

ANÁLISIS DE DATOS AGROPECUARIOS

IVÁN RAMÍREZ-MORALES / BERTHA MAZON-OLIVO



Análisis de Datos Agropecuarios

Iván Ramírez-Morales
Bertha Mazon-Olivo

Coordinadores



Primera edición en español, 2018

Este texto ha sido sometido a un proceso de evaluación por pares externos con base en la normativa editorial de la UTMACH

Ediciones UTMACH

Gestión de proyectos editoriales universitarios

302 pag; 22X19cm - (Colección REDES 2017)

Título: Análisis de Datos Agropecuarios. / Iván Ramírez-Morales
/ Bertha Mazon-Olivo (Coordinadores)

ISBN: 978-9942-24-120-7

Publicación digital

Título del libro: Análisis de Datos Agropecuarios.

ISBN: 978-9942-24-120-7

Comentarios y sugerencias: editorial@utmachala.edu.ec

Diseño de portada: MZ Diseño Editorial

Diagramación: MZ Diseño Editorial

Diseño y comunicación digital: Jorge Maza Córdova, Ms.

© Editorial UTMACH, 2018

© Iván Ramírez / Bertha Mazón, por la coordinación

D.R. © UNIVERSIDAD TÉCNICA DE MACHALA, 2018

Km. 5 1/2 Vía Machala Pasaje

www.utmachala.edu.ec

Machala - Ecuador

Advertencia: “Se prohíbe la reproducción, el registro o la transmisión parcial o total de esta obra por cualquier sistema de recuperación de información, sea mecánico, fotoquímico, electrónico, magnético, electro-óptico, por fotocopia o cualquier otro, existente o por existir, sin el permiso previo por escrito del titular de los derechos correspondientes”.



César Quezada Abad, Ph.D
Rector

Amarilis Borja Herrera, Ph.D
Vicerrectora Académica

Jhonny Pérez Rodríguez, Ph.D
Vicerrector Administrativo

COORDINACIÓN EDITORIAL

Tomás Fontaines-Ruiz, Ph.D
Director de investigación

Karina Lozano Zambrano, Ing.
Jefe Editor

Elida Rivero Rodríguez, Ph.D
Roberto Aguirre Fernández, Ph.D
Eduardo Tusa Jumbo, Msc.
Irán Rodríguez Delgado, Ms.
Sandy Soto Armijos, M.Sc.
Raquel Tinóco Egas, Msc.
Gissela León García, Mgs.
Sixto Chilinguina Villacis, Mgs.

Consejo Editorial

Jorge Maza Córdova, Ms.
Fernanda Tusa Jumbo, Ph.D
Karla Ibañez Bustos, Ing.

Comisión de apoyo editorial

Índice

Capítulo I

Ciencia de datos en el sector agropecuario 12
Iván Ramírez-Morales; Bertha Mazon-Olivo ;Alberto Pan

Capítulo II

Obtención de datos en sistemas agropecuarios 45
Salomón Barrezueta Unda; Diego Villaseñor Ortiz

Capítulo III

Internet de las cosas (IoT) 72
Dixys Hernández Rojas; Bertha Mazon-Olivo; Carlos Escudero

Capítulo IV

Matemáticas aplicadas al sector agropecuario 101
Bladimir Serrano; Carlos Loor; Eduardo Tusa

Capítulo V

Estadística básica con datos agropecuarios 127

Irán Rodríguez Delgado; Bill Serrano; Diego Villaseñor Ortiz

Capítulo VI

Estadística predictiva con datos agropecuarios 218

Bill Serrano; Irán Rodríguez Delgado

Capítulo VII

Inteligencia de negocios en el sector agropecuario 246

Bertha Mazon-Olivo; Alberto Pan; Raquel Tinoco-Egas

Capítulo VIII

Inteligencia Artificial aplicada a datos agropecuarios 278

Iván Ramírez-Morales; Eduardo Tusa; Daniel Rivero

Introducción

El análisis de datos es un proceso complejo que trata de encontrar patrones útiles y relaciones entre los datos a fin de obtener información sobre un problema específico y de esta manera tomar decisiones acertadas para su solución.

Las técnicas de análisis de datos que son exploradas en el presente libro son actualmente utilizadas en diversos sectores de la economía. En un inicio, fueron empleadas por las grandes empresas a fin de incrementar sus rendimientos financieros.

El libro se basa en la aplicación de la especialización inteligente, de este modo, gracias al trabajo colaborativo, se combina al sector agropecuario con las tecnologías, matemáticas, estadística y las ciencias computacionales, para la optimización de los procesos productivos.

La idea de descubrir la información oculta en las relaciones entre los datos, incentiva a encontrar aplicaciones para el sector agropecuario, por ejemplo los obtenidos de una producción avícola, o los datos que se generan durante los procesos de fermentación, los parámetros físicos y químicos del suelo, del agua y de las plantas, los datos de sensores, de espectrometría, entre otros.

En la actualidad, este sector se ha mantenido con su producción habitual sin un destacado repunte ni diferenciación, a pesar de existir herramientas científicas que han permitido desarrollar dispositivos tecnológicos y sus aplicaciones.

Este libro ha sido el resultado de la sistematización de las experiencias individuales de un equipo humano con objetivos comunes y una historia académica multidisciplinar, cuyos hallazgos de investigación han sido publicados en revistas científicas y conferencias de alto impacto. El área temática sobre la que se centra este texto es en técnicas de extracción, procesamiento y análisis de datos del ámbito agropecuario, se combinan para entregar al lector una obra de calidad y alto valor científico.

Así, el presente libro está concebido desde diferentes puntos de vista de profesionales agrónomos, informáticos, electrónicos, matemáticos, estadísticos y empresarios. Todos buscan un objetivo en común: “descubrir el conocimiento oculto en los datos que proporcione una ventaja competitiva”. Se aborda el ciclo completo del proceso de obtención de conocimiento a partir de datos crudos del sector agropecuario, con la finalidad de apoyar la toma de decisiones. Este ciclo involucra procesos de: selección de los datos (extracción, comunicación, almacenamiento), pre-procesamiento, transformación, aplicación de modelos y/o técnicas de análisis, presentación e interpretación de resultados. El enfoque temático del libro es el siguiente:

Capítulo 1: Ciencia de Datos en el sector Agropecuario.- En este capítulo se aborda una revisión desde los inicios del análisis de datos en el sector agropecuario hasta el progreso actual que se ha dado en esta área del conocimiento que se considera como la nueva revolución en la agricultura y la ganadería de precisión.

Capítulo 2: Obtención de datos en sistemas agropecuarios.- El enfoque del capítulo es la generación de datos crudos en los sistemas agropecuarios, aplicando métodos y técnicas básicas donde se registran información de: número de unidades producidas, cantidad de nutrientes, variables climáticas, muestreo y monitoreo de organismos vivos, entre otros.

Capítulo 3: Internet de las cosas (IoT).- Este capítulo aborda los sistemas de telemetría para obtención de datos y control de dispositivos, aplicando tecnologías como: redes de sensores inalámbricos (dispositivos electrónicos, sensores, actuadores y puertas de enlace), protocolos de comunicación, centros de procesamiento de datos (cloud computing) y aplicaciones IoT para el sector agropecuario.

Capítulo 4: Matemáticas aplicadas al sector agropecuario.- Este capítulo explica los procedimientos para la creación de modelos matemáticos determinísticos que representen procesos asociados al sector agropecuario, como una alternativa de solución en la ingeniería.

Capítulo 5: Estadística básica con datos agropecuarios.- El capítulo se enfoca en los atributos, escalas de medición de las variables, su influencia en la elección del procedimiento estadístico a desarrollar, así como, el papel de las medidas de resumen, estimación puntual y prueba de hipótesis en la investigación científica.

Capítulo 6: Estadística predictiva con datos agropecuarios.- El capítulo considera las principales técnicas de la estadística avanzada aplicada al sector agropecuario, con el propósito de establecer predicciones que permita tomar mejores decisiones.

Capítulo 7: Inteligencia de negocios en el sector agropecuario.- El capítulo comprende la obtención de conocimiento a partir de datos crudos con la finalidad de apoyar la toma de decisiones en empresas del sector agropecuario. Involucra procesos de extracción, transformación y almacenamiento de datos en nuevos almacenes (Data warehouse - Big Data), distribución y análisis de la información con técnicas: multi-dimensional OLAP y tableros de control (dashboards).

Capítulo 8: Inteligencia Artificial aplicada a datos agropecuarios.- El capítulo trata sobre las principales técnicas de machine learning aplicadas a los datos agropecuarios, entre éstas se destacan: las redes de neuronas artificiales, máquinas de soporte de vectores, vecinos más cercanos, análisis de componentes principales, entre otros.

