



Aplicación de las tecnologías de Inteligencia Empresarial

**Contexto: Caja de Seguridad Social para los
Profesionales del Arte de Curar (CPAC)**

Ing. Carolina Verónica Díaz

caro.vero.diaz@gmail.com

**Universidad Tecnológica Nacional (UTN)
Especialización en Ingeniería de Sistemas de Información
Santa Fe – ARGENTINA**

Resumen

*“Gestionar la información en las empresas es, hoy en día, una herramienta clave para poder sobrevivir en un mercado cambiante, dinámico y global. Aprender a competir con esta información es fundamental para la toma de decisiones, el crecimiento y la gestión de nuestra empresa. La disciplina denominada como **Business Intelligence** nos acerca a los sistemas de información que nos ayudan a la toma de decisiones en nuestra organización.” [5].*

La creciente internacionalización de los mercados, y la consiguiente intensificación de la dinámica competitiva, se convierte en un verdadero reto de gestión. La capacidad para tomar decisiones con rapidez, basadas en un adecuado conocimiento de la realidad de la empresa así como del mercado y sus tendencias, ha pasado a convertirse en una nueva fuente de ventaja competitiva.

La C.P.A.C. (Caja de Seguridad Social para los Profesionales del Arte de Curar de la Provincia de Santa Fe) sin estar ajena a esta situación dispone de sistemas de información más o menos sofisticados y convenientes de analizar y optimizar. En este trabajo se investiga la forma en que esta disciplina a través de sus herramientas y tecnologías permite conocer el estado real de la institución, obtener información en tiempos oportunos y solventar nuevos requerimientos a corto y largo plazo. El objetivo es introducirla en el enfoque de Inteligencia Empresarial o Business Intelligence (BI) y presentarle los procesos, las herramientas y las tecnologías que le son propias (el qué); mostrarle las oportunidades de creación de valor que puede aportarle (el porqué); y proporcionarle una guía para su implementación (el cómo).

Tabla de Contenidos

CAPITULO I.....	5
1. Introducción y Preguntas de Investigación	5
1.1 Introducción	5
1.2 Importancia de la Investigación	6
1.3 Contextualización	6
1.4 Objetivos	8
1.5 Organización del Trabajo	8
CAPITULO II.....	9
2. Inteligencia de Negocios y Data Warehouse	9
2.1 Características Generales de la Inteligencia de Negocio	9
2.1.1 Introducción al Business Intelligence	9
2.1.2 Definición.....	9
2.1.3 Diferencia entre datos, información y conocimiento	10
2.1.4 Beneficios.....	11
2.1.5 Componentes de la Inteligencia de Negocios	11
2.2 Data Warehousing y Data Warehouse	12
2.2.1 Definición.....	12
2.2.2 DataMart.....	14
2.2.3 Immon vs Kimball	15
2.2.4 Arquitectura del Data Warehousing.....	16
2.2.4.1 OLTP.....	17
2.2.4.2 Load Manager	17
2.2.4.3 Data Warehouse Manager	18
2.2.4.4 Query Manager	20
2.2.4.5 Herramientas de Consulta y Análisis	21
2.2.4.6 Usuarios.....	22
CAPITULO III.....	24
3. Desarrollo de la Metodología	24
3.1 Metodología de la Investigación	24
3.2 Guía Metodológica de BI	24
3.2.1 Análisis de Requerimientos	25
3.2.1.1 Aplicación Etapa 1 al Caso de Estudio	27
3.2.2 Análisis de los OLTP	28
3.2.2.1 Aplicación Etapa 2 al Caso de Estudio	28
3.2.3 Modelo Lógico del Data Warehouse	37
3.2.4 Integración de Datos	42
3.3 Herramienta de Desarrollo	46
3.3.1 Características Principales.....	46
3.3.2 Arquitectura del Analysis Services	47
3.3.3 Herramientas del Analysis Services	47
3.3.4 Objetos Fundamentales del Analysis Services.....	47
3.3.5 Nuevas Características del Analysis Services en SQL Server 2008	48
3.4 Análisis de Resultados	48
CAPITULO IV	54
4. Evaluación de Resultados	54
4.1 Análisis de las variables y sus componentes	54
4.1.1 Calidad de la Información o Calidad de los Datos.....	55
4.1.2 Calidad del Sistema	55
4.1.3 Calidad del Servicio.....	56
4.1.4 Uso del Sistema	56
4.1.5 Satisfacción del Usuario	56
Conclusiones y Trabajos Futuros	57
Referencias Bibliográficas	58

Índice de Figuras

Figura 2.1: Concepto de BI	10
Figura 2.2: Principales Componentes de la Inteligencia de Negocios	12
Figura 2.3: Data Warehouse, variante en el tiempo	14
Figura 2.4: DW, no volátil.....	14
Figura 2.5: Top-Down	15
Figura 2.6: Bottom-Up.....	15
Figura 2.7: Arquitectura del DWH	17
Figura 2.8: Cubo Multidimensional.....	19
Figura 2.9: Esquema en Estrella.....	19
Figura 2.10: Esquema Copo de Nieve	20
Figura 2.11: Esquema Constelación	20
Figura 2.12: Toma de decisiones de acuerdo a un tipo de usuario.....	23
Figura 3.1: Ejemplo de Modelado Conceptual	28
Figura 3.2: Duplicidad de Información e inconsistencia en el origen de datos.....	30
Figura 3.3: Características de la base de datos operacional cp_prod	30
Figura 3.4: Tablas de Estado Previsión	31
Figura 3.5: Tablas que indican ubicación.....	32
Figura 3.6 Tablas de profesiones y especialidad.....	34
Figura 3.7: Tabla del estado de los afiliados en OS y Categoría en Previsión	35
Figura 3.8: Tabla para contratos.....	36
Figura 3.9: Tabla para deuda y cobranza	37
Figura 3.10: Cuatro pasos para el modelado dimensional.....	38
Figura 3.11: Modelado Dimensional. Fact Table: Afiliados_Titulares	39
Figura 3.12: Modelado Dimensional. Fact Table: Afiliados_Obra_Social	40
Figura 3.13: Modelado Dimensional. Fact Table: Afiliados_Beneficiarios.....	40
Figura 3.14: Modelado Dimensional. Fact Table: Beneficios	41
Figura 3.15: Modelado Dimensional. Fact Table: Deudores	41
Figura 3.16: Modelado Dimensional. Fact Table: Contrato	42
Figura 3.17: Modelado Dimensional. Fact Table: Proporción	42
Figura 3.18: Carga Inicial.....	43
Figura 3.19: Versión Microsoft SQL Server Management Studio.....	48
Figura 3.20: Page	52

Índice de Cuadros

Cuadro 2.1: Utilización de la Inteligencia de Negocios por parte de las empresas.....	12
Cuadro 2.2: Diferencia datos operacionales e informativos	13
Cuadro 2.3: Comparación de las Características Esenciales de los Modelos de Inmon y Kimball	16
Cuadro 2.4: Análisis de los distintos tipos de usuarios	22
Cuadro 3.1: Tabla Actividad.....	31
Cuadro 3.2: Tabla Grupos_Actividad	32
Cuadro 3.3: Tabla Leyendas Estado Previsión	32
Cuadro 3.4: Tabla Localidades	33
Cuadro 3.5: Tabla Departamentos y Distrito Electoral	33
Cuadro 3.6: Tabla Profesiones	34
Cuadro 3.7: Tabla Especialidades	35
Cuadro 3.8: Categorías de aportes y MPA	35
Cuadro 3.9: Categorías de aportes y MPA	35
Cuadro 3.10: Resultados antes de aplicar Drill-down	49
Cuadro 3.11: Resultados después de aplicar Drill-down	49
Cuadro 3.12: Resultados después de aplicar Drill-Across	50
Cuadro 3.13: Resultados antes de aplicar Pivot.....	51
Cuadro 3.14: Resultados después de aplicar Pivot	51
Cuadro 3.15: Resultados después de aplicar Page – Colegio Médicos.....	52
Cuadro 3.16: Resultados después de aplicar Page – Colegio Odontólogos.....	52
Cuadro 3.17: Resultados después de aplicar Drill Through.....	53
Cuadro 4.1: Principales Variables – Impacto de la Data Warehouse e Inteligencia de Negocios.....	55

CAPITULO I

1. Introducción y Preguntas de Investigación

1.1 Introducción

En la mayor parte de las organizaciones, la informática ha sido exitosamente aplicada para automatizar tareas repetitivas, rutinarias y bien definidas, y en general ha constituido una verdadera solución para aumentar la eficiencia de los procedimientos vinculados a transacciones como facturación, cobranzas, mantenimiento de stock, etc. Estos sistemas se los conoce como **Sistemas Transaccionales u Operacionales (OLTP: On Line Transaccional Processing)**.

Actualmente la mayoría de las organizaciones y empresas poseen y generan diariamente una enorme cantidad de datos imposibles de analizar a simple vista. La mayor parte de estos datos generados no aportan la información necesaria a la toma de decisiones empresarial, pues para poder usarlos es necesario que se transformen en conocimiento útil para quienes dispongan de ellos. Estos datos, llamados **datos operacionales**, son creados durante la ejecución de operaciones tradicionales de captura de datos y servicios básicos de tratamiento de datos. Son producidos por los sistemas OLTP y almacenados en archivos o base de datos¹. Frecuentemente contienen valores incorrectos, son muy detallados y son de uso mínimo en los negocios debido a su gran volumen, ubicación y formatos. Son específicos para cada aplicación y usualmente son almacenados de manera separada.

Cualquier organización al día de hoy necesita disponer de estrategias y herramientas de **inteligencia empresarial**, y en definitiva, de las tecnologías de la información para poder obtener la mayor cantidad de información útil en el menor tiempo posible a partir de todos los datos que se generan, y transformarlos de esta forma en un activo que preste beneficios y se pueda compartir, facilitando así la toma y la corrección de las decisiones del negocio. Además, esto supone un ahorro de tiempo y dinero en el análisis y el estudio de cualquiera de las actividades de la entidad, evitando de esta manera el costoso acceso a datos de diferentes procedencias o departamentos, la generación de informes a partir de complicadas herramientas o de forma manual, así como reducir el riesgo empresarial.

Por este motivo, el interés de este trabajo se centra en analizar de qué manera se puede aprovechar el potencial que ofrecen estas tecnologías de Inteligencia Empresarial en el contexto de la Caja de Seguridad Social para los Profesionales del Arte de Curar de la Provincia de Santa Fe (C.P.A.C). El objetivo final es en definitiva orientarla en la mejora de la explotación de la información que manejan en sus SI (Sistemas de Información), y con la finalidad de dotarla de una base de datos corporativa (Data Warehouse) que almacene no sólo datos sino también información que *aporte conocimiento y añada valor*, al tiempo que sirva de ayuda en la toma de decisiones prestando apoyo en una estrategia centrada en la mejora de la calidad del servicio que se presta a los afiliados profesionales apuntando a la *integración y accesibilidad* en los datos informativos.

Integración se refiere a identificar y utilizar la mejor fuente de datos realizando las transformaciones necesarias para construir los registros en el formato del Data Warehouse y crear un archivo de salida conteniendo todos los registros nuevos a ser cargados en el mismo. El objetivo de un Data Warehouse es consolidar información proveniente de diferentes bases de datos operacionales y hacerla disponible para la realización de análisis de datos de tipo gerencial. Los datos del mismo son el resultado de transformaciones, chequeos de control de calidad e integración de los datos operacionales. Se incluyen también totalizaciones y datos pre-calculados en base a datos operacionales.

Accesibilidad de los datos, estableciéndolos como la fuente principal de este concepto. Lo primero que debe garantizar este tipo de herramientas y técnicas es el acceso de los usuarios a los datos con independencia de la procedencia de estos. Su prioridad es el acceso interactivo e inmediato a información estratégica de un área de negocio. Las operaciones preponderantes no son las transacciones, como en las bases de datos operacionales, sino consultas que involucran gran cantidad de datos y agrupaciones de los mismos.

¹ Una **base de datos** es un conjunto de datos persistentes que es utilizada por los sistemas de aplicación de alguna empresa dada. Una base de datos se diseña, construye y rellena con datos para un propósito específico. Dispone de un grupo pretendido de usuarios y algunas aplicaciones preconcebidas en las que esos usuarios están interesados.

Ha motivado la preparación de este trabajo el hecho de que este tema no ha sido abordado con la intensidad ni con la profundidad que amerita en la C.P.A.C. El sector gerencial de esta organización se encuentra muy desprotegido y en un estado de alta vulnerabilidad al interactuar en un cambiante mercado del conocimiento y de información. Se dispone de un esquema de base de datos, que no facilita la recopilación de información a presentarse a nivel gerencial, razón por la cual surge la importancia de la presente investigación de modo que permita establecer un esquema de base de datos que apoye al actual, donde la información sea accesible en cuanto a *rapidez* y *eficacia*, para que de esta manera favorezca el desarrollo de nuevos reportes.

1.2 Importancia de la Investigación

Las compañías de la actualidad son juzgadas no únicamente por la calidad de sus productos o servicios, sino también por el grado en que comparten la información con sus clientes, empleados y socios. Sin embargo, la mayoría de las organizaciones, como es el caso de la C.P.A.C, tienen abundancia de datos, pero una penuria de conocimiento. Es por ello que en este trabajo se determina una metodología de Inteligencia Empresarial para englobar todos los sistemas de información y así obtener de ellos no solamente datos, sino una verdadera inteligencia que le confiera a la organización una ventaja competitiva garantizando aspectos como accesibilidad, integración tecnológica y toma de decisiones.

Accesibilidad para que los usuarios tengan la posibilidad de consultar la información y manipular diferentes variables. *Integración Tecnológica* de modo que el usuario final pueda consultar diferentes fuentes de datos sobre una interfaz que le permita manipular las variables de consulta para elaborar reportes e indicadores. *Toma de decisión* brindando la posibilidad al usuario final de visualizar mediante herramientas gráficas indicadores que ayudan en la toma de decisiones.

En definitiva es indudable la necesidad de instaurar una solución de BI que se adapte a los requerimientos de la organización basada en el análisis de las fuentes y que proporcione a los usuarios del sistema una manera universal de acceder, ver y utilizar la información. Podrá permitir a los distintos sectores y departamentos de la institución obtener resultados benéficos como: generar reportes globales o por secciones, conocer la realidad de los aportantes a la previsión y obra social, crear escenarios con respecto a una decisión, compartir información entre los departamentos, análisis multidimensionales, generar y procesar datos, en definitiva, mejorar el servicio al afiliado.

1.3 Contextualización

La Caja de Seguridad Social para los Profesionales del Arte de Curar de la Provincia de Santa Fe (C.P.A.C) se fundó en el año 1958. Nuclea a todos los profesionales de los sectores relacionados a la salud "*Arte de Curar*". Es una organización que brinda a sus afiliados y a su grupo familiar la capacidad de tener una obra social (*O.S.P.A.C: Obra Social para los Profesionales del Arte de Curar*) y realizar sus aportes jubilatorios para que en el futuro puedan percibir un beneficio jubilatorio. La afiliación a la Caja es automática y obligatoria y por lo tanto constituye una carga legal para todos los profesionales universitarios de disciplinas básicas o auxiliares del Arte de Curar que ejercen su profesión en forma autónoma a partir de su fecha de inscripción en la matrícula. La población de profesionales de la salud corresponde a distintos colegios: Médicos, Odontólogos, Terapistas Ocupacionales, Psicólogos, Obstetras, Técnicos Radiólogos, Nutricionistas, entre otros.

Es una persona jurídica pública no estatal, con autonomía económica y financiera y administrada por sus afiliados. Tiene por objeto asegurar, organizar, implementar y administrar un sistema de seguridad social, siendo sus objetivos primordiales los sistemas de previsión y de obra social, todos fundados en el *principio de solidaridad* complementado con el de equidad. El gobierno, conducción y administración está a cargo de un Directorio integrado por ocho (8) miembros titulares, con igual número de directores suplentes, todos egresados de los Colegios Profesionales del Arte de Curar.

Según la diferenciación que realiza Eroles [10] sobre los distintos tipos de empresas, la C.P.A.C se enmarca en la clasificación de la mediana empresa, ya que cuenta con más de 100 empleados en la Casa Central Santa Fe y alrededor de 80 trabajadores más en la delegación Rosario.

En año 1998, la organización adquirió un sistema transaccional hecho a medida por una consultora de la ciudad de Rosario que, hasta la actualidad permite dar soporte a todas las tareas diarias. Es un sistema informático que a lo largo del tiempo ha tenido que ir adaptándose y ha sufrido algunas migraciones de datos para no quedar en desuso frente a los cambios tecnológicos. Hoy en día, brinda asistencia a distintos departamentos y sectores de la institución: Secretaría, Fiscalización y

Control de Aportes, Facturación de Obra Social, Bancos, Tesorería, Estampillado y Otros Recursos, Caja, etc.

Debido a la existencia limitada de reportería para control a nivel gerencial, no es posible contar con información oportuna, que permita apoyar y sustentar la toma de decisiones. Los requerimientos de información no se solventan de manera eficaz motivo por el cuál, surge la necesidad de contar con un nuevo esquema de información que apoye y preste las facilidades necesarias para alcanzar mejores resultados al momento de obtener información. Se puede denotar como un aspecto adicional, el problema presentado con la rotación de personal lo cual genera inconvenientes debido a desconocimientos de los desarrollos realizados a lo largo del tiempo.

Los sistemas transaccionales actuales de la C.P.A.C tienen las siguientes deficiencias:

- 1) **No permiten el análisis en forma eficiente de la información:** encierran los datos de la organización, pero sin embargo no se explotan debidamente y en toda su extensión. Hay graves inconvenientes al momento de obtener información que permita realizar la toma de decisiones en base a datos de calidad. Esto se debe a que las actividades de procesamiento de datos sumariados se realiza muchas veces con herramientas genéricas (como ser planillas de cálculo), que no han sido diseñadas con este propósito exclusivo. De esta manera el proceso de búsqueda y recolección de datos se vuelve muy complejo.
- 2) **Solo existen herramientas de consultas y reportes rígidos:** existen dificultades a la hora de realizar consultas complejas adaptadas a los tiempos de la institución. Hay una cuantiosa dispersión de los datos, lo que provoca lentitud en el proceso de consultas detalladas; como anexos, estructuras y reportes históricos.
- 3) **Baja integración de los datos:** ya que los datos operacionales están disgregados en distintos sistemas de la organización, desarrollados por diferentes personas y en diferentes momentos. Hay dispersión de plataformas e infraestructura (Sistemas Operativos, Bases de Datos, Lenguajes de Programación, etc) con cambios dinámicos de los datos.
- 4) **Existe una fuerte dependencia del área de sistemas para obtener cada informe gerencial.** Actualmente ante un requerimiento específico de información ad hoc, el personal de sistemas es requerido para realizar dicho trabajo.
- 5) **La explotación de la herramienta informática se ve limitada puesto que no se cuenta con manuales técnicos,** lo que dificulta que el personal de sistemas pueda realizar nuevas adecuaciones y/o mantenimientos.

Por otro lado, hoy en día, la C.P.A.C está atravesando un periodo de grandes cambios que han sido propiciados por causas de diferente índole; por un lado: 1) el crecimiento acelerado de los nuevos profesionales matriculados del Arte de Curar; 2) la situación económica actual que exige tener un gran conocimiento de la conformación de los grupos de afiliados (activos y pasivos) con el fin de promover nuevas políticas para el sostenimiento económico de la organización y 3) la necesidad de información cada vez más demandante requerida por los distintos colegios profesionales.

Frente a este contexto, surge la necesidad de contar con un sistema de ayuda a la toma de decisiones, que integre los datos de la organización y que permita su análisis. En los últimos años han surgido herramientas que permiten aprovechar el potencial encerrado en esos grandes cúmulos de datos. En efecto, las bases de datos transaccionales encierran la información detallada de los procesos ejecutados por la organización, los cuales debidamente actualizados permiten reconocer patrones y tendencias de comportamiento sumamente útiles para las decisiones gerenciales vinculadas al planeamiento y definición de estrategias.

Existen tecnologías que representan una oportunidad para poner los datos a disposición de la gerencia en esta organización y construir verdaderos sistemas de ayuda a la toma de decisiones. Estos sistemas contribuyen sin duda una herramienta compleja que responde a la estrategia de crecimiento de cada organización. Los datos almacenados representan un activo fundamental para la institución y como tales deben resguardarse bajo un sistema que los integre y que permita un acceso fácil y rápido a la información correcta.

1.4 Objetivos

Business Intelligence o Inteligencia Empresarial es una disciplina de reciente desarrollo, de firme y ascendente uso por diversas organizaciones. Está integrada por varias áreas de conocimiento, entre ellas métodos específicos de tratamiento y uso intensivo de datos. Su objetivo enunciado es “servir como soporte para la toma de decisiones”. En este trabajo, se investiga acerca de esta disciplina y sus tecnologías y se presenta el valor que puede aportar a la C.P.A.C, además de guiarla en la implementación, así como sus limitaciones. Es posible aprovechar todo este potencial integrando la información de diversas fuentes de datos y facilitando su análisis desde diversas perspectivas adquiriendo accesibilidad en los datos informativos. La idea central es que los datos de la institución dejen de ser solo simples datos, para convertirse en información que enriquezca las decisiones.

Objetivos Generales

- Proponer una solución de Business Intelligence en la C.P.A.C, establecer su guía metodológica de implementación, presentar un plan inicial de integración de datos, brindar una herramienta de desarrollo factible de utilizar y evaluar los resultados obtenidos con el objetivo final de orientarla en la mejora de la explotación de la información que manejan en sus Sistemas de Información.

Objetivos Específicos:

- Estudiar y analizar las diferentes técnicas y conceptos sobre Inteligencia Empresarial para aplicarla en el planteamiento y diseño de la solución.
- Instaurar una guía metodológica que se adapte a los requerimientos de la organización.
- Presentar un plan inicial de integración de datos.
- Proponer una herramienta de desarrollo para realizar consultas e informes adaptados a los tiempos de la institución.
- Estudiar cuales son las variables que se deben tener en cuenta en la evaluación de resultados.

1.5 Organización del trabajo

De manera general, este trabajo está organizado en las siguientes secciones: a continuación se presenta el capítulo dos donde se realiza una amplia revisión de literatura sobre la Inteligencia Empresarial y Data Warehouse. Luego, en el capítulo tres se desarrolla la guía metodología propuesta, se realiza un plan inicial de integración de datos, se presenta una herramienta para el desarrollo y se analizan los resultados obtenidos. Y en el último capítulo, el número cuatro, se evalúan las principales variables que intervienen en el momento de la evaluación de resultados. En la parte final se establecen las conclusiones, trabajos futuros y se presentan las referencias bibliográficas.

CAPITULO II

2. Inteligencia de Negocios y Data Warehouse

2.1 Características Generales de la Inteligencia de Negocios

2.1.1 Introducción al Business Intelligence

Business Intelligence es un término ideado por Gartner Group (*consultora internacional especializada en Tecnologías de Comunicación e Información*) en los años 80 para describir la capacidad de una organización para acceder y explotar la información residente en una base de datos, de manera que los usuarios puedan analizar esa información y adquirir en base a ella, teoría y conocimientos que apoyen la toma de decisiones del negocio.

Poseer un conocimiento proveniente de información comprensible, detallada, relevante y útil es vital para lograr y sostener una ventaja competitiva en el mundo empresarial. Para transformar los datos y convertirlos en información, y ésta a su vez, ser aprovechada como conocimiento, se necesitan distintas técnicas y procesos. A todos estos procesos de tratamiento de datos se les atribuye el término de **Business Intelligence** (BI, en adelante) o **Inteligencia Empresarial**. La combinación de los datos operacionales que han sido modificados, depurados, consolidados y organizados desde diversas fuentes se denominan **datos informativos**. Este tipo de información es específica para un conjunto de usuarios del negocio que lo hacen significativo y útil para su análisis. Son reorganizados por temas de negocio y se encuentran disponibles para su análisis durante periodos largos.

2.1.2 Definición

Business Intelligence se puede definir como un término de la administración que se refiere a aplicaciones y tecnologías que son usadas para obtener, lograr acceso y analizar datos e información sobre el funcionamiento de la empresa, los cuales pueden ayudar a obtener un conocimiento amplio de los factores que afectan su desempeño y de esa manera tomar mejores decisiones. Davenport y Harris [9] indican que la Inteligencia de Negocios consiste en la obtención, administración y reporte de los datos orientados a la toma de decisiones, y las técnicas analíticas y procesos computarizados que se usan para el análisis de la misma.

