

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO MULTIDIMENSIONAL
PARA LA TOMA DE DECISIONES DE LA EMPRESA “GRUPO NATURAL SER”**

**CRISTIAN ELÍAS MAYORGA CHAVES
LILIANA CATALINA PÉREZ FONSECA**



**UNIVERSIDAD DE CUNDINAMARCA
FACULTAD DE INGENIERÍA
PROGRAMA INGENIERÍA DE SISTEMAS
FUSAGASUGÁ**

2018

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO
MULTIDIMENSIONAL PARA LA TOMA DE DECISIONES DE LA EMPRESA
“GRUPO NATURAL SER”**

CRISTIAN ELIAS MAYORGA CHAVES

LILIANA CATALINA PEREZ FONSECA

**DOCUMENTO PROYECTO DE GRADO PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS**

DIRECTORES

ANGELA PATRICIA ARENAS AMADO

JOSE FERNANDO SOTELO CUBILLOS

UNIVERSIDAD DE CUNDINAMARCA

FACULTAD DE INGENIERÍA

PROGRAMA INGENIERÍA DE SISTEMAS

FUSAGASUGÁ

2018

Copyright © 2018 por Cristian Mayorga & Catalina Pérez Todos los derechos reservados.

Nota de Aceptación

Firma de Director de Proyecto

Firma de Director de Proyecto

Firma de Jurado

Firma de Jurado

Agradecimientos

Agradecemos a la Ingeniera Ángela Patricia Arenas A. y al Ingeniero Fernando Sotelo.

Por su orientación, colaboración y asesoría en el desarrollo de este proyecto.

Liliana Catalina Pérez Fonseca, Cristian Elías Mayorga Chaves

Agradezco primeramente a Dios.

Por ser mi guía y fortaleza en los momentos difíciles, por brindarme experiencias, grandes bendiciones y sabiduría a lo largo de mi vida, para poder cumplir todas mis metas propuestas.

Agradezco a mi madre Sandra Liliana Fonseca Sanabria.

Por haber estado ahí a lo largo de mi infancia, llenándome de enseñanzas y consejos de vida, por haberme brindado todo ese amor incondicional, aunque fuese un corto tiempo sé que en este momento estarías orgullosa de mí.

Agradezco a mi padre Thomas Pérez Escamilla.

Porque a pesar de nuestra distancia física, sé que cuento contigo, y aunque no hemos pasado muchos momentos juntos, sé que te alegra ver a una hija fuerte y que sale adelante sin dejarse doblegar por los problemas.

Agradezco a mis abuelos Rosa Oliva Samacá de Fonseca y Juan Bautista Fonseca Sanabria.

Por brindarme su apoyo incondicional cada día en mi crecimiento personal, por tenerlos aún conmigo, a pesar de las dificultades de salud, sé que Dios nos bendice grandemente. Los Amo Mis Viejitos.

Agradezco a mi tío Fernando Pérez Escamilla.

Por sus consejos brindados a lo largo de mi vida y su apoyo incondicional.

Agradezco a mis Amigos.

Porque sé que cuento con ustedes en todo momento, por brindarme su apoyo, sus consejos y su cariño a lo largo de estos años y por el ánimo para seguir luchando por mis sueños. Los Quiero.

Liliana Catalina Pérez Fonseca

Agradezco primeramente a Dios

Por darme la oportunidad de hacer realidad esta meta y estar siempre a mi lado apoyándome y guiando en cada camino.

Agradezco a mis padres María Alicia Chaves Rodríguez y Elías Mayorga Muñoz.

Por estar ahí en cada momento dándome su apoyo, educación y lecciones de vida a lo largo de este camino. A mi madre por estar siempre ahí dándome una palabra de ánimo y confiar en cada una de mis decisiones A mi padre por sus consejos y por enseñarme que las cosas se consiguen con trabajo y dedicación.

Agradezco a mi hermana Eliana Marcela Mayorga Chaves

Por su carisma y cariño desde que éramos niños y por ser mi confidente en tantas travesías.

Agradezco a mis tíos Wilson Chaves y Maritza León.

Por la instrucción brindada y por el apoyo en este proceso.

Agradezco a mi familia y amigos

Que de alguna u otra manera han aportado para que pueda alcanzar esta meta.

Cristian Elías Mayorga Chaves

TABLA DE CONTENIDOS

Abreviaturas	xviii
Resumen.....	1
INTRODUCCIÓN	6
2. PLANTEAMIENTO DEL PROBLEMA	8
2.1 Descripción del Problema.....	8
3. JUSTIFICACIÓN	10
4. OBJETIVOS	11
4.1 Objetivo General.....	12
4.2 Objetivos Específicos.....	12
5. MARCO REFERENCIAL.....	13
5.1 Empresa Grupo natural Ser.....	13
Misión	13
Visión.....	13
Objetivos de la Empresa	13
Estructura Organizacional.....	13
6. MARCO TEÓRICO.....	16
6.1 Inteligencia de Negocios (Business Intelligence).....	16
6.2 Arquitectura Inteligencia de Negocios.....	16
6.2.1 Bodega de Datos (Data Warehouse)	16
6.2.2 ETL (Extracción, Transformación y Carga)	18
6.3 Bases de Datos OLTP y OLAP.....	19

	viii
6.3.1 OLTP (On-Line Transactional Processing)	19
6.3.2 OLAP (On-Line Analytical Processing)	19
6.3.4 Diferencias entre OLTP y OLAP.....	20
6.4 Análisis OLAP	20
6.4.1 Modelos de Datos Multidimensionales.....	21
6.4.2 Tablas de Dimensiones y Hechos	24
6.4.3 Sistemas ROLAP, MOLAP, HOLAP Y DOLAP.....	26
6.4.4 Cubo OLAP	27
7. MARCO TECNICO.....	30
7.1 SQL Server Management Studio (SSMS)	30
7.2 Servidor Web de Plataforma XAMPP	31
7.3 Lenguaje de Código Abierto PHP.....	31
7.4 HTML	32
7.5 Herramienta SQL Server Data Tools 2013.....	32
7.6 Complemento de Excel: Power Pivot	33
7.7 Herramienta Power BI	33
8. CRISP-DM: METODOLOGÍA PARA EL DESARROLLO DEL MODELO MULTIDIMENSIONAL	34
8.1 COMPRENSIÓN DEL NEGOCIO	38
8.1.1 Determinar los Objetivos del Negocio.....	38
8.1.2 Valoración de la Situación.....	39
8.1.3 Determinar los Objetivos de los Resultados	40
8.1.4 Realizar el Plan del Proyecto	41

8.2 COMPRENSIÓN DEL LOS DATOS	42
8.2.1 Recolectar los Datos Iniciales	42
8.2.2 Descripción de los Datos	46
8.2.3 Exploración de los Datos	46
8.2.4 Verificar la Calidad de los Datos	47
8.3 PREPARACIÓN DEL LOS DATOS	49
8.3.1 Seleccionar los datos	49
8.3.2 Limpiar los datos.....	51
8.3.3 Construir los datos	51
8.3.4 Integración de datos	51
8.4 DESARROLLO (MODELADO).....	53
8.4.1 Diseño Conceptual del Data Warehouse.....	53
8.4.2 Construcción del Data Warehouse.....	61
8.4.3 Diseño del Aplicativo Auxiliar para la Carga de los Datos al Data Warehouse.....	63
8.4.4 Construcción del Aplicativo Auxiliar de Carga de Datos.....	64
8.4.5 Proceso de ETL.....	72
8.4.6 Esquema del Modelo.....	81
8.4.8 Construcción del Modelo Multidimensional en SQL Server Data Tools – Visual Studio	82
8.4.9 Construcción del Modelo Multidimensional en Excel Power Pivot.....	91
8.4.10 Construcción del Modelo Multidimensional en Power BI	94
8.4.11 Evaluación del Modelo	96
8.5 EVALUACIÓN DEL MODELO MULTIDIMENSIONAL	107

	x
8.5.1 Evaluación de resultados en Power Pivot	107
8.5.2 Evaluación de resultados en Power BI.....	110
8.5.4 Determinar los próximos pasos.....	112
8.6 DESPLIEGUE	113
8.6.1 Capacitación en la Empresa “Grupo Natural Ser”	114
8.6.2 Planificación de la monitorización y del mantenimiento.....	116
8.6.3 Requerimientos para la implementación.....	117
GLOSARIO	119
ANEXO 1.....	123
ANEXO 2.....	124
CONCLUSIONES	135
LISTA DE REFERENCIAS.....	¡Error! Marcador no definido.

LISTA DE TABLAS

Tabla 1. Diferencias entre OLTP y OLAP.....	20
Tabla 2. Cronograma de actividades.....	41
Tabla 3. Diccionario de datos Dimensión Región	55
Tabla 4. Diccionario de datos Dimensión Ciudad	55
Tabla 5. Diccionario de datos Dimensión Cliente	55
Tabla 6. Diccionario de datos Dimensión Producto	56
Tabla 7. Diccionario de datos Dimensión Tiempo	56
Tabla 8. Diccionario de datos Dimensión Vendedor	56
Tabla 9. Diccionario de datos Dimensión Estado.....	56
Tabla 10. Diccionario de datos Ventas	57
Tabla 11. Costo de la solución informática	118

LISTA DE ILUSTRACIONES

Ilustración 1. Organigrama de la empresa Grupo Natural Ser.....	13
Ilustración 2. Arquitectura de un Data Warehouse.....	17
Ilustración 3. Proceso ETL	18
Ilustración 4. Esquema Estrella.....	22
Ilustración 5. Esquema Copo de nieve.....	23
Ilustración 6. Esquema Constelación.....	24
Ilustración 7. Logo SQL Server 2014.....	30
Ilustración 8. Logo Plataforma XAMPP.....	31
Ilustración 9. Logo del Lenguaje PHP	32
Ilustración 10. Logo del Lenguaje HTML.....	32
Ilustración 11. Logo SQL Server Data Tools	33
Ilustración 12. Logo de Excel Power Pivot	33
Ilustración 13. Logo de Power BI.....	34
Ilustración 14. Ciclo de vida de la Metodología CRISP-DM	36
Ilustración 15. Informe agrupado por producto y vendedores.....	42
Ilustración 16. Informe comparativo por cliente.....	43
Ilustración 17. Informe de número de clientes por producto	44
Ilustración 18. Informe de ventas detallado por producto	44
Ilustración 19. Ventas agrupadas por vendedor.....	45
Ilustración 20. Ventas por vendedor agrupado por producto.....	45
Ilustración 21. Monto total por productos.....	46
Ilustración 22. Porcentaje de las ventas totales por vendedor	46

Ilustración 23. Monto total de las ventas en los dos años	47
Ilustración 24. Monto total de las ventas en las dos ciudades	47
Ilustración 25. Informe ventas por vendedor agrupado por producto con ventas anuladas.....	48
Ilustración 26. Informe ventas por vendedor agrupado por producto con obsequios a clientes ...	48
Ilustración 27. Ventas agrupadas por vendedor, selección de elementos	49
Ilustración 28. Ventas agrupadas por producto, selección de elementos.....	49
Ilustración 29. Ventas agrupadas por vendedor, selección de atributos	50
Ilustración 30. Ventas agrupadas por producto, selección de atributos	50
Ilustración 31. Data set Grupo Natural Ser	52
Ilustración 32. Diseño conceptual del Data Warehouse	54
Ilustración 33. Datos Dimensión Región	57
Ilustración 34. Datos Dimensión Ciudad	58
Ilustración 35. Datos Dimensión Clientes	58
Ilustración 36. Datos Dimensión Productos.....	59
Ilustración 37. Datos Dimensión Tiempo	59
Ilustración 38. Datos Dimensión Vendedores	60
Ilustración 39. Datos Dimensión Estado.....	60
Ilustración 40. Datos Ventas	61
Ilustración 41. Creación del Data Warehouse en SQL Server.....	61
Ilustración 42. Creación de las Dimensión y tabla de Hechos en el Data Warehouse	62
Ilustración 43. Creación del Diagrama Entidad - Relación	62
Ilustración 44. Visualización del Esquema del Data Warehouse	63
Ilustración 45. Librería PHPExcel	63

Ilustración 46. Estructura del Aplicativo Auxiliar	64
Ilustración 47. index.php – parte 1.....	65
Ilustración 48. Index.php – parte 2	65
Ilustración 49. Index.php – parte 3	66
Ilustración 50. validacion.php.....	67
Ilustración 51. Llamado de funciones y limite de memoria.....	67
Ilustración 52. Estructura HTML de la validación de datos en el Aplicativo Auxiliar	68
Ilustración 53. Variables de lectura del archivo Excel en el Aplicativo Auxiliar.....	68
Ilustración 54. Arreglo que recorre los campos del archivo	69
Ilustración 55. Validaciones de los campos	69
Ilustración 56. Impresión de los errores.....	70
Ilustración 57. Encabezados de la tabla	70
Ilustración 58. Arreglo que recorre los datos para cargarlos en la DB	71
Ilustración 59. Impresión de cada campo	71
Ilustración 60. Conexión del Aplicativo Auxiliar con SQL Server	71
Ilustración 61. Código que carga los datos en la DB	71
Ilustración 62. Inicio Aplicativo Auxiliar	72
Ilustración 63. Selección y carga del archivo Excel en Aplicativo auxiliar	73
Ilustración 64. Selección de archivo Excel “Datos”	73
Ilustración 65. Validación de datos en Aplicativo Auxiliar.....	74
Ilustración 66. Error en validación de datos en Aplicativo Auxiliar	74
Ilustración 67. Corrección de errores en Aplicativo Auxiliar.....	75
Ilustración 68. Carga de datos en Aplicativo Auxiliar.....	75

Ilustración 69. Impresión de datos cargados en Aplicativo Auxiliar	76
Ilustración 70. Volver a pagina de inicio en Aplicativo Auxiliar	76
Ilustración 71. Datos de Región en el Data Warehouse.....	77
Ilustración 72. Datos de Ciudad en el Data Warehouse.....	77
Ilustración 73. Datos de Clientes en el Data Warehouse	78
Ilustración 74. Datos de Producto en el Data Warehouse.....	78
Ilustración 75. Datos de Tiempo en el Data Warehouse.....	79
Ilustración 76. Datos de Vendedores en el Data Warehouse	79
Ilustración 77. Datos de Estado en el Data Warehouse	80
Ilustración 78. Datos de Ventas en el Data Warehouse	80
Ilustración 79. Esquema Estrella para Modelo Multidimensional.....	81
Ilustración 80. Creación del Proyecto en SQL Server Data Tools	82
Ilustración 81. Selección del origen de datos.....	82
Ilustración 82. Selección de Data Warehouse.....	83
Ilustración 83. Elección de cuenta de servicio.....	83
Ilustración 84. Finalización de selección de origen de datos.....	84
Ilustración 85. Vistas del origen de datos	84
Ilustración 86. Selección de tablas para la creación del Modelo Multidimensional – Paso 1	85
Ilustración 87. Selección de tablas para la creación del Modelo Multidimensional – Paso 2	85
Ilustración 88. Estructura previa del Modelo Multidimensional	86
Ilustración 89. Creación del cubo OLAP – Paso 1	86
Ilustración 90. Creación del cubo OLAP – Paso 2	87
Ilustración 91. Selección de la Tabla de Hechos	87

Ilustración 92. Selección de las medidas para el Cubo OLAP	88
Ilustración 93. Selección de las Dimensiones.....	88
Ilustración 94. Estructura general del cubo OLAP	89
Ilustración 95. Esquema Estrella del Modelo Multidimensional en SQL Server Data Tools	89
Ilustración 96. Jerarquías del Modelo Multidimensional	90
Ilustración 97. Acceso a Power Pivot en Excel	91
Ilustración 98. Selección de origen de datos.....	91
Ilustración 99. Selección de servidor y del Data Warehouse.....	92
Ilustración 100. Selección de Dimensiones y Tabla de Hechos	92
Ilustración 101. Esquema Estrella del Modelo Multidimensional en Excel Power Pivot	93
Ilustración 102. Acceso a Power BI y selección de origen de datos.....	94
Ilustración 103. Selección del servidor.....	94
Ilustración 104. Selección de Dimensiones y Tabla de Hechos	95
Ilustración 105. Esquema Estrella del Modelo Multidimensional en Power BI.....	95
Ilustración 106. Dashboard Excel Power Pivot	97
Ilustración 107. Cuadros de Valores.....	97
Ilustración 108. Segmentador de Fecha	98
Ilustración 109. Segmentadores de Datos.....	98
Ilustración 110. Gráficos.....	99
Ilustración 111. Top 10 de productos más vendidos.....	99
Ilustración 112. Top 10 de los mejores vendedores.....	100
Ilustración 113. Mapa 3D – Paseo 1	100
Ilustración 114. Mapa 3D – Paseo 2	101

Ilustración 115. Dashboard Power BI – Página 1	102
Ilustración 116. Cuadros de Valores	102
Ilustración 117. Segmentadores de Datos	103
Ilustración 118. Segmentador de Fecha	103
Ilustración 119. Gráficos	104
Ilustración 120. Dashboard Power BI – Página 2	104
Ilustración 121. Top 10	105
Ilustración 122. Mapa	105
Ilustración 123. Tablas Dinámicas	106
Ilustración 124. Consulta 1 – Dashboard Excel Power Pivot	107
Ilustración 125. Consulta 2 – Dashboard Excel Power Pivot	108
Ilustración 126. Consulta 3 – Dashboard Excel Power Pivot	109
Ilustración 127. Consulta 1 – Dashboard Power BI	110
Ilustración 128. Consulta 2 – Dashboard Power BI	111
Ilustración 129. Consulta 3 – Dashboard Power BI	112
Ilustración 130. Evidencia Capacitación Empresa 1	114
Ilustración 131. Evidencia Capacitación Empresa 2	115
Ilustración 132. Evidencia Capacitación Empresa 3	115
Ilustración 133. Evidencia Capacitación Empresa 4	116

Abreviaturas

BI: Business Intelligence

BA: Business Analytics

ETL: Extraction, Transformation and Loading

DWH: Data Warehouse

HTML: HyperText Markup Language

PHP: Hypertext preprocessor

SQL: Structured Query Language

DB: Data Base

OLAP: On-Line Analytical Processing

OLTP: On-Line Transaction Processing

MOLAP: Multidimensional On-Line Analytical Processing

CRISP-DM: Cross Industry Standard Process for Data Mining

Resumen

En el proyecto Análisis, Diseño e Implementación de un Modelo Multidimensional para la Toma de Decisiones de la empresa **Grupo Natural Ser**, se realiza una solución informática por medio de una arquitectura de Inteligencia de Negocios (Business Intelligence) y una etapa de Minería de Datos. Esta etapa se denomina preprocesamiento de los datos, en la cual seleccionamos, limpiamos, transformamos y construimos los datos que obtuvimos de los informes de ventas suministrados a partir del software transaccional que actualmente maneja la empresa.

La solución brinda el soporte en el análisis de información para la toma de decisiones a nivel gerencial o ejecutivo en la compañía.

Respecto a la metodología seguida para el desarrollo del proyecto se aplicó CRISP-DM, siendo ésta una de las mejores prácticas reconocidas en la gestión de volumen de datos (IBM, 2012). Dentro de las fases de la metodología partimos en la Comprensión del Negocio, donde analizamos cómo se desarrollan las ventas en la empresa iniciando con la idea de que la empresa comercializa productos naturales en Colombia. Continuamos con la fase de Comprensión de los Datos donde revisamos los informes que nos suministraron (ventas en Medellín y Bogotá), resaltamos los datos claves y fundamentales para analizar las ventas; es decir el Core del negocio. Luego de analizar los datos procedemos a prepararlos, donde unimos los datos claves en un conjunto de datos y posteriormente los estructuramos en tablas de acuerdo al modelo de la Bodega de Datos a realizar. Se plantea una base de datos multidimensional buscando cumplir con el objetivo primordial: fortalecer el área de inteligencia de negocios, por lo cual diseñamos la bodega de datos en la herramienta de SQL Server 2014 Management Studio en la cual se

ingresan los datos preparados. Para automatizar el proceso de carga de los datos, desarrollamos un aplicativo en PHP, el cual lee el archivo Excel (donde se encuentran los datos organizados), valida que los datos estén correctos y de ser así los carga en la bodega de datos.

En este punto nos encontramos en la parte del desarrollo del modelo, por lo cual utilizamos SQL Server Data Tools para la creación del mismo y buscando que en un futuro cuando la empresa cree el área de TI, conformen una base tecnológica con estas herramientas. Ya para la implementación de la solución, utilizamos 2 herramientas con muy buen desempeño y soporte para analizar y visualizar la información presentada. Estas herramientas son Power Pivot (complemento user roles and could view the information with smartpone de Excel) donde podemos gestionar el modelo de datos y da un soporte robusto al análisis de los datos mediante tablas y gráficos dinámicos, o lo que podemos denominar cubo OLAP. La segunda herramienta es Power BI, en la cual se puede gestionar el Modelo de Datos, crear informes, generar gráficos y tablas dinámicas para visualizar los datos almacenados en la bodega de datos, además de ser una herramienta empresarial de gran utilidad porque entre algunas de sus principales funcionalidades está el almacenamiento del archivo en un servidor de internet y la asignación de roles a los usuarios dependiendo su cargo dentro de la organización, para que los reportes puedan ser visualizados desde cualquier lugar, incluyendo en un smartpone.

Finalmente, se realiza la capacitación al personal encargado de la empresa Grupo Natural Ser, la cual cubre el manejo de la información de las ventas (solución de negocio) y el manejo general de la solución tecnológica. Además, con esto se elaboran y entregan para la empresa dos cartillas digitales: manuales de usuario final y usuario técnico, que contienen la información de la solución informática.

Palabras Claves: Gestión de Datos, Inteligencia de Negocios, Analítica de Negocios, Minería de Datos, Bodega de Datos, Modelo Multidimensional.

Abstract

In the development of the project "Análisis, Diseño e Implementación de un Modelo Multidimensional para la Toma de Decisiones de la empresa Grupo Natural Ser", we make a informatic solution about of an architecture of business intelligence and one stage of data mining. The stage is data preprocessing, in this we select, clean, transform and build data, that we had get of sales report from transactional software that the company currently uses.

The point of this project is that the solution provide support to information analysis and decision making from management of company.

We search comply with the objective through CRISP-DM, this being one of the best recognized practices in data volume management (IBM, 2012). In the phases of methodology we start with business understanding, in this part we analyze about the sales development in the company, and with the idea that it market its products in almost the entire country. In the next phase we continue with data understanding, where we review the reports of Medellin and Bogota, and stand out the key data for analyze the sale; that is to say the most important part of business. After to analyze the data, we prepare it, and join the key data in a data set and structure the data based on the data warehouse that we would do. Now the challenge was built a multidimensional data base, looking comply the most important objective: strengthen the business area. For that we design the data warehouse with SQL Server 2014 Management Studio, in that the prepared data is entered. To automate the data loading process, we development a auxiliary application in PHP. the application read the Excel file (in this are the organized data), validate that all data are correct and load in the data warehouse.

In this point we are in the model development, which is why we use SQL Server Data Tools, for do it, and searching that in the future the company could be create the TI area with a base of this tools. In the deployment of solution, we use two tools that give a good support to analyze the presented information. These tools are Power Pivot (complement of excel) where we can management the data model and give a good support to data analysis with dinamic tables and grafics, or olap cube. The second tool is Power BI, in which you can manage the Data Model, create reports, generate graphs and dynamic tables to visualize the data stored in the data warehouse, as well as being a very useful business tool because among some of Its main features are the storage of the file on an internet server and the allocation of roles to users depending on their position within the organization, so that reports can be viewed from anywhere, including on a smartphone.

In the end of project, we make a training in the company to the personal tha use the information of sales, with a explication about the operation of the solution. Also with this, we elaborate and give for the company two digital booklets: user and technical manual, that have the informacion about the solution of businees intelligence.

Key Words: Data Management, Business Intelligence, Business Analytics, Data Mining, Data Warehouse, Multidimensional Model.

INTRODUCCIÓN

En la empresa Grupo Natural Ser no cuentan con una herramienta en donde puedan analizar de manera fácil y rápida la gran cantidad de datos que manejan, convertir estos datos en información y luego en conocimiento para implementar posteriores estrategias de negocio, buscando la manera más óptima de poder tomar decisiones que potencialicen el crecimiento del negocio.

Teniendo en cuenta el gran avance en las Tecnologías de la Información en donde estas pueden recolectar, almacenar y procesar datos de una organización, nace un conjunto de aplicaciones y procesos que facilitan la obtención de datos de diferentes fuentes como Sistemas de Información, Bases de Datos, ERP, CRM, entre otros; que manejan las empresas para poder recopilarlos y explotarlos para su posterior análisis e interpretación de los mismos. Esta temática es ahora conocida como Inteligencia de Negocios (Business Intelligence), la cual se presenta como una herramienta importante para la toma de decisiones ya que provee información oportuna y rápida para que la empresa se fortalezca operacional (datos), táctica (información) y estratégicamente (conocimiento) hacia el éxito.

El objetivo de este proyecto es la realización de una solución informática mediante la creación de un Modelo Multidimensional para la empresa **Grupo Natural Ser**, el cual permite visualizar y analizar la información de las ventas desde el año 2016 y hasta el mes de Octubre del año 2017, momento en que fueron suministrados los datos por parte de la empresa. Para ello, empleamos la Arquitectura de Inteligencia de Negocios (Business Intelligence) y algunos procesos de Minería de Datos (Data Mining), que en conjunto permiten reunir y transformar los datos en información estructurada para su posterior análisis, conocimiento valioso para la toma de decisiones estratégicas al interior de la organización.

Este modelo ayudará a la empresa a hacer un mejor uso de su información comercial a través de un modelo de analítica avanzada, que les permitirá tomar decisiones apoyadas en la tecnología orientada a la Inteligencia de Negocios, proyectándose con más eficiencia y eficacia en la venta de sus productos, y fortaleciéndose tecnológicamente para los cambios que hoy en día deben afrontar las organizaciones.

Igualmente, este proyecto servirá de base para todas aquellas personas (naturales o jurídicas), que deseen implementar una solución informática de Inteligencia de Negocios (BI) en el área de ventas y en otras áreas organizacionales, en donde se requiera el proceso de tratar e integrar los datos y convertirlos en información relevante para la toma de decisiones inteligentes, como lo representa el modelo multidimensional propuesto en el presente proyecto.

2. PLANTEAMIENTO DEL PROBLEMA

2.1 Descripción del Problema

Actualmente las empresas se enfrentan a diversos problemas a la hora de plantear estrategias para desenvolverse en el mercado. Si desean progresar tienen que utilizar tácticas, desarrollar fortalezas, establecer relaciones para buscar la forma óptima de aumentar sus ventas. Es decir, se necesita llegar al cliente y conocerlo mejor, puesto que cada cliente plantea una necesidad diferente. Entonces, se deja de hablar de “solo vender y nada más”, por el contrario, se plantean formas para pasar de una venta netamente transaccional, hacia la satisfacción del cliente y en un mejor escenario la transformación del mundo en base a lo que la empresa hace y aporta. Pero este proceso es largo y significa un esfuerzo significativo para una empresa, y para lograrlo se hace necesario implementar una serie de metodologías, herramientas, procesos, equipos, recursos, entre otros.

Frente a este desafío se plantea iniciar con un proceso que se denomina Business Intelligence (Inteligencia de Negocios) y algunas fases de Minería de datos (Data mining), cuyo objetivo es tratar e integrar un conjunto de datos para obtener información relevante en una organización, la cual posteriormente se convierte en un conocimiento valioso como lo puede ser el histórico de las ventas de una empresa (nuestro caso de estudio en particular, la compañía Grupo Natural Ser). Es diferente saber que se vendieron 100 productos en determinado lugar, a conocer cuál fue el cliente que más compró cierto producto en determinado periodo de tiempo y quien fue el vendedor encargado de la venta y en qué región, o por qué un grupo de clientes de una región dejó de comprar ciertos productos. Esa transición de datos a conocimiento es indispensable para que las empresas analicen su situación y se proyecten hacia el futuro con

estrategias de negocio como ya hemos mencionado. Pero en su gran mayoría, las compañías se quedan en el momento de los datos y no buscan esa transformación apoyándose en la tecnología, lo que hace que ese conocimiento no se alcance o se desconozca con el paso del tiempo, repercutiendo a futuro en el cumplimiento de las metas corporativas (planeación táctica y estratégica de la organización).

Para poder llevar a cabo BI en una organización se necesita personal especializado, herramientas adecuadas al entorno de la empresa y empezar a mitigar el problema al que se enfrentan. Por su parte, la empresa en la cual se plantea la solución de BI, “Grupo Natural Ser” es una fábrica y comercializadora de productos naturales a nivel nacional, con sede en la ciudad de Fusagasugá. Esta es una empresa joven que cuenta con un software contable y financiero limitado llamado World Office vinculado al sistema de gestión de base de Datos Microsoft Access, el cual solo muestra informes estándar de transacciones básicas de la empresa, los cuales se visualizan en Microsoft Excel con algunas estadísticas que no plantean una solución eficiente para transformar esos datos en conocimiento. Además, la empresa no cuenta con un personal especializado en este tipo de procesos, por consiguiente, no logra desarrollarse de la mejor forma óptima dentro del mercado local y nacional, perdiendo oportunidades que generan ventajas competitivas.

De esta forma, se hace necesario implementar una solución tecnológica basada en BI para que la empresa Grupo Natural Ser pueda transformar de forma eficiente, los datos en información y la información en conocimiento, y con esta capacidad informática potenciar a futuro la organización a través del conocimiento que se construye.

3. JUSTIFICACIÓN

La solución informática mediante un modelo multidimensional permite a la empresa **Grupo Natural Ser** empezar a manejar una herramienta con arquitectura en Inteligencia de Negocios (BI), teniendo en cuenta que hoy en día esta área de la gestión de datos es de gran utilidad al momento de organizar la información relevante en una organización, realizando el análisis de los datos y convertirlos en conocimiento importante para el crecimiento competitivo de la organización. Para ello nos enfocamos en el área comercial, exclusivamente en las ventas, sabiendo que la distribución de productos naturales es el área fuerte de la empresa.

Al momento de la implementación de esta solución informática les entregamos una herramienta que les proporciona de manera dinámica la información relacionada con las ventas realizadas desde el año 2016 en donde ellos pueden observar la cantidad de productos vendidos desde sus inicios, con cuantos clientes y vendedores cuentan, los meses en que han vendido más productos, entre otros análisis, para así sacar conclusiones de su papel como distribuidora, y respecto a esta información poder tomar decisiones para la gestión de procesos como promociones, publicidad, innovación en sus productos, entre otros.

Adicionalmente esta herramienta permite beneficios a las organizaciones como: Manejo rápido de la información, fácil entendimiento sin necesidad de requerir un conocimiento técnico especializado, los tiempos de respuesta son más rápidos; todo ello en pro de obtener un conocimiento del estado del negocio en un periodo determinado, que les permita ser un referente en la industria, apropiar mejores prácticas, generar innovación y hasta liderar en el sector.

4. DELIMITACIÓN

La solución informática basada en un Modelo Multidimensional para la toma de decisiones se desarrolló e implementó para la empresa productora y distribuidora de productos naturales “Grupo Natural Ser”, como caso de uso, en donde nos fueron suministrados:

- Informes sobre las ventas realizadas entre el 1 de enero del 2016 hasta el 31 de octubre del 2017 en las ciudades de Bogotá y Medellín.
- Información sobre la empresa, como: objetivo, misión, visión y estructura organizacional.

Documentos que son de uso exclusivo para el desarrollo del proyecto.

Se realizó una capacitación como alcance del proyecto de acuerdo a los objetivos planteados, el personal de la empresa conoció cuál fue el resultado de la solución informática y se les explicó los beneficios e importancia que ésta les trae para el análisis de sus datos, quedando a gusto con el trabajo realizado.

Este proyecto se elaboró en cumplimiento de la opción de “Trabajo de Grado” para optar el título de Ingeniero de Sistemas de la Universidad de Cundinamarca, Facultad de Ingeniería, Programa Ingeniería de Sistemas. Este proyecto se inició en el mes de septiembre del año 2017, con una duración de 8 meses.

5. OBJETIVOS

4.1 Objetivo General

Brindar una solución informática mediante un modelo multidimensional como soporte a la toma de decisiones dando apoyo a la gerencia de la empresa Grupo Natural Ser.

4.2 Objetivos Específicos

- Analizar el entorno del negocio enfocado en los objetivos de la empresa.
- Diseñar la solución informática con estándares de tipo OLAP, dirigida a la empresa.
- Implementar el modelo multidimensional en la empresa.
- Capacitar el personal de gerencia de la empresa Grupo Natural Ser.

6. MARCO REFERENCIAL

5.1 Empresa Grupo natural Ser

Es una empresa productora y distribuidora de productos naturales a nivel nacional, operando desde finales del año 2015, actualmente tiene presencia en gran parte del territorio nacional y se proyecta para abrir nuevos caminos de distribución a nivel internacional.

Misión

Colaborar con la restauración de la salud integral de nuestra sociedad a través de la educación y el servicio.

Visión

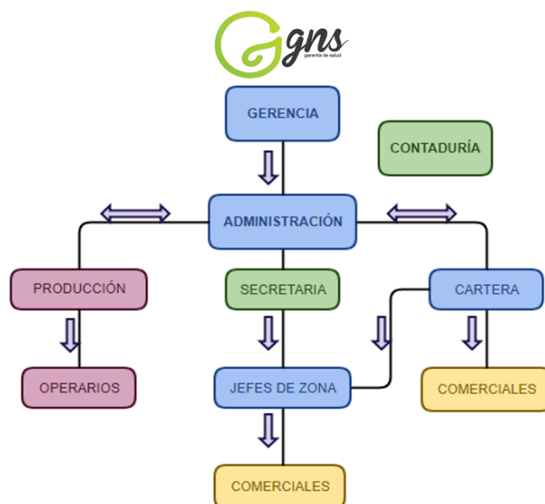
Lograr la renovación integral de la salud de nuestra sociedad.

Objetivos de la Empresa

Brindar atención integral en salud a las personas de todas las edades en su contexto social a través de la educación y medios naturales.

Estructura Organizacional

Ilustración 1. Organigrama de la empresa Grupo Natural Ser



7. ANTECEDENTES

Hoy en día las tecnologías de la información (TI) ocupan una parte importante dentro de las organizaciones empresariales ya que por medio de estas utilizan soluciones informáticas para poder sobresalir en un mercado competitivo, permitiendo la administración, gestión, coordinación y control de la empresa.

Una de esas soluciones informáticas es la implementación de un modelo multidimensional para toma de decisiones en las empresas ya que es de vital importancia puesto que permite orientar el rumbo de la misma, esto se logra gracias a los datos que operaciones anteriores han mostrado. Realizar un correcto análisis de estos datos brinda la oportunidad de crecer en el mercado en el que se desenvuelve la empresa.

A continuación, algunos proyectos que ya se han realizado acerca del tema:

- **“Diseño de un modelo multidimensional de un data mart del área de capacitación en el INEGI”**

Este proyecto respondía a la necesidad de contar con información del desarrollo de un área de esa institución para relacionarla con más tópicos, mediante un modelo multidimensional de data mart, lo cual puede ser un punto de referencia enfocado a nuestro proyecto.

La solución planteada en este proyecto de grado incluye:

- Desarrollo de un modelo de data mart para el área de capacitación.
- Implementación de 2 cubos multidimensionales
- Para la consulta de los data mart, se realizó una aplicación en la web

(Matínez, 2008).

- **“Utilización de información histórica para decisiones empresariales”**

Este proyecto está enfocado en el desarrollo de modelos multidimensionales, DWH (Data Warehouse), cubos y procesos de ETL (Extracción, Transformación y Carga), los cuales nos sirven para orientarnos en los procesos que debemos desarrollar en este proyecto.

La solución planteada en este proyecto de grado incluye:

- Desarrollo de un modelo de data mart análisis OLAP.
- Desarrollo de software en java para proceso de ETL.
- El análisis OLAP se realiza mediante un navegador.

(Peña Rivera & Suárez Daza, 2005).

- **“Modelo multidimensional para apoyo a toma de decisiones para el sistema contable de las firmas comisionistas de bolsa”**

Este proyecto está enfocado en el desarrollo de un modelo multidimensional utilizando cubos OLAP (On-Line Analytical Processing), cubo multidimensional, ETL (Extracción, Transformación y Carga), modelos de reportes para la toma de decisiones, que son temas de utilidad para el desarrollo del proyecto.

La solución planteada en este proyecto de grado incluye:

- Desarrollo de una base de datos multidimensional.
- Desarrollo de un modelo de reportes con la base de datos multidimensional

(CORREA MUÑOZ & GOMEZ GARCIA, 2014).

8. MARCO TEÓRICO

8.1 Inteligencia de Negocios (Business Intelligence)

Es un conjunto de procesos, metodologías y aplicaciones que buscan transformar los datos en información, y la información en conocimiento y así mejorar la toma de decisiones en procesos empresariales (S.L., Sinergia e Inteligencia de Negocio, 2016). Este proceso busca generar una ventaja competitiva de la empresa frente a las demás y se realiza depurando los datos de un sistema transaccional, que provee información sin estructura, y transformándolos en información estructurada lista para análisis, reporting etc.

La inteligencia se base en un ciclo de adquirir información, tomar una decisión y generar una acción, que posteriormente genera más información y sigue el proceso continuamente.

8.2 Arquitectura Inteligencia de Negocios

8.2.1 Bodega de Datos (Data Warehouse)

Es una base de datos caracterizada por integrar información de una o más fuentes diferentes para procesar los datos permitiendo un análisis de muchas formas distintas de esta información, y, obteniendo grandes velocidades de respuesta en dichos análisis. Esta técnica de almacenamiento permite gestionar la gran cantidad de datos que existen en una empresa. (S.L., Sinergia e Inteligencia de Negocio, 2016)

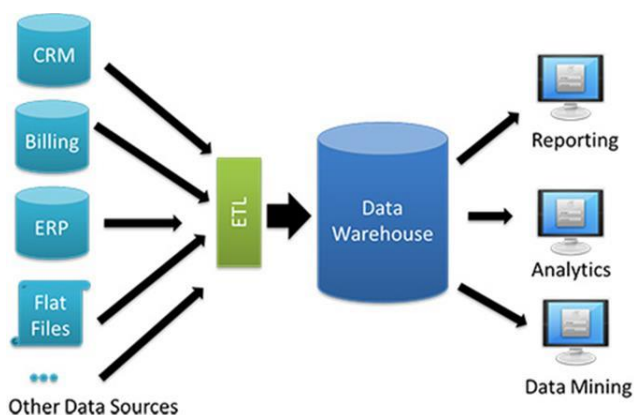


Ilustración 2. Arquitectura de un Data Warehouse

El término de Data Warehouse fue utilizado por primera vez por Bill Inmon y se traduce como “*almacén de datos*”. (S.L., Sinergia e Inteligencia de Negocio, 2016) Este almacén de datos está caracterizado por ser:

- **Integrado:** Los datos que son almacenados en un Data Warehouse deben ser integrados dentro de una estructura consistente, si existe alguna inconsistencia en estos datos esta debe ser eliminada. Generalmente la información suele estructurarse en diferentes niveles de detalles según las necesidades que el usuario desee.
- **Temático:** Los datos son organizados por temas para mejorar el entendimiento y facilitar el acceso a los datos por parte del usuario final.
- **Histórico:** En la mayoría de los sistemas operacionales los datos que muestran sobre el estado de la empresa son en tiempo presente, mientras que la información de la empresa que es almacenada en un Data Warehouse se puede utilizar para realizar un análisis de tendencias de manera histórica y esto permite el poder realizar comparaciones de la evolución del negocio.