07 Capítulo Inteligencia de negocios en el sector agropecuario

Bertha Mazon-Olivo; Alberto Pan; Raquel Tinoco-Egas

La Inteligencia de Negocios (BI) se encarga de obtener el conocimiento a partir del procesamiento de datos crudos, con la finalidad de apoyar la toma de decisiones en una organización en sus niveles tácticos y estratégicos. Los temas que se abordan son: fundamentos de sistemas de información e inteligencia de negocios, el diseño e implementación de almacenes de datos (Data warehouse y Data mart), el proceso de extracción, transformación y carga de datos (ETL) y las técnicas de visualización de datos OLAP y tableros de control (dashboards); y, se concluye con el análisis de un caso de estudio y su desarrollo práctico de una solución BI.

Bertha Mazon-Olivo: Ingeniera en Sistemas y Magíster en Informática Aplicada. Profesora Titular en la Universidad Técnica de Machala. Líneas de investigación: Internet de las Cosas, integración, procesamiento y análisis de datos. Es estudiante del programa doctoral en Tecnologías de la Información y las Comunicaciones en Universidad de Coruña, España. Sus líneas de investigación son: Internet de las cosas, Ciencia de datos y Desarrollo de Aplicaciones Informáticas. Cuenta con varias publicaciones indexadas.

Alberto Pan: Es Director Técnico de Denodo y Profesor Asociado de la Universidad de A Coruña. Reibió una Licenciatura en Ciencias de la Computación en la Universidad de A Coruña en 1996 y un Ph.D. en Informática por la misma universidad en 2002. Sus intereses de investigación están relacionados con la extracción e integración de datos y la automatización de la web. Alberto ha dirigido varios proyectos a nivel nacional y regional en el campo de la integración de datos y acceso a la Web oculta. También es autor y coautor de numerosas publicaciones en revistas científicas y actas de congresos.

Raquel Tinoco-Egas: Ingeniera en Gestión Empresarial Internacional. Máster en Desarrollo de Negocios Internacionales por la Universidad de Neuchâtel de Suiza. Estudiante doctoral en Análisis Económico y Estrategia Empresarial en la Universidad de Coruña. Profesora Titular en la Universidad Técnica de Machala. Investiga temas como la innovación y tecnología para el desarrollo empresarial. Cuenta con varias publicaciones.

Necesidades de información en una empresa agropecuaria

Una organización, dependiendo de su tamaño, objeto de actuación o negocio, genera cientos, miles o quizá millones de transacciones diarias que se traducen en datos; el sector agropecuario no es la excepción; el seguimiento de procesos productivos agrícolas o ganaderos, la gestión de recursos: financieros, humanos, materias primas, maquinaria, etc., generan muchos datos; si estos datos no son procesados, terminarán por ser olvidados y desaprovechados.

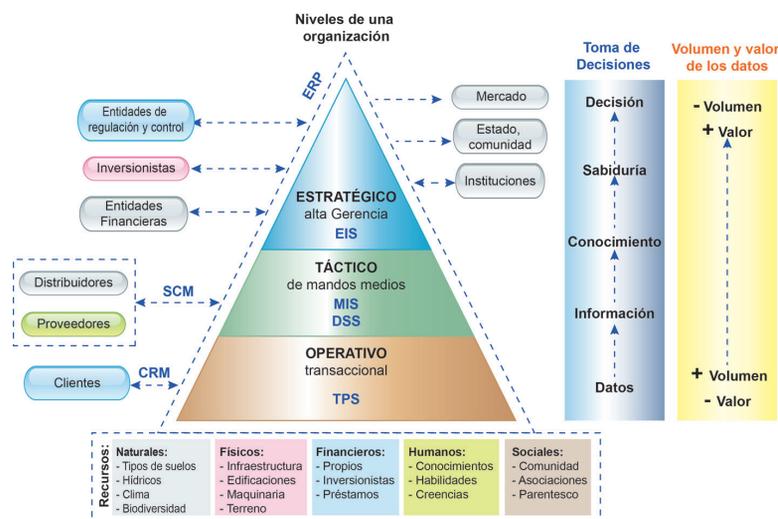
En la actualidad, con la Ciencia de Datos, es posible obtener el máximo provecho de la materia prima denominada “datos crudos” o “datos en bruto”; mediante el procesamiento de estos datos se genera información útil para la toma de decisiones en beneficio de la propia organización. Las disciplinas relacionadas con esta ciencia son: la Estadística, Inteligencia de Negocios, Minería de Datos, Inteligencia Artificial, Aprendizaje Automático, Bases de Datos y otras.

Para comprender la necesidad de la información en una organización del sector agropecuario, es necesario revisar el siguiente concepto:

Explotación Agropecuaria, comprende la producción agrícola y/o ganadera, que puede ser llevada a cabo mediante una gestión única, por una persona, un hogar, grupos familiares, asociaciones, cooperativas o empresas. La Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, 2016), la define como: “... *la unidad económica de producción agropecuaria bajo gestión única, que comprende todo el ganado mantenido en ella y toda la tierra dedicada total o parcialmente a fines agrícolas, independientemente del título, forma jurídica o tamaño. ...Las tierras de la explotación agropecuaria pueden constar de una o más parcelas, situadas en una o más áreas separadas en una o más divisiones territoriales o administrativas, siempre que todas las parcelas compartan los mismos recursos*”.

El proceso de toma de decisiones en una organización de producción agrícola (unidad de explotación agropecuaria) no se diferencia de otras empresas o instituciones. La Imagen 7.1, está basada en el libro de Laudon & Laudon (2012), ilustra los niveles de una organización, los tipos de sistemas de información, el proceso de conversión de datos en una decisión y finalmente la valoración de los datos según los criterio de volumen y valor. A continuación se describen en detalle cada componente del gráfico.

Imagen 7.1. Relación entre los sistemas de información y los niveles de una organización



Fuente: Elaboración propia

Niveles de una organización

En una organización, los tipos de sistemas de información se distribuyen en tres niveles:

- Nivel Operativo. En este nivel se ubican los sistemas de información que gestionan las transacciones diarias de una organización y son los generadores de “*datos crudos*” o en bruto. Las tareas, recursos y metas están predefinidos y bien estructurados. Los Sistemas de Procesamiento Transaccional (TPS) se encargan del control de

recursos y procesos, por ejemplo: producción, compras, ventas, inventarios, etc.

- Nivel Táctico. En el nivel táctico se ubican sistemas de información de mandos medios como son los Sistemas de Soporte de Decisiones (DSS) y los Sistema de Información Gerencial (MIS). Los usuarios de este nivel planifican, dirigen y controlan las acciones del nivel operativo, toman decisiones a mediano plazo con afectación según su área o departamento; generan reportes con “información” consolidada a partir de los datos crudos de acuerdo a su ámbito de acción, para mantener informados a los ejecutivos de mandos estratégicos.
- Nivel Estratégico. En este nivel se encuentra los sistemas de Información para Ejecutivos (EIS) que apoyan a la alta gerencia en la toma de decisiones a largo plazo y que afectan a toda la organización. La junta de accionistas, el gerente general o propietario, aprovechan la información de los niveles inferiores para su “conocimiento” y en base a su experiencia o “sabiduría”, son capaces de tomar la “decisión” más acertada.

Sistema de información (SI)

- Comprende un conjunto de personas, procedimientos, datos y tecnologías que apoyan las actividades de una organización. Según Laudon & Laudon (2012), lo definen como: “...conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar los procesos de toma de decisiones y de control en una organización”. Además, “...ayudan a los gerentes y trabajadores del conocimiento a analizar problemas, visualizar temas complejos y crear nuevos productos”.

Tipos de sistemas de información

Transaction Processing System (TPS). El sistema de Procesamiento Transaccional se encargan de recolectar, almacenar, procesar, calcular, ordenar, modificar y recuperar los datos

obtenidos de las transacciones diarias de una empresa.

Decision Support System (DSS). El sistema de Soporte de Decisiones, genera información a partir de los datos que provee el TPS y apoya las decisiones semi-estructuradas de los usuarios de mandos medios y estratégicos. Los DSS presentan información estadística, dinámica, multidimensional y consolidada acorde a los Indicadores Clave de Rendimiento (KPI's) de cada área de la organización, que regularmente se estructura en cubos OLAP (On-Line Analytical Processing).

Management Information System (MIS). El sistema de Información Gerencial comprende una colección de sistemas de información, que relacionados entre sí, apoyan las actividades de una organización en todos sus niveles. Las funciones principales de este sistema son la planeación, organización, dirección y control de las actividades del nivel operativo, apoyo a las decisiones de nivel táctico y a su vez, informar los avances y resultados de actividades, al nivel estratégico.

Executive Information System (EIS). El sistema de Información Ejecutiva, está orientado a los usuarios de alta gerencia, tiene la función de presentar de forma sencilla, el estado actual de los KPI's relevantes de la organización, basándose en otros sistemas como los DSS, cubos OLAP y varias fuentes de datos empresariales. La interfaz típica de estos sistemas son los cuadros de mando o tableros de control (dashboards) que presentan información estadística dinámica descriptiva, diagnóstica o predictiva.

