



UNIVERSIDAD PRIVADA DEL NORTE
Laureate International Universities

FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA DE SISTEMAS

DESARROLLO DE UN DATA MART EN EL ÁREA DE
ADMINISTRACIÓN Y FINANZAS DE LA MUNICIPALIDAD
DISTRITAL DE LOS BAÑOS DEL INCA

TESIS
PARA OPTAR EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS

AUTOR:
Bach. MELSI OCAS TERRONES

ASESOR:
Ing. ALEJANDRO HUGO PÉREZ QUIROZ

CAJAMARCA – PERÚ
2012

COPYRIGHT ©2012 by
Melsi Ocas Terrones
Todos los derechos reservados

UNIVERSIDAD PRIVADA DEL NORTE
Laureate International Universities

FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA DE SISTEMAS

ACEPTADA:

DESARROLLO DE UN DATA MART EN EL ÁREA DE
ADMINISTRACIÓN Y FINANZAS DE LA MUNICIPALIDAD
DISTRITAL DE LOS BAÑOS DEL INCA

AUTOR:

Bach. MELSI OCAS TERRONES

ASESOR:

Ing. ALEJANDRO HUGO PÉREZ QUIROZ

Aprobado por:

Ing. Patricia Uceda Martos
Presidenta del jurado

Ing. Manuel Asencio Zelada
Secretario del jurado

Ing. Oscar Zocón Alva
Vocal del jurado

Ing. Alejandro Hugo Pérez Quiroz
Asesor

Cajamarca, 21 de Octubre 2012

DEDICATORIA

A Dios por darme las fuerzas necesarias para seguir adelante, a mi familia por su aliento, consejos y apoyo brindado durante todos estos años.
A ellos va dedicada esta tesis.

AGRADECIMIENTO

A Dios por bendecirnos para poder hacer realidad este objetivo, a mi familia, por el amor y el apoyo incondicional la paciencia e invaluable ayuda durante este tiempo, y a mis profesores quienes siempre estuvieron prestos a brindarnos su ayuda, llenándonos de conocimientos y a todos aquellos que me brindaron su ayuda desinteresada.

LISTA DE ABREVIACIONES

OLAP	(On-Line Analytical Processing) Significa procesamiento analítico en línea
BI	Por sus siglas en inglés (Business Intelligence)
DBMS	Los sistemas de gestión de bases de datos en inglés (database management system), abreviado DBMS
EIS	(Executive Information Systems) Sistemas de Información para Directivos)
DSS	(Decision Support Systems) Sistemas de Ayuda a la toma de Decisiones
KLC	(Kimball Life Cycle) Significa ciclo de vida de Kimball
TI	Tecnologías de información (TI)
DW	Data Warehouse (DW)
BD	Base de Datos (cuya abreviatura es <i>BD</i>)
KPI	(Key Performance Indicator) Representa a los indicadores clave para la medición de rendimientos.

PRESENTACIÓN

Señores Miembros del Jurado:

De conformidad y cumpliendo lo estipulado en el Reglamento de Grados y Títulos de la Facultad de Ingeniería de la Universidad Privada del Norte, para Optar el Título Profesional de Ingeniero de Sistemas, pongo a vuestra consideración la presente Proyecto intitulado:

“DESARROLLO DE UN DATA MART EN EL ÁREA DE ADMINISTRACIÓN Y FINANZAS DE LA MUNICIPALIDAD DISTRITAL DE LOS BAÑOS DEL INCA”

El presente proyecto ha sido desarrollado durante los meses Junio, Julio, Agosto, Septiembre y Octubre Del año 2012, y espero que el contenido de este estudio sirva de referencia para otras Proyectos o Investigaciones.

Bach. MELSI OCAS TERRONES

RESUMEN

El presente trabajo de tesis desarrolla un Data Mart para el apoyo al proceso de toma de decisiones del área de administración y finanzas de la Municipalidad Distrital de los Baños del Inca.

Debido a que sus sistemas actuales no soportan el manejo adecuado de grandes volúmenes de información, tienen el problema de utilizar su información para emplearla en la toma de decisiones de la institución.

Para llevar adelante el desarrollo del Data Mart se utilizó la metodología de Kimball, conformada por las siguientes etapas:

- **Planificación:** En este proceso se determina el propósito del proyecto de DW/BI, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.
- **Análisis de requerimientos:** Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo.
- **Modelado Dimensional:** El proceso de diseño comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados.
- **Diseño del sistema de Extracción, Transformación y Carga (ETL):** El sistema de Extracción, Transformación y Carga (ETL) es la base sobre la cual se alimenta el Datawarehouse. Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos.
- **Especificación y desarrollo de aplicaciones de BI:** Las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis proporcionan información útil a los usuarios.

Para concluir con el proyecto, se realizó la contrastación de la hipótesis, las conclusiones y finalizando con las recomendaciones.

ABSTRACT

The present work of thesis develops a Data Mart for the support to the process of capture of decisions of the area of administration and finance of the Municipality Distrital of the Baths of the Inca.

Due to the fact that his current systems do not support the suitable managing of big volumes of information, they have the problem of using his information to use it in the capture of decisions of the institution.

To take forward the development of the Data Mart was in use Kimball's methodology, shaped by the following stages:

- **Planning:** In this process there decides the intention of DW/BI's project, his specific aims and the scope of the same one, the principal risks and an initial approximation to the needs of information.
- **Analysis of requirements:** it is necessary to to learn so much since one could on the business, the competitors, the industry and the clients of the same one
- **Shaped Dimensional:** The process of design begins with a dimensional high-level model obtained from the prioritized processes.
- **Design of the system of Extraction, Transformation and Load (ETL):** The system of Extraction, Transformation and Load (ETL) is the base on which the Datawarehouse is fed. If the system ETL is designed adequately, it can extract the information of the systems of origin of information; apply different rules to increase the quality and consistency of the same ones.
- **Specification and development of BI's applications:** BI's applications are the visible face of the business intelligence: the reports and applications of analysis provide useful information to the users.

To conclude with the project, there was realized the contrasting of the hypothesis, the conclusions and finishing with the recommendations.

Tabla de contenido

LISTA DE ABREVIACIONES	vi
OLAP (On-Line Analytical Processing) Significa procesamiento analítico en línea.....	vi
PRESENTACIÓN	vii
CAPITULO I: INTRODUCCIÓN.....	1
I.1. Problema de Investigación.....	2
I.1.1. Realidad Problemática.....	2
I.1.2. Formulación del problema	2
I.1.3. Justificación del problema.....	2
I.1.4. Limitaciones	2
I.2. Objetivos.....	2
I.2.1. Objetivo General	2
I.2.2. Objetivos Específicos.....	3
CAPITULO II: MARCO TEÓRICO.....	4
II.1. Antecedentes	5
II.2. Bases Teóricas	7
II.3. Marco Conceptual.....	25
CAPITULO III: MATERIALES Y MÉTODOS.....	27
III.1: Tipo de investigación	28
III.1.1. Según el propósito	28
III.1.2. Según el diseño de investigación.....	28
III.2: Diseño de investigación.....	28
III.2.1. Hipótesis	28
III.2.2. Material de estudio	28
III.2.3. Diseño de Contrastación.....	28
III.2.4. Técnicas, procedimientos e instrumentos.....	29
CAPITULO IV: RESULTADOS	30
IV.1. Planeamiento	31
IV.2. Requerimientos del negocio	37
IV.3. Diseño de arquitectura.....	42
IV.4. Instalación y selección del producto	42
IV.4.1. Instalación	42
IV.4.2. Selección del producto	46
IV.5. Modelamiento dimensional	47
IV.5.1. Modelo de Base de datos Transaccional	48
IV.5.2. Modelamiento Dimensional Estrella.....	49

IV.6. Diseño Físico.....	50
IV.6.1. Modelo de Base de Datos Dimensional y Datos	50
IV.6.2 Dimensiones y Tablas Hechos	56
IV.7. ETL.....	63
IV.7.1. Extracción.....	63
IV.7.2. Transformación	64
IV.7.3. Carga	72
IV.8. Diseño de aplicaciones BI.....	75
IV.8.1. Desarrollo de Cubos Multidimensionales	75
IV.8.2. Elaboración de Reportes.....	79
IV.8.3. KPI	81
CAPITULO V: DISCUSIÓN	83
V.1. Proceso.....	84
V.2. Resultado	85
V.3. Contrastación de la Hipótesis	87
Bibliografía.....	92

ÍNDICE DE CUADROS

Cuadro 1 – Recursos Informáticos	25
Cuadro 2 – Descripción de stakeholders	34
Cuadro 3 – Requisitos Humanos - Gestión	34
Cuadro 4 – Requisitos Humanos – Tecnologías información	34
Cuadro 5 – Requisitos Materiales.....	34
Cuadro 6 – Requisitos Equipamiento - Hardware	35
Cuadro 7 – Requisitos Equipamiento - Software	36
Cuadro 8 – Requisitos Equipamiento - Comunicaciones	36
Cuadro 9 – Riesgos del proyecto	36
Cuadro 10 – Conformación de equipos	37
Cuadro 11 – Conformación de equipos	37
Cuadro 12 – Conformación de equipos	38
Cuadro 13 – Atributos dim_concepto.....	57
Cuadro 14 – Atributos dim_contrivuyente	57
Cuadro 15 – Atributos dim_fuente_financiamiento	58
Cuadro 16 – Atributos dim_partida	58
Cuadro 17 – Atributos dim_prediosyotros	59
Cuadro 18 – Atributos dim_tenencia.....	59
Cuadro 19 – Atributos dim_tiempo	60
Cuadro 20 – Atributos dim_tipo_terreno	60
Cuadro 21 – Atributos dim_uso	61
Cuadro 22 – Granularidad fact_recaudacion	61
Cuadro 23 – Medidas fact_recaudacion	61
Cuadro 24 – Granularidad fact_recaudacion_predial	62
Cuadro 25 – Medidas fact_recaudacion_predial	62
Cuadro 26 – Especificación de ETL desde bdmunicipal a municipalidaddimension	63
Cuadro 27 – Permisos y accesos de los usuarios	63

ÍNDICE DE GRÁFICOS

Figura 1 - Tareas de la metodología de Kimball	17
Figura Tabla 1: matriz de procesos/dimensiones (bus matrix)2.....	19
Figura 2 – Diagrama de flujo del proceso dimensional de Kimball (Mundy & Thornthwaite 06)	20
Figura 3 –Lista de atributos (Mundy & Thornthwaite)	22
Figura 4 -Organigrama de la Municipalidad Distrital de Baños del Inca	33
Figura 5 –Figura de Cuadro de Dimensiones y Jerarquías	41
Figura 6 –Figura de Cuadro de Dimensiones y Medidas	41
Figura 7 – Diseño Arquitectura	42
Figura 8 - Configuración de Mysql	43
Figura 9 – Ejecución Pentaho Data Integration	44
Figura 10 –Pantalla de inicio de Pentaho Data Integration	44
Figura 11 – Pantalla de inicio de Pentaho Workbench	45
Figura 12 – Ejecución Pentaho Report Designer	45
Figura 13 – Pantalla de inicio de Pentaho Report Designer	46
Figura 14 – Diseño de la Base de Datos Transaccional	48
Figura 15 – Modelamiento Dimensional Estrella	49
Figura 16 – Modelo Dimensional Esquema municipalidaddimension	50
Figura 17 – Transformación de Dim_fuente_financiamiento	64
Figura 18 - Transformación de Dim_uso	64
Figura 19 - Transformación de Dim_partida	65
Figura 20 - Transformación de Dim_contribuyente	65
Figura 21 - Transformación de Dim_tenencia	65
Figura 22 - Transformación de Dim_tipo_terreno	66
Figura 23 - Transformación de Dim_concepto	66
Figura 24 - Transformación de Dim_prediosyotros	67
Figura 25 - Transformación de Dim_tiempo	67
Figura 26 – Transformación de Dim_recaudacion	69
Figura 27 – Transformación de Dim_recaudacion_predial	70
Figura 28 – Carga de Dimensiones	72
Figura 29 – Resultados de la carga de dimensiones	72
Figura 30 – Carga de tabla hechos dim_recaudacion	73
Figura 31 – Resultados de la carga de tabla hechos dim_recaudacion	73
Figura 32 – Carga de tabla hechos dim_recaudacion_predial	74
Figura 33 – Resultados de la carga de tabla hechos dim_recaudacion_predial	74
Figura 34 – Diseño del CUBO_RECAUDACION_PREDIAL	75
Figura 35 – Diseño del CUBO_RECAUDACION	76
Figura 36 - Página principal User Console Login	76
Figura 37 – Reporte de contribuyentes generado desde Cubo Recaudacion Predial	77
Figura 38 – Reporte de pagos de predios por año	77
Figura 39 – Reporte de contribuyentes y montos cancelados	78
Figura 40 – Reportes de Contribuyentes	79
Figura 41 – Reporte de Fuentes de Financiamiento	80
Figura 42 – Dashboard de Cartera de Predios	81
Figura 43 – Dashboard de Montos Cancelados de Predios	82
Figura 44 – Dashboard de Montos Cancelados de Recaudaciones	82
Figura 45 – Proceso Administrativo	84

Figura 46 – Grafico comparativo de resultados de la Pre – test y Post - test	86
Figura 47 – Resultado de la prueba nro. Reportes por mes	88
Figura 48 – Resultado de la prueba nro. de toma de decisiones por mes	89

CAPITULO I: INTRODUCCIÓN

I.1. Problema de Investigación

I.1.1. Realidad Problemática

Las operaciones del negocio en el área de Administración y Finanzas para la Municipalidad distrital de los Baños del Inca no muestran información procesada, únicamente se trabajan reportes para el área operativa que muchas veces conllevan a reprocesamiento por otras áreas, a fin de poder brindar información consolidada que dé respaldo a los procesos de toma de decisiones de la empresa, incrementando costos de tiempo, humano, entre otros para poder obtener los resultados.

I.1.2. Formulación del problema

¿En qué medida el desarrollo de un data mart ayudará a la Municipalidad Distrital de los Baños del Inca a elevar la eficiencia en la toma de decisiones en el área de Administración y Finanzas para el periodo 2012?

I.1.3. Justificación del problema

La presente investigación se justifica dado que las instituciones en la actualidad almacenan toda su de información en las bases de datos, sin embargo no se suele contar con una mayor accesibilidad a la información. Por lo tanto, es urgente desarrollar un data mart que permita un mejor control de la información para la gestión de la institución y así mismo incrementar los conocimientos para el uso de herramientas de gestión en el nivel gerencial de la organización.

I.1.4. Limitaciones

El presente proyecto se desarrollará para el área de Administración y Finanzas de la Municipalidad Distrital De Los Baños Del Inca.

La delimitación temporal que tendrá el desarrollo de la investigación está comprendida desde Junio a Octubre.

I.2. Objetivos

I.2.1. Objetivo General

Desarrollar un Data Mart para la Municipalidad distrital de los Baños del Inca en el área de Administración y Finanzas para que contribuya a elevar la eficiencia en la de toma de decisiones en la institución para el periodo 2012.

I.2.2. Objetivos Específicos

- Recolectar y analizar los requerimientos para el análisis de la información mediante fichas de observación.
- Diseñar la base de datos dimensional del área de Administración y Finanzas.
- Diseñar cubos y reportes necesarios para el mejor estudio de los datos.
- Demostrar que la utilización de un Data Mart permite optimizar la toma de decisiones.

CAPITULO II: MARCO TEÓRICO

II.1. Antecedentes

- Jair Pérez Pérez plantea en su tesis “Arquitectura técnica del data warehouse” (Guatemala, Marzo de 2006), que una estructuración diseñada a la medida en un manejador de base de datos, contribuye a mejorar los procesos administrativos del área de producción en una empresa productora y comercializadora de artículos de oficina.
- Eddy Fernández Ochoa en su tesis “Análisis, diseño e implementación de un data mart de clientes para el área de marketing de una entidad aseguradora”, en la Pontificia Universidad Católica del Perú, Lima, 08 de diciembre de 2009. Concluye que :
 - Es indispensable contar con los objetivos de la empresa a fin de establecer de manera correcta los indicadores que puedan reflejar una métrica válida del avance de lo planeado, solo así el Data mart constituirá una herramienta válida para el análisis cuantitativo y cualitativo.
 - La construcción de un modelo de datos OLAP permite realizar consultas a partir de información previamente procesada con lo cual se da flexibilidad al usuario a realizar diferentes consultas pre elaboradas. Sin embargo, es importante resaltar la relación que existe entre el desempeño de la consulta y la flexibilidad del análisis. El performance de la consulta se ve favorecida con la información pre procesada, sin embargo esto limita la capacidad de análisis ya que se tienen variables ya definidas que para ser modificadas requieren un reproceso.
 - La capa de datos BDS perteneciente a la arquitectura del presente proyecto aísla el entorno operacional para solo enfocarse en información consolidada que ayude a la toma de decisiones.
 - La capa de datos ODS perteneciente a la arquitectura del presente proyecto integró de manera exitosa información proveniente de distintos sistemas transaccionales en un solo repositorio que cuenta con un modelo de datos del negocio que soporta consultas de la información a un nivel transaccional.
- Paul Eduardo Caicedo Lescano (Quito,2010) comparo dos herramientas, Oracle y Cognos para inteligencia de negocios en su tesis “Comparación de las herramientas Cognos 8 BI y Oracle BI utilizadas en la elaboración de sistemas de toma de decisiones”, nos presenta las siguientes conclusiones:
 - Las herramientas de inteligencia de negocios nos permiten una integración de los datos de diferentes orígenes, así como la agilidad en la toma de decisiones.
 - Las herramientas de inteligencia de negocios nos hacen menos dependientes del área de tecnología.
 - Con el uso de estas herramientas de inteligencia de negocios los resultados son rápidos y visibles.

- La utilización de reportes mediante las herramientas de inteligencia de negocios se lo hace para tener detalle de la información.
- En la Tesis de Evelyn Roxana Torres Catalán (Guatemala, 2006) titulada “Arquitectura técnica del data warehouse”, se ha llegado a las siguientes conclusiones:
 - El beneficio obtenido del Data Warehouse es proporcional al grado de relación que este posea con las reglas, normas, métricas y metodologías de la empresa. Esto quiere decir que cuanto más definidos se encuentren los requerimientos y cuando más estén relacionados con las necesidades de la empresa, se obtendrá un mejor desempeño de la arquitectura que se diseñe en base a los lineamientos definidos, en los requerimientos.
 - El éxito de un proyecto de Data Warehouse no es representado por la tecnología con la que se apoye, si no con la calidad del diseño que se elabore, ya que, la tecnología, solamente representa un eslabón en la cadena de la elaboración de este tipo de sistemas y no así el punto clave del sistema.
 - El uso de herramientas de tecnología tales como Warehouse Builder y Cognos 8 BI, proporcionan un aumento en el rendimiento de las actividades involucradas al momento de crear un Data Warehouse, debido a que proporcionan interfaces accesibles a los usuarios finales, sobre acciones que presentan complejidad avanzada.

II.2. Bases Teóricas

1. Inteligencia de negocios:

El presente muestra la importancia de la inteligencia de negocios, la cual considera que “Uno de los recursos más importantes de cualquier organización es la información, gracias a ella se pueden tomar decisiones que permiten mejorar el rumbo de las empresas e instituciones. Hay que saber “escuchar” lo que nos dicen los datos que día con día genera una organización, si estamos interesados en descubrir las nuevas necesidades de nuestros usuarios, o bien, deseamos mejorar nuestros procesos.

La Inteligencia de Negocios es un concepto que hace referencia a las técnicas de análisis de datos destinados a encontrar información útil para la toma de decisiones, incluido el conjunto del software que aporta las interfaces y funciones necesarias que apoyan dicho proceso.

De esta forma, los sistemas analíticos o de toma de decisiones recuperan los datos de las operaciones diarias de la organización, generados por los sistemas transaccionales, y los procesan, con objeto de tenerlos disponibles para los tomadores de decisiones.

El desarrollo de sistemas analíticos es un tema complejo, con algunas décadas de evolución, y con aportaciones de áreas como la inteligencia artificial, que ha incorporado técnicas como la minería de datos (descubrimiento de patrones), el data warehousing (análisis multidimensional de información) y el OLAP”. (Romero, enterate, 2004)

Otra definición que nos ayuda a entender los conceptos de base de datos, es la que propone Josep Lluís Cano el cual define que: “Business Intelligence es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones. El primero que acuñó el término fue Howard Dresner que, cuando era consultor de Gartner, popularizó Business Intelligence o BI como un término paraguas para describir un conjunto de conceptos y métodos que mejoraran la toma de decisiones, utilizando información sobre qué había sucedido (hechos). Mediante el uso de tecnologías y las metodologías de Business Intelligence pretendemos convertir datos en información y a partir de la información ser capaces de descubrir conocimiento.

Para definir BI partiremos de la definición del glosario de términos de Gartner:

“BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.

El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar los cambios.

Las áreas incluyen clientes, proveedores, productos, servicios y competidores.”

Pero descompongamos detalladamente esta definición:

- Proceso interactivo: al hablar de BI estamos suponiendo que se trata de un análisis de información continuado en el tiempo, no sólo en un momento puntual. Aunque evidentemente este último tipo de análisis nos puede aportar valor, es incomparable con lo que nos puede aportar un proceso continuado de análisis de información, en el que por ejemplo podemos ver tendencias, cambios, variabilidades, etc.
- Explorar: En todo proyecto de BI hay un momento inicial en el que por primera vez accedemos a información que nos facilita su interpretación. En esta primera fase, lo que hacemos es “explorar” para comprender qué sucede en nuestro negocio; es posible incluso que descubramos nuevas relaciones que hasta el momento desconocíamos.
- Analizar: Pretendemos descubrir relaciones entre variables, tendencias, es decir, cuál puede ser la evolución de la variable, o patrones. Si un cliente tiene una serie de características, cuál es la probabilidad que otro con similares características actúe igual que el anterior.
- Información estructurada y datawarehouse: La información que utilizamos en BI está almacenada en tablas relacionadas entre ellas. Las tablas tienen registros y cada uno de los registros tiene distintos valores para cada uno de los atributos. Estas tablas están almacenadas en lo que conocemos como datawarehouse o almacén de datos. Más adelante lo definiremos con mayor precisión, pero se trata de una base de datos en las que se almacenan dichas tablas.
- Área de análisis: Todo proyecto de BI debe tener un objeto de análisis concreto. Nos podemos centrar en los clientes, los productos, los resultados de una localización, etc. Que pretendemos analizar con detalle y con un objetivo concreto: por ejemplo, la reducción de costes, el incremento de ventas, el aumento de la participación de mercado, el ajuste de previsiones de venta, el cumplimiento los objetivos de venta presupuestados, etc.

Comunicar los resultados y efectuar los cambios: Un objetivo fundamental del BI es que, una vez descubierto algo, sea comunicado a aquellas personas que tengan que realizar los cambios pertinentes en la organización para mejorar nuestra competitividad.

El origen de la Business Intelligence va ligado a proveer acceso directo a la información a los usuarios de negocio para ayudarles en la toma de decisiones, sin intervención de los departamentos de Sistemas de Información.

En el año 2005, en las conferencias del Simposio de Gartner en Australia, se llevó a cabo una encuesta informal a los asistentes de distintas presentaciones (150 técnicos y usuarios de negocio). La encuesta incluía 14 opciones de posibles definiciones de Business Intelligence e incluso permitía que los propios asistentes expresaran su propia definición. El 43% de ellos definía Business Intelligence como:

“El uso de información que permite a las organizaciones dirigir de la mejor forma, decidir, medir, gestionar y optimizar el alcance de la eficiencia y los resultados financieros.”

Un 16% afirmaban que Business Intelligence es: “La habilidad de proporcionar datos/información en un proceso (o aplicación) funcional para permitir mostrar un hecho específico y que en ese contexto puede originar una acción.”

Otro 16% veía Business Intelligence como: “El acceso al análisis de fuentes de información cuantitativa que permita mostrar a sus usuarios alinear mejor a las personas y los procesos con los objetivos del negocio.”

