SWAN
Service for
Web-based ANalysis

https://swan.web.cern.ch

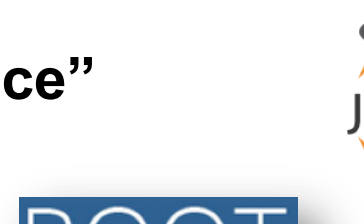
E. Tejedor, D. Piparo, P. Mató
on behalf of the ROOT team – EP-SFT
L. Mascetti, J. Moscicki, M. Lamanna – IT-ST

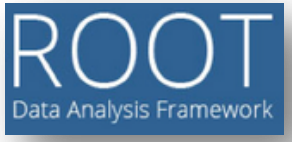
EP-SFT meeting

09/05/2016

Data analysis with ROOT "as a service"

Interface: Jupyter Notebooks





Goals:

- Use ROOT only with a web browser
  - Platform independent ROOT-based data analysis
  - Calculations, input and results "in the cloud"
- Allow easy sharing of scientific results: plots, data, code
  - Storage is crucial
- Simplify teaching of data processing and programming
- Potential integration with other analysis ecosystems: R, Python, …

The Service Interface

Notebook: A web-based interactive computing interface and platform that combines code, equations, text and visualisations.

Many supported languages: Python, Haskell, Julia, R … One generally speaks about a "kernel" for a specific language

In a nutshell: an "interactive shell opened within the browser"

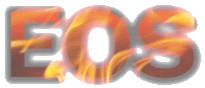
Also called:
"Jupyter Notebook" or "IPython Notebook"

The Service Backend

SWAN relies on production technologies at CERN:

- Authentication with CERN credentials (*IT-DI-CSO*)
- Infrastructure: virtual machines in OpenStack Cloud (*IT-OIS*)
- Software distribution (*EP-SFT, IT-ST*): CVMFS

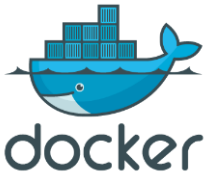
- Storage access (*IT-ST*): CERNBox, EOS
  - All data potentially available!



Plus some external technologies:

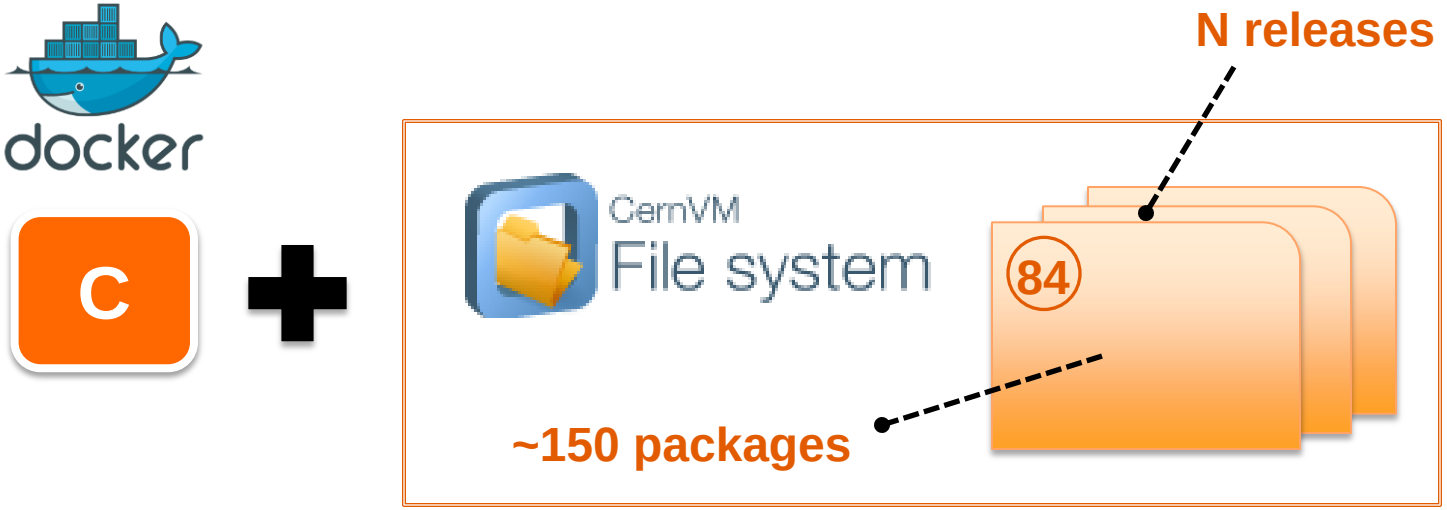
- JupyterHub

- Docker

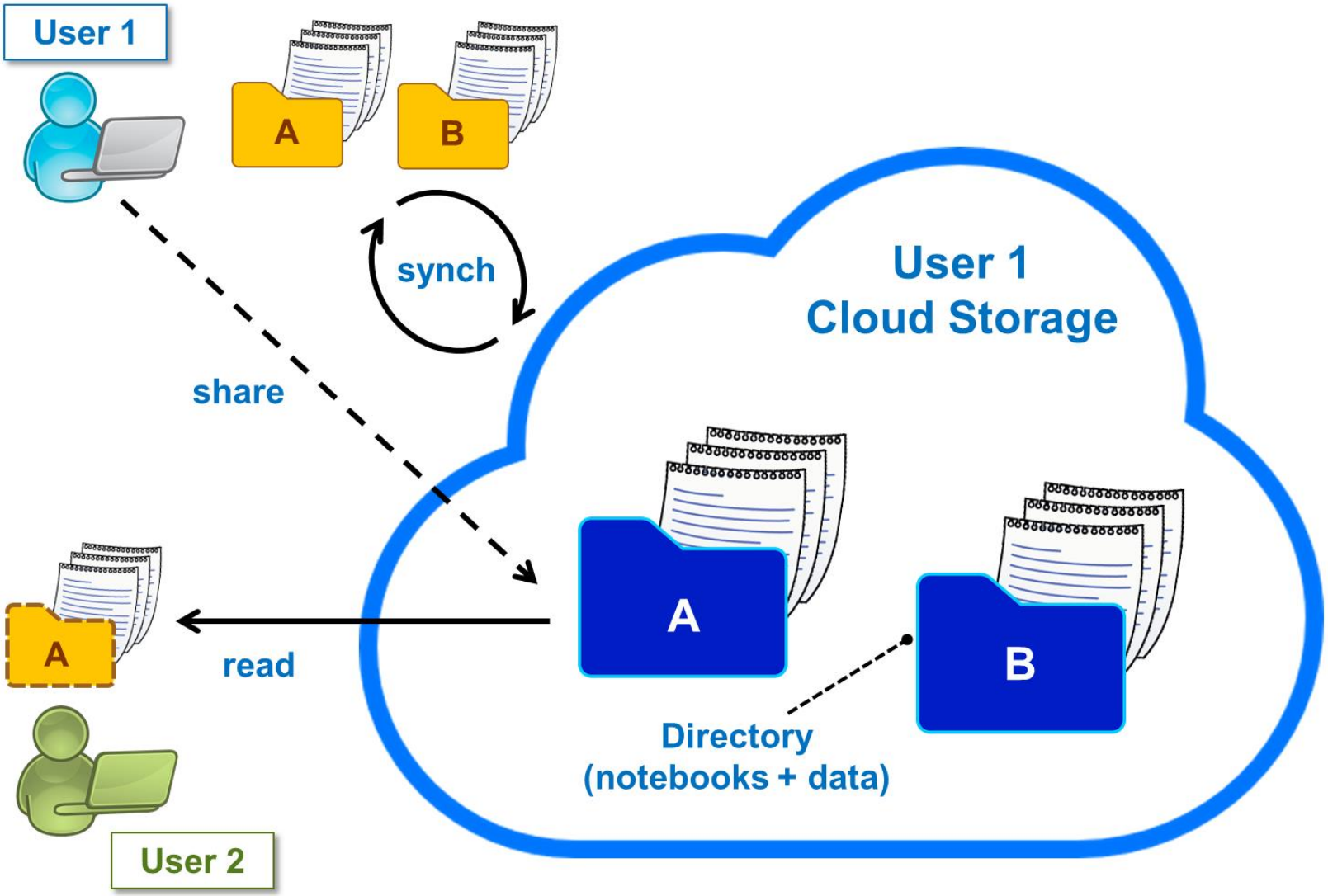


CVMFS + Docker

- Configure the software environment for a production service:
  - Docker: thin image
  - CVMFS: configurable environment via “views”
- Solves the problem of managing big images



Sync & Share


User 1
A
B
synch
User 1
Cloud Storage
share
A
B
A
read
Directory
(notebooks + data)
User 2

