



Engenharia de Dados na Prática

Pós Graduação em Engenharia de Dados e Inteligência Artificial

Jean Martins

<https://www.linkedin.com/in/jeanfmc>



O que esperar deste módulo?

01	Introdução ao Git e GitHub
02	Primeiros Passos com Python
03	SQL para Engenharia de Dados
04	Projeto Prático em Dados com Python, SQL, Git e GitHub

Introdução ao Git e GitHub

Conceitos

Versionamento e Repositórios

Git e Github

Estrutura do repositório

Principais comandos

"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.##\$%WHYDID
ICOMETOGRADSCHOOL?????.doc



Histórico de versões

Tornar disponível off-line

Nomear a versão atual

Ver histórico de versões

Ctrl+Alt+Shift+H

Histórico de versões

Todas as versões

JANEIRO

▶ 15 de janeiro, 20:14

Versão atual

● Martins Jean

▶ 15 de janeiro, 12:34

● Martins Jean

▶ 15 de janeiro, 00:49

● Martins Jean

14 de janeiro, 23:20

● Martins Jean

Revision history

Git: Not linked

Jan 12 2023, 02:11 AM -03

● jeanfelipemartinsdacosta@gmail.com

[Save now](#)

Jan 12 2023, 02:09 AM -03

● jeanfelipemartinsdacosta@gmail.com

Jan 12 2023, 02:07 AM -03

● jeanfelipemartinsdacosta@gmail.com

Jan 12 2023, 02:06 AM -03

● jeanfelipemartinsdacosta@gmail.com

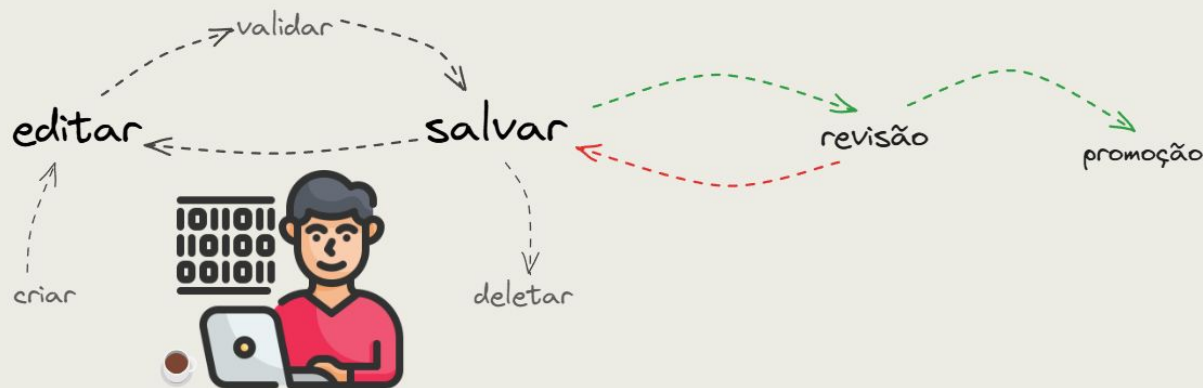
Jan 12 2023, 02:04 AM -03

● jeanfelipemartinsdacosta@gmail.com

VERSIONAMENTO

Porque a prática de versionamento é importante?

Rotina de Desenvolvimento



VERSIONAMENTO

Porque criar um repositório?

Duas pessoas trabalhando no mesmo arquivo?

Qual a última versão?

Qual versão está em produção?

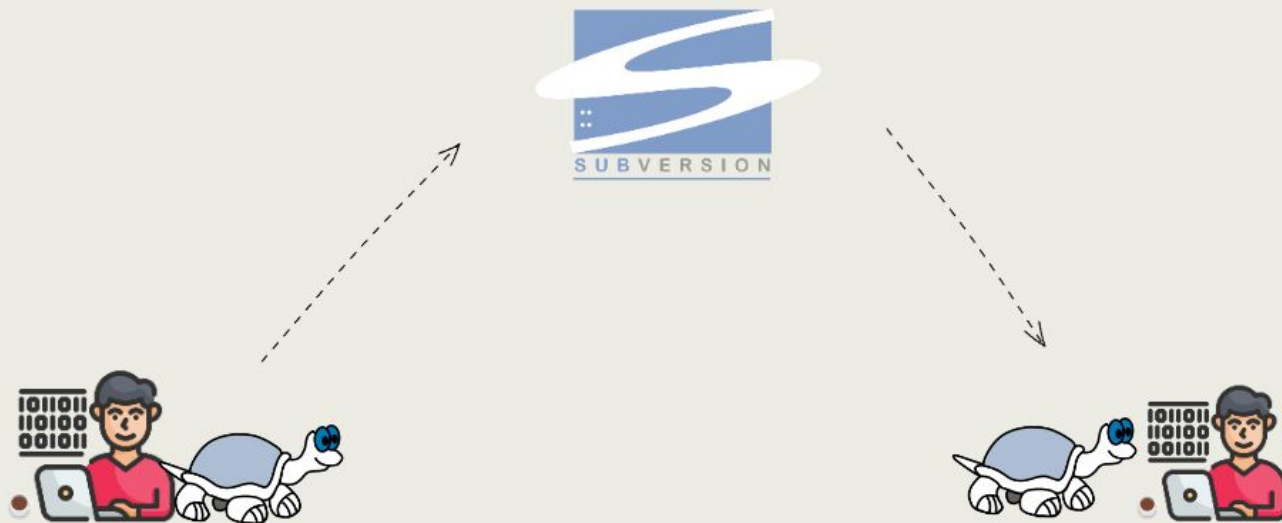
Onde eu salvo meus testes?



