



FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

ANÁLISIS DE DATOS PARA LA TOMA DE DECISIONES CON EL USO
DE TABLEROS DE CONTROL APLICADO A LOS REGISTROS DE
VENTAS DIGITALES DE UN E-COMMERCE.

AUTOR

Johny Patricio Ango Yanacallo

AÑO

2020



FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

ANÁLISIS DE DATOS PARA LA TOMA DE DECISIONES CON EL USO DE
TABLEROS DE CONTROL APLICADO A LOS REGISTROS DE VENTAS
DIGITALES DE UN E-COMMERCE.

Trabajo de Titulación presentado en conformidad a los requisitos establecidos
para optar por el título de Ingeniero Electrónico y Redes de Información.

Profesor Guía

Ms. William Eduardo Villegas Chilibingua

Autor

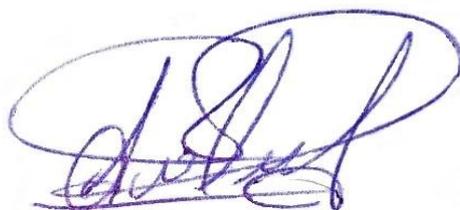
Johny Patricio Ango Yanacallo

Año

2020

DECLARACIÓN DEL PROFESOR GUÍA

“Declaro haber dirigido este trabajo, Análisis de datos para la toma de decisiones con el uso de tablero de control aplicado a los registros de ventas digitales en un e-commerce, a través de reuniones periódicas con el estudiante, Johny Patricio Ango Yanacallo, en el semestre 202010, orientando sus conocimiento y competencias para un eficiente desarrollo de tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación”

A handwritten signature in blue ink, consisting of stylized, overlapping loops and lines, positioned above a horizontal line.

William Eduardo Villegas Chiliqinga

Magister en Redes de Comunicaciones

C.I: 1715338263

DECLARACIÓN DEL PROFESOR CORRECTOR

“Declaro haber revisado este trabajo, Análisis de datos para la toma de decisiones con el uso de tablero de control aplicado a los registros de ventas digitales en un e-commerce, del estudiante, Johny Patricio Ango Yanacallo, en el semestre 202010, dando cumplimiento a todas las disposiciones vigentes que regulen los Trabajos de Titulación”



Iván Patricio Ortiz Garcés

Magister en Redes de Comunicaciones

C.I: 0602356776

DECLARACIÓN DE AUTORÍA DEL ESTUDIANTE.

“Yo, Johny Patricio Ango Yanacallo, declaro que el presente trabajo de titulación es de mi autoría, todas las fuentes utilizadas para el desarrollo han sido debidamente citadas y se ha respetado la normativa que protege a los autores”



Johny Patricio Ango Yanacallo

C.I: 1723781074

AGRADECIMIENTOS

Primeramente, agradezco a Dios por permitirme tener y compartir con mi familia; gracias a mi familia por su apoyo incondicional en cada decisión y proyecto. Por permitirme cumplir con excelencia el desarrollo de mi carrera, gracias por creer en mí y a Dios por permitirme vivir y disfrutar de cada día. Además, agradezco a mis profesores quienes aparte de transmitir su conocimiento y sabiduría, también me han sabido inculcar valores para convertirme en excelente profesional.

DEDICATORIA

Este proyecto de titulación va dedicado principalmente a los pilares de mi vida, mis padres, quienes han estado en todo momento a mi lado y me han apoyado en cada obstáculo de la vida. La finalización de mi carrera y del presente trabajo de titulación es gracias al esfuerzo, tiempo y dedicación que han puesto en mi en todo este tiempo.

RESUMEN

El presente proyecto de titulación se basa en la implementación de una arquitectura de inteligencia empresarial (BI) con herramientas Microsoft. Con la intención de analizar datos de una operación de ventas digitales asistidas de un retail. Esto con el objetivo de identificar comportamientos de compras, clientes fraudulentos y mejorar el proceso de control de fraude antes de una preventa. Estos datos se encuentran alojados en una base de datos transaccional y otros orígenes como archivos planos. Del mismo modo, se busca aportar a la toma de decisiones gerenciales y generar conocimiento para correcciones de problemas detectados. Todo esto gracias a la implementación de tableros gerenciales que serán desarrollados sobre Power BI como herramienta de visualización de datos.

Al tener un sistema de gestión de pedidos independiente del sistema comercial. Esta organización de comercio busca centralizar la información de los clientes potenciales. Con la intención de mantener una mejor visibilidad de la vida de los pedidos y encontrar patrones o comportamientos. De tal forma podrán tomar decisiones estratégicas con menos incertidumbre y que aporten valor agregado al negocio basándose en algo real.

Palabras Clave: BI, retail, fraude, patrones.

ABSTRACT

This degree project is based on the implementation of a business intelligence (BI) architecture with Microsoft tools. With the intent of data analysis of a digital assisted sales operation of a retailer. This with the objective of identifying shopping behaviors, fraudulent customers and improving the fraud control process before prevention. This data is housed in a transactional database and other sources as flat files. Similarly, it seeks to contribute to management decision making and generate knowledge for corrections of detected problems. All this thanks to the implementation of management boards that will be developed on Power BI as a data visualization tool.

By having an order management system independent of the commercial system. This trade organization seeks to centralize the information of potential customers. With the intention of maintaining a better visibility of the life of the orders and finding patterns or behaviors. In this way you can make strategic decisions with less uncertainty and that add value to the strategic business in something real.

Keywords: BI, retail, fraud, patterns.

ÍNDICE

1. CAPÍTULO I. INTRODUCCIÓN	1
1.1 Antecedentes	1
1.2 Justificación.....	3
1.3 Objetivos	3
1.3.1 Objetivo general.....	3
1.3.2 Objetivos específicos	3
2. CAPÍTULO II. MARCO TEÓRICO	4
2.1 Datos, Información y Conocimiento.....	4
2.1.1 Datos	5
2.1.2 Información	5
2.1.3 Conocimiento	6
2.2 Fuentes de datos	6
2.2.1 Calidad de datos	7
2.3 Almacén de datos	8
2.3.1 Características de un Almacén de datos.....	8
2.3.2 Ventajas de un almacén de datos	9
2.3.3 Diferencias entre un almacén de datos y una base de datos	9

2.4 Datamart.....	10
2.5 Cubos de Datos	11
2.5.1 Procesamiento de transacciones en línea	11
2.5.2 Procesamiento analítico en línea	12
2.6 Extracción, Transformación y Carga	12
2.6.1 Flujo de datos y Flujo de control	13
2.7 Minería de Datos	13
2.7.1 Arquitectura de Minería de datos	14
2.7.2 Algoritmos de minería de datos.....	15
2.8 Visualización de datos	19
2.8.1 KPIS	19
2.8.2 Informes y visualización de datos	20
3. CAPÍTULO III. ANÁLISIS Y CONFIGURACIÓN DE HERRAMIENTAS	21
3.1 Comparación de Herramientas BI	22
3.1.1 Servicios de Integración SQL Server	24
3.1.2 Servicios de Análisis SQL Server.....	25
3.1.3 Servicios de Reportería SQL Server	25
3.1.4 Diferencias entre Power BI y SSRS	26

3.1.5	Importancia de Power BI Microsoft	26
3.2	Instalación de Herramientas	27
3.2.1	Instalación SQL Server	27
3.2.2	Instalación Visual Studio y configuración paquetes SISS.....	30
3.3	Instalación ODBC MySQL Conector	33
4.	CAPÍTULO IV. DISEÑO DE PROTOTIPO DE INTELIGENCIA DE NEGOCIO.....	34
4.1	Diseño del Data Warehouse.....	34
4.1.1	Análisis y diseño del modelo dimensional	34
4.1.2	Tablas de Dimensiones.....	37
4.1.3	Tablas de Hechos	38
4.1.4	Definición de Procesos ETL.....	39
4.1.5	Diseño de cubos OLAP	41
4.2	Diseño de tablero de indicadores.....	43
5.	CAPÍTULO V. IMPLEMENTACIÓN Y ANÁLISIS DEL PROTOTIPO INTELIGENCIA DE NEGOCIO.	44
5.1	Infraestructura	44
5.2	Implementación del Data Warehouse	45
5.2.1	Tablas de dimensiones	46

5.2.2	Tabla de hechos	48
5.3	Conexión a fuente de datos	51
5.4	Desarrollo de procesos ETL	52
5.4.1	ODBC Origen.....	54
5.4.2	Conversión de Datos	56
5.4.3	Columna Derivada	57
5.4.4	Componente de Script	58
5.4.5	OLE DB Destino.....	59
5.5	Desarrollo cubo OLAP	63
5.6	Minería de datos	70
5.7	Desarrollo de Tablero de indicadores	72
5.7.1	Conexiones Orígenes de Datos	72
5.7.2	Reporte de ventas con fraude	74
5.7.3	Reporte intentos de compra con fraude	75
5.8	Análisis de resultados del caso desarrollado	76
5.8.1	Reportes	76
5.8.2	Minería de datos	80
6.	CONCLUSIONES Y RECOMENDACIONES.....	82

6.1 Conclusiones.....	82
6.2 Recomendaciones	83
REFERENCIAS	85
ANEXOS	89

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Pirámide Datos, Información y Conocimiento.	5
<i>Figura 2.</i> Esquema Flujo de control y Flujo de Datos.	13
<i>Figura 3.</i> Esquema Minería de datos.	14
<i>Figura 4.</i> Arquitectura minería de datos.	15
<i>Figura 5.</i> Técnicas de minería de datos.	16
<i>Figura 6.</i> Esquema de un árbol de decisión.	17
<i>Figura 7.</i> Algoritmo de agrupación.	18
<i>Figura 8.</i> Conjunto de reglas de un elemento.	19
<i>Figura 9.</i> Representación de Indicadores.	20
<i>Figura 10.</i> Tablero Power BI.	21
<i>Figura 11.</i> Cuadrante Mágico de GARTNER.	23
<i>Figura 12.</i> SQL Server Integration Services.	24
<i>Figura 13.</i> Informe Paginado vs Informe Móvil.	25
<i>Figura 14.</i> Esquema Power BI.	27
<i>Figura 15.</i> Centro de instalación SQL Server.	28
<i>Figura 16.</i> Selección de características SQL server.	28
<i>Figura 17.</i> Configuración del motor de base de datos.	29
<i>Figura 18.</i> Servicios Instalados.	30

<i>Figura 19.</i> Selección de cargas de trabajo.	30
<i>Figura 20.</i> Language Packs.	31
<i>Figura 21.</i> Instalación Extensión SSIS.	32
<i>Figura 22.</i> Comprobar Extensión SSIS.....	32
<i>Figura 23.</i> Conector ODBC MySQL.	33
<i>Figura 24.</i> Conexión ODBC MySQL.....	34
<i>Figura 25.</i> Modela Estrella.	35
<i>Figura 26.</i> Tablas de Dimensiones.....	38
<i>Figura 27.</i> Tablas de Hechos.	39
<i>Figura 28.</i> Diseño Extracción, Transformación y Carga.	39
<i>Figura 29.</i> Diseño Básico ETL.....	40
<i>Figura 30.</i> Estructura de un Cubo OLAP.	41
<i>Figura 31.</i> Cubo OLAP.....	42
<i>Figura 32.</i> Arquitectura del Prototipo.	44
<i>Figura 33.</i> Modelo Estrella.	45
<i>Figura 34.</i> Ventana de Consultas SSMS.....	49
<i>Figura 35.</i> Explorador de objetos.	49
<i>Figura 36.</i> Creación Diagrama Base Dimensional.....	50
<i>Figura 37.</i> Diagrama Base Dimensional.....	50

<i>Figura 38.</i> Parámetros ODBC.	51
<i>Figura 39.</i> Grupo de Seguridad AWS.	52
<i>Figura 40.</i> ETL en SSIS.	53
<i>Figura 41.</i> Editor de Origen ODBC.	54
<i>Figura 42.</i> Error Consulta SQL.	55
<i>Figura 43.</i> Consulta SQL Correcta.	55
<i>Figura 44.</i> Columnas Editor ODBC.	56
<i>Figura 45.</i> Editor de transformación conversión de datos.	57
<i>Figura 46.</i> Creación columna derivada de tiempo.	57
<i>Figura 47.</i> Editor de Transformación Columna de salida.	58
<i>Figura 48.</i> Script Id incremental.	58
<i>Figura 49.</i> Editor de destino OLE DB.	59
<i>Figura 50.</i> Asignación de columnas Editor destino OLE DB.	60
<i>Figura 51.</i> Ejecución paquete ETL.	60
<i>Figura 52.</i> Ejecución ETL Dimensiones.	61
<i>Figura 53.</i> Ejecución ETL Hecho Venta.	62
<i>Figura 54.</i> Ejecución ETL Hecho Pedido.	62
<i>Figura 55.</i> Creación Proyecto Analysis Services.	63
<i>Figura 56.</i> Explorar de Soluciones.	63

<i>Figura 57.</i> Configuración Fuente de Datos.....	64
<i>Figura 58.</i> Explorador de soluciones Data Source.	64
<i>Figura 59.</i> Vista de fuente de datos.....	65
<i>Figura 60.</i> Esquema relacional Vista de fuente de datos.....	65
<i>Figura 61.</i> Selección Método Creación Cubo.	66
<i>Figura 62.</i> Tablas de grupo de medidas.....	66
<i>Figura 63.</i> Resumen creación Cubo.....	67
<i>Figura 64.</i> Cubo Canal Online.....	67
<i>Figura 65.</i> Atributos Dimensión Callcenter.	68
<i>Figura 66.</i> Ejecución del Cubo.	68
<i>Figura 67.</i> Cubo Procesado.	69
<i>Figura 68.</i> Interfaz de administración Cubo.....	69
<i>Figura 69.</i> Método Estructura de minería de datos.....	70
<i>Figura 70.</i> Algoritmos de minería de datos.....	71
<i>Figura 71.</i> Columnas usadas en el modelo de minería de datos.	71
<i>Figura 72.</i> Visor de patrones del algoritmo de minería de datos.....	72
<i>Figura 73.</i> Tipo de orígenes de datos en Power BI.	73
<i>Figura 74.</i> Conexión Analysis Services.	73
<i>Figura 75.</i> Tablero Indicadores Ventas con Fraude.....	74

<i>Figura 76.</i> Tablero Indicadores Pedidos con Fraude.....	75
<i>Figura 77.</i> Indicador de Anillo Fraude por día.....	76
<i>Figura 78.</i> Indicador de Barras Fraude por Provincia.	77
<i>Figura 79.</i> Indicadores de fraude por producto.....	78
<i>Figura 80.</i> <i>Grafica de barras cantidad de pedidos por cliente</i>	78
<i>Figura 81.</i> Gráfico de Anillo cantidad de pedidos por campaña.....	79
<i>Figura 82.</i> Gráfico de pastel estados por pedidos.	79
<i>Figura 83.</i> Visualización algoritmo de agrupación.	80
<i>Figura 84.</i> Probabilidad de Fraude.....	81

ÍNDICE DE TABLAS

Tabla 1. <i>Diferencias entre DWH y Data Base</i>	19
Tabla 2. <i>Diferencias entre DWH y Datamart</i>	21
Tabla 3. <i>Forma no Normalizada</i>	46

1. Capítulo I. Introducción

1.1 Antecedentes

En la actualidad el término Inteligencia empresarial (BI) está en el apogeo de las tecnologías de la información en el contexto del análisis de datos. Este aporta valor agregado a los datos de las empresas que lo utilizan transformando la información en inteligencia procesable. Haciendo que cada vez la toma de decisiones estratégicas y tácticas dentro del giro de negocio tenga menos incertidumbre, mayor rapidez en base a información precisa. En contraste con lo mencionado, todo esto tiene una reseña histórica donde una serie de hitos vienen formando los cimientos de esta tecnología.

Desde la antigüedad el ser humano tiene la necesidad de recopilar información de sucesos que pasan en el día a día. En consecuencia, el desarrollo de las bases de datos surge ante la necesidad de almacenar grandes volúmenes de datos. Sobre todo, entre la década de los 60 desde el lanzamiento de las primeras computadoras que dan impulso al inicio de las primeras generaciones de bases de datos de red y jerárquicas.

Continuando con la época de los 70 donde Edgar Frank Codd (Científico informático) da aportaciones importantes a las bases de datos relacionales. Este hecho colabora al nacimiento de la segunda generación de bases de datos (DMS). Basándose en este trabajo se desarrolla el RSS o lo que actualmente se conoce como Corporación Oracle.

Este desarrolla un sistema de gestión con el mismo nombre. A su vez, se implementa el lenguaje SQL. Un lenguaje donde el usuario puede hacer consultas de los datos almacenados y alterar directamente toda la estructura de la base de datos. Estas características hacen que para la siguiente década se posicione como estándar en este tipo de industria.

De la mano con esto, se desarrollaron aplicaciones empresariales, estas tenían la principal característica de permitir la entrada de datos en los sistemas de

información. Ante esto, el flujo de entrada de datos aumentó exponencialmente en consecuencia el acceso a la misma fue difícil y lento. Es así como comenzó la proliferación de los datos y la necesidad de almacenar y acceder a los mismos fue evidente.

Por consiguiente, entra al juego el padre de los data WareHouse Bill Inmon (Científico informático estadounidense), quien publica el libro *Building the Data Warehouse* en 1992. Este se convierte en base fundamental de los almacenes de datos. Se establece un enfoque descendente. Donde en primera instancia se define un repositorio centralizado, y posteriormente se crean depósitos de datos que a su vez estos contienen subconjuntos de datos.

Por otro lado, Ralph Kimball publica *The Data Warehouse Toolkit* el mismo contiene un enfoque ascendente. Donde se propone el uso de depósitos de datos individuales para la creación de un almacén de datos (DW). Con estas bases se impulsa la era de la Inteligencia Empresarial 1.0 en este tiempo aparecen las primeras aplicaciones BI. La desventaja de estas primeras aplicaciones es que fueron incapaces de analizar grandes volúmenes de registros en un tiempo prudente. Además de esto, contaban con un número limitado de orígenes de datos y el precio era inaccesible para las pymes.

Otro rasgo es la evolución a Inteligencia Empresarial 2.0, en la que los usuarios finales tienen la capacidad de generar y compartir contenido. Con el uso de herramientas de análisis y visualización se generan indicadores que permiten tomar decisiones gerenciales. En el presente existen diversas formas de diseñar soluciones BI. Desde las soluciones tradicionales que usan muchas empresas hasta soluciones Cloud BI o Big Data que permiten tener datos integrados totalmente en la nube.

Cloud Computing e Inteligencia Empresarial son una pareja ideal; se trata de integrar la información correcta con las personas adecuadas en el momento adecuado ofreciendo un forma ligera, ágil, rápida y económica de acceder a las aplicaciones de BI. (Martín, 2015).

1.2 Justificación

Según el último estudio de la CECE al 2017, la penetración de internet en la población ecuatoriana se establece al 81%. En consecuencia, el comportamiento de compra digital crece exponencialmente, y las transacciones comerciales aumentan un 80% entre el 2012 y 2014. Esto orienta a las empresas a llevar un modelo de e-commerce y ventas digitales más estable. Donde la prioridad es dar una buena experiencia de compra a los clientes.

Por todo lo mencionado anteriormente, este retail tiene la necesidad de incluir un modelo BI. Donde el principal objetivo es tener una gestión adecuada del conocimiento que brindan los datos brutos de ventas. Con esto se busca desarrollar y monitorear indicadores (KPI) específicos que aporten valor a las decisiones del negocio.

Con la ayuda de esto se busca aumentar las ventas, ya que se pueden implementar estrategias comerciales en base a comportamientos de compra. Así mismo, se pueden optimizar las campañas de adquisición de clientes potenciales. Optimizar el proceso E2E, gracias a que se puede mantener un indicador concreto para esta acción en cada parte del flujo. Finalmente, modelar un control antifraude debido a que se podrá analizar el comportamiento de este tipo de clientes.

1.3 Objetivos

1.3.1 Objetivo general

- Desarrollar una arquitectura de análisis de datos basada en un modelo de inteligencia de negocios para el modelamiento de tableros de KPIS para la toma de decisiones en un ambiente de ventas digitales.

1.3.2 Objetivos específicos

- Diseñar de una arquitectura de análisis de datos aplicada a las ventas digitales de un E-commerce.

- Crear de un almacén de datos que garantice su tratamiento a través de algoritmos de minería de datos.
- Determinar los procesos de consulta de información mediante tableros de indicadores de ventas.

2. Capítulo II. Marco Teórico

La implementación de un prototipo de análisis de información empresarial requiere varios elementos que definen cada uno de los procesos que se ejecutan para su funcionamiento.

El concepto general de la inteligencia empresarial (BI). Se define como el conjunto de metodologías, aplicaciones y tecnologías que responden a la necesidad de desarrollar métodos más rápidos, y eficientes. Con el fin de tratar los datos y transformar los mismos en información útil para las organizaciones, enfocada a la toma de decisiones (Josep, 2012).

Algo muy importante en el análisis de datos, es comprender los tres factores por los que pasan los mismos (Datos, información y conocimiento). Por lo que inicialmente a cualquier desarrollo de BI, se deben identificar en qué parte están relacionados estos términos con los componentes de la inteligencia empresarial.

2.1 Datos, Información y Conocimiento

Es importante saber diferenciar entre estos términos, ya que al tratarse de algo similar lleva a la mala interpretación de estos. Por lo que vale la pena detallar cada concepto para su mejor comprensión.

La figura 1 muestra una jerarquía de tipo piramidal de los términos mencionados anteriormente. Donde los datos brutos son la materia prima que con el correcto tratamiento se convierten en Información y por ende en conocimiento para las organizaciones (Daverport & Prusak).



Figura 1. Pirámide Datos, Información y Conocimiento.

2.1.1 Datos

Los datos están representados como un conjunto discreto de valores, que por sí solos no dicen nada o aportan al por qué de las cosas. Por ende, no son útiles para la toma de decisiones. Sin embargo, es la materia prima de la información. Estos en su gran parte provienen de fuentes internas o externas de las organizaciones, pueden ser cualitativos o cuantitativos.

2.1.2 Información

La información se define como un grupo de datos que al ser procesados adquieren significado y utilidad para la toma de decisiones. Con esto se toma mejores decisiones disminuyendo incertidumbre y añadiendo valor ayudando a que el receptor de la información logre diferenciar entre grandes cantidades de datos.

A continuación, se plantea una fórmula sencilla que define a la información:

$$I = D + C + U$$

Donde,

I = información

D = Datos

C = Contexto (Añadir valor)

U = Utilidad (Disminuir incertidumbre)

2.1.3 Conocimiento

Finalmente, se define conocimiento a la fusión entre experiencia, valores, información y saber hacer (know-how). Esto sirve como base para la incorporación de nuevas experiencias e información, permite tomar decisiones y emitir juicios que llevan a conclusiones.

Es decir, con el tratamiento correcto en BI la información se transforma en conocimiento, así como los datos en información. Para esto, el tratamiento mencionado debe realizar acciones como: comparación, predicción, búsqueda de conexiones y mezcla con otros conocimientos.

2.2 Fuentes de datos

En una organización existen tres tipos de fuentes de datos: no estructurados, estructurados y semi estructurados. El primero conforma alrededor del 80% de la información de un entorno empresarial (Archivos de audio, video, artículos, correos electrónicos, entre otros). Por otro lado, el segundo son los datos que se encuentran en un formato específico y definido (bases relacionales, hojas de cálculos, entre otros). Finalmente, los semi estructurados no usan un formato fijo, pero contienen etiquetas u otros marcadores (XML, HTML) estos facilitan su comprensión.

Para sacar beneficio de los datos, y que estos generen oportunidades para el negocio, hay que saber que la procedencia de estos se da de cinco orígenes.

- **Datos Transaccionales:** Son los que se generan por una transacción, ejemplo las transacciones bancarias generan este tipo de datos.
- **Biométricos:** Se refieren a la identificación de una persona basada en su anatomía o rasgos personales.
- **Máquina a Máquina:** Esto se refiere a los datos que genera el internet de las cosas, mediante la conexión entre dispositivos, ejemplo un GPS.
- **Generados por los humanos:** Un ejemplo claro son los mails, grabaciones de un centro de atención telefónica o registros médicos.
- **Web y medios sociales:** Se refiere a la información que generan los clics en vínculos, búsquedas en Google, publicaciones en redes sociales o el mismo contenido generado en páginas web.

2.2.1 Calidad de datos

Se define como conjunto de información almacenada en una base de datos (DB) o almacén de datos (DW), esta a su vez deben cumplir los siguientes atributos:

- **Integridad:** Información correcta y completa dentro de una base de datos.
- **Accesibilidad:** Término que define al acceso de información, para que cumpla esto los datos deben ser de fácil acceso para el usuario.
- **Coherencia:** Datos consistentes, debe existir relación entre la data y el giro del negocio a analizar.
- **Disponibilidad:** Se refiere a la disponibilidad de la información cuando esta sea requerida.
- **Totalidad:** Datos completos y necesarios para poder realizar un análisis coherente.
- **Validez:** Datos o valores aceptables dentro de la estructura de la base de datos.
- **Precisión:** Datos verificables, de origen veraz.

Con la finalidad de disminuir la mala calidad de información, es necesario implementar mejores controles de captura de esta o a su vez darle un tratamiento previo a su análisis.

2.3 Almacén de datos

Término conocido en inglés como Data Warehouse, es una base de datos corporativa donde se centraliza la información de varias fuentes para posteriormente procesarla. Permite poder analizarla desde varias perspectivas con altas velocidades de respuesta. Es el primer paso para implementar una solución consistente de inteligencia de negocio.

Al usar estructuras de almacenamiento como: modelo estrella, copo de nieve, cubos relaciones, entre otras persistencias de datos, hacen que la información sea homogénea y fiable. De esta forma, brinda opciones de consultas y tratamiento jerarquizado de los datos fuera del entorno de los sistemas operacionales.

2.3.1 Características de un Almacén de datos

Un DW posee varias características que brindan ventaja al centralizar los datos de las organizaciones:

- **Integrado:** Los datos se integran en una estructura consistente. Es decir, las inconsistencias de los sistemas operacionales son eliminadas. Debido a que, el detalle de información es importante para resolver las necesidades de los usuarios.
- **Histórico:** El valor de una variable en el tiempo es importante en los DW ya que permite realizar análisis de tendencias o comportamientos a lo largo de la historia de la información, diferenciándose de un sistema transaccional que muestra información del presente.
- **No volátil:** Se refiere a que la información alojada en el DW únicamente se puede leer mas no modificar, esta es permanente hasta que se haga una actualización de los datos, los valores serán reemplazados sin ningún tipo de acción sobre los que ya existían.
- **Temático:** En el desarrollo de un proyecto de este tipo el DW está relacionado directamente con el giro del negocio para que el análisis de

la información brinde ventaja competitiva ante otras empresas con la misma visión.

2.3.2 Ventajas de un almacén de datos

Calidad, coherencia y precisión de datos, al tener información global e integrada del giro del negocio. Proporciona una herramienta potente para la toma de decisiones en cualquier zona estratégica de la empresa.

Es fácil aplicar técnicas de análisis, estadísticas y modelización de información almacenada, para transformar la información en conocimiento y que este aporte al negocio.

Inteligencia Histórica, se puede aprender del pasado y predecir situaciones en el futuro y mitigar el riesgo de una amenaza para la empresa.

Separación del procesamiento de análisis de las bases transaccionales, logrando un mejor desempeño entre ambos sistemas.

2.3.3 Diferencias entre un almacén de datos y una base de datos

La principal diferencia radica en que un DW está diseñado para en análisis de datos. Es decir, permite lectura de grandes volúmenes de información para comprender las tendencias y relaciones internas. Por otro lado, las DB son usadas para registrar y almacenar datos. Por ejemplo, los registros de detalles de una transacción. En la tabla 1 se complementan las características del DW.

Tabla 1.

Diferencias entre DWH y Data Base.

CARACTERÍSTICAS	DATA WAREHOUSE	BASE DE DATOS TRANSACCIONAL
Cargas de trabajo admitidas	Análisis, generación de informes.	Orienta al procesamiento de transacciones.

Origen de datos	Datos recopilados y normalizados desde varios orígenes.	Datos registrados tal cual, desde un único origen, como un sistema transaccional.
Registro de datos	Operaciones de escritura masivas normalmente en un cronograma en lotes predeterminado.	Optimizado para operaciones de escritura continua a medida que datos nuevos se encuentran disponibles.
Normalización de datos	Esquemas no normalizados, como los esquemas estrella o copo de nieve.	Esquemas estáticos con alto nivel de normalización.
Almacenamiento de datos	Optimizado para acceso simple y desempeño de consultas alta velocidad.	Optimizado para operaciones de escritura de alto procesamiento a un único bloque físico orientado a filas.
Acceso a los datos	Optimización para minimizar la E/S y maximizar el procesamiento de datos.	Grandes volúmenes de pequeñas operaciones de lectura.

2.4 Datamart

Básicamente es una base de datos departamental, permite almacenar datos de un equipo o una determinada unidad de negocios, ejemplo: Marketing, Ventas, Fianzas, etc. Además, brinda la capacidad de analizar información al detalle desde infinidad de perspectivas que afecten a los procesos de un departamento puntual. Este puede ser alimentado o cargado con información de un DW o a su vez integrar por sí mismo información de otras fuentes.

Las diferencias con un DW se las plantea en la tabla 2.

Tabla 2.

Diferencias entre DWH y Datamart.

CARACTERÍSTICAS	DATA WAREHOUSE	DATA MART
Ámbito	Centralizado, varios departamentos integrados juntos.	Descentralizado, área o departamento específico.
Usuarios	Toda la organización.	Un único departamento.
Origen de datos	Varios orígenes.	Un único origen o pocos o a su vez una porción de datos almacenados en un DW.
Tamaño	Grande, puede sobrepasar los gigabytes a petabytes.	Pequeño, generalmente alcanza decenas de gigabytes.
Diseño	De arriba hacia abajo.	De abajo hacia arriba.
Nivel de detalle de los datos	Datos completos y detallados.	Pueden incluir datos resumidos.

2.5 Cubos de Datos

2.5.1 Procesamiento de transacciones en línea

Término en inglés Online Transaction Processing (OLTP), son bases de datos transaccionales. Usadas en su mayoría en sistemas de gestión de relación con el cliente (CRM). Los cuales dependiendo del negocio involucran modificaciones a las estructuras relacionales de las bases de datos operaciones. Estas estructuras permiten el acceso a los datos dependiendo de las tareas del giro del negocio. Por ejemplo, un banco debe soportar enormes cantidades de transacciones por minuto y debe usar este tipo de estructura.

2.5.2 Procesamiento analítico en línea

Término en inglés Online Analytical Processing (OLAP), son bases de datos más orientadas a procesamiento analítico. Este análisis implica el procesamiento de grandes volúmenes de datos con el objetivo de extraer algún tipo de información como patrones o tendencias de ventas. Este es un comportamiento de un Data mart y tiene las siguientes características:

- Los datos son solo de lectura, brinda mayor velocidad a las consultas de estos.
- Los datos se pueden segmentar dependiendo del área del negocio.
- Historia de los datos a largo plazo.
- Se alimentan mediante ETL de varios orígenes de datos, información centralizada.

