



FACULTAD 8

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

**Título: Análisis, Diseño e Implementación del Mercado de Datos
Indicadores de la Educación
para la Oficina Nacional de Estadísticas.**

Autoras: Rosio Franco Capote

Isabel Cedeño Pupo

Tutora: Ing. Yusimy Rodríguez Ruíz

Ciudad de La Habana, junio 2010

“Año 52 de la Revolución”

DECLARACIÓN DE AUTORÍA

Declaramos que somos las únicas autoras del trabajo “Análisis, Diseño e Implementación del Mercado de Datos Indicadores de la Educación para la Oficina Nacional de Estadísticas” y autorizamos a la Facultad 8 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

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

Firma del Autor

Isabel Cedeño Pupo

Firma del Autor

Rosio Franco Capote

Firma del Tutor

Ing. Yusimy Rodríguez Ruiz

Datos de Contacto

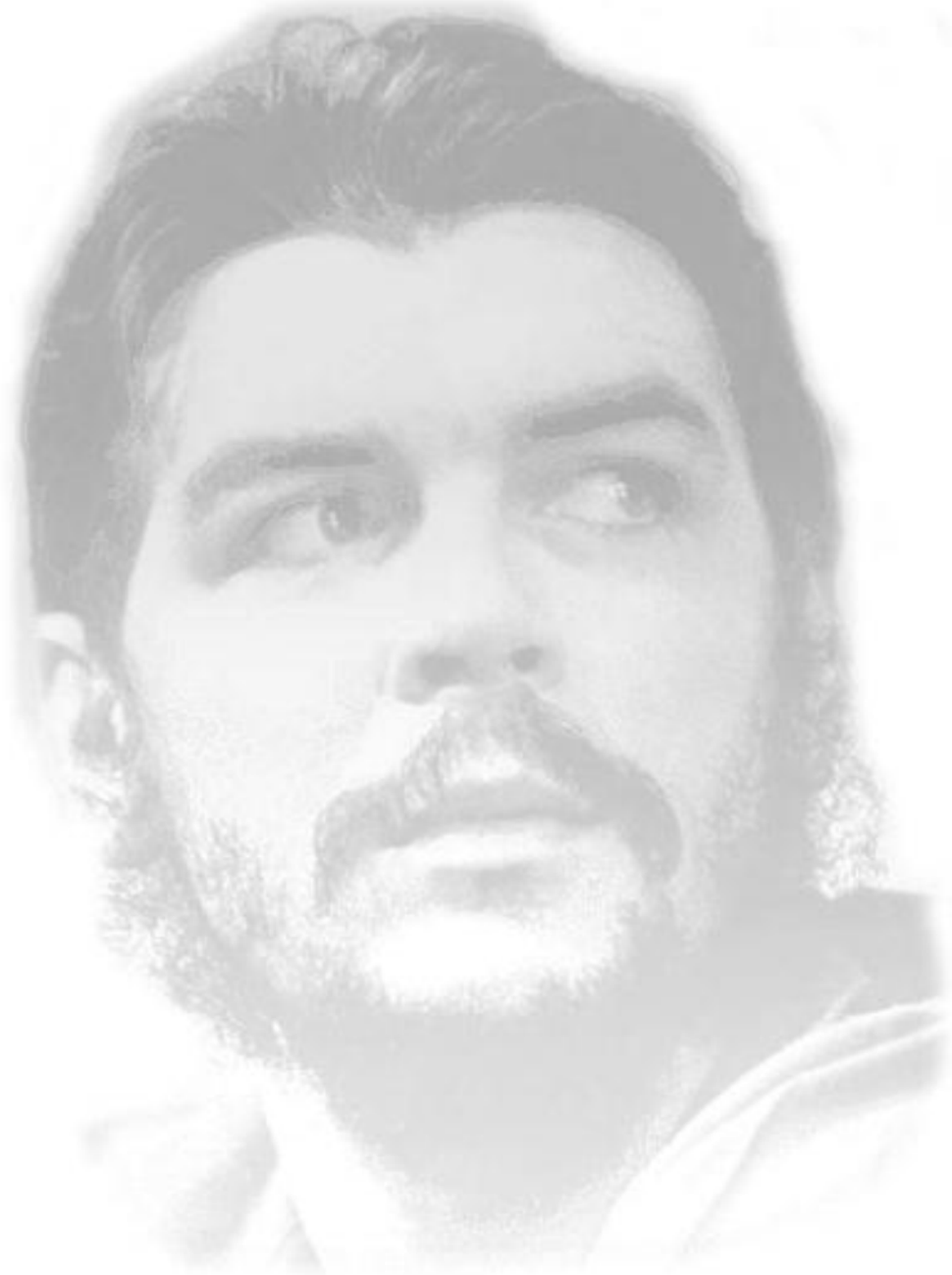
Ing. Yusimy Rodríguez Ruíz

Categoría Docente: Instructor (3 años de experiencia)

Especialidad: Ingeniero en Ciencias Informáticas

Correo electrónico: yrodriguezr@uci.cu

Centro de Trabajo: centro FORTES. Dpto. de producción de herramientas educativas. Universidad de las Ciencias Informáticas.



“La estadística es una de las disciplinas menos comprendida y conocida, pero de una gran importancia.”

El Che.

AGRADECIMIENTOS GENERALES

A nuestro Comandante en Jefe Fidel Castro por hacer nuestros sueños realidad.

Hay un conjunto de personas que nos han ayudado durante el paso por la universidad, brindándonos su amistad y apoyo para seguir adelante a pesar de las dificultades, a los cuales estaremos eternamente agradecidas:

A Roberto Dosagües: Fuiste más que nuestro profesor guía en primer año un gran amigo, una de las primeras personas que nos impulsó hasta aquí luego de ese curso de tantos conflictos, gracias por hacernos ver más allá de lo que podíamos ver.

A Tamara y Yaimy: Las secretarias docentes de la facultad 8, nuestras amigas, gracias por todo el apoyo y la ayuda que nos han dado en estos últimos 4 años.

A Isabel Lombillo: Por ser tan excepcional y habernos enseñado tanto además de habernos ayudado docente y personalmente.

También hay personas que han sido indispensables para la realización del trabajo de diploma, a los cuales damos innumerables agradecimientos:

A René: Por habernos ayudado tanto con tus conocimientos a cada momento que hizo falta.

A nuestra tutora Yusy, por demostrarnos su amistad, por todo el tiempo que nos ha dedicado, por las horas de sueño, hambre y cansancio que pasó junto a nosotras en la realización de este trabajo, por su ayuda y por habernos guiado como lo hizo hasta ahora.

A nuestra oponente Yanisbel González Hernández: Por habernos enseñado tanto, por su ayuda para realizar este trabajo y por sus recomendaciones que siempre fueron válidas.

A tribunal por habernos ayudado tanto en todo este tiempo.

Isabel y Rosio

AGRADECIMIENTOS

Quiero agradecer a todas esas personas que a lo largo de todos estos años en la universidad han estado conmigo y de una forma u otra me han ayudado a llegar hasta aquí.

A mi papá: Te agradezco que estés siempre conmigo, durante todos estos años de estudio, apoyándome y ayudándome en cada paso de este largo camino.

A mi mamá: Por estar siempre conmigo en las buenas y las malas.

A Yuri: Por ayudarme tanto y ser incondicional, además de ser mi compañero en los tres años más duros en esta universidad. En sentido general estos cinco años te considero mi mejor amigo, en gran parte te debo a ti haber llegado hasta el final.

A Manuel: Por haber estado dispuesto a ayudarme siempre que te necesité, desde el primer día.

A Rosio: Por ser mi amiga desde el primer día en esta universidad, conviviendo juntas todos estos años de una forma u otra y hoy por suerte estamos aquí compartiendo la mayor satisfacción del mundo, nuestro trabajo de diploma.

A mis amigos algunos de toda la vida como mi hermana Sandra, Arianna y su mamá Carmen que es casi una madre para mí, Felipe, Dachelys, Israel. También y no menos importantes esas personitas que conocí en la universidad que son Reñier mi amigote desde primer año, Taelen que a pesar de no haber continuado en la UCI te convertiste en una gran amiga sin importar la distancia y las dificultades, Elvis Doris, Ivet, Yanelis. El tiempo fue mucho más grato gracias a todos ustedes, siempre los llevaré en mi corazón.

Isabel

AGRADECIMIENTOS

A mi madre por apoyarme en todo, por educarme como lo ha hecho hasta ahora, por aconsejarme siempre y por ayudarme a ser la joven que soy.

A toda mi familia porque todos han contribuido en mi formación desde que era una niña. Todos me han apoyado, me han ayudado, me han aconsejado y han logrado que sea la joven que soy.

A mis abuelos Adela, Mencho, Emilio, Amelia y Tata por su dedicación conmigo, por el amor que siempre me han dado y por ser los mejores abuelos del mundo, en especial mi abuelito Mencho que se ha portado como un verdadero padre para mí.

A mis tíos Narcy y Juan Carlos porque han sido unos padres para durante estos 5 años, le doy gracias por tener que aguantar mis malacrianzas, por darme los gustos cada vez que iba para su casa, por estar dispuestos cada vez que los necesite y por muchas cosas más.

A mi tía Alexis por ser mi segunda madre y estar siempre al tanto de mí.

A mis primos Juanqui, Yordan y en especial a Yari que es una hermana para mí, por ser mi apoyo, mi amiga, por estar siempre conmigo en los buenos y malos momentos, por cuidarme cada vez que estaba enferma y no separarse de mí hasta que estuviera bien.

A Williams por ser el hermano que nunca tuve, mi paño de lágrimas, mi consejero, por haberme dado momentos de alegrías y por ser como es conmigo.

A Ale que me ha ayudado durante mi carrera, por ser amigo, por estar dispuesto siempre que lo necesité y por haberme dado tantos momentos agradables en todo este tiempo.

A mis amistades y mis compañeros de grupo que hemos compartido alegrías y tristezas, en especial a Isabel que además es mi compañera de tesis, Elvis Doris, Yanelis, Nory, Ivet, el bato, Karelia, el Yoa.

Rosio.

DEDICATORIA:

A mis padres y mi abuela por ser las personas que más amo en el mundo y que han estado conmigo toda la vida.

Isabel Cedeño Pupo.

A mi mami querida que es la luz de mi vida y mi razón de ser, es mi guía y mi apoyo, es quien se ha dedicado la vida entera a educarme y formarme para lograr lo que soy, a ella va dedicado este trabajo porque es merecedora de él y de mucho más.

A mi abuelita Adela porque su sueño era verme graduada y sé que dondequiera que esté va a estar feliz por esto.

Rosio Franco Capote

Resumen

La Oficina Nacional de Estadísticas es el órgano rector de la estadística en Cuba que se encarga de llevar el control de toda la información de los sectores de la economía y la sociedad, entre los que se encuentra la Educación. Con el fin de viabilizar la disponibilidad de la información para los órganos del estado y apoyar el proceso de toma de decisiones se realizó el análisis, diseño e implementación del mercado de datos Indicadores de la Educación. Para ello se hizo un estudio sobre las metodologías de desarrollo y se decidió tomar lo planteado por Ralfph Kimball incluyendo la creación de casos de uso para ajustarla a las políticas de desarrollo de la universidad. Durante su elaboración se definieron los conceptos de hechos y dimensiones que permitieron obtener el diseño de la solución. El mismo fue modelado mediante el esquema estrella y se desarrolló sobre la herramienta Visual Paradigm. El sistema gestor de base de datos que se eligió fue el PostgreSQL 8.4 y para su administración PgAdmin III 1.10. Como resultado de la implementación se obtuvieron los scripts Lenguaje de Definición de Datos, Lenguaje de Control de Datos y Lenguaje de Modelación de Datos que propiciaron la culminación del mercado de datos. La propuesta de solución fue evaluada de satisfactoria a través de un conjunto de pruebas para lograr la aceptación del cliente y un sistema con calidad.

Palabras Claves: mercado de datos, indicadores de la educación.

Índice de contenidos

INTRODUCCIÓN	5
CAPÍTULO 1 FUNDAMENTACIÓN TEÓRICA.....	8
1.1 TENDENCIAS EN EL MUNDO Y EN CUBA DE SISTEMAS QUE CONTROLAN ESTADÍSTICAS EN LA EDUCACIÓN.	8
1.2 TECNOLOGÍAS DE ALMACENAMIENTO DE DATOS.	10
1.2.1 Bases de Datos	11
1.2.2 Almacén de Datos	12
1.2.3 Mercado de datos	13
1.3 METODOLOGÍAS DE DESARROLLO	17
1.4 TIPOS DE MODELADO	18
1.4.1 Modelado Relacional	18
1.4.2 Modelado Multidimensional.....	19
1.4.3 Modelado Dimensional	19
1.5 HERRAMIENTAS DE MODELADO.....	21
1.5.1 Erwin	21
1.5.2 Visual Paradigm	21
1.6 SISTEMAS GESTORES DE BASES DE DATOS	22
1.6.1 Oracle.....	22
1.6.2 MySQL.....	22
1.6.3 SQL Server	23
1.6.4 PostgreSQL versión 8.4.....	24
1.7 PGADMIN III VERSIÓN 1.10 (HERRAMIENTA PARA LA ADMINISTRACIÓN DE BASES DE DATOS).....	24
CAPÍTULO 2 ANÁLISIS Y DISEÑO.....	27
2.1 DEFINICIÓN DEL NEGOCIO.....	27
2.2 TEMAS DE ANÁLISIS	29
2.3 REGLAS DEL NEGOCIO	29
2.4 NECESIDADES DE LOS USUARIOS	31
2.5 REQUISITOS DE INFORMACIÓN	31
2.6 REQUISITOS FUNCIONALES	33

2.7	REQUISITOS NO FUNCIONALES.....	33
2.8	CASOS DE USO DEL SISTEMA.....	35
2.8.1	<i>Casos de Uso de Información</i>	35
2.8.2	<i>Casos de Uso funcionales</i>	36
2.9	MATRIZ DIMENSIONAL (BUS).....	37
2.10	MODELO DE DATOS.....	39
2.11	MODELO DE DISEÑO.....	49
CAPÍTULO 3. IMPLEMENTACIÓN Y PRUEBAS.....		51
3.1	MODELOS DE DATOS FÍSICO.....	51
3.1.1	<i>Estructuras de Datos</i>	51
3.2	CARGA DE NOMENCLADORES.....	53
3.3	USUARIOS Y PRIVILEGIOS.....	54
3.3.1	<i>Usuarios y Roles</i>	54
3.3.2	<i>Privilegios</i>	54
3.4	GUÍA DE IMPLANTACIÓN.....	55
3.4.1	<i>Requerimientos</i>	55
3.4.2	<i>Secuencia de Pasos</i>	55
3.5	VALIDACIÓN Y PRUEBAS.....	56
3.4.1	<i>Listas de Chequeo Análisis</i>	56
3.4.4	<i>Pruebas de Implantación</i>	57
CONCLUSIONES GENERALES.....		59
RECOMENDACIONES.....		60
REFERENCIAS BIBLIOGRÁFICAS.....		61
BIBLIOGRAFÍA.....		64
ANEXOS.....		66
ANEXO1 TABLAS DE ESPECIFICACIÓN DE CASOS DE USO DEL SISTEMA.....		66
ANEXO 2 TABLAS DE ÍNDICES.....		71
ANEXO 3 ESQUEMAS Y TABLAS.....		72

Índice de tablas

TABLA 1 TIPOS DE ALMACENAMIENTO OLAP	15
TABLA 2 MATRIZ DIMENSIONAL	38
TABLA 3 ESPECIFICACIÓN DEL CASO DE USO DE INFORMACIÓN ANALIZAR INFORMACIÓN DE LOS CÍRCULOS INFANTILES.....	66
TABLA 4 ESPECIFICACIÓN DEL CASO DE USO DE INFORMACIÓN ANALIZAR INFORMACIÓN DE LAS ENSEÑANZAS PRIMARIA Y MEDIA	67
TABLA 5 ESPECIFICACIÓN DEL CASO DE USO DE EXTRAER EXCEL DE INDICADORES GENERALES DE LOS CÍRCULOS INFANTILES	68
TABLA 6 ESPECIFICACIÓN DEL CASO DE USO FUNCIONAL TRANSFORMAR Y CARGAR DATOS DEL EXCEL DE INDICADORES GENERALES DE LOS CÍRCULOS INFANTILES.....	70
TABLA 7 HECHO CÍRCULOS INFANTILES	72
TABLA 8 HECHO CONTROL DE ESCUELAS.....	73
TABLA 9 DIMENSIÓN AÑOS_CI	73
TABLA 10 DIMENSIÓN CENTROS_UNIVERSITARIOS.....	73
TABLA 11 DIMENSIÓN CINES	73
TABLA 12 DIMENSIÓN EDAD_ESTABLECIDA.....	73
TABLA 13 DIMENSIÓN EDUCACIONES	74
TABLA 14 DIMENSIÓN ESPECIALIDADES_EDUCACIÓN_TÉCNICA_PROFESIONAL	74
TABLA 15 DIMENSIÓN ESPECIALIDADES_UNIVERSITARIAS	74
TABLA 16 DIMENSIÓN ESTUDIANTES	74
TABLA 17 DIMENSIÓN ESTUDIANTES_EDUCACIÓN_TÉCNICA_PROFESIONAL	75
TABLA 18 DIMENSIÓN GEOGRAFÍA.....	75
TABLA 19 DIMENSIÓN ORGANISMOS	75
TABLA 20 DIMENSIÓN PERSONAL	75
TABLA 21 DIMENSIÓN POSGRADOS	75
TABLA 22 DIMENSIÓN TIEMPO_AÑOS.....	76
TABLA 23 DIMENSIÓN TIEMPO_CURSO_ESCOLAR.....	76
TABLA 24 DIMENSIÓN TIPO_ESTUDIO.....	76
TABLA 25 DIMENSIÓN ZONA_UBICACIÓN.....	76
TABLA 26 HECHO EDUCACIÓN POSGRADO	76
TABLA 27 HECHO ENSEÑANZA SUPERIOR	77
TABLA 28 HECHO ENSEÑANZA TÉCNICA PROFESIONAL.....	77

TABLA 29 HECHO ENSEÑANZAS PRIMARIA MEDIA77

TABLA 30 HECHO ESTUDIANTES PRIMARIA_MEDIA78

TABLA 31 HECHO INDICADORES ANUALES78

TABLA 32 HECHO NUEVO INGRESO MATRÍCULA EDUC_SUPERIOR.....79

TABLA 33 HECHO PERSONAL PRIMARIA MEDIA.....79

TABLA 34 HECHO TASAS ÍNDICES79

Introducción

Actualmente existe un constante crecimiento de las tecnologías de información proporcionado por la necesidad de almacenar grandes cantidades de datos. Esto trae consigo la creación de sistemas para el control de la misma. Por lo que se han creado empresas que brindan herramientas y sistemas cada vez más confiables y de mayor calidad. Estas impulsan el crecimiento de las exigencias de los usuarios, la competitividad en el mercado y se han insertado en las empresas mejoras en cuanto a organización, desarrollo tecnológico y recursos humanos. Con el objetivo de llevar un control de la información se han creado sitios en varios países que permiten consultar los datos de interés referentes a la economía, educación, salud, entre otros. Ejemplo de esto son varias de las funcionalidades existentes en el sitio oficial de la UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura).

