



Universidad de las Ciencias Informáticas

Facultad 2

**Sistema informático para el análisis de la información de la
gestión de los Recursos Humanos de la UCI**

**Trabajo de Diploma para optar por el Título de Ingeniero
en Ciencias Informáticas**

Autores:

Yusnelys Pedroso Malagon

Daynier Ramiro García Prats

Tutores:

Ing. Alexeis Joel Ochoa Reyes

Ing. Reydel Capote Coipel

Co-Tutor:

Adalennis Buchillón Soris

La Habana, junio de 2016

“Año 58 de la Revolución”



**“La innovación es
lo que distingue a
los líderes de los
seguidores”**

Steve Jobs
Steve Jobs

Declaración de Autoría

Declaramos que somos los únicos autores del sistema informático para el análisis de la información de la gestión de los Recursos Humanos y autorizamos a la Facultad 2 y a la Dirección de Capital Humano de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmamos la presente a los _____ días del mes de _____ del año_____.

Yusnelys Pedroso Malagon

Daynier Ramiro García Prats

Ing. Alexeis Joel Ochoa Reyes

Ing. Reydel Capote Coipel

Datos de Contacto

Tutores:

- ❖ **Ing. Alexeis Joel Ochoa Reyes** (ajochoa@uci.cu): Graduado de Ingeniería en Ciencias Informáticas en el año 2012 en la UCI, se desempeñó como desarrollador en el Centro de Informática Médica (CESIM). Ha realizado investigaciones vinculadas a la Informática Médica, Inteligencia Artificial y Minería de Datos. Ha participado en el desarrollo de sistemas de gestión de información para la Salud. Afiliado a la Sociedad Cubana de informática Médica (SOCIM) y de la Asociación Cubana de Reconocimiento de Patrones (ACRP).
- ❖ **Ing. Reydel Capote Coipel** (rcapote@uci.cu): Graduado de Ingeniería en Ciencias Informáticas del curso 2014-2015 en la UCI. Recién graduado en adiestramiento, sin experiencia docente. Pertenece al departamento de IGSW de la Facultad 2. Imparte las asignaturas de Subsistemas de Organizaciones, Matemática Discreta II y Fundamentos de Administración y Gestión de Organizaciones (Pregrado UCI).
- ❖ **Ing. Adalennis Buchillón Soris** (abuchillon@uci.cu): Graduada de Ingeniería en Ciencias Informáticas del curso 2011-2012 en la UCI. Especialista de Integración de Datos del Departamento de Desarrollo de Componentes del Centro de Tecnologías de Gestión de Datos (DATEC). Sin categoría docente.

De Yusnelys Pedroso Malagon:

La parte más emotiva de todas, “los agradecimientos” donde a veces se nos olvida alguien, o nos queda alguien por mencionar, pero tratare de mencionarlos a todos.

El primer agradecimiento, el más importante de todos, es para esas personas súper especiales, que son capaces de todo por nosotros: mamá y papí (Reyna y Ricardo), los reyes de mi vida. Que no haría por ellos, por verlos orgullosos y por demostrarles que el sacrificio valió la pena. Les agradezco por su eterno apoyo y por confiar en mí. Les doy las gracias por ser mis papás, porque sinceramente otros no hubiese querido. Son mis mejores amigos y los quiero un montón.

Los segundos mejores amigos, esos que siempre han compartido gran parte de nuestra vida, ya sea la tristeza, la alegría, el dolor, la soledad. Aquí no puede faltar mi agradecimiento a mi amiga Aloyma, nos conocemos desde pequeñas, estudiamos juntas y a pesar de la distancia me apoyó en todo momento y lo sigue haciendo todavía, miles de besos para ella.

Y qué decir de Bárbaro Guillermo, aquí no hay para cuando acabar y como le prometí, le dije que le iba a robar las palabras textuales que me dijo un día: “te quiero porque sé que debajo de toda esa dureza me quieres, pero no se... y no quisiera imaginarme el por qué no deseas demostrarlo y te quiero por lo que eres y no porque me lo demuestres, si fuera por eso ya no te querría”. Quiero que sepas que, aunque me sacas de quicio, siempre andamos peleando eres muy importante para mí; y espero seguir compartiendo contigo muchos años de mi vida. Te agradezco muchas cosas, tu apoyo, el haberme enseñado a programar, no fui tan buena alumna; pero me defendí. Arigato godaïmasu (Gracias) por estar siempre a mi lado y ayudarme con todo, gracias por ser mi Barbí aunque ya otros me han copiado el nombre. Daisukí (te quiero).

Le agradezco a mi querida Leske, mi amiga de todos los tiempos. A mi compañera de cuarto Yane. Las sorpresas que trae la vida, quien hubiera imaginado que íbamos a compartir tan buenos momentos. Gracias por los trucos de cocina y por todos los bellos momentos que vivimos. Gracias por enseñarme a bendecir la comida, las largas charlas por la madrugada, y por todos tus consejos. DTB (Dios te bendiga).

Le agradezco al malcriado de Curio, más conocido por Alfredo Viset, de este personaje tengo un sinfín de historias. Gracias por ponerme en forma y reventarme en la pista, eso me sirvió para ganarte y ser la 4ta a nivel UCI en 400 m. Gracias por tener el privilegio de entrenar con el Bolt de la nueva generación, la locomotora humana, campeón de campeones. Espero no te enredes y encuentres a la futura pronto, para que me invites a comer flamenco rosado, o guacaíca. Lo mejor para tí y que DTB (Dios te bendiga).

A mí queridísimo compañero de mesa, ahora compañero de tesis. El pepillo rubio de ojos claros. El chico más buscado por las redes sociales, ese Facebook se siente solo sin tí. En todo este tiempo has sido, mi amigo, mi hermano, incluso hasta mi novio dicen por ahí. Gracias por compartir este momento conmigo, sin tí esto no sería posible. Eres súper especial para mí, te quiero muchísimo.

A Damaris y Guillermo, que par, son como mis segundos papás. A ellos un agradecimiento enorme, por todo lo que han hecho por mí. Se puede decir que esta tesis es más suya que mía porque han estado pendientes de cada detalle; y sé que están presentes aquí conmigo mandándome su apoyo y cariño.

A esos tutores excelentes que tenemos: Ada, Alexei y Reydel, que han jugado un papel fundamental en la creación de este trabajo de diploma. Gracias por su dedicación y su tiempo para que todo saliera bien.

A la profe Rosa por soportarnos, a Bárbaro y a mí a pesar de tener su tiempo planificado. Gracias por su apoyo y ser nuestra querida profe. A los profes que nos han formado a lo largo de la carrera.

A Aurora y Bouza por estar siempre pendientes y preocuparse por mí. A la familia, mis hermanos, Yusniel por el que decidí cumplir mi sueño de estudiar aquí en la UCI, Alberto, mis cuñadas Yaliana, Cristina, Rosa. A la niñita María Alodya, que es como mi otra abuela, gracias a ella tengo los mejores consejos de vida. A mis abuelos, tíos, primos, a la gente del barrio.

A todas las bellas personas conocidas hasta ahora con las que compartí bellos momentos: Adyana mi estilista personal, la mejor de todas, Rafael, Yosberto, Luis, Joel (el entrenador) que lo conozco desde el primer día en la universidad y me ha ayudado mucho, me identifica con el nombre de papelito, a mi grupo de 5to. Besitos y gracias a todos.

Les agradezco a todas las personas que me decepcionaron en un momento determinado de mi vida; porque me enseñaron a ser fuerte y a darme cuenta de los tesoros de personas que tengo a mi lado, por darme esa fuerza y determinación en las cosas que hago.

Por último un agradecimiento importantísimo, que no por ser el último deja de ser especial. Un agradecimiento enorme para mí, por todo el sacrificio y los malos momentos ya pasados y superados para llegar hasta aquí, donde estoy ahora.

De Daynier Ramiro García Prats:

A mis queridos padres Mirurgia y Ramiro por todo el amor, esfuerzo y dedicación que me han dado, por los valores y enseñanzas que me han inculcado, y permitirme estar cumpliendo hoy uno de mis sueños; nunca tendré el tiempo para pagar el amor invertido en mí. Gracias a ambos, en fin, una sola cosa: los amo. A mi hermanito Jesús, sé que no he sido un buen hermano, pero solo quiero que sepas que te quiero.

A mi hermanita del alma Ofelia, por todos los momentos que hemos compartido, los buenos y los malos, los consejos y el sostén y apoyo que siempre me has dado. No tengo que decirte que te quiero porque espero lo sepas.

A Yase por la comprensión y apoyo, y más importante aún por el incondicional amor que me has mostrado. Sé que pocas veces no lo demuestro, pero no lo dudes te quiero y llegaste para quedarte. Hoy ya eres hija de mis padres.

A mis hermanos Raudel y Hamler, por convertirse en los hermanos que son para mí y todos los buenos momentos pasados juntos, así como las tantas veces que han estado para mí.

A Yusnelys por ser mi fiel compañera de mesa, de trabajos, mi amiga y convertirme en mi hermana y hoy estar cumpliendo este sueño junto a mí, mulatísima sabes que también te quiero mucho.

A mi familia de Guanabacoa por abrirme las puertas y hacerme sentir como en mi casa en especial a mi tía Virginia, a mis primas Milagríto, Yenni y Elisabeth.

A mi abuelita Elia, abu te quiero mucho.

A toda mi familia en general por el apoyo que siempre me han dado.

A mi primo Yadian que es mi hermano por los momentos pasados juntos.

A mi nuevo tío, Fernando, por adoptarme como su sobrino y los consejos que siempre me das, de verdad muchas gracias. A la familia de Ofelia, Raudel y Yase que las considero mía.

A mi amigo Reydel, que hoy además es mi tutor, gracias por todo, los regaños y consejos, muchas veces te pones bravo conmigo con razón, más para eso son los

amigos. A Adalennís, entraste en mi vida ayudándome, luego te convertiste en mi tutora y hoy te considero más que eso, gracias por tus enseñanzas y el tiempo dedicado, sin tí el camino hubiese sido más largo. Hoy le doy las gracias a Oslý por ponerte en mi camino.

A Alexei, por ser mi amigo y tutor, por los consejos y las bromas en momentos de tensión, y más que nada por la confianza depositada en mí.

A Yaneísí, Viset y Yusnelys por la unidad, la familia y los amigos que son para mí.

Al profesor Hugo, por las incontables veces que me ha ayudado y considerarme su hijo. A todos los profesores que han contribuido en mi desarrollo profesional en todos los niveles de enseñanza.

A Yazmín, nunca pensé que después de discutir tanto en el aula, íbamos a terminar siendo amigos, gracias por dejarme conocerte.

A mi amiga Yarítza, por siempre estar presente desde primer año, siempre serás mi negra favorita.

A Juan Carlos por llegar a convertirse en mi hermanito, y ser especial.

A todo el piquete de amigos de profesores en la secundaria básica en especial a Wendys, Roniel, Ernesto, Eduardo, Yazmín y Yoan. A Ethan, por lo que me enseñaste y los momentos compartidos. A mis amigos, Rubén, Bárbaro, Pedro, Dayan, Lian, Renides, Gustavo, Adrián Sosa, Eiler, Yarítza, Leandro, Niharby, Annie, Alejandro Alegrant.

A mi antiguo grupo y a los que han permanecido Jorge Raúl, Roniel Martín, Bryan, Judith, Hairo, Héctor Bobadilla, Ernesto García, Alejandro Adonis, Anabel, Luis Felipe, José Luis, Héctor. A mi grupo de 5to año, Yosamy, Rey Alberto, Wisbel, Roniel, Adrián, Dairon, Daniel, Heilly, Javier, Jorge Bárbaro, José Gabriel, Miguel Ángel, Néstor, William.

A todos los buenos amigos de la FEU.

Y a todos los buenos amigos y todos los que han estado a mi lado en estos 5 años.

De Yusnelys:

A mis padres Reyna y Ricardo, por su sacrificio, dedicación y amor, todos estos años.

A mis hermanos Alberto y Yusniel, por el ejemplo, y la confianza.

A Aloyma, Barbaro y Leske, por ser los mejores amigos.

A todos los que me apoyaron y me brindaron su cariño, durante todo este tiempo.

De Daynier:

A mis queridos padres, Mirurgia y Ramiro, los que con amor lo han dado todo por mí y mi hermano, por el estímulo, la fuerza, la inteligencia y el ejemplo que me han brindado para poder triunfar.

A mi hermanito Jesús Lisandro para que sigas mi ejemplo y no seas como yo.

A Yase porque quiero que te gradúes conmigo. Y a todos los que me quieren.

Resumen

Con el desarrollo de la informática han surgido avances en las Tecnologías de la Información y las Comunicaciones (TIC), existiendo numerosas aplicaciones que permiten analizar, procesar y almacenar la información, mediante el análisis de los datos asociados a las organizaciones; para poder predecir el comportamiento futuro de estas ayudando así a tomar decisiones, que garanticen el éxito de las mismas.

En la Dirección de Desarrollo de Capital Humano de la Universidad de las Ciencias Informáticas (UCI), se trabaja con el fin de resolver los problemas presentados por esta área a la hora de gestionar y controlar los recursos laborales. Para ello en la universidad se utiliza el módulo de personal y nóminas de ASSETS ULTIMATE. Actualmente el sistema está más orientado al pago de los trabajadores y no brinda suficientes informes estadísticos, lo que dificulta el trabajo de los especialistas de Capital Humano, haciendo sus actividades laborales más complejas, obstaculizando el proceso de toma de decisiones.

El objetivo de la investigación es crear un sistema informático para el análisis de la información de la gestión de los Recursos Humanos, apoyándose en los *Data Mart* (Mercado de Datos) para obtener informes estadísticos y patrones de comportamiento que respondan a la tipicidad de la UCI, apoyando así el proceso de toma de decisiones en la Dirección de Desarrollo del Capital Humano de la universidad.

Palabras clave: *Análisis; Capital Humano; Data Mart; Decisiones; Universidad de las Ciencias Informáticas*

Índice de Contenido

Introducción.....	13
Capítulo 1: Fundamentación Teórica.....	19
1.1 Data Warehouse	19
1.2 Data Mart	20
1.3 ¿Por qué utilizar un Data Mart y no un Data Warehouse?.....	21
1.4 Proceso de extracción, transformación y carga de los datos	21
1.5 Sistemas de Procesamiento Transaccional en Línea y Procesamiento Analítico en Línea...	22
1.6 Procesamiento Analítico en Línea	22
1.7 Metodologías para implementar un Data Mart.....	27
1.8 Sistemas existentes	29
1.8.1 Ámbito nacional.....	29
1.8.2 Ámbito internacional.....	30
1.9 Herramientas para construir un Data Mart.....	31
1.9.1 Sistemas Gestores de Base de Datos	31
1.9.2 Herramientas de integración de datos	33
1.9.3 Herramientas para el Proceso Analítico en Línea.....	35
1.10 Herramientas de Modelado.....	37
Conclusiones del Capítulo.....	38
Capítulo 2: Propuesta de diseño e implementación del Data Mart.....	40
2.1 Propuesta del sistema	40
2.2 Descripción de las fases de la metodología Hefesto.....	40
2.2.1 Fase1: Análisis de requerimientos	41
2.2.2 Fase 2: Análisis de los OLTP	44
2.2.3 Fase3: Modelo lógico del Data Mart.....	51

2.2.4 Fase 4: Integración de datos.....	53
Conclusiones del Capítulo.....	56
Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart.....	57
3.1 Proceso de análisis de la información del Data Mart.....	57
3.1.1 Jerarquías.....	57
3.1.2 Medidas.....	57
3.1.3 Consultas multidimensionales.....	58
3.1.4 Resultados del cliente STPivot.....	59
3.2 Gestión del cambio lento en las dimensiones.....	64
3.3 Gestión de los metadatos del proceso de integración.....	65
3.4 Captura de los cambios en los datos.....	66
3.5 Pruebas y validación.....	69
3.6 Calidad de datos.....	70
3.7 Auditoría de datos.....	71
Conclusiones del Capítulo.....	71
Conclusiones.....	72
Recomendaciones.....	73
Referencias Bibliográficas.....	74
Bibliografía.....	77
Anexos.....	80
1. Tablas de dimensiones.....	80
2. Tabla del hecho empleado.....	83
3. Transformaciones del Data Mart.....	83
4. Jerarquía de los atributos.....	86
5. Acta de aceptación.....	92
6. Entrevista.....	93

Introducción

Las tecnologías de la información se han convertido a gran velocidad en parte importante de la vida cotidiana. A esta idea se asocia lo que se llama, **sociedad de la información** que conforma el conjunto de recursos necesarios para manipular todo tipo de información permitiendo convertirla, almacenarla, administrarla, transmitirla y encontrarla. El hombre ha estado constantemente explicando el mundo con el que interactúa y cómo utilizar su capacidad racionalizada en función de mejorar las condiciones de vida y existencia; por lo que, con el desarrollo de la ciencia y las tecnologías informáticas, a la sociedad, se le han impuesto nuevos retos por lo que debe superarse rápidamente para poder brindar los servicios de excelencia. Mediante la informatización de los diferentes sectores de la sociedad, y el auge de las tecnologías se han podido crear herramientas informáticas más especializadas y más enfocadas a brindar mejores soluciones a los problemas existentes. Con el vertiginoso desarrollo de los sistemas informáticos la información aumenta considerablemente, por lo que las organizaciones han decidido perfeccionar sus herramientas informáticas y utilizan los datos ya almacenados para crear y a su vez obtener nuevos conocimientos. (Gorbe, Tomás Guillén 2016)

Mediante el análisis de los datos asociados a las organizaciones, se puede explicar el pasado, comprender el presente y de algún modo también predecir el comportamiento futuro de estas (García Martínez, Ramón 2002). El análisis de los datos permite inspeccionarlos, limpiarlos y transformarlos con el objetivo de resaltar información útil, lo que permite a las instituciones tomar mejores decisiones ya sea eligiendo entre las opciones disponibles o entre las formas diferentes para resolver un problema actual, garantizando el éxito. Precisamente en el ámbito organizacional se le concede un valor incalculable al desarrollo del Capital Humano; la correcta gestión de este elemento es considerado un factor crítico y a su vez estratégico, por cuanto determina el éxito o no, de cualquier desempeño empresarial. (Álvarez Trujillo, Cealys 2012)

El Capital Humano es un factor de producción que depende no solo de la cantidad, sino también de la calidad, del grado de formación y de la productividad de las personas involucradas en un proceso productivo. Además, designa el conjunto de Recursos Humanos que posee una empresa o institución económica, que gestionando eficientemente su información permite una buena organización,

planificación y control de todo el personal dentro de la empresa (Cañibano Sánchez, D.^a Carolina 2016).

La Universidad de las Ciencias Informáticas (UCI) fue creada como parte de una estrategia gubernamental orientada a la informatización de la sociedad cubana. Es un centro docente-productor que desarrolla aplicaciones y servicios informáticos, orientados a diversos sectores de la economía y los servicios, dentro y fuera de Cuba («Portada | Portal de la Universidad de las Ciencias Informáticas» 2015). Tiene la misión no solo de formar profesionales en la rama de la informática; sino también el compromiso de convertirse en una universidad productiva, estratégica para el desarrollo científico, económico y social del país.

La Dirección de Desarrollo del Capital Humano de la UCI es el área encargada de llevar a cabo los procesos de selección e integración del personal, formación y desarrollo, retribución y estimulación, organización del trabajo y evaluación del desempeño de los trabajadores (UCI 2014). Para la gestión y control de los recursos laborales y la retribución a los trabajadores; en la universidad al igual que en el resto de los Centros de Educación Superior del país, es utilizado el módulo de personal y nóminas de ASSETS ULTIMATE; sistema informático comercializado por la firma panameña D'MARCO S.A y distribuido en Cuba por INFOMASTER, entidad informática perteneciente a la Empresa Nacional de Producción y Servicios de la Educación Superior (Infomaster 2014). Garantizar la continuidad de los principales procesos de la universidad, así como lograr y mantener la estabilidad del claustro del centro han sido objetivos de la UCI desde su creación. Conocer con exactitud el comportamiento de los principales factores que inciden sobre los procesos de la universidad, es interés de la Dirección de Capital Humano con el fin de establecer sus proyecciones y líneas de trabajo.

La información de los trabajadores de la UCI se encuentra almacenada por ASSETS en sus bases de datos; analizarla a cualquier nivel se ha convertido en un proceso engorroso y en muchos casos prácticamente imposible por la carencia de herramientas y métodos que lo permitan. Aunque ASSETS garantiza la gestión de la información transaccional de los empleados, entre sus procesos definidos no cuenta con ningún procedimiento orientado a la realización de análisis de información basado en técnicas de análisis de datos; que tras su ejecución sean capaces de brindar resultados que describan el comportamiento de las variables analizadas, y a su vez puedan ser utilizados para trazar proyecciones apoyando de forma sustancial a la toma de decisiones (Jose, Ramírez Quintana, Jose Manuel y Ferri Ramírez, Cesar 2004).

Gracias a la estadística se obtienen datos que se organizan, se clasifican y se analizan hasta obtener conclusiones para, posteriormente, tomar decisiones. Los datos de carácter estadístico poseen un valor trascendental; las decisiones más importantes de las organizaciones se toman con base a su aplicación y el conocimiento obtenido a partir de ellos. Las instituciones a todos los niveles trazan sus proyecciones a corto, mediano y largo plazo, mediante el estudio de los datos asociados a los informes estadísticos (Infomaster 2014).

Actualmente por parte del Ministerio de Educación Superior (MES), el Ministerio de Trabajo y Seguridad Social, la Oficina Nacional de Estadística (ONEI) y la propia universidad, son solicitados a la Dirección de Capital Humano de la UCI un conjunto de datos e informaciones estadísticas básicas donde se expone el estado del claustro y de los trabajadores del centro. El sistema informático ASSETS ULTIMATE, dentro de sus opciones cuenta con un conjunto de informes y reportes predeterminados; sin embargo, las condiciones actuales demuestran que la obtención de los datos para la posterior confección de los informes estadísticos se ha convertido en un proceso complejo, debido al creciente número de solicitudes de información, el origen y denominación de estas y la incapacidad actual de ser obtenidas automáticamente mediante el uso de esta herramienta. Estos factores traen consigo que en muchos casos los datos necesarios para la confección de los informes solicitados, deban ser realizados total o parcialmente por los especialistas de la Dirección de Capital Humano de forma semiautomática teniendo como desventaja:

- ✓ La probabilidad de introducción de errores durante la confección de los informes.
- ✓ El diseño y ejecución de consultas SQL a nivel de datos con los problemas de seguridad asociados a este tipo de procedimiento.
- ✓ La utilización de diferentes fuentes para la obtención de las informaciones lo que provoca que para la obtención de un determinado dato no siempre se use el mismo procedimiento ni sea consultada la misma variable.
- ✓ La demora de la respuesta por el uso de métodos rudimentarios.

Por los elementos antes mencionados y en función de apoyar el proceso de análisis de la información de los trabajadores en la Dirección de Desarrollo del Capital Humano de la UCI, se identifica el siguiente **problema de investigación**: el sistema informático utilizado en la Universidad de las Ciencias Informáticas para la gestión de los Recursos Humanos no garantiza la obtención de informaciones estadísticas básicas de los trabajadores.

Para dar respuesta a la situación anterior, el presente trabajo toma como **objeto de estudio**, el proceso de informatización de la información de los Recursos Humanos. Se propone como **objetivo general**, desarrollar un sistema informático que, mediante la integración de diferentes técnicas de análisis de datos, aplicadas sobre la información almacenados de los Recursos Humanos de la UCI, agilice el proceso de obtención de las informaciones estadísticas básicas de forma automatizada; delimitándose como **campo de acción**, el proceso de informatización para la gestión de la información de los Recursos Humanos en la UCI.

A partir del objetivo general planteado para la investigación se derivan los siguientes **objetivos específicos**:

1. Elaborar el marco teórico de la investigación para obtener los conocimientos necesarios que faciliten el desarrollo de la solución.
2. Desarrollar un mercado de datos que permita realizar el análisis de datos, seleccionando las herramientas, tecnologías y metodologías necesarias para su construcción.
3. Implementar un componente de software que permita la obtención de informes estadísticos sobre la gestión de los Recursos Humanos en la UCI.
4. Validar la efectividad de la herramienta propuesta, para comprobar que cumple con la especificación del problema planteado.

Para brindar una solución apropiada a la situación planteada, se dará cumplimiento a las siguientes **tareas de la investigación**:

1. Revisión bibliográfica de los sistemas informáticos destinados a la realización de análisis de información.
2. Realización de la fundamentación teórica como resultado de la revisión bibliográfica e investigación del estado del arte.
3. Valoración de la arquitectura y marco de trabajo a ser utilizados para la implementación de los componentes de software propuestos.
4. Construcción de un mercado de datos a partir de la información almacenada de los Recursos Humanos de la Universidad.
5. Implementación de un componente de software que permita obtener automáticamente las informaciones estadísticas básicas de la Dirección de Capital Humano.

