Introduction to SWAN

Diogo Castro, Enric Tejedor, Pedro Maximino
On behalf of the SWAN team

https://cern.ch/swan

November 6th, 2024
CERN School of Computing on IT Services

SWAN in a nutshell

- Interactive analysis with a web browser
  - No local installation is needed
  - Based on Jupyter Notebooks
  - Calculations, input data and results “in the Cloud”
- Good for data analysis and exploration, but also for teaching
- Easy sharing of scientific results: plots, data, code
- **Added value: integration with CERN infrastructure and services!**

Integrating (CERN) services

The Notebook

Text

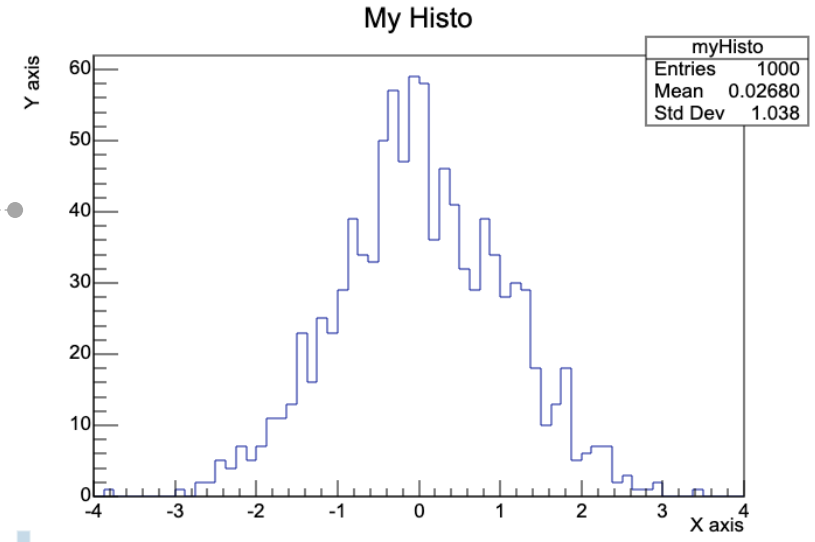
Code

Graphics

2 Displaying graphics

We can now draw the histogram. We will at first create a canvas, the entity which in ROOT holds graphics primitives. Note that thanks to JSROOT, this is not a static plot but an interactive visualisation. Try to play with it and save it as image when you are satisfied!

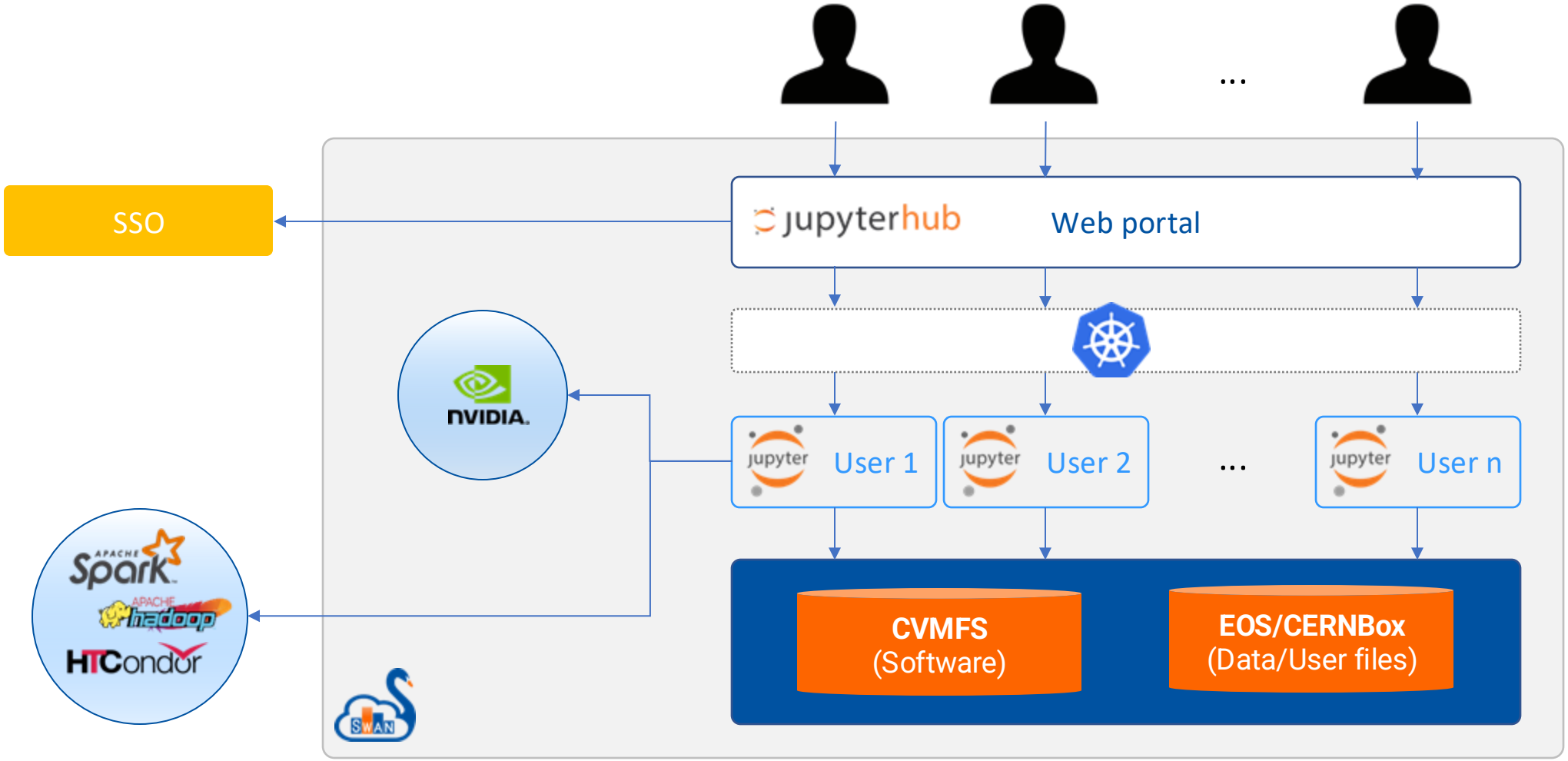
In [5]:
```
c = ROOT.TCanvas()
h.Draw()
c.Draw()
```



We'll try now to beautify the plot a bit, for example filling the histogram with a colour and setting a grid on the canvas.

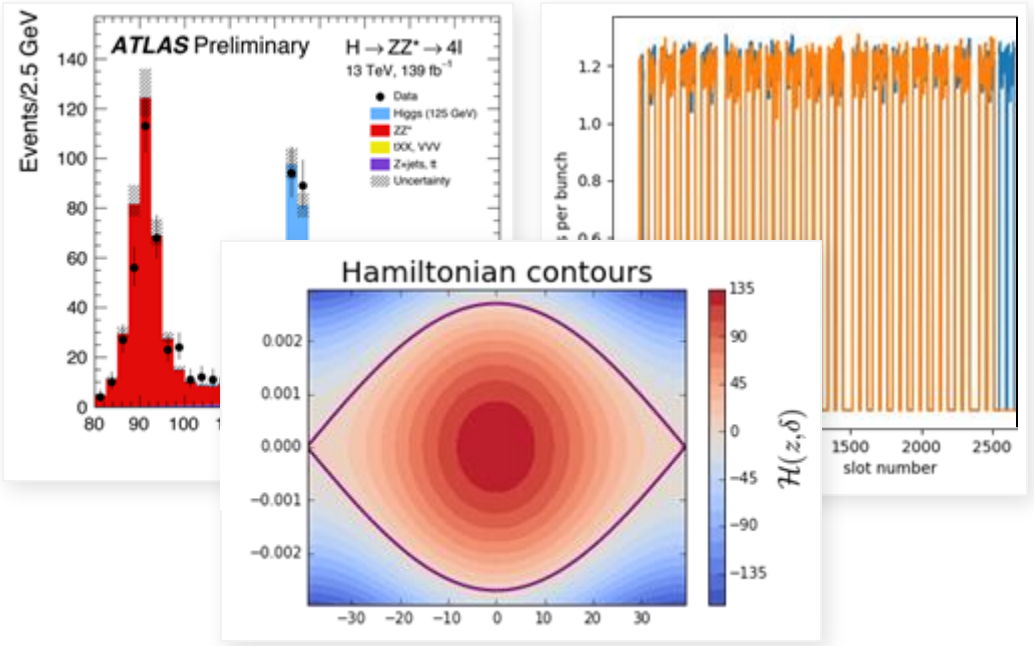
In [6]:
```
h.SetFillColor(ROOT.kBlue-10)
c.SetGrid()
h.Draw()
c.Draw()
```

