

UNIVERSIDAD HISPANOAMERICANA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

LICENCIATURA

Construcción de un DataMart comercial, con sistema de autoservicio de información para usuarios analíticos por medio de herramientas de punta.

Sustentante:
Oscar Alberto Arias Cortés.

TUTOR:
Erick López.

Julio 2019.

**UNIVERSIDAD HISPANOAMERICANA
CENTRO DE INFORMACION TECNOLOGICO (CENIT)
CARTA DE AUTORIZACIÓN DE LOS AUTORES PARA LA CONSULTA, LA
REPRODUCCION PARCIAL O TOTAL Y PUBLICACIÓN ELECTRÓNICA
DE LOS TRABAJOS FINALES DE GRADUACION**

San José, 09 de septiembre de 2020

Señores:
Universidad Hispanoamericana
Centro de Información Tecnológico (CENIT)

Estimados Señores:

El suscrito (a) Oscar Alberto Arias Cortés con número de identificación 402090224 autor (a) del trabajo de graduación titulado Construcción de un DataMart comercial, con sistema de autoservicio de información para usuarios analíticos por medio de herramientas de punta. Presentado y aprobado en el año 2020 como requisito para optar por el título de Licenciatura Informática con énfasis en administración de proyectos. Autorizo al Centro de Información Tecnológico (CENIT) para que con fines académicos, muestre a la comunidad universitaria la producción intelectual contenida en este documento.

De conformidad con lo establecido en la Ley sobre Derechos de Autor y Derechos Conexos N° 6683, Asamblea Legislativa de la República de Costa Rica.

Cordialmente,


402090224.
Firma y Documento de Identidad

Dedicatoria

Primero que todo quiero darle gracias a Dios por permitirme terminar mis estudios universitarios con esta licenciatura y pedirle también que me brinde más adelante las oportunidades para seguir estudiando y creciendo profesional y académicamente.

A mis papas que siempre me ayudaron con todo lo que pudieron y que sé que también sacrificaron muchas cosas por darnos los estudios a mis hermanos y a mí. Gracias por todo, hoy con más madurez entiendo muchos de los consejos que me dieron cuando recién inicié y durante el día a día y que quizás en su momento no entendí o no supe valorar.

A Marco también que muchas de las noches que me quede hasta tarde trabajando en la tesis y en temas de trabajo se queda despierto para preguntarme que porque no estaba dormido y porque todos los días me preguntaba que cuanto me faltaba y si hoy si me iba a dormir temprano.

A Taty también las gracias por siempre apoyarme y escucharme cuando muchas veces pensé que no lo iba a lograr, siempre tuvo palabras de apoyo y a pensar de todas las dificultades que hemos pasado siempre ha estado para mí.

A mis compañeros de trabajo y a mi jefe, que también siempre que requerí ayuda estuvieron ahí aun cuando era tarde o me costara solventar alguna dificultad, no dudaron en explicarme o colaborarame.

A todas las personas que de una u otra manera me ayudaron a llegar hasta aquí, gracias !!!

CARTA DEL TUTOR

Heredia, 12 de Junio del 2020

María Isabel Losilla.
Ingeniería en Sistemas.
Universidad Hispanoamericana

Estimada señora:

El estudiante Oscar Alberto Arias Cortés, cédula de identidad número 402090224, me ha presentado, para efectos de revisión y aprobación, el trabajo de investigación denominado Construcción de un DataMart comercial, con sistema de autoservicio de información para usuarios analíticos por medio de herramientas de punta el cual ha elaborado para optar por el grado académico de licenciatura.

En mi calidad de tutor, he verificado que se han hecho las correcciones indicadas durante el proceso de tutoría y he evaluado los aspectos relativos a la elaboración del problema, objetivos, justificación, antecedentes, marco teórico, marco metodológico, tabulación, análisis de datos, conclusiones y recomendaciones.

De los resultados obtenidos por el postulante, se obtiene la siguiente calificación:

a)	ORIGINALIDAD DEL TEMA	10%	10%
b)	CUMPLIMIENTO DE ENTREGA DE AVANCES	20%	15%
c)	COHERENCIA ENTRE LOS OBJETIVOS, LOS INSTRUMENTOS APLICADOS Y LOS RESULTADOS DE LA INVESTIGACION	30%	27%
d)	RELEVANCIA DE LAS CONCLUSIONES Y RECOMENDACIONES	20%	20%
e)	CALIDADM DETALLE DEL MARCO TEORICO	20%	20%
	TOTAL	100%	92%

En virtud de la calificación obtenida, se avala el traslado al proceso de lectura.

Atentamente,



Ing. Erick López Ch. M.R.I
109930088

CARTA DEL LECTOR

CARTA DE LECTOR

San José,

**Universidad Hispanoamericana
Sede Llorente
Carrera**

Estimado señor

La estudiante Oscar Arias Cortes, cédula de identidad 402090224, me ha presentado para efectos de revisión y aprobación, el trabajo de investigación denominado " Construcción de un DataMart comercial, con sistema de autoservicio de información para usuarios analíticos por medio de herramientas de punta ", el cual ha elaborado para obtener su grado de Licenciatura.

He revisado y he hecho las observaciones relativas al contenido analizado, particularmente lo relativo a la coherencia entre el marco teórico y análisis de datos, la consistencia de los datos recopilados y la coherencia entre éstos y las conclusiones; asimismo, la aplicabilidad y originalidad de las recomendaciones, en términos de aporte de la investigación. He verificado que se han hecho las modificaciones correspondientes a las observaciones indicadas.

Por consiguiente, este trabajo cuenta con mi aval para ser presentado en la defensa pública.

Atte.

Firma: *Randall Vargas V*
Nombre: Randall Vargas Villalobos
Cédula: 1-1140-0113
Carné: No Aplica

DECLARACIÓN JURADA

Yo, Oscar Alberto Arias Cortés, mayor de edad, portador de la cédula de identidad número 4-0209-0224, egresado de la carrera de ingeniería informática de la Universidad Hispanoamericana, hago constar por medio de éste acto y debidamente apercibido y entiendo las penas y consecuencias con las que se castiga en el código penal el delito de perjurio, ante quienes constituye el tribunal examinador de mi proyecto de tesis para optar por el título de licenciatura, juro solemnemente que mi trabajo titulado: **Construcción de un DataMart comercial, con sistema de autoservicio de información para usuarios analíticos por medio de herramientas de punta**. Es una obra original que respetado todo lo preceptuado por las leyes penales, así como la ley de derecho de autor y derechos conexos número 6683 del 14 de octubre de 1982 y sus reformas publicadas en la gaceta número 226 del 25 de noviembre de 1982; y sus reformas, incluyendo el numeral 70 de dicha ley que advierte; artículo 70. Es permitido citar a un autor, transcribiendo los pasajes pertinentes siempre que estos no sean tantos y seguidos, que puedan considerarse como una producción simulada y sustancial, que redunde en perjuicio del autor, de la obra original. Asimismo, quedo advertido que la universidad se reserva el derecho de protocolizar este documento ante un notario público. En fe de lo anterior, firmo en la ciudad de Heredia, a los once días del mes de junio del dos mil veinte.



4-0209-0224

Firma del estudiante.

Cédula.

Contenido

Capítulo I.....	1
Antecedentes Y Justificación Del Proyecto.....	1
1.1 Marco De Referencia Empresarial y Contextual	3
1.2 Justificación del Proyecto	3
1.3 Definición del problema	4
1.3.1 Problemática	4
1.3.2 Problema general	5
1.3.3 Problemas específicos	5
1.4 Objetivo General y Objetivos Específicos	6
1.4.1 Objetivo General.	6
1.4.2 Objetivos específicos.....	6
1.5 Alcance y Limitaciones.....	7
1.5.1 Alcance del Proyecto	7
1.5.2 Limitaciones.....	8
1.6 Cronograma.....	9
Capítulo II.....	11
Marco Teórico	11
2.1 Data Warehouse.....	12
2.2 Características de un DW	12
2.3 Arquitectura de un Data Warehouse.....	13
2.4 Business Intelligence	14
2.4.1 Usos de herramientas para Business intelligence:.....	17
2.4.2 ¿Como puede BI ayudar a las empresas?	17
2.5 Bases de datos OLTP y OLAP.....	19
2.5.1 Características de los datos transaccionales.....	21
2.6 Modelo multidimensional	21
2.7 Un cubo OLAP para cada necesidad.	24
2.7.1 ROLAP.....	25
2.7.2 MOLAP.....	26
2.7.3 HOLAP.....	27
3. Arquitectura de Business Intelligence.	27

3.1 William Inmon.....	27
3.1.1 La estructura del Data Warehouse	29
3.2 Ralph Kimball.....	29
3.3 Comparación entre Kimball e Inmon.....	31
3.4 Inmon y Kimball.....	32
4. ETL.....	33
4.1 ¿Para qué sirve este proceso?	35
4.2 Desafíos de los procesos ETL	37
5. Datamart	37
5.1 El futuro de los data marts es en la nube.	38
5.2 Comparativa entre Data Warehouse y Data Mart	39
6. Esquema Estrella	41
7. Copo de nieve.....	42
8. Tabla de hechos.....	43
8.1 La importancia de las dimensiones para el análisis.	44
9. Data Lake.....	45
9.1 Diferencias entre un Data Lake y un Data Warehouse	47
9.2 ¿Data lakes en la nube o local?	52
9.3 Los retos de los data lakes en la nube	53
9.4. Data lakes en Hadoop.....	53
9.5 Data lakes en AWS.....	54
9.6 Data lakes en Azure	55
10. Data Analytics.....	57
11. Herramientas de visualización.....	58
11.1 Tableau	59
11.2 Qlik	60
11.3 Google Charts	61
11.4 Power BI	61
11.4.1 Algunas de sus ventajas son:	62
Capítulo 3: Marco Metodológico.....	63
3.1 Enfoque de la Investigación:.....	64
3.2 Fuentes y Sujetos De Información	65
3.2.1 Fuentes de Información:.....	65

3.2.2	Sujetos de Información:.....	65
3.3	Técnicas y Herramientas de Recolección De Datos	66
3.4	Variables de Investigación	67
Capítulo IV.	Diagnóstico de la situación actual	71
4.1	Diagnóstico de la situación actual:	72
4.2	Diagnóstico operativo:	74
4.3.	Diagnóstico técnico:	76
4.4	Diagnostico percepción:	77
4.5	Análisis de brechas.	78
Capítulo V.	Propuesta de Proyecto.....	79
5.1	Diagrama de base de datos:	81
5.2	Tablas	83
5.2.1	Tablas Frontera.....	84
5..2.2	Tablas Stage.....	100
5.2.3	Tablas DWH	111
5.3	Jobs.....	124
5.4	Diseño de SQL Server Integration Services (SSIS)	124
5.5	Extracciones desde la base de datos	125
5.6	Proceso de transformaciones.	127
5.7	Base de datos multidimensionales (Cubos).....	142
5.7.1	Dimensiones	143
5.8	Medidas calculadas.	146
5.9	Dashboards.....	149
5.9.1	Dashboard 01 – Inventario	149
5.9.2	Dashboard 02 – Presencia por artículo.....	150
5.9.3	Dashboard 03 - Clientes.....	151
5.9.4	Dashboard 04 – Venta bruta vs venta Bruta Año anterior.....	152
5.9.5	Dashboard 05 – Venta bruta por provincia.....	152
5.9.6	Dashboard 06 – Unidades vendidas por grupo.....	153
5.9.7	Dashboard 07 - Efectividad Neta	154
5.9.8	Dashboard 08 - Presencia por artículo tabla.....	154
5.9.9	Dashboard 09 - Antigüedad de saldos	155
5.9.10	Dashboard 10- Clientes sin venta	156

5.9.11 Dashboard 11 - Comparativo ventas por vendedor.....	157
5.9.12 Dashboard 12 – Bonificaciones por vendedor.....	158
5.9.13 Dashboard 13 – Comisiones	158
5.9.14 Dashboard 14 – Inventario en almacén.....	159
5.9.15 Dashboard 15 – Descuento sobre la venta por vendedor.....	160
5.9.16 Dashboard 16 – Devoluciones en buen estado.....	161
5.9.17 Dashboard 17 – Devoluciones en mal estado.....	162
Capítulo VI.....	164
Conclusiones y Recomendaciones del Proyecto.....	164
6.1 Conclusiones.....	165
6.2 Recomendaciones	167

Capítulo I

Antecedentes Y Justificación

Del Proyecto

Introducción

En la actualidad cada día se torna más complicado sobrevivir en un mercado tan agitado y cambiante, es por esto por lo que las empresas deben buscar soluciones que les permitan seguir vigentes en este mercado, pero sobre todo obtener ventaja sobre sus competidores.

Es por esto por lo que, alimentos Jack's se ve en la necesidad de buscar una serie de herramientas o tecnologías que les permita dar el siguiente paso en la industria alimenticia y comercial.

El tener los datos de manera veraz, confiable y a unos pocos segundos permite a la compañía la toma de decisiones, pero sobre todo decisiones de calidad, que le permitan actuar no solo de manera reactiva si no de manera también de tener la posibilidad de actuar antes de que las cosas sucedan.

Maximizar el uso de la información obtenida a través del tiempo, saber administrarla, y comprender el entorno del negocio, permite tener una visión más amplia de las posibilidades y limitaciones que se tienen.

1.1 Marco De Referencia Empresarial y Contextual

El proyecto será realizado en la empresa alimentos Jack´s, empresa costarricense dedicada a la producción alimentaria desde los años 1962-1963 y desde entonces se han mantenido en el mercado nacional e internacional con gran variedad de productos alimenticios.

Misión de la empresa:

Producir alimentos de conveniencia buscando satisfacer el gusto de consumidores de bocadillos sabrosos y que provean una nutrición balanceada, dispuesta a cumplir con las exigencias cambiantes de los consumidores.

1.2 Justificación del Proyecto

El presente desarrollo e investigación de este proyecto busca facilitar de manera certera y oportuna a las diferentes áreas de toma de decisiones, llamase gerencias o direcciones para que siempre cuenten con los datos de manera veraz y veloz.

Desarrollar este proyecto permitirá tener un manejo más adecuado de los datos, ya que al tenerlos almacenados en un DataMart es más ágil el procesamiento y uso de estos, a su vez facilita la toma de decisiones en un ámbito tan importante y cambiante como lo es en las empresas de consumo masivo y en donde en muchas ocasiones solo se dispone de poco tiempo para tomar medidas o decisiones.

En algunas oportunidades las gerencias o direcciones deben solicitar a los diferentes departamentos de tecnología información para la toma de decisiones, sin embargo, sucede que la información no se tiene actualizada o a la mano, por esto, la creación de un DataMart permitirá que la información sea actualizada de manera diaria, además que, al ser una herramienta de autoservicio por medio de herramientas de punta, se eliminar la dependencia en gran parte de las áreas antes mencionadas.

Hacer las cosas de esta manera genera ventaja sobre la competencia, ya que permitirá tomar mejores y más acertadas decisiones que permitirán beneficiar a la entidad y también a los clientes finales de la organización.

1.3 Definición del problema

1.3.1 Problemática

En este momento la entidad tiene muchísima información almacenada en diferentes repositorios como bases de datos, archivos planos, Excel y sistemas transaccionales todos dispersos en diferentes áreas o carpetas, esto complica muchísimo el acceso a la misma ya que en muchas ocasiones se necesita tener los datos de manera inmediata y esto es imposible de la manera actual, con el desarrollo del proyecto, se busca cambiar esta situación drásticamente.

Actualmente las organizaciones se encuentra en| crecimiento y cambio constante, crecen las necesidades de los clientes y cartera de productos lo hace de la misma forma, pero no podemos dejar de lado que así como todo esto crece y cambia, la información que se almacena en la entidad también lo hace, en general las organizaciones no le toman importancia a la información que logran recopilar a través del tiempo pero esto debe cambiar, implementar un DataMart permitirá que se tenga información histórica y actual a la mano, siempre que sea precisa, de fácil acceso y sobre todo que pueda ser accesada rápidamente.

Tener la información de esta manera genera una herramienta y ventaja competitiva para la organización.

En los últimos años Business Intelligence o inteligencia de negocios, se ha convertido en una necesidad y no un área más dentro de las organizaciones por sus diferentes ventajas, dentro de las que se pueden mencionar:

- Permite analizar y comprender datos con análisis detallado.
- Permite mejorar los informes con diferentes reportes y gráficos adaptados para cada situación

- Permite tener diferentes perspectivas y diferentes grados de granularidad para analizar la información.
- Permite el autoservicio en la generación de informes para distintos usuarios.
- El acceso a la información es sencillo y rápido.
- Es un medio efectivo para obtener información de manera sencilla integrando los diferentes tipos de los datos y confeccionando los distintos formatos de informes.
- Permite a los directivos tener un panorama completo de los datos y a su vez transfórmalos en información valiosa para la organización.
- Al tener la información a la mano permite en dado caso replantearse los objetivos y metas trazadas.
- Analizar de forma interactiva información crítica de las diferentes unidades de negocio.
- Permitir identificar fácilmente problemas y oportunidades.
- Permitir a los usuarios obtener gran cantidad de información para analizar y establecer relaciones, comprender tendencias y soportar las decisiones
- Prevenir la potencial pérdida de conocimiento resultante de la masiva acumulación de información de lento y difícil acceso. (Muñiz, 2018. P.22)

1.3.2 Problema general

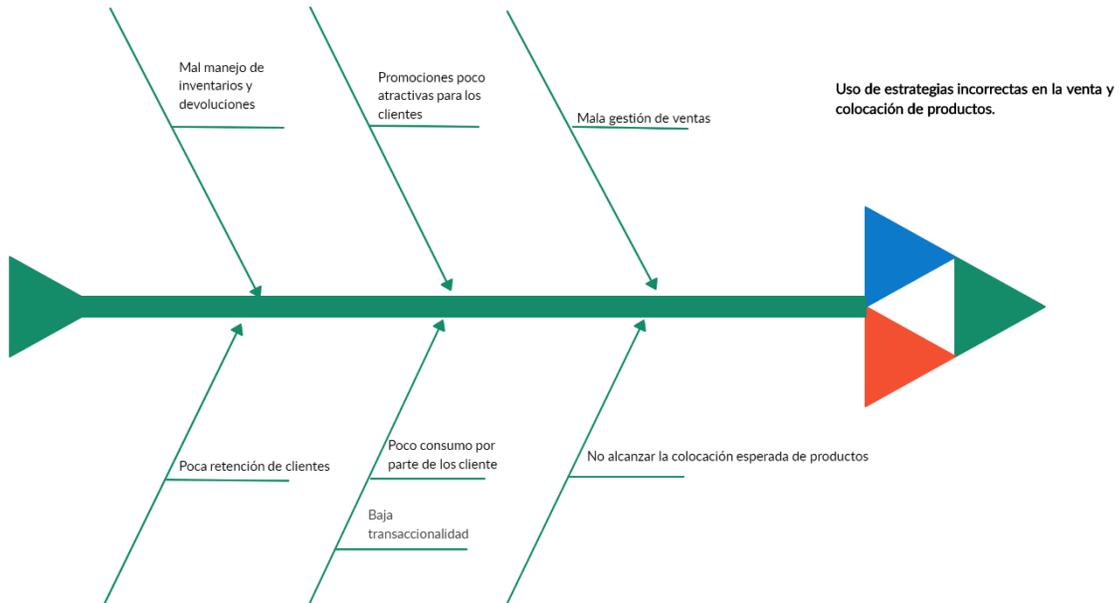
¿Puede la implementación de un DataMart ayudar a mejorar la toma de decisiones en la organización alimenticia en Costa Rica respecto a la colocación y venta de productos?

1.3.3 Problemas específicos

¿Pierde la entidad oportunidades de sacar ventaja sobre otras entidades al no tener acceso a la información de manera fácil y rápida?

¿Genera más trabajo y reportaría innecesaria la creación de un DataMart para la organización?

Figura 1. Diagrama de causas y efectos.



Fuente: Elaboración propia, 2019.

1.4 Objetivo General y Objetivos Específicos

1.4.1 Objetivo General.

- Desarrollar una herramienta analítica de inteligencia de negocios que facilite a las áreas de ventas la toma de decisiones mediante la creación de un DataMart e informes de autoservicio por medio de Power BI.

1.4.2 Objetivos específicos.

- Analizar las diferentes fuentes de datos de donde se va a tomar la información.

- Homologar las diferentes fuentes de datos que vamos a trabajar para la creación del DataMart.
- Implementar un DataMart comercial que facilite la toma de decisiones a las áreas analíticas.
- Desarrollar un grupo de Dashboards con la información obtenida del DataMart, que permita mostrar información correspondiente a las ventas de la compañía y la cantidad productos en inventario.

1.5 Alcance y Limitaciones

1.5.1 Alcance del Proyecto

El proyecto se desarrollará en el departamento de tecnologías de información con colaboración del area de ventas.

Se deberán homologar las fuentes de datos de donde se va a tomar la información, con el fin de tener la información de manera consistente para así poder realizar el desarrollo.

Se generará un DataMart para los procesos de toma de decisiones para las áreas de ventas, que permita observar las tendencias de consumo de los clientes, para así poder dirigir las estrategias y esfuerzos a segmentos específicos.

Se crearán reportes que permitan tener información precargada de manera diaria para que los encargados de las tomas de decisiones puedan obtener de manera rápida y sencilla la información que necesiten al momento.

El resultado serán modelos dimensionales que serán la base para crear un Cubo OLAP, con el cubo será la base o fuente para alimentar los indicadores y reportes que se mencionan anteriormente.

1.5.2 Limitaciones

Por políticas de la organización y en relación con la ley número 8968 de la república de Costa Rica que protege la información de las personas y al trato que se le da a la información, no se podrá revelar nombres, cifras exactas y alguna otra información considerada como privada entre la compañía y los clientes, por lo tanto, se deberá implementar una forma de encriptar estos datos y realizar el proyecto.

Todo el proyecto se desarrollará en Microsoft SQL Server y Microsoft Visual Studio por temas de licenciamiento de la organización.

1.6 Cronograma

<u>Nombre Tarea</u>	<u>Duración(días)</u>	<u>% Cumplimiento</u>
Planificación del proyecto	5	100%
Antecedentes y Justificación del proyecto	1	100%
Definición del problema general y específicos	1	100%
Alcances y Limitaciones	0,5	100%
Marco Empresarial	0,5	100%
Cronograma del Proyecto	2	100%
Diseño de Base de Datos	10	0%
Definición de Granulidad	1	0%
Definición de las tablas	2	0%
Creación de Tablas	2	0%
Dimensión de Tiempo	1	0%
Diseño de Modelo	3	0%
Tablas de Apoyo	1	0%
Diseño de ETL	4	0%
Proceso de Extracción	1	0%
Proceso de Transformación	1	0%
Proceso de Carga	1	0%
Proceso de Automatización	1	0%
Diseño del Cubo	1	0%
Reportes	4	0%
Definición de Reportes	1	0%
Creación de Reportes	3	0%
Pruebas	2	0%
Pruebas de Funcionamiento	2	0%
Implementación	2	0%
Capacitación de Usuarios	3	0%
Entrega a las Áreas	1	0%

Tabla 1. Cronograma del Proyecto.

Fuente: Elaboración propia.

Capítulo II

Marco Teórico

Introducción

En el siguiente capítulo se procederá con la explicación de metodologías, características, arquitecturas, herramientas, modelos de datos, repositorios de datos, que permitirán que la comprensión del proyecto sea más sencilla.

2.1 Data Warehouse

Podemos definir un Data Warehouse (DW) como un almacén o repositorio en donde se almacenan los datos provenientes de diferentes sistemas de la empresa, estos se pueden clasificar en físicos y lógicos y su valor reside en justamente en poder tener los datos de muchísimas fuentes distintas.

El concepto como tal, nace en los años 80 mediante la empresa IBM, los investigadores buscaban establecer un sistema que permitiera el manejo fluido y permanente de los datos y que a su vez estuviera organizado de acuerdo con las necesidades específicas. Los datos almacenados podrían ser de cualquier tipo y abarcar diferentes áreas de la empresa.

En el año 1985 aparece el primer Data Warehouse con el fin de organizar e integrar los datos que se encontraban de manera dispersa en múltiples bases de datos, el creador fue Bill Inmon para un banco en Colorado y a partir de este periodo se han ido perfeccionando hasta la actualidad. (Bembibre, 2009)

2.2 Características de un DW

Los Data Warehouse deben cumplir con las siguientes características:

- **Integrados:** Los datos como provienen de diferentes fuentes de información por lo cual deben pasar por un proceso de integración para quedar en una estructura, las inconsistencias deben

ser eliminadas en el proceso, también suele agruparse en diferentes niveles de detalle de acuerdo con las necesidades de los usuarios.

- **Agrupado por temas:** los datos deben ser agrupados por temas para facilitar el acceso y entendimiento para los usuarios finales del producto, por ejemplo, no existirá información de clientes o productos en más de una estructura ya que conservando todo en una tabla será más fácil de responder dado a que todo reside en el mismo lugar.
- **No volátil:** la razón de ser del DW es ser consultada para lectura y no para actualización de los datos por tanto la información debe ser permanente durante el tiempo. Se podrá ingresar nueva información, pero sin que esto signifique alterar la ya existente.
- **Histórico:** es uno de los factores más importantes de un data warehouse, permite el análisis de la historia y a su vez el estudio de tendencias, también se usan como datos comparativos.
- **Soporte al usuario final:** ayuda a que puedan acceder a la información con un lenguaje propio del negocio, indicando que significa la información almacenada y la importancia que tiene la misma. Con esto los usuarios pueden construir reportes, Dashboards e indicadores de acuerdo con su necesidad. (Data Warehouse Y Business Intelligence, 2018)

2.3 Arquitectura de un Data Warehouse

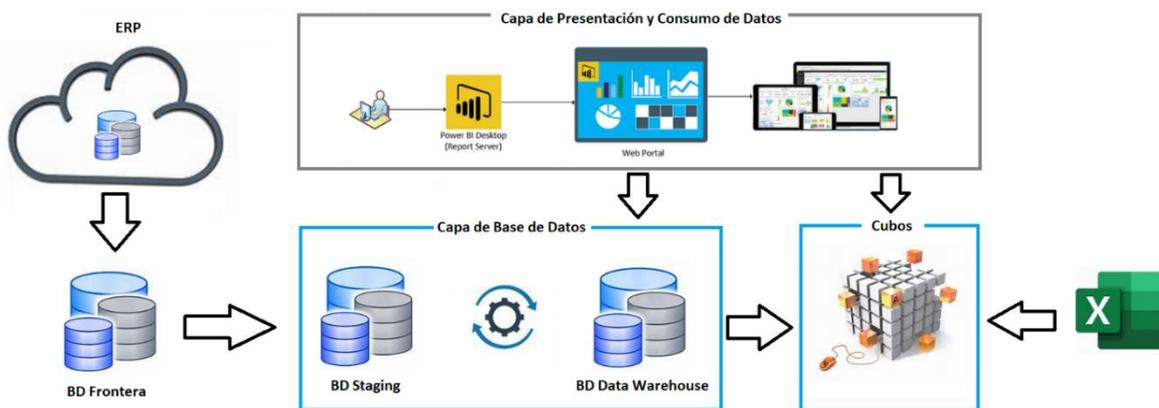
La arquitectura general de un DW se compone de varias partes que vamos a ir detallando a continuación:

Como se muestra en la figura 2, la primera parte corresponde a las fuentes que son necesarias para poblar la o las bases de datos. Conectado a las fuentes se encuentra el siguiente proceso que corresponde a la extracción y transformación de la información de las fuentes, posterior a esto se encuentra el proceso de carga a la base de datos, es importante destacar que durante el proceso de carga de la información se debe limpiar y depurar los datos que provienen de las fuentes.

Dicho proceso es vital para asegurar la calidad de la información y la correcta toma de decisiones con información correcta y verificada.

No en todos los casos las aplicaciones para reportería toman los datos directamente del Data Warehouse ya que no es óptimo para el usuario final, sino que lo toman de los data marts que son almacenes o repositorios más pequeños dedicados a un area o tema específico. Una vez que tenemos la información en este punto los datos son accedados a través de las diferentes herramientas para el usuario final, de las cuales hablaremos más adelante. (Data Warehouse, s.f)

Figura 2. Diseño del proyecto



Fuente: Elaboración propia, 2019.

2.4 Business Intelligence

En la época en la que nos encontramos ha aparecido un término con mucha fuerza y que llevo para quedarse, Business Intelligence principalmente en las empresas, Business intelligence, inteligencia de negocios o también conocido como BI, puede definirse como un conjunto de métodos y técnicas con los cuales se busca transformar los datos en información entendible y valiosa para quienes se encargan de tomar las decisiones. En resumen, transformar los datos recopilados a través del tiempo en conocimiento que debería ser aplicado en las organizaciones.

Es importante tener claro qué objetivos se desean alcanzar con la implementación de un proyecto de este tipo, ya que no tener claro esto, puede limitar el desarrollo y uso de las diferentes

herramientas con las que se va a trabajar. La incorrecta selección de las herramientas también puede generar una limitante al no poder quizás mostrar los indicadores esperados por los usuarios analíticos.

Al final lo que se espera con el uso e implementación de estos proyectos es la toma de decisiones bien informadas para marcar diferencia en el mercado.

En un artículo publicado en el año 1958 por Hans Peter Luhn se utiliza el termino inteligencia de negocios por primera vez y lo define como: “La capacidad de comprender las interrelaciones de los hechos presentados de tal forma que consiga orientar la acción hacía una meta deseada” (Inteligencia empresarial, s.f)

Algunas fechas destacadas del inicio de BI en las empresas:

- 1956: IBM inventa las unidades de disco duro, que llegaron a particulares y empresas de la mano de los PCs.
- 1958: el investigador de IBM Peter Luhn, utiliza el concepto Business Intelligence.

Sus investigaciones fueron cruciales para entender el potencial del análisis de datos para mejorar las decisiones de negocio. Se le considera el padre del término inteligencia de negocio.

- Década de 1970: SAP y otras empresas desarrollan aplicaciones de negocio que facilitaban la introducción de datos en bases de datos.
- Década de 1980: las bases de datos evolucionan y permiten almacenar datos de diversas fuentes en una única base de datos.
- Posteriormente en los años 90, se da una alzada en el desarrollo y comercialización de nuevas y más herramientas de inteligencia de negocios, esto debido a que las empresas pioneras empezaron a darse cuenta del potencial de este tipo de soluciones y deciden sacarle provecho.

El problema de las primeras herramientas de Business intelligence era que eran difíciles de usar, poco intuitivas y muy costosas, sobre todo. (Business Intelligence, Inteligencia de Negocios o Inteligencia Empresarial, 2019)

Figura 3. Business intelligence.



Recuperado de: <https://blog.euncet.es/business-intelligence-como-ayuda/>

A finales de los años 90 e inicios del 2000 fue cuando realmente se comenzó a entender el potencial del uso de las herramientas de Business Intelligence, las empresas fabricantes de software crecieron, con el pasar de los años las herramientas empezaron a mejorar, se hicieron más fáciles de usar para cualquier profesional sin necesidad de tener conocimientos informáticos.

El verdadero reto nació cuando se da la necesidad de integrar soluciones de BI rápidas y fáciles de utilizar para usuarios analíticos y no técnicos y que a su vez fuesen seguras y confiables.

En los últimos años la información ha crecido de manera exponencial llegando al punto de que en un año se ha producido más información que en los anteriores cinco mil, esto en el 2007.

En años recientes el uso de herramientas de BI ha aumentado de manera indudable en las empresas, así como también las fuentes de información papel, redes sociales, internet en fin un sin número de fuentes con información valiosa que debe ser aprovechada por las organizaciones para obtener ventaja sobre sus competidores.

2.4.1 Usos de herramientas para Business intelligence:

- Para gestión de datos: son utilizadas desde la depuración y estandarización de los datos provenientes de diversas fuentes hasta su extracción, transformación y carga a un determinado sistema de almacenamiento.
- Para recolección de datos: para minería de datos o data mining, sobre datos nuevos o ya existentes para aplicar diferentes técnicas de análisis.
- Para reportería: una vez que se tiene toda la información ayudan a visualizar e integrar de manera intuitiva y grafica dicha información. (¿Qué es Business Intelligence (BI) y qué herramientas existen?, 2017)

2.4.2 ¿Como puede BI ayudar a las empresas?

- Optimización del stock

Los sectores con una marcada estacionalidad generalmente tienen problemas para optimizar correctamente el stock. Por ejemplo, si las ventas de algún producto en específico se disparan en verano o durante las fiestas de fin de año, resulta difícil almacenar la cantidad correcta para maximizar los beneficios.

En base a esto, algunas empresas del sector de enlatados o alimento en general han logrado aumentar la rentabilidad cerca del 10% mediante técnicas de BI como, por ejemplo:

- Implementación de sistemas para la toma de decisión (DSS).
- Análisis detallado de datos históricos de ventas e inventarios de productos.

Los datos recopilados han hecho posible, un rediseño más eficaz y rentable en todo el proceso logístico y de almacenamiento del producto. (Calvo, 2019)

- Fidelización de clientes

Los procesos de BI son de suma importancia para detectar cuales clientes son más rentables, para ello es necesario el análisis correcto de una gran cantidad de datos para obtener un perfil:

- Edad
- Genero
- Ubicación geográfica
- Estado civil
- Cantidad de dependientes

Una buena forma de conseguir esta información puede ser crear una tarjeta de cliente frecuente o tarjeta de descuentos donde el cliente tenga que registrarse para obtener esta tarjeta.

- Reducción de costos.

Permite tener un mejor control y supervisión de los gastos y costos que tiene la organización en los diferentes procesos en los que se aplica la solución de BI.

- Ayuda a establecer metas reales.

La aplicación de business intelligence ayuda a los responsables de la toma de decisiones a establecer metas realistas, pudiendo observar y analizar la data histórica que se tiene y compararlos también con los datos actuales para de esta forma establecer metas y proyecciones posibles de alcanzar.

- Permite la mejora continua.

Implementar una herramienta de BI permitirá utilizar la información con la que se cuenta para convertirla en acciones importantes e inteligentes, permite también encontrar puntos débiles y errores y convertirlos en oportunidades y fortalezas para la organización.

2.5 Bases de datos OLTP y OLAP

Figura 4. OLTP vs OLAP.



Recuperado de <https://www.educba.com/oltp-vs-olap/>

El término OLTP puede definirse como una herramienta que es capaz de soportar el procesamiento, administración y mantenimiento diario de transacciones generadas por el negocio. Por sus siglas en inglés On Line Transaction Processing, corresponde a bases de datos orientadas al procesamiento de transacciones.

En la actualidad los sistemas y software gestores de datos son de máxima importancia para un mercado cada vez más activo en las operaciones comerciales y transaccionales.

Las aplicaciones OLTP generalmente se utilizan para almacenar datos o actualizar los ya existentes, con numerosos usuarios con consultas y actualizaciones de datos frecuentes y que requieren tiempos de respuesta rápidos. Por ejemplo, los sistemas OLTP de cajeros automáticos de las entidades bancarias.

Una consideración de los sistemas OLTP es que por su arquitectura son rápidos en la inserción y actualización, pero que son lentos cuando se sobrecargan de consultas, principalmente si son históricas. (Sistemas OLTP: procesamiento, administración y mantenimiento de transacciones, 2016)

2.5.1 ¿Qué beneficios me aportan los sistemas OLTP?

Según la EAE Business School (2018), los sistemas OLTP aportan una gran cantidad de beneficios para las empresas, de los cuales vamos a enumerar algunos.

- Optimización de la tecnología: la capacidad de optimizar y simplificar la administración de transacciones es una de las características más destacadas de los sistemas OLTP, incrementar la rapidez y precisión en la obtención de resultados es clave en los sistemas de este tipo.
- Menor esfuerzo: La prevención de problemas o anomalías de actualización, reduce los esfuerzos requeridos para llevar a cabo modificaciones y asegurar la consistencia de los datos que se emplean en las transacciones.

2.5.1 Características de los datos transaccionales.

Requisito	Descripción
Normalización	Muy normalizados
Esquema	Esquema durante la escritura, altamente aplicado
Coherencia	Coherencia alta, garantías ACID
Integridad	Integridad alta
Usa transacciones	Sí
Actualizable	Sí
Anexable	Sí
Carga de trabajo	Grandes escrituras, lecturas moderadas
Indización	Índices principales y secundarios
Tamaño de los datos	Pequeño a mediano tamaño
Modelo	Relacional
Forma de los datos	Tabular
Flexibilidad de consulta	Muy flexible
Escala	Pequeño (MB) a grande (algunos TB)

Tabla 2. Procesamiento de transacciones en línea.

