



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Software

Implementación de una solución de inteligencia de negocios utilizando la metodología Hefesto para las oficinas de contabilidad en universidades públicas

TESIS

Para optar el Título Profesional de Ingeniería de Software

AUTOR

Giancarlo Daniel GONZALES FARRO

ASESOR

Norberto Ulises ROMAN CONCHA

Lima, Perú

2018



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Gonzales, G. (2018). *Implementación de una solución de inteligencia de negocios utilizando la metodología Hefesto para las oficinas de contabilidad en universidades públicas*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Software]. Repositorio institucional Cybertesis UNMSM.



13^{ps} R
82^{ps} A

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SOFTWARE

Acta de Sustentación de Tesis

Siendo las 12:05 del día 27 de agosto del año 2018, se reunieron los docentes designados como miembros de Jurado de la Tesis, presidido por el Dr. Carlos Edmundo Navarro Depaz, Dra. Luzmila Elisa Pró Concepción (Miembro), y el Lic. Norberto Ulises Roman Concha (Miembro Asesor) para la sustentación de la Tesis intitulada: "IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS UTILIZANDO LA METODOLOGÍA HEFESTO PARA LAS OFICINAS DE CONTABILIDAD EN UNIVERSIDADES PÚBLICAS. CASO DE ESTUDIO: OFICINA DE CONTABILIDAD - UNMSM" por el Bach. Giancarlo Daniel Gonzales Farro, para optar el Título Profesional de Ingeniero de Software.

Acto seguido de la exposición de la Tesis, el Presidente invitó al Bachiller a dar respuesta a las preguntas establecidas por los Miembros de Jurado.

El Bachiller en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros de Jurado, el Bachiller obtuvo la nota de 17 (En letras) Diecisiete.

A continuación el Presidente del Jurado, CARLOS NAVARRO declara al Bachiller **Ingeniero de Software**.

Siendo las..... horas, se levantó la sesión.

Dr. Carlos Edmundo Navarro Depaz
Presidente

Dra. Luzmila Elisa Pró Concepción
Miembro

Lic. Norberto Ulises Roman Concha
Miembro Asesor

FICHA CATALOGRAFICA

IMPLEMENTACIÓN DE UN DASHBOARD DE RENDIMIENTO DE PERSONAL USANDO LA METODOLOGÍA HEFESTO PARA LAS OFICINAS DE CONTABILIDAD DE UNIVERSIDADES PÚBLICAS: CASO DE ESTUDIO OFICINA DE CONTABILIDAD – UNMSM

AUTOR: GONZALES FARRO, GIANCARLO DANIEL

ASESOR: ROMAN CONCHA, NORBERTO ULISES

LIMA – PERÚ, 2018

Título profesional: Ingeniero de Software

Área / Programa / Línea de Investigación:

Ingenierías / Tecnología de Información y Comunicación / Ingeniería de Software

Pregrado: Universidad Nacional Mayor de San Marcos - Facultad de Ingeniería de Sistemas e Informática - Escuela Profesional de Ingeniería de Software.

Formato 28 x 20 cm

Paginas: XIV, 82

Este trabajo está dedicado a mis padres Victor y Monselvy y mis hermanas Patty, Fiorella, Paola, Mariella y Adriana. Su confianza, aliento y apoyo constante me permiten hoy, alcanzar una de mis metas.

Agradecimientos

Al profesor Ulises Román, por su orientación y dedicación para que este trabajo cumpla con los objetivos trazados.

A todas aquellas personas que indirectamente me ayudaron para culminar este trabajo de investigación y que muchas veces constituyen un invalorable apoyo.

Y por encima de todo doy gracias a Dios.

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SOFTWARE

Implementación de un Dashboard de rendimiento de personal utilizando la metodología Hefesto para las oficinas de contabilidad en universidades públicas, Caso de estudio: Oficina de Contabilidad – UNMSM

Autor: Gonzales Farro, Giancarlo Daniel

Asesor: Román Concha, Norberto Ulises

Título: Título Profesional de Ingeniero de Software

Fecha: Enero de 2018

RESUMEN

Las Oficinas de Contabilidad de las Universidades Nacionales en gran mayoría no miden la información ni el desempeño de su personal en búsqueda de eficiencia y eficacia. La Oficina de Contabilidad de la Universidad Nacional Mayor de San Marcos (UNMSM) carece de información sobre el rendimiento de los trabajadores para la toma de decisiones estratégicas, esto es muy importante para saber dónde y cómo solventar alguna falla o mejorar las decisiones tomadas. El objetivo del presente trabajo es desarrollar un Data Mart a través de la metodología Hefesto, utilizando las herramientas Pentaho Data Integration y QlikView, dicho Data Mart nos permitirá almacenar solo la información requerida e ir eliminando aquellos datos que obstaculizan la labor del análisis. Gracias al desarrollo del Data Mart, la oficina de contabilidad pudo obtener los indicadores necesarios para la toma de decisiones respecto al rendimiento de su personal, mejorando así su desempeño y la distribución de trabajo.

Palabras claves: Data Mart, Pentaho, Data Integration, Hefesto, QlikView, Contabilidad.

NATIONAL UNIVERSITY OF SAN MARCOS
FACULTY OF SYSTEMS AND COMPUTER ENGINEERING
PROFESSIONAL SCHOOL OF SOFTWARE ENGINEERING

Implementation of a Dashboard using the Hefesto methodology for accounting offices in public universities, Case study: Accounting Office - UNMSM

Author: Gonzales Farro, Giancarlo Daniel

Advisor: Norberto Ulises, Roman Concha

Title: Professional Title of Software Engineer

Date: January 2018

ABSTRACT

The Accounting Offices of the National Universities, for the most part, do not measure the information or the performance of their personnel in search of efficiency and effectiveness. The Accounting Office of the National University of San Marcos (UNMSM) lacks information on the performance of workers for strategic decision making, this is very important to know where and how to resolve any failure or improve the decisions made. The objective of this work is to develop a Data Mart through the Hephaestus methodology, using the tools Pentaho Data Integration and QlikView, said Data Mart will allow us to store only the required information and to eliminate those data that hinder the work of the analysis. Thanks to the development of the Data Mart, the accounting office was able to obtain the necessary indicators to make decisions regarding the performance of its personnel, thus improving its performance and the distribution of work. of its personnel, improving its performance and the distribution of work.

Keys words: Data Mart, Pentaho, Data Integration, Hefesto, QlikView, Accounting.

Índice General

CARATULA EXTERNA O PASTA	vii
PÁGINA EN BLANCO.....	vii
CARATULA INTERNA	iii
FICHA CATALOGRAFICA	viiiv
DEDICATORIA	v
AGRADECIMIENTOS	vi
RESUMEN.....	vii
ABSTRACT	vviii
ÍNDICE GENERAL	ix
ÍNDICE DE FIGURAS.....	xii
ÍNDICE DE ILUSTRACIONES	xiii
ÍNDICE DE TABLAS	xiii
INTRODUCCIÓN	1
CAPITULO I PLANTEAMIENTO DEL PROBLEMA.....	3
1.1 ANTECEDENTES	3
1.1.1 DEL PROBLEMA	3
1.1.2 DE LA TÉCNICA	3
1.2 DEFINICIÓN DEL PROBLEMA	5
1.3 JUSTIFICACIÓN.....	5
1.4 OBJETIVOS	6
1.4.1 OBJETIVO GENERAL.....	6
1.4.2 OBJETIVOS ESPECÍFICOS.....	6
1.5 ALCANCES Y LIMITACIONES	6
1.6 PROPUESTA DE LA TESIS	7
CAPITULO II MARCO TEÓRICO	8

2.1	OFICINA GENERAL DE ECONOMÍA.....	8
2.2	OFICINA DE CONTABILIDAD.....	9
2.3	INTELIGENCIA DE NEGOCIOS.....	10
2.4	DATA WAREHOUSE.....	13
2.5	DATA MART.....	14
2.5.1	CLASIFICACIÓN.....	15
2.6	METODOLOGÍAS.....	16
2.6.1	METODOLOGÍAS KIMBALL.....	16
2.6.2	METODOLOGÍAS INMON.....	17
2.6.3	METODOLOGÍAS HEFESTO.....	20
2.7	HERRAMIENTAS.....	23
2.7.1	IBM DATA STAGE.....	23
2.7.2	JAVA – OCTOPUS.....	23
2.7.3	QLIK VIEW.....	23
2.7.4	MICROSTRATEGY.....	24
2.7.5	BUSINESS OBJECTS.....	24
2.7.6	PENTAHO COMMUNITY EDITION.....	24
CAPITULO III ESTADO DEL ARTE.....		26
3.1	METODOLOGÍAS PARA EL DESARROLLO DE SOLUCIONES BI.....	27
3.1.1	LA METODOLOGÍA DE KIMBALL PARA EL DISEÑO DE ALMACENES DE DATOS.....	27
3.1.2	DISEÑO DE UN ALMACÉN DE DATOS BASADO EN DATA WAREHOUSE ENGINEERING PROCESS (DWEP) Y HEFESTO.....	28
3.1.3	BUILDING AN EFFECTIVE DATA WAREHOUSE ARCHITECTURE.....	29
3.2	USO DE DATA MARTS COMO SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS.....	30
3.2.1	IMPLEMENTACIÓN DE UN DATA MART COMO UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE LOGÍSTICA.....	30
3.2.2	PROPUESTA DE DATA MART PARA EL ANÁLISIS DE LAS CONSECUENCIAS DE LA DINÁMICA EMPRESARIAL.....	31
3.2.3	ANÁLISIS COMPARATIVO.....	32
3.3	HERRAMIENTAS PARA EL DESARROLLO DE UNA SOLUCIÓN DE BUSINESS INTELLIGENCE.....	33
3.3.1	DATA WAREHOUSE PARA TOMA DE DECISIONES UTILIZANDO MICROSOFT SQL SERVER.....	33
3.3.2	UNA SOLUCIÓN BUSINESS INTELLIGENCE PARA LA GENERACIÓN DE INDICADORES Y CONTROL DE DESEMPEÑO.....	33
3.3.3	SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS, BAJO LA METODOLOGÍA DE RALPH KIMBALL PARA OPTIMIZAR LA TOMA DE DECISIONES EN EL DEPARTAMENTO DE FINANZAS.....	35

CAPITULO IV APOORTE TEÓRICO	37
4.1 SELECCIÓN DE LAS HERRAMIENTAS PARA EL DESARROLLO.....	37
4.1.1 SELECCIÓN DE LA METODOLOGÍA PARA EL DESARROLLO DEL DATA MART PROPUESTO	37
4.1.2 SELECCIÓN DE LA PLATAFORMA PARA EL DESARROLLO DEL DATA MART PROPUESTO.....	39
4.1.3 SELECCIÓN DE LA PLATAFORMA PARA LA PRESENTACIÓN DEL DATA MART PROPUESTO.....	40
4.2 DEFINICIÓN DE LA SOLUCIÓN	42
4.3 METODOLOGÍA A SEGUIR	43
CAPITULO V APOORTE PRÁCTICO	47
5.1 ANÁLISIS DE REQUERIMIENTOS.....	47
5.1.1 IDENTIFICAR PREGUNTAS	47
5.1.2 IDENTIFICAR INDICADORES Y PERSPECTIVAS DE ANÁLISIS.....	47
5.1.3 MODELO CONCEPTUAL.....	48
5.1.4 DISEÑO DE LA ARQUITECTURA	49
5.2 ANÁLISIS DEL PROCESAMIENTO DE TRANSACCIONES EN LÍNEA.....	51
5.2.1 DETERMINACIÓN DE INDICADORES.....	51
5.2.2 ESTABLECER CORRESPONDENCIAS.....	52
5.2.3 MODELO CONCEPTUAL AMPLIADO.....	54
5.3 MODELO LÓGICO DEL DATA MART	55
5.3.1 TIPO DE MODELO LÓGICO DEL DATA MART.....	55
5.3.2 TABLAS DE DIMENSIONES.....	55
5.3.3 TABLAS DE HECHOS.....	57
5.3.4 UNIONES	58
5.4 INTEGRACIÓN DE DATOS.....	59
5.4.1 CARGA INICIAL.....	60
5.4.2 VISUALIZACIÓN	68
CAPITULO VI ANÁLISIS DE RESULTADOS	74
6.1 RESULTADO DE LOS INDICADORES.....	74
6.1.1 A NIVEL DE FUENTE DE FINANCIAMIENTO	74
6.1.2 A NIVEL DE MODALIDAD DE COMPRA	74
CONCLUSIONES Y RECOMENDACIONES	75
7.1 CONCLUSIONES	75
7.2 RECOMENDACIONES	75
REFERENCIAS BIBLIOGRÁFICAS.....	77

ANEXOS	80
GLOSARIO DE TERMINOS	80

Índice de figuras

Figura 1 - Arquitectura de un Almacén de Datos	28
Figura 2 - El proceso unificado y proceso de ingeniería de almacenes de datos.....	29
Figura 3 - Modelo Bill Inmon	30
Figura 4 - Vista Toolbar de SQL Server	33
Figura 5 - Reporte Requerimientos por Ambiente	34
Figura 6 - Solución Propuesta.....	42
Figura 7 Arquitectura de la SoluciónFuente: Elaboración Propia	49
Figura 8 Establece correspondencias	53
Figura 9 - Modelo conceptual ampliado	54
Figura 10 - Tabla de dimensión Sectorista Fuente: Elaboración Propia.....	55
Figura 11 - Tabla de dimensión Dependencia Fuente: Elaboración Propia	56
Figura 12 - Tabla de dimensión Trabajador.....	56
Figura 13 - Tabla de Hechos Registro Contable.....	57
Figura 14 - Modelo Dimensional	58
Figura 15 - Conexión a Base de Datos I	61
Figura 16 - Conexión a Base de Datos II	61
Figura 17 - Entrada de datos a tablas I.....	61
Figura 18 - Entrada de datos II	62
Figura 19 - Ordenar datos I.....	62
Figura 20 - Ordenar datos II Fuente: Pentaho Data Integration - 2016	63
Figura 21 - Dar formato a datos I	63
Figura 22 - Dar formato a datos II	64
Figura 23 - Datos formatos a datos III	64
Figura 24 - Conexión de tareas.....	65
Figura 25 - Agrupar datos I	65
Figura 26 - Agrupar datos II Fuente: Pentaho Data Integration - 2016.....	66
Figura 27 - Insertar datos I	66
Figura 28 - Insertar datos II Fuente: Pentaho Data Integration - 2016	67
Figura 29 - Flujo de Trabajo Fuente: Pentaho Data Integration - 2016	68
Figura 30 - Conectarse al Data Mart I.....	68
Figura 31 - Conectarse al Data Mart II.....	69
Figura 32 - Conectarse al Data Mart III Fuente: QlikView Personal Edition - 2016.....	69

Figura 33 - Generar gráficos I	70
Figura 34 - Generar gráficos II	70
Figura 35 - Generar gráficos III	70
Figura 36 - Generar gráficos IV	71
Figura 37 - Registros procesados por año Fuente: QlikView Personal Edition - 2016	71
Figura 38 - Registros procesados por FF.....	72
Figura 39 - Registros procesados por MC.....	72
Figura 40 - Registros procesados por TO.....	73

Índice de ilustraciones

Ilustración 1- Los componentes de inteligencia de negocios.....	12
Ilustración 2 – Modelo Datawarehouse	13
Ilustración 3 - Componentes de la MetaData	19
Ilustración 4 - Pasos Metodología Hefesto	21
Ilustración 5 - MER Esquema Contabilidad	52

Índice de tablas

Tabla 1 - Investigaciones	26
Tabla 2 - Análisis Comparativo I	32
Tabla 3 - Tablas de Productos Pentaho	34
Tabla 4 - Puntajes Metodologías	38
Tabla 5 - Cuadro de evaluación de las metodologías	38
Tabla 6 - Cuadro de evaluación de plataformas de BI.....	40
Tabla 7 - Factor de Análisis	41
Tabla 8 - Cuadro de evaluación de plataformas de BI.....	42
Tabla 9 - Resultado de los indicadores	74

INTRODUCCIÓN

Inteligencia de Negocios hace referencia a la capacidad de organizar y explotar la información que se encuentra alojada en bases de datos de distintas fuentes, mediante procesos iterativos, los cuales permiten el análisis de información que facilita la interpretación y correcta comprensión de los datos, con el cual podemos obtener una ventaja competitiva.

En el Perú, la mayoría de las entidades educativas del sector público no cuentan con una solución de inteligencia de negocios, que les permita explotar su inmensa cantidad de información almacenada a lo largo de sus años de vida.

La Universidad Nacional Mayor De San Marcos, no es esquivada de dicho problema, por ello busca explotar su información en busca de indicadores que le permitan mejorar el rendimiento de su personal.

El presente trabajo implementa un DataMart como una solución para aplicar Inteligencia de Negocios en la oficina de Contabilidad de la Universidad Nacional Mayor de San Marcos y así medir el rendimiento del personal de dicha entidad.

El proyecto de tesis se divide en los siguientes capítulos:

En el capítulo 1 se describe el planteamiento del problema, donde se muestran anteriores investigaciones relacionadas a soluciones de inteligencia de negocios, mediante la creación de un Data Mart.

En el capítulo 2 se describe el marco teórico, donde se explican los conceptos fundamentales sobre Inteligencia de Negocios, Data Mart, Datawarehouse, Metodologías Business Intelligence, los cuales son abordados a lo largo de toda esta investigación.

En el capítulo 3 se describe el estado del arte, donde se describen y analizan algunas de las metodologías existentes para el desarrollo de soluciones de inteligencia de negocios, así mismo se presentan herramientas para el desarrollo de Data Marts y visualización de los mismos.

En el capítulo 4 se describe la metodología usada para la implementación de la solución de inteligencia de negocios, luego se explica las fases más importantes del desarrollo.

En el capítulo 5 se describe el caso de estudio y se explica el uso de la herramienta, sus resultados y validaciones.

En el capítulo 6 se realiza un análisis de los resultados obtenidos, mostrando las ventajas del desarrollo de esta tesis.

En el capítulo 7 se presentan las conclusiones generales y específicas, además de los trabajos futuros.

CAPITULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 ANTECEDENTES

1.1.1 DEL PROBLEMA

Según Alejandro Rojas Zaldivar (2014) en su investigación denominada "IMPLEMENTACIÓN DE UN DATA MART COMO SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS, BAJO LA METODOLOGÍA DE RALPH KIMBALL PARA OPTIMIZAR LA TOMA DE DECISIONES EN EL DEPARTAMENTO DE FINANZAS DE LA CONTRALORÍA GENERAL DE LA REPÚBLICA", para la Universidad de San Martín de Porres la cual propone la implementación de una solución de inteligencia de negocios para el Departamento de Finanzas de la Contraloría General de la República del Perú, de modo que, esta herramienta automatice el procedimiento de consulta de información y, que en un entorno amigable, permita a los usuarios acceder a información en menor tiempo y en un repositorio que facilite acceder a información histórica

El resultado de su investigación fue identificar los procesos que permiten a cabo la toma de decisiones en el Departamento de Finanzas de la CGR estableciendo las medidas que se utilizan en la realización de reportes para la toma de decisiones y se construyó un modelo OLAP utilizando la herramienta QlikView brindándole flexibilidad al usuario al realizar las diferentes consultas de información del Departamento de Finanzas.

1.1.2 DE LA TÉCNICA

Tesis desarrollada por Julio Yalan Castillo y Luis Palomino Paniara (2012), en su investigación denominada "IMPLEMENTACIÓN DE UN DATA MART COMO UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE LOGÍSTICA DE T-IMPULSO", para la Universidad Nacional Mayor de San Marcos, nos muestra la manera de implementar un Data Mart como solución de inteligencia de negocios, que permitirá desarrollar Inteligencia de Negocios sobre

el área de Logística de una empresa que brinda servicios de Tercerización y Outsourcing.

El resultado de su investigación concluyo que el Data Mart permitió apoyar al área de Logística en la toma de decisiones, a través de la entrega oportuna y relevante de información.

Tesis desarrollada por Sofia Anabel Bustos Barrera y Veronica Nathaly Mosquera Artieda (2013) en su investigación titulada “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN BUSINESS INTELLIGENCE PARA LA GENERACIÓN DE INDICADORES Y CONTROL DE DESEMPEÑO, EN LA EMPRESA OTECEL S.A., UTILIZANDO LA METODOLOGÍA HEFESTO V2.0”, para la Escuela Politécnica del Ejercito de Sangolqui – Ecuador nos muestra la implementación de una solución de inteligencia de negocios para el manejo de datos de la empresa de telefonía celular OTECEL SA, basado en la metodología Hefesto y haciendo uso de la herramienta Open Source Pentaho.

