

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato



"notebook" opensource per l'analisi di dati e la condivisione dei risultati

Giuseppe Chellini

GULP - Gruppo Utenti Linux Pisa

22 Ottobre 2022

Di cosa parleremo...

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Strumenti operativi per effettuare analisi dei dati e per preparare documenti efficaci per la presentazione dei risultati.

Avvertimento!

Non parleremo di linguaggi di programmazione specifici
ma di ambienti (IDE) a supporto delle analisi



Un po' di "storia"...

Introduzione

[Un po' di "storia"...](#)

[Architetture Software](#)

Jupyter/JupyterLab

[Installazione/Avvio](#)

[Applicazioni](#)

Apache Zeppelin

[Installazione/Avvio](#)

[Applicazioni](#)

Software Utilizzato

Negli anni '90 i software per l'elaborazione dati comunemente impiegati potevano essere:

- appositamente realizzati (in C, Fortran, etc...)
- tipo spreadsheet (Excel®, Origin Pro®, QtiPlot®, LabPlot, etc...)
- basati su interpreti specifici (Matlab®, Maple®, etc...)

Il loro utilizzo era fatto attraverso appositi IDE realizzati e commercializzati dalle società che curavano lo sviluppo dei software.



Un po' di "storia"...

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

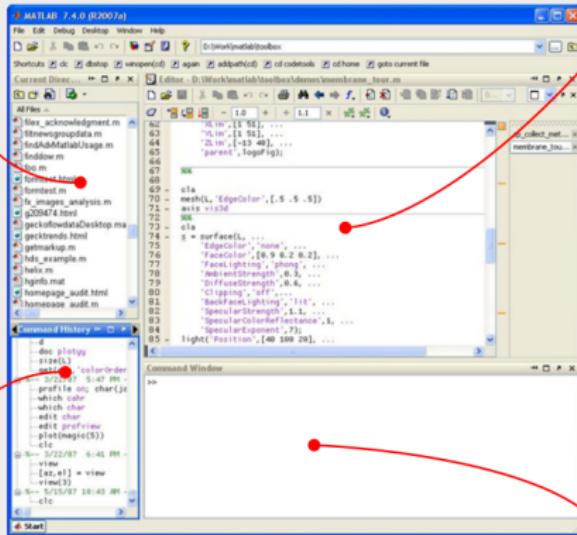
Applicazioni

Software Utilizzato



IDE di Matlab 7:

Finestra di navigazione



Storia dei comandi

Editore di script



interprete REPL

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Per quanto riguarda Python...

```
glisseppi@fedora:~$ python
Python 3.10.7 (main, Sep  7 2022, 00:00:00) [GCC 12.2.1 20220819
(Red Hat 12.2.1-1)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> a=2
>>> a
2
>>> type(a)
<class 'int'>
>>> 
```

Python nella Shell

```
IPython: home/giuseppe
IPython: home/giuseppe 65x20
[giuseppi@fedora ~]$ ipython
Python 3.10.7 (main, Sep  7 2022, 00:00:00) [GCC 12.2.1 20220819
(Red Hat 12.2.1-1)]
Type 'copyright', 'credits' or 'license' for more information
IPython 8.5.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: a=2

In [2]: a
Out[2]: 2

In [3]: type(a)
Out[3]: int

In [4]: 
```

IPython nella shell



Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato



Alcuni IDE attuali di tipo "classico"

Octave utilizza un linguaggio proprio, con una sintassi volutamente compatibile con quella dei Matlab®

<https://octave.org/>

Spyder utilizzato prevalentemente per lo sviluppo in ambiente Python

<https://www.spyder-ide.org/>

RStudio utilizzato prevalentemente per lo sviluppo in ambiente R e Python ma capace di operare con diversi linguaggi. Esiste una versione OS insieme ad altre versioni commerciali.

