

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 2

Trabajo de diploma para optar por el título de Ingeniero en Ciencias
Informáticas

Título:

Mercado de Datos para la toma de decisiones a partir de la información de los estudiantes generada por el Sistema de Gestión Universitaria y el Entorno Virtual de Enseñanza Aprendizaje de la Universidad de las Ciencias Informáticas.

Autores: Lisvet Marrero Pérez

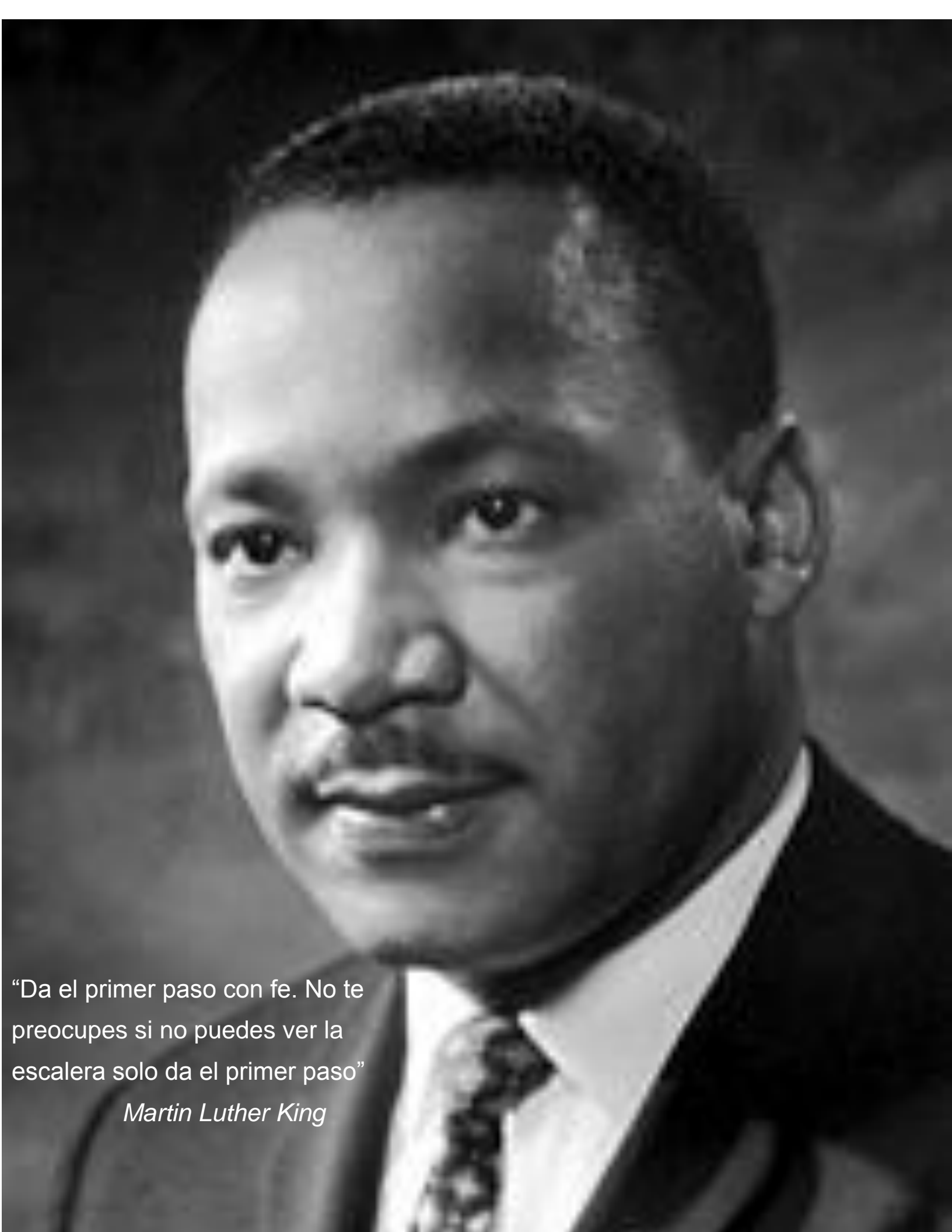
Pedro Pablo Ramos Pajón

Tutores: Ing. Lourdes Escalona Peral

Ing. Roanny Lamas López

La Habana, junio de 2015

“Año 57 de la Revolución”

A black and white portrait of Martin Luther King Jr. He is shown from the chest up, wearing a dark suit jacket, a white dress shirt, and a patterned tie. He has a serious expression and is looking slightly to the left of the camera. The background is a dark, textured grey.

“Da el primer paso con fe. No te preocupes si no puedes ver la escalera solo da el primer paso”

Martin Luther King

Declaración de Autoría

Declaramos que somos los únicos autores del presente trabajo “Mercado de Datos para la toma de decisiones a partir de la información de los estudiantes generada por el Sistema de Gestión Universitaria y el Entorno Virtual de Enseñanza Aprendizaje de la Universidad de las Ciencias Informáticas” y reconocemos a la UCI los derechos patrimoniales de la misma, con carácter exclusivo.

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

Autor: Lisvet Marrero Pérez

Autor: Pedro Pablo Ramos Pajón

Tutor: Ing. Lourdes Escalona Peral

Tutor: Ing. Roanny Lamas López

Agradecimientos

Se torna realmente difícil concentrar en solo una página a personas que de una manera u otra me han apoyado tanto. Comenzar agradeciendo a mis padres Ivet y César, por ser mis ejemplos, mis amigos incondicionales, por estar en el momento en que más los necesito y brindarme su apoyo en difíciles decisiones que en este largo andar de la vida me ha impuesto el destino. A mis abuelos Norys, Edita y Cecilio, por estar siempre al pendiente de mí, llamarme para desearme suerte en las pruebas y en todo lo que aconteciese en mi vida le doy gracias a ellos por todo. A mis tíos que son muchos pero en especial a mi tía Isa, Emília y a mi tío Ricardo que es el ejemplo más cercano que tengo de lo que es ser un profesional con prestigio, que ojalá algún día yo pueda estar a su altura, a él que estuvo al tanto del progreso de mi trabajo de diploma aconsejándome en todo momento. A mi prima Linet y su esposo Dani que gracias a ellos esta lejanía de la familia durante los cinco años se hizo menos dolorosa. En especial a mi hermanito Luisí que es quien me da el impulso de seguir adelante con el propósito de ser un buen ejemplo de hermana mayor y que ojalá siga mis pasos. A mis tutores Roanny y Lourdes, mil gracias por todos los desvelos y tiempo dedicado a nosotros. Made de más está decir que has sido como una madre para mí estos cinco años gracias por enseñarme lo que es un grupo unido y en el que tuve la oportunidad de conocer a todos las estupendas personas que hoy son más que amigos ya que los considero familia, gracias a todos ellos por los momentos inolvidables que hemos pasado juntos que son incontables y made que se cuide que en cualquier momento le hacemos otra fiesta sorpresa, de corazón, gracias a todos, en especial a mi novio Pedro Pablo al cual conocí en este grupo de amigos y que ocupa un lugar muy especial en mi vida gracias por soportar mis locuras en ocasiones un beso mi tití.

Lísvet Marrero Pérez

Agradecimientos

A mis padres María de la Caridad y Pedro Manuel por su educación, sacrificio y apoyo en todos los momentos de mi vida. Ha sido siempre su total dedicación hacia mí y su constante preocupación lo que me ha motivado a lograr lo que he logrado y a ser la persona que soy hoy en día.

A mi familia, por toda su ayuda y apoyo, en especial a mi tía Nancy por ser una persona muy especial para mí.

A mis tutores Lourdes y Roanny, por ser excelentes personas y excelentes profesionales, sin su ayuda y apoyo no hubiera sido posible este trabajo.

A todos los amigos y compañeros con los que he compartido durante estos cinco años, a los que hoy se encuentran cerca de mí y también a los que no, con todos he vivido momentos inolvidables y todos me han ayudado de una forma u otra, siempre los recordaré.

A todos los profesores que han contribuido en mi formación durante toda la carrera, en especial a Madelís que más que una profesora es una amiga.

A todas aquellas personas que fuera del ámbito universitario me han ayudado mucho durante los cinco años, en especial a Linet, Daniel, Rina y Barceló, nunca les podré agradecer todo lo que hicieron por mí.

A todas aquellas personas que de una forma u otra ayudaron en la elaboración de este trabajo.

A Lisvet, además de ser mi compañera de tesis por haber compartido conmigo estos años, eres muy importante para mí.

Pedro Pablo Ramos Pajón

Dedicatoria

Dedico especialmente este trabajo a mi hermanito Luisi como a él le gusta que le llamen ya que él es la personita más importante en mi vida y es el motor inspirador de mis logros. Este también es su fruto ojalá y yo tenga la satisfacción de que él también llegue hasta este nivel en la vida.

A todos los que creyeron y depositaron su confianza en mí, les dedico este logro.

Lísvet Marrero Pérez

Dedicatoria

A mis padres, que son lo más grande que tengo en la vida, por todo el amor que siempre me han brindado y por todos los esfuerzos y sacrificios que han realizado para cumplir este objetivo de convertirme en ingeniero.

A toda mi familia, por siempre apoyarme y estar pendientes de mí, en especial a la memoria de mi abuelo Manolo, a quien le hubiese gustado estar presente en estos momentos.

A mis amigos de Guayos y a los de la UCI, los que siempre ha estado a mi lado en las buenas y en las malas.

Pedro Pablo Ramos Pajón

Resumen

Durante el proceso de enseñanza aprendizaje, la tecnología juega un papel preponderante, destacándose en la Universidad de las Ciencias Informáticas (UCI) el uso del Entorno Virtual de Enseñanza y Aprendizaje (EVEA), el Repositorio de Objetos de Aprendizaje (RHODA), el Sistema de Gestión Universitaria (SGU), entre otras. Estas aplicaciones constituyen herramientas que les permiten a los profesores retroalimentarse del comportamiento de sus estudiantes y de esta forma implementar sus propias estrategias educativas.

Actualmente estos sistemas no son interoperables entre sí, por lo que la información se encuentra dispersa en varias bases de datos, imposibilitando el análisis integral de los estudiantes, de ahí que el presente trabajo de diploma tiene la finalidad de desarrollar un Mercado de Datos (Data Mart, por sus siglas en inglés) que permita la obtención de información dada por la interacción de los estudiantes con el EVEA y el SGU, con vistas a facilitar la toma de decisiones de directivos y profesores.

Para la construcción del Mercado de Datos (MD, por sus siglas en Español) se realizó un estudio de las metodologías, herramientas y tecnologías utilizadas en el desarrollo de este tipo de soluciones. Para de esta forma elegir luego de alcanzado cierto dominio en el tema los que respondan y sean más apropiados al entorno en que serán utilizados. Igualmente se definen los mecanismos de extracción, transformación y carga de los datos correspondientes al modelo propuesto y se realizan pruebas para validar la calidad de la solución.

Palabras Claves: Almacén de Datos (Data Warehouse), Entorno Virtual de Enseñanza Aprendizaje (EVEA), Mercado de Datos (Data Mart), Sistema de Gestión Universitaria (SGU), Toma de Decisiones.

Índice

Introducción.....	1
Capítulo 1: Fundamentación teórica de la investigación.....	6
1.1 Introducción.....	6
1.2 Introducción a los almacenes de datos.....	6
1.3 Data Mart (Mercado de Datos).....	7
1.4 Data Mart o Data Warehouse.....	8
1.5 Estudio del estado del arte.....	8
1.6 Metodologías para el diseño e implementación de un almacén de datos.....	10
1.6.1 Selección de la metodología.....	13
1.7 Sistemas OLTP y OLAP.....	13
1.7.1 Procesamiento Transaccional en Línea (OLTP).....	13
1.7.2 Procesamiento Analítico en Línea (OLAP).....	14
1.8 Modelado dimensional.....	16
1.9 Herramientas para la construcción del Data Mart.....	18
1.9.1 Sistemas Gestores de Bases de Datos.....	19
1.9.2 Herramientas de modelado.....	21
1.9.3 Herramientas para el proceso de extracción, transformación y carga.....	22
1.10 Herramientas de inteligencia de negocios.....	24
1.11 Servidor utilizado.....	25
1.12 Conclusiones parciales.....	26
Capítulo 2: Análisis, diseño e implementación del Data Mart.....	27
2.1 Introducción.....	27
2.2 Fases de la metodología Hefesto.....	27
2.2.1 Fase 1: Análisis de requerimientos.....	27
2.2.2 Fase 2: Análisis de los OLTP.....	30
Tabla 2. Determinar de indicadores.....	31
2.2.3 Fase 3: Modelo lógico del DM.....	34
2.2.4 Fase 4: Procesos ETL.....	38
2.3 Requisitos no funcionales.....	42
2.4 Conclusiones parciales.....	43

Capítulo 3: Visualización y validación.	44
3.1 Introducción.	44
3.2 Diseño de los cubos OLAP.	44
3.3 Visualización.	45
3.4 Validación de la solución.	47
3.4.1 Prueba de aceptación.	47
3.4.2 Pruebas de carga y estrés.	48
3.5 Conclusiones parciales.	50
Conclusiones Generales	51
Recomendaciones	52
Referencias Bibliográficas	53
Anexos	56

Introducción

Para la psicología, el comportamiento humano es todo lo que hace un individuo frente a los estímulos del entorno que le rodea. Cada interacción de una persona con su ambiente implica un comportamiento, donde inciden un conjunto de elementos como la cultura de las personas, las actitudes y aptitudes, sus creencias, motivaciones, entre otros.

En los tiempos actuales las Tecnologías de la Información y las Comunicaciones (TIC, por sus siglas en español), no solo han influido en los hábitos y patrones de conductas de los seres humanos, sino que además juegan un papel transformador en las formas de pensar, de trabajar, de educarse. Las mismas constituyen fuente de información, instrumentos didácticos y canal de comunicación.

En la educación, su uso no solo ha tenido gran impacto en la transformación de las formas, los métodos, y los medios de enseñanza, sino que también han influido en la formas de aprender. En la actualidad, los estudiantes juegan un rol más activo ante el aprendizaje, la tecnología les permite crear sus propios entornos de aprendizaje y llevar sus materiales a todas partes. Dicho así, los centros o instituciones educacionales no pueden estar ajenos a esta situación, por lo que constituye una preocupación el cómo perfeccionar el proceso de enseñanza aprendizaje teniendo en cuenta las potencialidades que brindan las tecnologías.

En la UCI, muchas son las aplicaciones puestas a disposición del proceso enseñanza aprendizaje (PEA) tal es el caso del Entorno Virtual de Enseñanza Aprendizaje (EVEA), disponible en <http://eva.uci.cu>. El mismo es soportado en la plataforma Moodle, y se destaca en la universidad por la utilidad de las funcionalidades que brinda. Puede ser utilizado como repositorio de información, para fomentar la comunicación síncrona y asíncrona entre estudiante - estudiante y estudiante – profesor a través de los foros de discusión y la realización de talleres. Brinda la posibilidad de realizar evaluaciones en línea, impulsar el auto aprendizaje a partir de la realización de los objetos de aprendizaje. Constituye una herramienta que permite a los profesores implementar estrategias para el proceso de enseñanza aprendizaje complementarias a las clases presenciales, así como diseñar cursos semi-presenciales o totalmente a distancia, disponiendo los estudiantes de un medio en el cual pueden obtener, utilizar o compartir materiales didácticos.

También se cuenta con el Sistema de Gestión Universitaria, disponible en <http://sgu.uci.cu>. Esta aplicación constituye una especie de expediente digital, donde se almacenan los datos

personales (nombre y apellidos, localidad de residencia, centro de procedencia, vías de ingreso, datos de los padres, entre otros), así como otros datos asociados al proceso docente entre los que se pudiera mencionar el registro de asistencia a clases, evaluaciones frecuentes, parciales y finales, promedio académico y bonificaciones.

Cabe destacar la importancia que tiene esta aplicación para el análisis de los resultados por facultad, por año, por grupos docentes, permitiendo además hacer valoraciones de estudiantes con mayor número de inasistencias a clases, identificar profesores donde se evidencia la mayor incidencia de estudiantes ausentistas. De forma general todos estos elementos permiten a los profesores y directivos ajustar las estrategias docentes cada curso académico, partiendo de las caracterizaciones del estudiantado y del claustro.

La Universidad además cuenta con una biblioteca virtual, disponible en <http://biblioteca.uci.cu>, la cual brinda un catálogo con todas las investigaciones realizadas hasta el momento, ya sean a nivel de pregrado o postgrado, además de un conjunto de artículos científicos que permiten a la comunidad universitaria estar actualizada de los avances de la ciencia acordes al perfil de la carrera.

Durante la ejecución de la actividad docente también se pudiera hacer uso de los medios audiovisuales publicados en el sitio Internos, disponible en <http://internos.uci.cu/>, el cual provee a estudiantes y profesores de conferencias, documentales y otra serie de materiales útiles para el PEA. Destacar además el papel de todas las comunidades de desarrollo existentes en la universidad, que también constituyen espacios que propician la comunicación, el intercambio de experiencias, y el desarrollo de habilidades de trabajo en equipo.

A pesar de la existencia de todas estas aplicaciones y recursos antes mencionados, se puede decir que no siempre se hace un buen aprovechamiento de la información que las mismas generan. El marco de esta investigación se centra en el análisis de la información procedente solamente del EVEA y el SGU, dada la relevancia de la información que estas producen y el impacto de su uso para el PEA. Además que se identifica que no se hace un uso adecuado de la misma, partiendo de las siguientes dificultades que así lo evidencian:

- Ambas aplicaciones cuentan con una base de datos propia lo que dificulta que se establezca un análisis transversal entre el comportamiento que manifiesta un estudiante en el EVEA y los resultados reflejados en el SGU. Por lo que actualmente preguntas como estas quedan sin solución:

- ¿Cuáles son los resultados de un estudiante que no asiste a clases, pero que está en constante interacción con el EVEA?
 - ¿Cuál es el resultado de los estudiantes que se manifiestan activos en la realización de objetos de aprendizajes, en los foros y talleres, con respecto a los resultados frecuentes y parciales reflejados en el SGU?
 - ¿Cómo se manifiestan los estudiantes en el EVEA, según el lugar de residencia, el nivel de escolaridad de los padres, el centro de procedencia?
- No existen mecanismos de recuperación de la información histórica del comportamiento de estudiantes en las aplicaciones mencionadas a lo largo de toda su carrera. Lo que imposibilita que se pueda determinar cómo fue el comportamiento de un estudiante en años anteriores, con respecto a las asignaturas que tienen precedencia, los profesores para poder consultar las trazas de años anteriores, deben tener el rol de administrador (generalmente no ocupan este rol), sin dejar de mencionar que la forma en que se muestran las trazas no es intuitiva y que muchos profesores no están en preparados para hacer este tipo de análisis.
 - Las vías que posibilitan que los profesores puedan arribar a conclusiones sobre el estado en que se encuentra el desarrollo docente educativo a partir de los datos que procesan los sistemas, aun son insuficientes. Los profesores no pueden generar reportes que les permitan hacer análisis de los estudiantes con mayores problemas de asistencia a clases y comparar con otras asignaturas del mismo año, lo mismo sucede con las comparaciones que pudiera hacer en el EVEA, no puede determinar cuáles son los estudiantes que más acceden a otras asignaturas, cuáles son los que mantienen una actitud activa, o que se mantienen en constante interacción con los distintos componentes o medios del entorno.

Partiendo del análisis de toda la problemática antes planteada, se identifica como **problema a resolver**: ¿Cómo facilitar la toma de decisiones de profesores y directivos de la Universidad de las Ciencias Informáticas a partir de la información generada de los estudiantes por el Entorno Virtual de Enseñanza Aprendizaje y el Sistema de Gestión Universitaria?

Se plantea como **objeto de estudio** el proceso de desarrollo de los mercados de datos, delimitándose como **campo de acción** el proceso de desarrollo de un mercado de datos para facilitar la toma de decisiones del claustro de profesores y directivos de la UCI.

Se identifica como **objetivo general**: Desarrollar un mercado de datos con la información generada de los estudiantes por el Entorno Virtual de Enseñanza Aprendizaje y el Sistema de Gestión Universitaria, que facilite la toma de decisiones de profesores y directivos de la UCI.

Para lograr el cumplimiento del objetivo se plantean las siguientes **tareas de investigación**:

1. Análisis de sistemas similares a la propuesta de solución, tanto a nivel nacional como internacional, para conocer aspectos regulares en la elaboración de un mercado de datos.
2. Definición de la metodología y herramientas para el desarrollo del mercado de datos.
3. Desarrollo del mercado de datos para dar solución a la problemática planteada.
4. Validación del mercado de datos propuesto a partir de técnicas existentes para verificar el cumplimiento de los requerimientos definidos.