Architecture



Main user communities

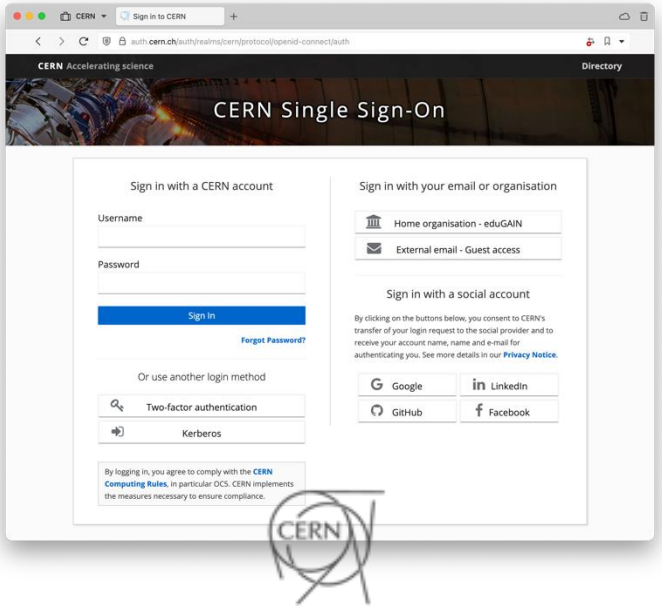
- Physics analysis
  - Usually last stages of analysis
  - Interactive, exploratory
  - Collision event data, ntuple-like, columnar
  - More and more with Machine Learning
- Non-physics analysis (e.g. ATS)
  - LHC studies: extract machine measurements, query machine settings
  - Beam dynamics simulation
  - Query and process LHC logs distributedly via Spark
  - Query and plot monitoring data in experiment DAQ systems
- Education
  - Many schools/workshops use SWAN for teaching



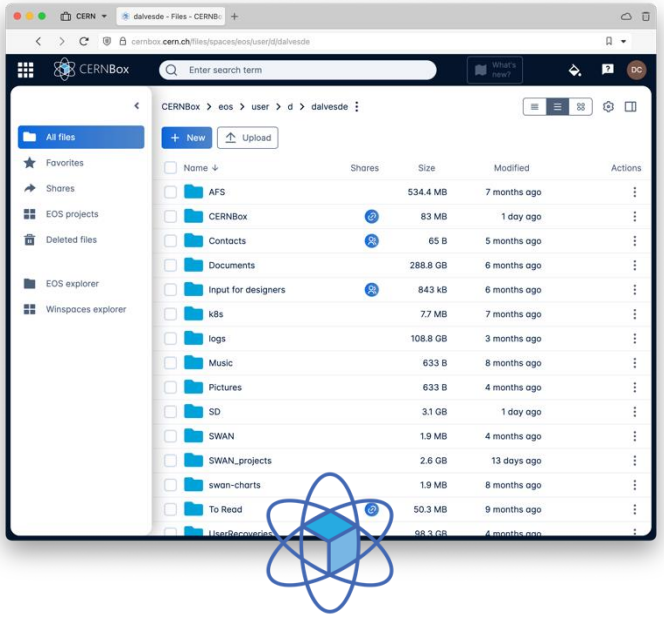
How to use it?

How to connect?

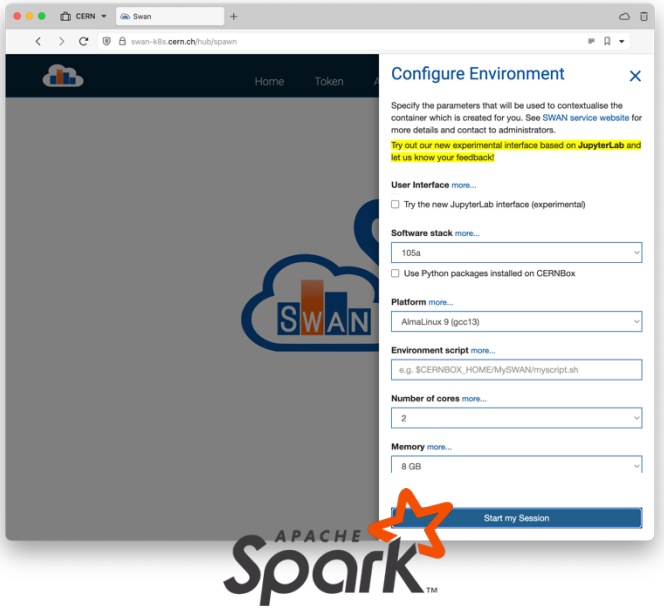
swan.cern.ch



Have a CERN account



Have a CERNBox space
(Open the service at least once!)



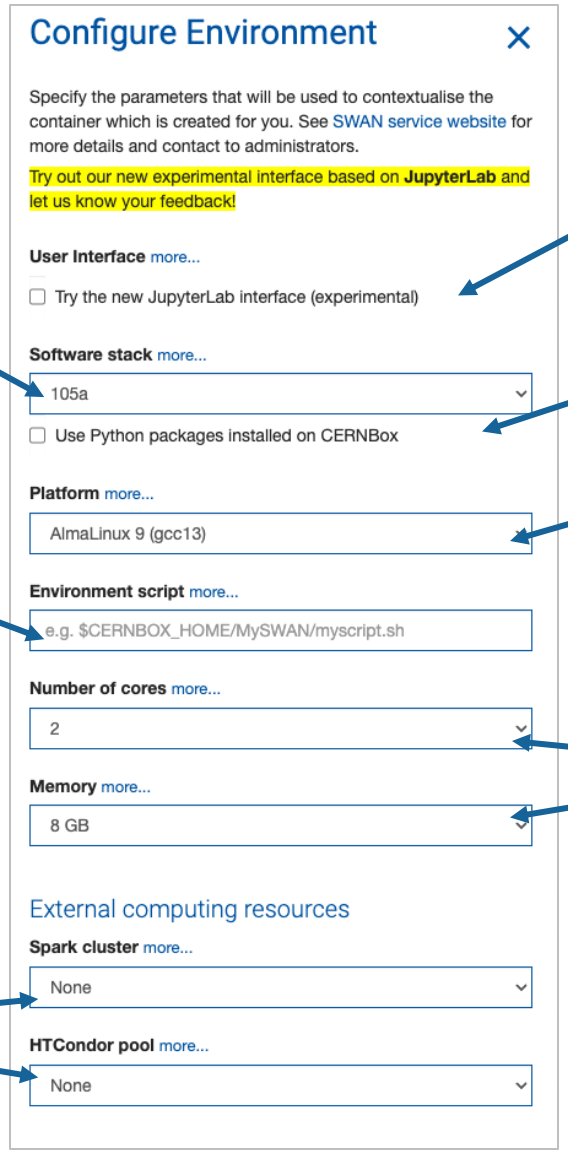
Optional:
Request access to Spark
(Service Now Request)

Start your session

Software Stack: what packages(versions) do you want?
http://lcginfo.cern.ch/

Personalise the environment? startup script!

Spark & HTCondor: do you want to offload computations to a CERN cluster? Which one?



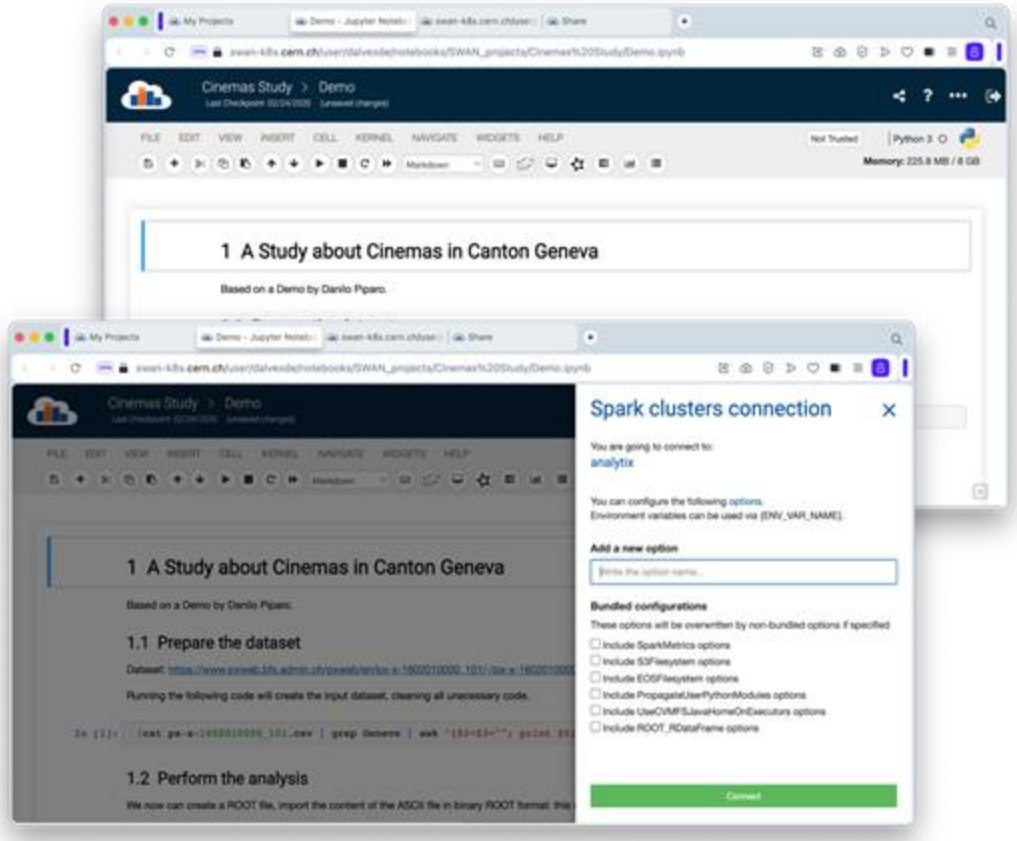
JupyterLab: Use the new JupyterLab interface?

