

UNIVERSIDAD NACIONAL DEL ALTIPLANO
**FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA,
ELECTRÓNICA Y SISTEMAS**
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS

**“SISTEMA DE SOPORTE DE DECISIONES CON TECNOLOGÍA
DATA WAREHOUSE PARA LA GESTIÓN DE LA INFORMACIÓN
DE LA EMPRESA MALLKU IMPORT SAC - JULIACA 2016”**

**PRESENTADO POR:
RONALD NILS GUILLÉN QUISCA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE: INGENIERO DE
SISTEMAS**

Puno – Perú

2017

UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA,
ELECTRÓNICA Y SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS

“SISTEMA DE SOPORTE DE DECISIONES CON TECNOLOGIA
DATA WAREHOUSE PARA LA GESTIÓN DE LA INFORMACIÓN
DE LA EMPRESA MALLKU IMPORT SAC - JULIACA 2016”

PRESENTADO POR:

RONALD NILS GUILLÉN QUISCA

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS



PUNO, PERÚ

2017

Universidad Nacional del Altiplano

FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA, ELECTRÓNICA
Y SISTEMAS

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

“SISTEMA DE SOPORTE DE DECISIONES CON TECNOLOGÍA DATA
WAREHOUSE PARA LA GESTIÓN DE LA INFORMACIÓN DE LA
EMPRESA MALLKU IMPORT SAC - JULIACA 2016”

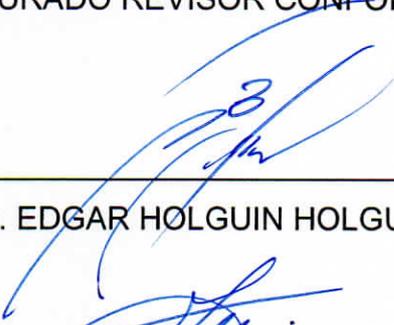
TESIS PRESENTADA POR:

RONALD NILS GUILLÉN QUISCA

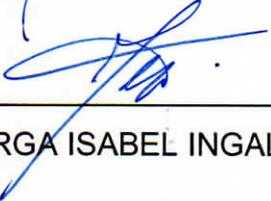
PARA OPTAR EL TÍTULO PROFESIONAL DE: INGENIERO DE SISTEMAS

APROBADA POR EL JURADO REVISOR CONFORMADO POR:

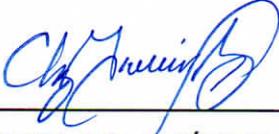
PRESIDENTE

: 
M.Sc. EDGAR HOLGUIN HOLGUIN

PRIMER MIEMBRO

: 
M.Sc. MARGA ISABEL INGALUQUE ARAPA

SEGUNDO MIEMBRO

: 
M.Sc. ADOLFO CARLOS JIMÉNEZ CHURA

ASESOR DE TESIS

: 
Mg. ELMER COYLA IDME

Puno - Perú

2017

Área: Sistemas de Información
Tema: Sistemas administrativo y de gestión

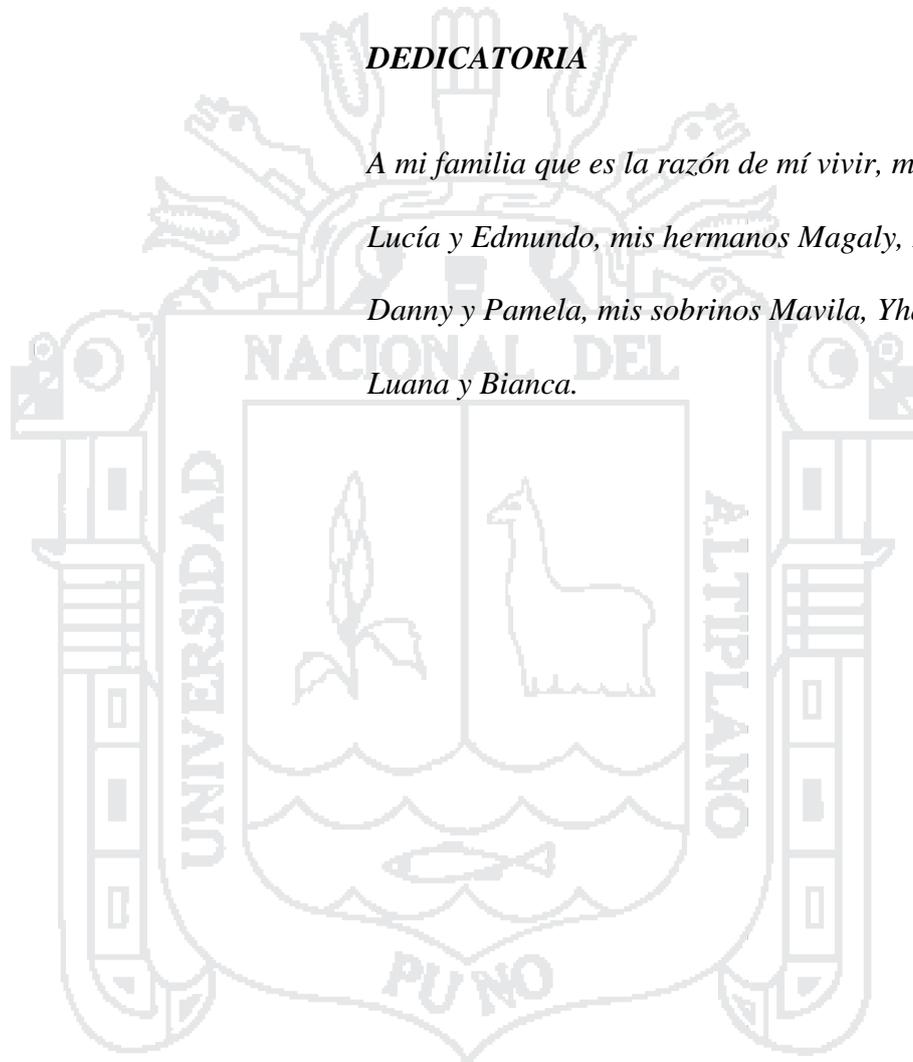
DEDICATORIA

A mi familia que es la razón de mí vivir, mis padres

Lucía y Edmundo, mis hermanos Magaly, Lisbeth,

Danny y Pamela, mis sobrinos Mavila, Yhanfranco,

Luana y Bianca.



AGRADECIMIENTOS

En primer lugar a Dios por guiar mi camino cada día de mi vida.

A mi hermana Lisbeth por iluminarme el camino a seguir con su esfuerzo y perseverancia.

A mi madre y mi padre por su gran corazón.

A la Universidad Nacional del Altiplano, gran alma mater.

A los docentes de la Escuela Profesional de Ingeniería de Sistemas quienes aportaron en mi formación profesional.

INDICE

RESUMEN	13
ABSTRACT.....	14
CAPITULO I.....	15
INTRODUCCIÓN.....	16
CAPITULO II.....	23
REVISIÓN DE LA LITERATURA	24
2.1 ANTECEDENTES DE LA INVESTIGACIÓN	24
2.2 BASE TEORICA	28
2.2.1. TIPOS DE SISTEMAS	28
2.2.2. SISTEMAS DE SOPORTE DE DECISIONES (DSS).....	29
2.2.3. INTELIGENCIA DE NEGOCIOS	31
2.2.4. GESTIÓN DE LA INFORMACIÓN.....	35
2.2.5. DATA WAREHOUSE	36
2.2.6. DATAMART.....	48
2.2.7. ETL (Extracción, Transformación y Carga).	49
2.2.8. OLAP (Procesamiento Analítico en Línea).	50
2.2.9. HERRAMIENTAS PARA IMPLEMENTAR INTELIGENCIA DE NEGOCIOS.....	57
2.2.10. PENTAHO.....	60
2.2.11. MYSQL.....	72
2.2.12. OTROS TÉRMINOS FUNDAMENTALES	74

2.3	MARCO CONCEPTUAL.....	77
CAPITULO III.....		83
MATERIALES Y MÉTODOS.....		83
3.1	METODOLOGÍA DE LA INVESTIGACIÓN.....	83
3.2	DISEÑO DE LA INVESTIGACIÓN.....	83
3.3	POBLACIÓN Y MUESTRA.....	84
3.3.1.	Población.....	84
3.3.2.	Muestra.....	84
3.4	MÉTODOS DE RECOPIACIÓN DE DATOS.....	85
3.4.1.	La entrevista.....	85
3.4.2.	Encuestas.....	85
3.5	METODO DE TRATAMIENTO DE DATOS.....	86
3.6	MATERIAL EXPERIMENTAL.....	86
3.6.1.	Metodología de desarrollo del Sistema.....	86
3.6.2.	Etapas de ingeniería y pasos de desarrollo.....	89
3.6.3.	Hardware y Software de Desarrollo.....	97
CAPITULO IV.....		100
RESULTADOS Y DISCUSIÓN.....		100
4.1.	APLICACIÓN DE LA METODOLOGÍA.....	100
4.1.1.	JUSTIFICACIÓN.....	100
4.1.2.	PLANIFICACIÓN.....	101
4.1.3.	ANÁLISIS DEL NEGOCIO.....	105

4.1.4. DISEÑO	116
4.1.5. CONSTRUCCIÓN.....	136
4.1.6. DESPLIEGUE	141
4.2. PRUEBA DE HIPÓTESIS	141
4.2.1. Evaluación del Sistema por la Empresa.....	141
4.2.2. Planteamiento de la hipótesis.....	146
4.2.3. Nivel de significancia.....	146
4.2.4. Zona de rechazo	146
4.2.5. Estadístico de prueba	147
4.2.6. Decisión	147
4.2.7. Resultado de la prueba de hipótesis	147
CAPITULO V.....	149
CONCLUSIONES.....	149
CAPITULO VI.....	152
RECOMENDACIONES.....	152
CAPITULO VII.....	154
REFERENCIAS.....	154
ANEXOS	157

INDICE DE FIGURAS

Figura N° 01. Tipos de sistemas de información.....	29
Figura N° 02. Arquitectura de almacenamiento de datos.....	40
Figura N° 03. Arquitectura Innom.....	41
Figura N° 04. Arquitectura Inmon - 3FN.....	42
Figura N° 05. Arquitectura Kimball - modelo dimensional.....	44
Figura N° 06. Arquitectura Kimball.....	45
Figura N° 07. Procesos ETL.....	50
Figura N° 08. Cubo OLAP.....	52
Figura N° 09. Clasificación de Pentaho.....	61
Figura N° 10. Plataforma Pentaho.....	62
Figura N° 11. Mondriam.....	63
Figura N° 12. Pentaho Reporting.....	65
Figura N° 13. Data Integration.....	67
Figura N° 14. Data Mining.....	68
Figura N° 15. Capa de metadatos.....	70
Figura N° 16. Arquitectura personalizada de Pentaho.....	72
Figura N° 17. Metodología utilizada para la implementación de BI.....	87
Figura N° 18. Etapas de ingeniería.....	88
Figura N° 19. Dependencias en los pasos de desarrollo.....	96
Figura N° 20. Actividades de planificación de proyectos.....	103
Figura N° 21. Modelo conceptual de ventas.....	108
Figura N° 22. Modelo conceptual compras.....	110
Figura N° 23. Datamart de ventas.....	111
Figura N° 24. Datamart de compras.....	112

Figura N° 25. Prototipo de datamart de ventas.....	113
Figura N° 26. Prototipo de datamarts de compras.....	114
Figura N° 27. Diseño de la dimensión cliente para ventas.....	118
Figura N° 28. Diseño de la dimensión producto para ventas.....	118
Figura N° 29. Diseño de la dimensión fecha para ventas.....	119
Figura N° 30. Diseño de la dimensión proveedor para compras.....	119
Figura N° 31. Diseño de la dimensión producto para compras.....	120
Figura N° 32. Diseño de la dimensión fecha para compras.....	120
Figura N° 33. Tabla de hechos para ventas.....	121
Figura N° 34. Tabla de hechos para compras.....	121
Figura N° 35. Unión de tablas para ventas.....	122
Figura N° 36. Unión de tablas para compras.....	122
Figura N° 37. Creación de un nuevo trabajo.....	137
Figura N° 38. Diseño de los procesos que tendrá el ETL.....	137
Figura N° 39. Carga de dimensiones.....	138
Figura N° 40. Carga del origen de datos a la tabla destino.....	138
Figura N° 41. Carga en las tablas hechos de ventas y compras.....	139
Figura N° 42. Configuración de esquema en OLAP.....	140
Figura N° 43. Zona de rechazo.....	146

INDICE DE TABLAS

Tabla N° 01. Operacionalización de variables.....	22
Tabla N° 02. Herramientas de ETL.....	59
Tabla N° 03. Herramientas de explotación.....	60
Tabla N° 04. Pasos específicos del proyecto versus pasos de toda la organización.....	97
Tabla N° 05. Características de hardware y software de la empresa.....	101
Tabla N° 06. Actividades de planificación del proyecto.....	104
Tabla N° 07. Incidencias del área de ventas.....	107
Tabla N° 08. Incidencias del área de compras.....	109
Tabla N° 09. Tabla fuente dimensión cliente para ventas.....	123
Tabla N° 010. Tabla estandarización de datos de cliente para ventas.....	123
Tabla N° 011. Tabla fuente de datos de cliente para ventas.....	124
Tabla N° 012. Tabla destino de cliente para ventas.....	124
Tabla N° 013. Tabla fuente dimensión producto para ventas.....	125
Tabla N° 014. Tabla estandarización de datos de producto para ventas.....	125
Tabla N° 015. Tabla fuente de datos de producto para ventas.....	126
Tabla N° 016. Tabla destino de producto para ventas.....	126
Tabla N° 017. Tabla fuente dimensión fecha para ventas.....	127
Tabla N° 018. Tabla estandarización de datos de Fecha para Ventas.....	127
Tabla N° 019. Tabla fuente de datos de fecha para ventas.....	128
Tabla N° 020. Tabla destino de fecha para ventas.....	128
Tabla N° 021. Tabla fuente dimensión proveedor para compras.....	129
Tabla N° 022. Tabla estandarización de datos de proveedor para compras.....	129
Tabla N° 023. Tabla fuente de datos de proveedor para compras.....	130

Tabla N° 024. Tabla destino de proveedor para compras.....	130
Tabla N° 025. Tabla fuente dimensión producto para compras.....	131
Tabla N° 026. Tabla estandarización de datos de producto para compras.....	131
Tabla N° 027. Tabla fuente de datos de producto para compras.....	132
Tabla N° 028. Tabla destino de producto para compras.....	132
Tabla N° 029. Tabla fuente dimensión fecha para compras.....	133
Tabla N° 030. Tabla estandarización de datos de fecha para compras.....	133
Tabla N° 031. Tabla fuente de datos de fecha para compras.....	134
Tabla N° 032. Tabla destino de fecha para compras.....	134
Tabla N° 033. Datamart ventas.....	135
Tabla N° 034. Tabla E – R de ventas.....	135
Tabla N° 035. Datamart compras.....	136
Tabla N° 036. Tabla E – R de compras.....	136
Tabla N° 037. Resultado de opinión con respecto al diseño de interfaz.....	142
Tabla N° 038. Resultado de opinión con respecto al manejo del sistema.....	142
Tabla N° 039. Resultado de opinión con respecto al mantenimiento y actualización de los procesos ETL.....	143
Tabla N° 040. Resultado de opinión con respecto a la creación de consultas y cuadros de dimensiones OLAP.....	143
Tabla N° 041. Resultado de opinión con respecto a la disponibilidad de la información.....	144
Tabla N° 042. Resultado de opinión con respecto a la aceptación del sistema....	144
Tabla N° 043. Cuadro comparativo del antes y después del sistema.....	145

INDICE DE ANEXOS

Anexo N° 01. Cuestionario de preguntas sobre los objetivos de negocio y similares..... 158

Anexo N° 02. Instalación de MySQL Server..... 159

Anexo N° 03. Instalación y configuración de Pentaho Data Integration..... 162

Anexo N° 04. Conexión de Data Integration a MySQL..... 163

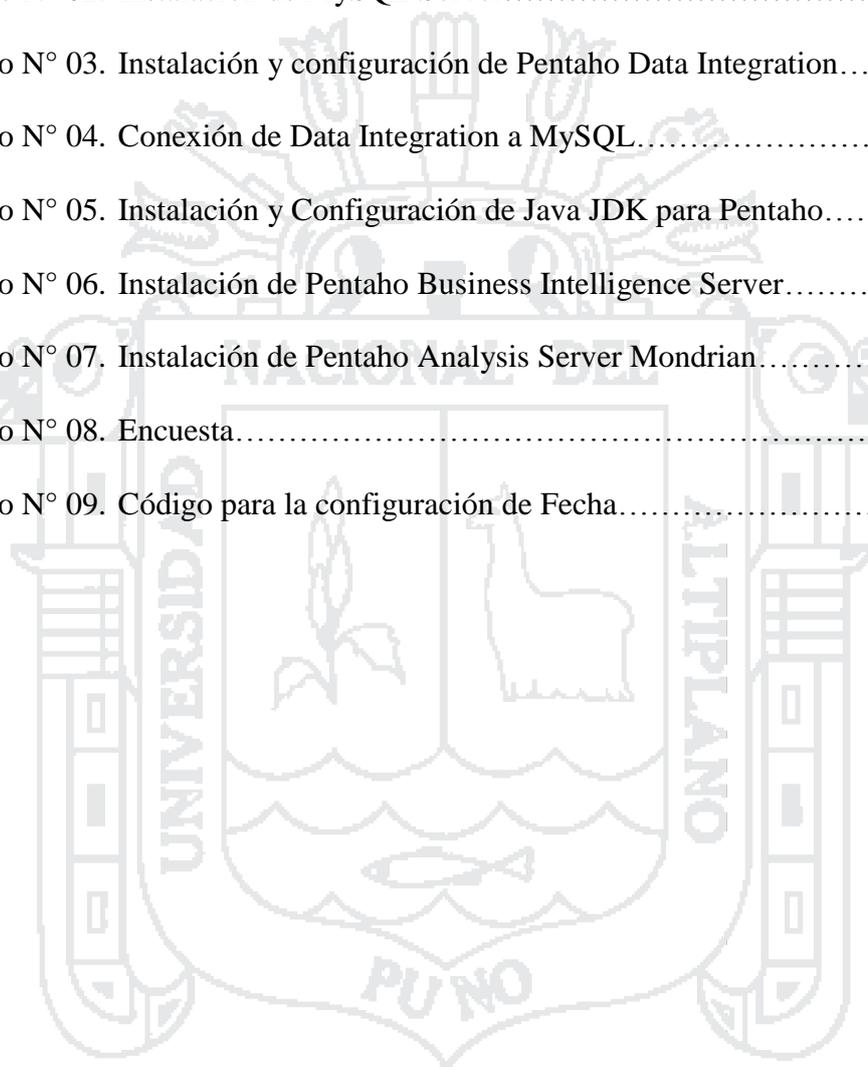
Anexo N° 05. Instalación y Configuración de Java JDK para Pentaho..... 165

Anexo N° 06. Instalación de Pentaho Business Intelligence Server..... 167

Anexo N° 07. Instalación de Pentaho Analysis Server Mondrian..... 169

Anexo N° 08. Encuesta..... 170

Anexo N° 09. Código para la configuración de Fecha..... 171



RESUMEN

La presente investigación tiene por finalidad el desarrollo de un Sistema de Soporte de Decisiones con Tecnología Data Warehouse que optimice la Gestión de la Información de la Empresa Mallku Import SAC, la empresa en mención tiene su centro de operaciones en la ciudad de Juliaca, región de Puno y se dedica al rubro de la importación de máquinas remanufacturadas como son fotocopiadoras, impresoras, repuestos, insumos y accesorios; donde sus operaciones diarias son la venta de estos productos. Encontrando como uno de los principales problemas la mala gestión de la información, el desaprovechamiento de información almacenada a diario, no generando información a partir de esto. Por lo cual se plantea un sistema de soporte de decisiones basado en inteligencia de negocios que tendrá como fin solucionar este problema. Para el desarrollo del software se utilizó la metodología de Moss Larissa que es basada en Inteligencia de Negocios de Roadmap y el enfoque de Ralph Kimball para la construcción del Data Warehouse y como herramientas de software se utilizó la suite de Inteligencia de Negocios “Pentaho” que viene a ser un conjunto de programas libres para generar inteligencia empresarial escrita en Java con un ambiente de implementación basado en Java, para evitar costos altísimos en las licencias de software. Finalmente se tiene como principal conclusión, que el sistema desarrollado ha optimizado la gestión de la información en un más del 83%, logrando validar y cumplir con el objetivo principal de la investigación.

Palabras Claves: Data Warehouse, Gestión de la Información, Inteligencia de Negocios y Soporte de Decisiones.

ABSTRACT

This research aims to develop a Decision Support System with Data Warehouse Technology that optimizes the Information Management of the Mallku Import SAC company, the company in question has its operations center in the city of Juliaca, region of Puno and is dedicated to the import of remanufactured machines such as photocopiers, printers, spare parts, supplies and accessories; Where their daily operations are the sale of these products. Finding as one of the main problems the mismanagement of information, the waste of information stored daily, not generating information from this. Therefore, a decision support system based on business intelligence is proposed, which will solve this problem. For the development of the software we used the methodology of Moss Larissa which is based on Roadmap Business Intelligence and Ralph Kimball's approach to building the Data Warehouse and as software tools we used the "Pentaho" Business Intelligence suite that Comes to be a set of free programs to generate business intelligence written in Java with a Java-based deployment environment, to avoid skyrocketing software licensing costs. Finally, the main conclusion is that the developed system has optimized the management of information by more than 83%, and has been able to validate and fulfill the main objective of the research.

Keywords: Business Intelligence, Data Warehouse, Decision Support and Information Management.

CAPÍTULO I



INTRODUCCIÓN

El presente trabajo de investigación contribuye al desarrollo de las MyPES de la región Puno como es el caso de la empresa Mallku Import SAC, una empresa creciente y consciente que para aumentar su competitividad y sostenibilidad en el mercado, necesita el uso de tecnologías para la toma de decisiones, la investigación plantea el desarrollo de un Sistema de Soporte de Decisiones con Tecnología Data Warehouse para la Gestión de la Información, para que mediante su uso las empresas puedan organizar su información y procesarla con las herramientas adecuadas para la toma de decisiones en cuanto al control eficiente de los productos, generación de reportes de ventas y registro de clientes para seguimiento post-venta, los cuales sin un control adecuado generan insatisfacción y pérdida de clientes.

En el mercado global y altamente competitivo de nuestros días, las compañías que sobresalen son aquellas que tienen un interés continuo por identificar cuáles son los factores más importantes para sus clientes, se enfocan en ellos y mejoran sus procesos para ofrecer el producto o servicio con la más alta calidad posible (Summers, 2006). Gracias a las nuevas tecnologías esto es posible teniendo la información en el momento y en forma oportuna. Las empresas importadoras tienen un gran mercado a nivel nacional, al igual que en la ciudad de Juliaca, que es conocida como el eje comercial del sur del país; por lo cual existen diversos negocios que se dedican a la importación trayendo una serie de productos los cuales por el bajo costo económico son adquiridos por la población, pero sin embargo muchas veces los productos no son rentables por lo cual son rematados o simplemente quedan en un almacén ocasionando su deterioro y pérdidas económicas lo cual será parte de este estudio. La empresa Mallku Import SAC, tiene su centro de operaciones en la ciudad de Juliaca, región de Puno y se dedica al rubro de la importación de máquinas remanufacturadas como son fotocopiadoras,

impresoras, repuestos, insumos y accesorios; donde sus principales proveedores son empresas de la ciudad de Lima y del país de los Estados Unidos, teniendo como marcas principales: Konica Minolta, Kyocera y Ricoh. Y entre sus operaciones diarias esta la venta de estos productos, como también el de brindar a sus clientes un servicio post-venta, siendo este servicio una garantía que da la empresa, como es el mantenimiento de las máquinas y/o la venta de algún repuesto o insumo que siempre debe estar disponible a los clientes y estos puedan seguir utilizando tanto los servicios y productos de la empresa. Uno de los objetivos de la empresa es que la población pueda adquirir estos productos para realizar sus actividades y operaciones diarias que beneficien a sus empresas o instituciones en las funciones de fotocopiado, impresión, escaneo entre otros y lo puedan hacer en forma mucho más económica ya que los productos nuevos cuestan más que el triple y para muchos es difícil adquirirlos. La empresa ya tiene más de dos años funcionando y no ha podido tener un crecimiento y posicionamiento en el mercado, viviendo día a día la incertidumbre y la rutina que ya están acostumbrados, el ingreso y salida de productos sin un control eficiente, sin poder tener un análisis del desarrollo de cuanto están avanzando ya que los informes muchas veces son tediosos de realizar por parte de los directivos y el personal encargado de la empresa, también se ve muchas veces la improvisación, como por ejemplo es el caso de los productos que se agotaron y son requeridos por los clientes, los cuales se compran a destiempo generando insatisfacción y pérdida de clientes. Donde se ve claramente una deficiente gestión de la información. Las operaciones diarias las registran en formatos de Excel y un pequeño sistema de ventas almacenando sus datos en Microsoft Acces, los cuales serán de gran beneficio para poder dar una solución de Inteligencia de Negocios que permita a la empresa poder cumplir a satisfacción sus metas y objetivos.

Según lo descrito anteriormente se ha llegado a plantear la siguiente pregunta

general de estudio:

¿En qué medida el desarrollo del Sistema de Soporte de Decisiones con Tecnología Data Warehouse optimizará la Gestión de la Información de la Empresa Mallku Import SAC - Juliaca 2016?

Y como preguntas del problema específicos se tiene:

- ¿En qué medida el análisis de los requerimientos del negocio permiten el diseño y construcción del modelo del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016?
- ¿En qué medida el diseño y construcción permiten la implementación y evaluación del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016?
- ¿En qué medida la implementación y evaluación permiten el desarrollo del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016?
- ¿En qué medida la validación del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016, logrará un resultado satisfactorio?

El proyecto de investigación es justificado de la siguiente manera:

La aplicación de las TICs en las empresas locales son poco conocidas aún más en las MYPES, donde las operaciones diarias las registran en documentos de Word o Excel

y en algunos casos en cuadernos. Son pocos los que cuentan con Sistemas Operacionales, y prácticamente ninguno con sistemas más avanzados, pero si hay mayor uso de tecnologías en cuanto a hardware y diferentes dispositivos teniendo por lo menos un par de equipos de cómputo y con servicio de internet, también se ve que existe mucha informalidad aun por parte de las empresas, que temen ponerse en la formalidad por evadir impuestos y sus negocios sean menos rentables, y los que entraron al proceso de la formalización como es el caso de esta empresa están de alguna manera sobreviviendo y llegando ajustadamente cada mes a cumplir no plenamente con sus objetivos, ni mostrando el crecimiento que como toda empresa anhela. En nuestro país actualmente existen varias empresas que se dedican a la importación de diversos productos, el cual ha venido incrementándose en los últimos años en forma acelerada debido a los TLC con diferentes países, haciendo que estos sean negocios rentables y al mismo tiempo generando mayor competitividad a lo cual no es ajeno la región de Puno donde se encuentra la Empresa Mallku Import SAC, exactamente en la ciudad de Juliaca. Como se dijo anteriormente que debido a la creciente competencia en el mercado como es la importación de máquinas en este caso fotocopiadoras, impresoras, repuestos, insumos, accesorios, y brindar un servicio de calidad y satisfacción al cliente no solo se requiere tener los mejores productos del mercado, el mejor personal calificado ya sea para el área de ventas, mantenimiento u otro, sino que para poder competir y tener un crecimiento mucho mayor se debe tener a conciencia de que existe un capital mucho mayor en la empresa que poco se da importancia, como son las operaciones diarias tal es el caso de la información que a diario se junta, y dar mayor provecho a esa información que se encuentra en los distintos registros. También existe en las MyPEs temor sobre los costos de las tecnologías de información y más si son avanzados como es el caso de este sistema que brinda una solución de Inteligencia de

Negocios, por lo que muchos no se atreven a investigar o indagar sobre el desarrollo de software por ser calificados de altísimos, no teniendo un presupuesto para solventar la construcción de ellos. La presente investigación propone la creación de un Sistema de Soporte de Decisiones con Tecnología Data Warehouse que permita optimizar la Gestión de la Información de la Empresa Mallku Import SAC, de la ciudad de Juliaca, brindar información completa, correcta, etc., haciendo uso de las herramientas de software libre como es la suite de Pentaho. Como también de concientizar a los encargados de la empresa quienes serán parte de esta investigación para que este tenga éxito y puedan colaborar con la información que se requiera.

Se tiene como objetivo principal de la investigación:

Desarrollar un Sistema de Soporte de Decisiones con Tecnología Data Warehouse que optimice la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.

Y como objetivos específicos tenemos los siguientes:

- Analizar los requerimientos del negocio, para diseñar y construir el Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.
- Diseñar y construir el modelo para implementar y evaluar el Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.
- Implementar y evaluar el Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.

- Validar el Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016, logrando un resultado satisfactorio.

Como hipótesis general tenemos:

El Sistema de Soporte de Decisiones con Tecnología Data Warehouse optimiza la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.

Y como hipótesis específicas las siguientes:

- El análisis de los requerimientos del negocio permiten el diseño y construcción del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.
- El diseño y la construcción permiten la implementación y evaluación del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.
- La implementación y evaluación permiten validar el Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.
- La validación del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016, muestra un resultado satisfactorio.

Tabla N° 01: Operacionalización de variables.

VARIABLES	DIMENSIONES	INDICADORES	ITEMS
1. INDEPENDIENTE Sistema de Soporte de Decisiones con Tecnología Data Warehouse	Análisis de requerimientos del negocio.	Satisfacción de los requerimientos.	N° de requerimientos identificados. N° de requerimientos desarrollados satisfactoriamente.
	Diseño y construcción del sistema.	Estética	Grado de atracción.
		Usabilidad	Facilidad de operación.
		Funcional	Facilidad de navegación.
2. DEPENDIENTE Gestión de la Información de la empresa Mallku Import SAC – Juliaca 2016	Implementación y evaluación del sistema.	Eficiencia para el mantenimiento y actualización de los procesos ETL.	Facilidad de aprender
		Eficiencia al crear consultas y cuadros dimensionales OLAP.	Nivel de precisión.
		Disponibilidad de la Información generada por el sistema.	Grado de satisfacción.
	Validación del sistema	Aceptación del sistema por el usuario.	- N° de usuarios satisfechos. - N° de usuarios no satisfechos.

Elaboración: Propia.

CAPITULO II



REVISIÓN DE LA LITERATURA

2.1 ANTECEDENTES DE LA INVESTIGACIÓN

- (Chávez, 2015). **Sistema de Soporte a la Toma de Decisiones basado en Inteligencia de Negocios para mejorar los procesos Comerciales del Importador Peruano**, Tesis presentado en la Universidad Católica Santo Toribio de Mogrovejo – Chiclayo – Perú, para optar el título profesional de Ingeniero de Sistemas.

El presente proyecto, propone la implementación de un sistema de soporte a la toma de decisiones basado en inteligencia de negocios para mejorar los procesos comerciales del importador peruano, donde el principal problema radica en la información desintegrada que se encuentra en diferentes formatos almacenados y además no se le da una debida orientación a dicho importador. El proyecto tiene como objetivo principal el correcto procesamiento y limpieza de datos para que de esta manera responda a los requerimientos del importador peruano.

Principales conclusiones a las que se llega:

- El software logro reducir en 70% los tiempos para el procesamiento y ordenamiento de la información, esto permitió al importador no desistir en dicha búsqueda de información dado que para lograr dicho propósito con el sistema tradicional se tiene que buscar esta información por separado.
- Con la implementación del sistema, también se logró conocer la tasa de variación o variación porcentual de los países, por consiguiente es un indicador importante debido a que permite conocer el crecimiento o

decrecimiento en un determinado país con respecto a las importaciones realizadas.

- Se logró conocer los diferentes precios CIF dentro de los cuales se encuentran el costo del producto, el flete del producto y el seguro del producto importado; esto le permite al importador analizar y reducir sus precios CIF de importación.
- (Sánchez, 2014) **Análisis de Información y Toma de decisiones para Administración de Negocios**. Tesis presentada en la Universidad Nacional Autónoma de México. Para optar el título profesional de Ingeniero en Computación.

La presente tesis tiene como objetivo mejorar la visión del negocio, en específico del área de Ventas, por medio del análisis de información, además de detectar por medio de indicadores la eficiencia del área. Esto es importante para la toma de decisiones en las empresas, ya que hoy día, pueden presentar pérdidas considerables por no interpretar de forma acertada las enormes cantidades de datos que generan, los cuales deberían de ser transformados en información relevante que ayude a mejorar la efectividad de la institución.

Principales conclusiones a las que se llega:

- La presente tesis expuso la implementación de un sistema de Business Intelligence, la cual puede ser aplicada a cualquier tipo de organización, tomando como sistema fuente un Data Warehouse montado en una base de Datos Oracle 11.2.0.1.0 y una herramienta de análisis Oracle Business Intelligence 11.1.1.6.0, el cual está basado en un Data Mart de Ventas. El sistema demostró que, mediante un Dashboard, puede determinar la

tendencia de la empresa con respecto al tiempo, y si esta va en aumento en los últimos años.

