

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author



Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Programa de Doctorado en Ingeniería de Proyectos: Medio ambiente,

Calidad, Seguridad y Comunicación

Tesis Doctoral

Metodología para la creación de conocimiento en los entornos

virtuales de enseñanza mediante herramientas de *Business*

Intelligence.

Caso de estudio en una institución educativa que emplea Moodle.

Por:

Kilian Tutusaus Pifarré.

Director:

Dr. Lázaro V. Cremades Oliver

Noviembre de 2015.

Resumen

Las organizaciones educativas buscan factores diferenciadores que las hagan más eficientes y eficaces mejorando de esta manera su competitividad en el mercado actual. El conocimiento que poseen sobre lo que está ocurriendo dentro de ellas supone un factor diferenciador que puede ser explotado mediante métodos y herramientas propios del *Business Intelligence*.

Muchas de las organizaciones educativas poseen entornos virtuales de aprendizaje que almacenan de forma automática cualquier interacción realizada por los usuarios. Estas interacciones se registran en sistemas digitales que no suele estar explotado debido, sobre todo, a las pocas funcionalidades y flexibilidad que disponen respecto al análisis y la personalización de indicadores.

En esta tesis se adaptan métodos del *Business intelligence* para captar el conocimiento residente en los datos almacenados en los entornos virtuales de aprendizaje según criterios de eficiencia y eficacia, poniendo el foco en cómo se transforman estos datos en conocimiento.

Como resultado de esta Tesis se contextualizan y aplican métodos de extracción de conocimiento en el marco del departamento de *e-learning* de una institución dedicada a impartir formación on-line. De esta forma, se definen las características propias de este tipo de instituciones y se aplica la experimentación empírica de la metodología propuesta estableciendo la viabilidad desde el punto de vista de la eficiencia y la eficacia.

Palabras Clave: Creación de conocimiento, Almacén de Datos, Moodle, Toma de Decisiones.

Abstract

Educational organisations search for differentiating factors in order to become more efficient and effective, so they can improve their competitiveness in the present market. Knowing what is happening within them is a differentiating factor that can be used by incorporating *Business intelligence* methods and tools.

Many educational organisations use learning virtual environments that automatically store data about any interaction of its users. These interactions are recorded in digital systems that are not usually used, mainly because of the few functionalities and flexibility of the analysis and customizing of the indicator.

This thesis adapts *Business intelligence* methods in order to compile a knowledge based in the data stored in the learning virtual environments by using efficiency and effectiveness criteria, it mainly focuses on how to transform the data into knowledge.

In its conclusion, this thesis contextualises and incorporates methods of knowledge compilation within the framework of the *e-learning* department of an online training institution. Thus, the characteristics of these kind of institutions are defined, the suggested methodology implements an empirical experimentation in order to establish the viability from an efficiency and effectiveness point of view.

Key words: Knowledge creation, data storage, Moodle, decision-making

Tabla de Contenidos

| | |
|--|-----------|
| Resumen | i |
| Abstract | ii |
| Tabla de Contenidos | iii |
| Lista de tablas | xiii |
| Lista de figuras | xix |
| Abreviaturas | xxiii |
| Capítulo 1 - Diseño metodológico..... | 1 |
| 1 Introducción | 1 |
| 2 Antecedentes | 3 |
| 3 Justificación de la Investigación | 8 |
| 4 Planteamiento del problema..... | 13 |
| 5 Hipótesis de la investigación | 17 |
| 6 Objetivos | 19 |
| 7 Alcance | 20 |
| 8 Entorno de Aplicación de la Propuesta..... | 22 |
| 9 Desarrollo y Estructura de la memoria | 25 |
| 10 Aporte a la sociedad del conocimiento | 26 |
| Capítulo 2 - Estado del arte | 29 |
| 1 Introducción | 29 |

| | | |
|------|---|-----|
| 2 | El conocimiento y su gestión | 29 |
| 2.1 | Introducción..... | 29 |
| 2.2 | Concepto de conocimiento | 30 |
| 2.3 | Naturaleza del conocimiento | 31 |
| 2.4 | Características del conocimiento | 31 |
| 2.5 | El conocimiento como recurso. | 38 |
| 2.6 | El conocimiento desde la perspectiva de las organizaciones..... | 39 |
| 2.7 | Relación entre conocimiento y valor agregado | 44 |
| 2.8 | Definiciones de gestión del conocimiento..... | 47 |
| 2.9 | La gestión del conocimiento en las organizaciones..... | 51 |
| 2.10 | Misiones y objetivos de la Gestión del Conocimiento | 55 |
| 2.11 | Componentes de la gestión del conocimiento | 63 |
| 2.12 | La gerencia del conocimiento..... | 66 |
| 2.13 | Procesos de la gestión del conocimiento | 68 |
| 2.14 | Modelos de gestión del conocimiento | 76 |
| 2.15 | Gestión de la información para generar conocimiento | 82 |
| 3 | <i>Business Intelligence</i> | 87 |
| 3.1 | Introducción..... | 87 |
| 3.2 | Definición de <i>Business Intelligence</i> | 87 |
| 3.3 | Utilidad del BI | 90 |
| 3.4 | Elementos de un BI..... | 98 |
| 3.5 | Fuentes de información..... | 101 |

| | | |
|-------|---|-----|
| 3.6 | ETL (<i>Extract, Transform and Load</i>) | 108 |
| 3.7 | <i>Data warehouse</i> : Almacenes de datos..... | 116 |
| 3.8 | Herramientas del BI para el uso de los datos | 128 |
| 3.9 | Principales herramientas del BI..... | 128 |
| 3.9.1 | Herramienta de análisis OLAP | 129 |
| 4 | <i>E-learning</i> , analítica del aprendizaje y su relación con el BI | 136 |
| 4.1 | Descripción de <i>e-learning</i> | 136 |
| 4.2 | Seguimiento en el <i>e-learning</i> | 138 |
| 4.3 | Los entornos virtuales de enseñanza y aprendizaje. | 139 |
| 4.3.1 | El entorno virtual de aprendizaje Moodle..... | 142 |
| 4.4 | Analítica del aprendizaje | 145 |
| 5 | Soluciones similares..... | 151 |
| 5.1 | ViMoodle..... | 151 |
| 5.2 | Extensión del navegador para análisis visual | 152 |
| 5.3 | PADA: <i>Dashboard of Learning Analytics of Dyslexia in Adults</i> | 153 |
| 5.4 | Cuadro de mando integral portable | 154 |
| 5.5 | MOCLog..... | 154 |
| 5.6 | GISMO | 155 |
| 5.7 | Otras soluciones..... | 155 |
| 6 | Herramientas de desarrollo | 155 |
| 6.1 | MySQL..... | 156 |
| 6.2 | Pentaho Data Integration | 156 |

| | | |
|-----|--------------------------------|-----|
| 6.3 | Pentaho Report Designer | 157 |
| 6.4 | Pentaho Schema Workbench | 157 |
| 6.5 | Pentaho Analyzer..... | 158 |
| 7 | Conclusiones | 159 |

Capítulo 3 - Propuesta de método para crear conocimiento a partir de los datos residentes en entornos virtuales de aprendizaje 161

| | | |
|-------|--|-----|
| 1 | Introducción | 161 |
| 2 | Esquema del método | 162 |
| 3 | Descripción de las fases | 171 |
| 3.1 | Análisis de requerimientos | 171 |
| 3.2 | Elección de las fuentes de datos | 171 |
| 3.3 | Diseño de la base de datos (o almacén de datos)..... | 172 |
| 3.4 | Integración de datos | 173 |
| 3.5 | Uso de la solución..... | 176 |
| 4 | Detalle de los procesos de cada fase | 177 |
| 4.1 | Análisis de requerimientos | 177 |
| 4.1.1 | Consideraciones según la estrategia escogida para construir el DW | 178 |
| 4.1.2 | Actores y funciones..... | 180 |
| 4.1.3 | Acciones para analizar los requerimientos..... | 181 |
| 4.1.4 | Identificar la información necesaria para la toma de decisiones..... | 182 |

| | | |
|--|---|------------|
| 4.1.5 | Identificar los indicadores para medir el conocimiento | 182 |
| 4.1.6 | Identificar posibles fuentes de datos y validar su viabilidad ... | 188 |
| 4.1.7 | Realizar el modelo conceptual | 189 |
| 4.2 | Análisis de las fuentes de datos | 190 |
| 4.3 | Establecer las correspondencias entre fuentes de datos y el modelo conceptual | 191 |
| 4.4 | Elaborar el modelo lógico del almacén de datos | 192 |
| 4.5 | Integración de los datos | 194 |
| 4.6 | Uso de la solución | 197 |
| 5 | Enfoque del método desde el punto de vista de la gestión del conocimiento..... | 198 |
| 6 | Beneficios del método para los roles más significativos de un entorno virtual de aprendizaje | 203 |
| 7 | Enfoque desde el punto de vista de la eficiencia y la eficacia | 205 |
| 8 | Conclusiones..... | 208 |
| Capítulo 4 - Aplicación del método..... | | 209 |
| 1 | Introducción | 209 |
| 2 | Entorno sobre el que se aplica la propuesta | 209 |
| 2.1 | Sistemas informáticos..... | 209 |
| 2.2 | Asignatura de estudio | 210 |
| 2.3 | Actividades de evaluación..... | 211 |
| 2.4 | Habilidades de evaluación..... | 212 |

| | | |
|-------|---|------------|
| 3 | Fase de análisis de requerimientos | 214 |
| 3.1 | Identificar preguntas | 214 |
| 3.2 | Identificar indicadores, dimensiones y resolución..... | 216 |
| 3.3 | Realizar el modelo conceptual..... | 220 |
| 4 | Fase de análisis y elección de las fuentes de datos | 223 |
| 4.1 | Conformar las métricas..... | 228 |
| 4.2 | Establecer correspondencias entre las fuentes de datos y el modelo conceptual | 232 |
| 5 | Fase de elaboración del modelo lógico del almacén de datos..... | 234 |
| 6 | Fase de integración de los datos..... | 236 |
| 6.1 | Implementar transformaciones y trabajos..... | 237 |
| 6.1.1 | Trabajos | 238 |
| 6.1.2 | Transformaciones | 240 |
| 6.2 | Implementación de la integración de los datos..... | 245 |
| 7 | Fase de uso de la solución..... | 246 |
| 7.1 | Conformación de los cubos multidimensionales | 246 |
| 7.2 | Análisis OLAP..... | 247 |
| 7.3 | Reportes predefinidos y a la medida..... | 252 |
| 7.4 | Cuadros de mando integral | 253 |
| 8 | Conclusiones | 254 |
| | Capítulo 5 - Resultados | 255 |
| 1 | Introducción | 255 |

| | | |
|-----|--|-----|
| 2 | Resultados de la fase de análisis de requerimientos | 255 |
| 2.1 | Resultados de la identificación de preguntas clave | 255 |
| 2.2 | Resultados de la identificación de los indicadores y las dimensiones | 256 |
| 2.3 | Resultados de la identificación del nivel de granularidad o resolución | 257 |
| 2.4 | Resultados de la definición del modelo conceptual | 257 |
| 3 | Resultados de la fase de selección de las fuentes de datos | 258 |
| 3.1 | Resultados de las fuentes de datos seleccionadas..... | 258 |
| 3.2 | Resultados de la correspondencia entre las fuentes de datos y el modelo conceptual..... | 260 |
| 3.3 | Resultados de las métricas que se han conformado..... | 260 |
| 3.4 | Resultados del modelo conceptual ampliado | 262 |
| 4 | Resultados de la fase de diseño del almacén de datos | 262 |
| 4.1 | Resultados del diseño del modelo lógico multidimensional del almacén de datos, de las tablas de hechos, de las tablas de dimensiones y de sus uniones..... | 262 |
| 5 | Resultados de la fase de integración de datos | 263 |
| 6 | Resultados de la fase de uso de la solución | 264 |
| 7 | Solución a los problemas detectados | 267 |
| 8 | Cumplimiento de los objetivos | 269 |
| 9 | Eficiencia y eficacia del método | 270 |

| | | |
|---|---|------------|
| 9.1 | Comparación entre el estado previo y el estado final | 271 |
| 9.2 | Mejoras en la eficiencia y eficacia respecto a la situación inicial. | 273 |
| 10 | Cumplimiento de las hipótesis | 277 |
| Capítulo 6 - Conclusiones | | 281 |
| 1 | Utilidad del método..... | 281 |
| 2 | Factores de eficiencia y eficacia | 283 |
| 3 | Organización de los datos | 284 |
| Capítulo 7 - Propuestas de continuación de la tesis..... | | 287 |
| Lista de referencias..... | | 289 |
| Apéndice | | 305 |
| 1 | Resultados de los análisis OLAP | 305 |
| 1.1 | Accesos por día de la semana | 305 |
| 1.2 | Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores) | 306 |
| 1.3 | Accesos por hora del día | 307 |
| 1.4 | Cantidad de estudiantes que accedieron por módulo del sistema, teniendo en cuenta el grupo de repetidores..... | 309 |
| 1.5 | Cantidad de Accesos por miembro del grupo. Porcentaje relativo de acceso al grupo por cada miembro. | 310 |
| 1.6 | Cantidad de días que accedió un usuario en todo el semestre y la mediana de accesos por día..... | 312 |

| | | |
|------|--|-----|
| 1.7 | Total de accesos, cantidad de días que accedió, promedio de acceso por día, desviación estándar y coeficiente de variación. | 313 |
| 1.8 | Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación. | 314 |
| 1.9 | Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo. | 316 |
| 1.10 | Resultado promedio de cada habilidad y el porcentaje de evaluación de dicha habilidad..... | 318 |
| 2 | Reportes | 320 |
| 2.1 | Notas Finales de los estudiantes agrupados por grupo | 320 |
| 2.2 | Notas de las actividades de evaluación en clase de un grupo determinado | 320 |
| 2.3 | Notas de la Primera Prueba Parcial | 321 |
| 3 | Cuadro de Mandos Integrales..... | 321 |
| 3.1 | Panel de Identificación | 322 |
| 3.2 | Panel de accesos semanales al sistema | 322 |
| 3.3 | Panel de accesos por hora del día | 323 |
| 3.4 | Panel de áreas de accesos | 323 |
| 3.5 | Panel de módulos de acceso | 323 |
| 3.6 | Panel de porcentaje de evaluaciones realizadas | 324 |
| 3.7 | Panel de resultado en principales evaluaciones..... | 325 |
| 3.8 | Panel de resultado en habilidades evaluadas | 325 |

Lista de tablas

| | |
|---|------|
| Tabla 1 Lista de abreviaturas..... | xxiv |
| Tabla 2 Resumen de la aplicación de métodos y sistemas de Business Intelligence (BI) en el e-learning según tres perfiles de usuario muy comunes en la educación. 9 | |
| Tabla 3 Relación entre herramientas de BI y preguntas que se responden | 20 |
| Tabla 4 Gestión del conocimiento y su relación con los tipos de formación (on-line vs presencial)..... | 21 |
| Tabla 5 Comparación entre el conocimiento tácito y explícito..... | 31 |
| Tabla 6 Síntesis de los objetivos de la gestión del conocimiento y su priorización según especificidad (1= baja, 4= alta)..... | 63 |
| Tabla 7 Comparativa Modelos de Gestión Del Conocimiento..... | 81 |
| Tabla 8¿Cómo puede ayudar un BI a las organizaciones?..... | 94 |
| Tabla 9 Beneficios de implementar un sistema de BI (Whittemore, 2003) | 98 |
| Tabla 10 Problemas, beneficios y características dela calidad de los datos..... | 107 |
| Tabla 11 Resumen de los subprocesos ETL. Adaptado de Cano, 2007..... | 114 |
| Tabla 12 Descripción de las características de un data warehouse según Inmon (1996) | 120 |
| Tabla 13 Arquitecturas propuestas por TDWI para el desarrollo de un data warehouse..... | 125 |
| Tabla 14 Resumen de las herramientas del BI | 129 |

| | |
|--|-----|
| Tabla 15 Otras soluciones similares estudiadas | 155 |
| Tabla 16 Resumen del método propuesto según sus fases y los materiales, métodos, resultados, actores y utilidad de cada fase | 170 |
| Tabla 17 Calidad en la información | 175 |
| Tabla 18 Implementación del análisis de requerimientos según las estrategias para crear un DW..... | 178 |
| Tabla 19 Actores y funciones del análisis de requerimientos | 180 |
| Tabla 20 Propuesta de información deseable para definir el modelo conceptual | 183 |
| Tabla 21 Propuesta de desglose de resoluciones deseable para definir el modelo conceptual..... | 183 |
| Tabla 22 Dimensiones y resoluciones propuestas para el análisis de los datos de un entorno académico..... | 185 |
| Tabla 23 Propuesta de resoluciones para la dimensión actividad según el uso que se hace del recurso en un entorno académico..... | 187 |
| Tabla 24 Propuesta de resoluciones para la dimensión actividad según el nivel de accesibilidad que se tiene sobre el recurso académico | 187 |
| Tabla 25 Distintos enfoques de modelo lógico según el esquema de datos empleado..... | 194 |
| Tabla 26 Resumen de los subprocesos ETL. Adaptado de Cano, 2007 | 197 |
| Tabla 27 Actores de un entorno virtual de aprendizaje y beneficios del método de BI. | 205 |

| | |
|---|-----|
| Tabla 28 Factores que inciden positivamente en la eficiencia y eficacia del método | 205 |
| Tabla 29 Indicadores globales de eficiencia y eficacia propuestos..... | 206 |
| Tabla 30 Factores de eficiencia y eficacia propuestos según las fases del método | 207 |
| Tabla 31 Tipología de las actividades de evaluación evaluadas | 212 |
| Tabla 32 Preguntas genéricas identificadas en el análisis de requerimientos | 216 |
| Tabla 33 Hechos, métricas y dimensiones desglosadas según las preguntas que definen los requisitos de información | 219 |
| Tabla 34 Resoluciones para cada binomio hecho-dimensión | 220 |
| Tabla 35 Tabla mdl_log | 224 |
| Tabla 36 Tabla mdl_user | 224 |
| Tabla 37 Tabla mdl_role | 224 |
| Tabla 38 Tabla mdl_role_assignments..... | 224 |
| Tabla 39 Tabla mdl_course | 225 |
| Tabla 40 Tabla mdl_course_modules..... | 225 |
| Tabla 41 Tabla mdl_groups..... | 225 |
| Tabla 42 Tabla mdl_groups_members | 226 |
| Tabla 43 Tabla mdl_assign..... | 226 |
| Tabla 44 Tabla mdl_assign_grades | 226 |

| | |
|--|-----|
| Tabla 45 Tabla mdl_grade_outcomes | 227 |
| Tabla 46 Tabla mdl_grade_items | 227 |
| Tabla 47 Tabla mdl_grade_grades | 227 |
| Tabla 48 Identificación de las métricas en la aplicación del modelo conceptual. | 232 |
| Tabla 49 Correspondencias entre las fuentes de datos y el modelo conceptual ... | 234 |
| Tabla 50 Resumen de hechos, métrica y dimensiones identificadas | 257 |
| Tabla 51 Resumen de los resultados de la fase de uso de la solución | 264 |
| Tabla 52 Análisis más importantes realizados y conocimiento creado | 266 |
| Tabla 53 Herramientas OLAP utilizadas según el conocimiento requerido detectado..... | 267 |
| Tabla 54 Resultados de la resolución de los problemas detectados y su impacto en la eficiencia y eficacia globales..... | 268 |
| Tabla 55 Resultado del cumplimiento de cada objetivo planteado en esta tesis .. | 270 |
| Tabla 56 Comparación entre la situación inicial y final tras haber aplicado el método. | 273 |
| Tabla 57 Mejoras en la eficiencia y eficacia respecto a la situación inicial. | 274 |
| Tabla 58 Resultados del análisis de indicadores de eficiencia y eficacia..... | 276 |
| Tabla 59 Resultado en el cumplimiento de las hipótesis..... | 279 |
| Tabla 60 Accesos por día de la semana..... | 305 |

| | |
|---|-----|
| Tabla 61 Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores)..... | 307 |
| Tabla 62 Accesos por hora del día | 308 |
| Tabla 63 Cantidad de estudiantes que accedieron por módulo del sistema, teniendo en cuenta el grupo de repetidores. | 310 |
| Tabla 64 Cantidad de Accesos por miembro del grupo. Porcentaje relativo de acceso al grupo por cada miembro | 311 |
| Tabla 65 Cantidad de días que accedió un usuario en todo el semestre y la mediana de accesos por día..... | 313 |
| Tabla 66 Total de accesos, cantidad de días que accedió, promedio de acceso por día, desviación estándar y coeficiente de variación | 314 |
| Tabla 67 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación. | 315 |
| Tabla 68 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo (sólo una parte). | 317 |
| Tabla 69 Resultado promedio de cada habilidad y el porcentaje de evaluación de dicha habilidad. | 319 |
| Tabla 70 Notas Finales de los estudiantes agrupados por grupo..... | 320 |
| Tabla 71 Notas de las actividades de evaluación en clase de un grupo determinado | 321 |

Lista de figuras

| | |
|--|-----|
| Figura 1 Espiral del conocimiento. Fases por las que pasa el conocimiento; cada fase lo incrementa y perfecciona | 34 |
| Figura 2 Espiral de conversión del conocimiento organizativo (Nonaka & Takehuchi, 1995)..... | 35 |
| Figura 3 Doble espiral del conocimiento | 36 |
| Figura 4 Foco en el conocimiento como recurso | 39 |
| Figura 5 Escala del conocimiento según North (2001) | 46 |
| Figura 6 La gestión del conocimiento (Ordóñez de Pablos, 2011) | 50 |
| Figura 7: el círculo vicioso del Conocimiento | 55 |
| Figura 8 Procesos estratégicos de la gestión del conocimiento (Probst, Raub, & Romhardt, 2001)..... | 69 |
| Figura 9 Modelo Bustelo y Amarilla..... | 86 |
| Figura 10 Business Intelligence y su relación con otros sistemas de información. Adaptado de Negash (2004) p. 179 | 89 |
| Figura 11 Arquitectura genérica de un almacén de datos | 99 |
| Figura 12 Tipos de fuentes de datos según su estructura y su relación con el BI (Estay Niculcar & Plaza Pimentel, 2012)..... | 103 |
| Figura 13 Puntos de control de calidad de datos. Adaptado de Beyer & Feinberg (2006) | 106 |

| | |
|--|-----|
| Figura 14 Características de un data warehouse según Inmon (1996) | 119 |
| Figura 15 ejemplo de un cubo OLAP..... | 131 |
| Figura 16 Ejemplo de cubo OLAP rotado | 132 |
| Figura 17 Ejemplo del proceso "dicing" en un cubo OLAP..... | 133 |
| Figura 18 Ejemplo de roll-up en un cubo OLAP..... | 134 |
| Figura 19 Ejemplo de drill-down en un cubo OLAP..... | 135 |
| Figura 20 Ejemplo de informes de actividad que proporciona Moodle | 144 |
| Figura 21 Reporte calificador de Moodle..... | 144 |
| Figura 22 Esquema del método propuesto | 163 |
| Figura 23 Ejemplo esquemático de un modelo conceptual | 190 |
| Figura 24 Ejemplo esquemático de un modelo conceptual ampliado | 192 |
| Figura 25 Representación gráfica de una tabla de dimensión | 193 |
| Figura 26 Relación entre el modelo lógico y las características de contexto y categorización que definen a la información..... | 201 |
| Figura 27 Conversión de datos en información según el método propuesto..... | 202 |
| Figura 28 Curso virtual de la asignatura | 211 |
| Figura 29 Ejemplo de habilidades de evaluación | 213 |
| Figura 30 Modelo conceptual para el hecho Acceso | 221 |
| Figura 31 Modelo conceptual para el hecho Evaluación..... | 221 |

| | |
|---|-----|
| Figura 32 Modelo conceptual para el hecho Evaluación de habilidades..... | 222 |
| Figura 33 Modelo lógico del hecho Acceso..... | 235 |
| Figura 34 Modelo lógico del hecho Evaluación..... | 235 |
| Figura 35 Modelo lógico del hecho Evaluación de habilidades..... | 236 |
| Figura 36 Arquitectura de integración de la solución | 237 |
| Figura 37 Trabajo para poblar el almacén de datos..... | 239 |
| Figura 38 Trabajo para poblar el data mart Acceso | 240 |
| Figura 39 Ejemplos de transformaciones | 241 |
| Figura 40 Transformación para la dimensión Tiempo | 242 |
| Figura 41 Transformación para el hecho Acceso..... | 243 |
| Figura 42 Transformación para el hecho Evaluación..... | 244 |
| Figura 43 Transformación para el hecho EvaluacionHabilidad | 245 |
| Figura 44 Implementación de los cubos multidimensionales | 247 |
| Figura 45 Índice de Acceso por Grupo | 248 |
| Figura 46 Accesos al campus por Semanas y Grupos..... | 250 |
| Figura 47 Realización de las evaluaciones por Tipo..... | 251 |
| Figura 48 Realización de las tareas agrupadas por temas de asignatura | 252 |
| Figura 49 Accesos por día de la semana | 306 |

| | |
|--|-----|
| Figura 50 Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores) | 307 |
| Figura 51 Accesos por hora del día | 309 |
| Figura 52 Cantidad de estudiantes que accedieron por módulo del sistema, teniendo en cuenta el grupo de repetidores. | 310 |
| Figura 53 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación. | 316 |
| Figura 54 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo (sólo una parte). | 317 |
| Figura 55 Resultado promedio de cada habilidad y el porcentaje de evaluación de dicha habilidad..... | 320 |
| Figura 56 Notas de la Primera Prueba Parcial..... | 321 |
| Figura 57 Ejemplo de panel de identificación..... | 322 |
| Figura 58 Ejemplo de panel de accesos semanales al sistema | 322 |
| Figura 59 Ejemplo de panel de accesos por hora del día..... | 323 |
| Figura 60 Ejemplo del panel de accesos por áreas | 323 |
| Figura 61 Ejemplo del panel de módulos de acceso..... | 324 |
| Figura 62 Panel de porcentaje de evaluaciones realizadas | 324 |
| Figura 63 Panel de resultado en principales evaluaciones | 325 |
| Figura 64 Panel de resultado en habilidades evaluadas..... | 325 |

Abreviaturas

| Abreviatura | Descripción |
|-------------|--|
| BBDD | acrónimo de Base de Datos. |
| BI | del inglés <i>Business intelligence</i> o inteligencia empresarial. |
| CKO | del inglés <i>Chief Knowledge Officer</i> o gerente del conocimiento. |
| CSV | del inglés <i>comma-separated values</i> , o valores separados por coma. Son un tipo de documento para representar datos en forma de tabla, en las que las columnas se separan por comas y las filas por saltos de línea. |
| DM | del inglés <i>Data mart</i> o almacén de datos departamental |
| DOC | del inglés <i>document</i> . Tipo de documento propio del Microsoft Word que se ha convertido en estándar de facto con el que pueden transferirse textos con formato o sin formato. |
| DSS | del inglés <i>Decision Support System</i> o sistema de soporte a las decisiones. |
| DW | del inglés <i>Data warehouse</i> o almacén de datos |
| ETL | del inglés <i>Extract, Transform and Load</i> o extraer, transformar y cargar. Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio. |
| FASMI | Del inglés <i>Fast, Analysis, Shared Multidimensional Information</i> o Información multidimensional, compartida para un análisis rápido. |
| HOLAP | del inglés <i>Hybrid OLAP</i> o OLAP implementado en una base de datos relacional y multidimensional simultáneamente. |
| MDDB | del inglés <i>Multi Dimensional Data Base</i> o base de datos multidimensional. |

| Abreviatura | Descripción |
|-------------|---|
| MDX | del inglés <i>Multi Dimensional eXpression</i> o expresiones multidimensionales. Es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP, se utiliza en Business Intelligence para generar reportes para la toma de decisiones basados en datos históricos, con la posibilidad de cambiar la estructura o rotación del cubo. |
| MOLAP | del inglés <i>Multidimensional OLAP</i> o OLAP implementado en una base de datos multidimensional. |
| OLAP | del inglés <i>On-Line Analytical Processing</i> o procesamiento analítico o en línea |
| PDF | del inglés <i>Portable Document Format</i> o formato de documento portátil. Es un formato de almacenamiento para documentos digitales independiente de plataformas de software o hardware. |
| ROLAP | del inglés <i>Relational OLAP</i> o OLAP implementado en una base de datos relacional. |
| SQL | del inglés <i>Structured Query Language</i> . Se trata de un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas |
| TDWI | del inglés <i>The data warehousing institute</i> . |
| TIC | acrónimo de Tecnologías de la Información y Comunicación |
| TXT | son un tipo de documento textual sin formatos añadidos. |
| XLS | formato de archivo del Microsoft Excel que contienen hojas de cálculo que permiten efectuar operaciones aritméticas y funciones con datos dispuestos en forma de tablas. |
| XML | del inglés <i>eXtensible Markup Language</i> o lenguaje de marcas extensible. es un lenguaje de marcas desarrollado por el <i>World Wide Web Consortium (W3C)</i> utilizado para almacenar datos en forma legible. |

Tabla 1 Lista de abreviaturas

Capítulo 1 - Diseño metodológico

1 Introducción

Hoy en día el mercado es cada vez más competitivo impulsando a las organizaciones a buscar factores diferenciadores que les ayuden a ganar posiciones según criterios de eficiencia y eficacia de sus procesos. Algunos de estos factores diferenciadores se hallan en el conocimiento que poseen dichas organizaciones respecto a lo que está ocurriendo dentro y fuera de ella.

Esto también es válido para las organizaciones educativas que emplean métodos virtuales de enseñanza/aprendizaje donde los ratios de abandono suelen ser muy elevados, impactando negativamente en las finanzas de estas organizaciones. Aunque esta tesis no pretende abordar el impacto económico de estos ratios de abandono es destacable indicar que existen factores que tratan de explicar estos ratios de abandono, algunos de ellos impredecibles y personales, otros en cambio causados por el desconocimiento del estado en el que se encuentran los cursos, los profesores y los alumnos durante la impartición de un curso según la modalidad *e-learning*.

En este sentido las organizaciones educativas están buscando crear mayor valor para sus alumnos, por ejemplo mediante acciones personalizadas proactivas

fruto del conocimiento que disponen de su estado, y de esta manera conseguir retenerlos e incluso fidelizarlos. Dados los esfuerzos que las organizaciones invierten en la búsqueda de ese mayor valor, se debería encaminar hacia una mayor eficiencia y eficacia en el proceso de creación de conocimiento útil para la toma de decisiones.

Una organización inteligente requiere que sus integrantes, responsables del éxito, promuevan una correcta gestión del conocimiento que les permita entre otras funcionalidades poder medir el estado de los indicadores que considere críticos para su desempeño, tales como análisis de los datos e información y que esto a su vez desemboque en una toma de decisiones estratégica y consciente, donde se destierre la improvisación y la subjetividad.

El camino que va desde los datos hasta la toma de decisiones atraviesa diversas etapas donde se utilizan diferentes métodos y tecnologías para agregar valor al proceso. En esta tesis se propone una metodología válida para crear el conocimiento residente en los datos de los entornos virtuales de aprendizaje poniendo el foco en cómo se transforman estos datos en conocimiento. En este sentido se analizan por un lado los procesos de creación de conocimiento desde el punto de vista teórico y se adaptan y aplican métodos y herramientas propias del *Business intelligence* para lograr el objetivo final de extraer conocimiento útil para la toma de decisiones.

Como resultado de esta Tesis se contextualizan y aplican métodos de extracción de conocimiento en el marco del departamento de *e-learning* de una

institución dedicada a impartir formación on-line. De esta forma, se definen las características propias de este tipo de instituciones y se aplica la experimentación empírica de la metodología propuesta estableciendo la viabilidad desde el punto de vista de la eficiencia y la eficacia.

Es importante destacar que el enfoque que se ha dado es de acuerdo a unas necesidades propias de un departamento de la institución educativa de estudio (procesos de accesos, evaluaciones y habilidades de los estudiantes propios del departamento de *e-learning*), es decir, no se ha considerado toda la organización ni todos los procesos. La metodología propuesta debe ser incremental y modular de forma que a lo largo del tiempo se incluyan más procesos y departamentos creando visiones del estado de la institución más globales y no tan específicas.

2 Antecedentes

Esta tesis aborda tres temáticas dos de ellas ampliamente estudiadas en el mundo científico, la gestión del conocimiento, la inteligencia de negocios y la tercera más nueva referente a la analítica en los entornos virtuales de aprendizaje como medio para impartir programas formativos según la metodología del *e-learning*.

A pesar de que existe numerosa literatura científica sobre las temáticas en las que se apoya esta tesis, existen pocos trabajos relacionados con el empleo de métodos y herramientas propios del *Business intelligence* aplicados a los entornos virtuales de aprendizaje desde la perspectiva de la gestión del conocimiento que

contextualicen este conocimiento en el marco de las dimensiones y métricas más comunes en las organizaciones educativas.

La Gestión del Conocimiento comienza a desarrollarse en la última década del siglo XX como respuesta no a las necesidades de información organizacionales, sino como respuesta a la necesidad de creación de un nuevo enfoque que ayude a las organizaciones a ser competitivas, innovadoras, creativas, inteligentes, aprender, a tener conocimientos.

La transformación de los datos en información y posteriormente en conocimiento es objeto de estudio de diversas disciplinas como Estadística, Inteligencia Artificial, Aprendizaje Automático, Bases de Datos, Inteligencia de Negocios, Recuperación de Información, Visualización, Gestión del Conocimiento, Reconocimiento de Patrones y Minería de Datos (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004)

En la literatura científica se pueden hallar numerosos trabajos que abordan la forma cómo se crea y difunde el conocimiento dentro de las organizaciones Nonaka (1991) (1994), Hedlund y Nonaka (1993), Hedlund (1994), Nonaka y Takeuchi (1995), Davenport y Prusak (1998). Estos y otros autores se citan en el capítulo correspondiente al marco teórico dentro del capítulo de gestión del conocimiento.

Por otro lado, el término *business intelligence* se popularizó en 1989 por Howard Dresner y junto con los aportes relativos al almacenamiento de datos propuestos por Ralph Kimball y William Inmon en la misma década, ha generado

una cantidad ingente de literatura científica que aborda las múltiples facetas del BI.

A efectos de esta tesis se ha prestado especial atención a los antecedentes relativos a la construcción y mantenimiento de los almacenes de datos por ser la base de cualquier proyecto de BI. En este sentido se han revisado trabajos de autores “gurú” tales como Kimball (*The datawarehouse Toolkit*, 1996) e Inmon (*Building the Data Warehouse*, 1996) y otros autores con publicaciones más recientes tales como Beyer & Feinberg (*The Basics of datawarehousing*, 2006), Eckerson & White (*Evaluating ETL and Data Integration Platforms*, 2003), Parr (*Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management*, 2000), Strange (*The Challenges of Implementing a datawarehouse to Achieve Business Agility*, 2001) y Watson (*Recent Developments in Data Warehousing*, 2002) entre otros. Adicionalmente esta tesis enfatiza la importancia de la calidad de los datos como medida de eficacia del sistema de BI, en este sentido se han revisado los siguientes antecedentes: Blumberg & Atre (*The Problem with Unstructured Data*, 2003), Bitterer (*Poor-Quality Data: The Sure Way to Lose Business and Attract Auditors*, 2006) y Eckerson (*Data Quality and the Bottom Line*, 2002)

Asimismo respecto a la gestión y beneficios de los sistemas de *Business Intelligence*, se han revisado trabajos de varios autores con publicaciones relativamente recientes tales como Eckerson & Howson (*Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale*,

2005), Fuchs (Aspects of ROI, 2003), Loshin (Business Intelligence: The Savvy Manager's Guide, Getting Onboard with Emerging IT, 2003), Negash (Business Intelligence, 2004), White (Now is the Right Time for the Real Time BI, 2004), Whittemore (The Business Intelligence ROI Challenge: Putting It All Together, 2003) e Inmon (Real-Time Decision Support Systems?, 2006), entre otros.

Otro elemento destacable afrontado en esta tesis relativo a los sistemas de *business intelligence* lo conforman las herramientas que se usan para explotar los datos. Aunque en esta tesis no se desarrolla la fase de uso de la solución sino que se centra más en la construcción de información a partir de datos contextualizados y categorizados se hace una revisión de los autores que han estudiado estas herramientas.

En este sentido los antecedentes se han encontrado en Eckerson y Howson (Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale, 2005), Romero y Ventura (Educational Data Mining: A Review of the State of the Art, 2010), Bouman & Van Dongen (Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQL, 2009), Hernández Orallo et al. (Introducción a la Minería de Datos, 2004), Hyde (Mondrian 2.2.2 Technical Guide, 2007), la comunidad virtual de Pentaho (Pentaho. Open source business intelligence, 2012), Rice (Business intelligence tools can help turn out savings in core cost areas, 2004) y Vercellis (Business Intelligence. Data Mining and Optimization for Decision Making, 2009), entre otros.

Finalmente se han revisado trabajos relacionados con la analítica del aprendizaje, término que apareció recientemente y que ha experimentado un auge en lo que se refiere a publicaciones en los últimos años. En este sentido numerosos autores han contribuido a este fin tales como Siemens (Open Learning Analytics: an integrated & modularized platform., 2011), Abell (Development of professional knowledge in learning to teach elementary science, 1999), Avouris et al. (Logging of fingertip actions is not enough for analysis of learning, 2005), Claros y Cobos (Pautas para la implementación de Analíticas de Aprendizaje en Entornos Colaborativos Centrados en la Interacción Social, 2013), Fidalgo (Learning analytics. Qué, cómo y para qué., 2012), Leony et al. (Learning analytics in the LMS: Using browser extensions to embed visualizations into a Learning Management System, 2013), Mazza et al. (MOCLog – Monitoring Online Courses with log data, 2012), Mejia et al. (PADA: a Dashboard of Learning Analytics for University Students with Dyslexia , 2013), Salinas et al. (Análisis de elementos que intervienen en el proceso de enseñanza aprendizaje en un entorno virtual de formación: Propuesta de un modelo didáctico, 2007), Uribe et al. (Moodle learning management system as a tool for information, documentation, and knowledge management by research groups, 2007), Van Barneveld et al. (Analytics in Higher Education: Establishing a Common Language, 2012), Vozniuk et al. (Towards portable learning analytics dashboards, 2013) y Zorrilla y Álvarez (MATEP: Monitoring and Analysis Tool for E-learning Platforms, 2008), entre otros.

3 Justificación de la Investigación

Las organizaciones educativas cuyo negocio se apoya en herramientas TIC para el aprendizaje tales como los entornos virtuales de aprendizaje desarrollan su actividad en un marco tecnológico que registra de forma automática, cualquier interacción realizada dentro de los mismos. Estas interacciones son convertidas en una ingente cantidad de datos, actuales e históricos, que se almacenan en sistemas informáticos, tales como bases de datos, que crecen desmesuradamente con el tiempo y cuya explotación puede resultar compleja y costosa, haciéndose ineficiente e ineficaz.

En estos datos reside un conocimiento en potencia que puede ser extraído y transferido a las personas encargadas de tomar decisiones dentro de estas instituciones, tales como profesores, coordinadores, directores académicos, gestores y hasta los propios estudiantes, que también pueden tomar decisiones, por ejemplo según su estado en relación con el estado de sus compañeros de clase.

La tabla inferior resume algunos de los beneficios de explotar estos datos según distintos perfiles de usuario muy comunes en las instituciones de educación formativa.

| Tipo de usuario | Beneficios de explotar los datos desde el punto de vista del conocimiento que se genera | Acciones que se toman a partir del conocimiento generado. | Resultados esperados |
|--------------------------------|---|---|--|
| Profesor / tutor | Conocer el estado de sus estudiantes. Conocer el estado de las asignaturas que tutorizan. | Personalización de acciones centradas en estudiantes individuales Corregir / mejorar el estado del alumno y de las asignaturas | Estudiantes fidelizados. Disminución del abandono. Asignaturas actualizadas y mejoradas. |
| Coordinador/Director académico | Conocer el estado del programa de estudios Conocer el estado de los profesores | Corregir el estado del programa y de los profesores. | Programa mejorado Profesores alineados con la política de la dirección académica |
| Estudiante | Conocer el estado de sus estudios en relación a otros estudiantes. Conocer cómo mejorar o potenciar sus habilidades para prevenir el abandono. | Corregir el estado del estudiante. Motivar mediante inputs veraces y válidos. | Estudiante más motivado y conectado con la red de estudiantes de la institución. |

Tabla 2 Resumen de la aplicación de métodos y sistemas de Business Intelligence (BI) en el e-learning según tres perfiles de usuario muy comunes en la educación.

La tabla superior refleja algunos de los beneficios potenciales de la gestión de los datos existentes para generar conocimiento, lo que justificaría el empleo de métodos y herramientas del *Business intelligence* para las organizaciones educativas que emplean TIC's. Más al detalle y atendiendo a los beneficios presentados por Fuchs (2003) y Whittemore (2003), referentes a la

implementación de estos sistemas en cualquier organización, esta justificación se hace más evidente:

- Beneficios tangibles, por ejemplo: reducción de costes, generación de ingresos, reducción de tiempos para las distintas actividades del negocio.
- Beneficios intangibles: el hecho de que tengamos disponible la información para la toma de decisiones hará que más usuarios utilicen dicha información para tomar decisiones y mejorar la posición competitiva de la organización.
- Beneficios estratégicos: Todos aquellos que facilitan la formulación de la estrategia, es decir, a qué clientes, mercados o con qué productos dirigimos.

Estay Niculcar & Plaza Pimentel (2012) a su vez indican que las ventajas que justificarían el empleo de un sistema de BI son:

- Mejorar la operativa del negocio
- Ganar cuota de mercado
- Mejorar el servicio al cliente y retener a aquellos que son más rentables
- Mejorar la eficacia de las campañas promocionales

Todas las anteriores justifican desde el punto de vista estratégico o competitivo el empleo de herramientas de *Business intelligence* para generar

conocimiento en las instituciones formativas que emplean entornos virtuales de aprendizaje.

Por otro lado, la justificación desde el punto de vista operativo del empleo de métodos y herramientas de BI en las instituciones educativas con apoyo de entornos virtuales de aprendizaje, queda resumida en los siguientes puntos:

- Mejora de la calidad de los datos considerados como fuentes del proceso de creación de nuevo conocimiento.
- Aumento de la eficiencia y eficacia en el proceso de generar y transmitir conocimiento.
- Aumento del uso de la información con el consecuente impacto positivo en la generación de nuevo conocimiento.

Más al detalle las ventajas operativas que justifican la implantación del método presentado en esta tesis son:

- Minimización de errores por procesos manuales de extracción y análisis de datos.
- Reducción del tiempo invertido en la captura de datos para el análisis.
- Reducción del tiempo invertido al no tener que repetir los procesos habituales.
- Reducción del tiempo invertido en la generación de análisis, tanto nuevos como predefinidos.

- Minimización de los problemas de rendimiento de los procesos transaccionales.
- Centralización de los entornos de información evitando el uso de entornos paralelos para distintos usuarios.
- Unificación y homogeneización de las fuentes de datos evitando distintas versiones de una misma realidad.
- Libertad de uso para los responsables de tomar decisiones evitando hacer partícipe al personal técnico especializado y potenciando la creación y difusión de nuevo conocimiento al situar a estos responsables en el centro de estos procesos.

Según lo que se desprende de lo anterior, esta tesis se justifica en el marco de la potenciación del capital intelectual, en la socialización del conocimiento y la generación de ventajas competitivas en las organizaciones educativas que usan entornos virtuales de aprendizaje. El hecho de acceder a esta información de forma eficiente aumenta la posibilidad de generar un conocimiento que ya reside potencialmente en los sistemas informáticos. Esto a su vez incide directamente sobre la competitividad de la institución debido a que el proceso de generar e intercambiar nuevo conocimiento impacta en la capacidad de competir dentro del marco de los mercados globalizados. En este sentido las organizaciones aumentan su competitividad aportando productos y servicios personalizados que respondan a las necesidades de los clientes, en este caso los alumnos.

No obstante no todas las organizaciones educativas son objeto directo de los beneficios presentados arriba. Esto se debe a la necesidad de disponer de antemano la información, almacenada en forma de datos dentro de los sistemas informáticos, que alimenten el sistema de BI. En este sentido las instituciones formativas que emplean entornos virtuales de aprendizaje son candidatas a poder emplearlo y maximizar los beneficios que aportan las herramientas y métodos del *Business intelligence* debido a que las fuentes de datos se auto-generan durante la actividad diaria de los usuarios, mientras que otras instituciones deben captarlas mediante otros métodos para poder generar beneficios al aplicar un sistema de BI. Por este motivo esta tesis se centra en las instituciones que emplean entornos virtuales de aprendizaje.

4 Planteamiento del problema

En la actualidad muchas organizaciones educativas cuyo negocio se apoya en herramientas TIC para el aprendizaje tales como los entornos virtuales de aprendizaje toman decisiones sustentadas en un proceso donde predomina el uso de la experiencia y la intuición, elementos que son importantes pero no suficientes.

Estas organizaciones disponen de muchos datos auto-generados los cuales no son aprovechados o en caso de serlo su aprovechamiento no es eficiente. En este sentido se puede encontrar organizaciones con las siguientes características respecto al uso que se dan a sus datos:

- Tienen datos y son malos.
- Tienen datos y pobre información.

- Tienen información y la usan poco.
- Tienen información y la usan, pero pudieran usarla de forma más eficiente.

Aun cuando estas instituciones estén usando los datos para generar conocimiento, este proceso suele ser realizado de forma poco eficiente, por ejemplo mediante el empleo de hojas de cálculo alimentadas por descargas puntuales de información en formato tabular desde los sistemas informáticos de aprendizaje. En este sentido, según Cano (2007) el empleo de otras herramientas, ajenas a las propias del BI, para la extracción de información necesaria para generar el conocimiento, tales como el empleo de hojas de cálculo, trae consigo los siguientes inconvenientes específicos:

- Posibilidad de cometer errores al introducir información proveniente de distintos sistemas.
- Inversión de gran cantidad de tiempo para la introducción de la información.
- Repetición del proceso para cada nuevo análisis
- Poca eficiencia en obtener detalles.
- Problemas de rendimiento en el entorno transaccional
- Se duplican los entornos de información de forma que cada uno de los usuarios de información querrá los informes con un diseño determinado y pueden aparecer “entornos paralelos” de información.

- Baja calidad de los datos con el riesgo de tener distintas versiones de una misma realidad:
- Desconocimiento por parte de los usuarios finales sobre dónde reside la información que necesitan.

Tal como se acaba de introducir, en los entornos virtuales de enseñanza donde se crean diariamente una cantidad ingente de registros de datos, uno para cada interacción que realizan los usuarios, estos datos son almacenados en bases de datos que no están preparadas para su consulta y filtrado, procesos necesarios para realizar búsquedas y acceso a la información, por lo tanto no están preparadas para generar y transferir el conocimiento que se esconde entre sus datos.

Este conocimiento en potencia, que puede ser extraído con métodos semiautomatizados, no se está aprovechando o en caso contrario, su aprovechamiento no es eficiente, impactando directamente sobre el rendimiento de los entornos virtuales haciéndolos inestables y aumentando el riesgo de privar de acceso a los usuarios del mismo. En este sentido muchas instituciones dedicadas a la formación o bien no generan conocimiento a partir de las interacciones de los usuarios o en caso de generarlo lo hacen de forma ineficiente y con el riesgo de colapsar los sistemas que forman parte del negocio principal de las instituciones dedicadas a la formación a distancia.

Los problemas específicos que explican por qué no se hace eficiente la generación de nuevo conocimiento son los siguientes:

- Baja disponibilidad de la información necesaria

- Distintas fuentes de datos no consolidadas
- Acceso a la información en términos de negocio
- Identificación de los factores críticos que afectan a los resultados futuros o pasados

Respecto a aumentar la eficiencia, según Zúmel (2008) el empleado europeo medio pierde una media de 67 minutos diariamente buscando información de la compañía, lo que equivale a un 15,9% de su jornada laboral, y según Rice (2004) los analistas financieros destinan un 80% del tiempo a agregar y normalizar manualmente información en hojas de cálculo Excel, y tan sólo un 20% a analizar la información.

Además el problema es mayor cuando se considera que en un entorno virtual de aprendizaje se genera un registro con muchos datos para cada acción del estudiante, con lo que la cantidad de datos que se tiene que procesar es tal que podría colapsar los sistemas de aprendizaje ante consultas complejas.

En este sentido los datos de los sistemas operacionales que realizan procesos transaccionales necesarios para mantenerlos operativos no están organizados de forma que se haga eficiente un análisis sobre los mismos. Los datos de estos sistemas están distribuidos de tal forma que se requiere de mucha pericia y recursos informáticos para obtener el conocimiento deseado, haciendo poco eficiente el proceso de extracción.

En el apartado 8 del presente capítulo, relativo al entorno de aplicación de esta tesis, se detalla el problema específico de la institución educativa donde se ha aplicado la metodología de *Business intelligence* desarrollada en este trabajo.

5 Hipótesis de la investigación

La hipótesis principal de esta tesis está sustentada en la siguiente afirmación:

“La aplicación de herramientas y métodos de Business Intelligence intensifica el uso de los datos que se generan en los entornos virtuales de aprendizaje aumentando la creación de conocimiento de forma eficiente y eficaz.”

Analizando la hipótesis anterior desde el punto de vista de lo que representa el conocimiento en las organizaciones, éste se puede definir como la información que posee valor para ella (Stewart, 1999), es decir aquella información que permite generar acciones asociadas a satisfacer las demandas del mercado (Porter & Millar, 1986), y apoyar las nuevas oportunidades a través de la explotación de las competencias centrales de la organización (Prahalad & Hamel, 1990).

Según lo anterior, si se puede generar información a partir de los datos que automáticamente se generan en los procesos transaccionales de los entornos virtuales de aprendizaje, esta información se orienta en términos de negocio o

utilidad y se puede transferir a las personas que toman decisiones, su uso implicaría la creación de nuevo conocimiento útil para la toma de decisiones.

Otras hipótesis, derivadas del análisis de la hipótesis anterior son las siguientes:

- Disponer de un sistema de BI hace que se exploten más los datos del entorno virtual de aprendizaje.
- Los datos explotados, cuando son sintetizados en términos de negocio o utilidad, se convierten en información que se transfiere eficientemente a los responsables de tomar decisiones.
- La información, cuando se consume por los responsables de tomar decisiones, se convierte en conocimiento de forma eficaz al estar sintetizada y orientada en términos de utilidad o de negocio.
- El conocimiento adquirido, resultante de la organización de los datos en términos de utilidad, mejora el proceso de toma de decisiones.
- La creación del conocimiento es eficaz en términos de disponer de una única verdad de los hechos para toda la institución, de evitar errores derivados del empleo de múltiples fuentes de datos y de almacenar los datos en términos de utilidad o negocio.
- La creación del conocimiento es eficiente en relación a otros métodos que se puedan emplear en términos de autonomía de los responsables de tomar decisiones, disminución de tiempos para

crear nuevo conocimiento, mejora del rendimiento de los sistemas informáticos OLTP y, captación de datos automatizado, entre otros.

- El conocimiento se difunde/comparte de forma más eficiente al estar la información centralizada en un único sistema informático accesible desde cualquier ordenador y al dar autonomía a los tomadores de decisiones para aplicar estrategias comunicativas automatizadas.

6 Objetivos

La presente investigación tiene como objetivo general demostrar que mediante métodos y herramientas propias del *Business intelligence* se intensifica el uso de los datos, almacenados en un entorno virtual de aprendizaje, para mejorar proceso de toma de decisiones por parte de los profesores.

Los objetivos específicos que ayudan a la consecución del objetivo general son:

- Caracterizar las especificidades de las organizaciones de formación a distancia dentro del marco de un proyecto de *Business Intelligence*.
- Adaptar las metodologías y herramientas del BI para generar un método que ayude a crear conocimiento en entornos virtuales de aprendizaje.

- Aumentar la disponibilidad de los datos para poder tomar decisiones según un conocimiento generado a través de su explotación.
- Extraer, limpiar y almacenar los datos procedentes de múltiples fuentes en un único repositorio.
- Organizar los datos dentro de un modelo multidimensional según criterios de utilidad.
- Proveer de un sistema centralizado donde se exploten los datos de forma autónoma por parte de los tomadores de decisiones.

7 Alcance

Los sistemas de *Business intelligence* proveen, a través de sus herramientas de análisis, muchas funcionalidades las cuales no se abordarán en esta tesis de forma íntegra. La tabla inferior recoge un resumen de las preguntas que resuelve y las herramientas de usuario final que utilizan este tipo de sistemas

| Conocimiento requerido | Herramientas de BI |
|------------------------|--|
| ¿Qué ha pasado? | Reportes (informes/ad-hoc/listados...) |
| ¿Qué pasa ahora? | Monitorización Cuadros de mando |
| ¿Qué pasará? | Minería de datos Simulaciones Redes neuronales |
| ¿Por qué ha pasado? | Análisis OLAP |

Tabla 3 Relación entre herramientas de BI y preguntas que se responden

En esta tesis se provee de un método que provee de respuestas a todas las preguntas de la tabla superior excepto a aquellas referentes a la predicción que utilizan herramientas de minería de datos, simulaciones o redes neuronales.

Asimismo, no todas las instituciones formativas son candidatas para aprovechar las bondades que este método pretende demostrar. En este sentido se podrían diferenciar tres tipos de instituciones educativas según el empleo de herramientas TIC para el aprendizaje:

| Tipo de formación | Descripción | Gestión del conocimiento |
|-------------------------------------|--|--|
| 100% on-line: | La acción formativa se desarrolla íntegramente en entornos informáticos que registran automáticamente cualquier acción realizada por sus usuarios. | Se facilitan la creación de nuevo conocimiento a partir de los datos que se registran automáticamente en durante la actividad académica de los usuarios |
| Formación 100% presencial | La acción formativa se desarrolla únicamente en sesiones presenciales | Al no capturarse de forma automática las interacciones que realizan los usuarios se deben registrar mediante métodos manuales los cuales pueden ser heterogéneos y estar distribuidos en islas de información. |
| Formación mixta (o <i>blended</i>) | La acción formativa se desarrolla mediante el empleo simultáneo de sesiones presenciales y herramientas TIC para el aprendizaje | Se dan las circunstancias referentes a la combinación entre la formación 100% presencial y on-line. |

Tabla 4 Gestión del conocimiento y su relación con los tipos de formación (on-line vs presencial)

A pesar de que se puede aplicar el método de BI a los tres tipos de formación expresados en la tabla superior, debido a las ventajas que supone

disponer de los datos almacenados de forma automática en bases de datos normalizadas, se aplicará el método propuesto a los aspectos formativos del entorno virtual de aprendizaje (no presenciales) dentro de un curso que utiliza una estrategia de formación mixta, también llamada *blended*.

Asimismo al aplicar el método no se evaluará a la institución formativa de forma íntegra sino que, para validar las hipótesis iniciales, se hace suficiente con acotar la aplicación del método al departamento académico y dar respuesta a tres hechos muy concretos: actividad, evaluaciones y habilidades evaluadas.

8 Entorno de Aplicación de la Propuesta

La propuesta se ha aplicado en un curso semipresencial impartido por la Fundación Universitaria Iberoamericana (FUNIBER) dentro del marco del departamento de *e-learning*.

Esta institución educativa emplea Moodle como entorno virtual de enseñanza/aprendizaje y dispone de un equipo de profesores encargados de dinamizar dicho entorno.

Respecto a los problemas que se han señalado en el apartado 4- Planteamiento del problema del presente capítulo esta institución históricamente ha explotado los datos del campus de forma manual usando plantillas Excel y atacando la base de datos del Moodle para sus consultas, con los consiguientes problemas asociados al rendimiento de los servidores, duplicidad de información, repetición del proceso ante cada nuevo análisis lo que desemboca en poca eficiencia y eficacia en el tratamiento de estos datos.

Más concretamente se tiene la siguiente situación de partida:

- El colectivo académico debía crear un conjunto de informes predefinidos que se repetían en el tiempo; ejemplos de lo anterior son los informes relacionados con los resultados en una determinada prueba parcial, con las calificaciones finales, con el registro de evaluaciones parciales, entre otros.
- Para la creación de estos informes se solicitaba al equipo técnico, que mantiene el entorno virtual de aprendizaje, la extracción de los datos en crudo que se volcaban a un archivo el cual era transferido a los interesados mediante el correo o a través de una carpeta compartida de la red privada de la institución.
- Como los sistemas informáticos que almacenan estos datos eran del tipo OLTP, y por lo tanto no están optimizados para ciertas consultas, suponía un riesgo crear un formulario para que los usuarios pudieran conseguir los resultados de los informes predefinidos de forma autónoma. Entre otros motivos el principal riesgo está asociado al rendimiento de los servidores ante tales consultas, lo cual pone en riesgo la integridad del servicio que se da al cliente (estudiantes). En este hubo ocasiones en las que se sobrecargó el servidor de producción debido a consultas concurrentes complejas y los estudiantes percibían una

ralentización del entorno virtual de aprendizaje y en el peor de los casos se quedaban sin acceso.

- Los profesores y otro personal interesado, tenían que hacer las agregaciones de forma manual, según su experiencia, mediante el empleo de aplicaciones de hojas de cálculo. La eficacia y eficiencia de este proceso están directamente relacionadas con la experiencia en el manejo de hojas de cálculo y la habilidad para interpretar los valores y resultados.
- En este proceso se invertía mucho tiempo y recursos. Casi el 80% del tiempo era dedicado a los procesos de conformar la información necesaria para generar el conocimiento deseado mientras que tan sólo el 20% se dedicaba a realizar el análisis de la información.
- Por otro lado, los interesados solicitaban al equipo técnico la elaboración de informes específicos a partir de los mismos datos que servían para generar los informes predefinidos o bien nuevos informes que requerían datos nuevos, incluso de la combinación de éstos con otras fuentes de datos externas al entorno virtual de aprendizaje.
- Respecto a la creación del conocimiento, venía limitada por los factores expresados arriba relativos al rendimiento de los servidores y la dependencia de terceros para crear o personalizar los informes. Asimismo existía un riesgo de generar distinto conocimiento para

un mismo hecho, debido entre otros motivos a la existencia de islas de información o de los procesos de análisis dependientes de la habilidad de los tomadores de decisiones en el manejo de las hojas de cálculo.

Todos los factores expresados arriba forman parte del entorno de aplicación de la propuesta y están planteados como problemas específicos que esta tesis pretende responder, además de los genéricos que se han planteado en el apartado 4- Planteamiento del problema del presente capítulo.

9 Desarrollo y Estructura de la memoria

La presente tesis se articula en forma de capítulos cada uno de los cuales recogen diferentes aspectos a ser tratados tal como se describe a continuación:

- Capítulo 1 – Diseño metodológico: se presentan los antecedentes, la justificación de la investigación, el planteamiento del problema, las hipótesis, objetivos, alcance, entorno de aplicación y el aporte de esta tesis a la sociedad del conocimiento.
- Capítulo 2 – Estado del arte y marco teórico: se aborda el conocimiento preexistente sobre el que se asienta esta tesis. En este sentido se hace especial énfasis en las temáticas de gestión del conocimiento y *Business Intelligence*, debido a que la metodología presentada tiene como objeto crear nuevo conocimiento o validar un conocimiento tácito preestablecido utilizando herramientas y métodos propios del BI

- Capítulo 3 – Propuesta de Método para crear conocimiento a partir de los datos residentes en entornos virtuales de aprendizaje de forma eficiente y eficaz: se adaptan métodos propios del BI para aplicarlos en organizaciones educativas que emplean entornos virtuales de aprendizaje con el objetivo de crear eficiente y eficazmente un conocimiento útil para la toma de decisiones.
- Capítulo 4 – Aplicación del método propuesto: se aplica el método diseñado en el capítulo anterior a una institución formativa que emplea Moodle como entorno virtual de aprendizaje y se obtienen resultados que servirán para inferir la utilidad de los métodos de BI en organizaciones formativas.
- Capítulo 5, 6 y 7 – Resultados, conclusiones y propuestas de continuación de la tesis: capítulos que recogen un resumen de los resultados de todos los capítulos anteriores haciendo especial énfasis en el análisis de los resultados obtenidos tras aplicar el método diseñado. Asimismo se aportan conclusiones y recomendaciones de continuación de la tesis, referentes a la implementación de sistemas de BI sobre entornos virtuales de aprendizaje.

10 Aporte a la sociedad del conocimiento

Esta tesis pretende aportar los siguientes puntos a la sociedad del conocimiento:

- Caracterizar aspectos relativos a las organizaciones formativas que inciden en los sistemas de *Business Intelligence*:
 - Propuesta de dimensiones y resoluciones para el análisis de los datos de un entorno de aprendizaje virtual.
 - Propuesta de resoluciones para la dimensión actividad según el uso que se hace del recurso en un entorno académico
 - Estudio de los beneficios del método para los roles más significativos de un entorno virtual de aprendizaje
- Proponer una metodología de BI personalizada para aplicarla en organizaciones educativas que empleen entornos virtuales de aprendizaje y aplicarla en la plataforma de aprendizaje virtual Moodle:
 - Análisis y elección de fuentes de datos provenientes de un entorno virtual de aprendizaje sobre la plataforma Moodle.
 - Ejemplos de transformaciones y trabajos ETL aplicadas en procesos de integración de datos de un entorno virtual de aprendizaje.
 - Ejemplos de análisis, reportes y cuadros de mando útiles para responder al conocimiento requerido por los profesores.
 - Ejemplos de acciones que se toman a partir del uso de las herramientas OLAP.
- Estudiar los métodos del BI desde las perspectivas de eficiencia y eficacia y de la gestión del conocimiento.

- Validar la utilidad de las herramientas y métodos del BI en las organizaciones educativas que emplean entornos virtuales de aprendizaje.

Capítulo 2 - Estado del arte

1 Introducción

En este capítulo se aborda el conocimiento preexistente sobre el que se asienta el siguiente capítulo de esta tesis. En este sentido se hace especial énfasis en las temáticas de gestión del conocimiento y *Business Intelligence*, debido a que la metodología presentada tiene como objeto crear nuevo conocimiento o validar un conocimiento tácito preexistente utilizando herramientas y métodos propios del BI. Asimismo un tercer apartado resume la literatura existente relativa a los entornos virtuales de aprendizaje y el análisis de los datos que almacenan.