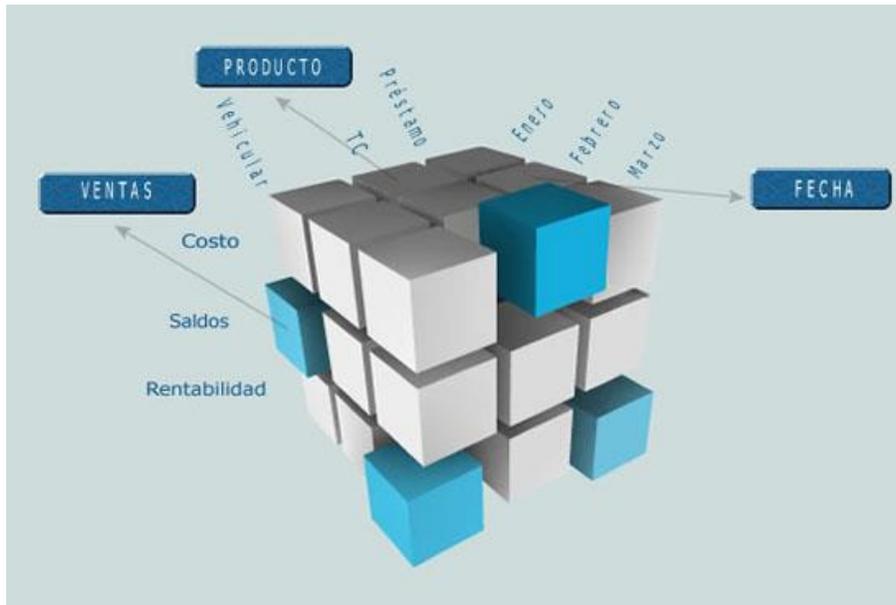
Fuente: (Procesamiento de transacciones en línea (OLTP), 2019)

2.6 Modelo multidimensional

Por facilidad de análisis de datos, los Data warehouse presenta los datos utilizando modelos multidimensionales, el modelo multidimensional acoge dos términos de suma importancia:

- **Medidas:** Una medida constituye el que analizar y son variables numéricas.
- **Dimensiones:** Las dimensiones responden al cómo y son variables cualitativas.

Figura 4. Sistemas OLAP



Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

Fuente: (Marker, 2019)

Cada una de las dimensiones aporta un campo determinado que posteriormente podrá ser comparado con la información almacenada en las otras dimensiones.

Gracias a la incorporación de las bases de datos multidimensionales y el nacimiento del nuevo concepto de Cubo OLAP, las herramientas de soluciones para sistemas Business Intelligence han avanzado notablemente en cuanto a las prestaciones que estas aplicaciones brindan a las organizaciones, donde la información es precisa, confiable y lista en el momento oportuno.

Las bases de datos multidimensionales pueden contener uno o más cubos.

Una característica importante de las herramientas OLAP es que brinda una solución ideal para llevar a cabo consultas a los datos de manera sencilla y veloz.

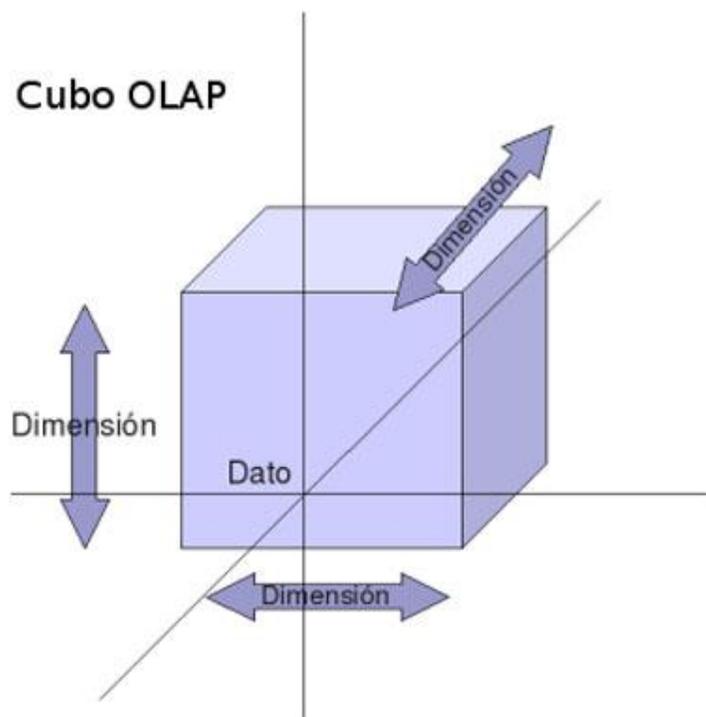
El objetivo principal de OLAP es agilizar de manera notable las consultas y evaluaciones de la gran cantidad de datos que producen diariamente las compañías, con la información proveniente de todos los departamentos o áreas de esta. (Marker, 2019)

Debido a su forma de funcionamiento y almacenamiento de la información, cuando los usuarios requieren realizar modificaciones en la estructura de este tipo de base de datos, deben rediseñar el Cubo OLAP, sin posibilidades de poder utilizar la estructura en la que se trabajó hasta el momento.

Para tener una idea más simple de la función de los Cubos OLAP dentro de una base de datos multidimensional, cabe destacar que cada una de las dimensiones o escalas del cubo OLAP corresponde básicamente a una jerarquía de datos. (Cubos OLAP de información para la toma de decisiones, 2016)

Un ejemplo claro de ello podría ser el siguiente caso: dentro de una escala temporal para incluir datos determinados a un periodo de tiempo, que llevará el nombre de “enero de 2019”, seguramente incluirá una dimensión denominada “Primer Trimestre de 2019”, la cual además incluirá otra dimensión llamada “Año 2019” y así sucesivamente, de acuerdo con las necesidades de cada empresa. (Marker, 2019)

Figura 5. Cubo OLAP



Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

Asimismo, también pueden utilizarse otras dimensiones del cubo para recabar información referente a situaciones geográficas, clasificación de los productos por categorías, gastos realizados por la empresa, y demás.

Esta confluencia de la información permite llevar a cabo un análisis completo de diversas situaciones, para hallar las soluciones correctas a los problemas de los negocios.

2.7 Un cubo OLAP para cada necesidad.

Gracias al auge y a la gran importancia que han tenido las herramientas OLAP, han nacido diferentes sistemas basados en su arquitectura que pueden brindar respuestas o soluciones distintas a las organizaciones.

De acuerdo con su estructura y funcionamiento los sistemas OLAP se han clasificado en distintas categorías, ROLAP, MOLAP, HOLAP, WOLAP, DOLAP, RTOLAP y SOLAP.

Figura 6. Sistema para cada necesidad.



Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

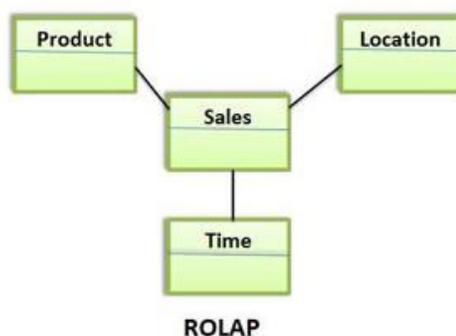
2.7.1 ROLAP

Significa Procesamiento Analítico Online Relacional y es uno de los sistemas OLAP más utilizados en la actualidad cuya característica principal es que se trata de un sistema capacitado para realizar manejo de gran cantidad de volúmenes de datos, aunque su principal deficiencia es que ofrece un rendimiento un tanto menor durante las consultas en comparación con los otros sistemas. Corresponde, básicamente a un sistema que ha sido construido utilizando herramientas OLAP y una base de datos relacional lo que permite un tiempo de carga menor que otros métodos basados en sistemas OLAP.

Otra de sus ventajas es que permite a cualquier usuario acceder de manera sencilla a la información contenida en la base de datos relacional, utilizando herramientas para la creación de reportes SQL, sin la necesidad de aplicaciones tipo OLAP.

En la actualidad, las herramientas más utilizadas del tipo ROLAP son Microsoft Analysis Services, MicroStrategy, Business Objects y el servidor Open Source denominado Mondrian. (Marker, 2019)

Figura 7. Cubo Rolap



Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

2.7.2 MOLAP

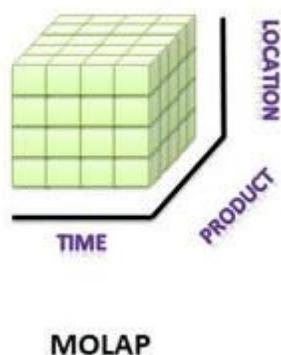
Es la gran competencia a ROLAP, sus siglas corresponden a Multidimensional Online Analytical Processing, es decir, corresponde a una herramienta que permite el procesamiento analítico multidimensional online.

Al utilizar una base de datos multidimensional, permite alcanzar un tiempo de respuesta superior y algunos sistemas MOLAP permiten comprimir datos con la finalidad de reducir bastante el espacio que utilizan en los discos.

Como uno de los inconvenientes es que puede presentarse redundancia en los datos y el inconveniente principal es que presenta dificultad para cargar gran cantidad de volumen de datos.

Algunas de las herramientas para sistemas MOLAP son Oracle OLAP, Microsoft Analysis Services, Essbase. (Marker, 2019)

Figura 8. Cubo Molap



Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

2.7.3 HOLAP

Corresponde a la combinación de sistemas ROLAP y MOLAP, sus siglas corresponden a Hybrid Online Analytical Process, queriendo decir que se trata de un sistema de procesamiento analítico híbrido Online.

Toma las características que más ventajas aportan de MOLAP y ROLAP, utiliza dos tipos diferentes de bases de datos para almacenar la información, una multidimensional y otra relacional, y son utilizadas dependiendo de los requerimientos de los datos que deben ser procesados.

Actualmente es uno de los sistemas más utilizados por muchísimas empresas ya que presenta ventajas debido al uso de las dos bases de datos. Si en la fusión de los sistemas MOLAP y ROLAP se maneja de manera correcta permite lograr una mayor velocidad de procesamiento de la información.

Como derivaciones de los sistemas OLAP se encuentran adicionalmente se encuentran WOLAP que es orientado a la web, DOLAP que es en base a aplicaciones de escritorio, RTOLAP que es una herramienta OLAP en tiempo real y por último SOLAP que es un sistema OLAP enfocado en lo espacial. (Marker, 2019)

3. Arquitectura de Business Intelligence.

3.1 William Inmon.

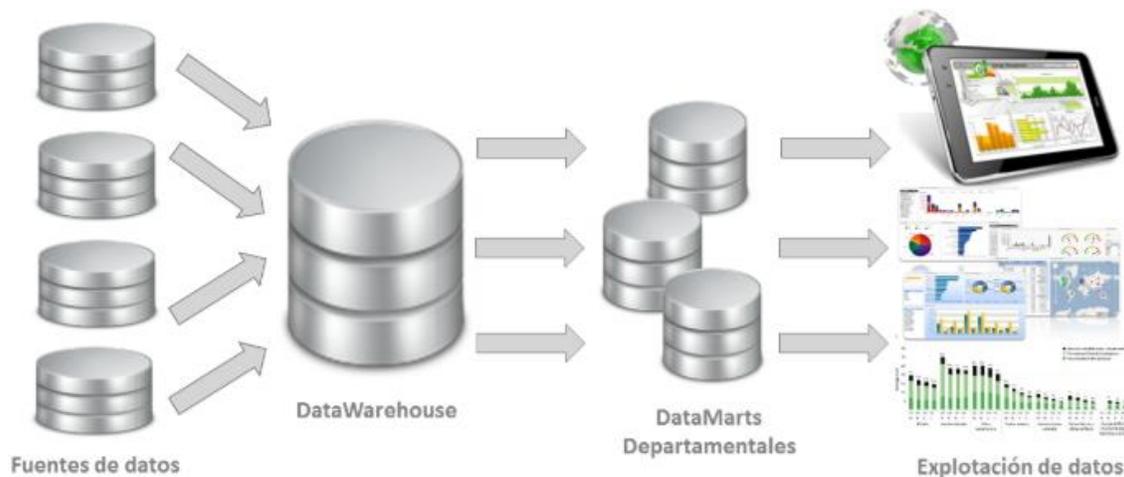
Según, William Inmon define un Datawarehouse como un almacén de datos único y global de todas las empresas.

En este enfoque, la premisa es que la información se almacene al máximo nivel de detalle, esto para garantizar la futura exploración de los datos, manteniéndose sin cambiar y no volátil, de

manera que los cambios que puedan sufrir los datos a través del tiempo quedan almacenados sin que puedan modificarse o eliminarse.

El Datawarehouse centraliza todos los datos de la compañía que a su vez alimenta pequeños Datamarts de áreas específicas, que serán el punto de alimentación para las diferentes herramientas de reportería, en este modelo se puede decir que cada departamento tendrá su DataMart abastecido con la información del Warehouse.

Figura 9. Inmon



Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

También suele conocerse como una metodología de trabajo “Top-Down” ya que primero se centra en una visión general de la empresa, para luego ir desgranándola en pequeños datos para los departamentos. (Dertiano, 2015)

3.1.1 La estructura del Data Warehouse

Según (Dertiano, 2015), indica que en cuanto a la estructura interna del Datawarehouse, para Inmon la prioridad es que el modelo de datos esté construido en la tercera forma normal. Brevemente esto quiere decir que, el proceso de normalización consiste en aplicar una serie de reglas a la hora de establecer las relaciones entre los diferentes objetos de la base de datos. Con realizar la normalización se consiguen muchos beneficios, como, por ejemplo, evitar la redundancia de los datos, mantener su integridad referencial, facilitar el mantenimiento de las tablas y disminuir el tamaño de la base de datos. Sin embargo, a diferencia de los DW desnormalizados, las consultas exigen el empleo de consultas mucho más complejas, lo que dificulta el análisis directo de la información y el uso de las herramientas de reportería.

3.2 Ralph Kimball

Kimball define su metodología como “Bottom-Up”, esto significa que se inicia la construcción a partir de pequeños componentes para luego ir incrementando o evolucionando a estructuras más grandes, para Ralph un Datawarehouse no es más que la unión de todos los datamarts que pueda tener una compañía.

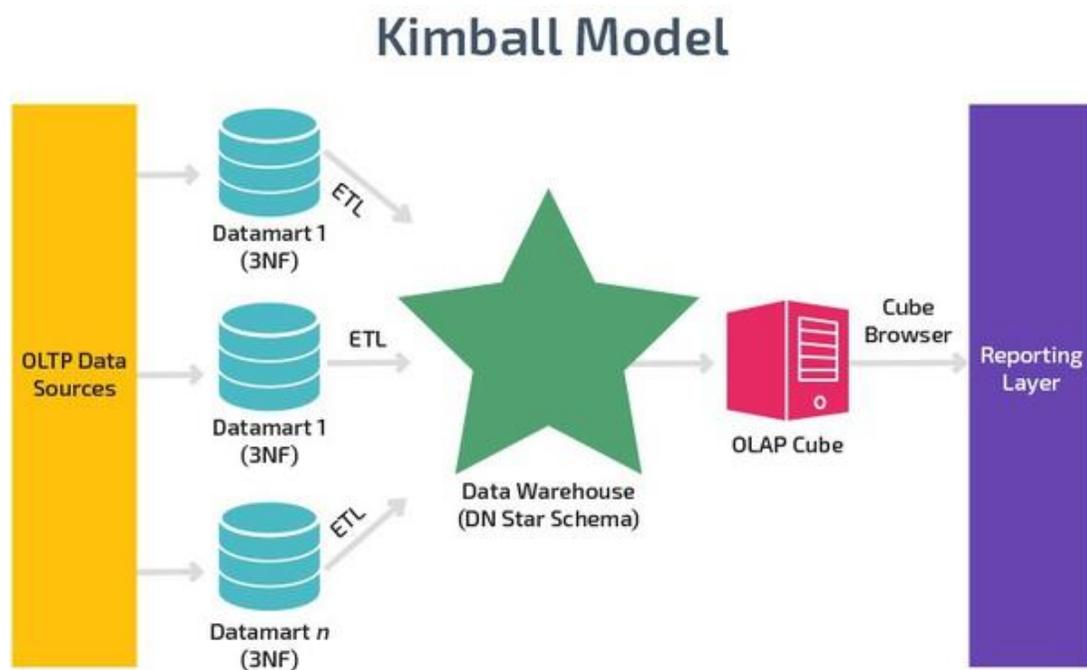
Su filosofía nace con la idea de que los Warehouse se origina por el interés, necesidad y esfuerzo de un departamento cualquiera, a medida de que otros departamentos se enteran de esto y entienden que también tienen la necesidad de tener sus propios datamarts, estos se irán relacionando entre si manteniendo una estandarización que Kimball denomina “dimensiones conformadas” que corresponden a dimensiones comunes entre los departamentos, es un punto clave decir que estas dimensiones que comparten los datamarts que se tengan deben tener garantizada la integridad. Para alcanzar esto es importante que las dimensiones tengan un estándar, un diseño consistente para ser usado por todos los datamarts.

La principal ventaja de este enfoque de almacén de datos es que, al estar formado por pequeños datamarts estructurados en modelos de datos dimensionales (esquemas de estrella o copo de nieve), especialmente diseñados para la consulta y generación de informes, el Datawarehouse al completo puede ser explotado directamente por las herramientas de Reporting y análisis de datos sin la necesidad de estructuras intermedias.

En cuanto a las cuestiones sobre la granularidad, a pesar de que este tipo de Datawarehouse suele presentar los datos agregados en base a las consultas e informes que haya que generar, Kimball insiste en la necesidad de que estas agregaciones estén complementadas con datos a mayor nivel de detalle.

El argumento es que las preguntas de negocio que puedan llegar a hacer los usuarios son impredecibles, de manera que el Datawarehouse tiene que estar preparado para dar respuesta a todas ellas, garantizando la exploración de los datos y la navegación a través de jerarquías desde datos agregados hasta información desagregada. (Dertiano, 2015),

Figura 10. Modelo Kimball.



Recuperado de <https://www.quora.com/Dimensional-modeling-as-popularized-by-Ralph-Kimball-is-perhaps-the-most-popular-data-warehouse-implementation-methodology-Yet-some-question-the-scalability-of-the-Kimballs-methodology-What-are-some-disadvantages-to-the-K-method>

3.3 Comparación entre Kimball e Inmon.

Existen marcadas divergencias entre ambos modelos a la hora de establecer los pasos a seguir para desarrollar un Data Warehouse. Como habréis podido observar tras exponer en entradas anteriores los enfoques de Inmon y Kimball, sus principios son tan dispares que no solo la estructura interna y alcance de éstos presentan variaciones, sino que también su intención y finalidad se ven afectados.

Sin embargo, a pesar de esta confrontación de enfoques e ideas y de las grandes diferencias que presentan ambos modelos, resulta muy atrevido decir que uno u otro es correcto o incorrecto ya que, según sea el caso en el que nos encontremos, nos puede encajar en mayor o menor medida implantar una de las dos perspectivas.

Aspectos para analizar

Por esta razón, hay una serie de aspectos que habrá que analizar antes de decantarnos por una de las opciones:

- Presupuesto para afrontar el proyecto.
- Plazos disponibles para la construcción del Datawarehouse.
- Conocimiento requerido para el equipo de desarrolladores.
- Alcance del Datawarehouse, ya sea para albergar los datos de toda la compañía o de determinadas áreas de negocio o departamentos.
- Complejidad de las labores de mantenimiento. (Dertiano, 2016)

Inmon		Kimball
Presupuesto	Coste inicial alto	Coste inicial bajo
Plazos	Requiere más tiempo de desarrollo	Tiempo de desarrollo inferior
Conocimiento	Equipo con especialización alta	Equipo con especialización media
Alcance	Toda la compañía	Departamentos individuales
Mantenimiento	Fácil mantenimiento	Mantenimiento más complejo

Tabla 3. Comparación Inmon-Kimball

Fuente: (Dertiano,2016)

3.4 Inmon y Kimball

Si recordamos lo expuesto en entradas anteriores, el *Datawarehouse* de Kimball está orientado a la consulta de la información, por lo que su estructura interna está especialmente diseñada para garantizar una explotación de los datos rápida y sencilla, no requiriendo usuarios especializados para ello. Por el contrario, el *Datawarehouse* de Inmon persigue la integración de todos los datos de la compañía, estando orientado hacia el almacenaje de grandes volúmenes de datos, por lo que su estructura interna normalizada se diseña para evitar la redundancia de datos, simplificar las labores de mantenimiento, etc. cuestiones que complican las consultas de la información, requiriendo que los usuarios finales estén mucho más especializados.

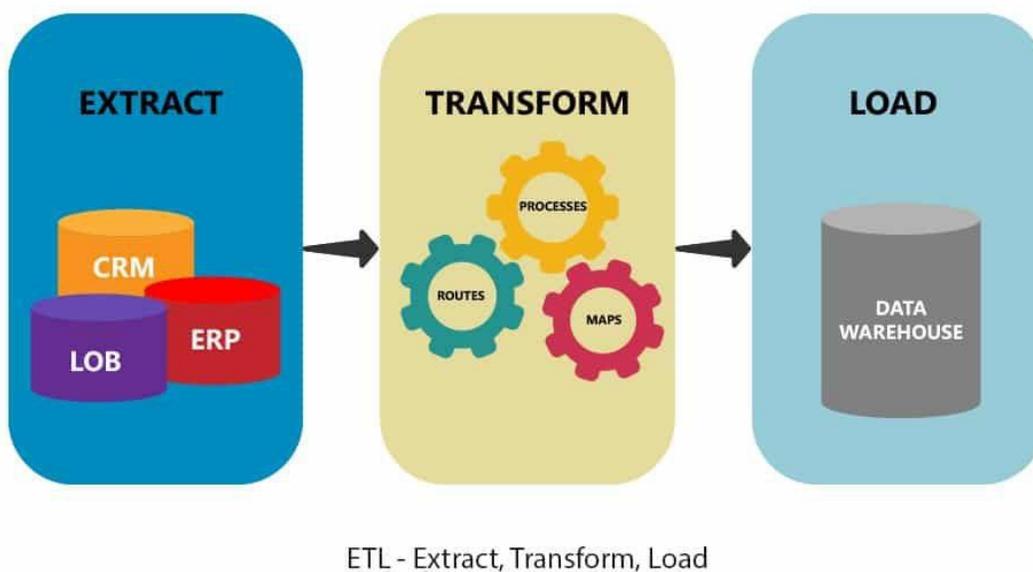
Así, podríamos decir que el enfoque de Kimball se ajusta más a proyectos pequeños en los que se persiga un sistema fácilmente explotable y entendible por el usuario y de rápido desarrollo, siendo el modelo de Inmon más apropiado para sistemas complejos de mayor envergadura. Todo proyecto tiene sus propias peculiaridades, siendo cada caso único e independiente, por lo que resulta necesario llevar a cabo un estudio de todas ellas antes de decantarnos por una solución u otra, de forma que podamos hacernos una idea sobre qué modelo se ajusta mejor a las condiciones de nuestro proyecto. (Dertiano, 2015)

4. ETL

La palabra ETL corresponde a:

- E: Extract (Extracción)
- T: Transform (Transformación)
- L: Load (Carga)

Figura 11. ETL, Extracción, Transformación, Carga



Recuperado de: <https://www.astera.com/es/tema/integraci%C3%B3n-de-datos-en-la-nube/herramientas-etl-y-lo-que-es-etl/>

El término de ETL tomó importancia en los años 70, cuando, las organizaciones comenzaron a utilizar diferentes repositorios de datos para almacenar diferentes tipos de información. Surgió la necesidad de integrar los datos y ETL se convirtió en el método universal para extraer datos de las diferentes fuentes y transformarlos antes de que sean cargados a un destino específico.

Aquí las tres fases del proceso de ETL:

Fase 1: Extracción

El objetivo de un proceso ETL es producir datos limpios y accesibles que puedan utilizarse para analíticas u operaciones comerciales. Los datos en bruto deben extraerse de una variedad de fuentes, por ejemplo:

- Bases de datos existentes
- Registros de actividad como el tráfico de red, informes de errores, etc.
- Rendimiento y anomalías de aplicaciones
- Incidencias de seguridad
- Otras actividades transaccionales que deben comunicarse para dar cumplimiento normativo
- Los datos extraídos en ocasiones se transfieren a otro destino como por ejemplo un data lake o un almacén de datos.

Fase 2: Transformación

La fase de transformación de ETL es donde se produce la operación más crítica. El resultado más destacado de la transformación pasa por aplicar las normas necesarias del negocio a los datos para cumplir con los requisitos de la organización. La transformación modifica los datos en bruto para que presenten los formatos de notificación correctos. Si los datos no se limpian, resulta más complicado aplicar las normas comerciales de notificación.

La transformación se efectúa mediante una serie de normas y reglamentos que se esbozan. Estos son algunos de los estándares que garantizan la calidad de datos y su accesibilidad durante esta fase:

- Normalización: definir qué datos entrarán en juego, cómo se formatearán y almacenarán, y otras consideraciones básicas que definirán las etapas sucesivas.
- Eliminación de duplicados: notificar los duplicados a los administradores de datos; excluyendo y/o eliminando los datos redundantes.
- Verificación: ejecutar comprobaciones automatizadas para cotejar información similar, como tiempos de transacción o registros de acceso. Las tareas de verificación permiten

seguir cribando los datos no utilizables y pueden alertar sobre anomalías en sus sistemas, aplicaciones o datos.

- **Clasificación:** maximizar la eficiencia de los almacenes de datos agrupando y clasificando elementos como los datos en bruto, audios, archivos multimedia y otros objetos en categorías. Las normas de transformación determinarán cómo se clasifica cada dato y dónde se trasladará a continuación. A menudo el proceso de ETL se emplea para crear tablas de agregación a efectos de informes resumidos. Para ello debemos clasificar y, posteriormente, agregar.

Fase 3: Carga

La última fase de un proceso de ETL típico es la carga de esos datos extraídos y transformados a su nuevo destino. Existen dos vías habituales de cargar los datos a un almacén de datos: la carga completa y la carga incremental.

La ejecución de un ciclo aislado de ETL o de una serie de ellos programada puede realizarse ejecutando una tarea desde una línea de comando o una interfaz GUI. Sin embargo, debemos estar atentos a varios frentes. Gestionar las excepciones, por ejemplo, puede resultar un proceso muy farragoso. Muchas veces las extracciones de datos pueden fallar cuando hay incidencias en uno o varios sistemas. La existencia de datos erróneos en un sistema puede afectar a datos que estén siendo extraídos de otro, por lo que el seguimiento y la gestión de errores son actividades fundamentales. (Pearlman 2019)

4.1 ¿Para qué sirve este proceso?

El ETL es un método potente que puede trabajar junto a herramientas de gestión e integración de datos para cumplir con los objetivos de la compañía. Algunos casos de usos incluyen:

- Migración de datos desde diferentes sistemas.
- Consolidación de datos como consecuencia de una fusión empresarial.

- Recolección y fusión de datos desde proveedores o partners externos.
- Integración de nuevas fuentes de datos como social media, videos, dispositivos conectados a internet de las cosas, entre otras.
- Analítica “Self-Service” para ofrecer la posibilidad de hacer decisiones basadas en los datos a perfiles de negocio y sin conocimientos técnicos.
- Integrarse con herramientas de *Data Quality* para asegurar que los datos sean confiables.
- Trabajar con metadatos para permitir la trazabilidad de los datos.
- Integrarse con sistemas transaccionales, almacenes de datos operativos, plataformas de Business Intelligence y sistemas de Master Data Management (MDM).

En resumen, los beneficios principales que una herramienta ETL puede proporcionar a las empresas son:

- Permitir extraer y consolidar datos de múltiples fuentes.
- Proporcionar un contexto histórico profundo sobre la empresa y negocio.
- Facilita el análisis y el reporte de datos de una forma sencilla y eficiente, mediante representación visual.
- Aumentar la productividad y facilitar el trabajo en equipo.
- Permitir adaptarse a la evolución de las tecnologías e integrar nuevas fuentes de datos con las tradicionales.
- Permitir la toma de decisión estratégica basadas en datos por parte de los directivos de la empresa. (Carisio, 2019)

4.2 Desafíos de los procesos ETL

Los procesos ETL son fundamentales para cada empresa. Sin embargo, se encuentran con importantes retos que tienes que superar para adaptarse a las nuevas necesidades:

- **Procesamiento de datos en tiempo real.** Cada día más se necesita tomar decisiones con mayor velocidad, lo que contrasta con el funcionamiento en *batch* de los sistemas ETL tradicionales, que tiene que adecuarse para operar lo más cercano posible al tiempo real.
- **Aumentar la velocidad del procesamiento de datos.** El aumento tanto de la cantidad como de la complejidad de los datos dificulta a veces las tareas de transformación. En este sentido, nació el concepto de ELT (*Extract, Load and Transformation*), que pospone la transformación en último lugar, realizándose ya en el sistema destino, y aprovechando de la potencia de cálculo del motor de la base de datos.
- **Integración de nuevas fuentes de datos.** En la actualidad, las empresas necesitan acceder a todo tipo de fuentes de datos heterogéneas: videos, redes sociales y hasta datos generados por máquinas (Internet de las cosas). Por esto, las herramientas ETL necesitan evolucionar y agregar nuevas transformaciones para soportar estas nuevas fuentes de datos y las que vendrán en futuro. (Carisio, 2019)

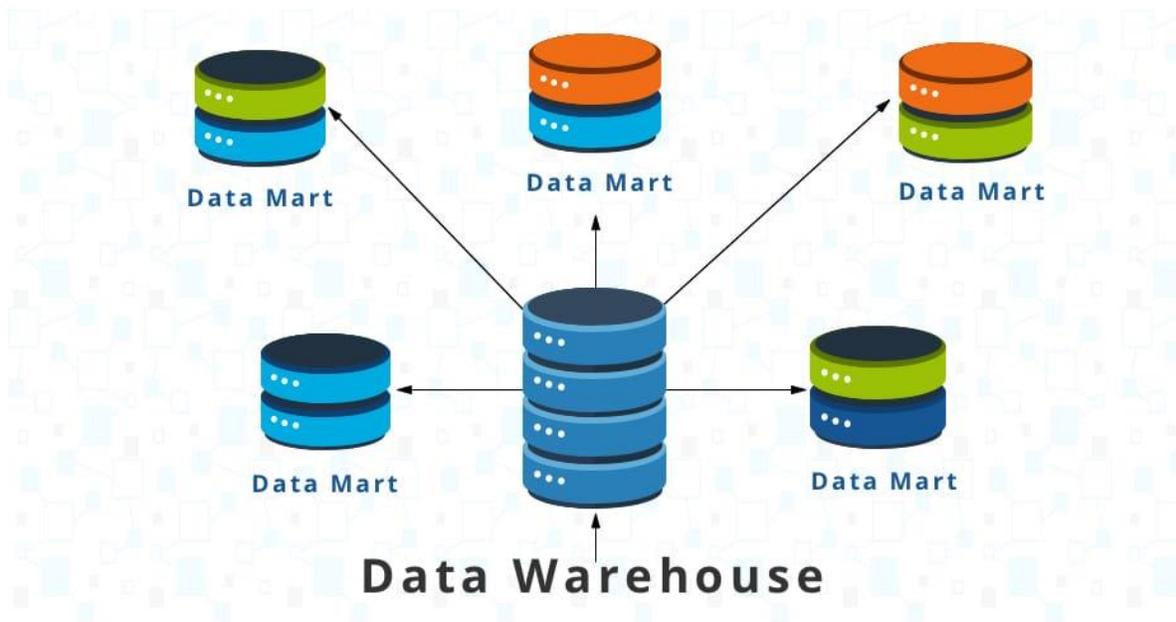
5. Datamart

Un datamart corresponde a una base de datos departamental, dedicada o especializada en un area del negocio en específico, también es conocido como una base de información departamental. Este almacén o repositorio permite que las organizaciones puedan acceder a datos vitales de forma sencilla.

Su función principal es ayudar a que un area particular dentro del negocio pueda tomar mejores decisiones. Los datamarts están orientados a la consulta, es consumido o consultado por herramientas OLAP que ofrecen una visión multidimensional de los datos.

De alguna manera podemos decir que los datamarts son segmentos de un Data Warehouse especializados en un área particular de la organización. (¿En qué consiste un data mart?, 2019)

Figura 12. Data mart frente a almacén de datos



Recuperado de: <https://es.talend.com/resources/what-is-data-mart/>

5.1 El futuro de los data marts es en la nube.

Incluso con la flexibilidad y eficiencia mejoradas que ofrecen los data marts, los sistemas de big data (y las grandes empresas) siguen creciendo demasiado para muchas soluciones locales. A medida que los almacenes de datos y los data lakes dan el salto a la nube, también lo hacen los data marts.

Con una plataforma cloud o en la nube compartida para crear y albergar datos, el acceso y los análisis ganan considerablemente en eficiencia. Se pueden crear clústeres temporales de datos para un análisis a corto plazo o combinar clústeres de largo recorrido para un trabajo más continuo. Las

tecnologías modernas también están separando el almacenamiento de los datos del procesamiento, lo que permite la máxima escalabilidad de cara a realizar consultas a los datos.

A continuación, se listan algunas ventajas más de los datamarts dependientes e híbridos basados en la nube:

- Arquitectura flexible con aplicaciones nativas en cloud.
- Único depósito que contiene todos los data marts.
- Recursos consumidos a demanda.
- Acceso en tiempo real a la información.
- Aumento de la eficiencia.
- Consolidación de recursos que reduce los costes.
- Analítica interactiva en tiempo real. (Pearlman, 2019)

5.2 Comparativa entre Data Warehouse y Data Mart

A continuación, vamos a ver más en detalle la diferencia entre Data mart y Data Warehouse.

criterio	Data Warehouse	Data mart
Uso	Ayuda a tomar una decisión estratégica.	Ayuda a tomar decisiones tácticas para el negocio.
Objetivo	El objetivo principal de Data Warehouse es proporcionar un entorno integrado y una imagen coherente de la empresa en un momento determinado.	Utilizado principalmente en una división de negocios a nivel de departamento.
Diseño	Complejo	Sencillo
Modelo dimensional	Puede o no puede usarse en un modelo dimensional. Sin embargo, puede alimentar modelos dimensionales.	Se construye enfocado en un modelo dimensional usando un esquema de inicio.
Gestión de datos	Incluye una gran área de las compañías, por lo que se tarda mucho tiempo en procesarla.	Son fáciles de usar, diseñar e implementar, ya que solo puede manejar pequeñas cantidades de datos.
Enfoque	El Data Warehouse se enfoca ampliamente en todos los departamentos. Es posible que incluso pueda representar a toda la empresa.	Data mart está orientado a un área de negocio y se utiliza a nivel de departamento.

Tipo de datos	Los datos almacenados en Data Warehouse siempre ofrecen más detalle en comparación con data mart.	Los Data Marts están contruidos para grupos de usuarios particulares. Por lo tanto, los datos son cortos y limitados.
Normalización	Los almacenes modernos están en su mayoría desnormalizados para proporcionar consultas de datos más rápidas y un buen rendimiento de lectura	No hay preferencia entre una estructura normalizada o desnormalizada.
Área temática	El objetivo principal de Data Warehouse es proporcionar un entorno integrado y una imagen coherente de la empresa en un momento determinado.	En su mayoría tienen solo un área temática, por ejemplo, detalle de ventas.
Almacenamiento de datos	Diseñado para almacenar datos de decisiones de toda la empresa, no solo datos de marketing.	Modelado dimensional y diseño de esquema en estrella empleado para optimizar el rendimiento al acceder a la información.
Tipo de datos	La variación en el tiempo y el diseño no volátil se aplican estrictamente.	Principalmente incluye estructuras de datos de consolidación para satisfacer las necesidades de informes y consultas del área temática.
Valor de los datos	Solo lectura desde el punto de vista de los usuarios finales.	Datos transaccionales agrupados alimentados directamente desde el Data Warehouse.
Alcance	Es más útil ya que puede traer información de cualquier departamento.	Data mart contiene datos, de un departamento específico de una empresa. Puede que haya Data Marts separados para ventas, finanzas, marketing, etc. Tiene un uso limitado
Fuente	Los datos provienen de muchas fuentes.	Los datos provienen de muy pocas fuentes.
Tamaño	El tamaño del Data Warehouse puede variar de 100 GB a más de un TB.	El tamaño de Data mart es inferior a 100 GB.
Tiempo de implementación	El proceso de implementación de Data Warehouse puede extenderse de meses a años.	El proceso de implementación de Data mart está restringido a unos pocos meses.

Tabla 4. Comparación entre Data Warehouse y Data Mart

Fuente: (Acosta, 2019)

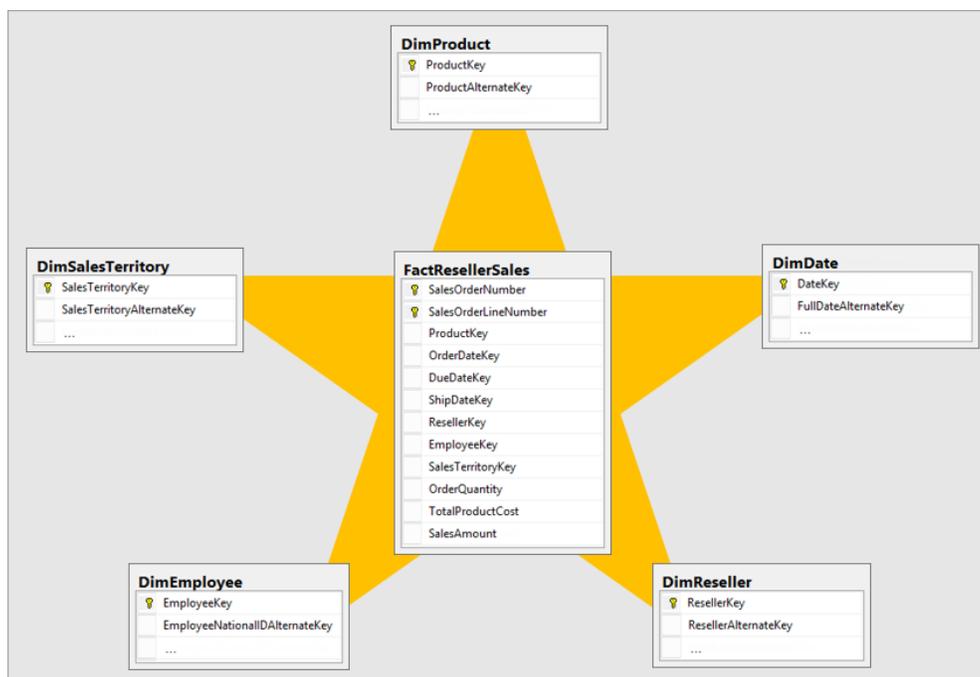
6. Esquema Estrella

Primeramente, debemos saber que un esquema de base de datos es una descripción de los objetos que incluye tablas, índices, vistas.

