### 1. ****Visão Geral da Arquitetura****

* **FastAPI**: Servidor de aplicação que fornece endpoints para extração de dados.
* **Apache Airflow**: Orquestrador para gerenciar a extração e o processamento dos dados em intervalos regulares.
* **Azure Blob Storage**: Armazenamento dos arquivos CSV extraídos e backups.
* **Azure Data Lake**: Armazenamento dos dados estruturados e processamento adicional.

### 2. ****Componentes da Arquitetura****

#### **2.1 FastAPI**

* **Endpoints**:
  + /tabela\_producao/{ano}: Para extrair dados de produção.
  + /tabela\_processamento/{ano}/{filtro}: Para extrair dados de processamento com filtros.
  + /tabela\_comercializacao/{ano}: Para extrair dados de comercialização.
  + /tabela\_importacao/{ano}/{filtro}: Para extrair dados de importação com filtros.
  + /tabela\_exportacao/{ano}/{filtro}: Para extrair dados de exportação com filtros.
* **Lógica de Extração**:
  + A API tenta primeiro acessar os dados diretamente. Se a página não estiver disponível ou não contiver dados válidos, a API faz o download do arquivo do Azure Blob Storage.

#### **2.2 Apache Airflow**

* **DAGs (Directed Acyclic Graphs)**: Para orquestrar as tarefas de extração e carregamento.
  + **Tarefa de Extração**: Utiliza a API FastAPI para solicitar dados periodicamente.
  + **Tarefa de Transformação**: Processa os dados obtidos, se necessário.
  + **Tarefa de Carregamento**: Envia os dados para o Azure Data Lake.
* **Scheduler**: Configurado para executar as DAGs em intervalos programados (diário, semanal, etc.).

#### **2.3 Azure Blob Storage**

* **Armazenamento de Backups**: Os arquivos CSV são armazenados no Azure Blob Storage para recuperação em caso de falha na extração da API.

#### **2.4 Azure Data Lake**

* **Armazenamento de Dados**: Os dados processados são armazenados no Azure Data Lake para consultas e análises posteriores.

### 3. ****Fluxo de Dados****

1. **Inicialização**:
   * O Apache Airflow inicia a execução da DAG programada.
2. **Extração de Dados**:
   * A tarefa de extração chama os endpoints da FastAPI.
   * Se os dados não estiverem disponíveis, a API acessa o Azure Blob Storage para obter os arquivos CSV de backup.
3. **Processamento e Armazenamento**:
   * Os dados obtidos são processados conforme necessário.
   * Os dados processados são carregados no Azure Data Lake.
4. **Monitoramento**:
   * O Airflow monitora as execuções das tarefas e fornece logs e relatórios.

### 4. ****Implementação****

* **FastAPI**: A estrutura do seu código já está implementada. Certifique-se de que as variáveis de ambiente estejam corretamente configuradas.
* **Apache Airflow**: Crie uma DAG que utilize operadores para fazer chamadas HTTP à sua API e para armazenar os dados no Azure Data Lake.
* **Azure**: Configure sua conta do Azure Blob Storage e Data Lake para garantir que você tenha acesso e permissões necessárias.

### 5. ****Consumo****

* **IA: Posterior a API ser executada e carregar os arquivos dentro do Azure LakeZone, os dados são armazenados e carregados dentro da área de consumo que utiliza do PostgreSQL para armazenamento.**
* **Objetivo: O modelo de de IA é um modelo preditivo que cria uma tendência para vendas dos anos seguintes, trazendo tendência de mercado e quantidade estimada e valor vendido estimado.**

Imagem da Arquitetura:

