JupyterHub
A solution for running remote technical workshops

Ludmila MARIAN, 14 December 2022
Outline

- Problem description and possible solutions
- What is Jupyter Notebook and JupyterHub
- JupyterHub setup for the ICTP Workshop
Joint ICTP-IAEA Workshop on Atomic Processes in Plasmas: Data-driven Research

13 - 17 December 2021
An ICTP - IAEA Virtual Meeting
Trieste, Italy

This School will assist Ph.D. students and other early-stage career researchers to develop their skills and understanding in the application of modern data science techniques to the calculation and evaluation of data relevant to the physics and behaviour of plasmas.

Description:
The properties of laboratory, industrial, astrophysical and fusion energy plasmas are determined by numerous atomic and molecular processes. Modelling such plasmas requires a large amount of data concerning its collisions, transitions, collisions with other particles and interactions, all of which is directly accessible by experiments. Plasma physicists therefore task to calculations and approximations to compile their data needs.

The last 10 years have seen a huge increase in the amount of computing power available to run the codes that calculate these data. Accompanying this expansion has been the development of so-called data science techniques which unify statistical methods, data analysis and machine learning.

The School will provide training for and knowledge transfer between scientists working in the fields of computational plasma physics and data science. Participants are expected to have some experience with basic programming in a language such as Python.

Topics:
- Plasma emission modelling
- Practical exercises using FLYCHK and the NIST LIBS database
- The principles of molecular dynamics modelling
- Practical exercises on the determination of plasma-material interaction parameters using LAMMPS
- Accessing, assessing and exploiting online data resources
- Managing atomic and molecular data
- Managing data, code and the usage descriptions and the JupyterLab environment for research

How to apply:
Order application: http://indico.ictp.it/event/9657/

Registration:
There is no registration fee.

Deadline:
6 December 2021

Further information:
http://indico.ictp.it/event/9657/
smr3638@ictp.it
Problem

How to effectively run **virtual** technical workshops?

- Each computer is the **same** (Operating System and Software Tools)
  
  Installation problems can be checked on the spot and **fixed**

- Each computer is **different** (Operating System, Version, Dependencies, Software Tools)

  Installation problems are **impossible** to troubleshoot
# Options for running remote workshops

1. Postpone them until they can be run in-person
   - *not an option for the IAEA*

2. Give participants remote access to computers
   - participants would need basic knowledge of using the chosen Operating System;
   - access to Terminal or Remote Desktop;
   - limited ways of transferring local files;

   **A. Physical computers (the ones existing at the ICTP headquarters)**
   - running Windows instead of Linux;
   - no admin rights for organizers;

   **B. Virtual computers (in the public cloud - AWS, Azure, Google)**
   - we needed several CPUs per participant, thus very expensive;
   - laborious installation

3. Use JupyterHub on the ICTP infrastructure
Using JupyterHub on the ICTP infrastructure

• **Used via the web browser** - no need for participants to install anything extra

• **Free of charge** - the ICTP offered us a cluster where they can install the JupyterHub

• **No system administration tasks** - they would be handled by the ICTP IT administrators; we would only be responsible with the installation of the specific software for the workshop

• **Modern tool for data science** - participants would become familiar with a modern tool used extensively in data science
Jupyter Notebook / JupyterHub
Free software, open standards, and web services for interactive computing across all programming languages

https://jupyter.org/
Jupyter: the Basics

• Open Source tool for **collaborating**.

• **Write & Share** from the **Browser**: code, text, equations, graphs.

• Makes your work more **transient, understandable, reproducible, and shareable**.

• Support for: **Julia**, **Python**, **R** + ~50 languages.
Jupyter: the Interface

- **Sections**
  - **File Browser**: folders or files available
    - Create new
    - Rename, Move, Download, Delete
    - Upload
  - **Launcher**
    - Create new notebooks, files
    - Start new terminals
    - Drag-and-drop tabs to arrange them as you wish
  - **Running Kernels and Terminals**
Jupyter: the Interface

• Sections
  • File Browser: folders or files available
    • Create new
    • Rename, Move, Download, Delete
    • Upload
  
• Launcher
  • Create new notebooks, files
  • Start new terminals
  • Drag-and-drop tabs to rearrange them as you wish