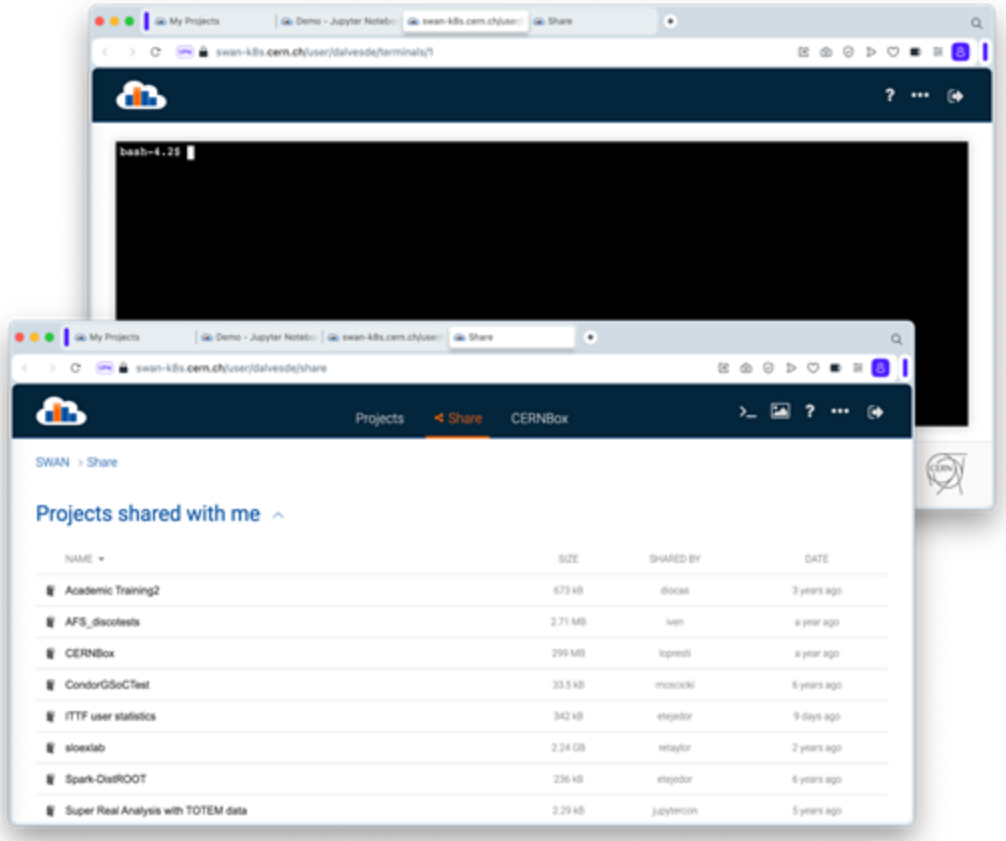
CERNBox: Include installed packages in Python path

Platform: what system/compiler?

How much **memory**? And **cores**?

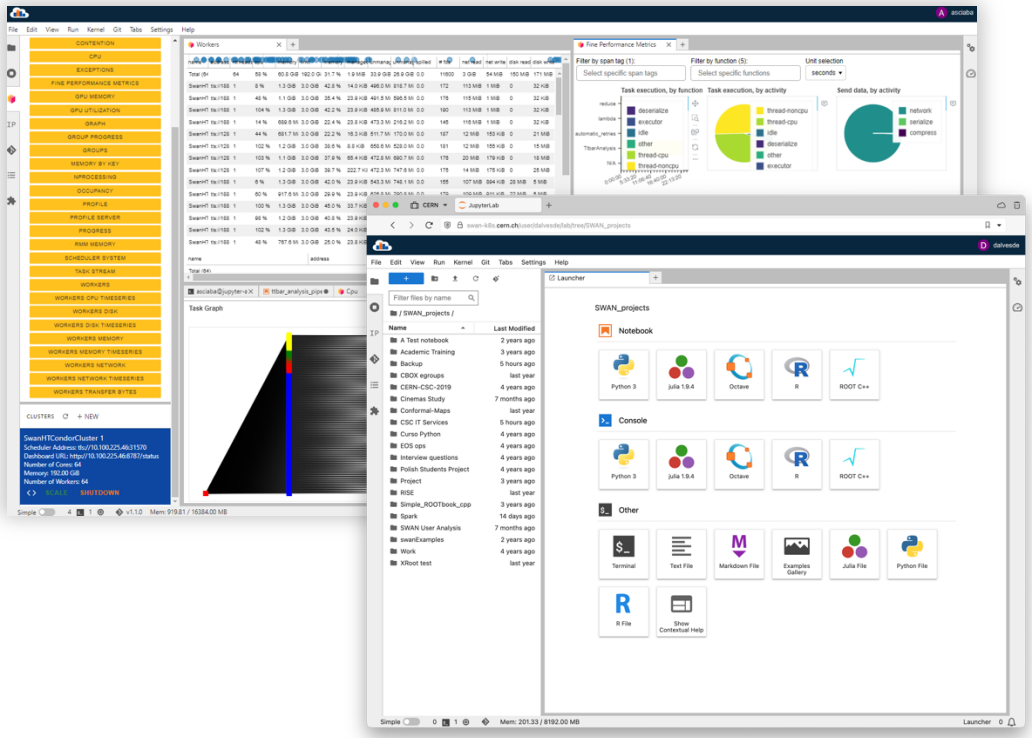
Classic UI


Cinemas Study > Demo
1 A Study about Cinemas in Canton Geneva
Based on a Demo by Danilo Piparo.
Spark clusters connection
You are going to connect to:
analytix
Add a new option
Bundled configurations
1.1 Prepare the dataset
1.2 Perform the analysis


Projects
Share
CERNBox
SWAN > Share
Projects shared with me
Academic Training2
AFS_discotests
CERNBox
CondorGSoCTest
ITTF user statistics
sloexlab
Spark-DistROOT
Super Real Analysis with TOTEM data

