

**UNIVERSIDAD GABRIELA MISTRAL
FACULTAD DE INGENIERIA**

**GESTIÓN DE LOS PROGRAMAS DE
EMPLEO DE EMERGENCIA EN CHILE
UTILIZANDO DATA WAREHOUSE**

Memoria para optar al título de Ingeniero de Ejecución en Informática

Autor: José Rivero Riquelme

Santiago - Chile
Septiembre, 2009

**UNIVERSIDAD GABRIELA MISTRAL
FACULTAD DE INGENIERIA**

**GESTIÓN DE LOS PROGRAMAS DE
EMPLEO DE EMERGENCIA EN CHILE
UTILIZANDO DATA WAREHOUSE**

Memoria para optar al título de Ingeniero de Ejecución en Informática

Autor : José Rivero Riquelme
Profesor Guía : Marcelo Ramírez Moraga
Profesor Integrante : Jorge Tapia Castillo

Santiago - Chile
Septiembre, 2009

INDICE

1. INTRODUCCIÓN	1
1.1. Hipótesis	3
1.2. Objetivo General	3
1.3. Objetivos Específicos.....	3
1.4. Alcances	4
2. MARCO TEORICO REFERENCIAL	5
2.1. Sistemas OLTP o Transaccionales.....	5
2.2. Data Warehouse	10
2.3. Data Mart	17
2.4. Diferencias entre Data Warehouse y OLTP	19
2.5. Herramientas OLAP	22
2.6. Modelo Dimensional.....	27
2.7. Modelando el Data Mart.....	30
2.7.1 Atributos:.....	30
2.7.2 Jerarquías:.....	31
2.7.3 Esquema:.....	32
3. CONTEXTO	35
3.1. La Unidad de Proempleo	35
3.2. Beneficiarios y ejecución de proyecto:	36
3.2.1 Beneficiarios	37
3.2.2 Ejecución de Proyectos	37
3.2.3 Procedimientos en la Ejecución de Proyectos	38
3.3. Descripción del Problema	43
3.4. Solución propuesta	45
3.5. Justificación de solución	47
4. DESARROLLO.....	49
4.1. Marco Metodológico.....	49
4.1.1 Planificación de Proyecto:.....	50
4.1.2 Definición de Requerimientos:	52
4.1.3 Diseño Lógico Dimensional:	53
4.1.4 Diseño Físico:	54
4.1.5 Data Staging ETL:.....	54
4.1.6 Diseño de Arquitectura:	55
4.1.7 Selección e Instalación de Productos:	56
4.1.8 Diseño Aplicaciones Usuarios:.....	58
4.1.9 Desarrollo Aplicaciones Usuarios:	60
4.1.10 Liberación:	67
4.1.11 Mantenición y Mejoras:.....	68
4.2. Modelo Entidad Relación para Proempleo.....	69
4.3. Modelo Lógico Multidimensional para Proempleo.....	72

4.4. Esquema para Proempleo	77
5. RESULTADOS.....	79
5.1. Problemas detectados	81
5.2. Conclusión	83
6. BIBLIOGRAFÍA	85
7. GLOSARIO DE TÉRMINOS.....	86

Índice de Figuras

<i>Figura 01, Ejemplo de tabla de una Base de Datos Relacional</i>	8
<i>Figura 02, Ejemplo de una hoja de cálculo</i>	9
<i>Figura 03, Representación de Sistema OLTP o Transaccional.</i>	9
<i>Figura 04, Representación de Tecnologías Data Warehouse</i>	15
<i>Figura 05, Diagrama del proceso de un Data Warehouse.</i>	16
<i>Figura 06, Representación de Data Mart</i>	18
<i>Figura 07, Comparación entre Data Warehouse y OLTP</i>	21
<i>Figura 08, Representación de Cubo multidimensional.</i>	22
<i>Figura 09, Estructura multidimensional de los datos</i>	23
<i>Figura 10, Análisis de los datos desde el punto de vista del gerente de producto.</i>	23
<i>Figura 11, Visión de los gerentes financiero y regional</i>	24
<i>Figura 12, Representación Cubo OLAP.</i>	25
<i>Figura 13, Representación de una Fact.</i>	27
<i>Figura 14, Preguntas de negocio.</i>	28
<i>Figura 15, Representación de dimensiones.</i>	28
<i>Figura 16, Hecho (Fact) de Transferencia.</i>	29
<i>Figura 17, Representación de atributos de la dimensión Geografía.</i>	30
<i>Figura 18, Representación de atributos de la dimensión Tiempo.</i>	31
<i>Figura 19, Representación de elementos de los Atributos.</i>	31
<i>Figura 20, Ejemplo de Jerarquía.</i>	32
<i>Figura 21, Construcción física de Esquema.</i>	32
<i>Figura 22, Construcción física del Data Warehouse/Data Mart.</i>	33
<i>Figura 23, Construcción del Data Warehouse o Data Mart.</i>	34
<i>Figura 24, Flujo del procedimiento para la Ejecución de Proyectos.</i>	42
<i>Figura 25, Ciclo de Vida Dimensional del Negocio.</i>	50
<i>Figura 26, Carta Gantt del Proyecto Data Mart Para Proempleo.</i>	52
<i>Figura 27, Arquitectura de la solución implementada.</i>	56
<i>Figura 28, Ejemplo de diseño; Costos por fueros y finiquitos.</i>	59
<i>Figura 29, Ejemplo de diseño; Transferencias Línea Intendencia.</i>	60
<i>Figura 30, Informe gráfico desarrollado por la herramienta de consulta OLAP.</i>	62
<i>Figura 31, Ejemplo de consulta dinámica generada por la herramienta OLAP.</i>	63
<i>Figura 32, Ejemplo de consulta dinámica generada por la herramienta OLAP.</i>	64
<i>Figura 33, Jerarquía proyecto y su navegación.</i>	65
<i>Figura 34, Jerarquía Beneficiario y su navegación.</i>	66
<i>Figura 35, Jerarquía Ejecutor y su navegación.</i>	66
<i>Figura 36, Jerarquía Tiempo y su navegación.</i>	67
<i>Figura 37, Modelo Entidad Relación del OLTP de Proempleo.</i>	70
<i>Figura 38, Representación de Jerarquía Beneficiarios.</i>	72
<i>Figura 39, Representación de Jerarquía Proyectos.</i>	73
<i>Figura 40, Representación de Jerarquía de Ejecutores.</i>	74
<i>Figura 41, Representación de Jerarquía de Tiempo.</i>	74
<i>Figura 42, Fact de Transferencias.</i>	75

<i>Figura 43, Fact de Rendiciones.</i>	<u>76</u>
<i>Figura 44, Modelo físico multidimensional (OLAP).</i>	<u>78</u>

Dedicatoria

A Jacqueline

Quien compartió el compromiso y la carga de la atención disminuida.

A Rodrigo y Karina

Quienes debieron asumir tardes de ausencias sin consentimiento y el poco tiempo compartido.

Agradecimientos

A mi profesor guía, señor Marcelo Ramírez Moraga por toda la ayuda y consejos prestados en la ejecución de este proyecto de Tesis, que se enmarca en dar solución a la problemática de la unidad de Proempleo.

Agradecer sin duda, el apoyo otorgado por las autoridades del Ministerio del Trabajo y Previsión Social y en especial a las jefaturas de la División de Administración y Finanzas, señor Ricardo León Espinoza, al jefe de Proempleo, señor Christian Álvarez Eugenin y al jefe de Informática, señor Carlos Abarzúa Castro. Quienes con su apoyo permitieron que este proyecto se pudiera realizar.

A mi familia, quien con su cariño y apoyo, permitieron que me pudiera dedicar a estudiar una vez más.

1. INTRODUCCIÓN

En la actualidad, las tecnologías de la información han automatizado los procesos de carácter típicamente repetitivo o administrativo, haciendo uso de lo que llamaremos sistemas de información operacionales o transaccionales. Entendemos por aplicaciones operacionales o transaccionales aquellas que resuelven las necesidades de funcionamiento de la organización. En este tipo de sistemas, los conceptos más importantes son la actualización y el tiempo de respuesta.

Una vez satisfechas estas necesidades más acuciantes, surge un nuevo grupo de requerimientos de información sobre los sistemas por parte de la organización, es decir obtener información que sirva de base para la toma de decisiones. Para atender a éstos, los responsables de los sistemas tropiezan con múltiples problemas. Cuando se realizan consultas masivas de información se perjudica el nivel de servicio del resto de los sistemas, dado que las consultas son costosas en cuanto a los recursos y tiempo y por otro lado es limitada la flexibilidad a la hora de navegar por la información.

Como solución a estos problemas de información surge el concepto de **Data Warehouse**. Un Data Warehouse es una colección de datos orientada a sujetos, integrada, variante en el tiempo, no volátil y que soporta el proceso de toma de decisiones.

Data Warehouse no es un producto que pueda ser comprado en el mercado, sino más bien un concepto que debe ser construido. Data Warehouse es una combinación de conceptos y tecnología que cambian significativamente la manera en que es entregada la información a la gente de negocios.

El Data Warehouse intenta responder a la compleja necesidad de obtención de información útil, sin el sacrificio del rendimiento de las aplicaciones operacionales,

debido a lo cual se ha convertido actualmente en una de las tendencias tecnológicas más significativas en la administración de información.

Los Data Warehouse (o almacenes de datos), generan bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente.

El término “Data Warehouse” se utiliza indistintamente para hablar de la tecnología en sí como también para uno de los componentes que la conforman, específicamente al repositorio de los datos (almacén o bodega). Ahora, con el propósito de facilitar el entendimiento por parte del lector, en adelante se hablará de Data Warehouse como el repositorio de los datos o almacén de datos, salvo que se indique lo contrario.

El concepto Data Mart es una extensión natural del Data Warehouse, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de Finanzas o Marketing. Permitiendo así un mejor control de la información que se está abarcando. Los Data Mart deben comprenderse como subconjuntos de datos de un Data Warehouse para áreas específicas.

En el desarrollo de este trabajo se pretende aclarar que hablar de Data Mart es conceptualmente lo mismo que hablar de Data Warehouse, solo que el Data Mart se diseña para un ámbito más reducido y se enfoca a un departamento de una gran empresa o bien a una pequeña empresa. Podemos entender a un Data Mart como un Data Warehouse pequeño. Por lo general se habla de Data Warehouse para referirse a un Data Mart, lo que conceptualmente es correcto, ya que un Data Warehouse tenemos que entenderlo como un conjunto de uno o más Data Mart. Por ejemplo: “El **Data Warehouse** del Ministerio del Trabajo es por ahora el **Data Mart** de Proempleo.”

Esta tesis trata sobre la implementación de *tecnologías Data Warehouse* para el diseño y desarrollo de un Data Mart (Data Warehouse), éste consolida y administra la información generada en la operación diaria, la que es mantenida por los sistemas informáticos operacionales de la unidad encargada de los programas de Empleo de Emergencia de la Subsecretaría del Trabajo del Ministerio del Trabajo y Previsión Social, denominado **Proempleo**.

1.1. Hipótesis

Implementando *tecnologías Data Warehouse* es posible resolver los requerimientos de información de la unidad Proempleo y que a la vez le sirva de base para la toma de decisiones a las autoridades del Ministerio del Trabajo y Previsión Social y que se traducen en aplicar mejoras a la *gestión de los Programas de Empleo de Emergencia en Chile*.

1.2. Objetivo General

Permitir mejorar la gestión y control de los procesos de transferencia y rendiciones de fondos de los distintos Programas de Empleo de Emergencia en Chile.

1.3. Objetivos Específicos

Generar un modelo de control de transferencias de fondos a los distintos organismos encargados de la administración de los Empleos de Emergencia del País, sobre los montos planificados o solicitados a transferir y los realmente transferidos.

Medir que los montos efectivamente gastados por las entidades receptoras, corresponda a los gastos generados por los proyectos que ejecutan.

Controlar el correcto otorgamiento de fondos a las entidades receptoras, sobre la base de las rendiciones efectuadas por los entes ejecutores de los programas de Empleo.

Como objetivo estratégico: permitir focalizar el otorgamiento del beneficio (Empleo), a los sectores más desprotegidos, utilizando criterios como etnias, género o por división política territorial.

1.4. Alcances

Desarrollar el diseño de un modelo analítico para gestionar las transferencias y rendiciones de fondos a los distintos organismos encargados de la administración de los Empleos de Emergencia del País.

Construcción de un sistema informático computacional que resuelve la implementación física del modelo descrito anteriormente.

Presentación del modelo lógico operacional de Proempleo.

Desarrollo del modelo lógico analítico (Data Mart).

Desarrollo de un prototipo funcional en el que se señalan los indicadores, reportes y pantallas de control.

2. MARCO TEORICO REFERENCIAL

2.1. Sistemas OLTP o Transaccionales

Es un tipo de sistema de información diseñado para recolectar, almacenar, modificar y recuperar todo tipo de información que es generada por las transacciones en una organización. Una transacción es un evento o proceso que genera o modifica la información que se encuentran eventualmente almacenados en un sistema de información.

Los sistemas transaccionales, también son conocidos como sistemas **OLTP**, por su sigla en inglés de Procesamiento de Transacciones en Línea (**On Line Transaction Processing**), facilitan y administran aplicaciones para entrada de datos, recuperación y procesamiento de transacciones, enfocados a gestionar un gran número de ellas en forma concurrente. Permiten insertar, actualizar, borrar y consultar una pequeña cantidad de registros.

Por lo general los sistemas OLTP o Transaccionales, son el primer sistema de Información que se implementa en la empresa. Se empieza apoyando las tareas a nivel operativo de la organización para continuar con los mandos intermedios y posteriormente la alta administración, conforme evolucionan.

Un ejemplo son los sistemas bancarios en las sucursales, los cajeros automáticos, los sistemas WEB de compras, entre otros.

Ejemplos de funciones concretas de un sistema OLTP o transaccional

Un sistema transaccional debe controlar las transacciones para mantener la seguridad y consistencia de los datos involucrados. Por ejemplo, un cliente utilizando la página web de su banco, transfiere dinero de una cuenta a otra

cuenta dentro del mismo banco; la cantidad de dinero que se descuenta de la cuenta emisora debe ser igual a la que se suma en la cuenta receptora. De no ser así, la acción (transacción) no se realiza.

Un sistema transaccional debe ser capaz de enmendar cualquier error ocurrido durante una transacción, pudiendo deshacer las operaciones realizadas, manteniendo los datos tal cual estaban antes del error.

También debe ser capaz de controlar y administrar múltiples transacciones, determinando prioridades entre éstas. Por ejemplo, un cliente está haciendo la reserva de un asiento en un vuelo, dicho asiento debe ser bloqueado temporalmente hasta que se concrete la transacción, porque otro cliente podría estar queriendo reservar el mismo asiento en el mismo momento.

Propiedades de los sistemas OLTP o transaccionales

- Automatizan tareas operativas en una organización, permitiendo ahorrar en personal.
- Suelen dirigirse especialmente al área de ventas, finanzas, marketing, administración y recursos humanos.
- Suelen ser los primeros sistemas de información que se implementan en una organización.
- Sus cálculos y procesos suelen ser simples.
- Los beneficios de este tipo de sistemas en una organización son rápidamente visibles.

- Estos sistemas son optimizados para almacenar grandes volúmenes de datos, pero no para analizar los mismos.

Características esperables de un sistema transaccional

- Rapidez: deben ser capaces de responder rápidamente, en general, la respuesta no debe ser mayor a un par de segundos.
- Fiabilidad: deben ser altamente fiables, de lo contrario podría afectar a clientes, al negocio, a la reputación de la organización, etc. En caso de fallas, debe tener mecanismos de recuperación y de respaldo de datos.
- Inflexibilidad: no pueden aceptar información distinta a la establecida. Por ejemplo, el sistema transaccional de una aerolínea debe aceptar reservas de múltiples agencias de viajes. Cada reserva debe contener los mismos datos obligatorios, con determinadas características.

La mayoría de los Sistemas transaccionales, almacenan la información que capturan, en bases de datos Relacionales, el cual es el modelo más utilizado en la actualidad.

El **modelo relacional** para la gestión de una base de datos, es un modelo de datos basado en la teoría de conjuntos. En este modelo todos los datos son almacenados en relaciones, y como cada relación es un conjunto de datos, el orden en el que estos se almacenen no tiene mayor relevancia. Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar por un usuario no experto.

Este modelo considera la base de datos como una colección de relaciones. De manera simple, una relación representa una tabla que no es más que un

conjunto de filas, cada fila es un conjunto de campos y cada campo representa un valor que interpretado describe el mundo real. Cada fila también se puede denominar tupla o registro y a cada columna también se le puede llamar campo o atributo.

Para una mejor comprensión, se puede pensar en la tabla como una hoja de cálculo. Las filas son el equivalente de filas de la hoja de cálculo. Las columnas equivalen a las columnas de una hoja de cálculo y el campo representa a una celda en la hoja de cálculo.

En las Figura 01 y Figura 02, se muestran ejemplos de analogía entre una tabla de base de datos relacional y de una hoja de cálculo.



Advertencia de seguridad Se ha deshabilitado parte del contenido

CodComuna	Descripcion	codRegion
01101	ARICA	1
01102	CAMARONES	1
01201	PUTRE	1
01202	GENERAL LAGO	1
01301	IQUIQUE	1
01302	HUARA	1
01303	CAMIÑA	1
01304	COLCHANE	1
01305	PICA	1
01306	POZO ALMONT	1
01307	ALTO HOSPICIO	1
02101	TOCOPILLA	2
02102	MARIA ELENA	2
02201	CALAMA	2
02202	OLLAGUE	2

Registro: 1 de 355 Sin filtro Busc

Figura 01, Ejemplo de tabla de una Base de Datos Relacional.

PROVINCIA								
	A	B	C	D	E	F	G	H
1	Transferencias Línea Intendencia							
2	PROVINCIA	COMUNA	Nº	1/2 JORNADA	3/4 JORNADA	1 JORNADA	Costo Mensual	COSTO AGOSTO
3	VIII REGION							
4								
5	CONCEPCION	LOTA	589	0	0	589	103.307.066	103.307.066
6		TALCAHUANO	525	0	0	525	92.081.850	92.081.850
7		PENCO	80	0	0	80	14.031.520	14.031.520
8		SAN PEDRO DE LA PAZ	120	0	20	100	20.246.860	20.246.860
9		CONCEPCIÓN	25	0	19	6	3.624.451	3.624.451
10		TOMÉ	247	0	0	247	43.322.318	43.322.318
11	ARAUCO	CURANILAHUE	153	0	0	153	26.835.282	26.835.282
12		LOS ALAMOS	189	0	0	189	33.149.466	33.149.466
13		LEBU	316	0	0	316	55.424.504	55.424.504
14		ARAUCO	48	0	0	48	8.418.912	8.418.912
15		CAÑETE	30	0	0	30	5.261.820	5.261.820
16		TIRÚA	65	0	0	65	11.400.610	11.400.610
17		CONTULMO	105	0	0	105	18.416.370	18.416.370
18	TOTAL		2.492	0	39	2.453	435.521.029	435.521.029
19								

Figura 02, Ejemplo de una hoja de cálculo.

La Figura 03, representa un sistema transaccional (OLTP) y el almacenamiento de datos en distintas bases de Datos Transaccionales.



Figura 03, Representación de Sistema OLTP o Transaccional.

2.2. Data Warehouse

Existen muchas definiciones para el Data Warehouse, la más conocida fue propuesta por Bill H. Inmon (considerado el padre de las Bases de Datos) en 1992: “Un Data Warehouse es una colección de datos orientados a temas, integrados, no volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”. En 1993, Susan Osterfeldt publica una definición que sin duda acierta en la clave del Data Warehouse: “Yo considero al Data Warehouse como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. Data Warehouse elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico”.

Esta última definición refleja claramente el principal beneficio que el Data Warehouse aporta a la organización: eliminar aquellos datos que obstaculizan la labor de análisis de información y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión.

Hasta ahora la manera tradicional de entregar la información es a través de emisión de reportes impresos desde los sistemas transaccionales, a veces con consultas directas en pantalla que permiten una extracción limitada de datos para suplir en alguna medida la acumulación de papel. Los problemas con la entrega de la información actual son muchos, incluyendo inconsistencia, inflexibilidad y carencia de integración a través de la organización.