Jourdan, Rainer y Marshall [17], indican que la Inteligencia de Negocios es un **proceso** y un **producto**. El primero compuesto de métodos que las organizaciones usan para desarrollar información aplicable o inteligencia de negocios, que les permita a las organizaciones salir adelante en un mundo muy competitivo y globalizado. Como producto es información que les permitirá a las organizaciones predecir el comportamiento de competidores, proveedores, clientes, tecnologías, adquisiciones, mercados, productos y servicios y el comportamiento en general del ambiente de negocios, con un cierto grado de precisión.

Según Krizan [20], la inteligencia es más que obtener información, ya que va relacionada con un individuo específico que la necesita, y es un conocimiento ya que requiere del involucramiento de un ser humano. La recolección de información produce datos, y es la mente humana la que la convierte en inteligencia al adaptarla a un contexto específico para un individuo particular. El proceso que produce inteligencia es la colección continua con verificación y análisis de la información que permite comprender el problema o la situación de una manera accionable de acuerdo a un usuario final.

En definitiva, Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Existe una frase muy popular que dice: "la inteligencia de negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción para la toma de decisiones". La figura 2.1 representa de manera gráfica este concepto:



Figura 2.1: Concepto de BI.

La definición del glosario de términos de Gartner [12], especifica que *“BI es un **proceso interactivo** para **explorar** y **analizar información estructurada** sobre un **área** (normalmente almacenada en un Data Warehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de Business Intelligence incluye la **comunicación** de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores.”*

Al descomponer esta definición se obtiene:

Proceso iterativo: se trata de un análisis de información continuado en el tiempo, no solo en un momento puntual. Nos puede aportar un proceso continuado de análisis de información en el que por ejemplo podemos ver tendencias, cambios, variabilidades, etc.

Explorar: en todo proceso hay un momento inicial que por primera vez accedemos a la información que nos facilita su interpretación. En esta primera fase lo que hacemos es “explorar” para comprender qué sucede en nuestra organización, incluso se podrían descubrir nuevas relaciones.

Analizar: se pretende descubrir relaciones entre variables, tendencias, es decir, cuál puede ser la evolución de las variables, o patrones. Si un afiliado a la CPAC tiene una serie de características, cuál es la probabilidad que otro con similares características actué igual que el anterior.

Información estructurada y Data Warehouse: La información que utilizamos en BI está almacenada en tablas relacionadas entre ellas. Estas tablas están almacenadas en lo que conocemos como **Data Warehouse** o **almacén de datos** (más adelante definiremos con precisión este término).

Área de análisis: Todo proyecto de BI debe tener un objeto de análisis concreto. Nos podemos centrar en el comportamiento de los afiliados y estudiar quienes son, de dónde provienen, quiénes y cómo pagan y establecer indicadores que permitan conocer las tasas de afiliación/desafiliación, tasas de beneficios, desagregación de los ingresos y egresos por dimensiones, etc.

Comunicar los resultados y efectuar los cambios: Un objetivo fundamental del BI es que, una vez descubierto algo, sea comunicado a aquellas personas que tengan que realizar los cambios pertinentes en la organización para mejorarla.

2.1.3 Diferencia entre datos, información y conocimiento

Los tres términos podrían utilizarse indistintamente y esto puede llevar a una interpretación libre del concepto de conocimiento.

Los *datos* se definen como la mínima unidad semántica y se corresponden con los elementos primarios de información. Por sí solos son irrelevantes en el proceso de toma de decisiones. No dicen nada sobre el origen de las cosas y no son orientativos para la acción. Sin una utilidad o un contexto no sirven como base para apoyar la toma de decisión.

La *información* es el conjunto de datos procesados que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quien debe tomar decisiones al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor:

- Contextualizando: conociendo el contexto y el propósito para el que se creó.
- Categorizando: conociendo las unidades de medida que ayudan a interpretarlos.
- Calculando: los datos pueden haberse procesado de manera matemática.
- Corrigiendo: eliminando errores e inconsistencias.
- Condensando: resumiendo en forma concisa.

Finalmente, el *conocimiento* es una mezcla de experiencia, valores e información que sirve como marco para la incorporación de nuevas experiencias e información y es útil para la acción. El

conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como: comparación con otros elementos, predicción de consecuencias y búsqueda de conexiones.

Según Krizan [20], la inteligencia es más que obtener información, ya que va relacionada con un individuo específico que la necesita, y es un conocimiento ya que requiere del involucramiento de un ser humano. La recolección de información produce datos, y es la mente humana la que la convierte en inteligencia al adaptarla a un contexto específico para un individuo particular. El proceso que produce inteligencia es la colección continua con verificación y análisis de la información que permite comprender el problema o la situación de una manera accionable de acuerdo a un usuario final.

2.1.4 Beneficios

Entre los beneficios más importantes que BI proporciona a las organizaciones, se pueden destacar los siguientes:

- **Reduce el tiempo** mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- **Información potente y asequible:** Sin necesidad de conocimientos técnicos cualquier usuario puede acceder a información útil organizada e integrada. Se podrá consultar y analizar los datos de manera sencilla e intuitiva.
- **Automatiza la asimilación de la información,** debido a que la extracción y carga de los datos necesarios se realizará a través de procesos definidos.
- **Proporciona herramientas de análisis** para establecer comparaciones y tomar decisiones.
- **Permite a los usuarios no depender** de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- **Posibilitar la formulación y respuesta** de preguntas que son claves para el desempeño de la organización.
 - Se **podrán detectar** situaciones fuera de lo normal.
 - **Eliminación del error humano:** Todos los informes se realizan a través de la plataforma evitando la generación de costosos documentos de forma manual.
 - **Automatización de circulación de información:** Planificación automática de envío de informes dentro de una periodicidad a las personas necesarias, esto permite una mejora en la comunicación.
 - **Sistemas de estructura escalable:** Esto hace posible que los sistemas crezcan de forma regulada.
 - **Ahorro de tiempo y dinero:** Estas soluciones permiten que cada sujeto cumpla su función y dedique tiempo para tareas más importantes, por tanto permite un aumento de la productividad. El tiempo es dinero.

2.1.5 Componentes de la Inteligencia de Negocios

Cuando se habla de Inteligencia de Negocios se deben considerar los diferentes elementos que la constituyen, dentro de los cuales están: la base de datos centralizada (*Data Warehouse*), el componente principal para la transformación de datos (*Procesamiento de Extracción, Transformación y Carga - ETL*), las reglas y tablas especializadas (*Data Store*), la versión limitada del Data Warehouse establecido para un grupo de usuarios o un área (*DataMarts*), el conjunto de herramientas que utilizará el usuario final (*Business Analytics*), las relaciones no conocidas entre las variables, que tienen que descubrirse mediante la minería de datos y metodologías complementarias como BPM (*Business Performance Management*), que sirven para monitorear el desempeño y obtener ventaja competitiva [30]. La figura 2.2 presenta un esquema sobre los componentes de la Inteligencia de Negocios. Más adelante se desarrolla la arquitectura del Data Warehouse con más detalle explicando cada uno de los módulos.

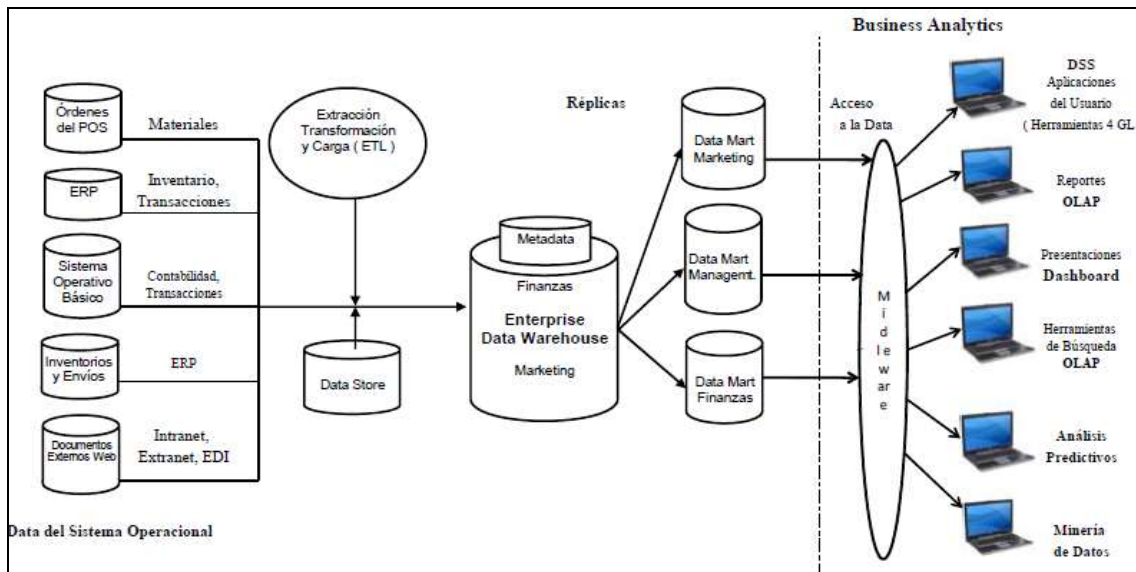


Figura 2.2: Principales componentes de la Inteligencia de Negocios
 Elaborado por Turban, Aronson, Liang y Sharda, 2007

Gonzales López [13] en su tesis de doctoral realiza, un estudio cualitativo exploratorio mediante entrevistas semiestructuradas donde indaga sobre las variables más relevantes y sus componentes, en el impacto que tiene un Data Warehouse (DW) y la Inteligencia de Negocios (BI) en las empresas en el caso de un país en vía de desarrollo como Perú. Mediante estas entrevistas se indagó a empresas proveedoras de DW y BI y a gerentes y usuarios de DW y BI de empresas que utilizaban este tipo de sistema. Con toda esa información, logra hacer una estimación de las principales aplicaciones que se le da a la Inteligencia de Negocios. Los resultados se aprecian en la siguiente tabla:

Actividad	Gerentes (10)	Usuarios (8)	Total (18)	%
Desarrollo de Modelos Predictivos	4	6	10	56%
Seguimiento de Campañas Comerciales	4	6	10	56%
Segmentación de Clientes	3	6	9	50%
Retorno de Inversión	5	0	5	28%
Análisis de resultados Financieros por Áreas	3	1	4	22%
Análisis de ventas	1	3	4	22%
Asignación parte variable de la Remuneración	3	1	4	22%
Mejor Conocimiento de los Clientes	4	0	4	22%
Mejora de Cobranzas	2	2	4	22%
Adquisición de Nuevos Clientes	3	0	3	17%
Rentabilidad de Clientes	3	0	3	17%
Seguimiento de Moras en Pago de Créditos	1	2	3	17%
Análisis de Mercado	0	2	2	11%
Seguimiento del Desempeño Operativo	2	0	2	11%
Análisis de Canales de Distribución	1	0	1	6%
Análisis de Riesgo	0	1	1	6%
Control y Seguimiento de Productos	1	0	1	6%
Seguimiento del cumplimiento de Metas	1	0	1	6%

Cuadro 2.1: Utilización de la Inteligencia de Negocios por parte de las empresas

2.2 Data Warehousing y Data Warehouse

2.2.1 Definición

Los datos de la empresa son activos estratégicos de mucha importancia, siendo su calidad una característica primordial. La función del Data Management (*Administración de los Datos*) es transformar los datos en información de calidad, para lo cual debe mejorar su calidad, combinar diversas fuentes e incrementar su valor. Sin embargo, hay diversos problemas con que nos

encontramos cuando comenzamos con el análisis: aumento exponencial de la cantidad de datos pertenecientes a diversas fuentes, datos externos también en aumento, aspectos de seguridad, calidad e integridad de los datos, excesivo número de herramientas para el manejo de los datos, y datos redundantes en varias partes de la organización [15].

Para llevar a cabo BI es de vital importancia contar con un proceso que satisfaga todas estas necesidades. Es necesario gestionar datos guardados para luego depurarlos, integrarlos y almacenarlos en un solo destino o base de datos que permita su posterior análisis y exploración. Este proceso se denomina *Data Warehousing*.

El Data Warehousing (DWH), es el encargado de extraer, transformar, consolidar, integrar y centralizar los datos que una organización genera en todos los ámbitos de su actividad diaria y/o información externa relacionada. Convertirá los datos operacionales de la empresa en una herramienta competitiva poniendo a disposición de los usuarios específicos la información pertinente, correcta e integrada, en el momento que se necesita. Pero para que el Data Warehousing pueda cumplir con sus objetivos, es necesario que la información que se extrae, se transforme, consolide y sea almacenada de manera centralizada en una base de datos con estructura multidimensional denominada *Data Warehouse* (DW).

Entre las herramientas con mayores posibilidades para lograr el objetivo planteado, se encuentran los DW y los Data Marts que pueden ayudar a las empresas a favorecer una estrategia de integración de la información, imprescindible para generar valor y se puedan apoyar decisiones claves que afecten al negocio con un mayor nivel de certidumbre, con la finalidad de superar los inconvenientes que presentan las bases de datos tradicionales que generalmente tienden a la fragmentación de la información.

Una de las definiciones más famosas sobre DW (almacén de datos), es la de Inmon [16], quien define: “*Un Data Warehouse es una colección de datos orientada a temas, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia*”. Contiene datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional.

La tarea fundamental del DWH es recolectarlos, unificarlos y depurarlos eliminando inconsistencias y conservando solo la información útil. Nos encontramos actualmente sumergidos en la era de la información y el almacén de datos permite que los gerentes tomen decisiones, basados en información confiable y oportuna. Es necesario transformar los datos operacionales en información útil para decidir. En todos los sectores económicos, en todas las empresas, la información se convierte en “quién corta la torta”. Disponer de la información útil y brindarla en el momento en que el usuario la necesite en un formato comprensible y utilizable, estos son los objetivos a lograr.

En el cuadro 2.2 se muestran las diferencias entre los datos operacionales y los datos informativos:

Datos Operacionales	Datos Informativos
Orientados a una aplicación.	Orientados a un tema.
Integración limitada.	Integrados (datos internos y externos).
Constantemente actualizados.	No volátiles.
Valor actual del dato.	Valores a lo largo del tiempo.
Soportan operaciones diarias.	Soportan decisiones de administración.

Cuadro 2.2: Diferencia datos operacionales e informativos

Orientado a temas: una primera característica del DW es que la información se clasifica en base a los aspectos que son de interés para la empresa, se excluye la información que no será utilizada para la toma de decisiones. Mientras que la información orientada a las aplicaciones contiene datos para satisfacer los requerimientos funcionales y de proceso.

Integración: la información del DW está siempre integrada (se emplean convenciones de nombres, medida uniforme de variables, entre otros). La mayor parte de los datos se alteran significativamente al ser seleccionados y movidos al DW. La información que reside en el ambiente operacional no es la misma que desde el punto de vista de integración. Los datos se filtran cuando pasan del ambiente operacional al de depósito y existen muchos datos que nunca salen del ambiente

operacional, solo los datos que se necesitan ingresarán al ambiente de DW. Se consolidan en una sola instancia y se analizan para asegurar su calidad y limpieza. A este proceso se lo conoce como Integración de Datos y cuenta con diversas técnicas y subprocesos para llevar a cabo sus tareas. Una de estas técnicas son los procesos *ETL: Extracción, Transformación y Carga de Datos* (Extraction, Transformation and Load).

De tiempo variante: debe contener datos históricos para usarse en la identificación y evaluación de tendencias. Utilizan un horizonte largo de tiempo mientras que las aplicaciones operacionales la información es más reciente y pueden actualizarse a partir del momento de acceso de acuerdo a la necesidad. La información del DW, una vez registrada correctamente, no puede ser actualizada. Se consideran como una serie larga de “**snapshots**” o **vistas instantáneas**. Es importante tener en cuenta la *granularidad* de los datos para evitar crecimientos incontrolables y desbordamientos de la base de datos. El intervalo de tiempo y periodicidad de los datos debe definirse de acuerdo a la necesidad y requisitos de los usuarios. Toda la información en el DW posee su propio sello de tiempo (figura 2.3).

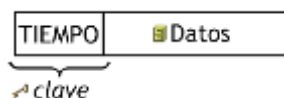


Figura 2.3: Data Warehouse, variante en el tiempo

No Volátil: la información es útil solo cuando es estable. Los datos operacionales cambian de momento a momento. Es esencial que los datos sean estables para el análisis y toma de decisiones. En el ambiente operacional, la información se actualiza regularmente, mientras que en el DW solo hay dos tipos de operaciones: la carga inicial de datos y el acceso a los mismos (figura 2.4). Una vez que el snapshots está hecho, los datos en el depósito no cambian.



Figura 2.4: DW no volátil

2.2.2 Data Mart

Data Mart (DM) es una versión limitada de un DW que es establecido para un grupo de usuarios o un área y está disponible para un número pequeño de personas. Los DW se construyen para servir a las necesidades de información de una organización completa, mientras que los DM sirven a las necesidades de información de un grupo o área específica de la misma. Ambas son herramientas que incorporan las mismas funciones, si bien la capacidad de un DW es superior en cuanto a que permite almacenar mayor volumen de datos que un DM, configurándose como un almacén central para toda la información del negocio.

Las dos tecnologías ayudan a incrementar la seguridad de los datos almacenados y a la vez otorgan una gran operatividad a los usuarios de los SI, pero se diferencian básicamente no sólo en cuanto al tamaño sino también en el coste de implantación. En este sentido, un DM resulta menos costoso de construir que un proyecto de DW; requiere menos ingeniería, menor coste del hardware y se necesitan menos recurso para desarrollarlo y mantenerlo. Por este motivo, muchos expertos recomiendan comenzar con un DM como paso previo para vincular a la empresa con esta tecnología sin asumir el mayor riesgo que implica el alto coste de implantación de un DW.

De acuerdo a las operaciones que se deseen o requieran desarrollar, los DM pueden adoptar las siguientes arquitecturas:

- **Top-Down:** Primero se define el DW y luego se desarrollan, construyen y cargan los DM a partir del mismo (figura 2.5). Esta forma de implementación cuenta con la ventaja de no tener que

incurrir en complicadas sincronizaciones pero requiere una gran inversión y una gran cantidad de tiempo de construcción.



Figura 2.5: Top-Down

- **Bottom-Up:** En esta arquitectura, se definen previamente los DM y luego se integran en un DW centralizado (figura 2.6). En muchas ocasiones los DM son implementados sin que exista el DW, ya que tienen las mismas características pero con la particularidad de que están enfocados a un tema específico. Luego de que hayan sido creados y cargados todos los DM, se procederá a su integración con el depósito.

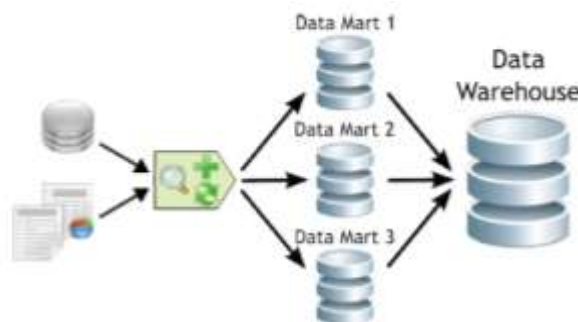


Figura 2.6: Bottom-Up

La potencialidad de estas herramientas, tanto de los DW como de los DM, reside en la capacidad que tienen para servir de apoyo a los **DSS (Sistemas de Ayuda a la toma de decisiones)** y a los **EIS (Sistemas de Información para Ejecutivos)**, al manejar información que presenta unas características propias y que no quedan cubiertas por los sistemas tradicionales de información que están más orientados hacia el tratamiento de información de gestión.

2.2.3 Inmon vs Kimball

Breslin [3] hace una comparación de los dos principales modelos usados para definir un Data Warehouse, el modelo de **Inmon** [16] y el de **Kimball** [18]. El primero es del tipo de desarrollo Top-Down, que adapta herramientas relacionales tradicionales de base de datos a las necesidades de una DW para toda la empresa, y el segundo modelo usa un tipo de desarrollo Bottom-Up con un modelamiento dimensional creando Data Marts por grupos de usuarios o áreas.

El modelo de Inmon tiene tres niveles, el Entity Relationship Diagrams (ERD), el data items set para cada departamento y el modelo físico, para luego usar la metodología de desarrollo espiral y la determinación de la granularidad.

Kimball utiliza un modelamiento dimensional con tablas de hechos y tablas de dimensiones. Las tablas de hechos o *fact tables* contienen métricas y las tablas de dimensiones o *dimensional tables* contienen atributos de las métricas de las tablas de hechos. Las tablas de dimensiones contienen grupos de datos repetidos, lo que viola las reglas de normalización, pero consigue un alto nivel de desempeño en la DW, permitiendo un fácil acceso al usuario final a una adecuada velocidad de respuesta. La estandarización de las Data Marts para lograr una Data Warehouse en su conjunto se logra mediante la arquitectura de bus, consiguiendo dimensiones conformadas estandarizadas. Los cuatro pasos del diseño dimensional son: seleccionar el proceso de negocio, establecer la granularidad, escoger las dimensiones e identificar los hechos para las tablas de hechos.

Las principales similitudes entre el modelo de Inmon y el de Kimball son el uso de datos relacionados con una fecha específica y el proceso de extracción, transformación y carga (ETL). Las diferencias van por el lado del desarrollo de la metodología, el modelamiento de los datos y la arquitectura.

En relación a la metodología de desarrollo y arquitectura, en el modelo de Inmon, el desarrollo es Top-Down con cierta complejidad técnica, por lo que tiene que ser manejado mayormente por los especialistas de IT, mientras que en el caso de Kimball, con su esquema de cuatro pasos, es mucho más accesible al usuario final, así como el concepto de las Data Marts, por su menor dimensión. En relación al modelamiento de la información, en el caso de Inmon la orientación es al tema o asunto mientras que en el caso de Kimball es al proceso de negocio, lo cual hace que se crucen las líneas entre varios departamentos de la empresa. En relación a las diferencias filosóficas, el modelo de Inmon ve a IT como el proveedor y desarrollador de la Data Warehouse, mientras que Kimball ve a un equipo conformado igualmente por personas de IT y de la parte de los usuarios finales.

El modelo de Inmon sería recomendable para empresas que tienen un equipo grande de especialistas en Data Warehouse, que van a desarrollar un proyecto amplio para toda la empresa, que va almacenar datos que no es exclusivamente métricas de negocios, y que puede esperar un plazo largo para ver los resultados (de cuatro a nueve meses). En el caso del modelo de Kimball, es recomendable para aplicar a proyectos que se realizan por partes (Data Marts), demandando la primera unos 90 días para luego demorar unos 60 a 90 días en cada una de las sub-siguientes.

En el cuadro 2.3 se pueden ver las principales características de ambos modelos:

Metodología y Arquitectura	Inmon	Kimball
Enfoque general	Arriba-abajo (Top-Down)	Abajo-arriba (Bottom-Up)
Estructura de la Arquitectura	Data Warehouse para toda la empresa (atómica). Suministra datos a las base de datos departamentales	Data Mart para cada proceso de negocio; Consistencia a través de todos las Data Marts, mediante un data bus y dimensiones conformadas
Complejidad del método	Bastante compleja	Bastante simple
Comparación con metodologías de desarrollo ya establecidas	Derivada de la metodología de Espiral	Proceso de 4 pasos, derivada de metodología RDBMS
Discusión del diseño físico	Bastante compleja	Bastante simple
Modelamiento de la Data		
Orientación de la data	Orientada al asunto o a la data	Orientada al proceso
Herramientas	Tradicional (ERD, DIS)	Modelamiento dimensional, derivado del modelamiento relacional
Accesibilidad del usuario final	Baja	Alta
Filosofía		
Audiencia primaria	Profesionales de IT	Usuarios finales
Objetivo	Dar una solución técnica eficiente basada en métodos y tecnologías de base de datos probados	Dar una solución fácil para el usuario final, de manera que pueda acceder a la data directamente a una velocidad razonable

Cuadro 2.3: Comparación de las Características Esenciales de los Modelos de Inmon y Kimball
 Elaborado por Breslin, 2004

2.2.4 Arquitectura del Data Warehousing

En este punto describiremos de manera más específica los componentes que intervienen en la arquitectura del Data Warehousing (DWH). El ambiente está formado por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema.

