

Almacén de datos para la vigilancia epidemiológica en la provincia Santiago de Cuba.

Quintana Casulo, Yasser¹
Goire Castilla, María Margarita²
Cantero Díaz, Alexander³

¹ Universidad de Oriente/Ingeniería Informática, Santiago de Cuba, Cuba, yasser.quintana@uo.edu.cu

² Universidad de Oriente/Ingeniería en Telecomunicaciones y Electrónica, Santiago de Cuba, Cuba, mgoire@uo.edu.cu

³ Universidad de Oriente/ Ingeniería en Telecomunicaciones y Electrónica, Santiago de Cuba, Cuba, acantero@uo.edu.cu

Resumen: Un problema crítico para el Centro Provincial de Higiene y Epidemiología de la provincia de Santiago de Cuba, es el análisis de la información de las diferentes áreas de la provincia para la toma de decisiones. Esta problemática es provocada por el gran volumen de datos sanitarios y estadísticos que maneja el Departamento de Vigilancia Epidemiológica y Análisis en Salud. Actualmente, toda esta información es tratada de forma semi-manual mediante el uso de documentos en Word, tablas en Excel y diversos sistemas estadísticos de forma independiente. Esta dispersión de la información en distintos sistemas, hace lento e ineficiente el análisis de los datos, lo que puede provocar poca efectividad en las acciones sanitarias a tomar. En este trabajo se presenta una propuesta de solución al problema expuesto que consistió en el diseño e implementación de un almacén de datos como herramienta de la inteligencia de negocio. El mismo debe garantizar una mejor organización, integración y acceso de la información, lo que admite realizar el estudio de la vigilancia epidemiológica y la toma de decisiones correspondientes. Para el diseño del almacén de datos se usó la metodología Hefesto y para la implementación se usó la Suite Pentaho BI. La solución planteada permitió integrar en un solo sistema los cálculos estadísticos, el análisis de los datos y los reportes de estos, propiciando un minucioso control epidemiológico. La propuesta se estructuró en 25 tablas: 18 dimensiones y 7 tablas de hechos.

Palabras clave: almacén de datos, vigilancia epidemiológica, inteligencia de negocio, toma de decisiones.

I. INTRODUCCIÓN

Con la informatización de la sociedad ha crecido a nivel mundial la capacidad de generación y almacenamiento de la información. Lo que ha provocado que esta no pueda ser analizada por los métodos tradicionales existentes, mientras mayor es la capacidad para almacenar más y más datos, mayor es la incapacidad para extraer información realmente útil que le sirva al usuario.

Ante esta problemática el hombre ha buscado como solución la Inteligencia de negocios, la clave de esta radica en las tecnologías que emplea para la exploración y análisis de los datos para una futura toma de decisiones, lo que se traduce en una ventaja competitiva. Las formas de organización de los datos se llevarán a cabo en los almacenes de datos los cuales son el centro de la arquitectura para los sistemas de información. Los almacenes de datos aseguran una vista única de los datos y una comprensión de los mismos; y están enfocado hacia la toma de decisiones por parte del personal.

El Centro Provincial de Higiene, Epidemiología y Microbiología (CPHEM) cuenta con el Departamento de Vigilancia Epidemiológica y Análisis en Salud (DVEAS), en este realizan diariamente una recepción de un gran flujo de información de 17 áreas de salud y los hospitales del municipio de Santiago de Cuba como: la vigilancia y lucha Antivectorial, la atención materno infantil, la salud ambiental, las enfermedades transmisibles y no transmisibles, y otros indicadores de salud. Pero el cumulo de información hace engorroso su examen y procesamiento y por eso el uso de sistemas basados en almacenes de datos sería una gran oportunidad. La metodología más adecuada es la que propone Bill Inmon que sea descendente, es decir, tener un almacén de datos y que mercados de datos se nutran de la información del almacén de datos (1).

Dicho departamento, logro del Ministerio de Salud Pública, ha dado pasos firmes hacia la vigilancia en salud, este como instancia integradora constituye un órgano de asesoría directa capaz de brindar el conocimiento oportuno y relevante de la situación de salud. Además, que deben constituir unidades de inteligencia epidemiológica, de análisis permanente, capaces de generar conocimientos a partir de la información existente y de mantener una visión integradora en el análisis de la situación de salud, con el reto de brindar respuestas eventuales o permanentes a los problemas de salud, con resultados útiles para las decisiones por los niveles ejecutivos (2).

