



UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE INGENIERÍA

**PROGRAMA ACADÉMICO DE INGENIERÍA DE SISTEMAS DE
INFORMACIÓN**

Solución de inteligencia de negocios que permita automatizar la
disponibilidad de la información enfocado en una empresa de
transporte

TESIS

Para optar el título profesional de Ingeniero de Sistemas de Información

AUTOR(ES)

Roque Vargas, Emmanuel Richard (0000-0002-4794-0425)

Cadillo Montesinos, Ricardo Benito Jean (0000-0001-7335-0157)

ASESOR

Barrientos Padilla, Alfredo (0000-0002-0029-4913)

Lima, Lunes, 30 de Julio de 2022

DEDICATORIA

***Emmanuel Roque:** este trabajo se lo dedico a mi abuelos que en vida siempre me
incentivaron a seguir mis sueños también a mis padres y
hermanos por apoyarme constantemente a no rendirme*

***Ricardo Cadillo:** este trabajo se lo dedico a mi familia, en especial a mis padres y
por apoyarme constantemente en cumplir mis metas*

RESUMEN

El presente proyecto de investigación tiene como objetivo implementar una solución de inteligencia de negocios (BI) aplicado a una empresa en el sector de transporte interprovincial en Barranca, dentro de sus necesidades abarca automatizar la disponibilidad del manejo de la información de la venta de sus pasajes. Parte de ello, implica la recolección de datos, análisis y procesamiento de la información que serán utilizado como parte de la salida de los reportes de ventas y tableros gráficos; el cual permita a la toma de decisiones de una forma más ágil y didáctica del proceso de venta de pasajes, evitando los retrasos de los procesos manuales.

Por consiguiente, para el desarrollo de la solución de inteligencia de negocios se analizaron diversas herramientas para la creación de flujo de procesamiento de ETL y además plataforma para la creación de tableros gráficos. Luego de realizar el análisis se optó por emplear las siguientes herramientas de Microsoft Power Automate para automatizar el proceso desde recibir los reportes, enviarlos a AWS S3 como repositorio en Cloud, después proceder a utilizar Pentaho Data Integration(PDI) para la transformación de los datos con la finalidad de poder almacenar las dimensiones de las tablas y la información en AWS Redshift y al final los reportes puedan mostrarse gráficamente en PowerBI. Para este caso, se empleó datos proporcionado por la organización de los cuales se ejecutaron pruebas de concepto con el objetivo de validar la solución.

Finalmente, de haber realizado la prueba de concepto se obtuvo como resultado esperado la disminución del tiempo de la disponibilidad de la información en el proceso de los reportes de ventas de pasajes. Asimismo, se adaptó a la organización una manera de trabajo más eficiente al generar los reportes y exportar la información en tableros gráficos más interactivos donde les permita poder analizar la información procesada, con la posibilidad de crear mejores estrategias para la toma de decisión en la organización.

Palabras clave: automatización de procesos, disponibilidad de información, inteligencia de negocios, sector transporte

ABSTRACT

The objective of this research project is to implement a business intelligence (BI) model applied to a company in the interprovincial transport sector in Barranca, within its needs it includes automating the availability of the information management of the sale of your tickets. Part of it implies the collection of data, analysis and processing of the information that will be used as part of the output of the sales reports and graphic dashboards, which allows decision-making in a more agile and didactic way of the ticket sales process, avoiding the delays of manual processes.

Therefore, for the development of the business intelligence model, various tools were analyzed for the creation of the ETL processing flow and also a platform for the creation of graphic dashboards. After performing the analysis, it was decided to use the following Microsoft Power Automate tools to automate the process from receiving the reports, sending them to AWS S3 as a repository in the Cloud, then proceeding to use Pentaho Data Integration (PDI) for data transformation. in order to be able to store the dimensions of the tables and the information in AWS Redshift and in the end the reports can be displayed graphically in PowerBI. For this case, data provided by the organization was used, of which concept tests were executed in order to validate the solution.

Finally, having carried out the proof of concept, the expected result was the reduction in the time of availability of information in the process of ticket sales reports. Likewise, a more efficient way of working was adapted to the organization by generating reports and exporting the information in more interactive graphic boards where it allows them to analyze the processed information, with the possibility of creating better strategies for decision-making in the organization.

Keywords: process automation, information availability, business intelligence, transport sector

INTRODUCCIÓN

Hoy en día las empresas de transporte tienen que contemplar un proceso de toma de decisiones rápido y ágil el cual se base en información confiable. Por ello, una de las maneras más factible de hacer todo esto es mediante el análisis de datos, puesto que toda empresa genera una gran cantidad de datos ya sea desde información de los clientes, empleados, márgenes de ventas y hasta los hábitos de consumo de sus clientes. Por lo general, estas informaciones no son tratadas o en algunos casos son ocupados como archivos de Excel. Pero si analizamos desde otra perspectiva estos reportes que se obtiene pueden dar como inputs importantes en el cual ayude a evitar problemas en la toma de decisiones. En eso se enfoca la Inteligencia de Negocios y el Big Data, extraer información dentro de los datos y crear visualizaciones interactivas con el fin de obtener respuestas rápidas. Por este motivo, el presente proyecto tiene la finalidad de diseñar una solución de inteligencia de negocios (BI) el cual será aplicado a una organización del sector transporte; el cual abarca la ruta interprovincial Barranca, Huacho, Chancay y Lima, esta solución ayudará a automatizar la disponibilidad de la información en las ventas diarias.

Para el desarrollo del proyecto se analizaron soluciones relacionado a Inteligencia de Negocios, Big Data y AWS. Asimismo, se realizó una búsqueda de artículos científicos relacionado a BI con la finalidad de aplicar las buenas prácticas. A partir de ello, se identificó los principales requerimientos y la necesidad del usuario final. Además, se identificó las herramientas a emplear AWS S3 para el repositorio de la información, para la extracción e integración de datos como por ejemplo Pentaho Data Integration (PDI) donde se AWS Redshift almacenara la información del data warehouse, con la finalidad de la generación de los reportes PowerBI. Igualmente, para la ejecución de pruebas de concepto; la empresa nos proporcionó la información de la organización con la finalidad de validar la solución.

Por último, habiendo realizado la prueba de concepto se obtuvo como resultado la disminución del tiempo al obtener la disponibilidad de información de los reportes de ventas. Adicionalmente, se proporcionó a la organización un flujo de trabajo más eficiente automatizando, usando Power Automate en el proceso de la inteligencia de negocio al momento de generar la transformación y exportación de sus reportes, monitorear los hábitos de las ventas de pasajes, y la posibilidad de crear estrategias para mejorar la gestión de la organización.

El presente trabajo constará de 6 capítulos de desarrollo. En el capítulo uno se detalla la descripción del proyecto, en planteamiento del problema y sus causas. En el capítulo dos se

describe los logros de los Students Outcomes los cuales son requeridas por el organismo de certificación internacional a los cuales se alinean la universidad. En el capítulo tres se describe el marco teórico del proyecto el cual contempla los conceptos y definiciones generales de la investigación. En el capítulo cuatro describe el estado del arte del proyecto, donde se mencionan artículos científicos d investigación y casos de estudio relacionado con el tema de BI. En el capítulo cinco describe el diseño y desarrollo del proyecto que comprende la solución de BI. Por último, en el capítulo 6 se muestran los resultados obtenidos de la solución, como también se brinda conclusiones y recomendaciones respecto a la solución implementada.

Tabla de Contenido

1.	DESCRIPCIÓN DEL PROYECTO	12
1.1	Objeto de estudio	13
1.2	Dominio del problema	13
1.3	Planteamiento de la Solución	14
1.4	Objetivos del Proyecto.....	14
1.4.1	<i>Objetivo General</i>	14
1.4.2	<i>Objetivos específicos</i>	14
1.5	Indicadores de éxito	15
1.6	Planificación del proyecto.....	15
1.6.2	Gestión del tiempo	16
1.6.3	Gestión de recursos humanos	18
1.6.4	Gestión de comunicaciones	19
1.6.5	Gestión del riesgo	19
2.	LOGROS DE LOS STUDENT OUTOMES	21
2.1	STUDENT OUTCOME (1)	22
2.1.1	<i>Descripción</i>	22
2.1.2	<i>Evidencia</i>	22
2.2	STUDENT OUTCOME (2)	22
2.2.1	<i>Descripción</i>	22
2.2.2	<i>Evidencia</i>	22
2.3	STUDENT OUTCOME (3)	23
2.3.1	<i>Descripción</i>	23
2.3.2	<i>Evidencia</i>	23
2.4	STUDENT OUTCOME (4)	23
2.4.1	<i>Descripción</i>	23
2.4.2	<i>Evidencia</i>	23
2.5	STUDENT OUTCOME (5)	23
2.5.1	<i>Descripción</i>	23
2.5.2	<i>Evidencia</i>	23
2.6	STUDENT OUTCOME (6)	24
2.6.1	<i>Descripción</i>	24
2.6.2	<i>Evidencia</i>	24
2.7	STUDENT OUTCOME (7)	24
2.7.1	<i>Descripción</i>	24
2.7.2	<i>Evidencia</i>	24
3.	MARCO TEÓRICO	25

3.1	Inteligencia de negocios	26
3.1.1	<i>OLTP</i>	27
3.1.2	<i>Salidas</i>	27
3.1.3	<i>OLAP</i>	27
3.1.4	<i>ETL</i>	27
3.1.5	<i>Data Warehouse</i>	28
3.1.6	<i>Dashboard</i>	29
3.1.7	<i>PowerBI</i>	29
3.2	Automatización de procesos	30
3.2.1	<i>Power Automate</i>	30
3.3	Herramientas de Amazon AWS	31
3.3.1	<i>Amazon AWS S3</i>	31
3.3.2	<i>Amazon AWS Redshift</i>	32
4.	ESTADO DEL ARTE	34
4.1	Artículo de Investigación 1	35
4.2	Artículo de Investigación 2	37
4.3	Artículo de Investigación 3	39
4.4	Artículo de Investigación 4	41
4.5	Artículo de Investigación 5	44
4.6	Artículo de Investigación 6	47
4.7	Artículo de Investigación 7	49
4.8	Artículo de Investigación 8	51
4.9	Artículo de Investigación 9	53
4.10	Artículo de Investigación 10	54
4.11	Artículo de Investigación 11	57
4.12	Artículo de Investigación 12	58
4.13	Artículo de Investigación 13	59
4.14	Artículo de Investigación 14	61
4.15	Artículo de Investigación 15	63
4.16	Artículo de Investigación 16	64
4.17	Artículo de Investigación 17	66
4.18	Artículo de Investigación 18	67
4.19	Artículo de Investigación 19	68
4.20	Artículo de Investigación 20	69
4.21	Conclusiones	70
5.	DESARROLLO DEL PROYECTO	71
5.1	Análisis	72

5.1.1	<i>Situación actual del turismo</i>	72
5.1.2	<i>Proceso del Sector Transporte</i>	72
5.1.3	<i>Análisis de la metodología de inteligencia de negocios</i>	75
5.1.4	<i>Análisis de la herramienta</i>	80
5.2	Diseño de la solución	83
5.2.1	<i>Entradas</i>	84
5.2.2	<i>Proceso</i>	84
5.2.3	<i>Output</i>	84
5.2.4	<i>Arquitectura lógica</i>	84
5.2.5	<i>Arquitectura Capa Negocios</i>	85
5.2.6	<i>Arquitectura Capa de aplicaciones</i>	86
5.2.7	<i>Arquitectura Capa de tecnología</i>	87
5.3	Construcción de la solución	88
5.3.1	<i>Plan de Proyecto</i>	88
5.3.2	<i>Proceso de desarrollo de software</i>	90
5.3.3	<i>Diagrama Específicos de subprocesos</i>	92
5.3.4	<i>Obtención de requisitos</i>	94
5.3.5	<i>Análisis del modelo dimensional</i>	96
5.3.6	<i>Esquema estrella</i>	97
5.3.7	<i>Creación de dimensiones y modelo físico</i>	98
5.3.8	<i>Proceso ETL</i>	100
5.3.9	<i>Plan de construcción</i>	108
6.	RESULTADO DEL PROYECTO	111
6.1	Mejora en la disponibilidad de información en el proceso de ventas	112
6.2	Estimación de tiempos de ejecución	112
6.3	Modelo de Calidad	113
6.3.1	<i>Ponderado de características</i>	113
6.3.2	<i>Definición de métricas calidad</i>	113
6.3.3	<i>Estimación de métricas calidad</i>	115
6.1	Visualización Dashboard	117
7.	CONCLUSIONES	119
	Recomendaciones	120
	Anexo	121
	Carta de Aceptación	121
	BIBLIOGRAFÍA	123

TABLA DE CONTENIDOS

Tabla 1: Indicadores de éxito	15
Tabla 2: Hitos del Proyecto	16
Tabla 3: Cronograma del Proyecto	17
Tabla 4: Matriz de riesgos	20
Tabla 5: Criterios de puntuación	80
Tabla 6: Análisis de las herramientas de visualización BI	80
Tabla 7: Análisis de las herramientas Integración BI	82
Tabla 8: Equipos de Trabajo	88
Tabla 9: Lista de Necesidades del usuario	89
Tabla 10: Lista de Necesidades del usuario	89
Tabla 11: Presupuesto Estimado	90
Tabla 12: Subprocesos y roles	91
Tabla 8: Caso de Uso 1	94
Tabla 9: Caso de Uso 2	94
Tabla 10: Caso de Uso 3	95
Tabla 11: Caso de Uso 4	95
Tabla 12: Caso de Uso 5	96
Tabla 13: Dimensiones de la solución	96
Tabla 14: Tabla hechos	97
Tabla 15: Dimensión Cliente	98
Tabla 16: Dimensión terminal	99
Tabla 17: Dimensión tiempo	99
Tabla 18: Dimensión buses	99
Tabla 19: Dimensión rutas	99
Tabla 20: Dimensión canal de venta	100
Tabla 21: Tabla hecho venta de pasajes	100
Tabla 22: Configuración entre PDI y Amazon Redshift	100
Tabla 23: Herramientas y tecnologías	108
Tabla 24: Lista de Intereses y plantilla	109
Tabla 25: Lista de Artefactos	109
Tabla 27: Ponderado de características	113
Tabla 28: Clasificación de las características	113
Tabla 29: Métrica Usabilidad	113
Tabla 30: Métrica Eficiencia	114
Tabla 31: Métrica Fiabilidad	114
Tabla 32: Métrica Eficiencia del Entendimiento	115
Tabla 33: Métrica de Rendimiento	116
Tabla 34: Métrica de eficiencia del entendimiento	116

TABLA DE CONTENIDOS

Ilustración 1: Organigrama de RRHH	18
Ilustración 2: Sistema Híbrido entre la Infraestructura de la arquitectura empresarial e Inteligencia de Negocios. ..	26
Ilustración 3: ETL: Extracción, transformación y carga de datos (2021)	28
Ilustración 4: Ejemplo de una tabla de dimensiones	29
Ilustración 5: Power Automate.....	30
Ilustración 6: Cuadrante de Gartner	31
Ilustración 7: Introducción AWS S3.....	32
Ilustración 8: Arquitectura Redshift.....	33
Ilustración 9: Modelo Propuesto por los autores.....	36
Ilustración 10: Alineación del proceso de negocio	38
Ilustración 11: Análisis de redes sociales asistido por Big data para el modelo de negocio.....	40
Ilustración 12: Diagrama de arquitectura de alto nivel del marco de Big Data basado en la nube.....	43
Ilustración 13: Diseño del modelo del autor	44
Ilustración 14: Diseño del modelo del autor	46
Ilustración 15: Diseño del modelo del autor	48
Ilustración 16: Diseño del modelo del autor	50
Ilustración 17: Digitalization process of complex B2B sales processes – Enablers and obstacles	52
Ilustración 18: Integración de los metamodelos.....	54
Ilustración 19: Modelo organizacional BI y la toma de decisiones	56
Ilustración 20: Arquitectura de Análisis de Datos	56
Ilustración 21: Modelo ETL de Biomarcadores.....	60
Ilustración 22: Implementación del caso Hong Kong.....	62
Ilustración 23: Datos Estadísticos de la recuperación de ventas turistas	72
Ilustración 24: Proceso de Venta de Pasajes	73
Ilustración 25: Proceso de Seguimiento de la venta de pasajes	74
Ilustración 26: Evaluación de las metodologías de BI	78
Ilustración 27: Cuadrante de Gartner	82
Ilustración 28: Solución propuesta.....	84
Ilustración 29: Arquitectura de la solución.....	85
Ilustración 30: Arquitectura de la capa de negocios	86
Ilustración 31: Arquitectura de aplicaciones	87
Ilustración 32: Arquitectura de capa tecnológica.....	87
Ilustración 33: Diagrama del proceso del desarrollo de software	92
Ilustración 34: Diagrama del proceso del análisis de requerimientos	93
Ilustración 35: Diagrama del proceso del diseño de la solución	93
Ilustración 36: Diagrama del proceso de construcción	93
Ilustración 37: Diagrama del proceso de la validación	93
Ilustración 38: Diagrama del modelo de datos.....	98
Ilustración 39: Configuración entre PDI y AWS Redshift.....	101

Ilustración 40: Job Principal de la solución 102

Ilustración 41: Job Cargar Dimensiones 102

Ilustración 42: Vista general de transformación Carga de la informacion 103

Ilustración 43: Vista general del envío de la información a Amazon Redshift 103

Ilustración 44: Ruta del archivo 104

Ilustración 45: Item de los datos..... 104

Ilustración 46: Vista previa de los datos 105

Ilustración 47: Vista secuencia incremental 105

Ilustración 48: Vista selección de variables 106

Ilustración 49: Vista de formato de variable..... 106

Ilustración 50: Vista generación de log..... 107

Ilustración 51: Vista de salida y carga de datos 108

Ilustración 52 Cantidad de Pasaje Vendido por Mes 117

Ilustración 53 Venta por Agencia..... 118

Ilustración 54 Ingreso Mensual por Rango de Edad 118

1. DESCRIPCIÓN DEL PROYECTO

En este capítulo se presenta el objetivo general del proyecto, objetivos específicos e indicadores de éxito para cada uno de ellos. Así mismo, se presenta el alcance del proyecto, plan de comunicaciones, plan de recursos humanos, plan de gestión del tiempo y los riesgos que pueden existir durante el desarrollo del trabajo.

1.1 Objeto de estudio

El proyecto se titula: **Solución de inteligencia de negocios que permita automatizar la disponibilidad de la información enfocado en una empresa de transporte.** El objeto de estudio comprueba los beneficios del uso de la inteligencia de negocios que permita automatizar la disponibilidad de la información de las ventas de pasajes. La misma se basará en la metodología Kimball para todo el procesamiento de la información de una empresa de transporte. De igual manera, el proyecto estará soportado por las herramienta de Power Automate para la automatización de la solución de BI, AWS S3 que servirá como repositorio de la información de las ventas, Pentaho PDI para la integración y transformación de los datos cuya informacion del data warehouse estará almacenado en AWS Redshift integrado con PowerBI para la presentación de los dashboard que le permitirá a la organización poder tomar decisiones mediante el reporte de índice de ventas que contienen las ventas mensuales, anuales, la segmentación del público por edades, canales de ventas, para que mediante este estudio la empresa pueda aplicar las mejoras correctivas en la organización.

1.2 Dominio del problema

El proyecto de solución se enfocará en la demora de la entrega de información para monitorear el comportamiento de las ventas periódicas en una empresa de transporte interprovincial, ya que en la actualidad el proceso es manual y ello podría traer consigo el error humano afectando los distintos reportes de ventas. Al ser un proceso manual cuenta con un mayor tiempo de trabajo y análisis de la información para que la organización pueda conocer mejor el índice de las ventas de sus clientes.

Causas encontradas del problema:

- Poco uso de la información generada por el negocio para monitorear el comportamiento de las ventas de pasajes
- Desconocimiento de las metodologías de inteligencia de negocios
- Sistemas desactualizados para el procesamiento de la información

1.3 Planteamiento de la Solución

En base al problema descrito anteriormente, se propone implementar una solución de inteligencia de negocios enfocado en medir el índice de las ventas en una empresa de transporte interprovincial que les permitirá poder conocer la evolución de sus ventas mediante métricas y análisis de datos utilizando herramientas de BI y AWS y Power Automate para la automatización, procesamiento y repositorio de la información. Al final los resultados se presentarán con reportes mediante dashboards en PowerBI que le permitirá a la empresa determinar su rentabilidad y mejorar en su estrategias aplicadas al negocio.

1.4 Objetivos del Proyecto

1.4.1 *Objetivo General*

Implementar una solución de inteligencia de negocios que permita automatizar la disponibilidad de la información del proceso de ventas enfocado en una empresa de transporte

1.4.2 *Objetivos específicos*

- **Objetivo Específico 1:** Analizar el uso de metodologías, técnicas y herramientas que permita automatizar la disponibilidad de la información contemplando inteligencia de negocio.
- **Objetivo Específico 2:** Diseñar una solución de inteligencia de negocios que permita medir el índice de las ventas en una empresa de transporte.
- **Objetivo Específico 3:** Validar los componentes de la solución de inteligencia de negocios mediante pruebas de concepto en una empresa de transporte que permita medir el índice de ventas

1.5 Indicadores de éxito

Se definen los indicadores de éxito con los que se evaluará el cumplimiento de los objetivos estratégicos del proyecto.

Tabla 1: Indicadores de éxito

Indicador de éxito		Objetivo
I1	Documento del Análisis de las herramientas y metodologías BI aprobado por el profesor de tesis	OE1
I2	Documento del diseño de la solución aprobado por el profesor de tesis	OE2
I3	Documento de las arquitecturas en capas aprobado por el profesor de tesis	OE2
I4	Documento de la construcción de la solución aprobado por el profesor de tesis	OE3
I5	Documento de las métricas de calidad	OE3
I6	Documento de la estimación de los casos de prueba	OE3

Fuente: Elaboración propia

1.6 Planificación del proyecto

1.6.1 Gestión del Alcance

El presente proyecto tiene como objetivo principal implementar una solución de inteligencia de negocios que permita automatizar la disponibilidad de la información del proceso de ventas enfocado en una empresa de transporte.

Asimismo, se analizará el uso de metodologías, técnicas y herramientas que permita automatizar las ventas contemplando inteligencia de negocio

- Definir el diseño de la solución de inteligencia de negocios soportado en la metodología Kimball apoyado en la herramienta de Power BI.
- Diseño del proceso de ventas y de las arquitecturas físicas y lógicas de la solución de BI.
- Validación de la solución por la empresa de transporte interprovincial mediante métricas de evaluación y dashboards que ayuden a la empresa poder seguir mejorando en su proceso
- Memoria del Proyecto de la solución.

1.6.2 Gestión del tiempo

En esta fase se busca definir los principales hitos del proyecto, con el fin de tener las fechas estimadas de los entregables y actividades para su respectivo seguimiento.

Tabla 2: Hitos del Proyecto

Hito del proyecto	Fecha Estimada	Entregables incluidos	Prioridad
Análisis el uso de metodologías, técnicas y que permita automatizar la disponibilidad de la información contemplando inteligencia de negocio.	17/06/2022	Informe del análisis del uso de metodologías, técnicas y herramientas de la inteligencia de negocios	Alta
Diseño de una solución de inteligencia de negocios que permita medir el índice de las ventas en una empresa de transporte.	18/07/2022	Informe del diseño de la solución de inteligencia de negocios	Alta
Validación de la solución Business Intelligence con la empresa del sector transporte	15/08/2022	Informe de la validación de la solución Business Intelligence con la empresa del sector transporte	Alta

Fuente: Elaboración propia

Tabla 3: Cronograma del Proyecto

Actividades	Artefacto	Rol	Inicio	Fin	% de Avance
Entrevista con el Gerente General	Acta de reunión	Jefe de Proyecto	19/06/2021	20/06/2021	100%
Elaboración del cronograma del proyecto	Cronograma de Proyecto	Jefe de Proyecto	19/06/2021	20/06/2021	100%
Formulación del acta de constitución del proyecto	Acta de Constitución	Jefe de Proyecto	19/06/2021	20/06/2021	100%
Definición del análisis de requerimientos	Documento de Análisis de requerimientos	Analista Funcional	20/06/2021	23/06/2021	100%
Análisis de la metodología de Inteligencia de Negocios	Documento de análisis de la metodología de Inteligencia de Negocios	Analista Funcional	02/07/2021	03/07/2021	100%
Análisis de herramientas	Documento de análisis de herramientas BI	Analista BI	04/07/2021	06/07/2021	100%
Hito 1			11/07/2021		100%
Elaboración del Diseño de la solución	Documento Diseño de la solución	Arquitecto	11/07/2021	12/07/2021	100%
Elaboración de la Arquitectura de la solución	Documento de la arquitectura de la solución	Arquitecto	15/07/2021	17/07/2021	100%
Diseño del Modelo Lógico Dimensional	Diagrama del Modelo Lógico Dimensional	Analista BD	18/07/2021	24/07/2021	100%
Hito 2			27/07/2021		100%
Construcción del ETL	Documento de la construcción ETL	Analista BI	28/07/2021	02/08/2021	100%
Construcción de las métricas de calidad	Documento de las métricas de calidad	Tester	02/08/2021	06/08/2021	100%
Hito 3			27/08/2021		100%
Resultados de la solución	Documento de los dashboard	Analista BI	01/09/2021	04/09/2021	100%
Elaboración de Entrevistas	Documento de las entrevistas	Analista BI	02/09/2021	04/09/2021	100%
Hito 4			11/09/2021		100%

Fuente: Elaboración propia

1.6.3 Gestión de recursos humanos

En esta fase se identifican y definen los roles y responsabilidades de las actividades a lo largo del desarrollo del proyecto.



Ilustración 1: Organigrama de RRHH

Fuente: Elaboración propia

Los roles involucrados en el desarrollo del proyecto son los siguientes:

En esta sección se debe describir brevemente cada rol del organigrama del proyecto, además se debe hacer referencia al documento de Descripción de Roles y Responsabilidades asociado. A continuación, se presenta el ejemplo de un rol.

- **Jefe de Proyecto:** Rol correspondiente al encargado del cumplimiento de todas las fases del proyecto cumpliendo lo establecido en el cronograma con el fin de cumplir con los objetivos establecidos por el cliente.
- **Analista Senior BI:** Rol correspondiente al encargado de la implementación de la solución de la inteligencia de negocios y aplicar la metodología Kimball
- **Analista BD:** Rol correspondiente al encargado de la BD que revisara la conexión entre Pentaho y Amazon AWS Redshift
- **Analista Arquitectura:** Rol correspondiente al encargado de la arquitectura física y lógica de la solución de inteligencia de negocios y responsable del mapeo del proceso de ventas
- **Analista Infraestructura:** Rol correspondiente al encargado de la infraestructura que brindará la instalación de Pentaho, aplicar parches de seguridad y será apoyo para

las conexiones cloud.

- **Tester:** Rol correspondiente al encargado de validar la solución con las pruebas de concepto y evaluar las métricas de calidad.

1.6.4 Gestión de comunicaciones

El Plan de Comunicaciones tiene como objetivo definir los lineamientos para la comunicación interna y externa respecto a todos los involucrados del proyecto. De este modo, se establece de manera clara y transparente las formas, medios y procedimientos de coordinación estándares.

- **Guías para reuniones**
 - Se maneja una agenda para cada reunión de acuerdo con los temas de interés a tratarse, así como también las tareas pendientes de reuniones anteriores.
 - Cada punto de la agenda contará con un nivel de prioridad a ser tratado.
 - Se cuenta con un horario específico para cada reunión, sin embargo, será necesario confirmar/recordar la reunión vía correo electrónico un día antes como máximo.
 - En caso se deba postergar una reunión, se deberá justificar y además acordar el nuevo día de esta.
 - Al finalizar cada reunión se deberá recopilar lo tratado, resaltando los acuerdos para posteriormente elaborar el acta de reunión.

1.6.5 Gestión del riesgo

El plan de gestión de riesgos contempla los riesgos y amenazas que pueden poner en peligro el desarrollo y cumplimiento del proyecto; esto con el fin de determinar el impacto y probabilidad de ocurrencia de cada uno de ellos, para así poder establecer estrategias de mitigación que podamos aplicar en caso de que los mismos logren materializarse:

En la siguiente tabla, se menciona los riesgos del proyecto y las estrategias de mitigación correspondientes.

Tabla 4: Matriz de riesgos

#	Riesgo	Probabilidad	Impacto	Estrategia de mitigación
1	Cambios en el alcance del proyecto.	Media	Alto	Seguir procedimiento de cambios.
2	Pérdida de información relevante del proyecto.	Media	Alto	Respaldar la información del proyecto utilizando herramientas de almacenamiento cloud.
3	Demora en la elaboración de los entregables del proyecto.	Alto	Alto	Gestionar la reprogramación de las fechas de entrega.
4	Recursos de la información no suministrada por la empresa	Media	Alto	Realizar con anticipación la solicitud de la información de la empresa
5	No cumplir con la satisfacción de las necesidades de la empresa.	Media	Alto	Programar reuniones periódicas con la empresa para presentarle el avance del proyecto.
7	Salida previa a la culminación del proyecto del contacto de la empresa.	Bajo	Alto	Coordinar reuniones con la persona que ocupará el cargo del contacto.

Fuente: Elaboración propia

2. LOGROS DE LOS STUDENT OUTOMES

El objetivo de este capítulo es describir como el presente proyecto se relaciona con cada uno de los Student Outcomes definidos por la acreditadora internacional ABET.

2.1 STUDENT OUTCOME (1)

2.1.1 Descripción

La capacidad de identificar, formular y resolver problemas complejos de ingeniería aplicando los principios de ingeniería, ciencia y matemática.

2.1.2 Evidencia

- En la fase del proyecto, se realizó la investigación sobre las herramientas y metodologías a usarse para la implementación de la solución de BI, además se evaluaron los costos de las herramientas a utilizar, los recursos necesarios para poder ejecutar el proyecto. **Capítulo 5, sección 5.1 Análisis y Capítulo 5, sección 5.3.1.6 Presupuesto Estimado.**
- Asimismo, en el proyecto se establecieron las métricas de calidad donde se efectuaron estimaciones aritméticas que efectúan el cumplimiento del y la eficiencia del modelo. **sección 6.3 Modelo de Calidad**

2.2 STUDENT OUTCOME (2)

2.2.1 Descripción

La capacidad de aplicar el diseño de ingeniería para producir soluciones que satisfagan necesidades específicas con consideración de salud pública, seguridad y bienestar, así como factores globales, culturales, sociales, ambientales y económicos.

2.2.2 Evidencia

- En el diseño de la solución de negocios le permite a la empresa poder automatizar la disponibilidad de la información de las ventas utilizando la inteligencia de negocios, para ello se propuso distintas herramientas que le permita al usuario poder tener la información de una manera fácil y sencilla. Por ello, en la fase de automatización se propuso la herramienta de Power Automate para que el proceso de la solución se pueda efectuar de manera automática, el mismo se encuentra con herramientas Amazon AWS que usan el entorno cloud, Pentaho PDI para la parte de la extracción, transformación para así proceder almacenarlos en nuestro repositorio cloud y el mismo sea integrado con PowerBI con la finalidad de que la empresa pueda revisar la evolución de la compañía mediante los ingresos que tienen en sus ventas pudiendo así poder tomar decisiones. **Capítulo 5, sección 5.2 Diseño de la Solución.**

2.3 STUDENT OUTCOME (3)

2.3.1 Descripción

Capacidad de comunicarse efectivamente con un rango de audiencias.

2.3.2 Evidencia

- Para la ejecución del proyecto nos tuvimos que reunir con el Gerente General de la empresa de transporte para entender la necesidad del negocio y poder aplicar la inteligencia de negocios en el proceso de ventas de pasajes.
- Cada módulo del proyecto de la solución fue presentado a los profesores de tesis

2.4 STUDENT OUTCOME (4)

2.4.1 Descripción

Capacidad de reconocer responsabilidades éticas y profesionales en situaciones de ingeniería. Hacer juicios informados, que deben considerar el impacto de las soluciones de ingeniería en contextos globales, económicos, ambientales y sociales.