A continuación se detallará cada uno de los componentes de la arquitectura del DWH, teniendo como referencia la figura 2.7.

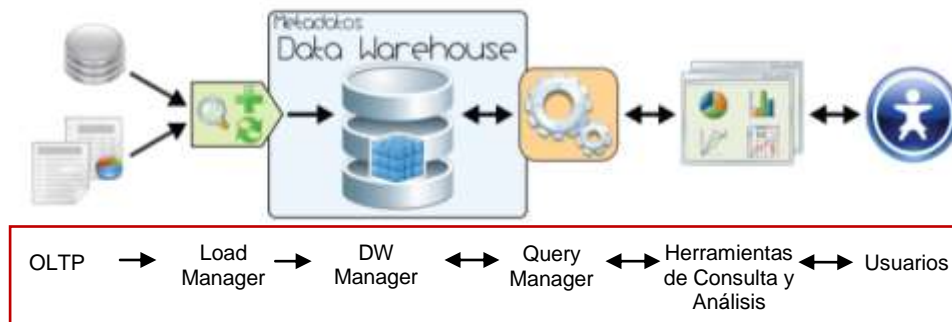


Figura 2.7: Arquitectura del DWH

2.2.4.1 OLTP

OLTP (On Line Transaction Processing), representa toda aquella información transaccional que genera la empresa en su accionar diario, además de las fuentes externas con las que puede llegar a disponer. Los datos son extraídos desde aplicaciones, base de datos, archivos, etc. Esta información generalmente reside en diversas fuentes de sistemas, orígenes y arquitecturas y tienen formatos muy variados.

2.2.4.2 Load Manager

La Integración de Datos agrupa una serie de técnicas y subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, control, integración, depuración de datos, carga y actualización del DW, etc. Es decir, todas las tareas que se realizarán desde que se toman los datos de los diferentes OLTP hasta que se cargan en el DW.

Alimentar un DW consiste en migrar y preparar los datos provenientes de los sistemas transaccionales o de fuentes externas para su posterior análisis en un entorno de decisión. El procesamiento de Extracción, Transformación y Carga (ETL), es el componente principal del proceso del DW y ocupa alrededor de un 70% del tiempo en un proyecto. Este consiste en la extracción (leer datos de una o más base de datos), la transformación (convertir los datos extraídos a la forma que necesita el DW), y la carga (ponerla en el DW). La transformación ocurre usando reglas o tablas especializadas, o combinando los datos.

La fase de **Descubrimiento y Extracción** es una tarea meramente de análisis de los sistemas fuentes (transaccionales). Es necesario haber analizado previamente los sistemas transaccionales para poder determinar de dónde extraer la información y con qué criterios. La ventaja que tenemos en este proyecto es que conocemos los sistemas fuentes ya que son los que se están utilizando en la actualidad. El propósito principal de la fase de extracción es capturar y copiar los datos requeridos de uno o más sistemas operacionales o fuentes de datos. Los datos que se extraen son colocados en un archivo intermedio con un formato definido, que luego será utilizado por la siguiente fase del proceso.

Transformación y Depuración: una de las claves de un proyecto de DW es el formateo y la preparación de los datos para su análisis. Para depurar hay que identificar los datos anormales, lo que equivale a realizar un análisis semántico de los datos. La transformación significa codificar datos que se han decodificado en forma inconsistente en diferentes fuentes (por ejemplo abreviaturas de estados, códigos, etc). Las funciones básicas consisten en leer los archivos intermedios generados por la fase de extracción, realizar las transformaciones necesarias, construir los registros en el formato del DW y crear un archivo de salida conteniendo todos los registros nuevos a ser cargados en el mismo. Es fundamental definir estándares para que todos los datos que ingresen al DW estén integrados.

Transferencia y Carga: consiste en extraer los datos de los sistemas fuentes y a partir de ellos cargar la base destino aplicándoles las transformaciones en el proceso. El objetivo de esta fase consiste en tomar los registros formateados por la fase de transformación y cargarlos en el DW, que es el contenedor para todos los datos informativos (actuales e históricos). Tenemos dos tareas relacionadas a esta función: por un lado la *carga inicial (Initial Load)*, que se refiere a la primera carga de datos para insertar los registros generados durante varios años; y por otro lado, la *actualización o mantenimiento periódico* donde se mueven pequeños volúmenes de datos y su frecuencia está dada

por la granularidad² del DW y los requerimientos definidos. El objetivo de esta última tarea es añadir al depósito aquellos datos nuevos que se fueron generando desde el último refresco.

2.2.4.3 Data Warehouse Manager

El DW Manager presenta las siguientes características y funciones principales:

- Se constituye al combinar un SGBD³ con software y aplicaciones dedicadas.
- Almacena los datos de forma multidimensional (a través de tablas de hechos y de dimensiones), gestionando diferentes estructuras de datos (Cubos Multidimensionales).
- Gestiona y mantiene metadatos.
- Transforma e integra los datos fuentes y del almacenamiento intermedio en un modelo adecuado para la toma de decisiones.
- Realiza todas las funciones de definición y manipulación del depósito de datos, para poder soportar todos los procesos de gestión del mismo.
- Ejecuta y define las políticas de particionamiento.
- Realiza copias de resguardo incrementales o totales de los datos del DW.

Base de datos Multidimensional

Una base de datos multidimensional almacena información a través de tablas de hechos y dimensiones. Provee una estructura de datos multidimensional que permite tener acceso flexible a los datos.

El cubo multidimensional es una estructura que permite realizar las diferentes y posibles combinaciones para visualizar los resultados de una organización hasta un determinado grado de detalle. Esta estructura es independiente del sistema transaccional de la organización y facilita consultar información histórica de manera rápida y eficiente, obteniendo la posibilidad de navegar y analizar los datos requeridos. La estructura del cubo está organizada como una matriz multidimensional donde los ejes se denominan **dimensiones** y representan los criterios de análisis, y a los datos almacenados se los llama **medidas** (hechos) que representan los indicadores o valores a analizar.

Las **tablas de hechos** contienen los indicadores que serán utilizados por los analistas del negocio para apoyar el proceso de toma de decisiones. Son datos instantáneos en el tiempo que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Las sumalizaciones no están referidas solo a sumas, sino también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, etc dependiendo de los requerimientos de información del negocio. Existen tres clases de medidas:

Aditivas: pueden ser combinadas a lo largo de cualquier dimensión.

Semiaditivas: no pueden ser combinadas a lo largo de una o más dimensiones.

No aditivas: no pueden ser combinadas a lo largo de ninguna dimensión.

Las **tablas de dimensiones** definen cómo están organizados los datos lógicamente y proveen el medio para analizar el contexto del negocio. Representan los aspectos de interés mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos. Cada tabla posee un identificador único y al menos un campo o dato de referencia que describe los criterios de análisis relevantes para la organización (datos de referencia). En un DW, la creación y el mantenimiento de una tabla de dimensión tiempo es obligatorio y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se está analizando. Se debe evaluar con cuidado la temporalidad de los datos, la forma en que trabaja la organización, los resultados que se desean obtener del almacén de datos relacionados con una unidad de tiempo y la flexibilidad que se desea obtener de dicha tabla.

² La **granularidad** se refiere al grado de detalle que el DW soporta. Cuando la granularidad es alta no se podrán realizar indagaciones de mucha precisión y pasar de datos resumidos a detalles individuales de transacciones. Con una granularidad baja habrá mucho más detalle, lo que ocupará más espacio en el DW y no permitirá una respuesta muy rápida, pero si sumamente detallada. En el proyecto se tendrá que definir con que granularidad se deberá trabajar.

³ Un **SGBD** es un software que facilita una serie de herramientas para manejar bases de datos y obtener resultados (información) de ellas. Además de almacenar la información, se le pueden hacer consultas sobre estos datos, obtener listados impresos, generar pequeños programas de mantenimiento de la BD, o ser utilizado como servidor de datos para programas más complejos. Centraliza los datos en un solo lugar para ser accedidos por múltiples usuarios y aplicaciones concurrentes. Posee herramientas para asegurar: independencia, integridad y seguridad de datos.

El cubo multidimensional representa o convierte los datos planos que se encuentran en filas y columnas en una matriz de N dimensiones. Los objetos más importantes del cubo son:

Indicadores: son sumalizaciones que se efectúan sobre algún hecho o expresiones basadas en sumalizaciones pertenecientes a una tabla de hechos.

Atributos: constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional.

Jerarquías: representa una relación lógica entre dos o más atributos.

En un cubo multidimensional como el de la figura 2.8, los atributos existen a lo largo de varios ejes o dimensiones y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando.

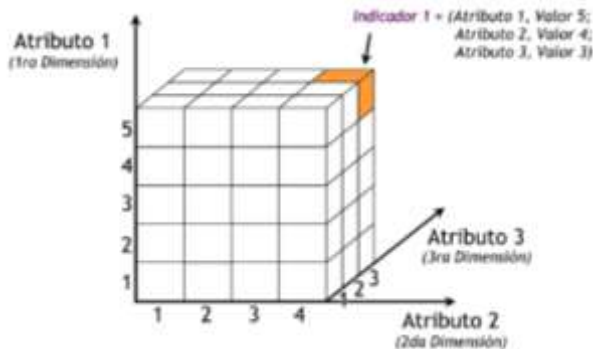


Figura 2.8: Cubo Multidimensional

Tipos de modelamiento de un DW

Esquema estrella

El esquema estrella, consta de una tabla de hechos central y de varias tablas o dimensiones relacionadas a esta a través de sus respectivas claves. Este modelo se representa en la figura 2.9 y está totalmente desnormalizado. Con ello hay redundancia y repetición de datos, pero se optimiza los tiempos de respuesta y es más simple de interpretar. Este modelo es soportado por casi todas las herramientas de consultas y análisis, y los metadatos⁴ son fáciles de documentar y mantener, sin embargo es menos robusto para la carga y es el más lento de construir.

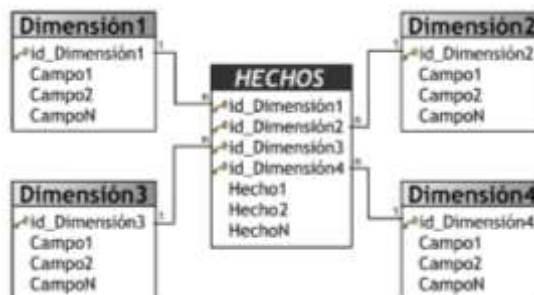


Figura 2.9: Esquema en Estrella

Esquema Copo de Nieve

Este esquema representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones. Existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas con una o más tablas de dimensiones (figura 2.10). Es más cercano a un modelo de Entidad Relación, que al modelo en estrella, debido a que sus tablas están normalizadas. Es un modelo muy flexible, hay menos esfuerzo de diseño y puede implementarse después de que se ha desarrollado un

⁴ Los **metadatos** son datos que describen o dan información de otros datos que en este caso, existen en la arquitectura del Data Warehousing. Brindan información de localización, estructura y significado de los datos, básicamente mapean los mismos. Se pueden distinguir tres tipos de metadatos: del proceso ETL, operacionales y de consulta.

esquema en estrella. Sin embargo, es más complejo en su estructura y el número de tablas puede ser inmanejable.

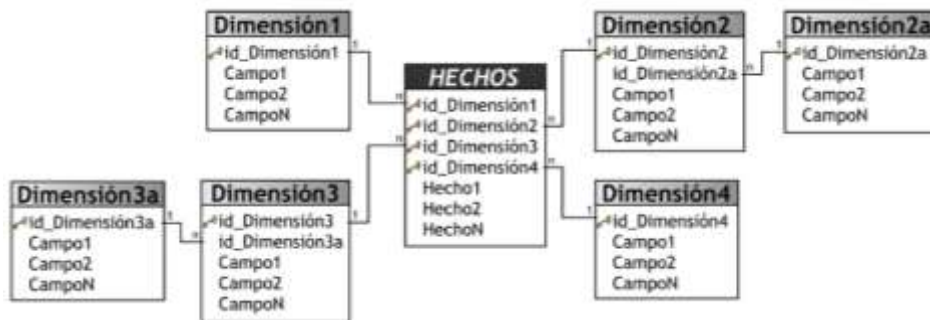


Figura 2.10: Esquema Copo de Nieve

Esquema Constelación

Este esquema está compuesto por una serie de esquemas en estrella. Está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones de la principal (figura 2.11). Además están relacionadas con sus respectivas tablas de dimensiones. Al tener más de una tabla de hechos, permite analizar más aspectos claves del negocio con un mínimo esfuerzo de diseño y permite reutilizar las tablas de dimensiones. No es soportado por todas las herramientas de consulta y análisis.

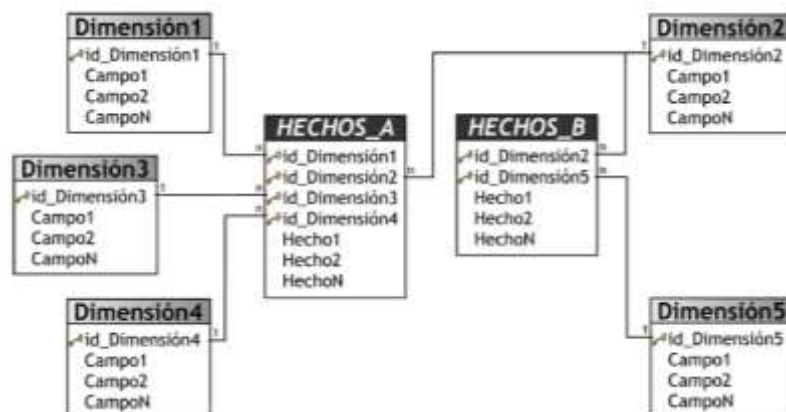


Figura 2.11: Esquema Constelación

2.2.4.4 Query Manager

Este componente realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, tales como joins y agregaciones, y de consultas propias del análisis de datos, como “drill-up” y “drill-down”. Recibe las consultas, las aplica a la estructura de datos correspondiente y devuelve los resultados obtenidos. Hay operaciones que son las que permitirán a los usuarios explorar e investigar en busca de respuestas, son:

Drill-down: Permite apreciar los datos en un mayor detalle, bajando por una jerarquía definida en un cubo. Brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, disgregando los grupos actuales.

Drill-up: Permite apreciar los datos en un menor nivel de detalle, subiendo por una jerarquía definida en un cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, agregando los grupos actuales.

Drill-across: Funciona de manera similar a drill-down, con la diferencia de que drill-across no se realiza sobre una jerarquía, sino que su forma de ir de lo general a lo específico es agregar un atributo a la consulta como nuevo criterio de análisis.

Roll-across: Funciona de manera similar a drill-up, con la diferencia de que roll-across no se hace sobre una jerarquía, sino que su forma de ir de lo específico a lo general es quitar un atributo de la consulta, eliminando de esta manera un criterio de análisis.

Pivot: Permite seleccionar el orden de visualización de los atributos e indicadores, con el objetivo de analizar la información desde diferentes perspectivas. Admite mover un atributo o indicador desde el encabezado de fila al de columna y viceversa, como así también cambiar el orden de los atributos de las filas y columnas.

Page: Presenta el cubo dividido en secciones, a través de los valores de un atributo, como si se tratase de páginas de un libro. Es muy útil cuando las consultas devuelven muchos registros y es necesario desplazarse por los datos para poder verlos en su totalidad.

Drill-through: Permite apreciar los datos en su máximo nivel de detalle. Esto brinda la posibilidad de analizar cuáles son los datos relacionados al valor de un indicador, que se ha sumariado dentro del cubo multidimensional.

2.2.4.5 Herramientas de consulta y análisis

Las herramientas de consulta y análisis son sistemas que permiten realizar la exploración de datos del DW. Utilizan metadata de las estructuras de datos que han sido creadas previamente (cubos multidimensionales) para devolver los resultados obtenidos. A través de una interfaz gráfica, los usuarios generan las consultas seleccionando los datos que desean obtener del DW. La herramienta recibe el pedido y construye la consulta que es enviada desde la herramienta de consulta y análisis al Query Manager. Este último, ejecuta la consulta sobre la estructura de datos (cubo multidimensional) realizando la extracción de información al DW manager. Finalmente devuelve los resultados obtenidos a la herramienta que lo solicitó.

Una de las principales ventajas de utilizar estas herramientas es que los usuarios no se deben preocupar por conocer cuáles son las características y funcionalidades de datos utilizadas, solo se deben enfocar en el análisis. Permiten: el acceso a la información de manera transparente, apoyo en la toma de decisiones obteniendo información oportuna, relevante y útil, y mayor interpretación y análisis de los usuarios sin poseer conocimientos técnicos.

Existen diferentes tipos de herramientas de consulta y análisis, y de acuerdo a las necesidades del negocio y requerimientos de información, se deberán seleccionar las más propicias al caso, entre ellas podemos destacar: reportes y consultas, OLAP, Dashboards, Data Mining y EIS.

OLAP

El **Procesamiento Analítico en Línea (OLAP- OnLine Analytical Processing)**, es el componente más poderoso del DW, ya que es el motor de consultas especializado del depósito de datos. Implica la organización multidimensional de la información desde las diferentes perspectivas o visiones de la institución y la utilización de herramientas especiales para visualizarla y que permiten operaciones de ordenamientos, navegación, pivoteo, etc. Su principal objetivo es el de brindar rápidas respuestas a complejas preguntas, para interpretar la situación del negocio y tomar decisiones. Cabe destacar que lo que es realmente interesante en OLAP, no es la ejecución de simples consultas tradicionales, sino la posibilidad de utilizar los operadores desarrollados con anterioridad como drill-up, drill-down, etc para explorar profundamente la información.

La ventaja de OLAP respecto de los reportes estáticos es que los usuarios pueden realizar consulta sobre la marcha, generando un alto grado de interactividad, permitiendo que los datos se analicen en tiempo real. En vez de recibir informes preparados por el Centro de Cómputos, los directivos y gerentes pueden crear sus propios informes, con la flexibilidad de manipular la información para descubrir tendencias y áreas problemáticas. En la actualidad los gerentes son más autosuficientes y no dependen del Departamento Sistemas para obtener informes personalizados.

Estas herramientas manejan una serie de consultas sobre estructuras multidimensionales (Cubos OLAP) cargadas previamente con los datos almacenados en las bases de datos corporativas tradicionales. Permiten realizar informes y obtener grandes cantidades de información a partir de lo que resultaría ser a modo rutinario una serie de complejas consultas sobre una base de datos. Al estar los datos precompilados sobre una estructura intermedia, el tiempo de respuesta de las consultas es menor, posee una enorme potencia de cálculo y técnicas de indexación especializadas. Esta tecnología es favorable en un sistema OLTP pero suele ser lenta si se realizan complejas consultas. Con estos sistemas es posible analizar la información almacenada en un Data warehouse, pero no es estrictamente necesario, ya que la información puede provenir de diferentes bases de

datos. El objetivo de estas herramientas es obtener una mejor comprensión de lo almacenado en las bases de datos.

Almacenamiento de Cubo

Todas ellas se basan en la tecnología OLAP que permite trabajar sobre los datos al pensarlos como hiper cubos con la información de la empresa. Existen diversas opciones en las que se negocia menor tiempo de acceso a cambio de mayor espacio utilizado en disco y viceversa. Existen tres modos posibles:

MOLAP: El objetivo de los sistemas MOLAP (**Multidimensional On Line Analytic Processing**) es almacenar físicamente los datos en estructuras multidimensionales. Para ello dispone de estructuras de almacenamiento específicas y técnicas de compactación de datos que favorecen el rendimiento del DW. Requiere que en una instancia previa se generen y calculen los cubos multidimensionales para que luego puedan ser consultados. El principal motivo de precalcular los datos de los cubos, es que posibilitan que las consultas sean respondidas con mucha rapidez, ya que los mismos no deben ser calculados en tiempo de ejecución, obteniendo de esta manera una muy buena performance. Tienen la desventaja de que cada vez que es necesario realizar cambio sobre algún cubo, se debe recalcular totalmente para que se lleven a cabo las modificaciones, provocando de esta manera una disminución importante en cuanto a flexibilidad. Además precisa más espacio físico para almacenamiento.

ROLAP: Este tipo de organización física se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento. ROLAP (**Relational On Line Analytic Processing**) cuenta con todos los beneficios de una SGBD Relacional. Los datos se encuentran en tablas de base de datos relacionales. Los cubos multidimensionales se generan dinámicamente al instante de realizar las diferentes consultas, haciendo de esta manera el manejo del cubo transparente a los usuarios. Es el OLAP más rápido en la inserción de datos y el de menor uso de espacio en disco. La principal desventaja de los sistemas ROLAP es que los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos. Esto provoca no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas.

HOLAP: Corresponde a la implementación híbrida que trata de obtener lo mejor de los dos modos anteriores. HOLAP (**Hybrid On Line Processing**) se comporta como un MOLAP para la información resumida (y constantemente utilizada), y como un ROLAP al solicitar un detalle de ésta. Este sistema combina las dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional. Los recursos promedios utilizados se ubican entre las dos opciones anteriores, al igual que el rendimiento.

2.2.4.6 Usuarios

Los usuarios que posee el DW son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio, es por ello que se hace tanto énfasis en la integración y limpieza de datos para conseguir que la información posea toda la calidad posible. Explora los datos en busca de respuestas para poder tomar decisiones proactivas. Los usuarios que acceden concurrentemente son pocos, por lo general son consultas complejas, no predecibles y no anticipadas, generan análisis sobre una gran cantidad de registros, sin tener tiempos de respuesta críticos y realizan una única operación: la consulta.

El cuadro 2.4 describe los usuarios de DW y los compara con los usuarios operacionales.

Característica	Usuarios de OLTP	Usuarios de DW
Acceso concurrente	Muchos	Pocos
Tipos de Consultas	Predefinidas	Complejas, no predecibles y no anticipadas
Registros consultados	Pocos	Muchos
Tiempo de Respuesta	Crítico	No Crítico
Acciones permitidas	Agregar, modificar, eliminar y consultar	Consultar

Cuadro 2.4: Análisis de los distintos tipos de usuarios

La gran mayoría de las organizaciones se encuentra organizada jerárquicamente en una forma piramidal (figura 2.12). En cada escalón se puede situar a los diferentes tipos de usuarios a la hora de tratar las herramientas que ofrece la Inteligencia de Negocio:

Dirección General: Aquí se sitúan los altos cargos de la empresa, quienes cargan con la mayor responsabilidad en la organización y, por tanto, no disponen de tiempo suficiente para dedicarlo al análisis de información. Es por esto que utilizan herramientas como scorecards o dashboards, donde obtienen una visión rápida y global de los movimientos del negocio, que además puede contener enlaces a informes más concretos.

Cargos Medios: A esta altura encontramos el perfil típico de la rama Administrativa de la empresa. Estos cargos disponen de herramientas para el análisis de ciertos indicadores del negocio a mayor detalle que sus jefes inmediatos, ya que deben responder con resultados e informes ante ellos. Las herramientas de OLAP y consultas ad-hoc permiten que este tipo de usuarios pueda representar de forma visual diferentes tipos de análisis tanto en informes de tablas como en gráficos, partiendo de una gran cantidad de datos distribuida en cubos multidimensionales para realizar consultas.

Operarios: Son los cargos más bajos de la empresa, pero aun así las herramientas de Inteligencia de Negocio les aportan cantidad de beneficios en su labor. Son usuarios de lo que se conoce como Informes Predefinidos, es un tipo estándar de informe con posibilidad de compartirlos con un formato típico corporativo, donde se incluyen gráficas y tablas de información obtenidas directamente de los procesos ETL.



Figura 2.12: Toma de decisiones de acuerdo a un tipo de usuario

CAPITULO III

3. Desarrollo de la Metodología

El propósito general del capítulo 3 es instaurar una guía metodológica de BI basada en el análisis de las fuentes de información y adaptada a los requerimientos de la CPAC. Lo que se busca es demostrar las ventajas del BI y motivar a los directivos y gerentes de su necesidad de implementación proporcionando a los usuarios del sistema una manera universal de acceder, ver y utilizar la información. También se presenta un plan inicial de integración de datos, una herramienta de desarrollo factible de utilizar y se realizan algunas operaciones sobre un modelo multidimensional creado a modo de prototipo.

