

## 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 Región. 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 regiones 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 regiones. Luego de conciliar todas las transacciones, la información actualizada de la cuenta es enviada a cada región.

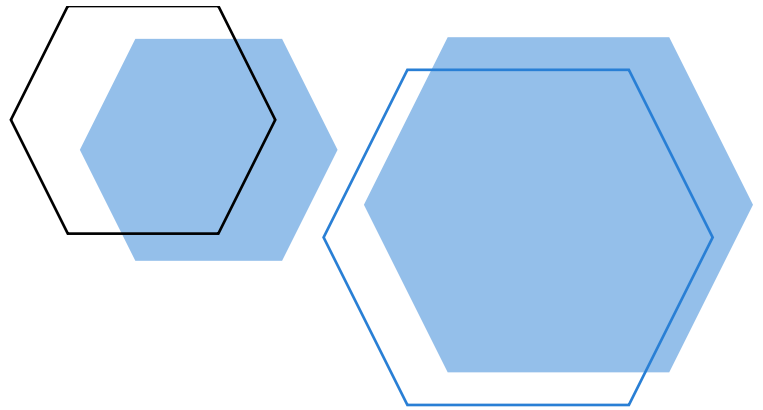
El área de negocio financiera ha identificado un problema recurrente al momento de tratar de acceder a la información requerida, y es que no hay un lugar donde esté la información integrada, con calidad y con la periodicidad requerida por el negocio. Por tal motivo ha solicitado la creación de un modelo 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 ganancia neta en la organización.
- La tendencia que tienen actualmente las diferentes regiones con respecto a sus ganancias y cumplimiento de sus targets.
- Identificar clientes en los últimos 3 meses de información 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 con compras en los últimos tres meses de información.
- 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>

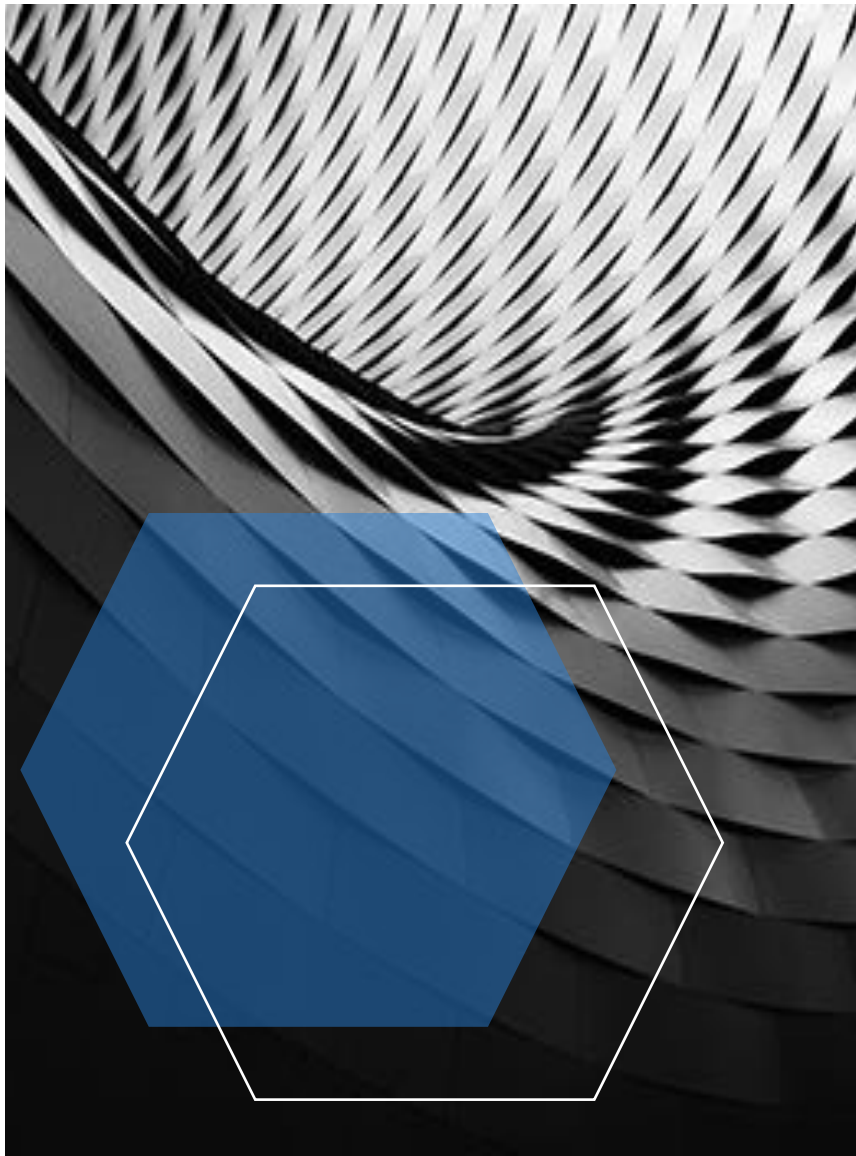
### Base de datos:

- 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.