Menos del 5% veía Business Intelligence cómo: “Herramientas y tecnologías (reporting8 y minería de datos) que ayudan a los analistas a trabajar la información.”

Los dos hechos más críticos de la encuesta fueron que los asistentes entendían que el valor de Business Intelligence iba más allá la distribución de información y que está fuertemente relacionado con la consecución de los objetivos de negocio.

Una definición más amplia es la que proponen en The datawarehouse Institute: “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 end10’, consultas, informes, análisis y las herramientas para mostrar información (estas son las herramientas de BI) y los procesos en el ‘front end’.” (Cano, 2007)

2. Base de Datos:

El siguiente concepto comprende de que se tratan los sistemas de base de datos, De acuerdo con C. J. Date: “es una colección de datos integrados, con redundancia controlada y con una estructura que refleje las interrelaciones y restricciones existentes en el mundo real; los datos que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción, únicas para cada tipo de dato, han de estar almacenadas junto con los mismos. Los procedimientos de actualización y recuperación, comunes, y bien determinados, habrán de ser capaces de conservar la integridad, seguridad y confidencialidad del conjunto de datos”. (DATE, Introducción a los Sistemas de bases de datos, 2001)

Otra definición que nos ayuda a entender las propiedades de una base de datos, de acuerdo con Ramez Elmasri Y Shamkant B. Navathe consideran que: “Las bases de datos y la tecnología de bases de datos tienen mucha culpa del uso creciente de los computadores. Es justo decir que las bases de datos juegan un papel fundamental en la mayoría de las áreas en las que se utilizan computadores, como en el ámbito empresarial, en el comercio electrónico, ingeniería, medicina, justicia, educación y bibliotecas. La expresión base de datos se utiliza tan a menudo que empezaremos por definir su significado. Nuestra primera definición es muy general.

Una base de datos es una colección de datos relacionados. Con la palabra datos nos referimos a los hechos (datos) conocidos que se pueden grabar y que tienen un significado implícito. Por ejemplo, piense en los nombres, números de teléfono y direcciones de las personas que conoce. Puede tener todos estos datos grabados en un

libro de direcciones indexado o los puede tener almacenados en el disco duro de un computador mediante una aplicación como Microsoft Access o Excel. Esta colección de datos relacionados con un significado implícito es una base de datos.

La definición anterior de base de datos es muy genérica; por ejemplo, podemos pensar que la colección de palabras que compone esta página de texto es una colección de datos relacionados y que, por tanto, constituye una base de datos. No obstante, el uso común del término base de datos es normalmente más restringido.

Una base de datos tiene las siguientes propiedades implícitas:

- Una base de datos representa algún aspecto del mundo real, lo que en ocasiones se denomina minimundo o universo de discurso (UoD, Universe 01 discollrse). Los cambios introducidos en el minimundo se reflejan en la base de datos.
- Una base de datos es una colección de datos lógicamente coherente con algún tipo de significado inherente.
- No es correcto denominar base de datos a un surtido aleatorio de datos.
- Una base de datos se diseña, construye y rellena con datos para un propósito específico. Dispone de un grupo pretendido de usuarios y algunas aplicaciones preconcebidas en las que esos usuarios están interesados.

En otras palabras, una base de datos tiene algún origen del que se derivan los datos, algún grado de interacción con eventos del mundo real y un público que está activamente interesado en su contenido. Los usuarios finales de una base de datos pueden efectuar transacciones comerciales (por ejemplo, un cliente que compra una cámara) o se pueden producir unos eventos (por ejemplo, un empleado tiene un hijo) que provoquen un cambio en la información almacenada en la base de datos. Al objeto de que una base de datos sea en todo momento precisa y fiable, debe ser un reflejo exacto del minimundo que representa; por consiguiente, en la base de datos deben reflejarse los cambios tan pronto como sea posible.

Una base de datos puede ser de cualquier tamaño y complejidad. Por ejemplo, la lista de nombres y direcciones a la que nos referíamos anteriormente puede constar de únicamente unos cuantos cientos de registros, cada uno de ellos con una estructura sencilla. Por el contrario, el catálogo computarizado de una gran biblioteca puede contener medio millón de entradas organizadas en diferentes categorías (por los apellidos del autor principal, por el tema, por el título del libro), y cada categoría ordenada alfabéticamente. El Departamento de tesorería de Estados Unidos (IRS, Internal Revenue Service) mantiene una base de datos de un tamaño y complejidad aún mayores para supervisar los formularios de impuestos presentados por los contribuyentes americanos.

Si asumimos que hay 100 millones de contribuyentes y que cada uno presenta una media de cinco formularios con aproximadamente 400 caracteres de información por cada uno, tenemos una base de datos de $100 \times 10^6 \times 400 \times 5$ caracteres (bytes) de información. Si el IRS conserva las tres últimas declaraciones de cada contribuyente, además de la actual tenemos una base de datos de 8×10^{11} bytes (800 gigabytes). Esta inmensa cantidad de información debe organizarse y administrarse para que los usuarios puedan buscar, recuperar y actualizar los datos que necesiten. Amazon.com es un buen ejemplo de una gran base de datos comercial.

Contiene datos de más de 20 millones de libros, CDs, vídeos, DVDs, juegos, ropa y otros productos. La base de datos ocupa más de 2 terabytes (un terabyte es 1012 bytes de almacenamiento) y se almacena en 200 computadores diferentes (denominados servidores). Cada día acceden a Amazon.com aproximadamente 15 millones de visitantes que utilizan la base de datos para hacer compras. La base de datos se actualiza continuamente a medida que se añaden libros y otros productos nuevos al inventario, mientras que el stock se actualiza al tiempo que se tramitan las compras. Alrededor de 100 personas son las responsables de mantener actualizada la base de datos de Amazon.

Una base de datos se puede generar y mantener manualmente o estar computarizada. Por ejemplo, el catálogo de cartas de una biblioteca es una base de datos que se puede crear y mantener de forma manual. Una base de datos computarizada se puede crear y mantener con un grupo de aplicaciones escritas específicamente para esa tarea o mediante un sistema de administración de bases de datos. En este libro sólo nos ocuparemos de las bases de datos computarizadas.

Un sistema de administración de datos (DBMS, database management system) es una colección de programas que permite a los usuarios crear y mantener una base de datos. El DBMS es un sistema de software de propósito general que facilita los procesos de definición, construcción, manipulación y compartición de bases de datos entre varios usuarios y aplicaciones. Definir una base de datos implica especificar los tipos de datos, estructuras y restricciones de los datos que se almacenarán en la base de datos. La definición o información descriptiva de una base de datos también se almacena en esta última en forma de catálogo o diccionario de la base de datos; es lo que se conoce como metadatos. La construcción de la base de datos es el proceso consistente en almacenar los datos en algún medio de almacenamiento controlado por el DBMS. La manipulación de una base de datos incluye funciones como la consulta de la base de datos para recuperar datos específicos, actualizar la base de datos para reflejar los cambios introducidos en el mundo y generar informes a partir de los datos. Compartir una base de datos permite que varios usuarios y programas accedan a la base de datos de forma simultánea.

Una aplicación accede a la base de datos enviando consultas o solicitudes de datos al DBMS. Una consulta normalmente provoca la recuperación de algunos datos; una transacción puede provocar la lectura o la escritura de algunos datos en la base de datos.

Otras funciones importantes ofrecidas por el DBMS son la protección de la base de datos y su mantenimiento durante un largo periodo de tiempo. La protección incluye la protección del sistema contra el funcionamiento defectuoso del hardware o el software (caídas) y la protección de la seguridad contra el acceso no autorizado o malintencionado. Una gran base de datos típica puede tener un ciclo de vida de muchos años, por lo que el DBMS debe ser capaz de mantener el sistema de bases de datos permitiendo que el sistema evolucione según cambian los requisitos con el tiempo.

No es necesario utilizar software DBMS de propósito general para implementar una base de datos computarizada. Podríamos escribir nuestro propio conjunto de programas para crear y mantener la base de datos; en realidad, podríamos crear nuestro propio

software DBMS de propósito especial. En cualquier caso (utilicemos o no un DBMS de propósito general), normalmente tenemos que implantar una cantidad considerable de software complejo. De hecho, la mayoría de los DBMS son sistemas de software muy complejos.

Como colofón de nuestras definiciones iniciales, denominaremos sistema de bases de datos a la combinación de base de datos y software DBMS.” (Ramez Elmasri & Shamkant B. Navathe, 2007)

3. Data Mart:

El presente define el concepto de data mart, de acuerdo con C.J. Date considera que un data mart es: “Por lo general, los sistemas operacionales tienen requerimientos de rendimiento estrictos, cargas de trabajo predecibles, pequeñas unidades de trabajo y una alta utilización. Por el contrario, los sistemas de apoyo para la toma de decisiones tienen por lo general requerimientos de rendimientos variantes, cargas de trabajo impredecibles, grandes unidades de trabajo y utilización errática.

Estas diferencias pueden hacer que sea muy difícil combinar el procesamiento operacional y el de apoyo para la toma de decisiones dentro de un solo sistema, en especial con relación a la planeación de la capacidad, la administración de recursos y el perfeccionamiento del rendimiento del sistema. Por estas razones, en general los administradores de sistemas operacionales están poco dispuestos a permitir actividades de apoyo para la toma de decisiones en sus sistemas; ésta es la causa del familiar enfoque del sistema doble.

Nota: Comentamos adicionalmente que esto no siempre fue así; los primeros sistemas de apoyo para la toma de decisiones fueron ejecutados en los sistemas operacionales pero con una baja prioridad, o durante la llamada "ventana por lotes". Con recursos de computación suficientes, existen varias ventajas de este arreglo, y tal vez la más obvia es que evita todas las operaciones (posiblemente costosas) de copiado de datos, reformato y transferencia (etcétera) requeridas por el enfoque del sistema doble. De hecho, el valor de la integración de las actividades operacionales y de apoyo para la toma de decisiones está llegando a ser cada vez más reconocido. Sin embargo, más detalles de tal integración están fuera del alcance de este capítulo.

A pesar del párrafo anterior (al menos al momento de la publicación de este libro), permanece el hecho de que los datos de apoyo para la toma de decisiones necesitan por lo general ser recolectados a partir de una variedad de sistemas operacionales (a menudo dispares) y ser mantenidos en un almacén de datos propio en una plataforma independiente. Ese almacén de datos separados es un data warehouse. Data warehouses Al igual que los almacenes de datos operacionales (y los data marts, vea la siguiente subsección), un data warehouse es un tipo especial de base de datos. Al parecer, el término se originó a finales de los ochenta, aunque el concepto es de alguna manera más antiguo. La referencia define un data warehouse como "un almacén de datos orientado a un tema, integrado, no volátil y variante en el tiempo, que soporta decisiones de administración" (donde el término no volátil significa que una vez que los datos han sido insertados, no pueden ser cambiados, aunque sí pueden ser

borrados). Los data warehouses surgieron por dos razones: primero, la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo para la toma de decisiones; segundo, la necesidad de hacerlo sin afectar a los sistemas operacionales.

Por definición, las cargas de trabajo del data warehouse están destinadas para el apoyo a la toma de decisiones y por lo tanto, tienen consultas intensivas (con actividades ocasionales de inserción por lotes); asimismo, los propios data warehouses tienden a ser bastante grandes (a menudo mayores que 500GB y con una tasa de crecimiento de hasta el 50 por ciento anual). Por consecuencia, es difícil —aunque no imposible— perfeccionar el rendimiento. También puede ser un problema la escalabilidad. Contribuyen a ese problema (a) los errores de diseño de la base de datos, (b) el uso ineficiente de los operadores relacionales (que mencionamos brevemente en la sección 21.2), (c) la debilidad en la implementación del modelo relacional del DBMS, (d) la falta de escalabilidad del propio DBMS y (e) los errores de diseño arquitectónico que limitan la capacidad e imposibilitan la escalabilidad de la plataforma. Los puntos (d) y (e) están fuera del alcance de este libro; por el contrario, ya hemos explicado los puntos (a) y (b) en este capítulo, y el punto (c) se trata ampliamente en otras partes del libro.

Data marts Por lo general, los data warehouses están hechos para proporcionar una fuente de datos única para todas las actividades de apoyo para la toma de decisiones. Sin embargo, cuando los data warehouses se hicieron populares (a principio de los años noventa) pronto fue evidente que los usuarios a menudo realizaban amplias operaciones de informes y análisis de datos sobre un subconjunto relativamente pequeño de todo el data warehouse. Asimismo, era muy probable que los usuarios repitieran las mismas operaciones sobre el mismo subconjunto de datos cada vez que era actualizado. Además, algunas de esas actividades —por ejemplo, análisis de pronósticos, simulación, modelado de datos de negocios del tipo "qué pasaría si..." involucraban la creación de nuevos esquemas y datos con actualizaciones posteriores a esos nuevos datos.

De manera obvia, la ejecución repetida de tales operaciones sobre el mismo subconjunto de todo el almacén no era muy eficiente; por lo tanto, pareció obviamente una buena idea construir algún tipo de "almacén" limitado de propósito general que estuviera hecho a la medida de ese propósito. Además, en algunos casos sería posible extraer y preparar los datos requeridos directamente a partir de las fuentes locales, lo que proporcionaba un acceso más rápido a los datos que si tuvieran que ser sincronizados con los demás datos cargados en todo el data warehouse. Dichas consideraciones condujeron al concepto de data marts.

De hecho, hay alguna controversia sobre la definición precisa del término data mart. Para nuestros propósitos podemos definirlo como "un almacén de datos especializado, orientado a un tema, integrado, volátil y variante en el tiempo para apoyar un subconjunto específico de decisiones de administración". Como puede ver, la principal diferencia entre un data mart y un data warehouse es que el data mart es especializado y volátil. Por especializado queremos decir que contiene datos para dar apoyo (solamente) a un área específica de análisis de negocios; por volátil queremos decir que los usuarios pueden actualizar los datos e incluso, posiblemente, crear nuevos datos (es decir, nuevas tablas) para algún propósito.

Hay tres enfoques principales para la creación de un data mart:

- Los datos pueden ser simplemente extraídos del data warehouse; de hecho, sigue un enfoque de "divide y vencerás" sobre la carga de trabajo general de apoyo para la toma de decisiones, a fin de lograr un mejor rendimiento y escalabilidad. Por lo general, los datos extraídos son cargados en una base de datos que tiene un esquema físico que se parece mucho al sub conjunto aplicable del data warehouse; sin embargo, puede ser simplificado de alguna manera gracias a la naturaleza especializada del data mart.
- A pesar del hecho de que el data warehouse pretende proporcionar un "punto de control único", un data mart puede ser creado todavía en forma independiente (es decir, no por medio de la extracción a partir del data warehouse). Dicho enfoque puede ser adecuado si el data warehouse es inaccesible por alguna razón, digamos razones financieras, operacionales o incluso políticas (o puede ser que ni siquiera exista todavía el data warehouse; vea el siguiente punto).
- Algunas instalaciones han seguido un enfoque de "primero el data mart", donde los data marts son creados conforme van siendo necesarios y el data warehouse general es creado, finalmente, como una consolidación de los diversos data marts.

Los últimos dos enfoques sufren posibles problemas de desacoplo semántico. Los data marts independientes son particularmente susceptibles a tales problemas, debido a que no hay forma obvia de verificar los desacoples semánticos cuando las bases de datos son diseñadas en forma independiente.

Por lo general, la consolidación de data marts en data warehouses falla, a menos que (a) se construya primero un esquema lógico único para el data warehouse y (b) los esquemas para los data marts individuales se deriven después a partir del esquema del data warehouse. (Por supuesto, el esquema de este último punto puede evolucionar — suponiendo que se sigan buenas prácticas de diseño— para incluir el tema de cada nuevo data mart conforme sea necesario.)

Una nota sobre el diseño de data marts: Una decisión importante que hay que tomar en el diseño de cualquier base de datos de apoyo para la toma de decisiones es la granularidad de la misma. Aquí el término granularidad se refiere al nivel más bajo de agregación de datos que se mantendrá en la base de datos. Ahora bien, la mayoría de las aplicaciones de apoyo para la toma de decisiones requerirán tarde o temprano acceso a datos detallados y por lo tanto, la decisión será fácil para el data warehouse. Para un data mart puede ser más difícil. La extracción de grandes cantidades de datos detallados del data warehouse, y su almacenamiento en el data mart, puede ser muy ineficiente si ese nivel de detalle no se necesita con mucha frecuencia. Por otro lado, en algunas ocasiones es difícil establecer definitivamente cuál es el nivel más bajo de agregación que en realidad se necesita. En dichos casos, los datos detallados pueden ser accedidos directamente desde el data warehouse cuando se necesiten, manteniendo en el data mart los datos que de alguna manera ya fueron agregados. Al mismo tiempo, generalmente no se hace la agregación completa de los datos, debido a que las diversas formas en las que pueden ser agregados, generarían cantidades muy grandes de datos de resumen.

Un punto adicional: Debido a que los usuarios de los data marts emplean frecuentemente determinadas herramientas analíticas, el diseño físico a menudo está

determinado, en parte, por las herramientas específicas a usar. Esta circunstancia desafortunada puede conducir a "esquemas dimensionales" que no son soportados por las buenas prácticas de diseño relacional." (Date, 2011)

Otra definición sostiene que "Un Data mart es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

El Data mart es un sistema orientado a la consulta, en el que se producen procesos batch de carga de datos (altas) con una frecuencia baja y conocida. Es consultado mediante herramientas OLAP (On line Analytical Processing -Procesamiento Analítico en Línea) que ofrecen una visión multidimensional de la información. Sobre estas bases de datos se pueden construir EIS (Executive Information Systems, Sistemas de Información para Directivos) y DSS (Decision Support Systems, Sistemas de Ayuda a la toma de Decisiones). Por otra parte, se conoce como Data Mining al proceso no trivial de análisis de grandes cantidades de datos con el objetivo de extraer información útil, por ejemplo para realizar clasificaciones o predicciones." (Wikipedia, Wikipedia)

4. Modelo Dimensional:

El presente define que: "El modelado dimensional es una técnica de diseño lógico que presenta los datos de un modo estandarizado que es intuitivo para los usuarios y proporciona un acceso eficiente a la información.

La idea principal del modelado dimensional es que prácticamente toda la información de una organización puede ser representada como un hipercubo de datos de n dimensiones, donde cada celda del hipercubo contiene una medición y cada eje del cubo determina una dimensión de estudio de los datos. En la siguiente figura puede verse la representación de los datos de un Centro Hospitalario como un cubo tridimensional cuyas dimensiones representan las diferentes patologías por centro, el motivo de la espera y la fecha en las que se producen las esperas.

Por supuesto, en la mayor parte de los casos tres dimensiones no son suficientes para aportar la información necesaria. El número de dimensiones de un modelo dimensional suele variar entre 4 y 15 dimensiones, dependiendo del dominio de estudio. Los modelos con más de 15 dimensiones suelen tener dimensiones innecesarias que pueden ser combinadas entre sí, dando lugar a un número menor. Los modelos con demasiadas dimensiones complican en exceso la comprensión de los modelos por los usuarios (en la práctica es raro encontrar dominios en los que las mediciones estén afectadas por más de 15 variables independientes)." (Kimball, 1996)

Otra definición manifiesta que: "El modelo dimensional es una adaptación especializada del modelo relacional, solía representar datos en depósitos de datos, en un camino que los datos fácilmente pueden ser resumidos usando consultas OLAP. En

el modelo dimensional, una base de datos consiste en una sola tabla grande de hechos que son descritos usando dimensiones y medidas.

Una dimensión proporciona el contexto de un hecho (como quien participó, cuando y donde pasó, y su tipo). Las dimensiones se toman en cuenta en la formulación de las consultas para agrupar hechos que están relacionados. Las dimensiones tienden a ser discretas y son a menudo jerárquicas; por ejemplo, la ubicación podría incluir el edificio, el estado, y el país.

Una medida es una cantidad que describe el hecho, tales como los ingresos. Es importante que las medidas puedan ser agregados significativamente - por ejemplo, los ingresos provenientes de diferentes lugares pueden sumarse.

En una consulta (OLAP), las dimensiones son escogidas y los hechos son agrupados y añadidos juntos para crear un reporte.

El modelo dimensional a menudo es puesto en práctica sobre la cima del modelo emparentado que usa un esquema de estrella, consistiendo en una mesa que contiene los hechos y mesas circundantes que contienen las dimensiones. Dimensiones en particular complicadas podrían ser representadas usando múltiples mesas, causando un esquema de copo de nieve.

Un almacén de datos (data warehouse) puede contener múltiples esquemas de estrella que comparten tablas de dimensión, permitiéndoles para ser usadas juntas. La llegada levanta un conjunto de dimensiones estándar y es una parte importante del modelado dimensional.” (Wikipedia, Wikipedia, 2012)

5. Metodología de Kimball:

La presente muestra el ciclo de vida del proyecto de un data warehouse, la cual considera que “La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle) (Kimball et al 98, 08, Mundy & Thornthwaite 06). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos:

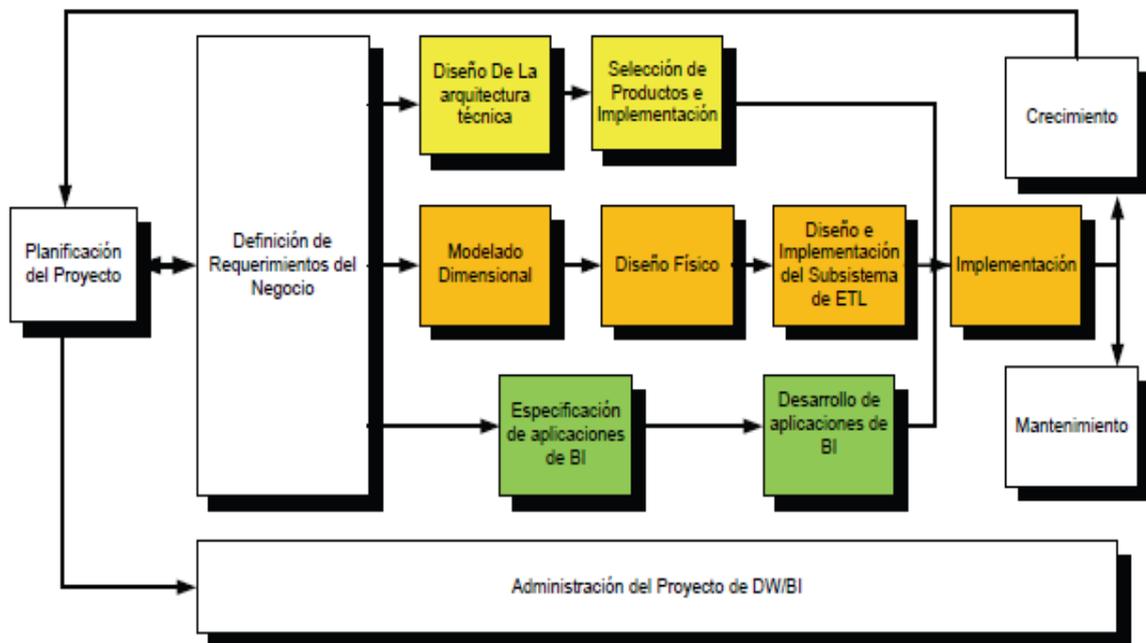
- **Centrarse en el negocio:** Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa
- **Realizar entregas en incrementos significativos:** crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los

incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.

- Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

La construcción de una solución de DW/BI(Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. Las tareas de esta metodología (ciclo de vida) se muestran en la figura 1.