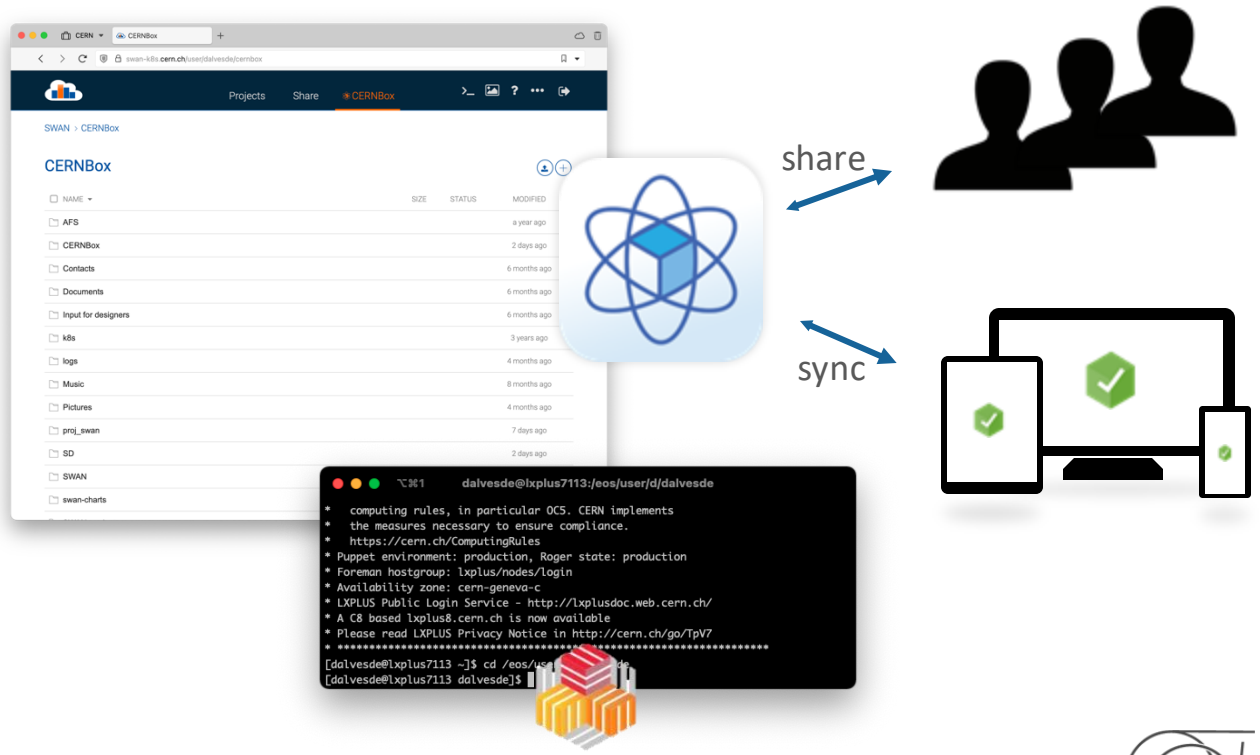
JupyterLab

- The latest interface proposed by Project Jupyter
  - Notebooks, terminals, …
  - … and virtually anything via extensions
- Started if users tick the box in the form
  - Will initially coexist with classic UI
  - To be made the default in the future
- Most SWAN extensions already migrated to JupyterLab



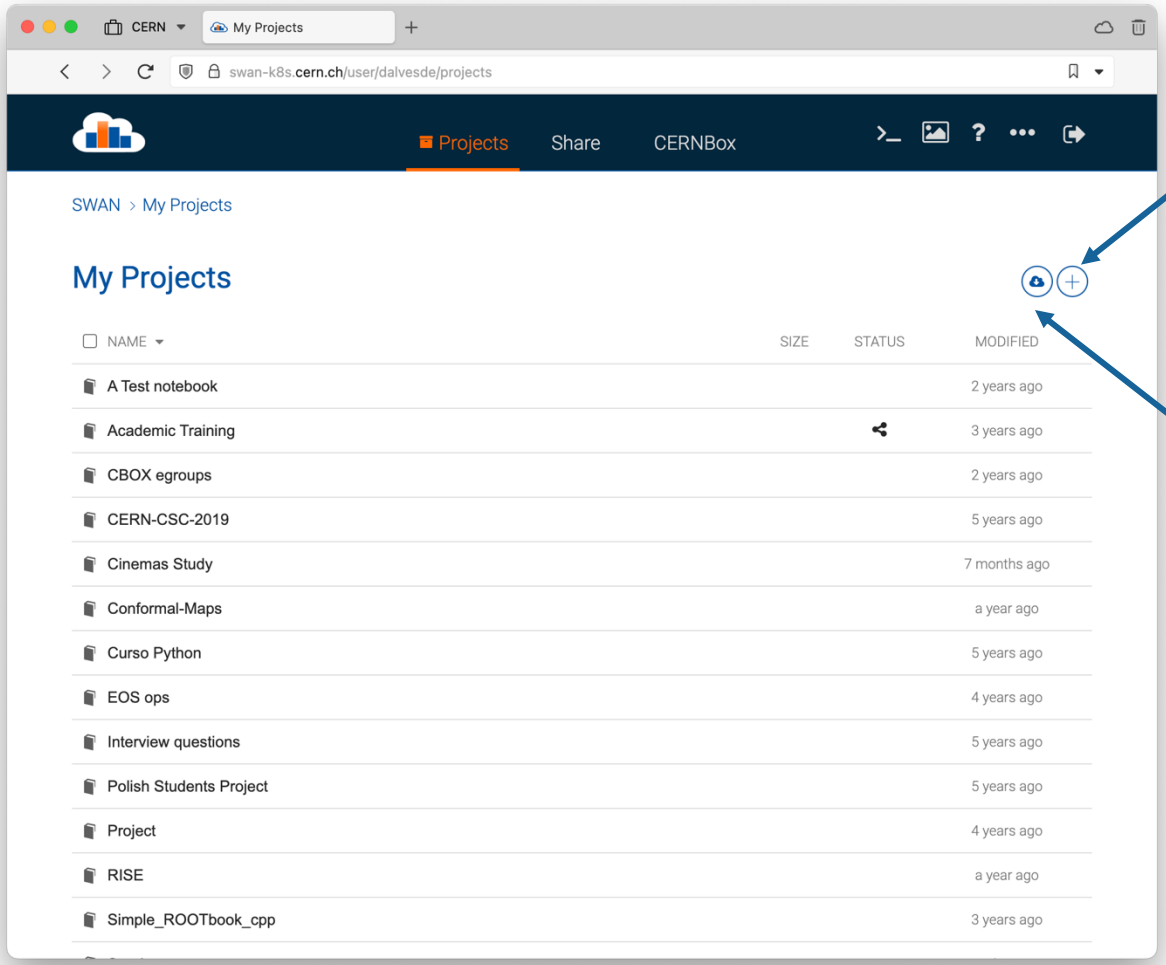
CERNBox is your home

- All the data our users need for their analysis
  - CERNBox as home directory
  - Experiment repositories, projects, open data, …
- Sync & Share
  - Files synced across devices and the Cloud
  - Simple collaborative analysis
- Data accessible from other services
  - E.g. lxplus



Projects

Folder with a set of notebooks, identified by a ".swanproject" file inside it



1. Create empty

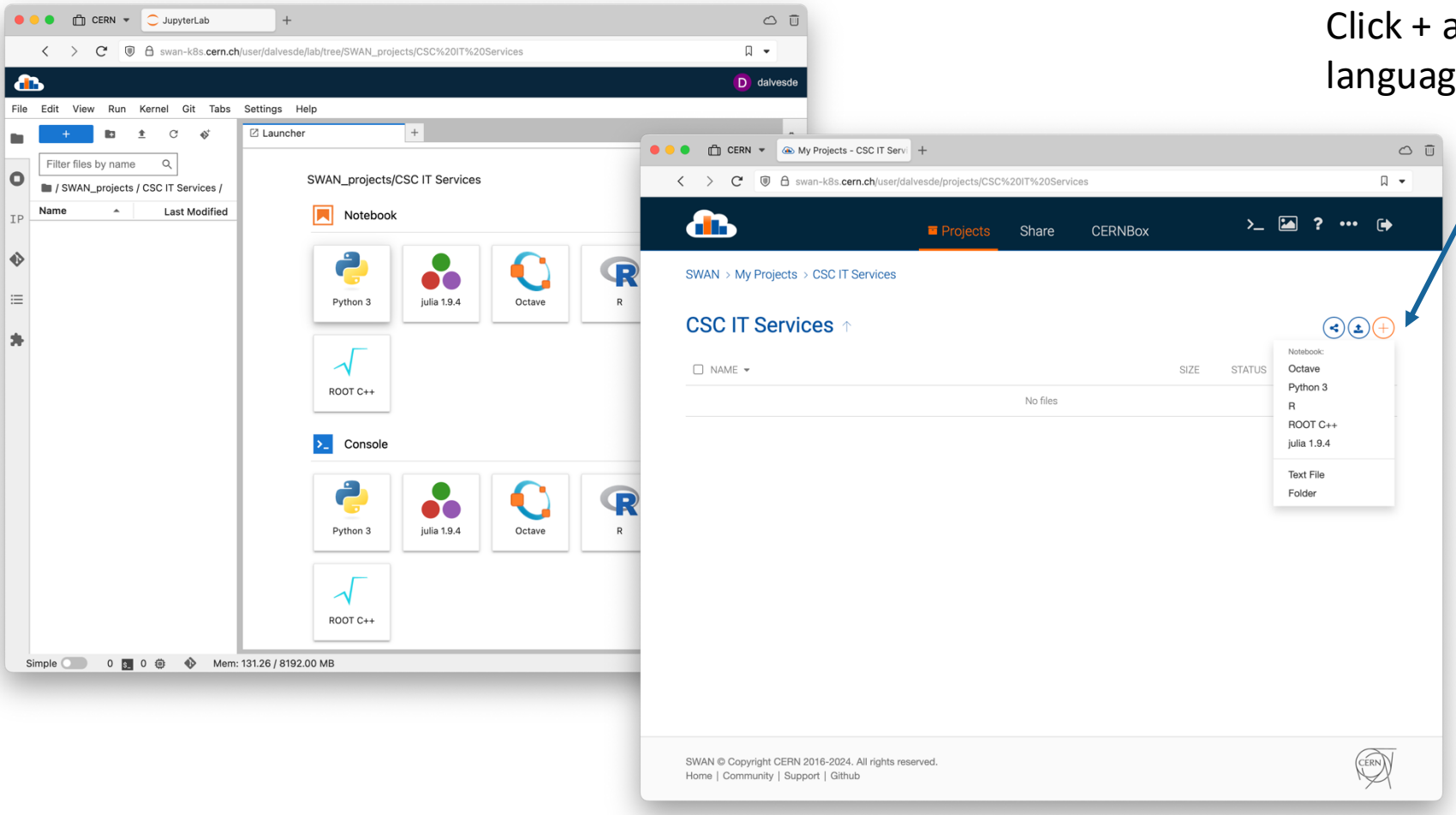
2. Download from our Gallery, GitHub, CERN GitLab, CERNBox or ROOT website

3.

