

Universidad de las Ciencias Informáticas

Facultad 7



**Trabajo de diploma para optar por el título de
Ingeniero en ciencias informáticas**

**Implementación de un Data Warehouse para el
control del Recurso Humano de la Salud**

Autores: Ariagna Ramón Cueto

Joannis Díaz García

Tutores: Lic. Roberto Acosta González

Ing. Yenisel Valido Pérez

Ciudad de La Habana, Junio de 2009

“Año del 50 aniversario del triunfo de la Revolución”

DECLARACIÓN DE AUTORÍA

DECLARACIÓN DE AUTORÍA

Declaramos que somos los únicos autores de este trabajo y autorizamos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los 23 días del mes de junio del año 2009.

Autora: Ariagna Ramón Cueto.

Autor: Joannis Díaz García.

Tutor: Lic. Roberto Acosta González

Tutora: Ing. Yenisel Valido Pérez.

DATOS DE CONTACTO

Tutores:

Lic. Roberto Acosta González: Graduado en la especialidad de Ciencias de la Computación en el año 2007. Imparte la asignatura de Programación. Ha presentado ponencias y trabajos en eventos científicos obteniendo diferentes reconocimientos y premios. Profesor con categoría docente Instructor. Es líder del Proyecto “Colaboración Médica” y dirige el grupo de trabajo de base de datos del Área Temática Sistema de Apoyo a la Salud en la Facultad número siete de la Universidad de las Ciencias Informáticas.

Correo electrónico: raglz@uci.cu

Ing. Yenisel Valido Pérez: Graduada de Ingeniería Informática en el Instituto Superior Politécnico “José Antonio Echeverría” (ISPJAE) y la Universidad de Cienfuegos “Carlos Rafael Rodríguez”. Profesora con categoría docente Instructor. Pertenece al departamento de Ingeniería y Gestión de Software, imparte la asignatura de Ingeniería de Software y se desempeña como jefa de la misma en la facultad número siete de la Universidad de las Ciencias Informáticas. Profesora del Área Temática Sistemas de Apoyo a la Salud, líder del proyecto Docencia Médica.

Correo electrónico: yenisel@uci.cu

Quisiera comenzar agradeciendo a mis padres por ser las personas que me dieron vida y sin las cuales esta realización no sería posible. A mi madre, que ha sido mi impulso de siempre, mi fuerza en todo momento y mi eterno apoyo, la mujer más hermosa e increíble del mundo, a ella le agradeceré siempre sus consejos y su amor. A mi padre, que es mi ejemplo y orgullo en la vida profesional, que nunca ha dejado de cuidarme y preocuparse. Esta tesis, es para que se sientan orgullosos de su hijo que tanto los quiere, por ellos han sido todos estos años de estudio.

A mi familia toda, a mi abuela María, que con su crianza me hizo la persona que soy hoy, en especial a mi hermano que es mi competencia más fuerte y una de las personas que más quiero, a mi hermana, a mi tía Lourdes, mi segunda madre, mis primos Rosa y Gilbert y a mi tío Gilberto por su afecto, por enseñarme e inculcarme la alegría de vivir.

Agradecer a mis amigos que no son menos importantes, a mis hermanos adoptivos, Yojan, Eric y Alejandro, que sin ellos esta escuela hubiese sido lo más aburrido y triste del mundo a mis amigos de siempre Reinel, Arianna (nuestra querida enanita) y Juliet, que siempre estuvieron ahí para cualquier cosa que necesitara. También a mis compañeros de aula, los de todos los años que pasé en esta lugar, a todos los miembros de Macagua 8 la gente más divertida que conozco, gracias por ayudarme a pasarla bien cada festival que estuvimos juntos.

Zuiero agradecer también a mi compañera de tesis, mejor no pudo haber sido. A mi tutor, el mejor a mi entender, que siempre nos apoyó y ayudó en todo momento gracias Robertón, especialmente al profesor Catalá por todo su esfuerzo, al tribunal de esta tesis y todo aquel que tuvo que ver durante todos estos años, para que yo pudiera estudiar sin costo alguno y sin preocupaciones que las que impulsan a hombres a seguir por el camino que entienden mejor, gracias a la revolución.

Por ultimo, pero no menos importante que cualquier persona de las anteriormente mencionadas, quisiera agradecer a un amigo especial que no esta entre nosotros hoy, pero del que estoy completamente seguro que todos recordaran siempre, a Manuel Alejandro Buch Meneses.

*Gracias a todos.
Joannis Díaz García.*

A mis padres que me han dado tantos buenos consejos y me han enseñado tantas cosas, ocultándome siempre sus tristezas y buscando todas las maneras posibles de hacerme el estudio fácil y la vida agradable. Ustedes que han sufrido todas mis amarguras, trabajando y llorando conmigo. Les doy gracias y les dedico este fruto, se que la distancia no ha sido fácil, pero valió la pena.

A mis segundos padres, mis suegritos y mi cuñadito lindo que con tanto amor y cariño me han tratado a lo largo de estos cinco años. Gracias por apoyarme en todo y darme fuerzas cuando lo necesitaba, agradezco por tenerlos siempre ahí, quiero que sepan que los quiero mucho. Sin ustedes este sueño hoy no hubiera sido posible.

A mi tía por sus acertados consejos, su apoyo y comprensión en todo momento, por poder contar con ella siempre, te agradezco mucho que estés hoy aquí conmigo.

A mi querida hermana y sobrinas que aunque no estén hoy, les doy gracias por todo el apoyo, cariño que me dan y la confianza que tienen en mi.

Quiero agradecer también a mi compañero de tesis (Turino), de veras que eres un muchacho como muy pocos, eres especial. Fue un placer haber trabajado a tu lado. A mi tutor, quiero que sepa que aprendí mucho con usted y fue un inmenso placer, gracias por su apoyo y comprensión en cada momento. A un profesor que puso mucho empeño en esta tesis, Catalá, gracias por tu trabajo, esfuerzo y dedicación eres parte de este resultado.

Agradecer a mis amigos que me han soportado a lo largo de estos años y me han tendido la mano en el momento preciso, Eli, Yaro, Suzy, gracias por estar ahí siempre que las necesito. Y a un amigo de todos que supo dar en vida lo mejor de él, Manuel Alejandro Buch Meneses, aunque no estés físicamente siempre vivirás en nuestros corazones, gracias por los buenos momentos que pasamos contigo amigo.

A mi novio por apoyarme y permitir conocer a esa familia maravillosa que son, le doy gracias en estos cinco años por ser más que mi pareja, mi amigo, te quiero.

Los quiero mucho a todos.

Ariagna Ramón Cueto.

Para mis padres y mi familia.

Joannis Díaz García.

A cuantos me han incitado y ayudado a aprender de un modo o de otro.

Ariagna Ramón Cueto.

RESUMEN

El presente trabajo tiene como objetivo desarrollar un Data Warehouse para consolidar la información relacionada con los temas de Colaboración y Docencia, específicamente sobre postgrados. Surge por la necesidad que presenta el Observatorio Nacional de Recursos Humanos de la Salud de interrelacionar estos dos temas para obtener una visión de la información histórica, de vital importancia para diferentes organismos del país que necesitan un control de sus colaboradores y su formación docente.

Actualmente esta interrelación se produce manualmente siendo esto un complejo trabajo y por ello no se obtiene toda la información necesaria y el histórico deseado en el momento requerido. El Data Warehouse para el control del Recurso Humano de la salud está basado en la metodología Hefesto, su modelamiento, análisis y posterior implementación son consecuentes con esta. El proceso de construcción consta de varias fases siendo las más importantes el proceso de extracción, transformación y carga junto al procesamiento analítico en línea.

El Data Warehouse brinda beneficios para el sistema de salud nacional, sobre todo para el Observatorio Nacional de los Recursos Humanos de la Salud. Su utilización permite a los especialistas: concebir al Recurso Humano como ente único en el sector de Salud. Conocer la cantidad, por ciento y total de estos, clasificándolos por diversos criterios y realizando combinaciones entre estos. Además de representar gráficamente los resultados. Todo ello posibilita una mejor decisión sobre la asignación de los RRHH donde realmente sea necesario según su formación postgraduada y experiencia en la actividad de la colaboración.

ÍNDICE

INTRODUCCIÓN	5
CAPÍTULO 1 .FUNDAMENTACIÓN TEÓRICA	9
1.1.4. Ministerio de Comercio Exterior (MINCEX):	10
1.2. Propósito de un Data Warehouse.	10
1.2.1. ¿Por qué Data Warehouse y no Sistema Operacional?	11
1.5. Modelamiento multidimensional.....	15
1.6. Proceso OLAP.....	19
1.6.1. ROLAP	21
1.6.2. MOLAP.....	22
1.6.3. HOLAP	22
1.7. Sistemas existentes.....	22
1.7.1. Nivel internacional	22
1.7.2. Nivel nacional	23
1.7.3. Universidad de la Ciencias Informáticas (UCI).....	23
1.8.1. OLTP	24
1.8.2. Herramientas de consulta y análisis.....	25
1.9. Metodología para el Diseño e Implementación del DWH.	25
1.10. Herramientas propuestas para el desarrollo del Data Warehouse	27
CAPÍTULO 2. ANÁLISIS Y DISEÑO	31
2.1. Análisis del proceso.....	31
2.2. Análisis de Requerimientos.....	32
2.2.1. Identificadores y perspectiva.....	33
2.2.1.1. Modelo Conceptual	33
2.2.2. Análisis de los OLTP	34
2.2.3. Elaboración del Modelo Lógico de la estructura del DWH.....	38
2.2.4. Proceso ETL.....	39
CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA	46
3.1. Determinar jerarquías	46
3.2. Medidas	47
3.3. Consultas.....	48
3.4. Mondrian.....	49
3.4.1. JPivot.....	49

3.4.2. JRubik	52
3.5. Despliegue de la solución	56
CONCLUSIONES	58
BENEFICIOS	59
RECOMENDACIONES	60
REFERENCIAS BIBLIOGRÁFICAS	61
BIBLIOGRAFÍA	64
ANEXOS.....	66
1. Ejemplos de Data Warehouse.....	66
1.1. Corporación CIMEX.....	66
1.2. Data Warehouse de CUBAENERGIA	67
2. Proceso ETL.....	67
2.1. Dimensión Especialidad.....	67
2.2. Dimensión Estado.....	67
2.3. Dimensión Localización	68
2.4. Dimensión Misión.....	68
2.5. Dimensión Postgrado.....	68

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

INTRODUCCIÓN

Con la informatización de la sociedad ha crecido a nivel mundial la capacidad de generación y almacenamiento de la información. Lo que ha provocado que esta no pueda ser analizada por los métodos tradicionales existentes, mientras mayor es la capacidad para almacenar más y más datos, mayor es la incapacidad para extraer información realmente útil. Mucha información importante, quedaba sepultada y disgregada, y los sistemas existentes no estaban preparados para el nuevo reto.

Mucho se ha hablado de la Era de la Información y sus ventajas; con las nuevas posibilidades se acortan las distancias y crecen los beneficios para quienes tienen acceso al gran caudal de datos. Sin embargo, lo que constituye un valioso recurso para todos, se ha tornado el gran problema de principios de siglo, manejar de forma óptima grandes volúmenes de información.

Los sistemas relacionales no están preparados óptimamente para extraer conocimiento de los datos almacenados, afirmación dada por cierta en el mundo de hoy. Los sistemas relacionales, por mucho que se les mejore y que se les quiera adaptar, no son los encargados de esa nueva fase de consulta, para eso ha surgido una nueva forma de organización de los datos, los Data Warehouse (DWH) también conocido como Almacén de Datos, estos son el centro de la arquitectura para los sistemas de información a partir de la década de los '90 hasta la fecha. [1]

Existen diferentes definiciones sobre DWH, la más conocida fue propuesta por Inmon (considerado el padre de las Bases de Datos, uno de los primeros autores en escribir sobre el tema de los almacenes de datos) en 1992: "Un Data Warehouse es una colección de datos orientados a temas, variante en el tiempo, integrados y no-volátiles, organizados para soportar necesidades empresariales". Puesto que los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí, los cambios producidos en ellos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones, la información no se modifica ni se elimina, una vez almacenado un dato, este se convierte en información de sólo lectura, y se mantiene para futuras consultas. Las bases de datos contienen los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes. [2]

Los Data Warehouse proveen dos beneficios empresariales reales: Integración y Acceso de datos. El almacén de datos elimina una gran cantidad de información inútil y no deseada, como también el

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

procesamiento desde el ambiente operacional clásico. Evidenciándose claramente el principal beneficio que este aporta, eliminar aquellos datos que obstaculizan la labor de análisis de información y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión. [3]

Los Data Warehouse surgen con la promesa del manejo y control de la información. Ellos aseguran una vista única de los datos, que pueden provenir de diversas fuentes. Gracias a esto, los usuarios finales no se ven en la necesidad de aprender y utilizar múltiples sistemas de acceso y manipulación de los datos. Un almacén de datos facilita la comprensión de los mismos, transformándolos en información útil, teniendo como bandera el apoyo a la toma de decisiones. Se puede comprender el carácter universal de esta herramienta, que es útil en toda empresa, entendiéndose a partir de aquí por empresa, compañía o negocio a toda organización de producción o de servicios con fines lucrativos o no.

Las empresas del mundo entero están cada vez más interesadas en integrar y centralizar sus datos para reducir costes y conseguir una visión completa de sus recursos humanos, productos, riesgos y finanzas, ya que en la actualidad se ha hecho fundamental el uso que se le da a los datos para la toma de decisiones.

Actualmente el Observatorio Nacional de Recursos Humanos de la Salud encuentra un obstáculo en el gran volumen, complejidad y diversidad de la información generada por los temas de colaboración y docencia postgrado. Para llevar a cabo la interrelación entre estos dos temas se manipula dicha información manualmente. Lo que limita el analizar e interpretar dichos datos y extraer conocimiento realmente útil, ya que resulta un trabajo complejo y muy agotador debido al gran volumen de información.

Cuba demuestra ante el mundo la posibilidad de brindar salud de calidad a todo aquel que lo necesite. Lo anterior queda evidenciado cuando se conoce que el país cuenta con más de 32 mil colaboradores de la salud prestando sus servicios en 76 países del mundo [4] y que de esta cifra muchos están involucrados en cursos de postgrado y otras investigaciones de carácter docente.

El prestigio de Cuba es cada día mayor y este sería mucho más efectivo si los colaboradores de la salud enviados a cumplir misión internacionalista tuviesen un mayor nivel de especialización y de grado académico. Todos estos datos de interés, junto a otras cifras se manejan frecuentemente por organismos del estado (Ver Fig. 1) como son el Ministerio de Relaciones Exteriores (MINREX),

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

Ministerio de Comercio Exterior (MINCEX), Ministerio de Salud Pública (MINSAP). Resultan necesarios para encomendar las futuras misiones a cumplir por el personal de salud dentro y fuera de las fronteras cubanas, sin embargo, la demora en el resultado de la información es bastante y los datos no están actualizados cercanos a la fecha de solicitud.



Fig. 1 Instituciones involucradas.

A partir de la situación descrita anteriormente se tiene como **problema a resolver** la insuficiente información consolidada que relacione los temas de colaboración y docencia postgrado en el Observatorio Nacional de Recurso Humano de la Salud.

El presente trabajo tiene como **objeto de estudio** el proceso de consultoría de la información en el Observatorio Nacional de Recursos Humanos de la Salud. A partir del objeto de estudio se determinó el **campo de acción** como el proceso de consultoría de la información de Recursos Humanos de la Salud en los temas de colaboración y docencia postgrado.

Por lo antes planteado, se definió como **objetivo de la investigación** implementar un Data Warehouse que contenga la información consolidada de los temas de colaboración y docencia postgrado para el Observatorio Nacional de Recursos Humanos de la Salud.

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

Para desarrollar el trabajo de diploma, se proponen las siguientes **tareas de la investigación**:

1. Asimilar las herramientas, tecnologías y tendencias actuales, propuestas para el desarrollo del Data Warehouse.
2. Utilizar la metodología HEFESTO para el diseño e implementación del Data Warehouse.
3. Realizar análisis del negocio de Docencia sobre la gestión de postgrado.
4. Identificar Perspectivas e Indicadores que relacionen los temas.
5. Implementar la base de datos utilizando el gestor de base de datos PostgreSQL v8.3 sobre Debian.
6. Realizar el proceso de ETL (Extracción, Transformación y Carga) utilizando una herramienta libre sobre Linux.
7. Realizar el proceso OLAP (Procesamiento Analítico en Línea) utilizando una herramienta libre sobre Linux.

El presente trabajo está estructurado en **3 capítulos**, distribuidos de la siguiente manera:

En el primer capítulo **Fundamentación Teórica** se realiza una observación sobre el estado del arte del tema tratado a nivel nacional, internacional y de la Universidad. Se describen las tecnologías, metodología y herramientas utilizadas para la solución del problema.

En el segundo capítulo **Análisis y Diseño** se muestran los resultados y pasos seguidos para el análisis según los describe la metodología utilizada y el diseño de la base de dato del DWH. El análisis de la metodología es un paso fundamental para el desarrollo del mismo y bajo esta guía se construyó un modelo conceptual a partir de los indicadores y perspectivas obtenidos del negocio. Para obtener una mayor visión de lo que se realiza, se llevó a cabo un modelo de procesos, el cual conforma toda una serie de pasos o iteraciones que serán la guía para concluir con el funcionamiento del Data Warehouse.

El tercer capítulo **Proceso Analítico en Línea** expone la construcción del cubo de datos, donde se definen las jerarquías de los atributos de las dimensiones a las que pertenecen, y la implementación de los indicadores o medidas. Se muestran las herramientas clientes disponibles para que los usuarios puedan consultar la información almacenada en el Data Warehouse, y es representado mediante el diagrama de despliegue los requerimientos tecnológicos y la integración entre el Data Warehouse y las base de datos de los Sistema de Colaboración y Docencia Médica.

CAPÍTULO 1 .FUNDAMENTACIÓN TEÓRICA

En este capítulo se realiza una observación sobre el estado del arte del tema tratado a nivel Internacional, Nacional y de la Universidad. Se describe la justificación de las tecnologías, metodología y herramientas utilizadas para la solución del problema.

1.1. Instituciones Relacionadas [5]

1.1.1. Observatorio Nacional de Recursos Humanos (RRHH) de la Salud:

Es el departamento encargado de supervisar y controlar el Recurso Humano de la Salud en Cuba.

1.1.2. Ministerio de Salud Pública (MINSAP):

El Ministerio de Salud Pública (MINSAP) es el Organismo rector del Sistema Nacional de Salud, encargado de dirigir, ejecutar y controlar la aplicación de la política del Estado y del Gobierno en cuanto a la Salud Pública, el desarrollo de las Ciencias Médicas y la Industria Médico Farmacéutica.

Tiene como funciones rectoras:

- Ejercer el control y la vigilancia epidemiológica de las enfermedades y sus factores de riesgo.
- Ejercer el control y la vigilancia sanitaria de todos los productos que pueden tener influencia sobre la salud humana.
- Regular y controlar la aprobación, ejecución y evaluación de las investigaciones biomédicas o de cualquier tipo que se realice directamente en seres humanos.
- Normar las condiciones higiénicas y el saneamiento del medio ambiente en aquellos aspectos que puedan resultar agresivos a la salud humana y controlar su cumplimiento a través de la inspección sanitaria estatal.
- Regular el ejercicio de la medicina y de las actividades que le son afines.
- Ejercer la evaluación, el registro, la regulación y el control de los medicamentos de producción nacional y de importación, equipos médicos y material gastable y otros de uso médico.