- **No Volátil:** Este almacén de datos existe para ser leído y no modificado, en donde solo permite la actualización de los datos, pero no realizar cambio alguno sobre los datos que ya existían. La información que contiene un Data Warehouse es permanente.

8.2.2 ETL (Extracción, Transformación y Carga)

Es un proceso en donde se permite mover datos de una o varias fuentes (bases de datos o archivos operacionales), extrayéndolos, transformándolos y cargándolos en una fuente nueva (base de datos o bodega de datos). (PowerData Especialista en Gestión de Datos, 2013)

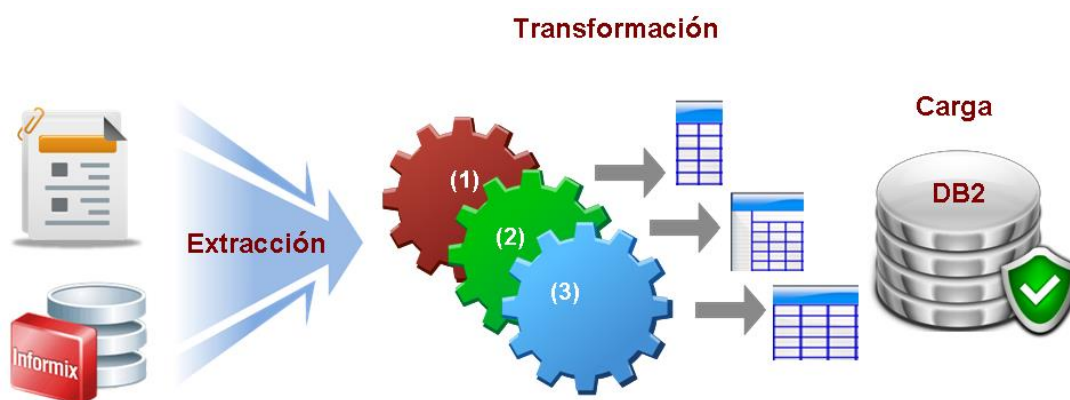


Ilustración 3. Proceso ETL

Fases de un proceso ETL

Proceso de Extracción: Es obtener datos de uno o varias fuentes como bases de datos, hojas electrónicas o archivos de texto (PowerData Especialista en Gestión de Datos, 2013). Para ello hay que seguir unos pasos como:

- Extracción de los datos
- Análisis de los datos
- Verificación de los datos y que estos cumplan con los requisitos, de no ser así, no se pueden utilizar los datos.

- Conversión de los datos a un nuevo formato para su transformación.

Proceso de Transformación: Es formatear y limpiar esos datos en caso de que sea necesario, para que estos puedan ser cargados (PowerData Especialista en Gestión de Datos, 2013).

Algunas formas de transformación son:

- Seleccionar algunas columnas
- Dividir los datos de una columna en varias columnas
- Unir datos de varias columnas en una columna
- Obtener nuevos valores de datos

Proceso de Carga: Es cuando se cargan los datos ya transformados a la nueva fuente ejemplos pueden ser Base de Datos, Data Mart, Data Warehouse (PowerData Especialista en Gestión de Datos, 2013).

8.3 Bases de Datos OLTP y OLAP

8.3.1 OLTP (On-Line Transactional Processing)

Son bases de datos orientadas a los procesos transaccionales. Una transacción genera un proceso atómico en donde se pueden realizar operaciones como insertar, modificar o eliminar datos. Estas bases de datos están limitadas por mostrar solo datos actuales o recientes. (S.L., Sinergia e Inteligencia de Negocio, 2016)

8.3.2 OLAP (On-Line Analytical Processing)

Son bases de datos orientadas a procesos analíticos. Pueden facilitar la lectura de grandes cantidades de datos para poder recolectar información relevante y útil de la empresa. Estas bases de datos se estructuran según el área de negocio y los datos son globalizados en toda la

organización y se puede evidenciar los datos de manera histórica. (S.L., Sinergia e Inteligencia de Negocio, 2016)

8.3.4 Diferencias entre OLTP y OLAP

Tabla 1. Diferencias entre OLTP y OLAP

CARACTERISTICAS	OLTP	OLAP
OBJETIVO	Soportar actividades transaccionales diarias	Consultar y analizar información estratégica y táctica
TIPO DE DATOS	Operacionales	Para la toma de decisiones
MODELO DE DATOS	Normalizado	Desnormalizado
CONSULTA	SQL	SQL mas extensiones
DATOS CONSULTADOS	Actuales	Actuales e históricos
HORIZONTE DE TIEMPO	60 – 90 días	5 – 10 años
TIPOS DE CONSULTAS	Repetitivas, predefinidas	No previsibles, dinámicas
NIVEL DE ALMACENAMIENTO	Nivel de detalle	Nivel de detalle y diferentes niveles de sumarización
ACCIONES DISPONIBLES	Alta, baja, modificación y consulta	Carga y consulta
NUMERO DE TRANSACCIONES	Elevado	Medio o bajo
TAMAÑO	Pequeño – mediano	Grande
TIEMPO DE RESPUESTA	Pequeño (segundos – minutos)	Variable (minutos – horas)
ORIENTACION	Orientado a las aplicaciones	Orientado al negocio
SELLO DE TIEMPO	La clave puede o no tener un elemento de tiempo	La clave tiene un elemento de tiempo
ESTRUCTURA	Generalmente estable	Generalmente varia de acuerdo a su propia evolución y utilización

(Bernabeu_Dario, 2009)

6.4 Análisis OLAP

Es la capacidad de poder contextualizar una o más variables (medidas) a partir de perspectivas (dimensiones). Este concepto de análisis OLAP o análisis multidimensional surge cuando un analista de una empresa requiere realizar un estudio del comportamiento del negocio, dándose cuenta la complicada tarea de mirar todas las variables que existen en una organización tales como: clientes, país, producto, tiempo, venta, costos, etc (LOGICALIS, 2015).

6.4.1 Modelos de Datos Multidimensionales

Estos modelos se utilizan para construir el modelo lógico conceptual de la bodega de datos, en donde se manejan variables de los diferentes indicadores de gestión y control del negocio (SENA). Existen tres tipos de modelos, como:

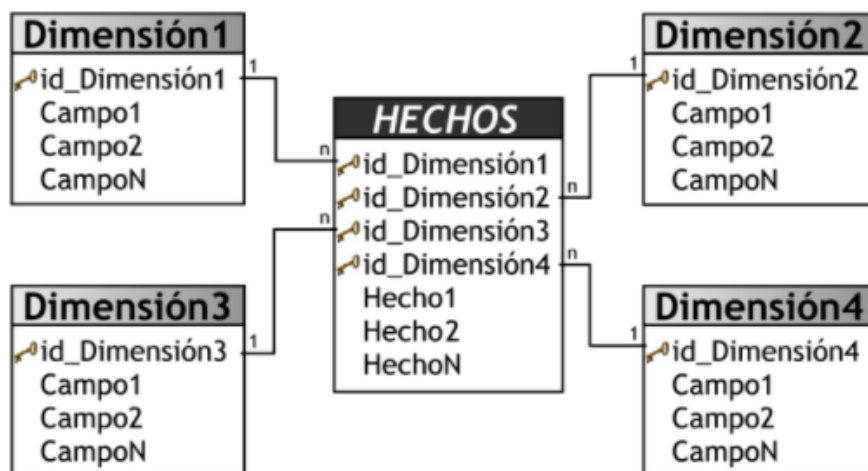
Esquema Estrella

Este esquema está conformado por una tabla central llamada tabla de hechos, que contiene los datos importantes para la realización del análisis multidimensional y se relaciona con unas tablas llamadas dimensiones. Este modelo ofrecer una gran velocidad al realizar el procesamiento del análisis y debe estar totalmente desnormalizado (SENA).

Principales características del esquema de estrella:

- Posee mejores tiempos de respuesta
- Su diseño es de fácil modificación
- Está estructurado de forma paralela de cómo los usuarios visualizan y manipulan los datos a como está diseñado
- Simplifica el análisis
- Facilita la interacción con herramientas de consulta y análisis

Ilustración 4. Esquema Estrella



Esquema Copo de nieve

Este esquema está conformado por una tabla central llamada tablas de hechos relacionada con las tablas de dimensiones, al igual que el esquema de estrella, agregando que las tablas de dimensiones a su vez se relacionan con otras tablas de dimensiones. Este modelo tiene un proceso de normalización, el cual reduce la redundancia de datos y disminuye el rendimiento en el análisis (SENA).

Es un modelo más acercado al modelo entidad - relación, ya que sus tablas están debidamente normalizadas, se puede implementar luego de haber diseñado un modelo de estrella.

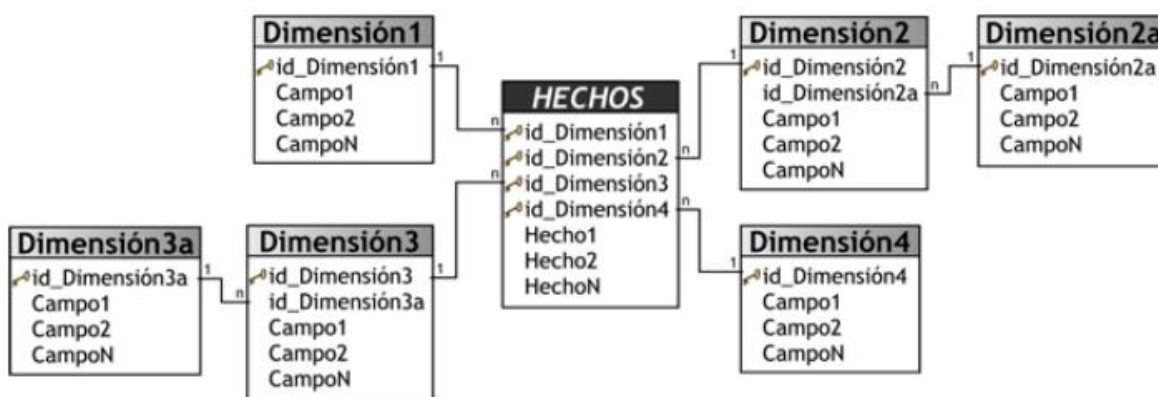
Principales características del esquema copo de nieve:

- Posee una mayor complejidad en su estructura
- Hace una mejor utilización del espacio
- Es muy útil en una tabla de dimensiones que contiene muchas tuplas
- Debido a la normalización de las tablas requiere menos esfuerzo de diseño
- Se pueden realizar clases de jerarquías en donde se podrá analizar de los datos más generales a los más detallados

Sin embargo, existen dos inconvenientes con el diseño de esta estructura, como:

- Debido a la cantidad de tablas que puede tener, puede volverse un modelo de datos inmanejable
- Al existir gran cantidad de relaciones entre tablas puede tener una disminución del desempeño de análisis

Ilustración 5. Esquema Copo de nieve



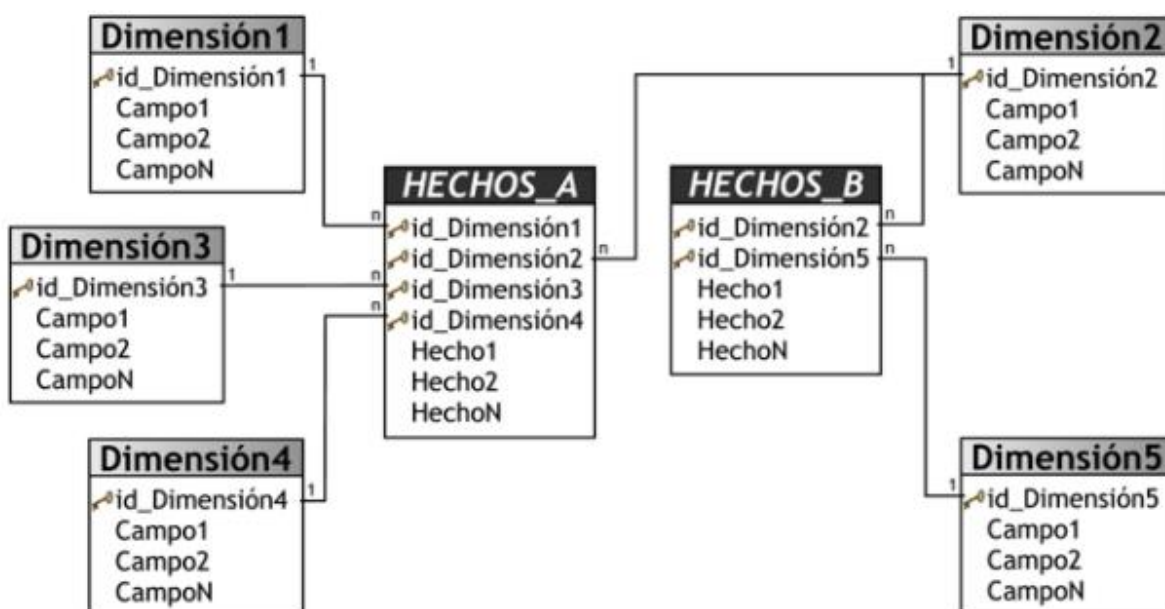
Esquema Constelación

Este esquema está conformado por una o más tablas centrales o tablas de hechos, siendo una la principal y otra ejerciendo un papel de auxiliar, haciendo un poco más complejo el modelo, y las tablas de dimensiones relacionadas a las tablas de hechos, no es necesario que las tablas de hechos compartan la relación con las mismas dimensiones, ya que la tabla de hechos auxiliar puede solo estar relacionada con alguna de las dimensiones de la tabla de hechos principal y con nuevas tablas de dimensiones (SENA).

Principales características de la estructura constelación:

- Al tener más de una tabla de hechos, permite analizar más de un aspecto clave del negocio.
- Se pueden reutilizar las tablas de dimensiones, ya que una sola tabla de dimensiones se puede relacionar con las dos tablas de hechos, la principal y la auxiliar

Ilustración 6. Esquema Constelación



6.4.2 Tablas de Dimensiones y Hechos

Tabla de hechos

Son las que contienen los hechos que serán utilizados para el análisis de los datos, estos son cuantitativos. Una de las características importantes de estas tablas es su nivel de detalle de la información almacenada en ella, otra característica es que se lleva la clave primaria de las tablas de dimensiones, las cuales son las que definen su nivel de detalle (Bernabeu_Dario, 2009).

Existen dos tipos de hechos en las tablas de hechos:

- **Hechos básicos:** Son los hechos que están representados sólo en la tabla de hechos.
- **Hechos derivados:** Son los hechos que se forman de la combinación de uno o más hechos con alguna operación matemática, se puede acceder a ellos por medio de consultas SQL sencillas.

Tablas de hechos agregadas

Son las que se generan después de procesar una consulta en la tabla de hechos, es decir que estas se agregan luego de darle la instrucción por medio de SQL y debe incluir operaciones matemáticas como sumalizaciones (Bernabeu_Dario, 2009).

Tablas de Hechos Preagregadas

Son las que se generan antes de procesar una consulta en la tabla de hechos, es decir que estas instrucciones se hacen a través de una tabla ya creada inicialmente y que haya sido sumariada por medio del proceso de ETL (Bernabeu_Dario, 2009).

Esta clase de tabla de hechos cuenta con unos beneficios, como:

- Reducir la utilización de recursos del hardware
- Reducir el número de registros que serán analizados por los usuarios
- Reducir el tiempo de respuesta en las consultas realizadas por parte de los usuarios

Tablas de Dimensiones

Son las que almacenan los detalles acerca de los hechos, cada una de estas tablas tiene los campos y atributos que describen los procesos del negocio, estas permiten filtrar y manipular la información almacenada en una tabla de hechos (Bernabeu_Dario, 2009). Cada una de las tablas posee una identificación única y por lo menos un dato referente a la información relevante de la organización.

Una tabla de dimensiones debe contener los siguientes campos:

- Clave principal
- Clave foránea
- Datos de referencia primarios y secundarios

6.4.3 Sistemas ROLAP, MOLAP, HOLAP Y DOLAP

Existen cuatro tipos de implementación de los modelos de datos dimensionales, estas estructuras deben estar desnormalizadas o semi desnormalizadas de manera que puedan facilitar el acceso a los datos y tener facilidad al momento de realizar las consultas. Estos tipos de sistemas son:

Sistema ROLAP

Construidos sobre bases de datos relacionales que emplean, principalmente, esquemas de copo de nieve o de estrella. Es un sistema apto para analizar grandes volúmenes de datos, disponen una facilidad de rendimiento (LOGICALIS, 2015). Los cubos multidimensionales en estos sistemas son generados de manera dinámica después de realizar diferentes consultas, los usuarios no tendrán que intervenir en la creación del cubo ni en el mantenimiento del mismo.

Este sistema tiene las siguientes características:

- Almacena información en una base de datos relacional
- Utiliza índice de mapas de bits
- Utiliza índices de unión (join)
- Optimiza las consultas
- Cuenta con extensiones SQL

Sistema MOLAP

Construidos sobre bases de datos multidimensionales, mejorando la rapidez en el almacenamiento de datos y eficiencia en la extracción de los mismos, estos sistemas utilizan estructuras de almacenamiento como Arrays, al momento de crear el cubo multidimensional requiere de una instancia previa para que los datos puedan ser consultados de manera más rápida (LOGICALIS, 2015).

Sistema HOLAP

Combinan los sistemas ROLAP con los MOLAP. Se recomienda almacenar los datos más recientes en un sistema MOLAP para mejorar la rapidez de los análisis, y los más antiguos o de menor uso en ROLAP, dada la facilidad que presenta para su almacenamiento (LOGICALIS, 2015).

Sistema DOLAP

Está orientado a los equipos de escritorio, recoge todos los datos que desea analizar desde una base de datos relacional y la guarda en el escritorio.

6.4.4 Cubo OLAP

Son estructuras multidimensionales que a partir de una base de datos relacional con una gran cantidad de datos almacenada permiten el análisis con una gran velocidad para ahorrar tiempo y recursos.

Estos cubos convierten los datos almacenados en filas y columnas a una matriz con una cierta cantidad de dimensiones (Bernabeu_Dario, 2009). Existen unos objetos importantes para la elaboración del cubo multidimensional, como:

Indicadores

Son las expresiones que se basan en sumalizaciones que pertenecen a un hecho (Bernabeu_Dario, 2009). Para su realización se pueden utilizar operaciones como: sumas, conteos, promedios, funciones matemáticas, estadística, entre otras.

Atributos

Son los campos o criterios de análisis que pertenecen a las tablas de dimensiones (Bernabeu_Dario, 2009).

Jerarquías

Es la relación lógica que existe entre dos o más atributos del cubo multidimensional, esta relación siempre tiene que ir de manera correspondiente como “padre-hijo” (Bernabeu_Dario, 2009). Estas jerarquías poseen las siguientes características:

- Pueden existir varias jerarquías en un solo cubo.
- Pueden estar compuestas por dos o más niveles.
- Si existe una relación “padre-hijo” siempre debe tener atributos consecutivos de un nivel a otro.

Una de las principales ventajas de hacer uso de las jerarquías en los cubos, es que se puede analizar de mejor manera los datos desde el nivel general hasta el nivel más detallado.

Relación

Es la manera como los atributos tiene interacción dentro de una jerarquía (Bernabeu_Dario, 2009). Existen dos tipos de relaciones:

- **Explícitas:** Se pueden modelar con atributos directos y tiene una línea continua con una jerarquía.

- **Implícitas:** Son relaciones de muchos a muchos, ya que este tipo de relaciones no funciona de manera directa.

Granularidad

Es la que representa el nivel de detalle en donde se desea almacenar la información de acuerdo al tipo de negocio que se esté analizando (Bernabeu_Dario, 2009).

9. MARCO TÉCNICO

7.1 SQL Server Management Studio (SSMS)

Es un sistema integrado creado por Microsoft que administra cualquier tipo de entorno SQL, este entorno se puede utilizar para acceder, configurar, administrar y desarrollar bases de datos relacionales y multidimensionales tales como Data Warehouse, en diferentes herramientas de Microsoft SQL Server. Esta herramienta incluye editores de scripts o herramientas gráficas. Este cuenta con 4 tipos de servidor que son:

- **Database Engine:** Esta herramienta gestiona el motor de base de datos relacional.
- **Analysis Services:** Esta herramienta es útil para crear modelos multidimensionales.
- **Reporting Services:** Esta herramienta permite crear informes, darles niveles de observación y se pueden visualizar en navegadores web u otras herramientas como lo puede ser Excel.
- **Integration Services:** Esta herramienta sirve para migrar datos de diferentes orígenes a un Data Warehouse; es decir facilita el proceso de ETL (Extracción, Transformación y Carga).

Ilustración 7. Logo SQL Server 2014



7.2 Servidor Web de Plataforma XAMPP

XAMPP es un servidor web de plataforma, software libre, de fácil acceso que permite interpretar lenguajes como PHP, servir como gestor de bases de datos como MySQL, el servidor web apache entre otros.

Ilustración 8. Logo Plataforma XAMPP



7.3 Lenguaje de Código Abierto PHP

PHP (Hypertext Preprocessor) es un lenguaje de programación de código abierto (Php, 2018) y multiplataforma orientado al desarrollo web de forma dinámica y puede ser trabajado junto con HTML, siendo esta su principal ventaja, tiene la posibilidad de conectar a diferentes herramientas de bases de datos. Este lenguaje posee 4 características (Red Gráfica Latinoamérica, s.f.):

- **Velocidad:** Rapidez al ser ejecutado y consume pocos recursos de una máquina.
- **Estabilidad:** Utiliza un sistema de administración de recursos propio.
- **Seguridad:** Tiene diferentes niveles de seguridad que se pueden configurar en el archivo .ini.
- **Simplicidad:** Tiene una gran cantidad de librerías y se le pueden agregar extensiones.

Ilustración 9. Logo del Lenguaje PHP



7.4 HTML

HTML (HyperText Markup Language) es un lenguaje de marcado para realizar páginas web (Wikipedia, s.f.). Este sirve para dar estructura a una página, agregar multimedia, imágenes, párrafos, todo esto por medio de etiquetas. Se puede crear o editar en un editor de texto básico.

Ilustración 10. Logo del Lenguaje HTML



7.5 Herramienta SQL Server Data Tools 2013

Esta herramienta permite realizar procesos de BI tales como Analysis Services Multidimensional And Data Mining, Integration Services entre otros los cuales facilitan el análisis de datos, la minería de datos, creación de cubo de datos, migración de datos etc. Esta herramienta se despliega en el entorno de visual studio y su conexión directa a datos es SQL Server. Puede trabajar con un proyecto de base de datos o directamente con una instancia de base de datos conectada.

Ilustración 11. Logo SQL Server Data Tools



7.6 Complemento de Excel: Power Pivot

Power Pivot es un complemento de Excel que permite trabajar con modelos de datos y da opciones para un análisis de datos eficaz. Además, soporta una gran cantidad de datos y da un amplio soporte para trabajarlos. También se cuenta con las tablas y gráficos dinámicos de Excel, haciendo de esta una gran herramienta para soluciones de Inteligencia de Negocios (Soporte Microsoft, s.f.).

Ilustración 12. Logo de Excel Power Pivot



7.7 Herramienta Power BI

Power BI son un conjunto de herramientas de tipo empresarial con el fin de analizar datos. Al igual que Power Pivot soporta la construcción del modelo además de generar una solución óptima puesto que brinda la opción de visualizar los datos desde el celular sin necesidad de tener un computador. Además, permite conectarse a diferentes fuentes de datos y creación de

informes muy robustos. Utiliza gobernanza y seguridad de datos por lo que puede escalar a toda la empresa (Microsoft Power BI, s.f.).

Ilustración 13. Logo de Power BI



8. CRISP-DM: METODOLOGÍA PARA EL DESARROLLO DEL MODELO MULTIDIMENSIONAL

CRISP-DM (Cross Industry Standard Process For Data Mining) como metodología orienta trabajos de minería de datos, conformada con algunas fases bidireccionales, las cuales permiten revisar parcial o totalmente las fases anteriores, generando una retroalimentación del proyecto en general. Como modelo de proceso se enfoca en ofrecer un ciclo vital de la minería de datos (IBM Knowledge Center, 2012).

Esta metodología nace en 1996 pero fue utilizada en un proyecto en 1997, siendo este dirigido por 5 empresas: Teradata, Daimler AG, Ohra, NCR Corporation y SPSS. Durante algunos años de le hizo cambios a esta metodología y actualmente es IBM la principal empresa que promueve el proceso de CRISP-DM (Wikipedia, s.f.).

Dentro de esta metodología se manejan seis fases que no necesariamente deben ser seguidas de manera estricta ya que cada una de estas se pueden personalizar de acuerdo a los requerimientos del proyecto a realizar. Estas fases son las siguientes:

Comprensión del Negocio: En esta fase es en donde se realiza la comprensión del negocio como tal, en donde se determinan los objetivos comerciales, se revisa la situación actual de la empresa, se determinan objetivos de lo resultado que desea obtener el cliente.

Comprensión de los Datos: En esta fase se hace una recopilación, descripción, exploración y verificación de los datos con los que se va a trabajar.

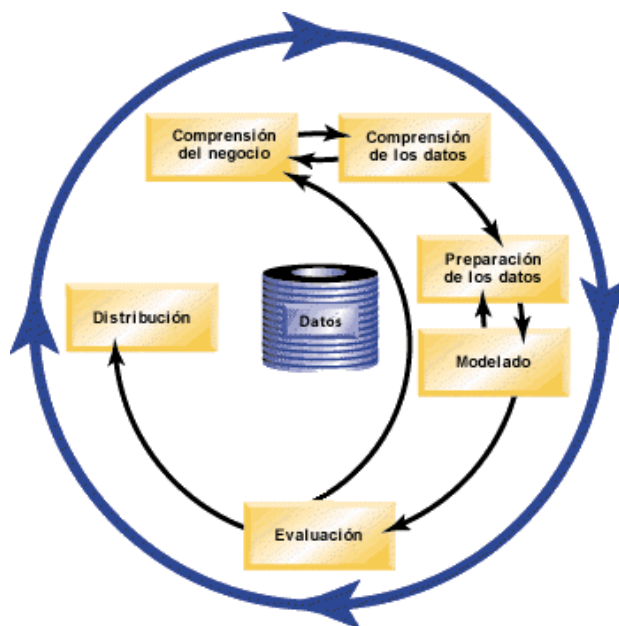
Preparación de los Datos: En esta fase se hace una selección, limpieza, construcción e integración de los datos que fueron recopilados en la fase anterior de modo que queden listos para el desarrollo del modelo y no haya errores.

Desarrollo o Modelado: En esta fase se realiza todo el desarrollo del proyecto, generación de diseños, estructuras, modelos, entre otros. También se realiza una evaluación previa cuando existen varias opciones de modelado, se verifica cual(es) de ellas cumplen con los objetivos del proyecto y del negocio.

Evaluación: Después de haber seleccionado la mejor opción de modelado, se realiza la evaluación de resultados, comprobando que funcione correctamente y esté listo para su distribución o entrega.

Distribución o Despliegue: Aquí es donde termina el proyecto, se pueden realizar capacitaciones, implementaciones del producto entre otros. Se deja la respectiva planificación de mantenimiento o un documento de que puede seguir disponible para posteriores mejoras.

Ilustración 14. Ciclo de vida de la Metodología CRISP-DM



Casos de Éxito:

Indagando acerca de esta metodología, encontramos algunos casos de éxito en donde ha sido utilizada. Estos son:

Using data mining for bank direct marketing: an application of the crisp-dm methodology

Este es un artículo desarrollado por un ingeniero de sistemas, un profesor de la escuela de negocios y un profesor asociado al departamento de sistemas de información, en las universidades correspondientes en Portugal. En este muestran cómo a través del paso del tiempo las campañas de marketing han perdido el impacto en el cliente, además de otros factores que pueden ser mejorados mediante el uso de BI y Data Mining. Utilizaron la metodología CRISP-DM para proponer un modelo que busque que las campañas sean más eficientes y el cliente se suscriba al depósito bancario (entorno donde se desarrolló el proyecto). Se demostró que el uso de la metodología fue de gran valor para proponer modelos basados en datos reales (Moro & Laureano).

Metodología crisp para la implementación Data Warehouse

Este artículo fue desarrollado por un Ingeniero de Sistemas y Magister en Teleinformática y en Economía y Docente de la Universidad Distrital Francisco José de Caldas, una Ingeniera de Sistemas Consultora de ORACLE y un Ingeniero de Sistemas Docente del Sena. En este resaltan la importancia de generar informes concisos y veraces como elemento fundamental de la toma de decisiones, surgiendo la necesidad de implementar Data Warehouse y demás elementos de la Arquitectura de Inteligencia de Negocios. Resaltan la importancia de la metodología CRISP-DM como una de las principales para rescatar específicamente Data Warehouse y Data Mining. Mediante esta metodología muestran el desarrollo óptimo de bodegas de datos y de modelos predictivos (SALCEDO PARRA, GALEANO, & RODRIGUEZ B., Enero - Junio de 2010).

Resultado

Es por ello que decidimos escoger esta metodología como guía para el desarrollo del Modelo Multidimensional ya que opta por las fases necesarias para alcanzar dicho objetivo. Además de que tiene gran enfoque en proyectos relacionados con BI, Data Mining, Data Warehouse, entre otros temas de Data Management objeto base de nuestro proyecto.

8.1 FASE 1: COMPRENSIÓN DEL NEGOCIO

8.1.1 Determinar los Objetivos del Negocio

La empresa “Grupo Natural Ser” productora y distribuidora de productos naturales, que tiene presencia en la mayor parte del país, desea poder realizar el análisis de ventas de sus productos de manera dinámica e interactiva, logrando así una mejor gestión de los datos relevantes sobre las ventas realizadas.

Objetivos comerciales

- Comprender por medio del análisis de los datos el estado actual y pasado de las ventas.
- Descubrir que entre los productos cual ha tenido mayor venta y en qué zona del país.
- Observar la evolución de la demanda histórica de las ventas.

Criterios de Éxito

- Visualizar las ventas de cada zona de forma dinámica e interactiva.
- Comparar las ventas entre productos.
- Comparar las ventas de manera histórica de acuerdo al crecimiento que ha tenido el negocio.
- Cumplimiento de la elaboración de la solución informática.

Compilación de la información de la empresa

- **Inventario de recursos:**

La empresa cuenta con un software transaccional World Office Enterprise el cual almacena la información de forma restringida en una base de datos de Microsoft Access, desde el año 2015 hasta la actualidad.

- **Descripción de la problemática:**

Como ya ha sido mencionado en el punto anterior la empresa cuenta con un software transaccional el cual almacena toda la información de la empresa, pero este programa no muestra la información de manera dinámica facilitando el análisis y gestionamiento de los datos más relevantes de la empresa, que en este caso sería la información del área de ventas. Decidimos enfocarnos en esta área ya que es el fuerte de la empresa y su recurso sostenible.

- **Descripción de la Solución:**

De acuerdo a la problemática que presenta la empresa pudimos notar que el software transaccional que ellos manejan, muestra una lista de informes, por ejemplo: Informe de ventas por cliente, Informe de ventas por producto, entre otros; de donde se selecciona el que se desea visualizar y lo descarga en un archivo plano.

La solución que planteamos es una herramienta informática que carga información desde esos archivos planos a una bodega de datos (base de datos) transfiriendo los datos al modelo multidimensional que permite visualizar y gestionar los datos más relevantes del área de ventas como fechas, productos, clientes, etc., de manera dinámica e interactiva para el gerente comercial de la empresa.

8.1.2 Valoración de la Situación

Inventario de recursos

- **Origen de datos:**

Los datos provienen directamente de los informes generados por el software transaccional, en formato xls. El programa los almacena en una base de datos de Microsoft Access, a la cual solo puede acceder el mismo proveedor de World Office.

Requisitos, supuestos y restricciones

Requisitos:

Para la implementación de la solución, se recomienda que el software a utilizar sea el siguiente:

- SQL Server 2014 Management Studio
- SQL Server Data Tools for Visual Studio 2013
- Microsoft Excel 2016
- XAMPP 5.6

Como requisito adicional, se recomienda que la empresa adquiera las licencias correspondientes para la implementación de esta solución.

Supuestos:

Al finalizar el proyecto se entregarán los resultados junto con una capacitación en la empresa sobre lo que se realizó durante el desarrollo del proyecto y manejo de la herramienta.

Restricciones:

Se acordó con la empresa realizar un acuerdo de palabra de confidencialidad de los datos, el cual especificaba que la información de las ventas solo la podíamos utilizar para el proyecto.

8.1.3 Determinar los Objetivos de los Resultados

Enfocado en los objetivos que se desean cumplir mediante la solución informática, se plantea:

- Identificar la fecha, producto y ciudad, donde se registraron la mayor cantidad de ventas.
- Identificar la variación de las ventas en los últimos 2 años en las ciudades de Medellín y Bogotá.
- Verificar el comportamiento de las devoluciones en los productos más críticos.

8.1.4 Realizar el Plan del Proyecto

Para facilitar la organización del proyecto y su tiempo estimado, se dividirá en las siguientes etapas:

Tabla 2. Cronograma de actividades

FASE	TIEMPO	RECURSOS	RIESGOS
Comprensión del Negocio	5 semanas	Todos los integrantes del proyecto	
Comprensión de los datos	5 semanas	Todos los integrantes del proyecto	Problemas tecnológicos, inconvenientes con los datos
Preparación de los datos	6 semanas	Todos los integrantes del proyecto	Problemas tecnológicos, inconvenientes con los datos
Modelado	10 semanas	Todos los integrantes del proyecto	Problemas de desarrollo de la herramienta y modelo
Evaluación	2 semanas	Todos los integrantes del proyecto	Problemas con los resultados del modelo
Distribución	2 semanas	Todos los integrantes del proyecto	Inconvenientes al implementar el modelo, problemas de comprensión del manejo de la herramienta.

8.2 FASE 2: COMPRENSIÓN DE LOS DATOS

8.2.1 Recolectar los Datos Iniciales

Los datos para los que la empresa nos suministró acceso, corresponden a dos (2) ciudades en las cuales presentan mayor cantidad de ventas y en un periodo comprendido del 1 de enero de 2016 hasta el 31 de octubre del 2017. Cabe resaltar que se utilizó datos reales para el planteamiento de la solución informática y de igual modo del modelo, dando la certeza de que la información presentada será correcta y de gran valor para la empresa a la hora de analizar la información.

Se nos otorgó los mismos datos tanto para la ciudad de Bogotá como la ciudad de Medellín. A continuación, describimos los datos que nos fueron suministrados en informes y tablas de Excel:

- **Informe Agrupado por Producto Y Vendedores**

Este informe nos muestra los productos vendidos y las unidades de estos, en el periodo de los 2 años.

Ilustración 15. Informe agrupado por producto y vendedores

Descripción	Cantidad	Precio Prom	Val Total Antes de Iva	Iva	Valor Total
AU ACID URIK	5,807.00	8,210.25	48,034,861.36	0.00	48,034,861.36
BIO BIO-CIR	6.00	18,800.00	112,800.00	0.00	112,800.00
BIOF BIOFYL	14.00	11,957.14	167,400.00	0.00	167,400.00
CHIA CHIA GRANO	42.00	4,410.00	185,220.00	0.00	185,220.00
CHD LICHO CHOCO YUSTON	91.00	23,442.98	2,133,302.02	0.00	2,133,302.02
COL COLAGEN HIDROLIZADO	289.00	9,665.11	2,793,217.50	0.00	2,793,217.50
DB DBT	2,643.00	13,542.35	35,792,435.72	0.00	35,792,435.72
DES DESPARASITOL KIDS	3,293.00	8,655.00	28,500,913.61	0.00	28,500,913.61
DIG DIGEST	284.00	12,737.95	3,601,578.96	0.00	3,601,578.96
DR DREVITAN ADULTO	11,970.00	10,093.21	119,810,269.57	0.00	119,810,269.57
DRP DREVITAN PLUS	315.00	14,706.67	4,632,600.00	0.00	4,632,600.00
DVE DVEYTE	50.00	11,730.00	586,500.00	0.00	586,500.00
EMA EXTRACTO DE MALTA	58.00	7,906.60	458,582.57	0.00	458,582.57
FAR FAROT	2,414.00	13,579.69	32,781,379.42	0.00	32,781,379.42
FEM FEMIGEN	73.00	11,773.97	859,500.00	0.00	859,500.00
FLOI FLORACOL ADULTO	685.00	17,429.47	11,990,600.00	0.00	11,990,600.00
GOL GOLESTROL GNS	3,698.00	20,532.56	75,929,416.66	0.00	75,929,416.66
GRA 1 GRADIGESTAR BREVA Y COCO	2.00	2,914.00	5,828.00	0.00	5,828.00
GRA 3 GRADIGESTAR COCO Y PIÑA	7.00	3,840.57	26,884.00	0.00	26,884.00
GRA 4 GRADIGESTAR COCO Y PIÑA	6.00	7,426.00	44,556.00	0.00	44,556.00
GRA 5 GRADIGESTAR PAPAYA Y ALMENDRA	11.00	3,880.18	43,782.00	0.00	43,782.00
GRA 6 GRADIGESTAR PAPAYA Y ALMENDRA	13.00	7,657.69	99,660.00	0.00	99,660.00
HGD HGD X 300 ML	52.00	12,524.78	651,288.79	0.00	651,288.79

- **Informe Comparativo por Cliente**

Este informe indica los clientes de la ciudad y el valor monetario que este compró en cada uno de los meses. Este informe está desde el 1 enero del 2016 hasta la 31 diciembre del 2016, puesto que el software de la empresa solo permitía exportar este informe en periodos completados.

