

DATAMART BASADO EN EL MODELO ESTRELLA PARA LA IMPLEMENTACIÓN DE INDICADORES CLAVE DE DESEMPEÑO COMO SALIDA DEL BIG DATA

Zerpa Héctor¹, García Richard², Izquierdo Henry³.

hzerpa@unexpo.edu.ve, {richard.garcia.lr, hizquier}@gmail.com
<https://orcid.org/0000-0003-0499-2089>, <https://orcid.org/0000-0001-6894-5834>,
<https://orcid.org/0000-0002-6679-8735>

Universidad Nacional Experimental Politécnica “Antonio José de Sucre”
(UNEXPO), Puerto Ordaz, Venezuela

Puerto Ordaz - Venezuela

Recibido (10/06/20), Aceptado (16/06/20)

Resumen: En un entorno de producción los procesos de toma de decisiones son importantes debido a los impactos que estos generan sobre otros procesos. Para tal fin resulta conveniente acceder a la información almacenada en los grandes almacenes de datos a través de un modelo menos complejo, los Datamarts. Un Datamart permite optimizar el proceso de aprovechamiento de la información, a través del agrupamiento de los factores de interés que inciden sobre un hecho o hechos en particular. Así pues, se realizó una investigación del tipo proyectiva y estableciendo como objetivo general el desarrollo de un Datamart basado en el modelo estrella, orientado hacia los modelos de sistemas productivos agropecuarios. La optimización del proceso de extracción y visualización de los datos almacenados en el Datamart, fue llevada a cabo a través de la implementación de un cubo OLAP. Utilizando herramientas de software como SQL Server Management para el diseño de la base de datos, el entorno de desarrollo integrado Visual Studio para la ejecución y diseño de los procesos de extracción, transformación y carga de los datos, y de Power BI como herramienta de Inteligencia de Negocio para la generación de informes y visualizaciones dinámicas de los indicadores establecidos

Palabras Clave: Datamart, Data warehouse, Base de Datos, Inteligencia de Negocio.

DATAMART BASED ON THE STAR MODEL FOR THE IMPLEMENTATION OF KEY PERFORMANCE INDICATORS AS BIG DATA OUTPUT

Abstract: In a production environment, decision-making processes are important because of the impacts they have on other processes. To this end, it is convenient to access the information stored in the large data warehouses through a less complex model, the Datamarts. A Datamart allows optimizing the process of information use, through the grouping of the factors of interest that affect a particular fact or facts. Therefore, a research of the projective type was carried out, establishing as a general objective the development of a Datamart based on the star model, oriented towards the models of agricultural production systems. The optimization of the extraction and visualization process of the data stored in the Datamart was carried out through the implementation of an OLAP cube. Using software tools such as SQL Server Management for the design of the database, the integrated development environment Visual Studio for the execution and design of the extraction, transformation and data loading processes, and Power BI as a Business Intelligence tool for the generation of reports and dynamic visualizations of the established indicators.

Keywords: Datamart, Datawarehouse, Database, Business Intelligence.

I. INTRODUCCIÓN

En el mundo actual, las tecnologías modernas generan un gran volumen de datos sobre todo caracterizados por su naturaleza no estructurada. Como por ejemplo los dispositivos de identificación por radiofrecuencia (RFID), web logs, los vehículos, las búsquedas por internet, las redes sociales, dispositivos GPS, entre otros.

Los sistemas transaccionales utilizados para almacenar los datos e información están estructurados básicamente para capturar los datos y operar; pero no están hechos para informar, debido a que no están diseñados para el análisis de datos complejos.

Desde el punto de vista del análisis o procesamiento de la información, los datos pueden ser moldeados o probados de cualquier manera que una empresa u organización considere adecuada. Con el propósito de identificar las áreas problemáticas, de tal modo que se puedan revertir las tendencias, realizar movimientos de negocio más inteligentes o ejecutar acciones que permitan despejar el camino hacia los objetivos planteados. Es por ello que la implementación de una bóveda de datos (Data warehouse) para la extracción de datos desde fuentes internas (bases de datos operacionales) o externas, transformación, consolidación, integración y chequeo de la integridad de los mismos, que permitan a una empresa o negocio monitorear y acceder a sus datos de las actividades diarias, se convierte en una necesidad vital a satisfacer.

