Integrated Science Platform at CERN

Enrico Bocchi
CERN – IT, Storage Group

Dec 16th, 2019

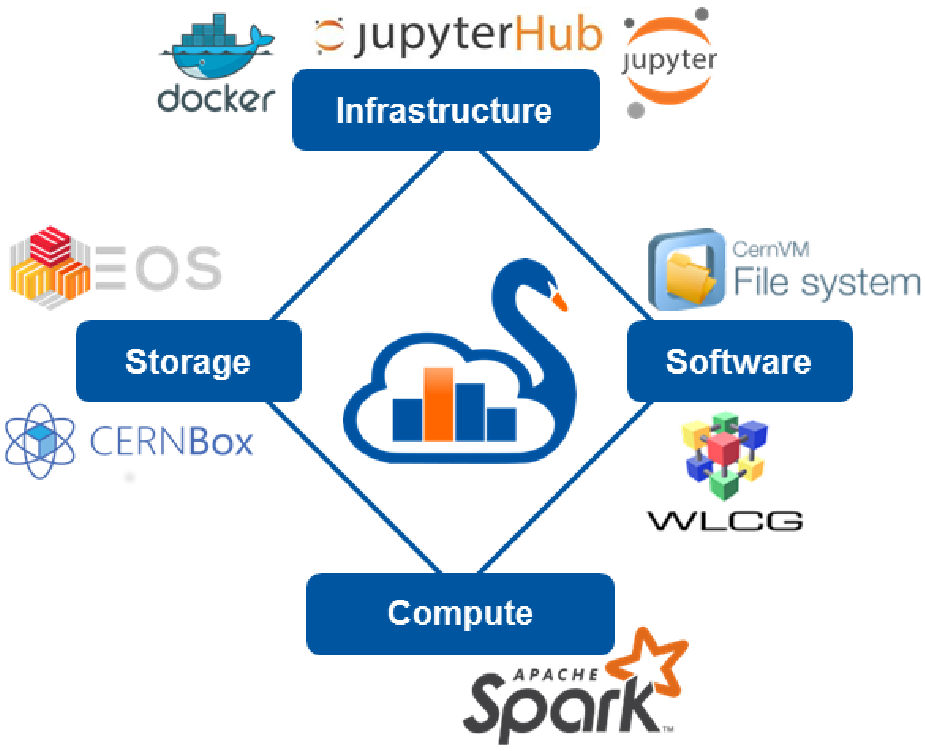
ESCAPE Tech Meeting, WP5

SWAN

The Jupyter Notebook Service at CERN

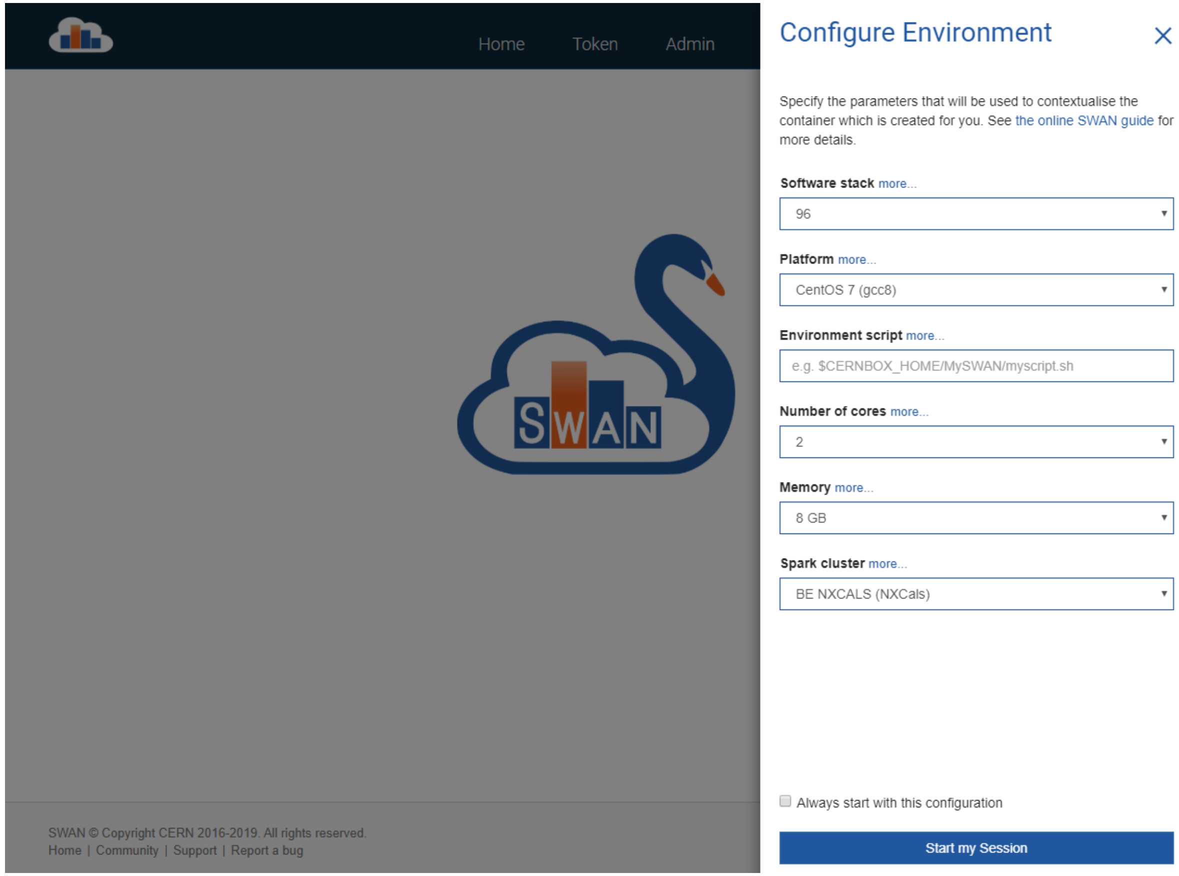
SWAN in a Nutshell

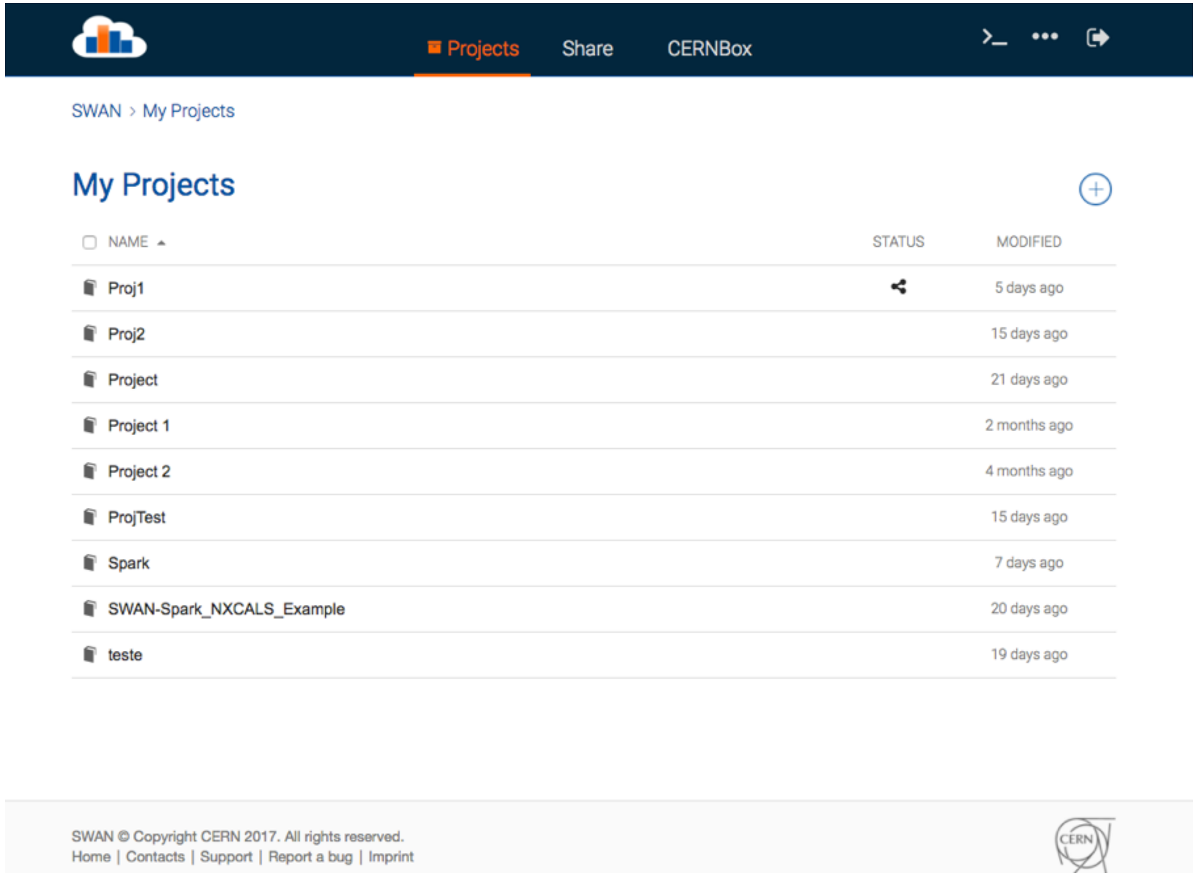
- CERN's Jupyter Notebook service
  - Based on upstream Jupyter Notebooks // JupyterHub
- Analysis only with a web browser
  - No local installation needed
  - Calculations, input data, and results "in the Cloud"
- Support for multiple analysis ecosystems and languages
  - Python, ROOT C++, R and Octave
- Easy sharing of scientific results: plots, data, code
- Integration with external resources
  - Storage, Software, mass processing power



