



TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE APIZACO

DESARROLLO DEL MODELADO DE UN ALMACÉN DE DATOS COMO PARTE INTEGRAL DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS

TESIS

PARA OBTENER EL GRADO DE:
MAESTRA EN SISTEMAS COMPUTACIONALES

PRESENTA:
ING. LOURDES LILIANA SUÁREZ MÉNDEZ

ASESORES:
DR. JOSÉ CRISPÍN HERNÁNDEZ HERNÁNDEZ
DR. PERFECTO MALAQUÍAS QUINTERO FLORES

APIZACO, TLAXCALA

SEPTIEMBRE 2018

Apizaco, Tlax., 01 de Agosto de 2018

No. de Oficio: DEPI/265/18

ASUNTO: Se Autoriza Impresión de Tesis de Grado.

ING. LOURDES LILIANA SUÁREZ MÉNDEZ
CANDIDATA AL GRADO DE MAESTRA
EN SISTEMAS COMPUTACIONALES
No. de Control: M11370869
PRESENTE.

Por este medio me permito informar a usted, que por aprobación de la Comisión Revisora asignada para valorar el trabajo, mediante la Opción: I Tesis de Grado por Proyecto de Investigación, de la Maestría en Sistemas Computacionales, que presenta con el tema: "DESARROLLO DEL MODELADO DE UN ALMACÉN DE DATOS COMO PARTE INTEGRAL DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS" y conforme a lo establecido en el Procedimiento para la Obtención del Grado de Maestría en el Instituto Tecnológico, la División de Estudios de Posgrado e Investigación a mi cargo le emite la:

AUTORIZACIÓN DE IMPRESIÓN

Debiendo entregar un ejemplar del mismo debidamente encuadernado y seis copias en CD en formato PDF, para presentar su Acto de Recepción Profesional a la brevedad.

Sin otro particular por el momento, le envío un cordial saludo.

ATENTAMENTE
EXCELENCIA EN EDUCACIÓN TECNOLÓGICA®
PENSAR PARA SERVIR, SERVIR PARA TRIUNFAR®



DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN.

JFCV/MZSH*mbn,
C.p. Expediente



SECRETARÍA DE EDUCACIÓN PÚBLICA
TECNOLÓGICO NACIONAL
DE MÉXICO
INSTITUTO TECNOLÓGICO DE APIZACO
DIVISIÓN DE ESTUDIO
DE POSGRADO E INVESTIGACIÓN



Apizaco, Tlax., 27 de Junio de 2018

ASUNTO: Aprobación del trabajo de Tesis de Maestría.

DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN.
P R E S E N T E.

Por este medio se le informa a usted, que los integrantes de la **Comisión Revisora** para el trabajo de tesis de maestría que presenta la **ING. LOURDES LILIANA SUÁREZ MÉNDEZ**, con número de control **M11370869**, candidata al grado de **Maestra en Sistemas Computacionales** y egresada del **Instituto Tecnológico de Apizaco**, cuyo tema es "**DESARROLLO DEL MODELADO DE UN ALMACÉN DE DATOS COMO PARTE INTEGRAL DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS**", fue:

A P R O B A D O

Lo anterior, al valorar el trabajo profesional presentado por la candidata y constatar que las observaciones que con anterioridad se le marcaron así como correcciones sugeridas para su mejora ya han sido realizadas.

Por lo que se avala se continúe con los trámites pertinentes para su titulación.

Sin otro particular por el momento, le envió un cordial saludo.

LA COMISIÓN REVISORA



DR. JOSÉ CRISPIN HERNÁNDEZ HERNÁNDEZ



DR. PERFECTO MALAQUIAS QUINTERO FLORES



DR. JOSÉ FEDERICO RAMÍREZ CRUZ



M.C. CARLOS DÍAZ GUTIÉRREZ

C p - Interesada

No pretendamos que las cosas cambien, si siempre hacemos lo mismo. La crisis es la mejor bendición que puede sucederle a personas y países, porque la crisis trae progresos. La creatividad nace de la angustia como el día nace de la noche oscura. Es en la crisis que nace la inventiva, los descubrimientos y las grandes estrategias. Quien supera la crisis se supera a sí mismo sin quedar "superado".

Quien atribuye a la crisis sus fracasos y penurias, violenta su propio talento y respeta más a los problemas que a las soluciones. La verdadera crisis, es la crisis de la incompetencia. El inconveniente de las personas y los países es la pereza para encontrar las salidas y soluciones. Sin crisis no hay desafíos, sin desafíos la vida es una rutina, una lenta agonía. Sin crisis no hay méritos.

Es en la crisis donde aflora lo mejor de cada uno, porque sin crisis todo viento es caricia. Hablar de crisis es promoverla, y callar en la crisis es exaltar el conformismo. En vez de esto, trabajemos duro. Acabemos de una vez con la única crisis amenazadora, que es la tragedia de no querer luchar por superarla.

Albert Einstein

Agradecimientos

Gracias Dios por todo lo que me permitiste que aprendiera, por permitir que adquiriera los conocimientos que tu creiste los más adecuados para este proyecto. Gracias por permitir que pusiera el corazón en cada actividad y en cada tarea, sin ti no lo hubiera logrado.

Gracias a mis hijos, Brandon Uriel y Jocelyn Faridi por ser testigos del gran esfuerzo y dedicación que se requiere para alcanzar las metas que uno se propone. Espero que en este ejemplo encuentren motivación hacia su futuro.

Agradezco de manera muy especial a mi esposo, Lic. Eliut Flores Jiménez, por compartir conmigo sus conocimientos, por el respaldo, la orientación y guiarme en el desarrollo de esta investigación.

Gracias a mi Director de tesis, Dr. José Crispín Hernández Hernández, por la oportunidad brindada para ingresar a la maestría, por la confianza depositada en su servidora para llevar a cabo el desarrollo de este trabajo de tesis y por los conocimientos compartidos.

De igual manera gracias al Dr. Malaquias, Dr. José Federico, M.C Carlos Diaz que a lo largo de mis estudios de maestría compartieron sus conocimientos y por todas aquellas tareas encomendadas que me ayudaron a enriquecer y fortalecer mis conocimientos.

Al mismo tiempo gracias a mi amigo, compañero y colega M.C Saúl Pérez Tirzo por su dirección, conocimiento, enseñanza, tiempo y colaboración para el desarrollo de este trabajo.

Una mención especial a mis amigos y compañeros Mary Carmen Cuecuecha, Saúl Pérez, Luis Herrera, Edith Vazquez y Lidia Corona por compartir juntos momentos de crisis y felicidad.

Un agradecimiento especial al Consejo Nacional de Ciencia y Tecnología CONACYT, por haber financiado mis estudios de maestría. Y a mi casa de estudios, el Instituto Tecnológico de Apizaco por recibirme nuevamente en sus instalaciones.

Infinitamente gracias.

Dedicado con todo mi amor a ...

A Dios

Por darme la sabiduría, foraleza, salud, motivación y ante todo fe para culminar esta etapa de mi vida. Que a pesar de todas las adversidades, nunca me abandonaste y me ayudaste a tomar mi cruz y continuar mi camino.

A mis hijos

A mi flaquito hermoso Brandon Uriel y a mi princesa hermosa Jocelyn Fardi, por ser el motivo e inspiración que me impulso a inicar esta etapa, ustedes son el sentido de mi vida. Los amo

A mis esposo

Eliut Flores, por ser mi compañero de vida, por ayudarme a ser mejor cada día en lo personal y en lo profesional. Por estar conmigo en los momentos críticos y felices. Por tu sabiduria y tenacidad.

A mis familiares

Primeramente a mi madre Micaela Méndez, por guiar mis primeros pasos y que apesar de nuestra diferencia de opiniones sigues siendo un apoyo y ejemplo para mi. A mi abuelita Joaquina Cruz, porque donde quiera que estés sigues siendo parte de cada uno de mis logros. A mis hermanos y cuñada Victor, Vianney, Ariadne y Celina respectivamente por su cariño, gran apoyo y comprensión. A mis sobrinos Abigail, Victor, Ricardo islas, y Sofia, Meredit y Alma Flores por quererme y regalarme momentos de felicidad.

PROTECCIÓN DE LA INFORMACIÓN

El contenido de este trabajo se encuentra protegido por la empresa Miracle Business Network S.A de C.V. La asesoría especializada y la estancia realizada por el tesista arriba mencionado para el desarrollo de esta investigación, cubre con las expectativas planteadas durante el desarrollo de la investigación. El acuerdo, desarrollo, seguimiento y resultados del mismo se encuentran a resguardo de la empresa como secreto industrial. No se debe utilizar el sistema y la documentación con fines de lucro o sin la autorización de la empresa, si se desea consultar más información, póngase en contacto a <http://www.mbn.com.mx>, con el representante autorizado de la organización de Miracle Business Network S.A de C.V.

RESUMEN

Este trabajo de tesis tiene por objetivo desarrollar un modelo que permita implementar un almacén de datos en la empresa MBN, con el propósito de apoyar la toma de decisiones estratégicas y tácticas de dicha empresa. Este proyecto surge de la necesidad de proporcionar a los dirigentes del negocio información que les permita realizar análisis profundo de ella y tomar las decisiones que permitan alcanzar los objetivos que los directivos plantean.

Para ello, se realizó la identificación de la información y las fuentes de datos que utiliza la empresa para la elaboración de los informes. Posteriormente, se realizó el análisis de la información obtenida. Después se identificaron las preguntas por cada área para dar lugar a la creación de los indicadores, y así diseñar el modelado conceptual y lógico, para concluir con el diseño de la base de datos multidimensional.

A lo largo de este trabajo se encontró primeramente la descripción de la justificación del proyecto, los objetivos general y específico, pregunta de investigación, las aportaciones, así como las limitaciones para la realización de este trabajo. Inmediatamente se presentan los trabajos relacionados a inteligencia de negocios por los trabajos que se han realizado en este área de investigación del manejo de la información.

Después se presenta el modelo propuesto enfocado al desarrollo de almacén de datos. Con la finalidad de conocer la base y la principal fuente de datos para el desarrollo del almacén, se hace especial énfasis a la empresa y a las normas MoProSoft y CMMI. Finalmente en las últimas secciones se muestran los pruebas, resultados y conclusiones respectivamente.

ABSTRACT

The aim of this thesis is to develop a model that allows the implementation of a data warehouse in the MBN business with the purpose of supporting the strategic and tactical decisions of this firm. This purpose arises from the need to provide business leaders with information that allows them to carry out an in-depth analysis of it and make the decisions that allow achieving the objectives that the managers raise.

Therefore, the identification of the information and the data sources used by the company for the composition of the reports was carried out. Afterwards, the analysis of the information obtained was carried out. Then the questions were identified for each area to create the indicators, and thus design the modeling conceptual and logical, to conclude with the design of the multidimensional database.

Throughout this work we find first the description of the justification of the project, the general and specific objectives, the research question, the contributions, as well as the limitations of the realization of this work. Following this, the works related to business intelligence are presented for the work that has been done in this area of information management research.

Afterwards, the proposed model focused on the development of the data warehouse is presented. In order to know the base and the main source of data for the development of the warehouse, special emphasis is placed on the business and the MoProSoft and CMMI standards. Finally in the last sections the tests, results and conclusions are shown respectively.

TABLA DE CONTENIDOS

	Page
LISTA DE TABLAS	xv
LISTA DE FIGURAS	xvii
1 Introducción	1
1.1 Contexto del problema	1
1.2 Descripción del problema	2
1.3 Justificación	2
1.4 Objetivos	2
1.4.1 Objetivo General	2
1.4.2 Objetivos Específicos	2
1.5 Pregunta de investigación	2
1.6 Limitaciones	3
1.7 Aportaciones	3
1.8 Organización de la tesis	3
2 Estado del arte	5
2.1 Antecedentes	5
2.1.1 Estudio comparativo de enfoques de modelado de almacenes de datos: Inmon, Kimball y Bóveda de datos	5
2.1.2 Propuesta de una metodología para un proceso de Data Warehousing	6
2.1.3 Diseño del almacén de datos de la biblioteca con el método esquema copo de nieve: caso de estudio: base de datos de la biblioteca del Campus XYZ	7
2.1.4 Un enfoque novedoso para la distribución inteligente de almacenes de datos	7
2.1.5 Hacia Soluciones de almacenamiento de datos basadas en NoSQL	8
2.1.6 Un almacén de datos espaciales sensible a la calidad para consultar datos hidroecológicos	9
2.1.7 Diseño e implementación de almacenamiento de datos con modelo de datos utilizando datos de servicios basados en encuestas	10
2.1.8 Beneficios de un almacén de datos clínicos con herramientas de minería de datos para recopilar datos para una prueba de radioterapia	11

2.1.9	Diseño e Implementación de un Almacén de Datos para Benchmarking en Rehabilitación Clínica	11
2.1.10	La aplicación del sistema de comercio electrónico basado en el almacén de datos	12
2.1.11	Integración entre Gestión de las Relaciones con los Clientes (Customer Relationship Management CRM) y almacén de datos	12
2.1.12	Diseño e implementación de un almacén de datos para una cadena de pequeños supermercados	13
2.1.13	Análisis, Diseño e Implementación de una base de datos departamental para el área de sismología del departamento de geofísica de la escuela politécnica nacional	13
2.2	Definición de términos	14
2.3	Almacén de datos	15
2.3.1	Definición y características	16
2.3.2	Orientado	16
2.3.3	Datos integrados	16
2.3.4	Datos Históricos o de tiempo variante	17
2.3.5	Datos No volátiles	17
2.3.6	Objetivos del almacén de datos	18
2.3.7	Estructura del almacén de datos	18
2.3.8	Uso del almacén de datos	20
2.3.9	Arquitecturas del almacén de datos	20
2.3.10	Base de datos multidimensional	23
2.3.11	Tablas de Dimensiones	24
2.3.12	Tablas de Hechos	24
2.3.13	Métricas	24
2.4	Modelos de almacén de datos	24
2.4.1	Modelo Estrella	24
2.4.2	Modelo Copo de Nieve	25
2.4.3	Modelo Constelación	26
2.5	Metodologías existentes	26
2.5.1	CRISP-DM	27
2.5.2	Hefesto	28
2.5.3	Metodología ciclo de vida de Kimball	30
2.5.4	Inmon	33
2.5.5	Rapid Warehousing Methodology (RWN)	34
2.5.6	Comparativa de metodologías	35
2.6	Herramientas Utilizadas	36
3	Modelo propuesto	39
3.1	Introducción	39

3.2	Hefesto: metodología propia para la construcción de un almacén de datos	39
3.2.1	Características	40
3.2.2	Empresa analizada	40
3.3	Paso 1) Análisis de requerimientos	40
3.3.1	Identificar preguntas	40
3.3.2	Identificar indicadores y perspectivas	41
3.3.3	Modelo conceptual	41
3.4	Paso 2) Análisis de los OLTP	41
3.4.1	Conformar indicadores	42
3.4.2	Establecer correspondencias	42
3.4.3	Nivel de granularidad	42
3.4.4	Modelo Conceptual ampliado	42
3.5	Paso 3) Modelo lógico del almacén de datos	43
3.5.1	Tipo de Modelo Lógico del almacén de datos	43
3.5.2	Tablas de dimensiones	43
3.5.3	Tablas de hechos	45
3.5.4	Uniones	48
3.6	Paso 4) Integración de Datos	48
3.6.1	Carga inicial	48
3.6.2	Actualización	49
3.7	Desarrollo del modelo propuesto	49
3.7.1	Descripción de la arquitectura	49
3.7.2	Metodología del desarrollo	51
3.7.3	Esquema copo de nieve	63
4	Modelo MoProSoft y CMMi : la empresa MBN	67
4.1	La empresa	67
4.2	Principales líneas de productos o servicios	68
4.2.1	Desarrollo de software a la medida (DMS)	68
4.2.2	Consultoría de negocios e implementación funcional (CNIF)	68
4.2.3	Desarrollo, Mantenimiento y Personalización Oracle (DMPO)	68
4.2.4	Consultoría en Administración de Base de Datos (CABD)	68
4.2.5	Licenciamiento y comercialización de productos(LCP)	69
4.3	MoProSoft	69
4.3.1	Categorías por las que está conformado MoProSoft son: Alta dirección, Gerencia y Operación	69
4.3.2	MoProSoft tiene 9 procesos que se agrupan por categorías	70
4.3.3	Niveles de Madurez	72
4.3.4	Relación entre procesos	72
4.3.5	Los roles	73

4.3.6	Definición general del proceso	74
4.4	CMMI	81
4.4.1	Objetivo de CMMI	82
4.4.2	Áreas de Procesos Cubiertos por CMMI-DEV	84
5	Resultados	89
5.1	Introducción	89
5.2	Instalación de las herramientas	89
5.3	Recursos	89
5.4	Datos para el llenado de las tablas dimensión y hechos	90
5.4.1	Datos para la dimensión colaboradores	90
5.4.2	Datos para la dimensión nómina	90
5.4.3	Datos para la dimensión costos operación	91
5.4.4	Datos para la dimensión proyectos	92
5.4.5	Datos para la dimensión tiempo	92
5.4.6	Datos para la tabla hechos	93
5.5	Creación y llenado de las tablas en el almacén de datos	93
5.6	Realizando consultas bajo el esquema estrella	93
5.7	Realizando consultas bajo el esquema copo de nieve	97
5.8	Resultados de tiempo de consulta entre esquema estrella y esquema copo de nieve	101
6	Conclusiones	105
A	Publicación	109
	BIBLIOGRAFÍA	117

LISTA DE TABLAS

TABLA	Page
2.1 Tabla ilustrativa de artículos relacionados	14
2.2 Rutas principales de la metodología de Kimball	33
2.3 Comparación de metodologías	35
4.1 Niveles de Madurez de MoProSoft	72
4.2 Roles en MoProSoft	74
4.3 Entradas de la Gestión de Proyecto	75
4.6 Productos Internos de Gestión de Proyectos	78
4.7 Entradas de Recursos Humanos y Ambiente de Trabajo	80
4.8 Salidas de Recursos Humanos y Ambiente de Trabajo	80
4.9 Productos Internos de Recursos Humanos y Ambiente de Trabajo	81
5.1 Recursos	90

LISTA DE FIGURAS

FIGURAS	Page
2.1 Base de datos operacional y almacén de datos.	17
2.2 Ambiente OLTP y almacén de datos.	18
2.3 Almacén de datos, estructura.	19
2.4 Arquitectura de almacén de datos básica.	21
2.5 Arquitectura de almacén de datos con área de preparación.	22
2.6 Arquitectura con área de ensayo y con almacén departamental.	22
2.7 Cubo Multidimensional.	23
2.8 Modelo Estrella.	25
2.9 Modelo Copo de Nieve.	25
2.10 Modelo Constelación.	26
2.11 Proceso Jerárquico CRISP-DM.	27
2.12 Ciclo de vida de CRISP-DM.	27
2.13 Pasos de la metodología Hefesto.	29
2.14 Rectángulos y elipse que representan el análisis de requerimientos.	29
2.15 Pasos de la metodología Kimball.	31
2.16 Pasos de la metodología Inmon.	34
2.17 Metodología de almacenamiento rápido.	34
2.18 MySQL Workbench.	36
2.19 XAMPP.	37
2.20 Windows 8.	37
3.1 Pasos de la metodología HEFESTO	39
3.2 Modelado conceptual.	41
3.3 Modelo Conceptual ampliado.	43
3.4 Diseño de tablas de dimensiones.	44
3.5 Jerarquía de "GEOGRAFÍA".	44
3.6 Normalización de "GEOGRAFÍA".	45
3.7 Tabla de hechos.	46
3.8 Caso 1, preguntas.	46
3.9 Caso 1, diseño de tablas de hechos.	47
3.10 Caso 2, preguntas.	47

3.11	Caso 2, diseño de tablas de hechos.	47
3.12	Caso 3, preguntas.	48
3.13	Caso 3, unificación.	48
3.14	Arquitectura almacén de datos.	50
3.15	área resaltada de la arquitectura almacén de datos.	51
3.16	Metodología de Hefesto adaptada.	52
3.17	Modelo conceptual de los proyectos.	54
3.18	Correspondencias.	55
3.19	Correspondencias.	58
3.20	Correspondencias.	59
3.21	Correspondencias.	60
3.22	Correspondencias.	60
3.23	Correspondencias.	61
3.24	Correspondencias.	61
3.25	Correspondencias.	62
3.26	Diseño físico del almacén de datos.	63
3.27	Dimensión colaboradores en copo de nieve.	64
3.28	Jerarquía de la tabla colaboradores.	64
3.29	Jerarquía de la tabla proyectos.	65
4.1	Logotipo de la empresa MBN.	67
4.2	Diagrama de categoría de procesos.	70
4.3	Relación entre procesos.	73
4.4	Niveles de mejora CMMI.	83
4.5	Categorías de Proceso en CMMI-DEV.	84
4.6	Áreas de proceso de CMMI por categoría de proceso y nivel de madurez.	85
4.7	Organigrama de la empresa MBN.	87
5.1	Datos para dimensión colaboradores.	90
5.2	Datos para dimensión nómina.	91
5.3	Datos para dimensión costos operación.	91
5.4	Datos para dimensión proyectos.	92
5.5	Datos para dimensión tiempo.	92
5.6	Datos para tabla hechos.	93
5.7	Esquema estrella.	94
5.8	Resultado de la consulta1, esquema estrella.	94
5.9	Resultado de la consulta2.	95
5.10	Resultado de la consulta3.	95
5.11	Resultado de la consulta4.	96
5.12	Resultado de la consulta5.	96

5.13	Resultado de la consulta6.	97
5.14	Esquema copo de nieve.	97
5.15	Resultado de la consulta 1, esquema copo de nieve.	98
5.16	Resultado de la consulta 2 esquema copo de nieve.	98
5.17	Resultado de la consulta3.	99
5.18	Resultado de la consulta 4 esquema copo de nieve.	99
5.19	Resultado de la consulta 5 esquema copo de nieve.	100
5.20	Resultado de la consulta 6 esquema copo de nieve.	100
5.21	Comparativa de tiempos de consulta de copo de nieve y estrella.	101
5.22	Gráfica de la consulta 1.	101
5.23	Gráfica de la consulta 2.	102
5.24	Gráfica de la consulta 3.	102
5.25	Gráfica de la consulta 4.	103
5.26	Gráfica de la consulta 5.	103
5.27	Gráfica de la consulta 6.	104

IT	Information Technology
DW	Data Warehouse
ETL	Extraction Transformation Load
BI	Business Intelligence
CMMI	Capability Maturity Model
Integration	
OLAP	On Line Analytical Processing
CRISP-DM	Cross Industry Standard Process for Data Mining
KPI	Key Performance Indicators

INTRODUCCIÓN

Los desafíos en los distintos sectores tiene clientes cada vez más exigentes, cambios cada vez más rápidos y una competencia cada vez más fuerte. De manera que las empresas están produciendo profundas transformaciones. En consecuencia cada vez más empresas consideran que es indispensable administrar sus datos aplicando inteligencia de negocios y obtener información adecuada, por tal razón hacer frente a estos desafíos. Los datos con los que las empresas disponen son datos producidos de sus procesos establecidos y sus sistemas operacionales o bien del exterior.

Las empresas cuentan con grandes volúmenes de datos, sin embargo los datos por si solos exponen una mínima parte de información. Las empresas pueden obtener ventajas de sus datos. Es por ello por lo que se requiere disponer de información con calidad, que aporte conocimiento para facilitar la toma de decisiones.

1.1 Contexto del problema

Por lo tratado en el párrafo anterior se establece que se requiere de soluciones tecnológicas que les proporcionen, por un lado, indicadores que les permitan medir el desempeño de sus procesos de gestión general, y por otro lado, patrones o reglas que les permitan describir o predecir comportamientos en sus movimientos o transacciones de los procesos de negocios. Todo esto para apoyar la toma de decisiones estratégicas de una manera mejor preparada (Castillo,2017).

Para esto, se hace uso de tecnologías que permitan el análisis de grandes volúmenes de datos, que permitan generar conocimiento con los datos que produce la empresa de sus procesos. Se puede mencionar que se logra haciendo uso de herramientas como ETL, los almacenes de datos o DW, OLAP, minería de datos, entre otras. En conjunto las herramientas de las que se acaba de hacer mención forman parte de lo que actualmente se conoce como inteligencia de negocios. La inteligencia de negocios cuenta con una arquitectura que gracias a su conjunto de componentes es la que ayuda a los objetivos estratégicos y tácticos de la toma de decisiones. La arquitectura

de inteligencia de negocios sera tratada a profundidad en el Capítulo 3. Como parte central de la arquitectura se encuentra el almacén de datos, es donde se integran y se concentran los datos operativos, de varias fuentes, y los proporcionan a los usuarios en determinados formatos para ser analizados y que al final aportaran conocimiento para ayudar a la empresa a la hora de tomar decisiones y que éstas sean estratégicas.

Este trabajo de tesis solo se enfoca a la creación del almacén de datos, por dicha razón, no se tratarán temas de proceso de extracción, transformación y carga de datos. La explotación de los datos tampoco es tema a tratar en esta tesis; se hace mención de estos procesos porque forman parte de la arquitectura de la inteligencia de negocios. En la figura 3.15 del Capítulo 3 se muestra la arquitectura de inteligencia de negocios en ella se muestra la fase o área que es interés para este trabajo.

1.2 Descripción del problema

Actualmente en la empresa Miracle Business Network S.A. de C.V. (MBN) la información se encuentra dispersa, hay heterogeneidad de datos y se encuentran en diferentes tipos de fuentes. En la empresa MBN se necesita realizar el análisis de los datos que se producen en ella, modelarlos e integrarlos bajo un esquema de inteligencia de negocios y obtener información útil para la toma de decisiones.

1.3 Justificación

Cada vez más las empresas u organizaciones confían en los almacenes de datos, para integrar información y así poder implementar inteligencia de negocios para orientar a la empresa al crecimiento e innovación en el sector TI y con ella consolidarse como una de las empresas más competentes del estado.

1.4 Objetivos

1.4.1 Objetivo General

Proponer el modelado de un almacén de datos que permita la integración de los datos para generar información que será utilizada para la toma de decisiones.

1.4.2 Objetivos Específicos

- Analizar la información que es obtenida bajo los estándares de calidad.
- Diseñar el modelado multidimensional de los datos.
- Modelo listo y preparado para la explotación de la información.

1.5 Pregunta de investigación

¿Es posible realizar la toma de decisiones a partir de la creación de un almacén de datos bajo estándares de calidad de desarrollo de software, como parte integral de un sistema de inteligencia de negocios?

1.6 Limitaciones

Este caso práctico se enfocará en el aspecto técnico-tecnológico y no se llevará a cabo un análisis de los aspectos económicos y sociales que intervienen en el proceso de diseño e implementación de un almacén de datos. Es decir se pondrá un modelo de almacén de datos para la empresa MBN, se implementará bajo la infraestructura actual de la empresa MBN y se utilizarán herramientas de código libre, por lo tanto no se requerirá hacer una inversión adicional?

1.7 Aportaciones

En este trabajo se realizó un estudio de las metodologías que con más frecuencia son utilizadas para la construcción de almacén de datos, primero se realizó un estudio por separado de cada una de ellas, explicando que es lo que cada una ofrece para la construcción del almacén de datos, posteriormente se hace una comparativa de características entre ellas. Finalmente se explica cual es la que se elige para conveniencia del desarrollo de esta tesis. Una segunda aportación es el diseño de un almacén de datos para la empresa que es el objeto de caso de estudio que la ayudará a integrar datos de diferentes fuentes para que al ser procesados y analizados le brinden una toma de decisiones eficaz y estratégica. Como tercer aportación se desarrolló un modelado estrella y copo de nieve de los datos, con el fin de ofrecer una comparativa entre ambos. Se definieron indicadores que miden ciertos comportamientos en los procesos de la empresa y que a partir de estos la empresa tomara decisiones más precisas.

1.8 Organización de la tesis

En el resto de esta tesis, se presenta en el Capítulo 2, en estado del arte. En la Capítulo 3 modelo propuesto para el desarrollo del almacén de datos de este trabajo. En el Capítulo 4 Modelo MoProSoft yCMMi: caso de estudio MBN. En el Capítulo 5 se muestran los resultados obtenidos y finalmente en el Capítulo 6 las conclusiones a las que se llegaron en la realización de este trabajo y trabajos a futuro.

ESTADO DEL ARTE

Un almacén de datos fundamentalmente tiene como principio, almacenar datos históricos para ser utilizados para la toma de decisiones estratégicas. Dichos sistemas son específicamente de consulta y están enfocados a extraer datos históricos almacenados. Es por ello por lo que un almacén de datos no es una base de datos tradicional en la que el usuario final puede hacer modificaciones como inserciones, borrados y actualizaciones sobre la base de datos. Un almacén de datos permite las herramientas que pueden realizar el análisis de los datos históricos son las herramientas OLAP, basadas en un modelado multidimensional, que a su vez se basa el análisis en cubos.

Desde su aparición en los noventa, existe una definición de almacén de datos proporcionada por Bill Inmon (Inmon,2005), considerado uno de los pioneros y padre de esta tecnología: "un almacén de datos en una colección de datos orientados por tema, integrados, variantes en el tiempo y no volátiles que se emplea como apoyo a la toma de decisiones estratégicas". Según Ralph Kimball(Kimball, 2008), considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos, un almacén es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis.

2.1 Antecedentes

A continuación se describen los trabajos relacionados con el diseño e implementación de almacén de datos:

2.1.1 Estudio comparativo de enfoques de modelado de almacenes de datos: Inmon, Kimball y Bóveda de datos

(Lamia Yessad,Aissa Laboid, 2016), realizan un estudio comparativo entre tres enfoques Inmon,Kimball y Bóveda de datos. Primero realizan un estudio por separado de cada enfoque, posteriormente realizan una comparativa entre cada uno de ellos y finalmente incluyen una recomendación para seleccionar el mejor enfoque. Como resultado del estudio individual se da la definición del almacén de datos que cada uno de los enfoques presenta en el cual Kimball lo define como una copia de de datos estructurados específicamente para consultas y análisis

donde el objetivo es proporcionar información ayudar a la toma de decisiones a una empresa. Por su parte para Inmon un almacén de datos ésta orientado a temas, integrado, de colección variante y no volátil, es un conjunto de datos que da soporte a los administradores de la toma de decisiones. Dan Linstedt el creador de la Bóveda de datos la define como: orientada a los detalles, seguimiento histórico y conjunto único vinculado a tablas de normalizadas que soportan una o más áreas funcionales del negocio. Los enfoque de Inmon y Kimball ambos usan ETL para alimentar los almacenes de datos. La similitud entre Bóveda de datos y Kimball reside en la implementación de la solución y del uso del área de preparación para la recuperación y la sincronización. El enfoque Bóveda de datos y el enfoque Inmon están de acuerdo con el hecho de que el almacén de datos es el repositorio más grande de la compañía. En cuanto a filosofía el objetivo de Inmon son los profesionales de TI que proveen una solución técnica completa de objetivos basada en métodos probados y tecnologías. Para Kimball los usuarios finales son los que proporcionan una solución que facilita la integración directa de los datos. La Bóveda de datos proporciona un solución completa y solida, ésta basada en métodos probados. Los resultados que se obtuvieron en cuanto la metodología y la arquitectura es que en la estructura de la arquitectura Inmon el nivel atómico es el que alimenta al almacén de datos departamental. Para Kimball el conjunto de almacenes de datos departamentales es el que constituye al almacén de datos. La Bóveda de datos alimenta al almacén de datos departamental. En cuanto a la complejidad Inmon es bastante complejo, Kimball es simple al igual que la Bóveda de datos. A lo que se refiere del desarrollo de la metodología Inmon se inspira en la metodología de espiral, Kimball se basa en un proceso de cuatro pasos y la Bóveda de datos se basa en una metodología ágil. Los costos el enfoque de Inmon el costo inicial es alto y subsecuente el costo del desarrollo baja. En Kimball el costo inicial es bajo y en cada paso subsecuente el costo aumenta. Y en Bóveda de datos el costo del proyecto es menor que los dos enfoques precedentes. En tiempo Inmon es de larga duración. Kimball corta duración y Bóveda de datos corta duración. En la integración de datos ETL, Inmon las reglas de transformación deben implementarse en los procesos de ETL. Kimball las reglas de transformación deben implementarse en el proceso de ETL. En Bóveda de datos la separación de los satélites y las claves empresariales reduce la complejidad. En la etapa de modelado de datos Inmon maneja datos, Kimball maneja procesos y Bóveda de datos hace uso de ambos. Las herramientas para el modelado Inmon usa el modelo clásico de herramientas diagramas de entidad relación (ERD: Entity-Relationship Diagrams) y conjunto de elementos de datos (DIS: Data Item Set).

2.1.2 Propuesta de una metodología para un proceso de Data Warehousing

(Wilson Castillo, 2017), hace una propuesta de una metodología para el desarrollo y proceso de un Data Warehousing. Diseña una plantilla para la obtención de indicadores, es basada en un modelo para requerimientos de información para desarrollo de software. Una vez determinados y especificados los indicadores que medirán la productividad académica de una universidad, procede a desarrollar la metodología elaborada. La elaboración de la metodología se basa en un conjunto de diversos enfoques de dos metodologías clásicas de almacén de datos que son: la propuesta por Kimball y la metodología Hefesto. De kimball considera las 4 etapas: definición de requerimientos del negocio, modelado dimensional, diseño físico y diseño e implementación de ETL. La metodología de Hefesto consta de 4 etapas: análisis de requerimientos, análisis de OLTP, modelo lógico del almacén de datos y la integración de los datos. Todas las etapas son consideradas en un orden diferente. Kimball: Definición de requerimientos + Hefesto: análisis de requerimientos + Hefesto: análisis de los OLTP es igual a la

etapa de análisis de la nueva metodología. Kimball: modelado dimensional + Kimball: diseño físico + Hefesto: modelo lógico del almacén de datos es igual a diseño modelo multidimensional de la nueva metodología. Kimball: diseño e implementación del subsistema ETL + Hefesto: integración de datos es igual al proceso ETL + V de la nueva metodología. La V en este último proceso significa la Validación de los indicadores. En esta actividad se analizan y verifican los indicadores por medio de consultas SQL se realizan consultas, cálculos y validaciones necesarias. Se trata de verificar de que los indicadores sean consistentes con los datos depositados en el repositorio temporal, las fuentes originales y el almacén de datos. Se observaron buenos resultados de la propuesta de la metodología, permite entregar informes cuantitativos y cualitativos del estado actual de la gestión universitaria de las diferentes áreas de la institución. La plantilla fue clave para la especificación de los indicadores, y al mismo tiempo actuó como interfaz entre los usuarios y el equipo de desarrolladores del almacén de datos. Finalmente hacen uso de una herramienta OLAP QLIKVIEW para la explotación y visualización de los indicadores y es así como dan respuesta a los requerimientos que son motivo de la creación de la propuesta de la metodología. Por lo tanto los resultados se muestra en gráficas y en cuadros de mando para que sean analizados e interpretados por los usuarios finales.

2.1.3 Diseño del almacén de datos de la biblioteca con el método esquema copo de nieve: caso de estudio: base de datos de la biblioteca del Campus XYZ

(Akhmad Dahlan, 2017), realiza el diseño de un almacén de datos y agregan el proceso de ETL para la biblioteca de una universidad. Preparan el modelo de datos usando el método de esquema de copo de nieve. En tres etapas desarrollaron el modelo primero definiendo los datos de acuerdo al proceso de circulación de los libros en la biblioteca es decir en el pedir prestado y devolver los libros, segundo definieron el origen de los datos y los requisitos de información, finalmente en la tercera etapa crean el modelo de datos dimensional en forma de copo de nieve, mejorando así el diseño de ETL. Para este proceso lo llevan a cabo a partir de la base de datos de la biblioteca hacia el almacén de datos sin antes eliminar el ruido de los datos, segundo haciendo la población de las tablas del almacén de datos. Como resultado concluyen cual es el proceso de construir un almacén de datos con el método de esquema de copo de nieve a través de varios pasos continuos como es diseñar el esquema, realizar ETL y hacer cubo de datos dimensional y que el modelado y la fabricación de un sistema de base de datos transaccional faciliten la preparación del almacén de datos.