- Se logró tener un sistema confiable que brinda la información necesaria para la toma de decisiones. El enfoque dado está dirigido al área de ventas pero no obstante se puede implementar la misma para cualquier área en la que se requiera hacer análisis.
- El aprendizaje obtenido consta de poder ver los datos de otra manera, transformarlos en información que sea útil y proporcionarla en el momento que sea necesaria, quienes toman decisiones en las organizaciones hoy en día han de tener, si quieren obtener ventajas competitivas, un acceso rápido y fácil a la información estratégica del negocio.
- (Moreno, 2013). **Análisis, Diseño e Implementación de Datamarts para las áreas de Ventas y Recursos Humanos de una Empresa dedicada a la exportación e importación de Productos Alimenticios.** Tesis presentada en la Pontificia Universidad Católica del Perú, para optar por título profesional de Ingeniero Informático.

El presente proyecto nos muestra como estudio una empresa dedicada a la comercialización de productos alimenticios, que en los últimos años ha crecido considerablemente y ha obtenido grandes ganancias; sin embargo arrastra dos problemas claramente marcados relacionados a las áreas de ventas y recursos humanos.

Para la implementación del presente proyecto se optó por usar un híbrido de herramientas siguiendo todos los pasos que implica una solución de Inteligencia de Negocios: análisis, diseño y construcción de los

Datamarts, creación y programación de los procesos ETL, creación de los cubos, creación de los informes y finalmente la implementación de la plataforma BI (Web). Para el proceso ETL se usó la herramienta Kettle de la suite de Pentaho permitiendo de este modo ahorrar costos de licencia de software ya que esta es una herramienta que se encuentra bajo licencia GNU GPL. Para el diseño y explotación del cubo se usó la herramienta SQL Server 2008 de Microsoft (Analysis Services y Reporting Services) el cual permite explotar los reportes no sólo vía web sino también crear reportes adicionales que se necesite mediante tablas dinámicas en excel. Esta última es la herramienta con la cual los usuarios finales (personal administrativo) ya se encuentran familiarizados ya que es su herramienta del día a día; lo cual se convierte en una ventaja al momento de realizar la explotación del cubo.

Principales conclusiones a las que se llega:

- La mejor posibilidad para desarrollar el presente proyecto es el esquema de Kimball dado que no es necesario la creación de un data warehouse, simplemente se debe de extraer la data de las diferentes bases de datos existentes de la organización y con ello ir armando los Datamarts para las diferentes áreas del negocio.
- Para realizar una óptima implementación de un sistema de soporte a decisiones es muy importante las reuniones con el usuario final. Como mínimo deben de existir tres reuniones: una primera reunión donde se detalle las necesidades del usuario; una segunda reunión donde se fije las dimensiones junto con los indicadores y medidas necesarias y una tercera reunión donde se fije los prototipos de reportes a implementar.

2.2 BASE TEORICA

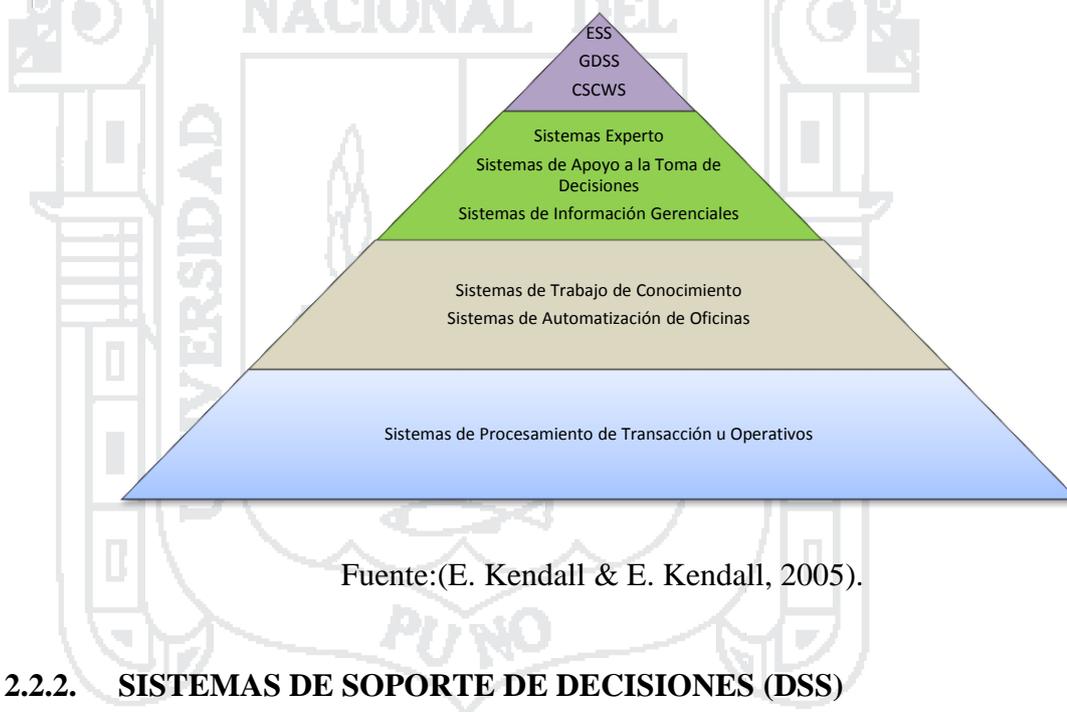
2.2.1. TIPOS DE SISTEMAS

Según (E. Kendall & E. Kendall, 2005) los sistemas de información se desarrollan con diversos propósitos, según las necesidades de la empresa. Los sistemas de procesamiento de transacciones (TPS, Transaction Processing Systems) funcionan al nivel operativo de una organización, los sistemas de automatización de la oficina (OAS, Office Automation Systems) y los sistemas de trabajo del conocimiento (KWS, Knowledge Work Systems) apoyan el trabajo al nivel del conocimiento. Los sistemas de información gerencial (MIS, Management Information Systems) y los sistemas de apoyo a la toma de decisiones (DSS, Decisión Support Systems) se encuentran entre los sistemas de alto nivel. Los sistemas expertos aplican el conocimiento de los encargados de la toma de decisiones para solucionar problemas estructurados específicos. Los sistemas de apoyo a ejecutivos (ESS, Executive Support Systems) se encuentran en el nivel estratégico de la administración. Los sistemas de apoyo a la toma de decisiones en grupo (GDSS, Group Decisión Support Systems) y los sistemas de trabajo corporativos apoyados por computadora (CSCWS, Computer-Supported Collaborative Work Systems), descritos de manera más general, auxilian la toma de decisiones semiestructuradas o no estructuradas a nivel de grupo.

En la figura N° 01 se muestra la diversidad de sistemas de información que podrían desarrollar los analistas. Observe que en la figura estos sistemas se representan de abajo hacia arriba, indicando que los TPS apoyan el nivel operativo, o más bajo, de la organización, mientras que los

ESS, GDSS y CSCWS soportan el nivel estratégico, o más alto, apoyando la toma de decisiones semiestructuradas o las no estructuradas. En este libro se emplean de manera indistinta los términos sistemas de información gerencia!, sistemas de información (IS, Information Systems), sistemas de información computarizados y sistemas de información de negocios computarizados, para denotar sistemas de información computarizados que apoyan el rango de actividades de negocios más amplio mediante la información que producen.

Figura N° 01: Tipos de sistemas de información.



2.2.2. SISTEMAS DE SOPORTE DE DECISIONES (DSS)

Los Sistemas de Soporte de decisiones (DSS, Decisión Support Systems) constituyen una clase de alto nivel de sistemas de información computarizada. Los DSS coinciden con los sistemas de información gerencial en que ambos dependen de una base de datos para abastecerse de datos. Sin embargo, difieren en que el DSS pone énfasis en el apoyo a la toma de decisiones en todas sus fases, aunque la decisión definitiva es responsabilidad exclusiva del encargado de tomarla. Los sistemas de

apoyo a la toma de decisiones se ajustan más al gusto de la persona o grupo que los utiliza que a los sistemas de información gerencial tradicionales. En ocasiones se hace referencia a ellos como sistemas que se enfocan en la inteligencia de negocios. (E. Kendall & E. Kendall, 2005).

Un Sistema de Soporte a la Decisión (DSS) es una herramienta de Business Intelligence enfocada al análisis de los datos de una organización. En principio, puede parecer que el análisis de datos es un proceso sencillo, y fácil de conseguir mediante una aplicación hecha a medida o un ERP sofisticado. Sin embargo, no es así: estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas, etc. (Sinnexus, 2007).

El DSS es una de las herramientas más emblemáticas del Business Intelligence ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de los programas de gestión. Estas son algunas de sus características principales: (Sinnexus, 2007).

- **Informes dinámicos, flexibles e interactivos**, de manera que el usuario no tenga que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
- **No requiere conocimientos técnicos**. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos, haciendo drag&drop o drill through. Por tanto, para examinar la

información disponible o crear nuevas métricas no es imprescindible buscar auxilio en el departamento de informática.

- **Rapidez en el tiempo de respuesta**, ya que la base de datos subyacente suele ser un data warehouse corporativo o un data mart, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos están optimizadas para el análisis de grandes volúmenes de información
- **Integración entre todos los sistemas/departamentos de la compañía**. El proceso de ETL previo a la implantación de un Sistema de Soporte a la Decisión garantiza la calidad y la integración de los datos entre las diferentes unidades de la empresa. Existe lo que se llama: integridad referencial absoluta.
- **Cada usuario dispone de información adecuada a su perfil**. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
- **Disponibilidad de información histórica**. En estos sistemas está a la orden del día comparar los datos actuales con información de otros períodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio.

2.2.3. INTELIGENCIA DE NEGOCIOS

Para (Howson, 2009) la inteligencia de negocios permite a las personas de todos los niveles de una organización tener acceso, interactuar y analizar información para administrar el negocio, mejorar el rendimiento, descubrir oportunidades y operar eficientemente. Sin

personas para interpretar información y actuar con base en ella, la inteligencia de negocios nada logra. Por esta razón, la inteligencia de negocios es menos sobre tecnología que creatividad, cultura e individuos considerando la información como un activo crucial. La tecnología posibilita la inteligencia de negocios, pero a veces, demasiada importancia puesta en la tecnología puede sabotear iniciativas de inteligencia de negocios. Son las personas quienes harán de los esfuerzos BI un enorme éxito o un rotundo fracaso.

Un depósito de información puede o no ser componente de la arquitectura de la inteligencia de negocios, pero depósito de información no es sinónimo de inteligencia de negocios. De hecho, aunque exista un almacén de información, solamente se puede decir que la compañía utiliza inteligencia de negocios una vez que ponga en manos de los usuarios herramientas para llegar a la información y hacerla útil.

Según (Díaz, 2010) se entiende por Inteligencia de Negocios al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización. Algunas de las tecnologías que forman parte de Business Intelligence son:

- Data warehouse.
- Reporting.
- Análisis OLAP (On-Line Analytical Processing).
- Análisis visual.
- Análisis predictivo.
- Cuadro de mando.

- Cuadro de mando integral.
- Minería de datos.
- Gestión del rendimiento.
- Previsiones.
- Reglas de negocio.
- Dashboards.
- Integración de datos (que incluye ETL, Extract, Transform and Load).

2.2.3.1. Beneficios de un sistema de inteligencia de negocio.

Según (Díaz, 2010) la implantación de estos sistemas de información proporciona diversos beneficios, entre los que podemos destacar:

- Crear un círculo virtuoso de la información (los datos se transforman en información que genera un conocimiento que permite tomar mejores decisiones que se traducen en mejores resultados y que generan nuevos datos).
- Permitir una visión única, conformada, histórica, persistente y de calidad de toda la información.
- Crear, manejar y mantener métricas, indicadores claves de rendimiento (KPI, Key Performance Indicator) e indicadores claves de metas (KGI, Key Goal Indicator) fundamentales para la empresa.
- Aportar información actualizada tanto a nivel agregado como en detalle.

- Reducir el diferencial de orientación de negocio entre el departamento TI y la organización.
- Mejorar comprensión y documentación de los sistemas de información en el contexto de una organización.
- Mejorar de la competitividad de la organización como resultado de ser capaces de:
 - a) Diferenciar lo relevante sobre lo superfluo.
 - b) Acceder más rápido a información.
 - c) Tener mayor agilidad en la toma de las decisiones.

2.2.3.2. ¿Cuándo es necesaria la inteligencia de negocios?

Según (Díaz, 2010) Existen situaciones en las que la implantación de un sistema de Inteligencia de Negocios resulta adecuada.

Destacamos, entre todas las que existen:

- La toma de decisiones se realiza de forma intuitiva en la organización.
- Identificación de problemas de calidad de información.
- Uso de Excel como repositorios de información corporativos o de usuario. Lo que se conoce como Excel caos.
- Necesidad de cruzar información de forma ágil entre departamentos.
- Evitar silos de información.
- Las campañas de marketing no son efectivas por la información base usada.

- Existe demasiada información en la organización para ser analizada de la forma habitual. Se ha alcanzado la masa crítica de datos.
- Es necesario automatizar los procesos de extracción y distribución de información.

2.2.4. GESTIÓN DE LA INFORMACIÓN

En el contexto de las organizaciones, la gestión de la información se puede identificar como la disciplina que se encargaría de todo lo relacionado con la obtención de la información adecuada, en la forma correcta, para la persona indicada, al coste adecuado, en el momento oportuno, en el lugar apropiado y articulando todas estas operaciones para el desarrollo de una acción correcta. En este contexto, los objetivos principales de la Gestión de la Información son: maximizar el valor y los beneficios derivados del uso de la información, minimizar el coste de adquisición, procesamiento y uso de la información, determinar responsabilidades para el uso efectivo, eficiente y económico de la información y asegurar un suministro continuo de la información (Pérez, 2009).

La Gestión de la Información mantiene una estrecha relación con la disciplina de la Gestión del Conocimiento en el contexto organizacional. Los objetivos de la Gestión de Información se centran en aquellos procesos relacionados con el almacenamiento, el tratamiento y la difusión del conocimiento explícito que se encuentra representado en los documentos. Sin embargo, en este contexto, la Gestión del Conocimiento iría un poco más allá que la Gestión de la Información. Ésta se encargaría

de convertir todo el conocimiento en conocimiento corporativo y de difundirlo adecuadamente. Se ocuparía, principalmente, de las decisiones pragmáticas y estratégicas relativas a la creación, la identificación, la captura, el almacenamiento y la difusión el conocimiento integrado en una organización. Y, el desarrollo de estas operaciones se implementaría en sintonía con la dimensión humana de esos procesos y respetando y rediseñando los elementos organizativos necesarios (Pérez, 2009).

2.2.5. DATA WAREHOUSE

2.2.5.1. Definición

Según (Rojas, 2008) las Tecnologías de la Información (IT) han cambiado sustancialmente la forma de hacer negocios de las empresas. En un entorno donde la competitividad, la globalización, la consolidación de industrias, un ciclo de vida más corto de los productos, saturación de mercados, etc. La información juega cada vez un papel más preponderante. La información referente a mercados, competidores, clientes, incluso la relativa a los indicadores de rendimiento de la propia compañía, se ha convertido en un recurso clave. El problema radica en que las empresas disponen de una gran cantidad de datos, pero muy poca información. Varias razones motivan estos hechos: islas de información, carencia de arquitectura, gestión, responsabilidad, posesión de los datos, deficiencia en calidad, contenido, accesibilidad, fiabilidad de la información, múltiples y diversas aplicaciones operacionales, existencia de fuentes de información externa, etc. Gran parte del producto generado por tecnologías de información, no es información, sino solo datos brutos. Son generados por sistemas que fueron ideados para recogerlos, pero no para analizarlos. Los

datos adquieren la categoría de información cuando disponen de una estructura inteligente. A su vez, esta información se convertirá en conocimiento si se le añade las ideas, intuición, capacidad del analista, es decir, conocimiento tácito

Según (Díaz, 2010) un data warehouse es un repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización independientemente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios, con las propiedades siguientes: estable, coherente, fiable y con información histórica. Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes). Las bases de datos relacionales son el soporte técnico más comúnmente usado para almacenar las estructuras de estos datos y sus grandes volúmenes.

Un data warehouse (Sinnexus, 2007) es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

2.2.5.2. Diferencia entre sistemas operacionales y data warehouse

Los usuarios de un sistema operativo giran las ruedas de la organización. Toman órdenes, firman nuevos clientes y registran quejas. Los usuarios de un sistema operativo casi siempre tratan un registro a la

vez. Repetidamente realizan las mismas tareas operacionales una y otra vez (Kimball & Ross, 2002).

Los usuarios de un data warehouse, por otro lado, observan las ruedas del giro de la organización. Cuentan las nuevas órdenes y las comparan con las órdenes de la semana pasada y preguntan por qué los nuevos clientes se inscribieron y de qué se quejaron los clientes. Los usuarios de un almacén de datos casi nunca tratan una fila a la vez. Más bien, sus preguntas a menudo requieren que centenares o miles de filas sean buscados y comprimidos en un conjunto de respuestas. Para complicar aún más las cosas, los usuarios de un almacén de datos cambian continuamente el tipo de preguntas que piden (Kimball & Ross, 2002).

2.2.5.3. Características

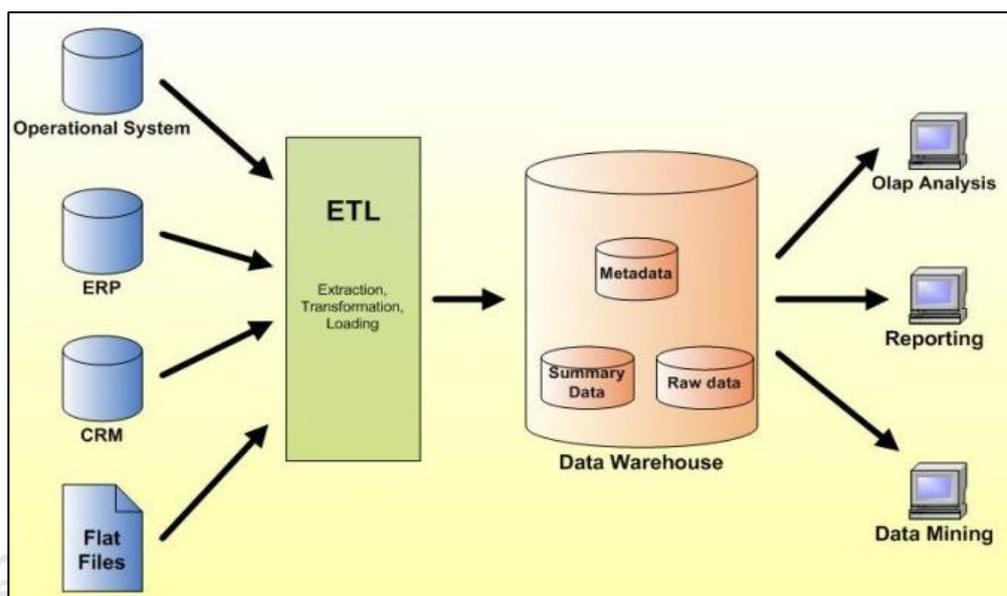
Un data warehouse presenta las siguientes características (Sinnexus, 2007):

- **Integrado:** Los datos almacenados en el data warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Temático:** Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla

del data warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

- **Histórico:** El tiempo es parte implícita de la información contenida en un data warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el data warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el data warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- **No volátil:** El almacén de información de un data warehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del data warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Según (Díaz, 2010) frecuentemente el data warehouse está constituido por una base de datos relacional, pero no es la única opción factible, también es posible considerar las bases de datos orientadas a columnas o incluso basadas en lógica asociativa.

Figura N° 02: Arquitectura de almacenamiento de datos.

Fuente: (datawarehouse4u.info, 2009).

2.2.5.4. Arquitectura

Según (Dertiano, 2015) existen otros enfoques en cuanto a la estructura interna y construcción del data warehouse, pero los más importantes son los de Bill Inmon y Ralph Kimball, que se definen a continuación:

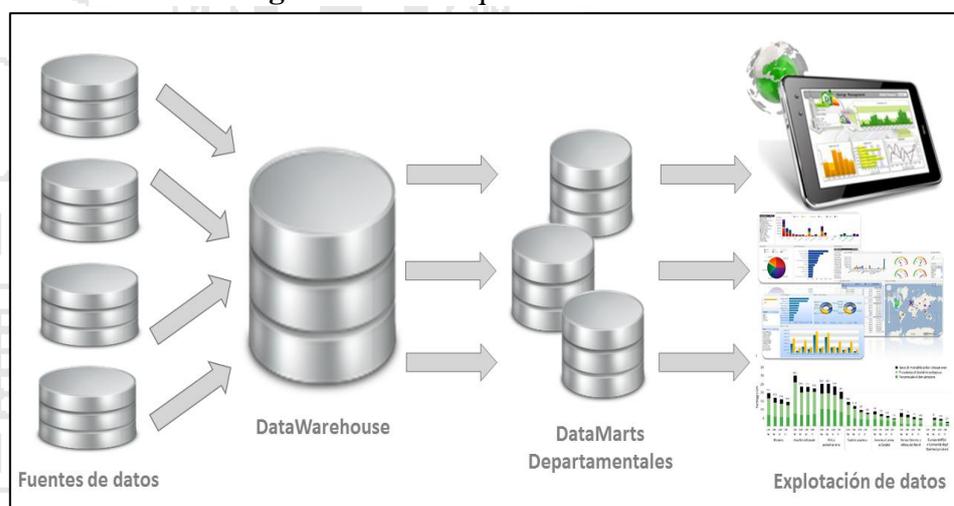
- **Enfoque de Bill Inmon.**

Para él, un data warehouse ha de entenderse como un almacén de datos único y global para toda la empresa. Un repositorio que centralice los datos de los diferentes sistemas operacionales de las organizaciones para que éstos queden validados e integrados en una única base de datos.

En este modelo, la premisa es que la información se almacene al máximo nivel de detalle (garantizando la futura exploración de los datos), permaneciendo invariable y no volátil, de manera que los cambios que sufran los datos a lo largo del tiempo queden registrados sin que puedan modificarse o eliminarse.

Estas son las claves fundamentales de la arquitectura defendida por Inmon, conocida como “Corporate Information Factory (CIF), donde el data warehouse centraliza todos los datos de la compañía para alimentar, a continuación, pequeños datamarts temáticos, que serán los puntos de acceso para las herramientas de reporting. En este sentido, cada departamento tendrá su propio datamart, abastecido con la información del datawarehouse, listo para su análisis y explotación .

Figura N° 03: Arquitectura Innom.



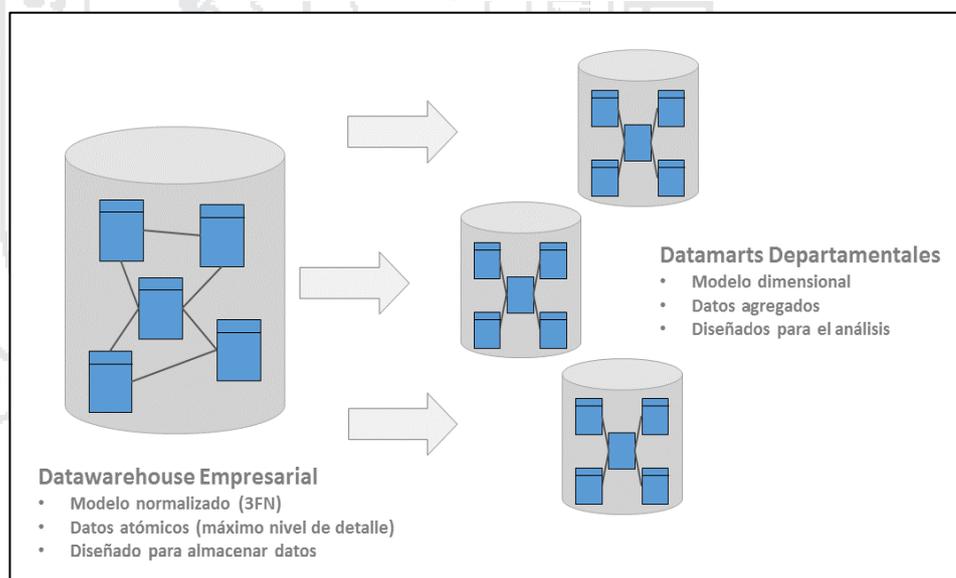
Fuente: (Dertiano, 2015).

Este enfoque de Inmon suele denominarse como una metodología de trabajo “Top-Down”, ya que se centra primero en una visión global de la compañía, para ir desmembrándola en pequeños sets de datos departamentales. Así, con esta arquitectura, todos los datamarts de la organización están conectados al datawarehouse, evitándose la aparición de incongruencias y anomalías al comparar los datos entre distintos departamentos.

En cuanto a la estructura interna del data warehouse, para Inmon la prioridad es que el modelo de datos esté construido en tercera forma normal. Por dar una breve explicación de lo que esto significa, el proceso

de normalización consiste en aplicar una serie de reglas o normas a la hora de establecer las relaciones entre los diferentes objetos dentro de la base de datos. Con este proceso de normalización se consiguen muchos beneficios, como evitar la redundancia de los datos, mantener su integridad referencial, facilitar el mantenimiento de las tablas y disminuir el tamaño de la base de datos. Sin embargo, a diferencia de los data warehouse desnormalizados, las consultas exigen el empleo de queries mucho más complejas, lo que dificulta el análisis directo de la información y el uso de las herramientas de reporting. De ahí, la necesidad de construir los datamarts que, como ya comenté, están basados en modelos dimensionales de estrella o copo de nieve, diseños fácilmente explotables por estas herramientas de análisis de datos.

Figura N° 04: Arquitectura Inmon - 3FN.



Fuente: (Dertiano, 2015).

- **Enfoque de Ralph Kimball.**

Al contrario que Inmon, Kimball defiende una metodología de trabajo “Bottom-up”. Con esto quiere decir que el procedimiento a seguir para construir un data warehouse es empezar en un principio por

pequeños componentes para ir evolucionando a estructuras y modelos superiores. Y esto es así porque para Kimball un data warehouse no es más que la unión de los diferentes datamarts de una organización.

Su filosofía se centra en que, en la mayoría de las organizaciones, la construcción de un data warehouse se origina por el interés y esfuerzo de un departamento. Es por esto por lo que en su primera versión este data warehouse no es más que un datamart departamental.

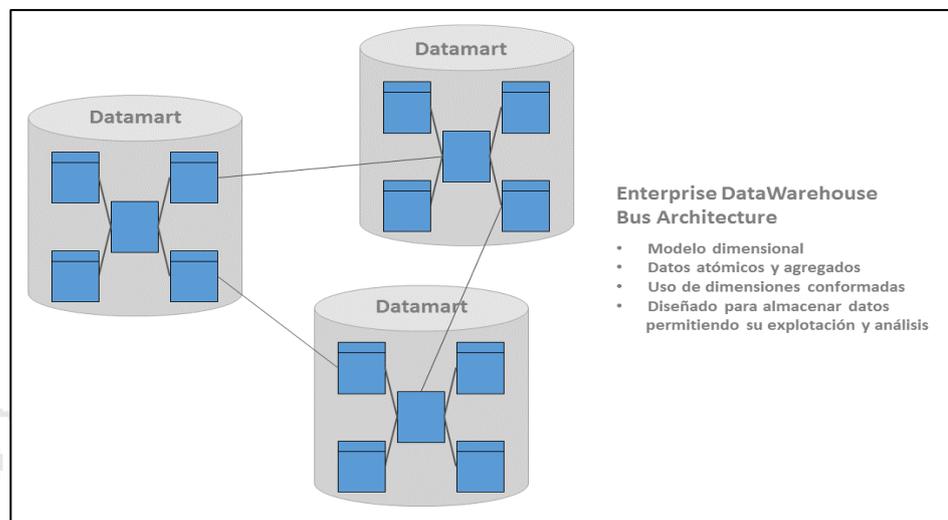
A medida que otros departamentos necesiten sus propios datamarts, éstos se irán combinando con el primero manteniendo una metodología de estandarización mediante lo que Kimball denomina “dimensiones conformadas”, que serán las dimensiones comunes entre los diferentes departamentos. La clave radica en que estas dimensiones han de ser compartidas por los distintos datamarts que existan en la organización, garantizándose así la integridad de los mismos y dando lugar al conglomerado de estructuras que para Kimball conforman el data warehouse.

Para lograr este resultado es importante que estas dimensiones conformadas tengan un diseño consistente y apto para todos los datamarts, de forma que al crearse uno nuevo reutilice las dimensiones ya definidas, pudiendo incluir o no otras dimensiones nuevas.

La principal ventaja de este enfoque de almacén de datos es que, al estar formado por pequeños datamarts estructurados en modelos de datos dimensionales (esquemas de estrella o copo de nieve), especialmente diseñados para la consulta y generación de informes, el datawarehouse al

completo puede ser explotado directamente por las herramientas de reporting y análisis de datos sin la necesidad de estructuras intermedias.

Figura N° 05: Arquitectura Kimball - modelo dimensional.



Fuente: (Dertiano, 2015).

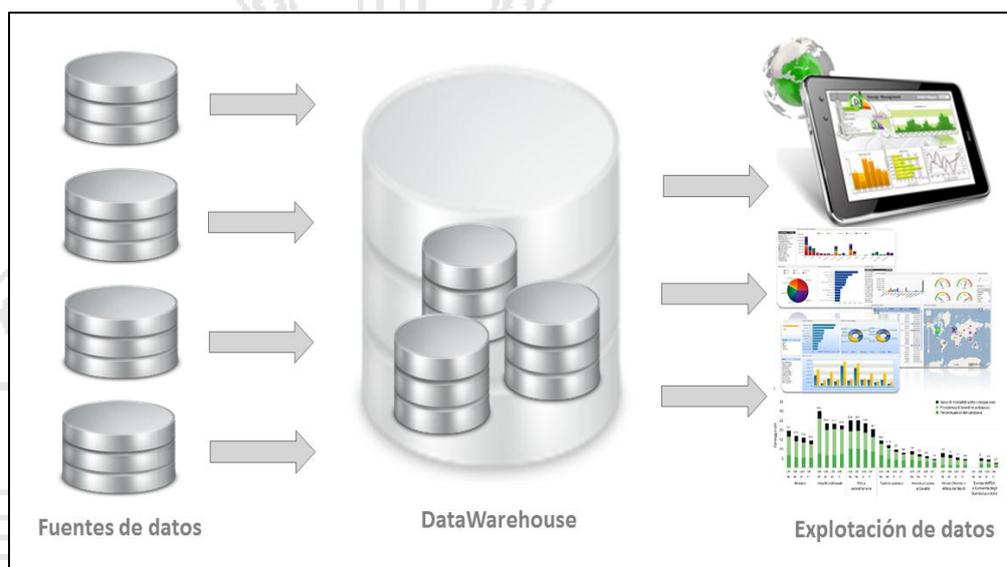
En cuanto a las cuestiones sobre la granularidad, a pesar de que este tipo de datawarehouse suele presentar los datos agregados en base a las consultas e informes que haya que generar, Kimball insiste en la necesidad de que estas agregaciones estén complementadas con datos a mayor nivel de detalle. El argumento es que las preguntas de negocio que puedan llegar a hacer los usuarios son impredecibles, de manera que el datawarehouse tiene que estar preparado para dar respuesta a todas ellas, garantizando la exploración de los datos y la navegación a través de jerarquías desde datos agregados hasta información desagregada.

A este tipo de arquitectura Kimball lo denomina como “Data Warehouse Bus Architecture” y los cuatro pasos fundamentales que se han de seguir para construir este tipo de base de datos son, en primer lugar, la identificación del proceso de negocio que se pretenda estudiar, la definición de la granularidad de los datos, la selección de las

dimensiones y atributos y, por último, la identificación de los hechos o métricas.

El esquema de arquitectura en base a los fundamentos de Ralph Kimball sería el de la siguiente imagen:

Figura N° 06: Arquitectura Kimball.



Fuente: (Dertiano, 2015).

2.2.5.5. COMPONENTES.

Según (Kimball & Ross, 2002) existen cuatro componentes separados y distintos a considerar cuando exploramos el entorno del Data Warehouse:

a) **Sistemas Operativos de Origen.**

Estos son los sistemas operativos de registro que capturan las transacciones del negocio. Los sistemas fuente deben considerarse como fuera del almacén de datos porque presumiblemente tenemos poco o ningún control sobre el contenido y el formato de los datos en estos sistemas heredados operativos. Las principales prioridades de los sistemas fuente son el procesamiento de rendimiento y disponibilidad. Las consultas con los sistemas de origen son consultas

estrechas, de un registro a la vez que son parte del flujo de transacción normal y severamente restringidas en sus demandas en el sistema operativo.

b) Área de Preparación de Datos.

El área de almacenamiento de datos del almacén de datos es tanto un área de almacenamiento como un conjunto de procesos comúnmente denominados extracción-transformación-carga (ETL).