REPOSITÓRIOS

Arquiteturas dos repositórios

Sistemas Centralizados



REPOSITÓRIOS

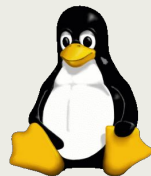
Origem do Git



Linus Torvalds



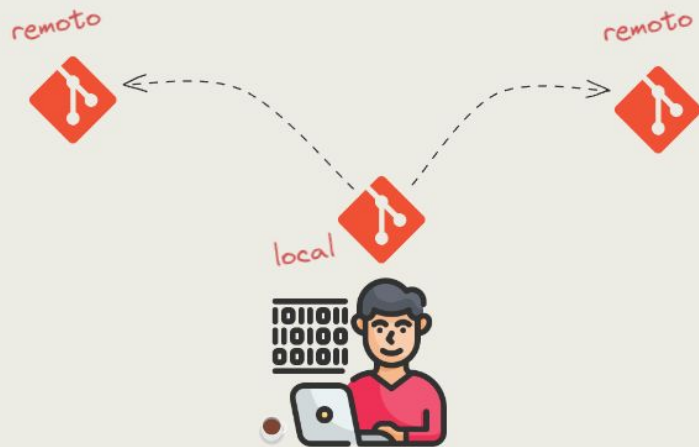
2005



REPOSITÓRIOS

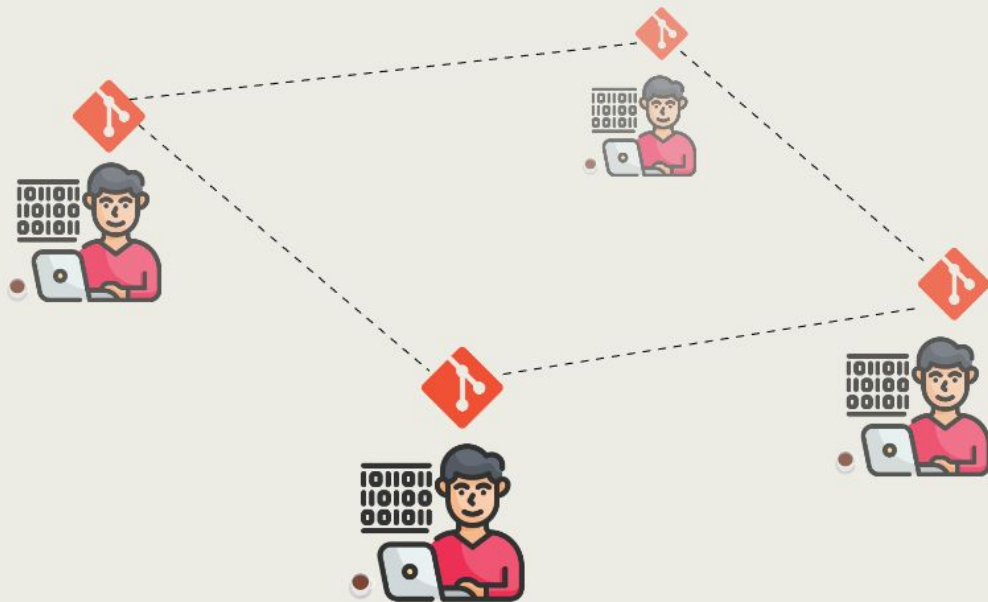
Repositórios Descentralizados

Sistemas Descentralizados



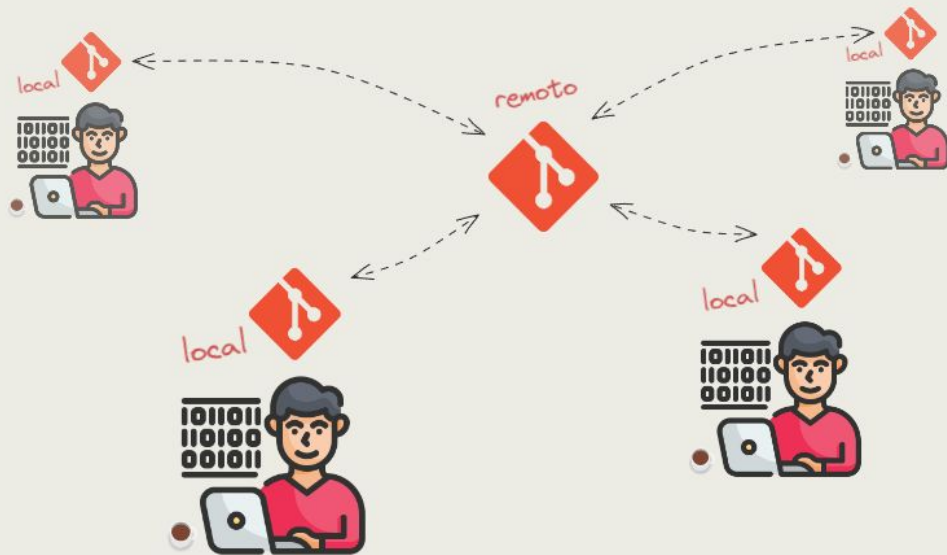
REPOSITÓRIOS

Repositórios Descentralizados



REPOSITÓRIOS

Repositórios Descentralizados



REPOSITÓRIOS

Plataformas Git

Plataformas de Gerenciamento
de Repositórios GIT



GitHub



GitLab



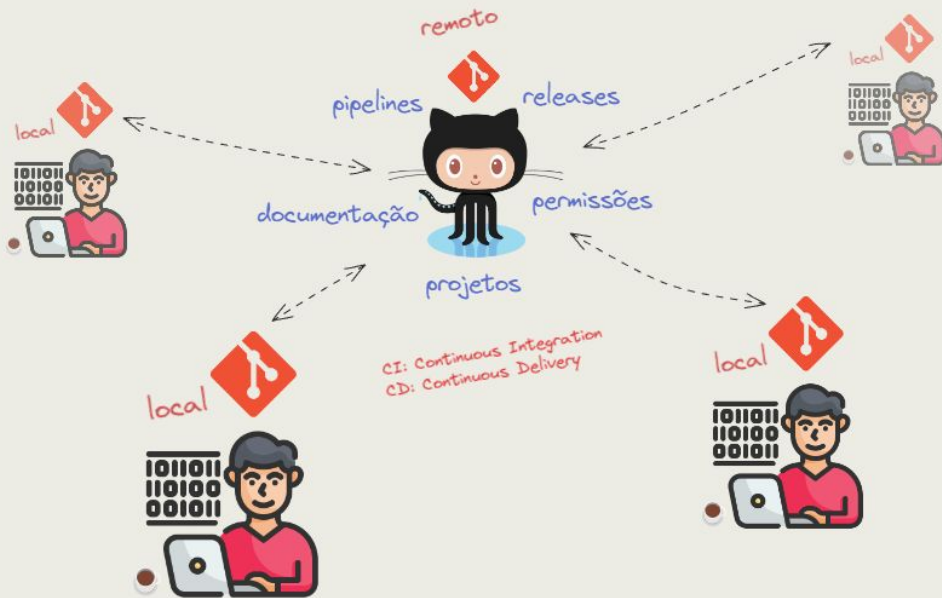
BitBucket



Azure DevOps
(Repos)

