**Universidad Nacional de Misiones**

**Facultad de Ciencias Exactas, Químicas y Naturales**

**Tesis de grado Licenciatura en Sistemas de Información**

**Solución OLAP Para Analizar Las Actividades Productivas Agropecuarias**.

“Estudio de caso sobre los productores agropecuarios del departamento General Manuel Belgrano”

**Autor:**Fernando Esteban Acosta

**Director:** Dra. Antonia Elisabeti Godoy

**Co-Director:** Lic. Jesús Germán Andrés Pautsch

**Año: 2016**

*Dedicatoria*

**Resumen**

Actualmente el término Business Intelligence es ampliamente conocido y utilizado mundialmente. Sin embargo, en la provincia de Misiones y en especial en los sectores productivos agropecuarios, aún no se han logrado importantes desarrollos en este campo. Las autoridades de los municipios del departamento General Manuel Belgrano mediante proyectos llevados a cabo con la FCEQyNde la Universidad Nacional de Misiones, han buscado avanzar en esta área en pos de posicionar sus productos en los mercados agropecuarios, Implementando estrategias que ayuden a su vez, a que sus productores mejoraren su calidad de vida participando en el desarrollo de actividades productivas de dicho mercado.

Con la presente tesis se pretende aplicar técnicas de inteligencia de negocios que sirvan de soporte para la toma de decisiones tanto a las autoridades para conocer la situación que atraviesa su municipio y así gestionar sus recursos en forma adecuada permitiendo con ello ofrecer a los productores oportunidades de acceso a los distintos mercados con bajo un nivel de incertidumbre.

**Palabras claves:***Información - Producción Agropecuaria - BI - OLAP - DataWarehouse*

**Abstract**

Nowadays the term Business Intelligence is widely known and worldwide used. Nevertheless, in the state of Misiones and especiallyin the agricultural productive sectors, have not been achieved important developments in this field, yet. The authorities of the municipalities of the Department General Manuel Belgrano through projects carried out with the FCEQyN of the Nacional University of Misiones, have sought to make progress in this area in pursuit of set their products in agricultural markets, by implementing strategies to help to turn their producers will improve their quality of life by participating in the development of productive activities of that market.

This thesis aims to apply techniques of business intelligence that serve as support for the decision-making both to the authorities to know the situation in their municipality and manage their resources appropriately thereby giving producers opportunities of access to the several markets with a low level of uncertainty.

**KeyWords:***Information - Agriculture Production - BI - OLAP - DataWarehouse*

Índice

[Capítulo 1 15](#_Toc446475774)

[1.1. Justificación 16](#_Toc446475775)

[1.1.1. Contexto de la Tesis 16](#_Toc446475776)

[1.1.2. Motivación 16](#_Toc446475777)

[1.2. Objetivos Generales 18](#_Toc446475778)

[1.3. Objetivos Específicos. 19](#_Toc446475779)

[1.4. Estructura del documento 19](#_Toc446475780)

[Capítulo 2 21](#_Toc446475781)

[2.1. Business Intelligence 22](#_Toc446475783)

[2.2. OLAP 23](#_Toc446475784)

[2.3. Data Warehouse 24](#_Toc446475785)

[2.3.1. Enfoques para la construcción de un DW 24](#_Toc446475786)

[2.4. Data Mart 26](#_Toc446475787)

[2.5. Modelo Multidimensional de un DW 26](#_Toc446475788)

[2.5.1. Tabla de Dimensiones 27](#_Toc446475789)

[2.5.2. Tabla de Hechos 27](#_Toc446475790)

[2.6. Esquemas para el Modelo de un DW 27](#_Toc446475791)

[2.6.1. Esquema en estrella. 27](#_Toc446475792)

[2.6.2. Esquema de Copo de Nieve. 28](#_Toc446475793)

[2.6.3. Esquema Constelación. 29](#_Toc446475794)

[Capítulo 3 31](#_Toc446475795)

[3.1. Productores del Departamento General Manuel Belgrano 32](#_Toc446475797)

[3.2. Unidades Productivas 33](#_Toc446475798)

[3.3. Análisis de Situación 34](#_Toc446475799)

[Capítulo 4 35](#_Toc446475800)

[4.1. Del Proyecto de Tesis 36](#_Toc446475802)

[4.2. Metodología Hefesto 36](#_Toc446475803)

[4.3. Características de la Metodología 37](#_Toc446475804)

[4.4. Fases de la Metodología Hefesto 38](#_Toc446475805)

[4.4.1. Análisis de Requerimientos 38](#_Toc446475806)

[4.4.1.1. Identificar preguntas: 38](#_Toc446475807)

[4.4.1.2. Identificar Indicadores y perspectivas 38](#_Toc446475808)

[4.4.1.3. Modelo Conceptual 39](#_Toc446475809)

[4.4.2. Análisis de los OLTP 39](#_Toc446475810)

[4.4.2.1. Conformar los indicadores. 40](#_Toc446475811)

[4.4.2.2. Establecer Correspondencias 40](#_Toc446475812)

[4.4.2.3. Nivel de granularidad 40](#_Toc446475813)

[4.4.2.4. Modelo conceptual Ampliado 41](#_Toc446475814)

[4.4.3. Modelo Lógico del DW 41](#_Toc446475815)

[4.4.3.1. Tipo de Modelo Lógico 42](#_Toc446475816)

[4.4.3.2. Tablas de Dimensiones 42](#_Toc446475817)

[4.4.3.3. Tablas de Hechos 42](#_Toc446475818)

[4.4.3.4. Uniones 43](#_Toc446475819)

[4.4.4. Integración de Datos 44](#_Toc446475820)

[4.4.4.1. Carga Inicial 44](#_Toc446475821)

[4.4.4.2. Actualización 44](#_Toc446475822)

[4.5. Selección de la herramienta de BI 45](#_Toc446475823)

[4.6. Recursos Disponibles 46](#_Toc446475824)

[Capítulo 5 49](#_Toc446475825)

[5.1. Modelo de Datos Multidimensional 50](#_Toc446475827)

[5.2. Análisis de Requerimientos 51](#_Toc446475828)

[5.2.1. Identificar Preguntas 51](#_Toc446475829)

[5.2.2. Identificar indicadores (In) y perspectivas (Pe) 52](#_Toc446475830)

[5.2.2.1. Análisis de Rentabilidad del Productor 52](#_Toc446475831)

[5.2.2.2. Análisis de Evolución de la Producción 52](#_Toc446475832)

[5.2.2.3. Análisis de Preferencias Productores 52](#_Toc446475833)

[5.2.2.4. Análisis de Posibles Relaciones entre Tipo de Producción y UP 53](#_Toc446475834)

[5.2.3. Modelos Conceptuales 53](#_Toc446475835)

[5.2.3.1. Análisis de Evolución de la Producción 53](#_Toc446475836)

[5.2.3.2. Análisis de Rentabilidad de la Producción 54](#_Toc446475837)

[5.3. Análisis de los OLTP 54](#_Toc446475838)

[5.3.1. OLTP Productores 54](#_Toc446475839)

[5.3.2. Conformar Indicadores 55](#_Toc446475840)

[5.3.3. Establecer Correspondencias 57](#_Toc446475841)

[5.3.4. Correspondencias fuente de Datos y Modelo Conceptual 58](#_Toc446475842)

[5.3.5. Nivel de Granularidad 59](#_Toc446475843)

[5.3.5.1. Diseño de las Perspectivas 61](#_Toc446475844)

[5.3.6. Modelo Conceptual Ampliado 64](#_Toc446475845)

[5.4. Modelo Lógico del DW 64](#_Toc446475846)

[5.4.1. Tipo de Modelo Lógico del DW 65](#_Toc446475847)

[5.4.2. Tablas de Dimensiones 65](#_Toc446475848)

[5.4.3. Tablas de Hechos 68](#_Toc446475849)

[5.4.4. Uniones 69](#_Toc446475850)

[5.5. Integración de Datos 69](#_Toc446475851)

[5.5.1. Carga Inicial 69](#_Toc446475852)

[5.5.2. Actualización 70](#_Toc446475853)

[5.6. Creación de Cubos Multidimensionales 71](#_Toc446475854)

[Capítulo 6 73](#_Toc446475859)

[6.1. Análisis De Los Cubos 74](#_Toc446475861)

[6.1.1. Análisis De La Evolución De La Producción 74](#_Toc446475862)

[6.1.1.1. Análisis De La Producción Por Zonas 75](#_Toc446475863)

[6.1.1.2. Resultado del Análisis de la Producción 80](#_Toc446475864)

[6.1.1.3. Participación de cada localidad en las Mayores Producciones 81](#_Toc446475865)

[6.1.1.4. Resultados del Analisis de Participación 84](#_Toc446475866)

[6.1.2. Análisis de la Rentabilidad de la Producción 84](#_Toc446475867)

[6.1.3. Resultado del Análisis de la Rentabilidad 88](#_Toc446475868)

[6.2. Conclusiones e interpretaciones de los usuarios 89](#_Toc446475869)

[Capítulo 7 90](#_Toc446475870)

[6.3. Conclusiones 91](#_Toc446475871)

[6.4. Futuras líneas de investigación 92](#_Toc446475872)

[ANEXOS 94](#_Toc446475873)

[Anexo 1 96](#_Toc446475874)

[Anexo 2 105](#_Toc446475875)

[Anexo 3 113](#_Toc446475876)

[Bibliografía 119](#_Toc446475877)

Índice de Figuras

[Figura 1: Esquema en Estrella 29](#_Toc446475991)

[Figura 2: Esquema de Copo de Nieve 29](#_Toc446475992)

[Figura 3: Esquema Constelación 30](#_Toc446475993)

[Figura 4: Modelo Conceptual Hefesto 40](#_Toc446475994)

[Figura 5: Modelo Conceptual Ampliado Hefesto 42](#_Toc446475995)

[Figura 6: Correspondecia Perspectiva y Tabla de Dimensiones- Hefesto 43](#_Toc446475996)

[Figura 7: Correspondecia Indicadores y Tabla de Hechos- Hefesto 44](#_Toc446475997)

[Figura 8: Uniones Entre Tabla de Hechos y Dimensiones - Hefesto 44](#_Toc446475998)

[Figura 9: Componentes de Pentaho. Fuente: *Pentaho Open Source Business Intelligence* 47](#_Toc446475999)

[Figura 10: Modelo Coneptual Analisis de la Producicón 54](#_Toc446476000)

[Figura 11: Modelo Coneptual Analisis de la Rentabilidad 55](#_Toc446476001)

[Figura 12: Elementos a Utilizar del Modelo E/R del Sistema OLTP 58](#_Toc446476002)

[Figura 13: Modelo Conceptual Ampliado del Análisis de Producción y Rentabilidad 65](#_Toc446476003)

[Figura 14: Modelo Multidimensional para el Análisis de la Producción 70](#_Toc446476004)

[Figura 15: Cubo para el Análisis de Producción 72](#_Toc446476005)

[Figura 16: Selección de dimensiones y medidas para el análisis de la producción por zonas 76](#_Toc446476006)

[Figura 17: Total de kg producidos y hectáreas cultivadas de cada producto por año 77](#_Toc446476007)

[Figura 18: Total de kg producidos y hectáreas cultivadas en cada localidad 77](#_Toc446476008)

[Figura 19: Selección de dimensiones y medidas para graficar la producción 78](#_Toc446476009)

[Figura 20: Total de Hectareas cultivadas por producto 79](#_Toc446476010)

[Figura 21: Total de kilogramos producidos por producto 80](#_Toc446476011)

[Figura 22: Total de kilogramos producidos por producto 81](#_Toc446476012)

[Figura 23: Filtro para el análisis de cada producto](file:///C:\Users\afer\Dropbox\Trabajo Final\Proyecto Final\Dcumentos\Libro de Tesis\Informe Final 22-03.docx" \l "_Toc446476013) 82

[Figura 24: Participación de cada localidad en base a las hectáreas de tabaco cultivadas 83](#_Toc446476014)

[Figura 25: Participación de cada localidad en base a las hectáreas de mandioca cultivadas 84](#_Toc446476015)

[Figura 26: Selección de medidas y dimensiones para el analisis de rentabilidad 85](#_Toc446476016)

[Figura 27: Rentabilidad de la producción de cada localidad por año 86](#_Toc446476017)

[Figura 28: Proyección de rentabilidad de la producción en San Antonio 87](#_Toc446476018)

[Figura 29: Proyección de rentabilidad de la producción en Cmte. Andresito. 88](#_Toc446476019)

[Figura 30: Proyección de rentabilidad de la producción en Bdo. De Irigoyen 89](#_Toc446476020)

[Figura 31: Trabajo para la carga del cubo para el análisis de la producción y rentabilidad 107](#_Toc446476021)

[Figura 32: Transformación para la carga de Productores 108](#_Toc446476022)

[Figura 33: Transformación para la Carga de Productos 108](#_Toc446476023)

[Figura 34: Transformación para las Zonas Territoriales 109](#_Toc446476024)

[Figura 35: Transformación para la Clasificación de Productos 110](#_Toc446476025)

[Figura 36: Transformación para la Fecha 111](#_Toc446476026)

[Figura 37: Transformación para la Producción 111](#_Toc446476027)

Índice deTablas

[Tabla 1: Correspondencia OLTP Productores y Perspectivas del Modelo Conceptual 61](#_Toc446466459)

[Tabla 2: Correspondencia OLTP Productores e Indicadores del Modelo Conceptual 61](#_Toc446466460)

[Tabla 3. Campos Candidatos Perspectiva Productos 62](#_Toc446466461)

[Tabla 4: Campos Candidatos Perspectiva Zona Territorial 62](#_Toc446466462)

[Tabla 5: Campos Candidatos Perspectiva Productores 63](#_Toc446466463)

[Tabla 6: Campos Candidatos Perspectiva Clasificación de Productos 64](#_Toc446466464)

[Tabla 7: Campos Candidatos Perspectiva Tiempo 64](#_Toc446466465)

[Tabla 8: Diseño de la Perspectiva Productos 65](#_Toc446466466)

[Tabla 9: Diseño de la Perspectiva categoría de Productos 65](#_Toc446466467)

[Tabla 10: Diseño de la Perspectiva categoría de Productos 66](#_Toc446466468)

[Tabla 11: Diseño de laPerspectiva categoría de Productos 66](#_Toc446466469)

[Tabla 12: Diseño de laPerspectiva categoría de Productos 67](#_Toc446466470)

[Tabla 13: Diseño de la Dimensión Productos 68](#_Toc446466471)

[Tabla 14: Diseño de la Dimensión Productores 69](#_Toc446466472)

[Tabla 15: Diseño de la Dimensión Zona Territorial 69](#_Toc446466473)

[Tabla 16: Diseño de la Dimensión Clasificación Productos 70](#_Toc446466474)

[Tabla 17: Diseño de la Dimensión Tiempo 70](#_Toc446466475)

[Tabla 18: Diseño de la Tabla de Hechos 71](#_Toc446466476)

[Tabla 19: Valores para el cálculo de rentabilidad para la producción de mandioca 104](#_Toc446466477)

[Tabla 20: Valores para el cálculo de rentabilidad para la producción del tabaco 104](#_Toc446466478)

Capítulo 1

Introducción

* 1. Justificación
     1. Contexto de la Tesis

En el presente trabajo de tesis se toma como caso de estudio los productores agropecuarios del Departamento General Manuel Belgrano ubicado al Noreste de la Provincia de Misiones. En la cual predominan las unidades productivas de pequeñas dimensiones, en las que el cultivo se realiza a escala micro. Por su clima y calidad de suelo, ofrece buenas condiciones para la producción de frutas, hortalizas, productos cárnicos, madera, etc., e incluso a futuro,con una correcta administración, podría constituirse en un polo de producción de alimentos.

La ausencia de información en cuanto a la oferta y demanda de la producción agropecuaria y registro de productores, además de la falta de medios para transportar los productos a lugares alejados de sus unidades productivas, dificulta la comercialización.

Para Godoy y Otros, (2013 y 2014), las excelentes ventajas comparativas por las condiciones naturales de la región no son aprovechadas por estos pequeñosproductores. El 80% de los productos que se consume en el mercado, proviene de otras provincias y/o países y se observa aquí una población de escasos recursos económicos. No se visualizan cambios sin la aplicación de acciones concretas y planificadas, tanto por parte de los productores como de las instituciones responsables del sector, para revertir la problemática.

Según Godoy y Otros (2013), a partir de la crisis del año 2001, surgieron planes sociales y apoyo al sector agropecuario, con un alto costo para el Estado pero bajo impacto de desarrollo socioeconómico en la región. Esto se replica en el sector cooperativo donde se promocionó, mediante subsidios, la creación de cooperativas agropecuarias, sin que éstas logren el desarrollo, alrededor del 90% de las mismas se encuentran sin o con escasa actividad económica.

* + 1. Motivación

En la actualidad Sostienen Godoy y Otros, (2013) y Acosta (2012)que, gracias a los avances dela informática, generar y acceder a lainformación no debería ser un problema, ya que ésta se puede crear en un punto y estar disponible en otro distante por medio de la web.Durante las tareas de relevamientorealizados en este proyecto, se constató que en la zona no todos los productores tienen acceso a la tecnologías de Información, pero la información puede centralizarse y estar disponible a través de la secretaria dependiente de cada municipio.

Del trabajo de campo realizado se ha logradorelevar información sobre variables que,una vezorganizadas, estructuradas y almacenadas en una bases de datos (BD),permitirála implementación de herramientas de análisis de datos, con el fin de generarinformes con un alto grado de fiabilidad,para que sirvan a los responsables,del proceso de toma de decisión.Una de las alternativas disponibles para el análisis de información, son las sentencias SQL[[1]](#footnote-2)en las que se pueden realizar consultas sobre BD, recuperando fácil y rápidamente toda la información almacenada sobre un criterio de interés.

Aun con la potencialidad de las consultas SQL, existe información oculta que puede ser de gran importancia para la toma de decisiones, difícil de reconocer por el gran volumen y variedad de datos almacenados. Por lo tanto, para lograr extraer dicha información sería necesaria la utilización de técnicas más avanzadas como el OLAP[[2]](#footnote-3).La implementación del OLAP tiene un amplio desarrollo en el campo de los negocios orientados a la toma de decisiones. Sin embargo, según la revisión bibliográfica, su aplicación conjunta en Sistemas de Gestión Agropecuariaes aun escaza.

Entre los trabajos hallados se encuentra una propuesta de solución OLAP en las PyMEs, que propone un modelo multidimensional común para empresas del sector yerbatero del nordeste argentino (Zamudio, 2011). Plataforma web para la gestión Agro-ganadera eimplementación de técnicas OLAP como apoyo a la toma de decisiones (David L. la Red Martínez, Lisandro Yaccuzzi, 2014).Modelo de datos multidimensional y análisis OLAP para la producción agropecuaria (Radulescu, y Otros, 2009). Procesamiento Analítico On-line en la agriculturamediante el uso de cubos multidimensionales (Chaturvedi y Otros, 2008).

En el escenario planteado y a partir de los diversos proyectos desarrollados, bajo la dirección de la Dra. Godoy en el Departamento General Manuel Belgrano, entre los que figuran: La marca Andresito Produce, la página Web y el sistema de información soporte para la logística de la comercialización, surge la necesidad de la creación de herramientas que favorezcan a las autoridades en los procesos de toma de decisión con el fin de posicionar sus productos en los mercados y así ampliar las oportunidades de los productores, hoy muy limitadas por la ausencia de información centralizada que permita conocer la situación y actividades que desarrolla cada productor del departamento.

Las autoridades de la municipalidad de Andresito, en conjunto con los demás municipios y la secretaria del Agro y la Producción, tienen por objeto apoyar al sector productivo e implementar herramientas que sirvan para la logística de la comercialización de los productos agropecuarios, es imprescindible para ello la implementación de herramientas que permitan almacenar la información referente a todo el sector productivo y permita su posterior extracción para mediante herramientas de análisis, diseñar soluciones de Inteligencia de Negocios que una vez automatizada, seria operada por la Secretaria de Agro y la producción de cada municipio permitiendo acceder a información referida al funcionamiento de la organización del municipio, la producción, administración de ayudas económicas y subsidios, y demás actividades diarias.

* 1. Objetivos Generales

El objetivo de este proyecto, es construir un Data Warehouse (DW), para generar una solución OLAP que permitan reconocer características de los perfiles de los productores del Departamento General Manuel Belgrano y las principales actividades desarrolladas por estos, con lo cual se permitan distintos niveles de análisis como ser, rentabilidad de las actividades productivas agropecuarias, volúmenes de producción y mercados posibles, como soporte para el proceso de toma de decisiones de los productores y de las autoridades del municipio.

* 1. Objetivos Específicos.
* Identificar las necesidades de información de las autoridades municipales en el contexto de producción Agropecuaria.
* Reconocer las áreas de interés de los usuarios y los procesos de negocio asociados.
* Definir los indicadores claves de las áreas de interés y las perspectivas que demarcan su análisis.
* Establecer un modelo lógico multidimensional que permita acceder de maneraeficiente a la información.
* Analizar las fuentes de información asegurando la automatización de los procesos de carga que faciliten su acceso desde la herramientatecnológica.
* Construir un prototipo que permita el análisis de los datos a partir de cubos OLAP.
  1. Estructura del documento

La tesis se divide en seis capítulos y tres anexos. A continuación se describe brevemente el contenido de los mismos.