6. Elaboración de pruebas que certifiquen la calidad de la solución y el cumplimiento del objetivo propuesto por la presente investigación.

Los **métodos científicos** que sustentan el presente trabajo de investigación son:

Teóricos:

Análisis histórico-lógico: se utilizó para realizar la valoración de sistemas o módulos informáticos existentes, y así conocer el funcionamiento y sus estructuras para el análisis de los datos de Recursos Humanos permitiendo el desarrollo de la presente investigación.

Analítico-sintético: se utilizó en el estudio de cada herramienta y tecnología, seleccionando lo esencial de cada una, viendo la relación existente entre ellas y el funcionamiento por si solas para el desarrollo de la investigación.

Inductivo-Deductivo: se utilizó en el momento de realizar casos de pruebas, a partir de la respuesta arrojada se llegó a conclusiones.

Modelación: se utilizó en la confección de modelos y diagramas que ayudan a la comprensión de los procesos a desarrollar dentro de la investigación.

Empírico:

Análisis Documental: se utilizó para establecer un marco teórico conceptual sobre la gestión de la información de los Recursos Humanos de la UCI.

Entrevista: se utilizó para determinar a través de las preguntas, los requerimientos necesarios para la construcción del sistema. (Barchini, Graciela Elisa 2004)

Como resultado de la investigación se pretende contar con un sistema informático que permita la integración de diferentes técnicas de análisis de datos, aplicadas sobre la información almacenada en los Recursos Humanos de la UCI, en aras de agilizar el proceso de obtención de las informaciones estadísticas básicas de forma automatizada. Mediante un nivel superior de análisis el sistema permitirá obtener modelos de comportamiento que sean capaces de describir las características de los indicadores de la gestión de los Recursos Humanos de la UCI, contribuyendo así al apoyo de la toma de decisiones de la Dirección de Desarrollo del Capital Humano.

El documento se encuentra estructurado por capítulos de la forma siguiente:

- ✓ **Capítulo 1: Fundamentación Teórica**

Se incluye el estado del arte del tema tratado tanto a nivel internacional como nacional. Se describen los conceptos fundamentales asociados al dominio del problema y se hace un análisis entre las soluciones existentes y la propuesta que se realiza. Se hace un estudio de los mercados de datos exponiendo sus elementos fundamentales como definición y características. También se describen las principales características de las tecnologías, metodologías, lenguajes y herramientas utilizadas para el desarrollo de la aplicación, utilizados para dar solución al problema planteado.

✓ **Capítulo 2: Propuesta de diseño e implementación del Data Mart**

Se describen los pasos a seguir para el diseño del DM haciendo uso de la metodología seleccionada, y de las técnicas de análisis de datos como el proceso ETL, donde se limpian, transforman y cargan los datos; para diseñar el cubo multidimensional. Se identifican las necesidades del cliente mediante el levantamiento de requisitos que debe cumplir el sistema, se definen las reglas del negocio, y el modelado de los datos con sus elementos tales como dimensiones, hechos y medidas.

✓ **Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart**

Se explica la construcción del cubo de datos, donde se definen las jerarquías y atributos de las dimensiones, además de la implementación de los valores numéricos llamados medidas.

Capítulo 1: Fundamentación Teórica

En el presente capítulo se incluye el estudio del estado del arte del tema tratado tanto a nivel internacional como nacional. Se describen los conceptos fundamentales asociados al dominio del problema y se hace un análisis entre las soluciones existentes y la propuesta que se realiza. Se hace un estudio de los *Data Mart* exponiendo sus elementos fundamentales como definición y características. También se tiene como objetivo describir las principales características de las tecnologías, metodologías, lenguajes y herramientas utilizadas para el desarrollo de la aplicación, utilizados para dar solución al problema planteado.

1.1 Data Warehouse

En el contexto de la informática, un almacén de datos (del inglés *Data Warehouse*) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. La información histórica que el Data Warehouse (DW) contiene, permite analizar tendencias y efectuar predicciones con la ayuda de herramientas de inteligencia empresarial. El DW asegura que toda la organización utilice los mismos datos y éstos se transformen en valiosa información con la que adoptar decisiones. (Febles Rodríguez, Juan Pedro y González Pérez, Abel 2010)

Entre sus principales características se encuentran:

- ✓ Orientado al tema: La información 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.
- ✓ Integrado: Integra datos recolectados de diferentes sistemas operacionales de la organización y o fuentes externas.
- ✓ De tiempo variante: Los datos son relativos a un período de tiempo y estos deben ser integrados periódicamente, los mismos son almacenados como fotos que se corresponden a un período de tiempo.
- ✓ No volátil: Los datos que son almacenados no sufren ninguna actualización solo son incrementados. El período cubierto para un DW va de 2 a 10 años. (Lázaro 2012)

Como se puede apreciar el DW tiene un efecto positivo sobre toda la vida de la empresa, va desde el desarrollo de los Recursos Humanos, mejoras tecnológicas, retos en el dominio de las nuevas tecnologías, y por supuesto toma de decisiones más eficientes (Febles Rodríguez, Juan Pedro y González Pérez, Abel 2010).

1.2 Data Mart

El *Data Mart* (DM) es una versión especial del DW, son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades. Es consultado mediante herramientas OLAP (*On line Analytical Processing* - Procesamiento Analítico en Línea) que ofrecen una visión multidimensional de la información. Se puede decir que los DM son pequeños DW centrados en un tema o un área de negocio específico dentro de una organización.

Clases de Data Marts

Data Marts Dependientes

Los datos que se utilizan para poblar un DM provienen del DW. Esta estrategia es particularmente apropiada cuando un DW crece muy rápidamente y los distintos departamentos requieren sólo una pequeña porción de los datos contenidos en él.

Data Marts Independientes

Los datos que se utilizan para poblar el DM provienen de los sistemas operacionales y/o fuentes externas.

Problemática:

El hecho de tener varios DM independientes, sin ningún tipo de integración, puede hacer que las tareas de administración y mantenimiento se conviertan en un lastre. Se puede dar el caso de que algunos DM necesiten los mismos datos para dar respuesta a determinadas preguntas, por lo que tenemos redundancia de datos, ya que cada DM se alimenta con sus procesos de extracción y transformación propios. (Bernabéu, Ricardo Darío 2007)

1.3 ¿Por qué utilizar un Data Mart y no un Data Warehouse?

Un DM cubre las necesidades de un determinado departamento dentro de la organización, y el costo de su uso es inferior por lo que se lleva menor tiempo para construirlo y ponerlo a funcionar; mientras que un Data Warehouse cubre las necesidades de la organización en su conjunto, y el costo de su uso es mayor por lo que construirlo y ponerlo a funcionar llevaría más tiempo. La información que se desea gestionar es la del departamento de Recursos Humanos de la UCI, por lo que se hace necesario implementar un DM y no un DW.

1.4 Proceso de extracción, transformación y carga de los datos

ETL - este término viene del inglés de las siglas *Extract-Transform-Load* que significan Extraer, Transformar y Cargar y se refiere a los datos en una empresa. ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos, DM o bodega de datos. ETL forma parte de la Inteligencia Empresarial (*Business Intelligence*), también llamado "Gestión de los Datos" (*Data Management*). La idea es que una aplicación ETL lea los datos primarios de unas bases de datos de sistemas principales, realice transformación, validación, el proceso cualitativo, filtración y al final escriba datos en el almacén y en este momento los datos son disponibles para analizar por los usuarios. Se trata del proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos (denominada DM o DW) con el objeto de analizarlos. También pueden ser enviados a otro sistema operacional para apoyar un proceso de negocio. (BI 2015)

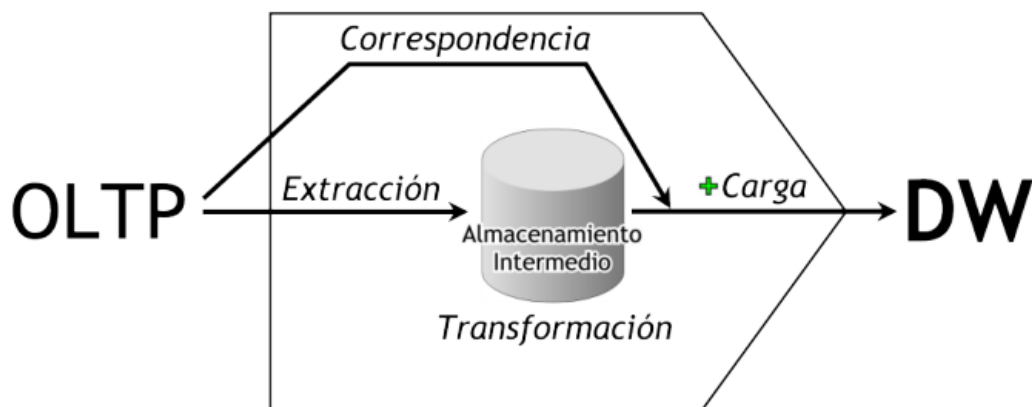


Figura 1. Proceso ETL (Bernabéu, 2010)

1.5 Sistemas de Procesamiento Transaccional en Línea y Procesamiento Analítico en Línea

Los sistemas de bases de datos operacionales realizan transacciones y resuelven consultas, por esta razón se denominan sistemas de Procesamiento Transaccional en Línea (*OLTP*, por sus siglas en inglés). Es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos, recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor que suelen ser utilizados por empresas con una red informática distribuida. Una transacción es un proceso que puede ser validado o abortado y puede involucrar operaciones de consultas, inserción, modificación y borrado de datos. Por otra parte, un DM se vincula más con el entorno decisional de la empresa, que se encuentra separado de la parte operacional y cuyo objetivo es proporcionar datos para su análisis y posterior toma de decisiones, así surgen los sistemas de Procesamiento Analítico en Línea (*OLAP*, por sus siglas en inglés). Estos sistemas organizan y presentan datos según la necesidad del usuario. (Bernabéu, Ricardo Darío 2007)

1.6 Procesamiento Analítico en Línea

Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos, entre otros. Este sistema es típico de un DM.

Existen varias arquitecturas para los sistemas OLAP:

- ✓ OLAP multidimensional (*MOLAP*, por sus siglas en inglés).

La arquitectura 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. Utiliza una arquitectura de dos niveles: las bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

- ✓ OLAP relacional (*ROLAP*, por sus siglas en inglés).

La arquitectura ROLAP, accede a los datos almacenados en un DW para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales. Utiliza una arquitectura de tres niveles. La base

de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

Para construir el DM de la solución se selecciona el sistema ROLAP, debido a que la arquitectura soporta mejor el análisis OLAP contra las bases de datos relacionales, coincidiendo esta arquitectura con el origen de los datos. Además, permite el uso total de la seguridad e integridad de los datos, para grandes volúmenes de información. (BI 2015)

Modelado Multidimensional:

Es una técnica utilizada especialmente por el DM; presenta la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios, además permite el acceso a la información mucho más rápido por parte de los manejadores de bases de datos. La información se almacena a través de tablas, dimensiones y hechos.

- ✓ Dimensiones: Objetos del negocio con los cuales se puede analizar la tendencia y el comportamiento del mismo. Las definiciones de las dimensiones se basan en políticas de la compañía, e indican la manera en que la organización interpreta o clasifica su información para segmentar el análisis facilitando la observación de los datos.
- ✓ Hechos: Concepto de interés primario para el proceso de toma de decisiones, corresponde a eventos que ocurren dinámicamente en el negocio de la empresa y contiene los hechos, indicadores o medidas del negocio que se desean analizar.
- ✓ Medidas: Características cualitativas o cuantitativas de los objetos que se desean analizar en las empresas. Las medidas cuantitativas están dadas por valores o cifras porcentuales. Por ejemplo, las ventas en dólares, cantidad de unidades principales, cantidad de unidades de productos vendidos, entre otros.

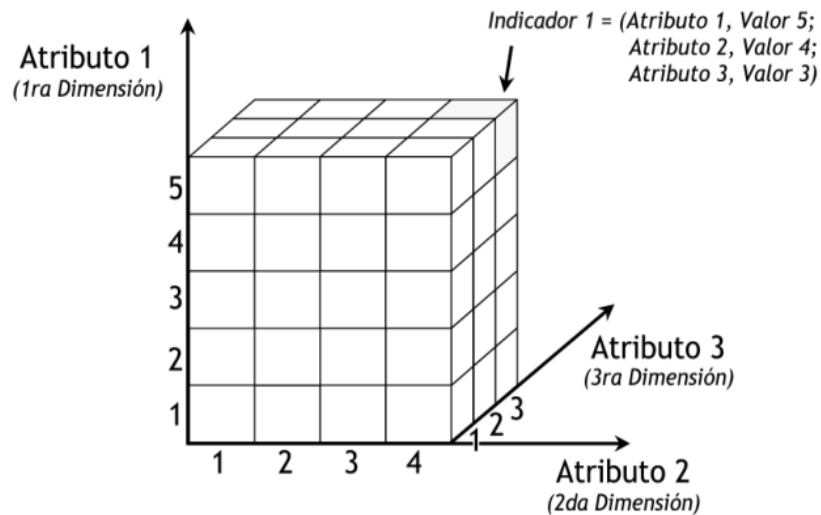


Figura 2. Modelo dimensional (Bernabéu, 2010)

Las bases de datos dimensionales implican tres variantes posibles de modelación, las cuales se mencionan a continuación:

✓ Esquema estrella.

Se llama así porque el diagrama se asemeja a una estrella. Contiene en el medio una tabla de hechos que se encuentra conectada a una o varias tablas de dimensiones. Las tablas de dimensiones solo se relacionan con la tabla de hechos y no existen relaciones entre dimensiones. Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que, en la tabla de hechos, la clave principal estará compuesta por las claves principales de las tablas dimensionales o una propia. Entre sus ventajas más significativas está que es el esquema más simple de interpretar, posee los mejores tiempos de respuesta, su diseño es fácil de modificar y simplifica el análisis. Sin embargo, es el menos robusto para la carga y es el más lento de construir.

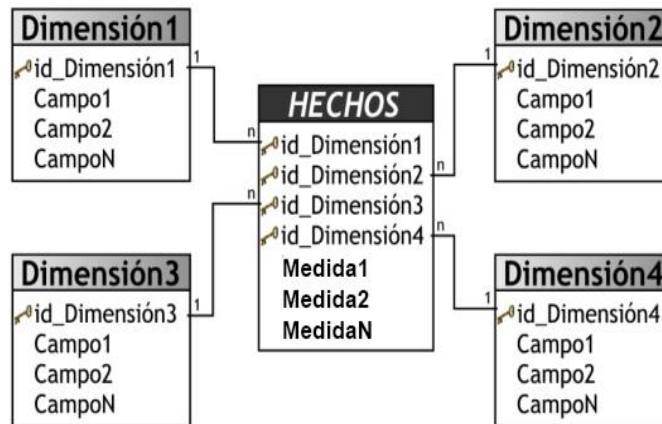


Figura 3. Esquema Estrella (Bernabéu, 2010)

- ✓ Esquema copo de nieve.

Es una extensión del esquema estrella en donde se presentan ramificaciones a partir de las tablas de dimensiones y no sólo a partir de la tabla de hechos. En esta forma de esquema, las tablas de dimensión pueden tener relación con otras tablas de dimensión. Tiene como desventaja que, de existir muchas tablas de dimensiones, cada una de ellas con varias jerarquías, pueden crearse bastantes tablas llegando a ser inmanejables. Además, su desempeño puede verse reducido si existen muchas uniones y relaciones entre tablas.

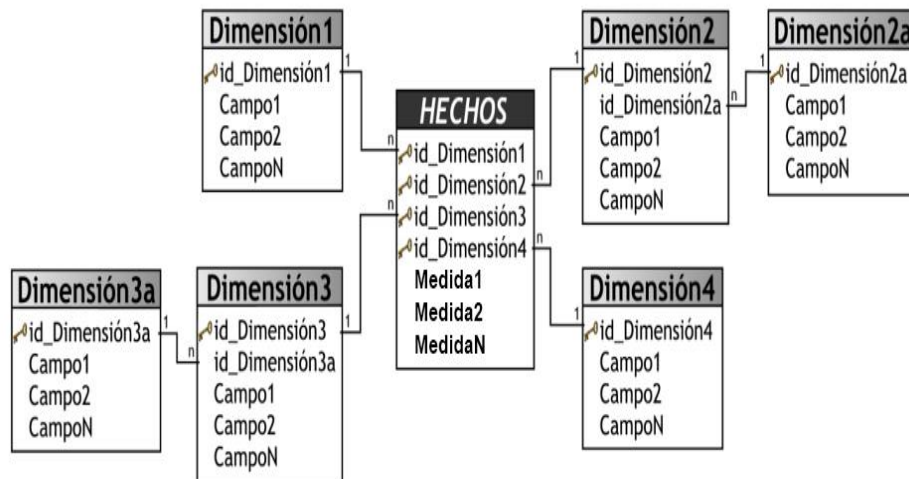


Figura 4. Esquema Copo de Nieve (Bernabéu, 2010)

✓ Esquema constelación de hechos.

Este modelo está compuesto por una serie de esquemas en estrella y lo conforman una tabla de hechos principal y una o más tablas de hechos auxiliares que están relacionadas con sus respectivas tablas de dimensiones, vinculándose las tablas de hechos auxiliares con algunas dimensiones asignadas a la tabla de hecho principal y también con nuevas tablas de dimensiones. Este esquema se destaca porque al tener más de una tabla de hechos facilita el análisis de más aspectos importantes del negocio y contribuye a la reutilización de las tablas de dimensiones, ya que pueden utilizarse para varias tablas de hechos.

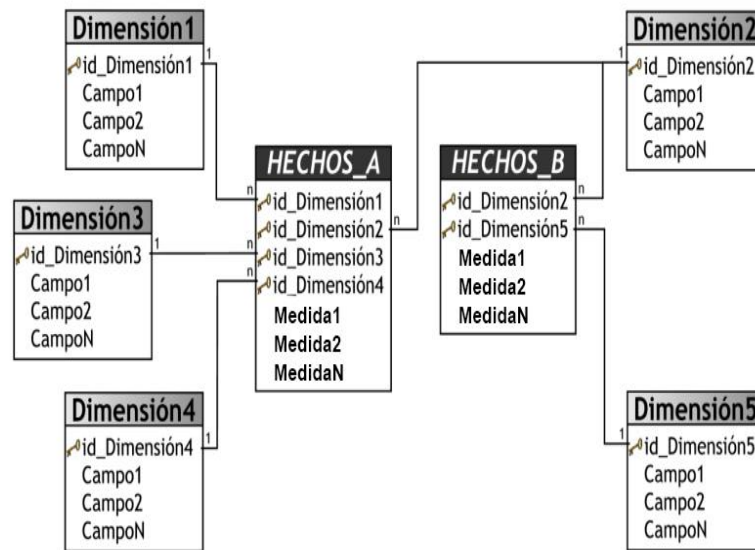


Figura 5. Esquema Constelación de Hechos (Bernabéu, 2010)

Se propone utilizar el modelado multidimensional estrella pues es ideal por su simplicidad y velocidad para ser usado en análisis multidimensionales. Además, el diseño de esquemas en estrella permite implementar la funcionalidad de una base de datos multidimensional utilizando una base de datos relacional. Otra razón para utilizar este tipo de esquema es su sencillez desde el punto de vista del usuario final. Es el esquema con mejor rendimiento y velocidad; que permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto. (Bernabéu, Ricardo Darío 2007)

1.7 Metodologías para implementar un Data Mart

Para construir un DM, han surgido un conjunto de metodologías que agilizan y rigen los procesos de desarrollo de los sistemas. Algunas de las metodologías más utilizadas para construir un DM son:

✓ Kimball

La metodología Kimball, llamada así en homenaje a su autor (Ralph Kimball), está basada en una arquitectura ascendente (*Bottom-Up*, por sus siglas en inglés); donde se definen previamente los DM y luego se integran en un DW centralizado. La arquitectura hace énfasis en los Data Marts. Contiene datos atómicos y sumariados. Ilustra la secuencia de tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación. Brinda una vista general que nos indica dónde estamos parados, por dónde pasamos y hacia dónde debemos dirigirnos. Kimball ofrece varias maneras de abordar los elementos para las etapas de desarrollo, y deja claro lo que se debe hacer, pero no cómo lograrlo, lo que provoca demoras en los resultados. (Yupa, Elsi Elizabeth Ilbay 2009)

✓ Inmon

La metodología Inmon, llamada así en homenaje a su autor (Bill Inmon), está basada en una arquitectura descendente (*Top-Down*, por sus siglas en inglés). Hace énfasis en los Data Warehouse, está compuesto por varios niveles de áreas de interés y DM dependientes. Contiene datos del Data Warehouse a nivel atómico, y datos del Data Marts sumariados. Esta metodología puede tener una implementación tardada, y es recomendada cuando se hace demasiado difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande. No es muy recomendable para proyectos sencillos pues va de lo general, el Data Warehouse, a lo más específico, el DM. (Kimball/Inmon 2012)

✓ Hefesto

Es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos, por lo que se encuentra en continua evolución.

Propone cuatro pasos fundamentales para su aplicación:

1. Análisis de requerimientos.
 - Identificar preguntas.
 - Identificar indicadores y perspectivas.

- Modelo Conceptual.
- 2. Análisis de los OLTP.
 - Conformar indicadores.
 - Establecer correspondencias.
 - Nivel de granulación.
 - Modelo conceptual ampliado.
- 3. Modelo lógico del DM.
 - Tipo de modelo lógico del DW.
 - Tablas de dimensiones.
 - Tablas de hechos.
 - Uniones.
- 4. Integración de datos.
 - Carga inicial.
 - Actualización.

Esta metodología cuenta con las siguientes características:

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

Se define Hefesto como la metodología a utilizar ya que permite construir el DM de forma sencilla, intuitiva y ordenada. Posee métodos, pasos lógicos que se relacionan durante las etapas del proceso de confección. Agiliza el proceso de desarrollo del DM debido a que propone no entrar en fases extensas de reunión de análisis y requerimientos ni fases de despliegue muy largas. (Bernabeu, Ricardo Dario 2007)

1.8 Sistemas existentes

Hoy en día las grandes empresas cuentan con sistemas informáticos que les ayuda con el proceso de toma de decisiones tanto a nivel nacional como internacional; debido al gran consumo de información, haciendo un mejor uso de sus datos ante un mercado competitivo.

1.8.1 Ámbito nacional

ASSETS ULTIMATE es un Sistema de Gestión Integral estándar y parametrizado que permite el control de los procesos de compras, ventas, producción, taller, inventario, finanzas, contabilidad, presupuesto, activos fijos, útiles y herramientas y Recursos Humanos. Como sistema integral todos sus módulos trabajan en estrecha relación, generando automáticamente al Módulo de Contabilidad los Comprobantes de Operaciones por cada una de las transacciones efectuadas, esto permite que se pueda trabajar bajo el principio de Contabilidad al Día. Dispone, además, de métodos novedosos para administración y planificación de inventarios, así como una amplia gama de análisis y consultas que le permitirán no sólo conocer exactamente la situación actual, sino proyectar decisiones futuras. Es un sistema flexible, amigable, con ayuda en línea que puede ser instalado en una microcomputadora o sobre varias, funcionando en ambiente multiusuario incluidas estaciones remotas. Proporciona opciones de seguridad que le permiten limitar el acceso a los diferentes procesos del sistema de acuerdo con el perfil de cada usuario. (Infomaster 2014)

Herramienta GREHU es un software que se encuentra implantado en gran variedad de empresas de Cuba. Cuenta con varios módulos o subsistemas como: inventario de personal, selección y contratación, puesto y caso, evaluación del desempeño, nóminas, pago por resultados, entre otros. El empleo de esta herramienta en los departamentos de personal y en los de economía ha implicado una más ágil realización del trabajo por parte de sus trabajadores. (López Paz, Carlos Ramón 2006)

Data Mart para la Unidad Central de Cooperación Médica (UCCM) mantiene un registro actualizado de los colaboradores y misiones médicas en el exterior y envía información detallada y

consolidada a los niveles superiores del Ministerio de Salud Pública (MINSAP), al Ministerio de Relaciones Exteriores (MINREX) y al Consejo de Estado. El DM brinda fácil acceso a los datos y permite que exista una fuente única de información relevante para la toma de decisiones. Los directivos pueden obtener una representación gráfica para una mejor comprensión y estudio en el momento de tomar decisiones. (Hidalgo López, Leydis 2013)

1.8.2 Ámbito internacional

Sistema automatizado para la selección de talento humano basado en perfiles de puesto de trabajo por competencias para el personal administrativo del colegio universitario de Caracas.