Como resultado de la implementación de la solución de inteligencia de negocios se logró reducir el tiempo de respuesta en la recuperación de la información a ser analizada por Jefes e Ingenieros del Área de Pruebas, que mediante reportes dinámicos prediseñados y la vista de análisis obtienen acceso a los indicadores y control de las tareas.

Según Alvarez Calvopiña Cristian Dario (2014) y en su investigación denominada “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART UTILIZANDO HERRAMIENTAS OPEN SOURCE PARA LA UNIDAD ADMINISTRATIVA Y FINANCIERA DE LA ESPE”, para la Universidad de las Fuerzas Armadas ESPE, se buscó desarrollar un conjunto de Data Marts (Data Warehouse) que serviría para simplificar procesos, mejorar el manejo y la consolidación de la información gestionada por la Dirección Financiera y Talento Humano utilizando la metodología Ralph Kimball y como herramienta BI Open Source se empleó Pentaho Community

Los resultados mostraron que, mediante la consolidación y depuración de información, se pudo reducir el tiempo de respuesta en la generación de informes, logrando un análisis eficiente de la información de las consultas hechas a las bases de datos.

1.2 DEFINICIÓN DEL PROBLEMA

La Oficina de Contabilidad de la Universidad Nacional Mayor de San Marcos, maneja gran cantidad de información, pero esta información no está organizada de acuerdo a las necesidades de los jefes de oficina, una de estas necesidades es poder medir el rendimiento de su personal, de tal forma que cuando necesitan reportes o informes lo realizan de forma manual, a través de un documento Excel, que es solicitado al Sistema Quipucamayoc Módulo de Contabilidad.

1.3 JUSTIFICACIÓN

La implementación de una solución de Inteligencia de Negocios permitirá tener tiempos de respuestas menores en la obtención de datos procesados y el mejor manejo de los mismos, ayudando en la toma de decisiones. (Castillo, 2012)

Algunas iniciativas estratégicas de negocio en las que se pueden apalancar beneficios que justifiquen soluciones de BI son (Almeida, 2015):

- Crear una nueva capacidad del negocio como un nuevo producto o servicio
- Atacar una desventaja competitiva, o crear una nueva ventaja competitiva.
- Incrementar los ingresos, incrementando ventas o reduciendo costos.
- Incrementar la satisfacción del cliente.
- Incrementar la satisfacción del empleado.
- Cumplir con nuevas regulaciones.
- Mejorar la seguridad.
- Reducir el tiempo para entregar un producto o servicio.

Por ello se busca reorganizar y procesar la información de la base de datos de la Oficina de Contabilidad, para superar las carencias de información oportuna para la toma de decisiones, a través de reportes y consultas determinadas.

1.4 OBJETIVOS

1.4.1 OBJETIVO GENERAL

Diseñar e Implementar un Dashboard para la oficina de contabilidad de la UNMSM por medio de la herramienta Open Source Pentaho, utilizando la metodología Hefesto.

1.4.2 OBJETIVOS ESPECÍFICOS

- Identificar las necesidades de la oficina de contabilidad para realizar un análisis y visualización de su información.
- Seleccionar la metodología para el desarrollo de la solución de inteligencia de negocios.
- Implementar un Data Mart utilizando las herramientas Pentaho Community Edition.
- Utilizar la herramienta QlikView para diseñar y crear el dashboard correspondiente al Data Mart.

1.5 ALCANCES Y LIMITACIONES

Se realizará la implementación de la Solución de Inteligencia de Negocios, para la obtención de reportes automatizados con información precisa, en el área de pruebas del Sistema Quipucamayoc Modulo de Contabilidad de la UNMSM.

Se realizará el diseño de un Data Mart para la automatización y depuración de la información en la base de datos con la metodología seleccionada.

La solución solo será utilizada por la Oficina de Contabilidad de la UNMSM, por motivos que la información provista por esta entidad pública es solo de contabilidad, sin intervención de los demás Departamentos.

La cantidad de tiempo disponible para completar la solución será limitado, ya que solo una persona desarrolla el proyecto.

Por motivos de políticas internas de seguridad de información del Sistema Quipucamayoc, los datos que se utilizaran han sido tomados desde el servidor de pruebas.

1.6 PROPUESTA DE LA TESIS

La solución que se plantea es la implementación de una solución de inteligencia de negocios que este conformada por la información obtenida de la base de datos Oracle del Sistema Quipucamayoc Modulo de Contabilidad, obteniendo un análisis eficiente y simplificado de la información, disminuyendo considerablemente el tráfico generado en las consultas hacia la base de datos y por consiguiente liberando recursos dentro del servidor.

Para ello la utilización de herramientas de software libre, Pentaho Data Integration, junto con herramientas optimizadas para la visualización de datos, QlikView, de la mano con la Metodología Hefesto para el desarrollo del Data Mart.

CAPITULO II

MARCO TEÓRICO

En el presente capítulo se define con mayor detalle los conceptos relevantes que abarca esta tesis. El marco teórico incluye una breve descripción de lo que significa Inteligencia de Negocios presentando además algunos de sus elementos y beneficios. Así mismo, también se describe algunas de las metodologías más conocidas para el desarrollo de una solución de Inteligencia de Negocios. Finalmente, se describe algunos conceptos sobre herramientas para Inteligencia de Negocios, conceptos que serán de ayuda para la implementación de la presente tesis.

2.1 OFICINA GENERAL DE ECONOMÍA

La Oficina General de Economía es el órgano dependiente de la Dirección General de Administración que tiene a su cargo la administración de los recursos económicos y financieros de la Universidad¹.

La Oficina General de Economía tiene las siguientes funciones (UNMSM, 2015):

- a) Conducir y desarrollar los procesos de centralización, revisión y registro de las operaciones contables de la Universidad en concordancia con las disposiciones legales y normativas vigentes.
- b) Conducir y desarrollar los procesos de formulación, revisión e interpretación de los estados financieros de la Universidad.
- c) Hacer cumplir los procedimientos de formulación de las conciliaciones de las cuentas bancarias de la Universidad.
- d) Controlar la ejecución del gasto de las asignaciones presupuestarias de la Universidad, en concordancia con la Ley de Presupuesto General de la República y demás disposiciones vigentes sobre el particular.

¹ <http://www.unmsm.edu.pe/home/organizacion/economia>

- e) Administrar y controlar la utilización de los recursos financieros de la Universidad.
- f) Conducir y desarrollar el proceso de adquisiciones de bienes requeridos por los órganos de la Universidad.
- g) Desarrollar y controlar las actividades de recuperación de bienes requeridos por los órganos de la Universidad.
- h) Establecer los lineamientos para la adecuada gestión de los sistemas de Contabilidad, Abastecimiento y Tesorería, así como de los servicios prestados por el Autoseguro de Accidentes, con sujeción a las normas vigentes.
- i) Emitir opinión técnica en materia de gestión económica y financiera.
- j) Desempeñar las demás funciones que en el ámbito de su competencia le asigne el Director General de Administración.

Su estructura orgánica es la siguiente:

- Oficina de Contabilidad
- Oficina de Tesorería
- Oficina de Abastecimiento
- Oficina de Autoseguro de Accidentes

Donde nos enfocaremos en la Oficina de Contabilidad.

2.2 OFICINA DE CONTABILIDAD

La Oficina de Contabilidad es el órgano dependiente de la Oficina General de Economía que tiene las siguientes funciones (UNMSM, Reglamento de Organización y Funciones (R.O.F.), 2016):

- a) Implementar las normas y procedimientos contables a nivel institucional, así como mantenerlos permanentemente actualizados.
- b) Efectuar los procesos de consolidación, registro y mantenimiento actualizado de las operaciones contables de la Universidad en concordancia con lo establecido en las disposiciones legales y normas vigentes.

c) Establecer procesos de formulación, análisis e interpretación de los estados financieros de la Universidad.

d) Controlar la ejecución del gasto de las asignaciones presupuestarias en concordancia con la Ley de Presupuesto General y las disposiciones complementarias que se emiten.

e) Administrar el desarrollo de las actividades para la elaboración de la información contable y de la ejecución del gasto de las asignaciones presupuestarias, requeridas por la Alta Dirección de la Universidad, así como por los organismos externos de control, de acuerdo a las disposiciones legales y normas establecidas.

f) Administrar el desarrollo de las actividades de conciliaciones de los procesos bancarios.

g) Administrar el desarrollo de las actividades de archivo y control de la información y documentación contable.

h) Realizar las demás funciones que le asigne el Jefe de la Oficina General de Economía.

2.3 INTELIGENCIA DE NEGOCIOS

“Una interesante definición para inteligencia de negocios o BI, por sus siglas en inglés, según el Data Warehouse Institute, lo define como la combinación de tecnología, herramientas y procesos que me permiten transformar mis datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados” (Oracle, 2015).

En el “Boletín de Asesoría Gerencial” proporcionado por Espíñera, Sheldon y asociados, indican que la definición de inteligencia de negocios incluye una amplia categoría de metodologías, aplicaciones y tecnologías que permiten reunir, acceder, transformar y analizar los datos, transacciones e información no estructurada (interna y externa), con el propósito de ayudar a los usuarios de una compañía a tomar mejores decisiones de negocio. (Espíñera, 2015).

“BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores” (Lluís Cano, 2012).

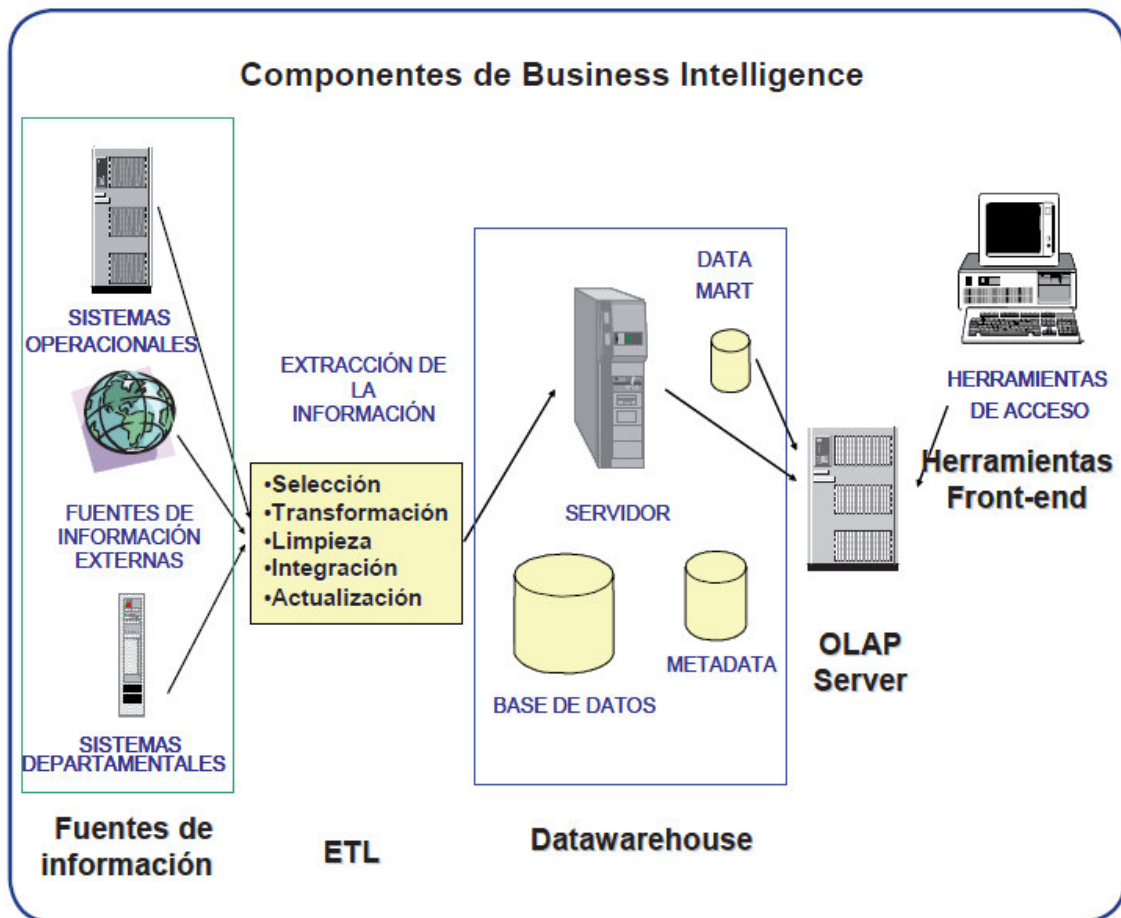
En el libro Business Intelligence: Josep Lluís Cano, menciona que el objetivo básico de inteligencia de negocios es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones.

Asimismo, describe tres tipos de beneficios que se pueden obtener a través del uso de inteligencia de negocios, los cuales son:

- Beneficios tangibles: por ejemplo, reducción de costos, generación de ingresos, reducción de tiempos para las distintas actividades del negocio.
- Beneficios intangibles: el hecho de que tengamos disponible la información para la toma de decisiones hará que más usuarios la utilicen para tomar decisiones y mejorar nuestra posición competitiva.
- Beneficios estratégicos: la formulación de estrategias, especificarán a qué clientes, mercados o con qué productos dirigirnos.

A su vez, el mismo autor, nos indica los componentes de inteligencia de negocios (Ver Ilustración 1):

Ilustración 1- Los componentes de inteligencia de negocios



Fuente: Business Intelligence: competir con información - Lluís Cano, 2012.

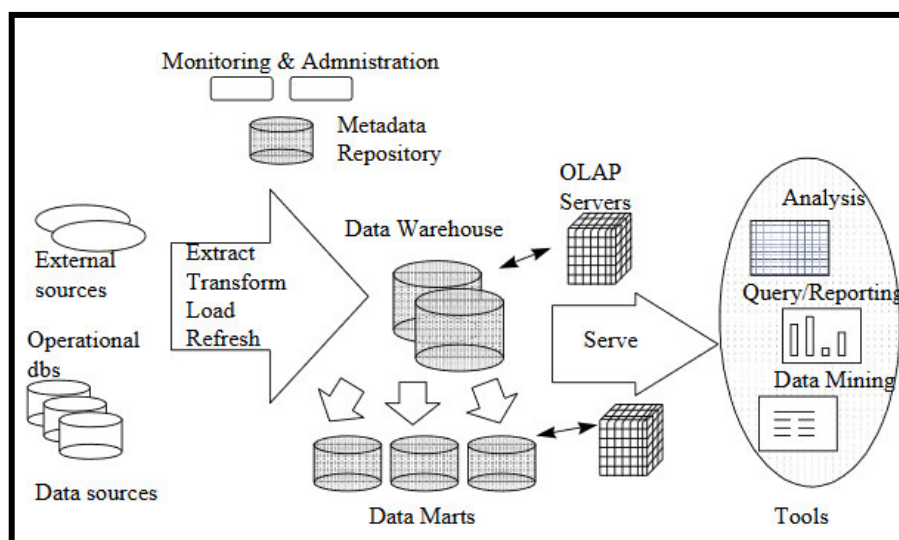
- Fuentes de información, de las cuales partiremos para alimentar de información el Data Warehouse.
- Proceso ETL de extracción, transformación y carga de los datos en la Data Warehouse. Antes de almacenar los datos en una Data Warehouse, éstos deben ser transformados, limpiados, filtrados y redefinidos. normalmente, la información que tenemos en los sistemas transaccionales no está preparada para la toma de decisiones.
- La propia Data Warehouse: se busca almacenar los datos de una forma que maximice su flexibilidad, facilidad de acceso y administración.
- El motor OLAP, nos provee capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos. En la actualidad, existen otras alternativas tecnológicas al OLAP.

Con las definiciones propuestas se concluye que Business Intelligence brinda herramientas que ayudan a la toma de decisiones, puesto que presenta la información de manera ordenada, rápida y oportuna, permite a las empresas profundizar en su análisis, ya que se tiene indicadores claves que ayudan a la gestión y manejo de la misma.

2.4 DATA WAREHOUSE

“Un almacén de datos (del inglés data warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.” (Bustos Barrera & Mosquera Artieda, 2013)

Ilustración 2 – Modelo Datawarehouse



Fuente: An Overview of Data Warehousing and OLAP Technology - **Surajit Chaudhuri & Umeshwar Dayal, 1997.**

Datawarehouse almacena una gran cantidad de información histórica del negocio, aísla los sistemas operacionales de las necesidades de información para la gestión. Un cambio en los sistemas operacionales no debe afectar al DW/DM.

El DataWarehouse se alimenta a partir de los datos operacionales mediante las herramientas ETL (extract, transform, load) (Ver ilustración 2).

Extraer: hace referencia a la captura de información desde los sistemas operacionales, necesidad de integración:

- Bases de datos
- Ficheros planos
- Documentos impresos

Transformar: hace referencia a la adaptación de los datos fuente al formato destino definido en el DataWarehouse:

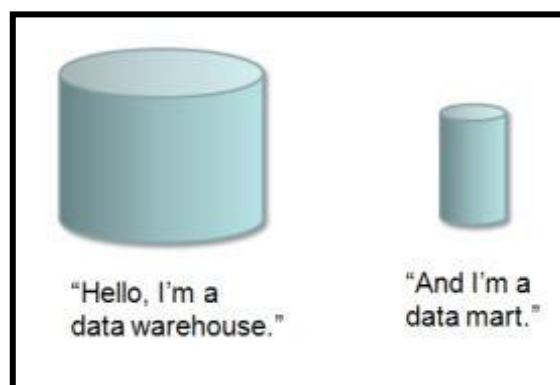
- Agregar datos numéricos
- Transponer información
- Componer unos textos a partir de otros

Cargar: hace referencia al proceso en que los nuevos datos son finalmente almacenados en el DataWarehouse en su formato definitivo. (Bustos Barrera & Mosquera Artieda, 2013)

2.5 DATA MART

Un Data Mart 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 Data Mart puede ser alimentado desde los datos de un datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información (Ver ilustración 3).

Ilustración 3 - Datawarehouse or Data Mart



Fuente: Data Warehouse Vs Data Mart - **Standen, 2015.**

Por tanto, para crear el Data Mart 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 datawarehouse, 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.

Los Data Marts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas (Sinnexus, 2015):

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos

2.5.1 CLASIFICACIÓN

Se observa la clasificación del Data Mart según (Inmon, 2002).

2.5.1.1 DATA MART DEPENDIENTE

Los Data Mart dependientes son aquellos que reciben los datos desde una Data warehouse. En este tipo de Data Mart la fuente de los datos es única.

2.5.1.2 DATA MART INDEPENDIENTE

Son aquellos que toman sus datos directamente desde los sistemas transaccionales y no dependen de otros Datawarehouse. Este tipo de Data Mart se alimenta generalmente de las organizaciones.

2.5.1.3 DATA MART HÍBRIDO

Los Data Mart híbridos permiten combinar las fuentes de datos de un Data Warehouse corporativo con otras fuentes de datos tales como sistemas transaccionales y/o operacionales.

2.6 METODOLOGÍAS

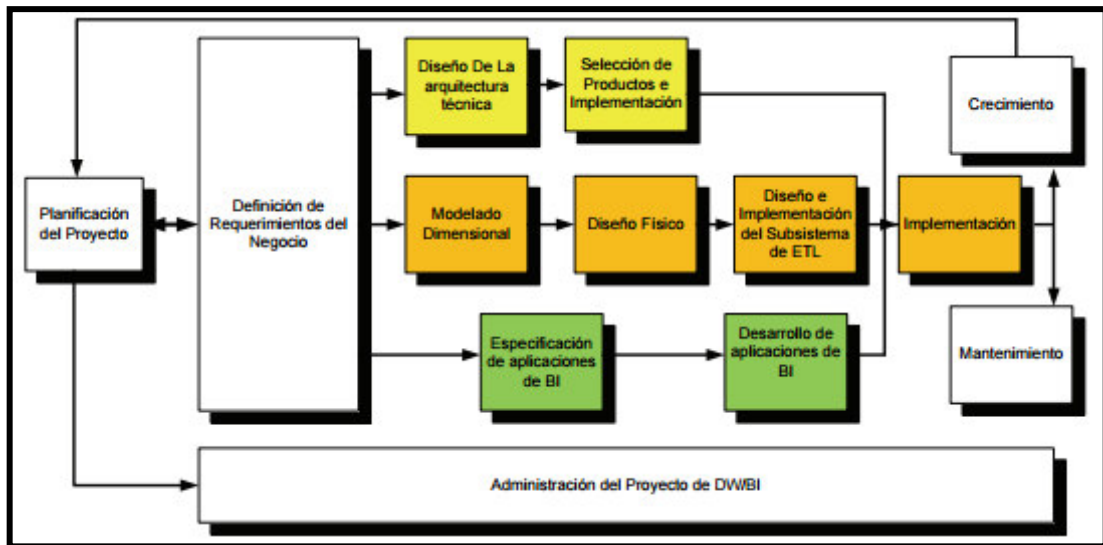
2.6.1 METODOLOGÍAS KIMBALL

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos:

- *Centrarse en el negocio:* Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- *Construir una infraestructura de información adecuada:* Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- *Realizar entregas en incrementos significativos:* Crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.
- *Ofrecer la solución completa:* Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

La construcción de una solución de DW/BI (Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. Las tareas de esta metodología (ciclo de vida) se muestran en la ilustración 4.

