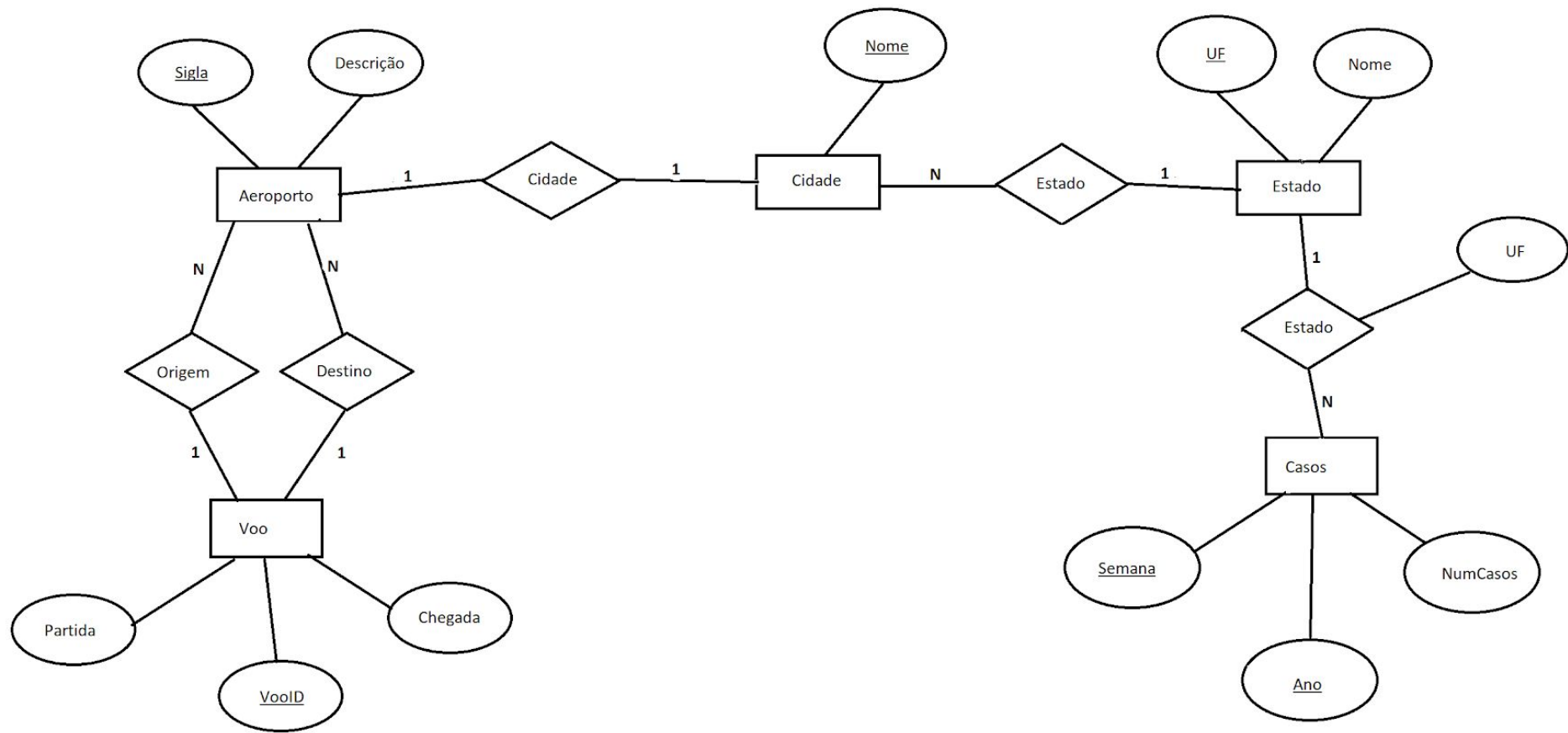


Estágio 04

Pedro Jun Novais(Ra:204678)
Thomas Gomes Ferreira(Ra:224919)

Modelo Conceitual



Dificuldades do estágio 3

Dificuldades

jupyter:

Não conseguimos utilizar o jupyter como plataforma de execução do código, tanto python como sql. Tivemos muitos problemas de execução e não conseguimos importar os dados direto do drive, assim, optamos por fazer localmente.

git:

O git não aceita arquivos únicos maiores que 25 MB, assim colocamos os dados no drive devido ao tamanho de alguns arquivos que facilmente ultrapassam o tamanho limite.

dados de clima:

Para a terceira fonte de dados, não conseguimos encontrar uma fonte de dados mais completa informando os dados climáticos. Mesmo após muita procura, ficamos sem encontrar, e decidimos deixar de lado, pelo menos temporariamente, focando nos outros dados.

Segundo Modelo Lógico

Grafo

- Simplifica o problema
- Facilita o manuseio dos dados
- Facilita a visualização dos dados
- Se encaixa muito bem para o csv de voos
- Fácil de se obter a partir da tabela

Metodologia

Preparando os dados

A partir da tabela de voos gerada anteriormente, iremos extrair as informações de rotas aéreas existentes, que representam as arestas dos grafos. Já considerando os nós, precisamos apenas dos nomes de cidades. Assim, obteremos um csv que pode ser carregado no cypher, gerando arestas e nós.

Outros dados relacionados na fase anterior, que não funcionam muito bem como grafo, serão preparados para que possam ser inseridos de alguma forma nessa etapa. Assim, queremos obter uma análise visual do problema, relacionando com as rotas aéreas.

Por exemplo, enquanto mostrar rotas entre dois estados, mostrar a quantidade de casos que existem nesses dois estados, para um período pré-determinado.

Justificativa e plataformas

Essa parte, iremos utilizar o modelo de grafos pois é mais adequado quando o foco vai estar nas rotas aéreas. Isso já havia sido pensado no começo do trabalho, em que obtemos as fontes de dados com modelos diferentes em mente. No caso, os dados de voos podem ser facilmente transformados em um grafo de rotas entre cidades.

Para preparar os dados, iremos utilizar scripts em python, obtendo os csvs que serão importados no neo4J/cypher. Caso alguma preparação mais complicada for necessária, iremos utilizar o orange, assim como na etapa anterior.

Para os grafos, iremos utilizar o neo4J/cypher, assim como foi trabalhado em aulas recentes.