Figura 1 - Tareas de la metodología de Kimball



Fuente: Kimball et al 98,08,Mundy & Thornthwaite

De la figura 1, podemos observar dos cuestiones. Primero, hay que resaltar el rol central de la tarea de definición de requerimientos. Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes. También tiene influencia en el plan de proyecto (nótese la doble fecha entre la caja de definición de requerimientos y la de planificación). En segundo lugar podemos ver tres rutas o caminos que se enfocan en tres diferentes áreas:

- *Tecnología (Camino Superior)*. Implica tareas relacionadas con software específico, por ejemplo, Microsoft SQL Analysis Services.
- *Datos (Camino del medio)*. En la misma diseñaremos e implementaremos el modelo dimensional, y desarrollaremos el subsistema de Extracción, Transformación y Carga (*Extract, Transformation, and Load - ETL*) para cargar el DW.

- *Aplicaciones de Inteligencia de Negocios (Camino Inferior)*. En esta ruta se encuentran tareas en las que diseñamos y desarrollamos las aplicaciones de negocios para los usuarios finales.

Estas rutas se combinan cuando se instala finalmente el sistema. En la parte de debajo de la figura se muestra la actividad general de administración del proyecto. A continuación describiremos cada una de las tareas.

Planificación

En este proceso se determina el propósito del proyecto de DW/BI, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.

En la visión de programas y proyectos de Kimball, Proyecto, se refiere a una iteración simple del KLC (Kimball Life Cycle), desde el lanzamiento hasta el despliegue.

Esta tarea incluye las siguientes acciones típicas de un plan de proyecto:

- Definir el alcance (entender los requerimientos del negocio).
- Identificar las tareas
- Programar las tareas
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos
- Elaboración de un documento final que representa un plan del proyecto.

Además en esta parte definimos cómo realizar la administración o gestión de esta subfase que es todo un proyecto en sí mismo, con las siguientes actividades:

- Monitoreo del estado de los procesos y actividades.
- Rastreo de problemas.
- Desarrollo de un plan de comunicación comprensiva que dirija la empresa y las áreas de TI.

Análisis de requerimientos:

La definición de los requerimientos es en gran medida un proceso de entrevistar al personal de negocio y técnico, pero siempre conviene tener un poco de preparación previa. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo. Hay que leer todos los informes posibles de la organización; rastrear los documentos de estrategia interna; entrevistar a los empleados, analizar lo que se dice en la prensa acerca de la organización, la competencia y la industria. Se deben conocer los términos y la terminología del negocio.

Parte del proceso de preparación es averiguar a quién se debe realmente entrevistar. Esto normalmente implica examinar cuidadosamente el organigrama de la organización. Hay básicamente cuatro grupos de personas con las que hablar desde el principio: el directivo responsable de tomar las decisiones estratégicas; los

administradores intermedios y de negocio responsables de explorar alternativas estratégicas y aplicar decisiones; personal de sistemas, si existen, la gente que realmente sabe qué tipos de problemas informáticos y de datos existen; y por último, la gente que se necesita entrevistar por razones políticas.

Por otra parte, a partir del análisis se puede construir una herramienta de la metodología denominada matriz de procesos/dimensiones (Bus Matrix en inglés).

Una dimensión es una forma o vista o criterio por medio de cual se pueden resumir, cruzar o cortar datos numéricos a analizar, datos que se denominan medidas (measures en inglés).

Esta matriz tiene en sus filas los procesos de los negocios identificados, y en las columnas, las dimensiones identificadas.

Un ejemplo de esta matriz se puede observar en la tabla 1. Cada X en la intersección de las filas y columnas significa que en el proceso de negocio de la fila seleccionada se identifican las dimensiones propuestas.

Figura Tabla 1: matriz de procesos/dimensiones (bus matrix)2

Proceso de Negocio	Dimensiones					
	Tiempo	Producto	Empleados	Clientes (Revendedores)	Geografía de ventas	Importes
Proyección de ventas	X	X	X	X	X	X
Compras	X	X	X	X	X	X
Control de llamadas	X	X	X	X	X	
...						

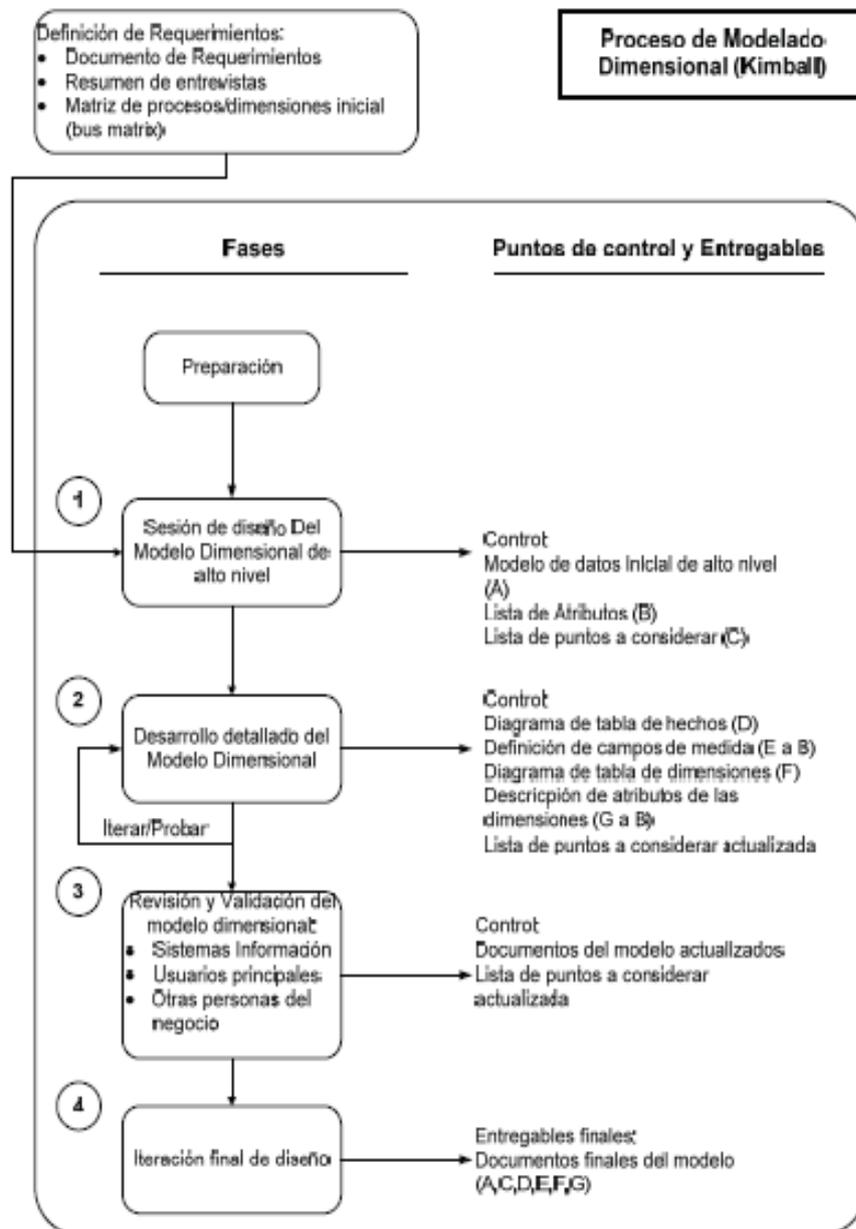
Fuente: Kimball et al 98,08,Mundy & Thornthwaite

Finalmente se busca priorizar los requerimientos o procesos de negocios más críticos.

Modelado Dimensional

La creación de un modelo dimensional es un proceso dinámico y altamente iterativo. Un esquema general se puede ver en la figura 2.

Figura 2 – Diagrama de flujo del proceso dimensional de Kimball (Mundy & Thornthwaite 06)



Fuente: Kimball et al 98,08,Mundy & Thornthwaite

El proceso de diseño comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz descrita en el punto anterior.

El proceso iterativo consiste en cuatro pasos:

1. Elegir el proceso de negocio.
2. Establecer el nivel de granularidad.
3. Elegir las dimensiones.
4. Identificar medidas y las tablas de hechos

1. Elegir el proceso de negocio

El primer paso es elegir el área a modelar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requerimientos y de los temas analíticos anotados en la etapa anterior.

2. Establecer el nivel de granularidad

La granularidad significa especificar el nivel de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales. La sugerencia general es comenzar a diseñar el DW al mayor nivel de detalle posible, ya que se podría luego realizar agrupamientos al nivel deseado. En caso contrario no sería posible abrir (drill-down) las sumalizaciones en caso de que el nivel de detalle no lo permita.

3. Elegir las dimensiones

Las dimensiones surgen naturalmente de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos. Una forma de identificar las tablas de dimensiones es que sus atributos son posibles candidatos para ser encabezado en los informes, tablas pivot, cubos, o cualquier forma de visualización, unidimensional o multidimensional.

4. Identificar las tablas de hechos y medidas

El último paso consiste en identificar las medidas que surgen de los procesos de negocios. Una medida es un atributo (campo) de una tabla que se desea analizar, sumalizando o agrupando sus datos, usando los criterios de corte conocidos como dimensiones. Las medidas habitualmente se vinculan con el nivel de granularidad del punto 4.3.2., y se encuentran en tablas que denominamos tablas de hechos (fact en inglés). Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requerimientos. Un registro contiene una medida expresada en números, como ser cantidad, tiempo, dinero, etc., sobre la cual se desea realizar una operación de agregación (promedio, conteo, suma, etc.) en función de una o más dimensiones. La granularidad es el nivel de detalle que posee cada registro de una tabla de hechos.

5. Modelo gráfico de alto nivel

Para concluir con el proceso dimensional inicial se realiza un gráfico denominado modelo dimensional de alto nivel.

6. Identificación de atributos de dimensiones y tablas de hechos

La segunda parte de la sesión inicial de diseño consiste en completar cada tabla con una lista de atributos bien formada. Una lista de este tipo se muestra en la figura 3. Esta lista o grilla se forma colocando en las filas los atributos de la tabla, y en las columnas la siguiente información:

- Características relacionadas con la futura tabla dimensional del almacén de datos (target), por ejemplo tipo de datos, si es clave primaria, valores de ejemplo, etc. Por razones de espacio no describiremos todas las columnas, para mayor información puede consultarse la referencia (Mundy & Thornthwaite 06).

- El origen de los datos (*source*, por lo general atributos de las tablas transaccionales).
- Reglas de conversión, transformación y carga (*ETL rules*), que nos dicen cómo transformar los datos de las tablas de origen a las del almacén de datos.

Figura 3 –Lista de atributos (Mundy & Thornthwaite)

Table Name: DimOrderInfo		Target											
Table Type	Dimension	Column Name	Description	Datatype	Size	Key?	FK To	NULL?	Default Value	Unknown Member	Example Values	SCD Type	Source System
View Name	OrderInfo	Extended Property?	Y				Y				Y	Y	Y
Description	OrderInfo is the "junk" dimension that includes miscellaneous information about the Order transaction	OrderInfoKey	Surrogate primary key	smallint		PK ID		N		-1	1, 2, 3, 4, ...		ETL Process
Used in schemas	Orders	BKSalesReasonID	Sales reason ID from source system	smallint				N		-1			AW
Generate script?	Y	Channel	Sales channel	char	8					Unknown	Reseller, Internet	1	AW
		SalesReason	Reason for the sale, as reported by the customer	varchar	30					Unknown		1	AW
		SalesReasonType	Type of sales reason	char	10					Unknown	Marketing, Promotion, Other	1	AW
		AuditKey	What process loaded this row?	int		FK	Audit Dim	N		-1		1	Derived

SCD Type		Source					ETL Rules		Comments
Y	Y	Y	Y	Y	Y	Y	Y	Y	
								Standard surrogate key	
1	AW	Sales	SalesReason	SalesReasonID	int			Convert to char; left-pad with zero. 'R' for reseller row.	We need to insert a single row for Reseller
1	AW	Sales	SalesReason	Derived				'Internet' for real sales reasons. 'Reseller' for reseller row.	
1	AW	Sales	SalesReason	Name	nvarchar(50)			Convert to varchar; 'Reseller' for reseller row.	
1	AW	Sales	SalesReason	ReasonType	nvarchar(50)			Convert to varchar; 'Reseller' for reseller row.	
1	Derived							Populated by ETL system using standard technique	

Fuente: Kimball et al 98,08,Mundy & Thornthwaite

7. Implementar el modelo dimensional detallado

Este proceso consiste simplemente en completar la información incompleta de los pasos anteriores. El objetivo en general es identificar todos los atributos útiles y sus ubicaciones, definiciones y reglas de negocios asociadas que especifican cómo se cargan estos datos. Para este cometido se usa la misma planilla del punto anterior.

8. Prueba del modelo

Si el modelo ya está estable, lo que se hace habitualmente es probarlo contra los requerimientos del negocio. Haciendo la pregunta práctica de ¿Cómo podemos obtener esta información en particular del modelo? Para las pruebas podemos usar diseños de reportes estructurados, de usuarios actuales, diseños de cubos prospectivos, etc.

9. Revisión y validación del modelo

Un vez que tenemos confianza plena en el modelo, ingresamos en esta etapa final, lo cual implica revisar el modelo con diferentes audiencias, cada una con diferentes conocimientos técnicos y del negocio. En el área de sistemas deberían revisarlo los programadores y analistas de los sistemas, y el DBA si existe. También debería revisarse con usuarios y personas del negocio que tengan mucho conocimiento de los procesos y que quizás no hayan participado del diseño del modelo. Finalmente podemos hacer un documento que enuncie una serie de preguntas del negocio (tomadas a partir de los requerimientos), y las conteste por medio del modelo.

10. Documentos finales

El producto final, como se puede ver en la Figura 2, son una serie de documentos (solo mencionamos los más importantes), a saber:

- Modelo de datos inicial de alto nivel
- Lista de atributos
- Diagrama de tablas de hechos
- Definición de campos de medida
- Diagrama de tablas de dimensiones
- Descripción de los atributos de las dimensiones
- Matriz DW (o DW Bus Matrix) completa.

En esta parte, intentamos contestar las siguientes preguntas:

- ¿Cómo puede determinar cuán grande será el sistema de DW/BI?
- ¿Cuáles son los factores de uso que llevarán a una configuración más grande y más compleja?
- ¿Cómo se debe configurar el sistema?
- ¿Cuánta memoria y servidores se necesitan? ¿Qué tipo de almacenamiento y procesadores?
- ¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?
- ¿Qué necesitan instalar los diferentes miembros del equipo de DW/BI en sus estaciones de trabajo?
- ¿Cómo convertir el modelo de datos lógico en un modelo de datos físicos en la base de datos relacional?
- ¿Cómo conseguir un plan de indexación inicial?
- ¿Debe usarse la partición en las tablas relacionales?

Diseño del sistema de Extracción, Transformación y Carga (ETL)

El sistema de Extracción, Transformación y Carga (ETL) es la base sobre la cual se alimenta el Datawarehouse. Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en el DW en un formato acorde para la utilización por parte de las herramientas de análisis.

Especificación y desarrollo de aplicaciones de BI

Una parte fundamental de todo proyecto de DW/BI está en proporcionarles a una gran comunidad de usuarios una forma más estructurada y por lo tanto, más fácil, de acceder al almacén de datos.

Proporcionamos este acceso estructurado a través de lo que llamamos aplicaciones de inteligencia de negocios (Business Intelligence Applications).

Las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio.

Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama informes estándar y aplicaciones analíticas.

1. Informes estándar

Los informes estándar son la base del espectro de aplicaciones de BI. Por lo general son informes relativamente simples, de formato predefinido, y parámetros de consulta fijos. En el caso más simple, son informes estáticos pre almacenados. Los informes estándar proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área determinada de la empresa. Este tipo de aplicaciones son el caballo de batalla de la BI de la empresa.

Son informes que los usuarios usan día a día. La mayor parte de lo que piden las personas durante el proceso de definición de requisitos se clasificaría como informes estándar. Por eso es conveniente desarrollar un conjunto de informes estándar en el ciclo de vida del proyecto.

Algunos informes estándares típicos podrían ser:

- Ventas del año actual frente a previsión de ventas por vendedor.
- Tasa de renovación mensual por plan de servicio.
- Tasa quinquenal de deserción por unidad académica.
- Tasas de respuestas de correo electrónico por promoción por producto (marketing).
- Recuento de audiencia y porcentaje de la audiencia total por la red de televisión por día de la semana y hora del día (Sistema de marketing televisivo).
- Reclamos del año actual hasta la fecha frente a previsión, por tipo de vehículo.
- Volumen de llamadas por producto como un porcentaje del total de ventas.

2. Aplicaciones analíticas

Las aplicaciones analíticas son más complejas que los informes estándar. Normalmente se centran en un proceso de negocio específico y resumen cierta experiencia acerca de cómo analizar e interpretar ese proceso de negocio. Estas aplicaciones pueden ser muy avanzadas e incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades o cuestiones subyacentes en los

datos. Otra característica avanzada en algunas aplicaciones analíticas es que el usuario puede pedir cambios en los sistemas transaccionales basándose en los conocimientos obtenidos del uso de la aplicación de BI. En el otro extremo del espectro, algunas aplicaciones analíticas se venden como soluciones cerradas o enlatados, y son independientes de las aplicaciones particulares de la empresa.

Algunas aplicaciones analíticas comunes incluyen:

- Análisis de la eficacia de la promociones.
- Análisis de rutas de acceso en un sitio Web.
- Análisis de afinidad de programas.
- Planificación del espacio en espacios comerciales.
- Detección de fraudes.
- Administración y manejo de categorías de productos. ” (Rivadera)

II.3. Marco Conceptual

Sistema Técnico:

- **Recursos Informáticos:**

Cuadro 1 – Recursos Informáticos

Base de datos	Postgresql 8.3.1
Servidor	-Sistema Operativo: Windows XP service pack 2 -Memoria Física (RAM): 4Gb. -Espacio en disco: 581Gb. -Procesador: Intel Core 2 Duo 2.16 Ghz. -Tarjeta video: NVIDIA GFORCE 1Gb. -Puerto LAN: Tarjeta de Red 10/100/1000Mbps.
Computadoras (2 Pc´s)	--Sistema Operativo: Windows XP service pack 2 -Memoria Física: 2 GB de Ram - Espacio en disco: 160 GB. -Tarjeta gráfica: NVIDIA GeForce 6150 SE -Procesador: AMD Athlon 64 X2 4050e Dual-Core. -Puerto LAN: Tarjeta de Red 10/100/1000Mbps.
RED	Topología estrella
Sistema	Sistema Predial

Fuente: Elaboración propia

- **Agentes:**
 - **Operarios:**
 - Técnico de sistemas.
 - **Gestor:**
 - Gerente de Administración y Finanzas .
 - **Usuarios:**
 - Usuarios Administración Tributaria

CAPITULO III: MATERIALES Y MÉTODOS

III.1: Tipo de investigación

III.1.1. Según el propósito

- Investigación Tecnológica

III.1.2. Según el diseño de investigación

- Pre experimental: Co relacional.

III.2: Diseño de investigación

III.2.1. Hipótesis

III.2.1.1.Planteamiento de la hipótesis

El desarrollo de un Data Mart ayuda a elevar la eficiencia en la toma de decisiones en el área de Administración y Finanzas de la Municipalidad Distrital de Baños del Inca en no menos de un 10%.

III.2.1.2.Variables

VARIABLE DEPENDIENTE:

Eficiencia en la toma de decisiones en el área de Administración y Finanzas.

VARIABLE INDEPENDIENTE:

El desarrollo de un Data Mart.

III.2.1.3.Operacionalización de variables

MATRIZ DE CONSISTENCIA DEL PROYECTO DE INVESTIGACIÓN
(Ver anexo 1)

III.2.2. Material de estudio

III.2.2.1.Población

La población seleccionada para este trabajo es el personal que labora en la municipalidad del área de administración y finanzas.

III.2.2.2. Muestra

Población muestral.

III.2.3. Diseño de Contrastación

Es el diseño de contrastación Pre-Experimental consta de tres etapas: pre test- la aplicación de la variable independiente y post test.

Formalización:



Donde:

O1: La toma de decisiones en el área de Administración y Finanzas antes de implementar el Datamart (variable dependiente).

X: El desarrollo del Data Mart (variable independiente).

O2: La toma de decisiones en el área de Administración y Finanzas después de implementar el Datamart.

III.2.4. Técnicas, procedimientos e instrumentos

III.2.4.1. De recolección de datos

Se aplicará una ficha de observación mediante un cuestionario, sobre el funcionamiento de la base de datos y el manejo de la misma. (Ver anexo 2 y anexo 3)

III.2.4.2. De procesamiento de datos

Se aplicará a las variables un análisis de datos cuantitativos mediante estadística descriptiva.

CAPITULO IV: RESULTADOS

La metodología utilizada para el desarrollo de la investigación es de Ralph Kimball, que consta de las siguientes etapas:

IV.1. Planeamiento

a) Misión de la institución

La Municipalidad de Los Baños del Inca promueve la gobernabilidad democrática en la localidad, asegurando la mayor participación ciudadana en la formulación de las políticas públicas locales y en la gestión y fiscalización de las acciones municipales, desarrollando al máximo sus capacidades para brindar bienes y servicios públicos de alta calidad con la mayor eficiencia e eficacia, haciendo un uso responsable, transparente y estratégico de los recursos públicos, promoviendo sinergias con las inversiones de otras Instituciones del Estado y del sector privado, para mejorar la calidad de vida de los ciudadanos en la jurisdicción.

b) Visión

Que Los Baños del Inca, reconocida como Primera Maravilla del Perú, cuente con una industria agropecuaria competitiva, ambientalmente sustentable, esté integrado socialmente, y se impulse a trabajar en base a la igualdad de oportunidades y la realización misma de las personas que integran esta comunidad.

c) Valores

Transparencia:

Claridad en nuestra gestión institucional, en los procesos operativos, administrativos y con los contribuyentes brindándoles la información y confianza para fomentar responsabilidad ciudadana; tenemos valores complementarios que consolidan a este como: La Honestidad, Lealtad, Equidad y Justicia.

Vocación de servicio:

Disposición del personal a la atención eficiente y cordial a los contribuyentes, tenemos valores complementarios que consolidan a este como: La Responsabilidad, Puntualidad, Compromiso, Orden, Pro actividad e Identificación.