Su objetivo principal es ofrecer una solución que permita agilizar las consultas y evaluaciones a grandes cantidades de datos que se producen constantemente en empresas.

2.6 Extracción, Transformación y Carga

Para definir este concepto (ETL), se analizará la problemática que enfrentan las organizaciones ante recopilar datos de múltiples orígenes, en varios formatos y moverlos a un DW. Para resolver esta problemática se han desarrollado varias herramientas, servicios y procesos. Los mismos que ayudan a centralizar la información desde varios orígenes y en distintos formatos en uno o más almacenes de datos mediante el siguiente proceso.

Extraer datos de varias fuentes de información. Transformar los mismos de acuerdo con las reglas comerciales del giro del negocio. Para finalmente, cargar en un almacén de datos de destino.

Este trabajo se ejecuta en un motor especializado, con frecuencia se usan tablas de preparación para contener temporalmente los datos mientras se transforman, y finalmente se cargan en el destino (Microsoft, 2019).

2.6.1 Flujo de datos y Flujo de control

El flujo de control es importante en la ejecución de los ETL, este garantiza el orden en el procesamiento de tareas utilizando restricciones de precedencia. Cuando se habla de precedencia se refiere a que cada tarea es dependiente de la anterior. Por ejemplo: si se tiene “Tarea A” y “Tarea B”, la segunda se ejecuta cuando la “Tarea A” sea completada.

Por otro lado, el flujo de datos se ejecuta como tarea dentro de un flujo de control como se puede observar en la figura 2. Los datos se extraen de un origen, estos se transforman y cargan en repositorio. Por lo tanto, la salida de información “Tarea A” es la entrada de la “Tarea B”. Sin embargo, no es regla que un flujo de datos se ejecute detrás de otro, estos se pueden ejecutar en paralelo.

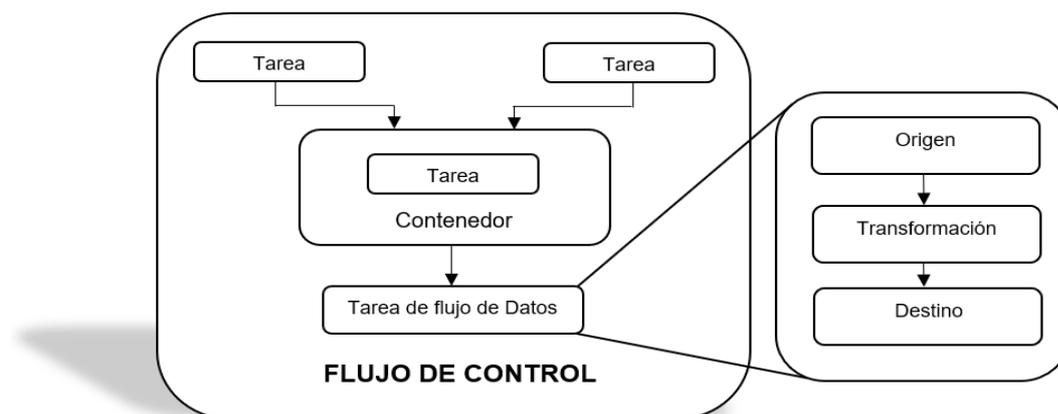


Figura 2. Esquema Flujo de control y Flujo de Datos.

2.7 Minería de Datos

La minería de datos o en inglés Data Mining es definida como el proceso de descubrir conocimiento en grandes cantidades de datos almacenados (Jiawei Han, Micheline Kamber 2001).

Es decir, se entiende como la acción de extraer conocimiento comprensible de grandes cantidades de datos de diversos orígenes y almacenados en distintos formatos. Estos deben pasar por un proceso detallado en la figura 3 desencadenando en la extracción del conocimiento y aplicación de este en el negocio.

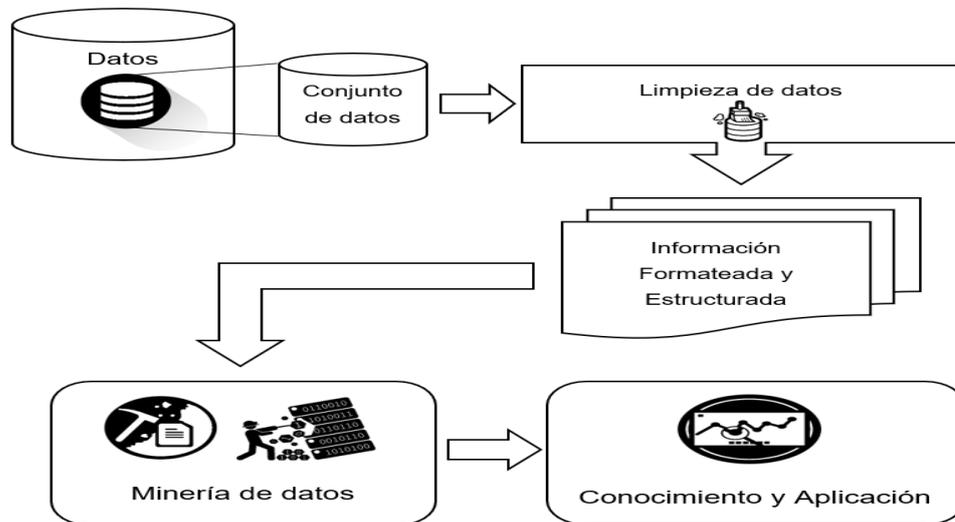


Figura 3. Esquema Minería de datos.

2.7.1 Arquitectura de Minería de datos

La arquitectura de la minería de datos consta de varios elementos que se ejecutan secuencialmente para llegar a un objetivo final conocido como conocimiento. Dichos elementos se listan a continuación y se usa la figura 4 para complementar los mismos.

- Orígenes de datos
- Almacenes de datos o Bases de datos
- Algoritmos de Minería de datos
- Evaluación de patrones
- Interfaces gráficas
- Base de conocimiento

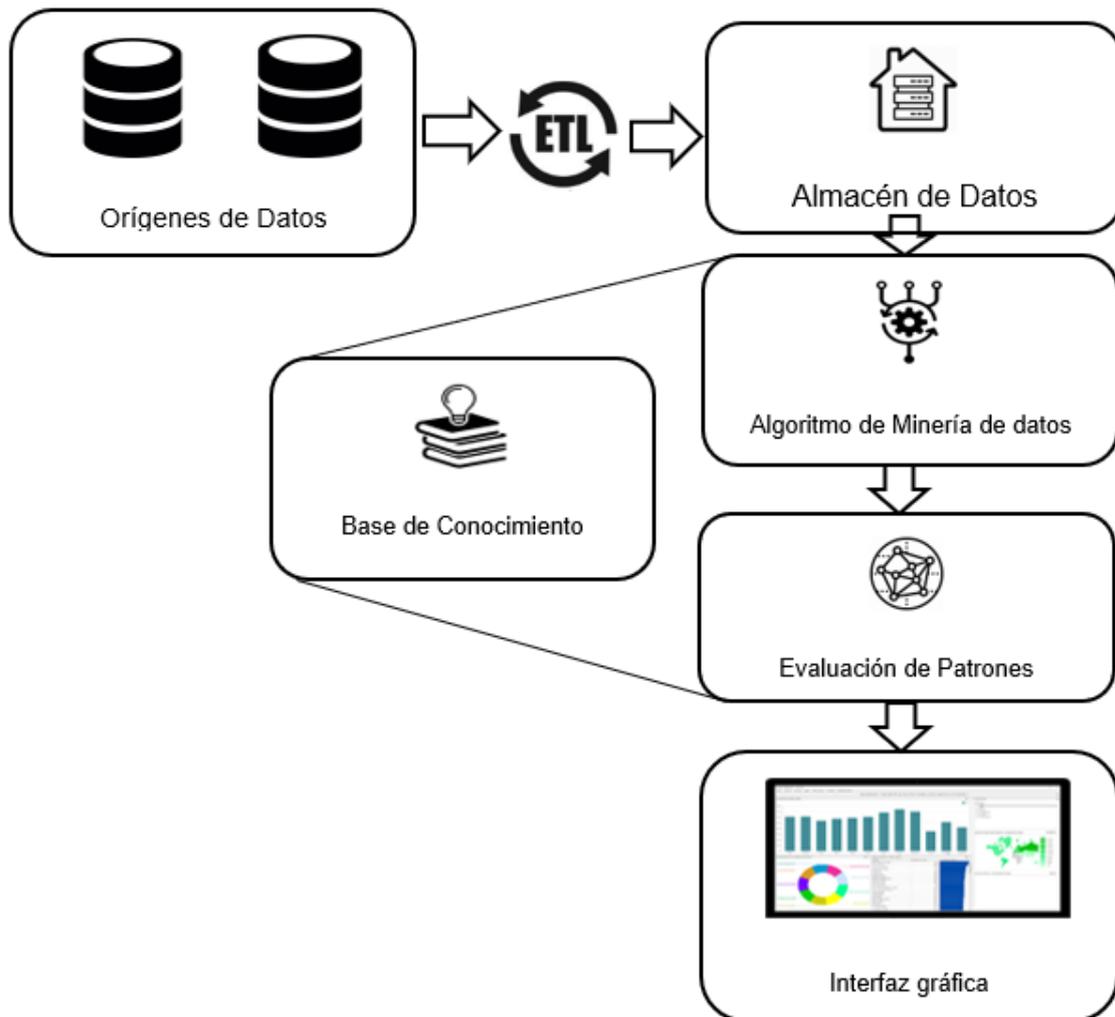


Figura 4. Arquitectura minería de datos.

2.7.2 Algoritmos de minería de datos

La minería de datos usa varias técnicas o algoritmos para descubrir conocimiento en grandes volúmenes de datos. Dicho algoritmo primero analiza los datos en busca de patrones o tendencias. De esta forma, define una serie de parámetros óptimos para la creación de un modelo eficiente de minería de datos. Finalmente, este se aplica en todo el conjunto de datos con la finalidad de extraer estadísticas detalladas y patrones procesables.

A continuación, en la figura 5 se muestra un esquema de las técnicas de minería de datos que existen.



Figura 5. Técnicas de minería de datos.

2.7.2.1 Árboles de decisión

Los árboles de decisión en inglés (Decision Trees), es la técnica más común que se usa para la minería de datos debido a su estructura simple. Como se puede observar en la figura 6, la raíz actúa como una condición, cada respuesta conduce a datos específicos que permiten determinar la decisión final.

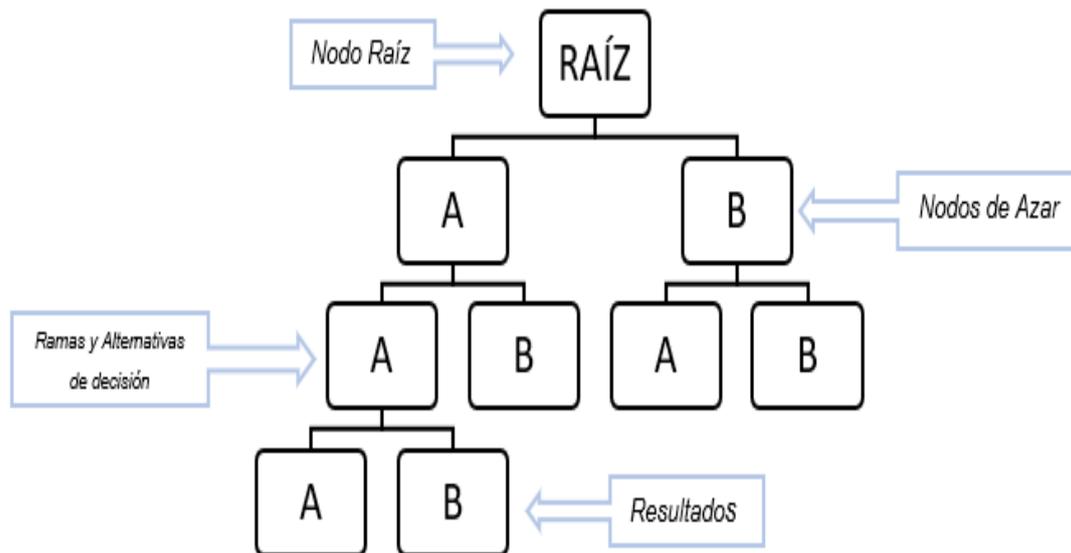


Figura 6. Esquema de un árbol de decisión.

2.7.2.2 Patrones Secuenciales

Al utilizar la minería de patrones secuenciales (Sequential Patterns), se pueden descubrir comportamientos o patrones estadísticamente relevantes en datos que están representados de manera secuencial. Esta técnica es uno de los métodos más usados para el descubrimiento de patrones. Debido a la frecuencia con que aparecen estos datos en aplicaciones reales.

2.7.2.3 Agrupamiento

Conocido en inglés como clustering, es un procedimiento que utiliza técnicas iterativas para agrupar una serie de datos, estos comparten propiedades comunes o características similares. Esta técnica de agrupación permite la exploración, identificación de anomalías en los datos y por ende la creación de predicciones. La figura 7 muestra gráficamente las agrupaciones que hace esta técnica de agrupación.

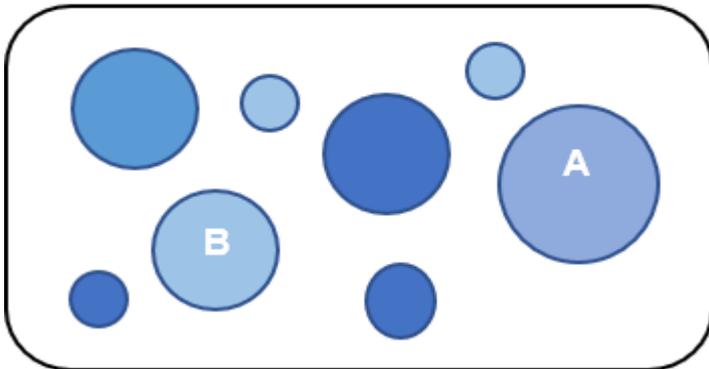


Figura 7. Algoritmo de agrupación.

A = Personas que compran por internet

B = Personas que compran físicamente

2.7.2.4 Regresión

La técnica de regresión (Regression), se utiliza para predecir variables continuas, basándose en otros atributos del conjunto de datos. Es decir, se puede utilizar para calcular una línea de tendencias en los datos de ventas de una organización.

2.7.2.5 Asociación

En inglés association o también conocida como técnica de relación, se utiliza para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. Estos contienen identificadores para casos individuales y para los elementos que contienen los casos. Los conjuntos de elementos es básicamente un grupo de elemento de un caso. Es decir, este modelo está compuesto por una serie de conjuntos de elementos y reglas que describen cómo dichos elementos se agrupan dentro de los casos.

La figura 8 muestra una serie de reglas en un conjunto de elementos. Las reglas que se identifican con el algoritmo permiten predecir posibles compras de un cliente en el futuro.

Rule
Road Bottle Cage = Existing, Cycling Cap = Existing -> Water Bottle = Existing
Mountain-200 = Existing, Mountain Tire Tube = Existing -> HL Mountain Tire = Existing
Mountain-200 = Existing, Water Bottle = Existing -> Mountain Bottle Cage = Existing
Touring-1000 = Existing, Water Bottle = Existing -> Road Bottle Cage = Existing
Road-750 = Existing, Water Bottle = Existing -> Road Bottle Cage = Existing
Touring Tire = Existing, Sport-100 = Existing -> Touring Tire Tube = Existing

Figura 8. Conjunto de reglas de un elemento.

Tomada de (Microsoft, 2019).

2.7.2.6 Clasificación

La técnica de clasificación e inglés classification, en términos generales predicen una o más variables discretas, basándose en los demás atributos del conjunto de datos.

2.8 Visualización de datos

2.8.1 KPIS

Las empresas generalmente evalúan sus resultados constantemente mediante indicadores o las siglas KPI (*Key Performance Indicator*). Por ejemplo, la evolución y el resultado de todas las estrategias de marketing que se están implementando en una agencia. Es una unidad de medida que aporta información sobre el funcionamiento de un aspecto determinado de la estrategia del negocio en un periodo de tiempo. De esta forma, se pueden tomar decisiones en base a algo que esté dando resultados positivos, o a su vez permite aplicar acciones correctivas.

Las características responden a las siglas en inglés SMART:

- Específicos (Specific)
- Medibles (Measurable)

- Alcanzables (Attainable)
- Realistas (Realistic)
- Definidos en el tiempo (Time)

las mismas definen las funciones que deben cumplir los KPI que se seleccionen para la estrategia de una empresa. La figura 9 muestra como los indicadores apoyan en los procesos de una organización.



Figura 9. Representación de Indicadores.

Tomada de (José Facchin, 2019).

2.8.2 Informes y visualización de datos

Para que un prototipo BI sea útil debe generar tableros con KPI coherentes. Esto con la finalidad de que los responsables del negocio puedan entender todas las estrategias que se están usando para tomar decisiones eficientes. Es decir, en la visualización de los datos se representa la información de la manera más intuitiva para conseguir comunicación simple, clara y efectiva. Así mismo, los KPI definidos sirven para monitorear el rendimiento de un departamento, o a su vez monitorear operaciones y procesos.

Estos tableros se los pueden desarrollar con herramientas de Inteligencia del negocio, herramientas SaaS o información manualmente distribuida. Estas dependen de la solución que se adapte más a las necesidades del negocio.

Un ejemplo de estas herramientas de reporting es Power BI. Esta herramienta será usada en el presente de desarrollo y se los explicará a profundidad más adelante. No obstante, en este apartado se permite mostrar un ejemplo de tablero desarrollado en dicha herramienta como se puede ver en la figura 10.

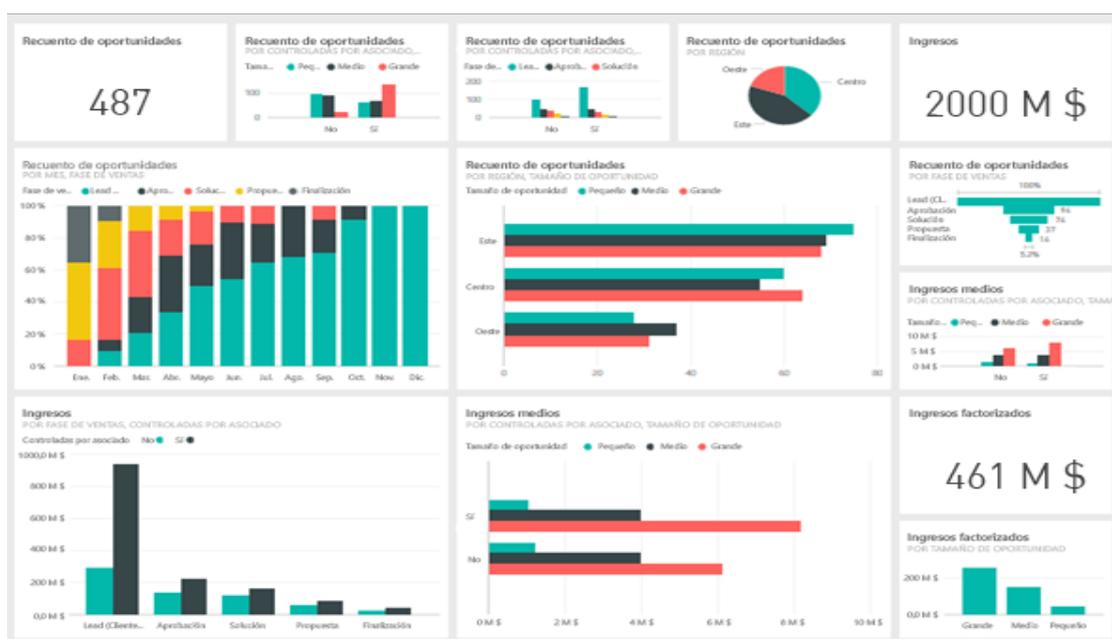


Figura 10. Tablero Power BI.
Tomada de (Microsoft, 2019).

3. Capítulo III. Análisis y configuración de Herramientas

El presente capítulo se basa en analizar herramientas que hacen posible el análisis de datos en un contexto empresarial. Esta se enfocará en realizar un breve análisis de herramientas “Open Source” frente al paquete ofrecido por Microsoft. Así mismo, una vez seleccionada la herramienta se procederá a dar una breve explicación de cómo se configura el ambiente de desarrollo para realizar los procesos BI necesarios con el fin de implementar el prototipo propuesto.

3.1 Comparación de Herramientas BI

Actualmente en el mercado existen diversas herramientas BI. Entre estas se encuentran la suite de Microsoft, IBM, SAP, OBI, Open Source, entre otras. Al comparar este tipo de herramientas se debe considerar las siguientes características:

- Rendimiento
- Escalabilidad
- Disponibilidad
- Integridad
- Precios
- Conexión a orígenes de datos
- Seguridad

En base a todas estas características se establece el cuadrante mágico de GARTNER, quien recientemente ha publicado su análisis de plataformas BI. Este cuadrante se lo puede ver de mejor manera en la figura 11, donde se detalla lo siguiente:

El aspecto de visión completa o *vision completeness* Microsoft está a un 26% arriba de Tableau. Por otro lado, en Habilidad para ejecutar o *ability to Execute* Tableau lleva la delantera con 1% sobre Microsoft. Sin embargo, solo es un punto de diferencia.

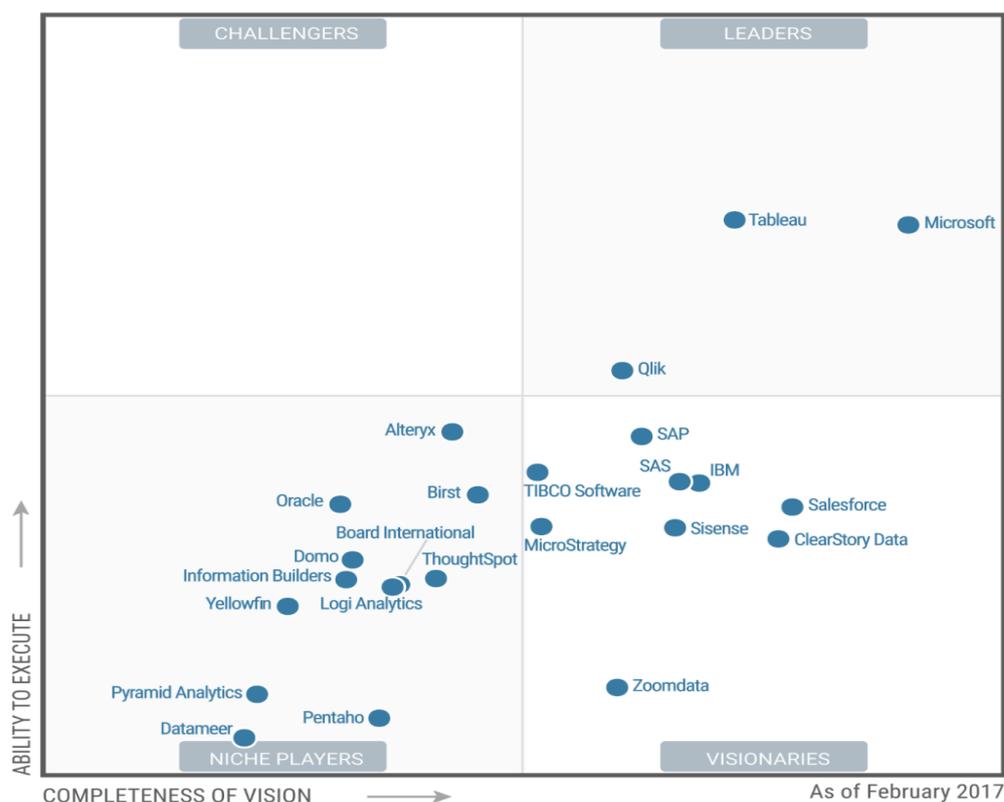


Figura 11. Cuadrante Mágico de GARTNER.

Tomada de (gravitar, 2019).

Ante esta comparativa se puede evidenciar que no hay mucha diferencia entre herramientas BI Microsoft y Open Source. Sin embargo, la decisión de utilizar la Suite de Microsoft es porque la fuente de los datos viene de una empresa multinacional que utiliza estas herramientas internamente. Por lo tanto, a continuación de detallarán estas herramientas en los siguientes puntos.

La suite de Microsoft BI consta de un conjunto de herramientas que brindan a las organizaciones a convertir datos brutos almacenados en información comprensible y extraíble. Actualmente es una herramienta muy popular en el mercado gracias a la gama completa de soluciones tales como:

- SSIS – SQL Server Integration Services.
- SSAS – SQL Server Analytical Services.
- SSRS – SQL Server Reporting Services.
- Power BI

Son usadas en cada etapa del modelo de inteligencia de negocio, dependiendo de la necesidad. De esta forma, con el correcto uso ayudan a convertir grandes cantidades de datos brutos en informes digeribles y de fácil entendimiento.

3.1.1 Servicios de Integración SQL Server

SSIS sigla que parte del término en inglés SQL Server Integration Services. Es una herramienta que sirve para la creación soluciones de integración y transformación de grandes cantidades de datos de nivel empresarial. Normalmente es usada para resolver problemas complejos del giro del negocio: descargando o copiando archivos, limpiando y minando datos, cargando o poblando almacenes de datos.

La integración de información puede ser de múltiples orígenes en diversos formatos como XML, archivos planos, bases relacionales para posteriormente cargarlos en uno o más destinos.

Gracias a la potente interfaz gráfica, la misma que posee una gran lista de tareas y transformaciones integradas. Se puede crear soluciones de integración sin escribir código tal y como se puede apreciar en la figura 12.

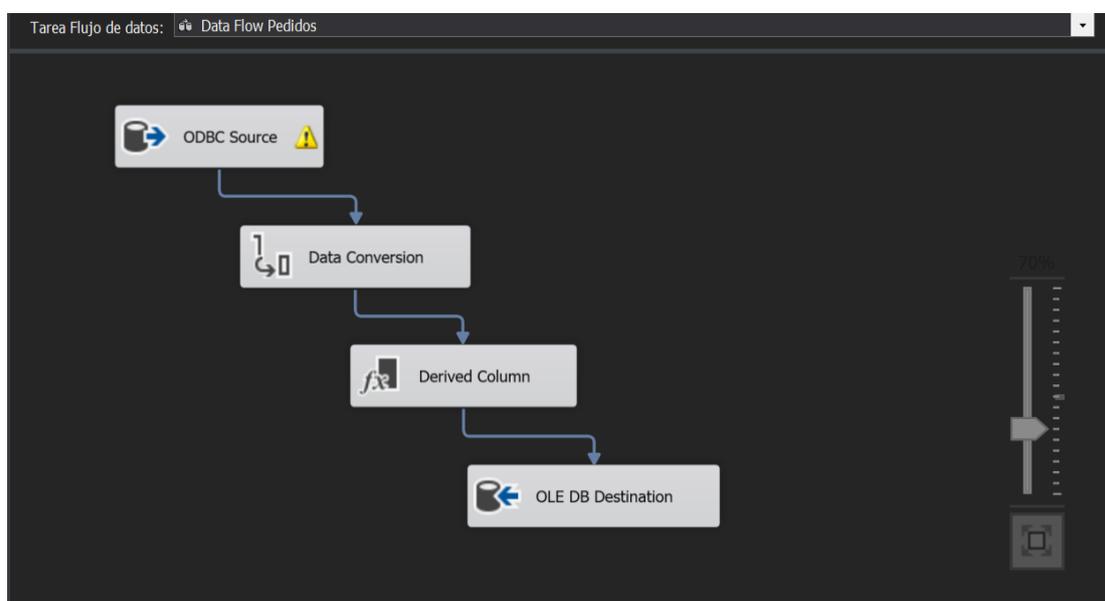


Figura 12. SQL Server Integration Services.

3.1.2 Servicios de Análisis SQL Server

SSAS sigla que proviene del término en inglés SQL Server Analytical Services. Es un motor de datos analíticos, comúnmente es usado en el soporte de decisiones o análisis empresariales. proporciona modelos de datos a nivel negocio para informes comerciales y aplicaciones de clientes como Excel, Power BI o informes en Reporting Services.

3.1.3 Servicios de Reportería SQL Server

SSRS sigla que proviene del término en inglés SQL Server Reporting Services. Esta herramienta proporciona una gama de herramientas y servicios locales donde se implementan y administran informes paginados o móviles. El primero se usa para crear informes de aspecto moderno. El segundo se basa en informes adaptables a diferentes dispositivos dependiendo del tamaño de la pantalla esto se puede apreciar en la figura 13.



Figura 13. Informe Paginado vs Informe Móvil.

Tomada de (Microsoft, 2019).

3.1.4 Diferencias entre Power BI y SSRS

Con frecuencia los desarrolladores de este tipo de tecnologías se preguntan sobre la diferencia entre estos dos conceptos.

La principal diferencia entre estos dos radica en que Power BI es un servicio de reportería en la nube de Microsoft Azure. Lo que significa que los informes no están localmente en los servidores de las empresas. En el caso de SSRS si se almacena localmente. Debido a que esta herramienta es instalada localmente junto con la licencia de SQL Server por ende los informes son desarrollados localmente.

Otra diferencia está en la forma en que se comparten los informes. Power BI actualmente consta de aplicaciones móviles nativas en IOS, Android, Windows. Estas permiten la exploración de datos off line a diferencia de SSRS que se lo hace de forma manual.

Finalmente, la curva de aprendizaje de SSRS para el personal técnico no se compara con la de Power BI que es una herramienta potente e intuitiva para el desarrollador.

Con todo lo mencionado anteriormente, se puede concluir que el uso entre estas dos depende del presupuesto del negocio. Debido a que empresas que están empezando en el mercado y no cuentan con la suficiente inversión para usar Power BI pueden optar por salir a producción con SSRS.

3.1.5 Importancia de Power BI Microsoft

Como se mencionó en el anterior punto, con Power BI se puede generar informes basados en la nube, su aplicación desktop es perfecta para crear y publicar informes. La principal característica se basa en la capacidad ilimitada para extraer datos de varios orígenes. Al tener varios componentes en su interfaz gráfica permite crear informes amigables para el usuario final tal y como se puede ver en la figura 14.

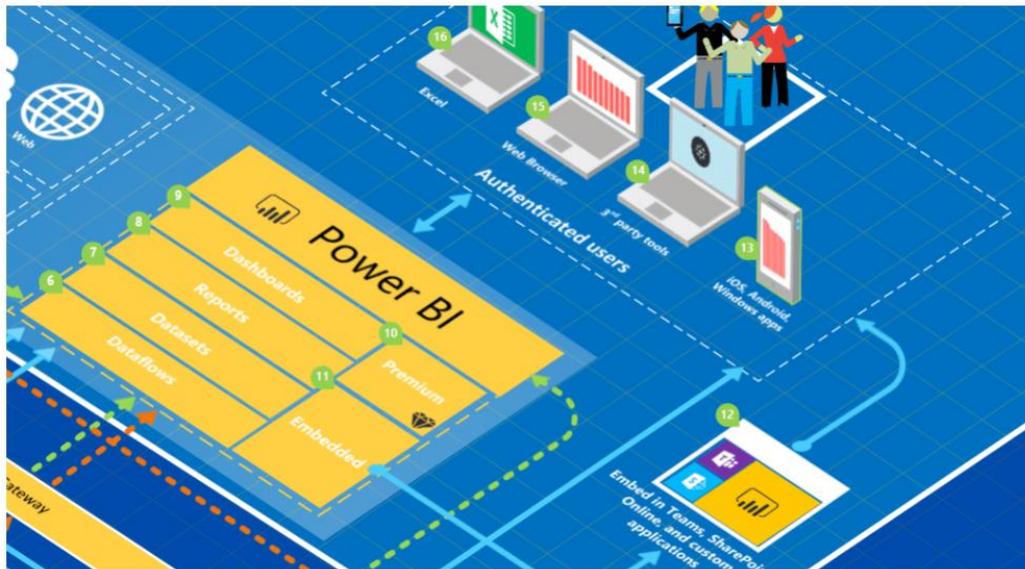


Figura 14. Esquema Power BI.

