# Grupo de Trabalho

## DEFINIÇÃO DE UM PROBLEMA (PODEM ESCOLHER "A GOSTO")

Vamos utilizar a base de dados da OLIST.

Cada csv será uma nova tabela (Imagem no Docker)

Passo 1: csv -> database -> s3

Passo 2: processamento (spark)

Passo 3: exploração (questionamentos, avaliações do dados) HIVE

Passo 4: Visualizações

## DEFINIÇÃO INGESTÃO EM DADOS (LOCAL DA ORIGEM, PODE JÁ EXISTIR OU PODE SER GERADO)

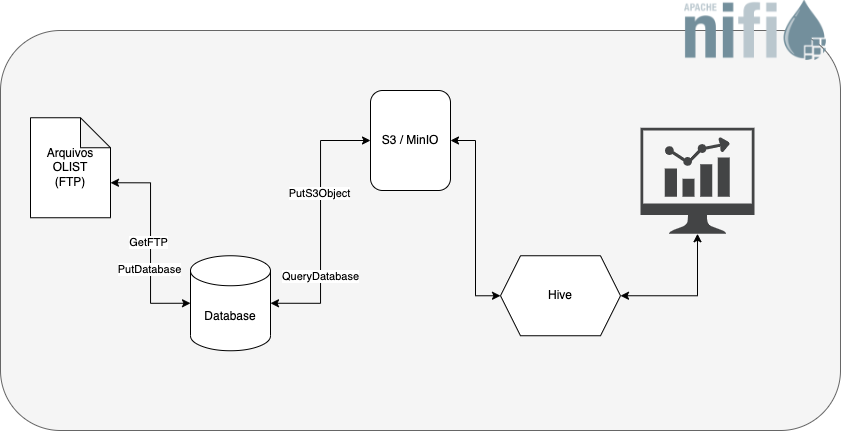
Origem dos dados será na S3/FTP/MinIO?

Processamento: Spark (agregações, etc)

Output para um banco de dados?

Dataviz?

## DEFINIÇÃO ARQUITETURA (DESENHO NO MÍNIMO DE 2 PARTES DO TODO INGESTÃO E ARMAZENAMENTO)



## IMAGEM DOCKER (SÓ SERÁ COBRADO QUANDO HOUVER A AULA QUARTA)

## 

## SUBIR NO GIT (SÓ SERÁ COBRADO QUANDO HOUVER A AULA QUARTA)