





**Matheus Jorge**  
Sou cientista de dados na  
Loft e escritor nas horas  
vagas



# AGENDA

- **Big Data: Introdução e Conceitos**
- **V de Volume**
- **Resumo**

A vertical bar with a gradient from bright green at the top to light blue at the bottom.

# Big Data: Introdução e Conceitos

# T Que bicho é esse?

- Da maneira simplista, Big Data pode ser definido como o conjunto de técnicas que permitem o armazenamento e a análise de grandes volumes de dados

Em 1 dia ...

  
**500M**  
de tweets

  
**95M**  
de fotos e vídeos  
publicados

  
**65B**  
de mensagens

**2B**  
de minutos  
de áudio

  
**3,5B**  
de pesquisas

# T Mas não é só volume

## Volume

V

A **quantidade** de dados gerada cresce exponencialmente a cada ano ...

## Velocidade

V

Esses dados são gerados e consumidos em **intervalos de tempo cada vez menores ...**

## Variedade

V

E nos mais **variados formatos** (texto, imagem, vídeo, áudio, etc.)

## Veracidade

V

Esse volume de dados que chegam cada mais rápido tornam difícil a **governança** e **garantia de veracidade** das informações



# T O que tem sido feito

## Dall-E 2

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

riding a horse lounging in a tropical resort in space playing basketball with cats in space

