

Facultad de Ingeniería y Ciencias Básicas
Universidad Central
Maestría Analítica de Datos
Curso de Bases de Datos
Bogotá, Colombia
{¹*jacevedoq*,²*jarguellt*,³*wlopezg*}@ucentral.edu.co
LaTeX2e

Proyecto de Bases de datos B2B

Javier Acevedo¹, Jose Ignacio Arguello², Willian Lopez³

October 19, 2024

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	4
2	Características del proyecto de investigación que hace uso de Bases de Datos (Max 500 Palabras) - (<i>Primera entrega</i>)	4
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2.1	Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	5
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	5
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	6
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	7
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? (Max 100 Palabras) - (<i>Primera entrega</i>)	7
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	7
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(<i>Primera entrega</i>)	8
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (<i>Primera entrega</i>)	8
4.2	Diagrama modelo de datos (<i>Primera entrega</i>)	9
4.3	Imágenes de la Base de Datos (<i>Primera entrega</i>)	9
4.4	Código SQL - lenguaje de definición de datos (DDL) (<i>Primera entrega</i>)	9
4.5	Código SQL - Manipulación de datos (DML) (<i>Primera entrega</i>)	9
4.6	Código SQL + Resultados: Vistas (<i>Primera entrega</i>)	9
4.7	Código SQL + Resultados: Triggers (<i>Primera entrega</i>)	9
4.8	Código SQL + Resultados: Funciones (<i>Primera entrega</i>)	9
4.9	Código SQL + Resultados: procedimientos almacenados (<i>Primera entrega</i>)	9

5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	10
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	10
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	10
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	11
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	11
6.2	Automatización de Datos (<i>Tercera entrega</i>)	11
6.3	Integración de Datos (<i>Tercera entrega</i>)	11
7	Proximos pasos (<i>Tercera entrega</i>)	12
8	Lecciones aprendidas (<i>Tercera entrega</i>)	13
9	Bibliografía	14

1 Introducción (Max 250 Palabras) - *(Primera entrega)*

En el contexto de una empresa manufacturera como la nuestra, donde el 80% de las ventas se concentra a través de distribuidores y mayoristas, es fundamental entender el comportamiento de ventas del cliente final. Dado que nuestra fortaleza radica en la fabricación y no en la distribución, el seguimiento de las ventas hasta el cliente final (sell-out) se convierte en una pieza clave para la optimización de recursos, la estrategia de mercado y la toma de decisiones comerciales. Este documento aborda la necesidad de desarrollar una base de datos robusta que nos permita analizar detalladamente el comportamiento de las ventas a nivel geográfico y de producto, con el fin de fortalecer nuestra posición en el mercado y optimizar nuestras estrategias de comercialización B2B.

2 Características del proyecto de investigación que hace uso de Bases de Datos (Max 500 Palabras) - *(Primera entrega)*

El proyecto se centra en el desarrollo de una base de datos integral que capture la información de sell-out en el contexto de una empresa manufacturera de PVC en Colombia. En un entorno donde el 80 por ciento de las ventas se canaliza a través de distribuidores y mayoristas, se hace necesario comprender mejor el comportamiento de las ventas al cliente final. La capacidad de rastrear y analizar esta información es crucial para optimizar los recursos y fortalecer las estrategias de comercialización.

2.1 Título del proyecto de investigación (Max 100 Palabras) - *(Primera entrega)*

B2B sell-out manufacturing.

2.2 Objetivo general (Max 100 Palabras) - *(Primera entrega)*

Desarrollar una base de datos integral que capture la información de sell-out, estructurada para permitir un análisis detallado a nivel geográfico (ciudades, departamentos, localidades, barrios) y de producto (categorías, subcategorías y materiales), con el fin de facilitar un análisis del comportamiento de ventas.

2.2.1 Objetivos específicos (Max 100 Palabras) - *(Primera entrega)*

- Estructurar la base de datos de sell-out que permita desglosar las ventas por ubicación geográfica, utilizando jerarquías claras como ciudades, departamentos, localidades y barrios.
- Organizar la información de producto en la base de datos bajo jerarquías definidas, como categorías, subcategorías y materiales, para facilitar un análisis profundo por tipo de producto.
- Generar reportes de comportamiento de ventas del cliente final que ofrezcan una visión clara y precisa de los patrones de compra por cliente, ciudad, producto en sus diferentes jerarquías.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

El ejercicio está dado por una base de datos a partir de facturas de ventas de un fabricante (sell-in) y facturas de venta de un distribuidor mayorista al cliente final (sell-out).