El propósito de este trabajo de investigación es el de desarrollar una versión especial de almacén de datos (Datamart) optimizada para la consulta y análisis de dichos datos, orientado hacia pequeños negocios o empresas que no pueden afrontar los altos costos de implementar un sistema tan amplio de almacenamiento de datos, como lo son los Data warehouse. Haciendo un particular enfoque de aplicación hacia el sector de la agricultura. Para posteriormente poder generar las métricas, KPI (key performance indicator) y gráficos que conformarán un panel de control o Dashboard, el cual será creado a través de herramientas de software de Inteligencia de Negocio.

La Inteligencia de Negocios (BI) es la tecnología que permite extraer, transformar y analizar los datos para generar escenarios, informes y pronósticos que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva.[1]

La investigación desarrollada es de tipo proyectiva, por ello, se lleva a cabo el diseño e implementación de un Data warehouse para el soporte de decisiones en empresas destinadas a la producción agrícola, a partir del cual se podrán centralizar los diferentes datos pertenecientes a la empresa para su posterior análisis a través

de las herramientas destinadas a la Inteligencia de Negocio y que permitirán al ejecutivo visualizar de manera concreta y efectiva dicha información que servirá de apoyo al momento de la toma de decisiones.

Una solución BI involucra la creación de nuevos almacenes de datos (data warehouse), que son alimentados mediante un proceso de ETL (Extraction Transformation and Loading), de distintas fuentes de datos para proporcionar la información oportuna a las aplicaciones BI y a los usuarios que toman decisiones. [1]

La estructura del presente artículo está conformada por el resumen, introducción, marco conceptual, resultados, conclusiones, recomendaciones y referencias. El resumen es una versión breve del trabajo completo. La introducción menciona de manera muy resumida el problema, el alcance, los antecedentes, resultados y conclusiones. El marco conceptual aborda los temas de interés tales como: inteligencia de negocios, Data warehouse, Datamart y la Metodología Hefesto. En los resultados se puede observar el diseño y desarrollo del almacén de datos y la generación de informes; por último, conclusiones, recomendaciones y referencias.

II. MARCO TEÓRICO

A. Big Data

Big Data representa los activos de información caracterizados por un volumen, velocidad y variedad tan altos que requieren tecnología específica y métodos analíticos para su transformación en valor [2].

Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido, debido a que lo que para una empresa determinada puede ser Big Data, para otra puede no serlo, por ello, para considerarlo una solución de esta índole se toma en cuenta que los datos presenten 4 características básicas: Variedad, Velocidad, Volumen y Veracidad. La variedad indica que la data puede estar en diferentes formatos, puede ser estructurada o no estructurada y venir de distintas fuentes, la velocidad indica la rapidez con la que se generan los datos, el volumen indica la cantidad que se genera y la veracidad indica la certeza de los datos [3].

B. Inteligencia de Negocio

La Inteligencia de Negocios es la tecnología que permite extraer, transformar y analizar los datos para generar escenarios, informes y pronósticos que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva. [1].

La Inteligencia de Negocios es un proceso y un producto. El primero compuesto de métodos que las orga-

nizaciones usan para desarrollar información aplicable o inteligencia de negocios, que les permita a las organizaciones salir adelante en un mundo muy competitivo y globalizado. Como producto es información que les permitirá a las organizaciones predecir el comportamiento de competidores, proveedores, clientes, tecnologías, adquisiciones, mercados, productos y servicios y el comportamiento en general del ambiente de negocios, con un cierto grado de precisión [4].

C.Data Warehouse y DataMart

En su definición, Inmon establecía que un Data Warehouse es una colección de datos orientados a temas, integrados, volátiles y variante en el tiempo, organizados para soportar necesidades empresariales [5].

Según la referencia [6] un Datamart se define como una especializada fuente para almacenar datos, siempre orientada a un área específica, volátil y que varía en el tiempo.