Tomada de (Microsoft, 2019).

3.2 Instalación de Herramientas

Una vez definidos los conceptos generales de la suite de herramientas BI de Microsoft, la mismas que serán usadas para el desarrollo del prototipo del presente proyecto, se procede con la instalación y configuración.

3.2.1 Instalación SQL Server

Una vez descargado el ejecutable .exe de la página oficial de Microsoft, se procede con la instalación, tal y como se puede ver en la figura 15.



Figura 15. Centro de instalación SQL Server.

Seleccionar en la opción que permite una nueva instalación independiente.

Los pasos que siguen son sencillos como: seleccionar la versión de instalación, aceptar los términos y condiciones, seleccionar las características que se desean instalar, esto se puede apreciar en la figura 16.

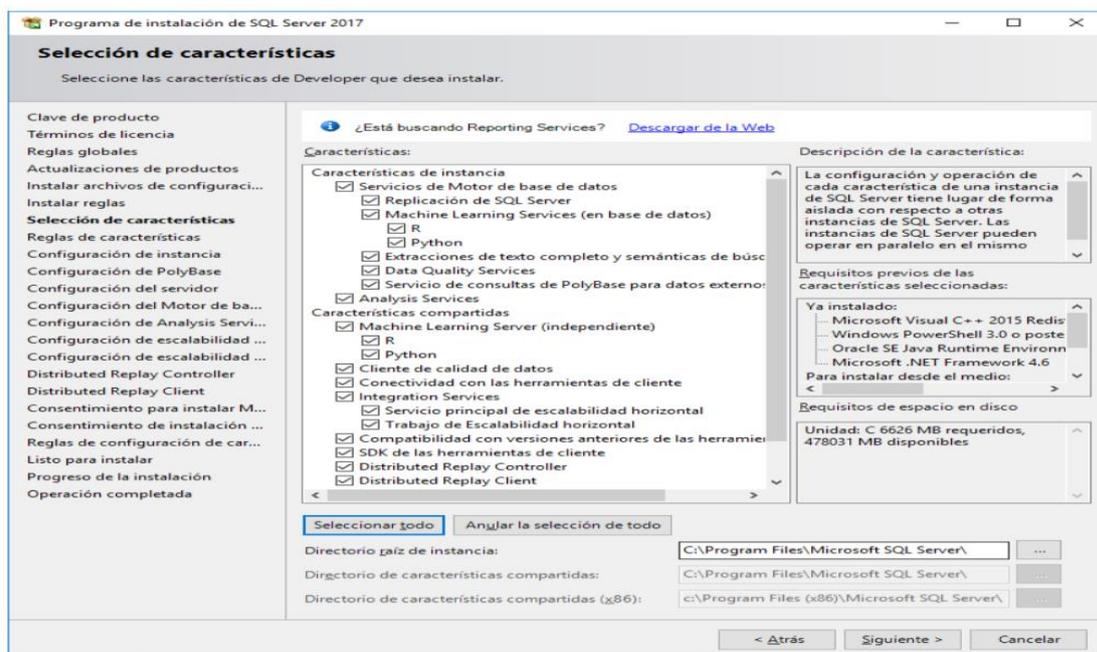


Figura 16. Selección de características SQL server.

A continuación, los siguientes pasos varían de acuerdo con las opciones previamente seleccionadas para instalar, se configura la instancia, es recomendable usar una con un nombre.

Después, configurar la Polybase, cuentas con las que iniciarán los servicios de SQL Server, desactivar sensibilidad de tildes y mayúsculas.

Configurar motor de bases de datos, directorios donde se almacenará la información Temp y FileStream, tal y como se puede ver en la figura 17.

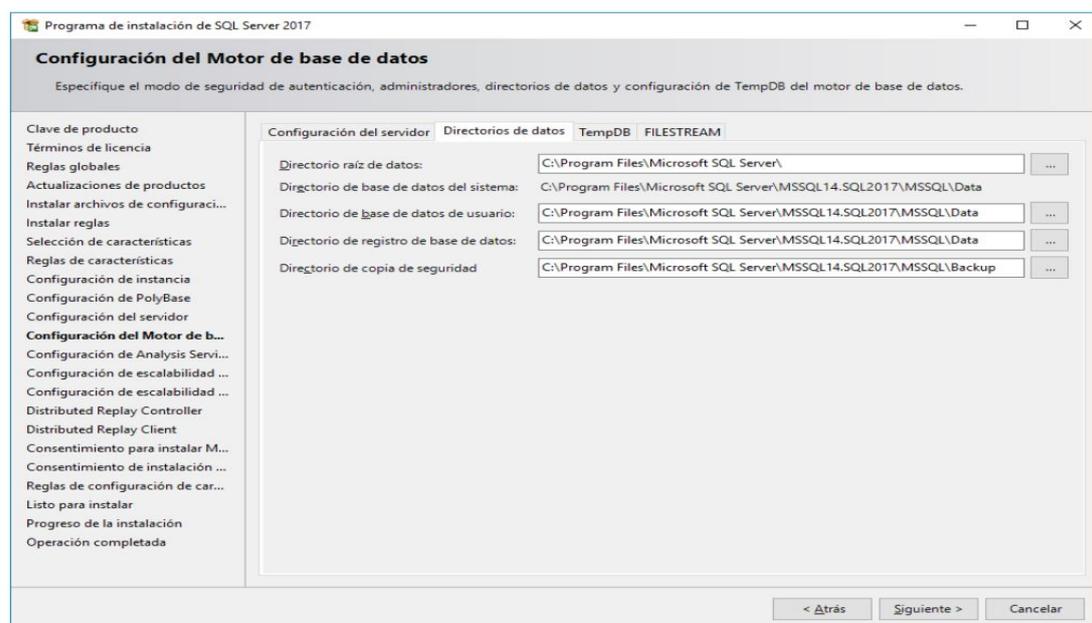


Figura 17. Configuración del motor de base de datos.

Para ir terminando, se configura Analysis Services, agregar el usuario que será el administrador del servicio.

Al finalizar la instalación, se debe comprobar que todos los servicios seleccionados previamente se hayan instalado correctamente, esto se puede evidenciar en la figura 18.

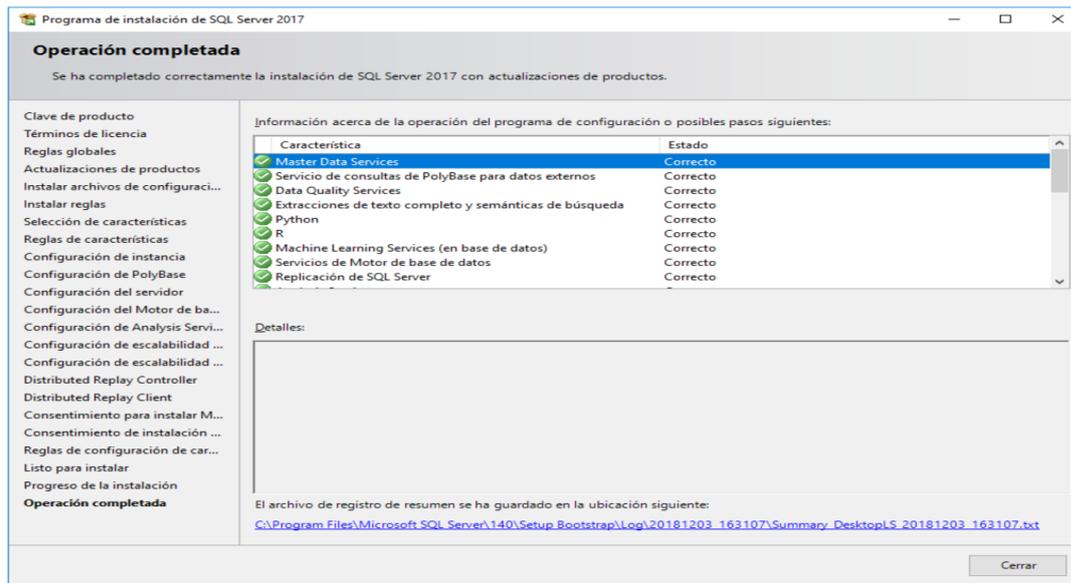


Figura 18. Servicios Instalados.

3.2.2 Instalación Visual Studio y configuración paquetes SISS

De igual manera, con el archivo .exe obtenido de la página oficial de Microsoft visual studio, se procede a ejecutar el asistente de instalación donde se puede seleccionar nuevos componentes y cargas de trabajo, tal y como se puede ver en la figura 19.

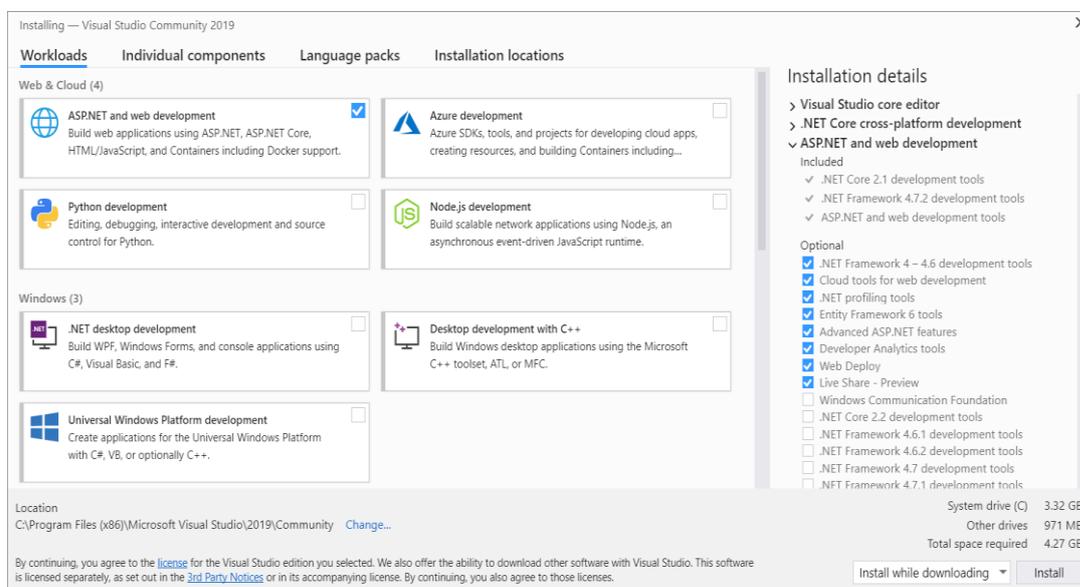


Figura 19. Selección de cargas de trabajo.

Además, se puede seleccionar el idioma, este es opcional según la figura 20.

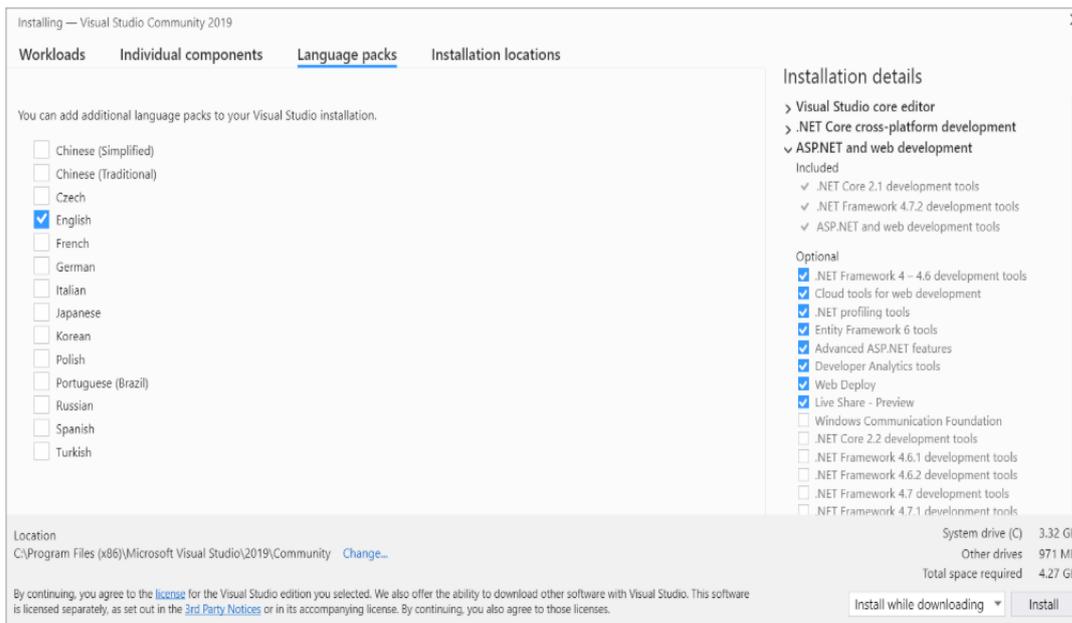


Figura 20. Language Packs.

Así mismo, opcionalmente se puede seleccionar la ubicación de instalación y finalmente se procede con la instalación, el asistente realiza la instalación automáticamente y se abre el programa.

El próximo paso será la configuración del paquete o extensión SSIS para realizar todas las integraciones que requiere el presente desarrollo.

Primero es necesario crear un nuevo proyecto en blanco para configurar la extensión SQL Server Integration Service, tal y como se puede ver en la figura 21.

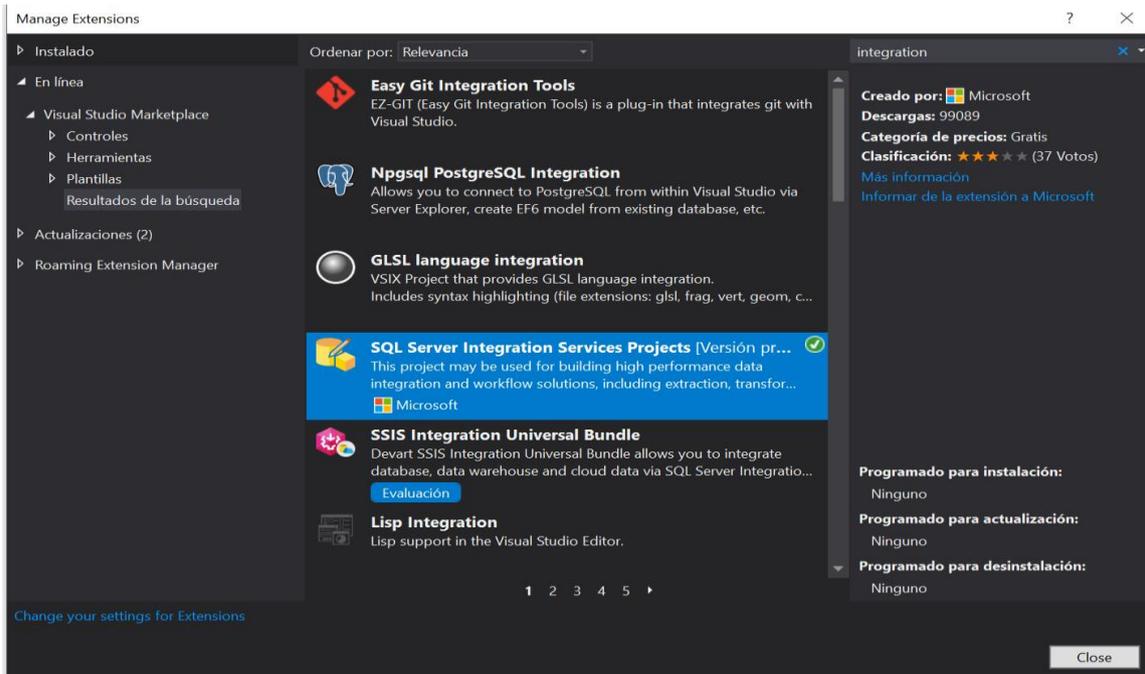


Figura 21. Instalación Extensión SSIS.

Una vez descargada, se ejecuta y se instala dentro de Visual Studio para poder hacer uso de esta, esto se puede ver en la figura 22.

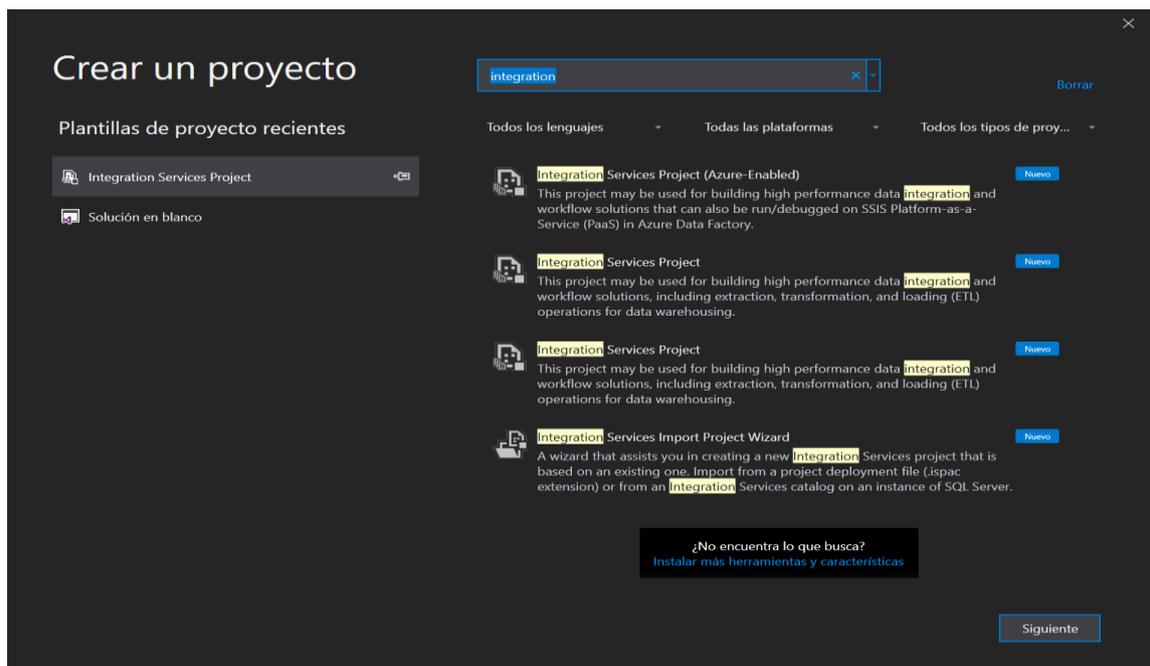
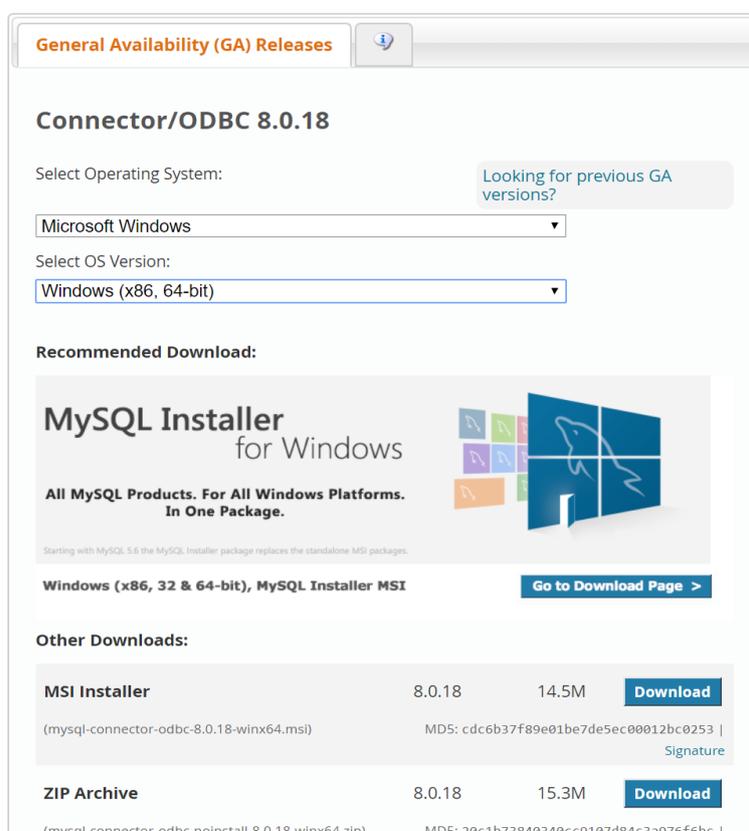


Figura 22. Comprobar Extensión SSIS.

Como se ha detallado a lo largo de este documento, estas herramientas sirven para integrar datos empresariales de distintos orígenes en diversos formatos, en el presente prototipo a desarrollarse se va a tomar como origen una estructura de base de datos open source MySQL por lo que se debe instalar un conector para que esta integración funcione, esto se detalla a profundidad en el siguiente punto.

3.3 Instalación ODBC MySQL Conector

Este complemento es muy importante para enlazar motores de bases Open Source como MySQL. La instalación es relativamente sencilla, basta con descargar el ejecutable de MySQL Community Downloads, tal y como se puede apreciar en la figura 23.



General Availability (GA) Releases

Connector/ODBC 8.0.18

Select Operating System: [Looking for previous GA versions?](#)

Microsoft Windows

Select OS Version: Windows (x86, 64-bit)

Recommended Download:

MySQL Installer for Windows

All MySQL Products. For All Windows Platforms. In One Package.

Starting with MySQL 5.6 the MySQL Installer package replaces the standalone MSI packages.

Windows (x86, 32 & 64-bit), MySQL Installer MSI [Go to Download Page >](#)

Other Downloads:

Download Type	Version	Size	Action
MSI Installer (mysql-connector-odbc-8.0.18-winx64.msi)	8.0.18	14.5M	Download
ZIP Archive (mysql-connector-odbc-noinstall-8.0.18-winx64.zip)	8.0.18	15.3M	Download

MDS: cdc6b37f89e01be7de5ec00012bc0253 | [Signature](#)

MDS: 20c1b73840340cc9107d84c3a976f6bc |

Figura 23. Conector ODBC MySQL.

Ejecutar y esperar que la instalación concluya, con esto se puede generar la conexión tal y como se puede apreciar en siguiente figura 24.

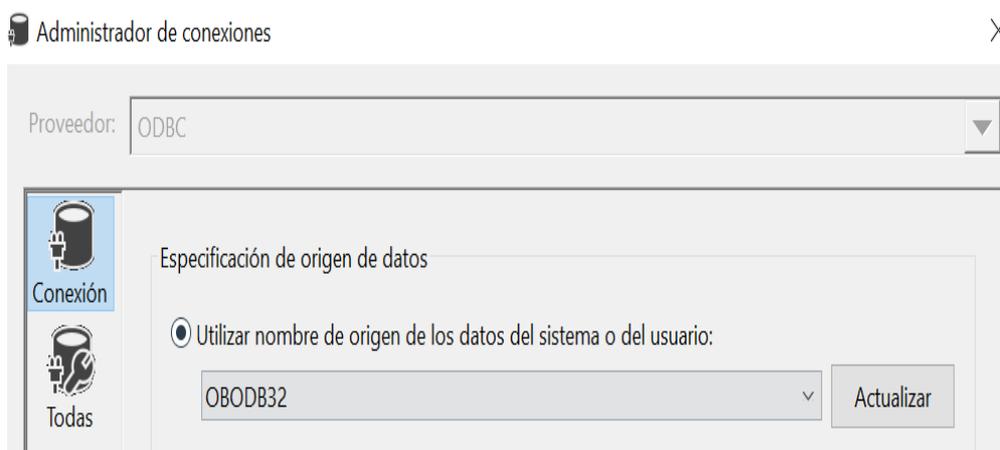


Figura 24. Conexión ODBC MySQL.

4. Capítulo IV. Diseño de prototipo de inteligencia de negocio.

Este capítulo se enfoca en todo el diseño del prototipo empezando por definir la estructura del almacén de datos. Dentro de este se definen las tablas con los campos que permiten el correcto análisis de la información. Siguiendo con los flujos de carga de data a dicha estructura (ETL) para finalmente establecer un cubo dimensional para la minería de datos. Todo esto será detallado a profundidad en los siguientes puntos.

4.1 Diseño del Data Warehouse

El diseño del DW consta de 3 fases empezando por el modelo dimensional, ETL para poblar la misma y el diseño de uno o varios cubos OLAP.

4.1.1 Análisis y diseño del modelo dimensional

Esta etapa inicial de construcción se basa en el modelado de los datos, en este se determinan las tablas de dimensiones y hechos. El prototipo utilizará un modelo de estrella donde tendrá dos tablas de hechos (ventas - pedidos) rodeada por tablas de dimensiones. La tabla que unirá las dos tablas de hechos es la de portabilidad ya que es el factor común entre las dos. La figura 25 muestra un esquema de todo lo mencionado en el párrafo anterior.

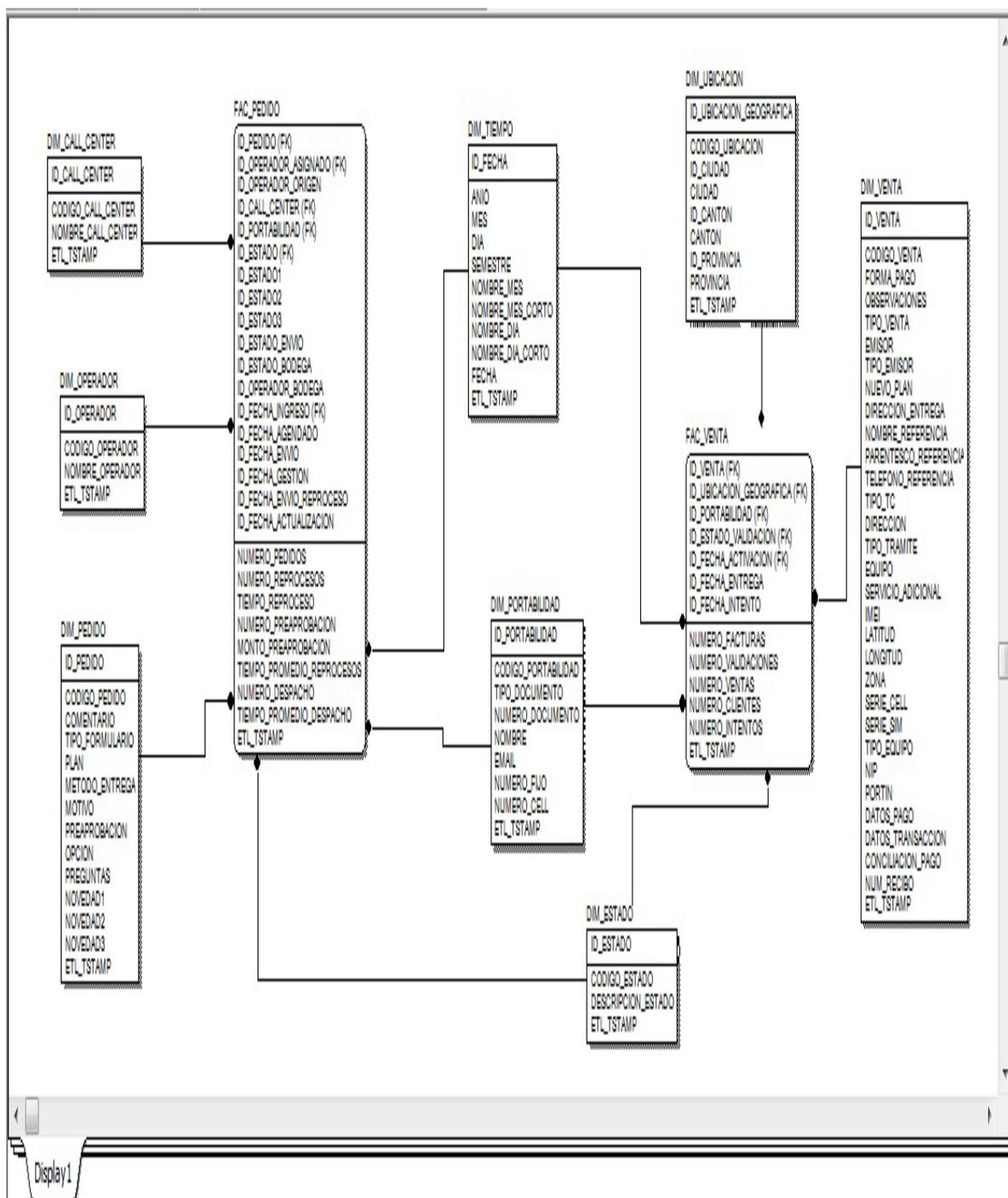


Figura 25. Modela Estrella.

Para llegar a este esquema se debe realizar el siguiente análisis. Primero, la forma no normalizada es donde se describen los datos tomados para el análisis del proceso de ventas asistidas. Como se puede ver la siguiente tabla que contiene la lista de cada campo con su respectivo detalle.

Tabla 3.

Forma no Normalizada.

CAMPO	DETALLE
campania	Nombre de campaña de ofertas
vendedor	Ejecutivo comercial de ventas
origen	Origen del pedido
callcenter	Proveedor de ventas
estado1	Tipificación estado primario
estado2	Tipificación estado secundario
estado3	Tipificación estado terciario
motivo	Tipificación motivo
direccionEntrega	Dirección entrega
provincia	Provincia de entrega
controlAgendado	Control de agendamiento del pedido
controlDespachado	Control de despacho del pedido
equipoSolicitado	Equipo solicitado por el cliente
planSolicitado	Plan solicitado por el cliente
utmCampaign	Urchin tracking module campania para trackeo de pedidos
utmSource	Urchin tracking module source para trackeo de pedidos
utmMedium	Urchin tracking module medium para trackeo de pedidos
utmContent	Urchin tracking module content para trackeo de pedidos
producto	Producto solicitado
origenUbicacionLP	Ubicación del elemento en el sitio web
numeroCel	Número celular cliente
numeroDocumento	Número cédula cliente
fechaIngreso	Fecha ingreso del pedido

fechaAgendado	Fecha en que se agenda el pedido para la entrega
fechaCitaEntrega	Fecha en que se cita para la entrega
fechaEnvioCourier	Fecha en que se envía a la entrega
fecha1G	Fecha de primera gestión del pedido
fechafin1G	Fecha fin de la primera gestión del pedido
intentos	Intentos de validación de fraude
cedula	Cedula cliente
tipo_lista	Tipo de lista de fraude
plan	Plan comprado
equipo	Celular comprado
numeroCelularFacturado	Número celular facturado
Dirección	Dirección de entrega
ciudad_venta	Ciudad registrada en la venta
ciudad_entrega	Ciudad de entrega

En base a lo anterior, la primera forma normal es donde se definen los nombres de los atributos atómicos, estos deben ser claros para poder separarlos en grupos repetitivos. Pasando a la segunda forma normal, que es donde se separan las llaves compuestas, previamente analizando los grupos de las llaves combinadas. Finalmente, la tercera forma normal donde se verifican las interdependencias entre los campos para proceder a analizar todas las tablas y determinar la tabla de hechos. La misma que es dependiente con todas las llaves foráneas definidas, todo este análisis entrega el siguiente resultado:

4.1.2 Tablas de Dimensiones

El modelo cuenta con 8 tablas de dimensiones, entre estas se encuentra la tabla de portabilidad que es la que enlaza las dos tablas de hechos. Todas estas tablas permitirán contextualizar las tablas de hechos dando diferentes perspectivas para su análisis. Estas están representadas por el acrónimo DIM que viene de

Dimensión seguido del nombre de la tabla a la que representa. Esto se puede apreciar de mejor manera en la figura 26.