Ilustración 4 - Tareas de la metodología de Kimball, denominada Business Dimensional Lifecycle



Fuente: The Data Warehouse Lifecycle Toolkit - Kimball, 2008.

De la ilustración 4, podemos observar dos cuestiones. Primero, hay que resaltar el rol central de la tarea de definición de requerimientos. Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes. También tiene influencia en el plan de proyecto (nótese la doble fecha entre la caja de definición de requerimientos y la de planificación). En segundo lugar, podemos ver tres rutas o caminos que se enfocan en tres diferentes áreas: (Rivadiera, 2010)

- *Tecnología (Camino Superior):* Implica tareas relacionadas con software específico, por ejemplo, Microsoft SQL Analysis Services.
- *Datos (Camino del medio):* En la misma se diseñará e implementará el modelo dimensional, y desarrollará el subsistema de Extracción, Transformación y Carga (Extract, Transformation, and Load - ETL) para cargar el DW.
- *Aplicaciones de Inteligencia de Negocios (Camino Inferior):* En esta ruta se encuentran tareas en las que se diseñará y desarrollará las aplicaciones de negocios para los usuarios finales.

2.6.2 METODOLOGÍAS INMON

En el capítulo dos del libro “Building the Data Warehouse”, Bill Inmon define al Data Warehouse (almacén de datos), de la siguiente manera:

“A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions. The data warehouse contains granular corporate data”. (Inmon, 2002)

Esta definición del Data Warehouse ha sido conceptualizada en términos de las características del repositorio de datos (Pérez, 2015):

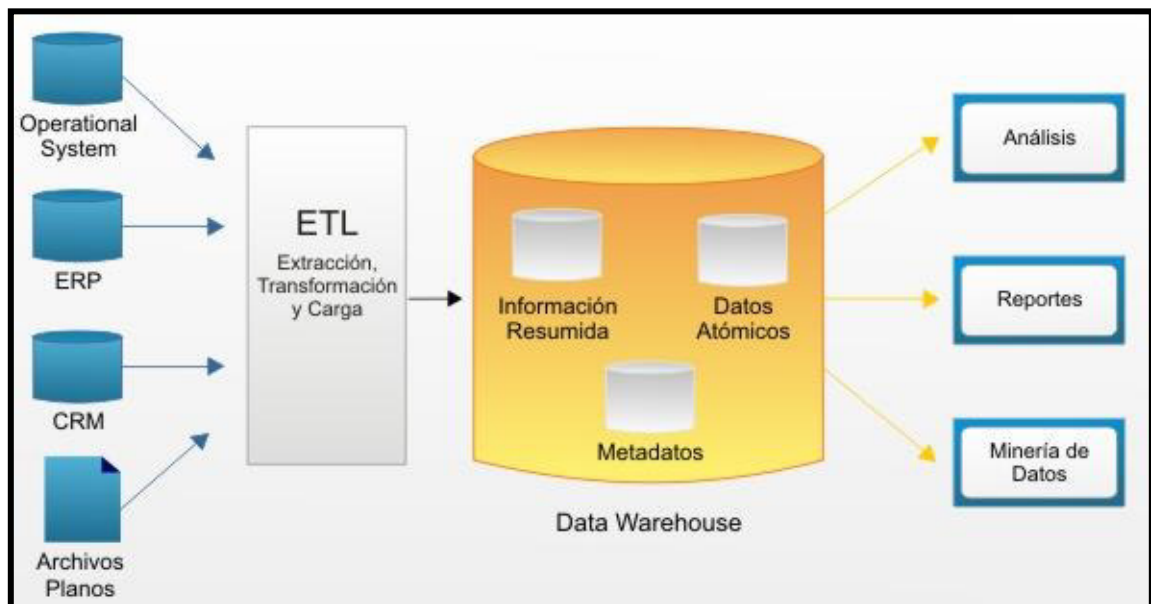
Orientado a temas: Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

Integrado: La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

No volátil: La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.

Variante en el tiempo: Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones (Ver ilustración 5).

Ilustración 5 - Arquitectura del Data Warehouse



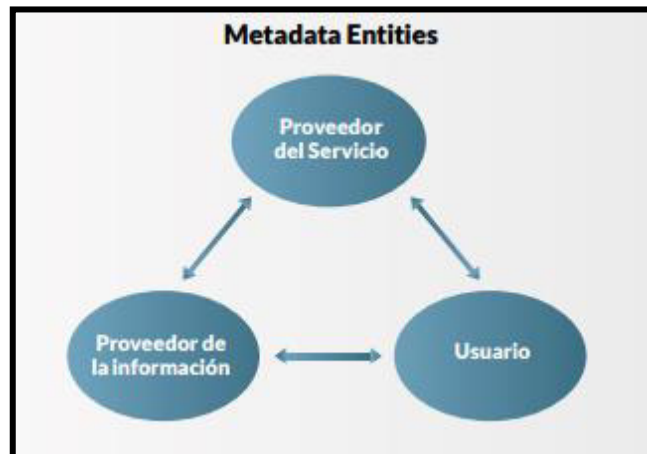
Fuente: Building the Data Warehouse - Inmon, 2002.

Otra característica del Data Warehouse es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la

información, su periodicidad de actualización, su fiabilidad, forma de cálculo, etc. (Ver Ilustración 6).

Los metadatos serán los que permiten simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

Ilustración 3 - Componentes de la MetaData



Fuente: Elaboración Propia

Estos son los objetivos que deben cumplir los metadatos, según las actividades de las personas a las que va dirigido:

Dar soporte al usuario final, ayudándole a acceder al Data Warehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de Business Intelligence como DSS², EIS³ o CMI⁴.

² DSS: Sistema de Apoyo a las Decisiones (DSS por sus siglas en inglés Decision support system). En general, podemos decir que un DSS es un sistema informático utilizado para servir de apoyo, más que automatizar, el proceso de toma de decisiones. La decisión es una elección entre alternativas basadas en estimaciones de los valores de esas alternativas. El apoyo a una decisión significa ayudar a las personas que trabajan solas o en grupo a reunir inteligencia, generar alternativas y tomar decisiones.

³ EIS: Sistema de Información Ejecutiva (Executive information system, EIS por sus siglas en inglés) es una herramienta de Inteligencia empresarial (Business Intelligence, BI), orientada a usuarios de nivel gerencial, que permite monitorizar el estado de las variables de un área o unidad de la empresa a partir de información interna y externa a la misma.

⁴ CMI: El concepto de Cuadro de Mando Integral – CMI (Balanced Scorecard – BSC) es un sistema de administración o sistema administrativo, que va más allá de la perspectiva financiera con la que los gerentes acostumbran evaluar la marcha de una empresa. Es una herramienta de administración de empresas que muestra continuamente cuándo una compañía y sus empleados alcanzan los resultados definidos por el plan estratégico.

Dar soporte a los responsables técnicos del Data Warehouse en aspectos de auditoria, gestión de la información histórica, administración del Data Warehouse, elaboración de programas de extracción de la información, especificación de las interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos, etc.

En términos generales el Data Warehouse es una herramienta que hace mucho más fácil el análisis de la información lo que permite facilitar la toma de decisiones (Pérez, 2015), o como el mismo Bill Inmon resume:

“The data warehouse is the heart of the architected environment, and is the foundation of all DSS processing. The job of the DSS analyst in the data warehouse environment is immeasurably easier than in the classical legacy environment because there is a single integrated source of data (the data warehouse) and because the granular data in the data warehouse is easily accessible”. (Inmon, 2002)

2.6.3 METODOLOGÍAS HEFESTO

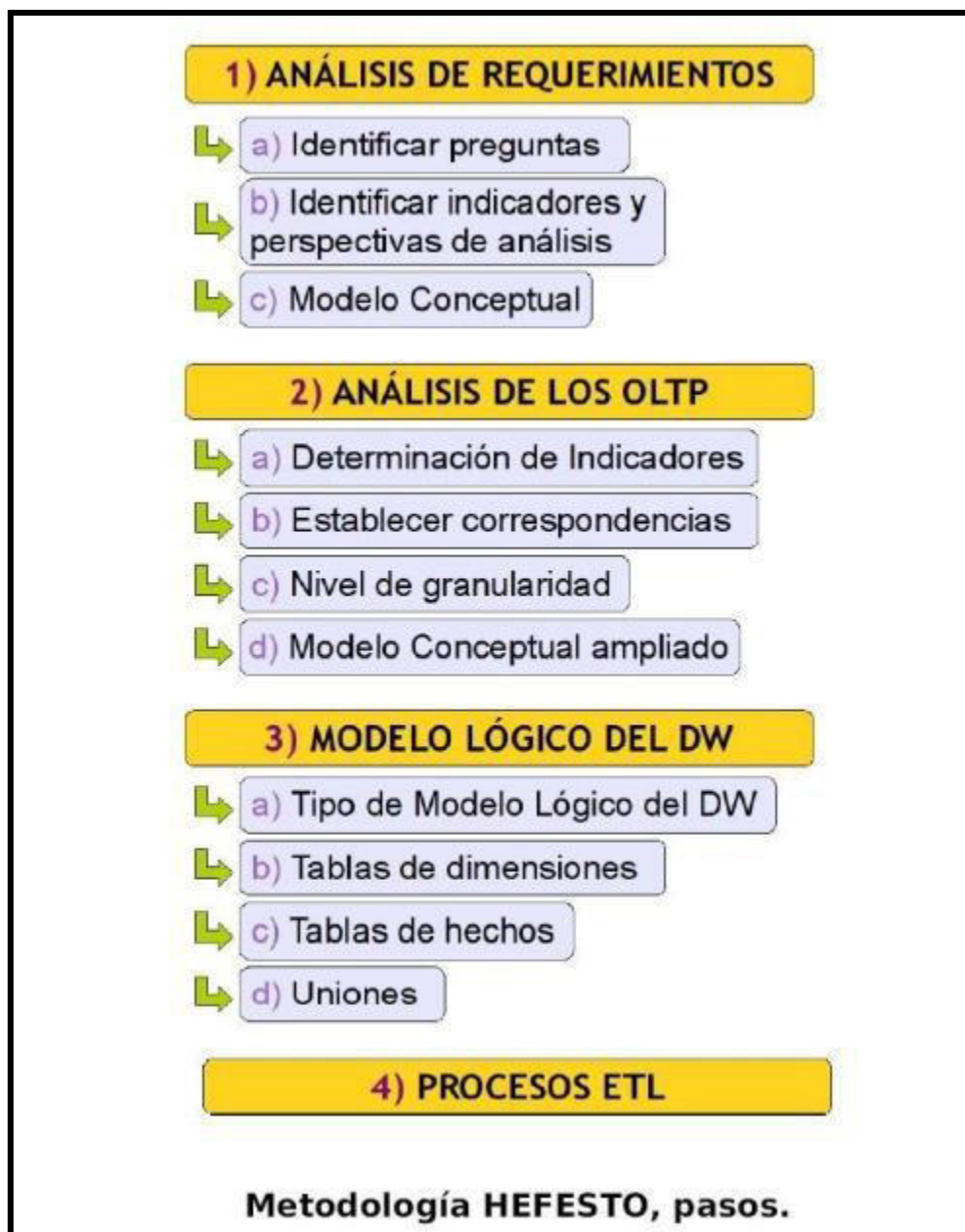
Es una metodología desarrollada recientemente por el Ing. Darío Bernabeu (Argentina), cuya finalidad es guiar el modelado del Data Warehouse de manera sencilla, ordenada y de forma intuitiva. Con la ayuda de esta metodología es posible obtener, en poco tiempo, una visión integral del problema y del modelo constructivo, suficientemente sólida, para desarrollar una implementación completa del sistema requerido. Estas, son las razones más fuertes que se tuvieron en cuenta para decidir su utilización en el proyecto de tesis.

La propuesta Hefesto fue impulsada en base a una amplia y exhaustiva investigación realizada, la comparación de metodologías existentes y experiencias previas aportadas por el mismo autor, en el estudio y confección de distintos almacenes de datos construidos. Actualmente la misma se encuentra en la versión 2.0.

2.6.3.1 PROCESO DE DISEÑO

Para llevar a cabo la construcción controlada de una solución de Inteligencia de Negocio, de manera de minimizar los riesgos del proyecto durante la ejecución, la metodología Hefesto sugiere distribuir el proceso en cuatro fases (Ver ilustración 7):

Ilustración 4 - Pasos Metodología Hefesto



Fuente: Metodología Hefesto – Dario, 2010.

Hefesto propone, en la primera etapa, identificar de las necesidades de los usuarios para conocer las preguntas claves del negocio. Luego, se continúa con el diseño de los indicadores que van a responder a estos interrogantes junto con las perspectivas de análisis requeridas. Más tarde, se construirá el modelo conceptual de datos del DW.

En etapas posteriores, se deberán analizar las fuentes de datos OLTP para determinar los factores constructivos de cada indicador, se señalarán las

correspondencias con las fuentes de datos y, por último, se determinará el campo de estudio de cada perspectiva.

Una vez hecho esto, se podrá continuar con la construcción del modelo lógico del depósito de datos. En esta etapa, deberá quedar definido el tipo de esquema utilizado en la implementación. También, se construirán las tablas de dimensiones y las tablas de hechos para, finalmente, efectuar las uniones correspondientes.

Por último, se definirán los procesos de extracción, transformación y carga para capturar los datos desde las diferentes fuentes de datos y, finalmente, almacenarlos en el DW.

2.6.3.2 VENTAJAS DEL MODELO HEFESTO

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen en la implementación.
- Es independiente de la estructura física que contengan el DW y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para Data Warehouse como para Data Mart.

2.7 HERRAMIENTAS

2.7.1 IBM DATA STAGE

Herramienta de IBM para crear y mantener rápidamente los Data Marts y los data warehouses. Soporta la extracción, la integración y la transformación de volúmenes altos de datos complejos.

Las soluciones con este software especializado requieren de un hardware de múltiples procesadores, esto permite que DataStage pueda brindar altos niveles de escalabilidad y trabajar eficientemente con grandes volúmenes de datos. Numerosos "conectores" soportan una amplia variedad de formatos de datos de origen y destino, que incluye a los famosos sistemas de gestión de bases de datos de IBM y OEM (licenciamiento por volumen), las fuentes de datos ODBC (Origen de Datos), aplicaciones de terceros, mensajes en tiempo real generados por programas de colocación de mensajes en colas, XML, servicios Web, y los formatos de archivos más utilizados. (IBM, 2015)

2.7.2 JAVA – OCTOPUS

Es una herramienta de extracción y carga dentro del framework de Java que permite realizar transformaciones definidas en fichero XML independiente de la base de datos a utilizar, se utiliza la conexión JDBC.

Diferentes tipos de bases pueden ser fuente de datos (MS SQL, Oracle, DB2, JDBC-ODBC con Excel y Access, Misal, CSV, y XML). (JAVA, 2015)

2.7.3 QLIK VIEW

Muestra la forma de manejo de los datos que se cargan a la memoria RAM para hacer más rápido el acceso a los datos, con respuestas inmediatas. La visualización de dashboards y reportes dinámicos. Visualiza datos en tablas, gráficos en 2-D o 3-D, puede hacer un Zoom in Zoom out de los gráficos y además revisar el detalle de los mismos. Se puede crear nuevas visualizaciones para una comprensión más profunda. (QLIK, 2015)

2.7.4 MICROSTRATEGY

Herramienta de Inteligencia de negocios que permite crear informes y análisis de datos almacenados en una base de datos relacional y de otras fuentes. Esta herramienta ofrece una flexibilidad y eficiencia en la explotación y minería de datos.

Permite el soporte para los tres niveles BI (del inglés Business Intelligence – Inteligencia de Negocios) que son el estratégico, analítico y operacional y, a la vez, ofrece un proceso sencillo de migración para fusionar aplicaciones de BI departamentales y de equipos de trabajo dentro de un marco que complementa la información de toda la organización. La seguridad implementada permite mantener autonomía e independencia en cada área y al mismo tiempo la compañía en general puede contar con información de negocios altamente eficiente, integral y consistente. Gradualmente permite fusionar aplicaciones de BI muy diversas para transformarlas en un activo homogéneo y confiable para la compañía. Presenta un diseño en web amigable e integridad con aplicaciones SAP BW metadatos, y la mejora de la presentación de informes para incorporar a las bases de datos heterogéneas. (MicroStrategy, 2015)

2.7.5 BUSINESS OBJECTS

Herramienta que elabora cubos para la extracción de data de sistemas transaccionales propietario. Su lenguaje de programación llamado ABAP, es el motor de una serie de herramientas para la dirección estratégica de empresas propietario de SAP. Las soluciones SAP para el almacenamiento de datos proporcionan una base sólida para los datos que le permitirá recopilar, almacenar, transformar y gestionar los datos en un almacén de datos empresariales escalable o en un centro de datos flexible y ágil para la eficiente explotación de datos. Ofrece amplia gama de servicios y soluciones para las distintas versiones de SAP existente desde SAP R/3 hasta SAP Business One. (Business Intelligence Consulting, 2015)

2.7.6 PENTAHO COMMUNITY EDITION

Pentaho es una plataforma de BI "orientada a la solución" y "centrada en procesos" que incluye todos los principales componentes requeridos para implementar soluciones basados en procesos. Las soluciones que Pentaho pretende ofrecer se componen de un entorno, al que podrá acceder vía web, de forma segura. Cada usuario visualiza todos los elementos habilitados para su

perfil, que incluirán informes, análisis OLAP y cuadros de mando con indicadores y tablas. Usted podrá generar nuevas vistas de análisis y nuevos informes y guardarlos asociados a su perfil, para consultarlos más tarde, así como exportarlos a Excel, PDF o realizar su impresión en papel. (OpenRed, 2015)

CAPITULO III

ESTADO DEL ARTE

En este capítulo se verá todo lo relacionado al estado del arte, primero se cuenta con un cuadro que resume las diferentes literaturas revisadas, así como también una breve descripción de cada una de las propuestas encontradas, un análisis comparativo de las mismas y finalmente las técnicas implementadas para las diferentes soluciones encontradas con su cuadro comparativo respectivo (Ver tabla 1).

Título del Paper	Autores
Implementación de un Data Mart como una solución de Inteligencia de Negocios para el área de logística de T-Impulso.	Julio Yalan Castillo, Luis Palomino Paniora (2012)
Análisis, diseño, construcción e implementación de un data warehouse para toma de decisiones y construcción de los KPI, para la empresa KRONOSCONSULTING CIA LTDA.	Ximena Lozada Peñafiel, Holger Cruz Tamayo, Washington Pérez Argudo, Andrés De La Torre Díaz (2014)
Análisis, diseño e implementación de una solución business intelligence para la generación de indicadores y control de desempeño, en la empresa Otecel S.A. utilizando la metodología Hefesto v2.0.	Anabel Bustos Barrera, Verónica Mosquera Artieda, Ing. Mauricio Campaña, Ing. Lorena Duque (2013)
Diseño de un Almacén de datos basado en Data Warehouse Engineering Process (DWEPE) y HEFESTO.	Castelán García Leopoldo, Ocharán Hernández Jorge Octavio (2012)
Implementación de un Data Mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el departamento de finanzas de la contraloría general de la república.	Alejandro Rojas Zaldívar (2014)
Análisis, diseño e implementación de un Data Mart utilizando herramientas open source para la unidad administrativa y financiera de la ESPE.	Alvarez Calvopiña Cristian Dario, Esparza Montes Diego Armando (2014)
Propuesta de Data Mart para el análisis de las consecuencias de la dinámica empresarial.	Campanaro Rosa Ingrassia Ramiro (2012)
Building an Effective Data Warehouse Architecture	James Serra (2013)
La metodología de Kimball para el diseño de almacenes de datos.	Gustavo R. Rivadera (2010)

Tabla 1 - Investigaciones

3.1 METODOLOGÍAS PARA EL DESARROLLO DE SOLUCIONES

BI

3.1.1 LA METODOLOGÍA DE KIMBALL PARA EL DISEÑO DE ALMACENES DE DATOS.

(Rivadera, 2010) Rivadera nos dice que la metodología más acorde a los negocios de nuestra región es la de Kimball, por cuanto proporciona un enfoque de menor a mayor, muy versátil, y una serie de herramientas prácticas que ayudan a la implementación de un DW. Es acorde a nuestras empresas porque se pueden implementar pequeños Data Marts en áreas específicas de las mismas (compras, ventas, etc.), con pocos recursos, integrándolos en un gran almacén de datos.