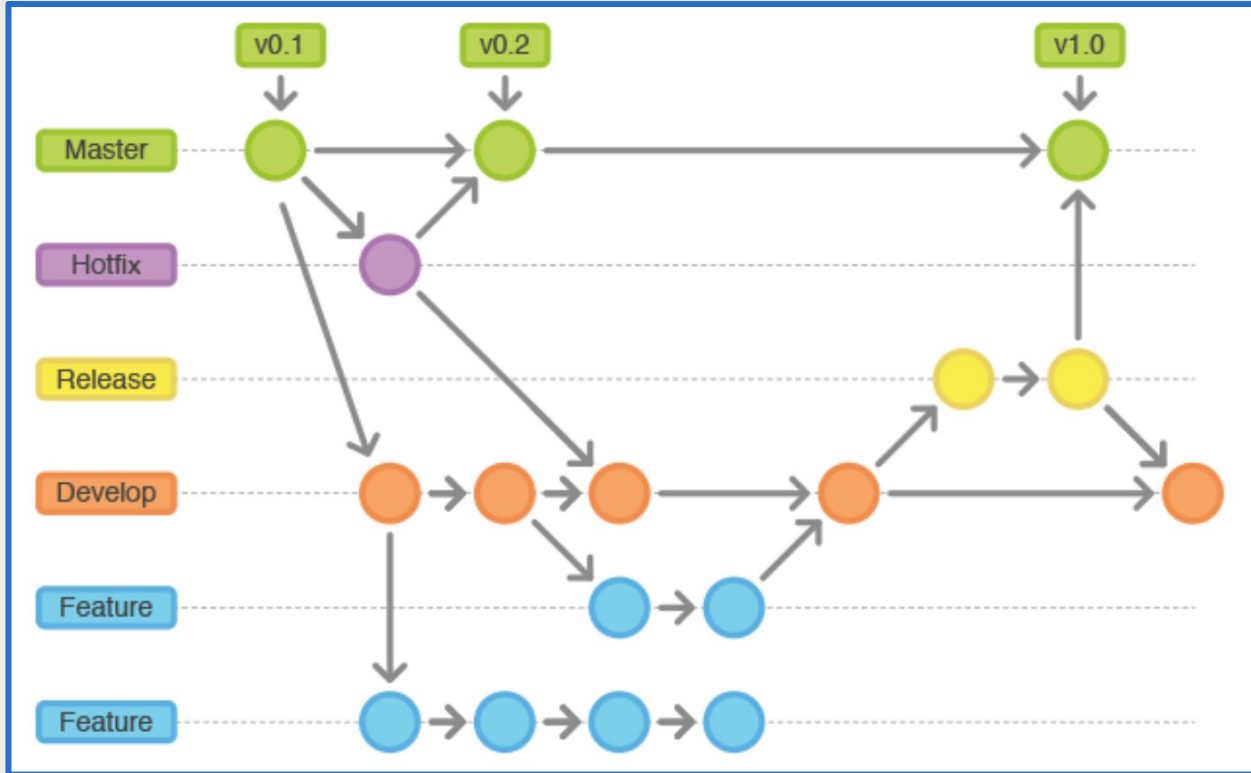
REPOSITÓRIOS

Plataformas Git



GIT

GitFlow



GIT OVERVIEW

Vamos iniciar o projeto de escrever dois capítulos a mais em um e-book, nossa equipe tem 2 escritores (um para cada capítulo) e 1 revisor.

Vamos utilizar o processo de desenvolvimento de código, utilizando a lógica dos repositórios Git.



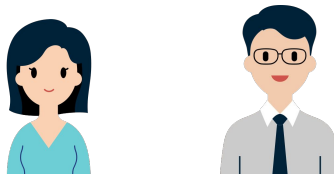
GIT OVERVIEW

Este gaveteiro vai representar nosso repositório, cada gaveta tem um propósito e representa uma branch.

Como é um e-book, temos uma gaveta para representar exatamente como o livro está nas lojas online, vamos chamar de main.



onde entregamos nosso material?



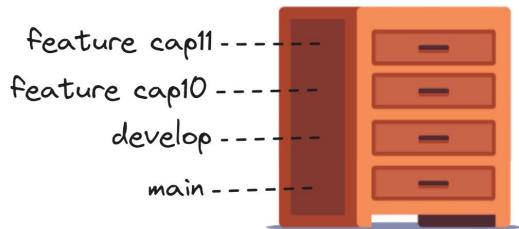
onde está o material para eu revisar?



GIT OVERVIEW

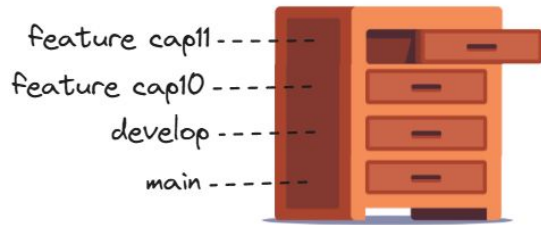
Temos uma gaveta (branch) para cada escritor (desenvolvedor) contribuir com o capítulo que está escrevendo.

O revisor só vai validar o que chegar na gaveta (branch) develop.



GIT OVERVIEW

O escritor vai enviando suas entregas diariamente colocando na gaveta do seu capítulo. Caso outro alguém precise continuar, todo material feito até aqui estará nesta gaveta.