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cómo almacenar datos para entender el comportamiento de los clientes finales de una fábrica de PVC colombiana a través del sell-out y que a partir de esto se pueda generar análisis y acompañar la estrategia de crecimiento en venta?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

La implementación de una base de datos centralizada permitirá a la fábrica identificar oportunidades de expansión y mejorar la eficiencia operativa al optimizar la gestión de la relación con los mayoristas.

Justificación: Al consolidar toda la información de ventas y productos en una única base de datos, se podrá tener una visión completa del rendimiento de los mayoristas, detectar áreas de mejora en la operación y crear estrategias más efectivas para fortalecer la relación comercial y expandir la presencia en el mercado final.

3 Reflexiones sobre el origen de datos e información

(Max 400 Palabras) - (Primera entrega)

El principal reto de este proyecto, en el que la información proviene de varios mayoristas con sus propios sistemas de facturación, es garantizar la homogeneidad y calidad de los datos. Cada mayorista maneja distintos criterios de codificación y categorización de clientes, productos y ventas, lo que exige un proceso riguroso de normalización. La diversidad en los formatos y estructuras de los datos puede generar inconsistencias, duplicaciones y errores si no se gestionan correctamente. Un ejemplo clave es la duplicidad de clientes: un mismo cliente final puede comprar a diferentes mayoristas, y cada uno puede asignarle una identificación distinta. Garantizar la unicidad de los clientes es esencial para un análisis preciso del comportamiento de compra, lo que a su vez permite identificar oportunidades de venta cruzada o personalización de ofertas.

Además, tras la facturación de la fábrica, los mayoristas pueden modificar productos o embalajes, lo que añade complejidad a la gestión de datos. Es necesario convertir esta información al lenguaje de la fábrica, creando equivalencias entre los códigos y descripciones de productos utilizados por los mayoristas y los propios de la fábrica. Sin esta conversión precisa, es probable que se cometan errores en la categorización de productos, lo que afectaría los análisis posteriores de ventas.

Otro reto importante es la obtención de datos detallados para el análisis geográfico. Los mayoristas deben proporcionar información con suficiente granularidad para permitir la segmentación por regiones, ciudades, localidades e incluso barrios. Este nivel de detalle es fundamental para identificar patrones de comportamiento por zona geográfica y ajustar las estrategias de distribución y marketing a las necesidades de cada mercado local. No obstante, la falta de estandarización entre los mayoristas respecto a la geolocalización de los clientes puede dificultar la implementación de una segmentación precisa.

Por lo tanto, un sistema de bases de datos eficiente debe integrar datos de diversas fuentes y gestionar la heterogeneidad de los mismos, asegurando su calidad para sustentar una toma de decisiones estratégica basada en información confiable. La clave es unificar, limpiar y estandarizar los datos, garantizando su unicidad y consistencia, además de traducirlos al lenguaje de la fábrica para hacer un análisis preciso de productos, clientes y zonas geográficas. Esto permitirá a la fábrica identificar oportunidades y amenazas en el mercado, optimizar la estrategia de *go to market* y fortalecer la relación con los mayoristas.

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (Primera entrega)

La base de datos de sell-out proviene de los datos de facturación suministrados por los mayoristas bajo acuerdos de confidencialidad. Esta información proporciona a la fábrica detalles clave sobre las ventas a clientes finales, permitiendo un análisis preciso del comportamiento de mercado. Los datos se estructuran para ofrecer insights geográficos (ciudades, departamentos, localidades, barrios) y jerárquicos del producto (categorías, subcategorías, materiales). Con esta base de datos, la fábrica puede tomar decisiones estratégicas, optimizando su estrategia de *go to market* y fortaleciendo la gestión comercial, a pesar de que el músculo logístico no es su principal fortaleza.

3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (Primera entrega)

Esta información es confidencial por consiguiente será usada con fines académicos únicamente.

3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? (Max 100 Palabras) - (Primera entrega)

Uno de los principales retos de la base de datos es asegurar la limpieza y consistencia de los datos, garantizando la unicidad e identificación precisa de los clientes finales. Dado que un cliente final puede comprar a múltiples mayoristas, y cada uno puede codificarlo de manera diferente, lo que complica la consolidación de la información. Además, es fundamental convertir los datos suministrados por los mayoristas al formato estándar de la fábrica, ya que, tras la facturación, pueden ocurrir variaciones en los productos o embalajes. Estos desafíos requieren un enfoque robusto en la normalización y estandarización de los datos para obtener análisis precisos.

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (Primera entrega)

El objetivo de utilizar un sistema de bases de datos en el proyecto es consolidar la información proveniente de diferentes mayoristas, lo que permitirá realizar un análisis profundo del mercado. A través de esta base de datos, se espera identificar oportunidades y amenazas mediante el análisis detallado de productos, segmentos de clientes y zonas geográficas. Esto facilitará la segmentación precisa del mercado, la detección de patrones de compra y la optimización de estrategias comerciales. Además, permitirá el acceso a datos confiables y actualizados, esenciales para la toma de decisiones informadas y para mejorar la competitividad en un entorno dinámico.