Un esquema de estrella es aquel que tiene una tabla central, llamada Fact o TH que contiene los datos relacionados al análisis, rodeada de las dimensiones, por eso es por lo que se le denomina modelo de estrella por su similitud.

Es un sistema muy utilizado por su simplicidad, velocidad de análisis y excelente rendimiento.

Figura 13. Modelo estrella.



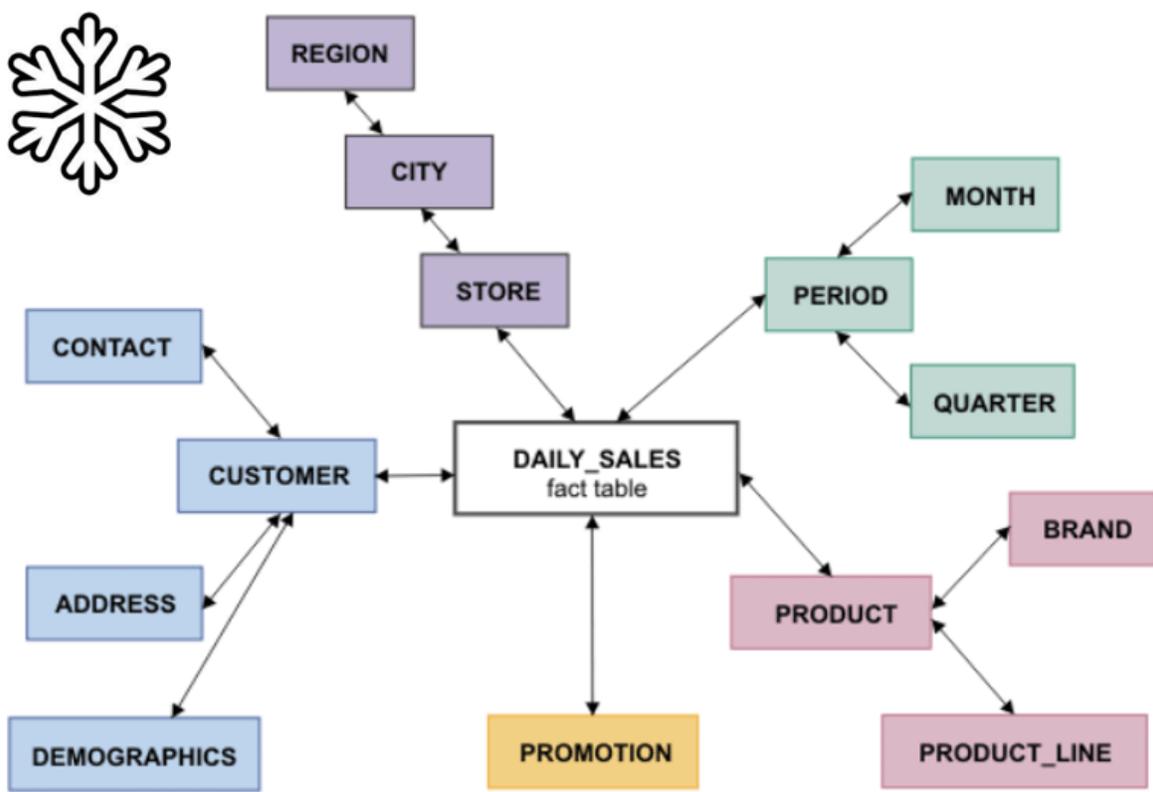
Recuperado de: <https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>

7. Copo de nieve

El modelo de copo de nieve es un esquema derivado del esquema de estrella, en el que las tablas de dimensión se normalizan en múltiples tablas, por esta razón la tabla de hechos deja de ser la única tabla que se relacionan con otras tablas y aparecen nuevas mediante joins.

Los Snowflake schem o copo de nieve son estructuras más complejas, se da cuando existen más de una TH o tabla Fact, sin que exista la posibilidad de reducir el número de tablas, parte de ello genera tener peor rendimiento al tener más joins con las dimensiones.

Figura 14. Modelo estrella



Recuperado de <https://blogs.solidq.com/es/business-analytics/data-warehouse-y-data-marts-esquema-en-estrella-11/>

Beneficios y aspectos negativos de cada tipo de esquema:

Modelo de estrella:

- Es más veloz y simple para el análisis.
- Para el usuario final las consultas no suelen ser complicadas.
- Es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que afecte el rendimiento de la base de datos.

Modelo copo de nieve:

- Con la normalización de las tablas se evita la redundancia de los datos y esto genera un ahorro de espacio.
- Generalmente se usan para Datawarehouse grandes y más complejos, pero debe tomarse en cuenta el factor tiempo ya que generalmente es crítico para los usuarios. (Ramírez, 2011)

Es importante antes de elegir una de las opciones, la labor de análisis y definición, una vez entendida la necesidad de los usuarios es más fácil adecuar y seleccionar la arquitectura que brinde mejor rendimiento.

8. Tabla de hechos

Corresponde a los indicadores del negocio, por ejemplo, en una compañía que se encarga de la venta de productos, los “hechos”, corresponden las ventas, las devoluciones y las compras de manera general, corresponde a las medidas numéricas que se incluirán en el cubo.

En el modelo o esquema de estrella es la tabla central, diferente en el modelo de copo de nieve.

Las tablas de hechos contienen llaves subrogadas de las dimensiones que definen el nivel de detalle y los indicadores, ningún otro campo o valor va en ellas. Es un tema importante definir y entender perfectamente que información queremos o debemos cargar y también el nivel de detalle de estos.

Por ejemplo, una tabla de hechos podría tener una estructura similar a esta:

ID_Fecha	ID_Cliente	Cantidad_Productos	Precio_Unitario	Descuento
----------	------------	--------------------	-----------------	-----------

Este ejemplo de tabla de hechos almacena datos de la fecha de la transacción, el cliente que realiza la compra, datos cuantitativos como la cantidad de productos, el precio unitario y el descuento en caso de que aplique. Toda esta es información importante, pero, por sí sola no aporta muchos datos para el análisis, para que tenga realmente un valor útil se necesitan las tablas de dimensión que contienen más información de los clientes, productos, entre otros. (Loshin,2018)

8.1 La importancia de las dimensiones para el análisis.

En la tabla de hechos o Fact table, los valores en los campos no son los datos “reales”, más bien, corresponden a llaves foráneas que apuntan o señalan a las claves primarias de las dimensiones, mismas que almacenan una importante cantidad de información de cada entidad a la que se hace referencia en la tabla de hechos.

Siguiendo con el ejemplo de la tabla anterior, la dimensión de cliente por ejemplo podría verse de la siguiente manera:

ID_Cliente	Nombre_Cliente	Provincia	Cantón	Distrito	Telefono
------------	----------------	-----------	--------	----------	----------

Sin duda, las tablas de dimensión pueden contener, muchísimos más atributos, más información geográfica, datos como el nacimiento, tipo de cliente, estado del cliente, y otras preferencias aportadas por el cliente. Una diferencia entre las tablas de dimensión y la tabla de hechos es que generalmente las dimensiones tienen un número corto de columnas mientras que las segundas son bastante amplias generalmente.

A manera de resumen podemos decir que,

Las tablas de hechos son:

- La tabla principal del modelo dimensional de la estrella.
- Contiene los campos llave que se relacionan a la dimensión.
- Contiene los datos cuantitativos del modelo o los valores numéricos.
- No permiten la redundancia de los atributos.
- Generalmente tiene millones de registros o muchos.

Las dimensiones son:

- Tablas desnormalizadas.
- La relación con las tablas de hechos se da mediante un campo clave.
- Almacenan la información importante de los campos llave almacenados en los hechos.
- No existe un límite definido para la cantidad de dimensiones.
- Las dimensiones pueden tener una o más relaciones jerárquicas.
- Generalmente no son tan extensas en cantidad de registros. (Loshin,2018)

9. Data Lake

Un data lake se puede definir como un repositorio de datos que contiene una gran cantidad de datos brutos y sin estructurar, tal cual vienen de la fuente, estos datos se mantienen en este repositorio hasta el momento en que se necesiten.

A los datos correspondientes al datalake no se le da ningún procesamiento previo, los datos provienen de diversas fuentes, ya sean bases de datos, documentos, registros de internet, redes sociales para ser analizados luego. (Balagueró, 2018)

Figura 15. ¿Como funciona un datalake?



Recuperado de <https://todobi.com/data-lakes-definiciones-y-plataformas/>

Figura 16. Aplicaciones más comunes.



Recuperado de <https://todobi.com/data-lakes-definiciones-y-plataformas/>

9.1 Diferencias entre un Data Lake y un Data Warehouse

- Un Data Lake conserva todos los datos:

Durante el desarrollo de un almacén de datos se gasta una cantidad importante de tiempo analizando las fuentes con las que se cuentan, entendiendo los procesos de negocio y alistando los datos. Como resultado, se obtiene un modelo de datos altamente estructurado que ya está listo para la generación de informes.

Una gran parte de este proceso incluye también la de toma de decisiones. Qué datos se van a incluir y cuáles no. Por lo general, si los datos no se utilizan para responder a preguntas específicas o no son imprescindibles en algunos informes serán excluidos. De esta manera, se simplifica el modelo y no desperdicia espacio.

Sin embargo, un Data Lake conserva todos los datos. No solo los que son vitales en ese preciso momento sino todos aquellos que están guardados que pueden hacer falta en algún momento. Lo que proporciona varios beneficios sobre todo en cuanto a la parte de análisis.

Este enfoque es posible porque el hardware para un Data Lake suele ser muy diferente del que se utiliza para un almacenamiento de datos regular. La comodidad, los servidores que hay disponibles y el almacenamiento más barato suponen una ampliación de un Data Lake a terabytes y petabytes bastante económico.

- Un Data Lake soporta todo tipo de datos:

Los almacenes de datos generalmente consisten en información extraída de sistemas transaccionales. Y, por lo tanto, incluyen métricas cuantitativas, así como otros atributos cualitativos.

Las fuentes de datos no tradicionales como pueden ser los registros del servidor web, la actividad de una determinada red social, imágenes o los textos suelen ignorarse. Y aunque se siguen encontrando nuevos usos para este tipo de datos, consumirlos y almacenarlos supone un costo elevado y resulta bastante complicado.

- El Data Lake apoya a todos los tipos de usuarios:

En la mayoría de las organizaciones, más del 80% de los usuarios son operacionales. Es decir, quieren obtener sus informes, ver sus métricas claves de rendimiento o trabajar sobre el mismo conjunto de datos en una hoja de Excel. Y para ellos, el Data Warehouse es ideal. Ofrece una estructura clara, fácil de usar, de comprender y, sobre todo, está diseñado para responder a su sinnúmero de preguntas.

Un 10%, en cambio, hace un análisis mayor sobre esos datos. Es decir, estos usuarios utilizan el Data Warehouse como una fuente más. Sin embargo, a menudo vuelven a los sistemas de origen para obtener datos que no están incluidos en el almacén. E, incluso, incorporan nuevos datos de fuera de la empresa.

Por último, el 10% restante hace un análisis profundo. Estos usuarios saben qué es un Data Lake y crean fuentes de datos totalmente nuevas basadas en la investigación. De hecho, estudian muchos tipos diferentes de datos hasta llegar a nuevas preguntas que intentan dar respuesta a sus necesidades.

Aunque pueden usar el Data Warehouse prefieren ignorarlo. Ya que, a menudo, su utilización supone tener un conocimiento que no disponen. Sin embargo, cuentan con ciencias de datos y utilizan herramientas de analítica avanzada.

El enfoque del Data Lake apoya de la misma manera a todos estos usuarios. La ciencia de datos “ir al lago” y trabajar con toda la información en bruto que necesiten. Mientras tanto, los otros usuarios harán uso de vistas más estructuradas de los datos proporcionados para uso.

- Los Data Lakes se adaptan fácilmente a los cambios:

Una de las principales desventajas de los Data Warehouse es que se tarda mucho tiempo en cambiarlos. Ten en cuenta que durante el desarrollo de la estructura del almacén se gasta un tiempo considerable. Sin embargo, aunque un buen diseño de este repositorio pueda adaptarse al cambio, el proceso de carga de datos es tan complejo que estos cambios necesariamente consumirán algunos recursos para desarrolladores.

Como puede suponerse hay muchas cuestiones comerciales que deben abarcarse. Y esto no puede esperar a que el equipo de almacén de datos adapte su sistema. Actualmente el usuario quiere respuestas aquí y ahora. Y, por eso mismo, se han creado muchos sistemas de autoservicio de inteligencia empresarial. Sistemas que auto responden para ahorrar tiempo y resultar más eficaces.

Para entender qué es un Data Lake, en cambio, debes tener clara una idea: Son los usuarios quienes tienen el poder de ir más allá de la propia estructura de este. Teniendo en cuenta que todos los datos se almacenan en bruto y siempre están accesibles, esta práctica es posible.

Se demuestra que el resultado de una exploración ha resultado útil y se quiere repetir, puede aplicarse un esquema más formal. De hecho, ahí es donde se podrían desarrollar sistemas de automatización y reutilización para extender los resultados a públicos más amplios.

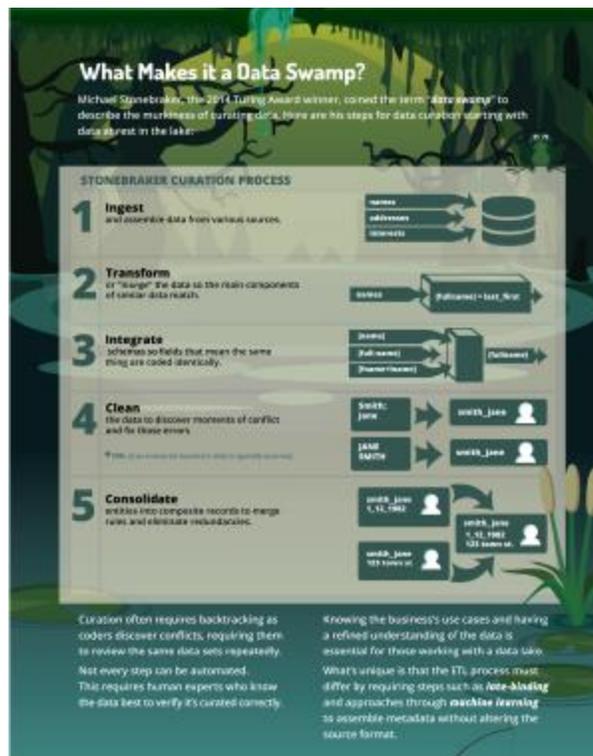
- Los Data Lakes proporcionan una visión más rápida:

Esta última diferencia es el resultado de las cuatro anteriores. Los Data Lakes contienen todo tipo de datos. Permiten a los usuarios acceder a ellos antes de transformarse y estructurarse. Y pueden llegar a sus resultados más rápido que utilizando el Data Warehouse.

Sin embargo, este acceso también tiene un precio. El trabajo típicamente realizado por el equipo de desarrollo de Data Warehouse no puede hacer uso de algunas fuentes de datos para realizar un análisis. Y esto deja a los usuarios como responsables de la exploración. Pueden usar los datos como mejor les parezca. Sin embargo, el primer nivel de usuarios mencionado anteriormente tal vez no quiera hacer esto. Quizás le interese simplemente obtener un informe con unos KPI's concretos y predefinidos.

En el Data Lake, estos usuarios de informes harán uso de vistas más estructuradas de los datos, algo más parecido al Data Warehouse. La diferencia es que estos puntos de vista existen principalmente como metadatos. Y se sitúan sobre los Data Lake, en lugar de tablas físicamente estructuradas que requieren de la ayuda de un desarrollador para ser modificadas. (¿Qué es un Data Lake, 2017)

Figura 17. What's Damming Data Lakes? Insights from Tableau, Qlik and Logi Analytics



5 Data Scientist Facts

1 By 2018 there is a projected 1.8 million shortage of US data professionals

2 A data scientist's top programming languages:

- R
- Python
- SAS
- Java
- Matlab
- SQL
- RPL
- Spark
- Hadoop

3

- 12% don't have a secondary degree
- 79% have graduate degrees
- 42% have master's degrees
- 28% have Ph.D.s

4 67% of data scientists were hired in that last four years

5 A data scientist's top eight responsibilities:

- Data cleaning & processing
- predictive modeling
- machine learning
- developing solutions
- testing theories
- designing statistical analysis
- identifying data sources
- wrapping & visualization

For those already maintaining a data warehouse, a data lake can be used as a "sandbox" for exploring data relations. For those new to storing and analyzing data, data lakes are more affordable with greater storage and provide flexibility for rapid innovation.

Data Lake vs. Data Warehouse

A data warehouse is what most businesses use to store their data. This only works with structured data. It offers less flexibility for identifying data trends, but more control over the data's quality.

One way to break down a data warehouse is into smaller "data marts". Think of a data mart as a selective fish market. Compared to the lake, the market offers to fish pre-packaged and sorted based on what daily visitors need, whereas in the lake, all fish occur naturally for fishermen to scoop out themselves when they hook it on-the-fly.

The fish within the lake run a higher risk of "sunkiness," but fish from the market are well refined while more carefully selected. In this same way, data from a data lake runs a higher risk of not being curated properly, and a data mart goes through a stricter process for ensuring quality.

	DATA LAKE	DATA MART
S	• Agile	• Efficient
O	• Flexible	• Governed
C	• Vast	• Rigid
N	• Multitier	• Unadaptable

Recuperado de <https://datafloq.com/read/damming-data-lakes-insights-tableau-qlik-logi/2162>

9.2 ¿Data lakes en la nube o local?

Los data lakes se suelen implantar localmente, con el almacenamiento en HDFS y el procesamiento (YARN) en clústeres de Hadoop. Hadoop es escalable, económica y ofrece un buen rendimiento con su ventaja inherente de la localidad de los datos (los datos y el almacenamiento residen en el mismo lugar).

Ahora bien, existen dificultades en la creación de infraestructuras locales:

- El espacio: los servidores voluminosos ocupan un espacio físico que acarrea precios más altos.
- La configuración: la adquisición de hardware y la configuración de centros de datos no es sencilla y puede tardar semanas o meses hasta ponerlos en funcionamiento.
- La escalabilidad: si existe la necesidad de ampliar la escala de capacidad de almacenamiento, se requiere tiempo y esfuerzo, debido al requisito de mayor espacio y las aprobaciones de costes de altos directivos.
- El cálculo de requisitos: dado que la escalabilidad no es sencilla en entornos locales, al principio del proyecto resulta importante calcular los requisitos de hardware correctamente. Como los datos crecen de manera no sistemática todos los días, este objetivo es muy difícil de lograr.
- El coste: los cálculos de los costes han demostrado ser más elevados en entornos locales que sus alternativas cloud.

Los data lakes en cloud, por otro lado, contribuyen a superar estas dificultades. Un data lake en cloud es:

- Más fácil y rápido de iniciar. En lugar de un enfoque explosivo, la cloud permite a los usuarios empezar paulatinamente.
- Rentable con un modelo de pago por uso.
- Más fácil de escalar al alza cuando las necesidades aumenten, lo que elimina la tensión de tener que calcular requisitos y obtener autorizaciones.
- Los ahorros en cuanto a espacio físico también se cuentan entre las ventajas económicas. (Pearlman, 2019)

9.3 Los retos de los data lakes en la nube

Es evidente que el uso de un data lake en cloud no está exento de dificultades. Algunas organizaciones prefieren no almacenar información confidencial y delicada en la cloud por posibles riesgos de seguridad. Mientras gran parte de los proveedores de data lakes en cloud avalan la seguridad y con el paso de los años han ido aumentando sus capas de protección, la insistente incertidumbre en lo que respecta al robo de datos persiste.

Otro reto de orden práctico es que algunas organizaciones ya tienen instalado un sistema de almacenamiento de datos para guardar sus datos estructurados. Pueden elegir migrar todos esos datos a la cloud o plantearse una solución híbrida con un motor de procesamiento compartido que acceda a los datos estructurados del almacén y a los datos no estructurados en cloud.

La gobernanza de datos es otro frente problemático. Un data lake, por mucho que contenga datos, no debería convertirse en un pantano por el que cueste recorrer.

Arquitectura de un data lake: Hadoop, AWS y Azure

Es importante recordar que un data lake tiene dos componentes: almacenamiento y procesamiento. Tanto el almacenamiento como el procesamiento pueden ubicarse localmente o en la cloud. Esto genera distintas combinaciones posibles al diseñar la arquitectura de un data lake.

Las organizaciones pueden decidir quedarse íntegramente en el entorno local, trasladar toda la arquitectura a la cloud, plantearse múltiples clouds o incluso un híbrido entre estas opciones.

No hay receta en este campo. Según cuales sean las necesidades de una organización, habrá varias opciones correctas. (Pearlman, 2019)

9.4. Data lakes en Hadoop

Hadoop es la opción que se asocia mayormente con los data lakes, por defecto.

Un clúster de Hadoop de servidores distribuidos resuelve las dudas sobre el almacenamiento de big data. En el núcleo de Hadoop encontramos su capa de almacenamiento, el HDFS (Sistema de

archivos distribuidos de Hadoop), que almacena y replica los datos por múltiples servidores. YARN (Otro gestor de recursos más) es el gestor de recursos que decide cómo programarlos en cada nódulo. MapReduce es el modelo de programación que utiliza Hadoop para dividir los datos en subconjuntos más pequeños y procesarlos en sus clústeres de servidores.

Más allá de estos tres componentes básicos, el ecosistema Hadoop engloba varias herramientas suplementarias, como Hive, Pig, Flume, Sqoop y Kafka que ayudan con la ingesta, la preparación y la extracción de datos. Los data lakes de Hadoop pueden configurarse localmente o en la cloud mediante plataformas de empresa como Cloudera o HortonWorks. Otros data lakes cloud, como Azure, envuelven sus funcionalidades alrededor de la arquitectura Hadoop.

Ventajas:

- Mayor familiaridad entre los profesionales informáticos.
- Más económicos, porque son de código abierto
- Muchas herramientas ETL disponibles para la integración con Hadoop
- Fáciles de escalar
- La localidad de los datos permite un procesamiento más rápido. (Pearlman, 2019)

9.5 Data lakes en AWS

AWS tiene un paquete exhaustivo de ofertas de producto para su solución de data lake.

Amazon Simple Storage Service (Amazon S3) se sitúa en el centro de la solución que proporciona la función de almacenamiento. Kinesis Streams, Kinesis Firehose, Snowball y Direct Connect son herramientas de ingesta de datos que permiten a los usuarios transferir cantidades ingentes de datos a S3. También existe un servicio de migración de bases de datos que contribuye a migrar datos locales existentes a cloud.

Además de S3, existe DynamoDB, una base de datos No-SQL de baja latencia, y Elastic Search, un servicio que ofrece un mecanismo simplificado para realizar consultas en el data lake. Los

Cognito User Pools definen la autenticación de usuarios y el acceso al data lake. Los servicios como Security Token Service, Key Management Service, CloudWatch y CloudTrail garantizan la seguridad de los datos. Para el procesamiento y la analítica existen herramientas como RedShift, QuickSight, EMR y el machine learning.

La enorme lista de ofertas de producto disponibles en AWS entraña una curva de aprendizaje inicial muy empinada. No obstante, las completas funciones de esta solución tienen un uso muy extendido en las aplicaciones de business intelligence.

Ventajas:

- Paquete de productos exhaustivo y repleto de prestaciones
- Flexibilidad a la hora de elegir productos a partir de requisitos particulares
- Precios bajos
- Nivel de seguridad y conformidad alto
- Separación del procesamiento y almacenamiento para poder ajustar su escala según convenga
- Colaboración con empresas de APN (AWS Partner Network), como Talend, que garantiza una etapa inicial de funcionamiento con AWS exenta de problemas. (Pearlman, 2019)

9.6 Data lakes en Azure

Azure es un data lake de Microsoft. Tiene una capa de almacenamiento y de analítica. La capa de almacenamiento se denomina Azure Data Lake Store (ADLS) y la analítica contiene dos componentes: Azure Data Lake Analytics y HDInsight.

ADLS está creada siguiendo la norma HDFS y tiene una capacidad de almacenamiento ilimitada. Es capaz de almacenar billones de archivos en un único archivo de más de un petabyte. ADLS permite guardar datos en cualquier formato y es seguro y escalable. Es compatible con cualquier aplicación que utilice la norma HDFS. De esta forma la migración de datos existentes es

mucho más sencilla y también facilita la conexión plug-and-play con otros motores de procesamiento.

HDInsight es un servicio analítico de data lakes en cloud. Creado sobre Hadoop YARN, permite acceder a los datos mediante herramientas como Spark, Hive, Kafka y Storm. Es compatible con la seguridad de nivel empresarial debido a su integración con Azure Active Directory.

Azure Data Lake Analytics es otro servicio analítico, pero su enfoque es distinto. En lugar de emplear herramientas como Hive, utiliza un lenguaje llamado U-SQL, que es una combinación de SQL y C#, para acceder a los datos. Es idóneo para procesamientos de big data por lotes, dado que ofrece mayor velocidad a menor coste (tan solo se paga por las tareas que se utilicen).

Ventajas:

- Tanto el almacenamiento como el procesamiento en cloud facilitan su manejo.
- Servicios analíticos robustos con funcionalidades potentes.
- Fácil de migrar desde un clúster Hadoop existente
- Muchos expertos en big data conocen Hadoop y sus herramientas, de modo que resulta fácil encontrar trabajadores cualificados.
- La integración con Active Directory garantiza que no son necesarios esfuerzos adicionales para gestionar la seguridad. (Pearlman, 2019)

10. Data Analytics

Figura 19. Data Analytics



Recuperado de: <https://fr.wegalanvize.com/risk/data-analytics-risk-assessments/>

Data Analytics corresponde a un enfoque que implica el análisis de los datos para obtener respuestas, al implementar Data Analytics las empresas pueden estar mejor preparadas para la toma de decisiones estratégicas, trabajar de manera más rápida y ágiles, que a su vez permite descubrir ideas y tendencias.

Los principales beneficios que trae Data Analytics son la eficiencia y velocidad, que esto brinda una ventaja competitiva que antes no se tenía.

El valor agregado que brinda Data Analytics se resumen en:

- Reducción de costo. Las tecnologías de big data como Hadoop y la analítica en la nube ofrecen ventajas de costo significativas cuando se trata de almacenar grandes cantidades de datos – además, pueden identificar formas más eficientes de hacer negocios.
- Toma de decisiones mejores y más rápidas. Con la velocidad de Hadoop y la analítica en memoria, combinadas con la posibilidad de analizar nuevas fuentes de datos, las empresas

pueden analizar información de inmediato – y tomar decisiones basadas en lo que han aprendido.

- Nuevos productos y servicios. Con la posibilidad de medir las necesidades y la satisfacción del cliente a través de la analítica viene el poder de dar a los clientes lo que desean. Davenport puntualiza que con la analítica del big data, más compañías están creando nuevos productos para satisfacer las necesidades de los clientes. (Analíticas del big data, s.f)

Figura 19. Big Data



Recuperado de <https://blog.es.logicalis.com/analytics/en-que-consiste-big-data-analytics-y-como-beneficia-a-tu-empresa>

11. Herramientas de visualización

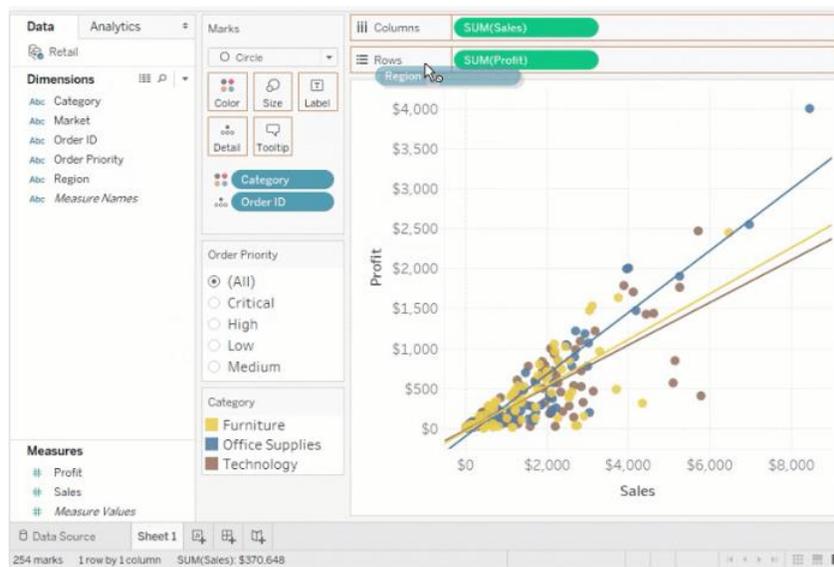
La visualización de datos permite interpretar la información de manera sencilla y visual, el consumo de información con la que contamos se ha multiplicado exponencialmente y esto genera que en muchas ocasiones no se dispone del tiempo suficiente para su correcta interpretación.

Es por esto por lo que las herramientas de visualización de datos toman un papel importante, a continuación, detallamos algunas de las principales herramientas con las que cuenta el mercado actual.

11.1 Tableau

En la actualidad es una de las herramientas más completas y de las más utilizadas, su interfaz permite generar visualizaciones sobre un gran volumen de datos, es sencilla de utilizar, además que cuenta con una versión gratuita. (Fernández, 2017)

Figura 21. Tableau

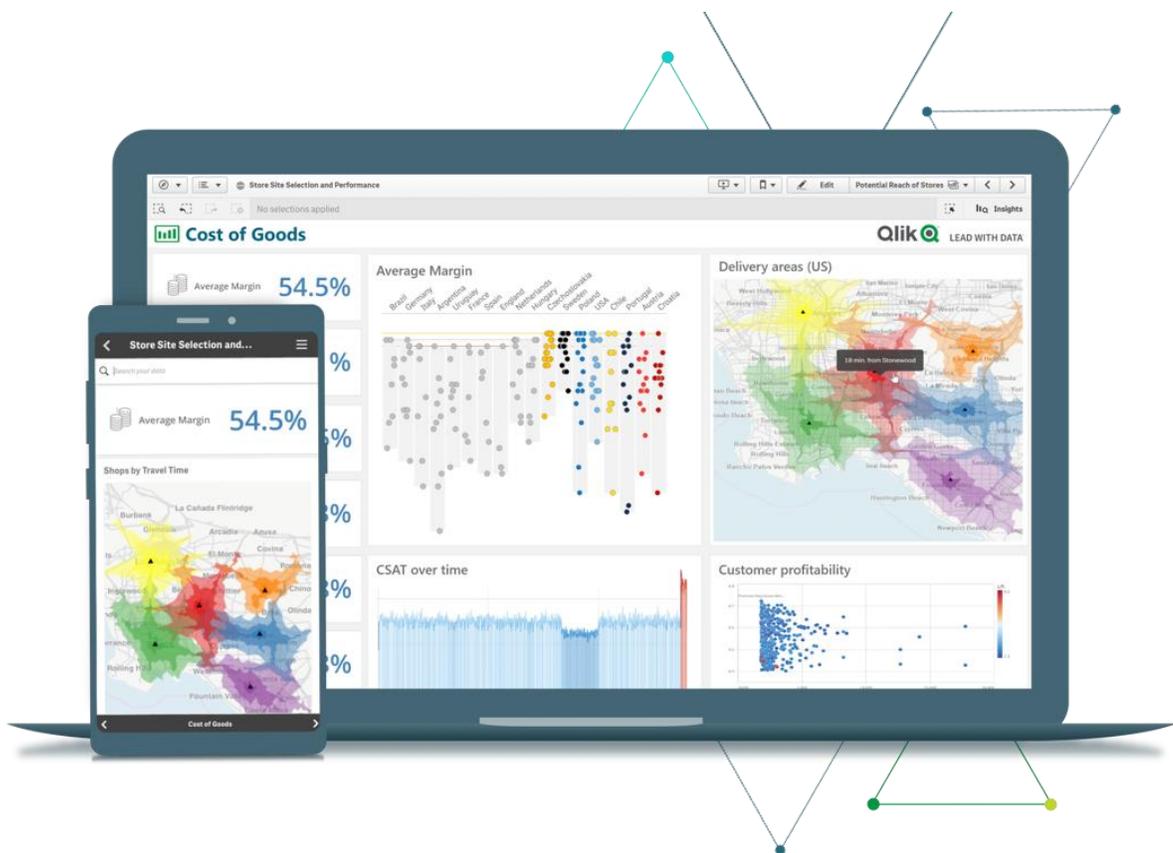


Recuperado de <https://nova-praxis.com/software-business-intelligence-epm/tableau-desktop-online/>

11.2 Qlik

Ofrece dos versiones del producto, Qlik View es el más popular y su mejor rasgo es su sencillez y que es muy personalizable, permitiendo a los usuarios tomar decisiones basadas en los datos, también existe QlikSense, una versión más “light” para crear visualizaciones flexibles e interactivas. (Fernández, 2017)

Figura 21. Qlik

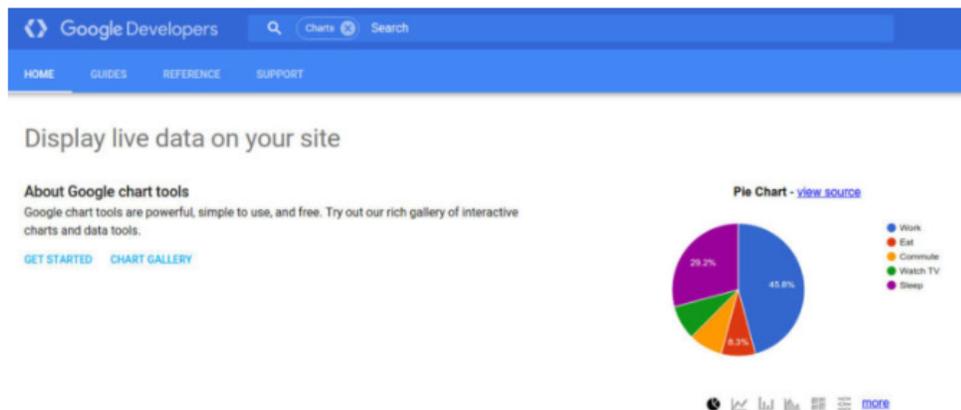


Recuperado de <https://www.qlik.com/es-es/products/qlik-sense>

11.3 Google Charts

Tiene como principal ventaja que es compatible con todos los navegadores comerciales, con esta herramienta se pueden crear gran cantidad de gráficos interactivos, tiene una interfaz muy agradable y sencilla.

Figura 22. Las 20 herramientas de visualización de datos Big Data del momento



Recuperado de <https://www.grapheverywhere.com/las-20-herramientas-de-visualizacion-de-datos-big-data-del-momento/>

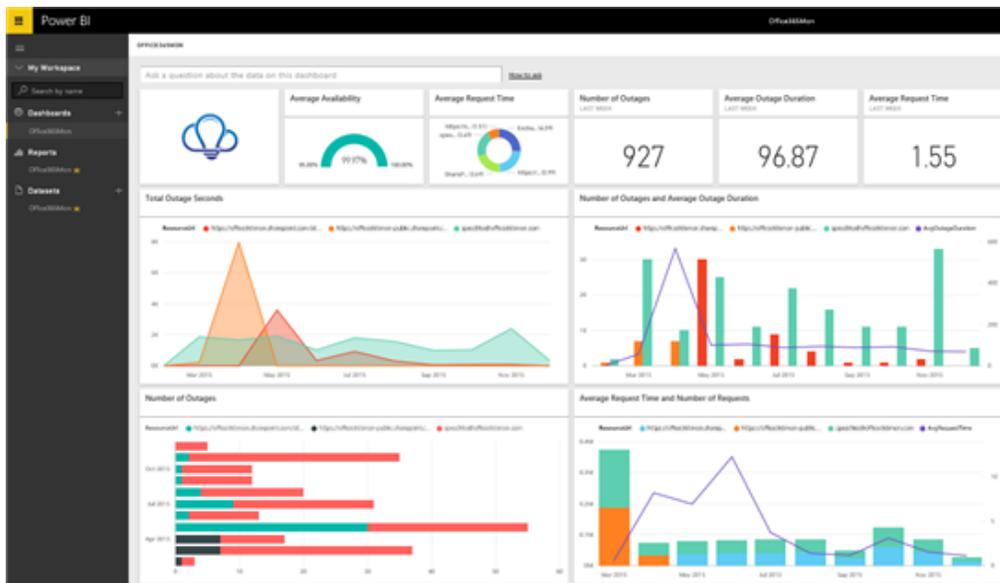
11.4 Power BI

Corresponde a un conjunto de herramientas de análisis empresarial, está integrada a las soluciones de Office 365, permite la conexión de cientos de orígenes de datos, permite generar informes bastante dinámicos y amigables.