Dia 1



Dia 2

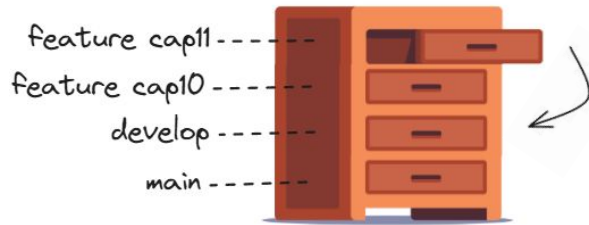


Dia 3



GIT OVERVIEW

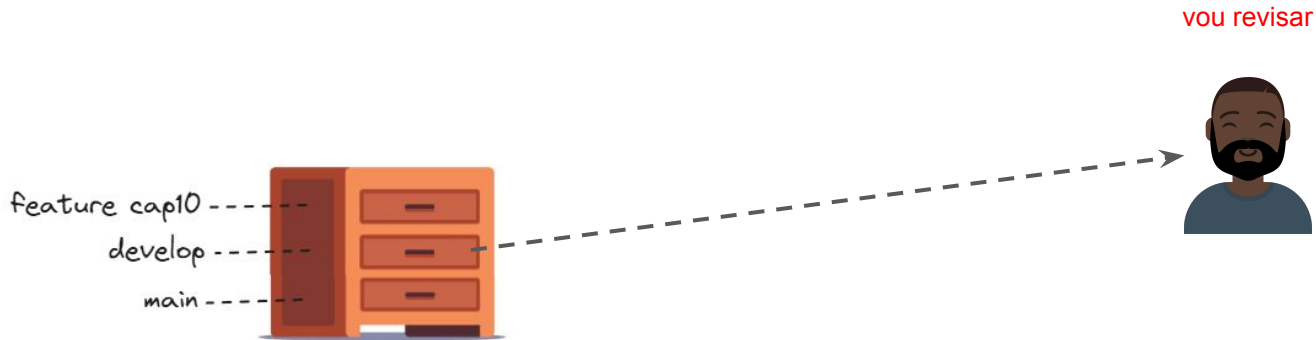
No momento que estiver pronto ele deve enviar o conteúdo de uma gaveta para a develop para ser revisada. Nos repositórios Git, para enviar de uma branch para outra, abrimos um pull request.



GIT OVERVIEW

Com o pull request aprovado e finalizado, podemos revisar esta nova entrega.

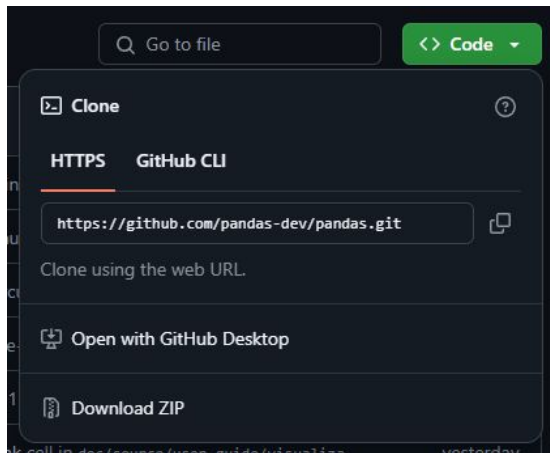
Quando tudo estiver pronto e revisado, podemos enviar para a branch main, atualizando nosso ebook.



