

DATA ANALYTICS

BANCOS DE DADOS PARA BIG DATA

AULA 05

SUMÁRIO

O QUE VEM POR AÍ?	3
HANDS ON	4
SAIBA MAIS.....	5
O QUE VOCÊ VIU NESTA AULA?	9
REFERÊNCIAS.....	10

EMANIP

O QUE VEM POR AÍ?

Bem-vindos e bem-vindas ao material de aula sobre análise de dados geoespaciais dentro do ambiente Google BigQuery! Nesta jornada, exploraremos as ferramentas poderosas fornecidas pelo Google BigQuery para lidar com dados geoespaciais e descobrir insights valiosos sobre o mundo ao nosso redor.

Nos últimos anos, a quantidade de dados geoespaciais gerados aumentou exponencialmente, graças ao avanço da tecnologia e à crescente adoção de dispositivos móveis e sensores. No entanto, o desafio reside em como extrair informações significativas desses dados massivos e complexos.

É aqui que o Google BigQuery entra em cena como uma solução eficaz para lidar com grandes volumes de dados geoespaciais. Com ele, você pode armazenar, consultar e analisar facilmente dados geoespaciais em larga escala, aproveitando a estrutura flexível de armazenamento e a poderosa linguagem de consulta SQL.

Neste material de aula, iremos abordar desde os conceitos básicos de análise de dados geoespaciais até técnicas avançadas de consulta e visualização no Google BigQuery. Exploraremos recursos como o tipo de dados geográficos do BigQuery, funções espaciais, índices espaciais e consultas complexas envolvendo dados geoespaciais.

Além disso, aprenderemos a integrar o BigQuery com outras ferramentas do ecossistema do Google Cloud Platform, como o Google Maps Platform, para criar visualizações interativas e mapas geoespaciais impressionantes.

Prepare-se para mergulhar em uma jornada emocionante pela análise de dados geoespaciais usando o Google BigQuery. Vamos começar a descobrir os segredos ocultos em nossos dados geoespaciais e desbloquear novos insights que podem impulsionar decisões informadas e estratégias impactantes. Vamos lá!

HANDS ON

Teremos a oportunidade de entender melhor o que são dados geoespaciais e as suas finalidades, então que tal explorarmos ainda mais como trabalhar com estes dados dentro do ambiente do Google BigQuery? Vamos lá!

EMANAP

SAIBA MAIS

ENTENDENDO SOBRE DADOS GEOESPACIAIS

Dados espaciais são informações que possuem um componente geográfico ou de localização. Esses dados são associados a coordenadas geográficas, como latitude e longitude, e representam características ou eventos que ocorrem em um determinado espaço geográfico.

Os dados espaciais descrevem a localização, forma, tamanho e distribuição de objetos ou fenômenos na Terra ou em outros corpos celestes. Eles podem ser representados de várias formas, como pontos, linhas, polígonos ou superfícies.

Existem diferentes tipos de dados espaciais, incluindo:

1. Pontos: representam uma localização específica no espaço, como a posição de uma cidade, uma loja ou um sinal de GPS.
2. Linhas: representam uma sequência ordenada de pontos conectados, como estradas, rios ou rotas de voo.
3. Polígonos: representam uma área fechada definida por uma série de linhas, como fronteiras de países, estados ou regiões.
4. Superfícies: representam dados que cobrem uma área contínua, como modelos digitais de terreno, mapas de elevação ou imagens de satélite.

Os dados espaciais são amplamente utilizados em várias áreas, como geografia, cartografia, geologia, ciências ambientais, transporte, planejamento urbano, agricultura, monitoramento de recursos naturais, análise de negócios e muitas outras. Eles permitem visualizar, analisar e tomar decisões com base na dimensão geográfica, relacionando informações a locais específicos.