El área de clasificación de datos es todo entre los sistemas de fuentes operacionales y el área de presentación de datos.

La extracción es el primer paso en el proceso de obtención de datos en el entorno del almacén de datos. Extraer significa leer y comprender los datos de origen y copiar los datos necesarios para el almacén de datos en el área de ensayo para manipulación adicional.

Una vez que los datos se extraen al área de ensayo, existen numerosas transformaciones potenciales, como la limpieza de los datos (corrección de errores ortográficos, resolución de conflictos de dominio, tratamiento de elementos ausentes o análisis en formatos estándar), combinando datos de múltiples fuentes, deduplicando Datos y asignación de claves de almacén. Estas transformaciones son precursoras para cargar los datos en el área de presentación del almacén de datos.

c) Área de Presentación de Datos.

El área de presentación de datos es donde los datos son organizados, almacenados y puestos a disposición para consultas directas por parte de los usuarios, escritores de informes y otras

aplicaciones analíticas. Dado que la zona de espera de la parte trasera está fuera de los límites, el área de presentación es el almacén de datos en lo que respecta a la comunidad empresarial. Es todo lo que la comunidad de negocios ve y toca a través de herramientas de acceso a datos.

Generalmente nos referimos al área de presentación como una serie de marts de datos integrados. Un mart de datos es una cuña de la tarta de área de presentación general. En su forma más sencilla, un data mart presenta los datos de un único proceso de negocio. Estos procesos de negocio cruzan los límites de las funciones organizativas.

d) Herramientas de Acceso a Datos.

El último componente principal del entorno del almacén de datos son las herramientas de acceso a los datos. Utilizamos el término herramientas de forma holgada para referirnos a la variedad de capacidades que se pueden proporcionar a los usuarios empresariales para aprovechar el área de presentación para la toma de decisiones analíticas. Por definición, todas las herramientas de acceso a datos consultan los datos en el área de presentación del almacén de datos. Consultar, obviamente, es el punto de usar el almacén de datos.

Una herramienta de acceso a datos puede ser tan simple como una herramienta de consulta ad hoc o tan compleja como una sofisticada aplicación de minería de datos o modelado. Las herramientas de consulta ad hoc, tan potentes como son, sólo pueden entenderse y utilizarse eficazmente por un pequeño porcentaje de la población

potencial de usuarios de negocios de data warehouse. La mayoría de la base de usuarios de negocios probablemente accederá a los datos a través de aplicaciones analíticas basadas en parámetros previamente construidas. Aproximadamente del 80 al 90 por ciento de los usuarios potenciales serán servidos por estas aplicaciones enlatadas que son plantillas esencialmente terminadas que no requieren que los usuarios construyan consultas relacionales directamente. Algunas de las herramientas de acceso a datos más sofisticadas, como las herramientas de modelado o previsión, realmente pueden cargar sus resultados en sistemas operativos de origen o en las áreas de puesta en escena / presentación del almacén de datos.

2.2.6. DATAMART.

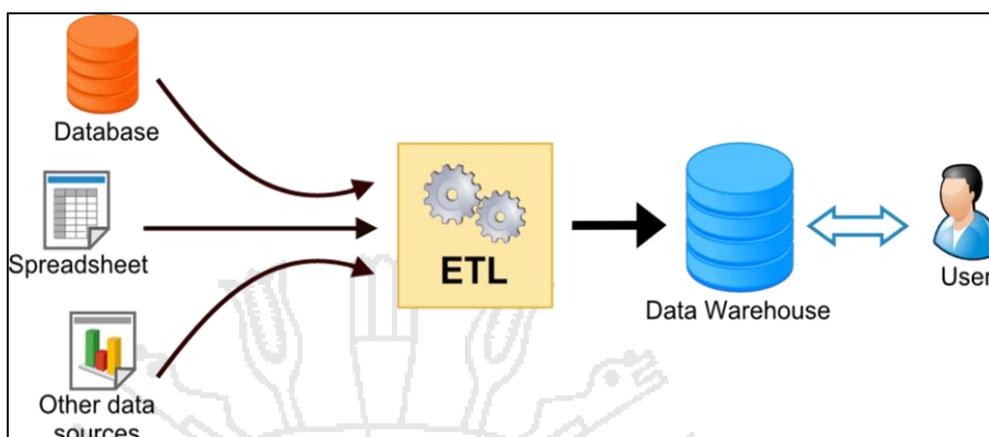
Un Datamart según (Sinnexus, 2007) es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un data warehouse, o integrar por sí mismo un compendio de distintas fuentes de información. Por tanto, para crear el datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio data warehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento.

2.2.7. ETL (Extracción, Transformación y Carga).

Según (Díaz, 2010), la ETL permite extraer datos del entorno origen, transformarlos según nuestras necesidades de negocio para integración de datos y cargar estos datos en los entornos destino. Los entornos origen y destino son usualmente bases de datos y/o ficheros, pero en ocasiones también pueden ser colas de mensajes de un determinado middleware, así como ficheros u otras fuentes estructuradas, semiestructuradas o no estructuradas. Está basada en técnicas de consolidación. Las herramientas de ETL en la práctica mueven o transportan datos entre entornos origen y destino, pero también documentan cómo estos datos son transformados (si lo son) entre el origen y el destino almacenando esta información en un catálogo propio de metadatos; intercambian estos metadatos con otras aplicaciones que puedan requerirlos y administran todas las ejecuciones y procesos de la ETL: planificación del transporte de datos, log de errores, log de cambios y estadísticas asociadas a los procesos de movimiento de datos. Este tipo de herramientas suelen tener una interfaz de usuario de tipo GUI y permiten diseñar, administrar y controlar cada uno de los procesos del entorno ETL.

Según (Larissa T. Moss & Shaku Atre, 2003) los datos de origen para las aplicaciones de BI provendrán de una variedad de plataformas, que son administradas por una variedad de sistemas operativos y aplicaciones. El propósito del proceso ETL es fusionar los datos de estas plataformas heterogéneas en un formato estándar para las bases de datos de BI en el entorno de soporte de decisiones de BI, como se muestra en la Figura N° 07.

Figura N° 07: Procesos ETL



Fuente: (SHISUBHA, 2015).

2.2.7.1. PREPARACION PARA EL PROCESO ETL.

El proceso de ETL comienza con los preparativos para reformatear, conciliar y limpiar los datos de origen (Larissa T. Moss & Shaku Atre, 2003):

- **Reformateo:** Los datos fuente que residen en varios archivos fuente diferentes y bases de datos fuente, cada uno con su propio formato, tendrá que ser unificado en un formato común durante el proceso ETL.
- **Conciliación:** La tremenda cantidad de datos en las organizaciones apunta a una asombrosa redundancia, que invariablemente resulta en asombrosas inconsistencias. Estos deben ser encontrados y reconciliados durante el proceso ETL.
- **Limpieza:** Los datos sucios encontrados durante el análisis de datos y prototipos tendrán que ser limpiados durante este proceso.

2.2.8. OLAP (Procesamiento Analítico en Línea).

Se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o

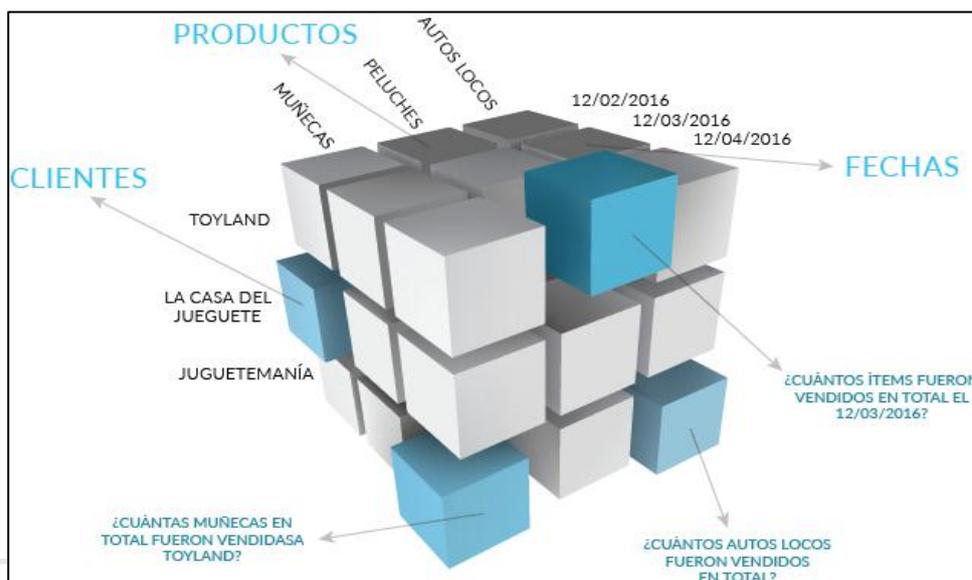
jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar, manipular datos y combinaciones de los mismos a través de consultas o incluso informes (Díaz, 2010).

OLAP es una capacidad enfocada en el análisis y exploración de información, mientras las herramientas de consulta y reportaje ponen mayor énfasis en el acceso a información, para propósitos de monitoreo. OLAP pasa del enfoque “qué” está pasando, a explorar “por qué” está pasando. Para descubrir el “por qué”, los usuarios pueden no saber exactamente qué información buscan y, más bien, navegarán, probarán un conjunto de información para descubrir detalles y patrones particulares (Howson, 2009).

OLAP proporciona análisis interactivo mediante diferentes dimensiones (geografía, producto, tiempo) y diferentes niveles de detalle (año, trimestre, mes). Para muchos usuarios, OLAP se ha convertido en sinónimo de capacidades de “prueba” y “pivote”. Sin embargo, muchos productos BI actualmente proporcionan capacidades de prueba y pivote, sin motor OLAP completamente funcional o base de datos OLAP auxiliar (Howson, 2009).

Conforme la tecnología y los usuarios evolucionan y maduran, las diferencias entre OLAP y reportaje se han desdibujado más. Los usuarios de OLAP quieren reportes altamente formateados basados en información multidimensional; un usuario de reportes quiere cambios inmediatos cuando ve problemas en la métrica particular de uno. No quieren se les obligue ejecutar una herramienta separada, mientras se mueven de reportaje a análisis y exploración (Howson, 2009).

Figura N° 08: Cubo OLAP.



Fuente: (Software, 2016).

Las siguientes características continúan diferenciando herramientas OLAP de herramientas de consulta comercial y reportaje (Howson, 2009):

- **Multidimensional.** Los usuarios analizan valores numéricos desde diferentes dimensiones, tales como producto, tiempo, y geografía. Un reporte, por otra parte, puede ser unidimensional, tal como lista de precios de productos en determinado momento en el tiempo.
- **Consistentemente rápido.** Conforme los usuarios navegan diferentes dimensiones y niveles dentro de una dimensión, OLAP significa rapidez (a la velocidad del pensamiento). Si un usuario hace doble clic para hacer cambios rápidos de Año a Trimestre, esperar 24 horas, 24 minutos o, incluso, 24 segundos por una respuesta no es aceptable. Por supuesto, los usuarios de reporte tampoco quieren reportes lentos, pero algunos toman este tiempo para correr y deben ser programados.

- **Altamente interactivo.** Cambios rápidos es una manera en que los usuarios interactúan con información OLAP. Girar da a los usuarios la habilidad de ver la información desde diferentes perspectivas tales como geografía o producto. Cortar permite filtrar la información dentro de estas dimensiones, tales como ventas para Nueva York solamente y luego Nueva Jersey; o estadísticas de crimen para Leeds solamente y luego Manchester. Este tipo de interacción dentro de los reportes sin OLAP va de inexistente a recientemente posible.
- **Varios niveles de agregaciones.** Para garantizar tiempos de consulta predecibles, los productos OLAP pre-añaden información de diferentes maneras. Al contrario, el reportaje puede estar en el nivel más bajo de detalle: en lugar de ventas por producto, se pueden tener objetos de línea individual para un número de pedido particular.
- **Cálculos interdimensiones.** Con dimensiones múltiples vienen cálculos complejos. En OLAP, se puede analizar el porcentaje de contribución o participación de mercado. Estos análisis requieren de ventas subtotales para un estado particular y luego calcular el porcentaje de contribución por el total de la región, país o mundo. Los usuarios pueden analizar este porcentaje de participación de mercado mediante otras dimensiones, tales como: actual contra el presupuesto, este año contra el pasado o para un grupo particular de productos. Dichos cálculos con frecuencia deben ser realizados en un orden particular, involucrando números de entrada que los usuarios pueden nunca ver. Sin embargo, los reportes detallados con

frecuencia dependen de subtotaes simples o cálculo de valores mostrados en el reporte mismo.

2.2.8.1. Tipos de OLAP.

Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Sin embargo, el solo describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente. Tenemos tres tipos entre los principales (Sinnexus, 2007):

- **MOLAP.**

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona un interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

La información procedente de los sistemas operacionales, se carga en el sistema MOLAP, mediante una serie de rutinas por lotes. Una vez cargado el dato elemental en la Base de Datos multidimensional (MDDDB), se realizan una serie de cálculos por lotes, para calcular los datos agregados, a través de las dimensiones de negocio, relleno la estructura MDDDB.

Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de accesos a las consultas. Una vez que el proceso de compilación se ha acabado, la MDDDB está lista para su uso. Los usuarios solicitan informes a través de la interface, y la lógica de aplicación de la MDDDB obtiene el dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos pre-compilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan pre-calculados y almacenados previamente.

- **ROLAP**

La arquitectura ROLAP, accede a los datos almacenados en un data warehouse para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan

mejor contra las bases de datos relacionales.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

El motor ROLAP se integra con niveles de presentación, a través de los cuáles los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el data warehouse se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por los modelos de datos. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos pre-calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del data warehouse, y soporta técnicas de

optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

- **HOLAP.**

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

2.2.9. HERRAMIENTAS PARA IMPLEMENTAR INTELIGENCIA DE NEGOCIOS

Según (Moreno, 2013) existen varias plataformas para implementar un sistema basado en negocios inteligentes teniendo los siguientes:

a) Websphere Data Integrator (Data Stage)

Data Stage es una herramienta que apoya a la recolección, integración y transformación de grandes volúmenes de datos, con estructuras de datos que van desde simples a muy complejas. Permite a las empresas resolver problemas de negocio a gran escala a través del procesamiento masivo de datos. Además permite la conectividad entre cualquier fuente de datos y cualquier aplicación.

b) SSIS

Microsoft Integration Services es una plataforma para la creación de soluciones empresariales de transformaciones de datos e

integración de datos. Integration Services sirve para resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenes de datos, la limpieza y minería de datos, y la administración de objetos y datos de SQL Server. Los paquetes pueden funcionar por separado o conjuntamente con otros paquetes para hacer frente a las complejas necesidades de la empresa. Integration Services puede extraer y transformar datos de muchos orígenes distintos, como archivos de datos XML, archivos planos y orígenes de datos relacionales, y, posteriormente, cargarlos en uno o varios destinos.

c) Sunopsis

Es un producto que posee un alto desempeño e integración efectiva. Trabaja con una arquitectura ELT (extraer, cargar y transformar) en lugar de la arquitectura tradicional ETL. Esto implica un mayor rendimiento y escalabilidad, junto con un mantenimiento más fácil para las soluciones de integración de datos.

d) Microstrategy

Es un software usado para el análisis de datos, informes y cuadros de mando, inteligencias móviles, minería de datos, la previsión, gestión de operaciones y toma de decisiones ejecutivas; permitiendo de esta forma mejorar y predecir el comportamiento del negocio.

e) Cognos

Software del grupo IBM cuyo uso permite comprender mejor el movimiento del negocio y tomar decisiones adecuadas.

f) Business Objects

Plataforma de Inteligencia de negocios que ofrece una amplia gama de herramientas y aplicaciones diseñadas para ayudar a optimizar el rendimiento empresarial conectando personas, información y empresas a través de redes de negocios.

g) Octopus

Es una herramienta de extracción, transformación y carga (ETL) de Microsoft basada en Java. Puede conectarse a cualquier fuente de datos JDBC y realizar transformaciones definidas en un archivo XML.

h) Pentaho

Es una herramienta de código abierto y de uso libre muy completa, pues incluye elaboración de reportes, cubos, data mining, ETL y una plataforma BI.

Tabla N° 02: Herramientas de ETL.

Herramienta	Permite ETL	Uso libre (Open Source)	Soporte de la Herramienta	Soporta múltiples plataformas
Data Stage	Si	No	Si	Si
SQL Server	Si	No	Si	Si
Sunopsis	Si	No	-	-
Pentaho	Si	Si	No	Si
Octopus	Si	Si	-	-

Fuente: (Moreno, 2013).

Tabla N°3: Herramientas de explotación.

Herramienta	Elaboración de Reportes	Uso libre (Open Source)	Análisis de la información en todas las dimensiones	Exportación a archivos Excel para visualizar la información	Drill and Down	Filtro y búsquedas personalizadas	Slice and Dice	Soporte de la Herramienta	Soporta múltiples plataformas
SQL Server	Si	No	Si	Si	Si	Si	Si	Si	Si
Microstrategy	Si	No	Si	Si	Si	Si	Si	Si	Si
Business Objects	Si	No	Si	Si	Si	No	Si	Si	Si
Cognos	Si	No	Si	Si	Si	Si	Si	Si	Si
Pentaho	Si	Si	Si	Si	Si	Si	Si	No	Si

Fuente: (Moreno, 2013).

2.2.10. PENTAHO.

PENTAHO es una de las suites más completas y maduras del mercado OSBI que existe desde el año 2006. Existen dos versiones: Community y Enterprise y están compuestas por diferentes motores incluidos en el servidor de Pentaho (Díaz, 2010):

- Pentaho Reporting: soporta informes estáticos, paramétricos y ad hoc.
- Pentaho Análisis: soporta OLAP (mediante Mondrian) y minería de datos (mediante Weka).
- Cuadros de mando: mediante CDF (Community Dashboard Framework).
- ETL: mediante la herramienta Kettle.
- Metadata: que proporciona una capa de acceso de información basada en lenguaje de negocio.
- Workflow: el servidor de Pentaho se basa en acciones que la mayoría de objetos de negocio permite lanzar.

Pentaho, es un proyecto iniciado por una comunidad Open Source, provee una alternativa de soluciones de BI en distintas áreas como en la arquitectura, soporte, funcionalidad e implantación. Estas soluciones al igual que su ambiente de implantación están basados en Java, haciéndolo flexible en cubrir amplias necesidades empresariales. A través de la integración funcional de diversos proyectos de Open Source permite ofrecer soluciones en áreas como: Análisis de información, reportes, tableros de mando conocido como “DashBoards”, flujos de trabajo y Minería de datos (Gravitar, 2016).

Para facilitar la comprensión y proveer una clasificación manejable sobre los diversos elementos que determinan el entorno Pentaho, hemos definido seis temas fundamentales de los cuales describiremos a manera de introducción en esta publicación (Gravitar, 2016).

Figura N° 09: Clasificación de Pentaho.



Fuente: (Gravitar, 2016).

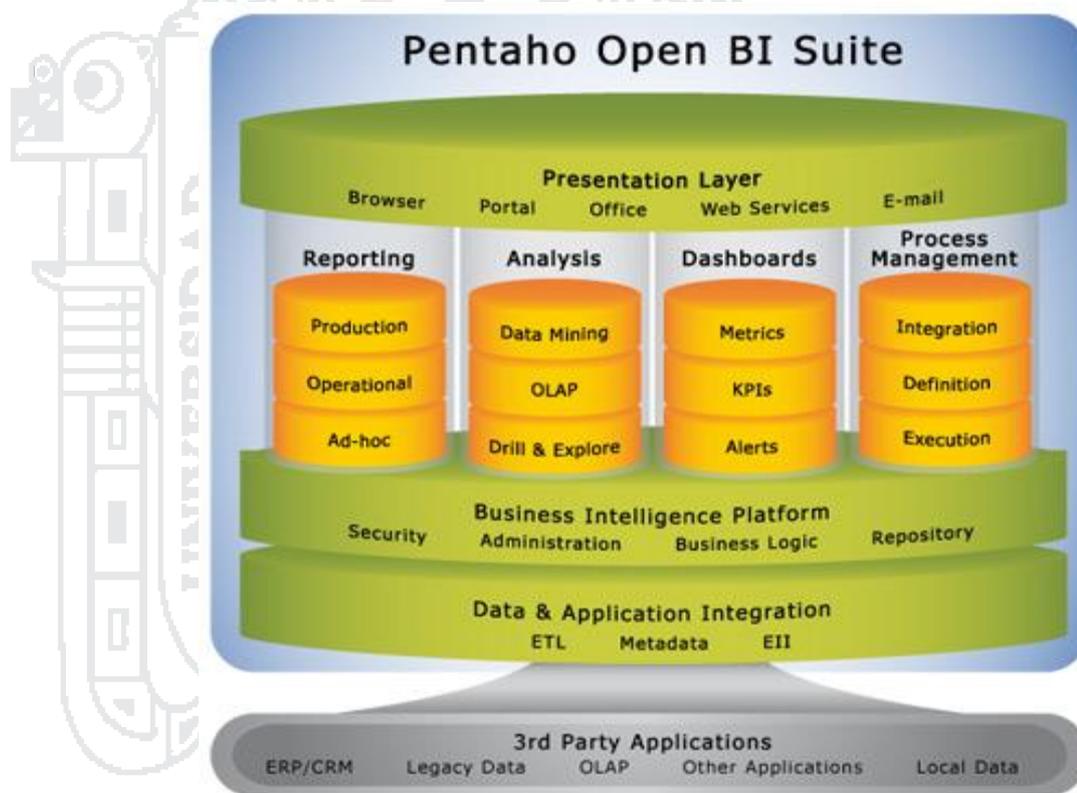
- **Plataforma BI**

Compuesta por componentes Open Source provee la arquitectura y la Infraestructura a la vez. Forma un proceso centralizado bajo un

marco de trabajo orientado a la solución de problemas empleando componentes de BI y permitiendo desarrollos completos para atender a soluciones de Inteligencia de Negocios.

En resumen la plataforma BI integra componentes Open Source mostrando una combinación de flujos de trabajo y administración de procesos. En la siguiente ilustración se muestra la arquitectura funcional de Pentaho:

Figura N° 10: Plataforma Pentaho.



Fuente: (Gravitar, 2016).

- **Herramientas y áreas de aplicación**

Bajo la integración de otros proyectos Open Source que brindan funcionalidad en áreas de BI, la comunidad Pentaho trabaja en formalizar estas herramientas y formar el Suite BI.

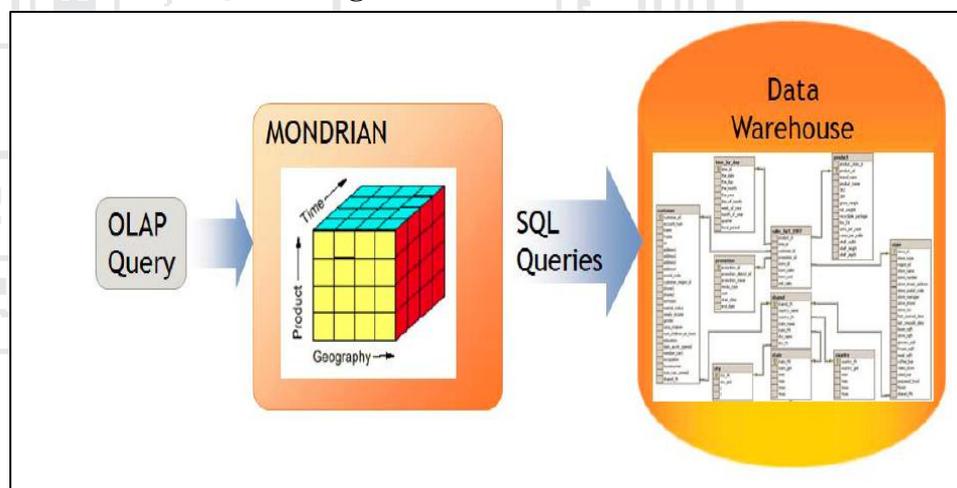
MONDRIAN

Ahora bautizado como “Pentaho Analysis Service” forma parte del motor OLAP integrado en el SUITE BI de PENTAHO

Un ejemplo rápido sobre el flujo de datos es:

1. El cliente manda una solicitud de consulta bajo la interfaz web JPivot.
2. Mondrian recibe la solicitud y bajo el esquema de metadatos que definen sus conceptos multidimensionales busca si ya tiene los datos en cache respondiendo rápidamente a la petición.
3. Si los datos no se encontraron en cache ejecuta las sentencias SQL para generar los datos.
4. Se almacenan los datos recibos en cache para agilizar posteriores consultas. Y finalmente se devuelve el resultado al usuario cliente a través de la interfaz.

Figura N° 11: Mondrian.



Fuente: (Haces, 2016)

PENTAHO REPORTING

La solución proporcionada por la plataforma Business Intelligence Open Source Pentaho e integrada en su suite para el desarrollo de informes se llama Pentaho Reporting.

Pentaho Reporting es un potente generador de informes: Permite la distribución de los resultados del análisis en múltiples formatos.

Existen tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios.

1. **Pentaho Report Designer:**

Editor basado en eclipse con prestaciones profesionales y de calidad y con capacidad de personalización de informes a las necesidades de negocio destinado a desarrolladores.

Incluye Asistentes para facilitar la configuración de propiedades. Está estructurado de forma que los desarrolladores pueden acceder a sus prestaciones de forma rápida:

Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe.

2. **Pentaho Report Design Wizard:**

Herramienta de diseño de informes, que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con menos conocimientos técnicos.

A través de pasos sencillos permite:

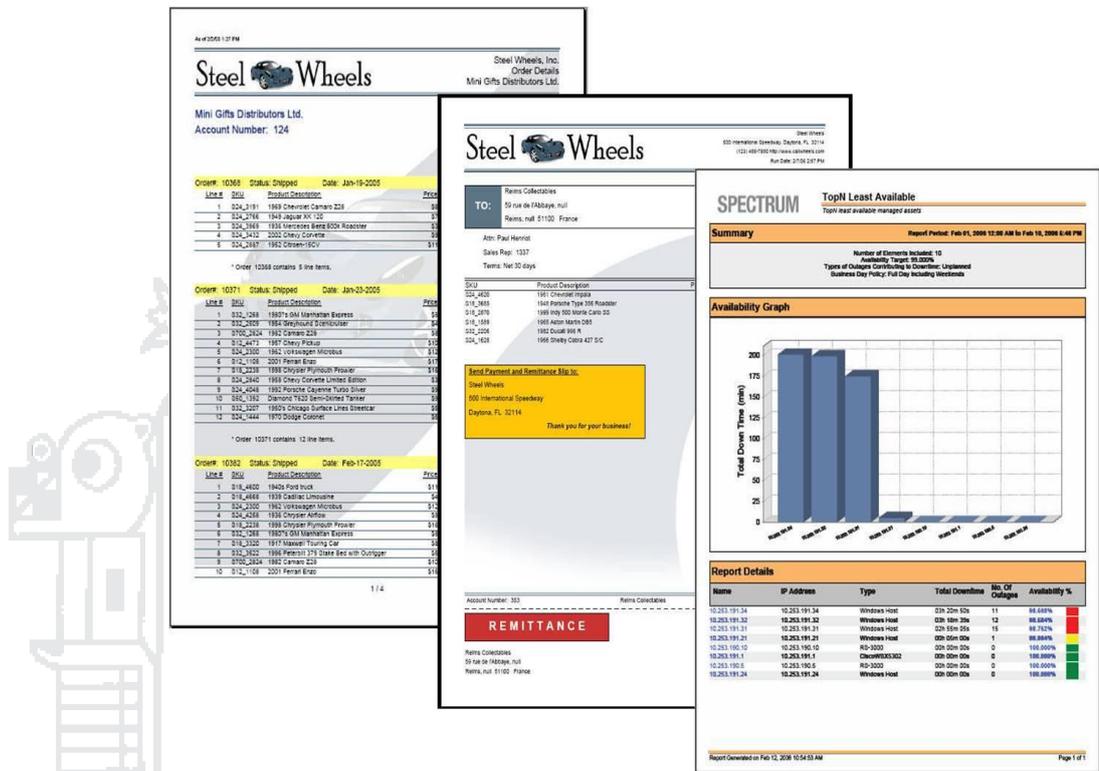
- Conectarse a todo tipo de bases relacionales.
- Integrar el resultado dentro del portal Pentaho.
- Posibilidad de montar codificación semafórica.

3. **Web ad-hoc reporting**

Es el similar a la herramienta anterior pero via web. Extiende la capacidad de los usuarios finales para la creación de informes a

partir de plantillas preconfiguradas y siguiendo un asistente de creación.

Figura N° 12: Pentaho Reporting.



Fuente: (Gravitar, 2016)

Pentaho Analysis

Ayuda a operar con máxima efectividad para ganar perspicacia y entender lo necesario para tomar optimas decisiones.

Las características generales son:

- Vista dimensional de datos (por ventas, por periodo).
- Navegar y explorar.
- Análisis Ad Hoc.
- Drill-down.
- Seleccionar un específico miembro para el análisis.
- Interactuar con alto rendimiento.
- Tecnología optimizada para rápida respuesta interactiva.

Pentaho Dashboards

Provee inmediata perspicacia en un rendimiento individual, departamental o empresarial. Para deliberar key metrics en una atractiva e intuitiva interfaz visual, Pentaho Dashboards a los usuarios de los negocios información crítica que necesitan para entender y mejorar el rendimiento organizacional.

- Identificación de unas Métricas Clave (KPI's, Key Performance Indicators).
 - Monitoreo/Métricas.
- Investiga detalles subyacentes.
 - Drill a reportes de soporte.
- Seguimiento de excepciones.
 - Alertas basadas en reglas del negocio.

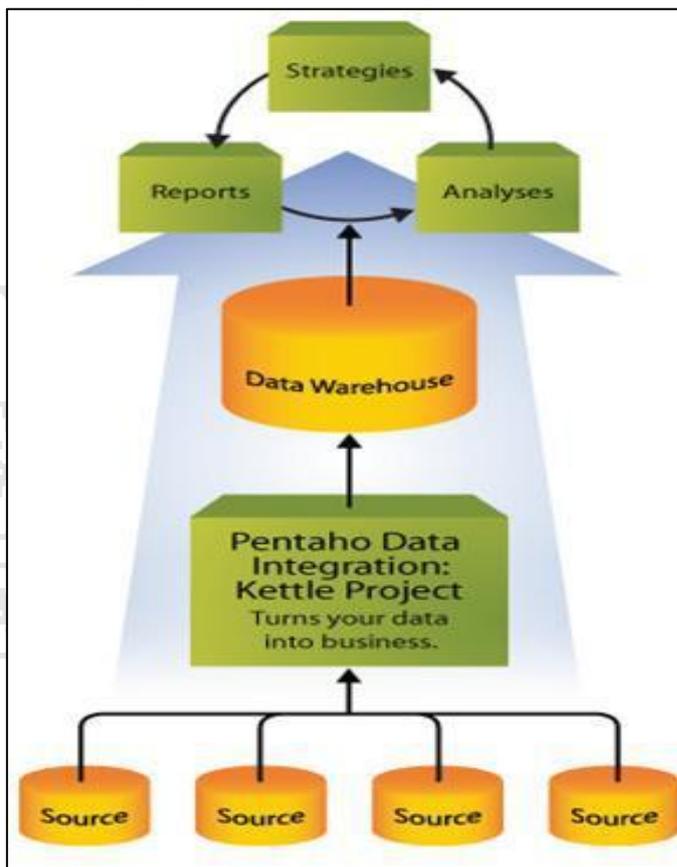
Pentaho Data Integration

Muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones TI hoy en día. Pentaho Data Integration permite una poderosa ETL (Extracción, Transformación y Carga).

El uso de kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

La arquitectura de Pentaho Data Integration viene representada por el siguiente esquema:

Figura N° 13: Data Integration.



Fuente: (Gravitar, 2016).