Enterprise Resource Planning (ERP). El sistema de Planificación de Recursos Empresariales, es un sistema integral, que gestionan recursos de toda la empresa como por ejemplo: recursos (naturales) para la producción, inventario de bienes y productos (físicos), logística, distribución, compras a proveedores y ventas a clientes (mercados), contable-financiero, recursos humanos y sociales. Los ERP se relacionan con otros sistemas como los CRM, SCM y MIS, DSS y EIS; estos sistemas pueden ser desarrollados a medida como sistemas independientes o como módulos de un ERP.

Customer Relationship Management (CRM). El sistema

de Gestión de la Relación con los Consumidores, se enfoca en servicios y estrategias de marketing para lograr la fidelización de los clientes que ya son parte de la empresa, así como para captar nuevos clientes. Permite llevar un control de reuniones, el registro del historial de acuerdos en procesos de negociación, contratos, convenios, etc. Por ejemplo las empresas inmobiliarias manejan este tipo de sistemas.

Supply Chain Management (SCM). El sistema de Gestión de la Cadena de Suministros, involucra a toda la logística que va desde el contacto con proveedores que suministran la materia prima necesaria para la producción de nuevos artículos; y, la organización de la cadena de distribuidores y vendedores para llegar al cliente o consumidor final.

Existen otros tipos de sistemas de información que contribuyen en las actividades de una organización, por ejemplo: los sistemas de ofimática, de mensajería o comunicación, de gestión documental, sistemas de comercio electrónico, de trabajo colaborativo o en grupo, etc.

Etapas para la obtención del conocimiento y toma de decisión

Como se observa en parte derecha de la Imagen 7.1, los datos que se obtienen de los sistemas de procesamiento transaccional, son procesados para generar información, la misma que al ser presentada de manera oportuna a la persona adecuada y con la experiencia necesaria, ésta es capaz de analizarla, convertirla en conocimiento y actuar con sabiduría tomando la decisión más acertada en beneficio de su organización.

Volumen y valor de los datos

En el nivel más bajo de la pirámide de la Imagen 7.1, los datos se obtienen en grandes cantidades de las transacciones diarias de la organización; éstos son considerados de poco valor porque no están procesados o no tiene una estructura enfocada en los Indicadores Clave de Rendimiento del negocio (KPI's). Los mismos datos al pasar por los niveles

táctico y estratégico se van procesando, resumiendo y consolidando de tal forma que su volumen disminuye pero su valor aumenta. Por ejemplo: supóngase que en promedio se generan 100 facturas de venta diariamente; al año serían un estimado de 36 000 facturas. Al gerente no le interesa las copias de todas las facturas, sino un reporte consolidado de ventas según el criterio específico; por ejemplo, puede ser el total de ventas por mes, trimestre, semestre o año; dicho reporte no pasará de una o dos páginas.

Los objetivos y usuarios de la información

En la Imagen 7.2, se aprecia los usuarios de la información categorizados por niveles:

Usuarios operativos. En el nivel operativo se ubican algunos trabajadores administrativos, de servicio y producción, por ejemplo: bodeguero, vendedores, auxiliares contables, operarios, etc. Este tipo de usuarios tienen como objetivos: realizar o ejecutar acciones encomendadas por su jefe inmediato superior, planificar acciones a corto plazo (diarias, semanales o mensuales) e informar periódicamente el resultado de sus actividades a su jefe.

Usuarios de mandos medios. En el nivel táctico se encuentran los directivos, gerentes o jefes de sucursales o departamentos; por ejemplo, administrador de finca o hacienda, el director financiero, el director de producción, el gerente de zona, etc. Los usuarios de este nivel se encargan de: realizar planes operativos a mediano plazo (máximo 1 año), organizar y coordinar actividades, recursos y el personal; dirigir, controlar y supervisar al personal y los resultados que generan; y finalmente, tomar decisiones por área o departamento.

Usuarios de alta gerencia. En el nivel estratégico se ubican la junta de accionistas, gerente general o propietario. Los objetivos a este nivel son: tomar decisiones que afectan a toda la organización, preparar planes (estratégicos) a largo plazo (3 o más años), dirigir, controlar y supervisar la implementación de estrategias enfocadas en incrementar la productividad, con una mejora de la calidad de sus productos o

Imagen 7.2. Objetivos y usuarios de la información en una organización

	Usuarios de la información	Objetivos de los usuarios respecto al manejo de la información
ESTRATÉGICO alta dirección	Usuarios de alta gerencia: - Junta de accionistas/socios - Gerente o director general - Propietario	Con alcance global: - Tomar decisiones que afectan a toda la organización - Planear a largo plazo (plan estratégico a 3 años o más) - Dirigir, controlar y supervisar la implementación de estrategias enfocadas a incrementar la productividad, mejorando la calidad con el menor costo posible
TÁCTICO de mandos medios	Usuarios de mandos medios. Director, jefe, gerente de sucursal o de departamento: - producción, - contabilidad y finanzas, - recursos humanos, - comercial y mercadeo,	Por área, departamento o zona: - Tomar decisiones con afectación en su área - Elaborar planes operativos (máximo 1 año) - Organizar y coordinar actividades, recursos y el personal - Dirigir, controlar y supervisar al personal y los resultados obtenidos (productos o servicios)
OPERATIVO transaccional	Usuarios operativos: algunos trabajadores administrativos, de servicio y producción: - bodeguero, vendedores - auxiliar contable, - operarios de maquinaria, etc.	Según el cargo: - Realizar o ejecutar acciones encomendadas por su jefe inmediato superior - Planificar actividades a corto plazo (diarias, semanal o mensual) - Informar resultados de su actividad o función

Fuente: Elaboración propia

Los problemas con los datos en una organización

Obtener un valor agregado de los datos para apoyar la toma de decisiones, es una tarea que requiere de conocimiento en el manejo de técnicas de Ciencia de Datos y herramientas tecnológicas. Algunas empresas desaprovechan sus datos y no analizan sus patrones de comportamiento en beneficio propio, debido al poco o inexistente conocimiento en el manejo de herramientas tecnológicas. A continuación se describen algunos problemas que tienen las empresas con sus datos:

- La falta de automatización e integración de sus departamentos y procesos hace que se ocupe el tiempo en tareas triviales de gestión manual de documentos.
- El sistema de procesamiento transaccional de la empresa puede ser limitado o tener algunos defectos que provoca la desconfianza de los usuarios en los informes o resultados de búsquedas.
- La alta gerencia y mandos medios pueden dar poca importancia a la información de un sistema informático, debido al inadecuado tratamiento y presentación.

- Los datos de las transacciones diarias sirven para realizar el proceso del momento y pronto éstos son olvidados en archivos impresos o en bases de datos electrónicas.
- Falta de conocimiento de la tecnología adecuada para el tratamiento de la información debido al presupuesto de la empresa.
- Los datos electrónicos pueden estar expuestos a perderse o ser robados por falta de seguridad.
- Falta de políticas de gobierno de datos.

Fundamentos de inteligencia de negocios

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. La información adecuada en el lugar y momento oportuno incrementa la efectividad de cualquier empresa y las del sector agropecuario no son la excepción, los datos crudos se generan de distintas áreas como: producción, mercadeo, ventas, finanzas, recurso humano, etc. 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 estructurados y no estructurados (por ejemplo: bases de datos relacionales, hojas de cálculo, archivos planos, etc.), para proporcionar la información oportuna a las aplicaciones BI (análisis multidimensional OLAP, consultas y reportes, tableros de control, minería de datos y otras aplicaciones personalizadas) y a los usuarios que toman decisiones.

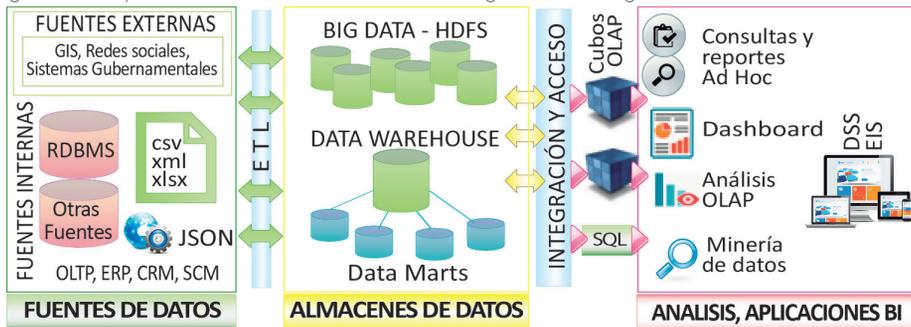
Según Mazon-Olivo et al. (2017), BI comprende un conjunto de estrategias y componentes que permiten transformar los datos operacionales en información y ésta en conocimiento útil para la toma de decisiones; es decir, facilita el monitoreo del cumplimiento de los objetivos organizacionales y admite el análisis de la información histórica, contribuyendo

a la creación de estrategias comerciales que generan ventajas competitivas en el mercado. BI tiene múltiples campos de aplicación, entre los más comunes están los sectores: comercial, empresarial, industrial, educativo, banca, turismo, y otros que requieren del análisis de sus datos para identificar tendencias o patrones que a su vez orientan la toma de decisiones.