Finalmente, en el próximo capítulo, el número 4, se realiza un análisis de las variables que intervienen en el momento de evaluar los resultados obtenidos y se describen cada una de ellas.

3.1 Metodología de la Investigación

A continuación se describe la metodología de la investigación aplicada en el desarrollo de este trabajo. Desde el punto de vista de la finalidad es de tipo **Aplicada** ya que resuelve el problema planteado proponiendo un sistema de ayuda a la toma de decisiones, que integra los datos de la organización y permite analizar la información. De acuerdo al alcance temporal es del tipo **Seccional** o **Sincrónico**, ya que se analiza el problema al día de hoy (vemos su estado actual).

Desde el punto de vista de la profundidad o alcance es del tipo **Descriptivo** ya que se conocen los factores del problema pero no se sabe cómo se relacionan. Se caracteriza el estado de situación y se investiga para elaborar un diagnóstico. En cuanto a la amplitud, este trabajo propone el **Estudio de Caso**, ya que se analiza el estado en particular de la CPAC frente a los requerimientos de información precisa.

Por lo que refiere a la naturaleza (que representa al grado de naturalidad o artificialidad del comportamiento de las variables), este trabajo es del tipo **No Experimental** o **Natural**, ya que realiza un análisis post facto, es decir después de los hechos. Más aún se puede decir que es del subtipo **De Campo** ya que se analiza una situación real.

3.2 Guía Metodológica de BI

En este apartado se describe una metodología de BI adaptada a los requerimientos de la institución, cuya propuesta está fundamentada en una amplia investigación que parte de la recolección de los requerimientos y necesidades de información de la institución y concluye en la presentación de la solución de BI con la exposición de algunos resultados. Este desarrollo ayudará a los distintos sectores y departamentos de la institución a obtener resultados benéficos como: generar reportes globales o por secciones, conocer la realidad de los aportantes a la previsión y obra social, crear escenarios con respecto a una decisión, compartir información entre los departamentos, análisis multidimensionales, generar y procesar datos, en definitiva, mejorar el servicio al afiliado.

El principal objetivo de esta guía es facilitar el arduo trabajo que significa construir una solución de BI y elaborarla de manera sencilla, ordenada e intuitiva. Básicamente se basa en cuatro pasos o etapas. A continuación se detallan cada una de ellas demostrando parte de los resultados obtenidos y ejemplificando cada concepto.

- 1- **Análisis de Requerimientos:** En primer lugar se comienza recolectando las necesidades de información y se obtienen las preguntas claves del negocio. La ventaja es que al trabajar en la institución hace más de 10 años, se conoce ampliamente el negocio y sus necesidades de información facilitando la toma de requerimientos. Luego se establecen los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construye el modelo conceptual de datos del DW.
- 2- **Análisis de los OLTP:** Se analizan los OLTP para determinar los indicadores, señalando las correspondencias con las fuentes de datos y seleccionando los campos de estudio de cada perspectiva.

- 3- **Construcción del Modelo lógico del DW:** Seguidamente se pasa a la construcción del modelo lógico del DW, en donde se define cuál es el tipo de esquema que se implementará, se define el nivel de granularidad y se confeccionan algunas tablas de dimensiones y de hechos.
- 4- **Integración de datos:** Por último, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc, se definen políticas y estrategias para la Carga Inicial del DW y su respectiva actualización.

A continuación se desarrolla cada etapa propuesta y se presentan los resultados obtenidos.

3.2.1 Análisis de Requerimientos

En primer lugar se identifican los requerimientos de los usuarios que expliciten los objetivos de la institución, sus metas y estrategias. Para ello se recopilan una serie de informes realizados a lo largo del tiempo y se establecen las prioridades, identificando los indicadores y perspectivas. Algunos reportes son planillas de cálculo mientras que otros son reportes de los sistemas transaccionales. En general, y sobre todos los primeros, tienen formatos preestablecidos y es muy dificultosa su actualización. Algunos se conforman luego de complejas consultas contra la base de datos operacional. De todas formas, el acopio de las necesidades de información podría llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

El análisis de los requerimientos de los usuarios es el punto de partida de esta metodología, ya que son ellos los que guían la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y cualidades. Es fundamental tener en cuenta que dicha información, es la que proveerá el soporte para desarrollar los pasos sucesivos, por lo cual, es muy importante prestar especial atención al relevar los datos. Es imprescindible que se formulen preguntas complejas sobre el negocio, que incluyan todas las variables de análisis que se consideren relevantes, ya que son las que permitirán estudiar la información desde diferentes perspectivas.

Se trata de una tarea meticulosa donde hay que actuar como consultor de estrategia de negocio además de como desarrollador, ya que hay que atender al detalle, a través de sucesivas reuniones con los interesados y/o departamentos involucrados, así como analizar los sistemas y métodos de obtención de información que se están utilizando previamente a la implantación de la solución.

Una forma de asegurar de que se ha realizado un buen análisis de la información, es corroborar que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la empresa que se está estudiando. Esta fase es la más importante dentro de la implantación de una solución de BI, ya que, si se falla en este proceso, las posteriores fases no tendrán sentido y obtendremos una solución errónea. Es la que establece el rumbo que tomará dicha solución en la empresa.

A la hora de iniciar la fase de la toma de requerimientos es necesario plantearse como punto de partida los objetivos a conseguir con cualquier solución de BI y tener en cuenta los diferentes fallos que se han presentado a lo largo del tiempo a la hora de llevar a cabo un proyecto de BI para no caer en ellos.

Requerimientos Genéricos

En cualquier solución de BI hay una serie de requerimientos genéricos, donde se deben tener en cuenta los siguientes puntos:

- Proveer de un sistema intuitivo y orientado hacia un usuario final con pocos conocimientos técnicos donde éste sea capaz de generar sus propios reportes y análisis.
- Tener una sola versión de la información. A través de los procesos ETL, la información queda preparada a gusto del usuario final y se presenta siempre actualizada de la misma manera en cualquier ocasión.
- Proveer información de toda la institución en un solo sistema. Si bien la institución maneja fundamentalmente una sola fuente de datos, la información se encuentra dispersa y repetida en varias tablas de la base operacional. Es fundamental realizar una integración de datos.

Requerimientos Funcionales

Los requerimientos funcionales son aquellos que marcan la base para tomar decisiones en las posteriores etapas de un proyecto de BI. Es muy común que muchas empresas decidan su proyecto de BI en base a las herramientas disponibles.

- Debe proveer la facilidad de drill down, slice and dice, fórmulas avanzadas...etc.
- Se necesitan mecanismos para controlar la seguridad de los datos por departamento, área o gerencia así como una distribución organizacional jerárquica de la información. Es fundamental establecer una plataforma securizada y por medio de los roles definir la posibilidad de distribuir la información contenida en la plataforma de forma jerárquica según el cargo y departamento del usuario.
- Como mejora se podría definir un mecanismo de notificaciones y alertas, con criterios y reglas configurables.

Analizando de forma objetiva estos requerimientos se puede observar que la mayoría de plataformas de Business Intelligence en el mercado actual cumplen esta serie de objetivos, aunque esto no es suficiente para que la solución desarrollada cumpla con el objetivo principal que es la de aportar valor al negocio.

Requerimientos de Datos

Otra vía de toma de requerimientos es la de analizar los reportes que se están sacando actualmente con las diversas herramientas para conseguir incluir todo lo necesario en la solución, de manera que los usuarios finales puedan realizar los reportes sin problemas. Ellos son los que conocen de primera mano su departamento y su área de trabajo, así como los reportes que se realizan. Es muy común y entendible que dichos usuarios marquen la pauta de los datos que se requieren y en el formato en que los quieren.

Aunque este escenario es importante para que los usuarios dispongan de toda la información necesaria, por sí sola no es una estrategia de toma de requerimientos completa, debido a que el resultado final no sería el aportar un valor de negocio sino una manera de automatizar las tareas de reporting.

Requerimientos de Negocio

Una manera de poder dar valor al negocio es hacer requerimientos que dependan del mismo. Para ello es necesario definir requerimientos donde se establezcan los objetivos de la solución que sean medibles, controlables y confiables. La toma de requerimientos sobre este escenario puede resultar más compleja que en los casos anteriores, ya que cada decisión debe aportar valor al negocio de manera directa o indirecta según una serie de objetivos:

- Definir la estrategia de negocio en procesos que sean medibles y controlables.
- Identificar la información necesaria para cada proceso de negocio, detallando su forma de análisis, seguimiento, técnicas de análisis, así como requerimientos tecnológicos para su cumplimiento.

Fallos comunes en las soluciones de Business Intelligence

Además de tener en cuenta los diferentes requerimientos debemos tener presente una serie de fallos en los que se suele caer a la hora de llevar a cabo un proyecto de BI [34]. Los más comunes son:

- Los Data Warehouses crecen en gran tamaño muchas veces no planificado.
- Los usuarios carecen de tiempo para dedicarle al proyecto o tienen falta de conocimientos necesarios para apoyarnos en la toma de requisitos.
- Se suelen producir retrasos en la entrada a producción del sistema debido a una mala planificación.
- Presupuesto erróneo, el precio presupuestado es escaso en comparación con la complejidad de lo que se quiere desarrollar.
- Mala elección de software y hardware sin tener en cuenta los criterios técnicos.
- No se realizan pruebas de concepto para determinar si el proyecto es viable.
- Los datos de origen no están limpios. Esto implica procesos ETL más costoso, mayor tamaño de la base de datos y peor rendimiento.
- Escasa involucración de los usuarios finales, la cual les lleva a sentir cierta frustración cuando se les presentan unos resultados que no son los esperados.

3.2.1.1 Aplicación Etapa 1 al Caso de Estudio

Las necesidades de información en la CPAC son muchas. Las mismas abarcan casi todas las actividades de la empresa, pero como para dar comienzo a esta investigación y demostrar la necesidad de un DW a corto plazo, se escogieron indicadores con sus variables o perspectivas asociadas para la toma de decisiones. Se tuvo en cuenta para la elección de los identificadores que sean realmente efectivos, en general valores numéricos. En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores. Cabe destacar que el tiempo es muy comúnmente una perspectiva. Las preguntas de negocio identificadas son:

- Cantidad de afiliados titulares según los distintos estados en previsión (Activos, Activos en Trámite de Cancelación a la Afiliación, Cancelados en la Afiliación, Cancelados de Matrículas, Fallecidos, Jubilados, etc): por circuito, profesión, colegio, edad, sexo, departamento, distrito electoral, categoría, escala de aportes por mes.
- Cantidad de afiliados titulares y familiares a la obra social por sexo, estado en obra Social, plan, con derecho y sin derecho por mes.
- Cantidad de afiliados beneficiarios por tipo de beneficio, sexo, profesión, colegio, edad al otorgamiento del beneficio por mes.
- Cantidad de beneficios otorgados por tipo de beneficio y mes.
- Cantidad de altas y bajas de beneficiarios por tipo de beneficio y mes.
- Total de deuda previsional y de obra social en Módulo Previsional de Aportes (MPA) por afiliado, circuito y colegio por mes.
- Total de ingresos mensuales por afiliado y mes.
- Cantidad de pagos de chequeras por mes.
- Total de deuda de contrato y cantidad de cuotas adeudadas por afiliado, colegio y circuito por mes.
- Promedio de módulos de beneficios de jubilación ordinaria, pensión, retiro voluntario y jubilación por incapacidad por mes.
- Cantidad de contratos por tipo de contrato que están al día y atrasados por mes (deuda vencida y exigible).
- Total de ingresos según los distintos rubros de deuda por mes.
- Total de deuda generada por mes y actividad.
- Proporción de activos/pasivos (beneficiarios) por mes y colegio.
- Entre otros.

En síntesis, los indicadores identificados son:

- Cantidad de afiliados titulares
- Cantidad de afiliados a la Obra Social
- Cantidad de beneficiarios
- Cantidad de beneficios
- Total de altas de beneficiarios
- Total de bajas de beneficiarios
- Monto total de deuda previsional en pesos
- Monto total de deuda previsional en MPA (Módulos Previsional de Aportes)
- Monto total de ingreso en pesos
- Cantidad de chequeras pagas
- Monto total de deuda de Obra Social en pesos
- Promedio de módulos de beneficios
- Cantidad total de contratos
- Cantidad de contratos al día
- Cantidad de contratos al atrasados
- Monto total de deuda generada en pesos
- Monto total de deuda generada en MPA
- Proporción Activos/Pasivos.

Y las perspectivas de análisis identificadas son:

- Estado previsión, Grupo Actividad, Actividad
- Provincia, Departamento, Distrito electoral, Localidad, Circuito
- Tipo Afiliado
- Colegio, Profesión, Especialidad

- Sexo
- Categoría, Escala de Aporte
- Edad
- Año, Semestre, Trimestre, Mes
- Estado Obra Social
- Plan, Sub-Plan
- Derecho Obra Social
- Tipo Beneficio
- Tipo Contrato
- Afiliado
- Rubro Deuda

Muchas veces resulta muy representativo generar un modelo conceptual a partir de los identificadores y perspectivas obtenidas que describa en alto nivel la estructura del DW. La información es representada a través de objetos, relaciones y atributos. Sin poseer demasiados conocimientos previos, permite comprender cuáles serán los resultados que se obtendrán, las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos. A la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, estos se ubican a la derecha del esquema (figura 3.1).

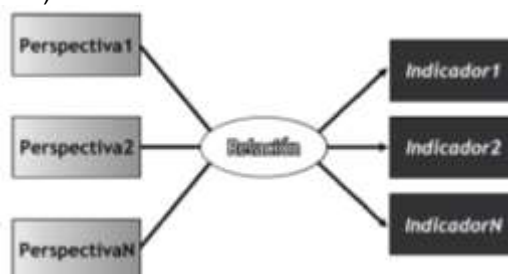


Figura 3.1: Ejemplo de Modelado Conceptual

En este trabajo no se desarrolla el modelado conceptual del caso de estudio pero es conveniente tenerlo en cuenta para trabajos similares. Puede ser muy útil en el momento de realizar el análisis de los requerimientos ya consigue aclarar los conceptos cuando se analiza y explora un nuevo negocio.

3.2.2 Análisis de los OLTP

El objetivo de este paso es examinar las fuentes de datos OLTP disponibles para determinar cómo serán calculados los indicadores y establecer las respectivas correspondencias entre el modelo conceptual y las fuentes de datos. Una vez determinada las relaciones con los OLTP, se deberán seleccionar los campos que contendrá cada perspectiva, ya que a través de ellos se examinarán y filtrarán los indicadores. Es muy importante conocer en detalle que significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc. Al momento de seleccionar los campos que integrarán cada perspectiva, debe prestarse mucha atención, ya que esta acción determinará la granularidad de la información encontrada en el DW.

3.2.2.1 Aplicación Etapa 2 al Caso de Estudio

Se trabaja con la fuente de datos que utilizan los sistemas operacionales de la CPAC. Las tareas de los distintos departamentos de la institución se informatizan a través de un sistema que es fundamental para llevar a cabo sus tareas diarias. Este sistema, como se explicó con anterioridad, fue adquirido por la organización en el año 1998 y actualmente permite dar soporte a todas las tareas diarias. Es un sistema informático que a lo largo del tiempo ha tenido que ir adaptándose y ha sufrido algunas migraciones de datos para no quedar en desuso frente a los cambios tecnológicos. Hoy en día, brinda asistencia a distintos departamentos y sectores de la institución: Secretaría, Fiscalización y

Control de Aportes, Facturación de Obra Social, Bancos, Tesorería, Estampillado y Otros Recursos, Caja, etc.

El sistema de gestión de base de datos utilizado en la institución es Microsoft SQL-Server 2008. Originalmente se desarrolló en la versión de SQL Server 6.5 y luego se fue migrando hasta alcanzar la versión actual.

La característica más desfavorable del diseño de este origen de datos es que **ninguna de las tablas está desarrollada utilizando claves foráneas**⁵, solo contienen claves primarias. Esta característica produce muchas inconsistencias en los datos ya que el motor no controla la consistencia de los mismos. Los desarrolladores que codificaron el sistema en el año 1998, optaron por este tipo de diseño para facilitar la migración de datos que existían en archivos que se generaban de utilizar un sistema bajo el lenguaje de datos RPG⁶. El sistema desarrollado realizaba los controles de información a través de la programación misma.

Con el correr del tiempo, los desarrolladores fueron cambiando. Algunos se dedicaron a otros proyectos, otros cambiaron de empresa y se fue incorporando nuevo personal. Por este motivo de rotación continua de personas es que hay mucha duplicación de información e inconsistencia en los datos.

Las necesidades de mayor información se relacionan fundamentalmente a conocer el universo de afiliados. Cuando se quieren clasificar por distintas perspectivas, al tener tanta duplicación de información es difícil obtener resultados consistentes. Una muestra de este problema se puede visualizar en la figura 3.2. Debido a que el sistema fue desarrollado a través de módulos independientes es que hay 3 entidades o tablas donde se almacenan los afiliados a la CPAC: una de ellas denominada *afiliados_titulares* donde se guardan los matriculados en los distintos colegios, otra denominada *afiliados_obra_social* donde se almacena nuevamente el afiliado matriculado (titular) más todo su grupo familiar (hijos, cónyuges, padres, hermanos y/o personas a cargo), y finalmente la tercera denominada *afiliados_beneficiarios* donde está por tercera vez el afiliado titular (matriculado) y/o algunos de los familiares si tienen beneficio de pensión y/o de jubilación. Observando la estructura de las mismas es simple identificar la **duplicidad de información** y la **inconsistencia** ya que muchas veces hay procesos que actualizan alguna entidad dejando desactualizada alguna otra. Muchas veces para evitar en cierta forma estas inconsistencias de información se utilizan Triggers (Store Procedures) definidos en la base de datos para que ante la actualización de algún dato en alguna entidad dispare automáticamente la modificación en todos los lugares donde se encuentra el mismo dato.

⁵ En el contexto de bases de datos relacionales, una **clave foránea** o **clave ajena** (o Foreign Key FK) es una limitación referencial entre dos tablas. La clave foránea identifica una columna o grupo de columnas en una tabla (tabla hija o referendo) que se refiere a una columna o grupo de columnas en otra tabla (tabla maestra o referenciada). Las columnas en la tabla referendo deben ser la clave primaria u otra clave candidata en la tabla referenciada. Los valores en una fila de las columnas referendo deben existir solo en una fila en la tabla referenciada. Así, una fila en la tabla referendo no puede contener valores que no existen en la tabla referenciada. De esta forma, las referencias pueden ser creadas para vincular o relacionar información. Esto es una parte esencial de la normalización de base de datos. Múltiples filas en la tabla referendo pueden hacer referencia, vincularse o relacionarse a la misma fila en la tabla referenciada. Mayormente esto se ve reflejado en una relación uno (tabla maestra o referenciada) a muchos (tabla hija o referendo).

⁶ **RPG: Report Program Operator** fue introducido en 1960 como un lenguaje para duplicar rápidamente el enfoque de proceso utilizado con un equipo de tarjeta perforada. Este lenguaje fue desarrollado por IBM en 1964. A pesar de las aplicaciones de actualización de archivos, el RPG es un lenguaje de propósito limitado porque los programas objeto generados por el compilador de RPG siguen un ciclo de procesamiento básico. Una ventaja del RPG es la relativa facilidad para aprenderlo y usarlo. Dado que la lógica de la programación es fija, existen menos reglas formales que en otros lenguajes.

afiliados_titulares	Type
tipo_documento	char(1)
nro_documento	char(15)
apellido	varchar(25)
nombre	varchar(30)
apellido_materno	varchar(25)
fecha_nacimiento	datetime2
sexo	char(1)
nacionalidad	char(1)
estado_civil	char(1)
apellido_conyuge	varchar(25)
nombre_conyuge	varchar(30)
cuit	char(11)
nro_aporte prevision	char(6)
tipo_afiliacion	char(1)
categoria	char(2)
nro_dom_legal	char(2)
domicilio_legal	varchar(40)
piso_legal	char(3)
depto_legal	char(3)
monoblock_legal	char(3)
escalera_legal	char(3)
barrio_o_manzana_legal	varchar(10)
localidad_legal	char(5)
zona_dom_legal	char(2)
obser_domicilio_legal	varchar(60)
nro_dom_personal	char(2)
domicilio_personal	varchar(40)
piso_personal	char(3)
depto_personal	char(3)
monoblock_personal	char(3)

afiliados_beneficiarios	Type
tipo_doc_benef	char(1)
nro_doc_benef	char(15)
tipo_doc_titular	char(1)
nro_doc_titular	char(15)
apellido_benef	varchar(25)
nombre_benef	varchar(30)
fecha_nacimiento	datetime2
domicilio	varchar(40)
monoblock	char(3)
piso	char(3)
escalera	char(3)
depto	char(3)
barrio_o_manzana	varchar(10)
localidad	char(5)
telefono_personal	varchar(25)
fecha_ult_ctrol_sup	datetime2
tipo_doc_tit_cobro	char(1)
nro_doc_tit_cobro	char(15)
tipo_doc_apoderado	char(1)
nro_doc_apoderado	varchar(15)
banco	char(3)
sucursal	char(3)
localidad_sucur	char(5)
porcentaje_benef	numeric(5,2)
estado_benef	char(1)
observaciones	varchar(500)
operador_alta	varchar(10)
fecha_alta	datetime2
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2

afiliados_obra_social	Type
tipo_documento	char(1)
nro_documento	char(15)
apellido	varchar(25)
nombre	varchar(30)
tipo_documento_titular	char(1)
nro_documento_titular	char(15)
tipo_doc_tit_pago	char(1)
nro_doc_tit_pago	char(15)
grupo_familiar	char(2)
parentesco	char(2)
fecha_afil_ospac	datetime2
plan_os	char(1)
fecha_ingreso_plan	datetime2
carencia_plan	char(1)
fecha_inicio_carencia	datetime2
fecha_fin_carencia	datetime2
estado_afiliado_obra_social	char(1)
fecha_nacimiento	datetime2
sexo	char(1)
estudia	char(1)
incapacitado	char(1)
motivo_fallecimiento	char(6)
fecha_fallecimiento	datetime2
autorizan_recetas_excepcion	char(1)
otra_ob_soc1	varchar(5)
nro_otra_os_1	varchar(15)
otra_ob_soc2	varchar(5)
nro_otra_os_2	varchar(15)
operador_alta	varchar(10)
fecha_alta	datetime2

Figura 3.2: Duplicidad de información e inconsistencia en el origen de datos

Debido a este grave error de diseño y dependiendo de dónde se obtenga la información, los reportes gerenciales muchas veces contienen divergencias, según la persona que lo desarrolló o el momento en que se confeccionó provocando falta de credibilidad e incertidumbre en la toma de decisión. La gran mayoría de los reportes que se generan a diario contienen un alto grado de complejidad y dependen en gran medida de las capacidades técnicas de la persona que lo realice.

Seguidamente, en la figura 3.3, se visualizan algunas características generales de esta base de datos operacional. Una particularidad a resaltar es que posee nada más ni nada menos que 1100 tablas. Es una base de un tamaño aproximado de 160.000 MB.

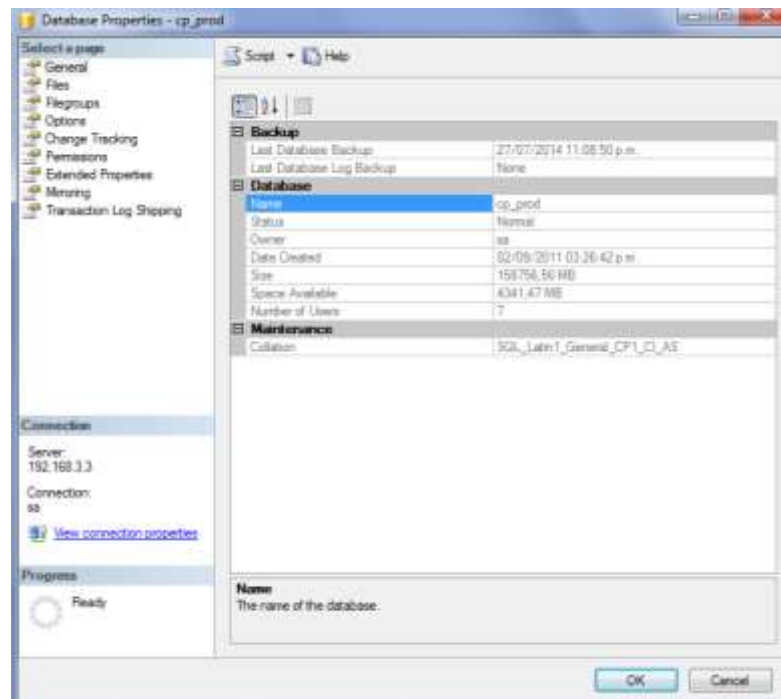


Figura 3.3: Características de la base de datos operacional cp_prod

A continuación se detallan algunas de las tablas de esta base de datos operacional que se utiliza como origen de datos para el armado del DW respondiendo a los requerimientos de información establecidos con anterioridad.