Data Mining

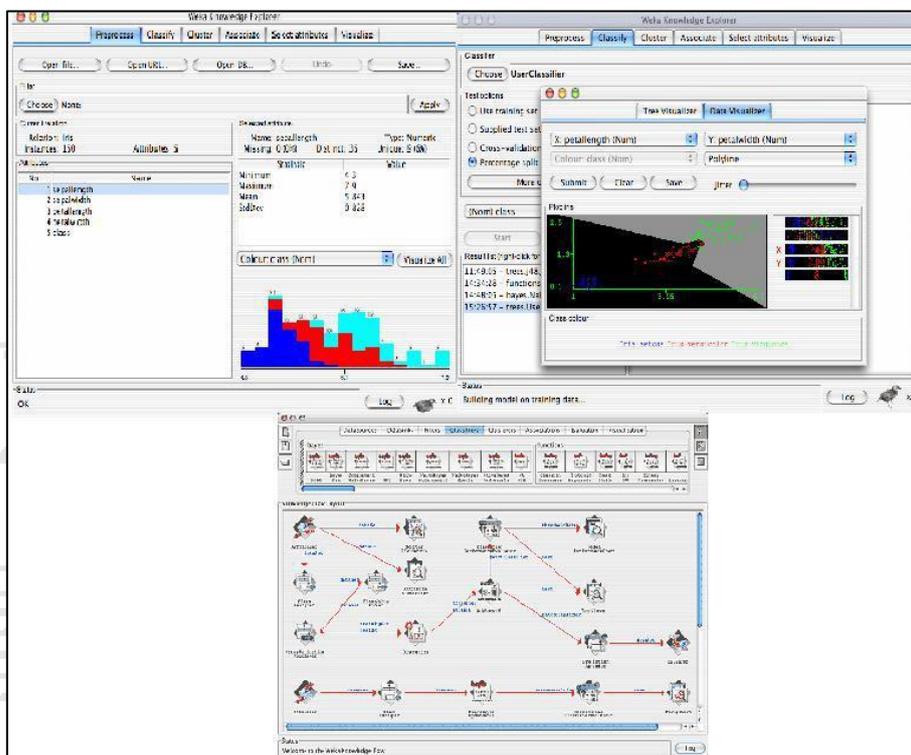
Es el proceso de correr datos en algoritmos completamente sofisticados, relevando significantes patrones y correlaciones que pueden estar escondidos. Esto puede ser usado para ayudar a entender lo mejor para el negocio y explotar el rendimiento de este en un futuro prediciendo completamente en el análisis.

Se caracteriza por:

- Descubrir patrones ocultos y correlacionales en los datos.
- Prevenir eventos futuros basados en patrones históricos.
- Contar con la tecnología de:
 - Poderoso motor de Data Mining.
 - Herramientas de diseño gráfico.

- Seguridad y conformidad.
- Servicios Web, Repositorios y definiciones basadas en XML.
- Rendimiento y escalabilidad.

Figura N° 14: Data Mining.



Fuente: (Gravitar, 2016)

- **Implantación (instalación)**

Plataforma BI de Pentaho

Para aquellos que cuentan con el Sistema Operativo Windows se recomienda optar por la Instalación Preconfigurada (PCI) con la aplicación de servidor JBoss y algunos ejemplos.

La plataforma requiere el paquete Java SDK 1.4 previamente instalado.

Adicionalmente, se requerirá del “J2EE Deployment Distribution” permitiendo construir una variedad de diferentes aplicaciones web.

Para que la instalación (PCI) sea complementada, se requerirá de

la instalación de MySQL en la maquina local con la configuración de omisión usando el puerto 3306.

Configuración

Correo: La plataforma envía mensajes a través del correo usando el SMTP Server. En la mayoría de los casos requiere el “ID” del usuario, contraseña y una dirección de correo válida.

Para modificar estas configuraciones la ruta de omisión es:

`/pentaho-demo/pentaho-solutions/system/smtp-
email/email_config.xml`

Puertos: El servidor tratará de usar el puerto 8080. En la siguiente ruta se ubica el archivo de configuración de puerto:

`/pentaho-preconfiguredinstall/server/default/deploy/jbossweb-
tomcat55.sar/server.xml`

Publicaciones: Herramientas del Cliente Pentaho publicadas en el servidor. Para disponer de las herramientas del cliente (“Report Design Wizard, Cube Design Wizard, etc”) y publicarlas al servidor se requiere de una contraseña.

Para configurar la contraseña edite el siguiente archivo:

`/pentaho-demo/pentaho-solutions/system/publisher_config.xml`

Despliegue: La instalación PCI de PENTAHO está diseñada para trabajar localmente en `http://localhost:8080/pentaho` como la URL de base.

Para accederlo a través de otra terminal, se requiere configurar el `web.xml`, ubicado en:

`/pentaho-demo/jboss/server/default/deploy/pentaho.war/WEB-INF`

```
<context-param>
```

```
<param-name>base-url</param-name>
```

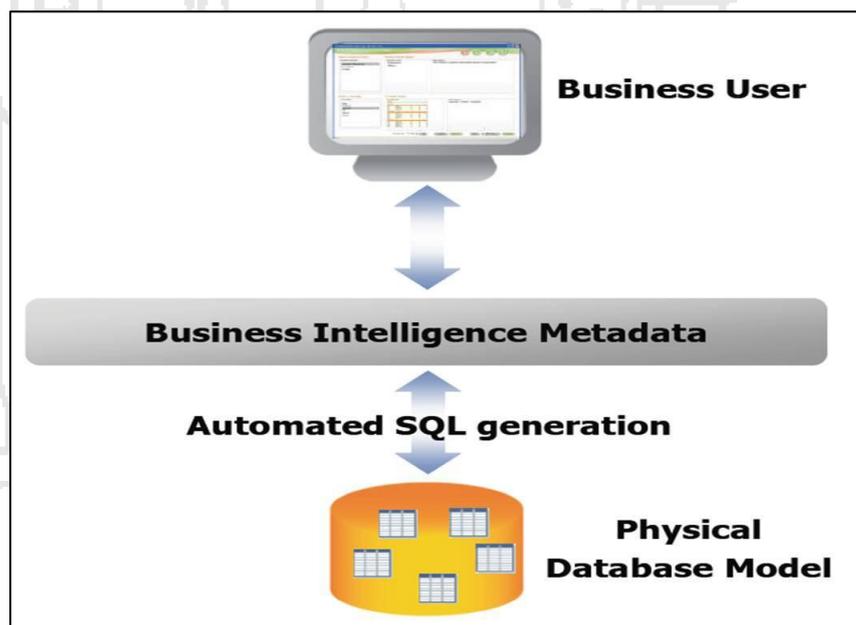
```
<param-value>http://myserver.pentaho.org:8080/pentaho</param-  
value>
```

```
</context-param>
```

- **Capacidades funcionales**

Capa de metadatos: Provee múltiples beneficios incluyendo seguridad integrada, optimización de consultas, amplia conectividad de base de datos y mantenimiento reducido. Entre sus funciones más básicas es proveer una abstracción lógica por capas entre usuarios y sofisticados esquemas de Base de Datos.

Figura N° 15: Capa de metadatos.



Fuente: (Gravitar, 2016).

Mantenimiento y control centralizado de TI

El Metadata de BI provee un control de seguridad gráfico, el acceso a datos y el manejo de cambios a la infraestructura sin alterar las

aplicaciones de la inteligencia de negocios.

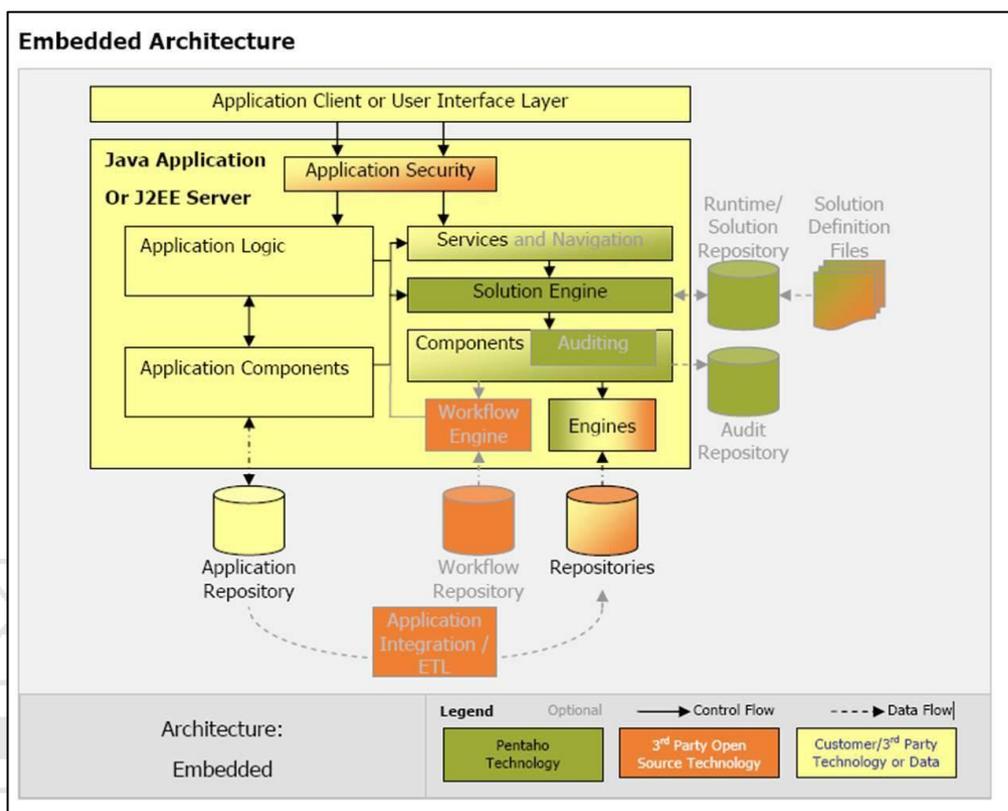
Adicionalmente la nueva plataforma BI de Pentaho permite el Soporte de Plataformas, Soporte de Integración de Datos y Soporte Usuario Final, donde sus capacidades a detalle podrán ser vistas dentro de la documentación que nos ofrece la comunidad Open Source de PENTAHO.

- **Personalización**

La personalización del despliegue de la plataforma Pentaho permite formar una arquitectura empotrada y solo manejable en componentes, motores, repositorios que se requiera de una configuración como:

- Motores de flujo de actividades, repositorios de flujos y repositorios de “runtime”
- Repositorios de Auditar
- Aplicaciones de integración
- Componentes de la interface del usuario
- Repositorio de soluciones y archivos de definición de soluciones

Figura N° 16: Arquitectura personalizada de Pentaho.



Fuente: (Gravitar, 2016).

- **Beneficios.**

Pentaho BI Open Source nos promete una gama de beneficios alentadores, haciendo mayor relieve en el costo, en la definición de estándares abiertos, en su flexibilidad, en sus funciones personalizadas y sobre todo en su arquitectura centrada en procesos, haciendo que las necesidades de la Inteligencia de Negocio sean atendidas con mayor facilidad.

2.2.11. MYSQL.

Según (Gilfillan, 2003) MySQL es la base de datos de código abierto más popular del mundo. Código abierto significa que todo el mundo puede acceder al código fuente, es decir, al código de programación de MySQL. Todo el mundo puede contribuir para incluir elementos,

arreglar problemas, realizar mejoras o sugerir optimizaciones. Y así ocurre. MySQL ha pasado de ser una "pequeña" base de datos a una completa herramienta y ha conseguido superar a una gran cantidad de bases de datos comerciales (lo que ha asustado a la mayor parte de los proveedores comerciales de bases de datos). Por lo tanto, su rápido desarrollo se debe a la contribución de mucha gente al proyecto, así como a la dedicación del equipo de MySQL.

MySQL es un sistema de administración de bases de datos relacional (RDBMS). Se trata de un programa capaz de almacenar una enorme cantidad de datos de gran variedad y de distribuirlos para cubrir las necesidades de cualquier tipo de organización, desde pequeños establecimientos comerciales a grandes empresas y organismos administrativos. MySQL compete con sistemas RDBMS propietarios conocidos, como Oracle, SQL Server y DB2 (Gilfillan, 2003).

MySQL incluye todos los elementos necesarios para instalar el programa, preparar diferentes niveles de acceso de usuario, administrar el sistema y proteger y hacer volcados de datos. Puede desarrollar sus propias aplicaciones de base de datos en la mayor parte de los lenguajes de programación utilizados en la actualidad y ejecutarlos en casi todos los sistemas operativos, incluyendo algunos de los que probablemente no ha oído nunca hablar. MySQL utiliza el lenguaje de consulta estructurado (SQL). Se trata del lenguaje utilizado por todas las bases de relacionales, que presentaremos en una sección posterior. Este lenguaje permite crear bases de datos, así como agregar, manipular y recuperar datos en función de criterios específicos (Gilfillan, 2003).

2.2.12. OTROS TÉRMINOS FUNDAMENTALES

A) INTEGRACION.

En un contexto empresarial, la integración puede darse en cuatro grandes áreas (Díaz, 2010):

- **Integración de datos:** proporciona una visión única de todos los datos de negocio, se encuentren donde se encuentren. Éste es el ámbito del presente documento y, en particular, en el contexto de la inteligencia de negocio.
- **Integración de aplicaciones:** proporciona una visión unificada de todas las aplicaciones tanto internas como externas a la empresa. Esta integración se consigue mediante la coordinación de los flujos de eventos (transacciones, mensaje o datos) entre aplicaciones.
- **Integración de procesos de negocio:** proporciona una visión unificada de todos los procesos de negocio. Su principal ventaja es que las consideraciones de diseño del análisis e implementación de los procesos de negocio son aislados del desarrollo de las aplicaciones.
- **Integración de la interacción de los usuarios:** proporciona una interfaz segura y personalizada al usuario del negocio (datos, aplicaciones y procesos de negocio).

B) OPEN SOURCE.

Las empresas y organizaciones Open Source han ido evolucionando para ofrecer una respuesta adecuada a las demandas del mercado teniendo en cuenta los siguientes factores (Díaz, 2010):

- Existen unos costes ocultos en la implantación de software open source. Es necesario contratar o formar personal especializado que dé soporte en la evaluación, en la integración, en la corrección de errores y en el ciclo de vida del producto, y que participe en la comunidad para que conozca la evolución de la aplicación. Tales costes inciden en horas, dinero o ambos.
- Ausencia de soporte y de un roadmap claro y preciso por parte de la organización que desarrolla el producto.
- Ciertas licencias open source no son business-friendly (limitando el uso a ámbitos académicos o organizaciones no comerciales).

El Open Source es una filosofía de desarrollo de software que cumple los siguientes principios (Díaz, 2010):

- **Abierto:** La comunidad tiene libre acceso, uso y participación del código fuente, así como la posibilidad de uso de foros para proporcionar feedback (retroalimentación).
- **Transparencia:** La comunidad tiene acceso al roadmap (hoja de ruta), documentación, defectos y agenda de las milestones (hitos).
- **Early & Often:** La información se publica de manera frecuente y pronto a través de repositorios públicos (incluyendo el código fuente).

El open source ya no es una tendencia emergente sino que es un enfoque que tiene un impacto profundo y que en años venideros tendrá una presencia importante en todos los sectores. (Díaz, 2010).

C) GRANULARIDAD.

Según (Alejandro, 2009) el concepto de granularidad parte del principio que es más fácil reutilizar unidades más pequeñas dado, que de este modo, es posible seleccionar aquellas partes que nos interesan y descartar aquellas que no son adecuadas en el contexto donde nos encontramos. Además la granularidad describe el nivel de detalle de la base de datos en datawarehouse. La determinación del nivel de granularidad es uno de los puntos más importantes del modelado lo cual impacta directamente en el tamaño de la base de datos.

Granularidad y jerarquías de bloqueo

El Microsoft SQL Server Database Engine (Motor de base de datos de SQL Server) admite bloqueo multigranular. Esta función permite que una transacción bloquee diferentes tipos de recursos. Para minimizar el costo del bloqueo, Database Engine (Motor de base de datos) bloquea automáticamente los recursos en el nivel apropiado para la tarea. Los bloqueos de menor granularidad, como es el caso de las filas, aumentan la simultaneidad. Sin embargo, se produce una sobrecarga mayor porque cuantas más filas se bloquean, más bloqueos se deben mantener. Los bloqueos realizados en una granularidad alta, por ejemplo en tablas, reducen la simultaneidad porque el bloqueo de toda una tabla restringe el acceso de otras transacciones a cualquier parte de la tabla. Sin embargo, produce una sobrecarga menor debido a que se mantienen menos bloqueos (Alejandro, 2009).

El Database Engine (Motor de base de datos) a menudo se ve en la obligación de adquirir bloqueos en distintos niveles de granularidad para

brindar una protección completa a un recurso. Este grupo de bloqueos en distintos niveles de granularidad se denomina jerarquía de bloqueos. Por ejemplo, para brindar protección completa a la lectura de un índice, probablemente sea necesario que una instancia del Database Engine (Motor de base de datos) adquiera bloqueos compartidos en filas y bloqueos con intención compartida en las páginas y la tabla (Alejandro, 2009).

2.3 MARCO CONCEPTUAL

2.3.1. INTELIGENCIA DE NEGOCIOS

Inteligencia de negocios significa diferentes cosas para distintas personas. Para un comerciante, significa investigación de mercado, algo que podría llamarse “inteligencia competitiva”. Para otra persona, “reportaje” puede ser un mejor término, a pesar de que inteligencia de negocios va mucho más allá del acceso a un reporte estático. “Reportaje” y “análisis” son términos utilizados con frecuencia para describir inteligencia de negocios. Otros utilizarán “comercio analítico” o “soporte de decisiones”, ambos con varios grados de adecuación. En qué difieren estos términos importa muy poco, a menos que la intención sea comparar acciones del mercado para diferentes tecnologías. Lo importante es utilizar la terminología más familiar para usuarios potenciales y que tenga connotación positiva (Díaz, 2010).

2.3.2. SISTEMA DE SOPORTE DE DECISIONES (DSS)

Los sistemas de apoyo a la toma de decisiones se ajustan más al gusto de la persona o grupo que los utiliza que a los sistemas de información gerencial tradicionales. En ocasiones se hace referencia a

ellos como sistemas que se enfocan en la inteligencia de negocios (E. Kendall & E. Kendall, 2005).

El DSS es una de las herramientas más emblemáticas del Business Intelligence ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de los programas de gestión (Sinnexus, 2007).

2.3.3. GESTION DE LA INFORMACION

La Gestión de la Información mantiene una estrecha relación con la disciplina de la Gestión del Conocimiento en el contexto organizacional. Los objetivos de la Gestión de Información se centran en aquellos procesos relacionados con el almacenamiento, el tratamiento y la difusión del conocimiento explícito que se encuentra representado en los documentos. Sin embargo, en este contexto, la Gestión del Conocimiento iría un poco más allá que la Gestión de la Información. Ésta se encargaría de convertir todo el conocimiento en conocimiento corporativo y de difundirlo adecuadamente. Se ocuparía, principalmente, de las decisiones pragmáticas y estratégicas relativas a la creación, la identificación, la captura, el almacenamiento y la difusión el conocimiento integrado en una organización. Y, el desarrollo de estas operaciones se implementaría en sintonía con la dimensión humana de esos procesos y respetando y rediseñando los elementos organizativos necesarios (Pérez, 2009).

2.3.4. DATA WAREHOUSE

El Data Warehouse debe servir de base para mejorar la toma de decisiones. El almacén de datos debe tener los datos correctos en él para apoyar la toma de decisiones. Sólo hay una salida real de un almacén de datos: las decisiones que se toman después de que el almacén de datos

haya presentado su evidencia. Estas decisiones entregan el impacto comercial y el valor atribuible al almacén. La etiqueta original que antecede al almacén de datos sigue siendo la mejor descripción de lo que estamos diseñando: un sistema de apoyo a la decisión (Kimball & Ross, 2002).

La comunidad empresarial debe aceptar el Data Warehouse si se considera que tiene éxito. No importa que hayamos construido una solución elegante utilizando los mejores productos y plataformas. Si la comunidad empresarial no ha abrazado el almacén de datos y ha continuado utilizándolo activamente seis meses después del entrenamiento, entonces hemos fallado en la prueba de aceptación. A diferencia de una reescritura del sistema operativo, en la que los usuarios empresariales no tienen otra opción que utilizar el nuevo sistema, el uso del almacén de datos es a veces opcional. La aceptación del usuario empresarial tiene más que ver con la simplicidad que con cualquier otra cosa (Kimball & Ross, 2002).

2.3.5. DATA MARTS

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento (Sinnexus, 2007).

2.3.6. ETL (Extracción, Transformación y Carga)

Los datos de origen para las aplicaciones de BI provendrán de una variedad de plataformas, que son administradas por una variedad de

sistemas operativos y aplicaciones. El propósito del proceso ETL es fusionar los datos de estas plataformas heterogéneas en un formato estándar para las bases de datos de BI en el entorno de soporte de decisiones de BI (Larissa T. Moss & Shaku Atre, 2003).

2.3.7. OLTP (Procesamiento Analítico en Línea)

OLAP es una capacidad enfocada en el análisis y exploración de información, mientras las herramientas de consulta y reportaje ponen mayor énfasis en el acceso a información, para propósitos de monitoreo. OLAP pasa del enfoque “qué” está pasando, a explorar “por qué” está pasando. Para descubrir el “por qué”, los usuarios pueden no saber exactamente qué información buscan y, más bien, navegarán, probarán un conjunto de información para descubrir detalles y patrones particulares (Howson, 2009).

2.3.8. PENTAHO

PENTAHO BI Open Source nos promete una gama de beneficios alentadores, haciendo mayor relieve en el costo, en la definición de estándares abiertos, en su flexibilidad, en sus funciones personalizadas y sobre todo en su arquitectura centrada en procesos, haciendo que las necesidades de la Inteligencia de Negocio sean atendidas con mayor facilidad (Gravitar, 2016).

2.3.9. MYSQL

MySQL es un sistema de administración de bases de datos relacional (RDBMS). Se trata de un programa capaz de almacenar una enorme cantidad de datos de gran variedad y de distribuirlos para cubrir las necesidades de cualquier tipo de organización, desde pequeños

establecimientos comerciales a grandes empresas y organismos administrativos. MySQL compite con sistemas RDBMS propietarios conocidos, como Oracle, SQL Server y DB2 (Gilfillan, 2003).



CAPITULO III



MATERIALES Y MÉTODOS

3.1 METODOLOGÍA DE LA INVESTIGACIÓN

Según (Hernández, Fernández, & Baptista, 2010). La investigación cuantitativa nos ofrece la posibilidad de generalizar los resultados más ampliamente, nos otorga control sobre los fenómenos, así como un punto de vista de conteo y las magnitudes de éstos. Asimismo, nos brinda una gran posibilidad de réplica y un enfoque sobre puntos específicos de tales fenómenos, además de que facilita la comparación entre estudios similares.

Por lo cual el presente proyecto corresponde a la metodología de investigación cuantitativa debido a su naturaleza que presenta y ser desarrollado en forma objetiva.

3.2 DISEÑO DE LA INVESTIGACIÓN

Según (Hernández et al., 2010). El diseño experimental se refiere a un estudio en el que se manipulan intencionalmente una o más variables independientes (supuestas causas-antecedentes), para analizar las consecuencias que la manipulación tiene sobre una o más variables dependientes (supuestos efectos-consecuentes), dentro de una situación de control para el investigador.

Por lo cual el diseño que se utilizara será el experimental ya que se tiene una variable independiente “Sistema de Soporte de Decisiones con Tecnología Data Warehouse” que causa efecto en la variable dependiente “Gestión de la Información de la Empresa Mallku Import SAC– Juliaca 2016”. Y dentro del experimental se opta por el cuasiexperimental.

Según (Hernández et al., 2010), en los diseños cuasiexperimentales los sujetos no se asignan al azar a los grupos ni se emparejan, sino que dichos

grupos ya están formados antes del experimento: son grupos intactos (la razón por la que surgen y la manera como se formaron es independiente o aparte del experimento).

3.3 POBLACIÓN Y MUESTRA

3.3.1. Población

Según (Hernández et al., 2010) delimitar la población que va a ser estudiada y sobre la cual se pretende generalizar los resultados. Es así que, una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones. Las poblaciones deben situarse claramente en torno a sus características de contenido, de lugar y en el tiempo.

De esta forma para el presente proyecto se ve la selección de la población conformada por un total de las 15 personas trabajadores de la Empresa Mallku Import SAC - Juliaca.

3.3.2. Muestra

Según (Hernández et al., 2010) básicamente categorizamos las muestras en dos: las muestras no probabilísticas y las muestras probabilísticas. Elegir entre una muestra probabilística o una no probabilística depende de los objetivos del estudio, del esquema de investigación y de la contribución que se piensa hacer con ella.

Por lo cual en la presente investigación se utilizó la técnica no probabilística con muestreo por conveniencia que es donde los sujetos son seleccionados según la conveniente accesibilidad y proximidad para el investigador. Teniendo como muestra en este caso cinco responsables de diferentes áreas y la gerencia general más

el dueño de la empresa, teniendo un total de seis personas.

3.4 MÉTODOS DE RECOPIACIÓN DE DATOS

Para la recolección de datos se utilizaron los siguientes métodos y técnicas:

3.4.1. La entrevista

Las entrevistas implican que una persona calificada (entrevistador) aplica el cuestionario a los participantes; el primero hace las preguntas a cada entrevistado y anota las respuestas. Su papel es crucial, es una especie de filtro. (Hernández et al., 2010).

Para lo cual se realizó un cuestionario previo que sirvió para guiar la entrevista cumpliendo el objetivo principal de la investigación, recabando información relevante para llevar a cabo la investigación. Las preguntas que se formularon están en el “Anexo A” y se detallará más adelante en la aplicación de la metodología la forma como se procedió.

3.4.2. Encuestas

Las encuestas fueron elaboradas por un conjunto de preguntas escritas por el investigador aplicadas a los involucrados en el desarrollo de la investigación, los cuales permitieron hacer las pruebas del sistema para su respectiva validación. El cuestionario que se utilizó para realizar estas encuestas está en el “Anexo H” y esto se desarrollará a detalle más adelante en la prueba de hipótesis.

3.5 METODO DE TRATAMIENTO DE DATOS

Para verificar la veracidad de la hipótesis, se aplicó la investigación experimental para medir el efecto de la variable independiente sobre la variable dependiente, teniendo la siguiente información:

- Recopilación de datos.
- Construcción de tablas de frecuencia.
- Interpretación de los resultados
- Prueba de hipótesis

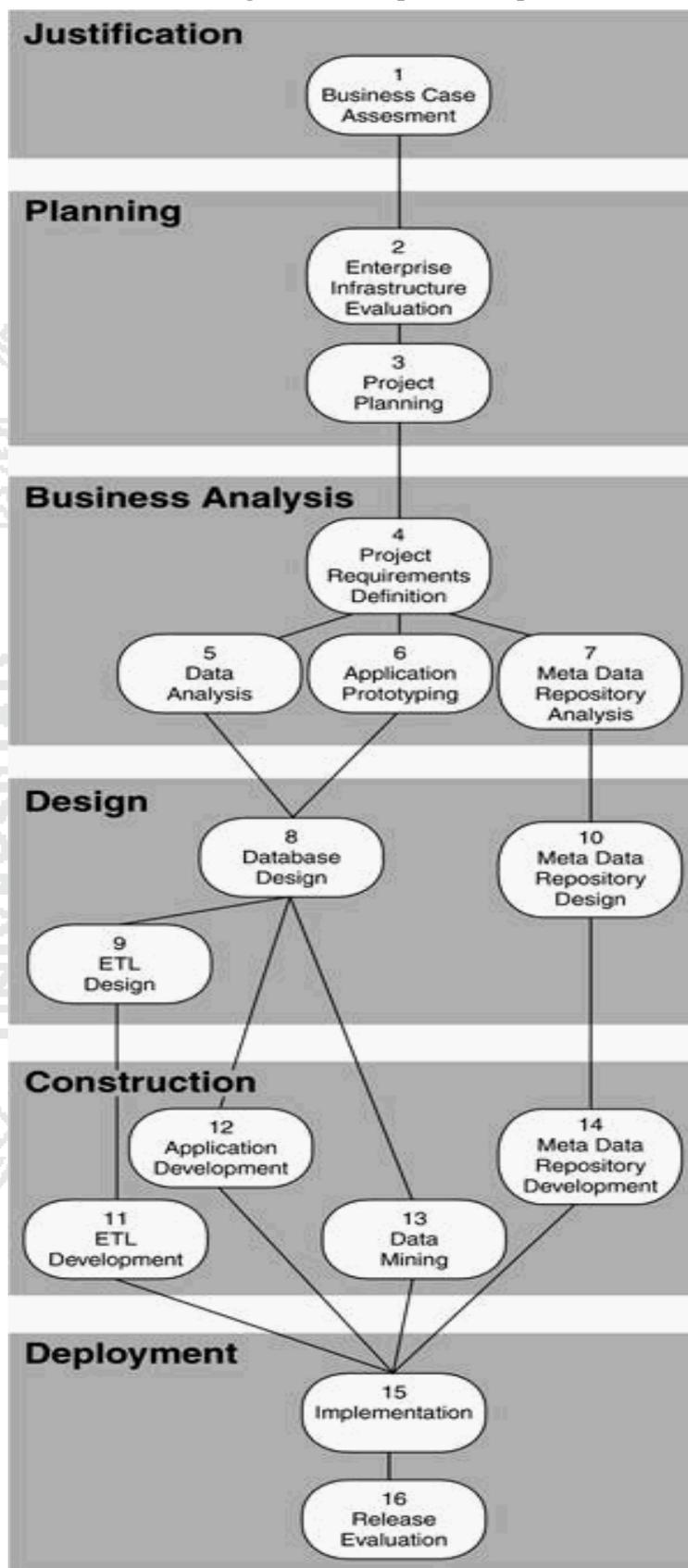
La recolección de los datos para el análisis se realizó mediante dos cuestionarios uno aplicado antes de la construcción del sistema y el otro después de la construcción del sistema.

3.6 MATERIAL EXPERIMENTAL

3.6.1. Metodología de desarrollo del Sistema

La metodología que se utiliza en la presente investigación es de Larissa T. Moss basada en inteligencia de negocios, el cual consiste en 6 etapas y 16 pasos como se muestra en la Figura N° 17, los cuales serán desarrollados más adelante.

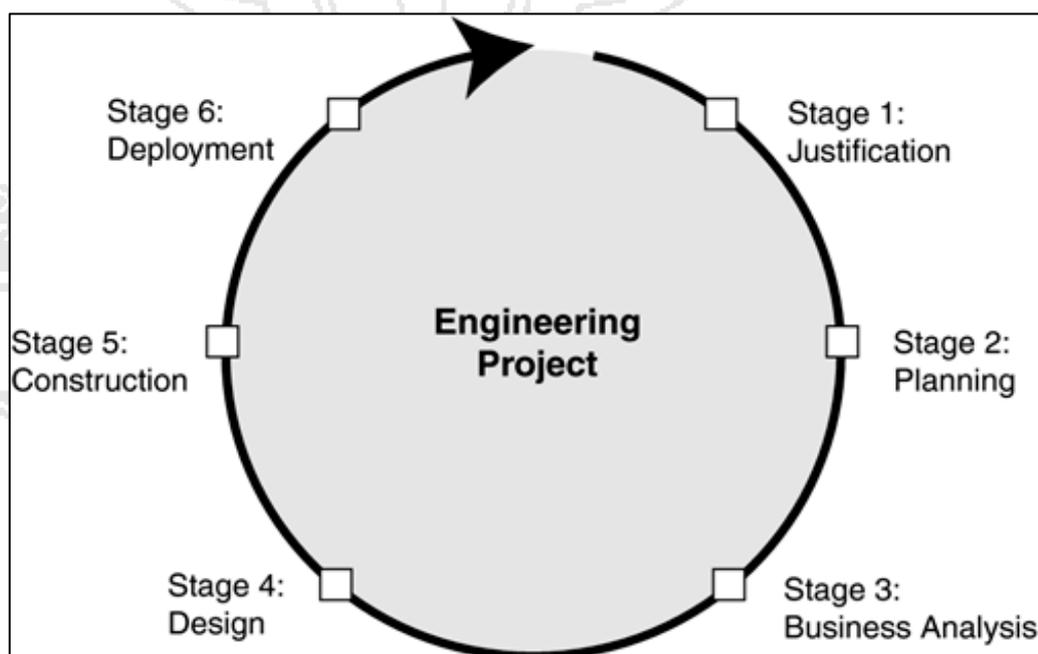
Figura N°17: Metodología utilizada para la implementación de BI.



Fuente: (Larissa T. Moss & Shaku Atre, 2003)

Según Moss, (2003, p.32) casi todo tipo de proyectos de ingeniería, tanto de ingeniería estructural como de ingeniería de software, pasa por seis etapas entre el inicio y la implementación, como se ilustra en la Figura 18. Por lo cual el presente proyecto se basará en esta metodología, siguiendo las diferentes etapas de ingeniería y pasos de desarrollo que plantea.

Figura N° 18: Etapas de ingeniería.



Fuente: (Larissa T. Moss & Shaku Atre, 2003)

Como indica la flecha de la figura 18, los procesos de ingeniería son iterativos. Una vez implementado, un producto se mejora continuamente en base a la retroalimentación de la comunidad empresarial que utiliza el producto. Cada iteración produce una versión de producto nueva a medida que el producto evoluciona y madura (Larissa T. Moss & Shaku Atre, 2003).

- **Etapas de Justificación:** Evaluar las necesidades del negocio que dan origen al proyecto de ingeniería.

- **Etapa de Planificación:** Desarrollar planes estratégicos y tácticos, que establecen cómo se desplegará y llevará a cabo el proyecto.
- **Etapa de Análisis del negocio:** Realizar un análisis detallado de los problemas y oportunidades del negocio para adquirir una comprensión sólida de los requisitos y llegar a una solución (producto).
- **Etapa de Diseño:** concebir un producto que resuelva el problema de negocio o provea oportunidades de negocio.
- **Etapa de Construcción:** construir el producto, el cual debe proporcionar un retorno de la inversión dentro de un lapso de tiempo definido.
- **Etapa de Implementación:** implementar el producto final, y medir su efectividad para determinar si la solución no cumple, cumple o excede con el retorno de la inversión esperado.