Partiendo de lo antes planteado podemos decir que el problema de la investigación es que, debido a la existencia de una amplia gama de datos estadísticos y sanitarios en el DVEAS, unidos a la carencia de un sistema único que permita su gestión en tiempo real y accesible para cualquier decisor dentro del sector y desde cualquier escenario, condiciona limitaciones para el apoyo a la toma de decisiones oportuna; por lo que el objetivo general vendría dado en la implementación de un sistema para el apoyo a la toma de decisiones basado en almacenes de datos que permita realizar análisis de la vigilancia epidemiológica en el DVEAS de Santiago de Cuba. Por lo que se definen como objetivos específicos:

1. Definir el estado actual de la toma de decisiones a partir de los datos estadísticos correspondientes a la vigilancia epidemiológica.

2. Implementar un almacén de datos que permita el análisis de la información estadística que se recoge en los distintos centros de salud.

3. Realizar los reportes diarios para la difusión de la información.

Teniendo en cuenta lo antes descrito podemos decir que, con el desarrollo de un sistema de soporte a la toma de decisiones, basado en un almacén de datos, que permita el análisis de la vigilancia epidemiológica, se aumentará la efectividad y la eficiencia en el análisis de la información, la realización de pronósticos y toma de decisiones para la gestión de problemas sanitarios en el orden epidemiológico.

II. MÉTODO

A. Inteligencia de Negocios

El término inteligencia de negocios surgió en el Grupo Gartner para abarcar las metodologías e ideas que conlleven al aumento de la eficiencia y la rentabilidad empresarial en el lugar de trabajo. La inteligencia de negocios es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un almacén de datos), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones (3).

En la Fig. 1 se representa la arquitectura de la inteligencia de negocios donde se refleja sus tres características más importantes: accesibilidad a la información, apoyo en la toma de decisiones y orientación al usuario final (4).



Fig. 1 Arquitectura de la Inteligencia de Negocios.

B. Almacenes de datos

Un almacén de datos, surge con el objetivo de hacer accesible la información que se tiene de una empresa tanto de meses como de años anteriores. Es un almacenamiento simple, completo y consistente de datos obtenidos desde una variedad de fuentes, disponibles para el usuario final de forma tal que puedan entenderlos y utilizarlos. Facilita la integración de sistemas de aplicación no integrados. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis que se necesita de la información. El objetivo fundamental es asistir al usuario en el entendimiento del pasado y contar con los elementos para la planeación del futuro de corto, mediano y largo plazo para la toma de decisiones. Entre sus principales características se puede destacar que: es orientado al negocio, integrado, variable en el tiempo y no volátil (5).

El uso de un almacén de datos en el desarrollo del sistema en cuestión fue de notable importancia por la necesidad de obtener información con una visión histórica de los diferentes indicadores de salud y se manejaron grandes volúmenes de datos para los análisis y comparaciones en el apoyo a la toma de decisiones.

Para el desarrollo y creación del almacén de datos se hace necesaria la utilización de una metodología de desarrollo y la más adecuada es la Hefesto. Esta metodología está bien fundamentada y explícita, permite crear un almacén de datos de manera metódica y sencilla, guiándose por pasos lógicos relacionados (6).

C. Modelo multidimensional

El modelo multidimensional asocia la información con los procesos del negocio dentro de dos categorías llamadas hechos y dimensiones. Cada modelo está compuesto por una tabla con múltiples llaves llamada tabla de hechos y un grupo de pequeñas tablas llamadas dimensiones. Cada tabla dimensión

tiene una llave primaria que corresponde exactamente con una de las múltiples llaves en la tabla de hechos. Las creaciones de los cubos multidimensionales permitirán el análisis con gran agilidad y rapidez reduciendo el tiempo y recursos.

La variante de base de datos multidimensional empleada en nuestro sistema es el esquema constelación la cual está formada por varias tablas de hechos y tablas de dimensiones como se ve en la Fig. 2.

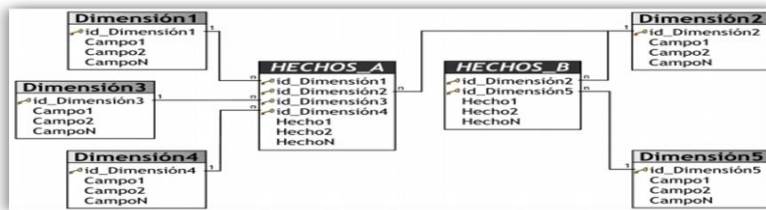


Fig. 2 Esquema Constelación.

El tipo de implementación usado es el ROLAP (Relational On Line Analytic Processing) ya que los indicadores, atributos y jerarquías se seleccionan para la construcción del cubo y el mismo se calcula dinámicamente para realizar consultas de manera transparente al usuario.

D. Suite Pentaho como herramienta de Inteligencia de Negocios.

La plataforma usada para la creación de consultas y ejecución de tareas fue Pentaho Data Integration en su versión 5,3. Esta herramienta posee un planificador de tareas para la ejecución automática y unión de estas, además permite la conexión a varios gestores de base de datos, para luego transformarlos y llevarlos a un formato factible en un almacén de datos, y así sean analizados.

El Schema Workbench en su versión 3,0,3 es una interfaz de diseño que permite crear y probar esquemas de cubos OLAP. Con esta aplicación, se configura una conexión con el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello la herramienta ofrece un editor de esquemas con la fuente de datos subyacente para su validación. Los modelos XML se pueden considerar como las estructuras en forma de cubo que utilizan hechos existentes y tablas de dimensiones.

La aplicación web empleada fue el Pentaho BI Server en su versión 4,8 la cual es 100% Java2EE y cuenta con una interfaz de usuario que permite analizar todas las informaciones. Además de proveer la arquitectura y la infraestructura necesarias para crear soluciones de inteligencia de negocio.

El visor OLAP empleado fue el STPivot en sustitución del JPivot, desarrollado para código abierto la idea era mejorar la experiencia del usuario de su interfaz e interacción con el mismo mediante el aprovechamiento de las bibliotecas de interfaz de usuario libre y las tecnologías como jQuery y Ajax.

El Pentaho Report Designer en su versión 3,4 permite trabajar con múltiples orígenes de datos y es un generador de informes que nos permite exportar en diferentes formatos.

El gestor de base de datos usada para almacenar toda la información histórica fue PostgreSQL en la versión 8,4 funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema.

III. RESULTADOS

El ciclo de vida del sistema se representa a través de la metodología Hefesto. La misma lleva una serie de pasos que se describen a continuación para la implementación de un almacén de datos.

A. Análisis y requerimientos

Para el análisis de los requerimientos se tuvieron en cuenta las declaraciones por parte del usuario de las condiciones que se deben cumplir para un buen desarrollo e implementación del sistema que expliciten los objetivos de su organización. Tienen como propósito definir e identificar cuáles son los indicadores y perspectivas que serán tomadas en cuenta para la construcción del almacén de datos.

En las entrevistas realizadas al usuario se definieron las actividades que se corresponden al departamento de epidemiología y se escogieron los procesos más importantes que se desarrollan. Algunas de estas actividades son:

- Cantidad de casos totales de brotes según un tipo de brotes del área de salud, municipio y provincia en una fecha determinada.
- Cantidad de casos totales de vectores según un tipo de vector del área de salud, municipio y provincia en una fecha determinada.

B. Modelo Conceptual

El siguiente paso consiste en la confección del modelo conceptual, que se realiza a partir de los indicadores y perspectivas obtenidas, lo que permitirá observar con claridad el alcance del almacén de datos. En la Fig. 3 se refleja una muestra de dicho modelo donde se ve la relación entre los indicadores y perspectivas según los descritos en los análisis y requerimientos. A la izquierda se colocan las perspectivas seleccionadas, a la derecha los indicadores y estos serán unidos por la relación que constituye el proceso o área de estudio elegida.