Analizar esta información puede presentar resultados distintos en cuestión de minutos, por lo que se deben extraer y almacenar fotografías de datos, para estos efectos, con la implicancia de un consumo adicional de recursos de cómputo. Llevar a cabo un análisis complejo sobre un sistema transaccional,

puede resultar en la degradación del sistema, con el consiguiente impacto en la operación del negocio.

Para una mayor comprensión, el **Data Warehouse** puede verse como un almacén o bodega, donde están almacenados todos los datos necesarios para realizar las funciones de gestión de la empresa, de manera que puedan utilizarse fácilmente según se requiera. El contenido de los datos, la organización y estructura son dirigidos a satisfacer las necesidades de información de analistas. Los sistemas transaccionales son dinámicos, en el sentido que constantemente se encuentran actualizando datos.

El Data Warehouse intenta responder a la compleja necesidad de obtención de información útil sin el sacrificio del rendimiento de las aplicaciones operacionales, debido a lo cual se ha convertido actualmente en una de las tendencias tecnológicas más significativas en la administración de información.

Uno de los objetivos principales, de este tipo de sistemas, es servir como base de información para la toma de decisiones. Los beneficios obtenidos por la utilización de este tipo de sistemas se basan en el acceso interactivo e inmediato a información estratégica de un área de negocios. Este acercamiento de la información al usuario final permite una toma de decisiones rápida y basada en datos objetivos obtenidos a partir de las bases de datos de la Organización. Estos beneficios aumentan cuanto más importantes son las decisiones a tomar y cuanto más crítico es el factor tiempo.

Hoy en día las organizaciones cuentan en su mayoría con la automatización de sus procesos, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea (operacionales). En esta información descansa el conocimiento de la organización, constituyendo un recurso corporativo primario y parte importante de su patrimonio.

La mayoría de las Empresas disponen de una gran cantidad de datos, pero muy poca información. Varias razones motivan estos hechos: islas de información, carencia de arquitectura, gestión, responsabilidad, posesión de los datos, deficiencia en calidad, contenido, accesibilidad, fiabilidad de la información, múltiples y diversas aplicaciones operacionales, existencia de fuentes de información externa, etc. Gran parte del producto generado por tecnologías de información, no es información, sino solo datos brutos. Son generados por sistemas que fueron ideados para recogerlos, pero no para analizarlos. Los datos adquieren la categoría de información cuando disponen de una estructura inteligente.

Las *tecnologías Data Warehouse*, permite recoger los datos de los diferentes sistemas operacionales y depurarlos, integrarlos y almacenarlos en una estructura de la cual se puede obtener información de ellos, muchos datos de los sistemas transaccionales no son de importancia como información para la toma de decisiones, por ejemplo: la hora en que se grabó un registro.

Un Sistema de Data Warehouse incluye funcionalidades tales como:

- Integración de bases de datos heterogéneas (relacionales, documentales, geográficas, archivos, etc.).
- Ejecución de consultas complejas no predefinidas visualizando el resultado en forma de gráfica y en diferentes niveles de agrupamiento y totalización de datos.
- Agrupar y desagrupar datos en forma interactiva.
- Análisis de problema en términos de dimensiones. Por ejemplo, permite analizar datos históricos a través de una dimensión tiempo.

- Control de calidad de datos para asegurar, no solo la consistencia de la base, sino también la relevancia de los datos en base a los cuales se toman las decisiones.

Profundizando en la definición del concepto, se dice que un Data Warehouse es una colección de datos **Orientada a sujetos, Integrada, Variante en el tiempo, No volátil** y que soporta el proceso de toma de decisiones. Un Data Warehouse soporta procesamiento informático, brindando una sólida plataforma de datos históricos, integrados de los cuales hacer análisis.

Es orientado a sujetos:

Un primer aspecto de un Data Warehouse es que está orientado a los mayores sujetos de la empresa. El mundo operacional está diseñado alrededor de aplicaciones y funciones, como por ejemplo pagos, ventas, entregas de mercadería, para una institución comercial. Un Data Warehouse está organizado alrededor de los mayores sujetos, como cliente, vendedor, producto y actividades. El mundo operacional concierne al diseño de la base de datos y al diseño de procesos. Un Data Warehouse está enfocado en modelar los datos y el diseño de la base de datos, exclusivamente. El diseño de procesos (en su forma clásica) no es parte del Data Warehouse.

Los datos son integrados:

El aspecto más importante del ambiente de un Data Warehouse es que sus datos están integrados. Cuando los datos son movidos del ambiente operacional, son integrados antes de entrar en el Warehouse. Por ejemplo, un diseñador puede representar el sexo como "M" y "F", otro puede representarlo como "0" y "1", o "x" e "y", y otro usar las palabras completas "masculino" y "femenino". No importa la fuente de la cual el sexo llegue al Data Warehouse, debe ser guardado en forma consistente; los datos deben ser integrados.

Es variante en el tiempo:

Los datos en el Warehouse son precisos para un cierto momento, no necesariamente ahora; por eso se dice que los datos en el Warehouse son variantes en el tiempo. La varianza en el tiempo de los datos de un Warehouse se manifiesta de muchas maneras. El Data Warehouse contiene datos de un largo horizonte de tiempo. Las aplicaciones operacionales, sin embargo, contienen datos de intervalos de tiempo pequeños, por cuestiones de performance (reducido tamaño de las tablas). Toda estructura clave en un Warehouse contiene implícita o explícitamente un elemento del tiempo. Esto no necesariamente pasa en el ambiente operacional. Los datos de un Warehouse, una vez almacenados, no pueden ser modificados (no se permiten actualizaciones). En el ambiente operacional, los datos pueden ser actualizados según sea necesario.

Es simple de manejar:

Actualizaciones, agregaciones y eliminaciones son efectuadas regularmente a los datos de los sistemas operacionales (OLTP). La manipulación de datos en un Warehouse, es mucho más sencilla. Solo ocurren dos operaciones, la carga inicial, y el acceso a los datos. No hay necesidad de actualizaciones (en su sentido general). Hay consecuencias muy importantes de esta diferencia de procesos con un sistema operacional: A nivel de diseño, en un Warehouse, no hay que controlar anomalías producidas por las actualizaciones, ya que no hay actualizaciones.

El Data Warehouse convierte entonces, los datos operacionales de una organización, en una herramienta competitiva, por hacerlos disponibles a los empleados que lo necesiten para el análisis y toma de decisiones.

La Figura 04, representa a las *tecnologías Data Warehouse*: A partir de diferentes fuentes y luego de un proceso de extracción y clasificación de los

datos, se crea el almacén de datos (Data Warehouse). Finalmente, por medio de un proceso de explotación se accede a los datos almacenados en el Almacén de Datos.

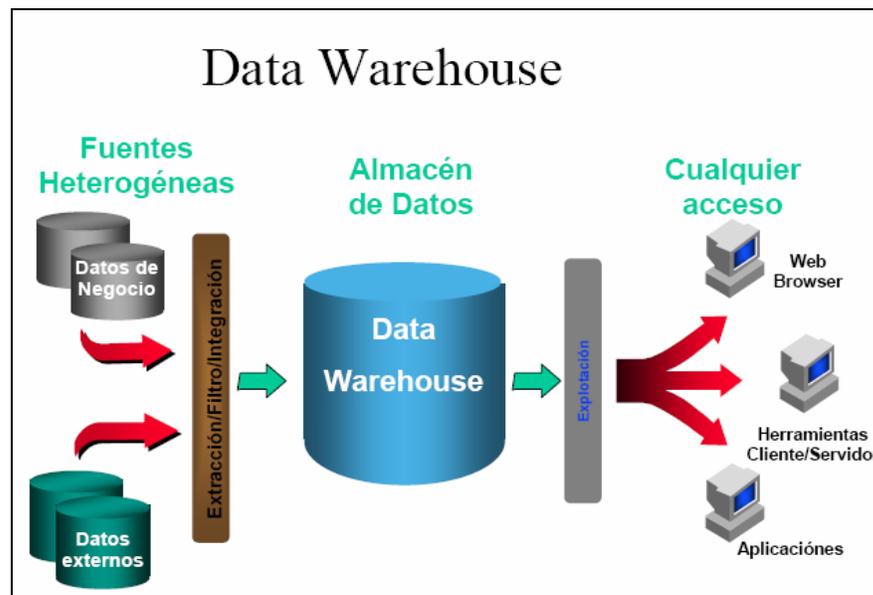


Figura 04, Representación de Tecnologías Data Warehouse.

Las Bases de Datos relacionales, modelo sobre el cual se basa OLTP (On Line Transaction Processing, Procesamiento Transaccional en Línea), tiene como objetivo guardar la integridad de la información necesaria para operar un negocio de la manera más eficiente. Sin embargo, este modelo no corresponde a la forma como el usuario percibe la operación de un negocio.

De hecho el Dr. Edgar Frank Codd, a quien se le reconoce como el creador del modelo relacional, dijo: "Aunque los RDBMS (siglas en inglés de **R**elational **D**atabase **M**anagement **S**ystem o Sistema de Gestión de Base de Datos Relacional o SGBDR), han sido tan beneficiosos para los usuarios, nunca han sido diseñados para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos".

El Data Warehouse se sustenta en un procesamiento distinto al utilizado por los sistemas operacionales, y se conoce como OLAP (Procesamiento Analítico En Línea), el cual surge como un proceso para ser usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio. A este procesamiento también se le conoce como análisis multidimensional o cubo OLAP.

La Figura 05, representa la alimentación del **Data Warehouse**, el motor de análisis (Cubos OLAP), y la extracción de información a través del análisis multidimensional.

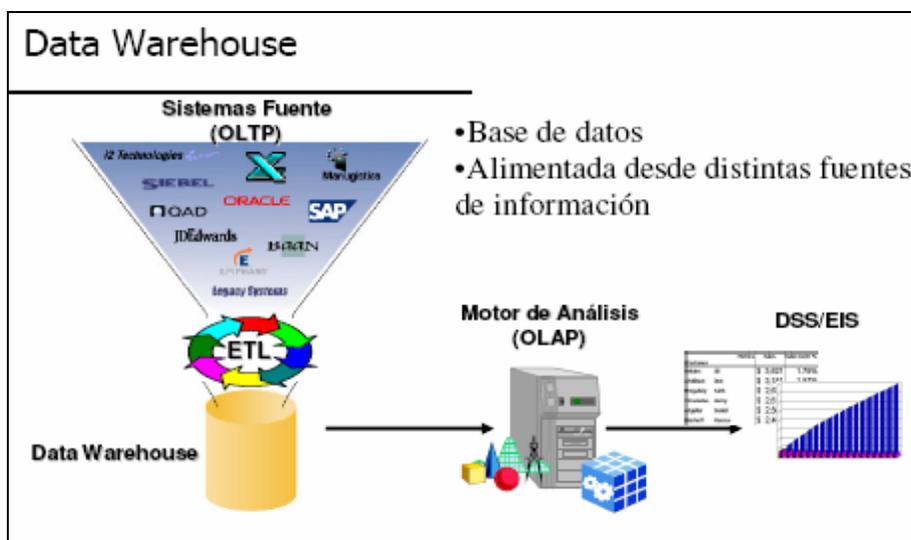


Figura 05, Diagrama del proceso de un Data Warehouse.

Ciertamente podemos decir que un Data Warehouse es un sistema que ofrece acceso fácil y rápido a información vital para la gestión y operación de una empresa, que permite acceso a datos históricos y detallados, que es un sistema donde la información es alimentada por múltiples sistemas fuentes, que es un sistema de análisis de información destinado a todo tipo de usuario y que la

extracción de información se realiza a través del análisis multidimensional, también conocido como cubos OLAP.

2.3. Data Mart

Un Data Mart es una versión especial de Data Warehouse. Son subconjuntos de los datos del Data Warehouse con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica.

Los Data Mart son diseñados para satisfacer las necesidades específicas de grupos comunes de usuarios (divisiones geográficas, divisiones organizacionales, departamentos dentro de la organización, etc.).

Al igual que el Data Warehouse, el **Data Mart** es un almacén de datos históricos especializado en el almacenamiento de los datos pero se diferencia en que éstos apuntan a un departamento o área de negocio específica de la organización. Se caracteriza por disponer la **estructura multidimensional o de cubo** para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento o área de negocio. Un Data Mart puede ser alimentado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Aunque hoy en día es difícil diferenciar a los Data Mart y Data Warehouse por su tamaño, algunas distinciones entre ellos son todavía importantes:

- Un Data Mart está enfocado a una sola área o grupo de usuarios, mientras que un Data Warehouse contiene información de diferentes sujetos y áreas de la corporación.

- Una organización puede tener un sólo Data Warehouse, pero varios Data Mart.
- Tanto el Data Warehouse como el Data Mart son sistemas orientados a la consulta. Mediante herramientas OLAP (**O**n **L**ine **A**nalytical **P**rocessing), se obtiene información de ellos y ofrecen una visión multidimensional de la información (Cubos OLAP).

Luego, un Data Mart, que compartiendo tecnología con el Data Warehouse (pero con contenidos específicos, volumen de datos más limitado y un alcance histórico menor), es la solución generalmente implementada en las organizaciones, que deciden incorporar la tecnología Data Warehouse, para satisfacer un área de negocio específico y carente de información estratégica. Con el tiempo el Data Warehouse, puede seguir creciendo con la incorporación de otros Data Mart de otras áreas de negocio de la Organización.

La Figura 06, representa a dos Data Mart, como subconjunto de Data Warehouse.

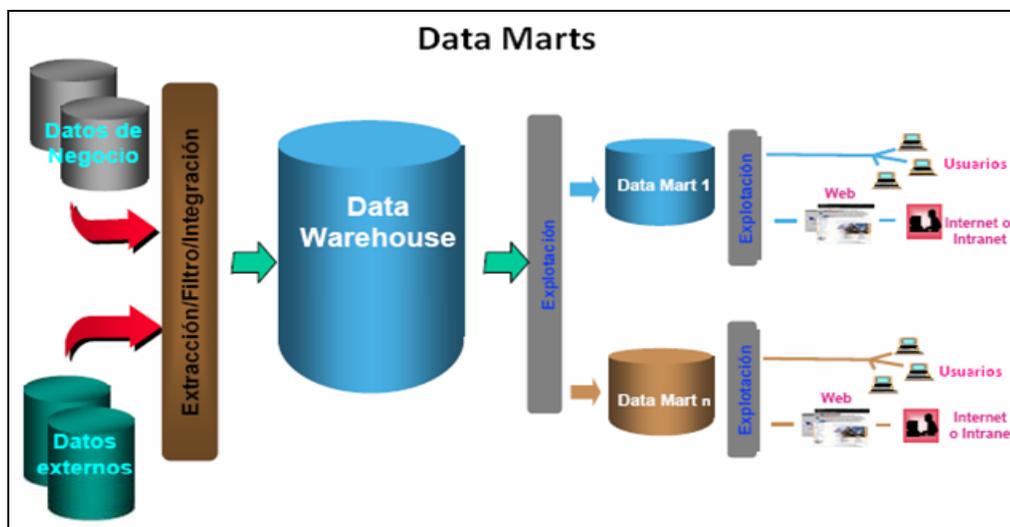


Figura 06, Representación de Data Mart.

2.4. Diferencias entre Data Warehouse y OLTP

En los títulos anteriores se ha explicado los conceptos de sistemas Data Warehouse y Sistemas Operacionales (OLTP), que son la base de alimentación de datos para los Data Warehouse, los que a su vez son accedido por medio de análisis multidimensional o de herramientas OLAP o cubo OLAP. A continuación se indicarán las diferencias más relevantes que existen entre estos dos mundos.

Mientras que las aplicaciones OLTP se caracterizan por estar actualizadas constantemente por varios usuarios a través de transacciones operacionales sobre datos individuales, las aplicaciones Data Warehouse (Consultadas por herramientas OLAP), son utilizadas por personal de niveles ejecutivos que requieren datos con alto grado de agregación y desde distintas perspectivas (dimensiones), como ser: totales de venta por región, por producto, por período de tiempo, etc.

Los sistemas transaccionales (OLTP) y las aplicaciones de Data Warehouse son polos opuestos en cuanto a sus requerimientos de diseño y sus características de operación.

Las aplicaciones de OLTP están organizadas para ejecutar las transacciones para los cuales fueron hechos, como por ejemplo: mover dinero entre cuentas, un cargo o abono, una devolución de inventario, etc. Por otro lado, un Data Warehouse está organizado en base a conceptos, como por ejemplo: clientes, facturas, productos, etc.

Otra diferencia radica en el número de usuarios. Normalmente, el número de usuarios de un Data Warehouse es menor al de un OLTP. Es común encontrar que los sistemas transaccionales son accedidos por cientos de usuarios simultáneamente, mientras que los Data Warehouse sólo por decenas. Los

sistemas de OLTP realizan cientos de transacciones por segundo mientras que una sola consulta de un Data Warehouse puede tomar minutos. Otro factor es que frecuentemente los sistemas transaccionales son menores en tamaño a los Data Warehouse, esto es debido a que un Data Warehouse puede estar formado por información de varios OLTP.

Las estructuras de los OLTP son muy estables, rara vez cambian, mientras las de los Data Warehouse sufren cambios constantes derivados de su evolución. Esto se debe a que los tipos de consultas a los cuales están sujetos son muy variados y es imposible preverlos todos de antemano.

La siguiente tabla muestra las diferencias entre OLTP y Data Warehouse

OLTP	Data Warehouse
Orientada a transacciones	Orientada a Conceptos
Actualizada en línea	Representa valores a un tiempo
Usuarios de nivel operativo	Usuarios de nivel gerencial
Acceso unidades a la vez	Acceso conjuntos de unidades a la vez
Orientado a operación	Orientado a análisis
Estructura estática	Estructura flexible
Sin redundancia	Con mucha redundancia
Alta probabilidad de acceso	Modesta probabilidad de acceso
Información bruta (Datos)	Información procesada (Información)
Actualizada en línea	Actualizada por períodos más extensos

La Figura 07, señala una comparación entre un Data Warehouse y los Sistemas Transaccionales tradicionales (OLTP).

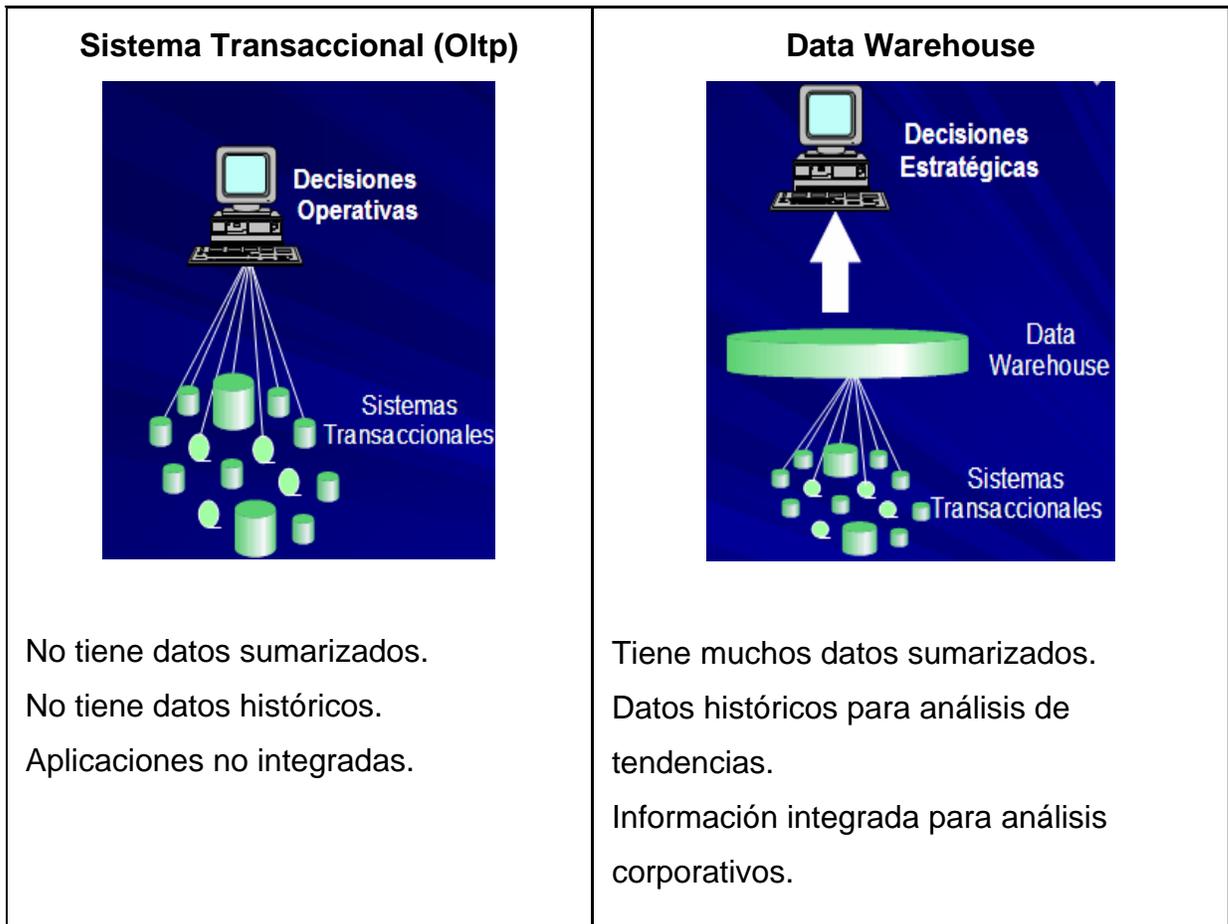


Figura 07, Comparación entre Data Warehouse y OLTP.