En la Universidad de las Ciencias Informáticas se creó el Centro de Tecnologías de Gestión de Datos (DATEC), el cual brinda apoyo a varias instituciones entre las que se encuentra la Oficina Nacional de Estadísticas (ONE) que busca soporte para organizar y mantener un control de grandes cantidades de datos.

La Oficina Nacional de Estadísticas es el órgano rector de la estadística en Cuba, tiene como objetivo fundamental captar, analizar y difundir los datos recogidos a lo largo y ancho del país. Cuenta con modelos estadísticos referentes a todos los sectores de la economía y la sociedad cubana. Entre estos se encuentra el de los Indicadores de la Educación el cual presenta todas las estadísticas de la educación en Cuba desde los círculos infantiles hasta el nivel superior; las mismas se guardan en ficheros Excel donde se registran datos desde 1959 hasta la actualidad. Estos son utilizados por especialistas y órganos del estado para estudiar y analizar las estadísticas de su interés. La forma de almacenar, recuperar y presentar la información proveniente de los indicadores de la educación impide realizar los principales reportes, cruces de variables, tasas, porcentajes, indicadores y demás aspectos de interés; dificultando así la disponibilidad de la información para órganos del estado y afectando el proceso para la toma de decisiones.

A partir de lo antes expuesto surge el **problema a resolver**: ¿Cómo viabilizar la disponibilidad de la información estadística de la educación para mejorar el proceso de toma de decisiones en la Oficina Nacional de Estadísticas? Para darle solución se trazó como **objetivo general**: desarrollar el análisis, diseño e implementación del Mercado de Datos para el control estadístico de los indicadores de la

educación en la Oficina Nacional de Estadísticas. En correspondencia con lo anterior surge como **objeto de estudio** las tecnologías de almacenamiento de datos. Para cumplir el objetivo general se definieron los siguientes **objetivos específicos**:

- Elaborar el marco teórico de la investigación.
- Realizar el análisis y diseño del Mercado de Datos Indicadores de la Educación para la Oficina Nacional de Estadísticas.
- Realizar la implementación y carga de los clasificadores del Mercado de Datos Indicadores de la Educación.
- Validar la solución desarrollada.

La investigación se enmarca en el **campo de acción**: Mercado de Datos estadísticos para los indicadores de la educación y al finalizar se espera (**resultados esperados**): viabilizar el acceso a toda la información estadística de la educación en Cuba para la consulta colectiva del gobierno, los especialistas de la Oficina Nacional de Estadísticas y personas externas que necesiten acceder a los datos; contribuir al mejoramiento del proceso de la toma de decisiones.

Como apoyo a los objetivos específicos se plantearon las siguientes **tareas de investigación**:

- Realizar el estado del arte de la investigación.
- Valorar las principales tecnologías que actualmente se emplean en el almacenamiento de datos.
- Realizar un estudio de la organización.
- Definir el tema de análisis.
- Identificar las necesidades de los usuarios.
- Estructurar el modelo dimensional.
- Transformar del modelo dimensional al diseño físico.
- Definir la granularidad del proceso del negocio.
- Montar los clasificadores para Mercado de Datos Indicadores de la Educación.
- Realizar las pruebas al Mercado de Datos.

Para el desarrollo de la investigación se utilizaron los **métodos científicos** teóricos y empíricos.

Los métodos teóricos empleados fueron:

- **Analítico–Sintético:** para analizar los problemas existentes en la organización donde se va a desplegar la aplicación y para lograr un producto con las características que se requieren.
- **Histórico – Lógico:** para hacer un estudio acerca de la evolución del desarrollo tecnológico de los almacenes de datos, enmarcándose fundamentalmente en los mercados de datos.
- **Modelación:** para definir el negocio durante en la fase de análisis, las tablas de hechos y dimensiones y por último fue útil en la elaboración del modelo lógico de diseño.

Se utilizó como método empírico la **Observación** con el fin de apreciar cómo la Oficina Nacional de Estadísticas maneja la información estadística de la educación en Cuba y para apoyar la realización del diseño teórico de la investigación.

El trabajo está estructurado en 3 capítulos:

Capítulo 1 Fundamentación Teórica

Presenta un estudio referente a la situación actual en que se encuentra el manejo de datos estadísticos relacionados con la educación en Cuba y otras partes del mundo, a las tecnologías de almacenamiento de datos, metodologías de desarrollo, los tipos de modelado y las herramientas fundamentales para elaborar la propuesta de solución.

Capítulo 2 Análisis y Diseño

En este capítulo se encuentran la definición del negocio, el tema de análisis, las reglas del negocio, los requisitos (de información, los funcionales y los no funcionales) y los casos de uso del sistema. Además, en el diseño se muestran la matriz dimensional, la descripción de las dimensiones, las medidas de cada tabla de hecho y el diseño de la solución.

Capítulo 3 Implementación y Prueba

En este capítulo realizó la implementación del sistema donde se generó el modelo físico de la base de datos, se cargan los distintos clasificadores y se concluye con la realización de pruebas y validaciones para comprobar la calidad del sistema por parte de los desarrolladores y el cliente.

CAPÍTULO 1 Fundamentación Teórica

Introducción

En la actualidad la información ha ocupado un lugar predominante en la actividad del hombre. Su uso y almacenamiento han crecido a gran ritmo debido al incremento constante de las tecnologías en diversos sectores, trayendo consigo el manejo de gran cantidad de datos. Es por esto que en Cuba y otras partes del mundo existen organizaciones que han incorporado dentro de su funcionamiento sitios web y bases de datos que permiten consultar la información, obtener mayores niveles de capacidad de desarrollo y apoyar el proceso de toma de decisiones. Tanto en el sector de la educación como en todos los de la sociedad, la información estadística juega un papel primordial que posibilita efectuar el estudio de situaciones que a diario se presentan y dar respuesta a problemas de una forma útil y significativa.

1.1 Tendencias en el mundo y en Cuba de sistemas que controlan estadísticas en la Educación.

La educación en la actualidad es un factor crucial en el desarrollo social y económico. Las estadísticas educativas han ocupado un lugar importante en el mundo, para ello se han desarrollado soluciones que ofrecen servicios y se encargan de controlar los sistemas estadísticos de la educación. A continuación se hace referencia a varias organizaciones que han creado programas informáticos y que permiten el acceso a ellos.

Organización de Estados Iberoamericanos para la Educación la Ciencia y la Cultura (OEI): posee un sitio Web que contiene diversos indicadores sociales enfocados en el tema de la educación de los países iberoamericanos.

Sistema Nacional de Planificación Educativa (SEP): sitio creado por los Estados Unidos Mexicanos. Tiene como objetivo proporcionar información de los centros escolares de educación básica, media y superior a la población, por medio de un sistema de información que articula los diferentes componentes educativos de este país. Expone distintos indicadores como: datos referentes a la educación, las escuelas, estudiantes, docencia, becas, padres, entre otros datos de interés, permite además acceder a GeoSep.

Sistema de Información Geográfico para la Planeación Educativa (GeoSep): sistema mediante el cual se puede localizar uno o varios centros educativos de todos los niveles y servicios en un mapa digital de la República Mexicana, así como otros programas que facilitan al usuario una mejor navegación. Posee una serie histórica y pronósticos de indicadores educativos según el nivel de enseñanza. Forma parte de SEP. (2)

Oficina Nacional de Estadísticas (ONE): esta institución posee un sitio web que facilita consultar diversas informaciones estadísticas de Cuba que sistematiza la Comisión Económica para América Latina (CEPAL). El acceso a estas bases de datos posibilita la elaboración de tablas que combinan datos de diferentes series y/o países y/o períodos. Cada una de las variables tiene asociada una descripción de las fuentes, así como las definiciones y otras características relevantes de las mismas. Con este servicio la ONE procura divulgar las buenas prácticas en materia de expansión de la información estadística, poniéndolas al servicio de todos sus usuarios. (1) Por otro lado permite también el acceso al sitio División de Estadísticas de las Naciones Unidas (UNSD) para que todos tengan la posibilidad de consultarlo. Además posee informaciones de varias organizaciones internacionales como por ejemplo: la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), la Organización de Desarrollo Industrial de las Naciones Unidas (UNIDO) y el Fondo de las Naciones Unidas para la infancia (UNICEF). Todas ellas cuentan con un sitio Web que permite consultar sus datos. A continuación se citan algunos ejemplos.

Comisión Económica para América Latina (CEPAL): su labor se extiende a los países del Caribe y se incorporó con el objetivo de promover el desarrollo social enfocado en factores económicos, brinda acceso a los datos mediante un sitio Web, tiene incorporada una base de datos para un mejor manejo de la información.

CEPALESTAT: es la base de datos Pobreza en América Latina creada por CEPAL, mediante la cual se pueden realizar diversas consultas, genera reportes en formato Excel y varios datos acerca de los indicadores que contiene.

División de Estadísticas de las Naciones Unidas (UNSD): ha puesto al servicio de la comunidad internacional una amplia base de datos en temas como educación, población, industria, energía, comercio y cuentas nacionales que abarcan a todos los países. La gran mayoría de estos datos han sido

proporcionados por los propios estados miembros mediante los cuestionarios que se les solicita y en otros casos parten de estimaciones o cálculos de los organismos internacionales. Por esta razón es posible que no coincidan determinados datos para un año referido o no aparezca información aun cuando el país la posee y está publicada en su sitio Web.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO): su principal objetivo es contribuir al mantenimiento de la paz y la seguridad en el mundo, promoviendo a través de la educación, la ciencia, la cultura y la comunicación la colaboración entre las naciones; posee un sitio oficial con informaciones acerca de los temas que maneja.

1.2 Tecnologías de almacenamiento de datos.

Desde la antigüedad al hombre se le hizo imprescindible el almacenamiento de datos históricos y análisis de estos para tomar decisiones futuras. Esta técnica era aplicada por los aztecas y mayas en su increíble calendario solar (3). Incluso los egipcios atesoraban registros de las primaveras con amplios desbordamientos del Nilo mediante este modo y podían saber si tendrían buena cosecha o no ese año. (4)

El estudio de datos relacionados con la gestión empresarial empezó cuando todavía la computación no llegaba a dar respuesta a estos problemas. Los directivos estudiaban enormes informes elaborados por comerciales y económicos compuestos de varias páginas de datos escrupulosamente resumidos. A partir estos problemas surge la era de la computación. El término de bases de datos fue escuchado por primera vez en 1963, en un simposio celebrado en California, Estados Unidos, a partir de la necesidad de almacenar grandes cantidades de datos para su posterior consulta, producidos por las nuevas industrias que creaban gran cantidad de información. (5)

En la actualidad existe gran cantidad de alternativas en línea que permiten hacer búsquedas orientadas a necesidades específicas de los usuarios. Una de las tendencias más amplias son las bases de datos, que permiten el almacenamiento de gran cantidad de información en distintos formatos, posibilitan una mayor visibilidad y acceso en el ámbito científico así como otros medios. Antes de hacer un mayor acercamiento respecto al tema es necesario hacer referencia a algunos conceptos esenciales para la comprensión de estas tecnologías.

Hecho

Es el objeto a analizar, posee atributos llamados de hechos o de síntesis y son de tipo cuantitativo. Sus valores se denominan medidas, se obtienen generalmente por la aplicación de una función estadística que resume un conjunto de valores en un único valor. Ejemplo de hechos: círculos infantiles, enseñanza primaria, enseñanza superior, entre otros. (20)

Dimensiones

Representan cada uno de los ejes en un espacio multidimensional. Suministran el contexto en el que se obtienen las medidas de un hecho. Algunos ejemplos son: tiempo, edad y escuelas. Las dimensiones se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. Los componentes de una dimensión se denominan niveles y se organizan por jerarquías. Por ejemplo, la dimensión tiempo puede tener niveles día, mes y año. (20)

Es importante acotar que los hechos se guardan en tablas de hechos y las dimensiones en tablas de dimensiones.

Hipercubo

Consiste en un conjunto de celdas, cada una se identifica por la combinación de los miembros de las diferentes dimensiones y contiene el valor de la medida analizada para dicha combinación de dimensiones. (20)

1.2.1 Bases de Datos

Según David Bolton (6) las bases de datos se definen como una aplicación que gestiona los datos y permite el almacenamiento y recuperación rápida de estos. Existen diferentes tipos de bases de datos pero las más utilizadas son las relacionales, que almacenan datos en las tablas donde cada fila tiene el mismo tipo de información.

Debido al crecimiento de las empresas los problemas de acumulación de información se agudizaron aún más y fue necesario expandirse en el área de almacenamiento de datos. Se requería de una enorme redundancia en los datos para respaldar múltiples ambientes de soporte de decisiones, situación nada deseable y que empeoraba al momento de reunir, limpiar o integrar nueva información. Es así como surge a finales de los años 80 el concepto de “almacén de datos”, cuando los investigadores de IBM

(International Business Machine), Barry Devlin y Paul Murphy desarrollaron el "Business Datawarehouse" (Negocios de almacenamiento de datos) con el propósito de proveer un modelo de arquitectura del flujo de datos de los sistemas operacionales a los ambientes de soporte de decisiones.

1.2.2 Almacén de Datos

Un almacén de datos se define como una colección de datos de gran volumen provenientes de diferentes fuentes; es un repositorio lógico de datos que recoge la información histórica de determinados departamentos de una empresa, con la misión de obtener información útil para el soporte de la toma de decisiones a partir de estos datos y apoyado en herramientas sofisticadas de análisis. (8) Brinda varios aportes como (7):

- Es de gran importancia para las empresas en el proceso de toma de decisiones, es una herramienta basada en información integrada y global del negocio.
- Presta la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Optimiza tecnológicamente y económicamente entornos de centro de información, estadística o de generación de informes con retornos de la inversión espectaculares.

Según Bill Inmon las principales características de un almacén de datos son (7):

- **Integrado:** los datos almacenados en el almacén de datos deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Temático:** sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del almacén de datos. De esta forma las

peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

- **Histórico:** el tiempo es parte implícita de la información contenida en un almacén de datos. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en estos sirve entre otras cosas, para realizar análisis de tendencias. Por tanto el almacén de datos se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- **No volátil:** el almacén de información de datos existe para ser leído pero no modificado, por lo cual la información es permanente, significando la actualización del almacén de datos como la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

En resumen, un almacén de datos constituye un repositorio lógico de datos donde se acumulan diferentes tipos de información para su posterior explotación. Posee un conjunto de características de integración de datos, seguridad y análisis. Puede reunir la información de varios departamentos de la organización y así apoyar el proceso de toma de decisiones de las empresas.