Innovación:

Es la capacidad de generar procesos eficientes mejorándolos constantemente y estar acorde con los avances de la tecnología, contamos con valores complementarios que consolidan a este como: Cambio y Creatividad.

d) Marco Institucional

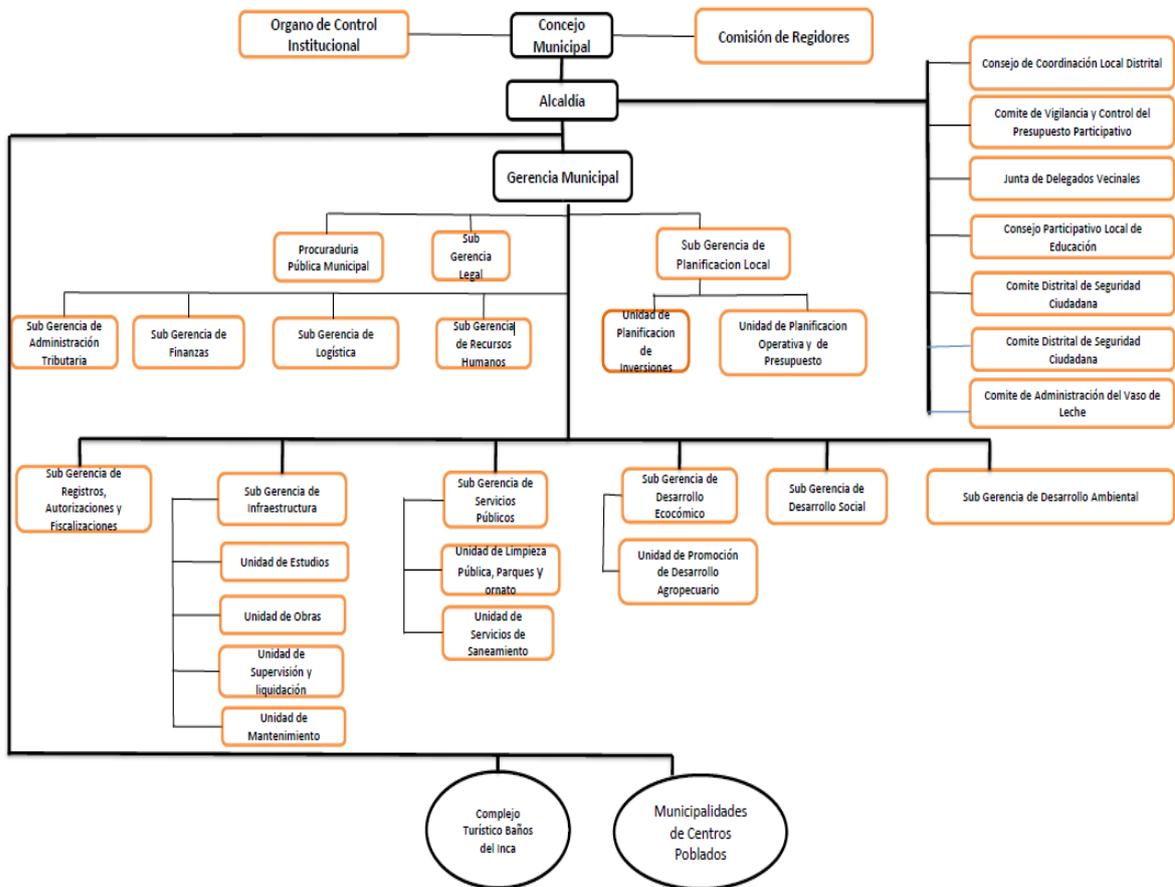
La Municipalidad Distrital de los Baños del Inca promueve la gobernabilidad democrática en la localidad, asegurando la mayor participación ciudadana en la formulación de las políticas públicas locales y en la gestión y fiscalización de las acciones municipales, desarrollando al máximo sus capacidades para brindar bienes y servicios públicos de alta calidad con la mayor eficiencia e eficacia, haciendo un uso responsable, transparente y estratégico de los recursos públicos, promoviendo sinergias con las inversiones de otras Instituciones del Estado y del sector privado, para mejorar la calidad de vida de los ciudadanos en la jurisdicción.

e) Datos Generales de la Institución

- Razón Social:
Municipalidad Distrital de Baños del Inca
- RUC:
20143625681
- Tipo Empresa:
Gobierno Regional, local
- Dirección Legal:
Av. Atahualpa Nro. S/n (Plaza de Armas)
- Distrito / Ciudad:
Baños del Inca
- Departamento:
Cajamarca
- Página Web:
<http://www.mdbi.gob.pe/index.php>

- Organigrama:

Figura 4 -Organigrama de la Municipalidad Distrital de Baños del Inca



Fuente: Municipalidad Distrital de Baños del Inca

f) Descripción del requerimiento

La identificación de los requerimientos se logró luego de analizar la información recolectada que se basó en entrevistas realizadas al personal involucrado en el proyecto.

- Elaborar un análisis de los impuestos recaudados en los últimos años.
- Estudiar el comportamiento de los clientes frente a las promociones que se realizan.
- Elaborar la cartera de clientes.
- Comparar los montos cancelados y los montos pronosticados en el año.
- Analizar el índice de morosidad de los últimos años.

g) Descripción de stakeholders

Cuadro 2 – Descripción de stakeholders

Stakeholder	Cargo	Función que desempeña en el proyecto
Cristian Paul Pajares Rabanal	Sub Gerente del área	Líder del Proyecto
Melsi Ocas Terrones	Analista dimensional	Responsable de TI
Melsi Ocas Terrones	Experto en ETL	Responsable de TI
Melsi Ocas Terrones	Experto en Cubos	Responsable de TI
Melsi Ocas Terrones	Experto en desarrollo	Responsable de TI
Melsi Ocas Terrones	Especialista en datos	Responsable de TI

Fuente: Elaboración propia

h) Requisitos

a. Humanos

i. Gestión

Cuadro 3 – Requisitos Humanos - Gestión

Persona	Cargo	Rol
Cristian Paul Pajares Rabanal	Sub Gerente de la Administración Tributaria	Encargado de la administración tributaria

Fuente: Elaboración propia

ii. Tecnologías Información

Cuadro 4 – Requisitos Humanos – Tecnologías información

Persona	Cargo	Rol
Farly Minchan Lezcano	Administrador DE BD	Administrar las bases de datos

Fuente: Elaboración propia

b. Materiales

Cuadro 5 – Requisitos Materiales

Material	Cantidad	Observación
Papel	1 millar	Documentar el proyecto
Impresora Funcional	1	Imprimir documentos para usos del proyecto
Libros	1	Analizar la metodología a desarrollar en el proyecto.

Fuente: Elaboración propia

c. Equipamiento
i. Hardware

Cuadro 6 – Requisitos Equipamiento - Hardware

Equipo	Configuración	Características Adicionales	Uso Proyecto(1) Uso Producto(2)
1 Servidor		-Sistema Operativo: Windows XP service pack 2 -Memoria Física (RAM): 4Gb. -Espacio en disco: 581Gb. -Procesador: Intel Core 2 Duo 2.16 Ghz. -Tarjeta video: NVIDIA GFORCE 1Gb. -Puerto LAN: Tarjeta de Red 10/100/1000Mbps.	2
2 Computadoras		-Sistema Operativo: WindowsXP service pack 2 -Memoria Física: 2 GB de Ram. - Espacio en disco: 160 GB. -Tarjeta gráfica: NVIDIA GeForce 6150 SE -Procesador: AMD Athlon 64 X2 4050e Dual-Core. -Puerto LAN: Tarjeta de Red 10/100/1000Mbps.	2
1 Laptop		-Sistema Operativo: Windows 7 service pack 1 -Memoria Física: 4 GB de Ram. - Espacio en disco: 581 GB. -Tarjeta gráfica : NVIDIA GeForce.	1

Fuente: Elaboración propia

ii. Software

Cuadro 7 – Requisitos Equipamiento - Software

Uso	Fabricante- Software- Versión	Características Adicionales	Uso Proyecto(1) Uso Producto(2)
Data Mart	Pentaho 4.3	Diseñar y construir cubos multidimensionales OLAP	1
Servicios Análisis	MicroOLAP Database Designer	Análisis de la estructura de la BD	1
Herramientas de Desarrollo	POSTGRES	BD de la Municipalidad Distrital de Baños del Inca	1
Herramientas de Desarrollo	Mysql	BD que contiene las dimensiones	1

Fuente: Elaboración propia

iii. Comunicaciones

Cuadro 8 – Requisitos Equipamiento - Comunicaciones

Equipo	Fabricante	Características Adicionales
1 Switch D-Link DES 1008D	D-Link	Comunica los diferentes equipos conectados a la red.

Fuente: Elaboración propia

i) Riesgos del proyecto

En el siguiente cuadro se realiza un resumen de los posibles riesgos que se pueden dar en el desarrollo del proyecto.

Cuadro 9 – Riesgos del proyecto

Riesgo	Descripción
1	Incumplimiento del cronograma.
2	Insuficiente información para el desarrollo del proyecto.
3	Problemas en la configuración entre las bases de datos y la herramienta de BI.
4	Poca colaboración del personal en las entrevistas a realizar.

Fuente: Elaboración propia

j) **Conformación del equipo de trabajo**

Cuadro 10 – Conformación de equipos

Personal	Cargo	Fono	Rol en el Proyecto
Christian Paul Pajares Rabanal	Sub Gerente Administración Tributaria		Representante del Negocio
Christian Paul Pajares Rabanal	Sub Gerente Administración Tributaria		Experto del Negocio
Farly Minchan Lezcano	Administrador BD		Auditor de Datos
Farly Minchan Lezcano	Administrador BD		Líder Tecnológico
Melsi Ocas Terrones			Analista Dimensional
Melsi Ocas Terrones			Desarrollador Dimensional

Fuente: Elaboración propia

IV.2. Requerimientos del negocio

- **Relacionado al proceso de producción**

En esta matriz se tienen los procesos de negocio identificados, y las dimensiones identificadas.

Cuadro 11 – Conformación de equipos

	Dimensión	Dimensión	Dimensión
Proceso del negocio	Contribuyente	Predios	Rubro
Planificación de recaudación	X	X	X

Fuente: Elaboración propia

- **Resultados del análisis en base a encuestas, cuestionarios, reportes, plan estratégico**

A partir de las encuestas realizadas al personal del negocio y técnico, se pudo identificar el proceso del negocio a ser estudiado.

Cuadro 12 – Conformación de equipos

Tema Analítico	Análisis o requerimiento inferido o pedido	Proceso de negocio de soporte	Comentarios
Planificación de Recaudación	Análisis histórico de predios	Recaudación Tributaria	Por contribuyente, por rubro
	Proyección de Recaudación	Recaudación Tributaria	Proceso que usa las recaudaciones como entrada

Fuente: Elaboración propia

- **Listado de requerimientos más comunes:**

Mostramos aquí los reportes más comunes solicitados por el área de Administración y Finanzas.

- Reportes de Predios recaudados semanales.
- Reportes de Clientes deudores al año.
- Reportes de Recaudaciones pendientes de pagos semanales.

- **Concluimos lo siguiente:**

- **Hoja de gestión**

PROCESOS	
Proceso	Recaudación de Impuestos
Objetivo	Brindar servicios de calidad en los procesos de recaudación, proporcionando recursos financieros a la Municipalidad Distrital de Baños del Inca para el cumplimiento de sus fines y fomentar conciencia tributaria en los contribuyentes.
Estrategias	Difusión de la cultura de cumplimiento voluntario. Gestión de cobranza virtual. Información masiva(lugares turísticos, playas de estacionamiento,etc) Campañas de fidelización y sensibilización para promover el cumplimiento voluntario. Programas para mejorar la información que se brinda a los ciudadanos.

Indicador	Indicadores	Medidas	Estados
	INDICADORES DE PREDIOS AL DIA	SUM(PREDIOS PAGADOS)	Mayor 63% 49% - 63 Menor 49%
	CLIENTES PUNTUALES	SUM (CLIENTES QUE PAGAN A TIEMPO)	Mayor 70% 49% - 70 Menor 49%

○ **Hoja de análisis**

Proceso	Recaudación de Impuestos
----------------	---------------------------------

Medidas	MONTOS CANCELADOS PREDIOS PAGADOS CARTERA DE CLIENTES
----------------	---

Interrogantes	Dimensiones	Formas De Analizar La Dimensión
----------------------	--------------------	--

¿Qué?	Concepto	Categoría	Rubro	Estado

	PrediosyOtros	Codigo Calle	Tipo NroCuadra	Motivo Area	
¿Cuándo?	Tiempo	Anual Mensual Temporada	Semestre Semanal	Trimestre Dia	
¿Dónde?	Partida	Tipo Cod_par Descripcion	Generica Especifica Estado	Subgenerica Sub especifica	
¿Cómo?	Fuente de Financiamiento	Tipo	Descripción		
	Tenencia	Tipo	Descripción		
	Tipo Terreno	Tipo	Descripción		
	Uso	Tipo	Descripción		
¿Quién?	Contribuyente	Tipo Ciudad	Nombre Grupo Dirección	Conyuge Direccion	Cod_per

Fuente: Elaboración propia

○ Matrices

Figura 5 –Figura de Cuadro de Dimensiones y Jerarquías

CUADRO DE DIMENSIONES Y JERARQUIAS					
DIMENSIONES	NIVELES				
	NIVEL 1	NIVEL 2	NIVEL 3	NIVEL 4	NIVEL 5
CONCEPTO	RUBRO	CATEGORIA			
	RUBRO				
	CATEGORIA				
	ESTADO	CATEGORIA			
TIEMPO	Día	Semanal	Mensual	Trimestral	Anual
	MES	ANUAL			
	ANUAL				
	TEMPORADA	ANUAL			
PARTIDA	Descripción Transaccional	Descripción Genérica			
	Descripción Genérica				
FUENTE DE FINANCIAMIENTO	Tipo	Descripción			
	Tipo				
CONTRIBUYENTE	Tipo	Grupo Dirección	Ciudad		
	Tipo				
PREDIOS Y OTROS	Tipo	Codigo	NroCuadra	Calle	Area
TENENCIA	Tipo	Descripción			
TIPO TERRENO	Tipo	Descripción			
USO	Tipo	Descripción			

Fuente: Elaboración propia

Figura 6 –Figura de Cuadro de Dimensiones y Medidas

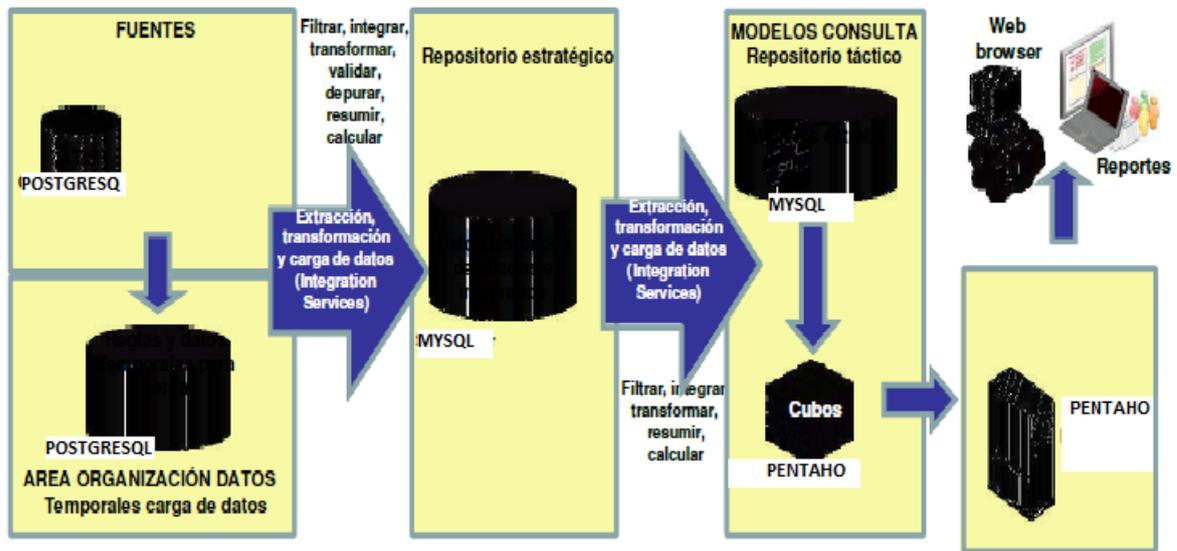
CUADRO DE DIMENSIONES Y MEDIDAS									
	CONCEPTO	TIEMPO	PARTIDA	FUENTE DE FINANCIAMIENTO	CONTRIBUYENTE	PREDIOS Y OTROS	TENENCIA	TIPO TERRENO	USO
MONTOS CANCELADOS	X	X	X	X	X	X	X	X	X
CARTERA DE CLIENTES	X	X	X	X	X	X	X	X	X

Fuente: Elaboración propia

IV.3. Diseño de arquitectura

En la figura N° 5 se ilustra la plataforma utilizada para la solución de inteligencia de negocios.

Figura 7 – Diseño Arquitectura



Fuente: Elaboración propia

Los componentes principales son:

Sistemas transaccionales: Donde se encuentran almacenadas las transacciones de la institución, la base de datos con la que contamos está en postgresql.

Base de Datos Mysql: Es la base de datos donde se guarda los registros temporales del negocio para la carga de la base de datos del datamart.

Pentaho: Esta herramienta nos permitió la integración de datos, incluidas la extracción, la transformación y la carga (ETL).

IV.4. Instalación y selección del producto

IV.4.1. Instalación

A continuación se muestran las instrucciones a tomar en cuenta para la instalación y configuración de la herramienta de inteligencia de negocios que se usa para la implementación del proyecto. Como repositorio del Data Mart se usa Mysql y como herramienta de análisis se tiene la plataforma Pentaho.

Instalación de Mysql

El gestor de base de datos Mysql es de software libre, para obtenerlo se descarga desde la página web: <http://www.mysql.com/downloads/mysql/>

El proceso de instalación y configuración es amigable ya que es realizado por un instalador wizard.

En este punto debemos prestar mucha atención ya que es donde definimos el password del administrador principal del Servidor MySQL, o sea el usuario root. También podemos indicar si queremos que nuestras bases de datos se accedan desde maquinas remotas con el usuario root, tal como se muestra en la siguiente figura:

Figura 8 - Configuración de Mysql



Fuente: Elaboración propia

Instalación de la plataforma Pentaho

Para poder obtener el Pentaho es necesario descargar el paquete completo de la página web: www.pentaho.org. Se accede al Spoon mediante el ícono Spoon.archivo por lotes y se muestra la siguiente figura:

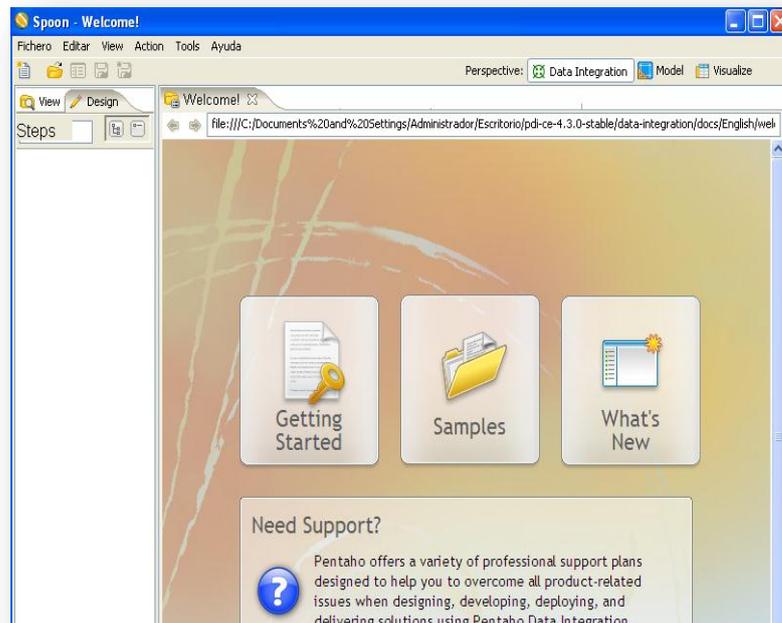
Figura 9 – Ejecución Pentaho Data Integration



Fuente: Elaboración propia

Luego aparecerá el área de trabajo para la ETL, la cual se muestra en la siguiente figura:

Figura 10 –Pantalla de inicio de Pentaho Data Integration

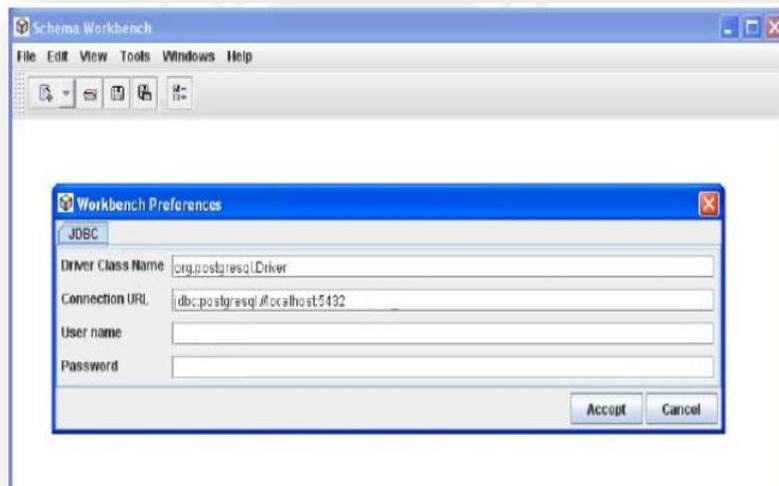


Fuente: Elaboración propia

Instalación de Pentaho Workbench

Descargar de la Web del Pentaho el archivo workbench y descomprimirlo en el lugar de su elección. Luego se ejecuta el archivo workbench.bat y nos muestra la siguiente figura:

Figura 11 – Pantalla de inicio de Pentaho Workbench

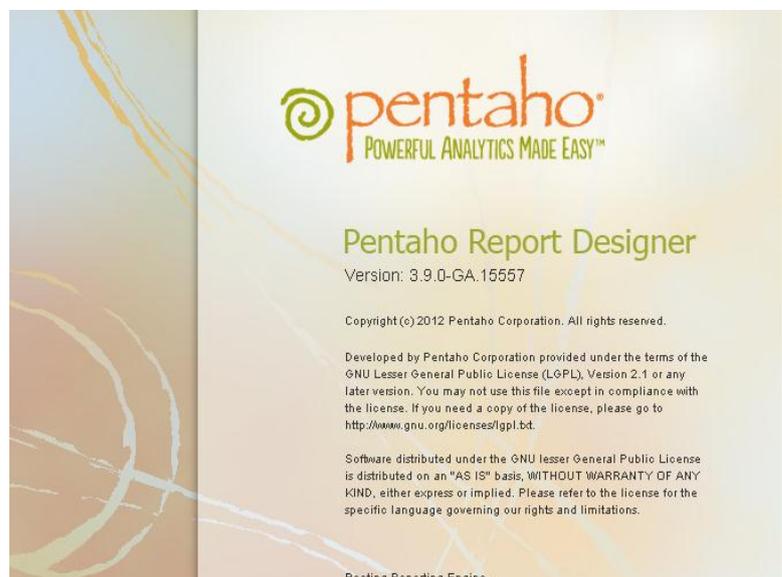


Fuente: Elaboración propia

Instalación de Report Designer

Descargar de la Web del Pentaho Report Designer , y descomprimirlo en el lugar de su elección. Luego se ejecuta el archivo startdesigner.bat y aparece la siguiente figura:

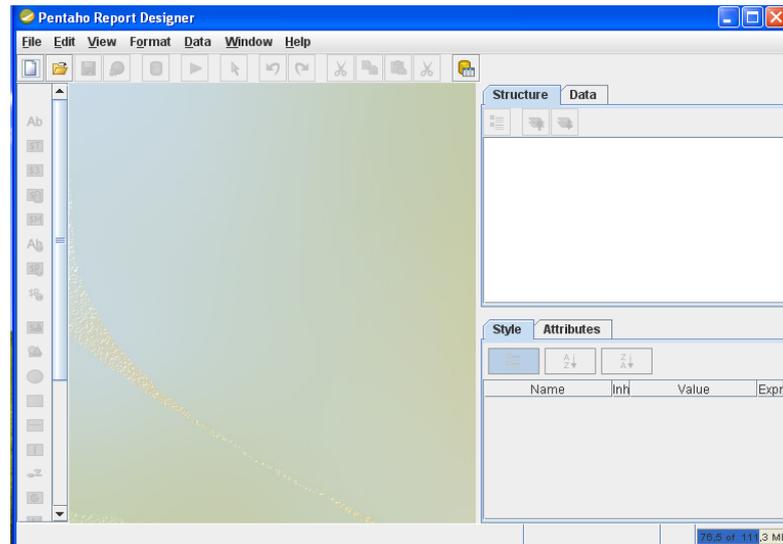
Figura 12 – Ejecución Pentaho Report Designer



Fuente: Elaboración propia

Luego aparecerá el área de trabajo de Pentaho Report Designer, la cual se muestra en la siguiente figura:

Figura 13 – Pantalla de inicio de Pentaho Report Designer



Fuente: Elaboración propia

IV.4.2. Selección del producto

Para la selección de las herramientas utilizadas en el desarrollo del proyecto, se han tomado en cuenta las siguientes consideraciones:

- La Institución tiene como uno de sus estándares el uso de herramientas Open Source.
- Información disponible para el aprendizaje de uso de todas las herramientas.

Las herramientas elegidas son:

- Mysql: el gestor de base de datos.
- Pentaho Data Integration: para realizar ETL.
- Pentaho Reporting: permite la distribución de los resultados del análisis en múltiples formatos - todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto.
- Pentaho workbench permite realizar el procesamiento de los cubos y explotación de los datos.