En el Capítulo 2: se abordan los fundamentos teóricos necesarios para la implementación de una solución OLAP, inicialmente se introduce en el concepto de *Business Intelligence* (sección 2.1), se continúa con un análisis de las técnicas OLAP y las alternativas para su implementación (sección 2.2). Posteriormente se describen los alcances de los almacenes de datos que se pueden emplear ya se un *Data Warehouse* (sección 2.3) o *Data Mart* (sección 2.4). Finalizando se define el concepto de modelo multidimensional de un DW (sección 2.5) y los distintos enfoques de diseño (sección 2.6).

El Capítulo 3 describe la problemática que enfrentan los productores de los distintos municipios del departamento General Manuel Belgrano (sección 3.1), en el contexto de la producción agropecuaria (sección 3.2) y la utilización de herramientas de inteligencia de negocios para dar soporte a la toma de decisiones (sección 3.3).

ElCapítulo 4 inicia con una breve conceptualización del trabajo de tesis (sección 4.1) para luego dar una introducción sobre la metodología seleccionada para abordar el trabajo de investigación (sección 4.2) describiendo sus principales características (sección 4.3) y las cuatro fases propuestas para su aplicación (sección 4.4): Análisis de Requerimientos (sección 4.4.1), Análisis de los OLTP (sección 4.4.2), Modelo Lógico del DW (sección 4.4.3) e Integración de Datos (sección 4.4.4). Seguidamente se describen las herramientas seleccionada (sección 4.5) y los Recursos Disponibles (sección 4.6).

El Capítulo 5 propone el desarrollo de la metodología seleccionada para la construcción del DW (sección 5.1), con elanálisis de los requerimientos de las autoridades municipales y susproductores (sección 5.2), el análisis del OLTP Productores (sección 5.3) y los demás Orígenes de Datos. Luego se presenta el modelo lógico del DW (sección 5.4), y los procesos parala integración de los datos (sección 5.5). Se cierra el capítulo con la creación de los cubos multidimensionales para su posterior aplicación de OLAP (Sección 5.6).

Finalmente el Capítulo 6 contiene las conclusiones obtenidas a partir del trabajo realizado y los posibles trabajos futuros.

El Anexo 1contiene el análisis y desarrollo realizado para poblar de datos del OLTP productores y así hacerlo funcionala fin que sirva como fuente de datos para el DW.

El Anexo 2 muestra los procesos de extracción, transformación y carga realizados para cargar las tablas de bases de datos.

El Anexo 3 contiene la estructura de los cubos.

Capítulo 2

Marco Teórico

En este capítulo se abordan los fundamentos teóricos necesarios para la implementaciónde una solución OLAP,inicialmente se introduce en el concepto de *Business Intelligence* (sección 2.1), se continua con un análisis de las técnicas OLAP y las alternativas para su implementación (sección 2.2). Posteriormente se describenlos alcances de los almacenes de datos que se pueden emplear ya se un Data Warehouse (sección 2.3) o Data Mart(sección 2.4).Finalizando se define el concepto de modelomultidimensional de un DW (sección 2.5) y los distintos enfoques de diseño (sección 2.6).

1. 1. Business Intelligence

El termino *Business Intelligence* (BI) tiene referencia en 1958 cuando el investigador H. P. Luhn (1958),propuso un sistema de inteligencia de negocio que aceptaba información en su formato original, que tras la automatización de procesamiento de datos, los diseminaba adecuada y rápidamente a los lugares correspondientes.

No fue sino hasta 1989 donde los sistemas de cómputo obtuvieron un desarrollo importante que Howard Dresner, considerado por muchos como el padre de la inteligencia de negocios, retomó las ideas de Luhn, para reformular la implicación de estos sistemas y elevar la categoría de los entonces llamados sistemas de soporte a las decisiones (DSS – *Decision Support Systems*). Él definióla inteligencia de negocios como “un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos” (Osorio, 2011).

Mediante la inteligencia de negocios se permite la extracción de conocimiento en base al análisis de los datos acumulados que, sirviendo como soporte a la toma de decisión y aprovechada por la expertise de los usuarios, se pueden lograr importantes ventajas competitivas.

El desarrollo de una solución de inteligencia de negocios requiere de las fuentes de datos,los procesos de limpieza y calidad de datos, la zona de almacenamiento y el análisis de los datos donde puede obtenerse desde una simple consulta y reporte, hasta la posibilidad de analizar información mediante técnicas OLAP, *dashboards*, minería de datos.

* 1. OLAP

El término OLAP propuesto originalmente por Codd (Codd,1993), con el fin de reflejar su diferencia con los sistemas OLTP[[3]](#footnote-4). Su significado deriva de los requisitos funcionales de los usuarios (analistas, administradores, y directivos) para visualizar, analizar y consolidar los datos de una organización en un modelo multidimensional.

El análisis multidimensional es solamente una de las características del OLAP, junto con una definición de los tipos de usuarios que lo utilizan y los requerimientos de negocio provistos por estos. Básicamente ofrece al usuario la posibilidad de responder a preguntas tales como “¿Qué pasaría si...?” y “¿por qué...?” en un contexto de análisis de datos organizacionales(Keneth y Otros, 2004). Esto es posible gracias a un análisis dinámico de datos provenientes de sistemas transaccionales u otras fuentes y que permite un entendimiento de los cambios ocurridos dentro del negocio (Zamudio, 2011).

Existen diferentes alternativas para la implementación de sistemas OLAP, estas son:

* + ROLAP:Se implementa sobre tecnología relacional. Utiliza un esquema en estrella cuyo nodo central representa a la tabla de hechos y sus extremos a las dimensiones (Kimball,1996). Con esta metodología, ante una consulta se genera el cubo correspondiente. Esta alternativa de generación de cubos se utiliza cuando no se posee gran capacidad de almacenamiento. Al generarse los cubos en tiempo de ejecución su rendimiento no es óptimo.
  + MOLAP:Tiene la estructura de *arrays* multidimensionales en la cuallos cubos son generados y almacenados antes de ser consultados por los expertos. Los datos son tomados de la tabla de hechos y las dimensiones,previamente calculadas y almacenadas. Si se dispone de suficiente espacio en disco, esta alternativa aumenta el rendimiento y mejora los tiempos de respuesta frente al ROLAP.
  + HOLAP:Es una combinación de las técnicas ROLAP y MOLAP. Los cubos frecuentemente consultados son generados y almacenados. Cualquier otra consulta debe generarse en tiempo de ejecución.
  1. Data Warehouse

Anteriormente se mencionó que los datos utilizados en OLAP provienen de sistemas transaccionales y otras fuentes, los cuales son tratados y volcados en un repositorio denominado DW.

Un DW es una gran base de datos que integra datos transaccionales de distintas áreas pertenecientes a una organización. Inmon (2005) define un DW como una *“colección de datos orientados a temas, integrados, no volátiles, y variantes en el tiempo para dar soporte a las decisiones de gestión”*. Sumada a esta definición, otro aspecto fundamental de un DW es la granularidad, la cual indica el nivel de detalle en que los datos pueden ser presentados. Este aspecto afecta directamente el tipo de consultas que pueden hacerse sobre los datos. Aunque existen distintos enfoques para adoptar la granularidad, el nivel requerido estará dado por el tipo de preguntas que deberán ser respondidas con dichos datos.

* + 1. Enfoques para la construcción de un DW

Respecto de la construcción del DW en general se recomienda un modo iterativo y ordenado, aunque existen dos enfoques bien diferenciados.Uno de ellos propuesto por Inmon, quien dice que *“el DW requiere una estructura que comience viendo un todo, y entonces vaya a lo particular”* (Inmon,2005). Esto implica una construcción que comience por un DW respecto de un tema mayor, luego abarcar los siguientes temas, y una vez que pueden diferenciarse correctamente los requisitos, se crean almacenes de datos departamentales, llamados *Data Mart* (DM), que responden a necesidades específicas de un grupo reducido de usuarios de la organización, reflejando un único proceso de negocios.

El otro enfoque, propuesto por Kimball (Kimball,2002), consiste en encarar la construcción desde una perspectiva inversa, esto implica comenzar por el desarrollo de DM y luego, a partir de éstos, abarcar los distintos procesos de negocios de la organización.

De los dos enfoques comentados anteriormente y en base experiencias en amplias investigacionesnace Hefesto, una metodología para la contruccion de DW(Bernabeu, 2010).

Hefesto se caracteriza por buscar soluciones capaces de adaptarse con facilidad y rapidez a los cambios. Para ello emplea ciclos cortos que culminan en la implementacion de soluciones que satisfagan necesidades particulares de los usuarios. Para ello Hefesto propone cuatro etapas basicas para la construccion de un DW.

Como primera etapa la metodologia propone un analisis de requerimientos, en la cual se busca identificar las necesidades de los usuarios empleando preguntas que expliciten los objetivos de suorganizacion. Prosteriormente estas preguntas serán analizadas a fin de distinguir los indicadores y perspectivas que serviran en la construccion del DW. Finalmente la etapa propone con la confeccion de un modelo conceptual que presenta el resultado del analisis.

La siguiente etapa consiste en el analisis de las distintas fuentes de información a fin de determinar la forma en la cual se calcularán los indicadores y para establecer sus correspondencias con el modelo conceptual creado en la fase anterior. Al finalizar se lograra un modelo conceptual ampliado con la información de los campos que conformaran los indicadores y las perspectivas.

En la tercera etapa Hefesto propone la confección del modelo logico de la estructura del DW. Para ello, define en primer lugar el tipo de esquema de DW a emplear (estrella, constelacion o copo de nieve). Luego plantea el diseño de las tablas de dimensiones y hechos, culminando con las relaciones entre estas.

Por ultimo la metodologia de Bernabeu propone realizar la prueba del modelo con datos utilizando las correspondientes tecnicas de limpieza, calidad de datos y procesos ETL[[4]](#footnote-5). Tras la carga inicial y la prueba exitosa del modelo, se definen las reglas y politicas a tomar en cuenta para proximas actualizaciones.

Procesos ETL

Los procesos ETL son actividades implementadas en la integración de Datos, mediante los cuales se busca que los datos provenientes de las distintas fuentes sean debidamente capturados cada ejemplar, analizarlos y los datos que formen parte de un mismo campo sean adaptados y estandarizados para poder ser cargados en la estructura de la DW.

* 1. Data Mart

Un *Data Mart* es un subconjunto del DW, por lo que mantiene las mismas características que este, pero su contexto de aplicación se enmarca en ofrecer soluciones de un área específica de la organización. Según Inmon (2005) “*un DM dependiente es un subconjunto lógico (vista) o un subconjunto físico (extracto) de un almacén de datos más grande*".

Desarrollar un DW puede conllevar un gran esfuerzo de tiempo y recursos, por lo cual la implementación de DM departamentales permite soluciones a requerimientos del momento. Sin embargo, cada DM debe ser bien planificado y orientado a los procesos para no disminuir el rendimiento e incrementar la redundancia de forma innecesaria (Chanduri, 2001).

* 1. Modelo Multidimensional de un DW

Un modelo multidimensional es una base de datos en donde su información se almacena en tablas de hechos y tablas de dimensiones,esta última con los valores descriptivos de las medidas contenidas en los hechos. Este tipo de modelos permite representar los datos de interés, desde las perspectivas que el encargado de la toma de decisiones contemple necesarios.

* + 1. Tabla de Dimensiones

Las tablas de dimensiones se conforman de un identificador único y los datos de referencia que describen los criterios de análisis relevantes para la organización. Estas permiten a los usuarios filtrar y manipular la información almacenada en las tablas de hechos.

* + 1. Tabla de Hechos

Contienen las medidas, que son los puntos de interés que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Conformado generalmente por datos cuantitativos y las claves de cada una de lastablas de dimensión relacionadas.Los hechos son datos instantáneos en el tiempo, que pueden ser filtrados, agrupados y explorados a través las condiciones definidas en las tablas de dimensiones.

* 1. Esquemas para el Modelo de un DW

Para abordar el modelado de un DW se puede optar por distintas formas de relacionar la tabla de hecho y las tablas de dimensiones(Bernabeu, 2010).

* + 1. Esquema en estrella.

El esquema en estrella propone un modelo con una tabla de hechos central relacionada a varias dimensiones. Requiere una desnormalización de los datos, tiene la ventaja de que ofrece tiempos de respuesta másrápidos ya que omite el uso de uniones (*join*) entre tablas en cada consulta. Como punto en contra tiene que genera redundancia pero el sacrificio de espacio no es significativo.

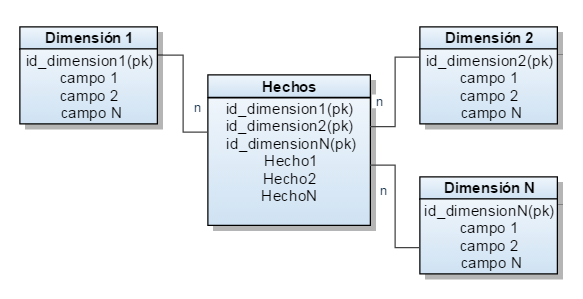


Figura 1: Esquema en Estrella

* + 1. Esquema de Copo de Nieve.

Similar al esquema en estrella con una tabla de hechos centrar y referencias de dimensiones, y estas a su vez se relacionan con otras tablas de dimensiones. Muy útil cuando se posee un gran volumen de datos que con el esquema de estrella implicaría un aumento considerable del espacio utilizado, este propone un uso más correcto del espacio. Como desventaja presenta que, ante una consulta,es necesario el uso de uniones cuyo procesamiento podrían afectar su desempeño.

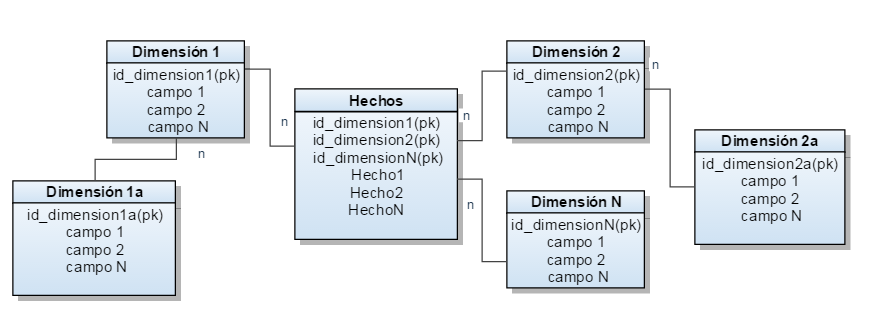


Figura 2: Esquema de Copo de Nieve

* + 1. Esquema Constelación.

Este modelo propone un conjunto de esquemas en estrella, con una tabla de hechos principal y otras tablas de hechos auxiliares que pueden ser calculadas de la principal. Los hechos se mantienen en el centro pudiendo cualquiera, principal o auxiliares, estar relacionadas por cualquiera de las dimensiones inclusive compartirlas. Al proponer más de una tabla de hechos, permite el análisis de más de un tema de interés con un mínimo de esfuerzo adicional, y permite la reutilización de las tablas de dimensión. Tanto su diseño como implementación implican un mayor nivel de complejidad por lo cual no todas las herramientas de consulta y análisis permiten su implementación.

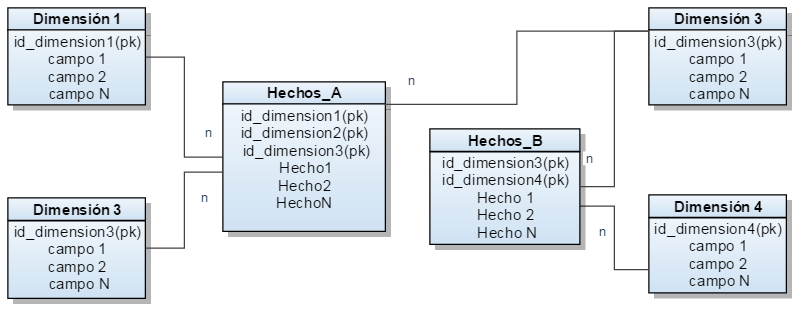


Figura 3: Esquema Constelación

Capítulo 3

Descripción Del Problema

En este capítulo se describe la problemática que enfrentan los productores de los distintos municipios del departamento General Manuel Belgrano (sección 3.1), en el contexto de la producción agropecuaria (sección 3.2) y la utilización de herramientas de inteligencia de negocios para dar soporte a la toma de decisiones (sección 3.3).

1. 1. Productores del Departamento General Manuel Belgrano

El departamento General Manuel Belgrano, posee una gran capacidad productiva y recursos naturales disponibles para su producción. Se desarrollan actividades referidas a la producción agropecuaria cuya difusión sistematizada es mínima y en algunos casos hasta se desconocen, situación que complica la toma de decisiones por parte de las autoridades del Municipio y por los mismos productores.

Los productores no poseen poder de negociación, si comercializan en forma atomizada, es por ello que conforman cooperativas agropecuarias; pero éstas se encuentran en una fase de desarrollo y con dificultades en su organización. La fuerte incidencia del sector informal en la producción agropecuaria; el bajo nivel de información y comunicación en el mencionado sector, deriva en la falta de oportunidades para crecer (Godoy y Otros, 2010).

La municipalidad de Comandante Andresito en conjunto con otros municipios, trabajan en el marco del desarrollo local y con la creación de la Secretaria del Agro y la Producción, tiene por objeto apoyar al sector productivo e implementar herramientas que sirvan para la logística de la comercialización de los productos agropecuarios (Rolón, 2012).

Dada las condiciones desfavorables con respecto a la información y el interés de las autoridades municipales de contar con ella, es de suma importancia registrar las actividades inherentes a la producción agrícola, la que se puede viabilizar mediante una base de datos. La misma será operada desde la Secretaria del Agro y la Producción del municipio, y brindara información referida al funcionamiento de una organización, como también de sus procesos y actividades diarias.

* 1. Unidades Productivas

El departamento General Manuel Belgrano posee una población que ha crecido de 11.987 habitantes a 33.152, entre los años 1.980 a 2.001 y según el censo del año 2010 la población actual es de 42.929(Censo,2010). Se verificó un crecimiento del 29,5 % respecto del censo anterior. El 70% de la población tiene menos de 20 años de residencia en dicho Departamento; las actividades y producción agropecuarias que desarrollan no poseen difusión sistematizada y en algunos casos hasta se desconocen.

Se verifica una importante incidencia del sector informal en la producción agropecuaria de los pequeños productores, con bajo nivel de información y dificultades en la planificación de su unidad productiva, que tiene como correlato la falta de desarrollo de este sector (Godoy y Otros, 2010).

Luego de la eliminación de la Secretaría de Planeamiento de la Provincia, y de la Comisión Reguladora de la Yerba Mate (C.R.Y.M.), son escasos los organismos que puedan brindar información actualizada de las principales actividades en la Provincia. La falta de disponibilidad de tales conjuntos de datos referentes al sector agropecuario es, probablemente, uno de los mayores obstáculos para su crecimiento y posicionamiento en la Provincia de Misiones

Por este motivo, es de suma importancia almacenar información correspondiente a los productores y sus principales actividades u operaciones con el objeto de contar con información a ser utilizada en los procesos de toma de decisión, tanto por los productores como por las autoridades responsables del sector.

En base a la aplicación de las encuestas del Programa de Reorganización de Asociación de Pequeños Productores del Departamento General Manuel Belgrano y Capacitación de sus Recursos Humanos, entre las actividades productivas frecuentadas por las unidades se encuentran.

* **Producción Agrícola:** Maíz – Poroto – Soja – C.Azúcar – Mandioca – Batata –Maní–Tabaco.
* **Forestación y cultivos complementarios:** Paraná – Toona – Paraíso – Pinus – Eucaliptus – Nativas - Yerba mate- Tung – Pasturas – Tartago.
* **Frutas:** Naranja - Mandarina - Pomelo - Limón - Palta - Banana - Kiwi - Nativas - Mamón – Ananás.
* **Verduras:** Tomate - Pepino – Rábano – Zapallo – Zapallito – Zanahoria- Remolacha – Apio – Morrón – Berenjena – Lechuga – Cebollita – Orégano – Perejil – Acelga – Espinaca – Repollo – Choclo.
* **Producción Avícola:** Aves para carne - Aves para huevos Vacunos.
* **Producción Porcina:** Porcinos - Madres – Padrillos - Lechones – Cachorros.
  1. Análisis de Situación

Actualmente, en el departamento General Manuel Belgrano se han trabajado con proyectos a cargo de quien dirige la presente tesis, que dieron lugar entre otras al desarrollo de un sistema OLTP. Sin embargo el mismo aún no ha podido ser implementado por cuestiones propias de cada municipio.A causa de ello no se cuenta con información centralizada acerca de quiénes son productores, cual es el volumen que estos producen, que extensión de tierra destinan a cada actividad, si han percibido algún beneficio económico para afrontar la producción, entre otras.Esta falta de información dificulta a cada municipio tomar decisiones respecto a cómo dirigir su productividad. Implementar una solución de inteligencia de negocios permitiría a las autoridades de cada municipio, implementar estrategias para orientar a sus productores en actividades cuya demanda les permita acceder a estos mercados, como así, generar planes o programas que ayuden a los productores a acceder a nuevas alternativas de producción disminuyendo perdidas al estado por subsidios cuyo fin no tenga un correcto destino.