2 El conocimiento y su gestión

2.1 Introducción

La sociedad del conocimiento tiene tres conceptos importantes: Aprendizaje Organizativo, Capital Intelectual y Gestión del Conocimiento, esta última en el proceso de obtención de resultados, contempla la planificación, la organización, la dirección y el control al igual que la capacidad y habilidades de las personas u organizaciones para transformar información en conocimiento, dando cabida al poder de innovación y a la creatividad, lo que impacta

directamente la toma de decisiones en nuevas propuestas o mejoramiento de las existentes.

Primeramente es necesario clarificar el significado del término conocimiento junto con sus características y la necesidad de gestionarlo para generar valor en las organizaciones. Posteriormente, analizaremos las herramientas para almacenar y tratar la información (repositorios y herramientas de Gestión de conocimiento), la infraestructura tecnológica sobre las que se asienta, los métodos de promoción de creación de conocimiento y los recursos necesarios para lograrlo.

2.2 Concepto de conocimiento

Pávez (2001) expone una definición globalmente aceptada:

“Conocimiento: Son las creencias cognitivas, confirmadas, experimentadas y contextualizadas del conocedor sobre el objeto, las cuales estarán condicionadas por el entorno, y serán potenciadas y sistematizadas por las capacidades del conocedor, las cuales establecen las bases para la acción objetiva y la generación de valor”

Conocimiento, en idioma español y de acuerdo con la Real Academia Española, se refiere a la acción y el efecto de conocer. El verbo conocer, por su parte, significa averiguar por medio de las facultades intelectuales, la naturaleza, cualidades y relaciones de las cosas. El concepto conocimiento puede abarcar dos niveles: la acción de conocer en lo cotidiano para lo cual no es necesario

esforzarse intelectualmente y, en el segundo plano, la acción de conocer donde se presenta el proceso racional de comprender las cosas. En español también existe otra palabra de significado similar: saber, que se refiere a tener noticia de una cosa; ser docto en alguna materia. Esta palabra se remite al término sabiduría que quiere decir conocimiento profundo en las ciencias, las artes o las letras. (Saavedra Fernandez)

2.3 Naturaleza del conocimiento

Una de las formas para clasificar y entender la naturaleza del conocimiento, es dividirlo entre conocimiento tácito y explícito, cada uno con unas características particulares (Nonaka & Takehuchi, 1995):

| Conocimiento Tácito (Subjetivo) | Conocimiento Explícito (Objetivo) |
|--|---|
| Conocimiento de las experiencias (Cuerpo) Conocimiento simultaneo (Aquí y ahora) Conocimiento Análogo (Práctica) | Conocimiento del raciocinio (Mente) Conocimiento secuencial (Allí y entonces) Conocimiento digital (Teoría) |

Tabla 5 Comparación entre el conocimiento tácito y explícito.

2.4 Características del conocimiento

Existen múltiples visiones sobre las características del conocimiento, de hecho cada autor que estudia el conocimiento realiza su propia clasificación, y aún no se tiene un consenso universalmente aceptado sobre sus características, organización y como abordar su estudio.

Por un lado Backler (1995), define el conocimiento como polifacético y complejo, lo sitúa en un punto abstracto entre: implícito y explícito, distribuido e individual, físico y mental, en desarrollo y estático, verbal y codificado.

Davenport & Prusak (1998) lo resumen como una mezcla fluida de experiencia, valores, información contextual y perspicacia.

Samiotis & Poulymenakou (2003) analiza el conocimiento desde una perspectiva humana dado que refleja las intenciones de los seres humanos que lo crean y lo interpretan. Los últimos estudios le han dado énfasis al aspecto humano de transformar información en conocimiento.

Para Andreu & Sieber (2000), lo fundamental son básicamente tres características:

- El conocimiento es personal, en el sentido de que se origina y reside en las personas, que lo asimilan como resultado de su propia experiencia (es decir, de su propio “hacer”, ya sea físico o intelectual) y lo incorporan a su acervo personal estando “convencidas” de sus significado e implicaciones, articulándolo como un todo organizado que da estructura y significado a sus distintas “piezas”;
- Su utilización, que puede utilizarse sin que el conocimiento “se consuma” como ocurre con otros bienes físicos, permite “entender” los fenómenos que las personas perciben (cada una “a su manera”, de acuerdo precisamente con lo que su conocimiento implica en un

momento determinado), y también “evaluarlos”, en el sentido de juzgar la bondad o conveniencia de los mismos para cada una en cada momento; y

- Sirve de guía para la acción de las personas, en el sentido de decidir qué hacer en cada momento porque esa acción tiene en general por objetivo mejorar las consecuencias, para cada individuo, de los fenómenos percibidos (incluso cambiándolos si es posible).

El conocimiento cambia pero lo hace, normalmente, a través de una serie de fases. Estas fases están identificadas, en lo que se conoce, como ciclos de vida del conocimiento o denominados espiral de conocimiento; debido a que refleja mejor la idea de que no hay una fase inicial y otra final, no hay un principio ni un fin sino una iteración continua que cada vez que se produce genera más conocimiento.

Según los autores se distinguen las siguientes espirales de conocimiento:

- Espiral simple: Creación de conocimiento.
- Espiral compuesta: Conocimiento y organización.
- Doble espiral: Organización del conocimiento y uso

El proceso de creación del conocimiento para Nonaka y Takeuchi (1995) relativo a la espiral simple es un proceso de interacción entre conocimiento tácito y explícito que tiene naturaleza dinámica y continua. Se constituye en una espiral permanente de transformación interna de conocimiento, desarrollada siguiendo cuatro fases que muestran en la siguiente figura:

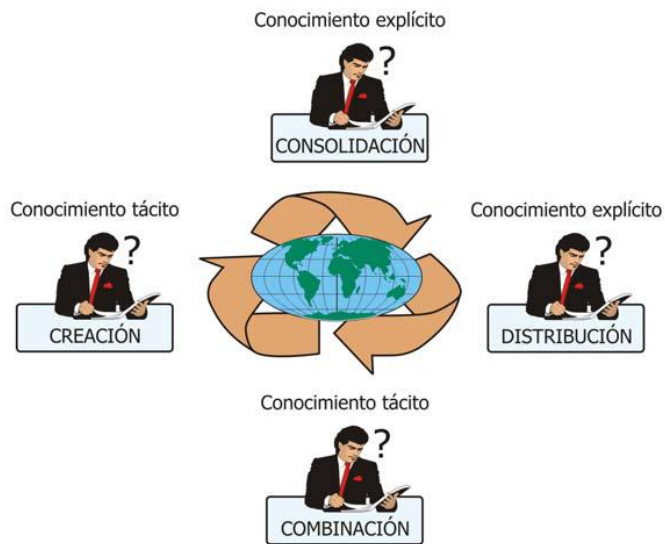


Figura 1 Espiral del conocimiento. Fases por las que pasa el conocimiento; cada fase lo incrementa y perfecciona

No obstante la espiral del conocimiento no es un proceso lineal y secuencial, sino exponencial y dinámico, que parte del elemento humano y de su necesidad de contrastar y validar sus ideas y premisas. Es lo que caracteriza la espiral compuesta, de esta forma, el individuo a través de la experiencia crea conocimiento tácito, el cual conceptualiza, convirtiéndolo en explícito individual. Al compartirlo, a través del diálogo continuo, con cualquiera de los agentes que intervienen en la organización se convierte en conocimiento explícito social. El siguiente paso consiste en internalizar las experiencias comunes, transformando el conocimiento explícito social en tácito individual.

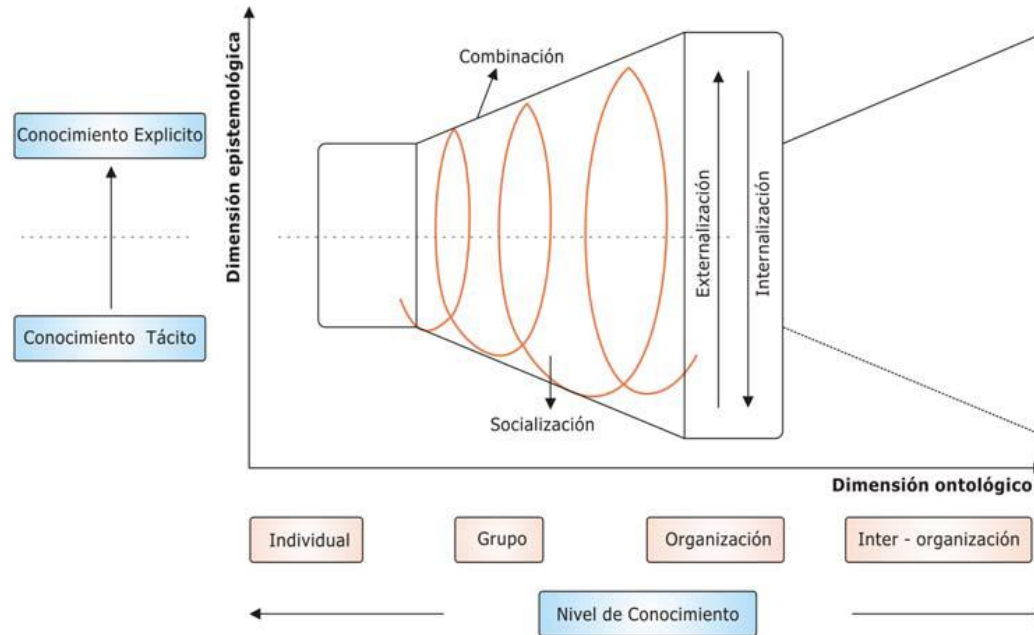


Figura 2 Espiral de conversión del conocimiento organizativo (Nonaka & Takeuchi, 1995)

El pilar sobre el que se fundamenta este proceso es el nuevo conocimiento, que se crea en términos de interrelación continua entre el de carácter tácito y explícito ya existente (en cualquier nivel de la organización) como consecuencia de un “proceso en espiral” ascendente, que se inicia a nivel individual, y posteriormente asciende al ámbito organizativo, alcanzando incluso el dominio inter organizativo. Este modelo ha sido propuesto por Nonaka (1991) (1994), Hedlund y Nonaka (1993), Hedlund (1994) y Nonaka y Takeuchi (1995), siendo éstos últimos los que culminaron su desarrollo, consolidándose por tanto como uno de los mecanismos de creación del conocimiento, junto al aprendizaje.

Otra visión del ciclo del conocimiento dentro de las organizaciones es el de la doble espiral. En este caso se presenta una espiral dentro de otra, coincidiendo algunas fases de ambas espirales

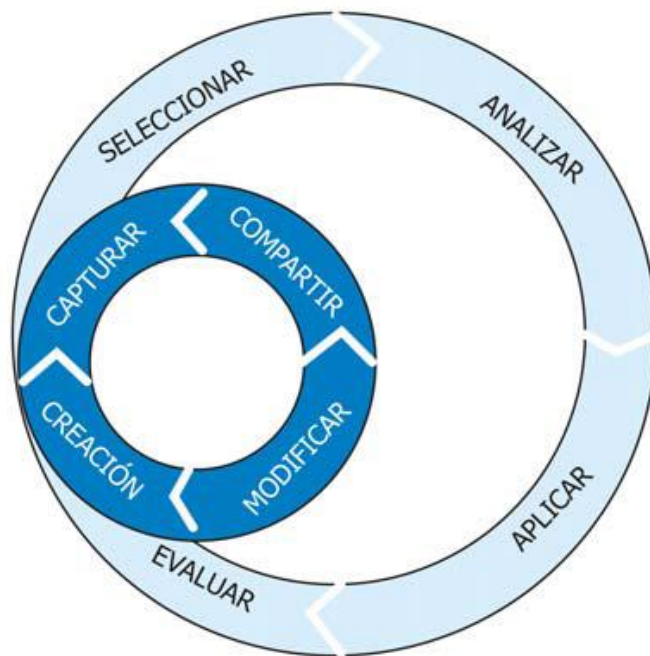


Figura 3 Doble espiral del conocimiento

La espiral interior representa el ciclo de creación y organización del propio conocimiento (espiral simple) y la espiral exterior representa la utilización del conocimiento dentro de la organización. Existen varios modelos que utilizan la doble espiral, uno de los primeros es el utilizado por PriceWaterhouse (1998):

- Espiral interna. Relativa al ciclo del conocimiento desde la organización del mismo, los procesos de esta espiral se basan en tareas propias de organizaciones de conocimiento, similar a los trabajos que se realizan en las bibliotecas. Esta espiral está dividida

en cuatro fases: creación de conocimiento (procesos y personas como generadores de conocimiento), captura (reflejar el conocimiento que tiene valor para la organización como explícito en soportes materiales), organización (clasificación y estructuración del conocimiento) y compartir (posibilitar canales de distribución fáciles y amigables para el conocimiento organizado en la fase anterior).

- Espiral externa. Relativa a la implantación de una estrategia de Gestión del Conocimiento dentro de las organizaciones que permita el acceso al mismo, su utilización y modificación. Las fases se refieren a la forma de utilizar el conocimiento dentro de las organizaciones. Esta espiral está dividida en cuatro fases: seleccionar, analizar, aplicar y evaluar. La primera fase se refiere al acceso a través de soportes de distribución de ese conocimiento por parte de los miembros de la organización y coincide con la fase de “crear” de la espiral interna. La segunda fase: “analizar”; el acceso lleva implícito la utilización de ese conocimiento para mejorar los procesos productivos o estratégicos y la valoración del uso del conocimiento y de la propia “calidad” del mismo; “aplicar” se basa en la traducción de ese conocimiento en acciones concretas (mejoras de procesos, nuevos rendimientos, ...); la fase de “evaluar” trata de actualizar y modificar los conocimientos, evidentemente

esta acción conduce a la fase de creación, por lo que se genera una nueva espiral.

2.5 El conocimiento como recurso.

Nonaka & Takehuchi (1995) investigaron de qué manera las compañías japonesas crean la dinámica de la Innovación. Para dicho estudio realizaron un análisis comparativo en el que descubren que las diferencias no se deben a sus procesos de manufactura, acceso a capital o relación cooperativa entre sus clientes, gobierno o proveedores. Tampoco al empleo vitalicio, u otras prácticas de recursos humanos. Que si bien son acciones importantes, no es lo medular, lo que si consideran distintivo es que las mismas se centran en la habilidad y pericia para la creación del conocimiento organizacional. Entendiendo este término como la capacidad de la compañía para crear conocimiento nuevo, difundirlo en la empresa e impregnarlo en sus productos, servicios y sistemas.

Las compañías japonesas, además están vinculadas estrechamente con su medio externo e interno, impulsan la formación de una base de conocimiento, la cual se utiliza en desarrollar nuevas tecnologías y productos. Este proceso se transforma en innovación, el cual a su vez, se transforma en una ventaja competitiva.

En síntesis, la diferencia consiste entonces en focalizar en la creación del conocimiento, no por el conocimiento en sí mismo, sino en su creación. Estas observaciones se reflejan en el siguiente esquema.

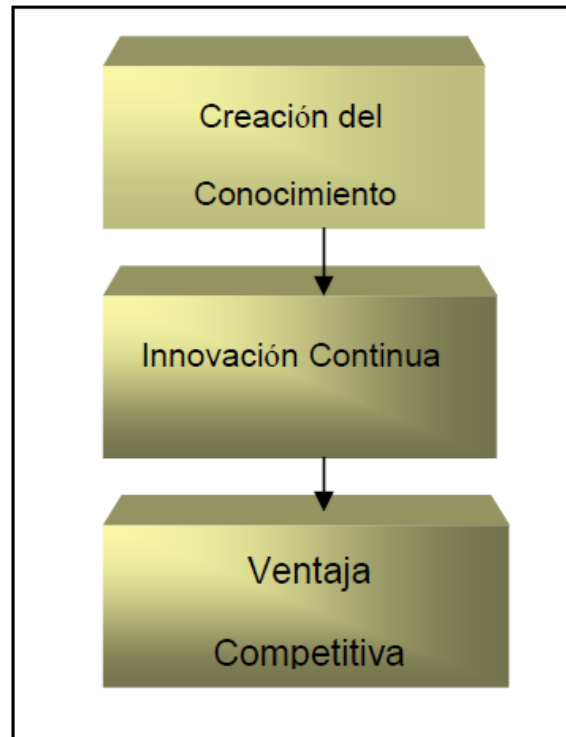


Figura 4 Foco en el conocimiento como recurso

2.6 El conocimiento desde la perspectiva de las organizaciones

La proliferación de entornos frecuentemente complejos e imprescindibles, sometidos a un estado de cambio permanente, acelerado e interdependiente, requiere que las organizaciones, como si de organismos vivos se tratasen, se adapten para sobrevivir.

Éstas deben ser ágiles, capaces de cuestionar su pasado y de hacer las cosas de forma diferenciada y, para ello, el conocimiento es la clave para mantener la actitud abierta hacia el cambio y hacia una mejora constante.

Tal y como afirma Nonaka (1991),

“En una economía donde lo único cierto es la inestabilidad, el conocimiento es una fuente segura de ventaja competitiva sostenible. Cuando los mercados cambian, las tecnologías proliferan, los competidores se multiplican y los productos pronto quedan obsoletos, las compañías con éxito son aquellas que crean nuevos conocimientos, los difunden rápidamente por toda la organización, y los aplican a nuevas tecnologías y productos.”

Según lo anterior las organizaciones no pueden confiar exclusivamente en el mantenimiento de sus conocimientos y habilidades actuales, sino que deben ser capaces de desarrollar y aplicar nuevos conocimientos, completando y conservando la capacidad de ajuste de sus características internas -productos, servicios, procesos productivos- a las transformaciones del entorno. Todo ello es el fundamento para la obtención de unos resultados superiores (Leonard-Barton, 1995)

Verdaderamente, resulta interesante observar que, en los últimos tiempos, se está generando una dinámica orientada a fomentar la capacidad de aprendizaje en la organización como uno de los elementos críticos para mantener a la organización en un estado de desarrollo y evolución permanentes. De hecho, la experiencia muestra claramente cómo aquellas organizaciones que no han desarrollado una capacidad de aprendizaje han visto reducida su capacidad competitiva. Incluso se ha reconocido que el aprendizaje es un comportamiento

inherente a toda organización, destinado a mejorar su capacidad de adaptación y anticipación a las exigencias del entorno, así como dotarla de una idiosincrasia propia y difícil de imitar fuera de su contexto. Duncan & Weiss (1979) y Nevis, A., DiBella., & Gould (1995)

Desde el punto de vista de las organizaciones, se puede definir el conocimiento como la información que posee valor para ella (Stewart, 1999), es decir aquella información que permite generar acciones asociadas a satisfacer las demandas del mercado (Porter & Millar, 1986), y apoyar las nuevas oportunidades a través de la explotación de las competencias centrales de la organización (Prahalad & Hamel, 1990). Es el activo para marcar la diferencia (Yadira & Magda León, 2001). El conocimiento organizacional también se define como lo que los integrantes de ella saben en su conjunto (Nonaka & Takehuchi, 1995). Esta visión establece que son las personas que integran la organización las que son las poseedoras del conocimiento, el cual articula el accionar de la organización y establece las bases para la “Memoria Organizacional” (Cross & Lloyd Baird, 2000).

Drucker (1993) plantea que es responsabilidad de la gerencia hacer productivo el conocimiento. El mismo autor dice que ha pasado de una situación referida al desarrollo intelectual de la persona a otro como factor de producción tomando como base su utilidad. Es por esto que la responsabilidad de la generación de políticas, estrategias y tácticas es función de la alta gerencia y la

organización debe involucrarse completa en el proceso de creación de conocimiento.

Tal y como apuntan algunos especialistas (Grant, 1991), (Schoemaker, 1992) proponen el uso adecuado de la información y, en especial, del conocimiento como la principal fuente de diferenciación en un mercado cada vez más competitivo y global, (Drucker, 1993) (Black & Synan, 1997), las organizaciones del futuro sólo podrán adquirir y mantener ventajas competitivas mediante el uso adecuado de la información y, sobre todo, del conocimiento. La aparición y creciente importancia del conocimiento como un nuevo factor de producción hace que el desarrollo de tecnologías, metodologías y estrategias para su medición, creación y difusión se convierta en una de las principales prioridades de las organizaciones en la sociedad del conocimiento. Así pues, si una organización desea ser competitiva de forma sostenida en el tiempo, ésta deberá identificar, crear, almacenar, transmitir y utilizar de forma eficiente el conocimiento individual y colectivo de sus trabajadores con el fin de resolver problemas, mejorar procesos o servicios y, sobre todo, aprovechar nuevas oportunidades de negocio.

Nonaka & Takehuchi (1995) establecen cuatro factores clave en torno a la creación de conocimiento organizacional:

- **Intención:** La organización debe tener la intención explícita de generar las condiciones óptimas que permitan el crecimiento de la espiral de conocimiento organizacional, apoyadas por el desarrollo de las

capacidades necesarias para llevar a cabo el proceso de Gestión del Conocimiento en torno a una visión compartida (Senge, 1990). Dentro de las intenciones se deben considerar los criterios necesarios para evaluar el valor y utilidad de los activos de conocimiento.

- **Autonomía:** La organización debe permitir algún nivel de autonomía en sus individuos (Senge, 1990), lo cual fomente las instancias de generación de nuevas ideas y visualización de nuevas oportunidades, motivando así a los participantes de la organización a generar nuevo conocimiento.
- **Fluctuación y caos creativo:** La organización debe estimular la interacción entre sus integrantes y el ambiente externo, donde los equipos enfrenten las rutinas, los hábitos y las limitaciones autoimpuestas con el objeto de estimular nuevas perspectivas de cómo hacer las cosas (Senge, 1990). El caos se genera naturalmente cuando la organización sufre una crisis o cuando los administradores deciden establecer nuevas metas.
- **Redundancia:** La organización debe permitir niveles de redundancia dentro de su operar. Esto genera que los diferentes puntos de vistas establecidos por las personas que conforman los equipos (Senge, 1990) permite compartir y combinar.

Finalmente, se puede concluir que el conocimiento es estratégico, es decir aquel que se diferencia del conocimiento base de toda la industria, puede ser

categorizado por la capacidad de apoyar una posición competitiva. Las categorías propuestas por Zack (Zack, 1998) son:

- **Conocimiento central:** Es el nivel de conocimiento de alcance mínimo requerido sólo para participar en el mercado. El poseer este nivel de conocimiento no asegura competitividad de largo plazo, pero la falta de él significará una base débil para la generación de barreras de entrada.
- **Conocimiento avanzado:** Es el nivel de conocimiento que le permite a la empresa ser competitiva. Esta categoría de conocimiento marca la diferencia en torno a la base de conocimiento en la cual está sostenida la industria.
- **Conocimiento innovador:** Es el nivel de conocimiento que le permite a la empresa liderar la industria y generar un nivel de diferenciación significativo como para ser sostenible en el tiempo. Este tipo de conocimiento podría generar un cambio de 'reglas del juego' en el contexto de la industria.

El conocimiento no es estático y lo que hoy se considera conocimiento innovador finalmente se convierte en el conocimiento central de mañana. Esto significa que para generar un crecimiento sostenido en torno a una posición competitiva es necesario un esquema de aprendizaje y adquisición del conocimiento continuo.

2.7 Relación entre conocimiento y valor agregado

Ponjuan (1998) presenta el enfoque de Valor Agregado establecido por el especialista norteamericano Robert Taylor. Este enfoque se fundamenta en la transferencia de información como respuesta intensiva a un proceso humano, tanto en las actividades formalizadas llamadas sistemas, como en el uso y usos de la información que son las salidas de estos sistemas.

El concepto de valor agregado se establece a partir de los procesos en torno a los elementos de la cadena informacional:

- Los datos (asociado a objetos) pueden, mediante procesos organizacionales, ser agrupados, clasificados, formateados, estructurados, etc. Cada uno de estos procesos le va agregando valor y los convierten en información.
- La información (asociada a un contexto), más experiencia, mediante procesos de análisis como: separación, evaluación, validación, comparación, entre otros, que le agregan valor la convierten en conocimiento informativo
- El conocimiento informativo (asociado a personas), más habilidades personales, modificado mediante procesos evaluativos que agregan valor como: opciones, ventajas y desventajas, pasa a constituir conocimiento productivo, denominado como inteligencia (Páez Urdaneta, 1992).
- El conocimiento productivo, debido a procesos decisionales de agregación de valor, como el planteamiento de metas, el compromiso,

la negociación o selección se transforma en un principio que conduce a la acción, lo cual nos puede llevar a un capital intelectual.

Una visión más detallada como la de North (2001) argumenta que nos falta una cuarta dimensión organizativa. Para convertirse en una organización basada en el conocimiento, los administradores y gestores tienen que aprender a manejar las herramientas y los conceptos de las nuevas estrategias asociadas a la gestión del conocimiento. De este modo, se explica claramente cómo se debe gestionar la información y el conocimiento, afirmando que la escala del conocimiento es la base para una empresa inteligente.

Para este autor la escala o pirámide del conocimiento es la siguiente:

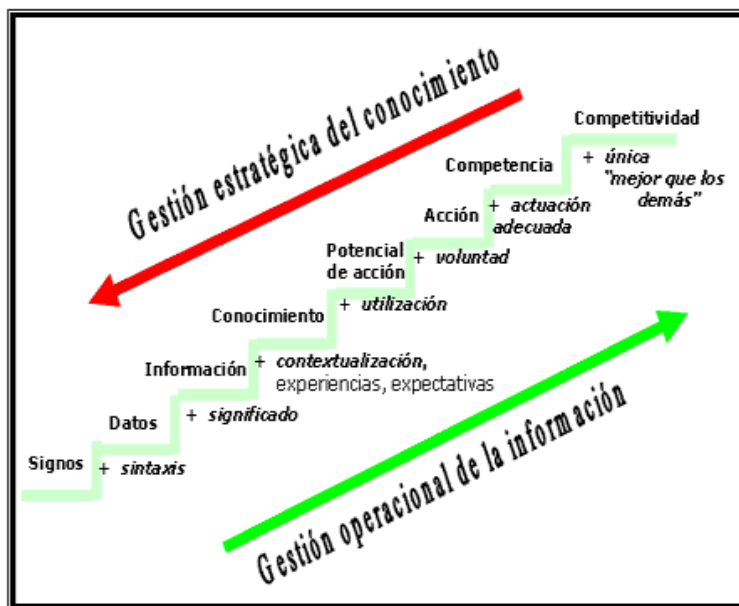


Figura 5 Escala del conocimiento según North (2001)

Parte de un concepto que ni siquiera algunos autores han tenido en cuenta y que es el de "signo": menos que un dato, una mera información sin sintaxis ni

significado. Cuando se sube un peldaño en la escala, se llega al dato con el único añadido de la sintaxis.

Un paso más importante es el de añadir significado a estos datos para convertirlos en información. Pero el paso definitivo en esta primera parte de la escala es, en nuestra opinión, el proveer de un contexto adecuado, unas determinadas expectativas y experiencias a dicha información y convertirla en conocimiento.

Aunque para llegar a la cumbre de la escala y conseguir una organización inteligente, que aprende y, lo que es más importante, se convierte en competitiva, se debe escalar el resto de peldaños (utilizar los conocimientos, tener voluntad y actuar de forma adecuada y ser únicos, originales y mejores que los demás, es decir, competitivos).

2.8 Definiciones de gestión del conocimiento

En la literatura sobre Gestión del Conocimiento nos encontramos múltiples definiciones de este concepto que recoge temas relacionados con la adaptación organizativa, supervivencia y competición en un entorno de cambios discontinuos

Según Jennex (2005) la gestión del conocimiento es:

“La práctica de forma selectiva de la aplicación de los conocimientos de las experiencias anteriores a la toma de decisiones de actividades actuales, con el propósito expreso de mejorar la eficacia de la organización”.

La Gestión del Conocimiento según Bueno (1998) es el conjunto de procesos que permiten utilizar el conocimiento como factor clave para añadir y generar valor.

Para Garvin (1994) incluye no solo los procesos de creación, adquisición y transferencia del conocimiento, sino el reflejo de ese nuevo conocimiento en el comportamiento de la organización.

Para Birkett (1995) supone traer a la superficie el conocimiento tácito, consolidarlo en formas mediante las cuales sea accesible ampliamente, y fomentar su continua creación.

Davenport (1994) considera que la Gestión del Conocimiento supone recoger, distribuir y utilizar conocimiento de forma eficiente.

Esencialmente comprende los procesos organizativos que buscan una combinación generadora de sinergias entre la capacidad de procesamiento de información y datos de las tecnologías de la información y la capacidad creativa e innovadora de las personas (Malhorta, 1998).

La gestión del conocimiento es un modo de pensar sobre gestión que incluye las experiencias pasadas (bibliotecas, bancos de datos, personas inteligentes) y creación de nuevos vehículos para intercambiar conocimiento (intranets, comunidades de prácticas, *networks* . . . etc.) (O'Dell & Grayson, 1998)

La Gestión del Conocimiento pretende poner al alcance de cada empleado la información que necesita en el momento preciso para que su actividad sea efectiva. En la actualidad, la tecnología permite entregar herramientas que apoyan

la gestión del conocimiento en las empresas, que apoyan la recolección, la transferencia, la seguridad y la administración sistemática de la información, junto con los sistemas diseñados para ayudar a hacer el mejor uso de ese conocimiento.

La gestión del conocimiento constituye un proceso integrador en el que convergen la gestión de la información, la tecnología y los recursos humanos y su implementación se orienta a perfeccionar los procesos de mayor impacto, mejor explotación del conocimiento en función de los procesos y su distribución en toda la organización, sobre la base del uso intensivo de las redes y las tecnologías.

Parece claro a la vista de las diferentes definiciones aportadas por varios autores que la Gestión del Conocimiento buscar mejorar la eficiencia en las organizaciones del siglo XXI y para ello debe ser capaz de (Gopal & Gagnon, 1995):

- Identificar las categorías de conocimiento necesario para apoyar la estrategia empresarial global.
- Evaluar del estado actual del conocimiento de la empresa, y
- Transformar la base de conocimiento actual en una nueva y poderosa base de conocimiento, rellenando las lagunas de conocimiento

Siguiendo a Tissen, Andriessen y Deprez (1998) se puede diferenciar entre gestión estratégica y gestión operativa del conocimiento. La gestión operativa utiliza las tecnologías de la información para organizar y distribuir la información

hacia y procedente de los empleados. La gestión estratégica es un proceso que relaciona el conocimiento de la empresa con:

- El diseño de estructuras organizativas que fomentan el conocimiento,
- La estrategia empresarial, y
- El desarrollo de profesionales del conocimiento.



Figura 6 La gestión del conocimiento (Ordóñez de Pablos, 2011)

2.9 La gestión del conocimiento en las organizaciones

Toda organización debe ser considerada como un sistema de acciones deliberadas, comprometido con su proceso de transformación y dirigido a la producción de determinados resultados es su entorno de actividad. Desde una perspectiva de la organización como sistema, su efectividad estará condicionada a la idoneidad de sus elecciones estratégicas a largo plazo, a la elección de sus procesos de transformación y a la adecuada coordinación de los mismos (Duncan & Weiss, 1979) (Shrivastava, 1983). Estas decisiones adoptadas en la organización y sus resultados, así como en el de las condiciones bajo las que tiene lugar esa relación, de manera que la efectividad de la organización está determinada por la calidad y singularidad de esos conocimientos. De acuerdo con ello el conjunto de saberes y conocimientos que se desarrollan en la organización son un elemento crítico para el despliegue coordinado de sus otros activos y capacidades (Cuervo, 1999), de tal forma que, globalmente, ayuden a la consecución de los objetivos y metas competitivas de la organización.

Las organizaciones se dieron cuenta de que sus activos físicos y financieros no tienen la capacidad de generar ventajas competitivas sostenibles en el tiempo, y descubren que los activos intangibles son los que aportan verdadero valor a las organizaciones. Son activos intangibles las capacidades que se generan en la organización cuando los recursos empiezan a trabajar en grupo, mucha gente en lugar de capacidades habla de procesos, o rutinas organizativas. Un activo intangible es todo aquello que una organización utiliza para crear valor, pero que

no contabiliza. Podemos enlazar con los conceptos desarrollados por la Teoría de Recursos y Capacidades. Esta teoría aparece en la década de los ochenta en el ámbito académico, y se puede considerar la precursora de la Gestión del Conocimiento, ya que se centra en analizar los recursos y las capacidades de las organizaciones como base para la formulación de su estrategia.

La Gestión del Conocimiento es, en definitiva, la gestión de los activos intangibles que generan valor para la organización. La mayoría de estos intangibles tienen que ver con procesos relacionados de una u otra forma con la captación, estructuración y transmisión de conocimiento. Por lo tanto, la Gestión del Conocimiento tiene en el aprendizaje organizacional su principal herramienta. La Gestión del Conocimiento es un concepto dinámico o de flujo.

El aprendizaje organizativo, la Gestión del Conocimiento y la Medición del Capital Intelectual son conceptos relacionados y complementarios. En pocas palabras, el aprendizaje organizativo es la base de una buena Gestión del Conocimiento, y la Gestión del Conocimiento es la base para la generación de Capital Intelectual y capacidades organizativas.

Se hace necesario enlazar bien las Competencias con Gestión del conocimiento y la Gestión del Conocimiento con todo lo que propones que es lograr la competencia Informacional. Recientemente, en el discurso sobre las competencias de los trabajadores en las organizaciones, se está abriendo un nuevo foco de atención: la competencia informacional. La competencia informacional proporciona las habilidades y los conocimientos necesarios para interactuar de

forma efectiva con la información; dichas habilidades han sido identificadas por distintos organismos e instituciones como parte de las competencias básicas en la cultura corporativa (Abell, 1999) (OECD, 1996). La importancia de la competencia informacional en las organizaciones ha originado la aparición de modelos y programas específicos para su formación (Skelton & Abell, 2001) (Goad, 2002). Así, las distintas maneras de interactuar con el mundo de la información convierten a dicha competencia en un elemento esencial para una implementación efectiva de programas de gestión del conocimiento. Nuestro trabajo tiene un doble objetivo: en primer lugar, describir los principales componentes de la competencia informacional en el puesto de trabajo, y, en segundo lugar, examinar en qué medida las competencias tecnológicas puede contribuir al éxito de la implementación de programas de gestión del conocimiento en las organizaciones.

A continuación se describen los factores por el cual es necesario implementar la gestión de conocimiento en una organización. (Macintosh, 1996)

- Los mercados son cada vez más competitivos y el ritmo de innovación se está incrementando.
- Las reducciones en dotación del personal, crea una necesidad de sustituir el conocimiento informal por métodos formales.
- Las presiones competitivas para reducir la fuerza laboral, la cual tiene valiosos conocimientos empresariales.

- Ha disminuido el tiempo disponible para adquirir experiencia y conocimiento
- El aumento de movilidad de la fuerza laboral conducen a la pérdida de conocimiento por parte de la organización.
- Cambios en la dirección estratégica puede resultar en la pérdida de los conocimientos en un área específica”.

La cantidad de Información disponible es ingente y su crecimiento exponencial. Las posibilidades de acceso a ella son cada vez mayores gracias a las TIC. Caracteriza a la sociedad de la información precisamente la capacidad de los ciudadanos para acceder a Información a través de las TIC. En consecuencia toda persona capaz de usar las TIC y obtener información a su través, puede considerarse integrada en la sociedad de la información.

El acceso a la Información es paso requerido para la obtención del conocimiento, pero no es suficiente. Se requiere procesarla y para ello se necesita capacidad y voluntad, es decir, saber hacer y esfuerzo. El resultado será la obtención de Conocimiento. Éste a su vez se convertirá en una nueva fuente de Información que nuevamente podrá dar lugar a Conocimiento adicional. Se genera así un Círculo Virtuoso del Conocimiento.



Figura 7: el círculo vicioso del Conocimiento

La generación de Conocimiento es clave para que la disponibilidad de Información revierta en un beneficio. Un estudiante puede tener a su disposición ingentes cantidades de información. Pero no llegará a ser un profesional cualificado, si no es capaz o no le interesa procesarla (comprendiéndola y asimilándola). Un ingeniero capaz de desarrollar una nueva tecnología, lo será en virtud del Conocimiento que habrá desarrollado a partir de la Información a la que habrá tenido acceso (COIT, 1996).

2.10 Misiones y objetivos de la Gestión del Conocimiento

Tres son las misiones principales de la gestión del conocimiento:

- Afianzar el conocimiento existente,
- Crear nuevos conocimientos, y
- Proporcionar las condiciones necesarias para que éste fluya por la organización.

Las propias definiciones de Gestión del Conocimiento se basan en estos aspectos:

“La Gestión del Conocimiento es la función por la que se planifica, coordina y controla los flujos de conocimiento que se producen en la empresa, en relación con sus actividades y con su entorno con el fin de crear competencias esenciales”

(Bueno, 1999)

El fin de los flujos mencionados por Bueno (1999) se basa en la repercusión de los mismos en la empresa y más concretamente en las competencias básicas siguientes:

- Capacidades para disponer, conocer y utilizar las tecnologías para producir o transformarlas en productos y servicios.
- Capacidades de los recursos humanos; experiencia, actitudes, habilidades y creatividad.
- Capacidades organizativas: procesos, conocimientos, gestión y definición de los valores de la organización.

Son varios los autores que definen la Gestión del Conocimiento como integración de los sistemas de información, procesos, personas, tecnologías y de los activos intangibles. La integración pasa primeramente por la organización y afianzamiento de los activos intangibles ya existentes; la organización es clave para evitar su pérdida y gestionarlos con mayor eficacia. Junto a esta idea de

organización y Gestión del Conocimiento también se asocia la característica dinámica del conocimiento (Davenport & Prusak, 1998), característica que hace referencia a la nueva creación del mismo. Esta idea hace que se consideren aspectos de aprendizaje, innovación y en general todos aquellos que permitan su creación.

La gestión del conocimiento tiene una misión y objetivos que deberían estar al alcance de cualquier organización (por muy pequeña que sea) y que no siempre están orientados a crear competitividad en el largo plazo. Por tanto podríamos enumerar otros alcances de la gestión del conocimiento (Gracia, 2006):

- Conocer el conjunto de activos intangibles que existen en la organización.
- Organizar los activos intangibles para una mejor gestión de los mismos.
- Reducir los tiempos de búsqueda del conocimiento generados en una organización.
- Integrar los distintos conocimientos que se generan en una organización.
- Cooperar en la creación del conocimiento.
- Compartir el conocimiento, adaptándolo a las necesidades temporales y estratégicas, actuales y futuras.
- Crear un patrimonio de conocimiento.

- Conseguir que el conocimiento que está en las personas esté en la organización.

Como vemos estos alcances podrían incluso servir para un sólo individuo, equipo de trabajo o pequeña organización.

Respecto a los objetivos específicos que persigue la gestión del conocimiento, varían según los autores consultados.

Según Davenport, De Long y Beers (1998) citado por Ordóñez de Pablos (2011) los objetivos de la gestión del conocimiento son:

- Crear almacenes de conocimiento, donde se recojan conocimiento de tres tipos:
 - conocimiento externo, como inteligencia competitiva,
 - conocimiento interno estructurado, como informes de investigación, y
 - conocimiento interno de tipo informal, con bases de datos sobre análisis de know-how, lecciones aprendidas, etc.
- Proporcionar acceso al conocimiento o facilitar su transferencia entre individuos, creando páginas amarillas que permiten la conexión, acceso y transferencia de conocimiento.
- Construir un entorno que permita la creación, transferencia y uso del conocimiento de forma más eficiente. La empresa puede estimular a los empleados para que participen en la base de conocimiento estructurado organizativo, retribuyéndoles en función

de sus contribuciones a la misma. La existencia de un “programa de auditoría de decisiones” posibilita que la empresa conozca si una determinada persona está aplicando este conocimiento en la toma de decisiones, y si la forma de aplicación es adecuada.

- Gestionar el conocimiento, realizar auditorías del capital intelectual organizativo e incluir los resultados en el Informe Anual, con un doble objetivo:
 - Que la organización se centre en cómo incrementar la utilización eficiente de los activos a lo largo del tiempo, y
 - Mostrar a los inversores del valor del conocimiento organizativo, y gestionar los activos intensivos en conocimiento de forma más eficiente, de modo que mejoren los resultados organizativos.

Según Koulopoulos y Frappaolo (2000) los objetivos son los siguientes:

- Vigilar el cambio y estimular la modificación constante –la innovación-, a una velocidad que, por lo menos, iguala la de la cambiante dinámica del mercado.
- No es controlar los pensamientos de la gente, sino centrarse en las formas que sirvan para dirigir la manera de usar el conocimiento y para construir sistemas y mecanismos que faciliten su expresión, y, con ello, compartir ideas y know-how.

- Desafiar la incapacidad de los sistemas de información tradicionales para codificar el conocimiento tácito.
- Compartir el conocimiento tácito.
- Mejorar los niveles de innovación a través de toda la organización (objetivo de la organización que aprende).
- Más allá de la gestión del conocimiento, el objetivo primario de los esfuerzos de conocimiento será la creatividad. .
- Reutilizar experiencias y procedimientos anteriores, con modificaciones para satisfacer las circunstancias actuales.

Para Lara (2001) los objetivos son:

- Acelerar el ritmo de aprendizaje de la organización.
- Olvidar las prácticas y hábitos que restan eficiencia.
- Posibilitar la viabilidad institucional de largo plazo al estimular el intercambio permanente con el entorno.
- Olvidar las prácticas y hábitos que restan eficiencia.
- Generar nuevas competencias individuales y organizacionales sostenibles evitando derrochar esfuerzos institucionales en otras direcciones.
- Incrementar el valor de los productos y servicios existentes mediante el empleo y reciclado más eficiente de los activos de conocimiento.

- Estimular la comunicació humana, desenvolupant la major quantitat possible de connexions mitjançant la creació d'una cultura amb horitzons compartits.
- Tener la capacitat de identificar y determinar qué información y conocimiento es importante para la organización.

Para Nonaka (1991) los objetivos son:

- Ubicar a la organización en una posición de competitividad basada en la innovación.
- Crear nuevo conocimiento.
- Desafiar la incapacidad de los sistemas de información tradicionales para codificar el conocimiento tácito y compartirlo.

Para Pávez (2001) los objetivos son:

- Crear sistemas y adaptar un entorno que sirvan para dirigir la manera de usar el conocimiento y que faciliten su expresión de tal manera que se pueda compartir y transferir de la manera más eficiente.
- Implantar estrategias orientadas al conocimiento.
- Ubicar a la organización en una posición de competitividad basada en la innovación.
- Medir logros y valorar económicamente el conocimiento.

- Incrementar el valor de los productos y servicios existentes mediante el empleo y reciclado más eficiente de los activos de conocimiento.
- Olvidar las prácticas y hábitos que restan eficiencia.

Gracia (2006) hace una síntesis de los objetivos de los autores citados anteriormente y los ordena según su generalización, evaluándolos en una escala del 1 al 4; el menor número indica mayor generalización y alcance y el mayor más especificidad. La tabla siguiente resume esta circunstancia.

| Objetivo | Nivel de Especificidad |
|--|-------------------------------|
| Reutilizar las experiencias y procedimientos anteriores para satisfacer las circunstancias actuales. | 1 |
| Generar nuevas competencias individuales y organizacionales sostenibles evitando derrochar esfuerzos institucionales en otras direcciones. | 1 |
| Estimular la comunicación humana, desarrollando la mayor cantidad posible de conexiones mediante la creación de una cultura con horizontes compartidos. | 1 |
| Medir logros y valorar económicamente el conocimiento. | 1 |
| Crear nuevo conocimiento. | 1 |
| Crear sistemas y adaptar un entorno que sirvan para dirigir la manera de usar el conocimiento y que faciliten su expresión de tal manera que se pueda compartir y transferir de la manera más eficiente. | 2 |
| Asumir una constante vigilancia del cambio y estimular la innovación. | 2 |
| Desafiar la incapacidad de los sistemas de información tradicionales para codificar el conocimiento tácito y compartirlo. | 2 |
| Mejorar los niveles de innovación a través de toda la organización. | 2 |
| Tener la capacidad de identificar y determinar qué información y conocimiento es importante para la organización. | 2 |
| Hacer que las personas desarrollen al máximo su creatividad. | 3 |

| Objetivo | Nivel de Especificidad |
|---|------------------------|
| Ubicar a la organización en una posición de competitividad basada en la innovación. | 3 |
| Acelerar el ritmo de aprendizaje de la organización. | 3 |
| Incrementar el valor de los productos y servicios existentes mediante el empleo y reciclado más eficiente de los activos de conocimiento. | 3 |
| Crear almacenes de conocimiento que guarden tres tipos de conocimiento: a) externo, b) interno estructurado e c) interno informal. | 4 |
| Ser un requisito de entrada para la competencia. | 4 |
| Olvidar las prácticas y hábitos que restan eficiencia. | 4 |
| Posibilitar la viabilidad institucional de largo plazo al estimular el intercambio permanente con el entorno. | 4 |
| Implantar estrategias orientadas al conocimiento. | 4 |

Tabla 6 Síntesis de los objetivos de la gestión del conocimiento y su priorización según especificidad (1= baja, 4= alta)

2.11 Componentes de la gestión del conocimiento

Con independencia del tipo de organización, de los objetivos estratégicos, misión y alcance de la Gestión del Conocimiento; existen una serie de componentes o dimensiones comunes (Gracia, 2006):

- Personas
- Procedimientos
- El ciclo del conocimiento
- Tecnologías
- La Organización

El conocimiento está en las personas, no en las organizaciones, ni en los sistemas de información, ni tan siquiera en los sistemas de Gestión del Conocimiento.

En este sentido el conocimiento que más suele interesar a las organizaciones es el referente a la experiencia; es clave que se pueda compartir esa experiencia (o parte de ella) por los distintos miembros de la organización para solucionar determinados problemas.

Por este motivo cualquier política de Gestión del Conocimiento debe tener en cuenta esta característica y no descuidar este factor, sin embargo, el tratamiento es distinto en función del tipo de organización:

- En grandes compañías y organizaciones, los sistemas de Gestión del Conocimiento se acompañan de incentivos para conseguir que las personas colaboren para crear conocimiento, para compartirlo y para transferirlo a la organización. Transferir conocimiento de las personas a la organización se denomina; paso de capital humano a estructural. Las personas que trabajan en este tipo de organizaciones son muy reacias a ceder el conocimiento a la organización.
- En pequeñas compañías, como empresas familiares, donde persona y empresa es una misma cosa, no hay ninguna resistencia a que el conocimiento quede en la organización; la motivación es que la empresa prospere o al menos, mantenga su ventaja competitiva.

Respecto a los procedimientos, éstos permiten que unos mismos recursos tangibles sean más rentables que otros, representan el know-how de la organización. Es clave para cualquier organización divulgarlos y que las personas los utilicen, modifiquen y mejoren. Los procedimientos de la organización siempre se utilizan para obtener una ventaja competitiva y optimizar la producción; normalmente los procedimientos se aplican diariamente en la organización y suelen ser de varios tipos (administrativos, organizativos, técnicos, operativos, desarrollo, gestión, etc.) pero los que menos incorporados y definidos están dentro de las organizaciones son todos los que están relacionados con el conocimiento, como creación, distribución, acceso, utilización, análisis, etc.

Respecto al ciclo del conocimiento, se debe tener en cuenta que es cambiante, la idea de ciclo lo refleja y además no sabemos ni cuando cambiará ni cómo. Por otra parte se debe tener en cuenta el resto de las características indicadas en el apartado 2.4: las fases, la cooperación para su creación, el uso lo perfecciona, etc.

Respecto a las tecnologías más adecuadas son las de la información (informática y comunicaciones). Estas tecnologías se clasifican en hardware: las redes y los ordenadores y el software (conjunto de programas de ordenador que gestionan tanto las comunicaciones como la información).

Finalmente, respecto a la organización, ésta debe disponer de la infraestructura necesaria (redes de ordenadores), una cultura organizacional (que favorezca la cooperación y divulgación del conocimiento), y necesita promover la

creación de uno nuevo (aprendizaje e innovación) así como capacitar a sus directivos para gestionarlo (dirección del conocimiento); es lo que se denomina organización basada en el conocimiento. Para ello los procesos que deben realizar las organizaciones deberían perseguir los siguientes objetivos:

- Hacer visible el conocimiento
- Incremento del conocimiento
- Suministrar la infraestructura
- Crear cultura del conocimiento

2.12 La gerencia del conocimiento

La gestión del conocimiento requiere de personal que desarrolle el rol estratégico de implementar y gestionar el conocimiento dentro de las organizaciones: el gerente de conocimientos o, en inglés, *Chief Knowledge Officer* (CKO). De hecho, la gerencia del conocimiento es el resultado de alto grado de consolidación de la cultura de la organización (Nieves & León, 2001)

En el estudio realizado por Michael Earl e Ian Scott (1999), se investigaron las características de este nuevo rol estratégico en las organizaciones que han adoptado expectativas de desarrollo del conocimiento dentro de ellas.

Descubrieron una serie de roles, tales como 'Director de capital intelectual', 'Vicepresidente de bienes intelectuales', 'Director de aprendizaje organizacional', 'Gerente de aprendizaje', entre muchos otros. Sin embargo, la finalidad objetiva de todos estos títulos apuntan en una sola dirección: el desarrollo del conocimiento como una fuente de ventajas competitivas sustentables.

Earl define al CKO como el encargado de "iniciar, impulsar y coordinar los programas de Gestión del conocimiento". Sin embargo, una definición tan sencilla puede llevar a confusiones tales como entender que los proyectos de Gestión del Conocimiento deben estar a cargo del Gerente TI (CIO) (Visión tecnológica) o del Gerente de RRHH (CHRO) (Visión organizacional).

La diferencia medular entre el CKO y el CIO en el objeto propio de Gestión: mientras que el CIO tiene como objetivo supervisar el despliegue de las TI, el CKO se centra en maximizar la creación, el descubrimiento y la diseminación de conocimientos en la organización (Management en Administración de la Información, 2000).

Sin duda, será necesario determinar si este nuevo puesto ejecutivo tiene fundamentos sostenibles para su implementación. Skyrme (1999) estableció una serie de situaciones en que el CKO será necesario. Algunas de ellas son:

- Maximizar el retorno de las inversiones en conocimiento, tales como nuevas contrataciones, procesos y capital intelectual.
- Explotar los activos intangibles, tales como el know-how, patentes y relación de clientes.
- Repetir los éxitos pasados y compartir mejores prácticas. Mejorar la innovación (Comercialización de ideas).
- Evitar la pérdida de conocimiento y las fugas producidas por las reestructuraciones organizacionales.

Un factor relevante descubierto por Earl (1999) fue la personalidad distintiva de los CKO. "Se destacaban por poseer un carácter vivaz, entusiasta y por la facilidad para transmitir su entusiasmo a los demás".

Algunas características del perfil de este tipo de profesionales son:

- Vivaces, entusiastas y capaz de transmitir su entusiasmo a los demás.
- Curiosos y reflexivos.
- Flexibles y abiertos a trabajar con cualquier persona.
- Abiertos a que otros asumieran el liderazgo y el reconocimiento de logros.
- Dispuestos a auspiciar proyectos.

Este tipo de perfil cuadra casi a la perfección con lo que Goleman (1999) define como un "Influenciador positivo", es decir, posee un manejo natural en su actuar emocional de conceptos tales como "influencia", "comunicación", "manejo de conflictos", "liderazgo" y "catalizador de cambio".

2.13 Procesos de la gestión del conocimiento

La Gestión del Conocimiento está compuesta por un grupo de procesos estratégicos que se producen en forma cíclica los cuales se enumeran a continuación (León, Ponjuán, & Rodríguez, 2004):

- Identificación del conocimiento
- Generación o Adquisición del conocimiento
- Desarrollo del Conocimiento

- Codificación o Transformación del Conocimiento
- Almacenamiento del Conocimiento
- Distribución o Compartir el Conocimiento
- Uso o Aplicación del Conocimiento
- Medición del Conocimiento

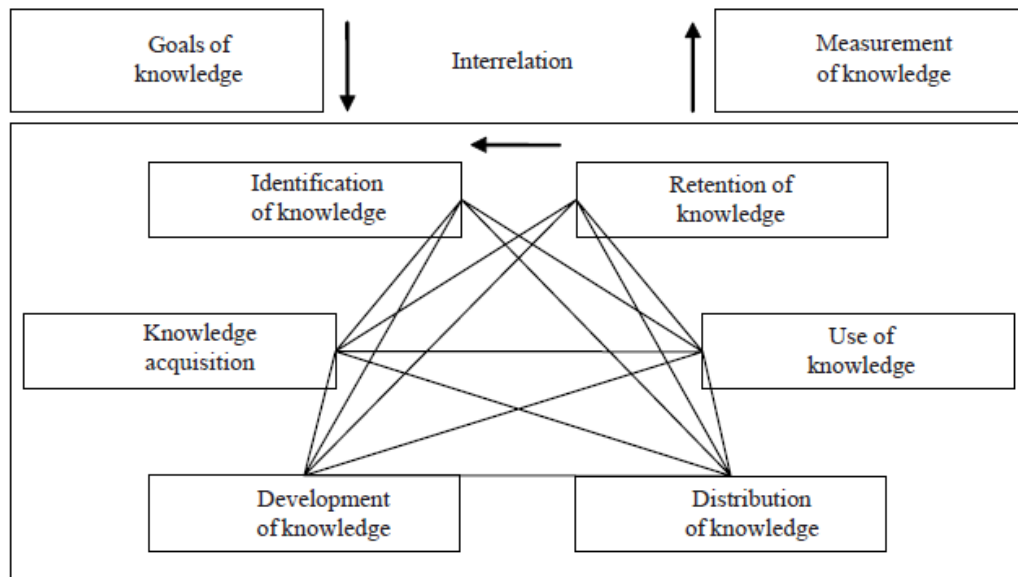


Figura 8 Procesos estratégicos de la gestión del conocimiento (Probst, Raub, & Romhardt, 2001)

La identificación del conocimiento organizacional permite determinar los vacíos de conocimiento que existen en la institución, las fuentes de conocimiento y las vías de intercambio y las reglas constituidas para ello.

A partir de estos resultados, puede determinarse si el activo se encuentra en la organización, en el entorno organizacional o no existe. Este diagnóstico es imprescindible, las organizaciones que gestionan conocimiento pueden enfrentar la

estrategia correcta en función de: anclar, adquirir o desarrollar respectivamente en cada caso. Este proceso se apoya en innumerables herramientas que permiten la eficiencia del proceso: los directorios, las páginas amarillas de expertos, los mapas de conocimiento, las topografías del conocimiento, tableros de comandos, evaluación de patentes, lecciones aprendidas, los mapas de activos del conocimiento, los mapas de fuentes del conocimientos, que se utilizan indistintamente en función de los objetivos propuestos, pero todos con resultados probados en diversos contextos.

Una vez identificado el conocimiento en la organización, este crece y se multiplica en la medida en que se utiliza. Esto exige a las organizaciones, que se encuentran en constante proceso de transformación, a trabajar intensamente para renovar su conocimiento. Es precisamente por eso, que la Gestión del Conocimiento no puede considerarse como un proceso aislado en la organización sino alineado con sus estrategias.

Este proceso consiste en la generación del conocimiento, lo cual se puede dar de diferentes formas, creación, adquisición, fusión, adaptación, etc. La creación de conocimiento es un proceso de desarrollo de las competencias y habilidades de los individuos que pertenecen a la organización, es un proceso donde se propicia el establecimiento de un ambiente que favorezca el surgimiento de nuevas ideas para fomentar la innovación y de esta forma, generar soluciones que contribuyan al progreso de los procesos, organización y sociedad en general.

También se puede hacer mediante la utilización de la tecnología en la creación de un mecanismo de conocimiento colectivo.

También es posible adquirir o capturar el conocimiento, obteniéndolo desde diferentes fuentes, pudiendo ser internas o externas. Igualmente y tomando en cuenta que el conocimiento se expresa por medio de la información y que esta debe registrarse en documentos que respalden el accionar de la organización, se apunta que todo sistema que gestiona conocimiento debe disponer para el desarrollo del proceso la adquisición o captura efectiva de los sistemas de información y de gestión documental. Algunos

ejemplos de creación de conocimiento son: mejores prácticas, *brainstorming*, mapas cognitivos, socios externos, benchmarking interno, externo, fusiones, imitaciones, replicación, adquisiciones, historias de éxito, talleres, bibliotecas, etc.

Si el conocimiento se halla en la organización este debe anclarse mediante estrategias que le permitan retenerlo y potenciarlo. En caso de que la organización carezca de un conocimiento específico necesario, debe buscarlo en su entorno para adquirirlo o simplemente desarrollarlo en su interior. En cualquier caso, el carácter de la fase de generación de conocimiento dependerá de los objetivos perseguidos.

Una vez generado o adquirido el conocimiento se transforma a un formato que hace posible su conservación y transferencia, esto es la codificación. No obstante, no todo el conocimiento es codificable, debido, en primer lugar a su dinámica, que no concede el tiempo y las vías inmediatas para codificarlo. En

ocasiones, la codificación se limita sólo a indicar quiénes son los expertos para que el interesado se remita a la fuente directamente.

Existen herramientas que resultan útiles para codificar cierta clase de conocimiento:

- Los diagramas de flujo o *workflow*.
- La creación de localizadores de expertos.
- Herramientas para el procesamiento que permitan indexar, filtrar, catalogar y clasificar al activo en cuestión.

Toda vez se ha codificado el conocimiento adquirido se procede a su almacenamiento. Este proceso consiste en sistematizar y almacenar el conocimiento para después poder distribuirlo al resto de la organización. Para las que tienen múltiples unidades es de tremenda utilizada, debido a que el adquirido en una puede ser de utilidad a las otras. Es importante destacar que se debe almacenar sólo aquel conocimiento que, en el momento necesario, se recordará su existencia y dónde se depositó. Todo el conocimiento que se va almacenando se conoce como memoria organizacional.

Igualmente que en el proceso de codificación, existen vías y herramientas que permiten el almacenamiento del conocimiento:

- Repositorio de conocimiento, bases de datos de contenido.
- Sistemas de gestión de documentos.
- Bases de datos de empleados (conocimientos, formación, experiencias, etc.).

- Páginas amarillas, *Data warehouse*, sistemas expertos.
- Literatura especializada.
- Manuales, entre otros.

El almacenamiento del conocimiento constituye un proceso esencial en la gestión del conocimiento. Si no es posible almacenar los conocimientos en la organización, se perderán los esfuerzos realizados en los procesos anteriores.

El almacenar también significa escribir la historia de la organización, su evolución, como una manera más de enfrentar los nuevos cambios y desafíos, que renovada y de manera constante, impone la sociedad moderna a sus instituciones.

El nuevo conocimiento organizacional sólo puede desarrollarse sobre la base del conocimiento previo. Ni los individuos ni las organizaciones borran sus experiencias anteriores con las nuevas, se apartan y no se utilizan en las circunstancias actuales, no obstante, permanecen como una opción. Para el almacenamiento del conocimiento, existen tres subprocesos fundamentales:

- Seleccionar, a partir de los múltiples sucesos que vive la organización, las personas y procesos que por su valor deben almacenarse.
- Guardar la experiencia en forma apropiada.
- Garantizar que la memoria organizacional se actualice constantemente.

Una vez almacenado, este conocimiento debe llegar al personal que lo necesite y por tanto se tienen que encontrar métodos para distribuirlo o compartirlo. En este sentido, las organizaciones enfrentan problemas para distribuir y colocar a disposición de sus miembros el conocimiento que ellos

necesitan. Es preciso considerar, que el conocimiento se transfiere mediante acciones personales y por tanto, este proceso puede realizarse desde un centro de distribución del conocimiento hacia uno o varios grupos específicos de individuos, entre y dentro de los grupos y equipos de trabajo de la organización o entre individuos. Para esto, se soportan en herramientas tecnológicas, crean determinadas plataformas, software que facilitan compartir y distribuir el conocimiento, aunque ello no significa que este último se utilice igualmente por todos los individuos en la organización. Se trata de proporcionar el conocimiento que necesita cada individuo para la realización de sus tareas específicas.

La difusión del Conocimiento puede partir de medios informales como la máquina de café, conversaciones, etc. o por otros más formales, como su reproducción, es decir, por medio de la formación, tutoriales, instalaciones de herramientas tecnológicas. Tanto esta como el desarrollo profesional forman parte de la reproducción del conocimiento que se cumple mediante la realización de actividades como son los eventos, los foros-debate, etcétera. Estas técnicas también favorecen a la conservación del conocimiento organizacional, porque al compartirse se evita que la ausencia de un individuo, por una u otra razón, prive a la organización de un conocimiento que necesita.

En el ciclo de los procesos estratégicos de la gestión del conocimiento, el uso del conocimiento se ubica casi al final; sin embargo, esta ubicación es relativa, debido a que los procesos de identificación, adquisición, desarrollo y distribución del conocimiento siempre se encuentran en consonancia con las necesidades de los

usuarios. Por eso, es necesario considerar un sistema de gestión de información que facilite información actualizada sobre las necesidades de los usuarios con vistas a lograr una eficiente gestión del conocimiento.

Para obtener una gestión efectiva del conocimiento, se deben crear plataformas de conocimientos, intranets, portales, escenarios, entre otras herramientas, con el objetivo de incentivar a los individuos a consumir información e incrementar su conocimiento.

Finalmente, el último proceso mencionado es el de la medición del conocimiento. Medir el conocimiento no significa calcular su valor monetario, sino evaluar en qué medida se cumplen o no los propósitos del conocimiento en la organización. Para esto, se aplican diferentes técnicas. El proceso de evaluación y medición del conocimiento puede dividirse en dos fases:

- Una, donde se observan los cambios en la base del conocimiento organizacional.
- Y otra, donde se interpretan estos cambios en relación con los objetivos de dicho conocimiento.

El problema fundamental para medir el conocimiento radica en las características que poseen los sistemas de contabilidad tradicionales, los cuales deben transformarse para poder contabilizar las operaciones con los activos intangibles; ellos sólo posibilitan otorgarle un valor financiero tangible al conocimiento una vez que este se haya incorporado a los bienes comercializables.

Probst (2001) afirma que: “La idea de que el conocimiento puede medirse induce a esperar objetividad donde sólo puede haber aproximación”. Por tanto, en este sentido, los sistemas de medición pueden sólo ofrecer aproximaciones sobre el comportamiento de este activo (el conocimiento) en la organización, debido a su propia naturaleza intangible.

