PROYECTO INGENIERIA DE DATOS

**EMPRESA INVERTRANS**

# INTEGRANTES:

Juan Nicolas Arévalo Lemus

William Alejandro Ramírez Montaña

Juan Sebastian Arroyo Zúñiga

# Bogotá D.C 25/02/2025

**RESUMEN EJECUTIVO**

Somos estudiantes de MACC e Ingeniería de Sistemas de la Universidad del Rosario. Nuestro proyecto de ingeniería de datos consiste en crear una base de datos relacional funcional SQL, así como una no relacional NoSQL para un cliente dueño de una empresa de transportes, hasta el día de hoy maneja toda la información relevante (facturas, conductores, historial de viajes) en un conjunto de hojas de Excel, lo cual es muy ineficiente ya que la información esta guardada en etiquetas no estandarizadas que hacen que sea muy difícil almacenar y acceder a la información de manera rápida y eficiente. Todo esto es muy perjudicial para el buen manejo de una empresa.

Es por eso que planeamos hacer las bases de datos -relacional y no relacional como alternativa- desde cero con los conceptos aprendidos durante todo el semestre en la clase de ingeniería de datos. Nuestro objetivo es brindarle al cliente una base de datos funcional al final del semestre para que reemplace el anterior sistema de manejo de información dentro de la empresa y mejore el funcionamiento de esta. La información que manejaremos será la misma de las hojas de Excel, haciendo un ETL (Combinar información de varias fuentes y convertirlas en un data set consistente) sobre esta para quitar variables repetidas y hacer la carga más manejable. Esto combinado con un workshop -un espacio de dialogo entre los ingenieros y el cliente en donde el cliente comparte que es lo que necesita y espera de la futura base de datos- son la base del trabajo. El proyecto será monitoreado por los propios miembros y la profesora haciendo uso de metodologías ágiles como Kanban para mejorar la carga de trabajo y la eficiencia.

**NECESIDAD Y CONTEXTUALIZACION DE LA NECESIDAD**

Nuestro cliente posee una empresa de transporte, cuenta con distintos vehículos y conductores con el fin de ofrecer viajes a distintas ciudades del país para transportar mercancía, todos los datos que ha obtenido la empresa a lo largo de los años se registraron en un Excel, lo cual no resulta eficiente puesto que, no se tiene un registro en tiempo real de la disponibilidad de conductores y vehículos, además, no se pueden realizar consultas como los clientes que son más representativos para la empresa o los viajes más lucrativos, lo que hace que la empresa no pueda aprovechar correctamente los datos que posee para tomar decisiones correctas.

**DEFINICIÓN DEL PROBLEMA**

El problema central radica en las limitaciones de Excel para gestionar datos dinámicos y complejos. Por un lado, actualizar campos como la disponibilidad de un vehículo o un conductor requiere un cambio manual cada que se realiza un viaje, lo que consume tiempo y es poco viable en la empresa de nuestro cliente, además, la falta de estandarización en el registro de gastos (como gasolina, peajes o hoteles) impide un control financiero efectivo, ya que no existen referencias de costos predefinidas que permitan detectar estafas por parte de los conductores

Otro aspecto preocupante es la imposibilidad de monitorear el desempeño de los recursos, ya que actualmente no existe la posibilidad de visualizar cuantas multas tiene un conductor o cuantas reparaciones ha tenido un vehículo en determinado periodo de tiempo.

# OBJETIVO GENERAL

Construir dos bases de datos una relación y otra no relacional que permita visualizar datos de forma sencilla para mejorar la toma de decisiones en una empresa de transporte

# OBJETIVO ESPECIFICO

* Construir un diagrama de clases para una base de datos relacional que permita entender como se gestionara la información
* Proporcionar una base de datos limpia y organizada que centralice la información de la empresa de transporte
* Garantizar la actualización automática de campos que varían con mucha frecuencia

ALTERNATIVAS DE SOLUCION

Soluciones similares existentes en la literatura y en el mercado.

Lista de posibles soluciones al problema

Las posibles soluciones debía ser una por cada integrante del equipo y debía ser un diagrama de flujo donde se explique cual es la solución al problema planteado.