A su vez nos muestra que el ciclo de vida de un proyecto de Data Warehouse, está basado en cuatro principios básicos:

1. *Centrarse en el negocio:* Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
2. *Construir una infraestructura de información adecuada:* Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
3. *Realizar entregas en incrementos significativos:* crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.
4. *Ofrecer la solución completa:* proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

3.1.2 DISEÑO DE UN ALMACÉN DE DATOS BASADO EN DATA WAREHOUSE ENGINEERING PROCESS (DWEPE) Y HEFESTO.

(Castelán García & Ocharán Hernández, 2012) Castelán García & Ocharán Hernández presentan una aproximación de modelo de desarrollo que se realiza mediante la integración de HEFESTO y DWEPE.

Este modelo de desarrollo consta de tres niveles en los cuales se integra cada una de las actividades a realizar y los artefactos por cada nivel. La propuesta se centra en los pasos del nivel conceptual debido al tiempo que requiere para el análisis de integración de los artefactos generados por ambas metodologías para la definición del modelo lógico y físico que no ha sido posible definir de una manera clara.

El uso de este modelo posibilita que el desarrollo de un AD esté bien estructurado, permitiendo considerar varios aspectos muy importantes en el desarrollo de este tipo de sistemas, como son la recopilación de requisitos de información y las OLTP, los procesos ETL, el AD como estructura física y la explotación de la información. En la figura 1, muestra la arquitectura que debe tener un almacén de datos.

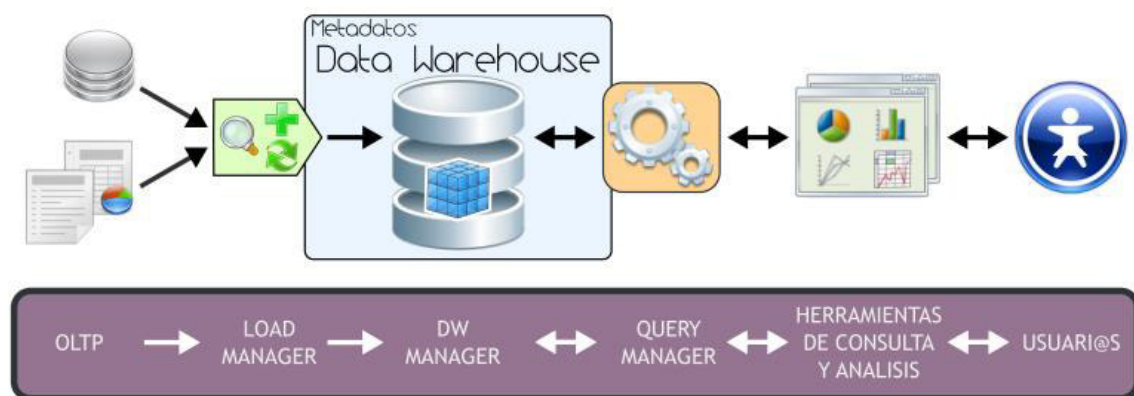


Figura 1 - Arquitectura de un Almacén de Datos

La necesidad que existe en las empresas por compartir y obtener un conocimiento sobre el mundo de información que guardan para poder realizar acciones para la toma de decisiones, hace posible el desarrollo de AD que sean guiados por procesos bien definidos y sustentados en metodologías que hacen uso de estándares como UML para el modelado, que garantiza que el diseño de cada elemento sea definido de forma única y poder ofrecer sistemas de calidad basados en una ingeniería de software que esté sustentada en la elaboración de

artefactos y seguimiento de procesos que sean claros que permiten entender la realización dicha actividad al momento de su diseño.

La figura 2 muestra las fases de la metodología DEWP.

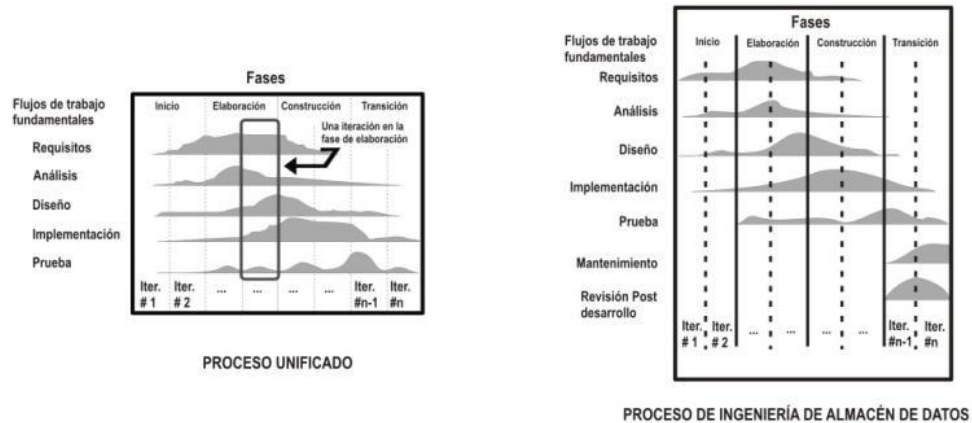


Figura 2 - El proceso unificado y proceso de ingeniería de almacenes de datos

3.1.3 BUILDING AN EFFECTIVE DATA WAREHOUSE ARCHITECTURE.

(Serra, 2013) Serra nos dice que la metodología Inmon no se cierra a un solo modelo y no solo eso, además su arquitectura mejora la trazabilidad decisional. Con ella podemos desgranar un valor en un KPI hasta una serie de análisis y reports que lo expliquen en detalle. Inmon es un modelo que mira a largo plazo y para una metodología ágil el largo plazo es secundario. Además que explica el modelo de Bill Inmon, el cual mostramos en la Figura 3 y que el autor lo caracteriza como:

- Al modelo de datos de la empresa, lo vuelve un data warehouse empresarial.
- Los usuarios tienen una participación pasiva.
- Las tablas normalizadas, son centralizadas y atómicas.
- Al utilizar Bill Inmon, se crean posteriormente mercados de datos dependientes, que se pueden utilizar para múltiples propósitos.

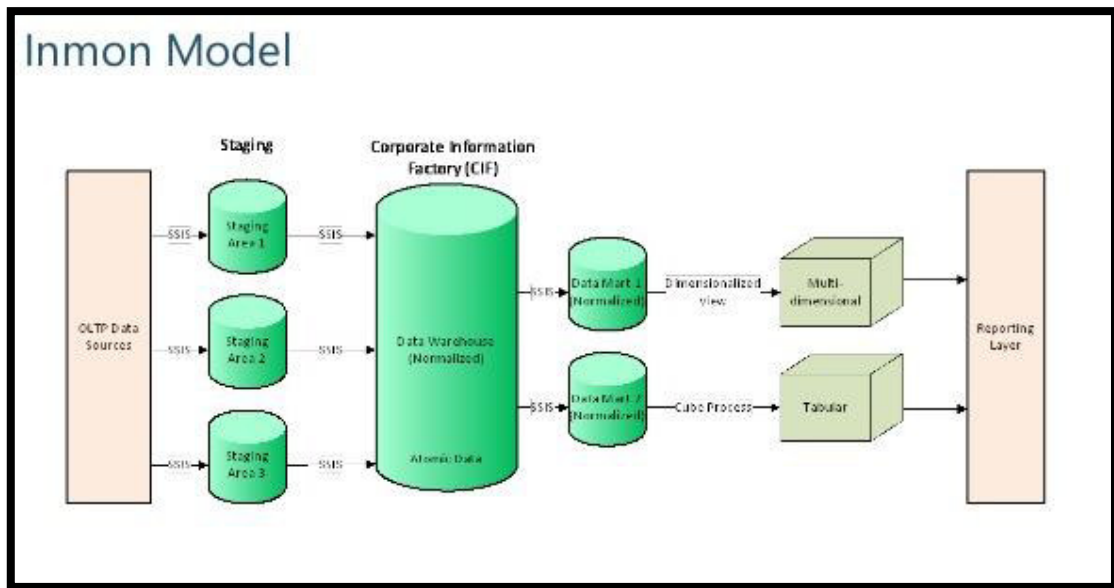


Figura 3 - Modelo Bill Inmon

3.2 USO DE DATA MARTS COMO SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS

3.2.1 IMPLEMENTACIÓN DE UN DATA MART COMO UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE LOGÍSTICA.

(Castillo, 2012) Castillo nos indica que mediante procesos de extracción; transformación y carga de data histórica, de forma automática se obtuvo un repositorio (Data Mart) que permitirá la explotación eficiente de la información.

El Data Mart permitió apoyar al área de logística en la toma de decisiones, a través de la entrega oportuna y relevante de información.

La implementación de un Data Mart redujo el tiempo en la elaboración de los reportes tanto al área de logística como al área de Informática, no se necesita tener demasiado conocimiento de los datos almacenados.

Como indica Castillo: El Data Mart se constituyó en una herramienta que nos ayuda a la elaboración de reportes y con ello a la toma de decisiones sobre el área de logística de la empresa T-Impulso.

3.2.2 PROPUESTA DE DATA MART PARA EL ANÁLISIS DE LAS CONSECUENCIAS DE LA DINÁMICA EMPRESARIAL.

(Campanaro & Ingressia, 2012) Los autores logran consensuar en la estructura y modelado de un Data Mart, a través de la estrategia Bottom up en contraposición a la Top-down, para obviar los inconvenientes propios de este último y obtener resultados más rápido y poder aprender y volcarlo en el diseño de los Data Marts por desarrollar.

A continuación, describen el procedimiento para el diseño lógico del Data Mart de acuerdo a las etapas identificadas en la metodología Hefestos.

1) *Análisis de Requerimientos.*

a. Identificar preguntas: Se tuvo en cuenta los aspectos que se querían conocer de los Jefes de Hogar para luego integrarlo con los indicadores de las GRI (Global Reporting Initiative).

b. Identificación de Indicadores y Perspectivas del análisis: Cantidad de Jefes de Hogar discriminados por Barrio, Sexo, Edad, Ocupación, Nivel de Estudios: al Inicio de su actividad laboral y Actual.

2) *Análisis de los OLTP*

a. Establecer correspondencia con los requerimientos: como se vio al describir al Data Warehouse, los datos que lo alimentan pueden provenir de distintas fuentes. En este caso, no proviene de un OLTP (Proceso de Transacciones en Línea) sino de una Planilla de Cálculo.

b. Seleccionar los criterios que integrarán cada perspectiva. Nivel de granularidad: Se entiende por granularidad, el nivel de detalle que tendrán las perspectivas: Barrios, Jefes de Hogar y Ramas de Actividad.

3) *Elaboración del Modelo Lógico de la Estructura del Data Warehouse*

A través de un modelo dimensional.

4) *Procesos ETL, Limpieza de Datos y Sentencias SQL.*

Terminado el Modelo deberán cargarse los datos. A este proceso se lo conoce con el nombre ETL (Extracción, Transformación y Carga):

Extracción: determinar los datos que se extraerán de las distintas fuentes, cada una de ellas requerirá su análisis particular.

Los datos que se extraen se colocan en un archivo intermedio, con un formato especial para ser utilizado en el proceso siguiente. Los registros rechazados deben guardarse en un archivo o log de rechazos para luego analizarlos y poder cargarlos al data warehouse (Ej. datos duplicados, datos vacíos, mal codificados, etc.).

Transformación: como su nombre indica consiste en transformar los datos extraídos al formato que requiere el data warehouse. Esto puede consistir en separar un campo fecha en día, mes y año, combinar campos en un único código, unificar codificaciones, tratamiento a dar a datos faltantes, etc.

Carga: consiste en tomar los datos extraídos y transformados y volcarlos al data warehouse.

3.2.3 ANÁLISIS COMPARATIVO.

	DATAWAREHOUSE	DAMART
Alcance	Construido para satisfacer las necesidades de información de toda la organización.	Construido para satisfacer las necesidades de un área de negocio específica.
Objetivo	Diseñado para optimizar la integración y la administración de los datos fuente.	Diseñado para optimizar la entrega de información de soporte de decisiones.
Características de los datos	Administra grandes cantidades de datos a nivel atómico.	Se concentra en administrar resúmenes y/o datos totalizados.
Pertenencia	Pertenece a toda la organización.	Pertenece al área de negocio al cual está orientado.
Administración	Es administrado por la unidad de sistema de la organización.	Es administrado por el personal de sistema de la unidad propietaria del Data Mart.

Tabla 2 - Análisis Comparativo I

Una vez realizada la comparación, se llega a la conclusión de que se debe realizar un Data Mart como diseño de almacén de datos (Ver tabla 2).

3.3 HERRAMIENTAS PARA EL DESARROLLO DE UNA SOLUCIÓN DE BUSINESS INTELLIGENCE

3.3.1 DATA WAREHOUSE PARA TOMA DE DECISIONES UTILIZANDO MICROSOFT SQL SERVER.

(Lozada Peñafiel, Cruz Tamayo, Perez Argudo, & De la Torre Diaz, 2014) Los autores muestran que SQL Server 2012 Edición Business Intelligence es una plataforma completa con la cual las organizaciones pueden crear y desplegar soluciones de BI seguras, escalables y manejables, además que fue seleccionada gracias al soporte que brinda Microsoft a todas sus herramientas (Ver figura 4).

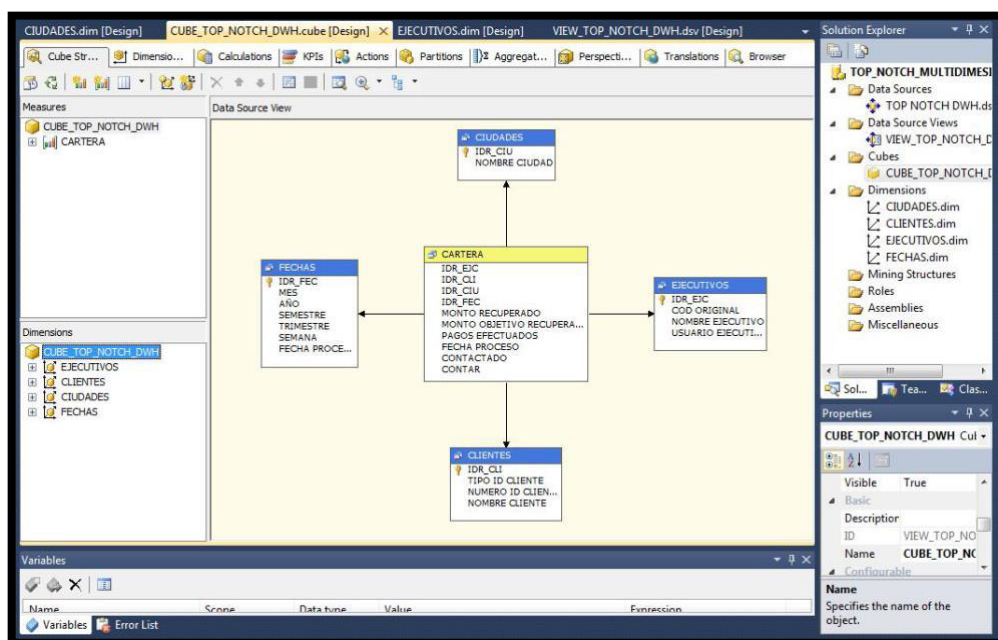


Figura 4 - Vista Toolbar de SQL Server

3.3.2 UNA SOLUCIÓN BUSINESS INTELLIGENCE PARA LA GENERACIÓN DE INDICADORES Y CONTROL DE DESEMPEÑO.

(Bustos Barrera, Mosquera Artieda, Campaña, & Duque, 2013) Los autores nos hablan que la herramienta de Pentaho Community (Data Integration), permite realizar la actualización de datos de manera sencilla, para que al momento de la obtención de los indicadores se obtenga resultados finales basándose en información al día.

También muestran una tabla con la descripción de sus productos (Ver tabla 3):

Producto	Nombre Código	Descripción
Pentaho Analysis Services	Mondrian	Servidor OLAP
Pentaho Reporting	Pentaho Reporting Engine	Generador de Reportes
Pentaho Data Mining	Weka	Minería de Datos
Pentaho Dashboard	Dashboard	Reporteador
Pentaho Data Integration	Kettle	Integración de Datos (ETL)

Tabla 3 - Tablas de Productos Pentaho

A través de esta herramienta realizan los procesos ETL, para elaborar un Data Source que les permite crear un Dashboard donde mostraran toda la información correspondiente de la empresa. Como se muestra en la figura 5.



Figura 5 - Reporte Requerimientos por Ambiente

3.3.3 SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS, BAJO LA METODOLOGÍA DE RALPH KIMBALL PARA OPTIMIZAR LA TOMA DE DECISIONES EN EL DEPARTAMENTO DE FINANZAS.

(Rojas Zaldivar, 2014) Rojas nos muestran las fases de la metodología Ralph Kimball, dando un resumen de cada fase:

Planificación del proyecto:

La planificación busca identificar la definición y el alcance del proyecto de Data Warehouse, también justificaciones del negocio y evaluaciones de factibilidad.

Definición de los requerimientos del negocio:

Un factor determinante en el éxito de un proceso de Data Warehousing es la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes niveles de usuarios.

Modelo dimensional:

La creación de un modelo dimensional es un proceso dinámico e altamente iterativo.

El proceso de diseño comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz de requerimientos.

Diseño físico:

Se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y seteos específicos del ambiente de la base de datos.

Diseño e implementación del subsistema de ETL:

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de Data Warehouse.

Las principales subetapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga, se define como proceso de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del modelo físico acordado.

Ante ello, el autor utiliza Qlik View por tres razones:

- 1) *Explotación de datos:* La plataforma de Business Discovery QlikView permite a los usuarios explorar los datos, hacer descubrimientos y revelar datos que pueden emplearse para ayudarles a resolver problemas de negocio de una forma nueva.
- 2) *La tecnología central de qlikview:* El segundo factor clave que distingue a QlikView en el mercado es su tecnología. QlikView ya es conocido por ser pionero de Business Intelligence en memoria, lo que es importante porque permite obtener una enorme mejora del rendimiento frente a sus competidores. Ningún sistema de Business Intelligence del mercado es más eficaz y rápido, no sólo en las consultas, sino, sobre todo, en el desarrollo del proyecto.
- 3) *El modelo de adopción de business discovery:* El tercer diferenciador clave de QlikView es el modelo de adopción de Business Discovery en el cliente, que ayuda a las empresas a resolver problemas importantes de negocio, que el BI tradicional no puede resolver. También impulsa un modelo asociativo y muy dinámico, en donde los cuadros de mando departamentales mejoran el funcionamiento de cada unidad de negocio y crean valor.

CAPITULO IV

APORTE TEÓRICO

En el este capítulo se desarrolla la selección y justificación de las herramientas tecnológicas a utilizar a través de una evaluación comparativa, finamente se explica su adaptación y la metodología que se va a seguir.

4.1 SELECCIÓN DE LAS HERRAMIENTAS PARA EL DESARROLLO

Considerando la evaluación comparativa se selecciona y se justifica las herramientas que van a ser utilizadas para desarrollar la solución.

4.1.1 SELECCIÓN DE LA METODOLOGÍA PARA EL DESARROLLO DEL DATA MART PROPUESTO

Para realizar la evaluación de las herramientas para la implementación del Data Mart presentado en el marco teórico, se describen los siguientes criterios de evaluación definidos con su respectiva justificación según las prioridades consideradas en el presente trabajo de investigación.

Criterio 1: Adaptable sobre cualquier tecnología

Se requerirá que la herramienta pueda ser plasmada en cualquier tecnología de inteligencia de negocios.

Criterio 2: Flexibilidad

Se requerirá que la metodología sea flexible antes cualquier cambio que surja en la etapa de desarrollo.

Criterio 3: Comunicación con el cliente

Se requerirá que la metodología permita una constante comunicación con los clientes, para validar los entregables y el avance de la implementación.