Arquitectura de una solución de inteligencia de negocios

Para el desarrollo de una solución de inteligencia de negocios es necesario comprender su arquitectura, en la Imagen 7.3 es posible apreciar 3 capas: la capa de fuentes de datos, la capa de almacenes de datos y la capa de análisis y aplicaciones BI. Entre la primera y segunda capas se requiere el proceso de extracción transformación y carga (ETL) que suministre los datos crudos o en bruto (raw data) al almacén de datos (data warehouse). Para el análisis de datos se requieren aplicaciones BI que accedan al data warehouse mediante cubos OLAP y procesos de integración de datos.

Imagen 7.3. Arquitectura de una solución de Inteligencia de Negocios



Fuente: (Mazon-Olivo et al., 2017)

La capa de datos fuente

Los datos en bruto existen en gran cantidad y diversidad de formatos y pueden clasificarse según:

El origen:

- Internos. Son de la propia empresa, pueden presentarse en bases de datos, archivos o reportes de sistemas informáticos (ERP, CRM, SCM)
- Externos. Se obtienen de otras organizaciones (redes sociales, sistemas de información geográfica, sistemas gubernamentales) mediante web services, descarga de sitios web, por email, reportes de sistemas informáticos.

El formato:

- Estructurados. Se obtienen de bases de datos relacionales, otras data warehouse
- Semi-estructurados. Por lo general son archivos en formatos: CSV, JSON, XML, HTML, etc.
- No estructurados. datos de archivos como PDF, imagen, sonido, video, etc.

El tamaño:

- Volúmenes de datos normales. Cientos o miles de registros medidos en KB o MB.
- Grandes volúmenes de datos (big data). Millones de registros medidos en GB, TB o PB (Peta Byte), generalmente se encuentran en sistemas clusterizados con mecanismos de procesamiento y almacenamiento distribuido.

La capa de almacenes de datos

Comprende los grandes repositorios de metadatos, en este ámbito se encuentran los Data Warehouse y los Big Data.

Data Warehouse (DW). Es considerado un almacén o bodega de datos estructurados que contiene información temática, histórica e integrada según los indicadores clave de desempeño que se hayan previsto en una organización (Rosado G. & Rico B., 2010). Un DW es temático debido a que está conformado por áreas o grupo de datos de una organización. Cada tema del DW es representado por un Data mart.

Big Data. Representa un conjunto de recursos de información de gran volumen, que se obtienen a altas velocidades y con una variedad de formatos, que demandan un almacenamiento escalable y eficiente, formas innovadoras de procesamiento de información para mejorar el análisis, la comprensión y toma de decisiones (NIST, 2015). Una Big Data es una oportunidad para que las organizaciones obtengan ventajas competitivas en el mundo actual digitalizado y globalizado (De Mauro, Greco, & Grimaldim, 2015). Un término asociado a Big Data es Data Lake, que comprende grandes conjuntos de datos (big data sets), también llamados lagos de datos, que se etiquetan para realizar consultas o buscar patrones. Los tipos de Big Data en función de sus prestaciones pueden ser: 1) de alto rendimiento y 2) distribuidos, por lo general de bajo rendimiento. Ejemplos de sistemas de alto rendimiento son: Teradata, HP Vertica, IBM Netezza, Oracle Exadata (Moniruzzaman & Hossain, 2013; Țăranu, 2015). Un ejemplo de sistema distribuido es Hadoop¹ (HDFS es Hadoop Distributed File System), considerado como la plataforma más utilizada para el procesamiento y almacenamiento distribuido de datos (Sawant & Shah, 2013; Țăranu, 2015; White, 2015); ejemplos de versiones de plataformas que integran Hadoop son Hortonworks², Cloudera³, AWS⁴, Microsoft Azure⁵, etc.

Las V's de la Big Data según (NIST, 2015):

- Volumen. Característica enfocada al tamaño del conjunto de datos (data set). Para esto se requiere tecnologías más eficientes enfocadas en la recolección, almacenamiento y procesamiento de datos. Ejemplos de generadores de grandes volúmenes de datos son: aplicaciones de la Web 2.0 como las redes sociales, el Internet

¹ Apache Hadoop: <http://hadoop.apache.org/>

² Hortonworks: <https://es.hortonworks.com/>

³ Cloudera: <https://www.cloudera.com/>

⁴ AWS Amazon Web Services: <https://aws.amazon.com/es/>

⁵ Microsoft Azure: <https://azure.microsoft.com/es-es/>

de las Cosas (Internet of Things), los objetos inteligentes (Smart objects), entre otros.

- Velocidad. Se refiere a la tasa de flujo de datos. La rapidez con la que se producen los datos se debe en gran medida a la concurrencia de generadores de datos (personas y objetos), muchos de esos datos se transmiten en tiempo real.
- Variedad. Se refiere a la diversidad de formatos de datos provenientes de múltiples repositorios y dominios; por ejemplo: datos estructurados (Bases de datos relacionales, Data Warehouse), semi-estructurados (archivos de datos en formatos JSON, XML, HTML, xls etc.) y no estructurados (archivos de imágenes, videos, audios, etc.).
- Variabilidad. Representa el cambio de características de los datos. Los requisitos de seguridad y privacidad pueden cambiar dependiente de la naturaleza y el tiempo que toma llevar a cabo funciones como: recogida, procesamiento, agregación y almacenamiento de datos. La gobernabilidad también puede cambiar a medida que las organizaciones responsables se fusionan o incluso desaparecen.
- Veracidad. Los datos recopilados deben ser confiables, es decir, se refiere a que se debe hacer un control de calidad de los datos antes de realizar alguna operación con ellos.
- Otras *v's* o características también atribuidas a big data son: valor, viabilidad y visualización.

Data Mart. Comprende un subconjunto de datos enfocados en el análisis de un tema, área o ámbito específico en una organización (Mosquera & Hallo, 2014). El conjunto de data marts comprende un data warehouse. Algunos ejemplos de data mart en una empresa agropecuaria aplicados al proceso productivo puede ser: *prod_parcelas*, *prod_cultivos*, *prod_siembra*, *prod_cosecha*, *prod_ingresos*, *prod_costos*, *inventario*, *recursos-humanos* y otros.

Componentes de un Data Mart:

- Tabla de hechos (table fact). Es la tabla central del data mart, que contiene los datos o medidas (indicadores claves de desempeño del negocio KPI's) que se utilizan para análisis y las claves de las tablas de dimensiones. Es común nombrar las tablas de hechos en función del área/tema que representa el data mart; por ejemplo, si el data mart representa los *cultivos agrícolas* que se han producido o que están en proceso de producción, la tabla de hechos puede ser: *th_produccion_cultivos*.
- Medida o KPI (Key Performance Indicator). Representa un valor numérico que contribuye al análisis de un hecho. Por ejemplo: *densidad de siembra (número de plantas por hectárea)*, *edad del cultivo (días o años)*, *rendimiento de un cultivo (toneladas métricas por hectárea TM/ha)*, *costos de producción por hectárea*, *rentabilidad de un cultivo (ingresos – egresos)*, etc.
- Tablas de Dimensiones: Corresponden a las perspectivas o vistas a través de las cuales es posible analizar las medidas del negocio o KPI's. Por ejemplo: para la medida *densidad de siembra (número de plantas por hectárea)*, las posibles dimensiones pueden ser: *especie de cultivo (banano, cacao, café, etc.)*, *clase de cultivo (frutales, hortalizas, cereales, etc.)*, *tipo de cultivo según duración (ciclo corto, perennes)*, *tiempo de la siembra (año, semestre, trimestre y mes)*, tipo de suelo, etc.

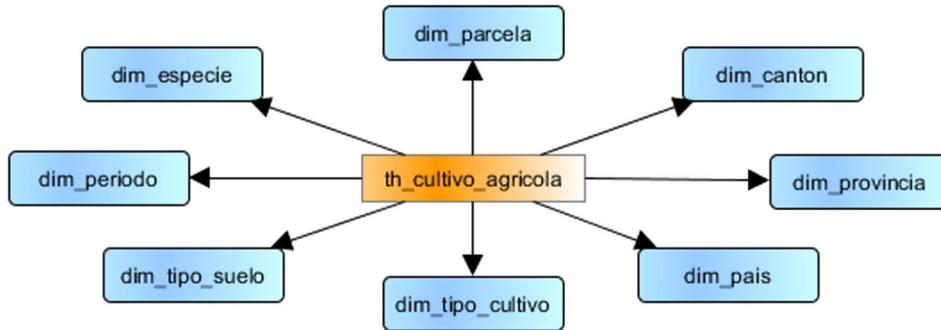
Tipos de esquemas de diseño de un data marts y data warehouse.