Cada uno de estos procesos estratégicos, que interactúan en la gestión del conocimiento, es susceptible de medirse por medio de diversos indicadores con el objetivo de determinar en qué medida se cumplen o no con eficiencia y tomar medidas correctivas en caso necesario. Esto, sin duda, permite potenciar una adecuada Gestión del Conocimiento que contribuye directamente al incremento del capital intelectual en las organizaciones.

El conocimiento parte de admitir y conocer su variabilidad y sus causas, estas son imposibles de conocer sin medición. Conocer esto es precisamente la clave para administrar el proceso, para conquistar los objetivos de excelencia que se plantea una entidad particular.

2.14 Modelos de gestión del conocimiento

En la literatura existen tantos modelos como autores han estudiado el tema; todos con elementos comunes y diferenciadores, a partir de sus propios aportes, veamos qué es un modelo de gestión del conocimiento.

Para Rodríguez Gómez (2007) un modelo de gestión del conocimiento es:

“Herramienta para representar de forma simplificada, resumida, simbólica, esquemática la gestión del conocimiento; delimitar algunas de sus dimensiones; permitir una visión aproximada; describir procesos y estructuras, orientar estrategias y aportar datos importantes.”

Existen multitud de modelos para la creación y gestión del conocimiento, así como diversas y variadas perspectivas para su estudio, análisis y comprensión.

La multidisciplinariedad inherente al estudio de la Gestión del Conocimiento supone la existencia de diferentes perspectivas para el desarrollo y el estudio de los sistemas y modelos de gestión del conocimiento. A pesar de la existencia de incontables modelos, la revisión de algunos de ellos y la literatura especializada en este ámbito (Davenport y Prusak, 2001; Davenport, De Long y Brees, 1997; Wiig, 1997; Rivero, 2002; Alavi y Leidner, 1999), nos permite agruparlos en tres tipos según el núcleo, los objetivos, la metodología, los participantes, etc., (Edvinsson, 1997) alrededor del cual se desarrollan:

- Almacenamiento, acceso y transferencia de conocimiento: modelos que no suelen distinguir el conocimiento de la información y los datos y que lo conciben como una entidad independiente de las personas que lo crean y lo utilizan. Este tipo de modelos de Gestión del Conocimiento se centran en el desarrollo de metodologías, estrategias y técnicas para almacenar el conocimiento disponible en la organización en depósitos de fácil acceso para propiciar su posterior transferencia entre los

miembros de la organización (por ejemplo: páginas amarillas del conocimiento, archivos de información de las personas, etc.).

- Socioculturales: modelos centrados en el desarrollo de una cultura organizacional adecuada para el desarrollo de procesos de gestión del conocimiento. Intentan promover cambios de actitudes, fomentar confianza, estimular la creatividad, concienciar sobre la importancia y el valor del conocimiento, promover la comunicación y la colaboración entre los miembros de la organización, etc.
- Tecnológicos: modelos en los que destaca el desarrollo y la utilización de sistemas (por ejemplo: *data warehousing*, *intranets*, sistemas expertos, sistemas de información, web, etc.) y herramientas tecnológicas (por ejemplo: motores de búsqueda).

Los modelos de gestión del conocimiento ilustran el aprendizaje organizacional como proceso que integra los conocimientos, habilidades, y actitudes para conseguir cambios o mejoras de conductas en la organización, es decir, el aprendizaje organizacional. Este proceso de cambio requiere herramientas o mecanismos que permitan convertir el conocimiento de las personas y equipos de la empresa en conocimiento colectivo.

Los modelos de gestión de conocimiento tienen por objetivo servir de herramienta para identificar, estructurar, organizar y valorar los activos intangibles que se aplican en la gestión del conocimiento, dichos modelos aportan un valor importante en la organización sobre todo en las instituciones ricas en

conocimiento, cada uno tiene sus fases y procesos de creación del conocimiento.

Estos modelos buscan también responder a cómo los componentes estructurales de la organización interactúan con el recurso humano.

Existen diversas y variadas perspectivas para el estudio, análisis y comprensión de los Modelos de Gestión del Conocimiento analizados. Arambarri (2012) realiza un análisis comparativo de un conjunto de modelos de Gestión de Conocimiento concluyendo lo siguiente:

- Cada modelo tiene su propia base de creación, iniciando por la diferenciación entre el conocimiento tácito y explícito.
- Los modelos de Gestión del Conocimiento deben seguir una serie de pasos para su desarrollo y su implementación dando como resultado la Gestión del Conocimiento.
- Las estrategias que emplean los modelos para la creación de conocimiento son mapas de conocimiento, trabajo en equipo, almacenes de conocimiento, comunidad de aprendizaje, trabajo colaborativo, entre otros.
- La cultura organizacional es muy importante en los modelos de gestión del conocimiento sobre todo para el diseño e implementación de cualquier proceso de Gestión del Conocimiento; y, si es así, qué tipo de cultura nos proponen como idónea para el desarrollo de procesos de creación y gestión del conocimiento.

- Las tecnologías de información y las comunicaciones en los modelos de gestión del conocimiento solo algunos modelos lo utilizan, pero los que la utilizan dan a conocer que las TICs, tienen un papel fundamental en los procesos de Gestión del Conocimiento.
- La relación existente entre la Gestión de la Información, e incluso de la documentación como el modelo Bustelo y Amarilla, y la Gestión del Conocimiento.
- El proceso mismo que reconoce diversas etapas desde la creación hasta el uso y transferencia del conocimiento que dan lugar al aprendizaje en la medida en la que denotan la activación del ciclo del conocimiento. Modelos como los planteados por Nonaka Takeuchi y Goñi resaltan esta fortaleza.
- La convergencia, de manera articulada y sistémica de elementos asociados a la estrategia, la estructura y la cultura como los modelos KPMG y KMAT. Estos resaltan la importancia de cada elemento asociado a la responsabilidad de los individuos pero también a la responsabilidad que tiene la organización de proveer aspectos de tipo infraestructura que soporten adecuadamente la Gestión del Conocimiento.
- Por último, la organización vista como un sistema de relaciones y conexiones, es algo a los que se hace mención pero que se trata más a

profundidad en modelos como el de Bueno, dejando claro como a partir de la Gestión del Conocimiento se construye cultura organizacional.

A continuación se relacionan los elementos fundamentales identificados en cada uno de los modelos (Arambarri, 2012)

| Modelo | Características |
|---------------------------|---|
| Nonaka Takeuchi | El Conocimiento se genera mediante dos espirales de contenido: epistemológico y ontológico. Diferenciación entre conocimiento tácito y explícito. |
| KPMG | Explicativo de la influencia, de manera sistémica, de un conjunto de variables que determinan la capacidad de aprendizaje organizacional. |
| Andersen | Énfasis en el nivel de responsabilidad individual (explicitar el conocimiento) y el organizacional (crear infraestructura posibilitante). |
| KMAT | Liderazgo, cultura, tecnología y medición como variables facilitadoras del proceso. |
| Rotación del Conocimiento | El proceso que se sigue con el conocimiento es cíclico; está en permanente retroalimentación. |
| Bustelo y Amarilla | Relación entre la gestión de la documentación y la gestión de la información |
| Integración de Tecnología | Describe el proceso de integración tecnológica en diferentes capas |
| GC Organizativo | La organización con sistema de relaciones y conexiones que hacen que se produzca conocimiento |

Tabla 7 Comparativa Modelos de Gestión Del Conocimiento.

En esta tesis no se desarrollarán todos los modelos sino aquellos que indiquen directamente con las tecnologías de gestión de la información que permiten generar conocimiento. Según la tabla superior el modelo que mejor se ajusta con el objetivo de esta tesis es el de Bustelo y Amarilla, que se describe a

continuación, dentro del marco de la gestión de la información para generar conocimiento.

2.15 Gestión de la información para generar conocimiento

El conocimiento implica una transformación de la información por parte del sujeto que conoce, es una construcción personal y social que está basada en la información pero que no queda reducida a ella.

Lo anterior nos obliga a hacer una reflexión entre las diferencias que existen entre información y conocimiento:

“El conocimiento, a diferencia de la información, es una característica humana; puede haber información que no sepa nadie, pero no puede existir conocimiento que no sepa nadie. Un conocimiento concreto solo sobrevive en mentes que sean capaces de entenderlo” (Shirky, 2010)

De la Rica (1999) centra la definición de la gestión del conocimiento en el uso de la información indicando que la gestión del conocimiento es un proceso sistemático de búsqueda, selección, organización y presentación de información para mejorar la comprensión y el conocimiento de los empleados y miembros de una organización en un área específica. Asimismo indica que el conocimiento es identificar, estructurar y sobre todo utilizar la información para obtener un resultado y que conocimiento requiere aplicar la intuición y la sabiduría, propios de la persona, a la información.

Malhotra (1997), a su vez la define como la combinación de sinergias entre datos, información, sistemas de información, y la capacidad creativa e innovadora de seres humanos.

Andersen (1999) ha ido más lejos y ha sido capaz de sintetizar el concepto conocimiento en una fórmula:

$$K=[P+I]s$$

El conocimiento organizativo (K) es la capacidad de las personas (P) para interpretar, entender y utilizar la información (I). Una capacidad (K+I) que se multiplica exponencialmente en función de la capacidad de compartir el conocimiento (s) que existe en la organización.

Siguiendo con las diferencias entre información y conocimiento y respecto a su gestión, no son procesos simultáneos. La Gestión de la Información es previa a la Gestión del Conocimiento. De hecho, la información se convierte en conocimiento cuando “alguien” la ha contextualizado, deliberadamente o no, de forma que gracias a este proceso mejora su capacidad de actuar de forma inteligente. Se trata del llamado proceso cognitivo. Poder actuar es lo que separa a la información del conocimiento. Así, lo que para una persona puede ser información, (stock o disposición de datos), para otra es conocimiento (capacidad para la acción). (Arambarri, 2012)

En este sentido según el mismo autor la mayoría de los fracasos de implantar servicios basados en la Gestión del Conocimiento se debe a la arquitectura del mismo. Si la arquitectura del conocimiento es fija, se trata como

una base de datos y no se contempla el ciclo del conocimiento, con lo cual se está diseñando un sistema de gestión de la información, no un sistema de Gestión del Conocimiento.

Uno de los principales problemas de los sistemas de Gestión del Conocimiento es cómo se incorpora el conocimiento tácito a la organización y cómo se clasifican los contenidos dentro del almacén. Según varios autores como Nonaka (1991) y Jackson (2000) ésta es la principal diferencia entre los sistemas que trabajan con información y con conocimiento.

Sedeño Prado (2001) plantea que la gestión del conocimiento es un proceso continuo de adquisición, distribución (en tiempo y forma, a los que la necesiten) y análisis de la información que se mueve en el entorno de una organización para hacer más inteligentes a sus trabajadores (más creativos e innovadores) y de esta forma ser más precisos en la toma de decisiones, dar una respuesta más rápida a las necesidades del mercado, obtener un desarrollo sostenible y ser más competitivos en este entorno tan turbulento, cambiante y lleno de incertidumbre.

Pérez Montoro-Gutiérrez (2008) indica que el conocimiento organizacional se concibe como la capacidad de tomar decisiones y de actuar a partir de la información, que se ha logrado acumular y organizar. Asimismo indica que el conocimiento conlleva información, pero no hay que confundir estos términos, ni interpretarlos como sinónimos, concluyendo que la gestión del conocimiento recoge la información como pieza clave, aunque no como pieza única

Según Bustelo y Amarilla (2001), además de la gestión de la documentación existe, en estrecha relación con ella, la gestión de la información, distribuida en bases de datos corporativas y aplicaciones informáticas, que no se conceptúan como documentos; pero que son una importante fuente de información registrada y lo representan.

Desde este punto de vista, sin una adecuada gestión de la información, es imposible llegar a la gestión del conocimiento. Las propuestas de la Gestión del Conocimiento representan el modelo de gestión que se basa en gran parte en gestionar adecuadamente la información. Es por lo tanto, el paso previo que cualquier organización debe dar antes de tratar de implantar un sistema de gestión del conocimiento.

Existen varios componentes que son necesarios para dar el salto de la gestión de la información a la del conocimiento según estas autoras.

- En primer lugar, la Gestión del Conocimiento es un modelo de gestión de toda la organización. En dependencia de la organización, los sistemas de gestión de la información, cada vez más importantes en la medida que las tecnologías ofrecen nuevas posibilidades, se adaptan a los modelos de gestión existentes.
- Para gestionar el conocimiento es necesario considerar que este no se produce sólo por la gestión de la información, sino que deben intervenir procesos y personas. En una organización, puede existir un perfecto

modelo de gestión de la información, pero si los individuos no lo utilizan es imposible que se cree conocimiento.

Si se considera el análisis de las autoras de este modelo, resulta entonces evidente por qué otra de las tendencias muy involucradas en la definición de la Gestión del Conocimiento es la que proviene de la gestión de los recursos humanos. La gestión de la motivación, del talento, del trabajo en equipo y, sobre todo, la creación de un ambiente de trabajo que facilite compartir ideas, es una tarea a la que difícilmente se accede mediante la gestión de la información, pero que forma parte indispensable de este proceso.



Figura 9 Modelo Bustelo y Amarilla

3 *Business Intelligence*

3.1 **Introducción**

En el mundo actual las organizaciones han acumulado una importante cantidad de datos, generados por los diferentes sistemas de información y procesos que mantiene. Con el objetivo de utilizar de una forma ágil esos datos para convertirlos en información que genere conocimiento se ha desarrollado la Inteligencia de Negocio (BI por sus siglas en inglés: *Business Intelligence*). En este apartado se revisan los aspectos del BI que inciden en el desarrollo del modelo que se ha adaptado y se presenta en el siguiente capítulo.

3.2 **Definición de *Business Intelligence***

Cuando se habla de *Business intelligence* (en adelante BI), una de las definiciones globales que se acostumbra a dar es: BI significa emplear los datos recopilados por una organización con el fin de soportar una mejor toma de decisiones de negocio, accediendo, analizando y cubriendo nuevas oportunidades.

Por ejemplo Negash (2004) lo define como

“La combinación entre la recolección de información, su almacenamiento y la gestión del conocimiento con herramientas de análisis con la finalidad de proporcionar información compleja e importante para los encargados de la planeación y los responsables de la toma de decisiones”.

Dentro de esta definición está presente la idea que los sistemas de *Business intelligence* se proveen de información relevante para la toma de decisiones, al

momento adecuado y en la forma correcta de tal manera que los responsables de la toma de decisiones puedan utilizarlo.

Otra definición muy similar la proporciona Parr (2000) cuando define *Business intelligence* con los siguientes términos:

“Business Intelligence se define como la habilidad corporativa para tomar decisiones. Esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento” (Parr, 2000)

Ambas definiciones coinciden en que el objetivo de un sistema BI es tomar decisiones a partir del tratamiento de los datos que posee una organización.

Para Vercellis (2009), la Inteligencia de Negocio puede ser definida como un conjunto de modelos matemáticos y metodologías de análisis para explotar los datos existentes con el objetivo de generar información y conocimiento útil para los procesos complejos de toma de decisiones.

Otras definiciones más amplias detallan las tecnologías que se suelen emplear en el BI:

“Business Intelligence (BI) es un término paraguas que abarca los procesos, las herramientas, y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los

negocios. BI abarca las tecnologías de datawarehousing los procesos en el 'back end', consultas, informes, análisis y las herramientas para mostrar información (estas son las herramientas de BI) y los procesos en el 'front end'." (Eckerson & Howson, Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale, 2005)

En este caso se introducen los conceptos de *datawarehousing*, procesos en el “*back-end*” y herramientas para mostrar la información. En otras palabras BI incluye los conceptos de *data warehouse*, herramientas de análisis empresarial y gestión del conocimiento y documental (Loshin, 2003)

Estos son los elementos que componen los sistemas de BI y que se desarrollarán más adelante. A modo de resumen la siguiente imagen sirve para ver qué elementos pueden configurar un sistema de BI según Negash (2004).

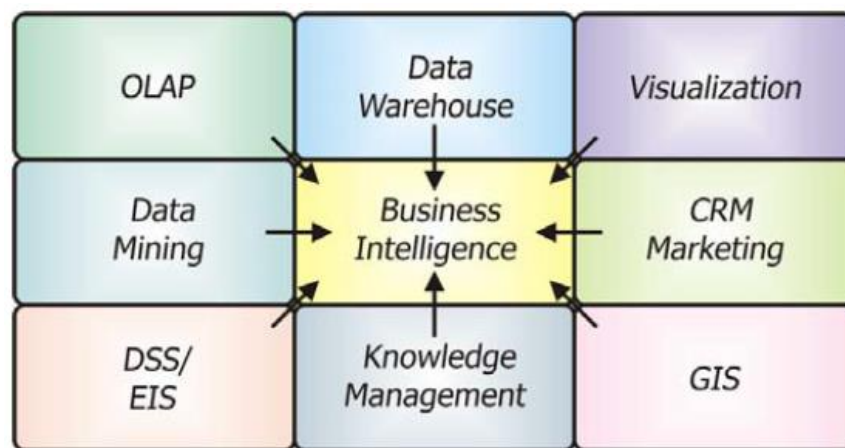


Figura 10 Business Intelligence y su relación con otros sistemas de información.

Adaptado de Negash (2004) p. 179

3.3 Utilidad del BI

En el apartado anterior, donde se han propuesto diversas definiciones de BI, se ha introducido la idea de que la utilidad de implementar un sistema BI radica en poder tomar decisiones según los datos que se generan en las instituciones. No obstante resulta evidente admitir que las instituciones que no disponen de un sistema BI también pueden tomar decisiones en base a los datos que poseen, entonces, ¿cuál es la utilidad real de un sistema de BI?, ¿por qué implementarlo si ya se están tomando decisiones sin un sistema BI, por ejemplo usando las hojas de cálculo como por ejemplo el empleo de Excel?

Según Cano (2007) el empleo de otras herramientas tales como el empleo de hojas de cálculo trae consigo los siguientes inconvenientes:

- Posibilidad de cometer errores: al introducir la información proveniente de distintos sistemas podemos cometer errores.
- Inversión de gran cantidad de tiempo: debemos invertir una cantidad de tiempo considerable en la introducción de la información.
- Repetición del proceso para cada nuevo análisis: cada vez que queramos hacer un análisis deberemos repetir el proceso, además cuando se produzcan modificaciones en el sistema transaccional, deberemos actualizar nuestra hoja de cálculo.
- Poca eficiencia en obtener detalles: si alguien necesita más detalle de las ventas, probablemente no dispondremos del mismo y quizás

respondamos que nos llevaría mucho tiempo introducir toda la información a nivel de detalle.

- Problemas de rendimiento en el entorno transaccional: cuando hagamos las consultas para extraer la información del entorno transaccional, penalizaremos el rendimiento de las aplicaciones.
- Se duplican los entornos de información: cada uno de los usuarios de información querrá los informes con un diseño determinado: Pueden aparecer “entornos paralelos” de información.
- Baja calidad de los datos: si no hay un acuerdo común, es posible que tengamos distintas versiones de una misma realidad: La cifra de ventas de uno de los informes igual no coincide con la de otro.
- Desconocimiento por parte de los usuarios finales: los usuarios finales no siempre saben dónde reside la información que necesitan.

Estos inconvenientes desembocan sobretodo en un aumento del tiempo para obtener los informes y un riesgo de obtener conclusiones incorrectas debido a la baja calidad de los datos. En este sentido Estay Niculcar & Plaza Pimentel (2012) indican que la meta principal del *Business intelligence* es la reducción del tiempo y mejoramiento de la calidad de la información. En este caso se hace énfasis en los beneficios que aporta BI respecto a la mejora de la calidad de la información y del tiempo invertido.

En este sentido diversos estudios concluyen que el costo de no implementar un sistema de toma de decisiones conlleva una pérdida monetaria

cuantificable para la organización. Respecto a la mejora del tiempo invertido algunos ejemplos de las conclusiones de estos estudios se citan a continuación:

“El empleado europeo medio pierde una media de 67 minutos diariamente buscando información de la compañía, lo que equivale a un 15,9% de su jornada laboral. Para una organización de 1.000 empleados que gane unos 50.000 euros al día esto equivale a 7,95 millones de euros al año de salario perdido, todo ello por la búsqueda de información para tomar una decisión” (Zúmel, 2008)

“En muchos hospitales los analistas financieros destinan un 80% del tiempo a agregar y normalizar manualmente información en hojas de cálculo Excel, y tan sólo un 20% a analizar la información relevante”. (Rice, 2004)

Estudios como los anteriormente llevan a concluir que una de las utilidades directas es obtener la información de forma eficiente y esto se traduce en invertir menos tiempo en prepararlas que en analizarlas. En este sentido no resulta lógico que se invierta el 80% del tiempo preparando información y tan solo el 20% analizándola: los porcentajes deberían ser al revés, es decir, destinar la mayor parte del tiempo a analizar la información y tan sólo una pequeña parte del tiempo a prepararla entendiendo que es en la toma de decisiones cuando se aporta valor, no en la preparación de la información. (Cano, 2007)

Respecto al mejoramiento de la calidad de la información apuntada por Estay Niculcar & Plaza Pimentel (2012) en su definición de BI un estudio afirma (Vesset & McDonough, 2006) que solo el 15% de los directivos que toman las decisiones confían en que la mayoría de los informes y cuadros de mando utilizados en sus organizaciones aportan la información relevante a las personas adecuadas en el momento adecuado.

En este sentido Inmon (2006) afirma lo siguiente:

“Las organizaciones actúan bajo la suposición de que la información de la que disponen es precisa y válida. Si la información no es válida, entonces no pueden responder de las decisiones basadas en ella.”

Todo lo anterior apunta a que la calidad de la información incide directamente en la confianza que se tenga por parte de los responsables de tomar decisiones y por tanto, si se consigue mejorar dicha calidad también mejorará la confianza de los directivos, haciendo que las decisiones que tomen sean más válidas. No obstante el mero hecho de implementar un sistema de BI no necesariamente implica que haya un aumento de la calidad de los datos, este es un proceso implícito a la implantación de cualquier sistema de BI y que se detalla en apartados posteriores.

Hasta aquí se han afrontado dos beneficios de implementar un sistema BI uno directo referente a la mejora de la eficiencia respecto al tiempo invertido en

obtener los informes, y otro no tan directo relativo a la aceptación de la información que se presenta fruto de la mejora de la calidad de las fuentes de datos que implícitamente debe realizarse antes de ejecutar el sistema. No obstante hay muchos otros beneficios que se pueden aplicar a los sistemas BI. En este sentido Estay Niculcar & Plaza Pimentel (2012) afrontan esta circunstancia planteando una serie de preguntas, problemas a resolver y objetivos a perseguir por parte de la entidad susceptible de implementar un BI.

| El BI en las organizaciones | |
|---|--|
| Preguntas que pueden resolverse mediante un sistema BI | ¿Quiénes son los clientes más rentables?, ¿Cómo retenerlos? ¿Cómo expandirse con nuevos productos, mercados y canales de forma eficaz? ¿Cuáles son las ventajas por región, producto y época? ¿Cómo mejorar el nivel de servicio? ¿Cómo mejorar la eficiencia? |
| Problemas que justifican la implantación de un sistema BI | Baja disponibilidad de la información necesaria Distintas fuentes de datos no consolidadas Acceso a la información en términos de negocio Identificación de los factores críticos que afectan a los resultados futuros o pasados |
| Objetivos que persigue un sistema BI | Mejorar la operativa del negocio Consolidar y homogeneizar las fuentes de información disponibles Ganar cuota de mercado Mejorar el servicio al cliente y retener a aquellos que son más rentables Mejorar la eficacia de las campañas promocionales |

Tabla 8: ¿Cómo puede ayudar un BI a las organizaciones?

Otros autores ofrecen otro tipo de clasificación de los beneficios. A continuación se muestran dichos beneficios ordenados según una clasificación más amplia (Fuchs, 2003) (Whittemore, 2003):

- Beneficios tangibles, por ejemplo: reducción de costes, generación de ingresos, reducción de tiempos para las distintas actividades del negocio.
- Beneficios intangibles: el hecho de que tengamos disponible la información para la toma de decisiones hará que más usuarios utilicen dicha información para tomar decisiones y mejorar la nuestra posición competitiva.
- Beneficios estratégicos: Todos aquellos que nos facilitan la formulación de la estrategia, es decir, a que clientes, mercados o con que productos dirigirnos.

La tabla inferior adaptada de Whittemore (2003) recopila ejemplos de cada uno de los beneficios indicados arriba desglosando los beneficios tangibles en reducción de costes y generación de ingresos.

| Tipo de beneficio | Ejemplos |
|---|--|
| Beneficios tangibles: generación de ingresos | <p>Mejorar la adquisición de clientes y su conversión mediante el uso de la segmentación.</p> <p>Reducir la tasa de abandono de clientes, incrementar su fidelidad, teniendo en cuenta cuál es su valor.</p> <p>Incrementar los ingresos por crecimiento de las ventas.</p> <p>Aumentar los resultados, consiguiendo que nuestros clientes actuales compren más productos o servicios.</p> <p>Evitar las pérdidas producidas por las ventas de nuestros competidores.</p> <p>Aumentar la rentabilidad por el acceso a información detallada de productos, clientes, etc.</p> <p>Conocer mejor cuáles son las características demográficas de nuestra zona de influencia.</p> <p>Hacer crecer la participación de mercado.</p> <p>Reducir el tiempo de lanzamiento de nuevos productos o servicios.</p> <p>Mejorar aquellas actividades relacionadas con la captura de datos.</p> <p>Analizar la cesta de la compra y la afinidad de venta entre los productos.</p> <p>Facilitar la adopción de los cambios en la estrategia.</p> <p>Proveer el autoservicio de información a trabajadores, colaboradores, clientes y proveedores.</p> <p>Medir la efectividad de las campañas rápidamente y ser capaces de hacer los ajustes durante el ciclo de vida de las mismas.</p> <p>Optimizar las acciones de marketing.</p> <p>Identificar clientes rentables en segmentos no rentables.</p> <p>Analizar a la competencia cuando establecemos precios.</p> <p>Crear nuevas oportunidades.</p> |

| Tipo de beneficio | Ejemplos |
|---|---|
| Beneficios tangibles, reducción de costes | <p>Negociar mejores precios con los proveedores, identificar a los proveedores más importantes, gestionar descuentos por cantidades de compra; permitir el análisis del cumplimiento de los proveedores y asignar las compras de acuerdo con ello; medir el nivel de calidad, servicio y precio.</p> <p>Reducir o reasignar el personal necesario para llevar a cabo los procesos.</p> <p>Incrementar la productividad con información más inmediata y mejor.</p> <p>Aumentar el control de costes.</p> <p>Disminuir los gastos.</p> <p>Eliminar ineficiencias y reducir los costes operativos al generar “una sola versión de la verdad”.</p> <p>Menguar pérdidas detectando fraude.</p> <p>Reducir los créditos incobrables.</p> <p>Hacer bajar las reclamaciones.</p> <p>Eliminar las sobreproducciones de productos.</p> <p>Proveer inventarios “<i>just-in-time</i>”; gestionar los inventarios de productos acabados; mejorar la gestión de materias primas y productos acabados; reducir los costes de mantenimiento del stock.</p> <p>Acortar los tiempos de respuesta a las peticiones de informes.</p> <p>Analizar los problemas, reparaciones y defectos y proveer la información para hacer un seguimiento y corregir los problemas recurrentes.</p> <p>Reducir el tiempo para recoger la información para cumplir con las normativas legales.</p> <p>Evaluar el rendimiento de los activos y generar las alertas cuando el mantenimiento preventivo se debe llevar a cabo.</p> <p>Proveer el suministro dentro del plazo.</p> <p>Hacer un seguimiento de los problemas de los productos desde el inicio de su vida hasta el fin.</p> <p>Retirar los viejos equipos para disminuirlos costes de actividad.</p> <p>Reducir las devoluciones de productos.</p> <p>Analizar la productividad de los empleados.</p> <p>Dar soporte a las reclamaciones de facturación.</p> <p>Abaratar los costes de las acciones de marketing.</p> <p>Hacer un mejor seguimiento de las acciones delictivas.</p> <p>Saber qué comisiones se deben pagar.</p> |

| Tipo de beneficio | Ejemplos |
|-------------------------|--|
| Beneficios intangibles: | Optimizar la atención a los clientes. Aumentar la satisfacción de los clientes. Mejorar el acceso a los datos a través de consultas, análisis o informes. Información más actualizada. Dotar a la información de mayor precisión. Conseguir ventajas competitivas. Ahorrar costes. Menor dependencia de los sistemas desarrollados. Mayor integración de la información. |
| Beneficios estratégicos | Mayor habilidad para analizar estrategias de precios. Mayor habilidad para identificar y nutrir a aquellos clientes con mayor potencial. Mejorar la toma de decisiones, realizándola de forma más rápida, informada y basada en hechos. Mayor visibilidad de la gestión. Dar soporte a las estrategias. Aumentar el valor de mercado. |

Tabla 9 Beneficios de implementar un sistema de BI (Whittemore, 2003)

De lo anterior se concluye que existen muchos beneficios algunos más directos y “fáciles” de cuantificar y otros no tan directos, de difícil cuantificación monetaria pero igualmente importantes a la hora de medir el impacto positivo que genera un sistema BI.

3.4 Elementos de un BI

Las definiciones aportadas en los apartados precedentes conllevan a inferir que la Inteligencia de Negocio es un área interdisciplinar y entre los temas que estudia se encuentran la organización empresarial, los diferentes sistemas de información, la construcción de almacenes de datos, así como las técnicas y modelos matemáticos aplicados al análisis de datos para la toma de decisiones.

Esta disciplina pone todo su esfuerzo en general valor (información y conocimiento) para la organización en pos de apoyar la toma de decisiones de una forma efectiva.

De forma más detallada un sistema de BI está compuesto por los elementos reflejados en la siguiente figura:

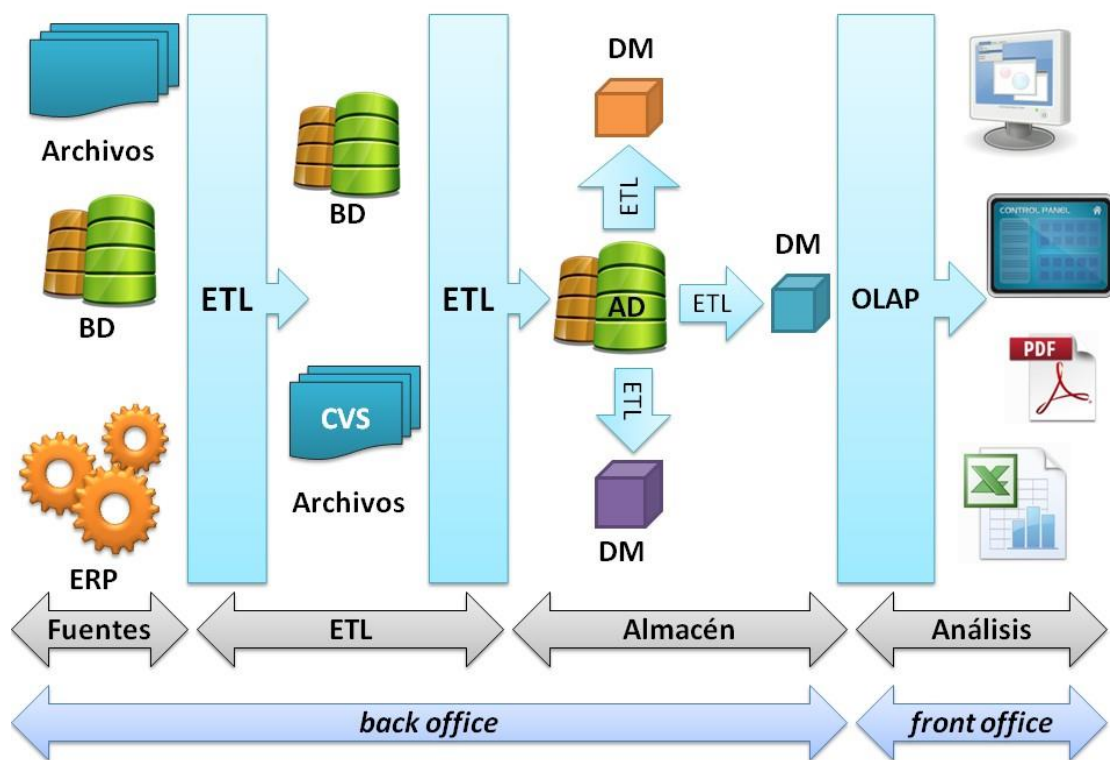


Figura 11 Arquitectura genérica de un almacén de datos

De forma resumida se describe cada una de las partes de la imagen a continuación:

- Las fuentes de los datos, que pueden ser múltiples en cantidad y tipos. Representan el origen del proceso desde donde se extraen los

datos iniciales que contienen la información necesaria para generar el conocimiento que se precisa.

- Procesos ETL (del inglés *Extract, Transform and Load*) donde se extraen, transforman y limpian los datos fuente para ser cargados en el almacén de datos. Normalmente, la información que tenemos en los sistemas transaccionales no está preparada para la toma de decisiones por este motivo deben realizarse estas operaciones.
- Los procesos ETL a menudo contienen un área de estacionamiento usada para posicionar los datos extraídos y realizarles una transformación y limpieza inicial.
- *Data warehouse* o almacén de datos (AD) que se nutre de los procesos ETL que trabajan sobre las fuentes de datos.
- Subconjuntos de datos llamados *Data marts* (DM) con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.
- Herramientas OLAP (Procesamiento Analítico en Línea por su acrónimo en inglés: *Online Analytical Processing*), los reportes, cuadros de mando integral, herramientas de minería de datos y otras. Estas se encargan de presentar la información de la forma requerida por el usuario.
- Las herramientas de visualización, que nos permitirán el análisis y la navegación a través de los mismos.

La Figura 11 está dividida en *back office* y *front office*. Para Kimball & Ross (2002) el back office representa la tecnología y los procesos utilizados para construir y mantener el almacén de datos, mientras que el *front office* es el área que utilizan los usuarios finales (Bouman & Van Dongen, 2009)

Los almacenes de datos, en cualquiera de sus arquitecturas, dan soporte a aplicaciones OLAP, sistemas de soporte a la toma de decisiones (DSS por sus siglas en inglés: *Decision Support System*) y aplicaciones de minería de datos.

Entre sus beneficios están: disponer de los datos en un solo lugar, contar con datos actualizados, históricos y de rápido acceso, y ser un repositorio donde los términos son de fácil entendimiento y están estandarizados. Por estas razones son considerados como un punto de arranque para construir un proyecto de inteligencia empresarial.

Los siguientes apartados explican con más detalle cada uno de los elementos presentados en este apartado.

3.5 Fuentes de información

Las fuentes de información suelen ser las siguientes:

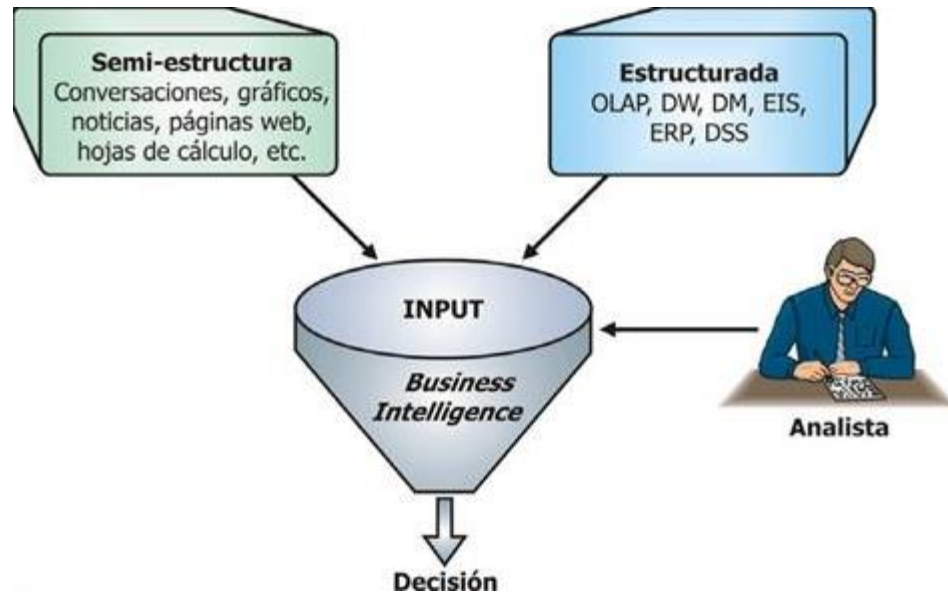
- Sistemas operacionales o transaccionales, que incluyen aplicaciones desarrolladas a medida, ERP, CRM...
- Sistemas de información departamentales: previsiones, presupuestos, hojas de cálculo, etcétera.

- Fuentes de información externa a la organización. En algunos casos es interesante incorporar información referente, por ejemplo, a población, número de habitantes, etc.

Como se puede imaginar el número de fuentes de información distintas de las que cargamos la información contribuye a la complejidad del proceso de carga dentro del DW. El número de fuentes no es el único factor de complejidad sino que el hecho de acceder a distintas bases de datos requiere distintas habilidades. Si el número de bases de datos a las que debemos acceder es elevado, puede provocar que tanto las definiciones como las codificaciones en los distintos entornos sean diferentes, lo que añadirá dificultad a nuestro proyecto; por ello, un aspecto clave será conocer el modelo de información transaccional y el significado de cada uno de sus elementos. En este sentido las aplicaciones no están suficientemente documentadas para ser interpretadas correctamente. En la mayoría de los casos son aplicaciones que han sido modificadas a lo largo del tiempo por distintos programadores, y normalmente no se han actualizado (Cano, 2007).

Debido a la diversidad de fuentes que alimentan a una solución de *Business intelligence* hace que sea más complejo el análisis de las informaciones. Estas poseen en algunos casos una estructura dada por el mismo entorno que las contiene, por ejemplo, una base de datos en donde las informaciones se graban en forma de filas y columnas (tablas de datos) o en otros casos las informaciones que procesa un sistema de *Business intelligence* proviene de fuentes que tienen una estructura diferente por ejemplo: e-mail, faxes, cartas, etc. Para el primer caso en

donde las informaciones presentan una estructura se las conocen como fuentes de datos estructuradas; caso contrario se los conocen como fuentes de datos semiestructuradas.



*Figura 12 Tipos de fuentes de datos según su estructura y su relación con el BI
(Estay Niculcar & Plaza Pimentel, 2012)*

Blumberg y Atre (2003) indican que el 60% de los directores de Sistemas de Información y de los de Tecnología de las compañías consideran a los datos semiestructurados como críticos para mejorar las operaciones de las organizaciones y clave vital para el desarrollo de nuevos negocios.

Adicionalmente Blumberg y Atre (2003) indican que no es fácil indagar información en las fuentes de datos semiestructuradas utilizando la tecnología existente en las bases de datos convencionales ya que la variedad es bastante amplia; entre las que se pueden citar están:

- Procesos de negocio
- E-mails
- Gráficos
- Archivos con imágenes
- Cartas
- Memos, noticias, etc.

Blumberg y Atre (2003) indica que Merrill Lynch estima que el 85% de toda la información de los negocios existe en forma de datos semiestructurados comúnmente capturados en formato de hojas de cálculo, es decir, no se encuentran almacenados bajo la estructura de una base de datos.

Independientemente de si la fuente de los datos es estructurada o no, asegurar la calidad de los datos es fundamental, (Inmon W. , 2006). Es necesario asegurar que la calidad de los datos es máxima. Si en el *data warehouse* hay errores, éstos se propagarán a lo largo de toda la organización y son muy difíciles de localizar. Además, pueden ocasionar que se tomen decisiones erróneas que afecten a los resultados de la organización. Los costes derivados de que la calidad de los datos no sea la correcta pueden llegar a ser muy elevados. Si los usuarios perciben que no tenemos suficiente calidad de datos rápidamente desprestigiarán el proyecto de *Business Intelligence*.

Asumir que la calidad de los datos es buena puede ser un error fatal en los proyectos de *Business intelligence* (Beyer & Feinberg, 2006) por tanto debería

establecerse un control o conjunto de controles en el proyecto que localizara los errores en los datos y no permitiera la carga de los mismos.

Los errores relativos a la calidad de los datos pueden tener un doble origen (Cano, 2007):

- Fallos en los sistemas transaccionales que sirven de fuente de datos, por ejemplo permitir que los usuarios puedan introducir datos sin ningún tipo de control es un riesgo que incide sobre la calidad de los mismos.
- Errores ocurridos en el proceso de ETL o al integrarlos en el *data warehouse*.

Según Beyer & Feinberg (2006) el control para evaluar la calidad de los datos debe hacerse tanto en el proceso de carga, una vez cargados mediante un proceso de auditoría y durante el uso del sistema BI por parte de los usuarios finales.

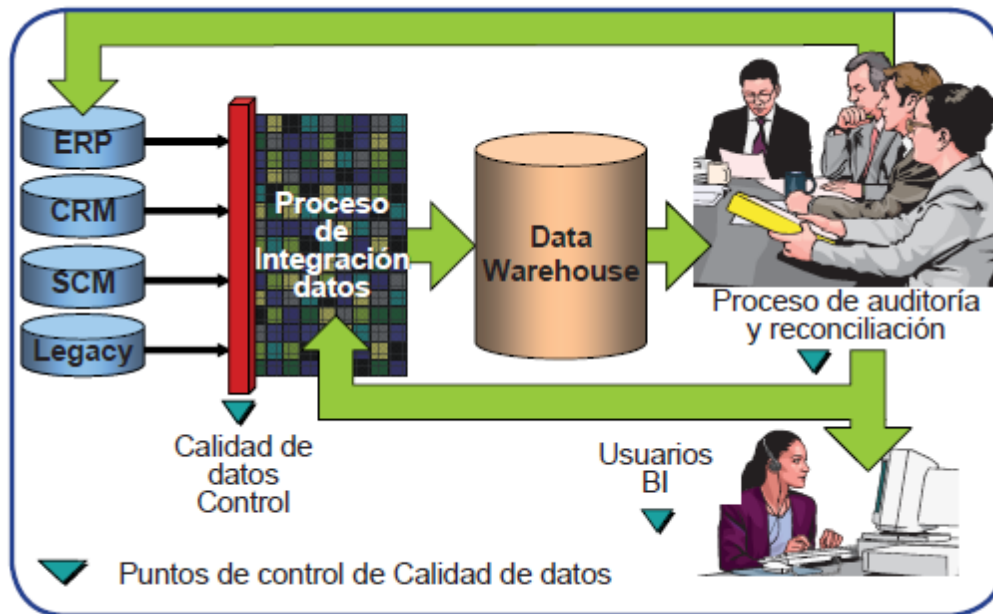


Figura 13 Puntos de control de calidad de datos. Adaptado de Beyer & Feinberg (2006)

Respecto a la implementación de un sistema de calidad de los datos, según una encuesta (Eckerson W. W., 2002) de *The data warehouse Institute* no hay demasiadas organizaciones que tengan un plan de calidad de datos implementado: El 48% de las organizaciones encuestadas no tenían un plan para gestionar o mejorar la calidad de los datos a pesar de que eran plenamente conscientes de los problemas que les ocasionaba la baja calidad de datos y de los beneficios que les reportaría si tuvieran mejor calidad.

La tabla inferior resume los resultados obtenidos en la encuesta.

| Tipo | Descripción |
|---|--|
| Problemas de la baja calidad de los datos | Tiempo extra para reconciliar los datos (87%). Pérdida de credibilidad en el sistema (81%). Costes extra, por ejemplo duplicidades en <i>mailing</i> (72%). Insatisfacción de clientes (67%). Retrasos en el desarrollo de nuevos sistemas (64%). Pérdidas de ingresos (54%). Problemas de conformidad (38%). Otros (5%). |
| Beneficios de tener datos con alta calidad | Simple versión de la verdad (19%). Incrementos en la satisfacción de los clientes (19%). Mayor confianza en los sistemas de análisis (17%). Reducción de costes (13%). Menor tiempo para reconciliar los datos (12%). Incremento de ingresos (9%). Otros (12%). |
| Características que deberían cumplir los datos para tener una buena calidad | Precisión: ¿Representan los datos con precisión una realidad o una fuente de datos que se pueda verificar? Integridad: ¿Se mantienen constantemente la estructura de los datos y las relaciones a través de las entidades y los atributos? Coherencia: ¿Son los elementos de datos constantemente definidos y comprendidos? Totalidad: ¿Están todos los datos necesarios? Validez: ¿Son los valores aceptables en los rangos definidos por el negocio? Disponibilidad: ¿Están los datos disponibles cuando se necesitan? Accesibilidad: ¿Se puede acceder a los datos fácil y comprensiblemente? |

Tabla 10 Problemas, beneficios y características de la calidad de los datos

Según Bitterer (2006) los problemas de calidad de datos son un problema de negocio, no de los departamentos de tecnología.

Las recomendaciones que deberíamos seguir, según Cano (2007), para mejorar la calidad de los datos son las siguientes:

- Conocer tus datos es la clave para el éxito en muchos negocios e iniciativas de tecnología:
 - Realizar una auditoria incluyendo una evaluación de la calidad.
 - Conocer dónde están los datos y su nivel de calidad.
 - Incluir la calidad de los datos en la estrategia de metadatos.
- Establecer un programa formal de calidad de datos:
 - Construir el acuerdo para aplicarla en toda la gestión de las fuentes de datos.
 - Establecer acciones de calidad de datos en la gestión de la información de la organización.
- Desarrollar las habilidades necesarias y organizar un equipo, tanto a nivel de los usuarios de negocio como de los de tecnología.
- Definir las políticas y las métricas de la calidad de datos:
 - Definir los estándares de direcciones, como calcular el beneficio, los ingresos, etc.
 - Establecer y usar métricas para alcanzar la calidad de los datos.
- Implementar tecnologías de calidad de datos, reconociendo que tan sólo son una parte de la solución.

3.6 ETL (*Extract, Transform and Load*)

Los procesos ETL son los componentes más importantes y de mayor valor añadido en una infraestructura que implique la integración de varias fuentes de datos. En consecuencia, representan un pilar fundamental tanto de simples proyectos de recopilación como de soluciones complejas de inteligencia de negocio, especialmente si se requiere mucha precisión o actualización en los datos.

Las fuentes de datos, que pueden ser diversas y con diversos formatos tal como se ha apuntado en el apartado 3.5 del presente capítulo, son extraídas y transformadas mediante herramientas procesos ETL con el fin de recuperar sus datos y alimentar de forma correcta el *data warehouse*.

Extraer, Transformar y Cargar (ETL por sus siglas en inglés: *Extract, Transform and Load*) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en otra base de datos, *data mart* o almacén de datos para analizar, o en otro sistema operacional para apoyar un proceso de negocio (Dataprix, 2014).

El proceso ETL normalmente es un proceso manual donde el administrador tiene que aplicar una cantidad de tareas de pre-procesamiento de datos tales como limpieza, identificación de usuarios, auto llenado de rutas, transacción, identificación transformación, enriquecimiento, integración y reducción de datos (Kimball & Caserta, 2002)

Como se puede suponer este proceso es complejo, al tener una componente manual muy elevada. De hecho según Eckerson & White (2003) consume entre el 60% y el 80% del tiempo de un proyecto de *Business Intelligence*, por lo que es un

proceso muy costoso y, tal como se ha apuntado en el apartado 3.5 referente a la calidad de los datos en la vida de todo proyecto de BI es crítico que su calidad sea excelente, de lo contrario todo el sistema puede dejar de tener validez.

Asimismo existen muchas fuentes de datos que son poco o nada útiles para alimentar el sistema BI por lo que será necesario realizar un análisis de las mismas y determinar cuáles son las mejores fuentes de información durante la fase de extracción. En este sentido el principal objetivo de la extracción es extraer tan sólo aquellos datos de los sistemas transaccionales que son necesarios y prepararlos para el resto de los subprocesos de ETL. (Cano, 2007)

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. En la mayoría de los casos los datos extraídos requerirán alguna manipulación y transformación. Algunas de estas operaciones pueden ser: sustituir algunos términos por otros (por ejemplo el sexo M por 1, F por 0), obtener nuevas

columnas (por ejemplo desglosar una fecha en día, mes y año) o unir datos de múltiples fuentes o columnas de una tabla.

La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los almacenes de datos mantienen un historial de los registros de manera que se pueda hacer una auditoría de calidad de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

Existen dos formas básicas de desarrollar el proceso de carga:

- **Acumulación simple:** La acumulación simple es la más sencilla y común, consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportando el resultado como una única transacción hacia el *data warehouse*, almacenando un valor calculado que consistirá típicamente en una sumatoria o un promedio de la magnitud considerada.
- **Rolling:** El proceso de Rolling por su parte, se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones

de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etc.) (Dataprix, 2014).

La tabla inferior resume todos los subprocesos o fases de la ETL

| Subproceso | Descripción | Objetivo | Problemas | Resultado |
|----------------|---|--|--|---|
| Extracción | Recuperación de los datos de las distintas fuentes de información. | Extraer tan sólo aquellos datos de los sistemas transaccionales que son necesarios y prepararlos para el resto de los subprocesos de ETL | Distintos/as: Fuentes, BBDD Plataformas tecnológicas, Protocolos de comunicaciones, Juegos de caracteres, Tipos de datos | Datos en bruto que no han sido depurados y que deben ser limpiados. |
| Limpieza | Depurar, corregir, estandarizar, relacionar y consolidar los datos extraídos de las fuentes de datos. | Conseguir datos sin duplicados, homogéneos, filtrados y corregidos. | Distintos/as formatos en fechas, abreviaturas... Errores en los datos detectables mediante algoritmos y fuentes de datos externas. Datos demasiado agregados que se pueden desagregar. Datos duplicados | Datos depurados desagregados |
| Transformación | Cambios de formato, sustitución de códigos, valores derivados y agregados | Transformar los datos según las reglas de negocio y los estándares prefijados. | Los datos están desagregados y teniendo en cuenta el nivel de granularidad que se le quiera dar se realizan agregaciones. | Datos depurados, recalculados y agregados. |

| Subproceso | Descripción | Objetivo | Problemas | Resultado |
|---------------|--|---|--|---|
| Integración | Proceso en el que los datos depurados, recalculados y agregados se carga en el <i>data warehouse</i> | Cargar los datos en el <i>data warehouse</i> y asegurar de que no haya errores en este proceso. | El principal problema en esta fase es que falle el proceso de carga y los datos cargados no coincidan con los datos transformados del subproceso anterior. | Datos cargados en el <i>data warehouse</i> sin errores. |
| Actualización | Proceso de actualizar los datos del <i>data warehouse</i> | Realizar actualizaciones según una periodicidad definida | Carga del sistema: hay que definir el mejor momento para no saturar los sistemas informáticos. Todos aquellos derivados de los subprocesos anteriores. | Datos actualizados periódicamente |

Tabla 11 Resumen de los subprocesos ETL. Adaptado de Cano, 2007

Nótese que los subprocesos identificados en la Tabla 11 Resumen de los subprocesos ETL. Adaptado de Cano, 2007 referents a la extracción, limpieza, transformación e integración de datos se repiten en cada ciclo de actualización de los datos. En este sentido es importante mencionar que si bien los sistemas transaccionales se actualizan normalmente en tiempo real para la actualización de los *data warehouses* los plazos suelen ser mensuales, semanales o diarios.

No es muy habitual que los datos se tengan que cargar en el *data warehouse* continuamente, lo que se define en la literatura como “Real Time”. Debemos decidir los plazos en función de las necesidades de los usuarios de negocio.

Colin White (2004) en su artículo define el concepto de “*Right Time*” como aquel que implica que diferentes situaciones de negocio y eventos requieren diferentes respuestas o tiempos para actuar. En algunos casos necesitaremos respuestas en tiempo real, mientras que en otros un retraso de minutos, horas e incluso días es aceptable.

En el mercado existen herramientas ETL para realizar los subprocesos de extracción, transformación y carga. Estas herramientas deben contar con los siguientes atributos o funcionalidades (Eckerson & White, Evaluating ETL and Data Integration Platforms, 2003):

- Diseño gráfico: entorno visual que permite a los desarrolladores establecer la relación entre las fuentes de datos, las transformaciones, los procesos y las tareas para desarrollar la carga.

- Gestión de los metadatos: proveer un repositorio donde definir, documentar y gestionar la información del proceso ETL y su ejecución.
- Extracción: Extracción de la información mediante conectores, como ODBC, SQL nativos de los distintos motores de bases de datos o ficheros planos.
- Transformación: Deberían proveer de librerías de transformación que permitan a los desarrolladores transformar los datos origen en los destino con las nuevas estructuras y crear las tablas de agregación para mejorar el rendimiento.
- Carga: Utilizar adaptadores para poder insertar o modificar los datos en el *data warehouse*.
- Servicios de transporte: Las herramientas ETL utilizan las redes y sus protocolos (por ejemplo: FTP) para mover los datos entre las distintas fuentes y los sistemas destino.
- Administración y operación: Las herramientas ETL deben permitir a los administradores programar, ejecutar y monitorizar los trabajos de ETL, los resultados, gestionar los errores, recuperar los fallos y reconciliar los resultados con los sistemas originales.

3.7 *Data warehouse*: Almacenes de datos

En la actualidad, una importante cantidad de entidades han abrazado las tecnologías de la información para sustentar el proceso de gestión de su organización. Las bases de datos se han convertido en un elemento inseparable de

los diferentes sistemas, permitiendo obtener informes complejos, resúmenes e incluso estadísticas sobre la información almacenada con la meta de apoyar el proceso de toma de decisiones. En este sentido cada día es más necesario distinguir dos usos diferentes de los sistemas de información: el procesamiento transaccional y el procesamiento analítico.

Los almacenes de datos responden al procesamiento analítico y surgieron por dos razones: primero, la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyar la toma de decisiones; segundo, la necesidad de hacerlo sin afectar los sistemas operacionales. Generalmente, implican recopilar datos desde diversos orígenes para crear un origen añadido de información y poder generar informes. Un almacén de datos es una vista consolidada de los datos empresariales, optimizada para la creación de informes y análisis. Los datos y la información se extraen de orígenes de datos de producción a medida que se generan (información en tiempo real) o en fases periódicas (información latente). De este modo se facilita la realización de consultas, comparando directamente los datos en vez de acceder a cada origen de datos por separado.

Existen múltiples definiciones y enfoques sobre lo que un *data warehouse* es y su función. A continuación se muestran algunas de ellas:

Una definición inicial la proporciona Watson (Watson H. J., 2002)

*“Un data warehouse es una colección de información
creada para soportar las aplicaciones de toma de decisiones”*

Hernández Orallo (2004) describe a los almacenes de datos de forma más específica como:

“Tecnología que pretende proporcionar metodologías y tecnología para recopilar e integrar los datos históricos de una organización, cuyo fin es el análisis, la obtención de resúmenes e informes complejos y la extracción de conocimiento. Esta tecnología está diseñada especialmente para organizar grandes volúmenes de datos de procedencia generalmente estructurada (bases de datos relacionales, por ejemplo), aunque el concepto general es útil para la organización de pequeños conjuntos de datos en aplicaciones de minería de datos más modestas”

Más adelante, refiriéndose a la conceptualización de un almacén de datos, precisa:

“Un almacén de datos es un conjunto de datos históricos, internos o externos, y descriptivos de un contexto o área de estudio, que están integrados y organizados de tal forma que permiten aplicar eficientemente herramientas para resumir, describir y analizar los datos con el fin de ayudar en la toma de decisiones estratégicas”

Un almacén de datos para Inmon (1996) es:

*“Una colección de datos orientados al dominio,
integrados, no volátiles y variables en el tiempo, organizados
para dar apoyo al proceso de toma de decisiones”.*



Figura 14 Características de un data warehouse según Inmon (1996)

La tabla inferior refleja cada una de las características de la definición de

Inmon:

| Características de los datos | Descripción |
|--|---|
| Datos orientados al dominio (área, negocio...) | Cada parte del <i>data warehouse</i> está construida para resolver un problema de negocio, que ha sido definido por los tomadores de decisiones. |
| Datos integrados | La información debe ser transformada en medidas comunes, códigos comunes y formatos comunes para que pueda ser útil. La integración permite a las organizaciones implementar la estandarización de sus definiciones, por ejemplo: La moneda en la que están expresados los importes es común. |
| Datos no volátiles | La información se almacena para la toma de decisiones. No se va actualizando continuamente, sino periódicamente, de forma preestablecida. |
| Datos variantes en el tiempo | Se mantiene la información histórica y se almacena referida a determinadas unidades de tiempo, tales como horas, días, semanas, meses, trimestres o años. Ello nos permitirá analizar, por ejemplo, la evolución de un hecho en los periodos que queramos |

Tabla 12 Descripción de las características de un data warehouse según Inmon (1996)

Existe un distinto enfoque entre los autores respecto a la importancia y alcance de un *data warehouse* en relación al BI. Kimball (2002) considera que el *data warehouse* es la plataforma para el BI mientras que para Inmon (1996) es simplemente la base de datos donde se consolidan y almacenan los datos empresariales

Para Ralph Kimball, considerado el más fiel promotor del enfoque dimensional, un almacén de datos es una copia de los datos transaccionales, específicamente estructurados para la consulta y el análisis. (Kimball & Ross, 2002).

Las características de un *data warehouse* según Kimball (1996) son:

- El *data warehouse* da acceso a la información de la corporación o del área funcional. El alcance del *data warehouse* puede ser bien un departamento o bien corporativo.
- La información del *data warehouse* es consistente.
- La información en el *data warehouse* puede ser separada y combinada para analizar cada una de las posibles medidas del negocio.
- El *data warehouse* no es sólo información sino también las herramientas de consulta, análisis y presentación de la información.
- Es el lugar donde publicamos la información.
- La calidad de la información en el *data warehouse* es el motor del *business reengineering*.

Un concepto que liga ambas visiones (Inmon vs Kimball) sobre lo que se considera *data warehouse* podría ser el de *data warehousing*. Según Watson (2002) es el proceso completo de extraer información, transformarla y cargarla en un *data warehouse* y el acceso a esta información por los usuarios finales y las aplicaciones.

Este entorno de *data warehousing* nos debería permitir acceder a información que ha sido estructurada para hacer consultas, y estas consultas deberían permitir a los usuarios percibir el valor de esa información. Muchas veces el valor aparece en un proceso secuencial de consultas, análisis y más consultas. Normalmente, las primeras consultas devuelven una gran cantidad de información que será refinada posteriormente con nuevas consultas. No todos los caminos de análisis escogidos serán los adecuados, y algunos los abandonaremos volviendo a modificar la consulta inicial. Al analizar formularemos hipótesis, que serán confirmadas o desmentidas por la información residente en el *data warehouse*. (Cano, 2007)

Hasta ahora se ha considerado el *data warehouse* como una única y gran base de datos, no obstante pueden estar distribuidos en distintas bases de datos llamadas *data marts* que no es otra cosa que una versión especial de almacén de datos para un propósito más específico. Este tiene el objetivo de responder a un determinado análisis, función o necesidad del negocio. Como refiere Bernabéu (2010), los *data marts* son pequeños almacenes de datos centrados en un tema o área de negocio específico dentro de una organización.

El empleo de estos *data marts* surge como respuesta a las distintas realidades que frecuentemente ocurren en las empresas, relativas al esfuerzo que se requiere para construir un *data warehouse* centralizado: el trabajo de construir un *data warehouse* corporativo puede generar inflexibilidades, o ser costoso y requerir plazos de tiempo que las organizaciones no están dispuestas a aceptar

(Cano, 2007). Esto resulta evidente si se tiene en cuenta que normalmente los *data mart* almacenan información de un número limitado de áreas y se definen para responder a usos muy concretos. Entre las características que los describen encontramos las siguientes:

- Son más pequeños que los *data warehouses*
- Almacenan menos cantidad de información
- Tienen menos modelos de negocio
- Son utilizados por un número inferior de usuarios.

El hecho de que sea más fácil implementar un *data mart* que un *data warehouse* corporativo ha hecho que existan distintas visiones sobre cómo construir el almacén de datos. En este sentido Inmon y Kimball recomiendan dos estrategias completamente contrapuestas para construir el *data warehouse*:

- Inmon (1996) propone definir un *data warehouse* corporativo y a partir de él ir construyendo los modelos de análisis para los distintos niveles y departamentos de la organización; es decir, una estrategia de arriba abajo, desde la estrategia a lo más operativo.
- Kimball (2002) propone justo lo contrario, construyendo inicialmente distintos *data marts* que cubran las distintas necesidades de la organización, sin la necesidad de construir un *data warehouse*. El conjunto de *data marts* sería considerado como el *data warehouse* de la organización.

Lo cierto es que ambas estrategias son perfectamente válidas, tal como afirma Watson (2002):

Con la estrategia de definir un *data warehouse* corporativo definida por Inmon, el *data warehouse* es desarrollado en fases y cada una de las mismas debe ser diseñada para generar valor para el negocio. Se construye un *data warehouse* corporativo, del que se cuelga un *data mart* dependiente con una parte de la información del *data warehouse*. En fases posteriores se van desarrollando *Data marts* usando subconjuntos del *data warehouse*. Igual que los proyectos complejos, es caro, necesita mucho tiempo y es propenso al fracaso. Cuando tenemos éxito conseguimos un *data warehouse* integrado y escalable.

En cambio si optamos por la estrategia más común propuesta por Kimball, la de construir distintos *data mart*, el proyecto comienza con un *data mart* único al que posteriormente se irán añadiendo otros *data mart* que cubrirán otras áreas de negocio. Normalmente no requiere de grandes inversiones y es fácil de implementar, aunque conlleva algunos riesgos; de entre ellos, cabe destacar fundamentalmente dos: puede perpetuar la existencia del problema de “islas de información” y posponer la toma de decisiones que conciernen a la definición de criterios y modelos de negocio. Si seguimos esta estrategia debemos tener claro el plan de acción, es decir, qué áreas cubriremos y la integración de los distintos modelos. Esta estrategia se utiliza a veces como un paso previo al desarrollo de un *data warehouse* corporativo.