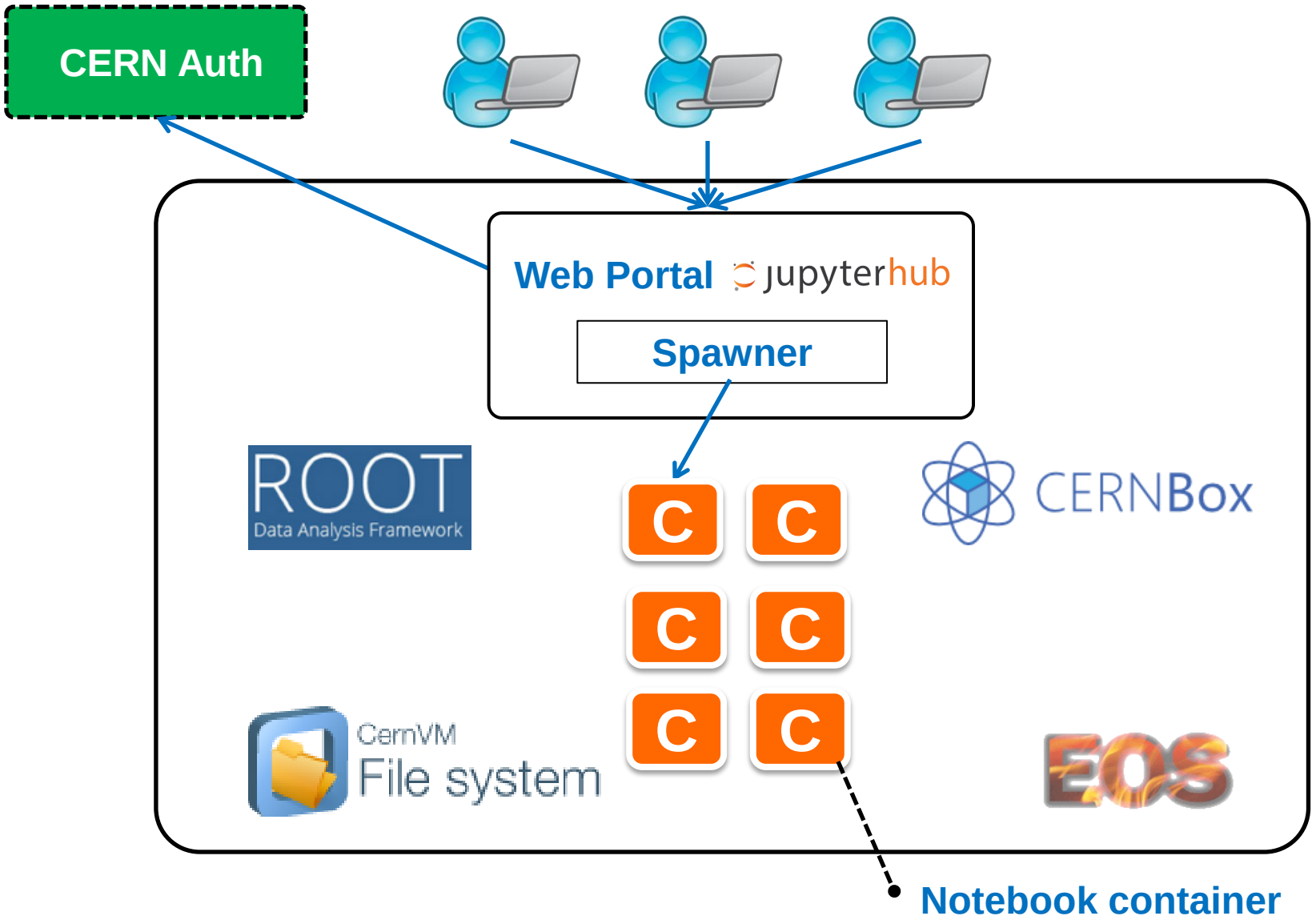
Pilot Service

- Released two weeks ago

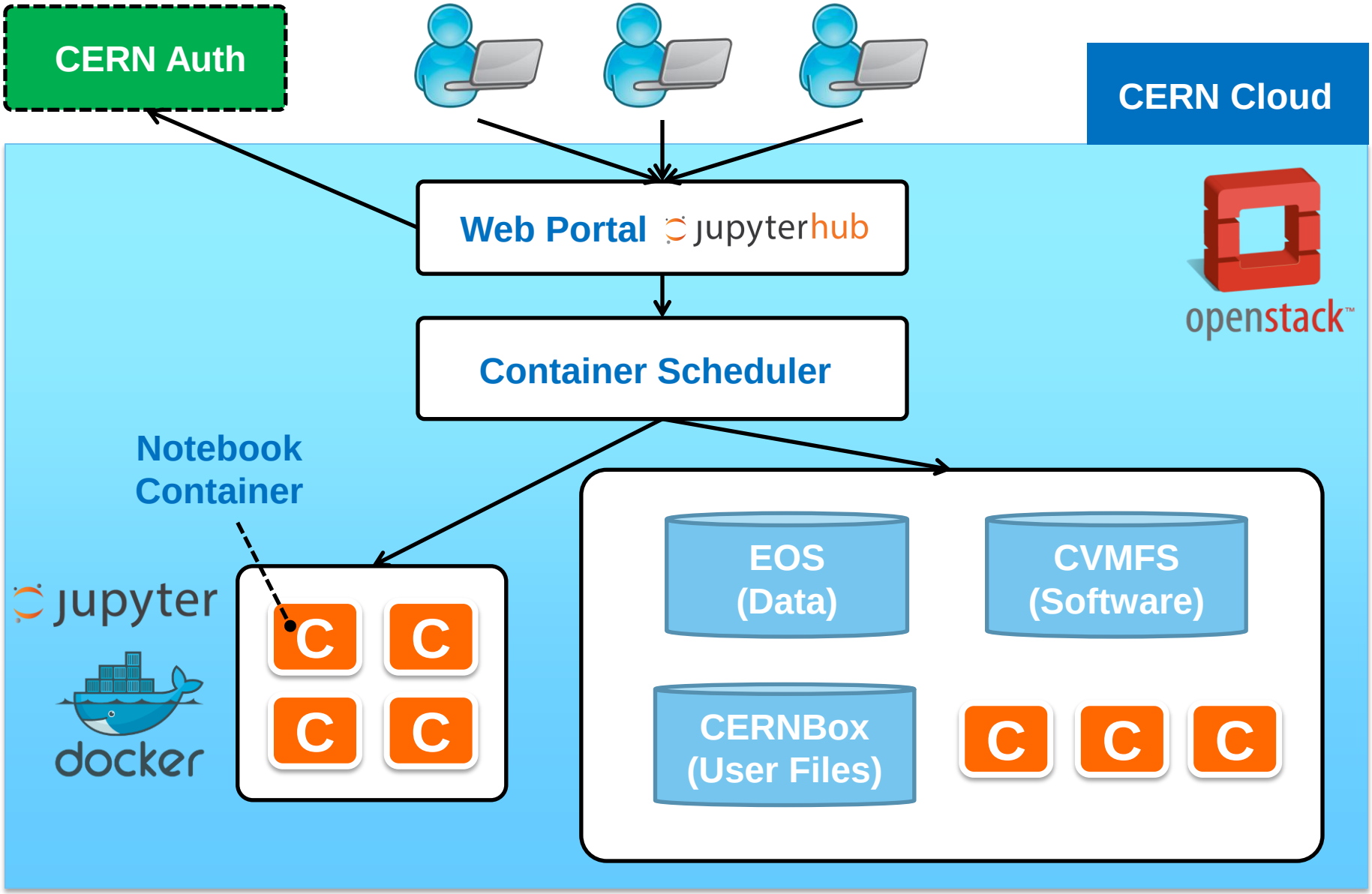
https://swan-beta2.cern.ch

- In beta testing phase: ~50 users, growing
  - Feedback from users already integrated
  - Will announce a second “release” today
- Automated configuration
  - Can create more instances if necessary
- Contributed to improve the performance and stability of the current EOS
- Access to Opendata, HEPData

Pilot Service: Architecture



Scaling out


CERN Auth
CERN Cloud
Web Portal
jupyterhub
Container Scheduler
openstack
Notebook
Container
jupyter
docker
C
C
C
C
EOS
(Data)
CVMFS
(Software)
CERNBox
(User Files)
C
C
C

The Demo

Summary, Plans

- First pilot service available for beta testing
  - ROOT C++ flavour integrated
  - CVMFS for software distribution
  - EOS mass storage + CERNBox synchronisation
  - Your feedback is very much appreciated!!
- Future work:
  - Advertisement
  - Improve experience with storage: response time, sharing
  - Exploit external resources (e.g. Spark clusters)
  - TMVA – ROOTbooks integration (GSoC)
  - Investigate CERN's container service
  - Open a production instance this summer