1.1.3. Ministerio de Relaciones Exteriores (MINREX):

La misión del Ministerio de Relaciones Exteriores es ejecutar la política exterior de Cuba, contribuir a su elaboración, promover y defender en las relaciones internacionales los principios, valores e intereses que la sustentan. La actuación del Ministerio se basa en los lineamientos e instrucciones que emanan de la Asamblea Nacional del Poder Popular como órgano supremo de la República, del Consejo de Estado y del Consejo de Ministros; es, según establece la Constitución de la República, el órgano rector de la política exterior de Cuba y trabaja en estrecha coordinación con los demás Organismos del Estado.

1.1.4. Ministerio de Comercio Exterior (MINCEX):

El Ministerio de Comercio Exterior es el organismo encargado, sobre el desarrollo de la base de las estrategias y la política establecida por el Estado y el Gobierno, de promover la inversión extranjera en Cuba. Dirige en coordinación con los organismos competentes, los procesos de negociación correspondientes y, coordinar y asesorar la instrumentación y cumplimiento de la política del Estado y el Gobierno en materia de colaboración económica.

1.2. Propósito de un Data Warehouse.

Un Data Warehouse o Almacén de Datos, surgió con el objetivo de hacer consultable la información que se tiene de una empresa tanto de meses como de años anteriores. Es un almacenamiento simple, completo y consistente de datos obtenidos desde una variedad de fuentes, disponibles para el usuario final de forma tal que puedan entenderlos y utilizarlos. [6]

Facilita la integración de sistemas de aplicación no integrados. Un DWH soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis que se necesita de la información. El objetivo fundamental es asistir al usuario en el entendimiento del pasado y contar con los elementos para la planeación del futuro de corto, mediano y largo plazo. Es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo registrada por las diversas aplicaciones operacionales [7].

1.2.1. ¿Por qué Data Warehouse y no Sistema Operacional?

Generalmente los OLTP (On-Line Transactional Processing) o procesamiento en línea transaccional, usan estructuras normalizadas, en las cuales se optimizan las inserciones y actualizaciones de artículos e incluso algunas selecciones, pero es menos probable que el sistema se organice de forma tal que produzca reportes eficientes para datos resumidos con cierta jerarquía. Una solución a este problema son los DWH, que usa los datos relevantes de fuentes existentes y los combinan en una estructura que ha sido optimizada para las selecciones. [8]

Se puede justificar su construcción y utilización en el Observatorio Nacional de Recursos Humanos de la Salud ya que la necesidad del mismo es obtener información con una visión histórica de la interrelación de colaboración y docencia que sea de ayuda para observar el comportamiento de estos dos temas, su propósito no es insertar ni modificar la información que es gestionada por las Unidades de Colaboración Médica y los Institutos de Ciencias Médica.

1.3. Objetivos fundamentales de un Data Warehouse: [9]

- **Hace que la información de la organización sea accesible:**

Los contenidos del Data Warehouse son entendibles, navegables, y el acceso a ellos es caracterizado por el rápido desempeño. Estos requerimientos no tienen fronteras y tampoco límites fijos. Cuando se habla de entendible, significa que los niveles de la información sean correctos y obvios. Navegables significa el reconocer el destino en la pantalla y llegar a donde se quiera. Rápido desempeño significa, menor tiempo de espera.

- **Es información adaptable y elástica:**

El Data Warehouse está diseñado para cambios continuos. Cuando se le hacen nuevas preguntas, los datos existentes y las tecnologías no cambian ni se corrompen.

- **Es un seguro baluarte que protege los valores de la información:**

No solamente controla el acceso efectivo a los datos, si no que da a los dueños de la información gran visibilidad en el uso de los datos, aún después de haber dejado el Data Warehouse.

- **Es la fundación de la toma de decisiones:**

Tiene los datos correctos para soportar la toma de decisiones. Solo hay una salida verdadera del DWH: las decisiones que son hechas después de que el Data Warehouse haya presentado las evidencias. La original etiqueta que preside el DWH sigue siendo la mejor descripción de lo que se quiere construir: un sistema de soporte a las decisiones.

1.4. Características del Data Warehouse [10]

Entre sus principales características se encuentran:

- **Orientado al tema**

La información en los Data Warehouse se clasifica en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados, están en contraste con los clásicos procesos orientados a las aplicaciones. El ambiente operacional está diseñado alrededor de las aplicaciones en este caso en los sistemas de Colaboración y Docencia con atributos tales como tipo_programa, nombres, salario, nota y pasaporte para una institución o ministerio como el Observatorio Nacional de Recursos Humanos o el MINSAP (ver Fig. 2). La base de datos combina estos elementos en una estructura que acomoda las necesidades de la aplicación. En el ambiente DWH se organiza alrededor de sujetos tales como colaborador, misión, programa y estado siendo sujetos involucrados en el proceso de consultoría de información en el observatorio de los recursos humanos de la salud.

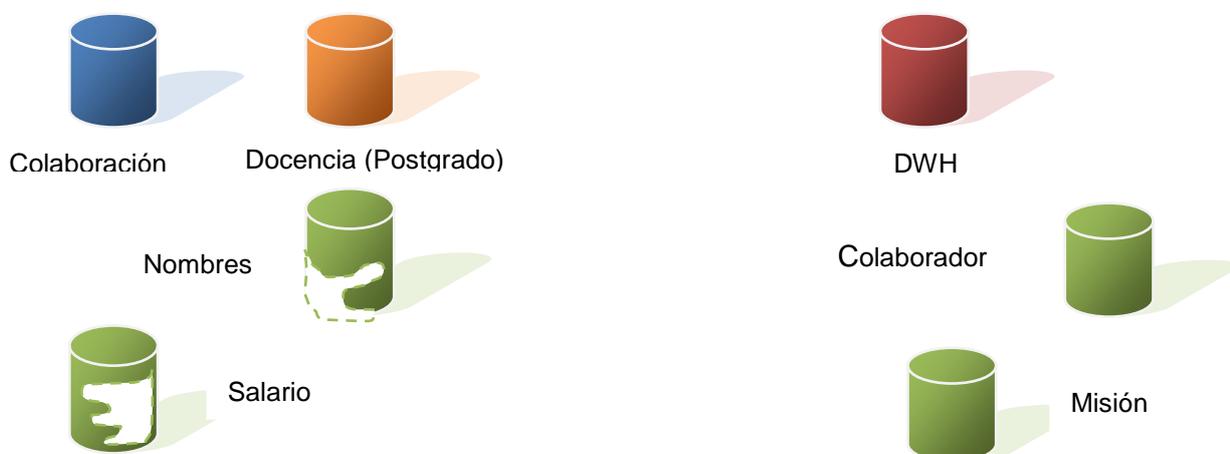




Fig. 2 Contraste entre los dos tipos de orientaciones.

- **Integrado**

Integra datos recolectados de diferentes sistemas operacionales de la organización y/o fuentes externas, ejemplo: Colaboración y Docencia Médica en el tema de postgrado. El aspecto más importante del ambiente DWH es que la información encontrada en el interior esté siempre integrada. Ejemplificando lo anterior los diseñadores de las aplicaciones a las que tributan los dos sistemas operacionales que se quieren integrar codifican de manera igual los identificadores propiciando que estos se integren sin tener que transformar en este caso. Vale destacar que no es importante el estilo o la forma en que llegan al almacén de datos, solamente la integración de estos campos.

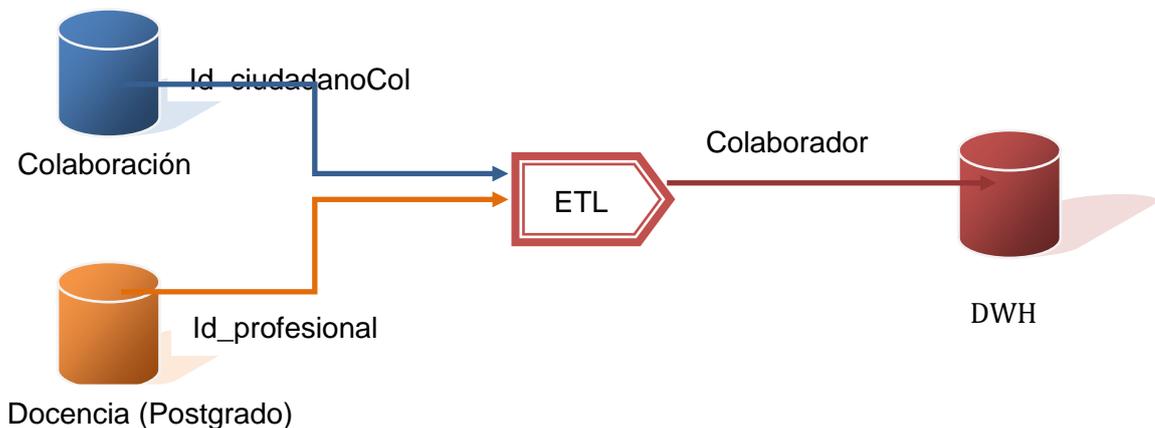


Fig. 3 Integración.

- **De tiempo variante**

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

Toda la información del DWH es requerida en algún momento. Esta característica básica, de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. En otras palabras, en los sistemas operacionales Colaboración y Docencia Médica en el tema de postgrado, cuando se quiere obtener la misión de un colaborador o el tipo de curso que está cursando, usted espera que los valores requeridos se obtengan a partir del momento de acceso. Como la información en el DWH es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante".

Los datos históricos son de poco uso en los OLTP ya que los datos obsoletos se van actualizando o eliminando o simplemente olvidando. La información del depósito por el contrario, si debe incluir los datos históricos para usarse en la identificación y evaluación de tendencias para la toma de decisiones, lo cual es uno de los principales objetivos del Observatorio Nacional de Recursos Humanos. (Ver Fig. 4).

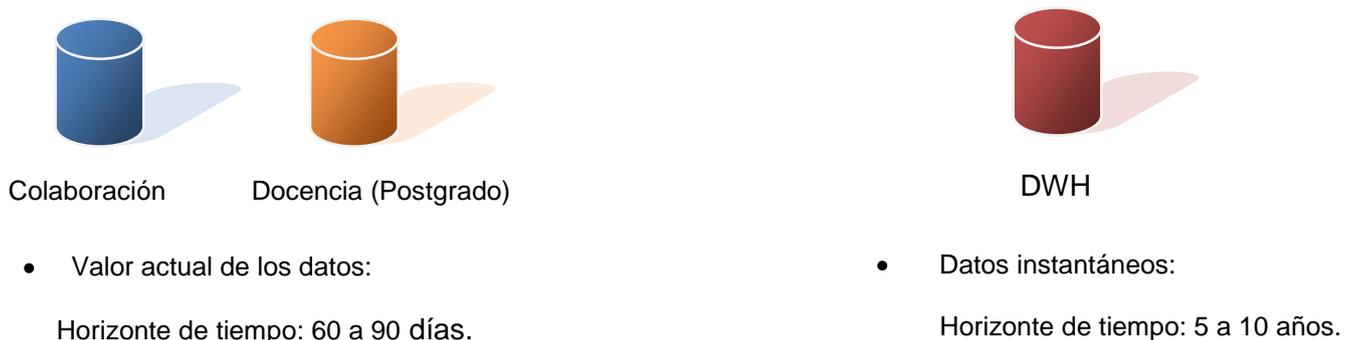


Fig. 4 Tiempo variante.

- **No volátil**

La información es útil sólo cuando es estable. Los datos en Colaboración y Docencia Médica en el tema de postgrado cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable. En la Fig. 5 se muestra que la actualización (insertar, borrar y modificar), se hace regularmente en el ambiente operacional sobre una base de registro por registro. Pero la manipulación básica de los datos que ocurre en el DWH que se desea implementar es mucho más simple. Hay dos únicos tipos de

operaciones: la carga inicial de datos y el acceso a los mismos por el Observatorio Nacional de Recursos Humanos de la Salud. No hay actualización de datos en el depósito, como una parte normal de procesamiento.

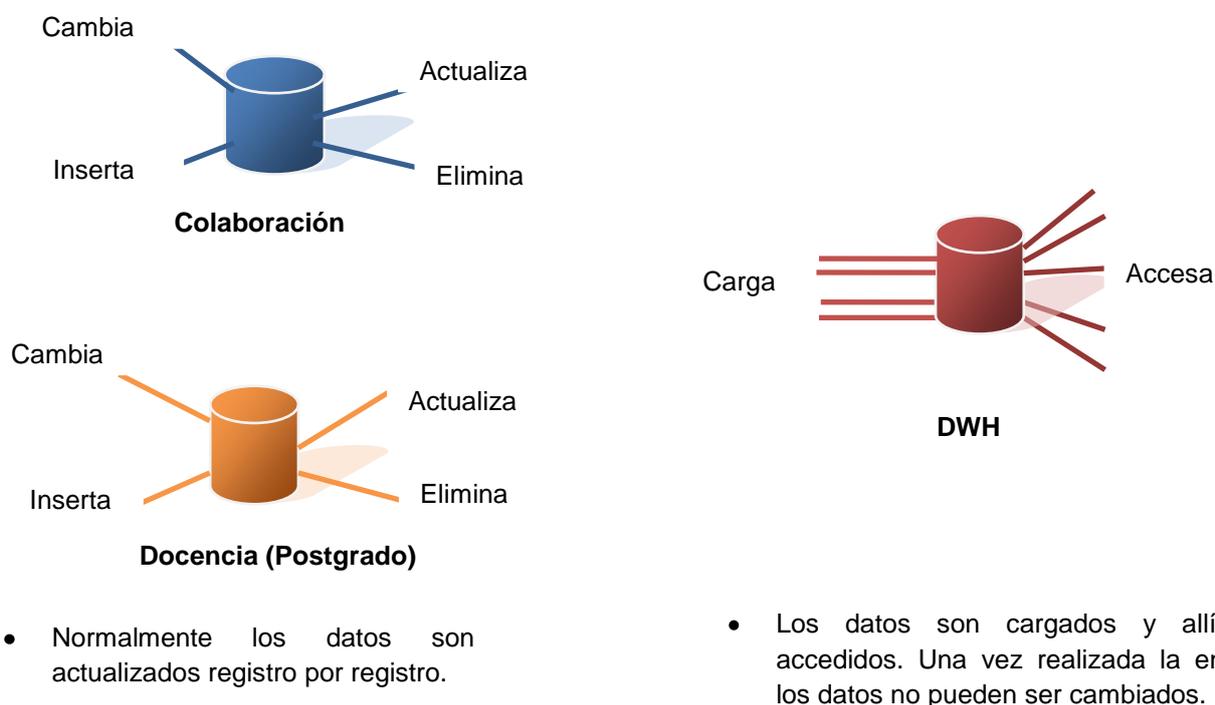


Fig. 5 No volátil.

1.5. Modelamiento multidimensional

Aquí se encuentra la principal diferencia entre los sistemas operacionales y el DWH: Cada uno de ellos es sostenido por un modelo de datos diferentes. Los sistemas operacionales se sustentan en el Modelo Entidad Relación (MER) y el DWH trabaja con el Modelo Multidimensional. [11] Los Data Warehouse gestionan el depósito de datos y lo organizan en torno a una base de datos multidimensional que tal y como lo indica su nombre almacena los datos en diversas dimensiones, que conforman un cubo multidimensional, en donde el cruce de los valores de los atributos de cada

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

dimensión a lo largo de las abscisas, determinan un hecho específico que se define como datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. (Ver Fig. 6) Esto ofrece al usuario su visión respecto a la operación del negocio, es más fácil para el usuario organizar la información e imaginarse en ella cortando y rebanando el cubo a través de cada una de sus dimensiones, para buscar la información deseada. [12]

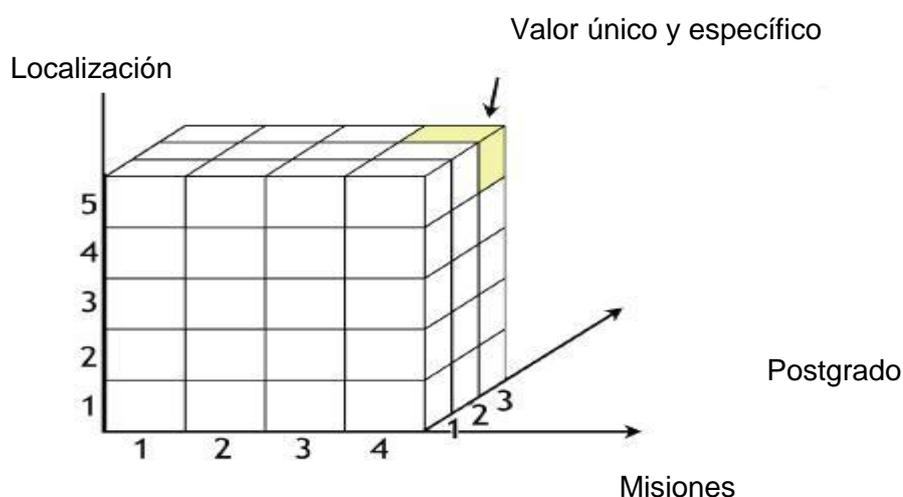


Fig. 6 Cubo Multidimensional.

El espacio se define a través de sus ejes coordenados (por ejemplo Localización, Misiones, Postgrado). Un punto cualquiera de este espacio quedará determinado por la intersección de tres valores particulares de sus ejes. Si se le asignan valores particulares a estos ejes. Se podría tener por ejemplo, la siguiente combinación: Localización = África, Misión = PIS (Programa Integral de Salud), Postgrado = Doctor en Oftalmología, se quiere saber la cantidad de Recursos Humanos RRHH que prestan servicios en África, en PIS y son Doctor en Oftalmología. La intersección de estos valores definirá un solo punto en el espacio, que será un valor específico y único para tal combinación (ver Fig. 6) [13]

Las bases de datos multidimensionales implican tres variantes posibles de modelación, que permiten realizar consultas de soporte de decisión.

- Esquema en estrella (Star Scheme).
- Esquema copo de nieve (Snowflake Scheme).
- Esquema constelación o copo de estrellas (Starflake Scheme).