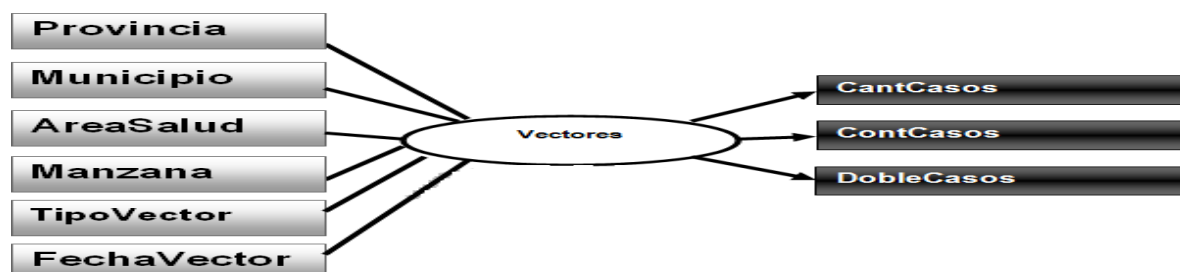


Fig. 3 Modelo conceptual de Vectores.

C. Análisis de los OLTP

En este paso se analizan los OLTP para determinar cómo serán calculados los indicadores y así establecer las correspondencias entre el modelo conceptual creado y las fuentes de datos, además definir los campos que se incluirán en cada perspectiva, obteniendo así un modelo conceptual ampliado. Toda esta información está recogida en otro documento, donde cada modelo conceptual, los indicadores, la correspondencia y el modelo conceptual ampliado están reflejados en dos hojas.

D. Modelo del almacén de datos.

Con la creación del almacén de datos ya se tienen las condiciones para la realización de cálculos estadísticos y análisis de los datos. El mismo está estructurado por 25 tablas: 18 dimensiones y 7 tablas de

hechos. Para que los datos puedan ser analizados a través de la creación de consultas se necesita una estructura de datos determinada, estos son los cubos multidimensionales, los cuales permiten el acceso flexible a los datos, para explorar y analizar sus relaciones, y consiguientes resultados. A continuación, se describe el proceso ETL y la estructura e implementación del cubo multidimensional.

E. Proceso ETL

Para la implementación del proceso ETL se necesita de la base de datos relacional, a partir de la cual se cargan los datos hacia la base de datos multidimensional. Este proceso tiene tres funciones específicas: extracción, transformación y carga, a continuación, se explican en las Fig. 4 y Fig. 5.