D.Cubo OLAP

Un cubo OLAP es un vector multidimensional, On-Line Analytical Processing o procesamiento Analítico en Línea, término acuñado por Edgar Frank Codd, es una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional.[7]

E.Metodología HEFESTO

La metodología HEFESTO, permite la construcción de Data Warehouse o Datamart de forma sencilla, ordenada e intuitiva. Como se puede apreciar a continuación, en el paso 1, se comienza recolectando las necesidades de información o requerimientos de los usuarios. En el paso 2, se obtienen las preguntas claves del negocio y se identifican los indicadores resultantes de los interrogativos y sus respectivas perspectivas o dimensiones de análisis, mediante las cuales se construirá el modelo conceptual de datos del Data Warehouse.

A continuación, se procede a realizar el modelo lógico del Datamart bajo el esquema estrella, copo de nieve o constelación. Se definirá la tabla de hechos la cual es la tabla central de un esquema dimensional, para luego efectuar sus respectivas uniones.

El esquema estrella utilizado en la investigación, está formado por una tabla de hechos principal Hechos, dicha tabla yace en el centro del modelo y está relacionada con sus respectivas tablas de dimensiones, como se muestra en la figura 1.

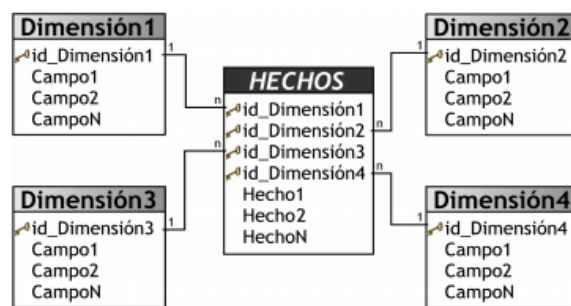


Figura 1. Esquema en estrella.

Fuente: Tomado de [8]

Las tablas de dimensiones representan cada uno de los ejes en un espacio multidimensional. La tabla de hechos constituye el objeto a analizar, poseen atributos de hechos que son del tipo cuantitativo cuyos valores se obtienen por aplicación de alguna función estadística que resumen un conjunto de valores en un único valor.

La arquitectura de un DataMart y/o Data Warehouse se basa en un modelo de datos llamado modelo multidimensional. Los almacenes de datos posibilitan una visión multidimensional de enormes cantidades de datos históricos provenientes de fuentes operacionales, suministrando la información necesaria para el apoyo a los procesos de toma de decisiones de una organización [8].

Por último, se define el proceso ETL (por sus siglas en inglés Extraction, Transformation and Load), el cual permite realizar la migración de datos desde múltiples fuentes, reformatearlos, limpiarlos, y cargarlos en otra base de datos, como es el caso de un DataMart, o Data Warehouse para analizar y apoyar un proceso de negocio.

III.DESARROLLO

Para dar inicio al proceso de diseño del almacén de datos principal, fue necesario la creación de un modelo adicional en representación del sistema transaccional. El modelo de base de datos relacional identificado como Finca, diseñado a partir de los requerimientos seleccionados por el investigador, fue necesario para efectos de realizar las pruebas del diseño e implementación del Datamart. Es conveniente aclarar que el modelo de base de datos relacional, es sólo una parte del modelo a desarrollar para estudiar los datos en una cadena de valor del sector de producción agrícola.

Diseño del Data Warehouse/Datamart.

Paso 1: Análisis de requerimientos.

A partir de un proceso de investigación y búsqueda de información en informes proporcionados por la

FAO se procedió a identificar las variables o aspectos relevantes para dar una representación a un modelo de datos aproximado en el área de producción agrícola:

- Las entidades presentes.
- Las relaciones existentes entre ellas.
- Nivel de relevancia para cada entidad seleccionada.
- Capacidad de brindar respuesta a los indicadores establecidos.

Paso 2: Conformar indicadores.

Identificar indicadores y preguntas de análisis.

Se procedió a la identificación y selección de los Indicadores Claves de Desempeño, correspondientes. La siguiente tabla a continuación muestra la lista de indicadores.

Tabla I. Tabla de indicadores y perspectivas.

