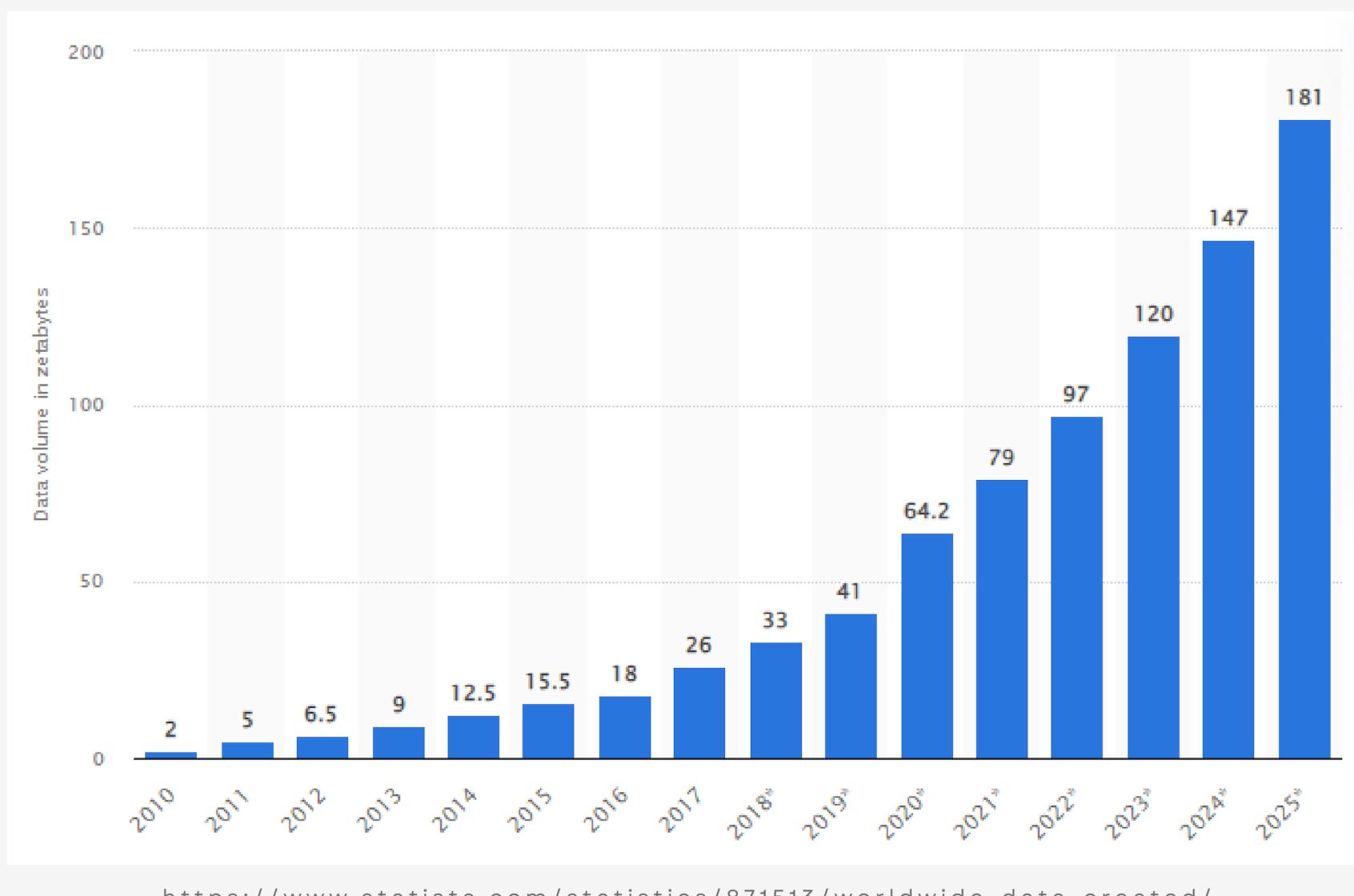


SQL PARA DATA SCIENCE

UMA INTRODUÇÃO

PRODUÇÃO DE DADOS NO MUNDO E SQL

O mundo tem passado por grandes mudanças atualmente e a produção de dados é um ponto crucial nesse processo de mudança em todos os campos de estudo. Estamos presenciando uma etapa da indústria onde a produção de conteúdo tem se tornado cada vez mais rápida, em uma escala nunca vista antes e demandando habilidades multidisciplinares para lidar com esse processo e manter vivo no mercado.



<https://www.statista.com/statistics/871513/worldwide-data-created/>

Como mostrado na figura acima, ano após ano a produção de dados vem aumentando. Segundo o gráfico em 2025 teremos 181 zetabytes de dados criados, capturados, copiados e consumidos globalmente. Nesse sentido o SQL se torna importante, pois, boa parte das indústrias armazenam dados em bases relacionais, isso quer dizer que possivelmente em sua trajetória profissional você terá que lidar com um banco de dados relacional para extrair insights, construir relatórios, painéis e /ou dashboards para seu cliente.