2.4.2 Evidencia

- Las fuentes empleadas en esta solución han sido citadas y referenciadas dentro de la bibliografía. **Capítulo 3, Marco Teórico y Capítulo 4, Estado del Arte.**
- La información de la empresa de transporte fue empleada de manera responsable y sin filtrar los datos de sus clientes **Capítulo 5, Desarrollo del Proyecto**

2.5 STUDENT OUTCOME (5)

2.5.1 Descripción

Capacidad de funcionar efectivamente en un equipo cuyos miembros juntos proporcionan liderazgo, crean un entorno de colaboración e inclusivo, establecen objetivos, planifican tareas y cumplen objetivos.

2.5.2 Evidencia

- En el proyecto se establecieron los objetivos generales y específicos y para verificar su cumplimiento se realizaron indicadores de éxito **Capítulo 1 sección 1.4 Objetivos del Proyecto.**
- Se realizó la planificación del proyecto en donde se encuentra detallado cada actividad a realizar en base a cada rol. **Capítulo 5, sección 5.3.1.7 Cronograma del Proyecto.**

2.6 STUDENT OUTCOME (6)

2.6.1 Descripción

Capacidad de desarrollar y llevar a cabo la experimentación adecuada, analizar e interpretar datos, y usar el juicio de ingeniería para sacar conclusiones.

2.6.2 Evidencia

- En el proceso de la solución se efectuó un análisis de la necesidad del negocio lo cual se evidencia en el **Capítulo 5.1 Análisis** y se verifica la funcionalidad mediante las pruebas de concepto y pruebas de calidad en donde se midieron los tiempos de ejecución en base a interacciones y tiempos de respuesta. **Capítulo 6, sección 6.3 Modelo de Calidad**

2.7 STUDENT OUTCOME (7)

2.7.1 Descripción

Capacidad de adquirir y aplicar nuevos conocimientos según sea necesario, utilizando estrategias de aprendizaje apropiadas.

2.7.2 Evidencia

- En los artículos de investigación y el análisis de herramientas de Integración y Visualización se decidió por Pentaho, para la automatización Power Automate, para los servicios de cloud se utilizó Amazon AWS S3 y Redshift y para la visualización de los reportes serian mediante PowerBI, Esto conllevó al aprendizaje de nuevas herramientas para la ejecución de la solución.

3. MARCO TEÓRICO

El desarrollo del presente capítulo presentará la definición de todos los términos y conceptos esenciales en el ámbito de la inteligencia de negocios. Dentro de la información presentada se detallan los conceptos básicos y específicos del contexto tecnológico del proyecto.

3.1 Inteligencia de negocios

Inteligencia de Negocios o por sus siglas en ingles business intelligence (BI), según el Data Warehouse Instute, “lo define como la combinación de tecnología, herramientas y procesos que me permiten transformar los datos almacenados en información” (Luis Joyanes Aguilar, 2019). En la actualidad, la inteligencia de negocios ha ido evolucionando con una visión más integral de las empresas impulsando el cambio de trabajar con la información, eliminando las ineficiencias y permitiendo adaptarse a los cambios de mercados.

Asimismo, una de las principales prioridades de los ejecutivos de TI por varios años ha sido el creciendo de las empresas de software porque les permite a las empresas dar una respuesta rápida y eficaz ante decisiones comerciales. (Oswaldo Moscoso, Jorge Castro, Joel Paredes, Sergio Luján, 2019)

Los sistemas de BI consisten en recopilar los datos, el preprocesamiento, la construcción de un data warehouse, análisis de datos y la visualización de resultados mediante reportes o gráficos dinámicos para la alta gerencia.

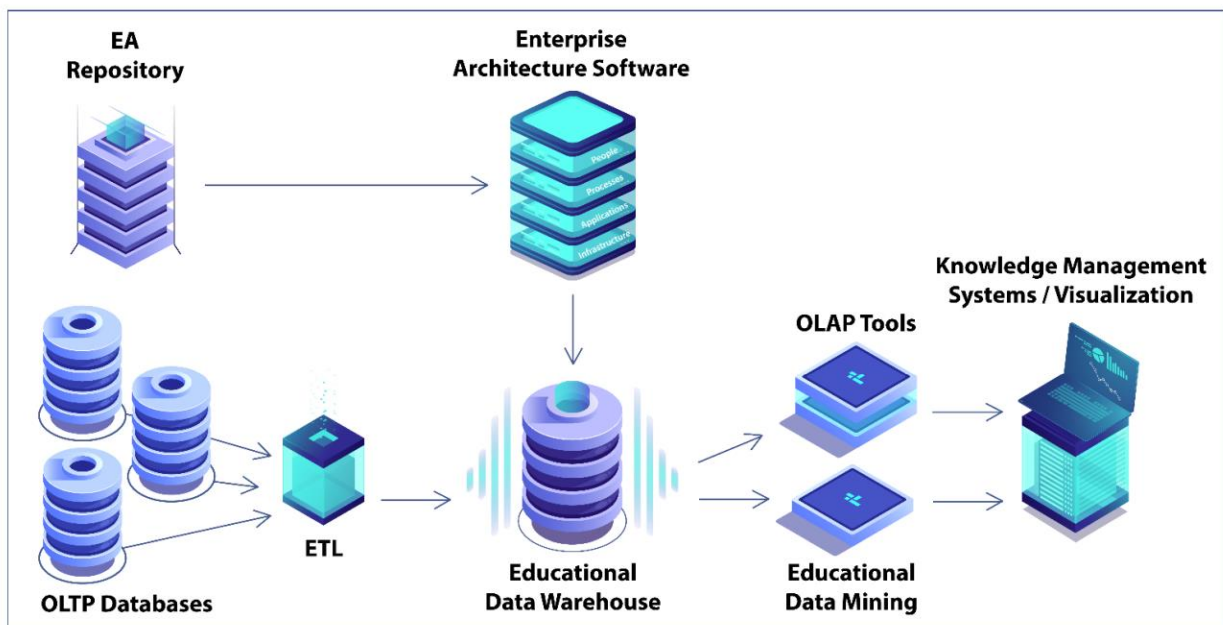


Ilustración 2: Sistema Híbrido entre la Infraestructura de la arquitectura empresarial e Inteligencia de Negocios.

Fuente: (Oswaldo Moscoso, Jorge Castro, Joel Paredes, Sergio Luján, 2019)

A continuación, se detalla las definiciones y conceptos relacionados.

3.1.1 *OLTP*

Es el procesamiento de transacciones en Línea facilitando la gestión de aplicaciones orientadas a transacciones, usualmente para el procesamiento de transacciones.

Entre las ventajas que tiene este sistema se encuentran la capacidad para manejar muchas solicitudes de transacciones simultáneamente (llamado concurrencia) y la capacidad de realizar copias de seguridad de manera confiable y continuar si falla parte del sistema (llamado atomicidad).

3.1.2 *Salidas*

Las principales salidas de la solución de BI propuesta son las siguientes:

Dashboard o Panel gráfico del proceso de ventas

En primer lugar, permite que todos los datos se muestren visualmente más atractivos, para cualquier persona de la organización, una manera más fácil de interpretar los resultados de las ventas. Asimismo, la creación de estos informes personalizados beneficia a las áreas de negocio interesadas lo cual mejora la comunicación.

En segundo lugar, habiendo automatizado el volumen de ventas de pasajes periódicamente el cual los dashboard mantendrá toda la información actualizada para generar versiones según la gerencia solicite y siendo esto reutilizado a futuro.

3.1.3 *OLAP*

Son bases de datos orientadas al proceso analítico, las cuales son utilizadas para realizar análisis con grandes cantidades de información de manera sencilla. Los análisis empleados realizan una gran lectura de datos con el principal objetivo de poder extraer información útil.

Fueron creados con las tablas de las bases de datos relacionales, los cubos OLAP trabajan en varios pasos desde la segmentación de información, filtro, profundizar y sintetizar, realizándose a una gran velocidad otorgándole al usuario los datos analizados en un menor tiempo.

3.1.4 *ETL*

Es un conjunto de técnicas, herramientas y tecnologías que permiten extraer los datos de distinto tipo de fuentes, transformarlos eliminando la redundancia e información innecesaria y cargarlos en otro sistema con la finalidad de obtener un reporte

accesible para la organización.

- Fase de extracción. - La fase de extracción consiste en obtener varias fuentes de datos estructurados o no estructurados sintetizando su información en una o varias estructuras con datos normalizados y homogeneizados para la fase de transformación.
- Fase de transformación. - La fase de transformación apunta en generar valor con los datos obtenidos de la fase de extracción en donde el principal objetivo es eliminar los datos irrelevantes incorrectos o duplicados trayendo consigo un nuevo conjunto de datos que aporte valor a la siguiente fase.
- Fase de Carga. - La última fase se encarga de recibir los datos procesados y cargarlos en el sistema de destino.

3.1.5 Data Warehouse

Es un repositorio que agrega y combina diferente tipo de información en un almacén de datos único y centralizado.

Un data warehouse en la mayoría de los casos se aloja en un servidor corporativo o en la nube. Los datos almacenados permiten a las empresas generar herramientas de visualización y análisis de datos mediante reportes permitiendo a las organizaciones poder tomar decisiones. (Kimball & Caserta, 2015)

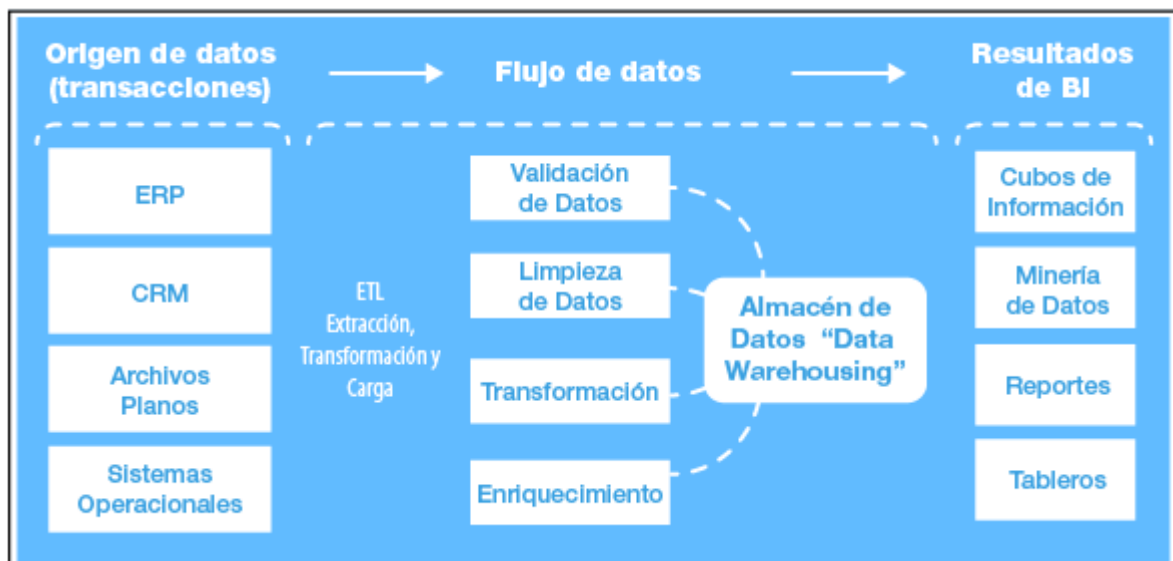


Ilustración 3: ETL: Extracción, transformación y carga de datos (2021)

Fuente: (Kimball & Caserta, 2015)

3.1.6 Dashboard

Es una herramienta de TI con la que toda la información obtenida con Business Intelligence se puede visualizar de forma didáctica y sencilla. Los dashboard muestran los KPI (indicadores clave de rendimiento) importantes de cifras relevantes para la compañía. Su principal objetivo es dar a conocer la información mediante gráficos para que sea más sencillo de entender.

Entre los formularios de representación puedes encontrar semáforos o diagramas de escala que son muy utilizados en Business Intelligence Dashboard con la finalidad de poder mapear diferentes tipos de métricas para poder procesar y evaluar datos de diferentes fuentes a tiempo real

3.1.7 PowerBI

Es un conjunto de herramientas, aplicaciones y conectores que brindan conocimiento al alcance de todos brindando los datos de forma segura y rápida, trae como consigo un sistema predictivo, inteligente y es de gran apoyo para traducir datos simples o complejos mediante gráficas. Los datos pueden ser extraídos desde una hoja de cálculo de Excel o una colección de almacenes de datos híbridos locales y basados en la nube. Power BI permite conectarse con facilidad a los orígenes de datos, visualizar y descubrir qué es importante y compartirlo con cualquiera o con todos los usuarios que desee. (“¿Que es PowerBI?”, 2020)

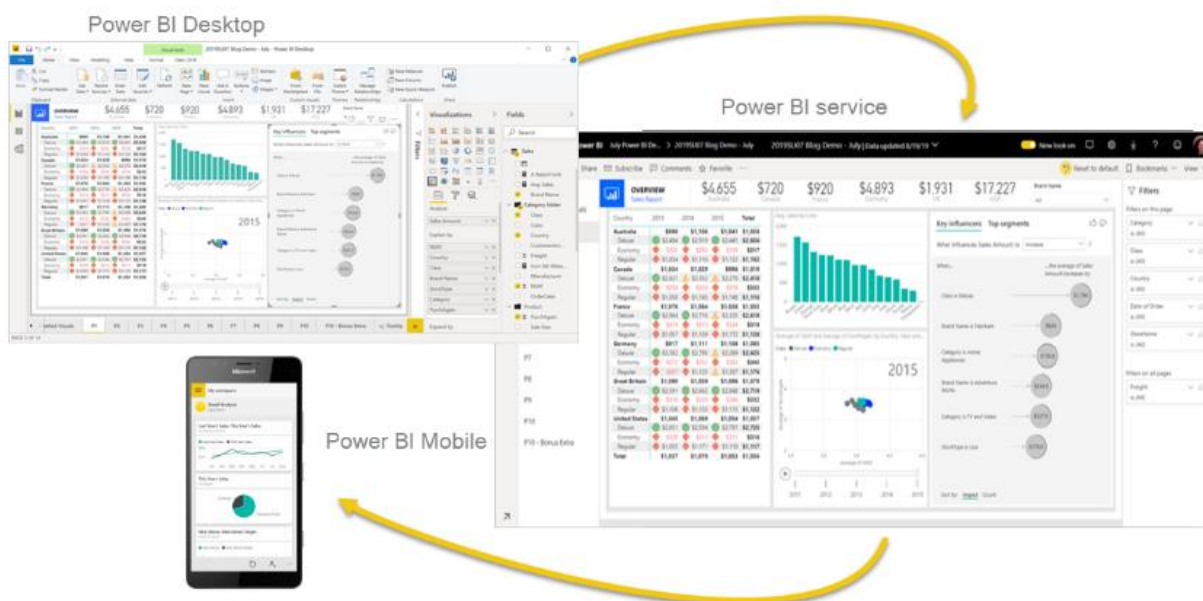


Ilustración 4: Ejemplo de una tabla de dimensiones

3.2 Automatización de procesos

3.2.1 Power Automate

Es el sistema empresarial de Microsoft que permite automatizar procesos y tareas de manera sencilla permitiendo disponer de un flujo funcional sin necesidad de desarrollar código o con muy poco código a esto se le conoce como Low Code. Asimismo, cuenta con plantillas predefinidas listas para usarse trayendo consigo de que cualquier usuario pueda utilizarlo. Asimismo, cuenta con integración con de Office 365 y aplicaciones de terceros lo cual la hace compatible con distintas plataformas. (Microsoft, 2019)

- **Ventajas**
 - Al contar con un proceso automatizado ahorra tiempo en realizar tareas individuales o manuales.
 - Integración con las aplicaciones de Microsoft y terceras.
 - Monitorizar la secuencia de las aplicaciones o desde el flujo creado.
- **Desventajas**
 - No permite iniciar un segundo flujo como acción continua al primer flujo.
 - No permite recrear un flujo o reconectarlo a nuevas listas

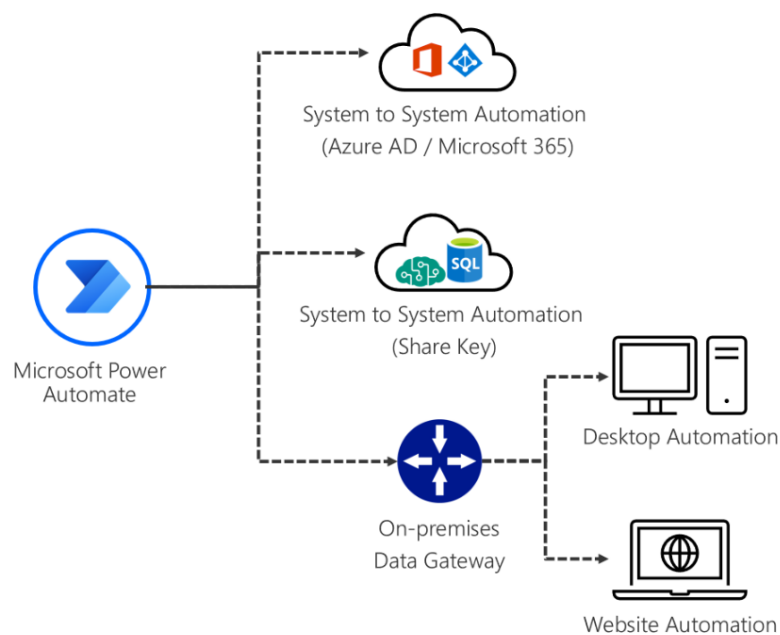
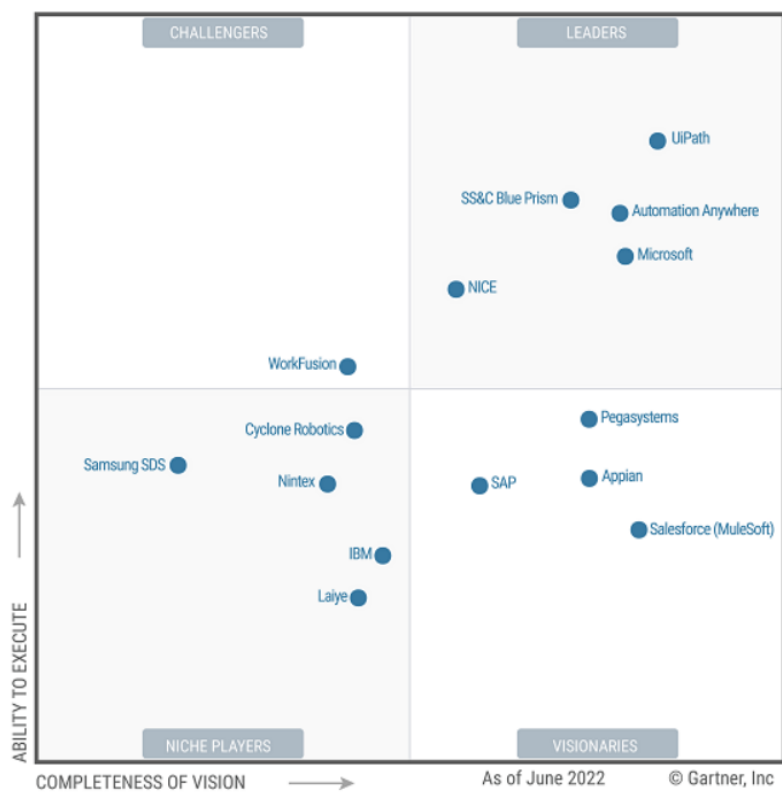


Ilustración 5: Power Automate

Fuente: (Microsoft, 2022)

El cuadrante mágico de Gartner podemos ver que en base de los 4 cuadrantes una de las mejores herramientas encontramos a Microsoft, RPA y Nice. En base a la tabla de criterios mencionadas anteriormente podemos escoger Microsoft Power Automate no solo por sus diversas funcionalidades, sino que al estar en el cuadrante de Gartner cuenta con un respaldo para poder aplicarlo en la solución. (“Gartner”, 2022)

Figure 1: Magic Quadrant for Robotic Process Automation



Source: Gartner (July 2022)

Ilustración 6: Cuadrante de Gartner

Fuente: (Gartner, 2022)

3.3 Herramientas de Amazon AWS

3.3.1 Amazon AWS S3

Es un servicio de Amazon AWS cuya función sirve de almacenamiento en la nube donde te permite guardar y recuperar datos cualquiera que sea su volumen en el cual el precio va acorde el almacenamiento. Asimismo, puede ser usado para almacenar web estático, almacenar imágenes, reportes, imágenes, etc.

En la actualidad, este servicio se posiciona como uno de los líderes del mercado por su popularidad, su facilidad de uso, escalabilidad, menor precio y su seguridad. (Amazon, 2021)

- **Ventajas**
 - La manipulación y administración de los archivos son sencillos e intuitivos tanto para profesionales y personas ajenas al rubro tecnológico
 - La API cuenta con varias capas de seguridad TI tanto físicas como operativas, además cuenta auditorías recurrentes con la finalidad de garantizar la seguridad de su infraestructura.
 - Cuenta con costos acorde al almacenamiento de los repositorios, esto quiere decir que uno paga solo por el uso del servicio de almacenamiento y transferencia.

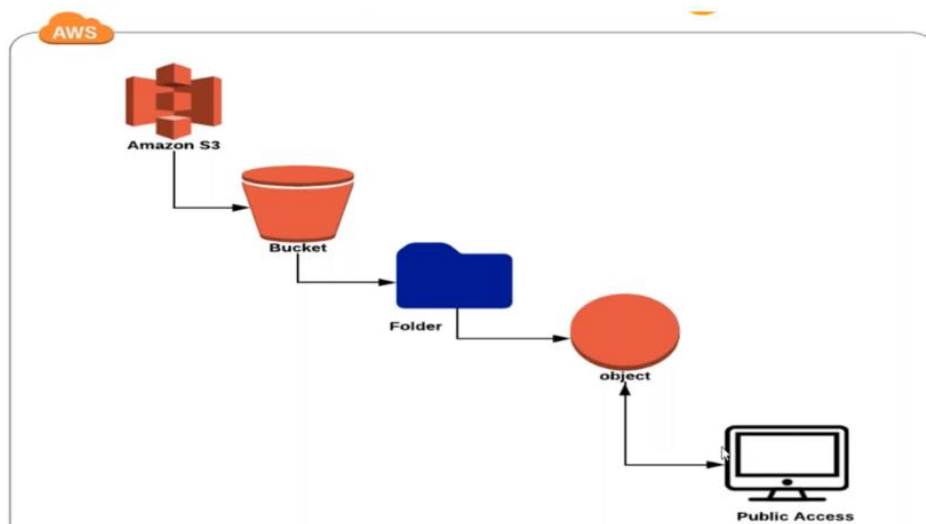


Ilustración 7: Introducción AWS S3

Fuente: (Amit Tyagi, 2021)

3.3.2 Amazon AWS Redshift

Es el servicio de data warehouse de Amazon AWS que permite realizar consultas en SQL de manera eficiente usando múltiples nodos en paralelo, el cual se encarga de ejecutar un subconjunto de consultas sobre los datos almacenados. (Amazon AWS, 2022)

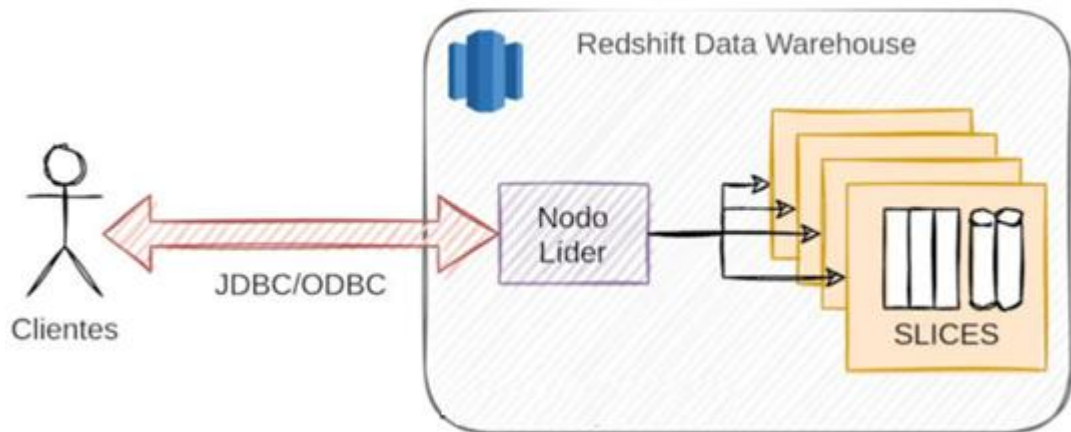


Ilustración 8: Arquitectura Redshift

Fuente: (Amazon, 2021)

- **Ventajas**
 - Cuenta con un almacenamiento de alto rendimiento
 - Se integra con herramientas de terceros
 - Cuenta con un cifrado SSL que permite encriptar sus datos cuando está en reposo

4. ESTADO DEL ARTE

En este cuarto capítulo se presenta una revisión de la literatura referida a la definición del Estado del Arte del proyecto, para esto se han revisado 20 artículos con un gran impacto de diferentes bases de datos. Mediante este estudio se permitirá poder llegar a la solución de la problemática descrita anteriormente.

4.1 Artículo de Investigación 1

Título: Business Intelligence for IT Governance of a Technology Company (Inteligencia de Negocios para el Gobierno TI de una Empresa de Tecnología).

Aporte:

Mediante el presente artículo, los autores presentan una investigación de caso de estudios para la definición de un modelo de inteligencia de negocios e indicadores clave de rendimiento relacionados (KPI) para apoyar la toma de decisiones relacionada con el riesgo. Por lo tanto, según las investigaciones el gobierno de TI involucra el liderazgo y las estructuras organizacionales y los procesos requeridos para asegurar que la TI de la organización se mantenga y se alinee con las estrategias y objetivos de la organización” (IT Governance Institute, 2011). Además, permite el intercambio de información dentro de la organización y facilita la identificación de áreas críticas de gobierno a monitorear.

Proceso: El primer proceso que se realiza en esta investigación, describir los antecedentes teóricos de la gestión del rendimiento, donde implica que el entorno del negocio dinámico ahora requiera que la estructura de TI brinde flexibilidad comercial, uso efectivo de sistemas y productos tecnológicos para el crecimiento, eficiencia de costos y soporte. Por ello, es necesario un gobierno centralizado para impulsar coordinar y respaldar el negocio. Es por ello que, medir el desempeño se relaciona con el reconocimiento de las variables que influyen en la ejecución de las tareas y la generación de resultados. Estos elementos son los controladores clave de rendimiento (KPD), que a menudo dan como resultado un mapa KPD. Por lo tanto, los autores indican que las organizaciones de hoy en día es crucial considerar cómo se debe compartir la información de la empresa en toda la empresa, puesto que es necesario reemplazar los informes basados en datos aislados, aplicaciones comerciales, almacenes de datos y aplicaciones de análisis, y la información debe usarse como base para cualquier decisión.

Asimismo, las organizaciones deben rediseñar todas sus prácticas de gestión de datos, que anteriormente se basaban en la intuición y la experiencia de los usuarios clave, con el fin de garantizar que BA las apoye sistemáticamente. Además, se debe comprender el valor de la tecnología de TI para compartir datos de la empresa. El desafío radica en lograr el equilibrio adecuado entre personas, procesos y tecnología.

Finalmente, los autores proponen a solución que da cuenta de un modelo de datos centralizado para respaldar la toma de decisiones holística en la gobernanza del

proyecto. La solución está destinada a ser utilizable en diferentes niveles organizacionales y compatible con los sistemas de TI para incorporar dinámicamente objetivos y KPI.

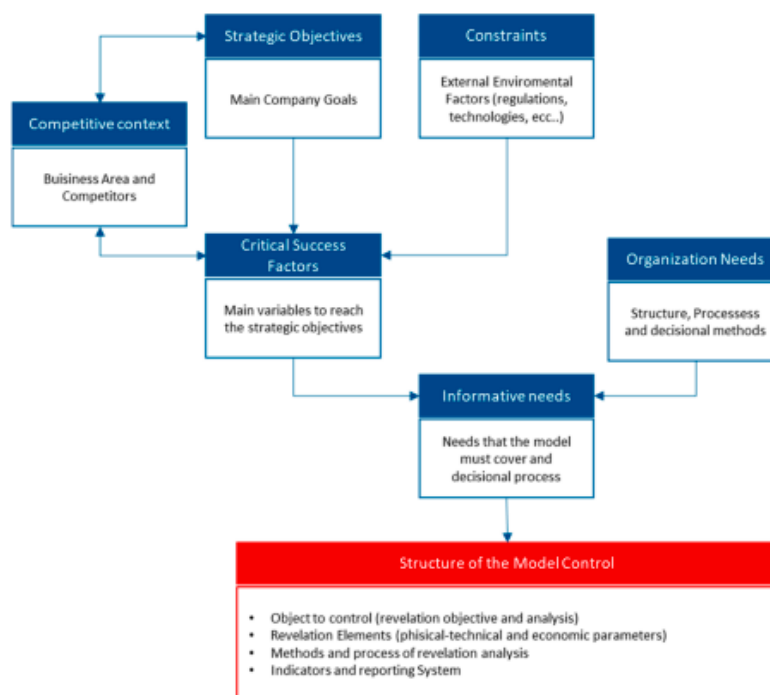


Ilustración 9: Modelo Propuesto por los autores

Fuente: (Vittoria Biagi, Riccardo Patriarca and Giulio Di Gravio, 2021)

Nota. Descripción general de un modelo de BI de gobernanza. Se visualiza el rectángulo rojo el cual representa la estructura objetivo del modelo, mientras que los rectángulos en azul muestran las áreas que influyen en la estructura final y las flechas las relaciones de los elementos.

Principal Resultado:

El enfoque propuesto por los autores demuestra que una solución de BI puede ayudar a tomar decisiones más rápidas y basadas en datos. Aunque la validación de este enfoque está más allá del alcance de este trabajo, el documento actual tiene como objetivo proporcionar evidencia ejemplar sobre la posibilidad de vincular la gobernanza con soluciones basadas en datos. Los resultados obtenidos por los autores en este trabajo abogan por el diseño de entornos de trabajo colaborativo y el abandono de la mentalidad de silos con respecto a los datos. A su vez, promueve el intercambio de datos y el monitoreo continuo del desempeño, particularmente en el área de gobierno, para garantizar el éxito de la empresa. Cabe resaltar que la solución de BI se

puede utilizar como un enfoque común para identificar cambios futuros y mejorar aún más la toma de decisiones.

4.2 Artículo de Investigación 2

Título: Business intelligence and organizational performance The role of alignment with business process management (Inteligencia de negocios y desempeño organizacional El rol de la alineación con la gestión de procesos de negocios)

Aporte:

Mediante el presente artículo, los autores realizan un estudio con el objetivo de mejorar la comprensión de cómo la madurez de BI se traduce en desempeño organizacional. Asimismo, La alineación de las iniciativas de BI y BPM parece una forma posible de crear valor empresarial con BI, especialmente porque BI permite la gestión y la medición del rendimiento de los procesos, lo que permite que la iniciativa de BI se centre más en el negocio. No obstante, los autores utilizaron un cuestionario para recopilar datos en organizaciones croatas y eslovenas con más de 50 empleados. De la misma manera, se desarrolló un instrumento de medición de alineación BI-BPM para este estudio utilizando el proceso recomendado de desarrollo y validación de escalas. Por otro lado, los autores obtuvieron dentro de sus estudios una mejor comprensión de cómo se materializa el impacto de BI en el rendimiento de la organización. Además, las iniciativas de BI y BPM deben estar alineadas para darle un valor comercial a BI.

Proceso:

El enfoque propuesto por los autores adopta los instrumentos de medición apropiados para las construcciones de madurez de Inteligencia de Negocios y el Rendimiento Organizacional. Por lo tanto, el proceso de recolección de datos tuvo dos fases. En primer lugar, se llevó a cabo una fase piloto el cual tiene como objetivo probar el cuestionario desarrollado en términos de la claridad de las preguntas y diseño de cuestionarios. Asimismo, en esta fase se llevó a cabo una serie de entrevistas, dando como resultado modificaciones menores del cuestionario. Durante la fase principal de recopilación de datos, de marzo a diciembre de 2016 se distribuyeron cuestionarios a las empresas que operan en Croacia y Eslovenia.

No obstante, los autores examinaron los datos más a fondo para detectar cualquier incoherencia, valores atípicos y valores faltantes, después de lo cual se eliminaron algunas respuestas de la base de datos. Teniendo en cuenta las propuestas de

investigación de este estudio, la base de datos final consistió en un total de 185 respuestas.

En resumen, el análisis de los datos y los resultados presentados proporcionan motivos para aceptar o rechazar las hipótesis de este estudio. Se puede concluir que se rechaza la primera hipótesis (H1), que afirma que la madurez de la inteligencia de negocios influye positivamente en la OP, ya que la influencia se reveló positiva, pero no estadísticamente significativa. A continuación, la segunda hipótesis (H2) se refinó mediante dos hipótesis: H2a, que indica que la madurez de BI influye positivamente en la alineación de BI-BPM, y H2b, que indica que la alineación de BI-BPM influye positivamente en OP. Los resultados presentados confirmaron ambas hipótesis. Por lo tanto, se puede concluir que esta investigación apoya la segunda hipótesis (H2), a saber, que el impacto de la madurez de BI en OP está mediado por la alineación BI-BPM.

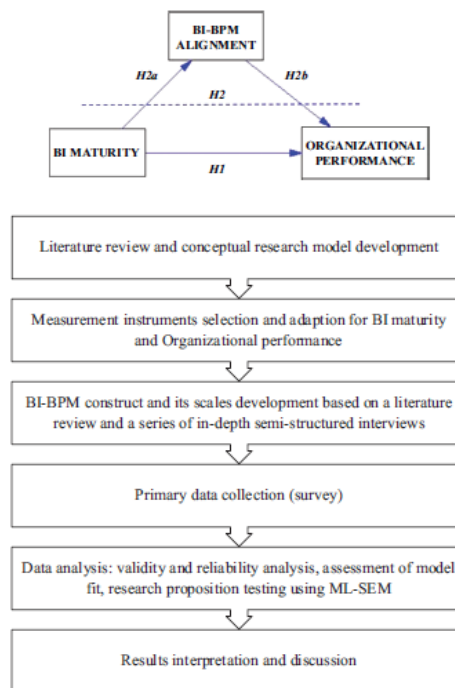


Ilustración 10: Alineación del proceso de negocio

Fuente: (Dalia Susa Vucec, Vesna Bosilj Vuksic and Mirjana Pejic Bach, 2020)

Nota. Una descripción general del modelo de investigación del impacto de la madurez de la inteligencia empresarial y su alineación con la gestión de procesos empresariales en OP.

Principal Resultado:

A nivel de estudio de los autores, los resultados muestran que el impacto de la madurez

de BI en OP es indirecto y está totalmente mediado por la alineación de BIBPM porque se confirma la hipótesis H2, pero no H1. Este hallazgo tiene importantes implicaciones teóricas y prácticas.

Primero, los autores desarrollaron y validaron el constructor de alineación BI-BPM utilizado en el estudio. Aunque algunas dimensiones de la alineación entre las iniciativas de BI y BPM se basa en los hallazgos de investigación donde se enfatizan la importancia de diferentes elementos al alinear las dos iniciativas. Tiene tres dimensiones La coordinación de iniciativas de BI y BPM, alineación de la terminología de BI y BPM y el uso de BI para medir el rendimiento del proceso.

En conclusión, los autores afirman que, para reducir la brecha entre la teoría y la práctica, sobre todo, se necesita educación y conciencia tanto en el lado del negocio como en el de TI. El movimiento de transformación digital también está trayendo cambios en esta área, ya que nos estamos dando cuenta de que la gerencia está cada vez más involucrada en TI y comprende mejor su papel en los negocios (Krotov, 2015; Dumeresque, 2014; Peppard, 2018), por lo que también podemos esperar mejoras en la alineación de BI BPM.

4.3 Artículo de Investigación 3

Título: Big data-assisted social media analytics for business model for business decision making system competitive analysis (Análisis de redes sociales asistido por Big data para el modelo de negocio para el análisis competitivo del sistema de toma de decisiones comerciales)

Aporte:

Los autores proponen un modelo de análisis de redes sociales para empresas asistido por Big Data (BD-SMAB) aumenta la conciencia y afecta a las tomas de decisiones en las estrategias de marketing. No obstante, utilizar el análisis de big data mejora la gestión de las empresas, permite evaluar a sus competidores en tiempo real y cambiar los precios, hacer tratos mejores que las ventas de sus competidores, analizar los comentarios desfavorables de los competidores y ver si pueden superar a ese competidor. En cuanto al método propuesto examina los impactos del análisis de las redes sociales en diferentes áreas, como bienes raíces, organizaciones y ferias comerciales de belleza. Posteriormente, ayuda a tomar mejores decisiones de marketing y desarrollar un enfoque estratégico. Por último, los autores presentan un

método BD-SMAB mejora la satisfacción y la experiencia del cliente y desarrolla el conocimiento de la marca.

Proceso:

El primer proceso que se realiza en esta investigación es la etapa de inteligencia en la cual se inspecciona el entorno para tomar decisión. Estos criterios comprenden los hechos esenciales, el conocimiento que se derivará y los resultados de estudios previos. Además, implica la recopilación y preparación de la información, en esta etapa se ha dividido en dos subtareas:

Decisión de comprensión y modelado el cual se describe y comprende el método, y luego el marco se refleja en los estándares de DMN. Los autores muestran una vista esquemática de análisis de redes sociales asistida por big data propuesta para el modelo empresarial (BD-SMAB). Por último, los autores han diseñado y analizado el modelo propuesto de análisis de redes sociales asistido por Big Data para empresas (BD-SMAB). Por lo tanto, esta investigación analiza los diferentes parámetros de simulación, como la mención en redes sociales, el índice de redes sociales, la compra de marca, etc.

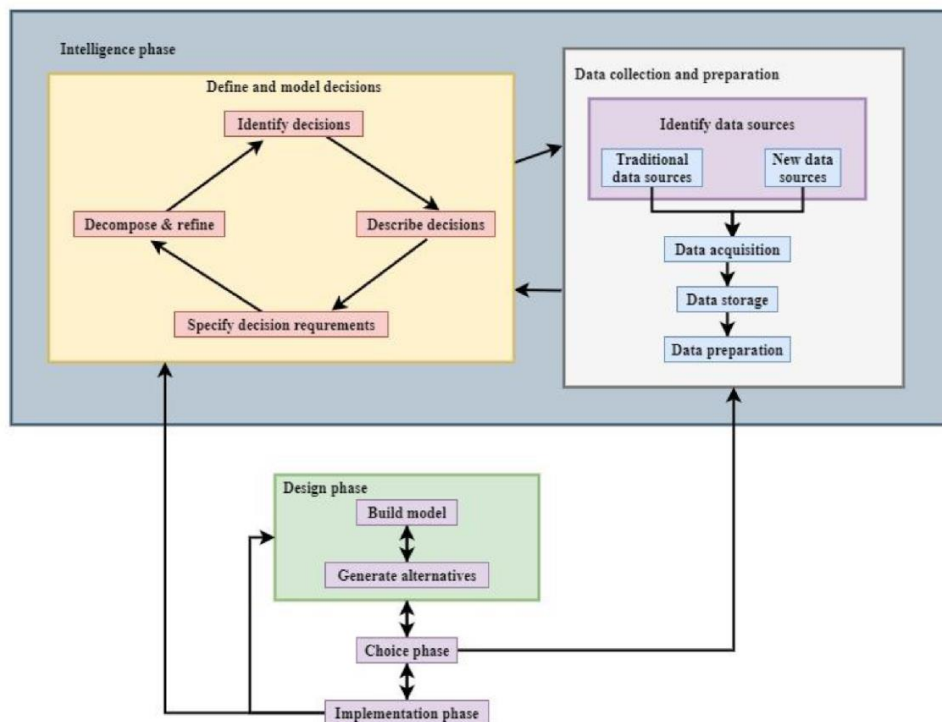


Ilustración 11: Análisis de redes sociales asistido por Big data para el modelo de negocio

Fuente: (Dalia Susa Vugec, Vesna Bosilj Vuksic and Mirjana Pejic Bach, 2020)

Nota. Una descripción general del modelo propuesto de análisis de redes sociales

asistido por Big Data para empresas (BD-SMAB)

Principal Resultado:

Tras evaluar el modelo se obtiene del estudio inicial un sistema que integra el mecanismo de toma de decisiones con big data y análisis y modelado. Los modelos de elección son las contribuciones más importantes del trabajo. El paradigma sugerido comienza con la comprensión y el modelado de elecciones. El enfoque analítico en las redes sociales se refiere a recopilar y evaluar sitios y blogs de redes sociales para la toma de decisiones. Por lo tanto, el enfoque analítico en las redes sociales se refiere a recopilar y evaluar sitios y blogs de redes sociales para la toma de decisiones.

Por último, el método propuesto de los autores examina los impactos del análisis de las redes sociales en diferentes áreas, como bienes raíces, organizaciones y ferias comerciales de belleza. Esta diversidad de estas empresas muestra los efectos de las redes sociales y cómo se pueden desarrollar decisiones positivas. Ayuda a tomar mejores decisiones de marketing y desarrolle un enfoque estratégico, además mejora la satisfacción y la experiencia del cliente y desarrolla el conocimiento de la marca.

4.4 Artículo de Investigación 4

Título: Performance Analysis of Big Data and Cloud Computing Techniques: A Survey

Aporte:

Los autores quieren dar a conocer como es el funcionamiento de un marco de trabajo en la nube como la agregación de componentes de servicios de desarrollo, middleware y bases de datos, esenciales para la computación en la nube, que permiten el desarrollo, la implementación y la administración potentes de aplicaciones basadas en la nube, lo que lo convierte en un modelo eficiente para escalar sus complejos recursos computacionales y asignados dinámicamente. Por este motivo, el autor proponer realizar encuesta de evaluación comparativa basada en el rendimiento de marcos de big data basados en la nube de empresas líderes como Amazon, Google, IBM y Microsoft, que ayudará a los investigadores, analistas de TI, lectores y usuarios comerciales a elegir los mejores Big Data basados en la nube. encuadre el marco de datos que mejor se adapte a sus necesidades. el trabajo garantiza el éxito en condiciones de resultados favorables.

Proceso:

Las plataformas de redes sociales en línea como Facebook, Whatsapp, Instagram y muchas otras aplicaciones comerciales en línea están dominando la era de los negocios, difundiendo diariamente enormes cantidades de datos (medidos en petabytes y exabytes). Este enorme volumen de datos, generado por mensajes, imágenes satelitales, redes sociales, correos electrónicos y más, puede ser tanto estructurado como no estructurado, y se conoce como Big Data. Como resultado, algunas empresas enfrentan la abrumadora tarea de rastrear y administrar una cantidad tan grande de datos, que los costosos almacenes de datos desarrollados por estas empresas se ven abrumados, lo que resulta en una gran carga de manejo. Para deshacer este cuello de botella, nació Big Data Analytics (BDA), en el que las organizaciones adquieren diversas herramientas, técnicas y métodos de big data para extraer los datos correctos de conjuntos de datos de este enorme material no estructurado. Apache Hadoop, que es de código abierto, es una de las herramientas elegidas por las empresas para proporcionar soluciones de descarga para datos de almacén y funciones de procesamiento. Hadoop, cuando se consolida con varios almacenes de datos, puede ser muy rentable y adecuado para las empresas. Otras empresas marco que están adoptando son programación Pig, Hive, Jaql, R y muchos otros. El análisis de estos grandes conjuntos de datos con BDA permite a las empresas de TI hacer predicciones prospectivas y mejorar la toma de decisiones, lo que ayuda a impulsar el crecimiento de los ingresos y a obtener una ventaja competitiva en el mercado. Sin embargo, la creciente popularidad y la amplia aceptación de los BDA en la industria de TI plantean muchos problemas y desafíos, dada la escala y el alto precio de los grandes datos. La nube proporciona una solución a este problema. La computación en la nube es un prototipo completo y fácil de usar para la entrega rápida de servicios a través de Internet, que permite el acceso bajo demanda a un grupo de recursos compartidos en la nube con una gestión de entrega mínima.

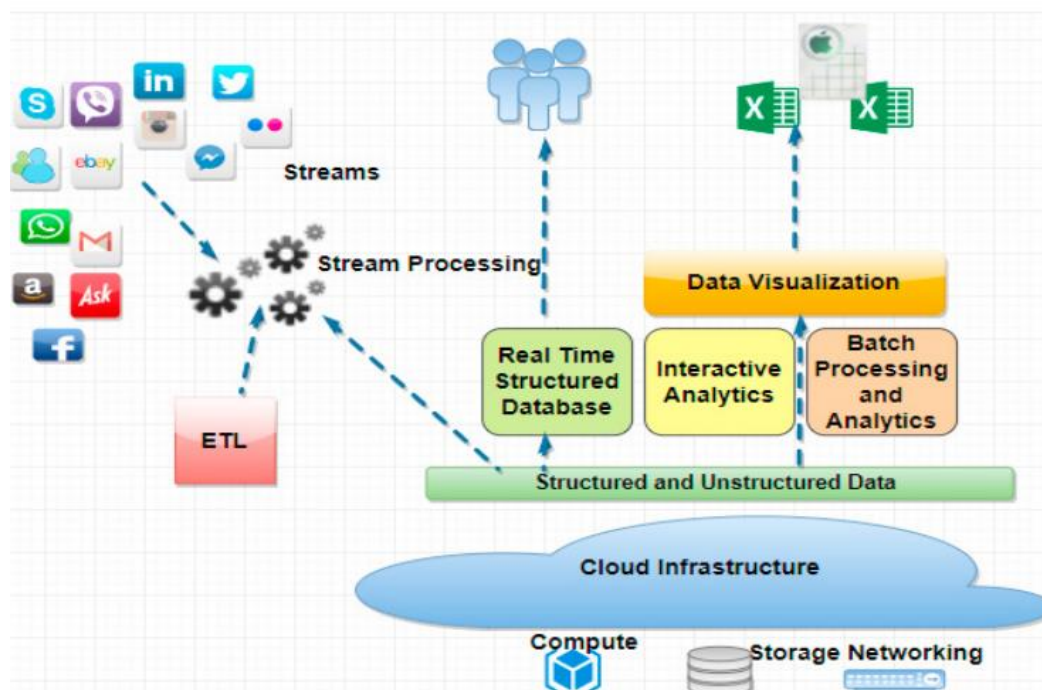


Ilustración 12: Diagrama de arquitectura de alto nivel del marco de Big Data basado en la nube

Fuente: (Subia Saif, Samar Wazir, 2018)

borrar

Principal Resultado:

La elección de un proveedor de plataforma en la nube adecuado depende de la perspectiva empresarial, las restricciones internas, los requisitos relacionados con el trabajo y el funcionamiento de las aplicaciones migradas. Donde AWS proporciona una plataforma bien integrada que promete distribuir sus servicios geográficamente, GCP opera un modelo de pago por segundo que confirma que paga exactamente por lo que usa, y como la amplia gama de servicios que proporciona tensorflow. varias aplicaciones de aprendizaje automático. MS Azure brinda la capacidad de ejecutar modelos basados en Microsoft RServer con soporte de Big Data y SLA del 99,99 % para una sola instancia de VM. IBM Cloud ofrece una solución completa con un sólido enfoque técnico para resolver los problemas de los clientes. Este estudio explora la comprensión profunda de esta marca sobre sus soluciones comerciales, servicios y características, así como sus aplicaciones y casos de uso, con comparaciones construidas entre ellos, donde beneficia a investigadores, analistas de TI, negocios de usuarios y lectores en más rápido. y fortalecer el apoyo a las decisiones, impulsando la innovación, la mejora y la implementación de estos valiosos marcos en un futuro

próximo.

4.5 Artículo de Investigación 5

Título: Data quality in ETL process: A preliminary study (Calidad de datos en el proceso ETL: un estudio preliminar)

Aporte:

El trabajo de investigación de los autores es demostrar las principales características de calidad de ETL. A la vez, brindan una descripción general de los enfoques de calidad de datos de proceso de ETL existente. Por ello, los autores presentan un estudio comparativo de algunas herramientas ETL comerciales para demostrar cuánto estas herramientas consideran las dimensiones de la calidad de los datos. Por último, los autores llevan a cabo experimentos utilizando una solución dedicada ETL (Talend Data Integration) y dimensiones de calidad de datos.

Proceso:

Durante la etapa de síntesis, los autores realizan la calidad del análisis multidimensional la cual se basa en la calidad de los datos extraídos, transformados y cargados. En primer lugar, se clasifican los enfoques existentes en enfoques orientados a procesos y a datos. Entre las quince dimensiones de calidad mostradas, para asegurar una mejor calidad en el proceso ETL, solo tres dimensiones (rendimiento, confiabilidad y reduplicación) han sido consideradas en trabajos existentes. Por otro lado, se realiza un comparativo basado en cuatro herramientas ETL para mostrar en qué medida estas herramientas consideran las dimensiones de calidad de datos. Para concluir, estas herramientas tienen muchos límites de calidad.

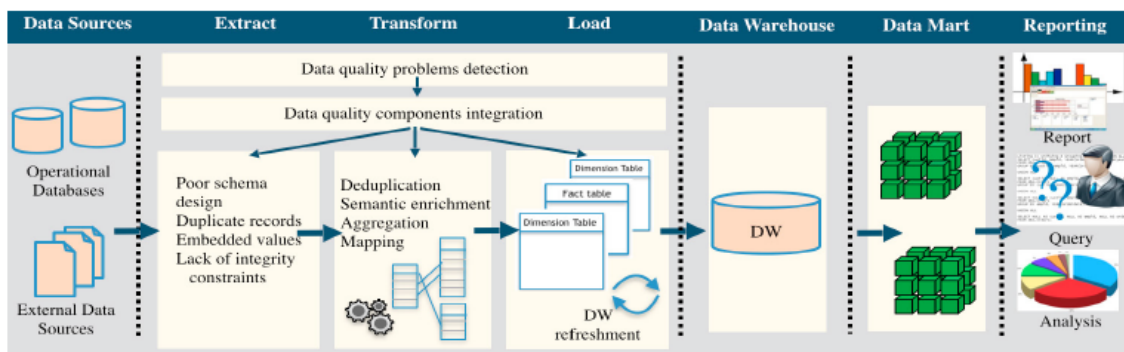


Ilustración 13: Diseño del modelo del autor

Fuente: (Ahmad Helmy Mohamad, Ghada Farouk Hassan, Ahmed S. Abd Elrahman, 2022)

Principal Resultado:

Los autores demuestran que comprender la semántica de los datos es de suma importancia en un contexto de integración de datos. Por lo tanto, una mayor integración del reconocimiento semántico sería de interés para resolver los problemas de calidad de datos (por ejemplo, conflictos de nombres y duplicación). El segundo punto para abordar es la detección de valores atípicos. De hecho, principalmente estamos estudiando la detección de valores atípicos en el flujo de documentos de estructura basada en JSON. La implementación de este enfoque está en progreso. Consiste en la detección de valores atípicos en tiempo real en el flujo de documentos JSON. Con la gran cantidad de datos generados por miles de millones de (por ejemplo, Internet de las cosas), el procesamiento de flujo se ha convertido en un potencial

Título: OLAP Patterns: A pattern-based approach to multidimensional data analysis (Patrones OLAP: un enfoque basado en patrones para el análisis de datos multidimensionales)

Aporte:

El aporte de los autores establece la noción básica de patrones OLAP y como esto se aplicó en el contexto de datos abiertos vinculados en la web semántica. Por lo tanto, el análisis de datos se desarrolló de manera ad hoc, por necesidad, adaptado a las necesidades específicas del proyecto agriProKnow, un subproducto para hacer frente a las incertidumbres relacionadas con los requisitos del proyecto. No obstante, los autores presentan una definición formal de los patrones OLAP, así como una conceptualización y un lenguaje de definición más expresivo, flexible y de aplicación general, mientras relaja algunas de las restricciones y suposiciones originales.

Proceso:

De manera similar a los métodos clásicos, el enfoque propuesto por los autores basado en patrones para el análisis de datos multidimensionales no modela directamente las consultas OLAP conceptualmente. Más bien, especifica una interfaz para cuándo y cómo redactar la consulta OLAP correspondiente.

Sin embargo, los autores comparten características comunes como la generación automática de código. Aunque no consideramos el MDA desde el principio, el enfoque propuesto se basa en patrones que se puede clasificar como ingeniería dirigida por modelos, ya que permite la ejecución automática de código a partir de un modelo de

patrón. Además, nuestro enfoque basado en patrones no realiza la transformación de modelo a modelo. Más bien, los patrones OLAP se transforman directamente en código específico de la plataforma. En cambio, nuestro enfoque basado en patrones asume que el modelo multidimensional conceptual también representa los datos almacenados. La transformación de consultas OLAP definidas en OCL es similar a la ejecución de patrones OLAP, tal y como apuntan Pardillo et al. use macros parametrizadas que se reemplazan por los fragmentos de código correspondientes.

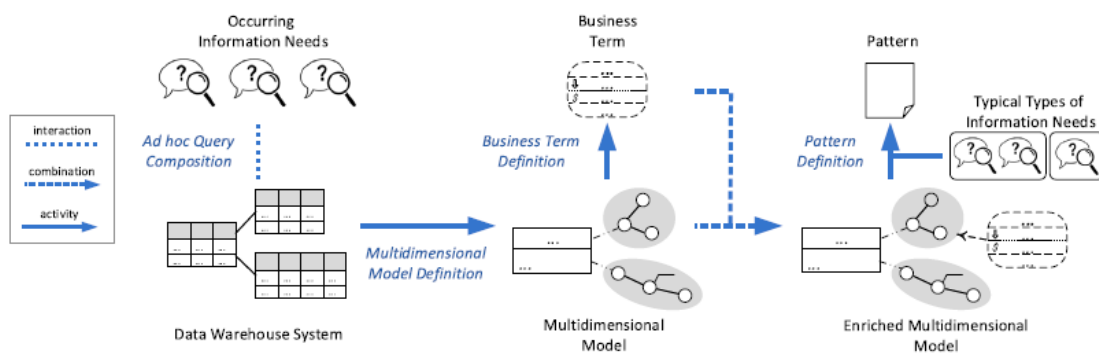


Ilustración 14: Diseño del modelo del autor

Fuente: (Ilko Kovacic, Christoph G. Schuetz , Bernd Neumayr, Michael Schrefl, 2022)

Principal Resultado: Los autores demuestran que preservar de manera explícita una separación de textura ayuda a lograr una mejor generalización que los enfoques de mapeo directo. La capacidad de generalización del modelo propuesto sigue presentando limitaciones, ya que presenta problemas para generalizar de manera correcta cuando una persona es procesada a una escala notablemente diferente a la del conjunto de entrenamiento. Además, los avatares texturizados presentan errores de estimación en las manos y la cara.

El principal punto que analizan los autores abarca las variables atómicas para parámetros y elementos derivados también pueden tener tipos creados por constructores de tipos familiares aplicados en tipos atómicos y complejos para admitir matrices (indexadas por números enteros), mapas (pares clave-valor), tuplas (binario, ternario, n-ario), y combinaciones de estos, como conjuntos de tuplas. En segundo lugar, además de las variables complejas, el enfoque del patrón OLAP puede ampliarse

mediante la opcionalidad de las variables atómicas y de colección. Las variables opcionales permitirían a los usuarios omitir enlaces de nombres a parámetros que no son necesarios para responder a una pregunta comercial en particular, lo que aumenta el alcance de los patrones OLAP. En tercer lugar, los conjuntos de valores presentados pueden extenderse por la escala de medida, es decir, valores nominales, ordinales, de intervalo y de razón, un tipo base con las condiciones que deben cumplir los valores, una enumeración de valores y por una unidad de medida: conversiones. entre estas unidades se pueden registrar por separado. En cuarto lugar, la expresividad de los patrones OLAP se puede aumentar ampliando las restricciones en patrones y términos comerciales para considerar las relaciones acumuladas entre niveles y describir las relaciones entre niveles y atributos.

Finalmente, la introducción de términos comerciales genéricos podría reducir la necesidad de definir términos comerciales desde cero, ya que se pueden definir una vez y especializarse varias veces a través de la creación de instancias, lo que genera varios términos comerciales (específicos) al proporcionar, además de los parámetros de contexto, parámetros que permiten la adaptación de restricciones y plantillas de expresión.

4.6 Artículo de Investigación 6

Título: Customer experience management in the age of big data analytics: A strategic framework (Gestión de la experiencia del cliente en la era del análisis de big data: un marco estratégico)

Aporte:

Los autores presentan un análisis en base a desarrollos recientes en el análisis de big data, con el fin de desbloquear los conocimientos del cliente para la gestión de la experiencia al cliente. Por lo general, la investigación de estos dos campos es escasa y existe la necesidad de un trabajo conceptual que brinde una descripción general de las oportunidades para usar análisis de big data en la gestión de la experiencia del cliente. Por lo tanto, los autores proponen desarrollar un marco estratégico para CXM (Gestión de la experiencia del cliente) basado en los conocimientos de CX (Experiencia del cliente) resultantes de BDA (Análisis de big data). Por último, se proporciona una guía paso a paso sobre cómo poner en marcha o implementar nuestro marco estratégico.

Proceso:

Los puntos que los autores se enfocan es señalar la gran cantidad de datos generados en diferentes puntos de contacto a lo largo del viaje del cliente y brindar orientación

sobre cómo las organizaciones pueden pasar de los datos de CX a las acciones de CX para CXM. Por lo tanto, proporcionan un marco integral e integrador en el que se desarrollan conocimientos de CX a través de análisis de CX de datos de CX. A lo largo del análisis los autores enfatizan que más datos, más análisis y más conocimientos no conducen necesariamente a una mejor CXM. Se considera una integración entre BDA y CXM en el cual destacan cómo una organización puede aplicar diferentes tipos de análisis de CX a diferentes tipos de datos de CX, con el objetivo de obtener diferentes tipos de conocimientos de CX que se pueden usar para tomar diferentes tipos de acciones de CX. Por último, se espera que el marco integrador pueda ayudar tanto a los profesionales como a los investigadores a reflexionar sobre las crecientes complejidades del uso de BDA para CXM.

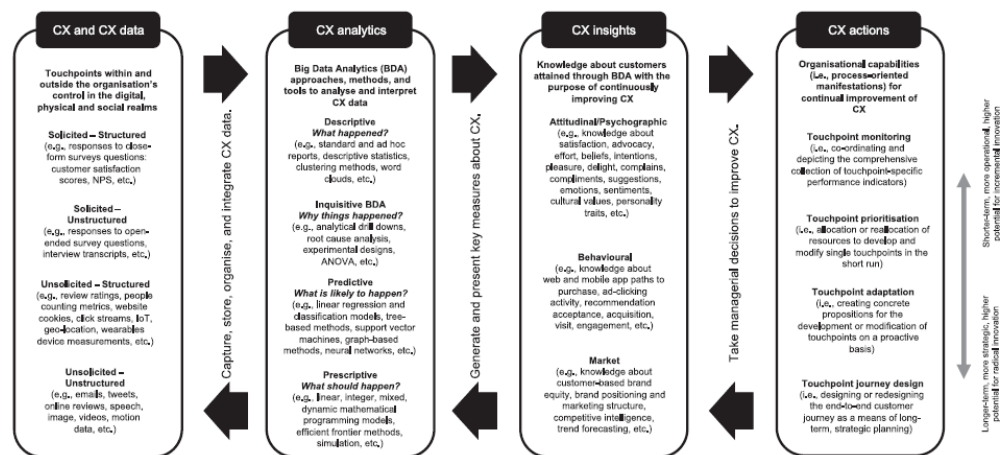


Ilustración 15: Diseño del modelo del autor

Fuente: (Maria Holmlunda, Yves Van Vaerenbergh, Robert Ciuchitaa, Annika Ravalde, Panagiotis Sarantopoulos, Francisco Villarroel Ordenese, Mohamed Zakif, 2020)

Principal Resultado:

Con el fin de que los autores obtengan diferentes tipos de información de experiencia al cliente y se pueda usar para tomar diferentes tipos de acciones. Por otro lado, se evidenció que las organizaciones se encuentran en diferentes etapas de madurez cuando se trata de gestión de experiencia al cliente impulsado por el análisis de big data, ya que las oportunidades y los desafíos de una organización novata, no nativa digital y experimentando difieren significativamente de los de una organización experimentada con una fuerte orientación CXM. Por lo que se recomienda desarrollar métricas de experiencia al cliente, medidas de calidad de datos de experiencia al cliente, cajas de herramientas de análisis de experiencia al cliente y requisitos de

información de experiencia al cliente ajustados a diferentes fases de madurez implicaría una contribución teórica y sería de utilidad práctica.

4.7 Artículo de Investigación 7

Título: Serverless computing for cloud-based power grid emergency generation

Aporte: El aporte de la solución demostró dos beneficios clave. Primero, la infraestructura de la nube es independiente y tolerante a fallas, lo que proporciona control y monitoreo de la red incluso si el EMS pierde la funcionalidad correspondiente o cuando los operadores necesitan trabajar de forma remota desde el centro de control. En segundo lugar, el diseño general se basa en eventos utilizando servicios en la nube sin servidor para responder a eventos de falla SCADA/EMS. Gracias a "sin servidor", los usuarios pueden evitar la carga del aprovisionamiento y el mantenimiento del servidor. El costo de usar servicios de nube pública para esta solución es extremadamente bajo porque está diseñado e implementado en base al modelo FaaS basado en eventos. Este trabajo también desarrolla un mecanismo integral de ciberseguridad para abordar los requisitos de infraestructura crítica para la red, que puede servir como un marco ejemplar para que otros operadores de red protejan sus servicios en la nube.

Proceso:

El autor presenta una solución sin servidor para realizar todos los despachos de emergencia basados en servicios de nube pública. En este trabajo, continuamos utilizando datos de PMU independientes de SCADA. A diferencia de [11], que desarrolló un esquema de orquestación basado en PMU que funciona en el centro de control para contrarrestar los efectos de las fallas de SCADA/EMS, este trabajo presenta una arquitectura completa completamente nueva con ingesta, procesamiento, almacenamiento, visualización y uso compartido de datos almacenados en la nube para garantizar la continuidad del negocio en caso de que las instalaciones del centro de control no estén disponibles o sean inaccesibles. Esta arquitectura centrada en la nube tiene dos ventajas principales sobre una arquitectura de solución tradicional que depende en gran medida de la infraestructura local. En primer lugar, la infraestructura de la nube es independiente y tolerante a fallas, lo que proporciona monitoreo y control

de red redundantes incluso si el EMS pierde la funcionalidad correspondiente o cuando los operadores necesitan trabajar de forma remota desde el centro de control de la red. En segundo lugar, la plataforma centrada en la nube se basa en el modelo FaaS basado en eventos. Esta arquitectura sin servidor libera a los clientes de la carga de la administración del servidor y elimina los costos de aprovisionamiento del servidor. Los usuarios solo pagan un costo operativo muy bajo para utilizar el servicio de activación de "emergencia" debido a su corta duración y baja frecuencia. Otra contribución de este trabajo es que propone e implementa un marco de ciberseguridad ejemplar que cumple con los requisitos de infraestructura energética crítica. Esto beneficiará a las empresas de servicios públicos que busquen soluciones basadas en la nube para sus casos de uso empresarial.