Criterio 4: Tiempo en el análisis y diseño

Se requerirá que la metodología tenga un tiempo promedio aceptable para el análisis y el diseño.

Criterio 5: Tiempo en construcción

Es necesario que la metodología a elegir, tenga un tiempo de construcción promedio aceptable, por tiempos de entrega y desarrollo del mismo.

Criterio 6: Más usada en el mundo

Se requerirá que la metodología sea altamente usada en el mundo, para tener un soporte comunitario para cualquier consulta o interrogantes.

Criterio 7: Facilidad de seguimiento

Es necesario que la metodología a elegir permita hacer un seguimiento constante del desarrollo del Data Mart.

Factores de análisis, puntaje:

VALOR	CONDICION	PUNTAJE
Bajo	Cumple la condición mínimamente	1
Medio	Cumple la condición de manera mediana	2
Alto	Cumple la condición de manera completa	3

Tabla 4 - Puntajes Metodologías

Obteniendo los resultados siguientes:

Nº	Factor de Análisis	Ralph Kimball	Bill Inmom	Ricardo Bernabeu
1	Adaptable sobre cualquier tecnología	1	1	1
2	Flexibilidad	2	3	3
3	Comunicación con el cliente.	3	3	3
4	Tiempo en el análisis y diseño	1	2	2
5	Tiempo en construcción	1	2	3
6	Fácil entendimiento principiantes	0	0	1
7	Más usada en el mundo	1	3	3
8	Facilidad de seguimiento	2	1	3
	TOTAL	11	15	19

Tabla 5 - Cuadro de evaluación de las metodologías

La metodología a emplear en este trabajo será la de Ricardo Bernabeu (Hefesto), ya que la misma presenta la mayor puntuación en la evaluación realizada.

4.1.2 SELECCIÓN DE LA PLATAFORMA PARA EL DESARROLLO DEL DATA MART PROPUESTO

Para realizar la evaluación de las plataformas de implementación del Data Mart presentadas en el marco teórico, se describen los siguientes criterios de evaluación definidos con su respectiva justificación según las prioridades consideradas en el presente trabajo de investigación.

Criterio 1: Licencia

Es necesario que la herramienta tenga una completa disponibilidad para su entero uso durante el desarrollo del software, es decir contar con una herramienta libre u open source.

Criterio 2: Permite procesos ETL

Se requerirá que la herramienta permita procesos ETL para la extracción, transporte y carga de los datos.

Criterio 3: Búsqueda de información

Se requerirá que la herramienta disponga de acceso completo a la información de sus componentes

Criterio 4: Interfaz Gráfica de Usuario (GUI)

Se requerirá que la herramienta tenga habilitada una interfaz gráfica de usuario.

Criterio 5: Soporte a diversas bases de datos

Es necesario que la herramienta a elegir, tenga soporte a las distintas bases de datos (Oracle, Mysql, SqlServer, etc).

Criterio 6: Exportación de datos

Es necesario que la herramienta a elegir permita exportar los datos, ya sea a Excel, XML, archivos planos.

Las características presentes en las diferentes herramientas serán marcadas denotadas con un estado SI, mientras que las ausentes con uno, NO.

Obteniendo los resultados siguientes:

CARACTERÍSTICAS	BUSINESS OBJETCS	SQL SERVER	IBM DATASTAGE	QLIK VIEW	PENTAHO
Software Libre / Open Source	NO	NO	NO	SI	SI
Permite procesos ETL	SI	NO	SI	NO	SI
Soporte a diversas bases de datos	NO	SI	NO	SI	SI
Exportación de datos	SI	NO	SI	SI	SI
Interfaz Gráfica de Usuario (GUI)	SI	NO	SI	SI	SI
Búsqueda de información	SI	SI	SI	SI	SI

Tabla 6 - Cuadro de evaluación de plataformas de BI

La plataforma a emplear en este trabajo será la de Pentaho, ya que la misma presenta todas las características consideradas como relevantes.

4.1.3 SELECCIÓN DE LA PLATAFORMA PARA LA PRESENTACIÓN DEL DATA MART PROPUESTO

Para realizar la evaluación de las plataformas de implementación del Data Mart presentadas en el marco teórico, se describen los siguientes criterios de evaluación definidos con su respectiva justificación según las prioridades consideradas en el presente trabajo de investigación.

Criterio 1: Licencia

Es necesario que la herramienta tenga una completa disponibilidad para su entero uso durante el desarrollo del software, es decir contar con una herramienta libre u open source.

Criterio 2: Complejidad de la herramienta

Se requerirá que la herramienta sea de fácil instalación además de ser intuitiva en su uso.

Criterio 3: Tiempo de respuesta

Se requerirá que la herramienta muestre los resultados en el menor tiempo posible.

Criterio 4: Integración con hojas de cálculo

Es necesario que la herramienta a elegir permita exportar los datos a un documento Excel.

Criterio 5: Soporte a diversas bases de datos

Es necesario que la herramienta a elegir, tenga soporte a las distintas bases de datos (Oracle, Mysql, SqlServer, etc).

Criterio 6: Cuadros de mandos analíticos

Se requerirá que la herramienta permita la creación de Dashboard elegantes y sofisticados para mostrar la información relevante fruto de la implementación del Data Mart.

Factores de análisis, puntaje:

VALOR	PUNTAJE
Muy malo	1
Malo	2
Regular	3
Bueno	4
Muy Bueno	5

Tabla 7 - Factor de Análisis

Obteniendo los resultados siguientes:

CARACTERÍSTICAS	BUSINESS OBJETCS	SQL SERVER	IBM DATASTAGE	PENTAHO	QLIK VIEW
Licencia	2	2	2	4	4
Complejidad de la herramienta	3	4	3	4	4
Tiempo de respuesta	3	3	3	3	5
Integración con hojas de cálculo	5	5	5	4	5
Soporte a diversas bases de datos	4	5	5	4	5
Cuadros de mandos analíticos	3	4	4	4	5
TOTAL	20	23	22	23	28

Tabla 8 - Cuadro de evaluación de plataformas de BI

La plataforma a emplear en este trabajo será la de QlikView, ya que la misma presenta la mayor puntuación en la evaluación realizada.

4.2 DEFINICIÓN DE LA SOLUCIÓN

En éste apartado se procederá a adaptar la plataforma seleccionada para el desarrollo del Data Mart para el Área Contable de la UNMSM y así dar solución al problema planteado.

Diseño de la solución

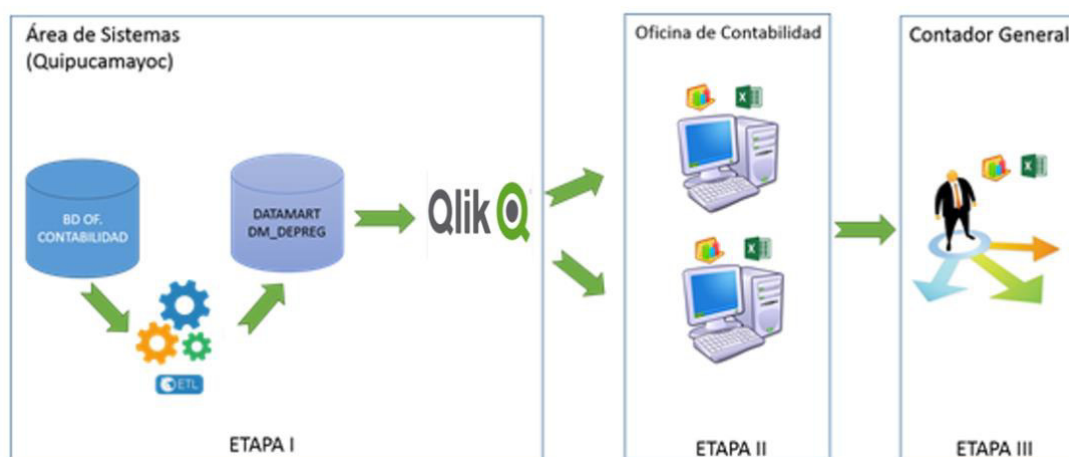


Figura 6 - Solución Propuesta

La figura 6 muestra la solución al problema planteado, se ha propuesto como una solución de tres etapas, en la Etapa I, se observa que los datos de la oficina de contabilidad ya se encuentran cargados en el Data Mart, eliminando

de esta manera la carga de trabajo del personal del Área de Sistemas para la obtención de datos, que consistía en tareas manuales de consultas extracción desde las bases de datos transaccionales y el tiempo empleado para pasar los datos obtenidos en archivos de formato Excel para el uso de los analistas de la Oficina de Contabilidad. Luego, en la Etapa II, se observa como ahora los analistas consultan directamente el Data Mart a través de la aplicación Pentaho que muestra reportes predefinidos que se pueden exportar a formato Excel. Así, se logra que los analistas puedan elaborar reportes personalizados con la información que realmente necesitan y en base a los criterios que consideran importantes, concentrando la mayor parte de su tiempo en realizar un adecuado análisis de la información y finalmente elaborar sus conclusiones y enviárselas al Contador General a través de los reportes solicitados para la toma de decisiones, y consecuentemente en la Etapa III, se logrará realizar la entrega de éstos en los tiempos estimados conteniendo información útil, con la cual se pueda decidir estrategias de negocio correctas.

4.3 METODOLOGÍA A SEGUIR

Para el desarrollo del Data Mart de la Oficina de Contabilidad se usará como base la metodología HEFESTO, con las siguientes fases (Dario, 2010):

1. Análisis de Requerimientos

Lo primero que se hará será identificar los requerimientos de los usuarios a través de preguntas que expliciten los objetivos de su organización. Luego, se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso

1.1. Identificar Preguntas

El primer paso comienza con el acopio de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

1.2. Identificar indicadores y perspectivas de análisis.

Una vez que se han establecido las preguntas de negocio, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

1.3. *Modelo Conceptual*

En esta etapa, se construirá un modelo conceptual¹ a partir de los indicadores y perspectivas obtenidas en el paso anterior. A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

1.4. *Diseño de la Arquitectura*

Se realizará la arquitectura del Data Mart que nos permitirá elaborar la solución de inteligencia de negocios.

2. Análisis del Procesamiento de Transacciones En Línea (OLTP)

Seguidamente, se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.

2.1. *Determinación de indicadores*

En este paso se deberán explicitar cómo se calcularán los indicadores. Que tipos de indicadores deseamos y la forma de obtener sus entradas.

2.2. *Establecer correspondencias.*

El objetivo de este paso, es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. La idea es, que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

2.3. *Nivel de granularidad*

Una vez que se han establecido las relaciones con los OLTP, se deben seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores.

2.4. *Modelo Conceptual ampliado.*

En este paso, y con el fin de graficar los resultados obtenidos en los pasos anteriores, se ampliará el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo.

3. **Modelo lógico del Data Mart**

A continuación, se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definirá el tipo de modelo que se utilizará y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizarán las uniones pertinentes entre estas tablas.

3.1. *Tipo de modelo lógico del DM*

Se debe seleccionar cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades de los usuarios. Es muy importante definir objetivamente si se empleará un esquema en estrella, constelación o copo de nieve, ya que esta decisión afectará considerablemente la elaboración del modelo lógico.

3.2. *Tablas de dimensiones*

En este paso se deben diseñar las tablas de dimensiones que formaran parte del DW. Para los tres tipos de esquemas, cada perspectiva definida en el modelo conceptual constituirá una tabla de dimensión. Para ello deberá tomarse cada perspectiva con sus campos relacionados y realizarse el siguiente proceso: Se elegirá un nombre que identifique la tabla de dimensión. Se añadirá un campo que represente su clave principal. Se redefinirán los nombres de los campos si es que no son lo suficientemente intuitivos.

3.3. *Tablas de hechos*

En este paso, se definirán las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

3.4. *Uniones*

Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

4. **Integración de datos**

Una vez construido el modelo lógico, se deberá proceder a poblarlo con datos, utilizando técnicas de limpieza, calidad de datos y procesos ETL.

4.1. *Carga Inicial*

Debemos realizar la Carga Inicial al DM, poblando el modelo de datos que hemos construido. Para lo cual debemos llevar adelante una serie de tareas, tales como limpieza de datos, calidad de datos y procesos ETL.

4.2. *Visualización (Creación de Dashboard)*

Vamos a realizar la construcción de los gráficos que se presentaran a los usuarios finales.

CAPITULO V

APORTE PRÁCTICO

5.1 ANÁLISIS DE REQUERIMIENTOS

Lo primero que se hará será identificar los requerimientos de los usuarios a través de preguntas que expliciten los objetivos de su organización.

5.1.1 IDENTIFICAR PREGUNTAS

Luego de reunirse y entrevistar a las personas encargadas del proceso de registro contable, se ha analizado los requerimientos y posteriormente se ha identificado las preguntas del negocio:

- a) ¿Cuál es el promedio de registros contables procesados de un determinado sectorista, de una determinada dependencia según un estado contable en un mes?
- b) ¿Cuál es el número de registros procesados realizados por un determinado sectorista en una unidad de tiempo de un estado inicial a un estado final?
- c) Total de registros contables procesados de los sectoristas de la oficina.
- d) ¿Cuál es la cantidad y el tiempo promedio de envío de registros contables de una dependencia hacia otra?

5.1.2 IDENTIFICAR INDICADORES Y PERSPECTIVAS DE ANÁLISIS.

Después de identificar las preguntas del negocio, en base a estas obtendremos los indicadores (valores numéricos) y las perspectivas (objetos):

- a) Promedio de registros contables procesados de un determinado sectorista, de una determinada dependencia según un estado contable en un mes

- b) Número de registros contables procesados realizados por un determinado sectorista en una unidad de tiempo de un estado inicial a un estado final
- c) Total de registros contables procesados de los sectoristas de la oficina.
- d) Tiempo promedio de envío de registros contables de una dependencia hacia otra

Los indicadores obtenidos en base a las preguntas son:

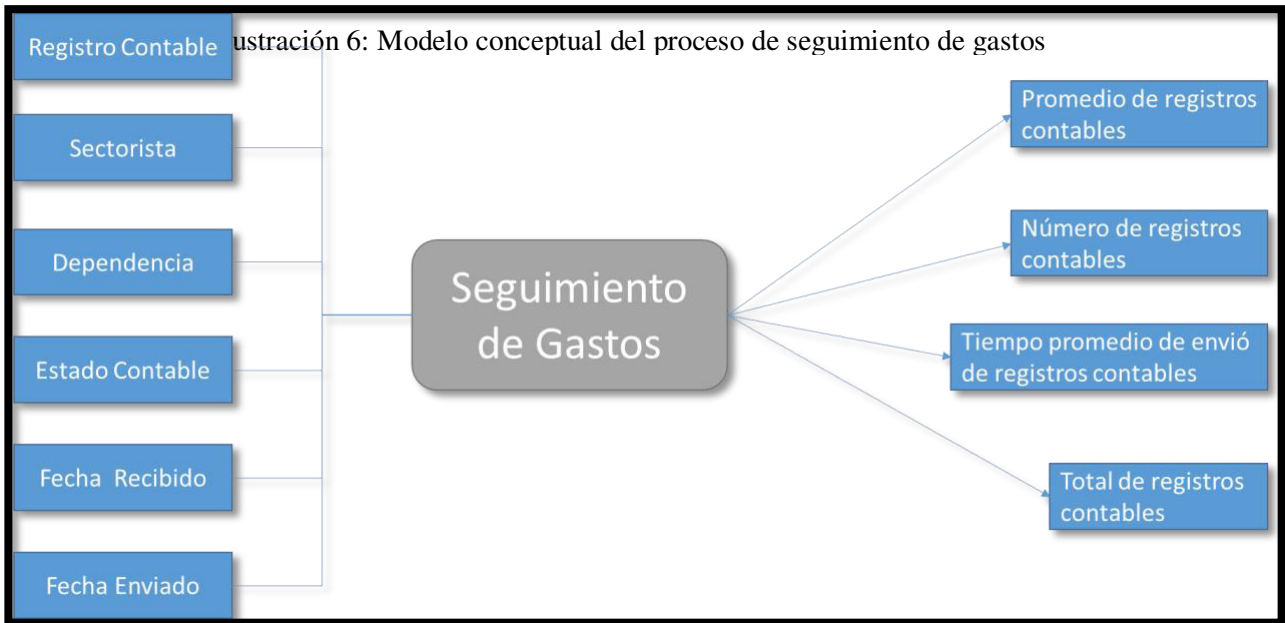
- a) Número de registros contables
- b) Promedio de registros contables
- c) Total de registros contables
- d) Tiempo promedio de envío de registros contables.

Las perspectivas que se deben considerar son

- a) Registro Contable
- b) Sectorista
- c) Dependencia
- d) Estado contable
- e) Fecha de recibido
- f) Fecha de enviado

5.1.3 MODELO CONCEPTUAL

En base a los indicadores y perspectivas, se elaborará un modelo conceptual, en el cual podremos observar con mayor claridad el alcance del proyecto en base al proceso de ejecución de gastos a través de registros contables. El modelo conceptual podemos verlo en la siguiente ilustración (Ver ilustración 6).

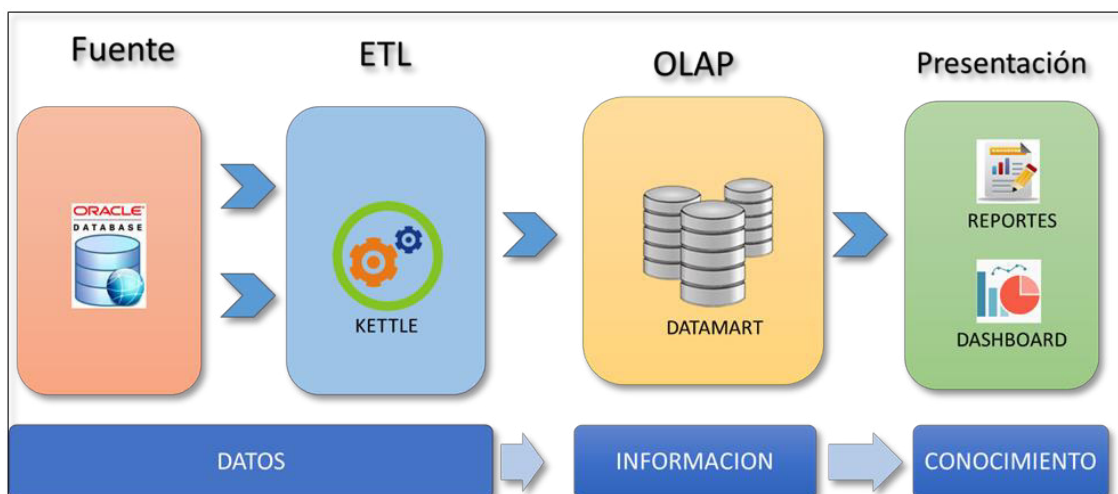


Fuente: Elaboración Propia

5.1.4 DISEÑO DE LA ARQUITECTURA

La arquitectura se divide en las tres fases de una solución de inteligencia de negocios, Datos de donde inicia todo proceso de BI que contiene la Fuente de Datos y los procesos ETL, Información donde se encuentra el procesamiento analítico en línea (OLAP) donde se encuentra el Data Mart a realizar y Conocimiento donde se presentan la información más relevante para los usuarios.

Figura 7 Arquitectura de la Solución



Fuente: Elaboración Propia

Se explicarán los módulos contenidos en la arquitectura presentada (Ver figura 7).

Vista Fuente

Los datos que se utilizarán para alimentar el Data Mart corresponden a la fuente de datos del Sistema Quipucamayoc. Este sistema tiene información relacionada a los procesos contables de la universidad que tiene una base de datos relacional en ORACLE 10G.

Vista ETL

Es la sección donde se agrupan una serie de procesos que llevan a cabo tareas relacionadas con la extracción, manipulación, control, integración, limpieza de datos, carga y actualización de los Data Marts, todas las tareas que se hagan desde que se toman los datos de la fuente en este caso la base relacional y archivos planos, hasta que se carguen en el sistema para su utilización en la construcción de los cubos o reportes.

Vista Procesamiento Analítico en Línea (OLAP)

Es el núcleo del sistema que es el repositorio central de la información donde se almacenan los datos que se están utilizando. En el Data Mart se almacenan los datos operacionales en estructuras multidimensionales que optimizan el acceso para las consultas y que son flexibles, además de contener toda la metadata de la información almacenada que ofrece información descriptiva sobre el contexto, la calidad, condición y características de los datos. En esta área se incluye el motor de cubos multidimensional que es el encargado de ejecutar las consultas realizadas por los componentes externos.