Selecciona el talento humano del personal administrativo adscrito al Colegio Universitario de Caracas, basados en perfiles de puestos de trabajo por competencias. Agiliza la búsqueda de información de una manera eficaz y rápida, lo que permitirá, la toma de decisiones, para la selección del talento humano que ocupe determinado cargo que necesite la Institución. (Martinez, 2010)

Análisis y diseño de un Data Mart para el seguimiento académico de alumnos en un entorno universitario. Proyecto de fin de carrera que utiliza el Data Mart en entornos universitarios, que describe con detalle el seguimiento académico de alumnos. (Rodríguez Sanz, 2010)

Valoración crítica

Después de haber realizado el análisis de los sistemas existentes, se concluye que del sistema nacional ASSETS ULTIMATE, entre sus procesos definidos no cuenta con ningún procedimiento orientado a la realización de análisis de información, basado en técnicas de análisis de datos que tras su ejecución sean capaces de brindar resultados que describan el comportamiento de las variables analizadas, y que puedan ser utilizados para trazar proyecciones apoyando la toma de decisiones. Los sistemas nacionales gestionan información de los trabajadores, pero como todo DM es único no analizan las mismas variables que emplea la universidad, por solo mencionar un ejemplo se tiene la variable “Docente” que indica si un trabajador es docente o no. Los sistemas analizados a nivel internacional, no gestionan información referente a los trabajadores, aunque brindaron información relevante para la construcción del DM. Por lo antes descrito se concluye que los sistemas existentes no brindan ninguna alternativa de solución al problema planteado.

1.9 Herramientas para construir un Data Mart

El sistema debe ser soportado por un conjunto de herramientas que cooperan entre sí para transitar por las etapas del proceso de análisis de datos, desde que se adquieren hasta que se visualizan los resultados.

Las herramientas son:

- ✓ Sistemas Gestores de Base de Datos.
- ✓ Herramientas de Integración de Datos.
- ✓ Herramientas para el proceso Analítico en Línea.

1.9.1 Sistemas Gestores de Base de Datos

Sistema Gestor de Base de Datos (SGBD) son un tipo de software específico, dedicados a servir de interfaz entre la base de datos, el usuario y las aplicaciones que lo utilizan. Las principales funciones que debe cumplir un SGBD, se relacionan con la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar redundancias e inconsistencias y mantener la integridad.

Entre los SGBD más conocidos se encuentran:

✓ Oracle.

Es un sistema de gestión de base de datos de tipo objeto-relacional (*ORDBMS*, por el acrónimo en inglés *de Object-Relational Data Base Management System*), desarrollado por *Oracle Corporation*. Se considera a *Oracle Database* como uno de los sistemas de bases de datos más completos, destacando: soporte de transacciones, estabilidad, escalabilidad, y soporte multiplataforma. Sus características principales son las siguientes: (Velasco 2011)

- Entorno cliente/servidor.
- Gestión de grandes bases de datos.
- Usuarios concurrentes.
- Alto rendimiento en transacciones.
- Disponibilidad controlada de los datos de las aplicaciones.
- Gestión de la seguridad.

✓ MySQL

Es un sistema de gestión de bases de datos relacional, multihilo y multiusuario bajo la Licencia Pública General (*GPL*, por sus siglas en inglés) de la GNU. Su diseño multihilo le permite soportar una gran carga de forma muy eficiente. Es una base de datos muy rápida en la lectura cuando utiliza motores no transaccionales, pero puede provocar problemas de integridad en entornos de alta concurrencia en la modificación de datos. (Hinz 2011). Las principales características de este gestor de bases de datos son las siguientes: (Pecos, 2010)

- Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo.
- Soporta gran cantidad de tipos de datos para las columnas.
- Gran portabilidad entre sistemas.

✓ PostgreSQL

Servidor de base de datos relacional, distribuido bajo licencia Distribución de Software Berkeley (*BSD*, por sus siglas en inglés) y con su código fuente disponible libremente. Incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. (PostgreSQL Global Development Group 2013). Las principales características de este gestor de bases de datos son:

- Implementación del estándar (lenguaje) SQL92/SQL99.
- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes (MAC, IP), cadenas de bits. También permite la creación de tipos propios.
- Incorpora una estructura de datos arreglos (array).
- Incorpora funciones de diversas índoles: manejo de fechas, geométricas, orientadas a operaciones con redes.
- Incluye herencia entre tablas (aunque no entre objetos, ya que no existen), por lo que a este gestor de bases de datos se le incluye entre los gestores objeto-relacionales.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.
- Soporta casi toda la sintaxis SQL (incluyendo sub consultas, transacciones, tipos y funciones definidas por el usuario).

- 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.

Luego del análisis realizado a los sistemas gestores de base de datos se determina que la herramienta a utilizar para la gestión de los datos en el DM es PostgreSQL, pues brinda estabilidad, integridad y seguridad en el sistema, siendo estas características esenciales para el manejo de información persistente y confiable de la base de datos al DM. Soporta mejor las acciones de procesamiento analítico en línea que Oracle. No se utiliza MySQL porque es rápida en la lectura cuando utiliza motores no transaccionales, y precisamente se desean realizar transacciones en línea, además puede provocar problemas de integridad en entornos de alta concurrencia en la modificación de datos. Con el PostgreSQL se utiliza PgAdmin que es la plataforma por defecto de este gestor de bases de datos. La interfaz gráfica que posee es compatible con todas las características de PostgreSQL y facilita su administración.

1.9.2 Herramientas de integración de datos

Las herramientas ETL son las encargadas de extraer datos de diferentes fuentes; la extracción se puede realizar de una base de datos, un excel, un XML (Lenguaje de Marcas extensibles) y otras fuentes. Estas herramientas también permiten transformar y limpiar datos que contengan errores, así como cargar los datos a cualquier sistema de almacenamiento, como por ejemplo un DW o un DM. Las principales herramientas de integración de datos son:

✓ Clover ETL.

Herramienta de código abierto, basado en Java, con un marco de integración de datos que pueden ser utilizados para transformar la información. Se puede utilizar independiente como una aplicación de servidor o puede ser empotrado en otras aplicaciones como una biblioteca de la transformación. Características principales de Clover ETL: («CloverETL Rapid Data Integration» 2008)

- Contiene una paleta de más de 40 componentes especializados de transformación
- Soporta los principales estándares de la industria de bases de datos (Oracle, MS SQL, DB2, entre otras) y varias variantes de código abierto (MySQL, PostgreSQL). Otras bases de datos cuentan con el apoyo a través de la capa de Conexión de Java a la Base de Datos (*JDBC*, por sus siglas en inglés).
- Lee y escribe datos en XML, excel y datos de longitud variable.

- Lectura y escritura de datos a través de protocolos FTP, SFTP, HTTP y HTTPS, se ejecuta en plataformas de 32 y 64 bits, Windows, Linux, Solaris y otros.

✓ Enhydra Octopus.

Herramienta de ETL basada en Java. Puede conectarse a cualquier base de datos usando Conexión de Java a la Base de Datos (*JDBC*, por sus siglas en inglés) y realizar transformaciones definidas en un archivo de lenguaje de marcas extensible (*XML*, por sus siglas en inglés). Es compatible con los modelos de datos por generar Identificadores de Objetos (*OIDs*, por sus siglas en inglés) para los nuevos objetos. Las claves naturales pueden utilizarse para insertar o actualizar los datos existentes y crear relaciones con los identificadores de objetos. (Softpedia 2012).

Características principales:

- Para acceder a las fuentes de datos que lo utilicen debe existir un manejador JDBC disponible.
- Herramienta poderosa capaz de normalizar datos, crear llaves artificiales, tablas y llaves primarias.
- Todos los trabajos que corren en Octopus son independientes de los proveedores de bases de datos.

✓ Pentaho Data Integration.

Pentaho Data Integration (PDI), ofrece capacidades de ETL mediante un enfoque innovador de metadatos. Con un ambiente intuitivo, gráfico, diseño de arrastrar y soltar (*drag and drop*, por sus siglas en inglés) y una arquitectura probada, escalable, basada en estándares, la integración de datos de Pentaho es la elección más demandada por las empresas. (Pentaho Solutions 2015). Uno de los objetivos de la integración de datos es que el proceso de ETL sea fácil de generar, mantener y desplegar. PDI se compone de cuatro herramientas: (Data PRIX 2010)

- SPOON: permite diseñar de forma gráfica la transformación.
- PAN: ejecuta un conjunto de transformaciones diseñadas con SPOON, conocidas como trabajos, creando dependencias entre las transformaciones.
- CHEF: permite, mediante una interfaz gráfica, diseñar la carga de datos incluyendo un control de estado de los trabajos.

- KITCHEN: permite ejecutar los trabajos por lotes diseñados con CHEF.

Características principales de PDI: (Ordaz, 2013)

- Tiene una interfaz visual con indicadores de las transformaciones.
- Es una aplicación escrita en Java con algunas características avanzadas escritas en Java Script.
- Basado en metadatos.
- Con respecto a escalabilidad, soporta la arquitectura de procesamiento en paralelo para distribuir las tareas de ETL a través de múltiples servidores.

Para el proceso de ETL se escoge la herramienta PDI, contando con características que facilitan la interacción con los datos, como puede ser su interfaz visual con indicadores de las transformaciones, la migración de datos entre aplicaciones y base de datos, la exportación de datos en archivos planos y la carga de forma masiva de las bases de datos. Se seleccionó porque es una herramienta completa que además del proceso ETL, también permite realizar la parte del procesamiento analítico en línea y la visualización del trabajo terminado. Es una herramienta que incluye todos los procesos anteriores, algo que no permiten las demás herramientas mencionadas.

1.9.3 Herramientas para el Proceso Analítico en Línea

Las herramientas OLAP proporcionan a las compañías un sistema confiable para procesar datos que luego serán utilizados para llevar a cabo análisis e informes que permitan mejorar las operaciones productivas, tomar decisiones inteligentes y optimizar la competitividad en el mercado. Las principales herramientas de procesamiento de la información son:

- ✓ Jedox Palo

Palo es una herramienta OLAP (*Online Analytical Processing*) de BI, que nos permite crear almacenar datos y generar informes en hojas de cálculo (vía Microsoft Excel / Open Office) o en el navegador Web. Esta herramienta está creada por la empresa Jedox y en las aplicaciones básicas son software libre, por lo que nos permite el uso básico a un costo cero, y además posee de herramientas más avanzadas de pago. («Jedox - Business Driven Intelligence» 2016)

- ✓ Olap4j

Interfaz de aplicación para el ambiente Java, que soporta la creación, almacenamiento y administración de datos para una aplicación OLAP (Morales, 2012). Permite consultar metadatos, y ejecutar las consultas de Expresiones multidimensionales (*MDX*, por sus siglas

en inglés), desde el entorno Java. Es la primera interfaz de aplicación Java que ha permitido a los programadores conectarse a servidores OLAP de diferentes proveedores. El primer piloto de aplicación es un consumidor de servicios web de Lenguaje de Marcado Extensibles de Análisis (*XMLA*, por sus siglas en inglés) genérico. Es capaz de conectarse a muchos motores de bases de datos en el mercado a través de servicios web de estilo Protocolo de Acceso a Objetos Simple (*SOAP*, por sus siglas en inglés). («Olap4j 1.0» 2011)

✓ Mondrian

Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho BI. Mondrian es un servidor OLAP *open source* que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente. Es decir, Mondrian actúa como “JDBC para OLAP” (Pentaho BI 2016). Permite analizar grandes conjuntos de datos que se encuentran almacenados en el DM, pues se encarga de recibir consultas dimensionales en lenguaje de Expresiones multidimensionales (*MDX*, por sus siglas en inglés) y devolver los datos del cubo que correspondan a la consulta. El cubo se representa como un conjunto de metadatos que definen cómo se han de mapear estas consultas dimensionales a sentencias SQL, para obtener de la base de datos la información necesaria para satisfacer la consulta dimensional (Pentaho a Hitachi Group Company 2016). Para acceder a las funcionalidades de Mondrian hay que hacer uso del cliente STPivot. Este cliente es un visor web OLAP, de código abierto, sobre la base del visor por defecto de JPivot, el cual es una librería de componentes Servidor de Páginas Java (*JSP*, por sus siglas en inglés), que se utiliza para construir tablas OLAP generadas de forma dinámica y permite a los usuarios realizar consultas OLAP por medio del lenguaje MDX. El objetivo de STPivot es mejorar la experiencia del usuario de JPivot mediante el aprovechamiento de las bibliotecas de interfaz de usuario libre y las tecnologías jQuery y Ajax (BI 2012). La herramienta que utiliza Mondrian para crear cubos OLAP es el Schema Workbench; que es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian, se utiliza para la creación de los archivos XML que se usan para la construcción de los cubos. Permite la ejecución de consultas MDX contra el esquema y la base de datos. (Espinosa, Roberto 2010)

De las herramientas propuestas se elige el servidor OLAP Mondrian, porque es la aplicación más importante de la plataforma Pentaho BI, que se seleccionó anteriormente. Además, cuenta con un conjunto de herramientas y plugins predeterminadas como el Schema Workbench que permite la

confección de cubos multidimensionales y se exploran dimensionalmente los datos con el visor web STPivot. Las otras herramientas no brindan este conjunto completo de herramientas que favorece el procesamiento analítico en línea.

1.10 Herramientas de Modelado

Pueden ayudar en los aspectos del ciclo de vida de desarrollo del software, en tareas como la realización del diseño del proyecto, cálculo de costes, implementación de una parte del código automáticamente con el diseño dado, compilación automática, documentación o detección de errores, entre otras. Las herramientas CASE (*Computer Aided Software Engineering*, Ingeniería de Software Asistida por Computadora) para el Lenguaje Unificado de Modelado (*UML*, por sus siglas en inglés) permiten analizar y diseñar orientado a objetos y abstraer el código fuente, a un nivel donde la arquitectura y el diseño se tornan más obvios y más fáciles de entender y modificar (SPARX System 2015).

Se llevó a cabo el análisis de las siguientes herramientas:

✓ Rational Rose.

Proporciona un conjunto de prestaciones controladas por modelo para desarrollar muchas aplicaciones de software, incluidas aplicaciones Ada, ANSI C++, C++, CORBA, Java, Java EE, Visual C++ y Visual Basic. El software permite acelerar el desarrollo de estas aplicaciones con código generado a partir de modelos visuales mediante el lenguaje UML (*Unified Modeling Language*). Es capaz de analizar la calidad del código y de generar código gracias a las prestaciones de sincronización configurable entre el modelo y el código. Permite la gestión granular y el uso de modelos con una característica de componentes de modelo que se puede controlar por separado. Proporciona el modelado UML para diseños de bases de datos e integra requisitos de datos y aplicación a través de diseños lógicos y físicos. (IBM 2016)

✓ Visual Paradigm.

Es una herramienta CASE: Ingeniería de Software Asistida por Computación. La misma propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y la documentación. Visual Paradigm ha sido concebida para soportar el ciclo de vida completo del proceso de desarrollo del software a través de la representación de todo tipo de diagramas. Se caracteriza por: (Pressman, Roger S. 2002)

- Disponibilidad en múltiples plataformas (Windows, Linux).
 - Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
 - Capacidades de ingeniería directa e inversa.
 - Modelo y código que permanece sincronizado en todo el ciclo de desarrollo.
 - Licencia gratuita y comercial.
 - Generación de código para Java y exportación como HTML.
 - Fácil de instalar y actualizar.
- ✓ Enterprise Architect.

Es una herramienta multi-usuario, diseñada para ayudar a construir software robusto y fácil de mantener. Además, permite generar documentación e informes flexibles y de alta calidad. Ofrece beneficios significativos para ayudar a construir modelos rigurosos de software, donde es posible mantener la trazabilidad de manera consistente. Proporciona trazabilidad completa desde el análisis de requerimientos y los artefactos de diseño, hasta la implementación y el despliegue. Es una herramienta de uso muy sencillo, que aborda el diseño y análisis UML y cubre el desarrollo de software desde la captura de requerimientos a lo largo de las etapas de análisis, diseño, pruebas y mantenimiento. (SPARX System 2015)

De las herramientas de modelado estudiadas se selecciona Visual Paradigm ya que esta herramienta soporta el ciclo de vida completo de un software; y como se desea implementar un DM con la metodología Hefesto que precisamente muestra paso por paso su ciclo de vida, ya sería una ventaja a la hora del análisis y el diseño de la aplicación por encima de las demás herramientas. Además, aumenta el conocimiento informático de una empresa ayudando así a la búsqueda de soluciones para los requisitos. También permite reutilizar el software, además de portabilidad y estandarización de la documentación usando metodologías propias de la Ingeniería de Software.

Conclusiones del Capítulo

En este capítulo se realizó un estudio teórico y conceptual acerca de los DM, a partir de búsquedas realizadas en la web, documentos, informes, revistas, artículos. Sobre la base del estudio y análisis de la información se llegó a la conclusión de que para implementar el DM se debía utilizar la metodología Hefesto, el PostgreSQL como gestor de base de datos, Pentaho Data Integration como herramienta para la integración de datos, Mondrian como motor de OLAP, STPivot para visualizar la

información gestionada del almacén y Schema Workbench para el diseño de los cubos. Como herramientas CASE se seleccionan Visual Paradigm para el diseño de los diagramas. También se realizó un estudio sobre la aplicación existente a nivel nacional que gestiona la información de los Recursos Humanos, sentando las bases para el análisis de las herramientas a utilizar en el desarrollo del DM e indagando sobre la ineficiencia del sistema informático existente en la UCI para garantizar niveles adecuados de análisis de la información que permita obtener estadísticas básicas del centro y apoye el proceso de toma de decisiones.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

En el presente capítulo se describen los pasos a seguir para el diseño del DM haciendo uso de la metodología seleccionada, y de las técnicas de análisis de datos como el proceso ETL, donde se limpian, transforman y cargan los datos; para diseñar el cubo multidimensional. Se identifican las necesidades del cliente mediante el levantamiento de requisitos que debe cumplir el sistema, se definen las reglas del negocio, y el modelado de los datos con sus elementos tales como dimensiones, hechos y medidas.

2.1 Propuesta del sistema

Se propone el desarrollo de un DM haciendo uso de la metodología Hefesto. Dando cumplimiento a sus fases se obtiene una herramienta que contribuye a la toma de decisiones referente a la gestión de la información de los Recursos Humanos de la UCI. Los especialistas de la Dirección de Desarrollo del Capital Humano utilizan las herramientas para analizar la información de sus trabajadores mediante reportes e informaciones estadísticas básicas del sistema.

2.2 Descripción de las fases de la metodología Hefesto

La metodología Hefesto cuenta con cuatro fases fundamentales que describen el desarrollo de un DM. La primera fase es “Análisis de requerimientos” donde se hacen preguntas claves del negocio y se identifican los requisitos y necesidades del cliente. Además, se identifican indicadores y perspectivas de análisis para la posterior confección del modelo conceptual de los datos. La segunda fase es el “Análisis de los OLTP”, donde se conforman los indicadores, se establecen correspondencias con los datos fuentes, se ve el nivel de granularidad y se realiza el modelo conceptual ampliado de los datos. La tercera fase es el “Modelo lógico del DM”, con sus dimensiones, hechos y uniones representando las jerarquías existentes entre las tablas. La cuarta fase es “Integración de datos” donde se definen los procesos de carga, transformación, extracción y limpieza de los datos fuente, a partir de la realización de la carga inicial y sus posteriores actualizaciones.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

2.2.1 Fase1: Análisis de requerimientos

Se realizó una entrevista (Ver Anexos: Entrevista) a los especialistas de la Dirección de Desarrollo del Capital Humano de la UCI, con el propósito de conocer sus necesidades como clientes, los resultados que desean obtener y los reportes que consideran importantes para el área. De la entrevista se obtienen las siguientes preguntas claves del negocio:

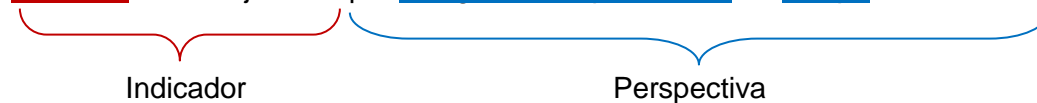
- ✓ Se desea conocer la cantidad de trabajadores por categorías ocupacionales en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por sexo y rangos de edades en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por tipos de contrato y cargo en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por rango de edades (menores de 35, de 35 a 50 y Mayores de 50), en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por categoría docente en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por grado científico en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por cargo y área en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por cargo y categoría de área en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por causa de baja en rango de fecha.
- ✓ Se desea conocer la cantidad de trabajadores por causa de alta en un rango de fecha.
- ✓ Se desea conocer la cantidad de trabajadores por profesiones en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores docentes por cargo en fecha determinada.
- ✓ Se desea conocer la cantidad de trabajadores por niveles de escolaridad.
- ✓ Se desea conocer la cantidad de Dirigentes docentes por Sexo, rangos de edad y categoría docente investigativa y grado científico.
- ✓ Se desea conocer la cantidad Dirigentes Administrativos por Sexo, rangos de edad y categoría docente investigativa y grado científico.
- ✓ Se desea conocer la cantidad de trabajadores por categoría ocupacional, sexo, rangos de edad y categoría docente investigativa y grado científico.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

Identificar indicadores y perspectivas

Una vez que se establecen las preguntas claves, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Para ello se debe tener en cuenta que los indicadores, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, entre otros. En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el tiempo es muy comúnmente una perspectiva. (Bernabéu, Ricardo Darío 2007) A partir de las preguntas identificadas en la entrevista realizada se obtuvieron los siguientes indicadores con sus correspondientes perspectivas:

Cantidad de trabajadores por categorías ocupacionales en tiempo determinado.



Cantidad de trabajadores por sexo y rangos de edades en tiempo determinado.



Cantidad de trabajadores por tipos de contrato y cargo en tiempo determinado.



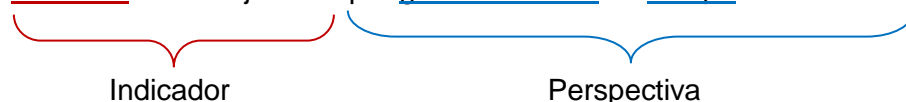
Cantidad de trabajadores por rango de edades (menores de 35, de 35 a 50 y Mayores de 50), en tiempo determinado.



Cantidad de trabajadores por categoría docente en tiempo determinado.



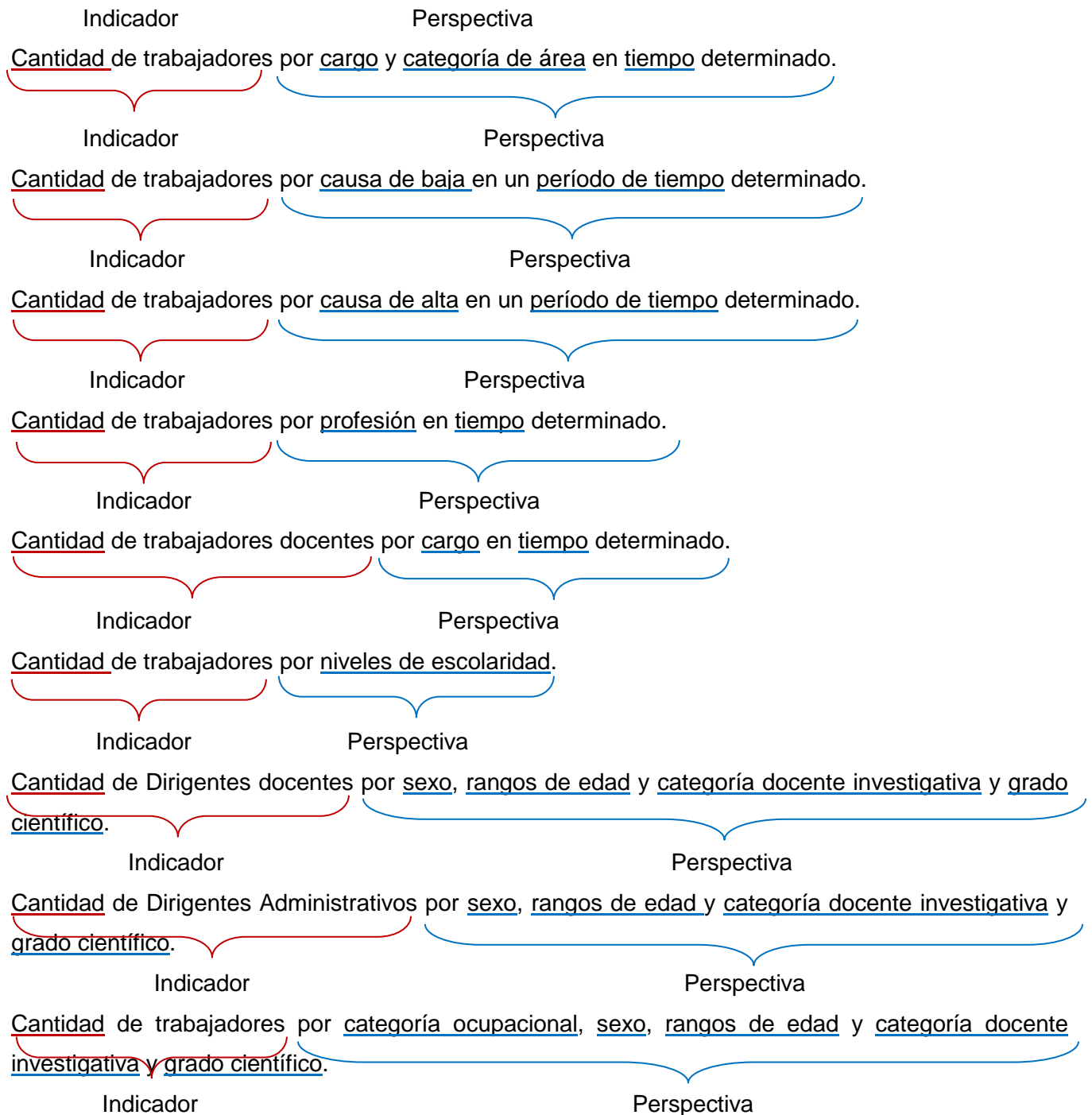
Cantidad de trabajadores por grado científico en tiempo determinado.