Los métodos científicos que sustentan el trabajo son:

Teóricos:

- Analítico sintético: Se hace un análisis de la documentación, herramientas y tecnologías sobre la construcción de los almacenes de datos y mercados de datos, permitiendo sintetizar los elementos necesarios a utilizar en el desarrollo del trabajo.

Empíricos:

- Entrevistas: Se realizan entrevistas para determinar las necesidades de los clientes y de esta forma identificar los requisitos de información necesarios para la confección del mercado de datos.
- Modelación: Permitted realizar todos los modelos y diagramas asociados al diseño del mercado de datos.

El resultado que se espera con el desarrollo de la investigación es un mercado de datos que facilite la toma de decisiones de directivos y profesores en la UCI, a partir de la información generada por los sistemas EVEA y SGU, y de esta forma trazar estrategias que apoyen el proceso de enseñanza aprendizaje.

El presente Trabajo de Diploma está estructurado en tres capítulos, conformados de la siguiente manera:

Capítulo 1: Fundamentación teórica de la investigación.

En este capítulo se abordan los principales conceptos y definiciones relacionados con los almacenes de datos y los mercados de datos, así como las principales características de los mismos. Asimismo, se definen y documentan la metodología y herramientas utilizadas en la confección de la solución al problema planteado.

Capítulo 2: Análisis, diseño e implementación del mercado de datos.

En este capítulo se describen los pasos a seguir para el diseño e implementación del mercado de datos, teniendo en cuenta la metodología de desarrollo seleccionada. Primeramente se

identifican las necesidades de información del cliente, luego se pasan a analizar los sistemas que constituyen fuentes de datos, seguidamente se diseña el modelo lógico del depósito, concluyendo con los procesos de extracción transformación y carga, que permiten poblar el MD.

Capítulo 3: Visualización y validación.

En este capítulo se realiza la implementación del subsistema de visualización, mediante una herramienta que permite consultar la información almacenada y analizarla mediante tablas de datos y gráficas. Luego se procede a la validación de la solución propuesta.

Capítulo 1: Fundamentación teórica de la investigación.

1.1 Introducción.

En el presente capítulo se estarán abordando definiciones y conceptos relacionados con almacenes de datos, se muestran los resultados del estudio del estado del arte o situación actual de los sistemas existentes tanto nacional como internacionalmente similares a la propuesta de solución. También se caracterizan las principales metodologías y herramientas usadas en el desarrollo de almacenes de datos así como la justificación de las seleccionadas a utilizar.

1.2 Introducción a los almacenes de datos.

Actualmente el flujo de información que manejan las grandes empresas y organizaciones es inmenso, esto hace que sea imprescindible para las mismas contar con mecanismos y herramientas para el análisis de esta información, con el objetivo de mejorar la toma de decisiones en las distintas áreas de la institución y así mantenerse competitivas en el mercado, de ahí la importancia de conocer y dominar el uso de los almacenes de datos o Data Warehouse.

Un **Almacén de Datos** es una gran colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, y cuya actividad se centra en la *Toma de Decisiones*, es decir, en el análisis de la información en vez de en su captura. Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, lo que permite el acceso a datos históricos; así los almacenes de datos proporcionan al usuario una interfaz consolidada única para los datos, lo que hace más fácil escribir las consultas para la toma de decisiones (1).

Ralph Kimball, conocido autor en el tema de almacenes de datos lo define como: “El Almacén de Datos es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis; es la unión de todos los Data Marts de una entidad” (2).

Bill Inmon, conocido como el padre de los Data Warehouse, definió este término como “la colección de datos, orientados a materias, integrados, cambiantes con el tiempo y no volátiles, para la ayuda al proceso de toma de decisiones de la dirección de una empresa” (3).

Se pueden mencionar como las principales características de los almacenes de datos las siguientes (4):

- **Orientado a temas.** Esta característica del almacén de datos permite responder a preguntas concretas de análisis de negocios acerca de un tema de interés. Con tema de

interés, se refiere, a las áreas de la organización sobre las que se requiere hacer un análisis, como podrían ser las ventas, compras, productos, entre otros. Por ejemplo para una compañía manufacturera sus áreas de interés podrían ser: productos, órdenes, costos de materiales y materias primas.

- **Integrado.** De todos los aspectos del almacén de datos, éste es el más importante, y está estrechamente relacionado con la característica de orientado a temas. Consiste en poner en un formato consistente los datos provenientes de diversas fuentes, resolviendo problemas tales como: conflictos de nombres, inconsistencias en unidades de medidas, etc. Cuando todo lo anterior se ha llevado a cabo, entonces, se dice que está integrado.
- **No volátil.** Esta característica significa, que una vez entrado los datos al almacén, estos no cambian, debido a que el propósito de un almacén de datos es permitir el análisis de los datos de lo que ha ocurrido. Los datos del Data Warehouse son almacenados (en grandes cantidades) y accedidos, pero no actualizados.
- **De tiempo variante.** Esta característica se enfoca en los cambios sobre el tiempo para descubrir tendencias e identificar patrones ocultos (sobre grandes cantidades de datos), útiles para los analistas de negocios. La variación en el tiempo implica que cada unidad de dato del Data Warehouse es exacta para algún momento en el tiempo.

Después de analizar las definiciones y características anteriormente planteadas se concluye que un Data Warehouse no es más que un conjunto de datos históricos coleccionados y seleccionados de distintas fuentes con el objetivo de ser analizados para contribuir en la toma de decisiones estratégicas y tácticas de una organización.

1.3 Data Mart (Mercado de Datos).

Un Data Mart es una forma simple de un Data Warehouse, ya que se especializa en el almacenamiento de los datos de un tema o área específica del negocio. Se caracteriza por disponer una estructura de datos que permita analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicha área. Un Data Mart puede ser confeccionado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información. Estos sistemas tienen como principales ventajas que los mismos son simples de implementar, además de que conllevan poco tiempo de construcción y despliegue.

Los Data Marts, tienen las mismas características de integración, no volatilidad, de tiempo variante y orientación temática que el Data Warehouse. Representan una estrategia de "divide y vencerás" para ámbitos muy genéricos de un DW.

Existen dos tipos básicos de Data Mart: dependientes e independientes. La clasificación se basa principalmente en el origen de datos de los cuales se alimenta el Data Mart. Los dependientes extraen los datos de un almacén de datos central que ya se ha creado. Los independientes, por el contrario, son sistemas autónomos construidos mediante la extracción de datos directamente desde fuentes externas u operacionales de datos, o ambos.

1.4 Data Mart o Data Warehouse.

Un Data Warehouse está pensado para el ámbito de una organización en su conjunto, mientras que un Data Mart se confecciona para cubrir las necesidades de un área o departamento específicos dentro de la misma. La confección de un Data Mart conlleva menos costo en cuanto a diseño e implementación y en cuanto a tiempo, que la requerida por un Data Warehouse.

La propuesta de solución que brinda el siguiente trabajo está dirigida a analizar la información relacionada con los estudiantes, que se obtiene de las aplicaciones EVEA y SGU de la UCI, por lo que se hace más factible la implementación de un Data Mart independiente.

1.5 Estudio del estado del arte.

Muchas empresas en el mundo utilizan los almacenes de datos como vía para mejorar la toma de decisiones, en la mayoría de estas organizaciones se han experimentado resultados positivos. Dadas las características de un almacén de datos, su aplicación puede tener variados fines, en una gran diversidad de entidades.

Twentieth Century Fox

Twentieth Century Fox utiliza inteligencia de negocios para predecir qué actores, argumentos y películas serán populares en cada vecindario. La compañía tiene ahorros de aproximadamente 100 millones de dólares alrededor del mundo cada año. Esa misma tecnología la utilizan para seleccionar los "trailers" (cortos previos a la presentación de una película) alternativos para cada película en cada cine y así maximizar las ventas (5).

Walt Disney

Walt Disney, una de las compañías más grandes del mundo y mejor valoradas según la prestigiosa revista Forbes, utiliza inteligencia de negocios que de momento ha implementado en su parque de Orlando en Florida. Cada año 100 millones de personas visitan sus parques temáticos. De esta forma, la actividad de dichos visitantes puede dar lugar a un gran número de datos si se registra y almacena correctamente pueden aportar información de suma importancia. El análisis del inmenso volumen de datos generados a diario puede proporcionar a Disney un

conocimiento sólido para la toma de decisiones estratégicas, ayudando de esta forma a mejorar sus parques temáticos haciéndolos más productivos, accesibles y, por tanto, rentables, ofreciendo al visitante aquello que necesita en cada momento (6).

Utilización de técnicas de Data Warehouse para la toma de decisiones en el Área Académica

Desarrollo del Data Warehouse UNIVERSIDAD NACIONAL DE LA MATANZA (UNLaM) para tomar decisiones que ayuden al mejoramiento de los procesos de enseñanza y aprendizaje, del rendimiento de los alumnos, de la calidad de los recursos humanos docentes y de la actualización de la infraestructura, equipamiento y bibliografía. El Data Warehouse UNLaM, en una primera etapa permitió obtener los indicadores relacionados a datos de los alumnos, requeridos por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) y por Proyecto de Mejoramiento de la Enseñanza en Ingeniería (PROMEI). Luego fue posible obtener indicadores de Graduados y del Cuerpo Docente (7).

Data Warehouse aplicado al estudio del rendimiento académico y de perfiles de alumnos.

El objetivo principal de este trabajo es encontrar perfiles de alumnos por medio de la aplicación de técnicas de Almacenes de Datos (DataWarehouses: DW) con datos académicos, socio económico y demográfico correspondientes a alumnos de sistema operativo (SO) de la Licenciatura en Sistemas de Información (LSI) de la FACENA de la Universidad Nacional del Nordeste (UNNE) (8).

En los ejemplos antes expuesto se ve un exitoso uso de la Inteligencia de Negocio lo que provoca una mejora visible y palpable en la empresa guiándola por el camino correcto a seguir en cuanto a qué deben producir según las tendencias, qué precio es el adecuado, qué tendrá más aceptación y otra infinidad de factores influyentes en una empresa para que tenga éxito y pueda ganar en la competencia del mercado.

La Inteligencia Institucional no es más que una especialización de la Inteligencia de Negocios, cuya principal diferencia radica en los indicadores que permiten medir el desempeño de las organizaciones. La Inteligencia Institucional es un conjunto de procesos y métodos que permiten transformar los datos de una organización en información significativa para su desarrollo estratégico, táctico y operativo. Una cuestión interesante es cómo la Inteligencia Institucional aplicada al seguimiento estratégico puede llegar a cuestionar las propias líneas estratégicas y a evidenciar las debilidades de algunos de los indicadores o de los objetivos definidos (9).

Ámbito Nacional

Cuba no está exenta de la utilización de las herramientas para el apoyo a la toma de decisiones, en diferentes ramas de la industria, el ejemplo que se muestra a continuación es uno de los trabajos que más se asemejan y puede contribuir al desarrollo de la presente investigación, este es un trabajo de diploma desarrollado en la UCI.

Almacén de datos operacional para contribuir a la toma de decisiones basado en el uso de los recursos de la Plataforma Educativa ZERA.

La Plataforma Educativa ZERA desarrollada en el centro FORTES perteneciente a la Facultad 4 de la Universidad de las Ciencias Informáticas posibilita la gestión de los procesos de enseñanza aprendizaje. En dicha investigación desarrollaron un almacén de datos para el análisis de los indicadores obtenidos de las necesidades de información, teniendo como fuente de datos la plataforma ZERA (10).

Luego de analizadas las investigaciones antes expuestas se puede decir que este estudio ayudó a la comprensión del uso y desarrollo de los AD concretando toda la teoría a casos reales con resultados reales, a pesar de que la mayoría de las investigaciones estudiadas no trabajan específicamente en el ámbito de la toma de decisiones del proceso educativo en una institución. Se puede decir que los últimos ejemplos son los más cercanos a los objetivos del presente trabajo, pero a pesar de su similitud, como es propio de los AD estos son únicos y exclusivos al proceso de negocio al que respondan.

1.6 Metodologías para el diseño e implementación de un almacén de datos.

Se entiende por metodología de desarrollo, una colección de documentación formal referente a los procesos, las políticas y los procedimientos que intervienen en el desarrollo del software. La finalidad de una metodología de desarrollo es garantizar la eficacia (p.ej. cumplir los requisitos iniciales) y la eficiencia (p.ej. minimizar las pérdidas de tiempo) en el proceso de generación de software (11).

Existen varias metodologías para el desarrollo de un MD, no se puede definir que metodología es la mejor todo depende de las características y necesidades que tenga la empresa u organización en que se quiera construir el MD, a continuación se dará una breve explicación de las metodologías más utilizadas actualmente:

Metodología de Kimball

Esta metodología adopta este nombre en honor a su creador Ralph Kimball, considerado el inventor del Modelo Dimensional y pionero en Data Warehouse e Inteligencia de Negocios. Define

un almacén de datos como "la unión de todos los Data Marts de una entidad". Defiende por tanto una metodología ascendente (bottom-up) a la hora de diseñar un almacén de datos.

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio. Este ciclo de vida del proyecto de Data Warehouse, está basado en cuatro principios básicos:

- Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado.
- Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- Realizar entregas en incrementos significativos: Crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. En esto la metodología se parece a las metodologías ágiles de construcción de software.
- Ofrecer la solución completa: Esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

La metodología de Kimball ofrece una amplia manera de abordar los elementos para las etapas de desarrollo, y deja claro qué se debe hacer, pero no cómo lograrlo, lo que provoca demoras en los resultados (12).

Metodología de Inmon

Desde el punto de vista arquitectónico, la mayor diferencia entre Inmon y Kimball es el sentido de la construcción del DW, Inmon plantea una metodología descendente (top-down) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los Data Marts se crearán después de haber terminado el Data Warehouse completo de la organización.

De forma contraria a la de Kimball, 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, pero no es muy recomendable para proyectos sencillos pues va de lo general, el DW, a lo más específico, el DM (13).

Metodología Hefesto

Es una de las metodologías más difundidas y utilizadas por su fácil implementación y aporte práctico. Hefesto es una metodología, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias adquiridas en procesos de confección de DM. La idea principal es comprender qué se realizará en cada paso para no caer en la rutina de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo; ni por qué (14).

La metodología Hefesto propone cuatro fases fundamentales para su aplicación, cada una con una serie de puntos que se deben realizar y que se resumen en el siguiente gráfico:



Figura 1. Fases de la metodología Hefesto.

Esta metodología presenta las siguientes características (15):

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.

- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, involucrando a los usuarios finales en cada etapa para que tomen decisiones respecto al comportamiento y funciones del almacén.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de la estructura física que contenga el almacén de datos y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para DW como para DM.

1.6.1 Selección de la metodología.

Después de analizar las metodologías antes mencionadas se selecciona Hefesto como metodología de desarrollo a utilizar, esta permite la construcción del MD de forma sencilla, ordenada e intuitiva. Hefesto es una metodología ideal para las personas que entran por primera vez en el mundo de los DW, pues permite construir un MD de manera metódica, guiándose por pasos lógicos relacionados sólidamente durante todas las etapas del proceso de confección. Dicha metodología agiliza el proceso de desarrollo del MD ya que propone no entrar en fases extensas de análisis y reunión de requerimientos, ni fases de despliegue muy largas (14).

1.7 Sistemas OLTP y OLAP.

Los sistemas de Procesamiento Transaccional y Analítico en Línea conocidos como OLTP y OLAP respectivamente, son conceptos de indispensable análisis para un mejor entendimiento y posterior desarrollo de los Data Warehouse, es de suma importancia reconocer cuando se está en presencia de uno u otro sistema.

1.7.1 Procesamiento Transaccional en Línea (OLTP).

Procesamiento Transaccional en Línea (On-Line Transaction Processing), representa toda aquella información transaccional que genera la organización en su accionar diario.

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado o invalidado), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales (16).

Las características principales de los sistemas OLTP son (17):

- Realizan transacciones en tiempo real del proceso de un negocio, con lo cual los datos almacenados cambian continuamente. Los sistemas OLTP en sus transacciones conducen procesos esenciales del negocio.
- Los sistemas OLTP son los responsables del mantenimiento de los datos, ya sea agregando datos, realizando actualizaciones o bien eliminándolos.
- Las estructuras de datos deben estar optimizadas para validar la entrada de los mismos, y rechazarlos si no cumplen con determinadas reglas de negocio.
- Para la toma de decisiones, proporciona capacidades limitadas ya que no es su objetivo, por lo tanto no es prioridad en su diseño. Si se quisiera obtener determinada información histórica relativa al negocio consultando un sistema OLTP, se produciría un impacto negativo en el funcionamiento del sistema.

1.7.2 Procesamiento Analítico en Línea (OLAP).

Los sistemas OLAP (On-Line Analytical Processing) 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, elaboración de informes complejos, etc. Este sistema es típico de los Data Marts (16).

Los sistemas OLAP proporcionan una alternativa a los sistemas transaccionales (OLTP), ofreciendo una visión de los datos orientada hacia el análisis y una rápida y flexible navegación por estos. Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes bases de datos o Sistemas Transaccionales. Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos.

Las siguientes son características que los sistemas OLAP poseen (17):

- Las bases de datos OLAP tienen un esquema que está optimizado para que las preguntas realizadas por los usuarios sean respondidas rápidamente.
- Las preguntas que se le hacen a un OLAP, deben permitir un uso interactivo con los usuarios.
- Un sistema OLAP está preparado para realizar informes complejos de una manera simple.

- OLAP proporciona una vista de datos multidimensional. Los cubos proporcionan una vista de los datos multidimensional que se extiende más allá del análisis de dos dimensiones que puede proporcionar una simple planilla de cálculo utilizada como tal.
- Los usuarios pueden cambiar fácilmente las filas, las columnas, y las páginas en informes de OLAP, pudiendo leer la información de la manera que se crea más conveniente para el análisis.

Existen varias arquitecturas para los sistemas OLAP como: MOLAP (OLAP Multidimensional), ROLAP (OLAP Relacional) y HOLAP (OLAP Híbrido).

MOLAP (Multidimensional On-line Analytical Process), estas herramientas utilizan bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. El almacenaje de MOLAP, provee excelente rendimiento y compresión de datos. El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato. El motor analítico es el responsable de la ejecución de los requerimientos OLAP y de proporcionar una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP (18).

ROLAP (Relacional On-line Analytical Process) arquitectura OLAP que crea vistas multidimensionales extrayendo los datos de bases de datos relacionales, simulan los datos multidimensionales usando sofisticadas técnicas de indexación. ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja el almacenamiento de los datos, el motor OLAP proporciona la funcionalidad analítica y alguna herramienta especializada es empleada para el nivel de presentación (19).

HOLAP. Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado (16).

En la solución propuesta se utiliza la arquitectura ROLAP debido ya que los datos se extraen de una base de datos soportada en el Sistema Gestor de Bases de Datos PostgreSQL donde se modela un diseño relacional que simula uno multidimensional a través de hechos y dimensiones.

1.8 Modelado dimensional.

El modelado dimensional brinda la posibilidad de visualizar los datos. La habilidad de visualizar algo como abstracción de un conjunto de datos en una manera concreta y tangible es el secreto de la comprensión. La modelación dimensional es un nuevo nombre para una técnica de hacer simples y comprensibles bases de datos. Cuando una base de datos puede ser visualizada como un “cubo” de tres, cuatro, o aún cinco o más dimensiones, las personas pueden imaginarse dividiendo y cortando en cubitos ese cubo a lo largo de cada una de esas dimensiones (20).

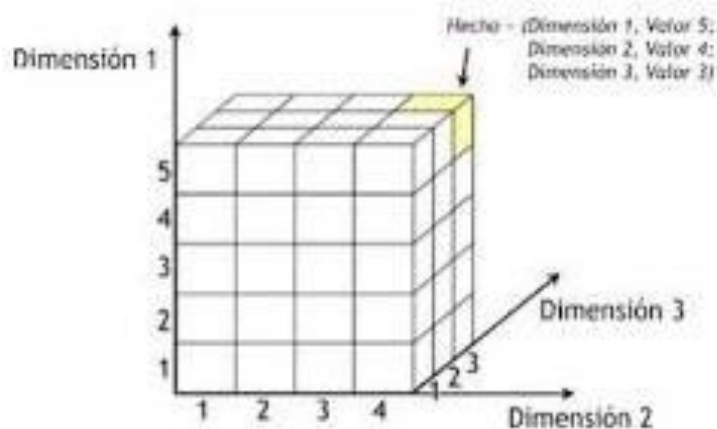


Figura 2. Cubo multidimensional.

Este modelo tiene en cuenta que para el análisis multidimensional los datos se representan como si estuvieran en un espacio n-dimensional (cubo de datos), permitiendo su estudio en términos de hechos sujetos al análisis y dimensiones que permiten diferentes puntos de vista por los que analizar esos hechos.