Vista Presentación

Es el área correspondiente a la interacción con el usuario, cuya funcionalidad es mostrar los datos almacenados de forma útil y transparente a través de las distintas herramientas. Este sistema se comunica directamente con el servidor de cubos a través de consultas, las cuales retornan la información requerida donde ésta es transformada y presentada para la visualización final. Los reportes requeridos en el proyecto se encuentran en esta área.

5.2 ANÁLISIS DEL PROCESAMIENTO DE TRANSACCIONES EN LÍNEA

5.2.1 DETERMINACIÓN DE INDICADORES

En éste paso se explicará cómo se van a calcular los indicadores encontrados para el proceso de seguimiento de gastos:

- “Número de registros contables”
 - Hecho: Registros Contables
 - Función de conteo: Count
 - Aclaración: el indicador número de registros contables representa el conteo de la cantidad de registros contables que se han procesado.
- “Promedio de registro contables”
 - Hechos: Registros Contables
 - Función de conteo: AVG
 - Aclaración: indicador del promedio de registros contables procesados por Sectorista.
- “Total de registros contables”
 - Hecho: Registros Contables
 - Función de conteo: Count
 - Aclaración: indicador del total de registros contables procesados por un sectorita.
- “Tiempo promedio”
 - Hechos: Registros Contables
 - Función de conteo: AVG
 - Aclaración: indicador del promedio de tiempo que demora en enviar un registro de un estado a otro.

5.2.2 ESTABLECER CORRESPONDENCIAS.

Modelo Entidad Relación del esquema utilizado en la oficina de contabilidad. (Ver ilustración 8)

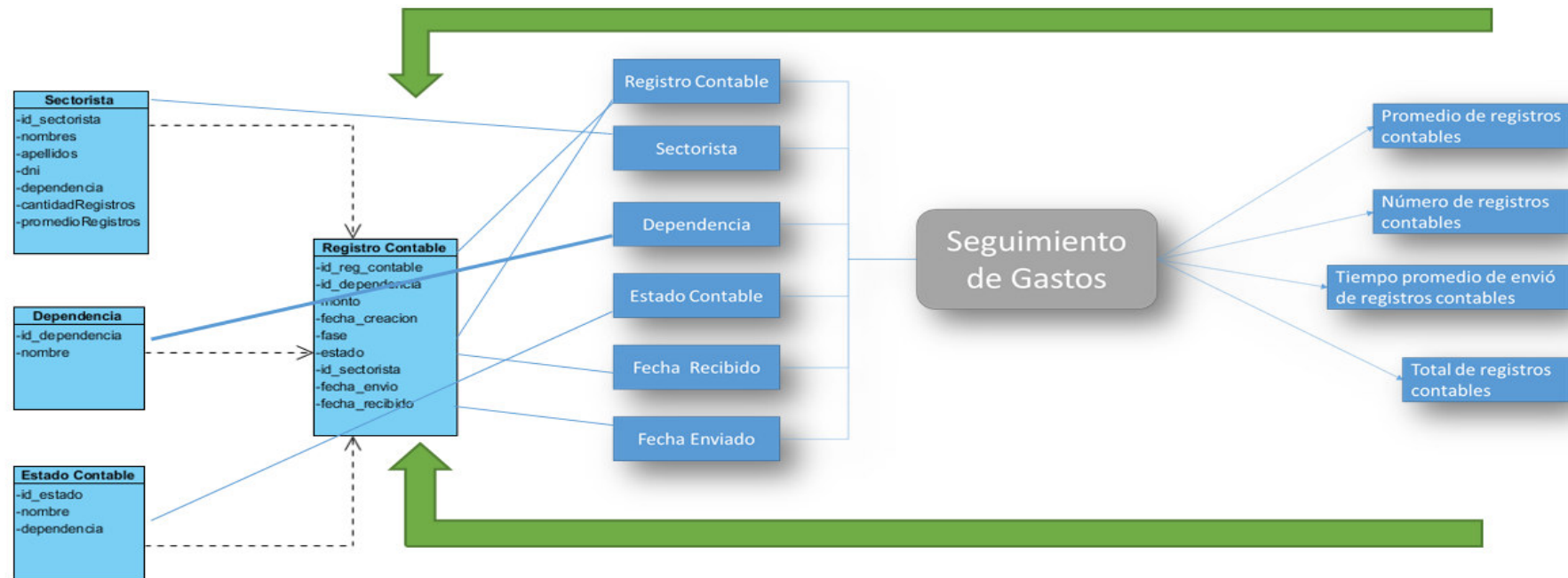
Ilustración 5 - MER Esquema Contabilidad



Fuente: Elaboración Propia

8) Se analizaron los OLTP para establecer las correspondencias entre el modelo conceptual y las fuentes de los datos: (Ver figura

Figura 8 Establece correspondencias

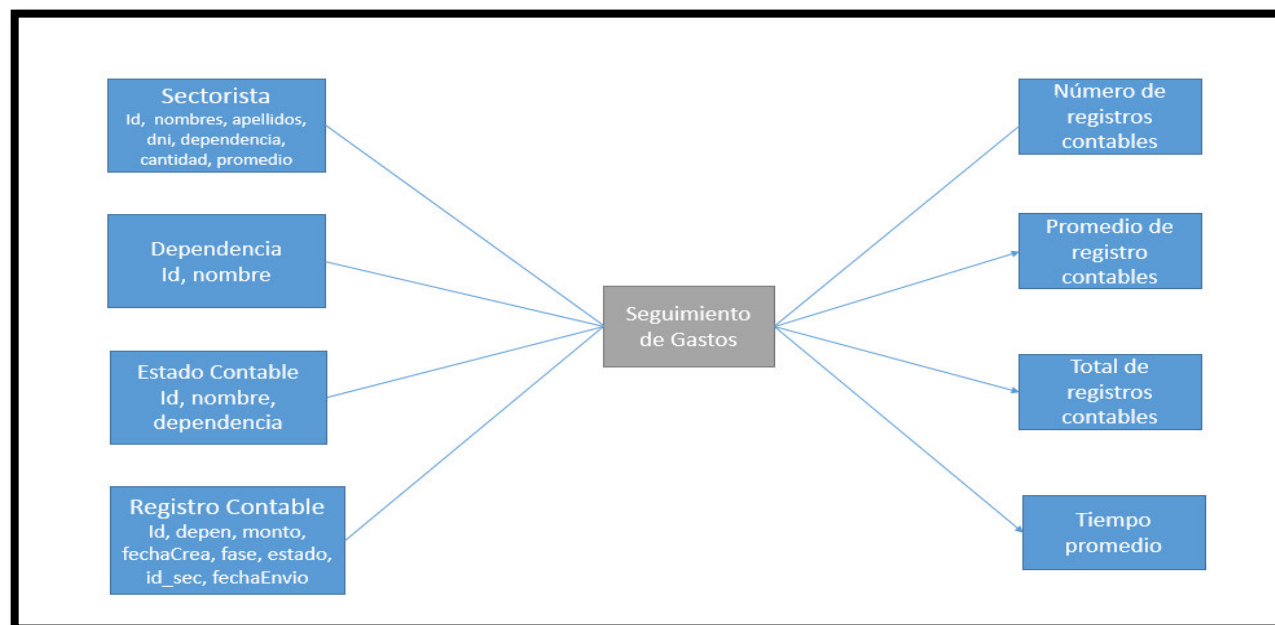


Fuente: Elaboración Propia

5.2.3 MODELO CONCEPTUAL AMPLIADO.

En éste paso se mostrará el modelo conceptual inicial, el cual ahora contendrá los campos seleccionados bajo cada perspectiva, y bajo cada indicador su fórmula de cálculo. (Ver figura 9)

Figura 9 - Modelo conceptual ampliado



Fuente: Elaboración Propia

5.3 MODELO LÓGICO DEL DATA MART

5.3.1 TIPO DE MODELO LÓGICO DEL DATA MART

Para éste trabajo se escogerá el modelo de DW tipo estrella, por su simplicidad.

5.3.2 TABLAS DE DIMENSIONES

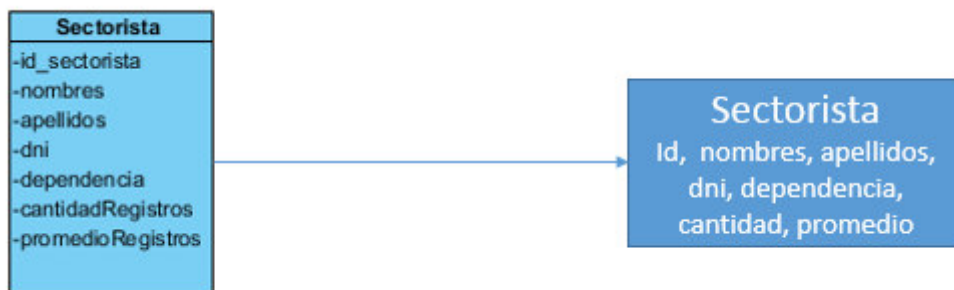
Diseñaremos las tablas de dimisiones, que van a formar parte del DW.

Perspectiva Sectorista

Esta dimensión tendrá los datos de un Sectorista (Ver figura 10)

- Se agrega el campo id_sectorista
- Se mantiene el campo nombre
- Se mantiene el campo apellidos
- Se mantiene el campo dni
- Se agrega el campo dependencia

Figura 10 - Tabla de dimensión Sectorista



Fuente: Elaboración Propia

Perspectiva Dependencia

Ésta dimensión contiene los datos de una Dependencia (Ver figura 11)

- Se agrega el campo id_dependencia
- Se mantiene el campo nombre

Figura 11 - Tabla de dimensión Dependencia



Fuente: Elaboración Propia

Perspectiva Estado Contable

Esta dimensión contiene la información de un Estado Contable (Ver figura 12)

- Se agrega el campo id_estado
- Se mantiene el campo nombre
- Se agrega el campo dependencia

Figura 12 - Tabla de dimensión Trabajador



Fuente: Elaboración Propia

5.3.3 TABLAS DE HECHOS

En ésta parte se va a definir la tabla de hechos, que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

- La tabla hechos tendrá el nombre Registro Contable
- Se agrega el campo id_reg_contable
- Se agrega el campo id_dependencia
- Se agrega el campo monto
- Se agrega el campo fecha_creacion
- Se agrega el campo fase
- Se agrega el campo estado
- Se agrega el campo id_sectorista
- Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: id_dependencia, id_sectorista, id_tiempo.
- Se crea la tabla de hechos Registro Contable con lo anteriormente mencionado (Ver figura 13):

Figura 13 - Tabla de Hechos Registro Contable

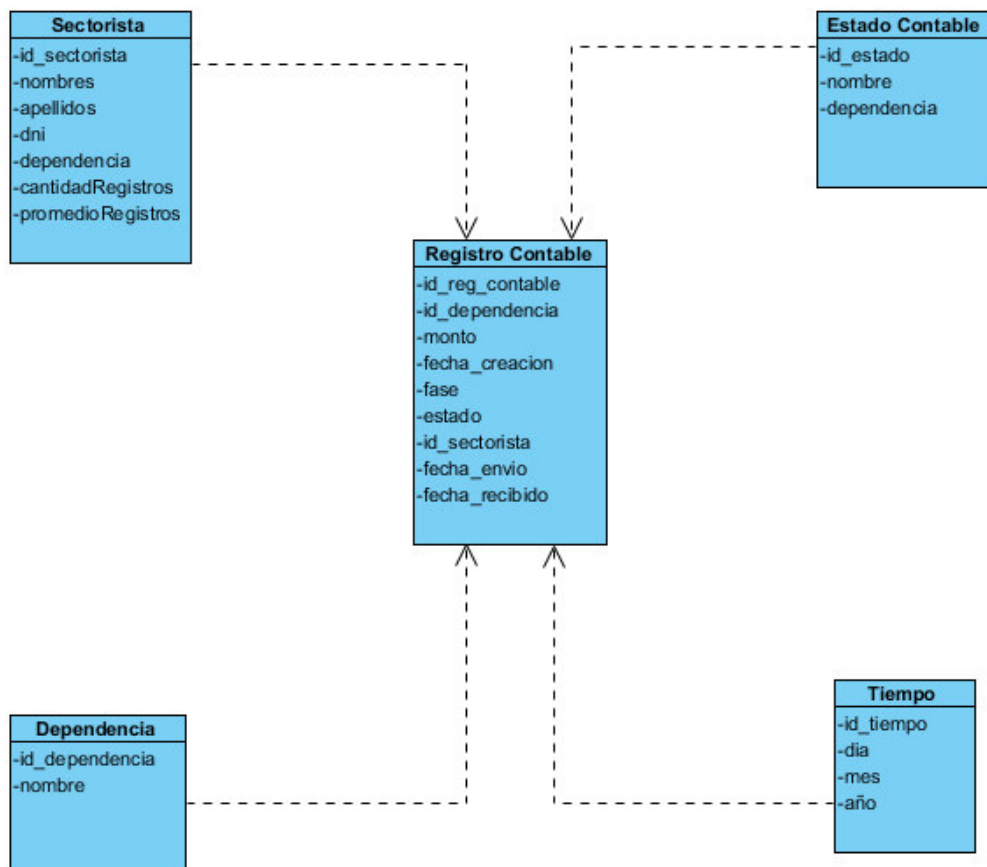
Registro Contable
-id_reg_contable
-id_dependencia
-monto
-fecha_creacion
-fase
-estado
-id_sectorista
-fecha_envio
-fecha_recibido

Fuente: Elaboración Propia

5.3.4 UNIONES

Ahora se realiza la unión entre las tablas de dimensiones y la tabla de hechos (Ver figura 14):

Figura 14 - Modelo Dimensional



Fuente: Elaboración Propia

5.4 INTEGRACIÓN DE DATOS

Una vez construido el modelo lógico, se deberá proceder a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc.; luego se definirán las reglas y políticas para su respectiva actualización, así como también los procesos que la llevarán a cabo.

Estándar de ETL

Para la creación de ETL el estándar que se llevara para asignar el nombre a los mismos es el siguiente:

- **Transformación:** Sera el nombre que se dará, para indicar que se trata de un proceso de ETL o transformación.
- **Nombre Proceso:** Ira el nombre del proceso al cual se aplicara los las distintas transformaciones.

ETL (Extracción, Transformación y Carga)

Lista de Procesos

Los tipos de entrada de los cuales se hará la carga de información serán directamente de la base de datos.

Fuentes de Entrada:

Las fuentes de entrada, son herramientas que permiten indicar que tipo de fuente contendrán los datos a ser transformados.

Valor Java Script Modificado:

Sirve para agregar validación o transformación a los campos, el nombre que se asignara a los objetos de este tipo será:

JS_: Representa un objeto Java Script.

Nombre Validador:

Representa el nombre del tipo de transformación que se está realizando.

Formato: JS_ Nombre Validador

Calculadora:

Esta herramienta, se utiliza para realizar transformaciones de campos, pudiendo hacer uso de las funciones propias de esta herramienta que ayudan a realizar una discriminación en los campos o llegar a un nivel de detalle más específico, en el caso que se requiera.

C_: Indica que se trata de la herramienta Calculadora.

Nombre Cálculos:

Indica que realizara se realizara con la herramienta calculadora.

Formato: C_ Nombre Cálculos

Almacén de Datos:

Esta herramienta, se utiliza para realizar la carga de los datos, de acuerdo a los campos a los que se les asociaran a cada Dimensión, los cuales se insertaran en la Base de Datos que alojaran, los resultados obtenidos del Proceso de transformación

Dim:

Es el indicador, para definir que se está realizando la carga de los datos a una Dimensión.

Nombre Dimensión: Indica el Nombre de la dimensión a la cual están asociados los datos cargados.

Formato: DimNombre Dimensión

5.4.1 CARGA INICIAL

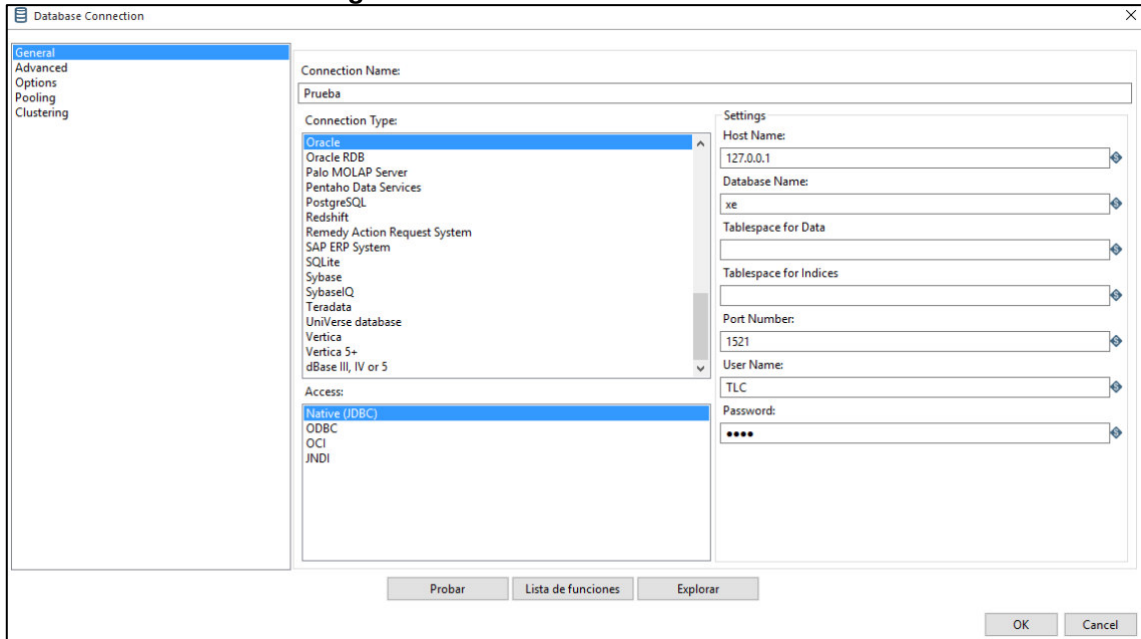
Se realiza la carga inicial del archivo OLTP⁵, utilizando técnicas de limpieza y calidad de datos, procesos ETL, con la herramienta Pentaho.

Conexión a la base de datos.

Se ingresa los datos de acuerdo a la imagen inferior (Ver figura 15).

⁵ Procesamiento de Transacciones En Línea

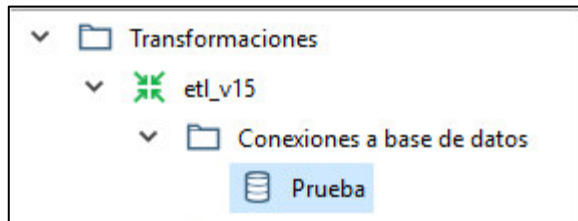
Figura 15 - Conexión a Base de Datos I



Fuente: Pentaho Data Integration - 2016

Al probar la conexión quedara de la siguiente manera (Ver figura 16):

Figura 16 - Conexión a Base de Datos II



Fuente: Pentaho Data Integration - 2016

Entrada de datos de las tablas.

Esta función permite obtener los datos de la tabla que se desea (Ver figura 17).

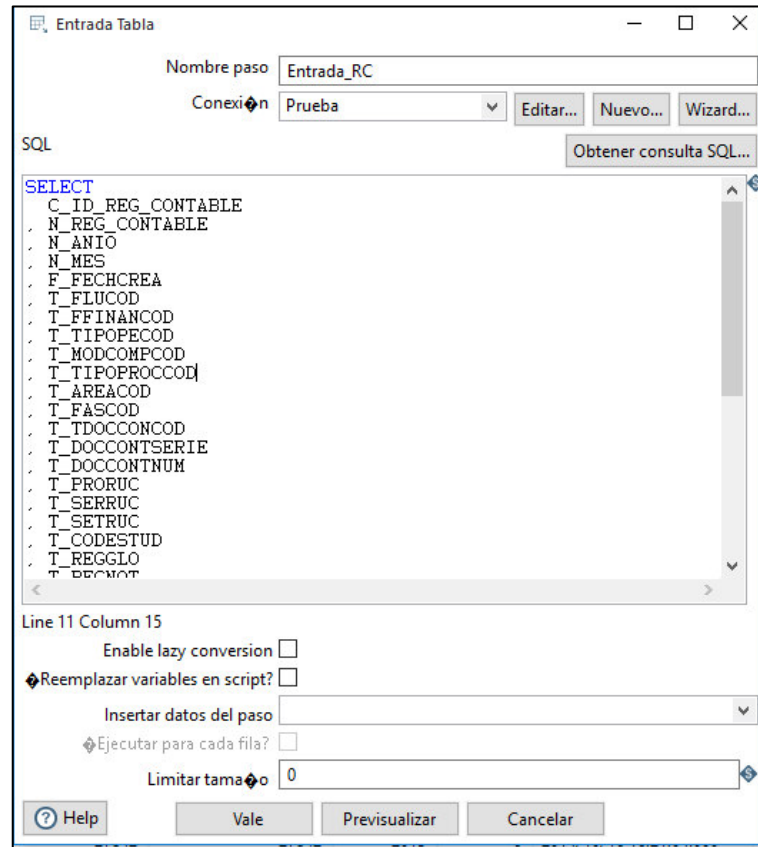
Figura 17 - Entrada de datos a tablas I



Fuente: Pentaho Data Integration - 2016

Se deben ingresar que tabla se desea obtener, haciendo doble clic (Ver figura 18):

Figura 18 - Entrada de datos II



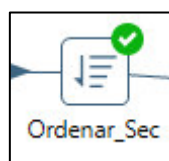
Fuente: Pentaho Data Integration - 2016

*Se realiza el mismo procedimiento para todas las tablas del OLTP de la base de datos de quipucamayoc.