as a children's book illustration in a minimalist style in a watercolor style

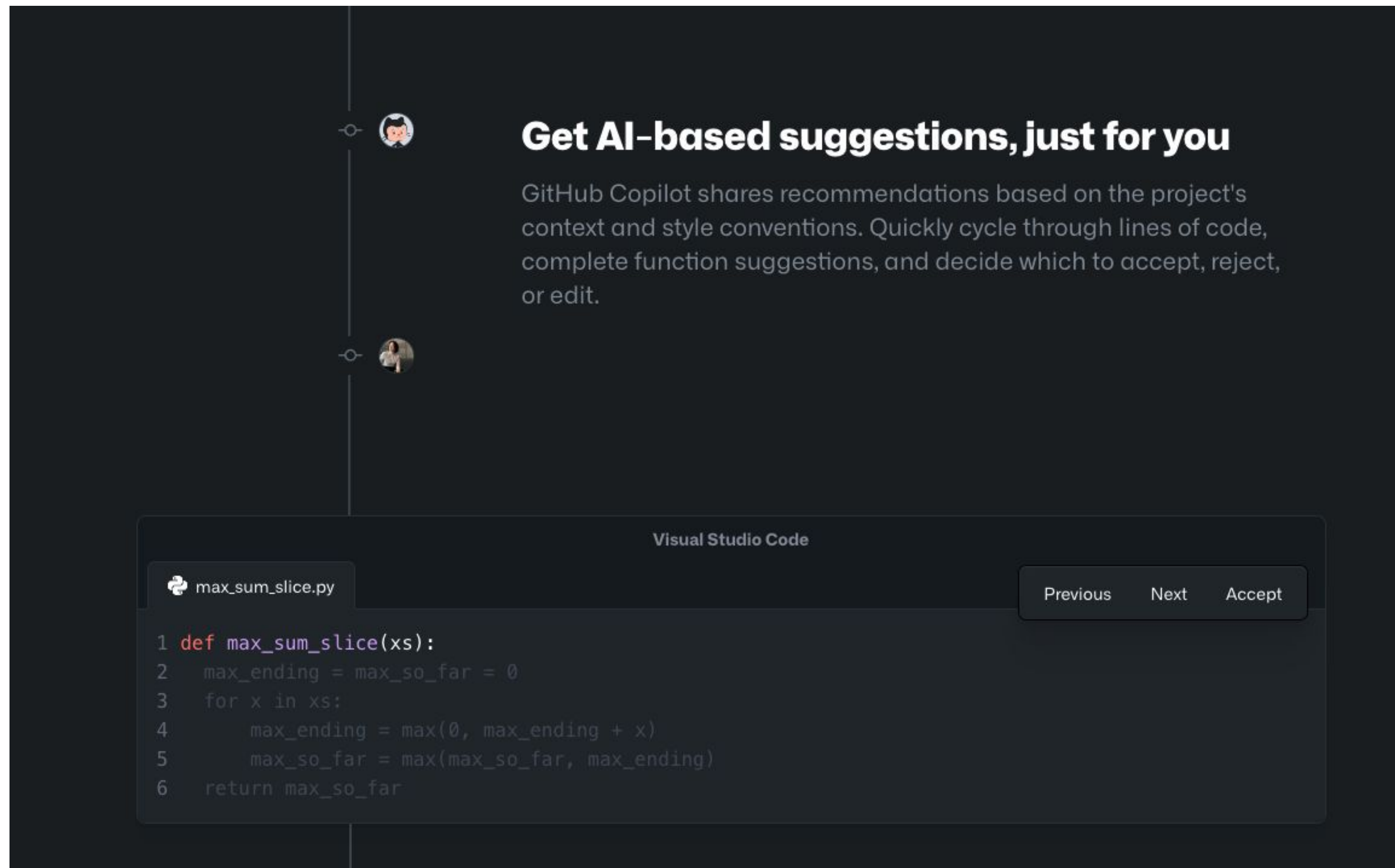


DALL-E 2



# T O que tem sido feito

## Github Copilot





# T O que tem sido feito

IoT

TRACTIAN

- Monitoramento Online
- **Tecnologia IoT**
- Gestão de Ativos
- Manutenção Preditiva

## Como funciona?

Toda falha apresenta sintomas antes de acontecer – e o Smart Trac não deixa passar nenhum. Receba **alertas automáticos** ao menor sinal de defeito e elimine os gastos com corretivas.

Saiba mais sobre os insights



T



**V de Volume**

# T Bancos relacionais vs NoSQL

- A chegada do Big Data exigiu a evolução dos bancos de dados para serem capazes de armazenar grandes volumes, dos mais variados tipos e processá-los em tempo razoável

## Relacionais

Dados armazenados em tabelas

Schema fixo e pré definido

Só armazenam com dados estruturados

Escala vertical



## NoSQL

Dados armazenados em diferentes formatos (chave-valor, documento, grafo, tabelas, etc.)

Schema dinâmico

Podem armazenar diferentes tipos de dados (tabelas, imagens, vídeos, documentos, etc.)

Escala Horizontal



# T Data Warehouse vs Data Lake

- Um novo conceito introduzido também foi o de Data Lake (ou lago de dados) para complementar os antigos Data Warehouses (ou armazéns de dados)

## Data Warehouse



Somente dados extraídos de sistemas transacionais e já processados e limpos

Informações estruturadas e com schemas bem definidos

Ideal para operação e acompanhamento de métricas e indicadores

ETL (Extract Transform Load)

## Data Lake



Todos os dados são salvos no seu formato bruto

Informações não estruturadas, semi-estruturadas ou estruturadas com schema dinâmico

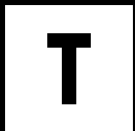
Ideal para quem precisa analisar a fundo os dados

ELT (Extract Load Transform)



# Feature Store

- Um conceito ainda mais recente é o de **Feature Store**;
- A Feature store é onde são armazenadas e organizadas as variáveis com o propósito explícito de serem usadas para treino ou previsão de modelos;
- É um modo de organizar o conhecimento de uma organização para que outras pessoas possam usar as mesmas variáveis já conhecidas sem o trabalho da criação manual.



# Resumo



T

## TAKEAWAY #1

**O volume de dados gerados está crescendo.**

E precisamos aprender a lidar com eles.

T

## TAKEAWAY #2

# Entenda os 4 Vs

Para não ser surpreendido nas suas aplicações

T

## TAKEAWAY #3

A vertical bar with a gradient from bright green at the top to light blue at the bottom.

# Estude sempre!

Novos conceitos estão sempre surgindo.