3.6.2. Etapas de ingeniería y pasos de desarrollo

Los proyectos de BI se organizan de acuerdo con las mismas seis etapas comunes a cada proyecto de ingeniería. Dentro de cada etapa de ingeniería, se llevan a cabo ciertos pasos para ver el proyecto de ingeniería hasta su terminación. Business Intelligence Roadmap describe 16 pasos de desarrollo dentro de estas etapas, como se describe a continuación (Larissa T. Moss & Shaku Atre, 2003):

a) Etapa de justificación

Paso 1: Evaluación del caso de negocio: se define el problema o la oportunidad del negocio, y se propone una solución de BI. Cada lanzamiento de una aplicación de BI debe justificar su costo y definir claramente sus beneficios, o la solución de un problema de negocio o el aprovechamiento de una oportunidad de negocio.

b) Etapa de planificación

Paso 2: Evaluación de infraestructura de la empresa: Ya que las aplicaciones de BI son iniciativas de toda la organización, esta debe crear una infraestructura para apoyarlas. Algunos componentes de la infraestructura pueden estar ya en su lugar, antes de que el primer proyecto de BI este en marcha. Otros componentes de la infraestructura pueden ser desarrollados con el tiempo, como parte de los proyectos de BI. Una infraestructura de la organización tiene dos componentes:

- **La infraestructura técnica:** que incluye hardware, software, middleware, sistemas de gestión de bases de datos, sistemas operativos, componentes de red, repositorios de metadatos, utilidades, etc.
- **Infraestructura no técnica:** que incluye estándares de metadatos, estándares de minería de datos, el modelo lógico empresarial (en evolución), metodologías, directrices, procedimientos de prueba, control de cambios y procesos, procedimientos para tareas administrativas y resolución de problemas, entre otros.

Paso 3: Planificación de Proyectos: los proyectos de BI son extremadamente dinámicos. Los cambios en el personal, en el presupuesto, en la tecnología, en los representantes del negocio y los patrocinadores, pueden afectar seriamente el éxito del proyecto. Por lo tanto, la planificación del proyecto debe ser detallada, y el progreso efectivo debe ser observado de cerca y reportado.

c) **Etapa de análisis del negocio**

Paso 4: Definición de requisitos del proyecto: administrar el alcance del proyecto es una de las tareas más difíciles en el transcurso del proyecto de BI. La necesidad de tener todo al instante es difícil de reducir, pero que se reduzca esta necesidad es uno de los aspectos más importantes en la negociación de los requisitos para cada entrega. Los integrantes de los equipos del proyecto deben saber que los requisitos cambian durante todo el ciclo de desarrollo, y los directivos deben conocer más sobre las posibilidades y las limitaciones de la tecnología de BI durante el desarrollo del proyecto.

Paso 5: Análisis de Datos: el mayor desafío de todos los proyectos de BI es la calidad de los datos de origen. Los malos hábitos desarrollados en las últimas décadas son difíciles de romper, y los daños provenientes de estos resultan muy caros, consumen mucho tiempo, y es tedioso encontrarlos y corregirlos. Además, el análisis de datos en el pasado se limitaba a la vista de una línea de negocio y nunca fue consolidada o conciliada con otros puntos de vista de la organización. Este paso requiere un porcentaje significativo del tiempo dedicado al calendario del proyecto completo.

Paso 6: Prototipo de la aplicación: el análisis de los resultados funcionales, que solía ser llamado análisis del sistema, se logra mediante los prototipos, por lo que se puede combinar con el diseño de aplicaciones. Las nuevas herramientas y lenguajes de programación permiten a los desarrolladores probar o refutar con relativa rapidez un concepto o una idea. Los prototipos también permiten a los empresarios

ver el potencial y los límites de la tecnología, lo que les da la oportunidad de ajustar los requisitos del proyecto y sus expectativas.

Paso 7: Análisis de repositorio de metadatos (Datawarehouse):

tener más herramientas significa tener más metadatos técnicos, además de los metadatos del negocio que suelen ser capturados mediante la ingeniería de software asistida por un ordenador de modelado de herramientas (CASE). Los metadatos técnicos necesitan ser asignados a los metadatos del negocio, y todos los metadatos deben ser almacenados en un repositorio de metadatos, estos últimos, pueden ser con licencia (comprados) o construidos. Los requisitos para que los tipos de datos sean capturados y almacenados, deben ser documentados en un modelo lógico de metadatos. Cuando se tienen las licencias de un producto de repositorio de metadatos, los requisitos documentados en este modelo lógico de metadatos deben ser comparados con el modelo metadatos del proveedor, si lo proporciona. Además, los requisitos para la entrega de los metadatos a la comunidad empresarial tienen que ser analizados.

d) Etapa de diseño

Paso 8: Diseño de bases de datos: uno o más objetivos de la base de datos de BI es almacenar de forma general y detallada los datos del negocio, dependiendo de las exigencias de la comunidad empresarial. No todos los requisitos de información son estratégicos y no todos son multidimensionales. Los esquemas de diseño de bases de datos deben coincidir con los requisitos de acceso a la información de la comunidad empresarial.

Paso 9: Diseño Extraer/Transformar/Cargar (ETL): el proceso ETL es el más complicado de todo el proyecto de BI, también es el menos glamoroso. Las ventanas de procesamiento ETL (ventanas de proceso por lotes) usualmente son pequeñas, sin embargo, debido a la mala calidad de la fuente de datos por lo general requiere mucho tiempo para ejecutar la transformación y los programas de limpieza. Acabar el proceso de ETL dentro del calendario previsto es un desafío para la mayoría de las organizaciones.

Paso 10: Diseño del repositorio de metadatos (Data Warehouse): si un repositorio de metadatos es comprado, lo más probable es que tenga que ser mejorado con características que fueron documentadas en el modelo lógico de metadatos, pero estas no se proporciona con el producto. Si se está construyendo un repositorio de metadatos, se debe tomar la decisión de si se diseña el repositorio de metadatos de la base de datos basado en entidad-relación u orientado a objetos. En cualquier caso, el diseño tiene que cumplir los requisitos del modelo lógico de metadatos.

e) **Etapa de construcción**

Paso 11: Desarrollo Extraer/Transformar/Cargar (ETL): muchas herramientas están disponibles para el proceso de ETL, algunas son sofisticadas y otras sencillas. Dependiendo de los requisitos para la limpieza y transformación de datos desarrollados en el paso 5, Análisis de Datos y en el Paso 9, Diseño ETL, una herramienta de ETL puede o no ser la mejor solución. En cualquier caso, se requiere con frecuencia el pre-procesamiento de los datos y la creación de ampliaciones para

complementar las capacidades de la herramienta de ETL.

Paso 12: Desarrollo de Aplicaciones: una vez que el prototipo concretó los requisitos funcionales, el verdadero desarrollo del acceso y el análisis de la aplicación puede empezar. El desarrollo de la aplicación puede ser una simple cuestión de la finalización de un prototipo operativo, o puede ser un esfuerzo de desarrollo que esté más involucrado con diferentes y más robustas herramientas de acceso y análisis. En ambos casos las actividades de desarrollo de aplicación front-end son realizadas generalmente en paralelo con las actividades de desarrollo de ETL back-end y el desarrollo del repositorio de metadatos.

Paso 13: Minería de datos: muchas organizaciones no utilizan el ambiente de BI en toda su extensión. Las aplicaciones de BI a menudo son limitadas a pre-escribir informes, algunos de los cuales incluso no son los nuevos tipos de informes, pero reemplazan los informes viejos. El retorno de la inversión real proviene de la información oculta en los datos de la organización, que sólo se puede descubrir con las herramientas de minería de datos.

Paso 14: Desarrollo del repositorio de metadatos: si se toma la decisión de construir un repositorio de metadatos en lugar de comprarlo, un equipo independiente se debe encargar del proceso de desarrollo. Esto se convierte en un sub-proyecto considerable en el proyecto global de BI.

f) Etapa de despliegue

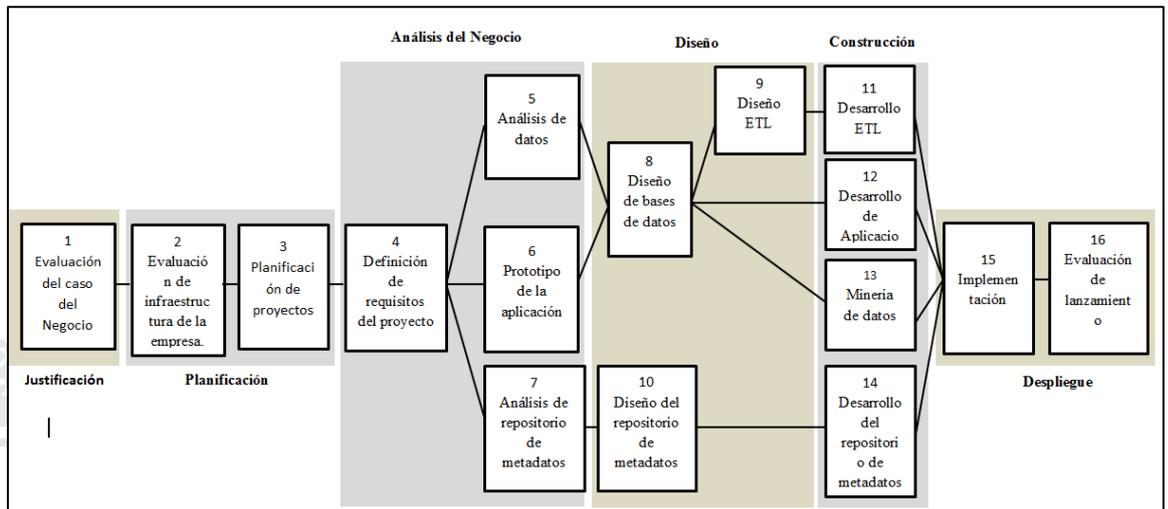
Paso 15: Implementación: una vez el equipo ha probado a fondo todos los componentes de la aplicación de BI, libera las bases de datos y aplicaciones. La formación está prevista para todo el personal del negocio y para otras personas que también utilizaran la aplicación de BI y el repositorio de metadatos. Las funciones de soporte que comienzan, incluyen operaciones desde mesa de ayuda, mantenimiento de las bases de datos de destino de BI, programación y ejecución de trabajos por lotes ETL, monitoreo del desempeño y puesta a punto de bases de datos.

Paso 16: Evaluación de lanzamiento: es muy importante beneficiarse de las lecciones y conceptos aprendidas de los proyectos anteriores para el lanzamiento de una aplicación. Cualquier incumplimiento de plazos, costos excesivos, conflictos y solución de conflictos deben ser examinados, y los procesos deben ajustarse antes de que comience la siguiente versión. Algunas herramientas, técnicas, pautas y procesos que no eran útiles deben ser reevaluados y ajustados o posiblemente descartados.

No es necesario realizar los pasos de desarrollo en secuencia, la mayoría de los equipos de proyecto los lleva a cabo en paralelo. No hay un orden natural de la progresión de una etapa de ingeniería a otra, sin embargo, existen ciertas dependencias entre algunas de las fases de desarrollo, como se muestra en la figura N° 19. Los pasos que se encuentran en el diagrama unos sobre otros, se puede realizar de forma simultánea, mientras que los pasos que aparecen a la derecha o a la

izquierda de cada uno, se llevan a cabo de manera relativamente lineal (con menos coincidencia) a causa de sus dependencias (Larissa T. Moss & Shaku Atre, 2003).

Figura N° 19: Dependencias en los pasos de desarrollo.



Fuente: (Larissa T. Moss & Shaku Atre, 2003).

Si bien algunos pasos de desarrollo son claramente específicos de cada proyecto, la mayoría de los pasos de desarrollo deben realizarse desde una perspectiva inter-organizacional. Por lo tanto, el enfoque de esas actividades de proyecto toma una dimensión multifuncional, y los revisores de esas actividades deben incluir representantes de negocios de otras líneas de negocio. La tarea principal de los representantes comerciales de las otras líneas de negocio es validar y ratificar las estrategias, políticas, reglas de negocio y estándares que se están utilizando o están siendo desarrollados durante el proyecto de BI. La Tabla 0.2 indica qué pasos son específicos del proyecto y cuáles son transversales a la organización (Larissa T. Moss & Shaku Atre, 2003).

Tabla 04: Pasos específicos del proyecto versus pasos de toda la organización.

Pasos de desarrollo	Proyecto-Específico versus Toda la Organización
1. Evaluación de casos de negocio	Toda la organización
2. Evaluación de Infraestructura Empresarial (Técnica y no técnica)	Toda la organización
3. Planificación de proyectos	Proyecto específico
4. Definición de los requisitos del proyecto	Proyecto específico
5. Análisis de los datos	Toda la organización
6. Prototipos de Aplicaciones	Proyecto específico
7. Análisis del repositorio de metadatos	Toda la organización
8. Diseño de base de datos	Toda la organización
9. Diseño ETL	Toda la organización
10. Diseño del repositorio de metadatos	Toda la organización
11. Desarrollo ETL	Toda la organización
12. Desarrollo de aplicaciones	Proyecto específico
13. Minería de datos	Toda la organización
14. Desarrollo de Repositorio de Metadatos	Toda la organización
15. Implementación	Proyecto específico
16. Evaluación de liberación	Toda la organización

Elaboración: Propia.

3.6.3. Hardware y Software de Desarrollo

Hardware

- Una PC
- Una laptop
- Una impresora
- Una memoria USB 8GB

Software

- Microsoft Word

- MySQL Workbench
- Pentaho Business Intelligence Server
- Pentaho Data Integration Kettle Spoom
- Pentaho Análisis Server Mondrian
- Pentaho Report Designer



CAPITULO IV



RESULTADOS Y DISCUSIÓN

4.1. APLICACIÓN DE LA METODOLOGÍA

4.1.1. JUSTIFICACIÓN

4.1.1.1 Evaluación del caso del negocio

La empresa Mallku Import SAC, dedicada al rubro de la importación para la venta de máquinas remanufacturadas como son fotocopiadoras, impresoras, repuestos, insumos y accesorios. Y ya viene más de dos años funcionando formalmente y no ha tenido un crecimiento en el mercado manteniéndose en forma regular, sin poder lograr sus metas y objetivos, preocupando al dueño y directivos de la empresa debido a que también existen muchas fallas en cuanto a sus operaciones como son de ventas, compras o mantenimiento que ocurren a diario.

La empresa tiene información que almacena a diario en archivos de Excel, Word y un pequeño sistema de ventas en Acces, pero solo hace eso almacenar la información sin tener ningún otro fin, para lo cual teniendo la base analizada sobre Sistemas basado en Inteligencia de Negocios que justamente hace uso de esta información que se almacena a diario en diferentes archivos, se plantea a la empresa poder desarrollar un Sistema de Soporte de Decisiones con tecnología Data Warehouse para optimizar la Gestión de la Información, donde el dueño, el gerente y los responsables de cada área puedan tener un soporte que les permita tomar decisiones acertadas, tener la información en el momento oportuno y adecuado. Y poder de esta forma alcanzar las metas y objetivos de la empresa.

4.1.2. PLANIFICACIÓN

4.1.2.1 Evaluación de infraestructura de la empresa

a) Infraestructura técnica

El Área de la Gerencia donde se realizara la implementación del sistema cuenta con los siguientes recursos:

Tabla N° 05: Características de hardware y software de la empresa.

Hardware
<ul style="list-style-type: none"> - 01 computadora de escritorio con las siguientes características: <ul style="list-style-type: none"> - Procesador Intel Corei5 - Memoria Ram 8 GB - Disco duro 1 TB - Video y audio integrado - Monitor LG 20" - Quemador CD/DVD - Mouse, teclado y parlantes - 01 laptop con las siguientes características: <ul style="list-style-type: none"> - Marca Toshiba - Procesador Corei5 - Memoria Ram 6 GB - Disco duro 1 TB - Quemador CD/DVD - 01 fotocopidora multifuncional kyosera FS-3640. - 01 impresora a color laser Ricoh Aficio SP C430dn - 01 Router de telefónica (que provee internet). - 01 Switch de 16 puertos (que distribuye la red e internet a los demás departamentos de la empresa).
Software
<ul style="list-style-type: none"> - Sistema Operativo Windows 8 - Microsoft Office 2013 - Adobe Reader 9 - Internet Explorer 9.1 - Mozilla Firefox 11 - Avast Antivirus - CCleaner - Nero Express

Elaboración: Propia.

Según la evaluación y los requerimientos mínimos para hacer la instalación del sistema se ve por conveniente que en la parte del hardware no es necesario hacer alguna adquisición de momento. Por lo cual en cuanto al software se hará las respectivas instalaciones que permitan que el sistema se implemente de manera correcta. Se precisa que la mayoría del software a utilizar es Open Source. Entre los cuales tenemos:

- MySQL server
- Servidor Tomcat
- Pentaho Suite

b) Infraestructura no técnica

La información obtenida por parte de la gerencia de la empresa es que no existe ningún sistema avanzado, ni de ningún tipo de DBMS. Va ser la primera vez que se va implementar un Sistema de este tipo.

4.1.2.2 Planificación del proyecto

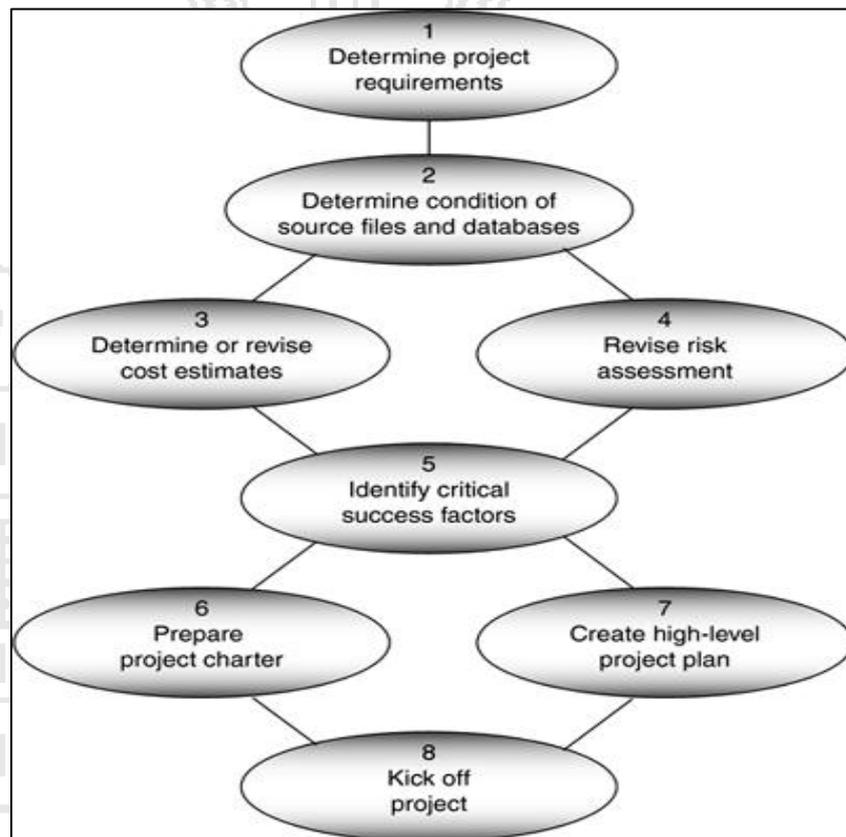
Para realizar una correcta planificación para el desarrollo del proyecto se ve por conveniente desarrollar las siguientes actividades (Larissa T. Moss & Shaku Atre, 2003):

- Determinar los requisitos del proyecto.
- Determinar la condición de los archivos de origen y bases de datos.
- Determinar o revisar las estimaciones de costos.
- Revisar la evaluación de riesgos.
- Identificar factores críticos de éxito.
- Preparar la carta del proyecto.
- Crear un plan de proyecto de alto nivel.

- Inicio del proyecto.

Las actividades de planificación del proyecto no necesitan realizarse de forma lineal. La Figura N° 20 indica qué actividades se pueden realizar simultáneamente (Larissa T. Moss & Shaku Atre, 2003):

Figura N° 20: Actividades de planificación de proyectos.



Fuente: (Larissa T. Moss & Shaku Atre, 2003).

Tabla N° 06: Actividades de planificación del proyecto.

Actividad	Días											
	1	2	3	4	5	6	7	8	9	10	11	
Determinar los requisitos del proyecto.	X	X										
Determinar la condición de los archivos de origen y bases de datos.			X	X								
Determinar o revisar las estimaciones de costos.					X	X						
Revisar la evaluación de riesgos.					X	X						
Identificar factores críticos de éxito.							X	X				
Preparar la carta del proyecto.									X			
Crear un plan de proyecto de alto nivel.									X	X		
Inicio del proyecto.												X

Elaboración: Propia

Se debe considerar que el desarrollo del proyecto puede tener algunos retrasos por algunos factores que se debe tomar en cuenta pudiendo ser estos:

- Tiempo: No cumplir con los plazos establecidos.
- Económico: Falta de presupuesto para emprender el proyecto.
- Social: Huelgas, paros.
- Salud: Que los responsables del proyecto sufran una descompensación.
- Climático y/o algún desastre natural.

También cabe resaltar que el proyecto y las fechas establecidas no siempre serán exactas sino son suposiciones que se tienen que asumir con cierto riesgo que no pueda perjudicar el proyecto para lo cual se ve la

experiencia en proyectos llevados a cabo anteriormente que sean similares.

4.1.3. ANÁLISIS DEL NEGOCIO

4.1.3.1 Definición de los requisitos del proyecto

Según (Larissa T. Moss & Shaku Atre, 2003) la recopilación de requisitos para una entrega concreta de un proyecto se centra en definir las necesidades explícitas de negocio del patrocinador comercial para quien se está desarrollando la aplicación de BI. Los requisitos del proyecto deben ser establecidos en términos comerciales y deben describir el problema de negocio a ser resuelto, así como los criterios de aceptación de la solución de BI.

Para poder recabar la información se trabajó con el personal que se seleccionó en la muestra, los cuales tienen mayor implicancia en la toma de decisiones conformado de la siguiente manera:

- Dueño de la empresa.
- Gerente general.
- Responsable de ventas.
- Responsable de compras.
- Responsable de mantenimiento.
- Responsable de contabilidad y finanzas.

Con este grupo de personas se trabajó en fin de obtener la información necesaria para el desarrollo del sistema, donde analizando la situación y teniendo en consideración el enfoque de Ralph Kimball que también es conocido como el enfoque “Bottom-up”, porque parte de la premisa que el Data Warehouse se construye a partir de un Datamart que está orientado a un proceso de negocio y normalmente está referido a un

tema puntual, para luego desarrollar otros Datamarts haciendo que las dimensiones sean comunes y sean las fuentes de integración.

Partiendo entonces de este panorama se realizó la obtención de requerimiento centrándonos principalmente en el objetivo principal más próximo de la empresa, donde se vio por conveniente partir del área de Ventas, seguidamente de Compras donde se realizó lo siguiente:

- a) Se realizaron entrevistas a los empleados según el cargo y responsabilidades de cada área, para lo cual se utilizó un cuestionario con preguntas de interés de la investigación. Ver **Anexo N° 1**.
- b) Se procedió a dar un recorrido para la revisión de los diferentes procesos que se realizan y tener una mejor visión para beneficio de la investigación.
- c) Se accedió a las diferentes fuentes y registros que maneja la empresa actualmente, siendo estos archivos de Excel, Word y un sistema de ventas en Acces, los cuales serán analizados y procesados y puedan ser de gran utilidad para el desarrollo del sistema.
- d) Finalmente se establecieron responsabilidades para cumplir con las actividades a realizar para que la construcción del sistema tenga éxito.

A continuación detallamos la información obtenida, creando un modelo conceptual para su mejor entendimiento de acuerdo a cada datamars:

- **Datamart de Ventas:**

Se comenzó el análisis por esta área debido a que por su función es considerada la más principal y la encargada de generar la mayor

recaudación de ingresos económicos para la empresa. Teniendo las siguientes incidencias:

Tabla N° 07: Incidencias del área de ventas.

ITEM	INCIDENCIAS
1	Actualmente vienen registrando las diferentes ventas diarias en un sistema hecho en Acces.
2	La información registrada es impresa y compartida en caso sea requerida por la gerencia para su análisis respectivo.
3	Existe un interés por mejorar el procesamiento de esta información para lograr las metas de esta área y de la empresa, tener información de los clientes, las veces que acuden a realizar una compra, que productos más compran y así tener en stock para que estos productos no falten y de esta forma poder aumentar las ventas y generar mayores ingresos. Complaciendo no también a los nuevos clientes y poder moderar los precios según el mercado que ya será cuestión de los encargados en tomar las decisiones.
4	Por lo cual se desea saber con precisión las cantidades vendidas por productos, por clientes y cada cierto periodo de tiempo. Como también los montos que generan.
5	Actualmente es complejo el procesamiento de esta información, no permitiéndoles cumplir con sus metas y objetivos.

Elaboración: Propia.

Del cuadro anterior determinamos del Ítem 4 los principales requerimientos formulando de la siguiente manera:

- Cantidad de unidades vendidas de cada producto a cada cliente por ciertos periodos de tiempo.
- Monto total que generan estas ventas de estos productos con sus respectivos clientes por ciertos periodos de tiempo.

Ahora determinamos los indicadores y perspectivas:

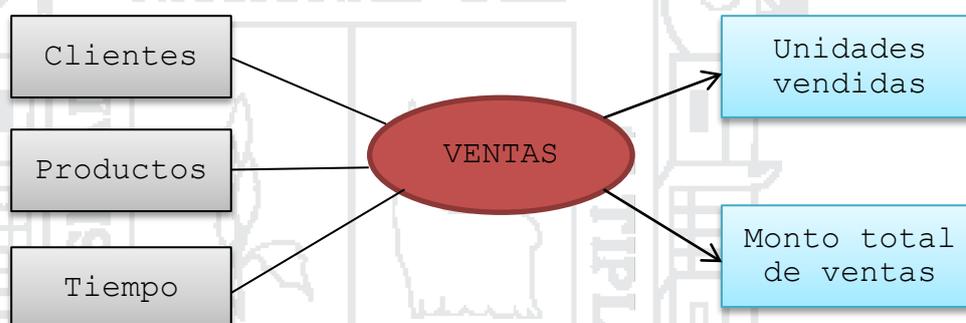
Indicadores:

- Unidades vendidas.
- Monto total de ventas.

Perspectivas:

- Clientes.
- Productos.
- Tiempo.

Figura N° 21: Modelo conceptual ventas.



Elaboración: Propia.

Datamart de Compras:

El área de compras es la responsable de adquirir los productos e insumos para que la empresa pueda realizar sus diferentes operaciones donde encontramos las siguientes incidencias:

Tabla N° 08: Incidencias del área de compras.

ITEM	INCIDENCIAS
1	Las compras las realizan de acuerdo a la solicitud o escases de algún producto.
2	Cuentan con un registro de proveedores con el detalle y descripción de cada producto a adquirir almacenado en Excel.
3	Muchas veces las compras las realizan cuando ya no existe el producto en stock.
4	Los productos son solicitados para su compra vía internet o telefónica haciendo luego el depósito a la cuenta respectiva del proveedor según sea el caso para luego ser enviados.
5	No existe un control claro de las cantidades que se están comprando por proveedor ni los montos exactos que implica estas compras durante ciertos periodos de tiempo, para poder tener un balance y comparaciones por proveedor y costos.

Elaboración: Propia.

Del cuadro anterior determinamos del ítem 5 los principales requerimientos:

- Cantidad de unidades compradas por producto, por proveedor y cada cierto periodo de tiempo.
- Monto total que genero estas compras de productos por proveedor cada cierto periodo de tiempo.

Indicadores

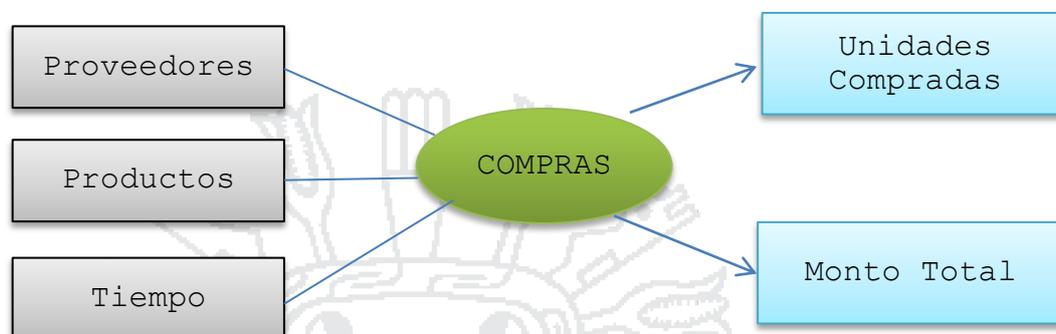
- Unidades compradas.
- Monto Total.

Perspectivas

- Proveedores.

- Productos.
- Tiempo.

Figura N° 22: Modelo conceptual compras.



Elaboración: Propia.

4.1.3.2 Análisis de datos

Para el análisis respectivo de los datos se toma cada datamarts de acuerdo al área para poder especificar los diferentes indicadores y perspectivas para su respectivo análisis según las especificaciones de los requerimientos vistos anteriormente, teniendo lo siguiente:

Datamart de Ventas:

- Las unidades vendidas de cada producto a cada cliente en un tiempo determinado.
- El monto total vendido por producto a cada cliente en un tiempo determinado.

Indicadores:

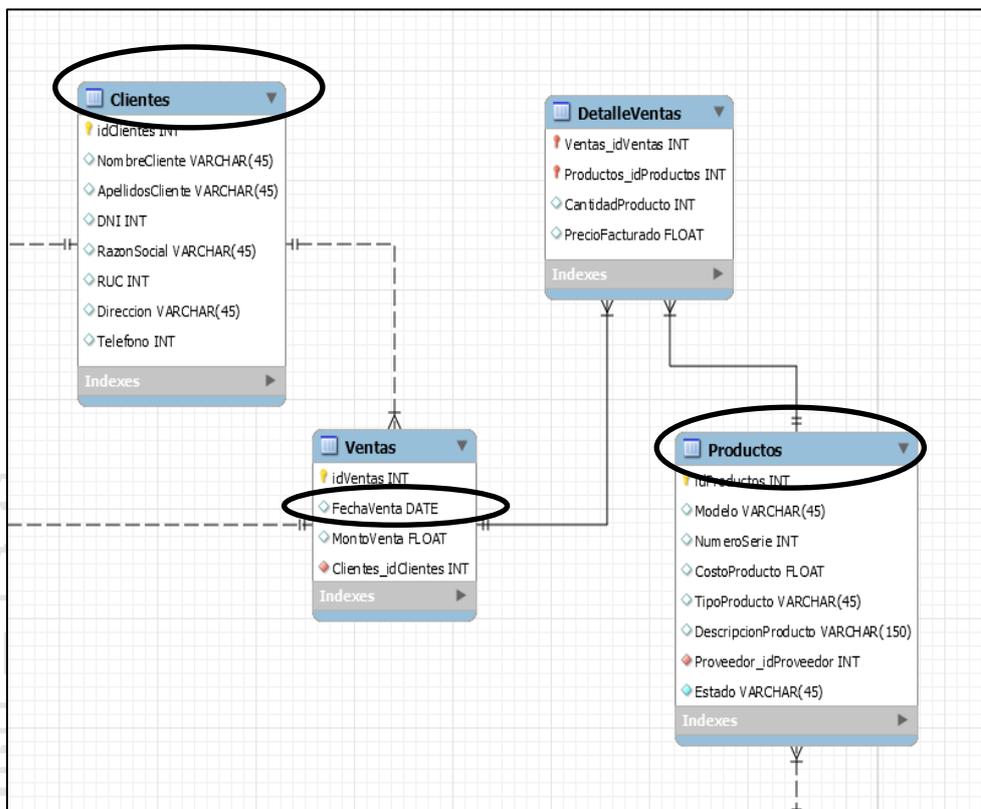
- Unidades vendidas.
- Monto total vendido.

Perspectivas:

- Clientes.
- Productos.

- Tiempo.

Figura N° 23: Datamart de ventas.



Elaboración: Propia.

Datamart Compras

- Cantidad de unidades compradas por producto, por proveedor y cada cierto periodo de tiempo.
- Monto total que genero estas compras de productos por proveedor cada cierto periodo de tiempo.

Indicadores

- Unidades compradas.
- Monto total.