DIM TIEMPO	
ID_FECHA	
FECHA	
ANIO	
MES	
DIA	
DIA_DEL_ANIO	
SEMANA_DEL_ANIO	
TRIMESTRE	
SEMESTRE	
NOMBRE_MES	
NOMBRE_MES_CORTO	
NOMBRE_DIA	
NOMBRE_DIA_CORTO	

DIM VENTA	
CODIGO_VENTA	
ID_VENTA	
TIPO_VENTA	
ETL_TSTAMP	
EQUIPO	
PROVINCIA	
CIUDAD	
LATITUD	
LONGITUD	

DIM PORTABILIDAD	
CODIGO_PORTABILIDAD	
NUMERO_CELL	
NUMERO_DOCUMENTO	
ID_PORTABILIDAD	
NOMBRE	
ETL_TSTAMP	

DIM OPERADOR	
CODIGO_OPERADOR	
NOMBRE_OPERADOR	
ID_OPERADOR	
ETL_TSTAMP	

DIM FRAUDE	
CODIGO_FRAUDE	
NUMERO_DOCUMENTO	
ETL_TSTAMP	

DIM PEDIDO	
CODIGO_PEDIDO	
PLAN_CELULAR	
EQUIPO_SOLICITADO	
ESTADO	
MOTIVO	
CAMPANIA	
ID_PEDIDO	
ETL_TSTAMP	

DIM CALL CENTER	
CODIGO_CALL_CENTER	
NOMBRE_CALL_CENTER	
ID_CALL_CENTER	
ETL_TSTAMP	

Figura 26. Tablas de Dimensiones.

4.1.3 Tablas de Hechos

El modelo dimensional cuenta con dos tablas de hechos que detallan el proceso del negocio (pedidos – ventas) y son importantes para el análisis completo. Como se puede apreciar en la figura 27 éstas contienen una clave principal que identifica la forma única de cada tabla. Claves externas que apuntan hacia las tablas de dimensiones. Medidas que representan columnas que contienen datos cuantificables, numéricos. Finalmente, Metadatos que permiten obtener información adicional sobre las filas (fecha).

FAC PEDIDO	
ID_PEDIDO	
ID_PORTABILIDAD	
ID_CALL_CENTER	
ID_OPERADOR_ASSIGNADO	
ID_FECHA_INGRESO	
ID_FRAUDE	
INTENTOS_GESTION	
ETL_TSTAMP	

FAC VENTA	
ID_VENTA	
ID_PORTABILIDAD	
ID_FECHA_VENTA	
ID_FRAUDE	
NUMERO_INTENTOS_VALIDACION	
ESTADO_VALIDACION	
CODIGO_VALIDACION	
ETL_TSTAMP	

Figura 27. Tablas de Hechos.

4.1.4 Definición de Procesos ETL

El siguiente paso será el análisis para la implementación de los procesos ETL.

Las extracciones de los datos se harán directamente de una réplica de una base de datos relacional MySQL estas se encuentra en AWS. La conexión local se hará mediante un conector ODBC, para extraer data de todas las tablas que involucran el proceso de ventas.

En la figura 28 se muestra un esquema de los ETL que se usarán para tratar los datos y poblar en el DW.

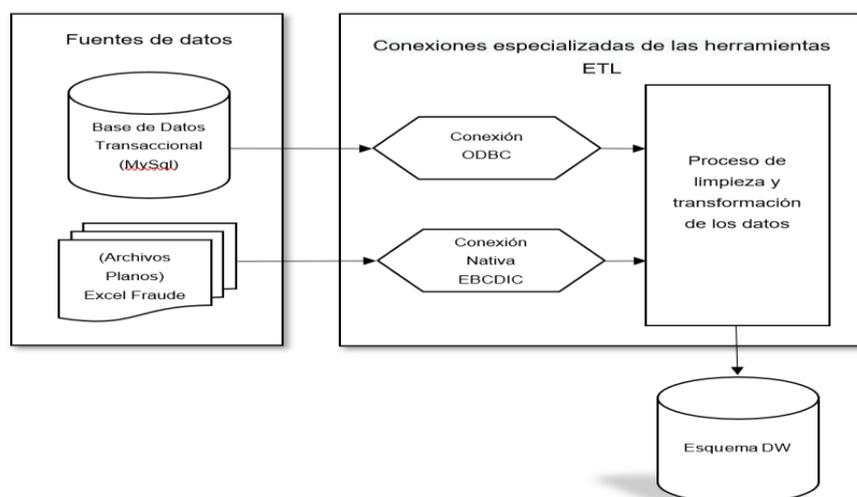


Figura 28. Diseño Extracción, Transformación y Carga.

En el diseño dimensional del DW existen 5 tablas que deben ser pobladas con datos de una base transaccional y archivos planos. Para esto, se usa un conector ODBC y mediante SQL se extraerán datos de las siguientes tablas:

- Pedidos
- Clientes
- Ventas
- Gestiones
- Proveedores
- Ejecutivos
- Campañas

Además, se usará un archivo plano excel que consta de toda la información de los clientes fraudulentos. Se podrá evidenciar la implementación de estos procesos en implementación de un caso real en el siguiente capítulo.

Tal y como se puede observar en la figura 29. Esta será la estructura básica de ETL que se usará para cada tabla y a la que se quiere llegar en el prototipo.

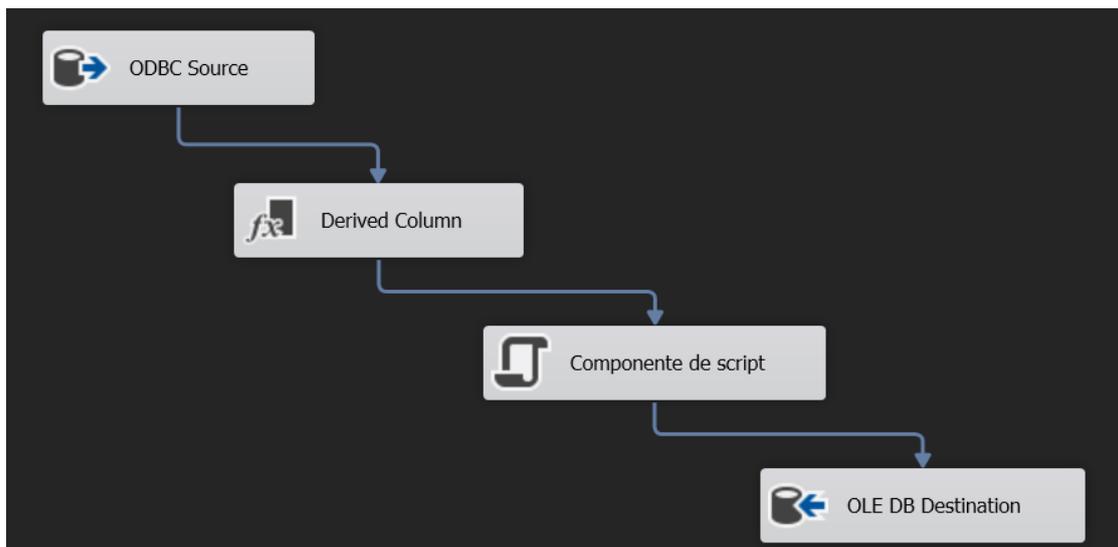


Figura 29. Diseño Básico ETL.

4.1.5 Diseño de cubos OLAP

Un cubo OLAP (On-Line Analytical Processing) se lo puede definir como un conjunto de coordenadas conocidas como dimensiones, estas determinan un valor de medida. Cuenta con una estructura jerárquica que permite o hace posible el análisis rápido de los datos, tal y como se puede apreciar en la figura 30.

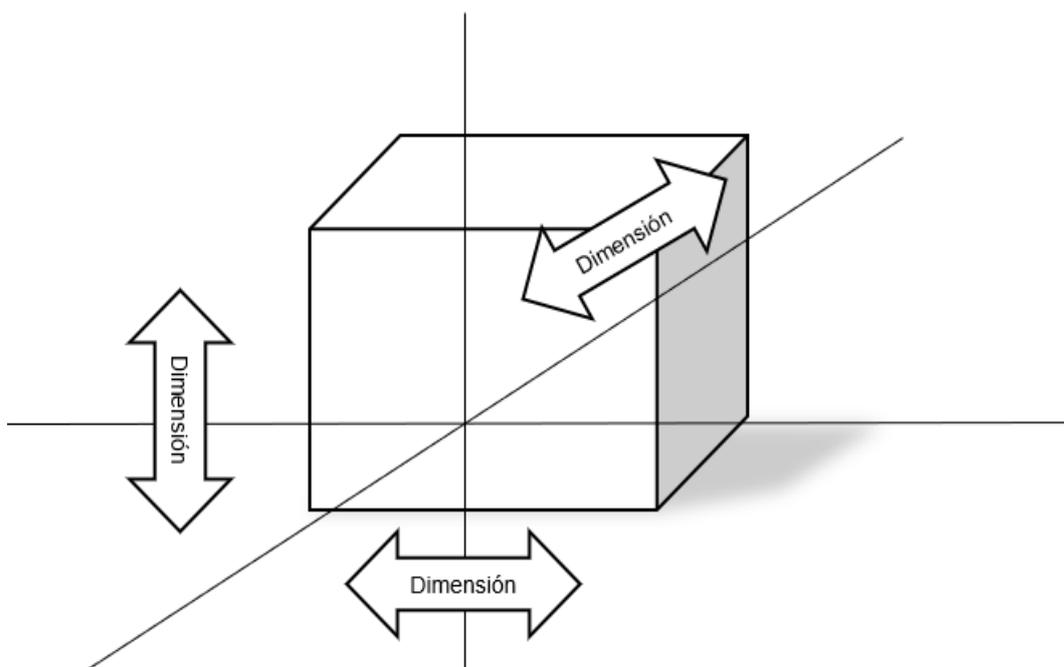


Figura 30. Estructura de un Cubo OLAP.

Cada coordenada o dimensión incorpora un campo específico para un tipo de datos, que posteriormente será comparado con la información contenida en las demás dimensiones. De esta forma, será posible analizar y desarrollar informes de la información relevante de la compañía mencionada.

Con el contexto general de lo que se puede realizar con un cubo se procede a realizar el diseño de un ejemplo con los datos definidos en anteriores puntos.

Inicialmente, dentro de una escala temporal se incluirán datos específicos de un periodo de tiempo definido (Ej.: diciembre 2018). Donde se incluirá una dimensión llamada "Cuarto Trimestre del 2018", esta a su vez tendrá otra

dimensión denominada “Año 2018”. Estas coordenadas son definidas de acuerdo con las necesidades de la organización y los datos que se requieran analizar.

Adicionalmente a esto, se pueden añadir otras dimensiones para complementar el análisis, en este caso se pueden mencionar algunas como: Ventas procesadas, no exitosas, exitosas, fraude, entre otras.

Con toda esta información poblada en el cubo. Se puede realizar un análisis completo de diversas situaciones. Ejemplo: comportamiento de clientes fraudulentos en un determinado de tiempo, porcentaje de ventas con un SLA de un minuto, etc. Con esto hallar soluciones antes los problemas detectados en el negocio.

En la figura 31 se muestra el diseño de un cubo que se utilizará en el prototipo de inteligencia empresarial.

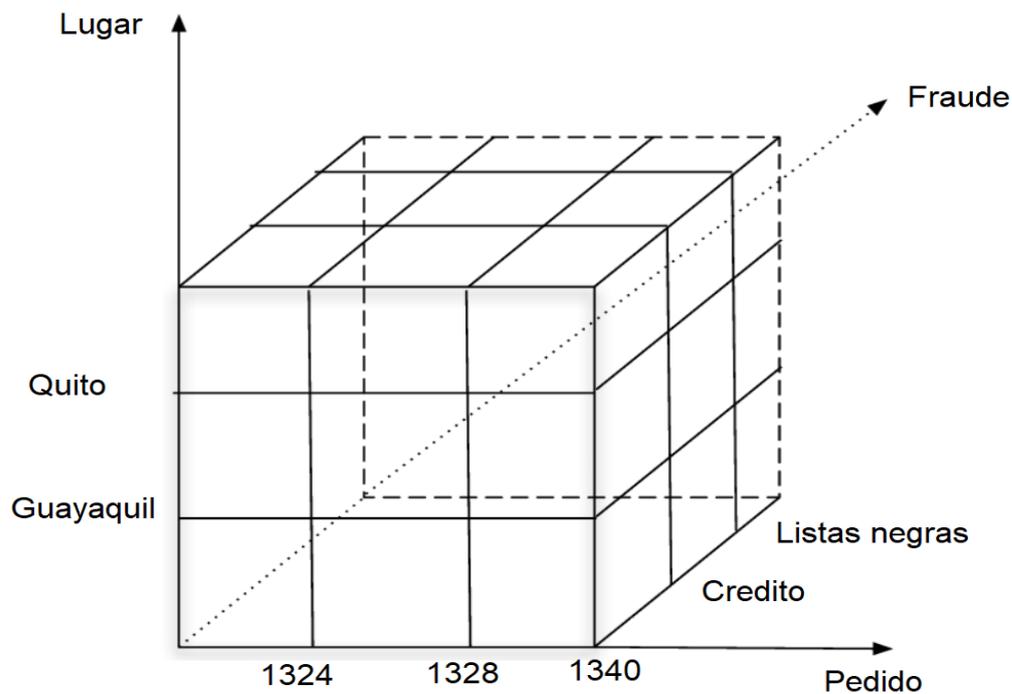


Figura 31. Cubo OLAP.

4.2 Diseño de tablero de indicadores

Un Tablero o en inglés Dashboard se lo define como un medio de comunicación visual. Donde se coloca información relevante, organizada y consolidada en una sola pantalla, con el objetivo de que pueda ser monitoreada con una sola mirada.

Para que el diseño e implementación de un Dashboard sea efectivo, este debe cumplir con la siguiente característica. Ayudar a las personas a tomar decisiones estratégicas mediante un número concreto representado en una métrica que mide algo esencial para la continuidad del negocio. Para resolver esto se debe manejar buenas prácticas de diseño y conocer los errores que normalmente se ejecutan cuando se desarrolla un tablero.

- Se debe optimizar el espacio para que los indicadores o KPIS ocupen una sola pantalla.
- Suministrar información real que provean valor y no presentar información compleja y sin sentido.
- Mostrar información con el mínimo detalle.
- Seleccionar un medio de visualización correcto.
- Los indicadores deben ser claros, eficientes y precisos.
- Organizar la información para que sea entendible y resaltar la más importante.
- No exceder la sobre decoración, utilizar fuentes y colores adecuados.

Siguiendo estas prácticas se garantiza que el tablero de indicadores sea efectivo y que junto a información consistente aporte valor a las decisiones estratégicas del negocio. Existen tres tipos de Dashboard:

- Analíticos.
- Estratégicos.
- Operacionales.

En este caso se utilizarán los dos primeros que permitirán: ordenar, comparar, filtrar, resultar, agregar, acceden por demanda a detalles, marcadores, etc.

5. Capítulo V. Implementación y análisis del prototipo inteligencia de negocio.

5.1 Infraestructura

La figura 32 muestra la arquitectura del caso en desarrollo. En esta se puede apreciar que los datos están alojados en un servicio de computación en la nube AWS. Este tiene configurado un servidor de base de datos de producción RDS con las siguientes características:

- Distribución Linux CentOS
- GB RAM
- vCPU
- Instancia de clase db.t2.large
- Storage 20 GiB
- Grupo de seguridades
- Entre otras configuraciones privadas

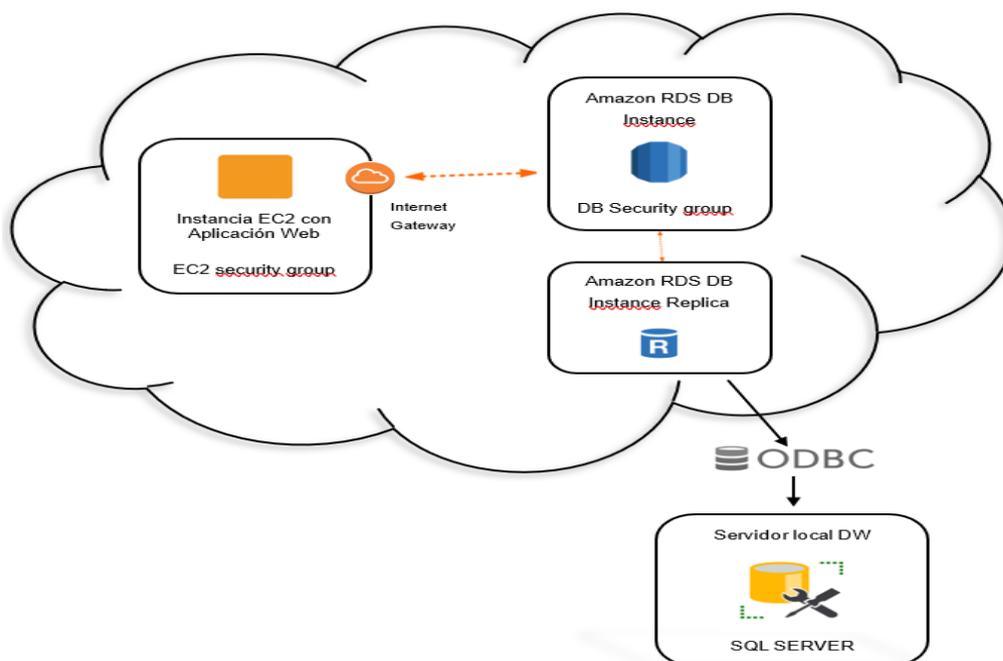


Figura 32. Arquitectura del Prototipo.

Además, la réplica RDS DB a la que se conectará el SSIS tiene las mismas características. Esta fue implementada para este caso práctico, para que la base de producción no sea usada directamente.

Finalmente, como se mencionó anteriormente el DW será implementado localmente en una base dimensional SQL Server. Donde se extraerán los datos con SSIS y los cubos serán realizados en la misma plataforma. De esta forma, posteriormente Power BI se conectará a estos para modelar los tableros gerenciales.

5.2 Implementación del Data Warehouse

Como se pudo apreciar en el capítulo 4 el diseño de la base dimensional dio como resultado un modelo tipo estrella con la siguiente estructura. Como se observa en la figura 33, esta estructura se divide en dos tablas de hechos con sus respectivas tablas de dimensiones a los extremos.

Adicional en el proceso ETL se agregó un campo *ETL_TSTAMP* en cada tabla de dimensiones, con la finalidad de guardar fechas de auditoria y mantener el control de cambios.

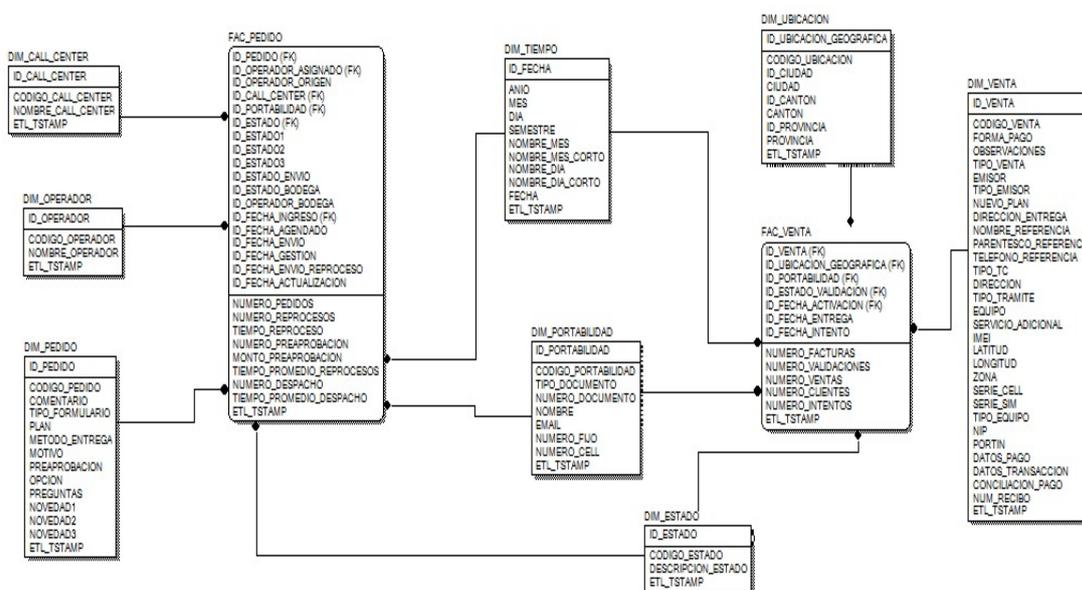


Figura 33. Modelo Estrella.

Ante este hecho, la implementación de cada tabla se la realizará con la herramienta SQL Server, mediante código o lenguaje SQL.

Empezando con la creación de la base de datos dimensional, dada por la siguiente sentencia SQL:

```
create database CanalOnlineV1;
```

Misma que abarcará todas las tablas se detallarán a continuación.

5.2.1 Tablas de dimensiones

Primero se implementará la tabla Dimensión Portabilidad con el alias DIM_PORTABILIDAD. Esto en base al análisis previo de los campos y tipos de datos que se realizó en el capítulo 4. Como se puede apreciar en el siguiente script se define la llave principal de la tabla. Seguido de cada campo con su respectivo tipo de dato, finalmente el campo de auditoria mencionado en el punto anterior.

```
/* DIMENSIÓN PORTABILIDAD */
CREATE TABLE DIM_PORTABILIDAD(
    CODIGO_PORTABILIDAD int NOT NULL ,
    NUMERO_CELL varchar(20) NULL ,
    NUMERO_DOCUMENTO varchar(20) NULL ,
    ID_PORTABILIDAD integer NOT NULL ,
    NOMBRE varchar(255) NULL ,
    ETL_TSTAMP datetime NULL
)
```

Después, se creará la Dimensión Tiempo o DIM_TIEMPO donde se almacenarán fechas importantes de los pedidos. Con este parámetro realizar un análisis de cada etapa de los pedidos basándose en fechas reales.

```
/* DIMENSIÓN TIEMPO */
CREATE TABLE DIM_TIEMPO
(
    ID_FECHA integer NOT NULL ,
```

```

FECHA                datetime NULL ,
ANIO                 int NOT NULL ,
MES                  int NULL ,
DIA                  int NULL ,
DIA_DEL_ANIO         int NULL ,
SEMANA_DEL_ANIO     int NULL ,
TRIMESTRE            varchar(20) NULL ,
SEMESTRE             varchar(20) NULL ,
NOMBRE_MES           varchar(20) NULL ,
NOMBRE_MES_CORTO    varchar(20) NULL ,
NOMBRE_DIA           varchar(20) NULL ,
NOMBRE_DIA_CORTO    varchar(20) NULL ,
)

```

Las 6 dimensiones siguientes son similares a las implementadas anteriormente, con sus respectivos campos que las caracterizan.

```

/* DIMENSIÓN CALLCENTER */
CREATE TABLE DIM_CALL_CENTER
(
    CODIGO_CALL_CENTER int NOT NULL ,
    NOMBRE_CALL_CENTER varchar(126) NOT NULL ,
    ID_CALL_CENTER     integer NOT NULL ,
    ETL_TSTAMP         datetime NULL
)

```

```

/* DIMENSIÓN OPERADOR */
CREATE TABLE DIM_OPERADOR
(
    CODIGO_OPERADOR int NOT NULL ,
    NOMBRE_OPERADOR varchar(128) NULL ,
    ID_OPERADOR     integer NOT NULL ,
    ETL_TSTAMP     datetime NULL
)

```

```

/* DIMENSIÓN PEDIDO */
CREATE TABLE DIM_PEDIDO
(
    CODIGO_PEDIDO int NOT NULL ,
    PLAN_CELULAR  varchar(255) NULL ,
    EQUIPO_SOLICITADO varchar(255) NULL ,
    ESTADO        varchar(255) NULL ,
    MOTIVO        varchar(255) NULL ,
    CAMPANIA      varchar(255) NULL ,
    ID_PEDIDO     integer NOT NULL ,
    ETL_TSTAMP   datetime NULL ,
)

```

```

/* DIMENSIÓN VENTA */
CREATE TABLE DIM_VENTA
(
    CODIGO_VENTA          integer NOT NULL ,
    ID_VENTA              integer NOT NULL ,
    TIPO_VENTA            varchar(255) NULL ,
    ETL_TSTAMP            datetime NULL ,
    DIRECCION_ENTREGA     text NULL ,
    DIRECCION              varchar(255) NULL ,
    EQUIPO                 varchar(150) NULL ,
    LATITUD                varchar(50) NULL ,
    LONGITUD               varchar(50) NULL ,
)

```

5.2.2 Tabla de hechos

Para empezar, las tablas de Hechos o conocida en el esquema estrella como FAC_VENTAS – FAC_PEDIDOS son implementadas de la siguiente forma.

```

/* HECHOS VENTA */
CREATE TABLE FAC_VENTA(
    ID_VENTA                integer NOT NULL ,
    ID_PORTABILIDAD         integer NOT NULL ,
    ID_FECHA_VENTA          integer NOT NULL ,
    NUMERO_INTENTOS_VALIDACION integer NULL,
    ESTADO_VALIDACION       varchar(100) NULL ,
    CODIGO_VALIDACION       varchar(100) NULL ,
    ETL_TSTAMP              datetime NULL ,
)

/* HECHOS PEDIDO */
CREATE TABLE FAC_PEDIDO(
    ID_PEDIDO                integer NOT NULL ,
    ID_PORTABILIDAD         integer NOT NULL ,
    ID_CALL_CENTER          integer NOT NULL ,
    ID_OPERADOR_ASIGNADO    integer NOT NULL ,
    ID_FECHA_INGRESO        integer NOT NULL ,
    INTENTOS_GESTION        integer NULL,
    ETL_TSTAMP              datetime NULL
)

```

Todas estas sentencias serán ejecutadas en la ventana de consultas de SSMS, tal y como se puede apreciar en la figura 34. Se ejecuta la porción de código seleccionado con F5 o el botón “Ejecute”, entregando un mensaje de Confirmación.

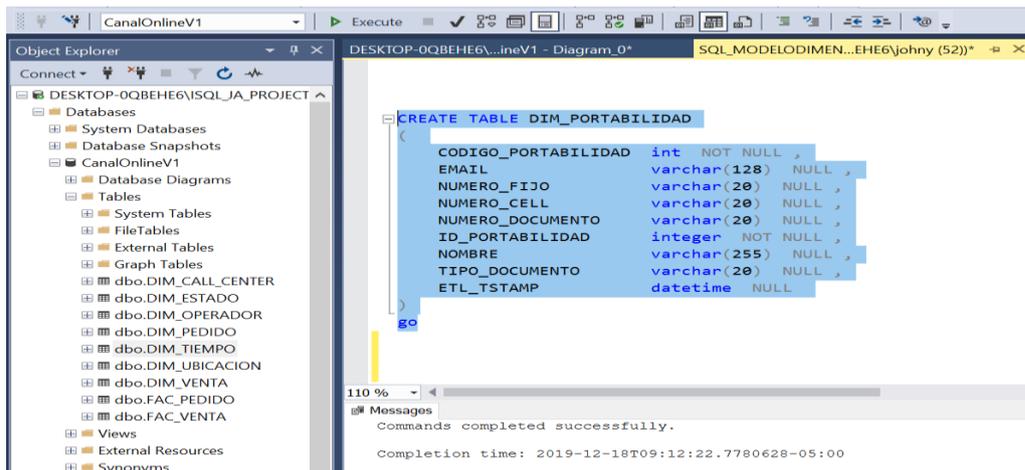


Figura 34. Ventana de Consultas SSMS.

En el explorador de objetos donde se puede apreciar el nombre de la base de datos CanalOnlineV1 y la tabla creada anteriormente, tal y como se puede ver en la figura 35.

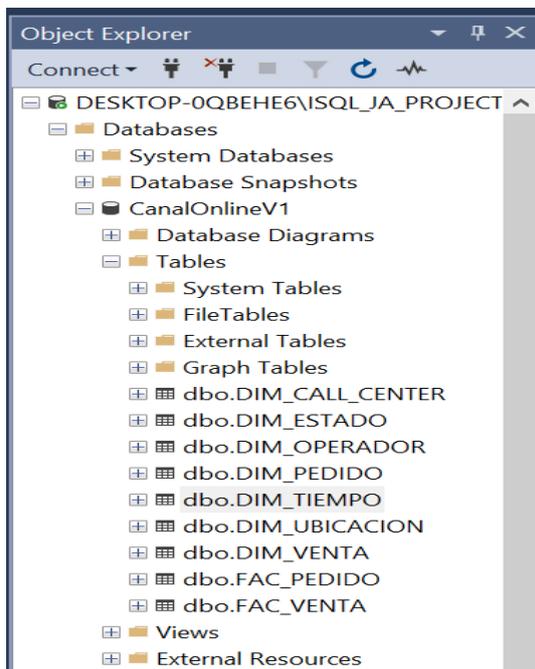


Figura 35. Explorador de objetos.

Después, en la misma herramienta se puede crear un esquema de la base dimensional, en base a las tablas creadas anteriormente. Tal y como se puede apreciar en la figura 36 basta con dos clics.

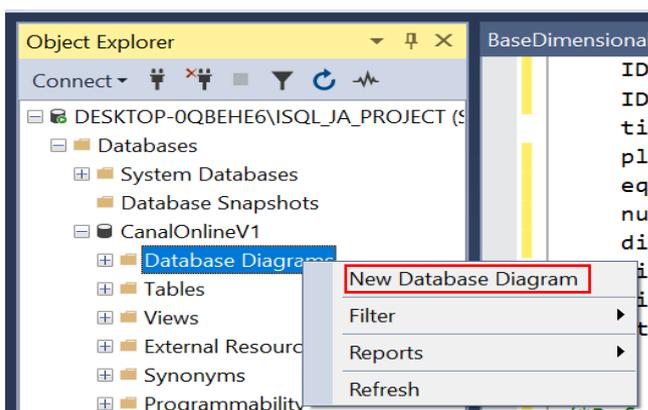


Figura 36. Creación Diagrama Base Dimensional.

Finalmente se obtiene el esquema representado en la figura 37, cabe recalcar que las relaciones son lógicas.