Cantidad de trabajadores por cargo y área en tiempo determinado.



Capítulo 2: Propuesta de diseño e implementación del Data Mart



Modelo Conceptual

Capítulo 2: Propuesta de diseño e implementación del Data Mart

Luego de identificar los indicadores y perspectivas se realiza el modelo conceptual de los datos del DM para que se pueda comprender el alcance del sistema y obtengan un alto nivel de definición de datos. La figura muestra a la izquierda las perspectivas y a la derecha el indicador.

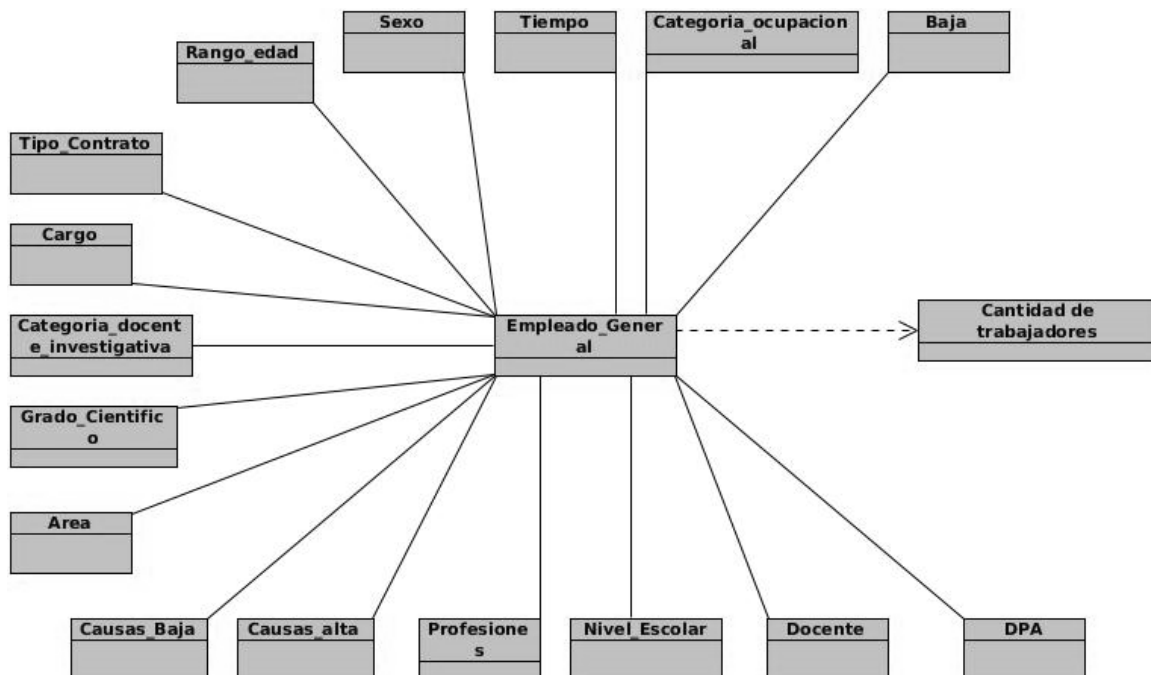


Figura 6. Modelo conceptual

2.2.2 Fase 2: Análisis de los OLTP

Se analizan las fuentes OLTP para determinar cómo se calculan los indicadores y establecer las correspondencias entre el modelo conceptual de la Figura 1 y las fuentes de datos. Se establecen los campos que se incluyen en cada perspectiva y se amplía el modelo conceptual con la información obtenida en este paso.

Conformar indicadores

Se obtiene el hecho que compone al indicador con su respectiva fórmula.

Hecho	Función de sumarización	Aclaración
Cantidad de trabajadores	distinct-count	Cuenta el total de trabajadores diferentes

Tabla 1. Conformación de indicadores

Capítulo 2: Propuesta de diseño e implementación del Data Mart

Establecer correspondencias

El objetivo es identificar las correspondencias entre el modelo conceptual y las fuentes de datos. En este caso para establecer las correspondencias entre la base de datos y el DM, las relaciones identificadas fueron las siguientes:

- ✓ Los campos “Desc_Categoria” y “Id_Categoria” de la tabla “RH_Categorias_Ocupacionales” se relaciona con la perspectiva “Categoria_ocupacional”.
- ✓ El campo “Sexo” de la tabla “Empleados_Gral” se relaciona con la perspectiva “Sexo”.
- ✓ Los campos “Id_CausaBaja” y “Des_CausaBaja” de la tabla “RH_Causas_Bajas” se relaciona con la perspectiva “Causas_baja”.
- ✓ Los campos “Id_CausaAlta” y “Des_CausaAlta” de la tabla “RH_Causas_Altas” de relaciona con la perspectiva “Causas_alta”.
- ✓ Los campos “Id_Profesion” y “Desc_Profesion” de la tabla “RH_Profesiones” se relaciona con la perspectiva “Profesión”.
- ✓ Los campos “Id_Nivel_Escolaridad” y “Desc_Nivel_Escolaridad” de la tabla “RH_Niveles_Escolaridad” se relaciona con la perspectiva “Nivel_Escolar”.
- ✓ Los campos “Id_Direccion”, “Desc_Direccion” y “Nota” de la tabla “RH_Unidades_Organizativas” se relaciona con la perspectiva “Area”.
- ✓ Los campos “Id_Empleado”, “No_CI” y “Fecha_Nacimiento” de la tabla “Empleados_Gral” se relaciona con la perspectiva “Rango_edad”.
- ✓ Los campos “Id_Cargo”, “Desc_Cargo” de la tabla “RH_Cargos” se relaciona con la perspectiva “Cargo”.
- ✓ El campo “Docente” de la tabla “Empleados_Gral” se relaciona con la perspectiva “Docente”.
- ✓ Los campos “Id_Grado_Cientifico” y “Desc_Grado_Cientifico” de la tabla “RH_Grados_Cientificos” se relaciona con la perspectiva “Grado_Cientifico”.
- ✓ Los campos “Id_Tipo_Contrato” y “Desc_Tipo_Contrato” de la tabla “RH_Contratos_Tipos” se relaciona con la perspectiva “Tipo_Contrato”.
- ✓ Los campos “Id_Empleado” y “Baja” de la tabla “Empleados_Gral” se relaciona con la perspectiva “Baja”.
- ✓ Los campos “Id_Categoria_DI” y “Desc_Categoria_DI” de la tabla “RH_Categorias_Docente_Invest” se relaciona con la perspectiva “Categoria_docete_investigativa”.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

- ✓ El campo “Fecha_Contratacion” de la tabla “Empleado_Gral” se relaciona con la perspectiva “Tiempo”.
- ✓ Los campos “municipio” y “provincia” de las tablas “Municipio” y “Provincia” se relacionan con la perspectiva “DPA”.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

Para comprender mejor la relación entre los campos de la base de datos y las perspectivas ver la tabla siguiente:

Campos	Perspectivas																	Nomencladores
	Categoria ocupacional	Sexo	Rango edad	Docente	Baja	Causa Baja	Causa alta	Profesión	Nivel escolar	Área	Cargo	Grado científico	Tipo contrato	Categoria docente investigativa	DPA	Tiempo		
	RH Categorías Ocupacionales	Empleados Gral				RH Causas Bajas	RH Causas Altas	RH Profesiones	RH Niveles Escolaridad	RH Unidades Organizativas	RH Cargos	RH Grados Científicos	RH Contratos Tipos	RH Categorías Docente Invest	RH Provincias	RH Municipios		
Desc_Categoria	X																	
Id_Categoria	X																X	
Sexo		X																
Id_CausaBaja					X												X	
Des_CausaBaja					X													
Id_CausaAlta						X											X	
Des_CausaAlta						X												
Id_Profesion							X										X	
Desc_Profesion							X											
Id_Nivel_Escolaridad								X									X	
Desc_Nivel_Escolaridad								X										
Id_Direccion									X								X	
Desc_Direccion									X									
Nota									X									
Id_Empleado			X		X												X	
No_CI			X														X	
Fecha_Nacimiento			X															
Id_Cargo										X							X	
Desc_Cargo										X								
Docente				X														
Id_Grado_Cientifico											X						X	
Desc_Grado_Cientifico											X							
Id_Tipo_Contrato												X					X	
Desc_Tipo_Contrato												X						
Baja					X													
Id_Categoria_DI													X				X	
Desc_Categoria_DI													X					

Tabla 2. Correspondencia con los OLTP

Nivel de granularidad:

Después de establecer las relaciones de los OLTP, se seleccionan los campos que tendrán cada perspectiva, por los cuales se examinarán y filtrarán los indicadores. Además, se presentan los datos de análisis disponibles para cada perspectiva; y se decide cuáles son los más relevantes para consultar los indicadores y cuáles no. La selección de los campos de cada perspectiva indica el nivel de granularidad de la información encontrada en el DM.

Perspectiva “Categoría_ocupacional”.

- ✓ “categoria_ocupacional_codigo”: Referido al identificador de la categoría ocupacional.
- ✓ “categoria_ocupacional_tipo”: Referido al tipo de categoría ocupacional de los trabajadores.
- ✓ “categoria_ocupacional_descripcion”: Referido al tipo de categoría ocupacional de los trabajadores, más detallado.

Perspectiva “Sexo”.

- ✓ “sexo_codigo”: Identificador del sexo.
- ✓ “sexo”: Sexo de cada trabajador (F o M)
- ✓ “sexo_descripcion”: Sexo de cada trabajador, más detallado (F-Femenino, M-Masculino)

Perspectiva “Causas_baja”.

- ✓ “causa_baja_codigo”: Identificador de la causa de baja.
- ✓ “causa_baja_tipo”: Tipo de la causa de baja.
- ✓ “causa_baja_descripcion”: Causa de baja, más detallado.

Perspectiva “Causas_alta”.

- ✓ “causa_alta_codigo”: Identificador de la causa de alta.
- ✓ “causa_alta_tipo”: Tipo de la causa de alta.
- ✓ “causa_alta_descripcion”: Causa de alta más detallada.

Perspectiva “Profesion”.

- ✓ “profesion_codigo”: Identificador de profesión.
- ✓ “profesion_tipo”: Tipo de profesión que posee el trabajador.
- ✓ “profesion_descripcion”: Profesión que posee el trabajador, más detallado.

Perspectiva “Nivel_Escolar”.

- ✓ “nivel_escolar_codigo”: Identificador del nivel escolar.
- ✓ “nivel_escolar_nombre”: Nivel escolar del trabajador.
- ✓ “nivel_escolar_descripcion”: Nivel escolar del trabajador, más detallado.

Perspectiva “Area”.

- ✓ “area_codigo”: Identificador de área.
- ✓ “area_nombre”: Área a la que pertenece el trabajador.
- ✓ “area_descripcion”: Área a la que pertenece el trabajador, más detallada.
- ✓ “area_tipo”: Tipo de área a la que pertenece el trabajador.
- ✓ “area_tipo_descripcion”: Tipo de área a la que pertenece el trabajador, más detallada.

Perspectiva “Rango_edad”.

- ✓ “rango_edad_codigo”: Identificador del rango edad.
- ✓ “rango_edad_nombre”: Rango de edades de los trabajadores.
- ✓ “rango_edad_descripcion”: Rango de edades de los trabajadores, más detallada.

Perspectiva “Cargo”.

- ✓ “cargo_codigo”: Identificador del cargo.
- ✓ “cargo_nombre”: Cargo que ejerce el trabajador.
- ✓ “cargo_descripcion”: Cargo que ejerce el trabajador, más detallado.

Perspectiva “Docente”.

- ✓ “docente_codigo”: Identificador de docente.
- ✓ “docente”: Indica si el trabajador es docente o no (1-Si, 0-No).
- ✓ “docente_descripcion”: Indica si el trabajador es docente o no docente.

Perspectiva “Grado_Cientifico”.

- ✓ “grado_cientifico_codigo”: Identificador del grado científico.
- ✓ “grado_cientifico_nombre”: Grado científico que posee el trabajador.
- ✓ “grado_cientifico_descripcion”: Grado científico que posee el trabajador, más detallado.

Perspectiva “Tipo_Contrato”.

- ✓ “tipo_contrato_codigo”: Identificador de tipo de contrato.
- ✓ “tipo_contrato_tipo”: Tipo de contrato que se le hace al trabajador.
- ✓ “tipo_contrato_descripcion”: Tipo de contrato que se le hace al trabajador, más detallado.

Perspectiva “Baja”.

- ✓ “baja_codigo”: Identificador de baja.
- ✓ “baja_estado”: Indica si el trabajador es baja o no (1-Si, 0-No.)
- ✓ “baja_descripcion”: Indica si el trabajador es baja, o si el trabajador no es baja.

Perspectiva “Categoria_docete_investigativa”.

- ✓ “categoria_docente_codigo”: Identificador de la categoría docente investigativa.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

- ✓ “categoria_docente_invest_tipo”: Tipo de categoría docente investigativa que posee el trabajador.
- ✓ “categoria_docente_invest_descripcion”: Categoría docente investigativa que posee el trabajador, más detallada.

Perspectiva “Tiempo”.

- ✓ “anno_codigo”: Identificador del año.
- ✓ “anno_nombre”: Número que identifica el año.
- ✓ “anno_numero”: Número que identifica el año.
- ✓ “mes_codigo”: Identificador del mes.
- ✓ “mes_nombre”: Nombre que identifica a cada uno de los meses.
- ✓ “mes_numero”: Número que identifica a cada uno de los meses.
- ✓ “dia_codigo”: Identificador del día.
- ✓ “dia_nombre”: Nombre que identifica el día de la semana.
- ✓ “dia_numero”: Número que identifica el día de la semana.
- ✓ “dia_fecha”: Fecha completa que abarca el año, el mes y el día.

Perspectiva “DPA”.

- ✓ “provincia_codigo”: Identificador de la provincia.
- ✓ “provincia_nombre”: Nombre que identifica a la provincia.
- ✓ “provincia_descripcion”: Nombre que identifica a la provincia, más detallado.
- ✓ “municipio_codigo”: Identificador del municipio.
- ✓ “municipio_nombre”: Nombre que identifica al municipio.
- ✓ “municipio_descripcion”: Nombre que identifica al municipio, más detallado.

Modelo Conceptual Ampliado:

Se amplía el modelo conceptual colocando en cada perspectiva los campos de interés seleccionados y en el indicador la función de sumarización.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

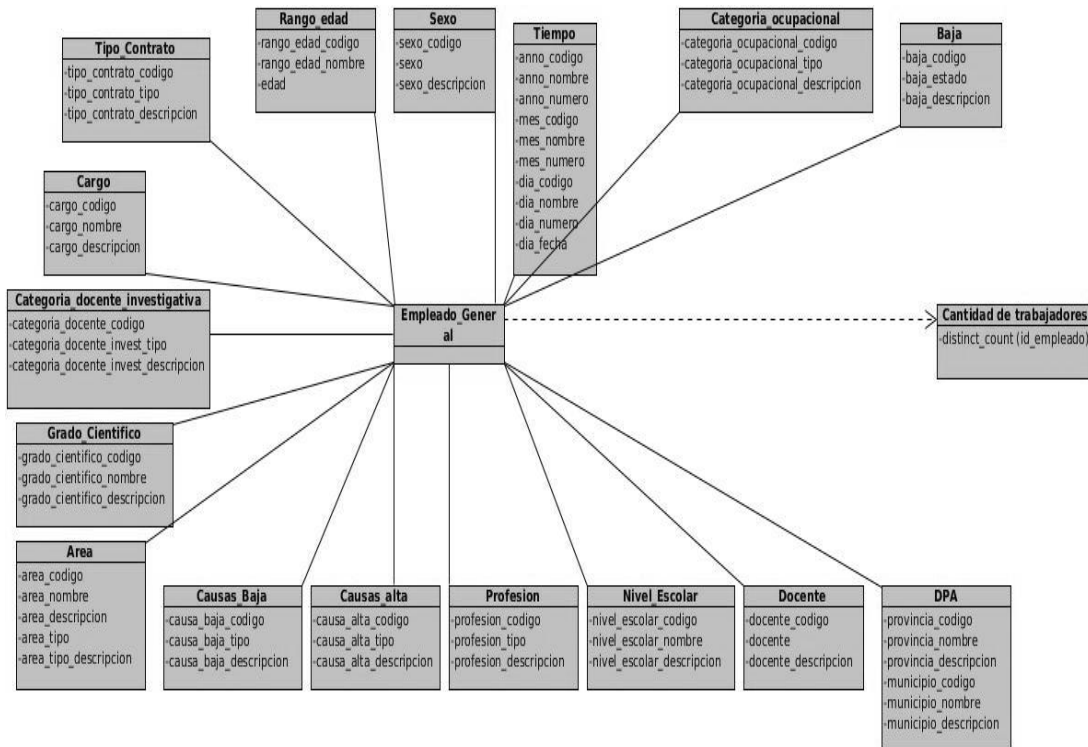


Figura 7. Modelo Conceptual Ampliado

2.2.3 Fase3: Modelo lógico del Data Mart

Se crea el modelo lógico del DM, teniendo como base el modelo conceptual creado. Se define el tipo de modelo a utilizar, para diseñar las tablas de hechos y de dimensiones. Después se realiza la unión, entre estas tablas.

Tipo de modelo lógico del Data Mart:

Para obtener la estructura del DM se selecciona el esquema en estrella, explicado en el capítulo anterior (Ver Figura3. Esquema Estrella).

Tablas de dimensiones:

Se diseñan las tablas de dimensiones que forman parte del DM. Cada perspectiva definida en el modelo conceptual constituye una tabla dimensión, escogiendo un nombre que la identifique. En la tabla de dimensión se añade el campo que representa su clave principal y se redefine los nombres de los campos que lo necesiten. A continuación, se presenta el ejemplo de la dimensión “Sexo”, para ver las demás dimensiones (Ver Anexos: Dimensiones del Data Mart):

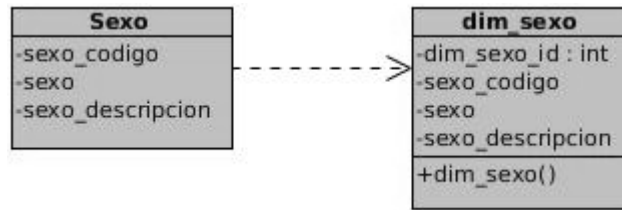


Figura 8. Dimensión Sexo

Tabla de hechos:

Se define la tabla de hechos, que contiene el hecho a través del cual se construye el indicador de estudio. También se le asigna un nombre a la tabla, que representa la información que se analiza, se define su clave primaria, además de la combinación de la llave primaria de las tablas de dimensiones.

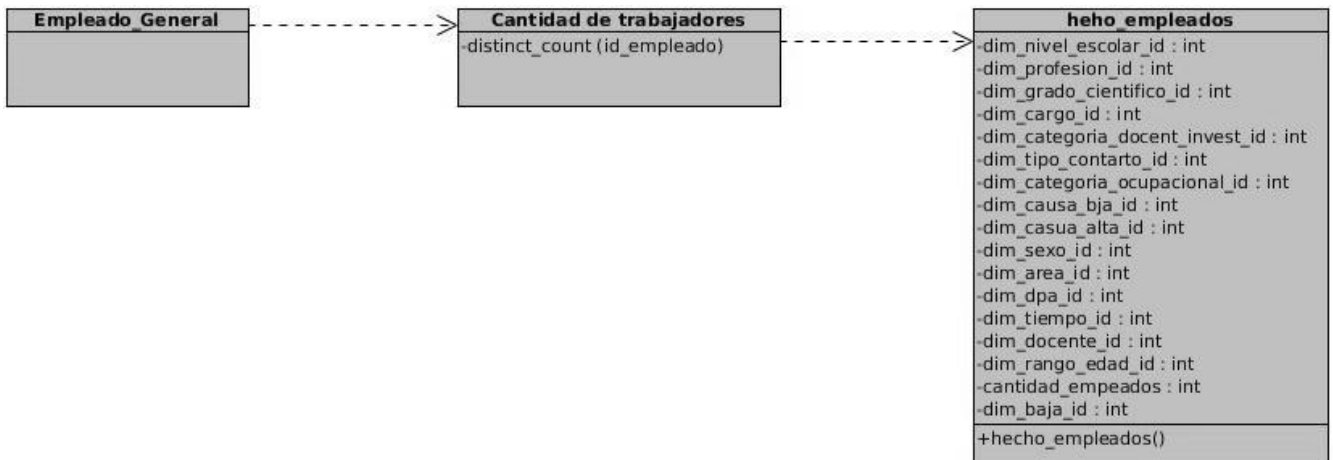


Figura 9. Diseño de la tabla de hechos empleados

Uniones:

Se realizan las uniones correspondientes entre las tablas de dimensiones y el hecho, mostrando la relación entre ellas.

Capítulo 2: Propuesta de diseño e implementación del Data Mart

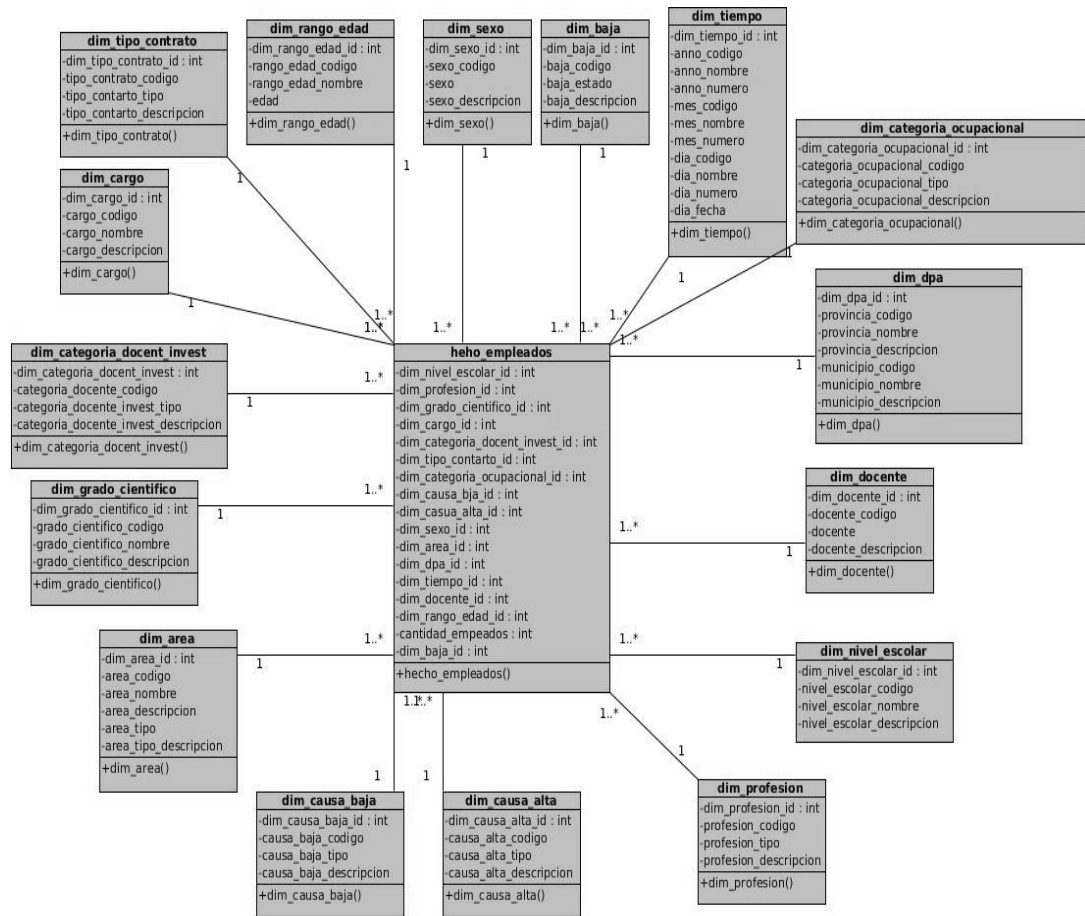


Figura 10. Uniones de las dimensiones con el hecho

2.2.4 Fase 4: Integración de datos

Después de construido el modelo lógico del DM, se procede a poblarlo de datos, utilizando las técnicas de limpieza, transformación y carga, conocidos como procesos ETL.