11.4.1 Algunas de sus ventajas son:

- Flexibilidad. Te permite extraer información importante para una amplia gama de escenarios.
- Optimizar, Limpiar, transformar y combinar datos de múltiples orígenes. Analizar en profundidad los datos y encontrar patrones.
- Innovación. Podrás crear informes sorprendentes con visualizaciones de datos interactivas.
- Personalizar. Diseñar tu informe mediante las herramientas de creación de temas, formato y diseño.
- Multiplataforma. Crear informes optimizados para dispositivos móviles. (Qué es Power BI, 2019)

Figura 23. Power BI



Recuperado de <https://powerbi.microsoft.com/es-es/blog/analyze-your-office365mon-data-in-power-bi/>

Capítulo 3: Marco Metodológico.

El presente proyecto se basa en la aplicación de la teoría de desarrollo de sistemas de información, específicamente de una solución de inteligencia de negocios, delimitada en este caso a la construcción de un Datamart comercial, lo anterior mediante de proceso de desarrollo evolutivo. Específicamente el modelo de prototipos y la metodología de diseño rápido de Aplicaciones o soluciones de software, para crear en un tiempo no muy prolongado un producto funcional que permita a alimentos Jack's el manejo de datos y producción de evidencias para el proceso de toma de decisiones acertadas, ágiles y confiables en base a tendencias históricas y proyecciones futuras. Esta es una investigación aplicada, que pretende resolver un problema identificado, apropiadamente delimitado o definido, mediante la implementación de un sistema informático para este problema, busca producir conocimiento nuevo y nuevas estrategias para el manejo de las ventas, inventarios y la penetración de los productos en los diferentes nichos de mercado.

3.1 Enfoque de la Investigación:

Se han identificado tres tipos de enfoque para una investigación, el enfoque cuantitativo, el cual representa un conjunto de procesos, es secuencial y probatorio; el enfoque cualitativo, el cual no se busca tener claridad sobre las preguntas de investigación o hipótesis antes de recolectar o analizar los datos, este tipo de enfoque puede desarrollar preguntas o hipótesis, durante o después de la recolección y el análisis de los datos. Finalmente, el enfoque mixto, es la integración sistemática de los métodos cuantitativo y cualitativo en un solo estudio con el fin de obtener una “fotografía” más completa del fenómeno. (Hernández, Fernández y Baptista, 2006)

El enfoque a utilizar es un enfoque cualitativo, ya que alimentos Jack's y en este caso las áreas específicas a analizar poseen una realidad única, que debe de ser apropiadamente analizada, contextualizada e interpretada, de acuerdo con la observación de procesos, documentos de control, y entrevistas realizadas a los recursos encargados del proceso de ventas, inventarios y penetración de mercado; para poder proponer una solución que se ajuste a este ambiente y a los recursos existentes, que complemente el proceso y minimice, en lo posible, la necesidad de procesos manuales para la inclusión de la información, el manejo de los datos y la creación de evidencias.

3.2 Fuentes y Sujetos De Información

3.2.1 Fuentes de Información:

Las fuentes de información son todos aquellos repositorios de conocimiento que se utilizaran, esta información será recopilada de una manera ordenada, estructurada y confiable. Son fuentes de información todos aquellos materiales utilizados para realizar esta investigación, tales como libros, tesis, proyectos, manuales de procesos y/o reglamentos que sean relevantes al tema.

3.2.2 Sujetos de Información:

Los sujetos de información para la investigación son aquellos que juegan un papel importante en los procesos de relacionados con el tema a investigar. Se han identificado dos sujetos de información claves para el entender los requerimientos de la solución a realizar, tanto desde un punto de proceso como de vista tecnológico.

El señor David Quirós encargado del departamento comercial y sus colaboradores encargados de las diferentes áreas de negocio involucradas (ventas, inventario, penetración de producto.)

La señora Nuria Marín, directora de sistemas, Jonathan Segura y Francisco Rodriguez analistas especializados en el manejo de información de la compañía. Entre sus funciones tienen desarrollar sistemas que ayuden al manejo de todos los procesos institucionales.

3.3 Técnicas y Herramientas de Recolección De Datos

La recolección de datos es parte crucial del proceso de investigación, Hernández, Fernández y Baptista (2006) se refieren a la recolección de datos como: “Recolectar los datos implica elaborar un plan detallado de procedimientos que nos conduzcan a reunir datos con un propósito específico.”

En el presente trabajo se aplicarán dos herramientas de recolección de datos: la entrevista y la observación.

Entrevista: Una persona calificada aplica un cuestionario a los participantes, realizando preguntas cada uno de los entrevistados y anotará las respuestas. (Hernández et al., 2006).

Se realizarán entrevistas a las fuentes de información anteriormente mencionadas para obtener un panorama del estado actual desde el punto de vista del entrevistado, y luego se contrapondrán estas respuestas con el estado ideal, es decir los requerimientos definidos por alimentos Jack’s para graduar ese estado ideal con el estado actual del proceso.

Observación: Esta consiste en la recopilación de datos de manera estructurada, válida y confiable de comportamientos y situaciones observables, mediante el uso de categorías y subcategorías. (Hernández et al., 2006).

Mediante la observación de los procesos utilizados en alimentos Jack’, así como en los documentos generados a raíz de estos, se tendrá una visión más cercana a la realidad actual, se podrán obtener datos de ventas, inventarios y penetración de mercado para obtener las evidencias requeridas y verificar si existe alguna mejora adicional que se pueda realizar a los procesos existentes.

3.4 Variables de Investigación

Las variables de investigación se definen en la Tabla 5, al ser el presente trabajo, una investigación aplicada de carácter cualitativo, las variables de investigación se agruparán en categorías para análisis.

Objetivo Específicos	Categorías de análisis	Descripción
<p>Precisar el proceso actual de recolección de datos, teniendo en cuenta los procesos manuales y automatizados si existiesen, sus interacciones actuales y la posterior manipulación de estos datos para la generación de evidencias.</p>	<p>Indagar en el proceso actual de recolección de datos.</p> <p>Clasificar los procesos actuales en manuales y automatizados.</p> <p>Determinar el proceso más importante entre los procesos actuales y los requeridos por alimentos Jack's.</p>	<p>Por medio de la información recopilada de los documentos utilizados, la observación de estos y una posterior entrevista con las fuentes primarias para determinar el estado real actual de los procesos.</p>
<p>Establecer los requerimientos necesarios junto con las áreas administrativas y de tecnologías de información que permitan el diseño funcional de una solución de software que soporte los diferentes procesos actuales.</p>	<p>Determinar los requerimientos funcionales de una posible solución de software.</p> <p>Determinar los requerimientos no-funcionales de una posible solución de inteligencia de negocios.</p>	<p>Por medio de las directrices de alimentos Jack's, se crearán y se ratificarán los requerimientos para la posible solución de inteligencia de negocios.</p>
<p>Diseñar la plataforma tecnológica que cumpla con los requerimientos funcionales y no-funcionales debidamente identificados y descritos para la solución de software.</p>	<p>Diseñar la solución de inteligencia de negocios. basados en los requerimientos funcionales, es decir los requerimientos de cómo ha de operar la solución de software para que esta resuelva el problema.</p> <p>Diseñar la solución de software basada en los requisitos no</p>	<p>Se utilizarán técnicas de modelado de software para diseñar la solución.</p>

	funcionales, es decir el rendimiento, disponibilidad, estabilidad, tolerancia a fallos.	
Desarrollar el producto funcional, basado en los diseños y requerimientos previamente definidos, debiendo estos de ser aceptados por la parte administrativa y de tecnología de alimentos Jack´s.	Esta parte de refiere a la creación de la solución de software, así como del ambiente necesario de hardware para que esta pueda operar apropiadamente.	Se utilizará técnicas ágiles de desarrollo y la construcción se hará bajo un ambiente de SQL Server 2017 por medio de sus componentes OLTP, SSIS, SSAS, SSRS, además de Power Bi como herramienta principal de auto servicio de información.

Tabla 5. Variables de investigación.

Fuente: propia.

En esta sección se describe el marco metodológico o de investigación empleado para el desarrollo del proyecto, así como las fuentes de información y las herramientas utilizadas para la recopilación de estas.

La presente tesis tiene como objetivo implementar un modelo que ayude a la toma de decisiones para las áreas comerciales, teniendo en cuenta de creación de reportes que funcionan como indicadores, los cuales ayudaran con la planificación estratégica para la toma de decisiones y planteamientos futuros para la organización.

La investigación utilizada es una investigación de campo, que permitirá por medio de la manipulación de los datos, obtener información que sea valiosa para Jack´s.

El desarrollo por realizar para este proyecto es una investigación, que buscar permitir a la organización mediante la obtención de información, las mejores condiciones posibles y las mejores prácticas o estrategias para que sean aplicadas en la implantación de las ventas.

Teniendo conocimiento pleno de las situaciones actuales y proyecciones futuras permitiría a la compañía plantearse nuevas o modificar las estrategias para aumentar la rentabilidad del negocio. Podrán tener información actualizada a tan solo unos clics.

Las técnicas de recopilación de datos que se utilizaron son:

- Observación: en el primer encuentro me dedique a observar el manejo de la información y como la obtienen.
- Entrevistas con los Stakeholders de la organización. Las mismas se realizaron personalizadas visitando las oficinas de la compañía.

Las entrevistas se realizan con los dos encargados de sistemas Francisco Rodriguez, Jonathan Segura, la encargada del proyecto Nuria Loaiza y David Quirós como el gerente de ventas, así como los encargados de los departamentos involucrados.

- Recopilación de toda la información que maneja la organización y que se encuentra en distintos formatos.
- Análisis de la información obtenida con el fin de determinar la mejor manera de manejar la información de la compañía.

Procesamiento de datos:

Una vez obtenida toda la información necesaria para realizar el proyecto, el procesamiento de los datos se realizará de la siguiente manera:

Utilizar la información recopilada en el diseño y modelado de un DataMart, el cual contiene la información de las ventas de la organización para posteriormente crear algunos reportes estratégicos para la toma de decisiones.

De modo que permita facilitar en gran manera la obtención y consulta de los datos por parte de los encargados de toma de decisiones. (Aimacaña, 2013)

Figura 24. Diagrama del proyecto.



Fuente: Elaboración propia (2019)

Capítulo IV. Diagnóstico de la situación actual

4.1 Diagnóstico de la situación actual:

Actualmente, alimentos Jack's cuenta con una herramienta de toma de decisiones que les ha sido funcional pero poco práctica, ya que lo que tienen es un archivo de Excel muy grande, el cual es elaborado por un analista de negocio. En este archivo lo que hacen es incluir información proveniente de múltiples reportes que se generan de los sistemas transaccionales, y que son descargados y manipulados por dicho analista para ir incorporando la información de manera histórica en el archivo de análisis.

Uno de los grandes problemas, es que hacer este proceso manual, le toma al analista de datos más de una semana en tabularlo y manipularlo antes de compartírselo a los usuarios que requieren la información para analizarla y tomar decisiones, y por el tiempo que toma, se hace únicamente a inicios de mes, con la información del cierre del mes anterior.

Como se puede notar, los usuarios claves de negocios y las diferentes gerencias, toman decisiones de forma reactiva, ya que por la forma en la que se hacen las cosas no pueden ir midiendo el comportamiento del negocio, ni las tendencias actuales para ir tomando decisiones mucho más asertivas.

Adicionalmente, los indicadores de rendimiento de la compañía se analizan hasta que ya no hay nada que hacer para el periodo que están analizando, es un análisis completamente "Post mortem" y no pueden encauzar el rumbo del negocio de forma dinámica, dependiendo de lo que sucede en el mercado.

Todo lo anterior, se hace para todas las áreas de negocio, pero en este caso, analizamos las áreas comerciales; entendiendo esto por los departamentos de ventas, inventarios, cuentas por cobrar e inteligencia de mercado. Este último departamento analiza el nivel de penetración de los productos por los diferentes segmentos de mercado, como lo son: supermercados, pulperías, sodas escolares, entre muchas otras.

Al analizar los procesos que son llevados a cabo por el departamento de inventario, notamos una complicación innecesaria y además imprecisa, ya que ellos necesitan saber cuántos artículos deben mandar a producir. Esto debería ser lo más exacto posible, ya que las máquinas de producción se utilizan para varios artículos diferentes, por lo cual, si estiman mal la cantidad a

producir y en un momento se están quedando con bajo inventario, deben hacer mucho trabajo para cambiar el producto que están preparando y regresar a otro producto que ya habían fabricado y que debería estar en la bodega. Esto eleva los tiempos de producción y por ende el costo de producción. El cálculo que realizan para el manejo de este inventario es tomando como referencia el mismo mes del año anterior y agregarle un porcentaje de holgura, así para cada periodo. Este porcentaje no corresponde a ningún criterio técnico ni científico, sino que queda completamente a discreción del encargado del área. Si, por el contrario, estiman que deben producir más de lo que realmente ocupaban, esto tiene la implicación de gasto de materiales de productos que permanecerán en bodega y eleva también los costos de almacenaje. Esto al final de cuentas, en cualquiera de los dos escenarios, tiene un impacto sobre el ROI (Retorno de la inversión).

Con todas las inquietudes del departamento de inventario ya claramente diagramadas, se procede a analizar el departamento de ventas, y se encuentra que el panorama no es muy diferente; Ellos tienen limitantes para realizar sus análisis, principalmente porque de forma reciente, la compañía implementó un nuevo ERP en la nube (Dynamics), y la información de venta de los ruterros (la venta que se realiza con los camiones en pulperías, escuelas, etcétera) se sigue manejando en un sistema antiguo que se encuentra OnPremise y que se llama Power Street. Por lo cual, si ellos el día de hoy necesitan saber cuánto se vendió ayer, no puede utilizar el archivo gigante de Excel, porque los datos se actualizan una vez al mes, y si desean generar reportes desde los sistemas transaccionales, deben generarlos de dos sistemas diferentes, y con formatos completamente diferentes, y como si fuera poco, con codificaciones diferentes, es decir, el código de un artículo en Power Street es numérico y el de Dynamics es alfanumérico y completamente diferente. Por lo que deben hacer una descarga de los reportes, exportarlos a Excel y manualmente homologar los artículos, clientes, rutas, etcétera, para posteriormente consolidar los datos y analizar el resultado.

Para ellos sería sumamente útil poder realizar proyecciones de venta, tomando como base las ventas del periodo actual y logrando de forma automática determinar cuántos son los días hábiles del mes y cuantos días han transcurrido, pero eso tiene complicaciones para ellos, ya que desearían que se pueda hacer con un calendario operativo, el cual cuanta de lunes a viernes como un valor de 1, los feriados y domingos cuentan como 0 y los sábados con un valor de 0.5. Pero si la proyección la va a revisar el área gerencial, se debe utilizar un calendario diferente, en el cual los cálculos

deben tomar de lunes a viernes como un valor de 1, los feriados, sábados y domingos con un valor de 0.

Adicionalmente, ellos desearían tener a la mano variables que les muestre cual es el crecimiento versus el mismo periodo del año anterior y versus el mes anterior. Estas variables deben ser con respecto a unidades vendidas y montos de venta. Actualmente hacer estos análisis son muy complejos y sumamente pesados, y pueden llegar a cierto nivel granular, ya que el Excel gigante no desgrana la información a nivel del artículo, ya que sería inmanejable.

Para cerrar las revisiones, se revisa el proceso que realiza el departamento de penetración de productos, y ellos básicamente toman como base la información que analiza el departamento de ventas para poder segmentar la información por producto y cliente. Una vez que realizan la segmentación, descargan el catálogo de productos de Power Street y el catálogo de productos de Dynamics, esto con el fin de determinar cuáles de los productos están como activos en producción, para limitar el análisis solo a esos productos, ya que, si incluyen productos inactivos, les afectará el porcentaje de penetración que están obteniendo en realidad. Exactamente el mismo proceso de descarga y unificación deben hacer con los clientes.

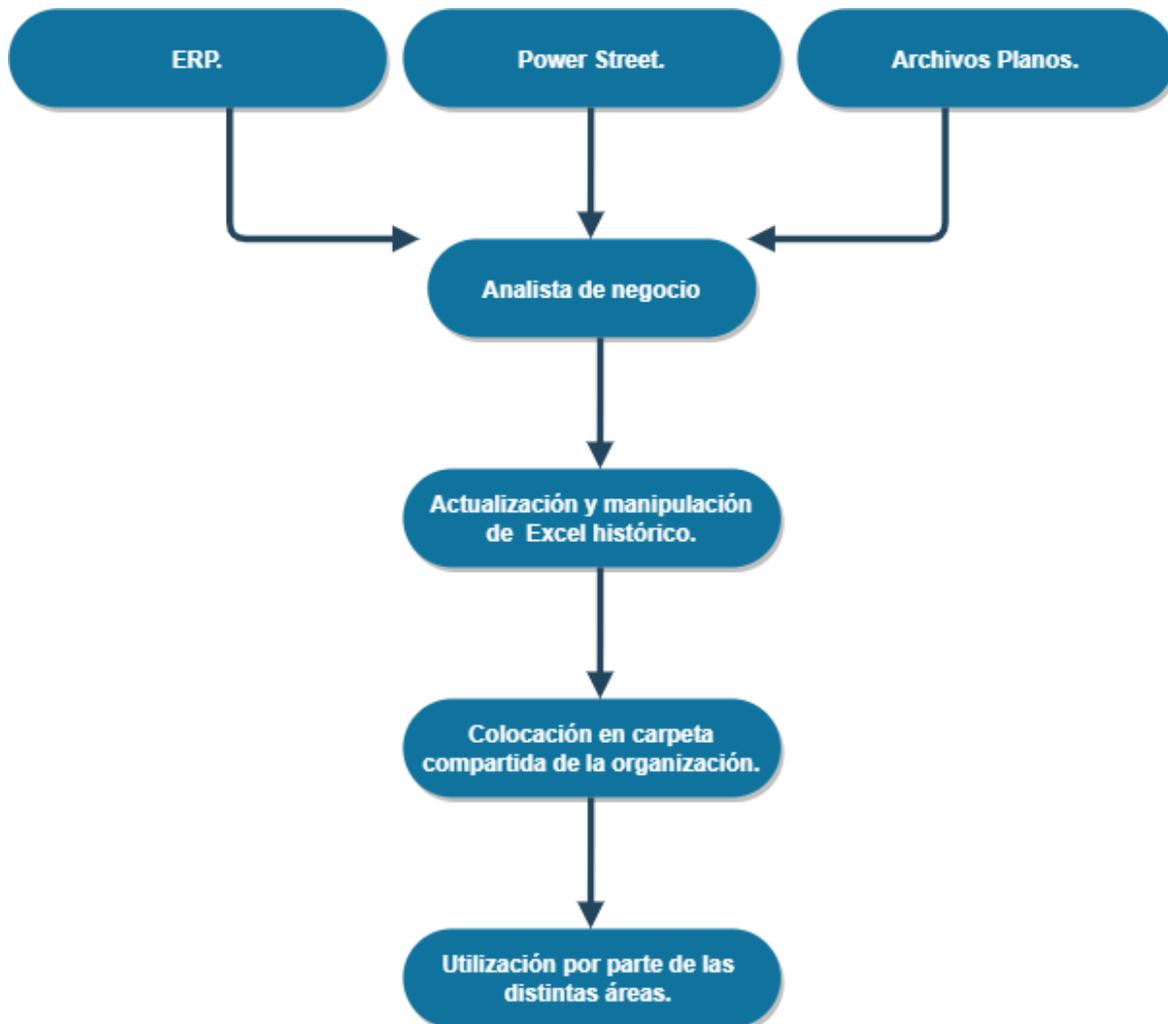
Algunos de los problemas que tienen, es que, si ventas comente un error en su tabulación de datos, y ellos realizan todo el trabajo, antes mencionado, y posteriormente les comunican del error, pierden todo el trabajo que habían hecho, y deben comenzar de cero. Y lamentablemente esto es algo común.

4.2 Diagnóstico operativo:

Un analista de negocio confecciona un Excel gigantesco, en el cual le incluye información que proviene de otros múltiples reportes que generan los sistemas transaccionales utilizados por alimentos Jack's, que son descargados y trabajados por el analista, con el fin de adicionar información histórica, este archivo gigante lleva mucho trabajo previo a colocarlo en las carpetas compartidas.

Posterior al trabajo del analista el gerente comercial la da una manipulación adicional a los datos incluidos en el Excel, para generar informes adicionales que son con los que toman decisiones como se mencionó antes “Post mortem”.

Figura 25. Diagrama situación actual.



Fuente: Elaboración propia, 2019

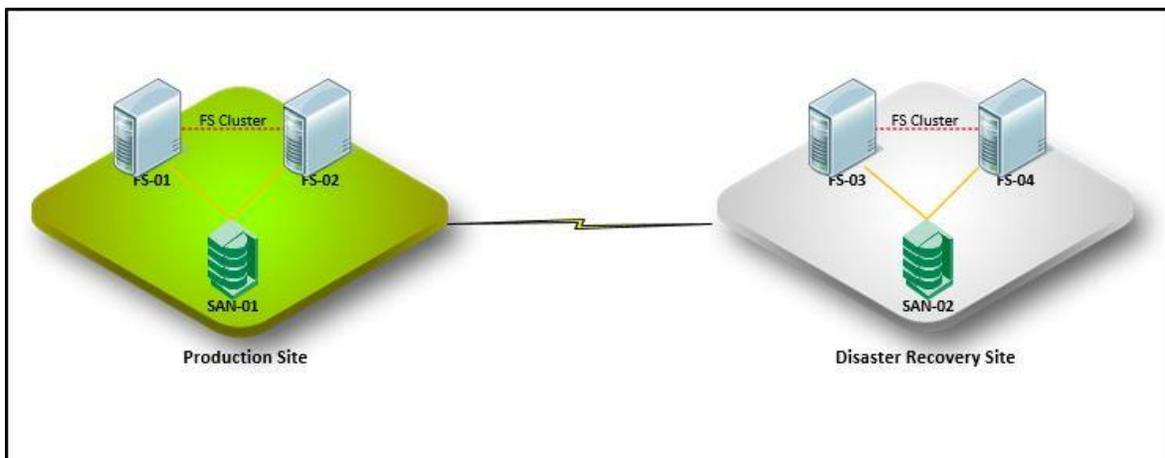
4.3. Diagnóstico técnico:

Actualmente todas las tareas se realizan de manera manual, parte del cambio con este desarrollo consiste en que se van a utilizar web services para extraer parte de la información. De igual manera cierta información siempre deberá ser obtenida mediante archivos de Excel y subida ya sea a las bases de datos o compartida a las gerencias por correo o carpetas compartidas. Como gestor de base de datos se utilizará SQL Server 2017 y Visual Studio 2017 para la realización de algunos procesos de carga.

La organización de manera general utiliza equipos Dell Optipex 7020, adicionalmente se utilizan Dell Inspiron 15 5584 para algunos colaboradores, como por ejemplo las contrapartes involucradas en este desarrollo.

Alimentos Jack's cuenta con un cuarto de servidores físico, además de servidores virtuales, todos con modo de replicación SAN to SAN.

Figura 26. Diseño SAN.



Recuperado de <https://pkjayan.wordpress.com/tag/file-services/>

4.4 Diagnostico percepción:

El criterio de los entrevistados en relación con los procesos actuales es que de la manera actual las tareas se vuelven tediosas y repetitivas a lo largo del tiempo, cada mes o cada vez que las gerencias necesiten determinada información deben empezar a extraer los datos para poder aportar esta información.

También concuerdan que existe la posibilidad de cometer errores durante el proceso que realizan, es por esto que se encuentran sumamente entusiasmados con la idea de poder llevar a cabo este desarrollo que les permita no solo tener más tiempo para atender otros tipos de solicitudes o tareas que generalmente tienen que postergar para poder estar a disposición de las gerencias en caso de que se requiera, sino, también para el desarrollo de nuevas ideas que aporten un valor adicional a la compañía.

La idea de automatizar estas tareas también genera cierta desconfianza o temor principalmente en las gerencias, que están muy acostumbradas a la forma en la que han trabajado durante muchísimo tiempo y ha sido una tarea que ha llevado tiempo y dedicación para poder generarles ese grado de confianza.

Con el tiempo los gerentes también se han involucrado bastante con las tareas realizadas y han cambiado drásticamente la percepción inicial, ahora se encuentran contentos e ilusionados con la gran cantidad de cosas que se pueden hacer con las herramientas que estamos utilizando para la creación de los informes.

Adicionalmente el modelo que se construirá les permitirá como compañía proyectar ventas y a partir de aquí gestionar de mejor manera el inventario y tomar decisiones proactivas.

Estas herramientas les permiten crear gran cantidad de nuevos reportes, que antes ni siquiera podían imaginar que podían crear por la dificultad de obtener datos, unificarlos y darles el tratamiento que estas herramientas ahora si les permiten.

4.5 Analisis de brechas.

Situación Actual	Brecha	Situación Deseada
Se ingresan archivos de forma manual a las tablas del servidor.	Establecer proceso de obtención de datos, a través de procesos automatizados para evitar los trabajos manuales y mitigar en la mayor forma posible los errores.	Obtener los datos de manera automática por medio de las diversas fuentes y evitar el trabajo manual.
Se da manejo de datos de manera manual.	Definir método que dentro de los procesos se establezcan los formatos y tipos de datos que son necesarios para no hacerlo de manera manual.	Que el procesamiento y el manejo de los datos sea automático dentro de los procesos desarrollados para evitar manipulación manual de los valores por parte de los usuarios.
Algunos datos sensibles son manejados en Excel, txt y pueden ser vistos por múltiples personeros de la organización.	Brindar permisos de acceso solo a los colaboradores que realmente deban ver la información. Definiendo un rol para determinar si debe contar con acceso a los datos.	Al automatizar los procesos menos recursos tendrán acceso a información sensible y con esto se protege un activo de la compañía y solo los colaboradores con determinado perfil puedan visualizar la información.

Tabla 6. Análisis de brechas

Fuente: propia.

Capítulo V. Propuesta de Proyecto

En este capítulo se especifican las características de orden técnico del modelo de Bases de Datos que se utilizará. Tiene el objetivo de brindar un detalle de los procesos creados con el fin de generar el producto requerido. También se especifica los paquetes que forman parte del proyecto “Modelo de Ventas”.

Definiciones, acrónimos y abreviaturas utilizadas en el proyecto.

FRONTERA: Base de datos fuente. Provee la información necesaria para alimentar las transformaciones del modelo de ventas.

STAGE: Base de datos de stage. Corresponde a la base de datos donde se carga la información volátil proveniente de las fuentes correspondientes, esto con la misma estructura que tienen en su origen.

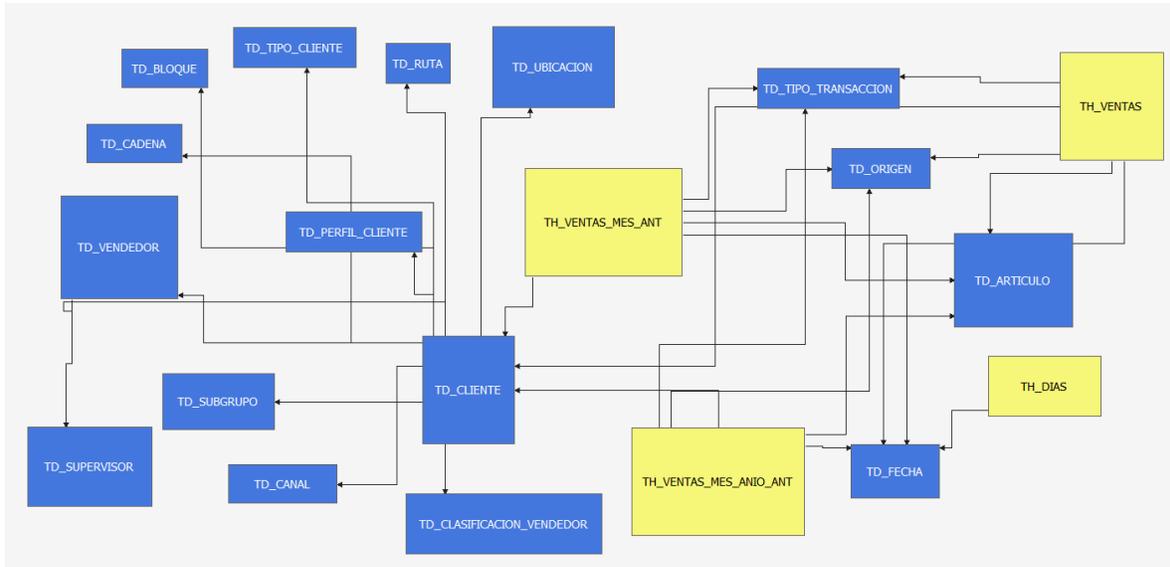
DWH: Base de datos del Data Warehouse. Es una base de datos de consulta. Almacena los objetos producto del proceso de carga de los catálogos e información de las Tablas de Hechos de los modelos (tablas de información cuantitativa e histórica).

Tablas de dimensión: Tabla que contiene en su estructura solo información de corte cualitativo. Los únicos campos numéricos que contiene son los campos de relación con otras tablas.

Tablas de hechos: Contiene la información de corte cuantitativo, en estas estructuras no debe existir información cualitativa. En este proyecto, la tabla de hechos contiene los campos para relacionarse con todas las dimensiones.

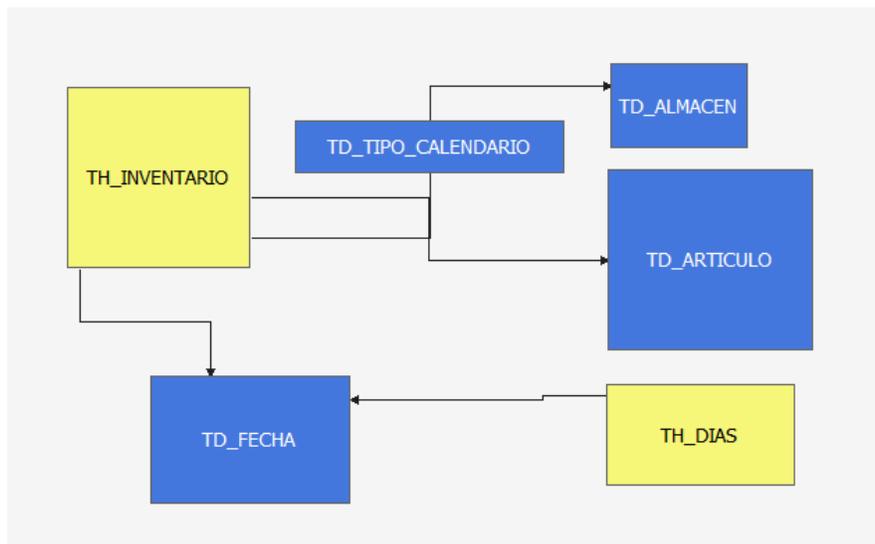
5.1 Diagrama de base de datos:

Figura 27. Diagrama entidad relación TH Ventas.



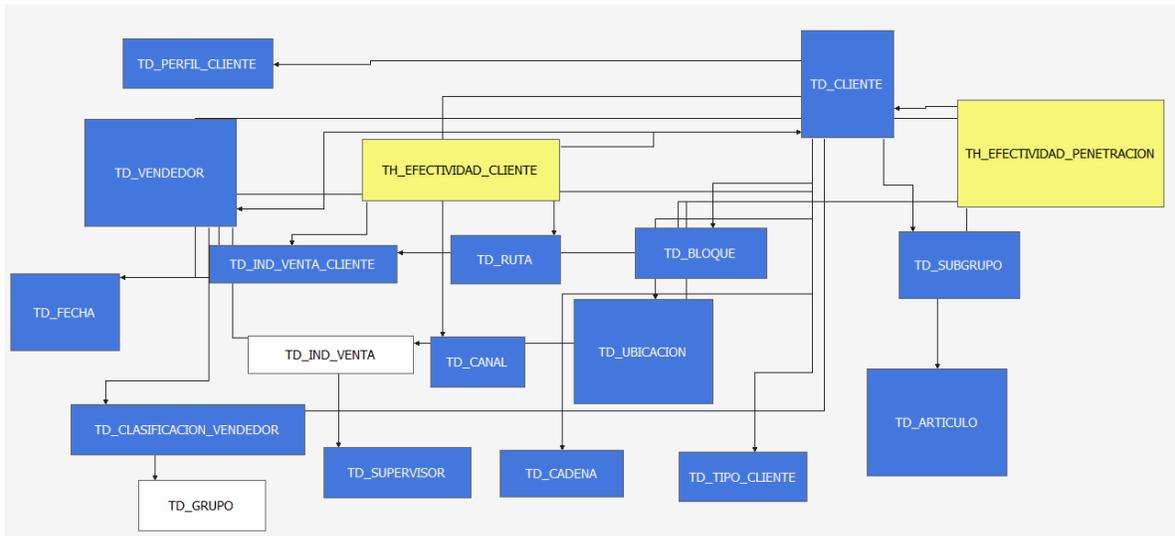
Fuente: Elaboración propia, 2019.

Figura 28. Diagrama entidad relación TH Inventario.



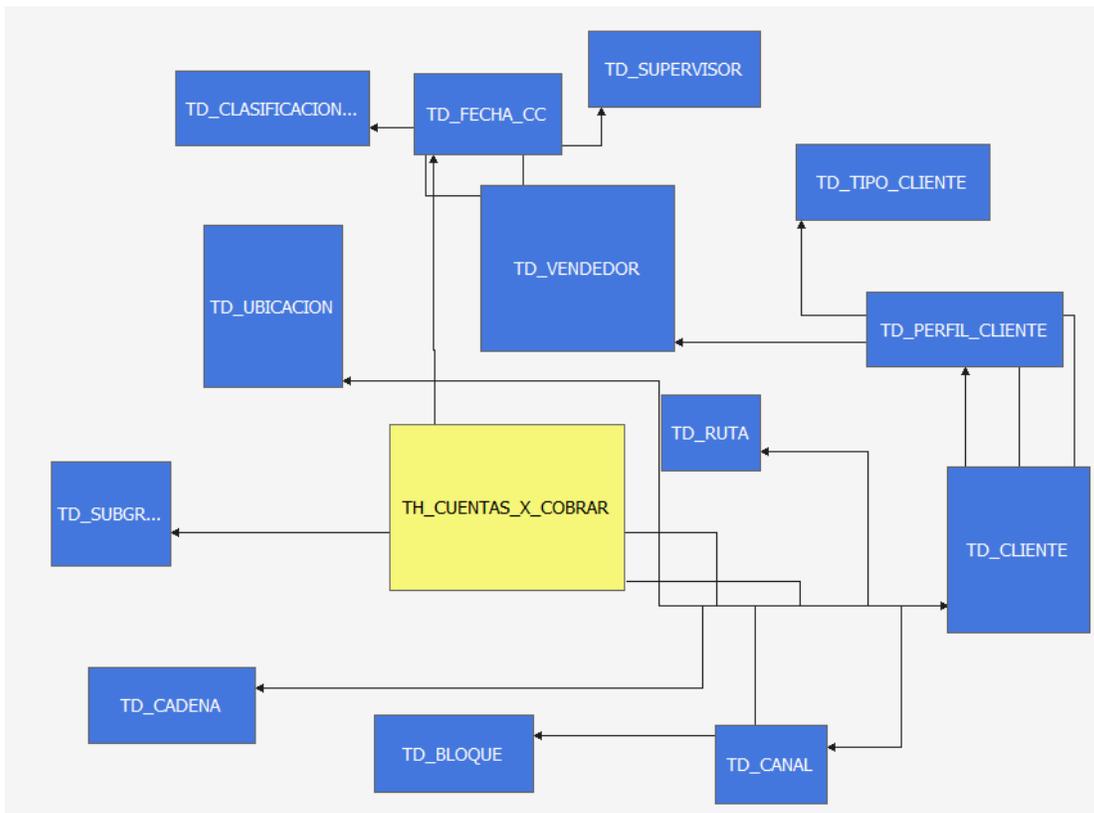
Fuente: Elaboración propia, 2019

Figura 29. Diagrama entidad relación TH Efectividad y penetración.



Fuente: Elaboración propia, 2019

Figura 30. Diagrama entidad relación TH Cuentas por cobrar.



Fuente: Elaboración, 2019

5.2 Tablas

A continuación, se detallan aquellas tablas que no se crean en los procesos de extracción o transformación, es decir, que son equivalentes a tablas del sistema y pertenecen al esquema fijos.

AUX_PARAMETROS

Descripción: Esta tabla almacena los diferentes parámetros que se requiere cargar información, por ejemplo:

Número de días a recargar

Cabe destacar, que la administración de esta tabla es de forma manual, no hay ningún proceso que la realice de forma automática.

Nombre Tabla:	AUX_PARAMETROS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_PARAMETRO	INT	4	
VALOR	VARCHAR	50	
DESCRIPCION	VARCHAR	250	

COMISION_VENDEDOR

Descripción: Esta tabla almacena los parámetros referentes a las comisiones por vendedor.