Además de las dos visiones, estrategias o arquitecturas propuestas por Kimball y Inmon referente a cómo organizar un *data mart*, el Instituto *Data Warehousing* (TDWI) distingue tres arquitecturas más. La tabla inferior recoge todas las arquitecturas que contempla TDWI:

| Arquitectura | Descripción |
|--|---|
| <i>Data marts</i> independientes | Cada <i>data mart</i> es construido y cargado individualmente; estos no están en conjugación ni comparten metadatos. |
| Conjunto de <i>data marts</i> | Es la solución defendida por Ralph Kimball (2002). En ella se propone la creación de varios <i>data marts</i> que comparten metadatos y que en su conjunto vienen a formar el almacén de datos de la organización. Se utiliza la estrategia ascendente (<i>Bottom-up</i>) donde se van añadiendo <i>data marts</i> de forma progresiva. Al final se debe lograr una conexión consistente entre todos ellos, lo cual suele tener sus problemas. |
| Almacén de datos general con <i>data marts</i> derivados | Es la solución propuesta por Bill Inmon (1996). En esta propuesta se crean los <i>data marts</i> después de haber creado un almacén de datos general. Se utiliza la estrategia descendente (<i>top-down</i>), donde se parte de una solución global (almacén de datos de la organización) para derivar en pequeñas soluciones para áreas específicas (<i>data marts</i>). Los problemas de consistencia entre los <i>data marts</i> no deben existir, pero esta propuesta tiene el inconveniente de que una organización necesita ir afrontando los costos de construcción de un almacén mediante su uso progresivo |
| Almacén de datos centralizado | Es similar a la solución anterior pero sin tener <i>data marts</i> derivados. Todos los usuarios acceden directamente al almacén de datos |
| Arquitectura federada | Corresponde a una solución para unificar múltiples almacenes de datos o <i>data marts</i> que ya existan en la organización. Una estrategia común en estos casos es construir un almacén de datos virtual donde todos los datos residan en sus fuentes originales pero estén integrados de una forma lógica haciendo uso de soluciones de software especiales para estos fines. |

Tabla 13 Arquitecturas propuestas por TDWI para el desarrollo de un *data warehouse*

Otro componente crítico de un *data warehouse* es el repositorio central de información o metadatos. Se trata de información adicional sobre los datos útil para dar significado a cada uno de los componentes y sus atributos que residen en el *data warehouse* (o *data mart*). La información que contienen los metadatos es útil para los departamentos de tecnología y los propios usuarios. Puede incluir definiciones de negocio, descripciones detalladas de los tipos de datos, formatos, de sus fuentes, de cómo se obtienen y otras características críticas para interpretar correctamente los datos y realizar las auditorías de calidad.

Las posibles utilidades de los metadatos dependen del tipo de usuario, por ejemplo el personal de los departamentos de Tecnología necesita saber los orígenes de la información: bases de datos de las que obtenemos los datos, qué transformaciones realizamos, criterios de filtros de información, nombre de las columnas y de las tablas, plazos de carga, utilización, etcétera.

En cambio los usuarios finales pueden necesitar saber las entidades y sus atributos, cómo han sido calculados, quiénes son los responsables de los datos, los informes disponibles y los flujos de distribución de la información, entre otras informaciones.

La construcción de los metadatos supone que se defina el significado de cada una de las tablas y cada uno de los atributos que se cargan en el *data warehouse*. Este es un punto complejo de todo proyecto, ya que obliga a que se definan los conceptos de negocio y se homogeneicen entre los distintos departamentos, filiales, etc. Obliga a que todos los componentes de la organización

hablen utilizando la misma terminología y con el mismo significado, lo cual no siempre es sencillo. Cuando alguien hable de “margen bruto” o “margen de contribución” deberá estar absolutamente definido para la organización. Evidentemente, organizaciones distintas tendrán normalmente definiciones distintas. (Cano, 2007)

Hasta aquí se han presentado las distintas definiciones de los *data warehouse*, sus posibles arquitecturas y otros conceptos necesarios para comprender la dimensión de estos sistemas de almacenamiento de datos. Para finalizar este apartado se presentan a continuación los factores que se deberían tener en cuenta cuando se está evaluando una alternativa tecnológica para la construcción de un *data warehouse* (Strange, 2001):

- Tamaño del *data warehouse*: es el volumen de datos que contiene el *data warehouse*.
- Complejidad de los esquemas de datos: si el modelo de datos es complejo, puede dificultar la optimización y el rendimiento de las consultas.
- Número de usuarios concurrente: éste es un factor determinante. Si distintos usuarios pueden lanzar consultas concurrentes (a la vez), el *data warehouse* debe gestionar sus recursos para poder dar respuesta a las distintas consultas.
- Complejidad de las consultas: si las consultas necesitan acceder a un número elevado de tablas y los cálculos a realizar son

complejos, podemos poner en dificultades al motor de la base de datos del *data warehouse*.

3.8 Herramientas del BI para el uso de los datos

Aunque esta tesis no se orienta hacia el diseño o estudio de las herramientas del *business intelligence* que explotan los datos almacenados en el almacén de datos, ampliamente desarrolladas en la literatura referente a la analítica del aprendizaje, se hace necesario presentar sus particularidades debido a que se utilizan en la última fase de la metodología que se presenta en el siguiente capítulo. En este sentido existen distintas tecnologías que nos permiten analizar la información que reside en un *data warehouse*, pero la más extendida es la OLAP (Procesamiento analítico en línea, del inglés *On Line Analytical Processing*) es la que se desarrollará por ser la empleada en esta tesis.

3.9 Principales herramientas del BI

Las principales herramientas (Eckerson & Howson, 2005) de *Business intelligence* se resumen en la siguiente tabla:

| Herramienta | Descripción |
|---|--|
| Generadores de informes | Utilizadas por desarrolladores profesionales para crear informes estándar para grupos, departamentos o la organización. |
| Herramientas de usuario final de consultas e informes | Empleadas por usuarios finales para crear informes para ellos mismos o para otros; no requieren programación. |
| Herramientas OLAP | Permiten a los usuarios finales tratar la información de forma multidimensional para explorarla desde distintas perspectivas y periodos de tiempo. |

| Herramienta | Descripción |
|---|---|
| Herramientas de <i>Dashboard</i> y <i>Scorecard</i> | Permiten a los usuarios finales ver información crítica para el rendimiento con un simple vistazo utilizando iconos gráficos y con la posibilidad de ver más detalle para analizar información detallada e informes, si lo desean. |
| Herramientas de planificación, modelización y consolidación | Permite a los analistas y a los usuarios finales crear planes de negocio y simulaciones con la información de <i>Business Intelligence</i> . Pueden ser para elaborar la planificación, los presupuestos, las previsiones. Estas herramientas proveen a los <i>dashboards</i> y los <i>scorecards</i> con los objetivos y los umbrales de las métricas. |
| Herramientas <i>datamining</i> | Permiten a estadísticos o analistas de negocio crear modelos estadísticos de las actividades de los negocios. <i>Datamining</i> es el proceso para descubrir e interpretar patrones desconocidos en la información mediante los cuales resolver problemas de negocio. Los usos más habituales del <i>datamining</i> son: segmentación, venta cruzada, sendas de consumo, clasificación, previsiones, optimizaciones, etc. |

Tabla 14 Resumen de las herramientas del BI

A continuación se hace especial énfasis en las herramientas OLAP que se han utilizado en esta tesis para realizar los análisis de los datos.

3.9.1 Herramienta de análisis OLAP

Los usuarios (Inmon W. H., 1996) necesitan analizar información a distintos niveles de agregación y sobre múltiples dimensiones: Por ejemplo, ventas de productos por zona de ventas, por tiempo, por clientes o tipo de cliente y por región geográfica. Los usuarios pueden hacer este análisis al máximo nivel de agregación o al máximo nivel de detalle. OLAP provee de estas funcionalidades y algunas más, con la flexibilidad necesaria para descubrir las relaciones y las tendencias que otras herramientas menos flexibles no pueden aportar.

Este tipo de análisis son llamados multidimensionales, porque facilitan el análisis de un hecho desde distintas perspectivas o dimensiones. Además es la

forma natural que se aplica para analizar la información por parte de los tomadores de decisiones, ya que los modelos de negocio normalmente son multidimensionales.

Durante el OLAP Council, conformado por algunos de los fabricantes de software OLAP, se resumieron las 12 reglas de Codd en lo que ellos llamaban el concepto FASMI que los productos OLAP deben cumplir. Este concepto viene de las siglas *Fast, Analysis, Shared Multidimensional Information*:

- *FAST* (Rápido): debe ser rápido, necesitamos lanzar consultas y ver los resultados inmediatamente.
- *ANALYSIS* (Análisis): debe soportar la lógica de negocio y análisis estadísticos que sean necesarios para los usuarios.
- *SHARED* (Compartido): tiene que manejar múltiples actualizaciones de forma segura y rápida.
- *MULTIDIMENSIONAL* (Multidimensional): tiene que proveer de una visión conceptual de la información a través de distintas dimensiones.
- *INFORMATION* (Información): debe poder manejar toda la información relevante y la información derivada.

La mejor forma de entender el concepto es representar gráficamente el OLAP en forma de cubos de información. Veamos un ejemplo:

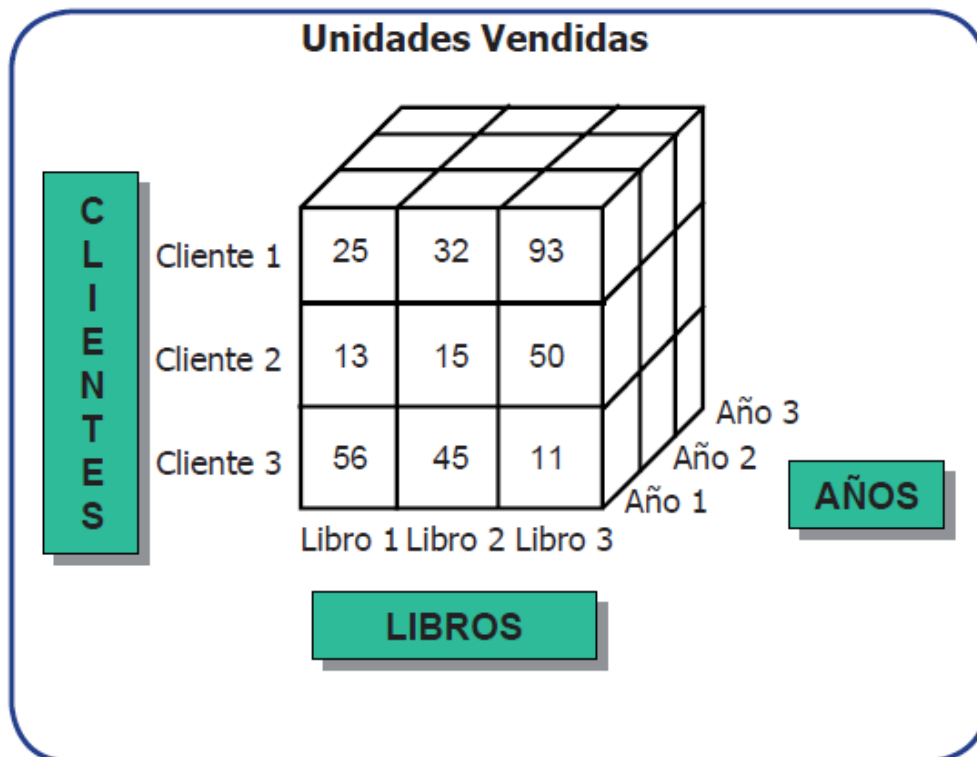


Figura 15 ejemplo de un cubo OLAP

En el cubo tenemos las unidades vendidas de cada uno de los libros, para los distintos clientes y en los distintos años. Este es el concepto de multidimensionalidad. Disponemos de las unidades vendidas de cada uno de los libros para cada uno de los clientes y en cada uno de los años: el contenido de un cubo individual son las ventas de un libro a un cliente en un año. Los contenidos de cada uno de los cubos individuales del cubo recogen lo que llamamos “hechos” (en nuestro ejemplo las unidades vendidas). En la actualidad, las soluciones OLAP permiten que cada una de los cubos individuales pueda contener más de un hecho.

Las herramientas OLAP nos permiten realizar muchas operaciones de forma sencilla tales como *slicing*, *dicing*, *roll-up* y *drill-down*. A continuación se explica qué son estas operaciones.

- *Slicing*: es la acción de rotar los cubos, es decir, cambiar el orden de las distintas dimensiones: En lugar de analizar por clientes, como en el caso anterior, quizás se esté interesado en analizarlo por libros, ya que los usuarios que lo quieren consultar son distintos y tienen distintas necesidades.

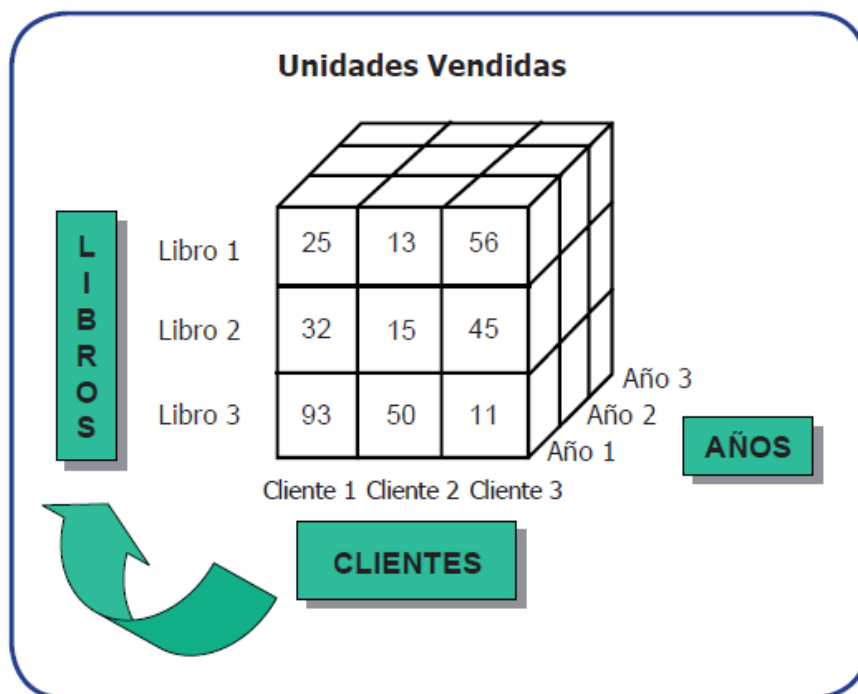


Figura 16 Ejemplo de cubo OLAP rotado

Tal como se puede observar se ha cambiado la dimensión “clientes” por la de “libros”.

- *Dicing*, es la operación de seleccionar sólo algunas de las celdas, por ejemplo: ¿Cuáles son las ventas al cliente 2, de los libros 1 y 2, en el año 1?

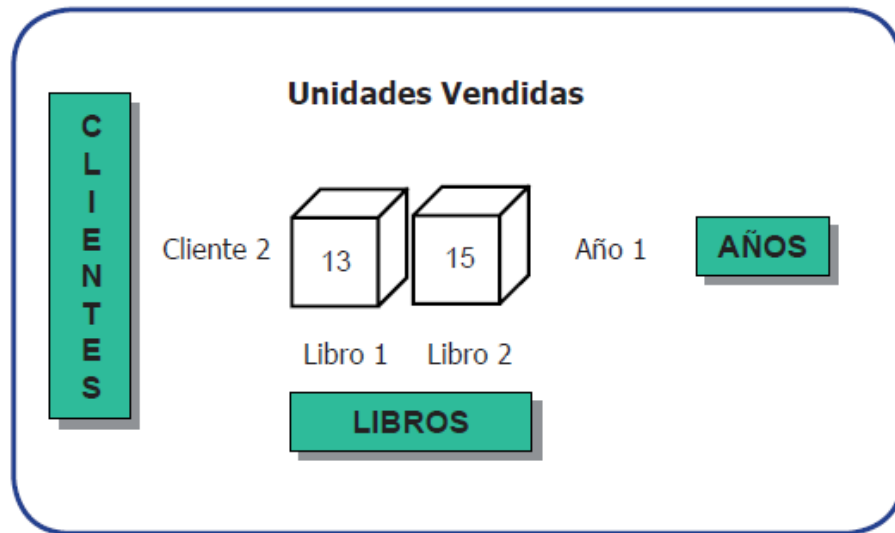


Figura 17 Ejemplo del proceso "dicing" en un cubo OLAP

- *Roll-up* es agrupar los resultados. Siguiendo con el ejemplo anterior, puede ser interesante conocer el total de libros:

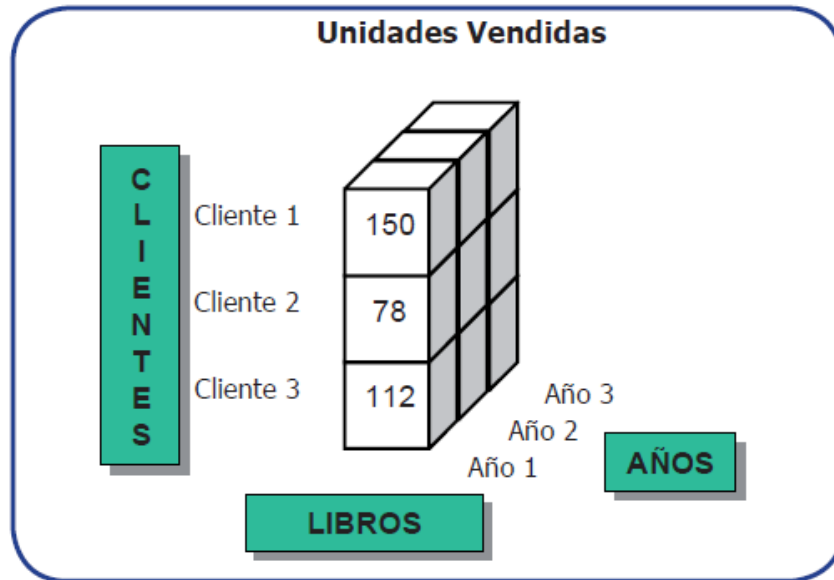


Figura 18 Ejemplo de roll-up en un cubo OLAP

- *Drill-down* es el contrario de *roll-up*, es decir, conocer el detalle de forma más granularizada. Según el ejemplo anterior, imagine que se tienen libros de dos materias distintas: El libro 1 y el libro 2 son de la materia A y el libro 3 de la materia B. Partiendo del cubo anterior de las ventas agregadas, bajamos a más detalle a través de la jerarquía “materias”. En ese caso se obtendría:

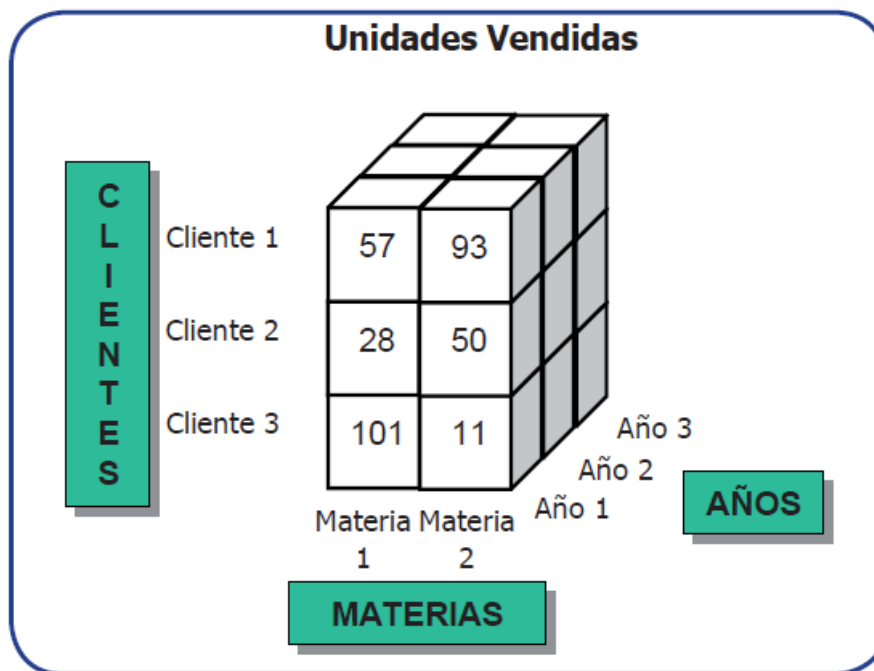


Figura 19 Ejemplo de drill-down en un cubo OLAP

Existen varios tipos de herramientas OLAP dependiendo de cómo acceden a los datos del *data warehouse* o de los *data marts* correspondientes: ROLAP, MOLAP y HOLAP

En ROLAP (*Relational OLAP*) se accede directamente a una base de datos relacional, habitualmente sobre un modelo lógico configurado en “estrella”. La principal ventaja es que no tiene limitaciones en cuanto al tamaño, pero es más lento que el MOLAP, aunque algunos productos comerciales nos permiten cargar cubos virtuales para acelerar los tiempos de acceso.

En MOLAP (*Multi Dimensional OLAP*) se accede directamente sobre una base de datos multidimensional (MDDDB del inglés *Multi Dimensional Data Base*). La ventaja principal de esta alternativa es que es muy rápida en los tiempos de

respuesta y la principal desventaja es que, si queremos cambiar las dimensiones, debemos cargar de nuevo el cubo.

Finalmente existe una solución híbrida entre las dos anteriores llamada HOLAP (*Hybrid OLAP*) donde se accede a los datos de alto nivel en una base de datos multidimensional y a los atómicos directamente sobre la base de datos relacional. En esencia utiliza las ventajas del ROLAP y del MOLAP.

4 *E-learning*, analítica del aprendizaje y su relación con el BI

En este apartado se revisarán los aspectos de los sistemas de *e-learning* que inciden en esta tesis y por lo tanto no se hará una revisión exhaustiva de las funcionalidades y características de este tipo de sistemas.

De forma introductoria se tiene que el propósito del BI en el *e-learning* es proveer a los usuarios información que les ayude a conseguir un mayor éxito y eficiencia (Lapuh Bele, Bele, Pirnat, & Anžin Lončarić, 2015) y el motivo de su empleo está debido a las pocas funcionalidades que proporcionan los entornos virtuales de aprendizaje para conseguir información de calidad y a tiempo de forma eficiente y eficaz.

4.1 Descripción de *e-learning*

El *e-learning* (del inglés *electronic learning*) ha significado un avance de los sistemas de educación a distancia (que se iniciaron con cursos por correspondencia y luego con apoyo audiovisual: radio, TV, vídeo...).

Inventados en el siglo XX con el propósito de proporcionar acceso a la educación a todos los que no podían acceder al sistema presencial, constituyen

un sistema especialmente satisfactorio para estudiantes con autodisciplina y perseverancia.

Asimismo ha evolucionado progresivamente en los últimos veinte años; a finales de la década de los 80 del siglo pasado fue precedido por el concepto de multimedia, la cual englobaba las imágenes, juegos y videos para ofrecer información y conocimientos a los estudiantes. A principios de los 90 -con el auge de internet- el hipertexto, el correo y otras herramientas como el chat se convirtieron en un medio para el aprendizaje y la enseñanza; esta etapa es conocida como “la primera ola de *e-learning*”.

Con la “segunda ola de *e-learning*” a principios del siglo XXI, cobraron auge los entornos virtuales de enseñanza y aprendizaje (EVE/A), sistemas *e-learning* o sistemas de gestión de aprendizaje (LMS, por sus siglas en inglés). Estos se han convertido en espacios para el intercambio de propuestas pedagógicas y reflexiones por parte de estudiantes y profesores.

Veamos cómo se puede definir el *e-learning*:

- Ampliación del entorno de aprendizaje, más allá de sus tradicionales límites físicos, geográficos y temporales, a través del uso de las tecnologías digitales en red. (Prieto & van de Pol, 2006)
- Aplicaciones y servicios que, tomando como base las Tecnologías de la información y la comunicación (TIC), se orientan a facilitar el proceso de enseñanza-aprendizaje. (Baelo Alvarez, 2009)

- Capacitación no presencial que, a través de plataformas tecnológicas, posibilita y flexibiliza el acceso y el tiempo en el proceso de enseñanza-aprendizaje mediante el uso de herramientas de comunicación síncrona y asíncrona. (García Peñalvo, 2005)

Nótese que en las tres definiciones proporcionadas arriba, sobre lo que se considera *e-learning*, el factor común es el uso de herramientas TIC como medio para aprender.

4.2 Seguimiento en el *e-learning*.

Desde el punto de vista del seguimiento y valoración del aprendizaje por el profesor en estos entornos informáticos se realiza, entre otras formas, mediante el seguimiento de estadísticas como por ejemplo el tiempo de utilización del aula virtual (Area & Adell, 2009). En este sentido el profesor toma decisiones antes y durante el desarrollo de la experiencia (Garrison & Anderson, 2005).

Las dos afirmaciones anteriores justifican el uso indicadores que le permitan al profesor medir el estado de sus estudiantes y de sus materias con el fin de tomar decisiones en el momento preciso y no al final de la acción formativa. Según lo que se plantea en esta tesis, estos indicadores se almacenan en forma de datos, que se generan ante cualquier interacción realizada en las herramientas TIC, los cuales deben ser seleccionados y procesados para convertirlos en información útil que al ser consumida por los profesores generen el conocimiento necesario para tomar las decisiones que ameriten en cada caso.

El lector en este punto podría suponer que las herramientas TIC que se emplean para la formación realizan estas funciones de forma automática y proveen a los profesores de los datos necesarios para tomar las decisiones. La realidad, tras años de experiencia en distintas herramientas TIC para la enseñanza y aprendizaje, es completamente distinta: aunque las herramientas TIC proveen de cierta información de seguimiento resumida en informes, éstos no resultan suficientes para atender necesidades particulares de conocimiento y deben buscarse alternativas para explotar estos datos en pos de conocer qué pasa, que pasó, qué pasará y por qué.

Si atendemos a las cifras de fracaso y abandono que proporciona Cebrián (2003) hay más de un 80% de fracaso en la gestión de cursos a distancia y más de un 60% de abandono de los mismos por parte de los estudiantes. Esto podría estar debido en parte por la falta de conocimiento sobre lo que está ocurriendo en el preciso momento en el que ocurre, y no al final de los periodos académicos más comunes en la educación (semestres, trimestres, año)

4.3 Los entornos virtuales de enseñanza y aprendizaje.

Salinas (2005) considera que los entornos virtuales de aprendizaje (entendidos como una herramienta TIC que se emplea en el *e-learning*) se basan en la idea de que la sofisticación de dichos entornos proporcionará la tan ansiada calidad del proceso enseñanza-aprendizaje. Esta sofisticación de la que habla Salinas (2005) podría encontrarse en la aplicación de tecnologías y métodos

propios del BI con el objetivo de disponer de información para la toma de decisiones.

Siguiendo con el análisis de los entornos virtuales de aprendizaje, Silva (2011) los define como:

“Aplicaciones informáticas diseñadas para facilitar la comunicación pedagógica entre los participantes de un proceso educativo, sea este completamente a distancia, presencial, o de una naturaleza mixta que combine ambas modalidades en diversas proporciones”

La dimensión pedagógica de un entorno virtual de aprendizaje está basada esencialmente en la interacción que se genera entre el docente y los alumnos (Salinas, Negre, Gallardo, Escandell, & Torrandell, 2007). La utilidad de métodos y herramientas de BI en este aspecto es doble: por un lado cualquier interacción realizada en registrada en forma de datos; por otro lado estos datos al ser explotados y combinados se convierten en información útil para conocer el estado de los estudiantes y poder realizar interacciones adicionales, por ejemplo acciones proactivas, y personalizadas con los estudiantes

Según Corrales (2008) la valoración de los aprendizajes en los entornos virtuales debe ser individualizada y enfocada en el alumno debiendo cumplir los siguientes criterios:

- Definición de objetivos.

- Retroalimentación durante todo el proceso de enseñanza-aprendizaje.
- Hacer uso de la tecnología para el diseño de pruebas y revisión de los resultados.
- Tomar en cuenta el ritmo y estilo de aprendizaje, así como las necesidades del alumno.

Analizando esta visión desde el punto de vista de lo que podría aportar el BI a su cumplimiento se tiene que:

- Los objetivos definidos conforman los requisitos de información que debe ser extraída del entorno virtual de aprendizaje mediante métodos de BI.
- La retroalimentación se hace eficaz cuando se dispone del conocimiento suficiente para evaluar la situación del estudiante. Este conocimiento puede ser proporcionado por las herramientas de análisis OLAP.
- Las tecnologías forman parte de un proyecto BI, en casi todas sus fases pero especialmente se hace visible en forma de herramientas OLAP de análisis, predicción segmentación, etc.
- Finalmente, criterios como ritmo y estilo de aprendizaje y necesidades del alumno pueden ser inferidas a partir de herramientas de BI.

4.3.1 *El entorno virtual de aprendizaje Moodle*

Aunque existen numerosos entornos virtuales de enseñanza esta tesis diseña y aplica un método para crear conocimiento sobre la plataforma tecnológica Moodle por este motivo se detalla en este apartado las características que inciden en el desarrollo de la solución sobre dicha plataforma.

Moodle (del inglés *Modular Object-Oriented Dynamic Learning Environment*) es un paquete de software para la creación de cursos y sitios web basados en internet. Es un proyecto en desarrollo que se distribuye como software libre (*Open Source*) bajo la licencia pública GNU (acrónimo recursivo, que en inglés significa: GNU no es Unix) (Moodle, 2015a).

El entorno virtual de enseñanza y aprendizaje Moodle es una aplicación que los educadores pueden utilizar para crear sus cursos y darle seguimiento a las acciones de los estudiantes. Este entorno cuenta con una comunidad de usuarios en línea que tiene como objetivos el desarrollo colaborativo de módulos y personalizaciones de la plataforma.

Moodle es una aplicación web diseñada bajo las ideas pedagógicas del constructivismo, que afirman que el conocimiento se construye en la mente del estudiante (Moodle, Pedagogia - MoodleDocs, 2015b). El corazón principal son los cursos, los cuales contienen actividades y recursos. Existe una veintena de actividades diferentes (foros, glosarios, wikis, tareas, cuestionarios, encuestas, etc.) y cada una de ellas puede ser personalizada a gusto del profesor. Este puede

instalarse en cualquier computadora en que pueda ejecutar PHP y soporte una base de datos MySQL.

Moodle mantiene un registro detallado de todas las actividades que realizan los estudiantes (Uribe, Melgar, & Bornacelly, 2007). Se registra cualquier “*click*” realizado por los usuarios durante la navegación por el entorno virtual y dispone de un sistema modesto de visualización de los registros, los cuales pueden ser filtrados por curso, participantes, días y actividad, algo insuficiente para ser empleado por aplicaciones informáticas de análisis. En este sentido, Moodle no almacena los registros en archivos de texto sino que su almacenamiento se realiza en una base de datos relacional facilitando la extracción y combinación de los datos que se almacenan. Hay aproximadamente 145 tablas interrelacionadas aunque no todas ellas contienen los datos que se necesitan para los análisis.

Según lo comentado arriba, Moodle consta de diferentes módulos, uno de estos permite explorar los accesos realizados por los usuarios a diferentes recursos y módulos del sistema. Los resultados obtenidos se muestran de forma tabular y pueden ser exportados a formato Excel o texto. En la figura inferior se ilustra lo anteriormente planteado.

Programación Web: Todos los participantes, Todos los días (Hora local del servidor)

Programación Web ▾ Todos los grupos ▾ Todos los participantes ▾ Todos los días ▾
 Todas las actividades ▾ Todas las acciones ▾ Mostrar en página ▾ [Conseguir estas bitácoras](#)

Mostrando 15666 registros

Página: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 ...157 (Siguiente)

| Hora | Dirección IP | Nombre completo del usuario | Acción | Información |
|------------------------------------|--------------|-----------------------------|-----------------------------------|------------------|
| mar 23 septiembre 2014, 10:57 p.m. | 127.0.0.1 | Admin Usuario | course report log | Programación Web |
| mar 23 septiembre 2014, 10:57 p.m. | 127.0.0.1 | Admin Usuario | course report log | Programación Web |
| mar 23 septiembre 2014, 10:56 p.m. | 127.0.0.1 | Admin Usuario | course view | Programación Web |

Figura 20 Ejemplo de informes de actividad que proporciona Moodle

Otro módulo, el relacionado a las calificaciones, permite mostrar de forma similar el resultado de los estudiantes en las diferentes actividades de evaluación. De igual manera puede ser exportado a formato Excel o texto. En la figura inferior se ilustra lo anteriormente planteado.

Reporte calificador

Grupos visibles 4301

| Apellido(s) ↑ Nombre | Dirección de correo | Programación Web | Trabajo Final. Orientaciones ↓ | Tarea#7 ↓ | Posicionamiento ↓ |
|----------------------|---------------------|------------------|--------------------------------|-----------|-------------------|
| | | - | 2.00 | Mal | |
| | | - | 2.00 | Mal | |
| | | - | 2.00 | Mal | |
| | | - | 2.00 | Mal | |
| | | - | 4.00 | Regular | |
| | | - | 2.00 | Mal | |
| | | - | 3.00 | Regular | |

Figura 21 Reporte calificador de Moodle

Los módulos anteriores, aunque muy útiles para los profesores, extraen sus datos directamente de la base de datos del entorno virtual lo que puede afectar negativamente su rendimiento, además de que no brindan facilidades gráficas en la visualización de la información, exploración y análisis de indicadores y reportes más personalizados.

Estos módulos que proveen información útil para conocer el estado de los estudiantes, tal como se ha comentado arriba, utilizan una mínima cantidad de los datos que se almacenan en las bases de datos de la plataforma los cuales están almacenados según el esquema normalizado que se puede consultar en su página web y que por motivos de extensión no se refleja en esta tesis (Moodle, 2015c). Se puede inferir según lo anterior que hay muchas posibilidades de obtener conocimiento a partir de estos datos más allá de lo que la plataforma virtual nos ofrece en sus módulos y es precisamente el eje sobre el cual se basa esta tesis.

4.4 Analítica del aprendizaje

El futuro de la educación se presagia en el uso de las nuevas tecnologías, dispositivos de cómputo para aprender en todas partes y en todo momento, aulas flexibles, y sistemas de apoyo continuo a la enseñanza y el aprendizaje. Estas tendencias se sustentan sobre un elemento muy importante: los grandes datos (*Big data*) y el análisis de estos.

En diversos campos como la medicina, el comercio y la investigación, por ejemplo, los datos son utilizados para tomar decisiones y mejorar los procesos; en la administración de organizaciones el término “Inteligencia de Negocio” responde

a este propósito. En la educación son muchas las posibilidades que se han abierto, en el aprovechamiento de los datos, con la introducción de los sistemas *e-learning*. El reto es entonces: ¿cómo utilizar los datos que se almacenan sobre los estudiantes y el uso que estos le dan a los contenidos en línea?

La Analítica del Aprendizaje es la medición, recolección, análisis y reporte de información acerca de los aprendices y sus contextos, para propósitos de entender y optimizar el aprendizaje y los ambientes en los cuales ocurre. La Analítica del Aprendizaje está enormemente comprometida con el mejoramiento y éxito del aprendiz (Siemens, y otros, 2011).

La analítica de aprendizaje, término que apareció por primera vez en 2009 en Horizon Reports publicación de The New Media Consortium, viene a convertirse en una herramienta imprescindible para estos tiempos, donde el aprendizaje electrónico ha ganado terreno. Cuando un estudiante realiza tareas en línea se pueden capturar innumerables datos (log), datos útiles, que posteriormente a ser analizados, la información y los conocimientos extraídos de ellos pueden servir para crear y mejorar recursos de aprendizaje. El análisis de las huellas dejadas por un aprendiz cuando interactúa con un sistema de enseñanza/aprendizaje puede reflejar: lo que le gusta, lo que no le gusta, lo que le queda por dominar, incluso en qué momento del día asimila mejor los contenidos.

Otras definiciones de Analítica de Aprendizaje son:

- La idea principal del *Learning analytics* (Analítica de Aprendizaje) se basa en analizar los datos que deja tras de sí el estudiante en los

procesos de formación y aprendizaje. El objetivo es utilizar dicha información para mejorar el propio proceso de aprendizaje. Los datos se pueden obtener a través de las distintas tecnologías que se utilizan actualmente en la formación; dichas tecnologías pueden ser corporativas (por ejemplo sistemas LMS tipo Moodle), propietarias de los alumnos (móvil, tableta, portátil) e incluso sociales (redes sociales, *blogs*, etc.).

- El uso de datos y modelos para predecir el progreso de los estudiantes y el rendimiento, así como la habilidad para actuar sobre esta información (Van Barneveld, Arnold, & Campbell, 2012).
- La Analítica del aprendizaje se refiere a la interpretación del gran número de datos dejados por los estudiantes producto de su progreso en el aprendizaje para evaluar el progreso académico, predecir actuaciones futuras e identificar los problemas potenciales. Los datos son recogidos por las acciones explícitas de los estudiantes, como pueden ser: la compleción de tareas, realización de exámenes, o desde acciones tácitas como la interacción social en línea, actividades extracurriculares, interacciones en foros de discusión y otras actividades que no forman parte directamente del proceso de aprendizaje del estudiante. El objetivo de la analítica del aprendizaje es permitir a la escuela y a los profesores diseñar las oportunidades educativas en dependencia de las necesidades y habilidades de cada estudiante. La analítica del

aprendizaje pretende aprovechar los últimos adelantos en la minería de datos, su interpretación y el modelado de las actividades de aprendizaje para llevar de una forma más eficaz la atención a las individualidades de cada estudiante (New Media Consortium, 2013)

Existe una amplia lista de investigaciones donde se ha aplicado el análisis de datos en el contexto del aprendizaje. En este sentido Romero y Ventura (2010) clasifican las investigaciones relacionadas con la analítica de aprendizaje en las siguientes once categorías:

- 1- Analizar y visualizar datos importantes para la toma de decisiones.
- 2- Ofrecer retroalimentación a los instructores.
- 3- Proveer recomendaciones para los estudiantes.
- 4- Predecir el desempeño en las tareas.
- 5- Mejorar o definir modelos de estudiantes.
- 6- Detectar conductas indeseables, como la baja motivación.
- 7- Identificar grupos o tipos de estudiantes teniendo en cuenta variables de interés.
- 8- Analizar el comportamiento social.
- 9- Desarrollar mapas de conceptos para modelado de conocimiento.
- 10- Asistir en procesos de diseño y construcción de herramientas o recursos educativos.

11- Ayudar a instructores y estudiantes al cumplimiento de planes y cronogramas de actividades de acuerdo a sus objetivos de aprendizaje.

Otros autores proponen algunos niveles que se deben tener en cuenta para la aplicación de la analítica de aprendizaje. Resumen, Exportación y Análisis, son los propuestos por Claros y Cobos (2013). Resumen se refiere a mostrar información representativa del sistema para lograr una entendimiento general de las acciones que en este se desarrollan. Exportación está relacionada a la capacidad de volcar los datos del sistema en formatos compatibles con herramientas externas que permitan aplicar métodos manuales de análisis más complejos. Finalmente, el último nivel permite analizar los indicadores específicos del sistema desde diferentes perspectivas.

Relacionado con lo anterior y con una visión más general, el Instituto SAS (empresa líder en el desarrollo de soluciones de inteligencia de negocio) presenta ocho niveles para la aplicación de las analíticas (SAS Institute, 2008):

- 1- Reportes estándares,
- 2- Reportes personalizados,
- 3- Procesamiento analítico en línea,
- 4- Alertas,
- 5- Análisis estadístico,
- 6- Análisis de pronósticos,
- 7- Modelos predictivos, y
- 8- Optimización.

El término de Analítica de Aprendizaje se ha bifurcado según el contexto de aplicación: la referente al entorno académico y la vinculada al entorno industrial o empresarial. Estas diferencias se evidencian en la siguiente idea presentada por Van Barneveld et al (2012):

“Desde la perspectiva del entrenamiento en la industria, la Analítica de Aprendizaje apunta a dos aspectos fundamentales –aprendizaje efectivo y excelencia operacional– que evidencian como el entrenamiento y el aprendizaje se están alineando a los objetivos de la organización.”

“La Analítica de Aprendizaje en la academia se centra en el estudiante, recogiendo datos de los sistemas de administración de cursos y otros sistemas de información con el objetivo de manejar el éxito estudiantil y procesos de intervención”

5 Soluciones similares

En esta sección se describen diferentes trabajos de investigación y herramientas que tienen características similares con la propuesta de solución que se exhibe en el presente informe. A continuación se presentan las soluciones similares que se encuentran en el mercado.

5.1 ViMoodle

ViMoodle es un sistema adaptativo e interactivo de analítica visual para los datos de Moodle. Para trabajar con este software se deben realizar dos operaciones principales; la primera consiste en configurar la conexión a los datos de Moodle (host, base de datos, usuario y contraseña) y la segunda se refiere a la visualización, filtrado y búsqueda de información. Esta herramienta, como su nombre lo indica, se apoya fuertemente en gráficas, iconos y representaciones visuales diversas. Los datos que se analizan, por ejemplo, se refieren al acceso a los diferentes cursos, foros, recursos y cuestionarios; el objeto de análisis lo conforman los profesores y estudiantes (Gómez Aguilar, Theron, & & García Peñalo, 2008).

5.2 Extensión del navegador para análisis visual

Se trata de una extensión, para los navegadores Mozilla Firefox y Google Chrome, que se incrusta dentro del escenario de trabajo del usuario –en la plataforma Moodle- con el objetivo de realizar un análisis visual de variables relacionadas a la interacción con el sistema. La herramienta se basa en una arquitectura cliente servidor y se nutre de un servicio web implementado en el servidor. El servicio web brinda los análisis y estos son embebidos en forma de

*widget*¹ o *portlet*² dentro del espacio de trabajo del usuario autenticado. Cada estudiante sólo tiene acceso al análisis de su actividad; los profesores pueden visualizar la actividad de cada uno de sus estudiantes y compararlas entre sí (Leony, Pardo, Valentin, Quiñones, & Delgado Kloos, 2013).

5.3 **PADA: Dashboard of Learning Analytics of Dyslexia in Adults**

En este trabajo se presenta un cuadro de mando integral, llamado PADA, para la visualización y la inspección de determinados problemas de lectura relacionados con la dislexia y las características del aprendizaje y la cognición de los estudiantes afectados. PADA es una herramienta basada en la web, diseñada para crear gráficos descriptivos necesarios para una mejor comprensión del estado del estudiante. PADA muestra a los estudiantes y profesores un modelo de usuario con el fin de ayudar a aumentar la concientización y promover la reflexión y la auto-regulación de las dificultades, las estrategias de aprendizaje y déficits cognitivos en la lectura. La herramienta se nutre de los datos obtenidos a partir de los log de un EVE/A. Estos log son extraídos hacia un almacén de datos para, mediante la librería JQuery y el lenguaje PHP, obtener un análisis visual y descriptivo de algunos indicadores relacionados a los estudiantes. También son

¹ Un *widget* es una pequeña aplicación que suele implementarse para realizar tareas de utilidad. Su objetivo es ser contenidos de otras aplicaciones.

² Los *portlets* son componentes modulares de las interfaces de usuario, gestionadas y visualizadas en un portal web

aplicadas técnicas de minería de datos para encontrar correspondencia entre los indicadores de los aprendices (Mejia, Bull, Vatrappu, Florian, & Fabregat, 2013).

5.4 Cuadro de mando integral portable

Esta solución consiste en una herramienta portable para visualizar información mediante cuadros de mando integral que acceden a los datos de diferentes entornos virtuales de aprendizaje. La herramienta ofrece vistas estáticas y dinámicas. Las vistas estáticas se refieren a la estructura de los cursos, las herramientas y los recursos de aprendizaje; la segunda a la colaboración entre los usuarios y el uso que estos le dan a las herramientas y los recursos de aprendizaje. Esta herramienta se integra a sistemas *e-learning* en forma de widget para lo cual necesita que la plataforma implemente cierta interfaz de programación (Vozniuk, Govaerts, & Gillet, 2013).

5.5 MOCLog

MOCLog (del inglés: *Monitoring Online Courses with log data*) es una herramienta para analizar y monitorear los datos que almacena el entorno virtual de enseñanza y aprendizaje Moodle. Esta analiza información referente al uso de las diferentes herramientas de Moodle como son las wiki, foros, chat, acceso a los recursos y a las tareas. Esta solución tiene una característica particular: presta atención a la interacción estudiante-estudiante, estudiante- profesor y las diferentes evaluaciones de los estudiantes en la plataforma. Utiliza bondades comunes como es el reporte de información y el uso de gráficas para ayudar a la comprensión de los diferentes indicadores (Mazza, Bettoni, Faré, & Mazzola, 2012).

5.6 GISMO

Se trata de una herramienta que monitorea las actividades de los estudiantes del entorno virtual de enseñanza/aprendizaje Moodle. La herramienta extrae datos de seguimiento del log de Moodle y representa los resultados en formato gráfico. Asimismo provee a los profesores la información en varios formatos diferentes tales como reportes y gráficos. (GISMO, 2010)

5.7 Otras soluciones

La tabla inferior recoge otras soluciones de análisis centradas en entornos virtuales de aprendizaje.

| Solución | Descripción |
|------------------|--|
| CourseVis | Recolecta los datos del registro de datos del entorno virtual de aprendizaje para posteriormente realizar informes. (Mazza & Dimitrova, 2007) |
| Sinergo/ColAT | Ofrece vistas interpretativas de la actividad desarrollada por estudiantes en entornos colaborativos de aprendizaje grupal y se asemeja al proceso de aprendizaje de los estudiantes. (Avouris, Komis, Fiotakis, Margaritis, & Voyiatzaki, 2005) |
| MATEP | Se trata de una interfaz web para el instructor que provee una serie de reportes según sus requerimientos. (Zorrilla & Álvarez, 2008) |
| Google Analytics | Se trata del motor analítico de Google que puede ser implementado en los entornos virtuales de aprendizaje. |

Tabla 15 Otras soluciones similares estudiadas

6 Herramientas de desarrollo

El desarrollo de cualquier producto tecnológico implica la utilización de herramientas que agilicen, mejoren y aporten calidad a la solución. Para dar cumplimiento al objetivo de la presente investigación fue necesario apoyarse en

diversas herramientas; el grupo de ellas más significativo fue el perteneciente a la plataforma de inteligencia de negocio de código abierto Pentaho.

La selección de Pentaho se debe a que esta cubre de forma amplia las necesidades de reporte y análisis de datos. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación basado en el mismo lenguaje. Esto hace que Pentaho sea una solución muy flexible para abarcar las, tanto típicas como sofisticadas, necesidades propias de una organización.

6.1 MySQL

Es un sistema de gestión de base de datos relacional de propósito general, muy rápido, robusto, multiusuario y multihilo. Su licencia es dual, con una versión propietaria y una comunitaria liberada bajo licencia GPL. Entre sus características principales están el ser probado con un amplio rango de compiladores diferentes y la implementación de sus funciones SQL utiliza una librería altamente optimizada, lo que contribuye a su rapidez. Este gestor, además de soportar grandes bases de datos, es la propuesta por la Suite Pentaho para el desarrollo de soluciones de inteligencia de negocio. (Bouman & Van Dongen, 2009).

6.2 Pentaho Data Integration

La herramienta de integración de datos de Pentaho entrega una poderosa capacidad de extracción, transformación y carga de datos usando una solución innovadora basada en metadatos. Esta herramienta provee un ambiente intuitivo, gráfico, y de tipo “*drag and drop*”, además de una arquitectura escalable, basada en altos estándares (Pentaho Community, 2012).

Algunas de sus características destacadas son el contar con una librería completa de más de 100 objetos de mapeo; ser 100% Java, lo que garantiza ser multiplataforma; soporta una amplia cantidad de fuentes de datos, incluyendo más de 30 plataformas propietarias y *open source*, archivos planos, documentos Excel y muchos más; está basada en repositorios, lo que facilita la reutilización de componentes de transformación, colaboración y conexiones; e incorpora además un calendario programado de transformaciones y trabajos (*scheduler*).

6.3 Pentaho Report Designer

Pentaho Report Designer es una aplicación gráfica, basada en el proyecto JFreeReport, para crear y publicar reportes para la plataforma Pentaho. Esta aplicación brinda las facilidades necesarias para lograr diseños de alta calidad que podrán ser publicados y consultados vía web por quienes lo requieran (Pentaho BI, 2013).

Entre las características de esta herramienta están: disponer de un diseñador gráfico de tipo “*drag and drop*” que proporciona un control completo de acceso a datos, diseño, agrupación, cálculos y gráficos; incorpora una ayuda paso a paso, que guía al usuario durante el proceso de diseño; provee plantillas predefinidas de alto nivel para agilizar el proceso de diseño; y permite además que los reportes sean publicados en el servidor de Pentaho donde los usuarios finales podrán verlos a través de su cuenta de Pentaho Analyzer.

6.4 Pentaho Schema Workbench

Schema Workbench es una herramienta visual para el diseño y prueba de cubos de datos para ser interpretados por el servidor Mondrian. Mondrian es un motor de procesamiento analítico en línea (OLAP) escrito en el lenguaje Java. Este motor recibe consultas MDX desde la vista frontal del usuario final, lee datos desde base de datos relacionales (implementación ROLAP) y responde con resultados multidimensionales (Hyde, 2007).

Schema Workbench tiene como objetivo facilitar la tarea de diseñar cubos OLAP. Su sencilla interfaz permite construir un XML con el diseño del cubo a través de opciones lógicas e intuitivas que no requieren conocer de manera avanzada este formato. Algunas de las características de esta aplicación son el tener un diseño intuitivo en la construcción de esquemas OLAP y de permitir crear, editar y publicar esquemas multidimensionales para ser utilizados por las herramientas de visualización de Pentaho.

6.5 Pentaho Analyzer

La herramienta de análisis de Pentaho brinda un poder de análisis significativo a los usuarios finales, permitiéndoles obtener ideas y entender las necesidades del negocio. Esta herramienta permite explorar de forma libre y completa la información a través de un análisis interactivo con distintos grados de profundidad (análisis OLAP); presenta multidimensionalmente los datos, lo que facilita seleccionar las dimensiones y medidas que se desean explorar; incorpora capacidades gráficas para presentar la información; y proporciona a los usuarios la posibilidad de diseñar sus propias consultas multidimensionales.

Esta herramienta incorpora un componente para la creación de reportes a la medida. Es un componente de interfaz web interactiva “*drag and drop*” para que el usuario genere sus propios reportes dependiendo de sus necesidades. Finalmente brinda un componente para crear tableros de control (*dashboard*) de forma rápida y fácil con el objetivo de controlar las estadísticas e indicadores de la organización.

7 Conclusiones

Hoy en día se tienen muchos entornos virtuales de aprendizaje que están diseñados con el objetivo de servir como plataforma a los integrantes de un proceso educativo y cuyas principales bondades se enmarcan en las herramientas para la creación de cursos, usuarios y la comunicación entre estos pero no en el seguimiento de lo que allí ocurre. Por muchos informes que puedan proveer las funcionalidades de los entornos virtuales de enseñanza éstos no suelen resultar suficientes para conocer el estado actual, pasado y futuro de los aspectos críticos de la educación on-line.

Por este motivo en este capítulo se ha estudiado cómo se crea y transforma el conocimiento y una vez se han asentado las bases del conocimiento organizativo se han estudiado métodos y herramientas propios del *Business intelligence* para extraerlo desde su estado más atómico (dato) combinarlo y contextualizarlo (información) y hacerlo llegar a los responsables, que en su acción de consumirlo crea un conocimiento útil para la toma de decisiones.

Respecto a las soluciones similares estudiadas, ninguna de ellas dispone de la flexibilidad que ofrecen las herramientas y métodos de BI en lo que se refiere a personalizaciones según la distinta realidad de las organizaciones, por mucho que se parezcan o pertenezcan al mismo ámbito, en este caso educativo. Asimismo muchas de las soluciones explotan directamente los datos desde los servidores OLTP con el consecuente riesgo de sobrecargarlos y aun cuando vuelquen los datos a otro servidor de forma periódica (a veces externo a la institución con el consiguiente problema de privacidad de los datos), se omiten todas las acciones de depuración y homogeneización que conducen a disponer de datos de calidad, confiables por los usuarios finales del sistema.

Finalmente se presentan las herramientas que se han empleado para aplicar el método que se propone en el siguiente capítulo.

Capítulo 3 - Propuesta de método para crear conocimiento a partir de los datos residentes en entornos virtuales de aprendizaje

1 Introducción

En este capítulo se propone un método de BI que tiene como propósito intensificar el uso de los datos, almacenados en un entorno virtual de aprendizaje, para la creación eficaz y eficiente de conocimiento útil para la toma de decisiones.

Para ello se han adaptado metodologías y herramientas propias del BI aplicadas sobre los datos almacenados en entornos virtuales de aprendizaje con el objetivo final de poder tomar decisiones a partir del conocimiento que se ha creado durante el uso de estos datos.

Para diseñar el método se han tenido en consideración los siguientes criterios:

- Los objetivos y resultados esperados en cada fase se deben distinguir fácilmente y deben ser sencillos de comprender.
- Debe ser capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.

- Debe reducir la resistencia al cambio involucrando al usuario final.
- Debe utilizar estructuras de datos sencillas de interpretar y analizar incluso por usuarios no expertos.
- Debe ser independiente de las herramientas que se utilicen para su implementación.
- Debe ser independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- Cuando se culmine una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo la siguiente.
- Cuando se culmine la última fase debe ser posible realimentarlo mediante el nuevo conocimiento adquirido.
- Se debe poder aplicar independientemente de la estrategia empleada en el almacén de datos.

Con estos requisitos en mente se desarrolla en los siguientes apartados el método seguido.

2 Esquema del método

Este método consta de fases diferenciadas descritas por métodos, materiales, resultados, utilidad, aporte al conocimiento y actores distintos para cada una de ellas. Estas fases se detallan de forma específica en los apartados posteriores de este capítulo y en este apartado se aborda el método de forma íntegra y esquemática.

La mejor forma de describir y entender el método es mediante un esquema gráfico que muestre todos los componentes y los relacione según las dependencias existentes. La figura de debajo pretende cumplir con este objetivo.

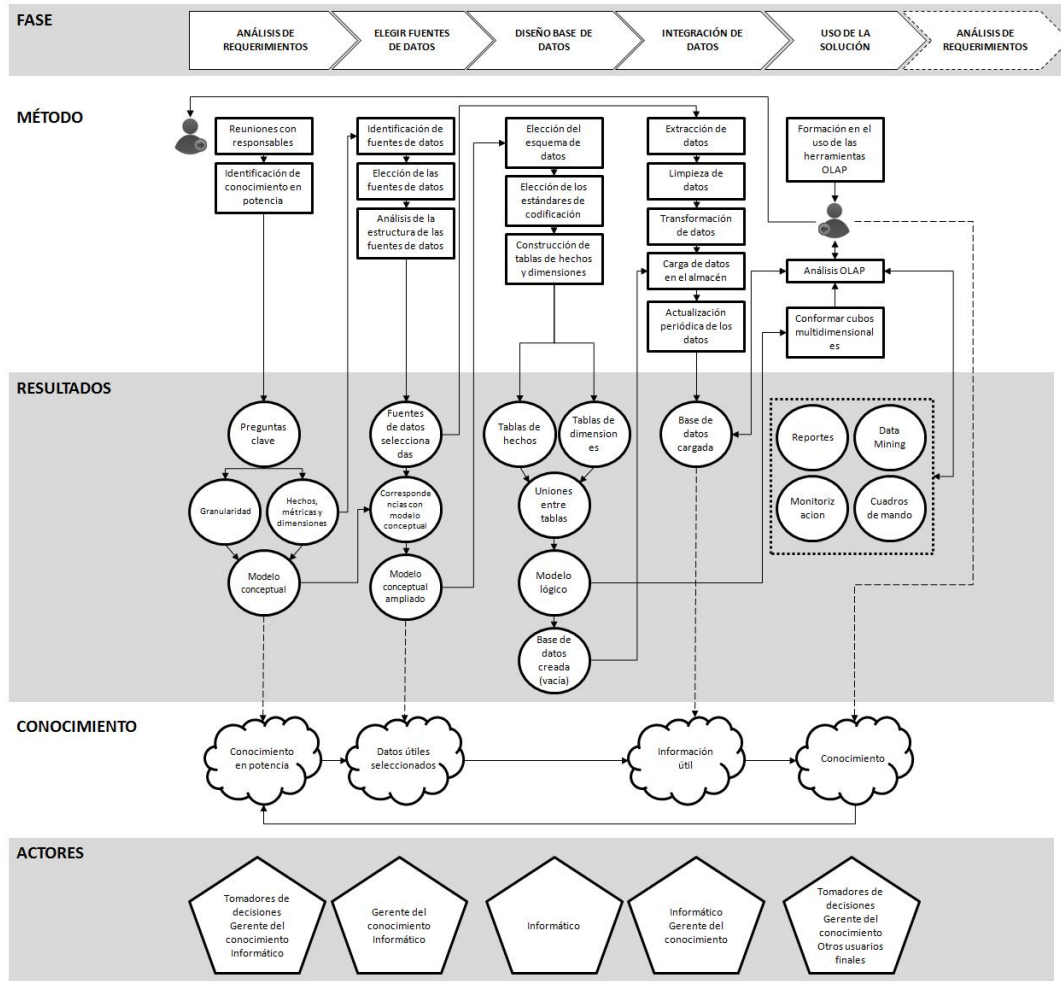


Figura 22 Esquema del método propuesto

De forma más detallada, la tabla inferior recoge el proceso mostrado en la figura anterior y lo amplía añadiendo otras visiones, relativas a la utilidad de cada

fase, a distintas visiones de la gestión del conocimiento y a su relación con la eficiencia y eficacia de cada una de las fases.

| Fase -> | Análisis de requerimientos (optativo) | Elección de las fuentes de datos | Diseño del almacén de datos | Integración de datos | Uso de la solución |
|---------|--|---|--|--|--|
| Método | <ul style="list-style-type: none"> • Reuniones con los responsables • Identificación de conocimiento en potencia mediante preguntas clave • Validar la existencia de las fuentes de datos | <ul style="list-style-type: none"> • Identificación de las fuentes de datos según los requerimientos detectados si los hubiera o, en caso contrario, según una estimación de datos útiles para generar nuevo conocimiento. • Análisis de la estructura de las fuentes de datos • Elección de las posibles fuentes de datos estructuradas y no estructuradas • Establecimiento de correspondencias entre las fuentes de datos y el modelo conceptual • Diseño de métricas a partir de las fuentes de datos • Definición del nivel de granularidad o resolución | <ul style="list-style-type: none"> • Elección del esquema de datos (arquitectura del almacén) • Construcción de tablas de hechos y dimensiones • Elección de los estándares de codificación | <ul style="list-style-type: none"> • Extracción de datos unitarios desde sus fuentes • Limpieza de los datos unitarios: depurar, corregir, estandarizar, relacionar y consolidar. • Transformación de los datos unitarios: cambio de formatos, sustitución de códigos, cálculo de valores agregados y derivados • Integración de los datos unitarios en el almacén de datos para conformar los datos útiles • Actualización de los datos útiles | <ul style="list-style-type: none"> • Reportes (consultas ad-hoc, listados por correo, informes estándar y personalizados) • Monitorización y cuadros de mandos • Análisis • <i>Data Mining</i> • Formación en el uso de las herramientas OLAP |

| Fase -> | Análisis de requerimientos (optativo) | Elección de las fuentes de datos | Diseño del almacén de datos | Integración de datos | Uso de la solución |
|------------|--|--|--|--|--|
| Materiales | <ul style="list-style-type: none"> • Salas físicas • Herramientas TIC de comunicación | <ul style="list-style-type: none"> • Bases de datos OLTP • Lenguajes para explorar las bases de datos • Mapa de la arquitectura de las bases de datos | <ul style="list-style-type: none"> • Herramientas de diseño de bases de datos multidimensionales • Bases de datos DM / DW | <ul style="list-style-type: none"> • Herramientas ETL o consultas a bases de datos • BBDD de datos unitarios • Plataformas • Protocolos de comunicación • Juegos de caracteres • Tipos de datos | <ul style="list-style-type: none"> • Herramientas OLAP |
| Resultados | <ul style="list-style-type: none"> • Identificar preguntas clave • Identificar indicadores (hechos y métricas) y dimensiones • Identificar el nivel de granularidad • Definir el modelo conceptual | <ul style="list-style-type: none"> • Fuentes de datos seleccionadas • Correspondencia entre las fuentes de datos y el modelo conceptual identificado. • Métricas conformadas • Modelo conceptual ampliado. | <ul style="list-style-type: none"> • Modelo lógico multidimensional del almacén de datos • Tablas de hechos • Tablas de dimensiones • Uniones entre tablas | <ul style="list-style-type: none"> • Datos útiles con calidad, • Estandarizados • Actualizados • Almacenados en un modelo multidimensional • Optimizados para responder a los requerimientos detectados en el paso 1 o estimados en el paso 2 | <ul style="list-style-type: none"> • Información combinada y presentada de forma que se genere conocimiento |

| Fase -> | Análisis de requerimientos (optativo) | Elección de las fuentes de datos | Diseño del almacén de datos | Integración de datos | Uso de la solución |
|------------------------|--|---|--|---|---|
| Preguntas que responde | <ul style="list-style-type: none"> • ¿Qué conocimiento se quiere generar? • ¿Qué resolución se necesita para medir el conocimiento requerido? • ¿Cómo se debería estructurar el sistema de datos? | <ul style="list-style-type: none"> • ¿Dónde están los datos unitarios que se necesitan para generar el conocimiento? • ¿Cómo se obtienen los datos unitarios que se necesitan para generar el conocimiento? • ¿Cómo combinar los datos unitarios para generar el conocimiento? • A partir de los datos unitarios ¿cómo se van a conformar las métricas para generar el conocimiento? • ¿Cuáles son las correspondencias entre el modelo conceptual y las fuentes de datos? | <ul style="list-style-type: none"> • ¿Cómo y dónde se almacenarán los datos para generar el conocimiento deseado? | <ul style="list-style-type: none"> • ¿Cómo se extraen y transforman los datos unitarios? • ¿Dónde y cómo se almacenan los datos útiles? | <ul style="list-style-type: none"> • ¿Qué pasó? • ¿Qué pasa ahora? • ¿Por qué pasó? • ¿Qué pasará? • ¿Qué decisión se tomará según las respuestas a las consultas anteriores? • ¿Qué nuevas preguntas hay que responder a partir de los resultados? |

| Fase -> | Análisis de requerimientos (optativo) | Elección de las fuentes de datos | Diseño del almacén de datos | Integración de datos | Uso de la solución |
|----------------------------------|--|--|--|--|--|
| Utilidad | <ul style="list-style-type: none"> • Detectar conocimiento en potencia y la resolución para su medición. | <ul style="list-style-type: none"> • Identificar las fuentes de los datos unitarios • Definir cómo se deben extraer • Conformar las métricas según las fuentes de datos detectadas • Establecer las correspondencias entre las fuentes de datos y el modelo conceptual • Ampliar el modelo conceptual | <ul style="list-style-type: none"> • Definir cómo y dónde se almacenarán los datos útiles | <ul style="list-style-type: none"> • Datos filtrados • Corregidos • Formateados • Estandarizados • Calculados • Combinados | <ul style="list-style-type: none"> • Conocimiento actual, pasado, futuro y nuevo para tomar decisiones |
| Actores | <ul style="list-style-type: none"> • Tomadores de decisiones • Gerente del conocimiento • Informático | <ul style="list-style-type: none"> • Gerente del conocimiento • Informático | <ul style="list-style-type: none"> • Informático | <ul style="list-style-type: none"> • Informático • Gerente del conocimiento | <ul style="list-style-type: none"> • Tomadores de decisiones • Gerente del conocimiento • Otros usuarios finales (i.e. alumnos) |
| Datos Información y conocimiento | Conocimiento en potencia | Datos unitarios | Datos útiles | Información útil | Conocimiento (creado y transferido) |

| Fase -> | Análisis de requerimientos (optativo) | Elección de las fuentes de datos | Diseño del almacén de datos | Integración de datos | Uso de la solución |
|--|--|--|---|--|---|
| Datos Información y conocimiento según Davenport y Prusak (1998) | Conocimiento requerido <ul style="list-style-type: none"> • Contextualización • Categorización | Datos | | Información <ul style="list-style-type: none"> • Contextualización • Categorización • Cálculo • Corrección • Condensación | Conocimiento <ul style="list-style-type: none"> • Comparación • Consecuencias • Conexiones |
| Conocimiento según Nonaka y Takeuchi (1995) | Externalización Tácito → Explícito <ul style="list-style-type: none"> • Contexto • Relevancia • Utilidad | | | Combinación Explícito → Explícito | Internalización Explícito → Tácito |
| Indicador de eficiencia | <ul style="list-style-type: none"> • Tiempo en reuniones | <ul style="list-style-type: none"> • Rendimiento de sistemas de origen • Tiempos y rendimiento de procesos ETL • Tiempo de análisis de las fuentes de datos | <ul style="list-style-type: none"> • Tiempo y rendimiento de procesos ETL. • Tiempos de ejecución de todas las tareas de esta fase. | <ul style="list-style-type: none"> • Rendimiento de sistemas de origen. • Tiempo y rendimiento de procesos ETL. • Tiempo de diseño del motor ETL. | <ul style="list-style-type: none"> • Tiempo de interpretación de la información • Tiempo de uso de las herramientas OLAP. • Número de informes creados (auto y manual) |

| Fase -> | Análisis de requerimientos (optativo) | Elección de las fuentes de datos | Diseño del almacén de datos | Integración de datos | Uso de la solución |
|-----------------------|--|--|--|--|---|
| Indicador de eficacia | <ul style="list-style-type: none"> • Actores útiles seleccionados • Preguntas útiles y/o transversales identificadas | <ul style="list-style-type: none"> • Personal informático adecuado seleccionado • Fuentes de datos necesarias identificadas • Indicadores de alta utilidad y/o transversalidad escogidos • Granularidad necesaria seleccionada | <ul style="list-style-type: none"> • Codificación elegida • Personal informático adecuado seleccionado | <ul style="list-style-type: none"> • Personal informático adecuado seleccionado • Tiempos de actualización necesarios escogidos • Calidad de los datos validada | <ul style="list-style-type: none"> • Usuarios finales formados • Sistemas automatizados útiles preparados • Plantillas visuales creadas. • Informes creados (auto y manual) |

Tabla 16 Resumen del método propuesto según sus fases y los materiales, métodos, resultados, actores y utilidad de cada fase

A continuación se detalla cada una de las fases y características presentadas en la tabla superior.