Carga inicial:

Para la carga inicial se utiliza la herramienta antes mencionada PDI, donde se realizan una serie de pasos para poblar el DM de datos, comenzando a partir de la confección de las transformaciones.

A continuación, se explica la transformación Sexo (Ver Figura 8. Dimensión Sexo), para ver las demás transformaciones (Ver Anexos: Transformaciones del Data Mart).

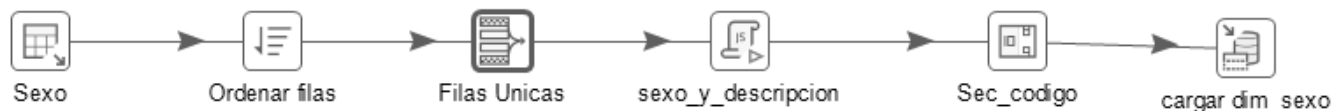


Figura 11. Transformación Sexo

En el caso de la dimensión sexo el proceso se inicia insertando en la dimensión el identificador nombre, si el sexo es F (Femenino) o M (Masculino). Para obtener los atributos, se procede de la forma siguiente:

- ✓ El primer componente llamado (Entrada Tabla) permite obtener la información desde la base de datos y se obtienen los campos asociados, se ordenaron por el identificador y el nombre a través del componente (Ordenar filas).
- ✓ A través del componente (Filas Únicas) se obtuvo una sola fila con el sexo del trabajador, por ejemplo, después de ordenar a todos los trabajadores y cargar sus sexos F (Femenino) o M (Masculino), las tuplas se repiten varias veces por lo que con el componente solo se toma la fila una sola vez, aunque aparezca varias veces.
- ✓ Con el componente (Java Script) se renombra el sexo F (Femenino) o M (Masculino) utilizando métodos de programación.
- ✓ Con el componente (Secuencia código) se crea una secuencia de código, generando la clave principal de la dimensión.
- ✓ Con el componente (cargar dimensión), se adicionan o actualizan los datos en la dimensión sexo.

Para la carga del hecho y sus dimensiones se ejecuta un trabajo que carga todas las dimensiones. Para iniciar el proceso se utiliza el componente (Star) el cual inicializa el proceso ETL por la dimensión causa baja y culmina con la transformación del hecho empleado general. Para el manejo de errores en cualquier etapa se encuentra el componente (DUMMY) que informa mediante el componente (Display MsgBox Info) al usuario sobre las acciones realizadas. Ver la siguiente figura para una mejor claridad de la explicación antes mencionada:

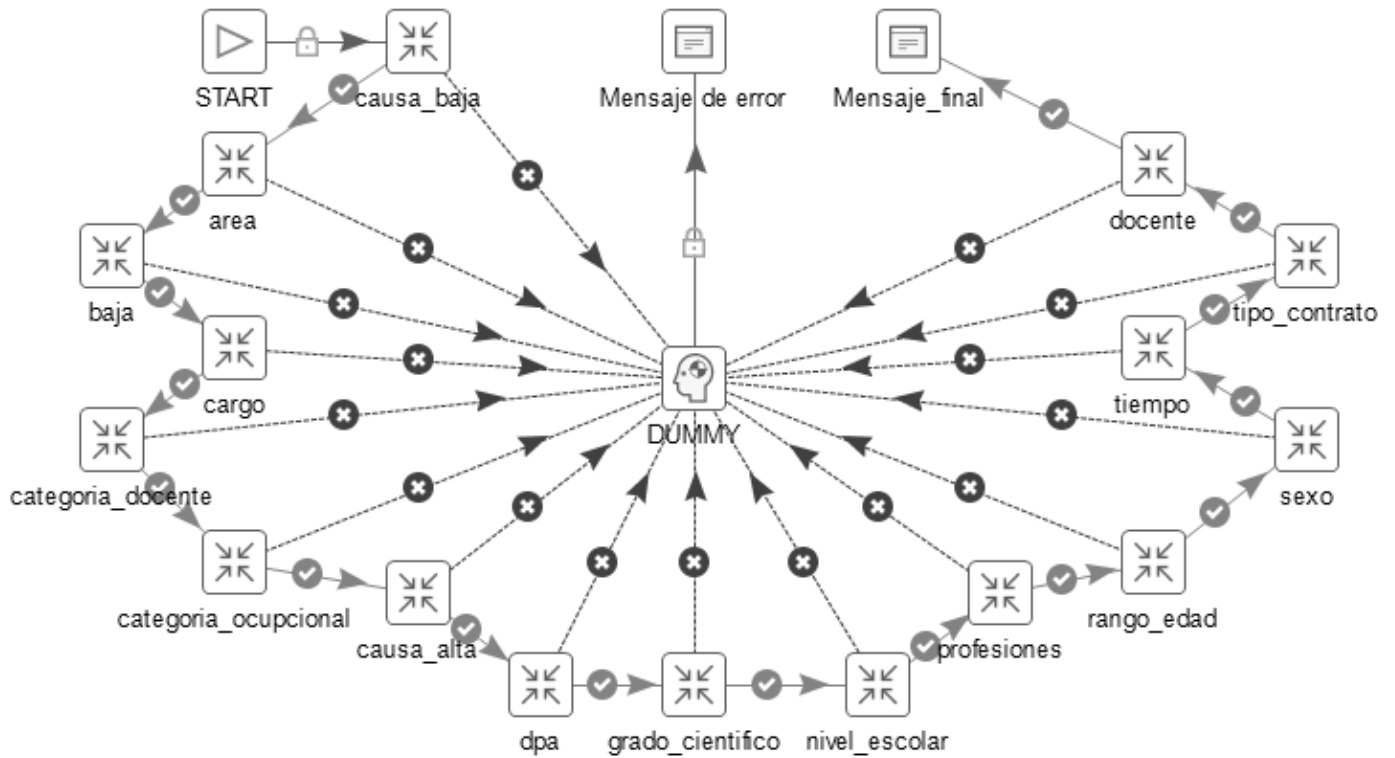


Figura 12. Ejecución del trabajo

Después de ejecutar las dimensiones se realiza otro trabajo que carga al hecho.

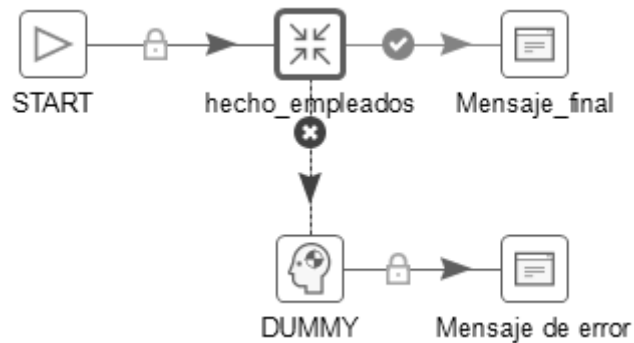


Figura 13. Trabajo que carga el hecho

Por último, se carga un trabajo general que se encarga de llamar a los trabajos anteriores, si se carga satisfactoriamente se cargan todos los datos del DM.



Figura 14. Trabajo carga general

Conclusiones del Capítulo

En el capítulo se desarrolló cada una de las etapas de la metodología Hefesto, obteniéndose las preguntas claves del negocio, identificando así los requerimientos del cliente. Se identificaron los indicadores y las perspectivas, que permitieron elaborar el modelo conceptual, para mostrar el alcance del DM. Se confeccionó el modelo lógico del DM, para realizar el proceso de extracción, transformación y carga de los datos.

Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart

En el presente capítulo se explica la construcción del cubo de datos, donde se definen las jerarquías y atributos de las dimensiones, además de la implementación de los valores numéricos llamados medidas.

3.1 Proceso de análisis de la información del Data Mart

3.1.1 Jerarquías

Una jerarquía representa una relación lógica entre dos o más atributos dentro de una misma dimensión. Posee las siguientes características: (Bernabeu, Ricardo Dario 2007)

- ✓ Pueden existir varias en una misma dimensión.
- ✓ Están compuestas por dos o más niveles.
- ✓ Se tiene una relación “1-n” entre atributos consecutivos de un nivel superior y uno inferior.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa. En la siguiente figura, se muestran las jerarquías de los atributos de algunas de las dimensiones que componen al DM, para ver las jerarquías de las otras dimensiones (Ver Anexos: Jerarquía de los atributos):

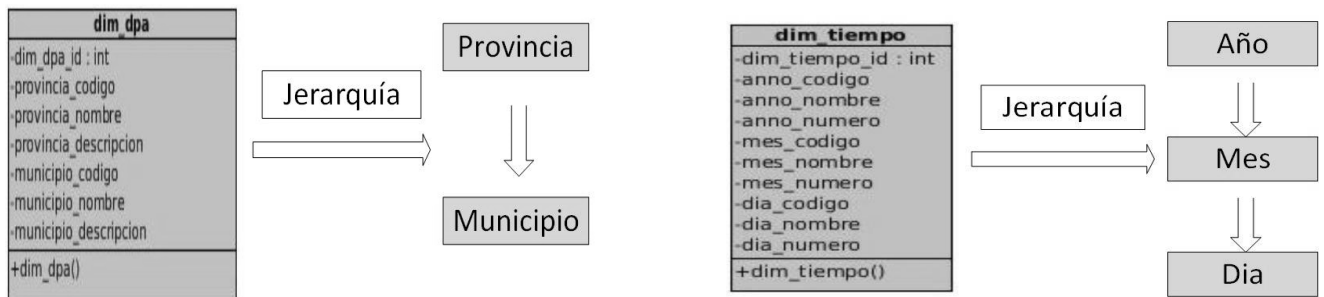


Figura 15. Jerarquía de los atributos

3.1.2 Medidas

Luego de establecer las jerarquías de los atributos de cada una de las dimensiones se elaboran las medidas para dar respuesta a los intereses de los especialistas de la Dirección de Capital Humano. Para ello se utilizan consultas MDX, lenguaje diseñado para realizar consultas multidimensionales sobre los cubos se ejecutan sobre el DM para obtener los indicadores que contribuirán a la toma de

decisiones. Con la herramienta Workbench se diseña el cubo de datos, agregándole las medidas y formulando las consultas para obtener la cantidad de trabajadores de la universidad, en dependencia de la información que sea necesario consultar. (Bustos, Jorge 2016)

A continuación, se muestra la medida cantidad de empleados, que se obtiene mediante la función de suma (sum) sobre el identificador de la tabla del hecho “Empleados”, donde se va sumando cada identificador de los trabajadores y se obtiene el total de empleados.

Measure for 'hech_empleados' Cube	
Attribute	Value
name	cantidad_empleados
description	
aggregator	sum
column	cantidad_empleados
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Numeric
formatter	
caption	cantidad_empleados

Figura 16. Medida cantidad de empleados

3.1.3 Consultas multidimensionales

Las consultas multidimensionales permiten consultar objetos multidimensionales, como el cubo multidimensional y devolver un conjunto de celdas multidimensionales que contengan los datos del cubo. Además, calcularán los datos numéricos y estos serán añadidos en los resultados a visualizar, utilizando la herramienta Mondrian.

Algunos ejemplos de consultas multidimensionales son:

- ✓ En esta consulta se desea conocer la cantidad de trabajadores por categorías ocupacionales en fecha determinada.

```
SELECT NON EMPTY {Hierarchize({[dim_tiempo.fecha].[anno].Members})} ON COLUMNS,
NON EMPTY
CrossJoin([dim_categoria_ocupacional].[categoria_ocupacional_descripcion].Members,
{[dim_baja].[No Baja]}) ON ROWS FROM [hech_empleados]
```

- ✓ En esta consulta se desea conocer la cantidad de trabajadores por categoría ocupacional, sexo, rangos de edad y categoría docente y categoría científica.

```
SELECT NON EMPTY CrossJoin({[dim_baja].[No Baja]},
CrossJoin([dim_rango_edad].[rango_edad_descripcion].Members,
[dim_grado_cientifico].[grado_cientifico_descripcion].Members)) ON COLUMNS, NON EMPTY
Hierarchize(Union(CrossJoin({[dim_categoria_ocupacional].[Ejecutivos]},
```

```
CrossJoin([dim_sexo].[sexo_descripcion].Members,
[dim_categoria_docent_invest].[categoria_docent_invest_descripcion].Members)),
Union(CrossJoin({[dim_categoria_ocupacional].[Tecnicos]},
CrossJoin([dim_sexo].[sexo_descripcion].Members,
[dim_categoria_docent_invest].[categoria_docent_invest_descripcion].Members)),
Union(CrossJoin({[dim_categoria_ocupacional].[Administrativos]},
CrossJoin([dim_sexo].[sexo_descripcion].Members,
[dim_categoria_docent_invest].[categoria_docent_invest_descripcion].Members)),
Union(CrossJoin({[dim_categoria_ocupacional].[Servicios]},
CrossJoin([dim_sexo].[sexo_descripcion].Members,
[dim_categoria_docent_invest].[categoria_docent_invest_descripcion].Members)),
CrossJoin({[dim_categoria_ocupacional].[Operarios]},
CrossJoin([dim_sexo].[sexo_descripcion].Members,
[dim_categoria_docent_invest].[categoria_docent_invest_descripcion].Members)))))) ON
ROWS
FROM [hech_empleados]
```

- ✓ En esta consulta se desea conocer la cantidad de Dirigentes docentes por Sexo, rangos de edad, categoría docente y categoría científica.

```
SELECT NON EMPTY CrossJoin({[dim_baja].[No Baja]}, CrossJoin({[dim_docente].[Docente]},
CrossJoin([dim_rango_edad].[rango_edad_descripcion].Members,
[dim_grado_cientifico].[grado_cientifico_descripcion].Members))) ON COLUMNS, NON
EMPTY CrossJoin({[dim_categoria_ocupacional].[Ejecutivos]},
CrossJoin([dim_sexo].[sexo_descripcion].Members,
[dim_categoria_docent_invest].[categoria_docent_invest_descripcion].Members)) ON ROWS
FROM [hech_empleados]
```

3.1.4 Resultados del cliente STPivot

Cuando se diseña el cubo multidimensional, toda su información se archiva en un archivo XML “lenguaje de marcas extensibles”, que se carga en el Mondrian, que es la herramienta utilizada para realizar el Proceso Analítico en Línea. En la siguiente figura (Figura 17) se muestra el cliente STPivot, disponible a través de la web, cuya página principal muestra un menú con todas las opciones, el cual permite:

- ✓ Filtrar los campos que desee.
- ✓ Añadir o quitar campos.
- ✓ Empleo de filtros de visualización.
- ✓ Observar la información desde diferentes perspectivas.
- ✓ Mostrar lo datos mediante tablas y gráficos.
- ✓ Exportar los datos a formatos PDF, excel, etc.

A continuación, un ejemplo de uno de los reportes que genera el cliente:

Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart

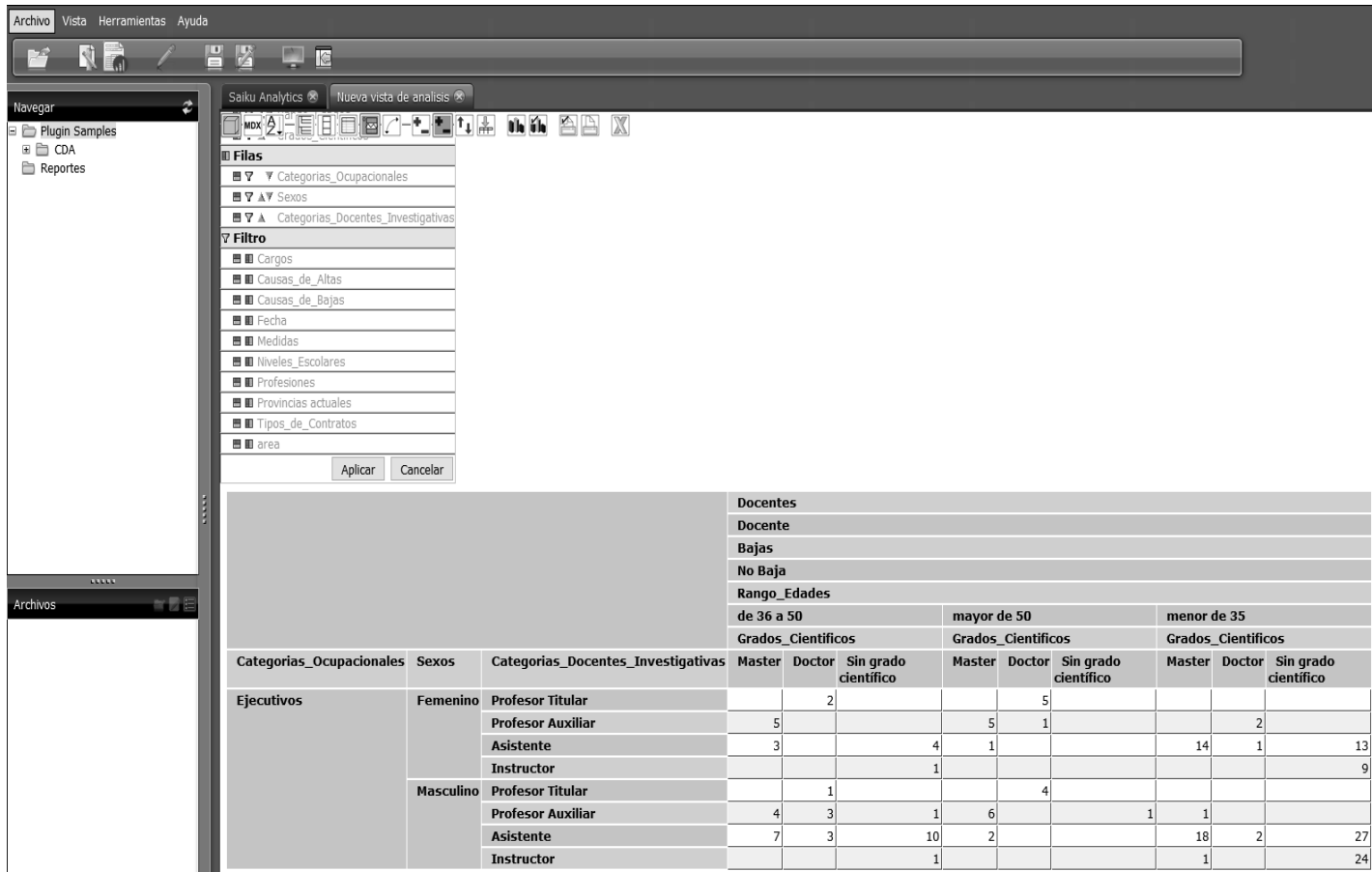
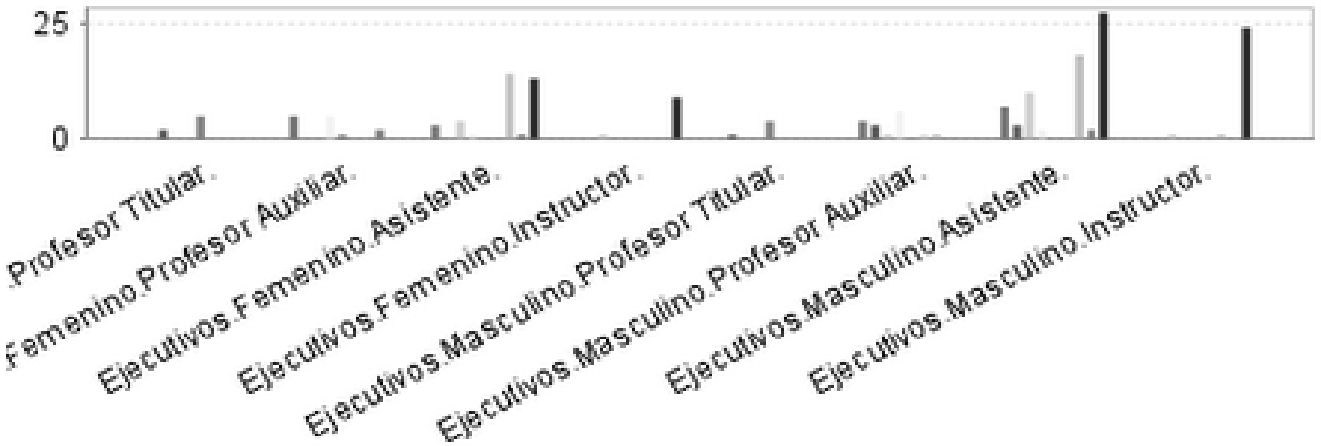


Figura 17. Reporte que muestra la cantidad de dirigentes docentes por sexo, rango edad, categoría docente y categoría científica

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.