Tablas que almacenan datos de los distintos *tipos de afiliados* (figura 3.4):

Afiliados Titulares: Contiene cada uno de los profesionales matriculados adheridos al Arte de Curar. Posee todos los datos particulares que lo identifican, información de su domicilio, estado actual en previsión, estado en obra social, plan elegido, subplan, categoría, actividad, etc (a continuación se analiza más en detalle cada uno de estos datos).

Afiliados Obra Social: Almacena el afiliado titular con todo su grupo familiar. Contiene datos como la fecha a partir de la cual tiene la obra social, el plan seleccionado, subplan, el estado, si posee otra obra social para complementariedad, entre otros. Como explicamos con anterioridad ciertos datos de los afiliados titulares se repiten en las dos entidades y ocasionan inconsistencia en la presentación de informes de acuerdo de dónde y cómo se obtenga la información a mostrar.

Afiliados Beneficiarios: Posee la información de cada uno de los beneficiarios jubilados y pensionados con su tipo de beneficio, la fecha de otorgamiento del mismo, la cantidad de módulos que le corresponde según su beneficio, datos del domicilio, etc. En el caso de los pensionados tiene datos del afiliado titular que originó el beneficio de pensión. Como se detalló anteriormente, los datos del titular fallecido también se encuentra en las dos entidades expuestas con anterioridad.

Tablas que guardan relación con el *estado en previsión* del afiliado titular:

actividades	Type
actividad	char(2)
descripcion	char(40)
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2
fila	varchar(13)
grupo_actividad	char(2)
activo_en_prevision	char(1)

leyendas_estados_prevision	Type
estado_actual	char(2)
subestado_actual	char(2)
estado_futuro	char(2)
leyenda	varchar(80)
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2
fila	varchar(13)

grupos_actividad	Type
grupo_actividad	char(2)
descripcion	varchar(40)
operador_alta	varchar(10)
fecha_alta	datetime2
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2
fila	varchar(13)

Figura 3.4: Tablas de Estado Previsión

Actividad: Hay disponibles 99 actividades (el campo actividad es un texto con longitud de 2). Indica el estado actual detallado del afiliado. Establece si está activo y debe abonar previsión, si pertenece a algún centro de trabajo mediante al cuál realiza sus aportes jubilatorios, si está cancelado y que tipo de cancelación tiene, si está fallecido o tiene un beneficio de jubilación, etc. A su vez todas estas actividades se agrupan en lo que se denomina *grupo actividad* (cuadro 3.1).

Grupos Actividad: Establece un nivel de agrupamiento de las actividades. Esta forma de asociación permite simplificar la identificación de los grupos de afiliados (cuadro 3.2).

Actividad	Descripción	Grupo Actividad
01	Activo	01
02	Jubilado	07
03	Fallecido que Generó Pensión	07
04	Fallecido	06
05	Desafiliado	06
06	C.A.M.O.S.	02
09	Afiliación De Oficio	01
10	Cancelación De Matricula	06
11	Cancelación De Matricula por Jubilación	07
12	Afiliados Con Más De Un Numero	06
13	Afiliados Incisos A C.P.A.C.	08
15	Empleados De C.P.A.C.	08
20	Circulo De Prest. Farmacias Rosario	02
21	Asociación Medica De Alcorta	02
22	Asociación Medica Del Dpto. San Lorenzo	02
23	Circulo De Farmacias De Rafaela	02
24	Circulo Odontológico De Rafaela	02
25	Asociación Bioquímicos Dpto San Lorenzo	02
26	Asociación Medica De San Jorge	02

Cuadro 3.1: Tabla Actividad

Grupo Actividad	Descripción
01	Activos (01+09)
02	Intergremiales (06 + 20 a 65)
03	Cancelados con OS (66 a 77)
04	Automáticos (80 a 83)
05	Cancelados sin OS (93 a 98)
06	Otros (04+05+10+12)
07	Beneficios (02+03+11)
08	Empleados e Incisos (13+15)

Cuadro 3.2: Tabla Grupos_Actividad

Leyendas Estados Previsión: Si bien los afiliados titulares poseen, como explicamos con anterioridad una actividad que indica su situación previsional, existe otro agrupamiento que identifica su estado y situación previsional (cuadro 3.3). Se compone de un estado actual, un subestado y un estado futuro, (este último tiene significancia si tiene algún trámite en suspenso que falta efectivizar o dar resolución de directorio).

Estado	Subestado	Estado Futuro	Leyenda
AC			Activo
AC		CA	Activo En Tramite De Cancelación En La Afiliación
AC		JU	Activo
BA			Baja
CA	AF		Cancelado En La Afiliación
CA	MA		Cancelado En La Matricula
CA	MA	AC	Cancelado En La Matricula
CA	MA	CA	Cancelado En La Matricula
CA	MA	JU	Cancelado En La Matricula En Tramite De Jubilación
DE			Desafiliado
FA	GP		Fallecido Que Genero Pensión
FA	NG		Fallecido
JU			Jubilado
JU	EA		Jubilado Por Edad Avanzada
JU	IT		Retiro Invalidez
JU	OR		Jubilación Ordinaria
PT			En Tramite De Activación
PT		AC	En Tramite De Activación
PT		CA	En Tramite De Afiliación y Cancelación

Cuadro 3.3: Tabla Leyendas Estado Previsión

Tablas que guardan relación con la *ubicación física del domicilio* de cada afiliado (figura 3.5):

The figure shows four screenshots of database tables with their fields and data types:

- localidades:** localidad (char(5)), nombre (char(30)), nom_abrev (char(12)), provincia (char(2)), cod_postal (char(4)), pais (char(2)), departamento (char(2)), circunscripción (char(1)), circuito (char(2)), telediscado (char(4)), operador_ult_modif (varchar(10)), fecha_ult_modif (datetime2), fila (varchar(13)), tipo_medio_envio (char(1)), medio_envio (char(3)).
- departamentos:** departamento (char(2)), nombre (varchar(20)), operador_ult_modif (varchar(10)), fecha_ult_modif (datetime2), fila (varchar(13)), distrito_electoral (char(2)).
- distritos_electorales:** distrito_electoral (char(2)), nombre (varchar(50)), operador_ult_modif (varchar(10)), fecha_ult_modif (datetime2), fila (varchar(13)).
- circuitos:** circuito (char(2)), nombre (varchar(20)), operador_ult_modif (varchar(10)), fecha_ult_modif (datetime2), fila (varchar(13)).

Figura 3.5: Tablas que indican ubicación

Localidades: Esta entidad almacena las distintas localidades (tal como lo expresa su nombre) que guardan relación con el domicilio del afiliado, de su grupo familiar o beneficiarios jubilados y pensionados. Las localidades tienen un agrupamiento que define la zona en donde se ubica el domicilio. Este agrupamiento se denomina *circuito* (cuadro 3.4).

Circuito: Agrupa las localidades según la ubicación física. A su vez los circuitos se pueden agrupar en un nivel superior denominado *circunscripción*. Esta última se identifica con el primer

número del circuito (solo se tienen circunscripción 1 y 2 que igualan a los valores Norte y Sur). Los valores para los circuito son los siguientes:

- 1.1 – Santa Fe Capital
- 1.2 – Norte de Santa Fe
- 2.1 – Rosario
- 2.2 – Sur de Santa Fe
- 3.1 – Fuera de la Provincia
- 3.2 – Sin Domicilio

Departamentos: Las localidades corresponden a distintos departamentos. Es otra forma de agrupamiento que se utiliza en gran medida cuando se analiza el grupo de afiliados (cuadro 3.5).

Distritos Electorales: Cuando se deben elegir nuevas autoridades, los afiliados activos deben votar como forma democrática de selección. Para el armado de listas y el recuento de votos se tienen en cuenta distintos distritos electorales. Son un grupo de localidades que pertenecen a un circuito en particular pero que están subdivididas por distrito. Es otra forma de agrupamiento.

Localidad	Nombre	Provincia	Cod Postal	País	Dpto	Circ.	Circuito
22180	CARRIZALES	SF	2218	AR	IR		22
22181	CLARKE	SF	2218	AR	IR		22
22220	DIAZ	SF	2222	AR	SR		12
22400	CORONDA	SF	2240	AR	SR		12
22410	LARRECHEA	SF	2241	AR	SR		12
22420	AROCENA	SF	2242	AR	SR	1	12
22421	DESVIO ARIJON	SF	2242	AR	SR		12
23491	MONTE OSCURIDAD	SF	2349	AR	SC	1	12
23520	AMBROSETTI	SF	2352	AR	SC	1	12
23521	HERSILIA	SF	2352	AR	SC	1	12
23540	SELVA	SE	2354	AR	SD		31
23900	SAN PEDRO	SF	2390	AR	CP	1	12
24000	S.FRANCISCO	CD	2400	AR	SD	1	31
24010	CASTELAR	SF	2401	AR	SM	1	12
24030	BAUER Y SIGEL	SF	2403	AR	CT	1	12
24031	COLONIA JOSEFINA	SF	2403	AR	CT	1	12
24032	PLAZA JOSEFINA	SF	2403	AR	CT	1	12
24050	SANTA C.DE SAGUIER	SF	2405	AR	CT	1	12
24051	COLONIA CELLO	SF	2405	AR	CT	1	12
24070	CLUCELLAS	SF	2407	AR	CT		12
24071	EUSTOLIA	SF	2407	AR	CT	1	12
24072	PLAZA CLUCELLAS	SF	2407	AR	CT	1	12
24073	ESTACION CLUCELLAS	SF	2407	AR	CT	1	12
24080	ESTACION CLUCELLAS	SF	2407	AR	CT	1	12
24090	ZENON PEREYRA	SF	2409	AR	CT		12
24130	FREYRE	CD	2413	AR	SD		31
24150	PORTENA	CD	2415	AR	SD		31

Cuadro 3.4: Tabla Localidades

Departamento	Nombre	Distrito	Descripción
BE	BELGRANO	09	BELGRANO
CA	CASEROS	10	CASEROS
CN	CONSTITUCION	11	CONSTITUCION
CP	LA CAPITAL	03	LA CAPITAL
CT	CASTELLANOS	01	CASTELLANOS
GA	GARAY	07	SAN JUSTO, SAN JAVIER Y GARAY
GL	GRAL. LOPEZ	12	GRAL. LOPEZ
GO	GRAL.OBLIGADO	02	GRAL. OBLIGADO
IR	IRIONDO	13	IRIONDO
LC	LAS COLONIAS	04	LAS COLONIAS
NJ	9 DE JULIO	05	SAN CRISTOBAL, VERA Y 9 DE JULIO
RO	ROSARIO	14	ROSARIO
SC	SAN CRISTOBAL	05	SAN CRISTOBAL, VERA Y 9 DE JULIO
SJ	SAN JUSTO	07	SAN JUSTO, SAN JAVIER Y GARAY
SL	SAN LORENZO	15	SAN LORENZO
SM	SAN MARTIN	08	SAN MARTIN
SR	SAN JERONIMO	06	SAN JERONIMO
SV	SAN JAVIER	07	SAN JUSTO, SAN JAVIER Y GARAY
VE	VERA	05	SAN CRISTOBAL, VERA Y 9 DE JULIO

Cuadro 3.5: Tabla Departamentos y Distrito Electoral

Tablas que almacenan datos en cuanto a *profesiones* y *especialidad* (figura 3.6):

profesiones	Type
profesion	char(3)
descripcion	varchar(40)
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2
fila	varchar(13)
orden_impresion	int
descripcion_colegio_listado	varchar(20)
prescribe	char(1)
porcentaje	numeric(5,2)

especialidades	Type
profesion	char(3)
especialidad	char(2)
descripcion	varchar(40)
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2
fila	varchar(13)

colegios	Type
circunscripcion	char(1)
colegio	char(3)
descripcion	varchar(60)
domicilio	varchar(60)
localidad	char(5)
telefono	varchar(25)
fecha_ultimo_informe	datetime2
contacto	varchar(40)
operador_alta	varchar(10)
fecha_alta	datetime2
operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2
ult_fec_env_carta	datetime2
fila	varchar(13)
descripcion_colegio_circ	varchar(60)

Figura 3.6: Tablas de profesiones y especialidad

Colegios: Una vez finalizada su carrera de grado, los profesionales de la salud se matriculan en su colegio correspondiente. Este último asigna un número de libro, folio y matrícula que es fundamental para el ejercicio de la profesión. Los colegios de la provincia de Santa Fe son 14 en total que a su vez se dividen de acuerdo a la circunscripción que explicamos con anterioridad.

Profesiones: Los colegios de profesionales abarcan distintas profesiones. A modo de ejemplo se muestran algunos datos en el cuadro 3.6 como para entender este punto:

Profesión	Descripción	Colegio
BIO	BIOQUIMICO	05-BIOQUIMICOS
DBI	DR. EN BIOQUIMICA	05-BIOQUIMICOS
DEN	MÉCANICO DENTAL	02-ODONTOLOGOS
DIE	DIETISTA	12-DIETISTAS
DQU	DR. EN QUIMICA	05-BIOQUIMICOS
ENF	ENFERMERIA	01-MEDICOS
FAR	FARMACEUTICO	03-FARMACEUTICOS
FIS	FISIOTERAPEUTA	06-KINESIOLOGOS
FON	FONOAUDIOLOGO	07-FONOAUDIOLOGOS
HEM	TECNICO HEMOTERAPIA	05-BIOQUIMICOS
KIN	KINESIOLOGO	06-KINESIOLOGOS
LBI	LICENCIADO EN BIOQUIMICA	05-BIOQUIMICOS
LEN	LICENCIADO ENFERMERIA	01-MEDICOS
MED	MEDICO	01-MEDICOS

Cuadro 3.6: Tabla Profesiones

Especialidades: Dentro de las distintas profesiones de la salud hay una distinción de especialidades. Es una forma detallada de conocer al afiliado en cuanto a su desarrollo profesional (cuadro 3.7).

Profesión	Especialidad	Descripción
MED	01	ALERGIA E INMUNOLOGIA
MED	02	ANATOMIA PATOLOGICA
MED	03	ANESTESIOLOGIA
MED	04	ANGIOLOGIA
MED	05	CARDIOLOGIA
MED	06	CIRUG. VASCULAR PERIFERICA
MED	07	CIRUGIA GENERAL
MED	08	CIRUGIA INFANTIL
MED	09	CIRUG. PLASTICA O ESTETICA
MED	10	CIRUGIA TORAXICA
MED	11	CITODIAGNOSTICO
MED	12	CLINICA MEDICA
MED	13	DERMATOLOGIA
MED	14	ENDOCRINOLOGIA
MED	15	ENFERMEDADES DE LA NUTRICION
MED	16	FISIATRIA
MED	17	GASTROENTEROLOGIA
MED	18	GERIATRIA

MED	19	GINECOLOGIA
MED	20	HEMATOLOGIA
MED	21	HEMOTERAPIA
MED	22	NEUMONOLOGIA

Cuadro 3.7: Tabla Especialidades

Estado de los afiliados en *Obra Social* y su *categoría* en previsión (figura 3.7):

leyendas_estados_obra_social	Type	categorias	Type
estado_os	char(2)	categoria	char(2)
afiliado_adherente	char(1)	descripcion	varchar(40)
leyenda	varchar(80)	modulos	numeric(11,5)
operador_ult_modif	varchar(10)	operador_ult_modif	varchar(10)
fecha_ult_modif	datetime2	fecha_ult_modif	datetime2
fila	varchar(13)	modulos_reparto	numeric(11,5)
		modulos_capital	numeric(11,5)
		periodo_desde	char(6)
		fila	varchar(13)

Figura 3.7: Tablas del estado de los afiliados en OS y Categoría en Previsión

Leyendas Estado Obra Social: Contiene los distintos estados de los afiliados en obra social. Indican si el afiliado se encuentra activo, cancelado, desafiliado, etc. En base a este estado del afiliado y de todo su grupo familiar, se determina el importe a abonar por los servicios sociales en base al plan elegido.

Categorías: Los aportes personales se efectuarán en concepto de cuota de inscripción (deberá abonarse dentro de los 30 días a contar desde la fecha de afiliación) y de aportes mensuales (según categoría). Las categorías son 5 que se determinan de acuerdo a la edad del aportante y de sus años de aporte (cuadro 3.8). Cada categoría establece la cantidad de *Módulos Previsionales de Aportes (MPA)* obligatorios. Esta última es una unidad se determina por resolución del directorio teniendo en cuenta la realidad económica de los aportantes y beneficiarios, en forma periódica. En la actualidad el MPA está establecido en \$60. Por ello un afiliado en categoría C le corresponde abonar mensualmente 18 MPA en concepto de aporte jubilatorio, es decir \$1080.

Categoría	Aporte Mensual
A	3 MPA
B	12 MPA
C	18 MPA
D	24 MPA
E	30 MPA

Cuadro 3.8: Categorías de aportes y MPA

La categoría obligatoria mínima que aportarán mensualmente los afiliados se determina de acuerdo a la clasificación expuesta en el cuadro 3.9.

Categoría Para los afiliados ...	
A	- En sus 5 primeros años de antigüedad y que tengan menos de 27 años de edad. - Con más de 35 años de antigüedad y que tengan más de 80 años de edad.
B	- En sus primeros 9 años de antigüedad y que tengan menos de 31 años de edad (no incluido en A) - Con más de 35 años de antigüedad y que tengan más de 69 años de edad (no incluido en A)
C	- En sus primeros 13 años de antigüedad y que tengan menos de 35 años de edad (no incluido en A y B) - Con más de 35 años de antigüedad y que tengan más de 64 años de edad (no incluido en A y B)
D	- En sus primeros 18 años de antigüedad y que tengan menos de 40 años de edad (no incluido en A,B y C) - Con más de 30 años de antigüedad y que tengan más de 59 años de edad (no incluido en A,B y C)
E	- En sus primeros 30 años de antigüedad y que tengan menos de 60 años de edad (no incluido en A,B,C y D) - No encuadrados en las categorías precedentes.

Cuadro 3.9: Categorías de aportes y MPA

Contratos de afiliados (figura 3.8):

Figura 3.8: Tablas para Contratos

Contratos Afiliados: Aquellos afiliados que no depositen los aportes personales mensuales de cualquier naturaleza, cuotas, contribuciones y todo otro crédito que le corresponda a la Caja incurren en mora. En base a esta situación, se puede obtener una financiación (contrato) para estar al día con las distintas obligaciones. La base de datos almacena información del afiliado que lo solicitó, la fecha en que se llevó a cabo, la fecha de aprobación, el monto financiado, la cantidad de cuotas elegida, la fecha del primer vencimiento, entre otras. Por otro lado tienen posibilidad de solicitar préstamos de dinero en efectivo con distintos fines. De igual forma se abonan mensualmente.

Tipos de Contratos: Como desarrollamos en el punto anterior, hay distintos tipos de contratos: pueden ser por deuda refinanciada, por algún subsidio recibido para llevar a cabo alguna prestación de obra social, por viajes de turismo, prestamos en efectivo, etc.

Estados Contratos: Los contratos tienen distintos estados: por ejemplo Pendiente, Autorizado, Refinanciado, Intimado, En Proceso Judicial, entre otros.

Cuotas a Pagar: Cuando el contrato se autoriza se genera una serie de cuotas (de acuerdo a la financiación elegida). Se almacenan para cada cuota la fecha de vencimiento, el importe en módulos que es diferente al MPA visto con anterioridad. Se trata de un *Módulo Previsional de Contrato de Financiación (MPCF)*. Este último a diferencia del anterior tiene un componente de interés de acuerdo a la fecha de vencimiento de cada cuota.

Cuotas a Pagar Aplicación: Cada una de las cuotas del contrato tienen diversas componentes. Tienen parte de Capital (Componente CAP), Obra Social (Componente COS), Gastos de Generación, Gastos de Envío, etc.

Deuda y cobranza (figura 3.9):

Tabla	Columna	Tipo
cobranza_x_banco	nro_compbte_recaud	char(8)
	fecha_deposito	datetime2
	tipo_cmpbte_pago	char(2)
	nro_cmpbte_pago	char(18)
	tipo_ingreso	char(1)
	importe_abonado	numeric(11,2)
	moneda	char(5)
	procesado	char(1)
	nro_orden_ingreso	char(4)
	tipo_depositante	char(2)
	concepto	char(3)
	tipo_documento	char(1)
	nro_documento	char(15)
	nro_afiliado_rpg	char(7)
	agencia	char(2)
	centro_trabajo	char(3)
	entidad	char(3)
	depositante	varchar(255)
	operador_alta	varchar(10)
	fecha_alta	datetime2
	operador_ult_modif	varchar(10)
	fecha_ult_modif	datetime2
	fila	varchar(13)
periodo_abonado	char(6)	
mesa_entrada_contr	char(2)	
tipo_contrato	char(2)	
Deuda_afiliado	tipo_documento	char(1)
	nro_documento	char(15)
	periodo	char(6)
	nro_afiliado_rpg	char(7)
	origen_deuda	char(2)
	nro_intimacion	varchar(8)
	importe_total	numeric(11,2)
	saldo_total	numeric(11,2)
	estado_deuda	char(2)
	saldo_p_cc	numeric(11,2)
	saldo_p_cr	numeric(11,2)
	saldo_p_si	numeric(11,2)
	saldo_p_sc	numeric(11,2)
	saldo_p_sr	numeric(11,2)
	saldo_o_cu	numeric(11,2)
	saldo_o_ss	numeric(11,2)
	saldo_o_tr	numeric(11,2)
	saldo_s_se	numeric(11,2)
	saldo_a_sd	numeric(11,2)
	operador_alta	varchar(10)
	fecha_alta	datetime2
	operador_ult_modif	varchar(10)
	fecha_ult_modif	datetime2
fila	varchar(13)	
movim_del_dia_ro	numeric(11,2)	
movim_del_dia_sf	numeric(11,2)	
deuda_generada	tipo_documento	char(1)
	nro_documento	char(15)
	periodo	char(6)
	categoria	char(2)
	actividad	char(2)
	medio_cobro	char(2)
	nro_afiliado_rpg	char(7)
	origen_deuda	char(4)
	d1_subsidio	numeric(11,2)
	d2_obra_social	numeric(11,2)
	d3_prevision	numeric(11,2)
	d3_reparto	numeric(11,2)
	d3_capital	numeric(11,2)
	d4_aguinaldo	numeric(11,2)
	d4_reparto	numeric(11,2)
	d4_capital	numeric(11,2)
	d5_transplante	numeric(11,2)
sub_extraordinario	numeric(11,2)	
sub_sepelio	numeric(11,2)	
sub_invalidez	numeric(11,2)	
profesion	char(3)	
operador_alta	varchar(10)	
fecha_alta	datetime2	
tipo_medio_envio	char(1)	

Figura 3.9: Tablas para deuda y cobranza

Deuda Generada: Todos los meses se generan las chequeras que los afiliados activos (que no tienen cancelación de matrícula o de afiliación) deben abonar. Esta chequera tiene distintas componentes: hay partes que corresponden a previsión, a obra social, subsidios, etc. Además hay diversos medios de envío y de cobro (de acuerdo a la situación del afiliado), entre otros datos.

Cobranza por Banco: Diariamente los cajeros de la institución realizan cobranzas de chequeras, contratos, estampillas, prestaciones de obra social, etc. Por otro lado hay una serie de procesos que permiten extraer los pagos realizados en las distintas entidades bancarias. Esta información se almacena identificando el origen de la cobranza, el periodo que se está abonando, la fecha que se realizó, los datos del contrato si corresponde, entre otros.