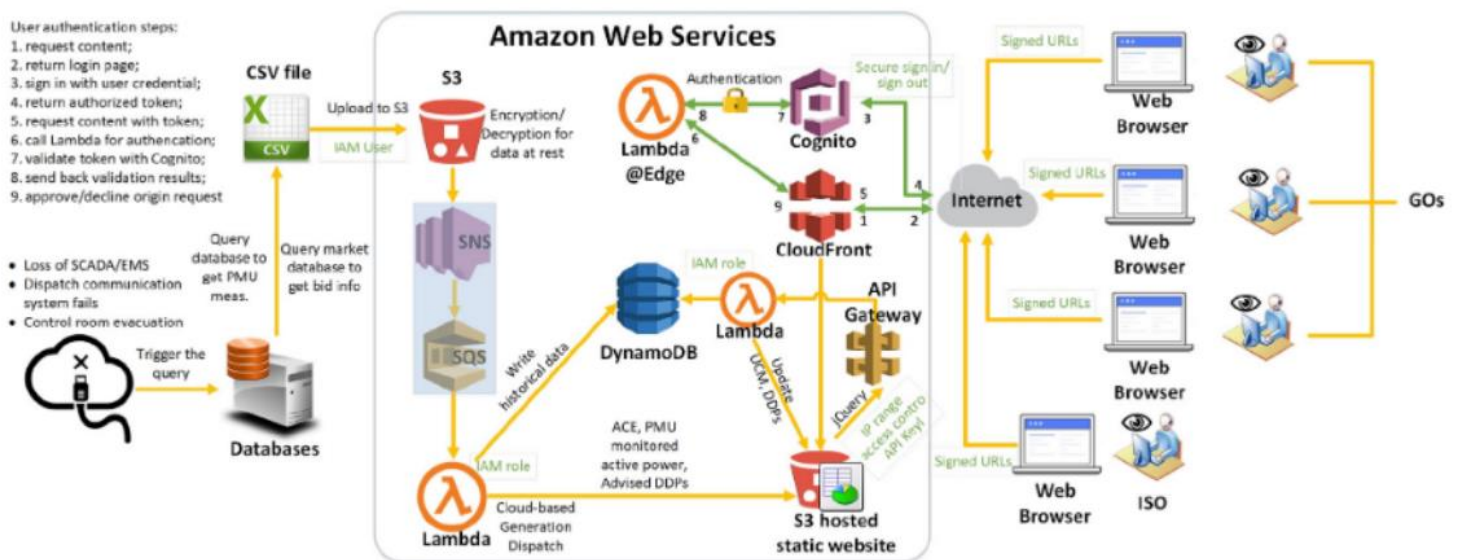


Ilustración 16: Diseño del modelo del autor

Fuente: (Song ZhangXiaochuan LuoEugene Litvinov (2021))

Principal Resultado:

La solución no solo puede realizar envíos rentables para mantener el equilibrio del sistema, sino que también permite compartir de forma rápida, precisa y segura la información de envío de emergencia, como señales de control, sistema de equilibrio de rendimiento de datos, sistema y operador. Con esta solución basada en la nube, se pueden evitar posibles errores humanos causados por el uso de conversaciones telefónicas para enviar manualmente de boca en boca. Cada transacción realizada a través de la plataforma se registra en la base de datos en un formato de datos bien definido, lo que facilita enormemente la auditoría y la rendición de cuentas. Más

importante aún, esta solución permite a los operadores monitorear y controlar la red incluso cuando tienen que trabajar de forma remota fuera de la sala de control, siempre que tengan una conexión a Internet.

4.8 Artículo de Investigación 8

Título: Digitalization process of complex B2B sales processes – Enablers and obstacles (Proceso de digitalización de procesos de ventas B2B complejos – Habilitadores y obstáculos)

Aporte:

La investigación de los autores es proporcionar una comprensión matizada del uso de la tecnología en procesos de ventas B2B complejos, puesto que digitalización difiere entre procesos de ventas B2B simples de productos físicos estandarizados y procesos de ventas B2B complejos que incluyen más ofertas de ventas basadas en soluciones (combinación de servicios y productos). El objetivo de este análisis es arrojar luz sobre la medida en que los vendedores utilizan la tecnología en procesos de ventas B2B complejos. Por lo tanto, este estudio propone condiciones de contorno para situaciones en las que la tecnología aumenta la productividad de los vendedores en procesos de ventas B2B complejos.

Proceso:

Los autores proponen que una empresa consultora proponga una hoja de ruta para el proceso de digitalización, incluida la digitalización del proceso de ventas y el tratamiento de los datos de ventas. A su vez, necesitan crear una nueva posición dentro de sus técnicas para promover la digitalización entre los empleados de la empresa y resolver cualquier duda sobre la digitalización (por ejemplo, cómo usar LinkedIn). Por lo que se recomienda convertir los obstáculos en facilitadores en la digitalización del complejo proceso de ventas B2B de la siguiente manera:

Hacer que la subdimensión cultural funcione en la empresa a través de valores y objetivos: alinee los valores de las personas con los objetivos a largo plazo de su organización para liberar todo el talento oculto entre el personal.

Hacer que la subdimensión organizacional funcione en la empresa a través de habilidades y expectativas: desarrollar las habilidades digitales profesionales del personal para que puedan realizar el trabajo que se espera de ellos; y dentro y más allá de la organización y la productividad – trabajar en los procesos internos y externos de digitalización de la organización para promover la productividad en el complejo

proceso de ventas B2B.

Hacer que la subdimensión tecnológico funcione en la empresa a través de la adecuación de las herramientas a los empleados y procesos organizacionales.

Hacer funcionar el subdimensión legal a través del conocimiento de leyes específicas y el conocimiento de los límites tecnológicos legales

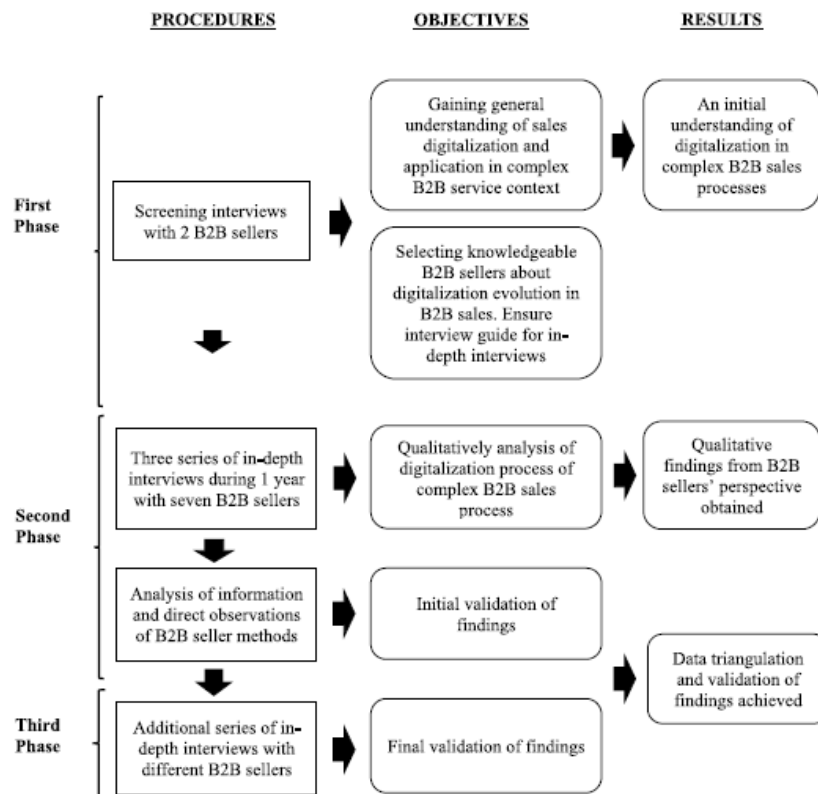


Ilustración 17: Digitalization process of complex B2B sales processes – Enablers and obstacles

Fuente: (Rocio Rodríguez, Goran € Svensson , Erik Jens Mehl, 2020)

Nota. Procedimientos metodológicos, objetivos y resultados.

Principal Resultado:

Los autores contribuyen a revelar subdimensiones en la digitalización del complejo proceso de ventas B2B, tales como aspectos organizacionales, tecnológicos, culturales y legales, así como de seguridad. A su vez, al conocimiento sobre el proceso de digitalización de las ventas directas y, específicamente, a la comprensión del proceso de digitalización de servicios complejos en contextos B2B. Por lo tanto, este estudio también contribuye a un enfoque para interconectar las funciones de marketing y ventas directas, lo que impacta en la organización interna de las empresas y su comunicación externa con los clientes y otras partes interesadas.

4.9 Artículo de Investigación 9

Título: Integration of business process data and organizational data for evidence-based business intelligence (Integración de datos de procesos comerciales y datos organizacionales para inteligencia de negocios basada en evidencia)

Aporte:

La investigación de los autores es el análisis de escenarios para los problemas de integración de datos y coincidencia de datos para vincular los datos de proceso y los datos organizacionales, considerando tanto los BP interorganizacionales como BP colaborativos interorganizacionales; y una propuesta basada en un metamodelo para la integración de varias fuentes de datos que contienen datos de procesos y datos organizacionales, generando un modelo unificado que se utilizará como entrada para extraer datos de procesos y datos organizacionales en conjunto.

Proceso:

Los autores hacen énfasis en la exploración de la relación entre la ciencia de datos y la ciencia de procesos a través de la minería de procesos, y en se define un cubo de proceso para analizar y explorar procesos de forma interactiva en función de una vista multidimensional sobre datos de eventos, tomando como entrada logs que ya contienen datos de proceso y datos organizativos. En este contexto, los autores emplean un algoritmo para fusionar registros de eventos de diferentes bases de datos para analizar procesos entre organizaciones, y los autores proponen construir múltiples modelos de puntos de vista a partir de bases de datos, proporcionando una visión holística de un proceso, de modo que es posible generar automáticamente muchos eventos posibles registros (uno para cada punto de vista de interés) para aplicar técnicas de minería de procesos. Todas estas obras observan el problema de la minería de procesos y no de la explotación conjunta de ambas fuentes de información. En cambio, se está tomando una visión integrada de los datos de ejecución en la organización, proporcionando una visión completa del comportamiento real de los procesos de la organización. Por ello, los autores proponen una integración integral de proceso y datos información en un formato consistente y unificado a través de la definición de un metamodelo. Sin embargo, se enfocan en la extracción de registros de eventos de lectura/escritura de una base de datos. Por lo tanto, las actividades a nivel empresarial están ocultas y el análisis se centra en el nivel inferior de las operaciones de la base de datos.

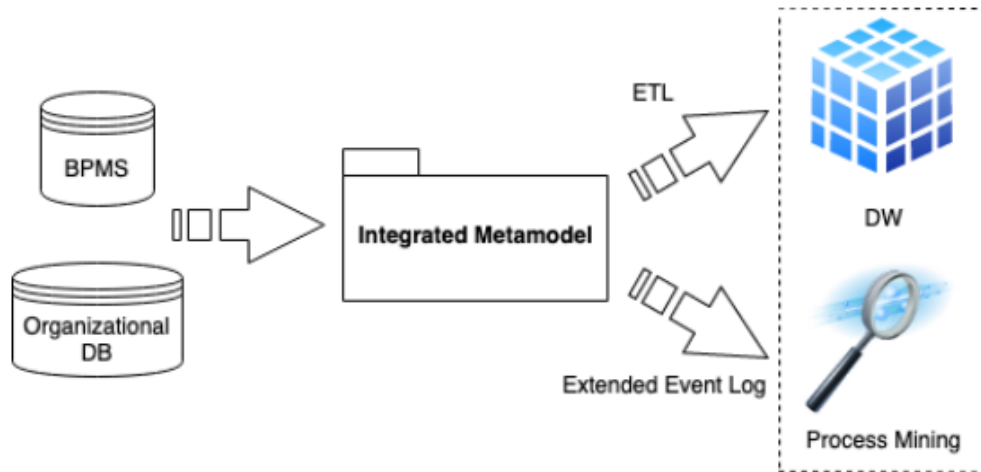


Ilustración 18: Integración de los metamodelos

Fuente: (Andrea Delgado, Daniel Calegari, Alexis Artus, Andrés Borges, 2021)

Principal Resultado:

Los autores contribuyen en proporcionar una organización de inteligencia de negocios basada en evidencia necesaria para mejorar su operación diaria, requiere una visión integrada de los datos del proceso y los datos organizacionales. En este paper se ha explorado cómo se puede lograr esta visión integrada. Primero se analiza el problema identificando varios escenarios para procesos colaborativos intra e interorganizacionales. Luego se proporciona una propuesta de solución inicial mediante la definición de un metamodelo integrado que podría ser la entrada para aplicar técnicas de minería de datos y procesos. También mostramos cómo funciona esto desde una perspectiva práctica a través de un ejemplo de aplicación. Desde una perspectiva general, se trabajó en muchos otros aspectos relacionados con el marco integrado. Por ejemplo, extraer información BPMS de un modelo de datos unificado, desacoplar el proceso ETL de un modelo de motor de proceso específico y la calidad de los datos desde un punto de vista metodológico y técnico.

4.10 Artículo de Investigación 10

Título: Organizational business intelligence and decision making using big data analytics (Inteligencia de negocios organizacional y toma de decisiones usando análisis de big data)

Aporte:

Los autores proponen un marco de Gestión Optimizada de Datos utilizando Big Data Analytics (ODM-BDA) para aumentar la eficacia organizativa inteligente y el análisis de la toma de decisiones en las organizaciones. No obstante, se introduce un método de retroceso en los entornos de inteligencia comercial y toma de decisiones para mejorar el fracaso del plan y la capacidad de asumir riesgos. La estrategia optimizada empinada se incorpora en el marco ODM-BDA para mejorar el plan de capacitación y administrar el dinero.

Proceso:

El proceso de la propuesta de los autores es la siguiente:

Modelos para servicios en la nube

El objetivo principal de la mayoría de las empresas es mejorar la experiencia del cliente. Otros objetivos incluyen marketing objetivo mejorado, reducciones de costos y eficiencia mejorada del proceso.

Tecnológicas en Big Data

El marco puede gestionar grandes cantidades de información restante o de difusión. La importancia de los datos es importante y crece constantemente en la toma de decisiones. Esto permite a las empresas crear nuevas oportunidades de negocio y generar más ingresos. Además, el autor indica que ayuda a predecir tendencias futuras. Por lo tanto, las empresas toman o deshacen decisiones comerciales basadas en datos. Al hacer predicciones sobre eventos futuros, los sistemas operativos de inteligencia empresarial deben procesar grandes cantidades de datos complejos. Además de la detección y el marketing de fraudes, las empresas utilizan tecnologías de análisis predictivo para la evaluación de riesgos, la gestión y la investigación de mercado.

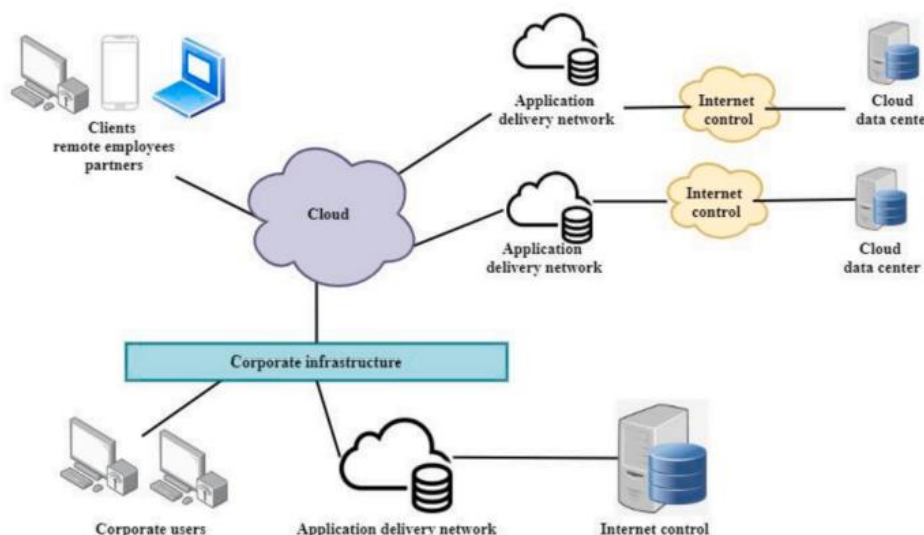


Ilustración 19: Modelo organizacional BI y la toma de decisiones

Fuente: (Yanfang Niu, Limeng Ying, Jie Yang, Mengqi Bao, C.B. Sivaparthipan, 2021)

Arquitectura en la nube de la gestión de datos optimizada propuesta mediante el marco de Big Data Analytics (ODM-BDA).

Implementación de análisis de servicios en la nube a gran escala

El análisis de datos basado en la nube es un sistema comercial que proporciona componentes del proceso de análisis de Big Data a través de una plataforma basada en la nube.

La información recopilada en un repositorio basado en la nube puede ayudar a las empresas a tomar decisiones. Los analistas usan big data basado en la nube para administrar múltiples registros con muchas características y proporcionan más información para funcionar. Permite la previsibilidad. Los grandes datos y la tecnología de la información a menudo permiten a los investigadores analizar regularmente información nueva, como sitios web populares.



Ilustración 20: Arquitectura de Análisis de Datos

Fuente: (Yanfang Niu, Limeng Ying, Jie Yang, Mengqi Bao, C.B. Sivaparthipan, 2021)

Nota. Describe la Arquitectura de análisis de datos de la gestión de datos optimizada propuesta mediante el marco de Big Data Analytics (ODM-BDA)

Principal Resultado: Los autores indican que las empresas han utilizado el análisis de datos durante mucho tiempo para mejorar su plan, maximizar la rentabilidad y promover sus métodos de toma de decisiones. La computación en la nube y la tecnología de big data ahora se reconocen generalmente como dos innovaciones predominantes. Asimismo, una empresa puede beneficiarse de la inversión en el análisis de la nube, y es importante asegurarse de que la preparación adecuada cubra todas las etapas de los elementos analíticos. Por lo tanto, los resultados experimentales muestran que nuestro método propuesto ODM-BDA tiene un valor de relación de precisión de 96,1 %, una tasa de error baja de 0,66 %, una tasa de eficiencia de 97,2 %, una tasa de precisión de 96,3 %, una tasa positiva verdadera de 80,6 % y menos retraso proporción del 52,8%.

4.11 Artículo de Investigación 11

Título: Big Data, Analytic Culture and Analytic-Based Decision Making – Evidence from Australia

Aporte: El autor investiga cómo la toma de decisiones gerenciales está influenciada por Big Data, análisis y cultura analítica. Los resultados de una encuesta transversal (n = 163) de gerentes senior de TI revela que Big Data Analytics crea un incentivo para que los gerentes basen más de sus decisiones en los conocimientos analíticos. Sin embargo, también encontramos que el principal impulsor de la toma de decisiones basada en análisis es la cultura analítica.

Proceso: Esta investigación es de naturaleza exploratoria, se seleccionó una encuesta transversal como la investigación más adecuada. Los procedimientos de encuesta utilizados siguen las sugerencias de Dillman. Como cada variable en las hipótesis es latente, la construcción de indicadores y escalas adecuadas fue esencial. Este proceso se guió por la literatura académica, pero cuando fue necesario, también se consultó la literatura profesional.

Dimensiones

En esta investigación, la sofisticación de BDA se operacionaliza a lo largo de dos dimensiones: Las herramientas analíticas se refieren a las aplicaciones de software que los profesionales analíticos utilizan en el análisis de datos, como hojas de cálculo,

herramientas de inteligencia comercial (BI), paquetes estadísticos, suites de minería de datos, herramientas de visualización de datos y herramientas informáticas de alto rendimiento.

Respuesta de las encuestas

La dispersión de las respuestas refleja la industria australiana, como lo muestra el coeficiente de variación (CoV). El CoV de las respuestas de la encuesta es de 844 mientras que la de la industria australiana es .843. El 84% de las respuestas provinieron de organizaciones con más de 100 empleados equivalentes a tiempo completo, y los CIO y otros gerentes senior de TI fueron casi igualmente representados (52,1% y 47,9%).

Principal Resultado: La investigación presentada en este documento intentó responder a las siguientes preguntas de investigación: (1) ¿Están los gerentes en las organizaciones con Big Data Analytics sofisticados tienen más probabilidades de basar sus decisiones en análisis (hechos, evidencia) que los gerentes de las organizaciones con poco análisis de Big Data? (2) ¿El nivel de cultura analítica organizacional influye en (a) la cantidad de información analítica que utilizan los gerentes en la toma de decisiones, y (b) el nivel de sofisticación de Big Data Analytics de una organización? Los resultados de nuestro análisis sugieren que la respuesta a estas preguntas es uniformemente "sí". Gerentes en empresas que usan herramientas y métodos BDA sofisticados tienden a basar sus decisiones más en análisis que los gerentes en BDA bajo empresas. Sin embargo, la cultura analítica en una organización es un predictor mucho más fuerte de la toma de decisiones basada en análisis que la sofisticación de las prácticas Análisis de Big Data.

4.12 Artículo de Investigación 12

. Título: Implementation of Business Intelligence for Sales Management

Aporte: Las empresas de hoy operan en un entorno socioeconómico cada vez más exigente. En tal situación, se ve obligado a adoptar un enfoque competitivo explotando al máximo la información a su disposición para desarrollar planes de acción apropiados y tomar decisiones efectivas. Los sistemas de soporte de decisiones brindan a las empresas herramientas para ayudar a las empresas a tomar decisiones basadas en técnicas y metodologías en el campo de las matemáticas, como la optimización, la

estadística y la teoría de la decisión. Los sistemas de apoyo a la toma de decisiones están formados por diferentes componentes, como almacenes de datos, herramientas ETL y herramientas de informes y análisis.

Proceso: Las empresas de hoy operan en un entorno socioeconómico cada vez más exigente. En tal situación, se ve obligado a adoptar un enfoque competitivo explotando al máximo la información a su disposición para desarrollar planes de acción apropiados y tomar decisiones efectivas. Los sistemas de soporte de decisiones brindan a las empresas herramientas para ayudar a las empresas a tomar decisiones basadas en técnicas y metodologías en el campo de las matemáticas, como la optimización, la estadística y la teoría de la decisión. Los sistemas de apoyo a la toma de decisiones están formados por diferentes componentes, como almacenes de datos, herramientas ETL y herramientas de informes y análisis.

4.13 Artículo de Investigación 13

Título: BICenter: A collaborative Web ETL solution based on a reflective software approach(BICenter: Una solución ETL web colaborativa basada en un enfoque de software reflexivo)

Aporte:

La investigación se enfoca en las técnicas que subyacen a este objetivo generalmente se conocen como canalizaciones de extracción, transformación y carga (ETL). Además, tiene como objetivo extender la canalización a otras fuentes de datos y requiere agregar nuevos conceptos y reglas a lo largo del tiempo, lo que implica una actualización continua de los scripts ETL. Por tal motivo, los autores presentan una aplicación ETL colaborativa basada en la web que permite a los usuarios diseñar, compartir y ejecutar canalizaciones ETL en varios centros. El sistema está respaldado por una interfaz fácil de usar en la que los usuarios sin conocimientos técnicos pueden crear canalizaciones de ETL sin necesidad de comprender los detalles de ETL y, lo que es más importante, sin tener acceso directo a los datos.

Proceso:

El proceso presentado por los autores se enfoca en BICenter el cual se desarrolló no solo para reemplazar las herramientas que se utilizan actualmente para los flujos de trabajo de ETL, sino también para ampliar la diversidad de soluciones utilizando un

enfoque basado en la web. Para seguir siendo compatible con las herramientas existentes, aprovecha las funciones de Kettle para administrar los procesos ETL localmente, en lugar de crear un nuevo núcleo. A pesar de tener un entorno colaborativo, BCenter también permite la ejecución y definición de flujos de trabajo ETL de forma remota, sin acceder a los servidores de origen y destino. Esta estrategia simplifica la gestión de los flujos de trabajo de ETL y la estrategia utilizada en la implementación proporciona una capa de seguridad para acceder a servidores privados. Estas características simplifican las tareas de los equipos de TI responsables del manejo de datos.

El software ahora se lanza como código abierto, sin embargo, se puede usar en entornos comerciales como una herramienta de apoyo. Actualmente, existe una iniciativa en un proyecto europeo para incentivar a las instituciones de salud a adoptar un modelo de datos común como se mencionó anteriormente. En este proyecto, algunas pequeñas y medianas empresas (PYMES) recibieron formación para ayudar a estas instituciones a migrar sus datos a este nuevo esquema. BCenter puede ser un gran activo para esta tarea, soportando todas las etapas del proceso ETL.

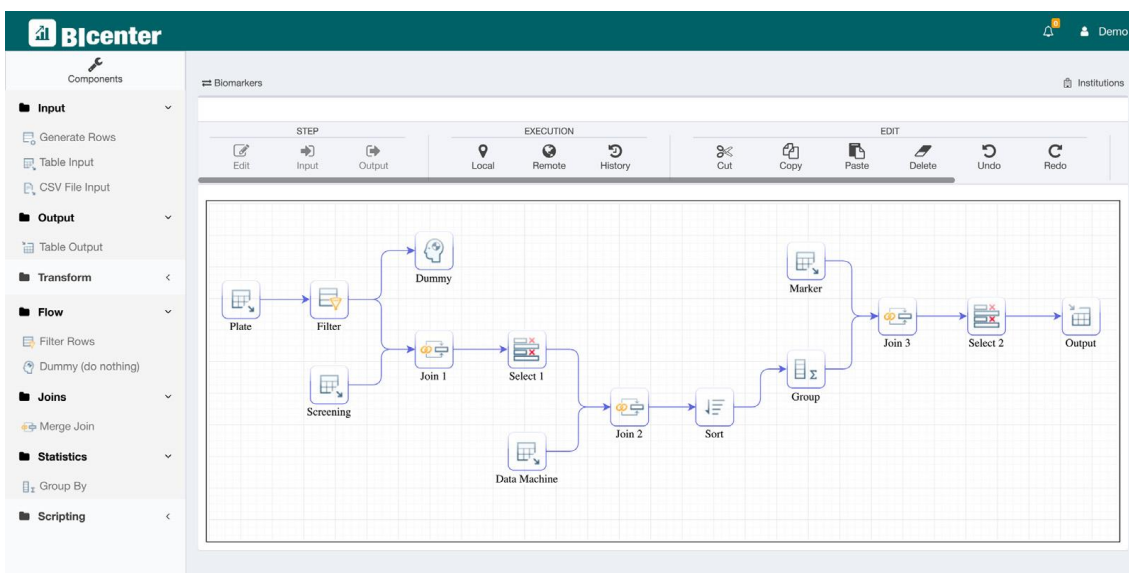


Ilustración 21: Modelo ETL de Biomarcadores

Fuente: (João R. Almeida, Leonardo Coelho, José L. Oliveira, 2021)

Nota. ETL pipeline para calcular promedios de biomarcadores.

Principal Resultado:

Los resultados muestran una solución ETL basada en web que replica los componentes disponibles en la API de Kettle, usando una descripción de metadatos, en JSON, y un enfoque reflexivo, usando Java Reflection. De esta manera, la adición de nuevos componentes se puede realizar sin ninguna codificación adicional. El objetivo principal al desarrollarlo fue crear un entorno de canalización ETL colaborativo y fácil de usar, sin perder las características bien establecidas que ya están disponibles en las herramientas de escritorio. El producto final ofrece un espacio de trabajo fácil de usar para construir y ejecutar los procesos ETL, sin necesidad de realizar una nueva implementación del sistema.

4.14 Artículo de Investigación 14

Título: Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores

Aporte:

La gestión de datos es uno de los procesos importantes realizados en XYZ Store para obtener información de ventas de productos. Como parte de sus operaciones, XYZ Store utiliza la aplicación Smile Invent para administrar los datos relacionados con las transacciones de venta de productos. Sin embargo, esta aplicación no puede ayudar a los gerentes a generar los informes requeridos. Por lo tanto, una forma de superar este problema es implementar una aplicación de Business Intelligence (BI) en la tienda XYZ utilizando vistas de panel interactivo. En la implementación de la aplicación BI, la hoja de ruta de BI es la base para realizar investigaciones basadas en la identificación de los problemas que se seleccionarán. Luego, la fase de planificación se lleva a cabo mediante la evaluación de la infraestructura y los proyectos de planificación. La fase de análisis luego se enfoca en realizar un análisis detallado de los problemas comerciales y las oportunidades de implementación de BI. A esto le sigue la fase de diseño mediante la realización del proceso de diseño de almacenamiento de datos y ETL utilizando Pentaho Data Integration (PDI). Luego se realiza la fase de implementación, es decir, la selección y el uso de herramientas de aplicación de BI para realizar la visualización de datos. Se espera que esta investigación pueda generar informes en forma de visualizaciones de paneles interactivos que los gerentes de las tiendas puedan usar para tomar mejores decisiones.

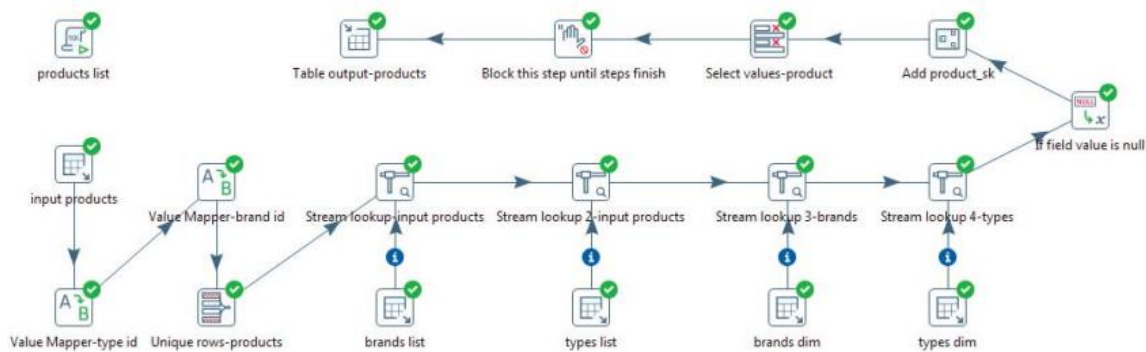


Ilustración 22: Diseño y ejecución de las dimensiones del producto ETL

Fuente: (Ricky Akbar, Meza Silvana, 2020)

Proceso:

La metodología utilizada para implementar la aplicación de BI se ha adoptado desde el ciclo de vida del proyecto de BI. Las etapas que comprende el ciclo de vida de un proyecto de BI son:

- 1). La fase de justificación es analizar y evaluar el caso de negocio, este es el primer paso a considerar para que los usuarios desarrollan BI,
- 2). La fase de planificación incluye dos actividades principales, que es la evaluación de la infraestructura empresarial y la planificación del proyecto,
- 3). La fase de análisis empresarial se enfoca en realizar la definición de requisitos del proyecto, el análisis de datos, la creación de prototipos de aplicaciones y el análisis del almacén de metadatos,
- 4). La fase de diseño, en esta fase se encuentra el diseño de la base de datos, el diseño ETL y el diseño del almacén de metadatos,
- 5). Construcción, los pasos que se toman en esta fase son el desarrollo de ETL, el desarrollo de aplicaciones, la extracción de datos y el desarrollo del almacén de metadatos,
- 6). En la fase de implementación, esta fase es la evaluación de la implementación y la versión. Estas etapas deben abordarse en orden.

Principal resultado:

El principal resultado es crear una aplicación de Business Intelligence (BI) para la gestión de datos de ventas mediante la visualización de paneles interactivos realizando primero el diseño de BI, realizando lo siguiente:

- 1). Diseño arquitectónico del almacén de datos, incluidos los diseños lógicos y físicos.
- 2). El proceso ETL utilizo el motor Pentaho Data Integration (PDI).

Una vez esté se encuentre completo, el siguiente paso se podría visualizar los datos usando la herramienta Public Tableau como un tablero del sistema y una pantalla gráfica. Basado en el caso de esta tienda XYZ, se cuenta con 1 panel de control para ventas y 11 paneles de visualización gráfica dependiendo de las necesidades del negocio.