Ilustración 16. Informe comparativo por cliente

JENNIFER MARITZA										
Informe Comparativo Por Cliente										
Entre 1/01/2016 Y 31/12/2016										
Cliente	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octub
5 ADRIANA	0.00	0.00	0.00	0.00	0.00	0.00	861,200.00	202,950.00	256,500.00	
6 ADRIANA	141,000.00	0.00	101,500.00	275,000.00	311,500.00	0.00	261,500.00	146,800.00	150,000.00	
7 ADRIANA RUTH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	51
8 AGUSTIN ENRIQUE	0.00	109,980.00	0.00	356,370.00	0.00	359,720.00	160,200.00	239,850.00	0.00	210
9 ALBA	54,000.00	163,560.00	0.00	117,500.00	0.00	0.00	234,060.00	196,200.00	0.00	168
10 ALBA	0.00	955,500.00	0.00	382,000.00	0.00	0.00	432,000.00	0.00	0.00	764,000.00
11 ALBAN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	96,000.00	0.00	
12 ALEJANDRO	0.00	0.00	84,000.00	0.00	0.00	0.00	0.00	0.00	0.00	
13 ALEXANDRA	0.00	0.00	0.00	78,000.00	0.00	0.00	0.00	0.00	1,111,000.00	600
14 ALVARO	0.00	0.00	198,000.00	196,000.00	0.00	0.00	0.00	0.00	0.00	
15 ANA CONSUELO	333,200.00	0.00	109,980.00	234,000.00	0.00	598,600.00	412,200.00	0.00	0.00	247
16 ANA CRISTINA DE LAS MERCEDES	0.00	0.00	227,000.00	0.00	0.00	167,000.00	210,000.00	0.00	213,500.00	
17 ANA ELVA	0.00	93,800.00	110,000.00	0.00	0.00	0.00	0.00	0.00	87,000.00	95
18 ANA RITA	156,600.00	0.00	142,410.00	136,350.00	159,750.00	0.00	0.00	0.00	0.00	118
19 ANA YOMARA	0.00	0.00	0.00	0.00	0.00	0.00	90,500.00	0.00	0.00	
20 ANATILDE	0.00	0.00	0.00	0.00	0.00	199,800.00	0.00	0.00	172,020.00	109
21 ANAYIBE	0.00	0.00	0.00	0.00	0.00	44,000.00	0.00	0.00	0.00	
22 ANDREA	74,260.00	118,910.00	320,822.00	265,750.00	173,900.00	139,672.00	326,500.00	352,048.00	86,010.00	548
23 ANDREA	0.00	0.00	0.00	115,620.00	115,620.00	0.00	115,620.00	173,700.00	115,620.00	
24 ANGELA BEATRIZ	0.00	0.00	0.00	290,250.00	0.00	0.00	0.00	0.00	416,980.00	
25 ANGELA	262,000.00	0.00	233,000.00	0.00	0.00	0.00	0.00	445,500.00	0.00	
26 ANGELA MARCELA	0.00	0.00	0.00	0.00	142,650.00	0.00	0.00	0.00	0.00	
27 ANGELA	0.00	116,560.00	123,000.00	0.00	244,240.00	190,820.00	98,700.00	83,190.00	0.00	142

- **Informe de Número de Clientes por Producto**

Ese informe nos muestra los clientes que han adquirido o solicitado determinado producto en el periodo establecido.

Ilustración 17. Informe de número de clientes por producto

Grupo	Descripción	Número de Clientes
PRODUCTOS NO FABRICADOS POR LA EMPRESA	AU ACID URIK	220.00
	BIO BIO-CPR	1.00
	BIOF BIOFYL	2.00
	CHIA CHIA GRANO	5.00
	CHO LICH D CHOCO YUSTON	32.00
	COL COLAGENO HIDROLIZADO	51.00
	DB DBT	152.00
	DES DESPARASITOL KIDS	214.00
	DIG DIGEST	38.00
	DR DREVITAN ADULTO	287.00
	DRP DREVITAN PLUS	27.00
	DVE DEVETE	3.00
	EMA EXTRACTO DE MALTA	27.00
	FAR FAROT	141.00
	FEM FEMIGEN	2.00
	FLOT FLORACOL ADULTO	60.00
	GOL GOLESTROL GNS	145.00
	GRA 1 GRADIGESTAR BREVA Y COCO	2.00
	GRA 3 GRADIGESTAR COCO Y PIÑA	3.00
	GRA 4 GRADIGESTAR COCO Y PIÑA	2.00
	GRA 5 GRADIGESTAR PAPAÑA Y ALMENDRA	5.00
	GRA 6 GRADIGESTAR PAPAÑA Y ALMENDRA	4.00
	HGD HGD	3.00

- **Informe de Ventas Detallado por Producto**

En este informe se puede visualizar las devoluciones que se generaron durante el periodo establecido. De igual forma muestra la fecha en que se hizo la devolución y el valor monetario que esta representa.

Ilustración 18. Informe de ventas detallado por producto

Documento	Fecha	Descripción	Dias	Cantidad	Valor Unit	Venta Bruta	Descto	Venta Neta	IVA	GTTotal
DMC 10	18/01/2016	TOLI TOLVOX ADULTO	0.00	-6.00	6,500.00	-39,000.00	0.00	-39,000.00	0.00	-39,000.00
DMC 11	06/05/2016	DB DBT	0.00	-9.00	15,000.00	-135,000.00	0.00	-135,000.00	0.00	-135,000.00
DMC 11	06/05/2016	DES DESPARASITOL KIDS	0.00	-6.00	9,500.00	-57,000.00	0.00	-57,000.00	0.00	-57,000.00
DMC 11	06/05/2016	DR DREVITAN ADULTO	0.00	-16.00	11,000.00	-176,000.00	0.00	-176,000.00	0.00	-176,000.00
DMC 11	06/05/2016	DRP DREVITAN PLUS	0.00	-10.00	16,500.00	-165,000.00	0.00	-165,000.00	0.00	-165,000.00
DMC 11	06/05/2016	FAR FAROT	0.00	-8.00	15,000.00	-120,000.00	0.00	-120,000.00	0.00	-120,000.00
DMC 11	06/05/2016	GOL GOLESTROL GNS	0.00	-7.00	25,000.00	-175,000.00	0.00	-175,000.00	0.00	-175,000.00
DMC 12	31/10/2016	FAR FAROT	0.00	-8.00	15,000.00	-120,000.00	0.00	-120,000.00	0.00	-120,000.00
DMC 13	02/08/2016	DB DBT	0.00	-2.00	15,000.00	-30,000.00	0.00	-30,000.00	0.00	-30,000.00
DMC 13	02/08/2016	DR DREVITAN ADULTO	0.00	-2.00	11,000.00	-22,000.00	0.00	-22,000.00	0.00	-22,000.00
DMC 13	02/08/2016	FAR FAROT	0.00	-1.00	15,000.00	-15,000.00	0.00	-15,000.00	0.00	-15,000.00
DMC 13	02/08/2016	ST STOMEK	0.00	-6.00	11,000.00	-66,000.00	0.00	-66,000.00	0.00	-66,000.00
DMC 14	27/11/2016	DB DBT	0.00	-1.00	15,000.00	-15,000.00	0.00	-15,000.00	0.00	-15,000.00
DMC 14	27/11/2016	GOL GOLESTROL GNS	0.00	-1.00	20,000.00	-20,000.00	0.00	-20,000.00	0.00	-20,000.00
DMC 14	27/11/2016	SF SF'LL CLOROFILA	0.00	-1.00	8,500.00	-8,500.00	0.00	-8,500.00	0.00	-8,500.00
DMC 14	27/11/2016	SFS SF'LL CLOROFILA	0.00	-2.00	11,000.00	-22,000.00	0.00	-22,000.00	0.00	-22,000.00
DMC 15	15/05/2017	GOL GOLESTROL GNS	0.00	-3.00	18,000.00	-54,000.00	0.00	-54,000.00	0.00	-54,000.00
DMC 16	15/05/2017	DB DBT	8.00	-1.00	15,000.00	-15,000.00	0.00	-15,000.00	0.00	-15,000.00
DMC 16	15/05/2017	DR DREVITAN ADULTO	8.00	-4.00	11,000.00	-44,000.00	0.00	-44,000.00	0.00	-44,000.00
DMC 16	15/05/2017	GOL GOLESTROL GNS	8.00	-1.00	25,000.00	-25,000.00	0.00	-25,000.00	0.00	-25,000.00
DMC 16	15/05/2017	OY OVAI	8.00	-1.00	9,000.00	-9,000.00	0.00	-9,000.00	0.00	-9,000.00
DMC 17	15/05/2017	DR DREVITAN ADULTO	0.00	-2.00	11,000.00	-22,000.00	0.00	-22,000.00	0.00	-22,000.00
DMC 17	15/05/2017	OY OVAI	0.00	-3.00	9,000.00	-27,000.00	0.00	-27,000.00	0.00	-27,000.00

- **Ventas Agrupadas Por Vendedor**

En este informe se muestra los vendedores de la ciudad evaluada y los clientes que le compraron a ese vendedor junto con la fecha. De igual forma muestra el monto monetario de esa venta.

Ilustración 19. Ventas agrupadas por vendedor

JENNIFER MARITZA							
Ventas Por Documento entre 1/01/2016 Y 31/12/2017							
Vendedor	CC/Nit	Cliente	Fecha	Documen	Venta Brul	Descuent	Venta Net
JENNIFER MARITZA	11440	IVAN DAVID	18/01/2017	FV 3486	33,000.00	0.00	33,000.00
JENNIFER MARITZA	11440	IVAN DAVID	17/02/2017	FV 3643	33,000.00	0.00	33,000.00
JENNIFER MARITZA	11440	IVAN DAVID	24/03/2017	FV 3784	49,500.00	0.00	49,500.00
JENNIFER MARITZA	79907	WILSON ENRIQUE	18/05/2017	FV 4050	267,900.00	42,864.00	225,036.00
JENNIFER MARITZA	11440	IVAN DAVID	09/06/2017	FV 4156	29,500.00	0.00	29,500.00
JENNIFER MARITZA	11440	IVAN DAVID	27/06/2017	FV 4218	45,400.00	0.00	45,400.00
JENNIFER MARITZA	11440	IVAN DAVID	11/08/2017	FV 4473	102,100.00	6,126.00	95,974.00
JENNIFER MARITZA	11440	IVAN DAVID	28/08/2017	FV 4557	103,400.00	8,204.00	97,196.00
JENNIFER MARITZA	11440	IVAN DAVID	18/09/2017	FV 4701	65,800.00	5,139.00	60,661.00
ADRIANA PATRICIA	51870	ROSA ALFA	12/01/2016	FV 1697	106,000.00	0.00	106,000.00
ADRIANA PATRICIA	52951	CAROLINA	12/01/2016	FV 1699	70,000.00	4,200.00	65,800.00
ADRIANA PATRICIA	50986	MARIA CANDELARIA	12/01/2016	FV 1699	108,000.00	6,480.00	101,520.00
ADRIANA PATRICIA	10324	JEISON ALBERTO	13/01/2016	FV 1700	166,000.00	16,600.00	149,400.00
ADRIANA PATRICIA	52177	SANDRA LILIANA	13/01/2016	FV 1701	261,000.00	26,100.00	234,900.00
ADRIANA PATRICIA	52272	EBER	13/01/2016	FV 1702	72,500.00	4,350.00	68,150.00
ADRIANA PATRICIA	19065	PAUL	13/01/2016	FV 1703	162,800.00	11,000.00	171,000.00
ADRIANA PATRICIA	52334	SUSEN YALIE	13/01/2016	FV 1708	107,000.00	6,420.00	100,580.00
ADRIANA PATRICIA	51772	NIUSA	14/01/2016	FV 1711	156,000.00	15,600.00	140,400.00
ADRIANA PATRICIA	52682	MARILEZ	14/01/2016	FV 1712	117,000.00	7,020.00	109,980.00
ADRIANA PATRICIA	79789	EMERSON DE JESUS	14/01/2016	FV 1720	44,000.00	0.00	44,000.00
ADRIANA PATRICIA	52061	JAKELINE	14/01/2016	FV 1721	84,500.00	5,070.00	79,430.00
ADRIANA PATRICIA	57985	SONIA LUCIA	15/01/2016	FV 1725	117,500.00	6,560.00	111,000.00
ADRIANA PATRICIA	28533	TILCA MARIA	15/01/2016	FV 1726	189,000.00	18,900.00	170,100.00
ADRIANA PATRICIA	43700	FLORES	15/01/2016	FV 1727	104,000.00	0.00	104,000.00

- **Ventas por Vendedor Agrupado por Producto**

En este informe podemos visualizar el producto que cada vendedor vendió, la cantidad de este, la fecha y el monto que representa.

Ilustración 20. Ventas por vendedor agrupado por producto

JENNIFER MARITZA REYES						
Ventas por Vendedor Agrupado Por Producto Entre 1/01/2016 Y 31/12/2017						
Nombres	Producto	Fecha	Documen	Cantid	Total Ingres	Precio Promed
ADRIANA PATRICIA	AU ACID URIC	12/01/2016	FV 1699	12.00	101,520.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	15/01/2016	FV 1725	9.00	69,900.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	19/01/2016	FV 1741	3.00	27,000.00	9,000.00
ADRIANA PATRICIA	AU ACID URIC	19/01/2016	FV 1745	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	19/01/2016	FV 1746	4.00	32,400.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	19/01/2016	FV 1748	7.00	59,220.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	20/01/2016	FV 1762	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	22/01/2016	FV 1773	6.00	50,760.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	25/01/2016	FV 1778	12.00	97,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	25/01/2016	FV 1780	6.00	50,760.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	25/01/2016	FV 1787	12.00	97,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	26/01/2016	FV 1788	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	27/01/2016	FV 1792	6.00	48,600.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	27/01/2016	FV 1794	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	29/01/2016	FV 1789	12.00	97,200.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	31/01/2016	FV 1806	4.00	32,400.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	01/02/2016	FV 1863	1.00	8,460.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	01/02/2016	FV 1868	6.00	54,000.00	9,000.00
ADRIANA PATRICIA	AU ACID URIC	04/02/2016	FV 1823	2.00	16,920.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	04/02/2016	FV 1824	4.00	32,400.00	8,100.00
ADRIANA PATRICIA	AU ACID URIC	05/02/2016	FV 1831	5.00	42,300.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	05/02/2016	FV 1832	3.00	25,380.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	05/02/2016	FV 1836	3.00	25,380.00	8,460.00
ADRIANA PATRICIA	AU ACID URIC	05/02/2016	FV 1837	6.00	48,600.00	8,100.00

8.2.2 Descripción de los Datos

Se cuenta con una gran cantidad de registros de ventas sumando ambas ciudades son 18.000 aproximadamente. Los tipos de dato son alfanuméricos ya que cuentan con fechas, nombres, precios, etc.

8.2.3 Exploración de los Datos

En cuanto a la exploración de los datos, se realizaron diferentes tipos de gráficos en Excel para conocer a fondo la información con la que disponía la empresa. Los gráficos se realizaron de la siguiente manera:

Ilustración 21. Monto total por productos

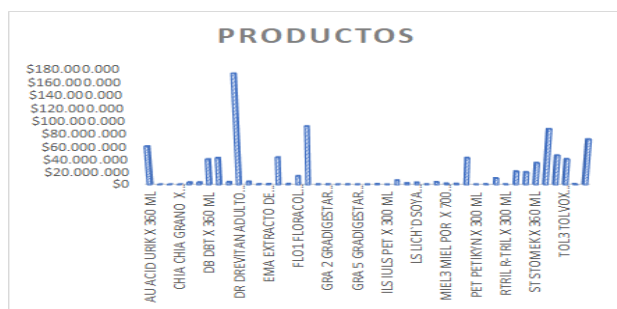


Ilustración 22. Porcentaje de las ventas totales por vendedor

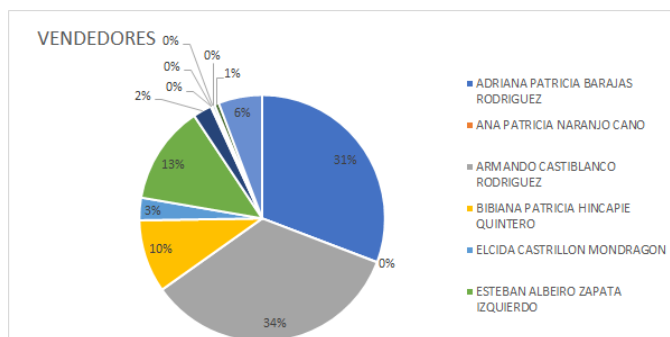


Ilustración 23. Monto total de las ventas en los dos años

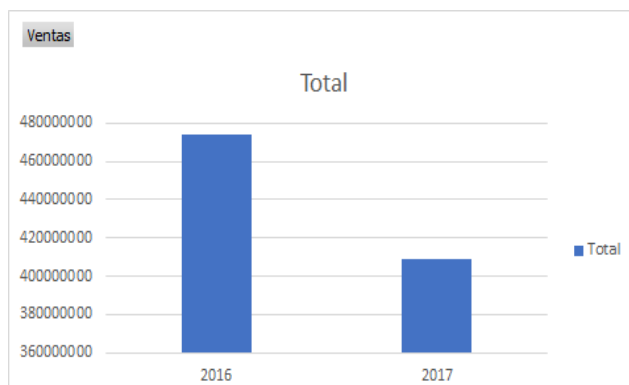
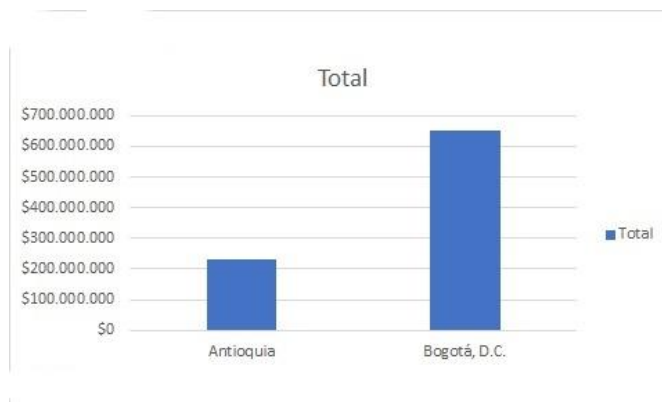


Ilustración 24. Monto total de las ventas en las dos ciudades



8.2.4 Verificar la Calidad de los Datos

Se encontraron datos en donde fueron registradas las ventas, pero en la columna cantidad y en las columnas de precios aparecen en 0.00, estas son ventas anuladas. De igual forma, aparece la cantidad registrada pero las columnas de precios en 0.00, estos son obsequios a los clientes.

Ilustración 25. Informe ventas por vendedor agrupado por producto con ventas anuladas

JENNIFER MARITZA						
Ventas por Vendedor Agrupado Por Producto Entre 1/01/2016 Y 31/12/2017						
Nombres	Productos	Fecha	Documen	Cantida	Total_Ingres	Precio Promed
392 ADRIANA PATRICIA	AU ACID URIK	22/04/2016 FV 2212		0.00	0.00	0.00
304 ADRIANA PATRICIA	DES DESPARASITOL KIDS	14/10/2016 FV 3098		0.00	0.00	0.00
1175 ADRIANA PATRICIA	DR DREVITAN ADULTO	18/01/2016 FV 1735		0.00	0.00	0.00
1188 ADRIANA PATRICIA	DR DREVITAN ADULTO	19/01/2016 FV 1751		0.00	0.00	0.00
1184 ADRIANA PATRICIA	DR DREVITAN ADULTO	20/01/2016 FV 1763		0.00	0.00	0.00
1280 ADRIANA PATRICIA	DR DREVITAN ADULTO	05/04/2016 FV 2119		0.00	0.00	0.00
1505 ADRIANA PATRICIA	DR DREVITAN ADULTO	22/05/2016 FV 2999		0.00	0.00	0.00
2032 ADRIANA PATRICIA	DR DREVITAN PLUS	22/05/2016 FV 2999		0.00	0.00	0.00
2316 ADRIANA PATRICIA	GOL GOLESTROL GNS	14/03/2016 FV 2019		0.00	0.00	0.00
2737 ADRIANA PATRICIA	OV OVAI	05/04/2016 FV 2119		0.00	0.00	0.00
3165 ADRIANA PATRICIA	SF SPYLL CLOROFLA	25/01/2016 FV 1779		0.00	0.00	0.00
3188 ADRIANA PATRICIA	SF SPYLL CLOROFLA	14/03/2016 FV 2019		0.00	0.00	0.00
3199 ADRIANA PATRICIA	SF SPYLL CLOROFLA	13/04/2016 FV 2152		0.00	0.00	0.00
3389 ADRIANA PATRICIA	SFS SPYLL CLOROFLA	25/01/2016 FV 1779		0.00	0.00	0.00
3635 ADRIANA PATRICIA	ST STOMEK	05/04/2016 FV 2119		0.00	0.00	0.00
3647 ADRIANA PATRICIA	ST STOMEK	31/01/2017 FV 3556		0.00	0.00	0.00
3727 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	18/01/2016 FV 1735		0.00	0.00	0.00
3731 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	19/01/2016 FV 1751		0.00	0.00	0.00
3734 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	20/01/2016 FV 1763		0.00	0.00	0.00
3741 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	25/01/2016 FV 1779		0.00	0.00	0.00
3797 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	14/03/2016 FV 2019		0.00	0.00	0.00
3818 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	05/04/2016 FV 2119		0.00	0.00	0.00
3852 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	28/04/2016 FV 2260		0.00	0.00	0.00
4082 ADRIANA PATRICIA	TOLI TOLVOX ADULTO	15/11/2016 FV 3225		0.00	0.00	0.00

Ilustración 26. Informe ventas por vendedor agrupado por producto con obsequios a clientes

JENNIFER MARITZA REYES						
Ventas por Vendedor Agrupado Por Producto Entre 1/01/2016 Y 31/12/2017						
Nombres	Productos	Fecha	Documen	Cantida	Total_Ingres	Precio Promed
349 ADRIANA PATRICIA	AU ACID URIK	23/05/2017 FV 4067		1.00	0.00	0.00
408 ADRIANA PATRICIA	AU ACID URIK	02/09/2017 FV 4432		1.00	0.00	0.00
496 ADRIANA PATRICIA	CHO LICH'D CHOCO YUSTON	31/09/2017 FV 4577		1.00	0.00	0.00
576 ADRIANA PATRICIA	DB DBT	26/04/2016 FV 2236		1.00	0.00	0.00
655 ADRIANA PATRICIA	DB DBT	20/01/2017 FV 3504		2.00	0.00	0.00
669 ADRIANA PATRICIA	DB DBT	03/03/2017 FV 3692		1.00	0.00	0.00
687 ADRIANA PATRICIA	DB DBT	25/04/2017 FV 3942		1.00	0.00	0.00
708 ADRIANA PATRICIA	DB DBT	29/09/2017 FV 4240		1.00	0.00	0.00
711 ADRIANA PATRICIA	DB DBT	30/06/2017 FV 4257		1.00	0.00	0.00
728 ADRIANA PATRICIA	DB DBT	27/07/2017 FV 4400		1.00	0.00	0.00
836 ADRIANA PATRICIA	DES DESPARASITOL KIDS	27/05/2016 FV 2414		2.00	0.00	0.00
900 ADRIANA PATRICIA	DES DESPARASITOL KIDS	06/10/2016 FV 3082		2.00	0.00	0.00
932 ADRIANA PATRICIA	DES DESPARASITOL KIDS	21/11/2016 FV 3259		2.00	0.00	0.00
937 ADRIANA PATRICIA	DES DESPARASITOL KIDS	30/11/2016 FV 3320		1.00	0.00	0.00
939 ADRIANA PATRICIA	DES DESPARASITOL KIDS	05/12/2016 FV 3334		1.00	0.00	0.00
1042 ADRIANA PATRICIA	DES DESPARASITOL KIDS	05/06/2017 FV 4123		2.00	0.00	0.00
1149 ADRIANA PATRICIA	DIG DIGEST	30/06/2017 FV 4257		1.00	0.00	0.00
1150 ADRIANA PATRICIA	DIG DIGEST	05/10/2017 FV 4754		1.00	0.00	0.00
1319 ADRIANA PATRICIA	DR DREVITAN ADULTO	25/04/2016 FV 2228		2.00	0.00	0.00
1393 ADRIANA PATRICIA	DR DREVITAN ADULTO	29/06/2016 FV 2563		2.00	0.00	0.00
1412 ADRIANA PATRICIA	DR DREVITAN ADULTO	13/07/2016 FV 2622		1.00	0.00	0.00
1564 ADRIANA PATRICIA	DR DREVITAN ADULTO	17/11/2016 FV 3243		1.00	0.00	0.00
1573 ADRIANA PATRICIA	DR DREVITAN ADULTO	21/11/2016 FV 3259		2.00	0.00	0.00
1579 ADRIANA PATRICIA	DR DREVITAN ADULTO	23/11/2016 FV 3278		1.00	0.00	0.00

8.3 FASE 3: PREPARACIÓN DE LOS DATOS

8.3.1 Seleccionar los datos

Selección de elementos: Utilizaremos los 18.000 registros de las ventas de las dos ciudades que corresponden a los años de 2016 hasta octubre del 2017. Utilizando solamente los informes de:

Ilustración 27. Ventas agrupadas por vendedor, selección de elementos

Vendedor	CC/NIT	Cliente	Fecha	Documento	Venta Bruta	Descuento	Venta Net
JENNIFER MARITZA	11440	IVAN DAVID	19/01/2017	FV 3436	33.000,00	0,00	33.000,00
JENNIFER MARITZA	11440	IVAN DAVID	17/02/2017	FV 3643	33.000,00	0,00	33.000,00
JENNIFER MARITZA	11440	IVAN DAVID	24/02/2017	FV 3784	48.500,00	0,00	48.500,00
JENNIFER MARITZA	79807	WILSON ENRIQUE	18/05/2017	FV 4050	267.900,00	42.884,00	225.036,00
JENNIFER MARITZA	11440	IVAN DAVID	09/06/2017	FV 4156	29.500,00	0,00	29.500,00
JENNIFER MARITZA	11440	IVAN DAVID	27/06/2017	FV 4216	45.400,00	0,00	45.400,00
JENNIFER MARITZA	11440	IVAN DAVID	11/08/2017	FV 4473	102.100,00	6.126,00	95.974,00
JENNIFER MARITZA	11440	IVAN DAVID	28/08/2017	FV 4557	103.400,00	6.204,00	97.196,00
JENNIFER MARITZA	11440	IVAN DAVID	18/09/2017	FV 4701	85.500,00	5.130,00	80.370,00
ADRIANA PATRICIA	51870	ROSA ALBA	12/01/2016	FV 1687	106.000,00	0,00	106.000,00
ADRIANA PATRICIA	52951	CAROLINA	12/01/2016	FV 1638	70.000,00	4.200,00	65.800,00
ADRIANA PATRICIA	20686	MARIA CANDELARIA	12/01/2016	FV 1699	108.000,00	6.480,00	101.520,00
ADRIANA PATRICIA	10324	JEISON ALBERTO	13/01/2016	FV 1700	166.000,00	16.600,00	149.400,00
ADRIANA PATRICIA	52117	SANDRA LILIANA	13/01/2016	FV 1701	261.900,00	26.190,00	234.900,00
ADRIANA PATRICIA	12272	EBER	13/01/2016	FV 1702	72.500,00	4.350,00	68.150,00
ADRIANA PATRICIA	19365	RAUL	13/01/2016	FV 1703	182.000,00	11.000,00	171.000,00
ADRIANA PATRICIA	52324	SUSEN YALILE	13/01/2016	FV 1708	107.000,00	6.420,00	100.580,00
ADRIANA PATRICIA	51772	NILSEA	14/01/2016	FV 1711	156.000,00	15.600,00	140.400,00
ADRIANA PATRICIA	52882	MARLUZ	14/01/2016	FV 1712	117.000,00	7.020,00	109.980,00
ADRIANA PATRICIA	79789	EMERSON DE JESUS	14/01/2016	FV 1720	44.000,00	0,00	44.000,00
ADRIANA PATRICIA	52081	JAKELINE	14/01/2016	FV 1721	84.500,00	5.070,00	79.430,00
ADRIANA PATRICIA	71895	SONIA LUCIA	15/01/2016	FV 1725	117.500,00	6.500,00	111.000,00
ADRIANA PATRICIA	28033	TILCIA MARIA	15/01/2016	FV 1726	188.000,00	18.800,00	170.100,00
ADRIANA PATRICIA	43700	FLORES	15/01/2016	FV 1727	104.000,00	0,00	104.000,00

Ilustración 28. Ventas agrupadas por producto, selección de elementos

Nombres	Productos	Fecha	Documento	Cantidad	Total Ingresos	Precio Promedio
ADRIANA PATRICIA	AJU ACID URUK	12/01/2016	FV 1659	12,00	101.520,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	15/01/2016	FV 1729	8,00	64.800,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	19/01/2016	FV 1741	3,00	27.000,00	9.000,00
ADRIANA PATRICIA	AJU ACID URUK	19/01/2016	FV 1745	2,00	16.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	19/01/2016	FV 1746	4,00	32.400,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	19/01/2016	FV 1748	7,00	56.700,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	20/01/2016	FV 1762	2,00	16.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	22/01/2016	FV 1773	6,00	50.760,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	25/01/2016	FV 1778	12,00	97.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	25/01/2016	FV 1780	6,00	50.760,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	26/01/2016	FV 1787	12,00	97.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	26/01/2016	FV 1788	2,00	16.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	27/01/2016	FV 1792	6,00	48.600,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	27/01/2016	FV 1794	2,00	16.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	29/01/2016	FV 1799	12,00	97.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	31/01/2016	FV 1806	4,00	32.400,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	01/02/2016	FV 1803	1,00	8.460,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	01/02/2016	FV 1808	6,00	50.760,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	04/02/2016	FV 1823	2,00	16.200,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	04/02/2016	FV 1824	4,00	32.400,00	8.100,00
ADRIANA PATRICIA	AJU ACID URUK	05/02/2016	FV 1831	5,00	42.300,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	05/02/2016	FV 1832	3,00	25.380,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	05/02/2016	FV 1836	3,00	25.380,00	8.460,00
ADRIANA PATRICIA	AJU ACID URUK	05/02/2016	FV 1837	6,00	48.600,00	8.100,00

Selección de atributos: Utilizaremos del informe de Ventas agrupadas por vendedor las columnas de Vendedor, CC/Nit y Cliente y del informe de Ventas por vendedor agrupadas por productos tomaremos las columnas de Producto, Fecha, Documento, Cantidad, Total Ingresos y Precio Promedio.

Ilustración 29. Ventas agrupadas por vendedor, selección de atributos

JENNIFER MARITZA							
Ventas Por Documento entre 1/01/2016 Y 31/12/2017							
Vendedor	CC/Nit	Cliente	Fecha	Documen	Venta Bruta	Descuentos	Venta Net
JENNIFER MARITZA	11440	IVAN DAVID	19/01/2017	FV 3496	33,000.00	0.00	33,000.00
JENNIFER MARITZA	11440	IVAN DAVID	17/02/2017	FV 3543	33,000.00	0.00	33,000.00
JENNIFER MARITZA	11440	IVAN DAVID	24/03/2017	FV 3784	49,500.00	0.00	49,500.00
JENNIFER MARITZA	79807	WILSON ENRIQUE	18/05/2017	FV 4050	267,900.00	42,664.00	225,236.00
JENNIFER MARITZA	11440	IVAN DAVID	09/06/2017	FV 4156	29,500.00	0.00	29,500.00
JENNIFER MARITZA	11440	IVAN DAVID	27/06/2017	FV 4218	45,400.00	0.00	45,400.00
JENNIFER MARITZA	11440	IVAN DAVID	11/08/2017	FV 4473	102,100.00	6,126.00	95,974.00
JENNIFER MARITZA	11440	IVAN DAVID	28/08/2017	FV 4557	103,400.00	6,204.00	97,196.00
JENNIFER MARITZA	11440	IVAN DAVID	18/09/2017	FV 4701	85,500.00	5,130.00	80,370.00
ADRIANA PATRICIA	51670	ROSA AURIA	12/01/2016	FV 1697	106,000.00	0.00	106,000.00
ADRIANA PATRICIA	52951	CAROLINA	12/01/2016	FV 1696	70,000.00	4,200.00	65,800.00
ADRIANA PATRICIA	20686	MARIA CANDELARIA	12/01/2016	FV 1699	108,000.00	6,480.00	101,520.00
ADRIANA PATRICIA	10324	JEISON ALBERTO	13/01/2016	FV 1700	166,000.00	16,600.00	149,400.00
ADRIANA PATRICIA	52177	SANDRA LILIANA	13/01/2016	FV 1701	261,000.00	26,100.00	234,900.00
ADRIANA PATRICIA	12272	EBER	13/01/2016	FV 1702	72,500.00	4,350.00	68,150.00
ADRIANA PATRICIA	19065	RAUL	13/01/2016	FV 1703	182,000.00	11,000.00	171,000.00
ADRIANA PATRICIA	52334	SUSEN YALILE	13/01/2016	FV 1708	107,000.00	6,420.00	100,580.00
ADRIANA PATRICIA	51772	NUBIA	14/01/2016	FV 1711	156,000.00	15,600.00	140,400.00
ADRIANA PATRICIA	52982	MARILUZ	14/01/2016	FV 1712	117,000.00	7,820.00	109,380.00
ADRIANA PATRICIA	79785	EMERSON DE JESUS	14/01/2016	FV 1720	44,000.00	0.00	44,000.00
ADRIANA PATRICIA	52061	JAKELINE	14/01/2016	FV 1721	84,500.00	5,070.00	79,430.00
ADRIANA PATRICIA	79789	SONIA LUCIA	15/01/2016	FV 1725	117,500.00	6,500.00	111,000.00
ADRIANA PATRICIA	58033	TILCIA MARIA	15/01/2016	FV 1726	189,000.00	18,900.00	170,100.00
ADRIANA PATRICIA	43700	FLORES ROJAS	15/01/2016	FV 1727	104,000.00	0.00	104,000.00

Ilustración 30. Ventas agrupadas por producto, selección de atributos

JENNIFER MARITZA						
Ventas por Vendedor Agrupado Por Producto Entre 1/01/2016 Y 31/12/2017						
Nombres	Productos	Fecha	Documen	Cantide	Total Ingres	Precio Promed
ADRIANA PATRICIA	AJ ACID URUK	12/01/2016	FV 1699	12.00	101,520.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	15/01/2016	FV 1729	8.00	64,800.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	19/01/2016	FV 1741	3.00	27,000.00	9,000.00
ADRIANA PATRICIA	AJ ACID URUK	19/01/2016	FV 1745	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	19/01/2016	FV 1746	4.00	33,840.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	19/01/2016	FV 1748	7.00	59,220.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	20/01/2016	FV 1762	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	22/01/2016	FV 1773	6.00	50,760.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	25/01/2016	FV 1778	12.00	97,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	25/01/2016	FV 1780	6.00	50,760.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	26/01/2016	FV 1787	12.00	97,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	26/01/2016	FV 1788	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	27/01/2016	FV 1792	6.00	48,600.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	27/01/2016	FV 1794	2.00	16,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	29/01/2016	FV 1799	12.00	97,200.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	31/01/2016	FV 1806	4.00	32,400.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	01/02/2016	FV 1803	1.00	8,460.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	01/02/2016	FV 1808	6.00	54,000.00	9,000.00
ADRIANA PATRICIA	AJ ACID URUK	04/02/2016	FV 1823	2.00	16,920.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	04/02/2016	FV 1824	4.00	32,400.00	8,100.00
ADRIANA PATRICIA	AJ ACID URUK	05/02/2016	FV 1831	5.00	42,300.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	05/02/2016	FV 1832	3.00	25,380.00	8,460.00
ADRIANA PATRICIA	AJ ACID URUK	05/02/2016	FV 1836	3.00	27,000.00	9,000.00
ADRIANA PATRICIA	AJ ACID URUK	05/02/2016	FV 1837	6.00	48,600.00	8,100.00

8.3.2 Limpiar los datos

Se hizo necesario llevar a cabo este proceso puesto que en los informes de donde obtuvimos los datos, existían espacios en 0 los cuales pueden dar resultados incorrectos a la hora de realizar el análisis de los datos.

De los datos que la empresa nos proporcionó encontramos dos tipos de errores como lo mencionamos en la sección de verificar la calidad de los datos, por consiguiente, eliminamos esos datos que hacían referencia a ventas anuladas y a obsequios que la empresa daba de acuerdo al monto vendido.

8.3.3 Construir los datos

Para la realización del Data Set requerimos agregar la columna ciudad, para la identificación de las ventas de cada ciudad (Bogotá y Medellín).

8.3.4 Integración de datos

Como ya mencionamos en la selección de datos, tomamos las diferentes columnas de los dos informes y los integramos en un Data Set que comprenden los siguientes campos:

Documento, Fecha, Ciudad, Vendedor, CC/Nit, Cliente, Productos, Cantidad, Total_Ingresos y Precio_Promedio; quedando de la siguiente manera:

Ilustración 31. Data Set Grupo Natural Ser

Documento	Fecha	Ciudad	Vendedor	CC/NIT	Ciudad	Productos	Cantidad	Total Ingreso	Precio Promovido
112 FV 233	05/01/2016	MEDELLIN	BIBIANA PATRICIA	71692	PODER NATURAL	AU ACID URIK	8.00	72.000.00	9000.00
113 FV 233	05/01/2016	MEDELLIN	BIBIANA PATRICIA	71692	PODER NATURAL	OV OVA	6.00	54.000.00	9000.00
114 FV 233	05/01/2016	MEDELLIN	BIBIANA PATRICIA	71692	PODER NATURAL	SF SPYLL CLOROFILA	5.00	42.500.00	8500.00
115 FV 233	05/01/2016	MEDELLIN	BIBIANA PATRICIA	71692	PODER NATURAL	ST STOMEK	8.00	56.000.00	11200.00
116 FV 233	05/01/2016	MEDELLIN	BIBIANA PATRICIA	71692	PODER NATURAL	TOL1 TOLVIX ADULTO	8.00	52.000.00	6500.00
117 FV 233	05/01/2016	MEDELLIN	BIBIANA PATRICIA	71692	PODER NATURAL	ZAR ZARZAN	12.00	84.000.00	7000.00
118 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	AU ACID URIK	2.00	18.000.00	9000.00
119 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	DES DESPARASITOL KIDS	1.00	9.500.00	9500.00
120 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	OV OVA	2.00	18.000.00	9000.00
121 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	SF SPYLL CLOROFILA	2.00	17.000.00	8500.00
122 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	TOL1 TOLVIX ADULTO	3.00	19.500.00	6500.00
123 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	TOL2 TOLVIX NIÑOS	3.00	19.500.00	6500.00
124 FV 234	07/01/2016	MEDELLIN	BIBIANA PATRICIA	71605	T N LA BOTANICA	ZAR ZARZAN	6.00	42.000.00	7000.00
125 FV 235	12/01/2016	MEDELLIN	BIBIANA PATRICIA	98540	T N EL CACIQUE	DR DREVITAN ADULTO	3.00	28.501.74	9500.58
126 FV 235	12/01/2016	MEDELLIN	BIBIANA PATRICIA	98540	T N EL CACIQUE	ZAR ZARZAN	4.00	26.320.00	6580.00
127 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	AU ACID URIK	3.00	26.380.00	8460.00
128 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	DR DBT	1.00	14.100.00	14100.00
129 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	DES DESPARASITOL KIDS	3.00	26.730.00	8950.00
130 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	DR DREVITAN ADULTO	6.00	57.003.48	9500.58
131 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	FAR FAROT	1.00	14.100.00	14100.00
132 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	GOL GOLESTROL GNS	1.00	23.500.00	23500.00
133 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	OV OVA	2.00	16.920.00	8460.00
134 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	SF SPYLL CLOROFILA	1.00	7.990.00	7990.00
135 FV 236	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	ST STOMEK	1.00	11.280.00	11280.00
136 FV 237	12/01/2016	MEDELLIN	BIBIANA PATRICIA	80001	MOLI S A	DR DREVITAN ADULTO	3.00	28.501.74	9500.58
137 FV 238	12/01/2016	MEDELLIN	BIBIANA PATRICIA	42536	T N LA SEGOVIANA	DES DESPARASITOL KIDS	1.00	8.930.00	8930.00
138 FV 238	12/01/2016	MEDELLIN	BIBIANA PATRICIA	42536	T N LA SEGOVIANA	GOL GOLESTROL GNS	2.00	47.000.00	23500.00
139 FV 238	12/01/2016	MEDELLIN	BIBIANA PATRICIA	42536	T N LA SEGOVIANA	OV OVA	1.00	25.390.00	8460.00

8.4 FASE 4: DESARROLLO (MODELADO)

8.4.1 Diseño Conceptual del Data Warehouse

Como paso inicial para la construcción del Data Warehouse es necesario realizar un diseño con el cual organizaremos los datos en una tabla de hechos y sus respectivas dimensiones, estas tablas contienen los atributos más relevantes de los informes de ventas que agrupamos en el Data Set, como necesitamos que el diseño sea similar a un modelo entidad - relación requerimos de la construcción de algunos campos para que esté normalizado y funcione correctamente.