El modelo dimensional distingue tres elementos básicos:

- **Hechos:** es la representación en el Data Warehouse de los procesos de negocio de la organización. Los hechos se podrán reconocer además porque siempre tienen asociada una fecha, y una vez registrados no se modifican ni se eliminan (para no perder la historia). Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Idealmente está compuesto por valores numéricos, continuamente evaluados y aditivos.
- **Medidas:** Son aquellos conceptos cuantificables que permiten medir nuestro proceso de negocio. Son características cualitativas o cuantitativas de los objetos que se desean analizar en las empresas. Las medidas, siempre son numéricas, se almacenan en las tablas de hechos.

- **Dimensiones:** Las dimensiones organizan los datos en función de un área de interés para los usuarios. Cada dimensión describe un aspecto del negocio y proporciona el acceso intuitivo y simple a datos. Representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos, es decir son las que alimentan a la tabla de hechos.

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

Esquema Estrella: Esta estructura es considerada las más básica y sencilla de diseñar con respecto a las existentes. Está definida por una tabla central o tabla de hechos y un conjunto de tablas organizadas alrededor de ésta, denominadas tablas de dimensiones, brindando la sensación de una estrella. Las tablas de dimensiones solo se relacionan con la tabla de hechos y no existen relaciones entre dimensiones.

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 (15).

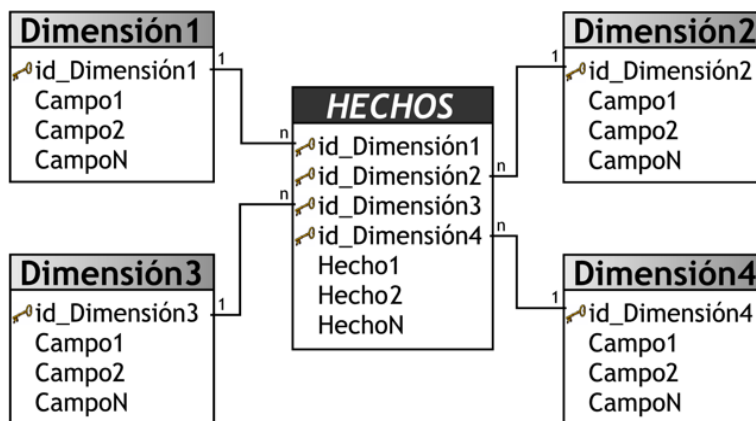


Figura 3. Esquema estrella.

Esquema Copo de Nieve: El modelo copo de nieve es una variación o derivación del modelo estrella. En este modelo la tabla de hechos deja de ser la única relacionada con otras tablas, ya que existen otras tablas que se relacionan con las dimensiones y que no tienen relación directa con la tabla de hechos.

A diferencia del estilo estrella, los datos de las dimensiones se reparten en múltiples tablas. Como ventaja principal de este estilo es el ahorro de espacio de almacenamiento en disco, pero en perjuicio de un aumento en la cantidad de tablas y disminución de rendimiento para la recuperación de información (15).

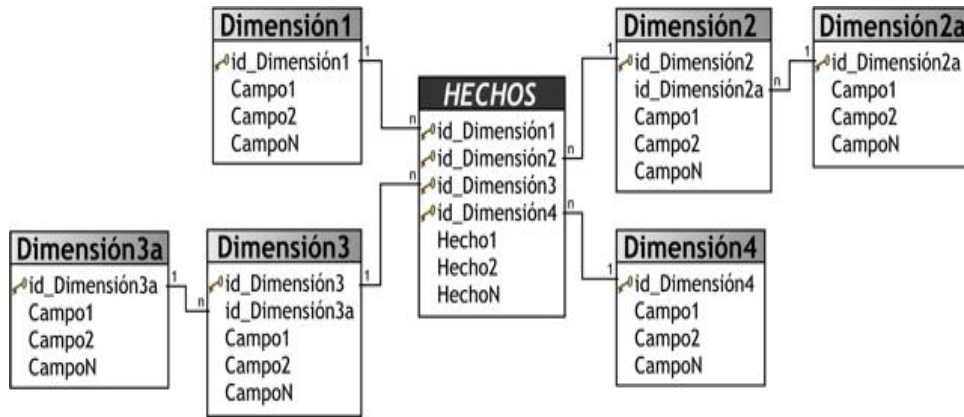


Figura 4. Esquema copo de nieve.

Esquema Constelación: 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.

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 (15).

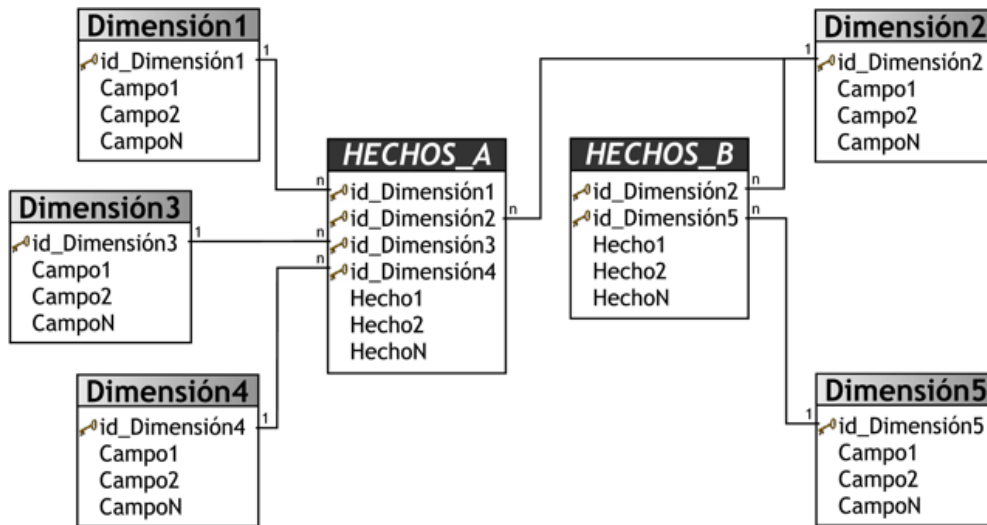


Figura 5. Esquema constelación de hechos.

1.9 Herramientas para la construcción del Data Mart.

Las herramientas para la construcción del Data Mart se dividieron en 3 grupos principales, primeramente los sistemas gestores de bases de datos, por otra parte las herramientas de modelado y por último las herramientas de integración de datos, realizando un estudio basado en soberanía tecnológica, usabilidad, así como documentación sobre el uso de las mismas.

Teniendo en cuenta también las características que presenta el sistema, este debe estar soportado por un conjunto de herramientas donde se establece una cooperación entre ellas, para transitar por las diferentes etapas del proceso de desarrollo.

1.9.1 Sistemas Gestores de Bases de Datos.

Oracle:

Es un sistema de gestión de base de datos relacional, se considera como uno de los sistemas de bases de datos más completos, destacando su soporte de transacciones, estabilidad y escalabilidad. Es multiplataforma y está distribuido bajo licencia propietaria, siendo esta una de sus mayores desventajas ya que su precio puede ser alto (teniendo en cuenta versiones y licencias) (21).

Es un SGBD con características objeto-relacionales, que pertenece al modelo evolutivo de SGBD. Presentando entre sus características principales las siguientes (22):

- Entorno cliente/servidor.
- Gestión de grandes bases de datos.
- Usuarios concurrentes.
- Alto rendimiento en transacciones.
- Sistemas de alta disponibilidad.
- Gestión de la seguridad.
- Autogestión de la integridad de los datos.
- Opción distribuida.
- Compatibilidad.
- Conectabilidad.
- Replicación de entornos.

MySQL:

Es un sistema gestor de bases de datos relacionales rápido, sólido y flexible, bajo la Licencia Pública General. Es idóneo para la creación de bases de datos con acceso desde páginas web dinámicas, así como para la creación de cualquier otra solución que implique el almacenamiento de datos, posibilitando realizar múltiples y rápidas consultas. Su diseño multihilo le permite soportar una gran carga de forma eficiente.

Algunas de las características principales de este SGBD son (23):

- Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo.

- Soporta gran cantidad de tipos de datos para las columnas.
- Dispone de API's¹ en gran cantidad de lenguajes (C, C++, Java, PHP, etc.).
- Gran portabilidad entre sistemas.
- Gestión de usuarios y contraseñas, manteniendo un muy buen nivel de seguridad en los datos.

PostgreSQL:

Es un sistema gestor de bases de datos relacionales orientadas a objetos, de código abierto, que permite trabajar con grandes volúmenes de datos; soporta gran parte de la sintaxis SQL y cuenta con un extenso grupo de enlaces con lenguajes de programación.

Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras bases de datos comerciales. PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando (24).

Entre sus principales características se encuentran (25).

- 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, etc. También permite la creación de tipos propios.
- Incorpora una estructura de datos arreglo.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- 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).
- Cuenta con un amplio conjunto de interfaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros.

¹ API: Application Programming Interface. Un API no es más que una serie de servicios o funciones que el Sistema Operativo ofrece al programador.

Luego del análisis de las herramientas antes descritas, se decide utilizar como sistema gestor de base de datos, el PostgreSQL, ya que el mismo al ser open source brinda la posibilidad de realizarle cualquier cambio o modificación a conveniencia sin tener que pagar por ello, también provee estabilidad, integridad y confiabilidad al Data Mart, además de soportar grandes cantidades de datos y de usuarios accediendo al sistema al mismo tiempo.

La herramienta a utilizar para la administración de la base de datos es PgAdmin. Herramienta diseñada para responder a la mayoría de las necesidades de los usuarios, desde escribir simples consultas, hasta desarrollar bases de datos complejas. El interfaz gráfico soporta todas las características de PostgreSQL y facilita enormemente la administración. Destacándose la característica de ser una herramienta con licencia open source, además de ser multiplataforma, o sea que se puede usar en Windows, Linux y otros sistemas operativos.

1.9.2 Herramientas de modelado.

Enterprise Architect:

Enterprise Architect es una herramienta comprensible de diseño y análisis UML, cubriendo el desarrollo de software desde el paso de los requerimientos a través de las etapas del análisis, modelos de diseño, pruebas y mantenimiento. Es una herramienta multi-usuario, basada en Windows, diseñada para ayudar a construir software robusto y fácil de mantener. Ofrece salida de documentación flexible y de alta calidad (26).

Entre sus características principales se encuentran (26):

- Extensiones personalizadas para modelado de procesos y más.
- Intuitivo y simple de usar.
- Documentación de alta calidad.
- Bajo costo de licencias.

Visual Paradigm:

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. 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. Está diseñado para una amplia gama de usuarios interesados en la construcción de sistemas de software de forma fiable a través de la utilización de un enfoque Orientado a Objetos (27).

Entre sus características principales se encuentran (27):

- Disponibilidad en múltiples plataformas (Windows, Linux).
- Diseño centrado en casos de uso y enfocado al negocio que generan un software de mayor calidad.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
- Capacidades de ingeniería directa e inversa.
- Licencia: gratuita y comercial.
- Generación de bases de datos - Transformación de diagramas de Entidad-Relación en tablas de base de datos.
- Generador de informes.

Luego del análisis de las herramientas antes descritas, se decide utilizar para el proceso de modelado la herramienta Visual Paradigm, debido a las grandes facilidades que brinda durante todo el proceso de desarrollo del Data Mart, permitiendo la creación de todos los artefactos y diagramas necesarios. Destacándose además que la misma posee licencia gratuita y es multiplataforma.

1.9.3 Herramientas para el proceso de extracción, transformación y carga.

CloverETL:

Es una plataforma de integración rápida de datos que ayuda a las organizaciones a obtener más valor de sus datos de forma rápida, sin gastos de infraestructura masivos o años de inversión del proyecto. Con CloverETL, se resuelven los retos de integración y manipulación de datos complejos, trabaja con proyectos de datos, formatos y protocolos, incluso cuando estos implican grandes volúmenes de datos (28).

Es un ambiente de transformación de datos de código abierto basado en Java utilizado entre otras funciones en el movimiento de datos entre cualquier número de diferentes sistemas y bases de datos. Es altamente configurable y muy rápido.

Características principales de CloverETL (29):

- 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).
- 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 muchos otros.

Enhydra Octopus:

Enhydra Octopus es una herramienta ETL basada en Java. Se puede conectar a cualquier fuente de datos JDBC (Conexión de Java a la Base de Datos) y realizar transformaciones definidas en un archivo XML. Esta herramienta brinda una forma muy genérica para transformar los datos. Puede definir transformaciones mediante la implementación de la interfaz del transformador o el uso de código JavaScript (directamente en el archivo XML trabajo de carga) (30).

Características principales de Enhydra Octopus (31):

- La característica principal de Octopus es el requerimiento de que 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:

Los datos que alimentan a un sistema DW provienen de diferentes fuentes, estas fuentes son los distintos sistemas operacionales que la organización posee, generalmente ni son homogéneos entre sí ni concuerdan exactamente, Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones hoy en día. Pentaho Data Integration permite una poderosa extracción, transformación y carga. El uso de esta solución permite evitar grandes cargas de trabajo manual, frecuentemente difícil de mantener y de desplegar (32).

Uno de los objetivos de la integración de datos es que el proceso de ETL sea fácil de generar, mantener y desplegar. Pentaho Data Integration se compone de cuatro herramientas (33):

- SPOON: permite diseñar de forma gráfica la transformación².
- PAN: ejecuta un conjunto de transformaciones diseñadas con SPOON, conocidas como trabajos, creando dependencias entre dichas transformaciones.
- CHEF: permite, mediante una interfaz gráfica, diseñar la carga de datos incluyendo un control de estado de los trabajos.

² Transformación: es el elemento básico de diseño de los procesos ETL en PDI. Una transformación se compone de pasos enlazados entre sí a través de los saltos. Los saltos constituyen el elemento a través del cual fluye la información.

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

Además de ser de código abierto, las características básicas de esta herramienta son (34):

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Macintosh, Linux.
- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones).

Luego del análisis de las herramientas antes descritas, se decide utilizar para el proceso de extracción, transformación y carga de los datos, la herramienta Pentaho Data Integration, además de la existencia de experiencias previas en la UCI, esta herramienta de código abierto y multiplataforma brinda libertad y flexibilidad para diseñar soluciones de la manera más apropiada para cada ambiente determinado, hoy en día constituye una herramienta muy utilizada mundialmente, debido a las posibilidades que brinda, además de ser relativamente simple de utilizar.

1.10 Herramientas de inteligencia de negocios.

Luego de obtener el mercado de datos completamente poblado con los datos necesarios se eligen utilizar las herramientas que a continuación se describen para cubrir el proceso necesario hasta la visualización de la información por parte de los usuarios.

Mondrian OLAP Server:

Mondrian es un servidor OLAP de código abierto que gestiona la comunicación entre una aplicación OLAP y la base de datos con los datos fuente.

Se encarga de recibir consultas dimensionales en lenguaje MDX (por sus siglas en inglés Multidimensional QueryExpression) y devolver los datos de un cubo, sólo que este cubo no es algo físico sino un conjunto de metadatos que definen como se han de “mapear” estas consultas que tratan conceptos dimensionales a sentencias SQL (35). Permite realizar consultas a un AD logrando que los resultados sean presentados mediante un navegador.

Schema Workbench:

Es una aplicación que permite crear visualmente y testear esquemas Mondrian de cubos OLAP. Se caracteriza por su potencia gráfica y capacidad multitarea.

En los ficheros de esquema XML, generados por Workbench, se describen la relación entre las dimensiones y medidas del cubo (modelo multidimensional) con las tablas y campos de la base de datos, a nivel relacional (36).

Con esta aplicación, se puede configurar una conexión JDBC con el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello el entorno ofrece un editor de esquemas con la fuente de datos subyacente para su validación y permite la ejecución de consultas MDX contra el esquema y la base de datos (35).

Saiku Analytics:

Es una aplicación web de código libre que proporciona un visor de consulta y análisis OLAP amigable e intuitivo. Saiku forma parte de Pentaho BI Suite Community Edition y se distribuye bajo licencia Apache. Es el nuevo nombre que se le dio a PAT (Pentaho Analysis Tool) el cual estaba destinado a ser el visualizador OLAP de la versión community de Pentaho, reemplazando a JPivot, una librería de componentes JSP³ que se utiliza para construir tablas OLAP generadas de forma dinámica (37).

Saiku es un navegador OLAP muy liviano e intuitivo, que saca provecho de muchos recursos que mejoran la experiencia del usuario. Permite a los usuarios construir sus propias vistas arrastrando y soltando campos y visualizando los resultados según se va construyendo la consulta, permitiendo analizar los datos en forma de tabla pivotante o en modo gráfico (37).

1.11 Servidor utilizado.

Apache Tomcat

También conocido como simplemente Tomcat o Jakarta Tomcat, es un servidor web multiplataforma que funciona como contenedor de servlets⁴ y que se desarrolla bajo el proyecto denominado Jakarta perteneciente a la Apache Software Foundation bajo la licencia Apache 2.0. Implementa las especificaciones de los servlets y de JSP de Sun Microsystem. Dicho servidor es mantenido y desarrollado por miembros de la fundación y voluntarios independientes, los cuales tienen libre acceso al código fuente bajo los términos establecidos por la Apache Software Foundation (38).

³ JSP: acrónimo de Java Server Pages es una tecnología Java que permite generar contenido dinámico para la Web, en forma de documentos HTML, XML o de otro tipo.

⁴ Servlets: Son programas que se ejecutan en un servidor Web y construyen páginas Web.

1.12 Conclusiones parciales.

Luego de la realización de este capítulo se obtienen como conclusiones el haber logrado un mayor entendimiento de las características, objetivos y usos de los almacenes de datos a nivel internacional y nacional gracias al estudio del estado del arte realizado. Se logra también, tener un amplio conocimiento y dominio sobre las principales metodologías, herramientas y tecnologías utilizadas en la confección de los mismos.

Se define utilizar como guía en el desarrollo del Data Mart la metodología Hefesto, apoyándose en las siguientes herramientas. Se decide utilizar como gestor de bases de datos el PostgreSQL, para el modelado de los artefactos el Visual Paradigm y como herramienta para el proceso ETL, Pentaho Data Integration, se decide utilizar Mondrian como motor OLAP, para el diseño de los cubos multidimensionales la herramienta Schema Workbench, como servidor web el Apache Tomcat y para la visualización de la información se define el Saiku Analytics.

Capítulo 2: Análisis, diseño e implementación del Data Mart.

2.1 Introducción.

En el presente capítulo se describe el proceso de análisis, diseño e implementación del Data Mart, siempre tomando en consideración las etapas y pasos brindados por la metodología de desarrollo de software escogida en el capítulo anterior. Se hace una descripción detallada de cada una de las fases de dicha metodología, partiendo inicialmente de la identificación de los requerimientos de información del usuario, hasta llegar a la definición de los procesos ETL para poblar completamente el Data Mart.

2.2 Fases de la metodología Hefesto.

La metodología Hefesto consta de 4 fases y cada una de ellas cuenta con una serie de pasos puntuales que guiarán al desarrollador en todo el proceso.

2.2.1 Fase 1: Análisis de requerimientos.

En esta primera fase, se comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio que permitan dar cumplimiento a los objetivos trazados. Luego del análisis de estas preguntas, se deben identificar los indicadores y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del DM.

Identificar preguntas claves.

Para el desarrollo del primer paso de la metodología se realizó una entrevista a la profesora principal de quinto año de la facultad 2 de la UCI, la Ing. Lourdes Escalona Peral con el propósito de obtener las preguntas claves, los resultados que se esperan obtener y los reportes que considere importantes.

De la entrevista realizada se obtuvieron las necesidades de información del cliente las cuales se pueden apreciar completamente en el **Anexo #1: Necesidades de información del cliente**, a continuación se brinda una representación de estos requerimientos ya agrupados para su estudio teniendo en cuenta el interés del cliente.

1- Se desea conocer el promedio general de los estudiantes por centro de procedencia, sexo, provincia, grado de escolaridad del padre, grado de escolaridad de la madre, promedio de procedencia, raza, ocupación de la madre, ocupación del padre.

2- Es de interés saber el promedio de ausencias de los estudiantes por asignatura, por provincia, sexo, centro de procedencia, en un tiempo determinado.

3- Se quiere saber el promedio de interacción en el EVEA de los estudiantes por asignatura, sexo, provincia, raza y situación escolar en un tiempo determinado.

4- Se desea conocer el tiempo de interacción en el EVEA de los estudiantes por asignatura, en un período determinado.

5- Se quiere saber la cantidad de entradas en el EVEA de los estudiantes por asignatura, provincia, sexo y raza.

6- Es de interés la cantidad de estudiantes matriculados en el EVEA por asignatura, provincia, sexo, centro de procedencia, vía de ingreso.