Slicer:



Slicer:

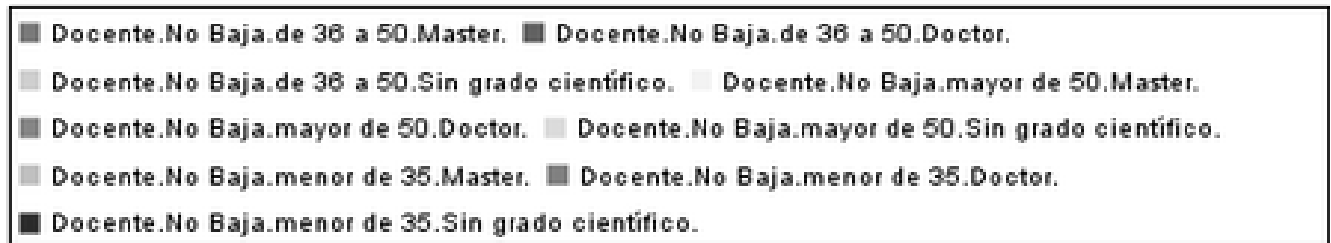


Figura 18. Información en gráfica de barra

También hay que agregar que el cliente STPivot utiliza un plugin llamado Saiku Reporting que se instala en el Pentaho BI Server. Este plugin permite armar consultas y generar gráficos. Saiku Reporting permite dar salida a los informes en varios formatos diferentes, manejando plantillas que agregan marca y temas, altamente personalizables utilizando la misma interfaz web. Está dirigido a la presentación de informes en papel, en lugar de análisis altamente interactivo. (Meteorite.bi 2016)

A continuación, se muestra el mismo reporte con el plugin integrado:

Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart

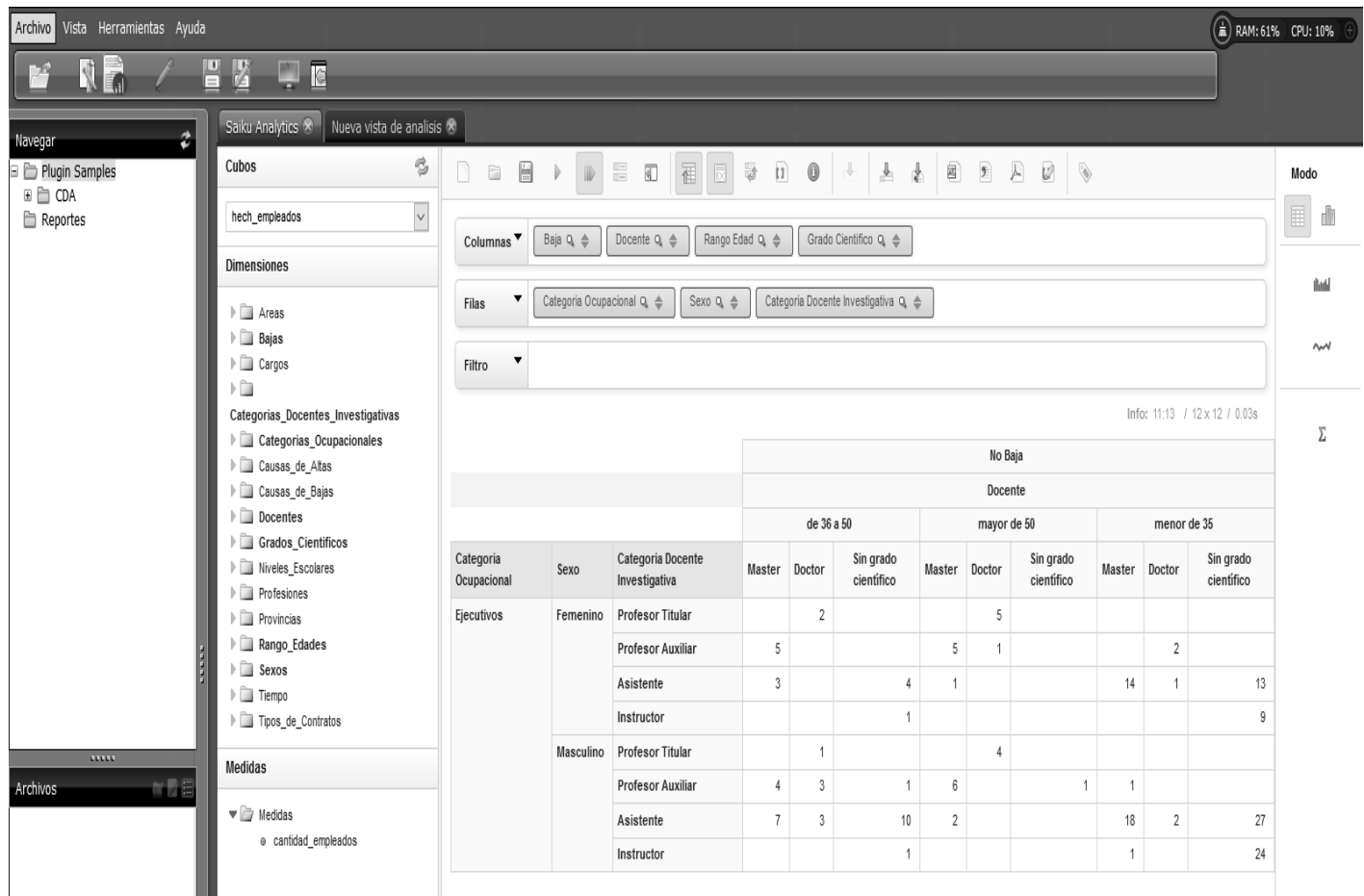


Figura 19. Reporte que muestra la cantidad de dirigentes docentes por sexo, rango edad, categoría docente y categoría científica utilizando el plugin Saiku Reporting

A continuación, se muestra la información en gráficas utilizando el plugin Saiku Reporting:

Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart

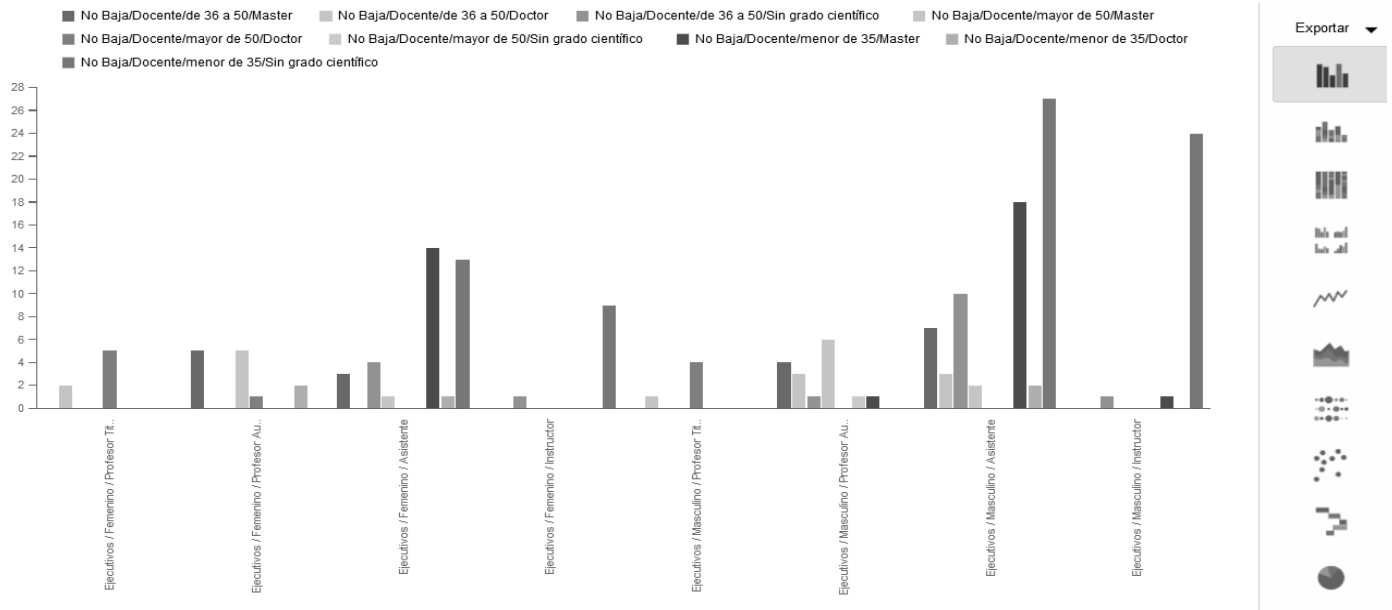


Figura 20. Información en gráfica de barra utilizando el plugin Saiku Reporting

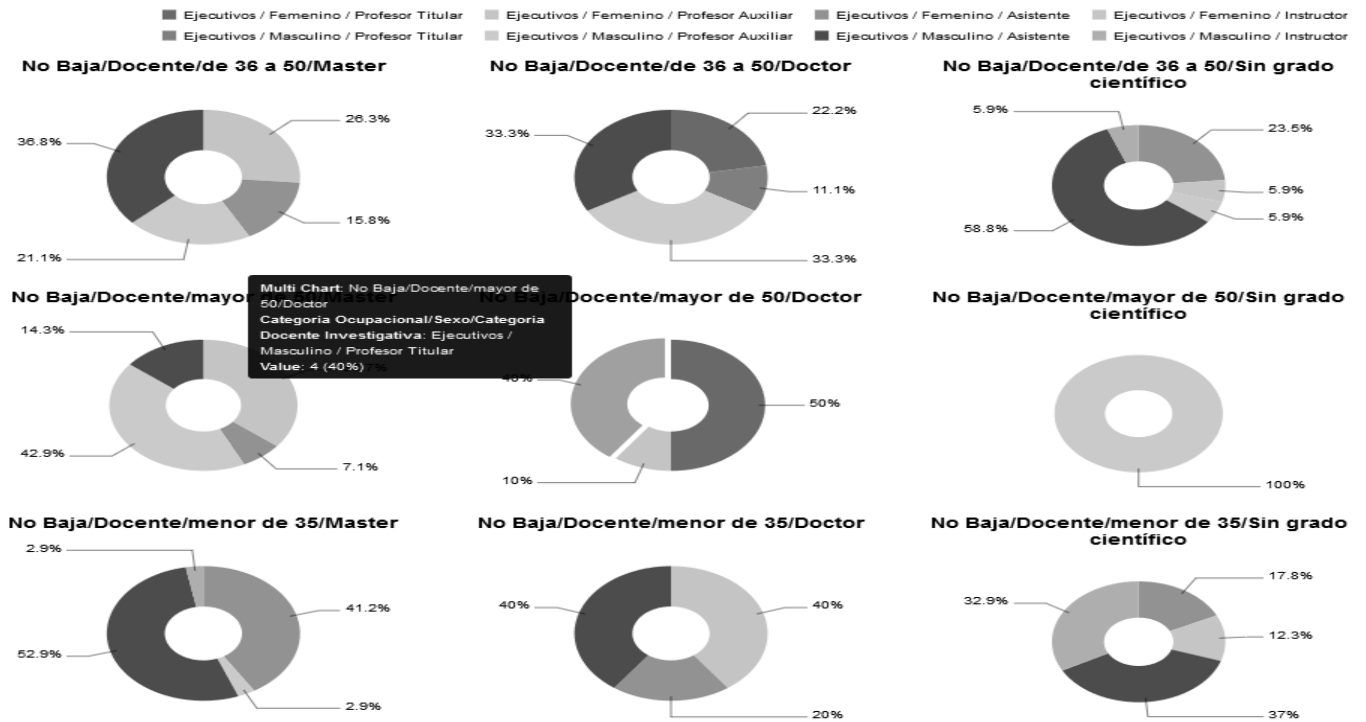


Figura 21. Información gráfica de pastel con el plugin Saiku Reporting

3.2 Gestión del cambio lento en las dimensiones

Las dimensiones de cambio lento o SCD (*Slowly Changing Dimensions*), son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, o implique a un solo registro o la tabla completa. Cuando ocurren estos cambios, se puede optar por seguir alguna de estas dos grandes opciones (Kimball, Ralph y Ross, Margy 2002):

- ✓ Registrar el historial de cambios.
- ✓ Reemplazar los valores que sean necesarios.

Inicialmente Ralph Kimball planteó tres estrategias a seguir cuando se tratan las SCD: tipo 1, tipo 2 y tipo 3; pero a través de los años la comunidad de personas que se encargaba de modelar bases de datos profundizó las definiciones iniciales e incluyó varios tipos SCD más, por ejemplo: tipo 4 y tipo 6.

A continuación, se detallará cada tipo de estrategia SCD:

- ✓ SCD Tipo 1: Sobre escribir.

Este tipo es el más básico y sencillo de implementar, porque si bien no guarda los cambios históricos, tampoco requiere ningún modelado especial y no necesita que se añadan nuevos registros a la tabla. En este caso cuando un registro presente un cambio en alguno de los valores de sus campos, se debe proceder simplemente a actualizar el dato en cuestión, sobre escribiendo el antiguo.

- ✓ SCD Tipo 2: Añadir fila.

Esta estrategia requiere que se agreguen algunas columnas adicionales a la tabla de dimensión, para que almacenen el historial de cambios. Las columnas que suelen agregarse son:

- I. Fecha Inicio: fecha desde que entró en vigencia el registro actual. Por defecto suele utilizarse una fecha muy antigua.
- II. Fecha Fin: fecha en la cual el registro actual dejó de estar en vigencia. Por defecto suele utilizarse una fecha muy futurista.
- III. Versión: número secuencial que se incrementa con cada nuevo cambio. Por defecto suele comenzar en "1".
- IV. Versión actual: especifica si el campo actual es el vigente. Este valor puede ser en caso de ser verdadero: "true" o "1"; y en caso de ser falso: "false" o "0".

Entonces, cuando ocurra algún cambio en los valores de los registros, se añadirá una nueva fila y se crea una nueva entrada en la tabla. Al nuevo registro se le asigna una nueva clave subrogada y a partir de este momento será el valor usado para futuras entradas, las antiguas usarán el valor anterior. En este modo se gestiona un versionado que puede incluir

fechas para indicar los períodos de validez, así como numeradores de registros o indicadores de registros activos o no. Este tipo permite guardar toda la información histórica en el almacén de datos.

✓ SCD Tipo 3: Añadir columna.

Esta estrategia requiere que se agregue a la tabla de dimensión una columna adicional por cada columna cuyos valores se desean mantener en un historial de cambios. De este modo en la nueva columna se coloca el valor antiguo antes de sobrescribir el valor actual con el nuevo. Este tipo presenta como principal desventaja que solo permite guardar un historial limitado de los datos, dependiendo del número de columnas que se creen.

✓ SCD Tipo 4: Tabla de historia separada.

Esta técnica se utiliza en combinación con otra y su función básica es almacenar en una tabla adicional los detalles de cambios históricos realizados en una tabla de dimensión. Esta tabla histórica indicará por ejemplo el tipo de operación se ha realizado (Insert, Update, Delete), sobre qué campo y en qué fecha. El objetivo de mantener esta tabla es el de contar con un detalle de todos los cambios, para luego analizarlos y poder tomar decisiones acerca de cuál técnica SCD podría aplicarse mejor.

✓ SCD Tipo 6: Híbrido.

Esta técnica combina las SCD Tipo 1, 2 y 3. Se denomina SCD Tipo "6", simplemente porque: $6 = 1 + 2 + 3$. Esta estrategia utiliza el Tipo 1 (sobrescribir) junto con el Tipo 2 (añadir filas) y el Tipo 3 (añadir columnas), añadiendo además una pareja adicional de columnas para indicar el rango de fechas al cual aplica cada fila en particular.

De acuerdo a la naturaleza del cambio se debe seleccionar qué tipo SCD se utilizará, en algunos casos resultará conveniente combinar varias técnicas. Para la solución se escogió la opción de registrar los cambios mediante el SCD Tipo 2, porque en caso de ocurrir algún cambio en una de las dimensiones en el futuro se seguirá teniendo acceso a la información histórica.

3.3 Gestión de los metadatos del proceso de integración

Los metadatos son datos que ayudan a identificar, describir y localizar recursos digitales, son información estructurada que describe y/o permite encontrar, gestionar, controlar y entender o preservar otra información; o sea, que no son más que datos sobre los propios datos (Kimball, Ralph y Ross, Margy 2002). Estos se utilizan para describir recursos y no se limitan a un tipo de formato, sino que cubren una amplia gama de recursos, además pueden describir una colección en general,

un recurso en particular o un solo elemento. Los metadatos presentan entre sus funciones básicas proporcionar una descripción de una entidad u objeto de la información a través de otra información necesaria para su manejo y preservación, además de proporcionar puntos de acceso a esa descripción y codificar la misma. Para comprender de manera global los metadatos es necesario clasificarlos, esta clasificación puede realizarse mediante grupos o categorías de acuerdo a los propósitos generales de cada marco de metadatos. A continuación, se presentan algunas de estas clasificaciones (Kimball, Ralph y Ross, Margy 2002):

- ✓ Metadatos administrativos: son utilizados para el manejo y administración de los recursos de información.
- ✓ Metadatos descriptivos y de descubrimiento: utilizados para describir, descubrir o identificar los recursos de información.
- ✓ Metadatos técnicos o modelos: están relacionados con la función de un sistema o el modo en que interrelacionan sus componentes.
- ✓ Metadatos de proceso: permiten obtener información de los procesos en que se ejecutan.
- ✓ Metadatos de negocio: posibilita obtener los datos y la información referente a los aspectos del negocio, como son los datos provenientes de la fuente.

Para la investigación se hizo uso de los metadatos de proceso para obtener la información correspondiente a los procesos de las transformaciones y los trabajos referentes a los subprocesos de ETL. Además, se utilizaron los metadatos de negocio con el fin de obtener los datos que puedan presentar errores en la fuente, posibilitando realizar el proceso de gestión de errores para luego presentarlos al cliente y definir su tratamiento.

3.4 Captura de los cambios en los datos

Para gestionar la captura de los cambios en los datos, en el mercado de dato se debe tener en cuenta como fluye la información desde la entidad informante que en este caso es Capital Humano de la universidad hasta la Oficina Nacional de Estadísticas e Información (ONEI), el Ministerio de Trabajo y Seguridad Social, el Ministerio de Educación Superior y la UCI. Mensualmente, la Dirección de Capital Humano actualiza la información de sus trabajadores, para verificar los datos de cada uno y en caso de cambio actualizar la información. La captura de los cambios se realiza mensual y es gestionado por la tabla de metadatos "md_gestion_cambio_datos_fuente". Esta tabla contiene la fecha del mes que corresponde cargar el hecho; en este caso la fecha es siempre los días 3 de cada mes, debido a que ese día es que realizan el cierre del mes. Una vez realizada esta carga, las fechas de inicio y fin se actualizan con las del mes siguiente al que se realizó la carga, permitiendo actualizar

Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart

el mes en el cual se realizará la próxima carga al mercado de datos. Para que se puedan guardar los cambios durante las cargas se implementó un trigger que es un disparador o acción definida en una tabla de la base de datos y ejecutada automáticamente por una función programada. Esta acción se activará, según se defina, cuando realicemos un INSERT, un UPDATE o un DELETE en la tabla. (PostgreSQL-es 2009). En la aplicación se define que ocurra después de cualquier INSERT o UPDATE. Se implementaron dos triggers de la siguiente manera:

- ✓ CREATE TRIGGER [dbo].[capurar_fecha_insercion]
ON [dbo].[Empleados_Gral]
AFTER INSERT
AS
Declare @fecha_op datetime,
@id_emp varchar(50);
BEGIN
Select @fecha_op = GETDATE();
Select @id_emp = (SELECT Id_Expediente FROM inserted)
INSERT INTO dbo.z_operaciones_datos_almacen(id_empleado,fecha_operacion) values
(@id_emp,@fecha_op)
SET NOCOUNT ON;
END
- ✓ CREATE TRIGGER [dbo].[capurar_fecha_update]
ON [dbo].[Empleados_Gral]
AFTER UPDATE
AS
Declare @fecha_op datetime,
@id_emp varchar(50);
IF (UPDATE(Id_Tipo_Contrato) OR UPDATE(Baja) OR UPDATE(Id_Cargo) OR
UPDATE(Id_Categoria) OR UPDATE(Id_Direccion) OR UPDATE(Id_CausaAlta) OR
UPDATE(Id_CausaBaja) OR UPDATE(Sexo)
OR UPDATE(Id_Provincia) OR UPDATE(Id_Municipio) OR UPDATE(Id_Nivel_Escolaridad)
OR UPDATE(Id_Profesion) OR UPDATE(Docente) OR UPDATE(Id_Categoria_DI) OR
UPDATE(Id_Grado_Cientifico))
Begin
Select @fecha_op = GETDATE();

Capítulo 3: Proceso Analítico en Línea y pruebas del Data Mart

```

Select @id_emp = (SELECT Id_Expediente FROM deleted)
INSERT INTO dbo.z_operaciones_datos_almacen(id_empleado,fecha_operacion) values
(@id_emp,@fecha_op)
SET NOCOUNT ON;
END;

```

Fue necesario crear una tabla en la fuente de datos llamada “z_operaciones_datos_almacen” para guardar el identificador del empleado y la fecha de la operación realizada sobre este. El trigger es una de las estrategias que se utilizó para realizar las cargas incrementales, la estrategia consiste en una invasión a la fuente de datos original.

Para poner a funcionar el DM se hace primeramente una carga histórica de los datos de los trabajadores, para ello se realizó la siguiente transformación:

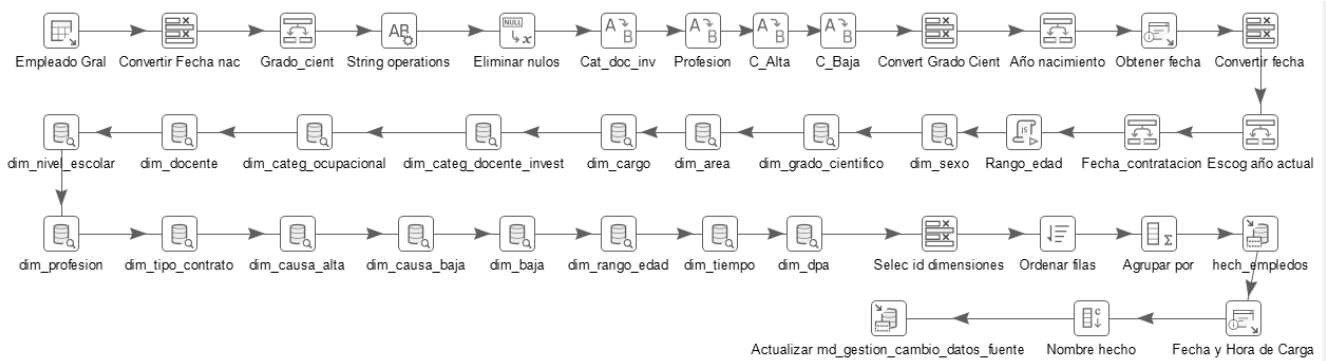


Figura 22. Carga histórica del hecho

Después de la carga histórica, se procede a la carga incremental, la cual se realiza los días 3 de cada mes, donde se verifica si existen cambios en los datos de los trabajadores y se procede a modificar la información. La figura siguiente, muestra la transformación de la carga incremental del hecho:

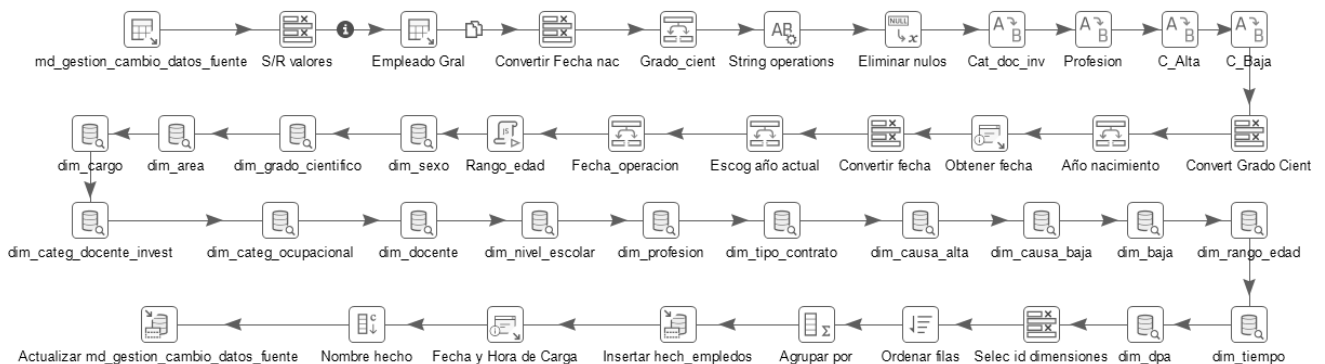


Figura 23. Carga incremental del hecho

3.5 Pruebas y validación

Todo proceso de creación de software está sujeto a fallos, es por esto que las pruebas de software constituyen una fase importante en el desarrollo de cualquier producto, ya que permiten comprobar que no existan fallos en la implementación del mismo, proporcionándole calidad al software. Para realizar las pruebas necesarias en el desarrollo de la solución se decidió utilizar el modelo V el cual es utilizado por DATEC (Centro de Tecnologías de Gestión de Datos), para garantizar la calidad del producto. A continuación, se muestra una representación del ciclo de vida en el modelo V. A la izquierda del mismo se puede detallar las etapas de desarrollo del software y a la derecha de este, las pruebas correspondientes a cada etapa.

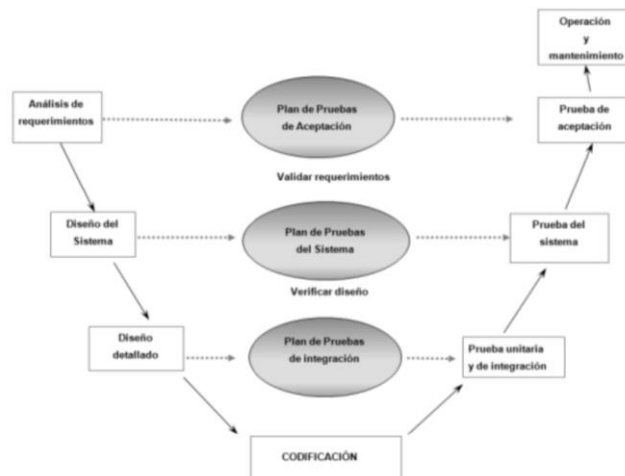


Figura 24. Modelo V

Pruebas a utilizar en el proceso de calidad del software, las cuales surgen a partir del modelo en V:

- ✓ Pruebas unitarias: esta prueba centra el proceso de verificación en la menor unidad del diseño del software, o sea, en algún componente del software o módulo.
- ✓ Pruebas de integración: esta prueba construye el sistema a partir de distintos componentes y lo prueba una vez estén todos estos componentes integrados, debe realizarse progresivamente.
- ✓ Pruebas de aceptación: estas pruebas se realizan para comprobar que el sistema cumple con las necesidades del cliente, y pueden ser de dos tipos:
 - Pruebas alfa: las realiza el usuario en presencia del equipo de desarrollo del proyecto utilizando para esto una máquina preparada con este fin.
 - Pruebas beta: son realizadas por el usuario una vez que el equipo de desarrollo le haya entregado una versión casi definitiva del software.

- ✓ Pruebas de regresión: estas pruebas consisten en volver a ejecutar un conjunto de pruebas ya ejecutadas anteriormente, de este modo se asegura que los cambios realizados no conducen a errores adicionales.

Del modelo en V se escogieron las pruebas de integración y de aceptación. Finalmente se realiza la prueba de aceptación con el cliente para comprobar que la aplicación cumple con los requerimientos de información planteados en el capítulo anterior; para ello se realizó un acta de aceptación junto con el cliente (Ver Anexos: Acta de aceptación). Otras pruebas que se le realizaron al DM para validar que cumplía con los requerimientos del cliente fueron con la herramienta BI Server y el plugin Saiku Reporting. Esta herramienta es una interfaz gráfica que ya tiene el Mondrian y se le instala el plugin, entonces carga el DM que se implementó y se comienza a probar, si la aplicación no crea los reportes requeridos significa que el DM está mal implementado y contiene errores internos. Cuando se utilizó la herramienta se comprobó que el DM funciona correctamente, y brinda los reportes necesarios para el cliente; puesto que el cliente probó la herramienta y comparó con la que utiliza actualmente (ASSETS ULTIMATE) arrojando resultados satisfactorios.

