Realizando a ingestão de dados de um bucket para Bigquery utilizando o dataflow para automatizar a carga de trabalho o inicio ao fim e logo após criar procudere para ser enviado a cada ingestão nova de dados que o bigquery receber.

1 – Criando um novo projeto na GCP.

Interface gráfica do usuário, Texto, Aplicativo, Tabela

Descrição gerada automaticamente

2 – Especificando o nome do novo projeto e selecionado uma organização para ficar atrelado a organização especifica.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

3 – Selecionando o novo projeto criado.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

4 – Criando o bucket pra inserir os dados e também os metadados para a transformação.

Cloud Storage -> Buckets

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

5 - Clicando em criar, nomear o buckets para o nome de sua escolha e clicar em criar.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

5 – Criando pastas dentro do buckets para armazenar dos dados, metadados e arquivos temporários.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

6 – Adicionando os arquivos nas pastas especificas.

Dados/ : Inserir o arquivo CSV que contém os dados da venda da hamburgueria

Metadados/ : Arquivos JSON que será o schema do banco de dados e o arquivo Javascripts que servirá para fazer a transformação dos dados

Temp/ -> Armazena os arquivos temporários

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

07 – Criação da tabela no BigQuery para receber os dados.

Interface gráfica do usuário, Aplicativo

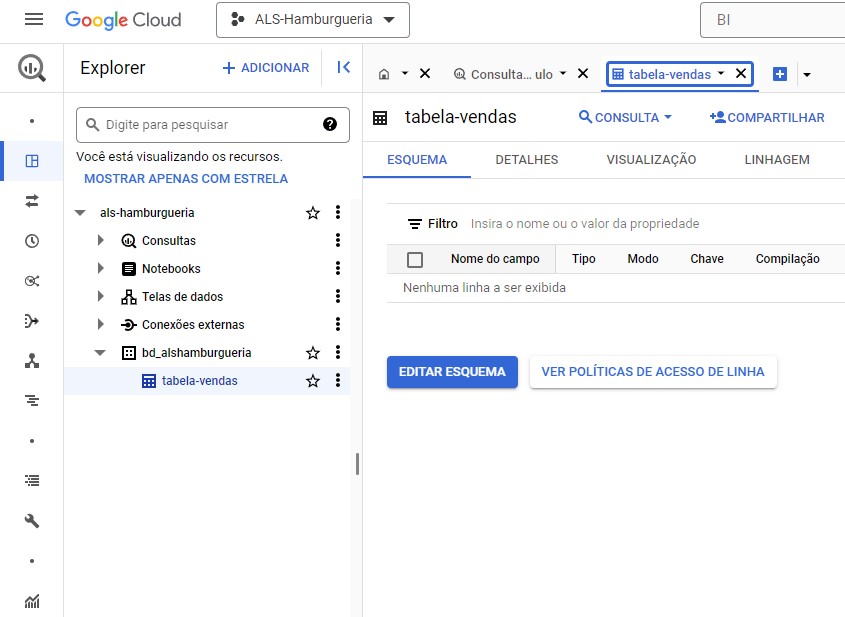
Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente



08 -Nessa etapa irei utilizar o dataflow para fazer a carga de trabalho na bigquery.

Acessando serviço

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

9 - Clicando em “criar job...”

Uma imagem contendo Texto

Descrição gerada automaticamente

10 – Definindo o nome e a região

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

11 – Selecionando o modelo do Dataflow – Text Files on cloud Storage to Bigquery

Interface gráfica do usuário, Texto, Aplicativo, Teams

Descrição gerada automaticamente

12 - Mas opções abaixo iremos localizar cada arquivo referente

Texto, Aplicativo, Email

Descrição gerada automaticamente

SOURCE -> Será os dados em CSV que iremos inserir na bigquery que está localizado no diretórios dados.

TARGET -> É o arquivo json , onde realizar toda transformação do csv e transformar os dados em um json.

Texto

Descrição gerada automaticamente

BIGQUERY OUTPUT TABLE -> A tabela no datawarehouse que irá receber os dados.

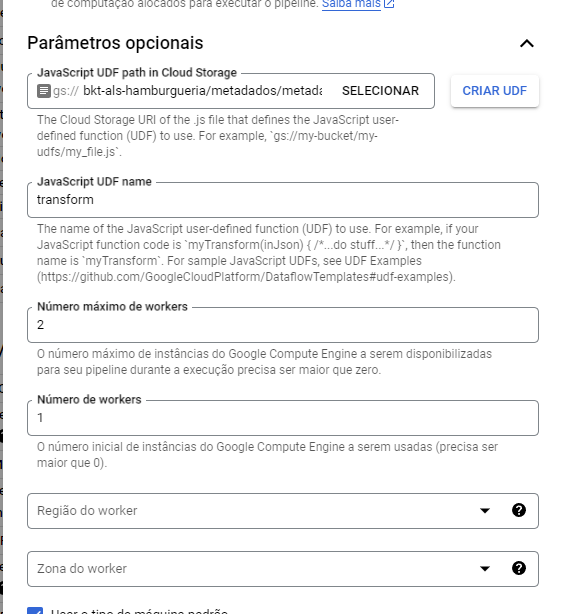
TEMPORARY -> Diretório para salvar os dados temporários que são processados no job

LOCAL TEMPORÁRIO -> Segue a mesma finalidade do caminho “ Temporary”

13 - Mais abaixo iremos inserir o arquivo Javascript que contém o schema da nossa tabela, configurar o numero de máximo e mínimo que workers que iremos utilizar.

Texto

Descrição gerada automaticamente



14 - Após essa configuração , podemos executar o job:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

15 - Ao termino, iremos ter essa imagem com todos os Jobs executados.

Diagrama

Descrição gerada automaticamente

16 - Com isso já possuímos os dados na tabela pronto para ser utilizado.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

17 - Realizando algumas consultas.

Interface gráfica do usuário

Descrição gerada automaticamente com confiança baixa

Interface gráfica do usuário, Aplicativo, Tabela

Descrição gerada automaticamente

Interface gráfica do usuário

Descrição gerada automaticamente

E também é possível utilizar o lookstudio utilizando essa base de dados e criar dashboard

Gráfico

Descrição gerada automaticamente.