Capítulo 4

Materiales y Métodos

En este capítulo se inicia con una breve conceptualización del trabajo de tesis (sección 4.1) para luego dar una introducción sobre la metodología seleccionada para abordar el trabajo de investigación (sección 4.2) describiendo sus principales características (sección 4.3) y las cuatro fases propuestas para su aplicación (sección 4.4): Análisis de Requerimientos(sección 4.4.1), Análisis de los OLTP (sección 4.4.2), Modelo Lógico del DW (sección 4.4.3) e Integración de Datos (sección 4.4.4). Seguidamente se describen las herramientas seleccionada (sección 4.5) y los Recursos Disponibles (sección 4.6).

1. 1. Del Proyecto de Tesis

Esta investigación de tipo aplicada, tendrá como fuentes de información el censo tabacalero del año 2013 y encuestas generadas en proyectos a cargo de la misma dirección (A.Godoy y Otros, 2012).

Como unidad de análisis se determinó forman parte los productores cuya unidad de producción se ubique dentro de la zona territorial del Departamento General Manuel Belgrano.

* 1. Metodología Hefesto

Hefesto es una metodología de licencia GNU[[5]](#footnote-6)que combina las mejores prácticas de propuestas existentes como Inmon y Kimball. Propone cuatro etapas para la creación de un almacén de datos a partir de la recolección de requerimientos y diseño del modelo multidimensional, seguido de la definicióndel esquema lógico, ya sea para la creación de un *Data Warehouse* o un *Data Mart,*seguido de los procesos de extracción, transformación y carga (ETL) que permitan su correcta integración de datos (Bernabeu, 2010). Luego de comparar las propuestas de cada metodología[[6]](#footnote-7), el enfoque seleccionado es Hefesto, ya que permite dar énfasis a los requerimientos de los usuarios permitiendo que los resultados persigan los objetivos definidos al momento del planteamiento del *Data Warehouse* (Haro-Valle y Otros,2014)(Brizuela y Otros, 2013).

Bernabeu propone con su desarrollo, enno caer en extensas fases de recolección, análisis y desarrollo, sino que busca entregar una primera implementación que satisfaga una parte de las necesidades, a fin de con ello demostrar las ventajas del DW y motivar con así a los usuarios.

* 1. Características de la Metodología
* Los objetivos y resultados esperados en cada fase se distinguen fácilmente y sonsencillos de comprender.
* Se basa en los requerimientos de los usuarios, por lo cual su estructura es capazde adaptarse con facilidad y rapidez ante los cambios en el negocio.
* Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cadaetapa para que tome decisiones respecto al comportamiento y funciones del DW.
* Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
* Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
* Es independiente de las herramientas que se utilicen para su implementación.
* Es independiente de las estructuras físicas que contengan el DW y de su respectivadistribución.
* Cuando se culmina con una fase, los resultados obtenidos se convierten en el puntode partida para llevar a cabo el paso siguiente.
* Se aplica tanto para Data Warehouse como para Data Mart.
  1. Fases de la Metodología Hefesto

Hefesto propone cuatro Etapas o Fases para la Construcción de un DW. A continuación se propone una descripción de cada una de ellas con sus correspondientes pasos.

* + 1. Análisis de Requerimientos

Lo primero será identificar los requerimientos de los usuarios mediante técnicas de relevamiento que permitan definir preguntas queexpliciten los objetivos de la organización.

Luego, se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas a tomar en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.

* + - 1. Identificar preguntas:

Se obtienen mediante técnicas de relevamiento, referencian las necesidades de información claves de alto nivelrepresentadasen consultas que los usuarios desean responder, deben ser complejasy abarcarcada posible condición al que esté sujeto el análisis.

* + - 1. Identificar Indicadores y perspectivas

De las preguntas definidas se procede a su descomposición a fin de reconocer indicadores y perspectivasde análisis. Los indicadores son usualmente valores cuantitativos (montos, cantidades, porcentajes), sobre los cuales radica el análisis. Mientras las Perspectivas, usualmente datos cualitativos, son los distintos puntos de vista sobre las cuales se pueden analizar los indicadores (tiempo, lugares, productos)

* + - 1. Modelo Conceptual

Como resultado de la primera fase, la metodología proponemediante una representación gráfica, las perspectivas (rectángulos a izquierda) que conformaran las dimensiones del cubo OLAP y los indicadores (rectángulos a derecha) que conformaran las medidas de la tabla de hechos. Ambos grupos unidos al ovalo central que representa la relación entre ellas, es decir el área de estudio elegida.

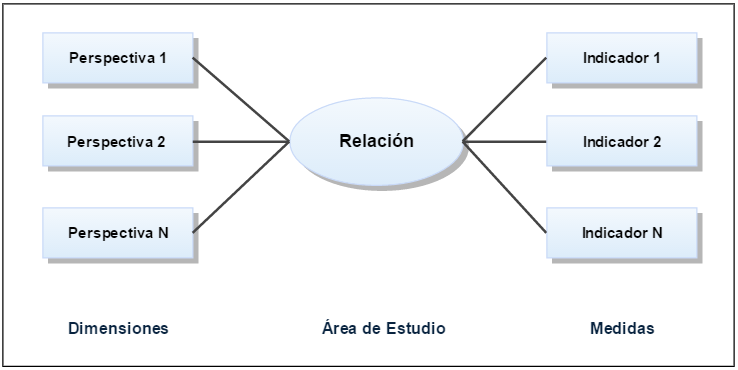


Figura 4: Modelo Conceptual Hefesto

* + 1. Análisis de los OLTP

La segunda fase propone se analicen las fuentes de datos a fin de determinar la forma en la que serán calculados los indicadores y establecer las correspondencias entre el modelo conceptual creado en la fase anterior y las fuentes de datos. Luego, se definirán los campos que se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida en esta fase.

* + - 1. Conformar los indicadores.

Se define la fórmula de cómo se debe calcular cada indicador en base a los siguientes conceptos:

* Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1+ Hecho2.
* Función que se utiliza para su agregación. Por ejemplo: SUM, AVG, etc.
  + - 1. Establecer Correspondencias

En este paso, se deben examinar las fuentes de datos a fin de establecer cuál es la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y cada elemento de las fuentes de datos. De este paso se debe poder establecer la correspondencia de cada elemento de las fuentes de datos con su correspondiente indicador o perspectiva del modelo conceptual.

* + - 1. Nivel de granularidad

Una vez establecidas las relaciones entre el modelo conceptual y las fuentes de datos, se deben seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores.

Es importante la participación de los usuarios en este paso, ya que con ellos se debe establecer el nivel de detalle requerido para cada campo.

* + - 1. Modelo conceptual Ampliado

Como resultado de esta fase se obtieneun gráficoque sintetiza los resultados obtenidos en los pasos anteriores, ampliando el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo.

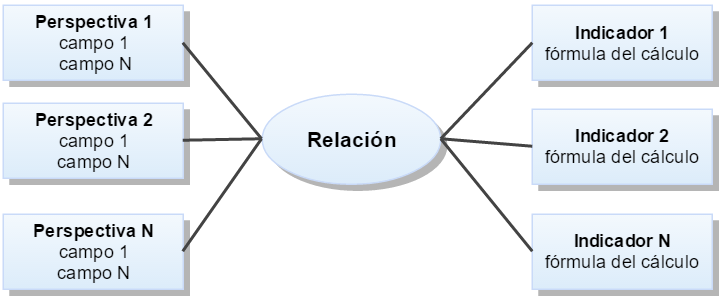


Figura 5: Modelo Conceptual Ampliado Hefesto

* + 1. Modelo Lógico del DW

En su tercera etapa, teniendo como partida base el modelo conceptual Ampliado, la metodología plantea la confección del modelo lógico del DW. Para ello, primero se defineel tipo de modelo a utilizar y luego se procede a diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas.

* + - 1. Tipo de Modelo Lógico

En este paso se seleccionael esquema que mejor se adapte a los requerimientos y necesidades de los usuarios para contener la estructura del depósito de los datos[[7]](#footnote-8).

* + - 1. Tablas de Dimensiones

El diseño de las tablas de Dimensiones que formaran parte del DW independientemente del esquema elegido, estarán definidas por cada perspectiva presente en el modelo conceptual ampliado. Para ello debe tomarse cada perspectiva con sus campos relacionados, Se elige un nombre que identifique la tabla de dimensión. Luego Se añade un campo que represente su clave principal. Finalmente se redefinen los nombres de los campos que no sean lo suficientemente intuitivos.

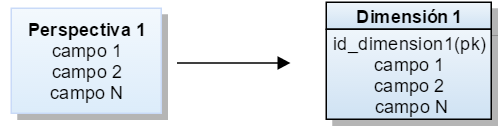


Figura 6: Correspondecia Perspectiva y Tabla de Dimensiones- Hefesto

* + - 1. Tablas de Hechos

Las tablas de hechos son las que contendrán las medidas y las referencias de las dimensiones a través de los cuales se construirán los indicadores de estudio. Inicialmente se debe asignar un nombre a la tabla de hechos que represente la informaciónanalizada, área de investigación, negocio enfocado, etc. Luego, se define su clave primaria, la cual se compone de la combinación de las clavesprimarias de cada tabla de dimensión relacionada. Finalmente se crean tantos campos de hechos como indicadores se hayan definido en elmodelo conceptual, pudiendo modificar el nombre de cada hecho a fin de facilitar su interpretación.

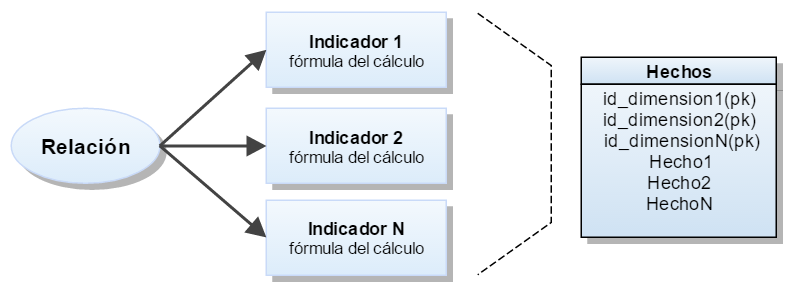


Figura 7: Correspondecia Indicadores y Tabla de Hechos- Hefesto

* + - 1. Uniones

En este paso se deben concretar las uniones correspondientes entre las tablas de dimensiones y la tabla de hechos.

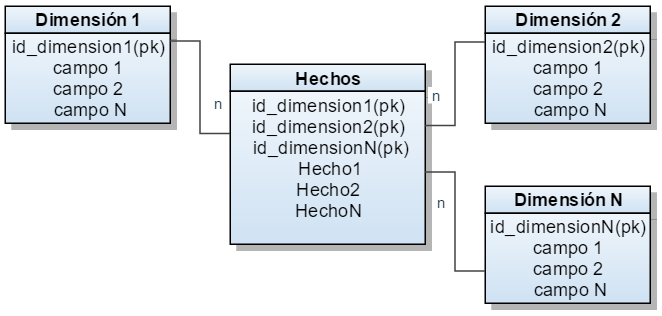


Figura 8: Uniones EntreTabla de Hechos y Dimensiones - Hefesto

* + 1. Integración de Datos

Una vez construido el modelo lógico, se procede a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc. Luego se definen las reglas y políticas para su respectiva actualización, como así también los procesos que lollevarán a cabo.

* + - 1. Carga Inicial

En este paso se procede a la carga inicial del DW, poblando el modelo de dato construido anteriormente. Para ello, se proponen una serie de tareas básicas como limpieza y calidad de datos, procesos ETL, etc. Para ello se cuentan con herramientas software que facilitan estas tareas.

Con este paso se busca no simplemente poblar con datos la DW, sino que los mismos sean completos y de calidad, evitando los campos faltantes o anómalos como también aquellos que no tendrán utilidad en análisis en cuestión.Se procede con la carga de los datos de cada dimensión y luego de las tablas de hechos respetando las correspondencias existentes entre cada elemento.

En resumen, la carga iniciales el conjunto de pasos a realizar para, extraer información de las fuentes de datos, establecer las actividades que validan la calidad de la información extraída y cargar estos campos validados en la DW.

* + - 1. Actualización

Con el DW poblado de datos, se deben establecer las políticas y estrategias de actualización o refresco de datos. Para lo cual se especifican inicialmente las tareas de limpieza y calidad de datos, procesos ETL, etc., a realizarse para la correspondiente actualización de los datos del DW. Luego de forma detallada, especificar las acciones que deberán realizarse en cada software para lograr dicha actualización.

* 1. Selección de la herramienta de BI

Para la implementación de la solución de BI se ha optado por la herramienta Pentaho en su versión *Community Edition[[8]](#footnote-9)*. Es una plataforma *Open Source*ampliamente utilizadaen el Área de BI (Marinheiro y Otros, 2013) con numerosos trabajos que validad sus ventajas en proyectos aplicados de BI(Haro-Valle y Otros, 2014) (Hernández y Otros, 2011) .Analistas de Gartner Group (2015), manifiestanque el 20% de clientes referentes afirman utilizar Pentaho para la recolección y análisis de datos, lo posiciona entre las primeras alternativas en crecimiento al momento de abordar proyectos de BI.

Pentaho es una suite de software centrada en procesos y orientada a la solución mediante flujos de datos, con componentes que permiten desarrollar soluciones integradas. Proporciona un espectro completo de herramientas de inteligencia de negocios capaz de abordar procesos como integración de datos, análisis OLAP, reportes, *dashboards*y minería de datos.

La figura 9 muestra una representación gráfica de los componentes que conforman la Suite Pentaho. Entre sus capas se identifican la de presentación en lo más alto y la capa de datos e integración de aplicaciones en lo más bajo. Las áreas funcionales más importantes reportes, análisis, *dashboards* y administración de procesos forman parte de la capa intermedia, mientras que la plataforma de BI es la encargada de brindar características básicas de seguridad y administración.

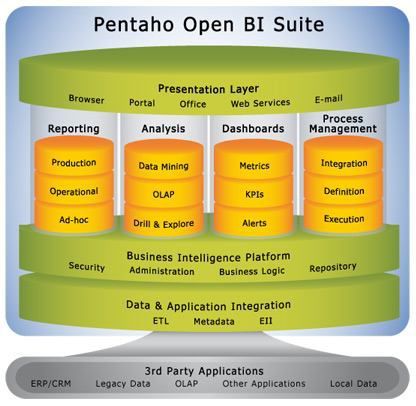


Figura 9: Componentes de Pentaho. Fuente: *Pentaho Open Source Business Intelligence*

Entre las principales herramientas que integranla Suite Pentaho ofrece una serie de productos como:

* *Mondrian (Open Source OLAP Server).*Servidor OLAP.
* *JFreeReport (Open Source Reporting).-* Herramientas de Reportes.
* *Kettle (Open Source Data* Integration).-Integración de Datos.
* *Weka (Open Source Data Mining).*Minería de Datos.

Debido a su completa gama de componentes que permiten desde la integración de datos, procesos ETL, hasta el diseño e implementación de análisis OLAP. La suite de Pentaho es la herramienta seleccionada para abordar el presente proyecto.

* 1. Recursos Disponibles

A continuación se describe brevemente el origen de los datos, siendo el mismo desarrollado en detalle en la sección 5.3.1.

Si bien en el Departamento General Manuel Belgrano, las autoridades de los municipios accedieron a posibilidad de implementar un sistema de información que les permitiría registrar y con ello centralizar la información referente a la producción (Godoy y Otros, 2012), su puesta en produccion aun no ha sido concretada por indoles propias de cada municipio.

Debido a que el presente trabajo nace a fin de complementar el sistema de información con herramientas de inteligencia de negocios que permitan facilitar el analisis de sus datos, se empleará la estructura de la base de datos del OLTP como fuente para la extracción de datos. La información utilizada fue relevada a fines de investigación y vinculación tecnlógica en el marco del desarrollo local de la zona de estudio. Se emplearon todos los datos de los productores obtenidasdel censo tabacalero del año 2013 (CT’13)del Ministerio del Agro e información obtenida de las distintas actividades de relevamiento mediante entrevistas y observacion.

Se integraron los datos procedentes de fuentes de información primarias y secundarias para conformar los datos,hasta que el OLTP comience a generar los propios. La información primaria se logro a través de entrevistas semi estructuradas, profundizadas en los puntos de interes identificados y los hechos cuya información era parcial o completamente desconocida, estas entrevistas se realizaron a informantes claves(autoridades de municipios, secretarias de agro, especialistas en el analisis de producción agropecuaria) y productores de la zona presentes en cada reunión.

Del análisis de los datos del CT’13, contenidos en una hoja de cálculo, se extrajo información referente a datos de identificación de cada productor, los productos que estos producen, la extensión territorial destinada a la actividad. Toda esta informaciónpertenece al registroanual, puntualmente del año 2013,por lo cual, la variable tiempo extraída de esta fuente se limita a dicha fecha.

De las entrevistas realizadas algunos temas de interés resaltantes se mencionaba, la necesidad de conocer la evolución de cada actividad agropecuaria, tanto por volúmenes como en zonas y años; permitir conocer la rentabilidad de cada producción, la situación social económica y educativa, entre otros. Temas estos cuyos datos se encontraban relevados en el censo.

Es debido a esta falta de información que se emplearon estrategias para completar la base de datos OLTP con información del CT’13 y estimaciones referente a los años 2012 y 2014. Dichas estimaciones fueron logradas en base al relevamiento bibliográfico de producción, precios y costos de los productos, de los informes publicados por entidades estatales como Ministerio del Agro y Producción, INTA, entre otros, entidades sindicalistas como UATRE y Notas de Diarios como LA NACION, entre otros. Los datos referidos a los preciosy los valores de inflación como y demás temas de interés con información obtenida de las entrevistas y del relevamiento de trabajos de investigación. Como ser volumen de producción, ingresos y costos de producción.

Capítulo 5

***Solución Propuesta***

Este capítulo se propone el desarrollo de la metodología seleccionada para la construcción del DW(sección 5.1) y el diseño de los cubos OLAP. Para ello se inicia con la etapa de análisis de los requerimientos (sección 5.2) dentro de la cual se identifican las necesidades de los usuarios (sección 5.2.1) en términos de identificadores y perspectivas (sección 5.2.2) que representaran el modelo conceptual(sección 5.2.3). En la siguiente sección (sección 5.3) se desarrolla la segunda etapa de Hefesto donde se analizaran los OLTP describiendo brevemente sus ventajas y limitaciones (sección 5.3.1). Luego se presentan los cálculos en base a los cuales se conformara cada uno de los indicadores (sección 5.3.2) y se establecen las correspondencias entre los elementos que componen el modelo conceptual con las del modelo de datos del sistema OLTP (sección 5.3.3), las cuales una vez establecidas se definenlos niveles de granularidad(sección 5.3.4). Al finalizar la segunda etapa se logra un modelo conceptual ampliado (sección 5.3.5) que sirve de base para el modelo lógico del DW(sección 5.4), cuyo desarrollo tiene lugar en la tercer etapa que inicia con la selección del tipo de esquema a implementar (sección 5.4.1) y continua con la definición de las tablas de dimensiones(sección 5.4.2), de las tablas de hechos (sección 5.4.3) y sus respectivas uniones(sección 5.4.4). La Integración de los Datos (sección 5.5) se trabaja en la última etapa de la propuesta de Bernabeu, en la cual establece los procesos para la carga Inicial de la DW (sección 5.5.1) y los necesarios para las posteriores actualizaciones (sección 5.5.2). Finalmente este capítulo cierra con la creación de los cubos multidimensionales para su posterior aplicación de OLAP (Sección 5.6).

1. 1. Modelo de Datos Multidimensional

Las cuatro etapaspropuestas por Hefesto son una alternativa relativamente nueva para la construcción de modelos multidimensionales. Sin embargo, su aplicación en el campo de BI es ampliamente aceptado, lo cual es en gran medida debido a que busca lograr una primera implementación que satisfaga parte de los requerimientos, mostrando con ello las ventajas de un modelo multidimensional y con ello motivar a los usuarios.

* 1. Análisis de Requerimientos

A través de entrevistas con especialistas en el análisis de producciónagropecuaria, Dra. Godoy e informantes claves dispuestos por las autoridades del municipio de Andresito se lograron identificar las siguientes necesidades de información.

* + 1. Identificar Preguntas
* ¿Cuál es el producto de cada zona que mayor rentabilidad ofrecea cada productor?

¿Cuál es el producto con mayor volumen de producción por productor de cada zona?

* ¿Cuál es la evolución de cada producción en la zona y quiénes son sus productores?
* ¿Existe relación entre el nivel de actividad agropecuaria y la dimensión de la superficie de cada unidad productiva de cada productor?
* ¿Cuál es la producción más elegida por los productores según las dimensiones de sus unidades productivas?
* ¿Cuál es el ratio de producción?
* ¿Cuál es el porcentaje, sobre el total de la superficie, ocioso de cada productor?
* ¿Cuál es la superficie ociosa de cada productor?
  + 1. Identificar indicadores (In) y perspectivas (Pe)
       1. Análisis de Rentabilidad del Productor

Total de Ingresos por cada producto de cada productor por zona y año

In Pe Pe Pe Pe

Total de Costos por cada producto de cada productor por zona y año

In Pe Pe Pe Pe

Volumen de Producción de cada producto por cada productor por zona y año

In Pe Pe Pe Pe

Superficie cultivada de cada producto por cada productor por zona y año

In Pe Pe Pe Pe

* + - 1. Análisis de Evolucióndela Producción

Mayor volumen de producción de cada producto por cada productor, zona y año

In Pe Pe Pe Pe

Mayor Superficie cultivada de cada producto por cada productor, zona y año

In Pe Pe Pe Pe

Porcentaje de la superficie sin cultivar que paso a cultivo por zona y año

In Pe Pe Pe

* + - 1. Análisis de Preferencias Productores

Cantidad de Productores que cultivan determinados productos por zona y año

In Pe Pe Pe

* + - 1. Análisis de Posibles Relaciones entre Tipo de Producción y UP

Porcentaje de superficie cultivada de la extensión total cada productor por zona y año

In Pe Pe Pe Pe

Mayor Ratio de producción de cada producto de cada productor por zona y año

In Pe Pe Pe Pe

Promedio de superficie de cada productor sin cultivar por zona y año

In Pe Pe Pe Pe

* + 1. Modelos Conceptuales
       1. Análisis de Evolución de la Producción

Del análisis realizado se ha logrado confeccionar el Modelo Conceptual (figura 10) referente al área de estudio Evolución de la Producción.