Open in SWAN

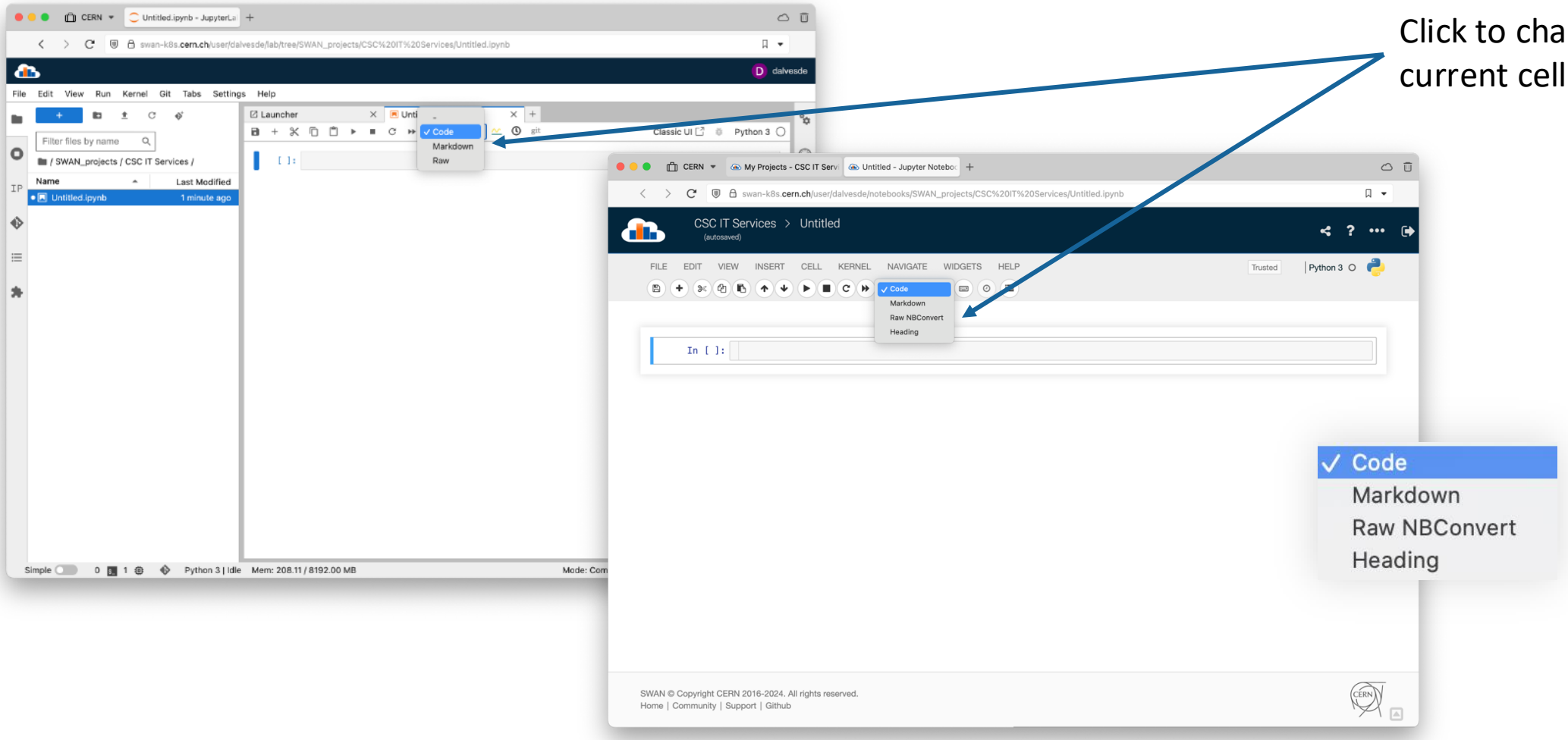
https://cern.ch/swanserver/cgi-bin/go?projurl=**<path to your repo>**

