

Introdução a Dados Estruturados

1. O que São Dados Estruturados?

Definição: Dados estruturados são dados organizados em um formato predefinido, geralmente em forma de tabelas. Eles aderem ao **Modelo Relacional**, utilizando SQL (Structured Query Language) para manipulação e consulta.

Importância:

- A estrutura rígida dos dados facilita consultas e operações, permitindo que engenheiros de dados realizem análises precisas e rápidas.

Exemplo de Criação de Tabela em BigQuery:

```
CREATE TABLE `seu_projeto.meu_dataset.clientes` (  
  cliente_id INT64,  
  nome STRING,  
  email STRING  
);
```

2. Componentes Fundamentais dos Dados Estruturados

Componentes Básicos:

1. **Tabela:**
 - Um conjunto de linhas que representam uma entidade. Exemplo: tabela de **clientes**.
2. **Coluna:**
 - Cada coluna representa um tipo de dado específico que faz parte do registro. Exemplos de colunas em uma tabela de **clientes** podem incluir **cliente_id**, **nome**, **email**.
3. **Linha:**
 - Uma linha consiste em um conjunto de valores que descrevem um único registro. Por exemplo, uma linha pode conter os dados de um cliente específico.

Exemplo SQL: Para visualizar os dados em uma tabela, você pode usar:

```
SELECT * FROM `seu_projeto.meu_dataset.clientes`;
```

3. Tipos de Dados Estruturados no BigQuery

Dentro do BigQuery, os tipos de dados estruturados incluem:

- **STRING:** Para texto (equivalente ao tipo VARCHAR em outros bancos, embora o BigQuery utilize STRING para textos).

- **INT64:** Para números inteiros.
- **FLOAT64:** Para números de ponto flutuante.
- **BOOLEAN:** Para valores verdadeiro/falso.
- **DATE/TIMESTAMP:** Para armazenar informações relacionadas a datas e horas.

Observação: BigQuery não aceita o tipo VARCHAR como em outros sistemas, mas a utilização de STRING funcionará da mesma forma para armazenar textos.

```
SELECT * FROM `seu_projeto.meu_dataset.INFORMATION_SCHEMA.COLUMNS`  
WHERE TABLE_NAME = "clientes";
```

5. Comparação com Outros Tipos de Dados

Dados Não Estruturados:

- Exemplos incluem documentos, emails, conteúdos multimídia. Esses dados não têm uma organização rígida, o que dificulta consultas diretas.

Dados Semiestruturados:

- Possuem uma estrutura mais solta e podem incluir elementos como JSON ou XML. Eles não se encaixam nos modelos tradicionais, mas têm algum nível de consistência.

6. Aplicação Prática: Utilizando BigQuery

Analisando Estrutura de Tabelas:

- Os alunos devem usar o **BigQuery** para criar uma tabela de clientes, adicionando dados conforme o exemplo anterior.

Ação Prática:

1. Crie uma tabela chamada **clientes** no BigQuery.
2. Insira alguns dados fictícios usando o SQL.
3. Execute uma consulta para visualizar os dados inseridos.

Exemplo de SQL para Inserção:

```
INSERT INTO `seu_projeto.meu_dataset.clientes` (cliente_id, nome, email)  
VALUES  
(1, 'José da Silva', 'jose@email.com'),  
(2, 'Maria Oliveira', 'maria@email.com');
```