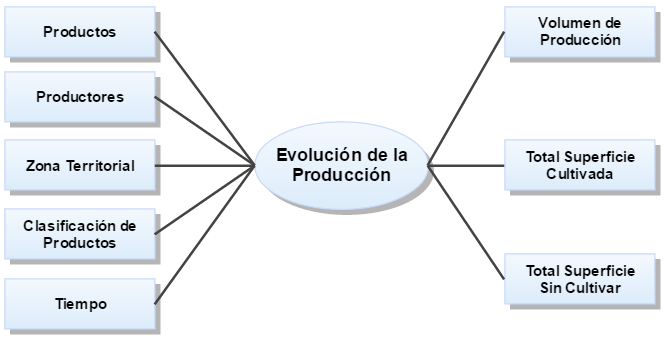


Figura 10: Modelo Coneptual Analisis de la Producicón

* + - 1. Análisis de Rentabilidad de la Producción

El siguiente Modelo Conceptual (figura 11)es referente al área de estudio Rentabilidad de la Producción. No obstante se incluirán al modelo de producción(figura 10) por emplear estos las mismas medidas, y según los usuarios seria de ayuda cruzar estos indicadores con las hectáreas producidas y volumen de producción a fin de ampliar las posibilidades de análisis.

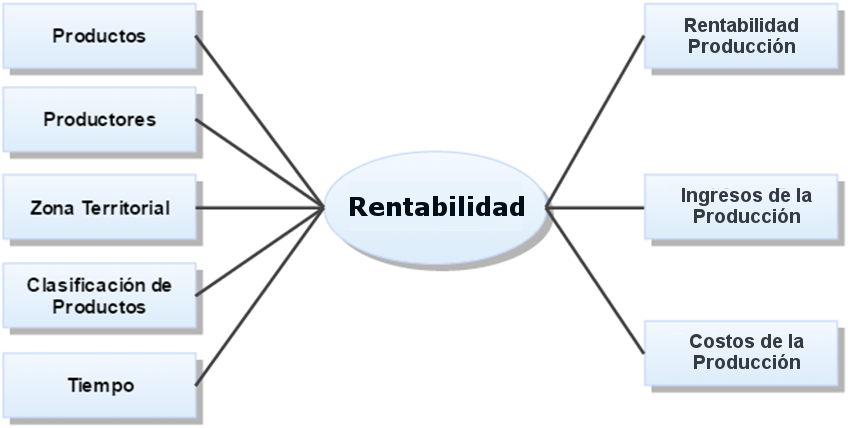


Figura 11:Modelo Coneptual Analisis de la Rentabilidad

* 1. Análisis de los OLTP

A continuación se establecerán las correspondencias entre el OLTP que registra la información de la producción y el modelo conceptual de la producción(figura 10) y rentabilidad (figura 11) a fin de determinar los datos que integraran el cálculo que poblara los indicadores.

* + 1. OLTP Productores

Actualmente los municipios del Departamento Gral. Manuel Belgrano, mediante programas a cargo por quien dirige esta tesis, han logrado acceder a un sistema OLTP desarrollado empleando las herramientas *Open Source* Siu-Toba en su versión 2.5 y PostgreSQLv9.1. Dicha aplicación permitiríaa cada municipio mantener centralizada, estandarizada y ordenada, información referente a la producción, sus productores, las características de las unidades productivas, situación educativa, social y ambiental, como así también información sobre ayudas económicas y subsidios que el estado les haya otorgado, los puntos de comercialización, e inclusive sus costos e ingresos(Godoy y Otros, 2012). Sin embargo el mismo aún no se ha logrado implementar por cuestiones particulares de cada municipio.

A causa de falta de persistencia de datos, y con el fin de ofrecer a los municipios una solución integral que aproveche los beneficios del sistema OLTP desarrollado, se ha optado por emplear la estructura de la base de datos como fuente, a la cual se poblaran con datos obtenidos de fuentes primarias y secundarias, entre las cuales se encuentran el CT’13 en formato de hoja de cálculo. Este censo tiene información de todos los productores registrados en el Ministerio del Agro y la Producción de la Provincia de Misiones. A pesar de que dicho censo recauda datos de distintas índoles, se tomaron como datos de interés todos los que sirvan,según los expertos, de apoyo para el análisis de la producción.

Respecto a los procesos de integración de datos necesaria para poblar el OLTP desde el Excel, se empleó la herramienta *Data Integration* de Pentaho con *scripts*SQL[[9]](#footnote-10).

Debido a que el CT’13, presenta datos de tan solo el 2013, fueron necesarios recolección de datos tanto con entrevistas como búsquedas bibliográficas para poder ampliar dicha información a los años 2012 y 2014.

* + 1. Conformar Indicadores

Los indicadores son las medidas que componen la tabla de hechos, a continuación se establecerá como se deberán calcular cada uno de los indicadores y el conjunto de hechos afectados al mismo.

Hecho: Volumen de Producción.

* Función: SUM(Volumen de Producción).
* Aclaración: Representa la sumatoria de la producción que se han producido de un producto en particular. Medida en Kilogramos.

Hecho: Total Superficie Cultivada.

* Función: SUM(Total Superficie Cultivada).
* Aclaración: Representa la sumatoria de la superficie destinada al cultivo de un producto en particular. Medida en Hectáreas.

Hecho: Total Superficie sin Cultivar.

* Función: SUM (Total Superficie -Total Superficie Cultivada).
* Aclaración: Representa la diferencia entre la superficie total y la superficie cultivada de una unidad productiva en particular.Medida en moneda Pesos.

Hecho: Ingresos de la Producción.

* Función: SUM (Ingresos de la Producción).
* Aclaración: Representa la sumatoria de los ingresos obtenidos por cada producción. Medida en moneda Pesos.

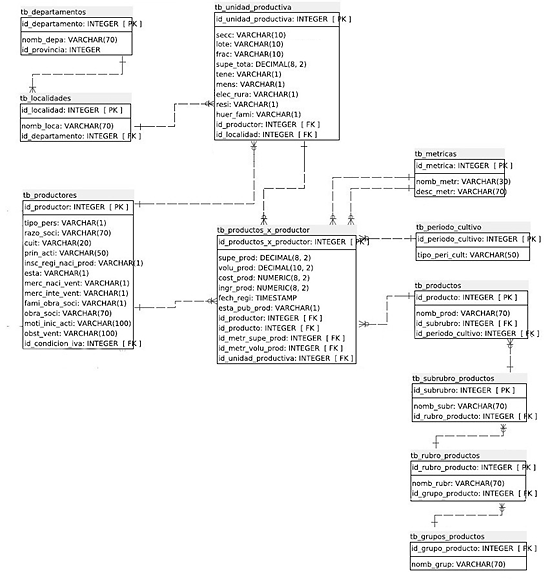
Hecho: Costos de la Producción.

* Función: SUM (Costos de la Producción).
* Aclaración: Representa la sumatoria de los costosabordadospara cada producción. Medida en moneda Pesos.

Hecho: Rentabilidad de la Producción.

* Función: SUM (Ingresos de la Producción – Costos de la Producción).
* Aclaración: Representa la diferencia entre los ingresos obtenidos y los costos abordados en cada producción. Medida en moneda Pesos.
  + 1. Establecer Correspondencias

Las correspondencias entre el OLTP y el modelo conceptual de la Producción (figura 11) se establecerán en base al siguiente modelo de datos, el cual presenta los componentes a utilizar del OLTP productores.

Figura 12: Elementos a Utilizar del Modelo E/R del Sistema OLTP

* + 1. Correspondencias fuente de Datos y Modelo Conceptual

Seguidamente se detallaran las correspondencias entre el OLTP y el modelo conceptual de la evolución de la Producción figura N a fin de determinar que cada elemento del modelo pueda ser referenciado con el OLTP. La figura N representa el proceso de producción del OLTP

|  |  |
| --- | --- |
| **Fuente OLTP** | **Perspectivas del Modelo** |
| **Tabla tb\_productos** | Productos |
| **Tabla tb\_productores** | Productores |
| **Tabla tb\_subrubro U tb\_rubro U tb\_grupo** | Clasificación de Productos |
| **Tabla tb\_Localidad U tb\_departamentos** | Zona Territorial |
| **Campo fech\_regi de la tabla tb\_productos\_x\_productor** | Tiempo |

Tabla 1: Correspondencia OLTP Productores y Perspectivas delModelo Conceptual

En cuanto a los indicadores se determinaron las siguientes correspondencias:

|  |  |  |  |
| --- | --- | --- | --- |
| **Campo** | **Tabla** | **Operación** | **Indicador** |
| **volu\_prod** | tb\_productos\_x\_productor | …… | Volumen de Producción |
| **supe\_prod** | tb\_productos\_x\_productor | …… | Superficie Cultivada |
| **supe\_tota** | tb\_unidad\_productiva | …… | Superficie Total de la UP |
| **supe\_sin\_cult** | tb\_productos\_x\_productor | Supe\_total-  Supe\_prod | Superficie sin Cultivar |
| **Ingr\_prod** | tb\_productos\_x\_productor | …… | Ingresos de la Producción |
| **cost\_prod** | tb\_productos\_x\_productor | …… | Costos de la Producción |
| **rent\_prod** | tb\_productos\_x\_productor | Ingr\_prod-  cost\_prod | Rentabilidad de la Producción |

Tabla 2: Correspondencia OLTP Productores e Indicadores del Modelo Conceptual

* + 1. Nivel de Granularidad

El nivel de granularidad determina el detalle con el cual se desea analizar un hecho, de las tablas referenciadas se analizaron los campos que conforman cada una de ellas y los posibles valores a integrar cada perspectiva.

Perspectiva Productos

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_producto** | Clave primaria de la tabla productos | Numérico |
| **nomb\_prod** | Nombre del producto de una actividad agropecuaria | Texto |
| **id\_subrubro** | Clave foránea que representa el subrubro al que pertenece cada producto -Vegetal, Legumbres, entre otros-. | Numérico |
| **id\_periodo\_cultivo** | Clave foránea que representa tiempo de inicio y fin del ciclo de producción de cada producto-mes a mes, mes año, año a año, entre otros-. | Numérico |

Tabla 3. Campos Candidatos Perspectiva Productos

Perspectiva Zona Territorial

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_localidad** | Clave primaria de la tabla localidades. | Numérico |
| **nomb\_loca** | Representa el nombre de la localidad. | Texto |
| **id\_departament** | Clave foránea que indica el departamento al que pertenece cada localidad. | Numérico |

Tabla 4: Campos Candidatos Perspectiva Zona Territorial

Perspectiva Productores

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_producto** | Clave primaria de la tabla productores | Numérico |
| **tipo\_pers** | Indica si un productor es una persona “F” Física o “J” Jurídica | Texto |
| **razo\_soci** | Es el nombre que representa a cada productor sea Física o Jurídica. | Texto |
| **cuit** | Numero de CUIT del Productor. | Texto |
| **prin\_acti** | Hace referencia a la producción que cada productor elige como su principal actividad. | Texto |
| **insc\_regi\_naci\_prod** | Si está o no inscripto en el registro nacional de productores. | Texto |
| **esta** | Si el Productor se encuentra en actividad o no. | Texto |
| **merc\_naci\_vent** | Si el Productor comercializa a nivel nacional. | Texto |
| **merc\_inte\_vent** | Si el Productor comercializa a nivel internacional. | Texto |
| **fami\_obra\_soci** | Si su familia posee obra social. | Texto |
| **obra\_soci** | El nombre de la obra social, si posee. | Texto |
| **moti\_inic\_acti** | Motivo que llevo al productor a formar parte de las actividades agropecuarias. | Texto |
| **obst\_vent** | Inconvenientes que el productor considere mencionar son problemas para comercializar sus productos. | Texto |
| **id\_condicion\_iva** | Clave foránea que referencia la condición de IVA Qué el productor posee. | Numérico |

Tabla 5: Campos Candidatos Perspectiva Productores

Perspectiva Clasificación de Productos

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_subrubro** | Clave primaria de la tabla Subrubros. | Numérico |
| **nomb\_subr** | Representa el nombre de la clasificación subrubro de cada producto. | Texto |
| **id\_rubro\_producto** | Clave foránea que indica el rubro al que pertenece cada subrubro de productos. | Numérico |

Tabla 6: Campos Candidatos Perspectiva Clasificación de Productos

Perspectiva Tiempo

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **fech\_regi** | De la tabla productos\_x\_productor, indica la fecha del momento en que se realizó el registro de los datos del productor y las actividades que este realizo durante el año | fecha |

Tabla 7: Campos Candidatos Perspectiva Tiempo

* + - 1. Diseño de las Perspectivas

Ya habiendo relevado y consultado con los expertos cada una de las alternativas admitidas de cada perspectiva para el análisis de los indicadores, se emplearan los siguientes datos de interés para formar parte de cada dimensión.

Diseño de laPerspectiva Productos

Las variables de interés seleccionadas para analizar los indicadores en base al producto son:

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **nomb\_prod** | Nombre del producto de una actividad agropecuaria | Texto |

Tabla 8: Diseño de laPerspectiva Productos

Diseño de laPerspectiva Categoría de Productos

Con el fin de permitir analizar los productos por categorías se propone la presente perspectiva compuesta por los siguientes campos

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **Subrubro** | De la tabla subrubros, este campo representa el nombre del subrubro al que pertenece cada producto- Verduras, Frutas, aves, entre otros-. | Texto |
| **Rubro** | De la tabla rubros a través de la tabla subrubro, este campo representa el nombre del rubro al que pertenece cada subrubro- Agricultura, Ganadería, entre otros-. | Texto |
| **Grupo** | De la tabla grupos, este campo representa el nombre del grupo al que pertenece cada rubro-Alimento, No alimento-. | Texto |

Tabla 9:Diseño de la Perspectiva categoría de Productos

Diseño de laPerspectiva Productores

En la siguiente tabla se exponen los campos seleccionados para analizar los indicadores en base a los productores.

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **razo\_soci** | Seleccionado para identificar a cada productor | Texto |
| **cuit** | Seleccionado para en caso de que la razón social se repita, poder distinguir un productor de otro. | Texto |

Tabla 10:Diseño de la Perspectiva categoría de Productos

Diseño de laPerspectiva Zona Territorial

La granularidad necesaria para analizar los indicadores por zonas requerida por los usuarios fue al nivel de departamentos y localidades. Por ello se propone una perspectiva que unifique estas dos variables de interés en una sola dimensión.

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **nomb\_loca** | Representa el nombre de la localidad. | Texto |
| **nomb\_depa** | De la tabla departamentos, este campo representa el nombre del departamento al que pertenece cada localidad. | Texto |

Tabla 11:Diseño de laPerspectiva categoría de Productos

Diseño de laPerspectiva Clasificación de Productos

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **nomb\_subr** | Nombre que identifica el subrubro de cada producto. | Texto |
| **nomb\_rubro** | De la tabla rubros, este campo representa el nombre del rubro al que pertenece cada subrubro. | Texto |
| **nomb\_grupo** | De la tabla grupos, este campo representa el nombre del grupo al que pertenece cada rubro. | Texto |

Tabla 12:Diseño de laPerspectiva categoría de Productos

Diseño de laPerspectiva Tiempo

Debido a que el registro de producción se releva anualmente, la perspectiva tiempo contendrá como variable el año y el mes del relevamiento.

* + 1. Modelo Conceptual Ampliado

De las perspectivas e indicadores obtenidos anteriormente se diseña el siguiente modelo

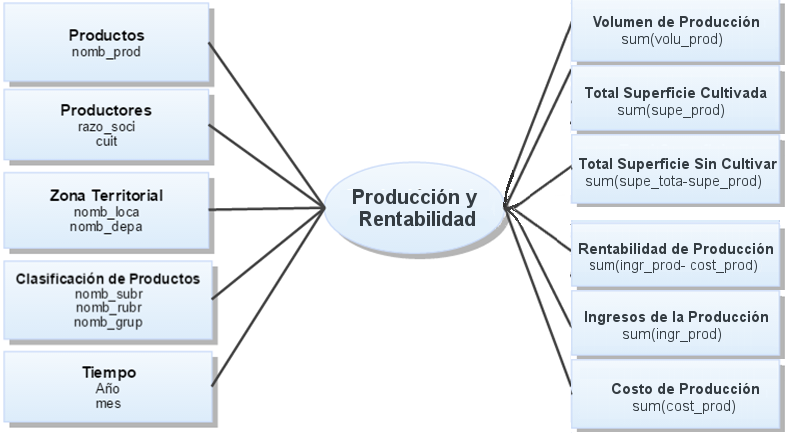


Figura 13:Modelo Conceptual Ampliado del Análisis de Producción y Rentabilidad

* 1. Modelo Lógico del DW

Ya confeccionado el modelo conceptual ampliado, se precede al diseño del modelo lógico de la estructura del DW. Se inicia con la selección del tipo de modelo lógico a emplear y se continúa con las acciones propias para el diseño de las tablas de dimensiones y hechos con sus correspondientes Uniones.

* + 1. Tipo de Modelo Lógico del DW

Para la selección del esquema a utilizar para soportar la estructura de la DW se ha tenido en cuenta que, si bien las perspectivas propuestas se podrían relacionar mejor a un esquema de copo de nieve, el sacrificio de ejecución de las uniones al momento de las consulta para sostener dicho esquema, en relación al ahorro de espacio por emplear un esquema jerárquico no es significante, por ello se opta por emplear este último.

* + 1. Tablas de Dimensiones

A continuación se presenta el diseño de las distintas tablas de dimensiones que formaran parte del DW.

Diseño de la Dimensión Productos

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_producto** | Clave que identificara unívocamente a cada producto, será una clave subrogada | Numérico |
| **producto** | Del campo nomb\_prod, se lo renombra para facilitar su interpretación. | Texto |

Tabla 13: Diseño de la Dimensión Productos

Dimensión Productores

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_productor** | Clave que identificara unívocamente a cada productor, será una clave subrogada | Numérico |
| **productor** | Del campo razo\_soci, se lo renombra para facilitar su interpretación. | Texto |
| **cuit** | Mantiene el mismo nombre, para en caso de ser necesario se diferencie un productor de otro con el mismo nombre | Texto |

Tabla 14: Diseño de la Dimensión Productores

Diseño de la Dimensión Zona Territorial

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_zona** | Clave que identificara unívocamente a cada zona, será una clave subrogada | Numérico |
| **localidad** | Del campo nomb\_loca, se lo renombra para facilitar su interpretación. | Texto |
| **departamento** | Del campo nomb\_depa, se lo renombra para facilitar su interpretación. | Texto |

Tabla 15: Diseño de la Dimensión Zona Territorial

Diseño de la Dimensión Clasificación Productos

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_clas\_prod** | Clave que identificara unívocamente la clasificación de cada producto, será una clave subrogada | Numérico |
| **subrubro** | Del campo nomb\_subr, se lo renombra para facilitar su interpretación. | Texto |
| **rubro** | Del campo nomb\_rubro, se lo renombra para facilitar su interpretación. | Texto |
| **grupo** | Del campo nomb\_rubro, se lo renombra para facilitar su interpretación. | Texto |

Tabla 16: Diseño de la Dimensión Clasificación Productos

Dimensión Tiempo

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_fecha** | Clave que identificara unívocamente cada fecha. | Fecha |
| **año** | Extraída del campo fech\_regi, solo contendrá el año. | Texto |
| **mes** | Extraída del campo fech\_regi, solo contendrá el mes. | Texto |

Tabla 17: Diseño de la Dimensión Tiempo

* + 1. Tablas de Hechos

Seguidamente se detallan las tablas de hechos, definido por las claves de cada dimensión y el conjunto de indicadores establecidos.

La tabla de hechos Producción según lo analizado y consensuado con los expertos estará dada de la siguiente forma.

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Tipo** |
| **id\_producto** | Clave foránea dimensión Producto. | Numérico |
| **id\_productor** | Clave foránea dimensión Productor. | Numérico |
| **id\_zona** | Clave foránea dimensión Zona Territorial. | Numérico |
| **id\_clas\_prod** | Clave foránea dimensión Clasificación Producto. | Numérico |
| **id\_fecha** | Clave foránea dimensión Tiempo. | Fecha |
| **ProducciónKg** | Del indicador volumen de producción, se lo renombra a fin de facilitar su comprensión. | Numérico |
| **Hectáreas Cultivadas** | Del indicador total superficie cultivada, se lo renombra a fin de facilitar su comprensión. | Numérico |
| **Hectáreas sin Cultivar** | Del indicador total superficie sin cultivar se lo renombra a fin de facilitar su comprensión. | Numérico |
| **IngresoProducción** | Del indicador ingreso de la producción, se lo renombra a fin de facilitar su comprensión. | Numérico |
| **CostoProducción** | Del indicador costo de la producción, se lo renombra a fin de facilitar su comprensión. | Numérico |
| **RentabilidadProducción** | Del indicador rentabilidad de la producción, se lo renombra a fin de facilitar su comprensión. | Numérico |

Tabla 18: Diseño de la Tabla de Hechos

* + 1. Uniones

Creadas las tablas de Dimensión y Hechos se realizan las uniones correspondientes obteniendo el siguiente grafico que representa el DW

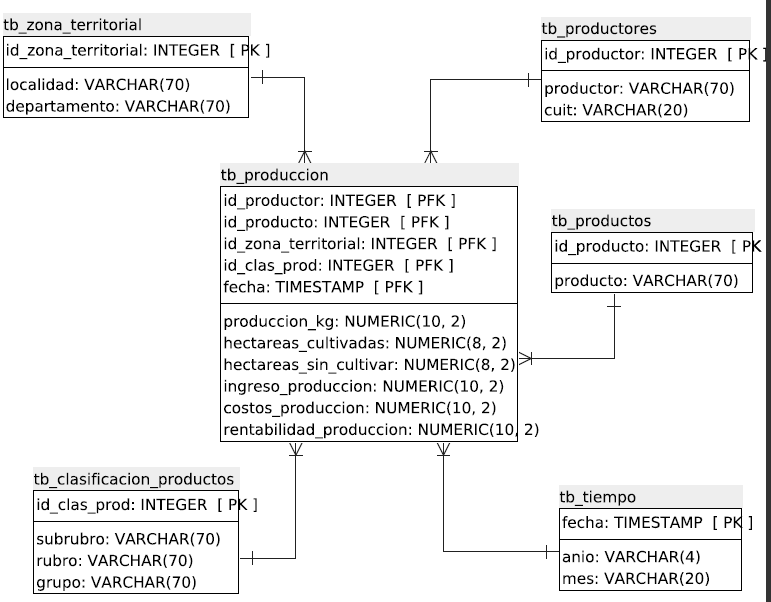


Figura 14: Modelo Multidimensional para el Análisis de la Producción

* 1. Integración de Datos

En esta etapa se procederá a poblar con datos la DW empleando las técnicas de limpieza y calidad de datos como procesos ETL, entre otros. Posteriormente se determinaran las políticas de actualización y los procesos para lograrlo.

* + 1. Carga Inicial

Ya habiendo generado la estructura del DW se procede a poblarlo con los datos de los OLTP. Para ello se proponen los siguientes procesos ETL implementados en la herramienta *data-integration[[10]](#footnote-11)* de la Suite Pentaho Solutions**.**Los procesos de extracción, transformación y carga pueden conllevar el 70% del esfuerzo en un proyecto de BI(Kimball y Otros, 2004).

En el Anexo 2 se presentan los procesos involucrados para la extracción, limpieza, carga del modelo multidimensional.

* + 1. Actualización

La estrategia de actualización de los datos del DW estipulado con los usuarios consistirá en el borrado de todos los datos de las tablas de dimensión Productores y tiempo, como así la tabla de hechos, los cualesse poblaran nuevamente mediante el trabajo empleado para la carga inicial.Esto es debido a que según lo analizado los cambios frecuentes son respecto a nuevos productores, y nuevas producciones. Mientras que las dimensiones de productos, zonas y clasificación por la naturaleza de su información, varían esporádicamente e inclusive podrían no hacerlo en años.

Esta tarea se realizará, en manera de prueba, por parte del analista, una vez definidos los periodos de necesidad por parte de los usuarios se establecerán las cargas automáticas en los periodos que se soliciten.

* 1. Creación de Cubos Multidimensionales

Para la creación de los Cubos se empleó la herramienta Mondrian Schema-Worckbench[[11]](#footnote-12) de la suite Pentaho. En ella se confecciona el esquema con los cubos, sus dimensiones y atributos como así también la tabla de hechos con los indicadores con la asignación de la función matemática definida anteriormente.

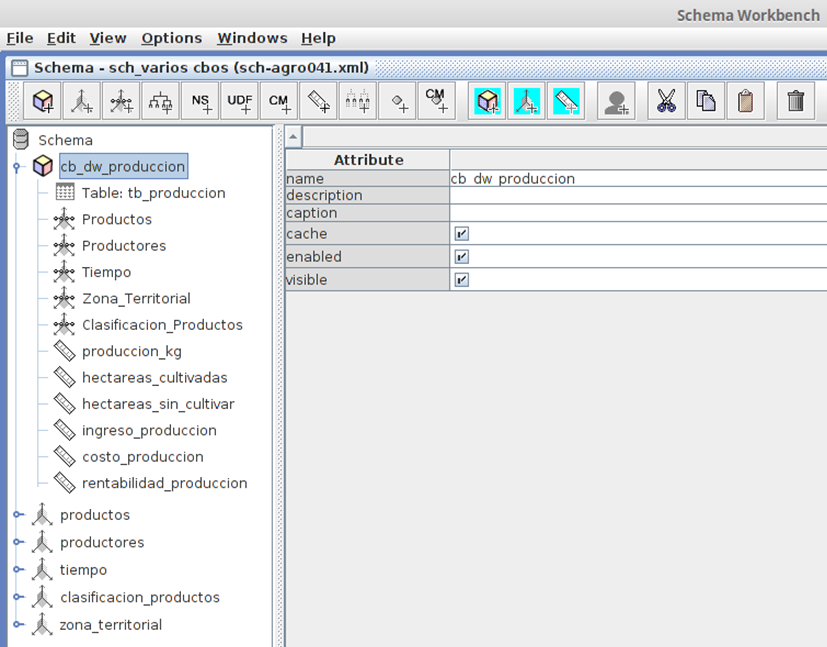


Figura 15: Cubo para el Análisis de Producción

La figura 15muestra la estructura del cubo para el análisis de Producción y rentabilidad, con las dimensiones y medidas definidas. Es en esta etapa donde se asigna a cada medida de la tabla de hechos la función que facilitará la interpretación de cada uno de ellos ante su cruzamiento con una o varias dimensiones. La función utilizada para todas las medidas será la de suma, excepto por la medida de ratio de producción, para la cual se empleara la de máximo.



Capítulo 6

Prueba Experimental



El siguiente capítulo muestra ejemplos de consultas al cubo de producción(Sección 6.1.1) y rentabilidad (Sección 6.2.1) con el fin de demostrar la utilización de la herramienta y las posibilidades del análisis multidimensional. En la primer consulta busca exponer un análisis de la evolución de la producción (sección 6.1.1) mediante la descomposición del mismo en dos puntos de vista. La primera busca analizar la producción por zonas (sección 6.1.1.1) y la segundase orienta a estudiar laparticipación de cada localidad frente a los productos con mayor evolución (sección 6.1.1.3). Finaliza el capítulo con una breve conclusión referente a la interpretación obtenida por los usuarios (Sección 6.3).

* 1. Análisis De Los Cubos

Para el análisis de los cubos se optó por utilizar la herramienta Saiku en la Suite Pentaho en lugar JPivot, la herramienta por defecto de Pentaho, por permitir esta una manipulación más amigable de las dimensiones y medidas mediante la interacción *drag and drop.*

* + 1. Análisis De LaEvolución De La Producción

Para el análisis de la evolución se propone la siguiente pregunta, que luego se la descompone en varias a fin de facilitar su posterior estudio.

Consulta 1:¿Cuál es la evolución de cada producción en la zona y quiénes son sus productores?

* Consulta 1.a:¿Cuál es la evolución de cada producción en la zona?
* Consulta1.b:¿De dónde son los productores de los productos con mejor evolución?
  + - 1. Análisis De La ProducciónPor Zonas

Conceptualización:

En este trabajo la evolución de la producción se entiende como el incremento desuperficie implantada y del volumen obtenido.

Dimensiones:

Departamento General Manuel Belgrano, todas sus localidades

Producto todos los productos

Tiempo Años 2012 a 2014

Medidas:

Produccion en Kg – Hectareas Cultivadas

Inicialmente como se muestra en la figura 16 se ingresan las medidas y las dimensiones comentadas en las filas y columnas del Saiku.



Figura 16: Selección de dimensiones y medidas para el análisis de la producción por zonas

De la ejecucion mediante estas medidasy dimensiones se obtiene como resultado lo presentado en la figura 17en la cual se pueden compararlos volumenesproducidospor cada producto y las extensiones destinadas para dicha produccioón durante losaños 2012 a 2014 de todo el departamento Gral. Manue Belgrano. Los valores se han distorcionado.

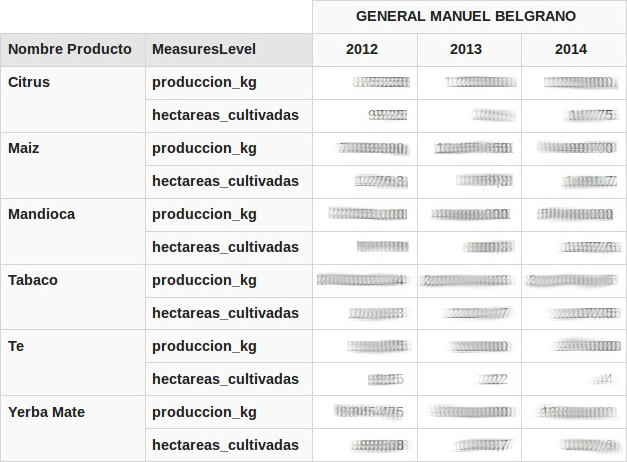


Figura 17: Total de kg producidos y hectáreas cultivadas de cada producto por año

A la actual consulta se la puedeaplicar un *drill down*con el fin de visualizar los volumenes y hectareas cultivadas por localidad como se expone en la figura 18, de esta forma se permite estudiar la evolucion que tuvo cada actividad productiva en cada una de las localidad que componen el departamento bajo estudio.



Figura 18:Total de kg producidos y hectáreas cultivadas en cada localidad

Representación Gráfica Hectáreas Cultivadas

Manteniendo el mismo punto de analisis, se puede representar su resultado en forma grafica, demostrando con ello la facilidad de uso y potencialidad de la herramienta. Es necesario definir cuales seran las categorias por las cuales se agrupara la medida en analisis a fin de obtener un grafico sencillo e intuitivo.Por ello, variando la posicion de las perspectivas entre filas y columnas, y seleccionado una de las dos medidas de analisis, ya que las mismas poseen metricas diferentres (figura 19).

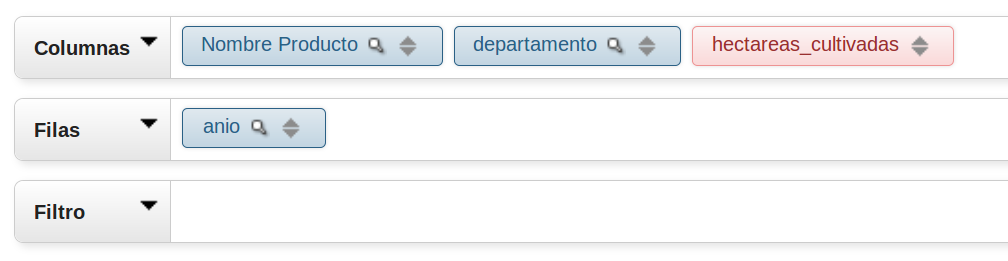


Figura 19: Selección de dimensiones y medidas para graficar la producción

En base a la medida y dimensiones seleccionadas. Se obtiene el grafico expuesto en la figura 20, en la cual se presenta de forma lineal la evolución, medida en hectareas cultivadas, que cada producto logro año a año en el departamento bajo estudio. Si bien en la figura no se pueden apreciar los valores, la herramienta ofrece la posibilidad de mediante el posicionamiento del puntero, ir visualizando los valores de cada linea.

No obstante, mediante ésta vista, puede evaluarse facilmente, que los valores que mayor crecimiento presentan, son la mandioca y el tabaco, esta primera con un crecimiento continuo en cada periodo partiendo de aproximadamente 500Ha en 2012 alcanzando las 1500Ha en 2014.

Mientras que el tabaco por su parte, presentó un notable crecimiento entre la campaña 2012/2013, pero que levemente cayó en la siguiente, perdiendo cerca de 80Ha de cultivo. Las extensiones destinadas al maiz y la yerba mate, se muestran con una importante suba entre el primer y segundo periodo. Sin embargo, caen fuertemente en el ultimo. Mientras que los citricos y el te, se mantuvieron según la grafica estables en cada año.

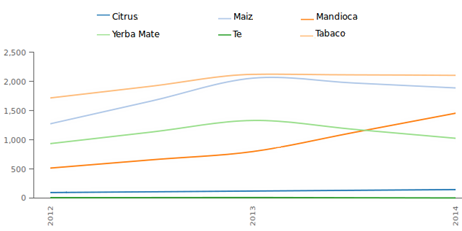


Figura 20: Total de Hectareas cultivadas por producto

Representación Gráfica Producción en Kilogramos

Ya habiendo graficado la evolucion según las hectareas, se podria proceder de la misma forma para el analisis gráfico en base al volumen de producción (figura 21). Variando unicamente la medida de Hectáreas Cultivadas por Produccion en Kg.

Según la figura 20la producción de la mandioca por hectáreases el producto con mayor evolución lo cual comose puede observar en la figura 21, se mantiene con respecto al volumen producido. Donde la mandioca crece practicamente con un valor constante partiendo de aproximadamente 23.000.000 kg en 2012, llegando cerca de los 60.000.000 Kg en 2014 lo cual se refleja con una linea practicamente constante en la gráfica.

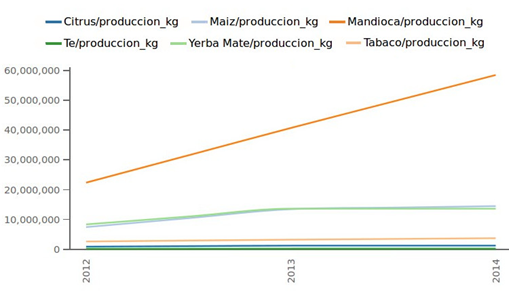


Figura 21: Total de kilogramos producidos por producto

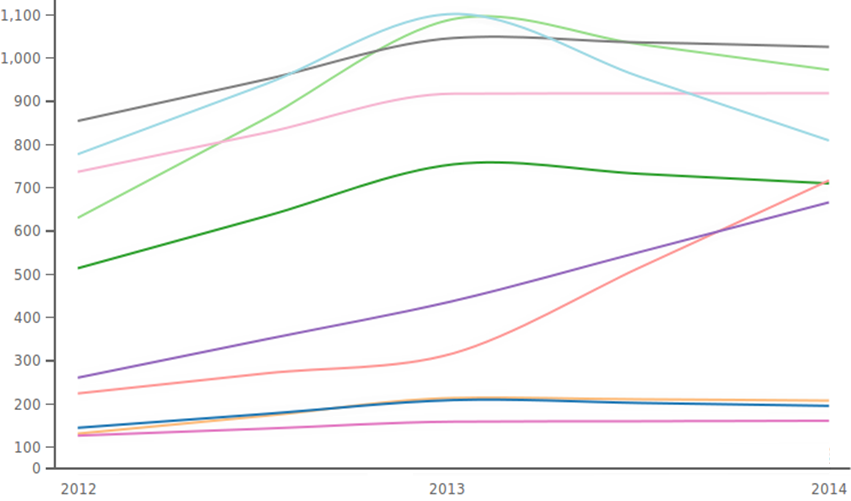
El tabaco que habia mostrado buen crecimiento según las hectáreas cultivadas (figura 20), analizado despues a partir mediante elvolumen de producción (figura 21) pareceria no haber sufrido cambios.

Esto se debe a la escala de valores utilizada, la cual varia enormemente entre los volumenes producido de cada producto impidiendo con ello lograr un analisis comparativo estas tecnicas gráficas. Solucion a esto seria quitar la mandioca del analisis a fin de utilizar una escala mas baja para la produccion de tabaco cuyos valores operan por debajo de los 10.000.000Kg.

Como se puede apreciar,las graficas muestran la evolucion de la producción en base a todo el departamento. Se podria incluir un nivel mayor de granularidad y mostrar la evolución lograda por cada localidad con respecto a cada producción (figura 22).

Donde las principales lineas son:

1. Ha de Tabaco en San Antonio
2. Ha de Yerba Mate en Cte. Andresito
3. Ha de Tabaco en Cte. Andresito
4. Ha de Maiz en San Antonio
5. Ha de Maiz en Cte. Andresito
6. Ha de Mandioca en San Antonio
7. Ha de Mandioca en Cte. Andresito



**4**

**7**

**6**

**3**

**2**

**5**

**1**

Figura 22: Total de kilogramos producidos por producto

* + - 1. Resultado del Análisis de la Producción

El analisis de la evolución del tabaco, mandioca, maiz, te y citrus, muestra que los mismostuvieron particulares situaciones, donde la produccion de la mandioca fue la que mayor comportamiento positivo obtuvo seguido del tabaco, el cual muestra un notorio crecimiento entre la campaña 2012/2013, pero al año 2014 se estabilió. La yerba mate y el Maiz fueron los productos que mas perdida presentaron en el año 2014, pese que en los primeros años mostraron buenas perspectivas. Los productos que se mantuvieron establescon leves oscilaciones en cada año, fueron los citricos y el té.

* + - 1. Participación de cada localidad en las Mayores Producciones

Consulta:

¿Cuál es la participación de cada localidad en laproducción?

Dimensiones:

Departamento General Manuel Belgrano, todas sus localidades

Producto Mandioca - tabaco

Productor productores según la zona y produccion

Tiempo Años 2013 a 2014

Medidas:

Produccion en Kg

Hectareas cultivadas

Como se observa en la figura 23 se emplea para este analisis el filtro en base a los productos tabaco y mandioca. Se emplearán tanto las medidas de producción en kg y hectáreas cultivadas como el filtro por un producto u otro en consultas distintas ya que si bien ambas son indicadores para analizar la producción, estas operan en metricas distintas que dificultarían suinterpretación gráfica.



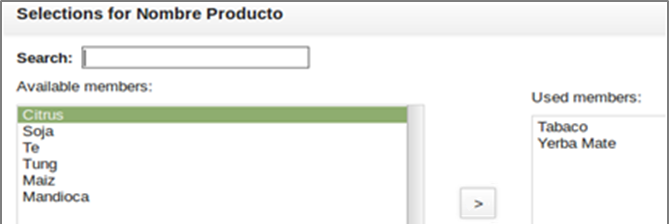


Figura 23: Filtro para el análisis de cada producto

Mayores Productores de Tabaco

De la selección de dimensiones y medidas propuesta se obtiene el siguiente set de gráficos (figura 24) en los que se observa el porcentaje de participación que tuvo cada localidad, en base a las hectáreas destinadas a la producción del tabaco en cada año.

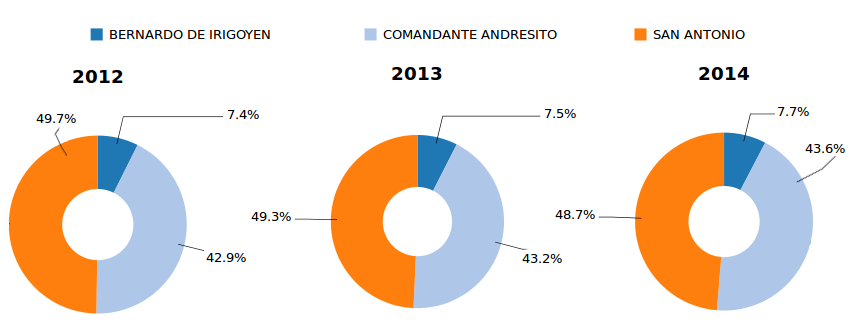


Figura 24: Participación de cada localidad en base a las hectáreas de tabaco cultivadas

Como se observa en la figura 24 los productores de San antonio fueron quienes mayor participación tuvieron en la producción tabacalera con casi la mitad del total de participacion sobre esta actividad. Comandante andresito no se encuentra tan lejano ya que mantiene un porcentaje del 40% del total de tierras destinadas a este cultivo.

Mayores Productores de Mandioca

Manteniendo el punto de análisis propuesto para anteriormente, variando únicamente el producto filtrado, se logra el siguiente set de gráficos (figura 25) que exponen la participación que cada localidad logro en cada año con respecto a la producción de mandioca.

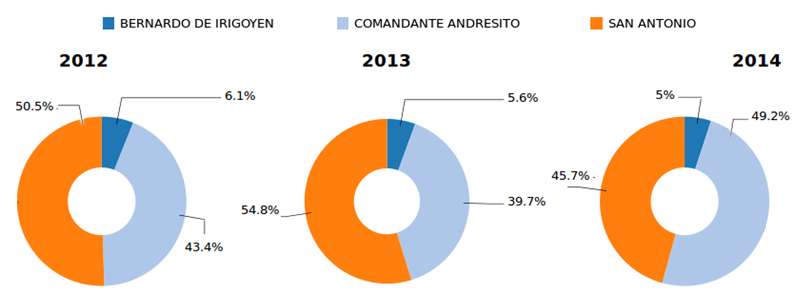


Figura 25: Participación de cada localidad en base a las hectáreas de mandioca cultivadas

Estos gráficos muestran que Comandante Andresito logro en 2014 liderar la producción de mandioca. Una importante aclaración de los especialistas es que en cada analisis se lo ve al municipio de Bdo. De Irigoyen muy por debajo en el desarrollo de estas actividades. Lo cual se debe a que el suelo de esta localidad es de composicion mas rocosa, lo que lleva a los productores a optar por actividades como la ganadería, entre otras, en logar de estos cultivos.