El área de negocio con base en su entregable identifico que hay cierta información que no ha cubierto en su requerimiento inicial y por ende agradece que ustedes lo realicen durante esta etapa.

El requerimiento solicitado es la creación de nueva información con base en los datos base que ustedes ya gestionan. Estas nuevas métricas son **Ganancia Neta**, **Impuestos**, **Valor logístico**, **Valor provisión COVID**.

De igual forma les facilita una serie de definiciones para que tengan en cuenta para la gestión de los datos que ustedes van a aprovisionar en el Data warehouse para BankMarket.

Definición	Reglas de negocio
Reglas de región	Son todas los temas que se necesitan cubrir para garantizar que la información sea consistente
Línea total	Se refiere al valor total de la orden
Geografía	Se tiene que validar muy bien la relación entre geografía y la región
Ordenes	Las ordenes se pueden repetir por cada región
Ganancia	Es el cálculo del (precio unitario * cantidad) - descuento
Ganancia Neta	Es el cálculo de la Ganancia - valor de logística - valor de provisión COVID
Impuestos	Es el 19% del valor de la ganancia
Valor logístico	Representa el 12% del precio unitario * cantidad . Este se refiere al costo de transporte y almacenamiento
Valor provisión COVID	Representa el 3% del precio unitario * cantidad. Esto se refiere a los sobrecostos generados por los elementos de bioseguridad requeridos para la operación.
Localización del Cliente	Un cliente puede tener una sola localidad. Si no se identifica en la dimensión debe ser tratado como un huérfano
Categoría	Una categoría puede tener una o más subcategorías y por ende uno o más productos. Si no se identifica en la dimensión debe ser tratado como un huérfano
Sub Categoría	Una sub categoría puede tener uno o más productos. Si no se identifica en la dimensión debe ser tratado como un huérfano
Región	Puede tener uno o más estados y a su vez uno o más ciudades asociadas. Si no se identifica en la dimensión debe ser tratado como un huérfano
Estado	Puede tener uno o más ciudades asociadas. Si no se identifica en la dimensión debe ser tratado como un huérfano
Ciudad	Tiene una región y un solo estado asociado. Si no se identifica en la dimensión debe ser tratado como un huérfano
Producto	Tiene una categoría y sub categoría asociada. Si no se identifica en la dimensión debe ser tratado como un huérfano
Cliente	Un cliente es único. Si no se identifica en la dimensión debe ser tratado como un huérfano
Tipo de envío	Un tipo de envío puede estar asociado a más de un modo de envío. Si no se identifica en la dimensión debe ser tratado como un huérfano
Modo de envío	Un modo de envío tiene un solo tipo de envío. Si no se identifica en la dimensión debe ser tratado como un huérfano
Huertano	Cualquier valor que proceda desde los sistemas fuentes y no se tenga referencia en la dimensión.

Definición	Reglas de negocio
Target	Es el presupuesto definido para cada región a cumplirse. Este se realiza con base en la ganancia neta. La información de su periodicidad es con base en la fecha de orden
Fecha de orden	Es el momento en el cual el cliente realiza la orden
Fecha de envió	Es el momento en el cual la orden generada se le ha enviado al cliente

El negocio define que toda la información que no cumpla con las reglas de negocio o no tenga integridad referencial se debería guardar en otro repositorio y notificar al área de negocio por medio de correo a su docente para que se pueda solucionar estos errores directamente desde la fuente de datos. La gestión de huérfanos se tendrá que realizar colocando en la dimensión respectiva como “huérfano de ” en caso por ejemplo que sea una ciudad , sería una huérfano de ciudad. Esto se reporta de igual forma como en el punto anterior.

Es por ello por lo que después de las validaciones de carga que ustedes realicen y aprovisionar su data warehouse, deberán enviar el reporte de errores por mail indicando en que tabla del data warehouse estarán los errores para que el área de negocio pueda ingresar , revisarlos y confirmarles fecha de solución estimada, con el fin de que ustedes puedan seguir avanzando en su trabajo.



# 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