



PÁG. 2

Requerimientos del negocio

PÁG. 3

Fuentes de datos

PÁG. 4

Procesos de datos requeridos

Visualización de información y
auto-servicio de datos

BANMARKET

INTRODUCCIÓN

BANMARKET es un gran Retail que cubre las necesidades de medio millón clientes. Cuenta con una importante red que cubre las 20 regiones comerciales en todo el país. Sus productos principales incluyen materiales de oficina, muebles, tecnología, inclusive abarrotes y productos de comidas rápidas.

La historia de BANMARKET se remonta a más de 50 años, cuando inició con una única sucursal. Sin embargo, en los siguientes años, tuvo un crecimiento sin precedentes debido a su enfoque en la satisfacción del cliente y en sus productos y servicios basados en las necesidades de sus clientes.

AUTOR DEL CASO

JULIAN GRIJALBA FACUNDO

julian@here4data
@here4data

Las operaciones de BANMARKET se dividen en 20 regiones. En la sede principal se encuentra el departamento de TI, que gestiona varios sistemas de software y aplicaciones, además del sitio web de BANMARKET. Varios analistas se encuentran también en la sede

principal, y ellos analizan los informes generados por los datos de sus operaciones diarias. Basados en estos análisis, los analistas y directivos se aseguran de que la empresa lleve a cabo sus operaciones de acuerdo con todas las regulaciones financieras y de comercio que los rigen.

Los sistemas de TI se dividen en un datacenter, ubicado en la sede principal, y en varias sucursales ubicadas en todo el país. En el centro de los sistemas de TI se encuentra un sistema OLTP sobre un potente servidor, que almacena la información de los clientes. Este sistema central maneja transacciones en las regiones y también se utiliza para generar varios informes a nivel operacional. Además, este sistema central utiliza su propia base de datos relacional especializada. Cada región cuenta con un sistema OLTP basado en servidores de cliente, con su propia base de datos relacional para procesar las transacciones diarias de los clientes. La empresa utiliza un modelo de procesamiento por lotes en el que cada región envía sus datos de transacciones diarias a la sede principal al finalizar la jornada laboral. Los datos de cada región se procesan en el sistema OLTP central, mediante la fusión de las transacciones de todas las sucursales. Luego de conciliar todas las transacciones, la información actualizada de la cuenta es enviada a cada sucursal.

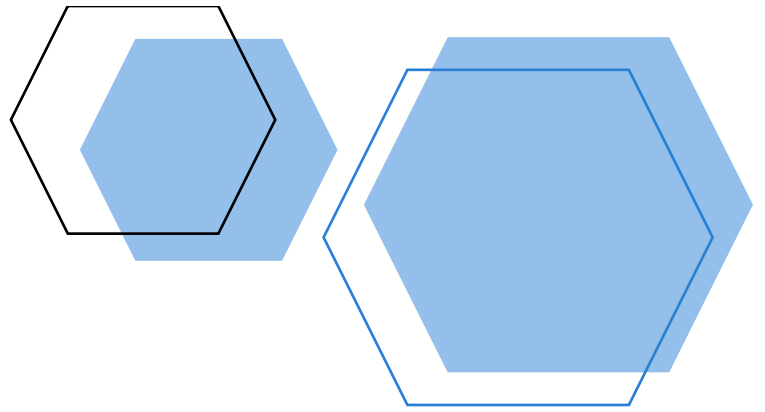
El área de negocio financiera ha identificado un problema recurrente al momento de tratar de acceder a la información requerida, y es que no hay un lugar donde esté la información integrada, con calidad y con la periodicidad requerida por el negocio. Por tal motivo ha solicitado la creación de un modelo de información que permita cubrir sus necesidades actuales y futuras de información, con la facilidad de crear análisis ad-hoc/periódicos desde las diferentes perspectivas que tienen sus datos y así garantizar una mejor toma de decisión como un seguimiento a la evolución del negocio.



