Grupo 2

Integrantes

- Amanda De Jesus Garcia
- Alter Vitor Soares Rodrigues
- Gabriela Monteiro Xavier
- Henrique Sales Kouyoumdjian
- Thiago Ruiz Aniceto

1) **Definição de um problema (Podem escolher “a gosto”)**

- **Contexto**: Somos uma empresa de e-commerce que possui vendas de vários produtos em diversas regiões no país e não tem um controle sobre os dados que são gerados de seus compradores.
- **Problema**: Identificar os melhores compradores por região e por produto em tempo real. Isso significa que a empresa precisa ter um controle eficiente sobre os dados, com o objetivo de identificar e acompanhar em tempo real os clientes que são considerados os melhores compradores em cada região e em relação a cada produto específico.

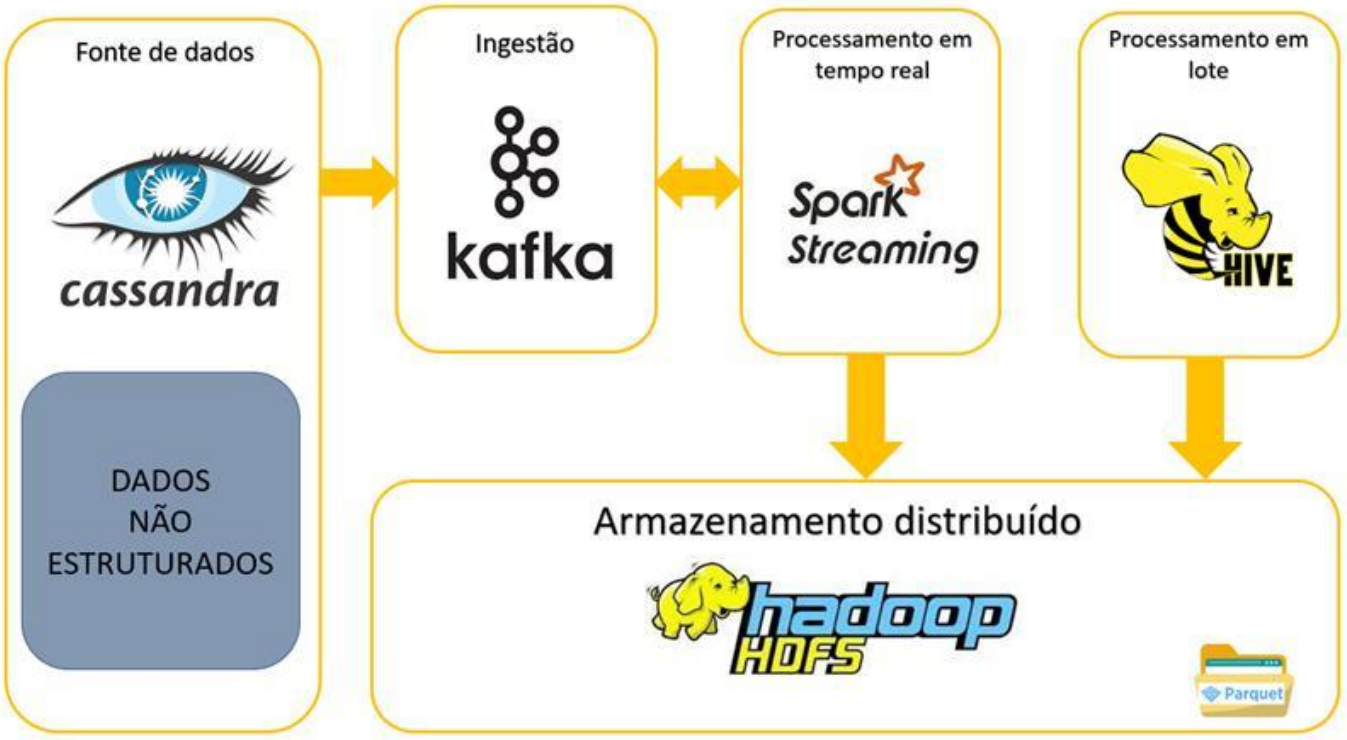
2) **Definição ingestão em dados (Local da origem, pode já existir ou pode ser gerado, qual ferramenta que vão usar)**

- **Apache Cassandra:** As transações de compra e venda são rastreadas e registradas no Apache Cassandra, que atua como **fonte de dados** do processo.
- **Kafka**: responsável por receber os dados do Cassandra e fazer a **ingestão** em tempo real. Ele atua como um intermediário entre as fontes externas e o processamento stream.
- **Spark Streaming**: **processa** os dados em tempo real da fonte Kafka e realiza as operações de transformação de dados, como filtragem, mapeamento e agregação por comprador, região e etc... Após esse

processamento, armazenamos os resultados no HDFS para fins de armazenamento seguro e durável.

- **HDFS**: O Spark Streaming envia os dados para o HDFS , que é um sistema de arquivos distribuído projetado para **armazenar** grandes volumes de dados, garantindo que os dados dos compradores sejam armazenados de forma segura. Utilizaremos o formato **Parquet.**
- **HIVE**: uma camada de consulta e análise de dados construída sobre o Hadoop que vai permitir **escrever** consultas para **extrair** informações dos dados armazenados no HDFS sobre os compradores.

3) **Definição arquitetura (desenho no mínimo 2 partes do todo ingestão e armazenamento como será)**



4) **Subir no GIT**