3 Descripción de las fases

Este apartado tiene como objetivo describir las fases y entender su utilidad. En este sentido no se detallan los procesos ni los resultados debido a que éstos se abordan en el siguiente apartado del presente capítulo.

3.1 Análisis de requerimientos

El proceso se inicia mediante un análisis de requerimientos en pos de detectar el conocimiento que se desea crear y acotarlo según perspectivas de análisis y nivel de granularidad. Más concretamente en esta fase se da respuesta a las siguientes preguntas:

- ¿Qué conocimiento se quiere crear?
- ¿Qué perspectivas o dimensiones describen el conocimiento?
- ¿Qué niveles de resolución son deseables para medirlo?
- ¿Cómo se relaciona el conocimiento deseado con las perspectivas o dimensiones?
- ¿Qué indicadores son útiles para medir los hechos que aportan conocimiento?

3.2 Elección de las fuentes de datos

Del análisis de requerimientos anterior se descomponen las necesidades de conocimiento (o hechos medibles) en dimensiones, métricas y niveles de resolución los cuales deben ser extraídos de fuentes de datos distribuidas que contienen datos heterogéneos de forma atómica.

En esta fase se analizan los sistemas o recursos que contienen los datos que mejor describen el conocimiento requerido. Para ello es necesario conocer bien los sistemas de almacenamiento de datos de la institución y otras posibles fuentes de datos que pueden estar desnormalizadas y/o desestructuradas.

Las preguntas que pretende responder esta fase son las que se describen a continuación:

- ¿Dónde están los datos unitarios que se necesitan para generar el conocimiento?
- ¿Cómo se obtienen los datos unitarios que se necesitan para generar el conocimiento?
- ¿Cómo combinar los datos unitarios para generar el conocimiento?
- A partir de los datos unitarios ¿cómo se van a conformar las métricas para generar el conocimiento?
- ¿Cuáles son las correspondencias entre el modelo conceptual y las fuentes de datos?

Concluyendo, mediante esta fase se identifican las fuentes de los datos unitarios, se define cómo se deben extraer, se conforman las métricas según las fuentes de datos detectadas, se establecen correspondencias entre las fuentes de datos y el modelo conceptual y se amplía dicho modelo que será útil para las siguientes fases

3.3 Diseño de la base de datos (o almacén de datos)

Posteriormente se crea una base de datos adicional, conformando un almacén de datos (*Data warehouse*) o varios almacenes de datos departamentales (*Data marts*) que almacenarán estos datos seleccionados en un formato multidimensional, de forma que los análisis se hagan sobre sistemas de datos distintos a los de origen.

Lo anterior es crítico respecto a la eficiencia del proceso ya que se evita la sobrecarga de estos últimos. Asimismo al seleccionar los datos en esta base de datos no se almacenan todos los datos de las fuentes de datos sino que sólo se almacenan aquellos que son necesarios y con el grado de resolución (o granularidad) que se ajuste a las necesidades detectadas en la primera fase.

En esta fase se pretende dar respuesta a las siguientes consultas:

- ¿Dónde se almacenan los datos seleccionados?
- ¿Cómo se almacenan los datos para generar el conocimiento deseado? Definir cómo y dónde se almacenarán los datos útiles.

Como resultado se dispone de un sistema informático donde se almacenarán los datos necesarios para los análisis posteriores.

3.4 Integración de datos

Antes de actualizar (o poblar por primera vez) el almacén se requiere realizar un proceso de limpieza de los datos que se extraen desde múltiples fuentes heterogéneas. Realizar correctamente este proceso asegurará que los datos tengan calidad e incidirá sobre la eficiencia y eficacia global del método, por ello resulta crítico dedicar muchos esfuerzos en esta fase.

En concreto en esta fase se realizan las siguientes tareas referentes a los datos unitarios:

- Extracción desde sus fuentes
- Limpieza:
 - depurar,
 - corregir,
 - estandarizar,
 - relacionar y
 - consolidar.
- Transformación:
 - cambio de formatos,
 - sustitución de códigos,
 - cálculo de valores agregados y derivados
- Integración en el almacén de datos para conformar los datos útiles
- Actualizaciones periódicas de los datos útiles.

Una vez cargados los datos se realizan tareas de validación de la calidad de los mismos, observando las fuentes y contrastándolas con los datos incluidos en el almacén. La tabla inferior se ha adaptado de Stair & Reynolds (2012) describe las características que debe tener la información y analiza cómo se puede cumplir cada una de ellas mediante el empleo de este método:

| Características de la información de calidad | Descripción | Materiales y Métodos |
|--|--|---|
| Accesible | Debe ser accesible por personal autorizado | Herramientas OLAP de <i>reporting</i> con control de acceso según roles de usuario. |
| Exacta | La información almacenada debe ser correcta | Procesos ETL Revisión posterior |
| Completa | Debe estar toda la información necesaria para medir los hechos. | Procesos ETL Revisión posterior |
| Económica | Debe ser económica de producir. | Aplicación del método. Inicialmente hay inversiones que son amortizadas durante el tiempo. |
| Flexible | Debe poder servir para múltiples propósitos | Las dimensiones compartidas cumplen esta característica al describir distintos hechos. Un ejemplo lo constituye la dimensión tiempo. |
| Relevante | Se debe almacenar aquella información que tenga una utilidad. | En el análisis de necesidades del método se realiza el filtro para encontrar información relevante. |
| Confiable | La información debe ser creída por los usuarios. | Si se hace partícipe a los usuarios desde un inicio y comprenden cómo se han calculado los indicadores es más probable que sea confiable. |
| Segura | Debe ser consumida sólo por usuarios autorizados | Herramientas OLAP de <i>reporting</i> con control de acceso según roles de usuario. |
| Simple | Cuanto más simple pueda ser más fácil de interpretar resulta y por lo tanto más útil es. | Procesos ETL y herramientas OLAP agregan la información simplificando muchos datos en uno. |
| Puntual | Debe ser entregada en el momento oportuno. | Herramientas OLAP para difundir reportes usando procesos automatizados y medios oportunos. |
| Verificable | Debe poder ser auditable | Al estar almacenada en bases de datos es verificable. |

Tabla 17 Calidad en la información

Con esta fase se consigue disponer de un almacén con datos necesarios (y solo los necesarios), con alta calidad y estructurados según un modelo lógico multidimensional para poder desarrollar análisis sobre los mismos de forma eficiente y eficaz.

3.5 Uso de la solución

Una vez se disponga del almacén de datos actualizado, y se haya asegurado la calidad de los datos que se incluyen en éste, se pueden realizar análisis sobre los mismos mediante herramientas especializadas para este fin.

Estos análisis se alimentan de los datos incluidos en el almacén y se ejecutan según unas normas inferidas a partir del modelo lógico diseñado en etapas anteriores. Estas normas son formateadas en esquemas o cubos de información de forma que son interpretadas por las herramientas OLAP encargadas de realizar los análisis en línea.

El usuario, previamente formado en el uso de las herramientas, tiene a su disposición un conjunto de funcionalidades encaminadas a crear conocimiento. Éstas de forma sintética se presentan como reportes del estado actual, pasado o futuro de uno o varios hechos combinados y organizados según jerarquías y dimensiones que aportan utilidad y comprensión a los mismos.

En este sentido es destacable remarcar el motivo del doble sentido de la flecha que une al usuario con las herramientas de análisis OLAP incluida en la

Figura 22. En concreto el doble sentido hace referencia a que el uso puede ser a demanda, donde el usuario debe realizar una acción para obtener el conocimiento deseado o, por el contrario, puede ser automático recibiendo resúmenes temporalizados mediante los canales de comunicación que se empleen en la organización.

Cuando el usuario recibe la información y la consume, si ésta es lo suficientemente útil y contextualizada según las necesidades del usuario, se genera un conocimiento válido para poder tomar decisiones respecto al hecho que se está midiendo. Esto a su vez genera nuevas preguntas que desean ser respondidas alimentando de nuevo el método y ampliando la información que se almacena en el almacén de datos.

4 Detalle de los procesos de cada fase

En este apartado se explica de forma detallada, para cada una de las fases descritas en el apartado anterior, los procesos y los resultados obtenidos, así como los actores que participan.

4.1 Análisis de requerimientos

Esta fase pretende responder a las siguientes preguntas:

- ¿Qué conocimiento se quiere generar?
- ¿Qué resolución se necesita para medir el conocimiento requerido?
- ¿Cómo se debería estructurar el sistema de datos?

Esta información es necesaria para diseñar el modelo lógico multidimensional de datos que se usa para generar las consultas por parte de las

herramientas OLAP de análisis. Además también resulta necesaria para escoger las fuentes de datos que poblarán el *Data warehouse*.

De forma más genérica, mediante este análisis de necesidades se pretende identificar las necesidades de información clave esenciales para la toma de decisiones de forma eficaz y eficiente.

4.1.1 Consideraciones según la estrategia escogida para construir el DW

Independientemente de la estrategia que se quiera emplear para crear el *Data warehouse* (conjunto de *data marts* o *data marts* derivados) esta información tarde o temprano será necesaria para implementar el método propuesto en esta tesis. En este sentido según las distintas estrategias consideradas para construir el *Data warehouse* se tienen distintas visiones sobre cómo afrontar esta fase. La tabla inferior resume este distinto enfoque según la estrategia escogida:

| Estrategia | Implementación del análisis de requerimientos |
|--|--|
| <i>Bottom-up</i> (Kimball, 1996) | Se hace necesario obtener esta información de forma muy detallada para escoger las mejores fuentes de datos y para construir los distintos modelos lógicos multidimensionales que conformarán los distintos <i>Data marts</i> . |
| <i>Top-down</i> (Inmon W. H., 1996) | Se podría obviar esta fase inicialmente siempre que se pudiera estimar qué información será útil en un futuro pero sin entrar a mucho nivel de detalle. Esta información quedaría almacenada en una base de datos relacional que luego, cuando se tuviera una necesidad concreta y tras un análisis de requerimientos, serviría para generar los distintos modelos lógicos multidimensionales. |

Tabla 18 Implementación del análisis de requerimientos según las estrategias para crear un DW.

Analizando la tabla superior, es lógico concluir que hay riesgos de ineficacia del método cuando se emplea la estrategia *Top-down*. Esto está debido a que inicialmente se recopilan datos según una estimación inicial de información potencialmente útil en un futuro. En este sentido si la estimación inicial no ha sido certera, existe un gran riesgo de no poder generar análisis específicos, debido por un lado a la inexistencia de estos datos en nuestro DW y por otro a niveles de granularidad altos. Por este motivo en caso de querer implementar esta estrategia, se sugiere que se capture casi toda la información disponible en las fuentes de datos y en el nivel de granularidad menor posible para luego disponer de toda la información necesaria en el análisis de los mismos. Esto implica que prácticamente se tiene una copia exacta de las fuentes de datos dentro del DW, implicando un aumento del coste en equipos informáticos y mantenimiento de los sistemas.

Debido a que tarde o temprano se requerirá realizar un análisis de requerimientos detallado, resulta lógico considerar que este análisis sea el primer paso para alimentar el DW de la organización y por lo tanto se sugiere emplear el enfoque de Kimball (1996) aplicando la estrategia *Bottom-up*, así se reducen los riesgos de tener que volver a capturar información desde las fuentes de datos que no se había considerado inicialmente, entre otras ventajas.

4.1.2 Actores y funciones

Aunque en el apartado 6 del presente capítulo se describen los actores que participan en este proyecto, se describen con más profundidad debido a la importancia que tienen en esta etapa. En este sentido los actores de la fase de análisis de requerimientos y sus funciones se propone que como mínimo sean los que quedan descritos en la tabla inferior:

| Actor | Funciones |
|-------------------------|--|
| Gestor del conocimiento | Coordinar las reuniones Asegurar que se recopila la información necesaria para las siguientes fases (hechos, métricas y dimensiones) Sugerir indicadores adicionales Orientar a los tomadores de decisiones sobre las dimensiones y métricas que ayudan a medir un hecho. Orientar a los tomadores de decisiones en plantear preguntas genéricas y transversales para todos los ámbitos académicos de la institución Puente entre el discurso de los tomadores de decisiones y el personal informático Proponer el modelo conceptual |
| Tomadores de decisiones | Definir el nuevo conocimiento deseable de capturar Definir el nivel de granularidad Definir las dimensiones y métricas ideales para la medición |
| Informáticos | Definir el modelo conceptual Sugerir modificaciones de indicadores para optimizar esfuerzos. Validar la factibilidad de extracción de indicadores según el análisis de las preguntas |

Tabla 19 Actores y funciones del análisis de requerimientos

Se considera importante que exista un miembro del equipo informático en esta fase para que entienda el motivo de las tareas que va a tener que desarrollar,

entre otros motivos debido a que puede aportar una visión encaminada a optimizar los esfuerzos futuros, por ejemplo sugiriendo cambios en la forma de calcular los indicadores justificados según criterios de eficiencia o eficacia tecnológica del sistema final.

Estas mismas funciones puede desarrollarlas el gerente del conocimiento quien está a medio camino entre los usuarios finales y los sistemas informáticos en los que se apoya este método, no obstante sus funciones son más de asegurar que se extraiga la información correcta para poder avanzar hacia las siguientes fases. Además, el equipo informático tendrá mucho más conocimiento a nivel de detalle respecto a las fuentes de datos sobre las que se apoya el método.

4.1.3 Acciones para analizar los requerimientos

El proceso que se emplea en esta fase se resume en las siguientes acciones:

1. Realizar preguntas a los tomadores de decisiones que expliciten los objetivos de su organización o las necesidades concretas de información.
2. Analizar estas preguntas para identificar hechos, métricas, dimensiones y nivel de resolución de los hechos.
3. Identificar posibles fuentes de datos y validar su viabilidad
4. Realizar el modelo conceptual a partir de la información anterior

En los siguientes apartados se detalla cada uno de estos pasos

4.1.4 Identificar la información necesaria para la toma de decisiones

Inicialmente se requiere obtener la información sobre qué nuevo conocimiento debe generarse. Esto puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

Durante este proceso es necesario tener en consideración los siguientes aspectos:

- Hacer partícipe a los tomadores de decisiones quienes serán los usuarios finales de las funcionalidades que derivan de este método.
- Enfocar las necesidades de información en los procesos principales que desarrollan los tomadores de decisiones
- .Se debe asegurar que la información debe estar soportada por alguna fuente de dato existente en la institución. En este sentido se sugiere hacer partícipe a personal del equipo técnico en las reuniones para que valide esta circunstancia.

4.1.5 Identificar los indicadores para medir el conocimiento

Para cada pregunta generada en el paso anterior se plantea hacer un análisis conjunto con los actores que participan en este proceso (ver apartado 4.1.6 de este capítulo)

Para esta tarea se propone el empleo de la siguiente tabla en la que se desglosa cada pregunta en un hecho, varias dimensiones y varias métricas. En este

sentido se sugiere que esta actividad se realice en la misma reunión después de haber identificado todas las preguntas clave.

| Pregunta a responder | Hecho | Métrica | Dimensiones |
|---|---------------|----------------------------|---------------------|
| Se trata de la pregunta a la que se quiere dar respuesta mediante este método | ¿Qué se mide? | ¿Con qué unidades se mide? | ¿Sobre qué se mide? |

Tabla 20 Propuesta de información deseable para definir el modelo conceptual

Nótese que las métricas, para que sean realmente efectivas son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: cantidades, saldos, promedios, sumatorias, fórmulas, etc.

Adicionalmente se requiere entrar al nivel de detalle de la resolución de cada par hecho-dimensión. Por ello se debe definir la resolución o nivel de granularidad siguiendo como modelo la siguiente tabla:

| Hecho-dimensión | Resoluciones |
|-----------------|--------------|
| | |

Tabla 21 Propuesta de desglose de resoluciones deseable para definir el modelo conceptual

Atendiendo a la estrategia de Kimball (1996) respecto a que un DW está compuesto por varios DM que comparten dimensiones, se propone a continuación un conjunto de dimensiones comunes que pueden ser compartidas por muchos hechos que suelen ocurrir en entornos virtuales de aprendizaje. En este sentido la tabla de debajo refleja las dimensiones compartidas y distintos niveles de

resolución para resolver la mayor parte de las consultas y análisis que se realicen sobre un entorno virtual de aprendizaje:

| Dimensión | Resoluciones propuestas |
|-----------------------|--|
| Tiempo | Año Mes Día Hora Minuto Día de la semana Tipo de día (laborable/festivo...) Semestre Trimestre |
| Usuarios | Id de usuario Nombre y apellidos Rol (estudiante, profesor, coordinador...) Grupo Sexo Edad País Ciudad |
| Acciones | Id de acción Nombre de acción Tipo de acción (creativa/consultiva, activa/pasiva...) Subtipo de acción (crear, responder, corregir, borrar, leer, visitar...) |
| Módulos y actividades | Id de módulo Nombre de módulo Tipo de módulo (social/académico, comunicativo/evaluativo/informativo, obligatorio/optativo...) Subtipo de módulo (Examen evaluado, Foro de preguntas frecuentes, foro de preguntas activas...) |

| Dimensión | Resoluciones propuestas |
|-----------|---|
| Cursos | Id de Curso Nombre de curso Abreviatura de curso Área / departamento del curso Tipo de curso (abierto/cerrado, teórico/práctico...) |

Tabla 22 Dimensiones y resoluciones propuestas para el análisis de los datos de un entorno académico

Respecto a los tipos y subtipos presentados en los niveles de resolución para las dimensiones de la tabla superior, éstos deben responder a las necesidades de la organización y por lo tanto deben ser definidos por los miembros de ésta, normalmente por los responsables de la toma de decisiones. Por ejemplo para los tipos de actividades se ha diferenciado las actividades activas de las pasivas, algo que puede ser útil para una organización pero no necesariamente para otra. En este sentido cada organización debe agrupar la información según segmentos y reglas de negocio que le sean de provecho.

No obstante algunas de estas dimensiones son muy transversales en todas las organizaciones educativas y puede resultar útil detallar posibles formas de organizar la información que se obtiene de los entornos virtuales de aprendizaje. Ejemplo de ello sería la dimensión referente a los módulos o actividades, que se usa en todas las organizaciones educativas. Por ello las tablas inferiores resumen algunas de las posibilidades para describir las actividades y realizar análisis teniendo en cuenta los atributos que describen esta dimensión.

| Descriptores | Resoluciones | | |
|--|---|---|--|
| Según el tipo de interactividad | Expositivo Activo Combinado | | |
| Según el tipo de actividad | Ejercicio/ Problema Cerrado Expositivo Juego Didáctico Lectura Guiada Metodología Lección Magistral Comentario de texto/imagen Actividad de discusión Caso Contextualizado Escenario real/virtual de aprendizaje | Problema abierto <i>Webquest</i> Experimento Proyecto real Herramienta Simulación Cuestionario Autoevaluación Examen Módulo formativo | |
| Según el nivel de interactividad | Alto / medio / bajo... u otra escala única para toda la institución formativa. | | |
| Según la dificultad | Alta / media / baja... u otra escala única para toda la institución formativa. | | |
| Según las competencias que desarrolla | Establecer previamente un mapa de competencias transversal y único para todos los departamentos de la institución formativa. | | |
| Según los procesos cognitivos que ocurren durante la elaboración de la actividad | Analizar Aplicar Colaborar Comprender Comparar Compartir Comprobar Comunicar Controlar Contextualizar Cooperar Crear Decidir Definir | Explicar Extrapolar Innovar Juzgar Motivar Observar Organizar Planificar Practicar Producir Discutir Diseñar Emparejar Evaluarse | Reconocer Redactar Reflexionar Relacionar Representar Resolver Simular Sintetizar Valorar Descubrir |

Tabla 23 Propuesta de resoluciones para la dimensión actividad según el uso que se hace del recurso en un entorno académico.

Otra visión distinta es enfocar el análisis de las actividades según criterios de accesibilidad, en este caso se podrían tener en consideración las siguientes resoluciones:

| Descriptores | Resoluciones |
|--|--|
| Modo de interacción para el acceso | Teclado Ratón Dispositivos apuntadores Reconocimiento de voz |
| Modo de presentación de la información dentro del objeto | Auditivo Visual Textual Táctil |
| Control sobre el tiempo de ejecución | Permite adaptar al visualización Tamaño de caracteres Tipo de fuente Color de fuente Color de fondo Formato de cursor Formato de selección Tratamiento y contraste de la imagen |

Tabla 24 Propuesta de resoluciones para la dimensión actividad según el nivel de accesibilidad que se tiene sobre el recurso académico

Los ejemplos de las tablas superiores implican un trabajo cooperativo entre todos los actores de la institución, incluyendo las instancias directivas en pos de definir un marco estable, único y transversal sobre el que clasificar los datos y poder construir un conocimiento válido en toda la institución, independientemente

del departamento que lo utilice. Evidentemente los ejemplos de arriba responden a una dimensión que sólo será compartida por los departamentos académicos y no necesariamente deben ser aplicados, sino que se usarán aquellos niveles de dimensiones que más se ajusten a las necesidades de información que se requieran.

4.1.6 Identificar posibles fuentes de datos y validar su viabilidad

Respecto a identificar las fuentes de datos, a pesar que es un proceso que ocurre más adelante se sugiere realizar una primera aproximación durante la fase de análisis de requerimientos, incluso durante la misma reunión en la que se están identificando los indicadores que se pretenden medir. En este sentido se sugiere que al menos un participante en esta fase provenga del departamento tecnológico y que dicho participante conozca bien las posibles fuentes de datos para afrontar este método.

Esto se entiende al considerar que este equipo técnico, que conoce las fuentes de datos y las operaciones de transformación de los mismos, puede inferir en la misma reunión la factibilidad o complejidad de la obtención de los mismos, y sugerir el rediseño de los indicadores a partir de otras fuentes de datos o de otros métodos de cálculo que también pudieran ser válidos.

Por este motivo, a medida que se van trabajando con la identificación de hechos, métricas, dimensiones y niveles de granularidad este integrante debería ir validando la factibilidad de poder extraer los datos y su dificultad, por si se

podieran identificar otros indicadores en sustitución de los iniciales, cuyo cálculo u obtención fueran más sencillas de implementar.

4.1.7 Realizar el modelo conceptual

En esta etapa se genera una aproximación al diseño de la solución final. El diseño conceptual, que podrá ser entendido fácilmente por personal no técnico, tiene el objetivo de mostrar el diseño del almacén en vista a las dimensiones y las métricas que se emplean para medir el hecho.

Un modelo conceptual no es otra cosa que una representación gráfica de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos. Es necesario en este sentido que se genere un modelo conceptual distinto para cada hecho que se quiera medir, pues cada uno tendrá sus dimensiones y métricas asociadas. Asimismo cada dimensión se puede presentar con la estructura que refleje el nivel de resolución o granularidad deseado.

La ventaja de representar la estructura de la base de datos de esta forma radica en que se puede observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además permite que pueda ser presentado ante los tomadores de decisiones de forma que se entienda y se pueda explicar con facilidad.

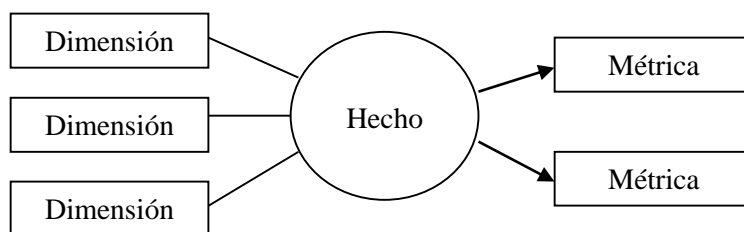


Figura 23 Ejemplo esquemático de un modelo conceptual

Es el gerente del conocimiento junto con el personal informático quien diseña los modelos conceptuales, no obstante los tomadores de decisiones también pueden aprender a realizarlos y presentar la información en este formato ante nuevas necesidades de generación de nuevo conocimiento.

4.2 Análisis de las fuentes de datos

Una vez diseñado el modelo conceptual se analizará en profundidad si se tienen las fuentes de datos suficientes para alimentarlo y en caso contrario se tratará de generarlas a partir de la combinación de otras fuentes.

Asimismo en esta fase se determina cómo serán calculados los indicadores y las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definen los campos que se incluirán en cada dimensión. Finalmente, se actualiza el modelo conceptual ampliándolo con la información obtenida en este paso.

Esta tarea amplía la sugerencia de realizar un primer análisis de las fuentes de datos, por parte del personal técnico informático, y su viabilidad durante las

reuniones iniciales donde se extraen los requerimientos de información para la toma de decisiones.

De forma resumida esta fase responde a las siguientes preguntas:

- ¿Dónde están los datos unitarios que se necesitan para generar el conocimiento?
- ¿Cómo se obtienen los datos unitarios que se necesitan para generar el conocimiento?
- ¿Cómo combinar los datos unitarios para generar el conocimiento?
- A partir de los datos unitarios ¿cómo se van a conformar las métricas para generar el conocimiento?

Los objetivos que persigue esta fase deberían ser los siguientes:

- Conformar las métricas, y

4.3 Establecer las correspondencias entre fuentes de datos y el modelo conceptual

Finalmente se obtiene información suficiente para conformar el modelo conceptual ampliado, con toda la información necesaria para el siguiente paso.

Debajo se presenta un ejemplo de dicho modelo conceptual ampliado. Nótese que se han añadido los campos que describen cada dimensión y las fórmulas de cálculo de cada métrica.

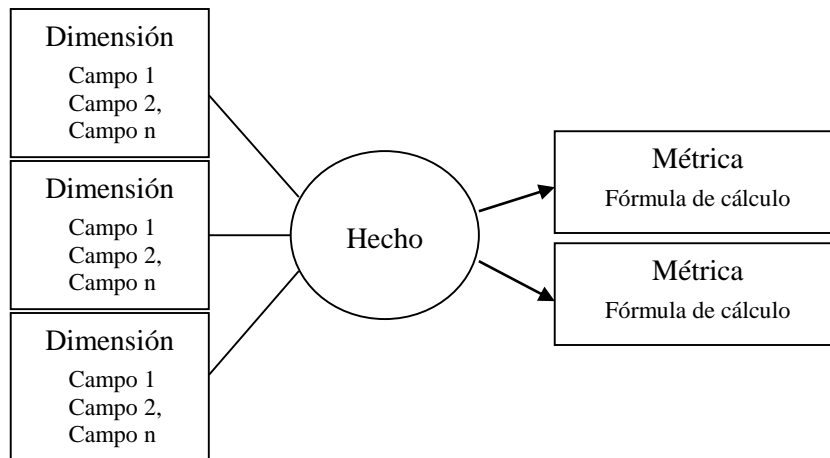


Figura 24 Ejemplo esquemático de un modelo conceptual ampliado

4.4 Elaborar el modelo lógico del almacén de datos

Una vez se tiene el modelo conceptual se transforma en un modelo lógico que responda a la siguiente pregunta: ¿Cómo y dónde se almacenarán los datos para generar el conocimiento deseado?

Según lo anterior se define cuál será el tipo de esquema que se implementará para después construir las tablas de dimensiones y las tablas hechos, y así determinar sus respectivas uniones, a partir del modelo conceptual creado.

En este sentido resulta muy importante definir si se empleará un esquema en estrella, constelación o copo de nieve, ya que esta decisión afectará la elaboración del modelo lógico.

Asimismo este es el momento para definir los estándares de codificación, que entre otros objetivos tiene el de establecer los nombres de la manera más clara posible y que reflejen su significado en el negocio.

Una vez definido el esquema y los estándares que se seguirán para codificar la información, se puede empezar a construir las tablas en el DW o DM que acogerán los datos seleccionados en la fase anterior. Respecto a las tablas que deben construirse son las siguientes:

- Tablas de dimensiones
- Tablas de hechos

Lo primero que se propone es la creación de las tablas de dimensiones. Para ello se tomará cada dimensión del modelo conceptual, con sus atributos relacionados, y se realizarán los siguientes trabajos:

1. Elección de un nombre que identifique la dimensión.
2. Añadir un campo que represente su clave principal.
3. Redefinir los nombres de los atributos en el caso de no ser lo suficientemente explicativos.

Gráficamente se podría representar, siguiendo el modelo conceptual presentado en la Figura 24, de la siguiente forma:

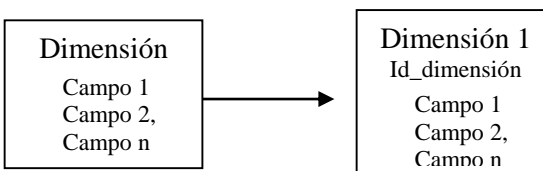


Figura 25 Representación gráfica de una tabla de dimensión

Una vez construidas las tablas de dimensiones se crean las tablas de hechos, las cuales se configuran de distinta manera según el esquema implementado.

La tabla inferior resume las recomendaciones de configuración según esta realidad:

| Tipo de esquema | Enfoque de configuración |
|--------------------------|---|
| Estrella y copo de nieve | Asignar un nombre que sea identificativo del hecho. La clave primaria resulta de la combinación de las claves primarias de cada dimensión asociada a ese hecho. Renombrar las métricas si no fueran suficientemente explícitas. |
| Constelación | Las mismas recomendaciones para los esquemas de estrella y copo de nieve. Cuando se comparten métricas pero distintas dimensiones, existirán tantas tablas de hechos como métricas compartidas haya. Cuando no se comparten métricas ni indicadores habrá tantas tablas de hecho como métricas se quieran medir. Cuando comparten métricas y dimensiones se asemeja al esquema de estrella y se procede de igual manera. |

Tabla 25 Distintos enfoques de modelo lógico según el esquema de datos empleado

Una vez se tengan creadas las tablas en el almacén de datos se puede empezar a cargar los datos previamente seleccionados, de esto trata el siguiente apartado.

4.5 Integración de los datos

Una vez construido el modelo de datos del almacén, se procede a poblarlo. Se definirán políticas y estrategias para la carga inicial y su respectiva

actualización utilizando técnicas de limpieza y calidad de datos incluidas en los procesos ETL. Debido a que existe mucha literatura respecto a estos procesos en este apartado se resumen los procesos previamente identificados y detallados en el apartado 3.6 del capítulo referente al marco teórico.

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. En la mayoría de los casos los datos extraídos requerirán alguna manipulación y transformación. Algunas de estas operaciones pueden ser: sustituir algunos términos por otros (por ejemplo el sexo M por 1, F por 0), obtener nuevas columnas (por ejemplo desglosar una fecha en día, mes y año) o unir datos de múltiples fuentes o columnas de una tabla.

La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad

de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. En este sentido los almacenes de datos mantienen un historial de los registros de manera que se pueda hacer una auditoría de calidad de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

La tabla inferior adaptada de Cano (2007) resume con más nivel de detalle las características de cada fase del proceso de integración de datos.

| Subproceso | Descripción | Objetivo | Problemas | Resultado |
|------------|---|--|--|---|
| Extracción | Recuperación de los datos de las distintas fuentes de información. | Extraer tan sólo aquellos datos de los sistemas transaccionales que son necesarios y prepararlos para el resto de los subprocesos de ETL | Distintos/as: Fuentes, BBDD Plataformas tecnológicas, Protocolos de comunicaciones, Juegos de caracteres, Tipos de datos | Datos en bruto que no han sido depurados y que deben ser limpiados. |
| Limpieza | Depurar, corregir, estandarizar, relacionar y consolidar los datos extraídos de las fuentes de datos. | Conseguir datos sin duplicados, homogéneos, filtrados y corregidos. | Distintos/as formatos en fechas, abreviaturas... Errores en los datos detectables mediante algoritmos y fuentes de datos externas. Datos demasiado agregados que se pueden desagregar. Datos duplicados | Datos depurados desagregados |

| Subproceso | Descripción | Objetivo | Problemas | Resultado |
|----------------|--|---|--|---|
| Transformación | Cambios de formato, sustitución de códigos, valores derivados y agregados | Transformar los datos según las reglas de negocio y los estándares prefijados. | Los datos están desagregados y teniendo en cuenta el nivel de granularidad que se le quiera dar se realizan agregaciones. | Datos depurados, recalculados y agregados. |
| Integración | Proceso en el que los datos depurados, recalculados y agregados se carga en el <i>data warehouse</i> | Cargar los datos en el <i>data warehouse</i> y asegurar de que no haya errores en este proceso. | El principal problema en esta fase es que falle el proceso de carga y los datos cargados no coincidan con los datos transformados del subproceso anterior. | Datos cargados en el <i>data warehouse</i> sin errores. |
| Actualización | Proceso de actualizar los datos del <i>data warehouse</i> | Realizar actualizaciones según una periodicidad definida | Carga del sistema: hay que definir el mejor momento para no saturar los sistemas informáticos. Todos aquellos derivados de los subprocesos anteriores. | Datos actualizados periódicamente |

Tabla 26 Resumen de los subprocesos ETL. Adaptado de Cano, 2007

4.6 Uso de la solución

En este apartado se engloban las tareas relativas a la configuración de las herramientas de análisis y las distintas funcionalidades que se explotan en esta tesis.

Inicialmente se configuran los cubos multidimensionales, según el modelo lógico y conceptual obtenidos en fases anteriores, para posteriormente diseñar los análisis, reportes y cuadros de mando que van a ser utilizados por los usuarios finales del sistema con el fin de obtener el conocimiento necesario y poder tomar las decisiones oportunas según los resultados mostrados en éstos.

Previamente al uso de los análisis, reportes y cuadros de mando se requiere realizar una formación a los usuarios en el manejo de las herramientas para maximizar su uso y minimizar el tiempo para obtener los resultados de los informes.

Debido al carácter aplicado de las tareas de este apartado se ha considerado mejor desarrollarlas en el capítulo siguiente relativo a la aplicación de la metodología propuesta en este capítulo. Esto es así porque su diseño está directamente relacionado con los requisitos de conocimiento detectados en la primera fase.

5 Enfoque del método desde el punto de vista de la gestión del conocimiento

La figura inferior resume esquemáticamente según cada fase, el enfoque del proyecto desde el punto de vista de la gestión del conocimiento.

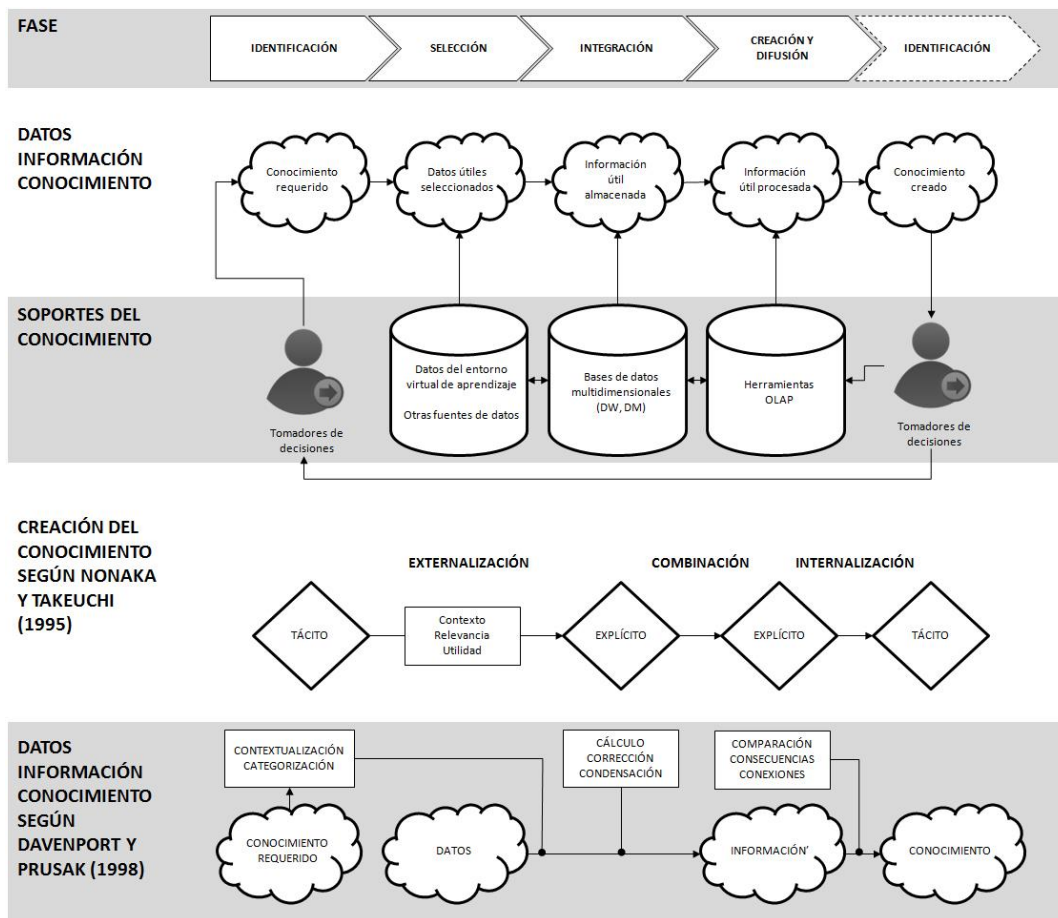


Figura 22 Fases del método según la perspectiva de la gestión del conocimiento

De forma genérica, el conocimiento se crea a partir de datos, de formatos heterogéneos y distribuidos en muchos sistemas, a los cuales se les aportan capas de contexto, relevancia y utilidad que los convierten en información. Ésta al ser consumida según criterios de comparación, consecuencias o conexiones crean un conocimiento útil para el tomador de decisiones.

En la fase inicial de identificación del conocimiento, que se realiza durante el análisis de necesidades, se externaliza el conocimiento tácito ubicado en los tomadores de decisiones, agregando contexto y categorización. Este proceso se

sintetiza en el modelo conceptual que se construye en esta fase y que sirve de base para la construcción del almacén de datos con estructura multidimensional.

Posteriormente se escogen las fuentes de datos que son representativas según criterios de relevancia y utilidad de los mismos y se transforman (cálculos, correcciones y condensaciones) e integran en una base de datos cuya estructura es multidimensional de forma que los datos iniciales se agrupan alrededor de un hecho en forma de dimensiones y métricas según los criterios mencionados anteriormente, representados en el modelo lógico.

Finalmente el usuario, mediante el empleo de herramientas de análisis OLAP, que permiten realizar comparaciones, conexiones y validar consecuencias, internaliza el conocimiento explícito contenido en la información en tácito.

Desde el punto de vista de la estructura de los datos representada en el modelo lógico, que sirve de base para realizar análisis OLAP tiene una correlación directa con lo expresado por Davenport y Prusak (1998) en su afán por describir el proceso de conversión de los datos en información y ésta en conocimiento. La figura siguiente refleja esta relación, representada de forma esquemática.

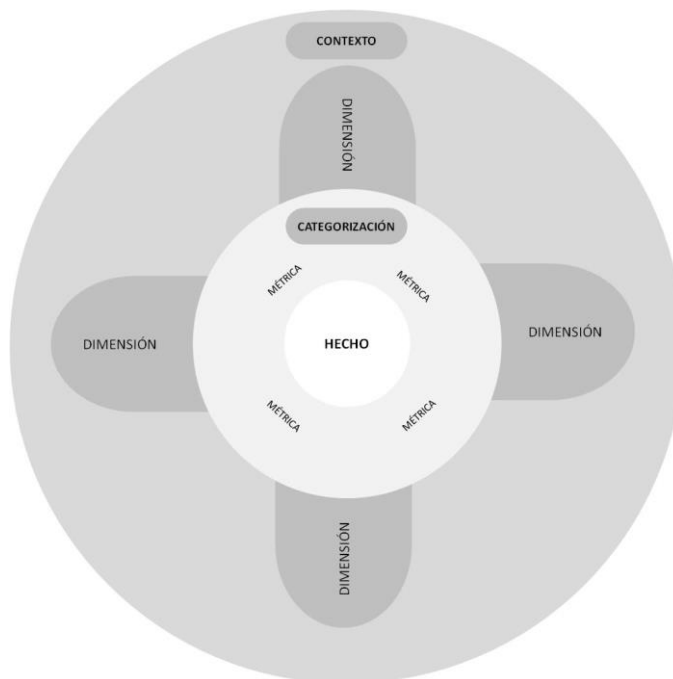


Figura 26 Relación entre el modelo lógico y las características de contexto y categorización que definen a la información.

En esta figura el hecho central correspondería a un dato único (calculado, corregido y condensado) el cual es categorizado por medio de las métricas y contextualizado según las dimensiones que le rodean convirtiéndolo en información.

Las métricas están intrínsecamente ligadas al hecho que se pretende medir, es decir, no se puede medir un hecho sin una métrica y viceversa, en este método tampoco puede existir un hecho que no se mida. Por ejemplo si se pretende conocer los accesos (hecho) que existe en un entorno virtual de aprendizaje se necesita aportar una unidad de medida y ésta está intrínsecamente vinculada al hecho. En este caso el acceso se podría medir por frecuencia: número de veces que

se accede; tiempo de cada acceso: cuánto tiempo ha estado en cada acceso; índice de acceso en relación a otros participantes o grupos, etc....

Por el contrario la vinculación de las dimensiones con el hecho es más laxa de forma que una dimensión puede contextualizar varios hechos, un ejemplo de ello sería la dimensión tiempo, que aplicaría a un gran conjunto de hechos pero es posible que no a todos.

La figura inferior representa esquemáticamente cómo los datos que se organizan dentro de un repositorio de datos multidimensional que conforma el DW han adquirido unas características tales que ya no son datos sin sentido sino que se han convertido en información contextualizada y categorizada.

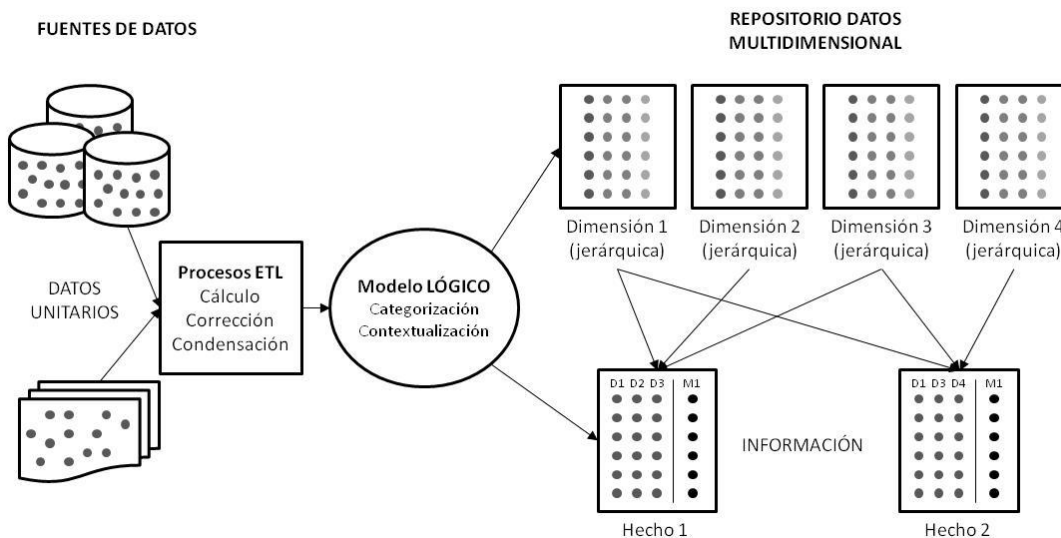


Figura 27 Conversión de datos en información según el método propuesto

En la parte izquierda de la figura se ubican las fuentes de datos que son múltiples y contienen datos en formatos heterogéneos. Estos datos son

seleccionados y transformados mediante procesos informáticos y posteriormente almacenados siguiendo un modelo lógico definido a partir del conocimiento tácito externalizado durante el análisis de necesidades que inicia todo el proceso.

En la parte central se expresan los procesos que transforman y dan sentido a los datos los cuales se almacenan físicamente siguiendo una estructura dada por la utilidad que se pretende conseguir. En este sentido se presentan dos hechos que quedan descritos mediante varias dimensiones (D) y una o varias métricas (M).

Desde el punto de vista expresado en la figura superior se podría considerar que las dimensiones y los hechos conforman la información que dispone la organización para poder generar el conocimiento la cual es almacenada en bases de datos cuya estructura es multidimensional. Siguiendo esta línea de pensamiento, el conocimiento se consigue explotando esta información mediante técnicas tales como comparar o conectar distintas dimensiones de un hecho o distintos hechos entre sí. Esto se realiza gracias a la estructura multidimensional de las tablas que componen las bases de datos y a herramientas OLAP que son capaces de explotarla.

6 Beneficios del método para los roles más significativos de un entorno virtual de aprendizaje

En esta sección se identifican los roles sobre los que puede aplicar una solución de BI aplicada a entornos virtuales de aprendizaje. La tabla de debajo resume para cada rol, sus funciones y tareas principales y cómo una solución de BI para la toma de decisiones le puede ser de utilidad.

| Rol | Funciones | Tareas | Beneficios de la solución según el conocimiento que se adquiere |
|------------------|--|--|---|
| Profesor | Mantener asignaturas actualizadas y mejoradas Atender aportes de alumnos en el menor tiempo posible | Corregir exámenes Responder consultas Actualizar recursos Mejorar asignaturas | Conocer el estado de las asignaturas según la actividad que se registra en ellas Conocer el uso que se da a los recursos de la asignatura Estar informado de los nuevos aportes de alumnos Prever amenazas (i.e. exámenes mal redactados) |
| Tutor | Detectar y corregir alumnos en estado incorrecto Responder solicitudes de alumnos | Seguimiento de alumnos | Conocer el estado actual y pasado de los alumnos en un curso según su actividad Prever tendencias Disponer de información ampliada sobre la actividad de los alumnos |
| Coordinador | Mantener los cursos actualizados y mejorados Asegurar la calidad del trabajo de los profesores | Seguimiento de profesores Seguimiento de cursos | Conocer el estado actual y pasado de los cursos Disponer de información sobre el estado de los profesores según su actividad y carga de trabajo Prever tendencias en el estado de cursos y de profesores Descubrir cargas de trabajo según previsiones de altas y bajas. |
| Administrador CV | Asegurar la calidad del trabajo del personal técnico del campus Conocer el estado de los servicios informáticos | Seguimiento de equipo técnico Seguimiento de servicios informáticos | Conocer el estado del personal técnico del campus según su actividad Conocer el estado de los servicios informáticos (errores, caídas...) |

| Rol | Funciones | Tareas | Beneficios de la solución según el conocimiento que se adquiere |
|-----------------------|---|--|---|
| Equipo técnico del CV | Mantener actualizado y operativo el campus Atender las incidencias de los usuarios | Soporte técnico Creación y actualización de cursos y recursos | Disponer información técnica de los equipos de los usuarios Detectar errores según curso o usuario |

Tabla 27 Actores de un entorno virtual de aprendizaje y beneficios del método de BI.

7 Enfoque desde el punto de vista de la eficiencia y la eficacia

De forma genérica los factores que se estima que inciden en la eficiencia y eficacia globales del proceso se recogen en la tabla inferior.

| | Factores |
|------------|---|
| Eficiencia | Autonomía de los tomadores de decisiones Reducción del tiempo para la generación de nuevos análisis (evitar introducción manual, evitar repeticiones, informes detallados) Mejora en el rendimiento de los sistemas de producción Centralización de la información en un único sistema informático accesible desde cualquier ordenador |
| Eficacia | Almacenamiento de información en términos de utilidad o negocio. Aumento en la calidad de los datos Reducción de errores relativos al empleo de distintos sistemas |

Tabla 28 Factores que inciden positivamente en la eficiencia y eficacia del método

Respecto a la medición de la eficiencia y eficacia globales del método se proponen los indicadores reflejados en la tabla inferior.

| | Indicador | Criterios de comparación |
|------------|---|---|
| Eficiencia | Conocimiento creado en términos de consumo de información (reportes consumidos y creados) | Tiempo invertido por el tomador de decisiones Rendimiento de los sistemas informáticos |
| Eficacia | Cantidad de información útil y con calidad disponible para su consumo. Acciones tomadas según el conocimiento creado | |

Tabla 29 Indicadores globales de eficiencia y eficacia propuestos.

De forma más específica en cada una de las fases se sugiere tener en consideración los factores expresados en la tabla inferior, relativos a los indicadores de eficiencia y eficacia, los cuales se han introducido en la Tabla 16.

| Fase | Factores de eficiencia | Factores de eficacia |
|----------------------------------|---|---|
| Análisis de requerimientos | <ul style="list-style-type: none"> • Preparación de las reuniones para minimizar el tiempo dedicado. • Selección de los actores que mejor conozcan las necesidades para reducir el tiempo en las reuniones. | <ul style="list-style-type: none"> • Selección de los actores que conozcan las necesidades de conocimiento crítico. • Selección de las preguntas correctas según criterios de utilidad y/o transversalidad. |
| Elección de las fuentes de datos | <ul style="list-style-type: none"> • Identificación de las fuentes de datos que maximicen el rendimiento de los sistemas de origen y minimicen el tiempo de extracción, limpieza y transformación de los datos de origen. • Selección del personal informático más preparado para minimizar el tiempo de análisis de las fuentes de datos • Identificación de indicadores de fácil obtención y cálculo. • Selección del nivel de granularidad mayor posible según las necesidades actuales y previstas. | <ul style="list-style-type: none"> • Elección del personal informático correcto para no obviar ninguna fuente de datos necesaria. • Elección de las fuentes de datos correctas • Selección de indicadores correctos (alta utilidad y/o transversalidad). • Selección de la granularidad necesaria |

| Fase | Factores de eficiencia | Factores de eficacia |
|-----------------------------|---|--|
| Diseño del almacén de datos | <ul style="list-style-type: none"> • Elección del mejor esquema según las necesidades para maximizar el rendimiento de los servidores y reducir los tiempos de respuesta. • Elección del personal informático formado en las tareas de esta fase para minimizar los tiempos de ejecución de las mismas | <ul style="list-style-type: none"> • Elección de la codificación correcta • Elección del personal informático adecuado para evitar errores en el diseño de la base de datos |
| Integración de datos | <ul style="list-style-type: none"> • Elección del horario de actualización de la base de datos que maximice el rendimiento de los sistemas de origen. • Elección la máxima periodicidad de actualización de los datos según las necesidades temporales del conocimiento que se requiere para maximizar el rendimiento de los sistemas de origen. • Diseño de los procesos de limpieza y carga que menor tiempo y mayor rendimiento de procesos ETL requiera. • Elección del personal informático más preparado en las tareas de esta fase para minimizar el tiempo de diseño de los mismos. | <ul style="list-style-type: none"> • Elección del personal informático adecuado para evitar errores en los procesos ETL. • Elección de los tiempos de actualización necesarios según las necesidades temporales del conocimiento requerido. • Validación de la calidad de los datos útiles. |
| Uso de la solución | <ul style="list-style-type: none"> • Simplificación de los resultados para minimizar el tiempo de interpretación de la información. • Formación de los usuarios antes del uso de las herramientas OLAP para que minimicen el tiempo de uso de las mismas | <ul style="list-style-type: none"> • Formación suficiente de los usuarios finales en las herramientas OLAP. • Diseño de sistemas automatizados para hacer llegar la información a los tomadores de decisiones en el momento requerido. • Elección de plantillas visuales para representar la información de forma que responda a las necesidades de conocimiento. |

Tabla 30 Factores de eficiencia y eficacia propuestos según las fases del método

8 Conclusiones

- Se ha presentado una metodología BI adaptada con el objetivo de extraer los datos de un entorno virtual de aprendizaje y convertirlos en conocimiento de forma eficiente y eficaz.
- Se han descrito las fases del proyecto detallando los aspectos críticos a tener en consideración en cada una de ellas.
- Se han presentado distintos factores que influyen en la eficiencia y eficacia del método.
- Se ha realizado un análisis del método desde el punto de vista de la gestión del conocimiento según Nonaka y Takeuchi y Davenport y Prusak.
- Se han analizado los beneficios potenciales que este método aporta a los roles más significativos de un entorno virtual de aprendizaje.
- Se han caracterizado dimensiones y resoluciones de aplicación para las organizaciones formativas que emplean entornos virtuales de enseñanza.

Capítulo 4 - Aplicación del método

1 Introducción

En este apartado se ha puesto en aplicación el método presentado en el capítulo anterior en un curso semipresencial impartido por la institución formativa FUNIBER y se han obtenido resultados en forma de reportes, análisis y un cuadro de mando.

2 Entorno sobre el que se aplica la propuesta

El colectivo de profesores de la asignatura presencial diseñó un curso virtual en la plataforma Moodle con el objetivo de apoyar las actividades presenciales así como soportar las actividades prácticas de laboratorio y el trabajo independiente.

2.1 Sistemas informáticos

La solución propuesta fue instalada en un servidor del departamento docente donde los profesores tienen acceso mediante la red. Desde cada terminal sólo era necesario disponer de un navegador para conectarse a la aplicación, independientemente del sistema operativo que tuviera instalado.

Debido a que el sistema fue utilizado por pocos profesores del mismo departamento, la computadora destinada como servidor no requirió disponer de altas prestaciones.

La configuración técnica de la solución fue la siguiente:

- Máquina virtual de Java, versión 6 o superior;
- Servidor de base de datos MySQL, versión 5.5 o superior.
- Espacio en disco duro de 3 gigabytes para almacenar los datos generados por la herramienta en la elaboración de las diferentes vistas de análisis, reportes y cuadros de mando integral.

El colectivo pedagógico de la asignatura estuvo conformado por tres profesores, donde uno de ellos ejerció como jefe de asignatura.

2.2 Asignatura de estudio

La propuesta de solución fue aplicada sobre una asignatura que se impartía mediante el uso simultáneo de un entorno virtual de aprendizaje y sesiones presenciales.

Esta asignatura contó con 79 estudiantes, 73 del año (distribuidos en 4 grupos) y 6 estudiantes de años anteriores que repetían los estudios. Estos estudiantes conformaron un grupo aparte y fueron tenidos en cuenta por la importancia de seguir sus actividades en pos de su éxito docente

La asignatura contó con 32 actividades, de las cuales 18 se impartieron en las mediante el entorno virtual de aprendizaje Moodle montado en la institución.

Las actividades y recursos montados en el curso virtual fueron agrupados por semanas, las cuales fueron visualizándose de manera progresiva. En la figura inferior se muestra una sección de la asignatura virtual.



Figura 28 Curso virtual de la asignatura

2.3 Actividades de evaluación

El módulo Tarea fue la herramienta más utilizada en el diseño del espacio virtual. El colectivo de asignatura elaboró una serie de actividades de evaluación, alrededor de todo el semestre, para controlar el aprendizaje de los estudiantes y contribuir al cumplimiento de los objetivos de la asignatura. Como actividades de evaluación se entienden las evaluaciones realizadas normalmente

en una asignatura, en ella se incluyen las evaluaciones parciales como los trabajos de control o las frecuentes como los trabajos independientes. En total se elaboraron un poco más de 30 actividades de evaluación, pero sólo fueron consideradas para la solución 27.

En las actividades el estudiante debe subir al entorno virtual un archivo como evidencia del trabajo realizado, el cual sería revisado posteriormente por su profesor.

Las actividades de evaluación son divididas en tipos específicamente escogidos para realizar los análisis posteriores:

| Tipo de actividades evaluadas | Descripción |
|--------------------------------------|--|
| Trabajo integral en clase | Se trata de las tareas que se desarrollan en clase. |
| Trabajo independiente, | Se trata de las actividades ubicadas en el entorno virtual de aprendizaje cuya elaboración es voluntaria pero muy recomendable |
| Trabajo de control | Se trata de una evaluación parcial que se desarrolla tanto en el campus. |
| Taller | Se trata de talleres virtuales donde los estudiantes aprenden mediante prácticas reales. |
| Tarea extra clase. | Se trata de tareas adicionales que el profesor plantea en clase |

Tabla 31 Tipología de las actividades de evaluación evaluadas


Asimismo, cada una de las actividades de evaluación está asociada a algún tema de la asignatura el cual también se usará para los análisis de la actividad de los estudiantes y profesores.


2.4 Habilidades de evaluación

Aparte de las actividades de evaluación, el colectivo pedagógico definió las habilidades que debían poner en práctica cada estudiante en la realización de cada actividad de evaluación. Cuando un profesor revisaba una evaluación de un determinado estudiante debía establecer el resultado en cada habilidad y la evaluación final de la actividad.

El colectivo no estableció ninguna relación restrictiva entre el resultado de las habilidades y el resultado de la evaluación asociada a tales habilidades, dejando que cada profesor decidiera sobre la base de su criterio y experiencia. En la figura inferior se ilustra las habilidades a evaluar dentro una actividad de evaluación.

Estado de la entrega

| | |
|------------------------|--|
| Estado de la entrega | Enviado para calificar |
| Status de calificación | Calificado |
| Fecha de entrega | martes, 18 marzo 2014, 2:25 a.m. |
| Tiempo restante | La tarea fue enviada 6 días 12 horas antes |
| Última modificación | martes, 11 marzo 2014, 1:52 p.m. |
| Envíos de archivo |  Tarea5.zip |
| Comentarios al envío | ▶ Comentarios (0) |

Calificación sobre 5  4.00 **Nota Final**

Inclusión de hojas de estilo: Excelente

Trabajo con selectores: Bien

Pseudoelementos: Bien

Calificación actual en el libro: 5.00



Habilidades evaluadas

Figura 29 Ejemplo de habilidades de evaluación

3 Fase de análisis de requerimientos

La fase del análisis de requerimientos se ha llevado a cabo recolectando, primeramente, las necesidades de información para la herramienta de analítica de aprendizaje, obteniendo de esta manera las preguntas claves que permiten guiar la solución propuesta.

De lo anterior se han obtenido un conjunto de preguntas las cuales han sido analizadas para identificar los indicadores, las diferentes dimensiones de análisis y su resolución

Finalmente con la información anterior se ha procedido a diseñar un modelo conceptual del almacén de datos.

A continuación se detalla cada una de las subfases realizadas.

3.1 Identificar preguntas

El gerente del conocimiento ha mantenido diversas reuniones con los usuarios finales de la herramienta, es decir aquellos usuarios que van a tomar decisiones a partir de la información generada al final del proceso. En concreto se ha involucrado a los perfiles siguientes:

- Profesores con altas habilidades tecnológicas y con un marcado carácter proactivo y constructivo.
- Coordinadores de área como responsables del trabajo de los profesores
- Directores académicos como responsables de los programas formativos.

Asimismo se ha hecho partícipe a un miembro del equipo informático de la institución, que será quien implementará la solución desde el punto de vista tecnológico.

Debido a que los datos almacenados en el Entorno Virtual de Enseñanza y Aprendizaje abarcan muchas posibles circunstancias, se decidió acotar el alcance del proyecto en tres hechos fundamentales:

- Accesos al entorno virtual,
- Evaluaciones realizadas y
- Habilidades evaluadas.