Indicadores Claves de Desempeño (KPI)	Perspectivas
<ul style="list-style-type: none"> - Representación porcentual de la cantidad por producto. - Ganancias por producto. - Ganancias obtenidas el mes anterior. - Ganancias obtenidas el trimestre anterior. - Ganancias obtenidas el año anterior. - Tendencia de la cantidad producida. - Tendencia de las ganancias generadas. - Ingresos obtenidos por el tipo de fertilización empleada. - Ingresos brutos por año. - Costos de producción por el tipo de fertilización empleada. - Tasa de crecimiento (cantidad y ganancias). 	Socioeconómico
	Tiempo
	Equipamiento
	Ecológico

Fuente: Elaboración propia.

El modelo conceptual ampliado está conformado por el análisis realizado a las entidades seleccionadas para tal fin como lo muestra la figura 2. Y por las relaciones establecidas entre cada una de ellas, como lo muestra la figura 2.

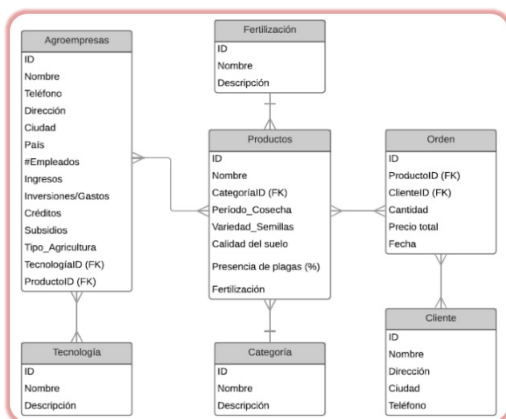


Figura 2. Modelo conceptual ampliado del Data wa-

rehouse.

Fuente: Elaboración propia.

Paso 3: Modelo lógico del Data warehouse/Data-mart.

Modelo conceptual.

Ya elaborado un modelo de datos base para la extracción de un modelo más centralizado y/o específico, se diseña bajo el esquema estrella el Datamart enfocado a albergar las medidas y perspectivas que permitirán el análisis sobre los indicadores.

El modelo se rige bajo dos características: esquema estrella (ver figura 1) y arquitectura (ver figura 3).

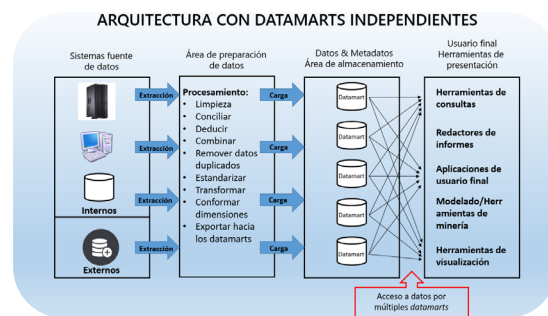


Figura 3. Arquitectura del diseño del Datamart.

Fuente: Elaboración propia.

El modelo lógico de la propuesta se elaboró a través del uso de la herramienta SQL Server Management Studio (SSMS) propiedad de la empresa Microsoft. Dentro de la cual se establecieron las relaciones entre el grupo de medidas o tabla de hechos y las dimensiones asignadas. La figura 4 muestra el modelo diseñado bajo los requerimientos preestablecidos:



Figura 4. Vista del Datamart diseñado.

Fuente: elaboración propia.

Paso 4: Proceso de extracción, transformación y carga (ETL) de la fuente de información al Data Warehouse.

El diseño del proceso ETL para la base de datos transaccional, se llevó a cabo con la utilización del componente Integration Services del entorno de desarrollo Visual Studio. Dentro del cual se estableció como origen, los datos almacenados en hojas de Excel para su posterior extracción en el poblamiento de las tablas. Una visualización del proceso ETL diseñado para cada una de las tablas es mostrada en la figura 5.

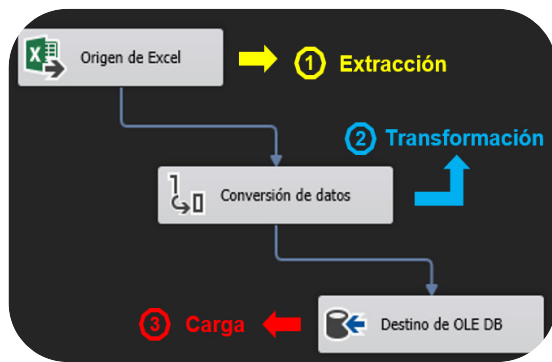


Figura 5. Proceso ETL para el DW.