2.1.4 Un enfoque novedoso para la distribución inteligente de almacenes de datos

En este trabajo, el crecimiento de los datos en una organización, genera a razón de este la creación de un almacén de datos centralizado, el cual es el responsable de almacenar todos los datos de la organización, responder todas las consultas y tomar decisiones. Entonces a medida que aumentan los datos y la cantidad de consultas surge la necesidad de distribuir los datos almacenados, lo que a su vez se desarrolla la necesidad de un almacén de datos distribuidos. Lo que pretenden alcanzar Abhay Kumar Agarwal y Neelendra Badal es proponer un almacén de datos distribuido e inteligente que utiliza un enfoque de diseño de arriba hacia abajo. Este IDDW no requiere de que se ubique un almacén de datos en cada ubicación sino que se puede construir un almacén en cualquier momento según la necesidad de esa ubicación, otra ventaja es que su arquitectura es tal que cada usuario registrado

en el sistema está identificado de manera única. Para la construcción del modelo de datos se basa en el uso de 3 ecuaciones matemáticas con la ecuación uno enumera al almacén de datos como y cuando se construye en cualquier ubicación en el IDW. Posteriormente después del modelo propuesto con las fórmulas (1-3) se genera la arquitectura del IDW. Para la construcción del IDW se apoya en 3 algoritmos. El algoritmo uno se usa para construir un almacén de datos de una ubicación en un nivel uno. El algoritmo dos se usa para construir un almacén de datos en varios niveles y el algoritmo tres se usa si el almacén de datos se forma en las ubicaciones de dos niveles no consecutivas uno después del otro (por ejemplo: en el nivel tres o en el nivel cuatro, solo después del nivel uno). Este enfoque novedoso es presentado en una universidad, en el país de la India. Obtienen una jerarquía de 8 niveles partiendo de que en el país de la India está formada por muchos estados; cada estado tiene muchas universidades; cada universidad tiene instituciones afiliadas; cada instituto tiene departamentos; cada departamento tiene categorías de personas; cada categoría de personas tiene puestos. Y consideran un nivel anual para formar una estructura jerárquica con el año en el nivel más superior y la posición es el nivel más bajo. Además, introducen un nivel más, en la jerarquía en el nivel más bajo, que genera un número único para cada persona registrada en el sistema y así cada persona puede identificarse de manera única dentro del sistema. Cada nivel de la estructura jerárquica de 8 niveles se nombra en consecuencia de arriba abajo. Se construye un almacén de datos por cada jerarquía, haciendo uso de la arquitectura jerárquica propuesta. Para la implementación del caso de estudio escribieron un programa en Java con Net Beans: 6.9 como un IDE, utilizando un servidor web Apache Tomcat 6.0.26 y para la compilación de las tablas se utilizan MySql 5.0.45. el hardware requerido para modelar el estudio consistió en un conmutador de 16 puertos y 8 maquinas con las siguientes especificaciones: procesador Core 2 Duo, 3.0 GHz y 1 GB de RAM. Las observaciones que aportan es que el sistema es capaz de almacenar y vincular millones de registros. Menciona que es un sistema altamente adaptable. Tiene una alta disponibilidad el almacén de datos en una ubicación con una sobre carga mínima desde el lugar de las entradas del usuario. Además, el sistema es lo suficientemente capaz de realizar un seguimiento de todos y cada uno de los registros sobre el usuario ingresado. En cuanto a confiabilidad el IDW con registros distribuidos en muchos almacenes de datos, el sistema completo nunca falla, sin embargo, si un almacén de datos falla solo esa parte del sistema falla relacionada con ese almacén de datos, por lo tanto, el sistema es altamente confiable en comparación con el almacén de datos centralizado. Y finalmente la escalabilidad del sistema es altamente escalable tanto vertical como horizontalmente.

2.1.5 Hacia Soluciones de almacenamiento de datos basadas en NoSQL

En este trabajo y de acuerdo con (Oditis, 2016), menciona que los almacenes de datos se usan en las empresas para el manejo de los datos producidos por los procesos que se llevan a cabo dentro de ellas. Las bases de datos no relacionales son aún muy inusuales en las tareas de procesamiento de datos. Por lo que se implementó un almacén de datos NoSQL basado en documentos. El almacén de datos basado en NoSQL fue desarrollado utilizando una tecnología de NoSQL, un almacén de datos departamental NoSQL y una herramienta de desnormalización. Se ha demostrado que la tecnología NoSQL permite enfoque ágil hacer uso de las herramientas de desnormalización y sincronización de datos semiestructurados y no estructurados para ser cargados en el almacén de datos departamental basado en NoSQL. Como resultado demuestran que las bases de datos relacionales NoSQL brindan flexibilidad, alta lectura/escritura, rendimiento así como la escala a conjuntos de datos muy grandes.

2.1.6 Un almacén de datos espaciales sensible a la calidad para consultar datos hidroecológicos

Desarrollan nuevos métodos de minería de datos para estudiar, comparar y hacer un uso efectivo de todos los parámetros disponibles relacionados con el estado de las aguas corrientes como información que describe usos y medidas tomadas. Más concretamente aborda dos cuestiones específicas: (1) obtener información sobre el funcionamiento de las aguas corrientes mediante el análisis de taxones (es decir animales acuáticos o plantas) que soportan índices biológicos, (2) que conectan las fuentes de presión, la fisicoquímica y la calidad biológica de las aguas corrientes. Todos estos datos tienen características espaciales y son complejos de estructurar e interconectar debido a su volumen y naturaleza. Se caracterizan por su alta heterogeneidad. Los datos obtenidos no son de la más alta calidad debido al tamaño, la complejidad de la red de agua francesa, los cambios en los parámetros y métodos. Ahora bien, debido al costo, tiempo y aporte humano necesarios para las campañas de muestreo, estos datos imperfectos recopilados son los únicos datos disponibles. (L. Berrahou , 2015), su objetivo es vincular todos los datos recopilados para crear una evaluación global de masas de agua. Para lo cual diseñaron una base de datos espaciotemporal y un almacén para explorar los datos y apoyar el proceso de decisión de evaluación. Su enfoque es integrar los procesos de calidad desde el modelado hasta el paso de implementación en función de la experiencia del dominio. La integración de procesos de calidad de datos es actualmente una de las principales prioridades para ayudar a tomar decisiones adecuadas. Dice (L. Berrahou , 2015), que en lugar de evaluar la calidad de los datos con métricas e integrar los datos de acuerdo con las reglas de limpieza como generalmente se hace en el proceso de ETL, se decidieron abordar los problemas de calidad de los datos en un nuevo marco de mejora.

El modelo de sistema de información espaciotemporal propuesto aquí integra varios aspectos de calidad de datos en un paquete dedicado. De acuerdo con el asesoramiento de expertos, la precisión espacial, temática y temporal de los datos son compatibles:

- La precisión espacial mide las relaciones que pueden existir entre las características de los objetos a fin de mantener las relaciones espaciales relevantes entre ellos.
- La precisión temática describe la precisión de los valores de los atributos codificados en la base de datos. Por ejemplo, los valores de las mediciones fisicoquímicas no tienen la misma confianza si se incluye un umbral de cuantificación o detección válido. El tema hidrológico está integrado en el enfoque para hacer diagnósticos más relevantes del estado hídrico ecológico.
- La precisión temporal mide la medida en que los datos recopilados permanecen constantes a lo largo del tiempo.

Con base a las definiciones anteriores de los principales elementos de calidad, consideraron el contexto de la toma de decisiones donde el experto necesita una gran cantidad de información en un curso de tiempo, así como datos precisos para tomar decisiones correctas a lo largo del proceso de análisis.

La arquitectura empleada se ajusta a los requisitos de calidad dentro de un sistema de información ambiental. La arquitectura propuesta puede transferir indicadores de calidad de datos de bases de datos de entrada al almacén

de datos espaciales. Las consultas están diseñadas para someterse a diferentes jerarquías, como suele ser el caso de un contexto multidimensional.

En el paso de la calidad de los datos la tabla hechos provee de parámetros de valores de acuerdo los diferentes niveles de jerarquías dimensionales.

- Dimensión espacial
- Dimensión temporal
- Dimensión parámetro

Otra dimensión espacial y temática ha sido desarrollada, la dimensión administrativa (una estación de muestra se encuentra en un pueblo o territorio de la aldea, que se incluye en un departamento, etc.). Cada registro en la tabla hechos puede ser consultado en varias dimensiones del cubo espaciotemporalmente analizado.

En los resultados la arquitectura fue enteramente desarrollada en una tecnología de código abierto usando PostgreSQL DBMS la cual es perfectamente adaptada para datos espaciales y Mondrian server para consultas de cubos, adecuada para datos espaciales porque admite dos tipos de datos espaciales: tipo de dato geometry y tipo de dato geography. Hacen uso de una arquitectura QUIDA, que es una arquitectura que se ajusta a los requisitos de calidad dentro de un sistema de información ambiental. Esta arquitectura se puede usar para transferir indicadores de calidad de datos, desde bases de datos de entrada hacia al almacén de datos espaciales. Posteriormente presta especial atención a los pasos de modelado de datos e integración de acuerdo con las dimensiones de calidad de datos y definición de los pesos de calidad como indicadores de datos auténticos.

2.1.7 Diseño e implementación de almacenamiento de datos con modelo de datos utilizando datos de servicios basados en encuestas

(Seah Selan, 2014), presenta un análisis de servicio basado en encuestas con el diseño e implementación de un marco de almacén de datos para la extracción de datos y la presentación de informes de inteligencia empresarial. El desarrollo del almacén de datos lo hace a través del diseño de modelado multidimensional para la creación de mercados de datos múltiples para el análisis de estos, desarrollado con un esquema estrella y dimensiones comunes para facilitar el desglose cruzado de los datos utilizados en el cubo OLAP, posteriormente diseñan el proceso de extracción, transformación y carga (Extract, Transform and Load (ETL)) para la población de datos desde las fuentes heterogeneas, hasta el almacén de datos, la información se proporciona a través de los archivos MS Access MBD. Para la visualización de los datos, los datos se analizaron y procesaron de acuerdo con las demandas. Para modelar el almacén de datos primero obtuvieron los temas comerciales sobre el tipo de análisis que pueden generar los informes. Luego realizan una verificación de la calidad de los datos y posteriormente implementan el proceso ETL en el desarrollo del almacén de datos. Los resultados obtenidos son el diseño e implementación del el almacén de datos utilizando su propio marco de almacén de datos y el diseño del modelado de datos.

2.1.8 Beneficios de un almacén de datos clínicos con herramientas de minería de datos para recopilar datos para una prueba de radioterapia

(Erik Roelofs a, 2013), desarrolla un portal de investigación para integrar las fuentes de datos médicos esenciales y ofrecer herramientas automatizadas de extracción de datos para fines de investigación. Para probar el rendimiento de este portal con herramientas de minería de datos en comparación con el proceso de recolección manual para pruebas clínicas. El rendimiento se mide en gastos de tiempo y calidad de datos para enfocarse en la hipótesis de que estos disminuirán y mejorarán, respectivamente, mediante el uso de un almacén de datos. Se diseñaron flujos de trabajo totalmente automatizado y personalizado en un marco llamado DIGITrans. Éste ofrece una interfaz de usuario y servicios en segundo plano para la generación, extracción y transporte de datos para fines clínicos y de investigación; ayudando así a la práctica clínica diaria con los flujos de trabajo de radioterapia 2D, mientras que para los usuarios e investigación, con los flujos de trabajo para extraer parámetros.

(Erik Roelofs a, 2013), descubrió que la recopilación por dos observadores y el monitoreo y la validación son vitales para la calidad de los datos porque resuelve los conflictos entre los observadores. Menciona que al utilizar enfoques inteligentes de búsqueda / minería de texto libre y técnica de aprendizaje automático para recuperar los datos en un depósito de datos, los datos pueden validarse automáticamente y pueden identificarse valores de datos verdaderos. Por otro lado, se menciona que utilizar el portal de investigación para la recopilación de datos de prueba con eCRF (electronic case report form) utilizando datos extraídos de bases de datos locales, posiblemente podría ser un mejor comienzo de la recopilación de datos e incluso reemplazar al primer observador. Las soluciones de tecnología de la información, como el almacén de datos de CAT (Computer Aided Theragnostics), mejoran la investigación clínica para la radioterapia al reducir el tiempo necesario para recopilar los datos necesarios y al mejorar la calidad de los datos recopilados. Se espera que las inversiones iniciales en la digitalización de datos se compensen debido a la flexibilidad del análisis de datos. Además, las investigaciones sucesivas pueden seleccionar fácilmente candidatos de prueba y extraer nuevos parámetros de las bases de datos existentes.

2.1.9 Diseño e Implementación de un Almacén de Datos para Benchmarking en Rehabilitación Clínica

(João Completo, 2012), llevo a cabo el diseño y desarrollo de un almacén de datos para respaldar un Sistema de Soporte de Decisiones (Decision Support Systems (DSS)) dinámico para Benchmarking en Rehabilitación Clínica. Para el proceso del desarrollo del almacén de datos hace uso de dos aspectos importantes para admitir un DSS dinámico para Benchmarking clínico el primero es la aplicación de la metodología que sigue el marco de trabajo del ciclo de vida empresarial de Kimball y segundo atender a las necesidades de los clientes de gestión de la salud, en la forma de un DSS para Benchmarking clínico. Hace uso de los elementos de un almacén de datos que involucra la extracción de datos On-Line OLTP, modelado de datos multidimensional y los almacenes de datos departamentales son cargados. Para facilitar el análisis y la visualización de los datos es por medio del procesamiento analítico OLAP. Para el diseño e implementación del almacén de datos.(João Completo, 2012) , aplicó la prueba de preparación Tornosol que permitio verificar el alto nivel de preparación que tenía la organización para avanzar con un proyecto de almacén de datos. Como resultado el uso de escalas de medición

funcional estándar y un sistema adecuado de clasificación de pacientes permite realizar un Benchmarking clínico bien abordado por la arquitectura del almacén de datos y los mecanismos de procesamiento Analítico en Línea OLAP.

2.1.10 La aplicación del sistema de comercio electrónico basado en el almacén de datos

(Hao, Hongwei, Zili, 2011), construyen un almacén de datos en el comercio electrónico para reflejar la información en profundidad para reducir costos y aumentar las ventas, y además, mejorar la competitividad de las empresas. El desarrollo del almacén de datos se emplearon las siguientes herramientas: el sistema de almacenamiento de datos que está compuesto por modelos de almacén de datos, preprocesamiento de datos y análisis de front-end (OLAP), modelado de los datos para diseñar el tema del análisis del sistema, el siguiente paso es la extracción, transformación y carga (ETL) y finalmente el análisis del almacén de los datos de acuerdo con los requisitos. El método que se eligió para el depósito de datos fue el modelado de datos multidimensional, el modelo usado fue modelo de estrella. Para el ETL se procesaron los datos e incluyó la eliminación de datos extracción de datos y carga de datos de acuerdo con el tema necesario y no para todos los datos. Para la gestión de los metadatos registró información importante del depósito de datos. También se consideró la gestión del índice y el análisis OLAP. (Hao, Hongwei, Zili, 2011), diseñaron para el almacén de datos en los sistemas de comercio electrónico una plataforma de almacenamiento que contendrá los datos según las necesidades del negocio, el modelo es un diseño "basado en datos". Diseñaron el modelo del almacén de datos del sistema de comercio electrónico, con el modelado multidimensional que consiste en las tablas dimensiones principales y la tabla de hechos y establecieron la relación entre ellos. Para el diseño del ETL (Hao, Hongwei, Zili, 2011), se centra en la limpieza de los datos para posteriormente extraerlos y cargarlos según el modelo del depósito de los datos. La función del metadato en su documento se refiere a la administración los datos, almacenar las reglas de limpieza y extraer reglas. Como resultado se obtuvo el modelado de los temas cliente y ventas, el modelado puede lograr la combinación perfecta original del sistema de comercio electrónico. La limpieza de datos se llevó a cabo con grandes cantidades de datos obteniendo velocidades de procesamiento de 10 millones de datos con un tiempo de procesamiento de 3 min. Al igual para el proceso de análisis de OLAP, no se cree que la velocidad de ejecución sea diferente.

2.1.11 Integración entre Gestión de las Relaciones con los Clientes (Customer Relationship Management CRM) y almacén de datos

(Abeer Khan, 2011), las organizaciones especialmente en Pakistán que han cambiado al almacenamiento de datos para CRM han sido investigadas analíticamente. La metodología de investigación empleada fue la investigación analítica. Los hechos y la información ya disponibles han sido extraídos y analizados para hacer una evaluación crítica. El almacén de datos es utilizado como una herramienta de comunicación con todas las partes del almacén de datos, está diseñado para recopilar datos de diversas fuentes de datos operativos, la información recopilada dentro del almacén de datos ayuda al descubrimiento de conocimiento para mejorar los procesos en la organización. La arquitectura de un almacén de datos también es estudiada en esta investigación ya que es la que permite llevar a cabo el desarrollo del almacén datos de forma ordenada y estructurada pues la conforman la capa de datos detallados de todos los sistemas fuentes relevantes. Todos los datos históricos y transaccionales residen dentro de

un almacén de datos centralizado basado en la extracción de datos y herramientas de acceso a la información. Como resultado obtienen diferentes beneficios, como un procesamiento de ETL minimizado, datos más oportunos y de alta calidad, alineaciones garantizadas con los objetivos del negocio, reducción del costo operativo, mejora en los servicios al cliente y la retención de clientes.

2.1.12 Diseño e implementación de un almacén de datos para una cadena de pequeños supermercados

(Xiewu, 2010), implementa un almacén de datos para resolver el embotellamiento de desarrollo de datos acumulativos y masivos que la base de datos de una cadena minorista de supermercados no puede almacenar. Se hace uso del modelo de almacenamiento estrella para desarrollar el almacén de datos, haciendo uso de la herramienta Microsoft SQL Server Analysis Services, la importación de los datos a las tablas dimensiones se realizó por medio de un proceso de extracción, transformación y cargar datos y para la tabla hechos se selecciona como recursos de datos y se adopta Microsoft Excel como herramienta de front-end para generar la conexión directa con las tablas dinámicas y los gráficos interactivos. Para el diseño del almacén de datos los desarrolladores primero investigan cada uno de los departamentos las encuestan son desde el representante corporativo, el líder del proyecto hasta los usuarios, los consumidores de datos y los gerentes de información hasta obtener un sólido respaldo. Con la información obtenida se crea la tabla dimensiones y la tabla hechos. Ésta última contendrá los datos de ventas, transacciones, inventarios, etc. El resultado fue que el sistema de almacenamiento fue llevado a cabo sin dificultades de lenguaje y técnicas y que los datos masivos y diversos ahora se almacenan sin dificultad.

2.1.13 Análisis, Diseño e Implementación de una base de datos departamental para el área de sismología del departamento de geofísica de la escuela politécnica nacional

Desarrollo (Michael Wladimir Vizuete Naranjo, 2006) una solución base de datos departamental para el departamento de Sismología de la Escuela Politécnica Nacional con el fin de mejorar el manejo de la información de dicha área para prevenir sismos que han producido un gran impacto en muchas comunidades ocasionando grandes pérdidas tanto humanas como económicas. Las plataformas empleadas es un sistema operativo Microsoft Windows XP profesional versión 5.1 con service park 2, para construir la base de datos y para la carga de datos se utilizó Microsoft SQL server 2000, la transformación multidimensional de los datos de la base de datos departamental de sismología se utilizó a través de la herramienta Analysis Services disponible en Microsoft SQL Server 2000, para obtener la información almacenada en la base de datos multidimensional es con Open I. Posteriormente se realiza la construcción de la base de datos, ya generada la base de datos se procede a extraer los datos de las distintas fuentes, más tarde se transformaron y cargaron los datos a la base de datos de la base de datos departamental de sismología. Después de tener cargados los datos en las tablas de dimensiones y hechos de la base de datos departamental se procedió a generar diferentes cubos con la herramienta de Analysis Services de Microsoft SQL 2000 y finalmente para la visualización de los datos utilizaron la herramienta Open I. Los resultados son satisfactorios gracias a las diferentes herramientas que facilitan las tareas del desarrollo de una base de datos departamental, aunque los datos son el tema principal y son los que sufren un proceso de transformación al pasar de un ambiente operacional a un ambiente de base de datos multidimensional.

En la tabla 2.1 se muestran los artículos más apegados al tema realizado en este trabajo.

Tabla 2.1: Tabla ilustrativa de artículos relacionados

Nombre artículo	Autores	Año	Resultados
Impacto de la inteligencia empresarial en la calidad de la toma de decisiones un modelo de mediación	Bernhard Wieder Maria-Luise Ossimitz	2015	Los resultados confirman por medio del PLS y el análisis de mediación que la calidad de gestión de inteligencia de negocios tiene efectos directos e indirectos positivos sobre la calidad de los datos.
Un enfoque novedoso para la distribución inteligente de almacenes de datos	Abhay Kumar Agarwal Neelendra Badal	2016	Se obtiene un IDWW que está modelado matemáticamente que ayuda a enumerar matemáticamente cada almacén de datos integrado en el IDWW
La metodología ATDM para apoyar el diseño e implementación de un almacén de datos empresarial	Brenda Scholtz, Charmain Cilliers, Cornel Ferreira	2015	Se identificaron tres áreas temáticas administración de computadoras, gestión de impresión y gestión de procesos. Se identificaron dos indicadores para cada campo comercial, también se identificaron preguntas comerciales para la implementación del almacén de datos
Proponen una nueva metodología para construir sistemas multidimensionales de almacenamiento de datos e inteligencia de negocios para empresas de servicios públicas	Jelena Lukic Milos Radenkovic a Marijana Destopovic Zrakib Aleksandra Labusb Zorica Bogdanovic	2016	La integración de los datos que provenían de diferentes fuentes de datos y en diferentes formas. Se obtuvo un almacén de datos que permitió a la empresa almacenar sus datos para poder realizar presentaciones de informes precisos.
Estudio comparativo de enfoques de modelado de almacenes de datos: Inmon, Kimball y Bóveda de datos	Lamia Yessad, Aissa Labiod	2016	Ninguno de los enfoques cumple con todos los requisitos a la perfección. La elección de un método de basa en el conocimiento de la empresa y sus políticas para lograr el enfoque favorable.
Inteligencia de negocios aplicada en tamaño pequeño para empresas de beneficios	Carmine D'Arconte	2018	La distribución de satisfacción del cliente se obtuvo como resultado una curva gaussiana que subdividirá la cartera de los clientes en cuatro segmentos.
Soluciones de inteligencia empresarial para pymes	Mihaela Filofteia Tutunea, Rozalia Veronica Rus	2015	Presenta tres tipos de soluciones de inteligencia de negocios: código abierto, soluciones SaaS y Cloud BI
Análisis de costo-beneficio de las metodologías de diseño del almacén de datos	Francesco Ditria Ezio Lefons Filippo Tangorra	2017	Las metodologías GrHyMM, MDBE y GQM presentan los mejores valores en términos de costos y beneficios. Obtienen una mayor calidad con menos esfuerzo al adoptar GrHyMM en lugar de las otras metodologías.

En la tabla superior se encuentra el nombre del artículo relacionado, los nombres o el nombre del autor, el año de publicación del artículo y los resultados obtenidos en los artículos.

2.2 Definición de términos

El hacer uso de los datos para generar información y que esa información se convierta en conocimiento y el conocimiento permite a las organizaciones tomar decisiones adecuadas para tener una mayor ventaja

competitiva, es el objetivo principal del almacén de datos. Es por ello por lo que es necesario abordar una serie de conceptos para tener un mejor entendimiento de lo que implica un almacén de datos. Los conceptos serán tratados a continuación:

Inteligencia de negocios Antes de dar una definición de lo que es inteligencia de negocios se debe recordar que los datos generados en las organizaciones son producto de sus operaciones que realizan día con día. Los datos generalmente son almacenados y administrados por sistemas transaccionales en bases de datos relacionales. Las organizaciones pueden sacar provecho de su información y obtener beneficios tales como: incrementar sus ganancias, enfocarse en retener a sus clientes actuales, así como también conseguir nuevos. Es por ello por lo que las organizaciones necesitan tener un mejor control, conocimiento y clasificación de sus clientes ya sea por su rentabilidad y permitan actuar a partir de ello.

Definición

La inteligencia de negocios (Business Intelligence-BI) permite llevar un tipo de segmentación, además no solo permite responder preguntas de qué está sucediendo o qué sucedió sino también permite predecir situaciones futuras. La inteligencia de negocios se puede definir como un concepto que integra por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración (Ing. Bernabeu, 2007).

La inteligencia de negocios no es un producto ni un sistema. Es una arquitectura y una colección integrada de aplicaciones tanto de apoyo a la toma de decisiones, como operacionales, así como bases de datos que facilitan a la comunidad de negocio, el acceso a los datos del negocio.

La definición antes expuesta puede representarse con la siguiente fórmula:

$$\text{Datos} + \text{Análisis} = \text{Conocimiento}$$

Dice (Ing. Bernabeu, 2007), existe una definición muy popular de inteligencia de negocios, que dice: "Inteligencia de Negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones".

La inteligencia de negocios remarca que las organizaciones deben hacer uso efectivo de información con el fin de mejorar la operación del negocio brindando información específica a los encargados de tomar decisiones en la organización. Aparte de lo ya mencionado es posible identificar y corregir situaciones antes de que se conviertan en problemas que generen potenciales pérdidas de control de la empresa. Entonces entre más inteligencia tenga la empresa sobre un negocio, mayor será su ventaja competitiva y se podrá tomar mejores decisiones.

La inteligencia de negocios tiene sus raíces en los Sistemas de Información Ejecutiva (Executive Information Systems (EIS)) y en los Sistemas para la Toma de Decisiones (Decision Support Systems (DSS)), pero ha evolucionado y se ha transformado en todo un conjunto de tecnologías capaces de satisfacer a una gran gama de usuarios junto a sus necesidades específicas en cuanto al análisis de información (Ing. Bernabeu, 2007).

2.3 Almacén de datos

El llevar a cabo inteligencia de negocios significa gestionar datos guardados en diferentes formatos, fuentes y tipos, para posteriormente integrarlos y almacenarlos en un solo lugar para examinarlos en un solo destino bien

conocido como base de datos, más adelante, es analizada y explotada, es necesario contar con un proceso que cubra todas estas necesidades. Este proceso recibe el nombre de almacenamiento de datos.

El proceso de almacenamiento de datos es el responsable de la extraer, transformar, integrar los datos que son generados en una organización como resultado de sus procesos de operación y actividad diaria incluso información externa relacionada. De esta manera permite el acceso y la explotación de la información requerida. Existe una gama de herramientas que permiten el análisis, con el fin de dar soporte a que las organizaciones puedan tomar decisiones adecuadas y estratégicas.

2.3.1 Definición y características

Un almacén de datos es una base de datos relacional que es más bien diseñada para consultas y análisis que para procesamiento de transacciones. Generalmente contiene datos históricos derivados de datos transaccionales, pero puede incluir datos de otras fuentes. Separa la carga de trabajo del análisis de la carga de trabajo de la transacción y permite que una organización consolide datos de varias fuentes (Lane, 2002).

En agregación a una base de datos relacional, el ambiente de un almacén de datos incluye una solución de extracción, transportación, transformación y carga (extraction, transportation, transformation, and loading (ETL)), un motor de procesamiento analítico en línea (an online analytical processing (OLAP)), herramientas cliente de análisis y otras herramientas que administran el proceso de reunir datos y entregarlos a los usuarios del negocio.

2.3.2 Orientado

Los almacenes de datos están diseñados para ayudar a analizar los datos. por ejemplo, para obtener más información acerca de los datos de ventas de la empresa, se puede construir un almacén que se concentre en las ventas. Haciendo uso de este almacén se pueden responder preguntas como "Quién fue el mejor cliente". Esta capacidad de definir un almacén de datos por tema, en este caso en cuestión las ventas, orientan al almacén de datos.

Esta primera característica del almacén de datos se refiere a que la información se clasifica en base a los intereses de empresa.

2.3.3 Datos integrados

La integración está estrechamente relacionada con el hecho de que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones o aplicaciones, tanto internos como externos deben ser consolidados en una instancia antes de ser agregados en el almacén de datos a este proceso se le conoce como ETL. La integración de los datos, (Lane, 2002), dice debe resolver tales problemas como nombrar conflictos e inconsistencias entre las unidades de medida, cuando logran esto, se dice que están integrados.

Los puntos de integración afectan casi todos los aspectos de diseño, y cualquiera que sea su forma, la información sera almacenada en el almacén de datos en un modelo globalmente aceptable y singular y la utilización de los datos no deba cuestionarse sobre la confiabilidad o solidez de estos.

2.3.4 Datos Históricos o de tiempo variante

Se denomina de tiempo variante debido al gran volumen de información que se manejará en el almacén de datos y cuando se le realice una consulta, los resultados deseados demorarán en originarse. El almacenamiento de datos históricos, es lo que permite al almacén de datos desarrollar pronósticos y análisis de tendencias y patrones. Ésto es en contraste con los sistemas de procesamiento de transacciones en línea (OLTP), donde los requisitos de rendimiento exigen que los datos históricos se muevan a un archivo.

2.3.5 Datos No volátiles

La información es útil para la toma de decisiones solo cuando es estable. Los datos una vez que entran en el almacén de datos no deben cambiar, eso significa que No es Volátil. Esto es lógico porque el propósito de un almacén es permitirle analizar qué ha ocurrido. Sin embargo, los datos operacionales varían momento a momento (Lane, 2002). En esencia el análisis y la toma de decisiones necesita una base de datos estable. Una base de datos en un ambiente operacional inserta, borra y modifica regularmente. A diferencia de un almacén de datos la manipulación básica de los datos es mucho más simple. En el almacén de datos hay dos únicos tipos de operaciones: la carga inicial de los datos y el acceso a los mismos. Es claro observar que no hay actualización de datos en un sentido general de actualización en el depósito como parte del procesamiento.

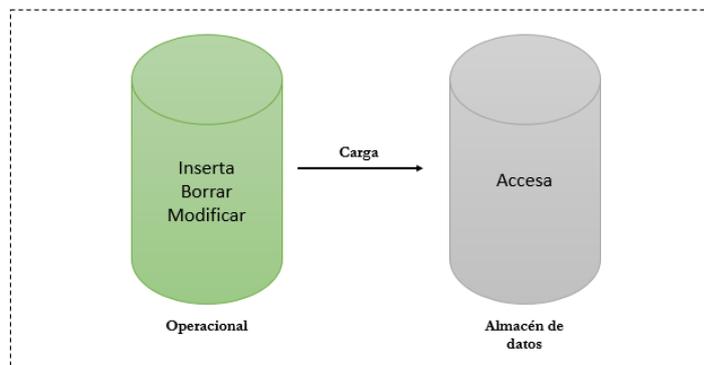


FIGURA 2.1. Base de datos operacional y almacén de datos.

Fuente: Diseño propio

Se observa en la figura 2.1 dos bases de datos a la derecha en una base de datos operacional la cual indica que en este tipo de bases de datos permiten insertar, borrar y modificar datos a diferencia de la base de datos del lado izquierdo que se refiere a un almacén de datos que solo permite el acceso a los datos históricos. También se observa que la base de datos operacional alimenta al almacén de datos.

Contraste entre OLTP y almacén de datos

En la figura 2.2 se ilustran las principales diferencias entre un sistema OLTP y un almacén de datos

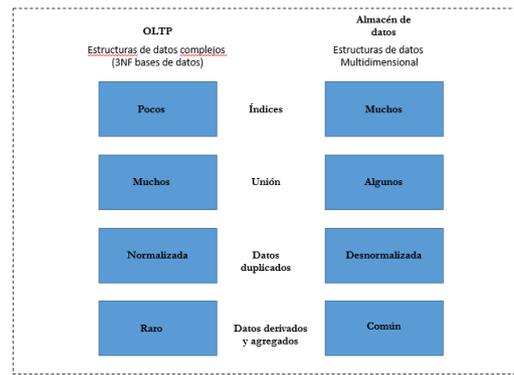


FIGURA 2.2. Ambiente OLTP y almacén de datos.

Fuente: Diseño propio

Una mayor diferencia entre los tipos de sistemas es que el almacén de datos generalmente no está en Tercera Forma Normal, un tipo común de normalización de datos en ambientes de OLTP. La imagen ilustra las diferencias entre las estructuras de datos de un OLTP y las estructuras de un almacén de datos. Los sistemas OLTP permite pocos índices, el almacén de datos muchos; OLTP permite muchas uniones, almacén de datos algunos; datos duplicados normalizados, almacén de datos desnormalizados; OLTP permite raramente datos derivados y agregados, almacén de datos es muy común.

2.3.6 Objetivos del almacén de datos

La información es vital para las empresas, los datos independientemente de cual sea su fuente deben organizarse, coordinarse, integrarse y almacenarse para dar al usuario del negocio una visión integrada y orientada al negocio. Con el almacén de datos y sus servicios lo hacen posible. Un almacén de datos es un área determinada que guarda en un único lugar toda la información útil proveniente de sistemas de producción y de fuentes externas.

2.3.7 Estructura del almacén de datos

Un almacén de datos se estructura en cuatro niveles diferentes de esquematización y detalle que los delimitan. En la figura número 2.3 se muestra mejor su respectiva estructura.

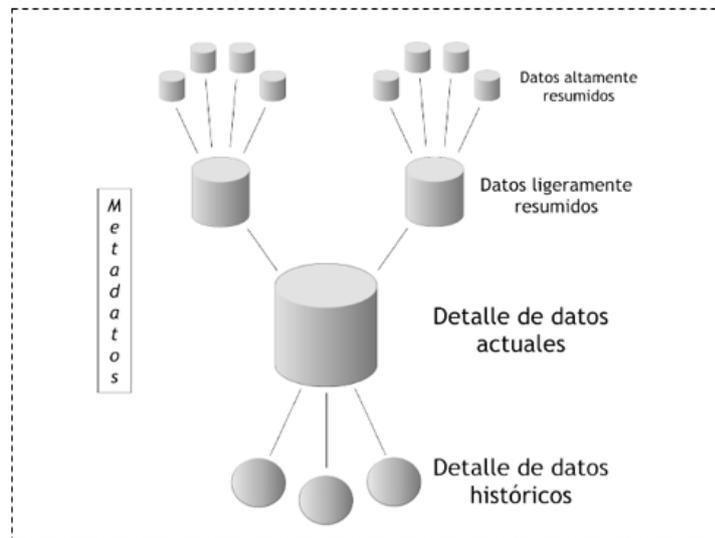


FIGURA 2.3. Almacén de datos, estructura.

Fuente: Hefesto

A continuación se explicarán cada uno de los tipos de datos:

- Detalle de datos actuales.

Reflejan los eventos más recientes. Los datos insertados regularmente son datos surgidos de los sistemas de producción. Los datos detallados normalmente pueden ser muy voluminosos y se necesitan de herramientas sofisticadas para gestionarlos y tratarlos. Cabe resaltar que el nivel de detalle almacenado en el Almacén de datos no es necesariamente idéntico al nivel de detalle gestionado en los sistemas operacionales. El dato insertado en el Almacén de datos puede ser un resumen de información obtenida del sistema de producción. Ahora bien, cuanto más fino es el nivel de detalle, es más posible segmentar estos datos dinámicamente. Hay que recordar que no porque un dato este disponible debe integrarse a este nivel. Sólo se almacenan los datos de detalle útiles y necesarios. Por ejemplo, aquí es donde se guardaría el detalle de una venta realizada en tal fecha.

- Datos resumidos.

Se utilizan con frecuencia ya que son los elementos de análisis que representan las necesidades de los usuarios. Representan ya un resultado de análisis y una síntesis de la información contenida en el sistema de decisión. Es por ello deben ser accesibles y comprensibles. Son los que provienen desde un bajo nivel de detalle y sumariza o agrupan los datos bajo algún criterio o condición de análisis. Por ejemplo, en este caso se almacenaría la sumarización del detalle de las ventas realizadas en cada mes.

- Datos altamente resumidos

son aquellos que representan el siguiente nivel de datos encontrado en el almacén de datos. Estos datos son compactos y de fácil acceso. Es decir, son aquellos que compactan aún más a los datos ligeramente resumidos. Por ejemplo, aquí se encontraría la sumarización de las ventas realizadas en cada año.