La sigiuente grafica representa todos los productores de mandioca por municipios, como se puede apreciar los productores de san Antonio a pesar de tener en el año inicio una produccion superior, año a año se reconoce fueron ampliando sus niveles de produccion ya sea en gran medida entre los años 2012 a 2013 y en menor medida entre 2013 y 2014

* + - 1. Resultados del Analisis de Participación

Los productores cuya producción presenta incrementos continuos en las areas de tabaco, forman parte mayormente del municipio de San Antonio seguido de Comandante andresito con producciones en kg similares.Mientras que en la producción de mandioca los productores de Comandante Andresitoson quienes mayor participacion han tenido. Por su parte Bdo. de Irigoyen por las caracteristicas de su suelo no frecuentan estas actividades, por lo cual sus valores con respecto a estos cultivos se visualizan siempre por debajo de los demas municipios.

* + 1. Análisis de laRentabilidad de la Producción

**Consulta:**

¿Cuál es el producto de cada zona que mayor rentabilidad ofrece a cada productor?

Dimensiones:

Departamento General Manuel Belgrano, todas sus localidades

Producto Mandioca - Tabaco

Tiempo Años 2012 a 2014

Medidas:

Producción en Kg

Hectáreas cultivadas

Rentabilidad producción

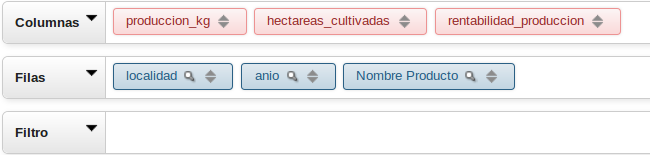


Figura 26: Selección de medidas y dimensiones para el analisis de rentabilidad

En base a la selección de la consulta realizada en la figura 26 y debido a que aún los municipios no han logrado generar registros referentes a los ingresos y costos de la producción, se ha podido realizar el analisis de la rentabilidad en base a dos productos, la mandioca y el tabaco, datos obtenidos de la empiria del articulo elaborado por Godoy y Otros (2015). En la figura 27 se visualizan los valores adquiridos por los productores de cada municipio por su produccion en cada año.

Se han incluido en el analisis las medidas de produccion en Kg y las hectareas cultivadas, debido a que según los especialistas se emplea la superficie como variable de analisispara determinar los costos de insumos y mano de obra, siendo el valor de produccion en Kg una variable que acompaña al ingreso, y en base a cruce de datos se realizan las tomasde disiciones referidas a la designación de tierras para cierto cultivo y la productividad lograda.

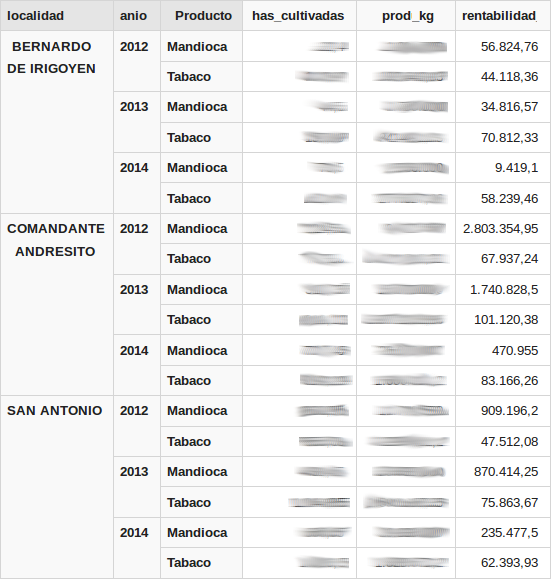


Figura 27: Rentabilidad de la producción de cada localidad por año

En la figura 27 se muestra la produccion obtenida por año, las superficies destinadas y la rentabilidad que alcanzada para la mandioca y el tabaco. De su análisis se aprecia que durante los tres periodos la rentabilidad de la mandioca fue muy superior a la obtenida por la producción de tabaco. Sin embargo, en cada año este producto fue decreciendo sus ganancias. Mientras que el tabaco desde 2012 a 2013 las aumento en un 60% aproximadamente y al llegar al 2014 bajo al igual que la mandioca.

A continuacion se presentan en tres gráficas la proyección de la rentabilidad obtenida por cada localidad en base a la producción de Tabaco y Mandioca.

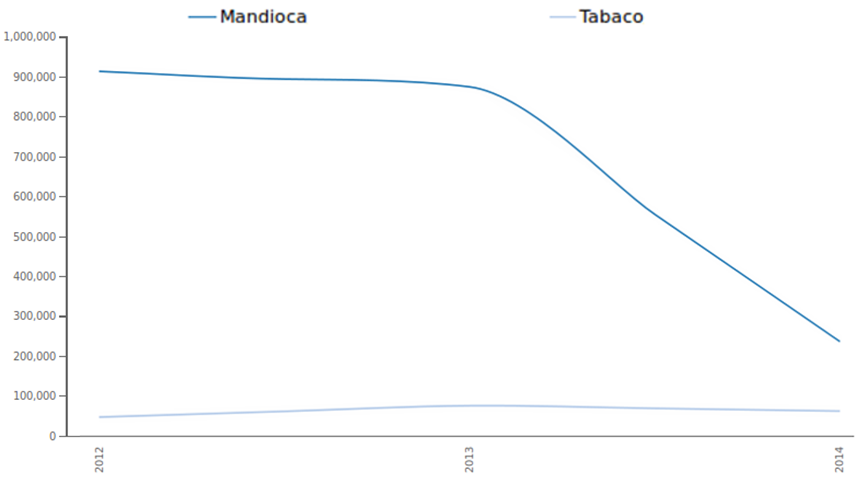


Figura 28: Proyección de rentabilidad de la producción en San Antonio

En la figura 28 se muestra el comportamiento que tuvo la medida rentabilidad en la localidad de San Antonio. Como se puede apreciar la grafica se corresponde con lo expuesto en las la figura 27, donde la produccion de mandioca cae continuamente pero de forma mas abrupta en a campaña del 2014, mientras que el tabaco muestra una leve mejora entre 2012 y 2013 pero decrece levemente al 2014.

Similar es la situacion que muestra la figura 29 referente a la rentabilidad en la localidad de Comandante Andresito. El comportamiento decreciente de la mandioca en esta localidad se muestra con mayor notoriedad, mientras que el tabaco muestra como en San Antonio incrementos entre 2012 y 2013, pero difiere de este en que logra estabilizarse en 2014.

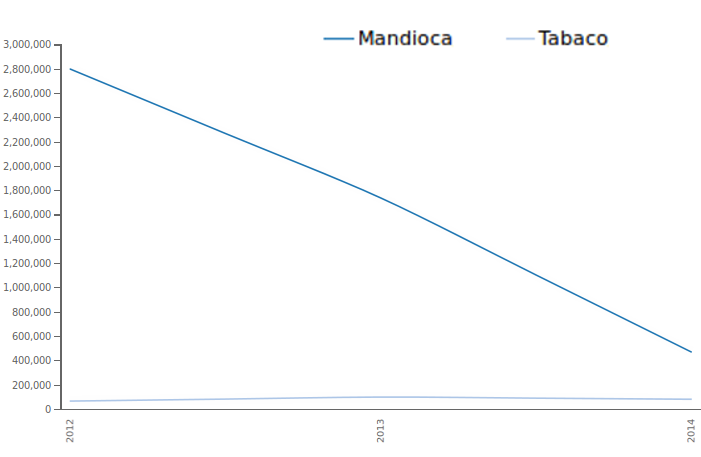


Figura 29: Proyección de rentabilidad de la producción en Cmte. Andresito.

La gráfica que representa la proyección de rentabilidad de Bernardo De Irigoyen (figura 30), muestra respecto de la mandioca un comportamiento similar al generado para Comandante Andresito, con una caída constante y abrupta desde el primer periodo. Por otro lado, el tabaco es más análogo al comportamiento observado en la localidad de San Antonio, con buenas proyecciones entre los dos primeros años, concluyendo con una notoria caída en el 2014.

A continuación se expone la gráfica de rentabilidad correspondiente a la localidad de Bernardo de Irigoyen

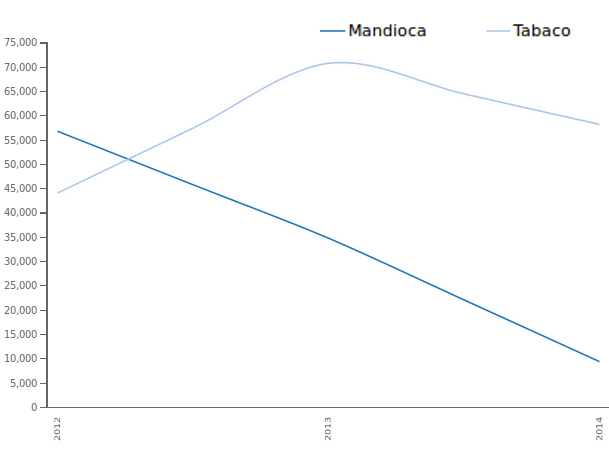


Figura 30: Proyección de rentabilidad de la producción en Bdo. De Irigoyen

* + 1. Resultado del Análisis de la Rentabilidad

Según especialistas, esto pudo deberse a que en el caso de la mandioca, se vivio un 2012 con una gran demanda que elevo en gran medida los precios de comercialización mientras que en sus siguientes años la inclusion de nuevos productores en 2013 bajo en aproximadamente la mitad de lo que se ofrecia en el periodo anterior. Empeorando su situación, en la campaña 2014, la aplicación de políticas de importación de productos derivados de la mandioca, llevaron a los productores a competir con precios inferiores al del mercado, que sumandos al aumento de los insumos por la de inflación del momento y al incremento de los gastos de mano de obra, se explica esta abrupta caida de valores.

En cuanto al tabaco su baja de rentabilidad en el ultimo periodo explican los expertos que puede deberse a los efectos causados por la inflacion y aumento de la cotización del dólar. Lo cual afeta directamente al costo de los insumos por ser estos adquiridos al valor de dicha moneda.

* 1. Conclusiones e interpretaciones de los usuarios

Debido a que los datos fueron calculados a partir de informacion de fuente primaria y secundaria, conlleva a un cierto grado de incertidumbre y éstos son estimados. Sin embargo, los valores a grandes razgos responden a la realidad de la poblacion bajo estudio.

Cabe aclarar que para el objeto de este trabajo y a efectos de visualizar un mayor analisis, en este caso se dejo a un plano relegado la importancia de la validez de los datos. Pero de todas maneras, una vez realizado el analisis con la informacion cargada, en cierta forma al abordar el problema se logro asimilar con naturalidad la concepcion ficticia de los datos, y se logrocomprender el comportamiento presentado por los mismos, acompanando medidas de acciones posibles.

Capítulo 7

Conclusiones

* 1. Conclusiones

Con el presente trabajo se ha podido construir un modelo multidimensional, el cual a través de cubos OLAP, se ha logrado explotar su contenido, medianteanálisis en múltiples perspectivas que permitieron interpretar la situación que atraviesan los sectores productivos, identificar sus características y delinear acciones de nivel gerencial para administrar de forma correctasus recursos (ayudas económicas, subsidios, créditos) y explotar con inteligencia las capacidades productivas de cada municipiobuscando el accesoa nuevos mercados y por sobre todo a orientar a cada productor en actividades que le permitan mejorar su calidad de vida.

En su desarrollo se han logrado identificar las necesidades recurrentes en cuanto a información de interés, como así el gran déficit en la recolección y centralización de la información de producción y sus actores.

En cuanto a los datos utilizados para poblar la DW, muchos de ellos basados en experiencias de los expertos, se intentó continuamente establecer las validaciones necesarias a fin de consolidar que el sistema OLTP empleado actualmente, una vez cargados con los datos del departamento, permita facilitar cualquier actividad de implementación de los cubos OLAP sobre dicho sistema.

Con las bases establecidas en este trabajo, y con la continua predisposición de las autoridades municipales se espera, una vez que logre poblar sus fuentes de información, puedan utilizar estas herramientas que les servirán de soporte para interpretar la situación que atraviesan sus productores y tomar las decisiones en favor de su comunidad.

* 1. Futuras líneas de investigación

La solución propuesta abarca particularmente a procesos de decisión referidas al análisis de producción, futuros trabajos deberían enfocarse no solo a refinar este modelo sino a implementar soluciones en cuanto a las condiciones sociales, económicas y educativas, datos estos, soportadas por el sistema OLTP, que aún no ha sido poblado con dichos valores.

Asimismo fomentar la utilización de estas herramientas, y particularmente este enfoque en otros departamentos y regiones de la provincia, que buscan mejorar la administración de sus recursos en pos de un correcto uso de sus suelos y sus productores

ANEXOS

Anexo 1

**Fuente de Datos OLTP**

Las actividades necesarias para poblar la BD de productores construida en postgreSQL v9.1 fueron logradas mediante la combinación de transformaciones diseñadas y ejecutadas con la herramienta PDI, permitieron exportar los datos del CT’13 y la posterior carga de los datos de producción, sus productores y las unidades productivas en el OLTP. Si bien hubiese sido posible extraer información referente a la educación de la familia del productor, situación de vivienda y equipamiento para la producción, estos no fueron incluidos en el análisis debido a que motivos que se explican a lo largo de este anexo.

**Limitantes de las fuentes de datos**

Entre las limitantes del CT’13 se tienen dos principales, el periodo de registro que se basa en tan solo el año 2013 y la falta de información referente al volumen producido, costos abordados e ingresos obtenidos.

**Periodo de Registro**

Debido a que los registros del CT’13 se corresponden únicamente al año 2013, se debió realizar un análisis de las fuentes bibliográficas tanto informes y publicaciones emitidas por entidades del agro y producción de la provincia, como así también, diarios y artículos a fin de conocer, a grandes rasgos, la situación que atravesaba el desarrollo de las distintas actividades relevadas en el CT’13, tanto desde el análisis de producción como de los beneficios y costos que implicaba abordar cada una de ellas.

**Fuentes secundarias**

De lo relevado se obtuvo un panorama en donde las producciones fueron en suba desde 2012 a 2014 para la Yerba Mate[[12]](#footnote-13), mandioca[[13]](#footnote-14), té[[14]](#footnote-15), tabaco[[15]](#footnote-16), mientras que el maíz aunque se proyectaba con gran crecimiento, vario levemente maíz[[16]](#footnote-17). Los productos que más decayeron fueron los cítricos por plagas que afectaron la provincia[[17]](#footnote-18).

De lo expuesto anteriormente se realizaron entrevistas telefónicas y vía e-mail con informantes claves (expertos, autoridades de los municipios, productores), quienes ajustaron los resultados obtenidos del relevamiento. En algunos casos las variaciones se ajustan a la realidad vivida por el departamento Gral. Manuel Belgrano.

En relación a la situación de cada producción en los años posterior y siguiente al del CT’13, existen diferencias en productos como la yerba mate y té, las cuales lograron un gran volumen en 2012 pero grandes caídas en los años 2013 y 2014, mientras que los cítricos mantienen lo analizado en cuanto a pérdidas durante cada campaña. En cuanto a los productos que mayor crecimiento de producción obtuvieron, tomando en cuenta los productos registrados en el CT’13, se encuentran la mandioca, el tabaco y el maíz.

De la segunda limitación, se encontró que los datos necesarios para el análisis de volumen, costos e ingresos eran totalmente inexistentes. Por ello se investigó en publicaciones tanto de entidades estatales como del Ministerio del Agro y la producción de la provincia de misiones[[18]](#footnote-19), del INTA[[19]](#footnote-20), entre otros, complementados con artículos de periódicos nacionales y provinciales[[20]](#footnote-21). Todo esto con el fin de completar los campos de los años 2012 y 2014.,datos que ayuden a comprender la situación enfrentada por los productores respecto a los costes afrontados e ingresos obtenidos. No obstante, debido a la diversificación de las zonas y situación de cada productor, esta información fue escasa donde información hallada se limitó únicamente a la producción de Mandioca y Tabaco.

**Volumen de producción**

En base al CT’13 pose información de las Hectáreas (Ha) que cada productor cultivo de: Mandioca, Tabaco, Te, Tung, Maíz, Yerba Mate y cítricos en general. Según información obtenida de fuentes del Municipio de Andresito, se estableció que en promedio de cada producto se producen:

* Mandioca: 40.000 Kg por Ha.
* Tabaco: 1517 Kg por Ha.
* Maíz: entre 4000 a 9000 Kg por Ha. Por recomendación del experto se definió, 6.500kg por Ha.
* Yerba Mate: entre 8000 a 12000 Kg por Ha. Por recomendación del experto se definió, 1000kg por Ha.
* Té: entre 15.000 a 20.000 Kg por Ha. Por recomendación del experto se definió, 17.500kg por Ha.
* Citrus: 10.000 Kg por Ha.

En base a estos valores, el cálculo para el volumen de producción de cada producto estará dado por:

Volumen de Producción en Kg = Hectáreas Cultivadas \* Kilogramos por Hectárea

Estos valores fueron aplicados tanto para el año 2013 como para los demás años que constituyen el análisis. Mientras que para la producción de Soja y Tung, no se pudieron obtener los valores para el cálculo, por lo que no fueron tomados en cuenta para integrar la BD.

El volumen de producción es una variable influenciada por la cantidad de Hectáreas cultivadas y por tanto por la cantidad de productores, y este último según fuentes del municipio y expertos. En el caso de la mandioca el número de productores fue creciendo entre los años 2012 al 2014 mientras que el tabaco aunque su comportamiento fue similar, sus cantidades fueron menores.

**Costos e Ingresos de la producción**

En el artículo publicado por Godoy y Otros (2015), se define lo siguientepara el cálculo de rentabilidad (costo e ingresos).

**Calculo para el Costo de la producción**

El cálculo para el costo los productos mandioca y tabaco, se definen de la siguiente forma:

Costo = (Val\_Jor\*Can\_Jor+ Cos\_Ins) \*Vol\_Pro

Donde

Val\_Jor = Valordel Jornal.

Can\_Jor = Cantidad de jornales

Cos\_Ins = Costo de Insumos

Vol\_Pro = Volumen de Producido en Kg

**Cálculo para el Ingreso de la Producción**

Los cálculos para obtenerel ingresoobtenidopor cada producto están dados según lo siguiente:

Ingresos de la mandioca = ( Pre\_Kg\*Vol\_Pro )

Ingreso del tabaco = ( Pre\_Kg\*Vol\_Pro )**+**( Vol\_Pro \* Pre\_Kg ) \* Pre\_FET

Donde

Pre\_Kg = Precio por Kilogramo

Vol\_Pro = Volumen Producido en Kg

Pre\_FET = Subsidio ofrecido a los tabacaleros del FET

**Rentabilidad de la Producción**

La rentabilidad está dada por el la diferencia entre los ingresos y costos de producción. Los valores para cada calculo fueron definidos en base a lo expuesto por los expertos y validada por publicaciones de las entidades reguladoras. En el siguiente cuadro se presentan los valores correspondientes a cada añopara el cálculo de los valores faltantes.

Cuadro de valores para el cálculode rentabilidad para laproducción de mandioca

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2012** | **2013** | **2014** |
| **Kilogramos por Hectárea** | 40.000 | | |
| **Valor del Jornal** | $ 145,15 | $157,33 | $205,23 |
| **Cantidad de jornales** | 31 | 31 | 40 |
| **Costo de Insumos por Hectárea22** | $ 234,00 | $ 300,00 | $ 371,70 |
| **Precio Kilogramos** | $ 2,30 | $ 1,00 | $ 0,45 |

Tabla 19: Valores para el cálculo de rentabilidad para la producción de mandioca

Cuadro de valores para el cálculo de producción de Tabaco

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2012** | **2013** | **2014** |
| **Kilogramos por Hectárea** | $ 1.517,00 | | |
| **Valor del Jornal** | $ 145,15 | $157,33 | $205,23 |
| **Cantidad de jornales** | 100 | | |
| **Costo de Insumos por Hectárea** | $ 5.785,00 | $ 7.250,00 | $ 8982,75 |
| **Precio Kilogramos** | $ 10,75 | $ 12,78 | $ 12,78 |
| **Ingreso Directo del FET** | 66,90% del valor de Ingreso | | |

Tabla 20:Valores para el cálculo de rentabilidad para la producción del tabaco

Los valores utilizados para el año 2013 fueron obtenidos del artículo publicado por Godoy y Otros (2015), mientras que para el año anterior y siguiente a este, se debieron combinar distintas fuentesa fin de establecer la mayor correspondencia posiblea la situación del momento.

En el año 2012 para la mandioca y el tabaco según fuentes de la municipalidad de Andresito no hubieron cambios en los Kg/Ha, en el Valor del FET, como así tampoco en la cantidad de jornales necesarios en cada actividad.

Con respecto al valor del Jornal, se tomó el establecido por la CNTA[[21]](#footnote-22) en cada año. Los insumos que adquiere el productor para el cuidado de la producción, para establecer sus costos, al estar valuados en dólares, se tomó la cotización del momento donde en el 2012 el dólar valuó a US$5,11[[22]](#footnote-23)dejando en $ 234,00 los insumos para la mandioca y $ 5.785,00 para el tabaco. Por su parte en el 2014 la moneda extranjera cotizo a US$7,14[[23]](#footnote-24), fijando a $ 371,70los insumos para la mandioca y $ 8.982,75 para el tabaco.