REQUERIMIENTO

Esta área ha compartido las fuentes de datos que tienen actualmente, con el objetivo de que su equipo de trabajo las pueda analizar, ya que es indispensable que esta información pueda ser unificada en el data warehouse. De igual forma, describe los siguientes requerimientos de información que usualmente aplican, para que los incluyan en el diseño del modelo de información:

- Analizar los clientes y productos que más generan valor en la organización.
- La tendencia que tienen actualmente las diferentes sucursales con respecto a sus ventas y cumplimiento de sus targets.
- Identificar clientes con compras en el primer cuartil y productos de baja rotación con el fin de intervenirlos y mejorarlos.
- Oportunidad de penetración en nuevas categorías para cada uno de los clientes.
- Pérdida de clientes, la cual se identifica cuando un cliente deja de comprar por más de 6 meses.



FUENTES DE DATOS

Las fuentes de datos disponibles son las siguientes:

Archivos planos:

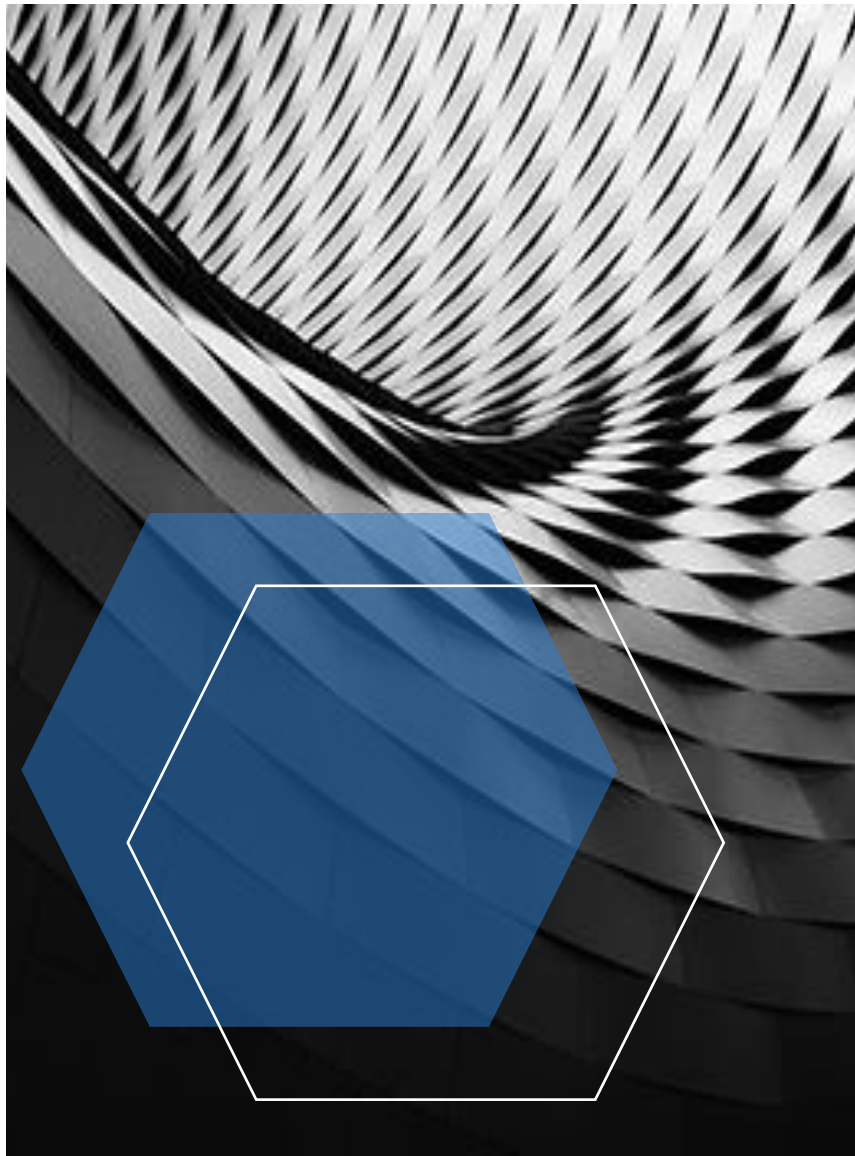
<https://github.com/here4data/DWH/tree/main/Sources>

Base de datos:

- here4data.database.windows.net
- Base de datos: EDWH
- Usuario: gateway
- Clave: Thegate13?