Justificación de Mysql

- El primer factor por la selección de Mysql es por el tipo de proyecto que se está utilizando, en el cual se busca economía y se están utilizando productos Open Source.
- Algunas de las características que más sobresalen de Mysql, que lo hacen óptimo para proyectos Web que no manejen escenarios críticos, con su rapidez de repuesta, bajo consumo de recursos y facilidad de uso.

Justificación de Pentaho

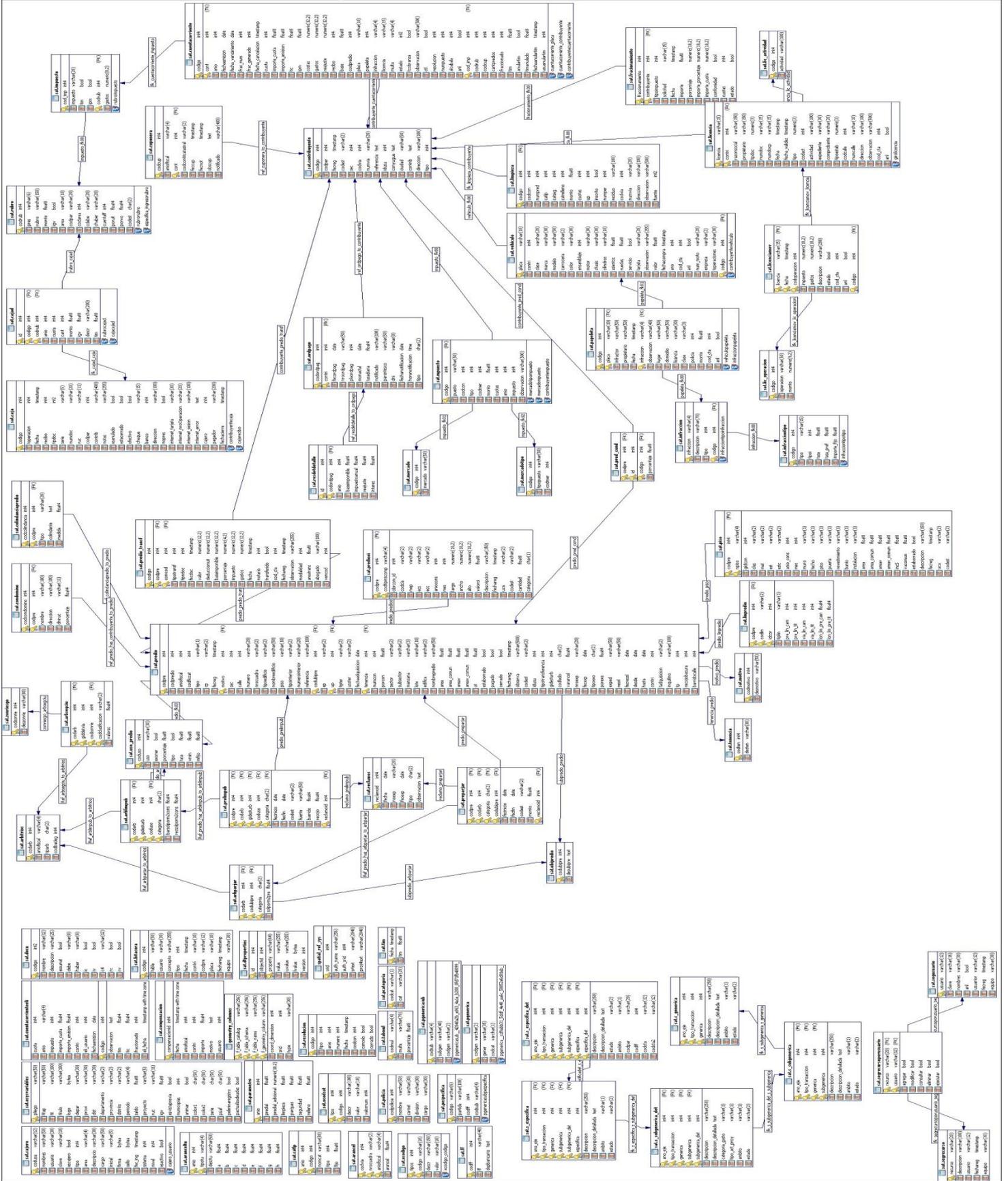
- Pentaho es una plataforma de BI y centrada en procesos, incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos y ha sido concebido desde sus inicios para ser una solución basada en procesos.
- Existe información detallada sobre la implementación de inteligencia de negocios haciendo uso de estas herramientas.

IV.5. Modelamiento dimensional

El diseño de los modelos de las bases de datos están orientados a modelar los procesos de acuerdo a los requerimientos de los usuarios. Analizando el modelo de base de datos del sistema transaccional podremos construir el modelo del data mart.

IV.5.1. Modelo de Base de datos Transaccional

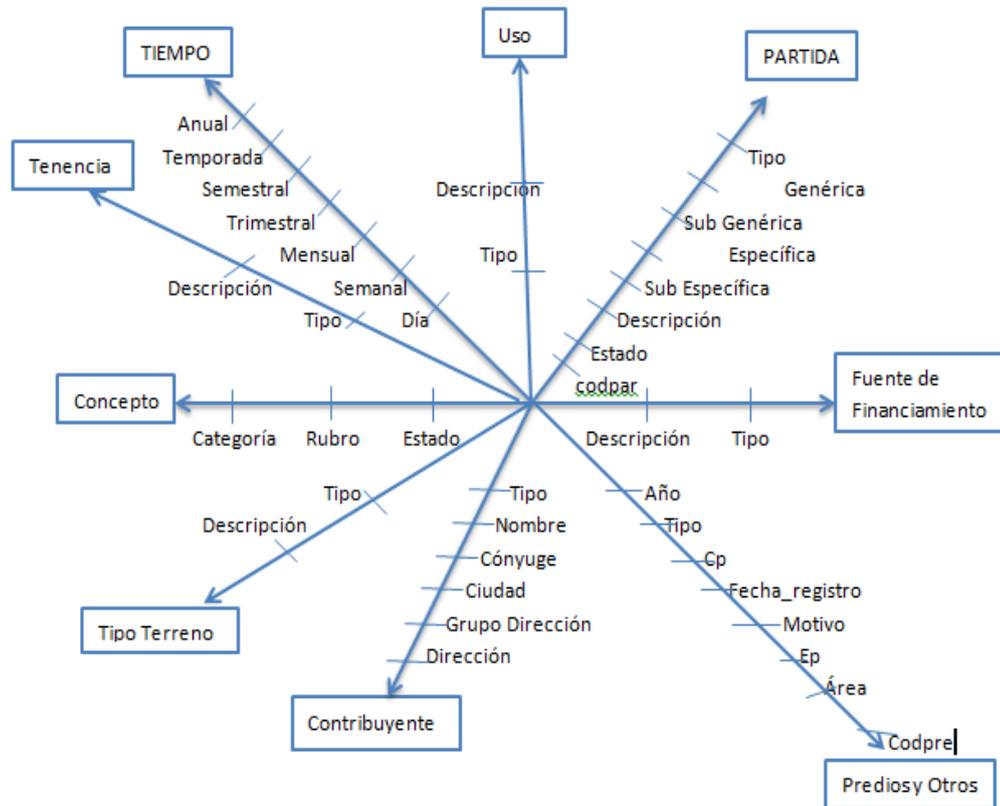
Figura 14 – Diseño de la Base de Datos Transaccional



Fuente: Municipalidad Distrital de Baños del Inca

IV.5.2. Modelamiento Dimensional Estrella

Figura 15 – Modelamiento Dimensional Estrella



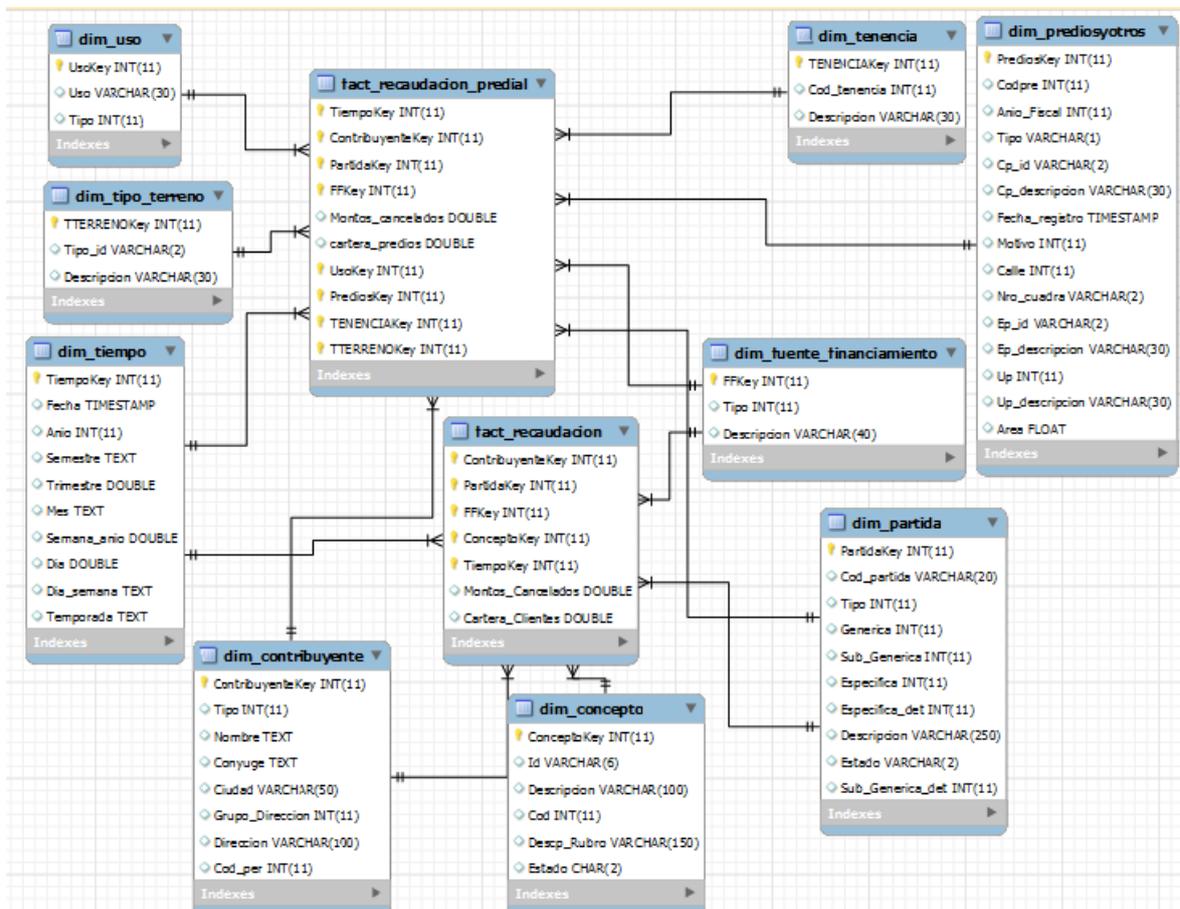
Fuente: Elaboración propia

IV.6. Diseño Físico

IV.6.1. Modelo de Base de Datos Dimensional y Datos

El esquema municipaladimension está conformado por nueve dimensiones y dos tablas de hechos, las tablas de hechos llamadas: fact_recaudacion_predial y fact_recaudacion. Y las dimensiones llamadas: dim_concepto, dim_contribuyente, dim_fuente_financiamiento, dim_partida, dim_prediosyotros, dim_tenencia, dim_tiempo, dim_tipo_terreno, dim_uso.

Figura 16 – Modelo Dimensional Esquema municipaladimension



Fuente: Elaboración propia

Teniendo listo la base de datos, podemos hacer una ingeniería directa para poder generar el script de nuestra base de datos dimensional a utilizar.

```
CREATE TABLE DIM_CONCEPTO
(
    ConceptoKey    INTEGER AUTO_INCREMENT PRIMARY KEY,
    Id             VARCHAR(6) NULL,
    Descripcion    VARCHAR(100) NULL,
    Cod            INTEGER NULL,
    Descp_Rubro   VARCHAR(150) NULL,
    Estado        CHAR(2) NULL
);
```

```
CREATE TABLE DIM_CONTRIBUYENTE
(
    ContribuyenteKey INTEGER AUTO_INCREMENT PRIMARY KEY,
    Tipo             INTEGER NULL,
    Nombre           TEXT NULL,
    Conyuge          TEXT NULL,
    Ciudad           VARCHAR(50) NULL,
    Grupo_Direccion INTEGER NULL,
    Direccion        VARCHAR(100) NULL,
    Cod_per          INTEGER NULL
);
```

```
CREATE TABLE DIM_FUENTE_FINANCIAMIENTO
(
    FFKey          INTEGER AUTO_INCREMENT PRIMARY KEY,
    Tipo           INTEGER NULL,
    Descripcion    VARCHAR(40) NULL
);
```

```

CREATE TABLE DIM_PARTIDA
(
    PartidaKey      INTEGER AUTO_INCREMENT PRIMARY KEY,
    Cod_partida    VARCHAR(20) NULL,
    Tipo           INTEGER NULL,
    Generica       INTEGER NULL,
    Sub_Generica   INTEGER NULL,
    Especifica     INTEGER NULL,
    Especifica_det INTEGER NULL,
    Descripcion    VARCHAR(250) NULL,
    Estado         VARCHAR(2) NULL,
    Sub_Generica_det  INTEGER NULL
);

```

```

CREATE TABLE DIM_PREDIOSYOTROS
(
    PrediosKey      INTEGER AUTO_INCREMENT PRIMARY KEY,
    Codpre          INTEGER NULL,
    Anio_Fiscal     INTEGER NULL,
    Tipo           VARCHAR(1) NULL,
    Cp_id          VARCHAR(2) NULL,
    Cp_descripcion  VARCHAR(30) NULL,
    Fecha_registro  TIMESTAMP NULL,
    Motivo         INTEGER NULL,
    Calle          INTEGER NULL,
    Nro_cuadra     VARCHAR(2) NULL,
    Ep_id          VARCHAR(2) NULL,
    Ep_descripcion  VARCHAR(30) NULL,
    Up             INTEGER NULL,
    Up_descripcion  VARCHAR(30) NULL,
    Area           FLOAT NULL
);

```

```

CREATE TABLE DIM_TENENCIA
(
    TENENCIAKey    INTEGER AUTO_INCREMENT PRIMARY KEY,
    Cod_tenencia   INTEGER NULL,
    Descripcion    VARCHAR(30) NULL
);

```

```

CREATE TABLE DIM_TIEMPO
(
    TiempoKey    INTEGER AUTO_INCREMENT PRIMARY KEY,
    Fecha        TIMESTAMP NULL,
    Anio         INTEGER NULL,
    Semestre     TEXT NULL,
    Trimestre    DOUBLE NULL,
    Mes         TEXT NULL,
    Semana_anio  DOUBLE NULL,
    Dia          DOUBLE NULL,
    Dia_semana   TEXT NULL,
    Temporada    TEXT NULL
);

```

```

CREATE TABLE DIM_TIPO_TERRENO
(
    TTERRENOKey  INTEGER AUTO_INCREMENT PRIMARY KEY,
    Tipo_id      VARCHAR(2) NULL,
    Descripcion   VARCHAR(30) NULL
);

```

```

CREATE TABLE DIM_USO
(
    UsoKey       INTEGER AUTO_INCREMENT PRIMARY KEY,
    Uso          VARCHAR(30) NULL,
    Tipo         INTEGER NULL
);

```

```

CREATE TABLE FACT_RECAUDACION
(
    ContribuyenteKey    INTEGER NOT NULL,
    PartidaKey          INTEGER NOT NULL,
    FFKey               INTEGER NOT NULL,
    ConceptoKey         INTEGER NOT NULL,
    TiempoKey           INTEGER NOT NULL,
    Montos_Cancelados   CHAR(18) NULL,
    Cartera_Clientes    CHAR(18) NULL
);

ALTER TABLE FACT_RECAUDACION
ADD PRIMARY KEY
(ContribuyenteKey,PartidaKey,FFKey,ConceptoKey,TiempoKey);

CREATE TABLE FACT_RECAUDACION_PREDIAL
(
    TiempoKey           INTEGER NOT NULL,
    ContribuyenteKey    INTEGER NOT NULL,
    PartidaKey          INTEGER NOT NULL,
    FFKey               INTEGER NOT NULL,
    Montos_cancelados   CHAR(18) NULL,
    cartera_predios     CHAR(18) NULL,
    UsoKey              INTEGER NOT NULL,
    PrediosKey          INTEGER NOT NULL,
    TENENCIAKey        INTEGER NOT NULL,
    TTERRENOKey        INTEGER NOT NULL
);

ALTER TABLE FACT_RECAUDACION_PREDIAL
ADD PRIMARY KEY
(TiempoKey,ContribuyenteKey,PartidaKey,FFKey,UsoKey,PrediosKey,TE
NENCIAKey,TTERRENOKey);

ALTER TABLE FACT_RECAUDACION
ADD FOREIGN KEY R_2 (ContribuyenteKey) REFERENCES
DIM_CONTRIBUYENTE (ContribuyenteKey);

ALTER TABLE FACT_RECAUDACION

```

```
ADD FOREIGN KEY R_3 (PartidaKey) REFERENCES DIM_PARTIDA  
(PartidaKey);
```

```
ALTER TABLE FACT_RECAUDACION  
ADD FOREIGN KEY R_4 (FFKey) REFERENCES  
DIM_FUENTE_FINANCIAMIENTO (FFKey);
```

```
ALTER TABLE FACT_RECAUDACION  
ADD FOREIGN KEY R_5 (ConceptoKey) REFERENCES  
DIM_CONCEPTO (ConceptoKey);
```

```
ALTER TABLE FACT_RECAUDACION  
ADD FOREIGN KEY R_6 (TiempoKey) REFERENCES DIM_TIEMPO  
(TiempoKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL  
ADD FOREIGN KEY R_9 (TiempoKey) REFERENCES DIM_TIEMPO  
(TiempoKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL  
ADD FOREIGN KEY R_10 (ContribuyenteKey) REFERENCES  
DIM_CONTRIBUYENTE (ContribuyenteKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL  
ADD FOREIGN KEY R_11 (PartidaKey) REFERENCES DIM_PARTIDA  
(PartidaKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL  
ADD FOREIGN KEY R_12 (FFKey) REFERENCES  
DIM_FUENTE_FINANCIAMIENTO (FFKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL  
ADD FOREIGN KEY R_8 (UsoKey) REFERENCES DIM_USO (UsoKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL
ADD FOREIGN KEY R_7 (PrediosKey) REFERENCES
DIM_PREDIOSYOTROS (PrediosKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL
ADD FOREIGN KEY R_13 (TENENCIAKey) REFERENCES
DIM_TENENCIA (TENENCIAKey);
```

```
ALTER TABLE FACT_RECAUDACION_PREDIAL
ADD FOREIGN KEY R_14 (TTERRENOKey) REFERENCES
DIM_TIPO_TERRENO (TTERRENOKey);
```

IV.6.2 Dimensiones y Tablas Hechos

Dimensiones

A continuación se definen las dimensiones, cumpliendo con la granularidad de las tablas hechos:

- a. dim_concepto.
- b. dim_contribuyente.
- c. dim_fuente_financiamiento.
- d. dim_partida.
- e. dim_prediosyotros.
- f. dim_tenencia.
- g. dim_tiempo.
- h. dim_tipo_terreno.
- i. dim_uso

a. dim_concepto

Descripción:

Contiene los diferentes rubros pertenecientes a las recaudaciones.

Atributos:

Cuadro 13 – Atributos dim_concepto

Nombre del Atributo	Descripción	Formato	Valor por Defecto
ConceptoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Id	Código del atributo descripción	Varchar(6)	Ninguno
Descripción	Nombre descripción	Varchar(100)	Ninguno
Cod	Código de Rubro	Int(11)	Ninguno
Descp_Rubro	Nombre de Rubro	Varchar(50)	Ninguno
Estado	Activo/ Inactivo	Char(2)	Ninguno

Fuente: Elaboración propia

b. dim_contribuyente

Descripción:

Contiene los datos personales de los contribuyentes

Atributos:

Cuadro 14 – Atributos dim_contribuyente

Nombre del Atributo	Descripción	Formato	Valor por Defecto
ConceptoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Id	Código del atributo descripción	Varchar(6)	Ninguno
Descripción	Nombre descripción	Varchar(100)	Ninguno
Cod	Código de Rubro	Int(11)	Ninguno
Descp_Rubro	Nombre de Rubro	Varchar(50)	Ninguno
Estado	Activo/ Inactivo	Char(2)	Ninguno

Fuente: Elaboración propia

c. dim_fuente_financiamiento

Descripción:

Contiene las diferentes fuentes de financiamiento.

Atributos:

Cuadro 15 – Atributos dim_fuente_financiamiento

Nombre del Atributo	Descripción	Formato	Valor por Defecto
ConceptoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Id	Código del atributo descripción	Varchar(6)	Ninguno
Descripción	Nombre descripción	Varchar(100)	Ninguno
Cod	Código de Rubro	Int(11)	Ninguno
Descp_Rubro	Nombre de Rubro	Varchar(50)	Ninguno
Estado	Activo/ Inactivo	Char(2)	Ninguno

Fuente: Elaboración propia

d. dim_partida

Descripción:

Contiene información desde tipo a específica.

Atributos:

Cuadro 16 – Atributos dim_partida

Nombre del Atributo	Descripción	Formato	Valor por Defecto
ConceptoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Id	Código del atributo descripción	Varchar(6)	Ninguno
Descripción	Nombre descripción	Varchar(100)	Ninguno
Cod	Código de Rubro	Int(11)	Ninguno
Descp_Rubro	Nombre de Rubro	Varchar(50)	Ninguno
Estado	Activo/ Inactivo	Char(2)	Ninguno

Fuente: Elaboración propia

e. dim_prediosyotros

Descripción:

Contiene la información de los predios

Atributos:

Cuadro 17 – Atributos dim_prediosyotros

Nombre del Atributo	Descripción	Formato	Valor por Defecto
PrediosKey	Llave primaria de la dimensión	Int(11)	Ninguno
Codpre	Código del predio	Int(11)	Ninguno
Anio_Fiscal	Año	Int(11)	Ninguno
Tipo	Rural / Urbano	Varchar(1)	Ninguno
Cp_id	Código de tipo propietario	Varchar(2)	Ninguno
Cp_descripcion	Nombre de tipo propietario	Varchar(30)	Ninguno
Fecha_registro	Fecha registro del predio	Timestamp	Ninguno
Motivo	Código Motivo	Int(11)	Ninguno
Calle	Número de calle	Int(11)	Ninguno
Nro_cuadra	Número de cuadra	Varchar(2)	Ninguno
Ep_id	Código Estado del predio	Varchar(2)	Ninguno
Ep_descripcion	Nombre Estado del predio	Varchar(30)	Ninguno
Up	Código Uso	Int(11)	Ninguno
Up_descripcion	Nombre Uso	Varchar(30)	Ninguno
Area	Area	Float	Ninguno

Fuente: Elaboración propia

f. dim_tenencia

Descripción:

Contiene información sobre la tenencia de terrenos.

Atributos:

Cuadro 18 – Atributos dim_tenencia

Nombre del Atributo	Descripción	Formato	Valor por Defecto
TenenciaKey	Llave primaria de la dimensión	Int(11)	Ninguno
Cod_tenencia	Código de tenencia	Int(11)	Ninguno
Descripción	Descripción del código de tenencia	Varchar(30)	Ninguno

Fuente: Elaboración propia

g. dim_tiempo

Descripción:

Contiene la fecha en la que el contribuyente realiza el pago de predios.

