

Tipo de artículo: Artículo original

## Mercado de Datos para el proceso de perfeccionamiento educacional en Cuba

### *Data Mart for the process of improvement in Cuba educational system*

Irela González Piñera <sup>1\*</sup> ,  
Amarelys Alvarez Pérez <sup>2</sup> ,  
Arodys Eugenio Dominguez Vaillant <sup>3</sup> .

<sup>1</sup> Departamento Integración de Soluciones. Centro de Representación y Análisis de Datos. Facultad de Ciencias y Tecnologías Computacionales. Universidad de las Ciencias Informáticas, Carretera a San Antonio Km 2 ½, Torrens, La Lisa, La Habana, Cuba. CP: 19370. [irela@uci.cu](mailto:irela@uci.cu)

<sup>2</sup> Departamento Integración de Soluciones. Centro de Representación y Análisis de Datos. Facultad de Ciencias y Tecnologías Computacionales. Universidad de las Ciencias Informáticas, Carretera a San Antonio Km 2 ½, Torrens, La Lisa, La Habana, Cuba. CP: 19370. [aaperez@uci.cu](mailto:aaperez@uci.cu).

<sup>3</sup> Dirección de Inversiones, Grupo Empresarial de la Industria Sidero-Mecánica. 10800. arodys83@gmail.com

\* Autor para correspondencia: [irela@uci.cu](mailto:irela@uci.cu)

#### Resumen

El software estadístico se ha convertido en el principal soporte para gestionar los grandes volúmenes de información que se generan diariamente en las organizaciones, las cuales hacen uso de la tecnología instalando poderosas herramientas de análisis de datos para obtener con mayor eficiencia información de los datos almacenados. Debido al gran cúmulo de información que genera el proceso de perfeccionamiento del sistema educacional cubano se decide desarrollar un Mercado de Datos que apoye el proceso de toma de decisiones en el Instituto Central de Ciencias Pedagógicas sobre dicho proceso. Para guiar el proceso de desarrollo se utilizó la metodología de Almacenes de Datos del Centro de Representación y Análisis de Datos. Para el diseño e implementación de la solución se utilizaron varias herramientas de software libre. Como herramienta de modelado Visual paradigm 8.0, para la extracción, transformación y carga de los datos el Pentaho Data Integration 7.1 y para la visualización de la información y los dashboard el Pentaho BI Server 6.0. Para comprobar el correcto funcionamiento del sistema se realizaron pruebas haciendo uso de las herramientas casos de prueba y manual de usuario. Las vistas de análisis y los cuadros de mando que conforman la capa de visualización del Mercado de Datos, permiten que los especialistas del Instituto Central de Ciencias Pedagógicas realicen un análisis certero de la información referente a las actividades que son evaluadas en el proceso de perfeccionamiento en el sistema educacional cubano.

**Palabras clave:** Almacenes de datos, Mercado de Datos, toma de decisiones, proceso de perfeccionamiento.

#### Abstract

*Statistical software has become the main support for managing the large volumes of information that are generated daily in organizations, which make use of technology by installing powerful data analysis tools to more efficiently obtain information from stored data. Due to the great accumulation of information generated by the improvement process of the Cuban educational system, it is decided to develop a Data Mart that supports the decision-making process in the Central Institute of Pedagogical Sciences on said process. To guide the development process, was used the Data Warehouse methodology of the Data*



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

*Representation and Analysis Center. For the design and implementation of the solution, were used several free software tools. As modeling tool Visual parading 8.0, Pentaho Data Integration 7.1 for data extraction, transformation and loading, and Pentaho BI Server 6.0 for displaying information and dashboards. To verify the correct operation of the system, tests were carried out using the test cases and user manual tools. The analysis views and the dashboards that make up the Data Mart visualization layer, allow the specialists of the Central Institute of Pedagogical Sciences to carry out an accurate analysis of the information regarding the activities that are evaluated in the process of improvement in the Cuban educational system.*

**Keywords:** Data Warehouse, Data Mart, decision-making, process of improvement.

**Recibido:** 06/12/2020

**Aceptado:** 15/02/2021

## Introducción

El mundo actual se caracteriza por la rapidez con la que se generan los datos y su amplia difusión. Esta dinámica es la que ha llevado a las Instituciones a impulsar reformas haciendo uso de las tecnologías de la información y las comunicaciones (TIC) como apoyo en el proceso de toma de decisiones. Las TIC han propiciado el desarrollo de la industria del software, especialmente el software estadístico, el cual se ha convertido en el principal soporte para gestionar los grandes volúmenes de información que se generan diariamente en las organizaciones. Además, los avances de la informática imponen tecnologías como una evolución necesaria en todos los campos donde la información se procesa, transmite y acumula. En este contexto surgen los almacenes de datos, los cuales tienen como característica la facilidad de uso y la gran capacidad de información. Estos están basados en una tecnología diseñada expresamente para el almacenamiento, consulta y análisis de datos y poseen una estructura de la información orientada al negocio y al usuario final.

En Cuba son muchas las entidades desarrolladoras de software, entre las que vale destacar por su labor en la informatización de la sociedad cubana a la Universidad de las Ciencias Informáticas (UCI). La UCI posee varios centros productivos entre los que se encuentra el Centro de Representación y Análisis de Datos (CREAD), cuyo objetivo se centra en crear bienes y servicios informáticos relacionados con la gestión de datos.