Esquema en Estrella

Se definió como variante el Esquema en Estrella: Consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. [14]

Tablas de Hechos

Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados para apoyar el proceso de toma de decisiones, donde los hechos son datos instantáneos en el tiempo. [15]

El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos que representan la información acerca de los datos, son fáciles de documentar y mantener. [16] En la Fig.7 se puede apreciar un esquema en estrella estándar:

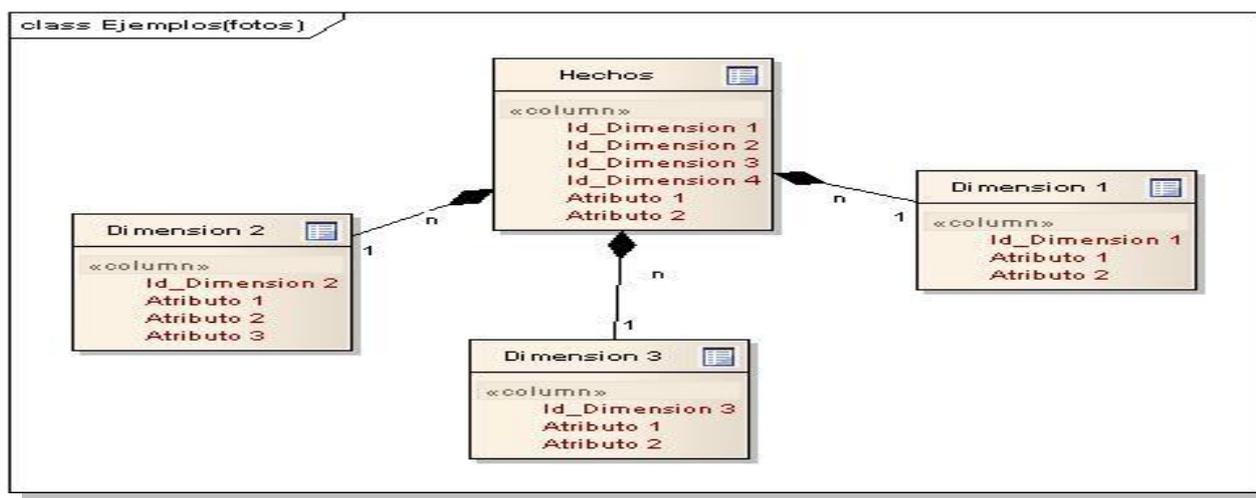


Fig. 7 Esquema en Estrella

A continuación se destacarán algunas características de este modelo, que ayudarán a comprender mejor el por qué de sus ventajas y de su selección para la confección de la base de datos multidimensional: [17]

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

Esquema Copo de Nieve

Este esquema representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones. [18]

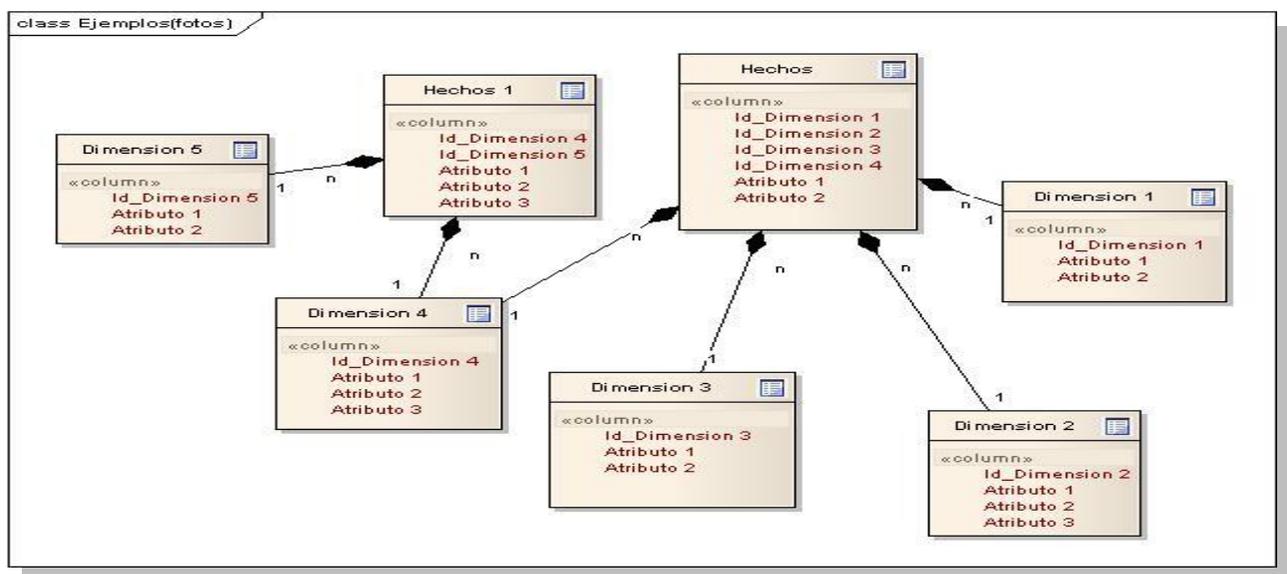


Fig. 8 Esquema Copo de Nieve.

Esquema Constelación

Este modelo está compuesto por una serie de esquemas en estrella, y tal como se puede apreciar (Ver Fig.9), 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. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones. [19]

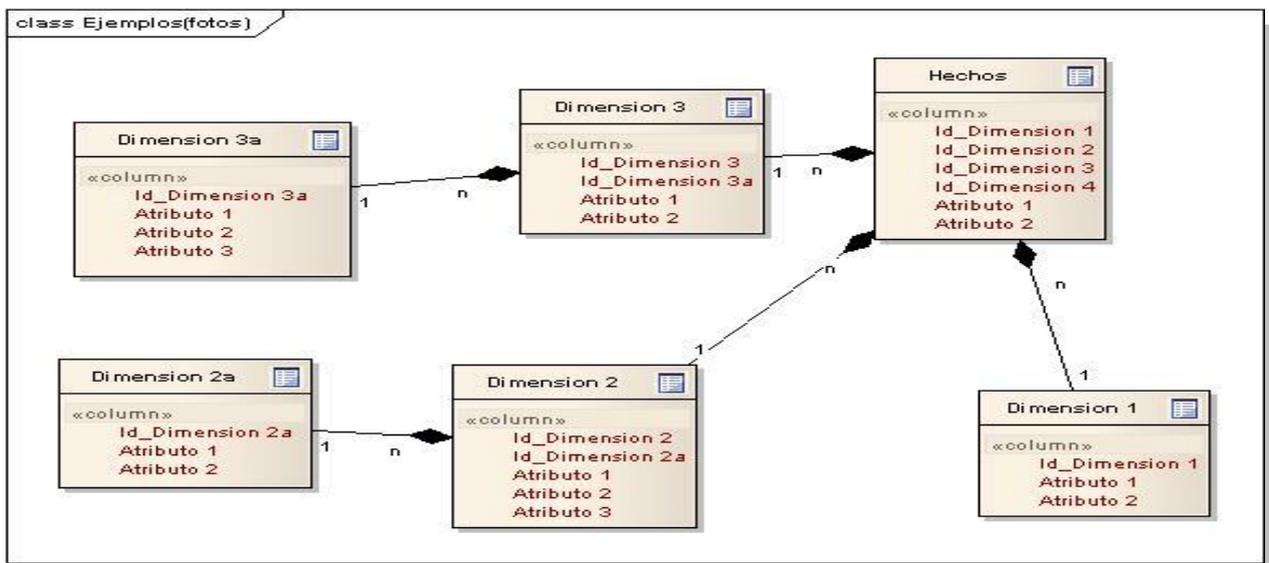


Fig. 9 Esquema Constelación.

1.6. Proceso OLAP

El procesamiento analítico en línea OLAP (On Line Analytic Processing), es la componente más poderosa de los DWH, ya que es el motor de consultas especializado del almacén de datos. Las herramientas OLAP, son una tecnología de software para el análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio. 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 operadores tales como drill-up, drill-down, drill through para explotar profundamente la información.

Este tipo de herramientas, puede analizar el negocio desde diferentes escenarios históricos, y proyectar como se ha venido comportando y evolucionando en un ambiente multidimensional, o sea, mediante la combinación de diferentes perspectivas, temas de interés o dimensiones. Esto permite

deducir tendencias, por medio del descubrimiento de relaciones entre las perspectivas que a simple vista no se podrían encontrar sencillamente. [20]

Drill-down: Permite apreciar los datos en un mayor detalle, bajando por la jerarquía de una dimensión, es decir por el orden aplicado a los atributos según el criterio seleccionado. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, lo que facilita y organiza aún más la búsqueda. [21]

Drill-down es ir de lo general a lo específico. Gráficamente:

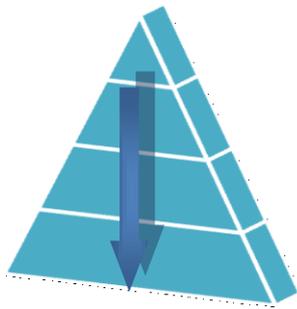


Fig. 10 Drill down

Drill-up: Permite apreciar los datos en menor nivel de detalle, subiendo por la jerarquía de una dimensión, es decir por el orden aplicado a los atributos según el criterio seleccionado. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, lo cual facilita y organiza aún más la búsqueda. [22]

Drill-up es ir de lo específico a lo general. Gráficamente:

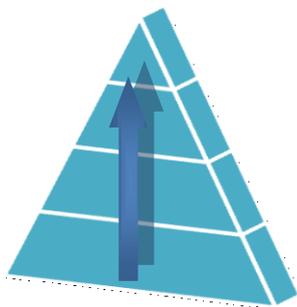


Fig. 11 Drill up

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

Drill throuhg: Permite expresar que existe un nivel de datos detallados, que resulta de ejecutar una consulta sobre la base de datos, recupera un conjunto de resultados de los datos de origen de una celda para obtener información más detallada.

Estas operaciones que se pueden realizar sobre modelos multidimensionales son las que verdaderamente les permitirán a los usuarios explorar y desmenuzar los datos en busca de respuestas.

Las herramientas OLAP requieren que los datos estén organizados dentro del depósito en forma multidimensional, por lo cual es que utilizan los cubos multidimensionales. Además de las características ya descritas, se pueden enumerar las siguientes: [23]

- Permite recolectar y organizar la información analítica necesaria para los usuarios y disponer de ella en diversos formatos, tales como tablas, gráficos, reportes, etc.
- Complementa las actividades de otras herramientas que requieran procesamiento analítico en línea.
- Presenta al usuario una visión multidimensional de los datos.
- Es transparente al tipo de tecnología que soporta el DWH, ya sea ROLAP, MOLAP o HOLAP.
- Permite a los usuarios, analizar la información basándose en más criterios que un análisis de forma tradicional.
- Permiten realizar agregaciones y combinaciones de los datos de maneras complejas y específicas, con el fin de realizar análisis más estratégicos.

1.6.1. 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, (procesamiento analítico en línea relacional), cuenta con todos los beneficios de una Sistema Gestor de Base de Datos (SGBD) Relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de DWH. [24]

1.6.2. MOLAP

Usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. El almacenaje de (MOLAP) Multidimensional Online Analytical Processing, procesamiento analítico multidimensional en línea, provee excelente rendimiento y compresión de datos. Como se dice, todo va en el cubo. Tiene el mejor tiempo de respuesta, dependiendo solo en el porcentaje y diseño de las agregaciones del cubo. En general este método, es muy apropiado para cubos con uso frecuente por su rápida respuesta. [25]

1.6.3. HOLAP

El desarrollo más reciente ha sido la solución OLAP híbrida (HOLAP) Hybrid Online Analytical Process, (procesamiento analítico en línea híbrido), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. Los cubos almacenados como HOLAP, son más pequeños que los MOLAP y responden más rápidos que los ROLAP. HOLAP es generalmente usado para cubos que requieren rápida respuesta, para sumalizaciones basadas en una gran cantidad de datos. [26]

Para el caso específico del Data Warehouse para el recurso Humano de la Salud se implementará el proceso analítico en línea utilizando ROLAP dado que la base de datos que se construirá como parte del almacén de datos esta soportada por el gestor de base de datos PostgreSQL, el cual es un sistema gestor de base de datos relacional.

1.7. Sistemas existentes

1.7.1. Nivel internacional

Las empresas del mundo entero están cada vez más interesadas en manejar de forma óptima el gran volumen de información que es generado en su accionar diario y así obtener una mejor visión para el apoyo a la toma de decisiones. Algunas de estas compañías son reconocidas internacionalmente por ser grandes corporaciones que manejan enormes cantidades de datos que le son imprescindibles para su correcto funcionamiento. En este caso se encuentran Adidas, Hewlett Packard y Sun, esta última con el Data Warehouse más grande del mundo y todas ellas gigantes corporativos que cuentan con el

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

capital suficiente para realizar una impecable inteligencia de negocio. Estas empresas construyen sus sistemas operacionales con software propietario, lo cuál no proporciona el enfoque más correcto a tener en cuenta, sin embargo brinda una visión muy clara de cómo ordenar la información y el orden de trabajo para la confección del Data Warehouse.

1.7.2. Nivel nacional

Las empresas cubanas no se quieren quedar atrás con el actual avance de las tecnologías, desean poseer un mejor manejo de sus datos y poseer mejor visión de los mismos. Ejemplos fehacientes de tal desempeño:

Data Warehouse implantado por la Corporación CIMEX que se dedica fundamentalmente a la Exportación e Importación de mercancías. Este centra su atención en la actividad del comercio, principalmente en la gestión de inventario, permitiendo una gestión de compra–venta eficiente, con una finalidad fundamental: “Disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida, aumentando las ganancias o utilidades de las empresas”. [\[Ver anexos 1.1\]](#)

Existencia en CUBAENERGIA de un Data Warehouse para coleccionar toda la documentación concerniente al sistema de información de CUBAENERGIA soportado básicamente en sus sistemas operacionales y aplicaciones en explotación. Los hechos modelados fueron: Ventas, Compras y Gestión de Recursos Humanos por Competencias y por Tiempo. [\[Ver anexo 1.2\]](#)

Estas empresas cubanas brindan un acercamiento mayor a la creación del Data Warehouse, los procesos de ETL y OLAP que son de gran valor para la investigación. Estos Data Warehouse de producción nacional ofrecen una visión detallada que contribuirá al desarrollo del presente trabajo aunque no estén enfocados precisamente al tema de recursos humanos.

1.7.3. Universidad de la Ciencias Informáticas (UCI)

Actualmente la Facultad 6 de esta Universidad se encuentra en el desarrollo de una tesis con el objetivo de llevar a cabo un Data Warehouse para el proyecto Lims de Control de Calidad. Esta acción resulta de vital importancia ya que existe una fuente cercana de gran ayuda para el intercambio de información, en cuanto a los procesos anteriormente descritos (ETL, OLAP) brindan así un mayor conocimiento respecto a la construcción del los DWH.

1.8. Arquitectura del Data Warehouse.

En este punto y teniendo en cuenta que ya se han detallado claramente las características generales del Data Warehouse, se describen y definen los componentes que intervienen en su arquitectura o ambiente.

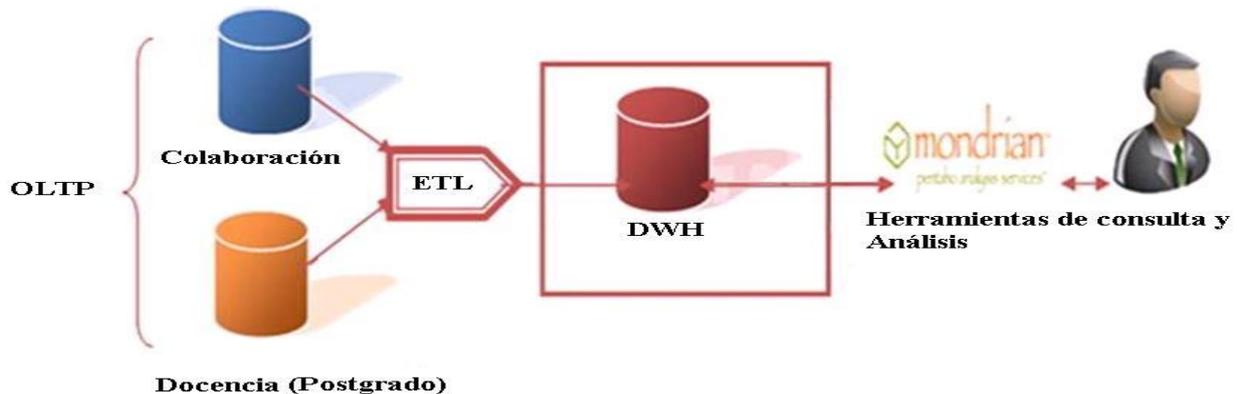


Fig. 12 Arquitectura del Data Warehouse.

1.8.1. OLTP

Procesamiento Analítico Transaccional (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, estas fuentes de información, son de características muy disímiles entre sí, en formato, procedencia, función, etc. [27]

Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran: [28]

- Archivos de textos.
- Hipertextos.
- Hojas de cálculos.
- Informes semanales, mensuales, anuales, etc.
- Bases de datos transaccionales.

Para poder extraer los datos desde los OLTP, para luego manipularlos, integrarlos y transformarlos, para posteriormente cargar los resultados obtenidos en el DWH, es necesario contar con algún sistema que se encargue de ello. Precisamente, los ETL (Extract, Transform and Load) (Extracción,

Transformación y Carga) son los que cumplirán con tal fin. ETL extrae los datos de las diversas fuentes que se requieran, los transforman para resolver posibles problemas de inconsistencias entre los mismos y finalmente, luego de su depuración se procede a la carga en el depósito de datos.

En síntesis, las funciones específicas de los ETL son tres:

- Extracción.
- Transformación.
- Carga.

Se explica en síntesis el accionar del proceso ETL, y la relación existente entre sus diversas funciones. En la Fig. 12 se puede apreciar mejor lo antes descrito: [29]

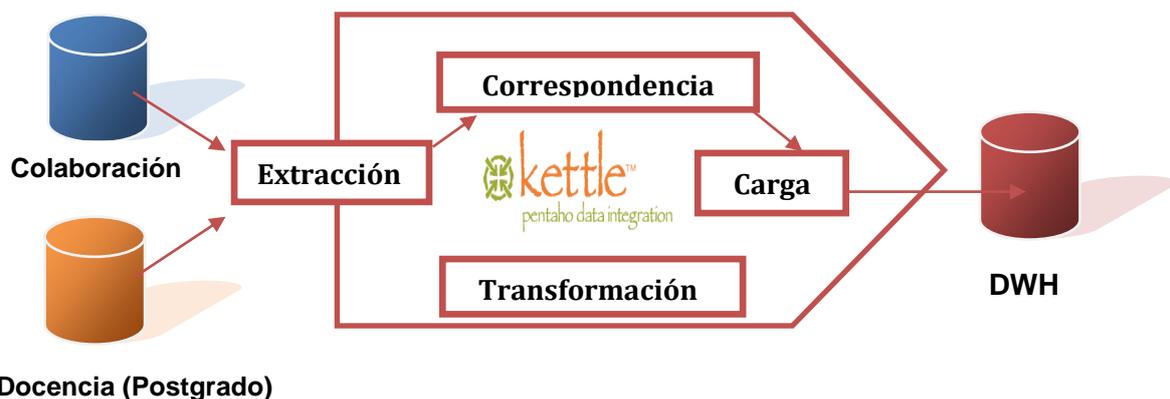


Fig. 13 Extracción, Transformación y Carga

1.8.2. Herramientas de consulta y análisis.

Las herramientas de consulta y análisis son sistemas que permiten al usuario realizar la exploración de datos del DWH. Básicamente constituyen el nexo entre el depósito de datos y los usuarios. A través de una interfaz gráfica y una serie de simples pasos, devuelve los resultados obtenidos que le fueron solicitados. Luego, estos resultados son expuestos ante el usuario en formatos que le son familiares. [30]

1.9. Metodología para el Diseño e Implementación del DWH.

La metodología a seguir es HEFESTO, que permite la construcción del Data Warehouse de forma sencilla, ordenada e intuitiva. Su nombre fue inspirado en el dios griego de la construcción y el fuego.

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

Hefesto es una metodología bien fundamentada [31] y explícita que permite facilitar un DWH de manera metódica y sencilla, guiándose por pasos lógicos relacionados sólidamente durante todas las etapas del proceso de confección. Se comienza recolectando las necesidades de información a partir de las preguntas, entrevistas o cuestionarios realizados a los usuarios.

Luego, se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del Data Warehouse, se analizarán los OLTP para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Una vez hecho esto, se pasará a la construcción del modelo lógico del depósito. Por último, se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente.

La metodología cuenta con las siguientes características: [32]

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



Fig. 14 Pasos de la Metodología Hefesto.

1.10. Herramientas propuestas para el desarrollo del Data Warehouse

1.10.1. La plataforma Pentaho Business Intelligence

La plataforma Pentaho Business Intelligence (BI) o Inteligencia de Negocio cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales, tanto las típicas como las sofisticadas y específicas al negocio. [33]

CAPÍTULO 1.FUNDAMENTACIÓN TEÓRICA

Incluye herramientas para realizar consultas, generación de informes y reportes, análisis interactivo, tableros de mando, ETL/integración de datos, data mining (minería de datos), y una servidor para la plataforma de BI que lo ha convertido en la suite BI de software libre más popular del mundo. Productos de Pentaho se utilizan en el Sistema de Mando Aéreo de US Army, Lifetime Networks, Terra Industries y Sun Microsystems. [34]

Los módulos que se utilizan de la plataforma Pentaho BI son: [35]

Análisis - Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas, generadas por Mondrian, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Además, con el Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP **server Mondrian**).