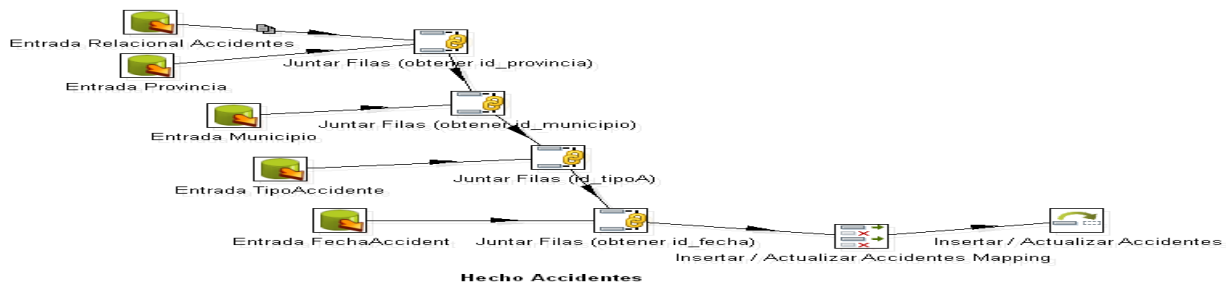


Fig. 4 Dimensión Hecho Accidente

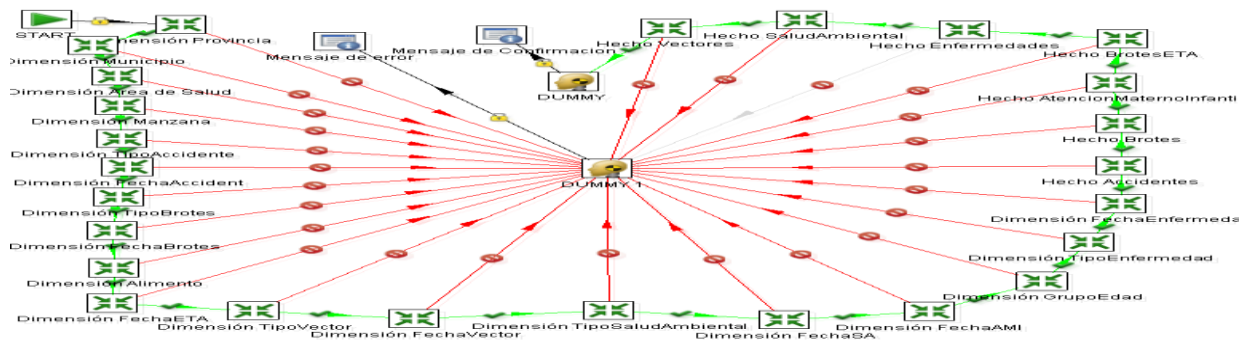


Fig. 5 Carga del trabajo

F. Proceso OLAP

Luego de terminar el proceso ETL, se conforma el cubo de datos, este es fundamental para realizar la navegación por los datos con el fin de extraer la información deseada de forma rápida y eficiente.

El cubo está diseñado para el esquema constelación el cual, como se muestra en la Fig. 8 tiene dimensiones globales como son provincia, municipio, área de salud, y manzana ya que las mismas son utilizadas en varios de los cubos diseñados. Estos cubos son los que representan cada tabla de hecho del almacén, los mismos al ser creados desde el esquema base se les crea la relación con las dimensiones que le corresponden y las respectivas jerarquías para cada una. En el caso de las dimensiones globales solo necesitan ser invocadas dentro del cubo especificando el id y el nombre.

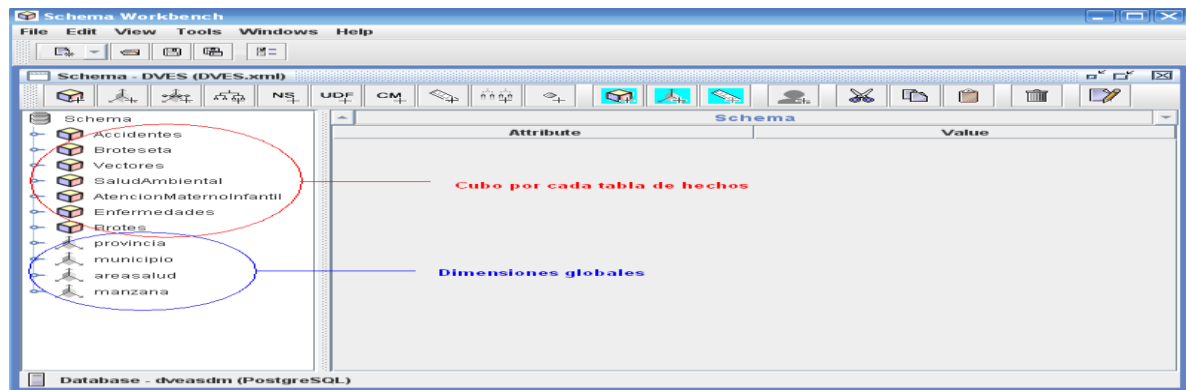


Fig. 8 Cubo OLAP

A partir de la obtención del cubo multidimensional se hará uso de la herramienta Pentaho BI Suite donde se publica el cubo anteriormente realizado y se procederá al análisis de los datos del departamento realizando consultas, donde se obtienen tablas y gráficos que permiten una mejor comprensión de los datos apoyando así al proceso de toma de decisiones. A continuación, aparecen las Fig. 9 y Fig. 10 algunas de estas interfaces, análisis y reportes.