Actualmente el centro CREAD se encuentra inmerso en la informatización del proceso de perfeccionamiento educacional que se realiza en el Instituto Central de Ciencias Pedagógicas (ICCP), debido a que los softwares informáticos se han convertido en un elemento indispensable para que los resultados de experimentación e investigaciones de las instituciones cubanas se realicen de forma certera. La informatización del ICCP se realiza con el fin de apoyar el proceso de toma de decisiones por parte de los especialistas del instituto, contribuyendo a que el cumplimiento de las tareas asignadas se logre en menor tiempo y con mayor calidad.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

El ICCP es una institución científica que ejerce la dirección, desarrollo y evolución de las actividades científicas y teórico-prácticas de los estudios e investigaciones que, en el campo pedagógico y psicopedagógico, se realizan por el Ministerio de Educación (MINED) de la República de Cuba (MINED,2020). Es la única entidad de ciencia e innovación tecnológica del país adscripta al MINED, y, en conjunto con éste, ha llevado a cabo un proceso de perfeccionamiento educacional iniciado en el año 2011.

Para llevar a cabo el proceso de perfeccionamiento educacional, el ICCP ha utilizado herramientas informáticas con el propósito de almacenar los datos referentes a las clases, las actividades metodológicas y actividades educativas llevadas a cabo en cada uno de los centros escolares involucrados en el experimento. Actualmente la información es recogida en ficheros Access por cada una de las provincias del país, siendo recibidos en el ICCP un total de 42 ficheros. Debido al elevado volumen de datos que son manejados, los especialistas del instituto presentan dificultades a la hora insertar, consultar y analizar la información recopilada anualmente. Estos procesos de revisión y análisis de los resultados de las evaluaciones por cada una de las provincias se realizan con lentitud y de manera engorrosa, principalmente para realizar el cruzamiento de información. El proceso de evaluación se realiza a través de planillas de forma manual. Además los datos añadidos a las tablas Access no se encuentran integrados pues presentan errores humanos como la introducción de información errónea, nula y filas duplicadas, lo cual dificulta la realización de comparaciones y evaluaciones. De igual manera se alteran los valores ya insertados, lo que trae consigo la pérdida de información, creando cambios significativos en los reportes estadísticos finales.

## Materiales y métodos

**Analítico-Sintético.** Este método se utilizó para seleccionar las herramientas y tecnologías a utilizar durante el desarrollo del MD. Además, contribuye al estudio de las soluciones existentes asociadas al dominio del problema, generándose nuevos conocimientos respecto a las funcionalidades que el MD debe cumplir.

**Modelación.** Permite representar el problema en cuestión mediante diagramas. Se utilizó para la representación del Modelo lógico de los datos, que facilitaron una mejor comprensión del negocio.

**Inductivo – Deductivo.** La utilización de este método permitió comparar las diferentes metodologías orientadas a AD y realizar una selección eficiente y acertada. Permitió además establecer conclusiones y obtener una idea del funcionamiento del sistema, partiendo de la información consultada para el mismo.

### Métodos empíricos:

**Entrevista no estructurada.** Se realizaron una serie de encuentros con especialistas del ICCP que participan directamente en el proceso de perfeccionamiento educacional, con el objetivo de obtener información actualizada



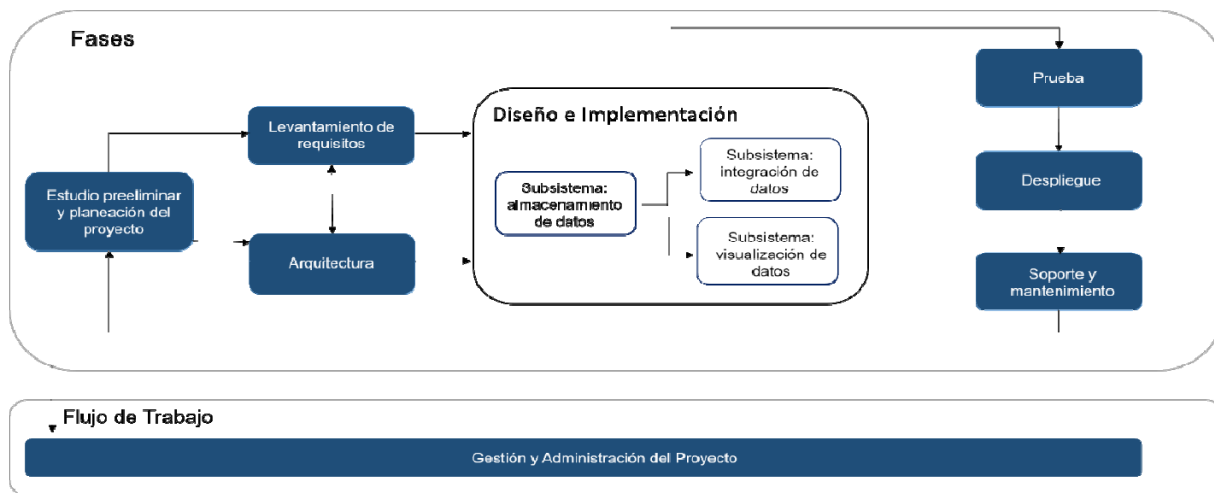
sobre cómo se manejan los datos captados en los ficheros Acces por cada una de las provincias del país y de esta forma definir las reglas del negocio correspondientes así como los requisitos de información y los requisitos no funcionales del MD a desarrollar.

En muchas ocasiones los términos Mercados de datos (MD) y Almacenes de Datos (AD) son utilizados indistintamente. La diferencia es solamente en cuanto al alcance. Mientras que un AD es un sistema centralizado con datos globales de la empresa y de todos sus procesos operacionales, un MD es un subconjunto temático de datos, orientado a un proceso o un área de negocio específica propia de la organización (Ramos, 2011).