• Running Kernels and Terminals
Jupyter: the Interface

• Sections
  • File Browser: folders or files available
    • Create new
    • Rename, Move, Download, Delete
    • Upload
  • Launcher
    • Create new notebooks, files
    • Start new terminals
    • Drag-and-drop tabs to arrange them as you wish

• Running Kernels and Terminals
Jupyter: Create a new notebook

New Jupyter Notebook
Click right to rename, start a console for it, delete it.
**Jupyter: the Cell**

- **Input Cell**: Type code or text for the kernel to execute. Drag-and-drop cells to reorder them.
- **Input Label**: Keeps track of the execution of code using a sequence of numbers [ ], [1], [2] .. [*].
- **Edit Functions**: Save, add new cell, cut/delete cell, copy a cell, paste a cell.
- **Run Cell**: Run a cell content, stop the run, restart the kernel, restart and rerun all the notebook.
- **Type of Input**: Select the type of input a specific cell will have: code, markdown, text.
- **Kernel**: Shows the kernel chosen for this notebook.
Jupyter: the Kernel

• Behind every notebook runs a kernel.

• Kernel state is persistent between cells and over time.

• The order of execution is stated in the label of each cell (be careful if you edit previous cells).

• Useful reset options in the menu:
  • **Restart Kernel** - clear all variables
  • **Restart Kernel and Clear** - clear also all the output
  • **Restart Kernel and Run All** - clear everything and rerun all cells
Jupyter: Markdown

### ICTP-IAEA Workshop

#### Notebooks

Let's **explore** some of the features of *JupyterLab Notebooks*.

For more details check the [documentation](https://jupyterlab.readthedocs.io/en/stable/).

Let's start by generating some *random numbers*:

```python
import random
randlist = random.sample(range(1,100), 10)
print(randlist)
```

Python 3.8.10 (default, Sep 28 2021, 16:18:42)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.27.0 — An enhanced Interactive Python. Type '?' for help.

```
[1]: import random
    randlist = random.sample(range(1,100), 10)
    print(randlist)

    [93, 57, 99, 88, 12, 84, 65, 62, 8, 49]
```

```
Jupyter: some useful Features

- Insert LaTex in the code cells

\[ z = \sqrt{a^2 + b^2} \]

- Run command line commands from inside the notebook

```bash
[15]: list = ! lrs
```

```
total 220K
-rw-r--r--  1 smr3638 grpsmr3638j  21K Dec 10 19:21 output.png
-rw-r--r--  1 smr3638 grpsmr3638j    584 Dec 10 19:41 text.csv
drvx-xr-x  1 smr3638 grpsmr3638j     36 Dec 10 19:44 src
-rw-r--r--  1 smr3638 grpsmr3638j   25K Dec 10 23:36 a35g.tf1.png
```
Jupyter: some useful Features

• *Tab* to complete methods or attributes available.

• Documentation on functions
Jupyter: tool for Collaboration

• Notebooks are automatically saved and stored on disk as `.ipynb` (JSON).

• Easily shareable: **Export Notebook As..**

• Add it to [GitHub](https://github.com) (automatically rendered).

• Use [binder](https://mybinder.org) to run the notebook in the cloud.

• Use [docker](https://www.docker.com) to package the notebook + dependencies.
A multi-user version of the notebook designed for companies, classrooms and research labs.
JupyterHub capabilities

JupyterHub allows users to access a server from any computer with internet using the web browser, without any local installation.

- Provides a Jupyter Notebook environment for thousands of users;
- Flexible authentication for users;
- Can be deployed on any infrastructure (public cloud, virtual machines, local server);
- Can be accessed from any computer with internet;
- Ready to use environment, with all dependencies installed (no more wasted time with installing and maintaining software);
- Administrator can login into the running instance of each user and troubleshoot;
- Jupyter and JupyterHub are open source.
JupyterHub setup for the ICTP Workshop
ICTP Workshop use case

- JupyterHub installed on an ICTP multi-CPU server and available at: https://jupyter.ictp.it/smr3638/
- Each participant had an account (with credentials communicated by the ICTP organisers in advance, via e-mail)
- All participants had access to exactly the same environment with all the software tools needed already installed
- All participants had access to a common folder (with data examples, tutorials, executables)
- On the first day we had an introductory presentation on Jupyter Notebooks and command-line tools
- One lecturer was presenting the tutorial, and another one was running it (just like a participant), to anticipate any problems
Thank you!

Try it: https://calc4-nds.iaea.org

Questions: L.Marian@iaea.org