2.5. Herramientas OLAP

OLAP, siglas en inglés para “On-Line Analytical Processing”, o Proceso analítico en línea, es el nombre formal para el análisis de los Data Warehouse o Data Mart y que por lo general se conocen como cubos multidimensionales o cubos OLAP. Son una forma más intuitiva de ver la información empresarial.

La Figura 08, representa a un cubo multidimensional.

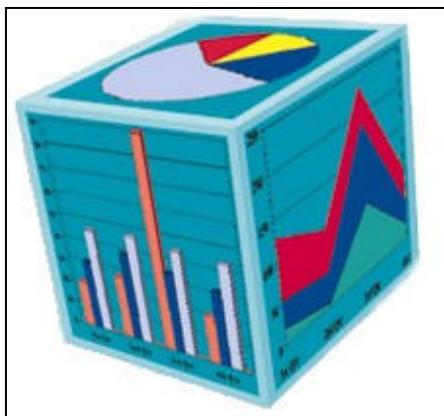


Figura 08, Representación de Cubo multidimensional.

Las herramientas OLAP son aplicaciones que se encargan de analizar datos del negocio para generar información táctica y estratégica que sirve de soporte para la toma de decisiones. Mientras que las transacciones OLTP utilizan Bases de Datos Relacionales u otro tipo de archivos, OLAP logra su máxima eficiencia y flexibilidad operando sobre Bases de datos Multidimensionales.

En el análisis multidimensional, los datos se representan mediante dimensiones, (Figura 09), como por ejemplo: producto, territorio y cliente. En general, las dimensiones se relacionan en jerarquías, por ejemplo: ciudad, estado, región, país y continente. El tiempo es también una dimensión estándar con sus propias jerarquías tales como: día, semana, mes, trimestre y año.

No es común que, por ejemplo, alguien dentro de la organización se pregunte: “¿cuánto se vendió?”.

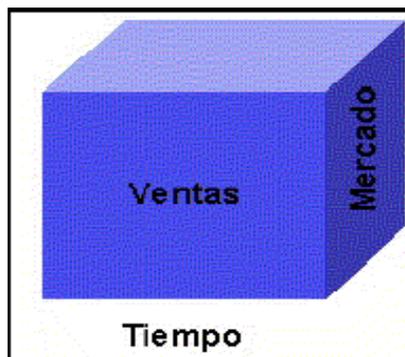


Figura 09, Estructura multidimensional de los datos.

En general, un Gerente de Ventas podría preguntarse: ¿Cuánto se vendió del producto “A” en el período “X” en la región “Y”?, como se señala en la Figura 10.

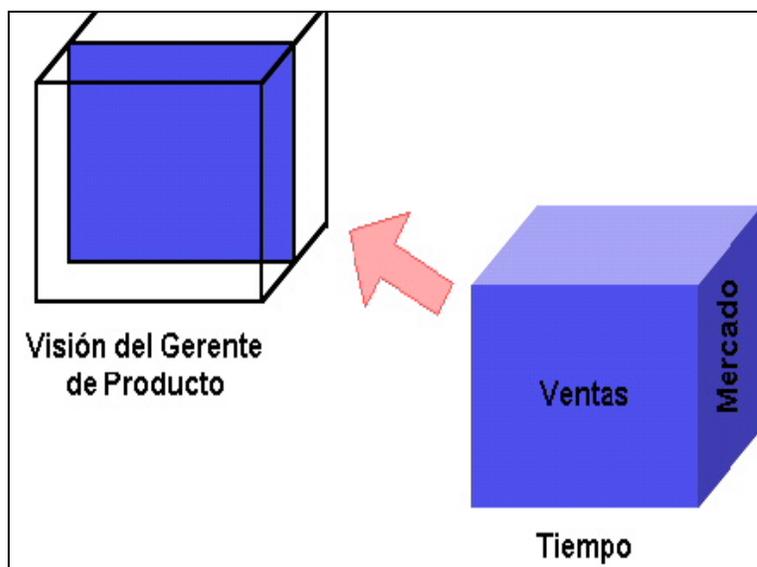


Figura 10, Análisis de los datos desde el punto de vista del gerente de producto.

En cambio, para un gerente de Finanzas la necesidad es diferente y su pregunta sería: ¿A cuánto ascendieron las ventas de todos los productos en todas las regiones al cierre del mes “M”? y para el caso de un gerente regional: ¿Cuánto fueron las ventas de todos los productos en el período J ó K en la región?, como se señala en la Figura 11.

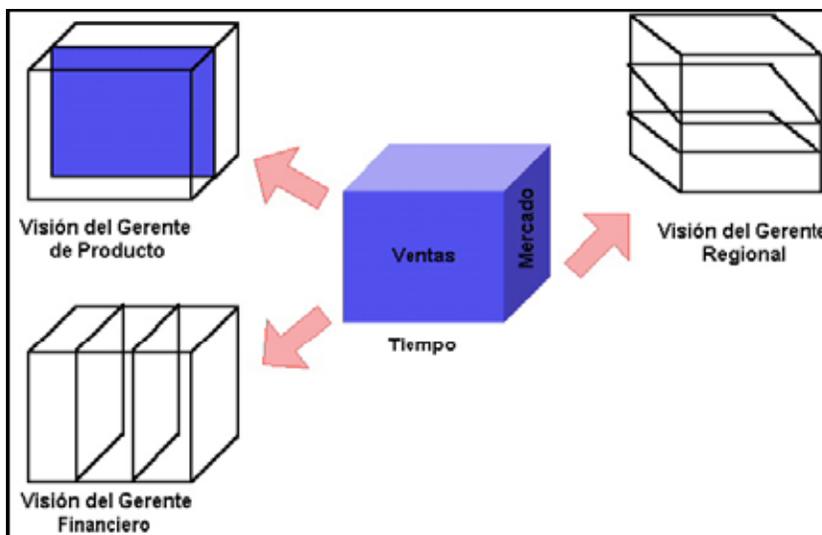


Figura 11, Visión de los gerentes financiero y regional

Podemos nombrar las siguientes características como las más sobresalientes de las herramientas OLAP:

- Estructura de datos transparente al usuario.
- Solo Consulta, trabajan sobre Data Warehouse o Data Mart.
- Consultas sobre grandes volúmenes de datos no predecibles.
- Información histórica.

- Alta redundancia de datos para facilitar la generación de consultas y obtener buenos tiempos de respuesta.

Con OLAP se puede ver un conjunto de datos del negocio de muchas y diversas formas sin mucho esfuerzo. Los archivos OLAP o cubos modelan los datos en **dimensiones**. Una dimensión es una clasificación de alguna actividad en una organización por la cual se puede medir su éxito. Por ejemplo, puede monitorear sus productos en zonas geográficas en un periodo de tiempo. Figura 12.

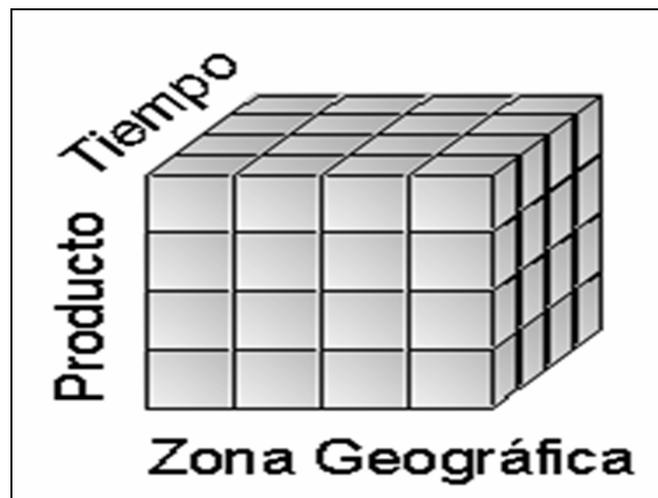


Figura 12, Representación Cubo OLAP.

OLAP se relaciona con hechos o dimensiones que normalmente contienen datos transaccionales que relacionan los productos, localizaciones y tiempos de una organización. Cada dimensión puede tener también alguna jerarquía. Por ejemplo, la dimensión temporal podría descender de años a trimestres, a meses, e incluso a semanas y días. Una dimensión geográfica podría ascender de la ciudad al estado, a la región, al país e incluso al hemisferio, si fuese necesario. Por ejemplo: la Intendencia de la quinta región, que elabora informes regionales para el mes de Noviembre, podría tener posteriormente interés en descender en

la profundidad de sus informes a una comuna particular de la región y sólo para una parte de los proyectos que ejecuta y para el 10 de Noviembre, y así sucesivamente. Los datos presentados en estas dimensiones, llamadas medidas, son generalmente agregados (por ejemplo, rendiciones totales o medias en pesos o unidades, o montos transferidos del presupuesto).

Durante algún tiempo ha habido muchas organizaciones que han acumulado cantidades considerables de datos, que podrían ser útiles para la planificación de los negocios. Podrían utilizarse las tendencias históricas y las proyecciones de futuro para analizar las alternativas de los negocios, y tomar decisiones más informadas que podrían ganar o mantener una ventaja competitiva. Durante la última década, los Data Warehouse han llegado a ser comunes en las grandes compañías, muchas de las cuales utilizan herramientas OLAP para realizar informes y sistemas de apoyo a la decisión. Estas aplicaciones OLAP abarcan una variedad de funciones organizativas. Los departamentos financieros utilizan OLAP para aplicaciones como realización de presupuestos, análisis de costes basado en la actividad (localizaciones), análisis de rendimiento financiero, y modelado financiero. Los análisis y pronósticos de ventas son dos de las aplicaciones OLAP que se hallan en los departamentos de ventas. Entre otras aplicaciones, los departamentos de marketing utilizan OLAP para análisis de investigación de mercados, pronóstico de ventas, análisis de la promoción, análisis de los clientes, y segmentación de mercados/clientes. Las típicas aplicaciones de las herramientas OLAP incluyen la planificación de la producción y el análisis de los defectos.

2.6. Modelo Dimensional

El modelado dimensional se basa en HECHOS que se conocen como “Facts” y es una alternativa al modelo relacional. Sus principales ventajas son:

- Enfocado en el negocio y sus actividades.
- Permite búsquedas a gran velocidad.

En la Figura 13, se diagrama un hecho (Fact), que representa a un número que está relacionado con una venta.

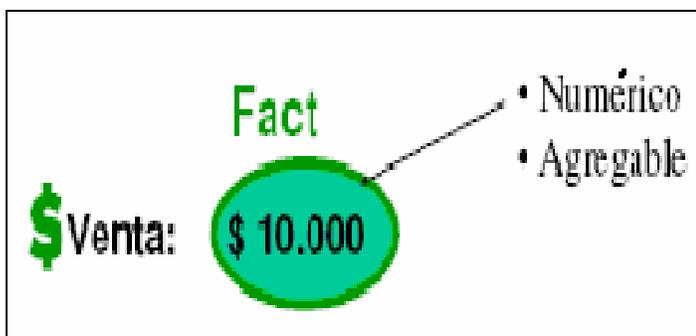


Figura 13, Representación de una Fact.

Generalmente una *Fact* es un dato numérico, que puede ser agregable para responder preguntas de negocio. Como por ejemplo: ¿Quién realizó la compra?, ¿Cómo fue pagada la compra?, etc. Como lo representa la Figura 14,

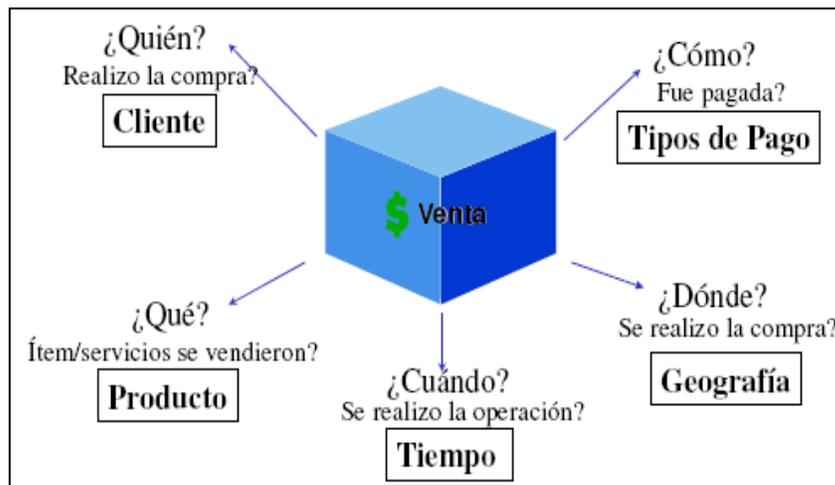


Figura 14, Preguntas de negocio.

Las dimensiones son las variables sobre las que se hace la consulta, cada dimensión tiene distintos valores. Las medidas son las preguntas o consultas que se hacen y sobre las que responderá el sistema de acuerdo a los valores de las dimensiones y los datos almacenados.

En la Figura 15, se representa a las dimensiones Geografía, Productos, Tiempo, Tipos de Pagos y Cliente.



Figura 15, Representación de dimensiones.

Ejemplo: Para el diseño de un modelo dimensional de Transferencias se considera al monto transferido como un hecho (Fact), luego la Fact principal se pasa a llamar “monto transferido”. Esta Fact tiene como medida la cantidad de montos transferidos. A esta medida se le conoce como MÉTRICA. Toda Fact a su vez cuenta con dimensiones (tablas) que la rodean.

Supongamos que se desea realizar una transferencia de fondos, para ello se debe considerar al *beneficiario*, al *proyecto* en el que trabaja el beneficiario, los *ejecutores* de los proyectos y a la fecha en que fue transferido el monto. Todo depende de la complejidad de la transferencia y de lo que se quiera mostrar. Por ejemplo, si nos interesa saber solo el nombre del beneficiario, nuestra dimensión beneficiarios solo tendrá el nombre y nada más. Sin embargo, si nos interesa la comuna donde vive, para posteriormente agrupar nuestras transferencias por comunas, entonces se deben considerar obligatoriamente a la comuna dentro de los campos de la tabla beneficiarios.

En la Figura 16, se representa un hecho (Fact), de una transferencia y las jerarquías asociadas.

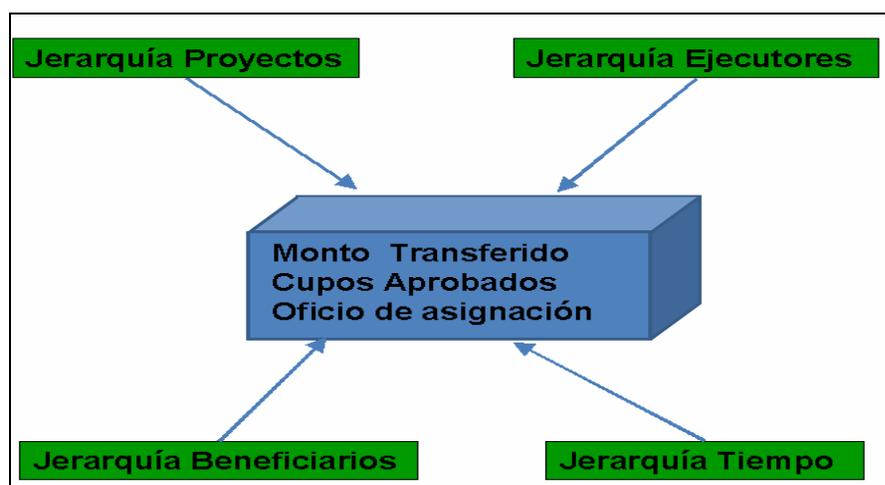


Figura 16, Hecho (Fact) de Transferencia.

2.7. Modelando el Data Mart

Para una mejor comprensión del modelamiento del Data Mart, es necesario definir tres nuevos conceptos, que son: **atributos**, **Jerarquía** y **Esquema**.

2.7.1 Atributos:

- Son las características de los miembros de las dimensiones.
- Pueden ser usados para clasificar, filtrar o buscar.
- Determinan los niveles de agregación, permiten realizar los resúmenes y las condiciones sobre los datos consultar.

Ejemplo:

Para la dimensión *Geografía*, los atributos pueden ser: Región, Provincia, Comuna, Ciudad. Como lo muestra la Figura 17,

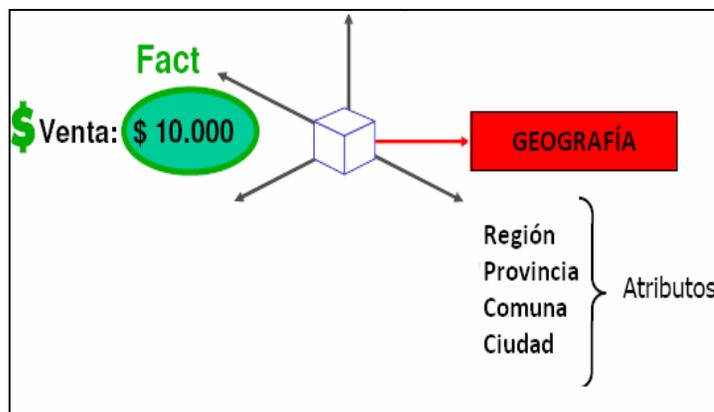


Figura 17, Representación de atributos de la dimensión Geografía.

Para la dimensión *Tiempo*, los atributos pueden ser: Año, Trimestre, Mes, Semana, Día. Como lo muestra la Figura 18,

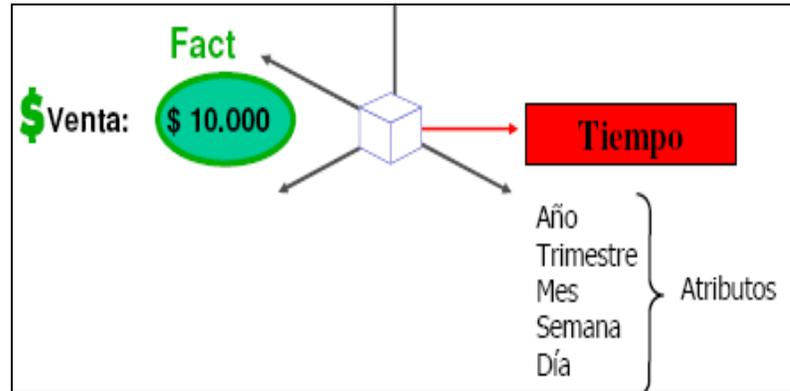


Figura 18, Representación de atributos de la dimensión Tiempo.

Los elementos de un atributo corresponden a los valores que este toma y que permite realizar calificaciones para una consulta. La Figura 19, muestra tres atributos con sus valores correspondientes.



Figura 19, Representación de elementos de los Atributos.

2.7.2 Jerarquías:

- Son niveles de agrupación dentro de una dimensión.
- También se conocen como los caminos de “drill-down” o “roll-up”.
- Son grupos de atributos relacionados.
- Reflejan las relaciones entre los datos del negocio.