4.15 Artículo de Investigación 15

Título: Big Data technologies: A survey

Aporte:

Exhiben una capacidad de respuesta lenta y falta de escalabilidad, rendimiento y precisión. Se ha trabajado mucho para abordar los complejos desafíos de Big Data. Como resultado, se han desarrollado muchos tipos de distribuciones y tecnologías. Este artículo es una revisión analítica de las tecnologías recientes desarrolladas para Big Data. Su objetivo es ayudarlo a seleccionar y aplicar la combinación correcta de diferentes tecnologías de big data en función de sus necesidades tecnológicas y requisitos de aplicaciones específicas. No solo proporciona una descripción general de las principales tecnologías de Big Data, sino que también compara las diferentes capas del sistema, como la capa de almacenamiento de datos, la capa de procesamiento de datos, la consulta de datos de la capa de acceso, la capa de acceso a los datos y la capa de acceso a los datos de gestión Clasificar y analizar las principales características, ventajas, limitaciones y usos de la tecnología.

Proceso:

Como resultado de varios proyectos de Big Data en todo el mundo, se han creado muchos nuevos modelos, marcos y tecnologías de Big Data para proporcionar más almacenamiento, procesamiento paralelo y análisis en tiempo real de diferentes fuentes heterogéneas. Además, se han desarrollado nuevas soluciones para garantizar la privacidad y seguridad de los datos. En comparación con las tecnologías tradicionales,

estas soluciones ofrecen mayor flexibilidad, escalabilidad y rendimiento. Además, el costo de la mayoría de las soluciones de hardware de procesamiento y almacenamiento está disminuyendo continuamente debido a los avances tecnológicos sostenidos (Purcell, 2013). Para extraer conocimiento del big data se han diseñado y propuesto diversos modelos, programas, software, hardware y tecnologías. Se esfuerzan por garantizar resultados más precisos y confiables para las aplicaciones de Big Data. Sin embargo, en un entorno de este tipo, puede llevar mucho tiempo y ser difícil elegir entre varias tecnologías.

Principal Resultado:

Las plataformas de big data actuales se basan en varias herramientas de análisis y procesamiento, así como en la visualización dinámica. Estas plataformas permiten extraer conocimiento y valor de un entorno dinámico complejo. También apoyan la toma de decisiones a través de recomendaciones y detectan automáticamente anomalías, comportamientos inusuales o nuevas tendencias. En este artículo, estudiamos las características de Big Data y discutimos en profundidad los desafíos que plantean los sistemas informáticos de Big Data. Además, hemos explicado el valor de Big Data Mining en diferentes campos. Adicionalmente, nos enfocamos en los componentes y tecnologías utilizadas en cada capa de la plataforma Big Data. También se comparan diferentes tecnologías y distribuciones en términos de capacidades, ventajas y limitaciones. También hemos clasificado los sistemas de big data en función de su funcionalidad y los servicios prestados a los usuarios finales.

4.16 Artículo de Investigación 16

Título: A Virtual Community of Practice Proposal for Business Intelligence Researchers

Aporte:

El aporte del autor es brindar una solución en la Comunidad Virtual de Práctica (VCoP) para que proporcione una importante herramienta de gestión del conocimiento, ya que esta se basa en objetivos compartidos e intereses comunes durante un largo período de tiempo, pudiendo desarrollar capital social, creando nuevos conocimientos, explotando el conocimiento tácito, estimular la innovación y difundir los resultados. Una organización solo puede ser un gran competidor en un mercado cambiante optimizando su desempeño, aprovechando continuamente las oportunidades emergentes, asumiendo riesgos y siendo flexible frente a las nuevas necesidades.

Las contribuciones de este estudio es proporcionar un método de comparación de árboles como un esquema de fusión para manejar diferentes tipos de información de contenido: atributos de taxonomía con estructura de árbol y etiquetas de folksonomía de forma libre; y diseñar un modelo único de paseo aleatorio en un gráfico heterogéneo con diferentes tipos de nodos y aristas. Al igual que los otros modelos de recomendación híbridos, el autor tiene como objetivo utilizar información heterogénea para mejorar la cobertura de recomendación manteniendo una alta precisión.

Proceso:

El autor propone una comunidad virtual de inteligencia de negocios mediante un modelo de práctica para su Universidad, capaz de apoyar a los profesionales compartiendo recursos: experiencias, problemas y soluciones, herramientas y Metodología. Esta comunidad facilitará el enriquecimiento de conocimiento de cada participante y aumentará su aprendizaje en el campo de la inteligencia de negocios y les permitirá realizar investigaciones originales con la ayuda del apoyo de los participantes. El autor cree que mediante un espacio de comunicación virtual debidamente construido y administrado será de gran ayuda para los profesores, estudiantes y aquellos interesados en las herramientas de BI. Esta iniciativa se enmarca en una tendencia más amplia de aprovechar los mecanismos favorables que brinda la sociedad del conocimiento, asumiendo un enfoque pragmático y emprendedor en la implementación de las funciones de aprendizaje, las técnicas de la universidad le permiten apalancar su conocimiento.

Principal Resultado:

El autor en base a lo descrito encuentra un gran potencial en el desarrollo de una comunidad virtual de habilidades de BI a través de listas de discusión por correo electrónico y foros de discusión, así como los desafíos de este enfoque. Tarea de transferencia de conocimiento asignada a los miembros de la comunidad; El medio ambiente no puede participar en este proceso. Este objetivo se puede lograr introduciendo el concepto de Web Semántica. Propone la existencia de conjuntos de datos estructurados y reglas de inferencia que conducirán a la inferencia automática.

Por este motivo, el autor considera esto como una investigación futura, ya que en artículo se han dado pasos desde el aprendizaje mediante la prueba del conocimiento hasta la construcción activa del conocimiento, el desarrollo del pensamiento y el estímulo de la innovación. El desarrollo de la sociedad de la información ha dado lugar al nacimiento de la sociedad del conocimiento, un paso intermedio en el camino hacia la formación de una sociedad consciente.

4.17 Artículo de Investigación 17

Título: Amazon Redshift and the Case for Simpler Data Warehouses

Aporte: Amazon Redshift es un sistema de datos a escala de petabytes rápido y completamente administrado. Warehouse Solution permite analizar de manera eficiente grandes volúmenes de datos utilizando las herramientas de inteligencia comercial existentes. Desde su lanzamiento en febrero de 2013, ha habido Amazon Web Services (AWS) que ha tenido el crecimiento más rápido, con varios miles de clientes y varios petabytes de datos bajo administración de La tasa de adopción de Amazon Redshift sorprendió a muchos participantes en la comunidad de almacenamiento de datos de aunque Amazon Redshift tenía un precio alto en el lanzamiento, con un costo de solo \$1000/TB/año, existen muchas tecnologías de almacenamiento de datos de código abierto y muchas herramientas comerciales de almacenamiento de datos que brindan una versión gratuita para el desarrollo o bajo un cierto uso limitado. Si bien Amazon Redshift proporciona una arquitectura MPP moderna, en columnas y escalable, también lo hacen muchos otros motores de almacenamiento de datos. almacenamiento local o en red Más sobre este texto fuente Para obtener más información sobre las traducciones, se requiere.

Proceso: Finalmente, debido a que administramos un clúster de miles de clústeres de Amazon Redshift, nos beneficiamos de una automatización que no tendría sentido desde el punto de vista económico para un DBA local individual. Una operación que requiere un minuto de costo por nodo vale varios días de inversión para eliminarla. La eliminación de la causa raíz de la falla que hace que un clúster individual falle y se reinicie una vez cada mil días bien vale nuestra inversión para reducir la paginación. A medida que crecemos, debemos mejorar continuamente la confiabilidad, la disponibilidad y la automatización de nuestra flota para administrar las operaciones, lo que hace que las interrupciones sean menos comunes para nuestros clientes. En este artículo, analizamos la arquitectura del sistema Amazon Redshift, la simplicidad como

principio rector y las lecciones que aprendimos durante los dos años de funcionamiento del servicio.

4.18 Artículo de Investigación 18

Título: Business Analytics using Microsoft Power BI and AWS Redshift

Aporte: Un almacén de datos es un repositorio que extrae datos de varias fuentes dentro de una organización para Business Reports y Business Analytics. Los informes generados a partir de consultas complejas en un almacén de datos se utilizan para mejorar los procesos de la empresa, realizar juicios óptimos e incluso anunciar posibles beneficios. A medida que los datos organizacionales crecen exponencialmente día a día, el almacén de datos interno se vuelve obsoleto y surge un nuevo concepto de almacén de datos basado en la nube. El jugador principal en este almacén de datos basado en la nube es AWS Redshift. Después de recopilar datos del almacén de datos, el siguiente paso es recuperar algunos datos analíticos y usarlos para generar información comercial y luego aumentar los ingresos y el rendimiento comercial. El análisis comercial anterior se puede realizar utilizando una variedad de herramientas. El líder del mercado a este respecto es Microsoft Power BI. En este artículo intentaremos analizar cómo funciona Microsoft Power BI en colaboración con AWS Redshift.

Proceso: AWS Redshift es una base de datos popular para Power-BI. Tanto Redshift como Power BI son líderes del mercado en sus respectivas áreas de análisis y almacenamiento de datos. Estamos tratando de reunir a los dos líderes en una plataforma para que los usuarios puedan aprovecharlos al máximo. Los pasos complejos para conectar y configurar Redshift y Power BI se han detallado en este documento.

Asimismo, Redshift cuando tenga varias bases de datos (al menos 5 o más) con un almacenamiento total de terabytes, entonces la relación calidad-precio es real. Entonces, si la seguridad es su principal preocupación, puede elegir tener su propia solución de almacenamiento de datos interna en lugar de buscar soluciones basadas en la nube. Del mismo modo, si no quiere que le molesten las interdependencias entre plataformas, puede usar AWS Quick Sight con AWS Redshift y, de manera similar, puede usar Microsoft SQL Datawarehouse con Microsoft Power BI.

4.19 Artículo de Investigación 19

Título: An Integrated Model of Business Intelligence & Analytics Capabilities and Organizational Performance

Aporte:

Un modelo de investigación para resaltar cómo las cuatro capacidades de BI&A influyen en el desempeño organizacional. adoptamos una vista jerárquica e integrada de las capacidades de B&A porque BI&A comprende tanto tecnología como componentes organizacionales y, por lo tanto, las organizaciones necesitan desarrollar, coordinar e integrar BI&A capacidades en diferentes niveles, de modo que proporcionen soluciones para todo el negocio y generen valor para partes interesadas

Proceso:

El proceso de la propuesta de los autores es la siguiente:

Operacionalización de Construcciones

Los autores hicieron una validación cruzada estos elementos iniciales del instrumento a través de una prueba previa con siete académicos y encuestados de la industria (senior gerentes en funciones de TI, marketing y operaciones involucradas en la implementación y uso de BI&A).

Desarrollo de marco de muestra

En segundo lugar, desarrollamos nuestro marco de muestra recopilando directorios de múltiples industrias, ciudades y estados del estado más grande (por producto interno bruto) en la India, del cual obtuvimos una lista inicial de aproximadamente 1500 empresas de los cuales solo pudieron rescatarse 790 empresas porque no constituían organizaciones de TI maduras con implementaciones de BI&A.

Procedimiento de recogida de datos

En tercer lugar, el autor administró el cuestionario tanto en línea como en persona a un encuestado de cada empresa (el ejecutivo de TI de más alto rango en la empresa con designaciones como director de información, director oficial de tecnología, director de TI y gerente de TI) en nuestra muestra. Los usuarios participaron voluntariamente y de forma anónima recopilando datos utilizables de 154 encuestados que utilizan BI&A en sus organizaciones en todo varias industrias (para una tasa de respuesta del 19,5%).

Técnicas de Análisis

El autor elige PLS como técnica de análisis principal, ya que permite estimar una serie

de factores interrelacionados. Además, PLS maneja mejor las construcciones formativas de varios niveles en comparación con SEM basado en covarianza (Lowry & Gaskin, 2014), y nuestro modelo contiene niveles múltiples construcciones formativas. PLS también puede manejar tamaños de muestra más pequeños. Utilizamos SmartPLS 3.0 para realizar el análisis.

Principal Resultado:

El autor proporciona una visión holística de la integración de BI&A en una organización, aunque con un sesgo hacia la perspectiva de la capacidad. Al tomar la perspectiva de las capacidades, puede resaltar el hecho de que BI&A no solo representa una moda pasajera o un rumor en la práctica y el discurso académico, sino que puede ayudar a obtener un mayor rendimiento organizacional. En general, adoptamos una perspectiva integradora enfoque de BI&A. Nos enfocamos en cuatro dimensiones de la capacidad de BI&A y las relacionamos con la organización actuación. Utilizamos la metodología de la encuesta para recopilar datos. Nuestros hallazgos brindan un marco integrador de capacidades para la implementación de BI&A y motivan a los gerentes a ver BI&A desde una perspectiva perspectiva de mejora del desempeño organizacional

4.20 Artículo de Investigación 20

Título: Using business intelligence solutions for forecasting in marketing researches

Aporte: Un almacén de datos es un repositorio que extrae datos de varias fuentes dentro de una organización para Business Reports y Business Analytics. Los informes generados a partir de consultas complejas en un almacén de datos se utilizan para mejorar los procesos de la empresa, realizar juicios óptimos e incluso anunciar posibles beneficios. A medida que los datos organizacionales crecen exponencialmente día a día, el almacén de datos interno se vuelve obsoleto y surge un nuevo concepto de almacén de datos basado en la nube. El jugador principal en este almacén de datos basado en la nube es AWS Redshift. Después de recopilar datos del almacén de datos, el siguiente paso es recuperar algunos datos analíticos y usarlos para generar información comercial y luego aumentar los ingresos y el rendimiento comercial. El análisis comercial anterior se puede realizar utilizando una variedad de herramientas. El líder del mercado a este respecto es Microsoft Power BI. En este artículo intentaremos analizar cómo funciona Microsoft Power BI en colaboración con AWS Redshift.

Proceso: La inteligencia empresarial se compone de herramientas y aplicaciones que aprovechan el software y los servicios para traducir datos en acciones inteligentes para decisiones estratégicas, tácticas y operativas. Soluciones de inteligencia empresarial que facilitan y amplían las ofertas de servicios a los investigadores de mercado, ahorrando el tiempo y el esfuerzo necesarios para identificar clientes, pronosticar la demanda y gestionar la producción. Más eficiente, capacidad de descubrir oportunidades para aumentar los ingresos. El propósito de este artículo es utilizar soluciones de inteligencia empresarial para realizar pronósticos en la investigación de mercado. Las soluciones inteligentes ayudan a los investigadores de mercado a lograr eficiencia, eficacia y diferenciación.

4.21 Conclusiones

En base al análisis de los artículos de investigación nos pudimos dar cuenta las importancias de poder aplicar la inteligencia de negocios en las empresas trayendo consigo grandes beneficios, ya que en primer instancia te permite poder reunirte con el negocio para poder realizar un análisis en los procesos de la empresa y ver los principales problemas. En base a ello, se escoge el proceso más crítico y se realiza una lista de requisitos, conocer las necesidades del negocio y en base a eso plantear una solución.

Asimismo, los autores proponen evaluar la infraestructura de la compañía y en base a eso poder empezar la fase de planificación y la evaluación de las herramientas. Además, si todo se realiza de la manera correcta de siempre involucrar al negocio y validar en conjunto que se cumplan los requisitos del proyecto, la creación de prototipos y el almacén de metadatos le permitirá a las empresas una mejor toma de decisiones, ya que estos se encontraran soportado en base técnicas y metodologías.

Por último, recomiendan emplear unificar las herramientas de BI en un entorno Cloud ya que generalmente son innovadoras y predominantes en el mercado y trae beneficios a largo plazo, ya que les permite a las empresas poder beneficiarse en tanto a costo como escalabilidad. En base a lo descrito, nos podemos dar cuenta que con ello las empresas no invertirán en servidores físicos porque en la actualidad esto conlleva costos tanto a nivel de seguridad como disponibilidad. Por este motivo, recomienda en utilizar para el almacenamiento del data warehouse Amazon Redshift y para la grafico de sus reportes podrían optar entre Tableau o Power BI permitiendo así poder aprovechar al máximo el análisis de los datos de una empresa.

5. DESARROLLO DEL PROYECTO

En el presente capítulo se mostrará los procedimientos seguidos para poder concluir de manera exitosa el desarrollo de nuestro proyecto en su totalidad. Asimismo, se comparten gráficos y tablas para mostrar lo desarrollado a lo largo del proyecto.

5.1 Análisis

5.1.1 Situación actual del turismo

Según Promperu las agencias de viajes y operadores turísticos se estima que para este año 2022 las agencias de viajes crezcan un 24% debido al programa de reactivación gradual del turismo interno buscaba incentivar a las personas que puedan viajar reactivando la económica en cada departamento.



Fuente: Promperu, SBS

Ilustración 23: Datos Estadísticos de la recuperación de ventas turistas

En el siguiente gráfico, se puede observar que la reactivación del turismo irá en aumento entre este año y el próximo a comparación del 2021 cuando estábamos en época de pandemia pudiendo las empresas del sector de transporte poder recuperar su capitalización en el mercado. Esto trae consigo que las empresas puedan invertir en nuevas tecnologías que les permita poder facilitar su trabajo.

Por otro lado, Gartner indica que, en el 2020, más del 40% de las tareas de análisis de datos serán realizadas por hardware. Además, para 2022, el 30% de las interacciones con los clientes se verán influidas por el análisis de datos, como la ubicación, los intereses y el comportamiento real

5.1.2 Proceso del Sector Transporte

En las siguientes imágenes podemos ver el proceso de la venta de pasajes de la empresa del sector transporte que se enfoca desde que el cliente compra un boleto hasta que se efectúa la venta de cada uno

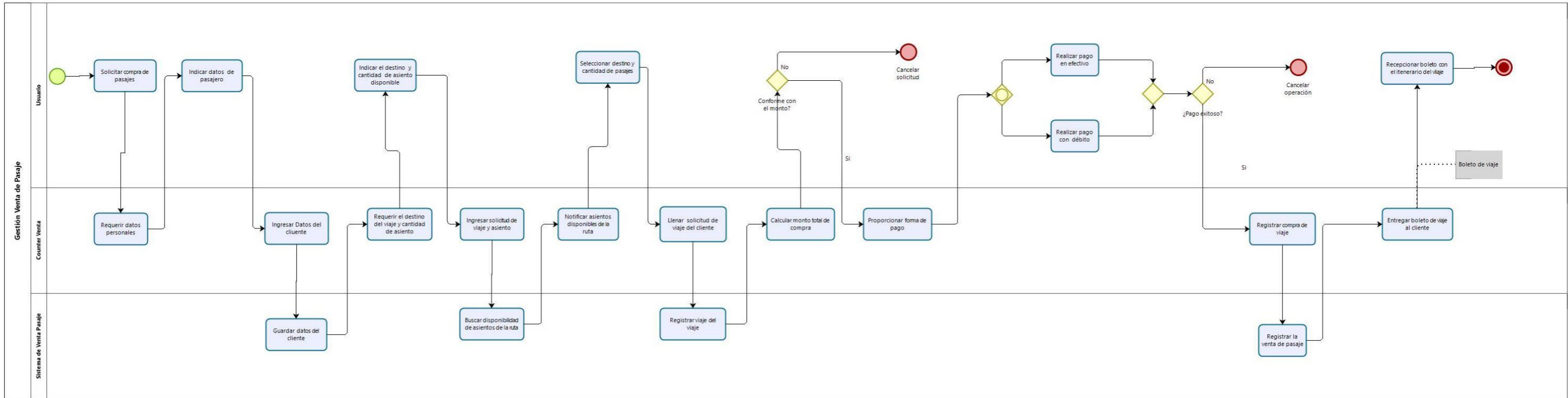


Ilustración 24: Proceso de Venta de Pasajes

Fuente: Elaboración Propia

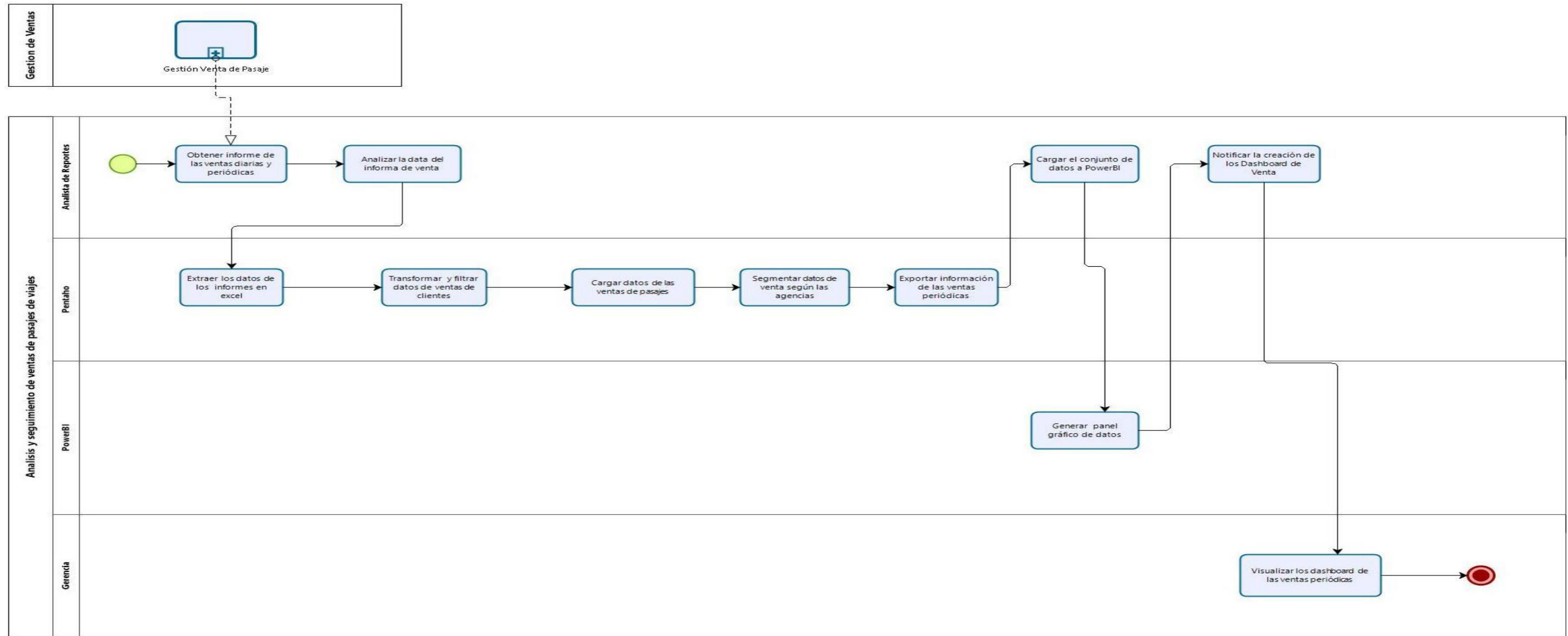


Ilustración 25: Proceso de Seguimiento de la venta de pasajes

Fuente: Elaboración Propia

5.1.3 *Análisis de la metodología de inteligencia de negocios*

Metodología Kimball

La metodología Kimball es una metodología que se usa para la construcción de un almacén de datos en otras palabras data warehouse, la cual consiste en un grupo de datos enfocados a una empresa u organización que sirve para la toma de decisiones en la entidad donde se aplique.

Asimismo, la metodología está basado en los siguientes cuatro principios:

Enfocarse en el negocio

Construir una infraestructura de información.

Realizar entregas periódicas que consiste en poder ir creando data warehouse en plazos entre 6 a 12 meses en este caso tiene cierta similitud a las metodologías ágiles

Ofrece una solución completa esto quiere decir que la metodología te ofrece todos los elementos necesarios que permite entregar valor al negocio, para ello deberá contar con un datawarehouse, herramientas de consulta, aplicaciones para informes y análisis avanzado.

Principales ventajas.

El desarrollo del modelado dimensional de Kimball es menor a otras metodologías debido a que es rápido de construir ya que no implica normalización.

No requiere un equipo con expertise alto con uno pequeño de diseñadores y planificadores porque al no contar con normalización en el modelado la optimización de las consultas es sencilla, predecible, rápida y controlable.

Solo se enfoca en las principales áreas estratégicas del negocio permitiendo poder realizar consultas más rápidas y pudiendo obtener reportes en un menor tiempo.

El costo que implica usar esta metodología es menor porque no requiere un gran equipo para el desarrollo y el análisis de datos, ya que solo se enfoca en las áreas de manera individual.

Principales desventajas.

No se puede manejar el informe de todas las áreas porque el modelado está enfocado en los procesos comerciales.

El proceso de ingresar grandes cantidades de datos heredados en el datawarehouse es complejo

Como el modelo de Kimball está orientado a los procesos comerciales, en lugar de centrarse en la empresa en su conjunto, no puede manejar todos los requisitos de informes de BI.

Fases

Planificación del proyecto - Esta fase se encarga de identificar la definición y alcance del proyecto de Datawarehouse, conocer las necesidades del negocio y evaluar la factibilidad sobre los recursos, roles, funciones, tiempo de duración y secuencialidad.

Definición de los requerimientos del negocio - Los diseñadores tienen que identificar los requerimientos y transformarlo en un diseño apropiado, ya que son la base para las tres etapas paralelas dando énfasis en la tecnología, los datos y las aplicaciones.

Modelado dimensional - Es una técnica del diseño de la base de datos enfocado en apoyar a las consultas de los usuarios finales en un datawarehouse donde se puede determinar la dimensionalidad de cada indicador, básicamente se enfoca en diferentes tipos de grados de detalle, concepto de negocio y la granularidad de cada indicador.

Diseño Físico - Se enfoca en la selección de estructuras que son esenciales para el diseño lógico. Principalmente en este proceso se definen los estándares de nombres, especificación de ambientes de base de datos y la indexación y estrategias de particionamiento en determinadas etapas.

Diseño y desarrollo de presentación de datos - Esta fase se enfoca en realizar ETL (Extracción, Transformación y Carga).

Diseño de Arquitectura Técnica - Para esta fase se tiene que tener en cuenta los ambientes del datawarehouse como: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas con ello se puede establecer el diseño de la arquitectura técnica del ambiente del datawarehouse.

Selección de Productos e Instalación - Se utiliza el diseño de la arquitectura técnica de la fase anterior como marco para después evaluar y

seleccionar los componentes, el motor de base de datos, la herramienta de ETL, etc.

Especificación de Aplicaciones para Usuarios Finales - Los distintos tipos de roles o perfiles determinan las interfases del datawarehouse. En esta etapa se diseñan los reportes o consultas avanzadas para los analistas, dashboard para la alta gerencia.

Desarrollo de Aplicaciones para Usuarios Finales - En base a las especificaciones de las aplicaciones de los usuarios finales, el desarrollo involucra realizar configuraciones al metadata y la creación de los reportes acorde a la necesidad del negocio.

Implementación - La implementación cuenta con la unificación de la tecnología, los datos y las aplicaciones de usuarios finales para que puedan acceder. En esta fase podemos encontrar factores extras que permiten asegurar el funcionamiento del modelado.

Metodología Inmon

La metodología Inmon utiliza la forma normalizada para la construcción de la estructura de entidad, relación evitando la redundancia de datos. Esto conlleva contar con una identificación clara de los requisitos comerciales.

Asimismo, esta metodología de warehouse propone la construcción de data marts separados de cada área. Por ello, todos los datos ingresados al datawarehouse están integrados entre sí.

Principales ventajas.

- El desarrollo del modelado dimensional de Kimball es menor a otras metodologías debido a que es rápido de construir ya que no implica normalización.
- Datawarehouse actúa como una fuente de verdad unificada para el negocio, porque al contar con el modelado de varias áreas le permite estar integrado.
- Cuenta con una redundancia de datos muy baja por la integración de sus datos por lo que hay menos probabilidades de irregularidades en la actualización de datos.

Principales desventajas.

La complejidad del modelado aumenta cuando comienzan a agregar más tablas de datos.

Cuenta con una redundancia de datos muy baja por la integración de sus datos por lo que hay menos probabilidades de irregularidades en la actualización de datos.

- Requiere un equipo con un expertise alto para poder modelar los datos de todas las áreas, realizar el desarrollo y análisis lo cual conlleva un mayor presupuesto.

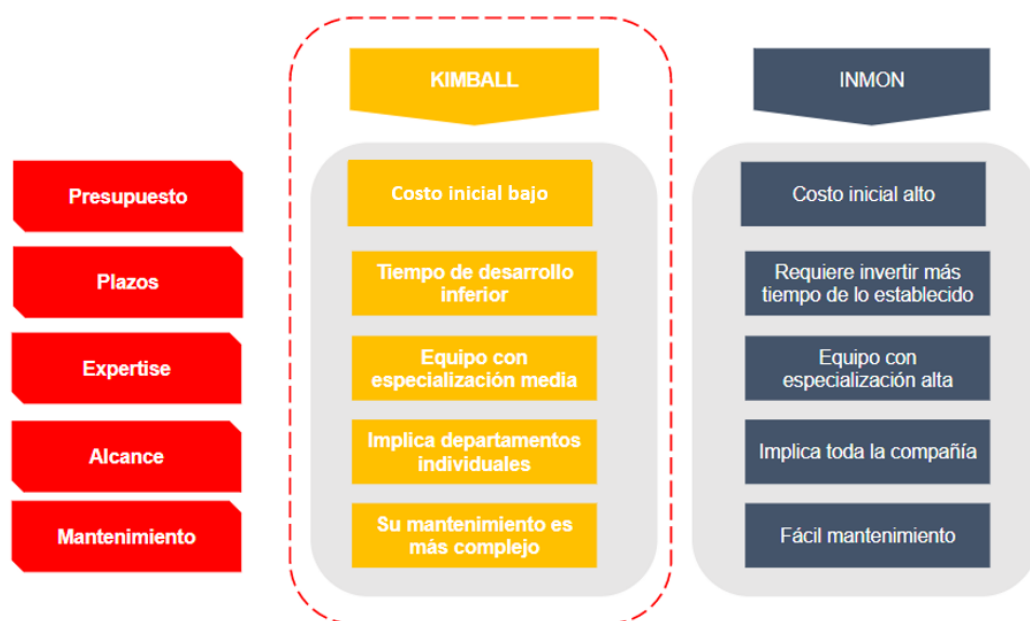


Ilustración 26: Evaluación de las metodologías de BI

Fuente: Elaboración propia

En base a lo descrito por ambas metodologías, para esta solución se emplea la metodología de Kimball porque nosotros nos centraremos en el proceso de ventas y la alta gerencia buscará mejorar sus informes para la toma de decisiones, lo que les permitirá enfocarse más al ámbito estratégico que al operativo.

Metodología Data Vault 2.0

Data Vault es un método innovador de modelado de datos para plataformas de almacenamiento de datos a gran escala. Inventado por Dan Linstedt, Data Vault está diseñado para proporcionar almacenamiento de datos empresariales mientras se superan las limitaciones de las técnicas de normalización y modelado dimensional (tercera forma normal). Combina el

almacén centralizado de datos sin procesar del método Inmon con las ventajas de construcción incremental de Kimball.

Ventajas

Distribución incremental: si bien tiene sentido crear un almacén de datos en el contexto de un modelo comercial general, Data Vault admite la distribución **incremental** completa. Al igual que el enfoque de diseño dimensional de Kimball, puede comenzar poco a poco y agregar fuentes adicionales con el tiempo.

- **Flexibilidad:** a diferencia del enfoque convencional de modelado de formularios de terceros, que puede ser inflexible, Data Vault no requiere reajustes a medida que se agregan fuentes adicionales. Dado que Data Vault almacena datos comerciales sin procesar y derivados por separado, admite fácilmente cambios en las reglas comerciales.
- **Reducción de la complejidad:** dado que la Bóveda de datos se basa en un enfoque de dos pasos, separa la reingeniería de la ingeniería de datos de la aplicación de reglas comerciales, lo que permite la separación de pasos potencialmente complejos. Además, la limpieza de datos se considera una regla comercial y se puede manejar independientemente del intento inicial de carga de datos.
- **Inclusión de datos sin procesar:** guardar datos sin procesar en un almacén de datos significa que el área de visualización se puede recrear con propiedades históricas que no estaban disponibles originalmente. Si el almacén de datos se implementa como vistas, esto puede ser tan simple como agregar una columna adicional a una vista existente.
- **Admite con elegancia el cambio de hora:** Al igual que la dimensión gradual del enfoque de Kimball, Data Vault admite con elegancia el cambio de hora. Sin embargo, a diferencia del diseño dimensional puro, Data Vault separa los datos comerciales sin procesar y los derivados y admite los cambios

que resultan en el sistema de origen y las reglas comerciales.