Nombre Tabla:	COMISION_VENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_COMISION	VARCHAR	20	
PORCENTAJE	DECIMAL	9	18

TD_ORIGEN

Descripción: Esta tabla almacena los atributos de la dimensión de origen.

Nombre Tabla:	TD_ORIGEN		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_ORIGEN	INT		
COD_ORIGEN	VARCHAR	3	
DES_ORIGEN	VARCHAR	50	

TD_TIPO_TRANSACCION

Descripción: Esta tabla almacena los atributos de la dimensión de tipo de transacción.

Nombre Tabla:	TD_TIPO_TRANSACCION		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_TIPO_TRANSACCION	INT		
COD_TIPO_TRANSACCION	VARCHAR	5	
DES_TIPO_TRANSACCION	VARCHAR	50	

A continuación, se detallan las tablas que son generadas en los procesos de extracción y transformación.

5.2.1 Tablas Frontera

Nombre Tabla:	AX365_DistrGeografica		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais_365	VARCHAR	10	
Provincia_365	VARCHAR	10	
Canton_365	VARCHAR	10	
Distrito_365	VARCHAR	10	
Pais_2009	VARCHAR	10	
Provincia_2009	VARCHAR	10	
Canton_2009	VARCHAR	10	
Distrito_2009	VARCHAR	10	
Pais_Pwst	VARCHAR	10	
Provincia_Pwst	VARCHAR	10	
Canton_Pwst	VARCHAR	10	

Distrito_Pwst	VARCHAR	10	
---------------	---------	----	--

Nombre Tabla:	CALENDARIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID	INT		
fecha	DATE		
valor	DECIMAL	5	5
valorGerencial	DECIMAL	5	5

Nombre Tabla:	INVENTARIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
Almacenid	VARCHAR		
CodigoArticulo	VARCHAR		
Saldo	DECIMAL	9	18
Fecha	DATETIME		
InventarioFisico	DECIMAL	9	18

Nombre Tabla:	TD_ALMACEN		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CODIGO	NUMERIC	5	4
ID_DNY	CHAR	10	
DESCRIPCION	CHAR	50	
CreationDateRec	DATETIME		

Nombre Tabla:	TD_ALMACEN_DYN365		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CODIGO	VARCHAR	15	
DESCRIPCION	VARCHAR	50	

Nombre Tabla:	TD_ARQUITECTURA		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_ARQ	CHAR	5	
DES_ARQ	CHAR	60	

Nombre Tabla:	TD_ARTICULO		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_ARTICULO	VARCHAR	20	
COD_CATEGORIA	VARCHAR	10	
COD_FAMILIA	VARCHAR	4	
DES_ARTICULO	VARCHAR	100	
DES_CATEGORIA	VARCHAR	60	
DES_FAMILIA	VARCHAR	100	
PESO	VARCHAR	41	
EMPAQUE	VARCHAR	41	
POCENTAJE	VARCHAR	41	
ACTIVO	NUMERIC	5	1

Nombre Tabla:	TD_ARTICULO_DYN365		
Nombre columna	Tipo de Dato	Longitud	Precisión
CodigoArticulo	VARCHAR	20	
Descripcion	VARCHAR	100	
CodigoFamilia	VARCHAR	4	
DescripFamilia	VARCHAR	100	
CodigoCategoria	VARCHAR	10	
DescripCategoria	VARCHAR	100	
Comision	NUMERIC	5	5
TipoArticulo	VARCHAR	10	
Empaque	NUMERIC	9	12
Activo	VARCHAR	5	
CreationDateRec	DATETIME		

Nombre Tabla:	TD_BLOQUE		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_BLOQUE	INT		
DES_BLOQUE	CHAR	100	

Nombre Tabla:	TD_CANAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_CANAL	NUMERIC	5	5
DES_CANAL	CHAR	100	

Nombre Tabla:	TD_CANTON		
Nombre columna	Tipo de Dato	Longitud	Precisión
departam	NUMERIC	5	5
codigo	NUMERIC	5	5
descrip	VARCHAR	100	
pais	NUMERIC	5	5

Nombre Tabla:	TD_CLASIFCLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
AA	INT	4	
MM	INT	4	
Cod_Cliente	VARCHAR	30	
Monto	DECIMAL	13	20
TipoCliente	VARCHAR	10	

Nombre Tabla:	TD_CLASIFVENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_CLASIF	CHAR	10	
DES_CLASIF	CHAR	100	

Nombre Tabla:	TD CLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
pais	NUMERIC	5	5
CodCliente	VARCHAR	30	
Sucursal	VARCHAR	30	
nombre	VARCHAR	200	
razon	VARCHAR	200	
Cedula	VARCHAR	20	
Direccion	VARCHAR	200	
Telefono	VARCHAR	100	
Cliente	VARCHAR	200	
bloque	NUMERIC	5	5
X	FLOAT	8	53
Y	FLOAT	8	53
Provincia	NUMERIC	5	5
Canton	NUMERIC	5	5
Distrito	NUMERIC	5	5
CanalCliente	NUMERIC	5	5
Inactivo	INT		
Canal	NUMERIC	5	5
Ruta	NUMERIC	5	5
TipoCliente	VARCHAR	10	
Grupo	NUMERIC	5	4
subgrupo	NUMERIC	5	4
Cadena	VARCHAR	5	0
fechaing	DATETIME		
ultimamodificacion	DATETIME		
perfilcuenta	INT		
codigoax	CHAR	10	
perfilcredito	INT		
empresa	NUMERIC	5	5
tipocuenta	NUMERIC	5	1
agencia	NUMERIC	5	5
distrib	NUMERIC	5	6
zona	NUMERIC	5	5
codigo	NUMERIC	5	7
correoelectronico	VARCHAR	200	

Nombre Tabla:	TD_CLIENTE_DYN365		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais	VARCHAR	5	
Agencia	INT		
CodCliente	VARCHAR	30	
Nombre	VARCHAR	200	
Razon	VARCHAR	200	
Cedula	VARCHAR	20	
Direccion	VARCHAR	300	
Telefono	VARCHAR	50	
Bloque	INT		
Provincia	VARCHAR	10	
Canton	VARCHAR	10	
Distrito	VARCHAR	10	
Canal	INT		
Inactivo	VARCHAR	5	
Ruta	INT		
TipoCliente	VARCHAR	10	
Grupo	INT		
SubGrupo	INT		
Cadena	VARCHAR	5	
FechaIngreso	DATETIME		
CreationDateRec	DATETIME		

Nombre Tabla:	TD_DISTRITO		
Nombre columna	Tipo de Dato	Longitud	Precisión
pais	NUMERIC	5	5
departam	NUMERIC	5	5
localidad	NUMERIC	5	5
codigo	NUMERIC	5	5
descrip	VARCHAR	100	

Nombre Tabla:	TD_FORMATO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID	INT		
SEGMENTO	VARCHAR	50	
SECUENCIA	INT		
GRUPOCTA	INT		
IDCANAL	CHAR	10	

IDSECCION	CHAR	10	
IDSUBSECCION	CHAR	10	
PORCENTAJE	INT		
COD_AREA	INT		

Nombre Tabla:	TD_GRUPO		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_GRUPO	NUMERIC	5	5
DES_GRUPO	CHAR	100	

Nombre Tabla:	TD_PERFILCLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_PERFIL	INT		
DES_PERFIL	CHAR		

Nombre Tabla:	TD_PORCENTAJE		
Nombre columna	Tipo de Dato	Longitud	Precisión
IDPORC	INT		
PORCENTAJE	DECIMAL	5	5
FECHA	DATE		

Nombre Tabla:	TD_PROVINCIA		
Nombre columna	Tipo de Dato	Longitud	Precisión
codigo	NUMERIC	5	5
descrip	VARCHAR	100	
pais	NUMERIC	5	5

Nombre Tabla:	TD_RUTA		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_RUTA	NUMERIC	5	5
DES_RUTA	VARCHAR	100	

Nombre Tabla:	TD_SUBGRUPO		
Nombre columna	Tipo de Dato	Longitud	Precisión
grupo	NUMERIC	5	4
codigo	NUMERIC	5	4
descripcion	CHAR	100	

Nombre Tabla:	TD_SUCURSAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_SUCURSAL	VARCHAR	30	
DES_SUCURSAL	VARCHAR	50	

Nombre Tabla:	TD_SUPERVISOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_SUPERVISOR	CHAR	20	
DES_SUPERVISOR	CHAR	100	

Nombre Tabla:	TD_TIPOCLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_TIPO	CHAR	10	
DES_TIPO	CHAR	60	

Nombre Tabla:	TD_VENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_VENDEDOR	VARCHAR	4	
NOMBRE	VARCHAR	100	
ID_RUTA	VARCHAR	41	
ID_CUENTA	NUMERIC	5	7
ID_BLOQUE	NUMERIC	5	5
ID_ALMACEN	NUMERIC	5	4
ID_SUPERVISOR	VARCHAR	20	
COMISION	VARCHAR	100	
ESQUINA	VARCHAR	20	
ID_CLASIFICACION	CHAR	10	

Nombre Tabla:	TH_ANTIGUEDAD		
Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
Saldo	NUMERIC	17	38
PorVencer15	INT		
PorVencer	NUMERIC	17	38
Vencido	NUMERIC	17	38
Vencido8	NUMERIC	17	38
Vencido15	NUMERIC	17	38
Vencido22	NUMERIC	17	38
Vencido30	NUMERIC	17	38
Vencido30mas	NUMERIC	17	38
PorcVencido	NUMERIC	17	38
zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5
fecha	DATE		

Nombre Tabla:	TH_VENTAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais	VARCHAR	2	
Origen	VARCHAR	3	
CodCliente	VARCHAR	20	
Nombre	VARCHAR	200	
Direccion	VARCHAR	200	
Sucursal	NVARCHAR	20	
CodigoProducto	VARCHAR	20	
TipoDoc	NVARCHAR	10	
Producto	VARCHAR	50	
NumDoc	VARCHAR	20	
TipoTransaccion	VARCHAR	3	
Paquetes	NUMERIC	9	10
VentaBruta	NUMERIC	9	18
Impuesto	NUMERIC	9	18
Descuento	NUMERIC	9	18
VentaNeta	NUMERIC	9	18
Fecha	DATETIME		
Telefono	VARCHAR	50	0

CustGroup	VARCHAR	10	0
TransTypePwSt	VARCHAR	10	0

Nombre Tabla:	TH_VENTASPS		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais	VARCHAR	2	
Origen	VARCHAR	3	
CodCliente	NUMERIC	5	7
razon	VARCHAR	200	
Cliente	VARCHAR	200	
Direccion	VARCHAR	200	
CodSucursal	VARCHAR	30	
CodigoProducto	VARCHAR	20	
TipoDoc	VARCHAR	4	
Producto	VARCHAR	100	
TipoTransaccion	VARCHAR	3	
Paquetes	NUMERIC	13	22
VentaBruta	NUMERIC	13	22
Impuesto	NUMERIC	13	22
Descuento	NUMERIC	13	22
VentaNeta	NUMERIC	13	22
fecha	VARCHAR	18	
Telefono	VARCHAR	20	
GustGrupo	VARCHAR	7	
TransTypePwSt	VARCHAR	3	
Empresa	NUMERIC	5	1
CodVendedor	VARCHAR	4	
Clavecarr	VARCHAR	5	
Serie	VARCHAR	5	
Nrodoc	NUMERIC	9	10

Nombre Tabla:	TIPOCAMBIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
Currency	NCHAR	6	
DATE	DATE		
TipoCambio	DECIMAL	9	10

Nombre Tabla:	TIPOCAMBIOHISTAX2009		
Nombre columna	Tipo de Dato	Longitud	Precisión
Currency	NCHAR	6	
DATE	DATE		
TipoCambio	DECIMAL	9	10

Nombre Tabla:	TM_AREAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID	INT		
COD_AREA	INT		
DESCRIP_AREA	NVARCHAR	100	
DPTO	CHAR	10	
SECCION	CHAR	10	
PORCENTAJE	DECIMAL	5	5

Nombre Tabla:	TM_COSTO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID	INT		
AA	INT		
MM	INT		
CODIGO	VARCHAR	20	
DESCRIPCION	VARCHAR	100	
EMBALAJE	INT		
CAJAS	NUMERIC	9	10
COLONES	NUMERIC	9	10
DOLARES	NUMERIC	9	10

Nombre Tabla:	TM_CTACBLE		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID	INT		
AA	INT		
MM	INT		
CODIGO	VARCHAR	100	
DESCRIPCION	VARCHAR	100	
CTA	VARCHAR	50	
DPTO	CHAR	10	
SECCION	CHAR	10	
SUB_SECCION	CHAR	10	
MONTO_REAL	NUMERIC	9	18

Nombre Tabla:	TM_FECHA		
---------------	----------	--	--

Nombre columna	Tipo de Dato	Longitud	Precisión
Fecha	VARCHAR	10	
Estado	INT		

Nombre Tabla:	TM_FINALCXC		
Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
plazo	INT		
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
descrip	VARCHAR	60	
fecha	DATETIME		
vence	DATETIME		
clavecorr	VARCHAR	5	
serie	VARCHAR	5	
nrodoc	NUMERIC	9	10
importe	NUMERIC	13	22
Cancelado	NUMERIC	17	38
zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5

Nombre Tabla:	TM_PERFIL		
Nombre columna	Tipo de Dato	Longitud	Precisión
codigo	INT		
descripcion	VARCHAR	100	
horasontime	SMALLINT		
prioridadfact	INT		
codigoak	VARCHAR	10	
descanalyzer	VARCHAR	150	
perfilcredito	INT		
listapreciofinal	INT		
categoriariesgo	VARCHAR	5	

Nombre Tabla:	TM_PERFILCREDITO
---------------	------------------

Nombre columna	Tipo de Dato	Longitud	Precisión
codigo	INT		
descripcion	VARCHAR	100	

Nombre Tabla:	TM_ROLLBACK		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID	INT		
AA	INT		
MM	INT		
MONTO	DECIMAL	9	18

Nombre Tabla:	TM_TEMPCANCELACIONES		
Nombre columna	Tipo de Dato	Longitud	Precisión
nromovOrigen	NUMERIC	9	15
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nromovCancela	NUMERIC	9	15
nromovCancelado	NUMERIC	9	15
ImporteCancelado	NUMERIC	13	22
ClavecorrC	VARCHAR	5	
serieC	VARCHAR	5	
nrodocC	NUMERIC	9	10

Nombre Tabla:	TM_TEMPCANCFINAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
nromovOrigen	NUMERIC	9	15
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
ImporteCancelado	NUMERIC	17	38

Nombre Tabla:	TM_TEMPCANCFINAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
nromovOrigen	NUMERIC	9	15
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
ImporteCancelado	NUMERIC	17	38
Nombre Tabla:	TM_TEMPCXC		

Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
tipocuenta	NUMERIC	5	1
tipomov	TINYINT		
nromov	NUMERIC	9	15
empresa	NUMERIC	5	5
serie	VARCHAR	5	
clavecorr	VARCHAR	5	
nrodoc	NUMERIC	9	10
fecha	DATETIME		
vence	DATETIME		
descrip	VARCHAR	60	
Plazo	INT		
Limitedinero	NUMERIC	9	16
importe	NUMERIC	13	22
zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5

Nombre Tabla:	TM_TEMPCXC		
Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
tipocuenta	NUMERIC	5	1
tipomov	TINYINT		
nromov	NUMERIC	9	15
empresa	NUMERIC	5	5
serie	VARCHAR	5	
clavecorr	VARCHAR	5	
nrodoc	NUMERIC	9	10
fecha	DATETIME		
vence	DATETIME		
descrip	VARCHAR	60	0
Plazo	INT		
Limitedinero	NUMERIC	9	16
importe	NUMERIC	13	22

zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5

Nombre Tabla:	TMCTASCTES		
Nombre columna	Tipo de Dato	Longitud	Precisión
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nromov	NUMERIC	9	15
tipomov	TINYINT		
empremov	NUMERIC	5	5
fecha	DATETIME		
vence	DATETIME		
descrip	VARCHAR	60	
empresa	NUMERIC	5	5
tipodoc	VARCHAR	4	
clavecorr	VARCHAR	5	
serie	VARCHAR	5	
nrocuota	NUMERIC	5	5
nrodoc	NUMERIC	9	10
importe	NUMERIC	13	22
moneda	NUMERIC	5	4
cancelado	NUMERIC	13	22
tipocambio	NUMERIC	13	22
gencheque	TINYINT		
sucursal	SMALLINT		
bancocheque	NUMERIC	5	5
seriecheque	VARCHAR	30	
numerocheque	NUMERIC	9	10
dtoxprontopago	NUMERIC	13	22
maxdiasprontopago	INT		
tipodocprontopago	VARCHAR	4	
porcdtofp	NUMERIC	5	6
ImpXProntoPago	NUMERIC	13	22
ConfirmaPtoPago	TINYINT		

5..2.2 Tablas Stage

Nombre Tabla:	AX365_DistrGeografica		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais_365	VARCHAR	10	
Provincia_365	VARCHAR	10	
Canton_365	VARCHAR	10	
Distrito_365	VARCHAR	10	
Pais_2009	VARCHAR	10	
Provincia_2009	VARCHAR	10	
Canton_2009	VARCHAR	10	
Distrito_2009	VARCHAR	10	
Pais_Pwst	VARCHAR	10	
Provincia_Pwst	VARCHAR	10	
Canton_Pwst	VARCHAR	10	
Distrito_Pwst	VARCHAR	10	

Nombre Tabla:	CALENDARIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
fecha	DATE		
valor	DECIMAL	5	5
valorGerencial	DECIMAL	5	5

Nombre Tabla:	INVENTARIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
fecha	INT		
Almacenid	VARCHAR	10	
CodigoArticulo	VARCHAR	20	
Saldo	DECIMAL	9	18
Fecha	DATETIME	8	23
InventarioFisico	DECIMAL	9	18

Nombre Tabla:	TD_ALMACEN		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CODIGO	NUMERIC	5	4
ID_DNY	CHAR	10	
DESCRIPCION	CHAR	50	

Nombre Tabla:	TD_ARQUITECTURA		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_ARQ	CHAR	5	
DES_ARQ	CHAR	60	

Nombre Tabla:	TD_ARTICULO		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_ARTICULO	VARCHAR	20	
COD_CATEGORIA	VARCHAR	10	
COD_FAMILIA	VARCHAR	4	
DES_ARTICULO	VARCHAR	100	
DES_CATEGORIA	VARCHAR	60	
DES_FAMILIA	VARCHAR	100	
PESO	VARCHAR	41	
EMPAQUE	VARCHAR	41	
POCENTAJE	VARCHAR	41	
ACTIVO	INT		

Nombre Tabla:	TD_BLOQUE		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_BLOQUE	INT		
DES_BLOQUE	CHAR	100	

Nombre Tabla:	TD_CANAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_CANAL	DECIMAL	5	5
DES_CANAL	CHAR	100	

Nombre Tabla:	TD_CANTON		
Nombre columna	Tipo de Dato	Longitud	Precisión
departam	DECIMAL	5	5
codigo	DECIMAL	5	5
descrip	VARCHAR	100	
pais	DECIMAL	5	5

Nombre Tabla:	TD_CLASIFVENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_CLASIF	CHAR	10	
DES_CLASIF	CHAR	100	

Nombre Tabla:	TD_CLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
pais	DECIMAL	5	5
CodCliente	VARCHAR	30	
Sucursal	VARCHAR	30	
nombre	VARCHAR	200	
razon	VARCHAR	200	
Cedula	VARCHAR	20	
Direccion	VARCHAR	200	
Telefono	VARCHAR	100	
Cliente	VARCHAR	200	
bloque	DECIMAL	5	5
X	FLOAT	8	53
Y	FLOAT	8	53
Provincia	DECIMAL	5	5
Canton	DECIMAL	5	5
Distrito	DECIMAL	5	5
CanalCliente	DECIMAL	5	5
Inactivo	INT		
Canal	DECIMAL	5	5
Ruta	DECIMAL	5	5
TipoCliente	VARCHAR	10	
Grupo	DECIMAL	5	4
subgrupo	DECIMAL	5	4
Cadena	VARCHAR	5	
fechaing	SMALLDATETIME		
ultimamodificacion	SMALLDATETIME		
perfilcuenta	INT	4	10
Codigo_AX	CHAR	10	
correoelectronico	VARCHAR	200	

Nombre Tabla:	TD_DISTRITO		
Nombre columna	Tipo de Dato	Longitud	Precisión
pais	DECIMAL	5	5
departam	DECIMAL	5	5
localidad	DECIMAL	5	5
codigo	DECIMAL	5	5
descrip	VARCHAR	100	

Nombre Tabla:	TD_GRUPO		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_GRUPO	DECIMAL	5	5
DES_GRUPO	CHAR	100	

Nombre Tabla:	TD_OBJETIVO		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_ART	VARCHAR	20	
FECHA	NCHAR	20	

Nombre Tabla:	TD_PERFILCLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_PERFIL	INT		
DES_PERFIL	CHAR	100	

Nombre Tabla:	TD_PORCENTAJE		
Nombre columna	Tipo de Dato	Longitud	Precisión
IDPORC	INT		
PORCENTAJE	DECIMAL	5	5
FECHA	DATE		

Nombre Tabla:	TD_PROVINCIA		
Nombre columna	Tipo de Dato	Longitud	Precisión
codigo	DECIMAL	5	5
descrip	VARCHAR	100	
pais	DECIMAL	5	5

Nombre Tabla:	TD_RUTA		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_RUTA	DECIMAL	5	5
DES_RUTA	VARCHAR	100	

Nombre Tabla:	TD_SUBGRUPO		
Nombre columna	Tipo de Dato	Longitud	Precisión
grupo	DECIMAL	5	4
codigo	DECIMAL	5	4
descripcion	CHAR	100	

Nombre Tabla:	TD_SUCURSAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_SUCURSAL	VARCHAR	30	
DES_SUCURSAL	VARCHAR	50	

Nombre Tabla:	TD_SUPERVISOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_SUPERVISOR	CHAR	20	
DES_SUPERVISOR	CHAR	100	

Nombre Tabla:	TD_TIPOCLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_TIPO	CHAR	10	
DES_TIPO	CHAR	60	

Nombre Tabla:	TD_VENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_VENDEDOR	VARCHAR	4	
NOMBRE	VARCHAR	100	
ID_RUTA	VARCHAR	41	
ID_CUENTA	DECIMAL	5	7
ID_BLOQUE	DECIMAL	5	5
ID_ALMACEN	DECIMAL	5	4
COMISION	VARCHAR	100	
ESQUINA	VARCHAR	20	
ID_SUPERVISOR	VARCHAR	20	

ID_CLASIFICACION	VARCHAR	10	
------------------	---------	----	--

Nombre Tabla:	TH_ANTIGUEDAD		
Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
Saldo	NUMERIC	17	38
PorVencer15	INT		
PorVencer	NUMERIC	17	38
Vencido	NUMERIC	17	38
Vencido8	NUMERIC	17	38
Vencido15	NUMERIC	17	38
Vencido22	NUMERIC	17	38
Vencido30	NUMERIC	17	38
Vencido30mas	NUMERIC	17	38
PorcVencido	NUMERIC	17	38
zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5
Fecha	DATE		

Nombre Tabla:	TH_VENTAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais	VARCHAR	2	
Origen	VARCHAR	3	
CodCliente	VARCHAR	20	
Nombre	VARCHAR	200	
Direccion	VARCHAR	200	
Sucursal	NVARCHAR	20	
CodigoProducto	VARCHAR	20	
TipoDoc	NVARCHAR	10	
Producto	VARCHAR	50	
NumDoc	VARCHAR	20	
TipoTransaccion	VARCHAR	3	
Paquetes	DECIMAL	9	10
VentaBruta	DECIMAL	9	18
Impuesto	DECIMAL	9	18
Descuento	DECIMAL	9	18

VentaNeta	DECIMAL	9	18
Fecha	SMALLDATETIME		
Telefono	VARCHAR	50	
CustGroup	VARCHAR	10	
TransTypePwSt	VARCHAR	10	

Nombre Tabla:	TH_VENTASPS		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais	VARCHAR	2	
Origen	VARCHAR	3	
CodCliente	DECIMAL	5	7
razon	VARCHAR	200	
Cliente	VARCHAR	200	
Direccion	VARCHAR	200	
CodSucursal	VARCHAR	30	
CodigoProducto	VARCHAR	20	
TipoDoc	VARCHAR	4	
Producto	VARCHAR	100	
TipoTransaccion	VARCHAR	3	
Paquetes	DECIMAL	13	22
VentaBruta	DECIMAL	13	22
Impuesto	DECIMAL	13	22
Descuento	DECIMAL	13	22
VentaNeta	DECIMAL	13	22
fecha	VARCHAR	18	
Telefono	VARCHAR	20	
GustGrupo	VARCHAR	7	
TransTypePwSt	VARCHAR	3	
Empresa	DECIMAL	5	1
CodVendedor	VARCHAR	4	
Clavecorr	VARCHAR	5	
Serie	VARCHAR	5	
Nrodoc	VARCHAR	50	

Nombre Tabla:	TIPOCAMBIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
Currency	VARCHAR	6	
DATE	DATE		
TipoCambio	DECIMAL	9	10

Nombre Tabla:	TIPOCAMBIOHISTAX2009		
Nombre columna	Tipo de Dato	Longitud	Precisión
Currency	NCHAR	6	
DATE	DATE		
TipoCambio	DECIMAL	9	10

Nombre Tabla:	TM_CANCELAC		
Nombre columna	Tipo de Dato	Longitud	Precisión
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nromov	NUMERIC	9	15
canc_tc	NUMERIC	5	1
canc_cuen	NUMERIC	5	7
canc_movi	NUMERIC	9	15
canc_impo	NUMERIC	13	22
SaldoAnteriorFact	NUMERIC	13	22
FechaCancelacion	DATETIME		

Nombre Tabla:	TM_FINALCXC		
Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
plazo	INT		
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
descrip	VARCHAR	60	
fecha	DATETIME		
vence	DATETIME		
clavecorr	VARCHAR	5	
serie	VARCHAR	5	
nrodoc	NUMERIC	9	10
importe	NUMERIC	13	22
Cancelado	NUMERIC	17	38
zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5
Nombre Tabla:	TM_LIMITEPEFILCREDITO		

Nombre columna	Tipo de Dato	Longitud	Precisión
perfilcredito	INT		
empresa	NUMERIC	5	5
moneda	NUMERIC	5	4
modo	TINYINT		
importe	NUMERIC	9	16
dias	INT		
vencimientos	INT		

Nombre Tabla:	TM_PERFIL		
Nombre columna	Tipo de Dato	Longitud	Precisión
codigo	INT		
descripcion	VARCHAR	100	
horasontime	SMALLINT		
prioridadfact	INT		
codigoak	VARCHAR	10	
descanalyzer	VARCHAR	150	
perfilcredito	INT		
listapreciofinal	INT		
categoriariesgo	VARCHAR	5	

Nombre Tabla:	TM_PERFILCREDITO		
Nombre columna	Tipo de Dato	Longitud	Precisión
codigo	INT		
descripcion	VARCHAR	100	

Nombre Tabla:	TM_TEMPCANCELACIONES		
Nombre columna	Tipo de Dato	Longitud	Precisión
nromovOrigen	NUMERIC	9	15
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nromovCancela	NUMERIC	9	15
nromovCancelado	NUMERIC	9	15
ImporteCancelado	NUMERIC	13	22
ClavecorrC	VARCHAR	5	
serieC	VARCHAR	5	
nrodocC	NUMERIC	9	10
Nombre Tabla:	TM_TEMPCANCFINAL		

Nombre columna	Tipo de Dato	Longitud	Precisión
nromovOrigen	NUMERIC	9	15
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
ImporteCancelado	NUMERIC	17	38

Nombre Tabla:	TM_TEMPCXC		
Nombre columna	Tipo de Dato	Longitud	Precisión
distribuidor	NUMERIC	5	6
agencia	NUMERIC	5	6
cuenta	NUMERIC	5	7
nombre	VARCHAR	200	
tipocuenta	NUMERIC	5	1
tipomov	TINYINT		
nromov	NUMERIC	9	15
empresa	NUMERIC	5	5
serie	VARCHAR	5	
clavecorr	VARCHAR	5	
nrodoc	NUMERIC	9	10
fecha	DATETIME		
vence	DATETIME		
descrip	VARCHAR	60	0
Plazo	INT		
Limitedinero	NUMERIC	9	16
importe	NUMERIC	13	22
zonavent	NUMERIC	5	5
zonarepa	NUMERIC	5	5

Nombre Tabla:	TMCTASCTES		
Nombre columna	Tipo de Dato	Longitud	Precisión
tipocuenta	NUMERIC	5	1
cuenta	NUMERIC	5	7
nromov	NUMERIC	9	15
tipomov	TINYINT		
empremov	NUMERIC	5	5
fecha	DATETIME		
vence	DATETIME		
descrip	VARCHAR	60	
empresa	NUMERIC	5	5
tipodoc	VARCHAR	4	
clavecorr	VARCHAR	5	
serie	VARCHAR	5	
nrocuota	NUMERIC	5	5
nrodoc	NUMERIC	9	10
importe	NUMERIC	13	22
moneda	NUMERIC	5	4
cancelado	NUMERIC	13	22
tipocambio	NUMERIC	13	22
gencheque	TINYINT		
sucursal	SMALLINT		
bancocheque	NUMERIC	5	5
seriecheque	VARCHAR	30	
numerocheque	NUMERIC	9	10
dtoxprontopago	NUMERIC	13	22
maxdiasprontopago	INT		
tipodocprontopago	VARCHAR	4	
porcdtofp	NUMERIC	5	6
ImpXProntoPago	NUMERIC	13	22
ConfirmaPtoPago	TINYINT		

5.2.3 Tablas DWH

Nombre Tabla:	AUX_CLIENTES_EXCLUIR_VENTAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_CLIENTE	VARCHAR	30	

Nombre Tabla:	AUX_EFECTIVIDAD_POR_CLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
IND_VENTA_GENERAL	DECIMAL	13	24
COD_CLIENTE	VARCHAR	30	

Nombre Tabla:	AUX_PAIS		
Nombre columna	Tipo de Dato	Longitud	Precisión
Pais_365	VARCHAR	10	
Pais_2009	VARCHAR	10	
Pais_Pwst	VARCHAR	10	

Nombre Tabla:	AUX_PARAMETROS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_PARAMETRO	INT	4	
VALOR	VARCHAR	50	
DESCRIPCION	VARCHAR	250	

Nombre Tabla:	AUX_TIPO_CAMBIO_VENTAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT	4	
TIPO_CAMBIO	DECIMAL	13	24

Nombre Tabla:	AUX_TOTAL_CLIENTES_ARTICULOS		
Nombre columna	Tipo de Dato	Longitud	Precisión
COD_CLIENTE	VARCHAR	30	
COD_ARTICULO	VARCHAR	20	

Nombre Tabla:	PORCENTAJE_CRECIMIENTO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_PORCENTAJE	INT		
PORCENTAJE	DECIMAL	13	28
ID_MES	INT		

Nombre Tabla:	PORCENTAJE_CRECIMIENTO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_ALMACEN	INT		
COD_ALMACEN	INT		
DES_ALMACEN	VARCHAR	50	0

Nombre Tabla:	TD_ALMACEN		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_ALMACEN	INT	4	10
COD_ALMACEN	INT	4	10
DES_ALMACEN	VARCHAR	50	0

Nombre Tabla:	TD_ARTICULO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_ARTICULO	INT		
COD_ARTICULO	VARCHAR	20	
COD_CATEGORIA	VARCHAR	10	
COD_FAMILIA	VARCHAR	4	
DES_ARTICULO	VARCHAR	100	
DES_CATEGORIA	VARCHAR	60	
DES_FAMILIA	VARCHAR	100	
PESO	VARCHAR	20	
EMPAQUE	VARCHAR	20	
PORCENTAJE	VARCHAR	20	
COD_ACTIVO	INT		
DES_ACTIVO	VARCHAR	25	
IND_OBJETIVO	VARCHAR	25	

Nombre Tabla:	TD_BLOQUE		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_BLOQUE	INT		
COD_BLOQUE	INT		
DES_BLOQUE	VARCHAR	100	

Nombre Tabla:	TD_CADENA		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CADENA	INT		
COD_CADENA	VARCHAR	5	
DES_CADENA	VARCHAR	100	

Nombre Tabla:	TD_CANAL		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CANAL	INT		
COD_CANAL	INT		
DES_CANAL	VARCHAR	100	0

Nombre Tabla:	TD_CLASIFICACION_VENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CLASIFICACION_VENDEDOR	INT		
COD_CLASIFICACION_VENDEDOR	VARCHAR	10	
DES_CLASIFICACION_VENDEDOR	VARCHAR	100	

Nombre Tabla:	TD_FECHA		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
FECHA	DATE		
ANIO	VARCHAR	4	
MES	VARCHAR	7	
DIA	VARCHAR	10	

Nombre Tabla:	TD_GRUPO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_GRUPO	INT		
COD_GRUPO	INT		
DES_GRUPO	VARCHAR	100	0

Nombre Tabla:	TD_PERFIL_CLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_PERFIL_CLIENTE	INT		
COD_PERFIL_CLIENTE	INT		
DES_PERFIL_CLIENTE	VARCHAR	100	

Nombre Tabla:	TD_RUTA		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_RUTA	INT		
COD_RUTA	INT		
DES_RUTA	VARCHAR	100	

Nombre Tabla:	TD_CLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_CLIENTE	INT		
COD_CLIENTE	VARCHAR	30	
ID_SUCURSAL	INT		
COD_SUCURSAL	VARCHAR	30	
ID_BLOQUE	INT		
COD_BLOQUE	INT		
ID_UBICACION	INT		
COD_PAIS	INT		
COD_PROVINCIA	INT		
COD_CANTON	INT		
COD_DISTRITO	INT		
IND_INACTIVO	INT		
ID_CANAL	INT		
COD_CANAL	INT		
ID_RUTA	INT		
COD_RUTA	INT		
ID_TIPO_CLIENTE	INT		
COD_TIPO_CLIENTE	VARCHAR	10	

ID_SUBGRUPO	INT		
ID_GRUPO	INT		
COD_GRUPO	INT		
COD_SUBGRUPO	INT		
ID_CADENA	INT		
COD_CADENA	VARCHAR	5	0
ID_PERFIL_CLIENTE	INT		
COD_PERFIL_CLIENTE	INT		
ID_VENDEDOR	INT		
COD_VENDEDOR	VARCHAR	4	
DES_INACTIVO	VARCHAR	20	
DES_CLIENTE	VARCHAR	200	
DES_RAZON	VARCHAR	200	
DES_CEDULA	VARCHAR	20	
DES_DIRECCION	VARCHAR	200	
DES_TELEFONO	VARCHAR	100	
DES_X	FLOAT	8	53
DES_Y	FLOAT	8	53
FECHA_INGRESO	DATE		
COD_AX	CHAR	10	
DES_CORREO	VARCHAR	200	

Nombre Tabla:	TD_SUBGRUPO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_SUBGRUPO	INT		
COD_SUBGRUPO	INT		
DES_SUBGRUPO	VARCHAR	100	0
COD_GRUPO	INT		
DES_GRUPO	VARCHAR	100	0

Nombre Tabla:	TD_SUPERVISOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_SUPERVISOR	INT		
COD_SUPERVISOR	VARCHAR	20	
DES_SUPERVISOR	VARCHAR	100	

Nombre Tabla:	TD_TIPO_CLIENTE		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_TIPO_CLIENTE	INT		
COD_TIPO_CLIENTE	VARCHAR	10	
DES_TIPO_CLIENTE	VARCHAR	100	

Nombre Tabla:	TD_UBICACION		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_UBICACION	INT		
COD_PAIS	INT		
DES_PAIS	VARCHAR	50	
COD_PROVINCIA	INT		
DES_PROVINCIA	VARCHAR	50	
COD_CANTON	INT		
DES_CANTON	VARCHAR	50	
COD_DISTRITO	INT		
DES_DISTRITO	VARCHAR	50	

Nombre Tabla:	TD_VENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_VENDEDOR	INT		
COD_VENDEDOR	VARCHAR	4	
DES_VENDEDOR	VARCHAR	100	
COD_RUTA	INT		
COD_BLOQUE	INT		
COD_ALMACEN	INT		
ID_SUPERVISOR	INT		
COD_SUPERVISOR	VARCHAR	20	
ID_CLASIFICACION_VENDEDOR	INT		
COD_CLASIFICACION_VENDEDOR	VARCHAR	10	
COD_COMISION	VARCHAR	20	
COD_ESQUINA	VARCHAR	20	
DES_ALMACEN	VARCHAR	50	

Nombre Tabla:	TEMP_TOTAL_ARTICULOS_ACT_HIST		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
COD_ARTICULO	VARCHAR	20	

PERIODO	INT		
---------	-----	--	--

Nombre Tabla:		TEMP_TOTAL_CLIENTES_ACT_HIST	
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
COD_CLIENTE	VARCHAR	30	0
PERIODO	INT		