7- Se desea conocer la cantidad de ausencias de un estudiante por situación escolar, vía de ingreso, provincia, en un tiempo determinado.

Identificar indicadores y perspectivas de análisis.

Los **indicadores** son valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: promedios, cantidades, sumatorias, fórmulas, entre otras (15).

Las **perspectivas** son objetos mediante los cuales se quiere examinar los indicadores, con el objetivo de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, entre otras (15).

De las entrevistas realizadas se obtuvieron las preguntas claves para el negocio, y a partir de estas preguntas se obtuvieron los siguientes indicadores con sus correspondientes perspectivas.

Indicador	Perspectivas
Promedio de notas	Estudiante Centro de procedencia Sexo Provincia Grado de escolaridad del padre Grado de escolaridad de la madre Promedio de procedencia Ocupación de la madre Ocupación del padre Raza
Promedio de ausencias	Estudiante Provincia Sexo Centro de procedencia En un tiempo determinado Asignatura

Promedio de interacción en el EVEA	Estudiante Asignatura Sexo Provincia Raza Situación escolar Tiempo
Cantidad de interacción en el EVEA	Asignatura Estudiante
Cantidad de entradas en el EVEA	Asignatura Estudiante Provincia Sexo Raza
Cantidad de estudiantes	Provincia Asignatura Sexo Centro de procedencia Vía de ingreso
Cantidad de ausencias	Tiempo Situación escolar Vía de ingreso Provincia

Tabla 1. Indicadores y perspectivas.

Modelo conceptual.

Luego de identificados los indicadores y las perspectivas, se construye el modelo conceptual, el cual constituye una descripción de la estructura de la base de datos, en la cual la información es representada a través de objetos y relaciones.

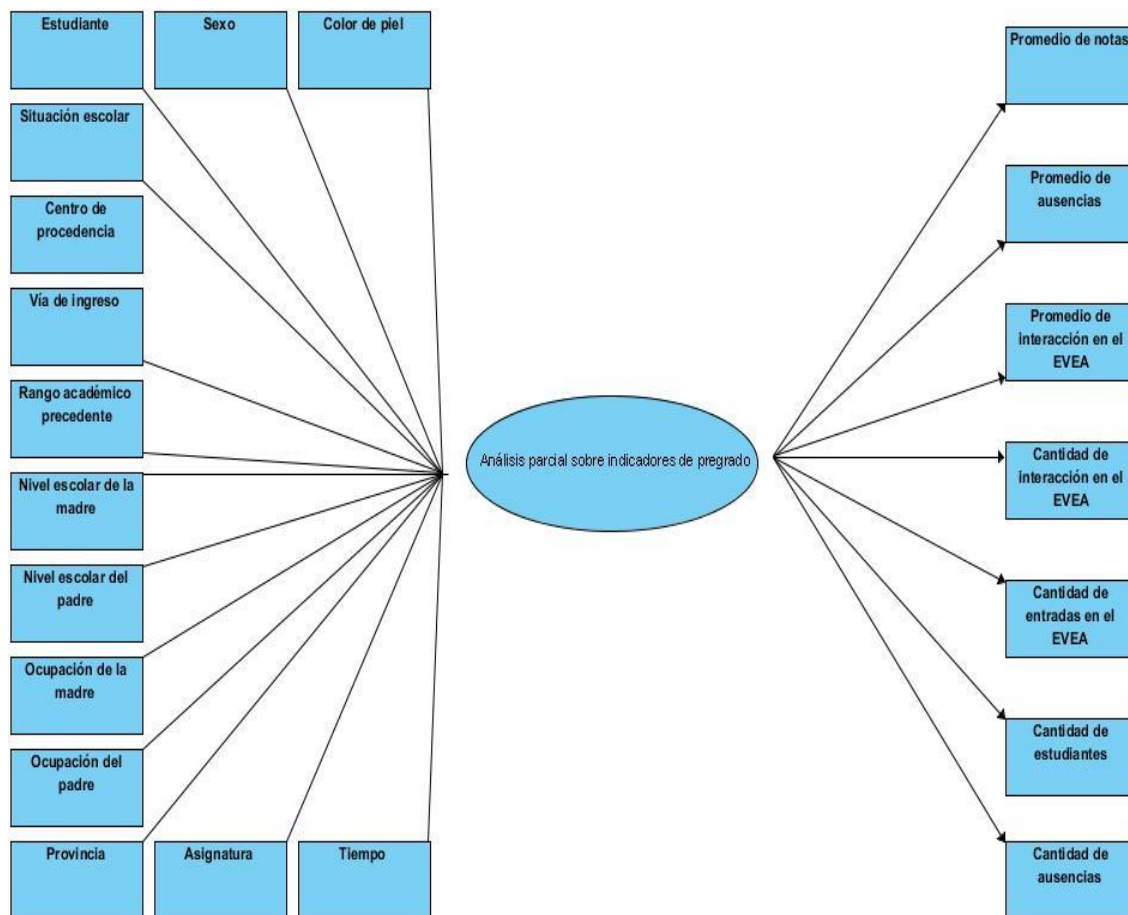


Figura 6. Modelo conceptual.

2.2.2 Fase 2: Análisis de los OLTP.

En esta fase primeramente se realiza un análisis de las fuentes de datos, con el objetivo de establecer una correspondencia entre estas y el modelo conceptual creado en el paso anterior, también se determina cómo se calculan los indicadores identificados, obteniéndose además los niveles de granularidad y finalmente se construye el modelo conceptual ampliado con la información obtenida hasta este paso.

Determinación de indicadores.

Indicador	Función de sumariación	Descripción
Promedio de notas	AVG	Promedio de notas de los estudiantes.
Promedio de ausencias	AVG	Promedio de ausencias de los estudiantes.
Promedio de interacción en el EVEA	AVG	Promedio de tiempo que han interactuado los estudiantes con el EVEA.

Cantidad de interacción en el EVEA	SUM	Tiempo total que ha permanecido el estudiante en el EVEA.
Cantidad de entradas en el EVEA	SUM	Cantidad de veces que ha entrado el estudiante en el EVEA.
Cantidad de estudiantes	COUNT	La cantidad de estudiantes matriculados en el EVEA.
Cantidad de ausencias	SUM	Cantidad de ausencias de los estudiantes.

Tabla 2. Determinar de indicadores.

Establecer correspondencias.

En este paso se establecen las correspondencias entre el modelo conceptual y los sistemas fuentes, para una mejor especificación y entendimiento del Data Mart. Para el desarrollo de la investigación se requería el uso de la base de datos del SGU, pero por razones de seguridad de la universidad no fue posible su adquisición. Por esta razón los sistemas fuentes analizados son la base de datos del EVEA y dos reportes en formato Excel del SGU brindados por la profesora principal de quinto año de la facultad 2 y por el jefe de departamento del centro de informatización el compañero Alexander Rodríguez Mompíe.

Perspectiva	Fuente	Campo
Estudiante	Reporte1.xls	-Nombre -Apellidos
Sexo	Reporte1.xls	-Sexo
Color de piel	Reporte Lourdes.xls	-Color de piel
Situación escolar	Reporte1.xls	-Situación Escolar
Centro de procedencia	Reporte1.xls	-Centro de Procedencia
Vía de ingreso	Reporte Lourdes.xls	-Vía de ingreso
Rango académico precedente	Reporte Lourdes.xls	-Índice académico
Nivel escolar de la madre	Reporte Lourdes.xls	-Nivel Escolar de la Madre o tutora
Nivel escolar del padre	Reporte Lourdes.xls	-Nivel escolar del padre o tutor
Ocupación de la madre	Reporte Lourdes.xls	-Ocupación de la madre o tutora

Ocupación del padre	Reporte Lourdes.xls	-Ocupación del padre o tutor
Provincia	Reporte1.xls	-Provincia
Asignatura	Reporte1.xls	Reporte1.xls
Tiempo	Reporte1.xls	-Fecha de cierre

Tabla 3. Establecer correspondencias.

Nivel de granularidad.

Luego de establecidas las relaciones entre los sistemas fuentes y el modelo conceptual definido y luego de determinar la forma en que se calculan los indicadores, en este paso se procede a definir los campos que contendrán cada perspectiva, ya que será a través de estos por donde se filtrarán y manipularán los diferentes indicadores. El resultado obtenido fue el siguiente.

Perspectiva	Detalle	Descripción
Estudiante	Nombre y apellidos	Nombre y apellidos del estudiante.
Sexo	Sexo	Sexo del estudiante.
Color de piel	Raza	Color de piel del estudiante.
Situación escolar	Promoción	Situación escolar en la que se encuentra el estudiante.
Centro de procedencia	Procedencia	Centro de estudio del que proviene el estudiante.
Vía de ingreso	Vía de ingreso	Vía por la cual ingreso el estudiante a la UCI.
Rango académico precedente	Promedio	Rango del índice académico con el que ingresó el estudiante a la UCI.
Nivel escolar de la madre	Nivel escolar	Nivel escolar de la madre o tutora del estudiante.
Nivel escolar del padre	Nivel escolar	Nivel escolar del padre o tutor del estudiante.
Ocupación de la madre	Ocupación	Ocupación de la madre o tutora del estudiante.

Ocupación del padre	Ocupación	Ocupación del padre o tutor del estudiante.
Provincia	Nombre	Nombre de la provincia del estudiante.
Asignatura	Nombre	Nombre del curso que se imparte a los estudiantes
Tiempo	Año Mes Día	Tiempo de cierre de un semestre

Tabla 4. Nivel de granularidad.

Modelo conceptual ampliado.

En este paso se procede a ampliar el modelo conceptual confeccionado en la fase anterior. Se grafican los resultados obtenidos en los pasos anteriores, colocando debajo de cada perspectivas los campos o atributos que esta posee, y debajo de los indicadores la forma en que lo mismos se calculan. El resultado obtenido fue el siguiente.

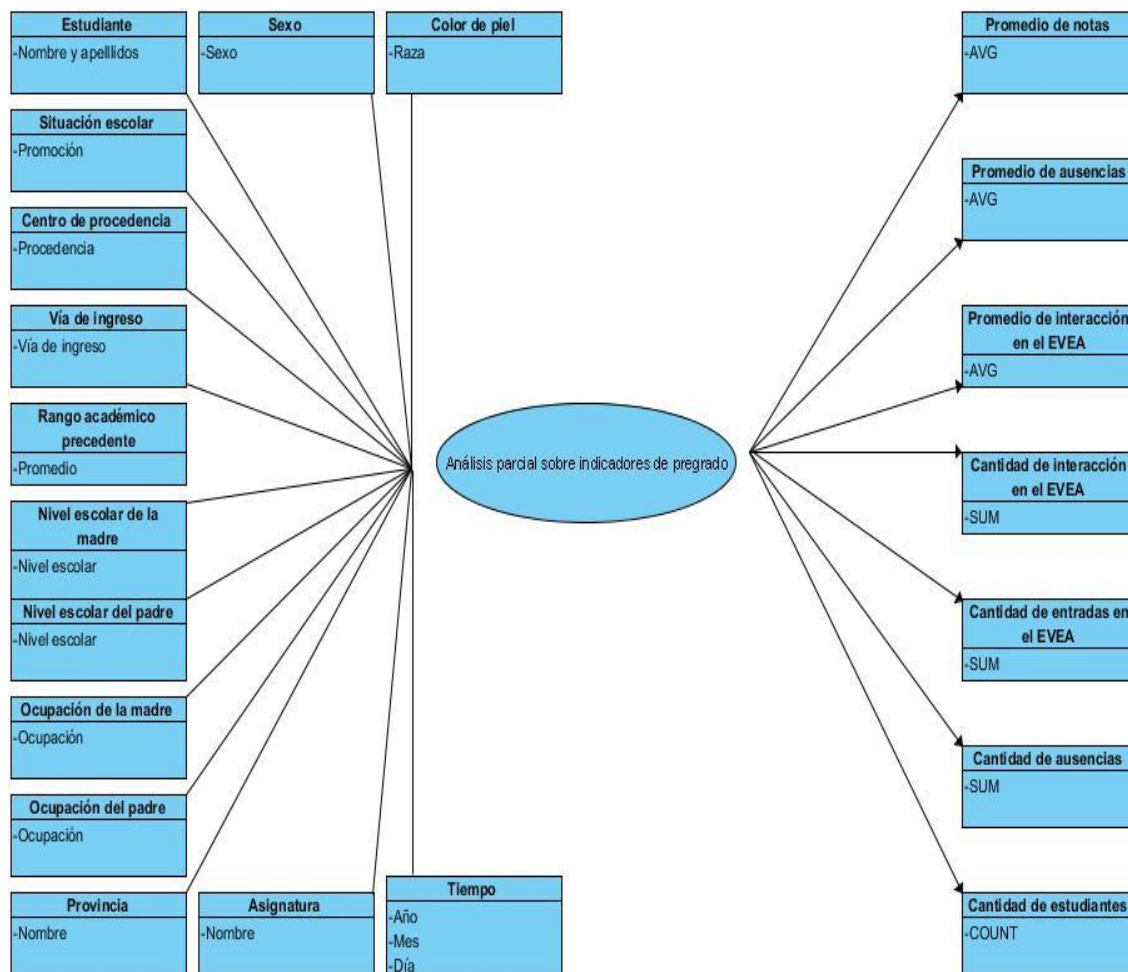


Figura 7. Modelo conceptual ampliado.

2.2.3 Fase 3: Modelo lógico del DM.

En esta fase primeramente se define el tipo de modelo lógico de la estructura del Data Mart, partiendo del modelo conceptual ampliado creado en la fase anterior. Una vez hecho esto se llevan a cabo las acciones propias del caso para diseñar las tablas de dimensiones y de hechos, y por último se definen las uniones correspondientes entre estas tablas.

Tipo de modelo lógico del DM.

Teniendo en cuenta el modelo conceptual ampliado obtenido en la fase anterior, y luego de estudiar los diferentes tipos de esquemas existentes para el modelado de las bases de datos multidimensionales, se define a utilizar para el diseño de la estructura lógica del Data Mart el esquema estrella, ya que es en este caso el que más se adapta a la situación existente. Las principales características y ventajas de este modelo se encuentran explicadas en el capítulo anterior.

Tablas de dimensiones.

Las dimensiones no son más que las características de un hecho, que permiten su posterior análisis en el proceso de toma de decisiones.

Una vez definido el tipo de esquema a utilizar en el diseño del DM se procede a diseñar las tablas de dimensiones que contendrá el depósito. Cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión.

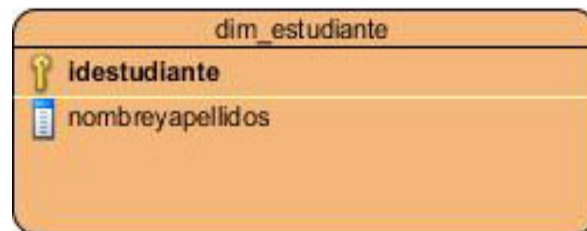


Figura 8. Dimensión dim_estudiante.

Todas las tablas de dimensiones definidas en el presente trabajo se pueden visualizar en el **Anexo #3: Tablas de dimensiones.**

Tablas de hechos.

Las tablas de hechos contienen las dimensiones y las medidas asociadas a estos. Las medidas son valores de datos numéricos que serán analizados por los usuarios.

En este paso se definen las tablas de hechos que contendrá el DM, las cuales contendrán los hechos que permitirán construir los indicadores de estudio.



Figura 9. Hecho cub_indicadorespregrado.

En el presente trabajo se definió una sola tabla de hechos denominada cub_indicadorespregrado y se definieron 14 tablas de dimensiones, ya que estas responden a las necesidades planteadas por el cliente.

Uniones entre dimensiones y hechos.

Una vez diseñadas y creadas las tablas de dimensiones y de hechos, se pasa a realizar las uniones correspondientes entre estas, definiéndose claramente las relaciones existentes entre las mismas. Con esta unión se obtiene el modelo lógico del DM, el cual se muestra a continuación.

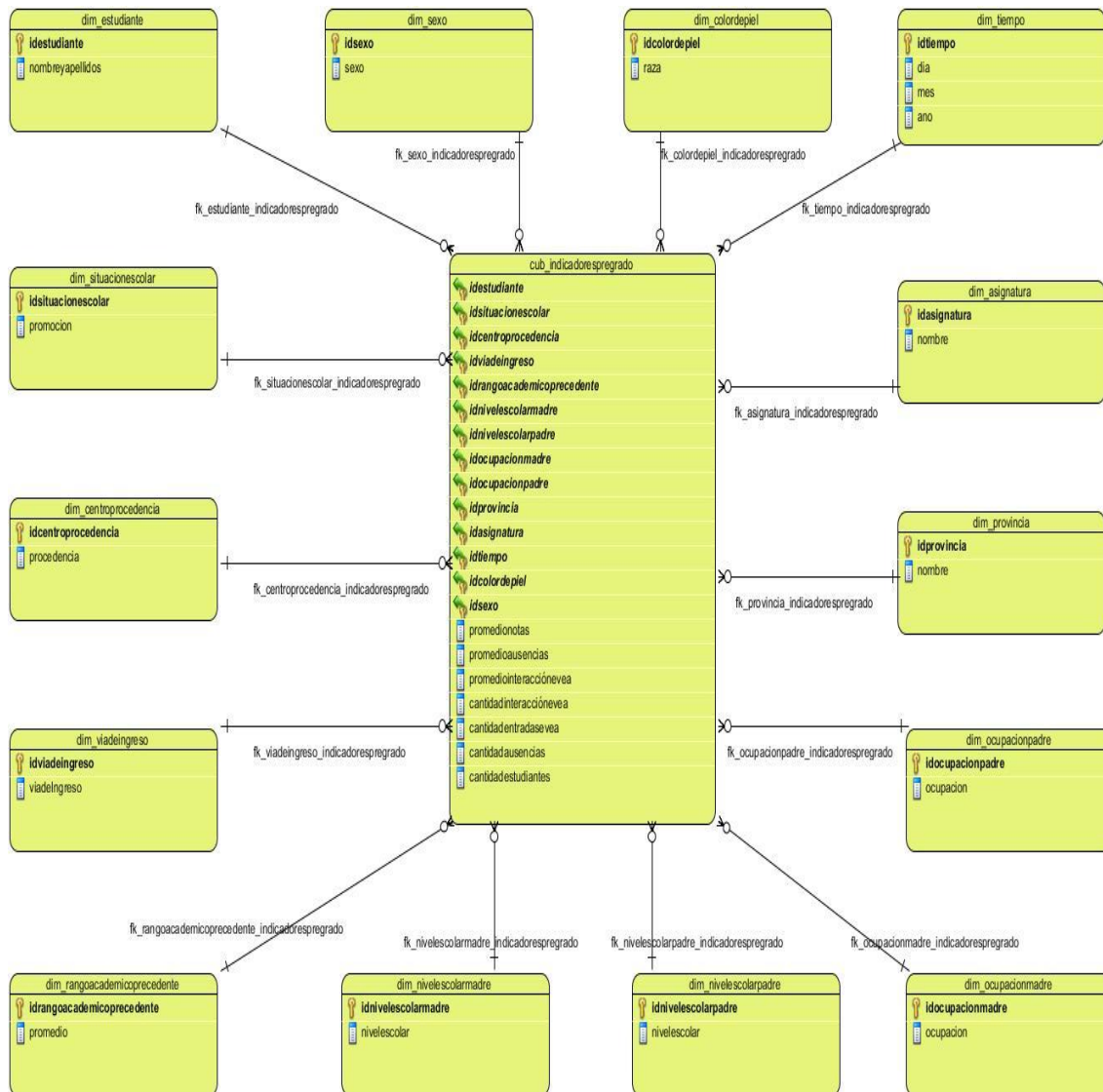


Figura 10. Modelo lógico del Data Mart.

Luego de obtenido el modelo lógico del DM, se construye el modelo físico del mismo, en el cual se especifican los tipos de datos de los diferentes campos del mercado. La herramienta Visual Paradigm permite exportar este modelo físico a una base de datos, con lo cual se obtiene la base de datos del mercado. A continuación se muestra el modelo físico obtenido.