- Metadatos

Representan la información acerca de los datos. De muchas maneras se sitúa en una dimensión diferente al

de los otros datos del almacén de datos, ya que su contenido no es tomado directamente desde el ambiente operacional. El metadato agrupa toda la información respecto al almacén de datos y los procesos asociados.

1. Al usuario
 2. A los equipos responsables de los procesos de transformación de los datos del entorno de producción hacia el almacén de datos.
 3. A los equipos responsables de los procesos de creación de los datos resumidos a partir de los datos detallados.
 4. A los equipos de administración de la base de datos. Información sobre la estructura de la base que implementa el almacén de datos.
 5. A los equipos de producción. Información sobre los procedimientos de carga, el historial de actualizaciones, etc.
- **Detalle de datos históricos**

Es uno de los objetivos principales del almacén de datos. Representa a los datos antiguos, que no son frecuentemente consultados. Estos datos se almacenan a nivel de detalle, normalmente sobre alguna forma de almacenamiento externa, ya que son muy pesados y en adición a esto, no son requeridos con mucha periodicidad. Por ejemplo, en este nivel, al igual que en el anterior, se encontraría el detalle de una venta realizada en tal fecha, pero con la particularidad de que el día en que se registró la venta debe ser lo suficientemente antigua, para que se considere como histórica.

2.3.8 Uso del almacén de datos

En diferentes casos puede ser utilizado el almacén de datos entre los principales se mencionan los siguientes: se tienen consultas complejas. Existen grandes cantidades de datos a los cuales deben ingresar los usuarios

2.3.9 Arquitecturas del almacén de datos

El almacén de datos y sus arquitecturas varían dependiendo las especificaciones de la situación de la organización. Tres arquitecturas comunes son:

- Arquitectura de almacén de datos (Básico)
- Arquitectura de almacén de datos (Con un área de ensayo)
- Arquitectura de almacén de datos (Con un área de ensayo y un almacén departamental)

Arquitectura de almacén de datos (Básico) En la figura 2.4 se muestra una arquitectura simple de almacén de datos. Los usuarios finales acceden directamente a los datos derivados de diversos sistemas fuente a través del almacén de datos.

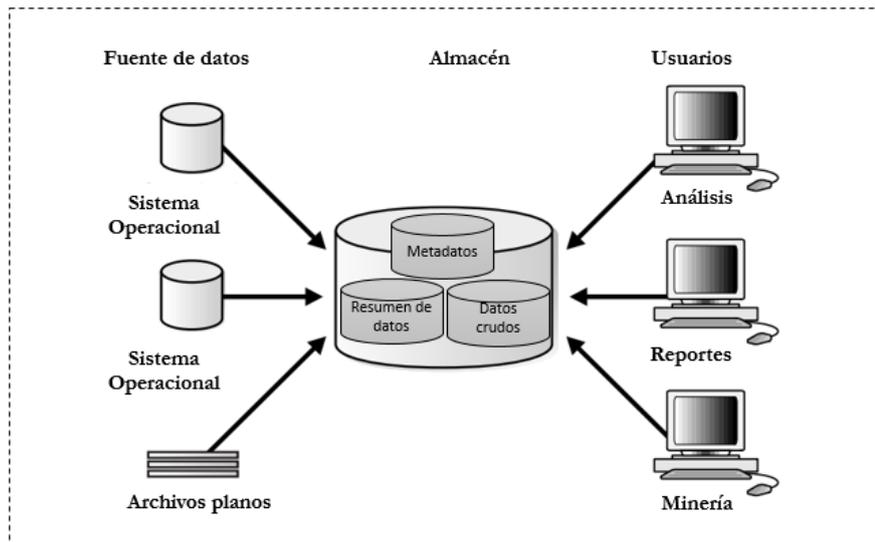


FIGURA 2.4. Arquitectura de almacén de datos básica. .

En la figura 2.4 los metadatos y los datos crudos o sin formato de un sistema OLTP tradicional están presentes como un tipo adicional de datos y resumen de datos. Los resúmenes son muy valiosos en los almacenes de datos, porque previenen largas operaciones de antemano. Por ejemplo, una consulta típica del almacén de datos es recuperar algo así como las ventas de agosto.

La figura representa que las fuentes de datos envían directamente los datos hacia al almacén de datos. Finalmente los usuarios acceden al almacén de datos para generar análisis, reportes y minaría de los datos.

Arquitectura de almacén de datos (Con área de ensayo)

En la figura 2.5 se necesita antes de poner los datos en el almacén de datos limpiarlos y procesarlos. Ésto se puede hacer automaticamente aunque la mayoría de los almacenes de datos usan una área de ensayo. Una área de ensayo simplifica la construcción resúmenes y gestión general del almacén. En la figura 2.5 se ilustra la típica arquitectura.

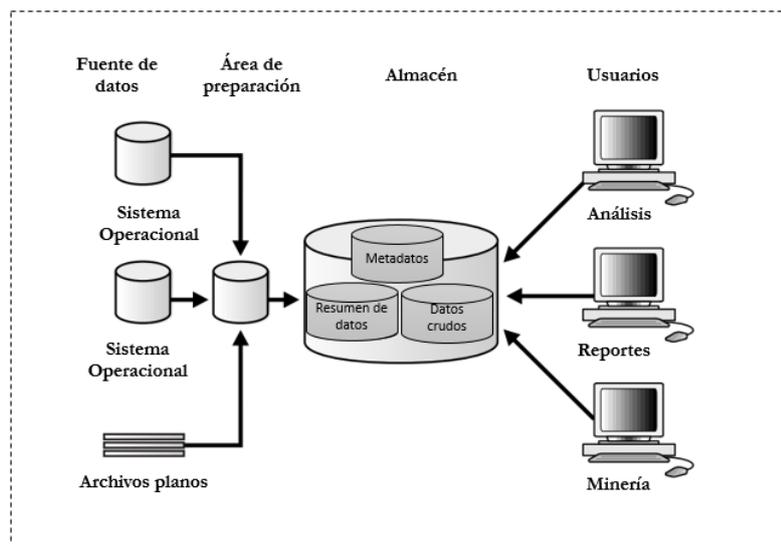


FIGURA 2.5. Arquitectura de almacén de datos con área de preparación.

La figura muestra que los datos extraídos de las fuentes de datos son enviados hacia una base de datos de ensayo en donde se llevan a cabo el proceso de limpieza de datos antes de ser enviados al almacén de datos. Posteriormente la base de datos de ensayo carga los datos al almacén de datos. Finalmente los usuarios acceden al almacén de datos para generar análisis, reportes y minería de los datos.

Arquitectura de almacén de datos (Con área de ensayo y un almacén departamental)

A pesar de que la arquitectura en la figura 2.5, es bastante común es posible que desee personalizar la arquitectura de su almacén para diferentes grupos dentro de su organización. Se puede hacer esto agregando mercados de datos que son sistemas diseñados para una línea en particular de negocios. La figura 2.6 ilustra un ejemplo donde las compras, las ventas y los inventarios están separados. En este ejemplo, un analista financiero podrá analizar datos históricos para compras y ventas.

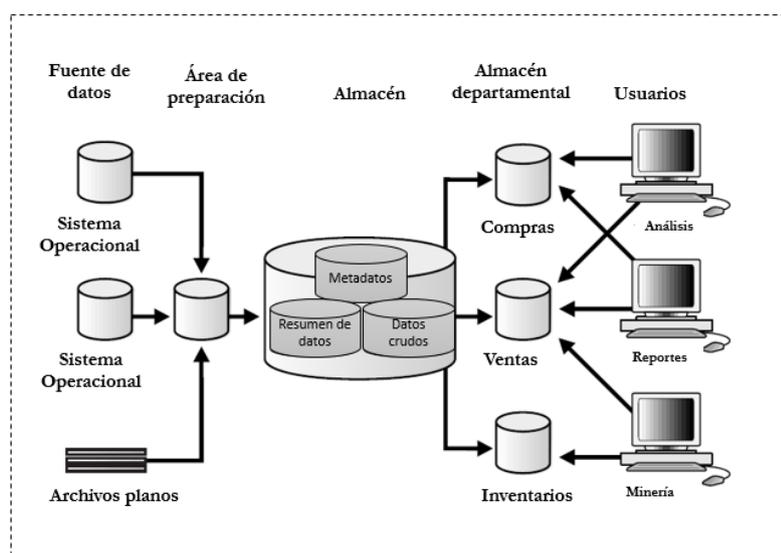


FIGURA 2.6. Arquitectura con área de ensayo y con almacén departamental.

Es importante recordar que los almacenes departamentales son una parte importante de muchos almacenes de datos. La figura ilustra que los datos extraídos de las fuentes de datos son cargados a una área de preparación o de ensayo, después son cargados al almacén de datos, la acción adicional en esta arquitectura se refiere a que una vez que son cargados los datos al almacén se pueden generar almacenes departamentales que permiten contener datos de una área en específico, ejemplo: datos exclusivamente de ventas. Finalmente los usuarios acceden a almacén para generar análisis, reportes y minería de los datos.

2.3.10 Base de datos multidimensional

Son aquellas bases de datos que proveen una estructura que permite tener acceso flexible a los datos, para explorar y analizar sus relaciones, y resultados. Cada dimensión del cubo corresponde a un criterio de análisis: tipo de parámetro, sitio, mes. las celdas del cubo se llaman medidas que almacenan las medidas promedio para cada tupla de (parámetro, sitio, mes) (L. Berrahou , 2015). Las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa la media, indicador o el hecho que se está evaluando.

En la figura número 2.7 se puede ver más claramente lo que se acaba de decir. En los almacenes de datos, los criterios de análisis están estructurados en jerarquías. También se resalta en la figura un hecho seleccionado para demostrar su correspondencia con los valores de las dimensiones. Como se puede observar de igual manera en la figura 2.7 la media a la que se hace referencia representa el cruce del valor "9" de "Dimensión 1", con el valor "6" de la "Dimensión 2" y con el valor "12" de la "Dimensión 3". Es así como el resultado del análisis está dado por los cruces matriciales de acuerdo con los valores de las dimensiones seleccionadas como se presenta en la figura 2.7.

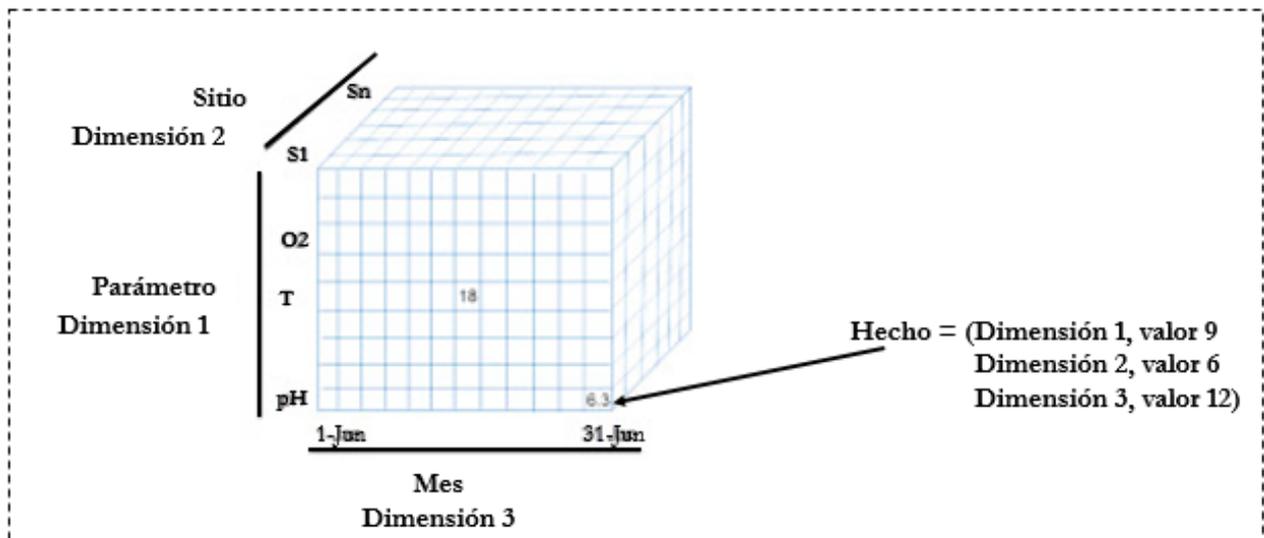


FIGURA 2.7. Cubo Multidimensional.

Por lo tanto, en la figura anterior se puede describir que existen tres dimensiones, "Dimensión 1 (parámetro)", "Dimensión 2 (sitio)", "Dimensión 3 (mes)", cada una con sus respectivos valores asociados.

2.3.11 Tablas de Dimensiones

Son la representación de como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Sin duda son las que representan los ejes del cubo y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos. Cada tabla posee un identificador único y al menos un atributo que describe los criterios de análisis relevantes de la empresa. Usualmente la cantidad de tablas de dimensiones, aplicadas en un tema de interés en particular, varían entre tres y quince, (Ing. Bernabeu, 2007).

2.3.12 Tablas de Hechos

La tabla de hechos está contenida por los hechos, medidas o indicadores que serán utilizados por los usuarios del negocio para dar soporte a la toma de decisiones. Por lo tanto, los hechos son datos filtrados, agrupados y explorados a través de las tablas de dimensiones. Los datos que están presentes en las tablas de hechos son los que constituyen el volumen del almacén. La tabla de hechos posee una clave primaria que está compuesta por claves primarias de las tablas de dimensiones relacionadas a este.

2.3.13 Métricas

Son los indicadores de negocio de un proceso de negocio. Son aquellos conceptos cuantificables que permiten medir el proceso del negocio.

2.4 Modelos de almacén de datos

Las bases de datos implican tres modelos, que permiten realizar consultas de soporte de decisión:

2.4.1 Modelo Estrella

El modelo estrella se caracteriza por tener una tabla de hechos y una serie de dimensiones en profundidad simple, esto quiere decir que no tiene relación con otras tablas más allá de la misma dimensión. Tiene como ventaja mayor velocidad de procesamiento, al prescindir de operaciones "join" para relacionar al resto de las tablas. Por otro lado, su desventaja es el almacenamiento por la desnormalización, pero ésto es poco crítico como se representa en la figura 2.8.

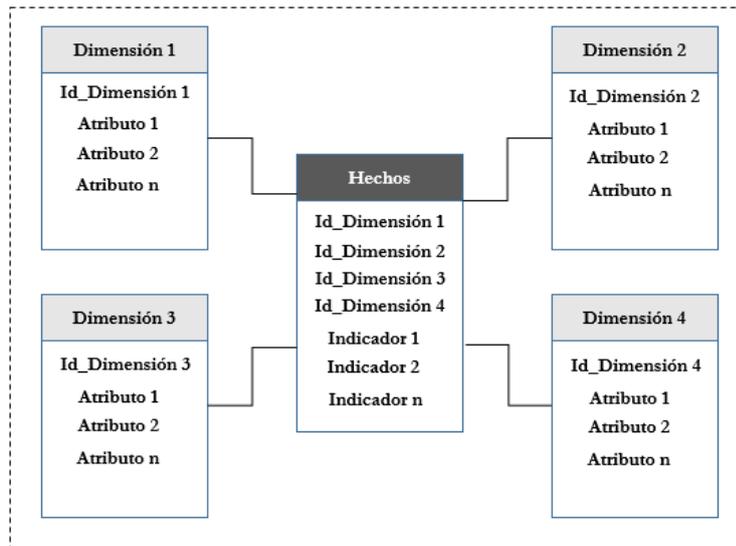


FIGURA 2.8. Modelo Estrella.

Fuente: Diseño propio

La figura permite apreciar que el modelo estrella está constituido en el centro por una tabla hechos y a su entorno por cuatro tablas dimensiones dentro de las cuales están conformadas por los atributos que daran respuesta a los indicadores contenidos en la tabla central hechos.

2.4.2 Modelo Copo de Nieve

El modelo copo de nieve mantiene relación de sus dimensiones con otras tablas, con el fin de clarificar lógicamente las relaciones jerárquicas. Su gran desventaja es el performance, dado por los numerosos "joins" necesarios para las consultas. Su ventaja es el espacio en disco, dada la normalización de datos, se presenta a continuación en la figura 2.9.

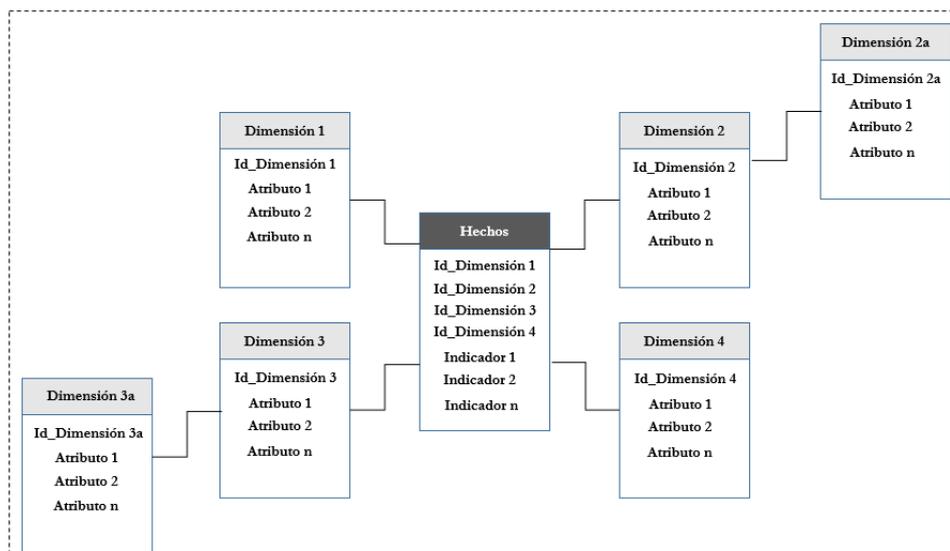


FIGURA 2.9. Modelo Copo de Nieve.

Fuente: Diseño propio

La figura ilustra al centro una tabla hechos que está conformada por dos secciones, la primera contiene su clave primaria conformada por el id (identificador) que representa su clave primaria de cada una de las tablas que tiene a su alrededor; la segunda sección contiene los indicadores que son las métricas de la organización. Al rededor de la tabla hechos se encuentran las dimensiones que también están conformadas por dos secciones la primera es su clave primaria y la segunda son los atributos que dan respuesta a los indicadores contenidos en la tabla hechos. Las tablas dimensiones tienen tablas hijas debido a la normalización que permite éste tipo de modelo.

2.4.3 Modelo Constelación

Este modelo representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones como se muestra en la figura 2.10.

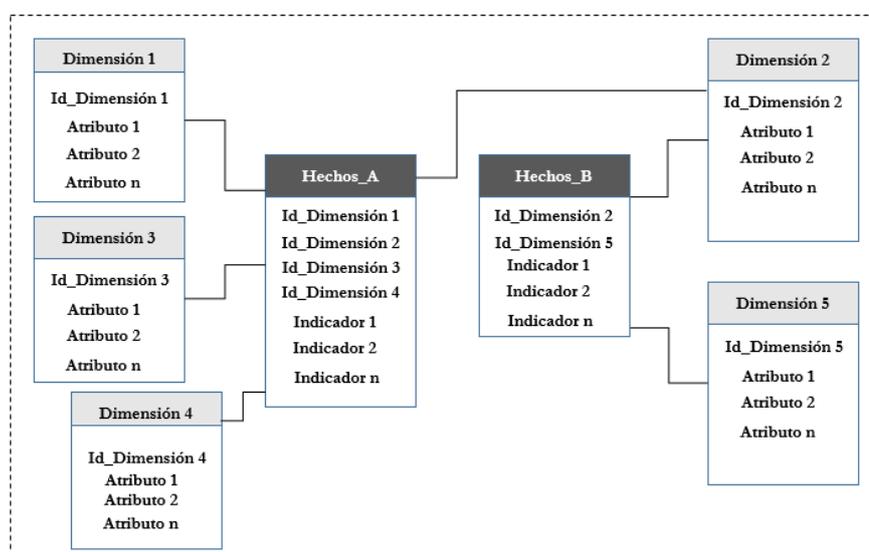


FIGURA 2.10. Modelo Constelación.

Fuente: Diseño propio

El modelo constelación como se observa en la figura 2.10 son dos a más tablas hechos y en su alrededor tiene tablas dimensiones, la diferencia a los modelos anteriores que es este esquema las tablas hechos pueden tener acceso a las dimensiones de la tabla hechos vecina.

2.5 Metodologías existentes

Existen diversas metodologías que están enfocadas a la inteligencia de negocios para el desarrollo de almacenes de dtos. La siguiente información se hace con el fin de tener una referencia sobre las metodologías existentes, entre las que se pueden mencionar:

2.5.1 CRISP-DM

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) fue desarrollada en términos de un modelo de proceso jerárquico, que consiste en conjunto de tareas descritas en cuatro niveles que van de lo general a lo particular: fases, tareas genéricas, tareas especializadas e instancias como se representa en la figura 2.11 que a continuación se presenta.

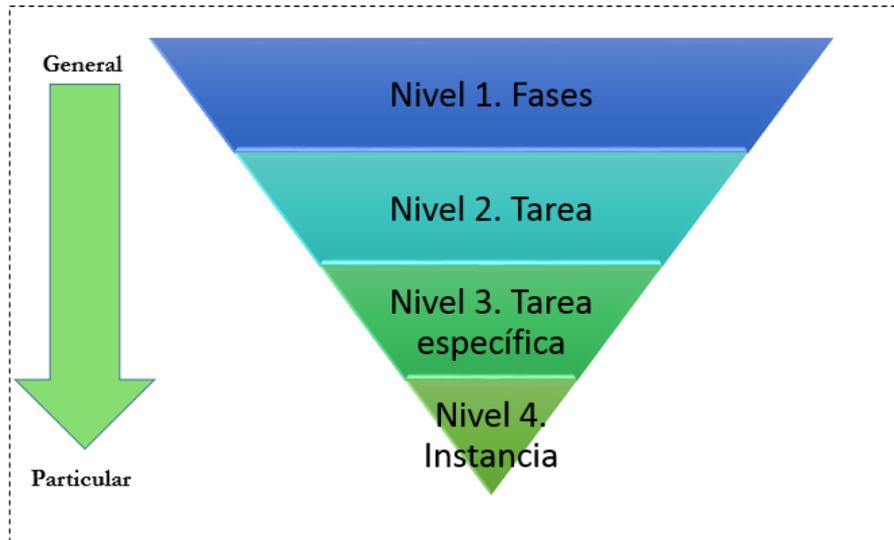


FIGURA 2.11. Proceso Jerárquico CRISP-DM.

Fuente: Diseño propio

En la minería de datos los proyectos siguen un ciclo de vida. Este contiene las fases de un proyecto, sus tareas respectivas y las relaciones entre estas tareas, sin embargo, las relaciones entre estos elementos pueden variar de acuerdo con los objetivos y necesidades que pueda presentar cada proyecto. En la figura 2.12 se simboliza la naturaleza cíclica de la metodología Crisp-DM.

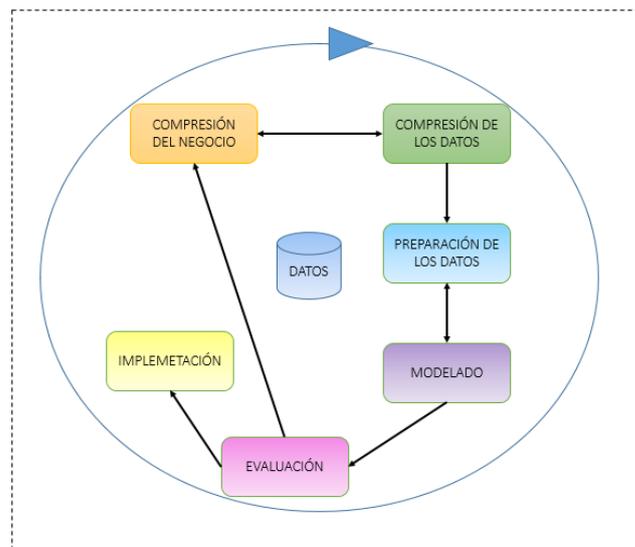


FIGURA 2.12. Ciclo de vida de CRISP-DM.

Fuente: Diseño propio

- Compresión del negocio

Lleva a cabo la comprensión de los objetivos del proyecto y exigencias desde una perspectiva de negocio. Es decir, entender que es lo que realmente quiere el cliente, descubrir factores importantes que puedan influenciar en el proyecto. En seguida evalúa la situación por ejemplo los recursos, restricciones, supuestos y otros factores importantes para cubrir el objetivo del análisis de datos y el plan del proyecto. Finalmente determina los objetivos de minería de datos, que se refiere a determinar los objetivos del proyecto de manera técnica.

- Compresión de los datos

Comienza con la colección de datos inicial y continua con las actividades que le permiten familiarizar primero con los datos, identificar los problemas de calidad de datos y explotar los datos.

- Preparación de los datos

Cubre todas las actividades necesarias para construir el conjunto de datos final (datos que serán vistos en el modelado). La selección de datos que podrán ser usados en el análisis de acuerdo con su relevancia para cumplir con los objetivos de la minería de datos. La limpieza de datos se lleva a cabo para elevar la calidad de los datos. Construcción de datos es la construcción de atributos derivados y registrados. Integración de los datos es la combinación de las tablas o registros. Y finalmente reformatean los datos que no es más que las modificaciones sintácticas que no modifican su significado.

- Modelado

Se selecciona una técnica de modelado. Evaluar el modelo de acuerdo a los criterios de éxito de minería de datos determinados en las fases anteriores.

- Desarrollo

La planeación de la implementación es la estrategia para llevarla a cabo. El plan de monitoreo y mantenimiento es la estrategia para monitorear y mantener la solución de la minería de datos. Producción del reporte final que se refiere al documento que resume los resultados y la presentación de los mismos. Revisión del proyecto son las lecciones aprendidas.

2.5.2 Hefesto

Es una metodología creada por el Ing. Bernabeu Ricardo Darío su última actualización es la versión 1.1 en abril de 2009 y disponible bajo la licencia GNU FDL. Es una metodología propia para la construcción de un almacén de datos, permite la construcción de un almacén de datos de forma sencilla, ordenada e intuitiva. Consta de cuatro fases: análisis de requerimientos, análisis de OLTP, modelo lógico del almacén de datos y proceso ETL. Puede ser utilizada en cualquier ciclo de vida que no involucre fases extensas de requerimientos. Lo que se busca es entregar una primera implementación que satisfaga las necesidades, para demostrar la ventajas del almacén de datos y motivar a los usuarios [9].

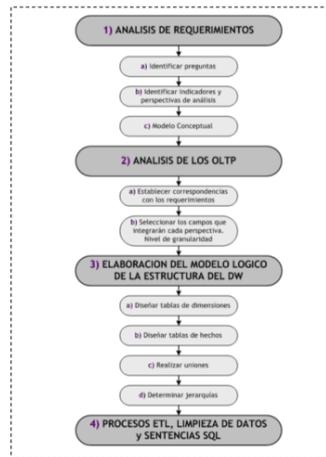


FIGURA 2.13. Pasos de la metodología Hefesto.

Fuente: Hefesto

A continuación se describen las cuatro fases:

La fase 1 Análisis de requerimientos, se refiere a obtener de los usuarios los requerimientos que permitirán definir los indicadores que le permitan medir el rendimiento de la empresa para la toma de decisiones y al cumplimiento de sus objetivos. En esta fase se realizan diferentes tareas una de ellas es para conocer las necesidades de información del usuario para alcanzar sus objetivos. La siguiente tarea es analizar lo obtenido de la tarea de conocimiento de las necesidades de información para identificar los indicadores y las dimensiones que le dan sentido a los indicadores. La tarea siguiente es crear un modelo conceptual que permita visualizar los hallazgos mencionados a continuación: “unidades vendidas de cada producto a cada cliente en un tiempo determinado”. En esta metodología se hace el uso de rectángulos y elipses, donde los rectángulos a la izquierda representan las dimensiones, la elipse indicará el proceso en cuestión y los rectángulos a la derecha representarán los indicadores como se muestra en la figura 2.13.

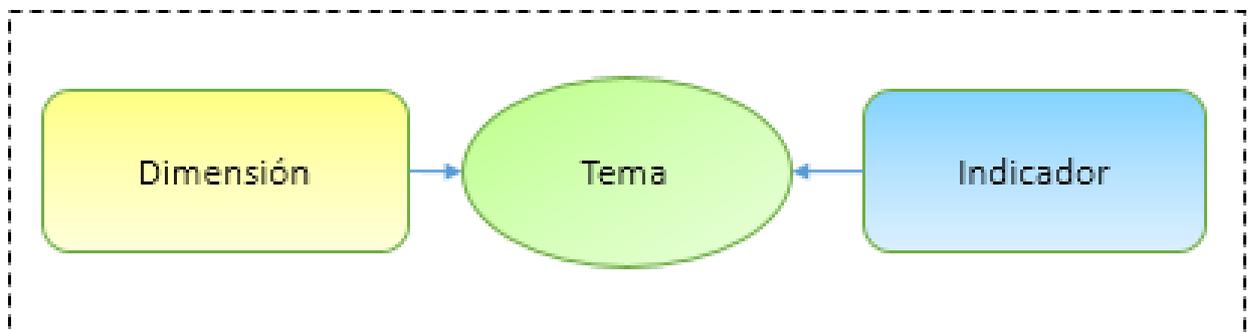


FIGURA 2.14. Rectángulos y elipse que representan el análisis de requerimientos.

Fuente: Diseño propio

La fase 2 Análisis de los OLTP, en esta etapa se determinan los indicadores, mediante definición de formulas y uso de funciones de sumarización que se utilizará para su agregación. La segunda actividad en esta fase es establecer correspondencias, ésta se refiere a realizar un mapeo de las fuentes de datos que proporciona la organización y el modelo conceptual realizado a partir de las necesidades de información. El objetivo es, que todos los elementos del modelo conceptual estén correspondidos a los OLTP. Al finalizar el mapeo en necesario realizar la siguiente tarea que es a nivel de granularidad y consiste más que nada en definir los campos que formaran parte de las dimensiones, para dicha tarea hay que apoyarse del usuario ya que el será el que defina los campos y podrá proporcionar una explicación de dichos campos, de esta manera se crea el diccionario de datos o en su caso apoyarse de alguno ya existente.

La fase 3 Modelo lógico del almacén de datos, para su construcción se realizan diferentes actividades, primero definir el modelo lógico, que bien puede ser estrella, copo de nieve y constelación. Cuando ya fue decidido el modelo lógico se avanza a la siguiente tarea o actividad que es el diseño de las tablas de dimensiones, ahora bien, tomando en cuenta el nombre de la tabla, claves primarias, nombres de campos. Posteriormente definen las tablas hechos que son las que contendrán los valores necesarios para determinar los indicadores, estas tablas tienen nombre muy específicos y representativos, se verifican las claves primarias. La última actividad finalmente consiste en hacer las relaciones indicadas en el modelo conceptual o dicho en otras palabras relacionar dimensiones y hechos.

La fase 4 Integración de los datos, ésta última actividad de esta fase consiste en poblar el almacén de datos, el cual se debe construir después del modelo lógico usándolo como guía para la construcción. La carga inicial es la primera actividad o tarea, es por medio de ETL el que permitirá garantizar la calidad de los datos. Para realizar una correcta carga primero se poblarán las dimensiones siguiendo su orden de jerarquía (de lo general a lo particular) y finalmente se cargaran las tablas hechos.

Cuando ya se obtuvo la primera carga, se procede a determinar la arquitectura la cual para actualizar el almacén de datos, para lo cual se debe continuar con la limpieza de datos.

2.5.3 Metodología ciclo de vida de Kimball

En la figura 2.15 se muestra el marco de trabajo de la metodología de Kimball, la cual se ilustran diferentes etapas durante el proceso de creación del almacén de datos. Hay ocho pasos que necesitan ser seguidos. Comenzando en encontrar definiciones de problemas, hacer estudios de literatura, exponer la metodología, analizar, diseñar e implementar el almacén de datos, presentación de informes y finalización de la conclusión.

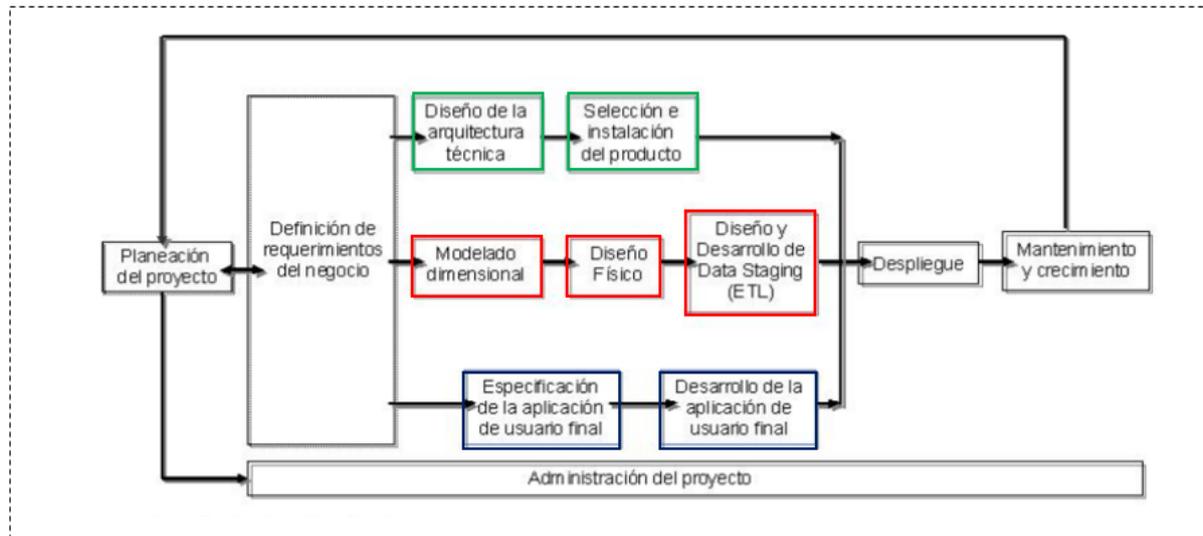


FIGURA 2.15. Pasos de la metodología Kimball.

Fuente: Kimball

La fase 1 planificación de proyectos, se busca la definición y el alcance del proyecto del almacén de datos, justificaciones del negocio y evaluaciones de factibilidad. Se enfoca en los recursos, perfiles, tareas, duraciones y secuencia.

El plan de proyecto que resulta identifica todas las tarea asociadas con el ciclo de vida dimensional del negocio e identifica las partes involucradas. La planeación del proyecto es independiente de los requerimientos del negocio tal y como se muestra en la figura 2.15 y esto se debe a que los requerimientos del negocio determinan el alcance del proyecto, define los recursos necesarios, entre otros. Por otro lado la planeación acota los requerimientos por recursos o por tiempo.

La fase 2 definición de los requerimientos, para que un almacén de datos tenga éxito es necesario tener un factor determinante que se refiere a la interpretación correcta de los diferentes niveles de requerimientos que se expresan por los diferentes niveles de usuarios. Los diseñadores de almacén de datos deben entender factores clave que guían al negocio para determinar efectivamente los requerimientos y traducirlos a consideraciones de diseño apropiadas.

Esta etapa es la base para las tres etapas subsecuentes que se enfocan a la tecnología, es decir, la herramienta de inteligencia de negocios que usara el usuario final; los datos, es decir, el modelado dimensional; y las aplicaciones, esto es altamente crítico y es el centro de atención del ciclo de vida dimensional el negocio. Los requerimientos del negocios son los que deben determinar el alcance del almacén de datos, es decir, los datos que debe contener, como debe ser organizado, cada cuando debe actualizarse, desde cuando accederan, y más.

La fase 3 modelado dimesional de negocio, organiza y presenta los datos definiendo dimensiones, es decir, líneas o áreas temáticas del negocio. Se puede mencionar como ejemplo que los modelos de las empresas es muy común encontrar las mediciones estructuradas por dimensiones como sucursal, producto y tiempo. Es de esta manera como se permite analizar la información a distintos niveles de agregación dentro de las diferentes dimensiones. En una dimensión se pueden definir los niveles de agregación o sumas de cada análisis a los cuales

se les da el nombre de atributos.

La fase 4 diseño físico, el diseño físico de las bases de datos se centra sobre la selección de las estructuras necesarias para soportar el modelado dimensional.