Nombre Tabla:		TH_CUENTAS_X_COBRAR	
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
ID_CLIENTE	INT		
COD_CLIENTE	VARCHAR	30	
SALDO_ACTUAL	DECIMAL	13	28
SALDO_POR_VENCER	DECIMAL	13	28
SALDO_VENCIDO	DECIMAL	13	28
SALDO_VENCIDO_8	DECIMAL	13	28
SALDO_VENCIDO_15	DECIMAL	13	28
SALDO_VENCIDO_22	DECIMAL	13	28
SALDO_VENCIDO_30	DECIMAL	13	28
SALDO_VENCIDO_30_MAS	DECIMAL	13	28

Nombre Tabla:		TH_DIAS	
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
DIAS_TOTALES	DECIMAL	5	8
DIAS_TRANSCURRIDOS	DECIMAL	5	8
DIAS_FALTANTES	DECIMAL	5	8
DIAS_TOTALES_GERENCIA	DECIMAL	5	8
DIAS_TRANSCURRIDOS_GERENCIA	DECIMAL	5	8
DIAS_FALTANTES_GERENCIA	DECIMAL	5	8

Nombre Tabla:		TH_EFECTIVIDAD_PENETRACION	
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT	4	
ID_ARTICULO	INT	4	

ID_CLIENTE	INT	4	
COD_CLIENTE	VARCHAR	30	
COD_ARTICULO	VARCHAR	30	
IND_VENTA	INT	4	
PERIODO	INT	4	
IND_VENTA_GENERAL	INT	4	

Nombre Tabla:	TH_INVENTARIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
ID_ALMACEN	INT		
COD_ALMACEN	INT		
ID_ARTICULO	INT		
COD_ARTICULO	VARCHAR	30	
UNIDADES_VENDIDAS_MES	DECIMAL	13	28
UNIDADES_VENDIDAS_MES_CEDI	DECIMAL	13	28
UNIDADES_INVENTARIO	DECIMAL	13	28
SALDO_INVENTARIO	DECIMAL	13	28
INVENTARIO_FISICO	DECIMAL	13	28
CIERRE_PROYECTADO	DECIMAL	13	28
CIERRE_PROYECTADO_GERENCIA	DECIMAL	13	28
CIERRE_PROYECTADO_CEDI	DECIMAL	13	28
CIERRE_PROYECTADO_GERENCIA_CEDI	DECIMAL	13	28

Nombre Tabla:	TH_VENTAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
ID_ORIGEN	INT		
ID_CLIENTE	INT		
ID_ARTICULO	INT		
ID_TIPO_TRANSACCION	INT		
ID_ALMACEN	INT		
COD_PAIS	INT		
COD_ORIGEN	VARCHAR	3	
COD_CLIENTE	VARCHAR	30	
COD_SUCURSAL	VARCHAR	30	
COD_ARTICULO	VARCHAR	30	

COD_TIPO_TRANSACCION	VARCHAR	5	
NUM_DOCUMENTO	VARCHAR	50	
COD_ALMACEN	VARCHAR	10	
VENTA_BRUTA	DECIMAL	13	28
VENTA_BRUTA_USD	DECIMAL	13	28
IMPUESTO	DECIMAL	13	28
IMPUESTO_USD	DECIMAL	13	28
DESCUENTO	DECIMAL	13	28
DESCUENTO_USD	DECIMAL	13	28
PAQUETES	DECIMAL	13	28
UNIDADES	DECIMAL	13	28
VENTA_NETA	DECIMAL	13	28
VENTA_NETA_USD	DECIMAL	13	28
COMISION	DECIMAL	13	28
COMISION_USD	DECIMAL	13	28
COMISION_REAL	DECIMAL	13	28
COMISION_REAL_USD	DECIMAL	13	28

Nombre Tabla:	TH_VENTAS_MES_ANIO_ANT		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
ID_ORIGEN	INT		
ID_CLIENTE	INT		
ID_ARTICULO	INT		
ID_TIPO_TRANSACCION	INT		
COD_PAIS	INT		
COD_ORIGEN	VARCHAR	3	
COD_CLIENTE	VARCHAR	30	
COD_ARTICULO	VARCHAR	30	
COD_TIPO_TRANSACCION	VARCHAR	5	
PORCENTAJE_CRECIMIENTO	DECIMAL	13	28
VENTA_NETA	DECIMAL	13	28
VENTA_NETA_USD	DECIMAL	13	28
VENTA_BRUTA	DECIMAL	13	28
VENTA_BRUTA_USD	DECIMAL	13	28
VENTA_META_NETA	DECIMAL	13	28
VENTA_META_NETA_USD	DECIMAL	13	28
UNIDADES_VENDIDAS	DECIMAL	13	28
PAQUETES	DECIMAL	13	28

Nombre Tabla:	TH_VENTAS_MES_ANT		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
ID_ORIGEN	INT		
ID_CLIENTE	INT		
ID_ARTICULO	INT		
ID_TIPO_TRANSACCION	INT		
COD_PAIS	INT		
COD_ORIGEN	VARCHAR	3	
COD_CLIENTE	VARCHAR	30	
COD_ARTICULO	VARCHAR	30	
COD_TIPO_TRANSACCION	VARCHAR	5	
VENTA_NETA	DECIMAL	13	28
VENTA_NETA_USD	DECIMAL	13	28
VENTA_BRUTA	DECIMAL	13	28
VENTA_BRUTA_USD	DECIMAL	13	28
UNIDADES_VENDIDAS	DECIMAL	13	28
PAQUETES	DECIMAL	13	28

Nombre Tabla:	THT_CUENTAS_X_COBRAR		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
COD_CLIENTE	VARCHAR	30	
SALDO_ACTUAL	DECIMAL	13	28
SALDO_POR_VENCER	DECIMAL	13	28
SALDO_VENCIDO	DECIMAL	13	28
SALDO_VENCIDO_8	DECIMAL	13	28
SALDO_VENCIDO_15	DECIMAL	13	28
SALDO_VENCIDO_22	DECIMAL	13	28
SALDO_VENCIDO_30	DECIMAL	13	28
SALDO_VENCIDO_30_MAS	DECIMAL	13	28

Nombre Tabla:	THT_DIAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
DIAS_TOTALES	DECIMAL	5	8
DIAS_TRANSCURRIDOS	DECIMAL	5	8
DIAS_FALTANTES	DECIMAL	5	8

DIAS_TOTALES_GERENCIA	DECIMAL	5	8
DIAS_TRANSCURRIDOS_GERENCIA	DECIMAL	5	8
DIAS_FALTANTES_GERENCIA	DECIMAL	5	8

Nombre Tabla:	THT_EFECTIVIDAD_PENETRACION		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
COD_CLIENTE	VARCHAR	30	
COD_ARTICULO	VARCHAR	30	
IND_VENTA	INT		
IND_VENTA_GENERAL	INT		
PERIODO	INT		

Nombre Tabla:	THT_INVENTARIO		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
COD_ALMACEN	INT		
COD_ARTICULO	VARCHAR	30	
UNIDADES_VENDIDAS_MES	DECIMAL	13	28
UNIDADES_VENDIDAS_MES_CEDI	DECIMAL	13	28
UNIDADES_INVENTARIO	DECIMAL	13	28
SALDO_INVENTARIO	DECIMAL	13	28
INVENTARIO_FISICO	DECIMAL	13	28
CIERRE_PROYECTADO	DECIMAL	13	28
CIERRE_PROYECTADO_GERENCIA	DECIMAL	13	28
CIERRE_PROYECTADO_CEDI	DECIMAL	13	28
CIERRE_PROYECTADO_GERENCIA_CEDI	DECIMAL	13	28

Nombre Tabla:	THT_VENTAS		
Nombre columna	Tipo de Dato	Longitud	Precisión
ID_FECHA	INT		
COD_PAIS	INT		
COD_ORIGEN	VARCHAR	3	
COD_CLIENTE	VARCHAR	30	
COD_SUCURSAL	VARCHAR	30	
COD_ARTICULO	VARCHAR	30	
COD_TIPO_TRANSACCION	VARCHAR	5	

NUM_DOCUMENTO	VARCHAR	50	
VENTA_BRUTA	DECIMAL	13	28
IMPUESTO	DECIMAL	13	28
DESCUENTO	DECIMAL	13	28
PAQUETES	DECIMAL	13	28
VENTA_NETA	DECIMAL	13	28

Nombre Tabla:		THT_VENTAS_MES_ANIO_ANT		
Nombre columna	Tipo de Dato	Longitud	Precisión	
ID_FECHA	INT			
ID_MES	INT			
ID_ORIGEN	INT			
ID_CLIENTE	INT			
ID_ARTICULO	INT			
ID_TIPO_TRANSACCION	INT			
COD_PAIS	INT			
COD_ORIGEN	VARCHAR	3		
COD_CLIENTE	VARCHAR	30		
COD_ARTICULO	VARCHAR	30		
COD_TIPO_TRANSACCION	VARCHAR	5		
VENTA_NETA	DECIMAL	13	28	
VENTA_NETA_USD	DECIMAL	13	28	
VENTA_BRUTA	DECIMAL	13	28	
VENTA_BRUTA_USD	DECIMAL	13	28	
UNIDADES_VENDIDAS	DECIMAL	13	28	
PAQUETES	DECIMAL	13	28	

Nombre Tabla:		THT_VENTAS_MES_ANT		
Nombre columna	Tipo de Dato	Longitud	Precisión	
ID_FECHA	INT			
ID_ORIGEN	INT			
ID_CLIENTE	INT			
ID_ARTICULO	INT			
ID_TIPO_TRANSACCION	INT			
COD_PAIS	INT			
COD_ORIGEN	VARCHAR	3		
COD_CLIENTE	VARCHAR	30		
COD_ARTICULO	VARCHAR	30		

COD_TIPO_TRANSACCION	VARCHAR	5	
VENTA_NETA	DECIMAL	13	28
VENTA_NETA_USD	DECIMAL	13	28
VENTA_BRUTA	DECIMAL	13	28
VENTA_BRUTA_USD	DECIMAL	13	28
UNIDADES_VENDIDAS	DECIMAL	13	28
PAQUETES	DECIMAL	13	28

Nombre Tabla:		TOTAL ARTICULOS_ACT_HIST		
Nombre columna	Tipo de Dato	Longitud	Precisión	
ID_FECHA	INT			
COD_ARTICULO	VARCHAR	20		
PERIODO	INT			

Nombre Tabla:		TOTAL ARTICULOS_ACT_HIST		
Nombre columna	Tipo de Dato	Longitud	Precisión	
ID_FECHA	INT			
COD_CLIENTE	VARCHAR	30		
PERIODO	INT			

Nombre Tabla:		Fijos.COMISION_VENDEDOR		
Nombre columna	Tipo de Dato	Longitud	Precisión	
COD_COMISION	VARCHAR	20		
PORCENTAJE	DECIMAL	9	18	

Nombre Tabla:		Fijos.TD_ORIGEN		
Nombre columna	Tipo de Dato	Longitud	Precisión	
ID_ORIGEN	INT			
COD_ORIGEN	VARCHAR	3		
DES_ORIGEN	VARCHAR	50		

Nombre Tabla:		TD_TIPO_TRANSACCION		
Nombre columna	Tipo de Dato	Longitud	Precisión	
ID_TIPO_TRANSACCION	INT			
COD_TIPO_TRANSACCION	VARCHAR	5		
DES_TIPO_TRANSACCION	VARCHAR	50		

5.3 Jobs

Durante el desarrollo del proyecto se creó un Job que se detalla a continuación.

DWH

Descripción: Este Job se localiza en la base de datos msdb. Es el encargado de ejecutar todos los paquetes de Integration Services necesarios, para la extracción y transformación de la información del modelo de ventas. Este Job está calendarizado para que se ejecute todos los días, a las 3:00 am.

5.4 Diseño de SQL Server Integration Services (SSIS)

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. En este caso se utiliza una base de datos almacenada en SQL (2017). La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Una parte intrínseca del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. De no ser así los datos son rechazados.

Un requerimiento importante que se debe exigir a la tarea de extracción es que ésta cause un impacto mínimo en el sistema origen. Si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que éste no pueda utilizarse con normalidad para su uso cotidiano. Por esta razón, las operaciones de extracción suelen programarse en horarios o días donde este impacto sea nulo o mínimo.

Las extracciones para este proyecto se realizan de la BD llamada “FRONTERA”, esta se encuentra en un motor de base de datos SQL 2017.

5.5 Extracciones desde la base de datos

EXT-SQL-XXXX.dtsx (XXXXXX Corresponde al nombre de tabla)

Este tipo de paquete extrae de una de las tablas de la base de datos “FRONTERA” y procesa la información de la tabla con el mismo nombre del paquete, esto en la Base de datos llamada STAGE, de la conexión con el mismo nombre. Se conserva la estructura de la tabla fuente.

El proceso que se realiza para la extracción de las tablas, provenientes de la base de datos del cliente, es el siguiente:

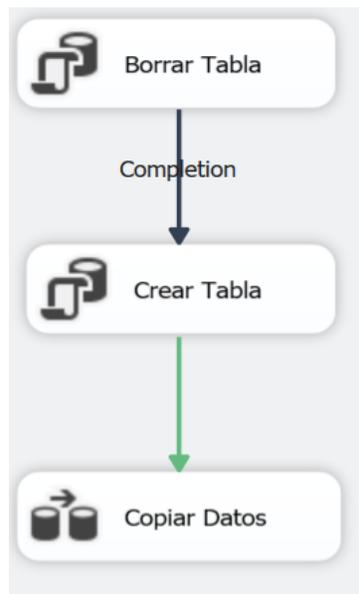
Se borra la tabla en la base de datos STAGE.

Se crea la tabla con la estructura que se definió con antelación de la fuente del sistema.

Se copian los datos que existen desde de la base de datos del cliente a la tabla de la base de datos STAGE.

- Se crean los índices necesarios para la optimización y acceso a las tablas. (Si es necesario)

Figura 31. Ejemplo proceso extracción.



Fuente: Elaboración propia, 2020.

A continuación, se listan todos los paquetes extractores cuya fuente es la Base de datos FRONTERA:

- EXT-SQL-TD_CANAL.dtsx
- EXT-SQL-AX365_DistrGeografica.dtsx
- EXT-SQL-TD_ARQUITECTURA.dtsx
- EXT-SQL-TD_ARTICULO.dtsx
- EXT-SQL-TD_BLOQUE.dtsx
- EXT-SQL-TD_CANTON.dtsx
- EXT-SQL-TD_CLASIFVENDEDOR.dtsx
- EXT-SQL-TD_CLIENTE.dtsx
- EXT-SQL-TD_DISTRITO.dtsx
- EXT-SQL-TD_GRUPO.dtsx
- EXT-SQL-TD_PERFILCLIENTE.dtsx
- EXT-SQL-TD_PROVINCIA.dtsx
- EXT-SQL-TD_RUTA.dtsx
- EXT-SQL-TD_SUBGRUPO.dtsx
- EXT-SQL-TD_SUPERVISOR.dtsx
- EXT-SQL-TD_TIPOCLIENTE.dtsx
- EXT-SQL-TD_VENDEDOR.dtsx
- EXT-SQL-TH_VENTAS.dtsx
- EXT-SQL-TH_VENTASPS.dtsx

- EXT-SQL-TIPOCAMBIO.dtsx
- EXT-SQL-TH_ANTIGUEDAD.dtsx
- EXT-SQL-CALENDARIO.dtsx
- EXT-SQL-TD_ALMACEN.dtsx
- EXT-SQL-INVENTARIO.dtsx
- EXT-SQL-TIPOCAMBIOHISTAX2009.dtsx
- EXT-SQL-TD_SUCURSAL.dtsx
- EXT-SQL-VENTAS_ANUALES.dtsx (Si se desea realizar una recarga histórica, se debe habilitar este paquete en el orquestador de extracciones)
- EXT-SQL-TD_OBJETIVO.dtsx

5.6 Proceso de transformaciones.

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna manipulación de estos.

TRA-TD_XXXX.dtsx (XXXXXX Corresponde al nombre de Dimensión)

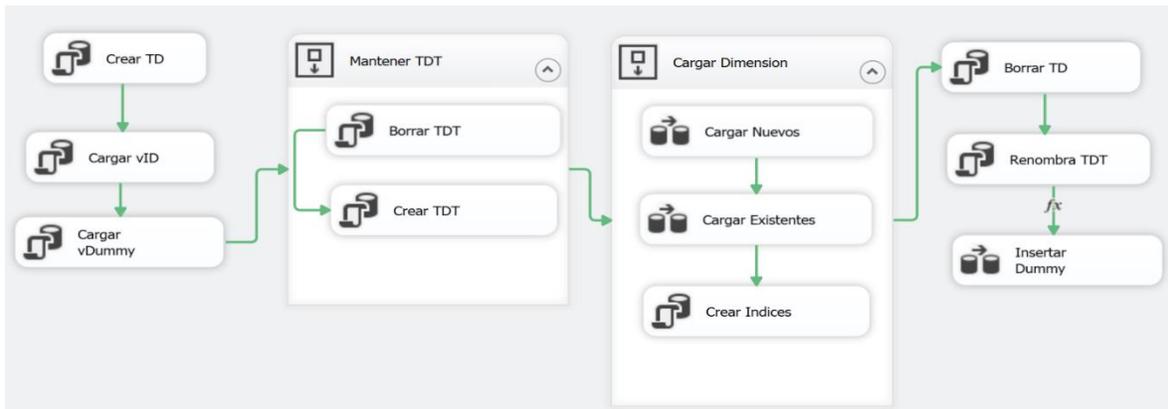
Se realiza la transformación de la dimensión “XXXX” a partir de los datos de la tabla “XXXX” en la base de datos STAGE. La ejecución de este paquete se realiza antes de la transformación de las tablas de hechos.

El proceso que se realiza en general para la carga de las dimensiones es el siguiente:

- Se crea la tabla de la dimensión, en caso de que esta no exista.

- Se obtiene el dato del máximo id de la tabla, a partir de este dato más 1 se ingresan los nuevos registros a la tabla.
- Se verifica si la tabla destino posee el registro base o por defecto (COD = '1, DES_XXX = 'NO DEFINIDO'), con lo que se actualiza el indicador que se usará posteriormente para ingresar este registro (sólo en caso de que no exista). Además, se valida si es necesario hacer un filtro por compañía.
- En el contenedor “*Cargar Dimension*” se realiza el siguiente proceso:
 - Se actualizan aquellos campos susceptibles a modificaciones.
 - Se obtienen los registros nuevos y se insertan en la dimensión.
 - Se crean los índices requeridos.
- Se inserta el registro base en caso de que no exista (valor dummy).

Figura 32. Ejemplo paquete transformación.



Fuente: Diseño propio.2020.

Los paquetes que se listan a continuación poseen el mismo comportamiento antes detallado:

- TRA-TD_CANAL.dtsx
- TRA-TD_SUPERVISOR.dtsx
- TRA-TD_TIPO_CLIENTE.dtsx
- TRA-TD_RUTA.dtsx
- TRA-TD_BLOQUE.dtsx

- TRA-TD_CADENA.dtsx
- TRA-TD_UBICACION.dtsx
- TRA-TD_CLASIFICACION_VENDEDOR.dtsx
- TRA-TD_PERFIL_CLIENTE.dtsx
- TRA-TD_SUBGRUPO.dtsx
- TRA-TD_ARTICULO.dtsx
- TRA-TD_VENDEDOR.dtsx
- TRA-TD_CLIENTE.dtsx
- TRA-TD_ALMACEN.dtsx
- TRA-TD_GRUPO.dtsx

TRA-TH_VENTAS.dtsx

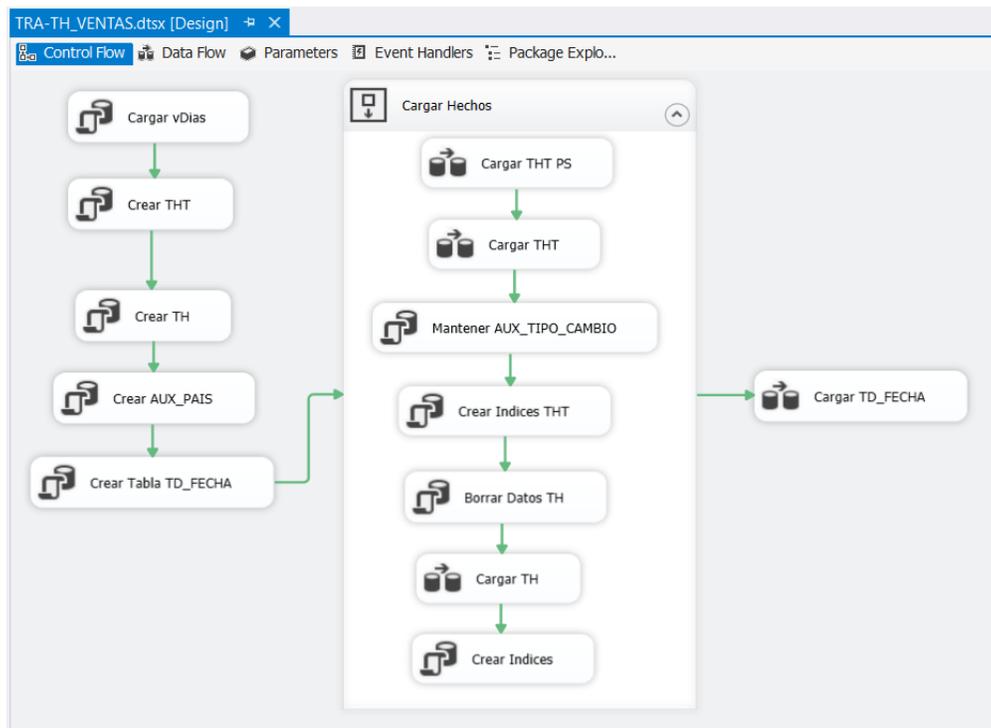
Ejecuta el proceso necesario para llenar la tabla “DWH.dbo.TH_VENTAS”.

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se obtiene la cantidad de días que se van a recargar.
- Se borra (si existe) y se crea la tabla temporal de hechos.
- Se crea la tabla de hechos sino existe.
- Se borra (si existe), se crea y se llena la tabla AUX_PAIS
- Se crea la tabla de dimensión “TD_FECHA” sino existe.
- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla temporal.
- Se crean los índices para la tabla de hechos Temporal.
- Se eliminan de la tabla de hechos aquellos registros que tengan una fecha que coincida con el rango del periodo que se va a cargar, obtenido de la tabla de hechos temporal.

- Se inserta la información de la tabla temporal a la tabla de hechos definitiva
- Se crean, si es que no existen los índices para la tabla de hechos.
- Se realizan la carga de la dimensión “TD_FECHA” con las fechas que se encuentran en la tabla “TH_VENTAS”.

Figura 33. Diseño paquete transformación TH Ventas.



Fuente: Diseño propio, 2020.

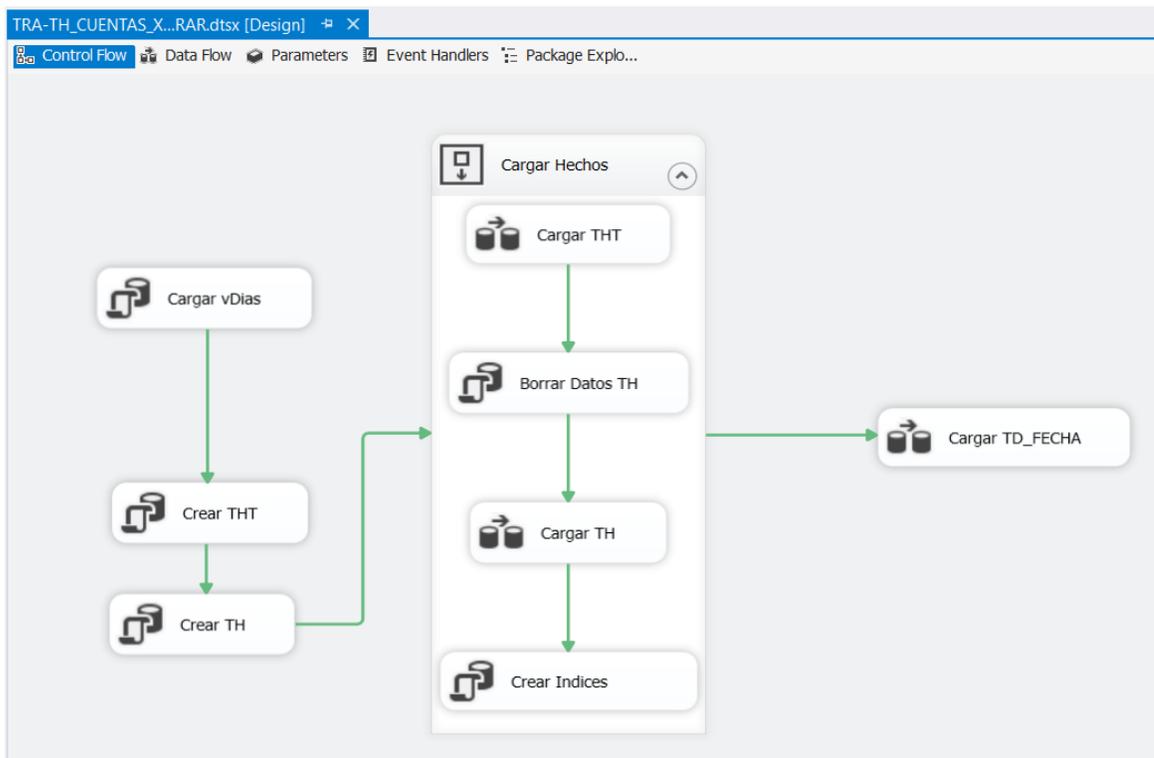
TH_CUENTAS_X_COBRAR.dtsx

Ejecuta el proceso necesario para llenar la tabla “DWH.dbo.TH_CUENTAS_X_COBRAR”.

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se obtiene la cantidad de días que se van a recargar.
- Se borra (si existe) y se crea la tabla temporal de hechos.
- Se crea la tabla de hechos sino existe.
- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla temporal.
- Se eliminan de la tabla de hechos aquellos registros que tengan una fecha que coincida con el rango del periodo que se va a cargar, obtenido de la tabla de hechos temporal.
- Se inserta la información de la tabla temporal a la tabla de hechos definitiva
- Se crean, si es que no existen los índices para la tabla de hechos.
- Se realizan la carga de la dimensión “TD_FECHA” con las fechas que se encuentran en la tabla “TH_CUENTAS_X_COBRAR”.

Figura 34. Diseño paquete transformación TH Cuentas por cobrar



Fuente: Elaboración propia, 2020.

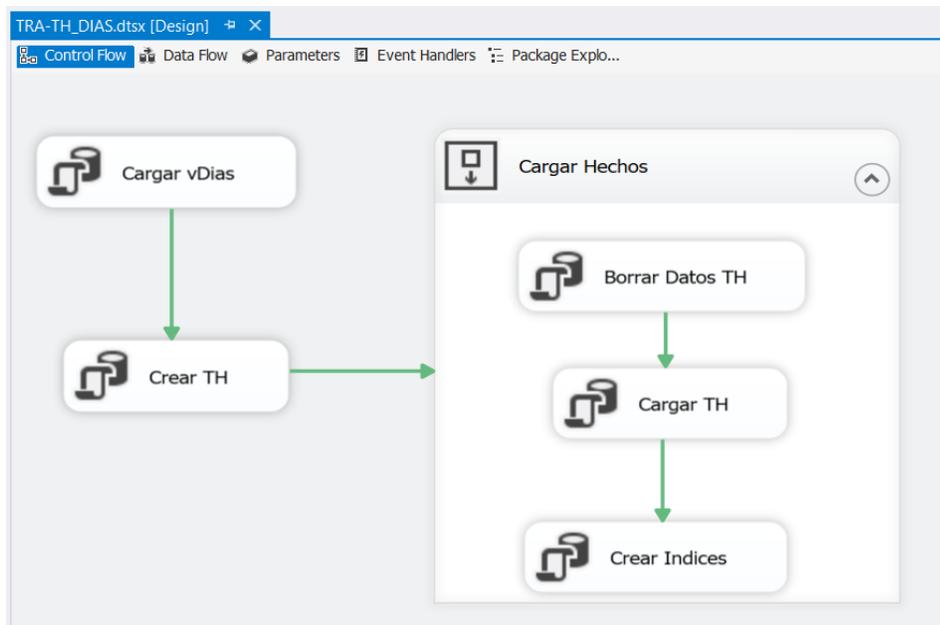
TRA-TH_DIAS.dtsx

Ejecuta el proceso necesario para llenar la tabla “DWH.dbo.TH_DIAS”.

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se obtiene la cantidad de días que se van a recargar.
- Se crea la tabla de hechos sino existe.
- Se eliminan de la tabla de hechos los días a recargar.
- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla de hechos definitiva.
- Se crean, si es que no existen los índices para la tabla de hechos.

Figura 35. Diseño paquete transformación TH Días.



Fuente: Elaboración propia, 2020.

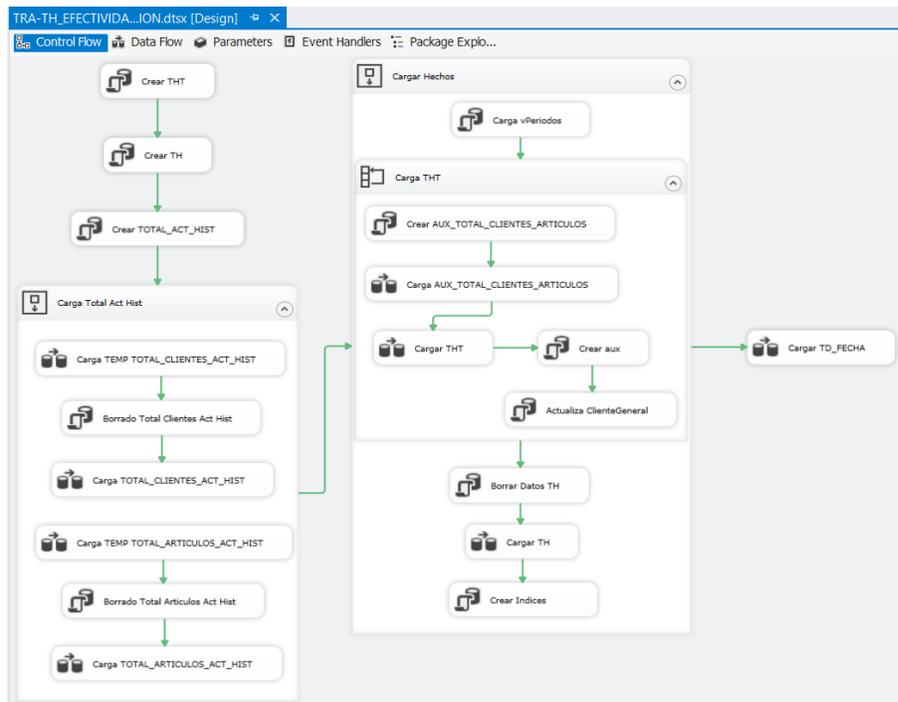
TRA-TH_EFECTIVIDAD_PENETRACION.dtsx

Ejecuta el proceso necesario para llenar la tabla “DHW.dbo.TH_ PENETRACION”

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se borra (si existe) y se crea la tabla temporal de hechos.
- Se crea la tabla de hechos sino existe.
- Se crean tablas auxiliares necesarias para la carga de información.
- Se realiza la carga de las tablas auxiliares.
- Se obtienen las fechas que se van a recargar.
- Se crea y se cargan tablas auxiliares de clientes y artículos basados en las fechas de carga.
- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla temporal.
- Se actualiza la columna indicadora de compra de la tabla de hechos temporal.
- Se eliminan de la tabla de hechos aquellos registros que tengan una fecha que coincida con el rango del periodo que se va a cargar, obtenido de la tabla de hechos temporal.
- Se inserta la información de la tabla temporal a la tabla de hechos definitiva
- Se crean, si es que no existen los índices para la tabla de hechos.
- Se realizan la carga de la dimensión “TD_FECHA” con las fechas que se encuentran en la tabla “TH_EFECTIVIDAD_PENETRACION”.

Figura 36. Diseño paquete transformación TH efectividad y penetración.



Fuente: Elaboración propia, 2020.

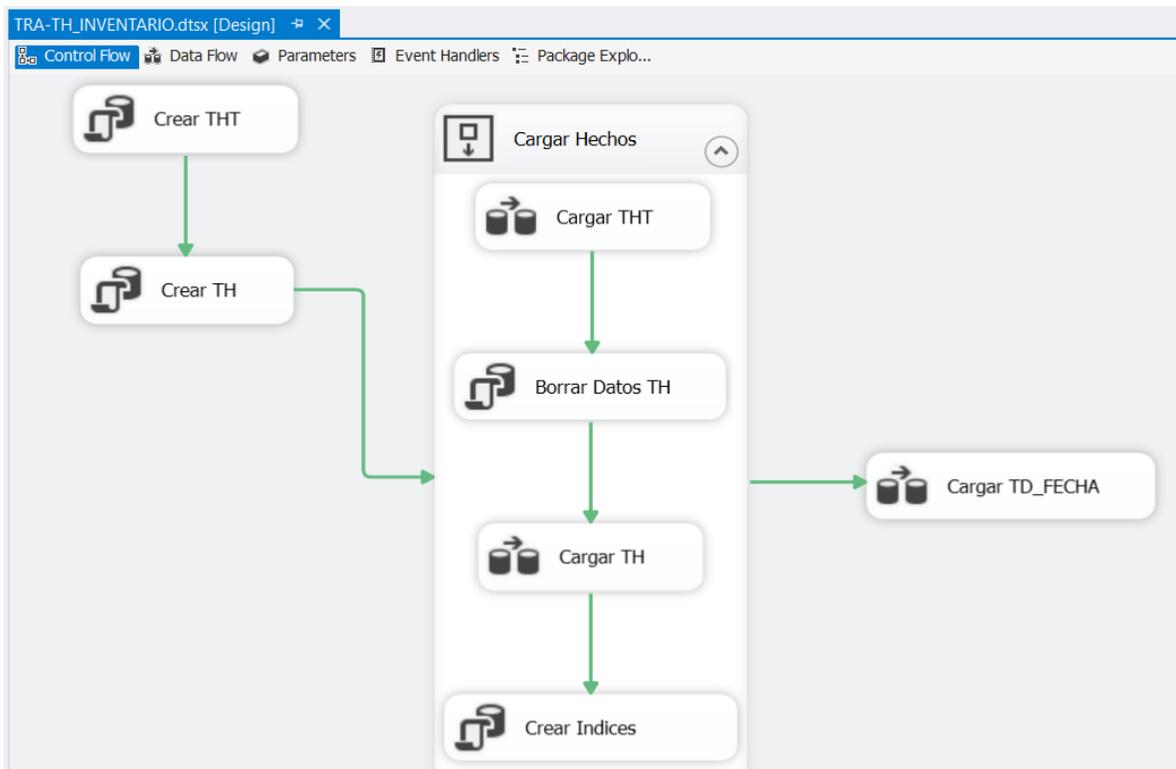
TH_INVENTARIO.dtsx

Ejecuta el proceso necesario para llenar la tabla “DWH.dbo. TH_INVENTARIO”.

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se borra (si existe) y se crea la tabla temporal de hechos.
- Se crea la tabla de hechos sino existe.
- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla temporal.
- Se eliminan de la tabla de hechos aquellos registros que tengan una fecha que coincida con el rango del periodo que se va a cargar, obtenido de la tabla de hechos temporal.
- Se inserta la información de la tabla temporal a la tabla de hechos definitiva
- Se crean, si es que no existen los índices para la tabla de hechos.
- Se realizan la carga de la dimensión “TD_FECHA” con las fechas que se encuentran en la tabla “TH_INVENTARIO”.

Figura 37. Diseño paquete transformación TH inventario.



Fuente: Elaboración propia, 2020.

TH_VENTAS_MES_ANIO_ANT.dtsx

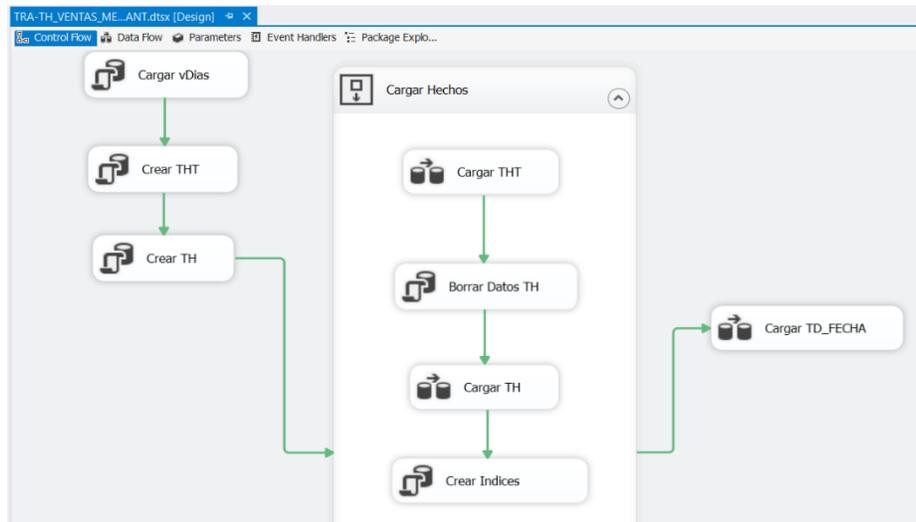
Ejecuta el proceso requerido para llenar la tabla “DWH.dbo.TH_VENTAS_MES_ANIO_ANT”

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se obtiene la cantidad de días que se van a recargar.
- Se borra (si existe) y se crea la tabla temporal de hechos.
- Se crea la tabla de hechos sino existe.
- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla temporal.