Por su parte plantea que un MD "...es un repositorio de información, similar a un AD, pero orientado a un área o departamento específico de la organización...". Este autor también afirma que un AD es la unión de los diferentes MD existentes en una organización (Kimball, 1998). Toda su filosofía se centra en que la construcción de un AD es originada por el interés y esfuerzo de un departamento determinado.

El centro CREAD tiene definida una metodología para el desarrollo de AD que se basa en la metodología de Kimball, reconocida internacionalmente considerando el empleo de MD para un departamento específico dentro de la organización. La metodología definida por el centro CREAD posee un ciclo de vida para el desarrollo de proyectos de AD, claramente definido a partir de la descripción del flujo de trabajo y las fases que presenta y una descripción de los procesos y actividades que se realizan a lo largo de todo el ciclo de vida de la metodología relacionándolos con los roles, artefactos y herramientas necesarios para su ejecución según las características de este tipo de soluciones. Su principal característica es que cubre todas las fases del ciclo de vida por las que pasa la construcción de un AD desde el levantamiento de requerimientos de información inicial hasta la implementación de la herramienta del negocio (Hernández, 2013). Esta metodología se divide en siete fases y un flujo de trabajo como se muestra en la Fig.1. Algunas de sus fases podrán ser implementadas de forma paralela como es el caso de la fase de levantamiento de requisitos y la fase de arquitectura, además durante la fase de diseño e implementación podrán ser desarrollados varios componentes al mismo tiempo, lo que permite un desarrollo mucho más ágil.





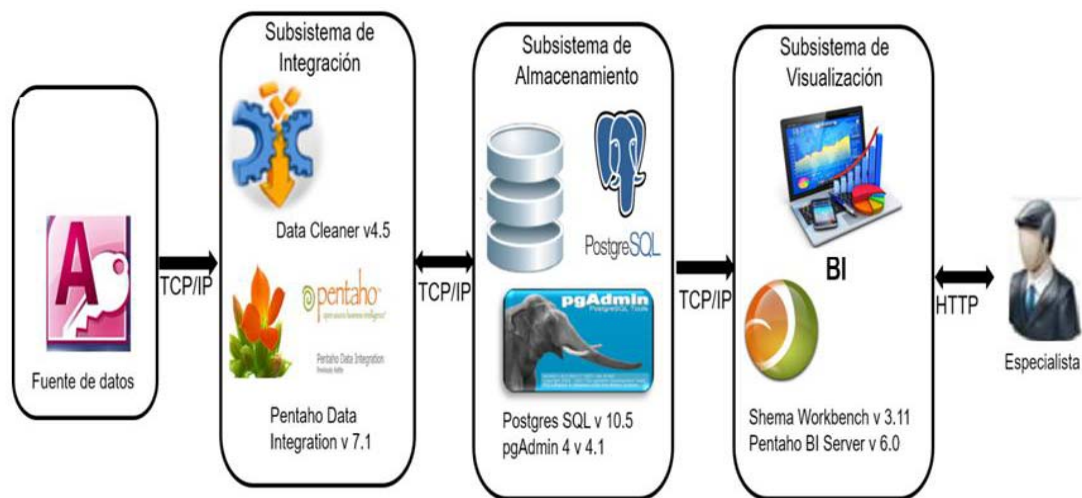
**Figura 1:** Ciclo de vida de la Metodología de Desarrollo para proyectos de AD. Fuente: Elaboración Propia.

El MD para el proceso de perfeccionamiento educacional del ICCP está ordenado mediante una arquitectura compuesta por tres subsistemas: integración, almacenamiento y visualización. A continuación, se presenta una descripción de cada uno de estos subsistemas.

- Subsistema de integración de los datos: es donde se realizan todos los procesos extracción, transformación y carga (ETL), constituido por varios subprocesos: limpieza, carga, extracción e integración de los datos provenientes de la fuente en ficheros Access.
- Subsistema de almacenamiento de los datos: es donde se guarda toda la información que ha sido transformada en el subsistema de integración. Contiene los esquemas y dentro de estos los hechos y dimensiones identificados en el MD.
- Subsistema de visualización de los datos: es donde se muestra toda la información almacenada al cliente, a través de cuadros de mando y vistas de análisis. Los mismos permiten al cliente realizar un análisis de toda la información procesada. Además de que se encarga principalmente de la visualización mediante gráficos de la información almacenada agrupándolas en libro de trabajo que se corresponden con el área de análisis identificada.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)



**Figura 2:** Arquitectura diseñada para el MD del proceso de perfeccionamiento educacional del ICCP. Fuente: Elaboración Propia.

La fuente de datos está compuesta por ficheros Access que contienen los datos del experimento. Primeramente, se encuentra el subsistema de integración donde se realizan los procesos de ETL. En estos procesos se limpian e integran los datos a través de transformaciones y trabajos utilizando herramientas como el Data Cleaner en su v4.5 para identificar varios aspectos de los datos como el tamaño máximo y mínimo de un campo y las distribuciones de los tipos de dato, además del Pentaho Data Integration en su v7.1 para la limpieza, las transformaciones y la carga de los datos. Este subsistema siguiendo un protocolo TCP/IP se relaciona con el subsistema de almacenamiento, y viceversa, en este último se almacenan todos los datos que se cargan desde las diferentes fuentes estadísticas. Los datos que son cargados al mercado se separan en esquemas de dimensiones y hechos, facilitando así la organización del mismo. Éste, recibe la información integrada y la almacena en la base de datos destino: dm\_iccp, la cual es soportada por el SGBD PostgreSQL v10.5 y administrada por los usuarios que tienen acceso a la información a través del pgAdmin4 v4.1; a su vez, a través del protocolo TCP/IP se relaciona con el subsistema de visualización, donde se muestran los datos de forma multidimensional al cliente, mediante el protocolo http. Además, el cliente puede acceder al sistema mediante el mismo protocolo. En este subsistema son mostrados los diferentes niveles de información mediante la herramienta Pentaho Schema Workbench v3.11 donde se crean los cubos OLAP y el Pentaho Business Intelligence Server v6.0 que muestra la información a través de vistas de análisis y cuadros de mando de tal forma que es visualizada multidimensionalmente y accedida por los usuarios autorizados.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

