Advanced Software for Data science : Julia, R, Python and Excel

Python

Colombe Becquart
colombe.becquart@tse-fr.eu

Python - 10h

Objectives

- Reviewing the basics of Python
- Introduction to the object model, inheritance and class methods
- Exception handling
- Creation and import of packages
- Handling of dataframes with Pandas
- Data Visualization
- Working with GitHub

No grading system but attendance required

Why Python is Essential for Data Scientists and Statisticians

- Widely Used in the Industry

- Rich Ecosystem of Libraries

- Ease of Learning and Use

- Strong Community Support

- Scalability and Performance

- Integration with Other Technologies (SQL databases, big data platforms, …)

Jupyter Notebook: Your Interactive Python Workspace

- Combines live Python code, visualizations, and narrative text in a single document.

- Interactive Learning: Write and run code with instant feedback.

- Documentation: Embed visuals and explanations to enhance analysis. Ideal for exploratory data analysis.

Jupyter Notebook: Your Interactive Python Workspace

- Combines live Python code, visualizations, and narrative text in a single document.

- Interactive Learning: Write and run code with instant feedback.

- Documentation: Embed visuals and explanations to enhance analysis. Ideal for exploratory data analysis.

Introduction to GitHub

What is GitHub?

- A platform for **version control** and **collaboration** using Git.
- Allows teams to manage and share code efficiently.
- Tracks changes in projects, making it easy to revert back to previous versions.

Why Should You Use GitHub?

- **Collaborate Easily:** Work together on projects, even with large teams.

- **Version Control:** Track every change in your code, with the ability to roll back.

- **Experiment Safely:** Create branches to test new ideas without affecting the main project.

- **Documentation:** Easily document your projects and analysis steps.

Key Concepts 1/3

- **Repository (Repo):** A project's workspace on GitHub where all the files and version history are stored.
- **Commit:** A saved snapshot of your project.
- **Push:** Upload local changes from your computer to the GitHub repository.
- **Fetch:** Download the latest changes from the GitHub repository to your local computer.
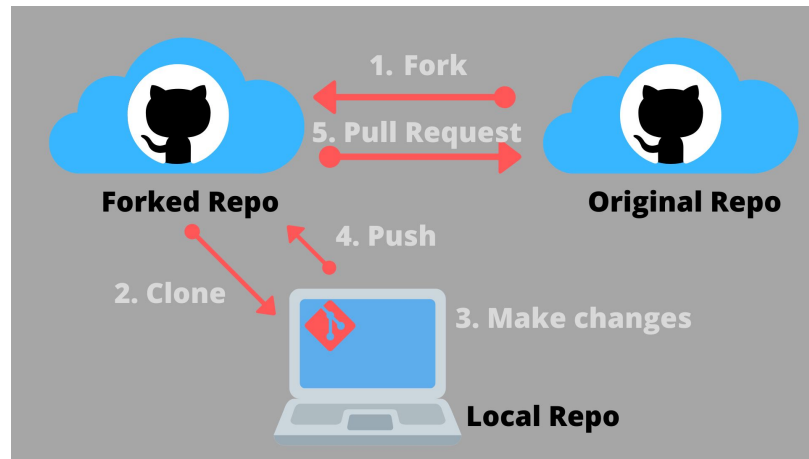

Source
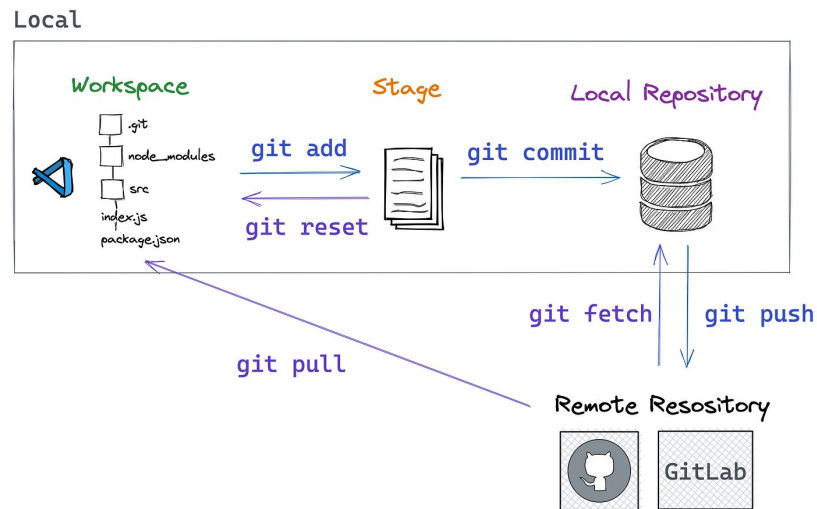
Key Concepts 2/3

- **Branch:** A version of your repo where you can work on new features separately.
- **Pull Request:** Propose changes to be reviewed and merged into the main project.
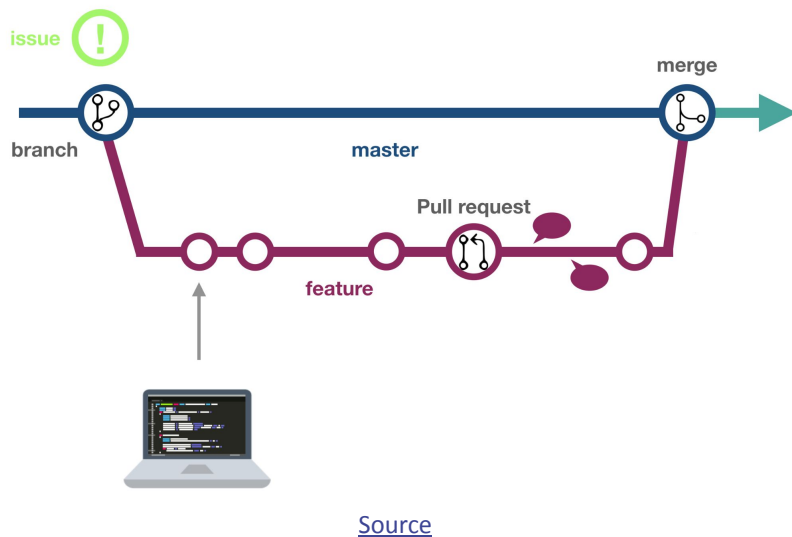
Source

Key Concepts 3/3

- **Fork:** Create a personal copy of someone else's repository for experimentation or contributing changes.
- **Clone:** Download a copy of a GitHub repository to your local machine for offline work.

[Source](#)

Getting Started

1. Create a GitHub account.

2. Download Github Desktop

3. Fork and clone the following repository: https://github.com/cbecquart/data-science-tutorials

4. Start Jupyter Notebook