- Se eliminan de la tabla de hechos aquellos registros que tengan una fecha que coincida con el rango del periodo que se va a cargar, obtenido de la tabla de hechos temporal.
- Se inserta la información de la tabla temporal a la tabla de hechos definitiva
- Se crean, si es que no existen los índices para la tabla de hechos.
- Se realizan la carga de la dimensión “TD_FECHA” con las fechas que se encuentran en la tabla “TH_VENTAS_MES_ANIO_ANT”.

Figura 38. Diseño paquete transformación TH Ventas mes anterior.



Fuente: Elaboración propia, 2020.

TH_VENTAS_MES_ANIO.dtsx

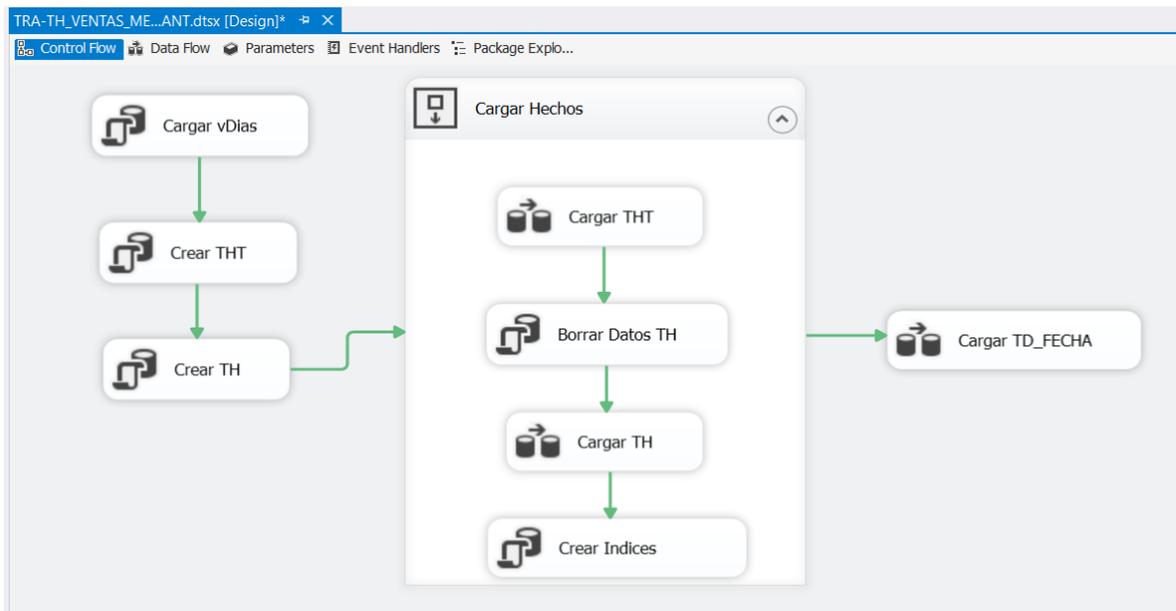
Ejecuta el proceso necesario para llenar la tabla “DWH.dbo.TH_VENTAS_MES_ANIO”.

El proceso que se realiza para la carga de la tabla de hechos es el siguiente:

- Se obtiene la cantidad de días que se van a recargar.
- Se borra (si existe) y se crea la tabla temporal de hechos.
- Se crea la tabla de hechos sino existe.

- Se realizan las transformaciones de los nuevos registros, estos se insertan en la tabla temporal.
- Se eliminan de la tabla de hechos aquellos registros que tengan una fecha que coincida con el rango del periodo que se va a cargar, obtenido de la tabla de hechos temporal.
- Se inserta la información de la tabla temporal a la tabla de hechos definitiva
- Se crean, si es que no existen los índices para la tabla de hechos.
- Se realizan la carga de la dimensión “TD_FECHA” con las fechas que se encuentran en la tabla “TH_VENTAS_MES_ANT”.

Figura 39. Diseño paquete transformación TH Ventas mes año anterior.

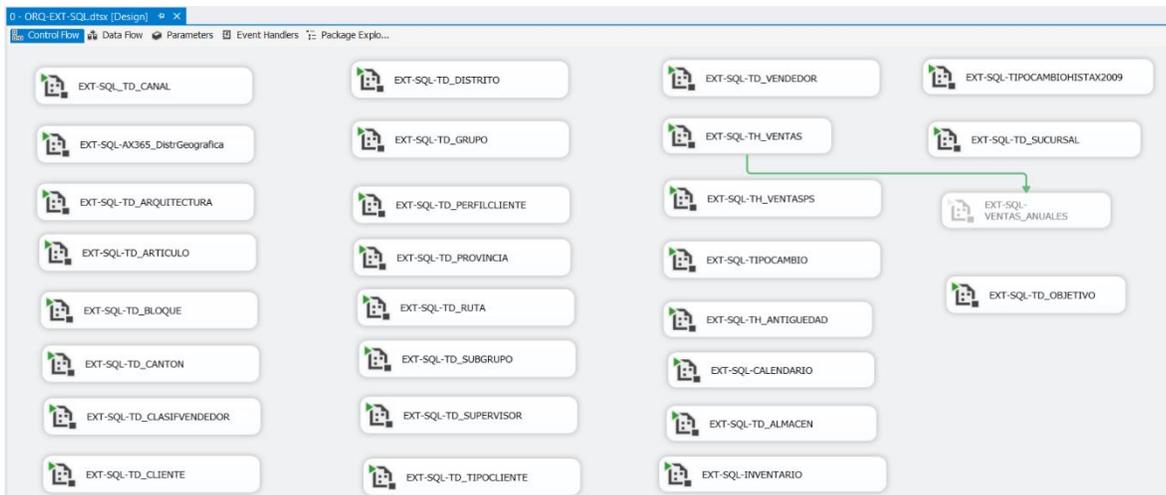


Fuente: Elaboración propia, 2020.

ORQ-EXT.dtsx

Descripción: Realiza el llamado a todos los paquetes de extracción, cuya fuente de datos es la base de datos FRONTERA.

Figura 40. Diseño paquete orquestador extracciones.

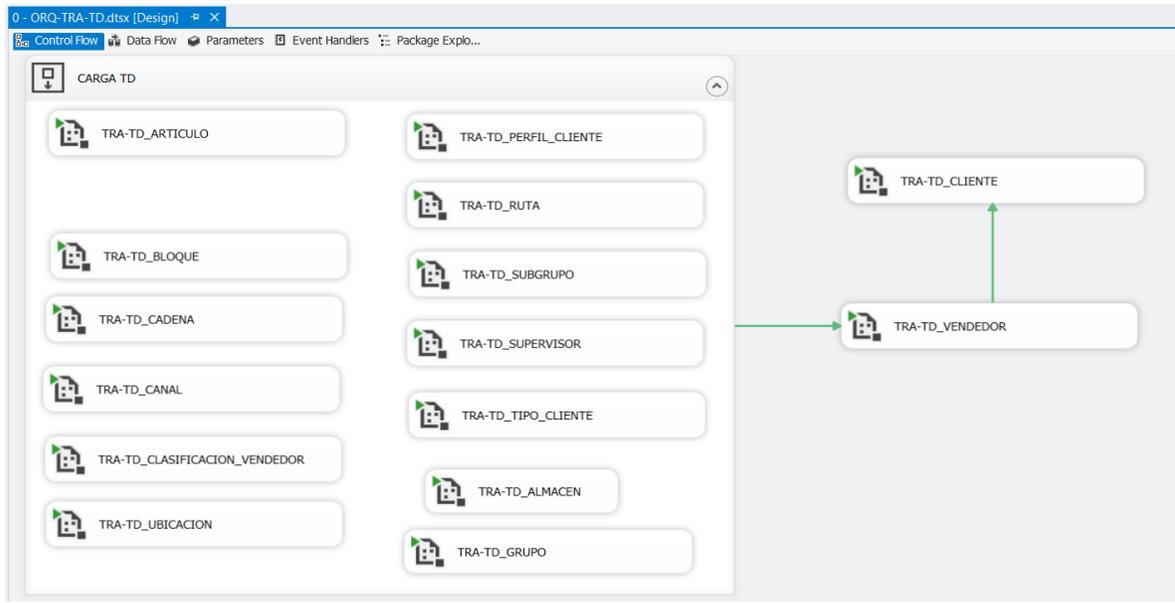


Fuente: Elaboración propia, 2020.

ORQ-TRA-TD.dtsx

Descripción: Es el paquete que se ejecutará para realizar el proceso de carga de todas las dimensiones del proyecto.

Figura 41. Diseño paquete orquestador transformaciones.

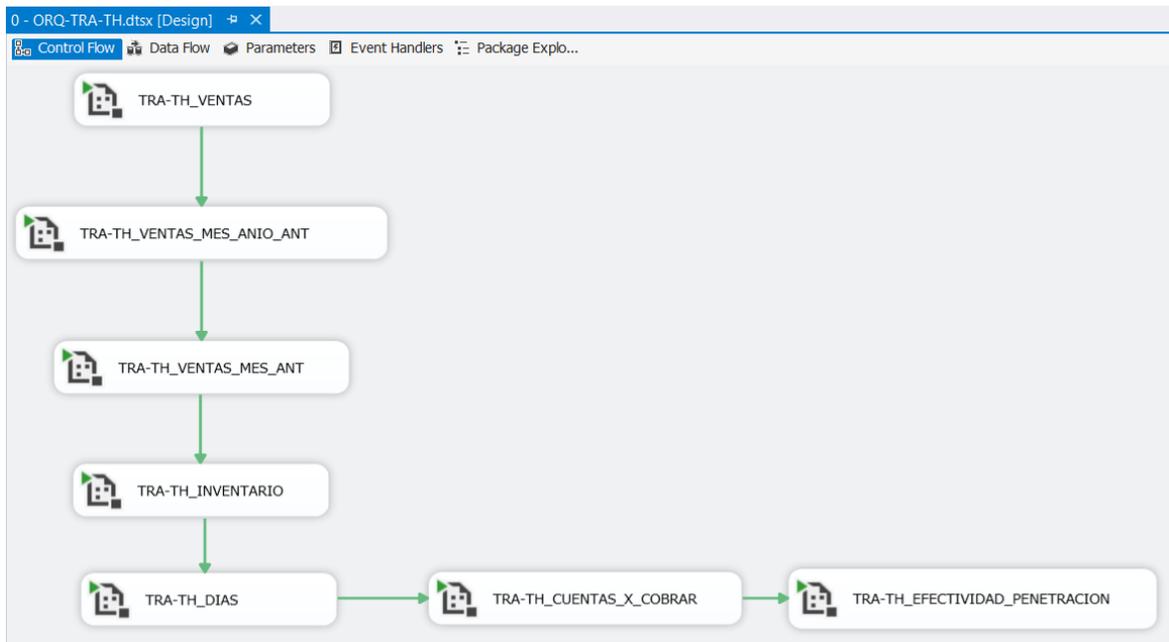


Fuente: Elaboración propia, 2020.

ORQ-TRA-TH.dtsx

Descripción: Es el paquete que se ejecutará para realizar el proceso de carga de todas las tablas de hechos del proyecto.

Figura 42. Diseño paquete orquestador transformación TH.

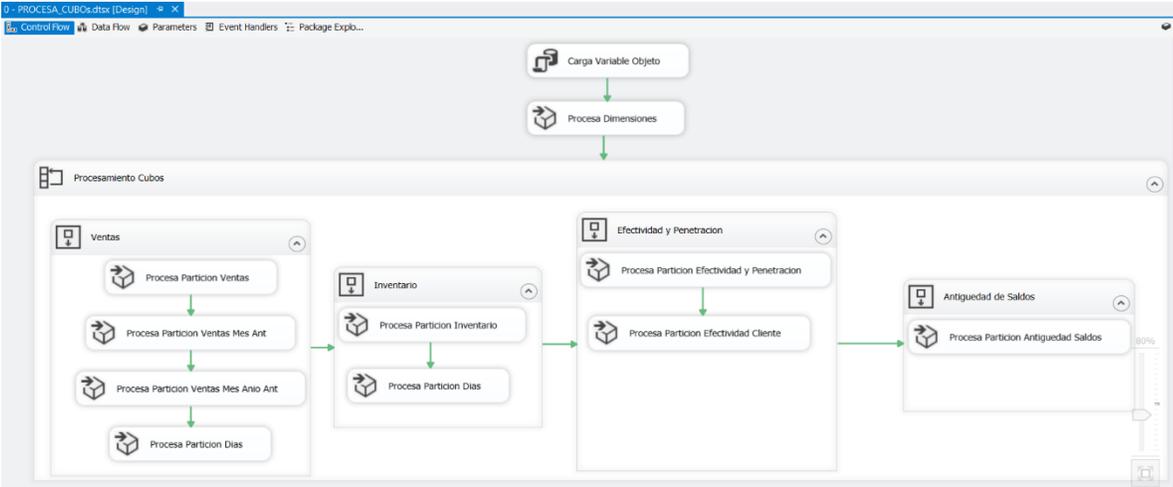


Fuente: Elaboración propia, 2020.

PROCESAR_CUBOs.dtsx

Descripción: Se realizan los procesos de actualización de los elementos que forman parte de los cubos.

Figura 43. Diseño paquete procesar cubo.

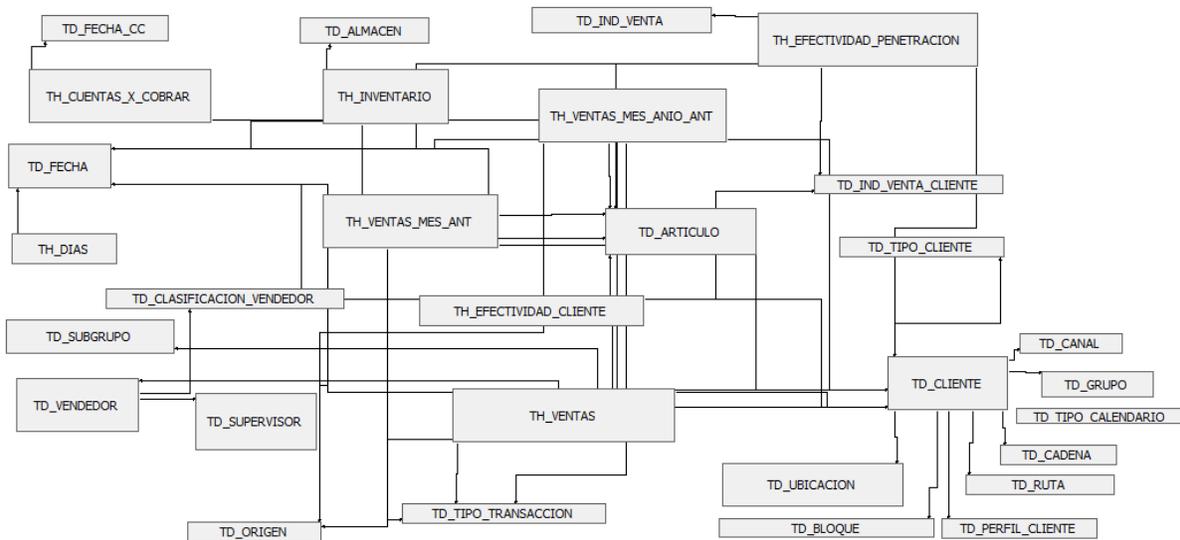


Fuente: Elaboración propia, 2020

5.7 Base de datos multidimensionales (Cubos)

Diagrama del cubo.

Figura 44. Diagrama estructura del cubo.



Fuente: Elaboración propia, 2020.

En este apartado del documento técnico se detallarán las partes que conforman los cubos Multidimensionales del modelo de ventas, estas son dos:

- Dimensiones
- Medidas

5.7.1 Dimensiones

Las dimensiones son las estructuras que nos permiten explotar la información del cubo y realizar los cruces de información que deseamos siempre y cuando la dimensión tenga relación con la medida o el dato cuantitativo que se va a trabajar. A continuación, se identifican las dimensiones del modelo.

- TD_ALMACEN
- TD_ARTICULO
- TD_BLOQUE
- TD_CADENA
- TD_CANAL
- TD_CLASIFICACION_VENDEDOR
- TD_CLIENTE
- TD_FECHA
- TD_GRUPO
- TD_ORIGEN
- TD_PERFIL_CLIENTE
- TD_RUTA
- TD_SUBGRUPO
- TD_SUPERVISOR
- TD_TIPO_CLIENTE
- TD_TIPO_TRANSACCION
- TD_UBICACION
- TD_VENDEDOR

Detalle de los cubos y las dimensiones con las que se relacionan.

Cubo	Ventas
Dimensiones con las que se relaciona	Origen de la transacción
	Cliente
	Articulo
	Tipo Transacción
	Bloques
	Cadena
	Canal
	Ruta
	Ubicación Geográfica
	Fecha
	Perfil Cliente
	SubGrupo
	Vendedor
	Tipo de Cliente
Clasificación del Vendedor	
Supervisor	

Cubo	Inventarios
Dimensiones con las que se relaciona	Fecha
	Articulo
	Almacén
	Tipo de Calendario

Cubo	Efectividad y Penetración
Dimensiones con las que se relaciona	Indicador de Venta Cliente
	Cliente
	Articulo
	Bloques
	Cadena
	Canal
	Ruta
	Ubicación Geográfica
	Fecha
	Perfil Cliente
	SubGrupo
	Vendedor
	Tipo de Cliente
	Clasificación del Vendedor
	Supervisor

Cubo	Antigüedad de Saldos
Dimensiones con las que se relaciona	Cliente
	Bloques
	Cadena
	Canal
	Ruta
	Ubicación Geográfica
	Perfil Cliente
	SubGrupo
	Vendedor
	Tipo de Cliente
	Clasificación del Vendedor
	Supervisor
	Fecha CC

5.8 Medidas calculadas.

Las medidas son la parte cuantitativa del cubo y estas se agrupan en grupos de medidas dependiendo de la tabla de hechos a la que pertenezcan y cálculo que se realice. En seguida se detallarán los grupos de medidas y las medidas existentes en el cubo.

Cubo	Ventas
Grupo de medidas	Ventas
	% Bonificaciones
	% Buen Estado
	% Cajas vs AA
	% Cajas vs MA
	% Crecimientovs Año Ant
	% Crecimientovs Mes Ant
	% Descuento
	% Días Facturados
	% Mal Estado
	% Proyección Venta Neta Cierre Mes vs AA
	% Unidades vs AA
	% Unidades vs AA
	Cajas Crecimiento vs Año Ant
	Cajas Crecimiento vs Mes Ant
	Cajas Vendidas
	Cajas Vendidas AA
	Cajas Vendidas MA
	Comisión
	Comisión Real
	Comisión Real USD
	Comisión USD
	Contador Clientes
	Crecimiento vs Año Ant
	Crecimiento vs Año Ant USD
	Crecimiento vs Mes Ant
	Crecimiento vs Mes Ant USD
	Cumplimiento
	Descuento

Descuento USD
Días del Mes
Días Facturados
Días Pendiente
Impuesto
Impuesto USD
Monto Bruto Bonificado
Monto Bruto Buen Estado
Monto Bruto Mal Estado
Proyección Venta Neta Cierre Mes
Proyección Venta Neta Cierre Mes USD
Unidades Bonificadas
Unidades Buen Estado
Unidades Crecimiento vs Año Ant
Unidades Crecimiento vs Mes Ant
Unidades Mal Estado
Unidades Producto
Unidades Vendidas
Unidades Vendidas AA
Unidades Vendidas MA
Venta Bruta
Venta Bruta AA
Venta Bruta AA USD
Venta Bruta MA
Venta Bruta MA USD
Venta Bruta Producto
Venta Bruta USD
Venta Neta
Venta Neta AA
Venta Neta AA USD
Venta Neta MA
Venta Neta MA USD
Venta Neta Meta
Venta Neta Meta UDS
Venta Neta USD

Cubo	Inventarios
Grupo de medidas	Inventarios
	Cajas Vendidas del Mes
	Cajas Vendidas del Mes CEDI
	Cierre Proyectado
	Cierre Proyectado CEDI
	Días del Mes
	Días Facturados
	Días Inventario
	Días Inventario CEDI
	Días Pendientes
	Inventario Físico
	Mese Inventario
	Mese Inventario CEDI
	Promedio Diario
	Promedio Diario CEDI

Cubo	Antigüedad de saldos
Grupo de medidas	Cuentas por Cobrar
	% Por Vencer
	% Vencido
	% Vencido 15
	% Vencido 22
	% Vencido 30
	% Vencido 30 +
	% Vencido 8
	Saldo Actual
	Saldo por Vencer
	Saldo Vencido
	Saldo Vencido 15
	Saldo Vencido 22
	Saldo Vencido 30
	Saldo Vencido 30 Mas
	Saldo Vencido 8

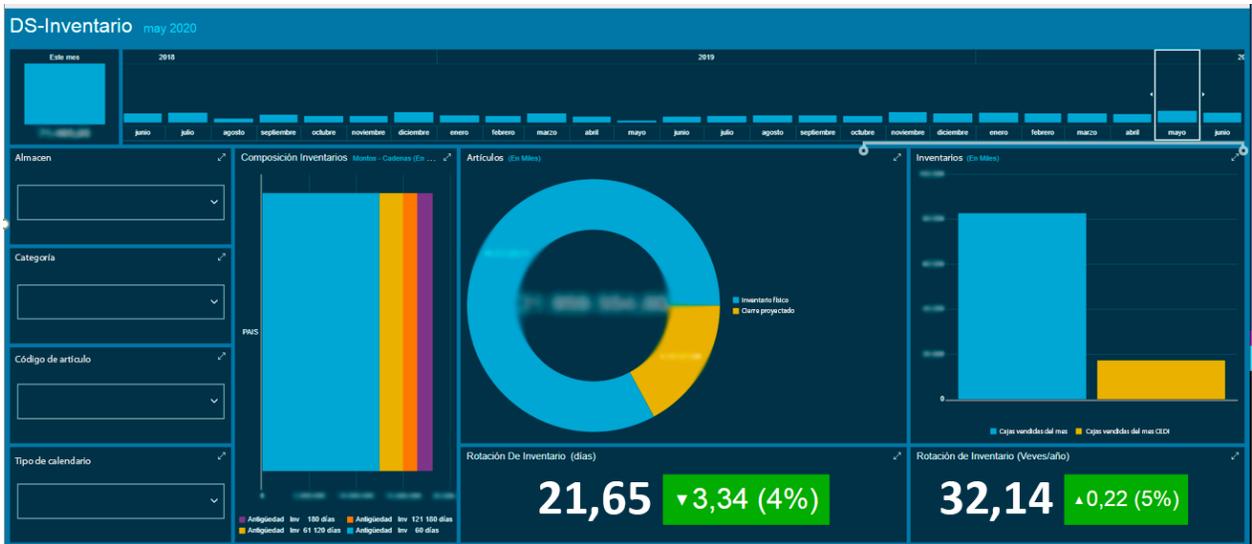
Cubo	Efectividad y penetración
Grupo de medidas	Efectividad y penetración
	% Presencia de Clientes
	Cant Total de Clientes
	Cantidad de Clientes sin Venta
	Cantidad Total de Clientes
	Efectividad de Clientes

5.9 Dashboards

A continuación, se detallan los Dashboards creados con las tablas y procesos mencionados anteriormente.

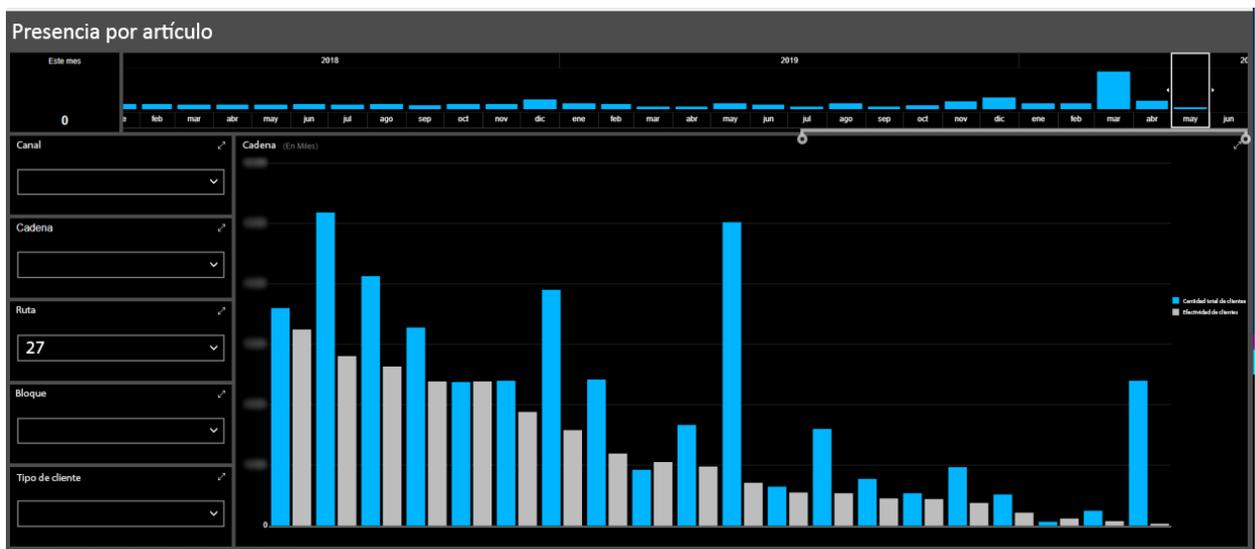
5.9.1 Dashboard 01 – Inventario

Detalla la cantidad de productos en inventario, la cantidad de cajas vendidas de ese producto en el periodo, la cantidad de cajas CEDI, un indicador de la rotación del inventario, la composición de los inventarios según la antigüedad. Permite filtrar por periodos, almacén, categoría, código del artículo y el tipo de calendario.



5.9.2 Dashboard 02 – Presencia por artículo

Corresponde a un Dashboard que permite mostrar la cantidad de clientes que tienen determinado artículo, la efectividad. Permite aplicar los filtros de periodo, canal, cadena, ruta, bloque, tipo de cliente.



5.9.3 Dashboard 03 - Clientes

Corresponde a un reporte detallado de los clientes de la compañía, muestra el código de cliente, la razón social, la dirección, cédula, teléfono y un contado de cliente.

Permite aplicar diversos filtros como bloque, el código de la ruta, el canal, tipo de cliente, el estado del cliente y código de cliente.

Alimentos Jack's Interamericana

Home > Informes Generales > COSTA RICA > Reporte Clientes

File View Edit in Power BI Desktop Export Refresh

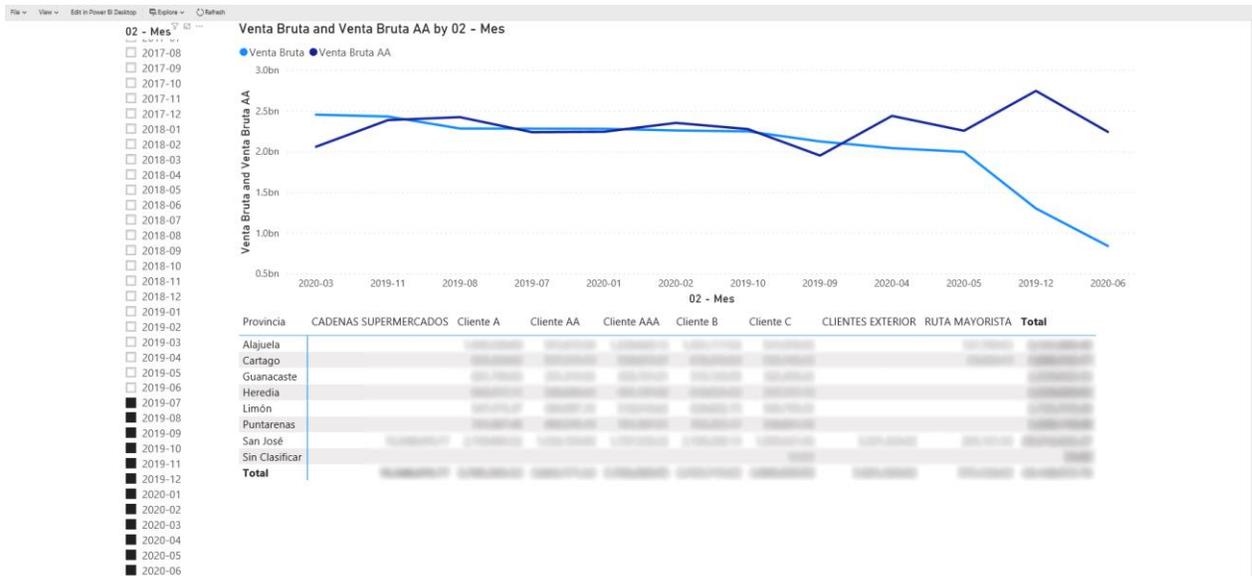
Bloque: All Código Ruta: All Canal: All Tipo de Cliente: All Código Cliente: [dropdown]

Indicador Inactivo: Activo [dropdown]

Código Cliente	Razón Social	Dirección	Cédula	Teléfono	Contador Clientes
89016		ZONA FRANCA METROPOLITANA BARREAL HEREDIA			1
9088		ZETILLAL 200N 200E DEL INA, IPIIS			1
110796		ZARCERO. LAS PALMITAS			1
108901		ZARCERO,CONTIGUO AL CEMENTERIO			1
102376		ZARCERO, ZAPOTE COSTADO O. PLAZA SAN JUAN DE LAJAS			1
102612		ZARCERO, TRAMOS LA PALMITA FTE A LA PARADA			1
118524		ZARCERO, TAPESCO DETRAS DE IGLESIA			1
111413		ZARCERO, PALMITA CENTRO			1
117405		ZARCERO, PALMIRA FTE A SALON COMUNAL			1
117938		ZARCERO, PALMIRA			1
101589		ZARCERO, LOS TRAMOS DE LAS PALMITAS			1
119415		ZARCERO, LAS BRISAS LOS ANGELES 20 E. ESCUELA			1
100647		ZARCERO, LAGUNA FTE A LA FARMACIA ROMI			1
118014		ZARCERO, LAGUNA			1
121479		ZARCERO, LA PALMITA			1
124186		ZARCERO, FRENTE AL JUZGADO			1
104250		ZARCERO, COSTADO S. DE PLAZA DEPORTES			1
118012		ZARCERO, 300 N. DE AGRISAL			1
104841		ZARCERO, 2 KMS S. DEL PARQUE			1
90057		ZARCERO, 100 N. 100 O. DE ERMITA B* SAN ANTONIO			1
104922		ZARCERO, 1.5 KM O. DEL PARQUE			1
110533		ZARCERO, 1 1/2 KM DEL COLEGIO			1
118821		ZARCERO EL BAJO			1
125496		ZARCERO DEL PARQUE 75 SUR			1
Total					15,193

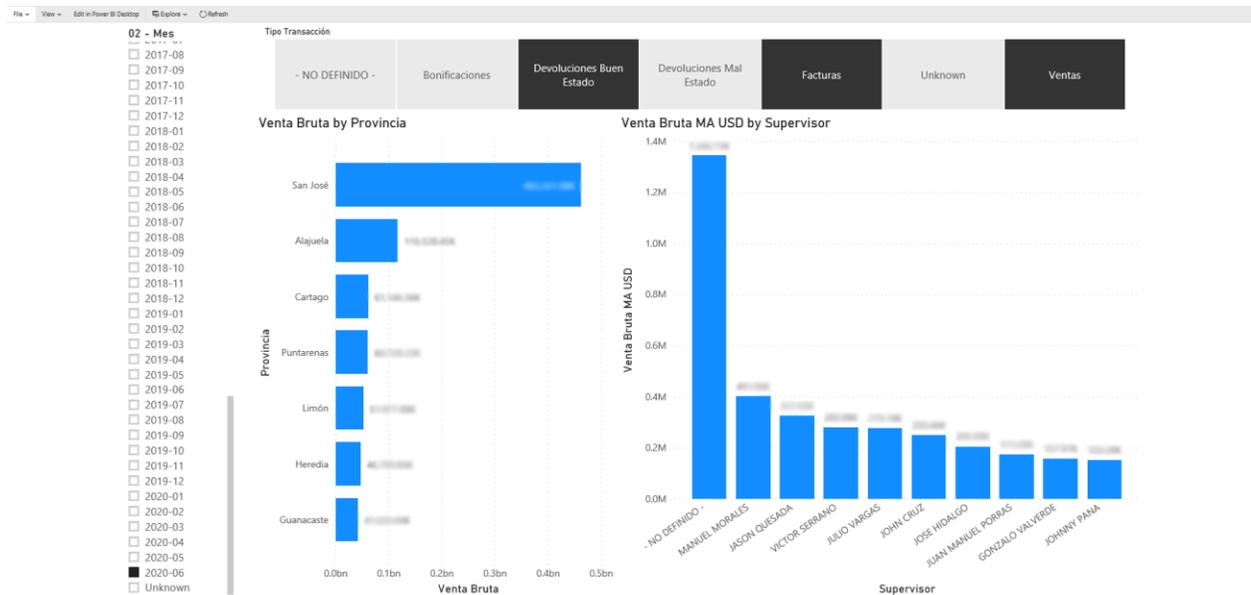
5.9.4 Dashboard 04 – Venta bruta vs venta Bruta Año anterior

Muestra la comparación de las ventas brutas del año contra las ventas brutas del año anterior.



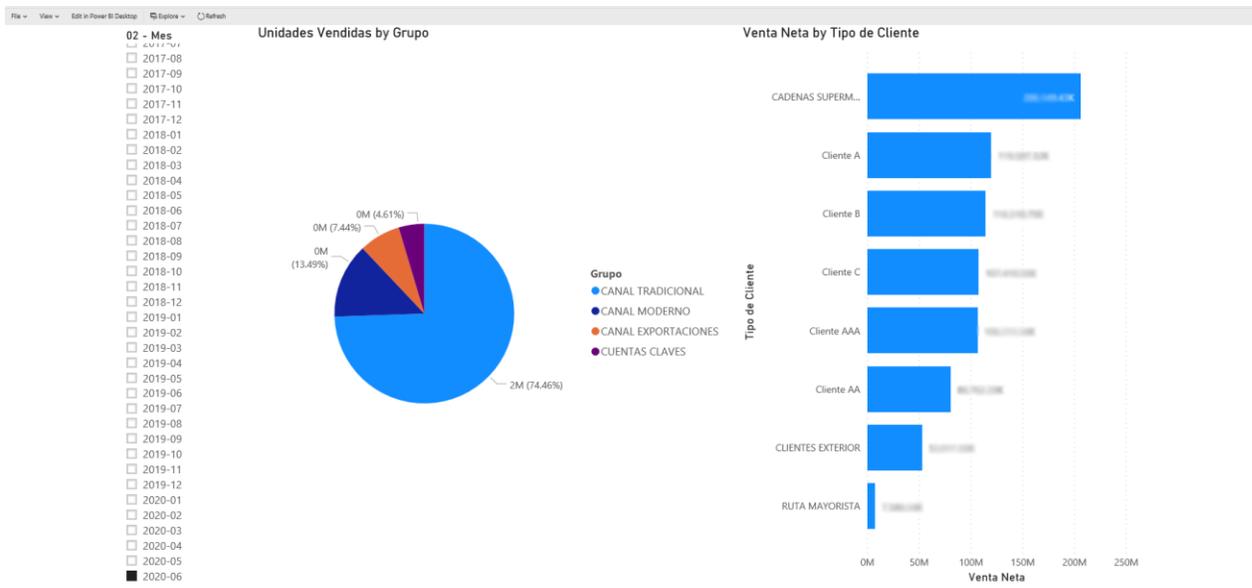
5.9.5 Dashboard 05 – Venta bruta por provincia

Detalla las ventas brutas por provincia y la venta bruta del mes anterior por supervisor.



5.9.6 Dashboard 06 – Unidades vendidas por grupo

Muestra las unidades vendidas por grupo de cliente en un grafico de pastel y la venta neta por tipo de cliente.



5.9.7 Dashboard 07 - Efectividad Neta

Corresponde a un reporte detallado de la efectividad de los vendedores de la compañía, muestra el subgrupo, código ruta, nombre del vendedor, cantidad de clientes, efectividad de clientes y el porcentaje de presencia.

Permite mostrar la información de acuerdo con los siguientes filtros, periodo, indicador de activo que corresponde al estado del artículo en la base de datos, en canal de venta, el bloque de venta, el tipo de cliente que corresponde a una categorización del cliente, el código de la ruta y el subgrupo al que pertenece.