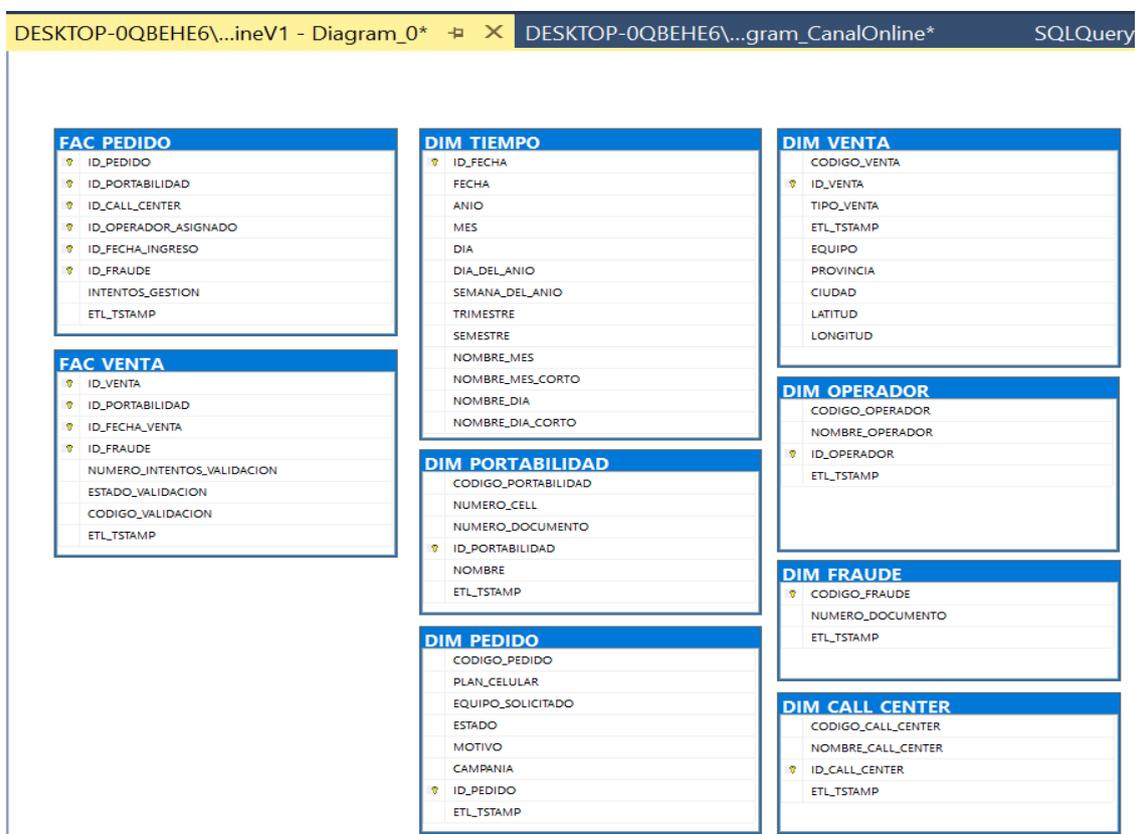


Figura 37. Diagrama Base Dimensional.

5.3 Conexión a fuente de datos

La conexión se la realizará mediante un conector ODBC MySQL, este permite unir dos gestores de datos diferentes, en este caso (MySQL – SQL Server). Una vez instalado este componente, se procede a configurar con los siguientes parámetros que se pueden ver en la figura 38.

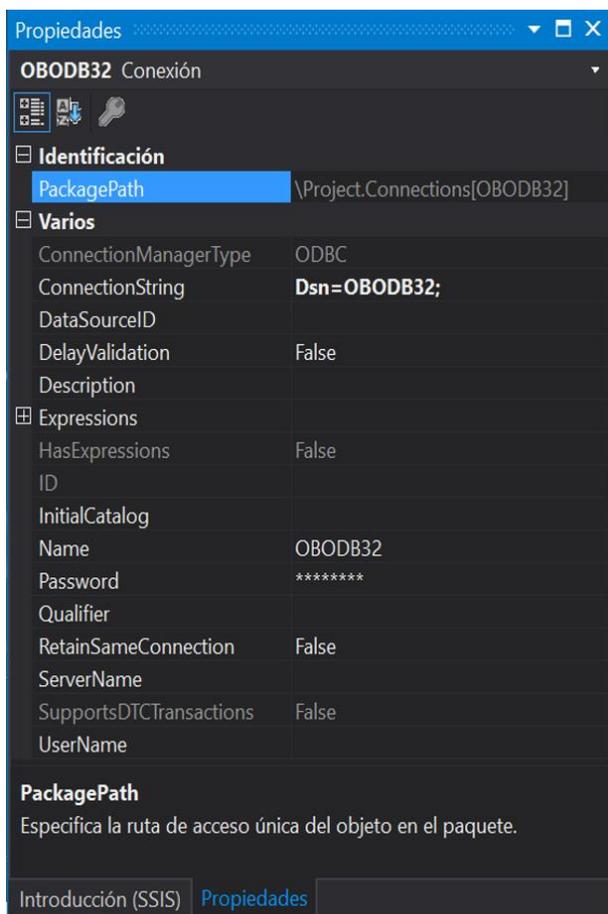


Figura 38. Parámetros ODBC.

Con esto ya se tendrá el enlace para extraer los datos. Sin embargo, por seguridad de la información AWS usa grupos de seguridades. Es decir, para poder acceder a los datos de la réplica se debe registrar (IP – Protocolo – Puerto) dentro del administrador de las instancias de Amazon. Tal y como se puede apreciar en la figura 39.

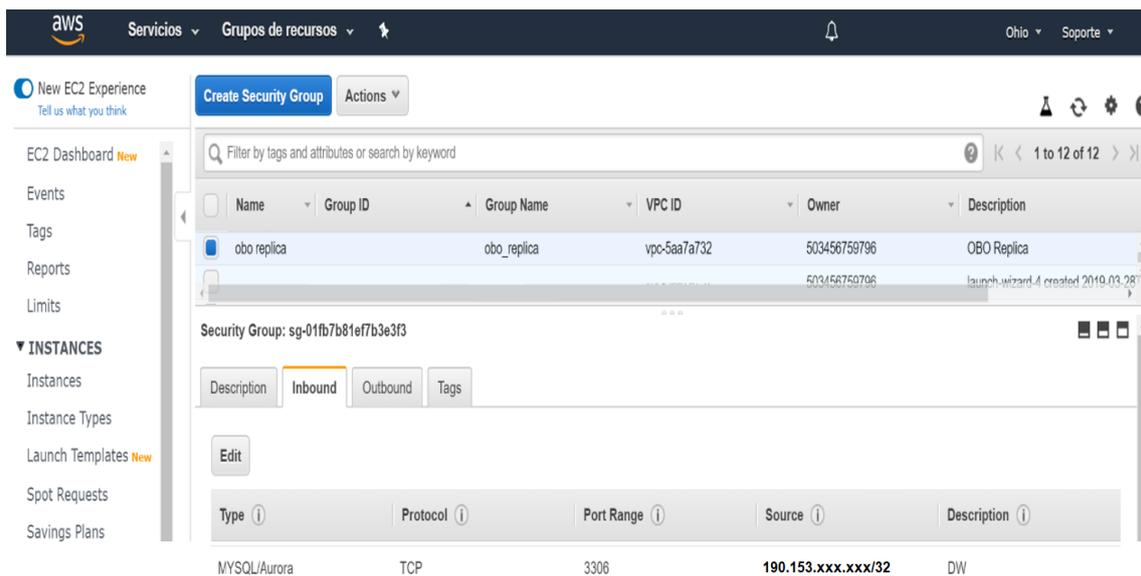


Figura 39. Grupo de Seguridad AWS.

Con estas configuraciones ya podemos acceder a la estructura de la base relacional y extraer los datos mediante procesos ETL.

De igual manera las conexiones a los archivos planos .xlsx serán con un elemento ODBC Source de SSIS.

5.4 Desarrollo de procesos ETL

Con la conexión estable a las fuentes de datos, se inicia la implementación de los ETL. Como se mencionó en subcapítulos del capítulo 4, las tablas que se usarán en la etapa de extracción son las siguientes:

- Pedidos
- Clientes
- Ventas
- Gestiones
- Proveedores
- Ejecutivos
- Campañas

Además, se utilizará una fuente de archivos planos .xlsx que contiene toda la información de fraude en base al número de documento del cliente. Posteriormente, en la etapa de transformación se realizarán tareas de definición de tipos de datos de cada campo, separación de información, limpieza, entre otros. Finalmente, la carga de los datos se realizará con una conexión OLE DB Destino a la estructura dimensional creada anteriormente.

En las siguientes figuras se muestra la manera de implementar cada ETL, empezando por la Dimensión Pedido. Primero se crea un flujo de control, este contendrá un flujo de datos, con la siguiente estructura presentada en la figura 40.

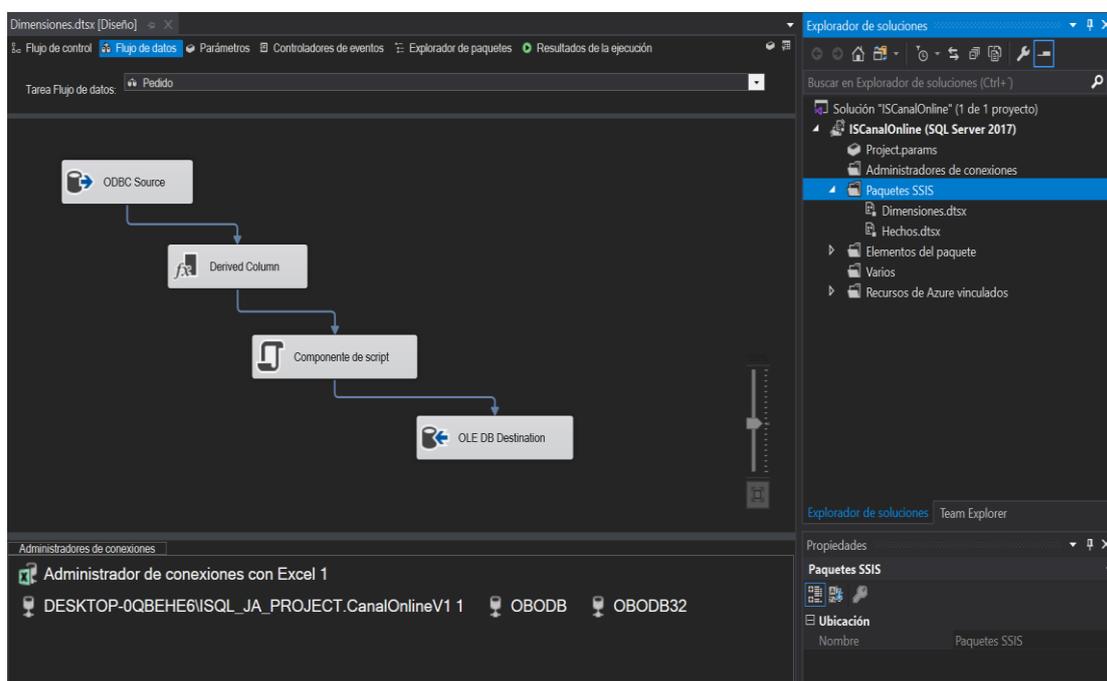


Figura 40. ETL en SSIS.

Al crear una solución SSIS (IntegrationServicesCOL) en Visual Studio, se puede acceder a cada componente para la elaboración de un ETL. Como se puede apreciar en la figura anterior, en la parte derecha se encuentra el “Explorador de soluciones”. En esta se puede observar el nombre de la solución y el directorio donde se almacena el archivo con la extensión dtsx. Este contiene todos los ajustes y configuraciones para la migración de los datos.

5.4.1 ODBC Origen

En el componente ODBC origen con la conexión establecida y la estructura del ETL implementada, se debe añadir todos los parámetros necesarios para poder extraer los datos. Tal y como se puede observar en la figura 41.

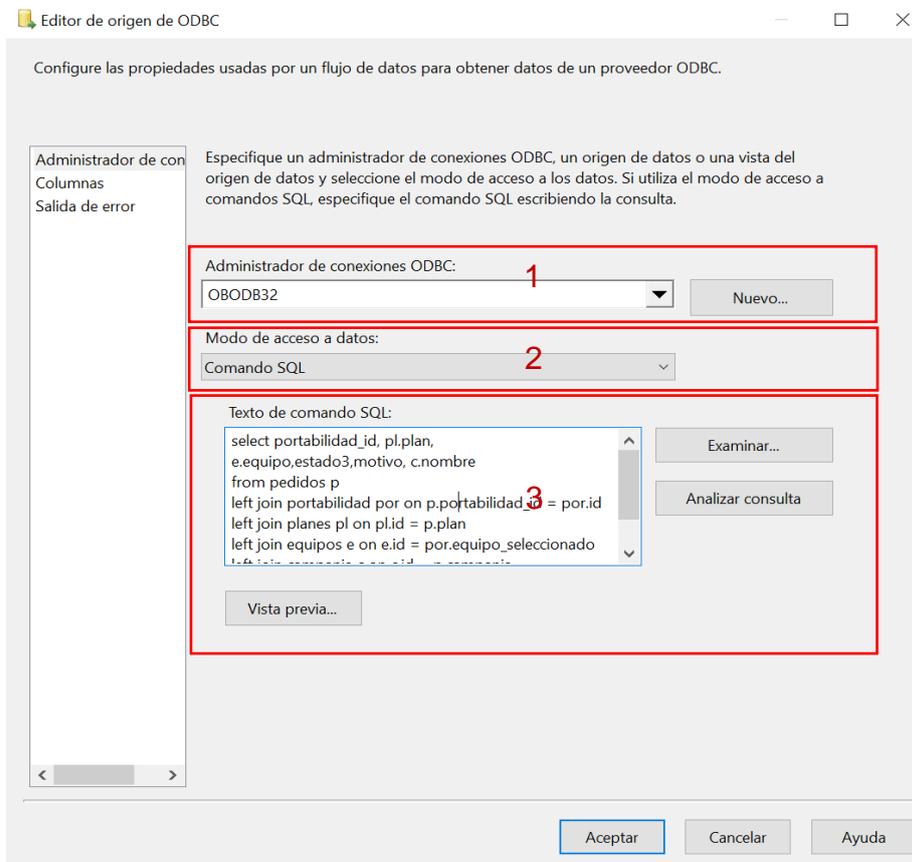


Figura 41. Editor de Origen ODBC.

1. Administrador de conexiones ODBC, se selecciona la conexión a la base de datos detallada en el punto 5.3
2. Método de acceso a datos, existen dos formas de acceder a los datos, directamente a la tabla o mediante un script sql.
3. Texto de comando SQL, si en el anterior punto se seleccionó la opción de acceder a datos mediante SQL. En este apartado se ingresa la sentencia SQL como se puede observar en la figura 42.

Continuando, en este mismo editor existe la manera de poder analizar la correcta sintaxis de la consulta SQL ingresar. Tal y como se puede apreciar en la figura 43 analizar la consulta la interfaz brinda un error de sintaxis. Sin embargo, cuando la sintaxis entregará un mensaje de éxito, esto se puede apreciar en la figura 44.

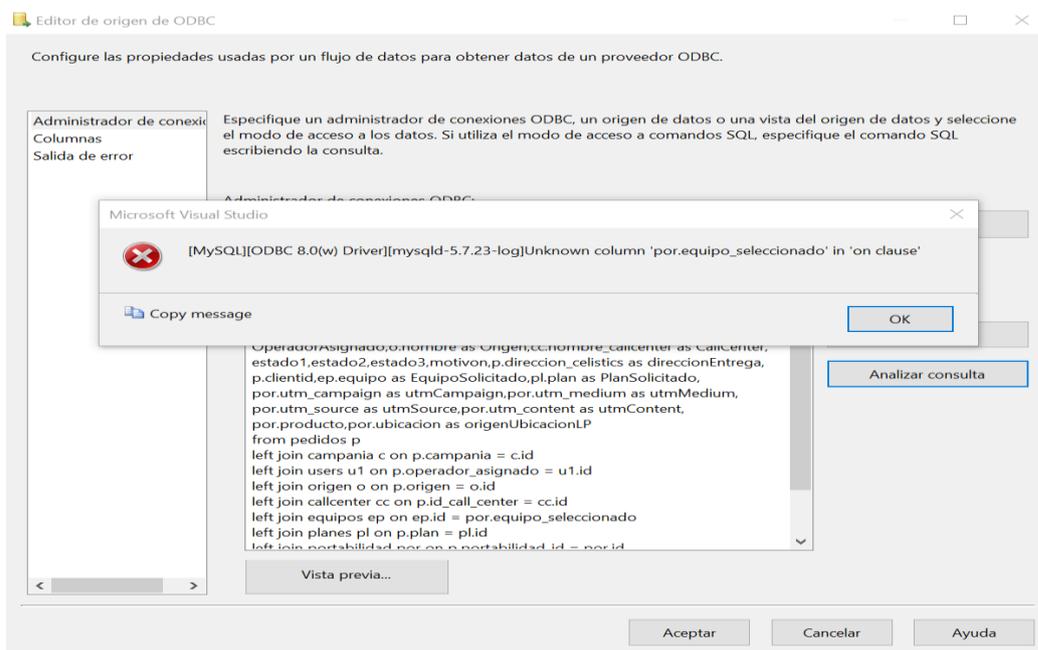


Figura 42. Error Consulta SQL.

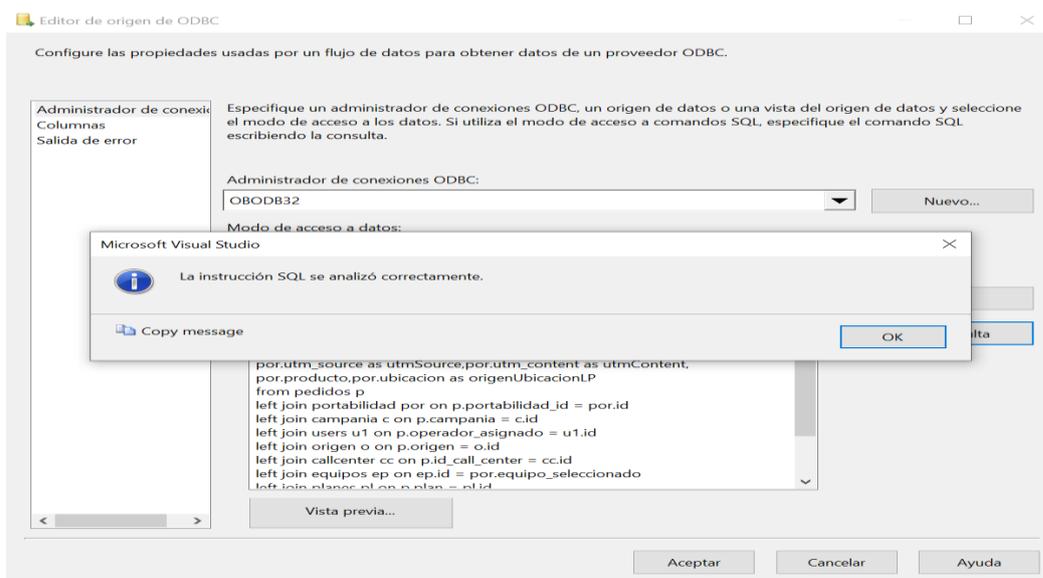


Figura 43. Consulta SQL Correcta.

Finalmente, en la sección de columnas se puede apreciar las entradas y salidas que entregará este elemento del ETL, tal y como se puede apreciar en la figura 44.

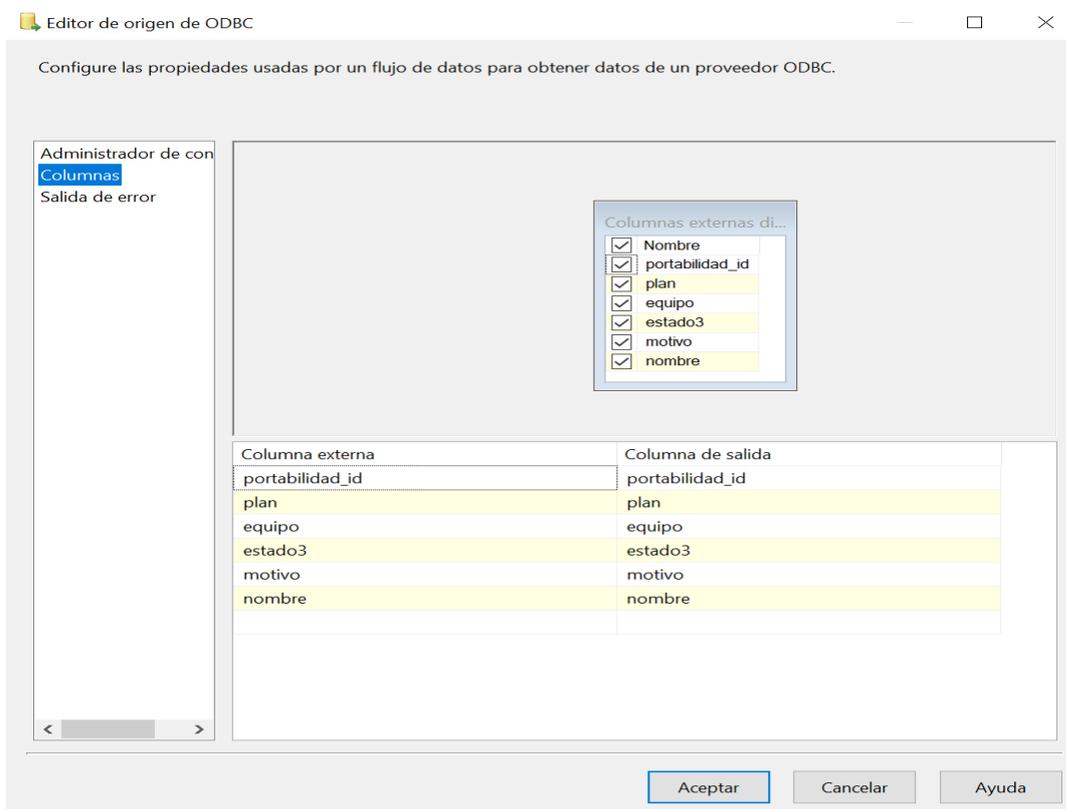


Figura 44. Columnas Editor ODBC.

5.4.2 Conversión de Datos

En parte del flujo básicamente se transforman los tipos de datos origen y para que puedan entrar en los campos definidos en el DW. Es decir, si desde la base de datos relacional, se envía el campo `client_id` como string, en esta parte se podrá transformar el mismo a un numeric.

En la figura 45 se puede apreciar el “Editor de transformación conversión de datos”, donde se realiza lo detallado en el párrafo anterior.

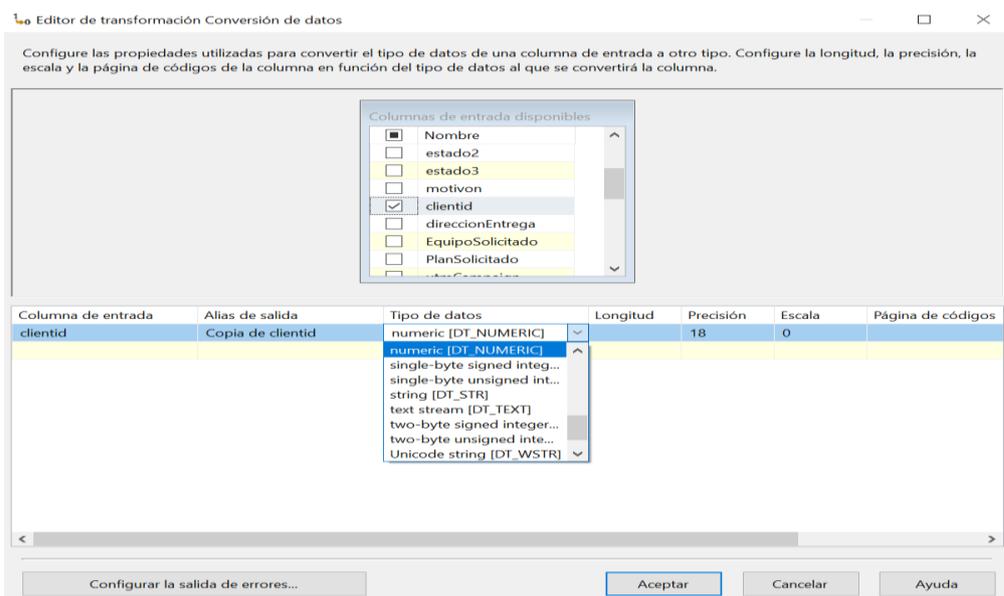


Figura 45. Editor de transformación conversión de datos.

5.4.3 Columna Derivada

Este componente que sirve para la generación de la fecha de auditoria en el ETL. Como se puede apreciar en la figura 46, únicamente se debe nombrar la columna y con una función de tiempo crear la fecha. En este se define el tipo de dato, la expresión y esta columna será la que llenará el campo de autoría en cada tabla de dimensión y hechos.

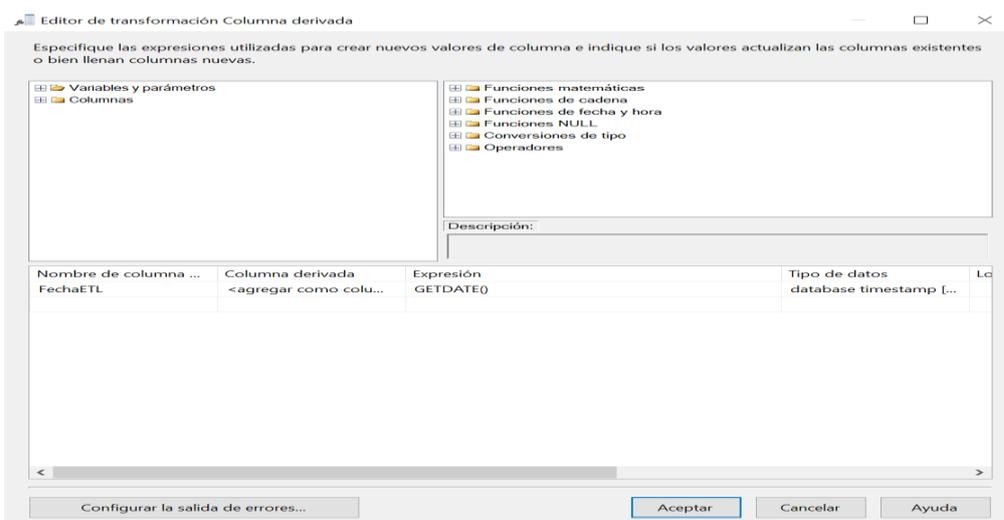


Figura 46. Creación columna derivada de tiempo.

5.4.4 Componente de Script

Este sirve para crear el id incremental en cada tabla de dimensión o tiempo. Como se puede apreciar en la figura 47, se debe definir una columna de salida.

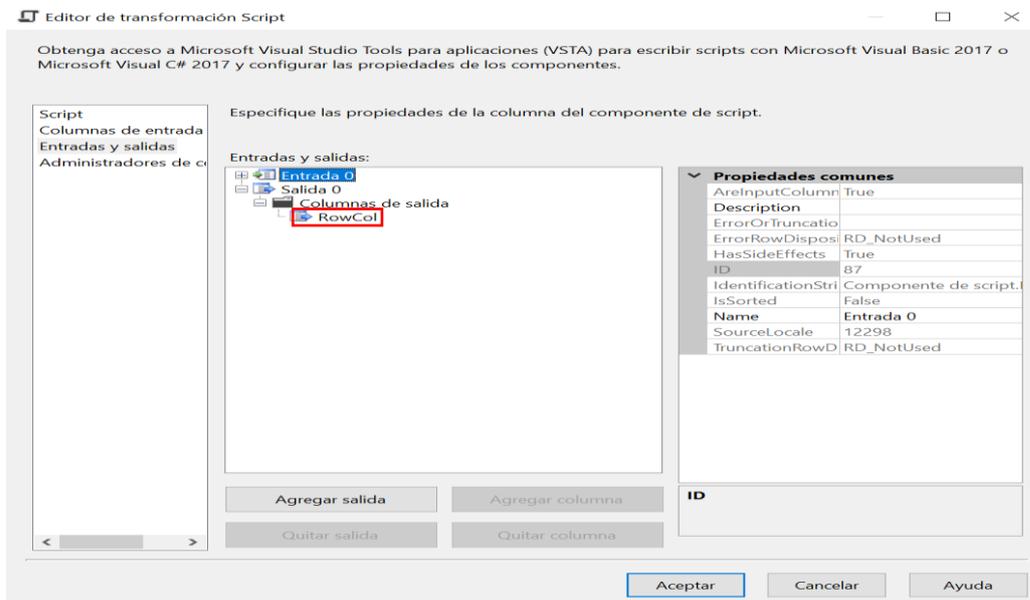


Figura 47. Editor de Transformación Columna de salida.

Al editar Script se debe ingresar la siguiente porción de código, la misma que se evidencia en la figura 48.

```

100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
public override void PostExecute()
{
    base.PostExecute();
    /*
     * Add your code here
     */
}

/// <summary>
/// This method is called once for every row that passes through the component from Entrada0.
///
/// Example of reading a value from a column in the the row:
/// string zipCode = Row.ZipCode
///
/// Example of writing a value to a column in the row:
/// Row.ZipCode = zipCode
/// </summary>
/// <param name="Row">The row that is currently passing through the component</param>
int row = 1;
public override void Entrada0_ProcessInputRow(Entrada0Buffer Row)
{
    Row.RowCol = row;
    row = row + 1;
}

```

Figura 48. Script Id incremental.

5.4.5 OLE DB Destino

En el extremo del ETL, se tiene el destino donde se hará la conexión para conectar los datos origen con la estructura del DW. En la figura 49 se pueden observar las configuraciones de este elemento, mismas que se las detallará a continuación.

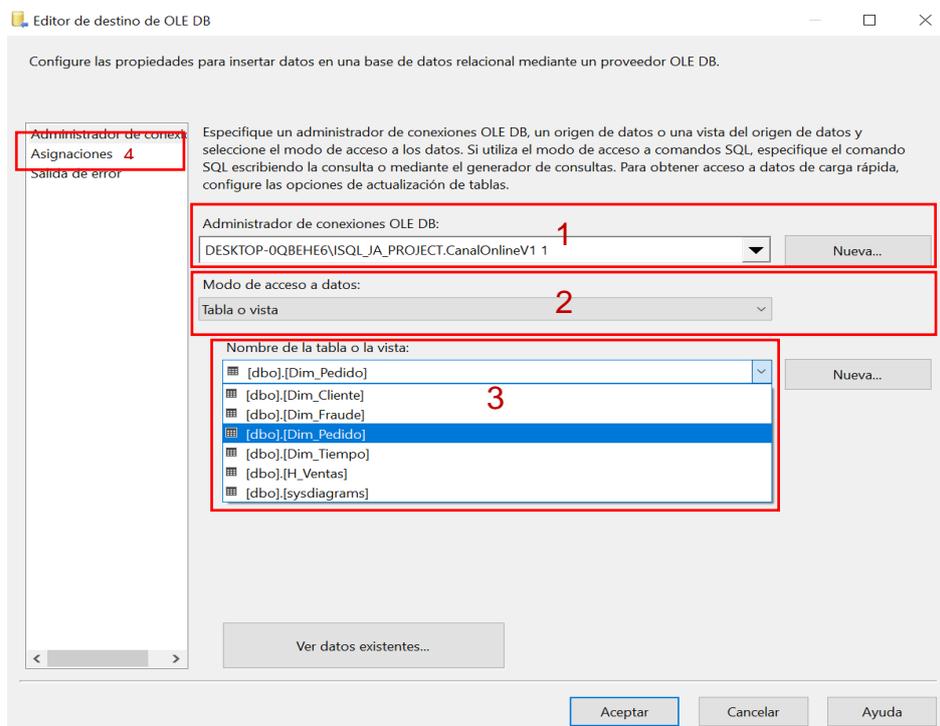


Figura 49. Editor de destino OLE DB.