Teniendo en cuenta el alcance inicial de aplicación de este proyecto, en las reuniones se ha identificado un conjunto de preguntas cuya respuesta debería generar nuevo conocimiento, útil para tomar decisiones

La tabla inferior recoge las preguntas genéricas identificadas, teniendo en cuenta que las preguntas planteadas en las reuniones inicialmente fueron más específicas. En este sentido la mayoría de preguntas indicadas debajo fueron planteadas en términos más específicos, por ejemplo según periodos de tiempo, lugar de acceso, cursos, etc.

No obstante la utilidad del sistema quedaría muy acotada si sólo se consideraran ciertos rangos de dimensiones para cada pregunta. En este sentido se encaminó a los tomadores de decisiones a que plantearan las preguntas en términos más genéricos puesto que la capa de dimensiones se aplicaría sobre éstas preguntas más adelante.

| Número | Pregunta identificada |
|--------|--|
| 1 | ¿Cuál es la cantidad de accesos realizados por un usuario? |
| 2 | ¿Cuál es la cantidad de acciones realizadas por un usuario? |
| 3 | ¿Cuál es la cantidad de módulos visitados por un usuario? |
| 4 | ¿Cuál es la cantidad de cursos visitados por un usuario? |
| 5 | ¿Desde qué lugares realiza los accesos un usuario? |
| 6 | ¿Cuál es la cantidad de evaluaciones que debió realizar un usuario? |
| 7 | ¿Cuáles son las evaluaciones realizadas a un usuario? |
| 8 | ¿Cuál es el resultado de las evaluaciones de un usuario? |
| 9 | ¿Cuál es el número de ocasiones en que un usuario debió realizar una actividad donde se evalúan sus habilidades? |
| 10 | ¿Cuál es el número de ocasiones en que un usuario realizó una actividad donde se evalúan sus habilidades? |
| 11 | ¿Cuál es el resultado de las evaluaciones de las habilidades de un usuario? |

Tabla 32 Preguntas genéricas identificadas en el análisis de requerimientos

3.2 Identificar indicadores, dimensiones y resolución.

Tras las reuniones se han obtenido un conjunto de preguntas las cuales han sido transformadas en indicadores y se les ha asociado las dimensiones sobre las que se realizan los análisis.

En esta fase, tal como se acaba de comentar, se han identificado los indicadores que son de interés para los usuarios de la solución (profesores, coordinadores y directores académicos). Las dimensiones se refieren a los objetos mediante los cuales se desea examinar las métricas con el fin de responder a las preguntas planteadas.

En la tabla inferior se muestran las preguntas con sus respectivos métricas y dimensiones. En esta tabla se muestran más preguntas que las mostradas en la

anterior esto se debe a que se han reflejado las preguntas derivadas para sustentar todos los indicadores y dimensiones seleccionadas. Nótese que la dimensión tiempo se ha omitido aun cuando siempre está presente en cualquier informe o consulta. Se ha resaltado su presencia en aquellas consultas con mayor incidencia temporal a efectos de resaltar su importancia.

| No | Pregunta | Hecho | Métrica | Dimensiones |
|-----|---|---------|------------|-----------------------------|
| 1 | ¿Cuál es la cantidad de accesos realizados por un usuario? | Accesos | Frecuencia | Usuario |
| 1.1 | ¿Cómo se comporta el acceso de un usuario según un determinado periodo de tiempo? | Accesos | Frecuencia | Usuario Tiempo |
| 1.2 | ¿Cuál es la cantidad de accesos realizados por un usuario según sus acciones? | Accesos | Frecuencia | Usuario Acción |
| 2 | ¿Cuál es la cantidad de acciones realizadas por un usuario? | Accesos | Frecuencia | Usuario Acción |
| 2.1 | ¿Cuál es la cantidad de acciones realizadas por un usuario según el módulo? | Accesos | Frecuencia | Usuario Acción Módulo |
| 3 | ¿Cuál es la cantidad de módulos visitados por un usuario? | Accesos | Frecuencia | Usuario Módulo |
| 3.1 | ¿Cuál es la cantidad de módulos visitados por un usuario según el curso? | Accesos | Frecuencia | Usuario Módulo Curso |
| 4 | ¿Cuál es la cantidad de cursos visitados por un usuario? | Accesos | Frecuencia | Usuario Curso |
| 5 | ¿Desde qué lugares realiza los accesos un usuario? | Accesos | Frecuencia | Usuario Lugar |
| 5.1 | ¿Cuál es la cantidad de lugares desde donde accede un usuario? | Accesos | Frecuencia | Usuario Lugar |

| No | Pregunta | Hecho | Métrica | Dimensiones |
|-----------|--|-----------------------------|--|---|
| 6 | ¿Cuál es la cantidad de evaluaciones que debió realizar un usuario? | Evaluaciones | Frecuencia | Usuario Actividad de evaluación |
| 6.1 | ¿Qué actividades de evaluación debió realizar un usuario? | Evaluaciones | Frecuencia | Usuario Actividad de evaluación |
| 7 | ¿Cuáles son las evaluaciones realizadas a un usuario? | Evaluaciones | Frecuencia | Usuario Actividad de evaluación |
| 8 | ¿Cuál es el resultado de las evaluaciones de un usuario? | Evaluaciones | Resultado Resultado máximo. Resultado mínimo. Promedio de realizadas. | Usuario Actividad de evaluación Resultado |
| 9 | ¿Cuál es el número de ocasiones en que un usuario debió realizar una actividad donde se evalúan sus habilidades? | Evaluaciones de Habilidades | Frecuencia | Usuario Habilidad |
| 9.1 | ¿Qué habilidades debió evaluar un usuario? | Evaluaciones de Habilidades | Frecuencia | Usuario Habilidad |
| 10 | ¿Cuál es el número de ocasiones en que un usuario realizó una actividad donde se evalúan sus habilidades? | Evaluaciones de Habilidades | Frecuencia | Usuario Actividad de evaluación Habilidad |
| 10.1 | ¿En qué actividades de evaluación se evaluaron las habilidades de un usuario? | Evaluaciones de Habilidades | Frecuencia | Usuario Actividad de evaluación Habilidad |

| No | Pregunta | Hecho | Métrica | Dimensiones |
|----|---|-----------------------------|--|---|
| 11 | ¿Cuál es el resultado de las evaluaciones de las habilidades de un usuario? | Evaluaciones de Habilidades | Resultado Resultado máximo. Resultado mínimo. Promedio de realizadas. | Usuario Actividad de evaluación Habilidad |

Tabla 33 Hechos, métricas y dimensiones desglosadas según las preguntas que definen los requisitos de información

Respecto al nivel de resolución o granularidad de cada binomio hecho-dimensión, se ha establecido que los valores son los que se reflejan en la tabla inferior:

| Hecho-dimensión | Resoluciones |
|--------------------------------------|--|
| Accesos-Usuario | Usuario individual (id y nombre) Grupo |
| Accesos-Tiempo | Mes Semana del curso Día Día de la semana Hora Minuto |
| Accesos-Acción | Acción unitaria (id y nombre de acción) |
| Accesos-Módulo | Módulo unitario (id y nombre de módulo) |
| Accesos-Curso | Curso unitario (id y nombre del curso) |
| Accesos-Lugar | Área Ip |
| Evaluaciones-Usuario | Usuario individual (id y nombre) Grupo |
| Evaluaciones-Actividad de evaluación | Tipo Tema Nombre |

| Hecho-dimensión | Resoluciones |
|-------------------------------------|---|
| Evaluaciones-Resultado | Tipo resultado |
| Habilidades-Usuario | Usuario individual (id y nombre) Grupo |
| Habilidades-Habilidad | Tema Nombre |
| Habilidades-Actividad de evaluación | Tipo Tema Nombre |

Tabla 34 Resoluciones para cada binomio hecho-dimensión

En total se han identificado los siguientes componentes necesarios para el diseño del modelo conceptual:

- Tres hechos: Accesos, evaluaciones y evaluaciones de habilidades
- Nueve dimensiones: tiempo, usuario, acción, módulo, curso, lugar, actividad de evaluación, tipo de resultado y habilidad, y
- Distintas métricas que varían según el hecho analizado: frecuencia como medida de cantidad de accesos, usuarios, módulos, acciones, cursos, lugares y evaluaciones y habilidades enviadas (y no enviadas); promedio de evaluaciones y habilidades; resultado de evaluaciones y habilidades; evaluación y evaluación de habilidad máxima y mínima; y finalmente porcentaje de evaluaciones y habilidades enviadas.

3.3 Realizar el modelo conceptual

Tras la recopilación de la información relativa a las dimensiones, hechos, métricas y resoluciones necesarias para responder a las preguntas de los tomadores

de decisiones, se ha diseñado un modelo conceptual distinto para cada hecho identificado en el paso anterior, según lo cual se han obtenido los siguientes modelos conceptuales:

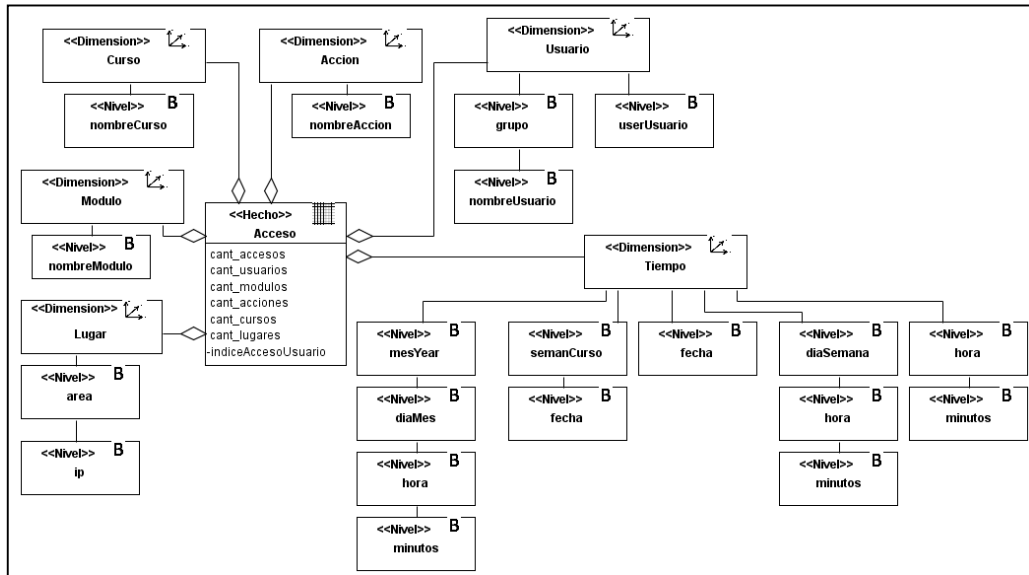


Figura 30 Modelo conceptual para el hecho Acceso

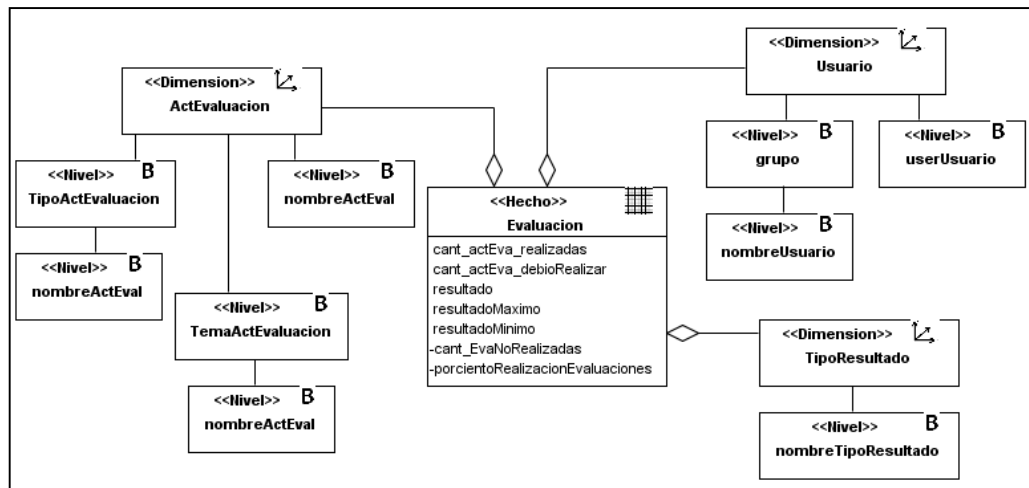


Figura 31 Modelo conceptual para el hecho Evaluación

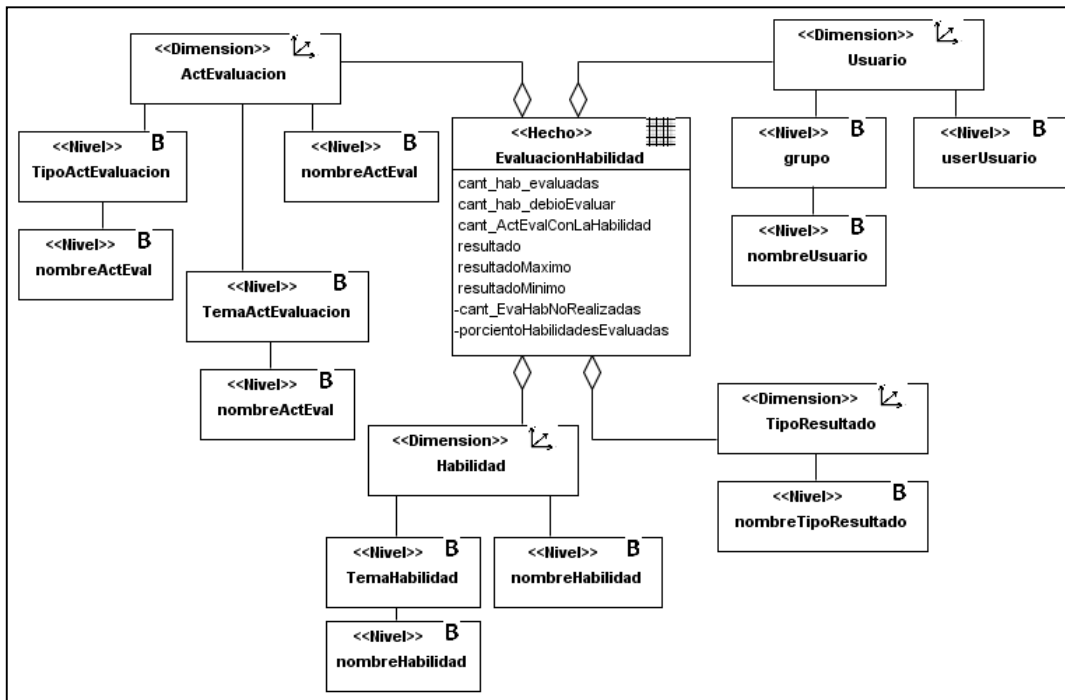


Figura 32 Modelo conceptual para el hecho Evaluación de habilidades

Nótese que los tres modelos comparten dimensiones, siguiendo la visión de Kimball (1996) quien considera al DW como un conjunto de DM que comparten sus dimensiones. Para este caso en concreto se comparten las siguientes dimensiones entre los hechos:

- Dimensión usuario: compartida por los tres hechos. Es algo que resulta obvio, pues en todos los hechos existen usuarios que los desarrollan y las características de éstos son comunes para cualquier hecho.

- Dimensiones tipo resultado y actividad de evaluación se comparten sólo entre los hechos relativos a las evaluaciones, ya sean evaluaciones simples o evaluaciones de habilidades.

4 Fase de análisis y elección de las fuentes de datos

La principal fuente de datos de la solución corresponde a la base de datos del Entorno Virtual de Enseñanza y Aprendizaje de FUNIBER que emplea el sistema Moodle. La base de datos de Moodle en su versión 2.3.4 está conformada por 299 tablas, la mayoría de ellas con más de 5 columnas. A continuación se presentan las tablas y aquellos campos que son de interés para la propuesta de solución. Más adelante se explican otras fuentes de datos utilizadas.

| Nombre: mdl_log | | |
|---|----------------|--|
| <i>Tabla que almacena los log del sistema, es decir, toda la actividad que se realiza sobre la plataforma. En esta tabla no se guarda información particular sobre la interacción de los usuarios con módulos</i> | | |
| Atributo | Tipo de | Descripción |
| id | BIGINT | Identificador del acceso |
| time | BIGINT | Tiempo en que se realizó el acceso en formato aaaa/mm/dd/ hh:mm. |
| userid | BIGINT | Identificador del usuario que accedió |
| ip | VARCHAR | Ip de la máquina desde donde se realizó el acceso |
| course | BIGINT | Identificador del curso al cual se accedió |
| module | VARCHAR | Nombre del módulo al cual se accedió |
| cmid | BIGINT | Identificador de la relación curso-módulo (mdl_course_modules) |
| action | VARCHAR | Nombre de la acción que se realizó |
| info | VARCHAR | Almacena información estrechamente relacionada al módulo y la acción ejecutada. Por ejemplo: Si la acción es crear un curso, se almacena la información del nombre y el id del curso. Otro ejemplo: si accede un usuario al sistema, se almacena el id del usuario que entró. En algunas ocasiones esta información es redundante porque ya está en los atributos userid |

Tabla 35 Tabla mdl_log

| Nombre: mdl_user | | |
|---|---------------------|---|
| <i>Tabla que almacena información sobre los usuarios del sistema.</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador del usuario |
| username | VARCHAR | Nombre del usuario |
| firstname | VARCHAR | Nombre de la persona a la que pertenece el |
| lastname | VARCHAR | Apellido de la persona a la que pertenece el usuario |
| firstaccess | BIGINT | Tiempo en que el usuario tuvo el primer acceso al sistema |
| lastaccess | BIGINT | Tiempo en que el usuario tuvo el último acceso al sistema |
| lastlogin | BIGINT | Tiempo en que el usuario tuvo el último login |
| currentlogin | BIGINT | Tiempo en que el usuario tuvo el login |
| lastip | VARCHAR | Ip del último acceso del usuario |
| timecreated | BIGINT | Tiempo en que fue creado el usuario |
| timemodified | BIGINT | Tiempo cuando fue modificado el usuario |

Tabla 36 Tabla mdl_user

| Nombre: mdl_role | | |
|---|---------------------|-----------------------|
| <i>Tabla que almacena los roles del sistema. Por ejemplo: Estudiante, profesor,</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador del rol |
| name | VARCHAR | Nombre del rol |

Tabla 37 Tabla mdl_role

| Nombre: mdl_role_assignments | | |
|---|---------------------|--|
| <i>Tabla que almacena la relación entre los usuarios y los roles del sistema.</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la relación entre el rol y el |
| roleid | BIGINT | Identificador del rol |
| userid | BIGINT | Identificador del usuario |

Tabla 38 Tabla mdl_role_assignments

| Nombre: mdl_course | | |
|---|---------------------|---------------------------------------|
| <i>Tabla que almacena los cursos del sistema.</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador del curso |
| fullname | VARCHAR | Nombre completo del curso |
| shortname | VARCHAR | Nombre corto (Abreviatura) del curso |
| summary | LongText | Descripción del curso |
| startdate | BIGINT | Tiempo de comienzo del curso |
| timecreated | BIGINT | Tiempo en que fue creado el curso |
| timemodified | BIGINT | Tiempo en que fue modificado el curso |

Tabla 39 Tabla mdl_course

| Nombre: mdl_course_modules | | |
|---|---------------------|--|
| <i>Tabla que almacena la relación entre los cursos y los módulos.</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la relación |
| course | BIGINT | Identificador del curso |
| module | BIGINT | Identificador del módulo |
| instance | BIGINT | Identificador de la instancia del módulo dentro de los de su tipo. Esto sucede porque un curso puede tener varios módulos del mismo tipo y cada uno tiene un identificador diferente. Este identificador debe coincidir con el id del tipo de módulo correspondiente. Por ejemplo: El fórum con instancia 2 debe |
| added | BIGINT | Tiempo en que fue adicionado |

Tabla 40 Tabla mdl_course_modules

| Nombre: mdl_groups | | |
|--|---------------------|---|
| <i>Tabla que almacena los grupos creados en el sistema</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador del grupo |
| courseid | BIGINT | Identificador del curso |
| name | VARCHAR | Nombre del grupo |
| description | LONGTEXT | Descripción del grupo |
| timecreated | BIGINT | Tiempo en el cual fue creado el grupo |
| timemodified | BIGINT | Tiempo en el cual fue modificado el grupo |

Tabla 41 Tabla mdl_groups

| Nombre: mdl_groups_members | | |
|--|---------------------|--|
| <i>Tabla que almacena la relación entre los grupos y usuarios.</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la tabla |
| groupid | BIGINT | Identificador del grupo |
| userid | BIGINT | Identificador del usuario |
| timeadded | BIGINT | Tiempo en que fue agregado el usuario al grupo |

Tabla 42 Tabla mdl_groups_members

| Nombre: mdl_assign | | |
|--|---------------------|---|
| <i>Tabla que almacena las actividades de evaluación existentes en los cursos</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la tabla |
| course | BIGINT | Identificador del curso |
| name | BIGINT | Nombre de la actividad |
| grade | BIGINT | Resultado máximo que puede alcanzar un estudiante |

Tabla 43 Tabla mdl_assign

| Nombre: mdl_assign_grades | | |
|--|---------------------|---|
| <i>Tabla que almacena las actividades de evaluación que tienen asignado los usuarios así como el resultado</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la tabla |
| assignment | BIGINT | Identificador de la actividad de evaluación |
| userid | BIGINT | Identificador del usuario |
| grade | DECIMAL | Resultado obtenido por estudiante en la actividad |

Tabla 44 Tabla mdl_assign_grades

| Nombre: mdl_grade_outcomes | | |
|--|---------------------|---------------------------------|
| <i>Tabla que almacena las habilidades de evaluación existentes en los cursos</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la tabla |
| courseid | BIGINT | Identificador del curso |
| shortname | VARCHAR | Nombre corto de la habilidad |
| fullname | LONGTEXT | Nombre completo de la habilidad |

Tabla 45 Tabla mdl_grade_outcomes

| Nombre: mdl_grade_items | | |
|---|---------------------|--|
| <i>Tabla que almacena los ítems para evaluar a un estudiante. Los ítems son las actividades de evaluación y</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la tabla |
| courseid | BIGINT | Identificador del curso |
| itemname | VARCHAR | Nombre del ítems |
| iteminstance | BIGINT | Identificador de la actividad de evaluación al |

Tabla 46 Tabla mdl_grade_items

| Nombre: mdl_grade_grades | | |
|--|---------------------|--|
| <i>Tabla que almacena los ítems que tienen asignado los usuarios así como el resultado que</i> | | |
| Atributo | Tipo de Dato | Descripción |
| id | BIGINT | Identificador de la tabla |
| itemid | BIGINT | Identificador del ítems |
| userid | BIGINT | Identificador del usuario |
| finalgrade | DECIMAL | Resultado que obtuvo el estudiante en el |

Tabla 47 Tabla mdl_grade_grades

En algunas de las tablas que se han mostrado existen datos que se repiten y valores que tienen diferentes significados en dependencia de otras columnas. Un ejemplo de esto, es que en la tabla mdl_grade_items están incluidas las actividades de evaluación presentes en la tabla mdl_assign, más las habilidades evaluadas. Por otra parte la columna finalgrade, de la tabla mdl_grade_grades, tiene un significado diferente en dependencia de lo que se esté evaluando: una habilidad o una actividad de evaluación; esto tiene que ver con la escala de evaluación usada en cada caso.

Aparte de los datos almacenados por Moodle, fueron necesarios tres archivos con extensión XML para complementar la información a analizar:

- El archivo datosIP.xml relaciona las diferentes subredes de la institución con las zonas a las que corresponden.
- El archivo datosEvaluacion.xml relaciona las actividades de evaluación con el tema del curso al que pertenecen y el tipo de actividad (trabajo de control, trabajo independiente, trabajo en clase, etc.).
- Por último, el archivo datosHabilidades.xml relaciona las habilidades con el tema del curso asociado.

4.1 Conformar las métricas

La tabla inferior muestra, para cada métrica, la dimensión a la que pertenece, una breve descripción y la forma de calcularla.

| Hecho | Métrica | Descripción | Cálculo |
|--------|---------------|---|--|
| Acceso | cant_accesos | Cantidad de accesos. Es la frecuencia o cantidad de ocurrencias de todas las | Función de agregación: count Aplicada a la dimensión Acción |
| Acceso | cant_usuarios | Cantidad de usuarios. Es la cantidad de usuarios diferentes que han accedido. | Función de agregación: distinct-count Aplicada a la dimensión Usuario |

| Hecho | Métrica | Descripción | Cálculo |
|------------|------------------------|--|---|
| Acceso | cant_modulos | Cantidad de módulos. Es la cantidad de módulos diferentes que han sido visitados. Es útil para | Función de agregación: distinct-count Aplicada a la dimensión Módulo |
| Acceso | cant_acciones | Cantidad de acciones. Es la cantidad de acciones diferentes realizadas. Es útil para | Función de agregación: distinct-count Aplicada a la dimensión Acción |
| Acceso | cant_cursos | Cantidad de cursos. Es la cantidad de cursos diferentes que | Función de agregación: distinct-count Aplicada a la dimensión Curso |
| Acceso | cant_lugares | Cantidad de lugares. Es la cantidad de lugares diferentes desde donde se realizan los accesos. Es útil para determinar | Función de agregación: distinct-count Aplicada a la dimensión Lugar |
| Acceso | indiceAccesoUsuario | Relación entre la cantidad de accesos y la cantidad de usuarios. Es útil para poder | cant_accesos/cant_usuarios |
| Evaluación | cant_actEva_realizadas | Cantidad de evaluaciones realizadas. Es útil para determinar la cantidad de evaluaciones de un estudiante o | Función de agregación: sum Aplicada a una columna de la tabla hecho que especifica si la evaluación fue realizada o no |

| Hecho | Métrica | Descripción | Cálculo |
|------------|---------------------------|--|---|
| Evaluación | cant_actEva_debioRealizar | Cantidad de evaluaciones que debieron ser realizadas. Es útil para determinar la cantidad de evaluaciones que debieron | Función de agregación: count Aplicada a una columna de la tabla hecho que especifica si la evaluación fue realizada o no |
| Evaluación | resultado | Resultado de las evaluaciones. Es útil para determinar el resultado promedio de las evaluaciones de | Función de agregación: avg Aplicada a una columna de la tabla hecho que especifica el resultado de la evaluación |
| Evaluación | resultadoMaximo | Resultado máximo de las evaluaciones. Es útil para determinar el resultado máximo de las evaluaciones de | Función de agregación: max Aplicada a una columna de la tabla hecho que especifica el resultado de la evaluación |
| Evaluación | resultadoMinimo | Resultado mínimo de las evaluaciones. Es útil para determinar el resultado mínimo de las evaluaciones de | Función de agregación: min Aplicada a una columna de la tabla hecho que especifica el resultado de la evaluación |
| Evaluación | cant_EvaNoRealizadas | Cantidad de evaluaciones no realizadas. Es útil para determinar la cantidad de evaluaciones | cant_actEva_debioRealizar - cant_actEva_realizadas |

| Hecho | Métrica | Descripción | Cálculo |
|----------------------|-----------------------------------|---|---|
| Evaluación | porcentajeRealizacionEvaluaciones | Porcentaje de realización de las evaluaciones. Es útil para tener una medida de las evaluaciones | $(\text{cant_actEva_realizadas} / \text{cant_actEva_debioRealizar}) * 100$ |
| Evaluación Habilidad | cant_hab_evaluadas | Cantidad de evaluaciones de habilidad realizadas. Es útil para determinar la cantidad de evaluaciones de | Función de agregación: sum Aplicada a una columna de la tabla hecho que especifica si la evaluación fue realizada o no |
| Evaluación Habilidad | cant_hab_debioEvaluar | Cantidad de evaluaciones de habilidad que debieron ser realizadas. Es útil para determinar la cantidad de evaluaciones de | Función de agregación: count Aplicada a una columna de la tabla hecho que especifica si la evaluación fue realizada o no |
| Evaluación Habilidad | cant_ActEvalConLaHabilidad | Cantidad de actividades de evaluación diferentes. Es útil para determinar la cantidad de | Función de agregación: distinct-count Aplicada a la dimensión Actividad de Evaluación. |
| Evaluación Habilidad | cant_EvaHabNoRealizadas | Cantidad de evaluaciones de habilidad no realizadas. Es útil para determinar la cantidad de | $\text{cant_hab_debioEvaluar} - \text{cant_hab_evaluadas}$ |

| Hecho | Métrica | Descripción | Cálculo |
|-------------------------|--------------------------------|---|--|
| Evaluación Habilidad | porcentajeHabilidadesEvaluadas | Porcentaje de realización de las evaluaciones de habilidad. Es útil para tener | $(\text{cant_hab_evaluadas} / \text{cant_hab_debioEvaluar}) * 100$ |
| Evaluación Habilidad | resultado | Resultado de las evaluaciones de habilidad. Es útil para determinar el resultado | Función de agregación: avg Aplicada a una columna de la tabla hecho que especifica el resultado de la evaluación de habilidad |
| Evaluación Habilidad | resultadoMaximo | Resultado máximo de las evaluaciones de habilidad. Es útil para determinar el resultado máximo de las | Función de agregación: max Aplicada a una columna de la tabla hecho que especifica el resultado de la evaluación de habilidad |
| Evaluación Habilidad | resultadoMinimo | Resultado mínimo de las evaluaciones. Es útil para determinar el resultado mínimo de las evaluaciones de habilidad de un estudiante, de | Función de agregación: min Aplicada a una columna de la tabla hecho que especifica el resultado de la evaluación de habilidad |

Tabla 48 Identificación de las métricas en la aplicación del modelo conceptual

4.2 Establecer correspondencias entre las fuentes de datos y el modelo conceptual

El objetivo de este paso es examinar los OLTP disponibles que contengan los datos requeridos y sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. La idea

principal es que todos los elementos del modelo conceptual estén correspondidos en los OLTP. La tabla inferior recoge estas correspondencias según los hechos escogidos en los anteriores pasos.

| Hecho | Correspondencias |
|--|---|
| Acceso | <p>La tabla mdl_course se relaciona con la dimensión Curso</p> <p>La tabla mdl_user se relaciona con la dimensión Usuario</p> <p>El campo module de la tabla mdl_log se relaciona con la dimensión Módulo porque en él se encuentran registrados todos los módulos sobre los cuales los usuarios realizan sus acciones</p> <p>El campo action de la tabla mdl_log se relaciona con la dimensión Acción porque en él se encuentran registrados todas las acciones realizadas por los usuarios</p> <p>El campo ip de la tabla mdl_log se relaciona con la dimensión Lugar porque en él se encuentran registrados todas las direcciones ip que fueron utilizadas para acceder. Esta información se complementa con la encontrada en el archivo datosIP.xml para agrupar las direcciones por zonas</p> <p>El campo time de la tabla mdl_log se relaciona con la dimensión Tiempo porque en él se encuentran registrados todos los momentos en los cuales se realizaron los accesos</p> <p>La tabla mdl_log se relaciona con el hecho Acceso. Esta tiene los datos para vincular todas las dimensiones</p> |
| Evaluación (no se vuelven a especificar las dimensiones compartidas con Acceso) | <p>La tabla mdl_assign se relaciona con la dimensión Actividad de Evaluación. Esta información se complementa con la encontrada en el archivo datosEvaluacion.xml para agrupar las actividades de evaluación por tema del curso y tipo.</p> <p>El campo grade de la tabla mdl_assign_grades se relaciona con la dimensión Tipo de Resultado porque en él se encuentran todos los posibles resultados que un estudiante puede adquirir en una evaluación.</p> <p>La tabla mdl_assign_grades se relaciona con el hecho Evaluación. Esta tiene los datos para vincular todas las dimensiones</p> |

| Hecho | Correspondencias |
|--|---|
| EvaluaciónHabilidad (no se vuelven a especificar las correspondencias con las dimensiones compartidas con Evaluación) | La tabla mdl_grade_outcomes se relaciona con la dimensión Habilidad. Esta información se complementa con la encontrada en el archivo datosHabilidades.xml para agrupar las habilidades por el tema del curso La tabla mdl_grade_grades se relaciona con el hecho EvaluacionHabilidad. Esta tiene los datos para vincular todas las dimensiones |

Tabla 49 Correspondencias entre las fuentes de datos y el modelo conceptual

5 Fase de elaboración del modelo lógico del almacén de datos

La propuesta de solución fue dividida en tres *data marts*, uno para cada aspecto del negocio: Acceso, Evaluación y EvaluacionHabilidad. Debido a que cada *data mart* tiene sólo una tabla hecho y varias dimensiones, que a su vez están relacionadas con otras, se decidió utilizar el esquema de Copo de Nieve. Este esquema se asemeja al modelo relacional por lo que se adapta mejor a la solución propuesta.

La aplicación de los estándares de codificación tuvo como principal premisa establecer los nombres de la manera más clara posible y que reflejen su significado en el negocio. Para el diseño de las tablas se decidió nombrar los campos mediante la concatenación del nombre de la tabla con el nombre propio del campo; solo en caso que los campos respondieran a una medida no se siguió esta nomenclatura, dejándolos solo con sus nombres propios. Los campos que se utilizaron para establecer una relación con otra tabla se nombraron igual que el campo de la tabla con la que están relacionados.

En las siguientes figuras se muestran los modelos lógicos de cada *data mart*, que en su totalidad conforman 18 tablas, 3 hechos y 15 dimensiones.

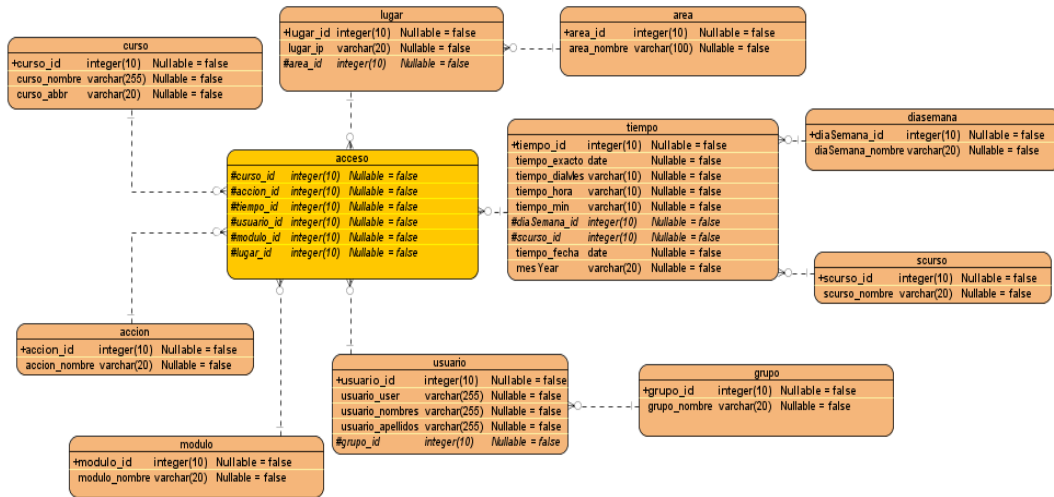


Figura 33 Modelo lógico del hecho Acceso

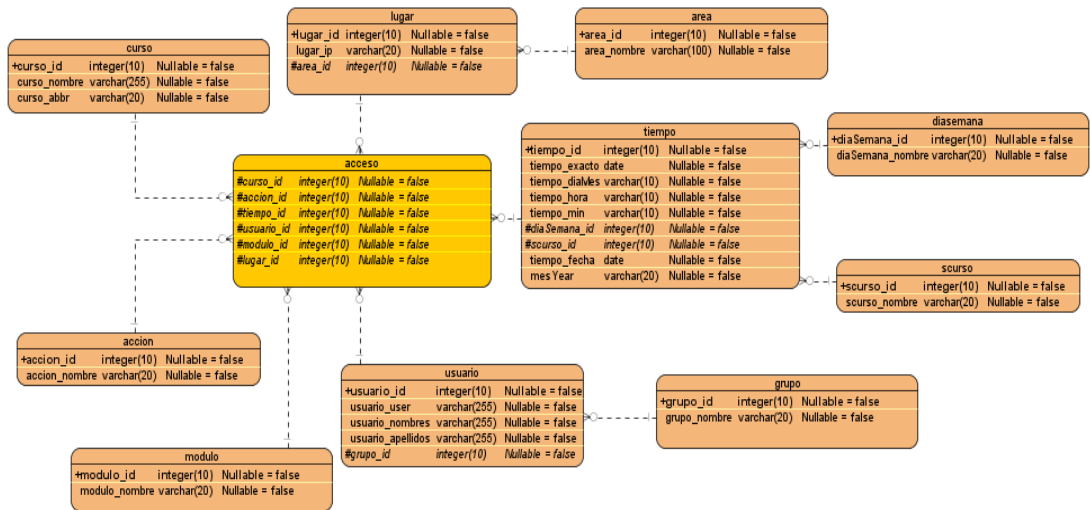


Figura 34 Modelo lógico del hecho Evaluación

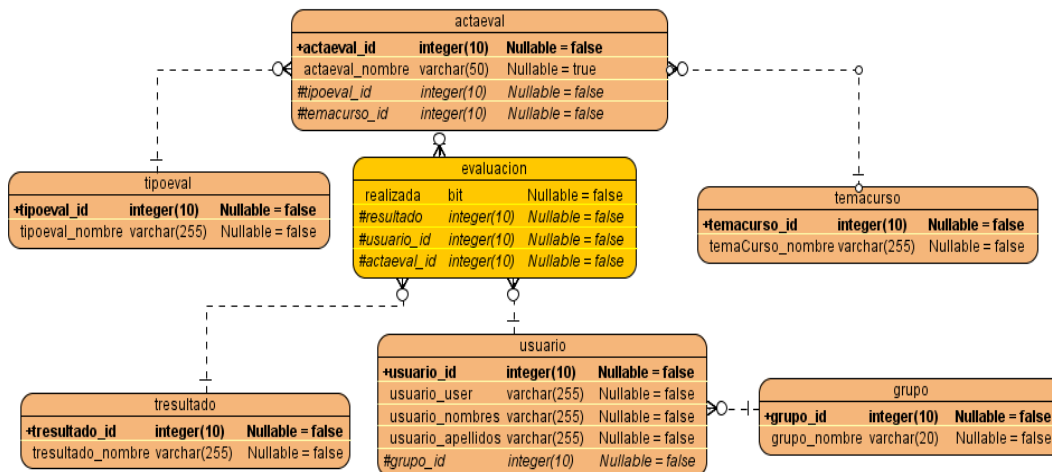


Figura 35 Modelo lógico del hecho Evaluación de habilidades

6 Fase de integración de los datos

En pos de definir una arquitectura de integración de datos se procede a resumir sus componentes:

- La base de datos del entorno virtual de enseñanza y aprendizaje de la institución montada sobre un servidor MySQL
- Los archivos datosIP.xml, datosEvaluacion.xml, datosHabilidades.xml contienen datos complementarios
- El proceso ETL, el cual es el punto intermedio entre la fuente de datos y el almacén. Se utiliza la herramienta Pentaho Data Integration
- El almacén de datos, destino donde se cargan los datos; este se encuentra montado sobre un servidor MySQL
- Como protocolo de comunicación se emplea TCP/IP

En la figura inferior se ilustran los componentes de la arquitectura de integración de datos anteriormente mencionados.

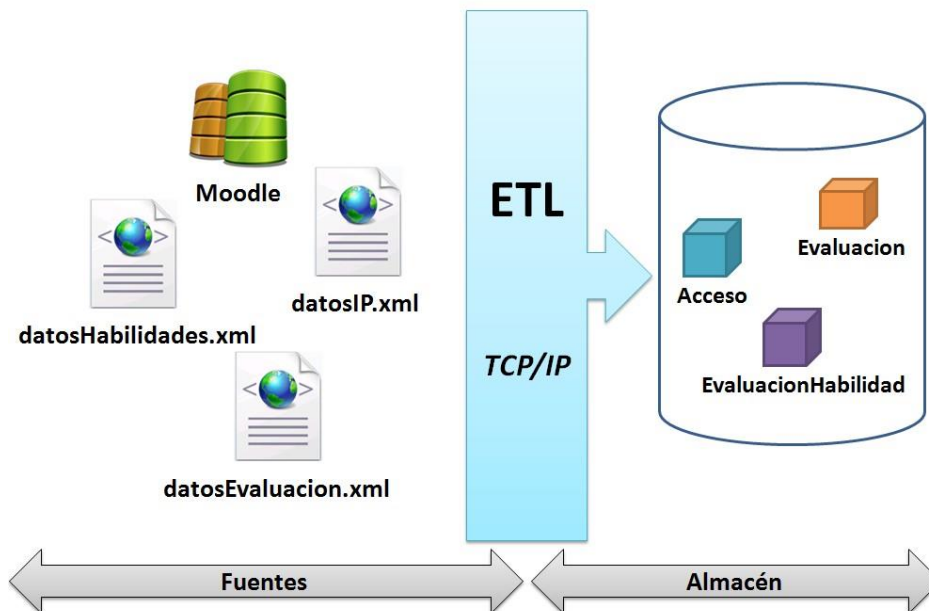


Figura 36 Arquitectura de integración de la solución

6.1 Implementar transformaciones y trabajos

La herramienta Pentaho Data Integration, mediante el componente visual Spoon, permite construir transformaciones y trabajos para desarrollar el proceso de extracción, transformación y carga. La transformación es el elemento básico de diseño de procesos ETL con Pentaho Data Integration; esta constituye un proceso que se compone de pasos o etapas, las cuales se conectan mediante saltos para compartir datos.

Los trabajos son conjuntos de transformaciones que tienen una relación conceptual y al ser agrupadas permiten ser ejecutadas de forma atómica, aunque hay que tener en cuenta que los trabajos pueden ejecutar otros tipos de tareas

utilitarias que no son encontradas dentro de una transformación. La herramienta de Pentaho brinda muchas facilidades para la implementación de los procesos ETL, una de ellas es la posibilidad de programar los trabajos para que puedan ser ejecutados en un momento determinado.

Como parte del desarrollo de la solución se implementaron 15 transformaciones y 4 trabajos, los cuales se resumen a continuación.

6.1.1 Trabajos

Con respecto a los trabajos, se diseñó uno para cada *data mart* y uno para el almacén. Los trabajos para poblar el almacén son ejecutados en dependencia del momento en que fue creado cada *data mart*. Esto se refiere a que el *data mart* Acceso se debe poblar primero, el de Evaluación de segundo y el de EvaluacionHabilidad de último. En la figura inferior se representa la secuencia seguida para poblar el almacén de datos.

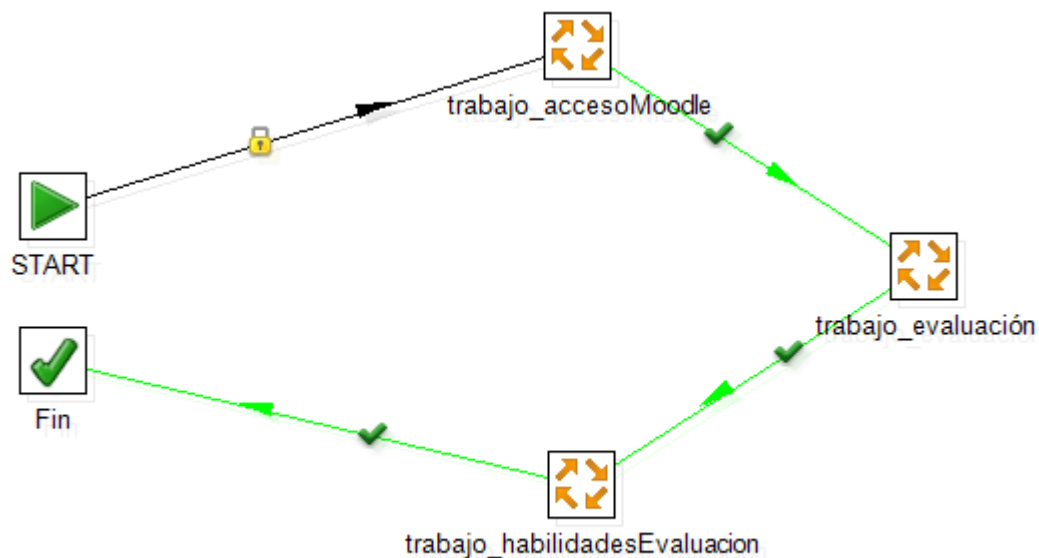


Figura 37 Trabajo para poblar el almacén de datos

Los trabajos para poblar cada *data mart* deben ejecutar las transformaciones en un orden definido, primero las dimensiones y al final la tabla hecho. Asimismo se tiene que tener en cuenta las relaciones que existen entre las dimensiones de forma que se cargan primero las no dependientes (ubicadas en los extremos del modelo lógico expresado en la Figura 33, Figura 34 y la Figura 35). Esto es así porque las dimensiones dependientes necesitan conocer de antemano las claves de las no dependientes de las, un ejemplo de ello es la dependencia que existe entre el lugar y el área, siendo el área independiente de cualquier otra dimensión y el lugar dependiente del área, por este motivo se carga antes el área y después el lugar, tal como se aprecia en la figura siguiente.

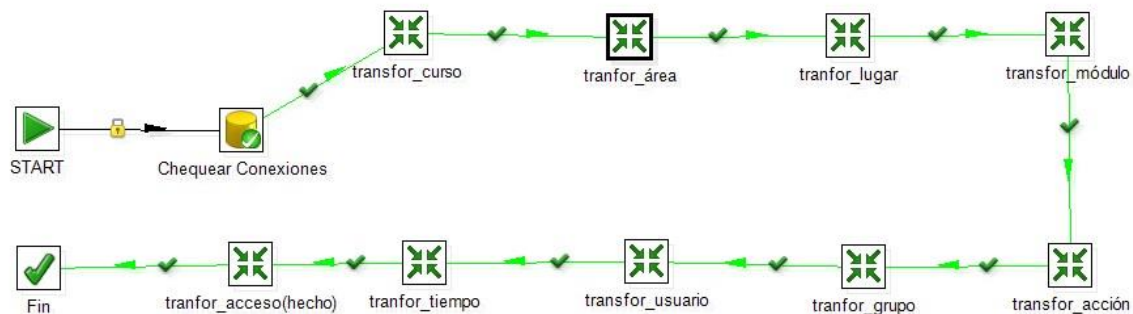


Figura 38 Trabajo para poblar el data mart Acceso

Un componente importante en el trabajo representado en la figura superior es el que chequea las conexiones utilizadas por las transformaciones. Los demás trabajos para el resto de los *data marts* fueron implementados de manera similar, teniendo en cuenta que no fue necesario cargar las dimensiones que son compartidas con trabajos ejecutados previamente. Este es el caso de las dimensiones usuario (compartida por los tres *data marts*) y las dimensiones tipo resultado y actividad de evaluación se comparten entre los hechos relativos a las evaluaciones, tal como se ha reflejado en el apartado 3.3 del presente capítulo.

A continuación se resumen las transformaciones más importantes que componen los trabajos recientemente mostrados.

6.1.2 Transformaciones

Algunas de las transformaciones más sencillas, las cuales consisten sólo en extraer los datos y cargarlos, son las que se muestran en la figura inferior:

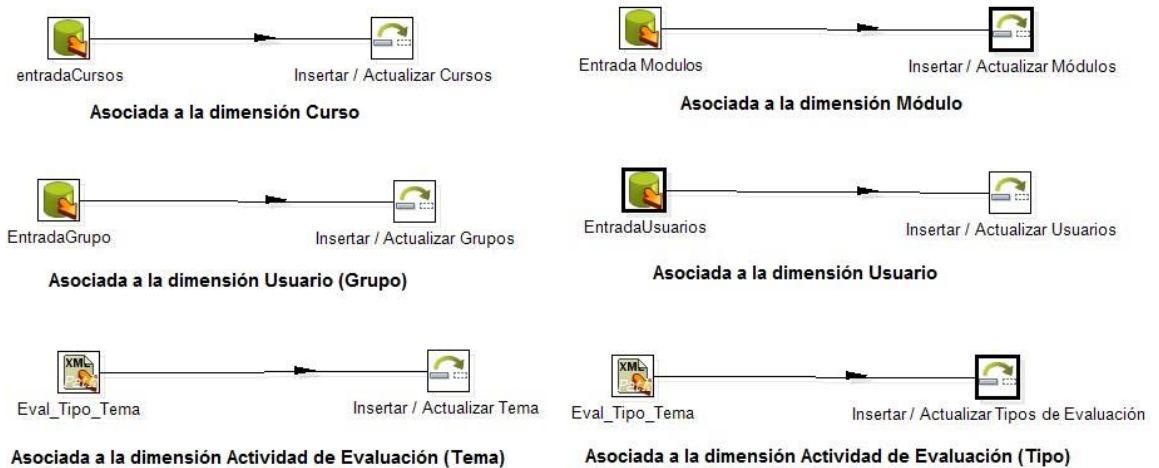


Figura 39 Ejemplos de transformaciones

En estas transformaciones se utilizaron tres componentes: Entrada de Tabla, Entrada de archivo XML e Insertar / Actualizar. Los dos primeros se utilizan para recibir datos y el tercero para almacenarlos o actualizarlos, según sea el caso, en una base de datos; este componente verifica la existencia del registro antes de cargarlo, lo que evita la sobre escritura.

Otras transformaciones sí conllevaron realizar cambios en los datos para adecuarse a los requerimientos del almacén, un ejemplo es la transformación para la dimensión Tiempo, la cual se muestra en la siguiente figura:

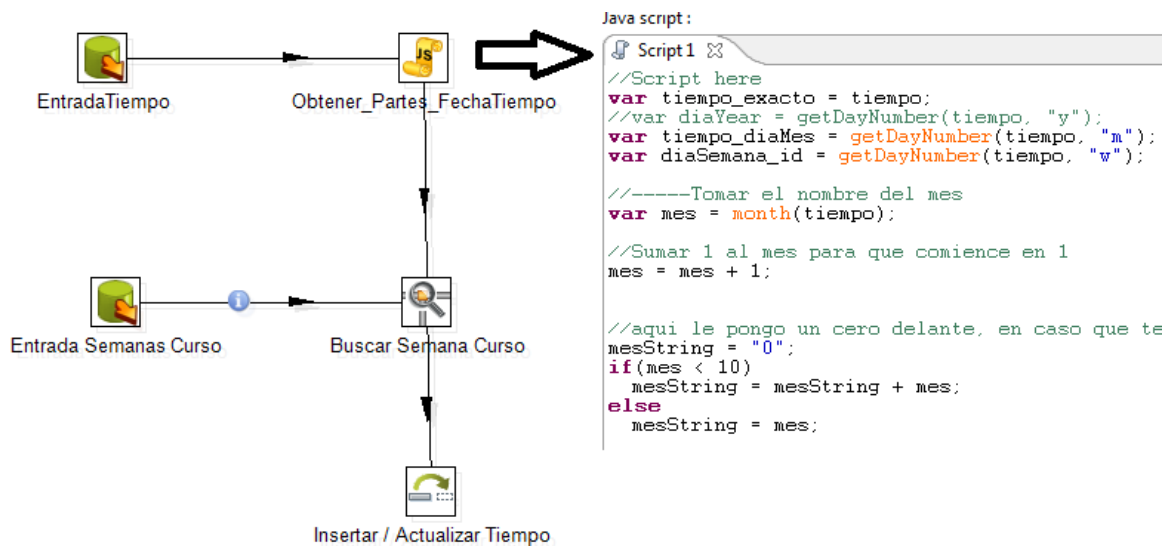


Figura 40 Transformación para la dimensión Tiempo

El tiempo fue cargado a partir del campo time de la tabla mdl_log; este campo tenía la fecha y la hora codificada en UNIX, por lo que fue necesario utilizar la función FROM_UNIXTIME() del lenguaje SQL para hacer el dato legible. Posteriormente, y como se ilustra en la figura superior, se utilizó un componente de Java Script para obtener las diferentes partes del tiempo.

Las transformaciones asociadas a las tablas hechos siempre son particulares por su capacidad de vincular todas las dimensiones, lo que en muchas ocasiones traer consigo realizar cambios, filtraciones y limpiezas. La transformación para el hecho Acceso no es un ejemplo de lo anteriormente planteado, por su parte las asociadas a los hechos Evaluación y EvaluacionHabilidad sí lo son. En las siguientes figuras se muestran tales transformaciones.

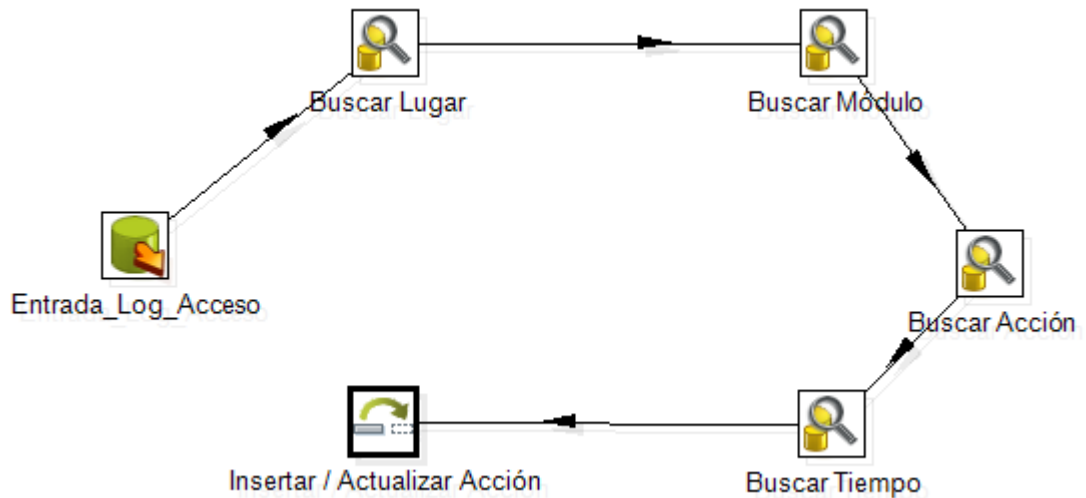


Figura 41 Transformación para el hecho Acceso

En la transformación de la figura superior se utilizó el componente Buscar en Base de Datos para encontrar las llaves de las dimensiones y conformar el hecho. Cabe resaltar que no fue necesario buscar las llaves de las dimensiones Usuario y Curso debido a que estos datos ya existen en la base de datos de Moodle.



Figura 42 Transformación para el hecho Evaluación

La transformación de la figura superior utiliza el componente Buscar en Flujo de Datos para encontrar las actividades de evaluación, en caso de no encontrarlas devuelve el valor -1, el cual es utilizado por el componente Filtrar Filas para descartar las actividades de evaluación que no son importantes para ser analizadas posteriormente. El componente de Java Script es utilizado para establecer si una evaluación fue realizada o no teniendo en cuenta el resultado obtenido y la etiqueta *submitted* existente en algunas evaluaciones. Esto se debe a que un estudiante pudo no haber subido la evaluación al Entorno Virtual de Enseñanza y Aprendizaje, por lo que no está marcado como *submitted*, sin embargo sí la entregó al profesor por otra vía y por tal motivo tiene registrado un resultado, en ese caso debe tomarse como que el estudiante realizó la evaluación.

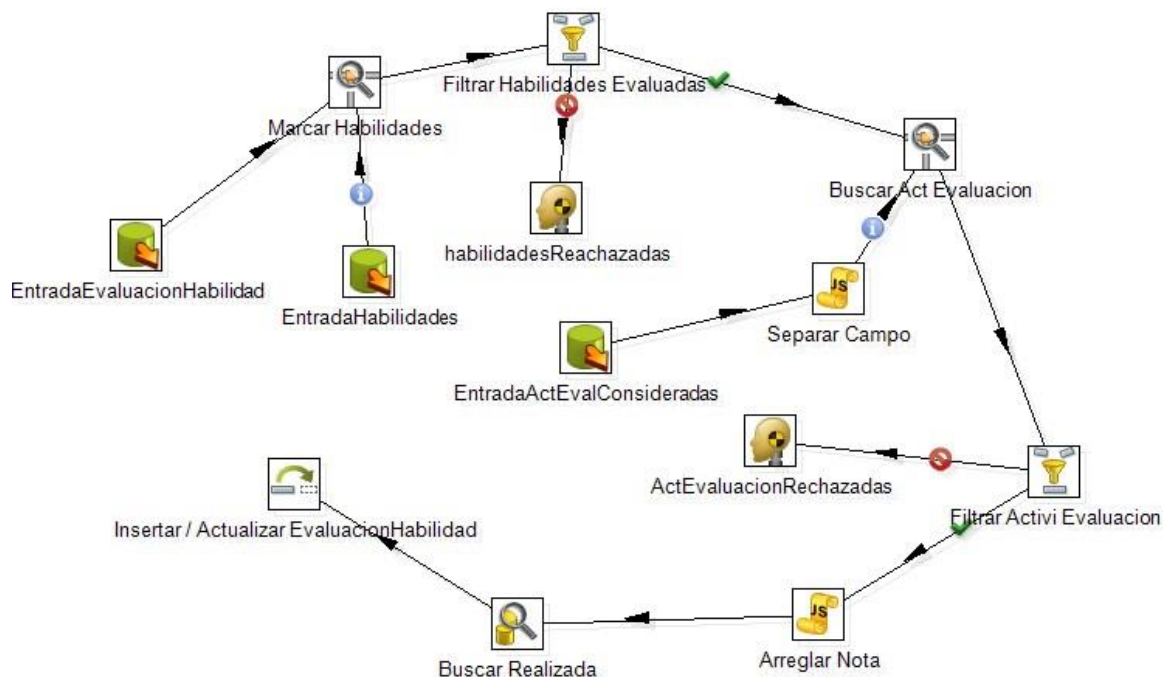


Figura 43 Transformación para el hecho EvaluacionHabilidad

La transformación de la figura superior es la que más componentes utiliza debido a que se cargan los datos desde las tablas mdl_grade_grades, mdl_grade_items, y mdl_assign. Estas tablas, aunque diferentes, comparten muchos de sus datos, por lo que es necesario realizar varias filtraciones para evitar repeticiones. De la misma forma que se realizó en la transformación de la Figura 2.12 no se tuvieron en cuenta las actividades de evaluación que no son importantes para ser analizadas posteriormente.

6.2 Implementación de la integración de los datos

Para realizar las tareas descritas en este apartado se configuró el trabajo descrito anteriormente (Figura 37 Trabajo para poblar el almacén de datos) en la computadora destinada como servidor para que fuera ejecutada a las 3:00 am de

cada día. Esta hora fue seleccionada para no penalizar el rendimiento del Entorno Virtual de Enseñanza y Aprendizaje.

Para programar la tarea fue necesario especificar la dirección donde se encuentra el archivo `Kitchen.bat` o `Kitchen.sh`, según sea el sistema operativo, seguido del trabajo (Figura 37) que fue implementado para poblar todo el almacén (`trabajo_almacenGeneral.kjb`).

7 Fase de uso de la solución

En esta fase se engloban las tareas de conformar los cubos multidimensionales, realizar los análisis, obtener los reportes y preparar los cuadros de mando integrales. A continuación se detallan cada una de las tareas anteriores.

7.1 Conformación de los cubos multidimensionales

Una vez se tienen las tablas pobladas con datos se procede a conformar el modelo multidimensional. Para ello se utilizó la herramienta Pentaho Schema Workbench.

Se conformaron 3 cubos, uno para cada *data mart*. En la figura inferior se ilustra una sección de los cubos implementados. Esta herramienta genera un archivo XML que permite publicar en el servidor Mondrian para poder realizar los análisis correspondientes.

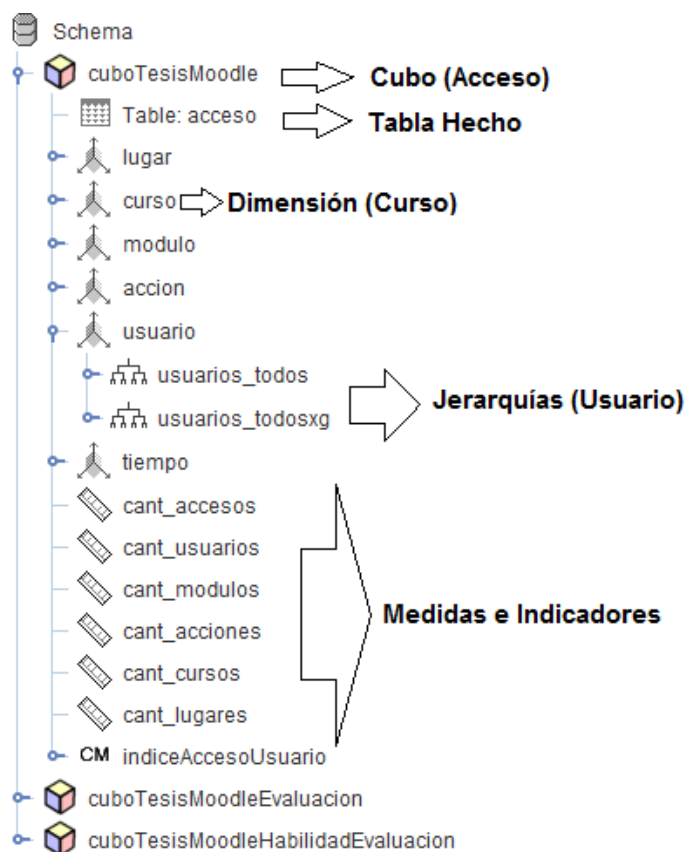


Figura 44 Implementación de los cubos multidimensionales

7.2 Análisis OLAP

Los cubos multidimensionales implementados por la herramienta Pentaho Schema Workbench fueron publicados en el servidor de Pentaho, listos para ser utilizados por Pentaho Analyzer. Debido al trabajo del colectivo de asignatura, se desarrollaron en todo el semestre poco más de 70 vistas de análisis con diferentes fines.

Para el uso de Pentaho Analyzer en la propuesta de solución se crearon 5 usuarios, uno para cada profesor y finalmente uno para el jefe de disciplina y otro para el jefe de departamento para que pudieran llevar a cabo el seguimiento

del trabajo en la asignatura. Estos cinco usuarios fueron divididos en dos roles: profesor y coordinador. A los profesores se les dio el permiso de crear vistas de análisis así como actualizarlas y eliminarlas, y al rol de coordinador se les permitió ejecutar una serie de vistas según sus necesidades de información.