SWAN in a Nutshell

SWAN: User Interface



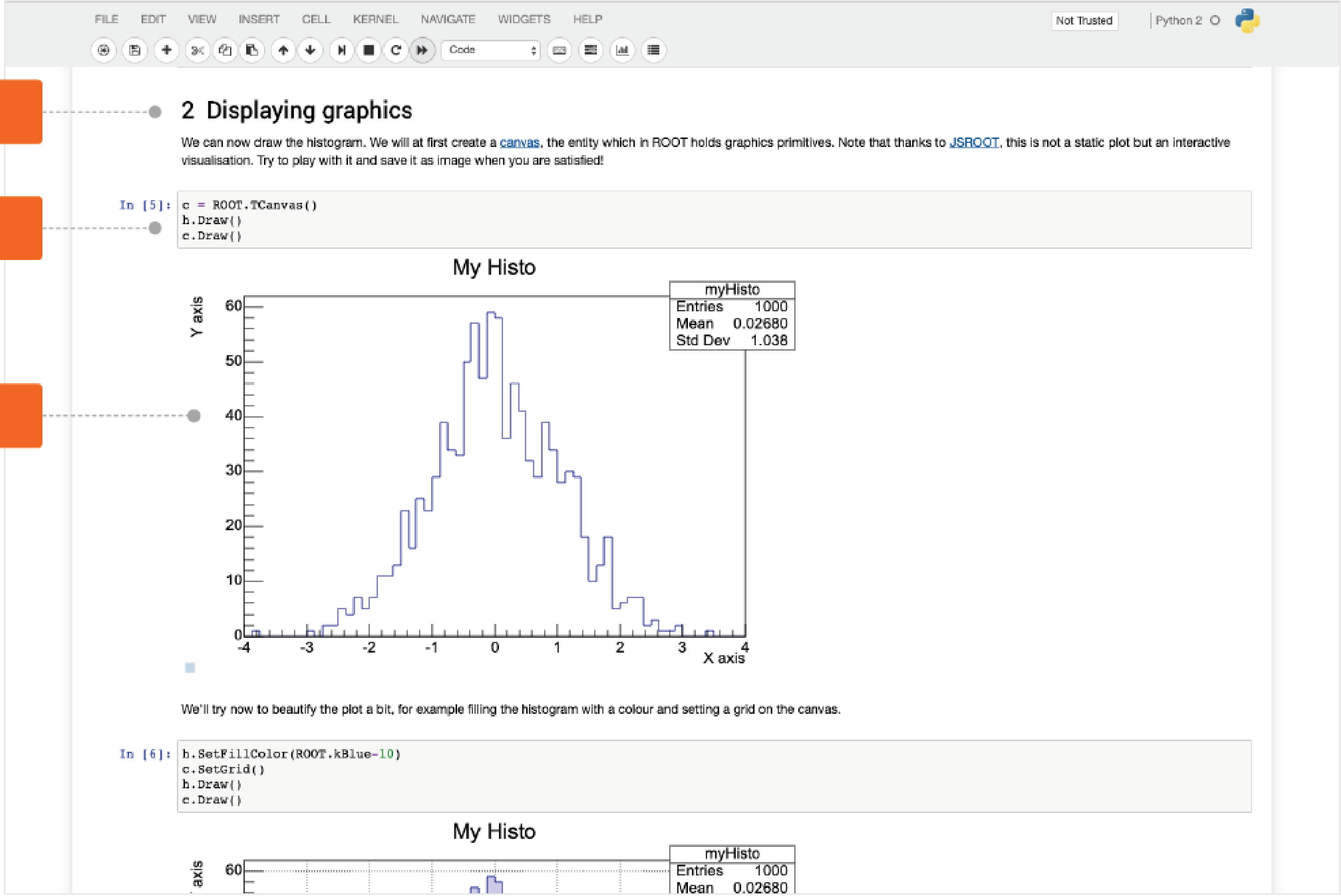


SWAN: User Interface

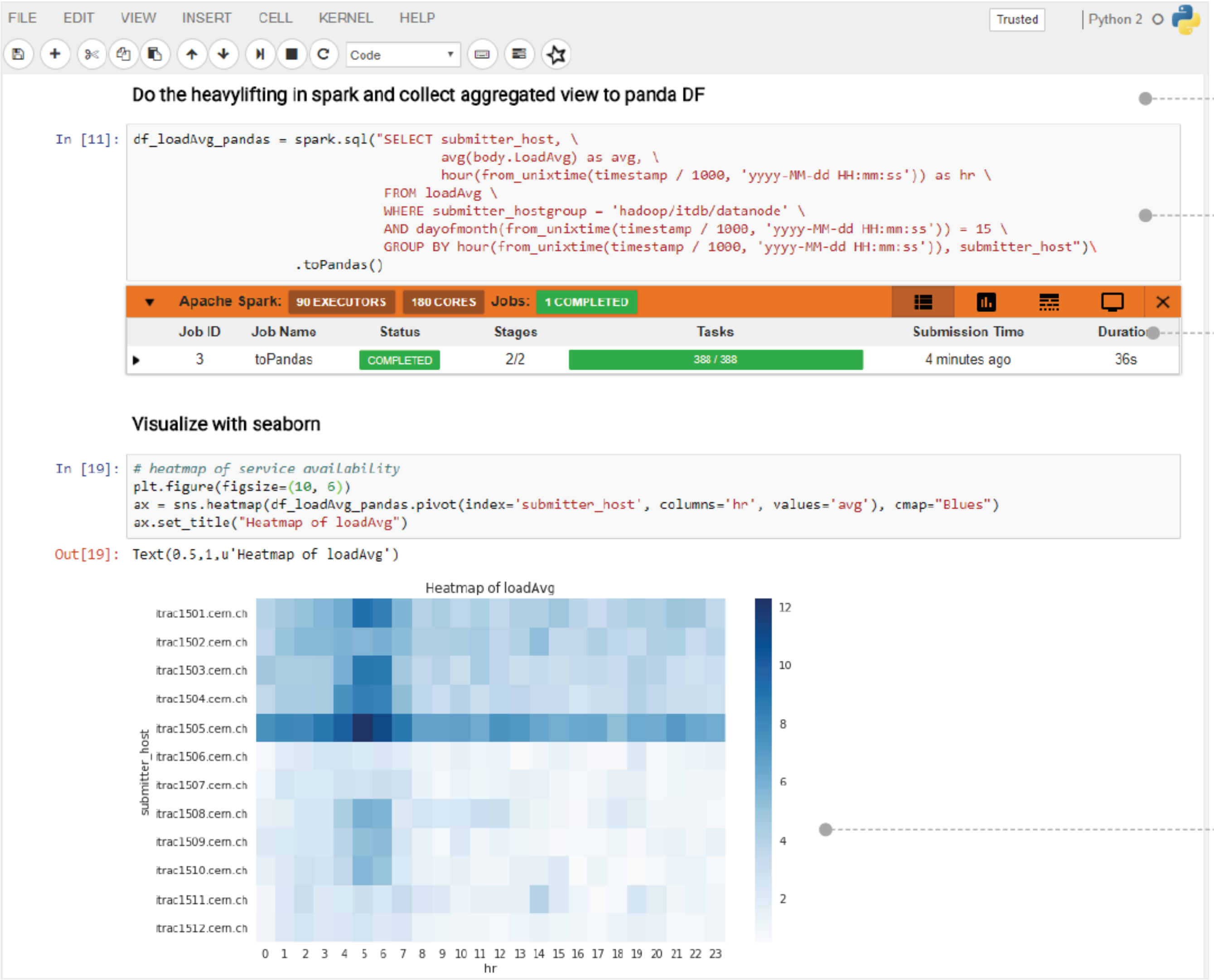
Text

Code

Graphics



SWAN: User Interface



Text

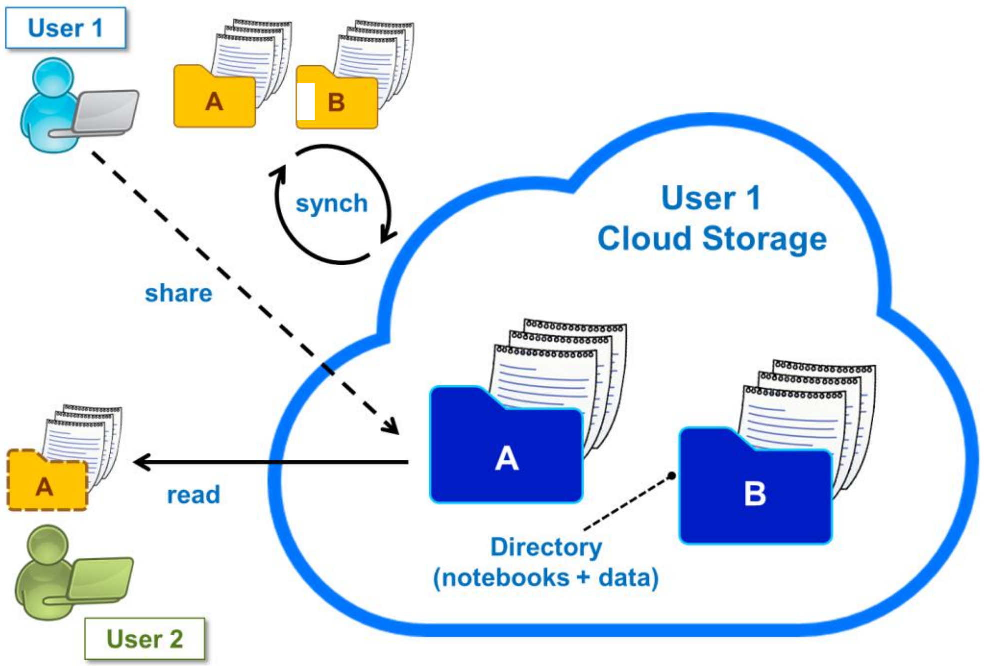
Code

Monitoring

Visualizations

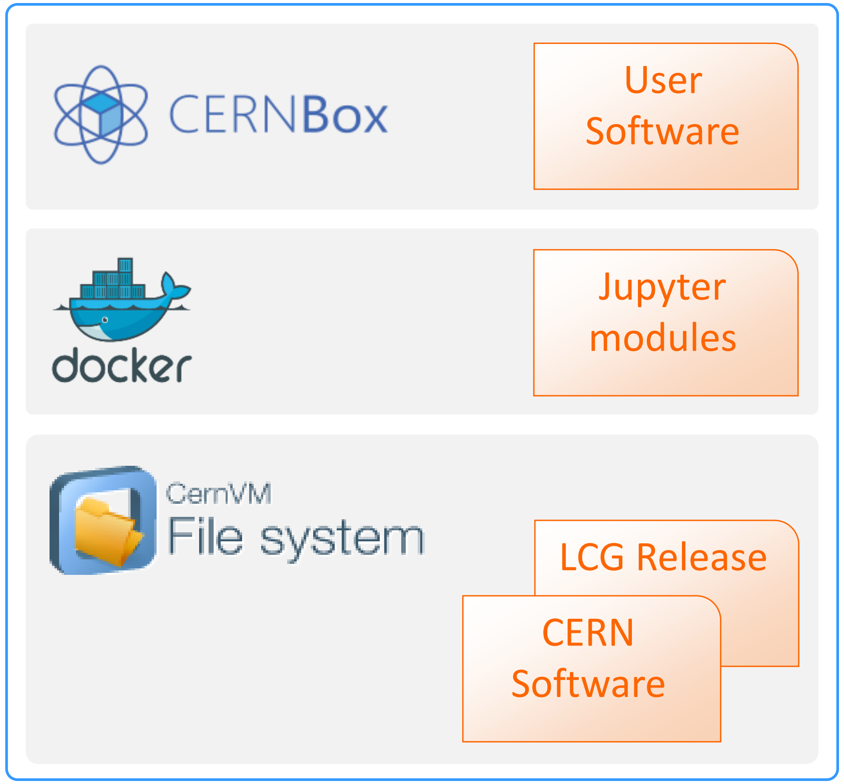
Storage: The Cloud as your Home

- CERNBox is SWAN's home directory
  - Based on EOS disk storage system
- Sync & Share
  - Files synced across devices and the Cloud
  - Collaborative analysis
- Sharing integration within SWAN UI
  - Users can share “Projects” (special kind of folder containing notebooks and other files)
  - Self contained