La fase 5 diseño y desarrollo de la área de ensayo, las principales sub-etapas de esta zona del ciclo de vida son: la extracción la transformación y la carga (ETL process). Los procesos de extracción son aquellos requeridos para obtener los datos que permitan efectuar la carga del modelo físico acordado. Los procesos de carga de datos son los procesos requeridos para poblar el almacén de datos.

La calidad de los datos es un factor determinante en el éxito de un proyecto de almacén de datos. Es una etapa en donde deben depurarse todos los inconvenientes relacionados con la calidad de los datos fuente.

La fase 6 diseño de la arquitectura técnica, para los ambientes de almacén de datos requieren de la integración de numerosas tecnologías, para ello se debe tener en cuenta tres factores importantes: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras; para poder establecer el diseño de la arquitectura técnica del ambiente del almacén de datos.

La fase 7 selección de productos e instalación, para este paso se hace uso de la arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura como la plataforma del hardware, el motor de la base de datos, la herramienta del ETL o el desarrollo pertinente, herramientas de acceso, entre otros, para ello se puede utilizar una matriz de evaluación de productos. Una vez evaluados y seleccionados los productos determinados se procede con la instalación y prueba de los mismos en un ambiente integrado del almacén de datos.

La fase 8 diseño de la aplicación de inteligencia de negocios, se identifican en esta etapa los diferentes roles o perfiles de usuarios para determinar los diferentes tipos de aplicaciones necesarias con base al alcance de los diferentes perfiles como son: gerencial, analista de negocio, vendedor entre otros.

La fase 9 desarrollo de aplicación de almacén de negocios, el desarrollo de las aplicaciones para los usuarios finales involucra configuraciones del metadata y construcción de reportes específicos. Cuando se ha concluido la especificación y se tiene posibilidad de trabajar con algunos datos de prueba se comienza el desarrollo de aplicación y se consideran los siguientes aspectos: i) selección de un enfoque de implementación, ii) desarrollo de aplicación, iii) prueba y verificación de datos, iv) documentación y Roll Out y finalmente v) mantenimiento.

La fase 10 implementación, representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio el usuario el negocio. Para el correcto funcionamiento de los elementos se requiere: la capacitación, el soporte técnico, la comunicación y las estrategias de retroalimentación. éstas son las consideraciones que se deben tener en cuenta antes de que cualquier usuario ingrese al almacén de datos.

La fase 11 mantenimiento y expansión, una vez finalizada la construcción e implementación del almacén de datos se debe estar preparado para administrar el mantenimiento y la expansión del mismo. Existen puntos a tomar en cuenta para mantener exitosamente el almacén de datos. Entre ellos se destacan: el continuo soporte y la constante capacitación a usuarios de negocios, manejo de la infraestructura (monitoreo de la base de datos, tráfico, entre otros) sintonización de rendimiento sobre las consultas, mantenimiento del metadata y procesos de ETL.

La fase 12 administración del proyecto, la administración del proyecto asegura que las actividades del ciclo de vida dimensional del negocio se lleven en forma y sincronización. La figura 2.15 indica como la administración acompaña todo el ciclo de vida. Entre sus actividades principales se encuentran: el monitoreo del estado del

proyecto y la comunicación entre los requerimientos del negocio, así como las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos.

La fase 13 puntos problemáticos de un almacén de datos, los principales puntos que pueden llegar a complicar un proyecto de almacén de datos se localizan en tres áreas: rutinas de carga, mantenimiento y en la sintonización.

Existen tres rutas principales en esta metodología:

Tabla 2.2: Rutas principales de la metodología de Kimball

No.	Rutas
1	Arquitectura y tecnología.
2	Datos.
3	Aplicaciones del usuario.

En la fase de planeación se establece la definición y alcance del proyecto del almacén de datos, incluye la operación y justificación del negocio. En la fase de definición del proyecto es donde se establece la base relacionada con la tecnología los datos y las aplicaciones de los usuarios. La ruta de mayor importancia es la de los datos, en ella se realiza el modelado dimensional, a partir de los requerimientos determinados y de las necesidades de los usuarios; el diseño físico su objetivo en definir las estructuras físicas para soportar el modelado dimensional; y en la etapa de ETL, en ella se diseña y desarrollan procesos para extraer, transformar y cargar datos. La tecnología es otra ruta que se debe considerar para establecer la arquitectura y visión general del marco de trabajo y la ruta de aplicación, en ella se definen un conjunto de aplicaciones de usuario final para consulta de datos.

Entonces a lo largo de ciclo de vida se debe seguir una administración general del proyecto para asegurar que los requerimientos del ciclo de vida se almacenen y se sincronicen [16].

2.5.4 Inmon

El enfoque que Bill Inmon considera para un almacén de datos es que debe responder a las necesidades de todos los usuarios en la organización, y no solo en un determinado grupo. Él menciona que el almacén de datos debe cumplir con las siguientes características:

- Dirigido a un área. Datos sobre un área específica en lugar de operaciones de la compañía.
- Integrado. Unión de diferentes fuentes de datos de manera coherente
- Variable en el tiempo. Todos los datos pertenecen a un periodo de tiempo determinado
- No volátil. Los datos no son eliminados

Inmon propone que su metodología debe ser iterativa la cual sigue un esquema contrario al clásico desarrollo de sistemas ya que en primer lugar con lo que trabaja es con datos, estos se integran para ser probados y programar de acuerdo a ellos para analizar los resultados y así comprender los requerimientos. La metodología consiste en la siguiente como se muestra en la figura 2.15.



FIGURA 2.16. Pasos de la metodología Inmon.

Fuente: Diseño propio

2.5.5 Rapid Warehousing Methodology (RWN)

Es una metodología iterativa y está basada en el desarrollo incremental de la creación del almacén de datos y se encuentra dividido en seis fases como se aprecia en (Ing. Eric Ismael Leonard Brizuela, 2013).

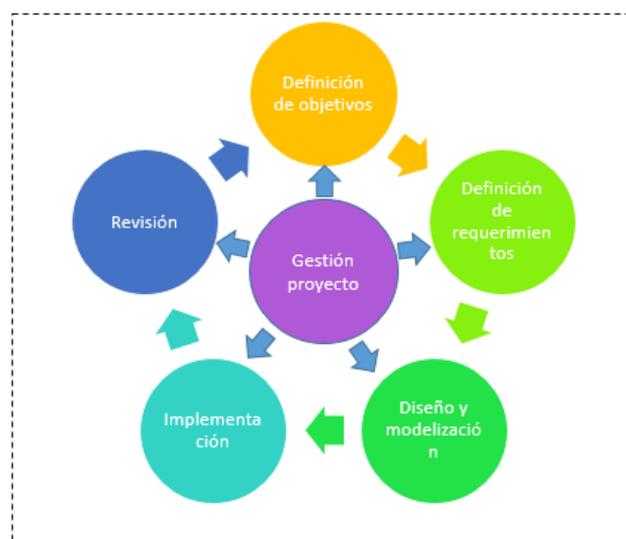


FIGURA 2.17. Metodología de almacenamiento rápido.

Fuente: Diseño propio

En la figura 2.16 se ilustra las fases de la metodología de almacenamiento rápido. Se detallarán a continuación cada una de ellas.

- **Definición de los objetivos:** se define al equipo del proyecto, el alcance del sistema y cuales serán las funciones que el almacén de datos realizará. De igual manera se definirán los parámetros por medio de los cuales se evaluarán el éxito del proyecto.
- **Diseño y modelización:** en esta etapa se identifican las fuentes de datos y se determinan las transformaciones necesarias, para que a partir de las fuentes definidas o establecidas, se obtenga el modeo lógico de datos del almacén de datos. El modelo se conforma por tablas dimensiones y tablas hechos que son las que permiten resolver las necesidades del negocio.
- **Implementación:** es la etapa donde se realiza la extracción y carga de datos, al igual que la explotación del almacén de datos por medio de los usuarios finales.

- Revisión: finalmente después de la implementación, se realiza una revisión del almacén de datos planteando preguntas que permitan, pasado unos 3 o 2 meses que fue puesto en marcha, marcar cuales son los aspectos a mejorar o potenciar en función a la utilización.

Cabe mencionar que esta metodología no incluye lo relativo a técnicas de análisis de la información, por lo que con su aplicación solo se obtendría el almacén de datos y no los multianálisis de los datos para apoyar la toma de decisión.

2.5.6 Comparativa de metodologías

Las metodologías descritas anteriormente son las metodologías más habituales que se emplean para el desarrollo del almacén de datos, como se muestra en la tabla 2.3, se puede seleccionar cual de ellas es la apropiada para la empresa teniendo como punto de partida el tipo de empresa a la que se aplica.

Tabla 2.3: Comparación de metodologías

Metodología	CRISP-DM	Hefesto	Kimball	Inmon	Rapid Warehousing Methodology (RWN)
Enfoque	Minería de datos	Almacén de datos	Almacén de datos	Almacén de datos	Almacén de datos
Características	Visión en el negocio	Construcción del almacén de datos	Diseño del almacén de datos	Conceptos de almacén de datos	Desarrollo incremental de un almacén de datos
Estructura	Comprensión comercial, comprensión de datos preparación de datos, modelado y evaluación	Análisis de requerimientos, análisis OLTP, modelo lógico del almacén de datos e integración de datos	Planeación del proyecto, requerimientos del negocio, diseño arquitectura técnica, selección e instalación de herramientas, modelado dimensional, diseño físico, especificaciones herramienta analítica, desarrollo herramienta analítica, área de staging, desarrollo y mantenimiento	Implementa almacén de datos, integración de Datos y pruebas, programación sobre datos, diseño DSS, análisis resultados, entendimiento	Definición de los objetivos, definición de los requerimientos, diseño y modelización, implementación y revisión.

En la tabla superior se presentan las metodologías estudiadas, así como, se menciona cual es el enfoque que tiene cada una de ellas, se describe sus características, y su estructura.

2.6 Herramientas Utilizadas

Para construir el almacén de datos se utilizaron las herramientas MySQL, es decir, son programas de código abierto, modificado por los desarrolladores de aplicaciones. Se tomo la decisión de hacer uso de código abierto porque son instrumentos estables, flexibles y maduros para la construcción del almacén de datos.

Las herramientas que se emplean en este trabajo son:

MySQL Workbench. es una herramienta visual unificada para arquitectos de bases de datos, desarrolladores y DBA. MySQL Workbench proporciona modelado de datos, desarrollo de SQL y herramientas integrales de administración para la configuración del servidor, administración de usuarios, respaldo y mucho más. MySQL Workbench está disponible en Windows, Linux y Mac OS X. Más información en: <https://www.mysql.com/products/workbench/>



FIGURA 2.18. MySQL Workbench.

Fuente: MySQL Workbench

MySQL Workbench ahora proporciona una solución completa y fácil de usar para migrar Microsoft SQL Server, Microsoft Access, Sybase ASE, PostreSQL y otras tablas, objetos y datos de RDBMS a MySQL.

XAMPP. XAMPP es una distribución de Apache completamente gratuita y fácil de instalar que contiene MariaDB, PHP y Perl. El paquete de instalación de XAMPP ha sido diseñado para ser increíblemente fácil de instalar y usar. Para más información en: <https://www.apachefriends.org/es/index.html>



FIGURA 2.19. XAMPP.

Fuente: XAMPP

XAMPP tiene más de 10 años – hay una gran comunidad detrás del proyecto. Puede participar uniéndose a los foros, suscribiéndose a nuestra lista de correo, uniéndose a nuestro Facebook, siguiendo las hazañas en Twitter, o añadiéndose en tus círculos de Google+

Windows 8. La principal novedad de Windows 8 es la nueva pantalla de Inicio, que adopta la interfaz Modern UI, ya presente otros productos de Microsoft, como Xbox, Windows Phone y servicios en línea como Outlook.com. La nueva interfaz nace con el mundo táctil en mente, pero sin olvidar a los usuarios de teclado y ratón. Para más información en: <https://www.microsoft.com/es-mx/software-download/windows8ISO>



FIGURA 2.20. Windows 8.

Fuente: Windows 8

Si necesita instalar o reinstalar Windows 8.1, puede usar las herramientas de la página anteriormente mencionada para crear sus propios medios de instalación con una unidad flash USB o un DVD.

MODELO PROPUESTO

3.1 Introducción

En el presente Capítulo se describe la metodología empleada para desarrollar la construcción del almacén de datos mediante la metodología llamada Hefesto que está enfocada exclusivamente a almacenes de datos.

3.2 Hefesto: metodología propia para la construcción de un almacén de datos

EL Ing. Bernabeu presenta con el nombre de HEFESTO las diferentes etapas por las que debe pasar todo un proceso de almacén de datos que se presenta en la siguiente figura. En el diagrama se presenta las secuencias de las tareas para el diseño desarrollo e implementación del almacén de datos. Esto representa una vista general de la ruta de un proyecto, en el cual cada bloque indica en dónde se encuentra ubicado, por dónde se pasara y hacia donde dirigirse.

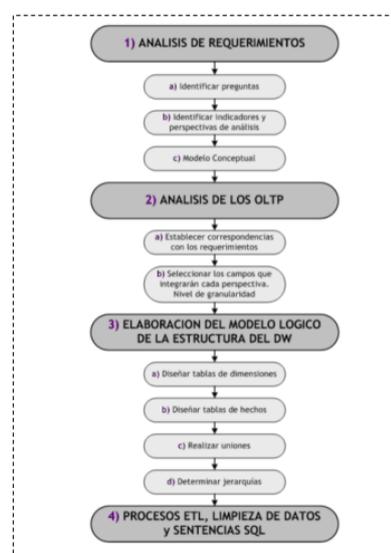


FIGURA 3.1. Pasos de la metodología HEFESTO

Fuente: Metodología Hefesto

Hefesto está fundamentada en una investigación muy amplia, comparación de metodologías existentes, experiencias propias en procesos de diseño de almacenes de datos. En primera instancia y como se aprecia, esta metodología da inicio con la recolección de las necesidades de la información de los usuarios y se obtienen las preguntas claves del negocio. Como consecuencia se identifican los indicadores resultantes de las entrevistas y sus perspectivas correspondientes, que dara paso a la construcción del modelo conceptual del almacén de datos.

3.2.1 Características

Las características con las que cuenta esta metodología son las siguientes:

- Objetivos y resultados esperados en cada fase son fáciles de comprender.
- Se basa en los requerimientos del usuario.
- Involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento.
- Utiliza modelos conceptuales y lógicos.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- No depende de las estructuras físicas que contenga el almacén de datos 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 mercado de datos y almacén de datos.

3.2.2 Empresa analizada

A las características principales de la empresa a la que se le ha de aplicar la metodología Hefesto, se le ha dedicado el Capítulo 4, particularmente para describirla detalladamente y así tener como base un ámbito predefinido y se comprenderá mejor cada decisión que se tome con respecto a la implementación y diseño del almacén de datos.

3.3 Paso 1) Análisis de requerimientos

Se identifican los requerimientos de los usuarios haciendo uso de preguntas que expliquen los objetivos de la organización. Se procede a analizar las preguntas a fin de identificar cuales seran los indicadores y las perspectivas que serán consideradas para la construcción del almacén de datos. Para finalizar se diseñará el modelo conceptual en donde se podra visualizar el resultado obtenido en este primer paso.

3.3.1 Identificar preguntas

Se comienza con la colección de las necesidades de la 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. El análisis de los requerimientos, es el punto de partida de

la metodología por medio de ellos se guiará la investigación hacia un desarrollo que refleje lo que se espera del almacén de datos en relación a sus funciones y cualidades. Como objetivo principal en esta fase es identificar necesidades de información para llevar a cabo las metas y estrategias de la empresa, y que facilitará la toma de decisiones.

3.3.2 Identificar indicadores y perspectivas

Cuando se concluyo el paso de establecer las preguntas de negocio, el siguiente paso es descomponerlas para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

3.3.3 Modelo conceptual

Se construye un modelo conceptual a partir de indicadores y perspectivas obtenidas en el paso anterior. Modelo conceptual: es la descripción de alto nivel de la estructura de la base de datos, donde la información se representa a través de objetos, relaciones y atributos.

Esté modelo permitirá observar con claridad ver los alcances del proyecto, y posteriormente trabajar con ellos, no solo posee alto nivel en la definición de los datos, sino que también permite que pueda ser presentado a los usuarios y explicado con facilidad.

La presentación gráfica del modelado conceptual de los datos será de la siguiente manera:

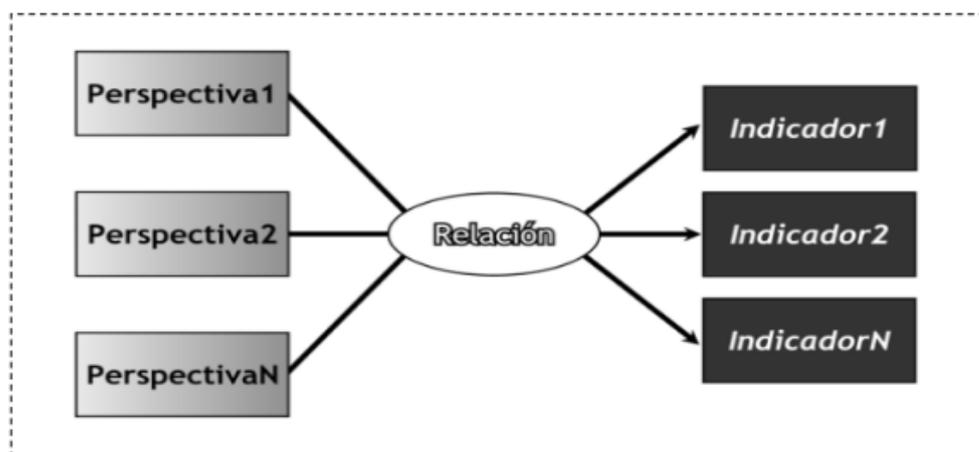


FIGURA 3.2. Modelado conceptual.

Fuente: Hefesto

La figura 3.2 se visualiza que a partir de las preguntas descompuestas se obtiene por el lado izquierdo las perspectivas y la relación que existe con los indicadores que intervendrán en el modelado.

3.4 Paso 2) Análisis de los OLTP

Se analizarán las fuentes para determinar como serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Seguidamente

se definirán que campos se incluirán en cada perspectiva. Finalmente se ampliará el modelo conceptual con la información obtenida en este paso.

3.4.1 Conformar indicadores

En este paso se deberán explicar como se calcularan los indicadores, definiendo los siguientes conceptos para cada uno de ellos:

- Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1 + Hecho2.
- Función de sumarización que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.

3.4.2 Establecer correspondencias

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

3.4.3 Nivel de granularidad

Cuando se establecieron las relaciones con los OLTP, se deben seleccionar los campos que contendrá cada perspectiva, ya que es a través de ellos por los que se examinarán y filtrarán los indicadores. Es muy importante tener en claro que significa cada campo y el valor de los datos encontrados en el OLTP, por lo cual es importante investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc.

Luego de exponer frente a los usuarios los datos existentes, explicando su significado, valores posibles y características, estos deben decidir cuales son los que consideran relevantes para consultar los indicadores y cuales no. Con la perspectiva "tiempo" se debe tener especial cuidado y es muy importante definir el ámbito mediante el cual se agruparán o sumarizarán los datos. Sus campos posibles pueden ser: día de la semana, quincena, mes, trimestres, semestre, año, etc.

Al momento de seleccionar los campos que integrarán cada perspectiva, debe prestarse mucha atención, ya que esta acción determinará la granularidad de la información encontrada en el almacén de datos.

3.4.4 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. Gráficamente:

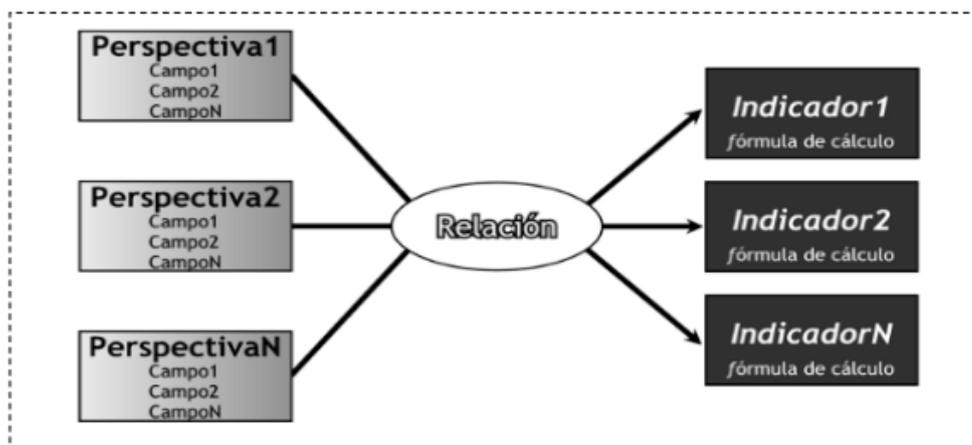


FIGURA 3.3. Modelo Conceptual ampliado.

Fuente: Hefesto

En la figura 3.3 se muestran que de cada perspectiva se colocaron los campos seleccionados y debajo de cada indicador se colocó la fórmula de cada indicador.

3.5 Paso 3) Modelo lógico del almacén de datos

En este paso se confeccionará el modelo lógico de la estructura del almacén de datos 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. Modelo lógico: representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.

3.5.1 Tipo de Modelo Lógico del almacén de datos

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.

3.5.2 Tablas de dimensiones

En este paso se deben diseñar las tablas de dimensiones que formarán parte del almacén de datos.

Para los tres tipos de esquemas, cada perspectiva definida en 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.

Se muestra la representación gráficamente en la siguiente figura:

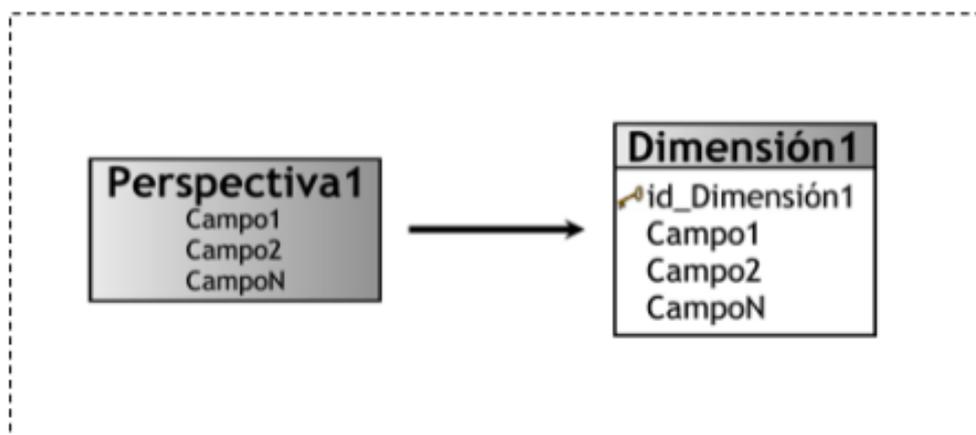


FIGURA 3.4. Diseño de tablas de dimensiones.

Fuente: Hefesto

En la figura 3.4 se muestra como se diseña una tabla dimensión se hace el traslado de la creación de las perspectivas y definición de campos a una tabla con un nombre, clave principal y los campos seleccionados.

Para los esquemas copo de nieve, cuando existan jerarquías dentro de una tabla de dimensión, esta tabla deberá ser normalizada. Por ejemplo, se tomará como referencia la siguiente tabla de dimensión y su respectivas relaciones padre-hijo entre sus campos:

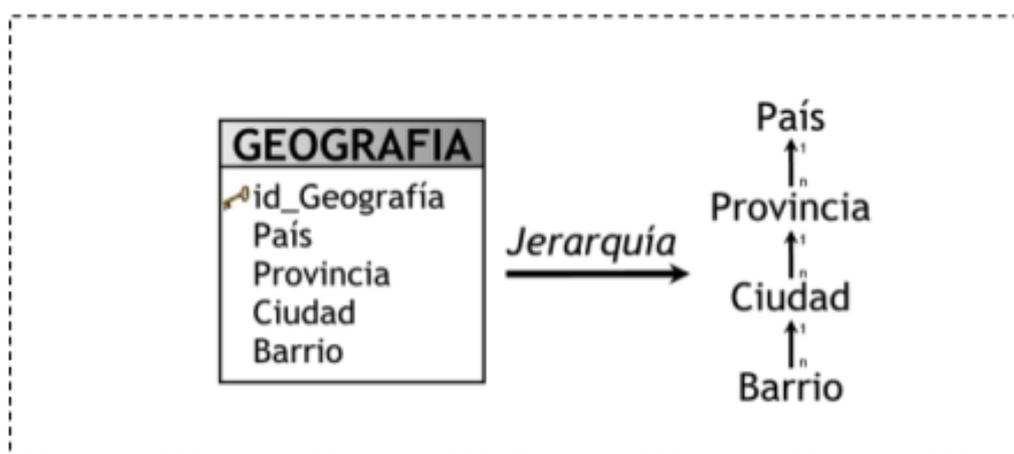


FIGURA 3.5. Jerarquía de "GEOGRAFÍA".

Fuente: Hefesto

En la figura 3.5 se aprecia una tabla desnormalizada, es decir, es una tabla con campos definidos. Dentro de la tabla geografía está conformada por niveles que van de lo general a lo específico.

Entonces, al normalizar esta tabla se obtendrá:

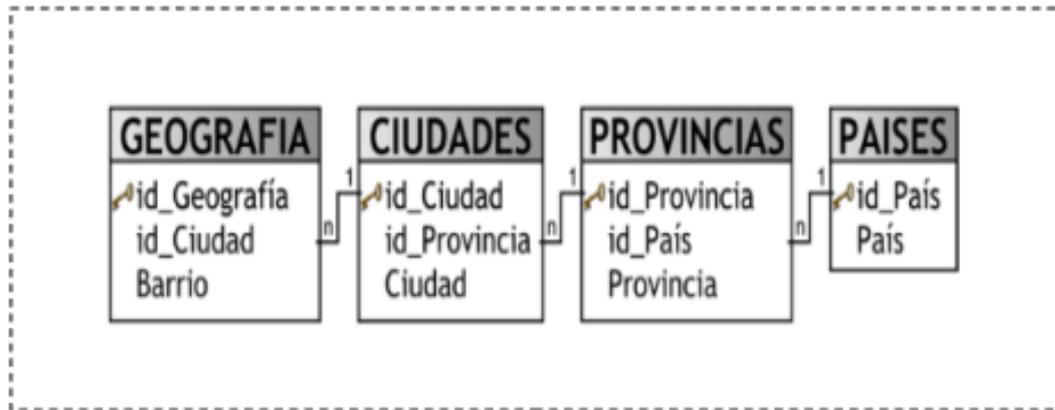


FIGURA 3.6. Normalización de "GEOGRAFÍA".

Fuente: Hefesto

La tabla de la figura 3.6 representa su normalización, es decir, los campos que estaban contenidos en una sola tabla ahora son una tabla y contienen su propia clave primaria y sus respectivos campos.

3.5.3 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.

- Se le deberá asignar un nombre a la tabla de hechos que represente la información analizada, área de investigación, negocio enfocado, etc.
- Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crearán tantos campos de hechos como indicadores se hayan definido en el modelo conceptual y se les asignará los mismos nombres que estos. En caso que se prefiera, podrán ser nombrados de cualquier otro modo.

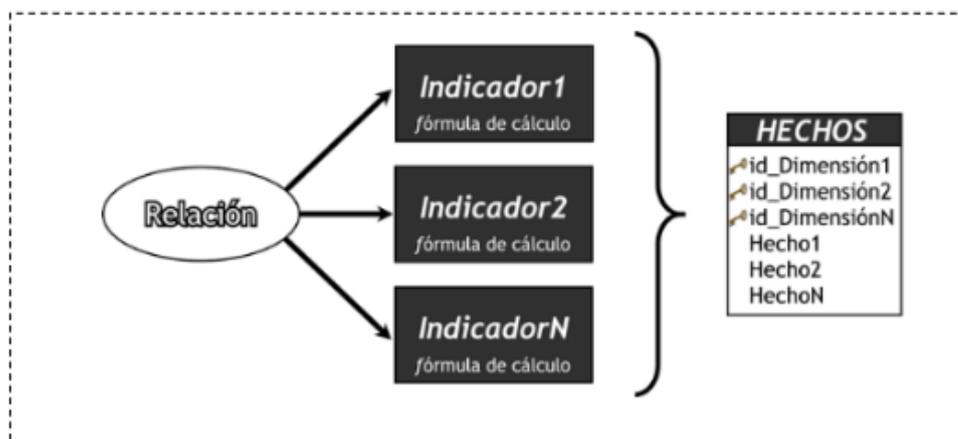


FIGURA 3.7. Tabla de hechos.

Fuente: Hefesto

La figura 3.7 indica como está compuesta una tabla hechos, se puede observar que tiene un nombre que representará la información analizada, en esta tabla su cuerpo está conformado por dos secciones, en la primera tiene una clave primaria que está compuesta por cada una de las claves primarias de las dimensiones que la rodean, y en la segunda sección contiene los indicadores definidos en el modelo conceptual.

Para los esquemas constelación se realizará lo siguiente:

- Las tablas de hechos se deben confeccionar teniendo en cuenta el análisis de las preguntas realizadas por los usuarios en pasos anteriores y sus respectivos indicadores y perspectivas.
- Cada tabla de hechos debe poseer un nombre que la identifique, contener sus hechos correspondientes y su clave debe estar formada por la combinación de las claves de las tablas de dimensiones relacionadas que se hayan definido en el modelo conceptual y se les asignará los mismos nombres que estos. En caso que se prefiera, podrán ser nombrados de cualquier otro modo.
- Al diseñar las tablas de hechos, se deberá tener en cuenta:

si en dos o más preguntas de negocio figuran los mismos indicadores pero con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición. Por ejemplo:

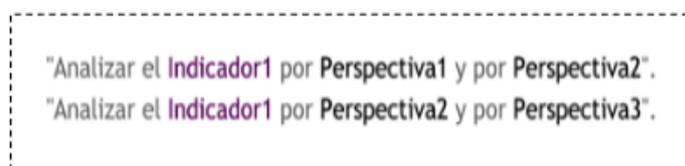


FIGURA 3.8. Caso 1, preguntas.

Fuente: Hefesto

La figura 3.8 muestra que un indicador puede ser consultado por dos o más tablas de hechos. Entonces se obtendrá:

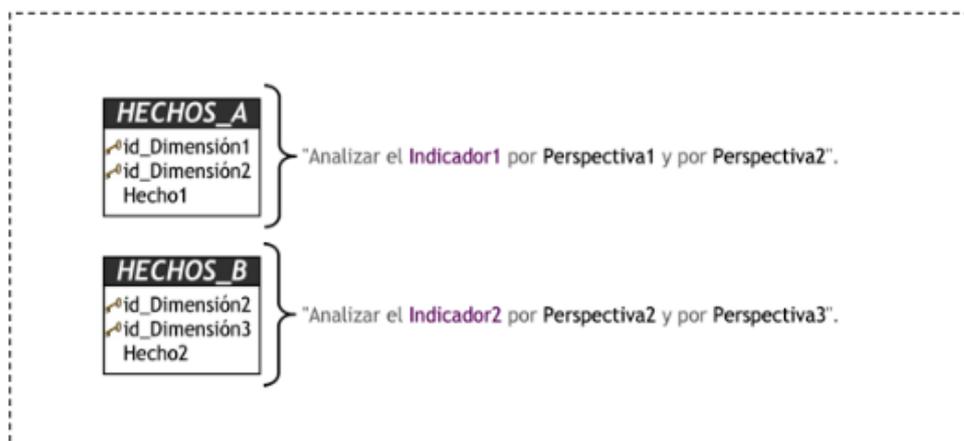


FIGURA 3.9. Caso 1, diseño de tablas de hechos.

Fuente: Hefesto

En la figura 3.9 se representa gráficamente lo que se expresó en la figura 3.8. La tabla hechos A analiza el indicador 1 con la dimensión 1 y la dimensión 2; el mismo indicador 1 está en la tabla hechos B pero se responde con las dimensiones 2 y 3.

Caso 2: si en dos o más preguntas de negocio figuran diferentes indicadores con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición. Por ejemplo:

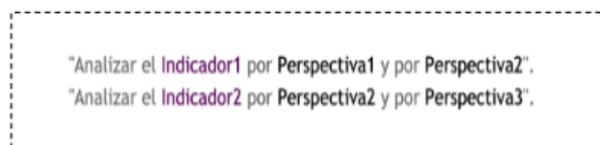


FIGURA 3.10. Caso 2, preguntas.

Fuente: Hefesto

Entonces se obtendrá:

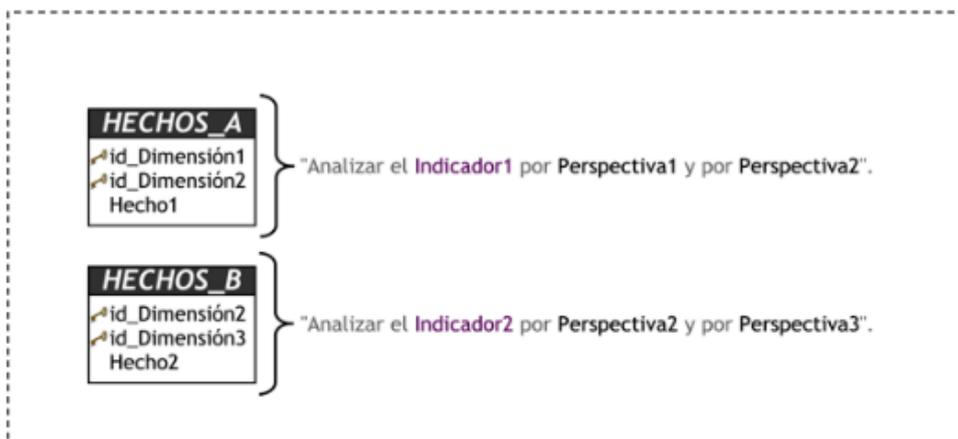
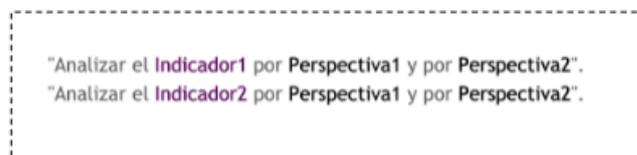


FIGURA 3.11. Caso 2, diseño de tablas de hechos.

Fuente: Hefesto

Caso 3: si el conjunto de preguntas de negocio cumplen con las condiciones de los dos puntos anteriores se deberán unificar aquellos interrogantes que posean diferentes indicadores pero iguales perspectivas de análisis, para luego reanudar el estudio de las preguntas. Por ejemplo:

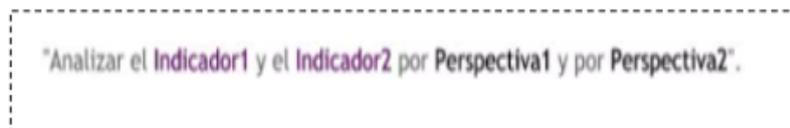


"Analizar el **Indicador1** por **Perspectiva1** y por **Perspectiva2**".
"Analizar el **Indicador2** por **Perspectiva1** y por **Perspectiva2**".

FIGURA 3.12. Caso 3, preguntas.

Fuente: Hefesto

Se unificarán en:



"Analizar el **Indicador1** y el **Indicador2** por **Perspectiva1** y por **Perspectiva2**".

FIGURA 3.13. Caso 3, unificación.

Fuente: Hefesto

3.5.4 Uniones

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

3.6 Paso 4) Integración de Datos

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

3.6.1 Carga inicial

La carga inicial del almacén de datos, en esta etapa se poblan el modelo de datos que se ha construido anteriormente, lo cual se lleva a cabo a través de tareas básicas como: limpieza de datos, ETL, etc.

Esta tarea puede realizarse con la ayuda de software que cumple con este fin y facilita este trabajo.

Se debe cargar el almacén de datos con datos unificados, al mismo tiempo se establecen condiciones y restricciones para utilizar datos de interés.

Cuando se trabaja con un esquema constelación, hay que tener presente que varias tablas de dimensiones serán compartidas con diferentes tablas de hechos, ya que puede darse el caso de que algunas restricciones aplicadas sobre una tabla de dimensión en particular para analizar una tabla de hechos, se puedan contraponer con otras restricciones o condiciones de análisis de otras tablas de hechos.

Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta siempre, la correcta correspondencia entre cada elemento. En el caso en que se este utilizando un esquema copo de nieve, cada vez que existan jerarquías de dimensiones, se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado.

Concretamente, en este paso se deberá registrar en detalle las acciones llevadas a cabo con los diferentes softwares. Por ejemplo, es muy común que sistemas ETL trabajen con "pasos" y "relaciones", en donde cada "paso" realiza una tarea en particular del proceso ETL y cada "relación" indica hacia donde debe dirigirse el flujo de datos. En este caso lo que se debe hacer es explicar que hace el proceso en general y luego que hace cada "paso" y/o "relación". Es decir, se partirá de lo más general y se irá a lo más específico, para obtener de esta manera una visión general y detallada de todo el proceso.

Es importante tener presente, que al cargar los datos en las tablas de hechos pueden utilizarse preagregaciones, ya sea al nivel de granularidad de la misma o a otros niveles diferentes.

3.6.2 Actualización

Cuando se haya cargado en su totalidad el almacén de datos, se deben establecer sus políticas y estrategias de actualización de datos.

Una vez realizado este paso, se tendrán que llevar a cabo las siguientes acciones:

- Especificar las tareas de limpieza de datos, calidad de datos, procesos ETL, etc., que deberán realizarse para actualizar los datos del almacén de datos.
- Especificar de forma general y detallada las acciones que deberá realizar cada software.

3.7 Desarrollo del modelo propuesto

En esta sección se detalla la creación del almacén de datos, es decir su estructura y funcionamiento. Siguiendo la metodología propuesta para la creación del almacén, a continuación se explica la arquitectura propuesta para la construcción del almacén de datos, inmediatamente con la recolección y análisis de los datos y requisitos, al igual que las necesidades de información del usuario, y finalizará con la confección del esquema conceptual y el esquema lógico.

3.7.1 Descripción de la arquitectura

Para dar una visión general del sistema a continuación en la figura 3.14 se explica la arquitectura común utilizada en los sistemas de almacén de datos. Como ya se menciona en Capítulos anteriores la inteligencia de negocios se refiere a tecnologías, herramientas y conjunto de técnicas, integración, análisis y procesamiento de información para permitir hacer mejor toma de decisiones. Hoy día la típica arquitectura de un almacén de datos (o almacén

departamental) consiste en consolidación de datos de diferentes bases de datos, y sirve a una variedad de consultas de front-end, reportes y herramientas analíticas. Por otro lado el back-end de la arquitectura son datos integrados para poblar el almacén extrayendo datos de fuentes operacionales heterogéneas y generalmente distribuidas; limpieza, integración y transformación de datos y cargarlos en el almacén de datos.

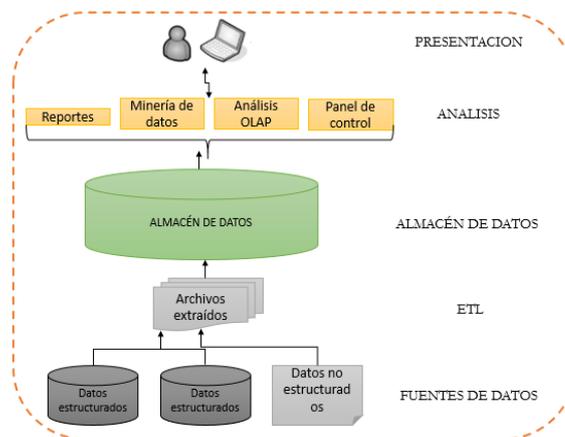


FIGURA 3.14. Arquitectura almacén de datos.

Fuente: propia

Los componentes de la arquitectura son las siguientes:

- Fuentes de datos: son los datos operacionales con los que se poblará el almacén de datos. El origen de estos datos son los sistemas transaccionales de la organización como también datos externos a ella.
- Procesamiento de los datos (ETL): son los procesos para sumarizar y realizar cambios en los datos operacionales. Los datos se limpian, transforman, combinan, quitan duplicados, se preparan para ser utilizados en el almacén de datos.
- Carga de datos en el almacén: se concentran los datos que permitirán llevar a cabo la toma de decisiones.
- Datos disponibles a los clientes (usuarios finales): los datos serán combinados y explotados por herramienta BI, para minería de datos, herramientas OLAP o la creación de informes finales.

caso de estudio Como se menciona en el Capítulo 1 sólo se trabajará con la parte central de la arquitectura propuesta que se refiere al modelado del almacén de datos. Como se muestra en la figura 3.15.

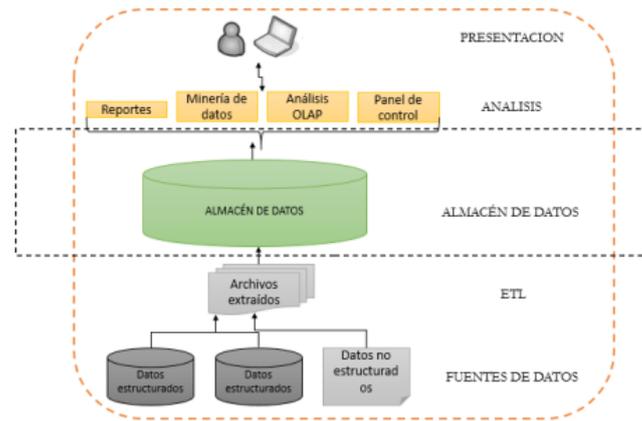


FIGURA 3.15. área resaltada de la arquitectura almacén de datos.

Fuente: propia

Las fuentes de datos que se utilizaron en la obtención de los datos, fue los artefactos producidos bajo la norma MoProSoft y CMMI. Toda la documentación que se explica en el Capítulo 4.

El proceso de ETL es la sección en donde se lleva a cabo el agrupamiento de una serie de procesos que realizan actividades relacionadas con la extracción, manipulación, control, integración, limpieza y carga de datos en el almacén de datos. Para efecto de esta tesis no se tratara este proceso.

En el almacén de datos se congregan los datos operacionales con una estructura multidimensional que permiten generar consultas flexibles. En la siguiente sección se abordará esté proceso detalladamente ya que es el motivo de esta tesis.

El módulo de presentación es el área en donde están los datos disponibles al cliente y se lleva a cabo la interacción con el usuario final, con la finalidad de mostrar los datos de forma útil y transparente por medio de diferentes herramientas. Para efecto de esta tesis no se tratara este proceso.

3.7.2 Metodología del desarrollo

El almacén es desarrollado de la mano de la metodología de Hefesto. Las etapas de la metodología de Hefesto seleccionadas para esta primera fase quedan ilustradas en la figura 3.16:

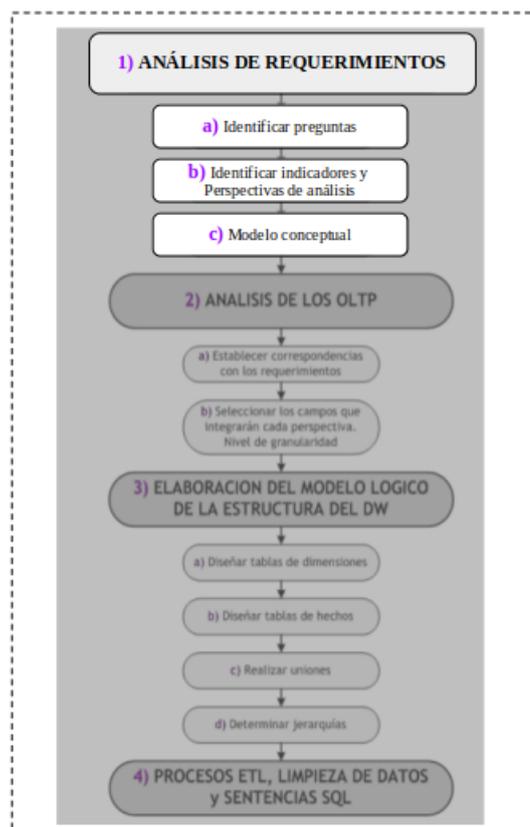


FIGURA 3.16. Metodología de Hefesto adaptada.

Fuente: Hefesto

La figura anterior representa la metodología de Hefesto adaptada para este trabajo de tesis, que ya fue explicada al inicio de este Capítulo. Las etapas contempladas se desarrollaron de la siguiente manera:

3.7.2.1 Análisis de requerimientos

a) Identificar preguntas

Como ya se explicó anteriormente este primer paso se inició con la recolección de los datos, este plan de recolección de datos consiste en recoger los datos necesarios que se encuentran en los documentos generados por las normas MoProSoft y CMMI las cuales son implantadas e implementadas en la empresa en cuestión.

Las técnicas empleadas para llevar a cabo la recolección de datos fue las entrevistas y posteriormente el análisis documental.

En este caso se llevó a cabo por medio de entrevistas en las áreas de la empresa. La entrevista consistió en preguntas abiertas, sin un orden preestablecido, lo más parecido a una conversación que permitió la espontaneidad, es decir las preguntas se realizaron de acuerdo a las respuestas que fueron surgiendo durante la entrevista.

Las áreas que se abordan en este trabajo de tesis corresponden exclusivamente a 4 de las 9 que define la norma MoProSoft Gestión de Proyectos (GPY), Recursos Humanos y Ambiente de Trabajo (RHAT), Desarrollo y Mantenimiento de Software (DMS) esta área por medio de CMMI, Administración de Proyectos Específicos

(APE) y el área que se encuentra fuera de las normas implantadas en la empresa pero que representa la columna vertebral de las principales actividades dentro de ella, se trata de la Gerencia de Finanzas.

Análisis de las fuentes de datos como parte de las entrevistas los responsables de las áreas proporcionaron documentación misma que fue analizada y estudiada cuidadosamente. De esta manera es como se obtienen e identifican las necesidades de información y las preguntas que facilitarán la toma de decisiones. Éstas se presentan como resultado del análisis de los documentos producidos bajo las normas, son preguntas que se proponen a la gerencia para que les ayude a medir el desempeño de los procesos que en estas áreas se realizan.

Las preguntas propuestas dan paso a crear los indicadores. Los indicadores representan de mejor modo el proceso de una área y que es exactamente lo que se desea analizar de la misma.

La empresa MBN es una pequeña empresa que se ha mantenido en el mercado con gran habilidad. Es una empresa ordenada que se rige bajo las normas antes mencionadas, de finanzas claras y muy alerta. En primer plano, esto sugiere que se apoye la gestión de desarrollo de proyectos. Por lo tanto en conversaciones con el cliente acerca de las potencialidades de la inteligencia de negocios llevaron a las preguntas en esa dirección: gestión del desarrollo de proyectos.

A continuación y en terminos concretos se presenta la lista de las preguntas que el cliente requiere dar respuesta:

- ¿Costo de venta por proyectos desarrollados, durante un tiempo determinado?
- ¿Colaboradores que participaron en el desarrollo de proyectos durante un tiempo determinado?
- ¿Gasto generado a la nomina por proyecto durante un tiempo determinado?
- ¿Cantidad vendida de proyectos a cada cliente durante un tiempo determinado?
- ¿Línea de negocio más rentable durante un tiempo determinado?
- ¿Costo de operación por colaborador durante un tiempo determinado?

Como se puede apreciar, las necesidades de información van a corde con lo que el cliente necesita, es la información requerida la que proveerá un ámbito para la toma de decisiones, que en este caso permitira analizar el comportamiento de los colaboradores con los proyectos y el comportamiento de los clientes con los proyectos y líneas de negocio de la empresa, ya que a estos son los que se pretende satisfacer ampliamente.

b) Identificar indicadores y perspectivas de análisis

Una vez que se identificaron las preguntas claves se descomponen para identificar los indicadores y las perspectivas de análisis que intervendrán. Por ende los indicadores son valores numéricos que representan lo que se desea analizar concretamente. La perspectivas son los objetos por medio de los cuales se examinarán los indicadores, para dar respuesta a las preguntas planteadas como por ejemplo: clientes, proveedores, sucursales, productos, entre otros.

A continuación, se presenta el análisis de las preguntas obtenidas en el paso anterior y se detallan cuales son sus respectivos indicadores y perspectivas

- Costo de venta por proyectos desarrollados durante un tiempo determinado
INDICADOR PERSPECTIVAS PERSPECTIVAS
- colaboradores que participaron en el desarrollo de proyectos durante un tiempo determinado
PERSPECTIVAS PERSPECTIVAS PERSPECTIVAS
- Gasto generado a la nomina por proyecto durante un tiempo determinado
PERSPECTIVAS PERSPECTIVAS PERSPECTIVAS
- Cantidad vendida de proyectos a cada cliente durante un tiempo determinado
INDICADOR PERSPECTIVAS PERSPECTIVAS PERSPECTIVAS
- línea de negocio más rentable durante un tiempo determinado
PERSPECTIVAS PERSPECTIVAS
- costo de operación por colaborador durante un tiempo determinado
PERSPECTIVAS PERSPECTIVAS PERSPECTIVAS

En síntesis los indicadores de interes son: costo de venta y cantidad vendida. Y las perspectivas de análisis son: proyectos, colaboradores, nomina, costos operación y tiempo.

c) Modelado conceptual

En esta sección se abordan los elementos multidimensionales que forman parte de la solución. Para el modelado se tomó como base la lista de indicadores, generados por las preguntas propuestas, que se determinaron de acuerdo al previo análisis que se hizo a cada uno de los artefactos producidos bajo las normas para las áreas mencionadas anteriormente, el cual son la base para la construcción y modelado del proyecto.

Se construirá el modelado a partir de los indicadores y perspectivas. Es por medio de este modelo que se podrá ver con claridad cual es el alcance del almacén de datos, posteriormente se trabajará sobre ellos y como se posee un alto nivel de definición de los datos, se presentará con facilidad a los usuarios.

La presentación gráfica del modelado conceptual de los datos se muestra en la figura 3.17.

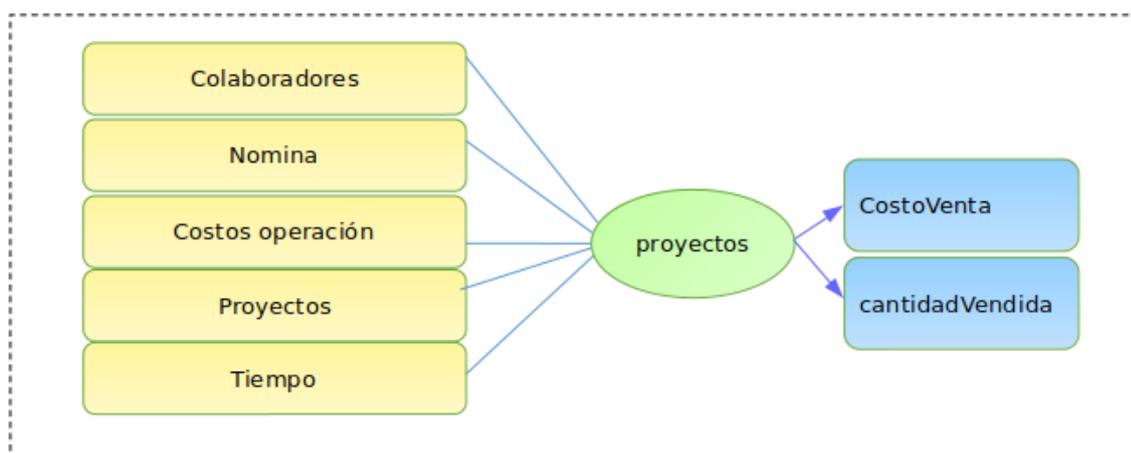


FIGURA 3.17. Modelo conceptual de los proyectos.

Fuente: propia

El modelo conceptual que se muestra es el modelo resultante de los datos que se han recolectado y como puede observarse, es clara la relación que existe entre ellas para alcanzar los resultados los indicadores establecidos. A la izquierda se colocan las perspectivas seleccionadas que son unidas al ovalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación constituye el proceso de estudio elegido. Ahora bien, de dicha relación y entrelazadas con flechas, de aquí se desprenden los indicadores o medidas, estos son los que se encuentran ubicados a la derecha del esquema.

3.7.2.2 Análisis de los OLTP

a) Establecer correspondencias con los requerimientos

En este punto la metodología Hefesto propone llevar a cabo el análisis de los OLTP disponibles en la empresa que contengan la información requerida, sus características, para así poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. El objetivo es, que todos los elementos del modelo conceptual esten correspondidos a los OLTP o las fuentes de datos que proporcionen la información. En el caso particular de la empresa en cuestión no cuenta con OLTP que proporcione información que ayude a la toma de decisiones, y como ya se comento al inicio del Capítulo las fuentes principales de datos para este proyectos fueron los documentos generados por MoProSoft y CMMI como se observa en la imagen 3.18.

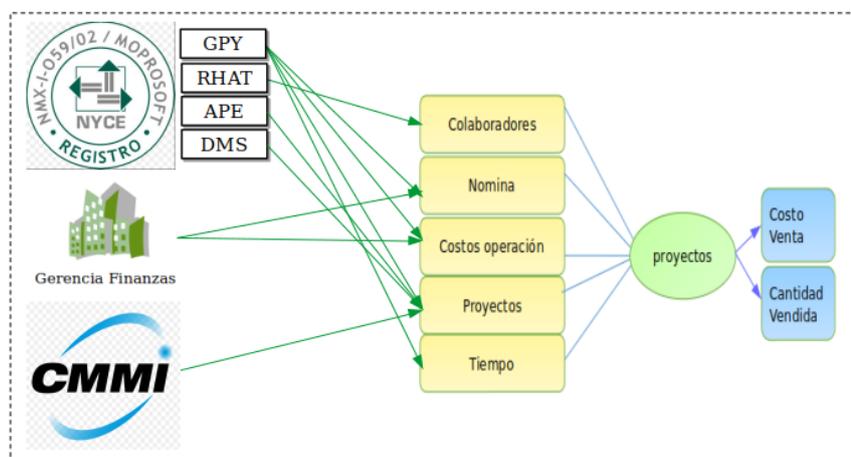


FIGURA 3.18. Correspondencias.

Fuente: propia

Las relaciones identificadas en la figura 3.18 fueron, de "MoProSoft que integra las áreas GPY, RHAT, APE, DMS" se relacionan de la siguiente manera:

- GPY, de esta área se obtiene información sobre la administración y las ventas de los proyectos y las relaciones con los clientes.
- RHAT, de esta área se obtiene la información relacionada con el comportamiento de los colaboradores.

- APE, esta área junto y de la mano con GPY proporciona información sobre la administración de los proyectos.
- DMS, proporciona información respecto a la administración de la documentación del ciclo de vida del proyecto.

Las relaciones identificadas con el área de finanzas son las siguientes:

- Nómina, proporciona la actividad económica con respecto a los colaboradores
- Costos operación, proporciona información con respecto a la inversión que se genera con respecto al integrar los colaboradores al desarrollo de los proyectos.

La relación identificada con CMMI:

- Proyectos, ya que se refiere a la buena práctica para el desarrollo de software, esto significa que para este proyecto proporciona información respecto al proceso relativo al desarrollo e implementación del software.

De esta manera queda representada la relación que existe entre la información de las áreas a la hora de dispararse un proyecto en la empresa.

a) Seleccionar los campos que integrarán cada perspectiva.

Una vez establecidas las relaciones, se examinan y establecen los campos que contendrán cada perspectiva, ya que es a través de ellos por los que se manipularán y filtrarán los indicadores. Por lo tanto en base a las correspondencias establecidas se analizaron, se definieron los campos, a través de dos métodos diferentes. Primero se examinarán los documentos como ya se menciono anteriormente para intuir los significados de cada campo, y luego se consultarán con el encargado sobre algunos puntos en los cuales no se comprendía su sentido.

Con respecto a la perspectiva "Colaboradores", los datos disponibles son los siguientes:

- id colaboradores: es la clave primaria de la tabla "colaboradores", y representa a un colaborador en particular.
- claveColaborador: representa la clave del colaborador, está formado por un conjunto numérico incremental.
- colaborador: se refiere a el nombre y apellidos del colaborador
- salario: es el dinero que recibe el colaborador de la empresa en concepto de paga.
- estatus: indica cual es la situación actual del colaborador en la empresa. Por ejemplo: puede ser activo o inactivo.
- fecha ingreso: la fecha en que se integro a colaborar con la empresa
- fecha termino: corresponde a la fecha en que dejo de laborar con la empresa.
- equipoTrabajo: se refiere al perfil que tiene el colaborador. Por ejemplo: Analista, Diseñador, Revisor, Responsable manuales, Programador, Diseñador de interfaz, Responsable de pruebas.

Con respecto a la perspectiva "Nómina", los datos disponibles son los siguientes:

- id nomina: es la clave primaria de la tabla "Nomina" y representa al pago correspondiente a un colaborador.
- numeroNominaColaborador: hace referencia al número de pago que se está realizando.
- colaborador: nombre del colaborador que se le realiza el pago.
- mesNomina: número de mes correspondiente al pago a realizar.
- primeraQuincenaMes: cantidad monetaria que se le paga al colaborador.
- fechaPagoNomina: fecha en que se realiza el pago.
- segundaQuincenaMes: cantidad monetaria que se le paga al colaborador.
- fechaPagoNomina: fecha en que se realiza el pago.

Con respecto a la perspectiva "Costos operación", los datos disponibles son los siguientes:

- id costosOperacion: es la clave primaria de la tabla "Costos operacion" y representa el gasto por colaborador al integrarse a un proyecto.
- colaborador: representa el nombre del colaborador que se integra a un proyecto.
- nominaIMSS: pago de nomina imss
- nominaHonorarios: pago de honorarios.
- prestaciones: prestaciones realizadas.
- viaticos: pago de viaticos
- fechaAsignacionCosto: fecha en que se realiza el costo
- lineaNegocio: linea de negocio a la que se realiza el cargo.
- proyecto: nombre del proyecto
- numeroProyecto: número asignado por parte de la empresa a un proyecto.

Con respecto a la perspectiva "Proyectos", los datos disponibles son los siguientes:

- id proyectos: es la clave primaria de la tabla "proyectos" y representa los sucesos ocurridos al desarrollar un proyecto.
- cliente: representa a un cliente en particular.
- fechaRegistroCliente: fecha en que el cliente contrató por primera vez un proyecto de la empresa.
- lineNegocios: forma en que la empresa categoriza sus productos y servicios.
- tipoCapacitacion: son las capacitaciones que se requirieron para llevar a cabo el proyecto
- capacitacion: nombre de la capacitación solicitada
- costoCapacitacion: el monto que se pago por la capacitación solicitada.
- cantidadPersonasCapacitacion: cuantas personas recibieron la capacitación
- fechaCapacitacion: fecha en que se llevó a cabo la capacitación
- tipoRecurso: tipo de recursos que se requirieron para el desarrollo del proyecto.

- conceptoRecurso: es la descripción del recurso solicitado.
- cantidadRecurso: es la cantidad de recursos solicitados.
- fechaEntregaRecurso: la fecha en que se entrego el recurso.
- costoInvertidoRecurso: el monto que se pago por el recurso.
- proyecto: nombre del proyecto
- tipoProyecto: es la clasificación que asigna la empresa a sus proyectos. Por ejemplo: interno y externo.
- fechaInicio: fecha en que despliega el proyecto.
- fechaFin: fecha en que concluye el proyecto.

Con respecto a la perspectiva "Tiempo", los datos disponibles son los siguientes:

- id tiempo: es la clave primaria de la tabla "Tiempo" y representa el periodo en que sucedieron los hechos con respecto a los proyectos y actividades relacionadas a dichos proyectos.
- fecha
- año
- trimestre

Se recolecto toda la información pertinente y fue verificada por el usuario cuales eran los datos que consideraban de interés. Tomando en cuenta lo anterior se acompleta el diseño del modelo conceptual como se muestra en la figura 3.19.

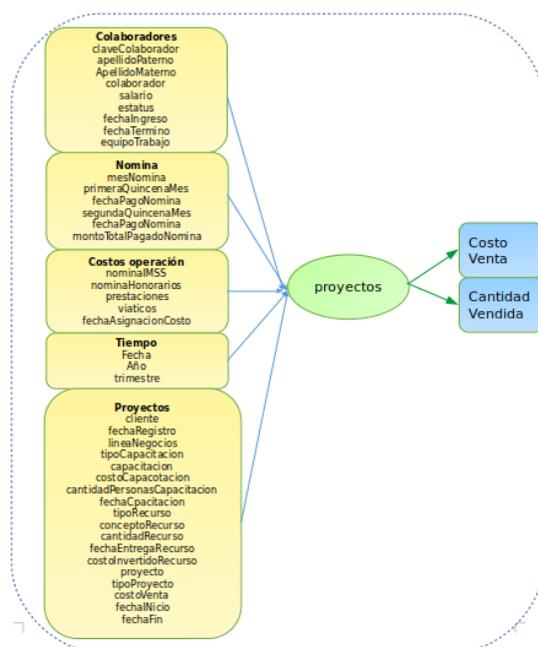


FIGURA 3.19. Modelo conceptual con atributos.

Fuente: propia

En la figura superior se aprecia que cada una de las perspectivas ya contiene atributos que fueron aprobados por el usuario.

3.7.2.3 Construcción del modelo lógico de la estructura del almacén de datos

Una vez que se concluyó el modelado conceptual del almacén de datos, se confeccionó el modelo lógico, teniendo como base el modelo conceptual.

El esquema que se seleccionó para contener la estructura del depósito de datos, es el esquema estrella debido a sus características, ventajas y diferencias con los otros esquemas.

a) Diseño de las tablas dimensiones

A continuación, se diseñan las tablas dimensiones. Lo primero que se hace para crear las dimensiones se toma cada perspectiva con sus atributos relacionados y lo que se les realiza es:

- se les asigna el nombre que identifique a la dimensión.
- se les añade el campo que representa su clave principal y finalmente
- se agregan los nombres de los atributos

Y gráficamente quedan representados así:

De la perspectiva "Colaboradores"

La figura 3.20 permite visualizar la transformación que sufre la perspectiva colaboradores para convertirse en una tabla dimensión, la cual ya cuenta con un nombre, una clave principal y los campos asignados.

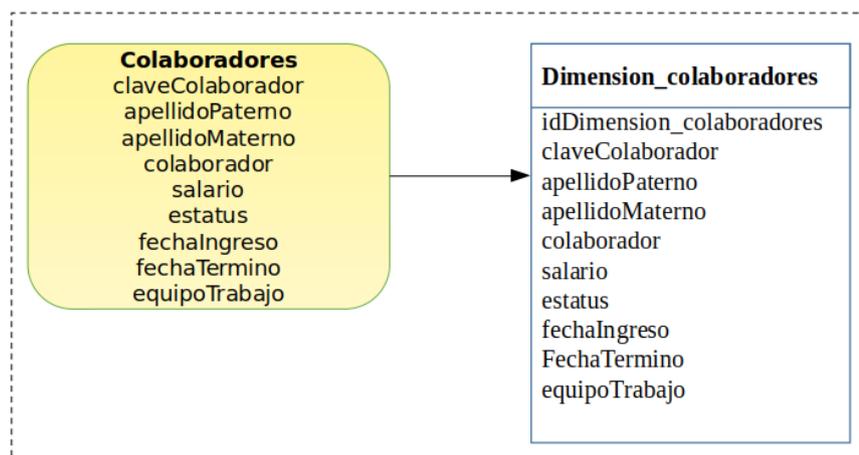


FIGURA 3.20. Dimensión colaboradores.

Fuente: propia

Los campos definidos para la perspectiva colaboradores, son los mismos que quedan establecidos para la tabla dimensión colaboradores la diferencia que predomina es que ahora la tabla muestra una clave primaria.

De la perspectiva "Nómina"

En la figura 3.21 se observa el cambio de la perspectiva nómina para transformarse en una tabla dimensión, que está conformada por un nombre, su clave principal y sus respectivos campos.

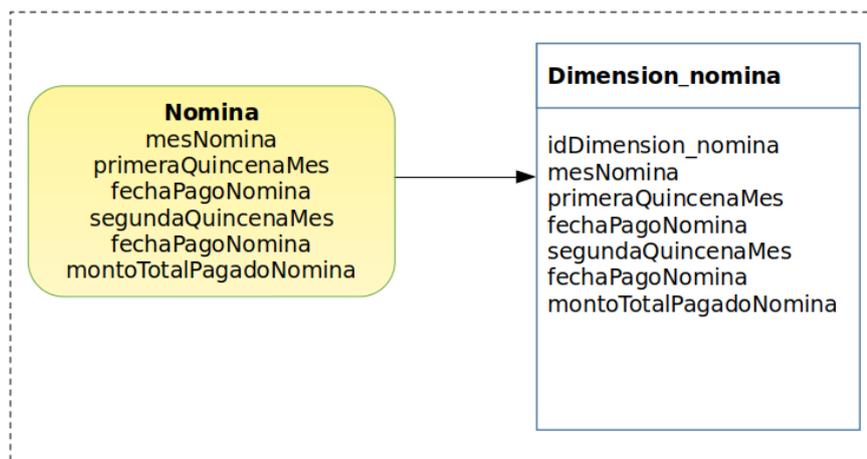


FIGURA 3.21. Dimensión Nómina.

Fuente: propia

Los campos para la tabla nómina no cambian son los mismos que se asignaron en la perspectiva, solo que ahora cuenta con un nombre y una clave primaria.

De la perspectiva "Costos operación"

La perspectiva costo operación es mostrada en la figura 3.22 en ella es mostrada la transición de dicha perspectiva a convertirse a una tabla dimensión.

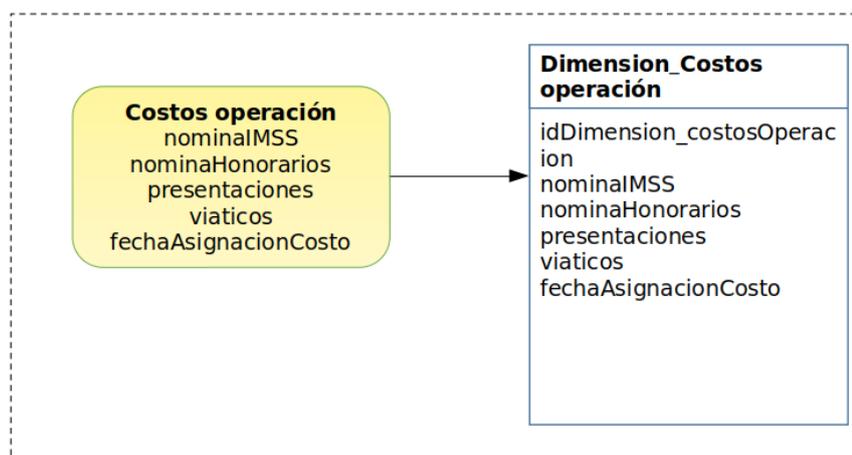


FIGURA 3.22. Dimensión Costos operación.

Fuente: propia

Pasar de una perspectiva a una tabla dimensión, conlleva a adquirir un nombre para darle una identificación, tener una clave primaria y agregar sus atributos.

De la perspectiva "Proyectos"

En la figura 3.23 se observa el cambio de la perspectiva proyectos para transformarse en una tabla dimensión, que está conformada por un nombre, su clave principal y sus respectivos campos.

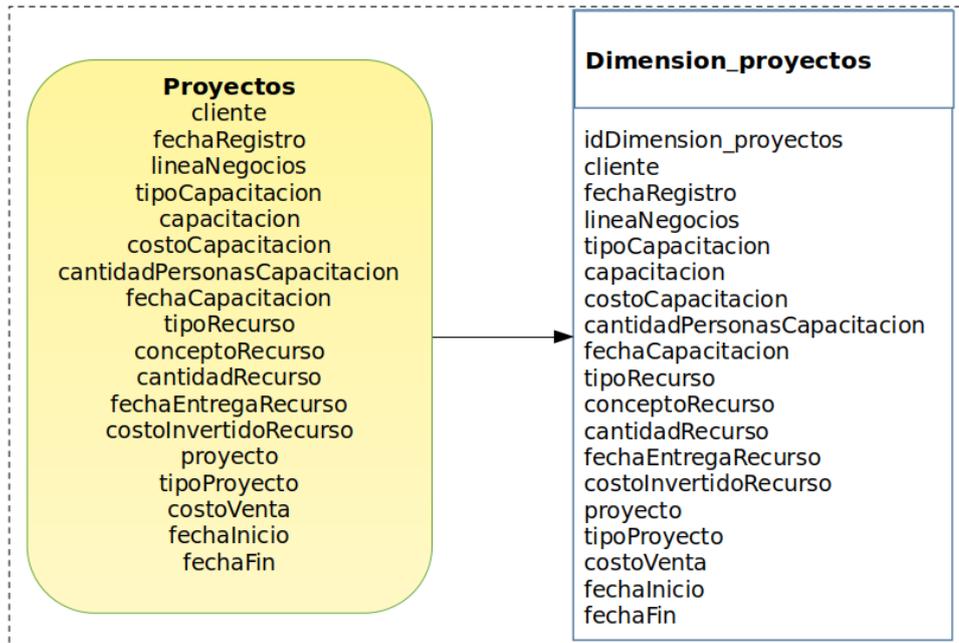


FIGURA 3.23. Dimensión Proyectos.

Fuente: propia

Los campos para la tabla proyectos no cambian son los mismos que se asignaron en la perspectiva, solo que ahora cuenta con un nombre y una clave primaria.

De la perspectiva "Tiempo"

La figura 3.24 permite visualizar la transformación que sufre la perspectiva tiempo para convertirse en una tabla dimensión, la cual ya cuenta con un nombre, una clave principal y los campos asignados.

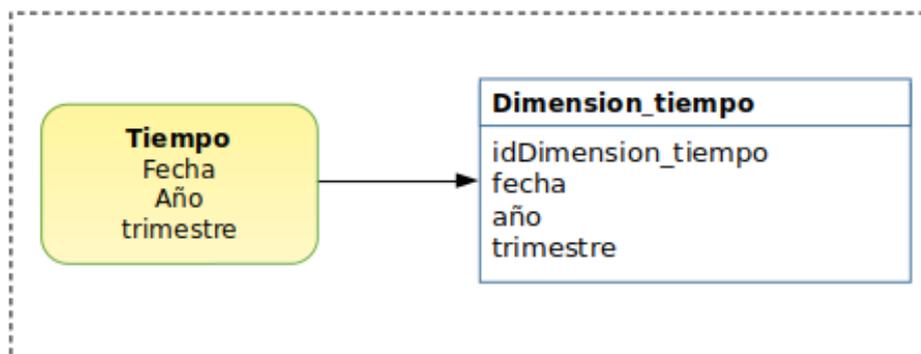


FIGURA 3.24. Dimensión Tiempo.

Fuente: propia

Los campos definidos para la perspectiva tiempo, son los mismos que quedan establecidos para la tabla dimensión tiempo la diferencia que predomina es que ahora la tabla muestra un nombre y una clave primaria.

b) Diseño de las tabla hechos

Es aquí en este paso donde se diseña la tabla hechos. Se realizó lo siguiente:

- se les asigna un nombre que identifique a la tabla hechos, haciendo referencia a los proyectos desarrollados. Por lo tanto el nombre que se le asigna es "proyectos".
- se les añade el campo que representa su clave principal, que está compuesta por las claves primarias de cada dimensión que se utilizara para generar la consulta : "idDimensionColaboradores", "idDimensionNomina", "idDimensionCostosOperacion", "idDimensionProyectos" e "idDimensionTiempo".
- se agregan los nombres de los indicadores: "costoVenta" y "cantidadVendida".

El diseño de la tabla hechos quedo formada de la siguiente como se muestra en la figura 3.25.