## Resultados y discusión

Para el desarrollo de la solución, es de vital importancia conocer las necesidades del cliente, por lo que fueron realizadas un conjunto de entrevistas a los especialistas del ICCP. Los requisitos de información fueron obtenidos analizando la información recopilada durante el proceso de perfeccionamiento educacional a nivel nacional desde el año 2011, contenida en los 42 ficheros Access. A partir de esto se identificó la necesidad de almacenar e integrar la información, de modo que las realizaciones de los reportes pertinentes sean más certeras y se contribuya a la toma de decisiones. Fueron identificados tres Áreas, los que permiten que los datos se organicen para facilitar su acceso y entendimiento por parte de los usuarios finales:

1. Actividades Educativas.
2. Actividades Metodológicas.
3. Clases.

**Análisis de los sistemas fuentes:** Los sistemas fuentes están constituidos por 42 ficheros Acces que almacenan los resultados de la evaluación de las actividades educativas, metodológicas y las clases de cada uno de los centros educacionales en los que se lleva a cabo el proceso de perfeccionamiento del sistema de educación. Haciendo uso de la herramienta Data Cleaner como parte del tratamiento a los datos almacenados, se analizaron las fuentes con el fin de comprender a profundidad todos sus detalles en cuanto a estructura y contenido, para determinar entonces las transformaciones que se le deben hacer a los mismos. Este proceso arrojó como resultado que los valores que se encontraban almacenados poseen tipos de datos como: textos, números enteros y fechas, la máscara de fecha mayormente utilizada es del tipo: día/mes/año además de que se encontraron valores nulos en algunas tuplas. Un correcto análisis de los sistemas fuente nos permite definir además las reglas del negocio, aspecto a tener en cuenta en el los procesos de integración y visualización de la información.

**Modelo de datos:** Durante el diseño del subsistema de almacenamiento son identificadas las tablas dimensiones y las tablas hechos, elementos imprescindibles para la realización de dicho subsistema. Las tablas de hechos contienen los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones y entre su contenido están los datos cuantitativos. Por su parte las tablas de dimensiones son elementos que contienen atributos utilizados para restringir y agrupar los datos que participan en el análisis. Las dimensiones poseen entre sus características principales la definición de jerarquías entre sus atributos, que tienen como objetivo plasmar explícitamente la forma en que se puede consolidar. Son las compañeras integrales de las tablas de hechos, ellas contienen la descripción textual del negocio. Para el MD de la presente investigación se identificaron 9 tablas de



hechos y 15 tablas de dimensiones. Una vez ya definidos los mismos se conforma el modelo de datos del MD propuesto, modelo que tiene como objetivo: representar la estructura lógica de la información, evidenciando además las relaciones existentes entre las tablas de hechos y las tablas dimensiones.

El modelo de datos para el desarrollo del MD para el proceso de perfeccionamiento educacional del ICCP, se basa en un esquema de constelación de hechos, siendo éste el más adecuado a las necesidades del experimento llevado a cabo por el ICCP, pues existen dimensiones compartidas para distintas tablas de hechos. En la figura 3 se muestra un fragmento del modelo de datos del MD propuesto.