Atributos:

Cuadro 19 – Atributos dim_tiempo

Nombre del Atributo	Descripción	Formato	Valor por Defecto
TiempoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Fecha	Fecha	Timestamp	Ninguno
Anio	Año	Text	Ninguno
Semestre	Semestre	Text	Ninguno
Trimestre	Trimestre	Double	Ninguno
Mes	Mes	Text	Ninguno
Semana_anio	Semana del año	Double	Ninguno
Dia	Día	Double	Ninguno
Dia_semana	Día de la semana	Text	Ninguno
Temporada	Temporada	Text	Ninguno

Fuente: Elaboración propia

h. dim_tipo_terreno

Descripción:

Contiene información sobre el tipo de terreno que posee un contribuyente

Atributos:

Cuadro 20 – Atributos dim_tipo_terreno

Nombre del Atributo	Descripción	Formato	Valor por Defecto
TTerrenoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Tipo_id	Código del tipo de terreno	Varchar(2)	Ninguno
Descripción	Nombre del tipo de terreno	Varchar(20)	Ninguno

Fuente: Elaboración propia

i. dim_uso

Descripción:

Contiene la información del tipo de uso que se le da al terreno.

Atributos:

Cuadro 21 – Atributos dim_uso

Nombre del Atributo	Descripción	Formato	Valor por Defecto
UsoKey	Llave primaria de la dimensión	Int(11)	Ninguno
Uso	Nombre del Uso	Varchar(30)	Ninguno
Tipo	Codigo de Uso	Int(11)	Ninguno

Fuente: Elaboración propia

Tablas Hechos

Existen dos tablas de hechos:

- a. fact_recaudacion
- b. fact_recaudacion_predial

A continuación se detalla cada una de las tablas hechos.

a. fact_recaudacion

Descripción:

Contiene la información general de las recaudaciones.

Granularidad:

Cuadro 22 – Granularidad fact_recaudacion

N°	Nombre de la Dimensión	Descripción	Llave Primaria
1	Dim_contribuyente	Datos contribuyente	Si
2	Dim_concepto	Información de rubro	Si
3	Dim_partida	Tipo partida	Si
4	Dim_fuente_financiamiento	Tipo fuente de financiamiento	Si
5	Dim_tiempo	Fecha en que se realizó el pago	Si

Fuente: Elaboración propia

Medidas:

Cuadro 23 – Medidas fact_recaudacion

N°	Nombre	Descripción
1	Montos_Cancelados	Cantidad de montos cancelados
2	Cartera_Clientes	Número de clientes puntuales

Fuente: Elaboración propia

b. fact_recaudacion_predial

Descripción:

Contiene la información solo de las recaudaciones prediales.

Granularidad:

Cuadro 24 – Granularidad fact_recaudacion_predial

Nº	Nombre de la Dimensión	Descripción	Llave Primaria
1	Dim_contribuyente	Datos contribuyente	Si
2	Dim_prediosyotros	Datos de predios	Si
3	Dim_partida	Tipo partida	Si
4	Dim_fuente_financiamiento	Tipo fuente de financiamiento	Si
5	Dim_tiempo	Fecha en que se realizó el pago	Si
6	Dim_uso	Tipo de uso de terreno	Si
7	Dim_tipo_terreno	Tipo de terreno	Si
8	Dim_tenencia	Numero de tenencia de terreno	Si

Fuente: Elaboración propia

Medidas:

Cuadro 25 – Medidas fact_recaudacion_predial

Nº	Nombre	Descripción
1	Montos_cancelados	Cantidad de montos cancelados
2	Cartera_predios	Número de predios puntuales

Fuente: Elaboración propia

IV.7.ETL

IV.7.1. Extracción

El proceso de extracción se da desde el sistema fuente llamado bdmunicipal al destino municipalidaddimension.

Cuadro 26 – Especificación de ETL desde bdmunicipal a municipalidaddimension

Fuente	BD bdmunicipal	
Destino	BD municipalidaddimension	
	bdmunicipal	BD municipalidaddimension
Transferencias	Sat.rubro	Dim_concepto
	Sat.contribuyente	Dim_contribuyente
	Sat.ff	Dim_fuente_financiamiento
	Sat.s_especifica_det	Dim_partida
	Sat.predio	Dim_prediosyotros
	Sat.tenencia	Dim_tenencia
	Sat.caja	Dim_tiempo
	Dominios.usoprerural	Dim_tipo_terreno
	Sat.uso_predio	Dim_uso
	Sat.caja, sat.predio	Dim_recaudacion
	Sat.caja, sat.predio	Dim_recaudacion_predial
Características de la transferencia	Transferir todos los campos seleccionados..	

Fuente: Elaboración propia

Especificar permisos y accesos de los usuarios

El siguiente cuadro nos indica los permisos y los accesos de los tipos de usuarios sobre las bases de datos.

Cuadro 27 – Permisos y accesos de los usuarios

Tipo de usuario		Permisos	
		Bases de Datos	
		BD Transaccional	BD Data mart
Tomador de Decisiones	Sub Gerente Administración Tributaria	N/A	Lectura a todas las tablas
Área Informática	Administrador Data mart	Lectura a todas las tablas	Lectura / Escritura en todas las Tablas

Fuente: Elaboración propia

IV.7.2. Transformación

La transformación ETL llamada TransformacionDimension es la cual contiene las transformaciones de las dimensiones,y cuenta con tres pasos:

- Entrada Tabla
- Ordenar Filas
- Salida Tabla

Dim_fuente_financiamiento:

Figura 17 – Transformación de Dim_fuente_financiamiento

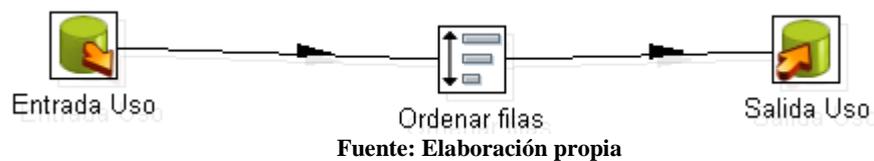


La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
  codff
, ff
FROM sat.ff
```

Dim_uso:

Figura 18 - Transformación de Dim_uso

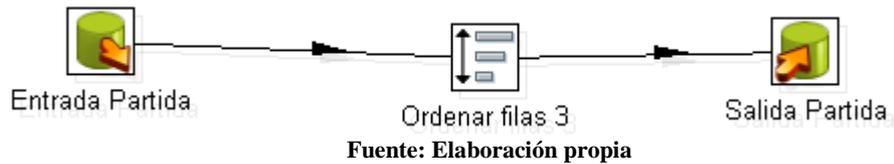


La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
  coduso
, uso
FROM sat.uso_predio
```

Dim_partida:

Figura 19 - Transformación de Dim_partida

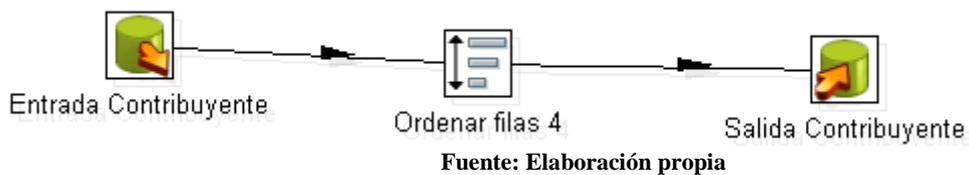


La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
tipo_transaccion, generica, subgenerica, subgenerica_det
, especifica, especifica_det, descripcion, estado, codpar
FROM sat.s_especifica_det
```

Dim_contribuyente:

Figura 20 - Transformación de Dim_contribuyente



La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
codper, codvia, conyugue, ciudad, contrib, direccion, tipo
FROM sat.contribuyente
```

Dim_tenencia:

Figura 21 - Transformación de Dim_tenencia

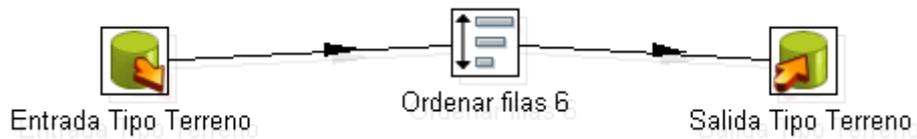


La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
  codten
, desten
FROM sat.tenencia
```

Dim_tipo_terreno:

Figura 22 - Transformación de Dim_tipo_terreno



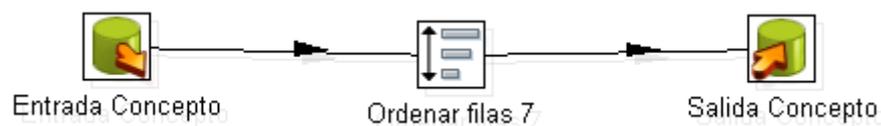
Fuente: Elaboración propia

La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
  codusoprerural
, desusoprerural
FROM dominios.usoprerural
```

Dim_concepto:

Figura 23 - Transformación de Dim_concepto



Fuente: Elaboración propia

La consulta para la extracción, limpieza y carga de datos es:

```
SELECT
  codrub, pres
, rubro, codest
FROM sat.rubro
```

Dim_prediosyotros:

Figura 24 - Transformación de Dim_prediosyotros



La consulta para la extracción, limpieza y carga de datos es:

```
SELECT codpredio,aniofiscal,fcereg,motivo,calle,nrocuadra,area,p.cp,  
cc.descontitcatastral,p.ep , ec.desecc, up, us.uso, p.tipo  
from sat.predio p JOIN dominios.contitcatastral cc  
ON(p.cp=cc.codcontitcatastral)  
JOIN dominios.ecc ec ON(p.ep= ec.codecc)  
JOIN sat.uso_predio us ON (p.up = us.coduso)
```

Dim_tiempo:

Figura 25 - Transformación de Dim_tiempo



La consulta para la extracción, limpieza y carga de datos es:

```

SELECT DISTINCT extract(YEAR FROM fecha) AS ANIO,

    CASE extract(QUARTER FROM fecha)
    WHEN 1 THEN '1'      WHEN 2 THEN '1'
    WHEN 3 THEN '2'      WHEN 4 THEN '2'
    END AS SEMESTRE,
extract(QUARTER FROM fecha) AS TRIMESTRE,
CASE extract(MONTH FROM fecha)
WHEN 1 THEN 'ENERO'    WHEN 2 THEN 'FEBRERO'
WHEN 3 THEN 'MARZO'    WHEN 4 THEN 'ABRIL'
WHEN 5 THEN 'MAYO'     WHEN 6 THEN 'JUNIO'
WHEN 7 THEN 'JULIO'    WHEN 8 THEN 'AGOSTO'
WHEN 9 THEN 'SEPTIEMBRE'
WHEN 10 THEN 'OCTUBRE'
    WHEN 11 THEN 'NOVIEMBRE'
    WHEN 12 THEN 'DICIEMBRE'
END
AS MES,
extract(WEEK FROM fecha) AS SEMANA_ANIO,
extract(DAY FROM fecha) AS DIA,

CASE extract(DOW FROM fecha )
WHEN 1 THEN 'LUNES'
WHEN 2 THEN 'MARTES'
WHEN 3 THEN 'MIERCOLES'
WHEN 4 THEN 'JUEVES'
WHEN 5 THEN 'VIERNES'
    WHEN 6 THEN 'SABADO'
    WHEN 7 THEN 'DOMINGO'
END AS DIA_semana,
CASE extract(MONTH FROM fecha)
WHEN 1 THEN 'VERANO'
WHEN 2 THEN 'VERANO'
WHEN 3 THEN 'VERANO'
WHEN 4 THEN 'OTONIO'
WHEN 5 THEN 'OTONIO'
WHEN 6 THEN 'OTONIO'
WHEN 7 THEN 'INVIERNO'
    WHEN 8 THEN 'INVIERNO'
    WHEN 9 THEN 'INVIERNO'
WHEN 10 THEN 'PRIMAVERA'
    WHEN 11 THEN 'PRIMAVERA'
    WHEN 12 THEN 'PRIMAVERA'
END AS TEMPORADA,fecha

FROM sat.caja

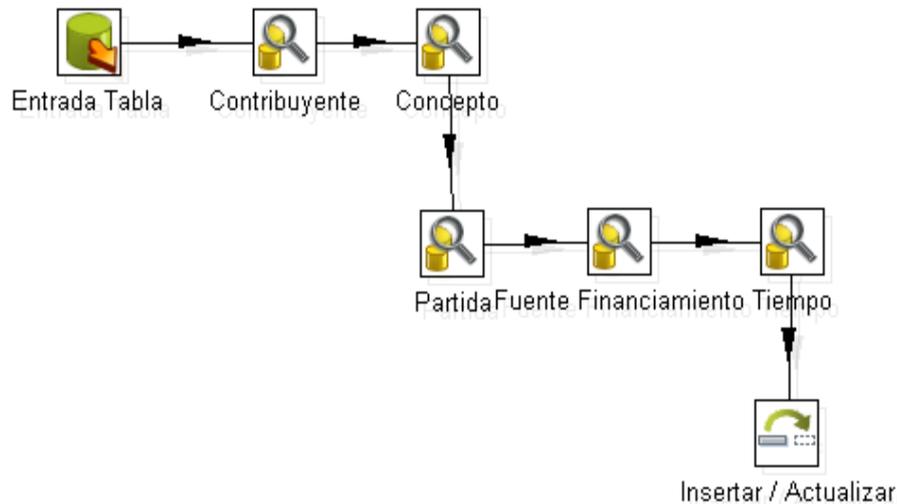
```

La transformación ETL de la tabla hechos dim_recaudacion llamada TransformacionRecaudacion cuenta con tres pasos:

- Entrada Tabla
- Búsqueda en Base de Datos
- Insertar / Actualizar

Dim_recaudacion:

Figura 26 – Transformación de Dim_recaudacion



Fuente: Elaboración propia

La consulta para la extracción, limpieza y carga de datos es:

```

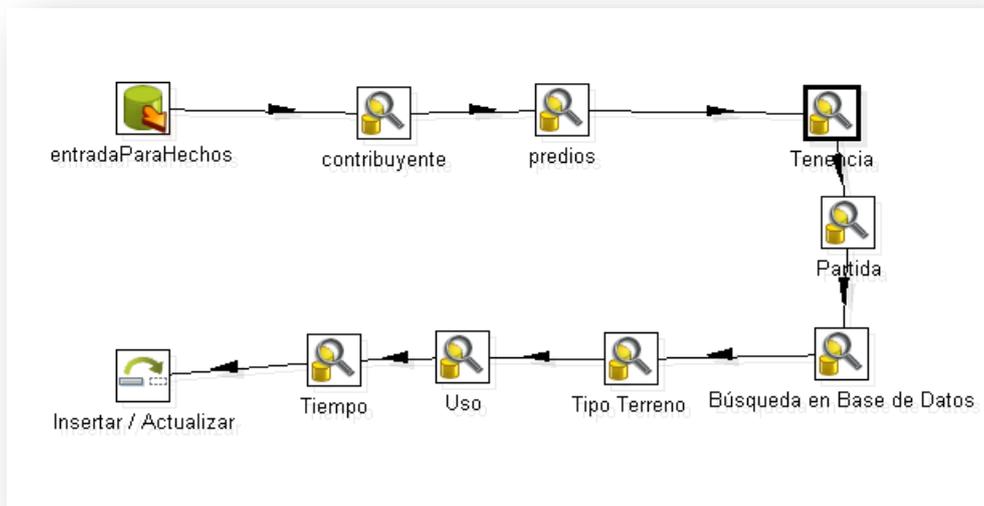
SELECT
c.fecha AS FECHA,cb.codper AS CODIGOPERSONA,rb.codrub AS
Codigorubro,rb.codpar AS CodigoPartida, ff.codff AS
Fuentefinanciamiento, sum(cd.monto) AS MONTOCANCELADO,
(SELECT count(DISTINCT(codper)) FROM sat.caja )AS
CARTERACLIENTES
FROM sat.caja c
  JOIN sat.cajad cd ON (c.codigo= cd.codigo)
  JOIN  sat.contribuyente cb ON(c.codper= cb.codper)
  JOIN  sat.rubro rb ON (rb.codrub= cd.codrub)
  JOIN  sat.s_especifica_det sed ON (sed.codpar= rb.codpar)
  JOIN  sat.ff ff ON(ff.codff= sed.codff)
  JOIN  sat.cuentacorriente cc ON (cc.cont= cb.codper)
  JOIN  sat.predio pr ON(pr.contri= cb.codper )
WHERE fecha<'2012-01-01'  AND cc.estado=1
GROUP BY
c.fecha,cb.codper,pr.codpredio
,rb.codrub,rb.codpar,
ff.codff
ORDER BY c.fecha
  
```

La transformación ETL de la tabla hechos dim_recaudacion_predial llamada TransformacionRecaudacion_Predial cuenta con tres pasos:

- Entrada Tabla
- Búsqueda en Base de Datos
- Insertar / Actualizar

Dim_recaudacion_predial:

Figura 27 – Transformación de Dim_recaudacion_predial



Fuente: Elaboración propia

La consulta para la extracción, limpieza y carga de datos es:

```

SELECT c.fecha AS FECHA,
cb.codper AS CODIGOPERSONA,pr.codpredio AS CODIGOPREDIO,
tt.codusoprerural AS TipoTerreno,te.codten AS TENENCIA,
us.coduso AS UsoPredio, ff.codff AS Fuentefinanciamiento,
sed.codpar AS CodPartida, sum(cd.monto) AS RESULTADO
, (select count(distinct(codper))from sat.caja) AS CARTERACLIENTES

FROM sat.caja c

JOIN sat.cajad cd ON (c.codigo= cd.codigo)
JOIN sat.contribuyente cb ON(c.codper= cb.codper)
JOIN sat.predio pr ON(pr.contri= cb.codper )
JOIN dominios.usoprerural tt ON(tt.codusoprerural=pr.tipter)
JOIN sat.tenencia te ON(te.codten=pr.tenencia)
JOIN sat.uso_predio us ON (us.coduso=pr.up)
JOIN sat.ff ff ON(ff.codff=8)
JOIN sat.s_especifica_det sed ON (sed.codpar='122111')
JOIN sat.cuentacorriente cc ON (cc.cont= cb.codper)

WHERE cc.estado=1
AND sed.codpar='122111'
AND fecha <= '2012-01-01'

GROUP BY
c.fecha,cb.codper ,pr.codpredio ,tt.codusoprerural ,te.codten
,us.coduso ,ff.codff
,sed.codpar
ORDER BY c.fecha

```

IV.7.3. Carga

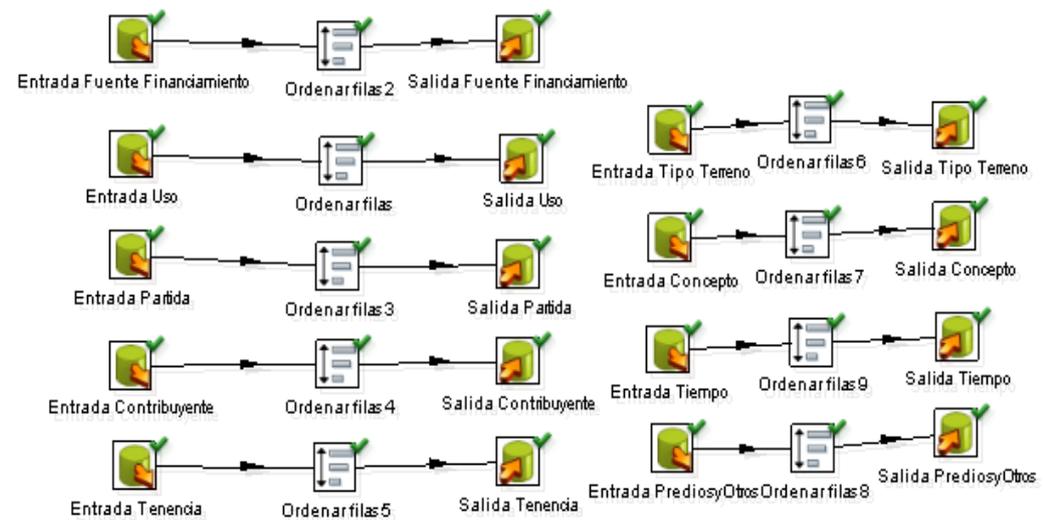
Los elementos usados para el proceso de carga son los siguientes:

- Entrada Tabla.
- Salida Tabla.
- Búsqueda en Base de Datos.
- Insertar / Actualizar.

Carga de Dimensiones

A continuación se muestra la carga de las dimensiones.

Figura 28 – Carga de Dimensiones



Fuente: Elaboración propia

Resultados de la carga de Dimensiones

Figura 29 – Resultados de la carga de dimensiones

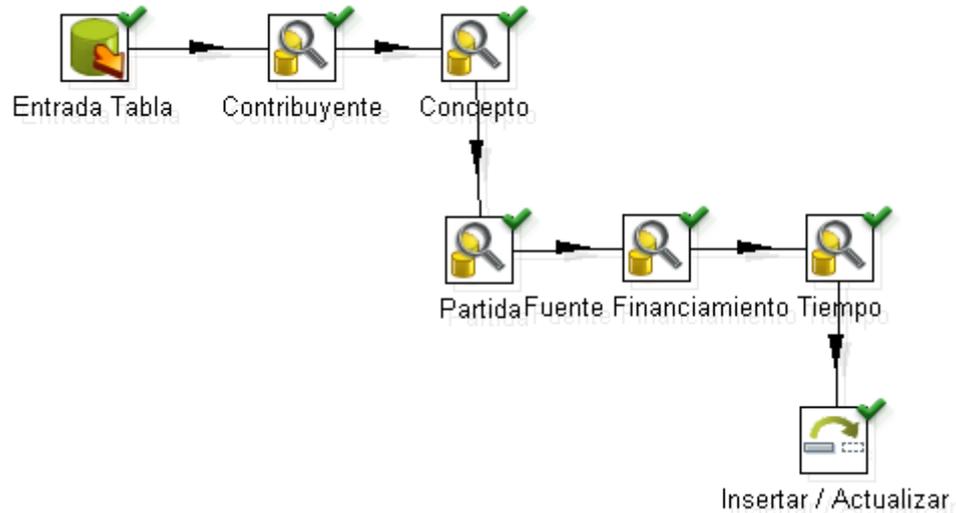
Execution Results													
Execution History Logging Step Metrics Performance Graph													
#	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo	Velocidad (r/s)	Pr[E]S
1	Entrada Fuente Financiamiento	0	0	14	14	0	0	0	0	0	Finalizado	2.5s	6
2	Ordenar filas 2	0	14	14	0	0	0	0	0	0	Finalizado	2.1s	6
3	Salida Fuente Financiamiento	0	14	14	0	14	0	0	0	0	Finalizado	5.8s	2
4	Entrada Uso	0	0	29	29	0	0	0	0	0	Finalizado	2.3s	12
5	Ordenar filas	0	29	29	0	0	0	0	0	0	Finalizado	2.1s	14
6	Salida Uso	0	29	29	0	29	0	0	0	0	Finalizado	5.8s	5
7	Entrada Partida	0	0	864	864	0	0	0	0	0	Finalizado	2.3s	369
8	Ordenar filas 3	0	864	864	0	0	0	0	0	0	Finalizado	2.1s	409
9	Salida Partida	0	864	864	0	864	0	0	0	0	Finalizado	20.5s	42
10	Entrada Contribuyente	0	0	8439	8439	0	0	0	0	0	Finalizado	3.0s	2,813
11	Ordenar filas 4	0	8439	8439	0	0	0	0	0	0	Finalizado	3.0s	2,798
12	Salida Contribuyente	0	8439	8439	0	8439	0	0	0	0	Finalizado	35.3s	239
13	Entrada Tenencia	0	0	8	8	0	0	0	0	0	Finalizado	2.3s	3
14	Ordenar filas 5	0	8	8	0	0	0	0	0	0	Finalizado	2.1s	4
15	Salida Tenencia	0	8	8	0	8	0	0	0	0	Finalizado	5.8s	1
16	Entrada Tipo Terreno	0	0	8	8	0	0	0	0	0	Finalizado	2.3s	3
17	Ordenar filas 6	0	8	8	0	0	0	0	0	0	Finalizado	2.1s	4
18	Salida Tipo Terreno	0	8	8	0	8	0	0	0	0	Finalizado	5.8s	1
19	Entrada Concepto	0	0	276	276	0	0	0	0	0	Finalizado	2.3s	118
20	Ordenar filas 7	0	276	276	0	0	0	0	0	0	Finalizado	2.1s	129
21	Salida Concepto	0	276	276	0	276	0	0	0	0	Finalizado	20.5s	13
22	Entrada PrediosyOtros	0	0	63000	63000	0	0	0	0	0	Finalizado	22.5s	2,800
23	Ordenar filas 8	0	63000	63000	0	0	0	0	0	0	Finalizado	1mn 4s	979
24	Salida PrediosyOtros	0	63000	63000	0	63000	0	0	0	0	Finalizado	1mn 5s	913
25	Entrada Tiempo	0	0	52294	52294	0	0	0	0	0	Finalizado	27.4s	1,906
26	Ordenar filas 9	0	52294	52294	0	0	0	0	0	0	Finalizado	58.0s	901
27	Salida Tiempo	0	52294	52294	0	52294	0	0	0	0	Finalizado	1mn 5s	801

Fuente: Elaboración propia

Carga de Tabla Hechos Dim_recaudacion:

A continuación se muestra la carga de la tabla hechos dim_recaudacion.