Mondrian es un motor de ROLAP desarrollado en Java que permite analizar grandes conjuntos de datos almacenados en un Data Warehouse. Posee un alto desempeño y análisis interactivo de volúmenes de información de cualquier tamaño, además explora dimensionalmente los datos y realiza cálculos avanzados utilizando el lenguaje MDX (multidimensional expressions o expresiones multidimensionales). [36]

Para acceder a las funcionalidades que brinda Mondrian se requiere de un cliente, es el caso de **JPivot** o **JRubik**.

JPivot: Es una librería de Java Server Pages (JSP) personalizados que presentan tablas OLAP y permite a los usuarios realizar operaciones tales como Drill Down y Drill Up además de consultas OLAP por medio del lenguaje MDX. Está diseñado para trabajar con motores OLAP como el de Mondrian por lo cual no utiliza las interfaces de programación de aplicaciones o (APIs) de Mondrian directamente sino que implementa su propio modelo OLAP. [37]

JRubik: Está basado en los componentes de JPivot y también se puede conectar a fuentes OLAP basadas en Mondrian. Las consultas son realizadas mediante el lenguaje MDX y cuenta con componentes similares a los de JPivot, solo que su interfaz no está basada en JSP sino en Swing, por lo que es una aplicación Desktop. [38]

Integración de Datos - se realiza con una herramienta Kettle ETL (Pentaho Data Integration) que permite implementar los procesos ETL.

1.10.2. Kettle 3.0.3

Kettle es una herramienta de código abierto compuesta por cuatro componentes fundamentales: SPOON para el diseño gráfico de las transformaciones, PAN para la ejecución de los trabajos y las transformaciones, CHEF para el diseño de la carga de datos y KITCHEN para la ejecución de los trabajos batch diseñados con CHEF. [39]

1.10.3. Schema Workbench 3.0.3

El Schema Workbench de Mondrian es un diseñador de interfaces que permite crear y configurar un cubo OLAP. El motor de Mondrian procesa las consultas MDX utilizando ROLAP. Los esquemas con los que trabaja son simplemente ficheros XML con toda la información acerca de los datos, los cuales serán utilizados por Mondrian para formar la estructura del cubo. Los ficheros XML pueden ser considerados como estructuras semejantes a las creadas anteriormente durante el diseño de la tabla de hechos y las dimensiones que conforman el almacén. Es importante conocer que no es necesario construir un cubo físico, solo basta con crear el modelo con los metadatos o sea con la información acerca de los datos. [40]

1.10.4. Enterprise Architect 7.0

Enterprise Architect es una herramienta de construcción y modelado de software de alto rendimiento con una trazabilidad completa desde los requisitos iniciales hasta las decisiones de diseño de software, Enterprise Architect 7.0 provee el tipo de visualización, colaboración eficiente y robusta requerida en los entornos de desarrollo de software que actualmente son altamente demandantes. Como una solución de modelado verdaderamente ágil, provee una sobrecarga de instalación baja, un rendimiento brillante y una interfaz intuitiva. Es una herramienta de análisis de negocio y UML orientada a objetos, provee el límite competitivo para el desarrollo de software, administración de proyecto, administración de requerimientos y análisis de negocio. Es una herramienta flexible para la plataforma de Windows y Linux, que será utilizada para realizar el diseño de los diferentes artefactos del almacén. [41]

1.10.5. PostgreSQL 8.3

Es el gestor de bases de datos de código abierto más avanzado hoy en día, ofreciendo control de concurrencia multi-versión, soportando casi toda la sintaxis SQL (incluyendo sub consultas, transacciones, tipos y funciones definidas por el usuario), contando también con un amplio conjunto de enlaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y muchos más. Es un sistema objeto-relacional, ya que incluye la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. [42]

El progreso continuo del gestor de datos de código abierto PostgreSQL brinda a los consumidores la opción de instalar una base de datos no privativa.

1.10.6. JBoss

JBoss Application Server es el servidor de aplicaciones de código abierto más ampliamente desarrollado del mercado. Combinando una arquitectura orientada a servicios, revolucionaria con una licencia de código abierto, JBoss AS puede ser descargado, utilizado, incrustado y distribuido sin restricciones por la licencia. [43]

En este capítulo se describen la metodología, los conceptos y las tecnologías que son utilizadas para el desarrollo del Data Warehouse. Al concluirlo se ha proporcionado una visión más clara de esta nueva manera de organización de los datos para la toma de decisiones, lo que resulta fundamental en el desarrollo del presente trabajo.

CAPÍTULO 2. ANÁLISIS Y DISEÑO

En este capítulo se muestran los resultados y pasos seguidos para el análisis según lo describe la metodología utilizada y se mostrará el diseño de la base de datos del DWH. El análisis de la metodología es un paso fundamental para el desarrollo del mismo y bajo esta guía se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidos del negocio. Para obtener una mayor visión de lo que se realiza, se llevó a cabo un modelo de procesos, el cual conforma toda una serie de pasos o iteraciones que se seguirán hasta concluir con el funcionamiento del Data Warehouse.

2.1. Análisis del proceso

Para una mejor comprensión de todo el proceso que se realiza, se lleva a cabo un diagrama de procesos, que no entra entre los pasos de la metodología propuesta (ver Fig. 15). Para ello se tuvo en cuenta el alcance del negocio y los objetivos que persigue esta investigación, además del análisis profundo de los sistemas operacionales involucrados en el proceso de implementación del DWH (ver Fig. 16,17, 18 y 19), objetivo imprescindible a la hora de determinar la relación entre los campos que forman estos sistemas, incluyendo tablas en su totalidad de ser necesario y los indicadores y perspectivas que definirán las dimensiones y las tabla de hecho(s) dentro del Data Warehouse (ver Fig. 20).

En la Fig.15 se muestra como fue analizado el conjunto de procesos que componen el flujo de trabajo a seguir para la construcción del Data Warehouse. El flujo se inicia cuando el usuario, el cual puede ser cualquiera de las organizaciones interesadas en el resultado del sistema, activa el evento Solicitud de información, éste no es más que la generalización de los objetivos que detonarán el inicio de una serie de procesos que llevarán al resultado final o meta.

El primero a analizar será el proceso de Extracción-Transformación-Carga, el cual tiene como entrada fundamental los recursos que se consumirán durante el mismo. Estos recursos son los sistemas operacionales involucrados en esta fase del flujo de trabajo, los cuales almacenan los datos que se extraerán para su transformación y posterior carga. El análisis en línea será el próximo proceso que se realizará, teniendo como entrada el propio almacén con las dimensiones requeridas y atributos

necesarios para la construcción del cubo de datos y las sentencias SQL, la salida es la información solicitada que será representada en forma gráfica, tabla, para su mejor comprensión y estudio a la hora de tomar decisiones.

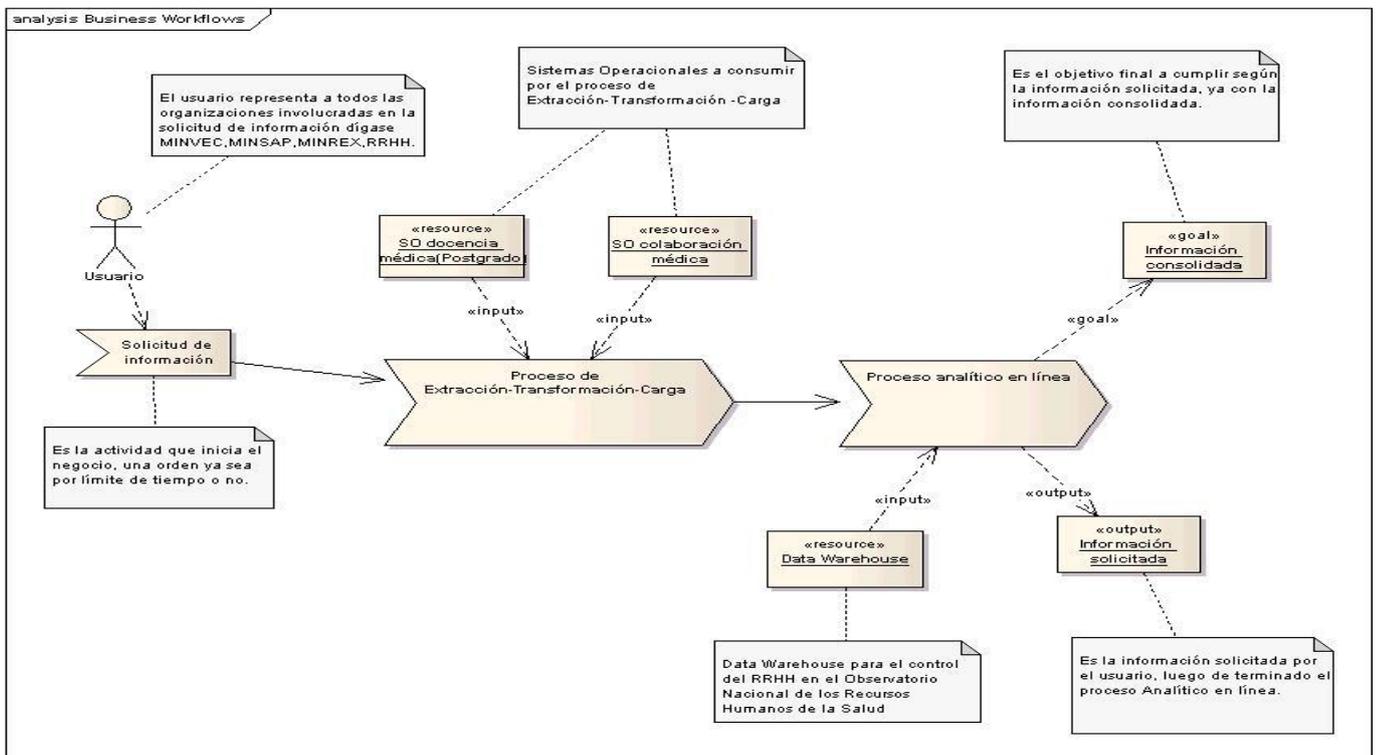


Fig. 15 Modelo de procesos.

Pasos y Aplicación de metodología

2.2. Análisis de Requerimientos

El primer paso comienza con obtener realmente las necesidades de información, lo cual puede llevarse a cabo de disímiles maneras ejemplo: entrevistas, cuestionarios, observaciones, etc. El objetivo fundamental de la misma es obtener e identificar las necesidades de información clave, que serán esencial para realizar las metas y estrategias trazadas para facilitar una eficaz y eficiente toma de decisiones.

Se realizó una entrevista con los usuarios en busca de sus necesidades de información, la actividad elegida fue la interrelación entre los temas de colaboración y docencia postgrado para obtener información realmente útil en relación con el proceso del tratado del colaborador como recurso humano

único en ambos temas y así posibilitar el estudio de su trayectoria para la toma de decisión. Luego se procedió a identificar que era lo que realmente interesaba de esta interrelación y cuáles eran los indicadores los que constituyen valores numéricos y representan lo que se desea analizar concretamente (cantidad, porcentaje, total), y las variables o perspectivas que se refieren a los temas mediante los cuales se quiere examinar los indicadores a tenerse en cuenta.

Algunos de los resultados obtenidos:

Se desean conocer cuántos RRHH se encuentra colaborando en país (X) y cuenta con la categoría de máster o doctor.

Se desea conocer el total de RRHH que se encuentra en espera de misión y es especialista en (Y).

Porcentaje que representa la cantidad de RRHH en misión con respecto al total que se encuentra en espera de misión y cuenta con alguna categoría científica.

2.2.1. Identificadores y perspectiva

Luego de haber establecido las preguntas se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Para cada uno se puede utilizar como indicador el representativo numérico cantidad, total o porcentaje.

Cantidad de Recursos Humanos (RRHH) que se encuentra colaborando en (país) y cuenta con la categoría de máster o doctor.

Total de RRHH que se encuentra en espera de misión y es especialista en (Y).

Porcentaje de de los RRHH en total con respecto a los RRHH en misión que se encuentra en espera de misión y cuenta con alguna categoría científica.

 Perspectivas

 Indicadores

2.2.1.1. Modelo Conceptual

En esta etapa se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas, lo que permitirá observar con claridad el alcance del DWH. En la Fig. 16 se muestra dicho modelo, a la izquierda se colocan las perspectivas seleccionadas, que serán unidas por un óvalo que representa y lleva el nombre de la relación que existe entre ellas. La relación constituye el proceso o área de estudio elegida.

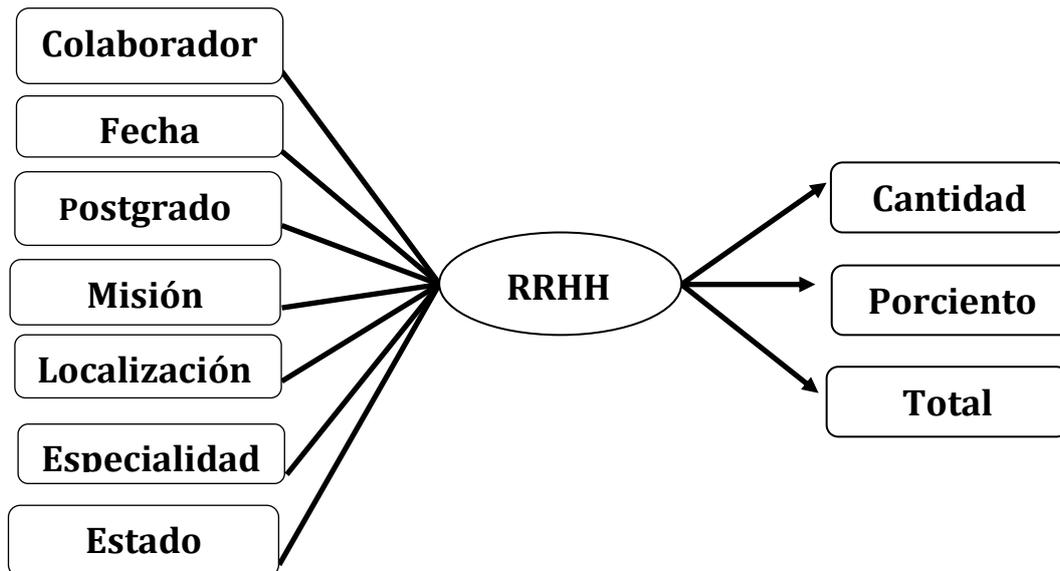


Fig. 16 Modelo Conceptual.

2.2.2. Análisis de los OLTP

Es el segundo paso a tener en cuenta, el objetivo de éste análisis es examinar los OLTP (Docencia Médica (esquema de postgrado, esquema especialidades) y Colaboración disponibles que contengan la información requerida. Las Fig. 17, 18 y 19 muestran los atributos pertenecientes a las tablas de las bases de datos de Docencia Médica y Colaboración respectivamente, en el caso de la última específicamente del esquema de Postgrado y Especialidades, lo cual facilita la comprensión del diseño de la nueva base de datos con la que se trabajará durante todo el proceso de confección del Data Warehouse. Los atributos resaltados serán incluidos como nuevos componentes de las dimensiones que forman el esquema en estrella (ver Fig.7) y que están reflejados en la Fig. 20 según el color de la dimensión con que se relacionan.

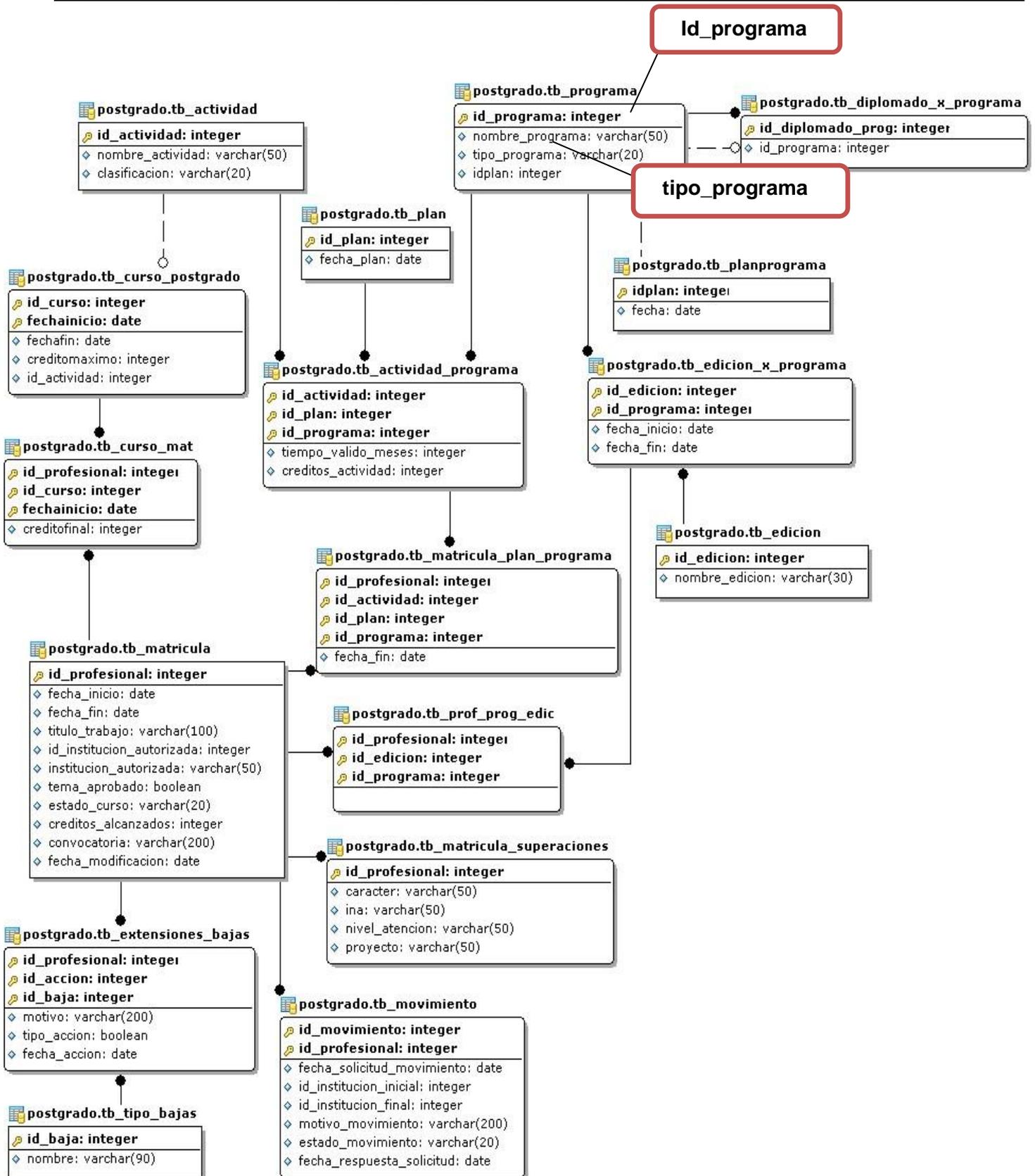


Fig.17 Diagrama de la base de datos Docencia Médica (Esquema de Postgrado).