Ordenar datos.

Esta función permite ordenar los datos de la tabla que se desea (Ver figura 19).

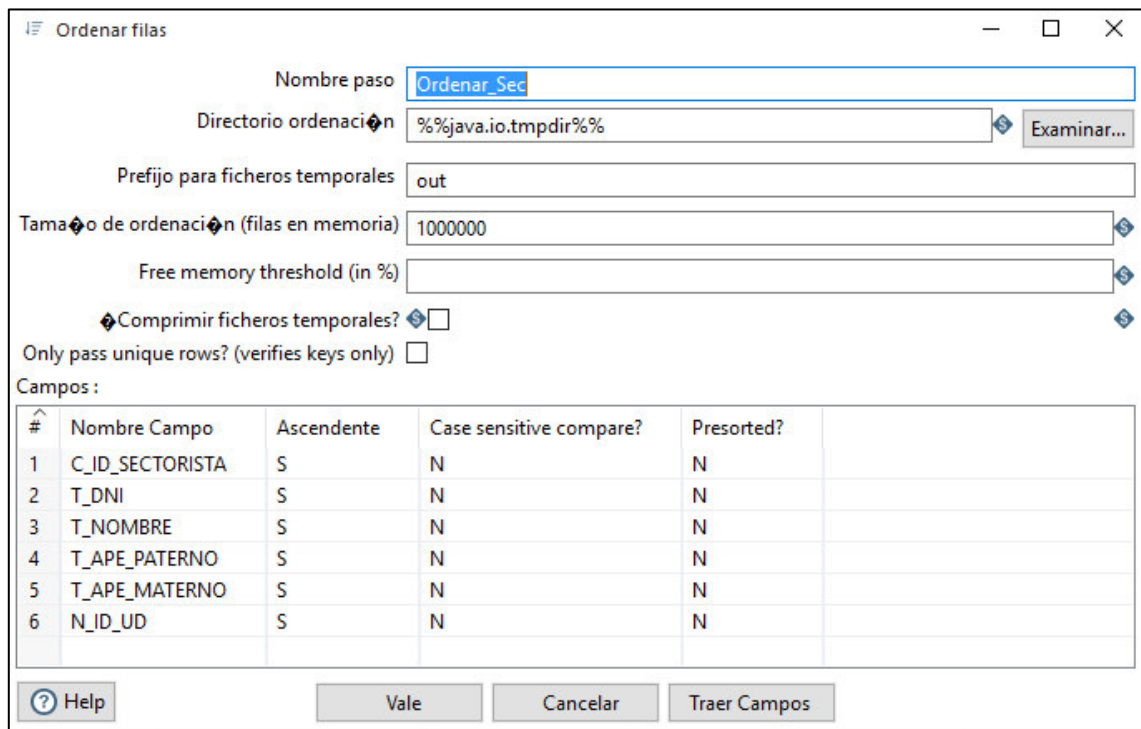
Figura 19 - Ordenar datos I



Fuente: Pentaho Data Integration - 2016

Se debe detallar que atributos serán los elegidos para ordenar (Ver figura 20):

Figura 20 - Ordenar datos II



Fuente: Pentaho Data Integration - 2016

Dar formato a datos.

Esta función permite dar un formato específico a los datos que se ingresaran en el Data Mart (Ver figura 21).

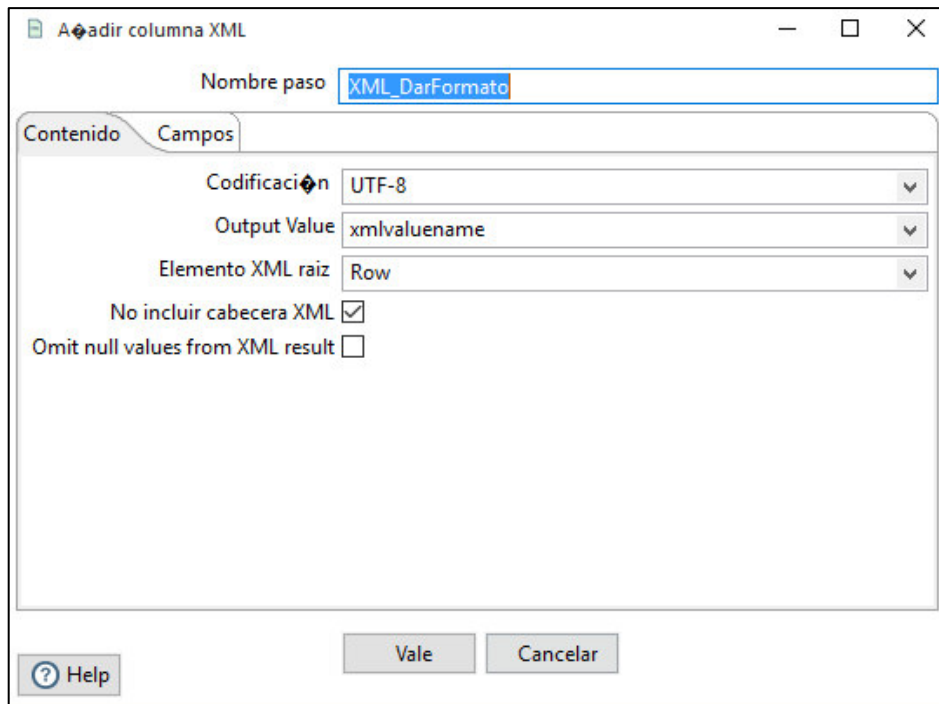
Figura 21 - Dar formato a datos I



Fuente: Pentaho Data Integration - 2016

En la pestaña contenido llenar de la siguiente manera (Ver figura 22):

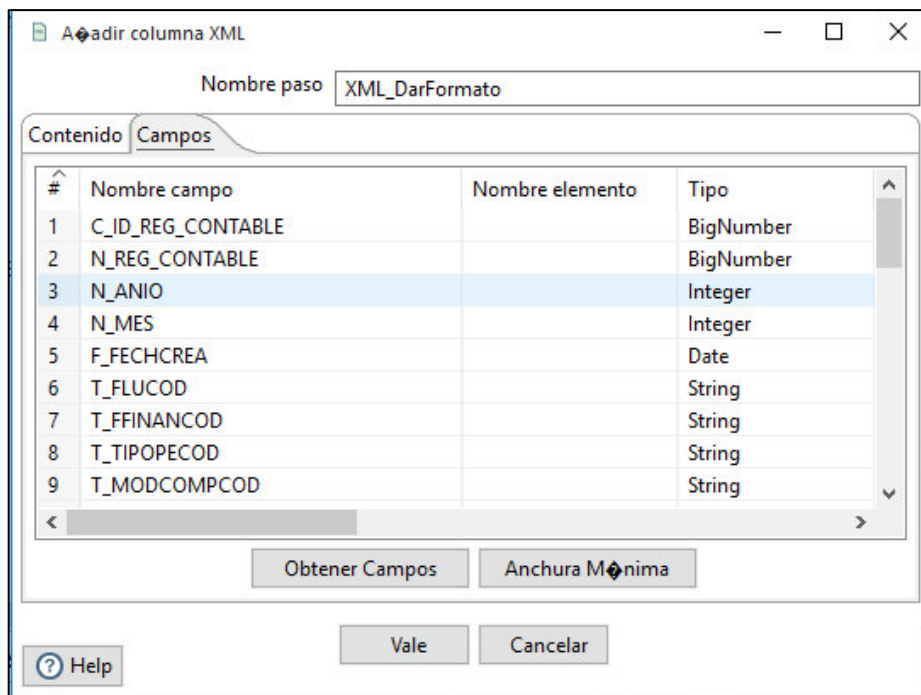
Figura 22 - Dar formato a datos II



Fuente: Pentaho Data Integration - 2016

En la pestaña campos selección el tipo de dato que serán los atributos (Ver figura 23):

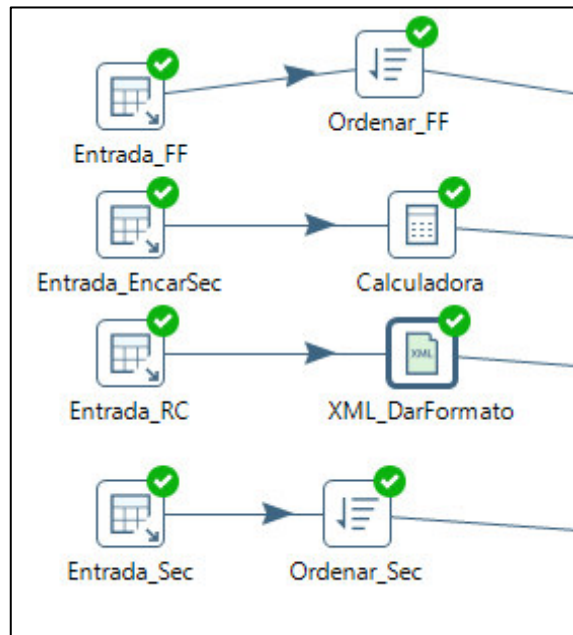
Figura 23 - Datos formatos a datos III



Fuente: Pentaho Data Integration - 2016

Ahora vamos conectando las tareas para ir probando si correctamente está funcionando nuestro flujo de trabajo (Ver figura 24):

Figura 24 - Conexión de tareas



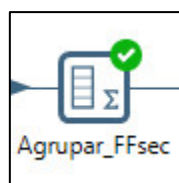
Fuente: Pentaho Data Integration - 2016

El check denota que la tarea se realizó con éxito.

Agrupar datos.

Esta función permite agrupar los datos (Ver figura 25).

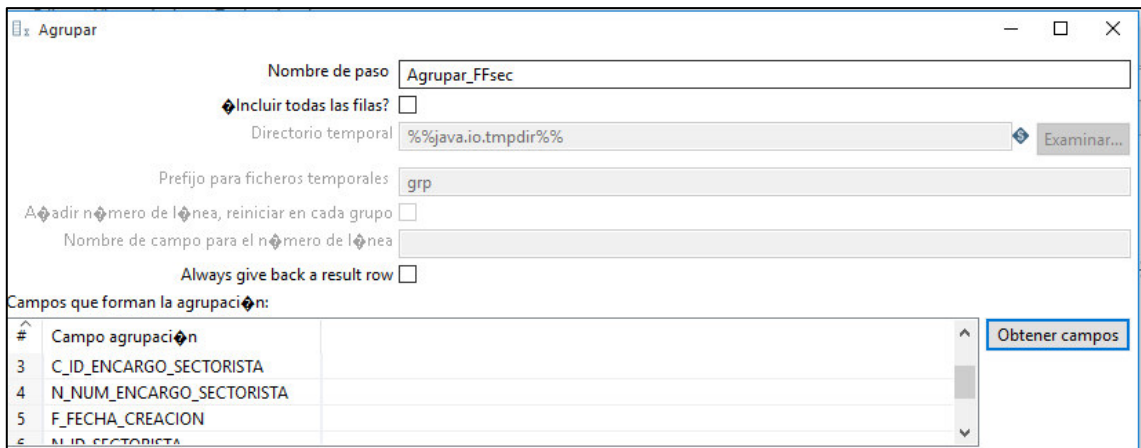
Figura 25 - Agrupar datos I



Fuente: Pentaho Data Integration - 2016

Completar los datos de la siguiente manera (Ver figura 26):

Figura 26 - Agrupar datos II

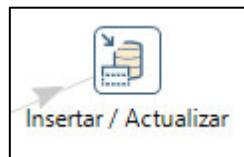


Fuente: Pentaho Data Integration - 2016

Insertar datos en otro esquema.

Esta función permite ingresar los datos, al nuevo esquema creado para el Data Mart (Ver figura 27).

Figura 27 - Insertar datos I



Fuente: Pentaho Data Integration - 2016

Se ingresa el primary key con el cual se mandara todas las filas a ingresar (Ver figura 28).

Figura 28 - Insertar datos II

Nombre de paso: Insertar / Actualizar

Conexión: Prueba [Editar...] [Nuevo...] [Wizard...]

Esquema destino: DW [Examinar...]

Tabla destino: DEMO_USERS [Examinar...]

Tamaño transacción (commit): 100

No realizar actualizaciones:

La clave(s) para realizar búsqueda de valor(es):

#	Campo de tabla	Comparador	Campo1	Campo2
1				

[Obtener Campos]

Campos de actualización:

#	Campo de tabla	Campo de Flujo	Actualizar
1	USER_ID		N

[Obtener campos de actualizaci...]

[Edit mapping]

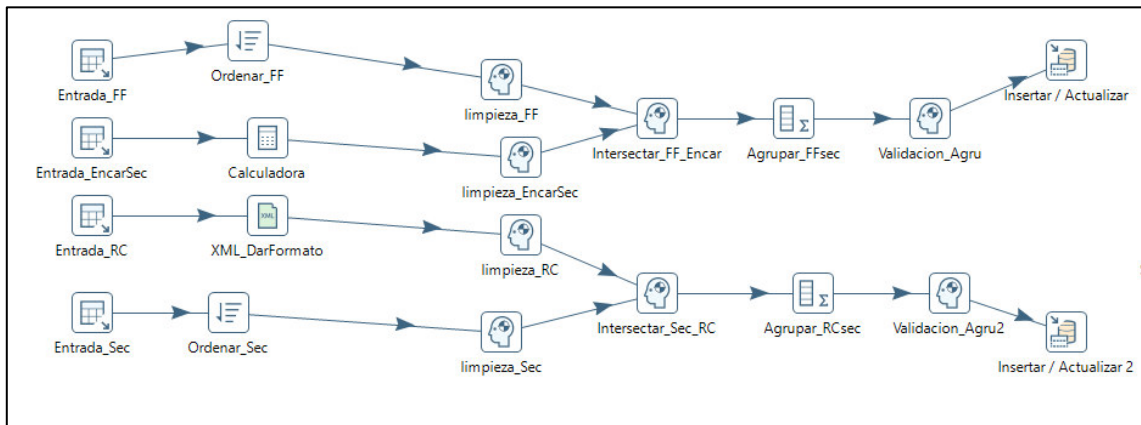
[?] Help [Vale] [Cancelar] [SQL]

Fuente: Pentaho Data Integration - 2016

Flujo de trabajo.

Se muestra como queda el workspace terminado (Ver figura 29).

Figura 29 - Flujo de Trabajo



Fuente: Pentaho Data Integration - 2016

5.4.2 VISUALIZACIÓN

Luego de haber realizado la extracción, limpieza, transformación y carga de la data a nuestro Data Mart, se visualizara en un dashboard elaborado con la herramienta QlikView.

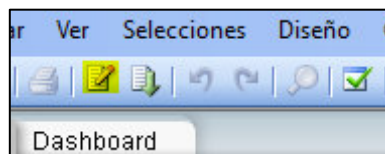
Conectarse al Data Mart.

Se deben conectar los datos del Data Mart a la aplicación, para eso debemos tener el ODBC de Oracle, seguir lo siguientes pasos:

Paso1:

Ir a la barra de herramientas y dar clic en el botón (Ver figura 30).

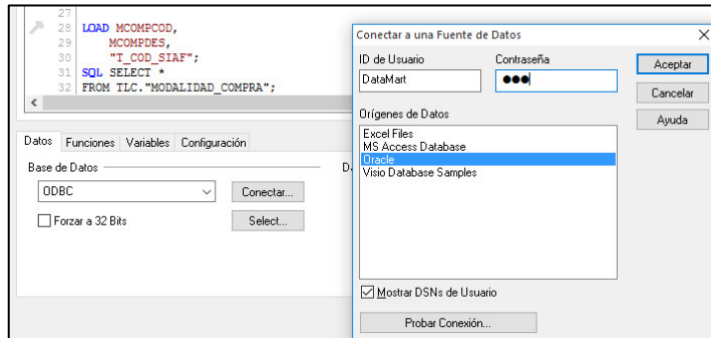
Figura 30 - Conectarse al Data Mart I



Fuente: QlikView Personal Edition - 2016

Luego en la parte de Base de datos, seleccionar ODBC y clic en conectar, ahora ingrese los datos de su conexión al Data Mart (Ver figura 31).

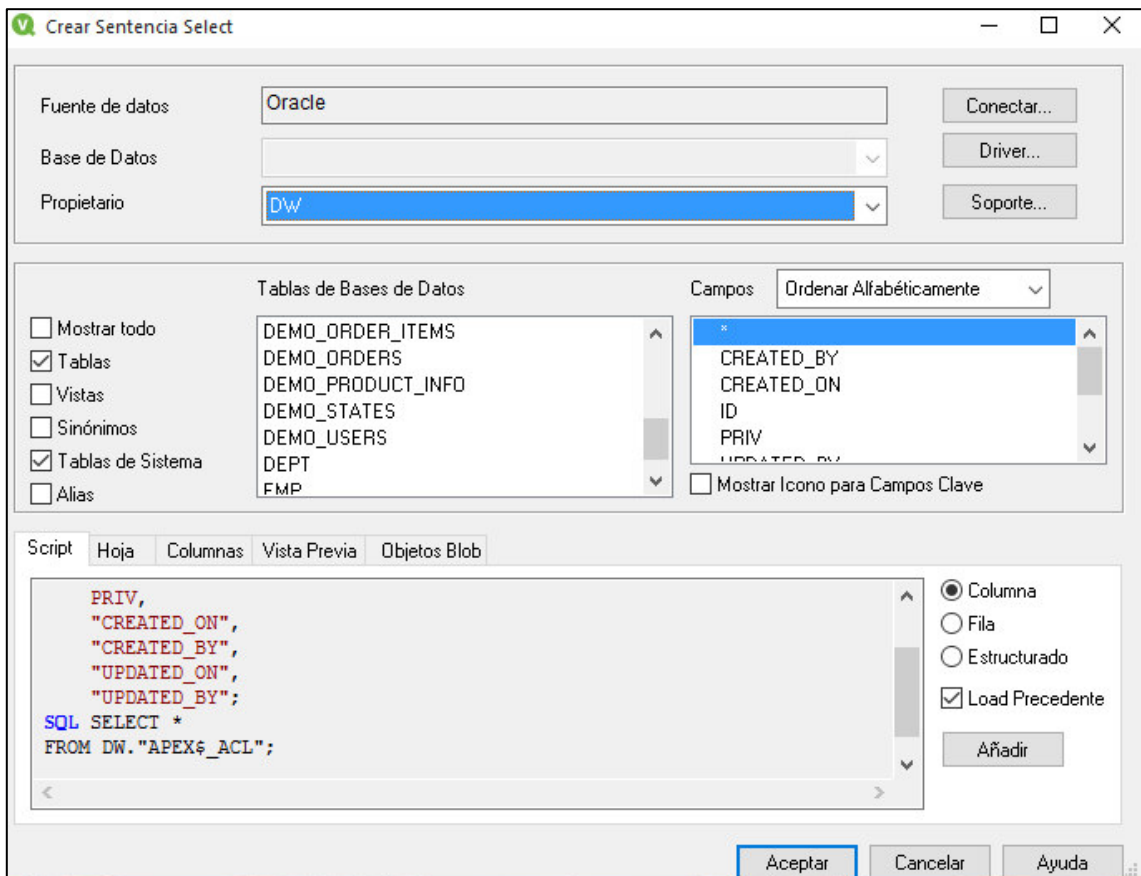
Figura 31 - Conectarse al Data Mart II



Fuente: QlikView Personal Edition - 2016

Luego presione "Select" e ingrese todas las tablas del Data Mart, con el botón añadir (Ver figura 32).