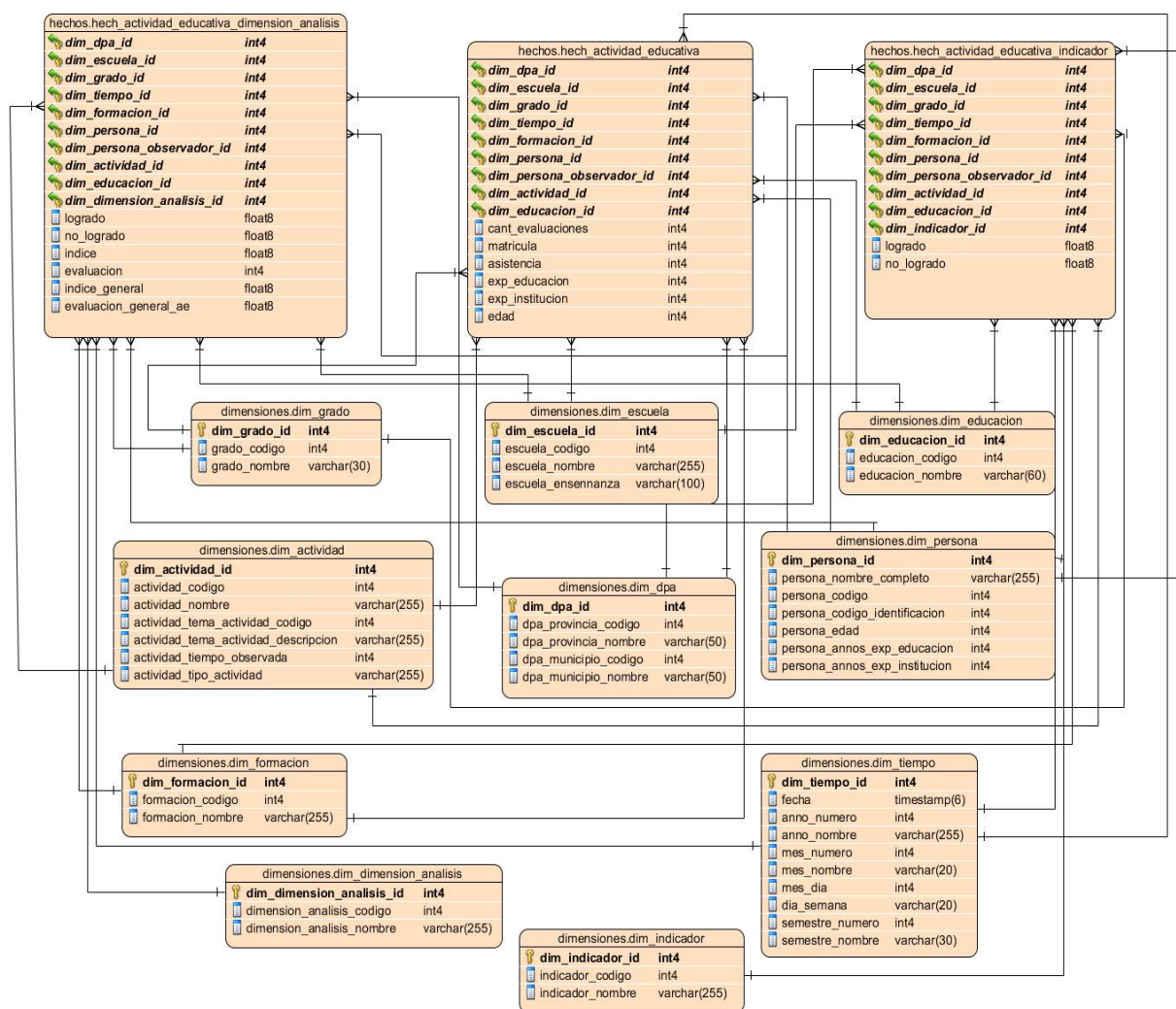


Figura 3: Modelo de datos.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)



A partir del modelo lógico de los datos se genera el modelo físico, constituido por una base de datos relacional en el sistema gestor de bases de datos PostgreSQL. La base de datos está compuesta por tres esquemas: hechos, dimensiones y metadatos.

**Proceso de integración de datos:** El proceso de integración de los datos consta de tres etapas fundamentales relacionadas entre sí: extracción, transformación y carga de los datos. En la etapa de extracción con el uso del componente Microsoft Acces Input se obtienen los datos desde los sistemas fuentes, luego haciendo uso de los componentes de transformación que ofrece la herramienta Pentaho Data Integration se transforman y limpian permitiendo llenar valores ausentes y corrigiendo errores y posteriormente son almacenados en los esquemas correspondientes mediante el componente Insertar/Actualizar. En la presente investigación se realizó un flujo de transformación para la carga de cada una de las tablas de hechos y dimensiones que forman parte del MD. A continuación se muestra el flujo de transformación de la dimensión indicador (*dim\_indicador*) y el hecho actividades educativas (*hech\_actividad\_educativa*).

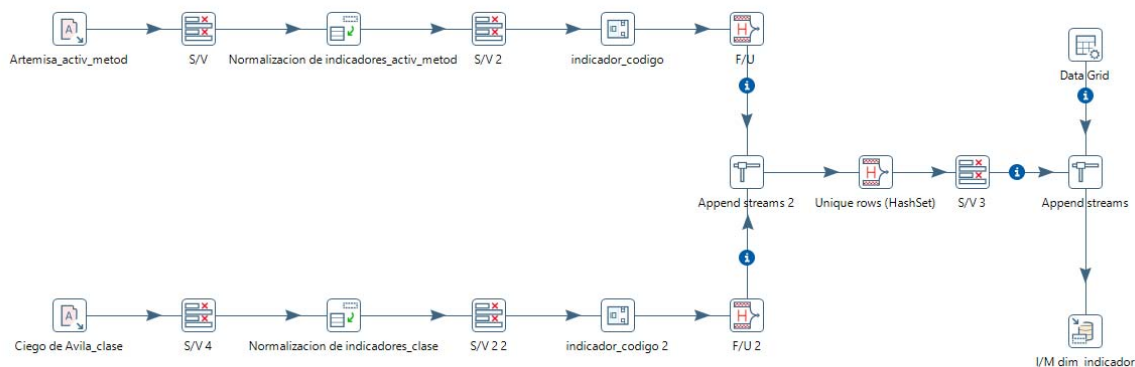


Figura 4: Transformación de la dimensión indicador.