Existen 2 tipos de esquemas de diseños para crear modelos de data marts, estos son: estrella y copo de nieve; la integración de varios esquemas estrella o copo de nieve dan lugar a un esquema en forma de constelación que representa el data warehouse. A continuación se explican con mayor detalle:

Esquema en Estrella. Contiene una sola tabla de hechos con los datos de análisis y las claves de todas las tablas de

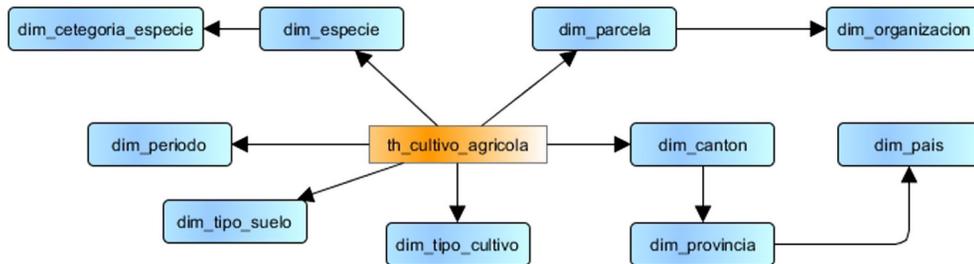
dimensiones. No existen relaciones entre las dimensiones. Ver Imagen 7.4.

Imagen 7.4. Esquema de un data mart en estrella



Esquema en Copo de Nieve. Similar al esquema en estrella con la diferencia de que las dimensiones pueden relacionarse creando jerarquías. Ver Imagen 7.5.

Imagen 7.5. Esquema de un data mart en copo de nieve



Esquema en Constelación. Integración de varios esquemas estrella o copos de nieve que representan un data warehouse. Su diseño puede llegar a ser complejo debido a la cantidad de relaciones entre las tablas de hechos y dimensiones, más si las dimensiones se comparten entre varios data marts. Es aconsejable manejar por separado el diseño de los data marts; sin embargo, en la implementación física del data warehouse es necesario optimizar y no repetir dimensiones.

El proceso de Extracción Transformación y Carga (ETL)

ETL de sus siglas en inglés Extraction, Transformation and Load, es el proceso que permite seleccionar datos desde múltiples fuentes, utilizando una herramienta de integración se pre-procesan y se cargan en un nuevo almacén de datos (data warehouse).

Las actividades ETL según Cornejo et al. (2014), se describen a continuación:

- **Extracción.** Consiste en la identificación de los datos fuentes, verificación de su calidad, lectura de datos crudos, obtención de agregados y establecimiento de la estructura de la metadata del data warehouse.
- **Transformación.** Comprende la aplicación de una serie de reglas de negocio sobre los datos extraídos para convertirlos en datos con el formato del data warehouse. Algunas de estas actividades son: limpieza, cambio de formato, generación de datos calculados, creación de nuevos datos o claves, filtrado, ordenación, asociaciones y agregaciones.
- **Carga.** Las actividades que se realizan son: integración de datos, pruebas de carga, carga (escritura) de datos en el data warehouse, gestión de errores y mantenimiento de metadata. La carga de datos normalmente es mediante procesos batch, sea por lotes, por registros, por totales, u otra forma.

Integración y virtualización de datos

La virtualización de datos consiste en el acceso a datos procedentes de distintos repositorios, ocultando la complejidad interna a las aplicaciones consumidoras. Las herramientas de integración y virtualización de datos, realizan dicha función, creando un único punto de acceso a los datos mediante una base datos lógica; y de esta manera facilitando el gobierno

de los datos (Van Der Lans, 2012). Ejemplos de este tipo de herramientas son: Denodo⁶, Informatica⁷, Cisco Data Virtualization⁸, etc.

Cubos OLAP. OLAP significa procesamiento analítico en línea. Un cubo OLAP representa el esquema o definición de la estructura multidimensional para el análisis de datos empresariales que se encuentra agregados y organizados en un data warehouse. Un cubo se construye definiendo una tabla de hechos, las medidas y las dimensiones. Las consultas a un cubo OLAP se realizan mediante el lenguaje MDX (MultiDimensional eXpressions o expresiones multidimensionales). Si un data warehouse se implementa en un Sistema Gestor de Base de Datos (DBMS) relacional, se considera un sistema ROLAP, si se lo implementa en un sistema de almacenamiento multidimensional es MOLAP y si se implementa en los dos tipos entonces es HOLAP (Morales, Cuevas, & Martínez, 2016; Rosado & Rico, 2010).

La capa analítica o de aplicaciones de inteligencia de negocios

En esta capa se ubican las aplicaciones y herramientas para el análisis descriptivo y predictivo de los datos, las más destacadas son: visores OLAP, tableros de control (dashboards), reportes y consultas Ad Hoc, Minería de datos y otras. Este tipo de aplicaciones se clasifican en dos tipos de sistemas que ya fueron mencionados en las secciones anteriores: los sistemas de soporte de decisiones (DSS) y los sistemas de información para ejecutivos (EIS). (Ghosh, Halder, & Sen, 2015; Gounder, Iyer, Professor-ccis, Mazyad, & Prof, 2016; Marinho & Bernardino, 2015; Vassell, Apperson, Calyam, Gillis, & Ahmad, 2016), como se puede observar en la Imagen 7.6. A continuación se explican con más detalle:

¹ <https://www.denodo.com/en>

² <https://www.informatica.com>

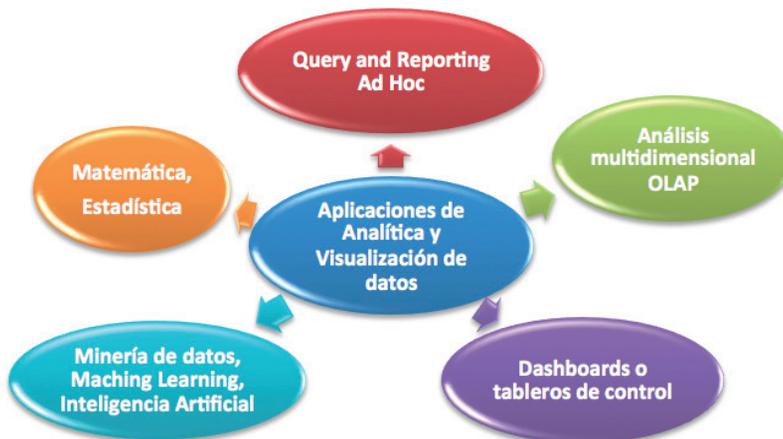
³ <http://www.compositesw.com/products-services/data-virtualization-platform/>

Consultas o Reportes Ad Hoc. Consiste en el diseño de consultas o reportes personalizados para resolver un problema específico sin posibilidad de generalizar.

Análisis multidimensional OLAP. Mediante visores de cubos OLAP, es posible generar consultas dinámicas de tipo MDX. Es decir, el usuario puede elegir los KPI's o medidas y dimensiones a visualizar. Las operaciones OLAP que se pueden realizar son:

- Roll-up (Agregación). Eliminación de un criterio de agrupación para el análisis.
- Drill-down (Disgregación). Introducción de un nuevo criterio de agrupación, disgregando los grupos actuales.
- Slice and dice. Consiste en seleccionar y proyectar datos en una consulta.
- Pivote. Rotar la visualización de los datos, transformando filas en columnas.

Imagen 7.6. Esquema de las aplicaciones BI



Dashboard (Tablero de control). Son interfaces visuales que resumen la información del negocio mediante los indicadores clave de desempeño (KPI) utilizando gráficos estadísticos, valores escalares, medidores, semáforos, etc.

Minería de datos (DM: Data Mining). Es el proceso de exploración mediante técnicas descriptivas y predictivas que permiten descubrir un conocimiento oculto (patrones) a partir un conjunto de datos (data sets) o bases de datos (KDD: Knowledge Discovery in Databases). DM aplica métodos y técnicas de estadística, inteligencia artificial, y aprendizaje automático.

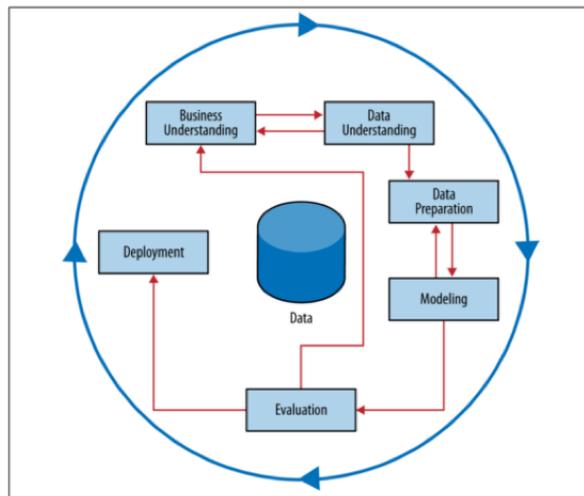
Metodologías para el desarrollo de sistemas BI y DM

Existen varias metodologías que proponen un ciclo de vida para crear soluciones BI y/o DM; entre las más destacadas se encuentran: Imon, Kimball, HEFESTO, CRISP-DM.