1. Administrador de conexiones OLE DB, se establece la conexión a la base dimensional del DW.
2. Método de acceso a datos, método para acceder a la tabla específica donde se almacenarán los datos.
3. Nombre de la tabla o vista, se define el nombre la tabla en la que llegan los datos.
4. Asignaciones, como se puede apreciar en la siguiente figura, en esta sección se definen automáticamente el vínculo entre los campos de las tablas origen – destino.

A continuación, se deberán asignar las columnas de entrada para. Esto se realiza en la tab “Asignaciones”, únicamente se debe seleccionar el destino y guardar la configuración, tal y como se puede observar en la figura 50.

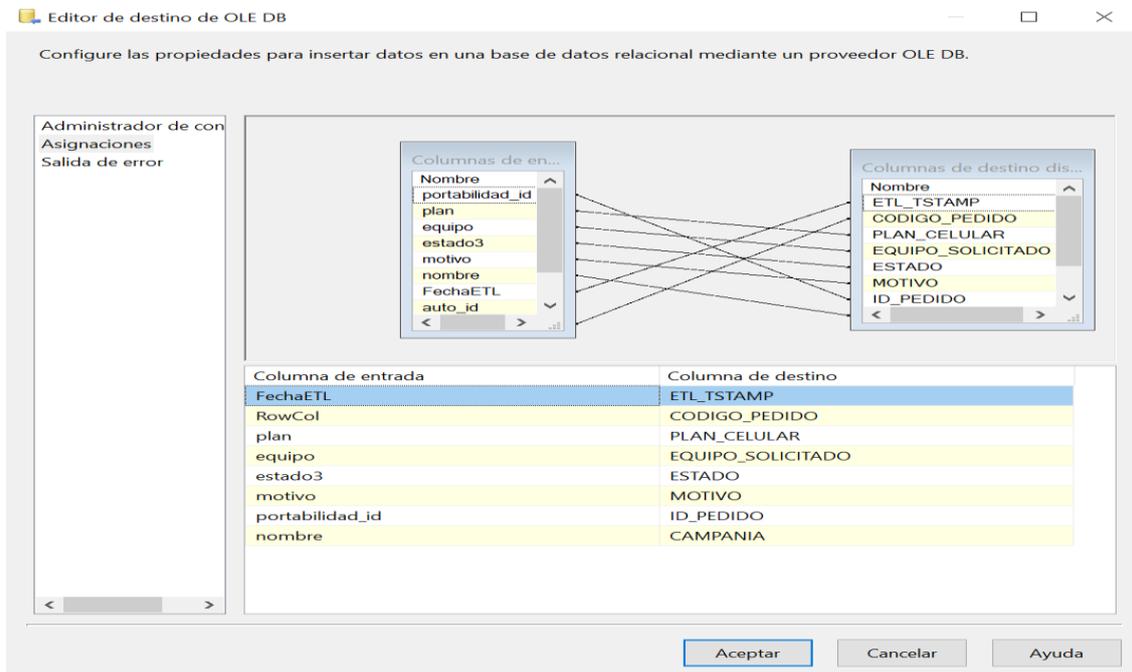


Figura 50. Asignación de columnas Editor destino OLE DB.

Finalmente, al ejecutar el paquete se puede observar en la figura 51 que cada tarea del flujo de datos se ejecuta correctamente.

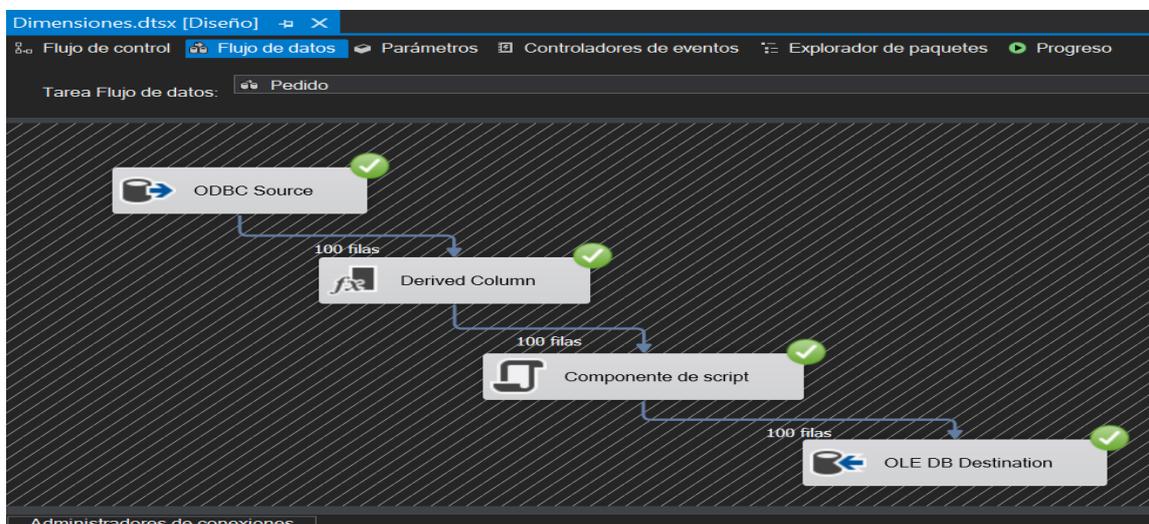


Figura 51. Ejecución paquete ETL.

Este proceso se realizará para poblar cada tabla de dimensión, todo en un solo paquete de SSIS como se puede apreciar en la figura 52.

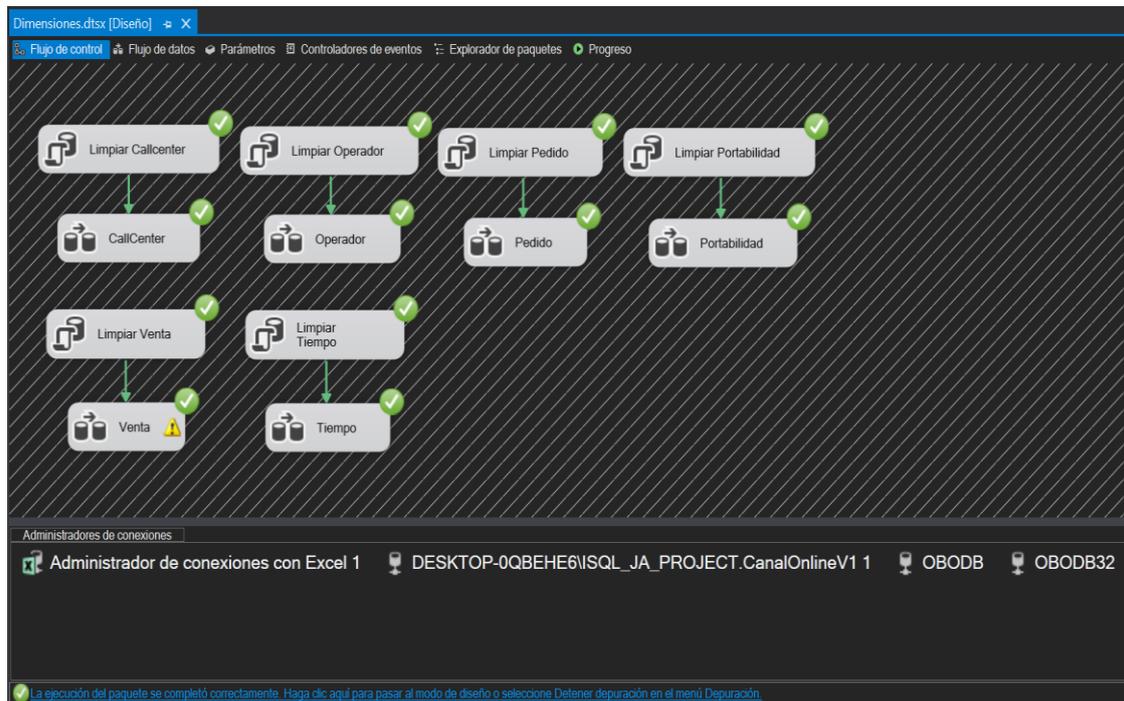


Figura 52. Ejecución ETL Dimensiones.

Respecto a las tablas de hechos, se deben realizar varios cruces con tablas de la base relacional y las dimensiones creadas previamente para obtener el ID y relacionarlas lógicamente. En la figura 53 se puede apreciar la ejecución para poblar la tabla Hecho Ventas. Mientras que en la figura 54 se puede evidenciar la ejecución de la tabla de Hecho Pedido.

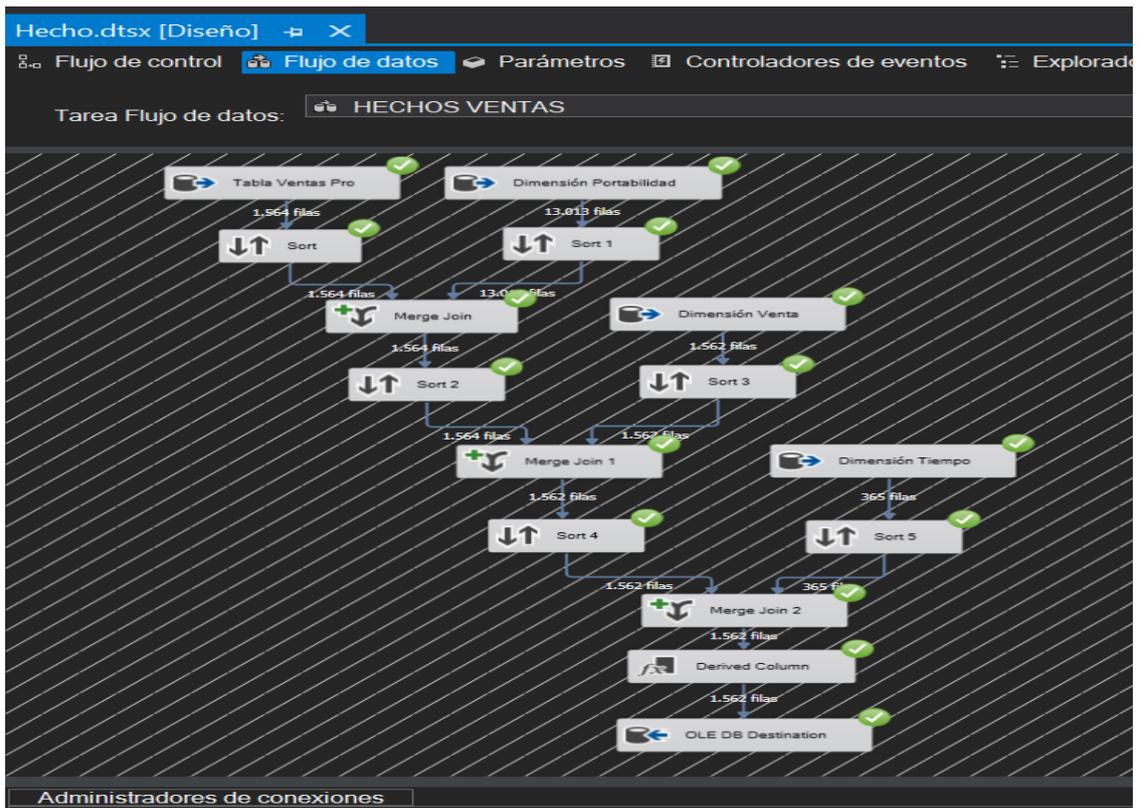


Figura 53. Ejecución ETL Hecho Venta.

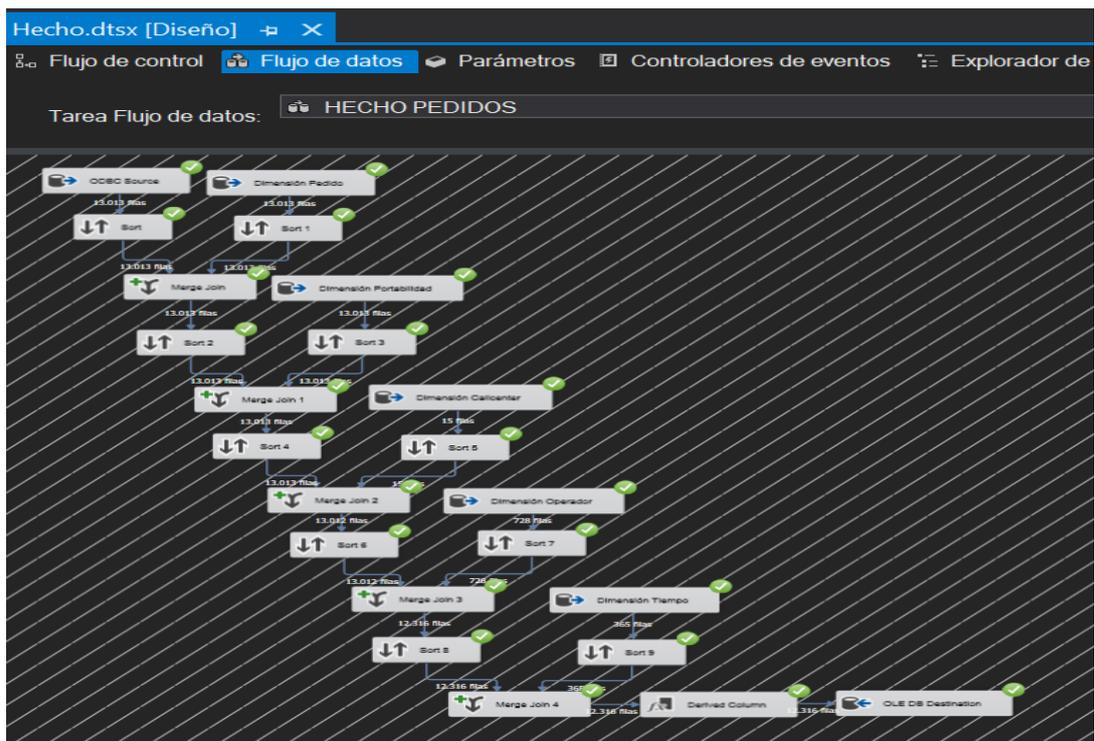


Figura 54. Ejecución ETL Hecho Pedido.

5.5 Desarrollo cubo OLAP

La creación de uno o varios cubos, se realiza dependiendo de las necesidades del negocio y del escenario de datos a analizar. Ante esto, se utilizará la herramienta Visual Studio con la extensión SSAS para creación de estos.

Con la extensión configurada correctamente se crea un nuevo proyecto de Servicio de Análisis, tal y como se puede ver en la figura 55.

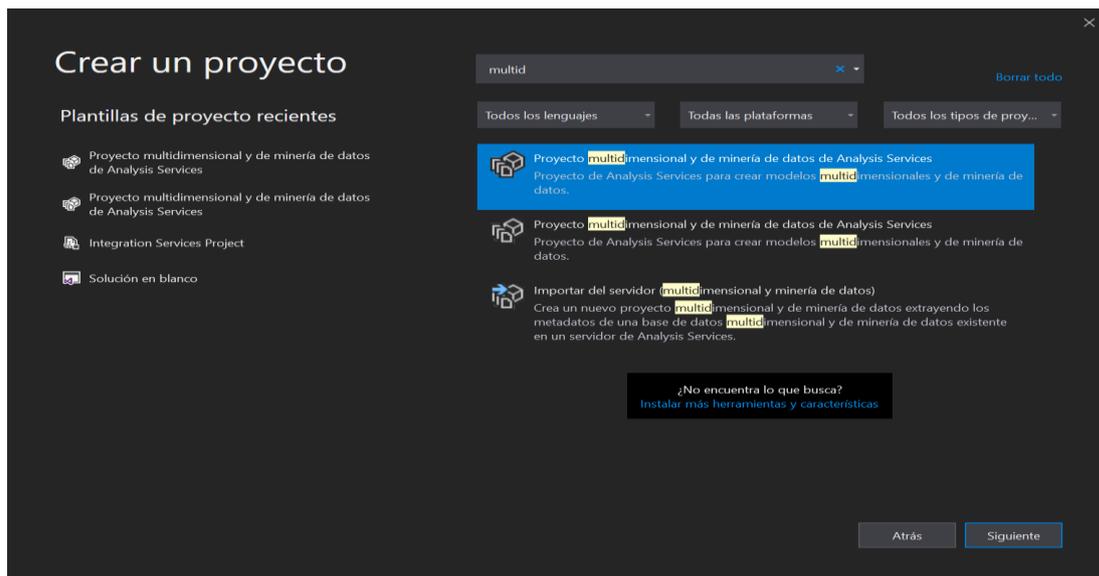


Figura 55. Creación Proyecto Analysis Services.

Al crear el proyecto en el explorador se puede visualizar el siguiente directorio, donde se configurarán algunos elementos para poder acceder a los datos. Esto se lo aprecia en la figura 56.

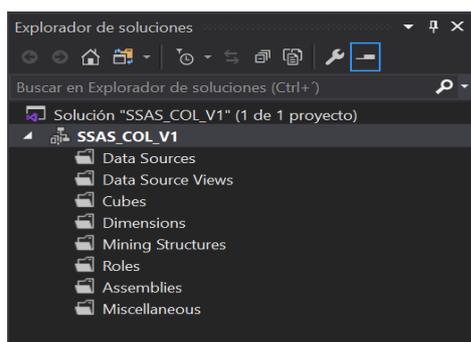


Figura 56. Explorar de Soluciones.

A continuación, se crea una Fuente de Datos, que traerá todas las tablas de Dimensiones y Hechos que se crearon en la base dimensional. La conexión se evidencia en la figura 57.

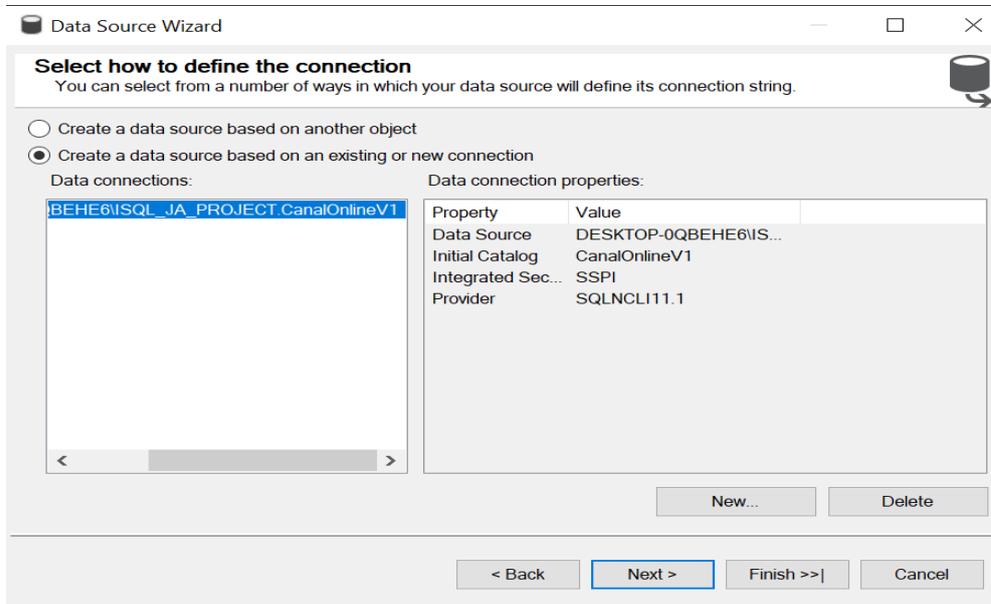


Figura 57. Configuración Fuente de Datos.

Al crear el paquete se crea la solución con el directorio de minería de datos, tal y como se puede ver en la figura 58.

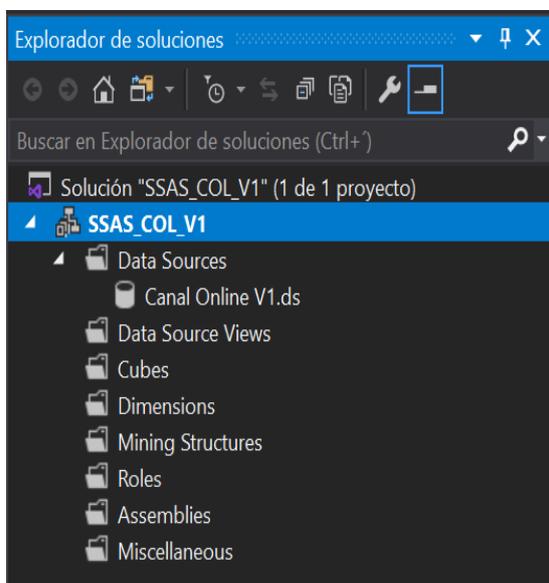


Figura 58. Explorador de soluciones Data Source.

Después se configura una vista de Fuente de datos, donde se añaden las tablas que se usarán en la creación del cubo o propiamente en la minería de datos. Tal y como se puede apreciar en la figura 59.

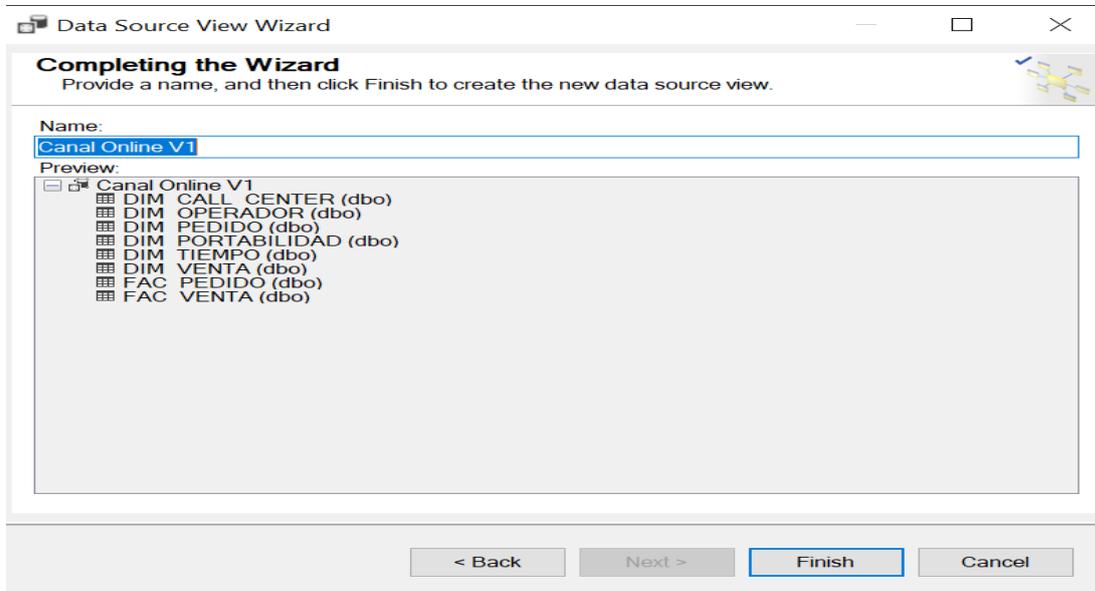


Figura 59. Vista de fuente de datos.

Al finalizar en la figura 60 se puede ver un esquema relacional con las tablas seleccionadas anteriormente.

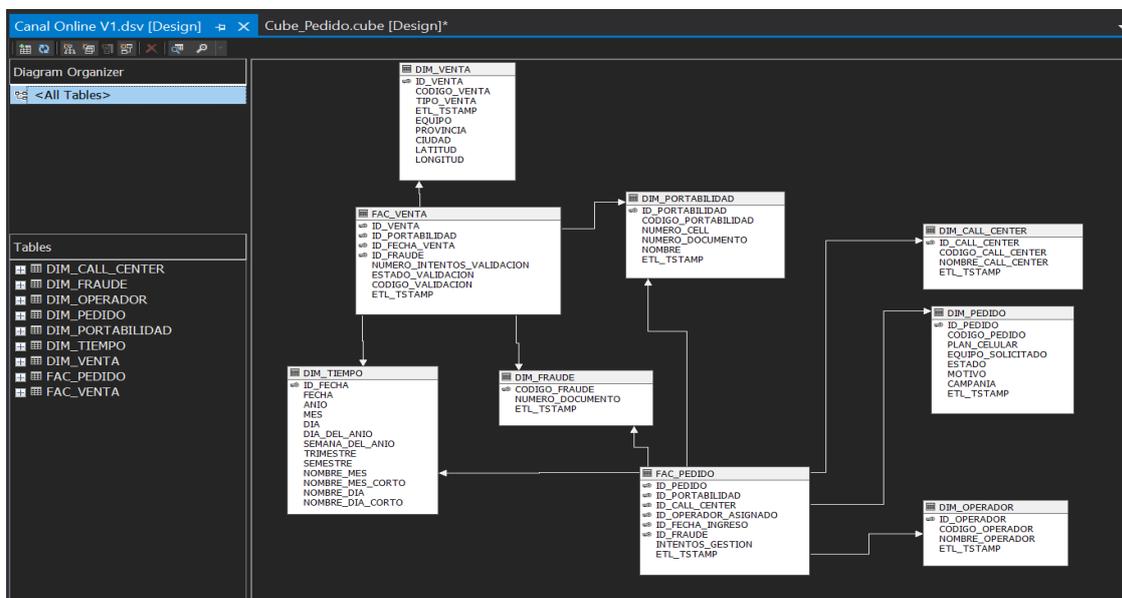


Figura 60. Esquema relacional Vista de fuente de datos.

Con los orígenes de datos cargados, se puede realizar un cubo. Cabe recalcar que la instalación de SQL server debe estar configurada en modo Analysis Services multidimensional para que esto funcione.

Inicialmente en la figura 61 se define la forma de creación del cubo, se puede crear usando tablas existentes o en blanco, en este caso el primero.

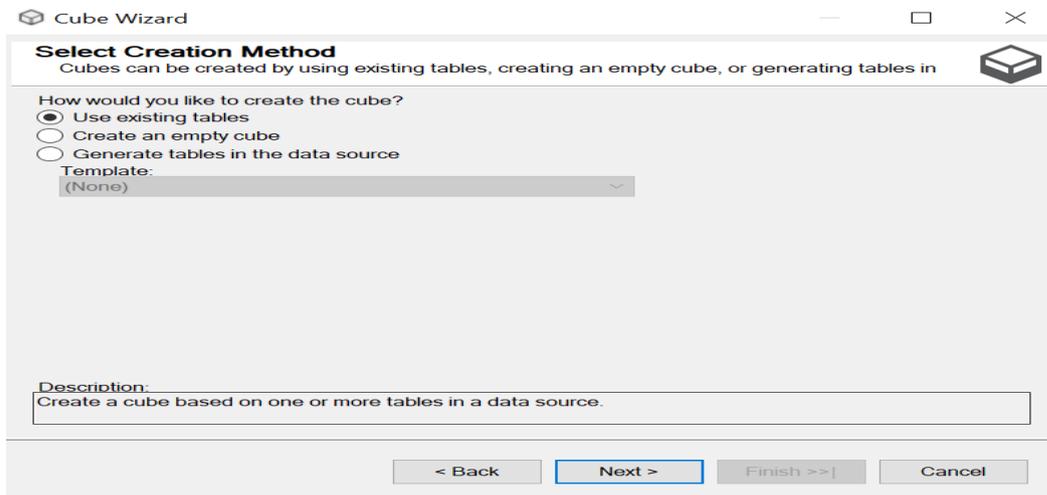


Figura 61. Selección Método Creación Cubo.

A continuación, como se puede ver en la figura 62, se seleccionan las tablas del grupo de medida en este caso Hechos Pedido.

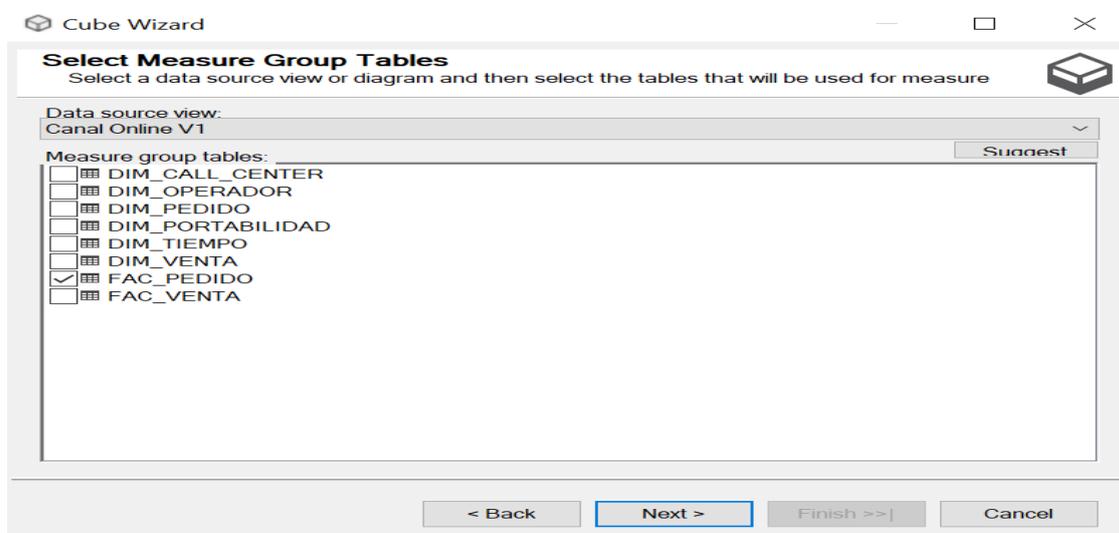


Figura 62. Tablas de grupo de medidas.

Después, en la figura 63 se puede ver el resumen con el grupo de medidas de la tabla de hecho y sus Dimensiones.