Deuda Afiliado: Si el afiliado no abona la totalidad de los módulos generados luego se convierte en deuda que el afiliado va acumulando. Cuando se decide pagar o elaborar un contrato como forma de pago, la misma se valoriza teniendo en cuenta el interés producido entre la fecha de generación y la fecha de pago.

3.2.3 Modelo Lógico del Data Warehouse

A la hora de diseñar la capa de metadatos que da origen a la solución hay que hacer hincapié e identificar aquellos conceptos y elementos utilizados en los informes que se extraen a través de la aplicación. Estos factores se deben adaptar a la estructura del Data Warehouse para que el usuario final sea capaz de generar los mismos informes que se estaban generando anteriormente.

El armado dimensional se puede llevar a cabo considerando sistemáticamente cuatro pasos en un orden determinado, [18].

1 - Seleccionar el proceso de negocio para modelar. Un proceso de negocio es un conjunto de actividades atómicas o compuestas (subprocesos) que definen la forma de trabajo de la organización. Puede verse como un conjunto de actividades ordenadas e interrelacionadas, que toman uno o más entradas y crean salida que es de valor al cliente, las cuales son realizadas con el propósito de alcanzar una meta organizacional de negocio, [34]. Escuchar a los usuarios es el medio más eficiente para la selección del proceso de negocio. Es importante recordar que cuando se habla de procesos de negocio no se hace referencia a un departamento de negocio específico de la organización. Por ejemplo, sería conveniente construir un solo modelo dimensional para manejar el grupo de afiliados en lugar de construir modelos separados para los departamentos de secretaría, de servicios sociales y de beneficios. Al centrarse en los procesos de negocio, en lugar de los departamentos, se puede entregar información consistente de forma más económica en toda la organización. Si se establecen modelos dimensionales departamentales, inevitablemente es muy probable que se dupliquen datos con diferentes etiquetas y terminología. Datos múltiples desembocan

en modelos dimensionales separados lo que será vulnerable a inconsistencias en los datos. La mejor manera de garantizar la coherencia es la publicación de los datos una vez. Esto también reduce el esfuerzo de desarrollo de los ETL (Extracción, Transformación y Carga), así como la carga de la gestión de datos y almacenamiento.

2. Declarar el grano de los procesos de negocio. Declarar el grano significa especificar exactamente lo que representa una fila en cada tabla de hecho. El grano transmite el nivel de detalle asociado con las mediciones. Es extremadamente importante que todos en el equipo de diseño estén de acuerdo sobre la granularidad elegida. Por otro lado, una declaración de grano inapropiado podría traer problemas futuros en la obtención de información. Declarar que el grano es un paso crítico que no puede tomarse a la ligera.

3. Escoger las dimensiones que se aplicarán a cada fila de la tabla de hechos. Dimensiones entran en la cuestión de cómo los gerentes/usuarios describen los datos que resultan del proceso de negocio. Se debe seleccionar un sólido conjunto de dimensiones que representan todas las descripciones posibles que tienen sobre los valores individuales en el contexto de cada medición. Si tenemos claro el grano, entonces las dimensiones generalmente se pueden identificar con bastante facilidad. Preferiblemente se deben desarrollar modelos dimensionales para la información más atómica capturada por un proceso de negocio. Datos Atómicos es la información más detallada recopilada; estos datos no se pueden subdividir. La dimensión fecha siempre existe en los DW para expresar una serie temporal. De hecho, la fecha suele ser la primera dimensión de la base de datos. A diferencia de la mayoría de las otras dimensiones, podemos construir la tabla de dimensión fecha con anticipación. Podemos poner entre 5 o 10 años de filas que representen la historia que hemos almacenado, así como de varios años en el futuro

4. Identificar los hechos numéricos que poblarán cada fila de las tablas de hechos. Se determinan a través de la pregunta "¿Qué estamos midiendo?" Los usuarios empresariales están muy interesados en el análisis de estas medidas. Todos los hechos candidatos guardan relación con el grano definido en el paso 2. Aquellos hechos que pertenecen claramente a un grano diferente deben estar en una tabla de hechos separados. Está claro que tenemos que tener en cuenta tanto las necesidades de nuestros usuarios de negocios como las realidades de nuestros datos de origen en conjunto para tomar decisiones con respecto a los cuatro pasos, como se ilustra en la Figura 3.10.

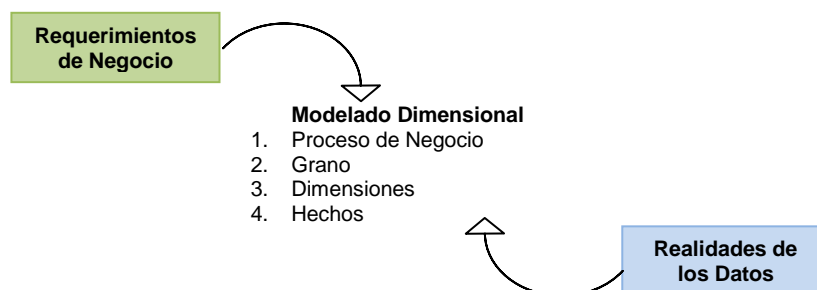


Figura 3.10: Cuatro pasos del modelado dimensional

Seguidamente se exponen los diseños de los modelos tipo estrellas siguiendo la metodología de Kimball para cumplir con todos los requerimientos de información planteados. Se compone de tablas de hechos (cuyos nombres comienzan con *FT: Fact Table*) y dimensiones (comienzan con *DT: Dimensional Table*) que marcarán la pauta para definir los cubos. Se presentan de manera separada por cuestiones físicas de espacio, pero las dimensiones son compartidas por los diversos esquemas.

El primer modelado dimensional refiere a los afiliados titulares de la CPAC (figura 3.11) y se definen las siguientes tablas:

De Dimensiones:

- *DT_Tiempo:* es la dimensión tiempo y se utiliza en todos los esquemas para diferentes tipos de fechas. Su ETL consiste, lógicamente, en generar una serie de fechas y otros campos que se asocien con los diferentes niveles de las jerarquías (Año, Trimestre, Mes...etc.). En este caso en particular y de acuerdo a los requerimientos, se eligió un intervalo de tiempo mensual.

- *DT_Estado_Prevision*: es la dimensión que agrupa los datos de previsión, con su actividad y grupo actividad.
- *DT_Distribucion_Geografica*: dimensión que identifica la zona del afiliado (como la provincia, departamento, distrito electoral, localidad, circuito).
- *DT_Colegios_Profesion*: dimensión que abarca datos del colegio, profesión y especialidad.
- *DT_Edad*: criterio de análisis que permite discriminar a los afiliados por edad.
- *DT_Sexo*: dimensión compartida por varios modelos que permite diferenciar a los afiliados por sexo.
- *DT_Categoria*: criterio de análisis que permite diferenciar a los afiliados por categoría y escala de aportes.

De Hecho:

- *FT_Afiliados_Titulares*: contiene los indicadores *cantidad_titulares* y *MPA_promedio* explorados a través de las dimensiones detalladas con anterioridad.

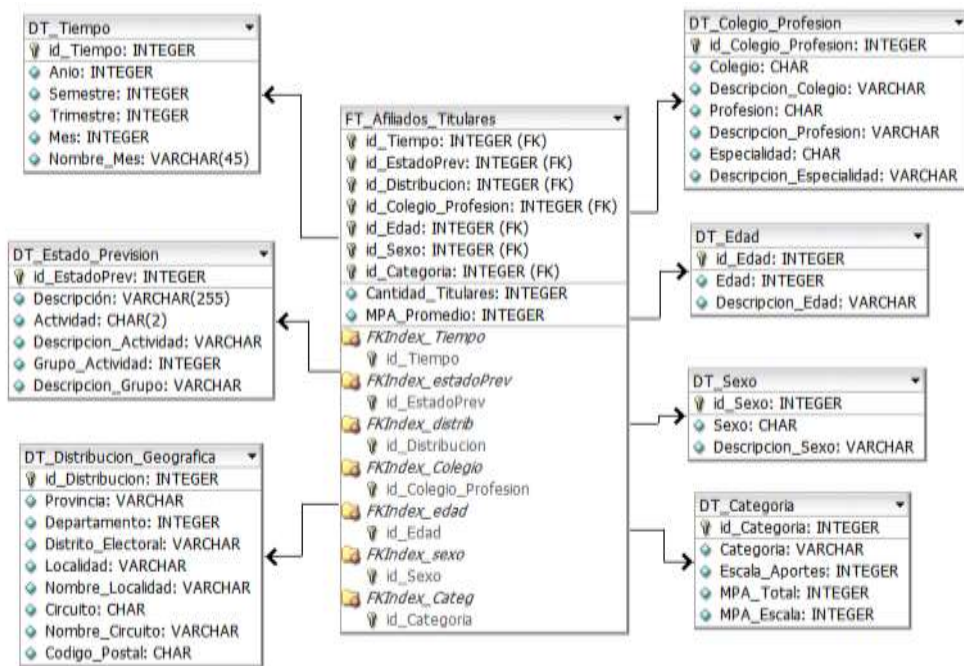


Figura 3.11: Modelado Dimensional. Fact Table: Afiliados_Titulares

La figura 3.12 representa el modelado dimensional que permite recuperar los afiliados a la obra social a través de diversas dimensiones:

- *DT_Tiempo*: ídem anterior.
- *DT_Tipo_Afiliado*: dimensión que permite diferenciar el tipo de afiliado titular de los familiares directos y a cargo.
- *DT_Estado_Obra_Social*: dimensión que permite analizar los afiliados a la obra social a través de los distintos estados de OS.
- *DT_Derecho_Os*: permite diferenciar a los afiliados con derecho de los sin derecho.
- *DT_Colegio_Profesion*: ídem anterior
- *DT_Sexo*: ídem anterior
- *DT_Plan_Obra_Social*: discrimina a los afiliados por tipo de plan.

De Hecho:

- *FT_Afiliados_Obra_Social*: contiene los indicadores *cantidad_titulares*, *cantidad_familiares* y *Total_Tarifas_OS* explorados a través de las dimensiones detalladas con anterioridad.

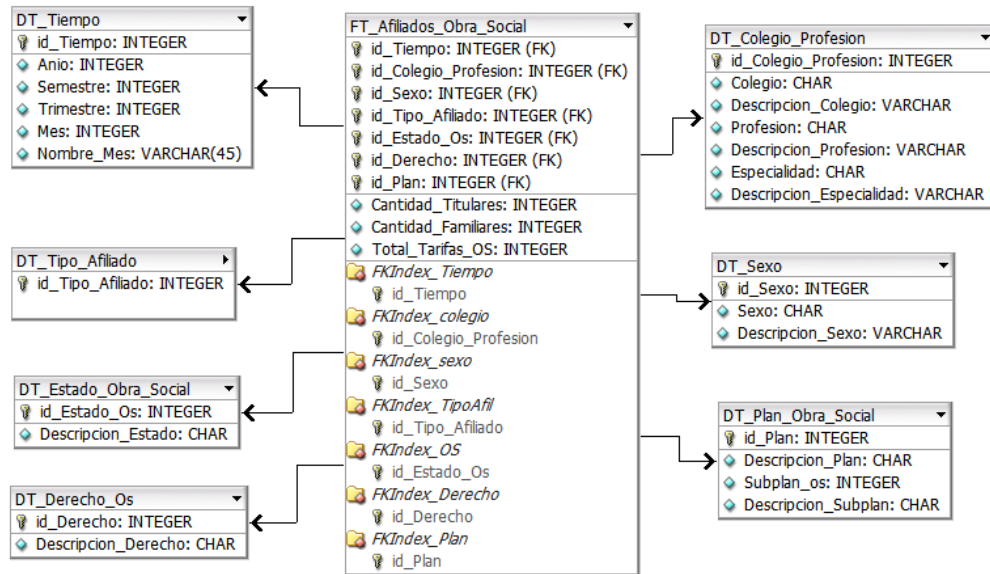


Figura 3.12: Modelado Dimensional. Fact Table: Afiliados_Obra_Social

La figura 3.13 representa el modelado dimensional que permite recuperar los afiliados beneficiarios (jubilados y pensionados) a través de diversas dimensiones:

- *DT_Tiempo*: ídem anterior.
- *DT_Tipo_Beneficio*: permite diferenciar los diferentes tipos de jubilaciones y pensiones.
- *DT_Distribucion_Geografica*: Idem anterior
- *DT_Colegio_Profesion*: Idem anterior
- *DT_Edad*: Idem anterior
- *DT_Sexo*: Idem anterior

De Hecho:

FT_Afiliados_Beneficiarios: contiene los indicadores *cantidad_titulares* y *MPA_promedio* explorados a través de las dimensiones detalladas con anterioridad

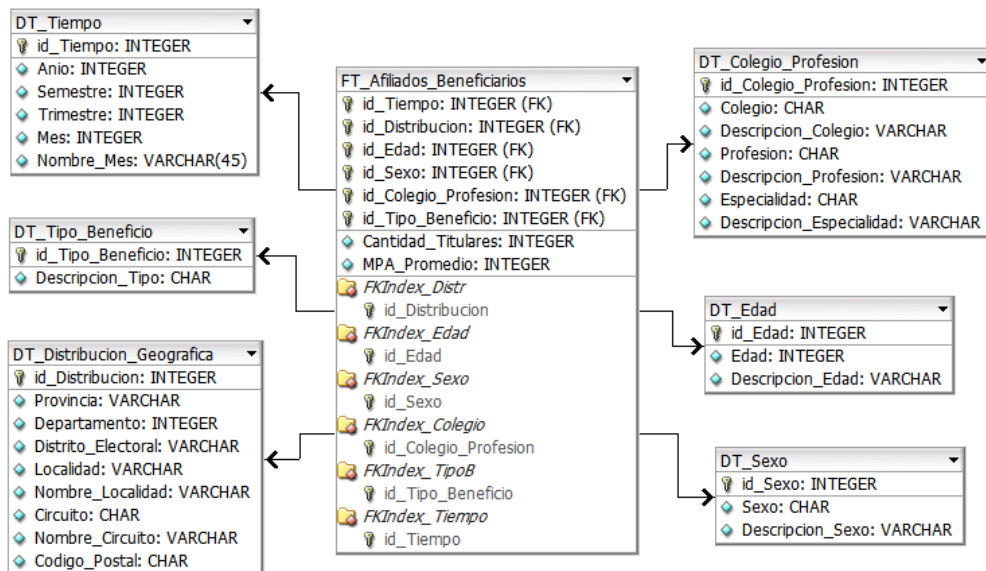


Figura 3.13: Modelado Dimensional. Fact Table: Afiliados_Beneficiarios

La figura 3.14 determina un modelado dimensional que permite discriminar los beneficios por tiempo y tipo de beneficio. Cabe destacar que el beneficio de pensión es único en un grupo familiar

mientras que los miembros o beneficiarios son cada una de las personas que conforman el grupo. En el caso de las jubilaciones no hay diferencia entre beneficios y beneficiarios ya que es un solo titular.

De Dimensiones:

- DT_Tiempo: ídem anterior.
- DT_Tipo_Beneficio: ídem anterior.

De Hecho:

- FT_Beneficios: contiene los indicadores *cantidad_altas*, *cantidad_bajas* y *promedio_MPB* explorados a través de las dimensiones detalladas con anterioridad

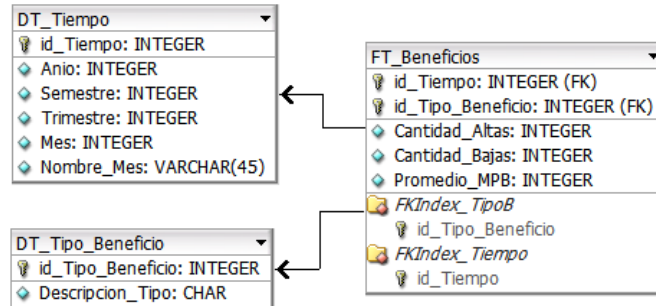


Figura 3.14: Modelado Dimensional. Fact Table: Beneficios

La figura 3.15 determina el modelado dimensional que permite discriminar los deudores de previsión, obra social y contratos a través de diversas dimensiones:

- DT_Tiempo: ídem anterior.
- DT_Distribucion_Geografica: ídem anterior.
- DT_Afiliado: es la dimensión que identifica al afiliado deudor. Esta dimensión admite una granularidad muy baja permitiendo mucho más detalle que los anteriores.
- DT_Colegio_Profesion: ídem anterior
- DT_Categoria: ídem anterior

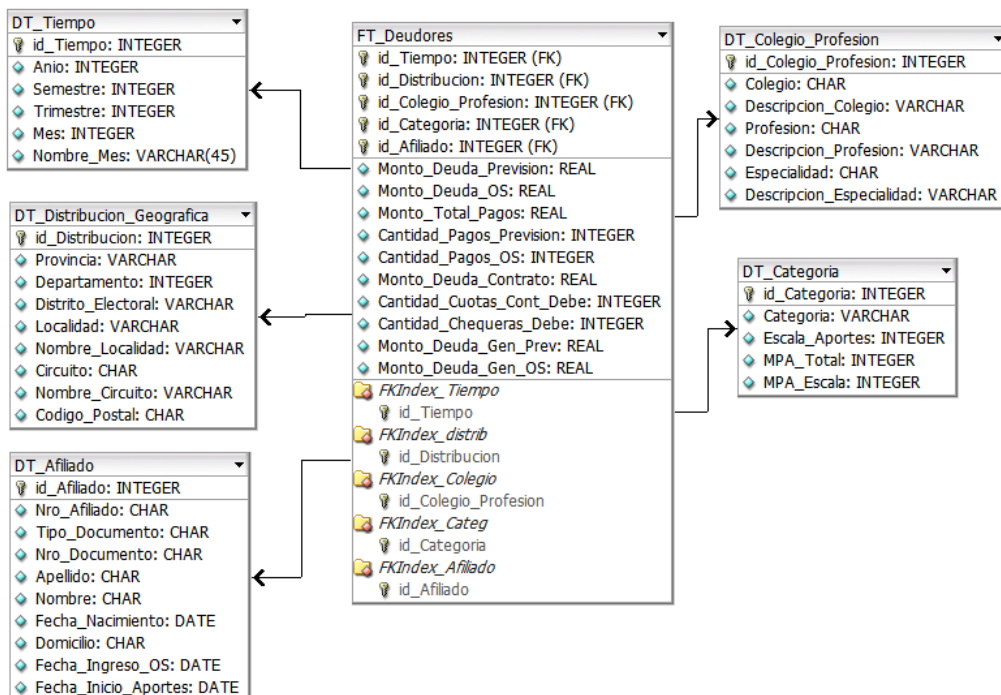


Figura 3.15: Modelado Dimensional. Fact Table: Deudores

De Hecho:

- *FT_Deudores*: contiene los indicadores *monto_deuda_prevision*, *monto_deuda_os*, *monto_total_pagos*, *cantidad_pagos_prevision*, *cantidad_pagos_os*, *monto_deuda_contrato*, *cantidad_cuotas_cont_debe*, *cantidad_chequeras_debe*, *monto_deuda_gen_prev* y *monto_deuda_gen_os*.

La figura 3.16 determina un modelado dimensional que permite diferenciar los contratos por tiempo, tipo de contrato y estado. Son distintas formas de financiar las deudas de previsión y obra social aunque también hay contratos por préstamos de turismo y de dinero en efectivo.

Las dimensiones definidas son:

- *DT_Tiempo*: ídem anterior.
- *DT_Tipo_Contrato*: identifica los distintos tipos de contratos.
- *DT_Estado_Contrato*: indica si el contrato está activo, finalizado, anulado, etc.

De Hecho:

- *FT_Contrato*: contiene indicadores que permite determinar la cantidad de contratos al día y los atrasados: *cantidad_al_dia*, *cantidad_atrasados*, *monto_deuda_vencida* y *monto_deuda_exigible*.



Figura 3.16: Modelado Dimensional. Fact Table: Contrato

Finalmente la figura 3.17 presenta el último esquema desarrollado en este trabajo aunque podrían crearse algunos más para análisis de la información. Este esquema representa la proporción de los beneficios sobre activos.

Dimensiones:

- *DT_Tiempo*: ídem anterior.
- *DT_Colegio_Profesion*: ídem anterior

Hechos:

- *FT_Proporcion*: Tiene un único indicador que permite determinar la proporción de activos sobre la cantidad de beneficios.

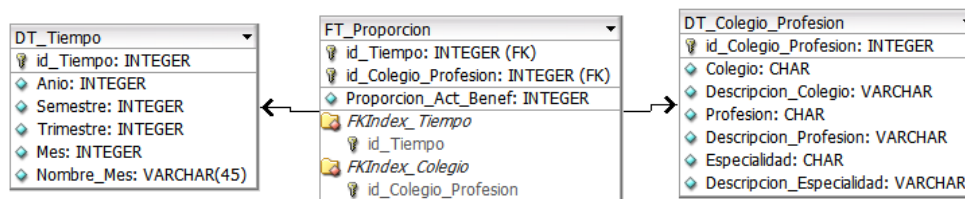


Figura 3.17: Modelado Dimensional. Fact Table: Proporción

3.2.4 Integración de datos

Una vez construido el modelo lógico, se procede a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos. Luego se definen reglas y políticas para su respectiva actualización. La

realización de estas tareas suelen contener una lógica realmente compleja. Afortunadamente, en la actualidad existen muchos softwares que se pueden emplear a tal fin, y que facilitan el trabajo.

El procesamiento de Extracción, Transformación y Carga (ETL), es el componente principal del proceso de DataWarehouse, y ocupa el 70 % del tiempo en un proyecto de data centralizado. Este consiste en la extracción (leer la data de una o más bases de datos), la transformación (convertir la data extraída a la forma que necesita el DW) y la carga (ponerla en el DW). La transformación ocurre usando reglas o tablas especializadas, o combinando los datos. Las herramientas de ETL también transportan datos entre las fuentes y los objetivos, documentan como cambian mientras se mueven, intercambian metadata con otras aplicaciones y administran el proceso de ejecución y operaciones. El proceso de ETL simplifica el mantenimiento de la DW y cuando los datos son manejados adecuadamente como un activo de la empresa, los esfuerzos del ETL se reducen notablemente y la información redundante es eliminada.

Se debe evitar que el DW sea cargado con valores faltantes o anómalos, así como también se deben establecer condiciones y restricciones para asegurar que solo se utilicen los datos de interés.

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

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

En este trabajo no se detalla la definición de todos los ETL para realizar la carga al DW porque sería demasiado extenso. A modo de ejemplo solo se desarrolla la carga inicial del primer esquema dimensional (figura 3.11). De la misma manera se debe proceder con los demás esquemas. Finalmente se concentra en los aspectos más importantes del ETL, obviando entrar en detalle de cómo se realizan algunas funciones y/o pasos.

El proceso de ETL planteado para la **Carga Inicial** es el siguiente:

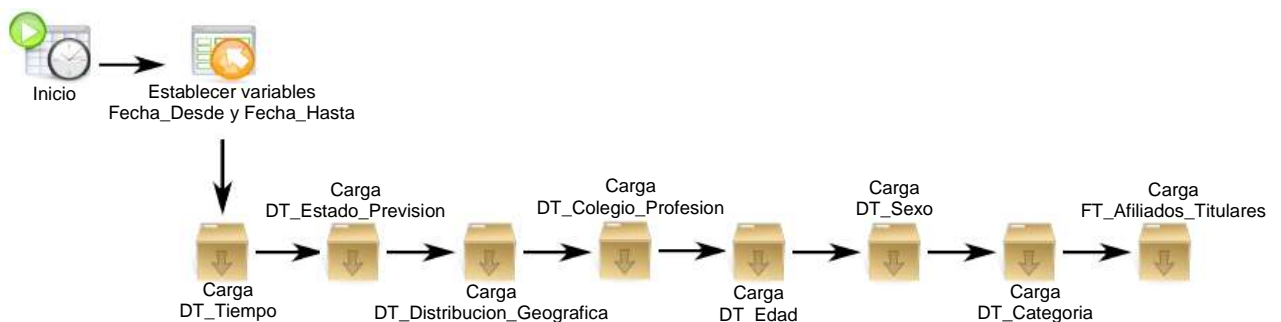


Figura 3.18: Carga Inicial

Las tareas que lleva a cargo este proceso son:

- Inicio: inicia la ejecución de los pasos en los momentos que se le indique.
- Establecer variables Fecha_Desde y Fecha_Hasta: establece dos variables globales que serán utilizadas posteriormente por algunos pasos.
 - o Para la variable Fecha_Desde se obtiene el valor de la fecha en que se matriculó el primer afiliado.
 - o Para la variable Fecha_Hasta se obtiene el valor de la fecha actual.
- Carga de dimensión DT_Tiempo: ejecuta el contenedor de pasos que cargará la dimensión DT_Tiempo.
- Carga de dimensión DT_estado_Prevision: ejecuta el contenedor de pasos que cargará la dimensión DT_estado_Prevision.