1.2.3 Mercado de datos

El mercado de datos o Datamart en inglés, es una especialización del almacén de datos. Está enfocado a un departamento o área específica, como por ejemplo los departamentos de economía o comercio, permitiendo así un mejor control de la información (9). Su principal beneficio radica en la centralización de los datos en un área específica dentro de la organización, consintiendo el acceso a estos por medio de un gran número de herramientas; acelerando las consultas y reduciendo el volumen de datos a recorrer; además divide los datos para imponer estrategias de control de acceso.(10)

En ocasiones las empresas necesitan que los usuarios hagan uso de un subconjunto de datos específico, para estos casos se utilizan los mercado de datos. Al manejar eficientemente la información de cada área de la empresa se pueden tomar mejores decisiones, efectuando acciones apropiadas con el fin de conseguir un mejor control sobre la producción empresarial y cumplir con varias de las metas trazadas al inicio de la investigación.

Existen dos tipos de mercados de datos.

Mercado de datos OLTP¹

Constituyen bases de datos orientadas al procesamiento de transacciones. (11) Puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

Cumplen las siguientes características (9):

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura. (Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las bases de datos de bancos o hipermercados diariamente).
- Los datos se estructuran según el nivel aplicación (programa de gestión a medida, sistema de información departamental).
- El historial de datos suele limitarse a los datos actuales o recientes.

Mercado de datos OLAP²

Se basa en los **cubos OLAP** que se construyen agregando según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo en función de la herramienta final que se utilice, cumpliendo con las siguientes características (7):

- El historial de datos es a largo plazo, normalmente de dos a cinco años.
- Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes mediante un proceso de extracción, transformación y carga (ETL).
- El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta con muy pocas inserciones, actualizaciones o eliminaciones.
- Los datos se estructuran según las áreas de negocio y los formatos de los datos están integrados de manera uniforme en toda la organización.

¹ *Procesamiento transaccional en línea*

² *Procesamiento analítico en línea*

OLAP puede trabajar con tres tipos de almacenamiento (9):

Tabla 1 Tipos de Almacenamiento OLAP

Tipos de Almacenamiento OLAP	CARACTERÍSTICAS
<p>ROLAP (Cubo Relacional OLAP)</p>	<p>Posibilita que las agregaciones de la partición se almacenen en vistas indizadas³ de la base de datos relacional que se especificó en el origen de datos de la partición.</p> <p>No permite que se almacene una copia de los datos del origen en una estructura multidimensional en las carpetas de datos de Analysis Services, en español Servicio de Análisis. En su lugar, cuando no se pueden derivar los resultados de la caché de consultas, se utilizan las vistas indizadas del origen de datos para responder a ellas. La respuesta a las consultas suele ser más lenta que con los modos de almacenamiento MOLAP u HOLAP que se mostrarán a continuación. El tiempo de procesamiento también suele ser más lento. No obstante, permite a los usuarios ver los datos y ahorrar espacio de almacenamiento al trabajar con conjuntos de datos grandes a los que no se suele consultar con frecuencia, como datos puramente históricos.</p>
<p>MOLAP (Cubo Multidimensional OLAP)</p>	<p>Da lugar a que las agregaciones de la partición y una copia de sus datos de origen se almacenen en una estructura multidimensional en Analysis Services. Está muy optimizada para maximizar el rendimiento de las consultas. La ubicación de almacenamiento puede estar en el equipo en donde se define la partición o en otro equipo que ejecute Analysis Services. Dado que una copia de los datos de origen reside en la estructura multidimensional, las consultas se pueden resolver sin necesidad de obtener acceso a los datos de origen de</p>

³ Consultas a las cuales se les asigna un índice para acelerar el proceso de búsqueda dentro de la base de datos.

	<p>la partición. Si se utilizan agregaciones, los tiempos de respuesta a las consultas pueden disminuir notablemente. Los datos de la estructura MOLAP de la partición están tan actualizados como el procesamiento más reciente de la misma.</p>
<p>HOLAP (Híbrido OLAP)</p>	<p>Combina atributos de los modos MOLAP y ROLAP. Al igual que MOLAP, hace que las agregaciones de la partición se almacenen en una estructura multidimensional Analysis Services. No almacena una copia de los datos de origen. Es el equivalente de MOLAP para las consultas que sólo tienen acceso a los datos de resumen de las agregaciones de una partición; las que tienen acceso a los datos de origen como por ejemplo, la obtención de detalles para una celda de un cubo atómico sin datos de agregación, deben recuperar la información de la base de datos relacional y no serán tan rápidas como lo serían si los datos de origen se hubieran almacenado en la estructura MOLAP. Con el modo de almacenamiento HOLAP, los usuarios suelen experimentar notables diferencias en cuanto a los tiempos de las consultas, según si se puede resolver desde la caché o las agregaciones frente a los propios datos de origen.</p>

Luego de hacer un estudio acerca de las tecnologías de almacenamiento de datos, se decidió utilizar los mercados de datos debido a las ventajas que proporciona dentro de la empresa para el control centralizado de información de un departamento; específicamente el almacenamiento OLAP de tipo ROLAP que brinda ventajas a los desarrolladores en cuanto a su entorno y además las herramientas necesarias para su utilización están disponibles son de fácil acceso para los mismos.

1.3 Metodologías de Desarrollo

Existen diferentes metodologías que definen y guían el ciclo de vida de desarrollo de la solución. Las tendencias más conocidas son las de Bill Inmon⁴ y Ralph Kimball⁵, que pretenden dar un acercamiento a una propuesta ideal para el desarrollo de almacenes de datos. Además se han desarrollado variantes de estas metodologías como son:

Metodología SQLBI: propone una metodología para la implementación de soluciones de inteligencia del negocio utilizando SQLServer mediante herramientas de la suite de Business Intelligence de Microsoft.

Metodología DM2: enfocada en las necesidades de información a nivel gerencial dentro de la empresa, de manera que los usuarios tengan acceso a la misma. Hace más rápido el tránsito de la fase de análisis a la de implantación incrementando el rendimiento del sistema.

Metodología CRISP-DM: consiste en un conjunto de tareas descritas en cuatro niveles de abstracción (de lo general a lo específico): fase, tarea genérica, tarea especializada e instancia de procesos.

HEFESTO: es aplicable tanto para Almacenes de Datos como para mercados de datos, utilizando modelos conceptuales y lógicos, sencillos de interpretar y analizar. Está basado en los requerimientos del usuario, por lo que su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio. Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del almacén de datos. (15)

Metodología para el Diseño Conceptual de Almacenes de Datos: presentada en la Tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez que plantea transformaciones para llevar desde un diagrama relacional a uno dimensional para obtener estructuras que conformarán un repositorio de datos e incorpora los casos de uso como guías en el proceso de desarrollo. (14)

⁴Escritor bastante prolífico sobre la construcción, utilización y mantenimiento del Datawarehouse. Ha escrito más de 650 artículos sobre el tema, y sus trabajos han sido publicados en la mayoría de las revistas de computación. Ha escrito 46 libros, y tan sólo de uno de ellos se han vendido medio millón de copias. (28)

⁵Autor ampliamente reconocido en materia de Datawarehouse Inteligencia del Negocio, es considerado como "Gurú del Datawarehouse" y reconocido por su plena convicción en que un Datawarehouse debe ser diseñado para ser entendible y veloz. (28)

Estas han venido desarrollándose para la implantación de este tipo de soluciones. Cada autor la orienta a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico. En dependencia de la problemática presentada es la política de selección para una u otra metodología.

La **Metodología de Inmon** propone construir primero el almacén de datos y luego los mercados de datos. Plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construidos de manera descendente los mercados de datos, estos se nutren del almacén de datos, quien se convierte en un complejo empresarial de bases de datos relacionales (12)

Kimball a diferencia de **Inmon** propone la elaboración del almacén de datos a través de la construcción de los mercados de datos, crea los conceptos hechos y dimensiones lo cual proporciona agilidad en el proceso de toma de decisiones. Existe una gran documentación acerca de esta lo que proporciona facilidades al equipo de desarrollo. Es una metodología madura donde se definen etapas, actividades, artefactos y roles. (12)

Para la realización del Mercado de Datos Indicadores de la Educación en la Oficina Nacional de Estadísticas (ONE) se va a emplear una metodología basada en lo planteado por Kimball pero acercándola a las tendencias y normas de la Universidad de las Ciencias Informáticas (UCI). Para robustecer la etapa de levantamiento de requisitos mediante la utilización de casos de uso, se empleó la Metodología para el Diseño Conceptual de Almacenes de Datos. (13)

1.4 Tipos de Modelado

En la realización del almacenamiento de datos es de gran importancia utilizar metodologías de diseño, ya que estas les sirven de soporte a las bases de datos mediante la utilización de procedimientos, técnicas y herramientas que ayudan a la generación de documentación. (19)

1.4.1 Modelado Relacional

Uno de los modelos matemáticos más importantes y actuales para la representación de las bases de datos es el enfoque relacional. Se basa en la teoría matemática de las relaciones, suministrándose por ello una fundamentación teórica que permite aplicar todos los resultados de dicha teoría a problemas tales

como el diseño de sub-lenguajes de datos y otros. En el modelo relacional, tanto los objetos o entidades, como las relaciones que se establecen entre ellos, se representan a través de "tablas" que en la terminología relacional se denominan relaciones.

Cada relación está compuesta de filas (las ocurrencias de los objetos) y se les denomina en la terminología relacional como tuplos (en realidad n-tuplos pero en muchos casos se suprime la n cuando no existe posibilidad de confusión) (21). No es lo más factible para la solución de la propuesta ya que no garantiza la recuperación óptima de los datos almacenados, además para la confección de los diagramas se tiene en cuenta la normalización⁶, lo cual no es necesario en la confección de los almacenes de datos. (13)

1.4.2 Modelado Multidimensional

En un modelo multidimensional los datos se organizan alrededor de los temas de la organización. La estructura de datos manejada en este modelo son matrices multidimensionales o hipercubos. Las dimensiones definen dominios como geografía, edad y tiempo.

- Los miembros de una dimensión se agrupan de forma jerárquica (dimensión geográfica: ciudad, provincia, país...).
- Cada celda contiene datos agregados que relacionan los elementos de las dimensiones.
- Es adecuado para resumir y organizar datos (hojas de cálculo).
- Está enfocado para trabajar sobre datos de tipo numérico.
- Es más simple y fácil de visualizar y entender que el modelado relacional.(20)

1.4.3 Modelado Dimensional

Se habla de modelos dimensionales al diseñar mercado de datos, bodega de datos y sistemas de inteligencia de negocios sobre tecnología Datawarehousing⁷. Teniendo un fin analítico de todos estos

⁶Desarrollado para obtener estructuras de datos eficientes que eviten las anomalías de actualización, fue pensado para aplicarse a sistemas relacionales aunque tiene aplicaciones más amplias. Es la expresión formal del modo de realizar un buen diseño. Provee los medios necesarios para describir la estructura lógica de los datos en un sistema de información.(21)

⁷ Es donde se utiliza como almacenamiento de datos el almacén.

sistemas, es necesario utilizar otras técnicas de modelado que permitan registrar las relaciones existentes entre los datos, para ofrecer una fuente de consulta que permita el análisis y la toma de decisiones (13). Posibilita que el procesamiento de información sea eficiente, es flexible para aceptar nuevos datos y se basa en el modelo estrella.

Esquema en estrella

El esquema en estrella es la arquitectura de almacén de datos más simple. En este diseño la tabla de variables (hechos) está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP (22). Mediante este esquema se puede encontrar la información de cada tabla de hechos y de cada tabla de dimensiones. Lo característico de la arquitectura en estrella es que sólo existe una tabla de dimensiones para cada dimensión y esta representa la segunda forma normal.

El centro del esquema se denomina hecho y está unido a múltiples dimensiones, estos pueden tener varias medidas que en conjunto pueden ser de un solo tipo: aditivas, no aditivas o semi-aditivas, donde el nivel de detalle que tendría una fila de datos al relacionar las medidas con las dimensiones asociadas se denomina granularidad, lo cual daría el mínimo nivel de detalle en el que un proceso de negocio es modelado para satisfacer la demanda de análisis que tendrán los usuarios sobre él. (23)

Esquema en copo de nieve

Es una variedad más compleja del esquema estrella, el afinamiento está orientado a facilitar el mantenimiento de dimensiones. Lo que distingue a la arquitectura en copo de nieve del esquema estrella es que las tablas de dimensiones en este modelo representan relaciones normalizadas (3NF) y forman parte de un modelo relacional de base de datos con varios usos del esquema en copo de nieve, el más común es cuando las tablas de dimensiones están muy grandes o complejas y es muy difícil representar los datos en esquema estrella. Tiene como desventaja que el vínculo de muchas tablas en las sentencias SQL puede llegar a ser muy complejo y difícil para mantener. (24)

Para el desarrollo del sistema se va a utilizar el modelado dimensional basado en el esquema en estrella porque permite modelar todas las estructuras dentro del mercado de datos, gracias a su simplicidad para la representación gráfica de la relaciones entre los hechos y las dimensiones; cumpliendo con los aspectos planteados en la metodología de desarrollo seleccionada.

1.5 Herramientas de modelado

En la actualidad existe un incremento de las tecnologías de información, creándose una constante competitividad por parte de los productores de herramientas para el control de la misma, lo cual genera grandes exigencias en cuanto al rendimiento y calidad de los productos.

Se han creado herramientas nombradas CASE (Ingeniería de Software Asistida por Computadora), útiles para la modelación de proyectos, flujos de información, entidades de datos así como requerimientos del sistema facilitando el proceso de planeación. Mucha de la información que se captura durante esta fase se empleará también durante el desarrollo y el mantenimiento del ciclo de vida del sistema. (25)

1.5.1 Erwin

Es una herramienta CASE para diseñar bases de datos que brinda productividad en su diseño, generación y mantenimiento de aplicaciones, desde un modelo lógico de los requerimientos de información hasta el modelo físico perfeccionado para las características específicas de la base de datos diseñada. Además, Erwin permite visualizar la estructura, los elementos importantes y optimizar el diseño de la base de datos. Genera automáticamente las tablas y miles de líneas de procedimientos almacenados y disparadores⁸ para los principales tipos de base de datos. Es una herramienta propietaria.

Erwin soporta principalmente bases de datos relacionales SQL y bases de datos que incluyen Oracle, MicrosoftSQL Server, Sybase. El mismo modelo puede ser usado para generar múltiples bases de datos o convertir una aplicación de una plataforma de base de datos a otra. (25)

1.5.2 Visual Paradigm

Es una herramienta profesional de modelado CASE que soporta el ciclo de vida completo del desarrollo del software comprendido en: análisis y diseño orientado a objetos, construcción, pruebas y despliegue. Es propietaria, posee varias versiones, como son: Enterprise, Professional, Standard, Modeler, Personal y Community (que es gratuita) además sus distintas ediciones tienen compatibilidad. Esta herramienta de modelado UML (Lenguaje Unificado de Modelado) ayuda a una rápida construcción de aplicaciones de

⁸ *Función del gestor de base de datos (triggers).*

calidad sin un elevado costo. Permite elaborar diagramas de clases, código inverso generando código a través de estos, así como documentación. Soporta aplicaciones Web aunque las imágenes y reportes generados no son de muy buena calidad. Posee generación de código para Java en HTML y es muy fácil de instalar y actualizar. (25)

Para el modelado de este sistema se utilizará la versión para UML 6.4 Enterprise Edition por las grandes posibilidades que brinda y en la Universidad de las Ciencias Informáticas (UCI) lugar donde se desarrolla la investigación, existe una licencia para el manejo del mismo lo que posibilita su utilización.

1.6 Sistemas gestores de bases de datos

Hoy en día las organizaciones enfrentan numerosos desafíos de información tales como: la necesidad de tomar decisiones más rápidas y orientadas a datos, así como de aumentar la productividad y flexibilidad del personal de desarrollo. La presión ejercida para reducir los presupuestos generales de informática escala una infraestructura para satisfacer exigencias cada vez mayores; por lo que se han creado diferentes herramientas que posibilitan la construcción de un almacén de datos conocidas como Sistemas Gestores de Bases de Datos.

1.6.1 Oracle

Es una herramienta de administración gráfica cómoda de utilizar, ayuda a analizar la información y efectuar recomendaciones concernientes a mejorar el rendimiento y la eficiencia en el manejo de aquellos datos que se encuentran almacenados apoyando el diseño y optimización de los modelos de datos. Es muy útil para los desarrolladores ya que los asiste con sus conocimientos de SQL y de construcción de procedimientos almacenados. Es ventajoso para documentar y mantener un registro periódico de las mantenciones, actualizaciones de hardware y software, cambios en las aplicaciones y en general todos aquellos eventos relacionados con cambios en el entorno de utilización de una base de datos. (14)No es una herramienta libre lo cual trae consigo limitaciones para su utilización.

1.6.2 MySQL

Es un sistema gestor de bases de datos muy conocido y ampliamente usado por su simplicidad y notable rendimiento. Tiene libre distribución en Internet bajo licencia GPL, es válido utilizarlo tanto para elaborar

aplicaciones comerciales como de entretenimiento. Debido a su facilidad de uso además del reducido tiempo de puesta en marcha permite contar con un elevado grado de estabilidad y rápido desarrollo.

Una de las desventajas de MySQL es que debido a que es de código abierto no es posible contratar algún tipo de mantenimiento o soporte. Si se da el caso de que el manejador de bases de datos falla, el usuario deberá resolver el problema sin poder contar con ayuda técnica de los desarrolladores del software. (15) Debido a las desventajas que posee este sistema gestor de bases de datos no fue seleccionado para su utilización.