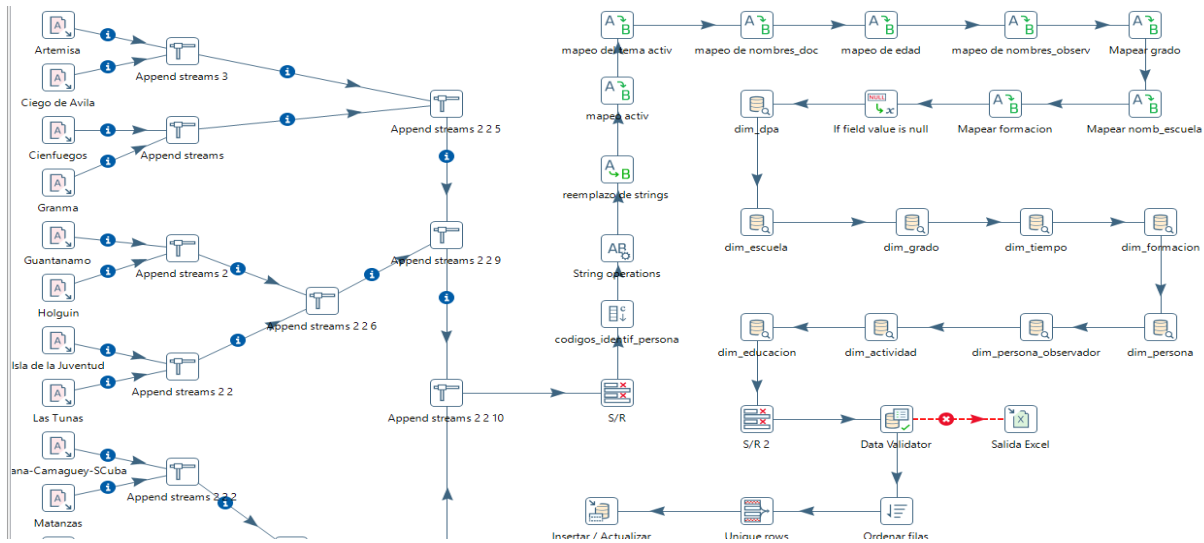


Figura 5: Transformación del hecho actividad educativa.

Una vez realizadas las transformaciones para cada uno de los hechos y dimensiones, se automatiza el proceso de carga mediante el uso de trabajos. Estos permiten ejecutar varias transformaciones o trabajos previamente diseñados y organizar una secuencia de ejecución de éstos. Los trabajos se encuentran en un nivel superior de las transformaciones. En la presente investigación se realizó 1 trabajo para la carga de las dimensiones( que contiene 5 trabajos en la secuencia de carga de las mismas), 1 para la carga de los hechos y un trabajo general para la carga de los 2 trabajos anteriores. A continuación se muestra el trabajo general.



Figura 6: Trabajo General.

**Subsistema de Visualización:** Luego de realizado el proceso de ETL, se procede a la creación de los cubos OLAP y las vistas de análisis para la visualización y análisis de la información. Mediante la herramienta Shema Worbench se crean los cubos OLAP, correspondiendo un cubo por cada hecho que contiene el MD (Ver Fig. 7). Luego haciendo



uso del Pentaho BI Server se podrán realizar vistas de análisis y cuadros de mando con el objetivo de obtener información relevante de los datos que se encuentran almacenados para contribuir a la toma de decisiones.

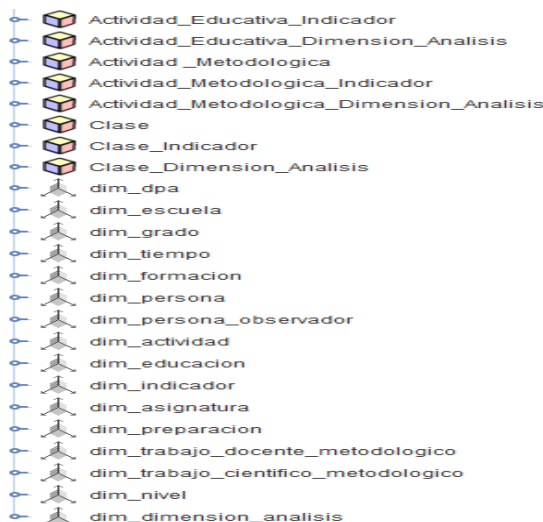


Figura 7: Implementación de los Cubos OLAP.

Con el objetivo de que los usuarios tengan una visión de cómo está estructurada la información referente al MD del ICCP se crea el mapa de navegación, que contribuye a definir el entorno de análisis del sistema. El Área de Análisis General (A.A.G) Proceso de Perfeccionamiento Educativo del ICCP, está dividida en tres Áreas de Análisis (A.A). El A.A Actividades Educativas, contiene cinco Libros de Trabajo (L.T). El A.A Actividades Metodológicas contiene cuatro L.T, y por último el A.A Clases que cuenta con cinco LT.

**A.A.G Proceso de Perfeccionamiento Educativo del ICCP:** Agrupa la información referente al proceso de perfeccionamiento educativo del ICCP.

**A.A Actividades Educativas:** Contiene toda la información referente a las actividades educativas observadas (dpa, escuela, grado, tiempo, formación, persona observada y persona observador, actividad y educación.)

**A.A Actividades Metodológicas:** Almacena la información referente a las actividades metodológicas (dpa, escuela, tiempo, asignatura, persona, nivel, preparación, trabajo docente metodológico).

**A.A Clases:** Contiene la información referente a las clases observadas (dpa, escuela, tiempo, asignatura, persona observada y persona observador, actividad, grado, formación y educación).

**LT Indicadores:** Contiene información acerca de los 28 indicadores que miden el proceso de perfeccionamiento educativo.



**LT Dimensiones:** Almacena información acerca de las dimensiones que se calculan apartir de un conjunto de indicadores determinado.

**LT Edad-Experiencia:** Recoge la información referente a los promedios de edad y experiencia de los involucrados en el experimento.

**LT Índices:** Contiene la información relacionada a los índices generales y las evaluaciones promedios de las dimensiones de análisis y las actividades educativas.