///

**ALCANCE**

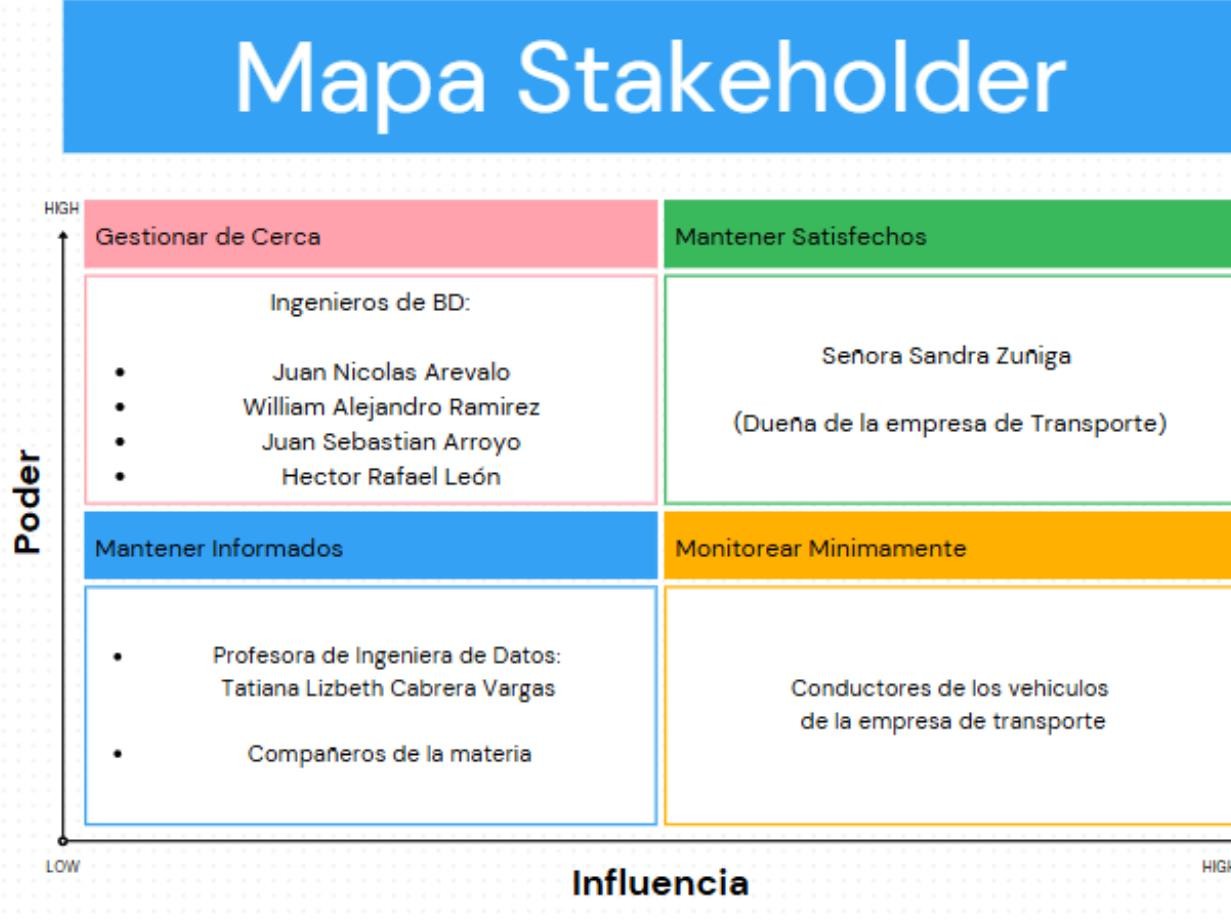
En el alcance es importante que se incluyan tiempos, cantidad de personas requeridas, tecnologías y herramientas a utilizar, así como lo que se pretende realizar en cada módulo a desarrollar. No hay un alcance definido.

**METODOLOGÍA**

Actualmente se esta trabajando con la metodología Kanban, la cual consiste en dividir las tareas en columnas que representan el estado de las mismas, en este caso backlog, en progreso, revisión y finalizadas, el enlace al tablero es el siguiente: [ing datos | Trello](https://trello.com/b/pqT8VsCh/ing-datos)

**Levantamiento de información y requerimientos funcionales a implementar**

… mil rqfs con las HU

**MAPA DE STAKEHOLDERS**

El mapa esta pensado de la siguiente manera: en prioridad alta ….

Modelamiento de base de datos SQL