1.6.3 SQL Server

Actualmente MySQL Server es una de las bases de datos de código abierto más utilizadas en el mundo, entre sus características se pueden destacar las siguientes (14):

- Gran rendimiento consistente así como alta fiabilidad y facilidad de uso.
- Múltiples motores de almacenamiento, permitiendo al usuario escoger la que sea más adecuada para cada tabla de la base de datos.
- Diseñado para servicios informativos de gran envergadura con un costo bajo.
- Funciona sobre múltiples plataformas.
- Su bajo consumo lo hace apto para ser ejecutado en una máquina con escasos recursos sin ningún problema.
- En cuanto a seguridad, ofrece un sistema de contraseñas y privilegios seguro mediante la verificación basada en el host y el tráfico de contraseñas está cifrado al conectarse a un servidor. Soporta gran cantidad de datos, tiene bases de datos de hasta 50 millones de registros.
- Se permiten hasta 64 índices por tabla. Cada índice puede consistir desde 1 hasta 16 columnas o partes de columnas. El máximo ancho de límite son 1000 bytes.
- Puede provocar problemas de integridad en entornos de alta concurrencia en la modificación.

Además de todas las características planteadas anteriormente, es importante destacar que es un software propietario, por lo que no fue la herramienta que se empleó en el desarrollo del mercado de datos que se realizó en este trabajo.

1.6.4 PostgreSQL versión 8.4

Contiene una gran cantidad de mejoras para que la administración, consulta y programación en PostgreSQL sea más fácil y productivo que versiones anteriores. Es una versión libre que continúa con el rápido desarrollo de las bases de datos de código abierto a escala mundial aumentando el rendimiento de las aplicaciones. Entre las principales mejoras se encuentran (16):

- Restauración de base de datos en procesos paralelos que acelera la recuperación de un respaldo hasta 8 veces.
- Privilegios por columna que permiten un control más granular de datos confidenciales.
- Posee nuevas herramientas de monitoreo de consultas que le otorgan a los administradores mayor información sobre la actividad del sistema.
- Mejora el rendimiento de las aplicaciones.
- Utiliza la licencia BSD (Berkeley Software Distribution) la cual sólo requiere que el código fuente licenciado mantenga la información de derechos de autor y licenciamiento. Esta licencia certificada por la OSI⁹ es altamente apreciada por su flexibilidad y por ser amigable para los negocios, por lo cual atrae a gran cantidad de usuarios.

Luego de analizar varios sistemas gestores de bases de datos, se llegó a la conclusión de utilizar PostgreSQL 8.4 porque ofrece todos los requisitos necesarios para la implementación eficiente del mercado de datos. Es muy estable y segura, además, la ONE desea llevar todas sus bases de datos a una plataforma libre.

1.7 PgAdmin III versión 1.10 (Herramienta para la administración de bases de datos)

- PgAdmin III es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL.
- Utiliza como licencia Open Source.

⁹Modelo de Interconexión de Sistemas Abiertos, creado por la ISO (Organización Estándar Internacional), posee varios niveles con el objetivo de establecer una relación entre desarrollador, hardware, aplicación, así como la arquitectura de la misma.

- Está escrita en el lenguaje de programación C++ usando la librería gráfica multiplataforma wxWidgets lo que permite que se pueda usar en distintos sistemas operativos como Linux, FreeBSD, Solaris, Mac OS X y Windows.(17)
- Es capaz de gestionar versiones a partir de PostgreSQL 7.3 ejecutándose en cualquier plataforma.

PgAdmin III está diseñado para responder a las necesidades de los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. El interfaz gráfico soporta todas las características de PostgreSQL además de ofrecer facilidades de administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor, un agente para lanzar scripts programados, soporte para el motor de replicación Slony-I y mucho más. La conexión al servidor puede hacerse mediante TCP/IP o Unix Domain Sockets (en plataformas Unix) y puede encriptarse mediante SSL¹⁰ para mayor seguridad.

En la versión 1.10 de esta herramienta se insertan varias mejoras como son:(18)

- Nuevas características importantes para la herramienta de consulta.
- Permite construir consultas de forma gráfica las cuales se pueden crear con el ratón permitiendo su uso de manera más fácil.
- Admite escribir scripts para ejecutar consultas, ya que posee otras secuencias de comandos.
- Tiene un mayor respaldo para las bases de datos, colas de recursos, las tablas externas y tablas con particiones son compatibles.

Conclusiones

Hasta este punto de la investigación se realizó un estudio acerca de las tecnologías de almacenamiento en varias partes del mundo, la Oficina Nacional de Estadísticas (ONE) es la entidad encargada de captar la información estadística de la educación en Cuba y agruparla por distintos indicadores. Para tener un control centralizado de los datos se decidió desarrollar un mercado de datos porque permite manejar la información de un área de la empresa, específicamente en el departamento de Educación en la Oficina Nacional de Estadísticas.

¹⁰ Secure Sockets Layer en español Protocolo de Capa de Conexión Segura (SSL) es un protocolo criptográfico que brinda seguridad a las comunicaciones Internet como navegar en la web.

Para el desarrollo de la propuesta se escogió la metodología de Kimball incluyéndole los casos de uso como guías en el proceso de desarrollo del software respetando así las necesidades de desarrollo y políticas de la UCI. Por su simplicidad para representar gráficamente la relación entre los hechos y las dimensiones se eligió el modelo en estrella para realizar el diseño del mercado de datos; quedando seleccionada como herramienta de modelado el Visual Paradigm UML 6.4 Enterprise Edition. El sistema gestor de base de datos elegido fue el PostgreSQL versión 8.4 por ser una herramienta libre y para su administración el PgAdmin III versión 1.10.

CAPÍTULO 2 Análisis y Diseño

Introducción

Para la elaboración del mercado de datos Indicadores de la Educación son necesarias las fases de análisis y diseño. El análisis es la base fundamental para el desarrollo del mercado de datos, ya que a partir del mismo se sientan las bases para los posteriores procesos de diseño e implementación; resulta una tarea compleja partiendo de que se necesita un estudio del proceso del negocio que se pretende informatizar para entender fácilmente lo que el usuario necesita. Posterior a la fase de análisis se inicia la fase de diseño que debe apoyar la solución de las necesidades planteadas. En ella se elabora el diagrama de diseño de la solución donde se definen las relaciones entre los hechos y las dimensiones.

2.1 Definición del Negocio

Durante la etapa de análisis se realizó una descripción de la ONE que es la entidad cliente. Esta oficina fue creada para proponer, organizar y ejecutar según corresponda, la aplicación de la política estatal en materia de estadísticas del país. Su misión fundamental es garantizar la producción de datos con calidad a través del Sistema Estadístico Nacional ejerciendo una adecuada dirección, ejecución y control de la captación de las cifras económicas y sociales, así como su adecuada difusión en concordancia con las necesidades financieras cubanas. Tiene como visión construir un sistema estadístico profesional apto para responder con calidad y oportunidad a las necesidades de información estadística cubana para enfrentar sus metas del desarrollo económico y social y su adecuado reflejo internacional. En el Sistema Estadístico Nacional participan coordinadamente los órganos, organismos, instituciones y entidades que elaboran estadísticas en Cuba y está integrado por 3 subsistemas:

- Sistema de Información Estadística Nacional (SIEN): está dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos de los más altos niveles del Estado y el Gobierno además de informar a los organismos internacionales. La ONE controla la información relacionada con la población y las entidades siendo la productora a través de su red.
- Sistema de Información Estadística Complementaria (SIEC): está dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos de los organismos a los efectos del control administrativo de sus entidades. La ONE revisa y ejerce su control metodológico

aunque no es la productora a través de su red. Controla la información de las entidades subordinadas y no subordinadas.

- Sistema de Información Estadística Territorial (SIET): dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos del Gobierno en el territorio y otros actores locales.

Entre las principales funciones y atribuciones de esta entidad se encuentran:

- Organizar y aprobar la producción de las estadísticas centralizadas y territoriales.
- Dirigir metodológicamente la actividad estadística de los órganos, organismos, instituciones y entidades estatales.
- Normar, velar y dictaminar acerca del funcionamiento de las estadísticas complementarias.
- Definir las atribuciones y responsabilidades de otros productores de estadísticas de interés nacional y su lugar en el sistema estadístico del país.
- Aprobar las normas metodológicas y de clasificación que se utilizan en las estadísticas centralizadas, territoriales y complementarias.
- Identificar las unidades de observación estadística y captar a través de la red territorial la información correspondiente a las estadísticas centralizadas.
- Dirigir los procesos y ejecutar según corresponda, los censos económicos y de población y encuestas económicas o sociales de carácter nacional. Aprobar la realización de este tipo de investigaciones estadísticas en el país.
- Llevar los registros estatales de empresas, unidades presupuestadas y otras entidades.
- Supervisar el trabajo estadístico de organismos y entidades organizando auditorías para velar por la autenticidad de la información.
- Centralizar y emitir las estadísticas del país.

El flujo de información de la ONE es efectuado mediante de una red de transmisión de datos por todo el país con más de 484 millones de reportes al año a través de las oficinas en los municipios, donde se digitan y revisan para ser enviadas posteriormente a la ONE donde procesan y emiten las informaciones

que se obtienen a través de los centros informantes, mediante modelos y encuestas de captación de información. También es recogida, digitada y procesada según la información de la población.

La información estadística de la educación en la ONE proviene del Ministerio de Educación (MINED) y del Ministerio de Educación Superior (MES), estos al igual que otros organismos le brindan la información de todos los niveles de enseñanza desde el triunfo de la Revolución hasta la actualidad. La ONE procesa esta información a través del Sistema de Información Estadística Complementaria (SIEC) cuyo funcionamiento fue explicado con anterioridad.

2.2 Temas de Análisis

Los temas de análisis brindan aportes para el desarrollo del mercado de datos ya que la realización de los mismos orienta el avance del cumplimiento de las tareas planteadas garantizando la utilidad y el éxito del diseño de las estructuras. La solución propuesta se orienta a las series¹¹ de información que realiza la ONE. Los desarrolladores de la presente investigación definieron un tema de análisis que es: la educación, el cual reúne todos los temas de análisis en concordancia con las necesidades de información identificadas para la realización del mercado de datos

La educación es una de las esferas sociales más importantes del país. Desde 1959 con el triunfo de la Revolución una de las primeras metas fue el desarrollo educacional de la población, a través de la campaña de alfabetización toda la población aprendió a leer y escribir. Hoy en día se mejoran los programas educativos y se crean nuevos centros para la preparación ciudadana con un grupo creciente de integrantes en los distintos niveles educativos, que comprenden desde los círculos infantiles hasta la enseñanza superior.

2.3 Reglas del Negocio

Otra de las tareas que se realizó fue la definición de las reglas del negocio. Estas constituyen un eslabón fundamental en la construcción del mercado de datos pues permiten realizar el trabajo de una forma más sencilla, especifican los formatos que deben cumplirse e indican cómo deben calcularse algunas medidas

¹¹ Documentos en formatos Excel donde la ONE almacena los datos de los indicadores, en este caso se hace referencia a los indicadores de la educación.

del almacén. En el mercado de datos indicadores de la educación se definieron cinco reglas del negocio. A continuación se muestra cada una de ellas:

Tasa neta de matrícula

La tasa neta de matrícula es el número de alumnos, del grupo de edad oficial de la enseñanza matriculados en ese mismo nivel expresado como porcentaje del total de la población de ese grupo de edad.

Se calcula mediante la siguiente fórmula $TNM = (Mn, e / Pe) * 100$

Donde: TNM: Tasa neta de matrícula en el nivel.

Mn, e: Matrícula en el nivel "n" con edades en el rango correspondiente.

Pe: Población del rango de edad que corresponde al nivel "n".

Tasa bruta de matrícula

La tasa bruta de matrícula es el número de alumnos, matriculados en un determinado nivel de educación, independientemente de su edad, expresado como porcentaje del grupo de población del grupo de edad teórica para ese mismo nivel.

Asistencia promedio a los círculos infantiles

Es la cantidad media de asistencia de los niños matriculados. La asistencia promedio se calcula sumando diariamente la asistencia de los niños matriculados y dividiendo el resultado obtenido entre los días del mes (excepto los domingos, días feriados y días cerrados¹²). A partir del promedio mensual se calculan los promedios trimestrales, semestrales y anuales.

Índice de paridad de género (IPG)

El índice de paridad de género indica la proporción entre alumnas y alumnos para un determinado indicador según el valor que alcance en caso de que:

- IPG=1 indica la paridad entre los dos sexos, o sea que hay la misma cantidad de ambos sexos.

¹² Días en los cuales no se abren los círculos infantiles debido a alguna afectación.

- $0 > IPG < 1$ significa que el valor del indicador para la población masculina es mayor que para la femenina.
- $IPG > 1$ representa que el valor del indicador para la población femenina es mayor que para la masculina.

Curso escolar

La regla curso escolar indica el formato en que debe estar definido el nombre del curso, este se puede definir de dos formas:

- 1- Año de inicio-año fin
- 2- Año de inicio/año fin

Cantidad de habitantes

Es una medida física del almacén que pertenece a otro mercado de datos pero que es necesaria para realizar el cálculo de la medida per cápita perteneciente al mercado de datos en cuestión.

La per cápita se calcula mediante la fórmula:

$$\text{per cápita} = \text{presupuesto de educación} / \text{cantidad de habitantes}$$

Donde:

Presupuesto de educación: es la cantidad de dinero valorado en millones de pesos de un determinado año.

Cantidad de habitantes: es el número de personas de ese mismo año.

La unidad de medida en que se expresa es (peso/habitantes).

2.4 Necesidades de los usuarios

Para el desarrollo del análisis en el proceso del negocio es necesario conocer lo que buscan y necesitan los usuarios. Es necesaria la implicación de los mismos en los proyectos ya que esto aumenta las probabilidades de que salga bien o no, además, se necesita conocer qué es lo que en realidad quiere el usuario para poder satisfacer sus necesidades, lo cual se realiza a partir de los requisitos de información.

2.5 Requisitos de información

Los requerimientos de información pretenden satisfacer las necesidades de informativas de los usuarios definidas a partir de la comparación entre las necesidades de información y las reglas del negocio con los

elementos disponibles en las fuentes. Son las funcionalidades que el sistema debe permitir que se cumplan. A continuación se muestran todos los requisitos de información identificados:

- RI.1 Obtener la matrícula final de los círculos infantiles.
- RI.2 Obtener la asistencia promedio de los círculos infantiles.
- RI.3 Obtener la cantidad de escuelas por educaciones.
- RI.4 Obtener la cantidad de escuelas por educaciones y provincias.
- RI.5 Obtener la cantidad de personal docente por educaciones.
- RI.6 Obtener la cantidad de personal docente frente al aula por educaciones.
- RI.7 Obtener la cantidad de personal docente frente al aula por educaciones y provincias.
- RI.8 Obtener la matrícula inicial por educaciones.
- RI.9 Obtener la matrícula inicial por educaciones y provincias.
- RI.10 Obtener la matrícula inicial (hembras) por educaciones.
- RI.11 Obtener la matrícula inicial (hembras) por educaciones y provincias.
- RI.12 Obtener el comportamiento de las tasas brutas y netas de matrícula.
- RI.13 Obtener la cantidad de becarios por educaciones.
- RI.14 Obtener la cantidad de seminternos por educaciones.
- RI.15 Obtener la cantidad de graduados por educaciones.
- RI.16 Obtener la cantidad de graduados por educaciones y provincias.
- RI.17 Obtener la cantidad de graduados (hembras) por educaciones.
- RI.18 Obtener la cantidad de graduados (hembras) por educaciones y provincias.
- RI.19 Obtener la matrícula inicial de la educación técnica y profesional por grupo de especialidades.
- RI.20 Obtener los graduados de la educación técnica y profesional por grupo de especialidades.
- RI.21 Obtener la matrícula inicial por grupo de especialidades de la educación superior.

- RI.22 Obtener la matrícula en los cursos dirigidos (enseñanza a distancia) por especialidades.
- RI.23 Obtener la matrícula inicial por organismos, centros y tipo de estudio.
- RI.24 Obtener el comportamiento de nuevo ingreso por grupo de especialidades de la educación superior.
- RI.25 Obtener la cantidad de graduados por ramas de la ciencia
- RI.26 Obtener la matrícula inicial de la educación superior por ramas de la ciencia.
- RI.27 Obtener la cantidad de graduados en la educación superior.
- RI.28 Obtener la cantidad de graduados en los cursos dirigidos (enseñanza a distancia) por especialidades.
- RI.29 Obtener la cantidad de graduados por organismos, centros y tipo de estudio de la educación superior.
- RI.30 Obtener el comportamiento de la educación de postgrado.
- RI.31 Obtener el comportamiento del presupuesto de la educación.

2.6 Requisitos Funcionales

Los requisitos funcionales son aquellos que definen la función que el sistema va a llevar a cabo, su correcta obtención puede asegurar el éxito desarrollo del mercado de datos. Se identificaron en sentido general dos requisitos que son aplicados a las 33 series facilitadas por la ONE.