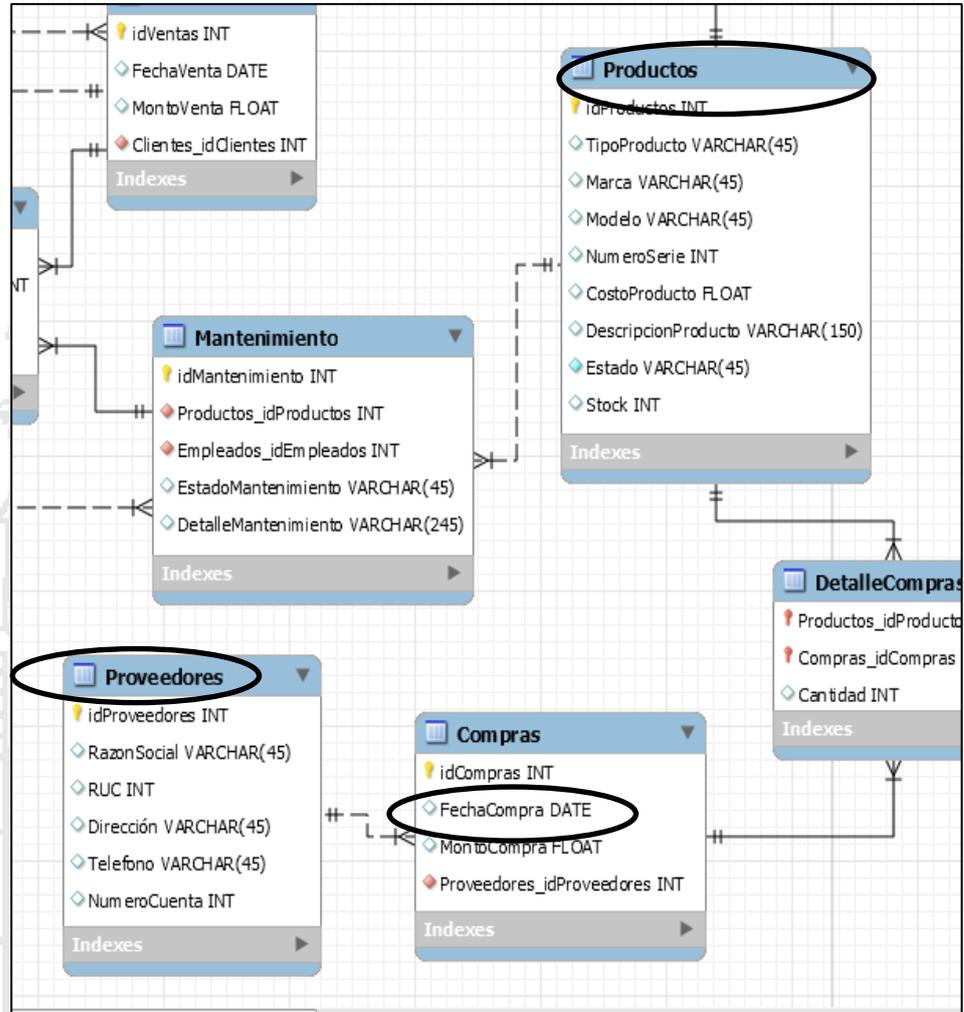
Perspectivas

- Proveedores.
- Productos.

- Tiempo.

Seleccionamos de donde se van a analizar y extraer los datos.

Figura N° 24: Datamart de compras.



Elaboración: Propia.

4.1.3.3 Prototipo de la aplicación

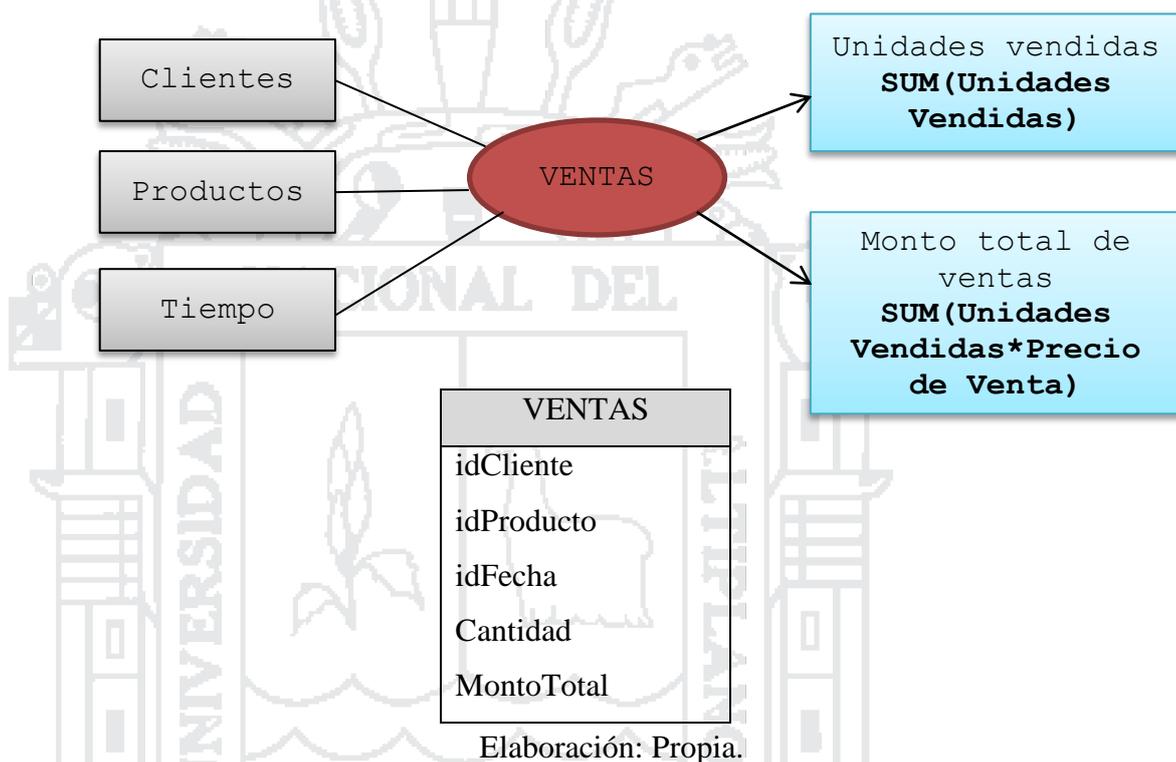
Mostraremos a continuación la confección de nuestras tablas de hechos para cada datamarts:

Datamarts de VENTAS:

- “VENTAS”, con la clave principal que estará conformado por las claves principales de las diferentes dimensiones ya antes definidas que son “idCliente”, “idProducto” e “idFecha”.

- Los hechos que se crearon y definieron anteriormente son “Cantidad” y “Monto Total”.
- En el grafico siguiente se especifica más ampliamente como estará conformado nuestra tabla de hechos de VENTAS.

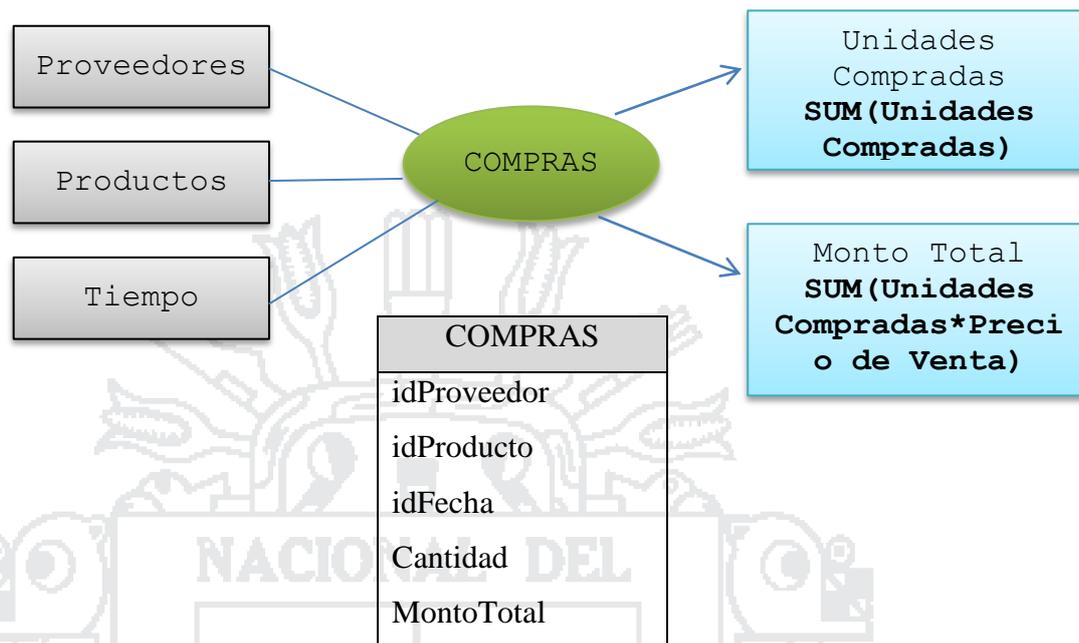
Figura N° 25: Prototipo de datamart de ventas.



Datamarts de COMPRAS

- “COMPRAS”, con la clave principal que estará conformado por las claves principales de las diferentes dimensiones ya antes definidas que son “idProveedor”, “idProducto” e “idFecha”.
- Los hechos que se crearon y definieron anteriormente son “Cantidad” y “Monto Total”.
- En el grafico siguiente se especifica más ampliamente como estará conformado nuestra tabla de hechos de COMPRAS.

Figura N° 26: Prototipo de datamarts de compras.



Elaboración: Propia.

4.1.3.4 Análisis de repositorio de metadatos (Data Warehouse)

Antes de avanzar a toda velocidad con importantes gastos en almacenes de datos, es prudente tomar un momento para evaluar la disposición de la organización a continuar. Sobre la base de nuestra experiencia acumulada de cientos de almacenes de datos, hemos identificado cinco factores que diferencian los proyectos que predominaban la navegación sin problemas en comparación con aquellos que implicaban una lucha constante. Estos factores son indicadores principales del éxito del almacén de datos. Usted no necesita altas calificaciones en cada factor para avanzar, pero cualquier déficit representa riesgos o vulnerabilidades. Vamos a describir las características en orden de importancia (Kimball & Ross, 2002).

El factor más crítico para el almacenamiento de datos exitoso es

tener un patrocinador de negocios fuerte. Los patrocinadores de negocios deben tener una visión para el impacto potencial de un almacén de datos en la organización. Deben ser apasionados y personalmente convencidos del valor del proyecto, mientras que realistas al mismo tiempo. De manera óptima, el patrocinador del negocio tiene un historial de éxito con otras iniciativas internas. Él o ella debe ser un líder políticamente astuto que puede convencer a sus pares para apoyar el almacén.

El segundo factor de preparación es tener una motivación empresarial fuerte y convincente para construir un almacén de datos. Este factor a menudo va de la mano con el patrocinio. Un proyecto de almacén de datos no puede simplemente ofrecer una capacidad agradable de tener; Necesita solucionar problemas críticos de negocio con el fin de reunir los recursos necesarios para un lanzamiento exitoso y una vida útil saludable. La motivación compulsiva típicamente crea un sentido de urgencia, ya sea que la motivación provenga de fuentes externas (por ejemplo, factores competitivos) o internas (por ejemplo, incapacidad para analizar el desempeño entre organizaciones después de adquisiciones).

El tercer factor al evaluar la preparación es la factibilidad. Existen varios aspectos de factibilidad, como la factibilidad técnica o de los recursos, pero la factibilidad de los datos es la más crucial. ¿Estamos recopilando datos reales en sistemas operativos reales de origen para dar soporte a los requerimientos del negocio? La factibilidad de los datos es una preocupación importante porque no hay una corrección a corto plazo si no estamos recopilando datos de fuentes razonablemente limpias en la

granularidad correcta.

Los siguientes factores no son proyecciones de proyectos, pero aún influyen en su probabilidad de éxito. El cuarto factor se centra en la relación entre las empresas y las organizaciones de TI. ¿En su empresa, la organización de TI entiende y respeta el negocio? Por el contrario, ¿entiende y respeta la organización de TI? La incapacidad de responder honestamente a estas preguntas no significa que no pueda continuar. Más bien, implica que usted necesita para mantener vigilante el negocio y los representantes de TI marchando al mismo tiempo. De muchas maneras, la iniciativa de almacén de datos puede ser una oportunidad para reparar la valla entre estas organizaciones, asumiendo que ambos entregan.

Según esta percepción para el análisis respectivo del repositorio de metadatos que es fundamental para el desarrollo del sistema, se ve por conveniente utilizar MySQL Server, que será instalado en la computadora que se encuentra en el área de gerencia y pueda servir de repositorio donde se almacenarán las diferentes tablas de hechos y dimensiones que se crearan más adelante.

4.1.4. DISEÑO

4.1.4.1 Diseño de la base de datos

Modelo lógico

Para confeccionar el modelo lógico de la estructura del Data Warehouse se tendrá como base los modelos conceptuales elaborados en la parte de análisis de requerimientos, para lo cual se tendrá como modelador el programa MySQL Server Workbench. Primeramente se definirá el modelo a utilizar y seguidamente se diseñaran las tablas de

dimensiones y de hechos para cada datamarts independientemente, para finalmente unir estas tablas.

a) Tipo de modelo lógico

Para el diseño lógico se opta por el esquema en estrella que según (Díaz, 2010) consiste en estructurar la información en procesos, vistas y métricas recordando a una estrella (por ello el nombre). A nivel de diseño, consiste en una tabla de hechos (lo que en los libros encontraremos como fact table) en el centro para el hecho objeto de análisis y una o varias tablas de dimensión por cada punto de vista de análisis que participa de la descripción de ese hecho. En la tabla de hecho encontramos los atributos destinados a medir (cuantificar): sus métricas. La tabla de hechos sólo presenta uniones con dimensiones.

b) Tablas de dimensiones

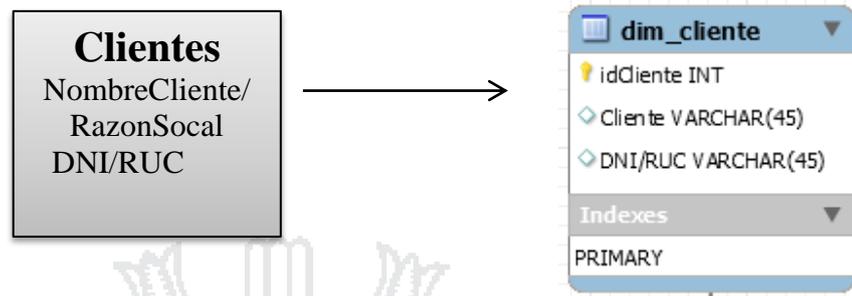
Vamos a comenzar el diseño de las tablas de dimensiones según los datamarts analizados anteriormente. Donde se partirá de las perspectivas encontradas en el análisis de requerimientos.

Datamart de ventas

- Cientes

- La nueva tabla de dimensión tendrá de nombre “dim_clientes”
- La clave principal que se le asignará es “idClientes”
- Se modificará el nombre de campo “NombreCliente/RazonSocial” por solo “Cliente” y el campo “DNI/RUC” seguirá “DNI/RUC”.

Figura N° 27: Diseño de la dimensión cliente para ventas.

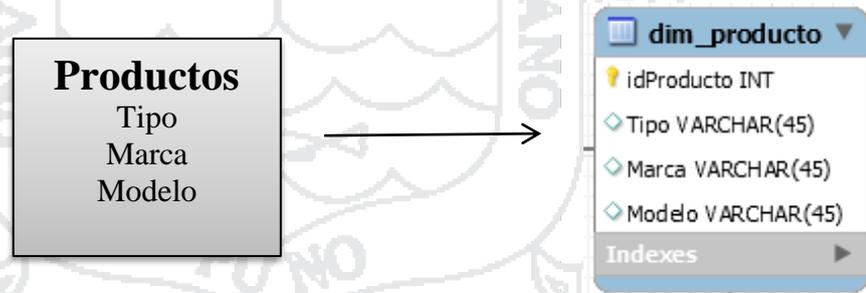


Elaboración: Propia.

- Productos

- La nueva tabla de dimensión tendrá el nombre de “dim_producto”
- La clave primaria que se le asignará es “idProducto”.
- Los campos permanecerán con sus nombres: “Tipo”, “Marca” y “Modelo”.

Figura N° 28: Diseño de la dimensión producto para ventas.

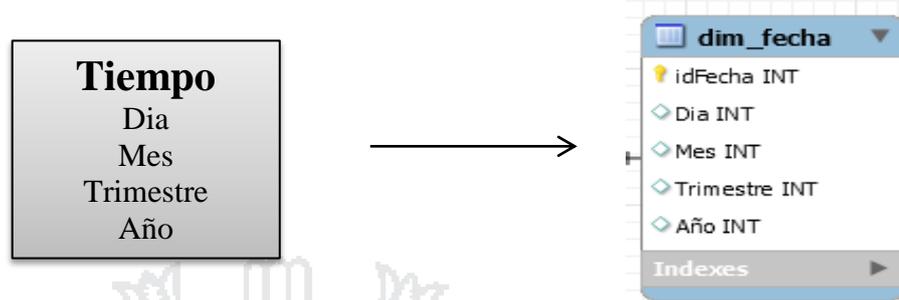


Elaboración: Propia.

- Tiempo

- La nueva tabla de dimensión tendrá el nombre de “dim_fecha”.
- La clave primaria que se le asignará es “idFecha”
- Los campos serán “Dia”, “Mes”, “Trimestre”, “Año”.

Figura N° 29: Diseño de la dimensión fecha para ventas.



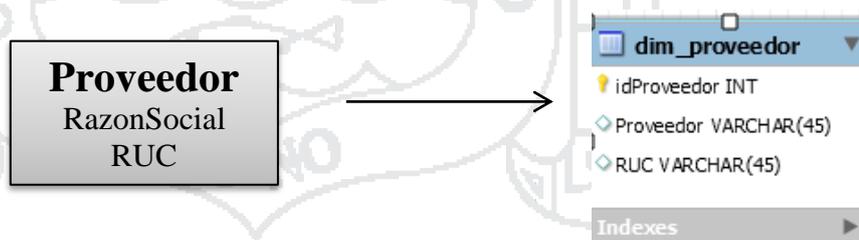
Elaboración: Propia.

Datamarts de compras

- Proveedores

- o La nueva tabla de dimensión tendrá el nombre de “dim_proveedor”.
- o La clave primaria que se le asignará es: “idProveedor”.
- o Tendrá un solo campo que será: “Proveedor”

Figura N° 30: Diseño de la dimensión proveedor para compras.

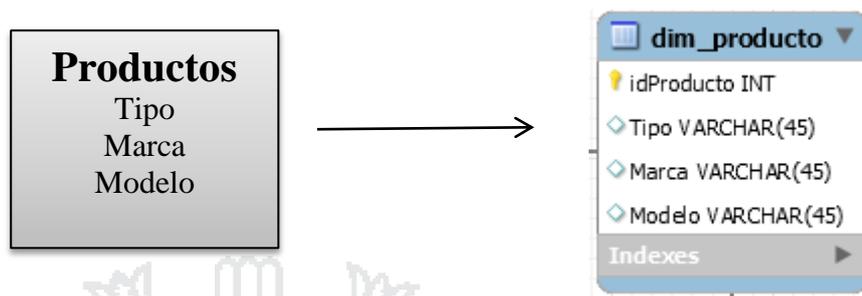


Elaboración: Propia.

- Productos

Se utilizara la dimensión Producto creada para el datarmart de ventas.

Figura N° 31: Diseño de la dimensión producto para compras.

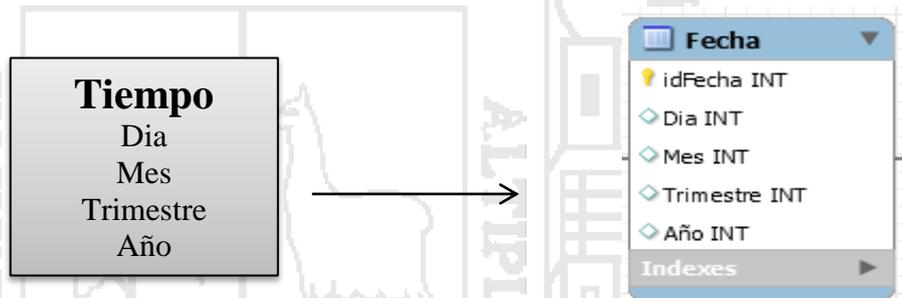


Elaboración: Propia.

- Tiempo

De igual forma también se utilizara la dimensión Fecha creada en el datamart de ventas.

Figura N° 32: Diseño de la dimensión fecha para compras.



Elaboración: Propia.

c) Tablas de hechos

Para la creación de las tablas de hechos el procedimiento será el mismo para cada datamarts, partiendo de los indicadores obtenidos en el análisis de requerimientos.

Datamarts de ventas

- La tabla de hechos tendrá como nombre “Ventas”.
- La clave principal será la combinación de las claves primarias de las tablas de dimensiones antes definidas: “idCliente”, “idProducto”, e “idFecha”.

- Se crearan dos hechos de acuerdo a los indicadores establecidos y se cambiaron como se indica: “Unidades Vendidas” por “Cantidad” y “Monto Total Vendido” por “MontoTotal”.

Figura N° 33: Tabla de hechos para ventas.



Elaboración: Propia.

Datamarts de compras

- La tabla de hechos tendrá como nombre “Compras”
- La clave principal será la combinación de las claves primarias de las tablas de dimensiones antes definidas: “idProveedor”, “idProducto” e “idFecha”.
- Se crearan dos hechos de acuerdo a los indicadores establecidos y se cambiaron como se indica: “Unidades Compradas” por “Cantidad”, “Monto Total” por “MontoTotal”.

Figura N° 34: Tabla de hechos para compras.



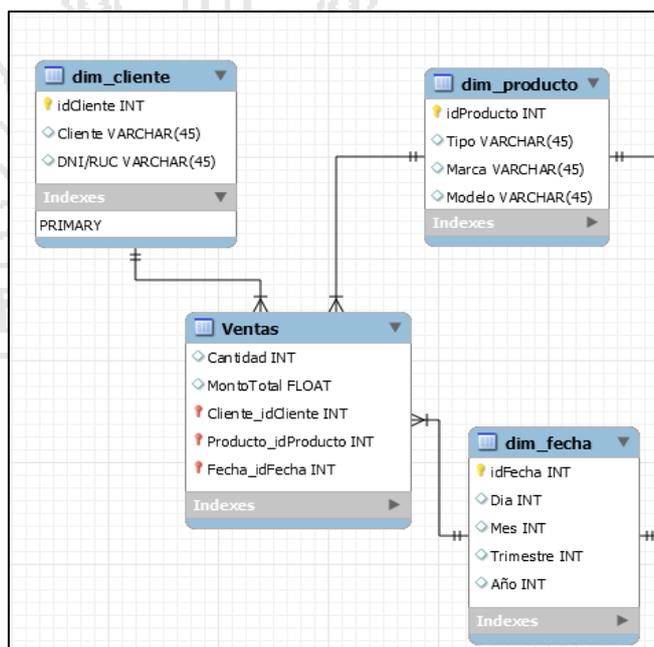
Elaboración: Propia.

d) Uniones

Finalmente se realizan las respectivas uniones entre las tablas de dimensiones con las tablas de hechos de acuerdo a cada datamarts.

Datamart de ventas

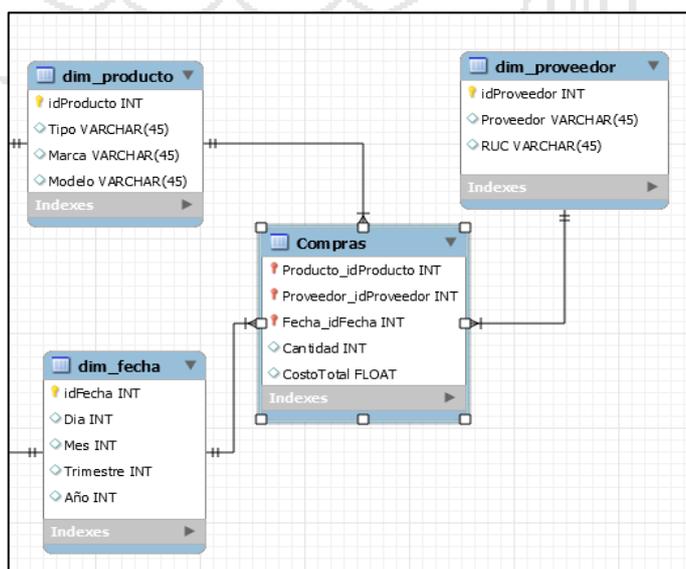
Figura N° 35: Unión de tablas para ventas.



Elaboración: Propia

Datamarts de compras

Figura N° 36: Unión de tablas para compras.



Elaboración: Propia.

4.1.4.2 Diseño del ETL

El proceso de Extracción, Transformación y Carga se realizara de igual forma como se vino realizando los anteriores procesos por cada datamarts.

Datamarts de Ventas

Carga de la dimensión cliente

a) Descripción

Se realizará la carga de la dimensión “dim_clientes” desde la tabla “Clientes” que se encuentra en la base de datos que maneja la empresa.

b) Descripción de las tabla fuente

Tabla N° 09: Tabla fuente cliente para ventas.

Tipo de Fuente	Nombre de la Tabla	Descripción
Base de datos de la empresa (MySQL)	Clientes	La tabla contiene todos los datos concernientes a los clientes de la empresa.

Elaboración: Propia.

c) Estandarización y limpieza de datos

Tabla N° 10: Tabla estandarización de datos de cliente para ventas.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
idCliente	PK	Int	Numérico	No debe ser nulo	NO TIENE
Cliente		Varchar(45)	Texto		NO TIENE
DNI/RUC		Varchar(45)	Texto		NO TIENE

Elaboración: Propia.

d) Fuentes de datos

Tabla N° 11: Tabla fuente de datos de cliente para ventas.

Tabla: Clientes				
Nombre	Llave	Tipo	Formato	Consideración importante
idClientes	PK	Int	Numérico	NO TIENE
NombreCliente/ RazonSocial		Varchar(45)	Texto	NO TIENE
DNI/RUC		Varchar(45)	Texto	NO TIENE
Direccion		Varchar(45)	Texto	NO TIENE
Telefono		Int	Numérico	NO TIENE

Elaboración: Propia.

e) Tabla destino

Tabla N° 12: Tabla destino de cliente para ventas.

Tabla: InterCliente		
Campo	Tipo	Mapeo
idCliente	Int	Clientes.idCliente
Cliente	Varchar(45)	Clientes.Cliente
DNI/RUC	Varchar(45)	Clientes.Empresa

Elaboración: Propia.

f) Proceso

i. Carga de registros de un archivo intermedio

Se extraen los datos de la tabla “Clientes” de acuerdo al mapeo y se carga en un archivo de texto intermedio “InterCliente”.

ii. Carga de la dimensión

Se extraen los valores del archivo “InterCliente” y se carga en la dimensión “dim_cliente”.

iii. Borrar el archivo intermedio

Finalmente se borra el archivo intermedio “InterCliente”

Carga de la dimensión producto

a) Descripción

Se realizará la carga de la dimensión “dim_producto” desde la tabla “Productos” que se encuentra en la base de datos que maneja la empresa.

b) Descripción de las tabla fuente

Tabla N° 13: Tabla fuente producto para ventas.

Tipo de Fuente	Nombre de la Tabla	Descripción
Base de datos de la empresa (MySQL)	Productos	La tabla contiene toda la información concerniente a los Productos de la empresa.

Elaboración: Propia.

c) Estandarización y limpieza de datos

Tabla N° 14: Tabla estandarización de datos de producto para ventas.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
idProducto	PK	Int	Numérico	No debe ser nulo	NO TIENE
Tipo		Varchar(45)	Texto		NO TIENE
Marca		Varchar(45)	Texto		NO TIENE
Modelo		Varchar(45)	Texto		NO TIENE

Elaboración: Propia.

d) Fuentes de datos

Tabla N° 15: Tabla fuente de datos de producto para ventas.

Tabla: Productos				
Nombre	Llave	Tipo	Formato	Consideración importante
idProductos	PK	Int	Numérico	NO TIENE
Tipo		Varchar(45)	Texto	NO TIENE
Marca		Varchar(45)	Texto	NO TIENE
Modelo		Varchar(45)	Texto	NO TIENE
Serie		Varchar(45)	Texto	NO TIENE
Descripción		Varchar(150)	Texto	NO TIENE
Precio		Float	Decimal	NO TIENE

Elaboración: Propia.

e) Tabla destino

Tabla N° 16: Tabla destino de producto para ventas.

Tabla: InterProducto		
Campo	Tipo	Mapeo
idProducto	Int	Productos.idProductos
Tipo	Varchar(45)	Productos.Tipo
Marca	Varchar(45)	Productos.Marca
Modelo	Varchar(45)	Productos.Modelo

Elaboración: Propia.

f) Proceso

i. Carga de registros de un archivo intermedio

Se extraen los datos de la tabla “Productos” de acuerdo al mapeo y se carga en un archivo de texto intermedio “InterProducto”.

ii. Carga de la dimensión

Se extraen los valores del archivo “InterProducto” y se carga en

la dimensión “dim_producto”.

iii. Borrar el archivo intermedio

Finalmente se borra el archivo intermedio “InterProducto”.

Carga de la dimensión fecha

a) Descripción

Se realizará la carga de la dimensión “dim_fecha” desde la tabla “Tiempo” que se encuentra en la base de datos que maneja la empresa.

b) Descripción de las tabla fuente

Tabla N° 17: Tabla fuente tiempo para ventas.

Tipo de Fuente	Nombre de la Tabla	Descripción
Base de datos de la empresa (MySQL)	Tiempo	La tabla es generada mediante un procedimiento de MySQL de las diferentes incidencias diarias de la empresa.

Elaboración: Propia.

c) Estandarización y limpieza de datos

Tabla N° 18: Tabla estandarización de datos de fecha para ventas.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
idFecha	PK	Int	Numérico	No debe ser nulo	NO TIENE
Dia		Int	Numérico		NO TIENE
Mes		Int	Numérico		NO TIENE
Trimestre		Int	Numérico		NO TIENE
Año		Int	Numérico		NO TIENE

Elaboración: Propia.

d) Fuentes de datos

Tabla N° 19: Tabla fuente de datos de fecha para ventas.

Tabla: Tiempo				
Nombre	Llave	Tipo	Formato	Consideración importante
idTiempo	PK	Int	Numérico	NO TIENE
Dia		Int	Numérico	NO TIENE
Mes		Int	Numérico	NO TIENE
Trimestre		Int	Numérico	NO TIENE
Año		Int	Numérico	NO TIENE

Elaboración: Propia.

e) Tabla destino

Tabla N° 20: Tabla destino de fecha para ventas.

Tabla: InterFecha		
Campo	Tipo	Mapeo
idFecha	Int	Tiempo.idFecha
Dia	Int	Tiempo.Dia
Mes	Int	Tiempo.Mes
Trimestre	Int	Tiempo.Trimestre
Año	Int	Tiempo.Anio

Elaboración: Propia.

f) Proceso

i. Carga de registros de un archivo intermedio

Se extraen los datos de la tabla “Tiempo” de acuerdo al mapeo y se carga en un archivo de texto intermedio “InterFecha”.

ii. Carga de la dimensión

Se extraen los valores del archivo “InterFecha” y se carga en la dimensión “dim_fecha”.

iii. Borrar el archivo intermedio

Finalmente se borra el archivo intermedio “InterFecha”.

Datamarts de Compras

Carga de la dimensión proveedor

a) Descripción

Se realizará la carga de la dimensión “dim_proveedor” desde la tabla “Proveedores” que se encuentra en la base de datos que maneja la empresa.

b) Descripción de las tabla fuente

Tabla N° 21: Tabla fuente proveedor para compras.

Tipo de Fuente	Nombre de la Tabla	Descripción
Base de datos de la empresa (MySQL)	Proveedores	La tabla contiene todos los datos concernientes a los proveedores de la empresa.

Elaboración: Propia.

c) Estandarización y limpieza de datos

Tabla N° 22: Tabla Estandarización de datos de proveedor para compras.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
idProveedor	PK	Int	Numérico	No debe ser nulo	NO TIENE
Proveedor		Varchar(45)	Texto		NO TIENE
RUC		Varchar(45)	Texto		NO TIENE

Elaboración: Propia.

d) Fuentes de Datos

Tabla N° 23: Tabla fuente de datos de proveedor para compras.

Tabla: Proveedores				
Nombre	Llave	Tipo	Formato	Consideración importante
idProveedores	PK	Int	Numérico	NO TIENE
RazonSocial		Varchar(45)	Texto	NO TIENE
RUC		Int	Numérico	NO TIENE
Direccion		Varchar(45)	Texto	NO TIENE
Teléfono		Varchar(45)	Texto	NO TIENE
Cuenta		Int	Numérico	NO TIENE

Elaboración: Propia.

e) Tabla destino

Tabla N° 24: Tabla destino de proveedor para compras.

Tabla: InterProveedor		
Campo	Tipo	Mapeo
idProveedor	Int	Proveedores.idProveedor
Proveedor	Varchar(45)	Proveedores.Proveedor
RUC	Varchar(45)	Proveedores.RUC

Elaboración: Propia.

f) Proceso

i. Carga de registros de un archivo intermedio

Se extraen los datos de la tabla “Proveedores” de acuerdo al mapeo y se carga en un archivo de texto intermedio “InterProveedor”.

ii. Carga de la dimensión

Se extraen los valores del archivo “InterProveedor” y se carga en la dimensión “dim_proveedor”.

iii. Borrar el archivo intermedio

Finalmente se borra el archivo intermedio “InterProveedor”.