provincia	municipio	areasalud	manzana	Tipo Vector	Fecha Vector	Cantidad Focos	Varianza Focos
All provincias	All municipios	All areasaluds	All manzanas	All tipovectores	All fechavectores	94.980	811,928
Santiago de Cuba	All municipios	All areasaluds	All manzanas	All tipovectores	All fechavectores	94.980	811,928
	Santiago de Cuba	All areasaluds	All manzanas	All tipovectores	All fechavectores	94.980	811,928
		28 de septiembre	All manzanas	All tipovectores	All fechavectores	3.892	949,677
		30 de noviembre	All manzanas	All tipovectores	All fechavectores	3.999	767,769
		Armando Garcia	All manzanas	All tipovectores	All fechavectores	5.261	841,252
		Boniato	All manzanas	All tipovectores	All fechavectores	7.349	766,773
		Camilo Torres	All manzanas	All tipovectores	All fechavectores	5.882	856,649
		Caney	All manzanas	All tipovectores	All fechavectores	5.452	727,764
		Cobre	All manzanas	All tipovectores	All fechavectores	4.831	779,133
		Cristo	All manzanas	All tipovectores	All fechavectores	5.706	766,148
		Finlay	All manzanas	All tipovectores	All fechavectores	5.542	830,528
		Frank Pais	All manzanas	All tipovectores	All fechavectores	4.717	707,756
		Grimau	All manzanas	All tipovectores	All fechavectores	5.580	882,271
		Hospital Clinico Q.	All manzanas	All tipovectores	All fechavectores	5.384	763,493
		Hospital Militar	All manzanas	All tipovectores	All fechavectores	4.297	771,431
		Hospital Provincial	All manzanas	All tipovectores	All fechavectores	5.228	838,669
		Jose Marti	All manzanas	All tipovectores	All fechavectores	5.076	733,493
		Jose Pais	All manzanas	All tipovectores	All fechavectores	3.961	767,441
		Municipal	All manzanas	All tipovectores	All fechavectores	3.513	935,423
		Pimental	All manzanas	All tipovectores	All fechavectores	3.743	911,525
		Siboney	All manzanas	All tipovectores	All fechavectores	5.567	742,262

Fig. 9 Análisis de vectores por área de salud

Output Type		PDF		View Report		Auto-Submit	
Edición Limitada		mayo 27, 2013					
Municipio	Area	manzana	ab	mes	semana	diatipovector	casos
Santiago de Cuba	28 de septiembre	7.002	2.013	1	2	17 Aedes	82
Santiago de Cuba	30 de noviembre	3.041	2.013	1	1	7 Aedes	67
Santiago de Cuba	30 de noviembre	4.155	2.013	1	2	17 Aedes	78
Santiago de Cuba	30 de noviembre	5.388	2.013	2	6	16 Aedes	37
Santiago de Cuba	30 de noviembre	7.677	2.013	4	13	3 Aedes	81
Santiago de Cuba	Armando Garcia	2.304	2.013	2	4	1 Aedes	82
Santiago de Cuba	Armando Garcia	3.867	2.013	4	15	20 Aedes	41
Santiago de Cuba	Armando Garcia	4.973	2.013	4	13	3 Aedes	61
Santiago de Cuba	Armando Garcia	5.252	2.013	5	17	2 Aedes	67
Santiago de Cuba	Armando Garcia	6.898	2.013	4	16	16 Aedes	3

Fig. 10 Reporte de vectores

IV. CONCLUSIONES

Con la presente investigación se logró:

- Un análisis minucioso del proceso que se lleva a cabo en la vigilancia epidemiológica.
- El estudio de las definiciones, herramientas y conceptos de un almacén de datos, lo cual fue de gran importancia como apoyo para la correcta implementación del almacén de datos en el análisis de las salas de situación de salud en la población.
- El análisis y diseño del almacén de datos para el análisis de la sala de situación de salud de la población de Santiago, guiado por la metodología Hefesto.
- La implementación de un almacén de datos que contribuye al análisis minucioso del control epidemiológico en la población de Santiago.

REFERENCIAS

1. Inmon WH. Building the Data Warehouse. 3rd ed. Wiley Computer Publishing; 2002.
2. Sánchez N. Las Unidades de Análisis y Tendencias en Salud dentro del Sistema de Vigilancia en Salud. Unidad de Análisis y Tendencias de Salud: Ciudad de La Habana; 2001. p. 2-11
3. Romero DL, Vargas CA. Diseño de prototipo para la implementación de un sistema BIG DATA [tesis]. Facultad de Ingeniería: Universidad Politécnica Gran Colombiano; 2015
4. Ríos, A. ¿Qué es la inteligencia de negocios? Inteligencia de Negocio [revista de internet] 2009. [acceso 3 febrero 2015]. Disponible en: <http://www.oracle.com/technology/global/lades/documentation/collaterals/Que-es-inteligencia-de-Negocios-Whitepaper-Angel-Rios.pdf>.
5. Quiroz JD, Reyes JA. Implementación de un BI para analizar la provisión de servicios de salud en los últimos 8 años, utilizando PostgreSQL y Pentaho [tesis]. Quito: Universidad Politécnica Salesiana; 2017
6. Bernabéu RD. HEFESTO: Metodología propia para la Construcción de un Data Warehouse; 2009