Fuente: Elaboración propia.

Diseño del cubo OLAP.

El diseño de una última etapa para la optimización del proceso de extracción y visualización de los datos almacenados en el Datamart, fue llevada a cabo a través de la implementación de un cubo OLAP. El cual ha sido diseñado con la utilización del componente Analysis Services integrado de igual manera en el entorno de Visual Studio.

Dicho proceso fue llevado a cabo a través de una serie de pasos intuitivos dentro de la interfaz de desarrollo (ver figura 6) para la creación de la base de datos multidimensional (ver figura 7).

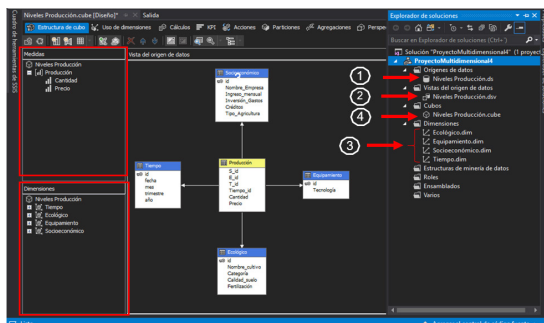


Figura 6. Interfaz de diseño del cubo OLAP en

Analysis Services.

Fuente: Elaboración propia.

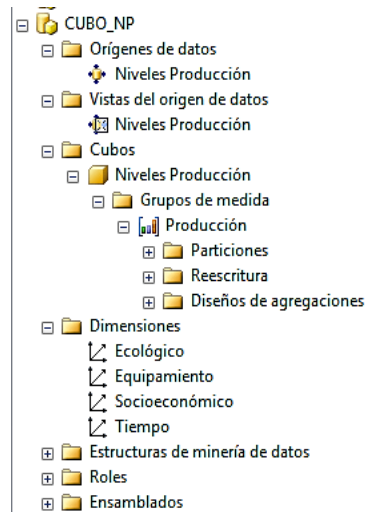


Figura 7. Vista previa del Cubo OLAP en SSMS alojado en el servidor de Analysis Services.

Fuente: Elaboración propia.

IV.RESULTADOS.

Finalizada la etapa de diseño de la BD multidimensional, ya es posible la creación de las visualizaciones correspondientes a los indicadores y reportes confinados en un mismo entorno o cuadro de mando. Para ello, la herramienta de Inteligencia de Negocio, Power BI ha sido la empleada para el diseño de los distintos gráficos e informes relacionados a la información alojada en el almacén de datos optimizado para la extracción y posterior consulta.

La figura 8 presenta una primera entrega de uno de los dos Dashboards diseñados para tal fin:

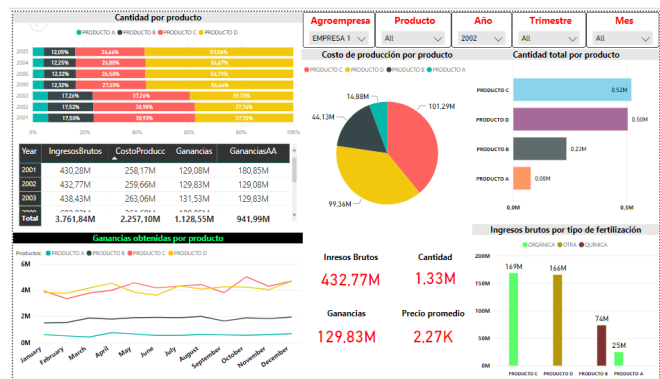


Figura 8. Dashboard 1.

Fuente: Elaboración propia.

A continuación, se presenta al detalle cada uno de los gráficos que componen el Dashboard 1.

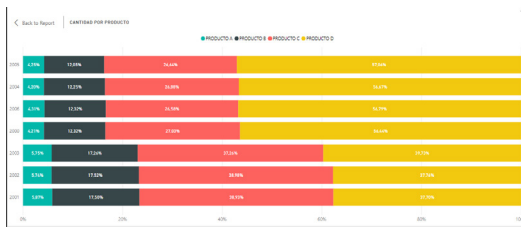


Figura 9. Representación porcentual por producto y año.
Fuente: Elaboración propia.