La Figura 20, muestra un ejemplo de jerarquías de Tiempo, Cliente y producto.

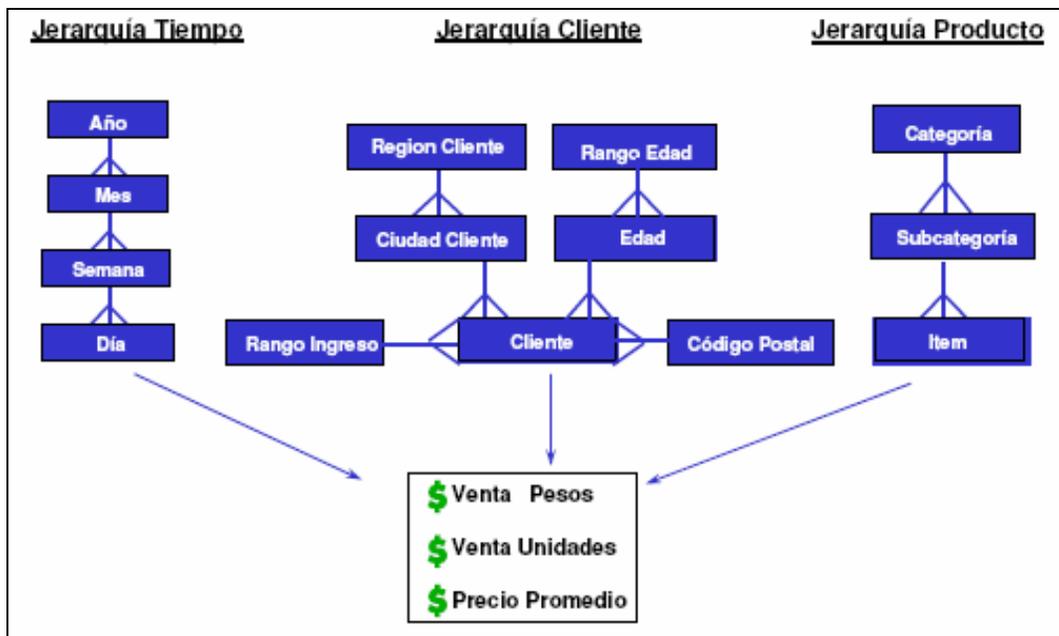


Figura 20, Ejemplo de Jerarquía.

2.7.3 Esquema:

Un esquema determina la representación física del Data Warehouse o Data Mart en tablas y columnas. Es en definitiva la componente estructural de la Bodega de Datos. Figura 21,

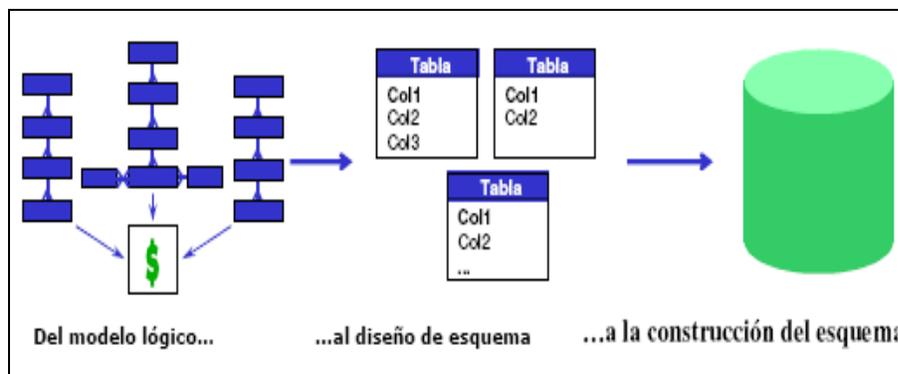


Figura 21, Construcción física de Esquema.

Las tablas y columnas en el Data Warehouse o Data Mart son utilizadas para alojar el modelo de datos Multidimensional (MDM). También conocido como Cubos OLAP.

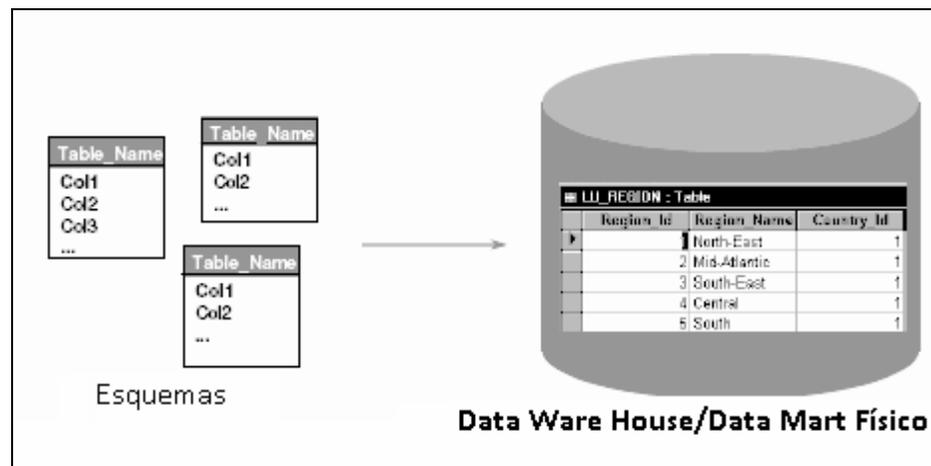


Figura 22, Construcción física del Data Warehouse/Data Mart.

Resumiendo:

Para la construcción física del Data Mart, debemos transitar desde los sistemas operacionales (OLTP), los que son soportados por bases de datos Relacionales y que gráficamente se representan por los diagramas de Entidad relación (ERD). A partir de estos diagramas y según las necesidades de información podemos diseñar el modelo lógico multidimensional (MDM), cual es la base para el diseño de esquemas, que se traducen en el modelo físico multidimensional también conocido como Data Warehouse o Data Mart.

En la Figura 23, se grafica la construcción física del Data Warehouse, desde los sistemas transaccionales hasta la creación de esquemas.

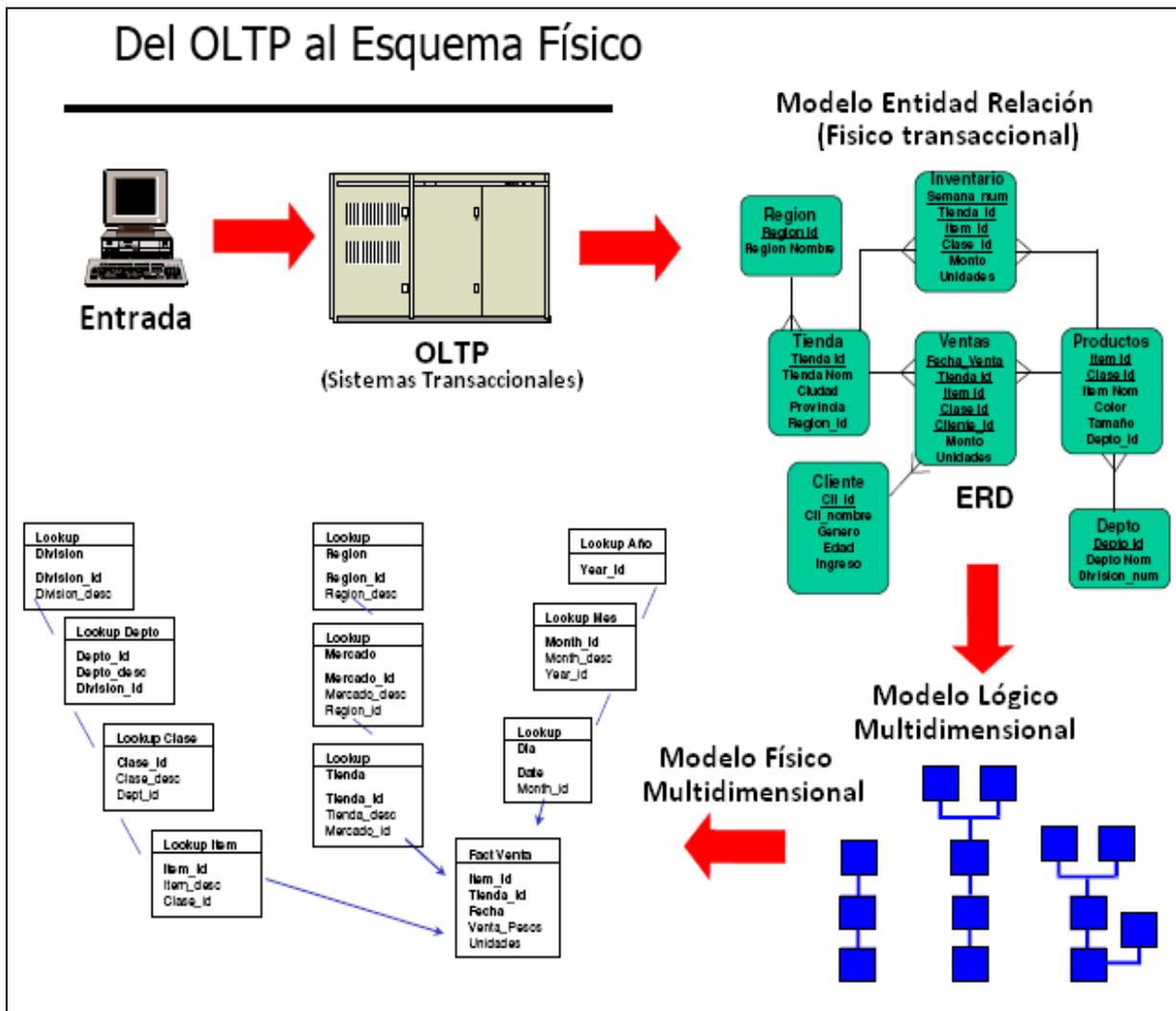


Figura 23, Construcción del Data Warehouse o Data Mart.

3. CONTEXTO

3.1. *La Unidad de Proempleo*

En Chile, las políticas activas de empleo han surgido típicamente como respuesta a procesos de crisis económica, como la que se está experimentando actualmente en el país y se prevé que nos alcanzará de lleno en el transcurso de este año. Es así como hacia fines de la década de los noventa, junto con la llegada de la crisis asiática, los instrumentos de acción pública sobre el mercado del trabajo lograron resurgir, con el objeto de paliar las altas tasas de desocupación inherentes a la crisis. Como consecuencia directa de las políticas implementadas para combatir la desocupación, *en el año 2000 la Subsecretaría del Trabajo creó la Unidad de Empleo, o Programa **Proempleo***, cuya función radica en la aplicación de las políticas activas al mercado del trabajo, a través de la supervisión y control de la ejecución de los Programas de Empleo con Aporte Fiscal (PEAF), lo cual se concreta a través de la coordinación efectiva de ellos tanto a nivel nacional, regional y local. Cabe mencionar que Proempleo divide su accionar en Programas de Empleo Directos y Programas de Empleos Indirectos.

El área de Empleos Directos se encarga de administrar y ejecutar el Programa Inversión en la Comunidad, el cual tiene por objeto el financiamiento de obras o acciones en el ámbito local, mediante proyectos intensivos en mano de obra, contratadas al efecto y que presenten un claro beneficio comunitario. El Programa se ejecuta a través de dos líneas de acción: Línea Intendencia y Línea Sence. La Línea Intendencia corresponde a la vía por la cual una entidad del sector público (principalmente los municipios) contrata trabajadores para realizar proyectos intensivos en mano de obra. A su vez, a través de la Línea Sence existe la posibilidad de realizar licitación para elegir las organizaciones, sin fines de lucro, que realizarán los proyectos, las que pueden provenir del sector privado y/o público. En su versión 2009, el Programa Inversión en la Comunidad ha

profundizado, con especial énfasis, el proceso de mejoramiento en la gestión programática, a través de la sistematización de la información del programa y una mayor coordinación entre el ente central y los organismos regionales dependientes del Programa.

Por otra parte, el área de Empleos Indirectos se encarga de administrar distintos programas que tienen por objeto mejorar los niveles de empleabilidad de trabajadores cesantes y activos. Un primer instrumento de Empleo Indirecto corresponde al Programa de Bonificación a la Contratación de Mano de Obra, el que se ejecuta mediante Sence a través de tres líneas, la Línea Regular, la Línea Adultos Chile Solidario y la Línea Jóvenes Chile Solidario. Un segundo instrumento de Empleo Indirecto es el Programa de Formación, Capacitación y Empleo, Profocap, ejecutado por la Corporación Nacional Forestal, CONAF. Un tercer instrumento de Empleo Indirecto lo constituye el Programa de Generación de Microemprendimiento Indígena Urbano, ejecutado por la Corporación Nacional de Desarrollo Indígena, CONADI. Un cuarto instrumento de Empleo Indirecto corresponde al Programa de Desarrollo de Competencias Laborales para Mujeres Chile Solidario, ejecutado por la Fundación para la Promoción del Desarrollo de la Mujer, PRODEMU. Por último, destaca el Programa de Servicios Sociales, ejecutado por la Fundación para la Superación de la Pobreza, y tiene por objeto ofrecer capacitación en servicios comunitarios a sectores vulnerables de la población.

3.2. Beneficiarios y ejecución de proyecto:

Como fue explicado anteriormente, el área de Empleos Directos se encarga de administrar y ejecutar el Programa Inversión en la Comunidad, el cual tiene por objeto el financiamiento de obras o acciones en el ámbito local, mediante proyectos que hacen uso intensivos en mano de obra.

Los actores principales de los Empleos Directos son:

3.2.1 Beneficiarios

Los beneficiarios del programa serán trabajadores o trabajadoras que cumplan los siguientes requisitos:

- a) Se encuentren desempleados.
- b) Sean jefes o jefas de hogar.
- c) Se encuentren inscritos en las respectivas OMIL (Oficina Municipal de Información Laboral), o en las instituciones que en la Municipalidad cumplan funciones similares a la OMIL, o en la Bolsa Nacional de Empleo creada al efecto por el Seguro de Desempleo.

3.2.2 Ejecución de Proyectos

Los órganos que intervienen en la ejecución de los proyectos son:

a) Intendencias Regionales:

Les corresponde, por una parte, presentar y postular ante la Subsecretaría del Trabajo, Programa Proempleo, área Inversión en la Comunidad, los proyectos a ejecutar en la Región; y por otra, administrar los recursos que esta última les transfiera para su ejecución, así como también, determinar los procedimientos de contratación de los ejecutores, suscribir con los ejecutores finales los contratos para la ejecución de los programas, mantener el control financiero y supervisar y controlar la ejecución del proyecto respectivo. Estas dos últimas funciones, las Intendencias pueden encomendarlas a otras instituciones públicas con presencia regional, mediante la suscripción de convenios de colaboración.

b) Subsecretaría del Trabajo:

Transfiere los recursos financieros necesarios a las Intendencias para la ejecución de los proyectos, según cobertura y de acuerdo a los decretos relacionados. Asimismo, le corresponde supervisar y fiscalizar en terreno la ejecución de los programas de empleo.

c) Ejecutor de los proyectos:

Podrán constituirse como ejecutores de los proyectos las Municipalidades o agentes privados, esto es, personas jurídicas con o sin fines de lucro, los que se relacionarán con las Intendencias regionales para la ejecución de los proyectos respectivos, mediante la celebración del contrato correspondiente.

3.2.3 Procedimientos en la Ejecución de Proyectos

a) Ordinario de Cobertura:

El procedimiento se inicia con el Ordinario de Cobertura; que consiste en el acto oficial mediante el cual el Subsecretario del Trabajo comunica a los intendentes regionales el Marco Presupuestario Regional que indica el monto de la cobertura del programa para la región y comuna respectiva. De acuerdo a la ley de responsabilidad fiscal y la ley de presupuesto del año correspondiente, el comité de seguimiento de los planes de empleo tiene la facultad de asignar coberturas adicionales y aumentos de jornada a las localidades que tengan una tasa igual o superior al 9% de desocupación.

b) Oficio de Asignación de Recursos:

Recibido el Ordinario de Cobertura el Intendente Regional deberá enviar un Oficio Ordinario de Asignación de Recursos, dirigido al Subsecretario del Trabajo. En dicho documento el Intendente solicitará se asignen recursos para la ejecución de una cartera de proyectos, incluyendo una descripción completa de ellos.

A dicho oficio deberá adjuntarse los siguientes documentos:

- i) Listado de Proyectos: La Intendencia deberá elaborar y adjuntar este listado expresando fielmente los datos y cifras generales de la cartera de proyectos presentada, indicando en forma resumida para cada proyecto; Nombre, monto solicitado expresado en pesos, número de empleos, duración expresada en meses, y comuna.
- ii) Ficha de Postulación de Proyectos: Para cada proyecto que se postule deberá adjuntarse una ficha con la información descriptiva pormenorizada, incluyendo datos de postulación, productos y/o servicios a realizar, el tiempo de ejecución, presupuesto total y responsables del proyecto.
- iii) Ficha de Registro de Ejecutores: En esta ficha deberán registrarse los datos que individualicen al órgano ejecutor. Además, en ella la Intendencia deberá declarar los documentos y la información que hubiere recabado de cada uno de los órganos ejecutores y que se encuentran en su poder.

c) Decreto :

Una vez revisados los documentos, se realiza el decreto de los proyectos por parte de la Subsecretaría del Trabajo. El decreto debe estar en concordancia con la información de proyectos remitidos de la Intendencia.

d) Contrato entre Intendencia y cada ejecutor:

Corresponde a cada Intendencia la responsabilidad de determinar la modalidad de contratación (licitación pública o privada, o por trato directo) y de celebrar el contrato con el ejecutor de los proyectos, velando por el resguardo y correcto uso de los recursos públicos dispuestos.

e) Listado de Beneficiarios del Proyecto:

Durante la primera quincena de ejecución de los proyectos la Intendencia Regional respectiva deberá remitir a la Subsecretaría del Trabajo, Programa Proempleo, área Inversión en la Comunidad, el Listado de Beneficiarios del Proyecto, en él debe incluirse la información que individualiza a cada una de las personas Beneficiarias del Programa en cada proyecto.

f) Ficha de Beneficiarios:

Antes del término del primer mes de ejecución de los proyectos la Intendencia Regional respectiva deberá remitir a la Subsecretaría del Trabajo, Programa Proempleo, área Inversión en la Comunidad, la Ficha de Beneficiarios que contiene la identificación detallada de la persona y la situación social y laboral del beneficiario.

g) Transferencia :

Una vez que se encuentre la documentación de los beneficiarios, se hace efectiva la transferencia de recursos de los proyectos. Esta se hace en forma parcelada dependiendo del período de ejecución.

h) Cumplimiento Resolución N° 759:

Además, según Resolución N° 759 (23/1 2/2003), de la Contraloría General de la República, publicada en el Diario Oficial el 17 de enero de 2004, la Intendencia Regional debe remitir a la Subsecretaría del Trabajo, Programa Proempleo, área Inversión en la Comunidad, los siguientes antecedentes:

- i) Comprobante de ingreso, las intendencias deberán enviar este comprobante cada vez que se les realice una transferencia de recursos, el incumplimiento de este requisito dificultará futuras transferencias.

- ii) Informe mensual de la inversión de los recursos, el que debe señalar, a los menos, el monto de los recursos recibidos en el mes, el monto detallado de la inversión realizada y el saldo disponible para el mes siguiente. Ante la ausencia o el retraso en la rendición de este informe no se procederá a realizar nuevas transferencias.

i) Rendición de cuentas:

Dentro de un plazo de 30 días contados desde la primera transferencia la Intendencia Regional deberá informar a la Subsecretaría del Trabajo, Programa Proempleo, área Inversión en la Comunidad, la situación financiera del proyecto mediante la rendición cuentas, respecto de la ejecución del proyecto, los gastos efectuados y los saldos por rendir del proyecto.

j) Finalizado el Proyecto:

Dentro de un plazo de 30 días contados desde la fecha de término de la ejecución de cada proyecto, la Intendencia Regional deberá informar a la Subsecretaría del Trabajo, Programa Proempleo, área Inversión en la Comunidad, el cierre del proyecto y deberá rendir cuentas respecto de la ejecución del proyecto, los gastos efectuados y los saldos que deba poner a disposición de la Subsecretaría del Trabajo.

La Figura 24, representa el flujo del procedimiento descrito.