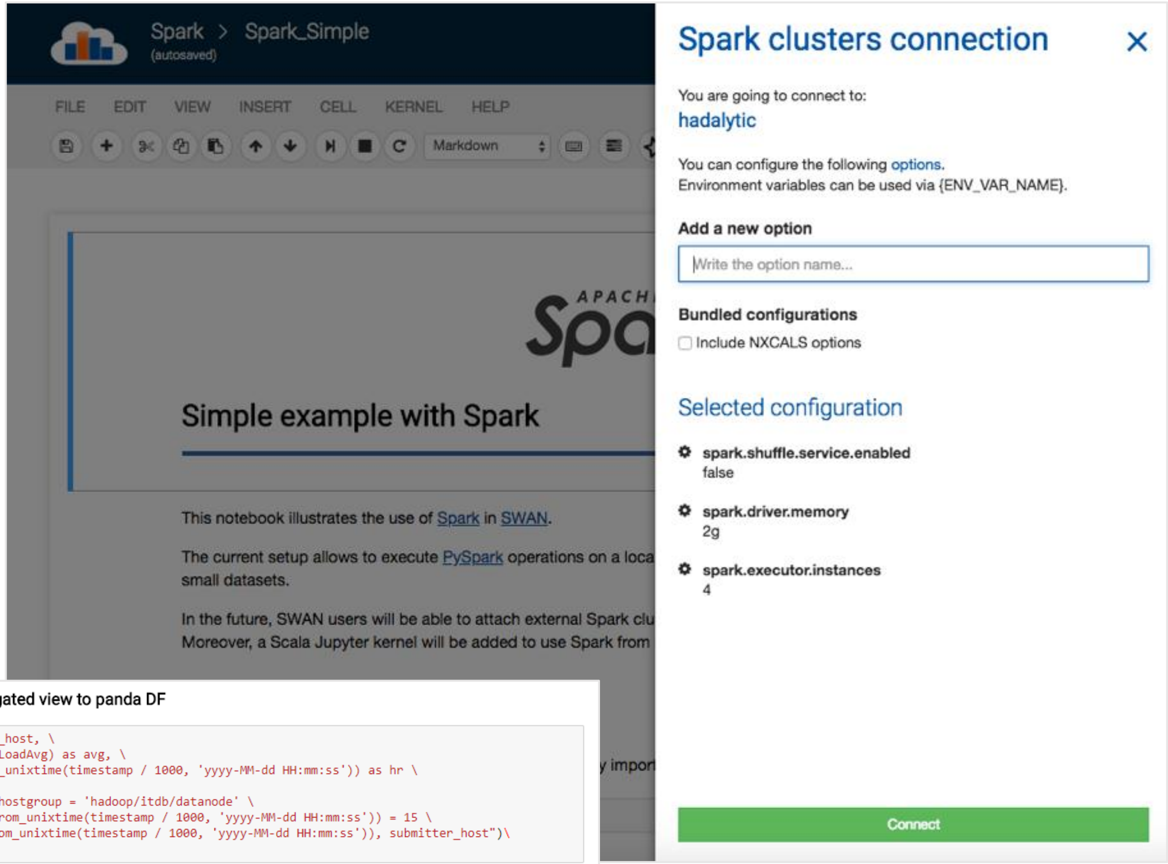
Software: Leverage on WLCG repositories

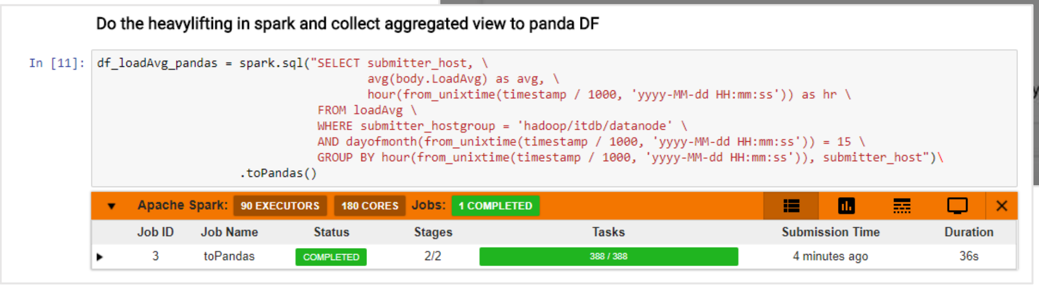
- Software distributed through CVMFS
  - Distributed read-only filesystem
  - ”LCG Releases” - pack a series of compatible packages
  - Reduced Docker Images size
  - Lazy fetching of software
- Possibility to install libraries in user cloud storage
  - Good way to use custom/not mainstream packages
  - Configurable environment



Mass Processing: Integration with Spark

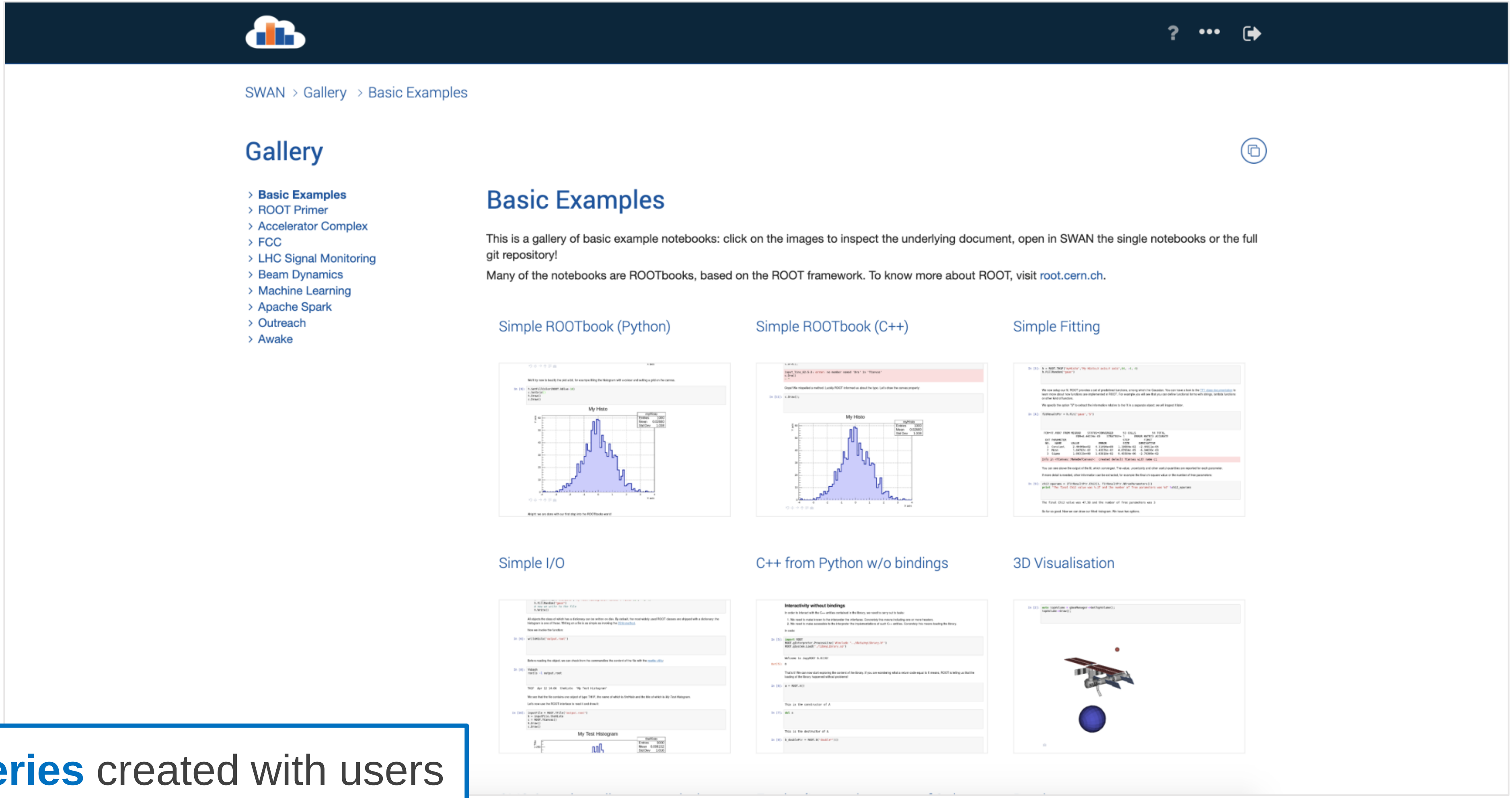
- Connection to Apache Spark Clusters
  - Spark: general purpose distributed computing framework
- Same environment across platforms (local/remote)
  - Software - CVMFS
- Graphical Jupyter extensions developed
  - Spark Connector
  - Spark Monitor