GIT OVERVIEW

Iniciando um projeto

git clone



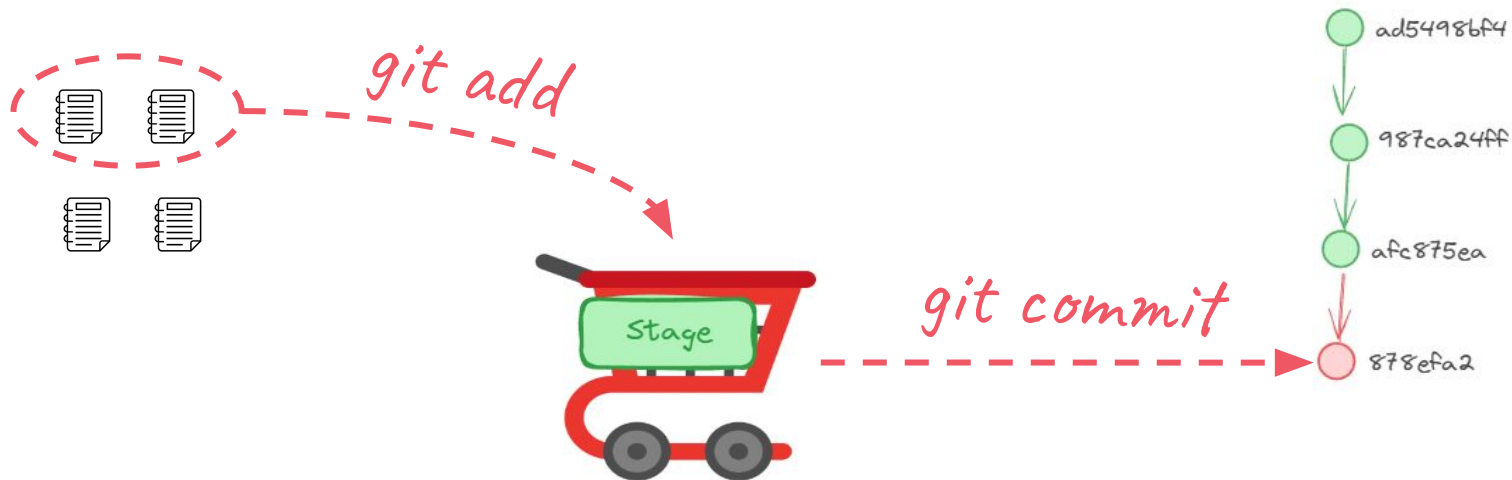
ou

git init

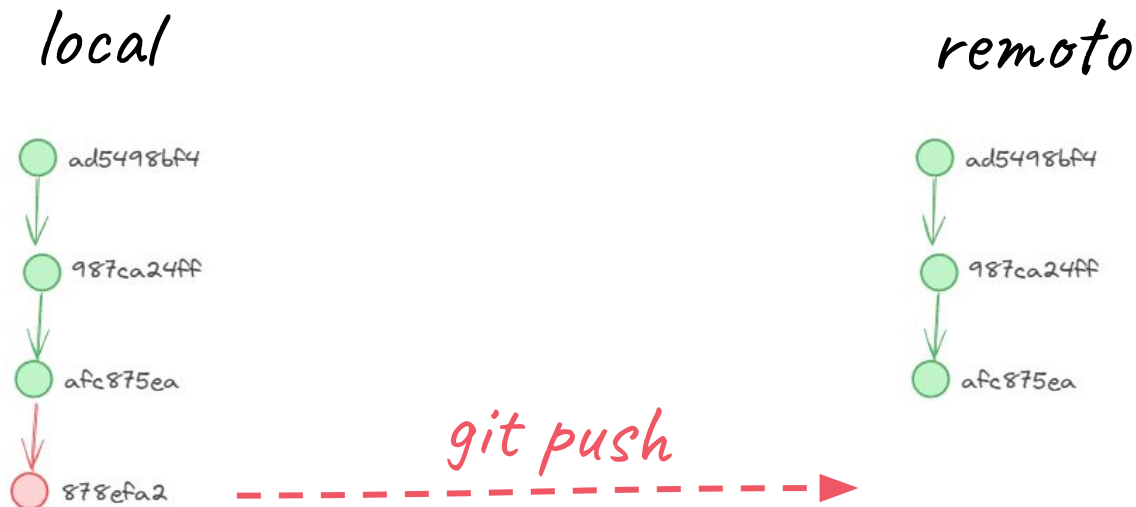


GIT OVERVIEW

Criando um commit



GIT OVERVIEW



<https://git-scm.com/downloads>



🔍 Type / to search entire site...

About

Documentation

Downloads

GUI Clients

Logos

Community

The entire [Pro Git book](#) written by Scott Chacon and Ben Straub is available to [read online for free](#). Dead tree versions are available on [Amazon.com](#).

Downloads



macOS



Windows



Linux/Unix



Older releases are available and the [Git source repository](#) is on GitHub.

GUI Clients