<https://www.rstudio.com/>

Problemi ricorrenti

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

ripetizione oltre a memorizzare le funzioni occorreva salvare dei file "script" con la sequenza delle operazioni per la loro eventuale ripetizione ed eventualmente lo stato della memoria

versione l'unico controllo di versione del file di esecuzione è basato sul salvataggio numerato degli script

visualizzazione l'unico modo per poter vedere nuovamente i risultati di un'esecuzione è ripeterla

condivisione occorre predisporre un documento apposito (.docx o altro) con i risultati delle analisi per condividerli con altre persone



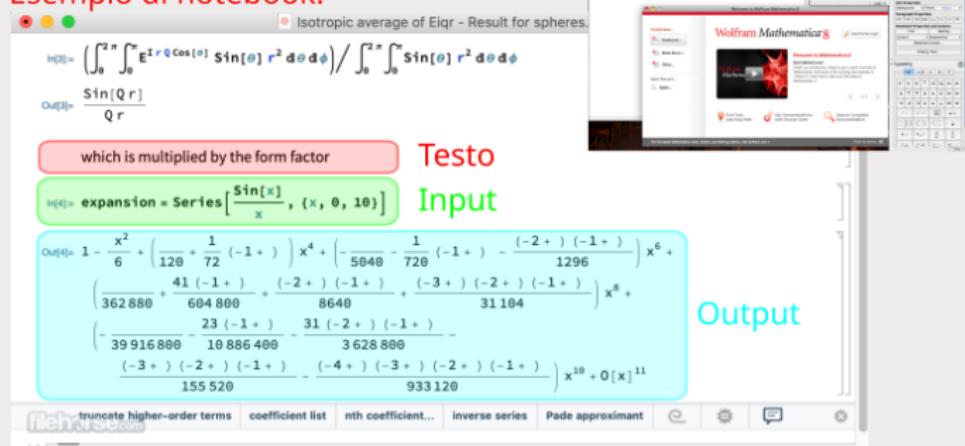
[Introduzione](#)[Un po' di "storia"](#)[Architetture Software](#)[Jupyter/JupyterLab](#)[Installazione/Avvio](#)[Applicazioni](#)[Apache Zeppelin](#)[Installazione/Avvio](#)[Applicazioni](#)[Software Utilizzato](#)

La modalità di lavoro "notebook"

Wolfram Mathematica® introdusse la modalità di lavoro mediante "notebook" intesi come ambienti in cui è possibile alternare frammenti di codice eseguibile a testo, organizzato per livelli, con formule ed immagini.

Schermata di avvio di Mathematica:

Esempio di notebook:



The screenshot shows the Mathematica interface with a notebook window. The first cell (Input) contains the code for calculating the isotropic average of Eqr for spheres. The second cell (Text) contains explanatory text about the form factor. The third cell (Input) shows the expansion of the function. The fourth cell (Output) displays the resulting series expansion.

which is multiplied by the form factor

Testo

Input

Output

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

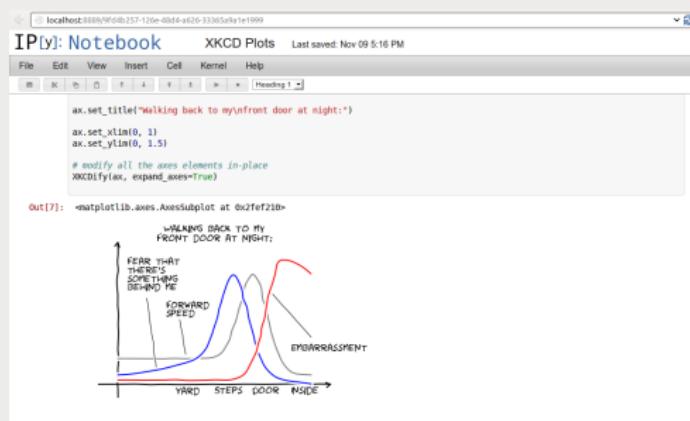
Applicazioni

Software Utilizzato

IPython Notebook

"The predecessor of Jupyter Notebook is IPython. It was born in 2001 as an afternoon hack by the hands of Fernando Perez, in the attempt to improve the Python REPL. It was a 250 lines Python script trying to mimic the Wolfram Mathematica prompt system to enable scientific computing with Python."