Figura 30 – Carga de tabla hechos dim_recaudacion



Fuente: Elaboración propia

Resultados de la Carga de Tabla Hechos Dim_recaudacion:

Figura 31 – Resultados de la carga de tabla hechos dim_recaudacion

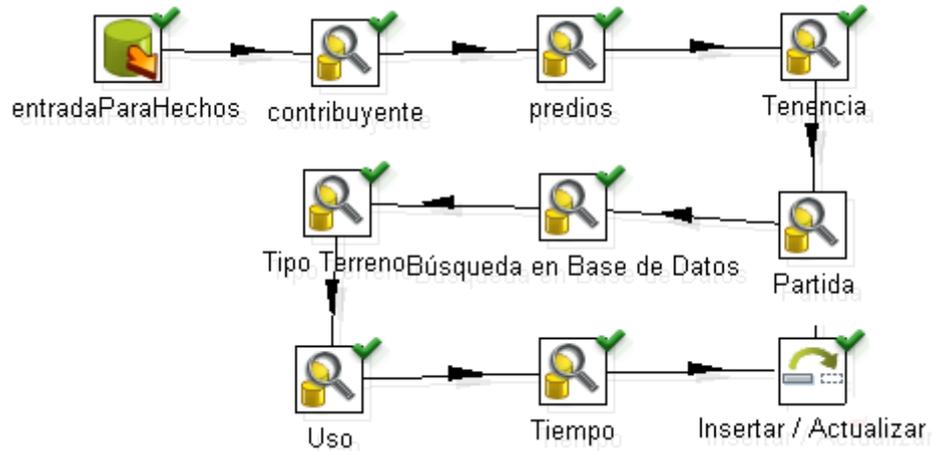
Execution Results													
#	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo	Velocidad (r/s)	Pr[E]S
1	Entrada Tabla	0	0	53717	53717	0	0	0	0	Finalizado	14mn 21s	62	-
2	Contribuyente	0	53717	53717	53717	0	0	0	0	Finalizado	14mn 26s	62	-
3	Concepto	0	53717	53717	53717	0	0	0	0	Finalizado	16mn 40s	54	-
4	Partida	0	53717	53717	53717	0	0	0	0	Finalizado	21mn 36s	41	-
5	Fuente Financiamiento	0	53717	53717	53717	0	0	0	0	Finalizado	27mn 42s	32	-
6	Tiempo	0	53717	53717	53717	0	0	0	0	Finalizado	34mn 52s	26	-
7	Insertar / Actualizar	0	53717	53717	53717	0	40664	0	0	Finalizado	34mn 52s	26	-

Fuente: Elaboración propia

Carga de Tabla Hechos Dim_recaudacion_predial:

A continuación se muestra la carga de la tabla hechos dim_recaudacion_predial.

Figura 32 – Carga de tabla hechos dim_recaudacion_predial



Fuente: Elaboración propia

Resultados de la Carga de Tabla Hechos Dim_recaudacion_predial:

Figura 33 – Resultados de la carga de tabla hechos dim_recaudacion_predial

Execution Results													
Execution History Logging Step Metrics Performance Graph													
	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo	Velocidad (/s)	Pri[E]S
1	Entrada Tabla	0	0	970	970	0	0	0	0	Finalizado	3m 24s	208	-
2	Contribuyente	0	970	970	970	0	0	0	0	Finalizado	4m 9s	171	-
3	Concepto	0	970	970	970	0	0	0	0	Finalizado	4m 42s	151	-
4	Partida	0	970	970	970	0	0	0	0	Finalizado	8m 32s	83	-
5	Fuente Financiamiento	0	970	970	970	0	0	0	0	Finalizado	13m 50s	51	-
6	Tiempo	0	970	970	970	0	0	0	0	Finalizado	20m 46s	34	-
7	Insertar / Actualizar	0	970	970	970	900	0	0	0	Finalizado	20m 46s	34	-

Fuente: Elaboración propia

IV.8. Diseño de aplicaciones BI

IV.8.1.Desarrollo de Cubos Multidimensionales

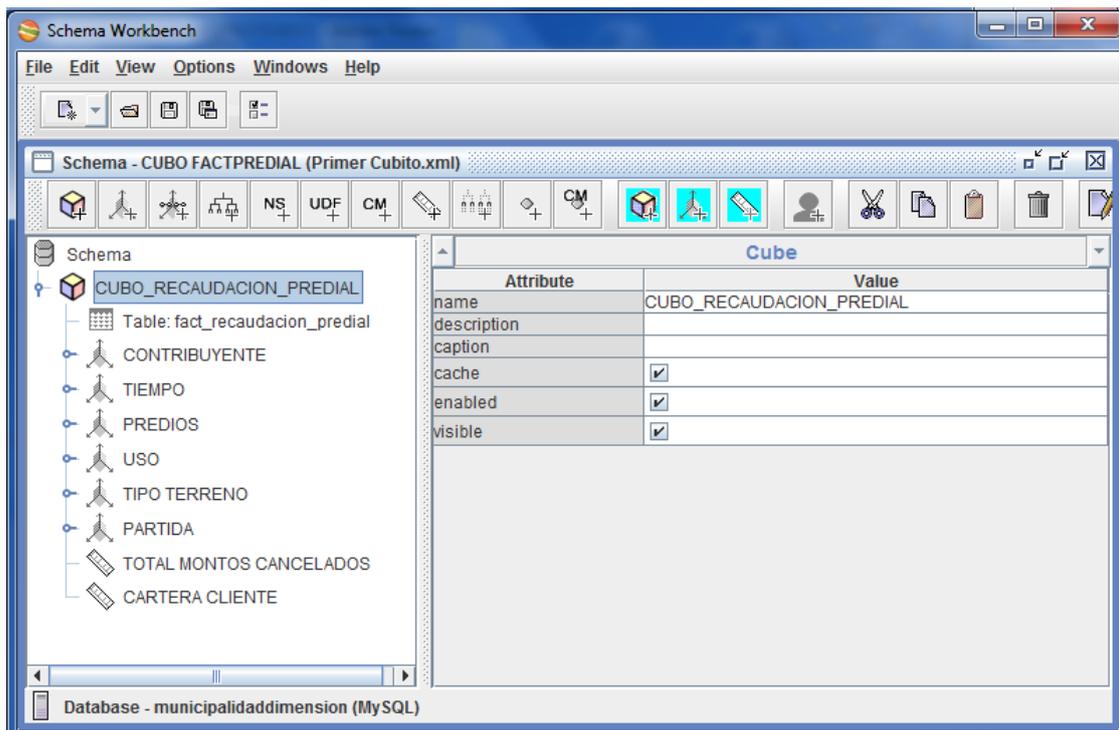
Luego que el data mart se encuentra poblado se procede a la creación de cubos, mediante la herramienta Schema workbench para ello se toma la información contenida en el data mart, para luego construir cubos acorde a los requerimientos de la empresa y como producto final se construyen los reportes.

El proyecto cuenta con dos cubos:

- CUBO_RECAUDACION_PREDIAL
- CUBO_RECAUDACION

Diseño del CUBO_RECAUDACION_PREDIAL:

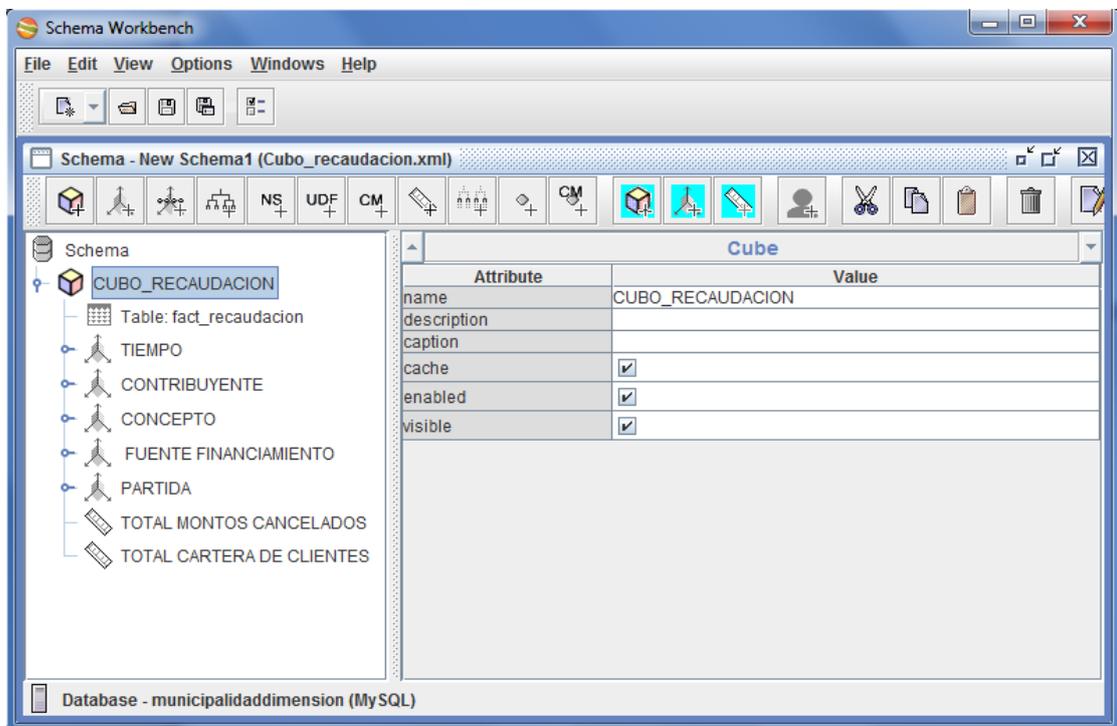
Figura 34 – Diseño del CUBO_RECAUDACION_PREDIAL



Fuente: Elaboración propia

Diseño del CUBO_RECAUDACION:

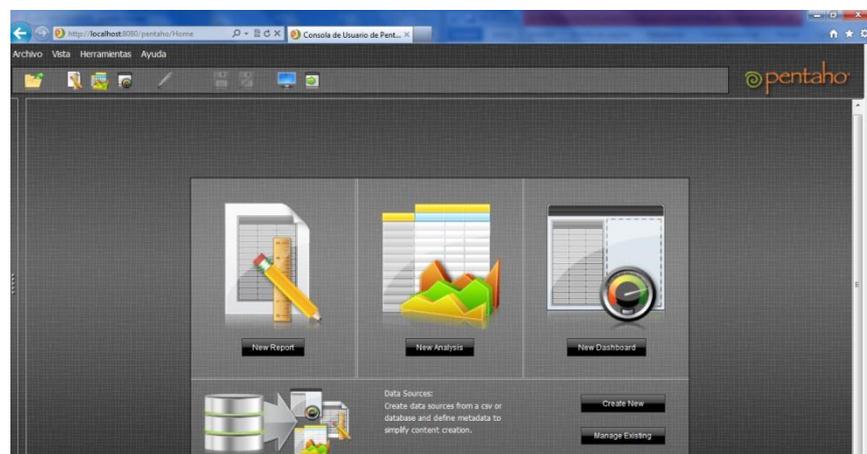
Figura 35 – Diseño del CUBO_RECAUDACION



Fuente: Elaboración propia

Para que los usuarios finales puedan visualizar los datos del cubo y reportes, harán uso del User Console Login.

Figura 36 - Página principal User Console Login



Fuente: Elaboración propia

Reportes Generados a partir del Cubo Recaudación Predial

- Reporte de contribuyentes generado a partir del Cubo Recaudación Predial

Figura 37 – Reporte de contribuyentes generado desde Cubo Recaudacion Predial

The screenshot shows the Pentaho Analyzer interface. The 'Available fields' pane on the left lists fields from the 'CUBO_RECAUDACION_PREDI' cube, including 'CONTRIBUYENTES', 'CODIGO DE PREDIO', and 'USO'. The 'Layout' pane shows these fields arranged in the report. The main report area displays an 'Unsaved Report' with the following data:

CONTRIBUYENTES	CODIGO DE PREDIO	USO	TOTAL MONTOS CANCELADOS
ABANTO CARRERA AURELIO	3716	26. Predio Rústico	204,6
ABANTO FUENTES LELIS GIULIANI	38	1. Casa Habitación	28.697,2
ABANTO MURRUGARRA SAANTIAGO	3977	26. Predio Rústico	14.694,4
ABANTO SALAZAR CARLOS MANUEL	3978	26. Predio Rústico	14.694,4
ABANTO VILLANUEVA JAJME FORTUNAT	846	1. Casa Habitación	3.964,8
ACOSTA ALCANTARA JOSE ALBERTO	3783	26. Predio Rústico	2.154
AEROPUERTOS DEL PERU S.A.	-4241	1. Casa Habitación	280
AGUILAR BARBA LUIS ADAN	4241	1. Casa Habitación	1.120
AGUILAR BARBA NELSON NAPOLEÓN	4217	24. Aeronavegación	1.095.456,6
AGUILAR JAVE WILLIAM RAMON	4309	19. Terreno sin Construir	6.343,2
AGUILAR REQUELME JUAN VICTOR	11694	19. Terreno sin Construir	2.819,2
	485	1. Casa Habitación	26.089,14
	4074	19. Terreno sin Construir	516,6
	3903	26. Predio Rústico	1.746
	3904	1. Casa Habitación	1.746
	3905	26. Predio Rústico	1.746

Fuente: Elaboración propia

- Reporte de pagos de predios por año

Figura 38 – Reporte de pagos de predios por año

The screenshot shows the Pentaho Analyzer interface. The 'Available fields' pane on the left lists fields from the 'CUBO_RECAUDACION_PREDI' cube, including 'AÑO', 'CODIGO DE PREDIO', and 'CONTRIBUYENTES'. The 'Layout' pane shows these fields arranged in the report. The main report area displays an 'Unsaved Report' with the following data:

AÑO	CODIGO DE PREDIO	CONTRIBUYENTES	TOTAL MONTOS CANCELADOS
	3840	LUICHO CHUNQUE ALEJANDRO	861
	3841	LUICHO CHUNQUE ALEJANDRO	861
	3842	LUICHO CHUNQUE ALEJANDRO	861
	3843	LUICHO CHUNQUE ALEJANDRO	861
	3844	NURÉÑA MONTENEGRO ANA NOEMÍ	4.935,84
	3845	NURÉÑA MONTENEGRO OLGA MARINA	922,8
	3846	NURÉÑA MONTENEGRO ANA NOEMÍ	4.935,84
	3849	TASILLA CHINGAY JOSE ANTONIO	172,2
	3850	VALENCY FLORES RAUL JOSAFAT	11.026,4
	3851	LLANOS ALAYA JUAN	861
	3852	CABELLOS INCIL MARIA JULIA	172,2
	3853	CABELLOS INCIL MARIA JULIA	172,2
	3854	CABELLOS INCIL MARIA JULIA	172,2
	3855	BURGA CHACON EUFEMIA	11.026,4
	3857	LEON CASTILLO VICTOR RAUL	28,7
	3858	LEON CASTILLO VICTOR RAUL	28,7

Fuente: Elaboración propia

Reportes Generados a partir del Cubo Recaudación

Figura 39 – Reporte de contribuyentes y montos cancelados

The screenshot shows the Pentaho Analyzer Report interface. On the left, there is a tree view of available fields for the 'CUBO_RECAUDACION' cube, including categories like CONCEPTO, CONTRIBUYENTES, FUENTE FINANCIAMIENTO, Measures, PARTIDA, and TIEMPO. The 'Measures' section has 'TOTAL MONTOS CANCELADOS' selected. The 'Layout' pane shows 'CODIGO DE CONTRIBUYENTE' and 'NOMBRE CONTRIBUYENTE' in the Rows section, and 'TOTAL MONTOS CANCELADOS' in the Measures section. The main area displays an 'Unsaved Report' with a table of 666 rows and 1 column. The table has three columns: 'CODIGO DE CONTRIBUYENTE', 'NOMBRE CONTRIBUYENTE', and 'TOTAL MONTOS CANCELADOS'. The data is as follows:

CODIGO DE CONTRIBUYENTE	NOMBRE CONTRIBUYENTE	TOTAL MONTOS CANCELADOS
5675	ASERRADERO Y CARPINTERIA GONZALES E.I.R.L.	46.611.780,1
7031	ASOCIACION "LAS ASAMBLEAS DE DIOS DEL PERU"	48
1072	ASOCIACION DE LOS TESTIGOS DE JEHOVA	13.264
6300	ASOCIACION DE PRODUCTORES AGROPECUARIOS DE	630,9
7616	ASOCIACION DE PRODUCTORES MICROCUENCA CHON	117,2
8246	ASOCIACION EVANGELICA JESUS DE NAZARETH DE T	1.024
6408	ASOCIACION PARA LA DEFENSA Y DESARROLLO SAN	1.472,1
4585	ASOCIACION PRO VIVIENDA SANTA BARBARA.	17.784
6942	ASOCIACION PROMOTORA DE SERVICIOS EDUCATIV	49.645,2
5450	ASOCIACION SANTA DOROTEA	1.200
5155	ATALAYA ATALAYA NICOLAS	5.754
6971	ATALAYA CABANILLAS FACUNDO	2.637
4122	AURAZO DIAZ HENRY MARTIN	2.667,5
8042	AURAZO REQUEJO IVAN EDUARDO	123.594,48
8469	AVELLANEDA ALARCON ROGER	32
9822	AVENDAÑO BENDEZU FERNANDO MARCELO	15.074,4

Fuente: Elaboración propia

IV.8.2. Elaboración de Reportes

Para generar reportes se utilizó la herramienta Pentaho Report Designer, **Para ello se toma la información contenida en el Datamart**, una vez los reportes estén terminados pueden ser exportados como archivos: PDF, rtf, html y otros.

- **Reportes de Contribuyentes**

Figura 40 – Reportes de Contribuyentes

octubre 05, 2012 @ 06:30

Reporte de Contribuyentes

Contribuyente	Key	Dirección	Montos_cancelados	Nombre	Anio
5	JR. INCA YUPANQUI Nº 250		11.652.945,025002865	INOSTROZA JAVE HILDA	2.007
6	CA. 02 URB. LAGUNA SECA Nº MZ-F LT-6		11.652.945,025002865	LINARES TERRONES BALDOMERO	2.007
7	CA. 05 URB. LAGUNA SECA Nº MZ-D LT-02		11.652.945,025002865	ABANTO FUENTES LELIS GIULIANI	2.007
8	CA. 05 URB. LAGUNA SECA Nº MZ-D LT-6A		11.652.945,025002865	SILVA DÍAZ MARÍA DE LOS ANGELES	2.007
9	PJE. TOPARPA Nº 135		11.652.945,025002865	MARQUEZ CÁCERES IDA VIOLETA	2.007
10	JR. YAHUAR HUACA Nº 155		11.652.945,025002865	RAICO GUTIERREZ MARIA MANUELA	2.007
11	JR. TUPAC INCA YUPANQUI Nº 145 -148		11.652.945,025002865	MESTANZA GUTIERREZ ANDRÉS	2.007
13	JR. MANCO INCA Nº 3/N		11.652.945,025002865	VENTURA LOZANO JUAN DEL CARMEN Y ESPOSA	2.007
14	AV. MANCO CAPAC Nº 313 - 316		11.652.945,025002865	VERA ROMERO ISAAC	2.007
15	JR. PACHACÚTEC Nº 425		11.652.945,025002865	CUZCO POTOSI JUAN EMILIO	2.007
16	AV. MANCO CAPAC Nº 375		11.652.945,025002865	GUTIERREZ EBAINÉ JORGE AMADO	2.007
18	JR. YAHUAR HUACA Nº 423		11.652.945,025002865	BARRANTES URTEAGA PURA MARÍA	2.007
19	JR. YAHUAR HUACA Nº 475		11.652.945,025002865	ARAUJO SÁNCHEZ GILBERTO	2.007
20	JR. CAPAC YUPANQUI Nº 300		11.652.945,025002865	MUÑOZ PABLO DANIEL HUMBERTO	2.007
21	JR. TUPAC INCA YUPANQUI 399		11.652.945,025002865	SALAZAR ALDAVE MARCO ANTONIO Y BOYD LLANOS MARTHA YSABEL	2.007
23	JR. PACHACÚTEC Nº 519		11.652.945,025002865	QUIROS ALCALDE DE CARRILLO JUSTA VIRGINIA	2.007
25	CA. 01 URB. HURTADO MILLER Nº MZ-J LT-16		11.652.945,025002865	HUALTIBAMBA CHINCHE PEDRO Y ESPOSA	2.007
26	JR. LLOQUE YUPANQUI Nº 251		11.652.945,025002865	SALINAS GUEVARA BALBINA	2.007
27	AV. MANCO CAPAC Nº 694		11.652.945,025002865	CABRERA CABELLOS FAUSTA	2.007

Fri Oct 05 18:30:44 EDT 2012

Fuente: Elaboración propia

- **Reporte de Fuentes de Financiamiento**

Figura 41 – Reporte de Fuentes de Financiamiento

octubre 16, 2012

Fuentes de Financiamiento

Descripcion	Tipo	Montos Recaudados
OTROS IMPUESTOS MUNICIPALES	8	3040.0
OTROS IMPUESTOS MUNICIPALES	8	3040.0
OTROS IMPUESTOS MUNICIPALES	8	3040.0
OTROS IMPUESTOS MUNICIPALES	8	3040.0
OTROS IMPUESTOS MUNICIPALES	8	3173.0
OTROS IMPUESTOS MUNICIPALES	8	3173.0
OTROS IMPUESTOS MUNICIPALES	8	3173.0
OTROS IMPUESTOS MUNICIPALES	8	3173.0
OTROS IMPUESTOS MUNICIPALES	8	5747.0
OTROS IMPUESTOS MUNICIPALES	8	973.0
OTROS IMPUESTOS MUNICIPALES	8	740.0
OTROS IMPUESTOS MUNICIPALES	8	761.0
OTROS IMPUESTOS MUNICIPALES	8	389.0
OTROS IMPUESTOS MUNICIPALES	8	389.0
OTROS IMPUESTOS MUNICIPALES	8	25057.0
OTROS IMPUESTOS MUNICIPALES	8	25696.0
OTROS IMPUESTOS MUNICIPALES	8	27039.0
OTROS IMPUESTOS MUNICIPALES	8	27361.0
OTROS IMPUESTOS MUNICIPALES	8	29220.0
OTROS IMPUESTOS MUNICIPALES	8	0.0
OTROS IMPUESTOS MUNICIPALES	8	0.0
OTROS IMPUESTOS MUNICIPALES	8	0.0
OTROS IMPUESTOS MUNICIPALES	8	0.0
OTROS IMPUESTOS MUNICIPALES	8	0.0
OTROS IMPUESTOS MUNICIPALES	8	27577.0
OTROS IMPUESTOS MUNICIPALES	8	27577.0
OTROS IMPUESTOS MUNICIPALES	8	27577.0
OTROS IMPUESTOS MUNICIPALES	8	27577.0
OTROS IMPUESTOS MUNICIPALES	8	25088.0
OTROS IMPUESTOS MUNICIPALES	8	25088.0

Tue Oct 16 23:18:03 COT 2012

Fuente: Elaboración propia

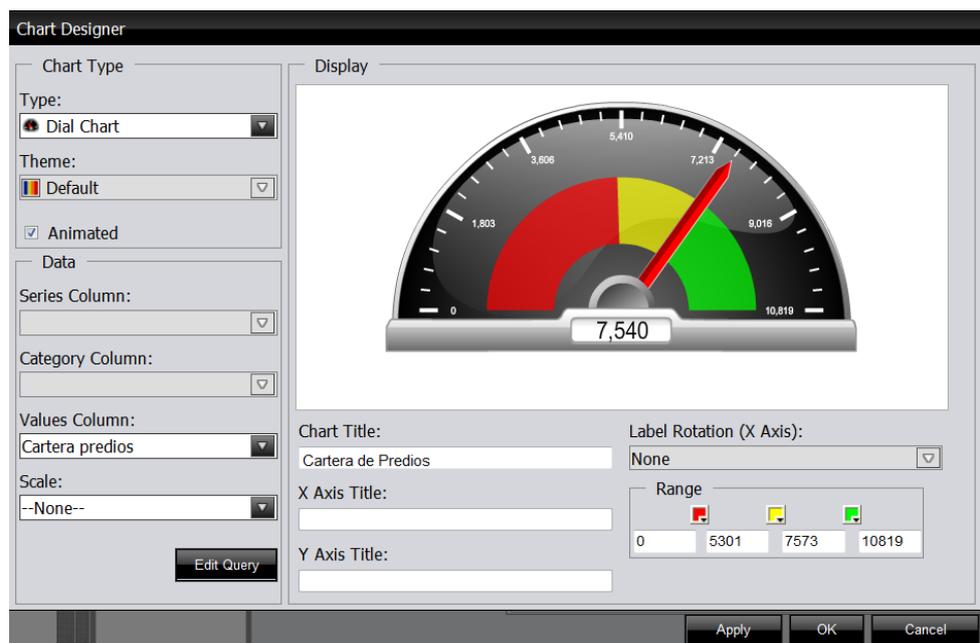
IV.8.3. KPI

Recaudación predial

El proyecto consistió en la implementación de un data mart, además se construyeron tableros de control con KPI, uno de estos tableros de control (dashboard) grafica: cartera de predios y montos cancelados de predios.