Git comes with built-in GUI tools ([git-gui](#), [gitk](#)), but there are several third-party tools for users looking for a platform-specific experience.

[View GUI Clients →](#)

Logos

Various Git logos in PNG (bitmap) and EPS (vector) formats are available for use in online and print projects.

[View Logos →](#)

Git via Git

If you already have Git installed, you can get the latest development version via Git itself:

```
git clone https://github.com/git/git
```

You can also always browse the current contents of the git repository using the [web interface](#).

Primeiros Passos com Python

Conceitos

Introdução ao Python

Principais comandos



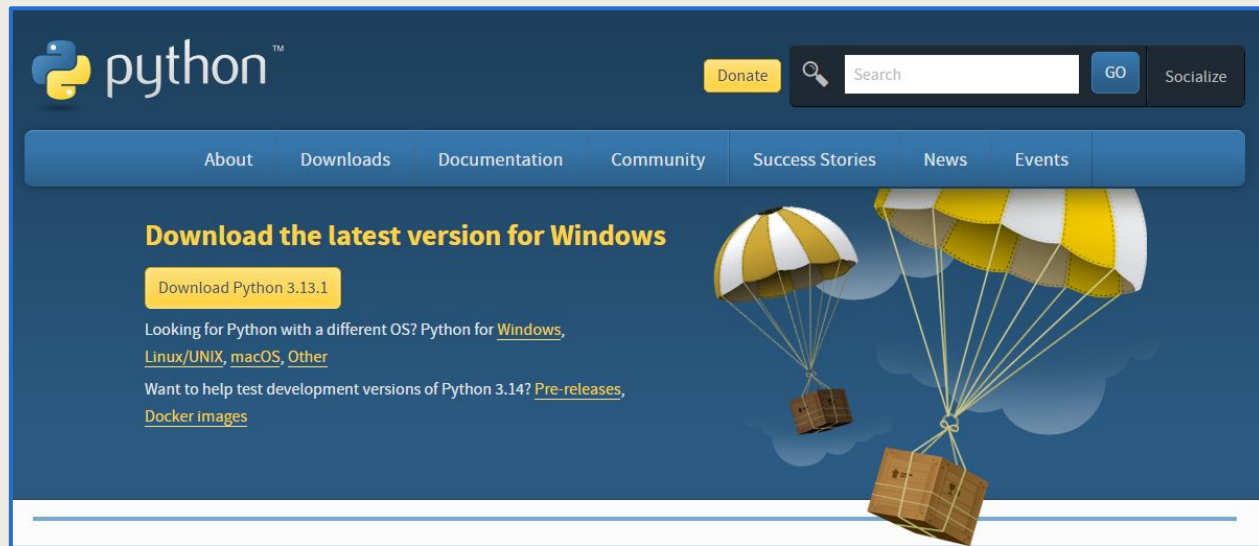
O que é Python?