El diseño conceptual lo construimos de la siguiente manera:

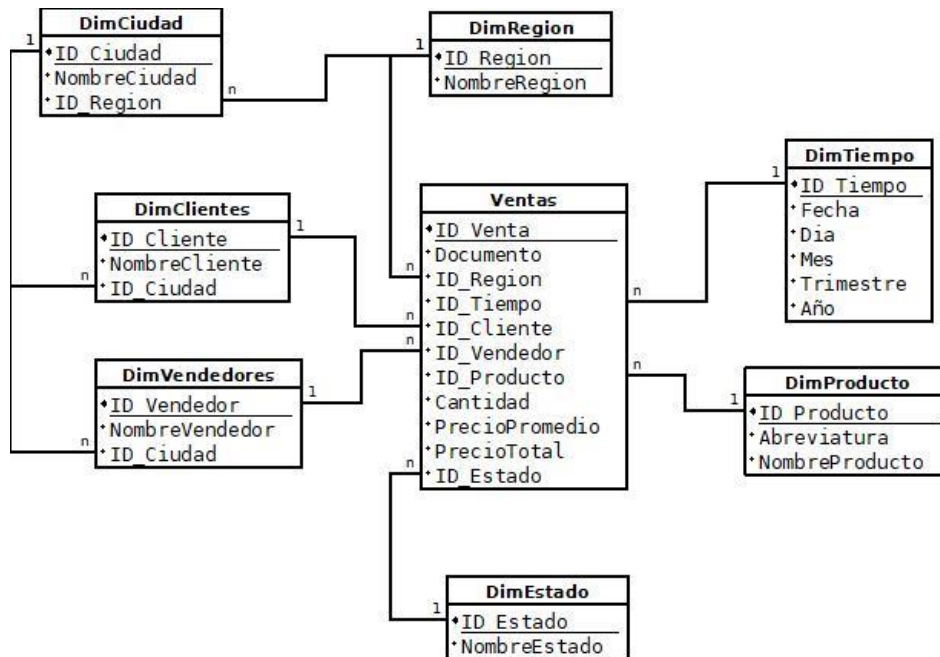
- Una tabla de hechos que es la que contendrá los datos que requerimos para realizar el análisis de las ventas, en ella construimos el campo de ID ventas.
- Las tablas de dimensiones que contienen los detalles de la tabla de hechos, estas son:
 - **Región:** Construimos esta dimensión para identificar las regiones en donde se realizaron las ventas.
 - **Ciudad:** Construimos esta dimensión para identificar en qué ciudad se encuentran los clientes y vendedores.
 - **Clientes:** Tomamos los campos ID Cliente y Nombre Cliente de los informes y construimos el ID Ciudad que relaciona cada cliente con la ciudad a donde pertenece.
 - **Vendedores:** Tomamos el campo de Nombre Vendedor del informe y construimos el ID Vendedor e ID Ciudad para relacionar cada vendedor con la ciudad donde distribuye los productos.
 - **Producto:** Tomamos el campo Nombre Producto como en el aparecía su abreviatura, decidimos hacer un nuevo campo Abreviatura y el ID Producto.

- **Tiempo:** Tomamos el campo Fecha y construimos a partir de este, el ID Tiempo, Día, Mes, Trimestre y Año.
- **Estado:** Construimos esta dimensión para poder identificar las devoluciones y ventas dentro de la tabla de hechos.

Todas las relaciones de las tablas de dimensiones a la tabla de hechos son de uno a muchos, por lo cual la tabla de hechos se lleva la llave primaria de las tablas de dimensiones.

A continuación, se muestra el diseño conceptual para el desarrollo de la bodega de datos:

Ilustración 32. Diseño Conceptual del Data Warehouse



Diccionario de Datos

Luego de ello requerimos de un diccionario de datos, para poder identificar la estructura y formato de los datos. Definimos el nombre de los campos, tipo de datos y su tamaño.

Tablas de Dimensiones

- **Tabla Región:** Contiene información de las regiones en donde se realizan las ventas.

Tabla 3. Diccionario de Datos Dimensión Región

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdRegion	Integer	Llave Primaria y código de la región	Si
NombreRegion	Varchar (50)	Nombre de la región	No

- **Tabla Ciudad:** Contiene información de las ciudades en donde se localizan los clientes y vendedores.

Tabla 4. Diccionario de Datos Dimensión Ciudad

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdCiudad	Integer	Llave Primaria y código de la ciudad	Si
NombreCiudad	Varchar (50)	Nombre de la ciudad	No
IdRegion	Integer	Código de la región a la que pertenece cada ciudad	No

- **Tabla Cliente:** Contiene información de cada uno de los clientes a quienes se les realiza la venta de los productos.

Tabla 5. Diccionario de Datos Dimensión Cliente

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdCliente	Bigint	Llave Primaria y numero de cedula del cliente	Si
NombreCliente	Varchar (200)	Nombre del cliente	No
IdCiudad	Integer	Código de la ciudad a la que pertenece cada cliente	No

- **Tabla Producto:** Contiene información de cada uno de los productos.

Tabla 6. Diccionario de Datos Dimensión Producto

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdProducto	Integer	Llave Primaria y código del producto	Si
Abreviatura	Varchar (50)	Abreviatura del nombre de cada producto	No
NombreProducto	Varchar (100)	Nombre del producto	No

- **Tabla Tiempo:** Contiene información del transcurso de los años en que se han realizado las ventas.

Tabla 7. Diccionario de Datos Dimensión Tiempo

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdTiempo	Integer	Llave Primaria y código de la fecha	Si
Fecha	Date	Fecha	No
Dia	Integer	Cada uno de los días del mes	No
Mes	Varchar (50)	Cada uno de los meses	No
Trimestre	Integer	Cada uno de los trimestres	No
Año	Integer	Cada uno de los años	No

- **Tabla Vendedor:** Contiene información de cada uno de los vendedores quienes han realizado la venta de los productos.

Tabla 8. Diccionario de Datos Dimensión Vendedor

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdVendedor	Integer	Llave Primaria y código del vendedor	Si
NombreVendedor	Varchar (200)	Nombre del vendedor	No
IdCiudad	Integer	Código de la ciudad a la que pertenece cada vendedor	No

- **Tabla Estado:** Contiene información del estado de la venta, como: Ventas y/o devoluciones.

Tabla 9. Diccionario de Datos Dimensión Estado

Campo-Nombre	Tipo/Tamaño	Descripción	Llave Primaria
IdEstado	Integer	Llave Primaria y código del estado	Si
NombreEstado	Varchar (50)	Nombre del estado	No

- Datos para la tabla de dimensión ciudad:

Ilustración 34. Datos Dimensión Ciudad

IdCiudad	NombreCiudad	IdRegion
11001	Bogotá, D.C.	11
05001	Medellin	05

- Datos para la tabla de dimensión clientes:

Ilustración 35. Datos Dimensión Clientes

IdCliente	NombreCliente	IdCiudad
43700	FLORES	11001
10	LILIA	11001
10000	MANUELA	05001
10026	JASBLEYDY	11001
10055	JOSE LEONARDO	11001
10101	ANGELICA MARIA	11001
10102	JUAN SEBASTIAN	11001
10123	LYDA JOHANNA	11001
10135	MONICA	11001
10142	MARLIETH	11001
10155	JOSE RIGOBERTO	11001
10171	T.N. EL ROMEDAL	05001
10184	DIANA MARCELA	11001
10184	DIANA	11001
10186	PEDRO	05001
10190	LUIS ALEJANDRO	11001
10190	LAURA ANDREA	11001
10204	T.N. FUERZA Y VIGOR	05001
10204	T.N. NATURALES SOL DE VIDA	05001
10207	NANCY ESTEFANIA	11001
10223	FRANCY	11001
10223	ROBINSON STIVEN	11001

- Datos para la tabla de dimensión productos:

Ilustración 36. Datos Dimensión Productos

IdProducto	Abreviatura	NombreProducto
1	AU	AU ACID URIK
2	BIO	BIO BIO-CIR
3	BIOF	BIOF BIOFYL
4	CHIA	CHIA CHIA GRANO
5	CHO	CHO LICH'D CHOCO YUSTON
6	COL	COL COLAGENO HIDROLIZADO
7	DB	DB DBT
8	DES	DES DESPARASITOL KIDS
9	DIG	DIG DIGEST
10	DR	DR DREVITAN ADULTO
11	DRP	DRP DREVITAN PLUS
12	DVE	DVE DREVETE
13	EMA	EMA EXTRACTO DE MALTA
14	FAR	FAR FAROT
15	FEM	FEM FEMIGEN
16	FLO1	FLO1 FLORACOL ADULTO
17	GOL	GOL GOLESTROL GNS
18	GRA1	GRA 1 GRADIGESTAR BREVA Y COCO
19	GRA2	GRA 2 GRADIGESTAR BREVA Y COCO
20	GRA3	GRA 3 GRADIGESTAR COCO Y PIÑA
21	GRA4	GRA 4 GRADIGESTAR COCO Y PIÑA
22	GRA5	GRA 5 GRADIGESTAR PAPAYA Y ALMENDRA

- Datos para la tabla de dimensión tiempo:

Ilustración 37. Datos Dimensión Tiempo

IdFecha	Fecha	Dia	Mes	Trimestre	Año
20160101	1/1/2016	1	Enero	1	2016
20160102	2/1/2016	2	Enero	1	2016
20160103	3/1/2016	3	Enero	1	2016
20160104	4/1/2016	4	Enero	1	2016
20160105	5/1/2016	5	Enero	1	2016
20160106	6/1/2016	6	Enero	1	2016
20160107	7/1/2016	7	Enero	1	2016
20160108	8/1/2016	8	Enero	1	2016
20160109	9/1/2016	9	Enero	1	2016
20160110	10/1/2016	10	Enero	1	2016
20160111	11/1/2016	11	Enero	1	2016
20160112	12/1/2016	12	Enero	1	2016
20160113	13/1/2016	13	Enero	1	2016
20160114	14/1/2016	14	Enero	1	2016
20160115	15/1/2016	15	Enero	1	2016
20160116	16/1/2016	16	Enero	1	2016
20160117	17/1/2016	17	Enero	1	2016
20160118	18/1/2016	18	Enero	1	2016
20160119	19/1/2016	19	Enero	1	2016
20160120	20/1/2016	20	Enero	1	2016
20160121	21/1/2016	21	Enero	1	2016
20160122	22/1/2016	22	Enero	1	2016

- Datos para la tabla de hechos ventas:

Ilustración 40. Datos Ventas

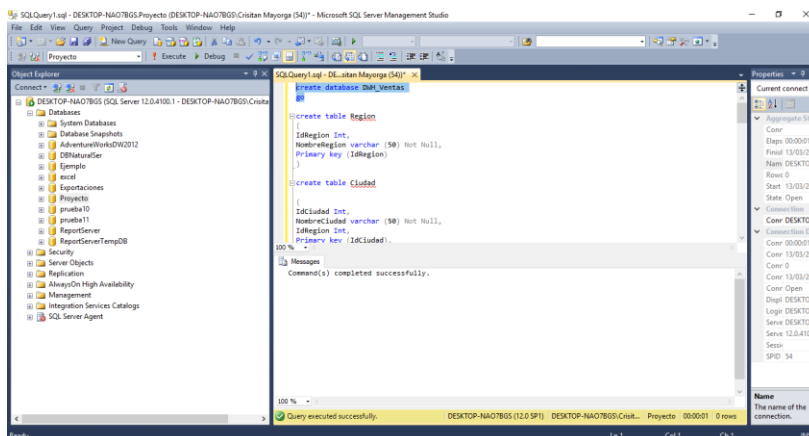
Document	IdRegion	IdFecha	IdCliente	IdVendedor	IdProducto	Cantidad	PrecioPromedi	PrecioTot	IdEstado
110 FV 1000	5	20160914	43275	6	5	1	\$ 24,900	\$ 24,900	2
111 FV 1000	5	20160914	43275	6	8	2	\$ 8,930	\$ 17,860	2
112 FV 1000	5	20160914	43275	6	10	6	\$ 9,501	\$ 57,003	2
113 FV 1000	5	20160914	43275	6	45	2	\$ 6,580	\$ 13,160	2
114 FV 1001	5	20160914	98559	6	10	2	\$ 9,501	\$ 19,001	2
115 FV 1001	5	20160914	98559	6	33	1	\$ 8,460	\$ 8,460	2
116 FV 1001	5	20160914	98559	6	43	2	\$ 6,110	\$ 12,220	2
117 FV 1001	5	20160914	98559	6	45	2	\$ 6,580	\$ 13,160	2
118 FV 1002	5	20160915	70750	6	10	3	\$ 10,107	\$ 30,321	2
119 FV 1002	5	20160915	70750	6	41	6	\$ 6,500	\$ 39,000	2
120 FV 1002	5	20160915	70750	6	43	6	\$ 6,500	\$ 39,000	2
121 FV 1002	5	20160915	70750	6	45	6	\$ 7,000	\$ 42,000	2
122 FV 1003	5	20160915	99434	6	10	1	\$ 9,501	\$ 9,501	2
123 FV 1003	5	20160915	99434	6	14	1	\$ 14,100	\$ 14,100	2
124 FV 1003	5	20160915	99434	6	17	2	\$ 23,500	\$ 47,000	2
125 FV 1003	5	20160915	99434	6	41	2	\$ 6,110	\$ 12,220	2
126 FV 1003	5	20160915	99434	6	43	3	\$ 6,110	\$ 18,330	2
127 FV 1004	5	20160915	70750	6	5	1	\$ 24,900	\$ 24,900	2
128 FV 1005	5	20160915	22095	6	33	2	\$ 8,460	\$ 16,920	2
129 FV 1005	5	20160915	22095	6	40	1	\$ 11,280	\$ 11,280	2
130 FV 1005	5	20160915	22095	6	41	1	\$ 6,110	\$ 6,110	2
131 FV 1005	5	20160915	22095	6	42	1	\$ 6,110	\$ 6,110	2

8.4.2 Construcción del Data Warehouse

Para la construcción del Data Warehouse utilizamos un script SQL. Ver Anexos. Y lo construimos de la siguiente manera:

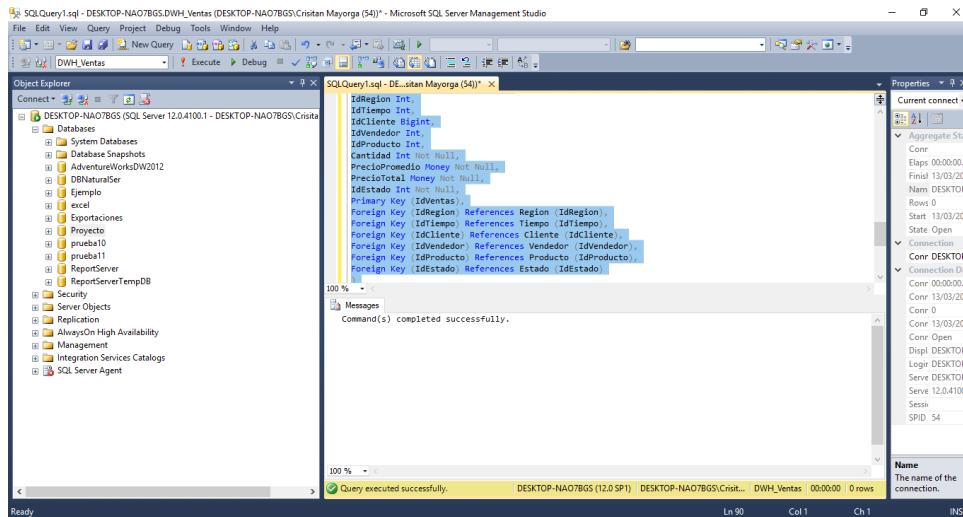
- Creamos la Bodega de Datos de ventas

Ilustración 41. Creación del Data Warehouse en SQL Server



- Creamos las tablas de las dimensiones y las tablas de hechos

Ilustración 42. Creación de las Dimensión y tabla de Hechos en el Data Warehouse



- Visualizamos el esquema entidad-relación de la bodega de datos

Ilustración 43. Creación del Diagrama Entidad - Relación

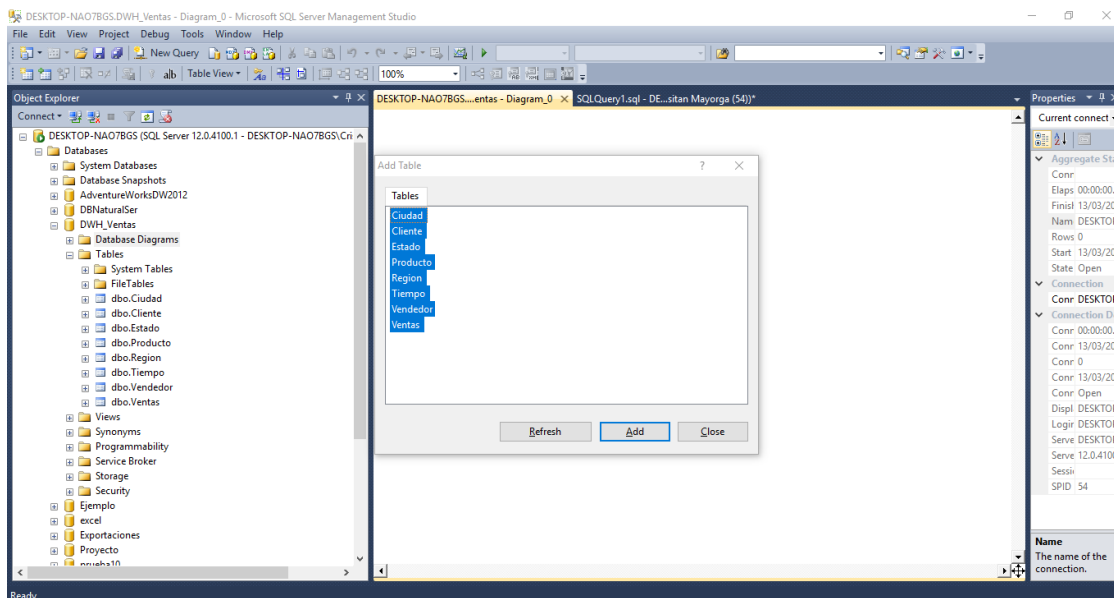
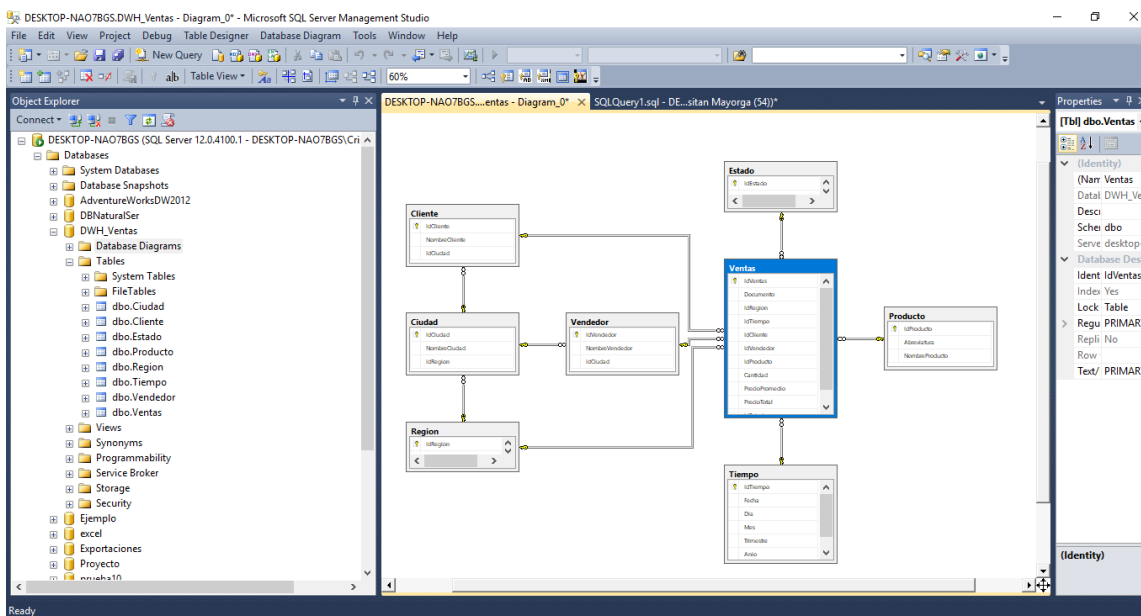


Ilustración 44. Visualización del Esquema del Data Warehouse

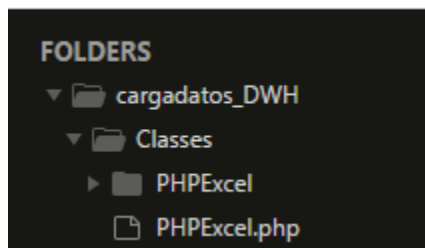


8.4.3 Diseño del Aplicativo Auxiliar para la Carga de los Datos al Data Warehouse

Para completar el proceso de ETL se hace necesario realizar un Aplicativo que cargue los datos al Data Warehouse. Este lo desarrollamos en PHP, el cual lee el archivo de Excel, valida los datos para evitar que existan datos nulos o en algunos casos datos erróneos y si todos los datos son correctos carga los datos en el Data Warehouse en SQL server.

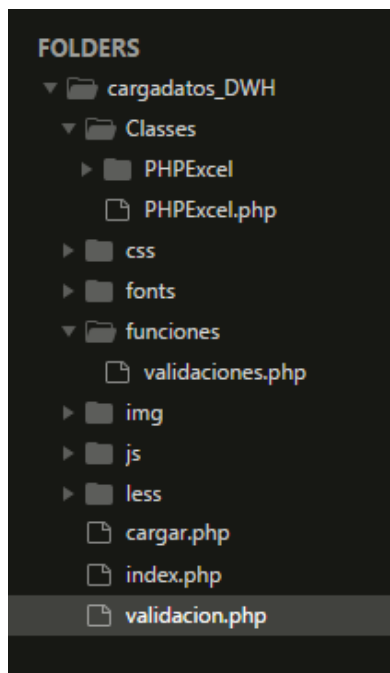
Para que el Aplicativo pueda leer el Excel se hacen necesarias la librería de 'PHPExcel':

Ilustración 45. Librería PHPExcel



La estructura del Aplicativo Auxiliar es la siguiente forma:

Ilustración 46. Estructura del Aplicativo Auxiliar



En donde la carpeta ‘Classes’ contiene la librería necesaria para leer el archivo Excel, la carpeta ‘funciones’ tiene el archivo necesario para la validación, el resto de carpetas hacen parte de la estructura Front-End. Los archivos cargar.php, index.php y validacion.php están desarrollados en HTML y PHP.

8.4.4 Construcción del Aplicativo Auxiliar de Carga de Datos

El Aplicativo se diseñó para cargar los datos hoja por hoja del Excel seleccionado, por lo cual desarrollamos un archivo index.php que el cual cuenta con tres secciones. En la primera sección se muestra el título del Aplicativo, en la segunda se carga el archivo Excel y en la tercera se validan los datos para reducir los errores y posteriormente permitir la carga en el Data Warehouse.

index.php

Ilustración 47. index.php – parte 1

```

1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4 <meta charset="utf-8">
5 <meta http-equiv="X-UA-Compatible" content="IE=edge">
6 <meta name="viewport" content="width=device-width, initial-scale=1">
7
8 <title>GRUPO NATURAL SER</title>
9
10 <!-- Custom styles for this template -->
11 <link href="css/main.css" rel="stylesheet">
12 <link href="https://fonts.googleapis.com/css?family=Montserrat" rel="stylesheet">
13 </head>
14
15 <body class="pagina-arriba" data-spy="scroll" data-target=".navbar-personalizada">
16
17 <nav class="navbar navbar-personalizada navbar-fixed-top">
18 <div class="container">
19 <div class="navbar-header pagina-scroll">
20 <button type="button" class="navbar-toggle collapsed" data-toggle="collapse" data-target="#navbar" aria-expanded="false"
21 <span class="sr-only">Toggle navigation</span>
22 <span class="icon-bar"></span>
23 <span class="icon-bar"></span>
24 <span class="icon-bar"></span>
25 </button>
26 <a class="navbar-brand" href="#">ONS</a>
27 </div>
28 <div id="navbar" class="collapse navbar-collapse navbar-right navbar-main-collapse">
29 <ul class="nav navbar-nav">
30 <li class="hidden">
31 <a href="#pagina-arriba"></a>
32 </li>
33 <li class="pagina-scroll">
34 <a href="#cargar">Cargar Excel</a>
35 </li>
36 <li class="pagina-scroll">
37 <a href="#cargar1">Validar Datos</a>
38 </li>
39 </ul>
40 </div><!--/.nav-collapse -->
41

```

Ilustración 48. Index.php – parte 2

```

45 <section id="intro" class="intro">
46 <div class="intro-cuerpo">
47 <div class="container">
48 <div class="row">
49 <div class="col-md-8 col-md-offset-2">
50 <h1 class="texto-bienvenidos">Replicativo Auxiliar</h1>
51 <p class="intro-texto"><strong><span class="color-verde">Grupo Natural Ser</span></strong> Carga de Datos Data Warehouse</p>
52 </div>
53 </div>
54 </div>
55 </section>
56
57 <section id="cargar" class="contenido-seccion text-left">
58 <h2 class="text-center">Cargar Excel</h2>
59 <p class="text-center">Aquí se cargan el archivo excel</p>
60 <strong>Grupo Natural Ser</strong></p>
61 <br>
62 <div class="row">
63 <div class="col-md-8 col-md-offset-5">
64 <ul class="list-inline">
65 <li class="list-inline">
66 <form name="importa" method="post" action="" enctype="multipart/form-data">
67 <div class="col-xs-4">
68 <div class="form-group">
69 <input type="file" class="filestyle" data-buttonText="Seleccione archivo" name="excel">
70 </div>
71 </div>
72 <div class="col-xs-2">
73 <input class="btn btn-default btn-file" type="submit" name="enviar" value="Cargar Excel"/>
74 </div>
75 <input type="hidden" value="upload" name="action" />
76 <input type="hidden" value="usuarios" name="mod">
77 <input type="hidden" value="masiva" name="acc">
78 </form>
79 </li>
80 </ul>
81 </div>
82 </div>
83 <div class="container">
84 
85 </div>

```

Ilustración 49. Index.php – parte 3

```

87 <section id="cargar1" class="contenido-seccion text-left">
88
89 <h2 class="text-center">Validar Datos</h2>
90 <p class="text-center">
91 <strong>Grupo Natural Ser</strong></p>
92 <br>
93 <div class="row">
94 <div class="col-md-8 col-md-offset-5">
95 <ul class="list-inline">
96 <li><a href="validacion.php" class="btn btn-success btn-lg"><span class="network-name">Validar Datos</span></a>
97 </li>
98 </ul>
99 </div>
100 </div>
101 <div class="container">
102 
103 </div>
104 </section>
105
106 <!-- Bootstrap core JavaScript
107 =====>
108 <!-- Placed at the end of the document so the page load faster -->
109 <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.12.4/jquery.min.js"></script>
110 <script src="js/bootstrap.js"></script>
111 <script type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/jquery-easing/1.3.2/jquery.easing.min.js"></script>
112 <script src="js/main.js"></script>
113 </body>
114 </html>
115 <?php
116 extract($_POST);
117 if (isset($_POST['action'])) {
118 $action=$_POST['action'];
119 }
120 if (isset($action)=="upload"){
121 //cargamos el fichero
122 $nombreadarchivo = $_FILES['excel']['name'];
123 $tipo = $_FILES['excel']['type'];
124 $destino = "cop_".$nombreadarchivo;//Le agregamos un prefijo para identificarlo el archivo cargado
125 if (copy($_FILES['excel']['tmp_name'],$destino)) echo "Archivo Cargado Con Éxito";
126 else echo "Error Al Cargar el Archivo";
127 }
128 }
  
```

El archivo index.php se estructuró por secciones con código HTML y al final del código agregamos un código PHP que es el que se encarga de guardar el Excel seleccionado por el usuario en la carpeta raíz.

Condición: El archivo debe llamarse “Datos” para que el Aplicativo pueda ejecutarse correctamente.

validación.php

Para validar los datos se creó el archivo validacion.php para que validara todos los datos antes de cargarlos en la base de datos.

Ilustración 50. validacion.php

```

1 <?php
2 require 'Classes/PHPExcel/IOFactory.php';
3 require_once 'funciones/validaciones.php';
4 ini_set('memory_limit', '1024M');
5
6
7 >
8
9 <!DOCTYPE html>
10 <html lang="en">
11 <head>
12 <meta charset="utf-8">
13 <meta http-equiv="X-UA-Compatible" content="IE=edge">
14 <meta name="viewport" content="width=device-width, initial-scale=1">
15
16 <title>GRUPO NATURAL SER</title>
17
18 <!-- Custom styles for this template -->
19 <link href="css/main.css" rel="stylesheet">
20 <link href="https://fonts.googleapis.com/css?family=Montserrat" rel="stylesheet">
21
22 </head>
23
24 <body class="pagina-arriba" data-spy="scroll" data-target=".navbar-personalizada">
25
26 <nav class="navbar navbar-personalizada navbar-fixed-top">
27 <div class="container">
28 <div class="navbar-header pagina-scroll">
29 <button type="button" class="navbar-toggle collapsed" data-toggle="collapse" data-target="#navbar" aria-expanded="false"
30   aria-controls="navbar">
31   <span class="sr-only">Toggle navigation</span>
32   <span class="icon-bar"></span>
33   <span class="icon-bar"></span>
34   <span class="icon-bar"></span>
35 </button>
36 <a class="navbar-brand" href="#">GNS</a>
37 </div>
38 <div id="navbar" class="collapse navbar-collapse navbar-right navbar-main-collapse">
39 <ul class="nav navbar-nav">
40 <li class="hidden">
41 <a href="#pagina-arriba"></a>

```

- Se llama la librería 'Classes/PHPExcel/IOFactory.php' y la carpeta de funciones 'funciones/validaciones.php', además de establecer un límite de memoria para superior al determinado para que el archivo pueda validar todos los datos.

Ilustración 51. Llamado de funciones y límite de memoria

```

<?php
require 'Classes/PHPExcel/IOFactory.php';
require_once 'funciones/validaciones.php';
ini_set('memory_limit', '1024M');

```

- Se inicia la estructura HTML similar a la del archivo index.php, es decir por secciones para mostrar los errores de cada hoja leída y la sección correspondiente. Por ejemplo, si se presenta un error en la hoja de 'Región', los errores se visualizarán en la sección correspondiente a 'Región'.

Ilustración 52. Estructura HTML de la validación de datos en el Aplicativo Auxiliar

```

<nav class="navbar navbar-personalizada navbar-fixed-top">
  <div class="container">
    <div class="navbar-header pagina-scroll">
      <button type="button" class="navbar-toggle collapsed" data-toggle="collapse" data-target="#navbar" aria-expanded="false"
        aria-controls="navbar">
        <span class="sr-only">Toggle navigation</span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
      </button>
      <a class="navbar-brand" href="#">GNS</a>
    </div>
    <div id="navbar" class="collapse navbar-collapse navbar-right navbar-main-collapse">
      <ul class="nav navbar-nav">
        <li class="hidden">
          <a href="#pagina-arriba"></a>
        </li>
        <li class="pagina-scroll">
          <a href="#region">Region</a>
        </li>
        <li class="pagina-scroll">
          <a href="#ciudad">Ciudad</a>
        </li>
        <li class="pagina-scroll">
          <a href="#cliente">Cliente</a>
        </li>
        <li class="pagina-scroll">
          <a href="#producto">Producto</a>
        </li>
        <li class="pagina-scroll">
          <a href="#tiempo">Tiempo</a>
        </li>
        <li class="pagina-scroll">
          <a href="#vendedor">Vendedor</a>
        </li>
        <li class="pagina-scroll">
          <a href="#estado">Estado</a>
        </li>
        <li class="pagina-scroll">
          <a href="#ventas">Ventas</a>
        </li>
      </ul>
    </div>
  </div>

```

- Se asignan las variables correspondientes para empezar a leer el Excel. En este caso se mostrará el ejemplo con la validación de producto.

Ilustración 53. Variables de lectura del archivo Excel en el Aplicativo Auxiliar

```

$erroresproducto = array(); //Inicia el arreglo donde se almacenaran los errores
$nombreArchivo = 'cop_Prueba.xlsx'; //Selecciona el archivo que se llama "cop_Datos.xlsx"
$objPHPExcel = PHPEXCEL_IOFactory::load($nombreArchivo); //Carga el archivo en la clase de PHPEXCEL_IOFactory
$objPHPExcel->setActiveSheetIndex(3); //Busca en la hoja numero 3, donde se encuentra los datos de Producto
$numRows = $objPHPExcel->setActiveSheetIndex(3)->getHighestRow(); //Lee el numero de filas que tiene la hoja

```

- Para que lea todas las líneas del Excel, se realiza un ciclo for de tal forma que lea fila por fila hasta que llegue a la última fila con datos. Después guarda los datos encontrados en cada celda del Excel y lo almacena en una variable. Como algunos valores de descripción no los cargaba correctamente, se hizo necesario utilizar una variable adicional donde almacenará el mismo valor, pero decodificado en un formato UTF-8.

Ilustración 54. Arreglo que recorre los campos del archivo

```

for($i = 2; $i <= $numRows; $i++){ //Inicializa el arreglo para leer los datos desde la fila 2 hasta que se complete
el numero de filas con datos

    $idproducto = $objPHPExcel->getActiveSheet()->getCell('A'.$i)->getCalculatedValue();//Asigna en la variable
    $idproducto el valor encontrado en la columna A y fila $i.
    $abrev = $objPHPExcel->getActiveSheet()->getCell('B'.$i)->getCalculatedValue(); //Asigna en la variable $abrev
    el valor encontrado en la columna B y fila $i.
    $nombreproducto = $objPHPExcel->getActiveSheet()->getCell('C'.$i)->getCalculatedValue(); //Asigna en la variable
    $nombreproducto el valor encontrado en la columna C y fila $i.
    $nombre_producto= utf8_decode($nombreproducto); //Asigna en la variable $nombre_producto el valor de
    $nombreproducto decodificado

```

- Para validar cada variable que ha captado el dato, iniciamos un condicional y llamamos la función correspondiente, en la cual asignamos la variable. De estar incorrecta almacena el error en el arreglo que previamente se definió. Una vez recorre y valida todos los datos se cierra el ciclo for.

Ilustración 55. Validaciones de los campos

```

    $opciones_producto = array(
        'options' => array(
            //Definimos el rango
            'min_range' => 1,
            'max_range' => 100
        )
    );
    if (!validaRequerido($idproducto)) { //Llama la función y le asigna la variable a examinar
        $erroresproducto[] = 'HOJA 4= El campo "id producto" esta vacio en la columna A y fila '.$i;
    }
    if (!validarEntero($idproducto, $opciones_producto)) { //Llama la función y le asigna la variable y el rango a
    examinar
        $erroresproducto[] = 'HOJA 4= El campo "id producto" es incorrecto en la columna A y fila '.$i;
    }
    if (!validaRequerido($abrev)) {
        $erroresproducto[] = 'HOJA 4= El campo "abreviatura" esta vacio en la columna B y fila '.$i;
    }
    if (!validaRequerido($nombreproducto)) {
        $erroresproducto[] = 'HOJA 4= El campo "nombre producto" esta vacio en la columna C y fila '.$i;
    }
}

```

- Luego imprimirá los errores de color rojo cada que encuentre uno almacenado en el arreglo

Ilustración 56. Impresión de los errores

```
<?php if ($erroresproducto): ?>
  <ul style="color: #f00;">
    <?php foreach ($erroresproducto as $error): ?>
      <li> <?php echo $error ?> </li>
    <?php endforeach; ?>
  </ul>
<?php endif; ?>
```

Al finalizar esta sección, continuará con la siguiente hoja hasta validar todos los datos de las hojas.

- En la última sección, se evalúa todos los arreglos establecidos para determinar si existe algún valor (error) en estos. De ser así, mostrará un mensaje de “CORREGIR DATOS” y un botón para volver al menú principal. Caso contrario mostrará un botón para cargar los datos en la base de datos.

Si al validar todos los datos no existe ningún error y da la opción para cargar los datos, el programa llama al archivo cargar.php.

cargar.php

Este archivo maneja la misma estructura del archivo ‘validacion.php’ con la diferencia de que ya no almacena los errores en un arreglo, sino que carga los datos en SQL Server.

- Asigna los títulos para imprimir los datos a cargar en SQL.