5.1.4 Análisis de la herramienta

En busca de las herramientas de inteligencia de negocios para aplicar en la solución ha sido necesario poder compararlas entre sí y que cumplan con los requisitos del sector de transporte. Las mismas fueron clasificadas de la siguiente manera:

Tabla 5: Criterios de puntuación

Criterios	Puntuación
No cumple con los requerimientos del proyecto	1-2
Regular	3-4
Alineado al proyecto	5

Fuente: Elaboración propia

En este paso estamos utilizando 5 herramientas de BI como: Power BI, Tableau, Qlikview Sense, Google Data Studio y SAS.

Los criterios de evaluación se realizaron en base a Acceso, Interfaz de Usuarios, Soporte y Documentación, Procesamiento de datos, Conexión de datos, autoservicio, Análisis Colaborativo, versiones ejecutadas y Precio. Entre todos estos criterios se hicieron las puntuaciones entre 1 al 5 ganando Power BI como herramienta de visualización para los reportes de la solución.

Tabla 6: Análisis de las herramientas de visualización BI

Criterios	Funcionalidades	Power BI	Tableau	Qlik View Sense	Google Data Studio	SAS
Acceso	Apertura pública de datos en la web	4	4	1	2	2
Interfaz de Usuario	Visualización de las plantillas de gráficos interactivos de datos	5	3	2	2	3
Soporte y Documentación	Permite visualizar documentación con ejemplo de uso de herramienta.	4	2	4	2	2
Procesamiento de datos	Permite el análisis y transformación de datos	4	4	3	2	2
Conexión de Datos	Almacenamiento en la nube y carga incremental de datos	4	4	4	2	3
Auto Servicio	Realizar consultas, filtros, drill up/down	4	4	3	2	3
Análisis Colaborativo	Visualización en tiempo real por los canales Web, Mobile.	5	5	2	1	3
Versiones Ejecutables	Soporta ejecutables para la versión de escritorio y nube	5	3	2	2	2
Precio	Gratuito y diferente versiones de pago	3	3	2	2	1
TOTAL		38	32	23	17	21

Fuente: Elaboración propia

El cuadrante mágico de Gartner podemos ver que en base de los 4 cuadrantes una de las mejores herramientas encontramos a Microsoft, Salesforce y Qlik. En base a la tabla de criterios mencionadas anteriormente podemos escoger Microsoft Power BI no solo por sus diversas funcionalidades, sino que al estar en el cuadrante de Gartner cuenta con un respaldo para poder aplicarlo en la solución.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Ilustración 27: Cuadrante de Gartner

Fuente: Gartner 2022

Asimismo, se realizó un cuadro comparativo entre las herramientas de ETL entre estas tenemos Apache NiFi, AWS Glue, Talend, Pentaho y AB Initio las cuales se clasificaron bajo los siguientes criterios:

Tabla 7: Análisis de las herramientas Integración BI

Criterios	Funcionalidades	Apache NiFi	AWS Glue	Talend	Pentaho	Ab Initio
Flexibilidad	Preprocesamiento de los datos	4	3	4	5	4
Usabilidad	Fácil manejo para el usuario	3	4	4	4	5
Transformaciones	Permite el análisis y transformación de datos	4	3	5	4	4
Fuente de Datos	Soporta grandes cantidades de datos	3	5	4	4	5
Sencillez	Fácil de instalar	4	4	4	4	4
Precio	Gratuito y diferente versiones de pago	3	3	2	5	2
TOTAL		21	22	23	26	22

Fuente: Elaboración propia

Para esta solución usaremos Pentaho porque al ser una herramienta con mayor flexibilidad y a comparación de otras ser una herramienta open source cumple con los requisitos para usarse con nuestra solución.

5.2 Diseño de la solución

En la elaboración del modelo propuesto, como primer paso se realizó el análisis y relevamiento de información necesaria para poder entender el proceso core de la organización como es de entender se debe mapear los procesos que involucra la solución. Por otro lado, se concertó reuniones de **entendimiento** junto con la gerencia de la empresa del sector de transporte interprovincial. Paso seguido, se realizó una arquitectura lógica que permita dar un mejor entendimiento de la solución de BI en su capa de negocio, la cual nos permitirá entender los objetivos generales de la organización y los principales procesos para su solución; la capa de aplicación, donde se define las herramientas con las que se va a soportar la solución y sus respectivos servicios asociados; por último la capa de tecnología, en la que se especifican los componentes necesarios a nivel de infraestructura y definición de Sistemas operativos, gestor de bases de datos, esquemas, entre otros. También se definió las salidas esperadas por parte de la solución, donde se puede identificar los resultados y las ventajas de la implementación de dicha solución. A continuación, se irá detallando punto por punto cada paso realizado para el desarrollo del presente proyecto. En el siguiente gráfico, se busca resumir de manera sencilla el diseño de la solución y sus componentes.

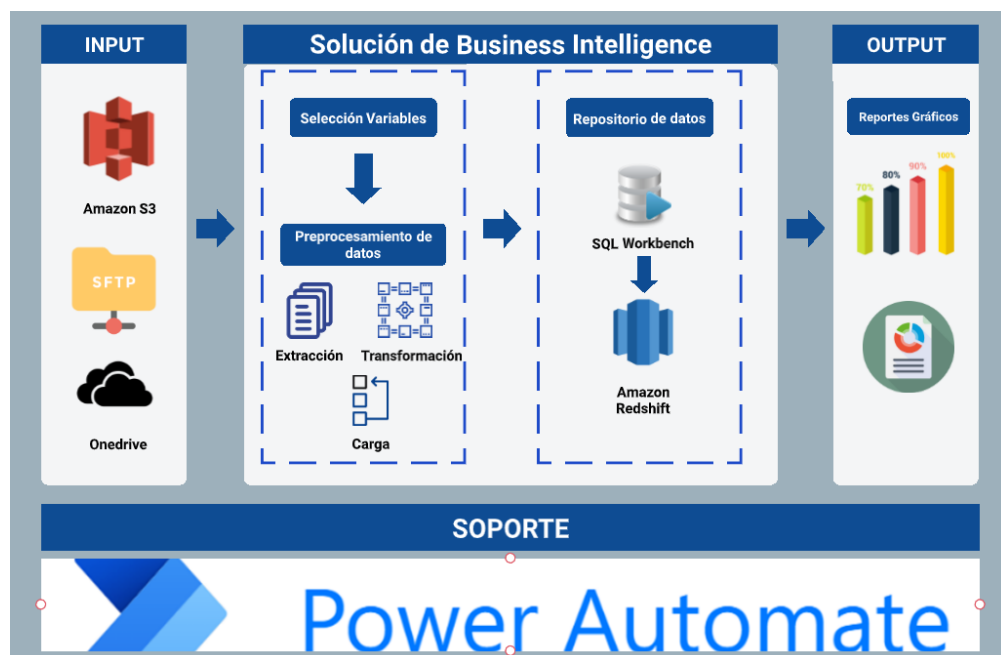


Ilustración 28: Solución propuesta

Fuente: Elaboración propia

5.2.1 Entradas

En primer lugar, se define las entradas de la solución de BI, por lo que se detalla a continuación:

Registro de Ventas de los pasajes terrestres diarias, quincenal, mensual.

Registro de Reservas realizada de los clientes de los pasajes.

Registro de Sucursal de ventas de pasajes terrestres

Los mismos estarán almacenados en Onedrive, los cual ingresaría al servicio SFTP, para que después sea almacenado en Amazon AWS S3 y pueda ser enviado a la herramienta Pentaho el cual será ejecutado con un proceso batchero.

5.2.2 Proceso

Cuando la información sea ingresada a Pentaho, la herramienta realizara el procesamiento de la información del proceso anterior. En esta etapa, Pentaho se integrará con Amazon AWS Redshift con SQL Workbench para poder almacenar en el data warehouse.

5.2.3 Output

En cuanto a los reportes gráficos estos serán soportados por PowerBI, lo cual le permitira a la empresa poder medir su índice de ventas y en base a esto el negocio analice el comportamiento que ha ido teniendo la empresa mediante segmentación de edades, sucursales y ventas en un determinado tiempo

5.2.4 Arquitectura lógica

En base a la propuesta de solución se analizó la situación actual de la organización. Por ello, como parte de este análisis se determinó la siguiente arquitectura el cual conlleva a la gráfica de la interacción de las capas o dominio según los lineamientos de la Arquitectura Empresarial de TOGAF, a continuación, se menciona dichas capas

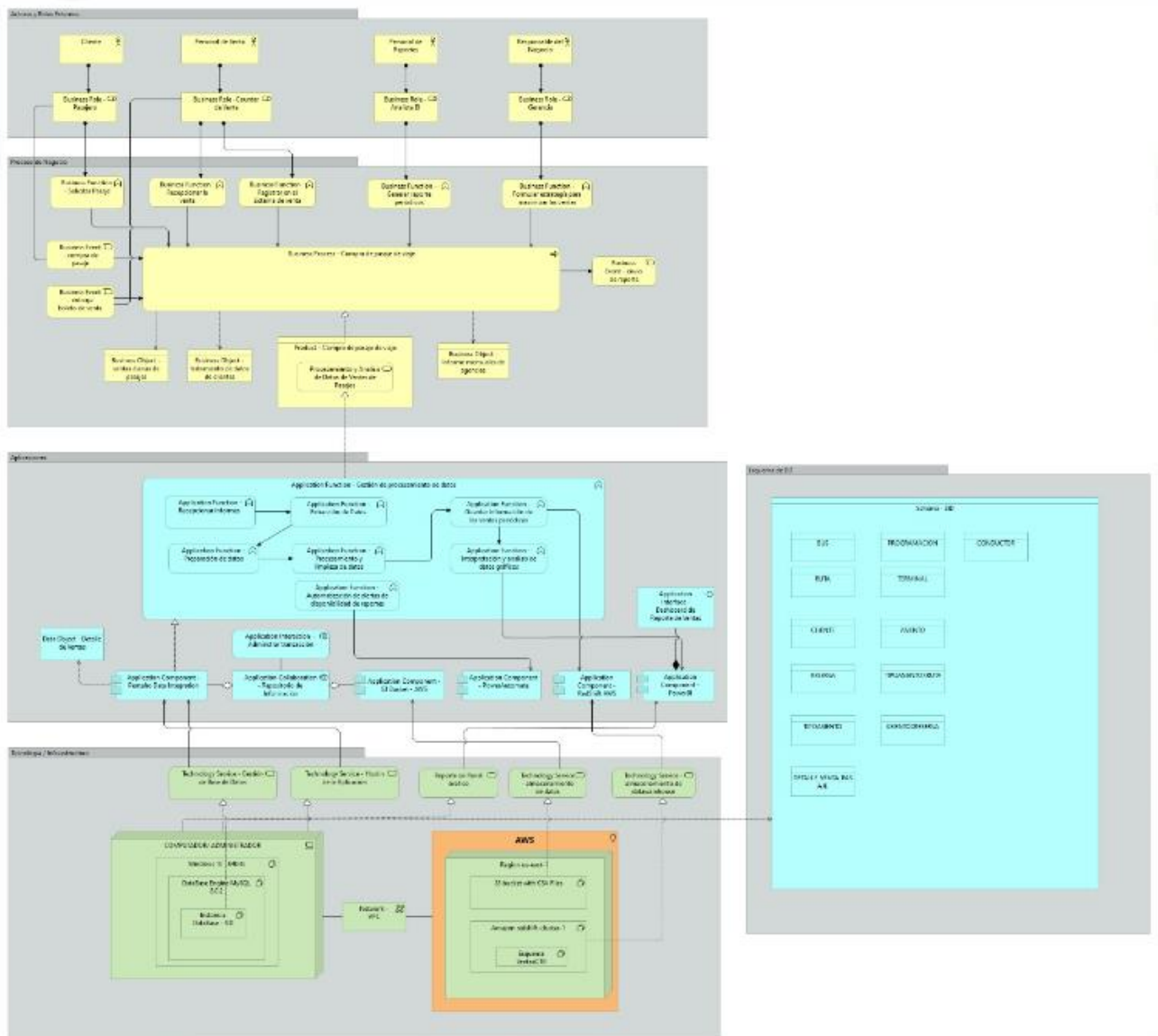


Ilustración 29: Arquitectura de la solución

Fuente: Elaboración propia

5.2.5 Arquitectura Capa Negocios

En esta capa de negocio se busca transmitir de manera gráfica principalmente los procesos, roles, actores, eventos y funciones que serán abordados en la solución propuesta. Dichas funcionalidades deberán estar alineadas a las estrategias que la organización ha definido previamente. Tal como se logra apreciar en la ilustración líneas abajo, se resume los procesos definidos y a su vez, estos son relacionados con los roles y los eventos que hacen que estos inicien.

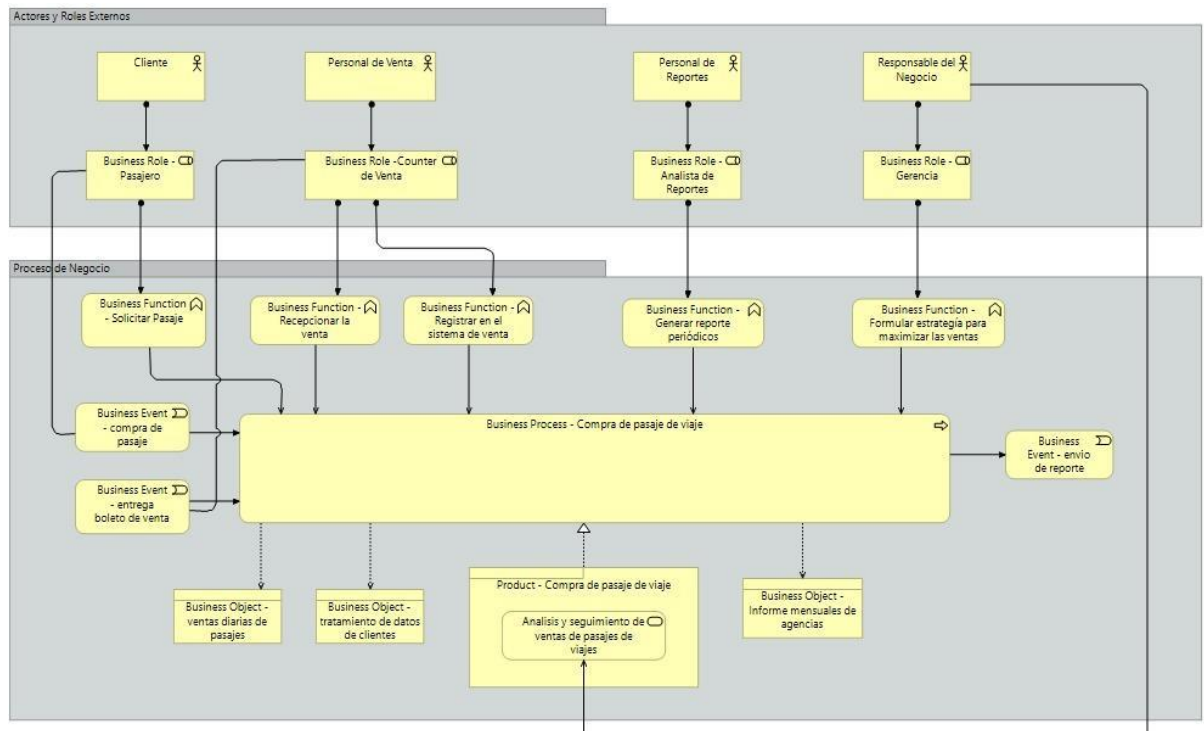


Ilustración 30: Arquitectura de la capa de negocios

Fuente: Elaboración propia

5.2.6 Arquitectura Capa de aplicaciones

En esta capa se describen los servicios de aplicaciones que serán consumidos por los procesos de la capa de negocios mediante software. Para el caso de la solución propuesta, el software soportado será Power Automate para la automatización del Proceso de BI el cual como input hará una integración con SFTP y Amazon S3 para el envío de la información como input, después irá al aplicativo de Pentaho Data Integration (PDI) el cual, al ser una herramienta de BI, permite realizar los ETLs, con ello se estaría accediendo a los servicios de:

- Recepción de datos
- Extracción de datos
- Procesamiento de datos
- Exportación de datos

A su vez, la siguiente aplicación estará integrada con Amazon Redshift y SQL Workbench para el almacenamiento de la información procesada en data warehouse y para la visualización de gráficos interactivos para la presente solución es PowerBI, la cual es una herramienta de visualización que permitirá acceder a los servicios de:

- Visualización de paneles gráficos de datos

- Análisis de conjunto de datos en tiempo real

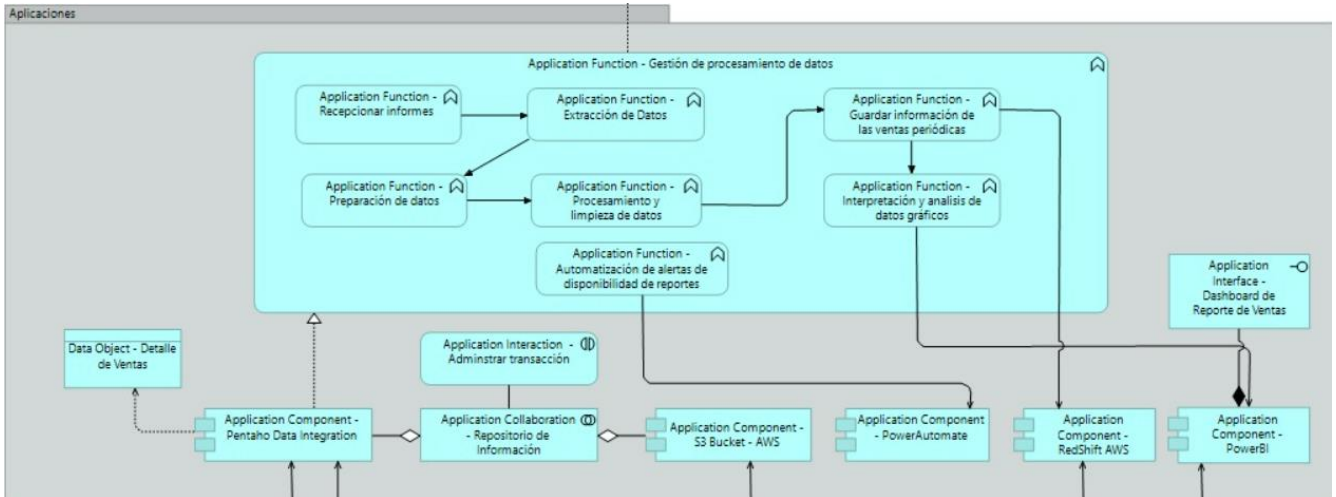


Ilustración 31: Arquitectura de aplicaciones

Fuente: Elaboración propia

5.2.7 Arquitectura Capa de tecnología

En esta sección se define la infraestructura que formará parte del soporte de las aplicaciones de la solución con la finalidad de que estas puedan ser ejecutadas. Para el proyecto, se estará trabajando en la laptop personal del analista de la empresa del sector transporte interprovincial, el cual se debe tener instalado el Sistema Operativo Windows 10 home 64 bits.

Para el desarrollo del proyecto se trabajará con un esquema y bases de datos independientes.

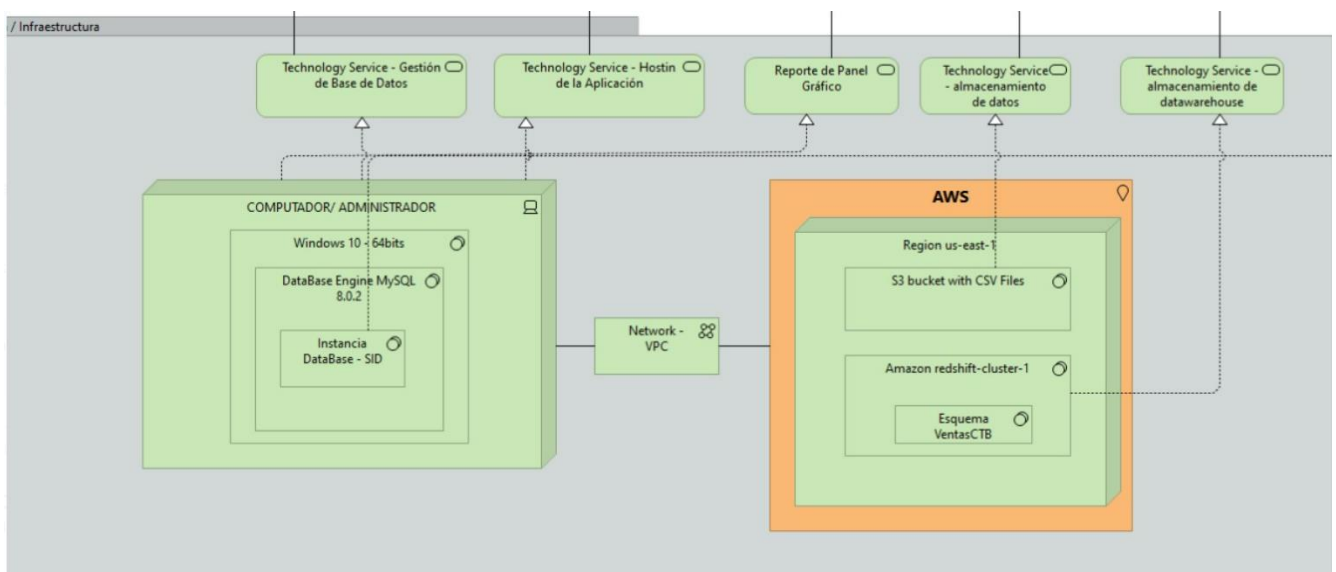


Ilustración 32: Arquitectura de capa tecnológica

Fuente: Elaboración propia

5.3 Construcción de la solución

A continuación, se detallará los pasos realizados para la construcción de la solución de BI:

5.3.1 Plan de Proyecto

5.3.1.1 Datos del Proyecto

NOMBRE DEL PROYECTO: Solución de inteligencia de negocios que permita automatizar la disponibilidad de la información enfocado en una empresa de transporte	SIGLAS: (ISNAP)
Inicio: 17/06/2022	Fin: 20/09/2022
DESCRIPCIÓN DEL PROYECTO: Para el desarrollo de la solución de inteligencia de negocios se analizaron diversas herramientas para la creación de flujo de procesamiento de ETL y además plataforma para la creación de tableros gráficos. Luego de realizar el análisis se optó por emplear las siguientes herramientas de Microsoft Power Automate para automatizar el proceso desde recibir los reportes, enviarlos a AWS S3 como repositorio en Cloud, después proceder a utilizar Pentaho Data Integration(PDI) para la transformación de los datos con la finalidad de poder almacenar las dimensiones de las tablas y la información en AWS Redshift y al final los reportes puedan mostrarse gráficamente en PowerBI. Para este caso, se empleó datos proporcionado por la organización de los cuales se ejecutaron pruebas de concepto con el objetivo de validar la solución.	

5.3.1.2 Datos del Cliente

NOMBRE DE LA EMPRESA:	Empresa del Sector transporte
RUBRO:	Transporte
DUEÑO DEL PROCESO:	Gerente General

5.3.1.3 Equipos de Trabajo

Tabla 8: Equipos de Trabajo

NOMBRE	ROL	CANTIDAD
Emmanuel Roque / Ricardo Cadillo	Jefe de proyecto	2
Luis Garcia	Analista Senior BI	1
Vanessa Saavedra	Analista de Arquitectura	1
Enrique Cortez	Analista de BD	1
Alexi Palomino Jaime Cordova Ricardo Gupioc Juan Luis Perez	Tester	4
Pedro Diaz	Analista Infraestructura IT	1

Fuente: Elaboración Propia

5.3.1.4 Lista de Necesidades del usuario

ID	Requisitos
1	Mostrar las ventas diarias por cada agencia de viaje
2	Mostrar las ventas acumulado por año
3	Mostrar el índice de ventas por meses
4	Mostrar las ventas por campaña

Tabla 9: Lista de Necesidades del usuario

Fuente: Elaboración Propia

5.3.1.5 Cronograma preliminar del proyecto e hitos

Tabla 10: Lista de Necesidades del usuario

Hito del proyecto	Fecha Estimada	Entregables incluidos	Prioridad
Análisis el uso de metodologías, técnicas y que permita automatizar la disponibilidad de la información contemplando inteligencia de negocio.	17/06/2022	Informe del análisis del uso de metodologías, técnicas y herramientas de la inteligencia de negocios	Alta
Diseño de una solución de inteligencia de negocios que permita medir el índice de las ventas en una empresa de transporte.	18/08/2022	Informe del diseño de la solución de inteligencia de negocios	Alta
Validación de la solución Business Intelligence con la empresa del sector transporte	20/09/2022	Informe de la validación de la solución Business Intelligence con la empresa del sector transporte	Alta

Fuente: Elaboración Propia

5.3.1.6 Presupuesto Estimado

Tabla 11: Presupuesto Estimado

Cargo	Número de personas	Remuneración Mensual (S/)	Cantidad de Meses	Remuneración Total (S/)
Jefe de Proyectos	1	4000	3	12,000
Analista Senior BI	1	3500	3	10,500
Analista Arquitectura	1	3500	3	10,500
Analista BD	1	3500	3	10,500
Tester	4	2000	3	24,000
Analista Infraestructura	1	3000	3	9,000
Total				76,500

Servicios	Mes	1er año	Moneda
Amazon S3	26.37	316.44	USD
Power BI	8.68	104	USD
Microsoft Power Automate	13.02	156.2	USD
Amazon Redshift	180	2160	USD

Fuente: Elaboración Propia

5.3.2 Proceso de desarrollo de software

5.3.2.1 Nombre del proyecto

Solución de inteligencia de negocios que permita automatizar la disponibilidad de la información enfocado en una empresa de transporte

5.3.2.2 Metodologías / Normas Utilizadas

- Metodología Kimball.- Se elige la metodología Kimball ante el costo bajo de implementación y el tiempo de desarrollo menor enfocándose solo a los proceso core de la empresa
- Guía de Buenas Prácticas del PMBOK.- Se eligió el marco de referencia de PMBOK, ya que nos brinda una guía de buenas prácticas a ejecutar en nuestra solución proveyéndonos de un conjunto de pasos y conocimientos para poder aplicarlo.
- Metodología RUP.- Se utilizó la metodología RUP porque le da un énfasis a los requerimientos, análisis, diseño y construcción lo cual se complementa con la metodología Kimball

5.3.2.3 Nombre del Proceso

- Proceso de Ventas de Pasajes

5.3.2.4 Lista de Subprocesos y roles

Tabla 12: Subprocesos y roles

SUBPROCESOS	ROL
Análisis y requisitos	Analista Funcional
Diseño de la solución	Arquitecto
Construcción de la solución	Analista BI / Analista BD/ Analista Infraestructura
Integración de los reportes	Analista BI
Validación de la solución	Tester

Fuente: Elaboración Propia

5.3.2.5 Proceso de desarrollo de software

Para la construcción del desarrollo se utilizaron unas de las fases de la metodología RUP, de las cuales nos enfocamos en recabar todos los requerimientos del negocio, para después poder realizar el diseño de la solución y la arquitectura. Asimismo, se procede con la construcción de la solución, con la integración de reportes y al último la validación con casos de prueba

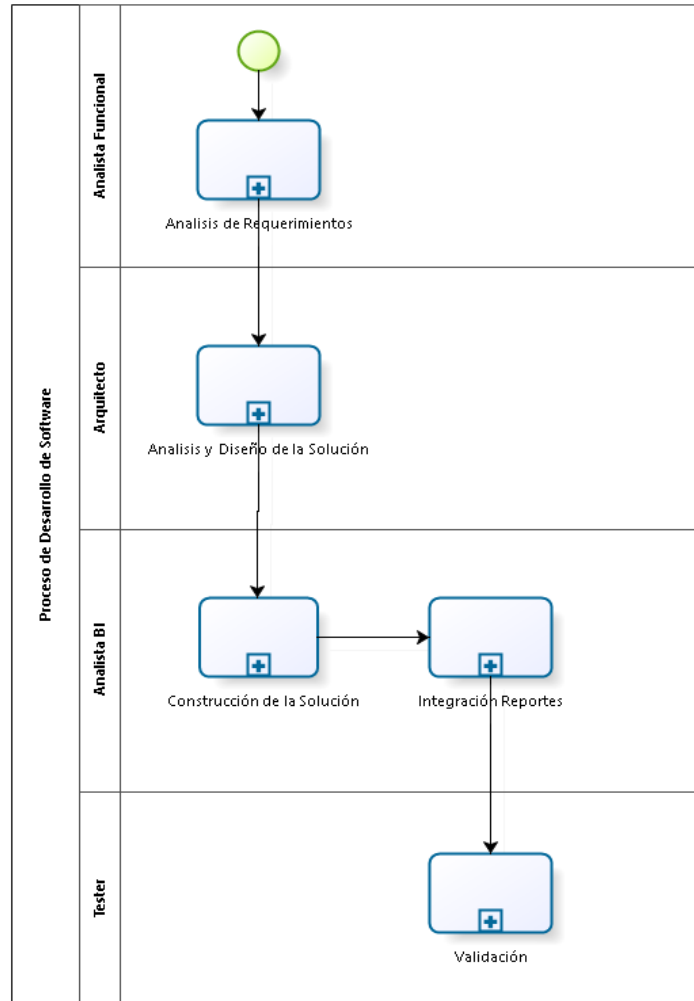


Ilustración 33: Diagrama del proceso del desarrollo de software

Fuente: Elaboración propia

5.3.3 Diagrama Específicos de subprocesos

Subproceso 1: Análisis de Requerimientos

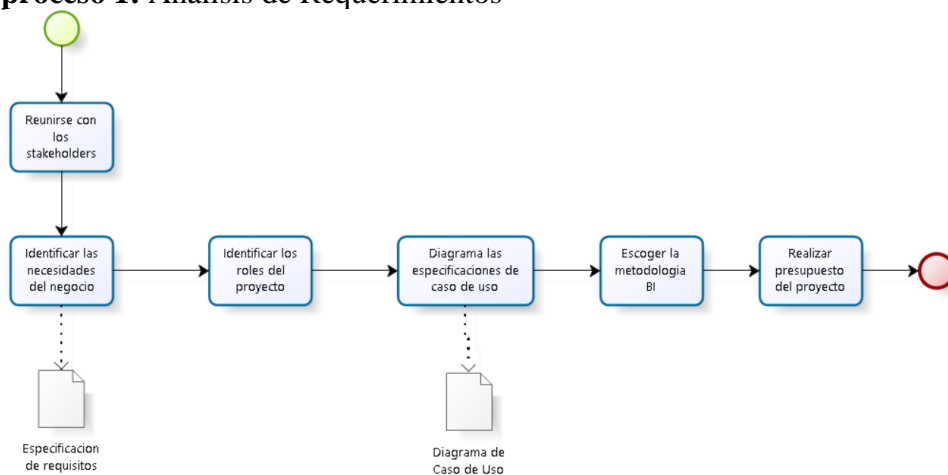


Ilustración 34: Diagrama del proceso del análisis de requerimientos

Fuente: Elaboración propia

Subproceso 2: Diseño de la solución

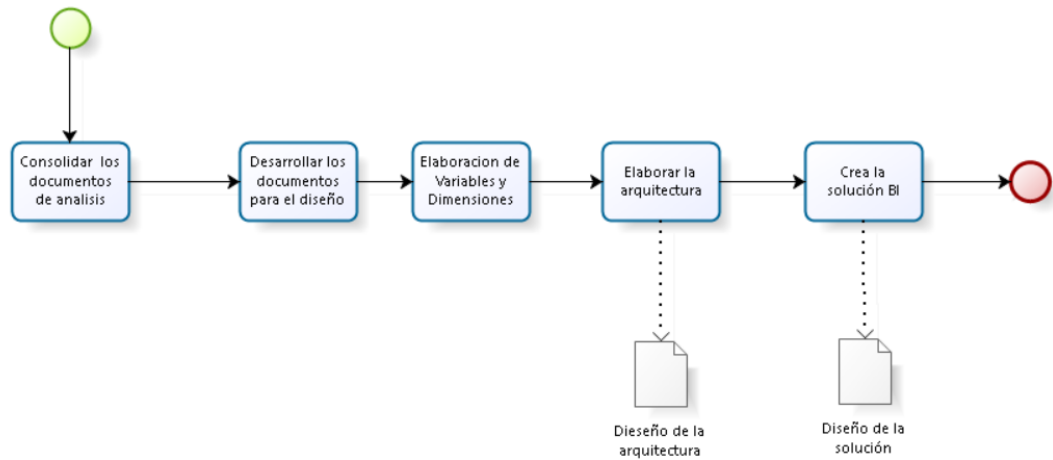


Ilustración 35: Diagrama del proceso del diseño de la solución

Fuente: Elaboración Propia

Subproceso 3: Construcción del software

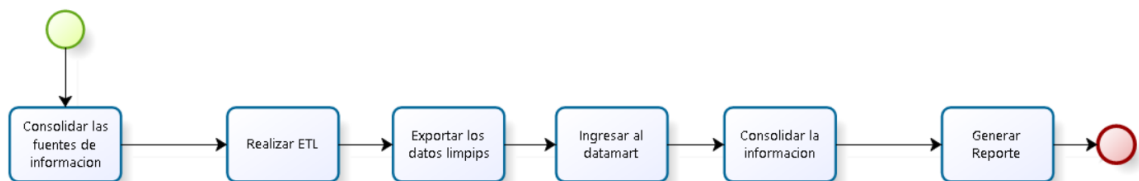


Ilustración 36: Diagrama del proceso de construcción

Fuente: Elaboración Propia

Subproceso 4: Validación

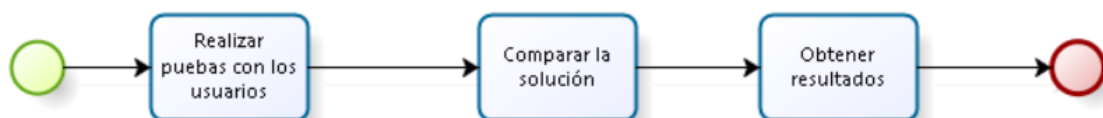


Ilustración 37: Diagrama del proceso de la validación

Fuente: Elaboración Propia

5.3.4 Obtención de requisitos

Tabla 8: Caso de Uso 1

CASO DE USO	Gestionar de Reporte de Ventas	
Descripción	El sistema deberá permitir al Asistente Administrativo poder descargar los reportes de ventas.	
Precondición		
Secuencia Normal	Paso	Acción
	1	El Sistema genera un reporte de las ventas para que lo pueda descargar el Asistente Administrativo
	2	El Asistente Administrativo podrá filtrar los reportes acordes a días, meses y años
Post condición	La descarga se realizó correctamente	
Excepciones	Paso	Acción
Importancia	Alta	

Fuente: Elaboración Propia

Tabla 9: Caso de Uso 2

CASO DE USO	Gestionar Programación de Viajes	
Descripción	El sistema deberá permitir al Asistente Administrativo poder gestionar la programación de los buses.	
Precondición		
Secuencia Normal	Paso	Acción
	1	El Asistente Administrativo ingresará la programación de la salida de los buses
	2	El Sistema generará las rutas y disponibilidad acorde al registro de programación
Post condición	La programación se realizó correctamente	
	Paso	Acción
	1	En el caso de que no se completen todos los datos necesarios el sistema emitirá un mensaje de error.