Porque Python é muito utilizada?

Onde é utilizado no ambiente de Dados?

O que precisamos para começar?

<https://www.python.org/downloads/>



The screenshot shows the Python.org website with a dark blue header. The Python logo is on the left, followed by a 'Donate' button, a search bar with a magnifying glass icon, and a 'GO' button. To the right of the search bar is a 'Socialize' button. Below the header is a navigation bar with links: 'About', 'Downloads', 'Documentation', 'Community', 'Success Stories', 'News', and 'Events'. The main content area has a dark blue background. On the left, it says 'Download the latest version for Windows' in yellow. Below this is a yellow button that says 'Download Python 3.13.1'. Further down, there are links for other operating systems: 'Looking for Python with a different OS? Python for [Windows](#), [Linux/UNIX](#), [macOS](#), [Other](#)'. Below that, there are links for development versions: 'Want to help test development versions of Python 3.14? [Pre-releases](#), [Docker images](#)'. On the right side of the main content area, there is an illustration of two yellow and white striped parachutes descending from the sky, each carrying a cardboard box. The background of the illustration shows stylized white clouds.

python™

Donate

Search

GO

Socialize

About Downloads Documentation Community Success Stories News Events

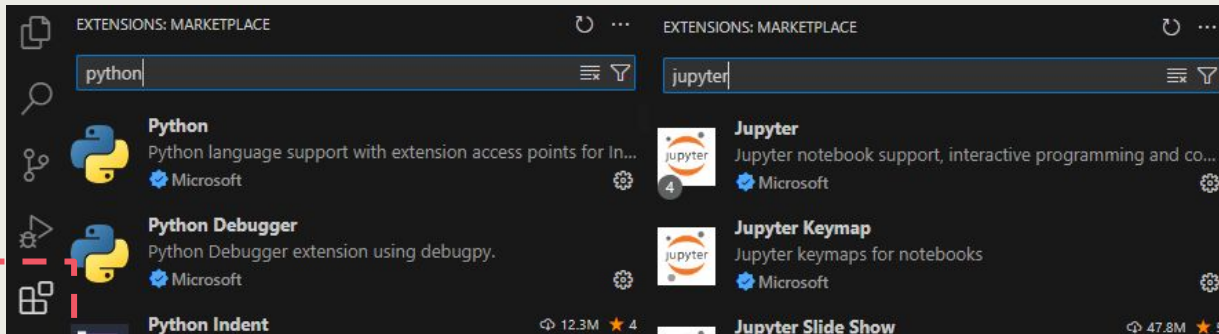
Download the latest version for Windows

Download Python 3.13.1

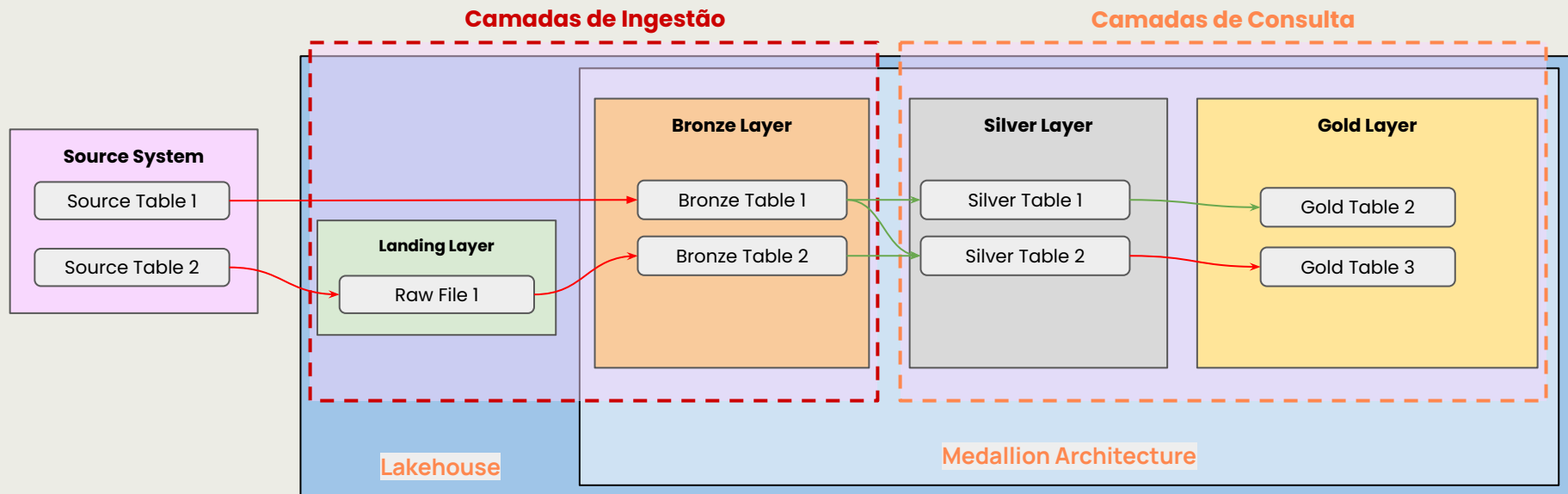
Looking for Python with a different OS? Python for [Windows](#),
[Linux/UNIX](#), [macOS](#), [Other](#)

