

O sistema público de transporte de São Paulo é utilizado por milhões de pessoas diariamente, e a eficiência no gerenciamento da frota de ônibus é essencial para garantir a qualidade do serviço. Para otimizar a operação do transporte público, é necessário implementar uma solução que permita o **monitoramento em near real-time** dos ônibus em circulação, além de fornecer **métricas e KPIs** relevantes para a tomada de decisão.

Objetivo: Criar um produto de dados em near real time ou batch que possibilite métricas, KPIs, monitoramento e acompanhamento dos ônibus em circulação no sistema público de transporte da cidade de São Paulo.

Entregáveis:

1. Desenho arquitetural da solução indicando as ferramentas utilizadas e fluxos propostos.
2. Explicação do por que de cada escolha realizada na arquitetura proposta.
3. Catálogo de metadados e documentação da aplicação.
4. Apresentação da solução proposta em perfeito funcionamento, com ingestão, processamento e entrega de dados, bem como o código fonte.
5. Apresentação dos entregáveis.

Montar apresentação:

1. Análise exploratória dos dados/documentação do data source.
2. Proposta do projeto.
3. Definição de arquitetura, o fluxo de dados e a explicação das escolhas de ferramentas.
4. Governança: Documentação da aplicação e um catálogo de metadados (pode ser uma planilha ou na própria apresentação)
5. Apresentar primeiro a parte teórica (apresentação) e depois mostrar funcionando. A aplicação tem que seguir a arquitetura do projeto.
6. Projeto funcionando deve conter: Ingestão de dados, armazenamento, processamento de dados e a camada de entrega (dataviz, api, etc...)
7. Entregar a apresentação e os códigos utilizados.

Especificação:

- Utilizar a API OLHO VIVO da SPTRANS para coletar os dados em near real time (a cada 2 minutos) da posição de todos os ônibus em circulação.
API: <https://www.sptrans.com.br/desenvolvedores>
ENDPOINTS:<https://www.sptrans.com.br/desenvolvedores/api-do-olho-vivo-guia-de-referencia/documentacao-api/>
- Utilizar o GTFS da SPTRANS para dados complementares (Dados Estáticos/Cadastrais).
- Enriquecer os dados de paradas com o endereço da localidade, através de latitude e longitude.
- Dados do GTFS: <https://gtfs.org/documentation/schedule/reference/>

- trips: Uma viagem é uma sequência de duas ou mais paradas que ocorrem durante um período de tempo específico.
 - stops: Paradas onde os veículos pegam ou deixam passageiros.
 - stop_times: Horários em que um veículo chega e parte das paradas para cada viagem.
 - shapes: Todos os trajetos no mapa. Regras para mapear caminhos de viagem de veículos, às vezes chamadas de alinhamentos de rotas.
 - routes: Uma rota é um grupo de viagens que são exibidas aos passageiros como um único serviço.
- Criar um data warehouse/data lake/data lakehouse com no mínimo 2 camadas de dados, onde os dados de entrada sejam armazenados da raw data e os dados tratados na camada trusted, etc...
- Deve ser aplicado regra de negócio para apresentação no produto de dados.
- Realizar discovery nas bases de dados para entender os possíveis cruzamentos de dados (API e GTFS) e as visualizações ou api que serão entregues como produto final.

OBS: Pode ser utilizado qualquer biblioteca open source ou base pública externa para enriquecimento e cruzamento de dados.

O que será avaliado:

- | |
|---|
| 1. Arquitetura e Design do Pipeline |
| 2. Implementação Técnica e Boas Práticas |
| 3. Governança, Qualidade e Documentação dos Dados |
| 4. Produto Analítico e Comunicação dos Resultados |
| 5. Contribuição Individual e Domínio Conceitual |

Grupo 1 – Apache Spark
 Cláudia dos Santos Silva
 Guilherme Reis Amorim Dias de Souza
 Renato Matsumoto
 Estela Maris Rodrigues

Grupo 2 – Data Lakehouse
 Luciano Rangel Pinheiro Neto
 Marília Naomi Nakayama
 Ambrósio Manuel Cosme

Grupo 3 – Machine Learning Ops
 Daniel Ramos Azevedo
 Rhaiza Rodrigues Seimoha
 Julio Cesar Frigo Silva

Grupo 4 – BigQuery Warriors
Isabella Piovan Fernandes
Lucas Silva do Carmo
Vitor Gabriel Vilar Silva

Grupo 5 – Data Mesh Council
Yuri Luis Pereira Knebel
André Luiz da Cruz Moreira
Erica Machado Lobato Gaya de Figueiredo
ROMULO PEDRO ARRIVABENE

Apresentação dos projetos: 21/01/2026