<https://delftswa.gitbooks.io/desosa-2017/content/jupyter/chapter.html>



Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

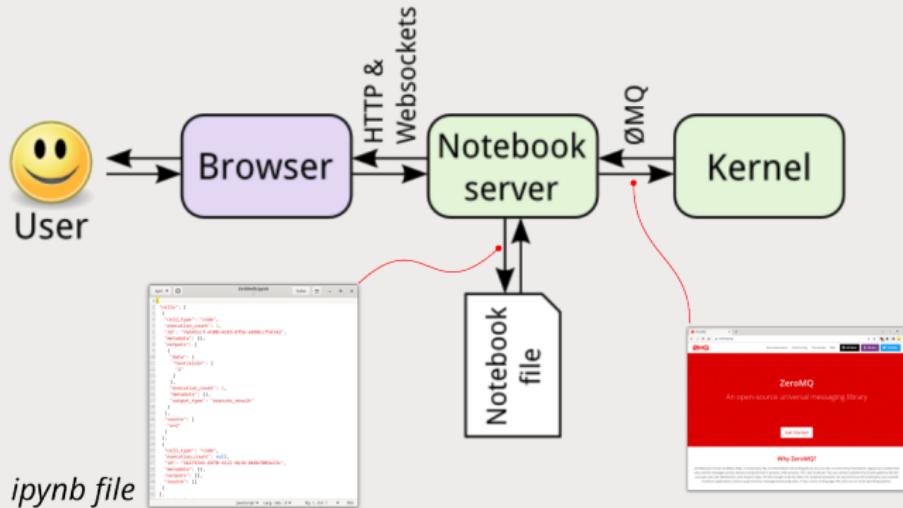
Applicazioni

Software Utilizzato

Architettura software dei "notebook"

(Riferimento specifico a Jupyter...)

Schema architetturale tipico ad alto livello:



<https://docs.jupyter.org/en/latest/projects/architecture/content-architecture.html>

Notebook

Giuseppe Chellini

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato



Jupyter Notebook / JupyterLab

<https://jupyter.org/>

The screenshot shows a web browser window with the URL <https://jupyter.org/> in the address bar. The page features a large background image of a swirling orange and brown storm on Jupiter. Overlaid on the image is the text "May 10–12, 2023" and "jupyterCON" in large letters, with "Cité des Sciences, Paris, France" and "CFP open Oct 31st" below it. In the foreground, there's a white rectangular area containing the Jupyter logo (an orange circle with the word "jupyter" inside) surrounded by various small icons representing different programming languages and tools like Python, R, and F#. At the bottom of the page, there's a blue navigation bar with the text "11/20" and buttons for "Indietro" (Back) and "Avanti" (Forward).

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Jupyter/JupyterLab

Installazione del software

Jupyter (usando "pip"):

Installazione

```
pip install notebook
```

Esecuzione

```
jupyter notebook
```

JupyterLab (usando "pip"):

Installazione

```
pip install jupyterlab
```

Esecuzione

```
jupyter-lab
```



Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Jupyter Notebook

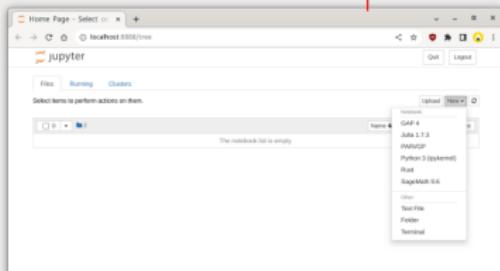
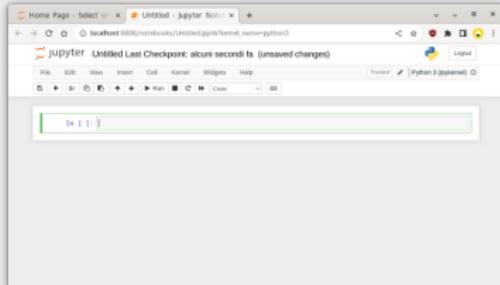
Esempio live di esecuzione