Metodología CRISP-DM (Cross Industry Standard Process for Data Mining)

Esta metodología es un estándar de la industria de la minería de datos y aplicaciones BI que describe un proceso organizado por fases para llevar a cabo un proyecto de minería de datos. Las actividades de esta metodología se observan en la Imagen 7.7 y se describen a continuación:

Imagen 7.7. Fases de la Metodología CRISP-DM



Fuente: (Provost & Fawcett, 2013)

- **Comprensión del negocio.** Consiste en realizar actividades como: identificar los requerimientos empresariales, el planteamiento del problema, los objetivos del proyecto de análisis de datos y el establecimiento de un plan de trabajo.
- **Comprensión de Datos.** En esta fase, se realizan actividades de identificación de los indicadores clave de desempeño, las dimensiones de análisis, familiarización con las estructuras de datos fuentes y verificación de su calidad.
- **Preparación de datos.** Comprende actividades de limpieza, transformación, formateo, selección e integración de datos. Estas actividades se ejecutan en reiteradas ocasiones hasta cumplir los requerimientos del negocio.
- **Modelado.** En esta fase se crean los modelos (diseños) de los conjuntos de datos o almacenes de datos en base a los indicadores de desempeño del negocio (KPI's) o requerimientos de análisis de datos. Con el propósito de refinar el diseño, se revisa y se corrige el (los) modelo(s) y si es necesario se vuelve a la fase de Preparación de datos.
- **Evaluación.** Cada uno de los modelos de datos con perspectiva de análisis, son evaluados según criterios de calidad.
- **Despliegue.** En esta actividad, se realizan tareas que pueden ser simples o complejas, como la generación de reportes (gráficos o tablas estadísticas) o la implantación de una plataforma de explotación de información que proporcione acceso controlado a los usuarios tomadores de decisiones en toda la organización.

Caso de estudio: diseño e implementación de un sistema de inteligencia de negocios aplicado al sector camaronero

Enunciado del caso de estudio

El consorcio de empresas ABC maneja N empresas que producen camarón; cada empresa tiene a cargo una camaronera que está ubicada en un país, provincia y cantón; Cada camaronera está organizada por piscinas. Cada piscina tiene un número de hectáreas de producción. El proceso de producción de camarón se denomina “corrida”, cada corrida tiene un número secuencial, fecha de inicio y fin, estado (concluido o en proceso) y una o más piscinas donde se siembra el camarón. Por cada piscina que es parte de una corrida, se registra el costo de inversión, el número de larvas de camarón sembradas, la cantidad en Kilogramos de camarón cosechados clasificados por talla (Grande, Mediano y Pequeño) y el total de ingresos por la venta de la cosecha (\$ por venta total y por talla). La empresa puede tener una o más corridas activas, pero una piscina no puede estar en más de una corrida activa en el mismo periodo de tiempo.

Aplicando un proceso metodológico de inteligencia de negocios, realizar el diseño de un data warehouse que satisfaga los requerimientos planteados; luego, el proceso ETL con datos simulados desde Excel; a continuación, la creación de Cubos OLAP; y finalmente, el diseño e implementación de un dashboard BI como aplicación EIS. Aplicar el proceso metodológico de CRISP-DM. En este caso se omite la actividad de Evaluación.

Comprensión del negocio

Requerimientos del negocio (preguntas del negocio)

1. ¿Cuántos Kilogramos de camarón se han producido por talla y empresa en el último año?

2. ¿Cuáles son los gastos (inversión), los ingresos y la utilidad por empresa en un año determinado?
3. ¿Cuántas hectáreas de producción de camarón tiene el consorcio por empresa, corridas activa y piscina en el último año?
4. ¿Cuántas hectáreas de producción de camarón dispone el consorcio por empresa, piscina y país?
5. ¿Cuál es el número promedio de larvas de camarón por hectárea que se siembran por empresa y camaronera?
6. ¿Cuál es la empresa que ha generado más utilidades por año?
7. ¿Cuál es la empresa que ha producido más camarón por talla y año?

Las preguntas se han obtenido en base a entrevistas aplicadas a algunos productores de camarón. Se ha resaltado el texto en color azul a aquellos posibles indicadores de desempeño o medidas del negocio, se ha subrayado las potenciales dimensiones, el color verde denota posible filtro aplicar a una dimensión y el color naranja denota sentido de orden.

Planteamiento del problema

Necesidad de diseñar una solución de inteligencia de negocios para el proceso de producción en un consorcio camaronero.

Objetivo

Diseñar una solución de inteligencia de negocios (Dashboard EIS) para el proceso de producción en un consorcio camaronero, empleando técnicas de análisis multidimensional OLAP y herramientas que permiten el diseño de un data warehouse, la ejecución del proceso ETL y la implementación de un dashboard BI (panel de control de los principales KPI's), con el propósito de mantener oportunamente informados a los ejecutivos que toman decisiones.

Plan de trabajo

El plan de trabajo para este caso de estudio, se divide en actividades, tareas, recursos, responsables y distribución del tiempo y costos según la metodología CRISP-DM. Por cuestión de limitaciones de espacio no se lo incluye en este texto.

Comprensión de los datos

En el Cuadro 7.1 se identifican los principales indicadores clave de desempeño (KPI's) con sus respectivas dimensiones.

Cuadro 7.1: Identificación de indicadores clave de rendimiento (KPI) y dimensiones

Data Mart	KPI	Código KPI	Dimensión
Estadística de piscinas	Número de hectáreas de producción de camarón	pis_hectareas	estado, empresa, país, periodo (año, semestre), piscina
Estadística de corridas, piscinas y siembra	Número de larvas de camarón sembradas por hectárea	cp_num_larvas	empresa, camaronera, país
Estadística de cosechas	Número de kilogramos de camarón cosechado	cose_num_kilos	talla, empresa, periodo (año, semestre)
	Gastos de producción (inversión)	cose_inversion	
	Total de ingresos por ventas	cose_total	
	Utilidad	cose_utilidad	

Fuente: Elaboración propia

En el Cuadro 7.2 se preparan los data marts con sus tablas de hechos, KPI's y funciones de agregación o fórmulas que se aplicarán a los datos. Las posibles funciones de agregación que se pueden aplicar son: sumariación (SUM), conteo (COUNT), promedio (AVG), mínimo (MIN), máximo (MAX). También es posible aplicar fórmulas matemáticas que combinan operadores y/o funciones de agregación.

Cuadro 7.2: Funciones de agregación por indicador clave de desempeño

Data Mart	Tabla de hecho	Función de agregación o fórmula por KPI
Estadística de piscinas	th_piscina	SUM (pis_hectareas)
Estadística de corridas, piscinas y siembra	th_corrida_piscina	AVG(cp_num_larvas)
Estadística de cosechas	th_cosecha	SUM(cose_num_kilos) SUM(cose_inversion) SUM(cose_total) SUM (cose_utilidad) = [SUM(cose_total) - SUM(cose_inversion)]

Fuente: Elaboración propia

Preparación de datos

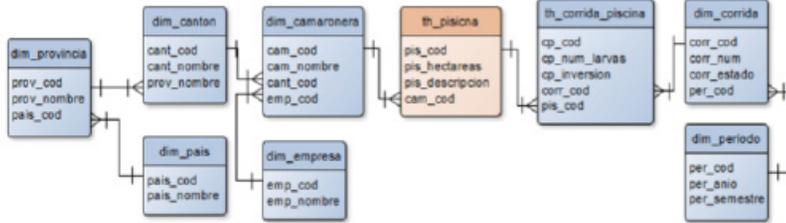
En esta fase lo ideal es tener acceso a datos reales de una empresa como una base de datos de un sistema transaccional. Debido a que se está tratando con un caso de estudio simulado, se trabajó con datos en una hoja de cálculo.

Modelado

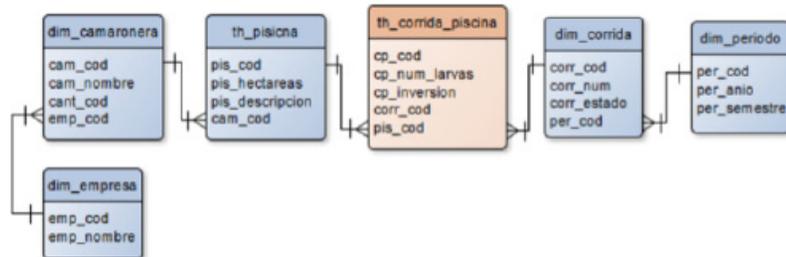
En la Imagen 7.8, se muestran los modelos lógicos de los data marts que conforman la data warehouse según el caso de estudio de control de producción del consorcio camaronero.