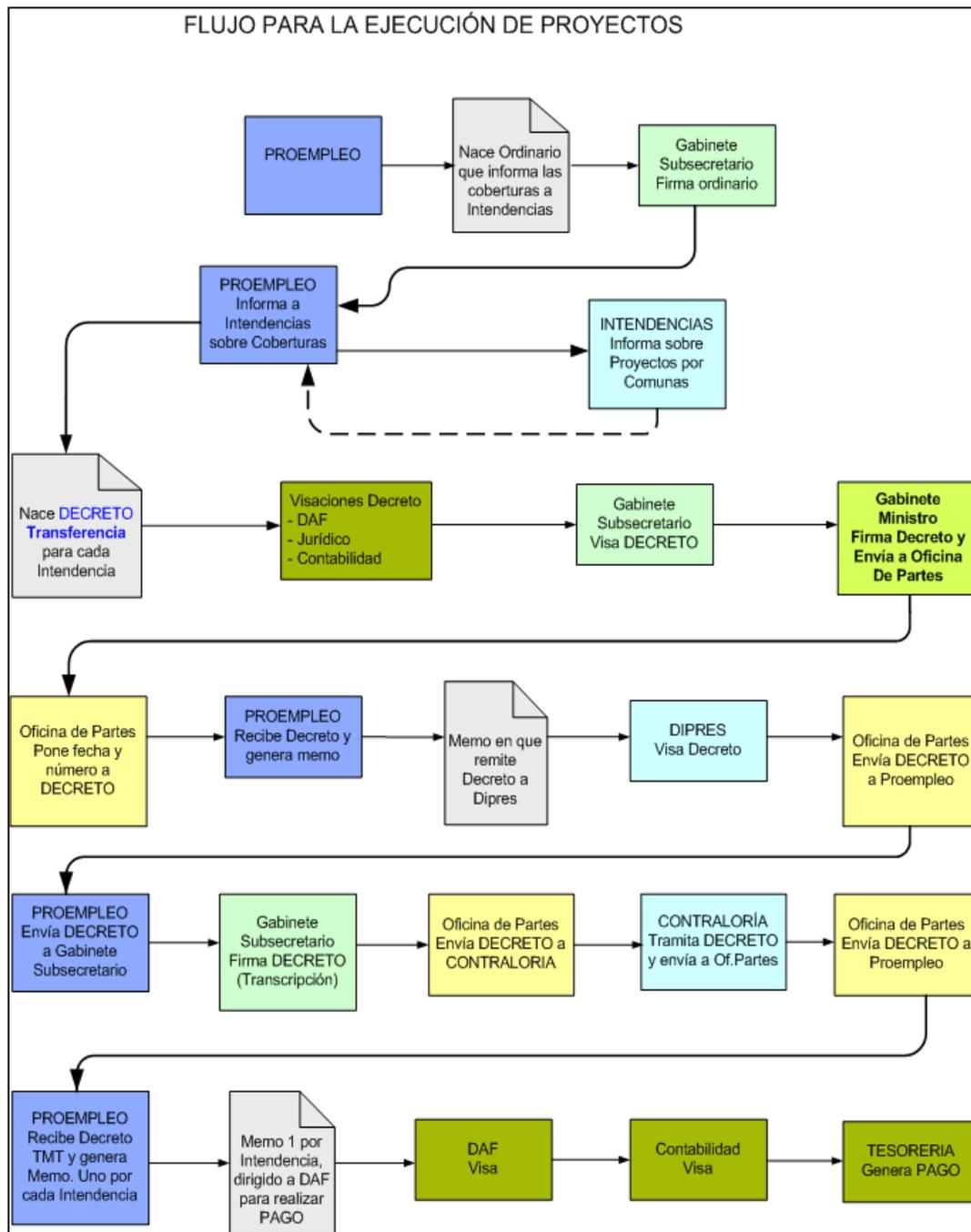


Figura 24, Flujo del procedimiento para la Ejecución de Proyectos.

3.3. Descripción del Problema

Dados los volúmenes de información que maneja Proempleo, la cantidad de procesos y procedimientos que debe controlar como parte de su operación diaria, la toma de decisiones tiene que involucrar una cantidad enorme de variables cuya movilidad en el tiempo no permiten tener una visión clara del entorno. Dichas variables de información tienen que ver con la administración relacionada con los procedimientos en la ejecución de los distintos proyectos de generación de empleo descritos en el capítulo anterior.

Para Proempleo, el mayor volumen de información que debe procesar se encuentra en los programas de Inversión en la Comunidad y de Bonificación a la contratación de Mano de Obra, por tratarse de los programas de mayor cobertura en regiones y montos.

Por lo general, las decisiones de la autoridad del Ministerio del Trabajo con respecto a proyectos de generación de empleo o las consultas de entidades contraloras del país (Contraloría General de la República, Cámara de Diputados, Senado de la República), sobre el empleo de emergencia, radica básicamente en la información sobre las transferencias de fondos, coberturas y rendiciones relacionadas justamente con los proyectos los programas de inversión en la comunidad a lo largo del país.

Proempleo cuenta con sistemas Operacionales (OLTP), que permiten el registro de los datos relativos al negocio; proyectos, beneficiarios, ejecutores, transferencias, coberturas y rendiciones. Pero como se indicó en el capítulo *Data Warehouse*, los sistemas transaccionales nunca han sido diseñados para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos.

Es por ello que el problema se puede definir como la falta de una herramienta informática que: proporcione información ejecutiva consolidada, permita resolver los requerimientos de información y que a la vez sirva de base para la toma de decisiones a las autoridades del Ministerio del Trabajo y Previsión Social.

Si lo anterior se cumple y se obtiene la herramienta informática, Proempleo podrá aplicar mejoras a la gestión de los procesos de transferencia y rendiciones de fondos que se comprometen y ejecutan en la generación *de los Programas de Empleo de Emergencia en Chile*.

3.4. Solución propuesta

Si no consideramos el estudio realizado en esta tesis, pareciera que utilizar los sistemas operacionales de Proempleo, es lo más razonable y sencillo para dar solución a parte de la problemática planteada en el título anterior, ya que se pueden realizar consultas y realizar reportes sobre los datos que estos sistemas operan y que sirvan para análisis.

Sin embargo, hemos dejado claro en el desarrollo de esta tesis, que sobre los sistemas operacionales, los reportes que se generan son estáticos y las consultas hay que generarlas cada vez que sean requeridas, siendo los resultados cada vez diferentes, degradando con este accionar la calidad de estos sistemas y aumentando la carga de trabajo a los programadores. Los informes así generados son costosos y muchas veces desfasados en el tiempo, porque alguien debe programar las consultas y luego ejecutarlas y en ese lapso podría, incluso, pasar días. Luego el informe que era necesario ayer, hoy, puede no serlo. Hay carencia de oportunidad de la información.

Por lo anterior y teniendo en cuenta que se ha definido como uno de los alcances de este trabajo el desarrollo de una herramienta para gestionar las transferencias y rendiciones de fondos a los distintos organismos encargados de la administración de los Empleos de Emergencia del País y que además sirva de apoyo a toma de decisiones, la solución propuesta es la implementación de la metodología de Data Warehouse. Específicamente la construcción de un Data Mart para la unidad de Proempleo del Ministerio del Trabajo y Previsión Social.

Fundamentando lo anterior es oportuno repasar algunos de los conceptos vertidos en esta tesis: Un Data Warehouse contiene la información de toda la Organización (Para este estudio, el Ministerio del Trabajo y Previsión Social), cualquier departamento puede acceder a la información de cualquier otro

departamento mediante un único medio, así como obligar a que los mismos términos tengan el mismo significado para todos. Un Data Mart almacena la información de un área o departamento específico, y un conjunto de Data Mart forman un Data Warehouse.

Es por ello que se puede afirmar que Data Mart es una herramienta que permite dar el soporte de información útil y consolidada, requerida por Proempleo, que es una unidad (o departamento) del Ministerio del Trabajo.

De lo anterior se desprende que, utilizar *tecnología Data Warehouse* en la construcción de un Data Mart para Proempleo, es la solución más óptima a la vista de los antecedentes aportados en este estudio, ya que se cubre de manera óptima las necesidades de información que necesita la unidad.

3.5. Justificación de solución

Como la mayoría de las organizaciones que necesitan información sintetizada para análisis y que sirvan de base para la toma de decisiones, Proempleo también tiene necesidad de este tipo de información, ya que, actualmente sus sistemas Transaccionales (que manejan grandes volúmenes de datos), no son capaces de entregar.

Parte de la problemática de Proempleo, en el sentido de obtención de información útil, es la falta de una herramienta informática que proporcione información ejecutiva consolidada.

Como parte de la investigación de esta tesis, ha quedado de manifiesto que utilizar los sistemas transaccionales para la obtención información sintetizada, y para análisis, es muy costoso, tanto en horas hombre como en la degradación de estos sistemas y que en muchas ocasiones la información obtenida directamente de estos sistemas es extemporánea.

Recordemos lo que dice Dr. Edgar Frank Codd, a quien se le conoce como el creador del modelo relacional: “Aunque los Sistema de Gestión de Base de Datos Relacional han sido tan beneficiosos para los usuarios, nunca han sido diseñados para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos”.

Como la solución propuesta es la **construcción de un Data Mart**, se quiere enfatizar que el término “Data Warehouse” se utiliza indistintamente para hablar de la tecnología en sí como también para uno de los componentes que la conforman, específicamente al repositorio de los datos. También dar énfasis a que un Data Mart es conceptualmente lo mismo que un Data Warehouse.

Aclarado lo anterior, es posible asegurar, basándose en el estudio realizado en este trabajo, que una de las tecnologías más utilizadas hoy en día para solucionar problemas en la obtención de información consolidada y que sirva de base para la toma de decisiones a las organizaciones, es la implementación de Data Warehouse, tanto en su contexto tecnológico como en el de bodega de datos.

A la vista de la solución propuesta y con los antecedentes expuestos, se justifica construir un Data Mart para Proempleo ya que con la implementación de esta solución informática, se resuelve el problema planteado.

4. DESARROLLO

4.1. Marco Metodológico

Para el desarrollo del Data Mart para la unidad de Proempleo del Ministerio del Trabajo y Previsión Social, se hará uso de la metodología Kimball.

La metodología de Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos (FACTS) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información.

“Ralph Kimball, es reconocido como uno de los padres del concepto de Data Warehouse, se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.”

El marco metodológico presentado por Ralph Kimball con el nombre de Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle BDL), ilustra las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. Este enfoque de implementación de Data Warehouse se ilustra en la Figura 25,

Este diagrama ilustra tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación de Data Warehouse.

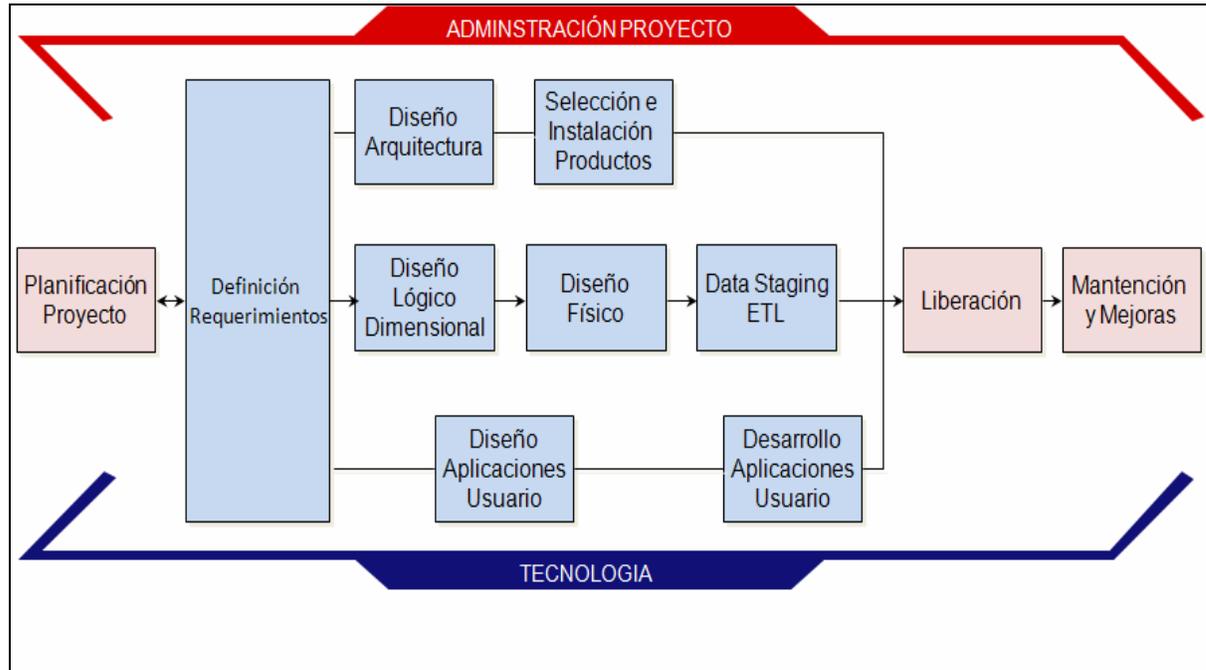


Figura 25, Ciclo de Vida Dimensional del Negocio.

A continuación se describe cada una de las etapas del Ciclo de Vida Dimensional del Negocio y los productos (Entregables), que estas etapas entregan.

4.1.1 Planificación de Proyecto:

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

La planificación del proyecto se focaliza sobre recursos, perfiles, tareas, duraciones y secuencias. El plan de proyecto resultante identifica todas las tareas asociadas con el Ciclo de Vida Dimensional del Negocio e identifica las partes involucradas.

La planificación del proyecto depende de los requerimientos de negocio, como se muestra en la Figura 25, que representa al ciclo de vida, con flecha de doble dirección entre estas actividades.

Tareas Representativas de esta etapa:

DEFINICIÓN PROYECTO
1 Realizar Setup del Proyecto
2 Definición Inicial de Alcances del Proyecto
3 Relevar las Justificaciones de Negocio
PLANIFICACION y ADMINISTRACION DE PROYECTO
1 Establecer la Identidad del Proyecto
2 Identificar Recursos del Proyecto
3 Preparar el plan de Proyecto Preliminar
4 Kick-Off Inducción Equipo Proyecto y Planificación
5 Revisión del Plan de Proyecto
6 Desarrollar Plan de Comunicación del Proyecto
7 Definir los Indicadores de éxito del Proyecto
8 Definir Procedimientos de Administración de Alcances
9 Puesta en Marcha la Administración de Proyecto

Productos de esta etapa: Carta Gantt del Proyecto. Figura 26,

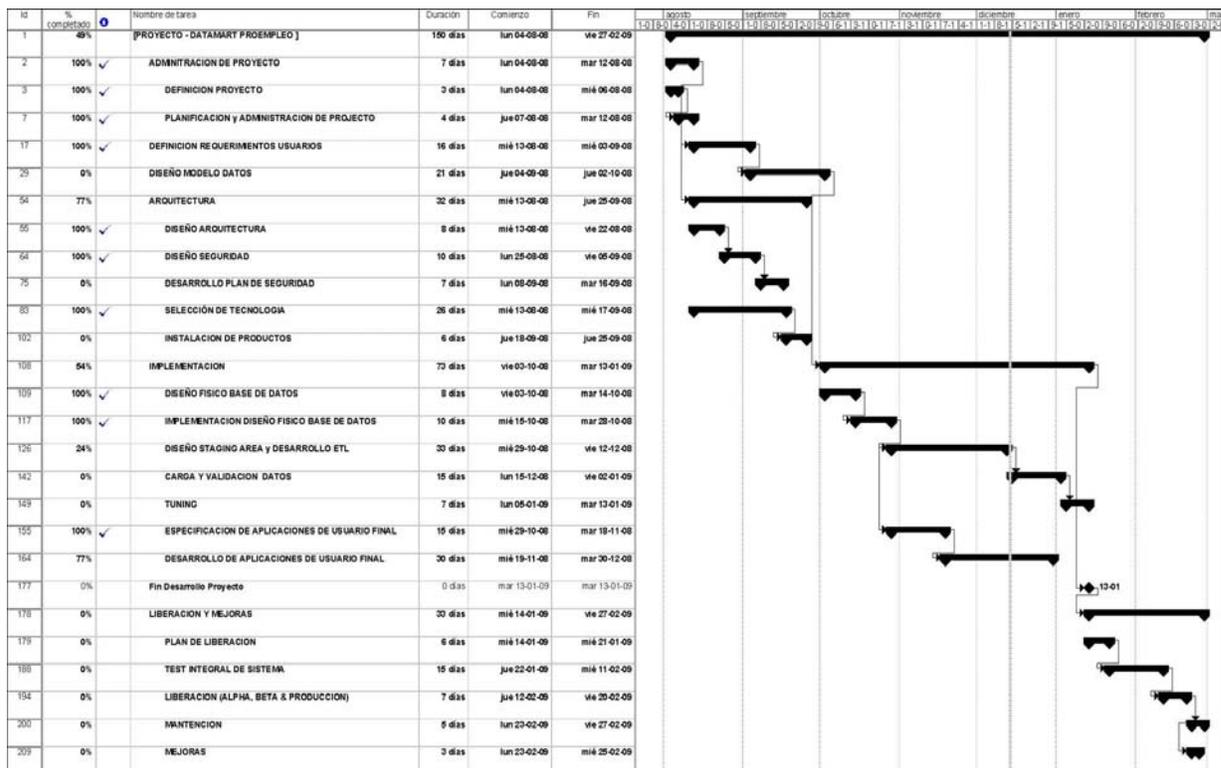


Figura 26, Carta Gantt del Proyecto Data Mart Para Proempleo.

4.1.2 Definición de Requerimientos:

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

Los diseñadores de los data Warehouse deben entender los factores claves que guían al negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.

La definición de los requerimientos del negocio establece la base para las tres etapas paralelas focalizadas en la tecnología, los datos y las aplicaciones.

Productos de esta etapa:

Set de preguntas de negocio expresadas por diferentes niveles de usuarios, las que indican los requerimientos de información de parte de ellos.

- ¿Cuántas personas se retiran de los programas de Empleo por edad y género?
- ¿Cuál es el costo por fueros y finiquitos por jornada y cantidad de beneficiarios de las comunas de una Región, de la línea Intendencia?
- ¿Quiénes son parientes dentro de los beneficiarios y a qué comunas pertenecen?
- ¿Cuál es la cantidad de beneficiarios y el monto transferido a las comunas de una región por tipo de jornada de la línea Intendencia?
- ¿Hay ejecutores que son beneficiarios?
- ¿Hay variación del índice de cesantía en las comunas v/s cupos de empleo de emergencia?
- ¿Las rendiciones mensuales de las comunas se encuadran con los montos asignados a ellas?

Muchas otras preguntas interesantes no se ha podido plasmar en esta tesis ya que podrían comprometer intereses estratégicos del Ministerio del Trabajo.

4.1.3 Diseño Lógico Dimensional:

La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos analíticos de los usuarios. Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales. Básicamente se comienza con una matriz donde se determina la dimensionalidad de cada indicador y

luego se especifican los diferentes grados de detalle (atributos) dentro de cada concepto del negocio (dimensión), como así también la granularidad de cada indicador (variable o métrica) y las diferentes jerarquías que dan forma al modelo dimensional del negocio (BDM) o mapa dimensional.

Productos de esta etapa: Modelo lógico Multidimensional, descrito en el capítulo “DESARROLLO” de la página 72.

4.1.4 Diseño Físico:

El diseño físico de las bases de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y seteos específicos del ambiente de la base de datos. La indexación y las estrategias de particionamiento son también determinadas en esta etapa.

Productos de esta etapa: El modelo Físico Multidimensional, descrito en el capítulo “DESARROLLO”, representado en la Figura 44 de la página 78.

4.1.5 Data Staging ETL:

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de Data Warehouse. Los principales procesos de esta etapa del ciclo de vida son: la extracción, la transformación y la carga (ETL). Se definen como procesos de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico acordado. Así mismo, se definen como procesos de transformación a los procesos para convertir o reescribir los datos fuente a fin poder efectuar la carga efectiva del Modelo

Físico. Por otra parte, los procesos de carga de datos son los procesos requeridos para poblar el Data Warehouse.