Avvio Shell

```
giuseppe@fedora:~$ jupyter notebook
```

Avvio Server

```
giuseppe@fedora:~/Dolphin/[Ingegneria Civile]/Presentazione LinuxDay2022-1/JPNotebook
giuseppe@fedora:~/Dolphin/[Ingegneria Civile]/Presentazione LinuxDay2022-1/JPNotebook ~ % jupyter notebook
[W 2022-10-19 20:01:28.572 LabApp]: password has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2022-10-19 20:01:28.572 LabApp]: password has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[I 2022-10-19 20:01:28.602 LabApp]: JupyterLab extension loaded from /home/giuseppe/.local/lib/python3.10/site-packages/jupyterlab
[I 2022-10-19 20:01:28.602 LabApp]: JupyterLab application directory is /home/giuseppe/.local/lib/python3.10/site-packages/jupyterlab
[I 2022-10-19 20:01:28.602 LabApp]: Serving notebooks from local directory: /home/giuseppe/Dolphin/[Ingegneria Civile]/Presentazione LinuxDay2022-1/JPNotebook
[I 2022-10-19 20:01:28.602 LabApp]: Jupyter Notebook 6.4.12 is running at:
[I 2022-10-19 20:01:28.602 LabApp]: http://localhost:8888/
[I 2022-10-19 20:01:28.602 LabApp]: Press Ctrl-C to stop this server and shut down all kernels (twice to skip confirmation).
```



Scelta interprete e apertura NB

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Jupyter Notebook

Esempio live di esecuzione

Dropbox (Ingegneria) | Notebook Test 1 - [jupyter] | +

File Edit View Insert Cell Kernel Widgets Help

Logout Python 3 (ipykernel)

Linux Day 2022

Questo è un testo di presentazione del notebook. Posso inserire anche del **quassotto** e dei codici.

In [1]: `a=2
a`

Out[1]: `2`

In [2]: `def funzione(x):
 return x**2
 # questo è un commento. Adesso applico la funzione
b = funzione(a)
b`

Out[2]: `3`

Posso inserire anche una figura utilizzando il linguaggio HTML, valido nel Markdown:

Introduci una descrizione per il grafico:

In [3]: `import matplotlib.pyplot as plt
plt.plot([1, 10, 64])
plt.xlabel('x')
plt.ylabel('y')
plt.title('ascissa')
plt.show()`

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

JupyterLab, evoluzione del Notebook

Esempio live di esecuzione

Screenshot of JupyterLab interface showing a notebook cell execution and a variable inspector.

The notebook cell contains the following Python code:

```

Linux Day 2022
Questo è un testo di presentazione del notebook. Posso inserire anche del grassetto e del corsivo.
(1) a=2
(1) 2
(2) def funzione(x):
    ...
    # questo è un commento. Adesso applica la funzione
    b = funzione(a)
    b
(2) 3
Posso inserire anche una figura utilizzando il linguaggio HTML valido nel Markdown:

```

The figure shows a graph of a function with two local maxima and one local minimum. The x-axis is labeled with t_0 and t_1 . The y-axis has horizontal dashed grid lines labeled "Max" and "Min".

The variable inspector on the right shows the following table:

NAME	TYPE	SIZE	SHAPE	CONTENT
a	int	28	-	2
b	int	28	-	3

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato



Apache Zeppelin

<https://zeppelin.apache.org/>

The screenshot shows the Apache Zeppelin homepage. At the top, there is a navigation bar with links for 'Quick Start', 'Download', 'Docs', 'Helium', 'Community', and 'Apache'. Below the navigation bar, the Apache logo is displayed next to the text 'Apache Zeppelin'. A large blue banner features the text 'Apache Zeppelin' and a brief description: 'Web-based notebook that enables data-driven, interactive data analytics and collaborative documents with SQL, Scala, Python, R and more.' Two buttons, 'GET STARTED' and 'DOWNLOAD', are located at the bottom of the banner. Below the banner, the word 'TECHNOLOGIES' is centered above three boxes. The first box contains the Apache Spark logo, the second contains the SQL logo, and the third contains the Python logo.

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato



Apache Zeppelin

<https://zeppelin.apache.org/>

A screenshot of a web browser displaying the Apache Zeppelin 0.10.1 documentation. The title bar says "Apache Zeppelin 0.10.1". The address bar shows "zeppelin.apache.org/docs/0.10.1/". The main header has the Zeppelin logo and navigation links: Quick Start, Usage, Setup, Interpreter, More, and a search icon. Below the header, the text "What is Apache Zeppelin?" is displayed, followed by a paragraph about its multi-purpose nature and support for 20+ language backends. To the right, there are three notebook preview cards showing pie charts, histograms, and line graphs. At the bottom of the page is a footer with a "Notebook" link and navigation icons for "Indietro" (Back) and "Avanti" (Forward).