CERN

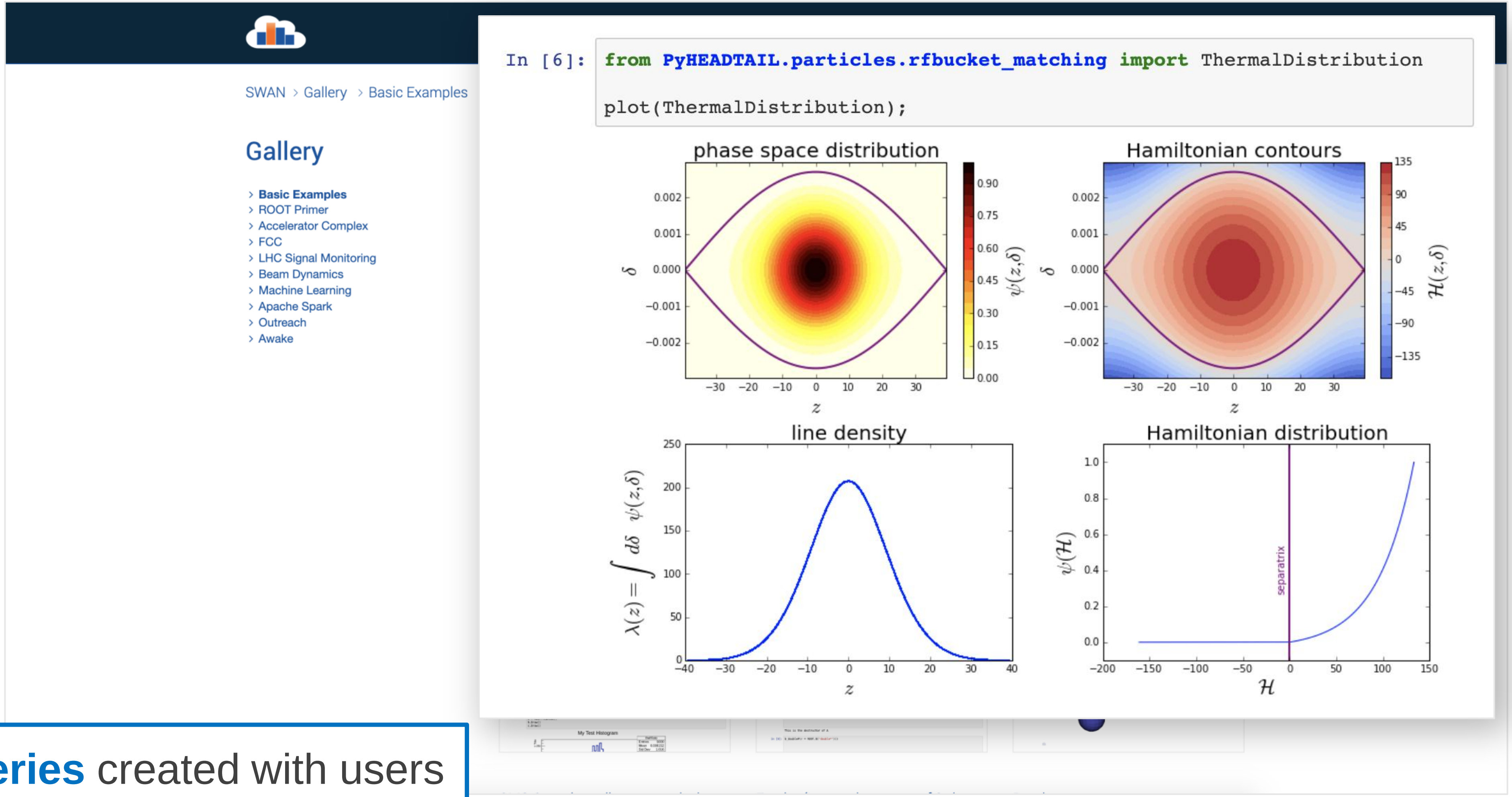
SWAN Galleries



70+ galleries created with users
10 categories

https://swan.web.cern.ch/content/basic-examples

SWAN Galleries

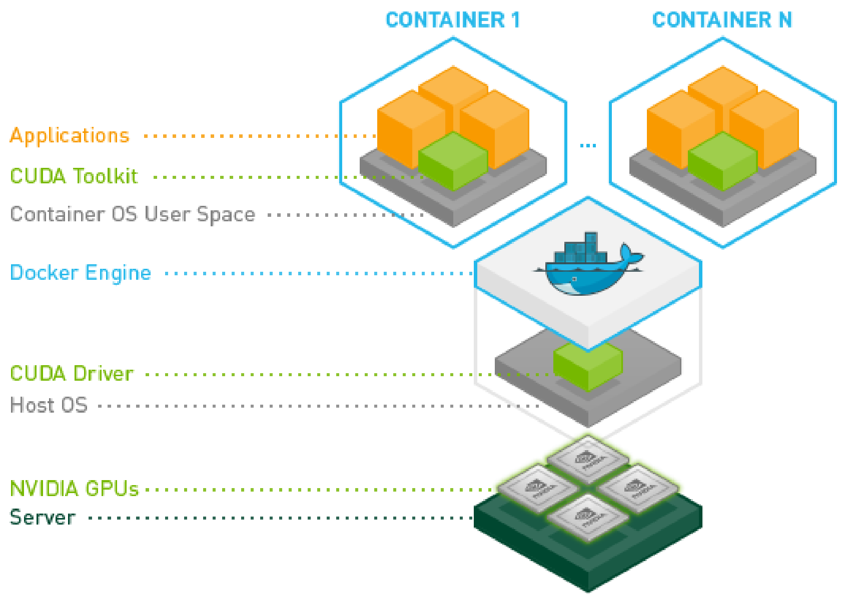


70+ galleries created with users
10 categories

https://swan.web.cern.ch/content/basic-examples

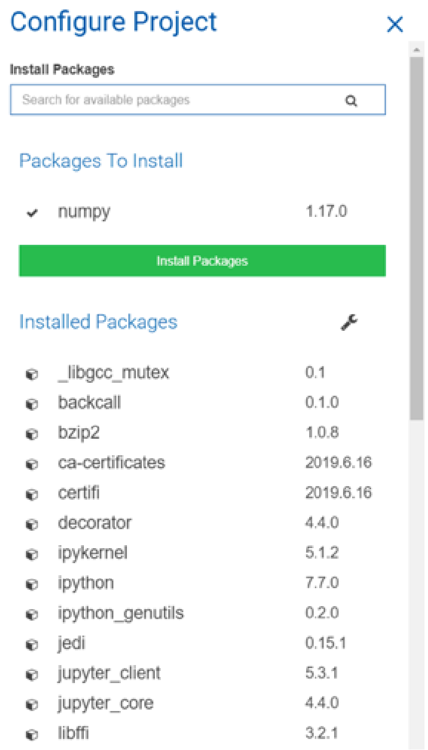
Upcoming: NVidia GPU Support

- Exploitation of container technologies to provide support for NVidia GPUs
- Prototype server for testing purposes
  - NVidia Tesla V100 PCIe 32GB
- All the packages are provided by CVMFS
  - Including CUDA enabled machine learning software stack
  - TensorBoard for interactive monitoring