Alimentos Jack's Interamericana

Home Efectividad y Penetración Efectividad Venta

File View Edit in Power BI Desktop Explore Refresh

02 - Mes: 2020-03

Indicador Activo: Artículo Activo

Canal: All

Bloque: All

Tipo de Cliente: All

Código Ruta: All

SubGrupo: All

SubGrupo	Código Ruta	Vendedor	Cantidad Total de Clientes	Efectividad de Clientes	% Presencia de Clientes
- NO DEFINIDO -	-1	- NO DEFINIDO -	2	1	50.00 %
- NO DEFINIDO -	999	- NO DEFINIDO -	2		
AM PM	997	Administracion central	102	88	86.27 %
AUTO MERCADO	-1	- NO DEFINIDO -	1	1	100.00 %
CURAZAO	928	- NO DEFINIDO -	1		
EL SALVADOR	921	- NO DEFINIDO -	1		
EL SALVADOR	946	- NO DEFINIDO -	1		
GESEA	-1	- NO DEFINIDO -	1	1	100.00 %
GONZALO VALVERDE	68	ALLAN ARIAS MARTINEZ	239	217	90.79 %
GONZALO VALVERDE	61	ANDREY MURILLO ARAYA	169	154	91.12 %
GONZALO VALVERDE	94	JOHAN ALPIZAR RODRIGUEZ	240	202	84.17 %
GONZALO VALVERDE	63	LORENZO PEREZ PEÑARANDA	225	190	84.44 %
GONZALO VALVERDE	62	MAURICIO GALARZA VALLADARES	221	199	90.05 %
GONZALO VALVERDE	30	YEXON ROJAS ROJAS	210	198	94.29 %
GUATEMALA	945	- NO DEFINIDO -	1		
HONDURAS	947	- NO DEFINIDO -	1		
JASON QUESADA	93	AARON RIVERA ZUÑIGA	321	311	96.88 %
JASON QUESADA	51	ADAN PICADO RIVERA	232	222	95.69 %
JASON QUESADA	39	DANILO MATA GONZALEZ	261	257	98.47 %
JASON QUESADA	38	JACOB GUZMAN LEITON	282	271	96.10 %
JASON QUESADA	45	JONATHAN MENDEZ ZUÑIGA	334	305	91.32 %
JASON QUESADA	36	JUAN CARLOS VILLALBA	273	263	96.34 %
Total			16,481	15,472	93.88 %

5.9.8 Dashboard 08 - Presencia por artículo tabla

Corresponde a un reporte que permite mostrar la cantidad de clientes que tienen determinado artículo, la efectividad, el porcentaje de clientes, la ruta, el supervisor y el vendedor. Permite aplicar los filtros de periodo, bloque, canal, tipo de cliente, indicador objetivo, código del artículo, supervisor, ruta y cadena.

Alimentos Jack's Interamericana

Home Efectividad y Penetración Presencia por Artículo

02 - Mes: 2020-05

Bloque: All

Canal: All

Tipo de Cliente: All

Indicador Objetivo: OBJETIVO

Código de Artículo: All

Supervisor: All

Ruta: All

Cadena: All

Supervisor	Código Ruta	Vendedor	Artículo	Cantidad Total de Clientes	Efectividad de Clientes	% Presencia de Clientes
VICTOR SERRANO	27		C.GALLETA COMP ARANDANO 25grs	285	4	1.40 %
VICTOR SERRANO	27		C.GALLETA COMP CHOCOLATE 25g	285	3	1.05 %
VICTOR SERRANO	27		C.PICARITA JALAPEÑA INTERM 30grs	285	35	12.28 %
VICTOR SERRANO	27		D.PICARITA JALAPEÑA MED 75grs	285	17	5.96 %
VICTOR SERRANO	27		E.GALLETA COMP ARAND 8 PACK 200grs	285		
VICTOR SERRANO	27		R.BOB CAFE 225GRS 20UD	285	2	0.70 %
VICTOR SERRANO	27		R.BONOBON CORONADO 270grs	285	18	6.32 %
VICTOR SERRANO	27		R.NIKOLO FRUTILLA 21grs 480ud	285	4	1.40 %
VICTOR SERRANO	31		C.GALLETA COMP ARANDANO 25grs	310	9	2.90 %
VICTOR SERRANO	31		C.GALLETA COMP CHOCOLATE 25g	310	12	3.87 %
VICTOR SERRANO	31		C.PICARITA JALAPEÑA INTERM 30grs	310	100	32.26 %
VICTOR SERRANO	31		D.PICARITA JALAPEÑA MED 75grs	310	50	16.13 %
VICTOR SERRANO	31		E.GALLETA COMP ARAND 8 PACK 200grs	310	1	0.32 %
VICTOR SERRANO	31		R.BOB CAFE 225GRS 20UD	310	2	0.65 %
VICTOR SERRANO	31		R.BONOBON CORONADO 270grs	310	46	14.84 %
VICTOR SERRANO	31		R.NIKOLO FRUTILLA 21grs 480ud	310	7	2.26 %
VICTOR SERRANO	35		C.GALLETA COMP ARANDANO 25grs	312	3	0.96 %
VICTOR SERRANO	35		C.GALLETA COMP CHOCOLATE 25g	312	28	8.97 %
VICTOR SERRANO	35		C.PICARITA JALAPEÑA INTERM 30grs	312	34	10.90 %
VICTOR SERRANO	35		D.PICARITA JALAPEÑA MED 75grs	312	5	1.60 %
VICTOR SERRANO	35		E.GALLETA COMP ARAND 8 PACK 200grs	312	27	8.65 %
VICTOR SERRANO	35		R.BOB CAFE 225GRS 20UD	312	5	1.60 %
VICTOR SERRANO	35		R.BONOBON CORONADO 270grs	312	33	10.58 %
VICTOR SERRANO	35		R.NIKOLO FRUTILLA 21grs 480ud	312	2	0.64 %
Total				16,678	5,869	35.19 %

5.9.9 Dashboard 09 - Antigüedad de saldos

Detalla el saldo actual de cliente, que monto está por vencer, cual está vencido, el porcentaje vencido a 8,15 días. Permite filtrar por periodo, bloque, código del cliente, cadena, subgrupo, código de la ruta, perfil del cliente, canal y día específico.

Alimentos Jack's Interamericana

Home > Informes Cuentas por Cobrar > Antigüedad de Saldos

02 - Mes: 2020-05 Bloque: AII Código Cliente: AII Cadena: AII SubGrupo: AII

03 - Día: 2020-05-31 Código Ruta: AII Perfil Cliente: AII Canal: AII

Código Cliente	Razon Social	Saldo Actual	Saldo por Vencer	Saldo Vencido	% Vencido	% Vencido 8	% Vencido 15
123535							100.00 %
553							100.00 %
71126							100.00 %
139549							100.00 %
548							100.00 %
97191							100.00 %
136435							100.00 %
108815							100.00 %
138446							100.00 %
149829							100.00 %
141041							100.00 %
72225							62.91 %
131498							-4.87 %
126223							55.19 %
137832							46.84 %
133113							53.16 %
61207							51.76 %
150249							40.39 %
116225							59.87 %
110920							36.68 %
111874							35.74 %
146826							64.99 %
139059							35.01 %
Total							51.34 %
							32.97 %
							7.62 %
							28.23 %
							51.71 %
							28.12 %
							-0.70 %
							28.08 %
							46.20 %
							24.88 %
							11.48 %
							1.42 %

5.9.10 Dashboard 10- Clientes sin venta

Muestra los clientes que en determinado periodo no tienen compras hacia la compañía, así como el tipo de cliente que corresponde, dirección y razon social.

Alimentos Jack's Interamericana

Home > Efectividad y Penetración > Clientes sin venta

01 - Año: 2020 | 02 - Mes: 2020-03 | Bloque: All | SubGrupo: All | Ruta: All

Código Cliente	Razón Social	Dirección	Tipo de Cliente	Cantidad de Clientes sin Ven
100			Cliente C	
100092			Cliente C	
1001			Cliente C	
100707			Cliente B	
100728			Cliente C	
100865			Cliente C	
101082			Cliente C	
101261			Cliente C	
101278			Cliente A	
101694			Cliente C	
101773			Cliente C	
101786			Cliente B	
102047			Cliente C	
102189			Cliente C	
102490			Cliente C	
102525			Cliente B	
102710			Cliente A	
103016			Cliente C	
103057			Cliente C	
103223			Cliente C	
103270			Cliente C	
103301			Cliente C	
Total				1,15

Indicador de Venta Cliente: Sin Venta por Cliente

5.9.11 Dashboard 11 - Comparativo ventas por vendedor.

Este reporte detalla la comparación de la venta neta del cliente respecto al mismo periodo del año anterior y mostrar ya sea su crecimiento o decrecimiento.

Alimentos Jack's Interamericana

Home > Informes Canal Tradicional > Comparativo ventas por vendedor

02 - Mes: 2020-04 | Canal: All | Cadena: All

Supervisor	Ruta	Vendedor	Venta Neta AA	Venta Neta	Crecimiento vs Año Ant	% Crecimiento vs Año Ant
- NO DEFINIDO -	- NO DEFINIDO -	- NO DEFINIDO -				-13.32 %
- NO DEFINIDO -	Clientes especiales	- NO DEFINIDO -				-59.80 %
- NO DEFINIDO -	Ruta 6 Alemora	- NO DEFINIDO -				-39.02 %
- NO DEFINIDO -	RUTA AM PM	Administracion central				46.14 %
- NO DEFINIDO -	RUTA MUSI	- NO DEFINIDO -				-43.36 %
GONZALO VALVERDE	Ruta 30	YEXON ROJAS ROJAS				-42.08 %
GONZALO VALVERDE	Ruta 61	ANDREY MURILLO ARAYA				-27.35 %
GONZALO VALVERDE	Ruta 62	MAURICIO GALARZA VALLADARES				-31.90 %
GONZALO VALVERDE	Ruta 63	LORENZO PEREZ PEÑARANDA				-25.06 %
GONZALO VALVERDE	Ruta 68	ALLAN ARIAS MARTINEZ				-2.34 %
GONZALO VALVERDE	Ruta 94	JOHAN ALPIZAR RODRIGUEZ				-30.74 %
JASON QUESADA	Ruta 26	JUAN CARLOS VILLALOBOS				16.80 %
JASON QUESADA	Ruta 38	JACOB GUZMAN LEITON				-5.67 %
JASON QUESADA	Ruta 39	DANILO MATA GONZALEZ				-0.24 %
JASON QUESADA	Ruta 40	MARVIN SANCHEZ MATA				-0.75 %
JASON QUESADA	Ruta 45	JONATHAN MENDEZ ZUÑIGA				-14.83 %
JASON QUESADA	Ruta 51	ADAN PICADO RIVERA				-4.95 %
JASON QUESADA	Ruta 56	LUIS CEDERO SEGURA				-5.05 %
JASON QUESADA	Ruta 93	AARON RIVERA ZUÑIGA				-4.22 %
JOHN CRUZ	Ruta 60	DOUGLAS VARELA QUINTANA				2.77 %
JOHN CRUZ	Ruta 64	JOSE DELGADO NAVARRO				-15.68 %
JOHN CRUZ	Ruta 65	GREVIN CASCANTE HERNANDEZ				-7.74 %
JOHN CRUZ	Ruta 66	VIDIAN PORRAS CORTES				4.60 %
JOHN CRUZ	Ruta 67	JONATHAN HERRERA MORA				-21.42 %
JOHN CRUZ	Ruta 69	DAVID ACUÑA ROJAS				-27.05 %
JOHN CRUZ	Ruta 97	JEINER ROMERO MORALES				-10.32 %

5.9.12 Dashboard 12 – Bonificaciones por vendedor.

Detalla el monto que cada vendedor recibe por la venta de los artículos, así como el porcentaje que representa. Permite aplicar los filtros de periodo, categoría, familia, código de artículo, el nombre del artículo, canal, grupo, tipo de cliente e indicador de inactivo.

Bloque	Código Ruta	Vendedor	Venta Bruta	Monto Bruto Bonificado	% Bonificaciones
ALLAN LOBO	-1				
DOUGLAS SOLIS	26				
	27				0.58 %
	28				
	30				0.14 %
	31				
	32				0.03 %
	33				
	34				0.19 %
	35				
	36				0.11 %
	37				
	38				0.11 %
	39				
	40				0.15 %
	43				
	45				0.22 %
	47				
	48				0.08 %
	50				
	51				0.12 %
	56				
	59				0.18 %
	60				

5.9.13 Dashboard 13 – Comisiones

Este reporte permite identificar el monto de comisión que genera cada uno de los artículos y además cuanta comisión generan los artículos por ruta.

Alimentos Jack's Interamericana

Home > Informes Gerenciales > COSTA RICA > Comisiones

Ruta: All

Articulo	Comisión Real
A.CEBOLLINOS POPUL 11grs	
A.POROPOZ POPUL 20grs	
A.SORBETO SURTIDO TIRA 22grs	
B. CEREAL SURTIDO TIRA CR 12X10	
B..CONITOS ANDYS 17grs	
B..NACHO QUESO S PICAANTE 26grs	
B..NACHO SABOR QUESO 26grs	
B.CHOKOS PEQ 30grs	
B.CRISPI CHOC PEQ 30grs	
B.CRISPI UVA PEQ 30grs	
B.FUSION CHOKO POFFIS 30g	
B.FUSION RODIT NARANIT 30g	
B.NARANITAS PEQ 30grs	
B.POFFIS PEQ 30grs	
B.RODITA FRESA PEQ 30grs	
B.TRIJUUELA MIEL PEQ 30grs	
C.CHARR.BBQ INTERM 15grs	
C.CHARR.LIM INTER15grs	
C.COCOSAS INTERM 40grs	
C.CORN CHIPS INTERM 35grs 96ud	
C.GALLETA COMP ARANDANO 25grs	
C.GALLETA COMP CHOCOLATE 25g	
C.MENEITO SUPER INTER 15grs	
C.MENEITOQUESO INTERM 25grs	
Total	

Ruta	Comisión Real
Cientes especiales	
Ruta 100	
Ruta 26	
Ruta 27	
Ruta 28	
Ruta 29	
Ruta 30	
Ruta 31	
Ruta 32	
Ruta 33	
Ruta 34	
Ruta 35	
Ruta 36	
Ruta 37	
Ruta 38	
Ruta 39	
Ruta 40	
Ruta 41	
Ruta 43	
Ruta 45	
Ruta 46	
Ruta 47	
Ruta 48	
Ruta 49	
Ruta 50	
Ruta 51	
Ruta 52	
Ruta 53	
Ruta 54	
Ruta 55	
Total	

SubGrupo: All

02 - Mes

2020-05

5.9.14 Dashboard 14 – Inventario en almacén.

Detalla la cantidad de productos en inventario físico en los almacenes, la cantidad de cajas vendidas de ese producto en el periodo, una estimación de cuantas unidades quedará al fin de periodo, el promedio de ventas diario del articulo y los dias que tiene de estar en inventario.

Alimentos Jack's Interamericana

Home | Informes Inventarios | Inventarios Almacenes

File View Edit in Power BI Desktop Explore Refresh

02 - Mes Almacén Categoría Código de Artículo

2020-04 All All All

Tipo de Calendario: Gerencial

Código de Artículo	Artículo	Inventario Físico	Cajas Vendidas del Mes	Cierre Proyectado	Promedio Diario	Días Inventario
0003	R.BONIFIC-DEVOLUC BONOBOB CHOC	0	0	0	0.00	0
0017	P.3 PACK CEREAL SURTIDO 510GRS	0	0	0	0.00	0
0024	N.BARRA SURTIDA1 CAJA 12 UDS	0	0	0	0.00	0
0029	E.POPI CARAMELO 130grs 10uds	0	0	0	0.00	0
0034	P.2 P PITA MEJITOS 325grs	0	0	0	0.00	0
0035	B..NACHO SABOR QUESO 26grs	0	0	0	0.00	0
0036	B..NACHO QUESO S PICANTE 26grs	0	0	0	0.00	0
0037	B..CONITOS ANDYS 17grs	0	0	0	0.00	0
0040	P.SET CEREALES 500grs	0	0	0	0.00	0
0043	B. CEREAL SURTIDO TIRA CR 12X10	0	0	0	0.00	0
0089	J.DOBLE PACK CJ 570grs	0	0	0	0.00	0
0107	R.MOGUL TIBURON 30grs	0	0	0	0.00	0
0130	E.MENEITO QUESO FAM 115grs	0	0	0	0.00	0
0131	D.MENEITO QUESO MED 56grs	0	0	0	0.00	0
0195	E.MEJITO JUMBO 300grs	0	0	0	0.00	0
0231	D.MENEITO CREMIDUL MED 56grs	0	0	0	0.00	0
0276	R.BONOBON LECHE 90grs	0	0	0	0.00	0
0290	R.BONOBON BLANCO 90grs	0	0	0	0.00	0
031699	P.2 P MENEI SUP PICARITA 185grs	0	0	0	0.00	0
031768	K.CHOKOS 300grs 14 ud	0	0	0	0.00	0
031769	K.NARANITAS 300grs 14ud	0	0	0	0.00	0
031770	K.POFFIS 300grs 14ud	0	0	0	0.00	0
031771	K.TRIJUELAS 300grs 14ud	0	0	0	0.00	0
031772	K.RODITA MASMEI.O 300grs.14ud	0	0	0	0.00	0
Total		118,768.62	155,338.40	163,504.18	0.00	0

5.9.15 Dashboard 15 – Descuento sobre la venta por vendedor.

Este reporte permite analizar por supervisor y ruta la venta bruta y el monto de descuento que se le aplicó, así como el porcentaje de descuento que corresponde.

Permite filtrar por periodo, categoría, familia de producto, código del artículo que se desea analizar, el canal de venta, el grupo, tipo de cliente e indicador para saber si el producto se encuentra activo en producción.

Alimentos Jack's Interamericana

Home > Informes Canal Tradicional > Descuento sobre las ventas por vendedor

02 - Mes

2020-05

Supervisor Ruta Venta Bruta Descuento % Descuento

Supervisor	Ruta	Venta Bruta	Descuento	% Descuento
- NO DEFINIDO -	RUTA AM PM			5.00 %
	RUTA MUSI			5.00 %
	Total			5.00 %
GONZALO VALVERDE	Ruta 30			9.33 %
	Ruta 61			9.42 %
	Ruta 62			9.85 %
	Ruta 63			9.14 %
	Ruta 68			9.61 %
	Ruta 94			9.21 %
	Total			9.40 %
JASON QUESADA	Ruta 26			9.16 %
	Ruta 38			9.31 %
	Ruta 39			10.00 %
	Ruta 40			5.37 %
	Ruta 45			8.26 %
	Ruta 51			6.03 %
	Ruta 56			7.59 %
	Ruta 93			10.00 %
	Total			8.69 %
	JOHN CRUZ	Ruta 64		
Ruta 65				10.56 %
Ruta 66				6.78 %
Ruta 67				10.33 %
Ruta 69				8.66 %
Ruta 97				9.70 %
Total				9.65 %
JOHNNY PANA	Ruta 70			7.19 %
	Ruta 71			8.71 %
	Ruta 72			10.00 %
	Ruta 73			10.00 %

5.9.16 Dashboard 16 – Devoluciones en buen estado.

Este reporte muestra por supervisor, ruta el monto de venta bruta, el monto correspondiente a las devoluciones en buen estado que realizan los clientes a la compañía y el porcentaje que corresponde este importe.

Alimentos Jack's Interamericana

Home > Informes Canal Tradicional > Devoluciones buen estado por vendedor

02 - Mes
2020-05

Categoría: All

Familia: All

Código de Artículo: All

Artículo: All

Canal: All

Grupo: All

Tipo de Cliente: All

Indicador Inactivo: All

Supervisor	Ruta	Venta Bruta	Monto Bruto Buen Estado	% Buen Estado
VICTOR SERRANO	Ruta 27			9.66 %
	Ruta 36			5.39 %
	Ruta 37			3.01 %
	Ruta 47			2.48 %
	Ruta 48			2.48 %
	Ruta 35			2.47 %
	Ruta 31			2.19 %
JUAN MANUEL PORRAS	Total			3.67 %
	Ruta 95			4.63 %
	Ruta 76			4.44 %
	Ruta 82			3.76 %
	Ruta 53			3.09 %
	Ruta 86			2.91 %
	Ruta 75			2.71 %
GONZALO VALVERDE	Ruta 46			2.42 %
	Total			3.47 %
	Ruta 62			5.19 %
	Ruta 61			5.19 %
	Ruta 30			2.50 %
	Ruta 63			2.01 %
	Ruta 94			1.71 %
MANUEL MORALES	Ruta 68			1.10 %
	Total			3.06 %
	Ruta 49			16.06 %
	Ruta 41			3.58 %
	Ruta 55			2.52 %
	Ruta 57			2.37 %

5.9.17 Dashboard 17 – Devoluciones en mal estado.

Este reporte muestra por supervisor, ruta el monto de venta bruta, el monto correspondiente a las devoluciones en mal estado que realizan los clientes a la compañía y el porcentaje que corresponde este importe.

02 - Mes		Supervisor	Ruta	Venta Bruta	Monto Bruto Mal Estado	% Mal Estado
2020-05		- NO DEFINIDO -	- NO DEFINIDO -			
Categoría	All		Cientes especiales			
Familia	All		PRODUCTOS ALIMENTICIOS PASCUAL S.A.			
Codigo de Articulo	All		Ruta 6 Alemora			
Articulo	All		RUTA AM PM			
Canal	All		RUTA MUSI			8.60 %
Grupo	All		Total			0.12 %
Tipo de Cliente	All	GONZALO VALVERDE	Ruta 30			3.93 %
Indicador Inactivo	All		Ruta 61			2.15 %
			Ruta 62			2.12 %
			Ruta 63			2.34 %
			Ruta 68			1.86 %
			Ruta 94			3.55 %
			Total			2.57 %
		JASON QUESADA	Ruta 26			1.89 %
			Ruta 38			1.20 %
			Ruta 39			1.64 %
			Ruta 40			1.53 %
			Ruta 45			2.25 %
			Ruta 51			3.13 %
			Ruta 56			2.22 %
			Ruta 93			1.72 %
			Total			1.92 %
		JOHN CRUZ	Ruta 60			1.78 %
			Ruta 64			2.07 %
			Ruta 65			1.90 %
			Ruta 66			1.48 %
			Ruta 67			2.27 %

Capítulo VI

Conclusiones y Recomendaciones del Proyecto

6.1 Conclusiones

Debido a la necesidad de alimentos Jack´s de tener una herramienta analítica de inteligencia de negocios que les permitiera a las áreas de ventas y comercial la toma de decisiones de manera clara, confiable y sobre todo ágil, fue que se originó el desarrollo de este proyecto de un DataMart comercial. Claro está que con este proyecto no se solventan todas las necesidades que pueda tener la organización en las diferentes áreas, pero si solventa por completo las dificultades de las áreas: comercial y ventas. Con la implementación de este proyecto, se logra la reducción del tiempo que los usuarios deben invertir tabulando y homologando información para lograr tener datos confiables para la toma de decisiones, además, que ahora cuentan con datos analíticos y cuantitativos al cierre del día anterior, logrando de esta forma tomar decisiones de forma proactiva, y no como anteriormente lo hacían, teniendo que esperar el cierre del mes para comenzar a tabular los datos y cuando ya los tenían listos para analizar, ya solo podían hacer una análisis forense, debido que no había nada que hacer para cambiar el rumbo de las ventas o los inventarios para el mes en cuestión. Era un análisis “post mortem” y las decisiones eran completamente reactivas.

Con la implementación de esta solución de Business Intelligence, se logró también dar portabilidad a la información, ya que anteriormente se tenía todo en un archivo de Excel, y eso era muy poco portable. Ahora, al tener una base de datos multidimensional y herramientas de punta como lo es Microsoft PowerBI, los reportes preestablecidos, más los que los usuarios van generando en su día a día, se van publicando en un repositorio Web, el cual es consultado por los usuarios desde sus diferentes dispositivos (computadoras, tabletas e incluso teléfonos inteligentes) con solo tener una conexión a internet.

Esto ha enriquecido el uso de la tecnología, ya que todos los usuarios, independientemente de su nivel de conocimiento tecnológico, tienen la información al alcance, únicamente presionando un clic. Ahora bien, para los usuarios más expertos y que fueron ampliamente capacitados, pueden no solamente consumir esos reportes, sino, que, por medio del uso de los cubos multidimensionales, puede hacer todos los análisis que se deseen, haciendo mezcla de todas las variables tanto cualitativas como cuantitativas, para ver tendencias, analizar de diferente forma su información y lo más importante, descubrir patrones entre sus datos que quizá antes era muy difíciles de encontrar.

Los mismos usuarios han logrado concluir, que el uso de esta tecnología les da una enorme ventaja e independencia del área de tecnología, ya que actualmente no deben estar solicitándoles reportes o descargas de información para procesarla posteriormente, sino que tienen total independencia para obtener información y lo más importante, sin afectar los sistemas transaccionales con consultas pesadas, dejando ahora los sistemas transaccionales para lo que realmente fueron construidos, que es procesar las diferentes transacciones de ventas y no así satisfacer la demanda de información por medio de consultas, ahora para este fin se tiene una solución tecnológica especializada con altos niveles de rendimiento y confiabilidad.

Los usuarios se encuentran altamente satisfechos, debido a que ahora se sienten que son realmente dueños de la información, y con libertad de consumirla para su toma de decisiones, ya que anteriormente parecía que los dueños de la información eran los del área de tecnología, quienes eran los encargados de custodiarla y entregarla a los usuarios cuando ellos la solicitaran, pero dependiendo de la cola de trabajo que tuvieran, lo cual no siempre era rápido.

Por lo anterior, todo el tiempo que los usuarios utilizaban anteriormente para tabular la información y unir todas las fuentes de datos diferentes que tenían, ahora pueden destinarlo a analizar los datos para tomar mejores decisiones, y pueden invertir ese tiempo en plantear estrategias para aumentar las ventas o administrar mejor sus inventarios o incluso generar promociones para incrementar las ventas de productos que no están teniendo el comportamiento esperado. Las diferentes áreas intervenidas con este proyecto dejarán de ser máquinas de molienda de datos y por fin se podrán concentrar en las actividades reales del departamento, logrando de esta forma aumentar sus ingresos y minimizar sus gastos, lo cual es sumamente importante en cualquier empresa, pero principalmente en un nicho de mercado tan dinámico y con tanta competencia como lo es el de alimentos Jack's.

6.2 Recomendaciones

Si bien es cierto que consideramos el desarrollo de este proyecto como un éxito absoluto, siempre quedan cosas que pueden mejorarse, y nuestras recomendaciones para alimentos Jack's en general son las siguientes:

1. Pasar el Datamart construido a un ambiente de nube. Esto porque la fuente principal de información se encuentra en AX el cual está en nube, y la comunicación sería más transparente e incluso el costo se reduciría ya que no tiene que bajarse los datos a un servidor OnPremise.
2. Ampliar el desarrollo de las soluciones de BI a otras áreas de la empresa. Incluso el área de comercial tiene cierta información que no está en el datamart de ventas, como por ejemplo lo que se conoce como ventas SellOut. Los cual es información que suministran las diferentes cadenas a Jack's. Pero también es recomendable que estas soluciones se propaguen a otras áreas, como, por ejemplo, producción, logística y compras.
3. Masificar las capacitaciones internas a todos los usuarios involucrados, para aumentar el número de usuarios expertos en el uso de la herramienta y con esto tener mayor cantidad de usuarios analíticos de información y no solo de consumo estándar.
4. Capacitar al personal técnico interno en la construcción de soluciones de BI para que puedan mantener y escalar la solución, así como propagar las herramientas a las otras áreas de la empresa.

Capítulo VII

Referencias bibliográficas

8.1 Bibliografía

Acosta, V.M. (17 de Mayo del 2019) Comparación entre Data Warehouse y Data Mart. Recuperado de <https://revistadigital.inesem.es/informatica-y-tics/diferencia-entre-data-mart-y-data-warehouse/>

Aimacaña, D. (2013). Análisis, Diseño E Implementación De Un Data Mart Académico Usando Tecnología De Bi Para La Facultad De Ingeniería, Ciencias Físicas Y Matemática (tesis de pregrado). Universidad Central del Ecuador, Quito, Ecuador.

Analíticas del big data. (s.f). SAS. Recuperado de https://www.sas.com/es_mx/insights/analytics/big-data-analytics.html

Bembibre, V. (2009). Definición de Datawarehouse. Definición ABC. Recuperado de <https://www.definicionabc.com/tecnologia/datawarehouse.php>

Bibliografía: EAE Business School. (2018). La tecnología OLTP: qué es y qué puede hacer por ti. España. Recuperado de <https://retos-operaciones-logistica.eae.es/la-tecnologia-oltp-que-es-y-que-puede-hacer-por-ti/>

Big Data Analytics: en qué consiste y cómo beneficia a tu empresa. (21 de agosto del 2014). Logicalis. Recuperado de <https://blog.es.logicalis.com/analytics/en-que-consiste-big-data-analytics-y-como-beneficia-a-tu-empresa>

Business Intelligence, Inteligencia de Negocios o Inteligencia Empresarial. (1 de Octubre del 2019) Profitline. Recuperado de <https://profitline.com.co/business-intelligence-inteligencia-de-negocios-o-inteligencia-empresarial/>

Business Intelligence: ¿Cómo ayuda a optimizar decisiones y rendimiento?. (s.f). Recuperado de <https://blog.euncet.es/business-intelligence-como-ayuda/>

Calvo, M. (2019). Business intelligence: ¿Por qué lo necesito en mi empresa? Recuperado de <https://www.captio.net/blog/algunos-ejemplos-practicos-de-uso-de-business-intelligence>

Carisio, E. (14 de Agosto del 2019). ETL: todo sobre el proceso de Extract, Transform and Load. Recuperado de [https://blog.mdcloud.es/que-es-etl-extraccion-transformacion-y-carga/#:~:text=El%20ETL%20es%20un%20m%C3%A9todo,los%20objetivos%20de%20nuestra%20compa%C3%B1a%20Da.&text=Integrarse%20con%20sistemas%20transaccionales%2C%20almacenes,Master%20Data%20Management%20\(MDM\).](https://blog.mdcloud.es/que-es-etl-extraccion-transformacion-y-carga/#:~:text=El%20ETL%20es%20un%20m%C3%A9todo,los%20objetivos%20de%20nuestra%20compa%C3%B1a%20Da.&text=Integrarse%20con%20sistemas%20transaccionales%2C%20almacenes,Master%20Data%20Management%20(MDM).)

Cubos OLAP de información para la toma de decisiones. (16 de Agosto del 2016). Evaluando software.com. Recuperado de <https://www.evaluandosoftware.com/cubos-olap-informacion-la-toma-decisiones/>

Data Analytics (28 de Marzo del 2019). Galvanize. Recuperado de: <https://fr.wegalvanize.com/risk/data-analytics-risk-assessments/>

Data Lakes: Definiciones y Plataformas. (30 de Octubre 2019). TodoBI. Recuperado de <https://todobi.com/data-lakes-definiciones-y-plataformas/>

Data Mart vs. Data Warehouse. (s.f). Recuperado de <https://www.quora.com/Dimensional-modeling-as-popularized-by-Ralph-Kimball-is-perhaps-the-most-popular-data-warehouse-implementation-methodology-Yet-some-question-the-scalability-of-the-Kimballs-methodology-What-are-some-disadvantages-to-the-K-method>

Data Warehouse Y Business Intelligence. (13 de Abril del 2018). Data Warehouse recuperado de <https://15257olga.wordpress.com/2018/04/13/datawarehouse-y-business-intelligence/>

Data Warehouse. (s.f) Capítulo 2. Recuperado de http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/ydirin_p_mm/capitulo2.pdf

Dertiano, V (09 de marzo del 2015). Arquitectura BI (Parte II): El enfoque de William H. Inmon. Recuperado de <https://blog.bi-geek.com/arquitectura-enfoque-de-william-h-inmon/>

Dertiano, V (09 de marzo del 2015). Arquitectura BI (Parte II): El enfoque de William H. Inmon. Recuperado de <https://blog.bi-geek.com/arquitectura-enfoque-de-william-h-inmon/>

Dertiano, V. (02 de Mayo del 2016) Arquitectura BI (Parte IV): Comparativa entre Inmon y Kimball. Recuperado de <https://blog.bi-geek.com/arquitectura-comparativa-inmon-y-kimball/>

Dertiano, V (06 de Abril del 2015) Arquitectura BI (Parte III): El enfoque de Ralph Kimball.
Recuperado de <https://blog.bi-geek.com/arquitectura-el-enfoque-de-ralph-kimball/>

Difference Between OLTP and OLAP. (2019). Recuperado de <https://www.educba.com/oltp-vs-olap/>

¿En qué consiste un data mart?, (06 de Febrero del 2019). Taled. Recuperado de <https://es.talend.com/resources/what-is-data-mart/>

Fernández, A. (28 de Julio del 2017). 5 Herramientas para la visualización de datos. Recuperado de <https://blogthinkbig.com/herramientas-para-la-visualizacion-de-datos>

<https://www.mdirector.com/marketing-digital/data-lake.html>

Inteligencia empresarial. (s.f). Recuperado de https://es.wikipedia.org/wiki/Inteligencia_empresarial

Las 20 herramientas de visualización de datos Big Data del momento, GRAPHEVERYWHERE.
Recuperado de <https://www.grapheverywhere.com/las-20-herramientas-de-visualizacion-de-datos-big-data-del-momento/>

Loshin, D. (27 de Marzo del 2018). Tablas de dimensión vs tablas de hechos: ¿Cuál es la diferencia?. Knowledge Integrity Inc. Recuperado de <https://searchdatacenter.techtarget.com/es/consejo/Tablas-de-dimension-vs-tablas-de-hechos-Cual-es-la-diferencia>

Microsoft Infrastructure, Virtualization & System Management. (25 de diciembre del 2013).
Recuperado de <https://pkjayan.wordpress.com/tag/file-services/>

Marker, G. (04 de Septiembre del 2019) Sistemas OLAP: Análisis empresarial. Cubos y tipos de OLAP. Tecnología más informática. Recuperado de <https://www.tecnologia-informatica.com/sistemas-olap-cubos/>

Muñiz, L. (2018). El Business Intelligence herramienta clave para mejorar la gestión empresarial. Recuperado el 07 de julio de 2019, de Sistema Control de Gestión: http://www.sistemacontrolgestion.com/Portals/1/Ebook_Mejorar_gestion_BI_SCG_Estrategia_v18.pdf#page=19&zoom=100,0,94

Pearlman, S. (19 de agosto del 2019). ¿En qué consiste un proceso de ETL (Extraer, Transformar y Cargar)?. Talend. Recuperado de <https://es.talend.com/resources/what-is-etl/>

Procesamiento de transacciones en línea (OLTP). (27 de Julio del 2019) Microsoft . Recuperado de <https://docs.microsoft.com/es-es/azure/architecture/data-guide/relational-data/online-transaction-processing>

Proceso ETL y los pasos para su implementación. (08 de Mayo 2020). Astera. Recuperado de <https://www.astera.com/es/tipo/Ofertas-%26-Solicitudes/proceso-etl-y-pasos/>

QlikTech International AB. (2020). Plataforma de analítica de datos. Recuperado de <https://www.qlik.com/es-es/products/qlik-sense>

¿Qué es Business Intelligence (BI) y qué herramientas existen? (10 de mayo del 2017). El Blog de Signaturit. Recuperado de <https://blog.signaturit.com/es/que-es-business-intelligence-bi-y-que-herramientas-existen>

¿Qué es un Data Lake? (09 de junio del 2017). MDirector. Recuperado de <https://www.mdirector.com/marketing-digital/data-lake.html>

Qué es Power BI. (Mayo 2019). Numsys. Recuperado de <https://www.nunsys.com/power-bi/>

Ramirez, A. (04 de Septiembre 2011) Business Intelligence. Recuperado de <http://r-pita.blogspot.com/2011/09/modelo-estrella-o-copo-de-nieve.html>

Ramos, S. (2017). Modelado Dimensional: Data Warehouse, Data Marts y Esquemas de Diseño (11). Solid. Recuperado de <https://blogs.solidq.com/es/business-analytics/data-warehouse-y-data-marts-esquema-en-estrella-11/>

Referencia: Balagueró, Th. (02 de octubre dl 2018). ¿Qué es y para qué sirve el Data Lake?. Deusto Formación. Recuperado de <https://www.deustoformacion.com/blog/gestion-empresas/que-es-para-que-sirve-data-lake>

Referencia: Microsoft, (2019). Understand star schema and the importance for Power BI. Recuperado de <https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>

Referencia: Palmer, Th. (11 de Diciembre 2015). Analyze your Office365Mon data in Power BI. Recuperado de <https://powerbi.microsoft.com/es-es/blog/analyze-your-office365mon-data-in-power-bi/>

Scavicchio, J. (s.f) What's Damming Data Lakes? Insights from Tableau, Qlik and Logi Analytics. Recuperado de <https://datafloq.com/read/damming-data-lakes-insights-tableau-qlik-logi/2162>

Sistemas OLTP: procesamiento, administración y mantenimiento de transacciones. (24 de Agosto del 2016). Recuperado de <https://www.evaluandosoftware.com/sistemas-oltp-procesamiento-administracion-mantenimiento-transacciones/#:~:text=Un%20sistema%20OLTP%20se%20puede,de%20disponibilidad%20C%20seguridad%20y%20confiabilidad.>

Tableau (s.f). Nova Praxis. Recuperado de <https://nova-praxis.com/software-business-intelligence-epm/tableau-desktop-online/>