Ilustración 57. Encabezados de la tabla

```
<thead>
  <tr>
    <th>Id Producto</th>
    <th>Abreviatura</th>
    <th>Nombre Producto</th>
  </tr>
</thead>

<tbody>
```

- De igual forma se realiza el ciclo for de tal forma que lea fila por fila hasta que llegue a la última fila con datos. Después guarda los datos encontrados en cada celda del Excel y lo

almacena en una variable. Como algunos valores de descripción no los cargaba correctamente, se hizo necesario utilizar una variable adicional donde almacenará el mismo valor, pero decodificado en un formato UTF-8.

Ilustración 58. Arreglo que recorre los datos para cargarlos en la DB

```
<?php for($i = 2; $i <= $numRows; $i++){
    $idproducto = $objPHPExcel->getActiveSheet()->getCell('A' . $i)->getCalculatedValue();//Asigna
    en la variable $idproducto el valor encontrado en la columna A y fila $i.
    $abrev = $objPHPExcel->getActiveSheet()->getCell('B' . $i)->getCalculatedValue(); //Asigna en la
    variable $abrev el valor encontrado en la columna B y fila $i.
    $nombreproducto = $objPHPExcel->getActiveSheet()->getCell('C' . $i)->getCalculatedValue(); //
    Asigna en la variable $nombreproducto el valor encontrado en la columna C y fila $i.
    $nombre_producto= utf8_decode($nombreproducto); //Asigna en la variable $nombre_producto el
    valor de $nombreproducto decodificado
?>
```

- Imprime los valores de las variables.

Ilustración 59. Impresión de cada campo

```
<tr>
<td><?php echo $idproducto;?></td>
<td><?php echo $abrev;?></td>
<td><?php echo $nombreproducto;?></td>
</tr>
```

- Establece la conexión con el servidor y con la base de datos.

Ilustración 60. Conexión del Aplicativo Auxiliar con SQL Server

```
$user = "";
$pass = "";
$server = "DESKTOP-NA07BGS"; // Selecciona el servidor
$dbase = "Example"; //Selecciona la base de datos
$conection_string = "DRIVER={SQL Server};SERVER=" . $server . " ; DATABASE=" . $dbase ;
$conection = odbc_connect($conection_string, $user, $pass); //Ejecuta la conexion de tipo odbc
a la base de datos
```

- Ejecuta la sentencia SQL para insertar los datos almacenados en las variables, en la base de datos y se cierra el ciclo for.

Ilustración 61. Código que carga los datos en la DB

```
$sqlquery="insert into Producto(IdProducto,Abreviatura,NombreProducto)values('$idproducto','$abrev','$
nombre_producto')"; //Alamcena en la variable $sqlquery la sentencia sql
odbc_exec($conection, $sqlquery); //Abre la conecion y ejecuta la sentencia
odbc_close($conection); //Cierra la conecion con la base de datos
```

Una vez carga todos los datos de esa hoja, pasa a la hoja siguiente hasta cargar todos los datos del Excel.

8.4.5 Proceso de ETL

Este proceso se llevó a cabo de la siguiente manera:

Extracción

La parte de extracción se explica en el apartado 9.2 ‘Compresión de los Datos’.

Transformación

La parte de transformación se explica desde el apartado 9.3 ‘Preparación de los Datos’, hasta el apartado 9.4 ‘Desarrollo(Modelado)’.

Carga

Para cargar los archivos hacemos uso del programa de la siguiente manera:

- Se inicia el programa y nos desplazamos a cargar Excel.

Ilustración 62. Inicio Aplicativo Auxiliar



- Seleccionamos el archivo Excel con el nombre de “Datos”.

Ilustración 63. Selección y carga del archivo Excel en Aplicativo auxiliar

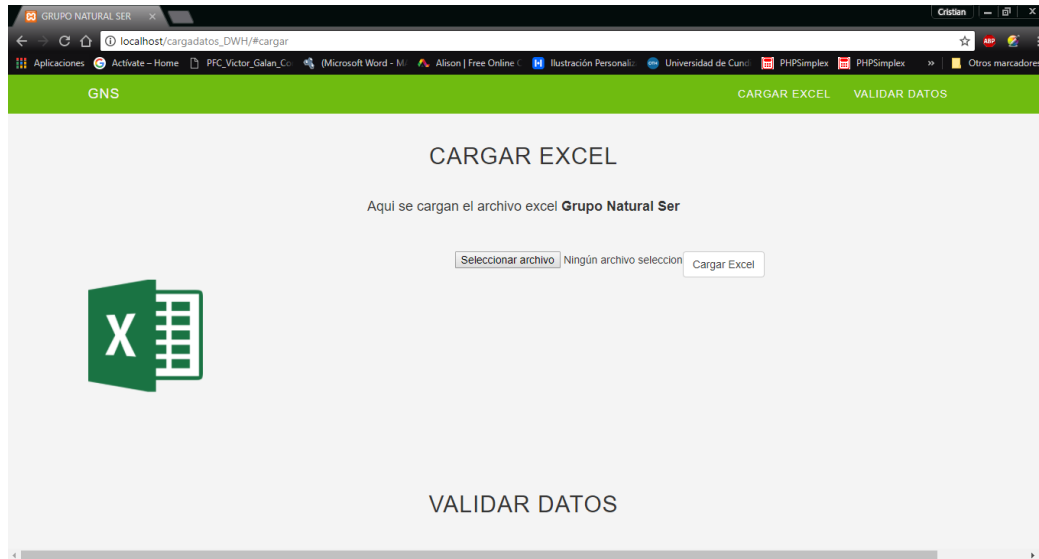
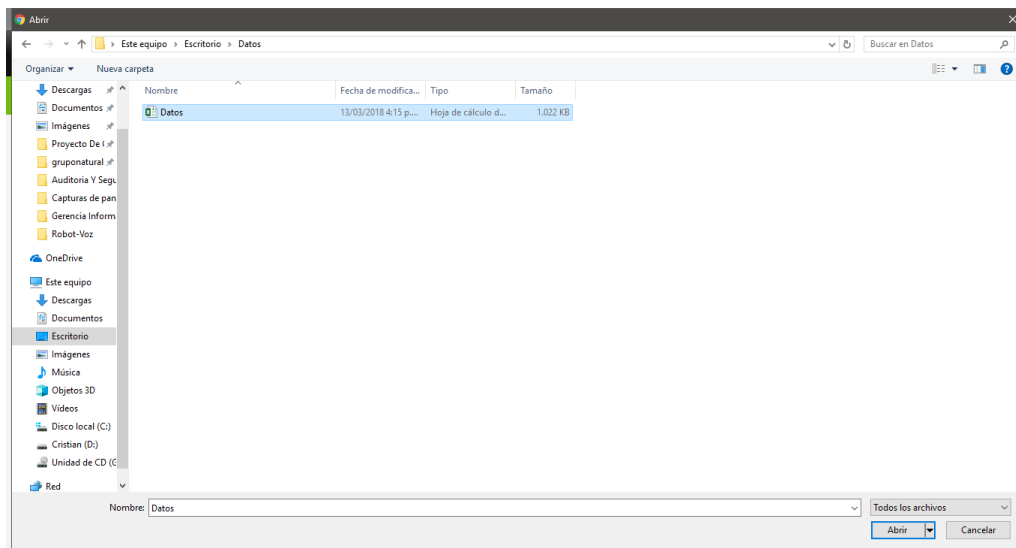
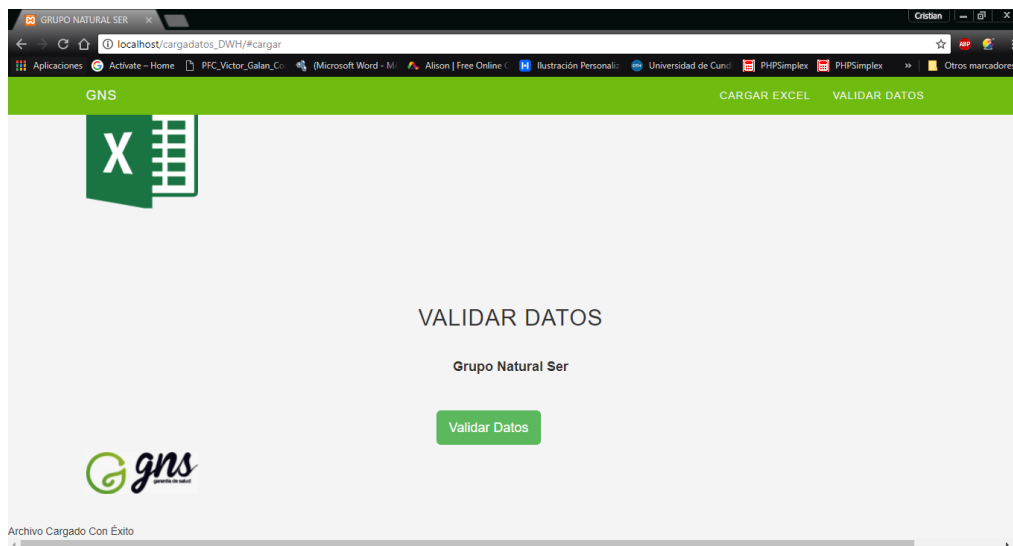


Ilustración 64. Selección de archivo Excel “Datos”



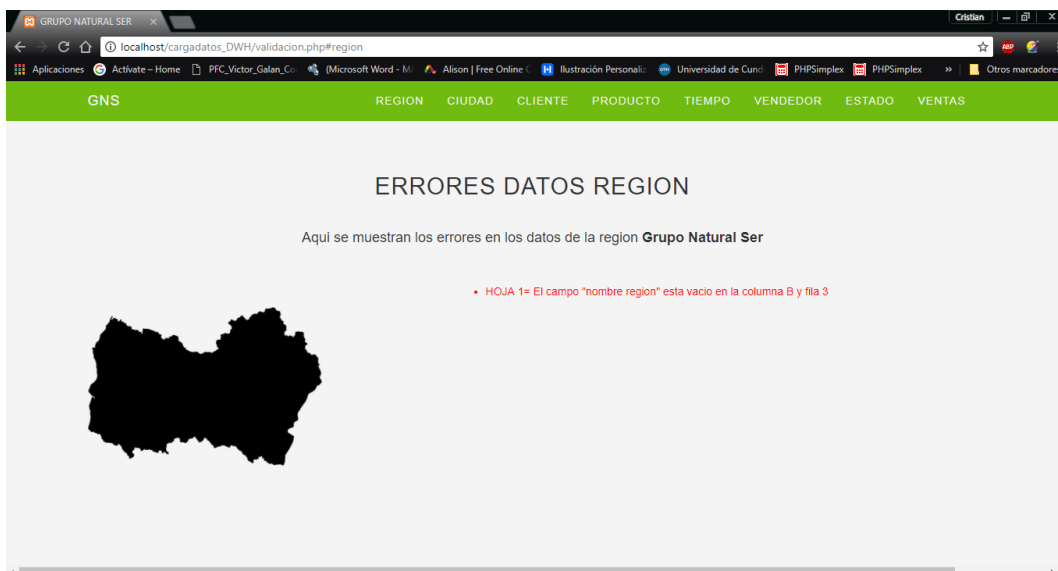
- Vamos a la sección de Validar Datos y hacemos clic en Validar Datos.

Ilustración 65. Validación de datos en Aplicativo Auxiliar



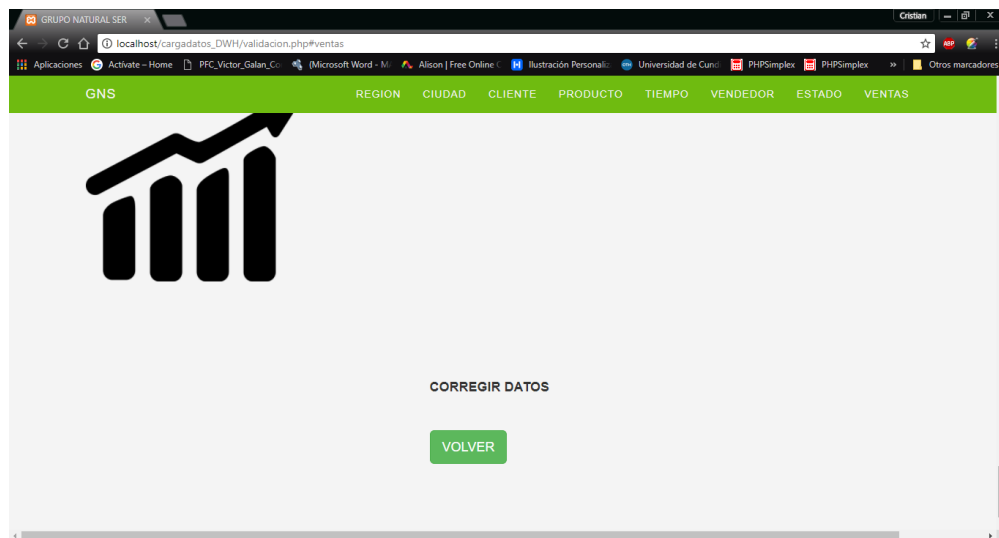
- Si se presenta algún error, se mostrará de la siguiente manera.

Ilustración 66. Error en validación de datos en Aplicativo Auxiliar



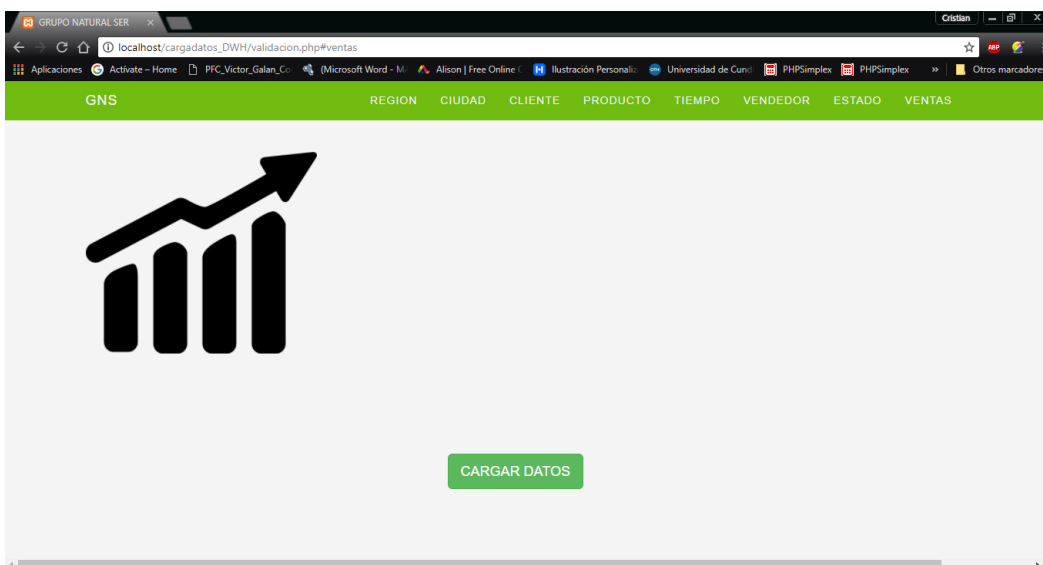
- Al final se encuentra el botón para volver a inicio pues no permite cargar los datos si tienen inconsistencias.

Ilustración 67. Corrección de errores en Aplicativo Auxiliar



- Si los datos no tienen ningún error, aparecerá el botón de Cargar Datos

Ilustración 68. Carga de datos en Aplicativo Auxiliar



- Al cargar los datos, estos se imprimen en tablas correspondientes a la sección.

Ilustración 69. Impresión de datos cargados en Aplicativo Auxiliar

The screenshot shows a web browser window with the URL `localhost/cargadatos_DWH/cargar.php#ciudad`. The page has a green header with the text 'GRUPO NATURAL SER' and a navigation menu with items: REGION, CIUDAD, CLIENTE, PRODUCTO, TIEMPO, VENDEDOR, ESTADO, VENTAS. The main content area is divided into two sections:

DATOS CARGADOS REGION

Grupo Natural Ser

Id Region	Nombre Region
5	Antioquia
11	Bogotá, D.C.

DATOS CARGADOS CIUDAD

Grupo Natural Ser

- Al final se muestra un botón para volver al inicio.

Ilustración 70. Volver a página de inicio en Aplicativo Auxiliar

The screenshot shows the same web browser window as in Illustration 69, but with a table of data displayed. The table has 11 columns: GNS, REGION, CIUDAD, CLIENTE, PRODUCTO, TIEMPO, VENDEDOR, ESTADO, VENTAS. The data rows are as follows:

GNS	REGION	CIUDAD	CLIENTE	PRODUCTO	TIEMPO	VENDEDOR	ESTADO	VENTAS
FV AC 8	11	20170925 35409	3	16	3	20000	60000	2
FV AC 9	11	20170925 35401	3	1	2	8370	16740	2
FV AC 9	11	20170925 35401	3	6	2	9540	19080	2
FV AC 9	11	20170925 35401	3	8	1	8820	8820	2
FV AC 9	11	20170925 35401	3	9	1	13950	13950	2
FV AC 9	11	20170925 35401	3	10	4	10215	40860	2
FV AC 9	11	20170925 35401	3	43	6	6120	36720	2
FV AC 9	11	20170925 35401	3	45	2	6480	12960	2

At the bottom center of the page, there is a green button labeled 'VOLVER'.

Una vez finalizado el proceso podemos visualizar los datos en el Data Warehouse.

Ilustración 71. Datos de Región en el Data Warehouse

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure for 'DESKTOP-NAO7BGS'. The central pane shows a query window with the following SQL code:

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [IdRegion]
, [NombreRegion]
FROM [DWH_Ventas].[dbo].[Region]

```

The Results pane below the query shows the following data:

IdRegion	NombreRegion
5	Antioquia
11	Bogotá, D.C.

The status bar at the bottom indicates 'Query executed successfully.' and '2 rows'.

Ilustración 72. Datos de Ciudad en el Data Warehouse

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure for 'DESKTOP-NAO7BGS'. The central pane shows a query window with the following SQL code:

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [IdCiudad]
, [NombreCiudad]
, [IdRegion]
FROM [DWH_Ventas].[dbo].[Ciudad]

```

The Results pane below the query shows the following data:

IdCiudad	NombreCiudad	IdRegion
5001	Medellín	5
11001	Bogotá, D.C.	11

The status bar at the bottom indicates 'Query executed successfully.' and '2 rows'.

Ilustración 73. Datos de Clientes en el Data Warehouse

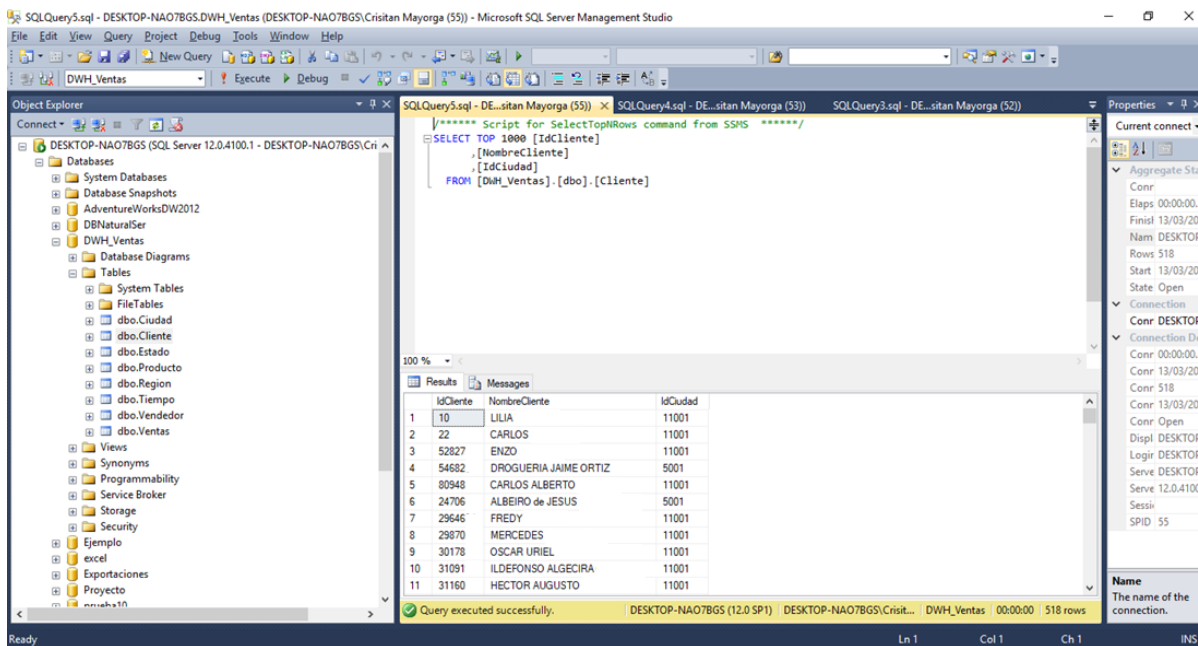


Ilustración 74. Datos de Producto en el Data Warehouse

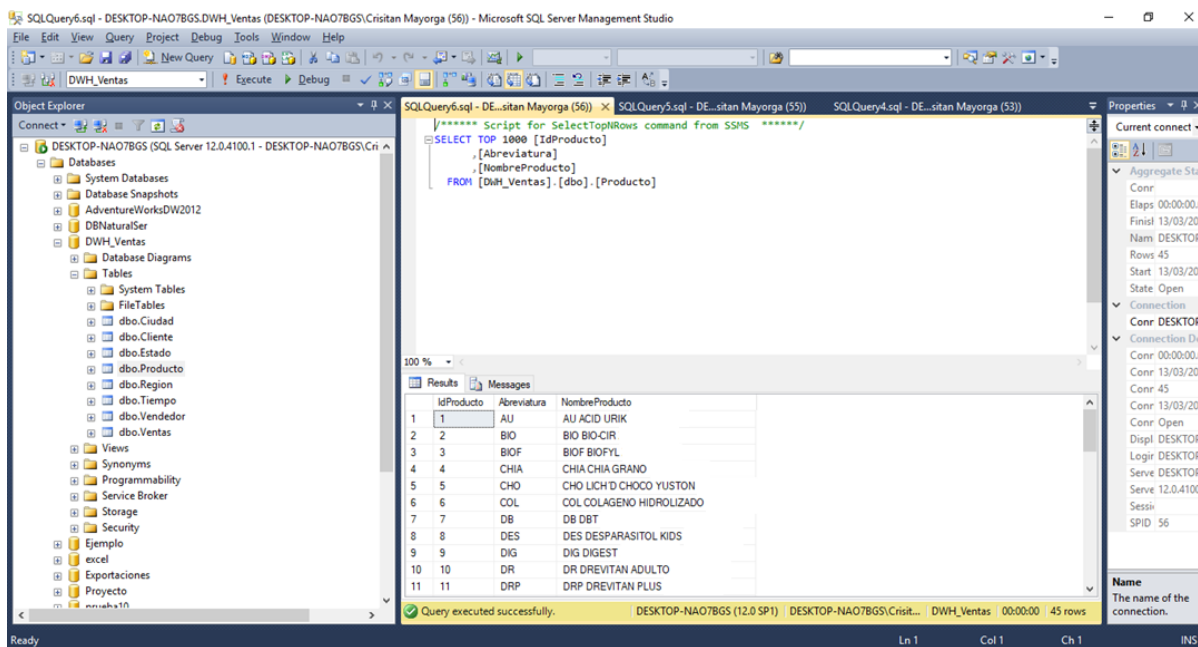


Ilustración 75. Datos de Tiempo en el Data Warehouse

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure for 'DESKTOP-NAO7BGS'. The central pane shows a SQL query executed in 'SQLQuery7.sql - DE...sitan Mayorga (57)'. The query is a 'SELECT TOP 1000' statement that retrieves columns: [IdTiempo], [Fecha], [Dia], [Mes], [Trimestre], and [Año] from the table [DWH_Ventas].[dbo].[Tiempo]. The Results pane at the bottom displays 10 rows of data, showing dates from 2016-01-01 to 2016-01-11. The Properties pane on the right shows connection details for 'DESKTOP-NAO7BGS'.

```

/***** Script for SelectTopNRows command from SSIS *****/
SELECT TOP 1000 [IdTiempo]
, [Fecha]
, [Dia]
, [Mes]
, [Trimestre]
, [Año]
FROM [DWH_Ventas].[dbo].[Tiempo]

```

IdTiempo	Fecha	Dia	Mes	Trimestre	Año	
1	20160101	2016-01-01	1	Enero	1	2016
2	20160102	2016-01-02	2	Enero	1	2016
3	20160103	2016-01-03	3	Enero	1	2016
4	20160104	2016-01-04	4	Enero	1	2016
5	20160105	2016-01-05	5	Enero	1	2016
6	20160106	2016-01-06	6	Enero	1	2016
7	20160107	2016-01-07	7	Enero	1	2016
8	20160108	2016-01-08	8	Enero	1	2016
9	20160109	2016-01-09	9	Enero	1	2016
10	20160110	2016-01-10	10	Enero	1	2016
11	20160111	2016-01-11	11	Enero	1	2016

Ilustración 76. Datos de Vendedores en el Data Warehouse

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure for 'DESKTOP-NAO7BGS'. The central pane shows a SQL query executed in 'SQLQuery8.sql - DE...sitan Mayorga (59)'. The query is a 'SELECT TOP 1000' statement that retrieves columns: [NombreVendedor] and [IdCiudad] from the table [DWH_Ventas].[dbo].[Vendedor]. The Results pane at the bottom displays 11 rows of data, listing various vendors and their cities. The Properties pane on the right shows connection details for 'DESKTOP-NAO7BGS'.

```

/***** Script for SelectTopNRows command from SSIS *****/
SELECT TOP 1000 [IdVendedor]
, [NombreVendedor]
, [IdCiudad]
FROM [DWH_Ventas].[dbo].[Vendedor]

```

IdVendedor	NombreVendedor	IdCiudad
1	ADRIANA PATRICIA	11001
2	ANA PATRICIA	5001
3	ARMANDO	11001
4	BIBIANA PATRICIA	5001
5	ELCIDA	5001
6	ESTEBAN ALBEIRO	5001
7	FRANCISCO ARTURO	11001
8	GNS	11001
9	JENNIFER MARITZA	11001
10	JOSE YOLIAN	5001
11	JOSUE DANIEL	5001

Ilustración 77. Datos de Estado en el Data Warehouse

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the 'DWH_Ventas' database structure. The central pane shows a query window with the following SQL code:

```

/***** Script for SelectTopNRows command from SSIS *****/
SELECT TOP 1000 [IdEstado]
,[Descripcion]
FROM [DWH_Ventas].[dbo].[Estado]

```

The Results pane below the query shows the following data:

IdEstado	Descripcion
1	Devolucion
2	Venta

The Properties pane on the right shows connection details for 'DESKTOP-NAO7BGS'.

Ilustración 78. Datos de Ventas en el Data Warehouse

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the 'DWH_Ventas' database structure. The central pane shows a query window with the following SQL code:

```

/***** Script for SelectTopNRows command from SSIS *****/
SELECT TOP 1000 [IdVentas]
,[Documento]
,[IdRegion]
,[IdTiempo]
,[IdCliente]
,[IdVendedor]
,[IdProducto]
,[Cantidad]
,[PrecioPromedio]
,[PrecioTotal]
,[IdEstado]
FROM [DWH_Ventas].[dbo].[Ventas]

```

The Results pane below the query shows the following data (first 10 rows):

IdVentas	Documento	IdRegion	IdTiempo	IdCliente	IdVendedor	IdProducto	Cantidad	PrecioPromedio	PrecioTotal	IdEstado
1	DMC 10	11	20160118	80852	7	41	-6	6500.00	-39000.00	1
2	DMC 10	5	20160701	71595	4	1	-3	8460.00	-25380.00	1
3	DMC 11	11	20160506	79487	1	7	-9	15000.00	-135000.00	1
4	DMC 11	11	20160506	79487	1	8	-6	9500.00	-57000.00	1
5	DMC 11	11	20160506	79487	1	10	-16	11000.00	-176000.00	1
6	DMC 11	11	20160506	79487	1	11	-10	16500.00	-165000.00	1
7	DMC 11	11	20160506	79487	1	14	-8	15000.00	-120000.00	1
8	DMC 11	11	20160506	79487	1	17	-7	25000.00	-175000.00	1
9	DMC 12	5	20160829	34057	4	17	-1	23500.00	-23500.00	1
10	DMC 12	11	20161031	53094	1	14	-8	15000.00	-120000.00	1

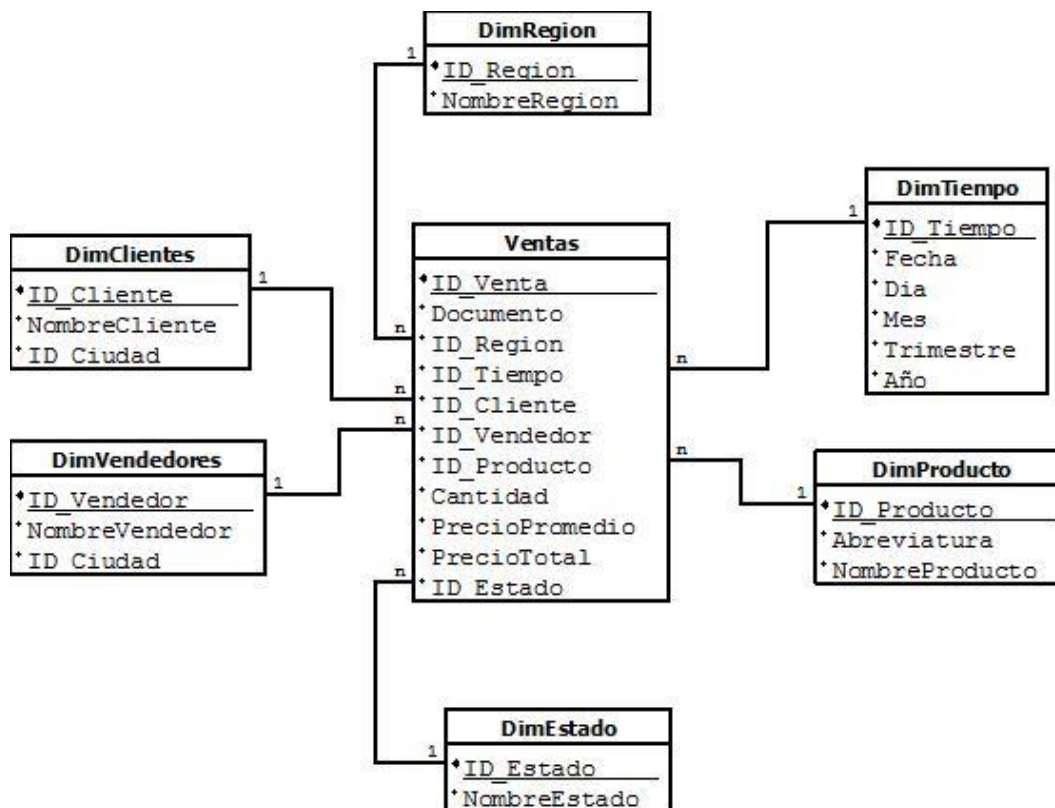
The Properties pane on the right shows connection details for 'DESKTOP-NAO7BGS'.

8.4.6 Esquema del Modelo

Para la creación de nuestro Cubo Multidimensional, escogimos el esquema estrella por lo siguiente:

- Tiene mejores tiempos de respuesta.
- Su diseño es modificable de manera fácil.
- Los usuarios visualizan el esquema a como este fue diseñado.
- Es interactivo con las herramientas de análisis y consulta.

Ilustración 79. Esquema Estrella para Modelo Multidimensional

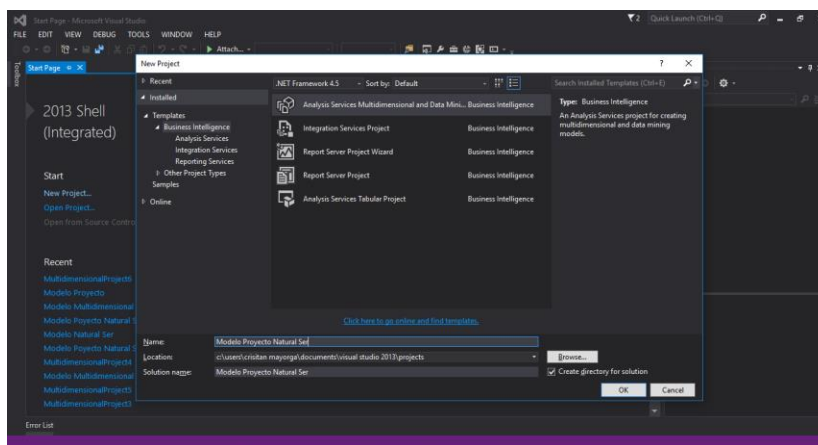


8.4.8 Construcción del Modelo Multidimensional en SQL Server Data Tools – Visual Studio

Para la construcción del modelo empleamos la herramienta SQL Server Data Tools y se creó de la siguiente manera:

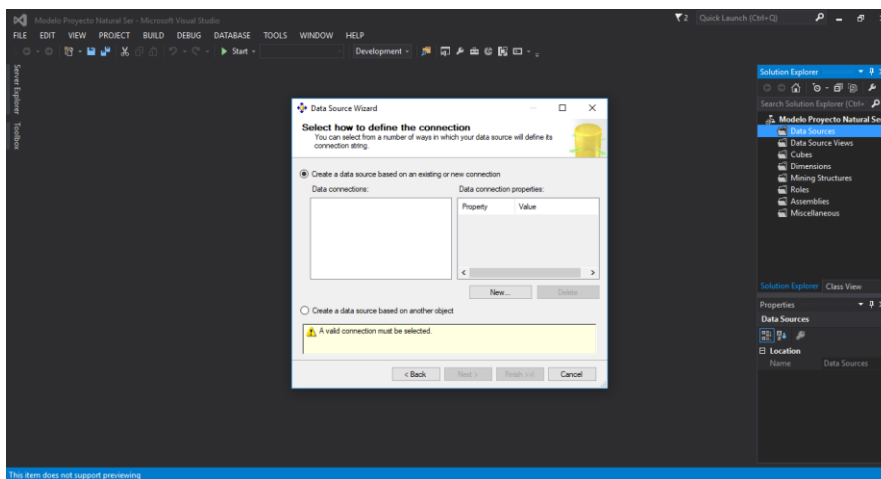
- Creamos el proyecto

Ilustración 80. Creación del Proyecto en SQL Server Data Tools



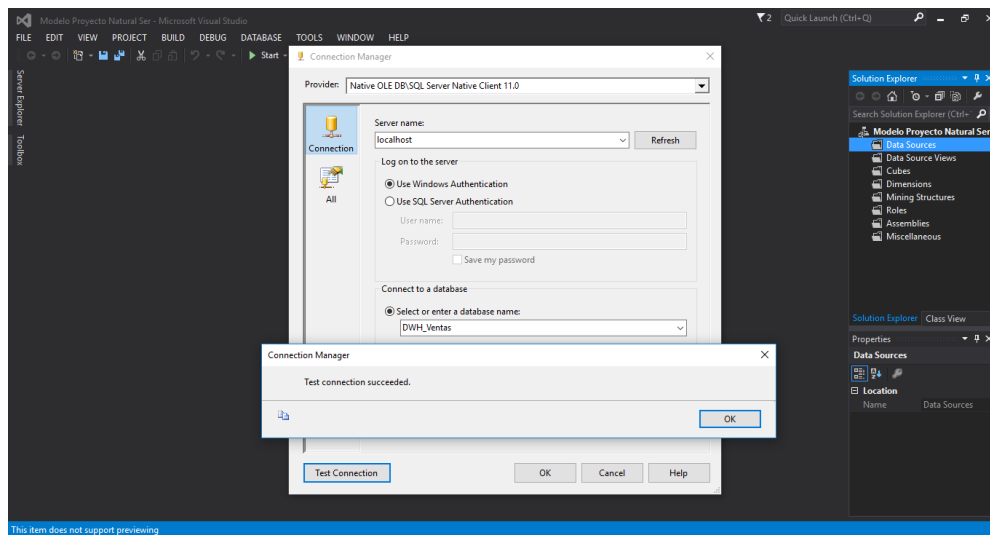
- Seleccionamos un origen de datos

Ilustración 81. Selección del Origen de Datos



- Se selecciona el DWH que tenemos en SQL server y probamos la conexión.

Ilustración 82. Selección de Data Warehouse



- Elegimos usar una cuenta de servicio por la configuración previa del programa.