Una de las primeras vistas elaboradas, para explorar los accesos al Entorno Virtual de Enseñanza y Aprendizaje, fue la relacionada al índice de acceso (Figura 45) por grupos de estudiantes. Esta permitió comparar el acceso de los diferentes grupos, independientemente de la cantidad de estudiantes que lo conformaban.

| | Medidas |
|--------------------------|-----------------------|
| Todos los UsuariosXGrupo | ● indiceAccesoUsuario |
| + 4301 | 180,00 |
| + 4302 | 105,13 |
| + 4303 | 218,32 |
| + 4304 | 111,53 |
| + Arrastre | 119,00 |

Slicer:

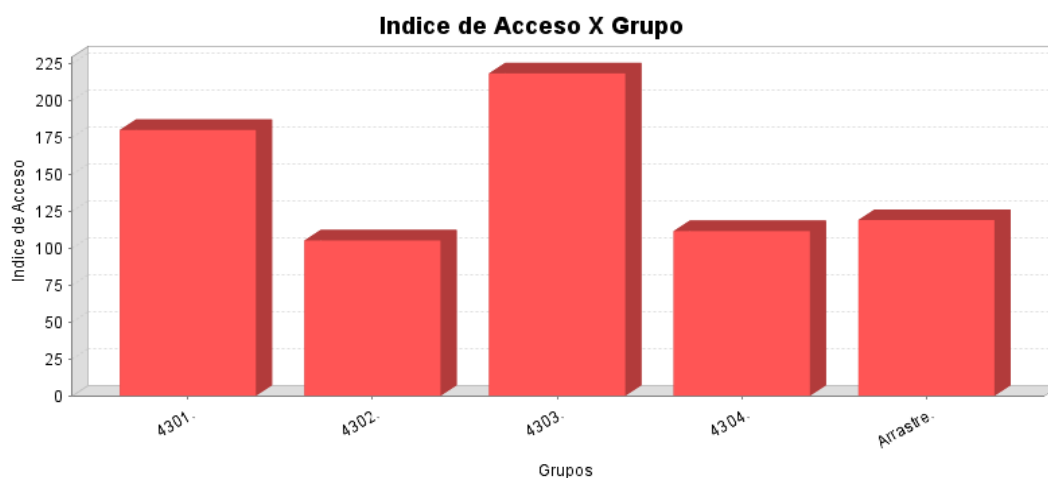


Figura 45 Índice de Acceso por Grupo

El colectivo de profesores, tras el análisis de estos resultados, se percató que los seis estudiantes del grupo de arrastre (correspondiente a los alumnos repetidores), en promedio, tuvieron un acceso comparable a los estudiantes de los otros grupos.

Asimismo en esta vista destacan los grupos 4301 y 4303 por mostrar un índice de acceso superior al resto de los grupos, circunstancia que llevó al colectivo de profesores a preguntarse por las diferencias que existían en estos dos grupos con respecto a los demás para que su participación en el entorno virtual fuera considerablemente superior.

La monitorización del acceso al entorno virtual fue otro indicador considerado por el colectivo de profesores como un elemento de alerta en cuanto a la realización del trabajo independiente por los estudiantes. Esta variable fue examinada de manera periódica por cada profesor de la asignatura y en general por el colectivo pedagógico. En la Figura 46 se muestran las vistas de acceso de cada uno de los grupos de estudiantes según la semana del curso. Otras vistas temporales fueron también examinadas pero por cuestiones de extensión del informe no serán ilustradas. Ver el apéndice 1.

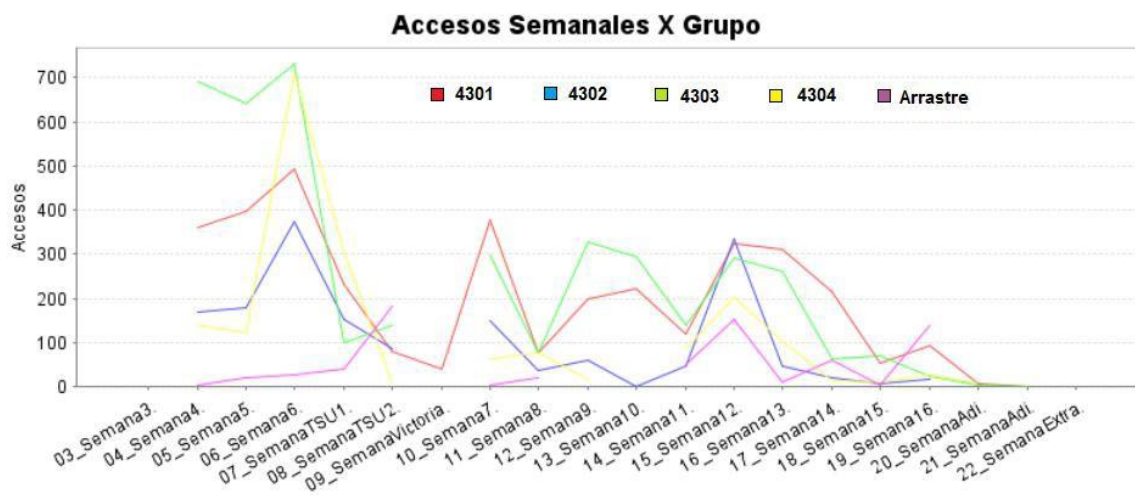


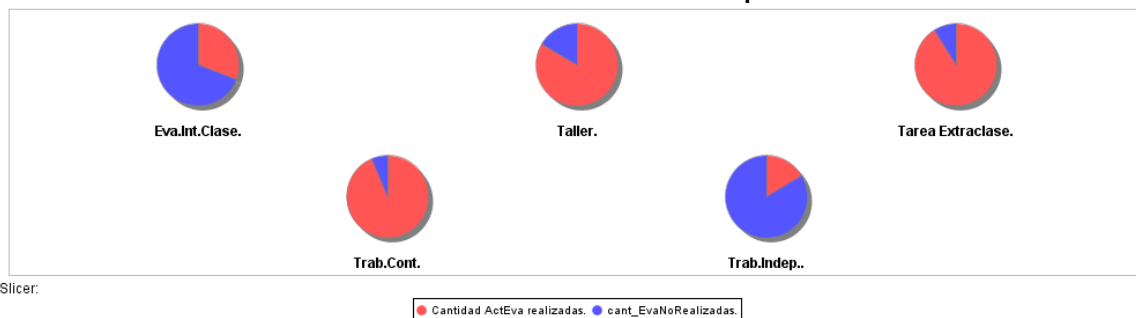
Figura 46 Accesos al campus por Semanas y Grupos

En el análisis de los datos referentes a las evaluaciones se corroboró uno de los principales problemas que el colectivo ya conocía desde cursos anteriores: las dificultades en el estudio habitual de los estudiantes y como consecuencia sus malos resultados en las evaluaciones frecuentes, incluso la no realización de las tareas orientadas por el profesor. En la Figura 47 se muestra en azul la cantidad de evaluaciones por tipo de evaluación que se dejaron de realizar. Nótese que la incidencia mayor recae en las evaluaciones en clase y el trabajo independiente.

| Medidas | | | |
|------------------------------|------------------------------|------------------------|-----|
| Todas las ActEvaluacionXTipo | ● Cantidad ActEva realizadas | ● cant_EvaNoRealizadas | |
| + | Eva.Int.Clase | 292 | 656 |
| + | Taller | 66 | 13 |
| + | Tarea Extraclase | 72 | 7 |
| + | Trab.Cont | 148 | 10 |
| + | Trab.Indep. | 142 | 727 |

Slicer:

Realización de las Evaluaciones X Tipo



Slicer:

Figura 47 Realización de las evaluaciones por Tipo

Esta información generó un conocimiento tal que se trazaron estrategias en pos de disminuir la cifra de evaluaciones no realizadas antes de que se terminara el curso.

Como era de esperar, los primeros indicios fueron identificados en el primer tema de la asignatura, por lo que el colectivo pedagógico se propuso que en los temas siguientes los resultados deberían mejorar y se aplicaron cambios.

En la Figura 48 se muestra como se comportó la realización de las evaluaciones por tema de la asignatura; aunque los porcentajes de actividades no realizadas bajaron, los resultados quedaron bastante lejos de lo esperado.

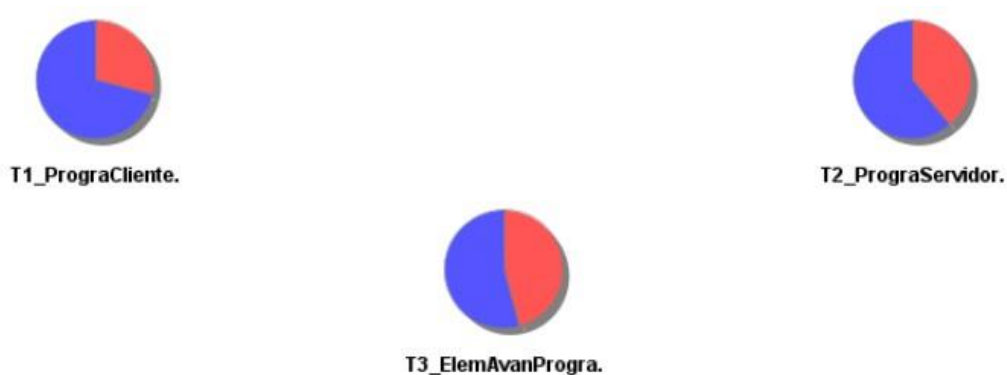


Figura 48 Realización de las tareas agrupadas por temas de asignatura

En el apéndice 1 se incluyen otros informes relativos a las evaluaciones.

7.3 Reportes predefinidos y a la medida

La solución desarrollada brindó a los usuarios del sistema alrededor de 20 diferentes tipos de reportes predefinidos, los cuales podían ser accedidos una vez que se autenticaran. Los reportes diseñados cubrieron las necesidades de información de los profesores así como de los agentes externos al colectivo. Los reportes pueden ser exportados a los formatos de archivo XLS, PDF, DOC, CSV y TXT.

En cursos anteriores, antes de utilizar la presente herramienta, el colectivo debía llenar un conjunto predefinido de informes: informes relacionados con los resultados en una determinada prueba parcial, informes de notas finales, informes de evaluaciones parciales, entre otros. Todos estos, ahora se pueden generar automáticamente por el sistema, lo que disminuye considerablemente los tiempos de respuestas del colectivo pedagógico a pedidos externos; un elemento que apoya el proceso de toma de decisiones si se tiene en cuenta que muchos de estos

informes son solicitados para ser analizados por actores externos al colectivo de profesores en aras de tomar acciones para mejorar los resultados obtenidos.

Además de los reportes predefinidos, los profesores pueden generar sus propios informes a la medida. Esta opción, aunque no fue muy explotada por los docentes, es también una vía alternativa para suplir algunas necesidades no cubiertas por los reportes predefinidos.

En el apéndice se incluye un apartado donde se pueden consultar ejemplos de los reportes predefinidos que brinda la solución.

7.4 Cuadros de mando integral

En aras de brindar a los profesores una visión más integral y particular sobre cada estudiante en los diferentes aspectos que fueron incorporados a la herramienta, estos tuvieron la oportunidad de acceder a un cuadro de mando integral que ofrece información sobre cada uno de los estudiantes que recibieron la asignatura.

Estos cuadros de mando integral, denominados “Vista General del Estudiante” brindaron información descriptiva sobre algunos indicadores y perspectivas importantes: dígame índice de acceso, visto desde la perspectiva semanal, hora del día y área de acceso; actividades realizadas así como las habilidades evaluadas y sus resultados. En toda la información mostrada siempre se refleja el indicador promedio para el grupo al que pertenece el estudiante y para todos los estudiantes de la asignatura; una forma de analizar a cada alumno desde una visión particular y grupal.

8 Conclusiones

- Se ha aplicado el método definido en el capítulo anterior en una asignatura impartida mediante el uso simultáneo de un entorno virtual de aprendizaje y sesiones presenciales con 79 estudiantes y 32 actividades, de las cuales 18 se impartieron en las mediante el entorno virtual de aprendizaje Moodle montado en la institución.
- Se han identificado 11 preguntas clave relativas a los aspectos de accesos al entorno virtual, evaluaciones realizadas y habilidades evaluadas las cuales se han descompuesto en 3 hechos 19 métricas y 15 dimensiones y se han definido el alcance en términos de resolución o nivel de granularidad para cada binomio hecho-dimensión.
- Se han establecido tres modelos conceptuales, uno para cada hecho que comparten las dimensiones siguiendo la estrategia de Kimball (1996)
- Se han seleccionado las fuentes de datos que mejor describían las necesidades detectadas y se han encontrado las correspondencias con el modelo conceptual.
- Se ha poblado y actualizado el almacén de datos mediante procesos ETL siguiendo una modelo de datos multidimensional utilizando la estructura de copo de nieve.
- Se crearon 20 reportes predefinidos, 10 reportes personalizados, 70 vistas de análisis y un cuadro de mando integral.

Capítulo 5 - Resultados

1 Introducción

En este apartado se resumen los resultados obtenidos tras la aplicación del método según cada fase los cuales están ampliamente detallados en el capítulo anterior referente a la aplicación de la metodología. Asimismo se realiza un análisis de los resultados globales del método desde el punto de vista de la eficiencia y eficacia en la creación de conocimiento.

2 Resultados de la fase de análisis de requerimientos

Tras realizar reuniones participadas por el gerente del conocimiento, los usuarios interesados y un miembro del equipo informático se esperaba obtener los siguientes resultados:

- Identificar preguntas clave
- Identificar indicadores (hechos y métricas) y dimensiones
- Identificar el nivel de granularidad
- Definir el modelo conceptual

Se ha conseguido obtener todos los resultados esperados. A continuación se detallan los resultados específicos de esta fase.

2.1 Resultados de la identificación de preguntas clave

Se han identificado 11 preguntas clave relativas a los aspectos de accesos al entorno virtual, evaluaciones realizadas y habilidades evaluadas.

Las preguntas identificadas fueron las siguientes:

- ¿Cuál es la cantidad de accesos realizados por un usuario?
- ¿Cuál es la cantidad de acciones realizadas por un usuario?
- ¿Cuál es la cantidad de módulos visitados por un usuario?
- ¿Cuál es la cantidad de cursos visitados por un usuario?
- ¿Desde qué lugares realiza los accesos un usuario?
- ¿Cuál es la cantidad de evaluaciones que debió realizar un usuario?
- ¿Cuáles son las evaluaciones realizadas a un usuario?
- ¿Cuál es el resultado de las evaluaciones de un usuario?
- ¿Cuál es el número de ocasiones en que un usuario debió realizar una actividad donde se evalúan sus habilidades?
- ¿Cuál es el número de ocasiones en que un usuario realizó una actividad donde se evalúan sus habilidades?
- ¿Cuál es el resultado de las evaluaciones de las habilidades de un usuario?

2.2 Resultados de la identificación de los indicadores y las dimensiones

Las preguntas identificadas en el apartado anterior, se han descompuesto en tres hechos, cinco métricas y nueve dimensiones representadas en la tabla inferior:

| Hechos detectados | Accesos Evaluaciones Evaluaciones de Habilidades |
|------------------------|--|
| Métricas detectadas | Frecuencia Resultado Resultado máximo. Resultado mínimo. Promedio de realizadas. |
| Dimensiones detectadas | Usuario Tiempo Acción Módulo Curso Lugar Actividad de evaluación Resultado Habilidad |

Tabla 50 Resumen de hechos, métrica y dimensiones identificadas

Cada métrica y dimensión aplica a uno o más hechos de forma que en la Tabla 33 del capítulo anterior se detalla dicha distribución.

2.3 Resultados de la identificación del nivel de granularidad o resolución

Se ha identificado el nivel de granularidad para cada binomio hecho-dimensión ajustado a las necesidades de conocimiento inferidas de las preguntas clave. La Tabla 34 del anterior capítulo recoge el detalle de los resultados de este apartado.

2.4 Resultados de la definición del modelo conceptual

Se han establecido tres modelos conceptuales, uno para cada hecho que comparten las dimensiones siguiendo la estrategia de Kimball (1996)

Estos modelos se han incluido en apartados anteriores, donde se han sintetizado en las siguientes figuras:

- Figura 30 Modelo conceptual para el hecho Acceso,
- Figura 31 Modelo conceptual para el hecho Evaluación y
- Figura 32 Modelo conceptual para el hecho Evaluación de habilidades

Las dimensiones compartidas para este método han sido:

- Dimensión usuario: compartida por todos los hechos.
- Dimensiones tipo resultado y actividad de evaluación se comparten sólo entre los hechos relativos a las evaluaciones, ya sean evaluaciones simples o evaluaciones de habilidades.

3 Resultados de la fase de selección de las fuentes de datos

En esta fase se esperaba obtener los siguientes resultados:

- Fuentes de datos seleccionadas
- Correspondencia identificada entre las fuentes de datos y el modelo conceptual.
- Métricas conformadas
- Modelo conceptual ampliado

A continuación se detallan los resultados obtenidos en cada uno de los anteriores

3.1 Resultados de las fuentes de datos seleccionadas

Tras analizar las necesidades de información se han seleccionado las siguientes fuentes de datos:

- Base de datos del Entorno Virtual de Enseñanza y Aprendizaje de FUNIBER que emplea el sistema Moodle. En concreto se han escogido las siguientes tablas que pueden ser consultadas en el apartado 4 del anterior capítulo:
 - Tabla 35 Tabla mdl_log
 - Tabla 36 Tabla mdl_user
 - Tabla 37 Tabla mdl_role
 - Tabla 38 Tabla mdl_role_assignments
 - Tabla 39 Tabla mdl_course
 - Tabla 40 Tabla mdl_course_modules
 - Tabla 41 Tabla mdl_groups
 - Tabla 42 Tabla mdl_groups_members
 - Tabla 43 Tabla mdl_assign
 - Tabla 44 Tabla mdl_assign_grades
 - Tabla 45 Tabla mdl_grade_outcomes
 - Tabla 46 Tabla mdl_grade_items
 - Tabla 47 Tabla mdl_grade_grades
- El archivo datosIP.xml que relaciona las diferentes subredes de la institución con las zonas a las que corresponden.

- El archivo datosEvaluacion.xml que relaciona las actividades de evaluación con el tema del curso al que pertenecen y el tipo de actividad (trabajo de control, trabajo independiente, trabajo en clase, etc.).
- Por último, el archivo datosHabilidades.xml que relaciona las habilidades con el tema del curso asociado.

3.2 Resultados de la correspondencia entre las fuentes de datos y el modelo conceptual.

En cada una de las fuentes de datos se ha encontrado la correspondencia entre éstas y el modelo conceptual, esto ha quedado detallado en la Tabla 49

Correspondencias entre las fuentes de datos y el modelo conceptual.

3.3 Resultados de las métricas que se han conformado

Finalmente se han conformado 19 métricas que aplican a distintos hechos las cuales han sido las siguientes:

-
- Cantidad de accesos
 - Cantidad de usuarios
 - Cantidad de módulos
 - Cantidad de acciones
 - Cantidad de cursos
 - Cantidad de lugares
 - Índice de acceso por usuario
 - Cantidad de actividades de evaluación realizadas
 - Cantidad de actividades de evaluación que debió realizar
 - Resultado
 - Resultado máximo
 - Resultado Mínimo
 - Cantidad de Evaluaciones no realizadas
 - Porcentaje de realización de evaluaciones
 - Cantidad de habilidades realizadas
 - Cantidad de habilidades que debió realizar
 - Cantidad de Actividades de evaluación con una habilidad dada
 - Cantidad de Evaluaciones de habilidades no realizadas
 - Porcentaje de habilidades evaluadas

La Tabla 48 Identificación de las métricas en la aplicación del modelo conceptual amplía los resultados obtenidos respecto a las métricas finales que se han conformado, detallando el hecho al que aplica, una breve descripción y la fórmula de calcularlas.

3.4 Resultados del modelo conceptual ampliado

El modelo conceptual ampliado es el resultado de la combinación del modelo conceptual y la conformación de las métricas descritos anteriormente en este capítulo.

4 Resultados de la fase de diseño del almacén de datos

En esta fase se esperaba obtener los siguientes resultados:

- Modelo lógico multidimensional del almacén de datos
- Tablas de hechos
- Tablas de dimensiones
- Uniones entre tablas

A continuación se describen únicamente los resultados del modelo lógico multidimensional del almacén de datos. Esto ya es suficiente para comprender los resultados de esta fase, pues este modelo describe la estructura real de la base de datos.

4.1 Resultados del diseño del modelo lógico multidimensional del almacén de datos, de las tablas de hechos, de las tablas de dimensiones y de sus uniones

Se diseñaron tres estructuras multidimensionales, una para cada hecho (acceso, evaluación y habilidad) que se quería medir, las cuales adoptaron el esquema de copo de nieve, debido principalmente a las relaciones existentes con otras tablas que consolidaban más eficientemente la información según un modelo relacional normalizado.

Como resultado se obtuvieron 18 tablas, 3 hechos y 15 dimensiones que pueden ser consultadas en las figuras siguientes:

- Figura 33 Modelo lógico del hecho Acceso
- Figura 34 Modelo lógico del hecho Evaluación
- Figura 35 Modelo lógico del hecho Evaluación de habilidades

En las figuras anteriores se muestran las tablas de hechos (centrales), las tablas de dimensiones (laterales) y las uniones entre ellas.

5 Resultados de la fase de integración de datos

Como resultado de esta fase se dispone del almacén de datos poblado y actualizado según el modelo lógico resultante de la fase anterior. En él se han incluido los datos útiles estandarizados y transformados según los criterios de calidad expresados en la Tabla 17 Calidad en la información.

La estructura de este almacén de datos es multidimensional al estar los datos organizados en tablas de hechos y dimensiones de manera que los elementos medibles (denominados hechos) puedan ser cuantificados utilizando las diferentes dimensiones.

6 Resultados de la fase de uso de la solución

En esta fase se engloban los resultados de las tareas relativas a la configuración y uso de las herramientas de análisis y las distintas funcionalidades que se han explotado para realizar esta tesis. La tabla inferior recoge los resultados detallados en el capítulo anterior y en el apéndice.

| Tarea | Resultados |
|--|---|
| Conformación de los cubos multidimensionales | Se conformaron 3 cubos, uno para cada <i>data mart</i> . |
| Análisis OLAP | Se desarrollaron 70 vistas de análisis con diferentes fines. Se crearon 7 usuarios, uno para cada profesor, uno para el jefe de disciplina y otro para el jefe de departamento. Se configuraron dos roles, profesor y jefe, cada uno de ellos con distintos permisos. |
| Reportes | Se crearon 20 diferentes tipos de reportes predefinidos Los profesores generaron un total de 10 reportes a la medida |
| Cuadro de mando integral | Se creó un cuadro de mando integral donde se recopilaba la información, relativa al usuario, siguiente: Índice de acceso, visto desde la perspectiva semanal, hora del día y área de acceso; Actividades realizadas Habilidades evaluadas y sus resultados |
| Formación de los usuarios en las herramientas OLAP | Se realizó una formación en el uso de las herramientas OLAP para realizar consultas a dicho sistema y crear nuevos reportes. |

Tabla 51 Resumen de los resultados de la fase de uso de la solución

Entrando más al detalle los reportes y análisis más importantes se pueden consultar en el capítulo anterior y el apéndice así como el cuadro de mando creado. Respecto a los análisis, la tabla inferior resume aquellos más importantes,

mostrados en el capítulo anterior y el apéndice y los relaciona con el conocimiento que se crea en cada uno de ellos.

| <i>Data mart</i> | Análisis | Conocimiento creado |
|---------------------------|--|--|
| Accesos | Índice de acceso por grupo | ¿Cuál ha sido el número de accesos totales para cada grupo? |
| | Accesos semanales por Grupos | ¿Cuál ha sido el número de accesos semanales para cada grupo? |
| | Accesos por día de la semana | ¿Cómo se comportaron los accesos teniendo en cuenta los días de la semana? |
| | Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores) | ¿Cómo se comportaron los accesos teniendo en cuenta los días de la semana y el grupo docente? |
| | Accesos por hora del día | ¿Cómo se comportaron los accesos teniendo en cuenta las horas del día? |
| | Cantidad de estudiantes que accedieron por módulo del sistema. | ¿Cuántos estudiantes visitaron los diferentes módulos? |
| | Porcentaje de Accesos por miembro del grupo. | ¿Cuál es el porcentaje de acceso de cada miembro relativo a los accesos totales del grupo? |
| | Cantidad de días que accedió un usuario en todo el semestre | ¿Cuál ha sido la cantidad de días que ha accedido un usuario? |
| | Mediana de accesos por día | ¿Qué mediana de accesos tiene un usuario? |
| | Total de accesos | ¿Cuál ha sido el total de accesos de un usuario? |
| Actividades de evaluación | Promedio de acceso por día, desviación estándar y coeficiente de variación. | ¿Qué medidas de tendencia central (promedio) y de dispersión (desviación estándar y coeficiente de variación) tienen los accesos de un estudiante? |
| | Realización de las evaluaciones por Tipo | ¿Cuál ha sido el grado de realización de las actividades según el tipo de las mismas? |
| | Realización de las tareas por tema de asignatura | ¿Cómo se comportó la realización de las evaluaciones por tema de la asignatura? |

| <i>Data mart</i> | Análisis | Conocimiento creado |
|------------------|---|--|
| | Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación | ¿Qué cantidad de estudiantes se evaluaron y no se evaluaron en cada actividad de evaluación y cuál ha sido el porcentaje de estudiantes evaluados? |
| | Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo | ¿Qué cantidad de estudiantes según sus grupos se evaluaron en cada actividad de evaluación realizada y la cantidad que no lo hicieron? |
| Habilidades | Resultado promedio de cada habilidad | ¿Cuál ha sido el resultado promedio de la evaluación de las habilidades? |
| | Porcentaje de evaluación de cada habilidad | ¿Cuál ha sido el porcentaje de habilidades evaluadas? |

Tabla 52 Análisis más importantes realizados y conocimiento creado

Las herramientas OLAP que han utilizado los usuarios finales han sido las de análisis, reporte y cuadro de mando. Según el conocimiento requerido que se ha detectado en la fase de análisis en forma de preguntas, se presenta a continuación una tabla que resume las herramientas usadas para responder cada una de las preguntas.

| Conocimiento requerido | Análisis | Reportes | Cuadro de mando |
|---|-----------------|-----------------|------------------------|
| ¿Cuál es la cantidad de accesos realizados por un usuario? | x | | x |
| ¿Cuál es la cantidad de acciones realizadas por un usuario? | x | | |
| ¿Cuál es la cantidad de módulos visitados por un usuario? | x | | x |
| ¿Cuál es la cantidad de cursos visitados por un usuario? | x | | |
| ¿Desde qué lugares realiza los accesos un usuario? | | | x |

| Conocimiento requerido | Análisis | Reportes | Cuadro de mando |
|--|----------|----------|-----------------|
| ¿Cuál es la cantidad de evaluaciones que debió realizar un usuario? | X | | X |
| ¿Cuáles son las evaluaciones realizadas a un usuario? | X | | |
| ¿Cuál es el resultado de las evaluaciones de un usuario? | | X | X |
| ¿Cuál es el número de ocasiones en que un usuario debió realizar una actividad donde se evalúan sus habilidades? | X | | X |
| ¿Cuál es el número de ocasiones en que un usuario realizó una actividad donde se evalúan sus habilidades? | X | | X |
| ¿Cuál es el resultado de las evaluaciones de las habilidades de un usuario? | | X | X |

Tabla 53 Herramientas OLAP utilizadas según el conocimiento requerido detectado.

7 Solución a los problemas detectados

La tabla inferior recoge los problemas que pretende resolver esta tesis y los resultados que se han obtenido respecto a su resolución.

| Problemas detectados | Resultados de la resolución de los problemas detectados. | Mejoras en eficiencia o eficacia |
|---|--|----------------------------------|
| Posibilidad de cometer errores al introducir información proveniente de distintos sistemas. | Se reducen los errores al centralizar la información dentro del almacén de datos | Mejora de eficacia. |
| Inversión de gran cantidad de tiempo para la introducción de la información. | Se minimiza el tiempo invertido al no ser necesario introducir la información cada vez que se requiere un informe o análisis | Mejora de eficiencia. |

| Problemas detectados | Resultados de la resolución de los problemas detectados. | Mejoras en eficiencia o eficacia |
|---|--|---|
| Repetición del proceso para cada nuevo análisis | No se repite el proceso en cada análisis al estar toda la información centralizada en un único lugar y los informes preparados de antemano. | Mejora de eficiencia. |
| Poca eficiencia en obtener detalles. | Se aumenta la eficiencia en obtener los detalles al estar estructurada la información de forma multidimensional pudiendo entrar en detalle mediante las utilidades de las herramientas OLAP | Mejora de eficiencia |
| Problemas de rendimiento en el entorno transaccional | Se minimizan los problemas derivados del rendimiento de los sistemas transaccionales al realizar las consultas a estos sistemas una vez al día, en vez de una consulta para cada análisis, y en un horario de mínimo uso de dichos sistemas | Mejora de eficiencia y eficacia |
| Se duplican los entornos de información de forma que cada uno de los usuarios de información querrá los informes con un diseño determinado y pueden aparecer “entornos paralelos” de información. | Al estar centralizada la información en un único lugar se evita usar datos de entornos paralelos que pudieran no estar actualizados o no ser correctos. Asimismo se homogeniza el diseño de los informes al configurar plantillas transversales a toda la institución. | Mejora de eficacia. |
| Baja calidad de los datos con el riesgo de tener distintas versiones de una misma realidad: | Se minimiza este problema al aplicar los procesos ETL donde se estandarizan y homogenizan los datos en pos de disponer de una única realidad. | Mejora de eficacia. |
| Desconocimiento por parte de los usuarios finales sobre dónde reside la información que necesitan. | Los usuarios al disponer de un único sistema de información centralizado ya no sufren este tipo de problemas. | Mejora de eficacia. |

Tabla 54 Resultados de la resolución de los problemas detectados y su impacto en la eficiencia y eficacia globales.

8 Cumplimiento de los objetivos

| Objetivos planteados | Resultados de cumplimiento de los objetivos |
|---|---|
| <p>Caracterizar las especificidades de las organizaciones de formación a distancia dentro del marco de un proyecto de <i>Business Intelligence</i>.</p> | <p>Se han caracterizado los siguientes aspectos de las organizaciones formativas respecto a su influencia en un sistema de <i>Business Intelligence</i>:</p> <ul style="list-style-type: none"> • Dimensiones y resoluciones propuestas para el análisis de los datos de un entorno de aprendizaje virtual. • Propuesta de resoluciones para la dimensión actividad según el uso que se hace del recurso en un entorno académico • Propuesta de resoluciones para la dimensión actividad según el nivel de accesibilidad que se tiene sobre el recurso académico • Beneficios del método para los roles más significativos de un entorno virtual de aprendizaje • Ejemplos de tipología de las actividades de evaluación • Propuesta de conocimiento requerido para profesores • Ejemplo de caracterización del conocimiento requerido por los profesores en hechos, métricas, dimensiones y resoluciones • Propuesta de modelo conceptual y lógico para almacenar los datos útiles para crear el conocimiento requerido por los profesores • Análisis y elección de fuentes de datos provenientes de un entorno virtual de aprendizaje sobre la plataforma Moodle • Análisis de las correspondencias entre las fuentes de datos del entorno virtual de aprendizaje y el modelo conceptual • Ejemplos de transformaciones y trabajos ETL aplicadas en procesos de integración de datos de un entorno virtual de aprendizaje • Ejemplos de análisis, reportes y cuadros de mando útiles para responder al conocimiento requerido por los profesores. • Ejemplos de acciones que se tomaron a partir del uso de las herramientas OLAP. |

| Objetivos planteados | Resultados de cumplimiento de los objetivos |
|--|--|
| Adaptar las metodologías y herramientas del BI para generar un método que ayude a crear conocimiento en entornos virtuales de aprendizaje. | <ul style="list-style-type: none"> • Se ha propuesto una metodología de BI personalizada para aplicarla en organizaciones que empleen entornos virtuales de aprendizaje • Se ha aplicado la metodología propuesta en una organización que emplea el entorno virtual de aprendizaje Moodle. |
| Aumentar la disponibilidad de los datos para poder tomar decisiones según un conocimiento generado a través de su explotación. | <ul style="list-style-type: none"> • Se ha aumentado el uso de informes antiguos y se han creado nuevos informes de forma autónoma por parte de los usuarios finales. |
| Extraer, limpiar y almacenar los datos procedentes de múltiples fuentes en un único repositorio. | <ul style="list-style-type: none"> • Se han realizado los procesos ETL sobre la base de datos del entorno virtual de aprendizaje Moodle y se han almacenado en un único repositorio conformando el almacén de datos de la institución formativa a partir de tres <i>Data marts</i> |
| Organizar los datos dentro de un modelo multidimensional según criterios de utilidad. | <ul style="list-style-type: none"> • Se ha diseñado el modelo multidimensional según criterios de utilidad detectados en la fase de análisis de necesidades y posteriormente se ha poblado con datos provenientes del entorno virtual de aprendizaje Moodle |

Tabla 55 Resultado del cumplimiento de cada objetivo planteado en esta tesis

9 Eficiencia y eficacia del método

En este apartado se revisan los resultados desde la perspectiva de la eficiencia y eficacia relativa al estado previo de implementación del método. Por este motivo en primer lugar se realiza una comparación entre el estado inicial y el final para poder contrastar los resultados para posteriormente medir la eficiencia y eficacia de la metodología según la variación en los indicadores utilizados.

9.1 Comparación entre el estado previo y el estado final

En este apartado se resume según los aspectos detectados en el apartado 8 del capítulo 1 (Entorno de Aplicación de la Propuesta) relativos al rendimiento de los servidores, la generación de informes y la creación de conocimiento.

La tabla inferior resume los resultados obtenidos tras la aplicación de la metodología al compararlos con la situación inicial.

| | Situación inicial | Situación final |
|------------------------------------|---|---|
| Rendimiento de los servidores OLTP | <p>La carga de los servidores OLTP se hacía muy alta ante cualquier consulta medianamente compleja:</p> <ul style="list-style-type: none"> • Las consultas se hacían en horarios aleatorios, según necesidades del tomador de decisiones, independientemente de la carga de los servidores. • Algunas consultas no se podían realizar inmediatamente debido a la carga de los sistemas OLTP. • Existía una correlación directa entre las alertas de alto riesgo (<i>warning</i>) de los servidores y el uso de consultas complejas sobre estos servidores. • Ante cualquier consulta se bloqueaban tablas de la base de datos de los servidores OLTP que impedían la escritura en las mismas, ralentizando el entorno virtual de aprendizaje hasta que se liberaban dichas tablas. Si la consulta era compleja el tiempo de bloqueo se hacía grande de forma que la velocidad de navegación se hacía lenta. • Cada informe o análisis requería de una consulta a estos servidores OLTP | <p>La carga de los servidores OLTP no queda afectada por la consulta de informes:</p> <ul style="list-style-type: none"> • Se han concentrado todas las consultas de los procesos ETL a servidores OLTP en un horario de mínima carga • Cada informe o análisis no genera una nueva consulta a los servidores OLTP ya que se realiza sobre el servidor que almacena el DW • No se dan casos de alertas de alto riesgo asociadas a las consultas. • Se pueden realizar consultas a cualquier hora del día independientemente de la carga de los servidores OLTP. |

| | Situación inicial | Situación final |
|---|---|--|
| Generación de informes (reportes, análisis, cuadros de mando) | <ul style="list-style-type: none"> • Debido a las limitaciones de rendimiento de los servidores OLTP no se permitía realizar consultas masivamente o bien consultas muy complejas. • Se requería hacer partícipe al equipo informático ante cualquier nuevo informe aumentando el tiempo de recursos humanos para cada nuevo informe | <ul style="list-style-type: none"> • Se permiten realizar consultas masivas y muy complejas. • Los usuarios finales pueden crear nuevos informes de forma autónoma. |
| Creación de conocimiento | <ul style="list-style-type: none"> • Creación de conocimiento con poca profundidad debido a los factores limitantes asociados a realizar consultas complejas en sistemas OLTP • Creación de conocimiento dependiente de otros usuarios Pocos mensajes proactivos de los profesores a los estudiantes por no disponer del conocimiento necesario sobre su estado • Acciones correctivas al finalizar la acción formativa. | <ul style="list-style-type: none"> • Creación de conocimiento con alto nivel de detalle • Creación de conocimiento autónomo • Aumento del número de mensajes proactivos de los profesores hacia los estudiantes. • Acciones correctivas durante la ejecución de la acción formativa. |

Tabla 56 Comparación entre la situación inicial y final tras haber aplicado el método.

9.2 Mejoras en la eficiencia y eficacia respecto a la situación inicial.

En este apartado se detallan las mejoras detectadas en eficiencia y eficacia y se dan indicadores que ayudan a validar tales mejoras.

La tabla inferior recoge las mejoras en eficiencia y eficacia del método observadas respecto a la situación inicial.

| | Mejoras en la eficiencia y eficacia | Indicadores |
|---|---|---|
| Rendimiento de los servidores OLTP | <p>Eficiencia:</p> <ul style="list-style-type: none"> • La carga de los servidores OLTP para cada consulta se ha minimizado haciendo que su rendimiento sea mejor. <p>Eficacia:</p> <ul style="list-style-type: none"> • Se obtienen resultados, independientemente del estado de carga de los servidores OLTP | Número de alertas de rendimiento críticas y nivel de carga de los sistemas OLTP durante el uso de las herramientas OLAP. |
| Generación y uso de informes (reportes, análisis, cuadros de mando) | <p>Eficiencia:</p> <ul style="list-style-type: none"> • El tiempo dedicado para consultar un informe predefinido o crear nuevos informes es menor. <p>Eficacia:</p> <ul style="list-style-type: none"> • Aumento del grado de detalle de los informes al poder realizar consultas complejas. • Personalización de informes para dar respuesta a nuevos conocimientos requeridos. | <p>Tiempo invertido en:</p> <ul style="list-style-type: none"> • Crear nuevos informes • Consultar informes predefinidos <p>Número de nuevos informes generados</p> |
| Creación de conocimiento | <p>Eficiencia:</p> <ul style="list-style-type: none"> • El tiempo para crear conocimiento es menor mientras que el rendimiento de los servidores es mayor. <p>Eficacia:</p> <ul style="list-style-type: none"> • Se crea conocimiento más amplio al haber aumentado el nivel de detalle. • Se crea conocimiento más específico al poder personalizar de forma autónoma los informes. | <ul style="list-style-type: none"> • Uso de informes. • Número de mensajes proactivos que se envían a los alumnos • Número de acciones tomadas durante la impartición de la acción formativa |

Tabla 57 Mejoras en la eficiencia y eficacia respecto a la situación inicial.

Para validar las mejoras indicadas en la tabla superior se ha realizado una comparación de los indicadores sobre dos ediciones del mismo curso que

comparten los mismos profesores pero distintos alumnos, algo que resulta insuficiente para asegurar con firmeza que los resultados son definitivos. Asimismo para el cálculo de los indicadores se ha evaluado la fase de uso de la solución independientemente del esfuerzo dedicado a la fase de análisis de necesidades y diseño del mismo. Por este motivo los resultados presentados en la siguiente tabla deben considerarse como orientativos, indicando una tendencia que debe ser validada con materiales y métodos más exhaustivos y amplios.

A pesar de lo indicado arriba, los resultados de la tabla inferior dan una idea de la utilidad del método empleado, sobre todo en aquellos indicadores que han experimentado una mejora significativa de eficacia y eficiencia durante la fase de uso de la solución.

| Indicadores | Semestre anterior | Semestre actual | Variación |
|--|--|--|--|
| Número de alertas de rendimiento críticas, durante el uso de las herramientas OLAP, del servidor que aloja el entorno virtual de aprendizaje | 7 alertas durante el semestre debidas a consultas complejas en horario de alto uso del entorno virtual de aprendizaje. | Ninguna alerta para consultas complejas en horario de alto uso del entorno virtual de aprendizaje. | Ausencia de alertas |
| Nivel de carga de los sistemas OLTP | Hasta un 70% | Hasta un 10% | Un mínimo de un 86% menos de carga |
| Tiempo para crear nuevos informes sin considerar el tiempo de reuniones ni de diseño y carga del DW | de 3 a 5 horas | Un máximo de una hora | Del 67% al 80% menos de tiempo invertido |
| Tiempo para consultar informes predefinidos | de 20 a 60 minutos | Un máximo de 10 minutos | Del 50% al 83% menos de tiempo invertido |

| Indicadores | Semestre anterior | Semestre actual | Variación |
|--|-------------------------------------|--|---|
| Nuevos informes creados | 3 | 10 | Aumento del 230% en la creación de informes nuevos |
| Uso de informes predefinidos para todos los profesores durante el semestre. | 18 (un informe al mes por profesor) | 72 (un informe a la semana por profesor) | Aumento del 300% en el uso de informes predefinidos |
| Número de mensajes proactivos que se envían a los alumnos por parte de todos los profesores durante el semestre. Se asume que todos los mensajes proactivos se originan a partir del uso de este método. | 13 | 25 | Aumento del 90% de mensajes proactivos enviados |
| Número de acciones tomadas durante la impartición de la acción formativa para todos los profesores durante el semestre. | 3 | 5 | Aumento del 70% de acciones tomadas. |

Tabla 58 Resultados del análisis de indicadores de eficiencia y eficacia

Según la tabla superior y atendiendo a las limitaciones de validez de estos resultados se podría inferir que los valores relativos a la carga de los servidores, del número de informes creados y de su uso podrían indicar que realmente existen ventajas de eficiencia del método. Asimismo si se considera que al usar el sistema se está generando conocimiento, también se puede considerar que el método es eficaz.

Respecto a los demás indicadores, aquellos relativos al tiempo invertido por parte de los profesores, aun cuando dan valores positivos y esperanzadores, al

no haberse contemplado los tiempos de las reuniones y preparación del almacén de datos no se puede concluir con firmeza que supongan una ventaja de eficiencia. No obstante al considerar que anteriormente para cada informe nuevo se tenían que realizar reuniones para el análisis de necesidades y que ahora con una única vez que se realicen aplicaría a muchos informes distintos, la eficiencia vendría dada por la amortización del tiempo invertido en la fase inicial según la cantidad de informes que se pueden obtener de los resultados de dicha fase. En otras palabras con una única vez que se capten las necesidades, se tendría acceso a muchos tipos de informes que podrían ser personalizados por los usuarios finales de forma autónoma amortizando el tiempo invertido a medida que se van personalizando nuevos informes de forma autónoma.

Respecto a los demás indicadores tales como el número de mensajes proactivos y el número de acciones tomadas, aunque han crecido, no se puede considerar algo significativo debido alcance de la muestra, donde sólo se han evaluado dos semestres.

10 Cumplimiento de las hipótesis

La tabla inferior recoge los resultados asociados a las hipótesis planteadas en esta tesis. En este sentido para la interpretación de los mismos se tiene que atender a los aspectos relativos a la validación de los resultados expuestos en el apartado anterior, de forma que dichos resultados deben considerarse como orientativos y no definitivos.

| Hipótesis | Validación |
|---|--|
| Disponer de un sistema de BI hace que se exploten más los datos del entorno virtual de aprendizaje. | Se ha constatado un incremento del número de informes nuevos creados y un aumento del uso de éstos y de los informes antiguos implicando un mayor consumo de los datos del entorno virtual de aprendizaje. |
| Los datos explotados al ser sintetizados en términos de negocio o utilidad se convierten en información que se transfiere eficientemente a los responsables de tomar decisiones. | Se ha analizado la relación entre el modelo multidimensional que incluye los datos contextualizados, corregidos, calculados, categorizados y condensados con el concepto de información de Davenport y Prusak (1998) y se ha constatado una disminución del tiempo en el uso de los informes y la creación de nuevos minimizando la carga de los servidores y eliminando las alertas de rendimiento de los mismos. |
| La información, al ser consumida por los responsables de tomar decisiones, se convierte en conocimiento de forma eficaz al estar sintetizada y orientada en términos de utilidad o de negocio. | Se ha considerado que la información almacenada en el almacén de datos, al haber sido seleccionada de antemano en la fase de análisis de necesidades, es de utilidad de forma que al ser consumida se responde a dicha utilidad y se convierte en conocimiento. Asimismo se ha constatado un aumento en el uso de los informes implicando creación de nuevo conocimiento y constatación del conocimiento tácito existente. |
| El conocimiento adquirido, resultante de la organización de los datos en términos de utilidad, mejora el proceso de toma de decisiones. | Se ha considerado que la información de utilidad es la antesala de la toma de decisiones y por lo tanto al haber generado este tipo de información se ha impactado positivamente en las decisiones tomadas. Asimismo se ha constatado un aumento de decisiones tomadas a partir del uso del método presentado en esta tesis. |
| La creación del conocimiento es eficaz en términos de disponer de una única verdad de los hechos para toda la institución, de evitar errores derivados del empleo de múltiples fuentes de datos y de almacenar los datos en términos de utilidad o negocio. | Se ha minimizado el riesgo de disponer de distintas versiones de una misma realidad al haber aplicado los procesos ETL donde se estandarizan y homogenizan los datos para toda la organización en un único repositorio de datos que actúa como fuente de todos los análisis y reportes que se realicen en la institución. |

| Hipòtesis | Validación |
|--|---|
| <p>La creación del conocimiento es eficiente en relación a otros métodos que se puedan emplear en términos de autonomía de los responsables de tomar decisiones, disminución de tiempos para crear nuevo conocimiento, mejora del rendimiento de los sistemas informáticos OLTP y, captación de datos automatizado, entre otros.</p> | <p>Se ha aumentado la autonomía de los usuarios finales en la creación y consulta de informes hecho que ha aumentado el número de consultas e informes nuevos y se ha reducido la carga de los servidores y el tiempo empleado para consultar los informes existentes y crear nuevos informes.</p> |
| <p>El conocimiento se difunde/comparte de forma más eficiente al estar la información centralizada en un único sistema informático accesible desde cualquier ordenador y al dar autonomía a los tomadores de decisiones para aplicar estrategias comunicativas automatizadas.</p> | <p>Se ha centralizado la información en el almacén de datos que se ha preparado para realizar consultas y se ha dado autonomía a los tomadores de decisiones hecho que ha simplificado el proceso de creación de nuevos informes y realización de consultas. Asimismo se ha registrado una disminución del tiempo para crear nuevos informes y ha disminuido la carga de los servidores OLTP.</p> |

Tabla 59 Resultado en el cumplimiento de las hipótesis

Capítulo 6 - Conclusiones

Este capítulo recoge las principales conclusiones a las que se ha llegado tras aplicar el método propuesto. En concreto se han dividido en tres grupos de conclusiones: utilidad del método, factores de eficiencia y eficacia y organización de los datos.

1 Utilidad del método

- Aplicar métodos y herramientas del BI en entornos virtuales de aprendizaje hace que se exploten más los datos de dichos entornos convirtiéndolos eficiente y eficazmente en información de utilidad para la toma de decisiones.
- Los usuarios finales del sistema utilizan más la información cuando son involucrados en el proceso de análisis de necesidades y disponen de autonomía en la creación de informes.
- Los datos explotados del entorno virtual de aprendizaje, al ser sintetizados en términos de negocio o utilidad durante el análisis de necesidades, se convierten en información que al ser consumida aporta nuevo conocimiento o transforma el conocimiento tácito preexistente en explícito.

- El conocimiento creado mediante métodos y herramientas BI es de alto nivel de detalle y al ser explotada de forma autónoma se consigue aumentar el número de mensajes proactivos personalizados y realizar más acciones correctivas durante la ejecución de la acción formativa.
- Los análisis que realizan los profesores sobre los datos del entorno virtual de enseñanza les permite conocer el estado de las asignaturas según la actividad que se registra en ellas, el uso que se da a los recursos de la asignatura, los aportes de alumnos y prever amenazas entre otras utilidades.
- La utilidad de los análisis para los tutores se da al poder conocer el estado actual y pasado de los alumnos en un curso según su actividad, prever tendencias y disponer de información ampliada sobre la actividad de los alumnos, entre otras.
- Conocer el estado actual y pasado de los cursos, disponer de información sobre el estado de los profesores según su actividad y carga de trabajo, prever tendencias en el estado de cursos y de profesores y descubrir cargas de trabajo según previsiones de altas y bajas son algunas de las utilidades de los análisis OLAP para los coordinadores académicos.
- Para los administradores y técnicos de los entornos virtuales de enseñanza el sistema les permite conocer el estado del personal técnico del campus según su actividad, conocer el estado de los servicios informáticos (errores, caídas...), disponer información técnica de los

equipos de los usuarios y detectar errores según curso o usuario, entre otras.

2 Factores de eficiencia y eficacia

- La creación del conocimiento es eficaz en términos de disponer de una única verdad de los hechos para toda la institución, de evitar errores derivados del empleo de múltiples fuentes de datos y de almacenar los datos previamente filtrados y corregidos según criterios de utilidad o negocio.
- La creación del conocimiento es eficiente en términos de rendimiento de los sistemas informáticos y tiempo invertido, experimentando mejoras sustanciales respecto a otros métodos que explotan los datos directamente desde los OLTP.
- Preparar las reuniones y seleccionar los actores que mejor conozcan las necesidades de conocimiento ayuda a reducir el tiempo invertido en el análisis de necesidades y mejora la eficacia de esta fase al identificar las preguntas correctas según criterios de utilidad.
- Identificar las fuentes de datos que maximicen el rendimiento de los sistemas de origen y minimicen el tiempo de extracción, limpieza y transformación de los datos de origen, seleccionar el personal informático más preparado, identificar indicadores de fácil obtención y cálculo y escoger el nivel de granularidad mayor posible según las

necesidades actuales y previstas son factores que inciden en la eficiencia de la fase de selección de las fuentes de datos

- Elegir el mejor esquema de datos según las necesidades detectadas y escoger el personal informático formado en las tareas de esta fase inciden directamente en la eficiencia en la fase de diseño del almacén de datos.
- Elegir el horario de actualización de la base de datos, la periodicidad de actualización de los datos, el personal informático más preparado y diseñar correctamente los procesos ETL inciden directamente en la eficiencia del proceso de integración de los datos.
- Simplificar los resultados incluidos en los informes y formar a los usuarios en las herramientas OLAP son factores que inciden en la eficiencia del uso de la solución.

3 Organización de los datos

- Hay riesgos de ineficacia en la generación de conocimiento cuando se emplea la estrategia *Top-down* de Inmon debido a que inicialmente se recopilan datos según una estimación inicial de información potencialmente útil en un futuro que podría ser incompleta. En cambio al emplear la estrategia *Bottom-up* se minimizan estos riesgos al modular y escalable. Además la inversión inicial es mucho menor cuando se emplea el segundo método, por estos motivos se ha empleado la estrategia propuesta por Kimball.

-
- El almacén de datos multidimensional debería ser llamado almacén de información debido a las características de los datos almacenados los cuales se han contextualizado y categorizado en una estructura multidimensional de hechos únicos y dimensiones compartidas.
 - El diseño de las dimensiones compartidas implica un trabajo cooperativo entre todos los actores de la institución, incluyendo las instancias directivas en pos de definir un marco estable, único y transversal sobre el que clasificar los datos y poder construir un conocimiento válido en toda la institución, independientemente del departamento que lo utilice.

Capítulo 7 - Propuestas de continuación de la tesis

En este capítulo se recogen las recomendaciones para ampliar lo presentado en esta tesis. En concreto las propuestas de continuación de la tesis son:

- Aumentar las fuentes de datos que alimenten la herramienta con el objetivo de tener una visión más integral del estudiante.
- Extender la solución al análisis de otros aspectos como: el trabajo con foros, cuestionarios, glosarios, wiki y demás herramientas presentes en el Entorno Virtual de Enseñanza y Aprendizaje.
- Extender la presente solución a otros niveles de análisis propuestos por el Instituto SAS como pueden ser: la generación de alertas, la confección de pronósticos, la realización de modelos predictivos y la optimización.
- Integrar técnicas de minería de datos a la solución de forma tal que sean útiles y accesibles a usuarios no especializados.
- Divulgar y extender el uso de la aplicación propuesta a otros colectivos de la universidad.

- Evaluar la actividad y estado de otros actores tales como profesores, tutores, coordinadores, con el fin de validar que su trabajo sea eficiente y medir cargas de trabajo, entre otras funcionalidades.
- Aparejar a cualquier utilización de la actual solución una capacitación con el objetivo de entender sus funcionalidades y aprovechar mejor sus potencialidades.
- Y finalmente, ampliar la solución a otros departamentos de la institución formativa tales como financiero, marketing y comercial para tener una visión más estratégica de lo que está ocurriendo en la institución.

Lista de referencias

- Abell, S. K. (1999). Development of professional knowledge in learning to teach elementary science. *Journal of Research in Science Teaching*, 121–139.
- Alavi, M., & Leidner, D. (2001). Review: knowledge management and knowledge management systems: conceptual foundations and research issues. *MIS Quarterly*, 107-136.
- Andersen, A. (1999). El management en el siglo XXI. Herramientas para los desafíos empresariales para la próxima década. *Granica*, Buenos Aires.
- Andreu, R., & Sieber, S. (2000). La Gestion Integral del Conocimiento y del Aprendizaje. *Economía Industrial*.
- Arambarri, J. (2012). *Metodología de Evaluacion y Gestion del Conocimiento dinamico por procesos utilizando como soporte TIC el Entorno Colaborativo de Trabajo basado en el modelo de creacion de Conocimiento de Nonaka-Takeuchi*. Córdoba: Universidad de Córdoba.
- Area, M., & Adell, J. (2009). e-Learning: Enseñar y Aprender en Espacios Virtuales. *Tecnología Educativa. La formación del profesorado en la era de Internet*, 391-424.

- Avouris, N., Komis, V., Fiotakis, G., Margaritis, M., & Voyiatzaki, G. (2005). Logging of fingertip actions is not enough for analysis of learning. *Proceedings of Workshop Usage Analysis in learning*.
- Backler, F. (. (1995). Knowledge, knowledge work and organizations: an overview and interpretation. *Organization Studies 16 (6)*, 1021-1046.
- Baelo Alvarez, R. (2009). El e-Learning, una respuesta educativa a las demandas de las sociedades del siglo XXI. *Pixel-Bit. Revista de Medios y Educación*, 87-96.
- Bernabéu, R. D. (2010). Hefesto: Metodología para la Construcción de un Data Warehouse. *HEFESTO - Data Warehousing: Investigación y Sistematización de Conceptos*.
- Beyer, M., & Feinberg, D. (2006). The Basics of datawarehousing. *Gartner*.
- Birkett, B. (1995). Knowledge Management. *Chartered Accountants Journal of New*, 14-18.
- Bitterer, A. (2006). Poor-Quality Data: The Sure Way to Lose Business and Attract Auditors. *Gartner*.
- Black, D. H., & Synan, C. (1997). The learning organization: The sixth dimension. *Management Accounting*.
- Blumberg, R., & Atre, S. (2003). The Problem with Unstructured Data. *DM Review Magazine*.
- Bouman, R., & Van Dongen, J. (2009). *Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQL*. Indiana: Wiley Publishing.

- Bueno, E. (1998). El capital intangible como clave estratégica en la competencia actual. *Boletín de estudios económicos*, 207-229.
- Bueno, E. (1999). Gestión de conocimiento, aprendizaje y capital intelectual. *Boletín del club Intelect.*
- Bustelo, C., & Amarilla, R. (2001). Gestión del conocimiento y gestión de la información. *Boletín del Instituto Andaluz de Patrimonio Histórico*, 226-230.
- Cano, J. L. (2007). *Business Intelligence: Competir con Información*. Fundación Banesto.
- Cebrián, M. (2003). *Innovar con tecnologías aplicadas a la docencia universitaria: Enseñanza virtual para la innovación universitaria*. Panamá: Nancea.
- Claros, I., & Cobos, R. (2013). Pautas para la implementación de Analíticas de Aprendizaje en Entornos Colaborativos Centrados en la Interacción Social. *Presentado en XV Simposio Internacional de Tecnologías de la Información y las Comunicaciones en la Educación, España*.
- COIT. (1996). *La evolución del Conocimiento y el Empleo*. Colegio Oficial de Ingenieros de Telecomunicación.
- Corrales, M. I. (2008). *Metodología de la formación abierta y a distancia*. México: Limusa.
- Cross, R., & Lloyd Baird, L. (2000). Technology is not enough: Improving performance by building. *Sloan Management Review*.

Cuervo, A. (1999). Las fusiones y adquisiciones de empresas: ¿Algo más que una moda? *Expansión*.

Dataprix. (2014). *Data warehouse manager. Metodología datawarehousing / Manual IT online*. Recuperado el marzo de 3 de 2014, de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>

Davenport, T. H. (1994). The Coming Soon: The CKO. *InformationWeek*.

Davenport, T., & Prusak, L. (1998). *Working Knowledge: how organizations manage what they know*. Harvard Business School Press.

Davenport, T., De Long, D., & Beers, M. (1998). Successful knowledge management projects. *Sloan management review*, 43-57.

De la Rica, E. (1999). *Gestión del conocimiento: caminando hacia las organizaciones inteligentes*. Recuperado el 05/02/2015, de <http://www.hacienda.go.cr/centro/datos/Articulo/Caminando%20hacia%20las%20organizaciones%20inteligentes.doc>.

Drucker, P. (1993). *Post-capitalist society*. Butterworth-Hainemann, Oxford.

Duncan, R., & Weiss, A. (1979). Organizational Learning: Implications for Organizational Design. *Research in Organizational Behavior*.

Earl, M., & Scott, I. (1999). What is a Chief Knowledge Officer? *Sloan Management Review*, 29-38.

Eckerson, W. (s.f.). *Four Ways to Build a Data Warehouse*.

Eckerson, W. W. (2002). Data Quality and the Bottom Line. *TDWI, Report Series*.

- Eckerson, W. W., & Howson, C. (2005). Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale. *TDWI Report Series*.
- Eckerson, W. W., & Howson, C. (2005). Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale. *TDWI Report Series*.
- Eckerson, W., & White, C. (2003). Evaluating ETL and Data Integration Platforms. *TDWI Report Series*.
- Edvinsson, L. (1997). Developing Intellectual Capital at Skandia. *Long Range Planning*.
- Estay Niculcar, C., & Plaza Pimentel, L. (2012). *Business Intelligence y Gestión Documental*. Barcelona: Fundación Universitaria Iberoamericana.
- Fidalgo, A. (2012). *Learning analytics. Qué, cómo y para qué*. Obtenido de <https://innovacioneducativa.wordpress.com/2012/11/10/learning-analytics-analiticas-de-aprendizaje-que-como-y-para-que/>
- Fuchs, G. (2003). Aspects of ROI.
- García Peñalvo, F. J. (2005). Estado actual de los sistemas e-learning. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*.
- Garnica Olmos, E. (1997). El rendimiento estudiantil: una metodología para su medición. *Revista Economía*, 7-26.
- Garrison, D., & Anderson, T. (2005). *El e-learning en el siglo XXI*. Octaedro.
- Garvin, D. A. (1994). Building a learning organization. *Business Credit*, 19-28.

GISMO. (2010). *GISMO*. Obtenido de <http://gismo.sourceforge.net/>

Goad, T. W. (2002). *Information literacy and workspace performance*. Westport: Quorum books.

Goleman, D. (1999). *La inteligencia emocional en la empresa*. Argentina: Ediciones Vergara.

Gómez Aguilar, D. A., Theron, R., & García Peñalo, F. (2008). *Visual analytics, for virtual education platforms*. Salamanca: Universidad de Salamanca.

Gopal, C., & Gagnon, J. (1995). Knowledge, information, learning and the IS manager. *Computerworld (Leadership Series)*, 1-7.

Gracia, S. (2006). *Gestión del conocimiento y aprendizaje organizacional*. Fundación Universitaria Iberoamericana.

Grant, R. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. *California Management Review*.

Hedlund, G. (1994). A model of knowledge management and the N-form corporation. *Strategic Management Journal*.

Hedlund, G., & Nonaka, I. (1993). Models of Knowledge Management in the West and Japan. *Implementing Strategic Process: Change, Learning and Cooperation*.

Hernández Orallo, J., Ramírez Quintana, J., & Ferri Ramírez, C. (2004). *Introducción a la Minería de Datos*. Valencia: Pearson Educación.

Hyde, J. (2007). *Mondrian 2.2.2 Technical Guide*.

Inmon, W. (2006). Real-Time Decision Support Systems? *Business Intelligence*.

-
- Inmon, W. H. (1996). *Building the Data Warehouse*. New York: John Wiley & Sons.
- Jackson, C. (2000). *Process to product: creating tools for knowledge Management*. Lisboa: International conference KM.
- Jennex, M. E. (2005). What is KM? *International Journal of Knowledge Management*.
- Kimball, R. (1996). *The datawarehouse Toolkit*. Wiley.
- Kimball, R., & Caserta, J. (2002). *The Data Warehouse ETL Toolkit*. Chichester: John Wiley & Sons.
- Kimball, R., & Ross, M. (2002). *The Data Warehouse. ETL Toolkit. The Complete Guide to Dimensional Modeling (Second Edition, Vol. I)*. Indiana: Wiley Publishing.
- Koulopoulos, T. M., & Frappaolo, C. (2000). *SMART. Lo fundamental y lo más efectivo acerca de la gerencia del conocimiento*. Santa Fé de Bogotá: McGraw-Hill Interamericana, S.A.
- Lapuh Bele, J., Bele, D., Pirnat, R., & Anžin Lončarić, V. (2015). Business Intelligence in e-Learning. *Proceedings of 2015 International Conference on Interactive Collaborative Learning (ICL)*.
- Lara, J. L. (2001). *El estado argentino y su capital en la era del conocimiento: un modelo para la acción*. Buenos Aires.
- León, M., Ponjuán, G., & Rodríguez, M. (2004). *Procesos estratégicos de gestión del conocimiento*.