Todos estos procesos son altamente críticos pues tienen que ver con la materia prima del Data Warehouse: **los datos**. La desconfianza y pérdida de credibilidad del Data Warehouse serán resultados inmediatos e inevitables si el usuario choca con información inconsistente. Es por ello que la calidad de los datos fuentes (OLTP), **son un factor determinante en el éxito de un proyecto de Data Warehouse**. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente.

Productos de esta etapa:

Procesos desarrollados en procedimientos almacenados de base datos y desarrollados por los analistas del sistema operacional como se explica en el título: Modelo Entidad Relación para Proempleo. De la página 69.

4.1.6 Diseño de Arquitectura:

Los ambientes de Data Warehousing requieren la integración de numerosas tecnologías. Se debe tener en cuenta tres factores: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de Data Warehousing.

Productos de esta etapa:

Arquitectura de la solución implementada, representada en la Figura 27,

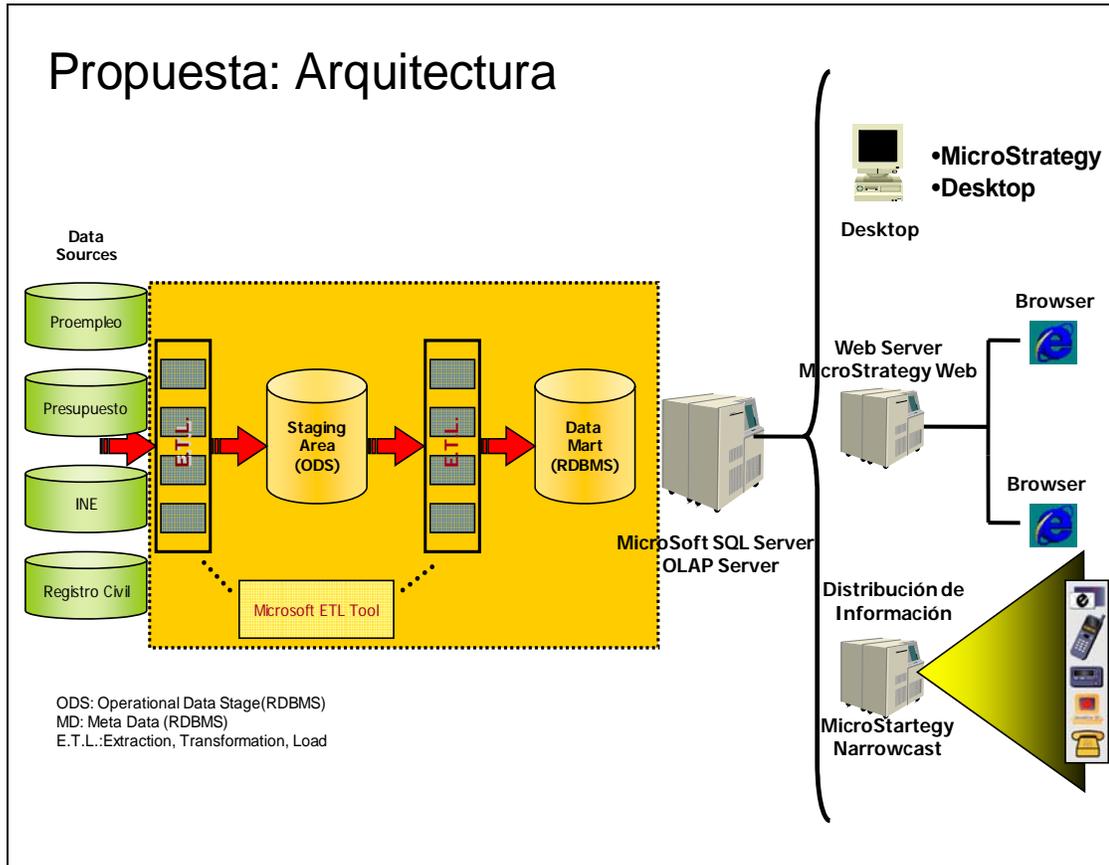


Figura 27, Arquitectura de la solución implementada.

4.1.7 Selección e Instalación de Productos:

Utilizando el diseño de arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura como ser la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc.

Una vez evaluados y seleccionados los componentes determinados, se procede con la instalación y prueba de los mismos en un ambiente integrado de Data Warehousing.

Las tareas realizadas que concluyeron con la selección de la plataforma, pasaron por las siguientes etapas de chequeo:

SELECCIÓN DEL PRODUCTO	
	(Repita este procedimiento para cada área de selección)
1	Desarrollar matriz de evaluación
2	Investigar productos candidatos
3	Elaborar una lista breve de producto
4	Evaluar las opciones de los productos
5	Prototipos Opcional (Puede repetir para diferentes productos)
	Seleccione los procesos de negocio/ Datos para la Evaluación
	Definir criterios de finalización
	Adquirir los recursos (internos y proveedores)
	Determinar la configuración de prueba
	Instalar requisitos previos de evaluación y componentes
	Capacitar el equipo de evaluación
	Desarrollar y afinar prototipo
	Pruebas de conducta
	Analizar y documentar los resultados
6	Determinar recomendaciones al producto
7	Presentar las conclusiones y resultados a la gestión
	Negociar contrato
9	Aceptación de los Usuario/Evaluación de Proyectos
INSTALACIÓN DEL PRODUCTO	
	(Repita para cada uno de los productos)
1	Planificación de la Instalación
2	Conoce Requisitos previos
3	Instalación de Hardware / Software

4	Pruebas al Hardware / Software
5	Aceptación de los Usuario/Evaluación de Proyectos

4.1.8 Diseño Aplicaciones Usuarios:

No todos los usuarios del Data Warehouse necesitan el mismo nivel de análisis. Es por ello que en esta etapa se debe identificar los roles o perfiles de los usuarios, lo que permitirá determinar las aplicaciones que deberá usar cada uno de ellos en sus niveles (gerencial, analista del negocio, vendedor, etc.).

Productos de esta etapa:

Reportes que den respuesta a preguntas de negocio.

En las Figura 28, **Figura 29**, se muestran ejemplos de reportes que responden a las preguntas:

- a. ¿Cuál es el costo por fueros y finiquitos por jornada y cantidad de beneficiarios de las comunas de una Región, de la línea Intendencia?
- b. ¿Cuál es la cantidad de beneficiarios y el monto transferido a las comunas de una región por tipo de jornada de la línea Intendencia?

V REGION	COMUNA	Costos por Fueros y Finiquitos - Línea Intendencia					
		Nº	1/2 JORNADA	3/4 JORNADA	1 JORNADA	Costo Mensual	COSTO NOV-DIC
VALPARAISO	VALPARAISO	968	362	0	606	153.929.232	307.858.464
TOTAL		968	362	0	606	153.929.232	307.858.464
FUEROS	VALPARAISO	19	4	0	15	3.115.508	6.231.016
	VIÑA DEL MAR						
TOTAL		1.955	728	0	1.227	3.115.508	6.231.016
FINIQUITOS	VALPARAISO	0	1	0	0	432.000	432.000
	VIÑA DEL MAR	0	0	0	0	0	0
TOTAL		0	1	0	0	432.000	432.000
TOTAL GENERAL		1.955	729	0	1.227	157.476.740	314.521.480

Figura 28, Ejemplo de diseño; Costos por fueros y finiquitos.

	COMUNA	Transferencias Línea Intendencia					
		Nº	1/2 JORNADA	¾ JORNADA	1 JORNADA	Costo Mensual	COSTO AGOSTO
VIII REGION							
CONCEPCION	LOTA	589	0	0	589	103.307.066	103.307.066
	TALCAHUANO	525	0	0	525	92.081.850	92.081.850
	PENCO	80	0	0	80	14.031.520	14.031.520
	SAN PEDRO DE LA PAZ	120	0	20	100	20.246.860	20.246.860
	CONCEPCIÓN	25	0	19	6	3.624.451	3.624.451
	TOMÉ	247	0	0	247	43.322.318	43.322.318
ARAUCO	CURANILAHUE	153	0	0	153	26.835.282	26.835.282
	LOS ALAMOS	189	0	0	189	33.149.466	33.149.466
	LEBU	316	0	0	316	55.424.504	55.424.504
	ARAUCO	48	0	0	48	8.418.912	8.418.912
	CAÑETE	30	0	0	30	5.261.820	5.261.820
	TIRÚA	65	0	0	65	11.400.610	11.400.610
	CONTULMO	105	0	0	105	18.416.370	18.416.370
TOTAL		2.492	0	39	2.453	435.521.029	435.521.029

Figura 29, Ejemplo de diseño; Transferencias Línea Intendencia.

NOTA: Los datos y montos señalados en las Figura 28, Figura 29, son de ejemplo y no corresponden a la realidad.

4.1.9 Desarrollo Aplicaciones Usuarios:

De acuerdo a los requerimientos expresados por los usuarios finales, el desarrollo de las aplicaciones para ellos involucra configuraciones del metadata y construcción de reportes específicos.

Óptimamente, estas aplicaciones se deben construir utilizando herramientas avanzadas de acceso a los datos, lo que entregaría un importante aumento de la productividad para el equipo de desarrollo de aplicaciones. Además, entrega un poderoso mecanismo para que los usuarios de negocios puedan modificar fácilmente las plantillas de informe.

Productos de esta etapa:

Reportes gráficos o pantallas desarrollados por la herramienta OLAP que den respuesta a preguntas de negocio.

En las figuras 30, 31 y 32. Se muestran resultados de pantallas gráficas y reportes que responden a las siguientes preguntas:

- ¿Cuál es el gasto total por regiones y por tipo de jornada que se ha transferido?

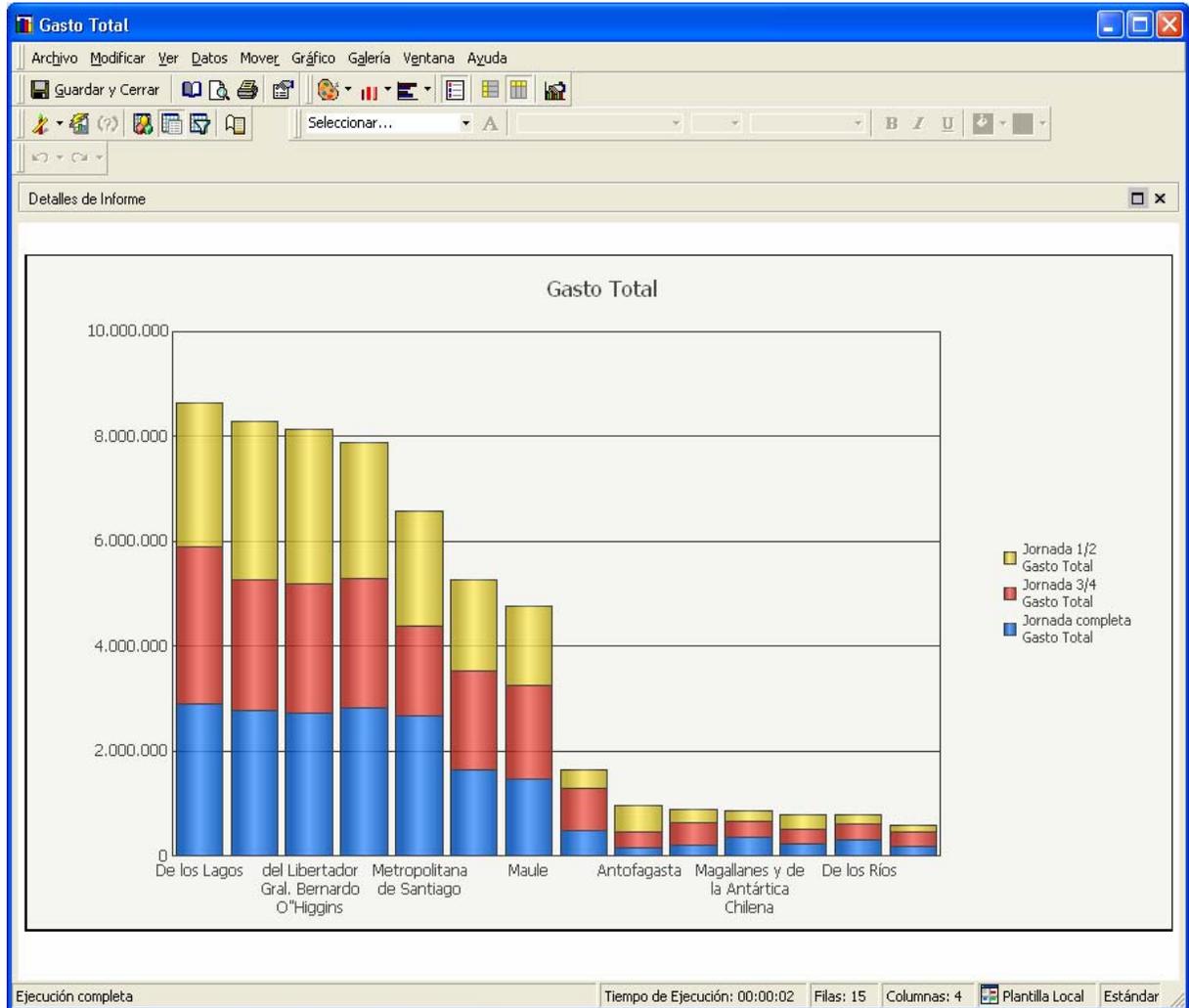
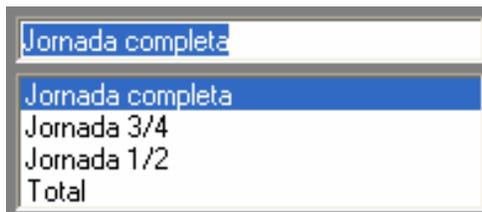


Figura 30, Informe gráfico desarrollado por la herramienta de consulta OLAP.

- ¿Cuál es el gasto total, de Finiquitos y Fueros Maternales por Comunas de las Intendencias y por tipo de jornada que se ha transferido?



Detalle Gasto Intendencias

Archivo Modificar Ver Insertar Formato Datos Cuadrícula Mover Ventana Ayuda

Guardar y Cerrar Colorido

Tipo Jornada: Jornada completa

Intendencia	Comuna Proyecto	Indicadores	Gasto Total	Finiquitos	Fuero Maternal
De los Lagos	LOS M				
del Libertador Gral. Bernardo O'Higgins	MARCHEL				
Metropolitana de Santiago	SAN JOSE		35.561	3.875	966
Metropolitana de Santiago	LA FLORIDA		477.432	52.046	12.976
De los Lagos	MAULLIN		299.563	32.652	8.147
Valparaíso	VILLA ALEMANA		135.937	14.820	3.696
del Libertador Gral. Bernardo O'Higgins	SAN VICENTE		125.204	13.648	3.406
De los Lagos	PUERTO VARAS		362.690	39.543	9.865
del Libertador Gral. Bernardo O'Higgins	PEUMO		86.541	9.437	2.354
De los Lagos	CALBUCO		42.471	4.630	1.153
del Libertador Gral. Bernardo O'Higgins	LAS CABRAS		289.443	31.553	7.868
Bío Bío	ALTO BIO BIO		322.312	35.135	8.761
Valparaíso	LA CRUZ		28.951	3.157	788
Antofagasta	MARIA ELENA		11.339	1.236	309
Valparaíso	ALGARROBO		168.993	18.421	4.596
del Libertador Gral. Bernardo O'Higgins	PUMANQUE		237.497	25.891	6.457
Araucanía	VICTORIA		11.985	1.306	326
Maule	MOLINA		53.898	5.876	1.465
Araucanía	TOLTEN		28.802	3.141	783
Bío Bío	LAJA		81.020	8.833	2.205
Valparaíso	EL TABO		136.119	14.841	3.703

Ejecución completa Tiempo de Ejecución: 00:00:01 Filas: 339 Columnas: 3 Plantilla Local Estándar

Figura 31, Ejemplo de consulta dinámica generada por la herramienta OLAP.

- ¿Cuál es el gasto total, de Finiquitos y Fueros Maternales por Comunas de las Intendencias y por tipo de proyectos que se ha transferido?

Intendencia	Comuna Proyecto	Tipo Proyecto	Indicadores	Gasto Total	Finiquitos	Fuero Maternal
Aisén del Gral. Carlos Ibáñez del Campo	COYHAIQUE	NA		96.855	10.558	2.633
Aisén del Gral. Carlos Ibáñez del Campo	COYHAIQUE	Total		96.855	10.558	2.633
Aisén del Gral. Carlos Ibáñez del Campo	LAGO VERDE	Adulto Mayor		862	94	23
Aisén del Gral. Carlos Ibáñez del Campo	LAGO VERDE	Jovenes		10.415	1.135	283
Aisén del Gral. Carlos Ibáñez del Campo	LAGO VERDE	NA		9.188	1.002	250
Aisén del Gral. Carlos Ibáñez del Campo	LAGO VERDE	Total		20.465	2.231	556
Aisén del Gral. Carlos Ibáñez del Campo	AYSEN	Adulto Mayor		10.954	1.194	298
Aisén del Gral. Carlos Ibáñez del Campo	AYSEN	NA		35.733	3.896	972
Aisén del Gral. Carlos Ibáñez del Campo	AYSEN	Total		46.687	5.090	1.270
Aisén del Gral. Carlos Ibáñez del Campo	CISNES	Adulto Mayor		30.443	3.319	828
Aisén del Gral. Carlos Ibáñez del Campo	CISNES	Jovenes		4.341	473	118
Aisén del Gral. Carlos Ibáñez del Campo	CISNES	NA		17.983	1.961	490
Aisén del Gral. Carlos Ibáñez del Campo	CISNES	Total		52.767	5.753	1.436
Aisén del Gral. Carlos Ibáñez del Campo	GUAITECAS	Adulto Mayor		5.764	628	157
Aisén del Gral. Carlos Ibáñez del Campo	GUAITECAS	Jovenes		13.379	1.458	363
Aisén del Gral. Carlos Ibáñez del Campo	GUAITECAS	NA		34.259	3.735	932
Aisén del Gral. Carlos Ibáñez del Campo	GUAITECAS	Total		53.402	5.821	1.452
Aisén del Gral. Carlos Ibáñez del Campo	CHILE CHICO	Adulto Mayor		50.237	5.478	1.367
Aisén del Gral. Carlos Ibáñez del Campo	CHILE CHICO	Jovenes		17.742	1.934	483
Aisén del Gral. Carlos Ibáñez del Campo	CHILE CHICO	NA		10.423	1.136	283
Aisén del Gral. Carlos Ibáñez del Campo	CHILE CHICO	Total		78.402	8.548	2.133
Aisén del Gral. Carlos Ibáñez del Campo	RIO IBAÑEZ	Adulto Mayor		8.476	924	230
Aisén del Gral. Carlos Ibáñez del Campo	RIO IBAÑEZ	Jovenes		19.781	2.156	537
Aisén del Gral. Carlos Ibáñez del Campo	RIO IBAÑEZ	NA		5.406	500	140

Figura 32, Ejemplo de consulta dinámica generada por la herramienta OLAP.

Como productos de esta etapa también se encuentran la representación gráfica de las jerarquías y los diseños de Navegación, realizadas por la herramienta OLAP definida en la etapa de Arquitectura, como lo muestran las figuras 33, 34, 35 y 36.