3.6 Calidad de datos

El proceso de calidad de datos es importante para el desarrollo del mercado de datos, ya que permite comprobar que los datos cargados no presentan errores. Una vez culminado el proceso de integración y carga de los datos, se realizó el perfilado de los datos que fueron cargados al mercado, con el objetivo que estos tuviesen la calidad requerida. Este proceso de perfilado permite obtener estadísticas e información sobre los datos, lo cual posibilita corregir problemas que pudiesen existir, como son valores escritos incorrectamente, duplicados o nulos. Mediante esta prueba interna al mercado de datos, detectaron algunas inconsistencias en la base de datos de Capital Humano, las inconsistencias fueron las siguientes:

- ✓ En la tabla “Empleado General” existían 13 empleados que tenían el identificador de la profesión que ejercían; pero el identificador que tenían no existía en la tabla “Profesiones”.
- ✓ El campo “Categoría docente Investigativa” en algunas tablas aparecía sin identificador y en otras aparecía con el identificador en valor “0” o “10”.
- ✓ Existían áreas que no tenían el valor que las identificara.

En este caso se le presentó estos errores al cliente para definir el tratamiento que se les debería dar. En caso de que sean datos de importancia, se define una nueva transformación para realizar la carga de estos al mercado de datos. Por último, se ejecuta la transformación y se verifica que los datos

fueron cargados correctamente al mercado. Las inconsistencias fueron resueltas satisfactoriamente, de la forma siguiente:

- ✓ A los 13 empleados se les puso que no tenían identificador, puesto que esa profesión se eliminó en el pasado antes de que los especialistas comenzaran a trabajar en Capital Humano.
- ✓ A los identificadores de “Categoría docente Investigativa” donde no tenían valores se le puso el valor “0”.
- ✓ A las áreas que no tenían un valor que las identificara se les puso el valor “P”, de Producción; debido a que esas áreas actualmente pertenecen a la producción.

3.7 Auditoría de datos

Para que la información sea confiable, se aplica auditoría a los datos cargados al mercado. Auditando los datos se puede conocer el nombre de las transformaciones ejecutadas, así como su estado, el número total de elementos en la entrada y en la salida, así como la cantidad de errores producidos durante la transformación, también es posible conocer la fecha en la cual se realizaron las transformaciones, entre otros datos. Cuando se aplican auditorías se emplean varios instrumentos y técnicas, entre los más utilizados se encuentran (Muñoz Razo, Carlos 2002):

- ✓ Cuestionarios.
- ✓ Entrevistas.
- ✓ Inventarios.
- ✓ Guías de evaluación.
- ✓ Listas de verificación o chequeo.
- ✓ Confirmación.
- ✓ Comparación.

Conclusiones del Capítulo

Con las herramientas propuestas en el primer capítulo se confeccionó el cubo de datos multidimensional, que consiste en la jerarquización de los atributos e implementación de los indicadores. Con el proceso OLAP los especialistas de Capital Humano podrán mejorar la toma de decisiones referente a la información de los trabajadores de la universidad, mediante los informes obtenidos en forma de tabla y gráficos.

Conclusiones

Con el desarrollo del presente trabajo se logró cumplir con el objetivo general propuesto, se obtuvo como producto de software un DM para el análisis de la información de la gestión de los Recursos Humanos de la UCI. El DM permite almacenar grandes volúmenes de información y facilita el proceso de toma de decisiones. Los siguientes resultados demuestran el cumplimiento de los objetivos propuestos en la investigación:

- ✓ Se seleccionó la metodología, herramientas y tecnologías a utilizar en el desarrollo del mercado de datos, lo cual permitió guiar el proceso de su construcción.
- ✓ Se llevó a cabo el análisis de los principales requerimientos de información de la Dirección de Desarrollo del Capital Humano, a través de preguntas que facilitaron la comprensión de los procesos del negocio, para la implementación del DM.
- ✓ Se diseñó e implementó un DM que integra diferentes técnicas de análisis de datos, las cuales fueron aplicadas sobre la información almacenada en los Recursos Humanos de la UCI, agilizando el proceso de obtención de informes estadísticos, que apoyan la toma de decisiones.
- ✓ Mediante la construcción del DM los especialistas de Recursos Humanos pueden obtener la información en un único sistema, a partir de reportes y gráficas de datos, además de exportar los informes a formato excel y PDF.

Recomendaciones

Después de haber apreciado los resultados obtenidos y basándose en la experiencia adquirida durante la realización de la investigación y con el propósito de mejorar la propuesta plasmada en este trabajo se recomienda:

- ❖ Realizar pruebas de rendimiento mensualmente, una vez cargados los datos, con el objetivo de comprobar que las cargas mensuales y el aumento de la cantidad de tuplas que se almacenan en el mercado de datos, no afectan su rendimiento.
- ❖ Utilizar la presente investigación como base para las capacitaciones a los especialistas de Recursos Humanos, con el objetivo de lograr una transferencia tecnológica y que los especialistas sepan interactuar con el sistema, para que de este modo aprovechen las ventajas que tiene el uso del almacén de datos en el apoyo a la toma de decisiones.
- ❖ Después que el sistema lleve 2 ó 3 años de explotación, aplicar nuevas técnicas de Inteligencia de negocios como la minería de datos.

Referencias Bibliográficas

Espinosa , Roberto. 2010. 17.3. Preparando el analisis dimensional. Definición de cubos utilizando Schema Workbench. *17.3. Preparando el analisis dimensional. Definición de cubos utilizando Schema Workbench.* [En línea] 4 de julio de 2010. [Citado el: 15 de febrero de 2016.] <https://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>.

Febles, Dr. Juan Pedro, y otros. 2015. *Importancia de la utilización de un Data Warehouse (DW) en las.* 2015.

López Paz, Carlos Ramón. 2006. *APLICACIÓN DEL CUADRO DE MANDO INTEGRAL Y EL DATA WAREHOUSE A LA GESTIÓN DE LOS RECURSOS HUMANOS.* Habana : s.n., 2006.

Muñoz Razo, Carlos. *Auditoría en sistemas computacionales.*

Alfárez Sánchez, José A . 2010. *Instalación, Configuración y Administración del Servidor de Aplicaciones JBOSS.* 2010.

Alvarez Trujillo , Cealys. 2012. *Tesis de Maestria: MODELO PARA LA GESTION INTEGRADA DEL CAPITAL HUMANO EN EL PROYECTO ERP CUBA.* Habana, Cuba. Universidad de las Ciencias Informaticas : s.n., 2012.

Aplicación de la minería de datos en la bioinformática. **Febles Rodriguez, Juan Pedro y Gonzalez Perez, Abel. 2010.** Ciudad de la Habana : s.n., 2010, Vol. 10. ISSN 1024-9435.

ASSETS. 2014. ASSETS Sistema de Gestion Integral. [En línea] Infomaster, 2014. [Citado el: 20 de noviembre de 2015.] <http://www.assets.co.cu>.

Barchini, Graciela Elisa. 2004. *Métodos "I + D" de la Informática.* Santiago del Estero, Argentina. : s.n., 2004.

Bernabeu, Ricardo Dario. 2010. Data Prix. [En línea] julio de 2010. <http://www.dataprix.com>.

BI. 2012. STPivot, Free and Open Source OLAP viewer. *STPivot, Free and Open Source OLAP viewer.* [En línea] 29 de octubre de 2012. [Citado el: 15 de febrero de 2016.]

Cañibano Sánchez, D.a Carolina . 2016. *El capital humano: factor de innovación, competitividad y crecimiento.* 2016.

DATA WAREHOUSING: Investigación y Sistematización de Conceptos-HEFESTO: Metodología propia para la Construcción de un Data Warehouse. **Bernabeu, Ricardo Dario. Miércoles 07 de Noviembre de 2007.** Cordoba, Argentina : s.n., Miércoles 07 de Noviembre de 2007.

Datamart para la Unidad Central de Cooperación Médica. **Hidalgo López, Leydis. 2013.** 2, Habana : s.n., 2013, Vol. 6. ISSN: 2306-2495 | RNPS: 2343 .

ETL-Tools.Info. Business Intelligence - Almacenes de Datos - ETL. [En línea]

Expresiones MDX en Analysis Services. **Bustos, Jorge. 2016.** 2016.

foundation, The apache software. 2016. Apache Tomcat. *Apache Tomcat*. [En línea] 2016. [Citado el: 15 de febrero de 2016.] <http://tomcat.apache.org/>.

Garcia Martinez, Ramon. 2002. *Algoritmos TDIDT aplicados a la Minería de Datos Inteligentes*. Buenos Aires : Servente M, 2002.

Gorbe, Tomás Guillén. 2016. Servicios TIC Grupo IFEDES S.A. [En línea] 2016. [Citado el: 16 de marzo de 2016.] <http://www.serviciostic.com/las-tic/definicion-de-tic.html>.

Hernandez Orallo, Jose, Ramirez Quintana, Jose Manuel y Ferri Ramirez, Cesar. 2004. *INTRODUCCION A LA MINERIA DE DATOS*. 2004.

Humble, Charles y Hirsch , Jai . 2011. Olap4j 1.0: a Java API for OLAP Servers. *Olap4j 1.0: a Java API for OLAP Servers*. [En línea] 24 de junio de 2011. [Citado el: 15 de febrero de 2016.] <http://www.infoq.com/news/2011/06/olap4j>.

IBM. 2016. Rational Rose Enterprise. *Rational Rose Enterprise*. [En línea] 2016. [Citado el: 15 de febrero de 2015.] <http://www-03.ibm.com/software/products/es/enterprise>.

Jedox AG. 2016. Jedox. [En línea] 2016. [Citado el: 15 de febrero de 2016.] www.jedox.com.

Kimball/Imon. 5 de mayo del 2012. *Enfoques de desarrollo DW*. 5 de mayo del 2012.

Lazaro, Rigoberto. 2012. Data Warehouse. [En línea] MARTES, 18 de SEPTIEMBRE de 2012.

Martinez, Dolores. 2010. *agilizará la búsqueda de información de una manera eficaz y rápida, lo que permitirá, la toma de decisiones, para la selección del talento humano que ocupe determinado cargo que necesite la Institución*. Caracas, Venezuela : s.n., 2010.

Meteorite.bi. 2016. Saiku Reporting. [En línea] 2016. <http://www.meteorite.bi/products/saiku-reporting>.

Pentaho a Hitachi Group Company. 2016. Mondrian. *Mondrian*. [En línea] 2016. [Citado el: 15 de febrero de 2016.] <http://community.pentaho.com/projects/mondrian/>.

Pentaho BI. 2011. El servidor OLAP Mondrian. *El servidor OLAP Mondrian*. [En línea] 2011. [Citado el: 15 de febrero de 2016.] <http://pentaho.almacen-datos.com/mondrian.html>.

PostgreSQL-es. 2009. PostgreSQL. [En línea] 11 de junio de 2009. [Citado el: 18 de mayo de 2016.] <http://www.postgresql.org.es/node/301>.

Pressman, Roger S. 2002. . *Ingeniería de Software, un enfoque práctico. Quinta edición*. s.l. : McGraw-Hill Companies, 2002. . ISBN: 8448132149..

SPARX System . 2015. Enterprise Architect. *Enterprise Architect*. [En línea] 03 de julio de 2015. [Citado el: 15 de febrero de 2016.] <http://www.sparxsystems.com.ar/products/ea/index.html>.

The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling. **Kimball, Ralph y Ross, Margy.** s.l. : John Wiley and Sons, Inc. ISBN 0-471-20024-7.

UCI. 2014. *Dirección de Desarrollo del Capital Humano. Mapa de Procesos.* Habana : s.n., 2014.

—. **2015.** Portal de la Universidad de las Ciencias Informáticas. [En línea] 2015. [Citado el: 7 de marzo de 2016.] <http://www.uci.cu/sites/default/files/Cat%C3%A1logo%202015.pdf>.

Yupa, Elsi Elizabeth Ilbay. octubre de 2009. *Propuesta metodológica para aplicar business intelligence caso practico "COHERVI S.A".* octubre de 2009.

Bibliografía

1. ALVAREZ TRUJILLO, CEALYS, 2012. Tesis de Maestría: MODELO PARA LA GESTIÓN INTEGRADA DEL CAPITAL HUMANO EN EL PROYECTO ERP CUBA. Universidad de las Ciencias Informáticas, La Habana, Cuba.:
2. BARCHINI, GRACIELA ELISA, 2004. Métodos «I + D» de la Informática. Santiago del Estero, Argentina.:
3. BERNABEU, RICARDO DARÍO, 2007. DATA WAREHOUSING: Investigación y Sistematización de Conceptos-HEFESTO: Metodología propia para la Construcción de un Data Warehouse. ,
4. BI, 2012. STPivot, Free and Open Source OLAP viewer. *STPivot, Free and Open Source OLAP viewer ~ Todo BI* [en línea]. [Consulta: 25 mayo 2016]. Disponible en: <http://todobi.blogspot.com/2012/10/stpivot-free-and-open-source-olap-viewer.html>.
5. BI, 2015. MOLAP, ROLAP, HOLAP. [en línea]. [Consulta: 26 mayo 2016]. Disponible en: http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
6. BUSTOS, JORGE, 2016. Expresiones MDX en Analysis Services.
7. CAÑIBANO SÁNCHEZ, D.A CAROLINA, 2016. El capital humano: factor de innovación, competitividad y crecimiento. S.I.:
8. CloverETL Rapid Data Integration. [en línea], 2008. [Consulta: 26 mayo 2016]. Disponible en: <http://www.cloveretl.com/>.
9. DATA PRIX, 2010. Dataprix TI | El portal sobre Software empresarial. Eventos IT, información, foro y directorio. [en línea]. [Consulta: 25 mayo 2016]. Disponible en: <http://www.dataprix.com/>.
10. ESPINOSA, ROBERTO, 2010. 17.3. Preparando el análisis dimensional. Definición de cubos utilizando Schema Workbench. *El Rincón del BI* [en línea]. [Consulta: 25 mayo 2016]. Disponible en: <https://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>.
11. FEBLES RODRIGUEZ, JUAN PEDRO y GONZALEZ PEREZ, ABEL, 2010. Aplicación de la minería de datos en la bioinformática, vol. 10. ISSN 1024-9435.
12. GARCIA MARTINEZ, RAMON, 2002. Algoritmos TDIDT aplicados a la Minería de Datos Inteligentes. Buenos Aires: Servente M.
13. GORBE, TOMÁS GUILLÉN, 2016. Definición de TIC. *Servicios TIC Grupo IFEDES S.A* [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://www.serviciostic.com/las-tic/definicion-de-tic.html>.
14. HIDALGO LÓPEZ, LEYDIS, 2013. Datamart para la Unidad Central de Cooperación Médica. vol. 6, no. 2. ISSN 2306-2495 | RNPS: 2343.

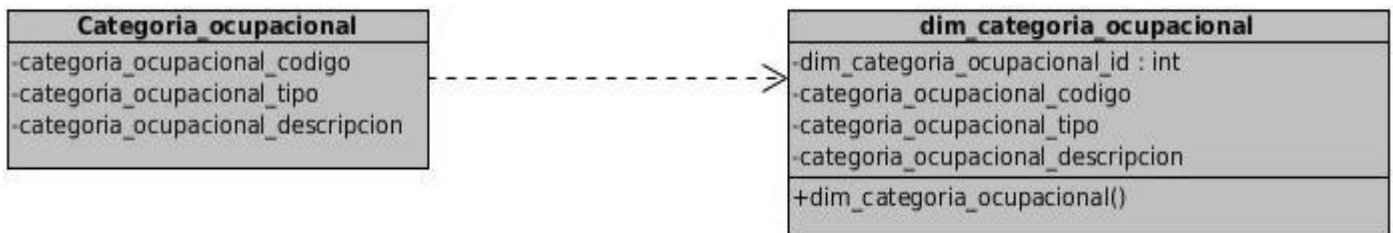
15. HINZ, 2011. MySQL. S.I.:
16. IBM, 2016. IBM - Rational Rose Enterprise. *Rational Rose Enterprise* [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://www-03.ibm.com/software/products/es/enterprise>.
17. INFOMASTER, 2014. ASSETS: Sistema de Gestión Integral. [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://www.assets.co.cu/>.
18. Jedox - Business Driven Intelligence. *Jedox* [en línea], 2016. [Consulta: 26 mayo 2016]. Disponible en: <http://www.jedox.com/en/>.
19. KIMBALL/IMON, 2012. Enfoques de desarrollo DW. S.I.:
20. KIMBALL, RALPH y ROSS, MARGY, 2002. The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling. *John Wiley and Sons, Inc*, ISSN ISBN 0-471-20024-7.
21. LAZARO, P. por R., 2012. Data Warehouse. [en línea]. [Consulta: 25 mayo 2016]. Disponible en: <http://rigobertlc.blogspot.com/>.
22. METEORITE.BI, 2016. Saiku Reporting | meteorite.bi. *Saiku Reporting* [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://www.meteorite.bi/products/saiku-reporting>.
23. MUÑOZ RAZO, CARLOS, [sin fecha]. *Auditoría en sistemas computacionales*. S.I.: s.n.
24. Olap4j 1.0: a Java API for OLAP Servers. *InfoQ* [en línea], 2011. [Consulta: 25 mayo 2016]. Disponible en: <https://www.infoq.com/news/2011/06/olap4j>.
25. PENTAHO A HITACHI GROUP COMPANY, 2016. Mondrian | Pentaho Community. *Mondrian* [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://community.pentaho.com/projects/mondrian/>.
26. PENTAHO BI, 2016. Mondrian - El servidor OLAP Open Source. *El servidor OLAP Mondrian* [en línea]. [Consulta: 25 mayo 2016]. Disponible en: <http://pentaho.almacendatos.com/mondrian.html>.
27. PENTAHO SOLUTIONS, 2015. Pentaho Data Integration - Pentaho BI Suite. [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <https://sites.google.com/site/pentahobisuite/home/caracteristicas/pentaho-dashboards/pentaho-data-integration>.
28. Portada | Portal de la Universidad de las Ciencias Informáticas. [en línea], 2015. [Consulta: 26 mayo 2016]. Disponible en: <http://www.uci.cu/sites/default/files/>.
29. POSTGRESQL-ES, 2009. Disparadores (triggers) en PostgreSQL | www.postgresql.org.es. [en línea]. [Consulta: 25 mayo 2016]. Disponible en: <http://www.postgresql.org.es/node/301>.
30. POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2013. PostgreSQL: The world's most advanced open source database. [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <https://www.postgresql.org/>.

31. PRESSMAN, ROGER S., 2002. *Ingeniería de Software, un enfoque práctico. Quinta edición.* Quinta. S.I.: McGraw-Hill Companies. ISBN 84-481-3214-9.
32. SOFTPEDIA, 2012. Enhydra Octopus Download. *softpedia* [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://www.softpedia.com/get/Internet/Servers/Database-Utils/Enhydra-Octopus.shtml>.
33. SPARX SYSTEM, 2015. Enterprise Architect - Herramienta CASE para diseño con UML y desarrollo de software. *Enterprise Architect* [en línea]. [Consulta: 26 mayo 2016]. Disponible en: <http://www.sparxsystems.com.ar/products/ea/index.html>.
34. UCI, 2014. Dirección de Desarrollo del Capital Humano. Mapa de Procesos. La Habana: Universidad de las Ciencias Informáticas.
35. VELASCO, 2011. Oracle Database. S.I.:
36. YUPA, ELSI ELIZABETH ILBAY, 2009. Propuesta metodológica para aplicar business intelligence caso práctico «COHERVI S.A». S.I.:
37. Magdalena Serventes, 2002. ALGORITMOS TDIDT APLICADOS A LA MINERÍA DE DATOS INTELIGENTE.
38. Miguel Rodríguez Sanz, 2010. ANÁLISIS Y DISEÑO DE UN DATA MART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO.
39. MSc. Emma R. Rizo Rizo et. al, 2015. Importancia de la utilización de un Data Warehouse (DW) en las empresas.
40. Gustavo R. Rivadera, 2010. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)
41. Ricardo Chinchilla Arley, 2010. Mercado de datos: conceptos y metodologías de desarrollo.
42. Alexeis Joel Ochoa Reyes, et. al, 2014. System for Processing and Analysis of Information Using Clustering Technique

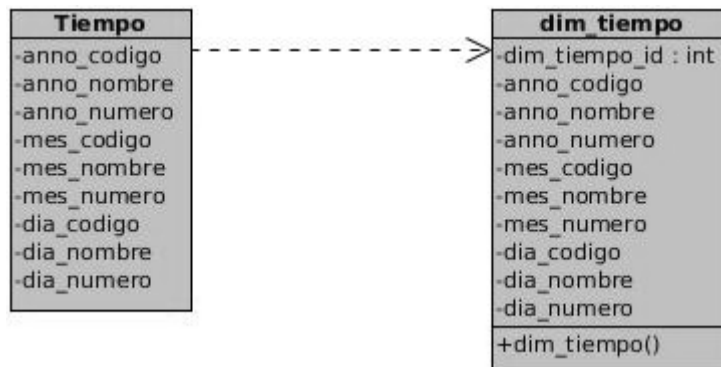
Anexos

1. Tablas de dimensiones

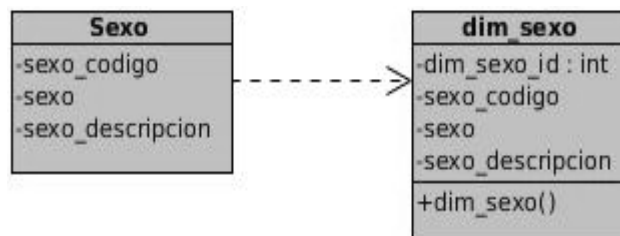
1.1 Dimensión Categoría ocupacional



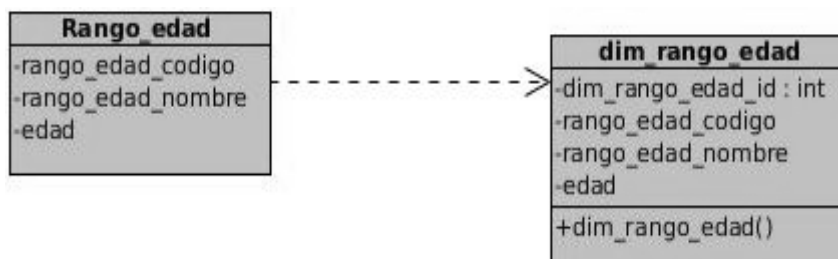
1.2 Dimensión tiempo



1.3 Dimensión sexo



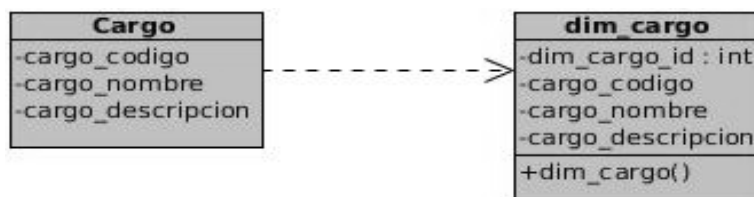
1.4 Dimensión Rango edad



1.5 Dimensión tipo contrato



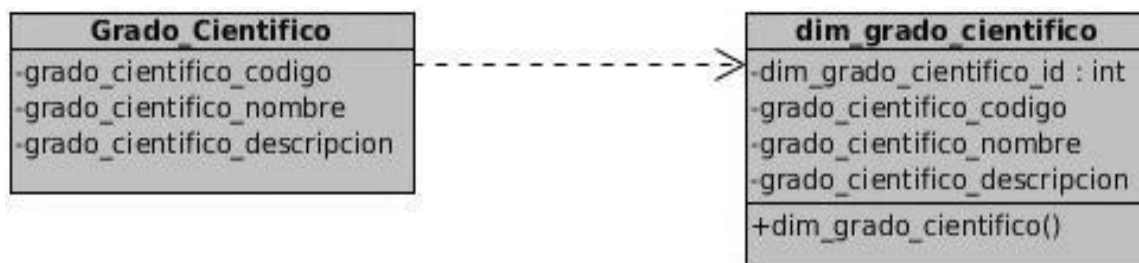
1.6 Dimensión cargo



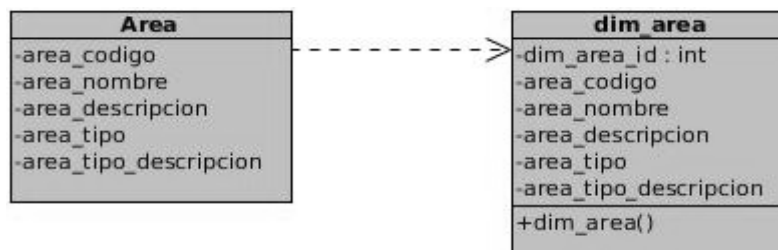
1.7 Dimensión categoría docente investigativa