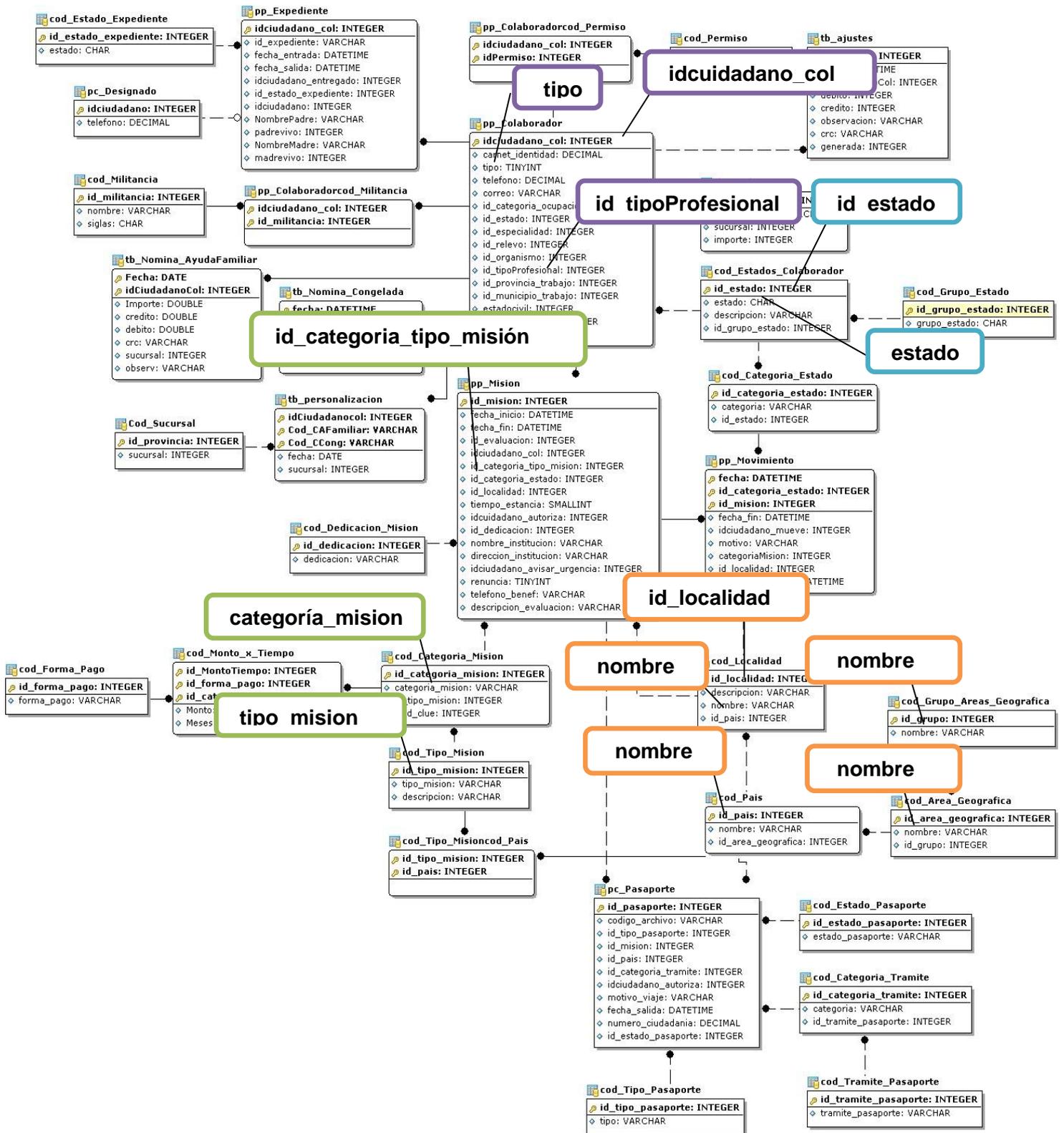


Fig.19 Diagrama de la Base de Datos Colaboración.

El modelo conceptual que se muestra en la Fig. 20 detalla según una escala de colores los atributos que guardan relación con su respectiva dimensión junto a la relación que se establece entre estos y los indicadores que conforman el negocio, los cuales serán el resultado final de realizar OLAP. Los indicadores reflejados en este modelo no guardan relación con ningún campo de los sistemas operacionales mostrados en las figuras 17, 18 y 19 de ahí que no aparezcan como atributos de la tabla de hechos RRHH (ver Fig.21) pues la información que se obtendrá como resultado del proceso ETL estará consolidada y viable para la ser analizada y explotada en el OLAP. Vale destacar que la perspectiva fecha fundamental en el DWH, será la misma que cuando se cargue la información.

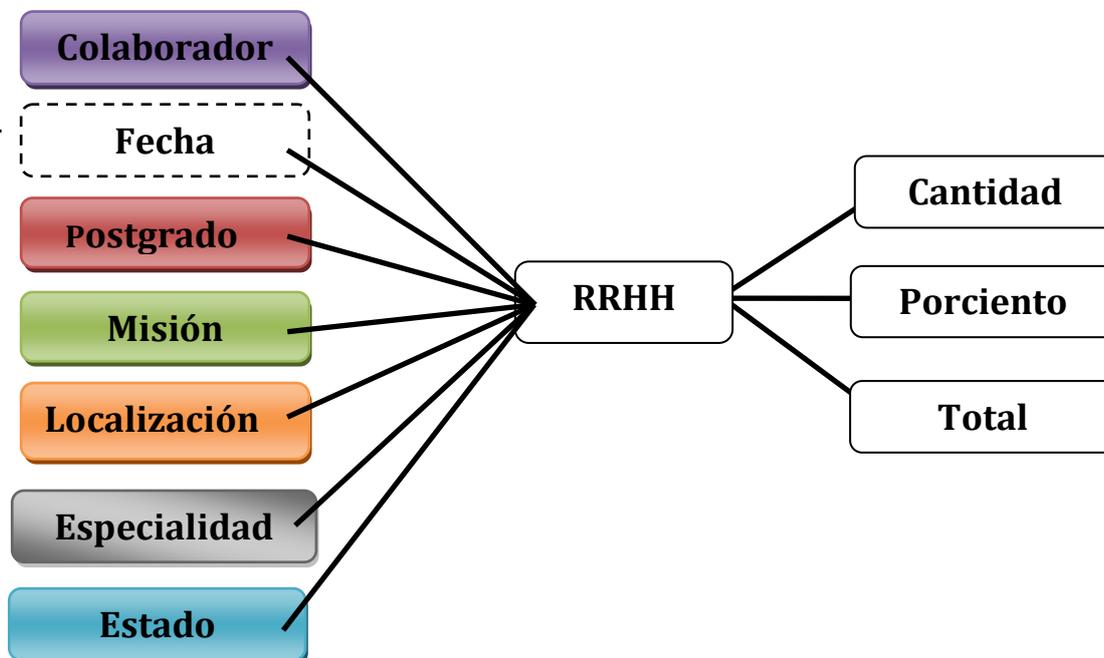


Fig. 20 Modelo Conceptual del DWH.

2.2.3. Elaboración del Modelo Lógico de la estructura del DWH.

En este paso se confecciona el modelo lógico que no es más que la representación de una estructura de datos, que puede procesarse y almacenarse en algún sistema gestor de base de datos en este caso PostgreSQLv8.3. Es muy importante definir en este punto el esquema que se empleará, el definido fue el esquema en estrella.

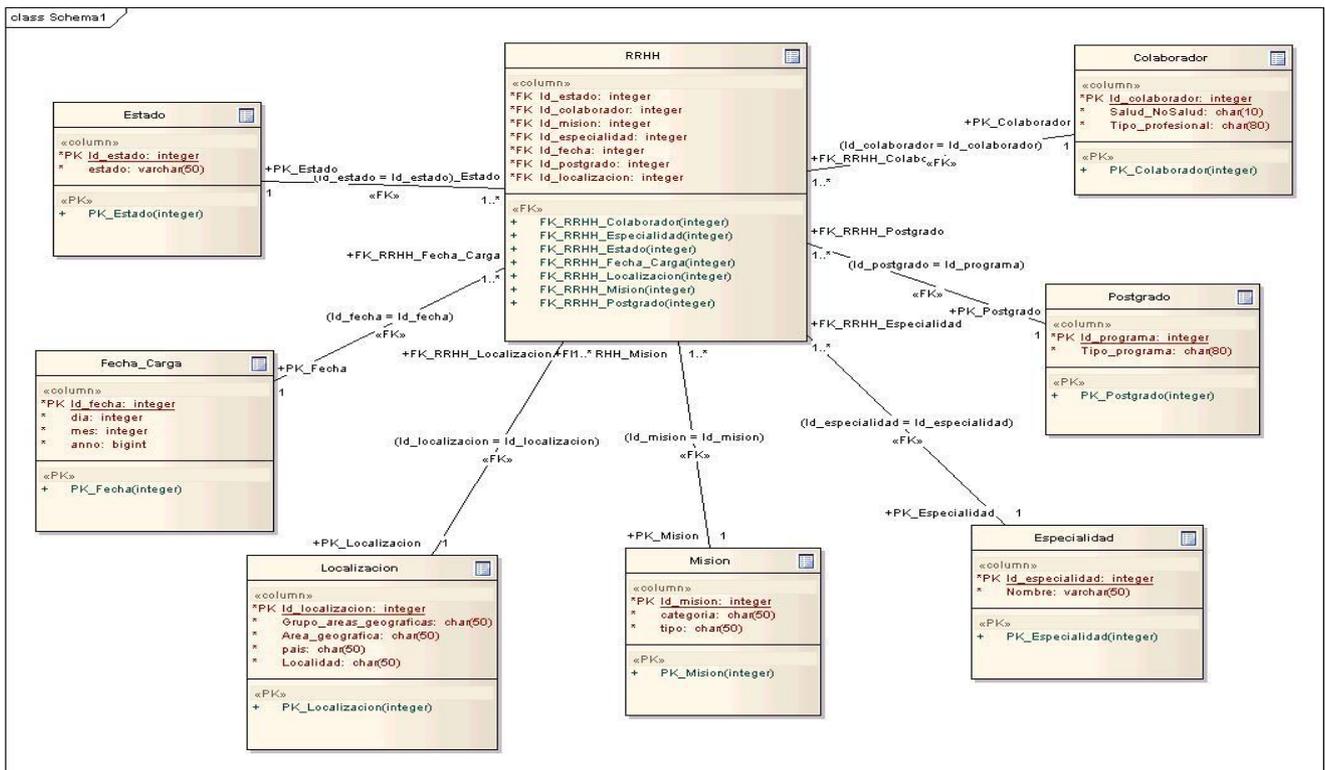


Fig. 21 Diseño de las dimensiones y tabla de hechos con sus atributos asociados, según esquema seleccionado.

2.2.4. Proceso ETL

Se construirá la base de datos que soportará las dimensiones y contendrá los datos después de su extracción y posterior transformación desde los sistemas operacionales que se desean integrar. Para llevar a cabo este proceso se trabajó con la herramienta descrita anteriormente de la suite Pentaho, Kettle.

Lo primero que se realiza es la extracción de los atributos necesarios de los sistemas operacionales (ver Fig. 16, 17 y 18), para llenar las dimensiones (ver Fig. 20). Como se puede ver en el caso específico de la dimensión Fecha_Carga, se necesita el día, mes, año en el que se cargarán los datos en el DWH. Para obtener dichos atributos (ver Fig. 21) se procedió de la siguiente manera:

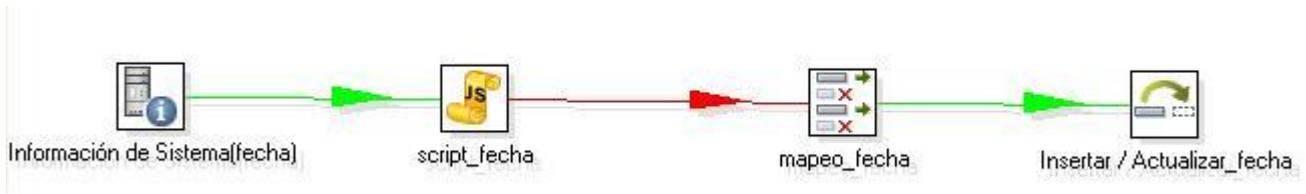


Fig. 21 Dimensión Fecha_Carga.

Se utilizó el componente (Información de Sistema) que permite obtener la fecha en la que se están cargando los datos. A través de un componente Script (ver Fig. 22) se especifica que valores tomarán las variables de salida de la función definida por el componente (Información de Sistema).

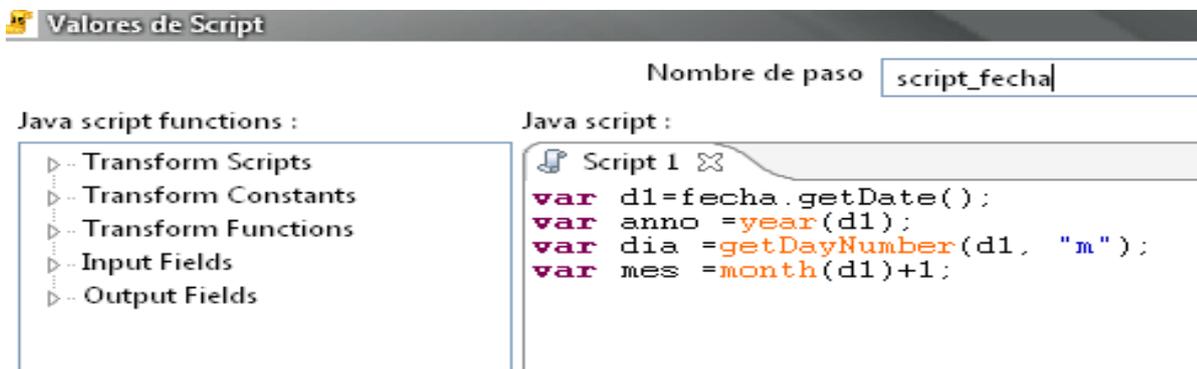


Fig. 22 Valores de Script.

Luego se realiza la carga a la dimensión requerida, para esto se utiliza el componente insertar/actualizar el cual tiene como objetivo insertar o actualizar en dependencia de lo especificado, en el caso tratado no es necesario actualizar las fechas, solo insertar para poder obtener así un historial de las mismas. Posterior se crea el mapeo (mapping) el cual desglosa la fecha obtenida en los atributos de la dimensión a llenar (ver Fig. 23).

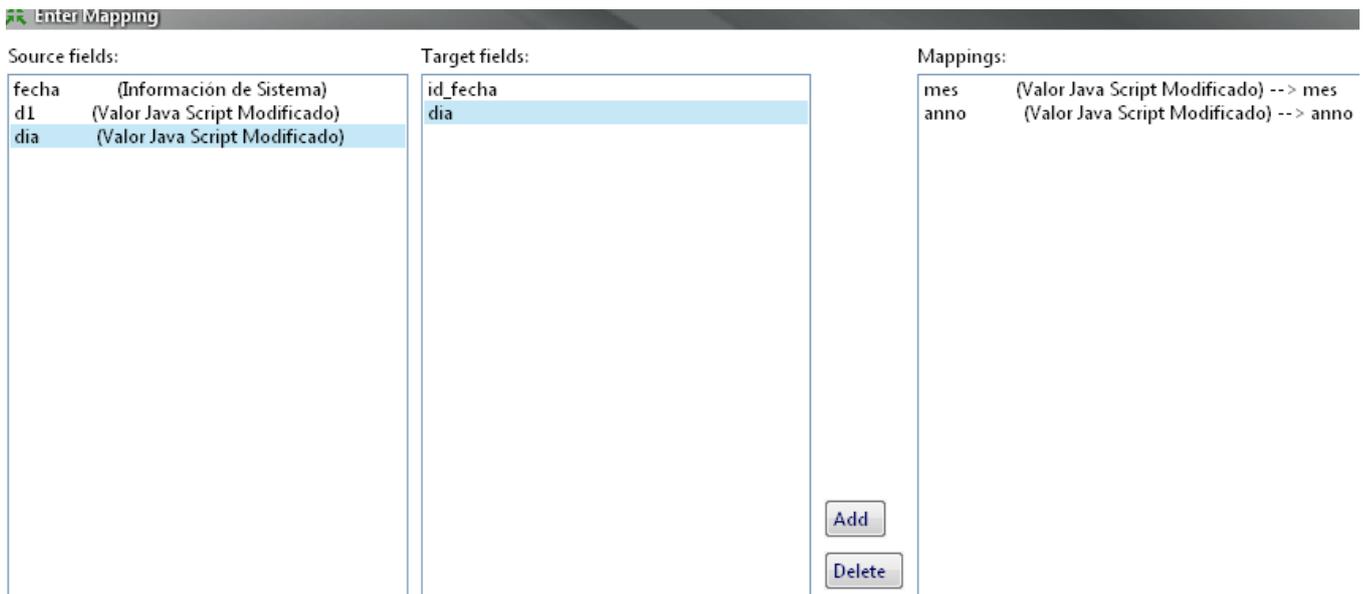


Fig. 23 Valores de Script.

En el caso de la dimensión colaborador, se extraerán los atributos del sistema operacional a través del primer componente (ver Fig. 24) utilizando sentencias SQL, cuyos resultados serán transformados con el componente Script para que lleguen al almacén en el formato deseado y se realizará la carga a través de los componentes mapeo_colaborador e insertar/actualizar_colaborador.

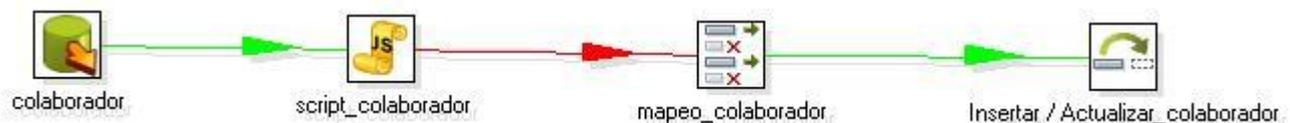


Fig. 24 Dimensión Colaborador.

El proceso realizado es similar para la carga de las 5 dimensiones restantes ([Ver anexos 2](#)).

Se procede a la carga de la tabla de hechos, la cual contiene todos los identificadores de las dimensiones que componen el almacén de datos. Los datos a extraer son cargados utilizando las transformaciones rrhh_mision y rrhh_nomision, las cuales cuentan con los componentes de entrada misiones, colaboración y fecha, realizando la integración entre las filas cargadas mediante el componente Búsqueda de Valor en Base de Datos (ver Fig. 26), el cual se encargará de que los datos a integrar sean los correctos, además se utilizará el componente Juntar Filas, que es el responsable de

realizar un producto cartesiano de los elementos que tiene como entrada y los filtra según una condición declarada (ver Fig. 27).

Los datos ya seleccionados serán filtrados para evitar que se repitan todos los identificadores con la fecha diferente, los cuales serán borrados en caso positivo, de lo contrario serán mapeados con los atributos de la tabla (RRHH) para su posterior almacenamiento en la misma. Primeramente se ingresarán en la tabla de hechos los colaboradores con una misión y luego los que no cuentan con misión, evitando que queden fuera en caso de que no se cumpla alguna de estas características (ver Fig. 28).

Nombre paso: Búsqueda especialidad

Conexión: dw_DocenciaMedica

Esquema de búsqueda: especialidades

Tabla de búsqueda: tb_especialidad_carrera

¿Habilitar cache?:

Tamaño de cache en filas (0=todas): 0

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

#	Campo de tabla	Comparador	Campo1	Campo2
1	id_profesional	=	id_col	

Valores a devolver de la tabla de búsqueda:

No procesar la fila si la búsqueda falla:

¿Producir error si se obtienen múltiples:

Ordenar por: _____

Vale Cancelar Obtener Campos Obtener Campos Búsqueda

Fig. 26 Componente Búsqueda de Valor en Base de Datos.