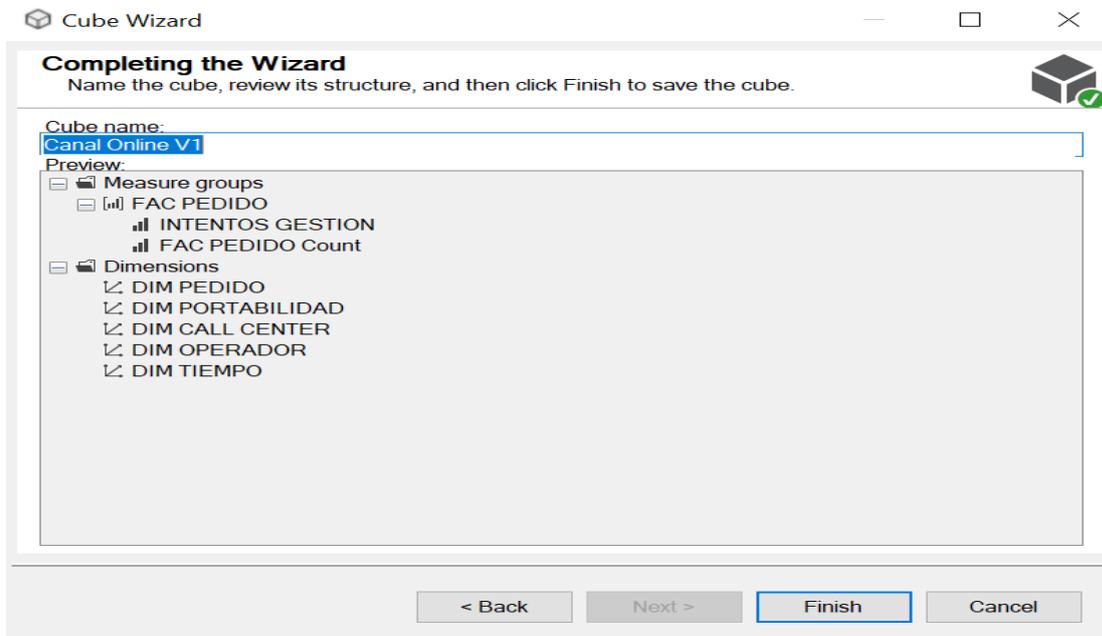


Figura 63. Resumen creación Cubo.

En la figura 64 se puede ver la tabla de hechos con sus respectivas Dimensiones lista para procesarse.

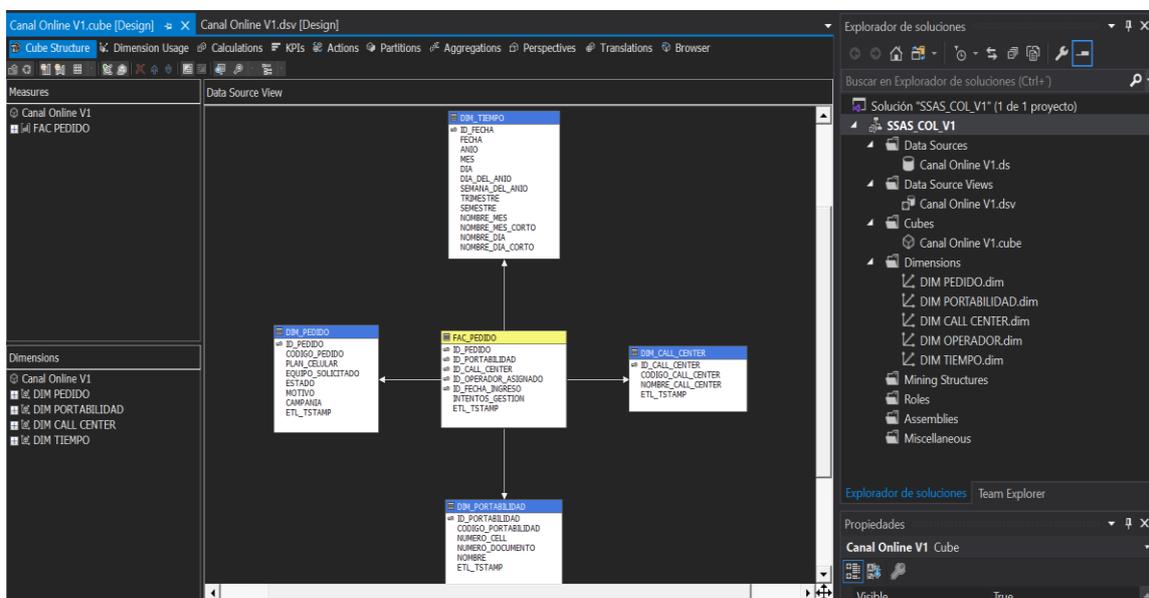


Figura 64. Cubo Canal Online.

Antes de procesar se debe editar cada una de las dimensiones para agregar los atributos que se utilizarán en el cubo tal y como se puede observar en la figura 65.

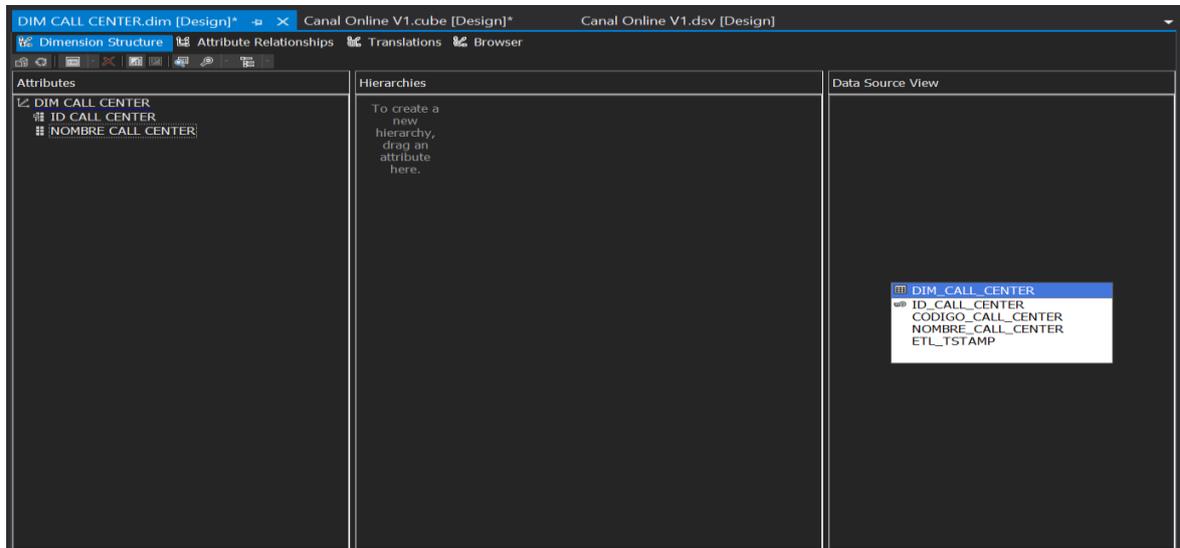


Figura 65. Atributos Dimensión Callcenter.

Finalmente, se procesa el cubo para implementarlo, para esto debe estar configurado el nombre de la instancia Multidimensional (DESKTOP-0QBEHE6\SQL_JA) en el proyecto, tal y como se puede ver en la figura 66 y 67.

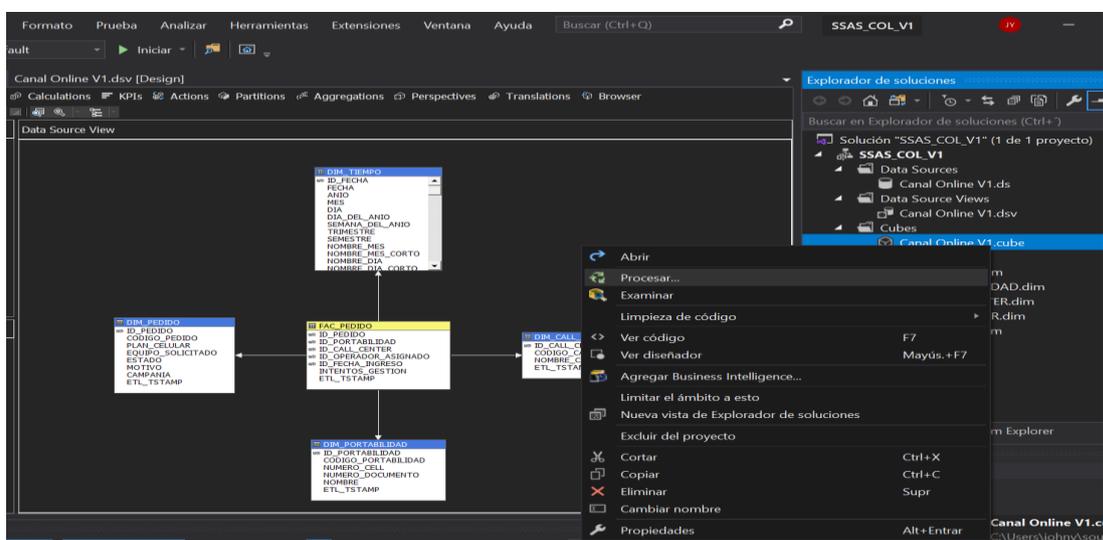


Figura 66. Ejecución del Cubo.

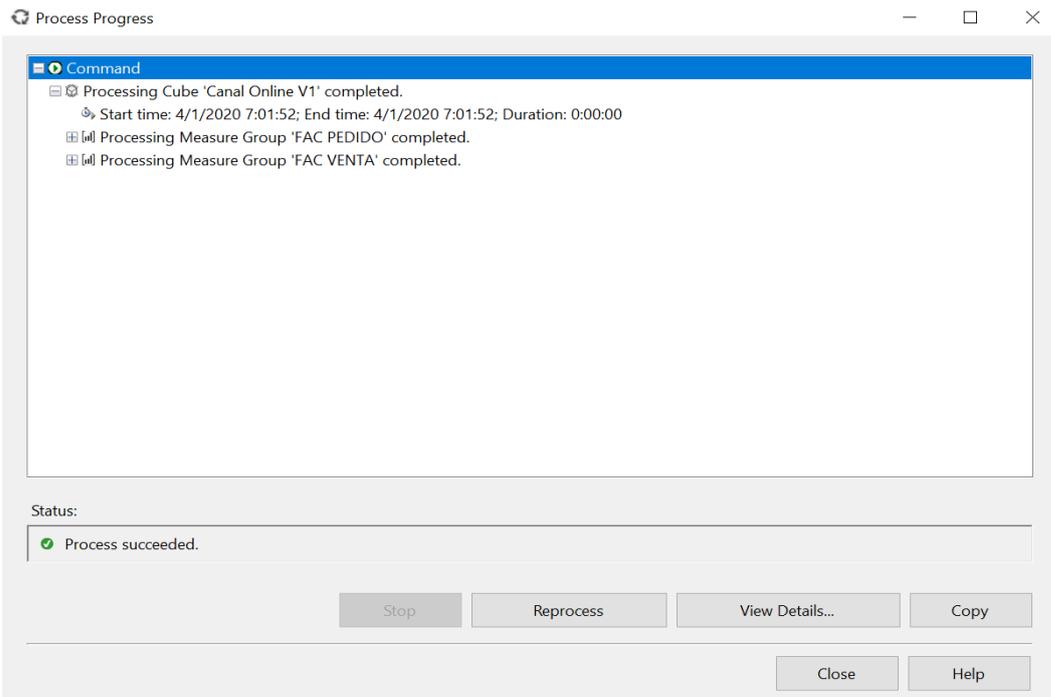


Figura 67. Cubo Procesado.

Después de procesar el cubo se obtienen sus medidas y distintas Dimensiones, al arrastrar medidas específicas de fraude se pueden ver los datos correspondientes. Tal y como se puede apreciar en la figura 68.

Dimension	Hierarchy	Operator	Filter Expression	Param...
<Select dimension>				
CIUDAD	CODIGO FRAUDE	TIPO VENTA	FAC VENTA Count	
COTOP...	332	Alta Pospago con equipo	1	
COTOP...	347	Portabilidad Pospago sin equipo	1	
EL ORO	681	Portabilidad Pospago sin equipo	1	
ESMER...	4455	Renovación	1	
GUAYAS	903	Renovación	1	
GUAYAS	943	Renovación	1	
GUAYAS	1166	Renovación	1	
GUAYAS	1624	Renovación	1	
GUAYAS	1734	Portabilidad Pospago sin equipo	1	
GUAYAS	2598	Renovación	1	
IMBAB...	2390	Renovación	1	
LOJA	2415	Renovación	1	
LOJA	2427	Upsell con equipo	1	
LOS RI...	2644	Alta Pospago sin equipo	1	
LOS RI...	2650	Renovación	1	
LOS RI...	2898	Renovación	1	
MANABI	2704	Renovación	1	
MANABI	2813	Renovación	1	

Figura 68. Interfaz de administración Cubo.

5.6 Minería de datos

La minería de datos se la realizará con SQL Server Analysis Services, se trabajará en la misma solución en la que fue creada el Cubo OLAP. Como se puede ver en la figura 69, al crear una nueva estructura de minería de datos se puede seleccionar entre un almacén de datos o un cubo existente. En este caso se usará el segundo.

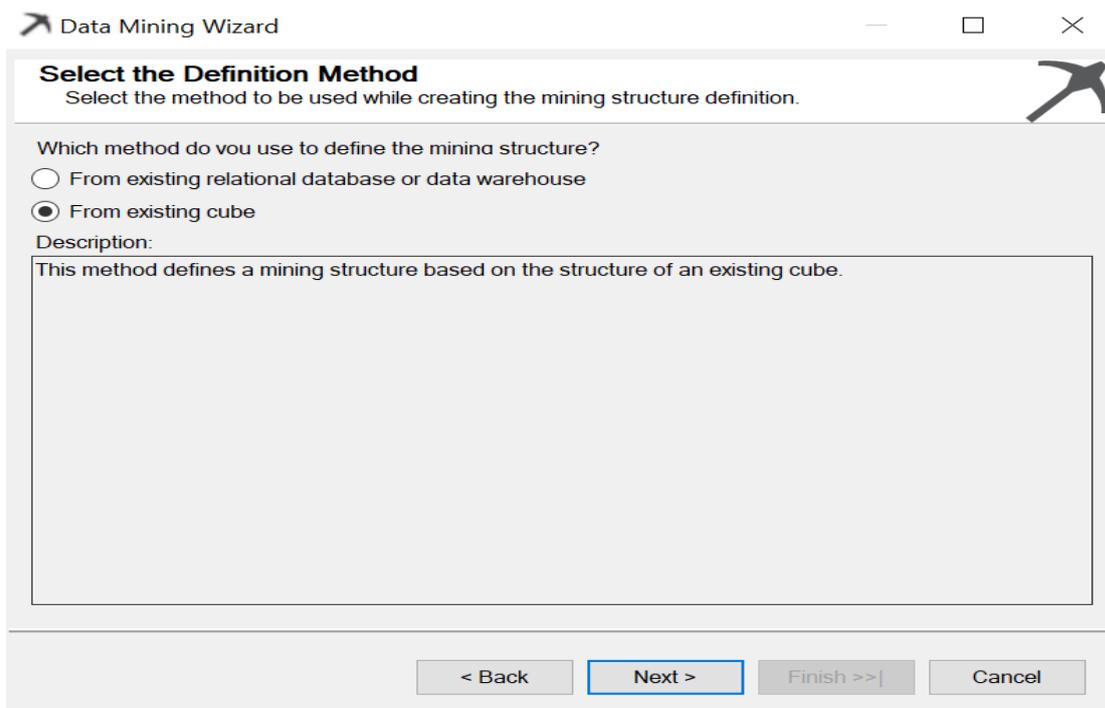


Figura 69. Método Estructura de minería de datos.

Después, se debe seleccionar el tipo de algoritmo de minería de datos que se utilizará. En la figura 70 se puede ver que Microsoft cuenta con varios algoritmos como (Árboles de decisión, Clústeres, entre otros). En este caso se utilizará el método de árboles de decisión para analizar fraude por ciudad.

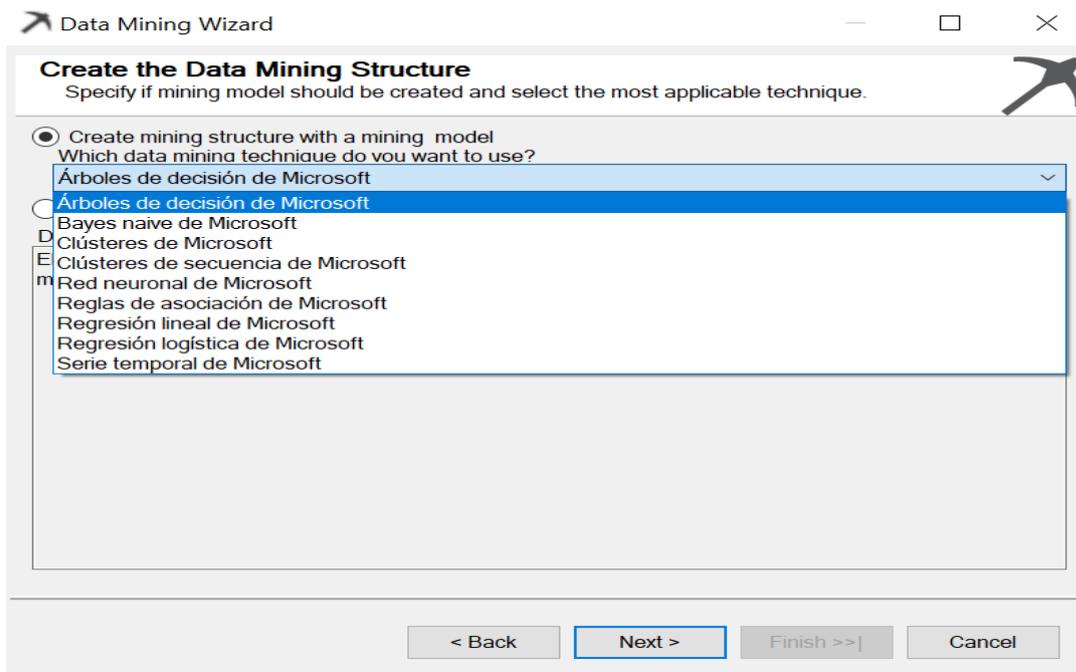


Figura 70. Algoritmos de minería de datos.

En la figura 71 se observa que se debe seleccionar las columnas para realizar la minería de datos, dando valores de entradas y una columna de predicción.

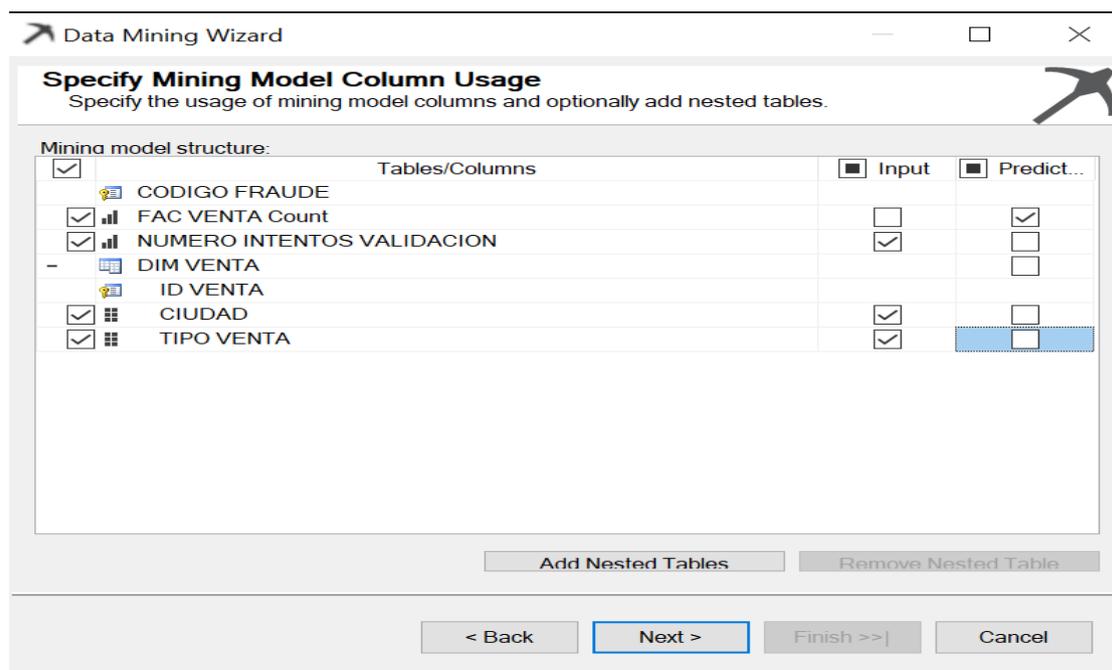


Figura 71. Columnas usadas en el modelo de minería de datos.

Finalmente, al procesar el algoritmo en la pestaña “Visor de minería de datos” de la figura 72 se puede observar gráficamente los patrones del modelo procesado.

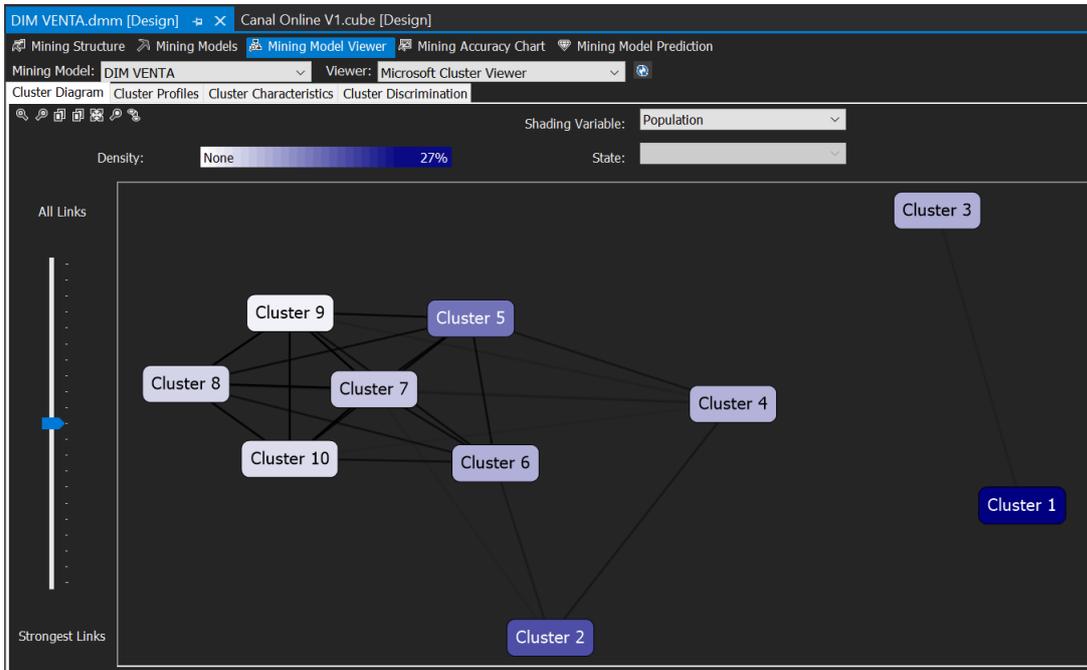


Figura 72. Visor de patrones del algoritmo de minería de datos.

5.7 Desarrollo de Tablero de indicadores

Para el desarrollo de indicadores se utilizará Power BI Desktop, una herramienta muy intuitiva y con componentes muy útiles para el desarrollo de tableros gerenciales. Basta con tener la conexión a la información y se empieza a diseñar la interfaz de manera sencilla.

5.7.1 Conexiones Orígenes de Datos

En Power BI se puede obtener datos desde varios orígenes. En la figura 73 se pueden observar algunos de los conectores disponibles en esta herramienta.

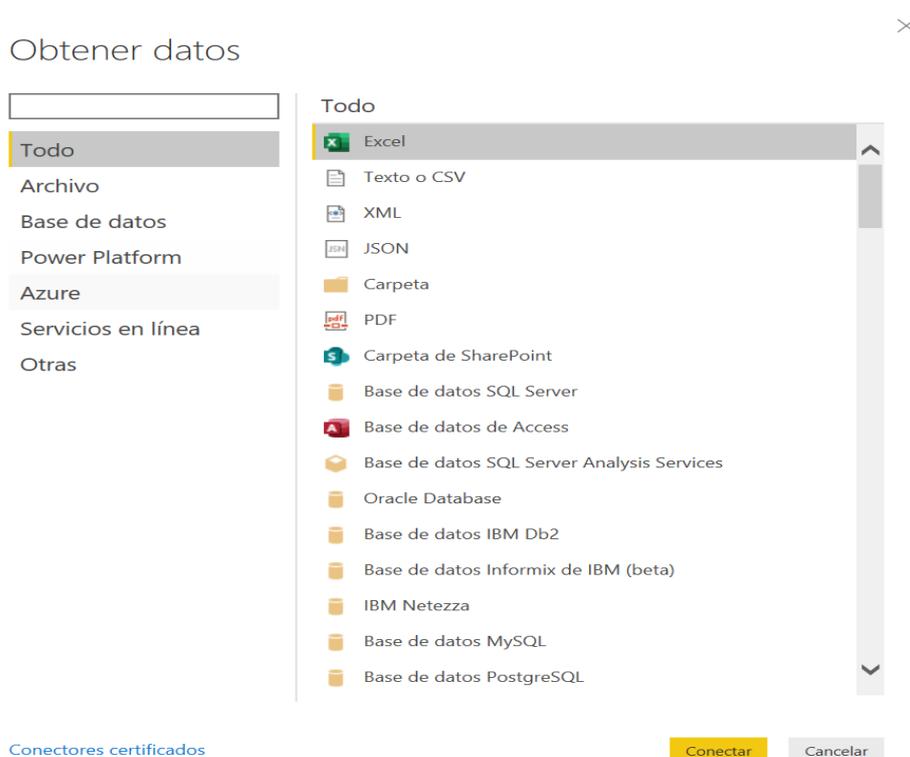


Figura 73. Tipo de orígenes de datos en Power BI.

Para el primer reporte se realizará una conexión al cubo OLAP creado anteriormente. Esta será mediante SQL Server Analysis Services, como se puede ver en la figura 74 se especifica el nombre del servidor.



Figura 74. Conexión Analysis Services.

Finalmente, con las dimensiones cargadas se puede desarrollar el tablero de Indicadores a medida.

5.7.2 Reporte de ventas con fraude

Este reporte se basa en el análisis de ventas fraudulentas del mes de diciembre, en la figura 75 se puede apreciar un tablero específico para KPIS de fraude.

En este se observa que el día con mayor índice de fraude es el lunes, las ventas fraudulentas se ejecutan en Quito – Guayaquil. El producto que mayor se ve afectado son las “Renovaciones” porque implica un equipo de por medio. El promedio de intentos de validación es de 13 en Renovaciones sin embargo se nota claramente que no es suficiente para mitigar el fraude.

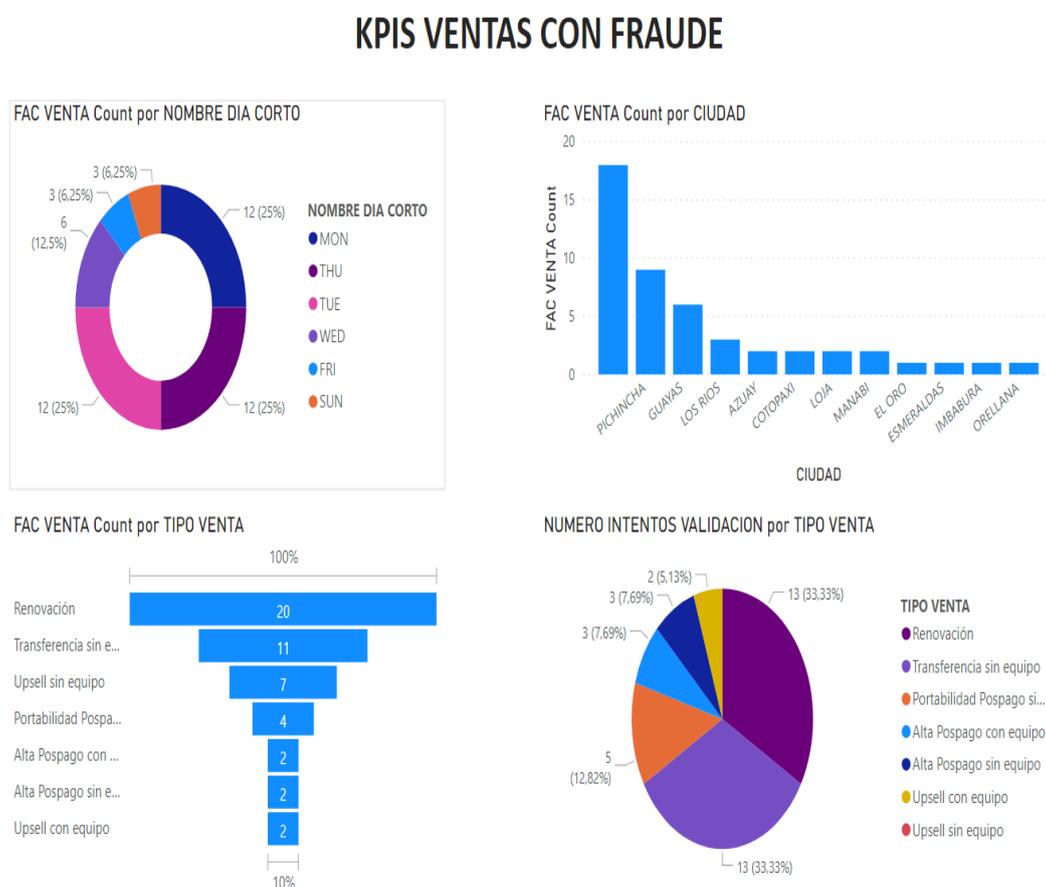


Figura 75. Tablero Indicadores Ventas con Fraude.

5.7.3 Reporte intentos de compra con fraude

Este reporte se basa en solicitudes ingresadas para una posible compra de un producto, es decir no necesariamente termina en una venta. Por lo tanto, en la figura 76 se puede ver un tablero con indicadores de pedidos con fraude. Cantidad de pedidos por cliente, cantidad de pedidos por campaña y en qué estado terminan estos.