Con respecto a los precios por kg de cada producto se obtuvieron en base a periódicos de la zona, entre los cuales misiones online público en marzo del 2012 el anuncio del Gobierno que fijo el precio del tabaco a $10,50 para ese año[[24]](#footnote-25), mientras que para el 2014 se mantuvo en $12,78. La mandioca por su parte fue variando en cada año iniciando con $2,30 el Kg, bajando a $1,00 en 2013 y volviendo a caer en 2014 a un precio de $0,45[[25]](#footnote-26). En este último año la mandioca además de bajar su precio incremento sus costos de mano de obra requiriendo 40 jornales22.

**Carga del OLTP**

Los procesos para la adaptación de los valores se realizaron con cursores implementados en el gestor de BD, en su mayoría para corregir los errores en los nombres de los productores, de localidades y departamentos, como los nombres de los productos y las superficies destinadas a cada actividad. Respecto a las tablas auxiliares, se emplearon sentencias SQL específicas a cada caso, entre los que implicaron carga de Países, Provincias, Métricas de Volumen y Superficies, entre otros.

Anexo 2

El siguiente anexo contiene los procesos ETL utilizados para la solución propuesta mencionada en la sección 5.5.1.

La carga del cubo Producción y Rentabilidad se lleva a cabo mediante el Archivo job\_cbo1\_gral, el cual procede a la carga de cada una de las dimensiones y hechos que componen el cubo.

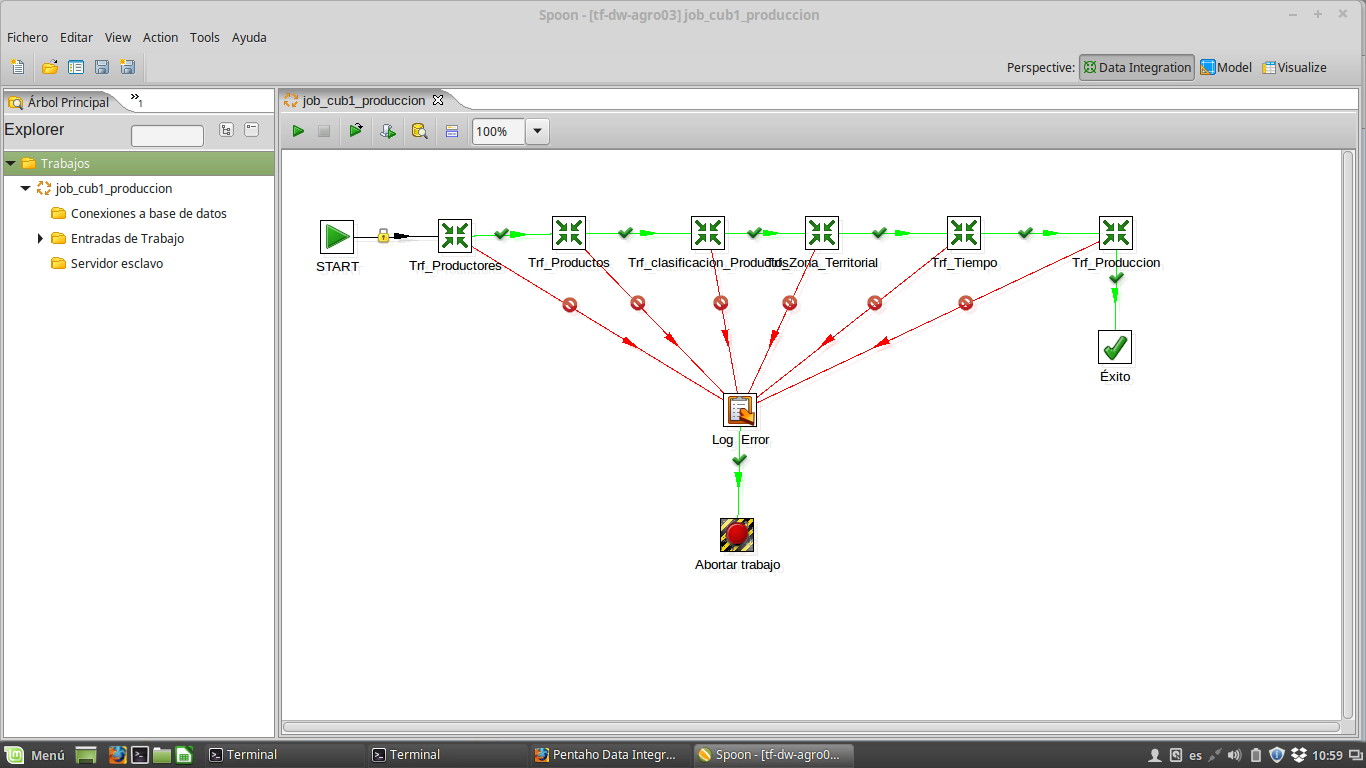


Figura 31: Trabajo para la carga del cubo para el análisis de la producción y rentabilidad

Con el objetivo de automatizar la carga de las tablas de dimensiones yhechos para el cubo Producción y rentabilidad, se diseñó el trabajo expresado en la figura31.

Transformación para la Carga de Productores

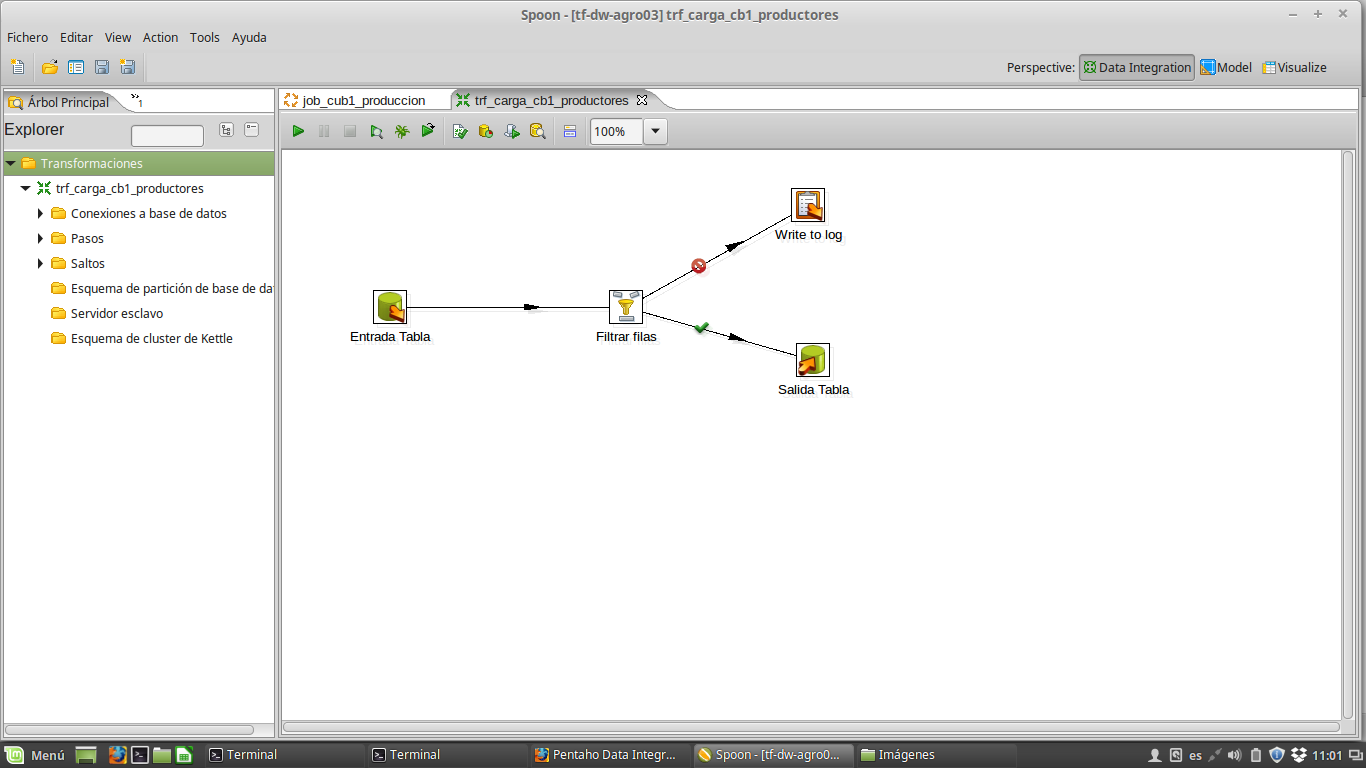


Figura 32: Transformación para la carga de Productores

La carga de los Productores emplea una consulta SQL que recupera todos los productores en base a los campos mencionados de razón social y cuit.

|  |
| --- |
| **SELECT** tb\_productores.id\_productor,  tb\_productores.razo\_soci,  tb\_productores.cuit  **FROM** tf.tb\_productores  **ORDER BY** tb\_productores.razo\_soci; |

Transformación para la Carga de Productos

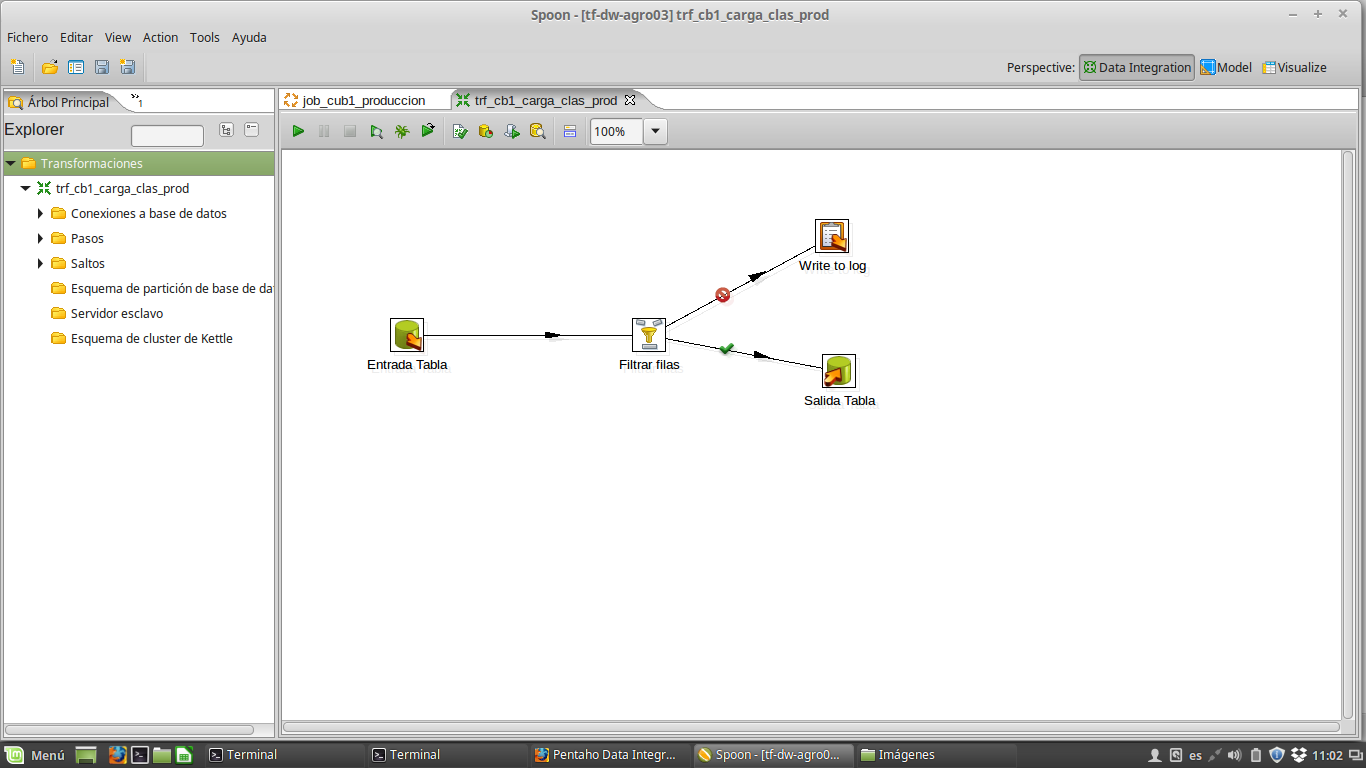


Figura 33: Transformación para la Carga de Productos

La carga de los Productos emplea una consulta SQL que recupera todos los nombres de productos registrados en el OLTP.

|  |
| --- |
| **SELECT** tb\_productos.id\_producto,  tb\_productos.nomb\_prod  **FROM** tf.tb\_productos  **ORDER BY** tb\_productos.nomb\_prod; |

Transformación para las Zonas Territoriales

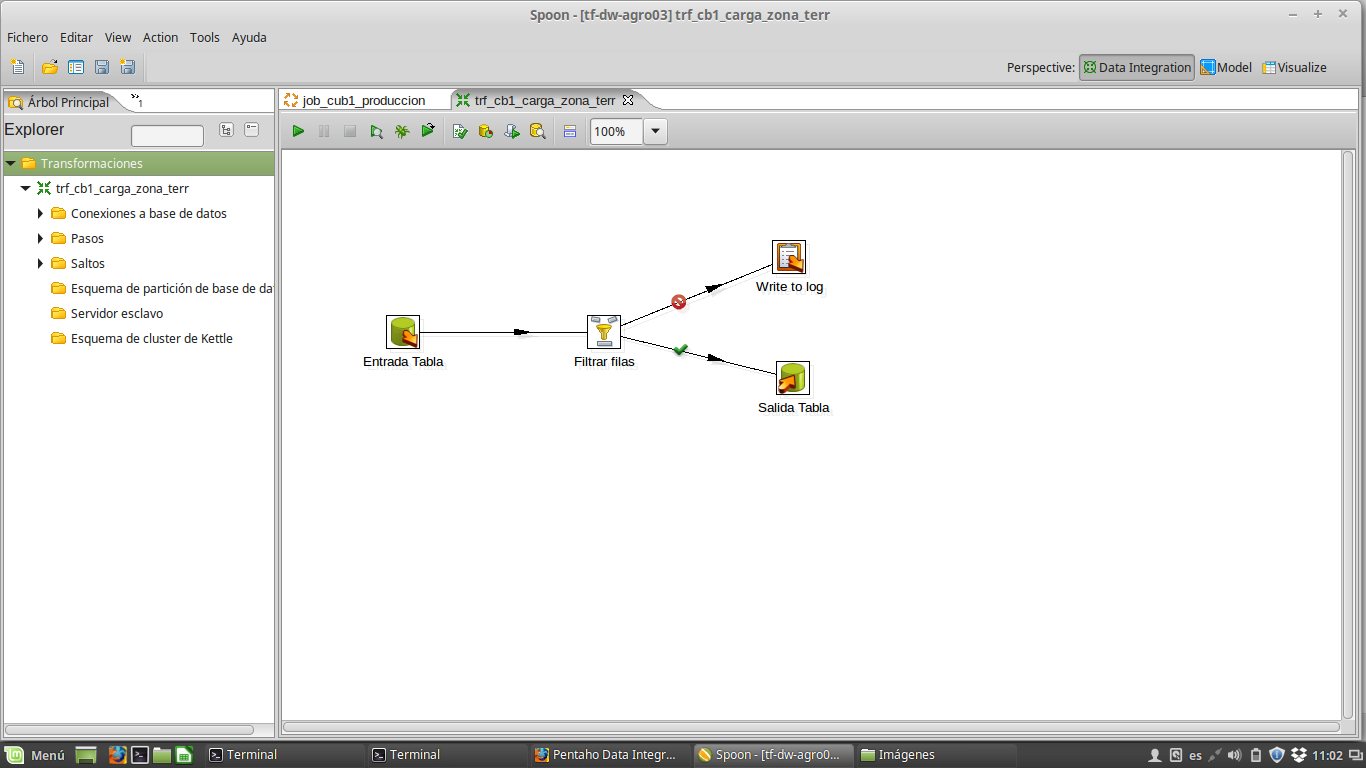


Figura 34: Transformación para las Zonas Territoriales

La carga de la dimensión Zona Territorial emplea una consulta SQL que recupera todos los nombres de departamentos y sus localidades registrados en el OLTP.

|  |
| --- |
| **SELECT**loc.id\_localidad,  loc.nomb\_loca,  dep.nomb\_depa  **FROM**tf.tb\_localidades loc  **JOIN**tf.tb\_departamentos dep **ON** dep.id\_departamento = loc.id\_departamento  **ORDER BY** tf.tb\_localidades.nomb\_loca; |

Transformación para la Clasificación de Productos

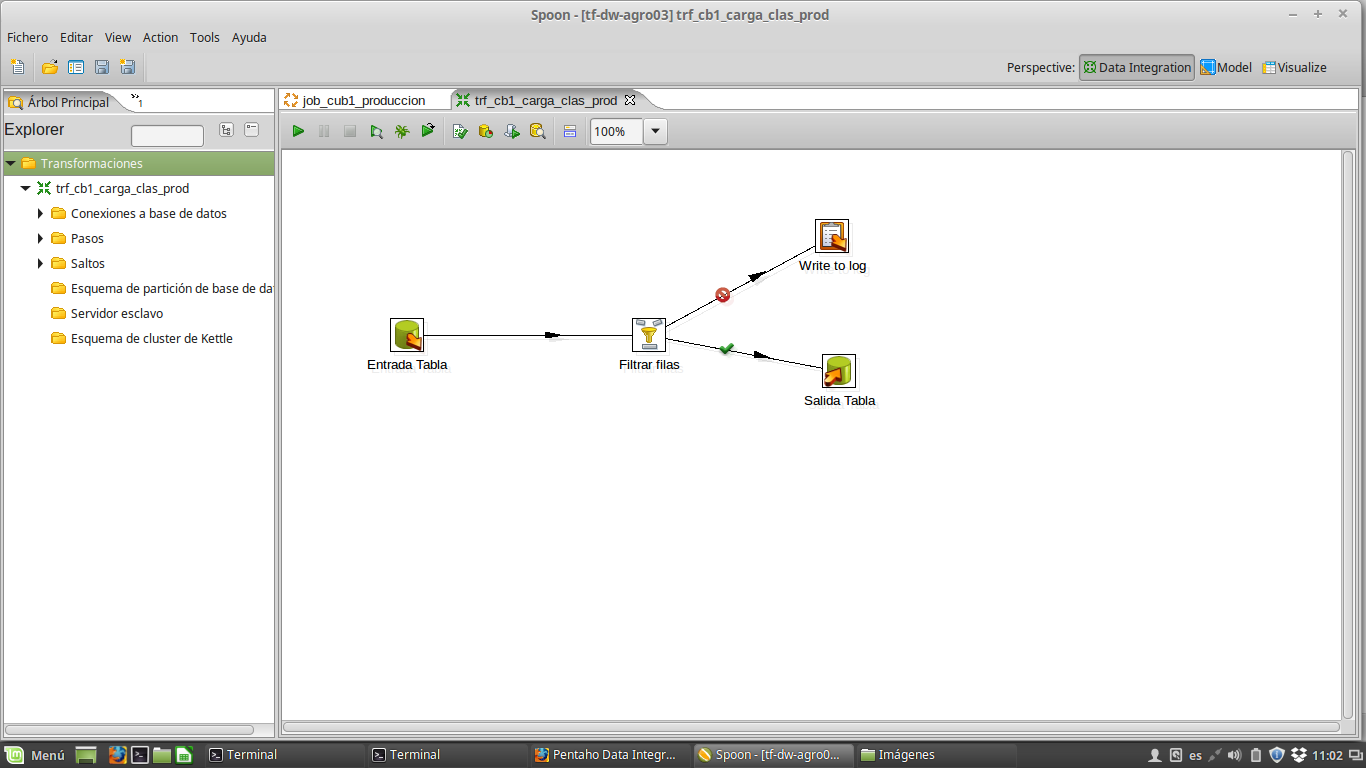


Figura 35:Transformación para la Clasificación de Productos

La carga de la dimensión Clasificación de productos emplea una consulta SQL que recupera todos los nombres de subrubros, rubros y grupos en los que se clasifican los productos registrados en el OLTP.

|  |
| --- |
| **SELECT**  sub.id\_subrubro,  sub.nomb\_subr,  rub.nomb\_rubr,  gru.nomb\_grup  **FROM** tf.tb\_subrubro\_productos sub  **JOIN** tf.tb\_rubro\_productos rub  **ON** sub.id\_rubro\_producto = rub.id\_rubro\_producto  **JOIN** tf.tb\_grupos\_productos gru  **ON** gru.id\_grupo\_producto = rub.id\_grupo\_producto  **ORDER BY** tf.tb\_subrubro\_productos. nomb\_subr; |