En los archivos planos se encuentra una copia de seguridad de la fuente de la base de datos, en ocasiones esta esta caída.

Dataset	Formato	Frecuencia de actualización	Descripción (campos)
LT_Customer	JSON	Diaria	Información básica de los clientes: Cliente ID, Nombre, Apellido y Segmento
tb_detalle_envios	SQL SERVER	Diaria	Información de los envíos: Numero Orden, Fecha Orden, Fecha Envio, Modo Envio, Cliente ID, SubTotal y Region ID
LT_detalle_ordenes	SQL SERVER	Diaria	Información de las órdenes que se procesan en el supermercado: Detalle Orden, Numero Orden, Producto ID, Linea Total, Cantidad, Descuento, Precio Unitario y Ganancia
LT_productos	CSV	Demanda	Información del producto: Producto ID, Numero Producto, Producto Sub Categoría ID y Nombre Producto
LT_productos_categoria	CSV	Demanda	Información de la categoría del producto: Categoría Producto ID y Descripción Categoría
LT_productos_sub_categoria	CSV	Demanda	Información de la subcategoría del producto: Subcategoría Producto ID, Categoría Producto ID y Descripción Subcategoría
lt_customer_localizacion	TXT	Semanal	Información de la localización de los Clientes: Cliente ID, País, Ciudad, Estado, Código Postal y Region
db_target_region	CSV	Trimestral	Información de los targets definidos en las regiones: Año, Mes, Region id, Target
lt_geografia	JSON	Demanda	Información geográfica: País, Estado, Ciudad
tb_modos_envio	XML	Demanda	Información de las formas en que se entrega los productos comprados: id_modos, Descripción, Tipo ID
tb_tipo_envio	XML	Demanda	Información de los tipos de envíos asociados al modo de envío: id_tipo, Descripción
Regiones	TXT	Demanda	Información de la distribución de las regiones comerciales : Region ID, Region Desc y Detalle region



PROCESOS DE DATOS REQUERIDOS

Un análisis de una muestra de estos Dataset sugiere que, antes de poder usarlos, se requiere refinar los datos, incluyendo el filtrado (filtering), la validación y la conversión de formatos, para lo cual usted debe plantear una solución con base en cada uno de los registros procesados.

Luego de que los datos han sido refinados, se deben extraer algunos campos de cada Dataset, y deben ser unidos en un registro único. Aunque la calidad de los Dataset es bastante buena, no se garantiza que todos los campos requeridos se completen en este único registro.

DATOS A CONOCIMIENTO

VISUALIZACIÓN DE INFORMACIÓN Y AUTO-SERVICIO DE DATOS

Al observar los datos refinados, algunos analistas creen que otros departamentos también pueden utilizar estos datos, tales como el de mercadeo, para desarrollar productos financieros atractivos y realizar análisis detallados. Para determinar la mejor forma de compartir esta información a perfiles no técnicos, es necesario generar una capa de información la cual se describe a continuación.

- Generar una serie de reportes que respondan a los requerimientos definidos por el área de negocio. En este diseño se debe considerar y aplicar las practicas buenas prácticas descritas para los procesos de datos.
- Construir los cálculos requeridos para cumplir con los requerimientos del negocio, como también todo el modelo que estará conectada a la data warehouse.
- Crean los productos visuales requeridos por el área de negocio