Ilustración 83. Elección de Cuenta de Servicio

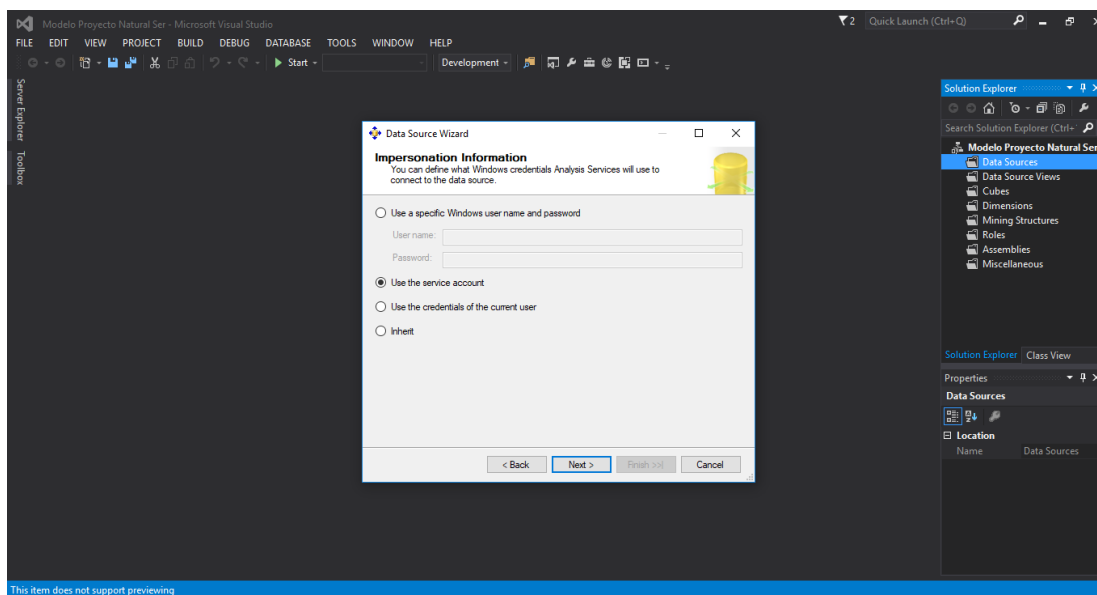
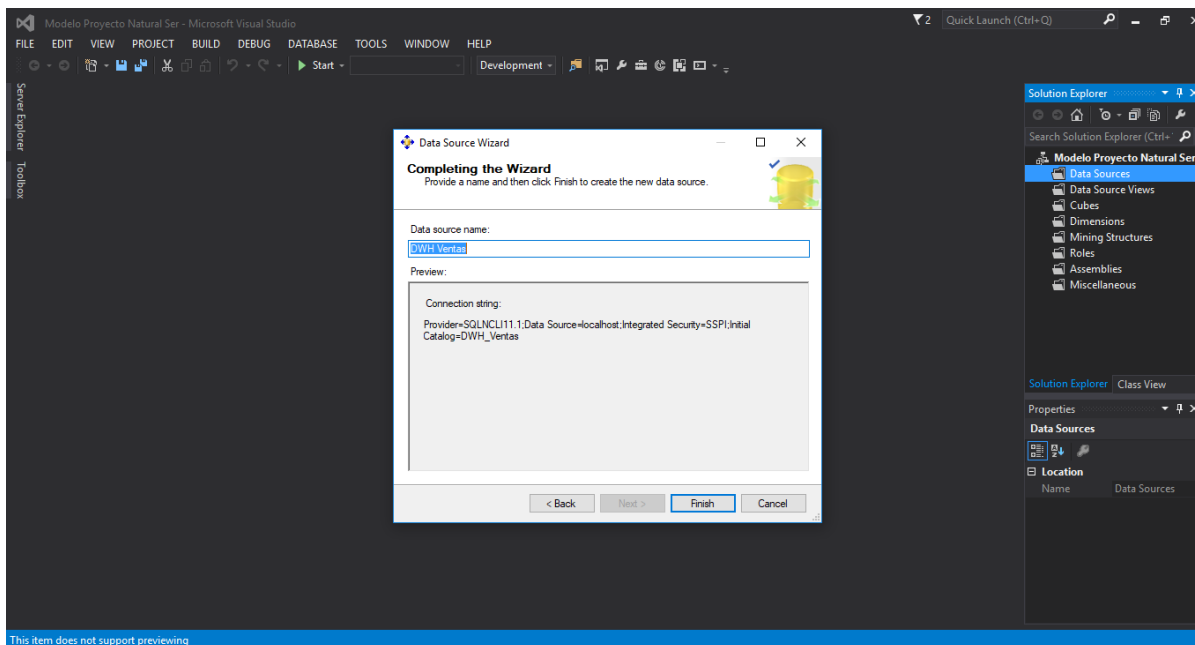
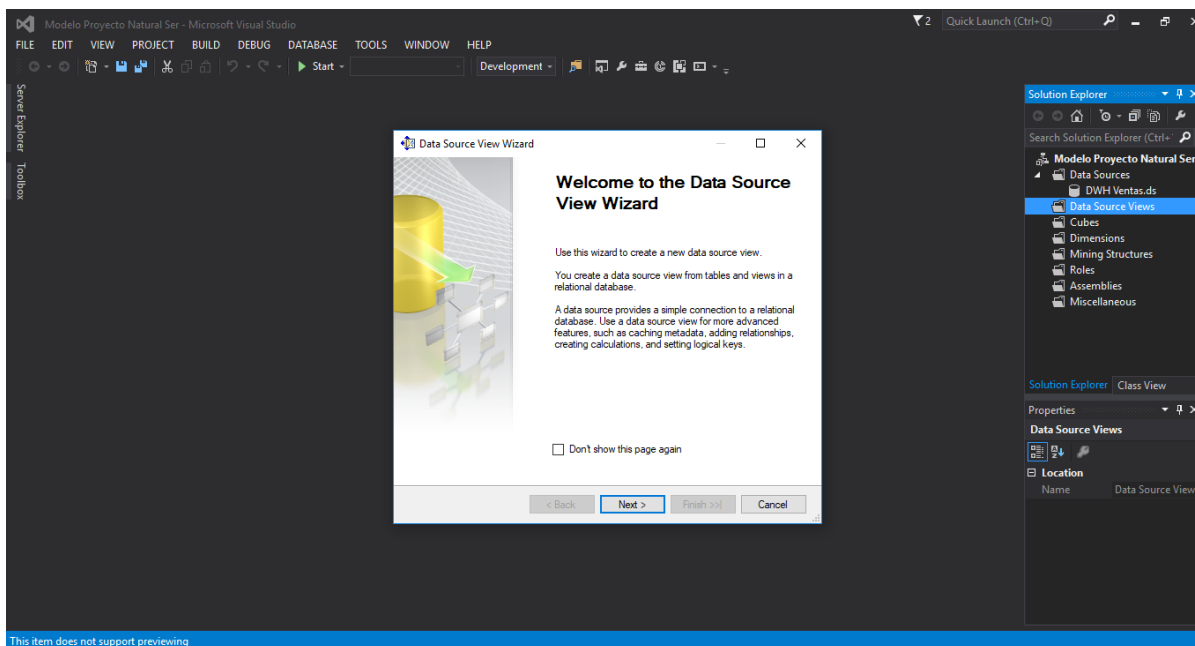


Ilustración 84. Finalización de Selección de Origen de Datos



- Seleccionamos las vistas del origen de datos

Ilustración 85. Vistas del Origen de Datos



- Seleccionamos las tablas con las que deseamos trabajar el modelo. En este caso no utilizaremos la tabla Ciudad.

Ilustración 86. Selección de tablas para la creación del Modelo Multidimensional – Paso 1

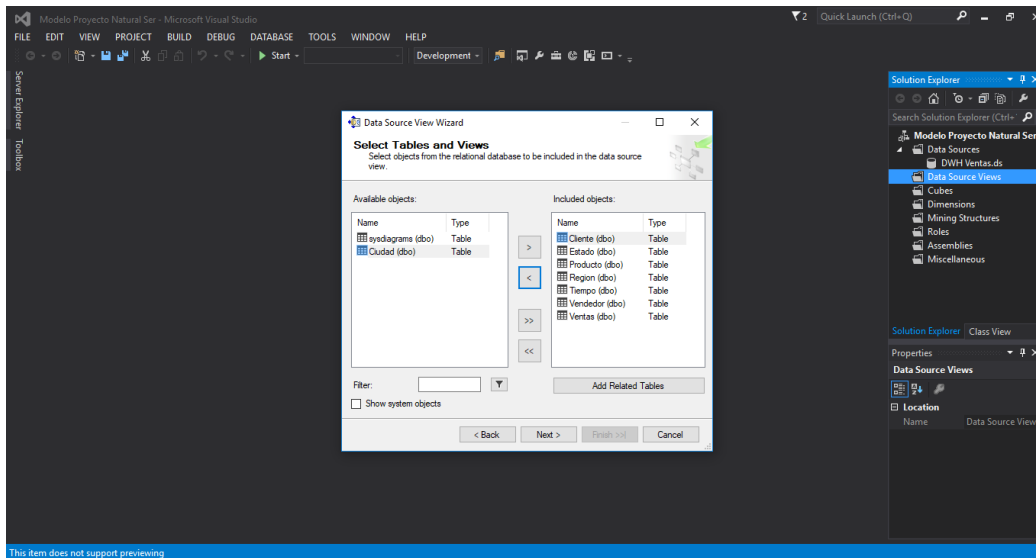
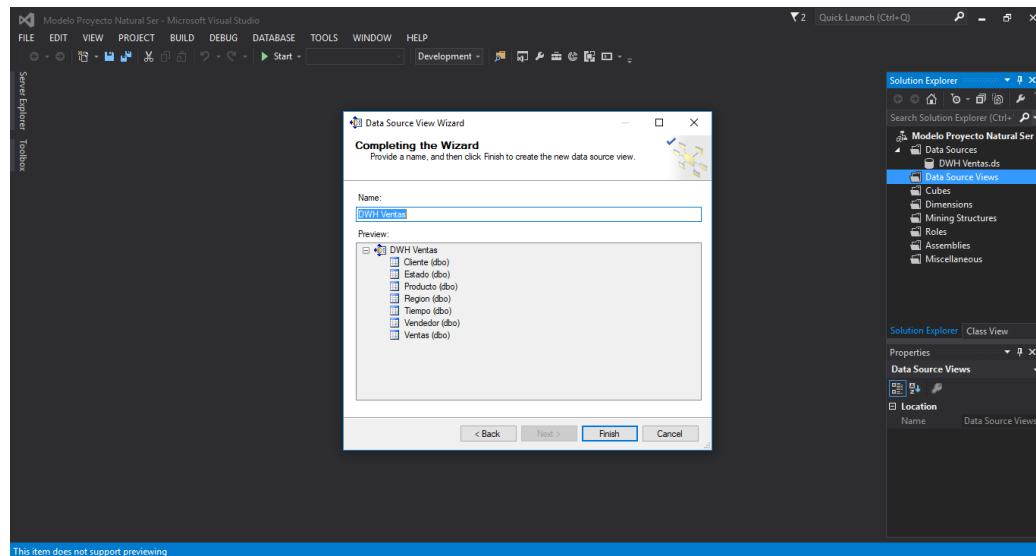
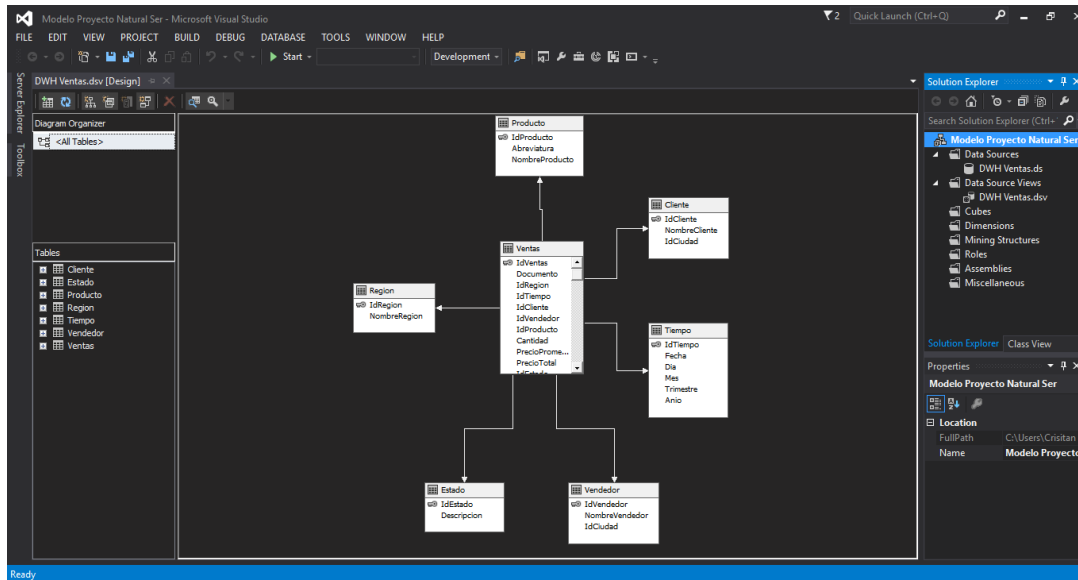


Ilustración 87. Selección de tablas para la creación del Modelo Multidimensional – Paso 2



- Nos muestra una estructura previa del modelo.

Ilustración 88. Estructura previa del Modelo Multidimensional



- Creamos el cubo

Ilustración 89. Creación del cubo OLAP – Paso 1

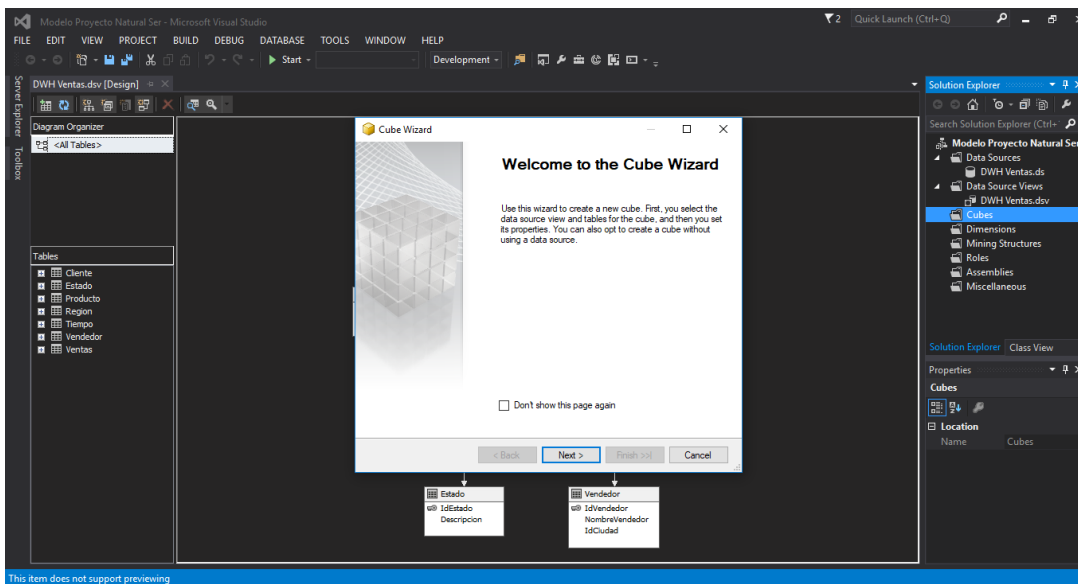
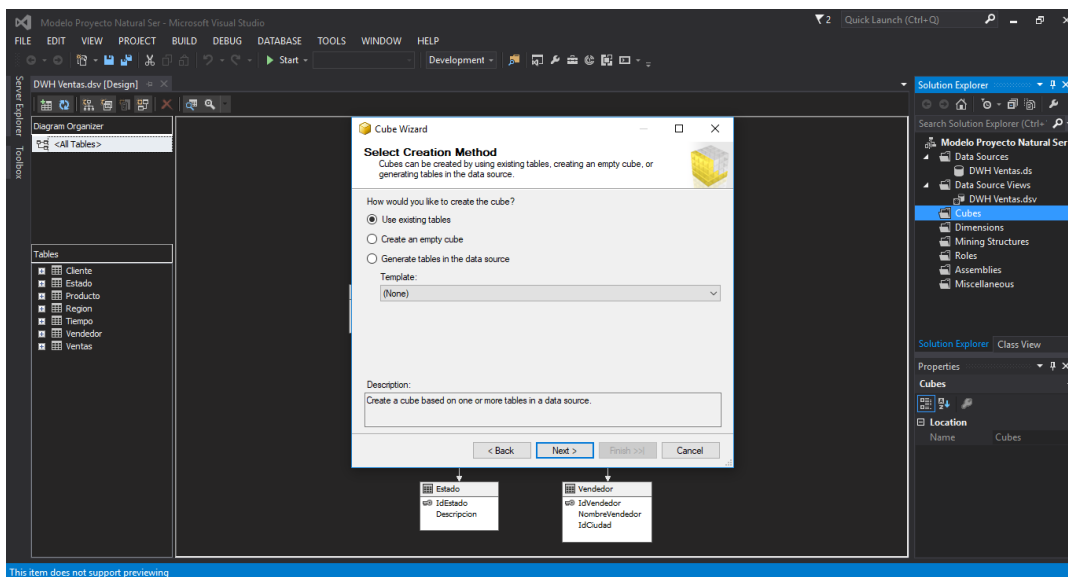
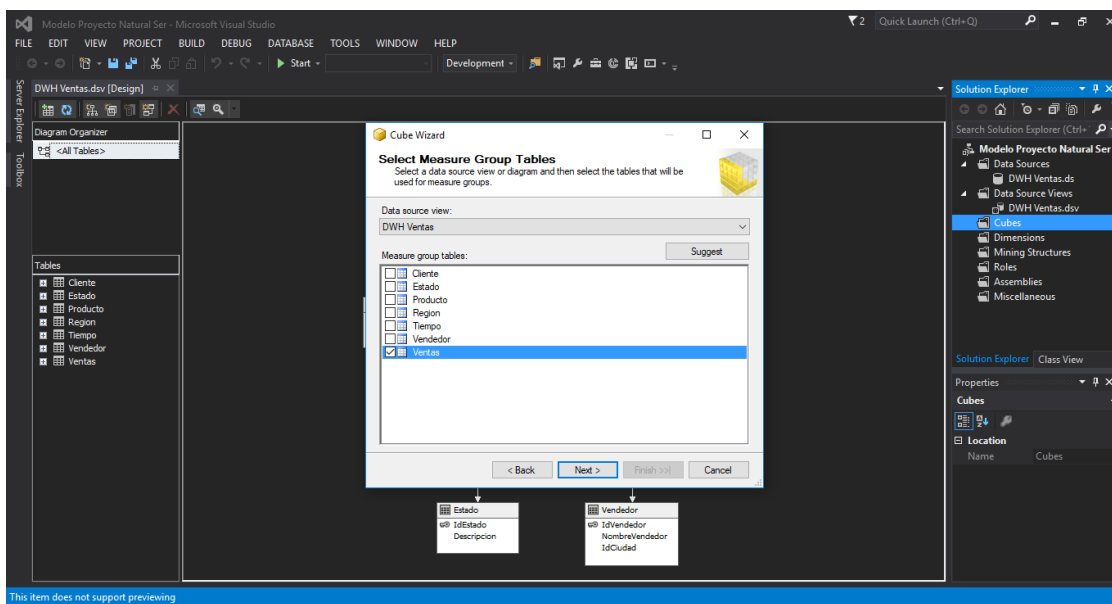


Ilustración 90. Creación del cubo OLAP – Paso 2



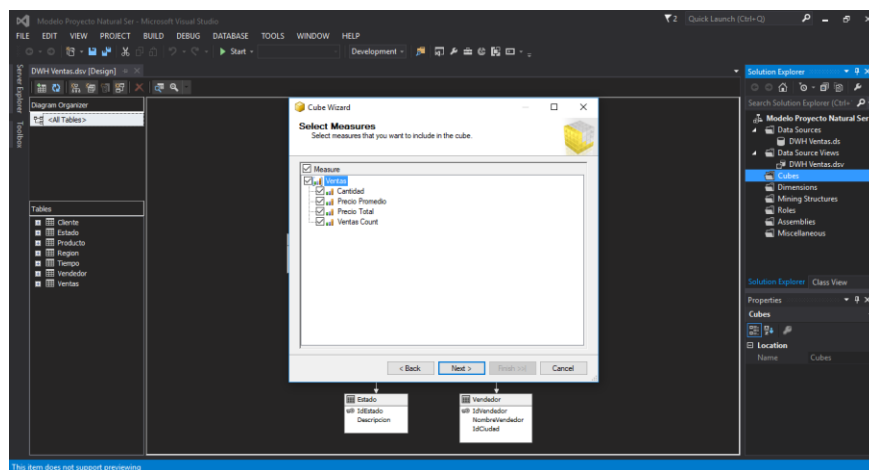
- Seleccionamos la tabla que designamos como el hecho de nuestro cubo: **Ventas**.

Ilustración 91. Selección de la Tabla de Hechos



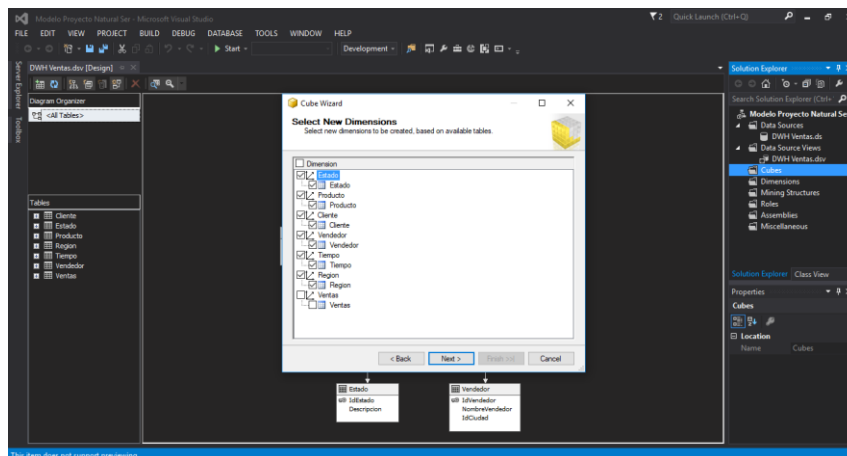
- Muestra las medidas que se utilizarán en el cubo.
 - ✓ Cantidad
 - ✓ Precio Promedio
 - ✓ Precio Total
 - ✓ Cantidad de Ventas

Ilustración 92. Selección de las Medidas para el Cubo OLAP



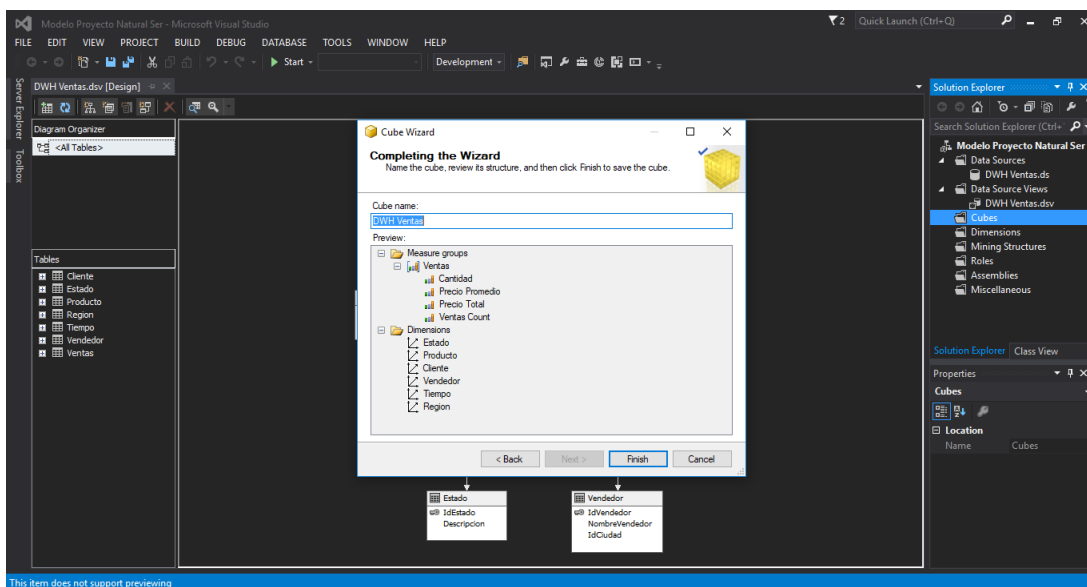
- Seleccionamos las dimensiones de nuestro modelo y no seleccionamos la tabla de Ventas puesto que será nuestro hecho.

Ilustración 93. Selección de las Dimensiones



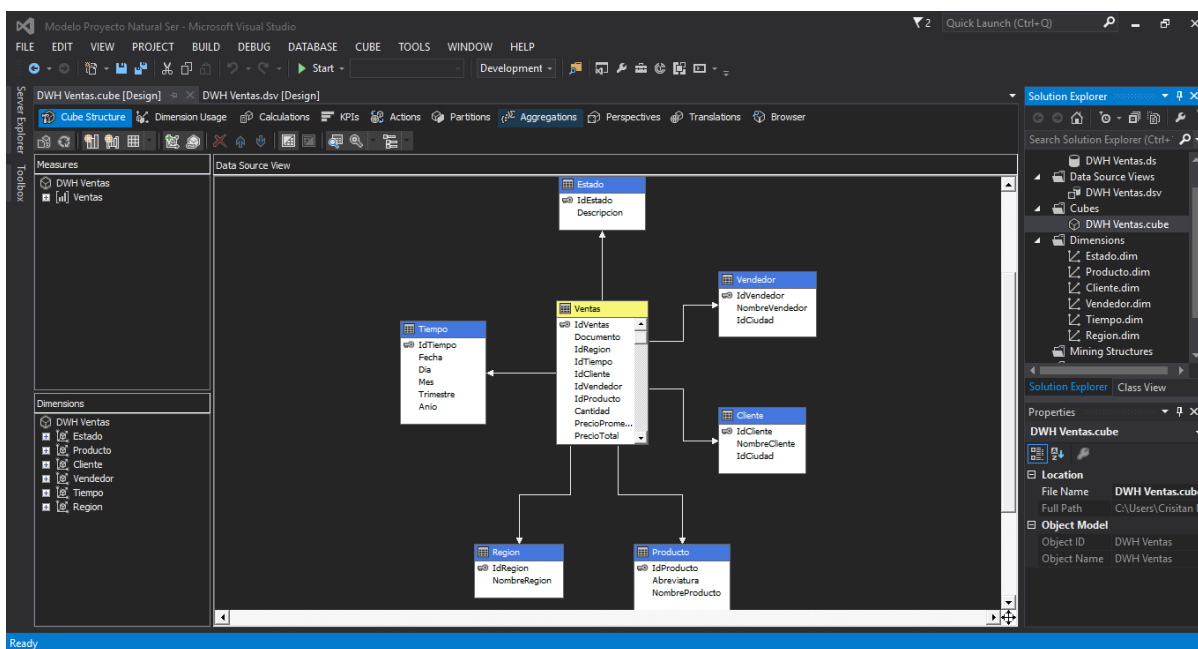
- Nos muestra la estructura general del cubo.

Ilustración 94. Estructura general del Cubo OLAP



- Visualizamos el cubo y podemos proceder a trabajar con cada dimensión.

Ilustración 95. Esquema Estrella del Modelo Multidimensional en SQL Server Data Tools



Jerarquías

Se crearon dos jerarquías en la dimensión tiempo:

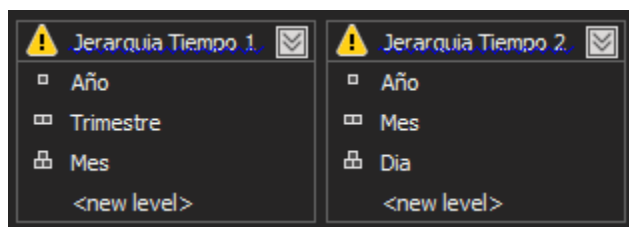
- La primera jerarquía se creó a partir de la relación entre los atributos año, trimestre y mes, en el cual el atributo año es el padre o atributo general y así se fue detallando, quedando de la siguiente forma:

AÑO → TRIMESTRE → MES

- La segunda jerarquía se creó a partir de la relación entre los atributos año, mes y día, en el cual el atributo año es el padre o atributo general y así se fue detallando, quedando de la siguiente forma:

AÑO → MES → DIA

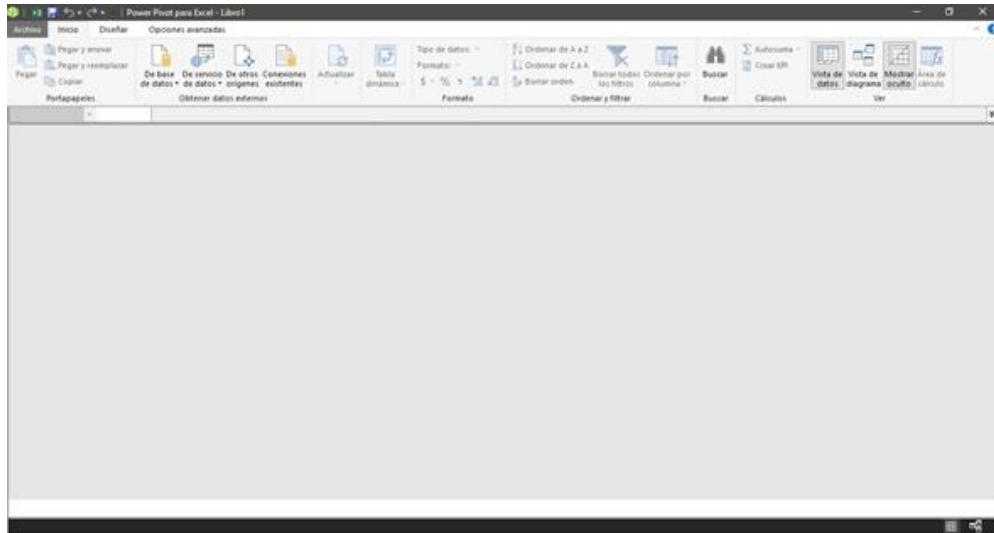
Ilustración 96. Jerarquías del Modelo Multidimensional



8.4.9 Construcción del Modelo Multidimensional en Excel Power Pivot

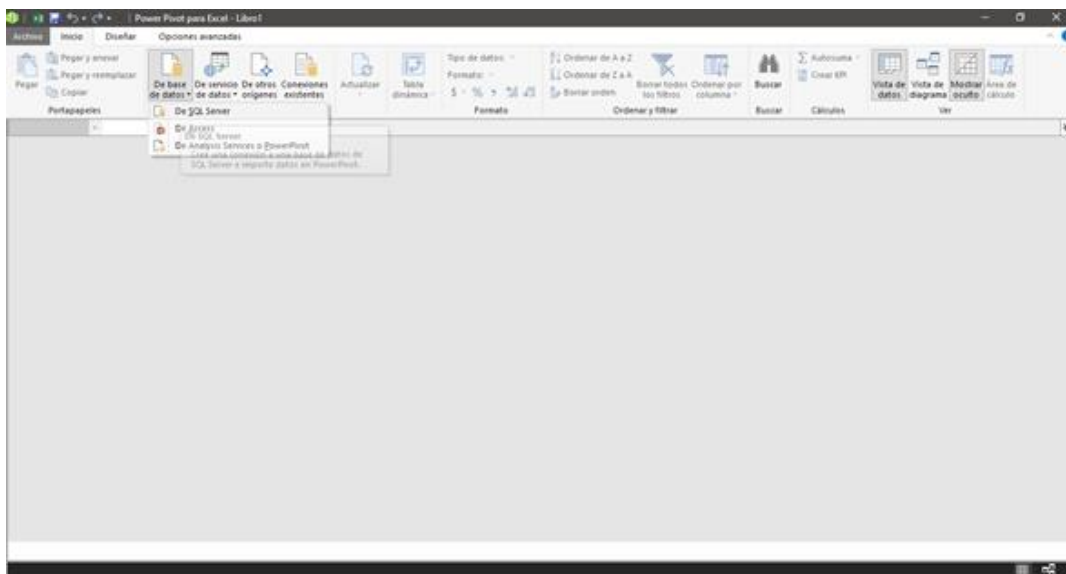
- Accedemos a Power Pivot

Ilustración 97. Acceso a Power Pivot en Excel



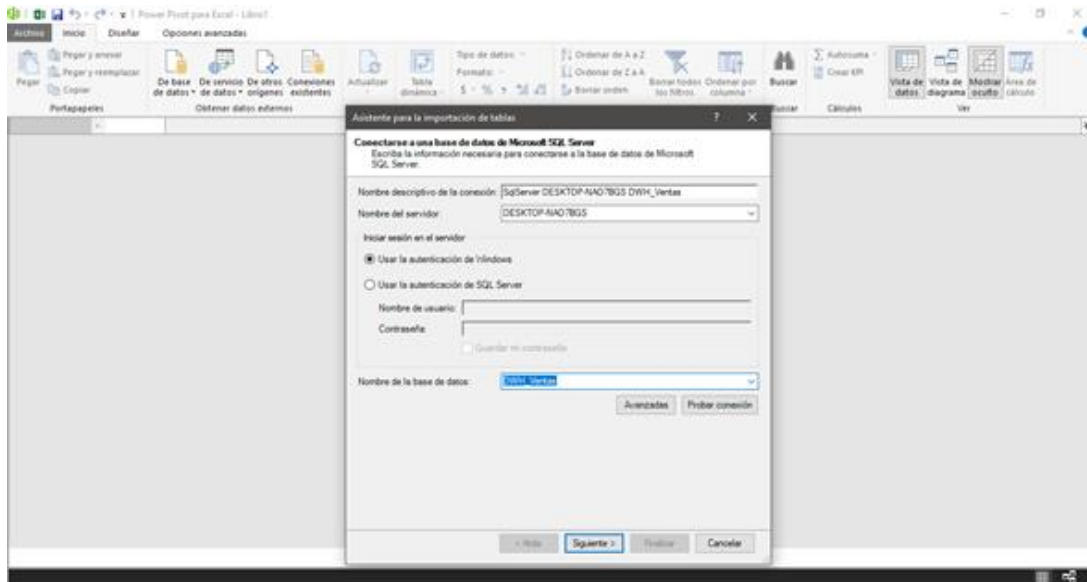
- Obtenemos datos de SQL Server

Ilustración 98. Selección de origen de datos



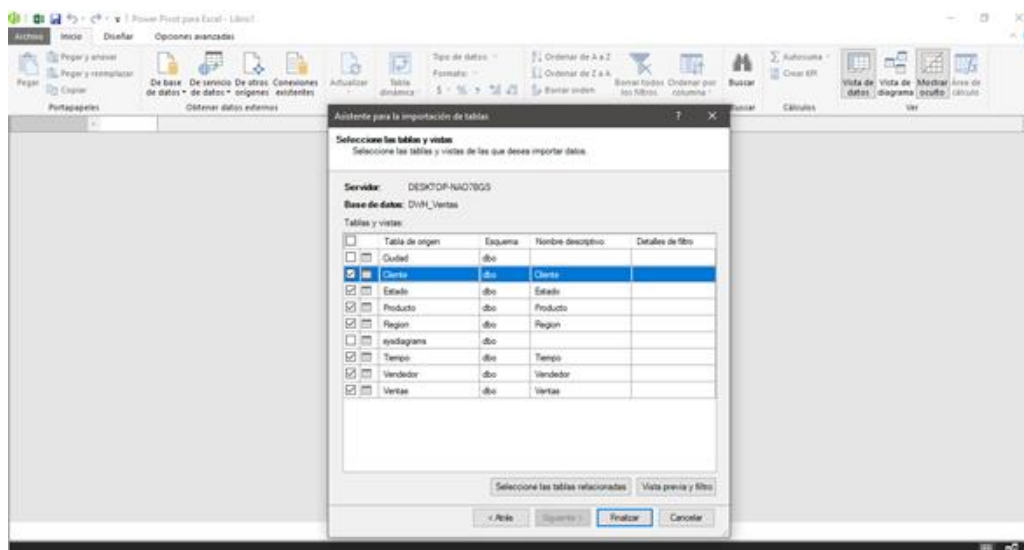
- Escribimos el nombre del servidor y el DWH para trabajar.

Ilustración 99. Selección de servidor y del Data Warehouse



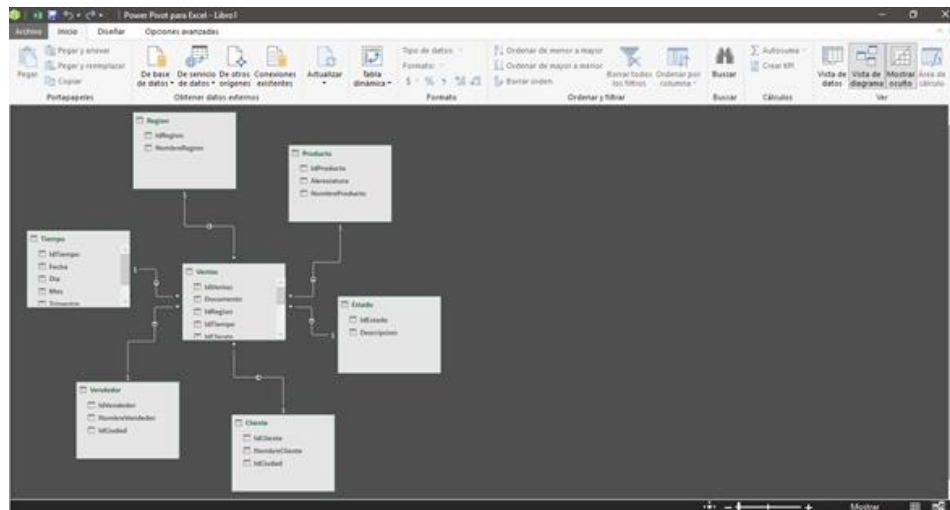
- Seleccionamos las tablas para el modelo

Ilustración 100. Selección de Dimensiones y Tabla de Hechos



- Nos muestra el modelo

Ilustración 101. Esquema Estrella del Modelo Multidimensional en Excel Power Pivot



Medidas:

Como se manejan valores negativos en la tabla de hechos, Power Pivot permite emplear fórmulas para crear medidas y así visualizar los datos como positivos.

