



O que estudar para se  
tornar um Engenheiro de  
Dados em Cloud?





O que estudar para se  
tornar um Engenheiro(a)  
de Dados em Cloud?



O que  
estudar?



---

As Linguagens Mais Usadas

---

Linguagem de Programação(Python)

---

Linguagem SQL

---

Bancos de Dados (Relacional e NoSQL)

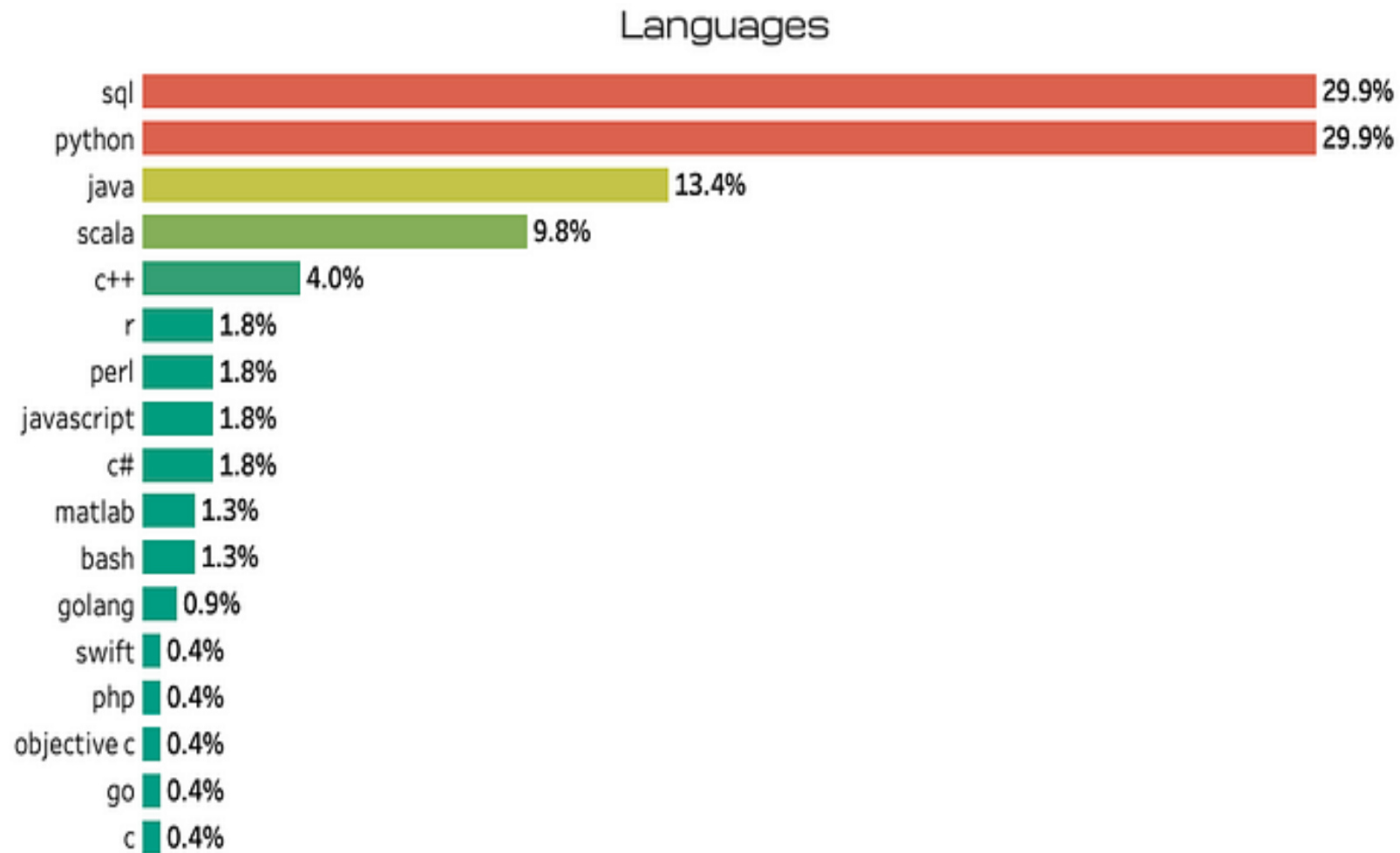
---

Modelagem de Dados

---

Provedores de Cloud

# As linguagens mais usadas



Fonte: Imagens da internet

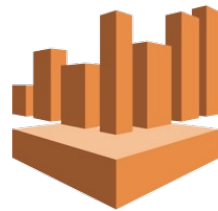
Fonte: <https://blog.devgenius.io/become-a-data-engineer-in-2022-analysis-of-over-1-000-faang-job-postings-38784fa727a8>



# Linguagem SQL



SQL (Structured Query Language) é uma linguagem indispensável na engenharia de dados, sendo utilizada para consulta e manipulação de banco de dados relacionais.