Carga de la dimensión producto

a) Descripción

Se realizará la carga de la dimensión “dim_producto” desde la tabla “Productos” que se encuentra en la base de datos que maneja la empresa.

b) Descripción de las tabla fuente

Tabla N° 25: Tabla fuente producto para compras.

Tipo de Fuente	Nombre de la Tabla	Descripción
Base de datos de la empresa (MySQL)	Productos	La tabla contiene toda la información concerniente a los Productos de la empresa.

Elaboración: Propia

c) Estandarización y limpieza de datos

Tabla N° 26: Tabla estandarización de datos de producto para compras.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
idProducto	PK	Int	Numérico	No debe ser nulo	NO TIENE
Tipo		Varchar(45)	Texto		NO TIENE
Marca		Varchar(45)	Texto		NO TIENE
Modelo		Varchar(45)	Texto		NO TIENE

Elaboración: Propia.

d) Fuentes de datos

Tabla N° 27: Tabla fuente de datos de producto para compras.

Tabla: Productos				
Nombre	Llave	Tipo	Formato	Consideración importante
idProductos	PK	Int	Numérico	NO TIENE
Tipo		Varchar(45)	Texto	NO TIENE
Marca		Varchar(45)	Texto	NO TIENE
Modelo		Varchar(45)	Texto	NO TIENE
Serie		Int	Numérico	NO TIENE
Precio		Float	Decimal	NO TIENE
Descripción		Varchar(150)	Texto	NO TIENE

Elaboración: Propia.

e) Tabla destino

Tabla N° 28: Tabla destino de producto para compras.

Tabla: InterProducto		
Campo	Tipo	Mapeo
idProducto	Int	Productos.idProductos
Tipo	Varchar(45)	Productos.Tipo
Marca	Varchar(45)	Productos.Marca
Modelo	Varchar(45)	Productos.Modelo

Elaboración: Propia.

f) Proceso

i. Carga de registros de un archivo intermedio

Se extraen los datos de la tabla “Productos” de acuerdo al mapeo y se carga en un archivo de texto intermedio “InterProducto”.

ii. Carga de la dimensión

Se extraen los valores del archivo “InterProducto” y se carga en la dimensión “dim_producto”.

iii. Borrar el archivo intermedio

Finalmente se borra el archivo intermedio “InterProducto”.

Carga de la dimensión fecha

a) Descripción

Se realizará la carga de la dimensión “dim_fecha” desde la tabla “Tiempo” que se encuentra en la base de datos que maneja la empresa.

b) Descripción de las tabla fuente

Tabla N° 29: Tabla fuente de tiempo para compras.

Tipo de Fuente	Nombre de la Tabla	Descripción
Base de datos de la empresa (MySQL)	Tiempo	La tabla es generada mediante un procedimiento de MySQL de las diferentes incidencias diarias de la empresa.

Elaboración: Propia.

c) Estandarización y limpieza de datos

Tabla N° 30: Tabla estandarización de datos de fecha para compras.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
idFecha	PK	Int	Numérico	No debe ser nulo	NO TIENE
Día		Int	Numérico		NO TIENE
Mes		Int	Numérico		NO TIENE
Trimestre		Int	Numérico		NO TIENE
Año		Int	Numérico		NO TIENE

Elaboración: Propia.

d) Fuentes de datos

Tabla N° 31: Tabla fuente de datos de fecha para compras.

Tabla: Tiempo				
Nombre	Llave	Tipo	Formato	Consideración importante
idTiempo	PK	Int	Numérico	NO TIENE
Dia		Int	Numérico	NO TIENE
Mes		Int	Numérico	NO TIENE
Trimestre		Int	Numérico	NO TIENE
Año		Int	Numérico	NO TIENE

Elaboración: Propia.

e) Tabla destino

Tabla N° 32: Tabla destino de fecha para compras.

Tabla: InterFecha		
Campo	Tipo	Mapeo
idFecha	Int	Tiempo.idFecha
Dia	Int	Tiempo.Dia
Mes	Int	Tiempo.Mes
Trimestre	Int	Tiempo.Trimestre
Año	Int	Tiempo.Año

Elaboración: Propia

f) Proceso

i. Carga de registros de un archivo intermedio

Se extraen los datos de la tabla “Tiempo” de acuerdo al mapeo y se carga en un archivo de texto intermedio “InterFecha”.

ii. Carga de la dimensión

Se extraen los valores del archivo “InterFecha” y se carga en la dimensión “dim_fecha”.

iii. Borrar el archivo intermedio

Finalmente se borra el archivo intermedio “InterFecha”.

4.1.4.3 Diseño del repositorio de metadatos (Data Warehouse)

Para hacer el diseño del repositorio de metadatos que vena hacer el lugar donde serán almacenados los datos una vez que sigan el proceso de Extracción, transformación y sean cargados en el repositorio. Para lo cual de acuerdo a cada datamarts se creara las tablas con sus respectivos campos.

Datamarts de Ventas

- La tabla que se conformara o llamada tabla de Hechos será llamado “Ventas”
- Los campos que conformaran esta tabla serán los que provendrán de las tablas de dimensiones los cuales en este caso son: Cliente, Producto y Fecha, como se muestra a continuación:

Tabla N° 33: Datamart ventas.

VENTAS
idCliente
idProducto
idFecha
Cantidad
MontoTotal

Elaboración: Propia.

- La tabla correspondiente de E-R será de esta forma:

Tabla N° 34: Tabla E – R de ventas.

Ventas				
idCliente	idProducto	idFecha	Cantidad	MontoTotal

Elaboración: Propia.

Datamarts de Compras

- La tabla que se conformara o llamada tabla de Hechos será llamado “Compras”.
- Los campos que conformaran esta tabla serán los que provendrán de las tablas de dimensiones los cuales en este caso son: Proveedor, Producto y Fecha, como se muestra a continuación:

Tabla N° 35. Datamart compras.

COMPRAS
idProveedor
idProducto
idFecha
Cantidad
MontoTotal

Elaboración: Propia.

- La tabla correspondiente de E-R será de esta forma:

Tabla N° 36: Tabla E – R de compras.

Compras				
idProveedor	idProducto	idFecha	Cantidad	MontoTotal

Elaboración: Propia.

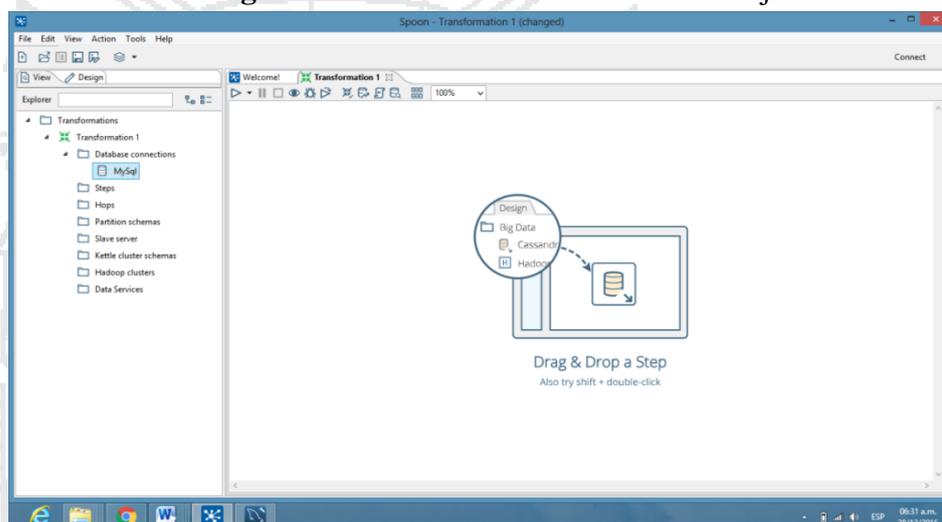
4.1.5. CONSTRUCCIÓN

4.1.5.1 Desarrollo del ETL

Antes de comenzar con el desarrollo primeramente se tendrá que instalar y configurar MySql, que es donde estará nuestra base de datos (ver Anexo N° 2). Y seguidamente para el desarrollo del ETL, se utilizará la herramienta Data Integration de Pentaho, teniendo los siguientes puntos:

1. Realizamos la instalación de Pentaho Data Integration. (ver **Anexo N° 3**) los pasos a seguir.
2. Inicializamos el programa creando seguidamente una conexión a la Base de Datos donde se almacenara los datos ya listos para que otras aplicaciones puedan hacer su uso. (ver **Anexo N° 4**) el proceso de la conexión. Seguidamente creamos un nuevo trabajo.

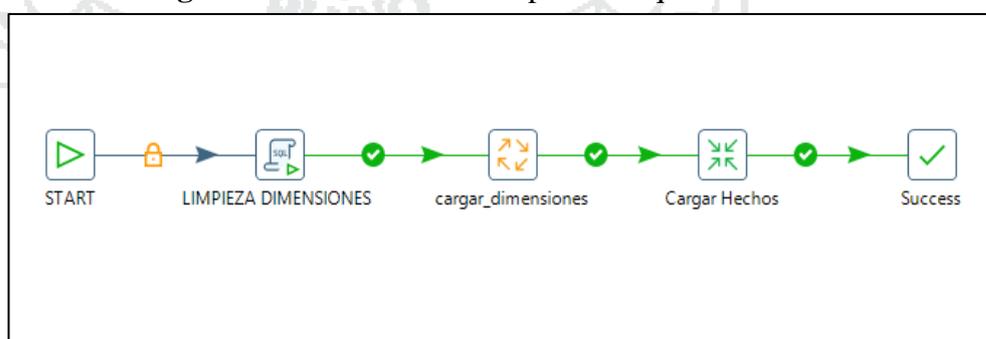
Figura N° 37: Creación de un nuevo trabajo.



Elaboración: Propia.

3. Luego diseñamos el proceso del ETL seleccionando en la pestaña respectiva “Design” y agregando los siguientes procesos.

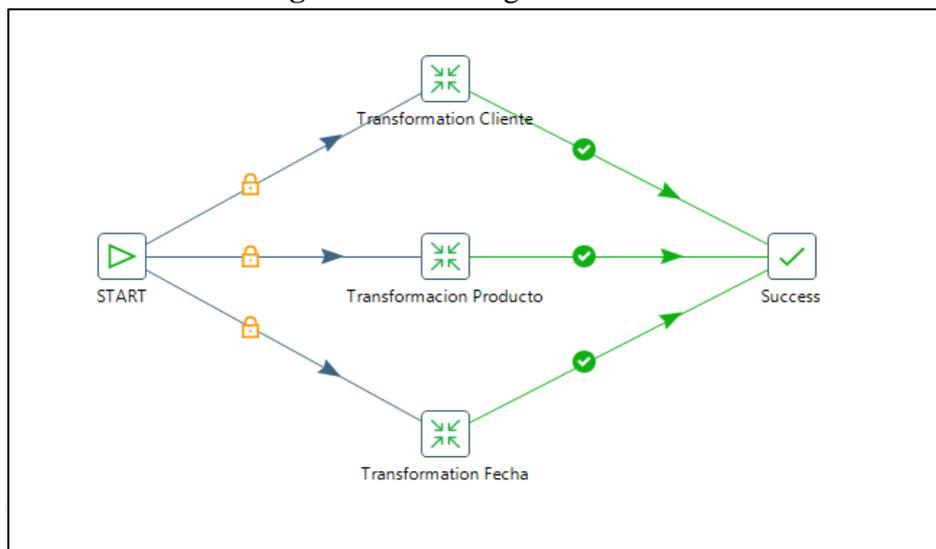
Figura N° 38: Diseño de los procesos que tendrá el ETL.



Elaboración: Propia.

4. En el sub trabajo “cargar_dimensiones” agregamos los siguientes procesos teniendo de esta manera:

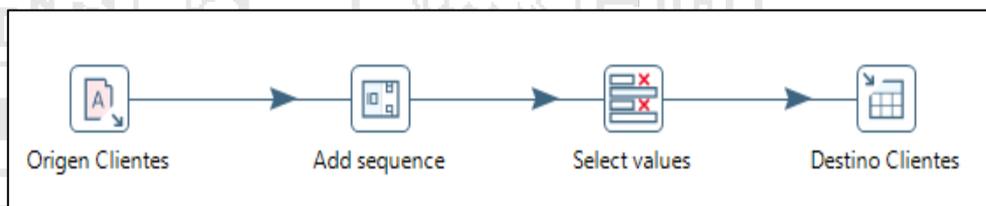
Figura N° 39: Carga de dimensiones.



Elaboración: Propia.

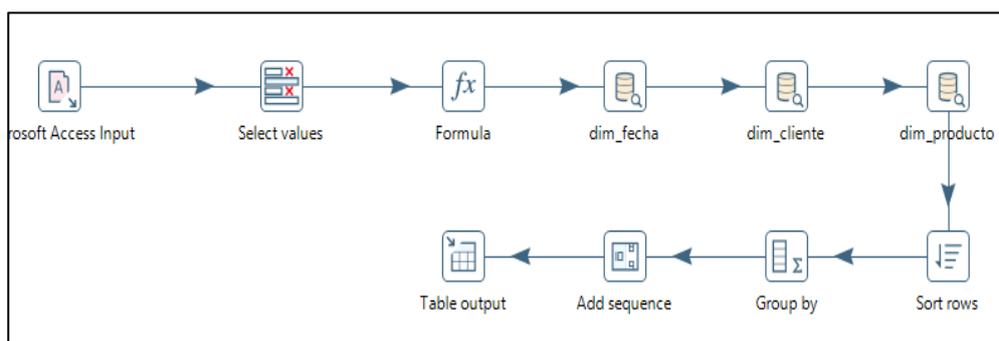
5. Configuramos el origen de datos haciendo doble clic, configuramos el nombre, el tipo y cargamos el archivo respectivo. Esto para cada dimensión.

Figura N° 40: Carga del origen de datos a la tabla destino.



Elaboración: Propia.

6. Finalmente los datos estarán listos en nuestra data warehouse “Mallku” en las tablas hechos de ventas y compras. Para la dimensión Fecha se implementó el código en java como se muestra en el **Anexo N° 9**.

Figura N° 41: Carga en las tablas hechos de ventas y compras.

Elaboración: Propia.

4.1.5.2 Desarrollo de aplicaciones

En esta etapa se realizan las instalaciones de las herramientas de Pentaho teniendo ya en el paso anterior la instalación y configuración de Pentaho Data Integration con MySQL. Para lo cual también se debe tener instalado Java JDK con la configuración de las variables respectivas (ver **Anexo N° 5**).

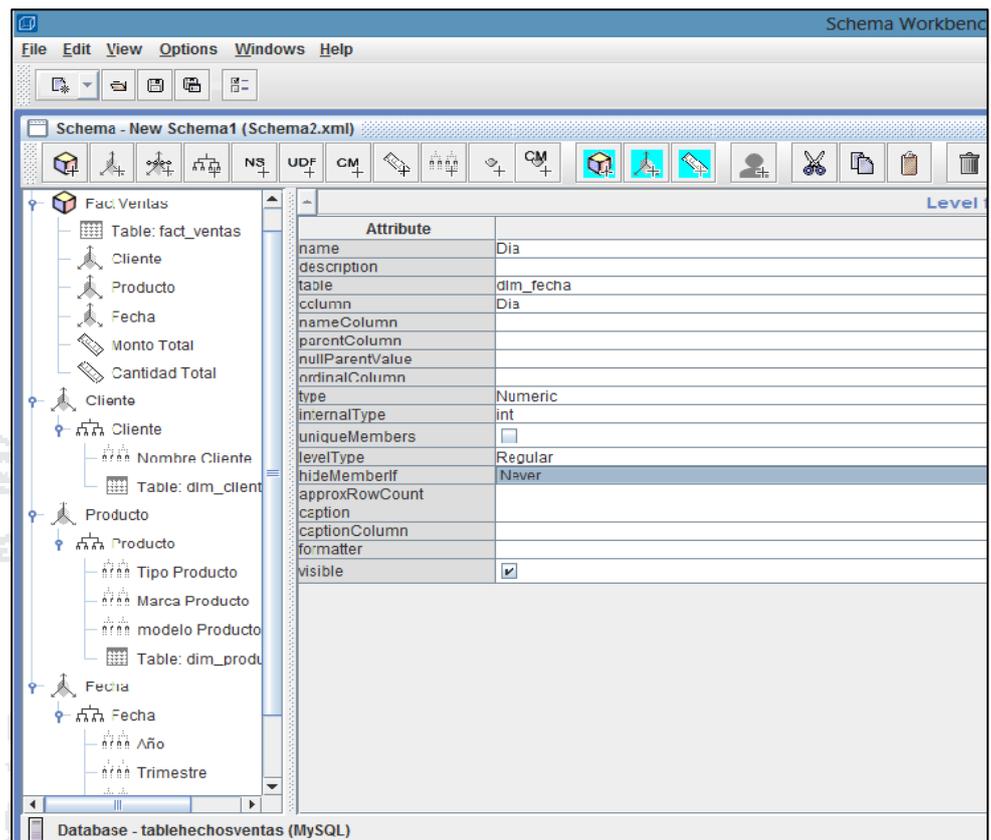
Las instalaciones se realizaran en el siguiente orden:

- Pentaho Analysis Server Mondrian (Schema Workbench) (ver

Anexo N° 6)

Luego se procede a configurar nuestro nuevo esquema como se muestra en la figura:

Figura N° 42: Configuración de esquema en OLAP.



Elaboración: Propia.

- Pentaho Business Intelligence Server (ver Anexo N° 7)

Donde serán cargados los esquemas creados para su respectivo análisis y creación de reportes.

4.1.5.3 Minería de datos

Esta etapa no será desarrollada en este proyecto de tesis.

4.1.5.4 Desarrollo del repositorio de metadatos (Data Warehouse)

El repositorio de metadatos como se explicó anteriormente en la implementación del proceso de ETL, será creado en el entorno de MySQL Server, donde también se especificó el nombre que es de “Mallku” y donde también se crearán de manera directa cada una de las tablas de dimensiones y hechos con sus respectivos campos obteniendo así el

repositorio de datos en forma íntegra.

4.1.6. DESPLIEGUE

4.1.6.1 Implementación

Una vez que se ha probado cada componente del proyecto, se comienza a instalar el motor de la base de datos y la plataforma Web. Se programa entrenamiento para los usuarios. En esta etapa comienza las funciones de soporte, que incluye mantener la base de datos, programar y correr los jobs ETL, monitorear el comportamiento del sistema y afinar la base de datos.

4.1.6.2 Evaluación de lanzamiento

La evaluación de lanzamiento constituye una actividad fundamental ya que asegura la calidad del proyecto. Por lo cual este paso se verá más adelante en la prueba de Hipótesis detalladamente.

4.2. PRUEBA DE HIPÓTESIS

4.2.1. Evaluación del Sistema por la Empresa

La evaluación del Sistema de Soporte de Decisiones se realizó en base a una encuesta (ver **Anexo N° 8**) y solo para las partes implicadas, que son los usuarios que se mencionó en la muestra que son 6 personas. Obteniendo el siguiente resultado:

A) Del diseño

- **¿Cómo califica el diseño de interfaz del sistema?**

Tabla N° 37: Resultado de opinión con respecto al diseño de interfaz.

Items	Usuarios	Porcentaje
Excelente	3	50%
Bueno	3	50%
Regular	0	0%
Malo	0	0%
Total	6	100%

Elaboración: Propia.

Interpretación: De los resultados a la pregunta efectuada se, resuelve que el 50% de los encuestados opina que la interfaz es excelente, el 50 % que es bueno. Por lo tanto se concluye que la interfaz es agradable al usuario.

- **¿Qué tan fácil le parece el uso o manejo del sistema?**

Tabla N° 38: Resultado de opinión con respecto al manejo del sistema.

Items	Usuarios	Porcentaje
Fácil	5	83.33%
Regular	1	16.67%
Difícil	0	0%
Total	6	100%

Elaboración: Propia.

Interpretación: De los resultados a la pregunta efectuada, se resuelve que el 83.33% de los encuestados opina que el manejo del sistema es fácil y un 16.67% es de manejo regular. Por lo tanto se concluye que el manejo del sistema es fácil para la mayoría de usuarios.

B) De la implementación y evaluación

- **¿Cómo considera Ud., el mantenimiento y actualización de los procesos de ETL (Extracción, Carga y Transformación)?**

Tabla N° 39: Resultado de opinión con respecto al mantenimiento y actualización de los procesos ETL.

Items	Usuarios	Porcentaje
Fácil	3	50%
Regular	3	50%
Difícil	0	0%
Total	6	100%

Elaboración: Propia.

Interpretación: De los resultados a la pregunta efectuada, se resuelve que para un 50% de los encuestados es fácil realizar el mantenimiento (actualización) de los procesos ETL y para un 50% es regular. Por lo tanto se concluye que para los usuarios es ligeramente fácil realizar el mantenimiento y actualización de los procesos ETL.

- **¿Cómo considera Ud., la creación de consultas y cuadros de dimensiones OLAP?**

Tabla N° 40: Resultado de opinión con respecto a la creación de consultas y cuadros de dimensiones OLAP.

Items	Usuarios	Porcentaje
Fácil	4	66.67%
Regular	2	33.33%
Difícil	0	0%
Total	6	100%

Elaboración: Propia.

Interpretación: De los resultados a la pregunta efectuada, se resuelve que para un 66.67% de los encuestados es fácil realizar la creación de consultas y cuadros de dimensiones OLAP y para un 33.33% es regular. Por lo tanto se concluye que para los usuarios es fácil realizar la

creación de consultas y cuadros de dimensiones OLAP.

¿Cómo considera la disponibilidad de la información generada por el sistema?

Tabla N° 41: Resultado de opinión con respecto a la disponibilidad de la información.

Items	Usuarios	Porcentaje
Excelente	3	50%
Bueno	2	33.34%
Regular	1	16.66%
Malo	0	0
Total	6	100%

Elaboración: Propia.

Interpretación: De los resultados a la pregunta efectuada se resuelve que un 50% considera que la disponibilidad de la información es excelente, un 33.34% que es bueno y un 16.66% que es regular. Por lo tanto, se concluye que la disponibilidad de la información generada por el sistema es correcta.

C) De la validación del sistema

¿El sistema propuesto, optimiza la gestión de la información para tomar mejores decisiones?

Tabla N° 42: Resultado de opinión con respecto a la aceptación del sistema.

Items	Usuarios	Porcentaje
Si	5	83.33%
No	1	16.67%
Total	6	100%

Elaboración: Propia.

Interpretación: De los resultados a la pregunta efectuada se resuelve que un 83.33% tiene aceptación del sistema y un 16.67% no. Por lo tanto, se concluye que el Sistema de Soporte de Decisiones si optimiza la Gestión de la Información para los usuarios.

Donde se tiene principalmente los siguientes aspectos del antes y después de la creación del sistema:

Tabla N° 43: Cuadro comparativo del antes y después del sistema.

Antes	Después
El registro de clientes y las operaciones de ventas se procesa en un sistema de Access donde difícilmente se puede hacer un análisis para la gestión de la información en tiempo real.	Se tiene información que fue extraída del sistema de Access donde la información generada permite optimizar la información como es el área de ventas y administración que permite saber sobre los clientes, según los productos que compra y por periodos de tiempo.
La información de proveedores es almacenada en registros de Excel, por lo cual es complejo poder realizar un análisis en tiempo real que permita gestionar la información como es el caso de saber a qué proveedor se le compra más y saber los costos de estos.	La información es extraída del registro de Excel, para que pueda ser procesada y genere información que permita optimizar la gestión de la información en el área de compras, mantenimiento y almacén, donde se tiene información del proveedor, los productos que se compra y por cada periodo de tiempo.

Elaboración: Propia.

4.2.2. Planteamiento de la hipótesis

Para el proyecto de investigación “Sistema de Soporte de Decisiones con tecnología Data Warehouse para la Gestión de la Información de la Empresa Mallku Import SAC - Juliaca 2016”, se supuso que el 80% de usuarios considera que el Sistema de Soporte de Decisiones con tecnología Data Warehouse, optimiza la Gestión de la información.

Establecemos la hipótesis:

H₀: $P \geq 0.80$; El Sistema de Soporte de Decisiones con Tecnología Data Warehouse Optimiza la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.

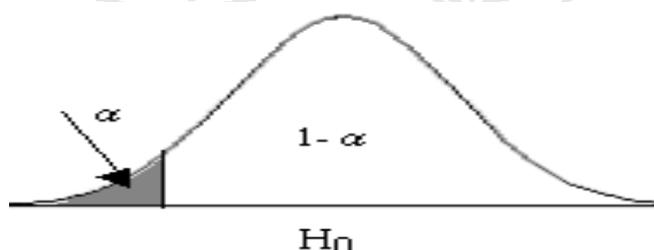
H₁: $P < 0.80$; El Sistema de Soporte de Decisiones con Tecnología Data Warehouse no Optimiza la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.

4.2.3. Nivel de significancia

El nivel de significancia será: $\alpha = 0.05$

4.2.4. Zona de rechazo

Figura N° 43: Zona de Rechazo.



Valor Crítico: **T = -1.64**

Elaboración: Propia.

4.2.5. Estadístico de prueba

Para hallar el estadístico de prueba utilizaremos la fórmula que corresponde a las muestras pequeñas y es:

$$T = \frac{\bar{P} - P_{H_0}}{\frac{S}{\sqrt{n}}} \quad \text{Ec.... (1)}$$

$$\bar{P} = 0.833$$

$$P_{H_0} = 0.80$$

$$n = 6$$

$$S = (0.80)(0.20) = 0.16$$

Reemplazamos los valores en la fórmula (1)

$$T = \frac{0.833 - 0.80}{\frac{0.16}{\sqrt{6}}}$$

$$T = 0.051$$

4.2.6. Decisión

$T = 0.051$, corresponde a la región de aceptación, por lo tanto, se acepta la hipótesis H_0 .

4.2.7. Resultado de la prueba de hipótesis

Se tiene como resultado la aceptación de la hipótesis planteada que efectivamente el Sistema de Soporte de Decisiones con tecnología Data Warehouse optimiza la Gestión de la Información de la Empresa Mallku Import SAC – Juliaca 2016.

CAPITULO V



CONCLUSIONES

PRIMERO: Se llevó a cabo con plena satisfacción el desarrollo del Sistema de Soporte de Decisiones con tecnología Data Warehouse para optimizar la Gestión de la Información de la Empresa Mallku SAC Juliaca – 2016, donde se tiene una optimización de más de un 83% en la gestión de la información, que permite mejorar la toma de decisiones, conocer a detalle la información sobre las ventas y compras por cliente, producto, proveedor y por periodo de tiempo.

SEGUNDO: En el caso del análisis de requerimientos del negocio, se logró determinar, de acuerdo a los objetivos de la empresa, se tiene como interrogantes a dar solución las siguientes: unidades vendidas por producto y en cierto periodo de tiempo, unidades compradas por proveedor y como también el monto total que generan ambas operaciones.

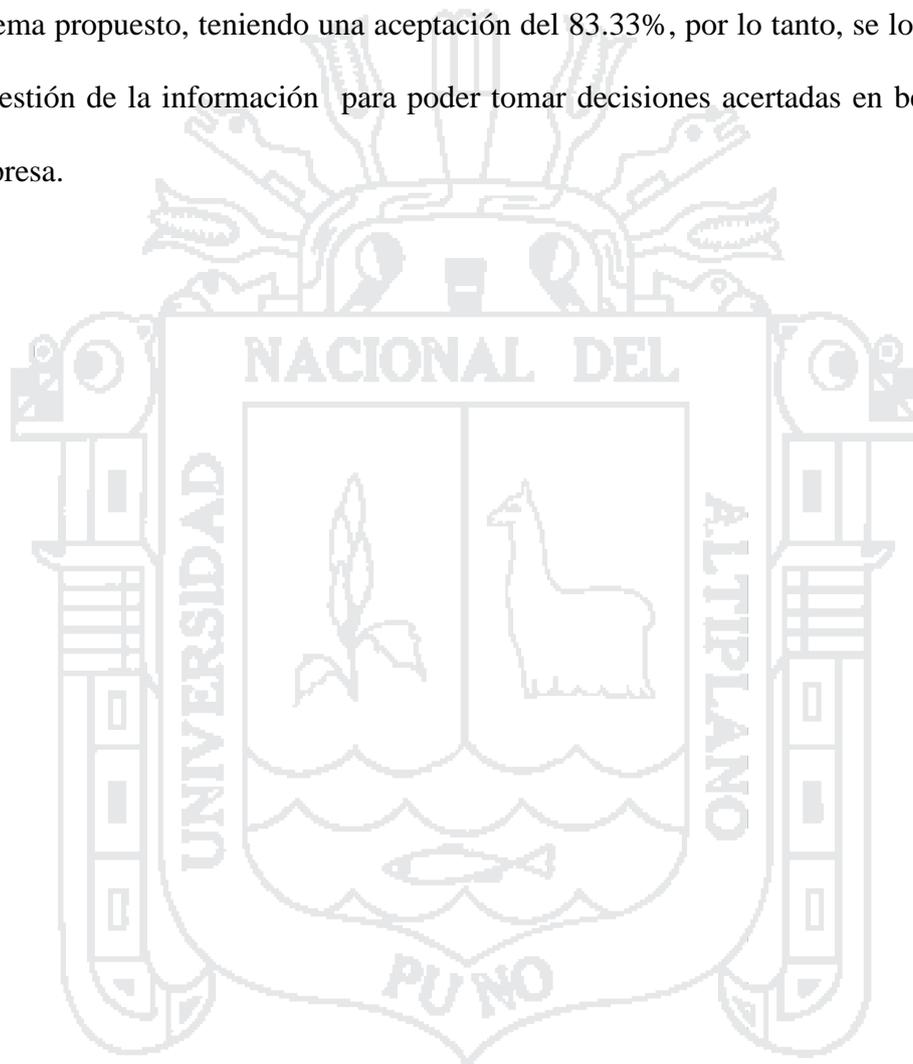
Es importante precisar los requerimientos claramente, para que el proyecto pueda emprender el camino correcto.

TERCERO: En el caso del diseño y construcción del modelo se utilizó la metodología de Larissa T. Moss basada en inteligencia de negocios y el enfoque de Ralph Kimball con el modelo estrella, como se muestra en el capítulo IV, debido a las características encontradas de acuerdo al análisis de los requerimientos, los cuales se ajustan a los este tipo de sistemas.

CUARTO: En la implementación y evaluación del sistema se tienen, las herramientas de software que se utilizaron como son: MySQL server para el modelamiento y el repositorio de la metadata, Spoom de Pentaho para el proceso de ETL (Extracción, Transformación y Carga), Workbench de Pentaho para la elaboración de los cubos multidimensionales OLAP (Procesamiento Analítico en Línea), Report Designer de

Pentaho para la elaboración de los reportes y la plataforma de Pentaho Server para que el usuario final pueda acceder y hacer uso del sistema.

QUINTO: Para la validación se realizaron las encuestas respectivas a los implicados en la investigación, donde se pudo observar y conocer la satisfacción de los usuarios con el sistema propuesto, teniendo una aceptación del 83.33%, por lo tanto, se logró optimizar la gestión de la información para poder tomar decisiones acertadas en beneficio de la empresa.



CAPITULO VI



RECOMENDACIONES

PRIMERO: Propiciar una mayor difusión para el conocimiento de las MyPEs, el uso de las nuevas tecnologías de información, como es el caso de los Sistemas de Soporte de Decisiones basada en inteligencia de negocios, con las herramientas adecuadas en su construcción es accesible poder desarrollar estos sistemas y nuestras empresas de la región puedan tener ese crecimiento y desarrollo que anhelan.

Como así también poder ampliar a futuro la optimización de las demás áreas según las prioridades del caso.

SEGUNDO: Es recomendable que para desarrollar este tipo de sistemas se tenga que tener una buena participación y comunicación con los trabajadores implicados de la empresa, para poder seguir el camino correcto, ya que estos sistemas trabajan tanto con los objetivos y estrategias que persiguen las empresas.

TERCERO: Es recomendable hacer uso de la metodología de Larissa T. Moss, el cual junto al enfoque de Ralph Kimball hacen que este tipo de sistemas puedan ser contruidos con las herramientas más adecuadas y correctas.

CUARTO: Es recomendable como se vio en esta investigación del uso de tecnologías open source, los cuales en su medida son beneficiosos para las MyPEs y puedan desarrollar e implementar este tipo de sistemas.

QUINTO: Es recomendable que las pruebas para la validación del sistema tengan la mayor disposición y participación de los usuarios finales del sistema.