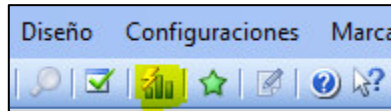
Figura 32 - Conectarse al Data Mart III



Fuente: QlikView Personal Edition - 2016

Ahora ya se puede comenzar a realizar los gráficos que irán en el dashboard, para ello nos vamos a la opción de asistente de gráficos (Ver figura 33).

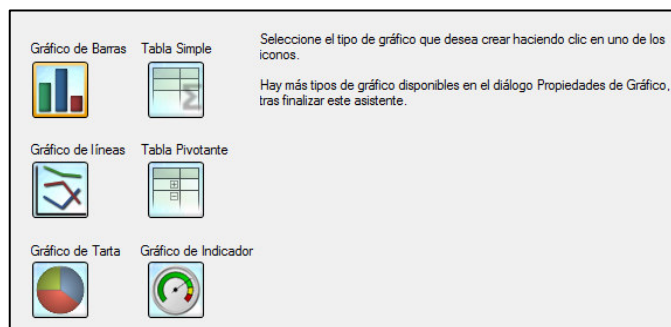
Figura 33 - Generar gráficos I



Fuente: QlikView Personal Edition - 2016

Tenemos que elegir el tipo de grafico que se elaborada (Ver figura 34).

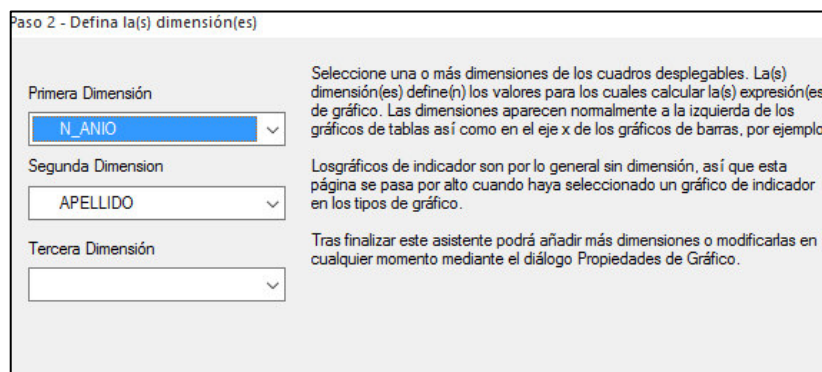
Figura 34 - Generar gráficos II



Fuente: QlikView Personal Edition - 2016

Elegimos las dimensiones que estarán presentes en el gráfico (Ver figura 35).

Figura 35 - Generar gráficos III



Fuente: QlikView Personal Edition - 2016

Seleccionamos el tipo de expresión que se utilizara (Ver figura 36).

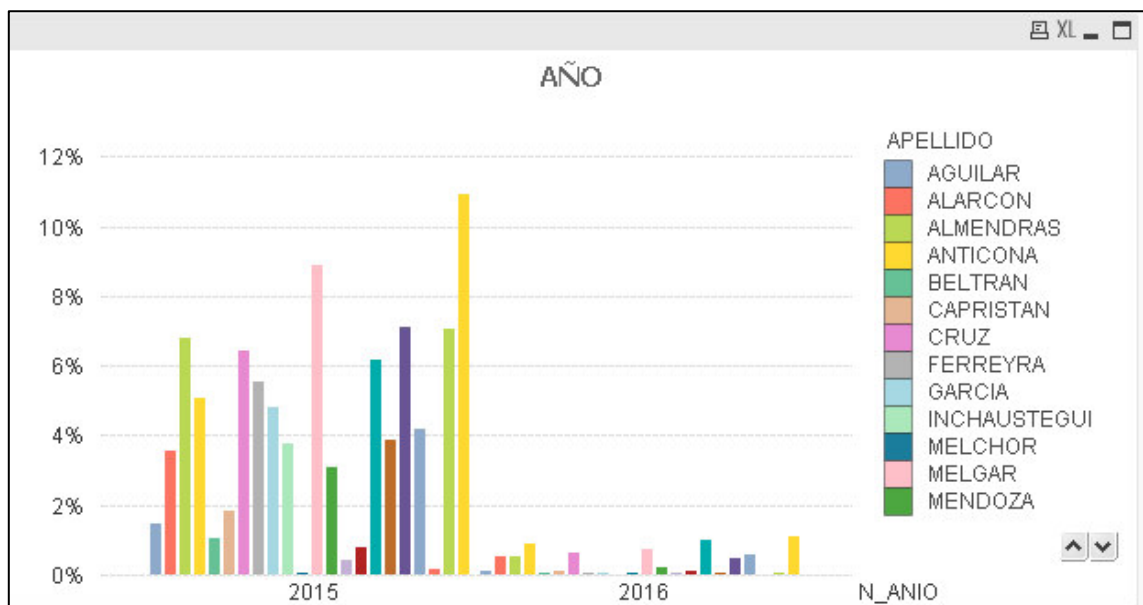
Figura 36 - Generar gráficos IV

Suma
 Media
 Contador
de
 Personalizar

Fuente: QlikView Personal Edition - 2016

Luego damos en Finalizar y tendremos el grafico para el dashboard (Ver figura 37).

Figura 37 - Registros procesados por año



Fuente: QlikView Personal Edition - 2016

Se realizará el mismo procedimiento para la realización de los demás gráficos.

Por fuente de financiamiento.

Este gráfico nos permitirá conocer cuáles son las FF más usadas por la UNMSM, para la realización de pagos, además el porcentaje y cantidad de registros que procesa un determinado sectorista (Ver figura 38).

Figura 38 - Registros procesados por FF

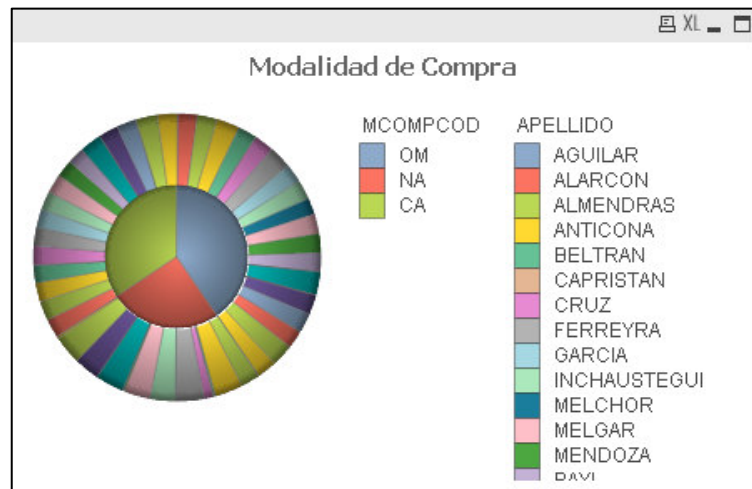


Fuente: QlikView Personal Edition - 2016

Por modalidad de compra.

Este gráfico nos permitirá conocer cuáles son las Modalidades de Compra más usadas por la UNMSM, para la realización de pagos, además el porcentaje y cantidad de registros que procesa un determinado sectorista (Ver figura 39).

Figura 39 - Registros procesados por MC

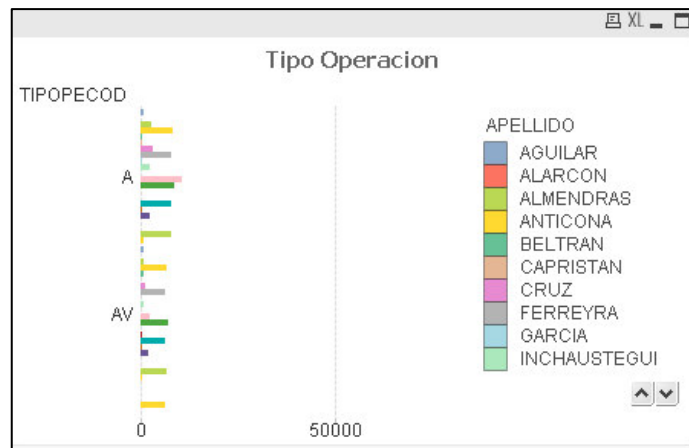


Fuente: QlikView Personal Edition - 2016

Por tipo de operación.

Este gráfico nos permitirá conocer cuáles son los Tipos de Operación (TO) más usados por la UNMSM, para la realización de pagos, además el porcentaje y cantidad de registros que procesa un determinado sectorista (Ver figura 40).

Figura 40 - Registros procesados por TO



Fuente: QlikView Personal Edition - 2016

CAPITULO VI

ANÁLISIS DE RESULTADOS

En esta sección se presentará los resultados que se dieron al explotar los datos transaccionales de la Oficina de Contabilidad de la UNMSM como parte de la solución de inteligencia de negocios desarrollada en esta tesis.

6.1 RESULTADO DE LOS INDICADORES

Por motivos de privacidad de información, no se mostrarán los resultados del personal de la oficina, pero mostraremos los resultados a nivel general.

6.1.1 A NIVEL DE FUENTE DE FINANCIAMIENTO

La fuente de financiamiento más utilizada fue Recursos Directamente Recaudados (09), en los años 2015 y 2016, siendo la menos utilizada la fuente Canon y Sobrecanon, Regalías, Renta de Aduanas y Participaciones (18), en los años 2015 y 2016.

6.1.2 A NIVEL DE MODALIDAD DE COMPRA

La modalidad de compra más utilizada fue Diferente a las Anteriores (OM), en los años 2015 y 2016, siendo la menos utilizadas la modalidad de compra No Aplicable (NA), en los años 2015 y 2016.

Tabla 9 - Resultado de los indicadores

Indicador Cantidad de Registros	Año	Fuente de financiamiento	Modalidad de Compra
Mayor % en	2015	09	OM
Menor % en	2016	18	NA

CONCLUSIONES Y RECOMENDACIONES

En el presente capítulo se indican las diferentes conclusiones deducidas como consecuencia del proyecto realizado, consejos y sugerencias para proyectos futuros.

7.1 CONCLUSIONES

- Se identificaron las necesidades de la oficina de contabilidad para poder realizar el análisis de su información y mediante ella, la visualización del rendimiento de su personal
- Se seleccionó la metodología Hefesto, para el desarrollo de la solución de inteligencia de negocios, que llevo a la creación de un Data Mart para la oficina de contabilidad.
- Se implementó una solución de inteligencia de negocios utilizando la herramienta Pentaho Community Edition, para la elaboración de los procesos ETL.
- Finalmente se utilizó satisfactoriamente las diferentes funcionalidades de Qlikview: herramienta para la presentación de información.

7.2 RECOMENDACIONES

- Se debe mantener reuniones constantes con los encargados de las bases de datos, para definir claramente los significados de los campos necesarios para obtener los indicadores y poder precisar que tablas contienen los datos requeridos para el análisis, puesto que se detectaron que los nombres de algunos campos no eran tan específicos y en otros casos los datos eran inconsistentes o no constaban, por lo que es necesario verificar que los datos entregados sean los requeridos para realizar el análisis antes de ejecutar la carga inicial.
- Se recomienda implementar la solución en un servidor diferente al que contenga la base de datos (OLTP) y que sea Oracle 11g en adelante.
- La solución propuesta se podría integrar o unirse con otros tipos de fuentes de datos (Sql Server, MySQL, Excel, etc).
- Si bien la solución esta implementada con la información del módulo de contabilidad, la misma puede ser implementada en diferentes módulos

del proyecto quipucamayoc.

- Analizar la posibilidad de incorporar sectoristas de otras dependencias como, por ejemplo, la oficina de abastecimiento, oficina de tesorería, entre otras.
- Clasificar los rechazos según el tipo anulación, extorno y devolución, para permitir desarrollar diferentes gráficos para el dashboard.
- Aplicar a los sectoristas un test para determinar su estilo de trabajo y encontrar en qué punto tienen inconvenientes en la hora de trabajar los registros contables.

REFERENCIAS BIBLIOGRÁFICAS

1. Almeida, G. (27 de Octubre de 2015). *Bitácora sobre Análisis de Negocios*. Obtenido de Cómo aplicar Business Analysis a soluciones de Data Warehouse o Business Intelligence: <http://gabrielalmeida.com.mx/1043/como-aplicar-business-analysis-a-soluciones-de-data-warehouse-o-business-intelligence/>
2. Business Intelligence Consulting. (04 de Noviembre de 2015). *SAP Business Objects*. Obtenido de <http://www.biconsulting.com.mx/sap-business-objects/>
3. Bustos Barrera, S., Mosquera Artieda, V., Campaña, M., & Duque, L. (2013). *Análisis, diseño e implementación de una solución de BI para la generación de indicadores y control de desempeño, en la empresa OTECEL, utilizando la metodología hefesto v2.0*. Ecuador.
4. Campanaro, R., & Ingressia, R. (2012).
5. Castelán García, L., & Ocharán Hernández, J. (2012). *Diseño de un Almacén de datos basado en Data Warehouse Engineering Process (DWEP) y HEFESTO*. Veracruz. Obtenido de <http://www.uv.mx/mis/files/2012/11/disenio-de-un-almacen-de-datos.pdf>
6. Castillo, J. Y. (2012). Implementación de un Data Mart como una solución de inteligencia de negocios para el área de logística de T-Impulso. *Revista de Investigacion de Sistemas e Informatica*, 2-2.
7. Dario, B. R. (2010). *HEFESTO*. Cordoba, Argentina. Obtenido de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>
8. Espñera, S. y. (27 de Octubre de 2015). *La Inteligencia de Negocios*. Obtenido de Business Intelligence: <http://www.pwc.com/ve/es/asesoria-gerencial/boletin/assets/boletin-advisory-edicion-10-2008.pdf>
9. Filofteia Tutuneeaa, M., & Veronica Rusa, R. (2012). *Business intelligence solutions for SME's*.
10. IBM. (11 de Noviembre de 2015). *InfoSphere DataStage*. Obtenido de <http://www-03.ibm.com/software/products/es/ibminfodata>

11. Inmon, B. (2002). *Building the Data Warehouse* (Tercera ed.). Wiley.
12. JAVA. (04 de Noviembre de 2015). *Enhydra Octopus Application*. Obtenido de <http://octopus.ow2.org/doc/3.0/manuals/OctopusManual.html#d0e14>
13. Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit. 2nd Edition*. New York.
14. Lluís Cano, J. (27 de Octubre de 2012). *Fundación IWITH*. Obtenido de Business Intelligence: Competir con Información: http://www.iwith.org/pdf/Libro_BI_Compertir_con_Informacion.pdf
15. Lozada Peñafiel, X., Cruz Tamayo, H., Pérez Argudo, W., & De la Torre Díaz, A. (2014). *Análisis, Diseño, Construcción e Implementación de un Data Warehouse para toma de decisiones y construcción de los KPI, para la empresa Kronosconsulting CIA LTDA*.
16. MicroStrategy. (04 de Noviembre de 2015). *MicroStrategy*. Obtenido de <https://www.microstrategy.com/es>
17. OpenRed. (04 de Noviembre de 2015). *Soluciones en Software Libre*. Obtenido de <http://www.openred.es/index.php/pentaho>
18. Oracle. (03 de Noviembre de 2015). *Business Intelligence*. Obtenido de http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf
19. Pérez, C. (2015). II. Bill Inmon. *VORTECH-IT*. Obtenido de <http://vortech-it.com/assets/cap-ii-bill-inmon.pdf>
20. Proyectos, G. d. (27 de Octubre de 2015). *Gestión de Proyectos*. Obtenido de Gestión de Proyectos
21. QLIK. (04 de Noviembre de 2015). *QlikTech International*. Obtenido de <http://global.qlik.com/es/explore/products/qlikview>
22. Rayner Huamantumba. (28 de Octubre de 2007). *Data Mart paso a paso*. Obtenido de La rueda tecnologica: [http://www.raynerhd.com/wp-content/uploads/rayner-Data Mart.pdf](http://www.raynerhd.com/wp-content/uploads/rayner-Data%20Mart.pdf)
23. Rivadera, G. R. (2010). *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. Buenos Aires. Obtenido de

<http://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadere-formateado.pdf>

24. Rojas Zaldivar, A. (2014). *Implementación de un Data Mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el departamento de finanzas de la Contraloría General de la República*. Chiclayo.
25. Roland Bouman, J. v. (2010). *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL*. WILEY .
26. Serra, J. (2013). *Building an Effective Data Warehouse Architecture. 24 Hours of Pass*.
27. Sinnexus. (27 de Octubre de 2015). *Business Intelligence Informatica Estrategica*. Obtenido de Datamar: http://www.sinnexus.com/business_intelligence/Data_Mart.aspx
28. Standen, J. (27 de Octubre de 2015). *Data Martist*. Obtenido de Data Warehouse vs Data Mart: http://www.Data_Martist.com/data-warehouse-vs-data-mart
29. Surajit Chaudhuri, & Umeshwar Dayal. (1997). *An Overview of Data Warehousing and OLAP Technology*. Obtenido de <http://www.cs.sfu.ca/CourseCentral/459/han/papers/chaudhuri97.pdf>
30. UNMSM. (16 de Noviembre de 2015). *Universidad Nacional Mayor de San Marcos*. Obtenido de UNMSM: <http://www.unmsm.edu.pe/home/organizacion/economia>
31. UNMSM. (25 de Enero de 2016). *Reglamento de Organización y Funciones (R.O.F.)*. Obtenido de <http://www.unmsm.edu.pe/transparencia/archivos/ROF-2011-UNMSM-RR.pdf>

ANEXOS

GLOSARIO DE TERMINOS

BI: Son las siglas en inglés de Inteligencia de Negocios (Business Intelligence), es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa.

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

OLAP: Es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

DW: Son las siglas de Data Warehouse. Es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

DSS: Sistema de Apoyo a las Decisiones (DSS por sus siglas en inglés Decision Support System). Es un sistema informático utilizado para servir de apoyo, más que automatizar, el proceso de toma de decisiones. La decisión es una elección entre alternativas basadas en estimaciones de los valores de esas alternativas. El apoyo a una decisión significa ayudar a las personas que trabajan solas o en grupo a reunir inteligencia, generar alternativas y tomar decisiones.

EIS: Sistema de Información Ejecutiva (Executive information system, EIS por sus siglas en inglés) es una herramienta de Inteligencia empresarial (Business Intelligence, BI), orientada a usuarios de nivel gerencial, que permite monitorizar

el estado de las variables de un área o unidad de la empresa a partir de información interna y externa a la misma.

CMI: El concepto de Cuadro de Mando Integral – CMI (Balanced Scorecard – BSC) es un sistema de administración o sistema administrativo, que va más allá de la perspectiva financiera con la que los gerentes acostumbran evaluar la marcha de una empresa. Es una herramienta de administración de empresas que muestra continuamente cuándo una compañía y sus empleados alcanzan los resultados definidos por el plan estratégico.

JDBC: Java Database Connectivity, más conocida por sus siglas JDBC, es una API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java, independientemente del sistema operativo donde se ejecute o de la base de datos a la cual se accede, utilizando el dialecto SQL del modelo de base de datos que se utilice.

CSV: Los archivos CSV (del inglés comma-separated values) son un tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas y las filas por saltos de línea.

QLIK: Qlik es una plataforma de análisis de datos. Con Qlik podemos analizar datos y descubrir cosas por nosotros mismos. Podemos compartir los conocimientos y analizar datos en equipos o en toda la empresa. Qlik nos permite formular y responder nuestras propias preguntas, así como seguir nuestro propio camino de investigación e indagación en los datos.

SAP: El Sistema SAP o “Systems, Applications, Products in Data Processing”, es un Sistema informático que le permite a las empresas administrar sus recursos humanos, financieros-contables, productivos, logísticos y más, las principales empresas del mundo utilizan SAP para gestionar de una manera exitosa todas las fases de sus modelos de negocios.

OLTP: Son las siglas en inglés de Procesamiento de Transacciones En Línea (OnLine Transaction Processing). Es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida.

HEFESTO: Es una metodología de desarrollo de Data Warehouse, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias en procesos de confección de almacenes de datos. Elaborado por Ruben Dario.

DWEP: La metodología denominada proceso de ingeniería para el desarrollo de almacenes de datos (DWEP) la cual está basada en el proceso unificado (PU), abarca los flujos de trabajo de requerimientos, análisis, diseño, pruebas, mantenimiento y revisiones posteriores al desarrollo; sus principales características son que es iterativa, dirigida por casos de uso y se basa en las etapas de desarrollo de PU, utilizando UML como lenguaje para modelado gráfico.