Imagen 7.8. Modelos lógicos de data marts: a) Información de piscinas y proceso productivo (corrida) por camaronera, b) Siembra de larvas en una piscina y c) Cosecha de camarón.

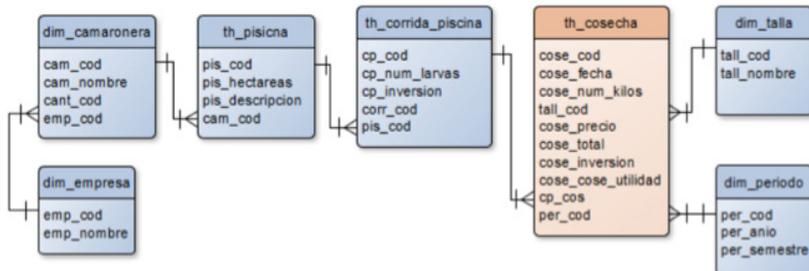
a) Data Mart 1: Información de piscinas y ciclos (corrida) de producción por camaronera



b) Data Mart 2: Siembra de larvas en una piscina camaronera



c) Data Mart 3: Cosecha de camarón



Fuente: Elaboración propia

Despliegue. Implementación de un dashboard en Power BI

Descarga e instalación de Power BI Desktop

Un dashboard o tablero de control es una herramienta que representa de manera gráfica los indicadores claves de desempeño (KPI) de una organización. En este caso se utilizó la herramienta Microsoft Power BI Desktop versión freeware, debido a que ofrece un conjunto de herramientas de inteligencia de negocios para el análisis de datos destinada a usuarios empresariales, profesionales de TI y desarrolladores.

Para su descarga, acceder al link en la página oficial del instalador PBIDesktop_x64.msi, <https://go.microsoft.com/fwlink/?LinkId=521662&clid=0x40a>.

Datos fuente. Los datos fuente utilizados para este caso de estudio, se pueden descargar del siguiente enlace:

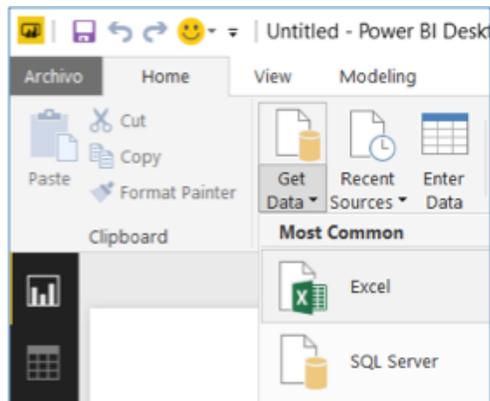
https://drive.google.com/file/d/1HYEuTuc_yU42g4y-ZpH7Mslj3Ghw9o2fH/view?usp=sharing

Creación de cubos OLAP y gráficos

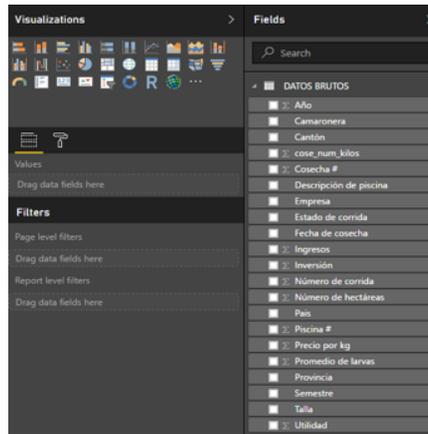
Primero se ingresará al entorno de trabajo de Power BI Desktop y se seguirán los pasos del Cuadro 7.3.

Cuadro 7.3: Pasos para la creación de gráficos estadísticos

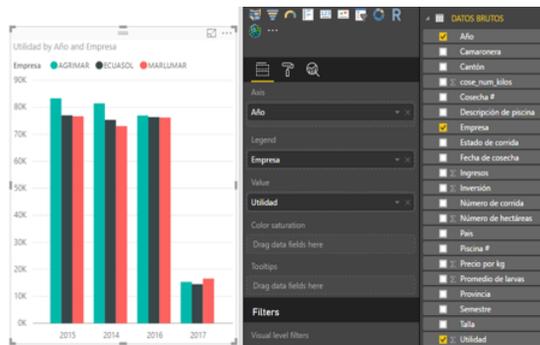
1. Dar clic en Get Data, a continuación en Excel y seleccionar el archivo con los datos fuentes.



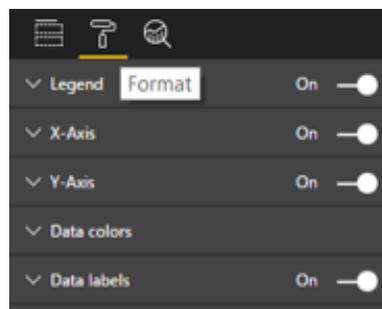
5. Para diseñar los gráficos estadísticos de acuerdo a las preguntas de negocio de cada data mart identificado, seleccionar el tipo de gráfico y luego las medidas y dimensiones que se deseen en la parte derecha de la ventana.



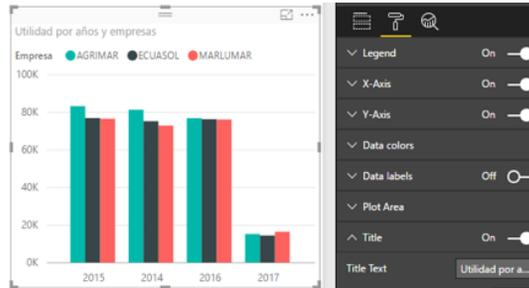
6. Una vez seleccionadas las medidas y dimensiones, el gráfico se va visualizando automáticamente, las medidas se añadirán a la sección de Valores, mientras que las dimensiones pueden ubicarse en Ejes o Leyenda, esto dependerá de la forma de visualización del gráfico.



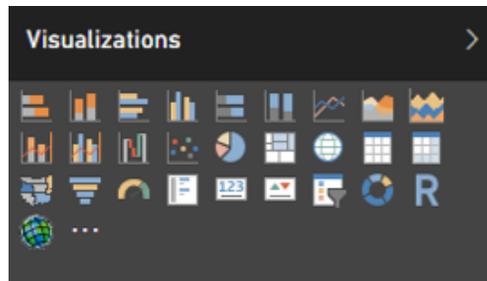
7. Se pueden mostrar los valores en las barras, dirigirse a la pestaña Formato y habilitar opción Etiquetas de datos.



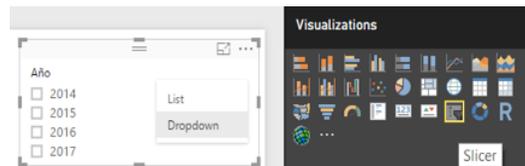
8. Para cambiar el título al gráfico ubicarse en Título de la pestaña Formato, se modifica el texto del título y listo.



9. Se puede cambiar el estilo del gráfico seleccionando otro diseño en Visualizaciones.



10. Para ubicar filtros se debe arrastrar la dimensión que se desee hacia un espacio en blanco del área de trabajo y seleccionar el ícono Slicer.



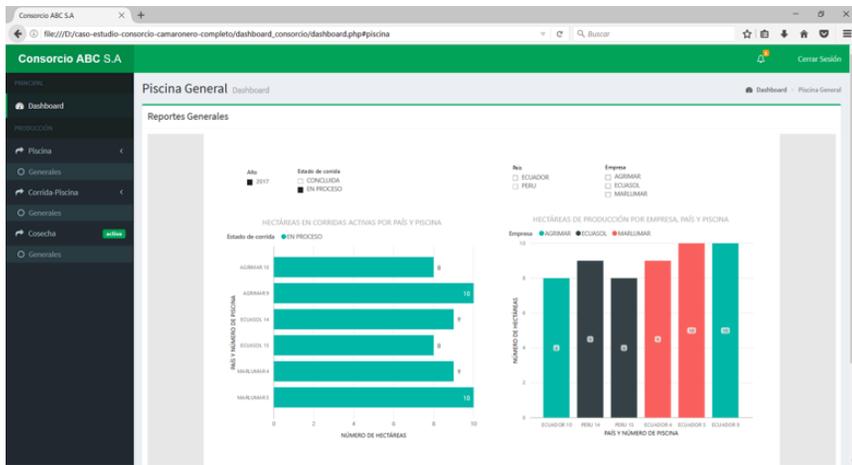
Dashboards BI

Un dashboard es un tablero de control que permite el manejo integrado de la información importante de la empresa que es útil para la toma de decisiones. En este caso de uso, primero se diseñaron los gráficos estadísticos y luego se publicaron en internet en la cloud de Power BI mediante una cuenta que se puede crear gratuitamente; sin embargo, esta cuenta es limitada y, si se requieren de más servicios o prestaciones se debe cancelar un valor mensual o anual. Para darle un estilo personalizado de aplicación web, se creó

el menú de opciones mediante código HTML y se embebieron los gráficos de Power BI. En la Imagen 7.9, se visualizan tres capturas de pantalla de la aplicación web correspondiente a los dashboard BI creados.