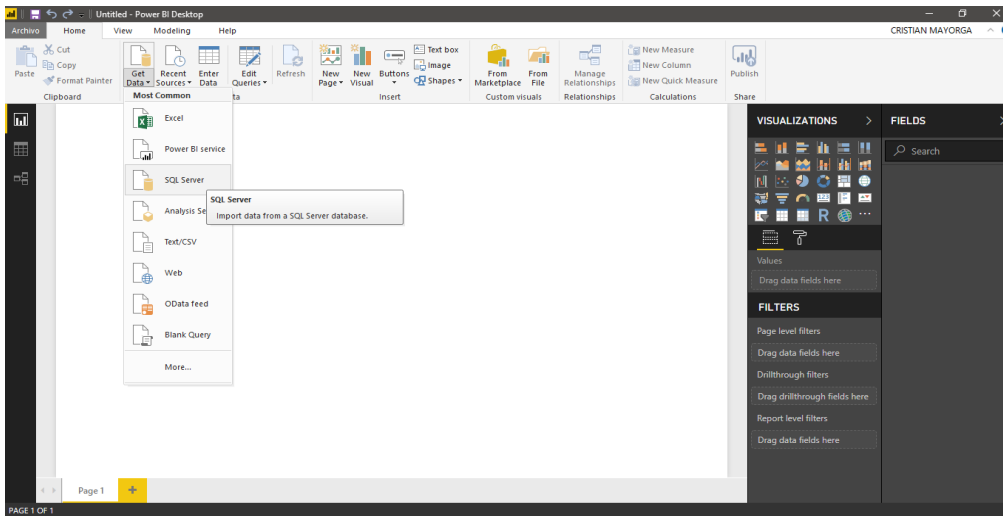
Medida Precio Total: =ABS(SUM(Ventas[PrecioTotal]))

Medida Cantidad: =ABS(SUM(Ventas[Cantidad]))

8.4.10 Construcción del Modelo Multidimensional en Power BI

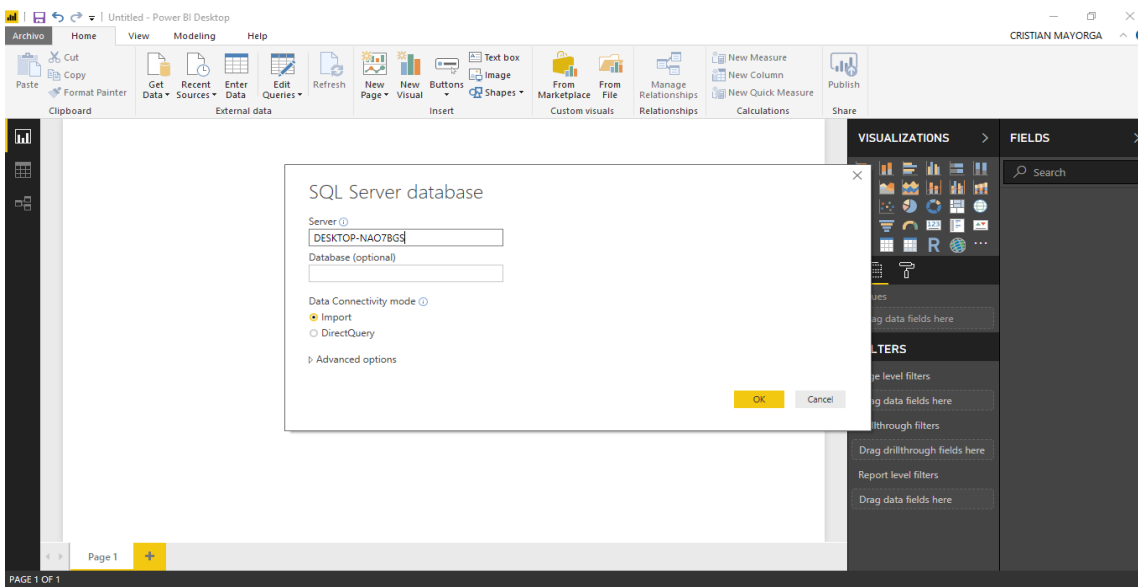
- Accedemos a Power BI y obtenemos datos de SQL Server

Ilustración 102. Acceso a Power BI y selección de origen de datos



- Escribimos el servidor

Ilustración 103. Selección del servidor



- Seleccionamos las tablas para el modelo y esperamos que cargue

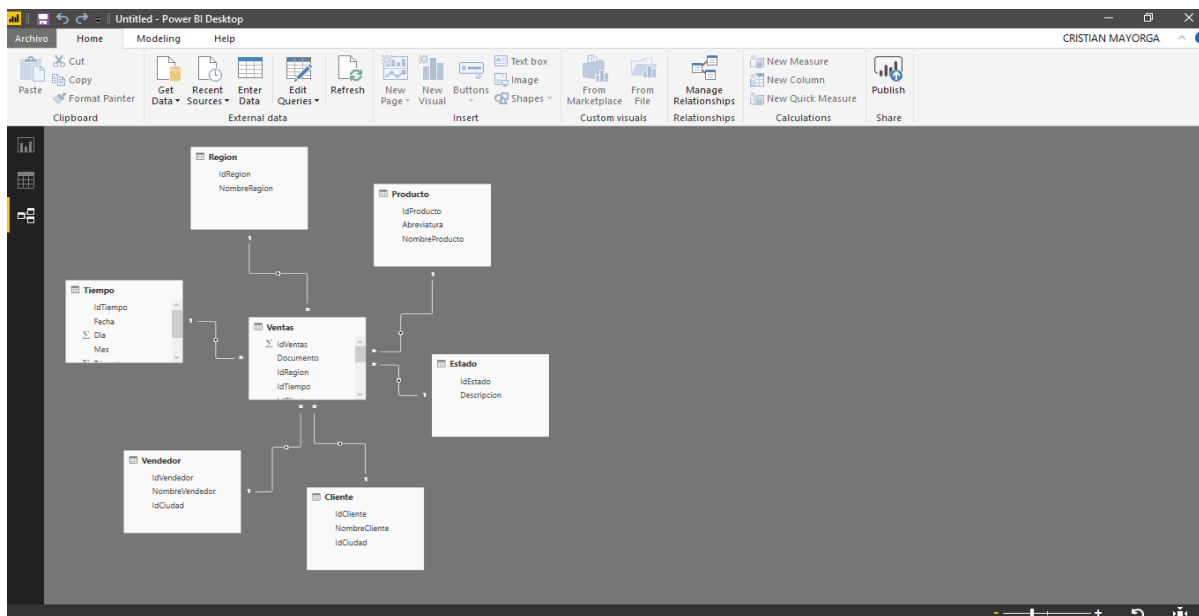
Ilustración 104. Selección de Dimensiones y Tabla de Hechos

The screenshot shows the Power BI Desktop interface. The Navigator window is open, displaying a tree view of data sources. The 'Ventas' table is selected. To the right, a preview of the 'Ventas' table is shown, downloaded on Sunday, April 8, 2018. The preview table has the following columns: IdVentas, Documento, IdRegion, IdTiempo, IdCliente, and IdVendedor.

IdVentas	Documento	IdRegion	IdTiempo	IdCliente	IdVendedor
60550	DMC 10	11	20160118	80852903	
60551	DMC 10	5	20160701	71595063	
60552	DMC 11	11	20160506	79487656	
60553	DMC 11	11	20160506	79487656	
60554	DMC 11	11	20160506	79487656	
60555	DMC 11	11	20160506	79487656	
60556	DMC 11	11	20160506	79487656	
60557	DMC 11	11	20160506	79487656	
60558	DMC 12	5	20160829	3405795	
60559	DMC 12	11	20161031	53094109	
60560	DMC 13	11	20160802	1023871824	
60561	DMC 13	11	20160802	1023871824	
60562	DMC 13	11	20160802	1023871824	
60563	DMC 13	11	20160802	1023871824	
60564	DMC 13	5	20160829	43531085	
60565	DMC 14	5	20160831	15438381	
60566	DMC 14	5	20160831	15438381	
60567	DMC 14	11	20161127	21055787	
60568	DMC 14	11	20161127	21055787	
60569	DMC 14	11	20161127	21055787	
60570	DMC 14	11	20161127	21055787	
60571	DMC 15	5	20160906	70130567	

- Nos muestra el modelo

Ilustración 105. Esquema Estrella del Modelo Multidimensional en Power BI



La construcción del modelo en esta herramienta es similar a Power Pivot y de igual forma se aplican las siguientes medidas a los datos.

Medidas

Como se manejan valores negativos en la tabla de hechos, Power BI permite emplear fórmulas para crear medidas y así visualizar los datos como positivos.

Medida Precio Total: Measure = ABS(SUM(Ventas[PrecioTotal]))

Medida Cantidad: Measure =ABS(SUM(Ventas[Cantidad]))

Medida Ventas: Measure = (COUNT(Ventas[Documento]))

8.4.11 Evaluación del Modelo

Como ya se vio anteriormente, contamos con tres opciones de herramientas para la creación y la presentación del modelo multidimensional, estas son: SQL Server Data Tools, Excel Power Pivot y Power BI. Enfocándonos en la experiencia de usuario del personal de la empresa decidimos mostrar los resultados en Excel Power Pivot y Power BI, por medio de los Dashboard para su fácil uso y entendimiento de los datos.

De igual forma dejamos planteado el modelo multidimensional en la herramienta SQL Server Data Tools, por si la empresa decide conformar un área de TI, con personal que tenga conocimiento avanzado en el uso de este tipo de herramientas.

A continuación, las herramientas seleccionadas con sus respectivos Dashboard y una breve explicación de cada una de sus partes.

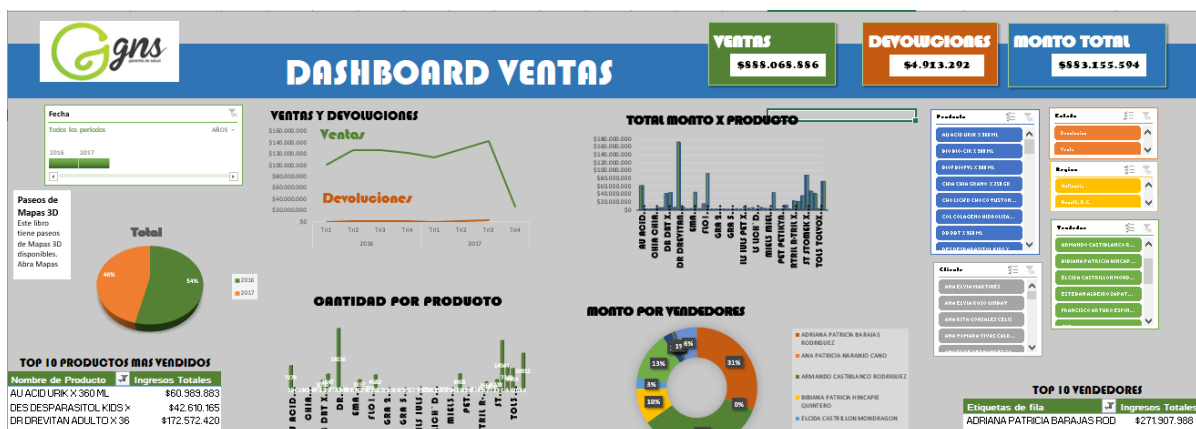
Dashboard Excel Power Pivot

Una vez realizado implementadas las medidas, realizamos el Dashboard donde visualizamos los datos, mediante tablas dinámicas, gráficos dinámicos, mapas en 3D, segmentación de datos etc.

El Dashboard en Excel, se construyó basados en tablas dinámicas, gráficos dinámicos, segmentación de datos, entre otros. Con esto podemos visualizar la información de forma precisa y ordenada para ayudar a los usuarios empresariales a tomar las mejores decisiones del negocio.

El Dashboard es el siguiente y cuenta con:

Ilustración 106. Dashboard Excel Power Pivot



- En la parte superior podemos ver: ventas, devoluciones y el monto total de las ventas en general. Estos datos irán cambiando a medida que se filtre la información.

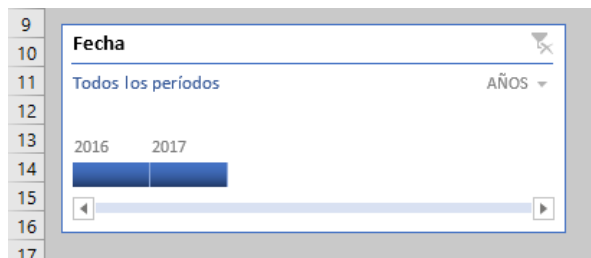
Ilustración 107. Cuadros de Valores



Segmentador de Fecha

Este segmentador permite visualizar los datos de acuerdo al año deseado o trimestre, mes día que se desee.

Ilustración 108. Segmentador de Fecha

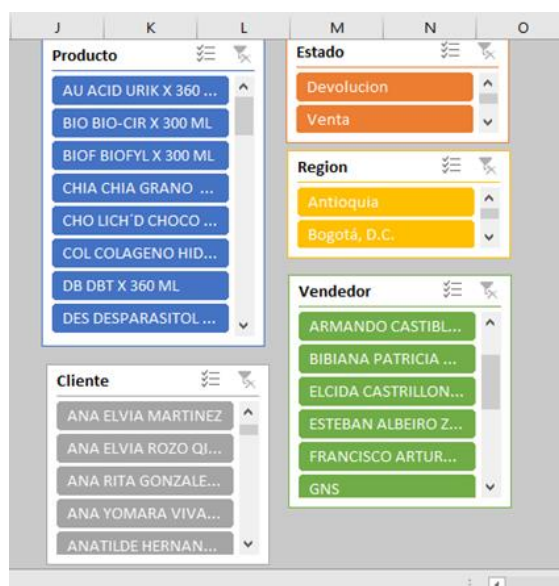


En ese mismo segmentador, utilizar un filtro donde podemos seleccionar: Años, trimestres, meses y días.

Segmentadores de Datos:

Permite segmentar los datos por un área específica. Al seleccionar alguno de estos, los gráficos y tablas cambian en torno a esa selección.

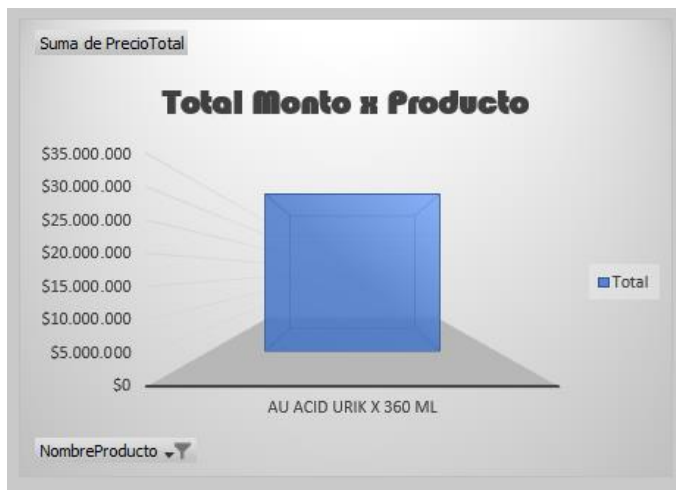
Ilustración 109. Segmentadores de Datos



Gráficos

En el centro del Dashboard encontramos varios gráficos los cuales cambian en base a las opciones de los segmentadores de datos y fecha que se seleccione.

Ilustración 110. Gráficos



Top 10

Además, el Dashboard cuenta con tablas que muestran el top 10 de productos más vendidos, el de los vendedores que más vendieron

Ilustración 111. Top 10 de Productos más vendidos

32	TOP 10 PRODUCTOS MAS VENDIDOS	
33		
34	Nombre de Producto	Ingresos Totales
35	AU ACID URIK	\$60.989.883
36	DES DESPARASITOL KIDS	\$42.610.165
37	DR DREVITAN ADULTO	\$172.572.420
38	FAR FAROT	\$43.559.651
39	GOL GOLESTROL GNS	\$92.304.372
40	OV OVAI	\$42.705.104
41	TOL1 TOLVOX ADULTO	\$88.042.789
42	TOL2 TOLVOX NIÑOS	\$46.652.520
43	TOL3 TOLVOX DIABETICO	\$40.798.197
44	ZAR ZARZAN	\$71.918.445
45	Total	\$702.153.546
46	Porcentaje de Ventas	79%
47		

Ilustración 112. Top 10 de los mejores Vendedores

TOP 10 VENDEDORES	
Etiquetas de fila	Ingresos Totales
ADRIANA PATRICIA	\$271.907.988
ARMANDO	\$304.032.514
BIBIANA PATRICIA	\$84.398.577
ELCIDA	\$26.084.609
ESTEBAN ALBEIRO	\$114.401.777
FRANCISCO ARTURO	\$21.816.180
GNS	\$1.000.046
JOSUE DANIEL	\$2.958.563
RAMIRO JOSE	\$5.208.851
WILSON	\$50.920.463
Total general	\$882.729.570

Mapa 3D

Se puede visualizar la información de ventas de las 2 ciudades en el mapa, de forma dinámica. Se puede observar de 2 formas: el crecimiento de la venta de cada producto o bien la variación de las ventas de cada producto, dando la ventaja de poder ver la fecha y la el valor del mismo.

Ilustración 113. Mapa 3D – Paseo 1

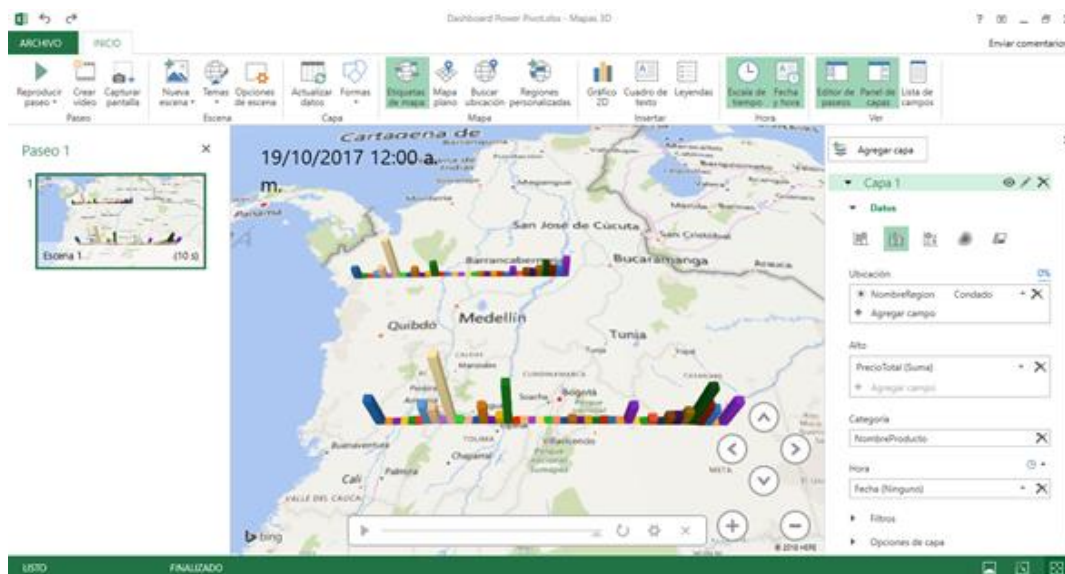
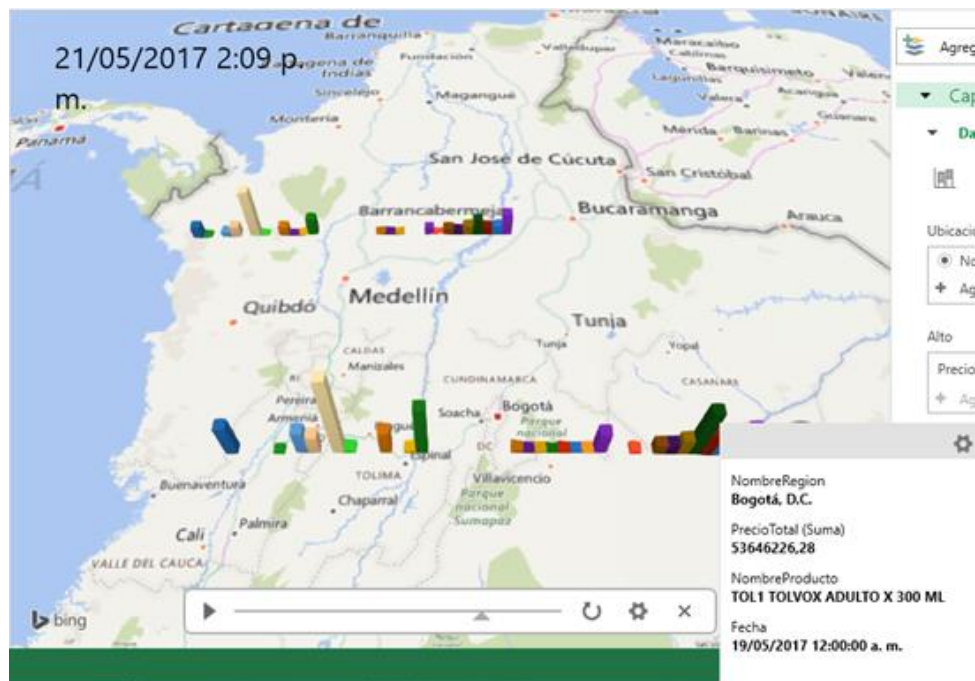


Ilustración 114. Mapa 3D – Paseo 2



Tablas Dinámicas

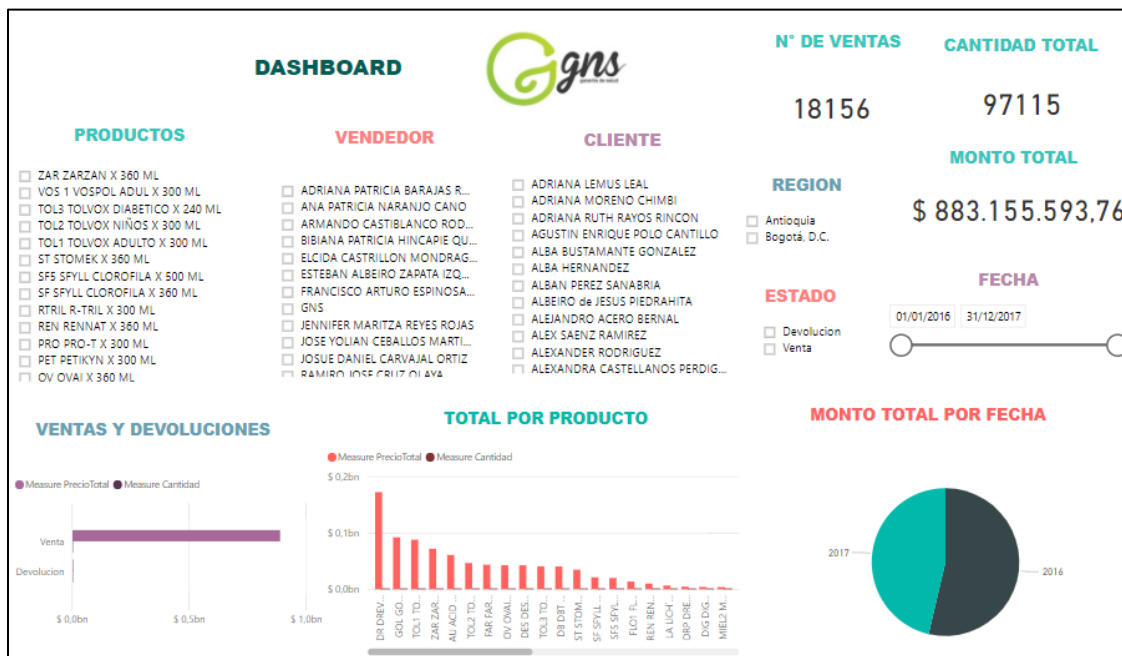
Adicional a todas estas características, podemos visualizar los datos en forma de tablas, donde el usuario puede crear su propio reporte para ver la información que desee.

Dashboard Power BI

El Dashboard en Power BI, se construyó basados en gráficos dinámicos, segmentación de datos, entre otros. Con esto podemos visualizar la información de forma precisa y ordenada para ayudar a los usuarios empresariales a tomar las mejores decisiones del negocio.

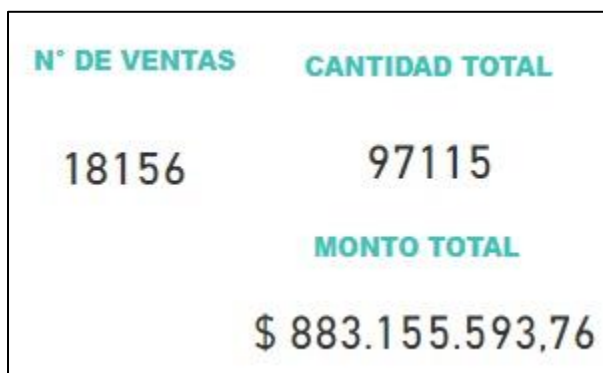
El Dashboard es el siguiente y cuenta con:

Ilustración 115. Dashboard Power BI – Página 1



- En la parte superior derecha podemos ver: el monto total y cantidad de las ventas en general, también la cantidad total de productos que se han vendido. Estos datos irán cambiando a medida que se filtre la información.

Ilustración 116. Cuadros de Valores



Segmentadores de Datos

Permite seleccionar cualquiera de los segmentos de datos como: Producto, Estado, Región, Cliente y Vendedor, al seleccionar alguno de los datos de los diferentes segmentos, los gráficos irán cambiando y mostrando la información de las ventas deseada.

Ilustración 117. Segmentadores de Datos

PRODUCTOS	VENDEDOR	CLIENTE	REGION
<input type="checkbox"/> ZAR ZARZAN X 360 ML	<input type="checkbox"/> ADRIANA PATRICIA BARAJAS R...	<input type="checkbox"/> ADRIANA LEMUS LEAL	<input type="checkbox"/> Antioquia
<input type="checkbox"/> VOS 1 VOSPOL ADUL X 300 ML	<input type="checkbox"/> ANA PATRICIA NARANJO CANO	<input type="checkbox"/> ADRIANA MORENO CHIMBI	<input type="checkbox"/> Bogotá, D.C.
<input type="checkbox"/> TOL3 TOLVOX DIABETICO X 240 ML	<input type="checkbox"/> ARMANDO CASTIBLANCO ROD...	<input type="checkbox"/> ADRIANA RUTH RAYOS RINCON	
<input type="checkbox"/> TOL2 TOLVOX NIÑOS X 300 ML	<input type="checkbox"/> BIBIANA PATRICIA HINCAPIE QU...	<input type="checkbox"/> AGUSTIN ENRIQUE POLO CANTILLO	
<input type="checkbox"/> TOL1 TOLVOX ADULTO X 300 ML	<input type="checkbox"/> ELCIDA CASTRILLON MONDRAG...	<input type="checkbox"/> ALBA BUSTAMANTE GONZALEZ	
<input type="checkbox"/> ST STOMEK X 360 ML	<input type="checkbox"/> ESTEBAN ALBEIRO ZAPATA IZQ...	<input type="checkbox"/> ALBA HERNANDEZ	
<input type="checkbox"/> SF5 SFYLL CLOROFILA X 500 ML	<input type="checkbox"/> FRANCISCO ARTURO ESPINOSA...	<input type="checkbox"/> ALBAN PEREZ SANABRIA	
<input type="checkbox"/> SF SFYLL CLOROFILA X 360 ML	<input type="checkbox"/> GNS	<input type="checkbox"/> ALBEIRO de JESUS PIEDRAHITA	
<input type="checkbox"/> RTRIL R-TRIL X 300 ML	<input type="checkbox"/> JENNIFER MARITZA REYES ROJAS	<input type="checkbox"/> ALEJANDRO ACERO BERNAL	
<input type="checkbox"/> REN RENNAT X 360 ML	<input type="checkbox"/> JOSE YOLIAN CEBALLOS MARTI...	<input type="checkbox"/> ALEX SAENZ RAMIREZ	<input type="checkbox"/> Devolucion
<input type="checkbox"/> PRO PRO-T X 300 ML	<input type="checkbox"/> JOSUE DANIEL CARVAJAL ORTIZ	<input type="checkbox"/> ALEXANDER RODRIGUEZ	<input type="checkbox"/> Venta
<input type="checkbox"/> PET PETIKYN X 300 ML	<input type="checkbox"/> RAMIRO JOSE CRUZ OLIVERA	<input type="checkbox"/> ALEXANDRA CASTELLANOS PERDIG...	
<input type="checkbox"/> OV OVAI X 360 ML			

Segmentador de Fecha

Este segmentador permite visualizar los datos de acuerdo a la fecha, permitiendo filtrar el año, trimestre, mes y/o día que se desee.

Ilustración 118. Segmentador de Fecha

FECHA

01/01/2016 31/12/2017

Gráficos

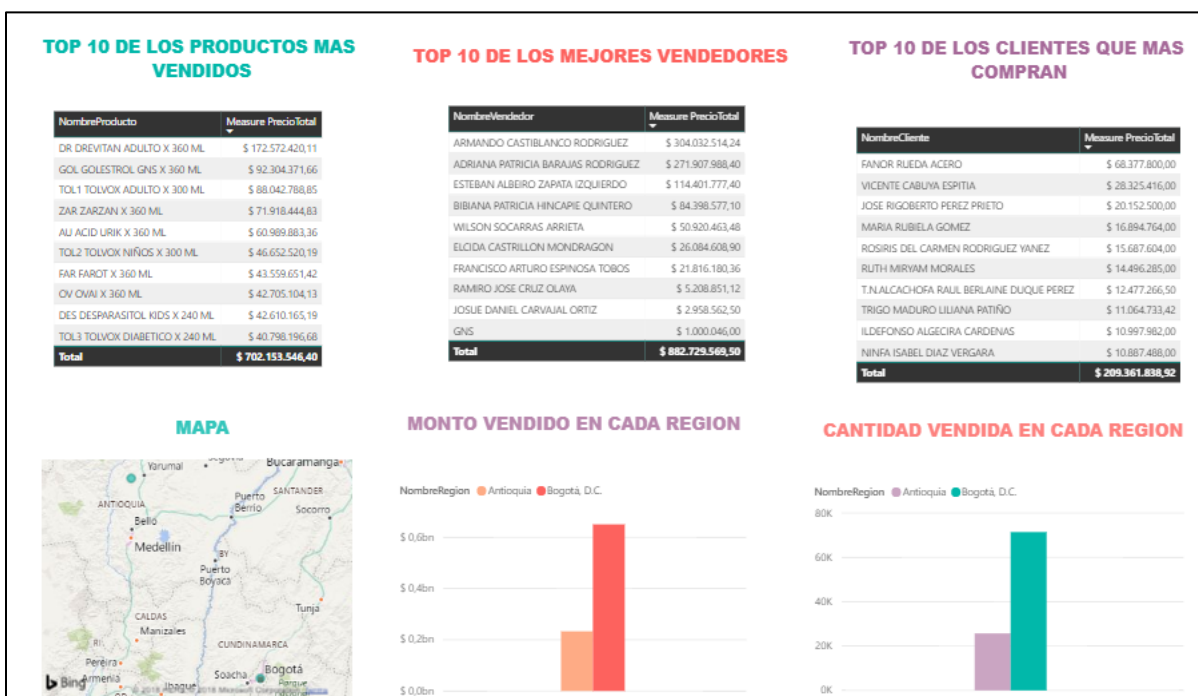
Estos cambian en base a las opciones de los segmentadores de datos y fecha que se seleccione, son los que se encuentran en la parte inferior del Dashboard.

Ilustración 119. Gráficos



- En la segunda página del Dashboard se puede encontrar:

Ilustración 120. Dashboard Power BI – Página 2



Top 10

Son tablas que muestran el top 10 de productos más vendidos, el de los mejores vendedores y los clientes que más han comprado.

Ilustración 121. Top 10



Mapa

Se puede visualizar la ubicación de los puntos de venta, en este caso el de las dos regiones, también se pueden observar dos gráficos que muestran el monto total de las ventas de cada ciudad y la cantidad de productos vendidos en estas.

Ilustración 122. Mapa



- Por último, en la tercera página del Dashboard se puede encontrar:

Tablas Dinámicas

Una serie de tablas que muestran la cantidad de productos vendidos de acuerdo a cada una de las dimensiones y el monto vendido por fecha y por estado (venta y devolución):

Ilustración 123. Tablas Dinámicas

CANTIDAD POR CLIENTE				CANTIDAD VENDIDO POR FECHA		MONTO VENDIDO POR FECHA	
Nombre Cliente	2016	2017	Total	Year	Measure Cantidad	Year	Measure PrecioTotal
FANCY RUEDA ACERO	3062	3159	6221	2016	53448	2016	\$ 474.124.240,36
VICENTE CABRIN ESPITA	1457	1265	2722	2017	43667	2017	\$ 469.031.353,40
T.NALCACHI RAUL BERLAINE DUQUE PEREZ	1018	640	1658	Total	97115	Total	\$ 883.155.593,76
JOSE RIGOBERTO PEREZ PRIETO	841	674	1515				
MARIA RUBIELA GOMEZ	978	464	1442				
ROSIRIS DEL CARMEN RODRIGUEZ VANEZ	660	730	1390				
MONICA GONZALEZ MONROY	512	782	1294				
RUTH MRYAM MORALES	442	782	1224				
NATURAL FOOD VEGETAL 3IN1 JIRO SANTIAGO	808	320	1188				
TRIGO MADURO LULIANA PATIÑO	555	583	1138				
MAURICIO CARDENAS	642	490	1132				
ILDEFONSO ALGECIRA CARDENAS	335	738	1073				
T.N. HORUS FRANCISCO LUIS CASTAÑO	455	519	974				
PEDRO BUSTOS CASA VERDE	293	670	963				
Total	53448	43667	97115				

CANTIDAD POR ESTADO				MONTO POR ESTADO	
Descripcion	Measure Cantidad	Measure PrecioTotal	Descripcion	Measure PrecioTotal	Measure PrecioTotal
Devolucion	386		Devolucion		\$ 4.913.292,32
Venta	97501		Venta		\$ 888.068.886,08
Total	97115		Total		\$ 883.155.593,76

CANTIDAD POR VENDEDOR				CANTIDAD POR PRODUCTO			
Nombre Vendedor	2016	2017	Total	Nombre Producto	2016	2017	Total
ADRIANA PATRICIA BARRAS RODRIGUEZ	18112	13947	32059	AJ ACID LRIX X 360 ML	4028	3212	7240
ANA PATRICIA NARANJO CANO	36		36	BIO BIO- CIR X 300 ML		6	6
ARMANDO CASTIBLANCO RODRIGUEZ	15999	13777	29776	BIOF BICFVL X 300 ML		14	14
BIBIANA PATRICIA HINCAPIE QUINTERO	8039	1530	9569	CHA CHA GRANO X 250 GR		42	42
ELIODA CASTRILLON MONDRAGON	1983	940	2923	CHO LICH O CHOCO MUSTOP X 600 GR	41	99	140
ESTEBAN ALBEIRO ZAPATA LOZUECO	3512	8943	12455	COL COLAGENO HIDROLIZADO X 250 GR		342	342
FRANCISCO ARTURO ESPINOSA TOBOS	2152	581	2733	DB DBT X 360 ML	1873	1160	3033
GNS	28	32	60	DES DESPARASITOL KIDS X 240 ML	2423	2395	4818
JENIFER MARITZA REYES ROJAS		65	65	DIG DIGEST X 360 ML		292	292
JOSE YOLIAN CEBALLOS MARTINEZ		11	11	DR DREVITAN ADULTO X 360 ML	10047	7942	17989
JOSUE DANIEL CARVAJAL CRIZ	194	110	304	DRP DREVITAN PLUS X 360 ML	339	1	338
RAMIRO JOSE CRUZ OLIVA	490		490	TRUF TRUFETE V 300 ML		46	46
WILSON SOCARRAS ARRIETA	2975	3731	6706	Total	53448	43667	97115
Total	53448	43667	97115				

8.5 EVALUACIÓN DEL MODELO MULTIDIMENSIONAL

En la evaluación del modelo se busca que este dé respuesta eficiente a las consultas realizadas, mediante una extracción de los datos para generar información. En este punto si deseamos saber la cantidad de productos vendidos, en la ciudad de Medellín y en el tercer trimestre del año 2016, el modelo mostrará esa información deseada.

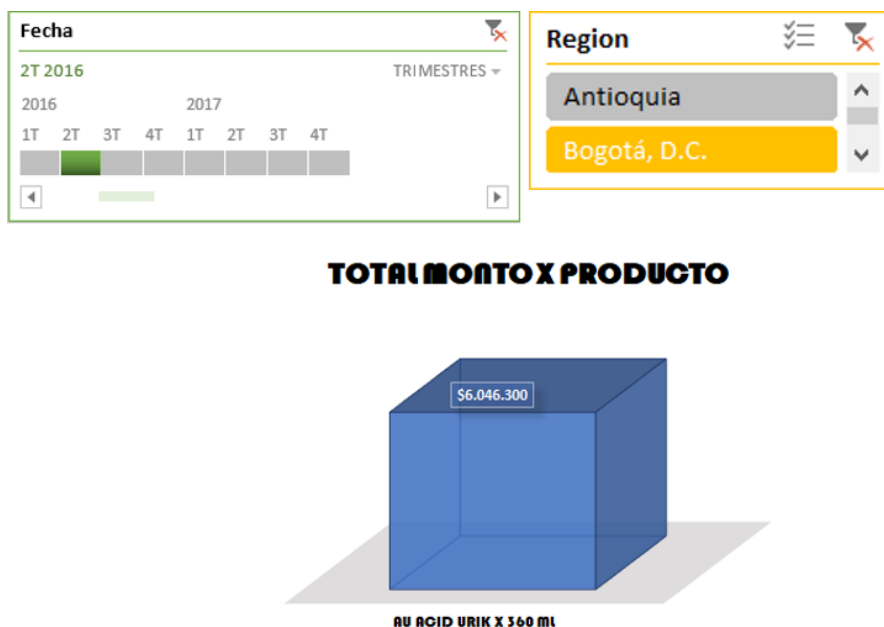
Para la evaluación realizaremos 3 consultas diferentes en cada herramienta:

8.5.1 Evaluación de resultados en Power Pivot

Consulta 1

¿Cuál fue el importe del producto Acid Urik en el segundo trimestre del año 2016 en la ciudad de Bogotá?

Ilustración 124. Consulta 1 – Dashboard Excel Power Pivot



En el segundo trimestre del año 2016 la empresa vendió en la ciudad de Bogotá un importe por 6'046.300 pesos del producto AU ACID URIK.

Consulta 3

¿Cuál fue la cantidad de unidades vendidas por el vendedor Esteban Albeiro Zapata Izquierdo del producto TOLVOX ADULTO en la ciudad de Medellín, entre marzo y agosto del año 2016?

Ilustración 126. Consulta 3 – Dashboard Excel Power Pivot



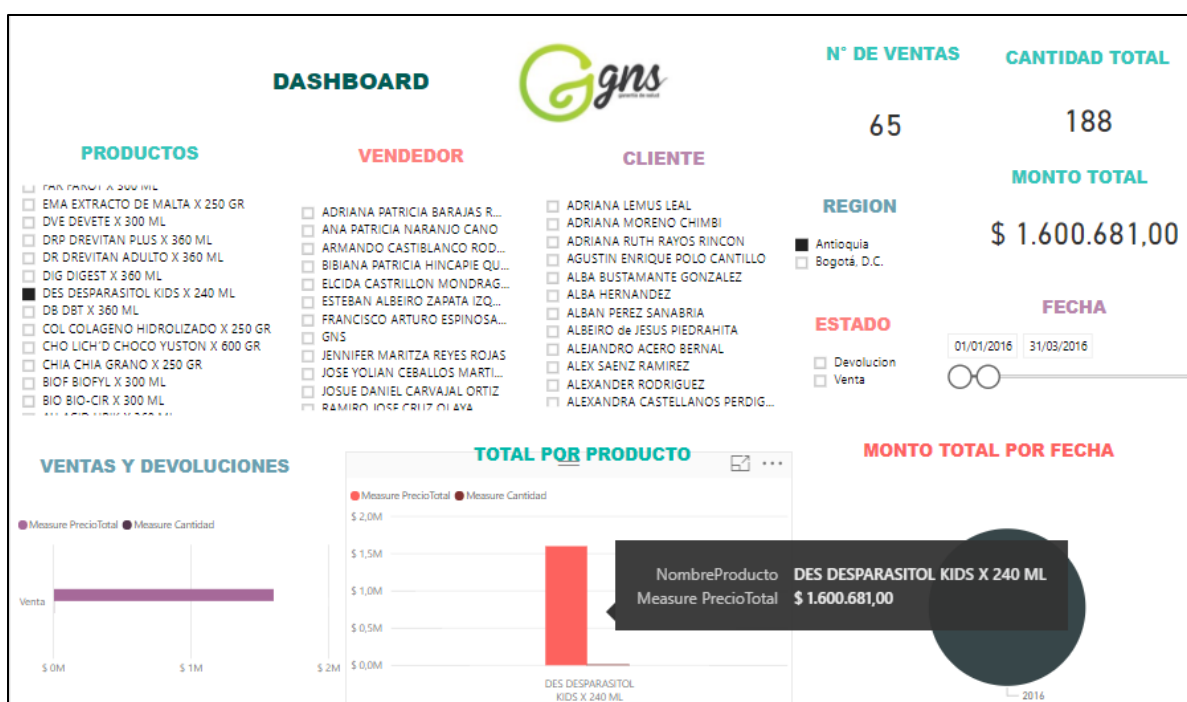
Entre los meses de marzo a agosto del año 2016, el vendedor Esteban Albeiro Zapata Izquierdo vendió un total de 307 unidades del producto TOLVOX ADULTO en la ciudad de Medellín.