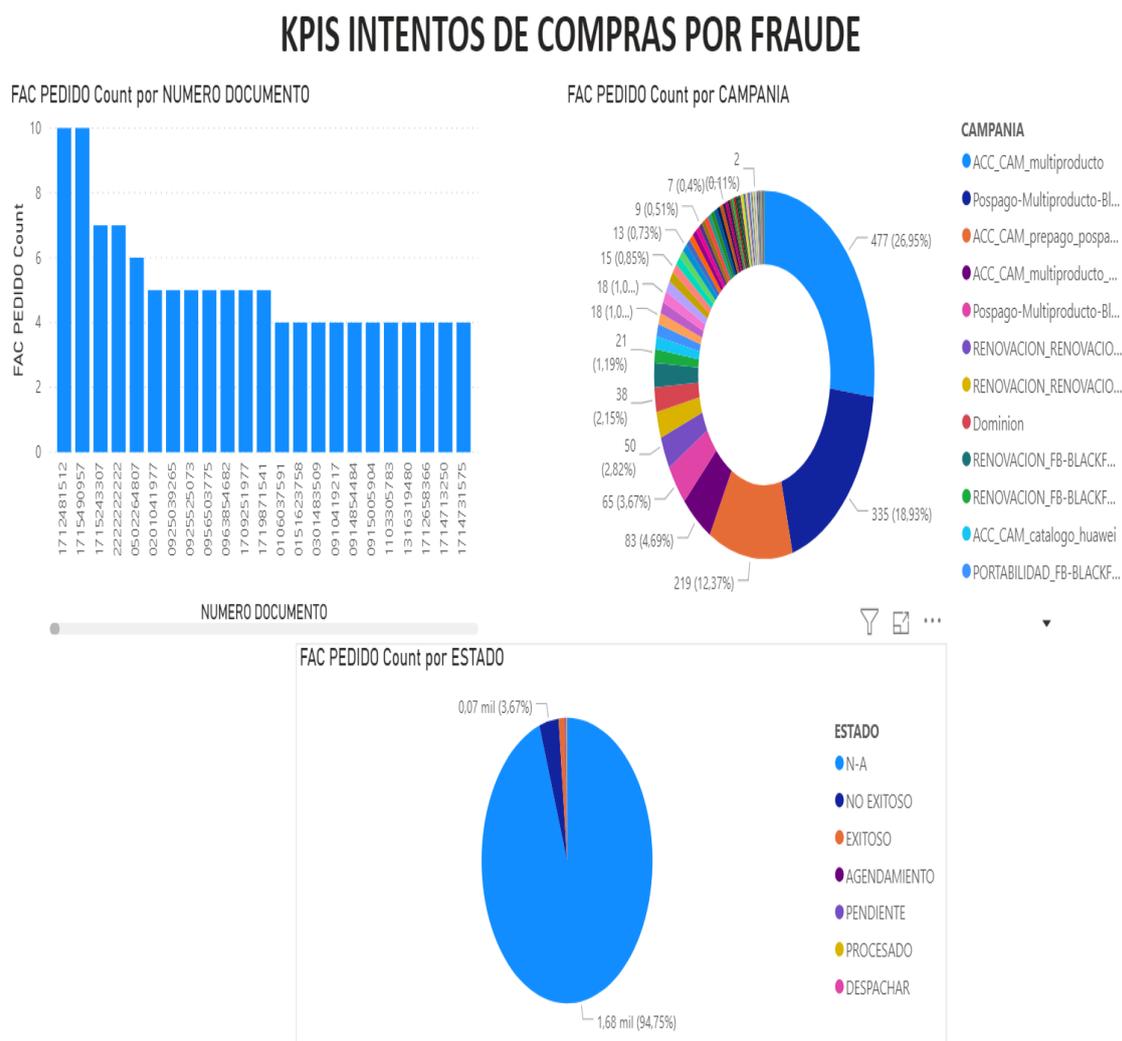


Figura 76. Tablero Indicadores Pedidos con Fraude.

5.8 Análisis de resultados del caso desarrollado

En este apartado se analizarán los indicadores propuestos en cada tablero, todo relacionado con fraude que es el foco de este prototipo de Inteligencia del negocio.

5.8.1 Reportes

El primer reporte se trata de indicadores de ventas que han cumplido todo el flujo, es decir ventas exitosas. Sin embargo, estas fueron reportadas como fraude siendo una pérdida para la compañía.

Dentro de este reporte el primer indicador es fraude por día. En la figura 77 se evidencia que el mayor porcentaje se encuentra en los tres primeros días de la semana. El lunes es el día con mayor índice por lo que deberá analizar los controles de fraude que se usan.

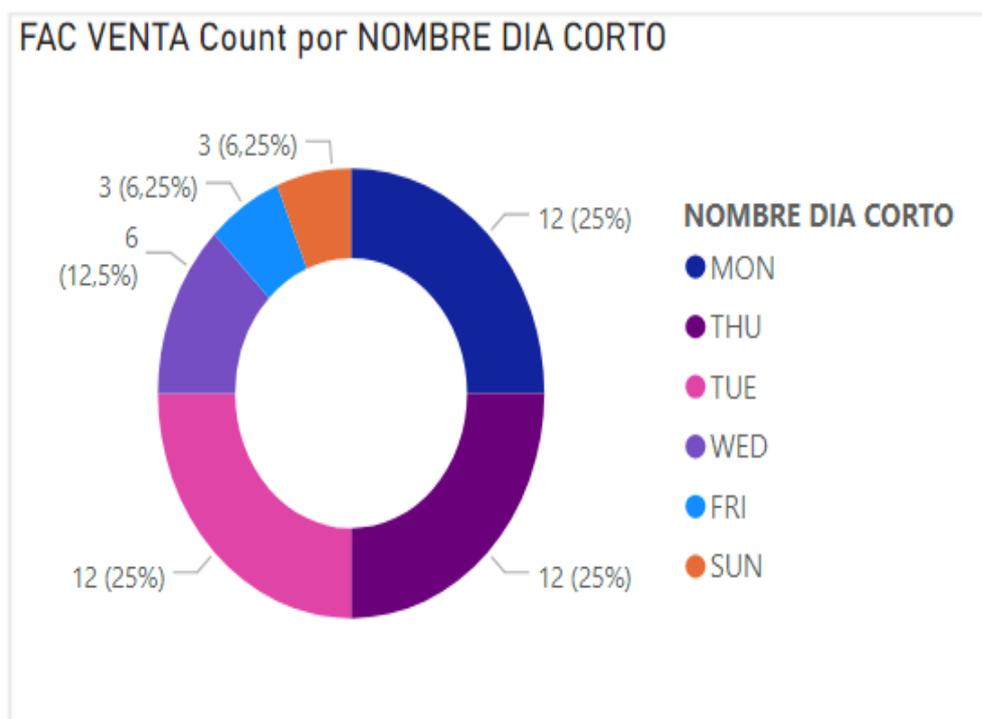


Figura 77. Indicador de Anillo Fraude por día.

El segundo indicador se basa en fraude por provincia, como se puede ver en la figura 78 las provincias que tienen mayor porcentaje de fraude es Pichincha y Guayas. Aunque este valor depende del volumen de ventas por cada provincia no quita la posibilidad de que es uno de los más importante para plantear controles. Esto debido a que se puede analizar la data de cada cliente respecto a direcciones ejemplo: un cliente que vive en Quito tiene menos probabilidad de que reciba su pedido en otra Ciudad.

FAC VENTA Count por CIUDAD

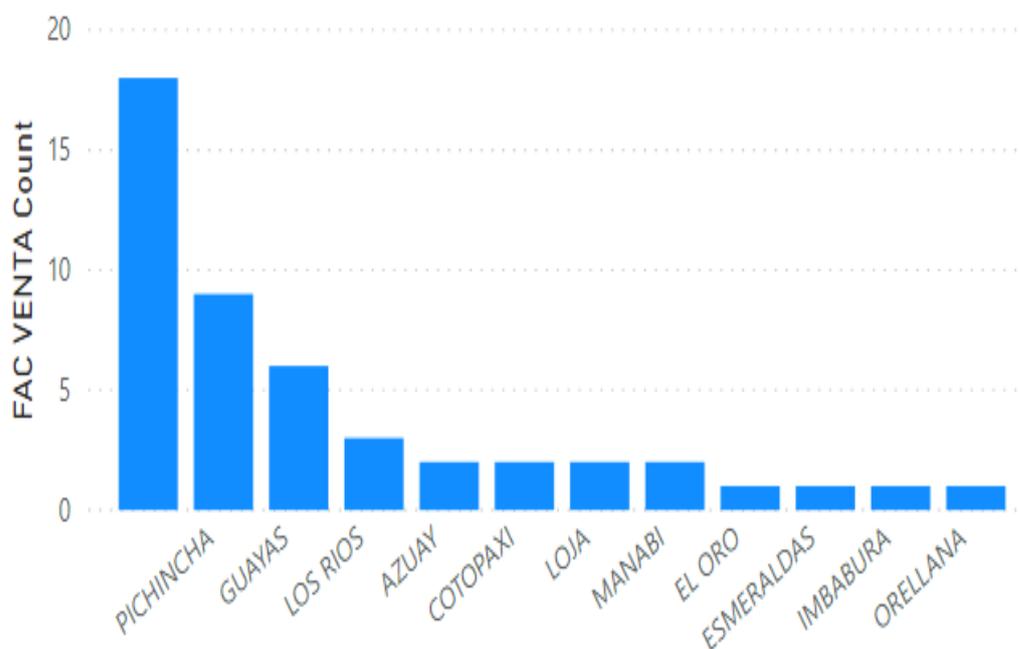
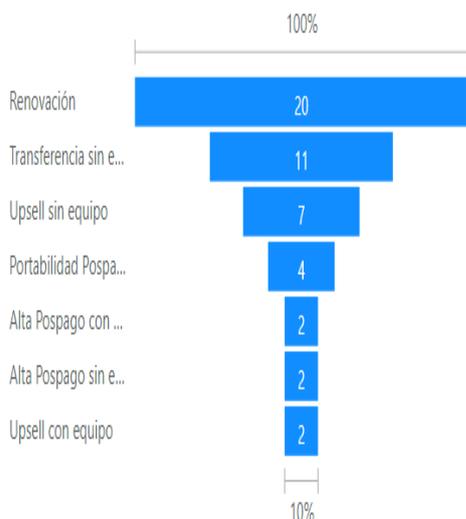


Figura 78. Indicador de Barras Fraude por Provincia.

Para cerrar el detalle del tablero de ventas, se analizan los productos con mayor índice de fraude. En la figura 79 se evidenciar el producto que encabeza la lista es Renovaciones ya que al tener un equipo de por medio tiende a ser buen objetivo. Así mismo, en el gráfico de pastel se detallan los intentos de validación de identidad por producto, este claramente indica que los controles de fraude que existen actualmente no están dando ningún resultado.

FAC VENTA Count por TIPO VENTA



NUMERO INTENTOS VALIDACION por TIPO VENTA

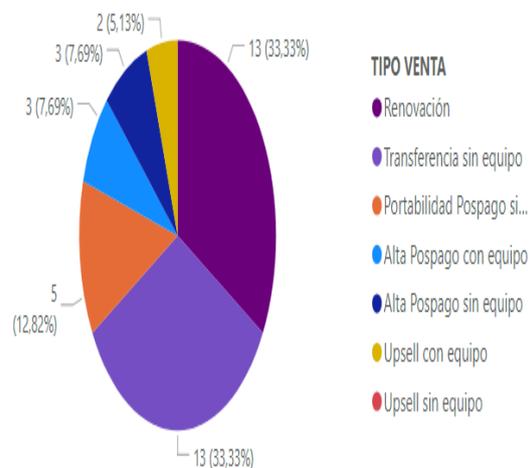


Figura 79. Indicadores de fraude por producto.

Por otro lado, el segundo reporte se basa en el análisis de indicadores de pedidos ingresados con fraude. Estos permiten tener una mejor visión del origen de este tipo de solicitudes.

Inicialmente, en la figura 80 se puede apreciar una gráfica de barras con los intentos de compra por cada cliente. Claramente se puede ver que se ingresan más de 3 solicitudes por cliente. En este caso al ser una venta asistida se puede levantar un control de duplicados por ciertas variables como celular para mitigar esta problemática.

FAC PEDIDO Count por NUMERO DOCUMENTO

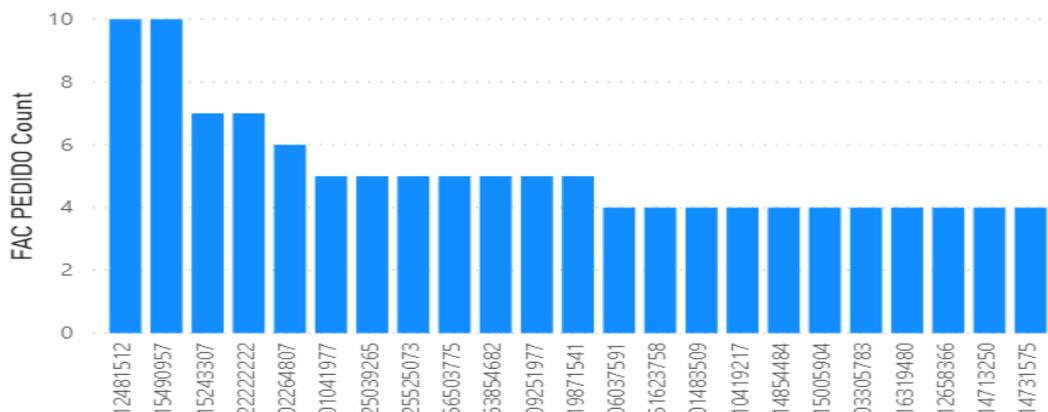


Figura 80. Grafica de barras cantidad de pedidos por cliente.

Otro indicador importante es la cantidad de pedidos ingresados por campaña o por página web. En la figura 81 se puede ver que ingresan pedidos con intentos de fraude desde varias campañas. Sin embargo, la campaña que más porcentaje de pedidos tiene es la “multiproducto”. Esto indica que la pauta utilizada para la misma deberá tener un primer filtro de clientes fraudulentos antes de enviar la publicidad.

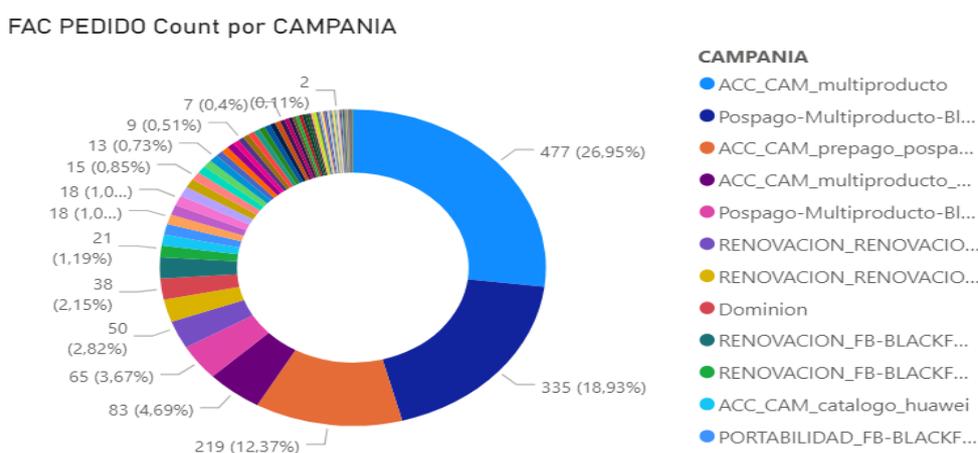


Figura 81. Gráfico de Anillo cantidad de pedidos por campaña.

Finalmente, la figura 82 muestra un gráfico de pastel donde se puede ver en qué estado termina cada pedido. En este claramente se puede ver en cuanto a pedidos brutos de fraude se convierte un 1.1% a ventas exitosas. Es un indicador muy importante ya que se puede levantar una funnel de fraude e ir implementando controles para disminuir este indicador.

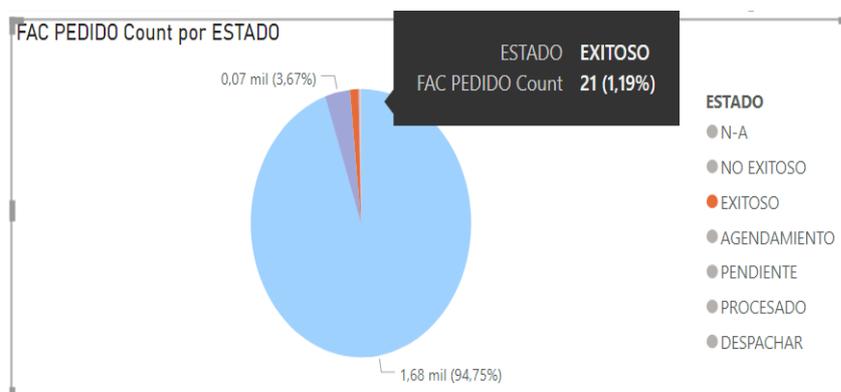


Figura 82. Gráfico de pastel estados por pedidos.

5.8.2 Minería de datos

Al ejecutar los algoritmos de minería de datos, se encontraron los siguientes resultados. En la figura 83 se puede ver gráficamente el algoritmo de agrupación o clustering, este representa presenta agrupaciones por ventas fraudulentas. Donde el cluster1 con mayor porcentaje de fraude se puede ver que es en Pichincha, seguido de Guayas cluster3.

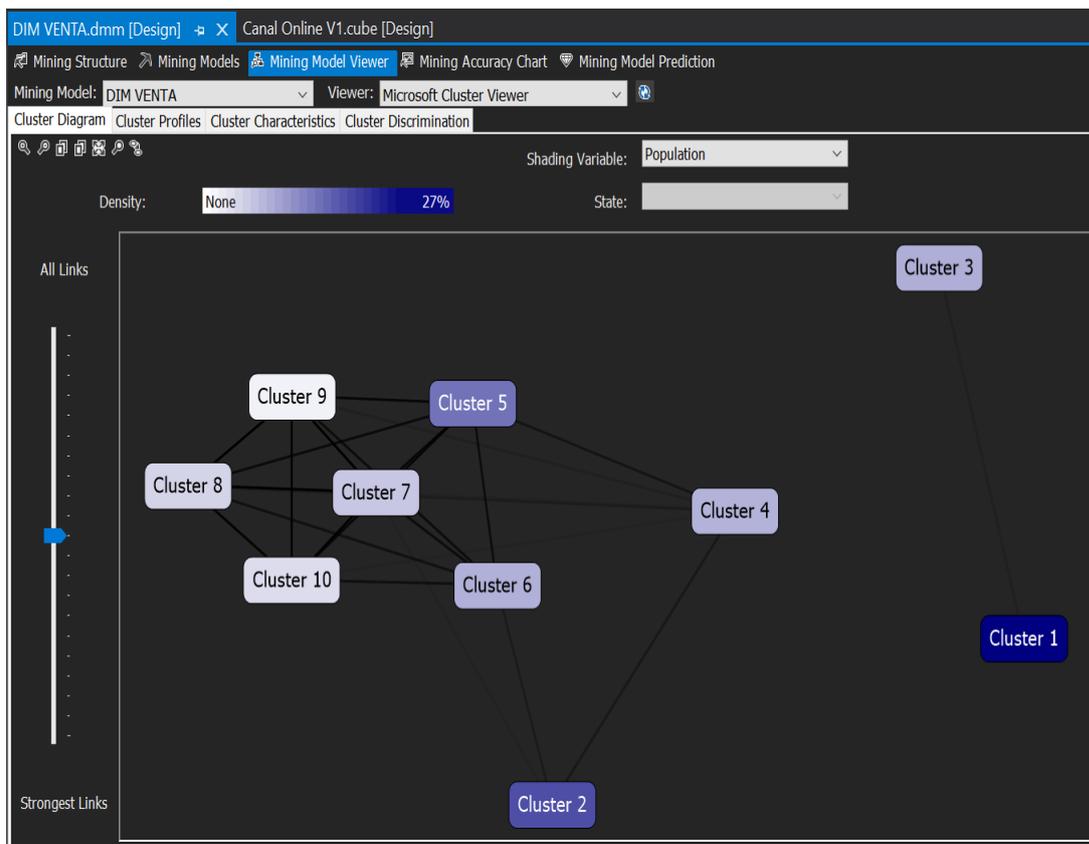


Figura 83. Visualización algoritmo de agrupación.

El algoritmo entrega una serie de probabilidades. Donde según la figura 84 se puede afirmar que la probabilidad de que en Pichincha se haga una venta fraudulenta es el 0.264%. Siendo la provincia con más índice de fraudes, por lo tanto, es la que tiene el mayor porcentaje. Esto hace que el negocio pueda tomar acciones en el escenario.

Cluster Model (000)

- Cluster 1 (001)
- Cluster 2 (002)
- Cluster 3 (003)
- Cluster 4 (004)
- Cluster 5 (005)
- Cluster 6 (006)
- Cluster 7 (007)
- Cluster 8 (008)
- Cluster 9 (009)
- Cluster 10 (010)

ATTRIBUTE_VALUE	SUPPORT	PROBABILITY	VARIAN
Missing	0	0	0
ORELLANA	29	0,00325367440816783	0
LOJA	55	0,00617076180859416	0
PICHINCHA	2356,00000000002	0,264332996746327	0
EL ORO	187	0,0209805901492202	0
GUAYAS	1508,00000000001	0,169191069224729	0
AZUAY	236,00000000001	0,0264781778423314	0

0776

Figura 84. Probabilidad de Fraude.

6. Conclusiones y Recomendaciones

6.1 Conclusiones

Con la implementación de este prototipo de inteligencia de negocio se pudo levantar un indicador de fraude global medido en base a las ventas con fraude sobre la cantidad de ventas mensuales. Actualmente se establece en un 17%. Con este indicador la empresa tiene una mejor visibilidad de esta problemática y se puede implementar mejores controles para disminuir este porcentaje.

Con los indicadores de fraude por día, campaña y producto. Se implementó en producción un control de validación de identidad en el flujo de ventas permitiendo disminuir 2 puntos porcentuales el fraude. Con eso se concluye que al tener visibilidad de los datos se pueden implementar mejoras logrando disminuir el porcentaje inicial. Cabe recalcar que esta es la primera acción realizada, es decir se pueden implementar mejores controles adicionales tales como validación de direcciones ya que en la minería de datos se encontró una probabilidad de 26% de fraude en Pichincha.

Centralizar la información en un solo almacén de datos es clave para el negocio a la hora de analizar la información, obteniendo de informes más dinámicos y alertas en base a indicadores que impacten el giro del negocio. Se logró que el área involucrada tenga mejor perspectiva de la información ya que se pueden implementar KPIS en cada punto estratégico del flujo de ventas con estos indicadores se pueden tomar decisiones estratégicas para mejorar procesos y la experiencia de compra del cliente.

Si se alimenta con más información al almacén de datos, es posible generar informes más robustos con la visión completa del funnel. Obteniendo mejores indicadores para la toma de decisiones con menor riesgo a equivocarse. Finalmente, se reduce la carga operativa ya que los informes manuales o cruces de archivos cada vez serán menores.

Es importante ser consciente de la curva de aprendizaje al momento de implementar una solución de inteligencia del negocio. Debido a que en la práctica nacen ciertas barreras funcionales o técnicas que deben ser resueltas en base a métodos teóricos planteados por expertos. Las soluciones de estas barreras hacen que el usuario adopte de mejor manera las soluciones BI.

La preparación de los datos en el DWH es la parte de inteligencia del negocio que demanda mayor esfuerzo y tiempo. En este punto es importante enfocarse y seguir cada recomendación que hacen empresas líderes de este sector tecnológico. Debido a que de esto depende el resultado comprensión de los datos al aplicar un modelo de minería de datos.

6.2 Recomendaciones

Dentro del diseño dimensional, las tablas de hechos son importantes y para la correcta creación de estas se debe conocer las necesidades del negocio. Determinar la información que se desea analizar y el motivo que impulsa el análisis de esta.

En la etapa inicial, previo a la implementación del prototipo se recomienda mantener reuniones con las personas involucradas para levantar requerimientos funcionales para lograr que todo el desarrollo satisfaga necesidades del negocio.

Es importante saber la cantidad de datos con las que se va a trabajar, para realizar una correcta selección de las herramientas de inteligencia de negocio que se va a usar. Para al momento de la implementación y paso a producción no tener limitantes de rendimiento.

A lo largo del desarrollo del este prototipo se puede apreciar que las herramientas utilizadas para la implementación son Microsoft. Por lo que una recomendación técnica es que se debe diferenciar la configuración de las instancias de bases multidimensionales. Al instalar se debe configurar una

instancia multidimensional para poder procesar los cubos creados y los algoritmos de minería de datos.

Para este tipo de prototipos es importante trabajar sobre una base de respaldo. Debido a que, al ser información sensible de la organización, esta puede ser alterada por lo que no se recomienda trabajar en las bases relacionales de producción. Como se evidenció, en este prototipo se creó una instancia replica en AWS para no afectar los datos de producción.

No necesariamente se debe conocer a profundidad sobre inteligencia del negocio para crear un ETL. Debido a que estas herramientas son técnicamente intuitivas, basta con tener fundamentos básicos de este tema. En la práctica se adquiere experiencia, mejorando el conocimiento para futuras implementaciones.

Respecto a los tableros gerenciales, siempre es importante tener una reunión funcional con las personas que harán uso de los indicadores que contiene el tablero. Para posteriormente implementarlos de manera sencilla y que sea de fácil comprensión para el usuario final.

Finalmente, se recomienda seguir cada consejo de expertos del tema ya sea técnicamente para el correcto uso de las herramientas de inteligencia del negocio. Así como con expertos funcionales del giro del negocio ya que los indicadores deben resolver las necesidades comerciales.

Referencias

- Amazon Web Services. (2019). *Datawarehouse*. Recuperado el 12 de diciembre de 2019 de <https://aws.amazon.com/es/data-warehouse/>
- Blaublog. (2019). Diferencias entre Power BI y SSRS. Recuperado el 12 de diciembre de 2019 de <http://blog.blauservices.com/?p=1452>
- Bussiness Intelligence. (2019). Barreras a la adopción de una solución Business Intelligence. Recuperado el 12 de diciembre de 2019 de <https://www.businessintelligence.info/definiciones/barreras-adopcion-business-intelligence.html>
- Data-Flair. (2019). *Data mining Architecture*. Recuperado el 13 de diciembre de 2019 de <https://data-flair.training/blogs/data-mining-architecture/>
- dblearner. (2018). Poblar la tabla de dimensión de tiempo para un datawarehouse. Recuperado el 14 de diciembre de 2019 de <http://dblearner.com/script-poblar-la-tabla-de-dimensin-de-tiempo-para-un-datawarehouse/>
- Defiebre . (2019). ¿Qué es el funnel o embudo de conversión? Recuperado el 25 de diciembre de 2019 de <https://www.40defiebre.com/que-es/embudo-conversion>
- Finances Online. (2019). *Business Intelligence Software*. Recuperado el 25 de diciembre de 2019 de <https://financesonline.com/>
- Gestión.org. (2019). Técnicas de minería de datos. Recuperado el 25 de diciembre de 2019 de <https://www.gestion.org/tecnicas-de-mineria-de-datos/>
- gravitar. (2018). SQL Server 2012 – Multidimensional vs Tabular. Recuperado el 28 de diciembre de 2019 de <https://gravitar.biz/bi/sql-server-2012-multidimensional-vs-tabular/>

- Grupokorporate. (2019). Los cinco tipos de fuentes de datos. Recuperado el 30 de diciembre de 2019 de <https://grupokorporate.com/los-cinco-tipos-de-fuentes-de-datos/>
- IBM. (2019). *Sequential patterns Data Mining*. Recuperado el 30 de diciembre de 2019 de https://www.ibm.com/support/knowledgecenter/en/SSHRBY/com.ibm.svg.im.dashdb.analytics.doc/doc/r_sequential_patterns.html
- Inaoep. (2019). Minería de datos. Recuperado el 30 de diciembre de 2019 de https://ccc.inaoep.mx/~jagonzalez/AI/Sesion13_Data_Mining.pdf
- JoseFacchin. (2019). *KPIS*. Recuperado el 30 de diciembre de 2019 de <https://josefacchin.com/indicadores-kpi-marketing-o-unidades-de-medida/>
- Medium. (2019). *How to do Data Mining on Cube in SSAS*. Recuperado el 30 de diciembre de 2019 de <https://medium.com/@shristibal1998/how-to-do-data-mining-on-cube-in-ssas-22db83c15cd2>
- Microsoft. (2019). *Error Procesamiento de Cubo OLAP*. Recuperado el 30 de diciembre de 2019 de <https://support.microsoft.com/en-ae/help/922673/error-messages-when-you-try-to-process-a-database-or-a-cube-in-sql-ser>
- Microsoft. (2019). ¿Qué es Microsoft Business Intelligence? Recuperado el 30 de diciembre de 2019 de <https://blablanegocios.com/todo-sobre-microsoft-business-intelligence/>
- Microsoft. (2019). *Flow Control*. Recuperado el 30 de diciembre de 2019 de <https://docs.microsoft.com/en-us/sql/integration-services/control-flow/control-flow?view=sql-server-ver15>

- Microsoft. (2019). *Asociation Data mining*. Recuperado el 30 de diciembre de 2019 de <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/microsoft-association-algorithm?view=sql-server-2014>
- Microsoft. (2019). *Data Flow Task*. Recuperado el 30 de diciembre de 2019 de <https://docs.microsoft.com/en-us/sql/integration-services/control-flow/data-flow-task?view=sql-server-ver15>
- Microsoft. (2019). *SSAS SQL Server Analytical Services*. Recuperado el 27 de diciembre de 2019 de <https://docs.microsoft.com/en-us/analysis-services/analysis-services-overview>
- Microsoft. (2019). *SSIS SQL Server Integration Services*. Recuperado el 19 de diciembre de 2019 de <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver15>
- Microsoft. (2019). *SSRS SQL Server Reporting Services*. Recuperado el 19 de diciembre de 2019 de <https://docs.microsoft.com/en-us/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-ver15>
- Sinnexus. (2019). *Datawarehouse*. Recuperado el 19 de diciembre de 2019 de https://www.sinnexus.com/business_intelligence/datawarehouse.aspx
- Sinnexus. (2019). Datos, Información, Conocimiento. Recuperado el 19 de diciembre de 2019 de https://www.sinnexus.com/business_intelligence/piramide_negocio.aspx
- Sinnexus. (2017). Cubos OLAP. Recuperado el 19 de diciembre de 2019 de https://www.sinnexus.com/business_intelligence/olap_avanzado.aspx
- Slideshare. (2019). Conceptos de almacenes de datos (Data Warehousing). Recuperado el 19 de diciembre de 2019 de <https://es.slideshare.net/fuylvlp/almacenes-de-datos-olap-y-minera-de-datos>

Solidq. (2019). Hechos y Dimensiones: Modelado Dimensional. Recuperado el 18 de diciembre de 2019 de <https://blogs.solidq.com/es/business-analytics/hechos-y-dimensiones-modelado-dimENSIONAL-12/>

Tecnología informática. (2019). Sistemas OLAP: Análisis empresarial cubos y tipos de OLAP. Recuperado el 18 de diciembre de 2019 de <https://tecnologia-informatica.com/sistemas-olap-cubos/>

Theadiar. (2019). Informes y visualización de datos en herramientas de inteligencia Empresarial. Recuperado el 16 de diciembre de 2019 de <https://www.thediar.com/es/what-we-do/business-intelligence>

Ula. (2019). Modelos de datos. Recuperado el 18 de diciembre de 2019 de <http://www.ing.ula.ve/~aguilar/actividad-docente/IN/transparencias/clase40.pdf>

ANEXOS

GLOSARIO DE ACRÓNIMOS Y TÉRMINOS

AWS Amazon Web Services (Cloud de amazon).

CECE Cámara Ecuatoriana de Comercio Electrónico.

CRM Customer Relationship Management.

DTSX Extensión de un archivo SSIS.

DMS Database Management System.

DB Database (Base de datos).

E2E End to End.

ETL Extract, transformation and Load.

FUNNEL Embudo de conversión utilizado en marketing Digital, define los distintos pasos por el que viaja un pedido.

GARTNER Empresa consultora y de investigación de las tecnologías de la información.

HTML HyperText Markup Language (Lenguaje de Marcas de Hipertexto).

IBM International Business Machines Corporation es una reconocida empresa multinacional estadounidense de tecnología y consultoría.

KNOW-HOW Sinónimo de experiencia empresarial.

KPI Key performance Indicator.

OBI Oracle Business Intelligence.

ODBC Open DataBase Connectivity.

RDS Relational Database Service.

RSS Relational Software System.

SQL Structure Query Language.

SAP Empresa multinacional dedicada al diseño de productos informáticos de gestión empresarial.

SLA Service Level Agreement – Nivel de servicio.

SCRIPT Archivo de órdenes o archivo de procesamiento por lote.

XML Extensible Markup Language (Lenguaje de Marcado Extensible).