Excepciones	2	En caso de que el Asistente Administrativo ya hubiera generado más de 2 con el mismo dato del conductor en un mismo día, el sistema emitirá mensaje señalando que el conductor ya cuenta con su totalidad de viajes programados.
Importancia	Alta	

Fuente: Elaboración Propia

Tabla 10: Caso de Uso 3

CASO DE USO	Gestionar Reservas	
Descripción	El sistema deberá permitir al Vendedor reservar un viaje.	
Precondición		
Secuencia Normal	Paso	Acción
	1	El Vendedor revisará los datos del cliente en el sistema
	2	El Vendedor revisará la disponibilidad de las rutas y los asientos disponibles
	3	El sistema generará una nueva reserva
Post condición	La reserva se realizó correctamente	
Excepciones	Paso	Acción
	1	En el caso de que no se completen todos los datos necesarios el sistema emitirá un mensaje de error.
	2	En caso de que el Vendedor no encuentre asientos disponibles el sistema emitirá un mensaje de error.
Importancia	Alta	

Fuente: Elaboración Propia

Tabla 11: Caso de Uso 4

CASO DE USO	Gestionar la información	
Descripción	El sistema deberá permitir al Analista de BI descargar las ventas.	
Precondición		
Secuencia Normal	Paso	Acción
	1	El Analista de BI podrá descargar las fuentes de las áreas de ventas
	2	El Analista de BI podrá consolidar la información para poder ingresarlo a la solución

Post condición	La información se descargó correctamente
Importancia	Alta

Fuente: Elaboración Propia

Tabla 12: Caso de Uso 5

CASO DE USO	Gestionar Integración de los datos	
Descripción	El sistema deberá permitir al Analista de BI poder integrar la información en un datamart con Amazon Redshift	
Precondición		
Secuencia Normal	Paso	Acción
	1	El Analista BI podrá ingresar la información en el datamart
	2	El sistema permitirá al Analista de BI puede agregar y segmentar la información
Post condición	La programación se realizó correctamente	
Excepciones	Paso	Acción
	1	En el caso de que las variables presenten algún campo erróneo aparecerá un mensaje indicando un error en los datos
Importancia	Alta	

Fuente: Elaboración Propia

5.3.5 Análisis del modelo dimensional

En base a la aplicación de Soluciones de Negocios en una empresa de sector transporte hemos identificado su proceso core de Ventas, los cuales necesitan generar reportes dinámicos y de impacto para la toma de decisiones donde se ha procedido a identificar las dimensiones de la solución

Tabla 13: Dimensiones de la solución

Dimensión	Descripción
Clientes	Dimensión que representa a los clientes que adquieran su boleto de viaje
Buses	Dimensión que representa a los

	buses de la empresa del sector transporte
Rutas	Dimensión que representa los puntos de origen y destino del transporte de la empresa
Canal de Venta	Dimensión que representa el canal de venta por donde se adquirió los boletos
Tiempo	Dimensión que representa los datos relacionados a las fechas y la secuencia cronológica

Fuente: Elaboración Propia

Para proceder a la construcción de un modelo dimensional es importante definir las tablas hechos (fact tables), puesto que será la tabla que contiene campos claves que se unen a las tablas de dimensión. Además, contempla las métricas o medidas y por lo general son valores numéricos.

Tabla 14: Tabla hechos

Fact	Descripción
Ventas pasajes	Representa las métricas respecto a las ventas de pasajes en una fecha establecida.

Fuente: Elaboración Propia

5.3.6 Esquema estrella

El tipo de esquema que se está empleando para este análisis de BI es el esquema de estrella, que consta de una sola tabla de hechos central y relacionada de tablas de dimensiones. Asimismo, es el más práctico a la hora de diseñar un datawarehouse.

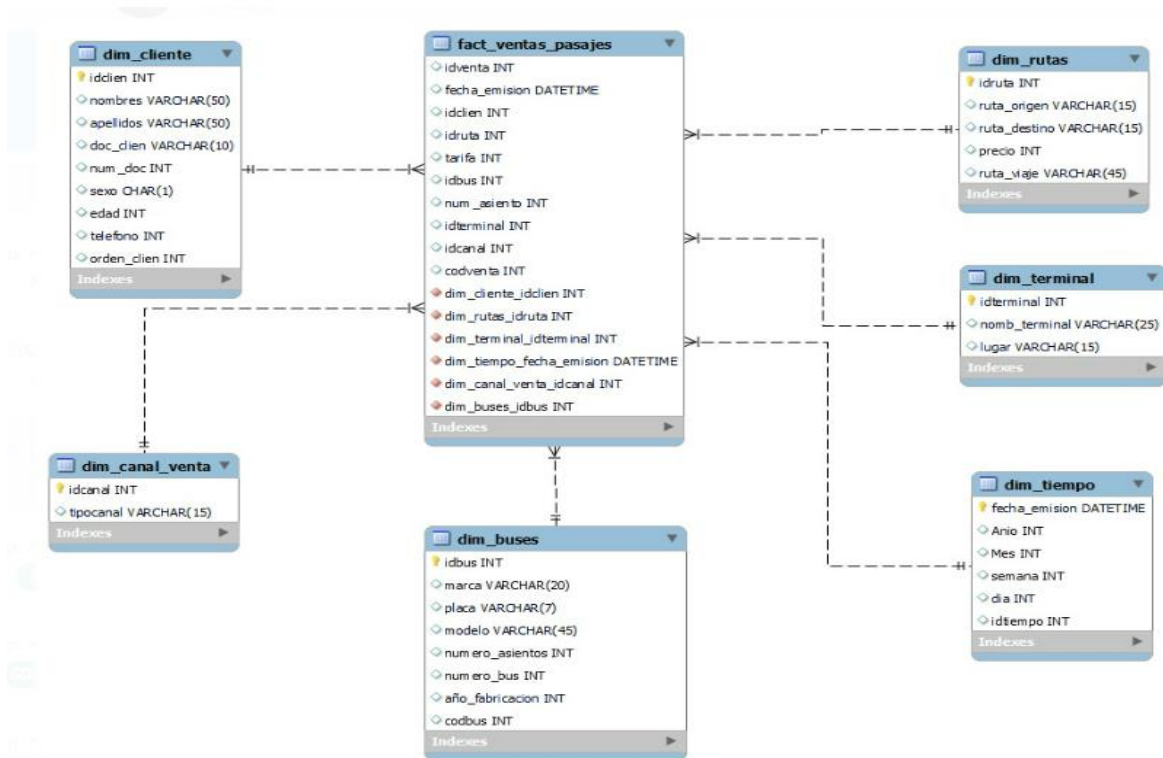


Ilustración 38: Diagrama del modelo de datos

Fuente: Elaboración Propia

5.3.7 Creación de dimensiones y modelo físico

En este proceso se definieron las dimensiones para proceder con la transformación de las dimensiones en base a la siguiente distribución.

Tabla 15: Dimensión Cliente

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
dim_cliente	idclien	int	PK	SI
dim_cliente	nombres	varchar (50)	-	SI
dim_cliente	apellidos	varchar (50)	-	SI
dim_cliente	doc_clien	varchar (10)	-	SI
dim_cliente	num_doc	Int	-	SI
dim_cliente	sexo	Char(1)	-	NO
dim_cliente	edad	int	-	NO
dim_cliente	teléfono	Int	-	NO

Fuente: Elaboración Propia

Tabla 16: Dimensión terminal

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
dim_terminal	idterminal	int	PK	SI
dim_terminal	nomb_terminal	varchar (25)	-	SI
dim_terminal	lugar	varchar (15)	-	SI

Fuente: Elaboración Propia

Tabla 17: Dimensión tiempo

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
dim_tiempo	idtiempo	int	PK	SI
dim_tiempo	fecha	date	-	SI
dim_tiempo	anio	int	-	SI
dim_tiempo	mes	int	-	SI
dim_tiempo	semana	int	-	SI
dim_tiempo	día	int	-	SI

Fuente: Elaboración Propia

Tabla 18: Dimensión buses

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
dim_buses	idbus	int	PK	SI
dim_buses	marca	varchar (15)	-	SI
dim_buses	placa	varchar (7)	-	SI
dim_buses	modelo	varchar (45)	-	NO
dim_buses	numero_asientos	int	-	NO
dim_buses	numero_bus	int	-	SI
dim_buses	año_fabricacion	int	-	SI

Fuente: Elaboración Propia

Tabla 19: Dimensión rutas

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
dim_rutas	idruta	int	PK	SI
dim_rutas	ruta_origen	varchar (15)	-	SI
dim_rutas	ruta_destino	varchar (15)	-	SI

dim_rutas	precio	int	-	SI
------------------	--------	-----	---	-----------

Fuente: Elaboración Propia

Tabla 20: Dimensión canal de venta

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
dim_canal_venta	idcanal	int	PK	SI
dim_canal_venta	tipocanal	varchar(15)	-	SI

Fuente: Elaboración Propia

Tabla 21: Tabla hecho venta de pasajes

Tabla	Columna	Tipo de Datos	Tipo de Llave	Requerido
fact_ventas_pasajes	idventa	int	PK	SI
fact_hechos_viajes	idclien	int	FK	SI
fact_hechos_viajes	idruta	int	FK	SI
fact_hechos_viajes	idterminal	int	FK	SI
fact_hechos_viajes	Idcanal	int	FK	SI
fact_hechos_viajes	idtiempo	int	FK	SI
fact_hechos_viajes	num_asiento	int	-	SI
tabla_hechos_viajes	tarifa	int	-	SI

Fuente: Elaboración Propia

5.3.8 Proceso ETL

En el proceso de ETL se necesitaron las ventas de la empresa del sector de transporte que son requeridas para el modelo. Para este caso se hizo una integración entre Mysql y Pentaho.

En primera instancia parte del proceso de la ejecución de ETL se desarrolla en la plataforma Pentaho Data Integration (PDI) el cual permitirá realizar la extracción y transformación de los datos, posterior a ello se procederá a almacenar toda la información en la base de datos local MySQL. Por lo tanto, se debe establecer la conexión entre ambos componentes, a continuación, se detalla los parámetros utilizados para la configuración correspondiente.

Tabla 22: Configuración entre PDI y Amazon Redshift

Campos	Valores
---------------	----------------

Connection name	PruebaRedshift
Hostname	redshift-cluster-1.c8dwdozpcl3.us-east-1.redshift.amazonaws.com:5439/ventasctb
Database Name	ventasctb
Port Number	5439
Username	awuser
Password	xxxx

Fuente: Elaboración Propia

✓ En la imagen se muestra que la conexión entre la base de datos y PDI fue exitosa.

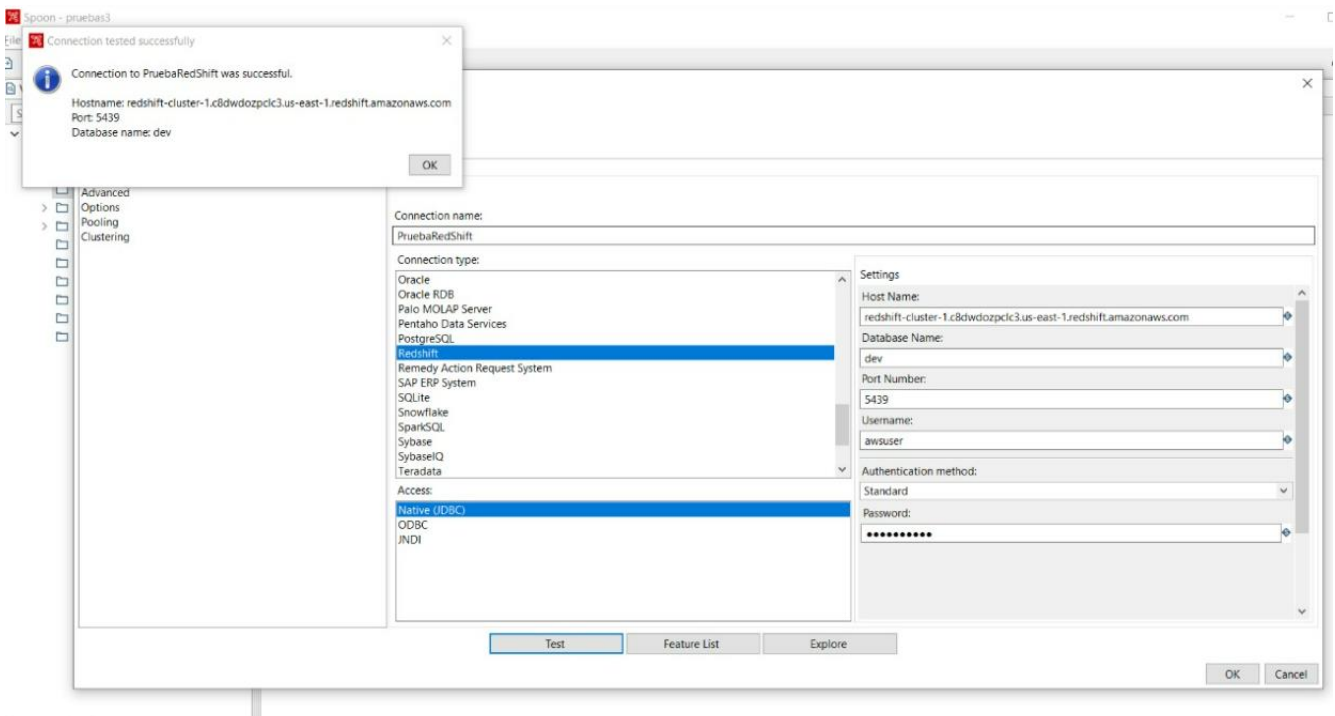


Ilustración 39: Configuración entre PDI y AWS Redshift

Fuente: Elaboración Propia

✓ **Job Principal de la solución ETL**

En la solución se utilizará el proceso de ETL mediante un Job Principal donde comenzará a ejecutarse de manera secuencial todas las transformaciones de la solución desde la limpieza de las tablas mediante un script generado, la carga de todas las dimensiones y enfocarla hasta la tabla hechos “fact_ventas_pasajes” del datamart para posterior proceder a exportar los datos a la plataforma de PowerBI y generar los dashboard dinámicos.

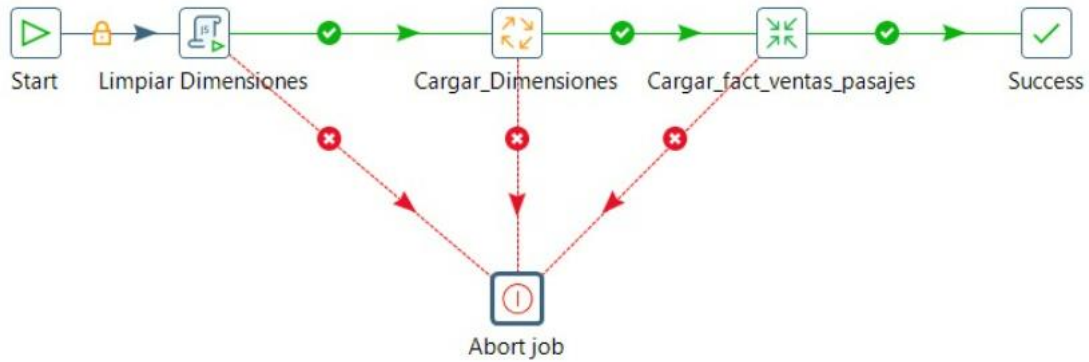


Ilustración 40: Job Principal de la solución

Fuente: Elaboración Propia

✓ **Job Cargar Dimensiones**

En el siguiente gráfico se realiza la secuencia de procedimientos que se realiza en la aplicación desde la preparación de la información de la Base de datos, la transformación de los datos, la carga de las dimensiones correspondientes a un Job anidado. Cada transformación seguirá una ruta por cada dimensión las cuales alberga la información en sus tablas

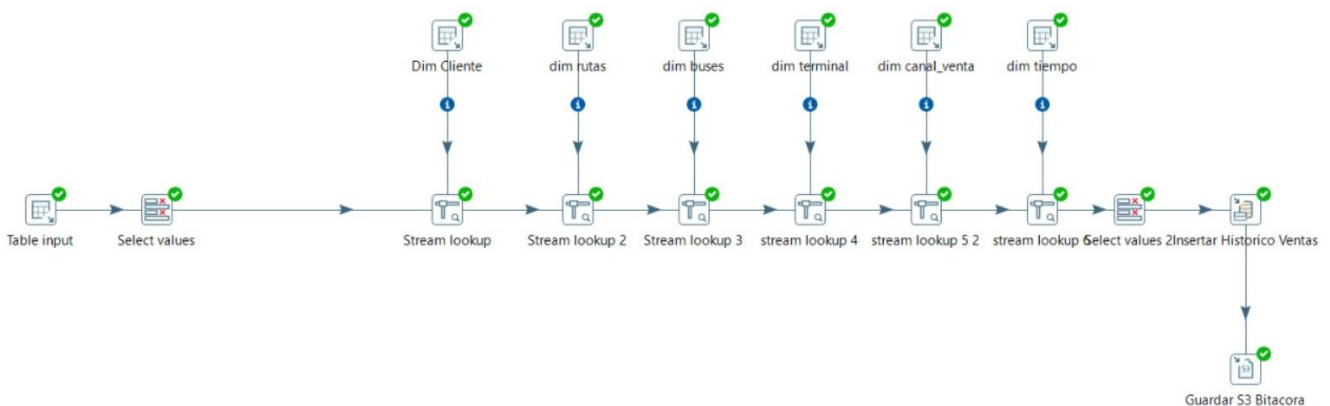


Ilustración 41: Job Cargar Dimensiones

Fuente: Elaboración Propia

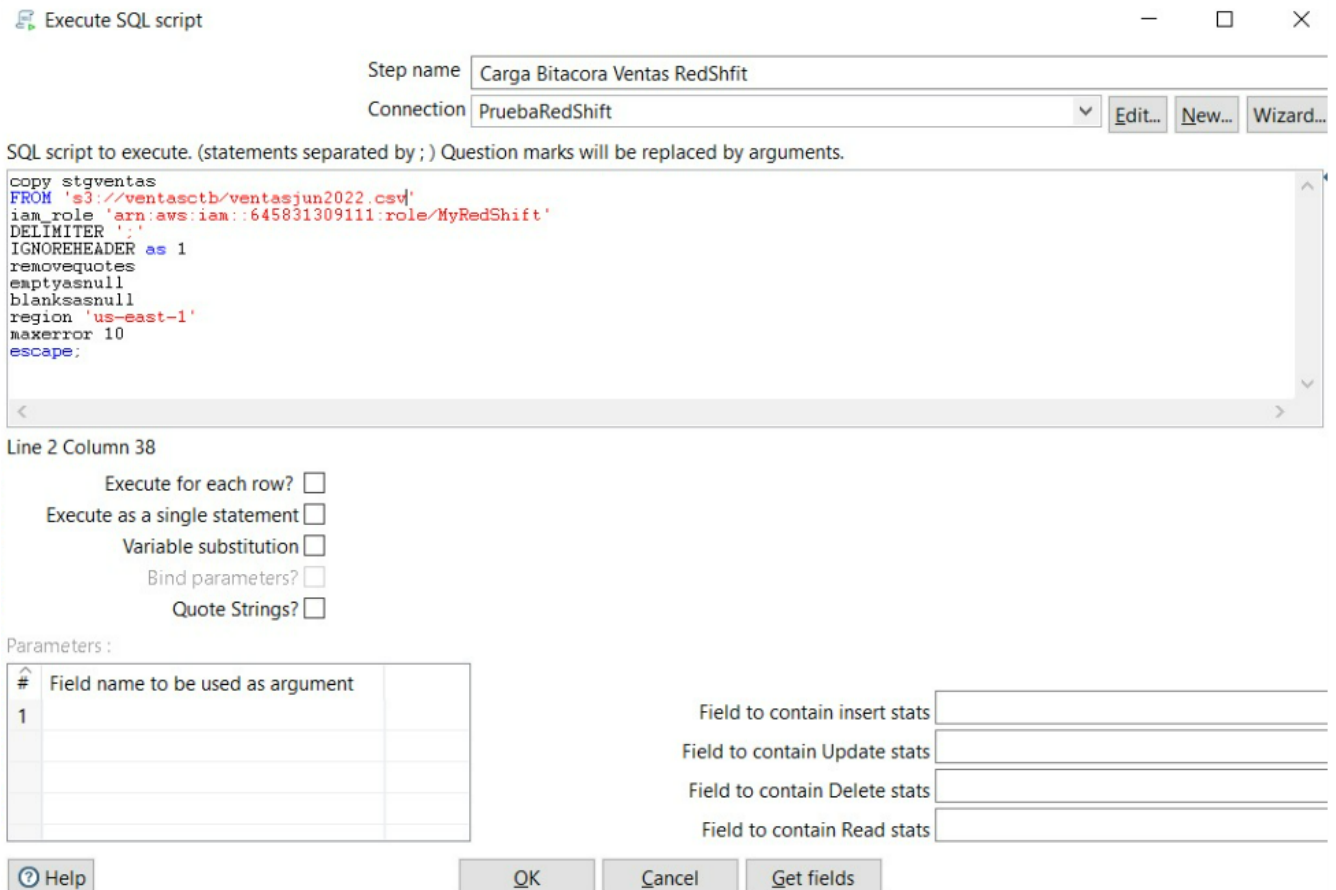
✓ **Componente Transformación**

La siguiente sección contempla la vista de la transformación principal para la carga de clientes; en este punto se muestra el flujo para realizar la carga de la dimensión clientes. A

continuación, mostraremos el detalle de los componentes utilizados.



Ilustración 42: Vista general de transformación Carga de la información



Fuente: Elaboración Propia

Ilustración 43: Vista general del envío de la información a Amazon Redshift

Fuente: Elaboración Propia

- ✓ Input de reportes en Excel de clientes

A continuación, definimos los parámetros del documento el cual se cargarán en un formato “Excel 2007 XLSX (Apache POI)”, ahora ubicamos la ruta donde se encuentra el archivo y

procedemos a agregar. En la pestaña “Sheets”, agregamos el ítem que contiene la información del reporte; al seleccionar nos muestra la pestaña “cliente”. Por último, para evidenciar que los datos que se cargarán son correctos seleccionamos la opción “Preview rows” con el fin de mostrar todos los datos a cargar.

- ✓ Ubicar el directorio del archivo

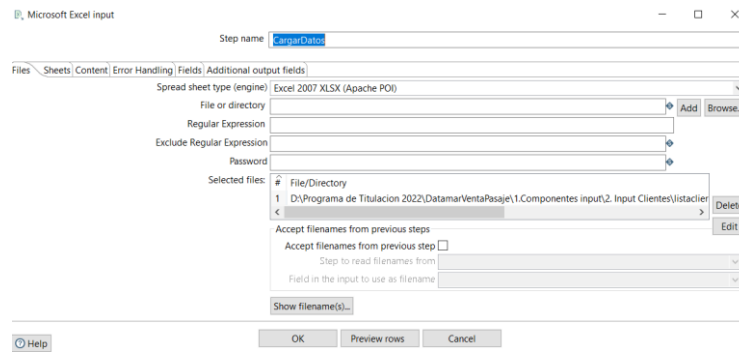


Ilustración 44: Ruta del archivo

Fuente: Elaboración Propia

- ✓ Seleccionar el ítem el cual contiene los datos

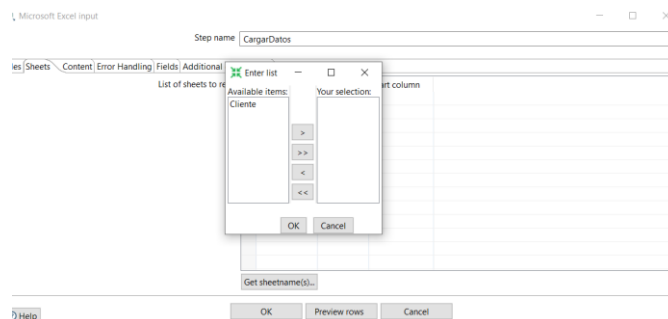


Ilustración 45: Ítem de los datos

Fuente: Elaboración Propia

- ✓ Vista previa de los datos a cargar

Examine preview data

Rows of step: CargarDatos (741 rows)

#	idcliente	Nombres	Apellidos	tipo_documento	nro_documento	sexo	edad	direccion	fecha_nacimiento	telefono
1	1.0	Juan Andres	Lopez Cabello	DNI	45345435.0	M	45.0	<null>	1977/05/24 00:00:00.0000	985645258.0
2	2.0	Jorge Luis	Mejia Flores	DNI	76754535.0	M	24.0	<null>	1998/07/05 00:00:00.0000	946407650.0
3	3.0	Erika Maria	Mora Rivera	DNI	34565678.0	F	34.0	<null>	1988/08/13 00:00:00.0000	922027249.0
4	4.0	Carlos Jose	Chumbez la Rosa	DNI	47245345.0	M	56.0	<null>	1966/11/10 00:00:00.0000	939998147.0
5	5.0	Mayra Daila	Rivera Santos	DNI	54356356.0	F	32.0	<null>	1990/04/02 00:00:00.0000	924185375.0
6	6.0	Marcial	Aspajo Guerra	DNI	8323754.0	M	30.0	<null>	<null>	914579988.0
7	7.0	Rolando	Escobar Colonia	DNI	2198781.0	M	39.0	<null>	<null>	982793664.0
8	8.0	Antonio	Padilla Agurto	DNI	66684904.0	M	25.0	<null>	<null>	995142950.0
9	9.0	Teodoro	Gaslac Torres	DNI	46107523.0	M	38.0	<null>	<null>	979781343.0
1.	10.0	Oswaldo	Franco Perez	DNI	19769235.0	M	47.0	<null>	<null>	957278805.0
1.	11.0	Jose Manuel	Vasquez Palomino	DNI	26707044.0	M	32.0	<null>	<null>	955860052.0
1.	12.0	Ricardo Lorenzo	Coronel Calderon	DNI	62786594.0	M	32.0	<null>	<null>	949382316.0
1.	13.0	Hoyos Fernando	Carrasco Hoyos	DNI	34445225.0	M	41.0	<null>	<null>	903383656.0
1.	14.0	Jorge Eduardo	Tejeda Mora Jorge	DNI	67937133.0	M	31.0	<null>	<null>	936647669.0
1.	15.0	Rusber Felipe	Lazaro Arquinigo	DNI	18105981.0	M	35.0	<null>	<null>	993048339.0
1.	16.0	Sense Lorenzo	Perez Sinosai	DNI	13815267.0	M	25.0	<null>	<null>	936359338.0
1.	17.0	Ramon Leonino	Alvarado Sanchez	DNI	11412483.0	M	36.0	<null>	<null>	980390186.0
1.	18.0	Miguel Abelardo	Huamani Valdivia	DNI	74753815.0	M	36.0	<null>	<null>	919378881.0
1.	19.0	Jose Vidal	Capcha Arango	DNI	75873039.0	M	47.0	<null>	<null>	977321512.0
2.	20.0	Manuel Jesus	Tapia Lazaro	DNI	17386887.0	M	29.0	<null>	<null>	915689062.0
2.	21.0	Hugo Alfredo	Ruiz Gutierrez	DNI	52004861.0	M	27.0	<null>	<null>	995577832.0
2.	22.0	Jorge Alexander	Minaya Gutierrez	DNI	77027347.0	M	44.0	<null>	<null>	964334300.0
2.	23.0	Markho Antonio	Moreno Soto	DNI	34462413.0	M	46.0	<null>	<null>	934276136.0
2.	24.0	Luis Alberto	Chivilchez Espinoza	DNI	40085481.0	M	45.0	<null>	<null>	927535705.0
2.	25.0	Paulo Cesar	Tupaypanqui Zulfiga	DNI	96576577.0	M	50.0	<null>	<null>	903285584.0

Close Show Log

Ilustración 46: Vista previa de los datos

Fuente: Elaboración Propia

- ✓ Agregar secuencia incremental

Se define una nueva variable para generar un valor incremental al id del cliente.

Ilustración 47: Vista secuencia incremental

Fuente: Elaboración Propia

- ✓ Selección de valores

En esta sección, podemos renombrar a las variables al momento de realizar la carga, además en la pestaña “Remove” se puede obviar los campos que no deseamos guardar en el proceso de carga.