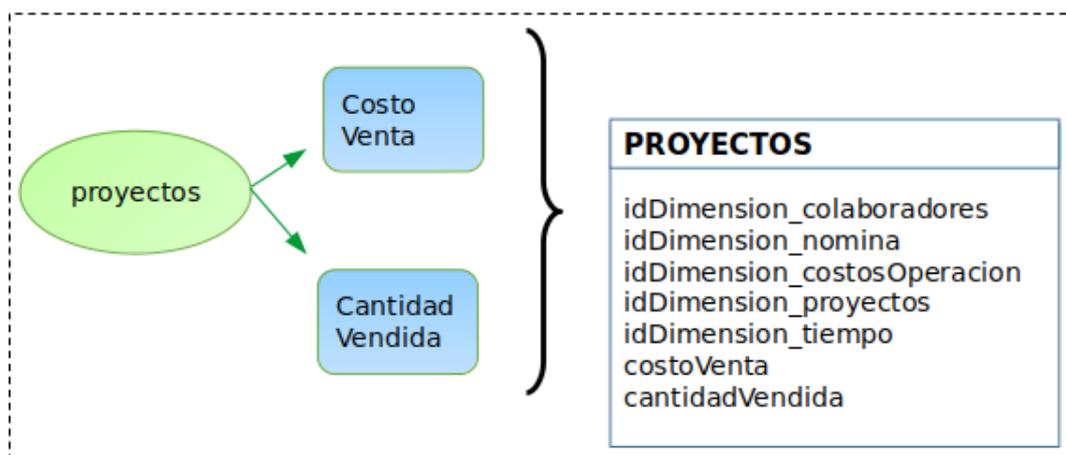


FIGURA 3.25. Diseño de la tabla hechos.

Fuente: propia

En la figura superior muestra que la tabla hechos queda constituida por nombre que corresponde al que se encuentra en el ovalo que es el tema central del almacén de datos, la clave primaria de la tabla no es más que cada uno de las claves principales de las tablas dimensión que ya fueron diseñadas y creadas, al final de la tabla se encuentran las medidas o indicadores.

c) Realizar uniones

Se realizan las uniones pertinentes, y todas las tablas dimensiones quedan unidas a la tabla hechos como se muestra en la figura 3.26.

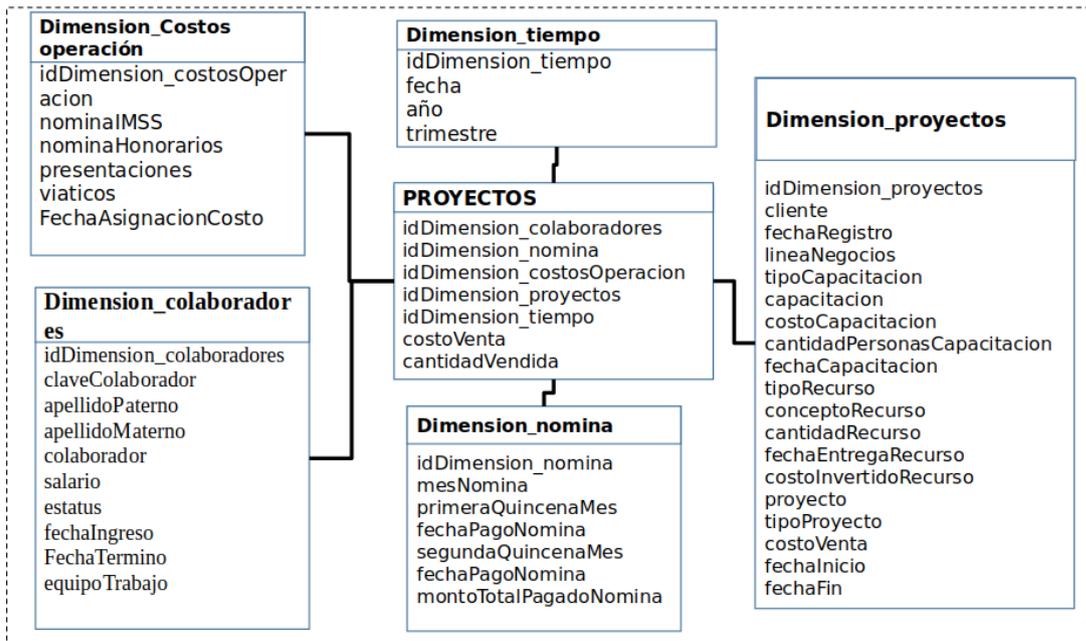


FIGURA 3.26. Diseño físico del almacén de datos.

Fuente: propia

En la figura también se muestra el resultado de como queda construido el modelo lógico del almacén de datos. Es de esta manera que queda formado el modelo estrella (elegido en 3.7.5 Construcción del modelo lógico de la estructura del almacén de datos).

3.7.3 Esquema copo de nieve

Como se menciona en el Capítulo 1 en la sección de aportaciones, se propone un segundo modelo de datos, para hacer una comparativa entre los esquemas estrella y copo de nieve tomando en cuenta los tiempos de ejecución de consultas con respecto uno del otro.

Con la finalidad de determinar que modelo es el más factible de implementar en la empresa.

Se tomará en cuenta el mismo proceso del esquema estrella a partir de: análisis de requerimientos: a) identificar preguntas, b) identificar indicadores y perspectivas de análisis; el índice c) modelado conceptual es en este paso donde se realiza la normalización de los datos, la característica principal que hace diferencia entre el esquema copo de nieve y esquema estrella como ya fue explicado a profundidad en el Capítulo 2.

Siguiendo las bases del modelado conceptual de datos de esquema estrella se parte de aquí para el modelado conceptual de copo de nieve.

Modelado conceptual copo de nieve

Para el modelado se tomó como base la lista de indicadores, generados por las preguntas propuestas, que se

determinaron de acuerdo al previo análisis que se hizo a cada uno de los artefactos producidos bajo las normas para las áreas mencionadas anteriormente, el cual son la base para la construcción y modelado de datos de este esquema.

La figura 3.27 muestra el diseño del esquema copo de nieve con los siguientes puntos:

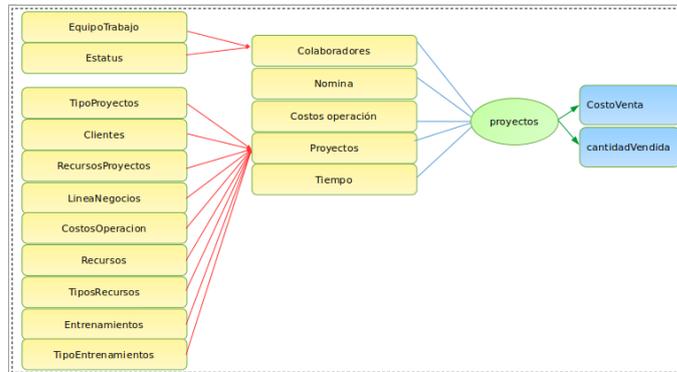


FIGURA 3.27. Dimensión colaboradores en copo de nieve.

Fuente: propia

- Algunas tablas de dimensiones en el esquema copo de nieve están normalizadas
- Con la normalización se dividieron los datos en tablas adicionales
- Entonces a diferencia del esquema estrella, la tabla de dimensiones de un esquema copo de nieve está normalizada. Por ejemplo la tabla colaboradores en el esquema estrella se normaliza y se divide en dos tablas dimensiones, equipo trabajo y estatus.
- La tabla de dimensión de colaboradores contiene los atributos clave primaria y las claves, idDimensionEquipoTrabajo y idDimensionEstatus como se muestra en la figura 3.28.

La jerarquización de la figura colaboradores se muestra en la tabla 3.28.

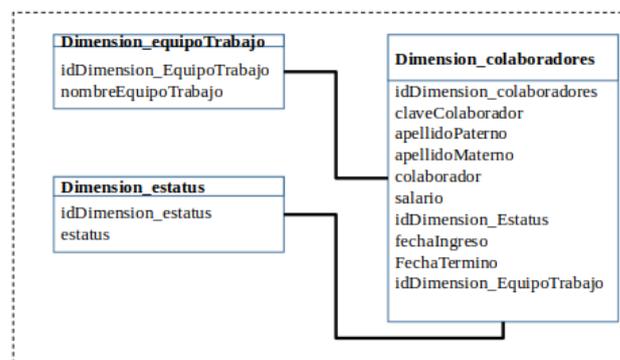


FIGURA 3.28. Jerarquía de la tabla colaboradores.

Fuente: propia

En la figura se aprecia como la tabla colaboradores fue normalizada dividiendola en dos tablas hijas equipo de trabajo y estatus estas tablas ahora contienen un nombre, un identificador el cual sera relacionada con la tabla padre y sus respectivos campos.

Debido a la normalización en el esquema copo de nieve, la redundancia se reduce y, por lo tanto, se vuelve fácil de mantener y de ahorrar espacio de almacenamiento.

La jerarquización de la figura colaboradoresse muestra en la tabla 3.29.

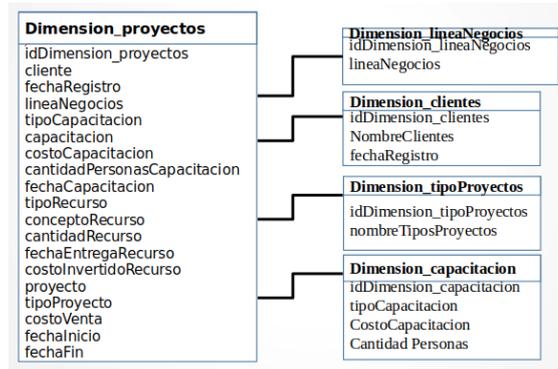


FIGURA 3.29. Jerarquía de la tabla proyectos.

Fuente: propia

En la figura se aprecia como la tabla proyectos fue normalizada dividiendola en dos tablas hijas líneas de negocios, clientes, tipo proyectos y capacitación estas tablas ahora contienen un nombre, un identificador el cual sera relacionada con la tabla padre y sus respectivos campos.

MODELO MoProSoft Y CMMi : LA EMPRESA MBN



FIGURA 4.1. Logotipo de la empresa MBN.

El proyecto se realizó en la empresa Miracle Business Network S.A de C.V. (MBN). ubicada en calle 37 No. 216-B La Loma Xicohtencalt en el estado de Tlaxcala.

4.1 La empresa

Miracle Business Network S.A de C.V. inicia operaciones en Febrero de 2000 en la ciudad de México, brindando servicios de administración de base de datos a ORACLE de México. En el año 2008 se toma como estrategia de largo plazo en MBN adoptar la norma MoProSoft-059-NYCE-2005 como una herramienta estratégica para dar estructura a la organización. En enero de 2015 MBN es evaluada de manera satisfactoria en SCAMPI-A por evaluadores del Software Engineering Institute (SEI), implementando la estructura de calidad MoProsoft alcanzando actualmente nivel dos en una escala para mejora y evaluación de los procesos de desarrollo y mantenimiento de sistemas y productos de software. MoProSoft fue desarrollado por la asociación mexicana para la calidad en la ingeniería de software a través de la facultad de ciencias de la Universidad Nacional Autónoma de México (UNAM). Recientemente logrando un nivel dos en la escala de madurez de capacidades o CMMi el cual es un modelo para la mejora y evaluación de procesos para el desarrollo, mantenimiento, y operación de sistemas de software, en los cuales se siguen distintos procesos de especificación y entendimiento de requerimientos de software. (MBN)

4.2 Principales líneas de productos o servicios

En la empresa MBN se desarrollan cinco unidades estratégicas de negocios:

4.2.1 Desarrollo de software a la medida (DMS)

Esta unidad estratégica de negocio (UEN) que a nivel de fabrica de software, desarrolla proyectos de software a la medida regidos por la norma de calidad MoProSoft y CMMI. Haciendo uso de diversas plataformas como java, .net, oracle, entre otros.

4.2.2 Consultoría de negocios e implementación funcional (CNIF)

Es un equipo de consultores que se encarga de implantar sistemas de información, teniendo como especialidad la implementación de ERP (Enterprise Resource Planning); es un software que permite a las empresas automatizar sus procesos de:

- Compras
- Ventas
- Inventarios
- Contabilidad
- Manufactura
- Recursos Humanos

4.2.3 Desarrollo, Mantenimiento y Personalización Oracle (DMPO)

Es un equipo calificado para realizar desarrollos, ajustes y customizaciones al ERP:

- SAP ABAP
- ORACLE Developer Suite
- ORACLE People Tools

4.2.4 Consultoría en Administración de Base de Datos (CABD)

La empresa MBN integra un portafolio de soluciones (Applications, Middleware y Database):

- Diseño de Bases de Datos
- Normalización de Bases de Datos
- Instalación de Bases de Datos
- Actualización de parches y Versiones
- Diseño y Configuración de Bases de datos en Cluster
- Diseño y Configuración de Bases de datos en DRP

- Migración de Plataforma de Bases de datos
- Programación de aplicaciones
- Configuración y adecuación de BladeCenter, FlexSystem, Pure y HexaData.
- Migraciones Heterogéneas con GoldenGate

4.2.5 Licenciamiento y comercialización de productos(LCP)

Hoy en día, cuentan con una gama de productos que cubren necesidades específicas y licencian productos de sus principales marcas como Oracle, sap, peoplesoft y citrix.

4.3 MoProSoft

Es un "Modelo de procesos para la industria de software". El Modelo de Procesos de Software fue desarrollado a solicitud de la Secretaría de economía para servir de base a la Norma Mexicana para la Industria de Desarrollo y mantenimiento de Software bajo el convenio con la Facultad de Ciencias, Universidad nacional Autónoma de México. Tomó referencias como: ISO9000: 2000, Nivel 2 y 3 de CMMI V.1.1, PMBOK y SWEBOK, (Allsoft, 2011).

MoProSoft está basado en procesos y orientado principalmente a las pequeñas y medianas empresas o a las áreas internas de desarrollo y mantenimiento de software. Por lo tanto, las pequeñas y medianas empresas que implementan MoProSoft trabajan bajo "buenas prácticas" . La adopción del modelo les permitirá elevar capacidad para ofrecer servicios con calidad y alcanzar niveles internacionales de competitividad.

Las características se refieren, como ya se menciono anteriormente que está orientado para el desarrollo y mantenimiento del software. MoProSoft facilita el cumplimiento de los requisitos de otros modelos como ISO 9000:2008 y CMMI, es fácil de entender y adoptar, es fácil de aplicar y ademas, comprende un documento de menos de 200 páginas lo que lo hace bastante práctico. Resulta acorde con la estructura de las organizaciones mexicanas con desarrollo o mantenimiento de software. Ayuda a mejorar los procesos, para contribuir a los objetivos de la organización. Tiene un bajo costo, tanto para su capacitación y, adopción como para su evaluación. Por lo que el modelo es muy accesible para que lo adopten las pequeñas y medianas empresas .

4.3.1 Categorías por las que está conformado MoProSoft son: Alta dirección, Gerencia y Operación

- Categoría alta dirección (DIR), contiene el proceso de Gestión de Negocio.
- Categoría Gerencia (GER), contiene los procesos de Gestión de Procesos, Gestión de Proyectos, Gestión de Recursos. Este último proceso contiene a su vez tres subprocesos que son: Recursos Humanos y Ambiente de Trabajo, Bienes Servicios e Infraestructura y Conocimiento de la Organización.
- Categoría Operación (OPE), esta categoría contiene los procesos de Administración de Proyectos Específicos y el de Desarrollo y Mantenimiento de Software.

4.3.2 MoProSoft tiene 9 procesos que se agrupan por categorías

Las categorías antes mencionadas contienen los procesos que conforman a MoProSoft como se muestra en la figura 4.2.

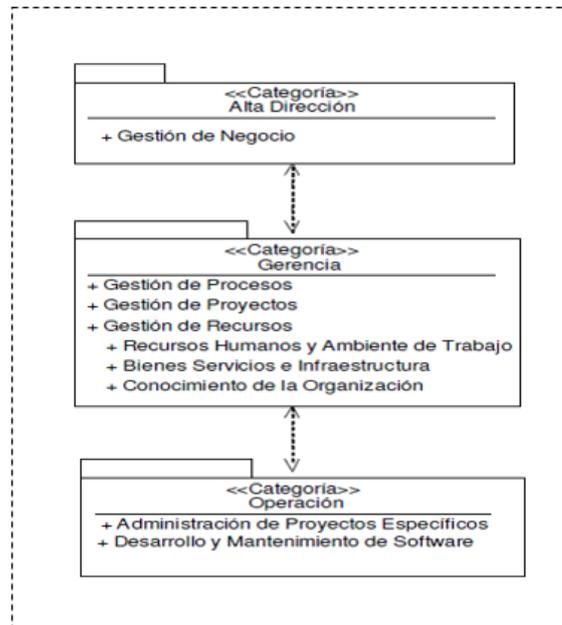


FIGURA 4.2. Diagrama de categoría de procesos.

En la figura 4.2 se observa la estructura del modelo de los procesos de MoProSoft que presenta tres categorías de procesos: Alta dirección, Gerencia y operación que reflejan la estructura de una organización. La categoría de alta dirección contiene el proceso de Gestión de Negocio. La categoría de Gerencia está integrada por los procesos de Gestión de Procesos, Gestión de Proyectos y Gestión de Recursos. Este último está constituido por los subprocesos de Recursos Humanos y Ambiente de Trabajo, Bienes, Servicios e Infraestructura y Conocimiento de la Organización. La categoría de Operación está integrada por los procesos de Administración de Proyectos Específicos y de Desarrollo y Mantenimiento de Software.

- Categoría Alta Dirección.
 - Gestión de negocios.
- Categoría Gerencia.
 - Gestión de Procesos.
 - Gestión de Proyectos.
 - Gestión de Recursos.
 - * Recursos Humanos y Ambiente de Trabajo.
 - * Bienes Servicios e Infraestructura.
 - * Conocimiento de la Organización

- Categoría Operación.
 - Administración de Proyectos Específicos.
 - Desarrollo y Mantenimiento de Software.

4.3.2.1 Gestión de Negocios

En este proceso el propósito de Gestión de Negocio es establecer la razón de ser de la organización, sus objetivos y las condiciones para lograrlos, para lo cual es necesario considerar las necesidades de los clientes, así como evaluar los resultados para poder proponer cambios que permitan la mejora continua.

4.3.2.2 Gestión de Procesos

Gestión de Procesos, el objetivo del proceso de Gestión de Procesos es establecer los procesos de la organización, en función de los procesos requeridos identificados en el plan estratégico. Así como definir, planificar, e implantar las actividades de mejora en los mismos.

4.3.2.3 Gestión de Proyectos

Gestión de Proyectos, para la Gestión de Proyectos el propósito es asegurar que los proyectos contribuyan al cumplimiento de los objetivos y estrategias de la organización.

4.3.2.4 Gestión de Recursos

Gestión de Recursos, en este proceso lo que se pretende es conseguir y dotar a la organización de los recursos humanos, infraestructura, ambiente de trabajo y proveedores, así como crear y mantener la base de conocimiento de la organización. La finalidad es apoyar el cumplimiento de los objetivos del plan estratégico de la organización.

- Recursos Humanos y Ambiente de Trabajo, el propósito de Recursos Humanos y Ambiente de Trabajo es proporcionar los recursos humanos adecuados para cumplir las responsabilidades asignadas a los roles dentro de la organización, así como la evaluación del ambiente de trabajo.
- Bienes Servicios e Infraestructura, el propósito de Bienes, Servicios e Infraestructura es proporcionar proveedores de bienes, servicios e infraestructura que satisfagan los requisitos de adquisición de los procesos y proyectos.
- Conocimiento de la Organización, el propósito de Conocimiento de la Organización es mantener disponible y administrar la base de conocimiento que contiene la información y los productos generados por la organización.

4.3.2.5 Administración de Proyectos Específicos

El propósito de la Administración de Proyectos Específicos es establecer y llevar a cabo sistemáticamente las actividades que permitan cumplir con los objetivos de un proyecto en tiempo y costo esperados.

4.3.2.6 Desarrollo y Mantenimiento de Software

El propósito de Desarrollo y Mantenimiento de Software es la realización sistemática de las actividades de análisis, diseño, construcción, integración y pruebas de productos de software nuevos o modificados cumpliendo con los requerimientos especificados.

4.3.3 Niveles de Madurez

MoProSoft se preocupa por ofrecer una guía a las organizaciones que quieren adoptar este modelo y lo hace presentando la versión de MoProSoft coloreada por niveles de madurez. Ahora bien, las partes coloreadas son las que sugieren el orden de implementación de las prácticas de los procesos de MoProSoft que van desde las prácticas básicas hasta las más avanzadas.

En la tabla 4.1 se muestra los niveles de capacidad de los procesos y los colores correspondientes a cada uno de ellos:

Tabla 4.1: Niveles de Madurez de MoProSoft

Nivel	Capacidad de proceso	Color
1	Realizado	Amarillo
2	Gestionado	Azul
3	Establecido	Verde
4	Predecible	Rosa
5	Optimizado	Ninguno

En la tabla 4.1 se visualiza la versión coloreada que fue generada para apoyar la implementación de MoProSoft durante las pruebas controladas en cuatro empresas y para la elaboración del paquete de evaluación de EvalProSoft.

A continuación se describe cada una de las partes coloreadas:

1. El color amarillo representa los productos y prácticas recomendadas a implementar primero.
2. Los productos y prácticas representadas por el color azul son las que ya están implementadas y controladas.
3. El color verde se refiere a el proceso realizado y gestionado se implementa por medio de un proceso definido.
4. El proceso establecido opera bajo límites definidos y conocidos y corresponde al color rosa.
5. El proceso predecible se mejora continuamente.

4.3.4 Relación entre procesos

Cada una de las áreas conformadas por las normas tienen un propósito: Gestión de proyectos, su propósito es asegurar que los proyectos contribuyan al cumplimiento de los objetivos y estrategias de la organización; el propósito de Recursos humanos y ambiente de trabajo se refiere a proporcionar los recursos humanos adecuados

para cumplir las responsabilidades asignadas a los roles, así como la evaluación del ambiente de trabajo; para administración de proyectos específicos tiene el propósito de establecer y llevar a cabo sistemáticamente las actividades que permitan cumplir con los objetivos de un proyecto en tiempo y costos esperados.

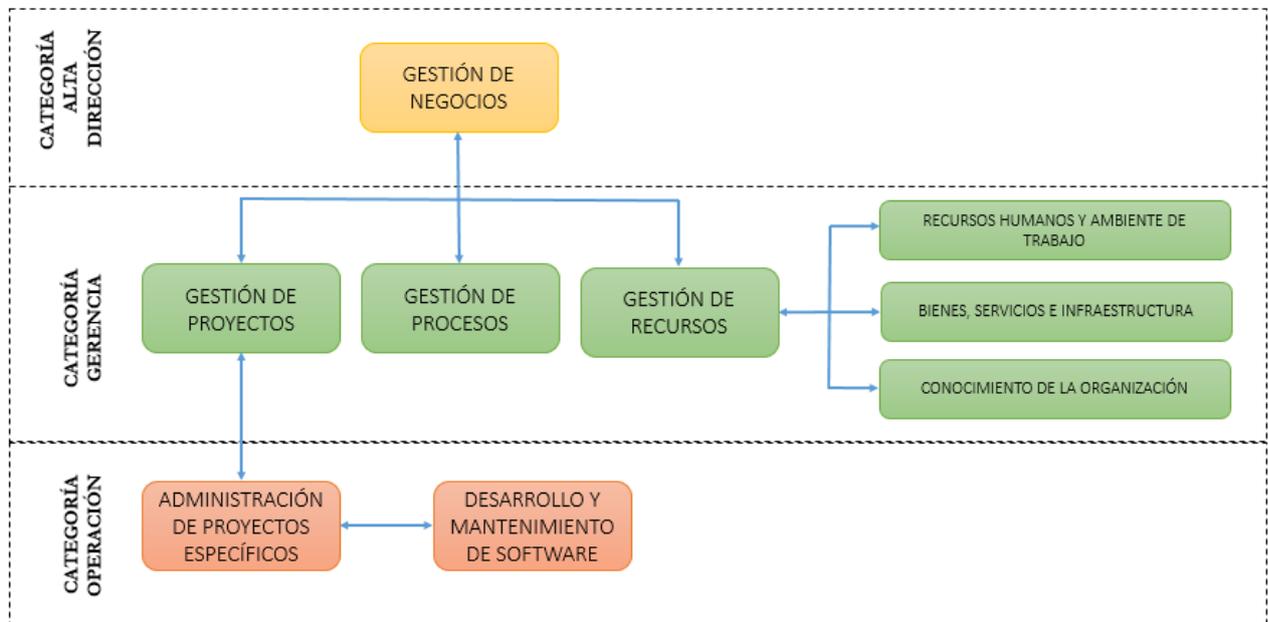


FIGURA 4.3. Relación entre procesos.

Fuente: Diseño propio

En la figura 4.3 se observa la relación que existe entre cada una de las gerencias los procesos de Gestión de Negocios tiene comunicación con los procesos de la Gerencia de Procesos, Gestión de Proyectos y Gestión de Recursos y es comunicación de bidireccional con Gestión de procesos. Por otro lado los procesos de Gestión de Recursos sostiene continua comunicación con los subprocesos: Recursos Humanos y Ambiente de Trabajo, Bienes servicios e Infraestructura y con Conocimiento de la Organización. A su vez los procesos de Gestión de proyectos mantienen comunicación directa con Administración de Proyectos Específicos y para finalizar Administración de proyectos específicos mantiene comunicación bidireccional con Desarrollo y Mantenimiento de Software.

4.3.5 Los roles

Rol, es responsable por un conjunto de actividades de uno o más procesos. Un rol puede ser asumido por una o más personas de tiempo parcial o tiempo completo.

Tabla 4.2: Roles en MoProSoft

Número	Rol
1	Cliente
2	Usuario
3	Grupo directivo
4	Responsable de proceso
5	Involucrado

Fuente: MoProSoft

La tabla 4.2 muestra los roles establecidos e por la norma MoProSoft, explicando cada uno de los roles expresados en ella.

- Cliente, es el que solicita un producto de software y financia el proyecto para su desarrollo o mantenimiento.
- Usuario, es el que va a utilizar el producto de software.
- Grupo directivo, son los que dirigen a una organización y son responsables por su funcionamiento exitoso.
- Responsable de proceso, es el encargado de la realización de las prácticas de un proceso y del cumplimiento de sus objetivos.
- Involucrado, otros roles con habilidades requeridas para la ejecución de actividades o tareas específicas. Por ejemplo: Analista, Programador, Revisor, entre otros.

4.3.6 Definición general del proceso

En esta sección se detalla con más profundidad las áreas de Gerencia de Proyectos, Gerencia de Recursos Humanos y Desarrollo y Mantenimiento de Software, fueron éstas las que se determinaron para trabajar y analizar su documentación. Una vez llevado a cabo el análisis se eligió la que de respuesta a los indicadores que ayudaran a la toma de decisiones.

4.3.6.1 Gerencia de Proyectos

El **nombre** del proceso es GES.2 Gestión de Proyectos, pertenece a la **categoría** Gerencia (GER). El **propósito** de esta gerencia es asegurar que los proyectos contribuyan al cumplimiento de los objetivos y estrategias de la organización. Para describir la gerencia de gestión de proyectos dice (Hanna Oktaba, Agosto 2005), que se ocupa de los proyectos internos, externos y también se ocupa de las oportunidades de proyectos de la organización. Para estos últimos lo que debe realizar es la generación y cierre de oportunidades de proyecto, llevar a cabo la presentación de propuesta y finalmente la firma de *Contrato*. Los proyectos internos se refieren a los proyectos para la propia organización o área interna de desarrollo), previo a su aprobación, se requiere evaluar diferentes alternativas de realización. Los proyectos externos e internos que ya fueron aprobados requiere de una planificación general y asignación de recursos, así como de un seguimiento y evaluación de desempeño.

La Gestión de Proyectos comprende la Planificación, realización, y la evaluación y control.

- Planificación

Se definen las actividades y los recursos requeridos por cada tipo de proyecto, que serán documentados en el plan de gestión de proyectos. Se elabora el plan de adquisiciones y capacitación. Establecer los Mecanismos de Comunicación con el cliente. Y por último se generan las Alternativas de Realización de proyectos Internos para los proyectos internos y se elige una nueva alternativa.

- Realización

Es la acción de llevar a cabo la ejecución de las actividades del plan de gestión de proyectos y su seguimiento, también de la mano el control de los Mecanismos de Comunicación con el cliente. Para cada proyecto se genera un Registro de Proyecto y la descripción de proyecto, se asigna al Responsable de Administración del Proyecto Específico y se entregan las metas cuantitativas para el proyecto. Ahora bien, en el caso de que el proyecto sea externo se elabora un contrato y a su termino se realiza el cierre de este. En esta etapa también se reciben y aprueban los Planes de Proyecto y se recolectan los Reportes de Seguimiento de los proyectos. Finalmente se recaudan los Comentarios y Quejas del Cliente.

- Evaluación y Control

Esta etapa comprende el análisis del Plan de Ventas, de los Reportes de Seguimiento y de los Comentarios y Quejas del Cliente, como consecuencia, se generan las Acciones Correctivas o Preventivas para los proyectos y se les da seguimiento hasta su cierre. Para mantener informado a Gestión de Negocio se genera el Reporte Cuantitativo y Cualitativo y el Reporte de Acciones Correctivas o Preventivas. Adicionalmente con base al Plan de Mediciones de Procesos se genera el Reporte de Mediciones y Sugerencias de Mejora en este proceso.

Los objetivos en primer lugar es cumplir con el Plan Estratégico de la organización mediante la generación e instrumentación de proyectos. Segundo mantener bajo control las actividades de Gestión de Proyectos mediante el cumplimiento del Plan de Gestión de Proyectos. Tercero proveer la información del desempeño de los proyectos a Gestión de Negocio mediante la generación del Reporte Cuantitativo y Cualitativo. Cuarto atender los comentarios y Quejas del Cliente mediante la definición y ejecución de Acciones Correctivas o Preventivas.

Entradas

Tabla 4.3: Entradas de la Gestión de Proyecto

Nombre	Fuente
Plan Estratégico: objetivos, estrategias, cartera de Proyectos, plan de Comunicación con el cliente	Gestión de Negocio
Plan de procesos: plan de Mediciones y procesos	Gestión de Procesos
Asignación de Recursos	Recursos Humanos y Ambiente de Trabajo
Plan del Proyecto	Administración de Proyectos Específicos
Reporte de Seguimiento	Administración de Proyectos Específicos

La tabla 4.3 se ilustra el nombre del producto o proceso que en este caso de Gerencia de Proyectos es el Plan Estratégico, Plan de Procesos, Asignación de Recursos, Plan del Proyecto, Reporte de Seguimiento. Se observa al lado izquierdo de la tabla cual es la fuente de referencia o el origen del producto o recurso, en este caso corresponde a Gestión de Negocio, Gestión de Procesos, Recursos Humanos y Ambiente de Trabajo, Administración de Proyectos Específicos y Administración de Proyectos Específicos respectivamente.

Salidas

Nombre	Descripción	Destino
Reporte Cuantitativo y Cualitativo	Elementos cuantitativos y cualitativos obtenidos a partir de la recopilación y análisis de Reportes de Seguimiento de los proyectos y del cumplimiento del Plan de Ventas.	Gestión de Negocio
Reporte de Acciones Correctivas o Preventivas Relacionadas con Clientes	Acciones establecidas para corregir o prever una desviación o problema sobre los Comentarios y Quejas del Cliente.	Gestión de Negocio
Reporte de Mediciones y Sugerencias de Mejora	Registro que contiene: mediciones de los indicadores del proceso de Gestión de Proyectos (ver Mediciones). Sugerencias de mejora al proceso de Gestión de Proyectos (métodos, herramientas, formatos, estándares, entre otros).	Gestión de Procesos
Continúa en la siguiente página		

Tabla 4.5 – continuación de la página previa

Nombre	Descripción	Destino
Plan de Adquisiciones y Capacitación	Descripción de los recursos y la capacitación requerida por los proyectos. Por ejemplo: perfil, cantidad de recursos humanos, fechas de incorporación al proyecto, requerimientos de capacitación, recursos de infraestructura, financieros, tecnológicos y materiales requeridos.	Gestión de Recursos
Contrato	Documento legal para la prestación de servicios con el cliente.	Conocimiento de la Organización
Registro de Proyecto	Información administrativa del proyecto, por ejemplo: nombre, responsable, fechas de inicio y terminación, cliente, precio, entre otros.	Conocimiento de la Organización
Lecciones Aprendidas	Registro de mejores prácticas, problemas recurrentes y experiencias exitosas, durante la implantación de este proceso.	Conocimiento de la Organización
Responsable de Administración del Proyecto Específico	Persona responsable de la administración de un proyecto específico.	Administración de Proyectos Específicos
Descripción del Proyecto	Descripción del propósito, del producto, objetivos, alcance, entregables, necesidad de negocio, supuestos y premisas, restricciones, entre otros.	Administración de Proyectos Específicos
Metas Cuantitativas para el Proyecto	Establece las metas cuantitativas que deberá cubrir el proyecto para tiempo y costo, entre otras.	Administración de Proyectos Específicos
Continua en la siguiente página		

Tabla 4.5 – continuación de la página previa

Nombre	Descripción	Destino
Acciones Correctivas o Preventivas	Acciones establecidas para corregir o prever una desviación o problema, tomando en cuenta los comentarios y quejas de clientes relacionadas con los proyectos.	Administración de Proyectos Específicos

En la tabla 4.5 se muestra las salidas de Gestión de Proyectos iniciando con el nombre del producto o recurso, posteriormente en la columna del medio se presenta una descripción y en la columna del lado izquierdo es el destino haciendo referencia al destinatario del producto o recurso.

Productos Internos

Tabla 4.6: Productos Internos de Gestión de Proyectos

Nombre	Fuente
Plan de Gestión de Proyectos	Plan de ventas: Contiene los objetivos, alcance, recursos, acciones y programa de trabajo para generar y cerrar oportunidades de proyectos. Plan de Proyectos: Descripción de las actividades para gestionar los proyectos externos e internos.
Acciones Correctivas o Preventivas	Acciones establecidas para corregir o prever una desviación o problema, relacionadas con la realización del Plan de Ventas o con los Mecanismos de Comunicación con los Clientes.
Comentarios y Quejas del Cliente	Registro de los comentarios y quejas del cliente.
Alternativas de Realización de Proyectos Internos	Descripción de diferentes opciones para llevar a cabo los proyectos internos. Incluye la decisión sobre la opción seleccionada.
Mecanismos de Comunicación con los Clientes	Información, medios, mensajes, responsables y mecanismos utilizados para comunicarse con los clientes
Reporte de Validación	Registro de participantes, fecha, lugar, duración y de defectos encontrados.

En la tabla 4.6 se observa cuáles son los productos internos de la Gerencia de proyectos en la columna derecha se presentan los nombres de cada uno de los productos generados y utilizados en el propio proceso y en la columna izquierda se presenta la descripción y características de cada uno de los productos.

4.3.6.2 Recursos Humanos y Ambiente de Trabajo

El **nombre** del proceso es GES.3.1 Recursos Humanos y Ambiente de Trabajo, pertenece a la **categoría** Gerencia (GER). El **propósito** de Recursos Humanos y Ambiente de Trabajo es proporcionar los recursos humanos adecuados para cumplir las responsabilidades asignadas a los roles dentro de la organización, así como la evaluación del ambiente de trabajo. (Hanna Oktaba, Agosto 2005), menciona en la descripción que en función del Plan Operativo de Recursos Humanos y Ambiente de Trabajo y Acciones Correctivas de Gestión de Recursos se realizan las actividades de preparación, instrumentación y generación de reportes.

- Preparación

Realiza las siguientes tareas: revisión del plan Operativo de Recursos Humanos y Ambiente de Trabajo y de las Acciones Correctivas. Definición de criterios. Elaboración del Plan de Capacitación. Elaboración de los formularios para la Evaluación de Desempeño y para la encuesta sobre el Ambiente de Trabajo.

- Instrumentación

En esta etapa se realizan las siguientes actividades: selección, asignación y aceptación de los recursos humanos (Asignación de recursos). Capacitación de recursos humanos de acuerdo a las necesidades actuales y futuras de los procesos y proyectos (Reporte de Capacitación). Evaluación de desempeño periódica de los recursos humanos asignados a los procesos y proyectos (Registro de Recursos Humanos). La evaluación del ambiente de trabajo y su registro es el (Reporte de Ambiente de Trabajo.)

- Generación de Reportes

Produce los siguientes reportes: reporte de Recursos Humanos Disponibles, Capacitación y Ambiente de Trabajo. Reporte de Mediciones y Sugerencias de Mejora.

Los **objetivos** para esta área en primer lugar es proveer a la organización de recursos humanos calificados mediante la selección y capacitación adecuada a los roles que se le asignen. El segundo de los objetivos se refiere a evaluar el ambiente de trabajo de la organización mediante la Encuesta sobre el Ambiente de Trabajo.

Las entradas es el producto interno o recurso que entran provenientes de una fuente referencia al origen del producto o recurso.