**amazon**  
ATHENA



**databricks**

# Linguagem de Programação(Python)



Como foi visto, Python é uma das linguagens mais utilizadas na engenharia de dados, esse sucesso é muito por sua facilidade de uso e grande diversidade de bibliotecas disponíveis.



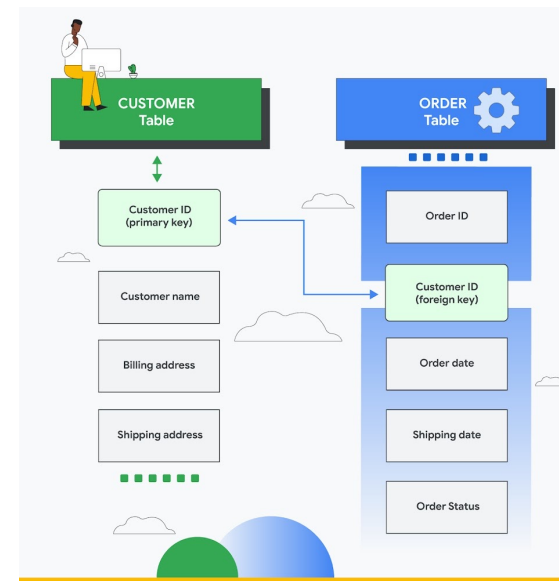
*NumPy*



# Bancos de Dados (Relacional e NoSQL)



Um banco de dados relacional é um conjunto de informações que organiza dados em relações predefinidas, em que os dados são armazenados em uma ou mais tabelas (ou "relações") de colunas e linhas, facilitando a visualização e a compreensão de como as diferentes estruturas de dados se relacionam. Desenvolvido por Edgar Frank Codd da IBM na década de 1970.



Fonte: Imagens da internet

Fonte: <https://cloud.google.com/learn/what-is-a-relational-database?hl=pt-br>

# Bancos de Dados (Relacional e NoSQL)



O NoSQL é uma tecnologia de banco de dados que armazena dados em esquemas flexíveis que escalam facilmente. Durante décadas, o modelo de dados predominante no desenvolvimento de aplicações foi o modelo de dados relacional que armazenava dados em tabelas feitas de linhas e colunas. O termo NoSQL foi primeiramente utilizado em 1998 como o nome de um banco de dados de código aberto que não possuía uma interface SQL. Seu autor, Carlo Strozzi.

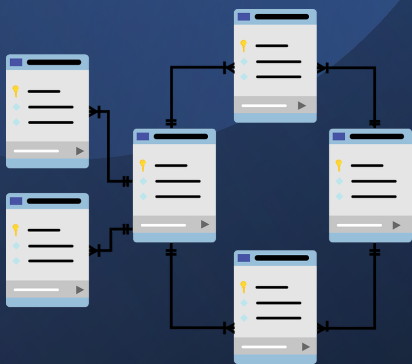
Grafos	Documentos	Chave-Valor	Colunar
Neo4J	MongoDB	Redis	Cassandra
Amazon Neptune	Amazon DocumentDB	Amazon DynamoDB	Amazon Keyspaces
Azure CosmosDB	Azure CosmosDB	Azure CosmosDB	Azure CosmosDB
GCP Knowledge Graph	GCP Cloud Firestore	GCP Bigtable	GCP Bigtable

Fonte: Imagens da internet

Fonte: <https://aws.amazon.com/pt/nosql/>, <https://cloud.google.com/learn/what-is-a-relational-database?hl=pt-br>, <https://db-engines.com/en/ranking>



# Modelagem de Dados



Captura → Processamento → Armazenamento → Consumo

A modelagem tem como objetivo facilitar o uso e gestão dos dados. Criar um modelo de dados é uma fase fundamental para design de bancos de dados.

Modelo Conceitual → Modelo Lógico → Modelo Físico

# Provedores de Cloud



Google Cloud

# Provedores de Cloud

A AWS, lançada pela Amazon em 2006, é frequentemente considerada a pioneira e líder no mercado de serviços em nuvem. Ela oferece uma ampla gama de serviços, desde armazenamento simples até recursos avançados de aprendizado de máquina. Sua extensa rede global de data centers permite alcançar uma latência mínima em todo o mundo.

O Microsoft Azure entrou no mercado em 2010 e rapidamente se tornou um concorrente sólido. A vantagem da Microsoft reside em sua integração com produtos amplamente utilizados, como o Windows Server e o Office 365.

O Google Cloud entrou na competição em 2011, trazendo a experiência do Google em infraestrutura de escala web. Ele se destaca por suas soluções de big data e machine learning.

Compare os serviços da AWS e do Azure com o Google Cloud:

<https://cloud.google.com/docs/get-started/aws-azure-gcp-service-comparison?hl=pt-br>