RF1. Extraer datos de los círculos infantiles.

RF2. Cargar y transformar datos de los círculos infantiles.

Y así sucesivamente para todas las series de los indicadores de la educación.

2.7 Requisitos No Funcionales

Fueron seleccionados una serie de requisitos no funcionales basados en las características del sistema que se va a realizar, tienen que ver con características que de una u otra forma pueden limitar al sistema, como por ejemplo la seguridad, el rendimiento, la fiabilidad, entre otros.

Software

RNF.1 El sistema debe ser multiplataforma.

RNF.2 Tener instalado en el servidor PostgreSQL para realizar las consultas a la bases de datos.

Hardware

RNF.3 Computadora/Procesador: con un procesador Pentium

RNF.4 Memoria: 512 MB de RAM (1 GB ideal)

RNF.5 Disco duro: 10 GB (mínimo)

RNF.6 Hardware de Red (Conectividad): para el acceso a los datos en el almacén, para la emisión de los datos y los servicios a las estaciones clientes.

Tiempo medio de fallos

RNF.7 El tiempo medio de fallos del proceso de integración es de 30 días.

Tiempo medio de reparación

RNF.8 El tiempo medio de reparación en el proceso de integración depende en gran medida de la categoría del error o falla. El tiempo estimado es de 12 horas.

Máximo de errores

RNF.9 La cantidad de errores en el proceso de integración define la calidad de los datos que se están almacenando; es por ello que es crítico definiéndose 1 errores/puntos de función.

Usabilidad

RNF.10 Dominar conocimientos básicos sobre Base de datos.

RNF.11 Tener conocimiento en análisis y tratamiento de la información.

RNF.12 Tiempo de 15 días para que usuarios normales y avanzados sean productivos operando el sistema.

RNF.13 Los usuarios deben recibir capacitación de las herramientas utilizadas en la confección de las tablas de salida y cómo interpretar la información de los mismos.

Disponibilidad

RNF.14 El área de almacenamiento del proceso de ETL (Extracción, transformación y carga) debe estar disponible las 12 horas del día lo que representa 50% horas de uso.

RNF.15 El sistema de integración será accedido para su mantenimiento 1 vez por mes. En este plazo de mantenimiento se validarán las estructuras de auditoría y se establecerán estrategias que permitan clasificar posibles errores y darles solución en caso de ser posible.

RNF.16 El sistema debe estar disponible 100% entre las 8:00 am y las 5:00 pm de lunes a viernes.

Utilización de Recursos

RNF.17 El tiempo de respuesta debe ser en aproximadamente 5 segundos.

RNF.18 El sistema debe permitir la concurrencia de usuarios sin que se afecte el tiempo de respuesta de las consultas.

Capacidad

RNF.19 En el proceso de integración solo estará conectado un solo usuario que tendrá la tarea de monitorear el proceso de integración de datos.

2.8 Casos de Uso del Sistema

Los casos de uso del sistema son una forma visual de representar información, ayudaron a describir lo que hace el sistema relacionado con el usuario. Tienen un escenario específico donde se representan acciones efectuadas dentro del sistema, inicializadas por un actor el cual es un trabajador del negocio. Para su mejor comprensión se hizo una representación gráfica que contiene los casos de uso y su relación con los usuarios que interactúan con él. Para la realización del mercado de datos fue necesario identificar casos de uso funcionales y de información, donde se realizó la especificación de cada uno de estos. (Ver Anexo 1)

2.8.1 Casos de Uso de Información

Los casos de uso de información se agruparon por el tipo de información que maneja la ONE y según las necesidades de información de los usuarios. (Ver figura 1)

Los actores o roles que van a interactuar con el sistema son:

Administrador: persona que dirige la ONE o algún departamento dentro de la misma, tiene permisos de consultar y modificar la información del sistema.

Consultor: persona que tendrá acceso a consultar la información del mercado de datos, puede ser un especialista de la ONE, un funcionario del gobierno o una persona externa que también necesite examinar los datos.

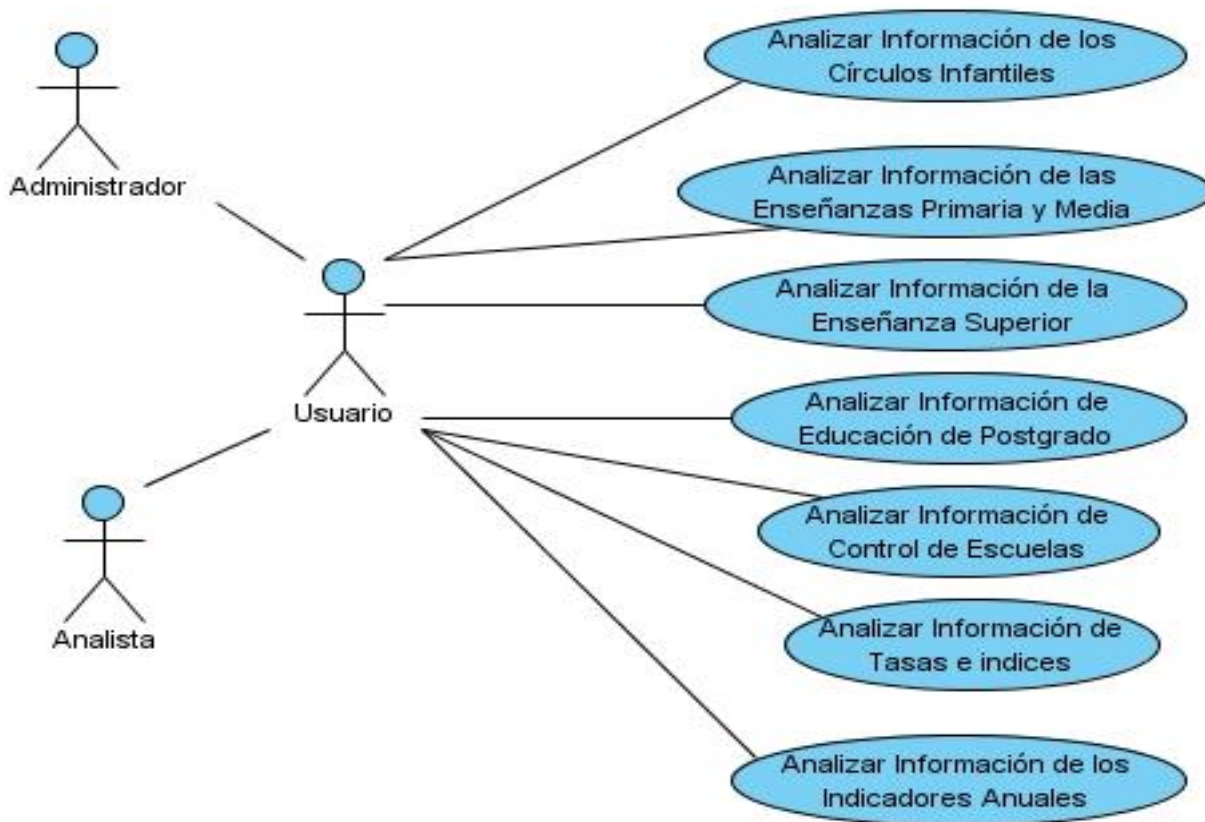


Figura 1 Diagrama de Casos de Uso de Información.

2.8.2 Casos de Uso funcionales

Los datos estadísticos de la educación se encuentran almacenados en series, a las cuales se les realizó el proceso de ETL (Extracción, Transformación y Carga). Para ello se definieron casos de uso funcionales que describen cómo extraer, cargar y transformar los datos del almacén. Este proceso se aplicó a cada

uno de los 33 ficheros que contienen los datos, obteniéndose un total de 66 descripciones de casos de uso. El usuario que interactúa con los casos de uso es el especialista de ETL¹³.

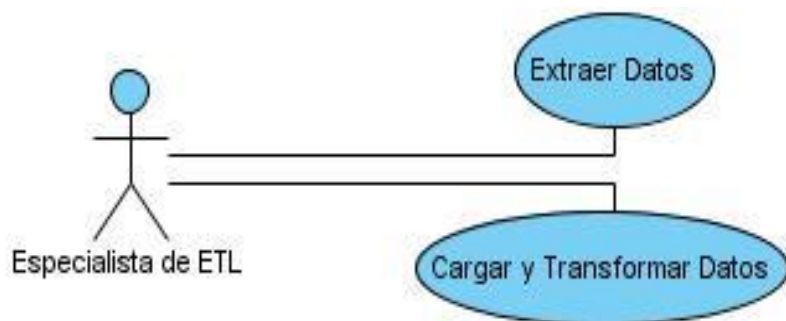


Figura 2 Diagrama de Casos de Uso Funcionales

2.9 Matriz Dimensional (BUS)

El primer paso en la etapa de diseño de un mercado de datos es la creación de la matriz dimensional o matriz BUS. En ella se refleja la relación que existe entre los hechos y sus dimensiones asociadas. Para esto es preciso definir inicialmente la lista de hechos y la de dimensiones.

Lista de hechos

- H1. Indicadores anuales
- H2. Círculos infantiles
- H3. Control de escuelas
- H4. Enseñanzas primaria y media
- H5. Estudiantes de las enseñanzas primaria y media
- H6. Personal de las enseñanzas primaria y media
- H7. Tasas e índices
- H8. Enseñanza superior

¹³ Encargado de realizar el proceso de extracción, transformación y carga del mercado de datos.

H9.Control de nuevo ingreso y matrícula en cursos dirigidos

H10.Educación de postgrado

H11.Enseñanza técnica y profesional

Lista de dimensiones

1. Tiempo por años
2. Tiempo por curso escolar
3. Años ci
4. Educaciones
5. Geografía
6. Zonas ubicación
7. Estudiantes
8. Personal
9. Especialidades educación técnica y profesional
10. Estudiantes educación técnica y profesional
11. Edad establecida
12. Cines
13. Organismos
14. Centros universitarios
15. Tipo de estudio
16. Especialidades universitarias
17. Postgrado

Tabla 2 Matriz Dimensional

<i>Matriz Dimensional</i>																	
Áreas de Análisis	Dimensiones																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
H1	X																
H2	X		X														

H3		X				X										
H4		X		X	X	X										
H5		X		X	X	X	X									
H6		X		X	X	X		X								
H7		X									X	X				
H8		X										X	X	X	X	
H9		X													X	
H10	X															X
H11		X							X	X						

2.10 Modelo de Datos

Dimensiones

Las dimensiones son las características de los hechos que permiten su análisis posterior en el proceso de toma de decisiones. Las relaciones entre ellas crean jerarquías que permiten un análisis jerárquico de los datos. En este trabajo se obtuvieron 17 dimensiones, un hecho puede estar relacionado con una o varias de ellas. A continuación se describirán cada una de estas.

Dimensión tipo de estudio

Controla la información referente al tipo de estudio según el horario en el que se recibe y la forma en que llega a los estudiantes.

Atributos asociados

- id_tipo_Est: identificador del tipo de estudio de la dimensión.
- Diurno: si los estudiantes reciben las clases por el día.
- Vespertino/Nocturno: si los estudiantes reciben las clases por las tardes o las noches.
- Por Encuentros: si los estudiantes reciben las clases mediante encuentros cada cierto tiempo.
- Enseñanza a distancia: si los estudiantes reciben las clases a distancia, es decir mediante correo electrónico, a través de foros, entre otros.

Dimensión organismos

Representa los diferentes organismos que contienen centros educacionales donde se estudian diferentes carreras técnicas o universitarias.

Atributos asociados

- id_organismo: identificador de la dimensión organismos.
- MINED: Ministerio de Educación
- MES: Ministerio de Educación Superior
- OTROS: hace referencia a otros organismos que también controlan la educación como son el Ministerio de Salud Pública (MINSAP), el INDER, entre otros excluyendo las escuelas de las Fuerzas Armadas Revolucionarias (FAR).

Dimensión zona ubicación

Se presenta la información referente a la zona donde se encuentran los centros educacionales.

Atributos asociados

- id_zona ubicación: identificador de la zona.
- Rural
- Urbano
- En el campo

Dimensión geografía

A lo largo de todo el territorio nacional existen centros educacionales, debido a esto en esta dimensión se encuentran todas las provincias del país.

Atributos asociados

- id_provincias: identificador de la dimensión geografía.
- Pinar del Río
- Ciudad Habana
- La Habana
- Matanzas
- Villa Clara
- Cienfuegos

- Ciego de Ávila
- Sancti Spíritus
- Camagüey
- Holguín
- Las Tunas
- Granma
- Santiago de Cuba
- Guantánamo
- La Isla de la Juventud

Dimensión estudiantes

Representa los estudiantes de los diferentes niveles educativos.

Atributos asociados

- id_estudiantes: identificador de estudiantes.
- Graduados: son alumnos que han finalizado satisfactoriamente los estudios correspondientes a un nivel o tipo de educación dado. Son los aprobados del grado o año de estudio terminal de un nivel o tipo de educación.
- Becarios: alumnos que reciben durante el tiempo que permanecen matriculados, educación, alimento, vestuario y albergue.
- Seminternos: alumnos que reciben durante el tiempo que permanecen matriculados, educación, alimento, en ocasiones vestuario y no están albergados. En la educación superior se calcula por la capacidad de comensales.

Dimensión personal

Número de personas oficialmente habilitadas en régimen de dedicación plena o parcial para orientar y alcanzar la experiencia de aprendizaje de los alumnos, cualquiera que sea su calificación profesional o el modo de proceso de aprendizaje: presencial o a distancia.

Comprende los maestros y profesores, directores, subdirectores, auxiliares pedagógicas y bibliotecarios. En la educación superior comprende el personal de nivel superior con categoría docente; no incluye los alumnos ayudantes ni a los instructores no graduados.

Atributos asociados

- id_personal: identificador del personal de la educación.
- Personal docente frente a aulas por educaciones: personal docente que cumple funciones de maestro o profesor, incluye a los directores y subdirectores independientemente que impartan clases o no y al personal contratado.
- Personal técnico profesional: personal técnico y profesional que cumple funciones de auxiliar en los círculos infantiles.

Dimensión educaciones

Existen tres niveles de enseñanza cada uno cuenta con diversos centros, así como las clasificaciones por año dentro de cada nivel.

Atributos asociados

- id_educaciones: identificador de los distintos niveles educativos de la educación.
- Nivel Primario
- Preescolar
- Primaria
- Nivel Medio
- Secundaria
- Preuniversitario
- Vocacional
- Formación de personal docente
- Técnico y Profesional
- Escuelas de Oficios u Obreros Calificados
- Escuelas de Trabajadores Sociales
- Juvenil
- Facultad Preparatoria de Idiomas

- Especial
- Nivel Superior
- Superior
- Sedes Universitarias
- Otras

Dimensión especialidades universitarias

La educación superior consta de varias especialidades en diversas ramas como son: la cultura, economía, ciencias exactas entre otras, éstas se encuentran dentro de esta dimensión.

Atributos asociados

- id_especialidades: identificador de las especialidades
- Ciencias Técnicas o Tecnología
- Ciencias Naturales
- Ciencias Médicas
- Ciencias Agropecuarias
- Ciencias Económicas
- Ciencias Sociales y Humanistas
- Ciencias Matemáticas
- Computación
- Economía
- Información Científica Técnicas y Bibliotecología
- Pedagogía
- Deporte y Cultura Física
- Arte
- Facultad Preparatoria
- Enseñanza dirigida
- Escuela del PCC
- Planificación y Economía Nacional
- Derecho

- Turismo
- Historia
- Contabilidad y Finanzas

Dimensión centros universitarios

La dimensión centros universitarios contiene todos los centros que existen en el país, a continuación se muestran cada uno de ellos.

Atributos asociados

- id_centros: identificador de la dimensión centros universitarios.
- Centro Universitario Pinar del Río
- Centro Universitario Matanzas
- Universidad de la Habana
- Universidad de Oriente
- Universidad Central de las Villas
- Universidad de Camagüey
- Centro Universitario José Antonio Echeverría (CUJAE)
- Instituto Superior Técnico de Cienfuegos
- Instituto Superior de Ciencias Agropecuarias
- Instituto Superior Agrícola de Ciego de Ávila
- Instituto Superior Técnico de Holguín
- Instituto Superior de Ciencias Agropecuarias Bayamo
- Instituto Minero Metalúrgico Moa
- Instituto Superior Pedagógico Pinar del Río
- Instituto Superior Pedagógico Ciudad de la Habana
- Instituto Superior Pedagógico Lenguas Extranjeras
- Instituto Superior Pedagógico Matanzas
- Instituto Superior Pedagógico Villa Clara
- Instituto Superior Pedagógico Camagüey
- Instituto Superior Pedagógico Holguín

- Instituto Superior Pedagógico Granma
- Instituto Superior Pedagógico Santiago de Cuba
- Instituto Superior Pedagógico Guantánamo
- Instituto Superior de Ciencias Médicas de La Habana
- Instituto Superior de Ciencias Médicas de Villa Clara
- Instituto Superior de Ciencias Médicas de Camagüey
- Instituto Superior de Ciencias Médicas de Santiago de Cuba

Dimensión especialidades educación técnica y profesional

Representa a las diferentes especialidades de la educación técnica y profesional, estas pueden ir variando con el tiempo y se le pueden agregar otras.