Ao trabalhar com dados espaciais, é possível realizar análises espaciais, como identificar padrões, calcular distâncias, realizar sobreposições de áreas, análise de proximidade, roteamento e muito mais. Essas análises podem fornecer insights valiosos para entender a distribuição espacial de eventos, padrões geográficos e relações entre diferentes fenômenos.

Com o avanço da tecnologia, os dados espaciais estão se tornando cada vez mais acessíveis e abundantes. Isso é impulsionado por fontes como sistemas de posicionamento global (GPS), imagens de satélite, sensores remotos, dispositivos móveis e redes de sensores. Esses avanços permitem uma melhor compreensão do mundo em que vivemos e o uso efetivo de dados geoespaciais para tomar decisões informadas em várias áreas de aplicação.

TRABALHANDO COM DADOS GEOESPACIAIS DENTRO DO GOOGLE BIGQUERY

Este laboratório usa um conjunto de dados disponível por meio do Google Cloud Public Dataset Program. Um conjunto de dados público é qualquer conjunto de dados armazenado no BigQuery e disponibilizado ao público em geral. Os conjuntos de dados públicos são conjuntos de dados que o BigQuery hospeda para você acessar e integrar em seus aplicativos. O Google paga pelo armazenamento desses conjuntos de dados e fornece acesso público aos dados por meio de um projeto. Você paga apenas pelas consultas que realiza nos dados (o primeiro 1TB por mês é gratuito, sujeito aos detalhes de preço da consulta).

O Citi Bike é o maior programa de compartilhamento de bicicletas do país, com 10.000 bicicletas e 600 estações em Manhattan, Brooklyn, Queens e Jersey City. O conjunto de dados que usamos aqui inclui viagens diárias do Citi Bike desde o lançamento do Citi Bike em setembro de 2013. Os dados foram processados pelo Citi Bike para remover viagens feitas pela equipe para manutenção e inspeção do sistema, bem como quaisquer viagens abaixo de 60 segundos em length, que são considerados inícios falsos.

EXEMPLOS DE ALGUMAS LINHAS DE DADOS

Você pode começar a explorar esses dados no console do BigQuery visualizando os detalhes da citibike_stationstabela.

Abra a IU da Web do BigQuery no Console do Google Cloud, selecione Menu de navegação (☰) > BigQuery.

A caixa de mensagem “Welcome to BigQuery in the Cloud Console” é aberta. Essa caixa de mensagem fornece um link para o guia de início rápido e lista as atualizações da interface do usuário.

Clique em “Concluído”.

Consulte algumas linhas de bigquery-public-data.new_york_citibike.citibike_stations para entender os dados armazenados na tabela. Adicione a seguinte consulta à área de texto do editor de consultas:

```
SELECT *
FROM `bigquery-public-data.new_york_citibike.citibike_stations`
LIMIT 10
```

Query results										
JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW				
Row	station_id	name	short_name	latitude	longitude	region_id	rental_methods	capacity	eightd	num_bikes_avai
1	258	DeKalb Ave & Vanderbilt Ave	4461.04	40.68940747	-73.96885458	71	CREDITCARD, KEY	3	fal...	2
2	261	Johnson St & Gold St	4668.08	40.69474881	-73.98362464	71	CREDITCARD, KEY	0	fal...	0
3	305	E 58 St & 3 Ave	6762.02	40.76095756	-73.96724467	71	CREDITCARD, KEY	0	fal...	0
4	383	Greenwich Ave & Charles St	5914.08	40.735238	-74.000271	71	CREDITCARD, KEY	39	fal...	0
5	403	E 2 St & 2 Ave	5593.02	40.72502876	-73.99069656	71	CREDITCARD, KEY	0	fal...	0
6	3085	Roebling St & N 4 St	5267.09	40.71469	-73.95739	71	CREDITCARD, KEY	25	fal...	25
7	3088	Union Ave & Jackson St	5300.06	40.7160751	-73.952029	71	CREDITCARD, KEY	19	fal...	19
8	3171	Amsterdam Ave & W 82 St	7360.10	40.78524672	-73.97667321	71	CREDITCARD, KEY	0	fal...	0
9	3201	Dey St	JC065	40.737711	-74.066921	70	CREDITCARD, KEY	18	fal...	18
10	3207	Oakland Ave	JC022	40.7376037	-74.0524783	70	CREDITCARD, KEY	26	fal...	24