- Carga de dimensión DT_Distribucion_Geografica: ejecuta el contenedor de pasos que cargará la dimensión DT_Distribucion_Geografica.
- Carga de dimensión DT_Colegio_Profesion: ejecuta el contenedor de pasos que cargará la dimensión DT_Colegio_Profesion.
- Carga de dimensión DT_Edad: ejecuta el contenedor de pasos que cargará la dimensión DT_Edad.
- Carga de dimensión DT_Sexo: ejecuta el contenedor de pasos que cargará la dimensión DT_Sexo.
- Carga de dimensión DT_Categoria: ejecuta el contenedor de pasos que cargará la dimensión DT_Categoria.
- Carga de tabla de hechos FT_Afiliados_Titulares: ejecuta el contenedor de pasos que cargará la tabla de hechos FT_Afiliados_Titulares.

A continuación se especifican las tareas llevadas a cabo por cada uno de los pasos propuestos en el ETL de la figura 3.18:

Carga DT_Tiempo: Para generar esta tabla de dimensión, infaltable en todo DW, existen varias herramientas y utilidades de software que proporcionan diversas opciones para su confección. Pero, si no se cuenta con ninguna, se puede realizar manualmente o mediante algún programa, llenando los datos en un archivo, tabla, hoja de cálculo, etc, y luego exportándolos a donde se requiera.

El procedimiento en general hace lo siguiente:

- Recibe como parámetros los valores de "Fecha_Desde" y "Fecha_Hasta".
- Recorre cada uno de los periodos (Año/Mes) que se encuentran dentro de este intervalo.
- Inserta los valores obtenidos en la tabla de dimensión DT_Tiempo.

Carga DT_Estado_Previsión: Este paso es un contenedor de pasos que incluye las siguientes tareas:

- Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión DT_Estado_Previsión. Se toma como fuente de entrada las tablas *actividades* y *grupos_actividad*. El campo *id_EstadoPrev* no se define en la sentencia porque es de tipo *identity* (campos numéricos incrementales).

```
select  a.actividad,
        a.descripcion,
        g.grupo_actividad,
        g.descripcion
from    actividades as a, grupo_actividad as g
where   a.grupo_actividad = g.grupo_actividad;
```

- Cargar DT_Estado_Previsión: almacena en la tabla de dimensión DT_Estado_Previsión los datos obtenidos en el paso anterior.

Carga DT_Distribucion_Geografica: Este paso es un contenedor de pasos que incluye las siguientes tareas:

- Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión DT_Distribucion_Geografica. Se toma como fuente de entrada las tablas *localidades*, *departamentos*, *circuitos* y *distritos_electorales*. El campo *id_Distribucion* no se define en la sentencia porque es de tipo *identity*.

```
select  l.provincia,
        dpto.nombre,
        d.nombre,
        l.nombre,
        c.circuito,
        c.nombre,
        l.cod_postal
from    localidades as l as a, departamentos as dpto,
        circuitos as c, distritos_electorales as d
where   l.departamento = dpto.departamento and
```



```
dpto.districto_electoral = d.districto_electoral and  
l.circuito = c.circuito;
```

- Cargar DT_Distribucion_Geografica: almacena en la tabla de dimensión DT_Distribucion_Geografica los datos obtenidos en el paso anterior.

Carga DT_Colegio_Profesion: Este paso es un contenedor de pasos que incluye las siguientes tareas:

- Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión DT_Colegio_Profesion. Se toma como fuente de entrada las tablas *profesiones*, *especialidades* y *colegios*. El campo *id_Colegio_Profesion* no se define en la sentencia porque es de tipo *identity*.

```
select      c.colegio,  
            c.descripcion,  
            p.profesion,  
            p.descripcion,  
            e.especialidad,  
            e.descripcion  
from        profesiones as p, especialidades as e, colegios as c  
where       p.profesion = e.profesion and  
            p.descripcion_colegio_listado = c.colegio;
```

- Cargar DT_Colegio_Profesion: almacena en la tabla de dimensión DT_Colegio_Profesion los datos obtenidos en el paso anterior.

Carga DT_Edad: Para generar esta tabla de dimensión, también existen varias herramientas y utilidades de software que proporcionan diversas opciones para su confección. Además se puede realizar manualmente o mediante algún programa, llenando los datos en un archivo, tabla, hoja de cálculo, etc, y luego exportándolos a donde se requiera. Se coloca un rango de edades de 0 a 110 años.

Carga DT_Sexo: Contiene los siguientes registros:

```
"id_sexo"; "sexo"; "descripción_sexo"  
"1"; "F"; "Femenino";  
"2"; "M"; "Masculino";
```

Carga DT_Categoria: Este paso es un contenedor de pasos que incluye las siguientes tareas:

- Obtener datos de OLTP: obtiene a través de tres consultas SQL los datos del OLTP necesarios para cargar la dimensión DT_Categoria. Se toma como fuente de entrada la tabla *categoría* y se ejecuta para cada escala de aportes (50%, 75% y 100%). El campo *id_Categoria* no se define en la sentencia porque es de tipo *identity*.

```
select      c.descripcion as categoria,  
            50 as escala_aportes,  
            c.modulos as MPA_Total,  
            c.modulos * 0.50 as MPA_Escala  
from        actividades as a;
```

- Cargar DT_Categoria: almacena en la tabla de dimensión DT_Categoria los datos obtenidos en el paso anterior.

Carga tabla de Hechos FT_Afiliados_Titulares: Este paso es un contenedor de pasos que incluye las siguientes tareas:

- Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la tabla de hechos FT_Afiliados_Titulares. Se toma como fuente de datos la tabla *Afiliados_Titulares*. Se aplica una preagregación a los hechos que forman parte de la tabla de hechos. Es por esta razón que se utiliza la cláusula *Group By*

para agrupar los registros. Finalmente la sentencia que contiene este paso tiene todas las relaciones con las tablas de dimensiones definidas y las dos funciones agregadas siguientes:

```
select  t.id_tiempo,
        e.id_EstadoPrev,
        d.id_distribucion,
        c.id_colegio_profesion,
        e.id_edad,
        s.id_sexo,
        cat.id_categoria,
        count(afi.nro_afiliado_rpg)
        avg(cat.mpa_escalas)
from    DT_tiempo as t,
        DT_Estado_Prevision as e,
        DT_Distribucion_Geografica as d,
        DT_Colegio_Profesion as c,
        DT_Edad as e,
        DT_Sexo as s,
        DT_categoria as cat,
        Afiliados_titulares as afi
where   t.anio = datepart(year,afi.fecha_matricula) and
        t.mes = datepart(month,afi.fecha_matricula) and
        e.actividad = afi.actividad and
        d.localidad = afi.localidad_legal and
        c.profesion = afi.profesion and
        e.edad = datediff(afi.fecha_nacimiento, getdate(), year) and
        s.sexo = afi.sexo and
        cat.categoria = afi.categoria and
        cat.escalas_aportes = afi.escalas_aportes
Group by t.id_tiempo,
        e.id_EstadoPrev,
        d.id_distribucion,
        c.id_colegio_profesion,
        e.id_edad,
        s.id_sexo,
        cat.id_categoria;
```

- Cargar FT_Afiliados_Titulares: almacena en la tabla de hechos FT_Afiliados_Titulares los datos obtenidos en el paso anterior.

Las políticas y **estrategias de actualización** o refresco de datos deben llevar a cabo las siguientes acciones:

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

A modo de ejemplo, las políticas de actualización que se pueden establecer son las siguientes:

- La información se refrescará todos los últimos días de cada mes ya que la dimensión DT_Tiempo se define de manera mensual.
- Los datos de la tabla de dimensión DT_Tiempo se cargarán de manera incremental teniendo en cuenta la fecha de la última actualización.
- Estas acciones se pueden realizar sobre un periodo de prueba, para analizar la manera más eficiente de generar las actualizaciones, basadas en el estudio de los cambios que se producen en los OLTP y que afectan al contenido del DW.

3.3 Herramienta de Desarrollo

3.3.1 Características Principales

La herramienta propuesta para el desarrollo es **SQL Server Analysis Services** que fue adquirido por la institución cuando obtuvo un paquete de productos Microsoft. Da la capacidad para realizar tareas de OLAP y minería de datos. Proporciona un único punto de acceso a los datos para la inteligencia de negocio de la organización (BI). Mediante la combinación de OLAP, las tecnologías de minería de datos y la información de indicadores claves de rendimiento (KPI: Key Performance Indicator) en un solo sistema unificado, simplifica el desarrollo de soluciones de BI y hace que resultados sean más accesibles a los usuarios.

- La funcionalidad de OLAP le permite consultar los datos agregados almacenados en cubos multidimensionales.
- La minería de datos puede identificar las relaciones dentro de los datos y evaluar las probabilidades de los resultados futuros basados en acciones pasadas.
- Analysis Services también combina los datos de OLAP y fuentes relacionales.
- El soporte de KPI en Analysis Services permite crear cuadros de mando (o cuadros de balance) para evaluar el desempeño contra los objetivos de negocio.

3.3.2 Arquitectura del Analysis Services

La arquitectura consta de dos componentes principales: la arquitectura del servidor y arquitectura del cliente. La arquitectura del servidor incluye todas las tareas de almacenamiento y procesamiento manejados por el servidor:

- Consiste en una aplicación que se ejecuta como un servicio.
- Las bases de datos de Analysis Services se almacenan en el servidor y se acceden a través del servicio de Analysis Services. Este servicio se conforma de los componentes que se encargan de las agrupaciones, las transacciones, cálculos, gestión de metadatos, seguridad y XML for Analysis (XMLA), así como numerosas otras funciones.

La arquitectura del cliente incluye las tareas de conexión y caché manejadas por el cliente.

- Permite a los clientes conectarse a cubos de Analysis Services utilizando ADO MD, MD.Net ADO, XML / A, o OLEDB para OLAP.
- Los usuarios pueden acceder al cubo desde una aplicación, como Microsoft Office Excel.

3.3.3 Herramientas del Analysis Services

Hay tres herramientas clave en el desarrollo y gestión de soluciones de Analysis Services de SQL Server: **Configuration Manager**, **SQL Server Management Studio** y **Business Intelligence Development Studio**. Se pueden realizar la mayoría de las tareas mediante el uso de más de un método, y se pueden realizar muchas tareas mediante el uso de múltiples herramientas.

- Business Intelligence Development Studio es una versión de Microsoft Visual Studio 2008. Es la principal herramienta para la creación de soluciones de BI.
- SQL Server Management Studio es la herramienta principal para el manejo de datos relacional de bases de datos y Analysis Services. Puede utilizar esta herramienta para desarrollar soluciones de Analysis Services, pero es más comúnmente utilizado para administrar bases de datos de Analysis Services en un entorno de producción.
- SQL Server Configuration Manager es una herramienta que puede utilizar para administrar los servicios de SQL Server y las configuraciones de red del servidor y del cliente.

3.3.4 Objetos fundamentales de Analysis Services

Las bases de datos OLAP se definen por los cubos, dimensiones y medidas. También es necesario definir la información de conexión y organización de las dimensiones y medidas. Los objetos se combinan para formar una base de datos de Analysis Services.

- Los orígenes de datos almacenan la información de conexión.
- Las vistas de origen de datos proporcionan una capa abstracta para los orígenes de datos.
- Las medidas son los valores numéricos, o hechos, que los usuarios analizan.
- Los grupos de medida son agrupaciones lógicas de las medidas.
- Las dimensiones proporcionan un contexto a las medidas. Las dimensiones se organizan en estructuras jerárquicas. Por ejemplo, una dimensión temporal podría contener día, mes, y los niveles del año.
- Los cubos combinan las dimensiones y medidas que forman una estructura multidimensional que contiene la suma de cada medida en la intersección de los miembros de dimensión. Los cubos son los objetos que los usuarios manipulan para acceder a los datos.

3.3.5 Nuevas características de Analysis Services en SQL Server 2008

Hay muchas nuevas características de Analysis Services de SQL Server 2008.

- Un nuevo diseñador de agregación hace que sea más fácil de explorar y modificar los diseños de agregaciones. Los diseños de agregaciones ahora se muestran agrupados por grupo de medida.
 - El diseñador de agregaciones actualizado y basado en asistentes le permiten modificar los parámetros de almacenamiento para las agregaciones en una o más particiones a la vez y con mayor facilidad establecer la configuración de uso de agregación.
 - Los nuevos mensajes de advertencia AMO alertan a los usuarios cuando salen de las mejores prácticas de diseño o errores lógicos.
 - El editor de dimensión tiene un diseñador de relaciones de atributos nuevos que hace que sea más fácil de examinar y modificar las relaciones de atributos.
 - La última versión del Asistente para dimensiones detecta automáticamente jerarquías entre padres e hijos, ofrece seguros de configuración de error por defecto, y es compatible con la especificación de las propiedades de miembro.
 - Un nuevo cuadro de diálogo de Columnas clave permite editar las columnas de clave más fácil.
 - La ficha Estructura de dimensión ahora trabaja con el diseñador de relaciones de atributos nuevo y es más fácil de usar, lo que hace que los atributos y las jerarquías de la modificación más fácil.
- La nueva estructura de almacenamiento proporciona un repositorio más robusto para la base de datos archivados.
 - La nueva copia de seguridad y la funcionalidad de restauración logra un mayor rendimiento.
 - Las nuevas extensiones de personalización permiten a los desarrolladores crear nuevos objetos de Analysis Services y la funcionalidad, y para poder proveer estos objetos y la funcionalidad de forma dinámica en el contexto de la sesión del usuario.
 - Los libros en línea tienen Bases de datos y aplicaciones de ejemplo disponibles en el sitio Web de SQL Server.



Figura 3.19: Versión Microsoft SQL Server Management Studio

3.4 Análisis de Resultados

A continuación se analizan las operaciones que se pueden realizar sobre los modelos multidimensionales. Se ejemplifica su utilización, para lo cual se utiliza el primer diagrama estrella planteado sobre los afiliados titulares. Estas operaciones son las que verdaderamente permitirán a los usuarios explorar e investigar los datos en busca de respuestas.

La operación **Drill-Down** permite apreciar los datos en un mayor detalle, bajando por una jerarquía definida en un cubo. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, disgregando los grupos actuales. Permite el desplazamiento de lo general a lo específico.

Para explicar esta operación se utiliza la siguiente representación tabular que indica la cantidad de afiliados titulares por colegio y categoría para el periodo 2014/07:

Colegio	A	B	C	D	E	Total
01-MEDICOS	648	959	1482	1936	6642	11667
02-ODONTOLOGOS	202	336	487	808	2199	4032
03-FARMACEUTICOS	60	110	132	213	1707	2222
04-VETERINARIOS	22	169	256	359	1147	1953
05-BIOQUIMICOS	51	59	91	185	1024	1410
06-KINESIOLOGOS	184	593	407	354	555	2093
07-FONAUDIOLOGOS	32	106	121	134	479	872
08-PSICOLOGOS	186	760	617	628	1618	3809
09-OBSTETRAS	7	11	16	13	20	67
10-TECNICOS OPTICOS		1	1	7	62	71
11-PSICOPEDAGOGOS	107	142	92	75	141	557
12-DIETISTAS	107	270	167	70	52	666
13-TEC. RADIOLOGOS	10	16	14	9	51	100
15-T. OCUPACIONALES	3	7	5	3		18
Total	1619	3539	3888	4794	15697	29537

Cuadro 3.10: Resultados antes de aplicar Drill-Down

Luego aplicando la operación Drill-Down se puede obtener más información sobre las cantidad de afiliados de acuerdo a la profesión de cada colegio. En el siguiente cuadro, se agregó un nuevo nivel de detalle y el valor 11.667 que pertenecía a la cantidad de afiliados del Colegio de Médicos, se dividió en seis filas. Esto se debe a que ahora se tendrá en cuenta el atributo profesión para realizar las sumarizaciones de la cantidad de afiliados.

Colegio	Profesión	Descripción	A	B	C	D	E	Total
01-MEDICOS	ENF	ENFERMERIA						5
	LEN	LICENCIADO ENFERMERIA	2	1	2			5
	MED	MEDICO	634	885	1381	1867	6581	11348
	TER	TERAPISTA OCUPACIONAL	11	69	96	67	48	291
	TIQ	TEC. SUPINSTRUM. QUIRURGICA	1	4	2	2	5	14
	TSH	TEC. SUPEN HEMOTERAPIA				1	3	4
01-MEDICOS Total			648	959	1482	1936	6642	11667
02-ODONTOLOGOS	DEN	MECANICO DENTAL					2	2
	ODO	ODONTOLOGO	202	336	487	808	2197	4030
02-ODONTOLOGOS Total			202	336	487	808	2199	4032
03-FARMACEUTICOS	FAR	FARMACEUTICO	60	110	132	213	1707	2222
03-FARMACEUTICOS Total			60	110	132	213	1707	2222
04-VETERINARIOS	VET	VETERINARIO	22	169	256	359	1147	1953
04-VETERINARIOS Total			22	169	256	359	1147	1953
05-BIOQUIMICOS	BIO	BIOQUIMICO	51	59	91	185	1019	1405
	DBI	DR. EN BIOQUIMICA					3	3
	DQU	DR. EN QUIMICA					1	1

Cuadro 3.11: Resultados después de aplicar Drill-down

Sin el modelado dimensional era muy difícil poder obtener más detalle sobre alguna característica en particular. Al tratarse de reportes rígidos, para obtener más información, se tenía que desarrollar un nuevo reporte con el nivel de detalle requerido y muchas veces los resultados no eran

los esperados ya que en el caso del ejemplo, cambiaban los totales al haber especialidades nulas o no especificadas. Además había una fuerte dependencia del área sistemas para poder obtener información más detallada. Con el modelado dimensional la accesibilidad de los datos es mucho más dinámica y se requiere menos tiempo para la elaboración de los informes.

Luego, **Drill-Up** es la operación contraria que permite apreciar los datos en menor nivel de detalle, subiendo por una jerarquía definida en un cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, agregando los grupos actuales. Drill-Up es ir de lo específico a lo general.

Por otro lado, la operación **Drill-Across** funciona de manera similar a Drill-Down, con la diferencia de que Drill-Across no se realiza sobre una jerarquía, sino que su forma de ir de lo general a lo específico es agregar un atributo en la consulta como nuevo criterio de análisis.

De esta forma el cuadro resultante de aplicar la operación Drill-Across sobre el cuadro 3.10 agregando la columna sexo, es el que sigue:

Colegio/Sexo	A		B		C		D		E		Total
	F	M	F	M	F	M	F	M	F	M	
01-MEDICOS	220	428	553	406	848	634	921	1015	2455	4187	11667
02-ODONTOLOGOS	126	76	200	136	304	183	472	336	1164	1035	4032
03-FARMACEUTICOS	47	13	82	28	89	43	154	59	1228	479	2222
04-VETERINARIOS	10	12	72	97	88	168	106	253	182	965	1953
05-BIOQUIMICOS	31	20	48	11	61	30	130	55	657	367	1410
06-KINESIOLOGOS	131	53	358	235	211	196	187	167	325	230	2093
07-FONAUDILOGOS	29	3	104	2	119	2	131	3	476	3	872
08-PSICOLOGOS	159	27	664	96	526	91	544	84	1312	306	3809
09-OBSTETRAS	7		11		16		13		20		67
10-TECNICOS OPTICOS			1		1		4	3	42	20	71
11-PSICOPEDAGOGOS	104	3	140	2	90	2	70	5	138	3	557
12-DIETISTAS	104	3	256	14	154	13	68	2	49	3	666
13-TEC. RADIOLOGOS	8	2	11	5	8	6	3	6	19	32	100
15-T. OCUPACIONALES	3		6	1	5		2	1			18
Total	979	640	2506	1033	2520	1368	2805	1989	8067	7630	29537

Cuadro 3.12: Resultados después de aplicar Drill-Across

Ahora tenemos el mismo resultado sobre el colegio de médicos (11.667 afiliados) pero discriminado por sexo y categoría.

Luego, **Roll-Across** funciona de manera similar a Drill-Up, con la diferencia de que no se hace sobre una jerarquía, sino que su forma de ir de lo específico a lo general es quitar un atributo de la consulta, eliminando de esta manera un criterio de análisis.

Por otro lado, la operación **Pivot** permite seleccionar el orden de visualización de los atributos e indicadores, con el objetivo de analizar la información desde diferentes perspectivas. Se tiene el cuadro 3.13 que indica la cantidad de afiliados por colegio, circuito y categoría:

Colegio	Circuito	Nombre	A	B	C	D	E	Total
01-MEDICOS	11	SANTAFE CAPITAL	86	238	273	312	820	1729
	12	NORTE DE SANTA FE	68	142	272	377	1062	1921
	21	ROSARIO	410	446	679	895	2978	5408
	22	SUR DE SANTA FE	78	108	201	250	966	1603
	31	FUERA DE LA PROVIN.	1	8	14	21	73	117
	32	DOM. DESCONOCIDO	5	17	43	81	743	889
01-MEDICOS Total			648	959	1482	1936	6642	11667
02-ODONTOLOGOS	11	SANTAFE CAPITAL	18	45	75	111	238	487
	12	NORTE DE SANTA FE	31	74	102	140	358	705
	21	ROSARIO	119	163	223	390	1083	1978
	22	SUR DE SANTA FE	33	47	81	151	396	708
	31	FUERA DE LA PROVIN.	1	4	3	3	10	21
	32	DOM. DESCONOCIDO		3	3	13	114	133
02-ODONTOLOGOS Total			202	336	487	808	2199	4032
03-FARMACEUTICOS	11	SANTAFE CAPITAL	5	11	8	18	140	182
	12	NORTE DE SANTA FE	13	39	49	62	392	555
	21	ROSARIO	21	26	35	61	605	748
	22	SUR DE SANTA FE	18	28	35	65	492	638
	31	FUERA DE LA PROVIN.					6	6

Cuadro 3.13: Resultados antes de aplicar Pivot

Luego aplicando la operación Pivot y reorientando la vista multidimensional se obtiene:

Circuito	Nombre	Colegio	A	B	C	D	E	Total
11	SANTAFE CAPITAL	01-MEDICOS	86	238	273	312	820	1729
		02-ODONTOLOGOS	18	45	75	111	238	487
		03-FARMACEUTICOS	5	11	8	18	140	182
		04-VETERINARIOS	2	23	26	41	108	200
		05-BIOQUIMICOS	6	15	21	50	201	293
		06-KINESIOLOGOS	19	59	26	37	55	196
		07-FONAUDIOLOGOS	1	1	4	6	27	39
		08-PSICOLOGOS	13	139	114	84	94	444
		09-OBSTETRAS	4	1	2	1	2	10
		10-TECNICOS OPTICOS					2	2
		11-PSICOPEDAGOGOS	13	22	20	23	39	117
		12-DIETISTAS	27	45	17	11	9	109
		13-TEC. RADIOLOGOS	5	2	3		8	18
		15-T. OCUPACIONALES	1	2	3	1		7
		SANTA FE CAPITAL Total			200	603	592	695
12	NORTE DE SANTA FE	01-MEDICOS	68	142	272	377	1062	1921
		02-ODONTOLOGOS	31	74	102	140	358	705
		03-FARMACEUTICOS	13	39	49	62	392	555
		04-VETERINARIOS	18	114	127	168	527	954
		05-BIOQUIMICOS	17	25	36	62	282	422
		06-KINESIOLOGOS	32	109	88	91	192	512
		07-FONAUDIOLOGOS	8	20	27	31	50	136
		08-PSICOLOGOS	17	117	111	119	137	582

Cuadro 3.14: Resultados después de aplicar Pivot

Pivot permite realizar las siguientes acciones:

- Mover un atributo o indicador desde el encabezado de fila al encabezado de columna.
- Mover un atributo o indicador desde el encabezado de columna al encabezado de fila.
- Cambiar el orden de los atributos o indicadores del encabezado de columna.
- Cambiar el orden de los atributos o indicadores del encabezado de fila.