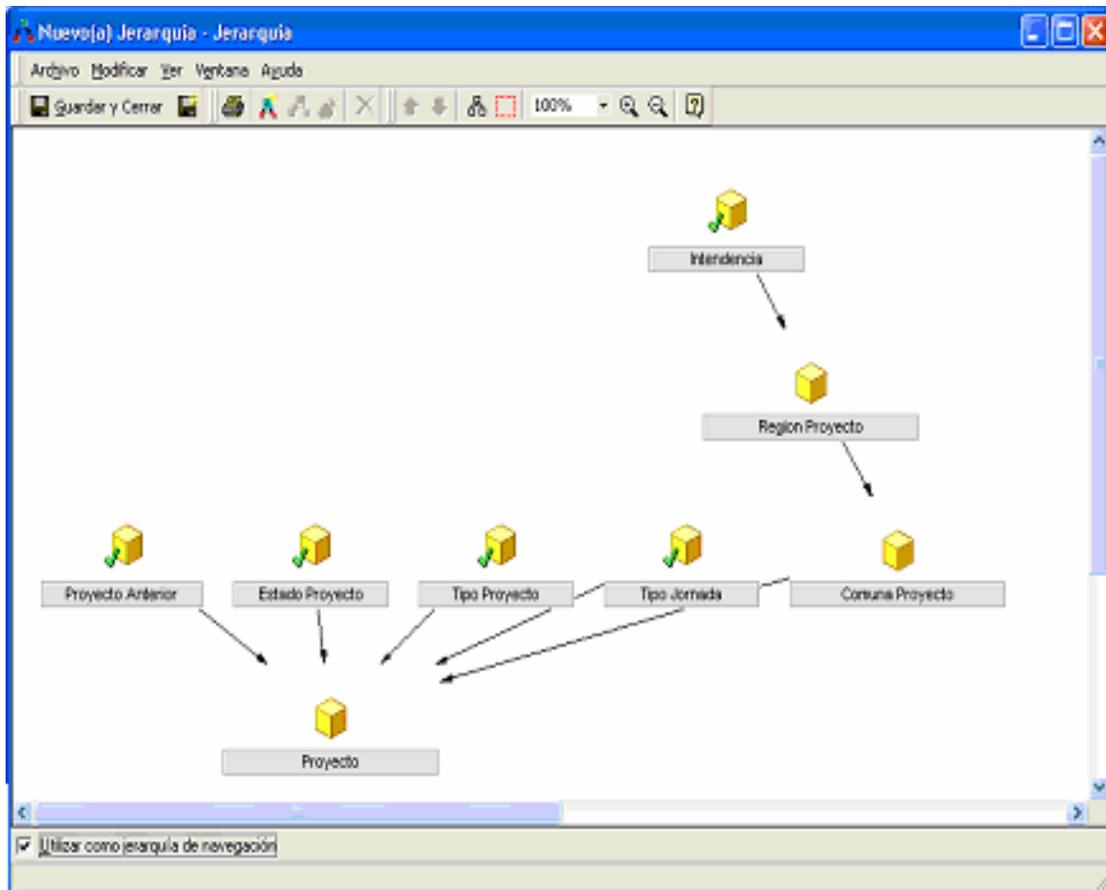


Figura 33, Jerarquía proyecto y su navegación.

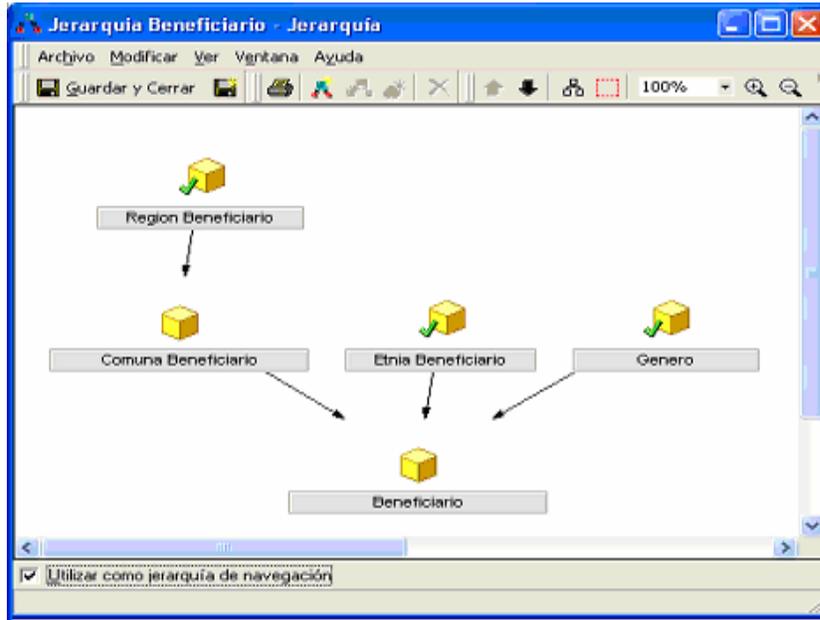


Figura 34, Jerarquía Beneficiario y su navegación.

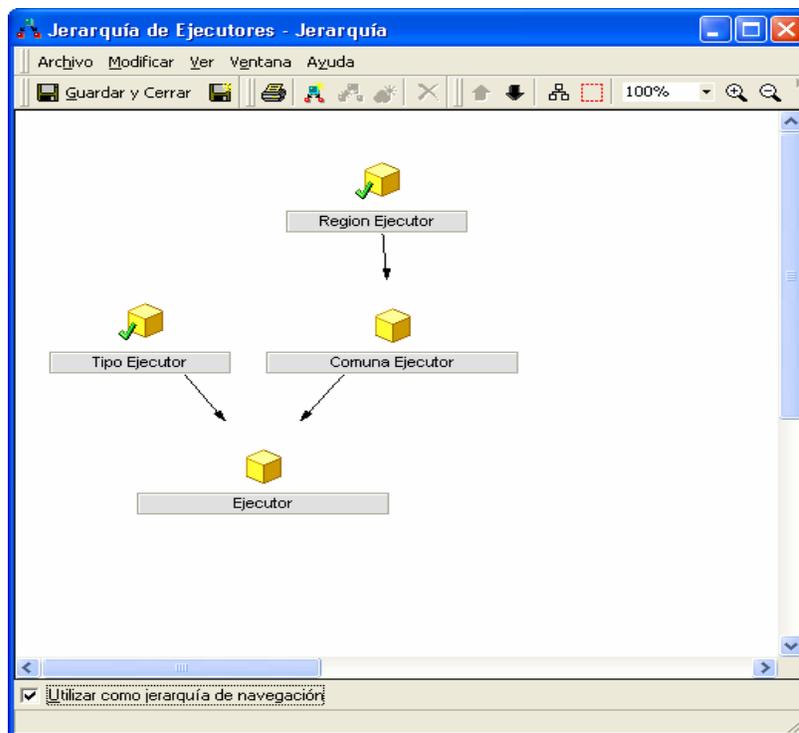


Figura 35, Jerarquía Ejecutor y su navegación.

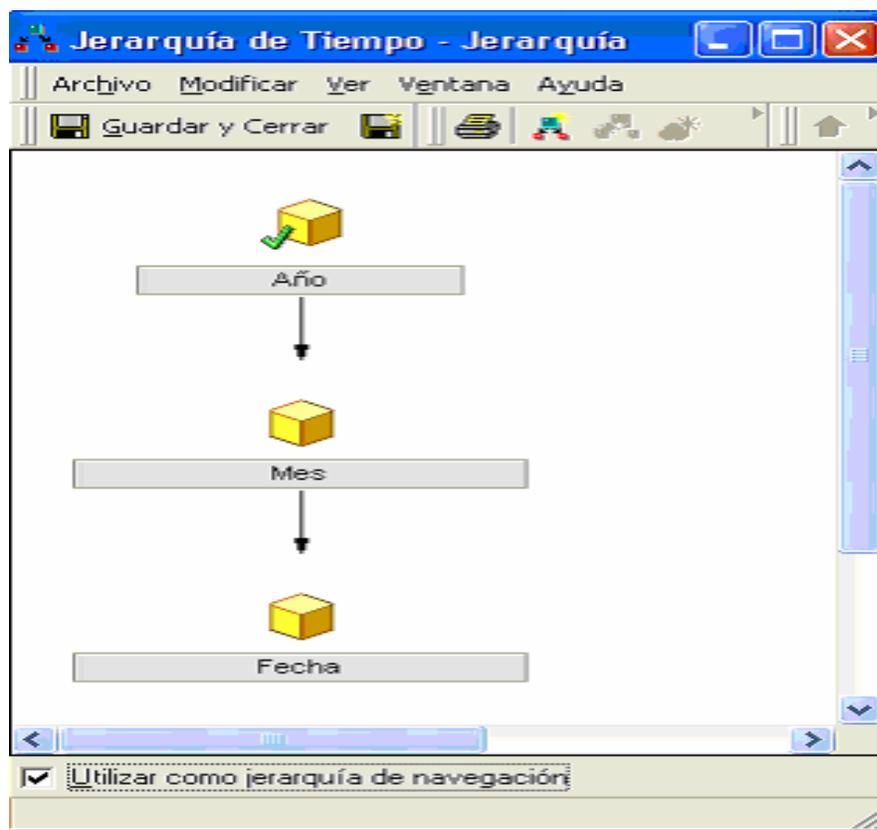


Figura 36, Jerarquía Tiempo y su navegación.

4.1.10 Liberación:

La liberación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesibles desde el escritorio del usuario del negocio. Hay varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de feedback. Todas estas tareas se deben tener en cuenta antes de que cualquier usuario pueda acceder al Data Warehouse.

Productos de esta etapa:

Producto final y Procesos en Marcha.

4.1.11 Mantenimiento y Mejoras:

Como se remarca en esta metodología (Ciclo de vida dimensional del negocio), Data Warehouse es un proceso de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral, pues acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los cambios de forma constante para poder seguir la evolución de las metas por conseguir.

Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito y no de falla. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

Una vez que se ha construido e implantado el Data Warehouse no hay tiempo para el descanso, rápidamente debemos estar preparados para administrar el mantenimiento y crecimiento del mismo.

4.2. Modelo Entidad Relación para Proempleo.

Las Bases de Datos Relacionales son las que se utilizan para alojar físicamente lo señalado por los Diagramas de Entidad Relación (ERD).

La Base de Dato Relacional de Proempleo y que soporta el Proceso Transaccional en Línea (OLTP) de Proempleo es SQLSERVER 2000. Montada sobre un servidor Dell, con procesador Pentium Xeon Biprocesado 3.0 Ghz con una RAM de 2 GB ECC. Arreglo de disco con 256 MB RAM Cache. 4 Discos 36 GB en Raid 5 de 10.000 rpm.

Es esta base de datos, la que nos proveerá de los datos necesarios para construir el Data Mart para Proempleo. La figura 37, muestra el modelo entidad relación de la base de datos.

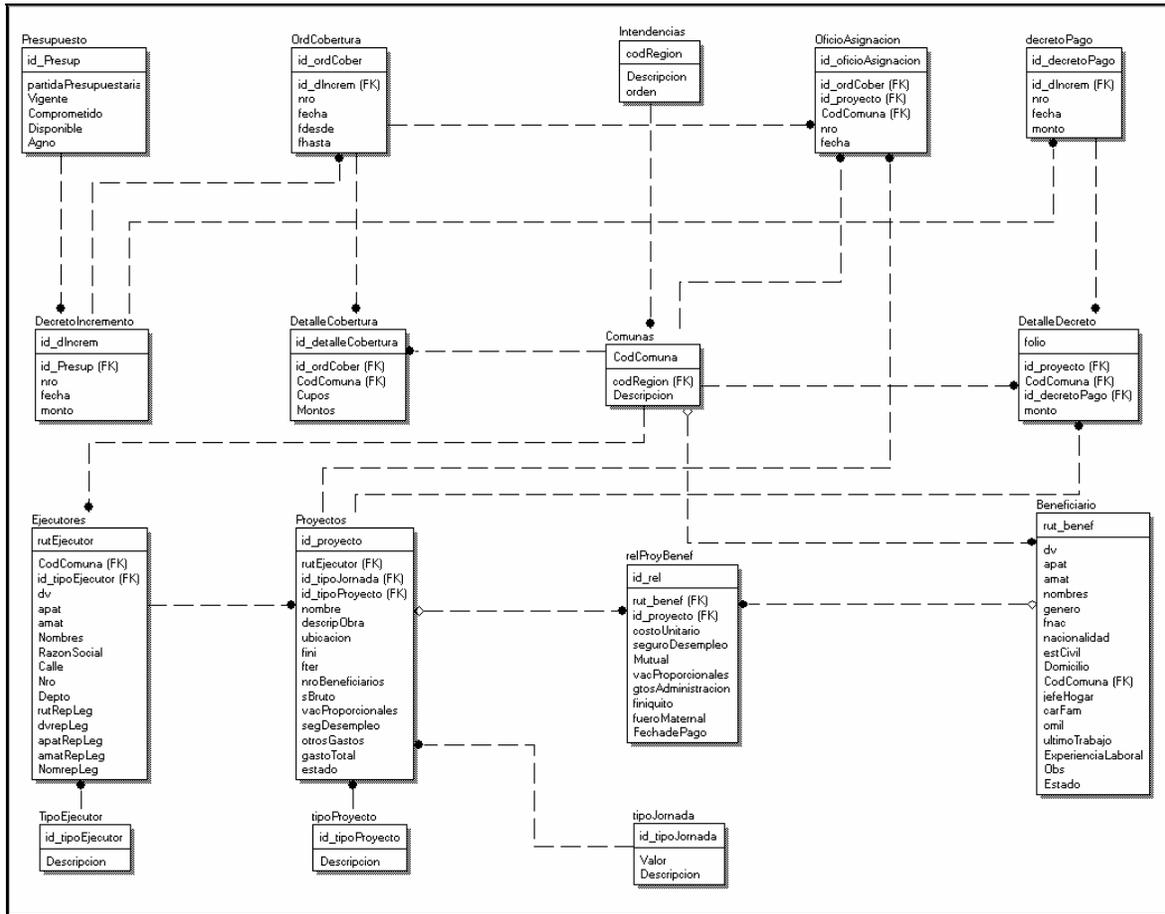


Figura 37, Modelo Entidad Relación del OLTP de Proempleo.

Este modelo es funcional y no está orientado al análisis de datos tal como lo hemos mencionado en este trabajo y es una característica de los modelos transaccionales (OLTP).

Los datos de esta base de datos deben ser pasados por un proceso de extracción, transformación y carga (Que está dado por la metodología “Data Staging ETL”).

Este proceso se conoce como ETL (del inglés Extraction, Transformation and Loading) y es el que nos va a permitir obtener como resultado la base de datos multidimensional, Data Mart.

Para la implementación del Data Mart, se definió que las áreas de desarrollo de los sistemas operacionales de Proempleo, son los que nos proveerán de los datos en la estructura del modelo físico OLAP (**Esquema**) desarrollado, para poder ser visualizado en el modelo multidimensional (Data Mart).

Mensualmente las áreas de desarrollo deben vaciar y transformar los datos de los sistemas operacionales de Proempleo, a la estructura física planteada para el Data Mart.

4.3. Modelo Lógico Multidimensional para Proempleo.

Como fue planteado en el capítulo "Marco Teórico Referencial", el Modelo multidimensional (MDM), define una visión lógica de datos, desde una perspectiva de negocio, no responde a la estructura física del Data Warehouse o Data Mart, es sólo una representación de negocio.

En las figuras 38, 39, 40 y 41. Se muestran las jerarquías definidas para el Modelo Lógico Dimensional desarrollado para Proempleo:

- **Jerarquía Beneficiarios:** refleja la relación de los datos del beneficiario con un lugar geográfico de su domicilio. El beneficiario pertenece a una comuna la que a su vez pertenece a una Región/Intendencia.

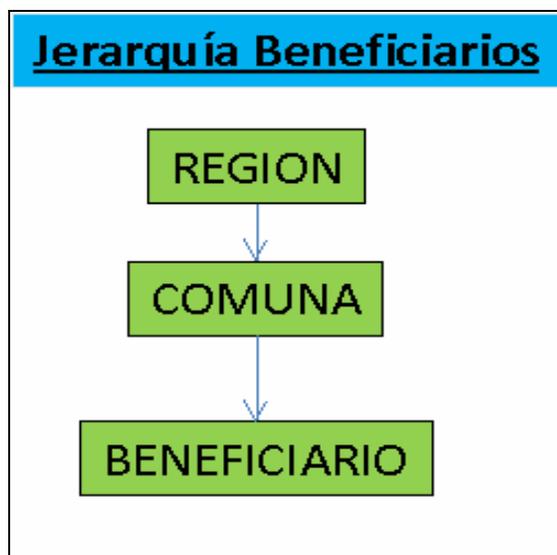


Figura 38, Representación de Jerarquía Beneficiarios.

- **Jerarquía Proyectos:** refleja la relación de los datos del Proyecto con el tipo de proyecto que se ejecuta, con la jornada asociada a él y con un lugar geográfico donde se desarrolla. El Proyecto pertenece a una comuna la que a su vez pertenece a una Región/Intendencia.

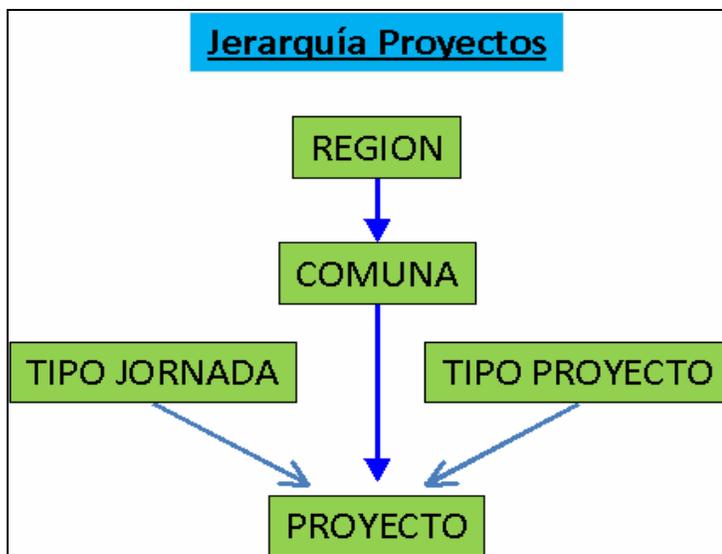


Figura 39, Representación de Jerarquía Proyectos.

- **Jerarquía Ejecutores:** refleja la relación de los datos del Ejecutor de un proyecto con el tipo de ejecutor catalogado y con un lugar geográfico de residencia. El Ejecutor pertenece a una comuna la que a su vez pertenece a una Región/Intendencia.

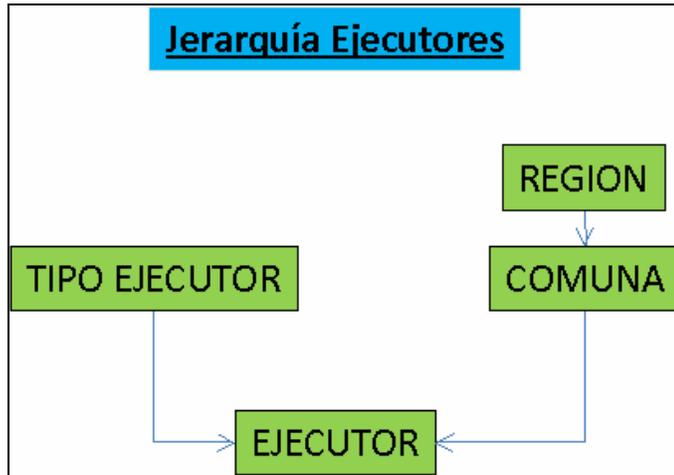


Figura 40, Representación de Jerarquía de Ejecutores.

- **Jerarquía Tiempo:** representa la forma en que se pueden agrupar los datos dentro de una dimensión temporal; se puede querer realizar una consulta agrupada por año, o por mes o día.

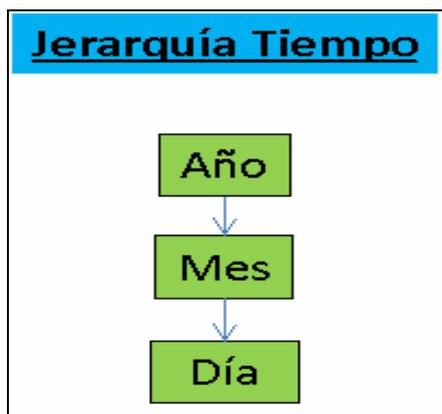


Figura 41, Representación de Jerarquía de Tiempo.

También se dijo en el capítulo "Marco Teórico" de este trabajo, que el modelado dimensional se basa en "hechos" que se conocen como Fact y que una de sus

principales ventajas es que se encuentra enfocado en el negocio y sus actividades.