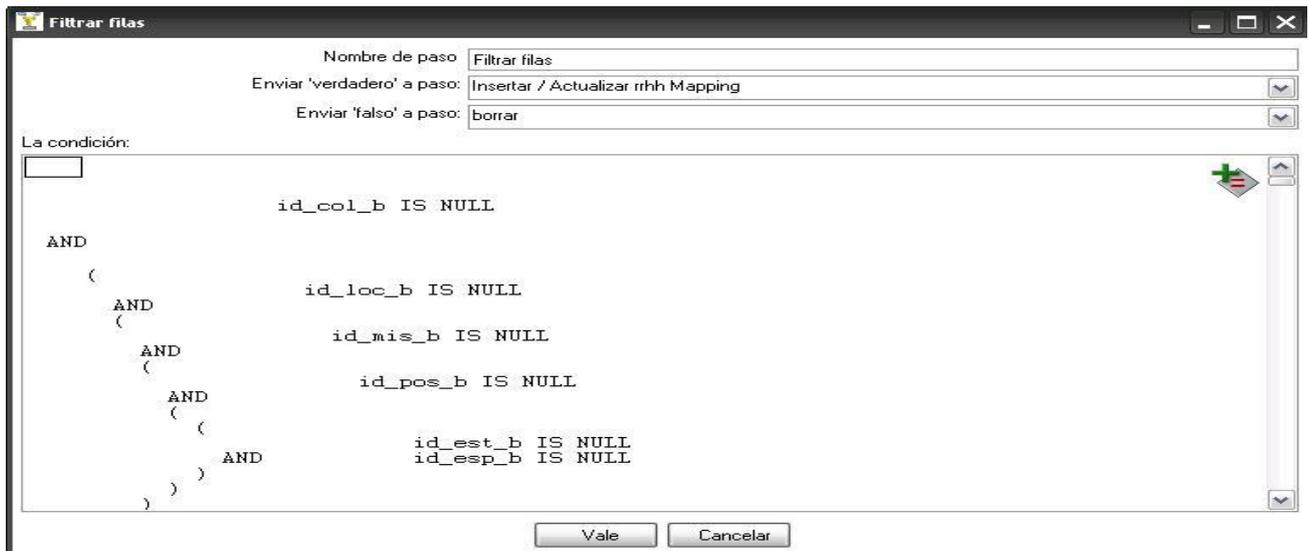


Fig. 27 Componente Filtrar Filas.

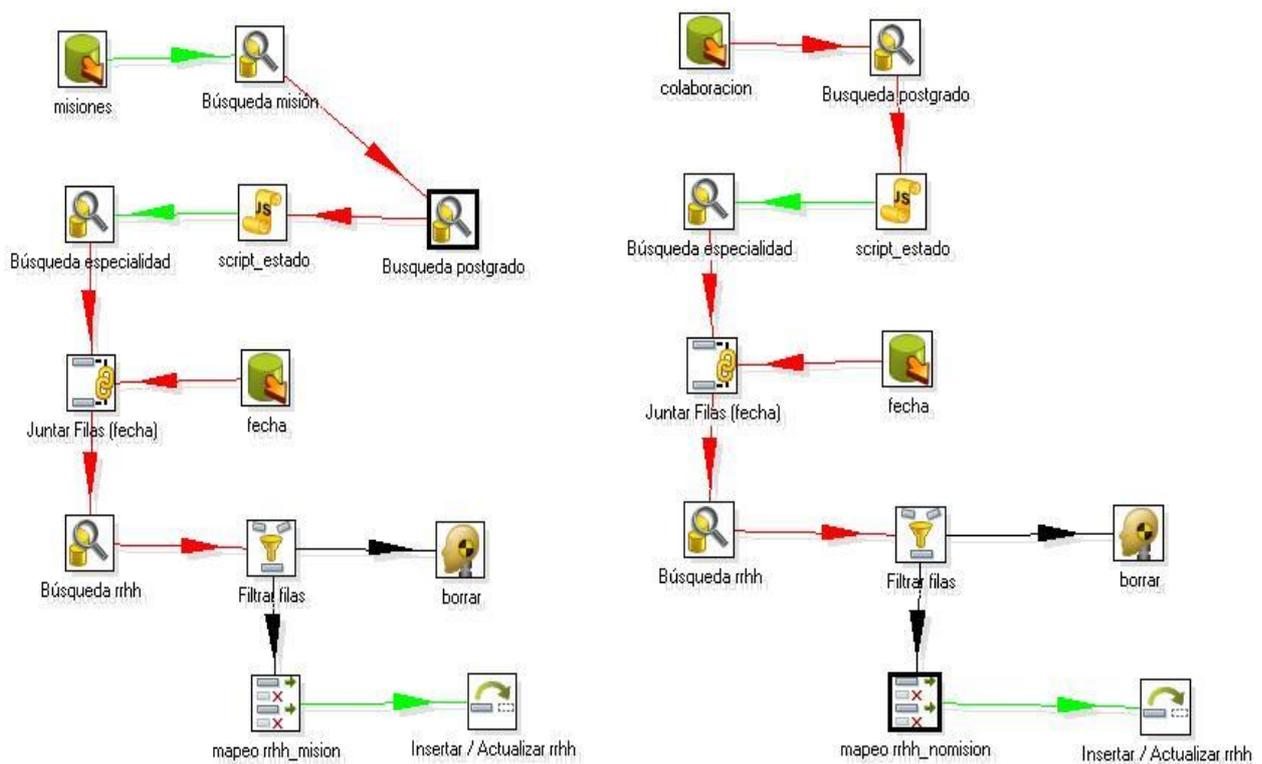


Fig. 28 Tabla de hechos.

Finalizada la carga de las dimensiones con la tabla de hecho(s), se pondrá en ejecución cada una de las transformaciones del proceso ETL, las que están dirigidas por el trabajo o job, que se ejecuta en función de en un tiempo seleccionado. Para ello se utiliza el componente Start (ver Fig. 29)



Fig. 29 Inicializador del trabajo.

Este componente iniciará las transformaciones una vez a la semana en el orden seleccionado, teniendo como transformación final el llenado de la tabla de hechos. En el trabajo se incluye otros componentes necesarios para casos de errores, estos son el DUMMY y el MsgBox Info (mensaje de error), el primero encargado de detener la transformación y el MsgBox Info de mostrar una advertencia al administrador del servidor en caso de que ocurra algún problema en la ejecución (ver Fig. 30).

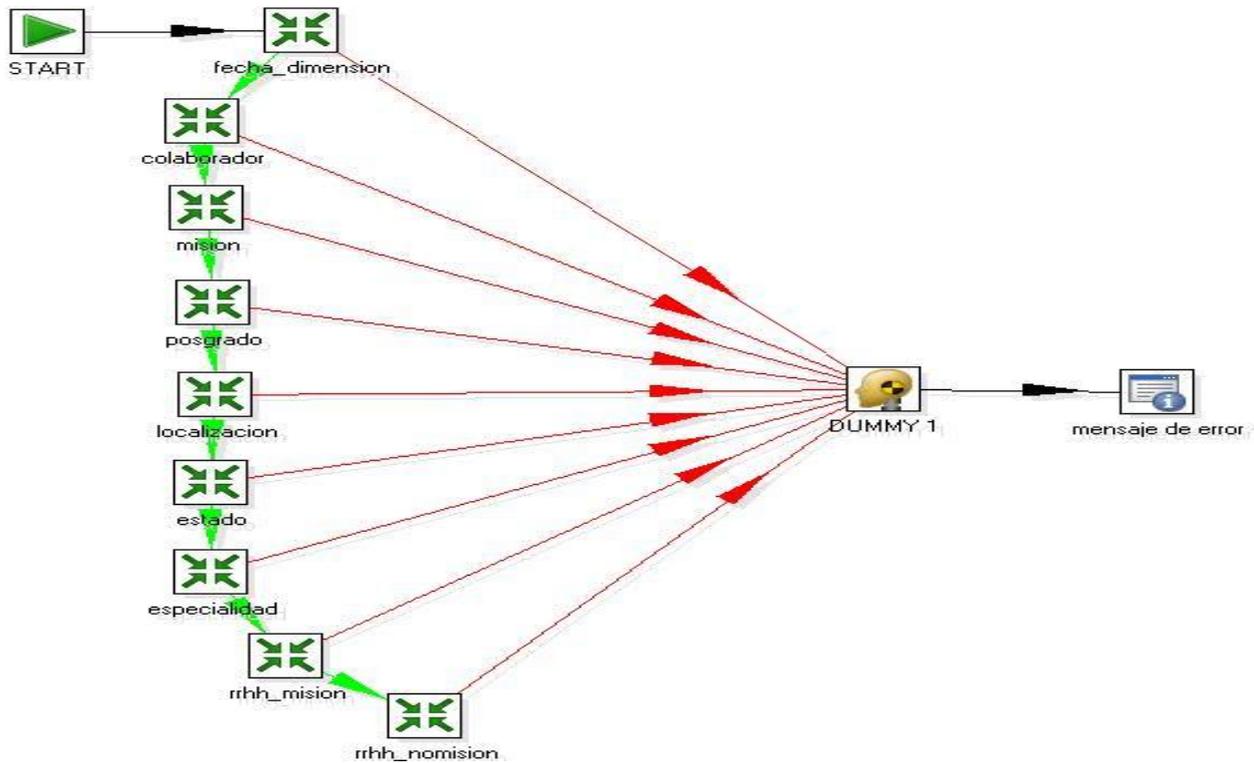


Fig. 30 Ejecución del trabajo.

En este capítulo se expusieron los resultados obtenidos para la construcción del DWH, siguiendo los pasos basados en la metodología definida. Se obtuvo la información necesaria para construir el almacén de datos, para así proceder a la realización del proceso ETL, proceso de vital importancia para el éxito del almacén. El cual se transformó y cargó hacia el Data Warehouse todos los datos seleccionados de los sistemas operacionales de Colaboración y Docencia Médica. El proceso ETL concluyó satisfactoriamente con la exitosa ejecución del trabajo como tarea final, de este parte del Data Warehouse para el control del RRHH. Este segundo capítulo, constituye la base para la construcción triunfante del procesamiento analítico en línea.

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

Luego de haber concluido el proceso de ETL se procede a conformar el cubo de datos, fundamental para realizar los cruces que permiten extraer, en forma rápida y eficiente, la información que se requiere.

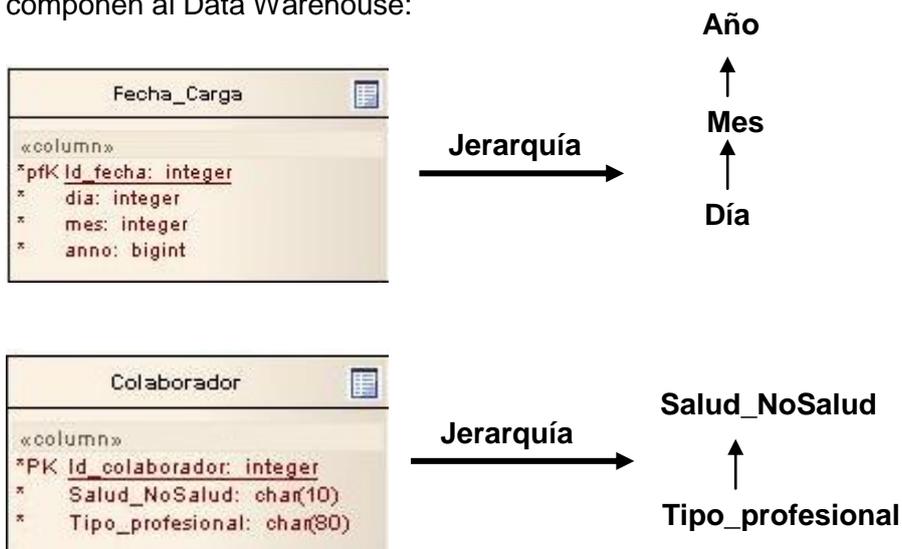
3.1. Determinar jerarquías

La construcción del cubo multidimensional, consiste en incorporar las dimensiones y establecer las jerarquías de sus atributos que siguen un orden preestablecido, de vital importancia para organizar los niveles dentro de una dimensión. La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado o del menor nivel de detalle hacia el más general.

Las jerarquías poseen las siguientes características:

- Pueden existir varias en una misma dimensión.
- Están compuestas por dos o más niveles.
- Se tiene una relación de uno a muchos entre atributos consecutivos de un nivel superior y uno inferior.

A continuación se muestran las jerarquías de los atributos de algunas de las dimensiones que componen al Data Warehouse:



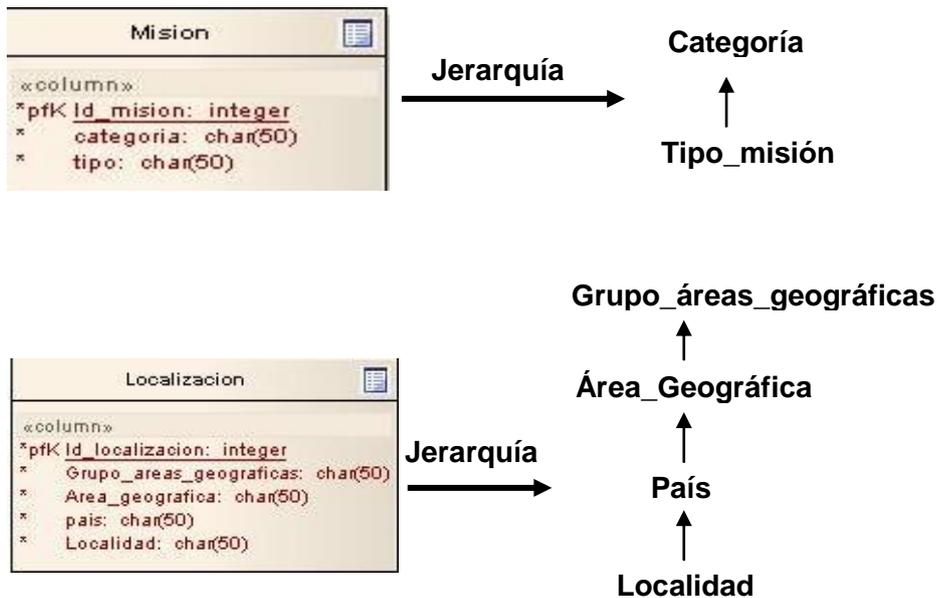


Fig. 31 Jerarquías de los atributos.

3.2. Medidas

Al establecer la jerarquía de las dimensiones se procede a crear las medidas que satisfacen las necesidades de información definidas en el epígrafe 2.2.1. Estas no se extraen de los sistemas operacionales, para obtenerlas se utilizan las consultas MDX (multidimensional expressions o expresiones multidimensionales) ejecutadas sobre el Data Warehouse que se encargan de brindar los indicadores que contribuirán a la toma de decisiones.

Las dimensiones y la tabla de hechos a las que se le calcularán las medidas se cargarán desde db_rrhh y se diseñará el cubo utilizando la herramienta Workbench (ver Fig., 32) añadiéndole las medidas y creando las consultas para obtener la cantidad, porciento y total de los Recursos Humanos según las necesidades de información del Observatorio Nacional.

A continuación se muestra la estructura del cubo multidimensional (cubo_datos) que contiene las dimensiones, jerarquías de los atributos y medidas.

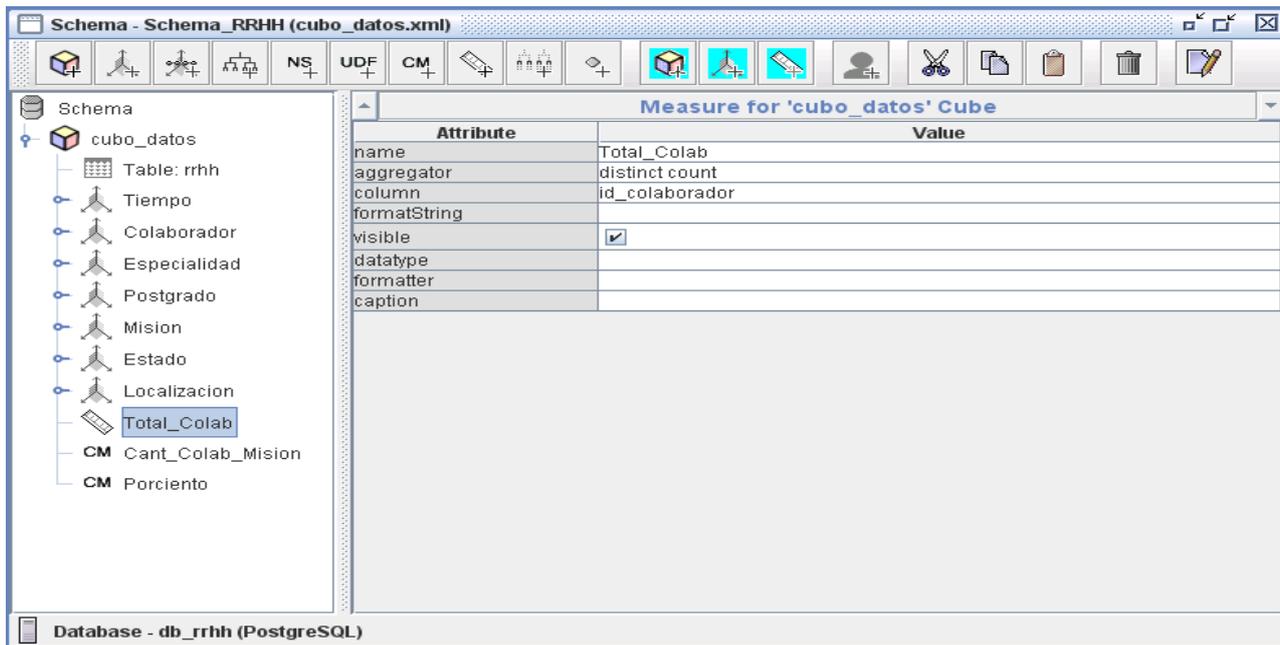


Fig. 32 Cubo OLAP

3.3. Consultas

Las expresiones multidimensionales permiten consultar objetos multidimensionales, como el cubo de RRHH (ver Fig. 33), y devolver conjuntos de celdas multidimensionales que contengan los datos del cubo. Las consultas calcularán los datos numéricos y estos serán añadidos en los resultados a visualizar utilizando las herramientas Mondrian por el Observatorio Nacional de los Recursos Humanos.

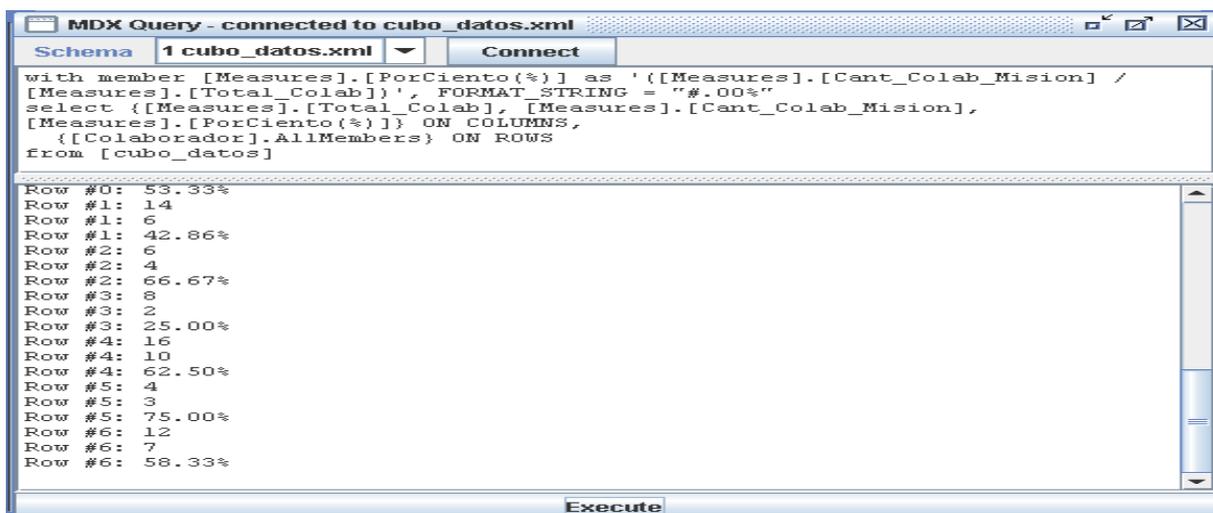


Fig. 33 Consultas MDX

3.4. Mondrian

3.4.1. JPivot

Al tener diseñado el cubo, es almacenado en un archivo XML que se cargará en el Mondrian, herramienta utilizada por el usuario para realizar el Proceso Analítico en Línea (OLAP). La cual será publicada en el servidor web JBoss, que está ejecutándose sobre el Sistema Operativo Linux distribución Debian etch 4.0.