- Leonard-Barton, D. A. (1995). *Modes of Technology Transfer. In Technology Management and Corporate Strategies: A Tricontinental Perspective*. Elsevier.
- Leony, D., Pardo, A., Valentin, L., Quiñones, I., & Delgado Kloos, C. (2013). *Learning analytics in the LMS: Using browser extensions to embed visualizations into a Learning Management System*. Madrid: Universidad Carlos III de Madrid.
- Loshin, D. (2003). *Business Intelligence: The Savvy Manager's Guide, Getting Onboard with Emerging IT*. Morgan Kaufmann.
- Macintosh, A. (1996). *Position Paper on Knowledge Assets Management*". *Artificial Intelligence Applications Institute*. Scotland: University of Edimburgh.
- Malhorta, Y. (1998). Knowledge management for the new world of business. *Asian Strategy Leadership Institute Review*.
- Malhotra, Y. (1997). Knowledge Management in Inquiring Organizations. *Proceedings of the Association for Information Systems Third Americas Conference on Information Systems*.
- Management en Administración de la Información. (2000). *El Diario*.
- Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, 125–139.
- Mazza, R., Bettoni, M., Faré, M., & Mazzola, L. (2012). MOCLog – Monitoring Online Courses with log data. *Presentado en 1st Moodle ResearchConference*, 132-139.

- Mejia, C., Bull, S., Vatrapu, R., Florian, B., & Fabregat, R. (2013). *PADA: a Dashboard of Learning Analytics for University Students with Dyslexia*. Girona: Institute of Informatics and Applications, University of Girona.
- Moodle. (2015a). *Acerca de Moodle - MoodleDocs*. Recuperado el 2015, de https://docs.moodle.org/all/es/Acerca_de_Moodle
- Moodle. (2015b). *Pedagogia - MoodleDocs*. Obtenido de <https://docs.moodle.org/all/es/Pedagog%C3%ADA>
- Moodle. (2015c). *Database Schema - Moodle Docs*. Obtenido de https://docs.moodle.org/dev/Database_Schema
- Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, 177-195.
- Nevis, E., A., DiBella., & Gould, J. (1995). Understanding Organizations as Learning Systems. *Sloan Management Review*.
- New Media Consortium. (2013). NMC Horizon Project Short List Higher Education Edition. *New Media Consortium*.
- Nieves, Y., & León, M. (2001). Gestión del conocimiento: una nueva perspectiva en la gerencia de las organizaciones.
- Nonaka, I. (1991). The knowledge creating company. *Harvard Business Review*, 96-104.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge. *Organization Science*, 14-37.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company*. EEUU: Oxford University Press.

North, K. (2001). *Organización basada en Conocimiento (La Cuarta Dimensión)*. Madrid: Taller “Estado del Capital Intelectual” del Foro del Conocimiento Intellectus.

O'Dell, C., & Grayson, J. J. (1998). *If we only know what we know*. New York: *The Free*.

OECD. (1996). *Resources for Lifelong Learning: What Might be Needed and How Might it be Found?* Paris: OECD.

Ordóñez de Pablos, P. (2011). *Gestión del Conocimiento en las Multinacionales. Relevancia Estratégica del Capital Humano*.

Páez Urdaneta, I. (1992). *Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo informacional : retos y oportunidades*. Caracas: Universidad Simón Bolívar. Consejo Nacional de Investigaciones Científicas y Tecnológicas.

Parr, O. (2000). *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management*. John Wiley & Sons.

Pávez, A. (2001). *Modelo de Implantación de Gestión de la Información para la Generación de Ventajas Competitivas*. Valparaíso: Universidad Técnica Federico Santa María.

Pentaho BI. (2013). *Introducción a Pentaho Business Intelligence*. Obtenido de <http://pentaho.almacen-datos.com/>

- Pentaho Community. (2012). Pentaho. Open source business intelligence. *Pentaho TM*.
- Pérez-Montoro Gutiérrez, M. (2008). *Gestión del conocimiento en las organizaciones: fundamentos, metodología y praxis*. Gijón: Trea.
- Plaz Landaeta, R. (1997). *Innovacion para el cambio. Capital Intelectual y Gestion del Conocimiento*. Caracas: Jornadas Tecnicas de ASOQUIM. Asociacion de Industriales del sector Quimico.
- Ponjuan, G. (1998). *Gestion de Informacion en las organizaciones: Principios, conceptos y aplicaciones*. Chile: Impresos Universitaria.
- Porter, M. E., & Millar, V. (1986). *Cómo obtener ventajas competitivas por medio de la información*. España: Harvard DEUSTO Business Review.
- Prahalad, C. K., & Hamel, G. (1990). The Core Competence of the Corporation. *Harvard Business Review*, 79-91.
- PriceWaterhouse. (1998). Guía para la caracterización del conocimiento estratégico. Modelo de Bartok.
- Prieto, D., & van de Pol, P. (2006). *e-Learning, comunicación y educación*. Radio Nederland Training Centre.
- Probst, G., Raub, S., & Romhardt, K. (2001). *Administre el conocimiento*. México: Pearson Educación.
- Rice, M. B. (2004). Business intelligence tools can help turn out savings in core cost areas. *Managed Healthcare Executive*, 40.

- Rivero, S. (2002). *Claves y Pautas para Comprender e Implantar la Gestion del Conocimiento. Un Modelo de Referencia*. Bilbao: Socintec.
- Rodríguez Gómez, D. (2007). Modelos para la creación y gestión del conocimiento: una aproximación teórica. *Educación*, 25-39.
- Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Systems, Man, and Cybernetics*, 601-618.
- Saavedra Fernandez, O. (s.f.). *La producción de conocimiento y las tecnologías de información: Valores, conceptos y herramientas*. EBSCO Information Services.
- Salinas, J. (2005). La gestión de los entornos virtuales de formación. *Seminario Internacional: La Calidad de la Formación en Red en el Espacio Europeo de Educación Superior*.
- Salinas, J., Negre, F., Gallardo, A., Escandell, C., & Torrandell, I. (2007). *Análisis de elementos que intervienen en el proceso de enseñanza aprendizaje en un entorno virtual de formación: Propuesta de un modelo didáctico*. Buenos Aires: Edutec.
- Samiotis, K., & Poulymenakou, A. (2003). Understanding knowledge management interventions: Evidence from supporting (e-)banking activities. *Knowledge and Process Management: Journal of Corporate Transformation*.
- SAS Institute. (2008). Eight levels of Analytics . *SASCOM Magazine*.
- Schoemaker, P. (1992). How to link strategic vision to core competences. *Sloan Management Review*.

- Sedeño Prado, Y. (2001). *La gestión del conocimiento, una nueva alternativa para el desarrollo de la economía latinoamericana*. Recuperado el 2015, de <http://www.gestiopolis.com/recursos/documentos/fulldocs/ger/kmcuba.htm>.
- Senge, P. (1990). *La quinta disciplina*. México: Ediciones Granítica.
- Shirky, C. (2010). *Excedente cognitivo*. Barcelona: Ed. Planeta.
- Shrivastava, P. (1983). A typology of organizational learning systems. *Journal of Management Studies*, 7-28.
- Siemens, G., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., & Baker, R. (2011). Open Learning Analytics: an integrated & modularized platform. *Canada: Solar. Society for Learning Analytics Research*.
- Silva Quiroz, J. (2011). *Diseño y moderación de entornos virtuales de aprendizaje*. Barcelona: Editorial UOC.
- Skelton, V., & Abell, A. (2001). Developing skills for the information services workforce in the knowledge economy. *Library and Information Commission Research Report*.
- Skyrme, D. (1999). *Knowledge Networking: Creating the Collaborative Enterprise*. Butterworth-Heinemann.
- Stair, R. M., & Reynolds, G. W. (2012). *Fundamentals of Information Systems*. Boston: Cengage Learning.
- Stewart, T. A. (1999). *Intellectual Capital: The new Wealth of Organizations*. EEUU: Doubleday.

Strange, K. (2001). The Challenges of Implementing a datawarehouse to Achieve Business Agility. *Gartner*.

Tissen, R., Andriessen, D., & Deprez. (1998). Value-based knowledge management. *Addison Wesley*.

Uribe, A., Melgar, L. M., & Bornacelly, J. A. (2007). Moodle learning management system as a tool for information, documentation, and knowledge management by research groups. *Profesional de la Informacion*, 468-474.

Van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in Higher Education: Establishing a Common Language. *EDUCAUSE*.

Van de Pol, P. (2006). *E-learning, comunicacion y educacion*. San Jose: Edicion RNTC.

Vercellis, C. (2009). *Business Intelligence. Data Mining and Optimization for Decision Making*. Milano: Wiley Publishing.

Vesset, D., & McDonough, B. (2006). Qlik Tech's Approach to Business Intelligence: Keep It Simple and Flexible. *IDC*.

Vozniuk, A., Govaerts, S., & Gillet, D. (2013). *Towards portable learning analytics dashboards*. Suiza: EPFL.

Watson, H. J. (2002). Recent Developments in Data Warehousing. *Communications of AIS*.

Watson, H., Annino, D., Wixom, B., K.L, A., & Rutherford, M. (2001). Current Practices in Data Warehousing. *Information Systems*, 47-55.

White, C. (2004). Now is the Right Time for the Real Time BI. *DM Review*.

-
- Whittemore, B. (2003). *The Business Intelligence ROI Challenge: Putting It All Together*.
- Wiig, K. (1997). Integrating Intellectual Capital and Knowledge Management. *Long Range Planning*.
- Yadira, N. L., & Magda León, S. (2001). La Gestión del conocimiento: una nueva perspectiva en la gerencia.
- Zack, M. (1998). *Developing a Knowledge Strategy*.
- Zorrilla, M. E., & Álvarez, E. (2008). MATEP: Monitoring and Analysis Tool for E-learning Platforms. *8th IEEE International Conference on Advanced Learning Technologies*, 611-613.
- Zúmel, P. (2008). Gestión del rendimiento. *Gestión del rendimiento*.

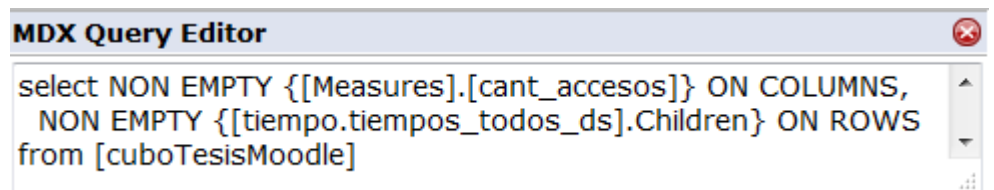
Apéndice

1 Resultados de los análisis OLAP

1.1 Accesos por día de la semana

La presente vista muestra cómo se comportaron los accesos teniendo en cuenta los días de la semana.

- Consulta MDX



```

MDX Query Editor
select NON EMPTY {[Measures].[cant_accesos]} ON COLUMNS,
  NON EMPTY {[tiempo.tiempos_todos_ds].Children} ON ROWS
from [cuboTesisMoodle]
  
```

- Resultado

| | Medidas |
|------------------|-----------------------|
| tiempos_todos_ds | ● Cantidad de Accesos |
| + 01_Domingo | 630 |
| + 02_Lunes | 1.505 |
| + 03_Martes | 5.089 |
| + 04_Miercoles | 2.019 |
| + 05_Jueves | 3.698 |
| + 06_Viernes | 1.839 |
| + 07_Sabado | 285 |

Tabla 60 Accesos por día de la semana

- Gráfica Alternativa

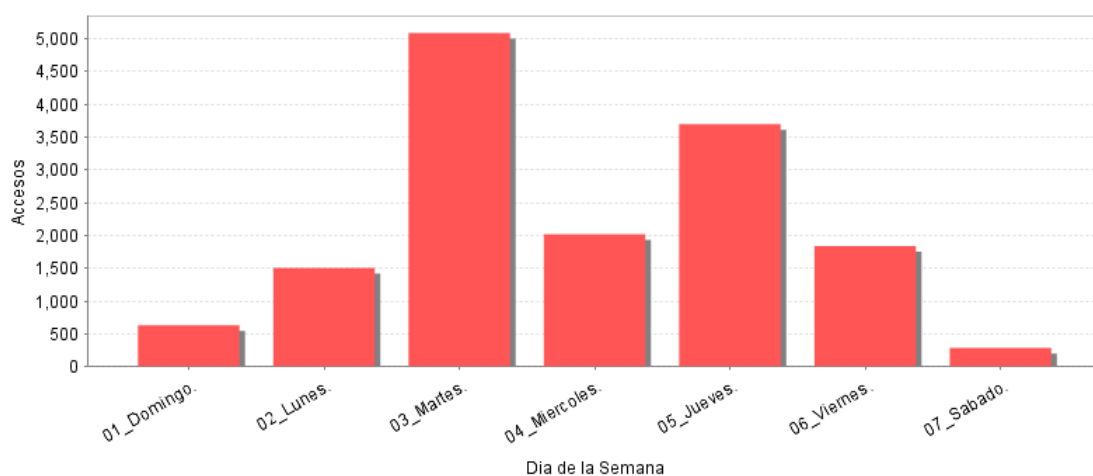


Figura 49 Accesos por día de la semana

1.2 Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores)

La presente vista muestra cómo se comportaron los accesos teniendo en cuenta los días de la semana y el grupo docente.

- Consulta MDX

```
MDX Query Editor
select NON EMPTY Crossjoin({[Measures].[cant_accesos]},
{[usuario.usuarios_todosxg].[4301], [usuario.usuarios_todosxg].[4302],
[usuario.usuarios_todosxg].[4303], [usuario.usuarios_todosxg].[4304]})
ON COLUMNS,
NON EMPTY {[tiempo.tiempos_todos_ds].Children} ON ROWS
from [cuboTesisMoodle]
```

- Resultado

| | Medidas | | | |
|------------------|--------------------------|----------|----------|----------|
| | Cantidad de Accesos | | | |
| | Todos los UsuariosXGrupo | | | |
| tiempos_todos_ds | ● + 4301 | ● + 4302 | ● + 4303 | ● + 4304 |
| + 01_Domingo | 174 | 61 | 264 | 25 |
| + 02_Lunes | 152 | 162 | 211 | 73 |
| + 03_Martes | 1.466 | 522 | 1.607 | 717 |
| + 04_Miercoles | 301 | 181 | 593 | 372 |
| + 05_Jueves | 976 | 607 | 1.005 | 288 |
| + 06_Viernes | 476 | 117 | 342 | 401 |
| + 07_Sabado | 55 | 32 | 126 | 20 |

Tabla 61 Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores)

• Gráfica Alternativa

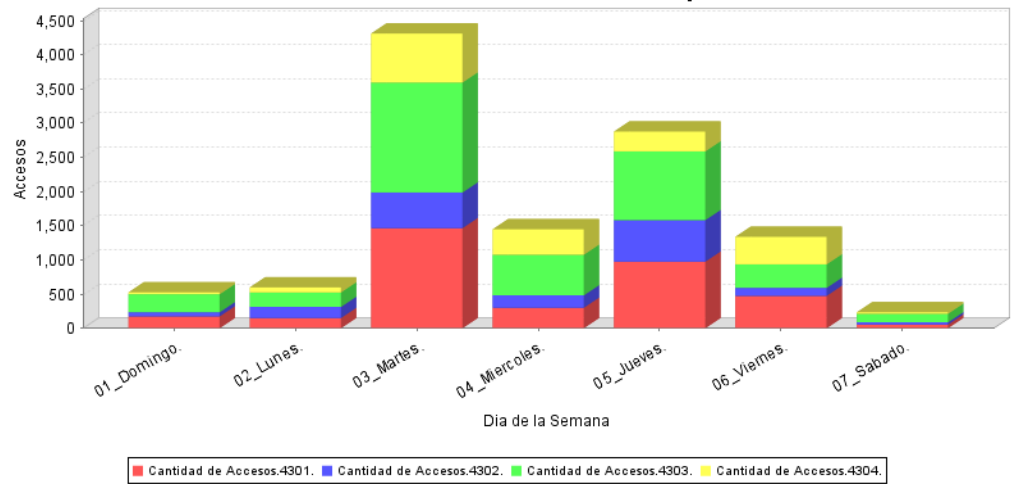


Figura 50 Accesos por día de la semana y por grupo docente (sin incluir estudiantes repetidores)

1.3 Accesos por hora del día

La presente vista muestra cómo se comportaron los accesos teniendo en cuenta las horas del día.

- Consulta MDX

```
MDX Query Editor
select NON EMPTY {[Measures].[cant_accesos]} ON COLUMNS,
NON EMPTY [tiempo.tiempos_todos_h_mi].[All
tiempo.tiempos_todos_h_mis].Children ON ROWS
from [cuboTesisMoodle]
```

- Resultado

| | Medidas |
|--------------------|-----------------------|
| tiempos_todos_h_mi | ● Cantidad de Accesos |
| + 00 | 287 |
| + 01 | 131 |
| + 02 | 59 |
| + 03 | 19 |
| + 04 | 7 |
| + 06 | 2 |
| + 07 | 327 |
| + 08 | 1.206 |
| + 09 | 2.027 |
| + 10 | 1.728 |
| + 11 | 1.895 |
| + 12 | 850 |
| + 13 | 859 |
| + 14 | 830 |
| + 15 | 979 |
| + 16 | 665 |
| + 17 | 413 |
| + 18 | 295 |
| + 19 | 182 |
| + 20 | 325 |
| + 21 | 796 |
| + 22 | 782 |
| + 23 | 401 |

Tabla 62 Accesos por hora del día

- Gráfica Alternativa

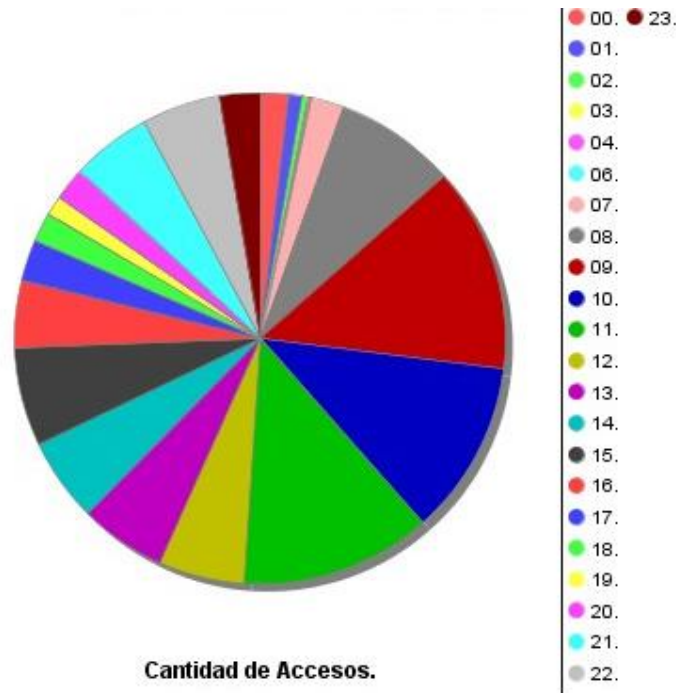


Figura 51 Accesos por hora del día

1.4 Cantidad de estudiantes que accedieron por módulo del sistema, teniendo en cuenta el grupo de repetidores.

La presente vista muestra cuántos estudiantes visitaron los diferentes módulos. Esto está referido a todo el semestre de clase, si alguna vez un estudiante visitó un módulo cualquiera, entonces estará contemplado en este resultado. Si se desea en un periodo determinado se puede realizar un filtro. Aquí se incluye los estudiantes del grupo de los repetidores.

- Consulta MDX

```

MDX Query Editor
select NON EMPTY {[Measures].[cant_usuarios]} ON COLUMNS,
NON EMPTY {[modulo.modulos_todos].Children} ON ROWS
from [cuboTesisMoodle]
where {[usuario.usuarios_todosxg].[4301], [usuario.usuarios_todosxg].
[4302], [usuario.usuarios_todosxg].[4303], [usuario.usuarios_todosxg].
[4304], [usuario.usuarios_todosxg].[Arrastre]}

```

- Resultado

| | Medidas |
|-------------------|------------------------|
| Todos los modulos | • Cantidad de Usuarios |
| assign | 78 |
| course | 78 |
| folder | 76 |
| forum | 12 |
| glossary | 22 |
| resource | 68 |
| user | 11 |

Tabla 63 Cantidad de estudiantes que accedieron por módulo del sistema, teniendo en cuenta el grupo de repetidores.

- Gráfica Alternativa

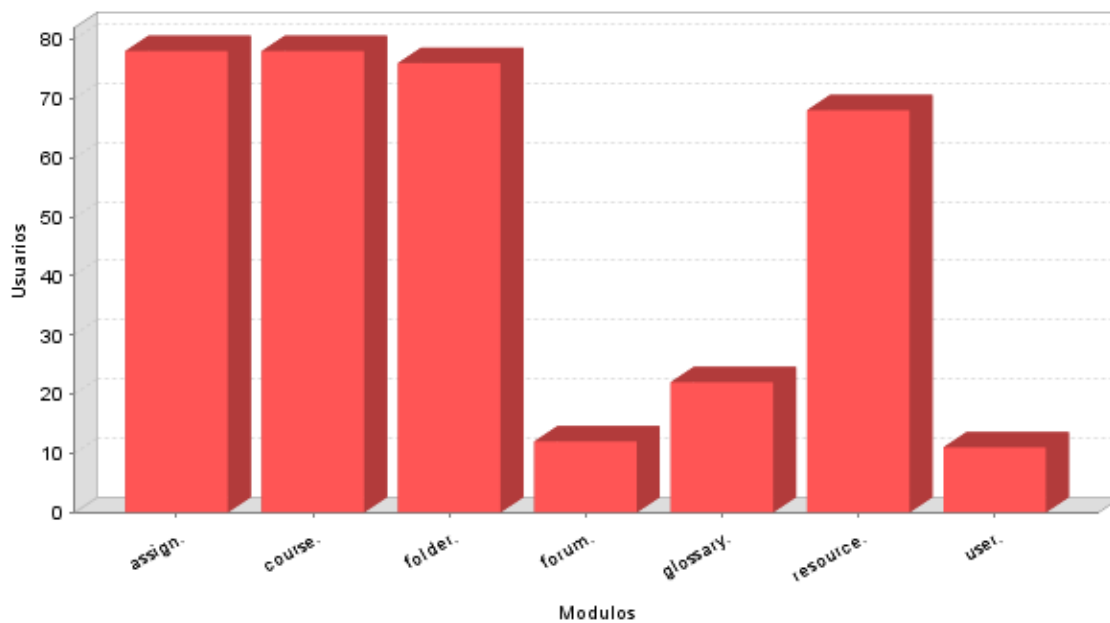


Figura 52 Cantidad de estudiantes que accedieron por módulo del sistema, teniendo en cuenta el grupo de repetidores.

1.5 Cantidad de Accesos por miembro del grupo. Porcentaje relativo de acceso al grupo por cada miembro.

La presente vista está más elaborada. Esta muestra la cantidad de accesos por cada grupo y sus miembros; se refleja además el porcentaje relativo de cada uno de ellos con respecto al grupo que pertenece. Esto significa, por ejemplo, que el grupo 4302 ha realizado 1682 accesos, lo que representa el 11.16% de todos los accesos al sistema; de igual forma el tercer estudiante ha realizado 58 accesos, lo que representa el 3.45 % de los accesos de su grupo.

- Consulta MDX

```

MDX Query Editor
with member [Measures].[Porcentaje Relativo Al Grupo] as '((([usuario.usuarios_todosxg].[nombresUsuario].CurrentMember, [Measures].[cant_accesos]) / ([usuario.usuarios_todosxg].[nombresUsuario].CurrentMember.Parent, [Measures].[cant_accesos])))', FORMAT_STRING = IIF((([Measures].[Porcentaje Relativo Al Grupo] > 0.60), "|#,#0.00%|style=red", "|#,#0.00%|style=green")
select NON EMPTY {[Measures].[cant_accesos], [Measures].[Porcentaje Relativo Al Grupo]} ON COLUMNS,
NON EMPTY Hierarchize({[usuario.usuarios_todosxg].[nombresUsuario].Members, [usuario.usuarios_todosxg].[grupoUsuario].Members}) ON ROWS
from [cuboTesisMoodle]
    
```

- Resultado (sólo una parte)

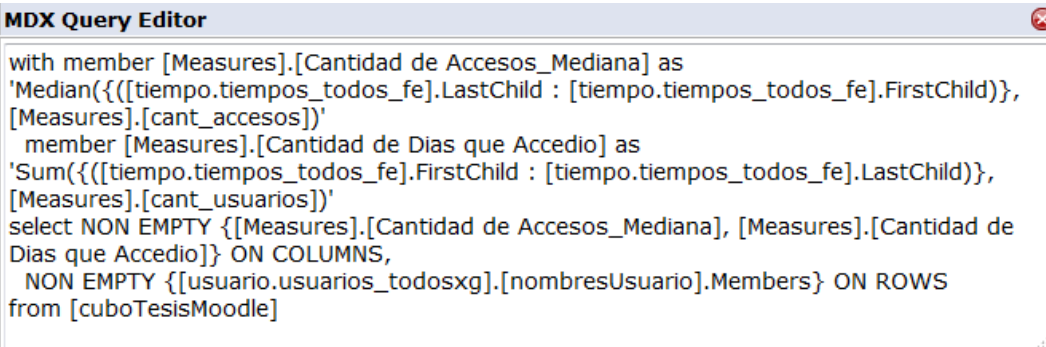
| | | |
|------|-------|--------|
| 4302 | 1.682 | 11,16% |
| + | 106 | 6,30% |
| + | 125 | 7,43% |
| + | 58 | 3,45% |
| + | 134 | 7,97% |
| + | 99 | 5,89% |
| + | 63 | 3,75% |
| + | 101 | 6,00% |
| + | 138 | 8,20% |
| + | 154 | 9,16% |
| + | 102 | 6,06% |
| + | 86 | 5,11% |
| + | 115 | 6,84% |
| + | 97 | 5,77% |
| + | 183 | 10,88% |
| + | 81 | 4,82% |
| + | 40 | 2,38% |

Tabla 64 Cantidad de Accesos por miembro del grupo. Porcentaje relativo de acceso al grupo por cada miembro

1.6 Cantidad de días que accedió un usuario en todo el semestre y la mediana de accesos por día.

La presente vista obtiene la cantidad de días en que un usuario accedió al sistema. Posteriormente, y teniendo en cuenta la cantidad de accesos realizados en cada uno de estos días, calcula la mediana de estos accesos. Todo esto significa, por ejemplo, que el último estudiante del resultado mostrado abajo accedió al sistema solo 22 días en todo el semestre y en no más de 11 de estos produjo una cantidad superior a 7 accesos y en no más de 11 también, produjo una cantidad inferior a esta cifra.

- Consulta MDX



```
with member [Measures].[Cantidad de Accesos_Mediana] as
'Median({([tiempo.tiempos_todos_fe].LastChild : [tiempo.tiempos_todos_fe].FirstChild)},
[Measures].[cant_accesos])'
  member [Measures].[Cantidad de Dias que Accedio] as
'Sum({([tiempo.tiempos_todos_fe].FirstChild : [tiempo.tiempos_todos_fe].LastChild)},
[Measures].[cant_usuarios])'
select NON EMPTY {[Measures].[Cantidad de Accesos_Mediana], [Measures].[Cantidad de
Dias que Accedio]} ON COLUMNS,
  NON EMPTY {[usuario.usuarios_todosxg].[nombresUsuario].Members} ON ROWS
from [cuboTesisMoodle]
```

- Resultado

| | Medidas | |
|--------------------------|-------------------------------|--------------------------------|
| Todos los UsuariosXGrupo | ● Cantidad de Accesos_Mediana | ● Cantidad de Dias que Accedio |
| + [Redacted] | 10 | 25 |
| + [Redacted] | 11 | 17 |
| + [Redacted] | 8 | 14 |
| + [Redacted] | 11 | 15 |
| + [Redacted] | 5 | 15 |
| + [Redacted] | 4 | 42 |
| + [Redacted] | 6 | 34 |
| + [Redacted] | 6,5 | 16 |
| + [Redacted] | 3 | 15 |
| + [Redacted] | 3 | 21 |
| + [Redacted] | 6 | 14 |
| + [Redacted] | 10 | 20 |
| + [Redacted] | 9 | 19 |
| + [Redacted] | 7 | 22 |

Tabla 65 Cantidad de días que accedió un usuario en todo el semestre y la mediana de accesos por día

1.7 Total de accesos, cantidad de días que accedió, promedio de acceso por día, desviación estándar y coeficiente de variación.

Esta vista es más integradora y agrupa medidas de tendencia central (promedio) y de dispersión (desviación estándar y coeficiente de variación).

- Consulta MDX

```

MDX Query Editor
with member [Measures].[Cant. Dias que Accedio] as 'Sum({([tiempo.tiempos_todos_fe].FirstChild :
[tiempo.tiempos_todos_fe].LastChild)}, [Measures].[cant_usuarios])'
member [Measures].[Total Accesos] as 'Sum({([tiempo.tiempos_todos_fe].[03-06-2014] :
[tiempo.tiempos_todos_fe].[07-16-2014]}), [Measures].[cant_accesos])'
member [Measures].[Prom. Acceso Por Dia] as '((([usuario.usuarios_todosxg].
[nombresUsuario].CurrentMember, [Measures].[Total Accesos]) / ([usuario.usuarios_todosxg].
[nombresUsuario].CurrentMember, [Measures].[Cant. Dias que Accedio]))', FORMAT_STRING =
"#,##0.00"
member [Measures].[Desv. Est] as 'Stdev({([tiempo.tiempos_todos_fe].LastChild :
[tiempo.tiempos_todos_fe].FirstChild)}, [Measures].[cant_accesos])'
member [Measures].[Coef. Varia] as '((([usuario.usuarios_todosxg].
[nombresUsuario].CurrentMember, [Measures].[Desv. Est]) / ([usuario.usuarios_todosxg].
[nombresUsuario].CurrentMember, [Measures].[Prom. Acceso Por Dia]))', FORMAT_STRING =
Iif((([Measures].[Coef. Varia] > 1.0), "|#,##0.00%|style=red", "|#,##0.00%|style=green")
select NON EMPTY {[Measures].[Total Accesos], [Measures].[Cant. Dias que Accedio], [Measures].
[Prom. Acceso Por Dia], [Measures].[Desv. Est], [Measures].[Coef. Varia]} ON COLUMNS,
NON EMPTY {[usuario.usuarios_todosxg].[nombresUsuario].Members} ON ROWS
from [cuboTesisMoodle]
    
```

- Resultado

| Todos los UsuariosXGrupo | Medidas | | | | |
|--------------------------|---------------|------------------------|----------------------|-----------|-------------|
| | Total Accesos | Cant. Dias que Accedio | Prom. Acceso Por Dia | Desv. Est | Coef. Varia |
| | 214 | 20 | 10,70 | 8,073 | 75,45% |
| | 191 | 19 | 10,05 | 6,069 | 60,37% |
| | 183 | 22 | 8,32 | 6,183 | 74,33% |
| | 209 | 17 | 12,29 | 14,894 | 121,15% |
| | 276 | 22 | 12,55 | 8,7 | 69,35% |
| | 42 | 7 | 6,00 | 2,309 | 38,49% |
| | 228 | 20 | 11,40 | 7,639 | 67,01% |
| | 65 | 11 | 5,91 | 4,158 | 70,37% |
| | 349 | 31 | 11,26 | 14,935 | 132,66% |
| | 106 | 13 | 8,15 | 7,128 | 87,42% |
| | 125 | 15 | 8,33 | 4,287 | 51,45% |
| | 58 | 9 | 6,44 | 4,333 | 67,24% |
| | 134 | 14 | 9,57 | 5,557 | 58,06% |

Tabla 66 Total de accesos, cantidad de días que accedió, promedio de acceso por día, desviación estándar y coeficiente de variación

1.8 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación.

Esta vista muestra la cantidad de estudiantes que se evaluaron en cada actividad de evaluación realizada y la cantidad que no lo hicieron. Muestra además el porcentaje de estudiantes evaluados.

- Consulta MDX

```

MDX Query Editor
with member [Measures].[%. Est. Eval] as '((([actEval.actEva_todas].
[nombreActEval].CurrentMember, [Measures].[cant_actEva_realizadas]) /
([actEval.actEva_todas].[nombreActEval].CurrentMember, [Measures].
[cant_actEva_debioRealizar]))', FORMAT_STRING = "#,##0.00%"
select NON EMPTY {[Measures].[cant_EvaNoRealizadas], [Measures].
[cant_actEva_realizadas], [Measures].[%. Est. Eval]} ON COLUMNS,
NON EMPTY {[actEval.actEva_todas].Children} ON ROWS
from [cuboTesisMoodleEvaluacion]
    
```

- Resultado (sólo una parte)

| | Medidas | | |
|----------------------------|------------------------|------------------------------|----------------|
| Todas las ActEvaluacion | • cant_EvaNoRealizadas | • Cantidad ActEva realizadas | • %. Est. Eval |
| 01_Tarea#1 | 59 | 20 | 25,32% |
| 02_Tarea#2 | 67 | 12 | 15,19% |
| 03_Tarea#3 | 69 | 10 | 12,66% |
| 04_Tarea#4 | 68 | 11 | 13,92% |
| 05_Tarea#5 | 55 | 24 | 30,38% |
| 06_Tarea#6 | 66 | 13 | 16,46% |
| 07_Tarea#7 | 35 | 44 | 55,70% |
| 08_Tarea#8 | 42 | 37 | 46,84% |
| 09_Tarea#9 | 58 | 21 | 26,58% |
| 10_Tarea#10 | 63 | 16 | 20,25% |
| 11_Tarea#11 | 32 | 47 | 59,49% |
| 12_Tarea#12 | 63 | 16 | 20,25% |
| 13_Tarea#13 | 69 | 10 | 12,66% |
| 14_Tarea#14 | 74 | 5 | 6,33% |
| 15_Tarea#15 | 63 | 16 | 20,25% |
| 16_Tarea#16 | 65 | 14 | 17,72% |
| 17_Primer a Prueba Parcial | 4 | 75 | 94,94% |

Tabla 67 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación.

- Gráfica Alternativa (sólo una parte)

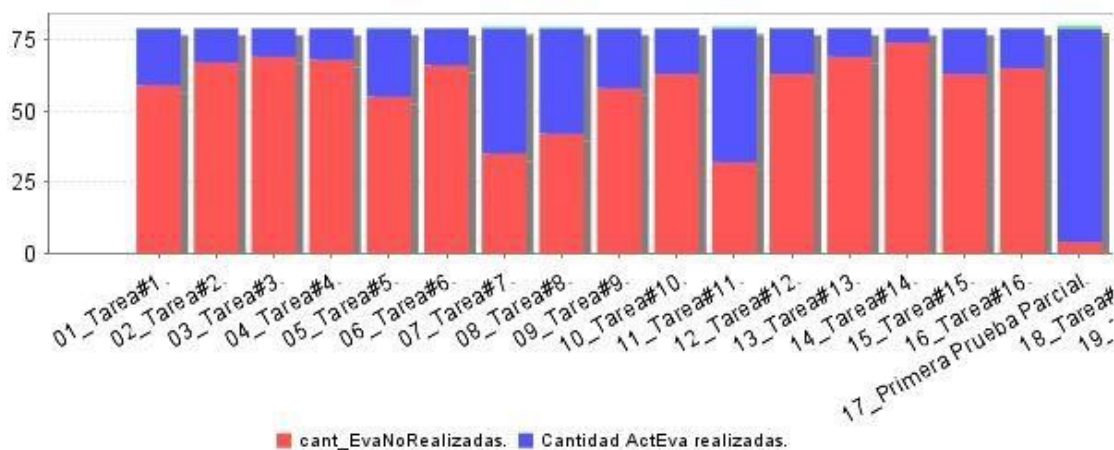


Figura 53 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación.

1.9 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo.

Esta vista muestra la cantidad de estudiantes que se evaluaron en cada actividad de evaluación realizada y la cantidad que no lo hicieron. La información se muestra desglosada por grupo.

- Consulta MDX

```

MDX Query Editor
select {[Measures].[cant_EvaNoRealizadas], [Measures].[cant_actEva_realizadas]} ON COLUMNS,
  Crossjoin({[actEval.actEva_todas].[nombreActEval].Members}, {[usuario.usuarios_todosxg].[4301],
[usuario.usuarios_todosxg].[4302], [usuario.usuarios_todosxg].[4303], [usuario.usuarios_todosxg].[4304]}) ON ROWS
from [cuboTesisMoodleEvaluacion]
    
```

- Resultado (sólo una parte)

| | | Medidas | |
|-------------------------|---------------------------|------------------------|------------------------------|
| Todas las ActEvaluacion | Todos los UsuariosXGrupos | ● cant_EvaNoRealizadas | ● Cantidad ActEva realizadas |
| 01_Tarea#1 | + 4301 | 18 | 2 |
| | + 4302 | 15 | 1 |
| | + 4303 | 4 | 15 |
| | + 4304 | 16 | 2 |
| 02_Tarea#2 | + 4301 | 19 | 1 |
| | + 4302 | 16 | 0 |
| | + 4303 | 10 | 9 |
| | + 4304 | 16 | 2 |
| 03_Tarea#3 | + 4301 | 20 | 0 |
| | + 4302 | 16 | 0 |
| | + 4303 | 11 | 8 |
| | + 4304 | 16 | 2 |

Tabla 68 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo (sólo una parte).

- Gráfica Alternativa (sólo una parte)

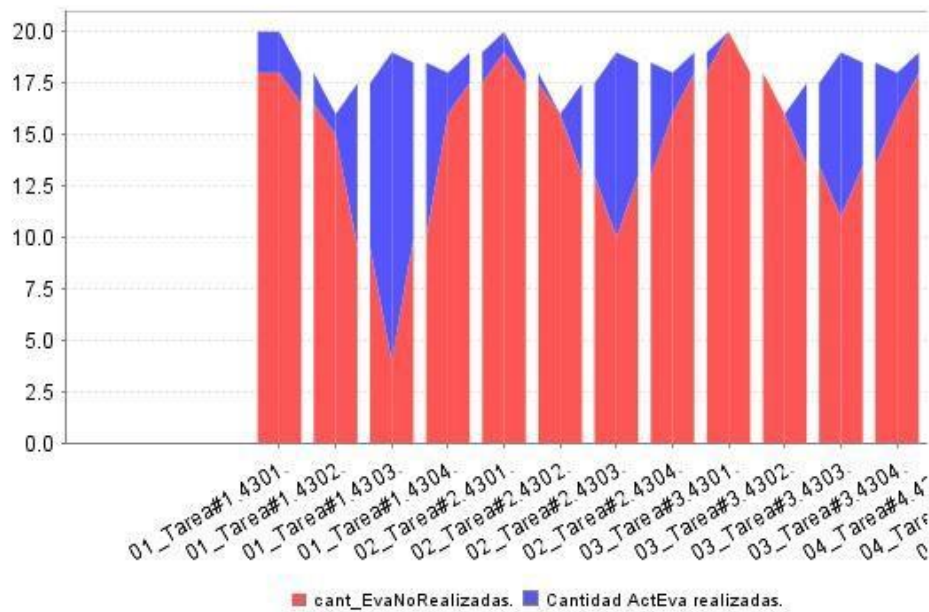
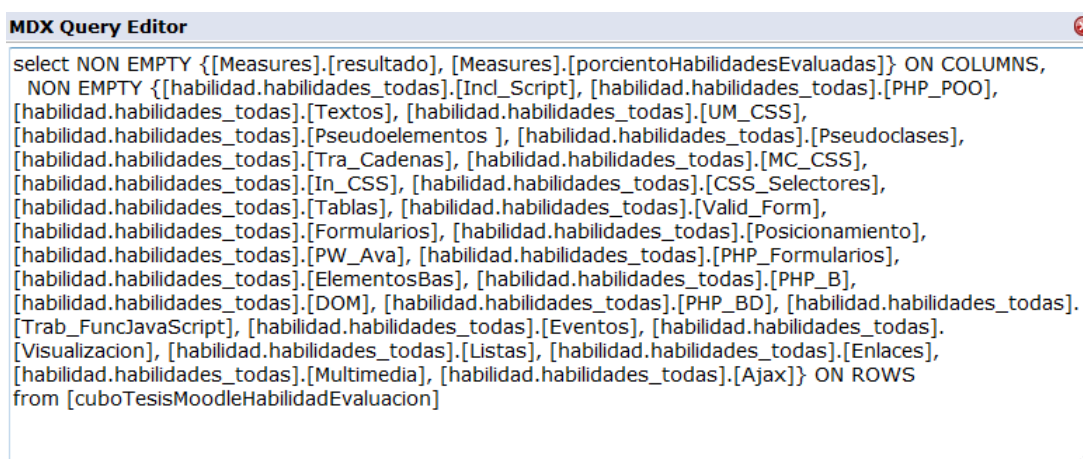


Figura 54 Estudiantes que se evaluaron y que dejaron de evaluarse por cada actividad de evaluación, desplegados por grupo (sólo una parte).

1.10 Resultado promedio de cada habilidad y el porcentaje de evaluación de dicha habilidad.

Esta vista muestra el resultado promedio que se ha obtenido en cada habilidad, la cual se obtiene de todas las evaluaciones registradas en esa habilidad. Esta vista muestra el porcentaje de evaluación de dicha habilidad: la cantidad de veces que dicha habilidad se evaluó dividido entre la cantidad de ocasiones que debió evaluarse. La información brindada por la actual vista puede sugerir que entre más se evaluó una habilidad mejor resultado se obtuvo en ella.

- Consulta MDX



```
select NON EMPTY {[Measures].[resultado], [Measures].[porcentajeHabilidadesEvaluadas]} ON COLUMNS,
NON EMPTY {[habilidad.habilidades_todas].[Incl_Script], [habilidad.habilidades_todas].[PHP_POO],
[habilidad.habilidades_todas].[Textos], [habilidad.habilidades_todas].[UM_CSS],
[habilidad.habilidades_todas].[Pseudoelementos ], [habilidad.habilidades_todas].[Pseudoclasas],
[habilidad.habilidades_todas].[Tra_Cadenas], [habilidad.habilidades_todas].[MC_CSS],
[habilidad.habilidades_todas].[In_CSS], [habilidad.habilidades_todas].[CSS_Selectores],
[habilidad.habilidades_todas].[Tablas], [habilidad.habilidades_todas].[Valid_Form],
[habilidad.habilidades_todas].[Formularios], [habilidad.habilidades_todas].[Posicionamiento],
[habilidad.habilidades_todas].[PW_Ava], [habilidad.habilidades_todas].[PHP_Formularios],
[habilidad.habilidades_todas].[ElementosBas], [habilidad.habilidades_todas].[PHP_B],
[habilidad.habilidades_todas].[DOM], [habilidad.habilidades_todas].[PHP_BD], [habilidad.habilidades_todas].
[Trab_FuncJavaScript], [habilidad.habilidades_todas].[Eventos], [habilidad.habilidades_todas].
[Visualizacion], [habilidad.habilidades_todas].[Listas], [habilidad.habilidades_todas].[Enlaces],
[habilidad.habilidades_todas].[Multimedia], [habilidad.habilidades_todas].[Ajax]} ON ROWS
from [cuboTesisMoodleHabilidadEvaluacion]
```

- Resultado

| | Medidas | |
|-----------------------|-------------|--------------------------|
| | ● Resultado | ● % Habiidades Evaluadas |
| Todas las Habilidades | | |
| Incl_Script | 3,608 | 59,49% |
| PHP_POO | 3,468 | 61,39% |
| Textos | 3,139 | 60,13% |
| UM_CSS | 3,099 | 48,73% |
| Pseudoelementos | 3,063 | 57,38% |
| Pseudoclasses | 3,008 | 56,12% |
| Tra_Cadenas | 2,937 | 57,59% |
| MC_CSS | 2,924 | 43,34% |
| In_CSS | 2,919 | 35,19% |
| CSS_Selectores | 2,864 | 41,59% |
| Tablas | 2,854 | 54,43% |
| Valid_Form | 2,842 | 57,59% |
| Formularios | 2,823 | 54,43% |
| Posicionamiento | 2,823 | 37,34% |
| PW_Ava | 2,765 | 33,03% |
| PHP_Formularios | 2,706 | 35,70% |
| ElementosBas | 2,658 | 30,79% |
| PHP_B | 2,656 | 33,92% |
| DOM | 2,633 | 44,30% |
| PHP_BD | 2,576 | 47,47% |
| Trab_FuncJavaScript | 2,541 | 32,91% |
| Eventos | 2,532 | 32,91% |
| Visualizacion | 2,5 | 23,42% |
| Listas | 2,367 | 15,19% |
| Enlaces | 2,24 | 15,19% |
| Multimedia | 2,228 | 12,66% |
| Ajax | 2,049 | 4,93% |

Tabla 69 Resultado promedio de cada habilidad y el porcentaje de evaluación de dicha habilidad.

- Gráfica Alternativa

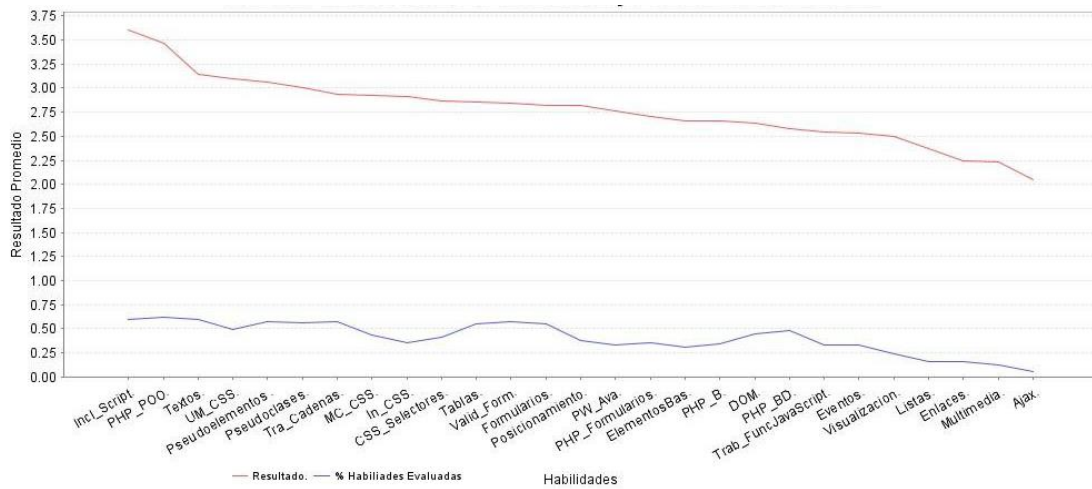


Figura 55 Resultado promedio de cada habilidad y el porcentaje de evaluación de dicha habilidad.

2 Reportes

2.1 Notas Finales de los estudiantes agrupados por grupo

| Grupo 4301 | |
|------------|--|
| Estudiante | Nota Final |
| Grupo | 3,35 ⇒ Resultado promedio del grupo |
| | 4 |
| | 2 |
| | 2 |
| | 4 |
| | 4 |
| | 4 |
| | 4 |
| | 4 |
| | 2 |
| | 4 |
| | 4 |

Tabla 70 Notas Finales de los estudiantes agrupados por grupo

2.2 Notas de las actividades de evaluación en clase de un grupo determinado

| Estudiantes | Eva. Int. Clase | T1 | T5 | T7 | T8 | T11 | T12 | T15 |
|-------------|-----------------|-------|------|-------|-------|-------|-------|-------|
| Todos | 2,762 | 2,481 | 2,81 | 3,317 | 3,076 | 3,481 | 2,582 | 2,481 |
| Grupo | 3,004 | 2,3 | 2,45 | 3,2 | 3,3 | 4,15 | 2,7 | 2,3 |
| | 3,5 | 2 | 2 | 5 | 5 | 5 | 2 | 2 |
| | 3,25 | 5 | 2 | 2 | 2 | 5 | 5 | 2 |
| | 2,5 | 5 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 3,167 | 2 | 5 | 2 | 2 | 4 | 5 | 5 |
| | 3 | 2 | 2 | 3 | 2 | 5 | 4 | 2 |
| | 3,333 | 2 | 2 | 3 | 4 | 5 | 2 | 2 |

Tabla 71 Notas de las actividades de evaluación en clase de un grupo determinado

2.3 Notas de la Primera Prueba Parcial

| Grupo | Matr. | 2 ptos. | 3 ptos. | 4 ptos. | 5 ptos. | Aus. | Apro. | %Prom. | %Cal. |
|----------|-------|---------|---------|---------|---------|------|-------|--------|--------|
| Todos | 79 | 24 | 39 | 9 | 7 | 4 | 55 | 69,62 | 20,253 |
| 4301 | 20 | 7 | 9 | 3 | 1 | 1 | 13 | 65 | 20 |
| 4302 | 16 | 4 | 11 | 1 | 0 | 0 | 12 | 75 | 6,25 |
| 4303 | 19 | 3 | 9 | 3 | 4 | 0 | 16 | 84,211 | 36,842 |
| 4304 | 18 | 6 | 8 | 2 | 2 | 3 | 12 | 66,667 | 22,222 |
| Arrastre | 6 | 4 | 2 | 0 | 0 | 0 | 2 | 33,333 | 0 |

Gráficas alternativas

Resultados X Grupos

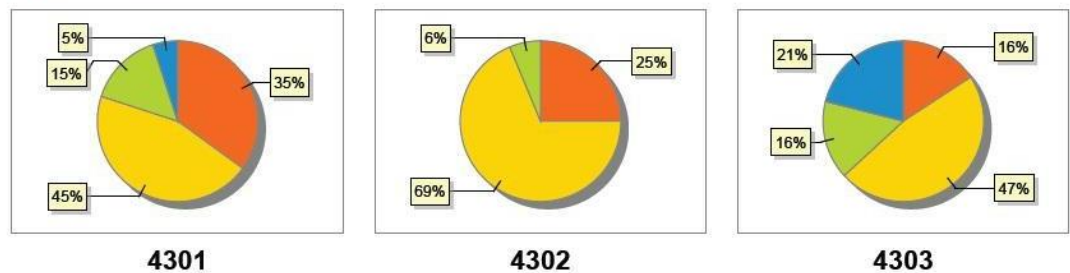


Figura 56 Notas de la Primera Prueba Parcial

3 Cuadro de Mandos Integrales

A continuación se muestran las secciones de los cuadros de mandos integrales que muestran información particular de cada estudiante. Cada cuadro de mando está dividido en 8 paneles. El primero refleja los datos para

identificar al estudiante y los restantes brindan información sobre su participación y resultados en la asignatura

3.1 Panel de Identificación

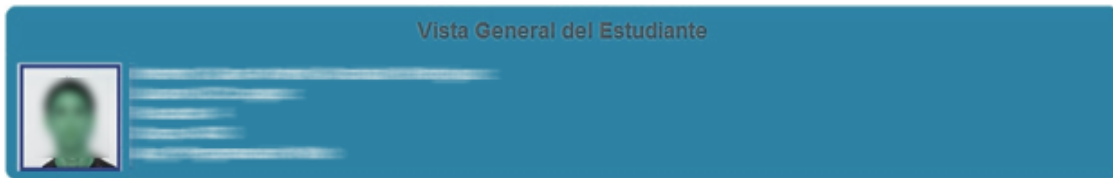


Figura 57 Ejemplo de panel de identificación

3.2 Panel de accesos semanales al sistema

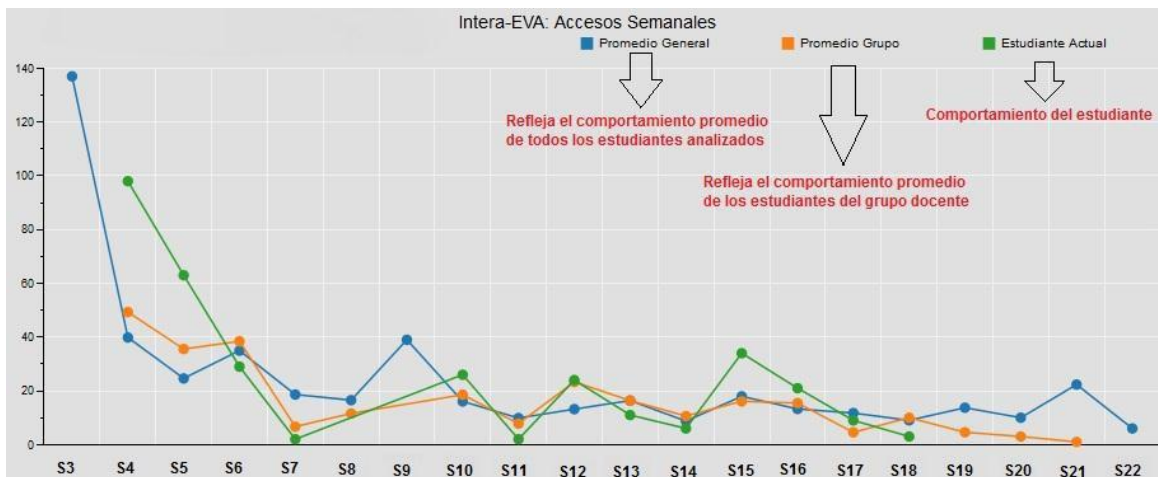


Figura 58 Ejemplo de panel de accesos semanales al sistema

3.3 Panel de accesos por hora del día

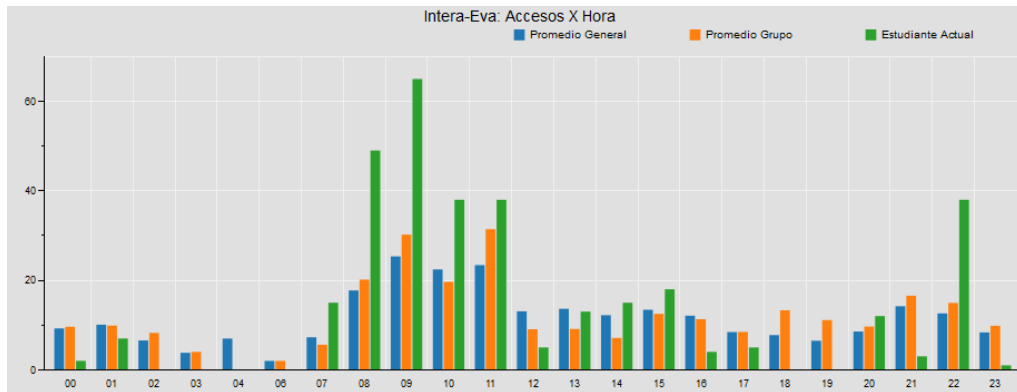


Figura 59 Ejemplo de panel de accesos por hora del día

3.4 Panel de áreas de accesos

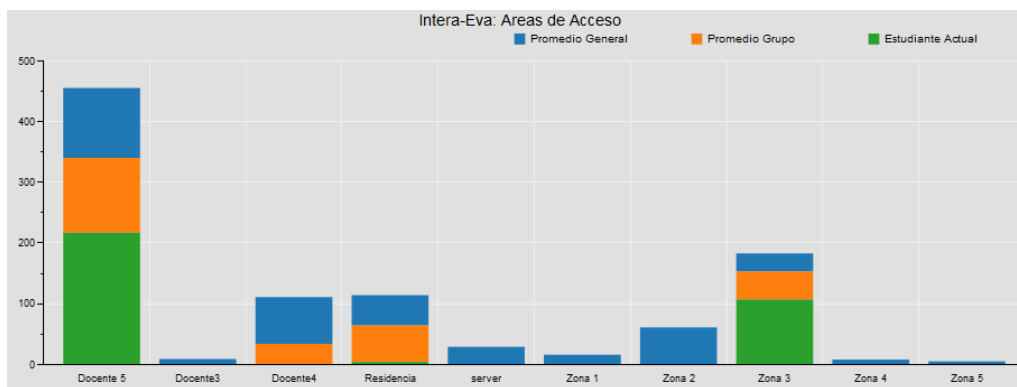


Figura 60 Ejemplo del panel de accesos por áreas

3.5 Panel de módulos de acceso

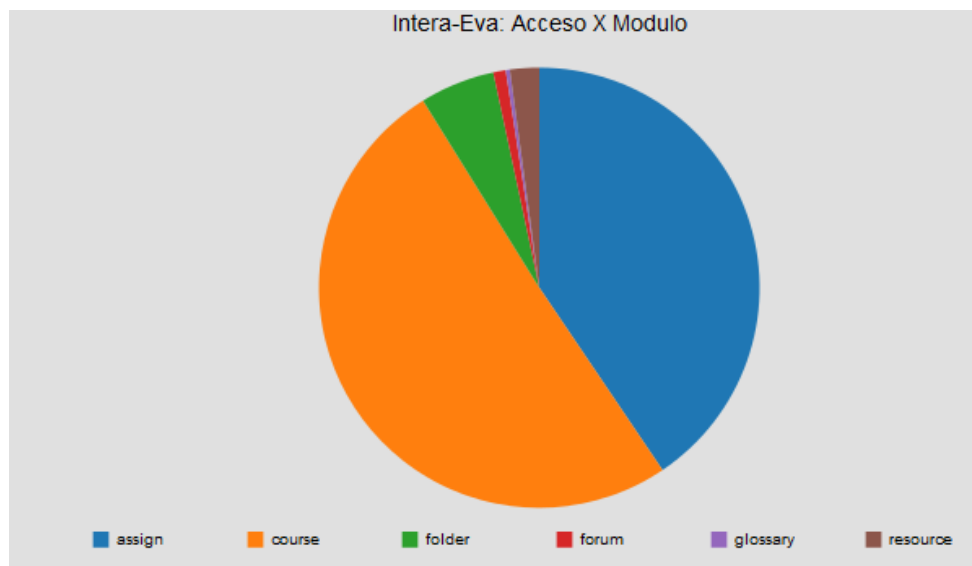


Figura 61 Ejemplo del panel de módulos de acceso

3.6 Panel de porcentaje de evaluaciones realizadas

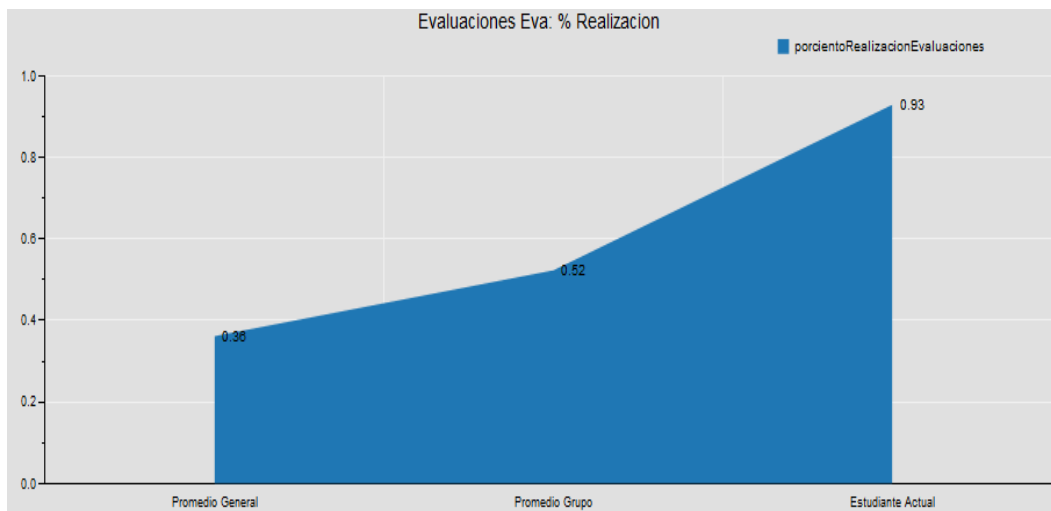


Figura 62 Panel de porcentaje de evaluaciones realizadas

3.7 Panel de resultado en principales evaluaciones

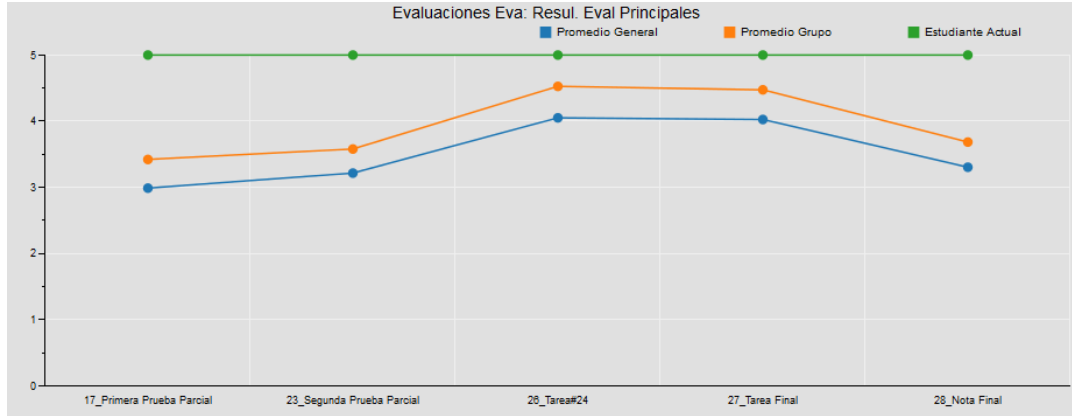


Figura 63 Panel de resultado en principales evaluaciones

3.8 Panel de resultado en habilidades evaluadas

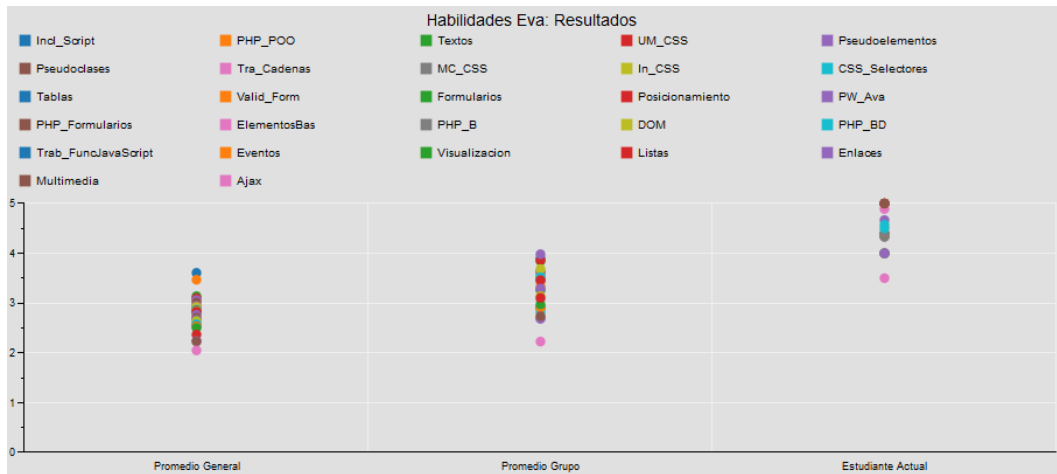


Figura 64 Panel de resultado en habilidades evaluadas