Creating notebooks

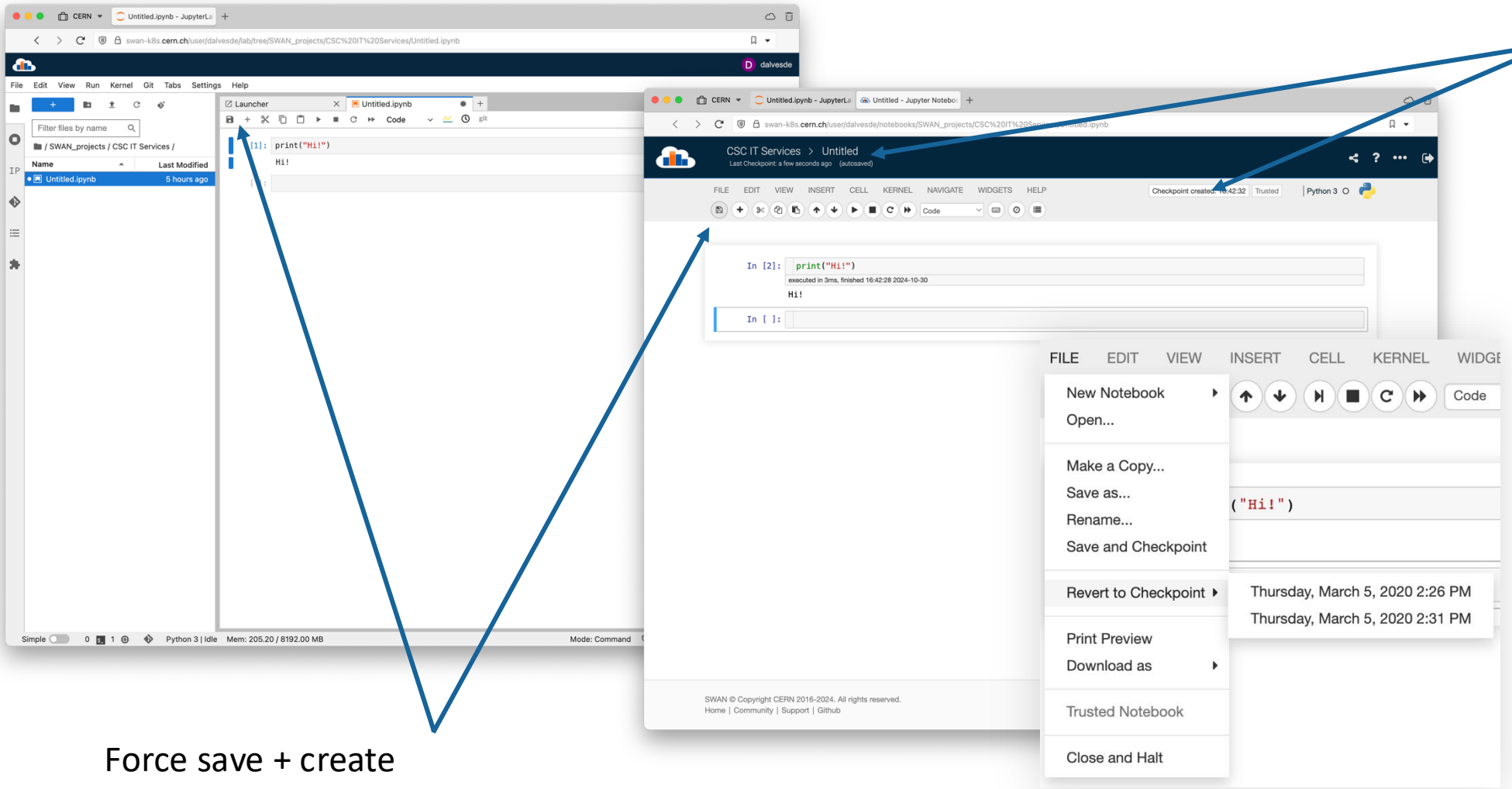


Click + and choose language

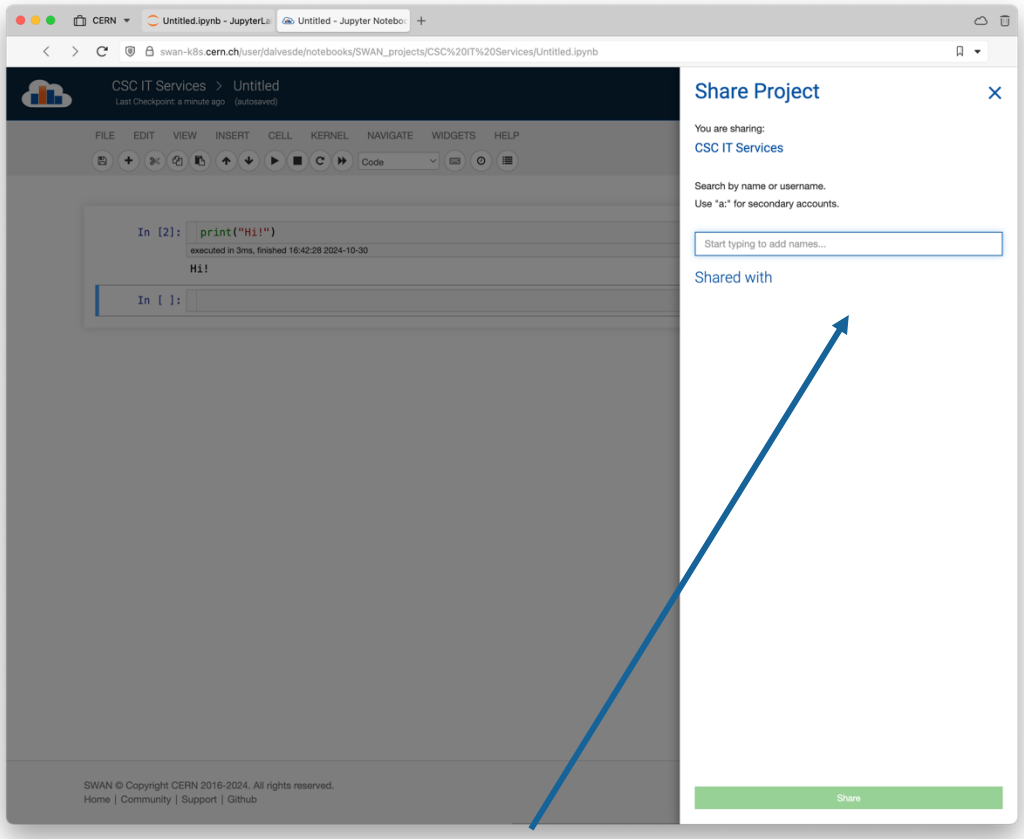
Notebook cells



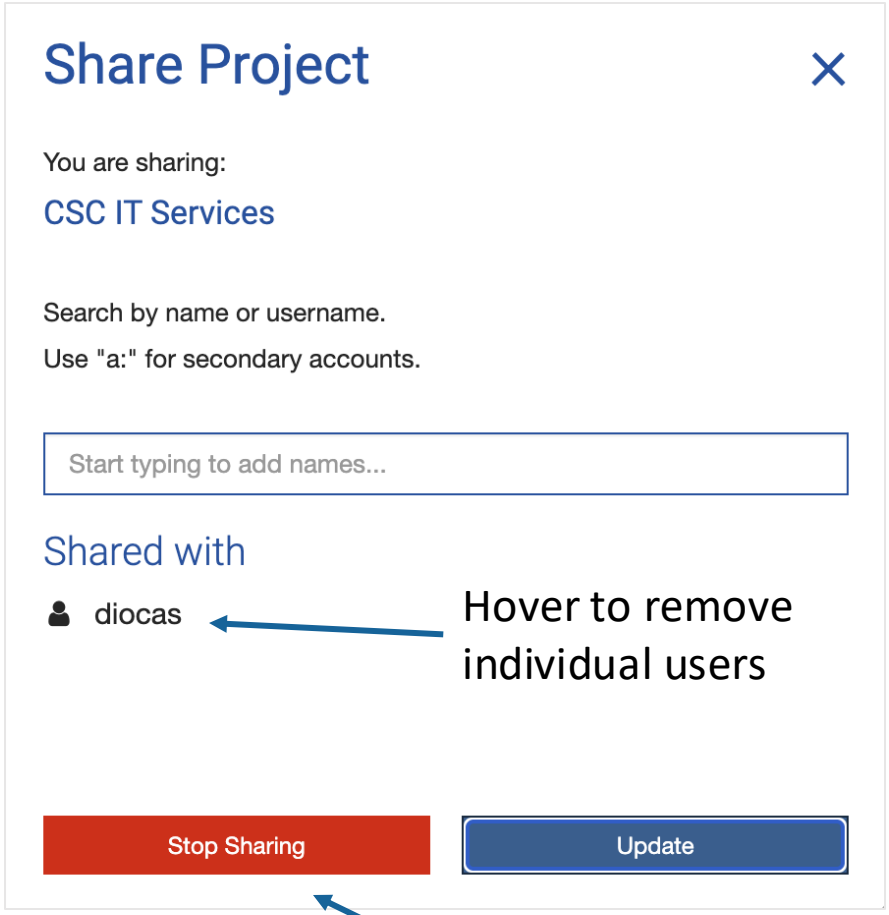
Saving notebooks



Sharing (Classic UI)



Search like from CERNBox (“a:” for secondary account)
(Egroups support released early next year)



GPUs & Machine Learning

GPUs in SWAN

- SWAN allows attaching a GPU to a user session
  - We currently offer 18x Tesla T4 + 4x A100
- The GPUs are used **interactively**
  - When starting their session, the user selects a CUDA software stack and gets a GPU
  - GPU-enabled packages can then be used in a notebook and computations offloaded to the GPU by default

```
In [1]: import tensorflow as tf

        tf.debugging.set_log_device_placement(True)

        # Create some tensors
        a = tf.constant([[1.0, 2.0, 3.0], [4.0, 5.0, 6.0]])
        b = tf.constant([[1.0, 2.0], [3.0, 4.0], [5.0, 6.0]])
        c = tf.matmul(a, b)

Executing op MatMul in device /job:localhost/replica:0/task:0/device:GPU:0
```

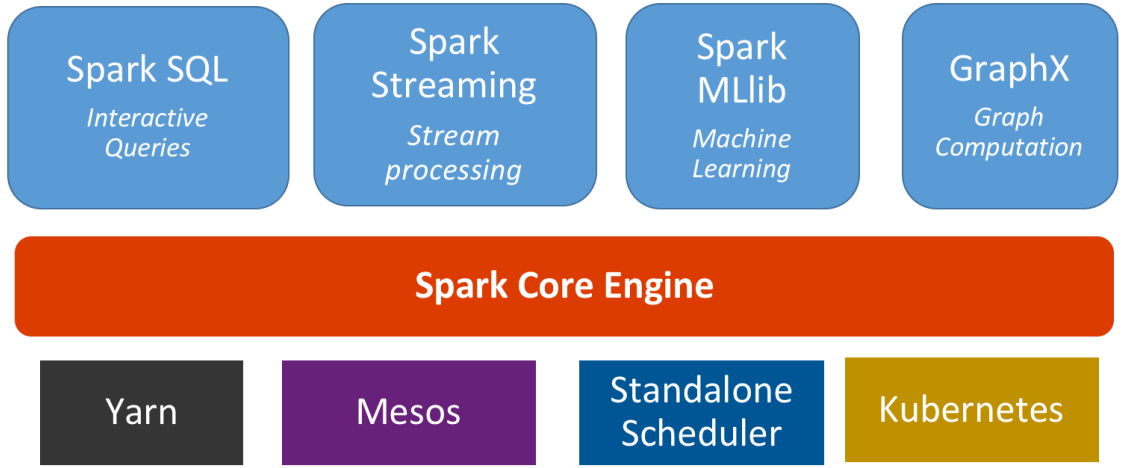
ML software on CVMFS

- Software provisioning for ML applications via CVMFS
  - **LCG CUDA** stacks with GPU-enabled software for ML