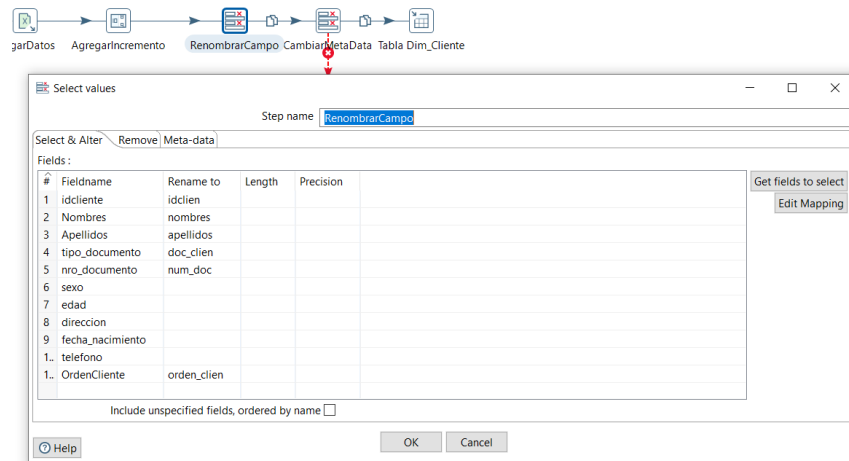


Ilustración 48: Vista selección de variables

Fuente: Elaboración Propia

En la pestaña de meta-data realizamos la definición del tipo de formato para cada campo y la longitud, ya que al momento de realizar la carga no presente ningún error de al generar la tabla en la base de datos MySQL.

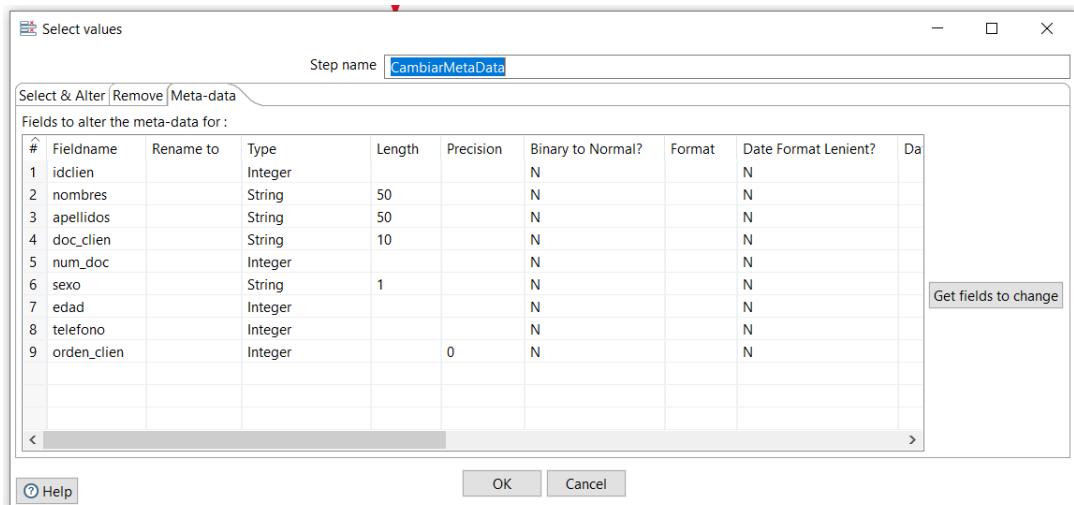


Ilustración 49: Vista de formato de variable

Fuente: Elaboración Propia

✓ Generación de Log

Por lo general en la vista de transformación se dispone de un log con el fin de monitorear que campos serán almacenado dentro del modelo.

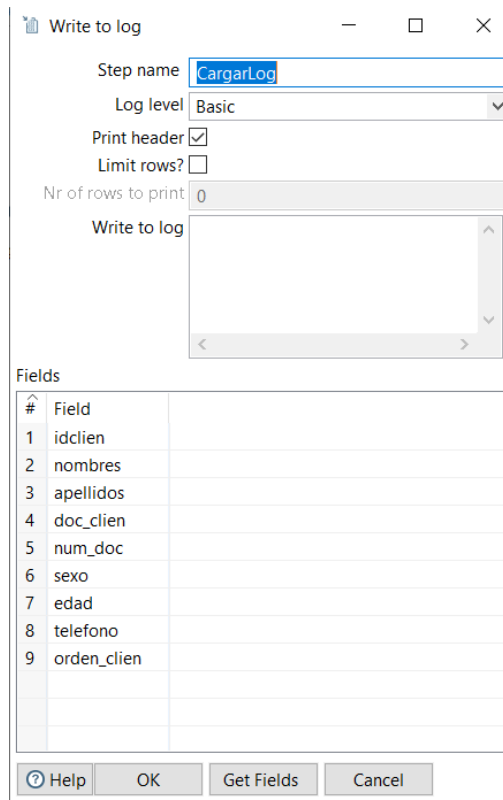


Ilustración 50: Vista generación de log

Fuente: Elaboración Propia

✓ Output

En la tabla de salida definimos la conexión de la base de datos donde se almacenará la carga de los registros de los clientes. Por ello; en la imagen se puede apreciar el nombre de la Conexión y la tabla en donde se cargará.

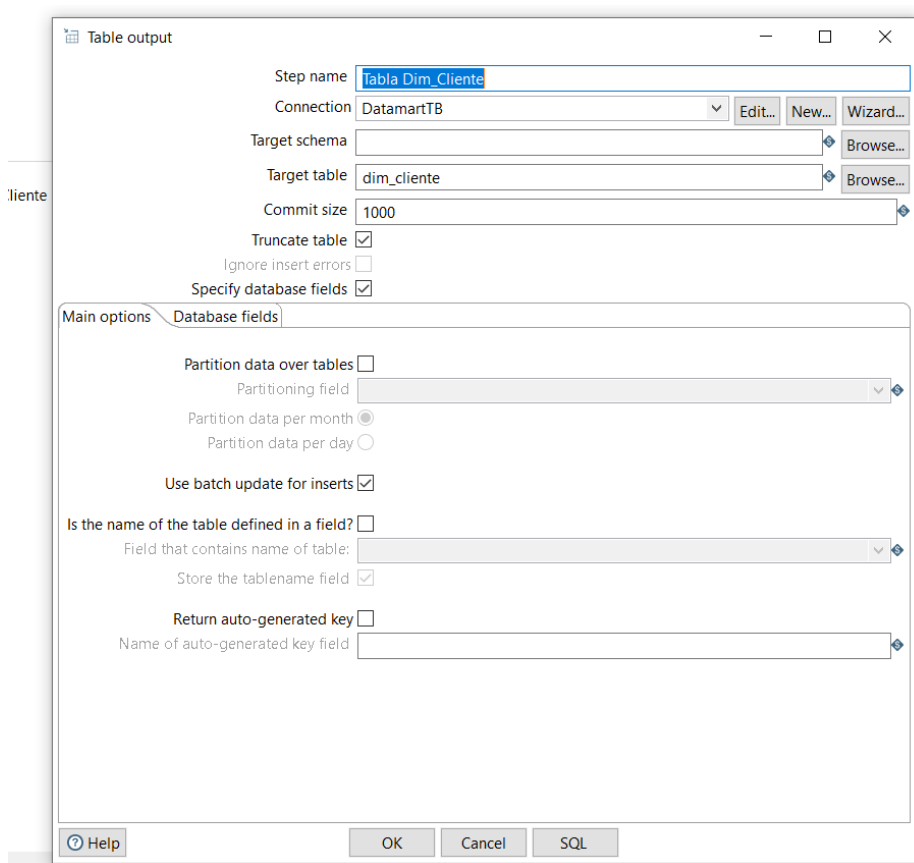


Ilustración 51: Vista de salida y carga de datos

Fuente: Elaboración Propia

5.3.9 Plan de construcción

5.3.9.1 Herramientas y tecnologías

Tabla 23: Herramientas y tecnologías

Programas / Herramientas	Descripción
Pentaho Data Integration	Es una herramienta que forma parte de la suite de multiplataformas de Open Source Pentaho permitiendo al usuario poder tomar decisiones
Amazon Redshift	Es un sistema de gestión de bases de datos relacional más extendido en la actualidad al estar basada en código abierto
Power BI	Es una herramienta que es utilizada para

	crear cuadros de mando que faciliten la toma de decisiones.
Power Automate	Es el sistema empresarial de Microsoft que permite automatizar procesos y tareas de manera sencilla permitiendo disponer de un flujo funcional sin necesidad de desarrollar código o con muy poco código a esto se le conoce como Low Code.
Amazon S3	Es un servicio de Amazon AWS cuya función sirve de almacenamiento en la nube donde te permite guardar y recuperar datos cualquiera que sea su volumen en el cual el precio va acorde el almacenamiento. Asimismo, puede ser usado para almacenar web estático, almacenar imágenes, reportes, imágenes, etc.

Fuente: Elaboración Propia

5.3.9.2 Lista de intereses y plantilla

Tabla 24: Lista de Intereses y plantilla

Estándar / Plantilla	Nomenclatura
Estándar de Base de datos Mysql	E-BD-MY.DOCX
Plantilla Analisis de Requisitos	P-AR.DOCX
Plantillas de métricas	P-MTR.DOCX
Estándar del ETL	E-ETL-BI.DOCX

Fuente: Elaboración Propia

5.3.9.1 Lista de artefactos

Tabla 25: Lista de Artefactos

Artefacto	Nomenclatura
Cronograma de Proyecto	BI-CP. XSLX
Acta de constitución del proyecto	BI-AC.DOCX

Documento de Análisis de requerimientos	BI-AB.DOCX
Documento de análisis de la metodología de Inteligencia de Negocios	BI-MT.DOCX
Documento de análisis de herramientas BI	BI-AH.DOCX
Documento Diseño de la solución	BI-DS.DOCX
Documento de la arquitectura de la solución	BI-AR.DOCX
Diagrama de Modelo Lógico de datos	BI-ML.DOCX
Documento de las métricas de calidad	BI-CS.DOCX
Documento de la construcción del ETL	BI-DMC.DOCX
Documento de los dashboard	BI-DS.DOCX
Documento de las entrevistas	BI-ENT.DOCX

Fuente: Elaboración Propia

6. RESULTADO DEL PROYECTO

En el presente capítulo se mostrará los procedimientos del resultado del proyecto. Asimismo, se comparten gráficos y tablas para mostrar lo desarrollado a lo largo del proyecto.

6.1 Mejora en la disponibilidad de información en el proceso de ventas

Los datos serán recopilados de los archivos CSV mediante Onedrive, el cual será enviado por el un servicio de SFTP para después almacenarlo en Amazon AWS S3 para después alimentar al proceso de Pentaho.

Cabe resaltar que todo el proceso estará soportado en Power Automate, esto trae consigo que el proceso sea automatizado y programado en una hora y fecha específica permitiendo así al Analista de BI solo este al pendiente de la solución si el archivo no se encuentra dentro de Onedrive si pasara ello, la solución no se ejecutaría.

Continuando con la solución, una vez la información pase por el flujo de trabajo de ETL la herramienta enviara toda la información a la data warehouse el cual se encontrara almacenado en Amazon AWS Redshift para después ser enviado a PowerBi para mostrar toda la información a través de gráficos que serán mostrados a la alta gerencia. Con ello, se puede medir el índice de ventas permitiendo a la organización saber cuándo invertir mediante los reportes de segmentación de edades, terminales e ingresos mensuales por terminal.

6.2 Estimación de tiempos de ejecución

En resumen, se mostrará el tiempo que comprende la ejecución de la herramienta al realizar el proceso completo del ETL desde la carga de datos y la exportación de los datos. Por lo general, se realizó un plan de prueba con la finalidad de analizar la performance necesaria para la solución de BI. Además, validar la funcionalidad, eficiencia de desempeño, usabilidad y fiabilidad. Se realizaron 9 iteraciones de los cuales contempla el flujo principal(Job) y se empleó la media como variable de prueba. Además, se simularon los reportes de venta del año 2022 entre los meses de Enero, febrero, Marzo, Abril, Mayo, Junio para visualizar el comportamiento de las ventas.

Tabla 26: Flujo de Proceso ETL

Flujo de Proceso ETL		Iteración	Tiempo promedio en procesar las ventas
Job Principal	Carga de Datos general	1	0.80
Job Carga de Dimensiones	Carga toda la dimensiones	2	0.38
Transformación	Dim Cliente	3	0.50
	Dim Buses	4	0.20
	Dim Rutas	5	0.20
	Dim Terminal	6	0.30
	Dim CanalVenta	7	0.12
	Dim Tiempo	8	0.70

	Tabla Hechos	9	0.60
Mediana			0.38

Fuente: Elaboración Propia

6.3 Modelo de Calidad

Para ponderar las características y sub-características se elaboró el modelo de calidad de su proyecto de software, según la ISO/IEC 9126:

6.3.1 Ponderado de características

Tabla 27: Ponderado de características

Clasificación	Valor
BAJA	1-3
MEDIA	4-6
ALTA	7-9

Fuente: Elaboración Propia

Tabla 28: Clasificación de las características

CARACTERÍSTICAS	CLASIFICACIÓN	VALOR	PESO	%	JUSTIFICACIÓN
FUNCIONALIDAD	Alta	9	0.35	35%	Por ser una aplicación de uso intenso de cálculos, debe ser lo más exacta posible, segura por trabajar con datos sensibles y altamente operable por el nivel de tecnificación de los usuarios.
FIABILIDAD	Alta	8	0.22	22%	La aplicación tiene que seguir operando en caso presente algún error en la transacción. Por este motivo, ante cualquier eventualidad errónea se generan logs que permitirá poder identificar la falla.
USABILIDAD	Alta	7	0.20	20%	La aplicación puede ser controlada y operada con facilidad debido a su interfaz amigable y los componentes que los incorpora al momento de hacer uso en la integración de la información.
EFICIENCIA	Media	4	0.15	15%	El tiempo de respuesta de la solución de BI en ejecución del Job Principal no debe ser mayor a 1 hora.
MANTENIBILIDAD	Baja	3	0.05	5%	Se requiere que en el futuro se pueda agregar nuevas funcionalidades a la solución de BI
PORTABILIDAD	Baja	2	0.03	3%	La aplicación solo va a estar soportada en Pentaho Data Integration.
		33	1.0000	100%	

Fuente: Elaboración Propia

6.3.2 Definición de métricas calidad

Tabla 29: Métrica Usabilidad

Código:	M10
Nombre de la Métrica:	Claridad de la descripción
Característica:	Usabilidad

Sub Característica:	Entendimiento
Propósito:	¿Qué proporción de las funciones (o tipos de funciones) es entendida después de leer la descripción del producto?
Método de aplicación:	Contar el número de las funciones que se entienden adecuadamente y comparar con el número total de funciones en el producto.
Medición, fórmula y elementos medibles:	$X = A / B$
	A = Nro. de funciones (o tipos de funciones) entendidas
	B = Total de Nro. de funciones (o tipo de funciones)
Tipo de medida:	A = Contador
	B = Contador
Interpretación del valor medido:	$0 <= X <= 1$
	Lo más cercano a 1 es lo mejor
Tipo de escala:	Absoluta
Entrada para la medición:	Reporte de operación

Fuente: ISO 12207

Tabla 30: Métrica Eficiencia

Código:	M11
Nombre de la Métrica:	Rendimiento
Característica:	Eficiencia
Subcaracterística:	Comportamiento en el tiempo
Propósito:	¿Cuántas tareas pueden ejecutarse satisfactoriamente en un determinado intervalo de tiempo?
Método de aplicación:	Medir el tiempo que toma completar la operación de las tareas medidas y mantener un registro de cada intento.
Medición, fórmula y elementos medibles:	$X = A / T$. A = Número de tareas completadas T = Intervalo de tiempo de observación.
Tipo de medida:	A = Contador T = Tiempo en segundos
Interpretación del valor medido:	$0 < X$ El mayor valor es lo mejor.
Tipo de escala:	Ratio
Entrada para la medición:	Reporte de operación

Fuente: ISO 12207

Tabla 31: Métrica Fiabilidad

Código:	M12
----------------	------------

Nombre de la Métrica:	Densidad de fallas contra los casos de prueba
Característica:	Fiabilidad
Subcaracterística:	Madurez
Propósito:	¿Cuántas fallas fueron detectadas durante el periodo de prueba definido?
Método de aplicación:	Contar el número de fallas detectadas y casos de prueba ejecutados.
Medición, fórmula y elementos medibles:	$X = A1 / A2$
	A1 = Número de fallas detectadas.
	A2 = Número de casos de prueba ejecutados.
Tipo de medida:	A = Contador
	B = Contador
Interpretación del valor medido:	$0 < = X$
	Cuando más pequeño el valor es lo mejor
Tipo de escala:	Absoluta
Entrada para la medición:	Reporte de operación
	Caso de Prueba

Fuente: ISO 12207

6.3.3 Estimación de métricas calidad

Tabla 32: Métrica Eficiencia del Entendimiento

NOMBRE DE LA MÉTRICA	EFICIENCIA DEL ENTENDIMIENTO		
CÓDIGO:	M10		
CARACTERÍSTICA	USABILIDAD		
SUBCARACTERÍSTICA	ENTENDIMIENTO		
Tester	A	B	1-(A/B)
	NÚMERO DE FALLAS	NÚMERO DE INTENTOS	RATIO
EMMANUEL	2	15	0.87
RICARDO	4	20	0.80
LUIS	4	35	0.89
PEDRO	6	20	0.70
PROMEDIO			0.82
<i>El valor más cercano a 1 (usuario no tuvo ninguna falla) es lo mejor</i>			

Fuente: Elaboración Propia

Interpretación:

Los datos obtenidos de la métrica corresponden a un promedio de ratio 0.82 por lo cual se encuentra dentro del rango establecido. Llegando a la conclusión que el dato registrado es aceptable.

Tabla 33: Métrica de Rendimiento

NOMBRE DE LA MÉTRICA	RENDIMIENTO		
CÓDIGO:	M11		
CARACTERÍSTICA	EFICIENCIA		
SUBCARACTERÍSTICA	COMPORTAMIENTO DEL TIEMPO		
Tester	A	B	(A/B)
	NÚMERO DE TAREAS COMPLETADAS	INTERVALO DE TIEMPO DE OBSERVACIÓN (segundos)	RATIO
EMMANUEL	10	15	0.67
RICARDO	12	17	0.71
LUIS	15	18	0.83
PEDRO	14	17.5	0.80
PROMEDIO			0.76
<i>El valor más cercano a 1 (usuario no tuvo ninguna falla) es lo mejor</i>			

Fuente: Elaboración Propia

Interpretación:

Los datos obtenidos de la métrica corresponden a un promedio de ratio 0.76 por lo cual se encuentra dentro del rango establecido. Llegando a la conclusión que el dato registrado es aceptable.

Tabla 34: Métrica de eficiencia del entendimiento

NOMBRE DE LA MÉTRICA	EFICIENCIA DEL ENTENDIMIENTO		
CÓDIGO:	M12		
CARACTERÍSTICA	FIABILIDAD		
SUBCARACTERÍSTICA	MADUREZ		
Tester	A	B	(A/B)
	NÚMERO DE FALLAS	NÚMERO DE CASO DE PRUEBA EJECUTADOS	RATIO
EMMANUEL	4	6	0.67
RICARDO	2	5	0.40
LUIS	3	5	0.60
PEDRO	2	4	0.50
PROMEDIO			0.54

El valor más cercano a 1 (usuario no tuvo ninguna falla) es lo mejor

Fuente: Elaboración Propia

Interpretación:

Los datos obtenidos de la métrica corresponden a un promedio de ratio 0.54 por lo cual se encuentra dentro del rango establecido. Llegando a la conclusión que el dato registrado es aceptable.

6.1 Visualización Dashboard

Una vez generado la ejecución de los ETL y ya habiendo procedido guardar en la BD en Amazon AWS Redshift. A continuación, realizamos una conexión en la plataforma de PowerBI con el fin de poder visualizar de manera dinámica los reportes y lo cual se presenta a la gerencia para realizar las tomas de decisiones de las ventas obtenidas.

En el presente grafico podemos observar la cantidad de pasajes vendido por cada mes lo cual le permite a la empresa poder ver el comportamiento del cliente y poder invertir más en sus campañas de marketing.



Ilustración 52 Cantidad de Pasaje Vendido por Mes

Fuente: Elaboración Propia

En el presente grafico podemos observar la cantidad que la agencia que tiene un mayor ingreso es de Lima a comparación de Barranca. Por ende, gracias a este grafico la empresa puede tomar acciones correctivas e indagar la disminución de sus ventas y poder ejecutar medidas correctivas.

DISTRIBUCION POR TERMINAL

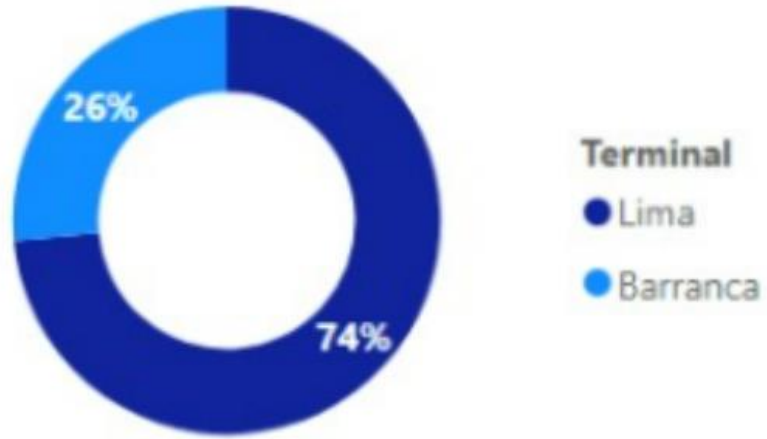


Ilustración 53 Venta por Agencia

Fuente: Elaboración Propia

En el presente grafico podemos observar el ingreso mensual por rango de edad teniendo consigo que las personas de un rango 20 a 35 años cuenta con una mayor demanda y con ello podrían empezar a organizar campañas personalizadas para las demás edades con la finalidad que puedan aumentar los ingresos.

Rango : INGRESOS MENSUAL POR RANGO DE EDAD

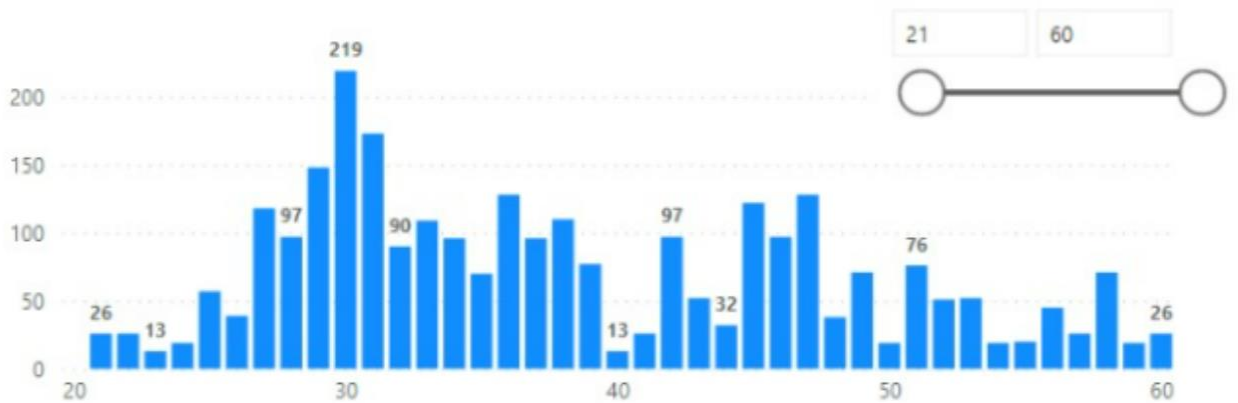


Ilustración 54 Ingreso Mensual por Rango de Edad

Fuente: Elaboración Propia

7. CONCLUSIONES

Frente a la evidencia recopilada en la primera fase del proyecto, se detectó que la principal falencia que abordaba el negocio era la demora en la entrega de información para monitorear el comportamiento de las ventas periódicas, ya que estas actividades de trabajo se realizaban de manera repetitiva y se invertían mucho tiempo para reprocesar la información. Por tal motivo, se ha desarrollado una solución de Inteligencia de negocio y pruebas de concepto con información de ventas de pasajes que simulen las ventas realizadas en el año 2022. Asimismo, se han generado métricas para medir el índice de ventas mediante una segmentación de edad, terminal y las ventas periódicas del año.

Por otro lado, se emplearon herramientas open source y servicios Cloud con el propósito de no realizar grandes inversiones en una infraestructura de TI física, además permita la escalabilidad en el tiempo y cuente con una disponibilidad de los servicios 24/7. A su vez para generar una automatización de las tareas programadas se optó en emplear Power Automate el cual cumpla la función de recibir alertas inmediatas al momento de realizar la carga de datos. Se utilizó AWS S3 como repositorio de los reportes de la organización y Amazon AWS Redshift para el almacenamiento de los datos transformados en un data warehouse.

En lo cual se concluye que la propuesta del proyecto fue aplicar una solución de BI en la empresa del sector de transporte trayendo como mejora la monitorización de la duplicidad de la información con el fin de evitarla y con ello mejorar en el proceso de toma de decisiones, ya que, al tener un proceso unificado y automatizado les permite el reordenamiento de la información pudiendo evitar un retrabajo y demora en el procesamiento de los datos. Por último, la solución les permite poder visualizar como la empresa ha ido creciendo en un periodo determinado, enfocarse en su público objetivo mediante el índice de ventas.

Recomendaciones

En la presente sección se indican las recomendaciones detectadas durante el desarrollo del presente proyecto.

Se sugiere que antes de realizar la solución se debe analizar las necesidades del negocio y verificar si se cuenta con un conjunto de datos para el procesamiento con la finalidad de brindar una mejor experiencia al usuario, ya que con ello la solución podrá brindarles una mejora toma de decisiones mediante los reportes en PowerBI en base a la medición del índice de ventas.

Asimismo, como principal recomendación sería mantener una continuidad del proceso de carga de datos se le indico a la organización subir su reporte a más tardar todos los lunes a las 8am, para que pueda ejecutarse el proceso automatizado.

Por último, se aconseja a la organización contar con una persona especializada que se encargue en la monitorización de las tareas programadas y evidenciar que la información se encuentre alojada en los repositorios de archivos.

Anexo

Carta de Aceptación

De nuestra consideración:

Por medio de la presente, tenemos el agrado de dirigimos a Ustedes, a fin de informarles sobre la solicitud para el uso de información de mi representada requerida por vuestro (a) alumno / egresado (a) **Ricardo Cadillo y Emmanuel Roque** para el desarrollo de su Tesis [] del Trabajo de Suficiencia Profesional [] Trabajo de Investigación [].

Al respecto, de manera expresa autorizamos que dicha información pase a ser de carácter pública dentro de los fines académicos que son propios de la naturaleza de este tipo de trabajos, entre los cuales está su publicación, una vez concluido el mismo, en el repositorio de la Universidad Peruana de Ciencias Aplicadas.

Sin otro particular, nos despedimos de Ustedes, expresándole las muestras de nuestra mayor consideración.

Atentamente,

NOMBRE DEL TRABAJO

**Memoria Emmanuel Roque y Ricardo Ca
dillo.docx**

AUTOR

Emmanuel Roque

RECuento DE PALABRAS

25427 Words

RECuento DE CARACTERES

138274 Characters

RECuento DE PÁGINAS

125 Pages

TAMAÑO DEL ARCHIVO

8.7MB

FECHA DE ENTREGA

Sep 5, 2022 5:48 PM GMT-5

FECHA DEL INFORME

Sep 5, 2022 5:59 PM GMT-5**● 8% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos

- 8% Base de datos de Internet
- 0% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 2% Base de datos de trabajos entregados

● Excluir del Reporte de Similitud

- Material bibliográfico
- Coincidencia baja (menos de 20 palabras)
- Bloques de texto excluidos manualmente

BIBLIOGRAFÍA

- Vittoria Biagi, Riccardo Patriarca and Giulio Di Gravio (2021). Business Intelligence for IT Governance of a Technology Company. *Data MDPI AG*. 12 de mayo de 2022.
- Andrea Delgado, Daniel Calegari, Alexis Artus, Andres Borges (2021). Integration of business process data and organizational data for evidence-based business intelligence. *CLEI electronic journal*. 12 de mayo de 2022.
- Maria Holmlunda, Yves Van Vaerenberghb , Robert Ciuchitaa , Annika Ravaldc , (2020). Customer experience management in the age of big data analytics: A strategic framework. *Journal of Business Research*, 12 de mayo de 2022.
- Definition of Dashboards - Gartner Information Technology Glossary. (s/f). Recuperado el 24 de octubre de 2020, de <https://www.gartner.com/en/information-technology/glossary/dashboard>
- Definition of Data Warehouse - Gartner Information Technology Glossary. (s/f). Recuperado el 24 de octubre de 2020, de <https://www.gartner.com/en/information-technology/glossary/data-warehouse>. 12 de mayo de 2022.
- Yanfang Niu, Limeng Ying, Jie (2021). Organizational business intelligence and decision making using big data analytics *Information Processing and Management*. 12 de mayo de 2022.
- Dalia Susa Vucec, Vesna Bosilj Vuksic (2020). Business intelligence and organizational performance The role of alignment with business process management. *Business Process Management Journal* . 12 de mayo de 2022.
- Kimball, R., & Caserta, J. (2015). The Data Warehouse ETL Toolkit. En *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference*. 12 de mayo de 2022.
- Rocio Rodríguez, Goran € Svensson , (2020). Digitalization process of complex B2B sales processes – Technology in Society . *On Smart Technology for Smart Nation, SmartTechCon 2017*. 12 de mayo de 2022.
- Manel Souibguia, Faten Atigui (2019). Data quality in ETL process: A preliminary study. Manel Souibguia, Faten Atigui. 12 de mayo de 2022.
- Leite, N., Pedrosa, I., & Bernardino, J. (2019). Open Source Business Intelligence on a SME: A Case Study using Pentaho. *Iberian Conference on Information Systems and*

Technologies, 12 de mayo de 2022.

Montalvo-Garcia, J., Quintero, J. B., & Manrique-Losada, B. (2020). Crisp-dm/smes: A data analytics methodology for non-profit smes. *Advances in Intelligent Systems and Computing*. 12 de mayo de 2022.

Pentaho Data Integration. (s/f). Recuperado el 15 de Junio de 2022, de <https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-data-integration.html>

Ilko Kovacic, Christoph G. Schuetz , (2022). OLAP Patterns: A pattern-based approach to multidimensional data analysis - *Data & Knowledge Engineering* 12 de mayo de 2022.

Usarat Thirathon, Bernhard Wieder, Zoltan Matolcsy, Maria-Luise Ossimitz . (2017). Big Data, Analytic Culture and Analytic-Based Decision Making – Evidence from Australia. *International Conference on ENTERprise Information Systems*. 12 de mayo de 2022.

Wang, C. S., Lin, S. L., Chou, T. H., & Li, B. Y. (2019). An integrated data analytics process to optimize data governance of non-profit organization. *Computers in Human Behavior*. 12 de mayo de 2022.

The Open Source Definition | Open Source Initiative. (s/f). Recuperado el 25 de octubre de 2020, de <https://opensource.org/osd>

João R. Almeida, Leonardo Coelho a, José L. Oliveira a. (2021). BICenter: A collaborative Web ETL solution based on a reflective software approach. *Journal Elsevier*. 12 de mayo de 2022.

Subia Saif, Samar Wazir (2018) Performance Analysis of Big Data and Cloud Computing Techniques: A Survey *International Conference on Computational Intelligence and Data Science*. 12 de mayo de 2022.

Song ZhangXiaochuan LuoEugene Litvinov (2021) Serverless computing for cloud-based power grid emergency generation. *Dispatch Electrical Power and Energy Systems Journal Elsevier*. 12 de mayo de 2022.

Ricky Akbar, Meza Silvana, Mohammad Hafiz Hersyah (2020). Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores. *International Conference on Information Technology Systems and Innovation*. 12 de mayo de 2022.