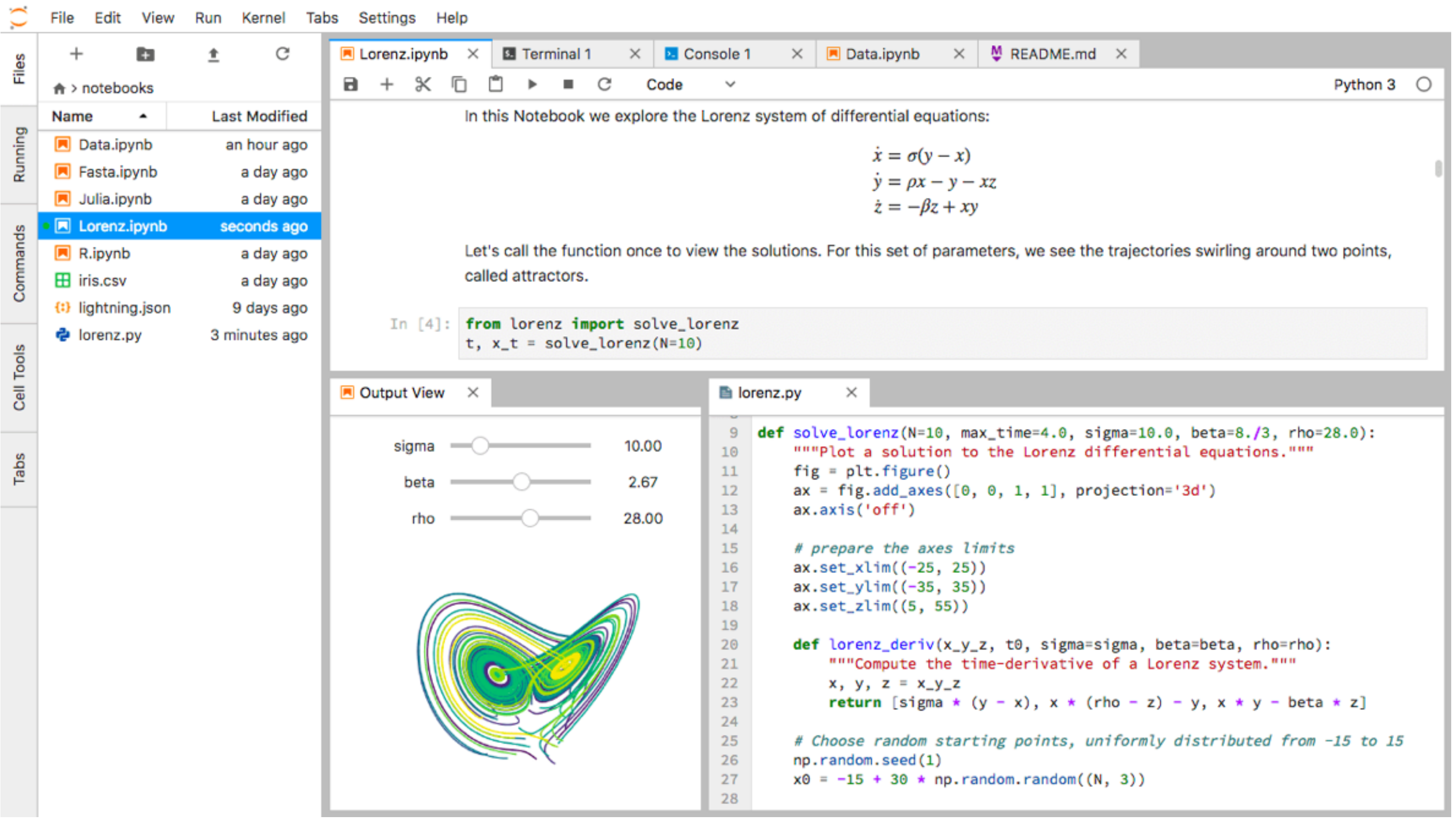
Upcoming: Configurable SW Environment

- Adding support for Conda environments
  - Linked to Projects
  - Sharable
- Easy installation of extra packages
  - Clone/import Projects and install the software automatically
- Custom images with SWAN
  - For higher customization of software environment
  - BYO Docker image
- Flexibility of Binder +
  Integrated Environment of SWAN



Upcoming: Jupyterlab

- Next-generation interface for Project Jupyter
  - “IDE-like” environment
  - Concurrent editing
- Next steps: integration of current extensions
  - SWAN Projects
  - CERNBox sharing integration
  - Spark Connector and Monitor
  - …



