

Termo de Abertura do Projeto

Projeto 5A – Engenharia de Big Data: Coleta e Estruturação de Dados da API do Spotify

Pipeline de Engenharia de Dados para Análise Musical no Spotify

1. Contexto

A indústria musical moderna depende fortemente de dados para decisões estratégicas. O Spotify oferece uma API com informações detalhadas sobre artistas, álbuns, faixas e métricas de áudio. O desafio consiste em extrair, armazenar e transformar esses dados semi-estruturados (JSON) em um **modelo dimensional analítico**, capaz de sustentar análises sobre popularidade e padrões musicais.

2. Objetivo

Construir um **pipeline de dados** que consuma a API do Spotify, armazene respostas JSON em um banco **NoSQL**, realize o **ETL via Spark no Databricks** e gere um **modelo dimensional** com tabelas fato e dimensão para exploração analítica.

3. Escopo

Amostragem Inicial

Coleta de dados para artistas de destaque nos principais gêneros:
Pop (Taylor Swift, Dua Lipa),
Rock (Imagine Dragons, Arctic Monkeys),
Rap/Hip-Hop (Drake, Kendrick Lamar),
MPB/Pop BR (Anitta, Jão),
Eletrônico (Calvin Harris, David Guetta).

4. Modelo Dimensional

Tipo	Tabela	Descrição
Fato	Artista_Musica	Métricas de popularidade e desempenho
Dimensão	Dim_Artista	Dados descritivos do

		artista
Dimensão	Dim_Album	Dados descritivos do álbum
Dimensão	Dim_Musica	Dados descritivos da música

5. Critérios de Sucesso

- Integração funcional com a API do Spotify;
- Estruturação completa dos dados em modelo dimensional;
- Dataset analítico consultável e documentado.