**Segunda Versão do projeto Arquitetura youtube, para a construção de um datalake**

**Introdução**

Esse é uma versão do outro projeto de extração de dados do youtube aonde construo um datalake com apache airflow, utilizando decoradores task, dag e taskgroup da biblioteca from airflow.decorators import task, dag, task\_group, e guardar os dados localmente.

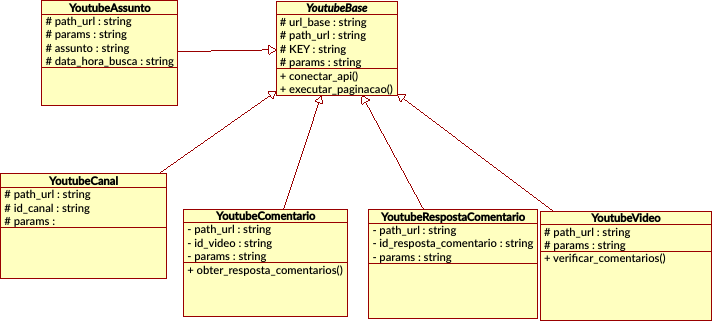
**Conceitos Rápidos**

* 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

**Atualizações**

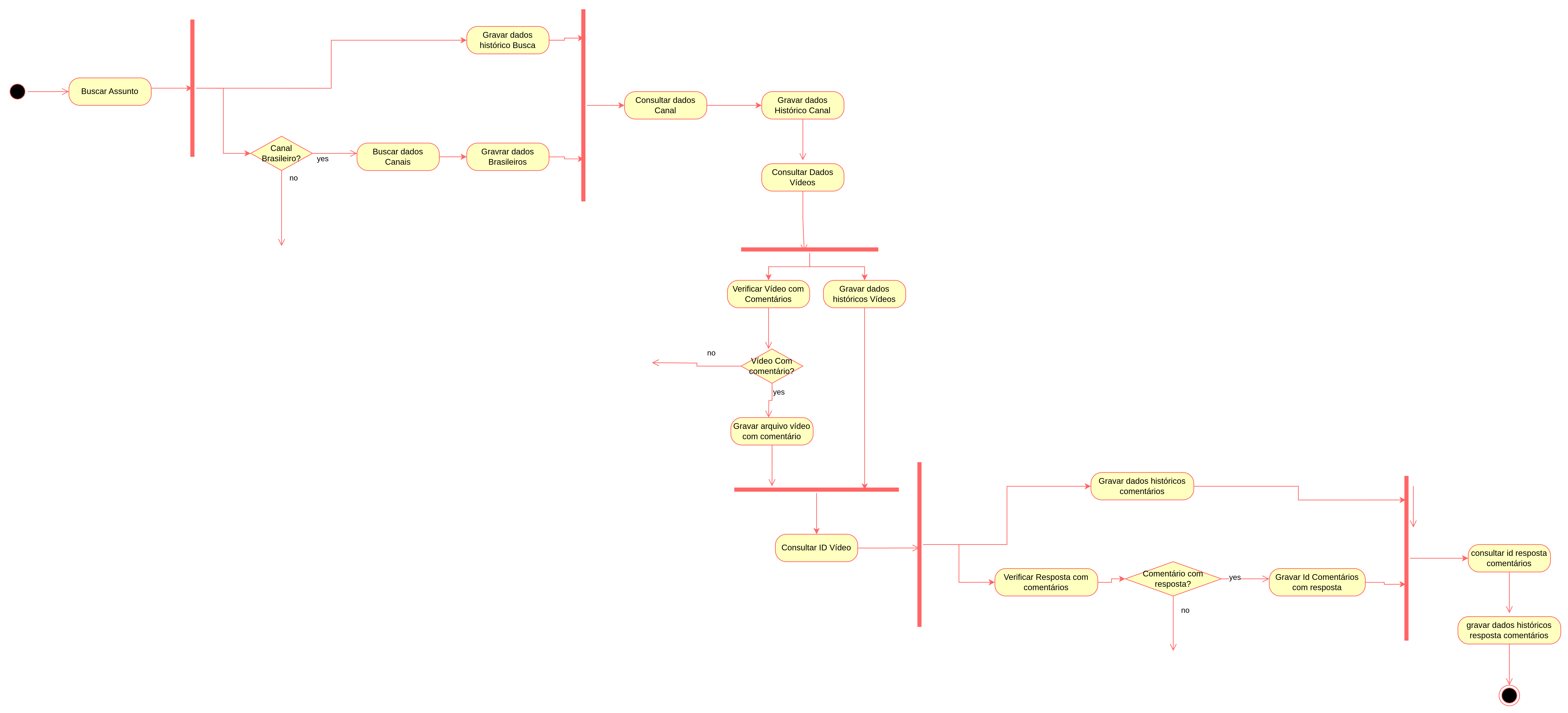
* Criação de arquivos separados para listar vídeos com comentários
* Verificação se o vídeo tem comentários
* Monitoramento de inscritos por Canal

**Diagrama de classe**



A figura abaixo mostra uma estrutura da chamada da API, temos uma classe, YoutubeBase, é uma classe abstrata, com os atributos base, como url, chave e params e o métodos comuns a todas as chamandas, executar\_api e executar\_paginacao.

**Diagrama de Atividade**



A figura abaixo mostra uma proposta de fluxo da DAG do apache airflow.

**Endpoints utilizados:**

**Pesquisar assunto do YouTube:**

**URL: /youtube/v3/search**

**Descrição:**URL para obter os vídeos por assunto. Salva todo o histórico de busca, independente do idioma do canal.

**Filtrar canais brasileiros:**

**URL:**/youtube/v3/channels

**Descrição:**Após salvar os registros, esta URL é usada para consultar o idioma do canal e gerar uma lista de vídeos dos canais.

**Obter dados dos vídeos:**

**URL:**/youtube/v3/videos

**Descrição:**Após obter a lista de vídeos brasileiros, esta URL é usada para obter os dados completos. Os campos viewCount, likeCount, favoriteCount, commentCount serão usados, sempre gravando os dados históricos.

**Extrair comentários:**

**URL:**/youtube/v3/commentThreads

**Descrição:**Depois de extrair a lista de vídeos, esta URL é executada para extrair os comentários, gravando a requisição original em um datalakes.

**Extrair respostas dos comentários:**

**URL:**/youtube/v3/comments

**Descrição:**Depois de extrair os comentários, cada comentário tem um ID que mostra se possui respostas. Esta URL é usada para extrair as respostas dos comentários.