La figura 9 muestra la representación porcentual que ocupan los distintos productos ofrecidos por una determinada empresa para cada año registrado.

Year	IngresosBrutos	CostoProducc	Ganancias	GananciasAA
2000	602,82M	361,69M	180,85M	
2001	430,28M	258,17M	129,08M	180,85M
2002	432,77M	259,66M	129,83M	129,08M
2003	438,43M	263,06M	131,53M	129,83M
2004	620,98M	372,59M	186,29M	131,53M
2005	614,66M	368,80M	184,40M	186,29M
2006	621,89M	373,13M	186,57M	184,40M
Total	3.761,84M	2.257,10M	1.128,55M	941,99M

Figura 10. Tabla comparativa entre Ingresos, Costos y Ganancias.
Fuente: Elaboración propia.

La tabla comparativa mostrada en la figura 10, permite visualizar datos anuales de algunos de los aspectos financieros más comunes e importantes.

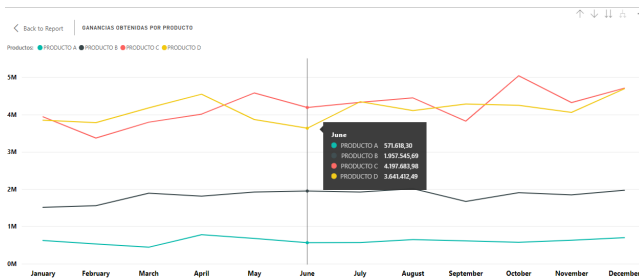


Figura 11. Ganancias mensuales obtenidas por producto.
Fuente: Elaboración propia.

En la figura 11 se visualiza un gráfico de líneas que representan las ganancias obtenidas de forma mensual por cada producto comercializado.

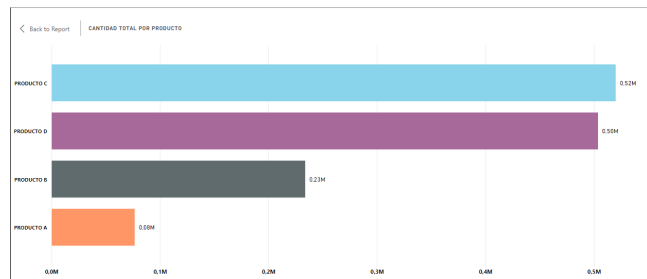


Figura 12. Cantidad total por producto.
Fuente: Elaboración propia.

El gráfico de barras mostrado en la figura 12, representa, de forma jerárquica, la cantidad total producida por producto.

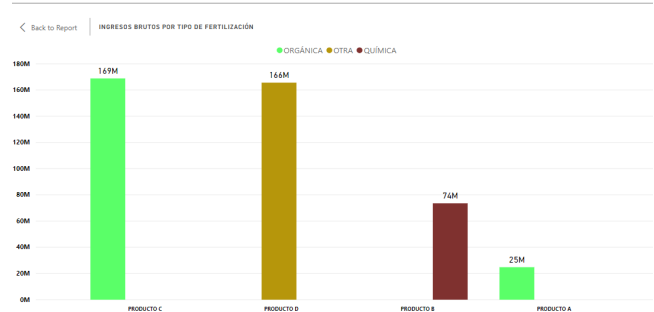


Figura 13. Ingresos brutos por tipo de fertilización.
Fuente: Elaboración propia.

El gráfico mostrado en la figura 13, no es más que una comparación de la variación de los ingresos brutos debido al tipo de fertilización empleada en la producción de los diferentes productos.

Seguidamente, la figura 14 muestra el diseño del segundo Dashboard implementado.