Por otro lado la operación **Page**, presenta el cubo dividido en secciones, a través de los valores de un atributo, como si se tratase de páginas de un libro. La figura 3.20 representa este concepto gráficamente:

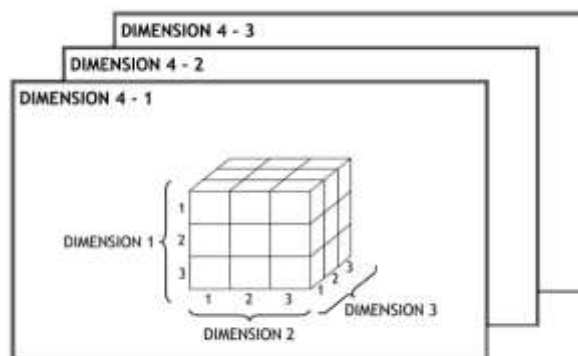


Figura 3.20: Page

Page es muy útil cuando las consultas devuelven muchos registros y es necesario desplazarse por los datos para poder verlos en su totalidad. Por ejemplo si se quiere analizar los distintos colegios de manera separada se obtendrían las siguientes páginas:

Página Nro 1: Colegio: 01 - Médicos

Circuito	Nombre	Sexo	A	B	C	D	E	Total
11	SANTA FE CAPITAL	F	35	140	147	145	255	722
		M	51	98	126	167	565	1007
12	NORTE DE SANTA FE	F	26	86	165	189	336	802
		M	42	56	107	188	726	1119
21	ROSARIO	F	136	254	395	426	1251	2462
		M	274	192	284	469	1727	2946
22	SUR DE SANTA FE	F	22	66	117	129	369	703
		M	56	42	84	121	597	900
31	FUERA DE LA PROVIN.	F		5	9	6	23	43
		M	1	3	5	15	50	74
32	DOM. DESCONOCIDO	F	1	2	15	26	221	265
		M	4	15	28	55	522	624
Total			648	959	1482	1936	6642	11667

Cuadro 3.15: Resultados después de aplicar Page – Colegio Médicos

Página Nro 2: Colegio: 02 - Odontólogos

Circuito	Nombre	Sexo	A	B	C	D	E	Total
11	SANTA FE CAPITAL	F	12	24	43	55	93	227
		M	6	21	32	56	145	260
12	NORTE DE SANTA FE	F	14	45	61	74	172	366
		M	17	29	41	66	186	339
21	ROSARIO	F	79	100	149	254	648	1230
		M	40	63	74	136	435	748
22	SUR DE SANTA FE	F	20	29	49	82	205	385
		M	13	18	32	69	191	323
31	FUERA DE LA PROVIN.	F	1	1	1		2	5
		M		3	2	3	8	16
32	DOM. DESCONOCIDO	F		1	1	7	44	53
		M		2	2	6	70	80
Total			202	336	487	808	2199	4032

Cuadro 3.16: Resultados después de aplicar Page – Colegio Odontólogos

Cuando existe más de un criterio por el cuál realizar Page, debe tenerse en cuenta el orden en que éstos serán procesados, ya que dependiendo de esto, se podrán obtener diferentes resultados sobre una misma consulta. El primer criterio utilizado para realizar Page condiciona los valores disponibles en el segundo y así sucesivamente.

Finalmente la operación **Drill-Through** permite apreciar los datos en su máximo nivel de detalle. Esto brinda la posibilidad de analizar cuáles son los datos relacionados al valor de un indicador, que se ha sumariado dentro del cubo multidimensional.

Colegio	Circuito	Nombre	Localidad	A	B	C	D	E	Total
01-MEDICOS	11	SANTA FE CAPITAL	30000 - SANTA FE	86	238	273	312	820	1729
		NORTE DE SANTA	121461 - SAN GENARO			2	1	7	10
			21470 - S.GENARO NORTE	1	1	2		4	8
			21480 - CASAS					1	1
			21481 - CENTENO			1		1	2
			22080 - MACIEL	2	1	1	2	4	10
			22081 - PUERTO GABOTO					1	1
			22120 - MONJE		3			2	5
			22220 - DIAZ					2	2
			22400 - CORONDA		2	1	3	17	23
			22420 - AROCENA					1	1
			22460 - BARRANCAS			2	1	3	6
			22480 - B.DE IRIGOYEN					3	3
			22520 - GALVEZ	2	4	7	7	17	37
			22530 - GESSLER				1		1
			22550 - LOPEZ					1	1
			22570 - COLONIA BELGRANO					1	1
			22580 - SANTA C.DE B.VISTA			1		2	3
			23000 - RAFAELA	8	20	42	71	207	348
			23001 - BELLA ITALIA			1			1
		23010 - RAMONA					1	1	
		23017 - VILA				1	1	1	

Cuadro 3.17: Resultados después de aplicar Drill Through

Cada una de estas operaciones mejora visiblemente el acceso a los datos y la elaboración de informes. Además el tiempo de respuesta es mucho más rápido al haber datos precalculados en los cubos multidimensionales. Finalmente es el mismo usuario el que se introduce en el nivel de detalle requerido.

CAPITULO IV

4. Evaluación de Resultados

4.1 Análisis de las variables y sus componentes

La medición de los beneficios de BI no es tan simple como parece, ya que muchos efectos favorables no son financieros, son intangibles, como la mejora de la calidad y la información actualizada. Adicionalmente hay un efecto retrasado entre la producción de la inteligencia y la obtención del resultado financiero. Cuando se mide el BI como un proceso, los profesionales de BI son los mayores usuarios donde se debe medir la información, y el objetivo es producir inteligencia de valor de una manera eficiente para necesidades específicas de los usuarios. Por lo tanto los aspectos más relevantes son la eficiencia del personal de BI, la efectiva asignación de recursos de valor, la calidad de los productos de BI, y la satisfacción del usuario.

En el apartado previo que hacía referencia a la importancia de esta investigación, y se ha establecido que los resultados finales se medirían en cuanto a la *calidad de información* obtenida con la aplicación de las tecnologías de BI, en cuanto al *tiempo de respuesta* de las consultas, en cuanto al *tiempo de elaboración de informes* y la *complejidad* en su preparación y en cuanto a las *posibilidades de explotación* que brinda el uso del Data Warehouse, siempre comparándolo con los rendimientos y respuesta que se tiene en la actualidad, con dependencia total del área de sistemas. Todo esto respaldado en una estrategia centrada en la mejora de la calidad del servicio que se presta a los afiliados profesionales apuntando a la *integración* y *accesibilidad* en los datos informativos.

Sin embargo, al llevar a cabo esta investigación, se ha detectado que hay muchas variables más para tener en cuenta al momento de evaluar los resultados. Se pueden resumir en: Calidad de la Información o Calidad de los Datos, Calidad del Sistema (Sistema de Información, en este caso Calidad del Sistema de DW e BI), Calidad del Servicio (Servicio que le dan los técnicos de DW o IT a los usuarios del sistema), Satisfacción del Usuario, Uso del Sistema, Impacto Individual (en el usuario), Impacto Organizacional, Utilidad Percibida (por el usuario), y facilidad de Uso Percibida (por el usuario). Las principales variables y componentes se pueden ver en la siguiente figura:

Principales Variables	Componentes
Calidad de la Información o Calidad de los Datos	Importancia Relevancia Comprensibilidad Claridad Contenido Formato Precisión Confiabilidad Actualidad Sin Errores
Calidad del Sistema	Contenido de la Base de Datos Facilidad de Uso Facilidad de Aprendizaje Facilidad de Acceso Flexibilidad del Sistema Confiabilidad del Sistema Exactitud del Sistema Tiempo de Respuesta Tiempo de Procesamiento
Calidad del Servicio	Servicio Tangible Confiabilidad en el Servicio Seguridad en el Servicio Compromiso en el Servicio Empatía en el Servicio
Satisfacción del Usuario	Satisfacción General Satisfacción con aspectos específicos Diferencia entre información necesitada y recibida Disfrute Satisfacción en la Toma de Decisiones

Uso del Sistema	Cantidad-Duración de Uso Uso por quién Nivel de Uso Recurrencia de Uso Volumen de Uso Motivación del uso
Impacto individual	Efectividad de la Decisión Comprensión de la Información Aprendizaje Exactitud de Interpretación Identificación del Problema
Impacto Organizacional	Reducción Costo Operativo Reducción de Staff Incremento de la Productividad Incremento de Ventas Incremento Participación de Mercado Incremento Utilidades Ratio Beneficio/Costo
Utilidad Percibida	Mejora en el Desempeño Satisfacción de Necesidades en el Trabajo Realización de mayor cantidad de trabajo Mayor Productividad
Facilidad de Uso Percibida	Flexibilidad en la Utilización Fácil de entender Fácil de recordar el manejo Provee de guía para el desempeño Facilidad de uso

Cuadro 4.1: Principales Variables – Impacto de la Data Warehouse e Inteligencia de Negocios

Finalmente es posible afirmar que hay tres grandes grupos de variables que impactan en el desempeño de la DW y BI, de acuerdo a relevancia son:

- 1- *Calidad de la Información* (Información Actualizada, Fuente de Datos Adecuada)
- 2- *Calidad del Sistema* (Herramientas, ETL de los Datos, Tiempo de Respuesta)
- 3- *Calidad del Servicio* (Entrenamiento del Usuario y Entrenamiento del personal de tecnologías de Información, Posibilidades de Explotación, Tiempo de Elaboración de Informes y Complejidad de su Preparación)

Estos tres grandes grupos están positivamente relacionados con nuestras variables de *Integración y Accesibilidad* y se relacionan de manera transitiva con dos variables que no hemos citados hasta el momento como son: el *Uso del Sistema* y la *Satisfacción de Usuario*.

4.1.1 Calidad de la Información o Calidad de los Datos

La variable de *Calidad de Información* se refiere a tener una fuente de datos adecuada en la institución que sea única y con información actualizada.

A través de los diversos estudios de investigación realizados en Sistemas de Información (SI), se han utilizado medidas diversas para evaluar la calidad del desempeño del sistema, y una de ellas ha sido la calidad de la información producida por el sistema en forma de archivos de información de salida. Gonzalez [13] en su investigación sobre el impacto de DW y BI en el desempeño de las empresas identifica ciertas variables relacionadas con la calidad de la información como: exactitud, precisión, actualidad, estar a tiempo, confiabilidad, completitud, ser conciso, formato y relevancia.

Se puede establecer que la calidad de los datos es un factor clave a lo largo del proceso de BI. La organización tiene que darse cuenta y ser consciente de cómo están sus datos. Se ha observado a lo largo de este trabajo de investigación que llevar a cabo un proceso de tal magnitud, requiere que su datos se unifiquen de manera consistente, ya que si los usuarios notan lo contrario, todo habrá fracasado.

Es un proceso que se va madurando, pero se ha podido demostrar que al trabajar con información en línea, en lugar de tener un grupo de personas construyendo reportes aislados, la gerencia lo nota inmediatamente al tener todos los días la información al cierre.

4.1.2 Calidad del Sistema

Esta variable refiere a tener un buen sistema de Extracción, Transformación y Carga de Datos, a tener una plataforma única y a tener una DW bien estructurada.

El *Tiempo de Respuesta* es una variable que representa a la Calidad del Sistema, específicamente en el caso de la Data Warehouse, en las cuales la obtención de la respuesta requiere de cierto tiempo de uso de la computadora. Se puede evaluar el impacto individual en el usuario en relación a realizar sus funciones más rápidamente, mejorar el desempeño, mejorar la productividad, realizar el trabajo más rápidamente y hacerlo más fácilmente. Se incrementa la productividad y ayuda a satisfacer las necesidades del cliente.

También se puede mencionar que cada vez más empresas tienen Data Warehouses y que el número de usuarios va aumentando, por lo que la variable Tiempo de Respuesta se va volviendo más relevante. Asiente una arquitectura sólida, totalmente estructurada, con indicadores e índices versátiles, que permiten flexibilizar la búsqueda con herramientas que facilitan la rapidez de consulta.

4.1.3 Calidad del Servicio

La variable de *Calidad del Servicio* se refiere al entrenamiento del usuario y del personal de IT.

Con la economía moderna cambiando gran parte de su actividad hacia los servicios, y en IS cambiando de ser un proveedor de información (o producto) a ser un proveedor de servicio (apoyo a los usuarios finales), es necesario medir no solamente la calidad del sistema y de la información, sino también del servicio que otorga.

Gonzalez [13] indica que la calidad del servicio debería medirse a través de la competencia técnica y actitud del staff de IS, el cumplimiento de fechas para la entrega de productos y servicios, el tiempo requerido para el desarrollo de sistemas, el tiempo requerido para cambios en el sistema, el soporte de mantenimiento de ventas, medios de entrada y salida con el centro de IS, comprensión del sistema por parte de los usuarios, y por el entrenamiento que se le da a los usuarios.

La información del DW es manejada por los usuarios, quienes tienen el control de los datos y tienen la responsabilidad de determinar y encontrar los datos que necesitan. Por tal motivo, el diseño del DW tiene que ser realizado de acuerdo a la perspectiva del usuario, lo que también influirá en el desarrollo de la misma. Existen muchos efectos y ventajas en el uso del DW y BI, los principales son: mejoras en los reportes de información, en búsquedas para el análisis detallado, obtención de alertas y análisis estadísticos, facilidad de uso, velocidad al momento de obtener información, acceder a mayor cantidad de información, mejorar la productividad, tomar mejores decisiones y mejorar los procesos.

4.1.4 Uso del Sistema

La variable *Uso del Sistema* de Información ha sido utilizada por muchos investigadores como una de las variables más importantes para indicar el éxito del sistema de información. Es fundamental que una vez implementado se tenga una buena intensidad de uso, ya que muchas veces se trabaja intensamente en proyectos que no se terminan utilizando. El éxito final se va a dar cuando los usuarios utilicen naturalmente el DW.

4.1.5 Satisfacción del Usuario

La *Satisfacción del Usuario* se refiere exactamente a que tan bien se siente el usuario del sistema al usar el mismo y es un indicador de qué tan bueno es. Ha sido una de las medidas más usadas desde hace buen tiempo. Hay factores que explican el motivo del gran uso de esta variable: el primero es debido a que tiene una gran validez de cara, ya que la relación entre un usuario del sistema que está satisfecho con el mismo y el hecho de que el sistema funcione bien, es muy probable. Se percibe en la cara de los usuarios al ver que existen herramientas que han solucionado un pesado trabajo que hacían antes.

Conclusiones y Trabajos Futuros

El presente estudio se centró en analizar de qué manera se puede aprovechar el potencial que ofrecen las tecnologías de Inteligencia Empresarial en el contexto de la CPAC. Luego se definió una guía metodológica adaptada a los requerimientos de la institución basada en el análisis de las fuentes de los sistemas de información actuales para obtener de ellos no solamente información, sino una verdadera inteligencia de datos. Se desarrolló una solución de BI siguiendo la guía propuesta, se propuso un plan inicial de integración, se presentó una herramienta de trabajo, y se pudo demostrar que es posible proporcionar a los usuarios una manera universal de acceder, ver y utilizar la información proporcionando una ventaja competitiva y garantizando aspectos como accesibilidad, integración tecnológica y toma de decisiones.

Esta metodología ayudará a los distintos sectores y departamentos de la institución a obtener resultados benéficos como: generar reportes globales o por secciones, conocer la realidad de los aportantes a la previsión y obra social, crear escenarios con respecto a una decisión, compartir información entre los departamentos, análisis multidimensionales, generar y procesar datos, en definitiva, mejorar el servicio al afiliado. Sin embargo, es mucho el camino que queda por recorrer.

Posteriormente se analizaron los factores significativos y sus componentes para determinar el éxito del DW y BI. Se pudo establecer que las principales variables que impactan en el desempeño de la DW y BI son: Calidad de la Información, Calidad del Sistema, Calidad del Servicio, Uso del Sistema y Satisfacción del Usuario.

Este trabajo es un buen punto de partida para que la institución conozca los beneficios de esta tecnología y evalúe posibles ampliaciones. Se ha observado que las necesidades de información son muchas. Para dar comienzo a esta investigación y demostrar la necesidad de un DW a corto plazo, se escogieron algunos indicadores con sus variables o perspectivas asociadas para la toma de decisiones. Sería interesante amplificar esta selección y seguir incorporando nuevas perspectivas y medidas que sean de utilidad para los tiempos actuales de la CPAC. Los mayores éxitos de las organizaciones que compiten mediante el uso de Business Intelligence se obtienen cuando somos capaces de aportar información sobre áreas estratégicas de la organización.

Finalmente, se pudo establecer que la participación de los usuarios en el proyecto es fundamental, ya que ellos serán los que consigan los éxitos con el uso de las soluciones de BI. Es primordial que se registren los éxitos obtenidos con el uso de las soluciones. El poder compartirlos asegurará la continuidad y los recursos necesarios para seguir avanzando.

Una posible extensión de este proyecto podría ser la elaboración de un **Data Mining** o **Minería de Datos**. Esta tecnología consiste en la extracción de información significativa y patrones de comportamiento que permanecen ocultos entre grandes cantidades de datos contenidas en el Data Warehouse. La Minería de Datos integra a ciencias como la Estadística y la Inteligencia Artificial con el fin de buscar patrones de comportamiento. Por un lado se pueden establecer modelos predictivos basados en el estudio detallado de cómo han evolucionado los datos existentes hasta ahora y a partir de esa información, extrapolar su posible funcionamiento futuro. Por otra parte, también es posible analizar la información disponible a través de técnicas de Data Mining para tratar de encontrar relaciones que regulan de forma oculta el funcionamiento conjunto de esos datos. Esta actividad llevaría a la elaboración de modelos descriptivos que explicarían con detalle por qué los datos muestran determinados valores y por qué cambian esos valores con el tiempo.

Referencias Bibliográficas

- [1] Apuntes Curso de Posgrado: Inteligencia de Negocios: Data Warehouse y Data Mining. 2009.
- [2] Apuntes Curso de Posgrado: Seminario de Integración: 2013.
- [3] BRESLIN, M. Data Warehousing Battle of Gigants: Comparing the Basics of the Kimball and Immon Models. Business Intelligence Journal, 2004.
- [4] CAMERON, S. & CONSULTING, H. Microsoft SQL Server 2008 Analysis Services. Step by Step. Microsoft Press. 2009.
- [5] CANO, J. Business Intelligence: Competir con Información. Depósito Legal: M-41185, 2007.
- [6] CHERRYTREE & CO. Business Intelligence - The Missing Link. www.cherrytreeco.com. 2000.
- [7] CUBILLO, J. La inteligencia empresarial de las pequeñas y medianas empresas competitivas de América Latina – algunas reflexiones. 1997.
- [8] DATE, J.C. Introducción a los Sistemas de Base de Datos. México. Pearson Educación. 2001.
- [9] DAVENPORT, T.H. & HARRIS, J.G. Competing on Analytics, The New Science of Winning. Boston: Harvard Business School Press. 2007
- [10] EROLES, A. La pequeña empresa exportadora. Panorama, México DF 1995.
- [11] FERNANDEZ CONCEPCION, R., LABRADOR RAMOS, K., FERNANDEZ LORENZO, A. & GUERRA CHICO, A. La inteligencia empresarial: Análisis teórico y pasos para su implementación en el contexto cubano. Universidad de Pinar del Río. 2010.
- [12] GARTNER, www.gartner.com. Gartner un una consultora Internacional especializada en Tecnología de Información y Comunicación, 2006
- [13] GONZALEZ LOPEZ R. Impacto de la Data Warehouse e Inteligencia de Negocios en el Desempeño de las Empresas: Investigación Empírica en Perú, como país en vía de desarrollo. Universitat Ramon Llull. Depósito Legal: B.29511-2012. 2012.
- [14] HASELDEN, K. Microsoft SQL Server 2008 Integration Services. 2009.
- [15] HOFFER, J.A, PRESCOTT, M.B. & MCFADDEN, F.R. Modern Database Management. Upper Saddle: Pearson Prendice Hall, 2005.
- [16] INMON, W.H. Building the Data Warehouse. Indianapolis. Wiley Publishing, Inc. 2005
- [17] JOURDAN, Z., RAINER R.K. & MARSHALL, T.E. Business Intelligence: An Analysis of Literature. Information System Management. 2008.
- [18] KIMBALL R. & ROSS, M. The Data Warehouse Toolkit, Second Edition, The Complete Guide to Dimensional Modeling. Ed. Wiley Computing Publishing, 2002
- [19] KIMBALL, R., ROSS, M., THORNTHWAIT, W., MUNDY, J. & BECKER, B. The Data Warehouse. Lifecycle Toolkit. Second Edition. Wiley, 2007.
- [20] KRIZAN, L. Intelligence Essencials for Everyone. Directions Magazine, October, www.directionsmag.com. 2006

- [21] LLORENTE, M., SIGURA, A., HADAD, A. & DROZDOWICZ, B. Sistema de Soporte a la Toma de Decisiones basado en datawarehouse para pacientes diabéticos. XIV Workshop de Investigadores en Ciencias de la Computación. 2012.
- [22] MENDEZ, A., MÁRTIRE, A., BRITOS, P. & GARCIA MARTINEZ, R. Fundamentos del Data Warehouse. Instituto Tecnológico de Buenos Aires. Reportes Técnicos en Ingeniería de Software. ISSN: 1667-5002. 2003.
- [23] ORTEGA CASTRO, I. Los sistemas de información gerencial y el esquema de la base de datos en la Asociación Mutualista Ambato. Facultad de Ingeniería en Sistemas, Electrónica e Industrial. 2014.
- [24] PADILLA, A. & MARTÍNEZ, R. Aplicación de las tecnologías datawarehouse en el contexto de la empresa turística de alojamiento hotelero. Universidad de Las Palmas de Gran Canaria, 1999.
- [25] PERALTA, V. Diseño Lógico de Data Warehouses a partir de Esquemas Conceptuales Multidimensionales. Universidad de la República, Uruguay. 2001.
- [26] PERALTA, V. & RUGGIA, R. Implementación de herramientas CASE que asistan en el Diseño de Data Warehouses. Universidad de la República, Uruguay. 2010.
- [27] ROJAS, M. Monografía de Adscripción: Data Warehouse. Licenciatura en Sistemas de Información. Corrientes-Argentina, 2009
- [28] ROZENFARB, A. Business Intelligence - Toma de Decisiones - Creación de valor. Marco Conceptual Formativo para el Informático. Centro de Altos Estudios en Tecnología Informática CAETI. Universidad Abierta Interamericana Buenos Aires – Argentina, 2009.
- [29] SABINO, C. Cómo hacer una tesis. Ed. Humanitas. México, 2005.
- [30] TURBAN, E., ARONSON, J.E., LIANG, T. & SHARDA, R. Decision Support Business Intelligence Systems. Upper Saddle River: Pearson Prendice Hall, 2007.
- [31] VALENGA, F., FERNANDEZ, E., MERLINO, H., RODRIGUEZ, D., PROCOPIO, C., BRITOS, P. & GARCIA MARTINEZ, R. Minería de datos aplicada a la detección de patrones delictivos en Argentina. Universidad de Morón. 2007.
- [32] VITT E., LUCKVICH M. & MISNER S. Business Intelligence Técnicas de Análisis para la Toma de Decisiones Estratégicas. España. McGraw Hill. 2002.
- [33] YUPA, E. Propuesta Metodológica para Aplicar Business Intelligence caso práctico “COHERVI S.A.”. Facultad de Informática y Electrónica. 2009.
- [34] WESKE, M. Business Process Management: Concepts, Languages, Architectures. Springer-Verlag Berlin Heidelberg. 2007.
- [35] ZAMUDIO, E., VIALEY, M. & KUNA, H. OLAP en las PyMEs. Un modelo multidimensional común para empresas del sector yerbatero del nordeste argentino. CACIC Congreso Argentino de Ciencias de la Computación. 2011.
- [36] ZAPATA, R. Implantación en una empresa de un sistema Business Intelligence SaaS / On Demand a través de la plataforma LITEBI. Universidad Politécnica de Valencia. 2009/2010.
- [37] ZORRILLA, M. Data Warehouse y OLAP. Universidad de Cantabria, 2010/11.