En la imagen 1 se muestra el modelo B2B, a partir del cual se pretende obtener el sistema manejador de bases de datos (SMBD).



The diagram illustrates the structure of the 'Puzzle' system, showing various classes and their relationships. The classes are organized into several groups:

- Variables:** Includes `Variables`, `Variables_1`, `Variables_2`, `Variables_3`, `Variables_4`, `Variables_5`, `Variables_6`, `Variables_7`, `Variables_8`, `Variables_9`, `Variables_10`, `Variables_11`, `Variables_12`, `Variables_13`, `Variables_14`, `Variables_15`, `Variables_16`, `Variables_17`, `Variables_18`, `Variables_19`, `Variables_20`, `Variables_21`, `Variables_22`, `Variables_23`, `Variables_24`, `Variables_25`, `Variables_26`, `Variables_27`, `Variables_28`, `Variables_29`, `Variables_30`, `Variables_31`, `Variables_32`, `Variables_33`, `Variables_34`, `Variables_35`, `Variables_36`, `Variables_37`, `Variables_38`, `Variables_39`, `Variables_40`, `Variables_41`, `Variables_42`, `Variables_43`, `Variables_44`, `Variables_45`, `Variables_46`, `Variables_47`, `Variables_48`, `Variables_49`, `Variables_50`, `Variables_51`, `Variables_52`, `Variables_53`, `Variables_54`, `Variables_55`, `Variables_56`, `Variables_57`, `Variables_58`, `Variables_59`, `Variables_60`, `Variables_61`, `Variables_62`, `Variables_63`, `Variables_64`, `Variables_65`, `Variables_66`, `Variables_67`, `Variables_68`, `Variables_69`, `Variables_70`, `Variables_71`, `Variables_72`, `Variables_73`, `Variables_74`, `Variables_75`, `Variables_76`, `Variables_77`, `Variables_78`, `Variables_79`, `Variables_80`, `Variables_81`, `Variables_82`, `Variables_83`, `Variables_84`, `Variables_85`, `Variables_86`, `Variables_87`, `Variables_88`, `Variables_89`, `Variables_90`, `Variables_91`, `Variables_92`, `Variables_93`, `Variables_94`, `Variables_95`, `Variables_96`, `Variables_97`, `Variables_98`, `Variables_99`, `Variables_100`.
- Matrices:** Includes `Matrices`, `Matrices_1`, `Matrices_2`, `Matrices_3`, `Matrices_4`, `Matrices_5`, `Matrices_6`, `Matrices_7`, `Matrices_8`, `Matrices_9`, `Matrices_10`, `Matrices_11`, `Matrices_12`, `Matrices_13`, `Matrices_14`, `Matrices_15`, `Matrices_16`, `Matrices_17`, `Matrices_18`, `Matrices_19`, `Matrices_20`, `Matrices_21`, `Matrices_22`, `Matrices_23`, `Matrices_24`, `Matrices_25`, `Matrices_26`, `Matrices_27`, `Matrices_28`, `Matrices_29`, `Matrices_30`, `Matrices_31`, `Matrices_32`, `Matrices_33`, `Matrices_34`, `Matrices_35`, `Matrices_36`, `Matrices_37`, `Matrices_38`, `Matrices_39`, `Matrices_40`, `Matrices_41`, `Matrices_42`, `Matrices_43`, `Matrices_44`, `Matrices_45`, `Matrices_46`, `Matrices_47`, `Matrices_48`, `Matrices_49`, `Matrices_50`, `Matrices_51`, `Matrices_52`, `Matrices_53`, `Matrices_54`, `Matrices_55`, `Matrices_56`, `Matrices_57`, `Matrices_58`, `Matrices_59`, `Matrices_60`, `Matrices_61`, `Matrices_62`, `Matrices_63`, `Matrices_64`, `Matrices_65`, `Matrices_66`, `Matrices_67`, `Matrices_68`, `Matrices_69`, `Matrices_70`, `Matrices_71`, `Matrices_72`, `Matrices_73`, `Matrices_74`, `Matrices_75`, `Matrices_76`, `Matrices_77`, `Matrices_78`, `Matrices_79`, `Matrices_80`, `Matrices_81`, `Matrices_82`, `Matrices_83`, `Matrices_84`, `Matrices_85`, `Matrices_86`, `Matrices_87`, `Matrices_88`, `Matrices_89`, `Matrices_90`, `Matrices_91`, `Matrices_92`, `Matrices_93`, `Matrices_94`, `Matrices_95`, `Matrices_96`, `Matrices_97`, `Matrices_98`, `Matrices_99`, `Matrices_100`.
- Datasets:** Includes `Datasets`, `Datasets_1`, `Datasets_2`, `Datasets_3`, `Datasets_4`, `Datasets_5`, `Datasets_6`, `Datasets_7`, `Datasets_8`, `Datasets_9`, `Datasets_10`, `Datasets_11`, `Datasets_12`, `Datasets_13`, `Datasets_14`, `Datasets_15`, `Datasets_16`, `Datasets_17`, `Datasets_18`, `Datasets_19`, `Datasets_20`, `Datasets_21`, `Datasets_22`, `Datasets_23`, `Datasets_24`, `Datasets_25`, `Datasets_26`, `Datasets_27`, `Datasets_28`, `Datasets_29`, `Datasets_30`, `Datasets_31`, `Datasets_32`, `Datasets_33`, `Datasets_34`, `Datasets_35`, `Datasets_36`, `Datasets_37`, `Datasets_38`, `Datasets_39`, `Datasets_40`, `Datasets_41`, `Datasets_42`, `Datasets_43`, `Datasets_44`, `Datasets_45`, `Datasets_46`, `Datasets_47`, `Datasets_48`, `Datasets_49`, `Datasets_50`, `Datasets_51`, `Datasets_52`, `Datasets_53`, `Datasets_54`, `Datasets_55`, `Datasets_56`, `Datasets_57`, `Datasets_58`, `Datasets_59`, `Datasets_60`, `Datasets_61`, `Datasets_62`, `Datasets_63`, `Datasets_64`, `Datasets_65`, `Datasets_66`, `Datasets_67`, `Datasets_68`, `Datasets_69`, `Datasets_70`, `Datasets_71`, `Datasets_72`, `Datasets_73`, `Datasets_74`, `Datasets_75`, `Datasets_76`, `Datasets_77`, `Datasets_78`, `Datasets_79`, `Datasets_80`, `Datasets_81`, `Datasets_82`, `Datasets_83`, `Datasets_84`, `Datasets_85`, `Datasets_86`, `Datasets_87`, `Datasets_88`, `Datasets_89`, `Datasets_90`, `Datasets_91`, `Datasets_92`, `Datasets_93`, `Datasets_94`, `Datasets_95`, `Datasets_96`, `Datasets_97`, `Datasets_98`, `Datasets_99`, `Datasets_100`.
- Puzzles:** Includes `Puzzle_1`, `Puzzle_2`, `Puzzle_3`, `Puzzle_4`, `Puzzle_5`, `Puzzle_6`, `Puzzle_7`, `Puzzle_8`, `Puzzle_9`, `Puzzle_10`, `Puzzle_11`, `Puzzle_12`, `Puzzle_13`, `Puzzle_14`, `Puzzle_15`, `Puzzle_16`, `Puzzle_17`, `Puzzle_18`, `Puzzle_19`, `Puzzle_20`, `Puzzle_21`, `Puzzle_22`, `Puzzle_23`, `Puzzle_24`, `Puzzle_25`, `Puzzle_26`, `Puzzle_27`, `Puzzle_28`, `Puzzle_29`, `Puzzle_30`, `Puzzle_31`, `Puzzle_32`, `Puzzle_33`, `Puzzle_34`, `Puzzle_35`, `Puzzle_36`, `Puzzle_37`, <

8

- 4.2 Diagrama modelo de datos (*Primera entrega*)
- 4.3 Imágenes de la Base de Datos (*Primera entrega*)
- 4.4 Código SQL - lenguaje de definición de datos (DDL) (*Primera entrega*)
- 4.5 Código SQL - Manipulación de datos (DML) (*Primera entrega*)
- 4.6 Código SQL + Resultados: Vistas (*Primera entrega*)
- 4.7 Código SQL + Resultados: Triggers (*Primera entrega*)
- 4.8 Código SQL + Resultados: Funciones (*Primera entrega*)
- 4.9 Código SQL + Resultados: procedimientos almacenados (*Primera entrega*)

VIERNES 18 DE OCTUBRE

- 5 Bases de Datos No-SQL (*Segunda entrega*)
 - 5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)
 - 5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (*Tercera entrega*)

6.1 Ejemplo de aplicación de ETL y Bodega de Datos (*Tercera entrega*)

6.2 Automatización de Datos (*Tercera entrega*)

6.3 Integración de Datos (*Tercera entrega*)

7 Proximos pasos (*Tercera entrega*)

8 Lecciones aprendidas (*Tercera entrega*)

9 Bibliografía