8.5.2 Evaluación de resultados en Power BI

Consulta 1

¿Cuál es el importe del producto DESPARASITOL KIDS en el primer trimestre del año 2016, en la ciudad de Medellín?

Ilustración 127. Consulta 1 – Dashboard Power BI



En el primer trimestre del año 2016 la empresa vendió en la ciudad de Medellín un importe por 1'600.681 pesos del producto DESPARASITOL KIDS.

Consulta 2

¿Cuál es el importe de las ventas en el cuarto trimestre del año 2016 en la ciudad de Bogotá?

Ilustración 128. Consulta 2 – Dashboard Power BI

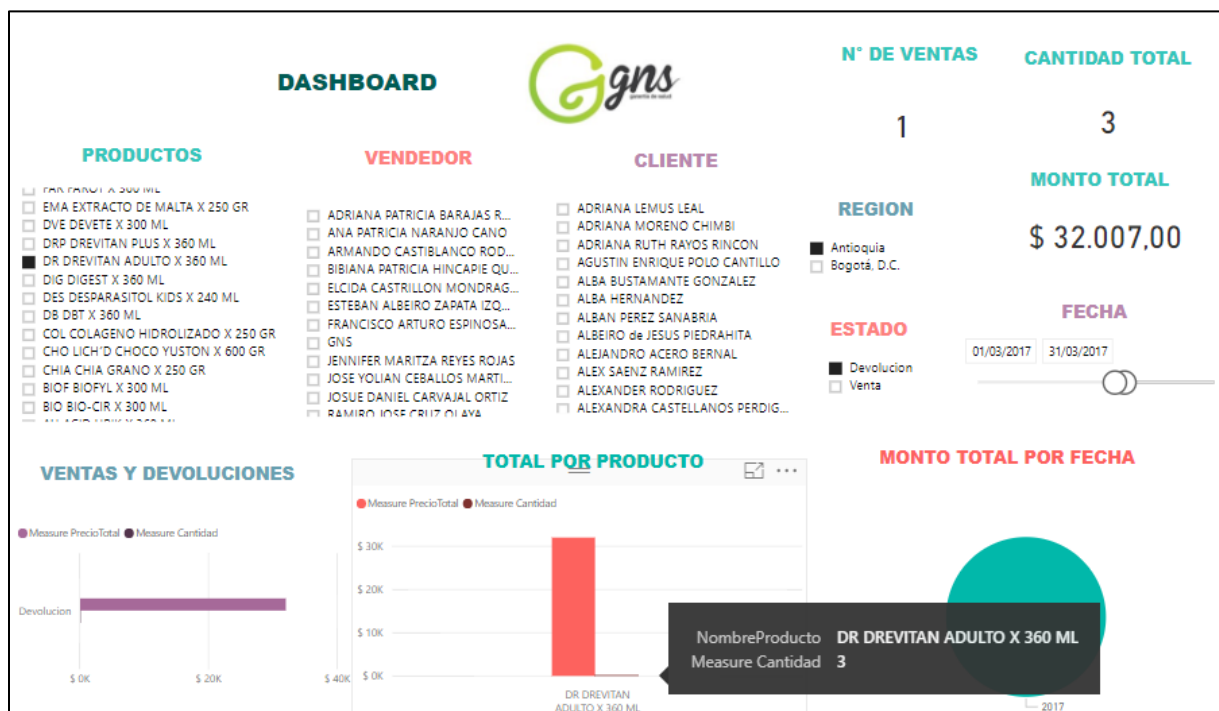


En el cuarto trimestre del año 2016, la empresa vendió en la ciudad de Bogotá un importe por 87'122.503 pesos.

Consulta 3

¿Cuál fue la cantidad devuelta del producto DREVITAN ADULTO en el mes de marzo del año 2017 en la ciudad de Medellín?

Ilustración 129. Consulta 3 – Dashboard Power BI



La cantidad devuelta del producto DREVITAN ADULTO en el mes de marzo de 2017 en Medellín fue de 3 productos.

8.5.4 Determinar los próximos pasos

Como se puede observar la solución BI, plantea un proceso eficiente a la hora de analizar la información, puesto que permite visualizarla desde las diferentes dimensiones propuestas y así proporciona al Gerente o Ejecutivo una visión del comportamiento de las ventas a través del tiempo; es decir el histórico de las ventas de la empresa. De esta forma puede tomar las mejores decisiones para alcanzar las oportunidades de mejora a corto y mediano plazo.

Como la solución BI está funcionando, se plantea implementar una solución BA (Business Analytics) donde si bien ya podemos mirar el histórico de las ventas (Inteligencia de Negocios Descriptiva), ahora podríamos avanzar hacia una Analítica de Negocios con una mirada hacia el futuro (Planificación y Simulación, Recomendación), mediante un uso más profundo de la Minería de Datos y basados en esto, proponer modelos predictivos de negocio, de modo que se pueda contar con un soporte complementario a la solución BI para tomar las mejores decisiones y aumentar la competitividad del negocio mediante su análisis detallado y tecnológicamente sostenible en el tiempo.

8.6 DESPLIEGUE

8.6.2 Cartillas Digitales

Se desarrollaron 2 cartillas que orienta la forma de uso correcto de la solución informática.

Cartilla “Manual Técnico”

Este manual técnico permite al usuario de la empresa conocer desde la configuración de la solución informática hasta su funcionamiento, resaltando cada uno de los pasos a tener en cuenta a la hora de implementarla, se hace lo más específico y entendible posible ya que no se cuenta con un experto en sistemas. Se requieren de unos programas especiales para que la solución informática funcione adecuadamente tales como SQL Server Management Studio, Xampp, Excel y Power BI, reuniendo la información necesaria para que el usuario conozca y utilice la solución informática de una forma adecuada.

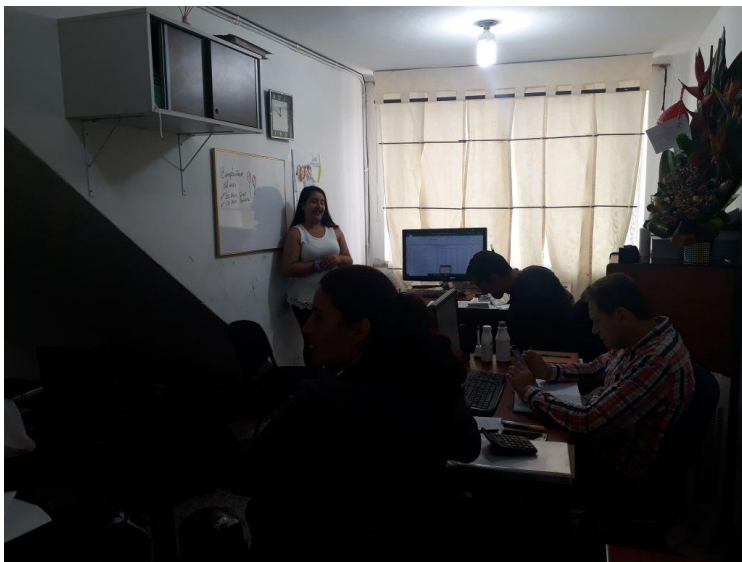
Cartilla “Manual de Usuario”

Este manual permite al usuario de la empresa conocer el manejo de la solución informática, resaltando cada uno de los pasos a tener en cuenta a la hora de la visualización e interacción con los Dashboards de las herramientas Excel y Power BI, se hace lo más específico y entendible posible. Reuniendo la información necesaria para que el usuario conozca y utilice la solución informática de una forma adecuada.

8.6.2 Capacitación en la Empresa “Grupo Natural Ser”

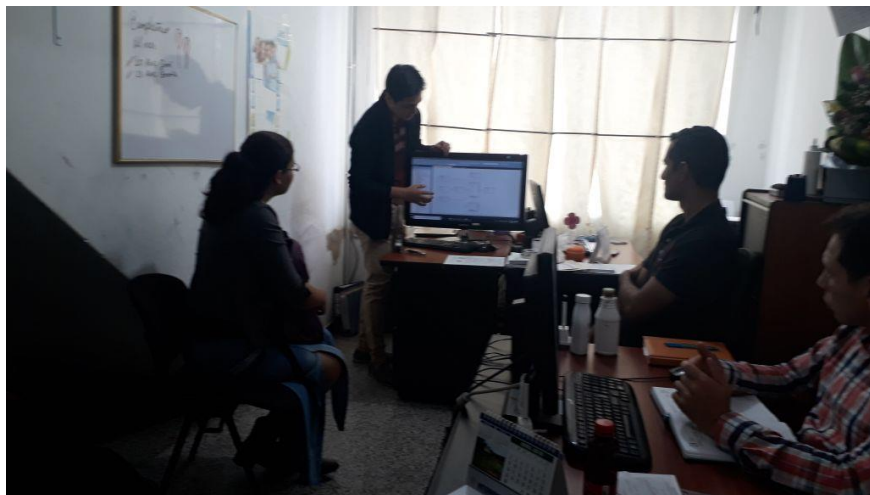
La capacitación en la empresa se llevó a cabo el día 24 de Abril del año 2018, en donde se explicó cada parte de la solución informática de BI planteada. Iniciamos explicando la extracción y transformación de los datos y el proceso que significó el haber recibido los datos suministrados en informes y su posterior transformación en un solo Data Set.

Ilustración 130. Evidencia Capacitación Empresa 1



Continuamos con la visualización del diseño de la Bodega de Datos y cómo pueden utilizarla como base tecnológica cuando la compañía cuente formalmente con área de TI.

Ilustración 131. Evidencia Capacitación Empresa 2



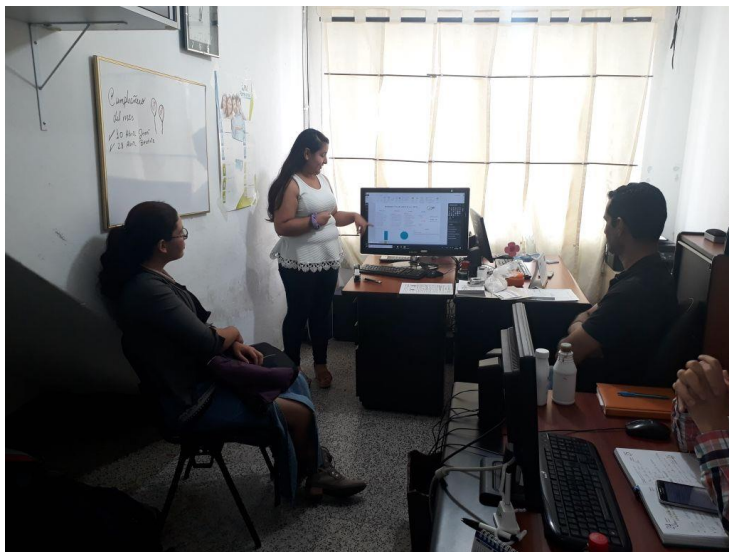
De igual forma presentamos el funcionamiento del aplicativo en lo que corresponde al cargue de los datos a la Bodega de Datos, explicando cada paso a seguir, como también los requerimientos técnicos mínimos a cumplir del archivo Excel (fuente de datos) para lograr un cargue exitoso.

Ilustración 132. Evidencia Capacitación Empresa 3



Explicamos la creación del Modelo Multidimensional y finalizamos con la demostración de los Dashboard tanto en Excel como en Power BI y la utilidad de cada uno.

Ilustración 133. Evidencia Capacitación Empresa 4



En la capacitación estuvieron presentes los Gerentes General y Comercial, el encargado del Area de Sistemas, el encargado del Area Financiera y una secretaria, quienes manifestaron su gusto por la Solución Informática implementada para la compañía y por la capacitación recibida, dado que esta Solución Informática les permitió conocer las bondades del manejo de una herramienta BI y los beneficios que les puede traer a la hora de analizar históricamente sus ventas.

8.6.3 Planificación de la monitorización y del mantenimiento

Para qué la solución informática esté funcionando correctamente, se recomienda realizar el proceso de Extracción y Transformación cada mes, puesto que son procesos que requieren tiempo. De igual forma durante ese periodo se recomienda realizar un archivo Excel nuevo con la misma estructura del anterior, pero alimentando solamente la hoja de ventas. De ser necesario agregar un nuevo cliente o un nuevo producto, entre otros, se debe adicionar en la hoja correspondiente y así se garantiza la integridad de los datos.

Si se desea alimentar siempre el mismo Excel, se deben realizar cambios en el aplicativo de tal forma que este lea el archivo Excel y lo compare con los datos presentes en la Bodega de Datos e ingrese solo los datos nuevos. De igual forma también se debería optimizar el programa reduciendo el tiempo de espera en la validación y en la carga de los datos a la Bodega de Datos.

8.6.4 Requerimientos para la implementación

Para la implementación de la solución en la empresa, se requiere de los programas previamente mencionados, junto con sus correspondientes licencias, lo que hace que el uso general de la solución sea seguro y eficiente para la empresa.

Entre las licencias y costos de la aplicación tenemos:

XAMPP

Software libre y se encuentra bajo la licencia GNU/GPL (*GNU General Public License*), esta “es una licencia de derecho de autor ampliamente usada en software libre y código abierto, y garantiza a los usuarios finales la libertad de usar, estudiar, compartir (copiar) y modificar el software.” por lo cual no genera un costo adicional a la solución.

SQL Server Management Studio

El precio para adquirir la licencia de ese motor de base de datos según la página oficial de Microsoft está en US\$3,717 dólares que en valor presente aproximado sería de \$10,430,942.00 pesos colombianos en la versión Standar.

Office 365

Este incluye el servicio necesario para utilizar tanto Excel como Power BI (el cual requiere una cuenta corporativa). El precio anual por usuario sería aproximadamente de \$269.400.00 pesos colombianos.

Aplicativo y Bodega de Datos

Teniendo en cuenta el diseño del aplicativo y el diseño de la Bodega de Datos, su precio se estima en \$2'000.000.00 de pesos colombianos.

Tabla 11. Costo de la solución informática

Herramienta	Precio
XAMPP	No aplica
SQL Server Management Studio	10'430.942
Office 365	269.400
Aplicativo y Bodega de Datos	2'000.000
Total	12'700.342

El costo estimado de la solución tecnológica sería de \$**12'700.342.00** de pesos colombianos.

GLOSARIO

Data Management: Es la implementación de políticas y procedimientos que las organizaciones utilizan para la gestión de sus datos comerciales. El Data Management Association (DAMA) lo define como: “La gestión de recursos de datos es el desarrollo y ejecución de arquitecturas, políticas, prácticas y procedimientos que gestionan apropiadamente las necesidades del ciclo de vida completo de los datos de una empresa.”

Business Intelligence: Es un amplio conjunto de herramientas, tecnologías, aplicaciones entre otras, que permiten recopilar, almacenar, analizar y proporcionar acceso a esos datos, transformados en información, a los usuarios empresariales con el fin de tomar mejores decisiones empresariales.

Business Analytics: Son un conjunto de herramientas, técnicas y habilidades que permiten ver el histórico comercial de una organización y así obtener información del rendimiento de la misma a futuro.

Data Mining: Es el proceso utilizado para encontrar anomalías, patrones y correlaciones en una gran cantidad de datos para predecir los resultados. Este proceso utiliza una gran cantidad de técnicas para aumentar los ingresos, reducir los costos, mejorar relaciones con el cliente entre otros.

Data Warehouse: Es una bodega que almacena grandes cantidades de datos que se recopilan e integran desde múltiples fuentes de datos

Modelo Multidimensional: Es una estructura para analizar los datos del negocio a través de varias dimensiones. De esta forma se logra aumentar el rendimiento de las consultas en los datos empresariales.

Cubo OLAP: Es una estructura de datos que supera a las bases de datos relacionales en el área de análisis de datos. Estos pueden manejar gran cantidad de datos y simultáneamente proporcionar acceso a esos datos.

Dimensión: Son tablas que almacenan los detalles acerca de los hechos, cada una de estas tablas tiene los campos y atributos que describen los procesos del negocio.

Hecho: Tabla que contiene las medidas que serán utilizados para el análisis de los datos.

Dato: Una unidad que hace referencia a una parte de algún tipo de información

Entidad: Representación de un objeto o concepto del mundo real en una base de datos.

Consulta: Forma en la que se puede acceder a los datos de una base de datos.

Base de Datos: Conjunto de datos que se relacionan en un contexto y se almacenan sistemáticamente.

Extracción de datos: Proceso en el cual se accede y obtiene datos desde alguna fuente.

Transformación de datos: Proceso en el que se aplican reglas de negocio para convertir los datos que se cargaran en la bodega de datos.

Carga de datos: Proceso en el que se cargan los datos al sistema después de ser transformados.

Esquema: Representación simbólica del diseño de una base de datos, bodega de datos, modelo multidimensional etc.

Análisis de datos: Proceso que consiste en inspeccionar, limpiar y transformar los datos para resaltar información útil.

Histórico: Hace referencia al estado de una empresa durante el tiempo que lleva activa.

Integrado: Hace referencia al proceso de unificar datos de diferentes fuentes en un solo conjunto de datos.

Tabla: Contiene datos determinados dependiendo del tipo del modelo de datos.

Clave Principal: Es un campo que identifica como única cada fila de una tabla.

Clave Foránea: Es un campo que hace referencia a la llave principal de otra tabla.

Indicador: Es la sumariación de las medidas de una tabla de hechos.

Atributo: Son campos de análisis de una dimensión.

Jerarquía: Es la forma de describir la relación jerárquica entre dos o más miembros de una dimensión.

Campo: Un espacio que contiene cierta capacidad de caracteres para almacenar un dato.

Sistema Rolap: Sistemas construidos sobre bases relacionales que es apto para analizar grandes volúmenes de datos.

Sistema Molap: Sistemas construidos sobre bases de datos multidimensionales que presenta ventajas en la rapidez de almacenamiento de datos, eficiencia de extracción de datos etc.

Sistema Holap: Sistemas híbridos entre ROLAP y MOLAP.

Sistema Dolap: Sistemas orientado a los equipos de escritorio, recoge todos los datos que desea analizar desde una base de datos relacional.

Relación: Describe un vínculo entre dos o más tablas de una base de datos entre otras.

Granularidad: Es el nivel de detalle de con el que se construye un modelo para el análisis de datos.

Servidor: Es un tipo de equipo informático que atiende a las peticiones de un cliente y da respuesta a estas.

Plataforma: Es un sistema que sirve como base para ejecutar determinados módulos de software o hardware.

Código Abierto: Es un modelo de desarrollo de software basado en colaboración abierta.

Bidireccional: Que puede analizarse en base a dos direcciones.

Metodología: Conjunto de métodos que se siguen en una investigación o proyecto.

Negocio: Ocupación en la que se busca generar ingresos.

Sistema Transaccional: Es un tipo de sistema de información encargado de almacenar, modificar, eliminar, etc., todo tipo de información.

Solución Informática: Proceso de desarrollo o aplicación de un recurso para solucionar problemas de tipo informático.

Modelo Entidad-Relación: Es un diagrama usado para representar las entidades relevantes de un modelo de datos y su relación.

Diccionario de datos: Un repositorio encargado de almacenar diferentes tipos de datos.

Microsoft Excel: Herramienta de hoja de cálculo utilizada para el análisis de datos.

Origen de datos o Fuente de Datos: Es la ubicación de donde se extraen o se crean los datos.

Archivo: Es un conjunto de bits que se almacenan y contienen datos.

ANEXO 1

Se anexa el script utilizado para crear la Bodega de Datos.

```

Script Data Warehouse
create database DWH_Ventas
go
create table Region (
  IdRegion Int,
  NombreRegion varchar (50) Not Null,
  Primary key (IdRegion)
)
create table Ciudad (
  IdCiudad Int,
  NombreCiudad varchar (50) Not Null,
  IdRegion Int,
  Primary key (IdCiudad),
  Foreign key (IdRegion) References Region (IdRegion)
)
create table Cliente (
  IdCliente Bigint,
  NombreCliente varchar (200) Not Null,
  IdCiudad Int,
  Primary key (IdCliente),
  Foreign key (IdCiudad) References Ciudad (IdCiudad)
)
create table Vendedor (
  NombreVendedor varchar (100) Not Null,
  IdCiudad Int,
  Primary Key (IdVendedor),
  Foreign key (IdCiudad) References Ciudad (IdCiudad)
)
create table Producto (
  IdProducto Int,
  Abreviatura varchar (50) Not Null,
  NombreProducto varchar (100) Not Null,
  Primary key (IdProducto)
)
create table Tiempo (
  IdTiempo Int,
  Fecha Date Not Null,
  Dia Int Not Null,
  Mes Varchar (50) Not Null,
  Trimestre Int Not Null,
  Anio Int Not Null,
  Primary Key (IdTiempo)
)
create table Estado (
  IdEstado Int,
  Descripcion Varchar (50) Not Null,
  Primary Key (IdEstado)
)
create table Ventas (
  IdVentas Int IDENTITY (1,1),
  Documento Varchar (50) Not Null,
  IdRegion Int,
  IdTiempo Int,
  IdCliente Bigint,
  IdVendedor Int,
  IdProducto Int,
  Cantidad Int Not Null,
  PrecioPromedio Money Not Null,
  PrecioTotal Money Not Null,
  IdEstado Int Not Null,
  Primary Key (IdVentas),
  Foreign Key (IdRegion) References Region (IdRegion),
  Foreign Key (IdTiempo) References Tiempo (IdTiempo),
  Foreign Key (IdCliente) References Cliente (IdCliente),
  Foreign Key (IdVendedor) References Vendedor (IdVendedor),
  Foreign Key (IdProducto) References Producto (IdProducto),
  Foreign Key (IdEstado) References Estado (IdEstado)
)

```

ANEXO 2

Se anexan las Actas de Reunión con la empresa “Grupo Natural Ser”.



UDEC
UNIVERSIDAD DE
CUNDINAMARCA

Página 1 de 2

ACTA No. 1 TRABAJO DE GRADO 21 DE SEPTIEMBRE DE 2017

CLASE SE REUNIÓN: Reunión ordinaria con la empresa sobre el proyecto
ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO
MULTIDIMENSIONAL PARA LA TOMA DE DECISIONES DE LA
EMPRESA “GRUPO NATURAL SER..

CLIENTE: Grupo Natural Ser

CIUDAD Y FECHA: Fusagasugá, 2017 – 9 - 21

HORA: 5:00 p.m.

LUGAR: Carrera 5ta Calle 3ra

ASISTENTES DE LA UNIVERSIDAD:

Estudiante Liliana Catalina Pérez Fonseca
Estudiante Pregrado
Estudiante Nelson David Romero Herrera
Estudiante Pregrado
Estudiante Cristian Elías Mayorga Chaves
Estudiante Pregrado

ASISTENTES DE LA EMPRESA:

Juan Carlos Ramírez
Área de Sistemas

OBJETIVOS DE LA REUNION:

- Analizar software contable de la empresa
- Comprender la información que manejan

RESUMEN TEMAS TRATADOS

En la primera reunión con la empresa, nos comentaron como manejan el área de ventas de esta, el software que utilizan, las personas que actúan sobre esta área, entre otras cosas.

En cuanto al software que manejan, es World Office edición Enterprise. Este software maneja toda la información de la empresa, desde ventas hasta

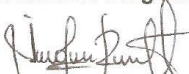


UDEEC
UNIVERSIDAD DE
CUNDINAMARCA

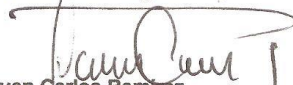
Página 2 de 2

producción, nomina etc. Es un software licenciado por lo cual no tenemos acceso a la base de datos, solo podemos acceder a la información que el sistema asume como informes. Esto hace que la empresa no pueda analizar la información con el enfoque deseado de minería de datos, lo cual abre las puertas para plantear la solución informática.


Liliana Catalina Pérez Fonseca
Estudiante Pregrado


Nelson David Romero Herrera
Estudiante Pregrado


Cristian Elías Mayorga Chaves
Estudiante Pregrado


Juan Carlos Ramírez
Área de Sistemas
86013792



UDEEC
UNIVERSIDAD DE
CUNDINAMARCA

Página 1 de 2

ACTA No. 2 TRABAJO DE GRADO 4 DE OCTUBRE DE 2017

CLASE SE REUNIÓN: Reunión ordinaria con la empresa sobre el proyecto ANALISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO MULTIDIMENSIONAL PARA LA TOMA DE DECISIONES DE LA EMPRESA "GRUPO NATURAL SER..

CLIENTE: Grupo Natural Ser

CIUDAD Y FECHA: Fusagasugá, 2017 – 10 - 4

HORA: 5:00 p.m.

LUGAR: Carrera 5ta Calle 3ra

ASISTENTES DE LA UNIVERSIDAD:

Estudiante Liliana Catalina Pérez Fonseca
Estudiante Pregrado
Estudiante Nelson David Romero Herrera
Estudiante Pregrado
Estudiante Cristian Elías Mayorga Chaves
Estudiante Pregrado

ASISTENTES DE LA EMPRESA:

Juan Carlos Ramírez
Área de Sistemas

OBJETIVOS DE LA REUNION:

- Entrevista con la empresa
- Comprensión del negocio de la empresa
- Analizar información que desean visualizar.

RESUMEN TEMAS TRATADOS

Hablamos con la gerente comercial de la empresa, quien nos comentó como se desenvuelve la empresa en el mercado, dado respuesta a preguntas como: como decidir donde abrir una nueva zona de venta, como manejan las ventas, la misión y visión se la empresa, como se enteran de la competencia

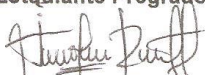


UDECA
UNIVERSIDAD DE
CUNDINAMARCA

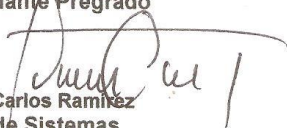
Página 2 de 2

etc. De esta forma comprendimos mejor el negocio y la información que les es útil a la hora de tomar decisiones.


Liliana Catalina Pérez Fonseca
Estudiante Pregrado


Nelson David Romero Herrera
Estudiante Pregrado


Cristian Elías Mayorga Chaves
Estudiante Pregrado


Juan Carlos Ramírez
Área de Sistemas
8603772.



UDECA
UNIVERSIDAD DE
CUNDINAMARCA

Página 1 de 2

ACTA No. 3 TRABAJO DE GRADO 19 DE OCTUBRE DE 2017

CLASE SE REUNIÓN: Reunión ordinaria con la empresa sobre el proyecto
ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO
MULTIDIMENSIONAL PARA LA TOMA DE DECISIONES DE LA
EMPRESA "GRUPO NATURAL SER..

CLIENTE: Grupo Natural Ser

CIUDAD Y FECHA: Fusagasugá, 2017 – 10 - 19

HORA: 5:00 p.m.

LUGAR: Carrera 5ta Calle 3ra

ASISTENTES DE LA UNIVERSIDAD:

Estudiante Liliana Catalina Pérez Fonseca
Estudiante Pregrado
Estudiante Nelson David Romero Herrera
Estudiante Pregrado
Estudiante Cristian Elías Mayorga Chaves
Estudiante Pregrado

ASISTENTES DE LA EMPRESA:

Juan Carlos Ramírez
Área de Sistemas

OBJETIVOS DE LA REUNION:

- Solicitar información para trabajar en el modelo

RESUMEN TEMAS TRATADOS

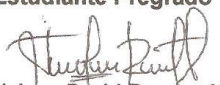
Se solicitó la información que el software de la empresa nos dejó visualizar, de este modo se nos otorgó la información de ventas de 2 ciudades: Bogotá y Medellín. Nos dieron 6 tipos de informes de cada ciudad, para un total de 12 con información desde unidades vendidas hasta monto vendido. Cada uno de estos agrupado por partes específicas.



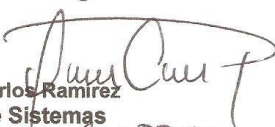
UDECA
UNIVERSIDAD DE
CUNDINAMARCA

Página 2 de 2


Liliana Catalina Pérez Fonseca
Estudiante Pregrado


Nelson David Romero Herrera
Estudiante Pregrado


Cristian Elías Mayorga Chaves
Estudiante Pregrado


Juan Carlos Ramírez
Área de Sistemas
86013792



UDEEC
UNIVERSIDAD DE
CUNDINAMARCA

Página 1 de 2

ACTA No. 5 TRABAJO DE GRADO 24 DE ABRIL DE 2018

CLASE SE REUNIÓN: Capacitación con la empresa sobre el proyecto
ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO
MULTIDIMENSIONAL PARA LA TOMA DE DECISIONES DE LA
EMPRESA "GRUPO NATURAL SER".

CLIENTE: Grupo Natural Ser

CIUDAD Y FECHA: Fusagasugá, 2017 – 04 - 24

HORA: 11:00 a.m.

LUGAR: Carrera 5ta Calle 3ra

ASISTENTES DE LA UNIVERSIDAD:

Estudiante Liliana Catalina Pérez Fonseca
Estudiante Pregrado
Estudiante Cristian Elías Mayorga Chaves
Estudiante Pregrado

ASISTENTES DE LA EMPRESA:

Yhon Jaime García Franco
Gerente

Jennifer Maritza Reyes Rojas
Gerente

Juan Carlos Ramírez
Área de Sistemas

OBJETIVOS DE LA REUNION:

- Capacitar a la empresa sobre la solución informática planteada en el proyecto

TEMAS TRATADOS

- Hablar sobre la solución y sus beneficios
- Explicación sobre el proceso de ETL de los datos.

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca
Teléfonos (091)8672144-8673273-8732512/30 Telefax: 8732554 - 8677898 – 8673826
Línea Gratuita 018000976000
www.unicundi.edu.co E-mail: unicundi@mail.unicundi.edu.co
NIT: 890.680.062-2



UDEC
UNIVERSIDAD DE
CUNDINAMARCA

Página 2 de 2

- Explicación de la funcionalidad del aplicativo auxiliar.
- Muestra de dashboard de ventas en Excel y Power BI

Yhon Jaime Garcia Franco
Gerente

Jennifer Maritza Reyes Rojas
Gerente

Juan Carlos Ramirez
Área de Sistemas

Liliana Catalina Pérez Fonseca
Estudiante Pregrado

Cristian Elías Mayorga Chaves
Estudiante Pregrado



**ACTA DE ENTREGA A SATISFACCION DE LA SOLUCIÓN INFORMÁTICA
TRABAJO DE GRADO 24 DE ABRIL DE 2018**

CLASE SE REUNIÓN: Entrega a satisfacción de la solución informática
ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN MODELO
MULTIDIMENSIONAL PARA LA TOMA DE DECISIONES DE LA
EMPRESA "GRUPO NATURAL SER".

CLIENTE: Grupo Natural Ser

CIUDAD Y FECHA: Fusagasugá, 2017 – 04 - 24

HORA: 11:00 a.m.

LUGAR: Carrera 5ta Calle 3ra

ASISTENTES DE LA UNIVERSIDAD:

Estudiante Liliana Catalina Pérez Fonseca
Estudiante Pregrado
Estudiante Cristian Elías Mayorga Chaves
Estudiante Pregrado

ASISTENTES DE LA EMPRESA:

Yhon Jaime García Franco
Gerente

Jennifer Maritza Reyes Rojas
Gerente

Juan Carlos Ramírez
Área de Sistemas

OBJETIVOS DE LA REUNION:

- Capacitar a la empresa sobre la solución informática planteada en el proyecto

TEMAS TRATADOS

- Entrega de la solución informática



UDEC
UNIVERSIDAD DE
CUNDINAMARCA

Yhon Jaime García-Franco
Gerente

Jennifer Maritza Reyes Rojas
Gerente

Juan Carlos Ramírez
Área de Sistemas

Liliana Catalina Pérez Fonseca
Estudiante Pregrado

Cristian Elias Mayorga Chaves
Estudiante Pregrado

ANEXO 3

Se anexa en la carpeta del archivo los manuales de usuario y técnico (Cartillas Digitales). De igual forma anexamos imágenes, el aplicativo, el artículo científico y las cartas correspondientes a la gestión del proyecto.

CONCLUSIONES

Implementar una solución BI en una empresa no solo se limita al componente técnico de la misma, sino al análisis profundo de cada aspecto que hará parte de la misma. Además, brindar el soporte necesario para que una empresa pueda no solo visualizar información, sino que además logre construir el conocimiento del negocio (en nuestro Caso de Uso en el área comercial de la compañía “Grupo Natural Ser”), un conocimiento que permite determinar los próximos pasos de la compañía, lo que significa aumentar su competitividad basada en una toma de decisiones estratégica y con oportunidad.

Muchas empresas desconocen la importancia de tener herramientas que les permitan hacer análisis de sus datos. Esta solución no solo brinda la opción de analizar datos sino que además brinda una base tecnológica a la empresa para un despliegue a futuro dedicando un área específica para BI, siendo esta de gran importancia como hemos resaltado a lo largo de este trabajo.

Se pudo observar la importancia de los datos para una empresa ya que estos presentan el activo de mayor valor al interior de una organización. En este proyecto se hizo un proceso extenso de ETL para poder convertir los datos en información, y por medio de los Dashboard poder visualizar esta información y convertirla en conocimiento valioso. Es por ello, que en el momento de la capacitación la solución informática tuvo una gran acogida, en especial para la Gerente Comercial y el encargado del área de Sistemas en la empresa “Grupo Natural Ser”, por ser una herramienta nueva representada en una solución tecnológica a empoderar de manera corporativa para potenciar el análisis de los datos, en este caso del área comercial, y así poder tomar decisiones posteriores para la distribución de sus productos a partir de información histórica.

Satisfactoriamente podemos concluir que el valor agregado de la implementación de esta solución tecnológica basada en BI y en Minería de Datos, está en el “conocimiento real” que logramos entregar a la organización Grupo Natural Ser a partir de un Modelo Multidimensional conceptualizado a la medida de su realidad, para generar una cultura organizativa de medición y evaluación sostenible en el tiempo, que además represente a futuro una decisión estratégica replicable a toda la compañía.

El desarrollo de esta solución dejó un aprendizaje amplio puesto que se aplicó la solución en un entorno real, en donde el mercado está cambiando al igual que las necesidades de las empresas. Al dar respuesta tecnológica a problemas actuales de las empresas, se da la oportunidad de aplicar los conocimientos adquiridos y ponerlos en práctica, como se realizó en este proyecto.

Una de las bases principales para el enfoque de este proyecto es el área de Data Management, la cual se conoció gracias a la Línea de Profundización de minería de datos con la que cuenta el programa. Esta Línea de Profundización permitió resolver, aplicar e implementar esta solución en un ambiente real donde el tema de Data Management, Inteligencia de Negocios entre otros, está tomando auge. De igual forma demuestra que el programa da respuesta a necesidades actuales no solo en el Desarrollo de Software sino también en estas áreas que son tan importantes para el desarrollo de las empresas.

En aspectos sociales, se generó un aporte a la compañía y de igual forma una solución aplicable en el sector productivo, es decir que la solución se puede replicar a cualquier tipo de organización. De esta forma las soluciones tecnológicas de este estilo dan soporte a necesidades de analítica, toma de decisiones buscando mejorar la visión y la planeación estratégica de las

empresas, las cuales son un aporte muy importante de la Ingeniería de Sistemas a la sociedad entre otros.

Bibliografía

- Bernabeu_Dario. (06 de 05 de 2009). *Datawarehouse manager*. Obtenido de DATAPRIX:
<http://www.dataprix.com/datawarehouse-manager>
- CEDEÑO, A. (2006). MODELO MULTIDIMENSIONAL. *Industrial*, Vol. XXVII.
- IBM. (2012). *Manual CRISP-DM de IBM SPSS*. Obtenido de
<http://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- IBM Knowledge Center. (2012). *CRISP-DM Help Overview*. Obtenido de IBM:
https://www.ibm.com/support/knowledgecenter/es/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm
- LOGICALIS. (24 de 07 de 2015). *Estructuras multidimensionales de los cubos OLAP*. Obtenido de LOGICALIS Business and technology working as one:
<https://blog.es.logicalis.com/analytics/cubos-olap-y-estructuras-multidimensionales-todo-lo-que-hay-que-saber>
- Microsoft Power BI. (s.f.). *Inteligencia empresarial sin precedentes*. Obtenido de Power BI:
<https://powerbi.microsoft.com/es-es/>
- Moro, S., & Laureano, R. (s.f.). USING DATA MINING FOR BANK DIRECT MARKETING:.
5.
- Php. (2018). *¿Qué es PHP?* Obtenido de <http://php.net/manual/es/intro-what-is.php>
- PowerData Especialista en Gestión de Datos. (13 de 06 de 2013). *Procesos ETL: Extracción, Transformación, Carga*. Obtenido de El valor de la gestión de datos:
<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/288859/procesos-etl-extracci-n-transformaci-n-carga>

Red Gráfica Latinoamérica. (s.f.). *El lenguaje de programación PHP*. Obtenido de Red Grafica

Latinoamerica: <http://redgrafica.com/El-lenguaje-de-programacion-PHP>

S.L., Sinergia e Inteligencia de Negocio. (2016). *¿Que es Business Intelligence?* Obtenido de

Sinnexus: http://www.sinnexus.com/business_intelligence/

S.L., Sinergia e Inteligencia de Negocio. (2016). *Bases de Datos OLTP y OLAP*. Obtenido de

Sinnexus: http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx

S.L., Sinergia e Inteligencia de Negocio. (2016). *Datawarehouse*. Obtenido de Sinnexus:

http://www.sinnexus.com/business_intelligence/datawarehouse.aspx

SALCEDO PARRA, O., GALEANO, R. M., & RODRIGUEZ B., L. (Enero - Junio de 2010).

Metodología crisp para la implementación. *Revista Tecnura*, paginas 35 - 48.

SENA. (s.f.). *SISTEMAS DE INTELIGENCIA DE NEGOCIOS*. Obtenido de

https://senaintro.blackboard.com/bbcswebdav/pid-47553315-dt-content-rid-14706754_4/institution/217213_tvirtual/OAAPs/OAAP2/aa4/oa1aa4/utilidades/oc.pdf

Sinergia e Inteligencia de Negocio S.L. (2016). *Persistencia MOLAP, ROLAP, HOLAP*.

Obtenido de Sinnexus:

http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx

Soporte Microsoft. (s.f.). *Power Pivot: análisis de datos eficaz y modelado de datos en Excel*.

Obtenido de Microsoft: <https://support.office.com/es-es/article/power-pivot-an%C3%A1lisis-de-datos-eficaz-y-modelado-de-datos-en-excel-a9c2c6e2-cc49-4976-a7d7-40896795d045>

Wikipedia. (s.f.). *Cross Industry Standard Process for Data Mining*. Obtenido de Wikipedia:

https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Wikipedia. (s.f.). *HTML*. Obtenido de Wikipedia: <https://es.wikipedia.org/wiki/HTML>