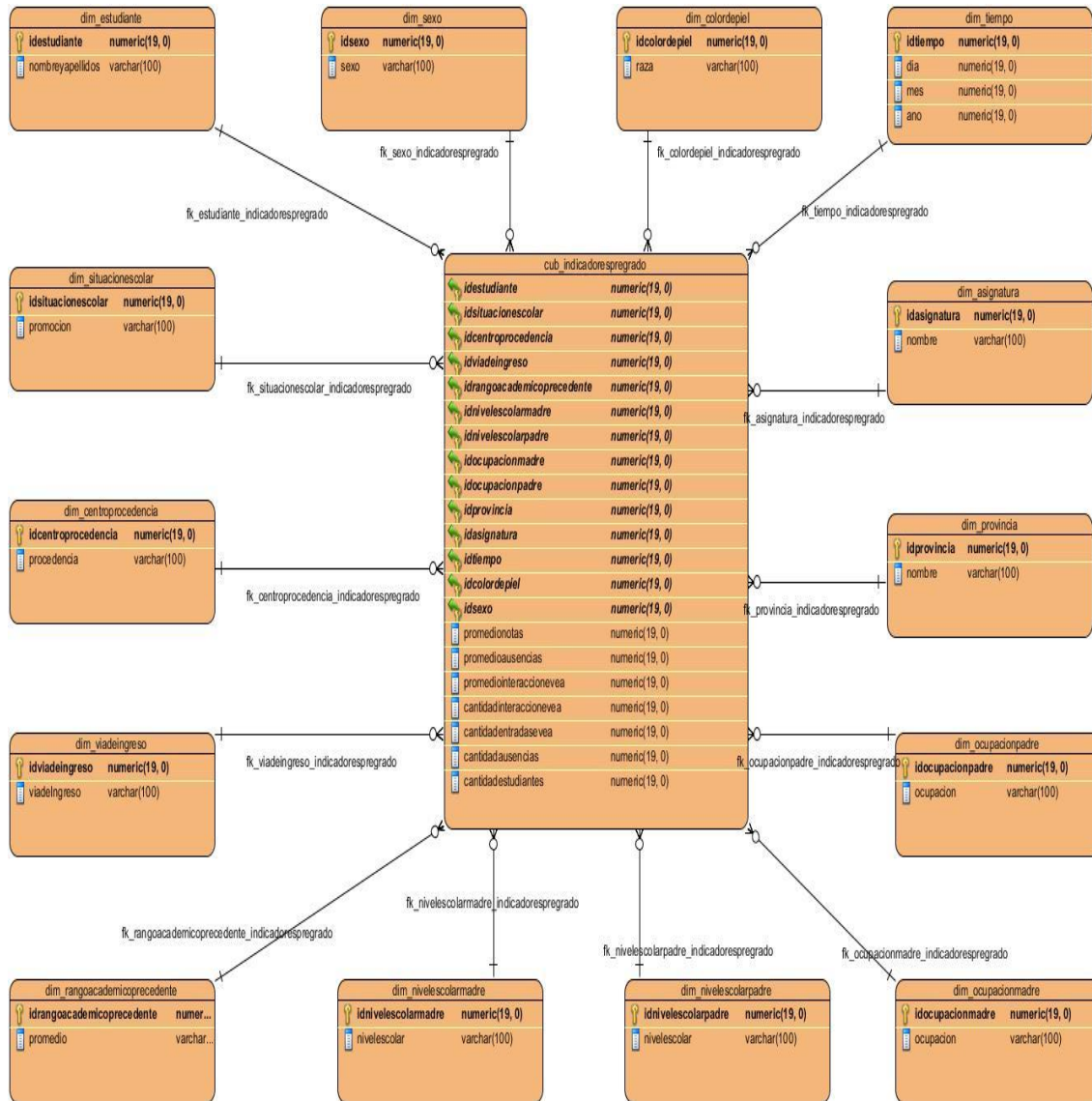


Figura 11. Modelo físico del Data Mart.

2.2.4 Fase 4: Procesos ETL.

Primeramente, se define de donde salen y hacia donde se insertarán, y si necesitan alguna restricción o condiciones para su entendimiento y cálculo. Lo siguiente, luego de tener establecidas las transformaciones (se realiza una transformación para cada dimensión y hecho definido), es la carga incremental de los datos hacia estos, para concluir con el diseño y la construcción de la automatización del sistema ETL.

Carga inicial y actualización:

En esta fase luego de tener el modelo físico del Data Mart con las dimensiones y hechos que se definieron se procede a incluirle los datos mediante la extracción, transformación y carga de los

mismos. Para llevar a cabo la carga inicial se emplea la herramienta Pentaho Data Integration descrita en el capítulo anterior, donde se realiza una serie de pasos para poblar de datos el Data Mart, iniciando este proceso a partir de la confección de las transformaciones. A continuación se describe la transformación Estudiante. Todas las transformaciones definidas en el presente trabajo se pueden visualizar en el **Anexo #2: Transformaciones**.

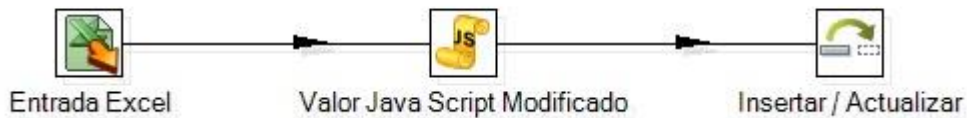


Figura 12. Transformación estudiante.

La transformación se inicia con el componente **Entrada Excel**, en el mismo se especifican de qué fuente se extraerán los datos y que columna se necesita de esa tabla.

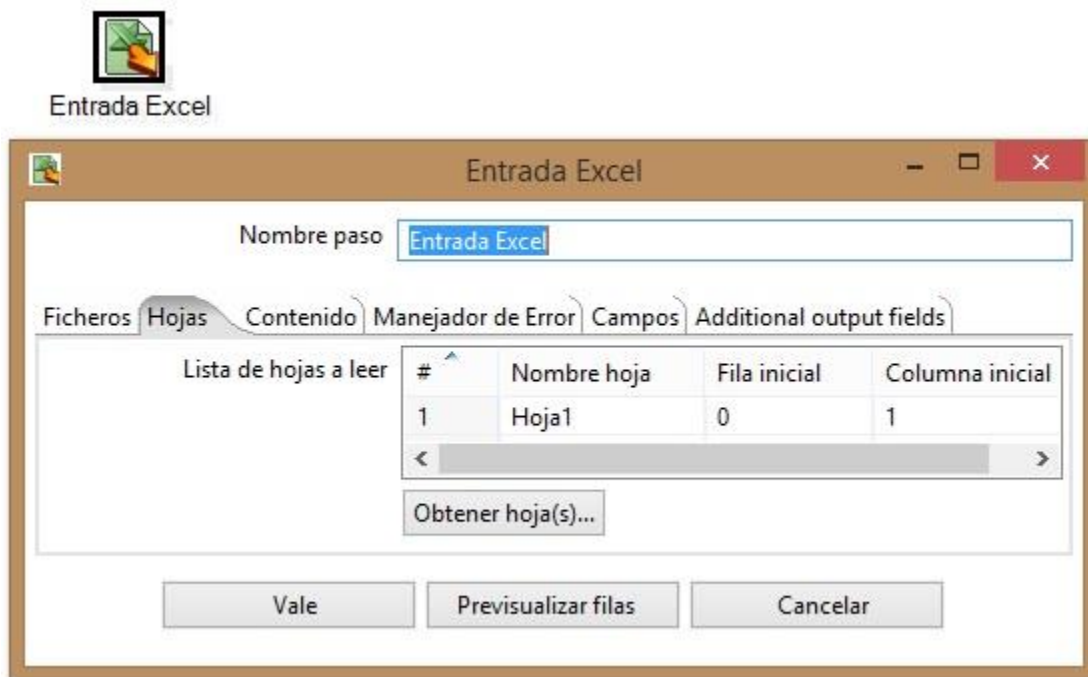


Figura 13. Componente Entrada Excel para la transformación Estudiante.

Luego se utiliza el componente **Valor Java Script Modificado**, este se utiliza para unir en una misma variable el nombre y apellidos del estudiante, ya que se definió que de esta forma sería como se insertaría en la dimensión dim_estudiante.

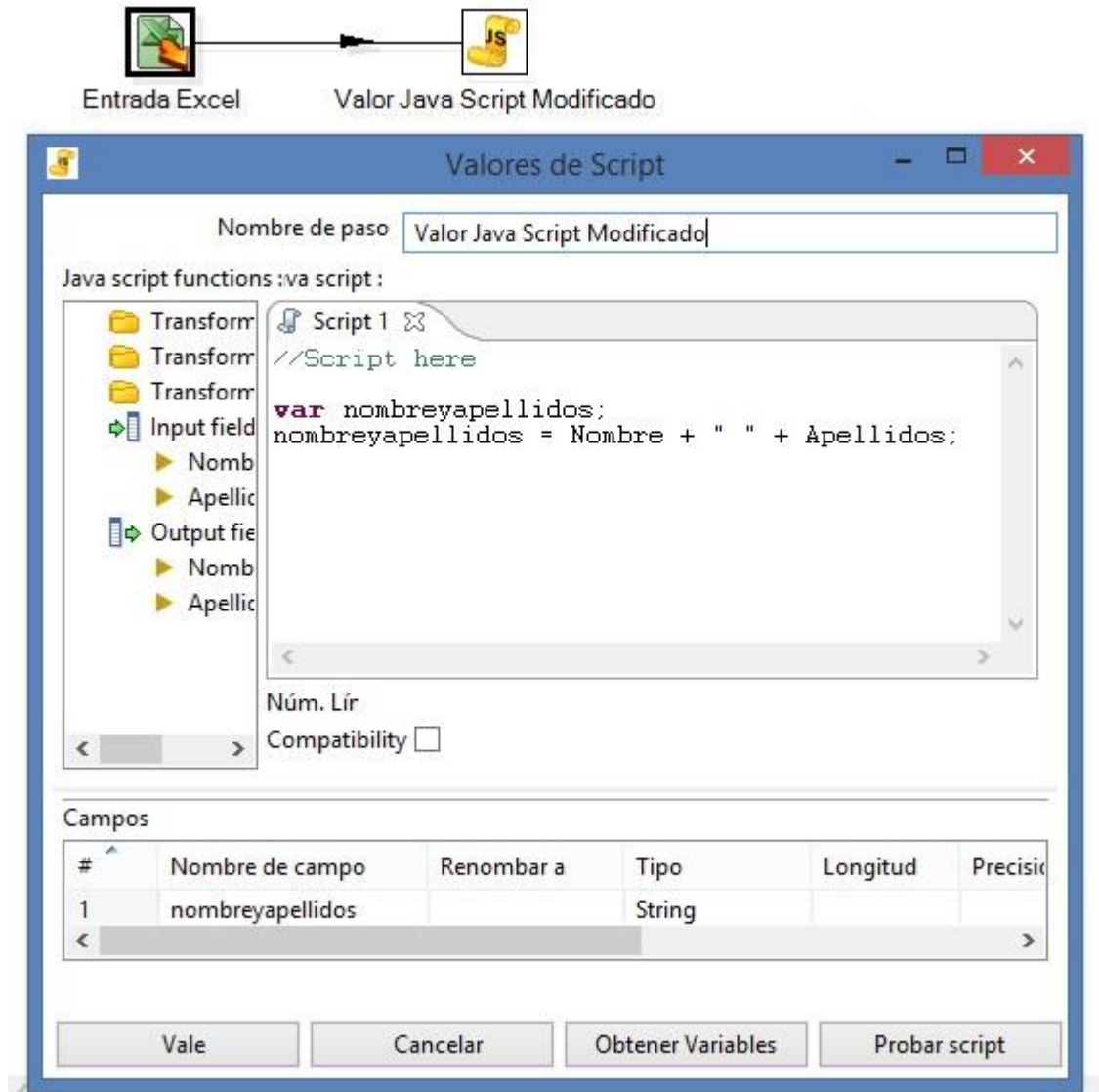


Figura 14. Componente Valor Java Script Modificado para la transformación Estudiante.

Concluye la transformación con el componente **Insertar/Actualizar**, el mismo permite insertar y/o actualizar en una tabla destino los valores de los campos seleccionados.

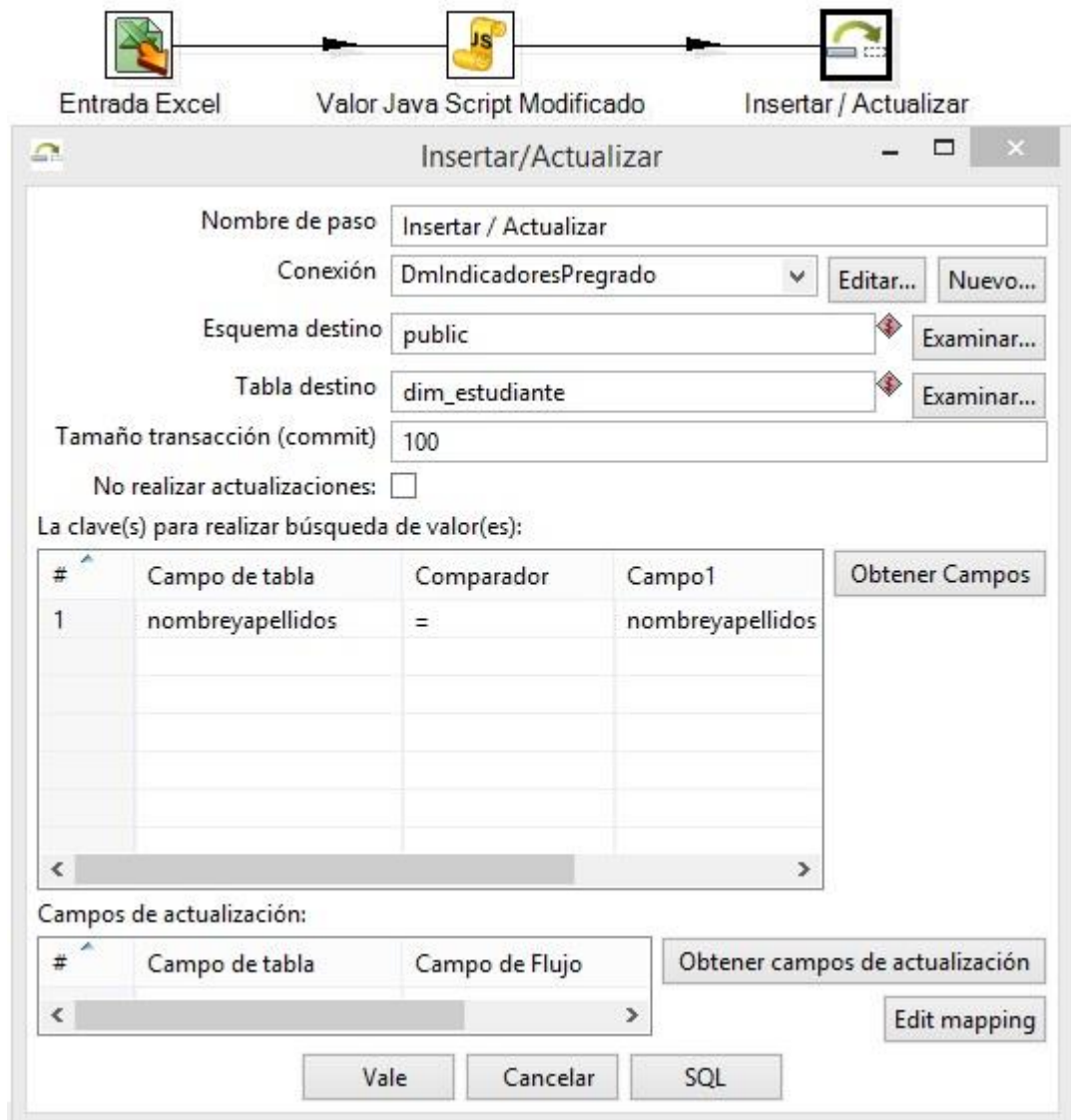


Figura 15. Componente Insertar/Actualizar para la transformación Estudiante.

Luego de concluida la realización de todas las transformaciones necesarias para la carga de los datos, se realiza la implementación del trabajo, este permitirá ejecutar todas las transformaciones que han sido diseñadas anteriormente en orden lógico, primero las dimensiones y luego los hechos. Al ejecutar el trabajo se ejecutan cada una de las transformaciones en el orden definido.

El Trabajo está compuesto por un componente **START** que da inicio a la ejecución del mismo, luego le siguen 15 componentes de **Transformaciones** que hacen referencia a las dimensiones del Data Mart, asociadas a cada una de estas transformaciones se encuentra el componente **Abort job**, para indicar si falla alguna de las transformaciones, y para finalizar se utiliza el componente **Success** para indicar que el trabajo se ejecutó correctamente.

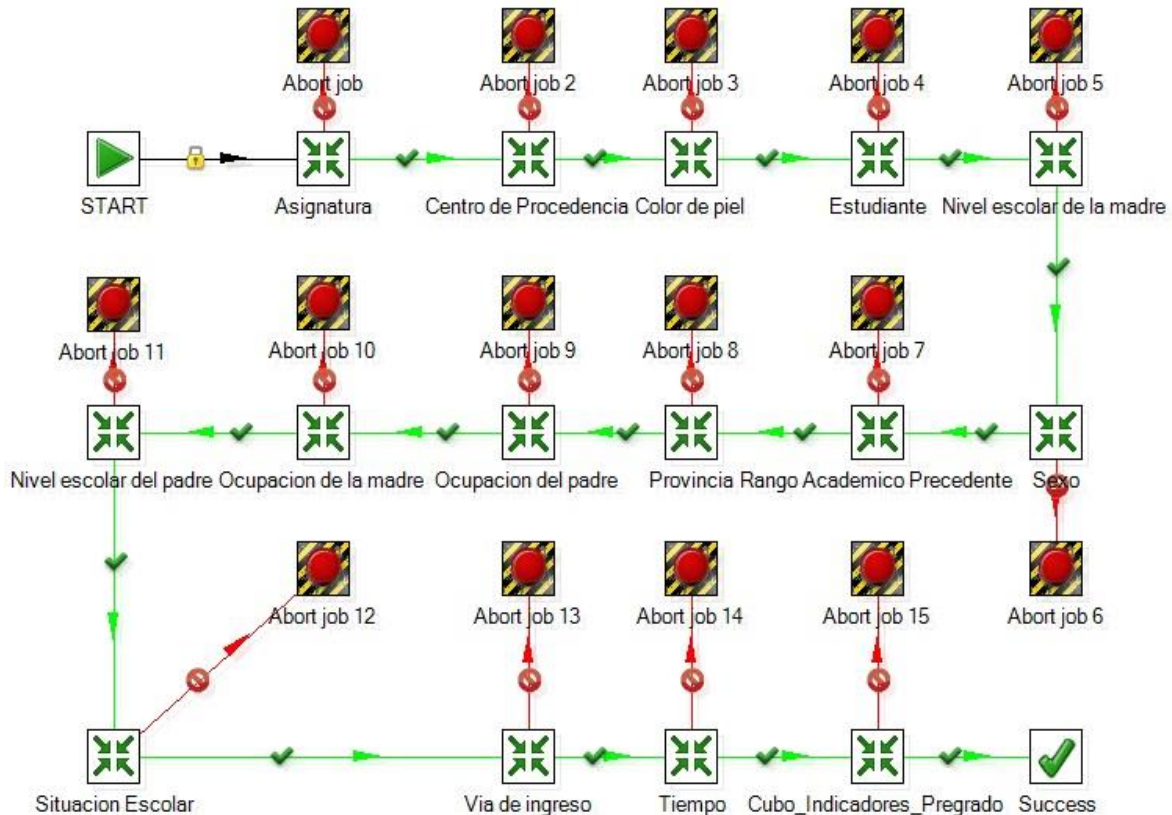


Figura 16. Trabajo para cargar las dimensiones y hechos del DM.

2.3 Requisitos no funcionales.

Estos requisitos describen las propiedades y cualidades que debe tener la solución. Representan las características del producto. Estos requisitos se clasifican en un grupo de categorías que suelen ser comunes para todos los desarrollos de software siempre dependiendo de las características del negocio.

Los requisitos no funcionales mínimos definidos para la propuesta de solución de la presente investigación son los siguientes:

Usabilidad:

RnF1 Las interfaces del sistema deben estar en idioma español, también deben ser intuitivas permitiendo el entendimiento por parte del usuario.

Confiabilidad:

RnF2 Asegurar la disponibilidad del sistema. El sistema debe estar disponible todo el tiempo, en caso de fallo, la recuperación del servicio no deberá exceder las 48 horas.

RnF3 Garantizar la persistencia de la información. Se debe realizar un respaldo total de los datos del MD con una frecuencia anual.

Restricciones de diseño:

RnF4 Utilizar las herramientas definidas durante la investigación:

- Como metodología de desarrollo, Hefesto.
- Como gestor de BD, PostgreSQL.
- Como herramienta de modelado, Visual Paradigm.
- Como herramienta para los procesos ETL, Pentaho Data Integration.
- Como herramienta para el diseño de los cubos OLAP, Schema Workbench.
- Como visor OLAP, Saiku Analytics.

Interfaces de hardware:

RnF5 La máquina donde radicará el servidor del sistema debe tener como mínimo 512 MB de RAM, microprocesador Pentium IV 2.8 GHZ o superior. Además debe contar con al menos 10 GB de espacio disponible en disco duro.