O QUE É SQL?

SQL é um acrônimo para Structered Query Language e foi desenvolvida com intuito de facilitar a consulta e gerenciamento de bancos de dados relacionais. Com o conhecimento adequado é possível realizar tarefas como criação, manutenção e consulta de dados em bases relacionais. Ainda usando SQL, podemos realizar transações como inserção, atualização, remoção, modificação de dados com comandos simples.

Existem algumas razões do motivo pelo qual o SQL é popular, não só em Ciência de Dados, mas em diversas áreas da computação. Isso porque a sintaxe é simples e é muito similar ao idioma inglês já que suas instruções são simples palavras em inglês. Além disso, como se tornou um padrão, diversos bancos de dados a utilizam como por exemplo:

- MySQL;
- Postgre
- SQL Server
- Oracle e muitos outros.

COMANDOS SQL MAIS UTILIZADOS

REALIZANDO CÁLCULOS COM SQL

```
SELECT 5200 + 13;
```

Realizando um único cálculo

```
SELECT 5200 + 13, 35 - 3, 80 * 2, 70/30;
```

Realizando múltiplos cálculos

```
SELECT 30+1+80, 50+30*80;
```

Realizando cálculos com múltiplos dígitos

```
SELECT 50*30 AS MULTIPLICACAO, 10+10+20 AS SOMA;
```

Renomeando cálculos

SELECIONANDO TABELAS, COLUNAS E LINHAS

```
SELECT * FROM TABELA;
```

Selecionando todos os valores de uma tabela

```
SELECT COLUNA_1, COLUNA_2, COLUNA_3 FROM TABELA;
```

Selecionando valores específicos de uma tabela

```
SELECT COLUNA_1, COLUNA_2, COLUNA_3 FROM TABELA LIMIT 10;
```

Selecionando apenas os 10 primeiros valores de uma tabela com colunas específicas.

COMANDOS SQL INTERMEDIÁRIOS

JOIN E CONSULTAS COMPLEXAS

```
SELECT COLUNA_1, COLUNA_2 FROM TABELA  
INNER JOIN TABELA_2 ON TABELA_1.COLUNA_1 = TABELA_2.COLUNA_1
```

Juntando tabelas com comando INNER JOIN

```
SELECT * FROM TABELA  
LEFT JOIN TABELA_2 ON TABELA_1.COLUNA_1 = TABELA_2.COLUNA_1
```

Juntando tabelas com comando LEFT JOIN

```
SELECT * FROM TABELA  
RIGHT JOIN TABELA_2 ON TABELA_1.COLUNA_1 = TABELA_2.COLUNA_1
```

Juntando tabelas com comando RIGHT JOIN

```
SELECT COLUNA_1, COLUNA_2 FROM TABELA  
FULL OUTER JOIN TABELA_2 ON TABELA_1.COLUNA_1 = TABELA_2.COLUNA_1
```

Juntando tabelas com comando FULL OUTER JOIN

```
SELECT COLUNA_1, COLUNA_2 FROM TABELA_1  
INNER JOIN (SELECT * FROM TABELA_2) TABELA_3 ON TABELA_1.COLUNA_1 =  
TABELA_2.COLUNA_1
```

Juntando tabelas usando subquery

```
SELECT TAB_1.COL_1, TAB_2.COL_1, TAB_3.COL_2 FROM TAB_1  
INNER JOIN TAB_2 ON TAB_1.COL_1 = TAB_2.COL_1  
LEFT JOIN TAB_3 ON TAB_1.COL_1 = TAB_3.COL_2
```

Juntando mais que duas tabelas

COMANDOS SQL AVANÇADOS

1

```
WITH DEPT_COUNT AS (
  SELECT DEPTNO, COUNT(*) AS DEPT_COUNT
  FROM   EMP
  GROUP BY DEPTNO)
SELECT E.ENAME AS EMPLOYEE_NAME,
       DC.DEPT_COUNT AS EMP_DEPT_COUNT
  FROM   EMP E
         JOIN DEPT_COUNT DC ON E.DEPTNO = DC.DEPTNO;
```

Otimização de consultas complexas, redução do uso dos recursos do banco de dados

2

```
SELECT DEPARTMENT_NAME, EMPLOYEE_NAME
  FROM  DEPARTMENTS D
        CROSS JOIN LATERAL (SELECT EMPLOYEE_NAME
                               FROM   EMPLOYEES E
                               WHERE  E.DEPARTMENT_ID = D.DEPARTMENT_ID)
 ORDER BY 1, 2;
```

Correlação esquerda dentro das visualizações em linha.

1 <https://oracle-base.com/articles/misc/with-clause>

2 <https://oracle-base.com/articles/12c/lateral-inline-views-cross-apply-and-outer-apply-joins-12cr1>