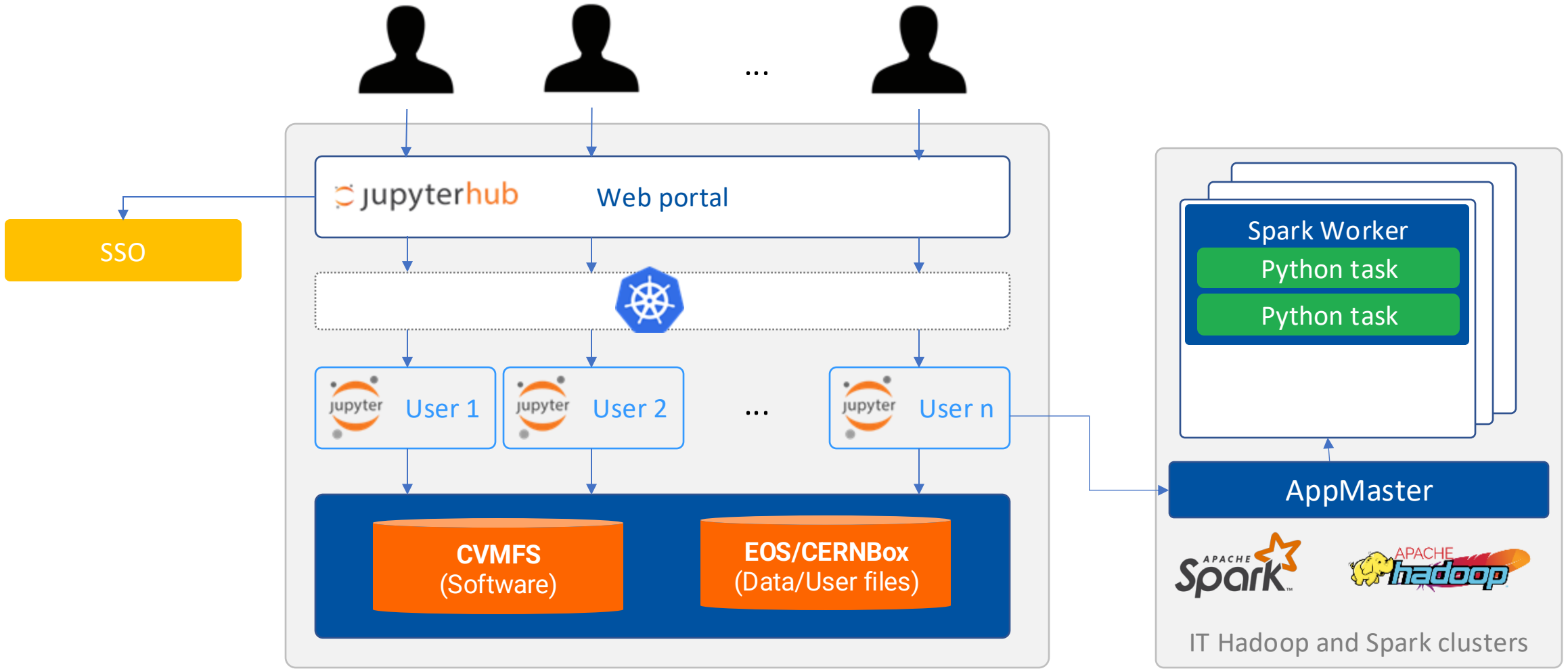
Apache Spark

What is Apache Spark?

- Apache Spark
  - An open-source parallel processing framework with expressive development APIs (in multiple languages) that allows for sophisticated analytics, real-time streaming and machine learning on large datasets
- Spark ecosystem

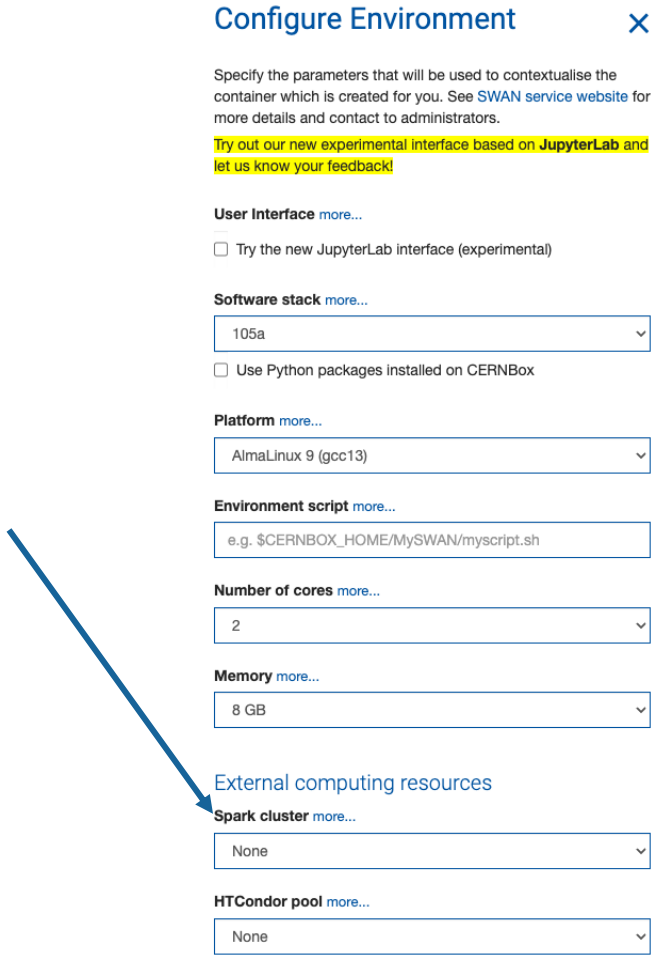


SWAN and Spark Architecture

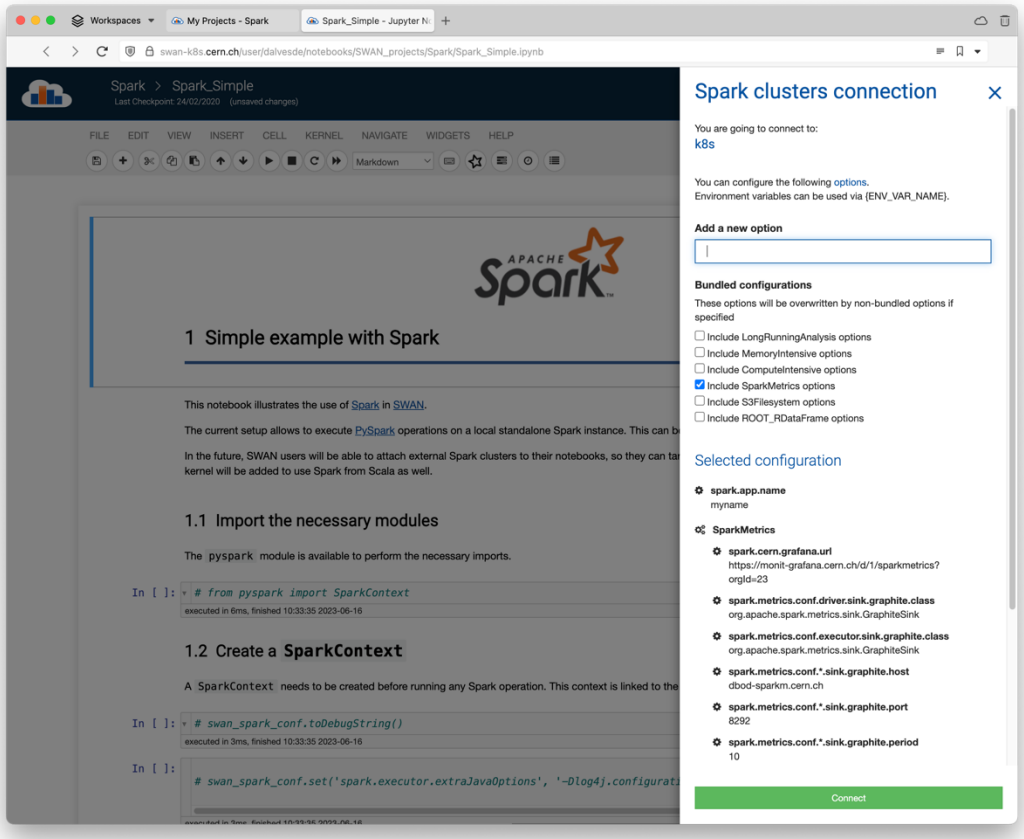


Spark Clusters

Cluster Name	Configuration	Primary Usage
analytix	48 nodes (Cores – 3456, Mem – 34.49TB, Storage – 19.84 PB)	General Purpose
nxcals	49 nodes (Cores – 2712, Mem – 24.34TB, Storage – 16.76 PB)	Accelerator logging (NXCALS) project dedicated cluster
Cloud containers	16 nodes (Cores 256, Mem – 1.87 TB, Storage – EOS)	General Purpose Compute ONLY