En la siguiente figura se muestra el cliente JPivot, disponible a través de la web, cuya página principal muestra un menú con todas las opciones que permite el filtrando por los campos que se desee (Ver Fig. 34), ajustar la visión de los datos, el empleo de filtros de visualización, añadir o quitar campos, observar la información desde diferentes perspectivas, cuyos resultados serán mostrados mediante tablas o gráficos (Ver Fig. 35 y 36), y exportarlos a formato PDF o EXCEL.

The screenshot shows the JPivot web interface. At the top is a toolbar with icons for MDX, filters, and visualization. Below it is a configuration panel with three sections: 'Columnas' (Measures), 'Filas' (Rows), and 'Filtro' (Filters). The 'Filtro' section is expanded, showing a list of dimensions: Especialidad, Estado, Localizacion, Mision, Postgrado, and Tiempo. Below the configuration panel are 'Aplicar' and 'Cancelar' buttons. Below the configuration panel is a data table showing the results for the 'Colaborador' dimension.

	Medidas		
Colaborador	Total_Colab	Cant_Colab_Mision	Porcentaje
Colaboradores	30	29	96,67%
No Salud	14	14	100,00%
Contador	6	6	100,00%
Informático	8	8	100,00%
Salud	16	15	93,75%
Estomatólogo	4	4	100,00%
Médico	12	11	91,67%

Fig. 34 Dimensión colaborador con las medidas o indicadores.

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

La siguiente imagen muestra la interrelación entre los colaboradores de acuerdo a su tipo, ya sea salud o no salud y de ellos solamente se escogió los que ostenten alguna maestría o doctorado y los que no hayan pasado ningún postgrado, igualmente son mostrados los indicadores (Total_Colab) que representa el total de colaboradores que estén o no cursando algún postgrado, (Cant_misión) es la cantidad de colaboradores que presentan al menos una misión y el (Porciento) es el porciento de colaboradores en misión del total que existen.

		Medidas		
Colaborador	Postgrado	•Total_Colab	•Cant_Colab_Mision	•Porciento
Colaboradores	Sin_Postgrado	22	22	100,00%
	Tipo_Doctorado	3	3	100,00%
	Tipo_Maestria	2	1	50,00%
No Salud	Sin_Postgrado	9	9	100,00%
	Tipo_Doctorado	1	1	100,00%
	Tipo_Maestria	1	1	100,00%
Contador	Sin_Postgrado	4	4	100,00%
	Tipo_Doctorado	1	1	100,00%
	Tipo_Maestria	1	1	100,00%
Informático	Sin_Postgrado	5	5	100,00%
Salud	Sin_Postgrado	13	13	100,00%
	Tipo_Doctorado	2	2	100,00%
	Tipo_Maestria	1		
Estomatólogo	Sin_Postgrado	3	3	100,00%
	Tipo_Doctorado	1	1	100,00%
Médico	Sin_Postgrado	10	10	100,00%
	Tipo_Doctorado	1	1	100,00%
	Tipo_Maestria	1		

Fig. 35 Información tabulada

Se muestra también esta información en forma de gráfica para que el usuario tenga la posibilidad de observarla desde otra perspectiva para un mejor entendimiento y comprensión para el apoyo a la toma de decisiones (Ver Fig.36).

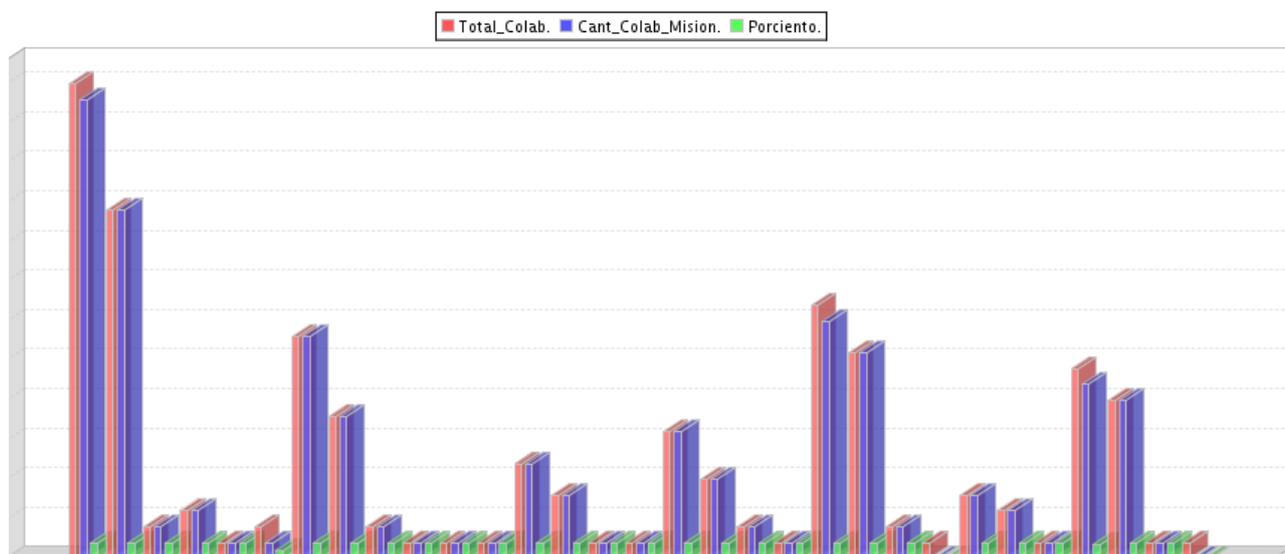


Fig.36 Información gráfica.

Otras de las operaciones realizadas para la captura de la información es el drill through, que permiten manejar más fácilmente la información seleccionada y la muestra de forma más explícita, exponiendo todos los campos de las dimensiones filtradas con sus datos, según la medida a la que se pide mostrar los datos (Ver Fig. 37).

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

	Medidas		
Estado	Cant_Colab	Cant_Colab_Mision	Por ciento
-Estados	30	29	96,67%
activo	29	29	100,00%
pasivo	30	29	96,67%

Drill Through Table for Cant_Colab								
▲ Anno	● Mes	● Dia	● Salud_NoSalud	● Tipo_Profesional	● Nombre_Especialidad	● Tipo_Programa	● Categoria	● Tipo_Mision
2.009,00	5	20	No Salud	Contador	Sin especialidad	Sin_Postgrado	CHR Brigadas Especiales	CHR
2.009,00	5	20	No Salud	Contador	Sin especialidad	Sin_Postgrado	CHR Brigadas Especiales	CHR
2.009,00	5	20	No Salud	Contador	Sin especialidad	Sin_Postgrado	Especialista Medicina Física y Rehabilitación	ME
2.009,00	5	20	No Salud	Contador	Sin especialidad	Sin_Postgrado	Especialista Medicina Física y Rehabilitación	ME
2.009,00	5	20	No Salud	Contador	Sin especialidad	Sin_Postgrado	Estomatólogo Contingente	ME
2.009,00	5	20	No Salud	Contador	Sin especialidad	Sin_Postgrado	Estomatólogo Contingente	ME

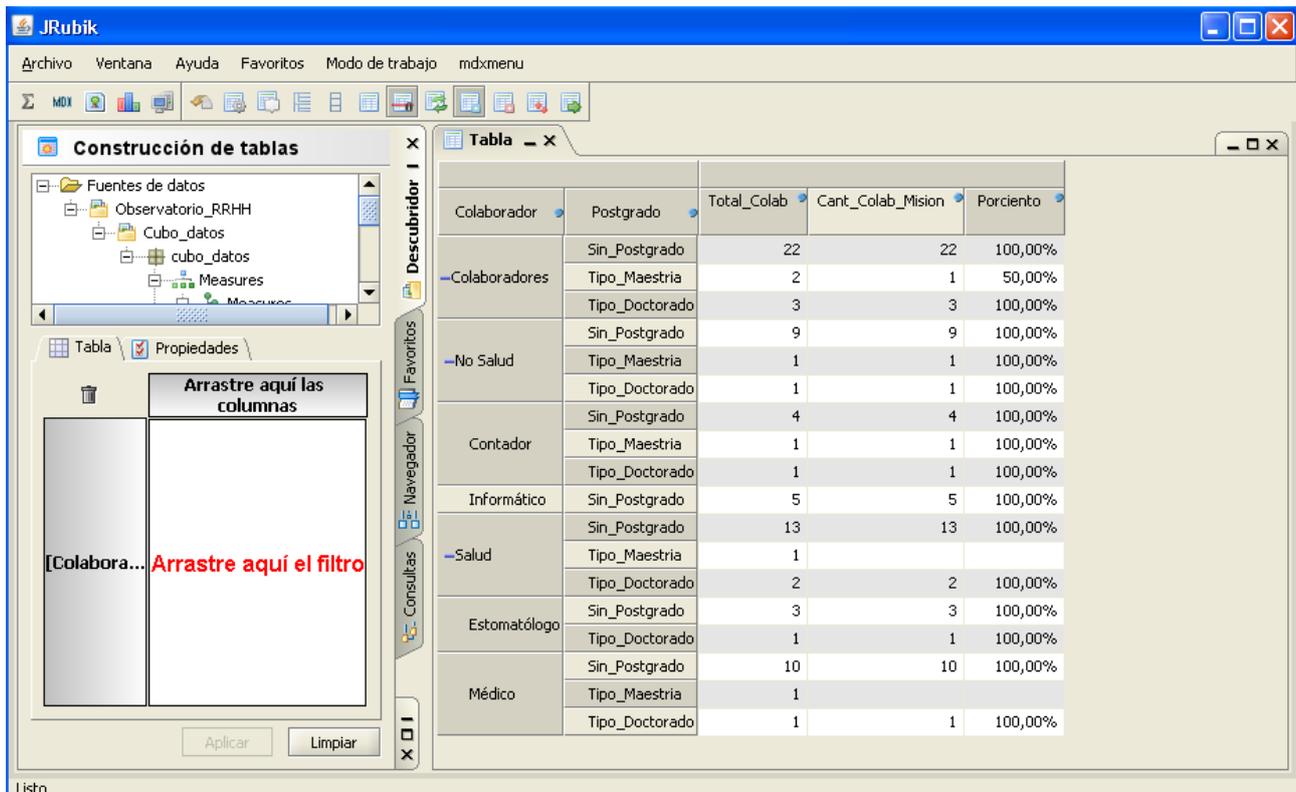
Fig. 37 Drill throughg

3.4.2. JRubik

Además de utilizar JPivot como cliente para Mondrian se utilizó el JRubik, una aplicación de escritorio que brinda también la posibilidad de combinar y filtrar las dimensiones y medidas en busca del resultado deseado (Ver Fig. 38), los mismos podrán ser exportados a formato PDF o EXCEL.

Las operaciones realizadas con el cliente JPivot son igualmente ejecutadas por el cliente JRubik. A continuación se muestran los resultados obtenidos con este último.

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA



The screenshot shows the JRubik software interface. On the left, the 'Construcción de tablas' (Table Construction) sidebar is visible, showing a tree view of data sources including 'Observatorio_RRHH', 'Cubo_datos', and 'Measures'. Below this, there are buttons for 'Aplicar' and 'Limpiar'. The main window displays a pivot table with the following data:

Colaborador	Postgrado	Total_Colab	Cant_Colab_Mision	Porcentaje
Colaboradores	Sin_Postgrado	22	22	100,00%
	Tipo_Maestria	2	1	50,00%
	Tipo_Doctorado	3	3	100,00%
No Salud	Sin_Postgrado	9	9	100,00%
	Tipo_Maestria	1	1	100,00%
	Tipo_Doctorado	1	1	100,00%
Contador	Sin_Postgrado	4	4	100,00%
	Tipo_Maestria	1	1	100,00%
	Tipo_Doctorado	1	1	100,00%
Informático	Sin_Postgrado	5	5	100,00%
	Tipo_Doctorado	1	1	100,00%
Salud	Sin_Postgrado	13	13	100,00%
	Tipo_Maestria	1		
	Tipo_Doctorado	2	2	100,00%
Estomatólogo	Sin_Postgrado	3	3	100,00%
	Tipo_Doctorado	1	1	100,00%
Médico	Sin_Postgrado	10	10	100,00%
	Tipo_Maestria	1		
	Tipo_Doctorado	1	1	100,00%

Fig. 38 Combinación de las dimensiones Misión y Colaborador.

Se brinda la posibilidad de realizar el drill through permitiendo apreciar los datos en un mayor nivel de detalle, con una interfaz más amigable (Ver Fig. 39).

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

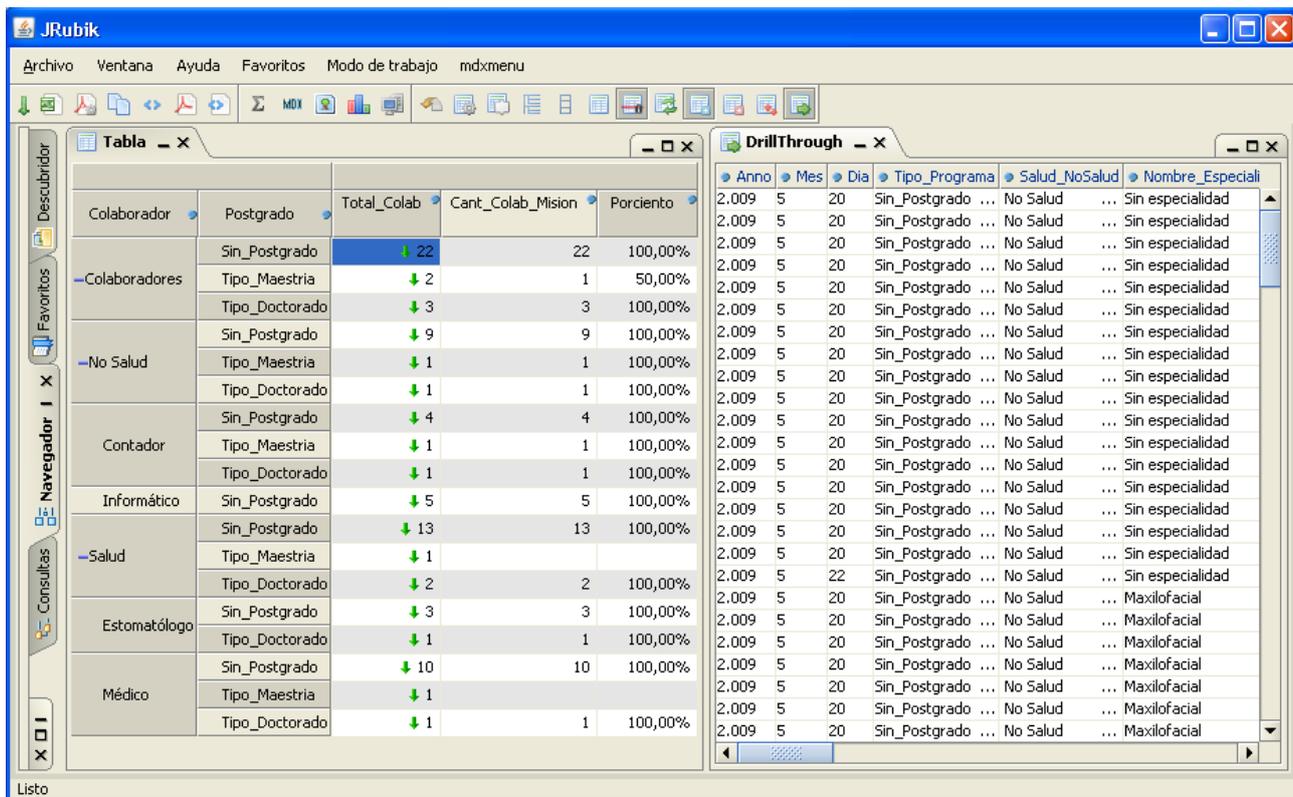


Fig. 39 Drill Through.

Otras de las opciones que tiene JRubik es que permite la presentación de los resultados, a través de diferentes gráficos, ya sea en forma tabular, anillo, pastel, (Ver Fig. 40, 41) etc. Ofreciendo una mayor posibilidad de comprensión e interpretación de la información para el apoyo a la toma de decisiones en el Observatorio Nacional de Recursos Humanos de la Salud.

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

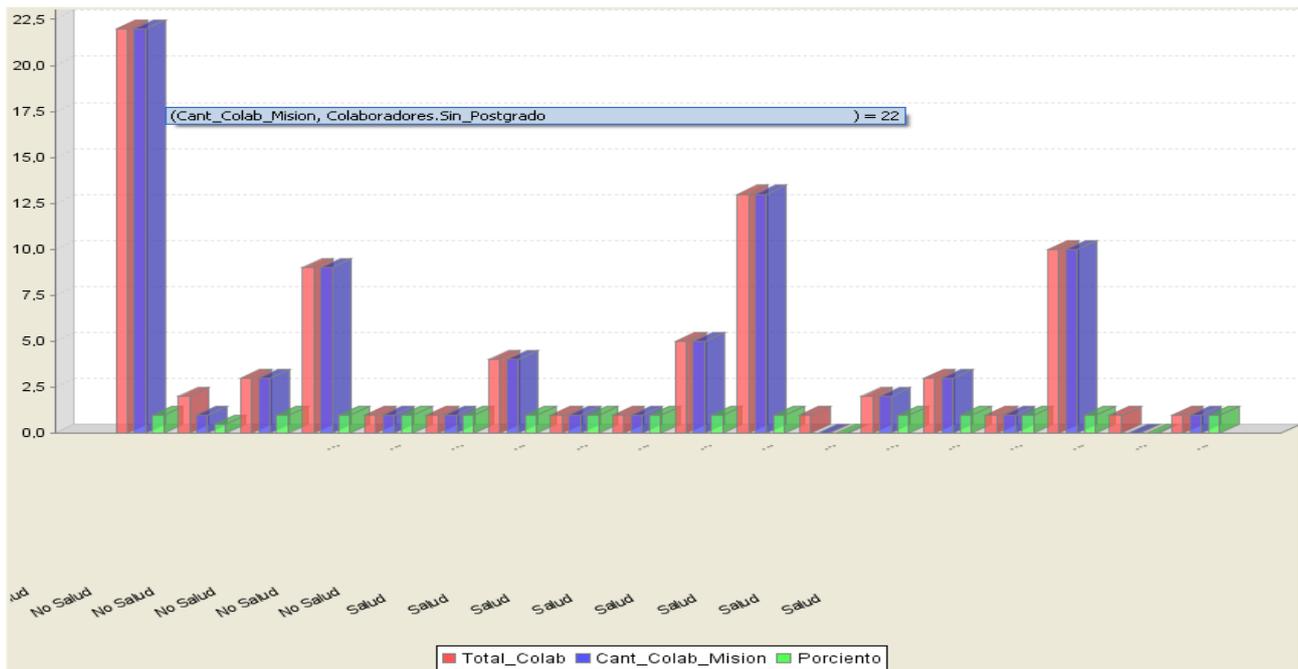


Fig. 40 Gráfica de barras.