ScienceBox

SWAN, CERNBox, EOS in Docker containers

ScienceBox

- Self-contained Docker-based software package

EOS + CERNBox + SWAN + CernVM File system

docker

ScienceBox

- Self-contained Docker-based software package

EOS + CERNBox + SWAN + CernVM File system

docker

docker-compose

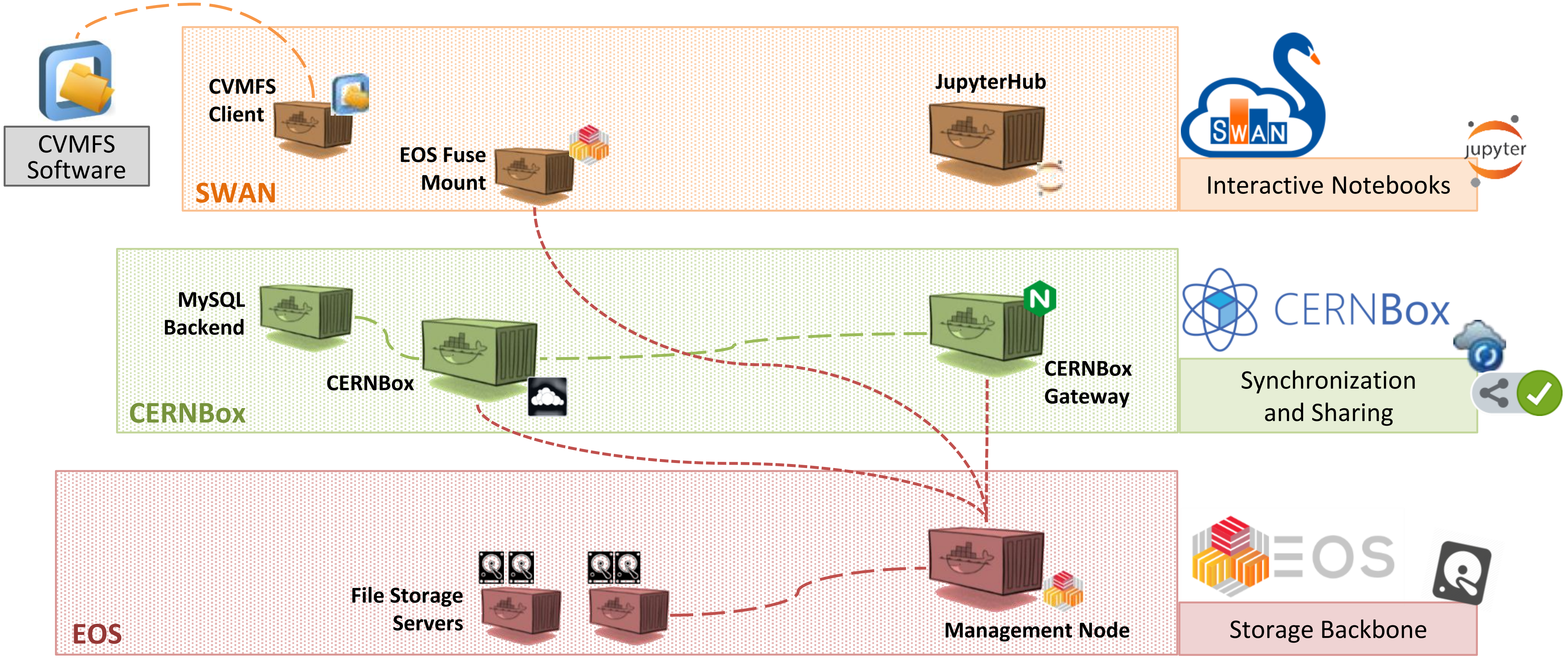
kubernetes

One-Click Demo Deployment
• Single-box installation • Download and run in 5 minutes https://github.com/cernbox/uboxed

Production-ready Deployment
• Scale out service capacity • Tolerant to node failures https://github.com/cernbox/kuboxed

CERN

ScienceBox Architecture



Use Cases: Up to University

- Allow students in high-schools to adopt tools used in science
  - SWAN – Full data analysis ecosystem in a web browser
  - CERNBox – Cloud storage for easy sharing and access form any device

Up To University

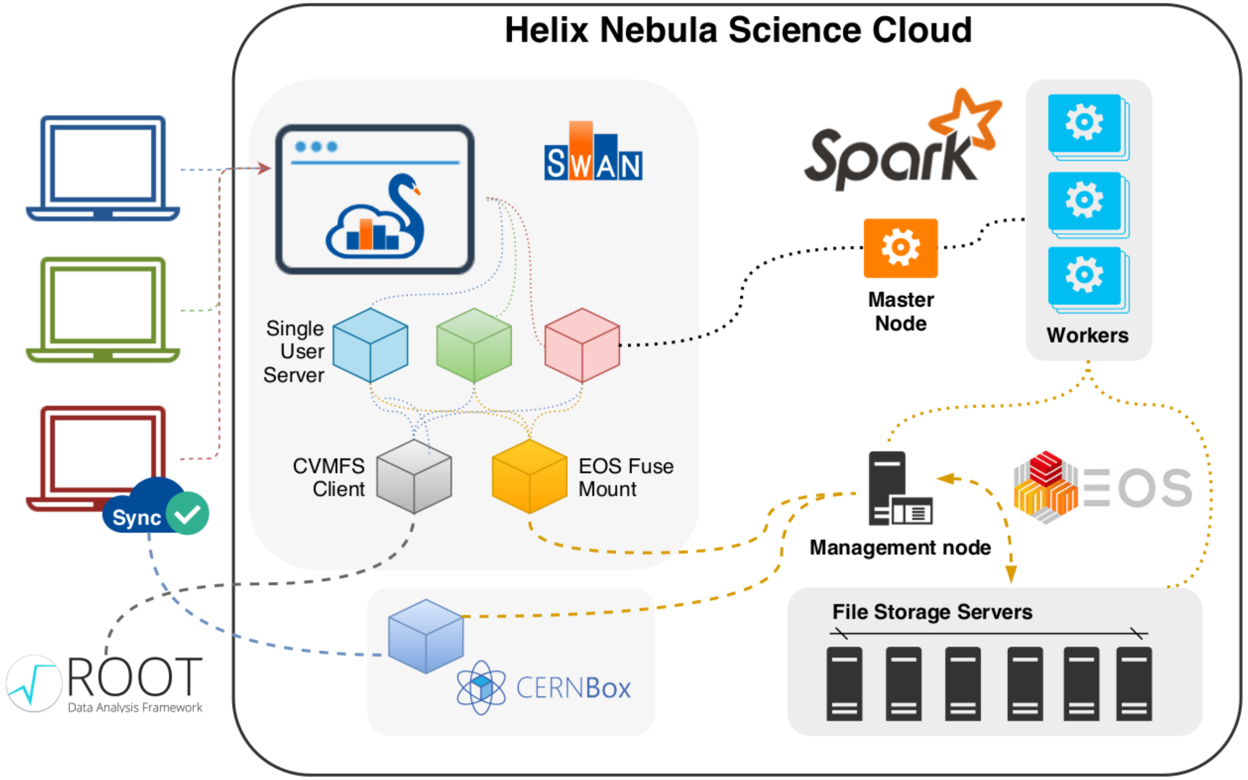
- ScienceBox in production for Up2U users for 1.5 years
  - Deployed at Poznan Supercomputing and Networking Center, Poland
  - Kubernetes on VMs, Ceph volumes for persistent storage
- Pilot service at CERN – http://up2u.cern.ch
  - CERNBox and SWAN on Kubernetes VMs
  - EOS on VMs and bare metal disks

openstack

CERN

Use Cases: Totem Analysis on Helix Nebula Cloud

- Deployment on commercial cloud
  - 2000+ CPUs
  - 10+ TB memory
  - Virtually unlimited block storage
- ScienceBox with Apache Spark for massive computations
- Full TOTEM Analysis
  - Dataset: 4.7 TB, 1153 files
  - Data imported via xrootd
  - Results synchronized back via CERNBox client



Big Data Tools and Cloud Services for High Energy Physics Analysis in TOTEM Experiment - V. Avati et al.
https://ieeexplore.ieee.org/document/8605741

More External Sites and Collaborators

- External SWAN deployments inspired by ScienceBox
  - Australia's Academic and Research Network (AARNET)
  - SURFSARA, The Netherlands
  - Joint Institute for Nuclear Research (JINR), Russia
  - Academia Sinica Grid Computing Centre (ASGC), Taiwan

Opportunities for Collaboration

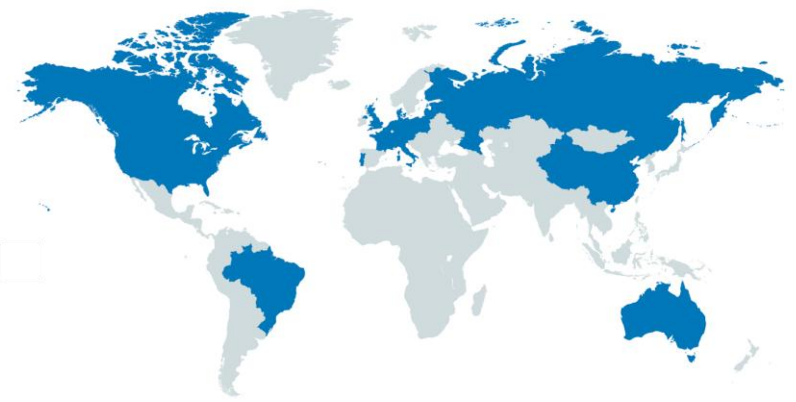
CS3MESH

- EU-funded project (coordinated by CERN)
  - 6M EUR, 12 partners, 2020-2022
- Goal: Global collaborative environment for research
  - Share documents, files, projects, data, …
  - Connected Application Hubs
  - Data Science Environments ➔ **SWAN**
- Federation of existing CS3 sites
  - 30+ sites (e.g. CERNBox, DesyBox, Universities, …)
  - 300K+ users
  - cs3community.org