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato



Apache Zeppelin

Installazione del software

The screenshot shows a web browser window with the URL zeppelin.apache.org/download.html. The page title is "Download Apache Zeppelin". It states that the latest release is 0.10.1, released on Feb 29, 2022. It provides three download options:

- Binary package with all interpreters ([Install guide](#)): [zeppelin-0.10.1-bin-all.tgz](#) (1.5g, pgp, md5, sha)
- Binary package with Spark interpreter and interpreter net-install script ([Interpreter installation guide](#)): [zeppelin-0.10.1-bin-netinst.tgz](#) (568 MB, pgp, md5, sha)
- Source: [zeppelin-0.10.1.tgz](#) (9.1 MB, pgp, md5, sha)

Using the official docker image

Make sure that `docker` is installed in your local machine.

Use this command to launch Apache Zeppelin in a container.

```
docker run -p 8080:8080 --rm --name zeppelin apache/zeppelin:0.10.1
```

To persist logs and notebook directories, use the `volume` option for docker container. You can also use `volume` for Spark and Flink binary distribution.

```
docker run -u $(id -u) -p 8080:8080 --rm -v $PWD/logs:/logs -v $PWD/notebook:/notebook \
-v /usr/lib/spark-2.4.7:/opt/spark -v /usr/lib/flink-1.12.2:/opt/flink \
-e FLINK_HOME=/opt/flink -e SPARK_HOME=/opt/spark \
-e ZEPPELIN_LOG_DIR=/logs -e ZEPPELIN_NOTEBOOK_DIR=/notebook' --name zeppelin apache/zeppelin:0.10.1
```

If you have trouble accessing `localhost:8080` in the browser, Please clear browser cache.

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Apache Zeppelin

Esempio live di esecuzione

2. IPython Visualization Tutorial

Overview

There're many visualization libraries available in python world. Visualization is tremendous useful in notebook environments, most of them could be used in Jupyter Notebook. And most of them can also be used in Apache Zeppelin Notebook, but maybe a little difference for some libraries. This tutorial will teach you how to use these popular visualization libraries in Apache Zeppelin. To be noted, you need to use Python interpreter to make these libraries work in Zeppelin. Python Interpreter is only available starting from Zeppelin 0.8.0.

We would cover the following visualization libraries:

- Matplotlib
- Pandas
- Seaborn
- Plotly
- Bokeh
- Holoviews
- Altair
- Flyby

Task 2 was last modified by anonymous at August 01, 2021, 7:00PM (changed)

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. The usage of Matplotlib in Zeppelin is the same as Jupyter Notebook. The key is to put `%matplotlib inline` before using Matplotlib. Below is one simple example, for more usage of Matplotlib, you can refer this [link](#).

Task 3 was last modified by anonymous at July 26, 2021, 8:16 AM (changed)

Matplotlib Example

```
[python:ipython]
%matplotlib inline

import matplotlib.pyplot as plt
plt.plot([1,2,3,4])
plt.xlabel('some numbers')
plt.show()
```

Task 3 was last modified by anonymous at July 26, 2021, 8:16 AM (changed)

Pandas

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Software per la presentazione

IMPORTANTE!!!

La presentazione è stata realizzata interamente
mediante SOFTWARE LIBERO.

- Latex con pacchetto Beamer

(<https://www.ctan.org/tex-archive/macros/latex/contrib/beamer>)

- editor per LaTeX: TexStudio (<http://www.texstudio.org/>)

- Beamer Template: Laughlin by Mohamed El Morabity

(https://fedoraproject.org/wiki/Templates_for_Presentations)

- Info descrittive: Wikipedia

(https://it.wikipedia.org/wiki/Pagina_principale)



Saluti

Introduzione

Un po' di "storia"...

Architetture Software

Jupyter/JupyterLab

Installazione/Avvio

Applicazioni

Apache Zeppelin

Installazione/Avvio

Applicazioni

Software Utilizzato

Domande?

Buon Linux Day!