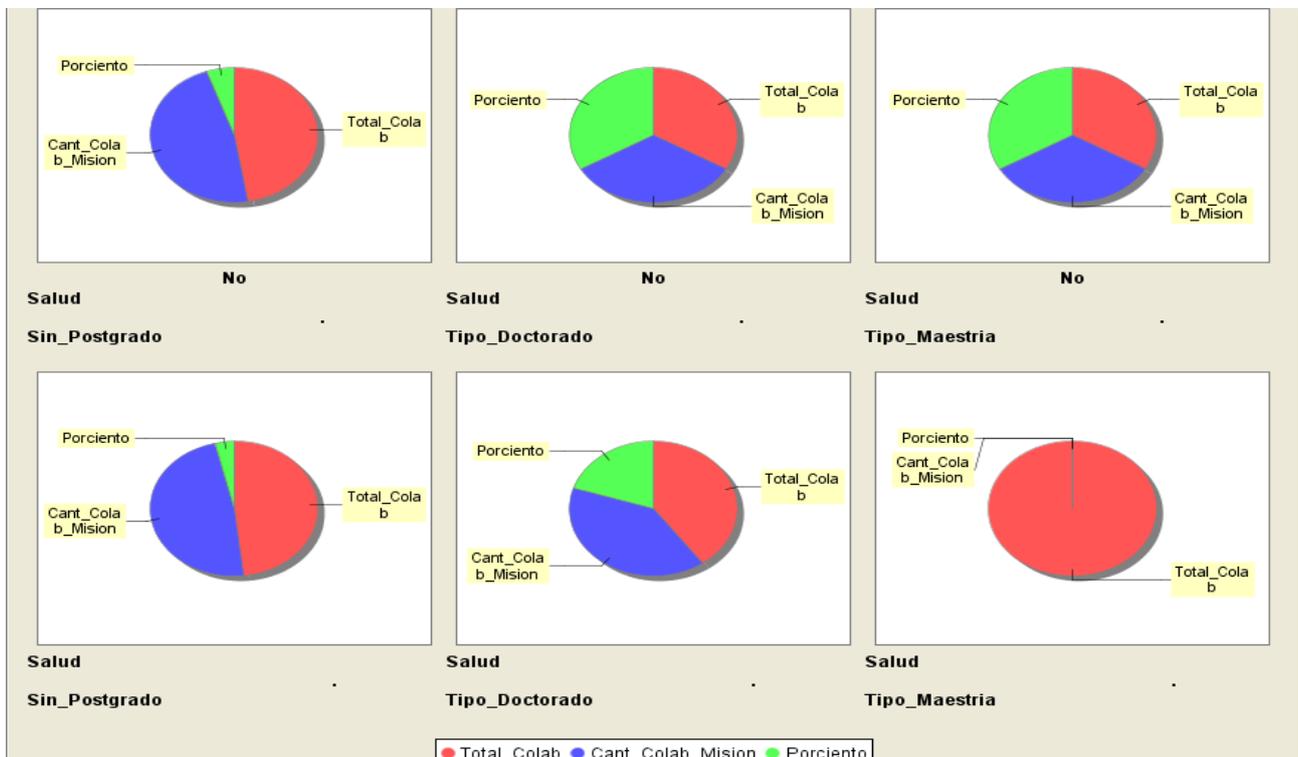


Fig. 41 Gráfica pastel por columnas.

3.5. Despliegue de la solución

El despliegue del Data Warehouse para el Recurso Humano de la Salud necesitará un total de 4 servidores, de los cuales tres sostendrán las bases de datos de los sistemas operacionales de Colaboración implementado sobre el gestor de base de datos MySQL 5.0, Docencia Médica y el Data Warehouse implementado sobre el gestor de base de datos PostgreSQL 8.3, y es necesario un servidor de aplicación web donde se ejecuta Pentaho Mondrian sobre JBoss.

La comunicación entre los servidores se realiza mediante el protocolo TCP/IP. Las PCs clientes pueden consultar el Data Warehouse a través de la herramienta Mondrian, conectándose por protocolo HTTP, existen PCs clientes que contienen una distribución de escritorio de Mondrian y la comunicación con el Data Warehouse se establece mediante el protocolo TCP/IP.

Los servidores de base de datos deben de contar con un mínimo de 150 Giga Bytes de capacidad de almacenamiento en disco y 1 Giga Bytes de memoria RAM, el servidor de aplicación se recomienda tenga 2 GB de memoria RAM y las PCs cliente deben contar con 60 Giga Bytes de capacidad de almacenamiento en disco y 256 Mega Bytes de memoria RAM.

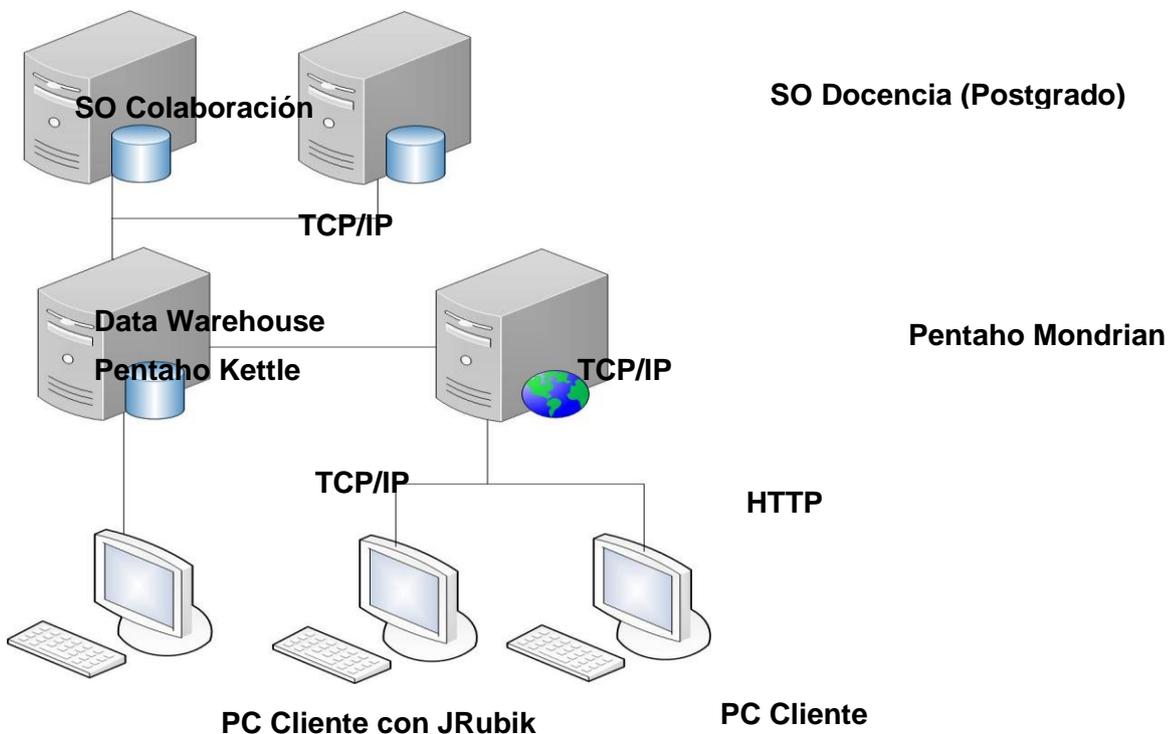


Fig. 42 Modelo de despliegue.

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

Con la utilización de las herramientas propuestas en el capítulo 1 fue implementado con éxito el cubo de datos multidimensional que consiste en la jerarquización de los atributos e implementación de los indicadores. Con el proceso OLAP los usuarios pueden mejorar la toma de decisiones sobre los Recursos Humanos por los informes obtenidos en forma de tabla y gráficos. Lo que permite satisfacer las disímiles necesidades de información en el nivel administrativo del Ministerio de Salud Pública representado por el Observatorio Nacional de los Recursos Humanos de la Salud.

CONCLUSIONES

A partir de los resultados obtenidos se puede concluir que:

- Con el estudio de los almacenes de datos actuales, se evidenció que la mayoría abordan los temas de Producción; Finanzas y Precios; Compra y Venta; no están vinculados con el proceso de consultoría de la información sobre los Recursos Humanos. No obstante, con su estudio se adquirió una visión de cómo resolver el problema, lo que permitió un mayor avance en la concepción e implementación de la solución.
- Las herramientas utilizadas en el desarrollo del trabajo han sido seleccionadas cuidadosamente respondiendo a las posibilidades tecnológicas de la institución.
- La implementación del Data Warehouse demostró ser la solución idónea en el nivel administrativo del Sistema nacional de salud, representado por el Observatorio Nacional de los RRHH de la Salud. Esta permite apoyar la toma de decisiones de los mismos, con relación a las actividades de Colaboración Médica y la Docencia en la formación postgraduada.

BENEFICIOS

Con la utilización del Almacén de datos los especialistas del Observatorio Nacional de los RRHH contarán con una herramienta que les permitirá:

- Concebir al Recurso Humano como ente único en el sector de la Salud.
- Conocer la cantidad, por ciento y total de los RRHH, clasificándolos por diversos criterios (estado actual, Categoría Científica, Localidad Geográfica Internacional, Tipo de Profesional) y realizando combinaciones entre estos.
- Representar gráficamente los resultados permitiendo una mejor interpretación de los mismos.
- Conocer los RRHH disponibles y sus características.
- Una mejor decisión sobre la asignación de los RRHH donde realmente sea necesario según su formación postgraduada y experiencia en la actividad de la colaboración.
- Disminuir considerablemente la demora en el envío de reportes a las instituciones solicitantes.

RECOMENDACIONES

Con el objetivo de mejorar el rendimiento y uso del Data Warehouse para apoyar la toma de decisiones y la planeación estratégica, se recomienda trabajar en las siguientes líneas:

- Utilizar un gestor multidimensional de base de datos sobre el Sistema Operativo Linux.
- Diseñar reportes para mejorar la captura de información sobre el almacén de datos, mediante la herramienta Pentaho Report.
- Utilizar la herramienta Pentaho WEKA sobre el Data Warehouse, para satisfacer las necesidades del nivel estratégico del Observatorio Nacional de los Recursos Humanos y de esta manera aplicar técnicas de Minería de Datos.

REFERENCIAS BIBLIOGRÁFICAS

1. **autores, Colectivo de.** Importancia de la utilización de un Data Warehouse (DW) en las empresa. [Online] 01 2009.
<http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf..>
2. Almacen de Datos. [Online] 2009.
http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos#Definici.C3.B3n_de_Bill_Inmon..
3. **Wolff, Gloria Carmen.** La Tecnología Datawarehousing. La Tecnología Datawarehousing. [Online] 2009. <http://www.inf.udec.cl/~revista/ediciones/edicion3/cwolff.PDF..>
4. Digital Granma Internacional. [En línea] <http://www.granma.cu/espanol/2007/junio/sabado16/32mil-colaboradores-salud-cubanos-76-paises.html..>
5. **Pérez de Armas, , Nerely Dialys.** nueva perspectiva de consulta para las empresa.
6. Idem[1]. [Online]
7. Idem[6]. [Online]
8. **Gutiérrez Echeverría, Damián.** Data WareHouse. [En línea] <http://www.monografias.com/trabajos17/data-warehouse/data-warehouse.shtml..>
9. Data Warehousing. [En línea] <http://www.sqlmax.com/dataw1.asp..>
10. **Herrera Basurto, Cristhian Kirs.** Apuntes Data Warehouse. [Online] 2007.
11. **Marrero Vázquez, Marlon, Nuñez, Yuri y Quinta Ladrón de Guevara, Liliani.** Módulo para el control de pago a los Colaboradores. 2008.
12. **Bernabeu, Dario Ricardo.** DATA WAREHOUSING:Investigación y Sistematización de ConceptosHEFESTO:Metodología propia para la Construcción de un DataWarehouse. Argentina : s.n., 2007.
13. Idem[12]. [Online]
14. Idem[12]. [Online]
15. Idem[12]. [Online]
16. Idem[12]. [Online]
17. Idem[12]. [Online]
18. Idem[12]. [Online]
19. Idem[12]. [Online]

20. Idem[12]. [Online]
21. Idem[12]. [Online]
22. Idem[12]. [Online]
23. Idem[12]. [Online]
24. Idem[12]. [Online]
25. Idem[12]. [Online]
26. Idem[12]. [Online]
27. ARQUITECTURA DEL DATA WAREHOUSING. [En línea] 28 de 07 de 2009.
<http://www.dataprix.com/es/arquitectura-del-data-warehouse>.
28. Idem[27]. [Online]
29. Idem[27]. [Online]
30. Idem[27]. [Online]
31. Metodología Hefesto. 2008.
32. Idem[31]. [Online]
33. **Gutiérrez Kafati, Elizabeth**. La plataforma Pentaho Open Source Business Intelligence. [Online] 06 25, 2008.
http://egkafati.bligoo.com/content/view/219538/La_plataforma_Pentaho_Open_Source_Business_Intelligence.html.
34. BUSINESS INTELLIGENCE SOLUTIONS. [En línea]
http://cba365.com.ar/cms/index.php?option=com_content&view=article&id=53&Itemid=65.
35. Idem[34]. [Online]
36. Herramientas y suites de la inteligencia empresarial.
37. Idem[36]. [Online]
38. Idem[36]. [Online]
39. Idem[36]. [Online]
40. [Online] <http://mondrian.pentaho.org/documentation/workbench.php>.
41. Enterprise Architect - Herramienta de diseño UML. [Online]
<http://www.sparxsystems.com.ar/products/ea.html>.

REFERENCIAS BIBLIOGRÁFICAS

42. Publican versión final 8.3 de PostgreSQL. [En línea]
<http://www.postgresql.org/about/press/presskit82.html.es..>
43. Nuestra Tecnología . [En línea] 2009. <http://www.naturasoftware.com/main.php?f=tecnologia>.

BIBLIOGRAFÍA

1. El Datawarehouse: Nueva perspectiva de aconsulta para las empresas. [En línea]
<http://www.monografias.com/trabajos16/datawarehouse/datawarehouse.shtml>.
2. **Coellar, Guillermo A.** DATA WAREHOUSE, ASPECTOS TÉCNICOS, CARACTERÍSTICAS, USOS, BENEFICIOS, COMPONENTES, HERRAMIENTAS OLAP. [En línea]
<http://fccea.unicauca.edu.co/old/datawarehouse.htm>.
3. INTRODUCCION AL CONCEPTO DATA WAREHOUSING. [En línea]
<http://www.ongei.gob.pe/publica/metodologias/Lib5084/11.HTM>.
4. Importancia de la utilización de un Data Warehouse (DW) en las empresas. [En línea]
<http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
5. Data WareHouse. [En línea] <http://www.monografias.com/trabajos17/data-warehouse/data-warehouse.shtml>.
6. Data Warehousing . [En línea] <http://www.sqlmax.com/dataw1.asp>.
7. Almacén de datos. [En línea] http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos.
8. Diseño Físico del Data Warehouse. [En línea]
http://www.betsime.disaic.cu/secciones/tec_ma_05.htm.
9. Arquitectura de un Datawarehouse. [En línea]
<http://informationmanagement.wordpress.com/2007/06/03/arquitectura-de-un-datawarehouse/>.
10. Arquitectura del datawarehouse . [En línea] <http://www.dataprix.com/es/arquitectura-del-data-warehouse>.
11. Conceptos Datwarehouse. [En línea] <http://dmi.uib.es/~bbuades/datawarehouse/sld002.htm>.
12. ETL. [En línea] <http://es.wikipedia.org/wiki/ETL>.
13. Pentaho BI. [En línea]
http://cba365.com.ar/cms/index.php?option=com_content&view=article&id=53&Itemid=65.
14. FORO - PostgreSQL. [En línea]
http://www.lawebdelprogramador.com/news/mostrar_new.php?id=72&texto=PostgreSQL&n1=394636&n2=1&n3=0&n4=0&n5=0&n6=0&n7=0&n8=0&n9=0&n0=0..
15. La_plataforma_Pentaho_Open_Source_Business_Intelligence. [En línea]
http://egkafati.bligoo.com/content/view/219538/La_plataforma_Pentaho_Open_Source_Business_Intelligence.html.
16. Imagen de PostgreSQL. [En línea] <http://postgresql.uptodown.com/>.

17. Viva Linux! [En línea] <http://www.vivalinux.com.ar/soft/postgresql-8.3.html>.
18. PostgreSQL 8.3 ya disponible . [En línea] <http://www.astra.es/noticias/postgresql-8-3-ya-disponible>.
19. Herramientas UML para el diseño y construcción de software. [En línea] <http://www.sparxsystems.com.ar/products/>.
20. Enterprise Architect. [En línea] <http://www.sparxsystems.es/New/products/ea.html>.
21. Data Warehouse. [En línea] http://html.rincondelvago.com/dwh_4.html.
22. Creación de los cubos OLAP Mondrian con el Pentaho Cube Designer. [En línea] <http://pentaho.almacen-datos.com/cube-designer.html>.
23. Persistencia MOLAP, ROLAP, HOLAP. [En línea] http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
24. Manual de Construcción de un Data Warehouse . [En línea] <http://www.ongei.gob.pe/publica/metodologias/Lib5084/INDEX.HTM>.
25. Operaciones en un Data Warehouse. [En línea] <http://www.ongei.gob.pe/publica/metodologias/Lib5084/152.HTM>.
26. Mondrian Schema Workbench. [En línea] <http://mondrian.pentaho.org/documentation/workbench.php>.
27. Herramientas OpenSource de DW. [En línea] https://www.u-cursos.cl/ingenieria/2007/1/IN830/1/material_docente/previsualizar?id_material=129812.
28. JBoss Application Server 4.2. [En línea] <http://www.osalt.com/es/jboss>.
29. Herrera Basurto, Cristhian Kirs. Apuntes de Data Warehouse. 2007.
30. Ing. Bernabeu, Ricardo Dario. DATA WAREHOUSING: Investigación y Sistematización de Conceptos- HEFESTO: Metodología propia para la Construcción de un Data Warehouse. 07 de Noviembre de 2007.
31. JRubik. [Online] http://rubik.sourceforge.net/spanish/jrubik_es.pdf.
32. Análisis del estado de Mondrian (OLAP Open Source). [Online] http://www.telefonica.net/web2/todobi/Mayo07/Analisis_de_Mondrian.pdf.

ANEXOS

1. Ejemplos de Data Warehouse

1.1. Corporación CIMEX

Datos en forma tabular - Microsoft Internet Explorer

DataWareHouse de la CORPORACION CIMEX

Parámetros: Vacíos en filas, Total en filas, Subtotal en filas, Promedio en filas, Valor medio en filas

Modos de navegación: Sustituir, Subtítulo de nivel, Solo al padre, Pestañas y herramientas, Expandir y Contracer

Modos de visualización: 2D, 3D

	Ingreso	Costos	Ingreso Venta	Utilidad Venta
1er Trimestre 2002				
Alimentos, Bebidas	\$538,209.57	\$1,093,605.01	\$558,091.23	
Automotriz y Combustibles	\$59,534.97	\$89,822.61	\$41,528.38	
Gastronomía	\$246.48	\$161,906.59	\$97,310.59	
SubTotal	\$587,991.02	\$1,345,334.21	\$696,930.60	
2do Trimestre 2002				
Alimentos, Bebidas	\$394,012.10	\$1,066,863.21	\$567,469.29	
Automotriz y Combustibles	\$38,296.20	\$68,094.90	\$32,400.17	
Gastronomía	\$2,009.36	\$158,567.05	\$95,075.40	
SubTotal	\$634,317.66	\$1,293,525.16	\$694,944.86	
3er Trimestre 2002				
Alimentos, Bebidas	\$537,632.52	\$914,991.09	\$499,637.36	
Automotriz y Combustibles	\$63,968.51	\$80,662.05	\$39,003.69	
Gastronomía	\$204.00	\$150,640.65	\$86,390.49	
SubTotal	\$601,805.03	\$1,146,293.79	\$627,654.54	

Elaborar consulta - Microsoft Internet Explorer

DataWareHouse de la CORPORACION CIMEX

Tipos de Cubo: Proceso

Cubos: AnalisisComercial

Hechos de AnalisisComercial

- AnalisisComercial
 - Área
 - CIMEX
 - Fecha
 - Todas las fechas
 - 2000
 - 2001
 - 2002
 - 1er Semestre 2002
 - 2do Semestre 2002
 - Producto

Fila(s): 2000 (Trimestre)

Columna(s): Alimentos, Bebidas; Automotriz y Combustibles; Gastronomía

Indicadores Relevantes: Vacío, Referencia cruzada, Total, SubTotal, Promedio, Valor Medio

Botones: Borrar, Reiniciar, Restringir Todo, Consultar, Cargar, Salvar, Web, Excel

1.2. Data Warehouse de CUBAENERGIA



2. Proceso ETL

2.1. Dimensión Especialidad



2.2. Dimensión Estado



2.3. Dimensión Localización



2.4. Dimensión Misión



2.5. Dimensión Postgrado