Atributos asociados

- id_especialidades: el identificador de las especialidades de la educación técnica y profesional
- Geología
- Minería y metalurgia
- Energética
- Construcción de maquinarias industria azucarera
- Química y alimentos
- Electrónica, automatización y comunicación
- Transporte
- Construcción
- Producción agropecuaria
- Economía
- Salud pública
- Arte

Dimensión años ci

Esta dimensión representa todos los años en que se encuentran los niños matriculados en los círculos infantiles, estos son:

Atributos asociados

- id_años: identificador de la dimensión.
- 1er Año
- 2do Año
- 3er Año
- 4to Año
- 5to Año
- 6to Año

Dimensión estudiantes educación técnica y profesional

Contiene los estudiantes que se encuentran en la educación técnica y profesional.

Atributos asociados

- id_est: identificador de la dimensión
- Técnico medio
- Obrero calificado

Dimensión tiempo por año

Mediante esta dimensión se almacena el tiempo por año, ejemplo: 1999, 2000, 2001.

Atributos asociados

- id_tiempo: identificador del tiempo por año
- Años

Dimensión tiempo por curso escolar

Contiene el tiempo de los cursos escolares, por ejemplo: 1990-1991, 2002-2003.

Atributos asociados

- id_tiempo: identificador del tiempo por curso escolar.
- Curso_escolar

Dimensión edad establecida

Contiene la edad establecida de los estudiantes de un nivel de enseñanza determinado.

Atributos asociados

- Id: identificador de la edad.
- rango_edad: está comprendido de la siguiente manera: 3 – 5; 6 – 11 y de 12 – 17 años de edad.

Dimensión Cines

Esta dimensión recoge los nomencladores de los distintos cines que existen y a su vez son diferentes niveles de enseñanza.

Atributos asociados

- id_cine: identificador de cine
- Cine 0: Nivel pre primario
- Cine 1: Nivel primario
- Cine 2 y 3: Nivel medio

Dimensión Posgrado

Representa los cursos de posgrados existentes.

Atributos asociados

- Id_posgrados: identificador de la dimensión
- Entrenamientos
- Diplomados
- Maestrías y especialidades
- Doctorados

Medidas

Las medidas son las propiedades de los hechos, pueden estar físicas en el almacén o pueden ser calculables. Casi siempre son numéricas y se emplean para el análisis. En este caso se encontraron 33 medidas para el desarrollo del mercado de datos.

1. Cantidad de personal técnico educacional en los círculos infantiles
2. Cantidad de madres beneficiadas en los círculos infantiles
3. Capacidad al final del año en los círculos infantiles

4. Asistencia promedio en los círculos infantiles
5. Matrícula final de los círculos infantiles
6. Cantidad de niños de 0-5 años matriculados por 10000 en esas edades
7. Cantidad de círculos infantiles
8. Cantidad de escuelas
9. Cantidad de graduados de las enseñanzas primaria y media
10. Matrícula Inicial de las enseñanzas primaria y media
11. Matrícula Inicial hembras de las enseñanzas primaria y media
12. Cantidad de personal de las enseñanzas primaria y media
13. Cantidad de estudiantes de las enseñanzas primaria y media
14. Cantidad de graduados de las enseñanzas primaria y media
15. Cantidad de graduados hembras de las enseñanzas primaria y media
16. Cantidad de Becarios de la enseñanza superior
17. Cantidad de graduados de la enseñanza superior
18. Cantidad de mujeres graduadas de la enseñanza superior
19. Cantidad de personal docente de la enseñanza superior
20. Matrícula inicial de la enseñanza superior
21. Matrícula en los cursos dirigidos
22. Cantidad de nuevo ingreso de la enseñanza superior
23. Cantidad de habitantes
24. Gasto público en educación
25. Percápita
26. Cantidad de cursos de posgrados
27. Cantidad de participantes
28. Matrícula de las tasas e índices
29. Población en edad escolar
30. Cantidad de niñas
31. Cantidad de niños
32. Matrícula inicial de la enseñanza técnica y profesional
33. Cantidad de graduados de la enseñanza técnica y profesional

A partir de la identificación de las dimensiones, los hechos y las medidas del almacén se obtuvo el diseño de la solución que a continuación se muestra, para tener una visión de la estructura lógica del mercado de datos.

2.11 Modelo de diseño

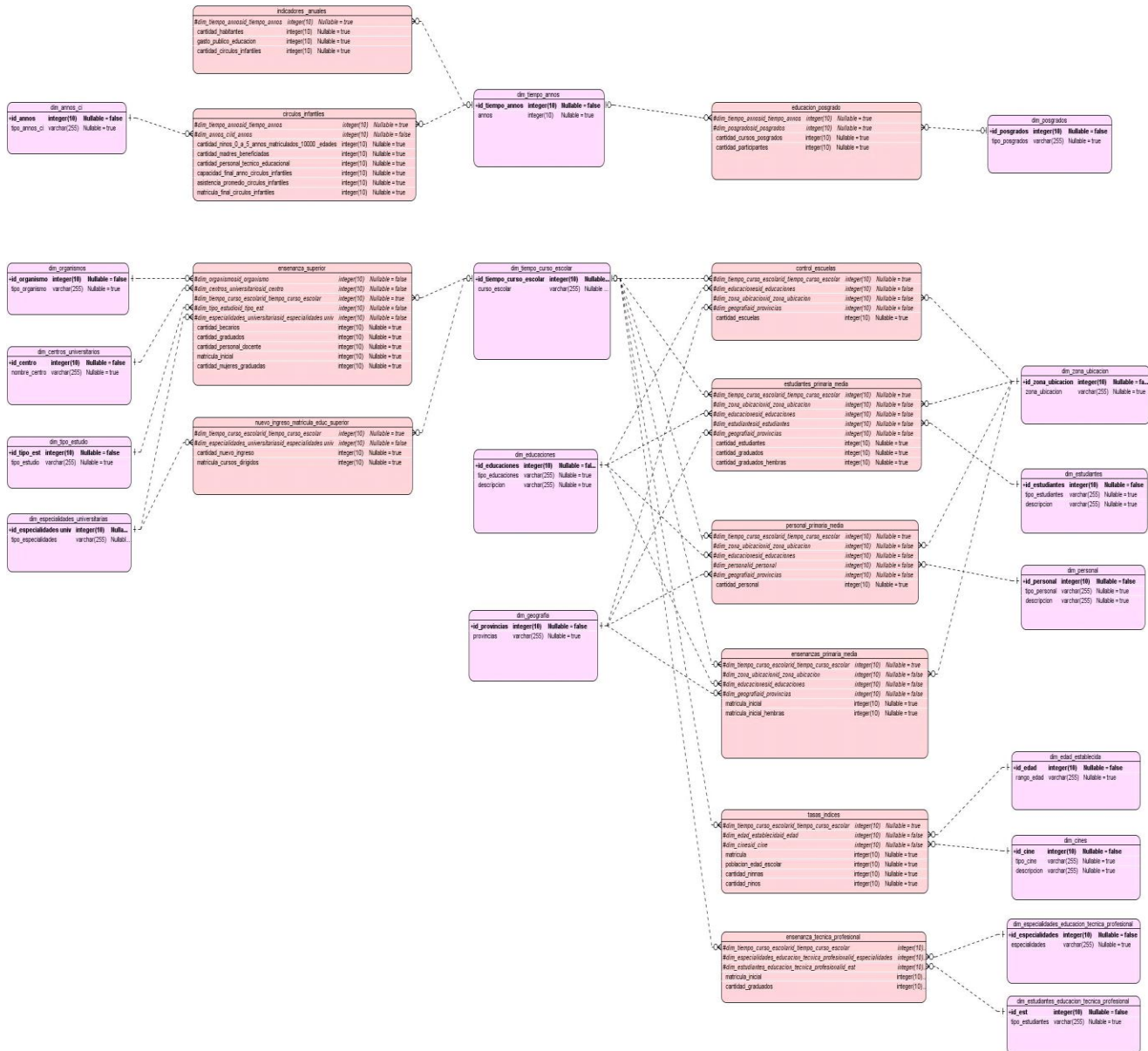


Figura 3 Modelo de diseño

Conclusiones

Para el mercado de datos se definió como tema de análisis la educación. Se especificaron seis reglas del negocio y fueron identificadas las necesidades del cliente las cuales se corresponden con los 31 requisitos de información. Se realizó la descripción de los requerimientos necesarios para cumplir con la demanda del cliente y garantizar el correcto funcionamiento del sistema. Se especificaron los casos de uso del sistema y se construyeron sus diagramas correspondientes. Sentadas las bases para iniciar la etapa de diseño, se construyó la matriz dimensional en la que se relacionaron las once tablas de hechos con las diecisiete dimensiones identificadas.

CAPÍTULO 3. Implementación y pruebas

Introducción

La implementación es la fase donde se crean las estructuras del mercado de datos, se definen los privilegios que se le asignan a cada usuario y se realizan los procesos de extracción, transformación y carga de los datos. De estos tres procesos sólo se abordará el último que es hasta donde quedó definido el alcance de la investigación. Con el objetivo de validar que la solución funcione con calidad se realizan las pruebas de implantación al mercado de datos y se aplican las listas de chequeo para validar los artefactos de las fases de análisis y diseño.

3.1 Modelos de Datos Físico

Un modelo de datos permite describir los elementos que intervienen en una realidad o en un problema dado y la forma en que se relacionan dichos elementos entre sí. Es un lenguaje utilizado para la descripción de una base de datos. Generalmente, permite describir las estructuras de datos de la base (el tipo de los datos que incluye la base y la forma en que se relacionan), las restricciones de integridad (las condiciones que los datos deben cumplir para reflejar correctamente la realidad deseada) y las operaciones de manipulación de los datos (agregado, borrado, modificación y recuperación de los datos de la base).

El modelo de datos físico no es más que estructuras de datos de bajo nivel implementadas dentro del propio manejador y proporcionan conceptos que describen los detalles de cómo se almacenan los datos en el ordenador. Este modelo se creó a partir del modelo lógico, el cual no es más que el modelo dimensional elaborado en el capítulo anterior. A través de este es que se inicia la implementación del almacén.

3.1.1 Estructuras de Datos

Las estructuras de datos son una colección de datos con el objetivo de facilitar su manipulación. Los datos almacenados están ordenados por tablas donde cada una de estas pertenece a un esquema, esto posibilita un mejor entendimiento y una organización adecuada de los mismos. A continuación se aborda sobre estos aspectos incluyendo además las restricciones, secuencias e índices.

Esquemas y Tablas

Se le denomina esquema a la estructura que maneja PostgreSQL para organizar sus tablas. El esquema define las tablas, los campos en cada tabla y las relaciones entre los campos y las tablas; estos no se modifican a menudo aunque los datos que se almacenan en la base de datos pueden cambiar frecuentemente ya que se actualizan, modifican e insertan. A partir de la especificación que se le hizo a las tablas de hechos y dimensiones en el diseño de la solución se definieron los esquemas en la implementación de la propuesta. A continuación se explican cada uno de ellos:

- Esquema Dimensiones: contiene agrupadas las 17 tablas dimensiones que almacenan información del tema de análisis de Educación.
- Esquema Hechos: agrupa las 11 tablas de hechos definidas en el capítulo 2. (Ver expediente del proyecto)

En las tablas es donde se guardan los datos recogidos por un programa. Su estructura es similar a la vista general de una hoja de cálculo. Para almacenar los datos de las tablas del sistema se utilizaron los ficheros Excel o series que guardan la información estadística de la educación en la ONE. (Ver Anexo 3)

Restricciones y Secuencias

Las restricciones son condiciones que obligan el cumplimiento de ciertas reglas de la base de datos. Proveen un método de implementar reglas en la base de datos y restringen los datos que pueden ser almacenados en las tablas. Algunas son definidas por el sistema gestor de base de datos y otras por los usuarios.

Las secuencias son basadas en las llaves primarias y foráneas, con atributos que se van a ir incrementando secuencialmente durante la entrada de datos a la base de datos, ejemplo de esto es, a las llaves primarias se le definió que el tipo de dato sería SERIAL y estas se van a ir incrementando de 1 en 1 a medida que le vayan introduciendo los datos.

Durante la construcción del almacén de datos se definió la restricción: la secuencia de las llaves debe ser incrementada según se definan siempre empezando por un valor mínimo hasta un máximo.

Índices

Las políticas de indexado son restricciones que el programador de base de datos define a la hora de indexar o crear un índice. Un índice es una estructura de disco asociada a una tabla o una vista que acelera la recuperación de filas de la tabla o de la vista. Contiene claves generadas a partir de una o varias columnas de la tabla o la vista. Dichas claves están almacenadas en una estructura (árbol b) que permite que SQL Server busque de forma rápida y eficiente la fila o filas asociadas los valores de cada clave.

Los índices fueron utilizados para optimizar las consultas y la política de indexado fue basada en las llaves primarias. En el Anexo 2 se muestra la tabla con los índices generados por el gestor, este la genera a la hora de crear una llave primaria, una tabla o cuando se identifica una columna según la llave primaria coincidiendo con las propiedades o características que poseen los índices agrupados.

A partir de las estructuras de datos del almacén se generó el Lenguaje de Definición de Datos (DDL). Este es un artefacto constituido por el script de creación de objetos de la base de datos y se construyó a partir de la transformación de los esquemas de datos. Este artefacto es un script que forma parte del expediente del proyecto.

3.2 Carga de nomencladores

El proceso de Extracción, Transformación y Carga de los datos (ETL) se basa en controlar la fuente, la transformación correspondiente y el destino de los datos en todo el proceso. Por cuestiones del alcance de la investigación solo se pudo realizar la carga de los nomencladores o clasificadores de las tablas dimensiones.

Los nomencladores son los valores de las dimensiones, se efectuó la carga de los mismos a partir de un conjunto de series facilitadas por la especialista de la ONE en la UCI. En el epígrafe 2.2.2 se muestran los clasificadores que fueron cargados en correspondencia con cada dimensión.

A partir de la realización del proceso de carga de los nomencladores se generó el artefacto Lenguaje de Modelación de los Datos (DML). Este es un script que debe ser capaz de generar los clasificadores existentes en la solución al ser cargados en la base de datos y forma parte del código fuente del expediente del proyecto.

3.3 Usuarios y privilegios

La definición de los usuarios del equipo que va a trabajar con el mercado de datos es muy importante pues mediante estos se garantiza la seguridad del sistema a partir de que cada persona tiene definido el trabajo que va a realizar y solo se le concederán los privilegios necesarios para esto. Postgres brinda la posibilidad de agrupar a los usuarios según las necesidades de permisos y accesos que necesitará cada rol.

3.3.1 Usuarios y Roles

Para facilitar la organización del trabajo de los desarrolladores y garantizar la seguridad del sistema se definieron los usuarios y roles que trabajarán con el mercado de datos. A continuación se muestra cada uno de ellos:

- **Administrador:** tiene acceso total al mercado de datos, apoya el rendimiento de la aplicación realizando la administración y configuración de la misma.
- **Analista:** su rol se basa en consultar la información del almacén.
- **Especialista de ETL:** su función es realizar el proceso de extracción, transformación y carga de los datos del almacén.

3.3.2 Privilegios

Los privilegios asignados a los usuarios del sistema se definen según en el rol que desempeñan. A continuación se muestra cada uno de estos:

- **Administrador:** posee los derechos de propietario, selección, actualización, inserción, eliminación, refrescar (Owner, Select, Update, Insert, Delete, Refresh) y Trigger¹⁴ sobre la estructura de la base de datos.
- **Analista:** solo tiene permisos para seleccionar (Select) los datos del almacén.

¹⁴ Procedimientos que se ejecutan cuando se cumple una condición establecida dentro de la base de datos al realizar operaciones de inserción (INSERT), actualización (UPDATE) o borrado (DELETE).

- **Programador ETL:** se le asignan los privilegios de seleccionar, actualizar, insertar, eliminar, refrescar (Select, Update, Insert, Delete, Refresh) y Trigger de los datos almacenados en el almacén.

El artefacto que se genera a partir de la definición de los privilegios es el Lenguaje de Control de Datos (DCL) que se encarga de reflejar todo el proceso de acceso a la base de datos; es utilizado en el control de la información y trata de que los usuarios accedan a los datos en dependencia de los privilegios que se le otorgaron. Este artefacto es un script que forma parte del expediente del proyecto.

En este punto el almacén cuenta con todos los datos almacenados en las tablas de dimensiones de manera consistente y con los privilegios asignados a cada uno de los usuarios. Conociendo los detalles del modelo de datos físico se procederá a la explicación de la implantación del sistema.

3.4 Guía de Implantación

La guía de implantación representa los pasos que hay que seguir para instalar el mercado de datos dentro de la organización y contiene los requerimientos necesarios para la instalación del almacén. Es necesario que se conozcan los requisitos antes de abordar los pasos para lograr la implantación.

3.4.1 Requerimientos

Los requerimientos necesarios para la realización de la aplicación son los requisitos no funcionales mencionados en el epígrafe 2.1.7, estos están divididos por distintas clasificaciones como son: hardware, software, fiabilidad, entre otros.