EUROPEAN OPEN SCIENCE CLOUD

CS3

Cloud Storage Services
for Synchronization and Sharing



CERN

CS3MESH Objectives

Collaborative Workflows

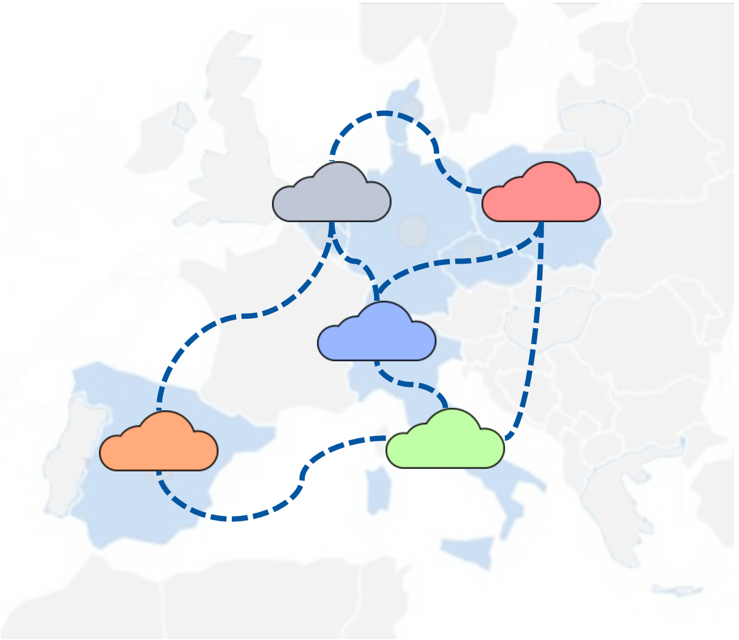
- Share, access, synchronize
- Metadata & Tagging, Open Data (OpenAIRE, Zenodo, …)
- Data Science: Jupyter Notebooks (SWAN, …)
- Concurrent editing
- On-demand data transfers (Direct, FTS, DTN, Rucio, …)

Interoperability

- Add thin layer on top of existing services
- Use existing fabric
- Use existing standards
  - Introduce new APIs only if needed
- Close collaboration with industry
- Integrate into upstream products

ScienceBox for CS3MESH

- ScienceBox is the reference platform for CS3MESH for distribution and deployment of cloud software
  - SWAN will become part of the core service for future European Open Science Cloud
- Developer community & upstream
  - Working together on SWAN
- Benefits for SWAN users at a scale
  - Share SWAN projects beyond the CERN borders
  - Work easily with your experiment collaborators inside/outside CERN



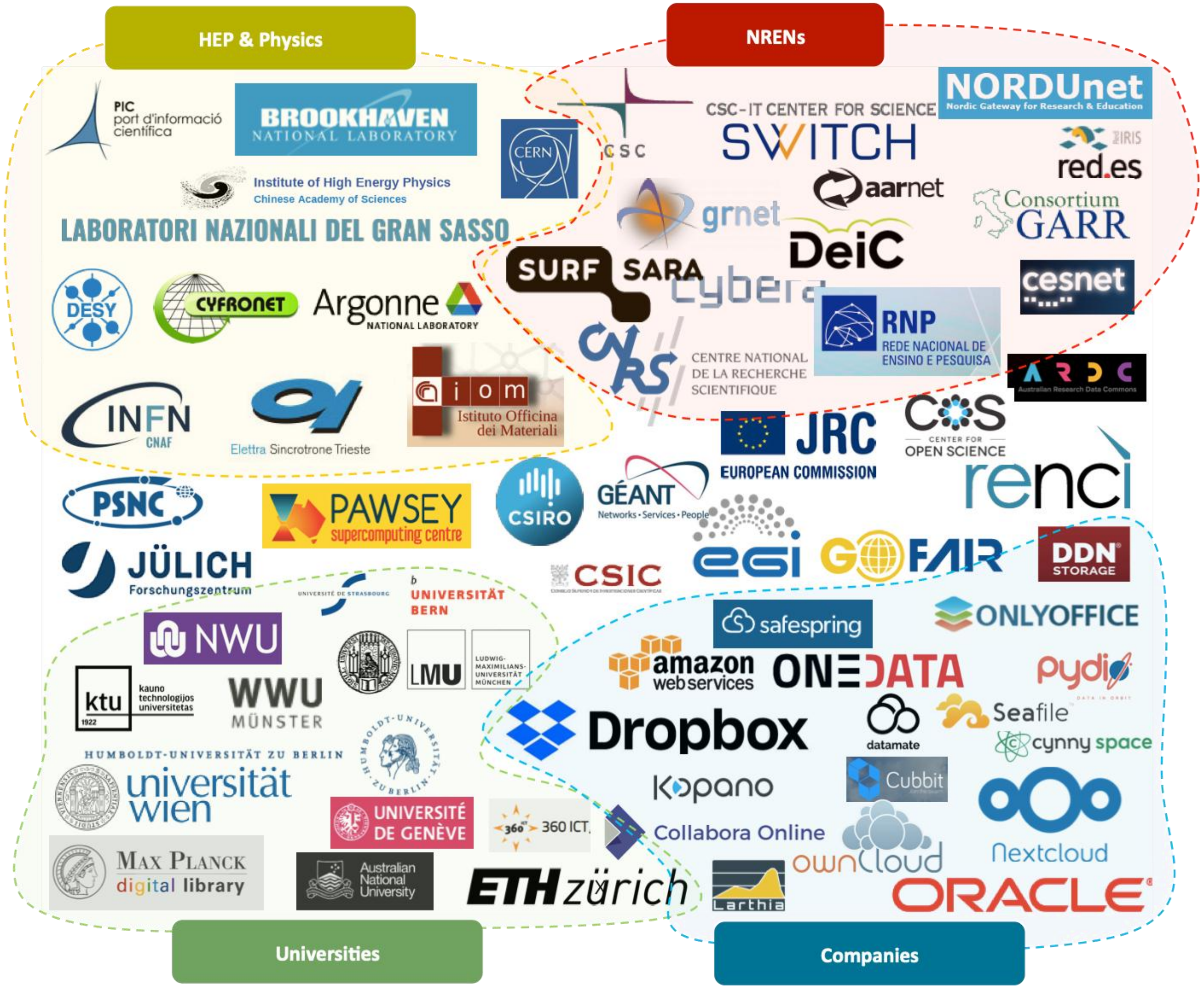
CS3 Workshop

https://cs3.deic.dk/

Collabora Online DDN AI · BIG DATA · HPC Kopano NetApp Nextcloud ONLYOFFICE ownCloud Seafile

CERN

CS3 Community



CERN

Where to find us

Where to find us

- Swan and Galleries
  - https://swan.web.cern.ch/
  - https://swan.web.cern.ch/content/basic-examples
- Science Box
  - https://sciencebox.web.cern.ch/
- CERNBox and EOS
  - https://cernbox.web.cern.ch/
  - https://eos.web.cern.ch/

Demo

Analysis of Cinemas in Geneva