Figura 1 – BigQuery

Fonte: elaborada pelo autor (2023), adaptada por FIAP (2023)

Três colunas nesta tabela são relevantes para este laboratório:

1. **longitude:** a longitude de uma estação. Os valores são longitudes WGS 84 válidas no formato de graus decimais.
2. **latitude:** a latitude de uma estação. Os valores são latitudes WGS 84 válidas no formato de graus decimais.
3. **num_bikes_available:** o número de bicicletas disponíveis para locação.

Encontre as estações com mais de 30 bicicletas disponíveis e, em seguida, execute uma consulta SQL padrão que encontre todas as estações Citi Bike na cidade de Nova York com mais de 30 bicicletas disponíveis para aluguel.

1. A consulta SQL padrão a seguir é usada para localizar as estações Citi Bike com mais de 30 bicicletas. Adicione esta consulta à área de texto do editor de consultas:

```
SELECT
```

```
  ST_GeogPoint(longitude, latitude) AS WKT,
```

```
  num_bikes_available
```

```
FROM `bigquery-public-data.new_york.citibike_stations`
```

```
WHERE num_bikes_available > 30
```

As cláusulas de consulta fazem o seguinte:

- `SELECT ST_GeogPoint(longitude, latitude) AS WKT, num_bikes_available` A `SELECT` cláusula seleciona a `num_bikes_available` coluna e usa a `ST_GeogPoint` - função para converter os valores nas colunas de latitude e longitude em `GEOGRAPHY` tipos (pontos).
- `FROM bigquery-public-data.new_york.citibike_stations` A `FROM` -cláusula especifica a tabela que está sendo consultada: `citibike_stations`.
- `WHERE num_bikes_available > 30` A `WHERE` - cláusula filtra os valores na `num_bikes_available` coluna apenas para as estações com mais de 30 bicicletas.

2. Clique no botão **Executar** para executar a consulta.

A consulta leva um momento para ser concluída. Após a execução da consulta, seus resultados aparecem no painel Resultados da consulta.

O QUE VOCÊ VIU NESTA AULA?

Nesta aula, tivemos a oportunidade de aplicar de maneira prática os conceitos de dados geoespaciais dentro do Google BigQueryBigQuery e entender como dados geográficos podem gerar valor através de manipulação com a linguagem SQL.

Não se esqueça de entrar na comunidade do curso no Discord! Lá você pode sanar dúvidas, interagir com colegas, receber avisos e muito mais.

EMANDA

REFERÊNCIAS

Getting Started with BigQuery GIS for Data Analysts, Google Cloud Skills

Boost, [s.d.]. Disponível em:

<https://www.cloudskillsboost.google/focuses/17817?locale=pt_BR&parent=catalog>

. Acesso em: 20 jun 2023.

Introduction to geospatial analytics, [s.d.]. Disponível em:

<<https://cloud.google.com/bigquery/docs/geospatial-data>>. Acesso em: 20 jun 2023.

JETHVA, HITESH. **Working with BigQuery GIS: 5 Useful Functions & Tools**,

2021. Disponível em: <<https://hevodata.com/learn/bigquery-gis/>>. Acesso em: 20 jun 2023.

PALAVRAS-CHAVE

Palavras-chave: GEOESPACIAL, BIGQUERY, GOOGLE CLOUD, SQL, DADOS, PYTHON.

EMENDAS

The background is a dark blue field filled with numerous small, light blue dots. Overlaid on this are several large, wavy, translucent lines in shades of blue and yellow. A vertical line with a small 'x' at the bottom is on the left. A circle containing the number '7' is in the upper center. A hexagon is in the lower right.

POSTECH