# Take Home Case Study - Pre Tech Interview Data Engineer

## Objective

This document aims to provide a **Data Engineer** case study roughly based on experiences with previous clients @ Mutt Data. It can be used as a data systems design exercise for candidates in an interview process.

The candidate should study this document and be prepared during the technical interview to answer the questions listed below. Note that not all questions must be answered, candidates should prioritize the ones that seem most important to them.

## Problem statement

The client is a company that operates a goods and restaurant orders delivery service. In order to allocate the correct amount of drivers and runners resources needed to fulfill their client's deliveries, they need to have an estimation, a forecast, of the number of delivery requests they will have for each hour of the next 7 days, in all the geolocations that they operate in.

A priori we know that the client does not have a main repository from which it consumes historical data in a normalized and robust way for analysis, reporting or any form of dashboard. We know that it has three main data sources:

- A mid volume, transactional SQL database with customers, sales, inventory and other business events,
- Another high volume NoSQL with diverse structures and complex objects from the delivery platform, like runners and branches information,
- And finally, a high volume events oriented / streaming system where delivery orders and requests are generated.

Queries and analyses are performed in an ad-hoc manner by each analyst with their own criteria and tools to carry out some analysis and build business indicators. Most of the time those analyses are irreproducible.

Your task as a consultant is to design a system that the customer can make use of as a turnkey solution.

## Explain the solution

First describe the solution considering 3 very distinct level of details depending on the potential stakeholder:

- High level: This is about the CEO of the company or some other non-technical decision making person high in the hierarchyTT of the client's company.
- Middle level: This could be a technical VP. They can understand a block diagram but they don't have time for details.
- Low level: This is the implementation level of detail that a Tech Lead would be interested in.

## Going into more technical details

- Which parts would the system have?
    - What kind or combination of data storage would you use?

    - Why would you choose each one?
    - Would you need a system scheduling solution? Which one? Why?
- What specific other technologies would you recommend?

**Suggestion:** Candidates can bring a System Diagram to the interview if it helps explaining the different architectural components.

# Bonus Section

## Data storage design

Let's say a feature store layer is needed for the implementations of some Machine Learning pipelines .

- What non-functional requirements would you define?
- Are there any standards that should be considered?
- How would you monitor it?
- What tooling would you use?
- What metrics would you be interested in?

## Scaling

Let's say that the solution has been online for a year. The client comes to you again and says that the business grew 100% last year and the both data storage layer and data processing handling issues. They expect 1,000% growth next year so they need a scalable solution.

- What would you change of the original architecture?
- Would you replace one or more of the services or technologies used?
- How would you know that you succeeded?