Transformación para la Fecha

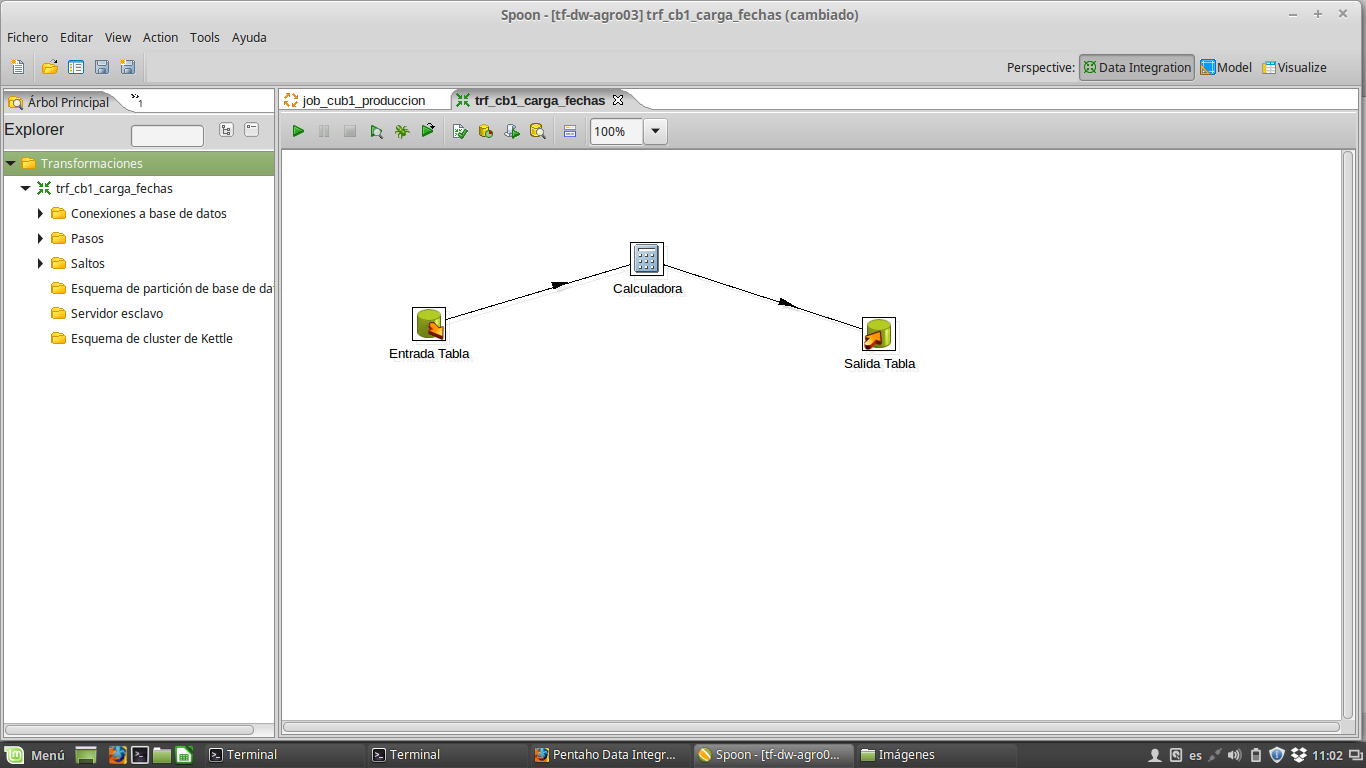


Figura 36:Transformación para la Fecha

La carga de la dimensión Tiempo emplea una consulta SQL que recupera todas las fechas en las que se registraron movimientos de producción por cada productor en registrados en el OLTP, una vez obtenidas las fechas se extrae el año y mes de dicho registro para incluirlas en la dimensión

|  |
| --- |
| **SELECTDISTINCT** tb\_productos\_x\_productor.fech\_regi  **FROM** tf.tb\_productos\_x\_productor  **ORDER BY** tb\_productos\_x\_productor.fech\_regi; |

Transformación para la Producción

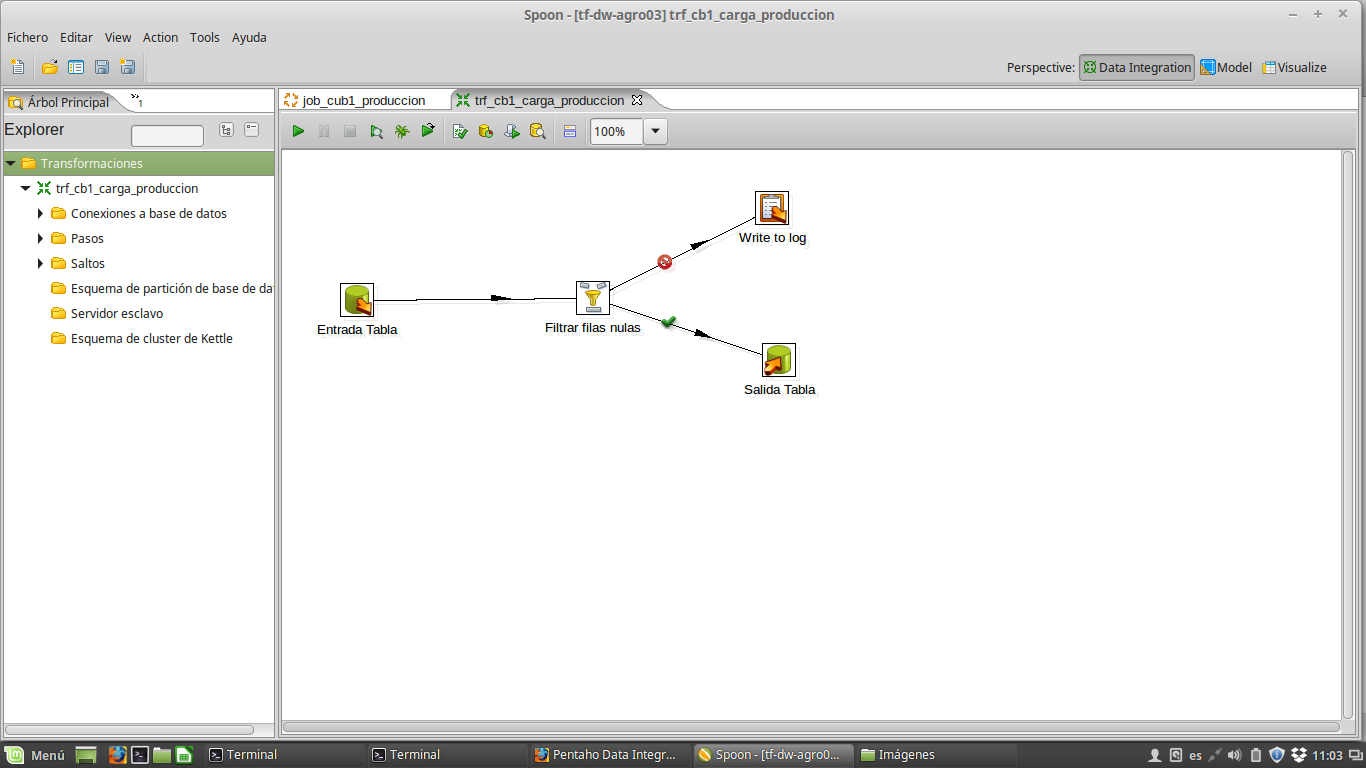


Figura 37:Transformación para laProducción

La carga de la tabla de hechos producción emplea una consulta SQL que recupera todos los registros de producción de cada productoralmacenado en el OLTP, una vez obtenidos los registros se aplican herramientas de filtro para evitar la carga de valores nulos en la tabla de hechos.

|  |
| --- |
| **SELECT** pxp.id\_productor,  pxp.id\_producto,  sub.id\_subrubro,  loc.id\_localidad,  pxp.supe\_prod,  pxp.volu\_prod,  unp.supe\_tota - pxp.supe\_prod **AS** supe\_sin\_cult,  pxp.ingr\_prod,  pxp.cost\_prod,  pxp.ingr\_prod - pxp.cost\_prod **AS** rent\_prod,  pxp.volu\_prod / pxp.supe\_prod **AS** rati\_prod,  pxp.fech\_regi  **FROM** tf.tb\_productos\_x\_productor pxp  **LEFT JOIN** tf.tb\_productores ptr **ON** ptr.id\_productor = pxp.id\_productor  **LEFT JOIN** tf.tb\_productos pro **ON** pro.id\_producto = pxp.id\_producto  **LEFT JOIN** tf.tb\_subrubro\_productos sub **ON** sub.id\_subrubro = pro.id\_subrubro  **LEFT JOIN**tf.tb\_cooperativas\_x\_productorcxp**ON**cxp.id\_productor = ptr.id\_productor  **LEFT JOIN**tf.tb\_unidad\_productivaunp**ON**unp.id\_productor = ptr.id\_productor  **LEFT JOIN**tf.tb\_localidadesloc**ON**loc.id\_localidad = unp.id\_localidad  **ORDER BY** pxp.id\_productor; |

Anexo 3

En este anexo se presenta el código XML necesario para la definición del Cubo de Producción y Rentabilidad

<Schema name="sch\_cbo1 agro">

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="productos">

<Hierarchy name="tb\_productos" visible="true" hasAll="true" primaryKey="id\_producto">

<Table name="tb\_productos" schema="cb1">

</Table>

<Level name="producto" visible="true" table="tb\_productos" column="producto" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="NombreProducto">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="productores">

<Hierarchy name="tb\_productores" visible="true" hasAll="true" primaryKey="id\_productor">

<Table name="tb\_productores" schema="cb1">

</Table>

<Level name="productor" visible="true" table="tb\_productores" column="productor" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="cuit" visible="true" table="tb\_productores" column="cuit" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="tiempo">

<Hierarchy name="tiempo" visible="true" hasAll="true" primaryKey="fecha">

<Table name="tb\_tiempo" schema="cb1">

</Table>

<Level name="anio" visible="true" table="tb\_tiempo" column="anio" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="mes" visible="true" table="tb\_tiempo" column="mes" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="clasificacion\_productos">

<Hierarchy name="tb\_clasificacion\_productos" visible="true" hasAll="true" primaryKey="id\_clas\_prod">

<Table name="tb\_clasificacion\_productos" schema="cb1">

</Table>

<Level name="grupo" visible="true" table="tb\_clasificacion\_productos" column="grupo" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="rubro" visible="true" table="tb\_clasificacion\_productos" column="rubro" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="subrubro" visible="true" table="tb\_clasificacion\_productos" column="subrubro" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="zona\_territorial">

<Hierarchy name="tb\_zona\_territorial" visible="true" hasAll="true" primaryKey="id\_zona\_territorial">

<Table name="tb\_zona\_territorial" schema="cb1">

</Table>

<Level name="departamento" visible="true" table="tb\_zona\_territorial" column="departamento" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="localidad" visible="true" table="tb\_zona\_territorial" column="localidad" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Cube name="cbo1\_produccion" visible="true" cache="true" enabled="true">

<Table name="tb\_produccion" schema="cb1">

</Table>

<DimensionUsage source="productos" name="Productos" visible="true" foreignKey="id\_producto" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="productores" name="Productores" visible="true" foreignKey="id\_productor" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="tiempo" name="Tiempo" visible="true" foreignKey="fecha" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="zona\_territorial" name="Zona\_Territorial" visible="true" foreignKey="id\_zona\_territorial" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="clasificacion\_productos" name="Clasificacion\_Productos" visible="true" foreignKey="id\_clas\_prod" highCardinality="false">

</DimensionUsage>

<Measure name="produccion\_kg" column="produccion\_kg" datatype="Numeric" aggregator="sum" visible="true">

</Measure>

<Measure name="hectareas\_cultivadas" column="hectareas\_cultivadas" datatype="Numeric" aggregator="sum" visible="true">

</Measure>

<Measure name="hectareas\_sin\_cultivar" column="hectareas\_sin\_cultivar" datatype="Numeric" aggregator="sum" visible="true">

</Measure>

<Measure name="ingreso\_produccion" column="ingreso\_produccion" datatype="Numeric" aggregator="sum" visible="true">

</Measure>

<Measure name="costo\_produccion" column="costos\_produccion" datatype="Numeric" aggregator="sum" visible="true">

</Measure>

<Measure name="rentabilidad\_produccion" column="rentabilidad\_produccion" datatype="Numeric" aggregator="max" visible="true">

</Measure>

<Measure name="ratio\_produccion" column="ratio\_produccion" datatype="Numeric" aggregator="avg" visible="true">

</Measure>

</Cube>

</Schema>

Bibliografía

Codd, E. F., Codd, S. B. Y Salley, C. T. *Providing OLAP (On-Line Analytical Processing) to User Analyst: An IT Mandate*.1993.

E. Zamudio, M. Vialey, H. Kuna. *OLAP en las PyMEs. Un modelo multidimensional común para empresas del sector yerbatero del nordeste argentino.* 2011.

S. Chaudhuri, U. Dayal. *An Overview of Data Warehousing and OLAP Technology*. 1997.

S. Chaudhuri, U. Dayal, and V. Ganti, “Database technology for decision support systems,” *Computer (Long. Beach. Calif).*, vol. 34, no. 12. Dec. 2001.

Godoy, A- Martínez, M. Carvallo Krieger, A. Franco, M. “*Propuesta de organización y gestión para el desarrollo de los productores de S. Antonio Misiones, Argentina*”, Editorial Universitaria, 2010.

Silberschartz, Korth&Sudarshan.*Fundamentos de Bases de Datos.* Libro, Quinta Edicion. 2005.

E.J.H Osorio. *Metodología Para El Desarrollo De Un Sistema De Inteligencia De Negocios Basada En El Proceso Unificado*. Tesis de Magíster en Ingeniería de sistemas y Computación, Departamento de ingeniería de sistemas e industrial, Universidad Nacional de Colombia, Bogotá, Colombia. 2011.

A.Godoy, M. Martíne.*Programa de Reorganización de Asociación de Pequeños Productores del Departamento General Manuel Belgrano y Capacitación de sus Recursos Humanos*. Resolución SPU Nº 885//09. Resolución FCE CD N°060/10. 2010.

A.Godoy, M. Martínez. *Proyecto de Investigacion. Codigo 16Q440.* Resolución FCEQyN CD N°191/10.

C. Radulescu, M. Radulescu, A. Rahoveanu.*A Multidimensional Data Model and OLAP Analysis for Agricultural Production*.2009.

A.Godoy, M. Martínez.*Programa de Reorganización de Asociación de Pequeños Productores del Departamento General Manuel Belgrano y Capacitación de sus Recursos Humanos*. 2010.

A.Godoy. “*Modelo de Diseño y Gestión de Políticas para los Micro-Productores Yerbateros de San Pedro, Misiones*”. Tesis Doctoral. 2006.

Inmon, William. H. *Building the Data Warehouse*. 4ta ed. Indianapolis: Wiley Publishing, Inc, 2005. 423 p. ISBN: 0-7645-9944-5

K.K.Chaturvedi, A. Rai, V. K. Dubey, P.K. Malhotra.*On-line analytical processing in agriculture using Multidimensional Cubes.* 2008

Kimball, Ralph Y Ross, Margy. *The data warehouse toolkit: the complete guide to dimensional modelling*. 2da. ed. Canada: Wiley Computer Publishing, 2002. 421 p.ISBN: 0-471-20024-7

L. Dejun , Z. Guangsheng. *Application of Data Mining Technology in Modern Agricultural Logistics Management Decision.*2008.

Luis Carlos Molina Félix, “*Torturando a los Datos Hasta que Confiesen*”. Coordinador del programa de Data Mining, Universidad Oberta de Catalunya (UOC), 2001.

Silberschartz, Korth&Sudarshan, “*Fundamentos de Bases de Datos”*, Libro, Quinta Edición. Mayo 2005.

R. D. Bernabeu. *HEFESTO: Metodología para la Construcción de un Data Warehouse.* 2010.

A.Godoy, M. D. Rolón,F. E. Acosta, Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales, “*Sistema de información para el cálculo de Indicadores para la producción Agropecuaria Sustentable*”, 2013

A.Godoy, H. Kuna, M. Rolón, Comité Ejecutivo De Desarrollo e Innovación Tecnológica, “*Sistema de información logístico de comercialización para productos agropecuarios y municipios del Departamento Gral. Manuel Belgrano*”, 2012

A.Godoy, M. Laveran.. *Indicatdores para la Sustentabilidad para la Produccion Agropecuaria del Noreste de Misiones. Codigo 16Q489.* Resolución FCEQyN CD N°426/12. Desde 2012 al 2015

EIL Brizuela, YC Blanco. Metodologías para desarrollar Almacén de Datos. *Revista de Arquitectura e Ingeniería.* 2013, vol.7 no.3 ISSN 1990-8830 / RNPS 2125.

A.GODOY. *Desarrollo de la Identidad de los productos agropecuarios, Comandante Andresito.* 2013. MinCyT. Resol N| 918/12

A.GODOY, G. Knass, N. Tessari, V. Naravaez. *La Rentabilidad Económica De Los Productores Del Noreste De Misiones*. 2015.

Keneth C. Laudon y Jane P. Laudon. Libro,S*istemas de Información Gerencial*. Octava Edición. Pearson. 2004.

E.N. Hernández, I.S. Mosquera. Pentaho: *software líder de Inteligencia de Negocio de códigoabierto*. Vol 10, No 2.Revista Telem@tica. ISSN 1729-3804.Departamento de Telemática. Facultad de Ingeniería Eléctrica. Instituto Superior Politécnico "José Antonio Echeverría".Cujae. 2011

A. Marinheiro, J. Bernardino, *Analysis of Open Source Business Intelligence Suites Análise de Suites Open Source Business Intelligence*. 2013.

Gartner Group, *Magic Quadrant for Business Intelligence and Analytics Platforms.*2015.

DuráSubiela, Sabina. *Sistemas de Información BI: Estado Actual y Herramientas de Software Libre*. UniversitatOberta de Catalunya. 2011.

R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit*. Wiley, 2004.

Instituto Nacional de Estadísticas y Censos. Republica Argentina. [en línea]. [Fecha de consulta: 10  Diciembre 2015]. Disponible en:[http://www.indec.gov.ar](http://www.indec.gov.ar/)

1. *StructuredQueryLanguaje* – Lenguaje de Consultas Estructurado (Silberschartz y Otros,2005) [↑](#footnote-ref-2)
2. *On-Line AnalyticalProcessing –*Procesamiento Analítico en Línea [↑](#footnote-ref-3)
3. (*OnLineTransactionProcessing –* Procesamiento de Transacciones en Línea) [↑](#footnote-ref-4)
4. Extract, transform and load – Extracción, Transformación y Carga [↑](#footnote-ref-5)
5. Licencia Pública General de GNU (*General PublicLicense*), para permitir suuso en software libre. [↑](#footnote-ref-6)
6. Ver Capítulo 2 Sección 3.1 [↑](#footnote-ref-7)
7. Ver Capitulo 2 Sección 6 [↑](#footnote-ref-8)
8. http://community.pentaho.com/ [↑](#footnote-ref-9)
9. Ver ANEXO 1:OLTP Productores [↑](#footnote-ref-10)
10. Pentaho Data Integration (PDI) o Kettle, es una potente e intuitiva herramienta que integra la Suite Pentaho para realizar los procesos de Extracción, Transformación y Carga. [↑](#footnote-ref-11)
11. Mondrian SchemaWorkbench es una utilidad de la suite Pentaho, la cual ofrece una interfaz de diseño que permite crear y probaresquemas de cubos OLAP visualmente. [↑](#footnote-ref-12)
12. Producción de la Yerba Mate <http://misionesonline.net/2015/05/06/agricultura-autorizo-un-leve-aumento-en-el-precio-de-la-yerba-mate/> [↑](#footnote-ref-13)
13. Producción de Mandioca http://intainforma.inta.gov.ar/?p=22175 [↑](#footnote-ref-14)
14. Producción Tealera http://misionesonline.net/2015/05/01/te-el-crecimiento-no-llega-al-productor/ [↑](#footnote-ref-15)
15. Producción de Tabacalera http://www.ambito.com/noticia.asp?id=769783 [↑](#footnote-ref-16)
16. Producción de Maíz http://www.misionesdigital.net/index.php/agro/record-historico-para-el-campo-119-millones-de-toneladas-cosechadas [↑](#footnote-ref-17)
17. Producción cítricos http://www.lanacion.com.ar/1731469-amenaza-a-la-citricultura-una-devastadora-plaga [↑](#footnote-ref-18)
18. [http://www.agro.misiones.gov.ar/index.php/component/search/?searchword=burley%202012&ordering=newest&searchphrase=all&limit=50](http://www.agro.misiones.gov.ar/index.php/component/search/?searchword=burley 2012&ordering=newest&searchphrase=all&limit=50) Noticia Nro. 6 [↑](#footnote-ref-19)
19. http://inta.gob.ar/documentos/el-cultivo-de-mandioca-tecnicas-para-elevar-la-produccion [↑](#footnote-ref-20)
20. http://misionesonline.net/2012/03/06/acordaron-precio-del-tabaco-burley-en-10-50-el-promedio-para-la-campana-2011-2012/ [↑](#footnote-ref-21)
21. Comisión Nacional del Trabajo Agrario(CNTA)del Ministerio de Trabajo, Empleo y Seguridad Social [↑](#footnote-ref-22)
22. http://www.lanacion.com.ar/1459169-el-dolar-oficial-y-los-paralelos-en-alza [↑](#footnote-ref-23)
23. http://www.lanacion.com.ar/1657500-cotizacion-dolar-hoy-blue-oficial-paralelo [↑](#footnote-ref-24)
24. http://misionesonline.net/2012/03/06/acordaron-precio-del-tabaco-burley-en-10-50-el-promedio-para-la-campana-2011-2012/ [↑](#footnote-ref-25)
25. Según Fuentes de la municipalidad de Andresito y publicación de Godoy y Otros (2015) [↑](#footnote-ref-26)