CAPITULO VII



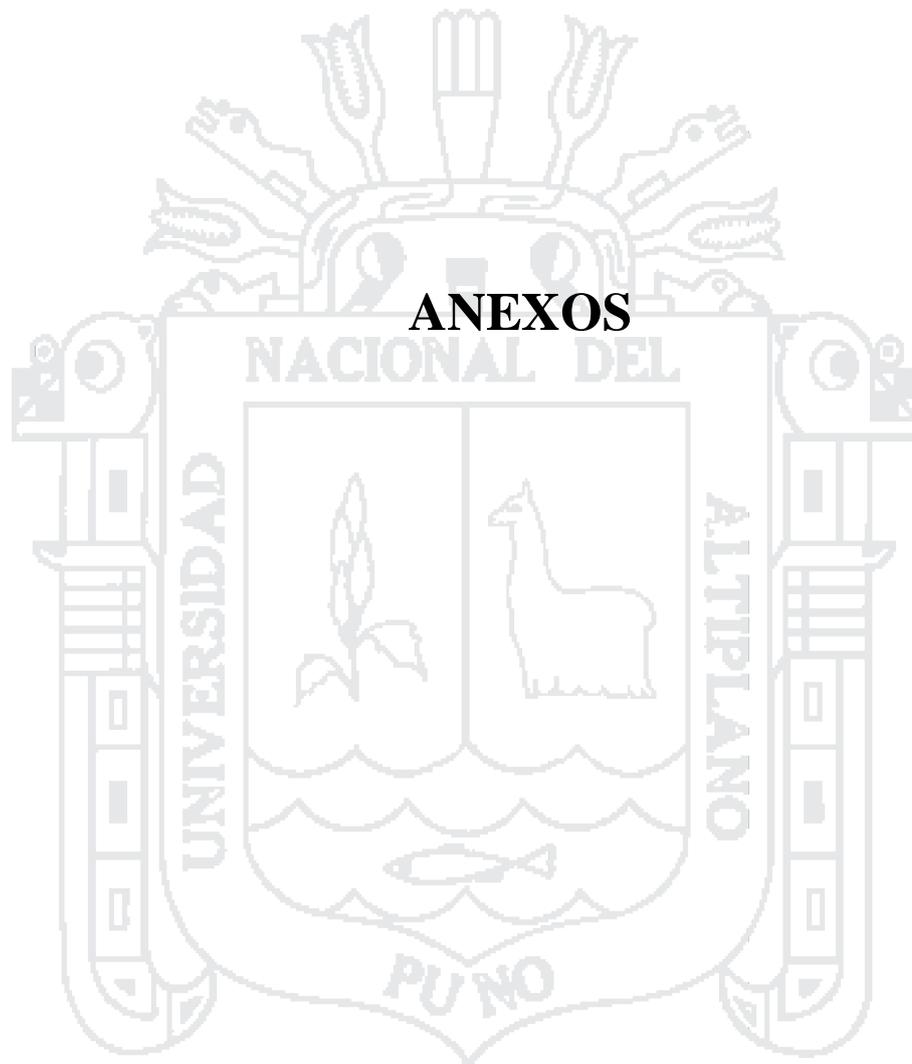
REFERENCIAS

- Alejandro.** (2009). Granularidad. Retrieved 08 de diciembre, 2016, from <http://alex-csvz-drugs-rock.blogspot.pe/2009/07/granularidad.html>
- Chávez, D. A.** (2015). Sistema de Soporte a la Toma de Decisiones basado en Inteligencia de Negocios para Mejorar los Procesos Comerciales del Importador Peruano. (Ingeniero de Sistemas y Computación), Universidad Católica Santo Toribio de Mogrovejo, Chiclayo – Perú.
- datawarehouse4u.info.** (2009). Almacen de datos. Retrieved 20 de octubre, 2016, from <http://datawarehouse4u.info/>
- Dertiano, V.** (2015). Arquitectura BI. Retrieved 21 de octubre, 2016, from <http://blog.mirai-advisory.com/>
- Díaz, J. C.** (2010). Introducción al Business Intelligence (1 ed.). Barcelona: Editorial UOC.
- E. Kendall, K., & E. Kendall, J.** (2005). Análisis y Diseño de Sistemas (6 ed.). México: Prentice Hall Inc.
- Gilfillan, I.** (2003). La Biblia de MySQL (1 ed.). Madrid: Anaya Multimedia.
- Gravitar.** (2016). Introducción a PENTAHO. Retrieved 02 de diciembre, 2016, from <http://gravitar.biz/bi/introduccion-pentaho-parte-1/>
- Haces, R.** (2016). Introduction to Mondrian Performance Tuning. Retrieved 01 de diciembre, 2016, from <https://support.pentaho.com/hc/en-us/articles/206021253-Best-Practice-Introduction-to-Mondrian-Performance-Tuning>
- Hernández, R., Fernández, C., & Baptista, M. d. P.** (2010). Metodología de la Investigación (5 ed.). México: McGraw-Hill.
- Howson, C.** (2009). Bussines Intelligence - Estrategias para una implementación exitosa (1 ed.). México: The McGraw-Hill Companies, Inc.

- Kimball, R., & Ross, M.** (2002). *The Data Warehouse Toolkit* (2 ed.). EEUU: John Wiley & Sons, Inc.
- Larissa T. Moss, & Shaku Atre.** (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision Support Applications* (3 ed.). Boston: Pearson Education.
- Moreno, R. H.** (2013). *Análisis, Diseño e Implementación de Datamarts para las áreas de Ventas y Recursos Humanos de una Empresa dedicada a la exportación e importación de Productos Alimenticios.* (Ingeniero Informático), Pontificia Universidad Católica del Perú, Perú.
- Pérez, M.** (2009). *Gestión de la información.* Retrieved 20 de octubre, 2016, from <http://glossarium.bitrum.unileon.es/Home/gestion-de-la-informacion>
- Rojas, C.** (2008). *Datawarehousing.* Retrieved 18 de octubre, 2016, from <http://html.rincondelvago.com/datawarehousing.html>
- Sánchez, L.** (2014). *Análisis de Información y Toma de decisiones para Administración de Negocios.* (Ingeniero en Computación), Universidad Nacional Autónoma de México, México.
- SHISUBHA.** (2015). *ETL SERVICES.* Retrieved 23 de octubre, 2016, from <http://www.srisubha.co.za/solutions/etl-services/>
- Sinnexus.** (2007, 2016). *Business Intelligence - Informática Estratégica.* Retrieved 20 de octubre 2016, from http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx
- Software, E.** (2016). *Cubos OLAP de información para la toma de decisiones.* Retrieved 23 de octubre, 2016, from <http://www.evaluandosoftware.com/cubos-olap-informacion-la-toma-decisiones/>

Summers, D. (2006). *Administración de la Calidad* (1 ed.). México: Prentice Hall Inc.





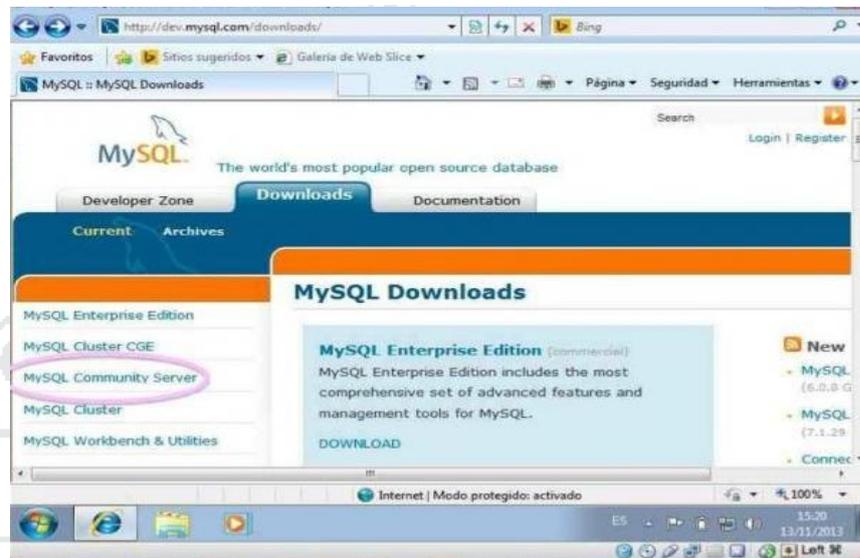
Anexo N° 1**CUESTIONARIO DE PREGUNTAS SOBRE LOS OBJETIVOS DE NEGOCIO
Y SIMILARES**

- Describa su trabajo y como se relaciona con el resto de la empresa
- Describa cuáles son sus responsabilidades
- ¿Cuáles son los objetivos de su organización?
- ¿Cuáles son los objetivos de esta área en específico?
- ¿Cómo mide actualmente su desempeño?
- ¿Utiliza indicadores? ¿Cuáles son sus indicadores de éxito?
- ¿Cómo sabe que lo está haciendo bien?
- A su modo de ver ¿Cuáles son los principales problemas que enfrenta hoy en día su empresa?
- ¿Cuáles son los principales problemas que enfrenta hoy su área y negocio?
- ¿Cómo identifica estos problemas?
- ¿Qué información utiliza usted actualmente para tomar sus decisiones?
- ¿Cómo usa esta información en sus decisiones?
- ¿Qué problemas tiene actualmente para conseguir la información que necesita?
- ¿Qué capacidades de análisis le gustaría tener?
- ¿Qué necesidades tiene de información histórica?
- ¿Qué reportes usa actualmente? ¿Cuál son los datos importantes en estos reportes?

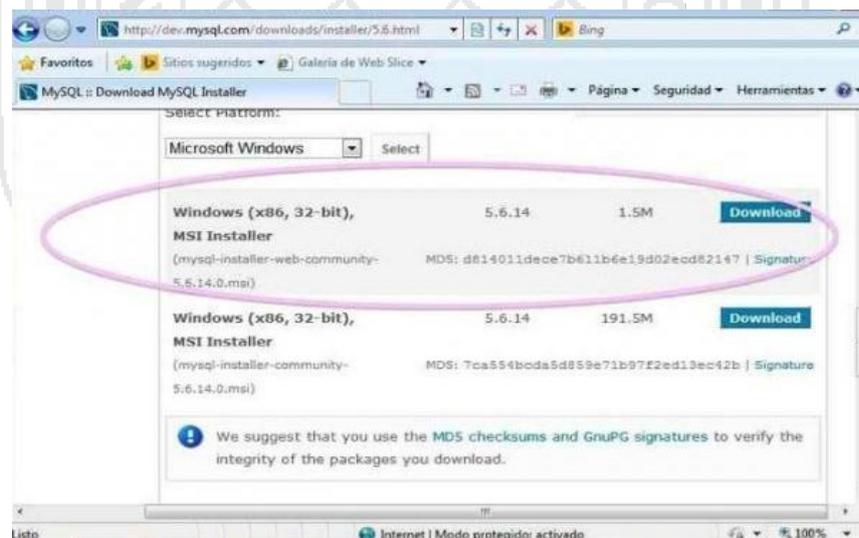
Anexo N° 2

INSTALACIÓN DE MYSQL SERVER

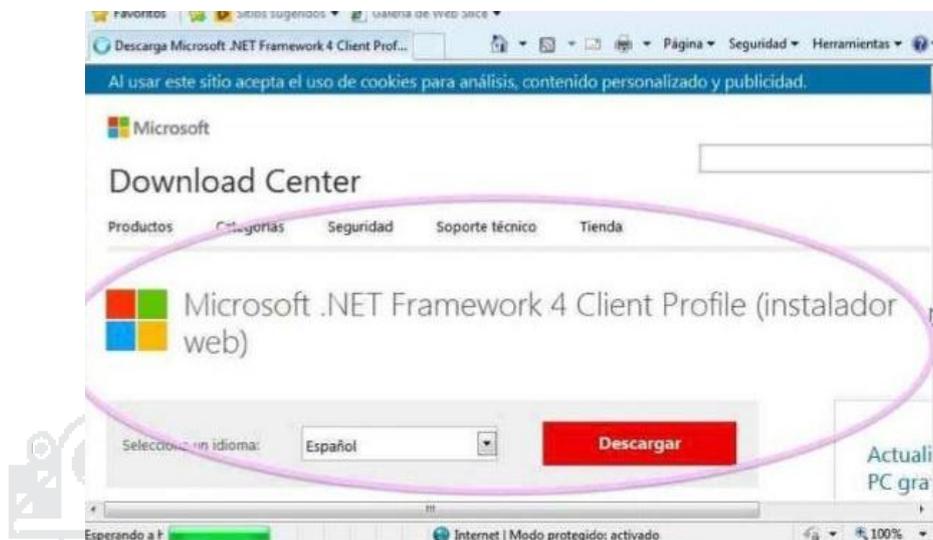
1. En la página web de **MySQL**, ve al apartado de "**MySQL Community Server**", que te dará acceso a las descargas del programa.



2. Después, tienes que elegir la versión que se adapta a las características de tu equipo y de tu sistema operativo, en este caso, lo necesitamos para **Windows**.



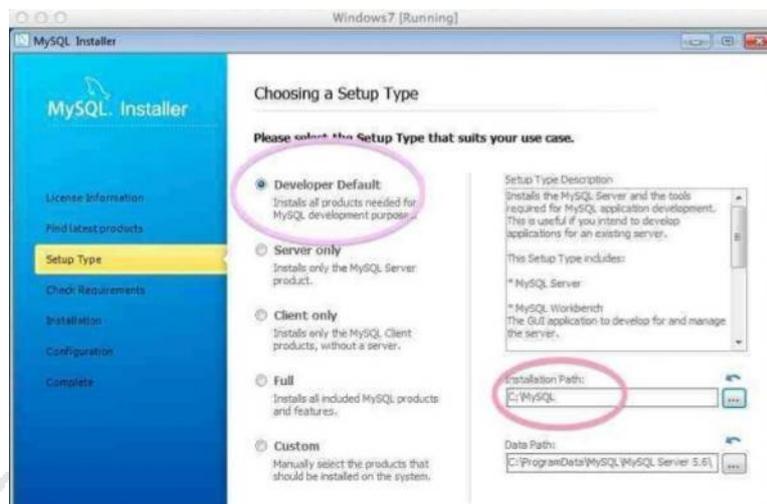
3. Cuando ya esté la descarga completada, en el caso de que no lo tengas aún, el instalador te avisará que necesitas descargar "Microsoft.NET Framework 4 Client Profile"



4. Una vez superado el paso anterior, has de elegir entre varias opciones. Dale a la de "Install MySQL Products".



5. Después, escoge la opción de "Developer Default" y cambia a "C:MySQL" la carpeta en la que quieres instalar el programa gestor de **bases de datos**.



6. A continuación, se te instalará **MySQL** junto a una serie de complementos que harán que puedas usar este programa con todas sus potencialidades.
7. En los siguientes pasos, tendrás unas opciones de configuración, puedes dejarlas tal como están. Eso sí, en las correspondientes a los usuarios, debes escribir una contraseña para el administrador y, si lo necesitas, añadir otros usuarios.



Anexo N° 3

INSTALACION Y CONFIGURACION DE PENTAHO DATA INTEGRATION**Pre-requisitos:**

Java Runtime Environment **JRE** 1.4 en adelante, configurada

correctamente(Path). Puede ser descargada

de <http://www.java.com/en/download/manual.jsp>

Crear una base de datos en blanco en cualquier motor de base de datos

Descomprimir el archivo .zip descargado en una carpeta creada previamente por ejemplo:

- /home/{usuario}/pentaho (para el caso de linux)
- [C:\pentaho](#) (En el caso de Windows)

Ejecutar Spoon para abrir la Herramienta

Para sistemas Unix (Linux, Solaris)

- Abrir una consola (terminal)
- ingresar: cd Pentaho (Ubicarse en la carpeta del programa)
- ingresar: chmod +x *.sh (asignar permisos de ejecución a todos los archivos con extensión sh)
- ingresar: sh spoon.sh

Para Windows

- Abrir una consola (cmd.exe)
- Ingresar: cd Pentaho (Ubicarse en la carpeta del programa)
- ingresar: Spoon.bat

Otra opción para el ingreso a la herramienta es acceder directamente al ejecutable Kettle.exe

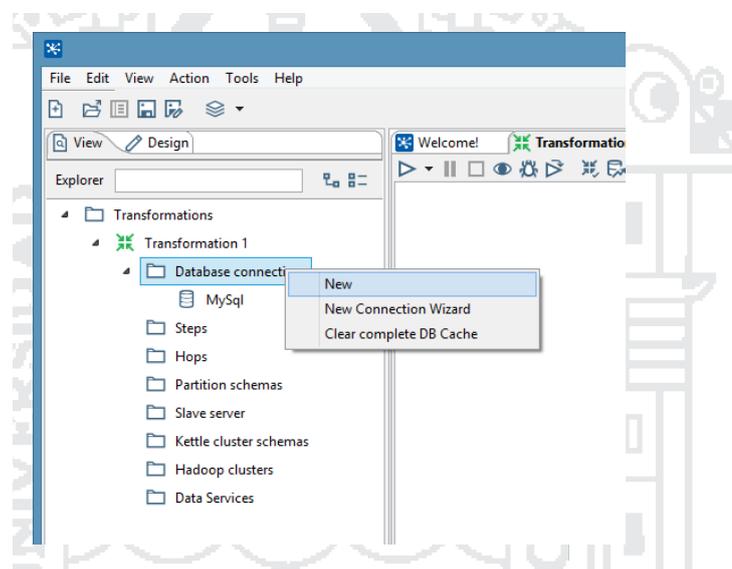
Anexo N° 4

CONEXIÓN DE DATA INTEGRATION A MYSQL

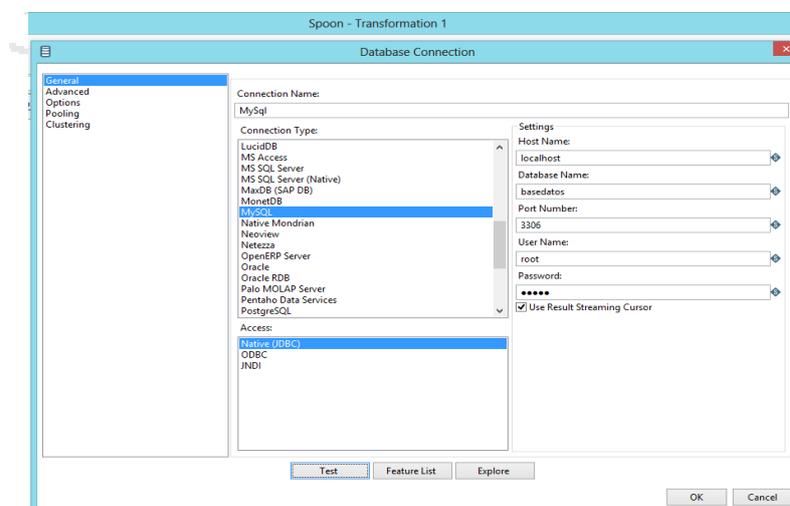
Esta sección describe cómo crear una nueva conexión de base de datos e incluye una descripción detallada de cada propiedad de la conexión.

Para crear una nueva conexión seleccionar en el panel izquierdo "Arbol Principal", hacer clic derecho en "Conexiones a bases de datos" y seleccionar "Nuevo" o "Asistente Nueva Conexión".

También se puede hacer doble clic en "Conexiones a bases de datos", o presionar F3.



Los siguientes apartados describen las opciones de configuración disponibles en cada pestaña del editor de conexión.



La pestaña General es en donde se configura la información básica sobre la conexión, tal como nombre de la conexión, tipo, método de acceso, nombre del servidor y acceso al mismo.

Lo primero que se debe hacer es dar un nombre a la conexión ("Connection Name").

Luego debe seleccionarse en el listado "Connection Type" el tipo de base de datos que utilizaremos. De acuerdo a lo que se elija, las opciones disponibles en "Setting" y "Access" irán variando.

Una vez que se han completado los datos de "Connection Name" , "Connection Type", "Connection Access" y "Connection Setting" es recomendable presionar el botón "Test" para verificar la correcta configuración de la conexión.

El botón "Explore" permite navegar interactivamente en la base de datos en cuestión, visualizar tablas, vistas y datos, generar DDL, etc.

El botón "Feature List " (Lista de funciones) muestra una tabla con variables y valores relacionados a la conexión actual.



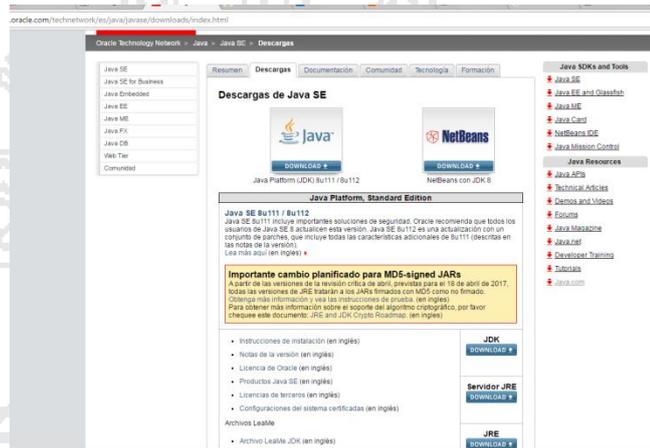
Anexo N° 5

INSTALACION Y CONFIGURACION DE JAVA JDK PARA PENTAHO

1. Primeramente descargamos el archivo de:

<http://www.oracle.com/technetwork/es/java/javase/downloads/index.html>

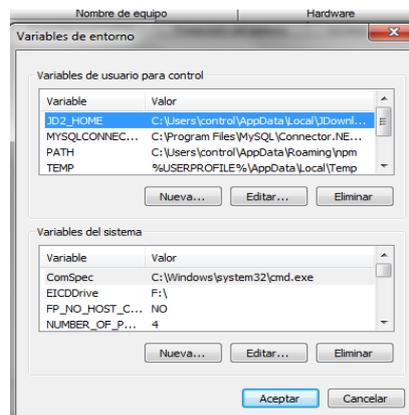
Especificando la versión y para qué sistema operativo.



2. Una vez descargado el archivo lo abrimos y continuamos con la instalación típica



3. Una vez instalado procedemos a realizar la configuración de las variables de entorno. Ingresando a Propiedades del equipo y configuración avanzada del equipo y seguidamente en variables de entorno.



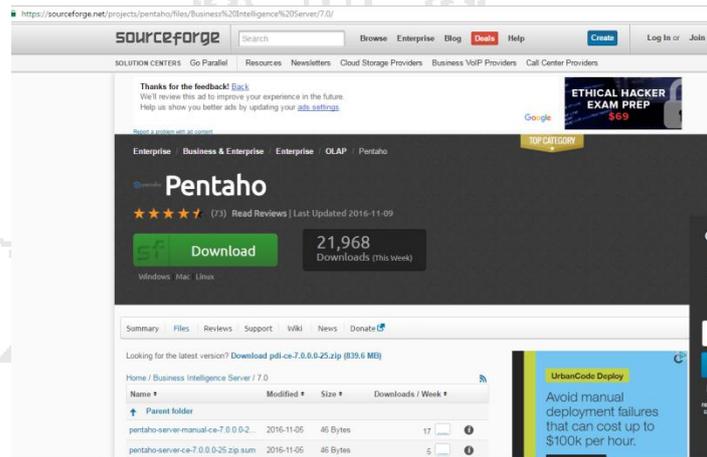
Anexo N° 6

INTALACION DE PENTAHO BUSINESS INTELLIGENCE SERVER

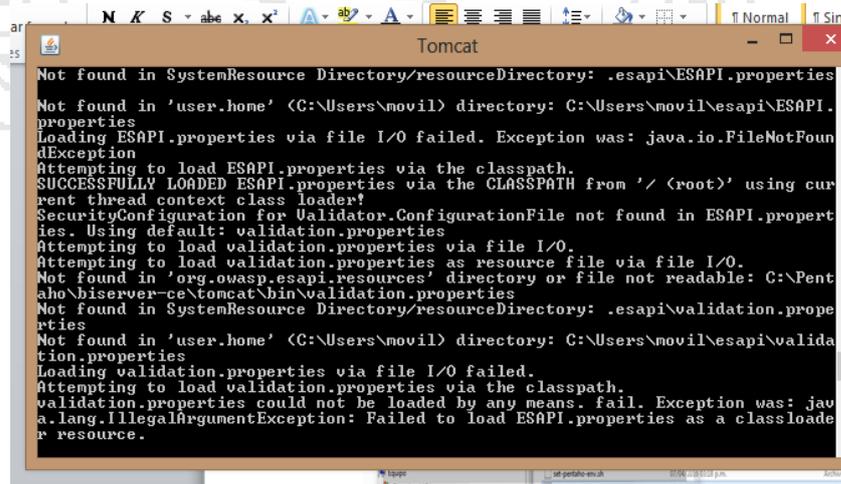
1. Descargar PENTAHO del sitio Web:

<https://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/>

2. Escogemos la opción donde dice pentaho-server-ce-7.0.0.0-25.zip



3. Luego descomprimos el archivo descargado en una carpeta creada con nombre “Pentaho” en la Unidad C.
4. Iniciamos el programa haciendo doble clic en “start-pentaho” que está en C:\Pentaho\biserver-ce.



```

Not found in SystemResource Directory/resourceDirectory: .esapi\ESAPI.properties
Not found in 'user.home' (C:\Users\movil) directory: C:\Users\movil\esapi\ESAPI.properties
Loading ESAPI.properties via file I/O failed. Exception was: java.io.FileNotFoundException
Attempting to load ESAPI.properties via the classpath.
SUCCESSFULLY LOADED ESAPI.properties via the CLASSPATH from '/' (root)' using current thread context class loader!
SecurityConfiguration for Validator.ConfigurationFile not found in ESAPI.properties. Using default: validation.properties
Attempting to load validation.properties via file I/O.
Not found in 'org.owasp.esapi.resources' directory or file not readable: C:\Pentaho\biserver-ce\tomcat\bin\validation.properties
Not found in SystemResource Directory/resourceDirectory: .esapi\validation.properties
Not found in 'user.home' (C:\Users\movil) directory: C:\Users\movil\esapi\validation.properties
Loading validation.properties via file I/O failed.
validation.properties could not be loaded by any means. fail. Exception was: java.lang.IllegalArgumentException: Failed to load ESAPI.properties as a classloader resource.
  
```

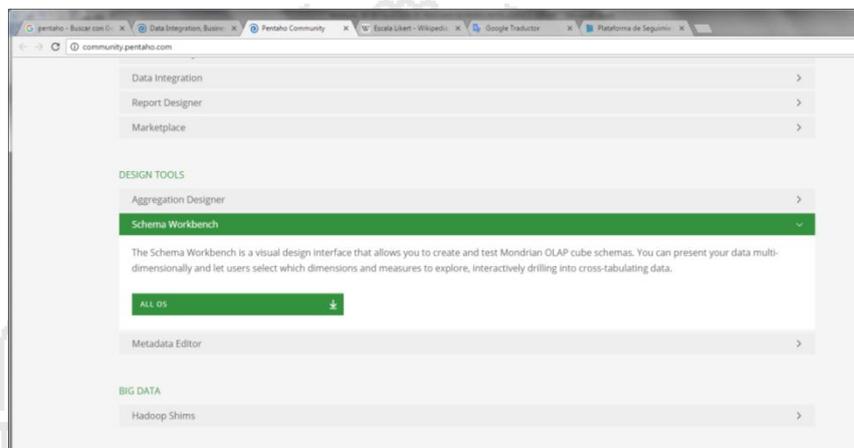
5. Finalmente ingresamos a la plataforma de Pentaho BI, mediante un navegador web en forma local. Poniendo “Localhost/8080”. Eh ingresamos en usuario “admin” y en contraseña “password”.



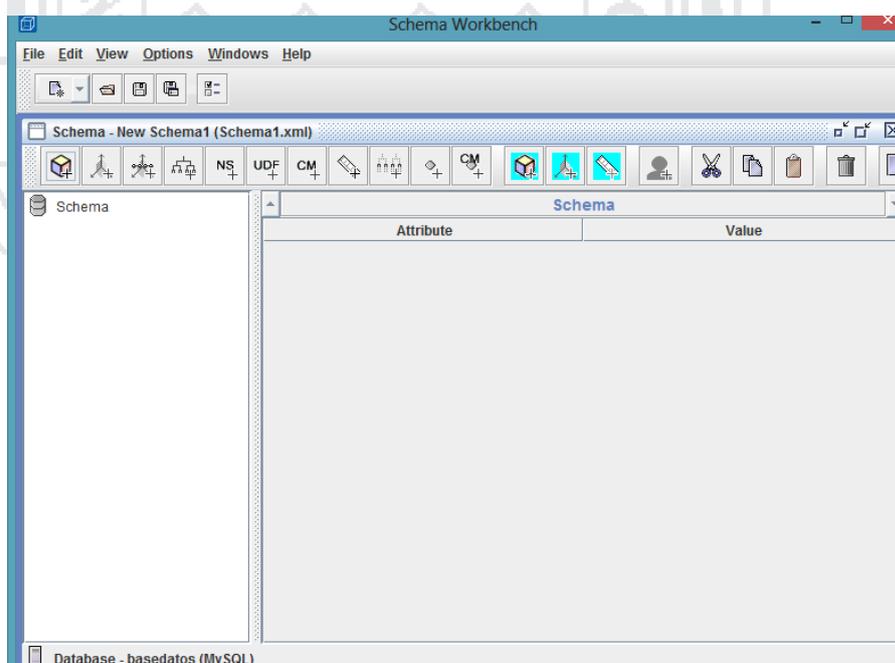
Anexo N° 7

INTALACION DE PENTAHO ANALYSIS SERVER MONDRIAN

6. Descargar PENTAHO del sitio Web: <http://community.pentaho.com/>
7. Escogemos la opción donde dice: Schema Workbench



8. Luego descomprimos el archivo descargado en una carpeta creada con nombre “Pentaho” en la Unidad C.
9. Iniciamos el programa haciendo doble clic en “workbench.bat” que está en C:\Pentaho\Schema Workbench.



Anexo N° 8**ENCUESTA**

Encuesta a usuarios del sistema, para validación del Sistema de Soporte de Decisiones con Tecnología Data Warehouse para la Gestión de la Información de la Empresa Mallku Import SAC.

1. ¿Cómo califica el diseño de interfaz del sistema?
 - a) Excelente
 - b) Bueno
 - c) Regular
 - d) Malo
2. ¿Qué tan fácil le parece el uso o manejo del sistema?
 - a) Fácil
 - b) Regular
 - c) Difícil
3. ¿Cómo considera Ud., el mantenimiento y actualización de los procesos de (ETL) Extracción, Carga y Transformación?
 - a) Fácil de aprender
 - b) Regularmente de aprender
 - c) Difícil de aprender
4. ¿Cómo considera Ud., la creación de consultas y cuadros de dimensiones OLAP?
 - a) Fácil
 - b) Regular
 - c) Difícil
5. ¿Cómo considera la disponibilidad de la información?
 - a) Excelente
 - b) Bueno
 - c) Regular
 - d) Malo
6. ¿El sistema propuesto, optimiza la Gestión de la Información para tomar mejores decisiones?
 - a) Si
 - b) No

Anexo N° 9

CODIGO DE CONFIGURACION DE FECHA

```

var Calendar = Packages.java.util.Calendar;

var Locale = Packages.java.util.Locale;

var SimpleDateFormat = Packages.java.text.SimpleDateFormat;

var year_for_week = (week_of_year == 1 && month == 12)?year+1:year;

var quarter = Math.floor((month-1)/3)+1;

var cal = Calendar.getInstance();

cal.set(year, month-1, day);

var d = cal.getTime();

var hoy = new java.util.Date();

var day_of_week_short_label_eng = new SimpleDateFormat("E", new
Locale("es","ES")).format(d);

var day_of_week_long_label_eng = new SimpleDateFormat("EEEE", new
Locale("es","ES")).format(d);

var month_short_label_eng = new SimpleDateFormat("MMM", new
Locale("es","ES")).format(d);

var month_long_label_eng = new SimpleDateFormat("MMMM", new
Locale("es","ES")).format(d);

var days_in_month = cal.getActualMaximum(Calendar.DAY_OF_MONTH);

var days_in_year = cal.getActualMaximum(Calendar.DAY_OF_YEAR);

var days_in_quarter;

switch(quarter){

    case 1:

        days_in_quarter = (days_in_year==366)? 31+29+31:

```

```
31+28+31;//ene,feb,mar
```

```
    break;
```

```
    case 2:
```

```
        days_in_quarter = 30+31+30; //abr,may,jun
```

```
        break;
```

```
    case 3:
```

```
        days_in_quarter = 31+31+30; //jul,agos,set
```

```
        break;
```

```
    case 4:
```

```
        days_in_quarter = 31+30+31; //oct,nov,dic
```

```
        break
```

```
    }
```

```
    var is_leap_year = (days_in_year == 366)?1:0;
```

```
    var is_weekend = (day_of_week==Calendar.SUNDAY || day_of_week ==  
Calendar.SATURDAY)?1:0;
```

```
    cal.set(Calendar.YEAR, year_for_week);
```

```
    var days_in_year_for_week = cal.getActualMaximum(Calendar.DAY_OF_YEAR);
```

```
    var is_leap_year_for_week = (days_in_year_for_week == 366)?1:0;
```

```
    cal.set(Calendar.YEAR, year_for_week_iso8601);
```

```
    var days_in_year_for_week_iso8601 =
```

```
    cal.getActualMaximum(Calendar.DAY_OF_YEAR);
```

```
    var is_leap_year_for_week_iso8601 = (days_in_year_for_week_iso8601 == 366)?1:0;
```