Interfaces de software:

RnF6 Se requiere para el funcionamiento de la aplicación, disponer en el servidor de la Máquina Virtual de Java 7.0. Los usuarios de la aplicación deberán contar con un navegador web, puede ser Internet Explorer versión 6 o superior, o Mozilla Firefox de su versión 10.0 en adelante.

2.4 Conclusiones parciales.

En este capítulo se realizó toda una iteración de la metodología Hefesto, logrando identificar como resultado de la entrevista con el cliente las necesidades de información del mismo, a partir de las preguntas claves definidas se obtuvieron los indicadores y perspectivas de análisis a tener en cuenta, a partir de estas se confeccionó el modelo conceptual del DM, proporcionando el alcance del mismo. Luego de confeccionado este modelo y luego de un análisis de las fuentes de datos se confeccionó el modelo lógico de la estructura del DM, a partir de la cual se diseñaron y llevaron a cabo los procesos de extracción transformación y carga de los datos. Obteniéndose como resultado final el Data Mart ya poblado.

Capítulo 3: Visualización y validación.

3.1 Introducción.

En el presente capítulo se describe el proceso de creación de los cubos OLAP, en los cuales se definen las dimensiones, los niveles y jerarquías de las dimensiones, las medidas físicas y las medidas calculables. Se lleva a cabo la visualización de los datos del DM a través de una herramienta que permite el análisis de los mismos mediante tablas y gráficos. Concluyendo con la aplicación de pruebas al mercado de datos para valorar el rendimiento del mismo así como para valorar también la satisfacción del cliente con el resultado obtenido.

3.2 Diseño de los cubos OLAP.

El diseño de los cubos de información se realiza utilizando la herramienta Pentaho Schema Workbench que permite crear visualmente y probar esquemas de cubos OLAP, como se explica en el Capítulo 1. Esta permite generar un fichero de configuración “.XML”, en el cual se definen los cubos, medidas, tablas y dimensiones, y para estas últimas la jerarquía que presenta, la tabla definida para la jerarquía y niveles. Las dimensiones se crearán fuera de los cubos, lo que posibilita que varios cubos puedan utilizar las mismas dimensiones.

En el presente trabajo se modeló 1 cubo, con las características correspondientes a la tabla de hechos y tablas de dimensiones definidas. A continuación se muestra un ejemplo de este diseño.

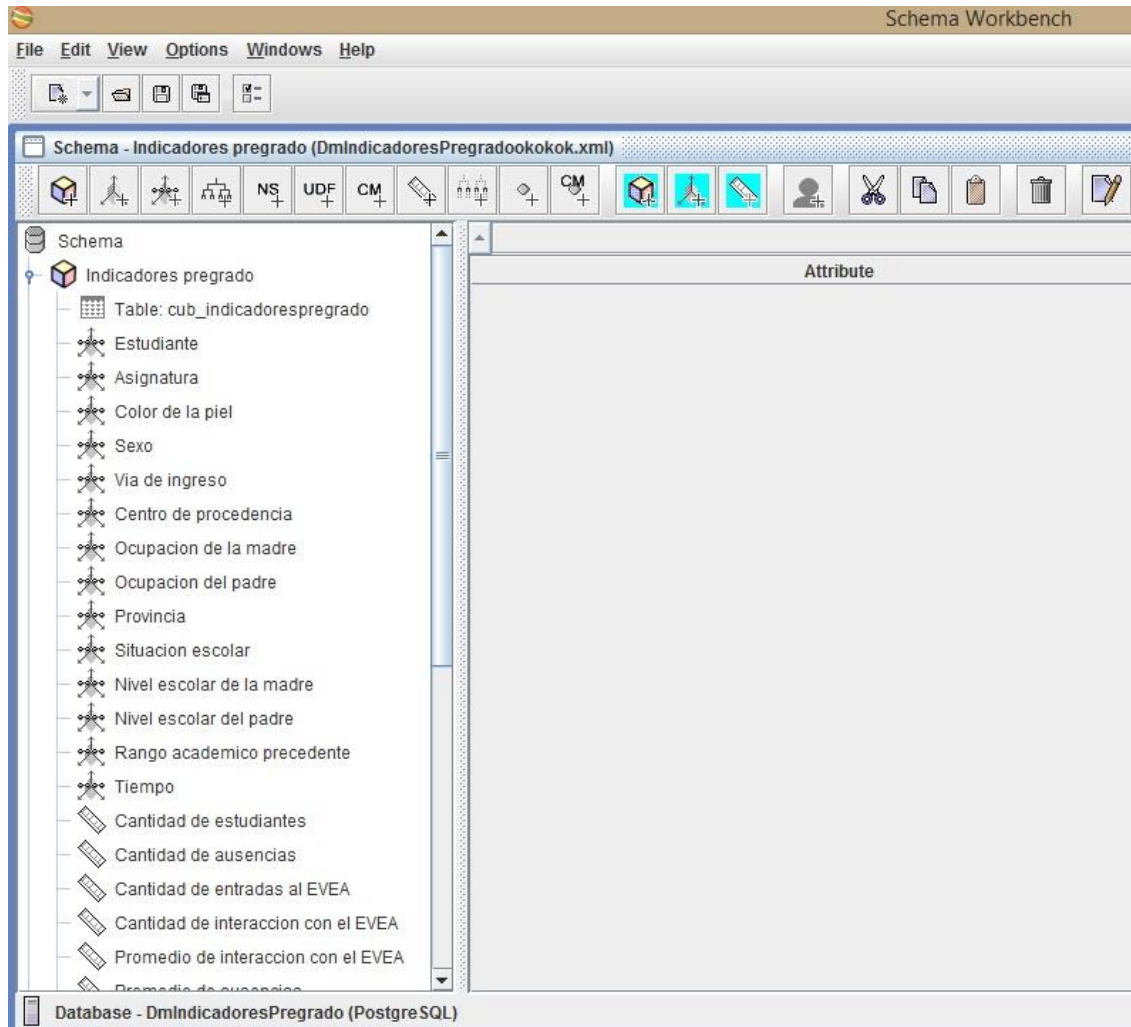


Figura 17. Diseño de los cubos OLAP en la herramienta Schema Workbench.

3.3 Visualización.

El archivo XML resultante del trabajo con la herramienta Schema Workbench es publicado en la herramienta Saiku Analytics, descrito anteriormente en el Capítulo 1, la cual es usada para la visualización del mercado de datos creado. Este es un visor OLAP muy amigable e intuitivo que les permite a los usuarios crear sus propias vistas arrastrando y soltando campos además de que permite el análisis de los datos en forma de tabla o gráficos.

Estas tablas o gráficos son construidas como resultado de los datos arrojados por las consultas MDX que constituyen en los sistemas OLAP el equivalente a las consultas SQL en las BD relacionales, estas consultas son ejecutadas por el servidor OLAP Mondrian.

A continuación se muestra la interfaz de Saiku, disponible a través de la web, cuya página principal muestra todas las opciones que posee el usuario para confeccionar las vistas de los datos las cuales le servirán de base en el análisis de los mismos.

El ejemplo muestra la solución a una de las necesidades de información del cliente, conocer la cantidad de ausencias de los estudiantes, atendiendo a la provincia de origen y el sexo de los mismos.

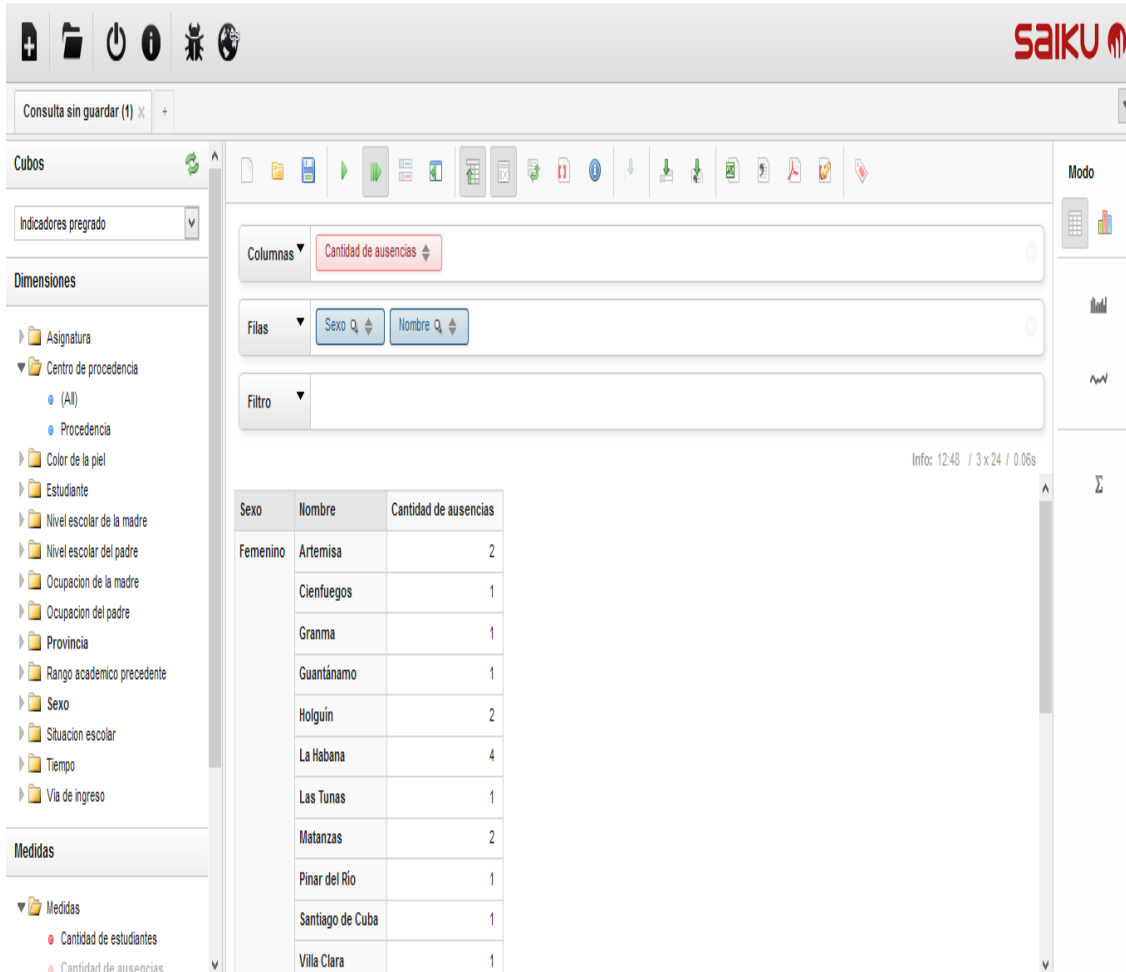


Figura 18. Ejemplo de reporte en forma de tabla brindado por Saiku.

Como se puede observar la información anterior fue brindada en forma de tabla, la siguiente figura muestra también esta misma información, pero en forma de gráfica, para de esta forma brindarle al usuario otra perspectiva de análisis de los datos, con el objetivo de lograr un mayor entendimiento de los mismos.

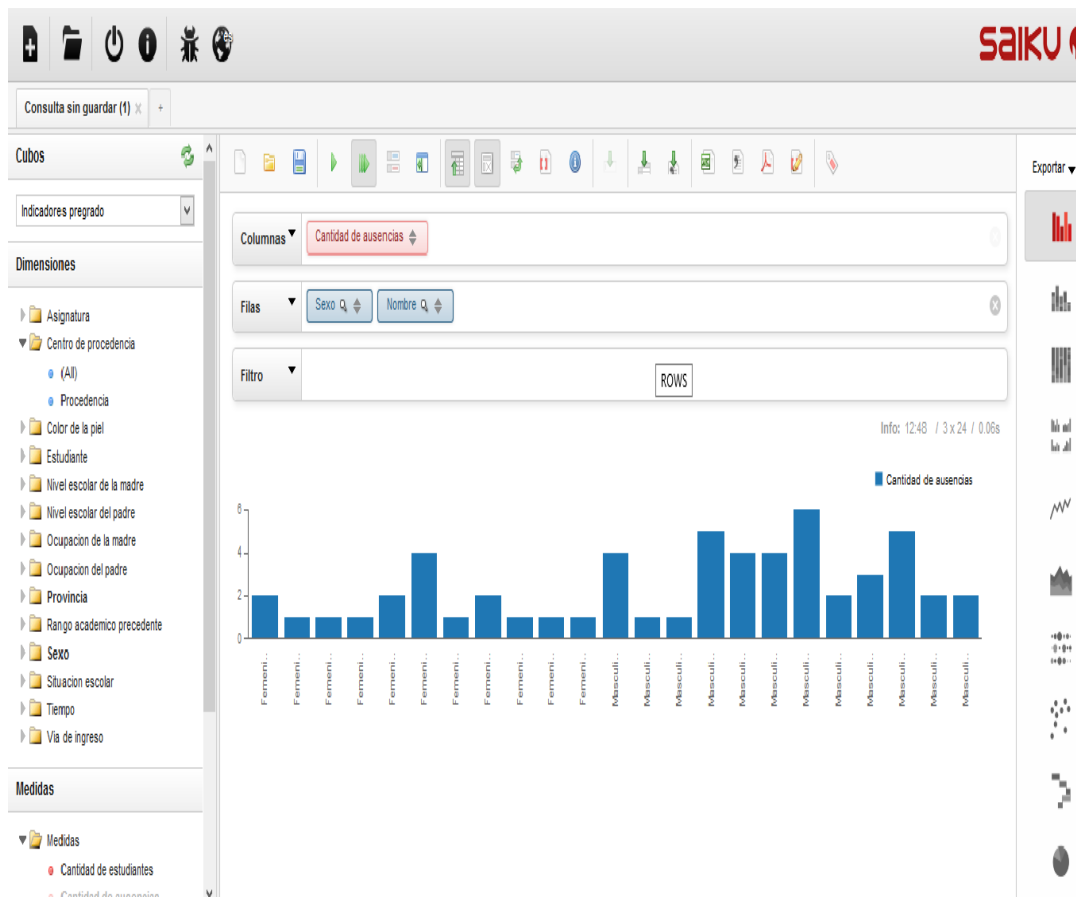


Figura 19. Ejemplo de reporte en forma de gráfica brindado por Saiku.

Para visualizar los demás reportes correspondientes a las necesidades de información definidas por el cliente, dirigirse al **Anexo #5: Reportes**.

3.4 Validación de la solución.

Una vez creado, poblado y visualizado el mercado de datos se procede a la validación del mismo mediante pruebas que verifiquen que se cumplan las necesidades del cliente y que el producto cuente con la calidad requerida.

3.4.1 Prueba de aceptación.

Luego de concluido el mercado de datos se hace necesario que el cliente apruebe el sistema de acuerdo con los requisitos planteados por el mismo y validar su conformidad con el producto.

La prueba de aceptación fue realizada por la Ing. Lourdes Escalona Peral (Cliente), quien confirmó que el Mercado de Datos para la toma de decisiones a partir de la información de los

estudiantes generada por el Sistema de Gestión Universitaria y el Entorno Virtual de Enseñanza Aprendizaje de la Universidad de las Ciencias Informáticas satisface todas las necesidades de información identificadas con anterioridad. Dichos resultados quedaron oficializados con la carta de aceptación del cliente la cual se puede ver en **Anexo #2: Carta de aceptación del cliente.**

3.4.2 Pruebas de carga y estrés.

Estas pruebas se realizan para saber el rendimiento y el tiempo de respuesta a peticiones, que se le envían al mercado de datos desarrollado, con el objetivo de determinar la velocidad y eficiencia del mismo.

Para la realización de estas pruebas se utilizó Apache Jmeter, una herramienta Java de código abierto diseñada para cargar pruebas de comportamiento funcional y medir el rendimiento (39).

Para evaluar los resultados se utilizó uno de los componentes provistos por la herramienta, el Summary Report. Este componente permite medir el rendimiento de una aplicación en términos de tiempos de respuesta y porcentaje de errores respecto al número total de peticiones.

Los datos más significativos que presenta el mismo son (40):

- Label: etiqueta de la muestra.
- #Muestras: cantidad de requerimientos utilizados.
- Media: tiempo promedio en milisegundos para un conjunto de resultados.
- Min: tiempo mínimo que demora en responder a un requerimiento.
- %Error: porcentaje de requerimientos con errores.
- Max: tiempo máximo que demora en responder a un requerimiento.

Pruebas de carga: Esta sub-categoría de pruebas de rendimiento se centra en determinar o validar las características de rendimiento del sistema o la aplicación cuando es sometido a cargas de trabajo y los volúmenes de carga prevista para la operación normal en producción (41).

Para llevar a cabo esta prueba se configuró la herramienta Jmeter simulando 200 usuarios conectados al mismo tiempo. Se implementó una consulta de complejidad alta, esta consulta resuelve uno de los requerimientos de información del usuario, la misma se realiza sobre 700 tuplas en la base de datos. Los resultados arrojados fueron los siguientes:

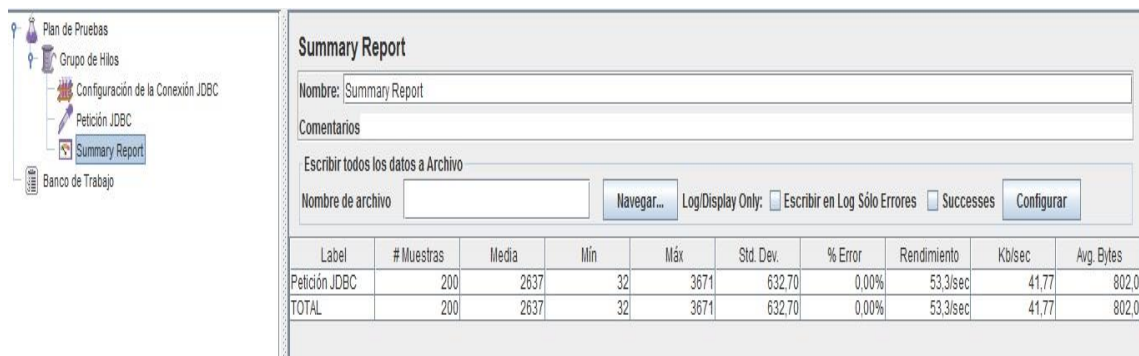


Figura 20. Resultados para 200 usuarios conectados simultáneamente.

Como se puede observar en los resultados arrojados por la herramienta, los tiempos de respuestas son adecuados, obteniéndose un tiempo medio de respuesta a peticiones de 2,6 segundos con un 0 % de requerimientos con error. Esto para un acceso de usuarios esperado, en tiempo real de explotación de la aplicación.

Pruebas de estrés: Pruebas centradas en validar características de rendimiento del sistema bajo condiciones superiores a lo previsto en las operaciones de producción (grandes volúmenes de carga) (41).

Para llevar a cabo esta prueba se configuró la herramienta Jmeter simulando 1000 usuarios conectados al mismo tiempo. Se implementó una consulta de complejidad alta, esta consulta resuelve uno de los requerimientos de información del usuario, la misma se realiza sobre 700 tuplas en la base de datos. Los resultados arrojados fueron los siguientes:

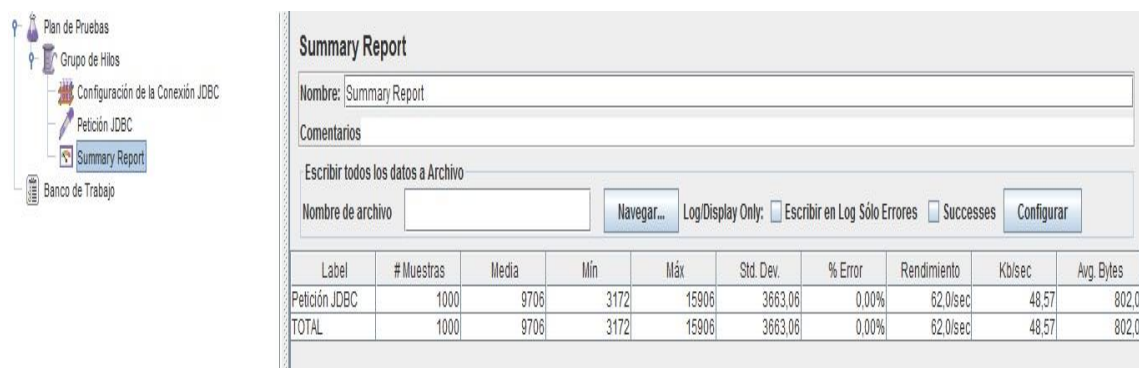


Figura 21. Resultados para 1000 usuarios conectados simultáneamente.

Como se puede observar, los tiempos de respuesta de la aplicación en este escenario de prueba son aceptables, aunque el sistema se encuentra bajo condiciones superiores a lo previsto en tiempo real de explotación. El porcentaje de error fue de 0 y el tiempo medio de respuesta a peticiones fue de 9,7 segundos.

