SWAN Elements Relevant for Analysis Preservation

www.swan.cern.ch

D. Piparo (EP-SFT) for the SWAN Team

The Concept

Provide at CERN "Data Analysis as a Service"

Interface chosen: Jupyter notebooks





- Do analysis only with a web browser
  - Platform independent , "in the Cloud"
- Allow easy sharing of scientific results: plots, data, code
  - Storage is crucial: mass & synchronised
- Simplify teaching of data processing and programming
  - Here SWAN shines!
- Ease reproducibility of results
- Access to larger computational resources
- Potential integration with several analysis ecosystems: R, Python, C++, Java…
  - Scientific software is crucial

The Notebook

- Interactive computing platform
- Mix code, documentation, plots, shell scripts
- Not enough alone to guarantee reproducibility
  - But provides useful tools to achieve it!

ROOT Notebook Functionalities | Control Panel | Logout

File Edit View Insert Cell Kernel Help | Python 2

Code | Cell Toolbar: None

Welcome to the Notebook Technology

This is a markdown cell. You can add LaTex code: $\sum_{n=-\infty}^{\infty} |x(n)|^2$

In [1]:
```
def thisFunction():
    return 42
```

In [2]:
```
thisFunction()
```

Out[2]: 42

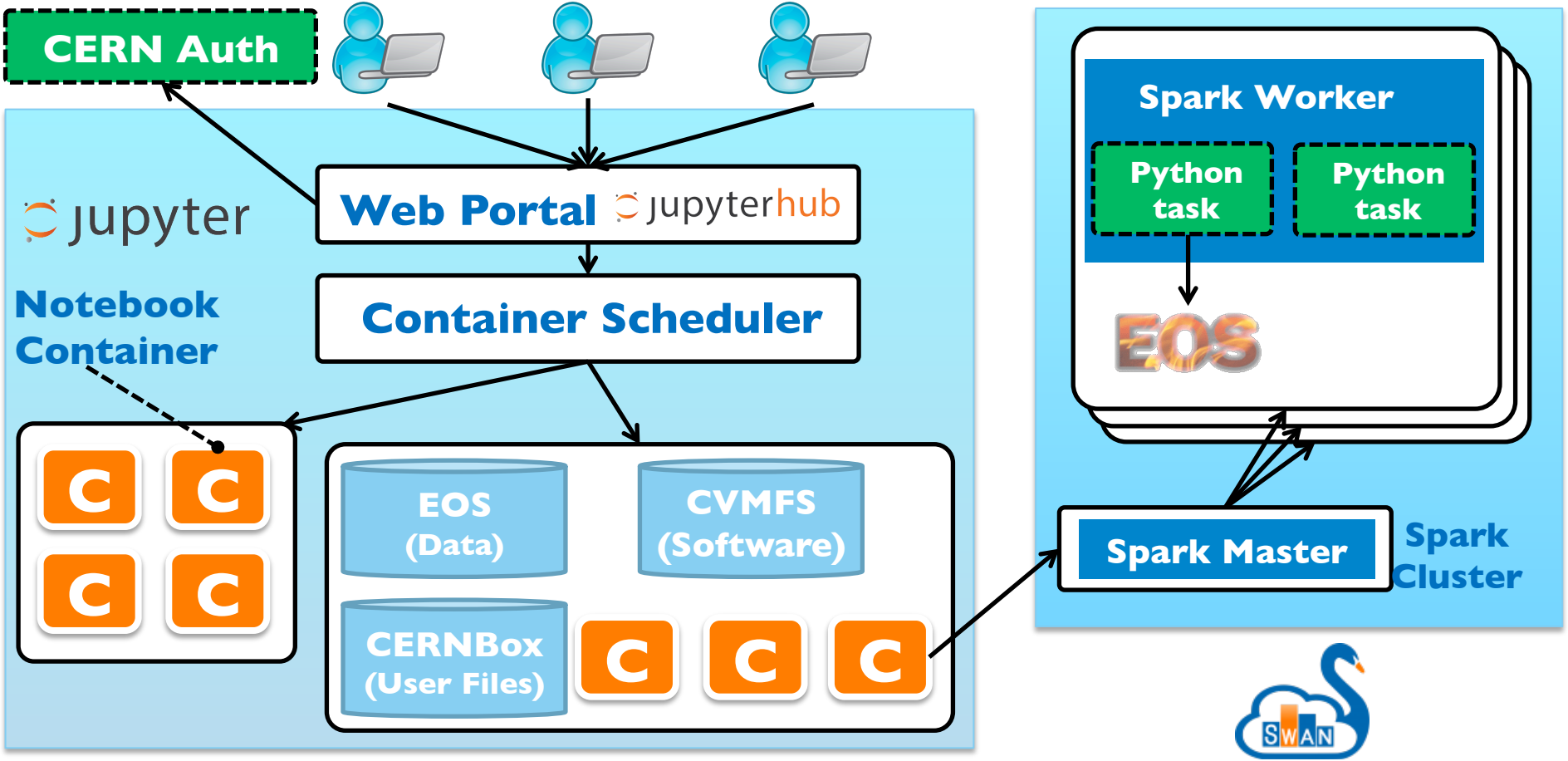
In [3]:
```
%%bash
curl rootaasdemo.web.cern.ch/rootaasdemo/SaasFee.jpg \
> SF.jpg
```

```
  % Total    % Received % Xferd  Average Speed   Time
  Time     Time  Current
                                 Dload  Upload   Total
  Spent    Left  Speed
100  128k  100  128k    0     0  2731k      0 --:--:--
--:--:-- --:--:-- 2787k
```

In [4]:
```
from IPython.display import Image
Image(filename="./SF.jpg",width=225)
```