**LT Participantes e Intervenciones:** Contiene la información que relaciona las cantidades de evaluaciones con la cantidad de intervenciones y los participantes en las mismas. La Fig. 8 muestra la arquitectura de información diseñada para el MD para el ICCP.

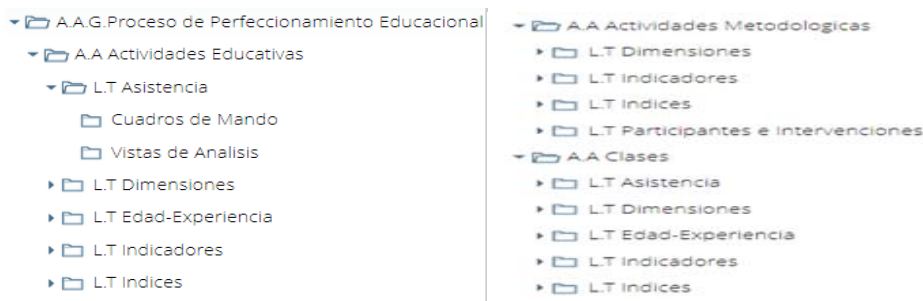


Figura 8: Arquitectura de información diseñada para el MD para el ICCP

Una vez poblado el MD se podrán realizar vistas de análisis y cuadros de mando con el objetivo de obtener información relevante de los datos que se encuentran almacenados para contribuir a la toma de decisiones. Los mismos se deben mostrar en formato PDF y Excel. A continuación, se muestran algunas vistas que se pueden observar en las siguientes figuras.



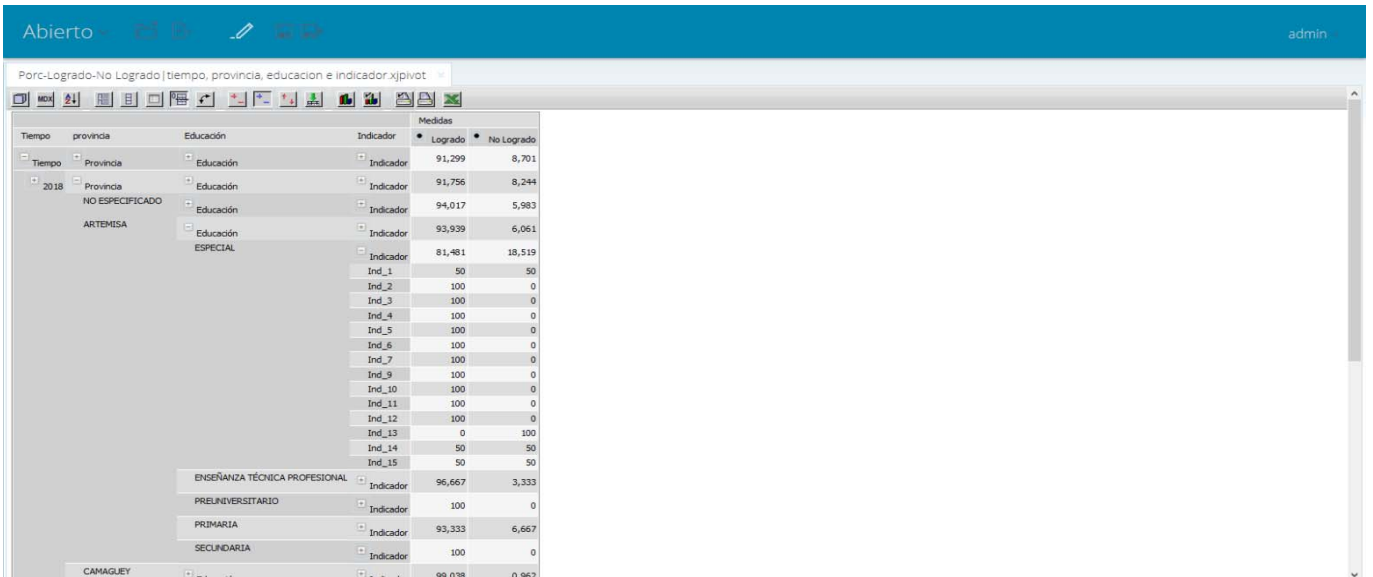


Figura 9: Vista de Análisis Porcentaje Logrado y No Logrado por tiempo, provincia, educación e indicador.

