Olá, que mostrar uma demonstração de um dashboard de dados, construido no dash e plotly obtidos usando apache airflow (python) e guardando dados em um datalake local.

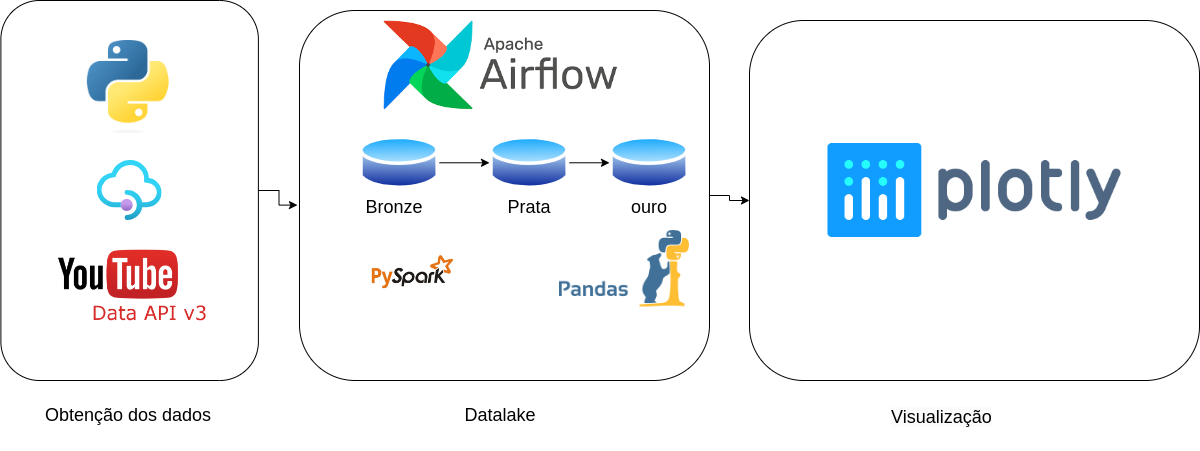
Requisitos da máquinas

Coloque aqui os requisitos

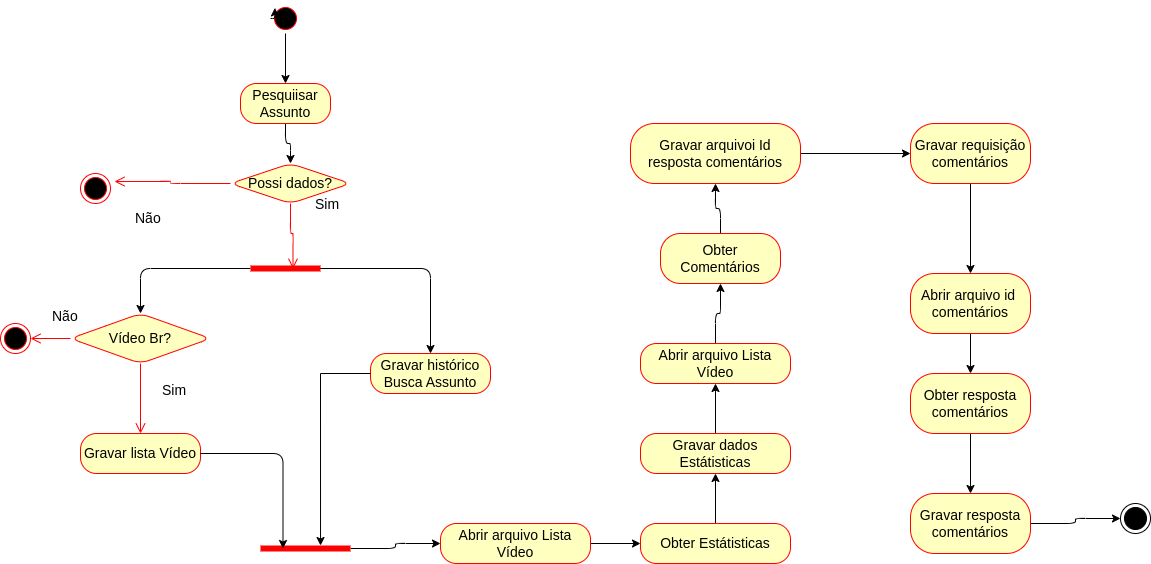
Conceitos Rápidos

* DAG: É uma representação do fluxo de trabalho, sem ciclos
* Operators: São classes no Airflow que representam uma tarefa específica.
* Hooks: São usados para fazer iterações entre os sistemas, seja API, banco de dados

A figura abaixo mostra a organização da extração dos dados da Api do youtube



A figura abaixo mostra uma proposta do diagrama de atividade para o fluxo da DAG



Em resumo, a DAG segue um padrão de execução sequencial. Umas das melhorias, seria a implementação, quando executar a extração de comentários, o vídeo deverá ter comentários antes de fazer a requisição da API.

**Diagrama de classe**

Uma das propostas do diagrama de classe, para a implementação da DAG, é implementação de uma classe base, que receba todos os parâmetros fixo da API, e cada endpoint, se tornaria uma classe herdando todos os atributos da classe mãe, mantendo seu comportamento especifico.

**Organização do dashboard**

O dashboard possui duas páginas divididas da seguinte maneira:

**Página Análise Assunto**

* Responsável por gerar as métricas de visualização de acordo com os assuntos selecionados. As métricas em questão são:
* Comparação do desempenho por dia do vídeo selecionando visualização, comentários e likes.
* Frequência dos vídeos publicados: Qual o dia da semana que possui o pico de vídeos publicados.
* Top 10 Vídeos com mais visualizações selecionando dia e comentários, visualizações e likes.
* Desempenho do canal por dia selecionando comentários, visualizações e likes.
* Desempenho do vídeo por dia selecionando comentários, visualizações e likes.
* Quantidade de palavras-chave no título.
* Taxa de engajamento por dia.
* Análise popularidade tags.

**Análise Trends**

* Top 10 Categorias Populares selecionando Visualizações, comentários e likes e dia.
* Top 10 Canais Populares selecionando Visualizações, comentários e likes e dia.
* Desempenho Categoria dia selecionando Visualizações, comentários e likes e dia.
* Top 10 Vídeo por Categoria.
* Top 10 Engajamento Canal.
* Top 10 Engajamento Vídeo.

Discussão das métricas

* Total de visualizações/comentários e likes por dia: A API do YouTube fornece esses dados de forma cumulativa. Para obter as visualizações/comentários e likes por dia, é necessário pegar o maior valor do dia e posteriormente subtrair o valor do dia atual ao do dia anterior.
* Taxa de engajamento: A taxa de engajamento é a quantidade de interações entre o público do canal e o conteúdo dele. A fórmula da taxa de engajamento é: Taxa de Engajamento = [(Número de Curtidas + Número de Comentários + Número de Compartilhamentos) / Número Total de Visualizações] \* 100

**Observações:**

Mesmo que o diagrama de atividade apresente, no final do processo, uma execução completa, uma proposta seria de implementar uma verificação de vídeos que contenham comentários antes de fazer todo o fluxo de comentários. Também pode ser desenvolvido, uma análise de inscritos que se inscreveram/desinscreveram por dia.