En las figuras 42 y 43, se muestran las Facts o *hechos* definidos para el Modelo Lógico Dimensional de este trabajo:

- Fact de Transferencia:

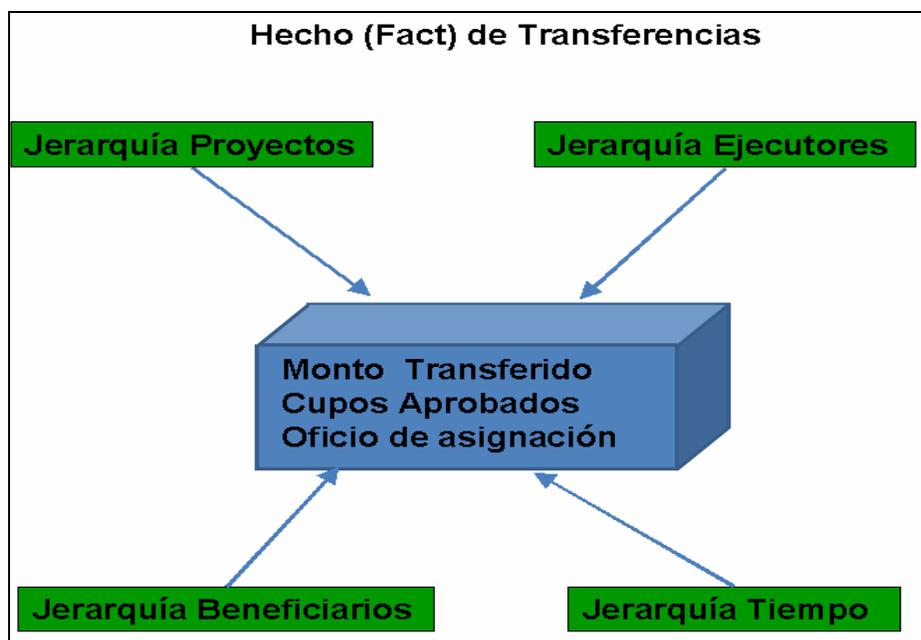


Figura 42, Fact de Transferencias.

Esta Fact, en el contexto de sus dimensiones, está definida para responder preguntas de negocio como por ejemplo; ¿Cuanto **monto** se ha **transferido** a los *Proyectos* de un *Ejecutor* que dentro de sus beneficiarios, más del 50% sean mujeres en el Mes de Marzo?

También está definida para responder a preguntas sobre los cupos aprobados por proyecto y por qué Oficio de Asignación.

Fact : Rendiciones

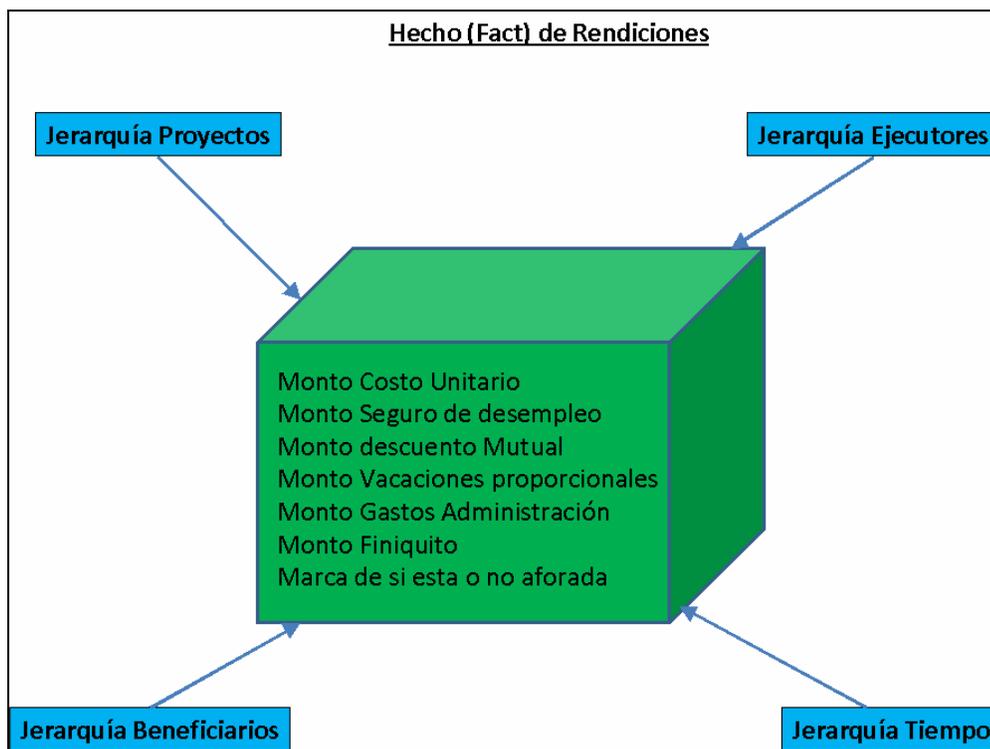


Figura 43, Fact de Rendiciones.

Esta Fact, en el contexto de sus dimensiones, está definida para responder preguntas de negocio como por ejemplo: ¿Qué *Proyectos* y de qué *Ejecutores* han finiquitado dentro de los *Beneficiarios* a más hombres que Mujeres dentro del *Mes* de Diciembre?

Este es solo un ejemplo de las distintas combinaciones posibles de consultas, ya que éstas son dinámicas y se van creando de acuerdo a las necesidades de información del momento.

4.4. Esquema para Proempleo

La figura 44, corresponde al **esquema** y es una representación física del Data Mart desarrollado para Proempleo, se puede apreciar la construcción de tablas y columnas, las que en definitiva forman el modelo de datos Dimensional del Data Mart.

Los esquemas normalizados (3FN), se conocen como “Esquema Copo de Nieve”, los esquemas altamente desnormalizados se conocen como “Esquema tipo estrella”.

Este esquema se forma a partir de los sistemas operacionales (OLTP) de Proempleo.

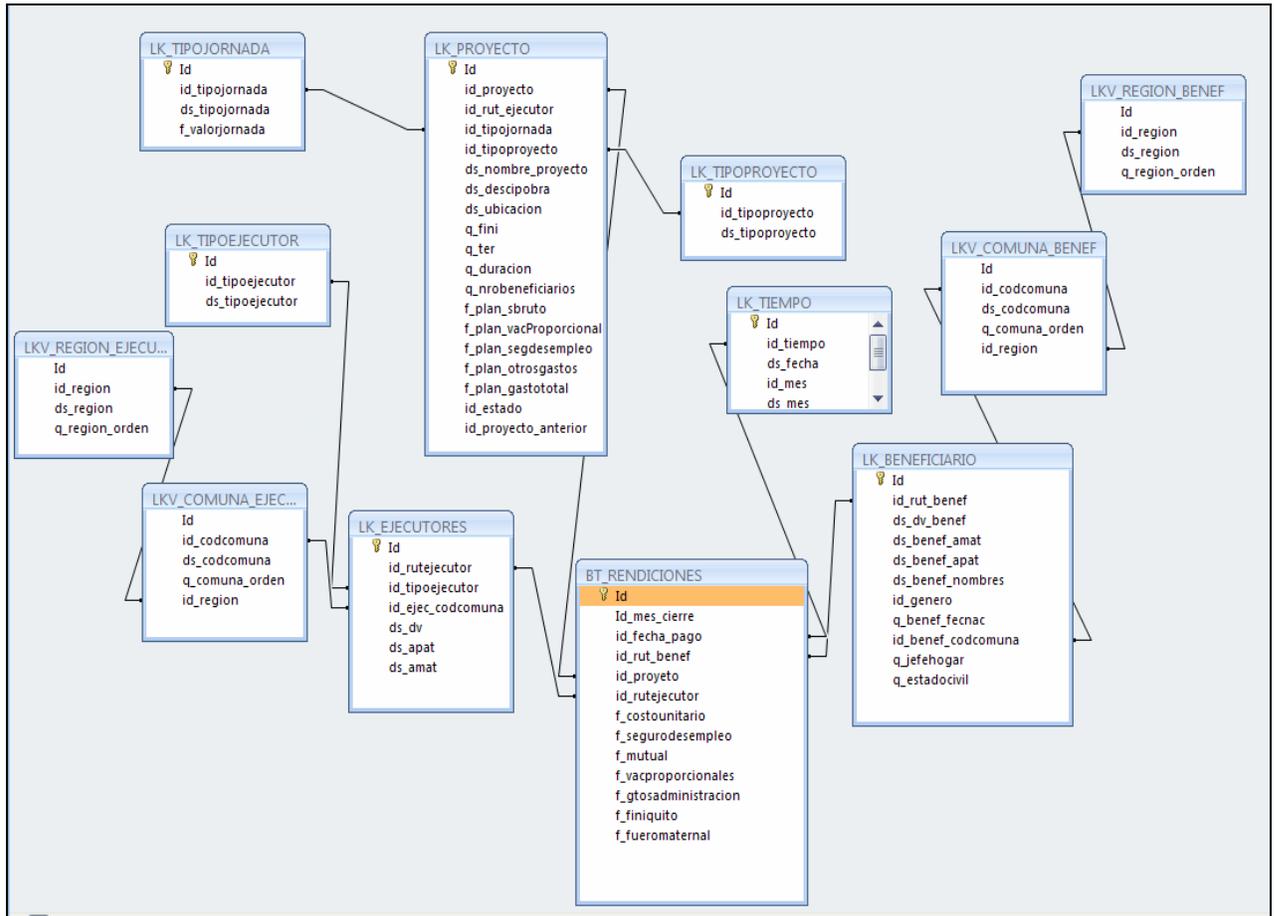


Figura 44, Modelo físico multidimensional (OLAP).

5. RESULTADOS

Este trabajo de tesis tuvo muchos resultados tangibles como intangibles dentro del Ministerio del Trabajo y Previsión Social, específicamente en la Unidad de Proempleo.

Los resultados intangibles se pueden resumir en la colaboración y disposición de los diferentes actores que se involucraron dentro de este proyecto de tesis, disposición toda, orientada al éxito de este trabajo y compromiso por parte de ellos en realizar las tareas tendientes a identificar las necesidades de información de negocio.

También como resultado positivo, es la disposición de la Autoridad a autorizar los viajes necesarios dentro del país con el fin de visitar las Intendencias, Comunas y Ejecutores relacionados con Proempleo, para recabar en terreno las necesidades de información a nivel operativo y poder levantar los requerimientos necesarios, para un buen desarrollo de este trabajo.

Los resultados tangibles más relevantes de este proceso de desarrollo son los siguientes:

- Se obtuvo toda la información pertinente para el desarrollo del Data Mart de Proempleo. Los requerimientos de información identificados entregaron la base sobre la cual se diseñó y modeló el Data Mart.
- Se logró que los administradores del sistema OLTP de Proempleo desarrollaran los programas y procedimientos almacenados necesarios para la extracción, transformación y carga de los datos fuentes al Data Mart (ETL).

- EL desarrollo del Modelo Lógico Multidimensional obedece a las especificaciones o requerimientos de las Jefaturas de Proempleo como de las Autoridades del Ministerio.
- Desarrollo de Esquema de Data Mart.
- Implementación física del Data Mart, es decir el paso del Esquema a Tablas físicas del Data Mart.
- Reportes de ejemplo de operaciones OLAP.

5.1. Problemas detectados

Durante el desarrollo de este trabajo, principalmente en los viajes a terreno que se realizaron, para el levantamiento de requerimientos, se detectó que no todos los usuarios de los sistemas operacionales de Proempleo saben realmente utilizarlos. Hay problemas de interpretación sobre algunas normativas que inciden directamente sobre la contratación del beneficiario.

Se da el caso, por ejemplo, que en Arica se crea un proyecto especial para las mujeres con fuero maternal y al proyecto no lo tipifican como “Proyecto con Fuero Maternal”. Luego el dato que se traspasa al Data Mart, no considera al proyecto en cuestión como un proyecto con fuero en Arica.

Lo anterior es muy significativo por cuanto puede hacer que el Data Mart no sea confiable ya que no maneja la información real, de la fuente operacional. Es necesario potenciar a los usuarios de los sistemas operacionales para que el Data Mart sea la herramienta de éxito que debe ser.

Es por lo anterior que se sugiere a las autoridades de Ministerio del Trabajo y en particular a las Jefaturas de Proempleo, se haga un esfuerzo para generar un programa de capacitación a la totalidad de usuarios de los sistemas operacionales de Proempleo, a lo largo del país, el que contemple difundir la necesidad e importancia de la recolección de información que les compete recabar día a día, la importancia que dicha información tiene para la alimentación del Data Mart y la importancia que tienen ellos como partícipes anónimos de tan relevante tarea.

Otro problema detectado, es la falta de experiencia del personal de Proempleo y del Ministerio en general, tanto en el área técnica como el de autoridades, en el uso de las tecnologías del Data Warehouse. Ello obliga a que en la implementación del proyecto se considere la realización de cursos de capacitación

dirigidos a orientar y operar con la nueva tecnología (Data Warehouse), a los actores relevantes.

5.2. Conclusión

La implementación de la tecnología de Data Warehouse en Proempleo, traducido en la construcción de un prototipo de Data Mart, cuyo propósito está orientado a agilizar la toma de decisiones y ayudar en el control y gestión de los programas de empleo de emergencia en el país, es sin duda un avance importante para las políticas que persiguen dar una mejor atención a los beneficiarios directos de los empleos.

Con la operación del Data Mart (construido como prototipo), en forma interactiva e intuitiva por los usuarios, Proempleo pudo obtener en forma práctica la información relevante para una mejor gestión de sus procesos, información impensable de obtener antes de este proyecto.

Con esta herramienta se mejora ostensiblemente el tan anhelado control y gestión de los procesos más importantes del negocio de Proempleo, los cuales son: transferencia de fondos y rendiciones de cuentas. Procesos muy delicados y complicados de manejar sin una herramienta de Data Warehouse, como la implementada.

Gracias al esfuerzo realizado en esta investigación, y que derivó en la construcción de un prototipo funcional de un Data Mart para Proempleo, se logró:

- Crear la necesidad real dentro del Ministerio del Trabajo, de implementar tecnologías Data Warehouse en la unidad de Proempleo.
- Generar expectativas de control y gestión, en las jefaturas, gracias a la información proporcionada por el prototipo Data Mart construido.

- Claridad en los usuarios, en cuanto a los requerimientos básicos de información que debe soportar un Data Mart para Proempleo.
- Crear experiencia en los actores relevantes, que participaron en este proyecto de tesis, para la construcción de un Data Mart.

Se ha adquirido la experiencia necesaria, por parte del memorista, para liderar en el futuro, la implementación del proyecto “*Gestión de los Programas de Empleo de Emergencia en Chile, utilizando Data Warehouse*” y se espera que esta metodología se pueda implementar pronto en el Ministerio del Trabajo, para una mejor gestión de los procesos en Proempleo.

6. BIBLIOGRAFÍA

The Data Warehouse, Lifecycle Toolkit. Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite

<http://www.alegsa.com.ar/Dic/sistematransaccional.php>

http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx

http://www.empresas.telefonica.es/catalogoTEE/aplicacionesnegocio/aplic_negoc/serv_gestion/data_warehouse/index.html

http://www.adictosaltrabajo.com/tutoriales/Datawarehouse/Datawarehouse_html_meb63b34.png

<http://blogs.baquia.com/todobi/post/2006/04/06/aque-es-olap->

<http://www.estadistico.com/arts.html?20010226#subcap10>

<http://www.inegi.gob.mx/inegi/contenidos/espanol/prensa/Contenidos/Articulos/tecnologia/relacional.pdf>

http://www.bitam.com/bitamweb/solutions_business_intelligence.php

<http://www.gestiopolis.com/delta/term/TER250.html>

<http://blogs.baquia.com/todobi/post/2006/06/14/inteligencia-negocio-open-source>

http://www.programacion.com/bbdd/articulo/man_cubeyrollup/#pre8

7. GLOSARIO DE TÉRMINOS

Análisis: Operación consistente en estudiar y descomponer los datos a fin de extraer de ellos los elementos esenciales y obtener un esquema de conjunto.

Análisis multidimensional: Análisis simultáneo de múltiples dimensiones de datos.

Atributo: Elemento de información, que describe una característica de una entidad.

Base de datos: Conjunto de datos no redundantes, almacenados en un soporte informático, organizado de forma independiente de su utilización y accesible simultáneamente por usuarios y aplicaciones.

Base de datos multidimensional: Una base de datos diseñada sobre un conjunto de dimensiones; se usa en el análisis multidimensional.

Consulta: Una petición formal y claramente especificada de información del Data Warehouse.

Data Mart: Subconjunto de datos de un Data Warehouse. También se entiende como una solución que, compartiendo tecnología con el Data Warehouse (pero con contenidos específicos, volumen de datos más limitado y un alcance histórico menor), permita dar soporte a una empresa pequeña, o un departamento o área de negocio de una empresa grande.

Data Warehouse: Es una combinación de conceptos y tecnología que cambian significativamente la manera en que es entregada la información a la gente de negocios. El objetivo principal es satisfacer los requerimientos de información

internos de la organización para una mejor gestión, con eficiencia y facilidad de acceso.

Dato: Unidad Básica de información.

Dimensión: Entidad independiente dentro del modelo multidimensional de una organización, son las variables sobre las que se hace la consulta

ETL (Extracción Transformación y Carga): Son los procesos por los que atraviesan los datos para ir desde el sistema OLTP al Data Mart/Data Warehouse.

La Extracción se refiere al mecanismo por medio del cual los datos son leídos desde la fuente original. Transformación; es la etapa por la que pueden atravesar los datos para estandarizarlos, normalizarlos. Carga; consiste en llevar los datos leídos y estandarizados al Data Mart/Data Warehouse.

Drill-Down (Taladrar): Es la habilidad para poder navegar de lo general a lo particular en la información presentada. Por ejemplo, en un informe de transferencias, se debe poder "taladrar" en los datos de cada región para obtener los datos por comunas.

Esquema tipo estrella: Un Esquema completamente desnormalizado, donde todos los niveles están en una misma tabla para toda la dimensión.

Fact: Es un dato numérico, que puede ser agregable para responder preguntas de negocio en un modelo multidimensional

Granularidad: Término que se usa en los Data Warehouse para expresar el nivel de detalle. A más alto nivel de granularidad más bajo nivel de detalle (mayor nivel de abstracción).

Hecho: Es un dato numérico, que puede ser agregable para responder preguntas de negocio en un modelo multidimensional.

Integridad: Condición de seguridad que garantiza que la información sea accedida y/o modificada, sólo por personas autorizadas.

Jerarquía: Formas de agrupar los elementos de una dimensión, reflejan las relaciones entre los datos del negocio.

Métrica: Es un indicador.

Metaproceso: Conjunto de procesos

Metadatos: Información que describe a un dato. En contexto de Data Warehouse, describe a un dato, precisando por ejemplo, su semántica, las reglas de gestión asociada, sus fuentes, etc.

Modelar Datos: Actividad de representar las categorías de datos y relaciones entre ellas como una abstracción en forma de diagrama.

Modelo Dimensional: El modelado dimensional se basa en HECHOS que se conocen como "Facts" y es una alternativa al modelado relacional.

OLAP: Siglas en inglés para "On-Line Analytical Processing", o Proceso analítico en línea es el nombre formal para el análisis de los Data Warehouse o Data Mart y que

por lo general se conocen como cubos multidimensionales. Son una forma más intuitiva de ver la información empresarial.

OLTP: Los sistemas transaccionales, son conocidos con esta sigla en inglés (**On Line Transaction Processing**), que significa; Procesamiento de Transacciones En Línea, estos sistemas facilitan y administran aplicaciones para entrada de datos y recuperación y procesamiento de transacciones, enfocados a gestionar un gran número de transacciones concurrentes. Permiten insertar, actualizar, borrar y consultar una pequeña cantidad de registros.

Redundancia: Repetición de los mismos datos en varios lugares.

Reporte: Documento que se produce en respuesta a una consulta.

Repositorio de datos: Lugar donde se almacenan los datos, también conocidos como bodega de datos categorizados o Data Warehouse.

Roll-Up: Es la operación inversa de Drill-down. Permite la navegación de lo particular a lo general.

SGBDR: Acrónimo de Sistema de Gestión de Base de Datos Relacional. Son un conjunto de aplicaciones que permites la creación, búsqueda y organización de los datos.

Semántica: Se refiere a los aspectos del significado o interpretación del significado de una determinada palabra, lenguaje o representación formal

Sumarización: Actividad de incremento de la granularidad de la información en una base de datos, reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.