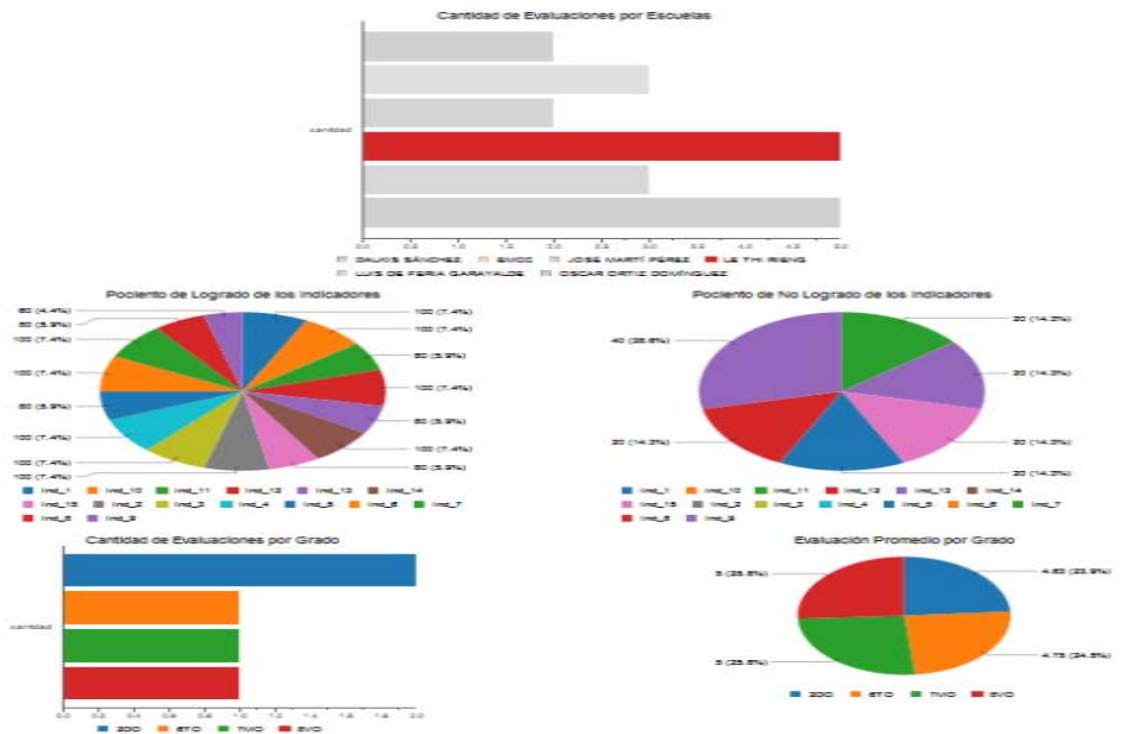


Figura 10: Cuadro de Mando Estadísticas los Resultados de los Indicadores.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

## Conclusiones

Al finalizar la presente investigación se concluye lo siguiente:

1. Las soluciones de AD facilitan el análisis de la información y tienen como principal objetivo servir de apoyo a la toma de decisiones.
2. Siguiendo el ciclo de vida de la metodología seleccionada se obtuvo un software con la calidad requerida.
3. Mediante los procesos de ETL se logró la transformación e integración de la información del proceso de perfeccionamiento educacional contenida en los 42 ficheros Acces.
4. La correcta carga de los datos permitió obtener un MD poblado y funcional.
5. En la capa de visualización se muestran vistas de análisis y cuadros de mando que permiten analizar la información desde diferentes perspectivas.

## Conflictos de intereses

Los autores de la presente investigación declaran que no poseen conflicto de intereses.

## Contribución de los autores

1. Conceptualización: Irela González Piñera, Amarelys Alvarez Pérez.
2. Curación de datos: Irela González Piñera, Amarelys Alvarez Pérez.
3. Análisis formal: Irela González Piñera , Amarelys Alvarez Pérez.
4. Investigación: Irela González Piñera , Arodys Eugenio Dominguez Vaillant.
5. Metodología: Irela González Piñera, Arodys Eugenio Dominguez Vaillant.
6. Administración del proyecto: Irela González Piñera.
7. Software: Irela González Piñera, Amarelys Alvarez Pérez, Arodys Eugenio Dominguez Vaillant.
8. Supervisión: Arodys Eugenio Dominguez Vaillant.
9. Validación: Arodys Eugenio Dominguez Vaillant.
10. Visualización: Arodys Eugenio Dominguez Vaillant.
11. Redacción – borrador original: Irela González Piñera, Amarelys Alvarez Pérez, Arodys Eugenio Dominguez Vaillant.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

12. Redacción – revisión y edición: Nombre y Apellidos del autor Irela González Piñera, Amarelys Alvarez Pérez, Arodys Eugenio Dominguez Vaillant.

## Financiamiento

La investigación no requirió fuente de financiamiento.

## Referencias

- Hernández, Y. G. (2013). Metodología de desarrollo para proyectos de almacenes de datos. Universidad de las Ciencias Informáticas. La Habana.
- Inmon, W. H. (2005). Building the Data Warehouse. John Wiley.
- Kimball, R. (2004). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Publishing.
- Kimball, R. (2008). The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems.
- Martínez, R. (2013). PostgreSQL-es. PostgreSQL-es. Obtenido de [http://www.postgresql.org/es/sobre\\_postgresql](http://www.postgresql.org/es/sobre_postgresql)
- Orallo, J. H. (2010). Data Warehouse, o Almacén de Datos, y Cubo OLAP. El Mejor Proyecto. Pentaho. (2015). Pentaho Suite. Obtenido de <https://sites.google.com/site/pentahounicah/informacion-general>
- Rivera Gómez, O. J. (2015). Sistema ETL para el almacén de datos de aspirantes de la UCLV, Universidad Central “Marta Abreu” de Las Villas.
- KIMBALL, R. y INMON, B., 2010. Ampliación de conceptos del Modelado Dimensional. El Rincón del BI [en línea]. [Consulta: 12 noviembre 2018]. Disponible en: <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modeladodimensional/>.
- SOFTWARE, E., 2017. ETL: Extracción, transformación y carga de datos. Evaluando Software [en línea]. [Consulta: 26 noviembre 2018]. Disponible en: <http://www.evaluandosoftware.com/etl-extracciontransformacion-carga-datos/>.
- Pressman, Roger S. Ingeniería de Software, un enfoque práctico. Quinta Edición. S.I.: McGraw-Hill Companies, 2002. ISBN: 8448132149
- Perdomo, C. R. B. (2009). Estudio de la interacción en la consulta y visualización de Almacenes de Datos. Trabajo de Maestría, UNIVERSIDAD AUTÓNOMA DE MADRID.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)