3.4.2 Secuencia de Pasos

A continuación se muestran los pasos para la implantación del almacén:

- 1º. Debe estar instalado el gestor de base de datos PostgreSQL 8.4.
- 2º. Debe estar instalada una herramienta de administración de base de datos, la cual ya fue definida: PgAdmin III versión 1.10.
- 3º. Se debe crear una base de datos nueva utilizando la herramienta de administración de bases de datos deseada.

- 4°. Luego de tener la base de datos lista se crean los usuarios: administrador, analista y especialista de ETL para obtener los permisos definidos con anterioridad.
- 5°. Luego se carga el script DDL Educación_BD, para crear la estructura física del almacén que se encuentra en el expediente del proyecto.
- 6°. Teniendo la estructura de la BD están creadas las condiciones para correr el script lenguaje de manipulación de datos DML Educación_BD, al ejecutar este quedan cargados los nomencladores en el almacén.
- 7°. Al estar la BD creada estructuralmente con los nomencladores cargados se debe correr el script lenguaje de control de datos DCL Educación_BD, para los usuarios y permisos definidos en el almacén.

3.5 Validación y pruebas

Los procesos de validación y pruebas son actividades que garantizan que el software funcione de manera eficiente y que esté presente la calidad requerida en el mismo. Para validar el funcionamiento del almacén se realizaron las pruebas de implantación.

3.4.1 Listas de Chequeo Análisis

Las listas de chequeo se realizaron con el objetivo de evaluar la calidad de los artefactos generados en la fase de análisis. Mediante estas listas se pueden cubrir todos los aspectos aplicables en el tema de análisis de la educación en la ONE. A continuación se muestran los nombres de las listas de chequeo aplicadas en el análisis que se encuentran en el expediente del proyecto.

- Lista de Chequeo Especificación de Requisitos
- Lista de Chequeo Evaluación de áreas de la organización
- Lista de Chequeo de la Herramienta para la recolección y análisis de la información

Luego de ser aplicadas las listas de chequeo a los artefactos que se generaron en la fase de análisis, se concluyó que los resultados fueron satisfactorios y se observa la solidez del análisis realizado.

3.4.2 Lista de Chequeo Diseño

Con vistas de lograr un buen diseño de la solución se aplicó la lista de chequeo del diseño. Es específica para cada sistema y pretende cubrir todos los temas aplicables en el diseño e implementación del tema de educación. Su nombre es Lista de Chequeo Modelo de Datos y forma parte del expediente del proyecto.

En esta lista de chequeo se encuentran una serie de estándares definidos con el objetivo de lograr un diseño estandarizado, definiendo pautas claves para la conformación de una estructura entendible.

3.4.3 Validación de requisitos por el cliente

La validación de los requisitos fue aprobada por la representante de la ONE en la UCI Elena Leonila Fernández García y por el jefe de la línea de almacenes de datos del Centro de Tecnologías de Gestión de Datos (DATEC) Ing. Asnioby Hernández López. Ellos realizaron una revisión detallada del diseño de la solución y de los requisitos, obteniéndose finalmente resultados satisfactorios en la propuesta.

3.4.4 Pruebas de Implantación

Se realizaron las pruebas de implantación para verificar la calidad del proceso de implementación y facilitar la identificación de los resultados. Seguidamente se muestra una tabla con las condiciones que se tienen que cumplir antes y después para la implantación del sistema.

Pre – Condiciones	Resultados Esperados	Pos-Condiciones
Antes de la implantación del sistema, se debe tener instalado el PostgreSQL y para su administración PgAdmin III.	Al tener las pre-condiciones específicas se obtienen las condiciones necesarias para la implantación del sistema en su totalidad.	Se crea la Base de Datos donde se montará el almacén.
Se debe tener creada la BD, para correr el script DDL.	Tener la estructura de la BD.	Se tienen las condiciones necesarias para proceder a definirle los usuarios y permisos.
Se debe correr el script DDL para la carga del DCL.	Tener los usuarios y permisos cargados en el almacén.	Se tiene el almacén creado listo para efectuar la carga de

		los nomencladores.
Se debe tener la BD creada así como la estructura de la misma para la carga del DML.	Tener los nomencladores cargados en las diferentes estructuras de BD.	Se tiene el almacén listo para su uso.

Conclusiones

En este capítulo se estructuró el Mercado de Datos Indicadores de la Educación en esquemas, tablas, índices, restricciones y secuencias que permitieron realizar el proceso de carga de nomencladores. El administrador, el analista y el especialista de ETL son los usuarios que tienen acceso sobre el mercado de datos y a los cuales se les asignaron determinados privilegios para garantizar la seguridad del sistema. La propuesta fue validada de satisfactoria por el cliente. También fueron evaluados los artefactos de las fases de análisis y diseño a través de las listas de chequeo y se aplicaron las pruebas de implantación para verificar que el mercado de datos tuviera la calidad requerida; en ambos casos se obtuvieron buenos resultados.

Conclusiones Generales

- Se decidió crear un mercado de datos para controlar la información del departamento de educación en la Oficina Nacional de Estadísticas haciendo uso de la metodología de Kimball ajustada a las políticas de la universidad.
- Se definieron herramientas para el desarrollo del mercado de datos como el Visual Paradigm para realizar el modelo de diseño, PostgreSQL 8.4 como gestor de base de datos y PgAdmin III 1.10 para la administración de la base de datos.
- A partir del análisis efectuado se sentaron las bases para realizar el modelo de diseño donde se crearon los hechos y las dimensiones que permitieron lograr de manera satisfactoria un diseño de la propuesta de solución validado por el cliente.
- El mercado de datos quedó correctamente estructurado y se garantizó su seguridad a través de los privilegios que fueron otorgados a los roles que interactúan con él.
- Las pruebas de implantación demostraron que se realizó una adecuada carga de nomencladores.

Recomendaciones

- Realizar el proceso de Extracción, Transformación y Carga (ETL).
- Hacer la capa de visualización a partir de la solución propuesta.
- Desplegar la aplicación en la Oficina Nacional de Estadísticas.

Referencias bibliográficas

1. **ONE.** Oficina Nacional de Estadísticas. *ONE*. [En línea] ONE, 2010. [Citado el: 19 de febrero de 2010.] Disponible en: www.one.cu.
2. **unesco.** Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. *unesco*. [En línea] unesco, 2010. [Citado el: 16 de febrero de 2010.] Disponible en: <http://www.unesco.org/new/es/unesco/>.
3. **Américas.** Américas Un Nouveau Monde. *Américas*. [En línea] 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://www.americas-fr.com/es/civilizaciones/calendario.html>.
4. **espiritastudio.** Estudios Espiritas. *espiritastudio*. [En línea] espiritastudio, 21 de marzo de 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://espiritastudio.blogspot.com/2009/03/el-ojo-de-horus-7parte-el-amanecer.html>.
5. **Valdés, Damián Pérez.** maestros del web. *maestros del web*. [En línea] maestros del web, 26 de octubre de 2007. [Citado el: 23 de febrero de 2010.] Disponible en: <http://www.maestrosdelweb.com/principiantes/%C2%BFque-son-las-bases-de-datos/>.
6. **Bolton, David.** about. [En línea] 2010. [Citado el: 1 de febrero de 2010.] Disponible en: <http://cplus.about.com/od/glossar1/g/databasedefn.htm>.
7. **sinnexus.** sinnexus. [En línea] enero de 2007. [Citado el: 5 de febrero de 2010.] Disponible en: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
8. **Méndez, Néstor Darío Duque.** Universidad Nacional de Colombia. Sede Manizales. *Universidad Nacional de Colombia. Sede Manizales*. [En línea] Universidad Nacional de Colombia. Sede Manizales, 2005. [Citado el: 09 de abril de 2010.] Disponible en: <http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060029/index.html>.
9. **Sinnexus.** sinnexus. [En línea] Ministerio de Hacienda Español, 2007. [Citado el: 15 de febrero de 2010.] Disponible en: http://www.sinnexus.com/business_intelligence/datamart.aspx.
10. **Mario Roberto Reyes Marroquín, Pablo Augusto Rosales Tejada.** *DESARROLLO DE UN DATAMART DE INFORMACIÓN ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS*. Guatemala : s.n., 2007.

11. **Microsoft Corporation.** msdn.microsoft.com. *MSDN*. [En línea] Microsoft Corporation, julio de 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://msdn.microsoft.com/es-es/library/ms175523.aspx>.
12. **Transforming Knowledge.** KLE. *Transforming Knowledge Into Action!* [En línea] Transforming Knowledge, 16 de julio de 2009. [Citado el: 21 de febrero de 2010.] Disponible en: <http://www.siskle.com/>.
13. **Sierra, Julio Ernesto Ortiz.** *Diseño e Implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. La Habana : s.n., 2009.
14. **Carrillo Ayala Alma Nelly, Cervantes Estrada Freddy, Gutiérrez Cervantes Miguel.** Universidad Autónoma de Morelos. *uaem*. [En línea] uaem, 2010. [Citado el: 16 de febrero de 2010.] Disponible en: www.uaem.mx/posgrado/mcruz/cursos/miic/sql.pdf.
15. **Creative Commons.** Dataprix Knowledge ia a Goal. *Dataprix*. [En línea] Creative Commons, 14 de Septiembre de 2007. [Citado el: 25 de febrero de 2010.] Disponible en: <http://www.dataprix.com/>.
16. **Jaime Casanova.** PostgreSQL. [En línea] 1 de julio de 2009. [Citado el: 23 de febrero de 2010.] Disponible en: <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>.
17. **WordPress.com.** Saforas'\$ the::beastieux. *Saforas'\$ the::beastieux*. [En línea] WordPress.com, 31 de enero de 2008. [Citado el: 12 de febrero de 2010.] Disponible en: <http://saforas.wordpress.com/2008/01/31/licencias-bsd/>.
18. **pgAdmin PostgreSQL Tools.** pgAdmin PostgreSQL Tools. *pgAdmin* . [En línea] 23 de abril de 2009. [Citado el: 24 de febrero de 2010.] Disponible en: <http://www.pgadmin.org/>.
19. **Ruth Hidalgo, Victor Samaniego.** Modelo Conceptual de la Base de Datos. *slideshare*. [En línea] SlideShare, 2009. [Citado el: 24 de febrero de 2010.] Disponible en: <http://www.slideshare.net/ruthamada/modelo-conceptual-de-la-base-de-datos-360327>.
20. **Fránquiz, Ing. Milagros Montalvo.** AHCIET. *Revista de Telecomunicaciones*. [En línea] Instituto de Investigación y Desarrollo de Telecomunicaciones, Cuba., enero/marzo de 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://www.ahciet.net/actualidad/revista/r.aspx?ids=10796&ids2=21861>.
21. **García, Lic. Rosa María Mato.** Diseño de Base de Datos. *Diseño de Base de Datos*. 1999.

22. **ETL-Tools.Info.** ETL-Tools.Info. *Esquema copo de nieve*. [En línea] ETL-Tools.Info, 2006 - 2010. [Citado el: 17 de febrero de 2010.] Disponible en: http://etl-tools.info/es/bi/almacenedatos_esquemacopo-de-nieve.htm.
23. **Hazbleydi C. Verástegui.** Modelado Dimensional de Datos. *DB System Ltda.* [En línea] 3 de abril de 2007. [Citado el: 17 de febrero de 2010.] Disponible en: http://api.ning.com/files/ei6*2AatbAfAFDwyLdYnMfJ7ZXvd3xwXakwY*gIlpDQk4Ykcb3K9af3G*OvF06uLSTXQwlltUimXRHhSRU8enhjUX2QDs-kR/MODELADODIMENSIONALDEDATOS_V2.pdf.
24. **ETL-Tools.Info.** ETL-Tools.Info. *Esquema de la estrella*. [En línea] ETL-Tools.Info, 2006 - 2010. [Citado el: 17 de febrero de 2010.] Disponible en: http://etl-tools.info/es/bi/almacenedatos_esquemarestrella.htm.
25. **CEDS e-learning.** Herramientas Case . *CEDS e-learning*. [En línea] 2010. [Citado el: 21 de febrero de 2010.] Disponible en: <http://ceds.nauta.es/informes/case04.htm>.
26. **Universidad Autónoma de Morelos.** Universidad Autónoma de Morelos. [En línea] enero de 2010. [Citado el: 18 de febrero de 2010.] Disponible en: <http://www.cert.org.mx/boletin/?vulne=5804>.
27. **Microsoft Corporation.** Microsoft Developer Network. *msdn*. [En línea] Microsoft Corporation, 2010. [Citado el: 8 de febrero de 2010.] Disponible en: <http://msdn.microsoft.com/es-es/library/ms174915.aspx>.
28. **Business Dictionary.com.** Business Dictionary. *BusinessDictionary.com*. [En línea] Business Dictionary.com. [Citado el: 23 de febrero de 2010.] Disponible en: <http://www.businessdictionary.com/definition/business-management.html>.
29. **Menéndez, Dr. Eugenio Santos.** Diseño y Optimización de Bases de Datos:DataWareHouse. *Diseño y Optimización de Bases de Datos:DataWareHouse*. [En línea] Universidad Politécnica de Madrid, 2009. [Citado el: 21 de febrero de 2010.] Disponible en: www.upm.es.
30. **Ing.Lilliam Vega Torres, Ing.Luis Rojas Díaz, Lic.Cecilia Placeres Villar.** *LA INTELIGENCIA DE NEGOCIO*. La Habana : Empresa de Tecnologías de la Información y Servicios Telemáticos Avanzados, 2008. Citado: 11 de febrero del 2010.

Bibliografía

1. **ONE.** Oficina Nacional de Estadísticas. *ONE*. [En línea] ONE, 2010. [Citado el: 19 de febrero de 2010.] Disponible en: www.one.cu.
2. **unesco.** Organizacion de las Naciones Unidas para la Educación, la Ciencia y la Cultura. *unesco*. [En línea] unesco, 2010. [Citado el: 16 de febrero de 2010.] Disponible en: <http://www.unesco.org/new/es/unesco/>.
3. **Américas.** Américas Un Nouveau Monde. *Américas*. [En línea] 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://www.americas-fr.com/es/civilizaciones/calendario.html>.
4. **espiritasestudio.** Estudios Espiritas. *espiritasestudio*. [En línea] espiritasestudio, 21 de marzo de 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://espiritasestudio.blogspot.com/2009/03/el-ojo-de-horus-7parte-el-amanecer.html>.
5. **sinnexus.** sinnexus. [En línea] enero de 2007. [Citado el: 5 de febrero de 2010.] Disponible en: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
6. **Sinnexus.** sinnexus. [En línea] Ministerio de Hacienda Español, 2007. [Citado el: 15 de febrero de 2010.] Disponible en: http://www.sinnexus.com/business_intelligence/datamart.aspx.
7. **Mario Roberto Reyes Marroquín, Pablo Augusto Rosales Tejada.** *DESARROLLO DE UN DATAMART DE INFORMACIÓN ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS*. Guatemala : s.n., 2007.
8. **Microsoft Corporation.** msdn.microsoft.com. *MSDN*. [En línea] Microsoft Corporation, julio de 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://msdn.microsoft.com/es-es/library/ms175523.aspx>.
9. **Transforming Knowledge.** KLE. *Transforming Knowledge Into Action!* [En línea] Transforming Knowledge, 16 de julio de 2009. [Citado el: 21 de febrero de 2010.] Disponible en: <http://www.siskle.com/>.
10. **Sierra, Julio Ernesto Ortiz.** *Diseño e Implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. La Habana : s.n., 2009.
11. **Creative Commons.** Dataprix Knowledge ia a Goal. *Dataprix*. [En línea] Creative Commons, 14 de Septiembre de 2007. [Citado el: 25 de febrero de 2010.] Disponible en: <http://www.dataprix.com/>.
12. **pgAdmin PostgreSQL Tools.** pgAdmin PostgreSQL Tools. *pgAdmin* . [En línea] 23 de abril de 2009. [Citado el: 24 de febrero de 2010.] Disponible en: <http://www.pgadmin.org/>.