3.5 Conclusiones parciales.

En este capítulo se llevó a cabo la confección de los cubos de información mediante la herramienta Schema Workbench, luego el archivo DmIndicadoresPregrado.xml resultante del trabajo con esta herramienta se publicó en Saiku Analytics (visor OLAP) encargado de llevar a cabo la visualización del mercado de datos creado.

Una vez terminado este proceso se llevó se validó la solución propuesta mediante una serie de pruebas enfocadas a medir la calidad del mercado de datos y la satisfacción del cliente, para ello se realizaron pruebas de aceptación, carga y estrés.

Conclusiones Generales

Luego de concluida la investigación, se obtiene como resultado el producto de software Mercado de Datos para la toma de decisiones a partir de la información de los estudiantes generada por el Sistema de Gestión Universitaria y el Entorno Virtual de Enseñanza Aprendizaje de la Universidad de las Ciencias Informáticas, y se arriban a las siguientes conclusiones:

- Se realizó un estudio de la aplicación de los almacenes de datos y mercados de datos como herramientas de apoyo a la toma de decisiones, concluyéndose que las aplicaciones analizadas son específicas para las áreas que fueron concebidas y que en la universidad no existe un mercado de datos que vincule el entorno EVEA con el SGU.
- El análisis de las metodologías y tecnologías para el desarrollo de los mercados de datos permitió identificar a Hefesto como una de metodologías más ágiles y fáciles de usar para personas con poca experiencia en el tema.
- El análisis de los requerimientos del cliente posibilitó la obtención del modelo conceptual del mercado de datos con la información del EVEA y el SGU a partir de la definición de los indicadores y sus perspectivas de análisis.
- Se realizó el diseño e implementación de la solución utilizando como guía la metodología seleccionada la cual permitió a través de sus fases la construcción del mercado de datos de manera ágil e intuitiva. Se tuvo como limitante que la fuente de información del SGU no fue completa, ya que solo se contó con un reporte parcial de la información en formato Excel.
- El diseño del cubo de información en la herramienta Schema Workbench, permitió la visualización de los datos deseados por el cliente mediante gráficos y tablas, haciendo uso de la herramienta Saiku.
- Se realizaron pruebas de carga y estrés a la solución propuesta utilizando la herramienta JMeter, determinándose que la aplicación brinda tiempos de respuesta satisfactorios ante las peticiones realizadas.
- Las pruebas de aceptación le permitió al cliente verificar el cumplimiento de los objetivos para los que fue creado el mercado de datos comprobándose que los reportes emitidos constituyen una herramienta válida para el apoyo a la toma de decisiones de profesores y directivos.

Recomendaciones

A partir del análisis de los resultados alcanzados con la presente investigación se recomienda:

- Ampliar las funcionalidades del mercado de datos teniendo como fuente de información la base de datos del Sistema de Gestión Universitaria.
- Actualizar con frecuencia las fuentes de información procedentes del EVA y el SGU para una mayor explotación del mercado de datos.
- Disponer de un servidor en la facultad que permita la publicación del mercado de datos y de esta forma brindar a profesores y directivos la posibilidad de acceder a los reportes.

Referencias Bibliográficas

1. **Velasco, Roberto Hernando.** Almacenes de Datos(Data Warehouse). [En línea] 13 de diciembre de 2014. <http://www2.rhernando.net/modules/tutorials/doc/bd/dw.html>.
2. **Kimball, R.** *The Data warehouse Lifecycle Toolkit*. New York : Wiley, 2006.
3. **Inmon, Bill.** *Building the Data Warehouse*. s.l. : Wiley, 1992.
4. **Sinnexus Business intelligence** . [En línea] [Citado el: 28 de febrero de 2015.] http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
5. **Monografias.** [En línea] [Citado el: 20 de enero de 2015.] <http://www.monografias.com/trabajos75/inteligencia-negocios/inteligencia-negocios3.shtml>.
6. **Lucentia.** [En línea] [Citado el: 25 de febrero de 2015.] <http://www.lucentia.es/walt-disney-tambien-usa-big-data/#more-577>.
7. **Ing. Lorena Romina Matteo, Lic. Julio César Bossero.** *Utilización de técnicas de Data Warehouse para la toma de decisiones en el Área Académica*. Buenos Aires : s.n., 2011. No 946.157.
8. **D. L. LA RED MARTINEZ, J. C. ACOSTA, L. A. CUTRO, V. E. URIBE, A. R. RAMBO.** *Data WareHouse y Data Mining aplicados al estudio del rendimiento academico y de perfiles de alumnos*.
9. **Oficina de cooperacion universitaria.** *LIBRO BLANCO INTELIGENCIA INSTITUCIONAL EN UNIVERSIDADES*. Madrid : Oficina de Cooperación Universitaria, S.A. ISBN: 978-84-695-8892-5.
10. **Almacén de datos operacional para contribuir a** . Urquizu, Ramel Vitier. La Habana : Universidad de las Ciencias Informaticas, 2013.
11. **Blanco-Cuaresma, Sergi.** Metodologias de Desarrollo. [En línea] [Citado el: 25 de febrero de 2015.] <http://www.marblestation.com/?p=644>.
12. **Rivadera, Gustavo R.** La metodología de Kimball para el diseño de. [En línea] [Citado el: 25 de febrero de 2015.] <http://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>.
13. **Rodríguez, Sanz Miguel.** Análisis y diseño de un Data Mart para el seguimiento . [En línea] 2010. [Citado el: 25 de febrero de 2015.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.
14. **Bernabeu, R. Dario.** DataPrix Datawarehouse Manager. [En línea] 2009. [Citado el: 26 de febrero de 2015.] www.dataprix.net/gl/datawarehouse-manager-34#x1-550003.4.7.1..
15. **Hefesto Data Warehousing:Investigación y sistematización de.** Ricardo, Dario Bernabeu. 2010, Vol. 2.

16. Bases de datos OLTP y OLAP. [En línea] [Citado el: 27 de febrero de 2015.]
[.http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx).
17. *Diseño e implementación de un mercado de datos para el análisis de eventos de*. Angel Rafael González Alvarez, Osmari Hechevarría Delis. s.l. : Universidad de las Ciencias Informaticas, 2011.
18. [En línea] [Citado el: 28 de febrero de 2015.]
http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
19. Molier, Francisco Javier Lopez. *Grupo A y B Informatica :Bloque específico*. Valencia : S.L. EDITORIAL MAD, 2005. ISBN 84-665-2080-5.
20. Kimball, Ralph. *El Juego de Herramientas del Almacén de Datos*.
21. Okamura, Hernan Calle. Sistemas gestores de base de datos. [En línea] 2012. [Citado el: 28 de febrero de 2015.] <http://es.slideshare.net/HernanOkamura/sistemas-de-gestores-de-base-de-datos-13332504>.
22. Velasco, Roberto Hernando. El SGBDR Oracle. [En línea] [Citado el: 28 de febrero de 2015.]
<http://www2.rhernando.net/modules/tutorials/doc/bd/oracle.html>.
23. Sitio oficial de my SQL. [En línea] [Citado el: 28 de febrero de 2015.] <http://dev.mysql.com/>.
24. PostgreSQL-es. Sobre PostgreSQL. [En línea] 2 de octubre de 2010. [Citado el: 26 de febrero de 2015.] http://www.postgresql.org.es/sobre_postgresql.
25. postgresQL investigación. [En línea] [Citado el: 3 de marzo de 2015.]
<http://es.scribd.com/doc/36570462/postgresql-investigacion#scribd>.
26. Herramienta de diseño UML. [En línea] [Citado el: 6 de marzo de 2015.]
<http://www.sparxsystems.com.ar/products/ea.html>.
27. Visual Paradigm. [En línea] [Citado el: 7 de marzo de 2015.] www.visual-paradigm.com/.
28. The Passion Behind CloverETL. [En línea] [Citado el: 4 de marzo de 2015.]
<http://www.cloveretl.com/about>.
29. Clover.ETL. [En línea] [Citado el: 4 de marzo de 2015.]
<http://macloc.blogspot.com/2008/10/cloveretl.html> .
30. Enhydra Octopus 3.2-2. [En línea] [Citado el: 4 de marzo de 2015.]
<http://www.softpedia.com/get/Internet/Servers/Database-Utills/Enhydra-Octopus.shtml>.
31. *Construcción de un Data Mart para el Sistema Integral de Gestión de Medicamentos*. Aguilar, Yuniesky Nueva. La Habana : Universidad de las Ciencias Informaticas, 2011.
32. Características de Pentaho. [En línea] [Citado el: 5 de marzo de 2015.]
<http://www.dataprix.com/723-caracter-sticas-pentaho>.
33. *Data Mart para la toma de decisiones referente a las Reacciones Adversas a Medicamentos, en el Ministerio de Salud Pública, desde el producto Synta*. Aliandy Sosa Bello, Dennier Salas Lóriga. La Habana : Universidad de las Ciencias Informaticas, 2013.

34. Pentaho Data Integration. [En línea] [Citado el: 5 de marzo de 2015.] <https://sites.google.com/site/pentahobisuite/home/caracteristicas/pentaho-dashboards/pentaho-data-integration>.
35. *El uso de la plataforma de inteligencia de negocio Pentaho en el modulo generador de reportes del sistema GREHUCORP*. Lisbany Labori Días, Carlos Ramón López Paz. La Habana : s.n., 2008.
36. Espinosa, Roberto. El Rincon del BI. [En línea] 4 de julio de 2010. [Citado el: 20 de febrero de 2015.] <https://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>.
37. SAIKU. SAIKU. [En línea] 31 de marzo de 2015. [Citado el: 2 de mayo de 2015.] <http://wiki.meteorite.bi/display/SAIK/Saiku>.
38. Apache tomcat. [En línea] [Citado el: 20 de marzo de 2015.] <http://apachefoundation.wikispaces.com/Apache+Tomcat>.
39. The apache software foundation . Apache jmeter. [En línea] 17 de abril de 2015. [Citado el: 15 de mayo de 2015.] <http://jmeter.apache.org/>.
40. F.Javier Diaz, Claudia M. Tzancoff Banchoff, Anahí S. Rodríguez, Valeria Soria. Usando Jmeter para pruebas de rendimiento. [En línea] [Citado el: 17 de mayo de 2015.] http://www.linti.unlp.edu.ar/uploads/docs/usando_jmeter_para_pruebas_de_rendimiento.pdf.
41. Globe testing. [En línea] [Citado el: 15 de mayo de 2015.] <http://www.globetesting.com/pruebas-de-rendimiento/>.

Anexos

Anexo #1: Necesidades de información del cliente.

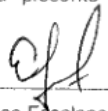
En el presente documento se exponen los requisitos de información definidos para el **Mercado de Datos para la toma de decisiones a partir de la información de los estudiantes generada por el Sistema de Gestión Universitaria y el Entorno Virtual de Enseñanza Aprendizaje de la Universidad de las Ciencias Informáticas**, el cual se desarrolla para apoyar la toma de decisiones de directivos y profesores en el proceso de enseñanza aprendizaje en la Universidad de las Ciencias Informáticas.


Requisitos de información:

- 1- Se desea conocer el promedio general de los estudiantes por centro de procedencia, raza y sexo.
- 2- Se desea conocer el promedio general de los estudiantes por provincia, grado de escolaridad del padre y grado de escolaridad de la madre.
- 3- Se desea conocer el promedio general de los estudiantes por provincia, ocupación de la madre y ocupación del padre.
- 4- Se desea conocer el promedio general de los estudiantes por centro de procedencia y promedio de procedencia.
- 5- Se desea conocer la cantidad de ausencias de los estudiantes por situación escolar, vía de ingreso y provincia, en un tiempo determinado.
- 6- Es de interés saber el promedio de ausencias de los estudiantes por asignatura y sexo en un tiempo determinado.
- 7- Es de interés saber el promedio de ausencias de los estudiantes por asignatura, por provincia y centro de procedencia, en un tiempo determinado.
- 8- Se quiere saber el promedio de interacción en el EVEA de los estudiantes por sexo y raza.

Para hacer constar la validez de la presente información firman los siguientes involucrados:


Lisvet Marrero Pérez
Autora


Ing. Lourdes Escalona Peral
Cliente


Pedro Pablo Ramos Pajón
Autor

9- Se quiere saber el promedio de interacción en el EVEA de los estudiantes por situación escolar, en un tiempo determinado.

10- Se desea conocer el tiempo de interacción en el EVEA de los estudiantes por asignatura, en un período determinado.

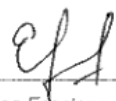
11- Se quiere saber la cantidad de entradas en el EVEA de los estudiantes por asignatura y provincia.


12- Es de interés conocer la cantidad de estudiantes matriculados en el EVEA por asignatura, centro de procedencia y vía de ingreso.

13- Es de interés conocer la cantidad de estudiantes matriculados en el EVEA por provincia, raza y sexo.

Para hacer constar la validez de la presente información firman los siguientes involucrados:


Lisvet Marrero Pérez
Autora


Ing. Lourdes Escaiona Peral
Cliente


Pedro Pablo Ramos Pajón
Autor

Anexo #2: Carta de aceptación del cliente.

La Habana 20 de mayo de 2015

Año 57 de la Revolución

Acta de aceptación:

Mediante la presente, certifico que los resultados obtenidos por los compañeros Lisvet Marrero Pérez y Pedro Pablo Ramos Pajón en la construcción del Mercado de Datos para la toma de decisiones a partir de la información de los estudiantes generada por el Sistema de Gestión Universitaria y el Entorno Virtual de Enseñanza Aprendizaje de la Universidad de las Ciencias Informáticas, están acordes a las funcionalidades que se esperaban del sistema.

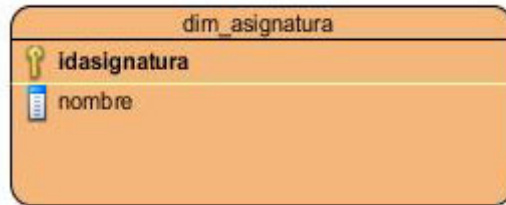
Los compañeros antes mencionados han demostrado tener un amplio dominio del tema y un alto sentido de responsabilidad en el cumplimiento de las tareas asignadas.

Para constancia de ello y en mi condición de profesora principal de quinto año de la facultad 2 de la Universidad de las Ciencias Informáticas firmo a los 20 días del mes de mayo del 2015.

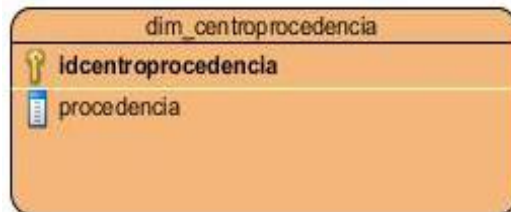


Ing. Lourdes Escalona Peral.

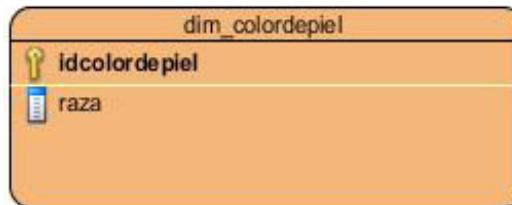
Anexo #3: Tablas de dimensiones.



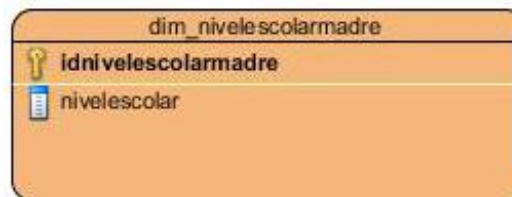
Dimensión Asignatura.



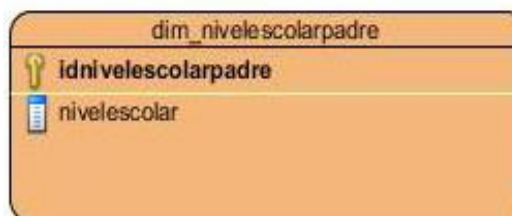
Dimensión Centro de procedencia.



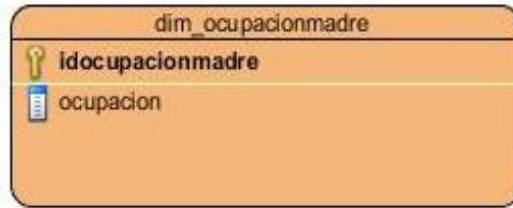
Dimensión Color de la piel.



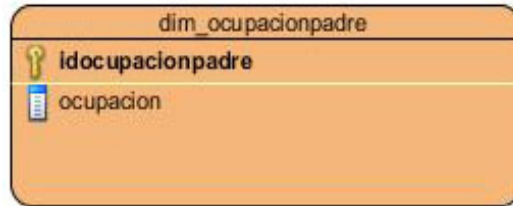
Dimensión Nivel escolar de la madre.



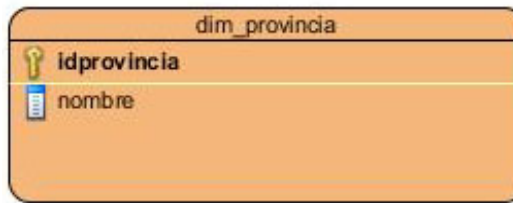
Dimensión Nivel escolar del padre.



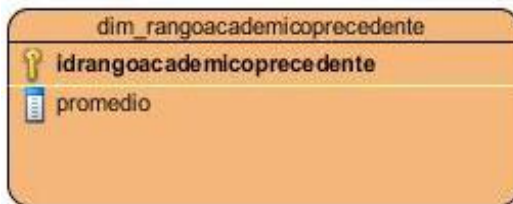
Dimensión Ocupación de la madre.



Dimensión Ocupación del padre.



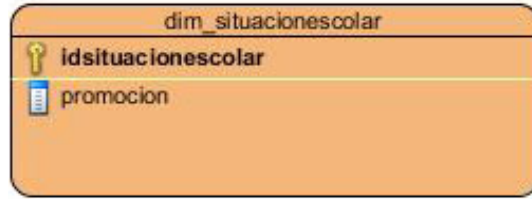
Dimensión Provincia.



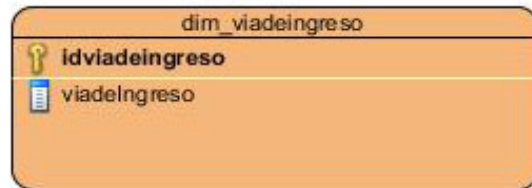
Dimensión Rango académico precedente.



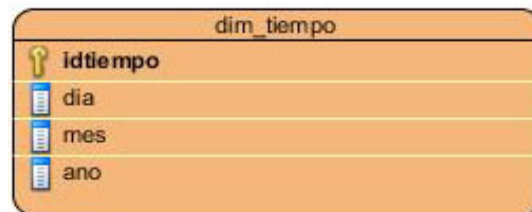
Dimensión Sexo.



Dimensión Situación Escolar.



Dimensión Vía de ingreso.



Dimensión Tiempo.

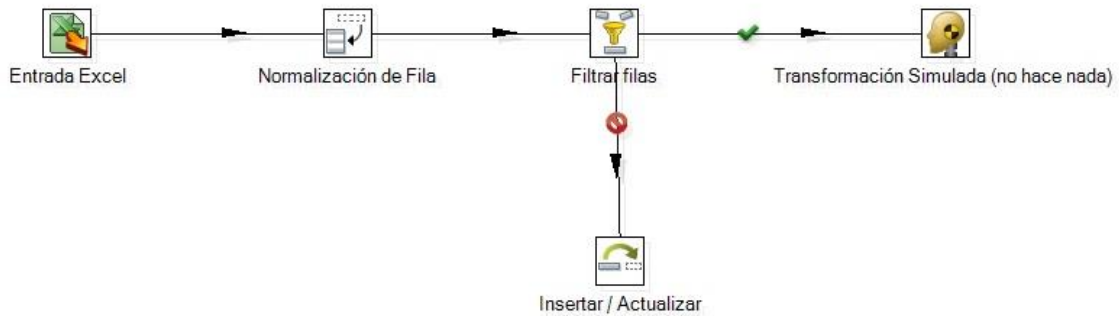
Anexo #4: Transformaciones.



Transformación Vía de ingreso.



Transformación Tiempo.



Transformación Asignatura.



Transformación Ocupación de la madre.



Transformación Ocupación del padre.



Transformación Nivel Escolar del Padre.



Transformación Nivel Escolar de la Madre.



Transformación Rango Académico Precedente.



Transformación Color de piel.



Transformación Provincia.