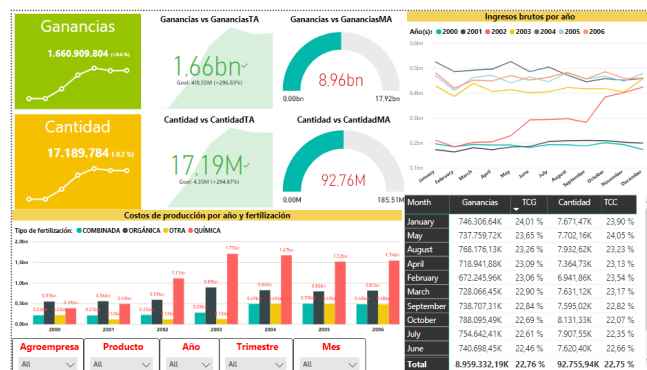


Figura 14. Dashboard 2.
Fuente: elaboración propia.

De una manera más detallada se muestran cada uno de los componentes visuales que conforman el Dashboard 2 a continuación.

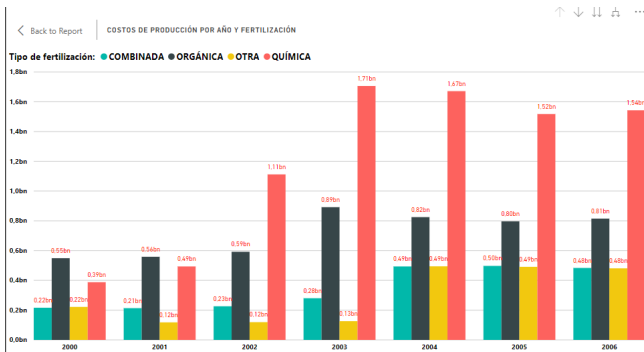


Figura 15. Costos de producción por año y tipo de fertilización.

Fuente: Elaboración propia.

La figura 15 ofrece la información asociada a los costos de producción por año en relación al tipo de fertilización empleada.

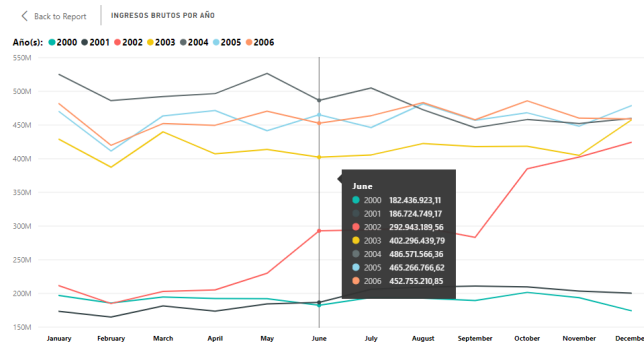


Figura 16. Ingresos brutos por año y mes.

Fuente: Elaboración propia.

En la figura 16 se puede observar una comparativa de los ingresos obtenidos en cada año de producción.

Month	Ganancias	TCG	Cantidad	TCC
January	746.306,64K	24,01 %	7.671,47K	23,90 %
May	737.759,72K	23,65 %	7.702,16K	24,05 %
August	768.176,13K	23,26 %	7.932,62K	23,23 %
April	718.941,88K	23,09 %	7.364,73K	23,13 %
February	672.245,96K	23,06 %	6.941,86K	23,54 %
March	728.066,45K	22,90 %	7.631,12K	23,17 %
September	738.707,31K	22,84 %	7.595,02K	22,82 %
October	788.095,49K	22,69 %	8.131,33K	22,07 %
July	754.642,41K	22,61 %	7.907,55K	22,35 %
June	740.698,45K	22,46 %	7.620,40K	22,66 %
November	769.616,98K	21,86 %	8.014,24K	21,76 %
December	796.074,76K	20,91 %	8.243,44K	20,70 %
Total	8.959.332,19K	22,76 %	92.755,94K	22,75 %

Figura 17. Tabla de tasas de crecimiento porcentua-

les (ganancias y cantidad).

Fuente: Elaboración propia.

La figura 17 ofrece los detalles de las tasas de crecimiento mensual tanto de las ganancias como de la cantidad total de uno o más productos.

V.CONCLUSIONES

Tomando en cuenta los objetivos dados del presente trabajo y los resultados obtenidos se concluye lo siguiente:

Luego de un proceso de investigación, consulta e interpretación de la información, se logró realizar una selección y agrupación de un conjunto de componentes o dimensiones de la entidad de estudio, de las cuales se extrajo atributos relevantes para los requerimientos y objetivo general del presente trabajo.

