Using Jupyter: From simple personal notebooks to large deployments with thousands of users

Juan Cruz-Benito @_juancb

Who am I?

Juan Cruz-Benito

- PhD in Computer Engineering by the University of Salamanca
- National award for Spanish young researchers 2019
- Senior Software Engineer @ IBM Researh Quantum & AI
- Lead at IBM Quantum Experience Cloud Notebooks
- Nerd

What is the Project Jupyter?

Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages

What is a Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

What can I do, personally, with a Jupyter Notebook

Play with code & data

In [ ]:

```
# This is a cell of a Jupyter Notebook
# here I can work on my things (DS, CS, or any other stuff related to code)
import numpy as np
import pandas as pd

dates = pd.date_range('20130101', periods=6)

display(dates)
```

In [ ]:
```
df = pd.DataFrame(np.random.randn(6, 4), index=dates, columns=list('ABCD'))

display(df)
```

In [ ]:
```
df.describe()
```

In [ ]:
```
df.plot()
```

Conduct research and share my results

I can publish my noteboks in GitHub or other platforms to let people review my research and results

Example: Code for the paper

```
Cruz-Benito, J., Vázquez-Ingelmo, A., Sánchez-Prieto, J. C., Therón, R., García-P
eñalvo, F. J., & Martín-González, M. (2017). Enabling adaptability in web forms b
ased on user characteristics detection through A/B testing and machine learning.
IEEE Access, 6, 2251-2265.
```

https://github.com/cbjuan/paper-ieeeAccess-2017 (https://github.com/cbjuan/paper-ieeeAccess-2017)

Even after sharing my notebooks in GitHub, I can let people to execute my code online

Example: https://mybinder.org/v2/gh/cbjuan/paper-ieeeAccess-2017/master (https://mybinder.org/v2/gh/cbjuan/paper-ieeeAccess-2017/master)

Develop and document software

Using for example, **nbdev** http://nbdev.fast.ai/ (http://nbdev.fast.ai/)

Example: https://github.com/cbjuan/nbdev-test (https://github.com/cbjuan/nbdev-test)

Some tricks when using Jupyter Notebooks

In [ ]:
```
## Minimal auto-completion
x = [1, 2, 3]
```

In [ ]:
```
# Type 'x.' and press TAB
x
```

In [ ]:
```
## Get inline documentation
x = [1, 2, 3]
```

In [ ]:
```
x.append?
```

What can't I do with regular Jupyter notebooks / servers?

- Share a Jupyter server with many* other users
- Create code ready for production
- You can do many other things, but some of them could be difficult to achieve

Some things people say against notebooks

- Discourages software development good habits
- Discourages modularity
- Discourages testability

- People can alter the order of executions

- Many others. Google "Jupyter Notebooks bad reviews" https://www.google.com/search?q=jupyter+notebooks+bad+reviews (https://www.google.com/search?q=jupyter+notebooks+bad+reviews)

OK. I want to use Jupyter Notebooks with friends or colleagues

What should I do? Use:

- Jupyter Enterprise Gateway https://jupyter-enterprise-gateway.readthedocs.io/en/latest/ (https://jupyter-enterprise-gateway.readthedocs.io/en/latest/)
- JupyterHub https://jupyterhub.readthedocs.io/en/stable/ (https://jupyterhub.readthedocs.io/en/stable/)

What is JupyterHub

JupyterHub is the best way to serve Jupyter notebook for multiple users. It can be used in a classes of students, a corporate data science group or scientific research group. It is a multi-user Hub that spawns, manages, and proxies multiple instances of the single-user Jupyter notebook server.

How much users can I manage with a JupyterHub?

Depending on your hardware resources and how do you deploy it. Theoretically, nearly infinite users :D

A real case

IBM Quantum Experience Cloud Notebooks

https://quantum-computing.ibm.com/ (https://quantum-computing.ibm.com/)

- +100K users in our Quantum platform
- Deployed on IBM Cloud using Kubernetes
- Individual storage for each user
- All the quantum libraries and typical data science/research stack installed
- Integration with our other cloud tools

Questions?

References:

- Project Jupyter https://jupyter.org/ (https://jupyter.org/)
- JupyterHub website https://jupyterhub.readthedocs.io/en/stable/ (https://jupyterhub.readthedocs.io/en/stable/)
- Joel Grus "I don't like notebooks" https://docs.google.com/presentation/d/1n2RlMdmv1p25Xy5thJUhkKGvjtV-dkAIsUXP-AL4ffI/preview (https://docs.google.com/presentation/d/1n2RlMdmv1p25Xy5thJUhkKGvjtV-dkAIsUXP-AL4ffI/preview)
- Austin Z. Henley "What's wrong with computational notebooks?" https://web.eecs.utk.edu/~azh/blog/notebookpainpoints.html (https://web.eecs.utk.edu/~azh/blog/notebookpainpoints.html)