Spark Connector

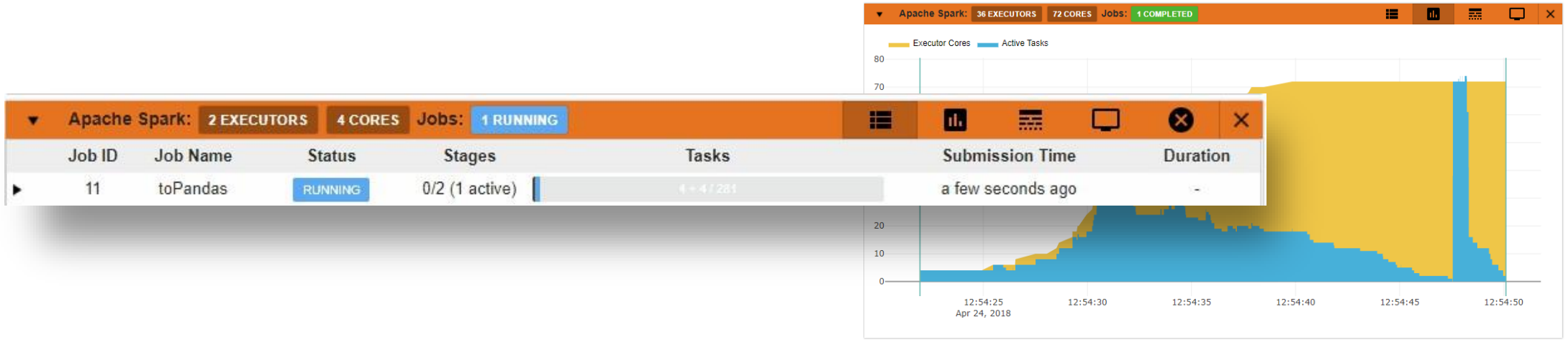


- **Spark Connector** – handling the spark configuration complexity
  - User is presented with Spark Session (Spark) and Spark Context (sc)
  - Ability to bundle configurations specific to user communities
  - Ability to specify additional configuration

Spark Monitor

- **Spark Monitor** – Jupyter notebook extension
  - For live monitoring of spark jobs spawned from the notebook
  - A graph showing number of active tasks & executor cores vs time
  - A timeline which shows jobs, stages, and tasks

Google Summer of Code



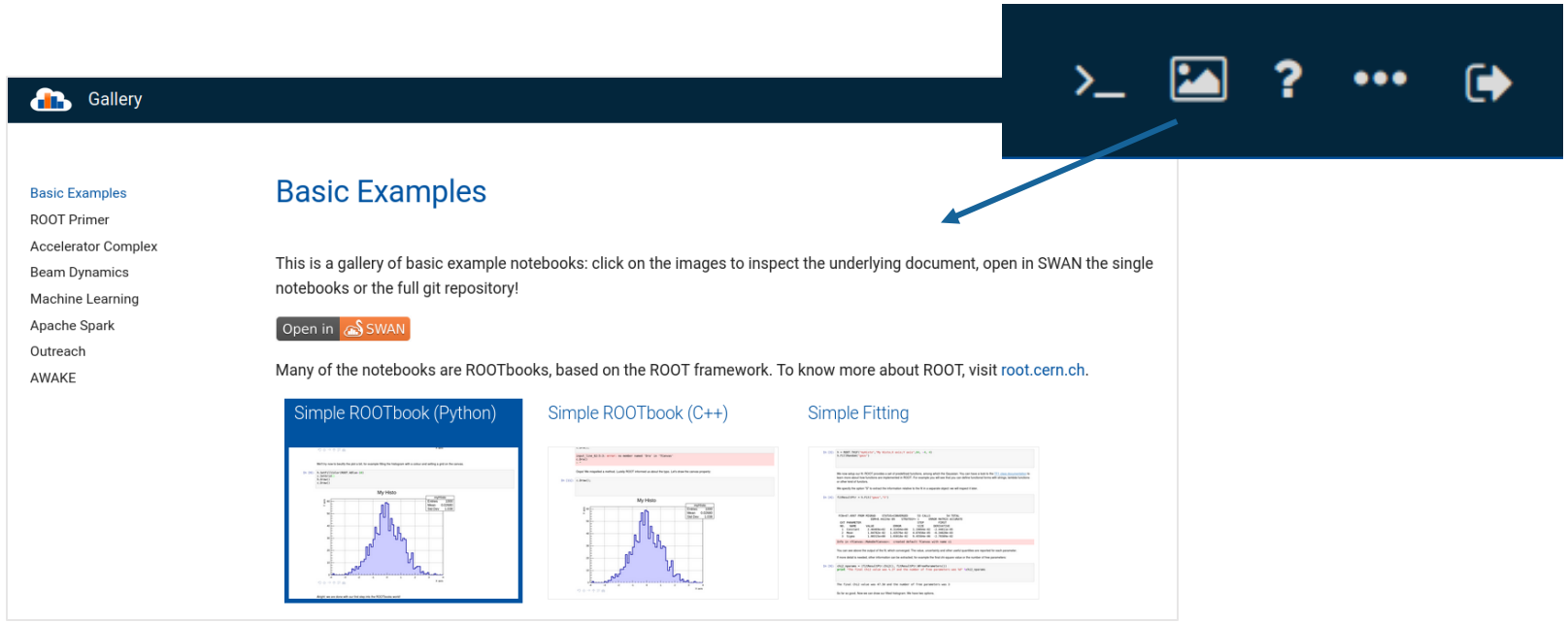
How to get started?

How to get started?

- A possible way to start using SWAN is by accessing some content (notebooks) that some other user created
- In order for a user to share content with others, SWAN offers several options:
  - For exploratory work, collaborative editing: use **sharing** feature (currently only available in the classic UI)!
  - To publish something that can be helpful to a broader audience: use **galleries** (https://swan-gallery.web.cern.ch/)!
  - For an event (training/course), for attendees to access its content: 

The galleries

- Galleries of sample notebooks for varied usage of SWAN
  - Quick way to be productive
  - Also accessible from https://swan-gallery.web.cern.ch/



Future work

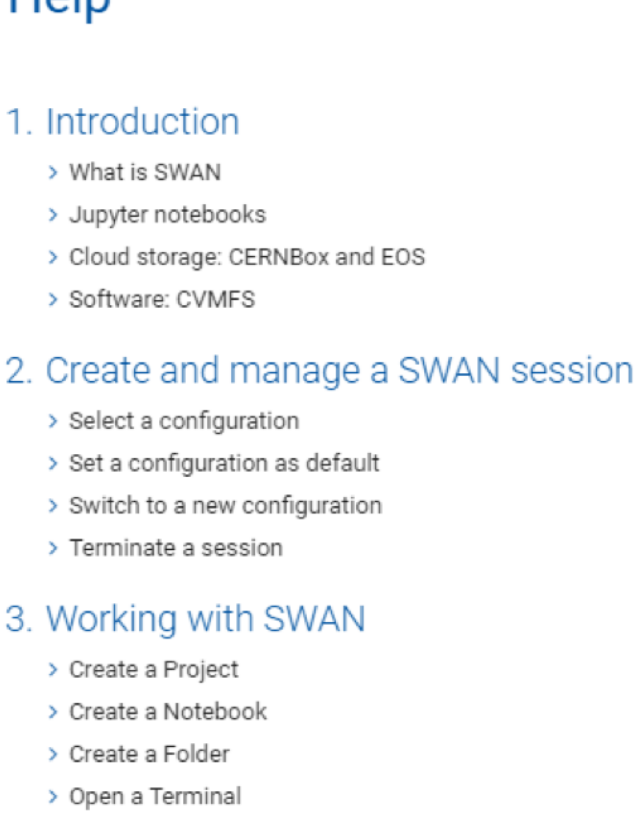
Future work

- Set **JupyterLab** as the default user interface
- Allow users to create **custom software environments** (i.e. independent of LCG stacks)
  - E.g. conda
- Support **experiment software stacks** for analysis
  - E.g. LHCb analysis environment
- Allow selection of **specific GPU models** when starting a session
- Deploy new **SWAN instance for ATS**
  - Exposed to devices in the Technical Network

Getting help & contact

How to get help?

- SWAN Community
  - https://cern.ch/swan-community
  - Find solution to the commonly encountered issues / questions on the usage of Jupyter notebooks, LCG releases, storage and Spark
  - Request improvements / new features to the service
  - E.g: How to install custom user packages
- Service Now
  - Report issues to the service
    - E.g: Unable to start a session
- Help on various features of the tool



Where to find us

- Contacts
  - swan-contact@cern.ch
  - http://cern.ch/swan
- Code repository
  - https://github.com/swan-cern

CERN

Introduction to SWAN

Thank you

Diogo Castro (diogo.castro@cern.ch)
Enric Tejedor (etejedor@cern.ch)
Pedro Maximino (pedro.maximino@cern.ch)