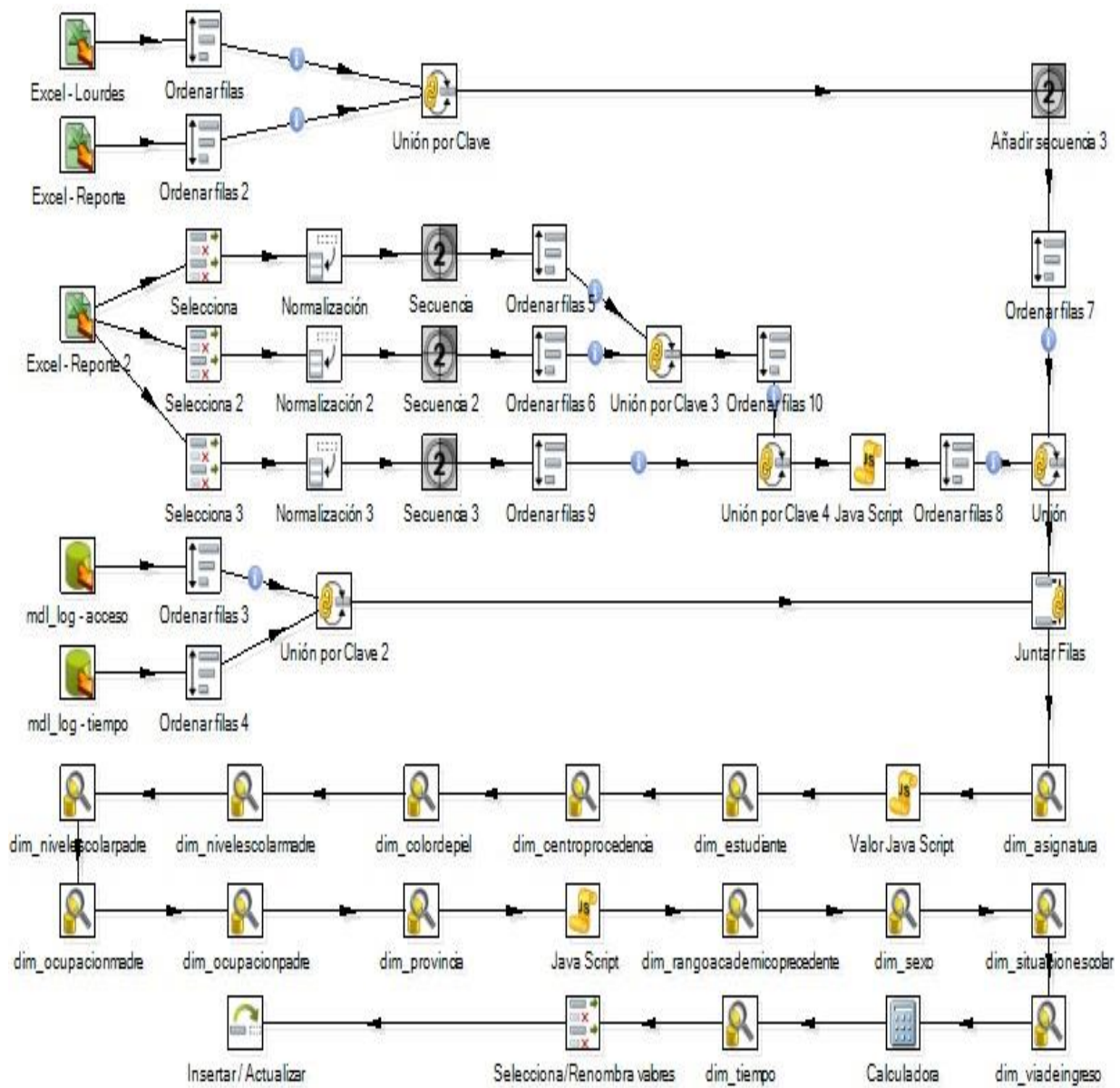
Transformación Situación Escolar.



Transformación Sexo.



Transformación Centro de Procedencia.



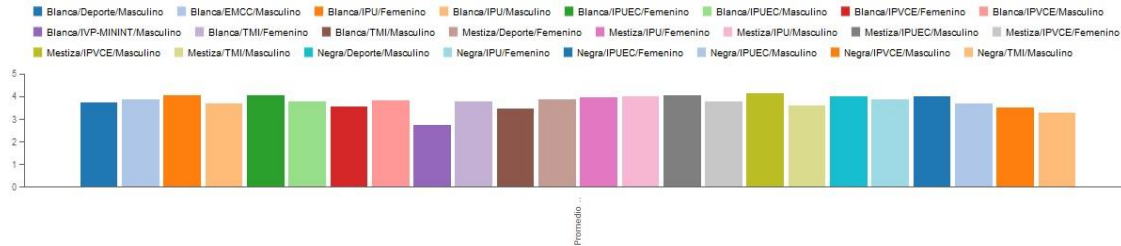
Transformación Cubo_Indicadores_Pregado.

Anexo #5: Reportes.

Info: 23.27 / 25 x 4 / 0.11s

MeasuresLevel	Blanca										Mestiza						Negra							
	Femenino				Masculino						Femenino			Masculino			Femenino		Masculino					
	IPU	IPUEC	IPVCE	TMI	Deporte	EMCC	IPU	IPUEC	IPVCE	IVP-MININT	TMI	Deporte	IPU	IPVCE	IPU	IPUEC	IPVCE	TMI	IPU	IPUEC	Deporte	IPUEC	IPVCE	TMI
Promedio de notas	4,062	4,059	3,538	3,767	3,727	3,857	3,681	3,769	3,825	2,75	3,473	3,857	3,952	3,786	4	4,045	4,158	3,59	3,857	4	4	3,714	3,5	3,294

Reporte 1 en forma de tabla.

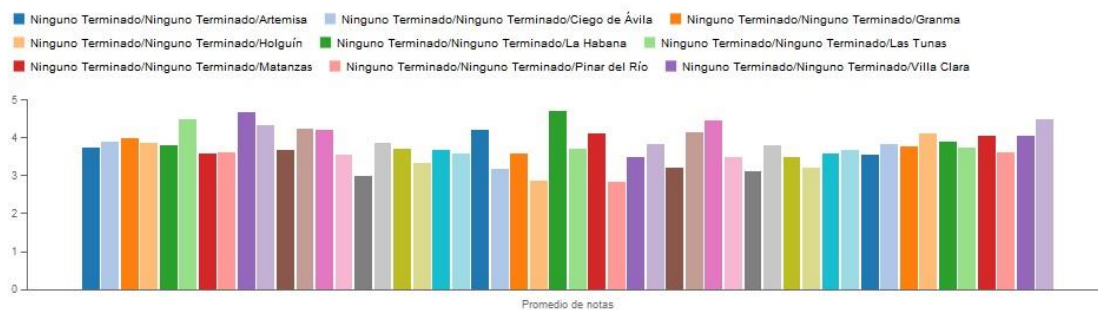


Reporte 1

en forma de gráfico.

MeasuresLevel	Ninguno Terminado										Obrero Calificado	Preuniversitario								
	Ninguno Terminado										Obrero Calificado	Preuniversitario							Secundi	
	Artemisa	Ciego de Ávila	Granma	Holguín	La Habana	Las Tunas	Matanzas	Pinar del Río	Villa Clara	La Habana	Artemisa	Camagüey	Granma	La Habana	Las Tunas	Matanzas	Pinar del Río	Santiago de Cuba	Villa Clara	La Habana
Promedio de notas	3,75	3,9	4	3,857	3,811	4,5	3,571	3,615	4,667	4,333	3,667	4,25	4,2	3,539	3	3,857	3,72	3,333	3,667	3,571

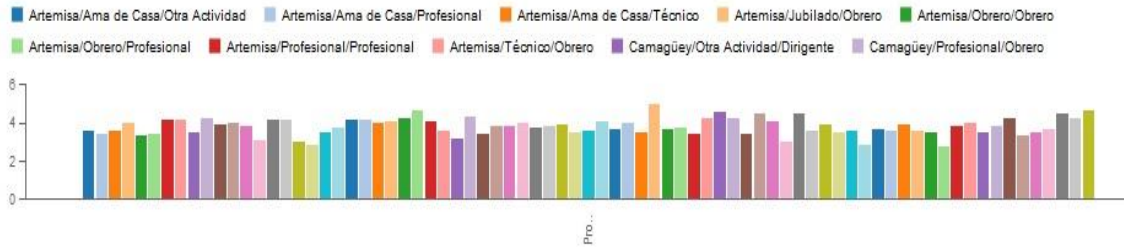
Reporte 2 en forma de tabla.



Reporte 2 en forma de gráfico.

MeasuresLevel	Artemisa								Camagüey		Ciego de Ávila	Cienfuegos				
	Ama de Casa		Jubilado	Obrero	Profesional	Técnico	Otra Actividad	Profesional	Técnico	Ama de Casa	Otra Actividad	Profesional	Obrero	Otra Actividad	Prof	
	Otra Actividad	Profesional	Técnico	Obrero	Obrero	Profesional	Profesional	Obrero	Dirigente	Obrero	Otra Actividad	Jubilado	Profesional	Cuentapropista	Obrero	Otra Actividad
Promedio de notas	3,615	3,429	3,571	4	3,333	3,429	4,2	4,143	3,556	4,25	3,9	4	3,833	3,125	4,167	4,167

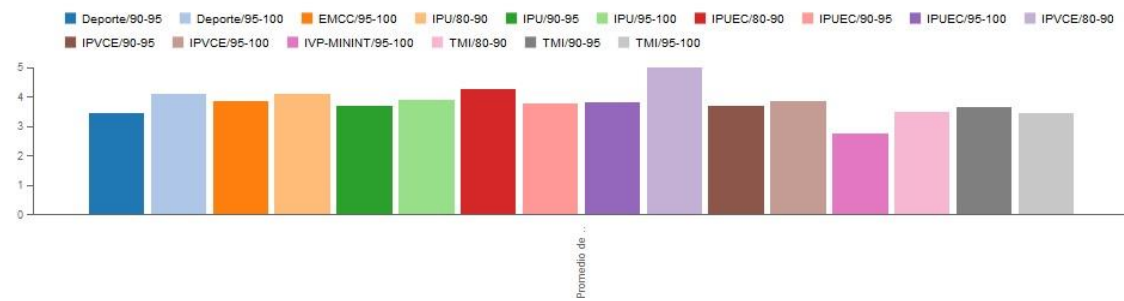
Reporte 3 en forma de tabla.



Reporte 3 en forma de gráfico.

	Deporte		EMCC	IPU			IPUEC			IPVCE			IVP-MININT	TMI		
MeasuresLevel	90-95	95-100	95-100	80-90	90-95	95-100	80-90	90-95	95-100	80-90	90-95	95-100	95-100	80-90	90-95	95-100
Promedio de notas	3,471	4,118	3,857	4,118	3,687	3,913	4,273	3,795	3,833	5	3,714	3,872	2,75	3,5	3,68	3,469

Reporte 4 en forma de tabla.



Reporte 4 en forma de gráfico.

MeasuresLevel	Nuevo Ingreso									Promovido									
	Instituto Politécnico			Preuniversitario			Cadetes MINFAR			Cadetes MININT		Concurso	Instituto Politécnico						
	La Habana	Cienfuegos	La Habana	Camagüey	La Habana	Santiago de Cuba	La Habana	Matanzas	La Habana	Artemisa	Holguín	La Habana	Las Tunas	Matanzas	Pinar del Río	Santiago de Cuba	Villa Clara	Artemisa	Camagüey
Cantidad de ausencias	1	1	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	2	

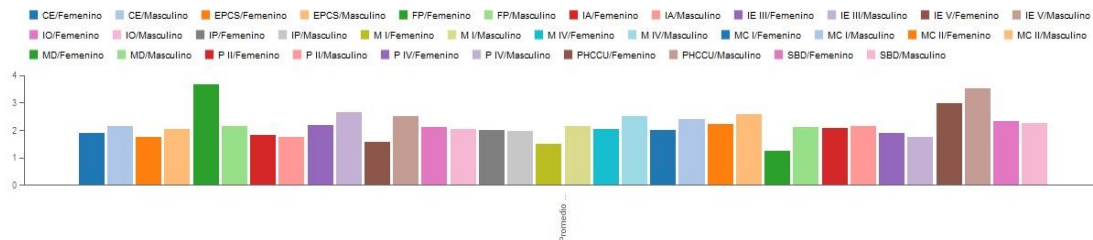
Reporte 5 en forma de tabla.



Reporte 5 en forma de gráfico.

MeasuresLevel	CE		EPCS		FP		IA		IE III		IE V		IO		IP	
	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino	Femenino	Masculino
Promedio de ausencias	1,909	2,156	1,75	2,045	3,667	2,167	1,833	1,767	2,188	2,673	1,583	2,533	2,111	2,062	2	1,964

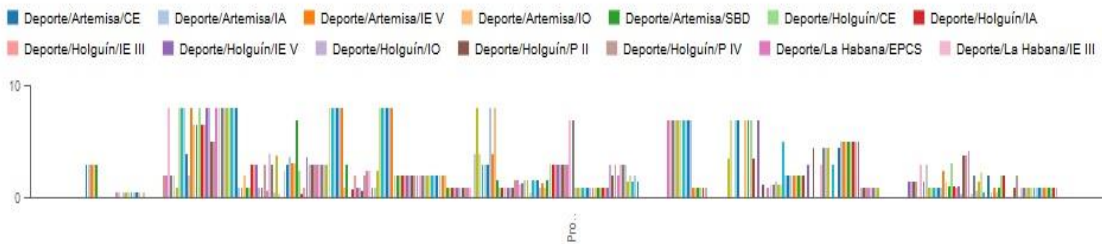
Reporte 6 en forma de tabla.



Reporte 6 en forma de gráfico.

MeasuresLevel	Deporte																EMCC																							
	Artemisa				Holguín				La Habana								Artemisa																							
	CE	IA	IE V	IO	SBD	CE	IA	IE III	IE V	IO	P II	P IV	EPCS	IE III	IO	IP	M I	M IV	MC I	MC II	MD	P II	P IV	SBD	EPCS	IE III	M IV	MC I	MC II	P II	SBD	CE	IA	IE III	IE V	IO	IP	M IV		
Promedio de ausencias	3	3	3	3	3	0	0	0	0	0	0	0	0,5	0,5	0	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0	0	0	0	0	0	0	0	0	2	2	8	2	2	1	8

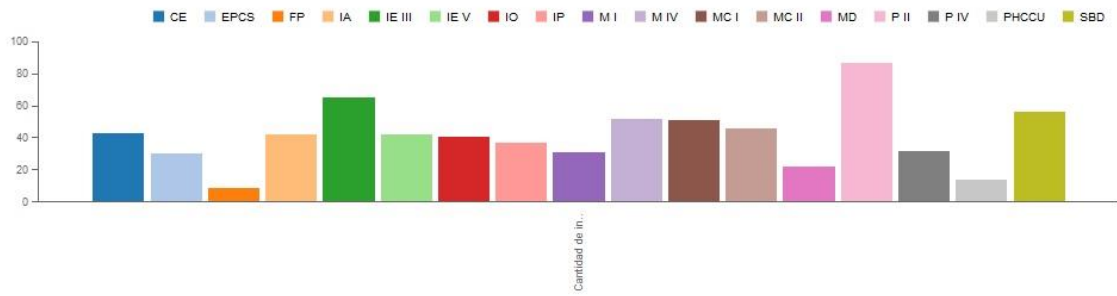
Reporte 7 en forma de tabla.



Reporte 7 en forma de gráfico.

MeasuresLevel	CE	EPCS	FP	IA	IE III	IE V	IO	IP	M I	M IV	MC I	MC II	MD	P II	P IV	PHCCU	SBD
Cantidad de interaccion con el EVEA	43	30	9	42	65	42	41	37	31	52	51	46	22	87	32	14	56

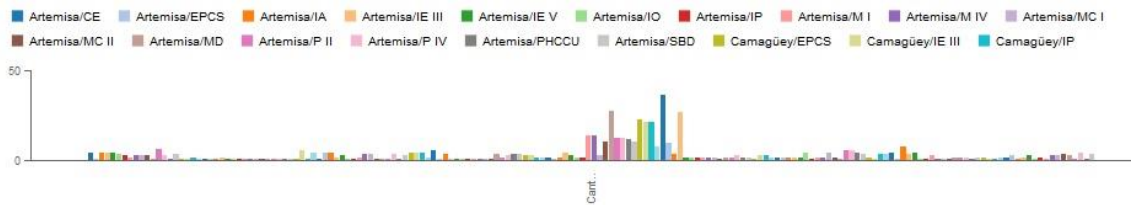
Reporte 8 en forma de tabla.



Reporte 8 en forma de gráfico.

MeasuresLevel	Artemisa															Camagüey										Ciego de Ávila															
	CE	EPCS	IA	IE III	IE V	IO	IP	M I	M IV	MC I	MC II	MD	P II	P IV	PHCCU	SBD	EPCS	IE III	IP	M I	M IV	MC I	MC II	MD	P II	PHCCU	SBD	CE	FP	IA	IE III	IE V	IO	M I	M IV	P II	P IV	PHCCU			
Cantidad de entradas al EVEA	5	1	5	5	5	4	3	2	3	3	3	1	7	3	1	4	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

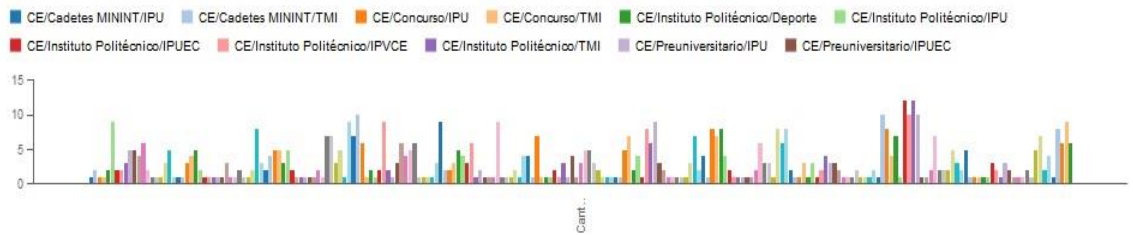
Reporte 9 en forma de tabla.



Reporte 9 en forma de gráfico.

MeasuresLevel	CE										EPCS															
	Cadetes MININT		Concurso		Instituto Politécnico			Preuniversitario			Cadetes MINFAR		Concurso		Instituto Politécnico				Preuniversitario							
Cantidad de interacción con el EVEA	IPU	TMI	IPU	TMI	Deporte	IPU	IPUEC	IPVCE	TMI	IPU	IPUEC	IPVCE	TMI	IPVCE	IPVCE	Deporte	EMCC	IPU	IPUEC	IPVCE	TMI	Deporte	IPU	IPUEC	IPVCE	TMI
	1	2	1	1	2	9	2	2	3	5	5	4	6	2	1	1	1	3	5	1	1	1	3	4	5	2

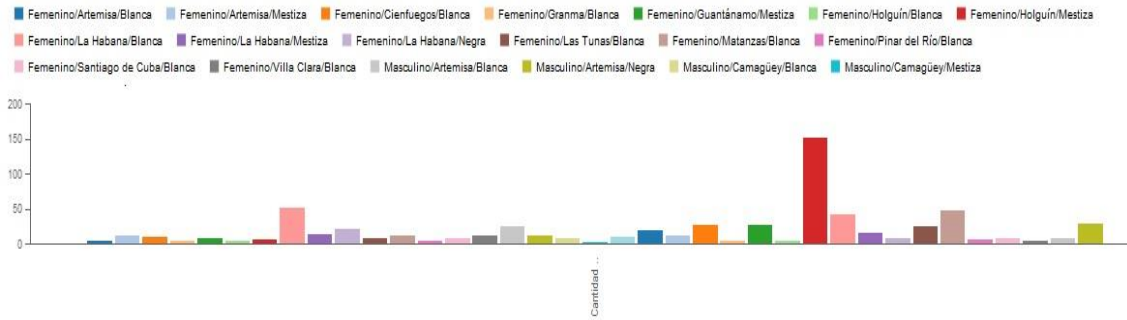
Reporte 10 en forma de tabla.



Reporte 10 en forma de gráfico.

MeasuresLevel	Femenino																				
	Artemisa		Cienfuegos		Granma	Guantánamo		Holguín		La Habana			Las Tunas	Matanzas	Pinar del Río	Santiago de Cuba	Villa Clara	Artemisa	Camagüey		Ciego de Ávila
	Blanca	Mestiza	Blanca	Blanca	Mestiza	Blanca	Mestiza	Blanca	Mestiza	Negra	Blanca	Blanca	Blanca	Blanca	Blanca	Blanca	Negra	Blanca	Mestiza	Blanca	
Cantidad de interacción con el EVEA	6	13	10	5	8	5	7	52	14	22	8	12	6	9	12	26	12	9	4	10	

Reporte 11 en forma de tabla.



Reporte 11 en forma de gráfico.