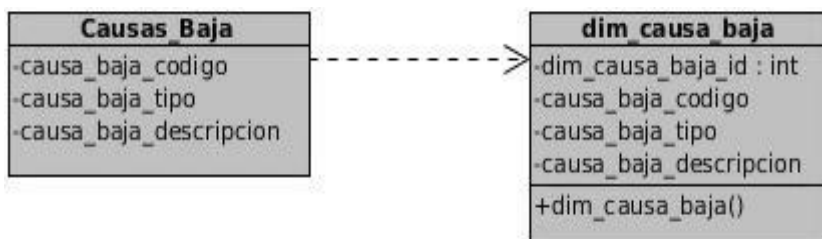
1.8 Dimensión grado científico



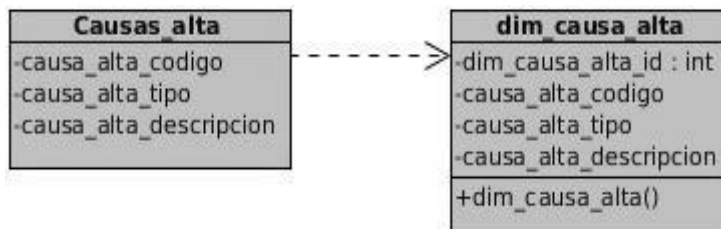
1.9 Dimensión área



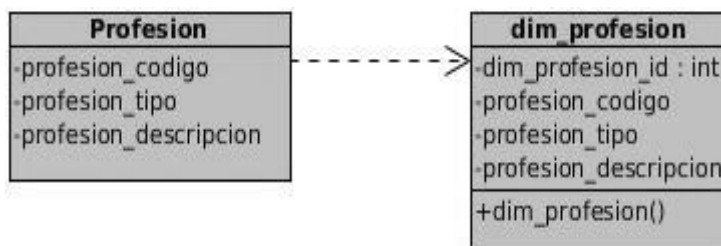
1.10 Dimensión causa baja



1.11 Dimensión causa alta



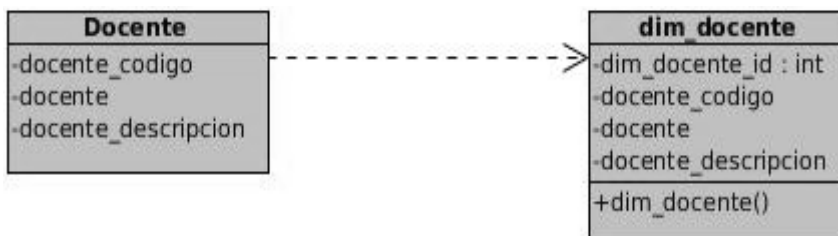
1.12 Dimensión profesión



1.13 Dimensión nivel escolar



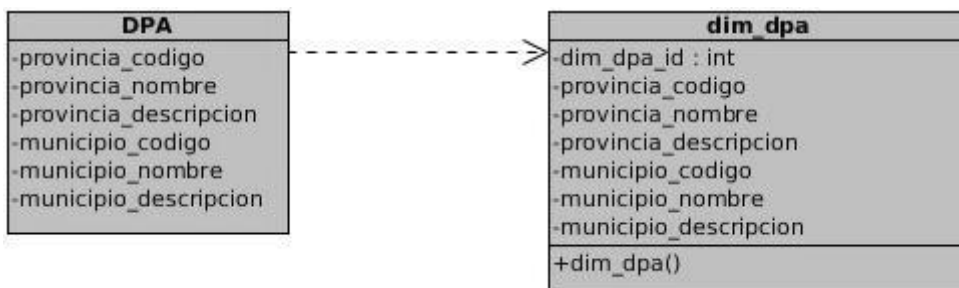
1.14 Dimensión docente



1.15 Dimensión baja



1.16 Dimensión DPA (División Político Administrativa)

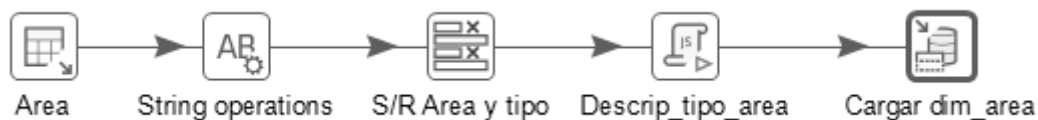


2. Tabla del hecho empleado



3. Transformaciones del Data Mart

3.1 Transformación área



3.2 Transformación baja



3.3 Transformación cargo



3.4 Transformación categoría docente investigativa



3.5 Transformación categoría ocupacional



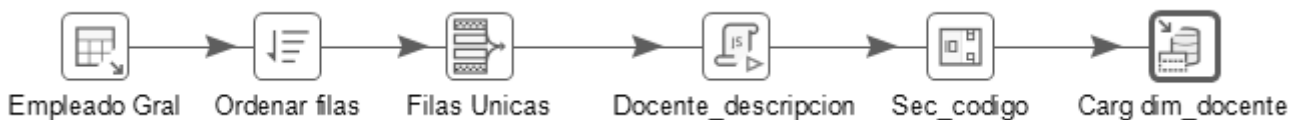
3.6 Transformación causa alta



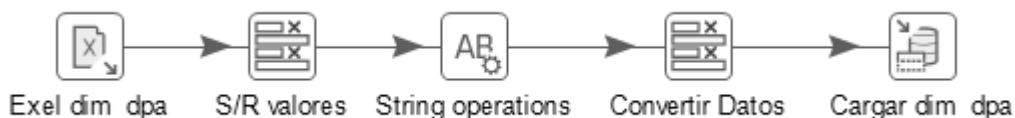
3.7 Transformación causa baja



3.8 Transformación docente



3.9 Transformación DPA (Distribución Político Administrativa)



3.10 Transformación grado científico



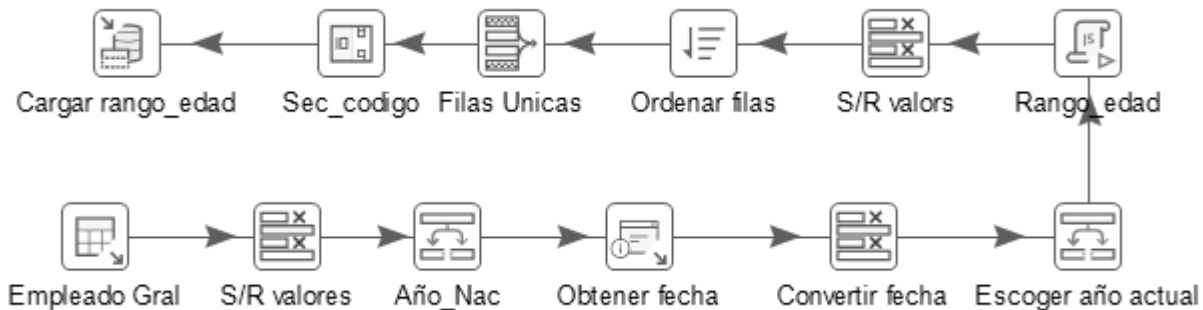
3.11 Transformación nivel escolar



3.12 Transformación profesión



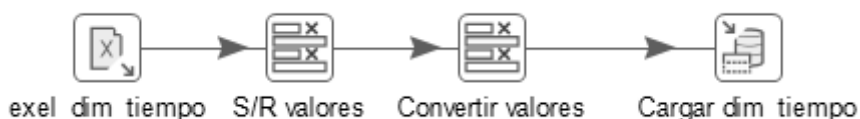
3.13 Transformación rango edad



3.14 Transformación sexo



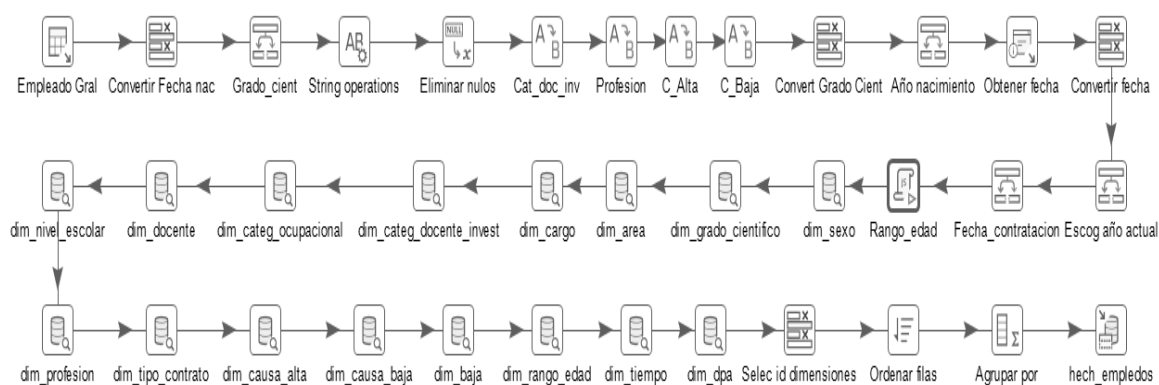
3.15 Transformación tiempo



3.16 Transformación tipo contrato

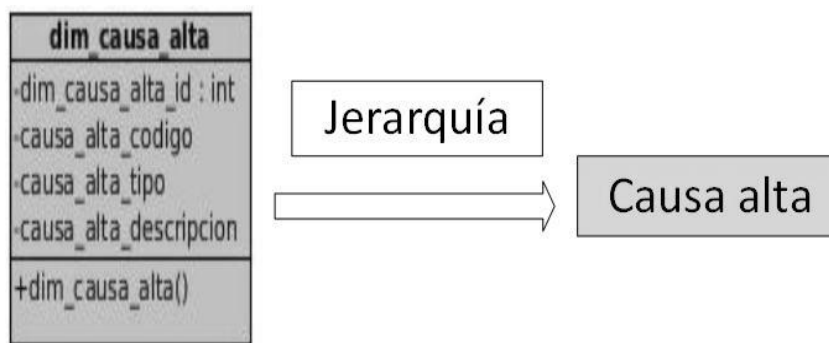


3.17 Transformación del hecho Empleado General



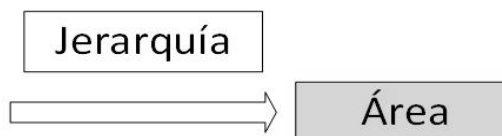
4. Jerarquía de los atributos

4.1 Causa Alta



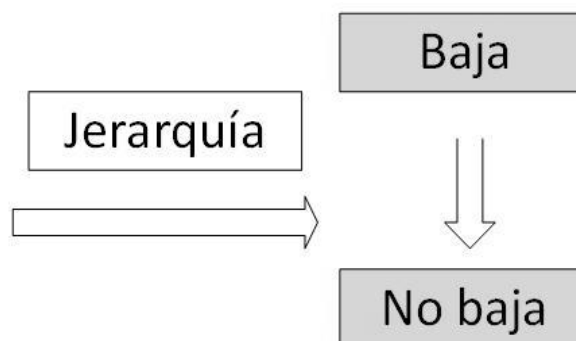
4.2 Área

dim_area
-dim_area_id : int
-area_codigo
-area_nombre
-area_descripcion
-area_tipo
-area_tipo_descripcion
+dim_area()



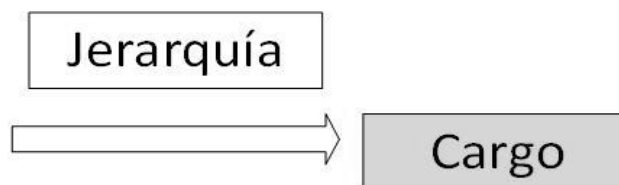
4.3 Baja

dim_baja
-dim_baja_id : int
-baja_codigo
-baja_estado
-baja_descripcion
+dim_baja()

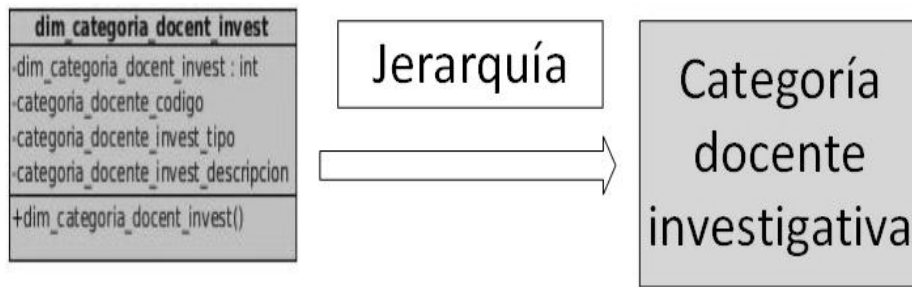


4.4 Cargo

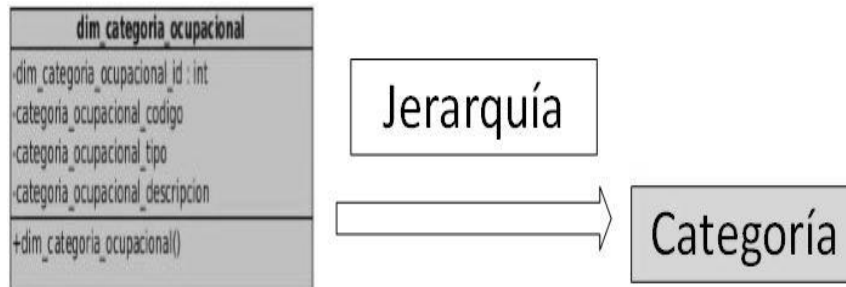
dim_cargo
-dim_cargo_id : int
-cargo_codigo
-cargo_nombre
-cargo_descripcion
+dim_cargo()



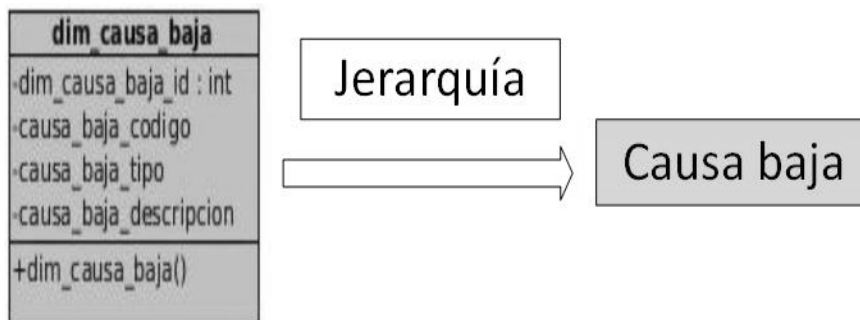
4.5 Categoría docente investigativa



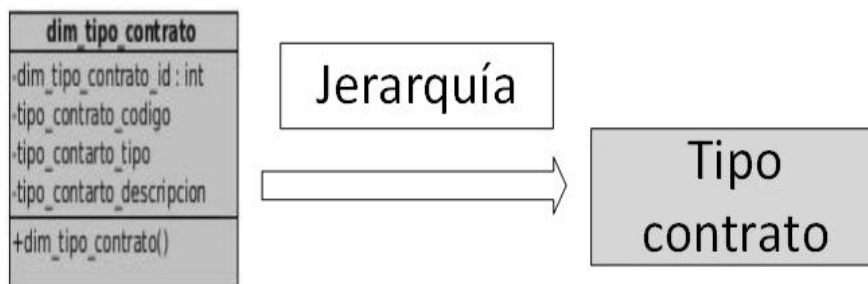
4.6 Categoría ocupacional



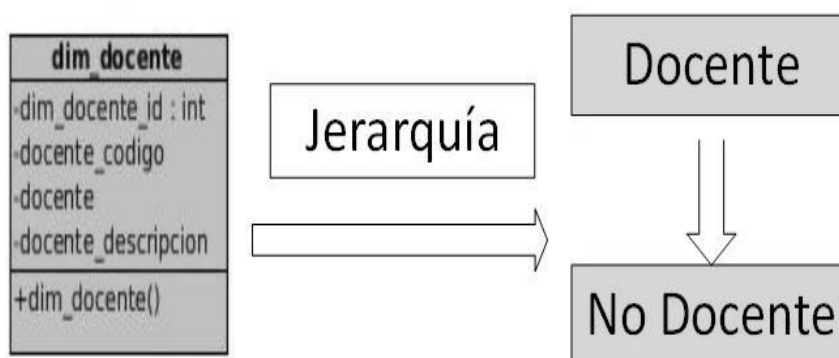
4.7 Causa baja



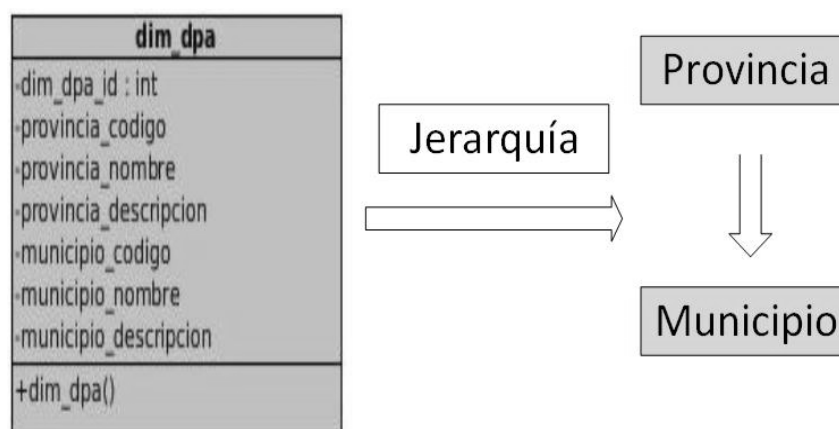
4.8 Tipo Contrato



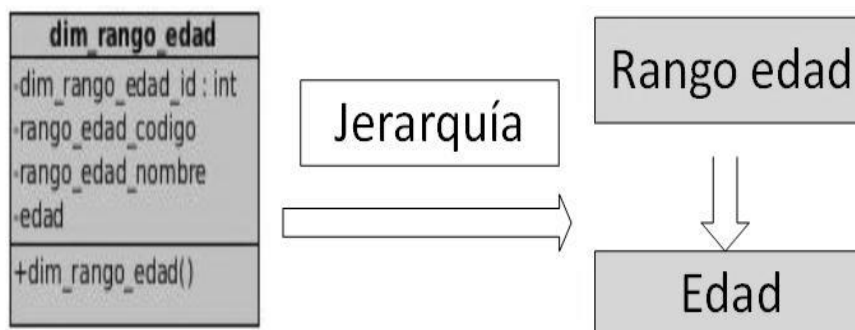
4.9 Docente



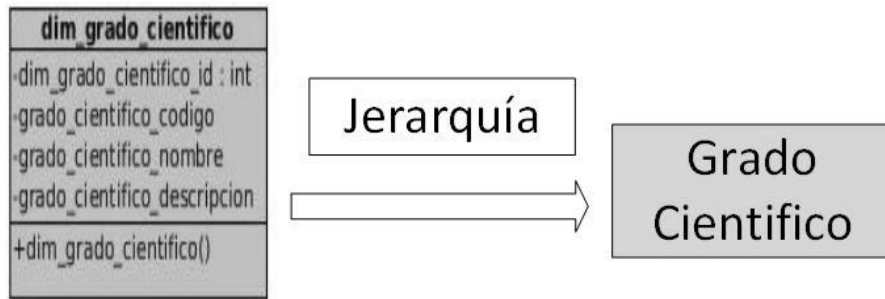
4.10 División Político Administrativa (DPA)



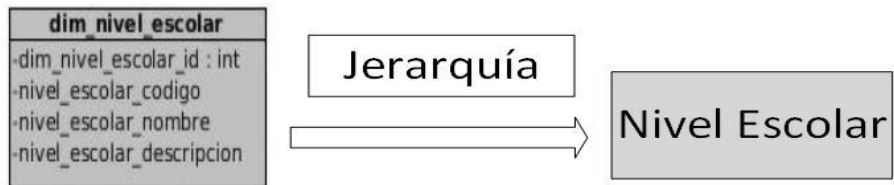
4.11 Rango edad



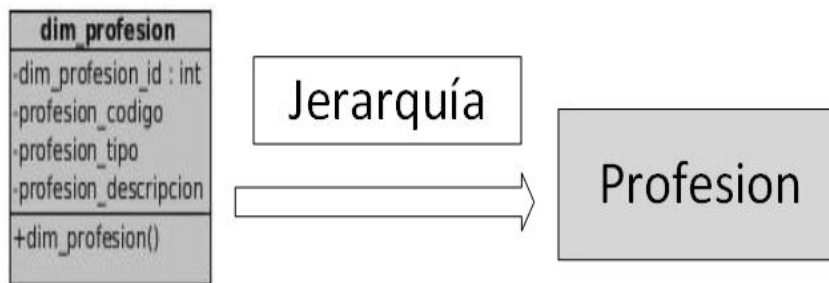
4.12 Grado científico



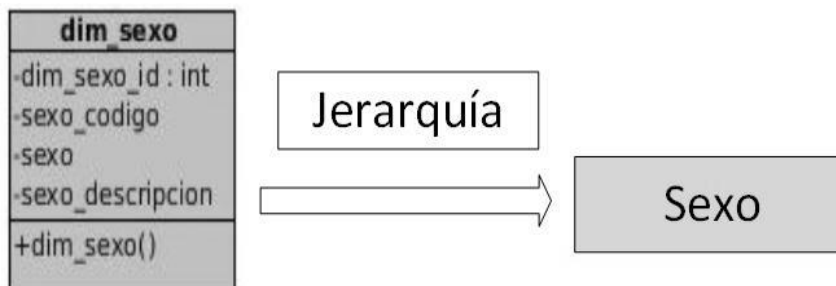
4.13 Nivel Escolar



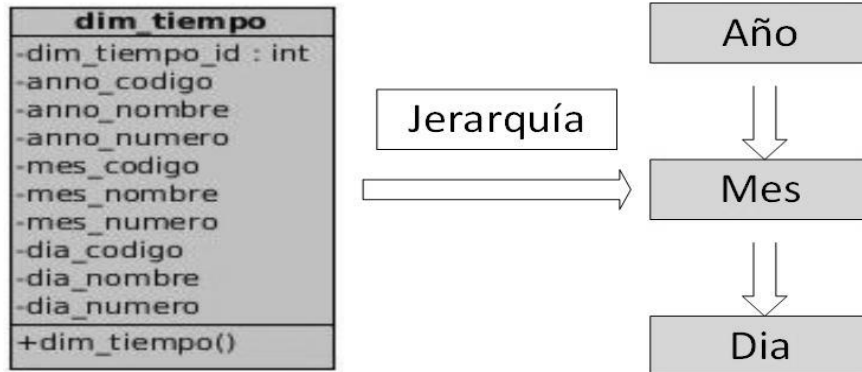
4.14 Profesión



4.15 Sexo



4.16 Tiempo



5. Acta de aceptación



Acta de aceptación

ACTA DE ACEPTACIÓN

Producto: Sistema informático para el análisis de la información de la gestión de los Recursos Humanos en la Universidad de las Ciencias Informáticas.

Involucrados en el proceso:

- **Estudiantes:** Yusnelys Pedroso Malagon.
Daynier Ramiro García Prats.
- **Tutores:** Ing. Mailin Ochoa Calzadilla.
Ing. Alexeis Joel Ochoa Reyes.

Observaciones del proceso:

Las no conformidades detectadas en el proceso de revisión fueron resueltas. Se comprobó la correcta implementación del Sistema informático para el análisis de la información de la gestión de los Recursos Humanos. Por tanto se acepta con fecha de 22 de mayo de 2016 la aplicación propuesta.

Lista de productos que son aceptados y que deben ser entregados:

- Diseño del mercado de datos: Capital Humano.
- Proceso de extracción, transformación y carga de los datos.
- Implementación de las vistas de análisis OLAP.

Entrega

Recibe

Nombre y Apellidos:

Yusnelys Pedroso Malagon.
Daynier Ramiro García Prats.

Nombre y Apellidos:

Ing. Alexeis Joel Ochoa Reyes

Cargo:

Estudiantes

Cargo:

Especialista Superior

Firma:

Firma:

Comentarios: Los productos aceptados deben ser entregados al cliente previo a la defensa.

6. Entrevista

Entrevista realizada a los especialistas de la Dirección de Desarrollo de Capital Humano de la UCI:

- ✓ ¿A qué información del área RRHH desean realizarle el análisis?

.....

- ✓ ¿Qué indicadores desean medir de la información analizada?

.....

- ✓ ¿Qué características de la información desea analizar?

.....

- ✓ ¿Cómo desea obtener la información una vez analizada?

.....

|