Want to help test development versions of Python 3.14? [Pre-releases](#),
[Docker images](#)

<https://code.visualstudio.com/>



Exercício



SQL para

Engenharia de Dados

Conceitos

Introdução a SQL

Principais comandos

Utilizando SQL com Python

SQL

SQL significa Structured Query Language (Linguagem de Consulta Estruturada). É uma linguagem padrão usada para gerenciar e manipular bancos de dados relacionais.

Permite a interação com bases de dados para:

- Consultar dados.
- Inserir novos registros.
- Atualizar informações existentes.
- Excluir registros.
- Criar e gerenciar a estrutura dos bancos de dados.

SQL

Por que usar SQL?

- **Versatilidade:** SQL é usado em quase todos os sistemas de bancos de dados, como MySQL, PostgreSQL, Oracle, SQL Server e SQLite.
- **Eficiência:** Ideal para lidar com grandes volumes de dados.
- **Padrão:** É uma linguagem universal, facilitando a transição entre diferentes sistemas de banco de dados.
- **Análise de Dados:** Com SQL, você pode extrair e transformar dados para relatórios ou análises.

SQL

Tipos de comandos

- DDL (Data Definition Language): Create Alter Drop Truncate
- DML (Data Manipulation Language): Insert Update Delete
- DQL (Data Query Language): Select
- DCL (Data Control Language): Grant Revoke

Joins

Um comando muito utilizado em SQL são os comandos Joins, eles servem para juntar os dados de diferentes tabelas.

Validar inadimplência

- contratos e pagamentos

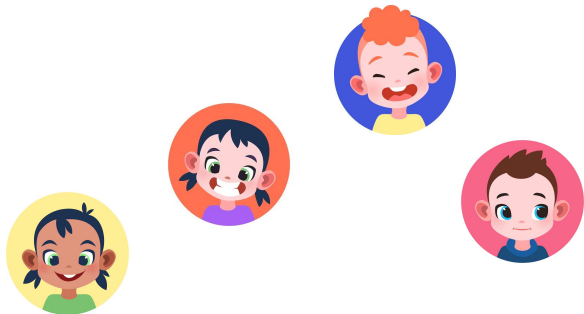
Analisar o total de vendas por categoria de produto

- pedidos e produtos

Listar funcionários sem departamento alocado

- funcionários e departamento

Joins

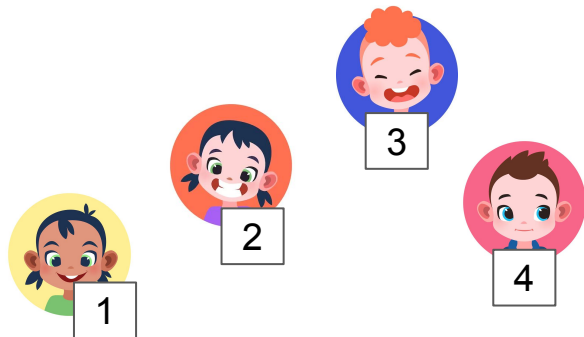


id
nome
idade

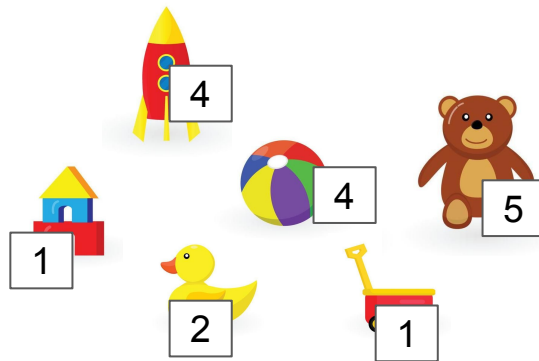


id
descricao
id_crianca

Joins



id
nome
idade



id
descricao
id_crianca

Joins



null



null



Joins - left join



null



Joins - inner join



Joins - inner join

```
SELECT *  
FROM crianca a  
JOIN brinquedob ON a.id = b.id_crianca
```

Projeto Prático em Dados com Python, SQL, Git e GitHub

Conceitos

Criação do repositório Git no Github

Criação dos dados

Recuperação e Inserção dos dados com SQL

Transformação dos dados com Python

Engenharia de Dados na Prática

Pós Graduação em Engenharia de Dados e Inteligência Artificial

Obrigado!