De los aspectos analizados para el diseño de la base de datos cuya función es la de representar un Data warehouse dentro de un entorno agropecuario, pudo ser posible dar respuesta a una aproximación adecuada para el desarrollo de un almacén de datos transaccional con características enmarcadas bajo las entidades o tablas que lo conforman. Además de brindar una mejor perspectiva del cómo podría ser el modelo de datos o servir de guía para el diseño de un modelo más completo y robusto.

El diseño del Datanart, el cual incluye las dimensiones o perspectivas desde las cuales podrán ser analizadas las medidas, sirve de contenedor para la información de interés y además permite visualizar un extracto de las fuentes de datos empleado.

Los datos generados de forma aleatoria y que figuran como transaccionales para el poblamiento de la base de datos, permiten plasmar el poder de la información cuando esta es seleccionada y explotada para fines de procesos de toma de decisiones en el marco de Big Data. Aun cuando solamente se han considerado datos provenientes desde fuentes transaccionales.

Los indicadores seleccionados como medidas del desempeño de la producción para cada una de las empresas, permiten realizar comparaciones con respecto al tiempo, detectar tendencias, evaluaciones del estado actual, históricos, etc.; los cuales cumplen con las exigencias mínimas permisibles en el marco de los procesos de toma de decisiones y de los esfuerzos realizados por la FAO para el desarrollo de indicadores.

La solución de Inteligencia de Negocio resultó ser adaptable a requerimientos de distinta índole. Pudiendo abarcar un sector lo suficientemente amplio como para dar respuesta a un buen número de situaciones que requieran de la optimización de los procesos de toma de

decisiones, aprovechamiento de la información, generación de reportes e informes dinámicos y del suministro de datos para los sistemas de información gerencial.

Se logro la implementación de estrategias durante el diseño de KPI, obteniendo concordancia con los recursos de información disponibles y las medidas a establecer.

Se utilizaron todas las fuentes de datos relevantes para la organización, de tal manera que la información almacenada en el Datamart contó con un mayor grado de variedad y capacidad de respuesta superior.

REFERENCIAS

- [1].B. Mazon, A. Pan and R. Tinoco. Análisis de Datos Agropecuarios. 1ra. Ed. Ecuador: UTMACH, 2018.
 [2].A. De Mauro, M. Greco & M. Grimaldi. “Una definición formal de Big Data basada en sus características esenciales. Revisión de la biblioteca”, Library Review, vol. 65 no. 3, pp. 122-135, Abril 2016.

[3].P. Muñoz. “Desarrollo de una arquitectura de Big Data para registros mercantiles”. Trabajo de grado, UCV, Caracas, 2016.

[4]. Z. Jourdan, R.K. Rainer & T.E. Marshall. 2Business Intelligence: An Analysis of Literature. Information System Management”, IEEE Engineering Management Review, vol. 25, no 2, p. 121-131, Marzo 2008.

[5].W.H. Inmon. Building the Data Warehouse. 3ra. ed. New York: Wiley, 2002.

[6].C. J. Date. Introducción a los sistemas de bases de datos. 7ma. ed. México: Pearson Educación, 2001.

[7]Cubo OLAP. (2020, mayo 10). Wikipedia. https://es.wikipedia.org/wiki/Cubo_OLAP.

[8]B.R. Dario. Data Warehousing: Investigación y sistematización de conceptos. Hefesto: Metodología propia para la construcción de un Datawarehouse. 1ra. ed. Cordoba: Argentina, 2010.

RESUMEN CURRICULAR



Héctor Zerpa. Ingeniero de Sistemas. MSc. Informática Gerencial. Estudiante del Doctorado en Ciencias de la Ingeniería. Docente – Investigador de la UNEXPO. Tutor de trabajos de pre y postgrado.



Henry Izquierdo. Docente Universitario con título de PhD. Docente de pregrado en Metodología de la Investigación (UNEXPO-UCV). Tutor de trabajos de pre y postgrado



Richard Gracia. Estudiante del último semestre de Ingeniería Electrónica de la UNEXPO. Asistente de investigación en el Centro de Investigación de redes e informática (CIRI).