Imagen 7.9. Capturas de pantalla de la aplicación web (dashboar BI) con estadísticas de producción de camarón de un consorcio camaronero. En a) estadística de piscinas, en b) estadísticas de siembra y en b) estadísticas de cosecha.

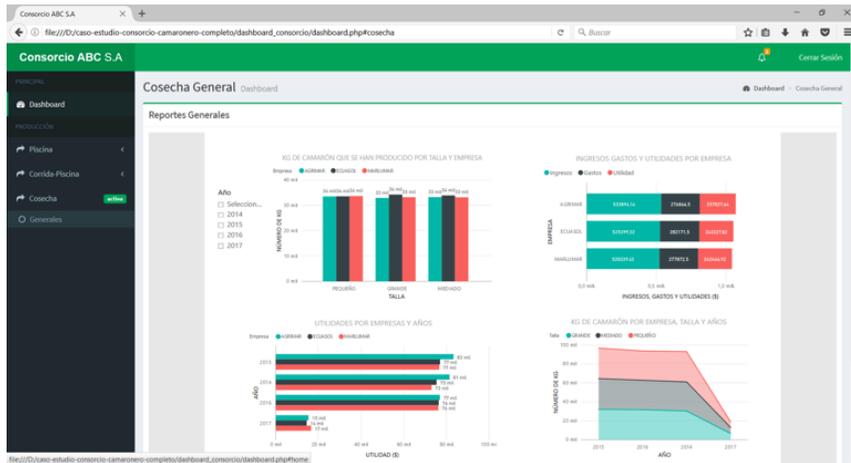
a) Estadística de piscinas



b) Estadística de siembra



c) Estadística de cosecha



En esta capítulo se presentó un panorama de lo que se puede hacer con la Inteligencia de Negocios en el sector agropecuario. Primero se explican los fundamentos teóricos de BI y mediante un caso de estudio enfocado en la producción de camarón, se evidencia una un ejemplo sencillo de dashboard BI.

Para el caso de estudio se utilizó datos en bruto simulados con el fin de explicar el proceso de construcción de una aplicación BI. El proceso metodológico se llevó a cabo mediante la metodología CRISP-DM. En la fase de Comprensión del Negocio se describió los requerimientos del negocio, mediante una serie de preguntas con una estructura particular (medida o KPI, dimensiones, filtros y sentido de orden). En la fase de Comprensión de los Datos, se determinó los principales data marts con sus respectivas medidas o indicadores clave de rendimiento (KPIs) y dimensiones. En la fase de Preparación de Datos, se creó el modelo de los data marts. Y en la fase de Despliegue, se construyó la aplicación BI mediante la herramienta Microsoft Power BI Desktop, que integra tres dashboards: a) estadística de piscinas, b) estadísticas de siembra b) estadísticas de cosecha.

Referencia Bibliográfica

- Cornejo, R., Navarrete, M., Valdivia, R., Aroca, P., & Aracena, S. (2014). Desarrollo de una base de datos integrada de Censo y encuesta mediante el uso de elementos de inteligencia de negocios y SIG. *Ingeniare. Revista Chilena de Ingeniería*, 22, 205-217. <http://doi.org/10.4067/S0718-33052014000200007>
- De Mauro, A., Greco, M., & Grimaldim, M. (2015). What is Big Data ? A Consensual Definition and a Review of Key Research Topics. *In International Conference on Integrated Information (IC-ININFO 2014)* (Vol. 1644, pp. 97-104). <http://doi.org/10.1063/1.4907823>
- FAO. (2016). *Programa mundial del censo agropecuario 2020. Volumen 1. Programa, definiciones y conceptos*. Retrieved from <http://www.fao.org/3/a-i4913s.pdf>
- Ghosh, R., Halder, S., & Sen, S. (2015). An Integrated Approach to Deploy Data Warehouse in Business Intelligence Environment. *In Proceedings of the 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT)* (p. 7). <http://doi.org/10.1109/C3IT.2015.7060115>
- Gounder, M. S., Iyer, V. V., Professor-ccis, A., Mazyad, A. Al, & Prof, A. (2016). A Survey on Business Intelligence tools for University Dashboard development. *2016 3rd MEC International Conference on Big Data and Smart City*.
- Laudon, K. C., & Laudon, J. P. (2012). *Sistemas De Información Gerencial*. (Pearson, Ed.) (12 Edición). México: Pearson Education.
- Marinheiro, A., & Bernardino, J. (2015). Experimental Evaluation of Open Source Business Intelligence Suites using OpenBRR, 13(3), 810-817.
- Mazon-Olivo, B., Rivas, W., Pinta, M., Mosquera, A., Astudillo, L., & Gallegos, H. (2017). Dashboard para el soporte de decisiones en una empresa del sector minero. *Conference Proceedings - Universidad Técnica de Machala*, 1, 1218-1229. Retrieved from <http://investigacion.utmachala.edu.ec/proceedings/index.php/utmach/article/view/219/191>
- Moniruzzaman, A., & Hossain, S. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. *International Journal of Database Theory and*

- Application*, 6(4), 1-14. Retrieved from http://www.sersc.org/journals/IJDTA/vol6_no4/1.pdf
- Morales, A., Cuevas, R., & Martínez, J. (2016). Procesamiento Analítico con Minería de Datos. *Revista Iberoamericana de Las Ciencias Computacionales E Informática*, 5.
- Mosquera, L., & Hallo, M. (2014). Data Mart Para El Sistema De Servicios Sociales Del Conadis. *Revista Politécnica*, 33(2).
- NIST. (2015). NIST Special Publication 1500-1 NIST. Big Data Interoperability Framework : Volume 1 , Definitions. *National Institute of Standards and Technology*, 1, 32. <http://doi.org/10.6028/NIST.SP.1500-1>
- Provost, F., & Fawcett, T. (2013). *Data Science for Business. What you need to know about Data Mining and Data-Analytic thinking*. O'Reilly Media.
- Rosado, A., & Rico, D. (2010). Business Intelligence :State of the Art. *Scientia Et Technica*, XVI(44), 321-326. Retrieved from <http://www.redalyc.org/articulo.oa?id=84917316060>
- Rosado G., A. A., & Rico B., D. W. (2010). Inteligencia de Negocios: Estado del Arte. *Scientia Et Technica*, XVI, 321-326. Retrieved from <http://www.redalyc.org/articulo.oa?id=84917316060>
- Sawant, N., & Shah, H. (2013). *Big Data Application Architecture A&A*. Apress. <http://doi.org/10.1007/978-1-4302-6293-0>
- Țăranu, I. (2015). Big Data Analytics Platforms analyze from startups to traditional database players. *Database Systems Journal*, VI(1), 23-32. Retrieved from http://www.dbjournal.ro/archive/19/19_3.pdf
- Van Der Lans, R. F. (2012). *Data Virtualization for Business Intelligence Systems. Revolutionizing Data Integration for Data Warehouses. Data Virtualization for Business Intelligence Systems*. Elsevier Inc. <http://doi.org/10.1016/B978-0-12-394425-2.00014-9>
- Vassell, M., Apperson, O., Calyam, P., Gillis, J., & Ahmad, S. (2016). Intelligent Dashboard for Augmented Reality based Incident Command Response Co-ordination. *IEEE Consumer Communications and Networking Conference*, 0-3. <http://doi.org/10.1109/CCNC.2016.7444921>
- White, T. (2015). *Hadoop: The Definitive Guide, 4th Edition*. O'Reilly (Vol. 54).

Análisis de Datos Agropecuarios

Edición digital 2017- 2018.

www.utmachala.edu.ec

Redes

Redes es la materialización del diálogo académico y propositivo entre investigadores de la UTMACH y de otras universidades iberoamericanas, que busca ofrecer respuestas glocalizadas a los requerimientos sociales y científicos. Los diversos textos de esta colección, tienen un espíritu crítico, constructivo y colaborativo. Ellos plasman alternativas novedosas para resignificar la pertinencia de nuestra investigación. Desde las ciencias experimentales hasta las artes y humanidades, Redes sintetiza policromías conceptuales que nos recuerdan, de forma empeñosa, la complejidad de los objetos construidos y la creatividad de sus autores para tratar temas de acalorada actualidad y de demanda creciente; por ello, cada interrogante y respuesta que se encierra en estas líneas, forman una trama que, sin lugar a dudas, inervará su sistema cognitivo, convirtiéndolo en un nodo de esta urdimbre de saberes.



UNIVERSIDADE DA CORUÑA

UNIVERSIDAD TÉCNICA DE MACHALA

Editorial UTMACH

Km. 5 1/2 Vía Machala Pasaje

www.investigacion.utmachala.edu.ec / www.utmachala.edu.ec

ISBN: 978-9942-24-120-7



9 789942 241207