- **Cartera de Predios**

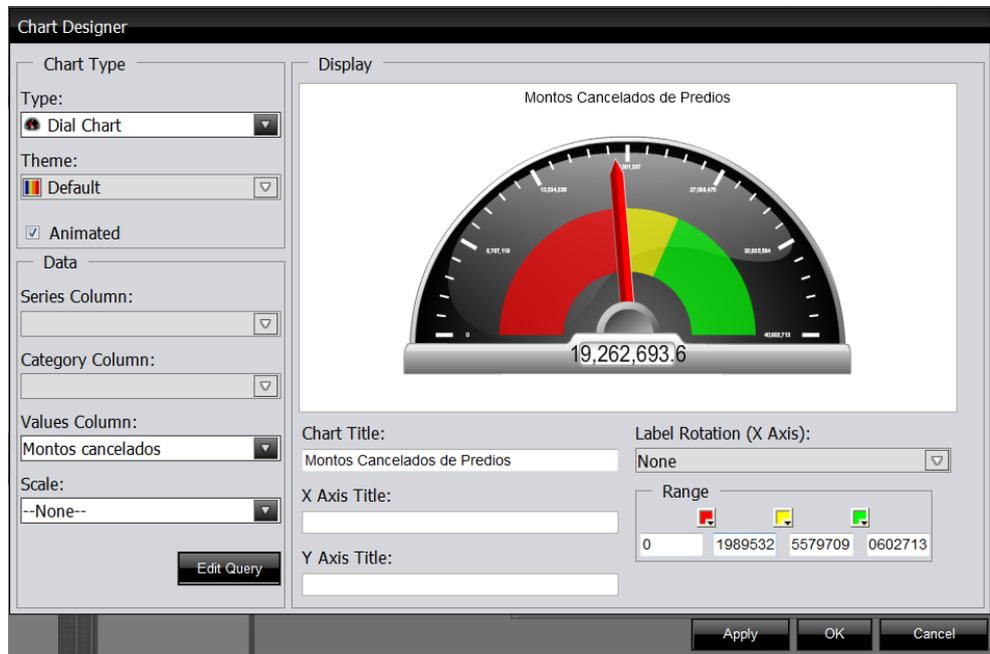
Figura 42 – Dashboard de Cartera de Predios



Fuente: Elaboración propia

- **Montos Cancelados de Predios**

Figura 43 – Dashboard de Montos Cancelados de Predios



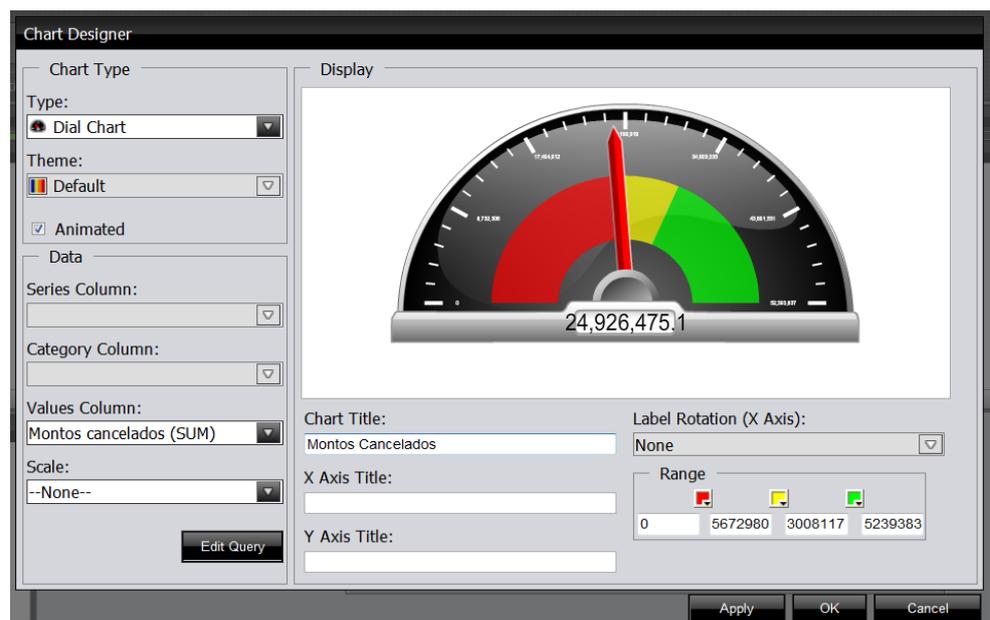
Fuente: Elaboración propia

Recaudación

En el siguiente dashboard se grafica montos cancelados de recaudaciones:

- **Montos Cancelados de Recaudaciones**

Figura 44 – Dashboard de Montos Cancelados de Recaudaciones



Fuente: Elaboración propia

CAPITULO V: DISCUSIÓN

El objetivo de este capítulo es mostrar los resultados obtenidos con la realización de la tesis, el estado actual del prototipo, su validación y las mejoras que se pueden hacer al mismo, así como las conclusiones que se obtuvieron de la aplicación a la institución.

Con esta tesis se comprueba la factibilidad de la aplicación en el proceso, mediante la metodología propuesta de Ralph Kimball, en la institución.

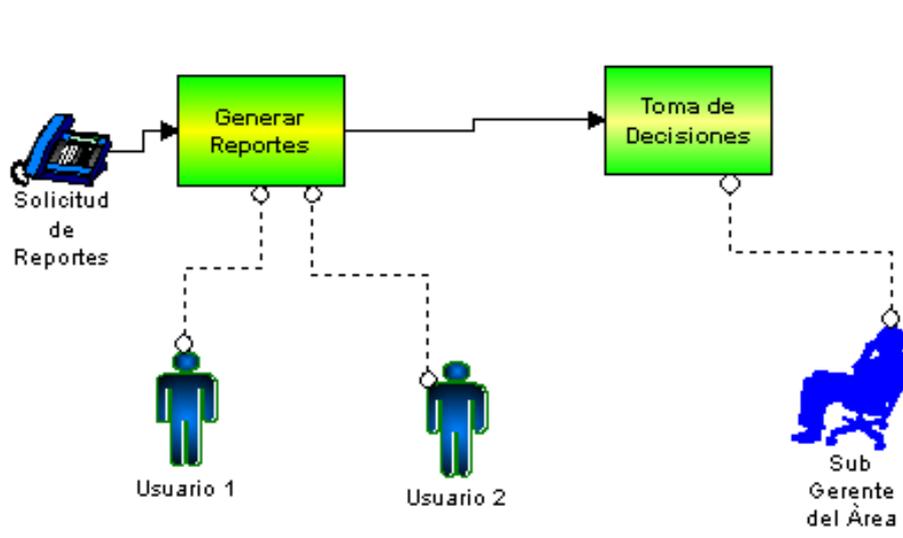
Dentro de las dificultades encontradas en el análisis del problema, se logró identificar que las principales dificultades que el área gerencial tenía para tomar decisiones eran:

- Tiempos largos para contar con información procesada y de calidad, por lo que tenían que hacer uso de más de un reporte para poder tomar decisiones.
- Dependencia del área de TI.

V.1. Proceso

A continuación se muestra el proceso administrativo en el área de Administración y Finanzas.

Figura 45 – Proceso Administrativo



Fuente: Elaboración propia

V.2. Resultado

Luego de la implementación del data mart se han realizado pruebas finales basados en la información que se obtiene de dicha herramienta, pasaremos a detallar los resultados obtenidos:



Donde:

- O1:** La toma de decisiones en el área de Administración y Finanzas antes de implementar el data mart (variable dependiente).
- X:** El desarrollo del data mart (variable independiente).
- O2:** La toma de decisiones en el área de Administración y Finanzas después de implementar el data mart.

En el pre test y post test se utilizó la distribución triangular, la cual es una herramienta que brinda mejores resultados para la interpretación:

Pre test:

Nro. De reportes por mes: T(59,60,61) horas

Nro. De Toma de decisiones por mes: T(95,96,97) horas

Post test:

Nro. De reportes por mes: T(23,24,25) horas

Nro. De Toma de decisiones por mes: T(47,48,49) horas

A continuación los resultados luego de la aplicación de distribución triangular:

O1:

Cuadro 28 - O1

Indicadores	Unidad de Medida	Valor
Nro. De reportes por mes	Número de Reportes	9
Nro. De Toma de decisiones por mes	Número de Decisiones	8

Fuente: Elaboración propia

O2:

Cuadro 29 - O2

Indicadores	Unidad de Medida	Valor
Nro. De reportes por mes	Número de Reportes	20
Nro. De Toma de decisiones por mes	Número de Decisiones	15

Fuente: Elaboración propia

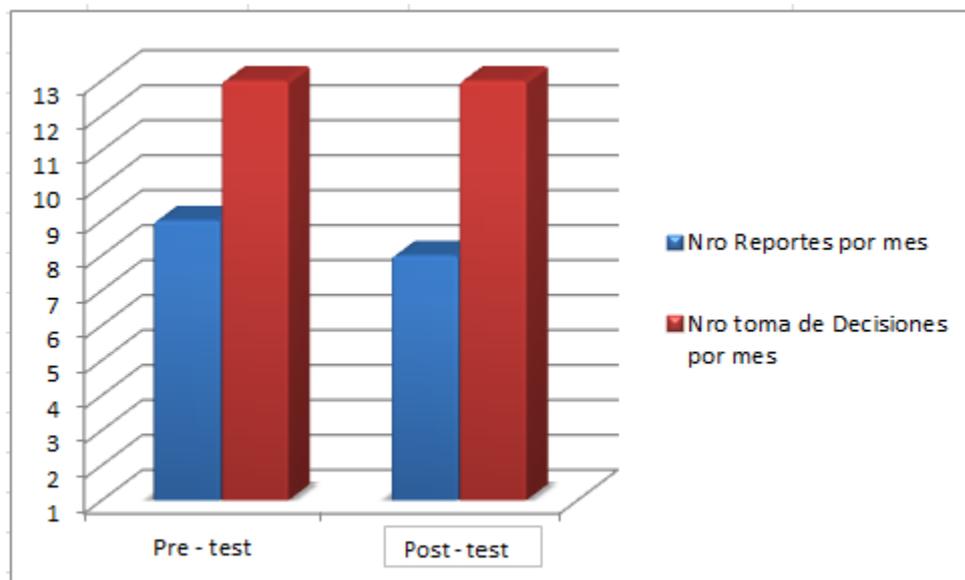
Comparativo de resultados de la Pre- test y Post-test

Cuadro 30 – Comparación de resultados entre el sistema actual y el Data mart

Indicadores	O1(Pre - test)	O2(Post - test)
Nro. De reportes por mes	9	13
Nro. De Toma de decisiones por mes	8	13

Fuente: Elaboración propia

Figura 46 – Grafico comparativo de resultados de la Pre – test y Post - test



Fuente: Elaboración propia

V.3. Contratación de la Hipótesis

Formulación De Las Hipótesis Nula Y Alterna

Hipótesis Nula (Ho)

H (0): El desarrollo de un Data mart ayuda a elevar la eficiencia en la toma de decisiones en el área de Administración y Finanzas de la Municipalidad Distrital de Baños del Inca en no menos de un 10%

Hipótesis alterna

H(A): El desarrollo de un Data mart no ayuda a elevar la eficiencia en la toma de decisiones en el área de Administración y Finanzas de la Municipalidad Distrital de Baños del Inca en no menos de un 10%.

Nro. De reportes por mes

Cuadro 31 – Nro. Reportes por mes

	Antes	Después	d
Usuario	9	13	-4
Usuario	9	13	-4
Total	18	26	-8

Fuente: Elaboración propia

A. Información para la prueba:

Nivel de Significancia: $1 - \alpha = 0.95$

Grados de Libertad: $N - 1 = 2 - 1 = 1$

Región de Aceptación: $t = 6.314$

B. Cálculos:

$$d' = \frac{\sum d}{N}$$

$$= -4$$

$$\sigma d = \sqrt{\frac{\sum (d-d')^2}{N-1}}$$

$$\sigma d = \sqrt{\frac{\sum (-8 - (-4))^2}{2-1}}$$

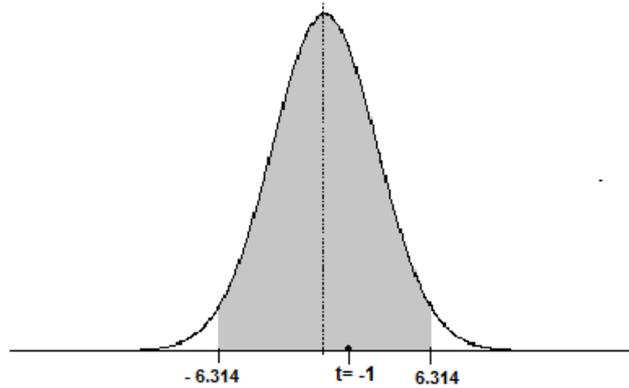
$$\sigma d = 4$$

C. Resultado de la Prueba

$$t = \frac{d'}{\frac{\sigma_d}{\sqrt{N}}}$$

$$t = -1$$

Figura 47 – Resultado de la prueba nro. Reportes por mes



Fuente: Elaboración propia

Nro. De Toma de decisiones por mes

Cuadro 32 – Nro. de toma de decisiones por mes

	Antes	Después	d
Usuario	8	13	-5
Usuario	8	13	-5
Total	16	26	-10

Fuente: Elaboración propia

A. Información para la prueba:

Nivel de Significancia: $1 - \alpha = 0.95$

Grados de Libertad: $N - 1 = 2 - 1 = 1$

Región de Aceptación: $t = 6.314$

B. Cálculos:

$$d' = \frac{\sum d}{N}$$

$$= -5$$

$$\sigma_d = \sqrt{\frac{\sum (d - d')^2}{N - 1}}$$

$$\sigma_d = \sqrt{\frac{\sum (-10 - (-5))^2}{2 - 1}}$$

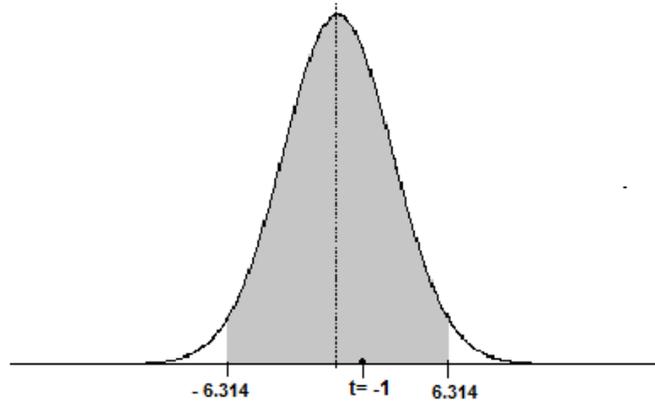
$$\sigma_d = 5$$

C. Resultado de la Prueba

$$t = \frac{d'}{\frac{\sigma d}{\sqrt{N}}}$$

$$t = -1$$

Figura 48 – Resultado de la prueba nro. de toma de decisiones por mes



Fuente: Elaboración propia

Aceptamos la hipótesis nula con un nivel de significancia del 5%, concluyendo que el desarrollo de un Data Mart proporciona una mejora a la eficiencia en la toma de decisiones en el área de Administración y Finanzas de la Municipalidad Distrital de Baños del Inca en más del 10%, ya que ahorra tiempo en la obtención de información requerida para la toma de decisiones

CONCLUSIONES

Se logró recolectar información de las partes interesadas del área gerencial y de sistemas con el fin de poder contar con los requerimientos claros.

Se logró analizar y diseñar el modelo multidimensional para el soporte de toma de decisiones, basado en un esquema en estrella con las dimensiones de contribuyente, tiempo, predios, fuente financiamiento, con sus respectivas jerarquías, hechos, y atributos., que incluían las jerarquías requeridas por los usuarios.

El uso de una interfaz de inteligencia de negocios con reportes permite un manejo intuitivo y sencillo a los usuarios finales para generar sus propios reportes y análisis, acorde a las necesidades del negocio en comparación del uso de hojas de cálculo o de los sistemas transaccionales utilizados.

La implementación del data mart en la institución contribuye a la mejor administración y gestión de la información, disminuyendo los tiempos del desarrollo de reportes y tiempos en la toma de decisiones que son necesarios para decisiones estratégicas basadas en información de calidad.

RECOMENDACIONES

Es importante contar con una base de datos transaccional con información consistente, para no tener problemas en el poblamiento dimensional.

Es muy importante desarrollar una buena fase de análisis para evitar que a lo largo del proyecto surjan problemas que ameriten una reestructuración de los procesos, mapeos o de los reportes mismos.

Tener todos los datos consistentes y ordenados en el data mart, brinda una fuente confiable y estandarizada para el desarrollo de futuros data marts.

Los datos que contiene el data mart permiten el diseño de otra serie de reportes para toma de decisiones en las áreas de administración y finanzas. Esta solución puede servir de base para cualquier institución del estado del mismo rubro.

El utilizar software libre como herramienta para instituciones del estado es recomendable debido a que estas instituciones no tienen cultura de inversión para invertir en software propietarios.

Se puede crear más adelante usuarios y darles permisos conforme el área Administración y Finanzas de la Municipalidad distrital de Baños del Inca lo requiera.

Se recomendaría continuar con el desarrollo de data marts en la Municipalidad para las demás áreas a fin de dar soporte integrado para la toma de decisiones en esta institución, en base a estudios o evaluación de prioridades.

FUENTES DE REFERENCIA

Bibliografía

- Cano, J. L. (2007). *Business Intelligence: Competir con información*. Madrid: Banesto, Fundación Cultural.
- DATE, C. (2001). *Introducción a los Sistemas de bases de datos*.
- DATE, C. (2010). *Introducción a las bases de datos*.
- Date, C. (2011). *Introducción a los sistemas de bases de datos - Séptima Edición*.
- Kimball, R. (1996). *The Data Warehouse Toolkit*. Wiley.
- Ramez Elmasri, & Shamkant B. Navathe. (2007). *Fundamentos de Sistemas de Bases de Datos*. Madrid: ADDISON WESLEY.
- Rivadera, G. R. (s.f.). *ucasal*. Recuperado el 28 de Julio de 2012, de <http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>
- Romero, L. M. (2004). *enterate*. Recuperado el 27 de Julio de 2012, de <http://www.enterate.unam.mx/Articulos/2004/octubre/inteligen.htm>
- Romero, L. M. (2004). *enterate*. Obtenido de <http://www.enterate.unam.mx/Articulos/2004/octubre/inteligen.htm>
- Wikipedia. (24 de Junio de 2012). *Wikipedia*. Recuperado el 28 de Julio de 2012, de http://es.wikipedia.org/wiki/Modelo_de_base_de_datos
- Wikipedia. (s.f.). *Wikipedia*. Recuperado el 8 de Agosto de 2012, de http://es.wikipedia.org/wiki/Data_mart

ANEXOS

ANEXO 1

MATRIZ DE CONSISTENCIA DEL PROYECTO DE INVESTIGACIÓN

HIPÓTESIS	DEFINICIÓN DE LA VARIABLE	DIMENSION DE LA VARIABLE	INDICADOR DE LA VARIABLE	ITEM DE LA VARIABLE
<p>El desarrollo de un Data Mart ayuda a elevar la eficiencia en la toma de decisiones en el área de Administración y Finanzas de la Municipalidad Distrital de Baños del Inca en no menos un 10%.</p>	<p>Variable 1 -Desarrollo de un data mart.</p>	<p>-Metodología utilizada. -Datos de los usuarios.</p>	<p>-Cumplimiento -Reporte de entregables</p>	<p>-¿Completa los pasos de la metodología? -¿Reporta los entregables de cada paso?</p>
	<p>Variable 2 -La toma de decisiones en el área de Administración y Finanzas.</p>	<p>-Información SAT.</p>	<p>--Cantidad de reportes. -Número de toma de decisiones.</p>	<p>¿Cuántos reportes se solicitan al mes? ¿Cuántas decisiones se toman al mes?</p>

ANEXO 2

FICHA DE OBSERVACIÓN No 1

APLICADA ANTES DE LA SIMULACIÓN DE LA PROPUESTA

I. INDICACIÓN: La presente ficha de observación tiene por finalidad recoger información con respecto a la validación del producto propuesto.

II. DATOS INFOMATIVOS:

1. Nombre de la institución visitada: Municipalidad Distrital de Baños del Inca

2. Nombre del observador: Melsi Ocas Terrones

3. Nombre del producto:

VARIABLE DEPENDIENTE			
	CANTIDAD DE REPORTES POR MES		
	00-9	10-19	20 A MAS
Nro. De reportes por mes	9		
	No DE TOMA DE DECISIONES POR MES		
	00-9	10-19	20 A MAS
Nro. De Toma de decisiones por mes	8		

ANEXO 3

FICHA DE OBSERVACIÓN No 2

APLICADA DESPUES DE LA SIMULACIÓN DE LA PROPUESTA

I. INDICACIÓN: La presente ficha de observación tiene por finalidad recoger información con respecto a la validación del producto propuesto.

II. DATOS INFOMATIVOS:

1. Nombre de la institución visitada: *Municipalidad Distrital de Baños del Inca*

2. Nombre del observador: *Melsi Ocas Terrones*

3. Nombre del producto:

VARIABLE DEPENDIENTE			
	CANTIDAD DE REPORTES POR MES		
	00-9	10-19	20 A MAS
Nro. De reportes por mes		13	
	No DE TOMA DE DECISIONES POR MES		
	00-9	10-19	20 A MAS
Nro. De Toma de decisiones por mes		13	