Entradas

Tabla 4.7: Entradas de Recursos Humanos y Ambiente de Trabajo

Nombre	Fuente
Plan de Procesos: plan de mediciones de procesos	Gestión de Procesos
Plan Operativo de Recursos Humanos y Ambiente de Trabajo	Gestión de Recursos
Acciones Correctivas	Gestión de Recursos

En la tabla 4.7 se describen las entradas de los productos de la Gerencia de Recursos Humanos y Ambiente de Trabajo en la columna derecha están los nombres del producto o recursos que entran a esta gerencia y en la columna izquierda se visualiza la fuente, es decir, la referencia al origen del producto o recurso.

Salidas

Tabla 4.8: Salidas de Recursos Humanos y Ambiente de Trabajo

Nombre	Descripción	Destino
Asignación de recursos	Personal competente con la formación apropiada, capacitación, habilidades y experiencia, de acuerdo con el rol a desempeñar.	Gestión de Procesos Gestión de Proyectos administración de Proyectos Específicos
Reporte de Mediciones y Sugerencias de Mejora	Registro que contiene: mediciones de los indicadores del proceso de Recursos Humanos y Ambiente de Trabajo (ver Mediciones), Sugerencias de mejora al proceso de Recursos Humanos y Ambiente de Trabajo (métodos, herramientas, formatos, estándares, entre otros).	Gestión de Procesos
Reporte de Recursos Humanos Disponibles, Capacitación y Ambiente de Trabajo	Reporte periódico que concentra la información de: estado de los recursos humanos Actividades de capacitación realizadas y planificadas Resultado de las encuestas sobre el ambiente de trabajo.	Gestión de Recursos
Plan de Capacitación	Descripción de las actividades de capacitación, incluyendo: cursos, talleres, calendario, instructores, logística, entre otros.	Conocimiento de la Organización.
Lecciones Aprendidas	Registro de mejores prácticas problemas recurrentes y experiencias exitosas, durante la implantación de este proceso.	Conocimiento de la Organización.

En la tabla 4.8 se muestra las salidas de Recursos Humanos y Ambiente de Trabajo iniciando con el nombre del producto o recurso, posteriormente en la columna del medio se presenta una descripción y en la columna del lado izquierdo es el destino haciendo referencia al destinatario del producto o recurso.

Productos Internos

Tabla 4.9: Productos Internos de Recursos Humanos y Ambiente de Trabajo

Nombre	Descripción
Registro de Recursos Humanos	Información del personal, incluyendo datos personales, formación, experiencia, roles asignados, capacitación, evaluaciones de desempeño, entre otros.
Reporte de Ambiente de Trabajo	Concentrado de las encuestas sobre las relaciones de trabajo, liderazgo, trabajo en equipo, horarios, infraestructura proporcionada, entre otros.
Reporte de Capacitación	Registro que puede contener los datos de capacitación proporcionada, fechas, número de asistentes, instructor, proveedor, evaluación, entre otros.
Evaluación de Desempeño	Evaluación del desempeño del recurso humano de acuerdo al rol asignado.
Encuesta sobre el Ambiente de Trabajo	Encuesta aplicada a los recursos humanos sobre el ambiente de trabajo.
Reporte(s) de Validación	Registro de participantes, fecha, lugar, duración y de defectos encontrados.

En la tabla 4.9 se observa cuáles son los productos internos de la Gerencia de Recursos Humanos y Ambiente de Trabajo en la columna derecha se presentan los nombres de cada uno de los productos generados y utilizados en el propio proceso y en la columna izquierda se presenta la descripción y características de cada uno de los productos.

4.4 CMMI

Uno de los retos a los que se enfrenta día a día las organizaciones de desarrollo de software es la administración de sus procesos y esto se debe a la falta de habilidad para llevarlo a cabo. CMMI es una guía que ayuda a las organizaciones a ganar el control sobre sus procesos y así desarrollar y mantener un buen software. CMMI para Desarrollo es un modelo de referencia que cubre las actividades para desarrollar tanto productos como servicios.

Es el acrónimo de (Capability Maturity Model Integration), en español Modelo de Madurez de Capacidad Integrado se refiere a los modelos que contienen las mejores prácticas que ayudan a las organizaciones a mejorar sus procesos. CMMI es un modelo de buenas prácticas utilizado como referencia mundial de calidad en el sector de las TI. CMMI fue publicado por el Software Engineering Institute (SEI) en 1994 con la finalidad de compartir con la industria las buenas prácticas o prácticas que han demostrado en uso ser exitosas para reducir riesgos habituales (L. Piedras Muñoz).

Para aumentar la habilidad de la administración de los procesos las organizaciones necesitan seguir las prácticas que CMMI proporciona. Poniendo en práctica la planeación, ingeniería y administración de desarrollo y mantenimiento de software lograran aumentar su habilidad.

4.4.1 Objetivo de CMMI

CMMI tiene el objetivo de orientar a las organizaciones para elegir las estrategias de mejora y hacer madurar el proceso actual y al mismo tiempo puedan identificar los puntos donde pudieran tener debilidades así poder atacarlas y de esta manera mejorar el proceso y la calidad del software.

Muy específicamente CMMI consiste en un conjunto de prácticas relacionadas con las actividades de desarrollo y mantenimiento aplicadas a productos y servicios, dichas actividades se enfocan en construir, mantener y gestionar el producto final.

Las organizaciones que aplican el modelo CMMI pueden garantizar los siguientes beneficios:

- Calidad
Desarrollar sistemas y servicios de la calidad con alta eficiencia y competitividad.

- Exactitud
Planificar el desarrollo, incluyendo personal, procesos, consumibles y equipos (plazos, costos, entregables)

- Proactividad
Controlar y prevenir los riesgos inherentes a proyectos de desarrollo y mantenimiento

Un nivel de madurez bien definida con una evolución hacia la consecución de un proceso de software maduro. Cada nivel de madurez está asociado a un conjunto de procesos que deben ser cubiertos como práctica estándar de una organización en cuestión para alcanzar dicho nivel de madurez.

El diagrama presentado en la figura 4.4 muestra los modelos de CMMI y los cinco niveles de madurez

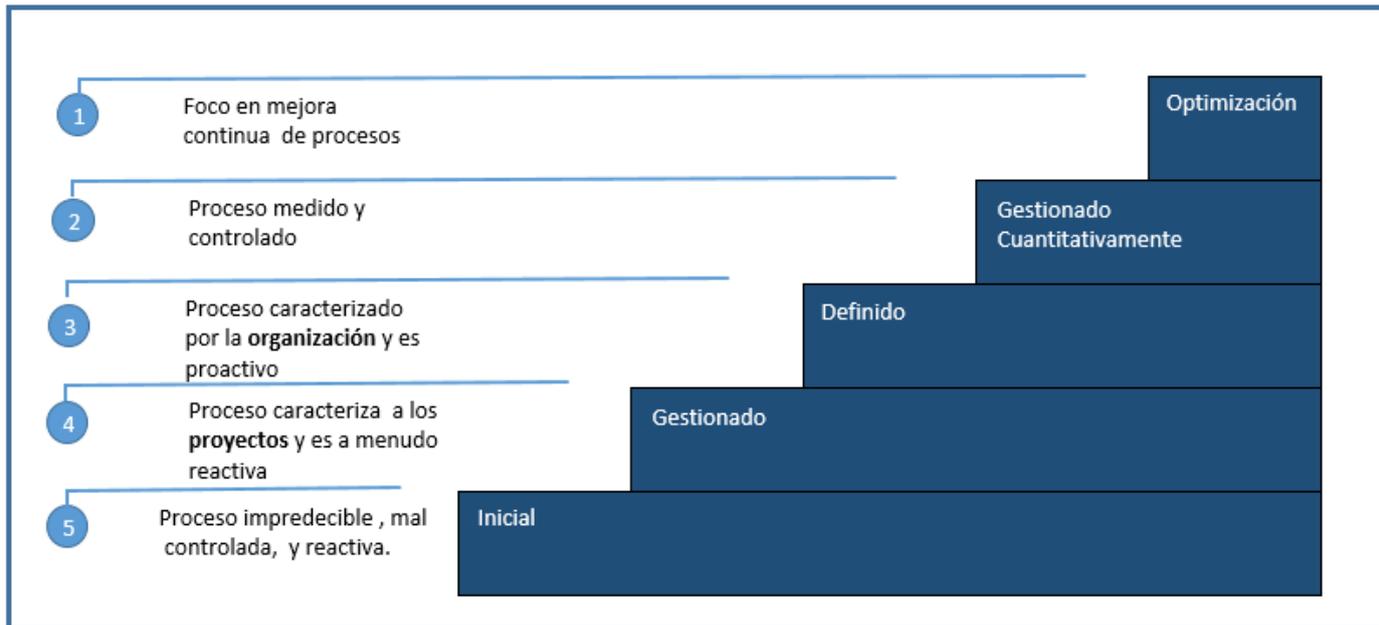


FIGURA 4.4. Niveles de mejora CMMI

Los niveles de madurez (ML) en CMMI a detalle:

- Madurez Básica Intermedia:

1. Inicial: el proceso se improvisa y depende de individuos. En el nivel de madurez 1, los procesos suelen ser ad hoc y caótico. La organización normalmente no proporciona un entorno estable. El éxito de estas organizaciones dependen de la competencia y de la disposición de las personas de la organización y no en el uso de procesos probados.
2. Gestionado: hay gestión de proyectos con mecanismos de control, aunque con procesos no necesariamente acordados. En el nivel de madurez 2, la organización ha logrado todos los objetivos genéricos y específicos del nivel de madurez 2 áreas de proceso. En otras palabras, los proyectos de la organización han asegurado que los requisitos son gestionados y de que los procesos se planifican, realizan, medido y controlado.
3. Definido: existen procesos estándares compartidos y guías de adaptación adecuadas. En el nivel de madurez 3, la organización ha alcanzado todos los objetivos específicos y de las áreas de proceso asignadas a los niveles de madurez 2 y 3.
 - Alta madurez:
4. Gestionado Cuantitativamente: entendimiento del rendimiento de proceso y una gestión cuantitativa. En el nivel de madurez 4, se seleccionan los que contribuyen de forma significativa al rendimiento del proceso en general. Estos sub-procesos están controlados mediante técnicas estadísticas y otras

técnicas cuantitativas.

5. Optimización: entendimiento del proceso y una gestión cuantitativa. En el nivel de madurez 5, una organización ha logrado todos los goalsof el proceso zonas asignadas a los niveles de madurez 2, 3, 4 y 5, y los objetivos genéricos asignados a los niveles de madurez 2 y 3.

4.4.2 Áreas de Procesos Cubiertos por CMMI-DEV

CMMI-DEV abarca 22 áreas de procesos asociados a cuatro categorías como se muestra en la figura 4.5:

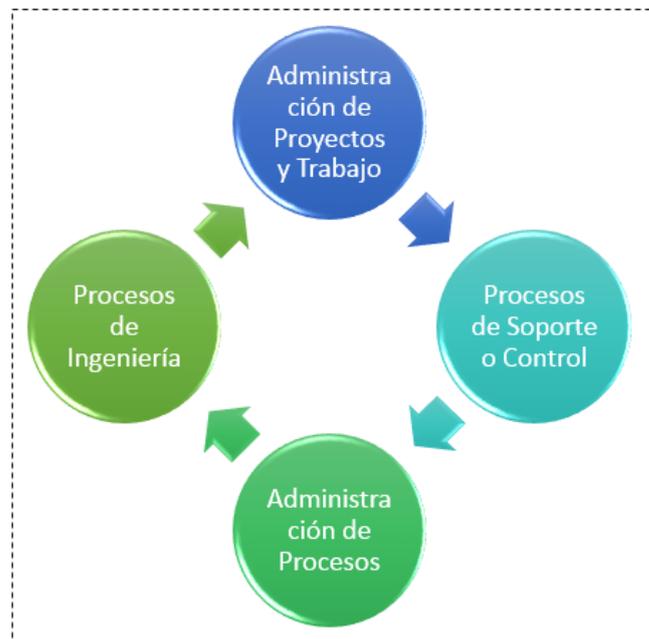


FIGURA 4.5. Categorías de Proceso en CMMI-DEV.

- Administración de Proyectos y Trabajo: comprende la planificación, control, prevención, subcontratación y gestión de amenazas que pueden afectar el desarrollo de un proyecto
- Procesos de Soporte o Control: son actividades y funciones transversales, tendientes a aumentar la calidad la satisfacción y eficacia del desarrollo
- Procesos de Ingeniería: se refiere la ingeniería de desarrollo de sistemas, gestión de servicios y adquisición de sistemas.
- Administración de Procesos: asociados con la mejora continua del proceso, capacidades y rendimiento en base a las necesidades del negocio.

En la Tabla número 4.6 se muestra la asociación del conjunto de procesos con cada nivel de madurez correspondiente.

Nivel de Madurez	Administración de Procesos	Administración de Proyectos	Procesos de Ingeniería	Procesos de Soporte
5	OPM – Gestión del Performance de la Organización			CAR – Análisis y Resolución Causal
4	OPP - Performance de Procesos de la Organización	QPM – Gestión Cuantitativa de Proyectos		
3	OPF – Enfoque en Procesos de la Organización OPD – Definición de Procesos de la Organización OT - Entrenamiento de la Organización	IPM – Gestión de Proyectos Integrada RSKM – Gestión de Riesgos	RD – Desarrollo de Requerimientos TS – Soluciones Técnicas VER – Verificación VAL – Validación PI – Integración de Productos	DAR – Análisis de Decisiones y Resoluciones
2		PP - Planeación de Proyectos PMC - Seguimiento de Proyectos REQM - Administración de Requerimientos SAM – Gestión de Acuerdos con Proveedores		CM - Administración de la configuración PPQA - Aseguramiento de la Calidad MA - Medición y Análisis
1	-	-	-	-

FIGURA 4.6. Áreas de proceso de CMMI por categoría de proceso y nivel de madurez.

4.4.2.1 Administración de Procesos

- Procesos incluidos en esta categoría son:
 - Organizational Process Performance (**OPP**)
 - Organizational Process Definition (**OPD**)
 - Organizational Process Focus (**OPF**)
 - Organizational Performance Management (**OPM**)
 - Organizational Training (**OT**)

4.4.2.2 Administración de Proyectos y Trabajo

- Procesos incluidos en esta categoría son:
 - Integrated Project Management (**IPM**)
 - Project Monitoring and Control (**PMC**)
 - Project Planning (**PP**)
 - Requirements Management (**REQM**)
 - Risk Management (**RSKM**)
 - Quantitative Project Management (**QPM**)
 - Supplier Agreement Management (**SAM**)

4.4.2.3 Procesos de Ingeniería

- Procesos incluidos en esta categoría son:
 - Requirements Development (**RD**)

- Technical Solution (**TS**)
- Product Integration (**PI**)
- Verification (**VER**)
- Validation (**VAL**)

4.4.2.4 Procesos de Soporte o control

- Procesos incluidos en esta categoría son:
 - Causal Analysis and Resolution (**CAR**)
 - Configuration Management (**CM**)
 - Decision Analysis and Resolution (**DAR**)
 - Measurement and Analysis (**MA**)
 - Process and Product Quality Assurance (**PPQA**)

Como nota adicional se agrega que cada área crea documentación adicional a las que establece las normas, para fortalecer las actividades de los procesos que se realizan día a día en la empresa.

Finanzas es una gerencia adicional en la empresa, no está determinada por MoProSoft, pero es la área clave para que la empresa pueda tomar decisiones en sus necesidades.

En la empresa MBN se enfocan principalmente en las prácticas que tienen que ver con el proceso de desarrollo de software, las cuales están asociadas al nivel de madures 2 del modelo y se refieren como CMMI-Dev 2.

Desde el inicio de operaciones la empresa MBN se ha caracterizado por adoptar la calidad como una forma de vida, ajustándose a distintas dinámicas y metodologías, comunicación efectiva, liderazgo, trabajo en equipo, entre otras.

La empresa MBN, ofrece servicios de Consultoría Especializada, Desarrollo de Software, OutSourcing y Centro Avanzado de Atención y Soporte, su visión es ser el mejor socio estratégico de sus clientes, colaboradores, proveedores, accionistas y asociados para ser valorados por la sociedad e instituciones. Su misión es ser una empresa innovadora comprometida con sus clientes en su logro de ser más rentables, competitivos y exitosos. Incorporan el talento que brinda el mejor servicio en la implementación tecnológica de soluciones de negocio.

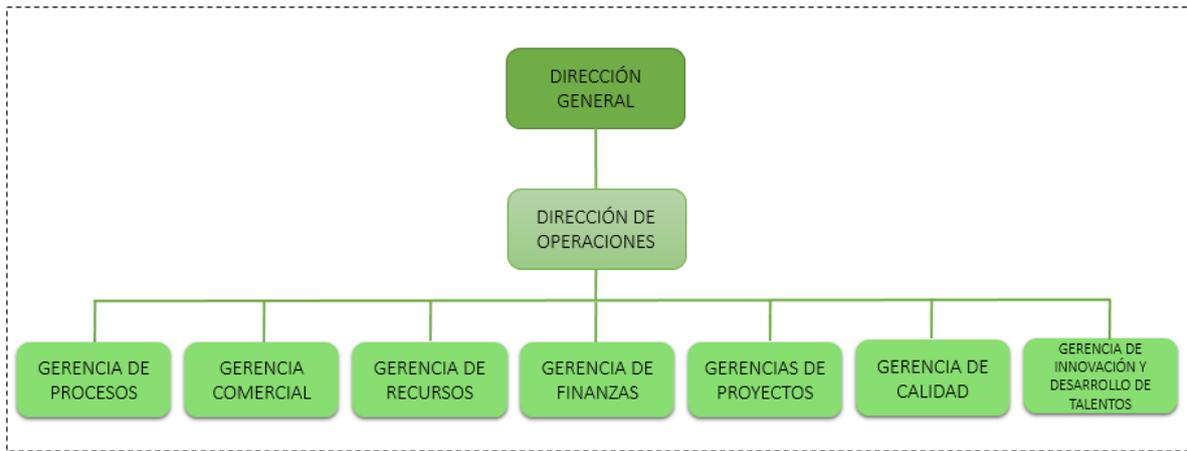


FIGURA 4.7. Organigrama de la empresa MBN.

La empresa está estructurada bajo la norma MoProSoft-059-NYCE-2005 como se muestra en la figura 4.7.

5.1 Introducción

Una vez diseñado el modelo lógico de las tablas dimensión y hechos, es momento de diseñar la base de datos real. Se procede a probar la base de datos con datos.

5.2 Instalación de las herramientas

Cuando se tienen los datos disponibles que son cargados al almacén de datos, se realiza la instalación de las diferentes herramientas empleadas para el desarrollo del almacén. La herramienta principal para los administradores de la base de datos, es una plataforma libre que permite el análisis y la administración de datos empresariales. Ya que provee escalabilidad, disponibilidad y seguridad al tiempo que simplifica la creación e implementación.

5.3 Recursos

Para fines de desarrollo de este proyecto, se conto con ciertos recursos, los cuales se muestran a continuación:

Tabla 5.1: Recursos

Concepto	Detalle
Fuente de datos	MoProSoft y CMMI.
Hardware	Procesador: Intel inside CORE i7 Memoria RAM: 10 GB Dico duro:
Software	Workbench Xampp.

5.4 Datos para el llenado de las tablas dimensión y hechos

En base a las tablas dimensión que se obtuvieron en el modelo conceptual se procedio a llenar cada una de las tablas, como ya se menciona no se realiza en este trabajo el proceso de ETL asi que el llenado es de manera artesanal, es decir manualmete. Para esta tarea se hace uso de la hoja de calculo Excel.

5.4.1 Datos para la dimensión colaboradores

La tabla colaboradores contiene información requerida y que con anterioridad fue analizada y valorada por el usuario de la empresa como se muestra en la figura 5.1.

A	B	C	D	E	F	G	H	I
claveColaborador	claveColaborador	nombreColaborador	apellidoPaterno	apellidoMaterno	Salario	estatus	fechaIngreso	fechaTermino
1	600001	Olivia	Roldan	Osorno	12000	activo	2015-09-18	1000-01-01
2	600002	Marisol	Vázquez	Aquino	3000	activo	2017-05-28	1000-01-01
3	600003	Alejandro	Osorno	Mendez	5000	activo	2015-01-15	1000-01-01
4	600004	Lucia	Ortega	Tapia	8000	inactivo	2017-03-30	02/06/2017
5	600005	Lazaro	Mendez	Guzman	4500	activo	2016-08-20	1000-01-01
6	600006	Liliana	Suarez	Suarez	7000	inactivo	2015-04-28	30/12/2017
7	600007	Saul	Lopez	Campos	26000	activo	2015-07-30	1000-01-01
8	600008	David	Hernandez	Ramirez	10000	inactivo	2016-02-14	25/02/2017
9	600009	Angelica	Contreras	Juarez	5000	activo	2015-08-21	1000-01-01
10	600010	Lorena	Fernandez	Diaz	5000	activo	2017-12-06	1000-01-01

FIGURA 5.1. Datos para dimensión colaboradores.

En la figura superior se muestran los datos con los que se cargan en la tabla dimensión colaboradores. Inicando con la clave primaria de la tabla, seguida de la clave del colaborador, así como su salario con el que fue registrado al ingresar a la empresa, el estatus actual y finalmente se muestra su fecha de ingreso y termino con la relación de ésta.

5.4.2 Datos para la dimensión nómina

Se presentan los datos de la tabla dimensión nómina, son los valores que se cargaran en la base de datos como se observa en la figura 5.2, los datos corresponden a los datos que determinan cuál es el gasto generado a la nómina.

A	B	C	D	E	F
idDimensionNomina	mesNomina	primeraQuincenaMes	fechaPagoNomina	segundaQuincenaMes	fechaPagoNomina
1	1	0	2015-01-09	4000	2015-01-23
2	2	4000	2015-02-10	4000	2015-02-25
3	3	4000	2015-03-10	4000	2015-03-25
4	4	4000	2015-04-10	4000	2015-04-24
5	5	4000	2015-05-08	4000	2015-05-25
6	5	2500	2015-05-08	2500	2015-05-25
7	6	4000	2015-06-10	4000	2015-06-25
8	6	2500	2015-06-10	2500	2015-06-25
9	7	4000	2015-07-02	4000	2015-07-16
10	7	2500	2015-07-02	2500	2015-07-16
11	8	4000	2015-08-10	4000	2015-08-25
12	8	2500	2015-08-10	2500	2015-08-25
13	8	4000	2015-08-10	4000	2015-08-25
14	8	0	2015-09-10	3000	2015-08-25
15	9	4000	2015-09-10	4000	2015-09-25
16	9	2500	2015-09-10	2500	2015-09-25
17	9	4000	2015-09-10	4000	2015-09-25
18	9	3000	2015-09-10	3000	2015-09-25
19	9	0	2015-09-10	2250	2015-09-25
20	10	4000	2015-10-09	4000	2015-10-23
21	10	2500	2015-10-09	2500	2015-10-23
22	10	4000	2015-10-09	4000	2015-10-23
23	10	0	2015-10-09	3000	2015-10-23

FIGURA 5.2. Datos para dimensión nómina.

En la figura 5.2 se observa que datos corresponden a cada uno de los campos que compone la tabla de nómina que corresponden a el mes en que se paga la nómina, al monto total pagado en la primera quincena del mes, en que fecha se pago esa primera quincena, el monto total pagado en la segunda quincena del mes y en que fecha se realizao el pago de esa segunda quincena.

5.4.3 Datos para la dimensión costos operación

La figura 5.3 corresponde a los datos de la tabla costos operación aqui se muestran los datos contenidos que corresponde al pago de costos del colaborador al ser integrado al desarrollo de un proyecto.

A	B	C	D	E	F
idCosteo	NominalMSS	NominaHonorarios	Prestaciones	Viaticos	fechaAsignacionCosteo
1	12000	12000	2000	1000	2015-01-12
2	22800	18000	2500	0	2015-06-05
3	33000	15000	1000	2000	2015-09-05
4	42950	1700	1000	2000	2015-09-05
5	53200	16000	1000	2000	2015-09-05
6	62800	18000	4000	1800	2016-01-02
7	72400	14000	4000	1800	2016-01-02
8	83000	15000	3000	2000	2016-05-12
9	92950	1700	3000	2000	2016-05-12
10	102800	18000	3000	2000	2016-05-12
11	113000	15000	2800	0	2016-10-12
12	122400	14000	2800	0	2016-10-12
13	132759	15000	2800	0	2016-10-12
14	142759	15000	3000	2800	2017-01-30
15	153200	16000	3000	2800	2017-01-30
16	163200	16000	1000	2000	2017-07-25
17	173100	16500	1000	2000	2017-07-25
18	182650	13500	1000	2000	2017-07-25
19	193000	15000	2800	0	2017-08-12
20	203200	16000	3000	0	2017-08-12
21	212900	12100	2900	0	2017-08-12
22	222759	15000	2800	0	2017-09-02
23	232100	16500	1000	0	2017-08-02

FIGURA 5.3. Datos para dimensión costos operación.

Estos datos que se observan corresponden a los campos propuestos para la tabla de los costos de operación correspondientes a pagos de nómina IMSS, nómina honorarios, prestaciones, viaticos y la fecha en que son asignados los gastos.

5.4.4 Datos para la dimensión proyectos

El la figura 5.4 se muestran los datos que el usuario considero necesarios e indispensables para formar parte de la tabla proyectos.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	idProyectos	cliente	fechaRegistro	lineaNegocio	tipoCapacitacion	capacitacion	costoCapacitacion	cantidadPersonas	fechaCapacitacion	tipoRecurso	conceptoRecurso	cantidad	fechaEntrega	costoInvertido	Proyecto	tipoProyectos
2	1	Coca-cola	2015-01-12	CABD	curso	MoProSoft	8000	8	2015-01-18	material	acapizaron	1	2015-01-12	8000	mercurio	externo
3	2	Coca-cola	2015-01-12	DMS	curso	Java	1200	12	2015-06-15	infraestructura	escritorios	1	2015-06-05	1200	venus	externo
4	3	P&G	2015-09-05	DMPO	capacitacion	Androit	25345	12	2015-09-10	equipo	computadora	1	2015-09-05	25345	Tierra	externo
5	4	Fabricas de P&G	2016-01-02	VRT	taller	C++	24300	5	2016-01-08	tecnologia	servidor	1	2016-01-02	19123	Marte	externo
6	5	P&G	2015-09-05	SER COMP	curso	Java	15000	3	2016-05-15	equipo	computadora	6	2016-05-12	243000	Jupiter	interno
7	6	Fabricas de P&G	2016-01-02	DMS	curso	Phyton	2340	6	2016-10-13	equipo	computadora	5	2016-10-30	15000	Saturno	externo
8	7	Telecable	2017-01-30	CABD	autoEstudio	CMMI	9000	1	2017-01-30	tecnologia	servidor pro	4	2017-01-30	17000	Urano	externo
9	8	Soriana	2017-07-25	DMPO	curso	MoProSoft	4290	3	2017-07-29	material	lapiceros	25	2017-07-25	234	Neptuno	interno
10	9	Coca-cola	2015-01-12	DMS	curso	modelado de	2220	5	2017-08-18	infraestructura	sillas	12	2017-08-12	9000	Pluton	externo
11	10	P&G	2015-09-05	DMS	curso	Phyton	3980	9	2017-09-12	material	libretas	50	2017-09-02	429	luna	externo

FIGURA 5.4. Datos para dimensión proyectos.

Los datos que se observan en la tabla proyectos corresponden a nombre del cliente, la fecha en que se registro el cliente, la línea de negocio a la que pertenece un proyecto, tipo de capacitación requerida en un proyecto, el nombre de la capacitación, el costo que genero la capacitación, la cantidad de personas que asistieron a la capacitación, la fecha en que se llevó a cabo la capacitación, los recursos que se necesitaron para desarrollar el proyecto, la descripción del recurso requerido, la cantidad requerida, la fecha en que se entrega el recurso, el costo invertido en el recurso, nombre del proyecto y el tipo del proyecto.

5.4.5 Datos para la dimensión tiempo

En la figura 5.5 ilustra los datos que se cargarón en la base de datos, las fechas van desde el año 2015 al 2017 es el período que la empresa requiere saber los últimos movimientos con respecto a sus proyectos.

	A	B	C	D	E
1	idDimension_Tiempo	Fecha	Año	Trimestre	Mes
2	1	2015-01-01	2015	T1	1
3	2	2015-01-02	2015	T1	1
4	3	2015-01-03	2015	T1	1
5	4	2015-01-04	2015	T1	1
6	5	2015-01-05	2015	T1	1
7	6	2015-01-06	2015	T1	1
8	7	2015-01-07	2015	T1	1
9	8	2015-01-08	2015	T1	1
10	9	2015-01-09	2015	T1	1
11	10	2015-01-10	2015	T1	1
12	11	2015-01-11	2015	T1	1
13	12	2015-01-12	2015	T1	1
14	13	2015-01-13	2015	T1	1
15	14	2015-01-14	2015	T1	1
16	15	2015-01-15	2015	T1	1
17	16	2015-01-16	2015	T1	1
18	17	2015-01-17	2015	T1	1
19	18	2015-01-18	2015	T1	1
20	19	2015-01-19	2015	T1	1
21	20	2015-01-20	2015	T1	1
22	21	2015-01-21	2015	T1	1
23	22	2015-01-22	2015	T1	1
24	23	2015-01-23	2015	T1	1

FIGURA 5.5. Datos para dimensión tiempo.

Los datos que se consideraron son los que corresponden fecha, año, trimestre y el número del mes.

5.4.6 Datos para la tabla hechos

La figura 5.6 muestran los datos que seran cargados en la tabla hechos de la base de datos. Son datos que reflejan el cruce de datos entre las diferentes tablas dimensiones.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R		
idDimension	cantidadVendida	costoVenta																	
1	1	1	1	1	1	1	1	12	1	1	1	1	1	1	1	1	1	330000.00	
1	2	2	2	1	2	7	1	156	2	1	1	1	1	2	2	2	1	185000.00	
2	3	3	3	2	3	16	1	248	3	2	2	2	7	3	3	1	141000.00		
2	3	3	3	3	4	17	1	248	3	3	1	2	7	3	3	1	141000.00		
2	3	3	3	4	5	18	1	248	3	2	1	2	7	3	3	1	141000.00		
3	4	4	4	1	6	35	1	367	4	1	1	3	3	4	4	1	270000.00		
3	4	4	4	5	7	39	1	367	4	4	1	3	3	4	4	1	270000.00		
2	5	5	5	1	8	58	2	498	5	1	1	1	2	3	3	1	225000.00		
2	5	5	5	3	9	60	2	498	5	1	1	1	2	3	3	1	225000.00		
2	5	5	5	6	10	63	2	498	5	5	2	1	2	3	3	1	225000.00		
3	2	6	6	2	11	91	1	651	6	2	2	1	4	3	3	1	370000.00		
3	2	6	6	5	12	94	1	651	6	5	1	1	4	3	3	1	370000.00		
3	2	6	6	7	13	96	1	651	6	6	1	1	4	3	3	1	370000.00		
5	1	7	7	1	14	111	1	761	7	1	1	4	5	4	4	1	155000.00		
5	1	7	7	4	15	114	1	761	7	2	1	4	5	4	4	1	155000.00		
4	3	8	8	5	16	161	2	937	8	4	1	1	1	1	5	1	384000.00		
4	3	8	8	8	17	164	2	937	8	3	2	1	1	1	5	1	384000.00		
4	3	8	8	9	18	165	2	937	8	1	1	1	1	1	5	1	384000.00		
1	2	9	9	2	19	167	1	955	9	2	2	1	6	2	6	1	237000.00		
1	2	9	9	4	20	169	1	955	9	2	1	1	6	2	6	1	237000.00		
1	2	9	9	10	21	211	1	955	9	7	1	1	6	2	6	1	237000.00		
2	2	10	10	7	22	181	1	976	10	6	1	1	4	1	7	1	122000.00		
2	2	10	10	8	23	182	1	976	10	3	2	1	4	1	7	1	122000.00		
2	2	10	10	10	24	211	1	976	10	7	1	1	4	1	7	1	122000.00		

FIGURA 5.6. Datos para tabla hechos.

La carga de los datos se refiere a las claves primarias correspondientes a cada una de las tablas dimensión como es: idDimensionclientes que es la clave primaria de la tabla dimensión clientes, al igual que la clave primaria de las tablas: idDimensionlineaNegocio, idDimensionrecursosProyecto, idDimensionentrenamientos, idDimensioncolaboradores, idDimensioncostosOperacion, idDimensionnomina, idDimensiontipoProyectos, idDimensiontiempo, idDimensionproyectosDesarrollados, idDimensionequipoTrabajo, idDimensionestatus, idDimensiontipoEntrenamientos, idDimensiontemasEntrenamientos, idDimensionrecursos, idDimensiontipoRecursos; cabe resaltar que el conjunto de las claves primarias de las tablas representan la clave primaria de la tabla hecho. En segundo plano la tabla muestra las medidas o los indicadores: cantidadVendida y costoVenta.

5.5 Creación y llenado de las tablas en el almacén de datos

Una vez que se tienen los datos disponibles que son cargados al almacén de datos y de la instalación de la herramintetas, se crean y cargan las tablas de la base de datos. Los datos contenidos en las tablas, son datos proporcionados por la empresa. Estas tablas permiten almacenar los datos relacionados a los colaboradores, nómina, costos operación, proyectos, tiempo y los hechos; sólo se encuentran aquellos datos que se requieren para dar respuesta a las preguntas solicitadas por el cliente es decir solo aquellos datos útiles.

5.6 Realizando consultas bajo el esquema estrella

Una vez concluida la creación de las tablas y el llenado de éstas es posible visualizar como quedo conformada la base de datos dimensional bajo el esquema estrella.

proyecto	colaboradorNomina	fecha
martes	Alejandro Osorno Mendez	2016-01-02
juovier	Alejandro Osorno Mendez	2016-05-12
urano	Alejandro Osorno Mendez	2017-01-30
tierra	Liliana Suarez Suarez	2015-09-05
saturno	Liliana Suarez Suarez	2016-10-12
pluton	Liliana Suarez Suarez	2017-08-12
tierra	Saul Lopez Campos	2015-09-05
juovier	Saul Lopez Campos	2016-05-12
tierra	Anoelica Contreras Juarez	2015-09-05
urano	Anoelica Contreras Juarez	2017-01-30
pluton	Anoelica Contreras Juarez	2017-08-12
martes	Olivia Roldan Osorno	2016-01-02
saturno	Olivia Roldan Osorno	2016-10-12
neptuno	Olivia Roldan Osorno	2017-07-25

FIGURA 5.9. Resultado de la consulta2.

En esta consulta número dos el cliente solicita saber que colaboradores estuvieron participando en determinado proyecto y en un tiempo determinado. Entonces se visualiza el nombre del proyecto, el nombre del colaborador y la fecha en la que se lleva a cabo el proyecto.

Consulta 3. ¿Gasto generado a la nómina por proyecto durante un tiempo determinado?

En la figura 5.16 se muestra el resultado de la consulta número tres.

proyecto	colaboradorNomina	fecha
martes	Alejandro Osorno Mendez	2016-01-02
juovier	Alejandro Osorno Mendez	2016-05-12
urano	Alejandro Osorno Mendez	2017-01-30
tierra	Liliana Suarez Suarez	2015-09-05
saturno	Liliana Suarez Suarez	2016-10-12
pluton	Liliana Suarez Suarez	2017-08-12
tierra	Saul Lopez Campos	2015-09-05
juovier	Saul Lopez Campos	2016-05-12
tierra	Anoelica Contreras Juarez	2015-09-05
urano	Anoelica Contreras Juarez	2017-01-30
pluton	Anoelica Contreras Juarez	2017-08-12
martes	Olivia Roldan Osorno	2016-01-02
saturno	Olivia Roldan Osorno	2016-10-12
neptuno	Olivia Roldan Osorno	2017-07-25

FIGURA 5.10. Resultado de la consulta3.

En esta consulta se requiere visualizar el gasto que genera a la nómina durante un periodo determinado de tiempo.

Consulta 4. ¿Cantidad vendida de proyectos a cada cliente durante un tiempo determinado?

En la figura 5.17 se muestra el resultado de la consulta número cuatro.

Proyecto	cliente	fecha
jupiter	P&G	2016-05-12
luna	P&G	2017-09-02
tierra	P&G	2015-09-05

FIGURA 5.11. Resultado de la consulta4.