13. **Fránquiz, Ing. Milagros Montalvo.** AHCIET. *Revista de Telecomunicaciones*. [En línea] Instituto de Investigación y Desarrollo de Telecomunicaciones, Cuba., enero/marzo de 2009. [Citado el: 10 de febrero de 2010.] Disponible en: <http://www.ahciet.net/actualidad/revista/r.aspx?ids=10796&ids2=21861>.
14. 22. **ETL-Tools.Info.** ETL-Tools.Info. *Esquema copo de nieve*. [En línea] ETL-Tools.Info, 2006 - 2010. [Citado el: 17 de febrero de 2010.] Disponible en: http://etl-tools.info/es/bi/almacenedatos_esquema-copo-de-nieve.htm.
15. 23. **Hazbleydi C. Verástegui.** Modelado Dimensional de Datos. *DB System Ltda.* [En línea] 3 de abril de 2007. [Citado el: 17 de febrero de 2010.] Disponible en: http://api.ning.com/files/ei6*2AatbAfAFDwyLdYnMfJ7ZXvd3xwXakwY*gllpDQk4Ykcb3K9af3G*OvF06uLSTXQwlltUimXRHhSRU8enhjUX2QDs-kR/MODELADODIMENSIONALDEDATOS_V2.pdf.
16. 24. **ETL-Tools.Info.** ETL-Tools.Info. *Esquema de la estrella*. [En línea] ETL-Tools.Info, 2006 - 2010. [Citado el: 17 de febrero de 2010.] Disponible en: http://etl-tools.info/es/bi/almacenedatos_esquema-estrella.htm.
17. 26. **Universidad Autónoma de Morelos.** Universidad Autónoma de Morelos. [En línea] enero de 2010. [Citado el: 18 de febrero de 2010.] Disponible en: <http://www.cert.org.mx/boletin/?vulne=5804>.
18. **Business Dictionary.com.** Business Dictionary. *BusinessDictionary.com*. [En línea] Business Dictionary.com. [Citado el: 23 de febrero de 2010.] Disponible en: <http://www.businessdictionary.com/definition/business-management.html>.
19. **Menéndez, Dr. Eugenio Santos.** Diseño y Optimización de Bases de Datos:DataWareHouse. *Diseño y Optimización de Bases de Datos:DataWareHouse*. [En línea] Universidad Politécnica de Madrid, 2009. [Citado el: 21 de febrero de 2010.] Disponible en: www.upm.es.
20. **Ing.Lilliam Vega Torres, Ing.Luis Rojas Díaz, Lic.Cecilia Placeres Villar.** *LA INTELIGENCIA DE NEGOCIO*. La Habana : Empresa de Tecnologías de la Información y Servicios Telemáticos Avanzados, 2008. Citado: 11 de febrero del 2010.

Anexos

Anexo1 Tablas de especificación de casos de uso del sistema

Tabla 3 Especificación del Caso de Uso de Información Analizar Información de los Círculos Infantiles

Caso de Uso:	Analizar Información de los Círculos Infantiles	
Tipo:	Información.	
Actores:	Usuario	
Resumen:		
Precondiciones:	Compleitud del almacén. Carga de los Datos.	
Referencias	R11, R12	
Prioridad		
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Negocio	
Entra al sistema.	Muestra opciones de reportes de los círculos infantiles.	
Selecciona reportes de círculos infantiles.	Muestra tipos de reportes de los círculos infantiles.	
Opciones de reportes de los círculos infantiles.		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
<i>Dimensiones disponibles relacionadas con el hecho círculos infantiles.</i> <ul style="list-style-type: none"> ✓ <i>Tiempo por años</i> ✓ <i>Años CI</i> 	<ul style="list-style-type: none"> ✓ <i>Cantidad de niños de 0 – 5 años matriculados por 1000 en esas edades.</i> ✓ <i>Cantidad de madres</i> 	<i>Anual</i>

	<p><i>beneficiadas.</i></p> <ul style="list-style-type: none"> ✓ <i>Cantidad de Personal Técnico y Profesional.</i> ✓ <i>Capacidad al final del año.</i> ✓ <i>Asistencia Promedio.</i> ✓ <i>Matrícula Final.</i> ✓ <i>Cantidad de Círculos Infantiles.</i> 	
<i>Prototipo de Interfaz</i>		
Poscondiciones	<p><i>Mostrar cantidad de niños de 0 – 5 años matriculados por 1000 en esas edades.</i></p> <p><i>Mostrar cantidad de madres beneficiadas.</i></p> <p><i>Mostrar cantidad de personal técnico y profesional.</i></p> <p><i>Mostrar capacidad al final del año.</i></p> <p><i>Mostrar asistencia promedio.</i></p> <p><i>Mostrar matrícula final.</i></p> <p><i>Mostrar cantidad de círculos infantiles.</i></p>	

Tabla 4 Especificación del Caso de Uso de Información Analizar Información de las Enseñanzas Primaria y Media

Caso de Uso:	Analizar Información de las Enseñanzas Primaria y Media
Tipo:	Información.
Actores:	Usuario
Resumen:	
Precondiciones:	Completitud del almacén. Carga de los Datos.
Referencias	RI5, RI6, RI7, RI8, RI9, RI10, RI11, RI15, RI13, RI14, RI15, RI16, RI17, RI18, RI19, RI20.

Prioridad	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Negocio
Entra al sistema.	Muestra opciones de reportes de las enseñanzas primaria y media.
Selecciona reportes de enseñanzas primaria y media.	Muestra tipos de reportes de las enseñanzas primaria y media.
Opciones de reportes de las enseñanzas primaria y media	
Perspectivas de análisis	Posibles resultados
	Medidas
	Periodicidad
<i>Dimensiones disponibles relacionadas con el hecho las enseñanzas primaria y media.</i>	✓ Cantidad de Anual
✓ Educaciones	✓ Matrícula Inicial.
✓ Estudiantes Educ Tec Prof	✓ Cantidad de
✓ Geografía	Personal.
✓ Sexo	✓ Cantidad de
✓ Especialidades técnica y profesional	Estudiantes.
✓ Estudiantes	
✓ Tiempo Curso Escolar	
✓ Personal	
Prototipo de Interfaz	
Poscondiciones	Mostrar cantidad de graduados. Mostrar matrícula inicial. Mostrar cantidad de personal. Mostrar cantidad de estudiantes

Tabla 5 Especificación del Caso de Uso de Extraer Excel de Indicadores generales de los círculos infantiles

Caso de Uso:	Extraer Excel de Indicadores generales de los círculos infantiles
---------------------	---

Tipo:	Funcional.
Actores:	Programador ETL.
Resumen:	
Precondiciones:	Disponibilidad de las fuentes.
Referencias	RF2
Prioridad	
Complejidad	
Flujo Normal de Eventos	
Sección ""	
Acción del Actor	Respuesta del Excel Indicadores generales de los círculos infantiles
1. Realiza la conexión al Excel de los indicadores generales de los círculos infantiles.	1.1 Responde a la solicitud de conexión.
2. Selecciona estructuras o archivos a extraer.	
3. Realiza la extracción.	
<i>Prototipo de Interfaz</i>	
Flujos Alternos	
Acción del Actor	Respuesta del Negocio
	1.1 No responde a solicitud de conexión.
2. Notifica el error a administrador Excel de los indicadores generales de los círculos infantiles	
Flujo Alternativo	
2. Si hay control de cambios, verifica si hay modificaciones. En caso de que no haya, va al paso 4 del flujo normal. En caso de que haya, va al paso 2 del flujo normal.	
<i>Prototipo de Interfaz</i>	
Poscondiciones	Extracción completada en área temporal. Estructuras del almacén disponibles.

Tabla 6 Especificación del Caso de Uso funcional Transformar y cargar datos del Excel de indicadores generales de los círculos infantiles

Caso de Uso:	Transformar y cargar datos del Excel de indicadores generales de los círculos infantiles	
Tipo:	Funcional.	
Actores:	Programador ETL.	
Resumen:		
Precondiciones:	Extracción completada en área temporal. Estructuras del almacén disponibles.	
Referencias	RF2	
Prioridad		
Flujo Normal de Eventos		
Sección “”		
Acción del Actor.	Respuesta del almacén.	
1. Selecciona estructuras del área temporal a transformar.		
2. Carga datos seleccionados en memoria.		
3. Aplica transformaciones pertinentes y genera datos de auditoría.		
4. Carga datos en el almacén.		
5. Actualiza las tablas de control de cambios. (Opcional)		
6. Inicia el proceso de actualización de las vistas materializadas.(Opcional)		
<i>Prototipo de Interfaz</i>		
Flujos Alternos		
Acción del Actor	Respuesta del Negocio	
<i>Prototipo de Interfaz</i>		
Poscondiciones	Datos del Excel de indicadores generales de los círculos infantiles transformados y cargados.	

Anexo 2 Tablas de índices

Índice	Tabla	Esquema	Tipo	Campo	Único	PK
PK19	circulos_infantiles	hechos	btree	id_tiempo_annos, id_annos	✓	✓
PK22	control_escuelas	hechos	btree	id_tiempo_curso_escolar id_educaciones id_provincias id_zona_ubicacion	✓	✓
PK8	dim_annos_ci	dimensiones	btree	id_annos	✓	✓
PK12	dim_centros_universitarios	dimensiones	btree	id_centros	✓	✓
PK13	dim_cines	dimensiones	btree	id_cine	✓	✓
PK4	dim_edad_establecida	dimensiones	btree	id_edad	✓	✓
PK15	dim_educaciones	dimensiones	btree	id_educaciones	✓	✓
PK7	dim_especialidades_educacion_tecnica_profesional	dimensiones	btree	id_especialidades	✓	✓
PK11	dim_especialidades_universitarias	dimensiones	btree	id_especialidades_univ	✓	✓
PK10	dim_estudiantes	dimensiones	btree	id_estudiantes	✓	✓
PK17	dim_estudiantes_educacion_tecnica_profesional	dimensiones	btree	id_est	✓	✓
PK14	dim_geografia	dimensiones	btree	id_provincias	✓	✓
PK3	dim_organismos	dimensiones	btree	id_organismos	✓	✓
PK29	dim_personal	dimensiones	btree	id_personal	✓	✓
PK9	dim_posgrados	dimensiones	btree	id_posgrados	✓	✓
PK1	dim_tiempo_annos	dimensiones	btree	id_tiempo_annos	✓	✓
PK2	dim_tiempo_curso_escolar	dimensiones	btree	id_tiempo_curso_escolar	✓	✓
PK5	dim_tipo_estudio	dimensiones	btree	id_tipo_est	✓	✓
PK32	dim_zona_ubicacion	dimensiones	btree	id_zona_ubicacion	✓	✓
PK20	educacion_posgrado	hechos	btree	id_posgrados id_tiempo_annos	✓	✓
PK23	ensenanza_superior	hechos	btree	id_organismos, id_centros, id_tipo_est id_especialidades_univ id_tiempo_curso_escolar	✓	✓
PK30	ensenanza_tecnica_profesional	hechos	btree	id_est, id_especialidades id_tiempo_curso_escolar	✓	✓

PK18	enseñanzas_primaria_media	hechos	btree	id_tiempo_curso_escolar id_educaciones id_provincias id_zona_ubicacion	✓	✓
PK27	estudiantes_primaria_media	hechos	btree	id_educaciones id_tiempo_curso_escolar id_provincias id_estudiantes id_zona_ubicacion	✓	✓
PK21	indicadores_anuales	hechos	btree	id_tiempo_anos	✓	✓
PK26	nuevo_ingreso_matricula_educ_superior	hechos	btree	id_especialidades_univ, id_tiempo_curso_escolar	✓	✓
PK28	personal_primaria_media	hechos	btree	id_tiempo_curso_escolar id_educaciones id_provincias, id_personal id_zona_ubicacion	✓	✓
PK24	tasas_indices	hechos	btree	id_tiempo_curso_escolar , id_cine, id_edad	✓	✓

Anexo 3 Esquemas y tablas

Tabla 7 Hecho círculos infantiles

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_tiempo_anos	integer	✓	
✓	✓	id_anos	integer	✓	
		matricula_final_circuitos_infantiles	integer		
		asistencia_promedio_circuitos_infantiles	integer		
		capacidad_final_año_circuitos_infantiles	integer		
		cantidad_madres_beneficiadas	integer		
		cantidad_personal_tecnico_educacional	integer		
		cantidad_ninos_0_a_5_años_matriculados _10000_edades	integer		

Tabla 8 Hecho control de escuelas

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_tiempo_curso_escolar	integer	✓	
✓	✓	id_annos	integer	✓	
✓	✓	id_educaciones	integer	✓	
✓	✓	id_provincias	integer	✓	
✓	✓	id_zona_ubicacion	integer	✓	
		cantidad_escuelas	integer		

Tabla 9 Dimensión años_ci

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_annos	serial	✓	✓
		tipo_annos_ci	varchar(10)		

Tabla 10 Dimensión centros_universitarios

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_centros	serial	✓	✓
		nombre_centro	varchar(60)		

Tabla 11 Dimensión cines

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_cine	serial	✓	✓
		tipo_cine	varchar(40)		
		descripción	varchar(30)		

Tabla 12 Dimensión edad_establecida

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_edad	serial	✓	✓
		rango_edad	varchar(10)		

Tabla 13 Dimensión educaciones

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_educaciones	serial	✓	✓
		tipo_educaciones	varchar(90)		
		descripción	varchar(30)		

Tabla 14 Dimensión especialidades_educación_técnica_profesional

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_especialidades	serial	✓	✓
		especialidades	varchar(80)		

Tabla 15 Dimensión especialidades_universitarias

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_especialidades_univ	serial	✓	✓
		tipo_especialidades	varchar(90)		

Tabla 16 Dimensión estudiantes

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_estudiantes	serial	✓	✓
		tipo_estudiantes	varchar(20)		

		descripción	varchar(30)		
--	--	-------------	-------------	--	--

Tabla 17 Dimensión estudiantes_educación_técnica_profesional

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_est	serial	✓	✓
		tipo_estudiantes	varchar(20)		

Tabla 18 Dimensión geografía

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_provincias	serial	✓	✓
		provincias	varchar(20)		

Tabla 19 Dimensión organismos

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_organismos	serial	✓	✓
		tipo_organismos	varchar(20)		

Tabla 20 Dimensión personal

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_personal	serial	✓	✓
		tipo_personal	varchar(60)		
		descripcion	varchar(30)		

Tabla 21 Dimensión posgrados

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
----------------	---------------	--------	--------------	---------	-------

✓		id_posgrados	serial	✓	✓
		tipo_posgrados	varchar(60)		

Tabla 22 Dimensión tiempo_años

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_tiempo_años	serial	✓	✓
		años	varchar(10)		

Tabla 23 Dimensión tiempo_curso_escolar

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_tiempo_curso_escolar	serial	✓	✓
		curso_escolar	varchar(40)		

Tabla 24 Dimensión tipo_estudio

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_tipo_est	serial	✓	✓
		tipo_estudio	varchar(60)		

Tabla 25 Dimensión zona_ubicación

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓		id_zona_ubicacion	serial	✓	✓
		zona_ubicacion	varchar(60)		

Tabla 26 Hecho educación posgrado

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único

✓	✓	id_posgrados	integer	✓	
✓	✓	id_tiempo_annos	integer	✓	
		cantidad_participantes	integer		
		cantidad_cursos_posgrados	integer		

Tabla 27 Hecho enseñanza superior

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_organismos	integer	✓	
✓	✓	id_centros	integer	✓	
✓	✓	id_tipo_est	integer	✓	
✓	✓	id_especialidades_univ	integer	✓	
✓	✓	id_tiempo_curso_escolar	integer	✓	
		cantidad_mujeres_graduadas	integer		
		cantidad_graduados	integer		
		cantidad_becarios	integer		
		matricula_inicial	varchar(18)		
		cantidad_personal_docente	integer		

Tabla 28 Hecho enseñanza técnica profesional

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_est	integer	✓	
✓	✓	id_especialidades	integer	✓	
✓	✓	id_tiempo_curso_escolar	integer	✓	
		matricula_inicial	integer		
		cantidad_graduados	integer		

Tabla 29 Hecho enseñanzas primaria media

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_tiempo_curso_escolar	integer	✓	
✓	✓	id_educaciones	integer	✓	
✓	✓	id_provincias	integer	✓	
✓	✓	id_zona_ubicacion	integer	✓	
		matricula_inicial	integer		
		matricula_inicial_hembras	integer		

Tabla 30 Hecho estudiantes primaria_media

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_educaciones	integer	✓	
✓	✓	id_tiempo_curso_escolar	integer	✓	
✓	✓	id_provincias	integer	✓	
✓	✓	id_estudiantes	integer	✓	
✓	✓	id_zona_ubicacion	integer	✓	
		cantidad_estudiantes	integer		
		cantidad_graduados	integer		
		cantidad_graduados_hembras	integer		

Tabla 31 Hecho indicadores anuales

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_tiempo_años	integer	✓	✓
		cantidad_circuitos_infantiles	integer		
		cantidad_habitantes	integer		
		gasto_publico_educacion	integer		
		percapita	varchar(10)		

Tabla 32 Hecho nuevo ingreso matrícula educ_superior

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_especialidades_univ	integer	✓	
✓	✓	id_tiempo_curso_escolar	integer	✓	
		matricula_cursos_dirigidos	integer		
		cantidad_nuevo_ingreso	integer		

Tabla 33 Hecho personal primaria media

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_tiempo_curso_escolar	integer	✓	
✓	✓	id_educaciones	integer	✓	
✓	✓	id_provincias	integer	✓	
✓	✓	id_personal	integer	✓	
✓	✓	id_zona_ubicacion	integer	✓	
		cantidad_personal	integer		

Tabla 34 Hecho tasas índices

Llave primaria	Llave foránea	Nombre	Tipo de dato	No Nulo	Único
✓	✓	id_tiempo_curso_escolar	integer	✓	
✓	✓	id_cine	integer	✓	
✓	✓	id_edad	integer	✓	
		matricula	integer		
		poblacion_edad_escolar	integer		
		cantidad_ninnas	varchar(10)		
		cantidad_ninos	integer		