El cliente requiere saber cuantos proyectos se vendieron a un solo cliente y en un tiempo determinado y así saber cual fue el cliente con el que más se colaboro.

Consulta 5. ¿Línea de negocio más rentable durante un tiempo determinado?

En la figura 5.18 se muestra el resultado de la consulta número cinco.

lineaNegocios	año	totalLineaNegocios
CABD	2015	1
DMPO	2015	3
DMS	2015	1

FIGURA 5.12. Resultado de la consulta5.

Como ya se explicó en el Capítulo 4 cada proyecto que se dispara en la empresa pertenece a una clasificación de línea de negocio propuesta por la empresa. Lo que el cliente desea saber es cuál línea de negocio que más proyectos tuvo durante un tiempo determinado. En la imagen muestra el nombre de la línea de negocio, el año y el total de proyectos por cada línea de negocio.

Consulta 6. ¿Costo de operación por colaborador durante un tiempo determinado?

En la figura 5.19 se muestra el resultado de la consulta número seis.

colaborador	nominaMSS	nominaHonorarios	prestaciones	viaticos	año	TotalCosteo
Alejandro Osomo Mendez	2000	12000	2000	1000	2015	17000
Alejandro Osomo Mendez	2759	15000	3000	2800	2017	23559
Alejandro Osomo Mendez	2800	18000	2500	0	2015	23300
Alejandro Osomo Mendez	2800	18000	4000	1800	2016	26600
Alejandro Osomo Mendez	3000	15000	3000	2000	2016	23000
Anaelica Contreras Juarez	3200	16000	1000	2000	2015	22200
Anaelica Contreras Juarez	3200	16000	3000	0	2017	22200
Anaelica Contreras Juarez	3200	16000	3000	2800	2017	25000
David Hernandez Ramirez	2800	18000	3000	2000	2016	25800
Lazaro Mendez Guzman	2759	15000	2800	0	2016	4118
Liliana Suarez Suarez	3000	15000	1000	2000	2015	21000
Liliana Suarez Suarez	3000	15000	2800	0	2016	41600

FIGURA 5.13. Resultado de la consulta6.

Se desea obtener información del costo operación invertido por colaborador en un tiempo determinado, es saber cuanto invirtió la empresa por colaborador al integrarse a un proyecto en específico.

5.7 Realizando consultas bajo el esquema copo de nieve

De la misma manera y tomando en cuenta las mismas necesidades de la información, se construyó el modelo copo de nieve.

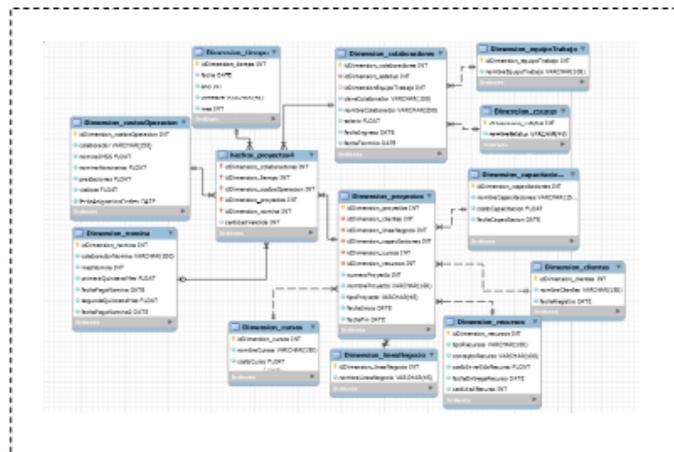


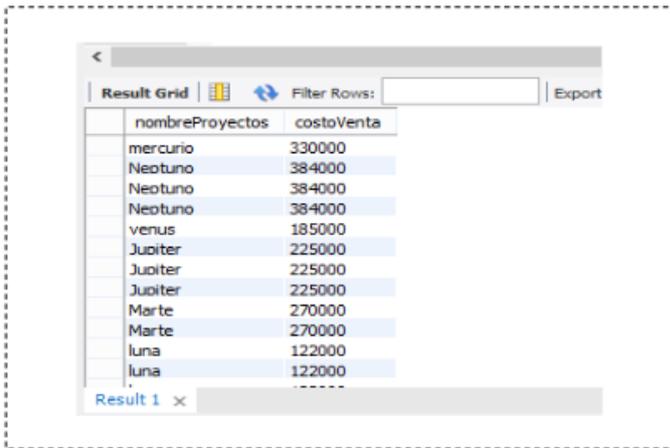
FIGURA 5.14. Esquema copo de nieve.

En la figura 5.20 se muestra el modelo copo de nieve en el que en primera instancia se tiene la tabla hechos proyectos con sus indicadores. Posteriormente se visualizan las tablas dimensiones algunas de ellas con jerarquías.

Para el modelo copo de nieve se muestran las imágenes de los resultados obtenidos:

Consulta 1. ¿Costo de venta por proyectos desarrollados, durante un tiempo determinado?

El resultado de la consulta número 1 se muestra a continuación en la cual se observa el nombre de los proyectos con sus respectivos costos que es lo requerido por el cliente.



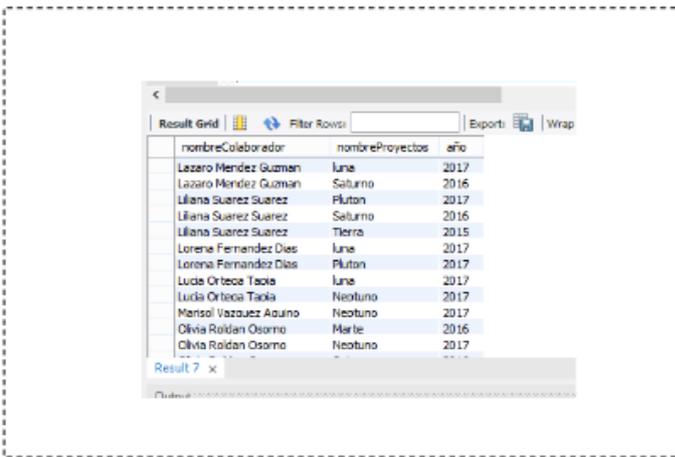
nombreProyectos	costoVenta
mercurio	330000
Neotuno	384000
Neotuno	384000
Neotuno	384000
venus	185000
Jupiter	225000
Jupiter	225000
Jupiter	225000
Marte	270000
Marte	270000
luna	122000
luna	122000

FIGURA 5.15. Resultado de la consulta 1 esquema copo de nieve.

La figura número 5.21 cual permite ver que proyecto es el que genero mayor ingreso y este es un indicador que ofrece gran valor para el cliente.

Consulta 2. ¿Colaboradores que participaron en el desarrollo de proyectos durante un tiempo determinado?

Para la consulta número dos, los resultados que muestra el esquema copo de nieve son los datos a la petición solicitada por el cliente.



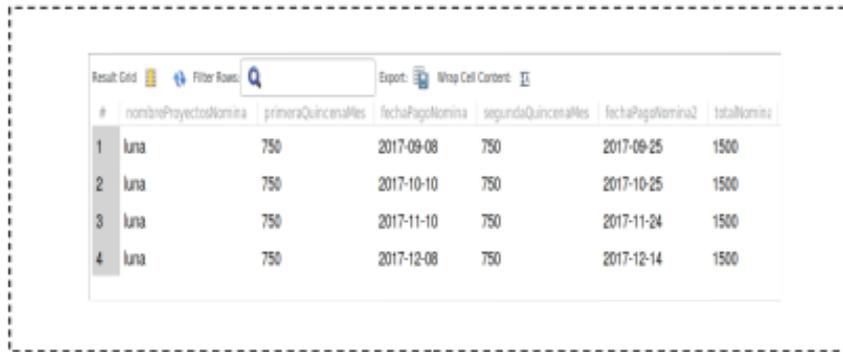
nombreColaborador	nombreProyectos	año
Lazaro Mendez Guzman	luna	2017
Lazaro Mendez Guzman	Saturno	2016
Liliana Suarez Suarez	Pluton	2017
Liliana Suarez Suarez	Saturno	2016
Liliana Suarez Suarez	Tierra	2015
Lorena Fernandez Dias	luna	2017
Lorena Fernandez Dias	Pluton	2017
Lucia Orteza Taola	luna	2017
Lucia Orteza Taola	Neotuno	2017
Mansol Vazquez Auano	Neotuno	2017
Olivia Roldan Osorno	Marte	2016
Olivia Roldan Osorno	Neotuno	2017

FIGURA 5.16. Resultado de la consulta 2 esquema copo de nieve.

En la figura 5.22 el resultado permite visualizar que colaboradores estuvieron integrados en determinados proyectos, este indicador permite a la empresa tener más control a la hora de realizar la asignación de sus empleados a los proyectos.

Consulta 3. ¿Gasto generado a la nomina por proyecto durante un tiempo determinado?

En la figura 5.23 y dando respuesta a la consulta número tres se presentan los resultados obtenidos.

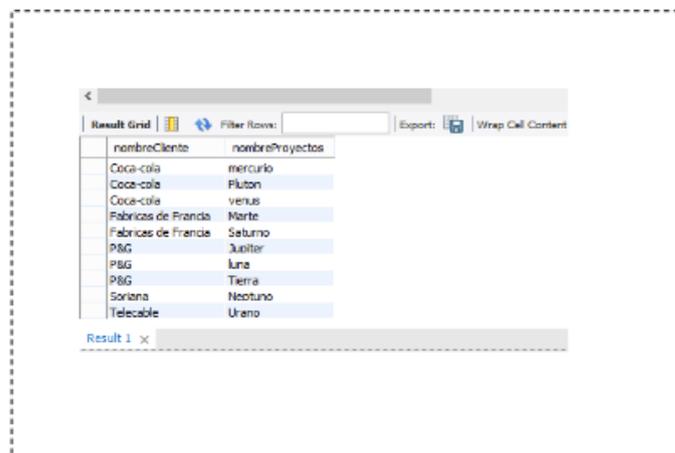


#	nombreProyectoNomina	primeraQuincenaMes	fechaPagoNomina	segundaQuincenaMes	fechaPagoNomina2	totalNomina
1	luna	750	2017-09-08	750	2017-09-25	1500
2	luna	750	2017-10-10	750	2017-10-25	1500
3	luna	750	2017-11-10	750	2017-11-24	1500
4	luna	750	2017-12-08	750	2017-12-14	1500

FIGURA 5.17. Resultado de la consulta3.

En esta consulta se muestra el detalle de los pagos de nómina de los colaboradores que fueron parte de un proyecto y cual fue el gasto generado.

Consulta 4. ¿Cantidad vendida de proyectos a cada cliente durante un tiempo determinado? En la figura 5.24 se observan los datos que dan resultado a la consulta número 4.



nombreCliente	nombreProyectos
Coca-cola	mercurio
Coca-cola	Pluton
Coca-cola	venus
Fabricas de Francia	Marte
Fabricas de Francia	Saturno
P&G	Jupiter
P&G	Luna
P&G	Tierra
Soriana	Neotuno
Telecable	Urano

FIGURA 5.18. Resultado de la consulta 4 esquema copo de nieve.

Permite conocer cuales fueron los clientes más frecuentes, esto le permite a la empresa establecer estrategias que permitan que se le ofrezcan promociones, obsequios lo cual implicaría un aumento en las ventas.

Consulta 5. ¿Línea de negocio más rentable durante un tiempo determinado? En la figura 5.25 se ilustran los resultados a la pregunta mencionada anteriormente, a solicitud del cliente.

nombreLineaNegocios	nombreProyectos	año
CABD	mercurio	2015
CABD	Urano	2017
DNFO	Neptuno	2017
DNFO	Tierra	2015
DHS	Luna	2017
DHS	Pluton	2017
DHS	Saturno	2016
DHS	venus	2015
SER COMP	Jupiter	2016
VIRT	Marte	2016

FIGURA 5.19. Resultado de la consulta 5 esquema copo de nieve.

Permite conocer cual línea de negocio es la que más ingresos está generando a la empresa, al mismo tiempo, esto permite conocer cuales están siendo menos rentables para poder realizar una investigación sobre porqué y atender las causas las cuales están generando estos resultados.

Consulta 6. ¿Costo de operación por colaborador durante un tiempo determinado?

Finalmente en la figura 5.26 se presenta la información obtenida para dar respuesta a la consulta número 6.

nombreColaborador	nominaMOS	nominaHonorarios	prestaciones	viaticos	año	sumaCosteo
Alejandro Osorno Mendez	2000	12000	2000	1000	2015	17000
Alejandro Osorno Mendez	2800	18000	2500	0	2015	23300

FIGURA 5.20. Resultado de la consulta 6 esquema copo de nieve.

El resultado de esta consulta permite saber a detalle cuanto invierte la empresa por empleado al ser integrado a un proyecto.

5.8 Resultados de tiempo de consulta entre esquema estrella y esquema copo de nieve

Los resultados de los tiempos de las consultas en los dos esquemas presentados se muestran en la figura número 5.21.

Resultados		
MySQL		
Windows		
Prueba	Estrella	Copo nieve
Consulta1	0	0.141
Consulta2	0	0.094
Consulta3	0	0.087
Consulta4	0	0.547
Consulta5	0	0.422
Consulta6	0	0.063

FIGURA 5.21. Comparativa de tiempos de consulta de copo de nieve y estrella.

A continuación se muestran las gráficas comparativas de los tiempos entre esquemas estrella y copo de nieve con cada una de las consultas. Para cada consulta se generó una gráfica. Las gráficas muestran los tiempos representados en segundos. Para el esquema estrella los tiempos van en relación a que la base de datos es desnormalizada y en el caso del esquema copo de nieve es una base de datos normalizada es decir cuenta con jerarquías en algunas de sus tablas dimensión, y esa es la razón por la cual ésta última genera más tiempo en las consultas.

En la figura 5.28 se observa el tiempo consumido en segundos para la primera consulta.



FIGURA 5.22. Gráfica de la consulta 1.

Para la consulta 1 se hizo uso de las tabla proyectos y la tabla tiempo. Para el modelo estrella el resultado en

tiempo de consulta es 0 segundos dado que es desnormalizado el modelo y para el modelo copo de nieve es su estructura normalizada consume un total 0.141 segundos en la consulta.

En la figura 5.29 se observa el tiempo consumido en segundos para la segunda consulta.

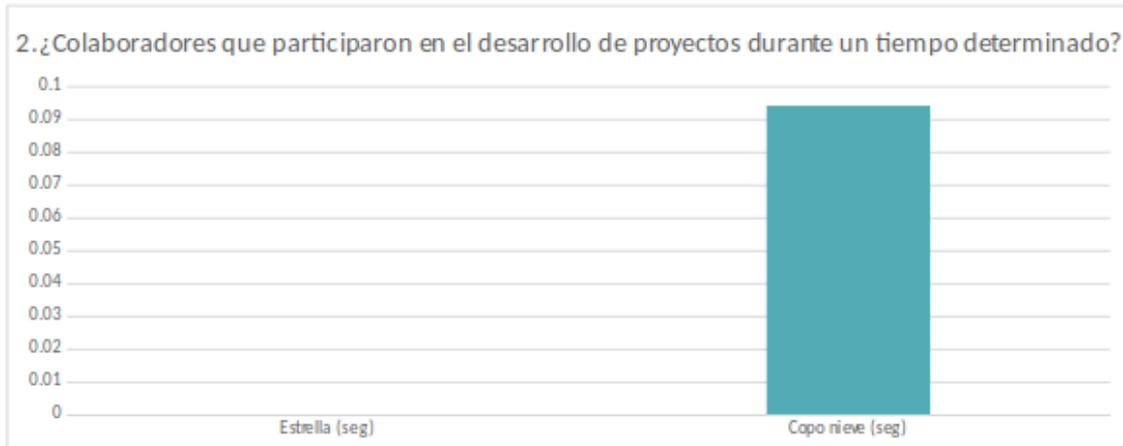


FIGURA 5.23. Gráfica de la consulta 2.

En la consulta 2 se requirió de las tablas colaboradores, la tabla proyectos, y la tabla tiempo. De igual manera en el esquema estrella el resultado en segundos para la consulta 2 fue 0 y 0.094 segundos para copo de nieve.

En la figura 5.30 se observa el tiempo consumido en segundos para la tercera consulta.



FIGURA 5.24. Gráfica de la consulta 3.

En la consulta 3 las tablas nómina, proyectos y tiempo son las que dan resultado a la pregunta tres. Para el esquema estrella esta consulta consumió un tiempo de 0 segundos, por lo que para el esquema copo de nieve requiero un tiempo de 0.087 segundos.

En la figura 5.31 se observa el tiempo consumido en segundos para la cuarta consulta.



FIGURA 5.25. Gráfica de la consulta 4.

Para esta consulta 4 las tablas que dieron respuesta a ésta fueron cliente, proyectos y tiempo. Para el esquema estrella dio un total de 0 segundos y en copo de nieve fue 0.547 segundos.

En la figura 5.32 se observa el tiempo consumido en segundos para la quinta consulta.

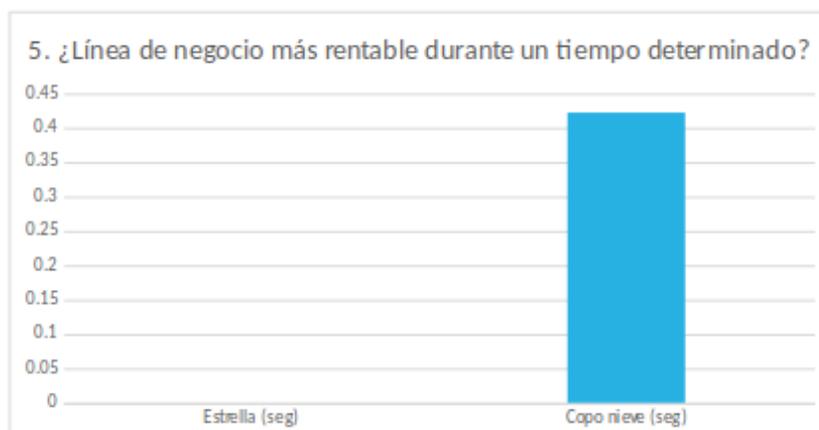


FIGURA 5.26. Gráfica de la consulta 5.

Las tablas que se requirieron para dar respuesta a la consulta 5 fueron línea negocios y tiempo. Para el esquema estrella los resultados fueron 0 segundos y para copo de nieve 0.0422 segundos.

En la figura 5.33 se observa el tiempo consumido en segundos para la sexta y última consulta.



FIGURA 5.27. Gráfica de la consulta 6.

Para esta consulta las tablas utilizadas fueron colaboradores y costos operación. Los tiempos consumidos para los esquemas estrella y copo de nieve fueron de 0 y 0.063 segundos respectivamente.

CONCLUSIONES

Las conclusiones derivadas de este trabajo de tesis confirman que dado a que las empresas de tamaño pequeño como lo son las pequeñas y medianas empresas, al igual que las grandes empresas también producen datos como el resultado de sus procesos que realizan día a día. La pequeñas y medianas empresas pueden adoptar a la inteligencia de negocios como una solución ante la problemática del análisis e integración de sus datos y a la toma de decisiones. Las herramientas inteligencia de negocios fueron creadas para las grandes empresas puesto que ellas pueden solventar el gasto que representa la implementación de éstas. Sin embargo y afortunadamente para la pequeñas y medianas empresas existe la versión libre de esta solución. Les ofrecen todos sus beneficios para el manejo y administración de sus datos; y así poder obtener el mayor provecho de sus datos. Pueden convertirlos en información y esa información en conocimiento para poder tomar mejores decisiones. y sin tener que realizar un gasto que pudiera afectar la economía de la empresa.

La inteligencia de negocios cuenta con una arquitectura básica que está diseñada para respaldar los indicadores clave de desempeño. En la clasificación de esta arquitectura se puede mencionar que está constituida por tres entornos: i) extracción, transformación y carga (ETL, por sus siglas en inglés extract, transform and load), ii) almacén de datos y iii) análisis empresarial. En la empresa Mirecle Business Network S.A de C.V, (MBN), se implementó la arquitectura de inteligencia de negocios, pero en este trabajo de tesis solo se desarrolló el almacén de datos; esta fase de la arquitectura es la más compleja en cuanto a desarrollo puesto que de ella depende dar respuesta a los indicadores a los que la empresa desea tener resultados y contentarse con la herramienta de inteligencia de negocios.

MBN, es una empresa que toma a la calidad como un estilo de vida es por ello que adopta a los estándares MoProSoft para darle estructura a la organización y CMMI como muestra de las buenas prácticas en el desarrollo de software. Es por medio de estos estándares que son producidos los datos de la empresa. Dentro de la empresa existe la gerencia de finanzas que es la columna vertebral de MBN; pero no se encuentra dentro de los estándares adoptados, y ésta también produce datos. Es de estas fuentes de donde se toman los datos para llevar a cabo el modelado de los datos.

Con respecto a la problemática de la empresa a la hora de generar informes finales y enfrentarse ante datos heterogéneos, en distintos formatos y aislados se comprueba que si los resuelve precisamente con la arquitectura de la inteligencia de negocios. Es por ello que el objetivo general de este trabajo de tesis si se cumple, dado que al modelar los datos se obtiene la integración de éstos obteniendo información útil y ayudando así a la empresa en la toma de decisiones.

Existe literatura que soporta el hecho de que se han realizado trabajos similares al de este trabajo de tesis. Todo trabajo de investigación debe estar sustentado y comparado por trabajos ya realizados. Por esa razón se hace una investigación de artículos que son sometidos bajo una estricta y rigurosa evaluación por expertos en la materia y que son aceptados en revistas científicas y congresos tanto nacionales como internacionales. Son artículos que explican que metodologías, herramientas, técnicas, procesos de inteligencia de negocios se emplearon para llevar a cabo la construcción de sus almacenes de datos es por ello que fueron de interés para la realización de esta tesis. Todos estos artículos presentados son lo último, la más avanzado en almacenes de datos. Ahora bien, para tener un mejor entendimiento de lo que se desarrolla es necesario incluir todos los conceptos necesarios para tener idea de lo que se está trabajando, tales conceptos como inteligencia de negocios, almacén de datos, los modelos que existen del almacén de datos, entre otros.

De igual manera se realizó un estudio de algunas de las metodologías existentes para el desarrollo de almacenes de datos entre las que se estudiaron fueron: CRISP-DM, Hefesto, Kimball, Inmon y Rapid Warehousing Methodology (RWN). Tomando en cuenta sus principales características se concluyó que la más apta para implementarse fue la metodología Hefesto puesto que esta enfocada a los almacenes de datos, solo consta de cuatro etapas, no genera un gasto para la implementación, esta enfocada a las pequeñas y medianas empresas, entre otras. Para la empresa MBN se le propuso hacer uso de software libre para llevar a cabo el desarrollo del almacén de datos y no le afectara económicamente para la obtención de alguna licencia, lo que se pudo comprobar es que las herramientas como: MySQL Workbench, XAMPP y Windows que son de licencia libre dieron excelentes resultados para el desarrollo de este trabajo de tesis.

Llevar a cabo el desarrollo de un almacén de datos es una tarea difícil, para realizar su desarrollo e implementación es necesario elegir una metodología; se necesita de un modelo conceptual que está basado en los requerimientos de información, un modelo lógico que está basado en una tecnología de base de datos. No existe una metodología que cubra todas las necesidades para el desarrollo de un almacén de datos, es por ello que se elige la más adecuada de acuerdo al tipo de empresa que se va a implementar el almacén de datos. Actualmente las metodologías híbridas son las que están teniendo más éxito a la hora de desarrollar e implementar un almacén de datos. Como se menciona en el párrafo anterior el modelo propuesto para el desarrollo del almacén de datos fue la metodología Hefesto.

El modelo propuesto permite desarrollar sin problema cada uno de los objetivos planteados:

- Analizar la información que es obtenida bajo los estándares de calidad.

La metodología propuesta ayudo al cumplimiento de este objetivo dado que la primera etapa llamada: Análisis de requerimientos, dio lugar a investigar sobre las necesidades de información de la empresa y en base a esas preguntas de negocio se llevo a cabo el análisis de las fuentes de datos, que como ya se aclaro las fuentes de datos son los artefactos producidos bajo las normas MoProSoft, CMMI y los de la gerencia de finanzas.

- Diseñar el modelado multidimensional de los datos.

Los datos obtenidos bajo los estándares MoProSoft(solo las áreas modeladas fueron GPY, RHAT, DMS Y APE), CMMI y los de la gerencia de finanzas son los que fueron modelados bajo los esquemas estrella y copo de nieve. Ambos modelos son recomendados para la empresa MBN, dado que esta catalogada como una pequeña y mediana empresa y por tal razón no genera grandes volúmenes de datos. Por esta razón de la cantidad de datos procesada el modelo estrella es perfecto ya que no genera demasiado tiempo entre consultas, es un esquema desnormalizado y la única desventaja es la redundancia de los datos. El esquema copo de nieve de igual manera y por la misma razón de la cantidad de datos procesada es adecuada para la empresa MBN la desventaja de este esquema es que es un esquema normalizado y consume más tiempo en el momento de generar las consultas.

- Modelo listo y preparado para la explotación de la información.

Finalmente los datos son cargados al almacén de datos y muestran de manera exitosa las respuestas a las preguntas de negocios realizadas por el cliente. Razón por la cual el almacén de datos esta listo para ser conectado a cualquier herramienta de inteligencia de negocios.

De acuerdo a las características de la empresa MBN, fue posible desarrollar el almacén de datos, partiendo de los resultados presentados de los dos esquemas propuestos, se determino que el esquema conveniente de adoptar fue el esquema estrella.

Trabajos a futuro

- Proponer indicadores que involucren a todas las áreas de la empresa.
- En determinado momento y contando con toda la información generada por todas las áreas de la empresa crear un almacén de datos con cruce de información de acuerdo a la estructura de la empresa
- Modelo listo y preparado para la explotación de la información.
- Realizar extracción de datos por medio de ETL.
- De acuerdo al estudio realizado de herramientas para inteligencia de negocios y presentado por: Lourdes Suárez, en Abril de 2018, publicado en la: International Journal of science and Research, con el título de: Comparative Analysis of Business Intelligence Tools for SME's, proponer la implementación de una herramienta de inteligencia de negocios para que los datos sean presentados y visualizados al usuario final por medio de gráficas, cuadros de mando, informes, entre otras.

APPENDIX



PUBLICACIÓN

INTERNATIONAL JOURNAL OF SCIENCE & RESEARCH



Online ISSN:2319-7064

Certificate of Publication

www.ijsr.net

This is to Certify that the paper ID: ART20181354 entitled

Comparative Analysis of Business Intelligence Tools for SMEs

Authored

By

Lourdes Suarez

has been published in Volume 7 Issue 4, April 2018

in

International Journal of Science and Research (IJSR)

The mentioned paper is measured upto the required standard.



Editor in Chief, International Journal of Science and Research, India



TOEFL ITP Score Report

Name of Institution: TOEFL TESTING CENTER PACHUCA

Name: SUAREZ MENDEZ LOURDES

Student Number:

DOB: 12/06/1980

Sex: F Degree:

Times Taken TOEFL: 1

Native Country: Mexico

Native Language: Spanish

Scaled Scores:

Listening Comprehension: 43

Test Date: 03/23/2018

Structure & Written Expression: 47

Form: TOEFL ITP

Reading Comprehension: 48

Total Score: 460

 ETS TOEFL ITP

The face of this document has a security background. The back contains a watermark. Hold at an angle to view.

The TOEFL® ITP Assessment Series is designed to be used for placement, progress monitoring, and exit purposes. TOEFL® ITP scores can also be used for admissions to programs and institutions where English is not the dominant language of instruction for content courses. Learn more at www.ets.org/toefl_itp/use.

119631-16573 • FB317C100 • Printed in U.S.A.

I.N. 770462

Copyright © 2012 by Educational Testing Service.

Student's File Copy
Do Not Copy



Miracle Business Network S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio

CARTA DE ACEPTACIÓN

Tlaxcala Tlax a 15 de mayo del 2017

MTRO. FELIPE PASCUAL ROSARIO AGUIRRE

DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO

PRESENTE

Por medio del presente le comunico que el C. Lourdes Liliana Suárez Méndez alumno del Instituto Tecnológico de Apizaco de la carrera de Ingeniería en Tecnologías de la Información y Comunicaciones, ha sido aceptado para realizar sus **Estancias profesionales bajo el proyecto "ANÁLISIS DE SOFTWARE ORIENTADO A LA INTELIGENCIA EMPRESARIAL PARA DETERMINAR LA ADMINISTRACIÓN DE LA INFORMACIÓN"** en nuestra empresa **MIRACLE BUSINESS NETWORK SA DE CV** específicamente en el área de **Dirección General** en el periodo que comprende del 02 de enero del 2017 al 30 de Junio de 2017 .

El C. Lourdes Liliana Suárez Méndez estará bajo la responsabilidad de la Mtra. Patricia Trejo Responsable de Conocimiento de la Organización de esta empresa, con el firme objetivo de poner en práctica sus conocimientos y desarrollar habilidades que engrandezcan su perfil profesional y faciliten su posicionamiento dentro del ámbito laboral.

Es para nosotros un placer trabajar con alumnos de excelencia, con deseos de superación y actitud positiva ante los retos globales que hacen de cada esfuerzo un paso más hacia el éxito.

ATENTAMENTE

Lic. Ariana Ramírez Moreno

Responsable de Recursos Humanos
y Ambiente de Trabajo



Miracle Business Network, S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio.

15 MAY 2017

R.F.C.MBN060316R19
Calle 37 No. 216 B La Loma Xicohténcatl
Tlaxcala, Tlax. C.P. 90070 Tel. 01 246 416 45 08



Miracle Business Network S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio

Asunto: CARTA DE LIBERACIÓN

Tlaxcala, Tlax., 11 de septiembre del 2017

Mtro. Felipe Pascual Rosario Aguirre

DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO

Con atención a:

Dr. José Federico Casco Vázquez

JEFE DEL DEPARTAMENTO DE DIVISIÓN DE ESTUDIOS

DE POSGRADO E INVESTIGACIÓN

P R E S E N T E

Me permito comunicar a usted, que la Ing. Lourdes Liliana Suárez Méndez, estudiante de la Maestría en sistemas computacionales en el Tecnológico Nacional de México, Instituto Tecnológico de Apizaco, con número de matrícula **M11370869**, ha concluido una **Estancia Técnica**, realizando el proyecto de **“Desarrollo del modelado de un almacén de datos como parte integral de un sistema de inteligencia de negocios”**., durante el período comprendido del 02 de Enero del 2017 al 30 de Junio del 2017, los días Miércoles, Jueves y Viernes con un horario de 9:00 a 18:00 horas, teniendo como jefe inmediato a la Maestra Patricia Trejo Xelhuantzi.

Sin más por el momento, se extiende la presente carta de liberación para los efectos correspondientes.

ATENTAMENTE

Lic. Vidalia Flores Mojica

Responsable de Recursos Humanos y Ambiente de Trabajo.



Miracle Business Network, S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio.

11 SEP 2017

R.F.C.MBN060316RI9

Calle 37 No. 216 B La Loma Xicohténcatl
Tlaxcala, Tlax. C.P. 90070 Tel. 01 246 416 45 08



Tlaxcala, Tlax., 08 Marzo de 2018

Asunto: Constancia de Satisfacción.

Mtro. Felipe Pascual Rosario Aguirre

DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO

Con atención a:

AT'N: Dr. José Federico Casco Vázquez

JEFE DEL DEPARTAMENTO DE DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN

P R E S E N T E:

Sirva la presente para enviarle un cordial saludo y notificarle que posterior a la recepción del proyecto de tesis de la Ing. Lourdes Liliana Suárez Méndez, alumna de la Maestría en Sistemas Computacionales, con número de matrícula **M11370869** de la institución que usted destacadamente dirige, se incluyó con el proyecto que lleva como título: **"DESARROLLO DEL MODELADO DE UN ALMACÉN DE DATOS COMO PARTE INTEGRAL DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS"**.

Siendo este desarrollado bajo la dirección del Dr. José Crispín Hernández Hernández catedrático de la citada maestría.

En virtud de que sea cubierto satisfactoriamente los objetivos establecidos del citado proyecto.

Tenemos a bien dar constancia de que dicho proyecto de tesis cubre y satisface las expectativas planteadas al inicio de este proyecto.

Sin más por el momento, se extiende la presente carta de satisfacción para los efectos correspondientes.



Miracle Business Network, S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio.

ATENTAMENTE

08 MAR 2018

R.F.C.MBN060316RI9

Calle 37 No. 216 B La Loma Xicohténcatl
Tlaxcala, Tlax. C.P. 90070 Tel. 01 246 416 45 00

Lic. Vidalia Flores Mojica

Responsable de Recursos Humanos y Ambiente de Trabajo.

BIBLIOGRAFÍA

- [1] (SEI), S. E. (2010). CMMI para Desarrollo, Versión 1.3 (SEI), S. E. (2010). CMMI para Desarrollo, Versión 1.3.
- [2] *Miracle Business Network*, "CMMI Dev 2", MBN
- [3] *L. Piedras Muñoz*, "Modelos de buenas prácticas CMMI", MBN, 15-5
- [4] *Hanna Oktaba (Director), Claudia Alquicira Esquivel, Angélica Su Ramos, Alfonso Martínez Martínez, Gloria Quintanilla Osorio, Mara Ruvalcaba López, Francisco López, Lira Hinojo*, "Modelo de Procesos para la Industria de Software MoProSoft", 1-186 (2005), Versión 1.3
- [5] *Abeer Khan, N. E. (2011)*. Integración entre Customer Relationship Management (CRM) y Data Warehousing. *ScienceDirect*, 1-11.
- [6] *Akhmad Dahlan, F. W. (2017)*. Diseño del almacén de datos de la biblioteca con el método esquema copo de nieve: Caso de estudio: base de datos de la biblioteca del Campus XYZ. *IEEE*, 1-5.
- [7] *Allsoft, S. d. (2011)*. Moprosoft. Obtenido de Allsoft: <http://www.allsoft.mx/recursos/AS-Moprosoft.pdf>
- [8] *Hao, Y., Hongwei, S., Zili, Z. (2011)*. La aplicación del sistema de comercio electrónico basado en el almacén de datos. *IEEE*, 1-4.
- [9] *Ing. Bernabeu, R. D. (2007)*. DATA WAREHOUSING: Investigación y Sistematización de Conceptos versión 0.1. Córdoba, Argentina..
- [10] *Ing. Eric Ismael Leonard Brizuela, M. Y. (2013)*. Metodologías para desarrollar Almacén de Datos. . *Revista de Arquitectura e Ingeniería.*, 12.
- [11] *João Completo, R. S. (2012)*. Diseño e Implementación de un Almacén de Datos para Benchmarking en Rehabilitación Clínica. *ScienceDirect*, 1-10.
- [12] *L. Berrahou. (2015)*. Un almacén de datos espaciales sensible a la calidad para consultar datos hidroecológicos. *ScienceDirect*, 1-10.
- [13] *Lane, P. (2002)*. Guía de Oracle9 i Data Warehousing, versión 2 (9.2). U.S. .
- [14] *Michael Wladimir Vizueté Naranjo, C. P. (marzo de 2006)*. Análisis, diseño e implementación de un base de datos departamental para el área de sismología del departamento de geofísica de la escuela politécnica nacional .
- [15] *Network, M. B. (s.f.)*. CMMi Dev 2. MBN.

- [16] *Abhay Kumar Agarwal (2016). Un enfoque novedoso para la distribución inteligente de almacenes de datos. ScienceDirect, 147-159*
- [17] *Oditis, Z. B. (2016). Hacia Soluciones de almacenamiento de datos basadas en NoSQL. ScienceDirect, 1-8.*
- [18] *Ralph Kimball, L. R. (s.f.). The Data Warehouse Lifecycle Toolkit. WILEY.*
- [19] *Seah, B. K., Selan, N. E. (2014). Diseño e implementación de almacenamiento de datos con modelo de datos utilizando datos de servicios basados en encuestas. IEEE Xplore, 1-7.*
- [20] *Xiewu, H. Z. (2010). Diseño e implementación de data warehouse de pequeños supermercados de cadena. IEEE, 1-3.*
- [21] *Yu, C. (2017). Diseño de sistema de análisis de decisión de venta basado en almacén de datos. IEEE Xplore, 1-6.*