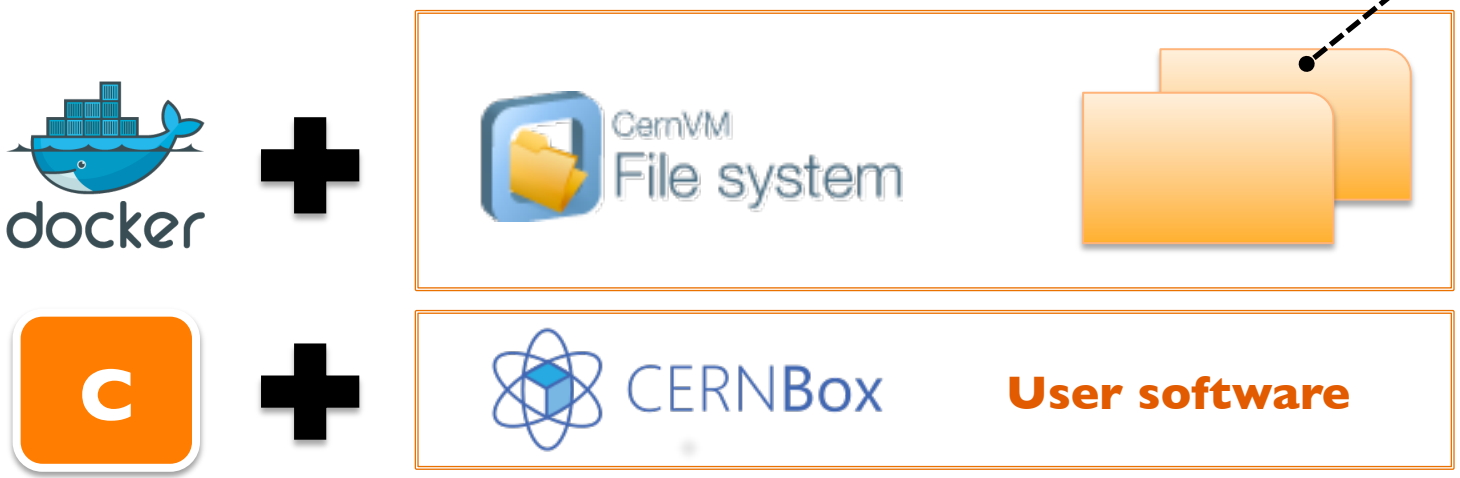
Out[4]:

SWAN Architecture

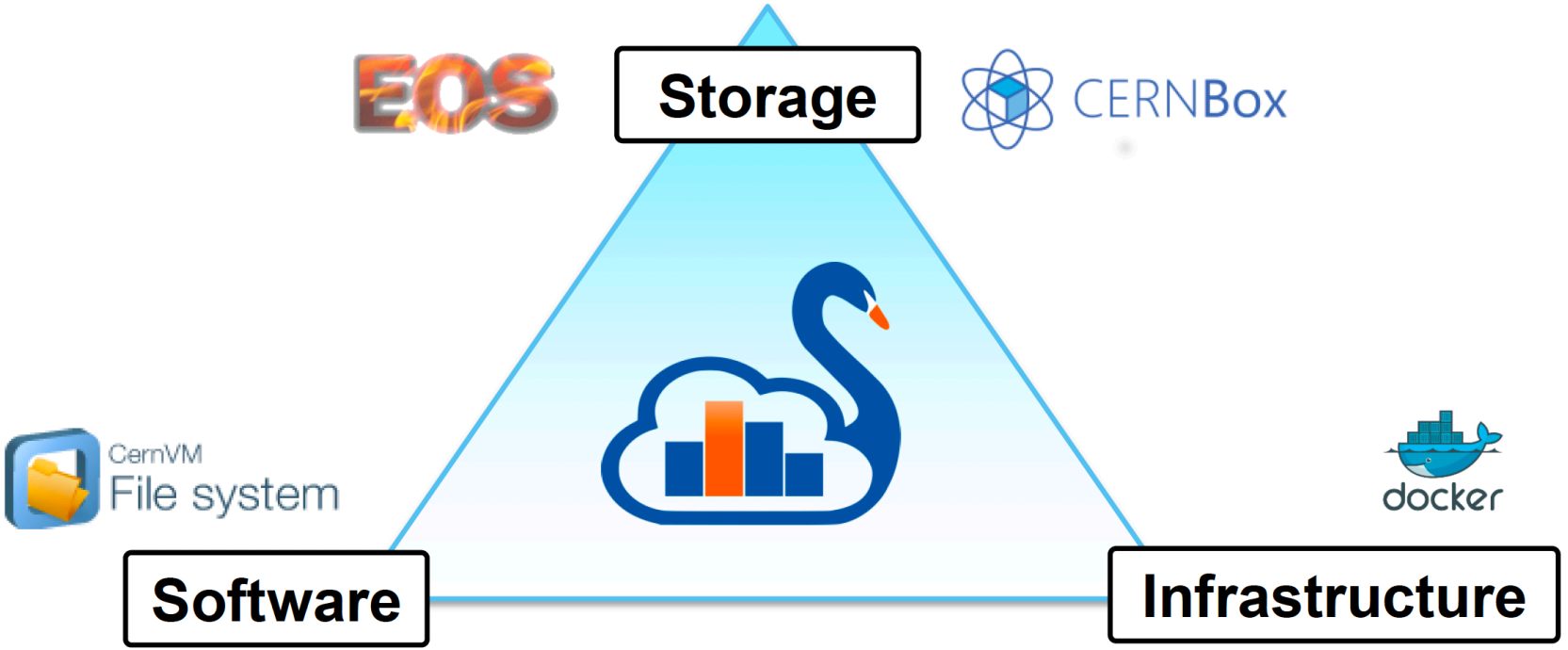


Software Environment

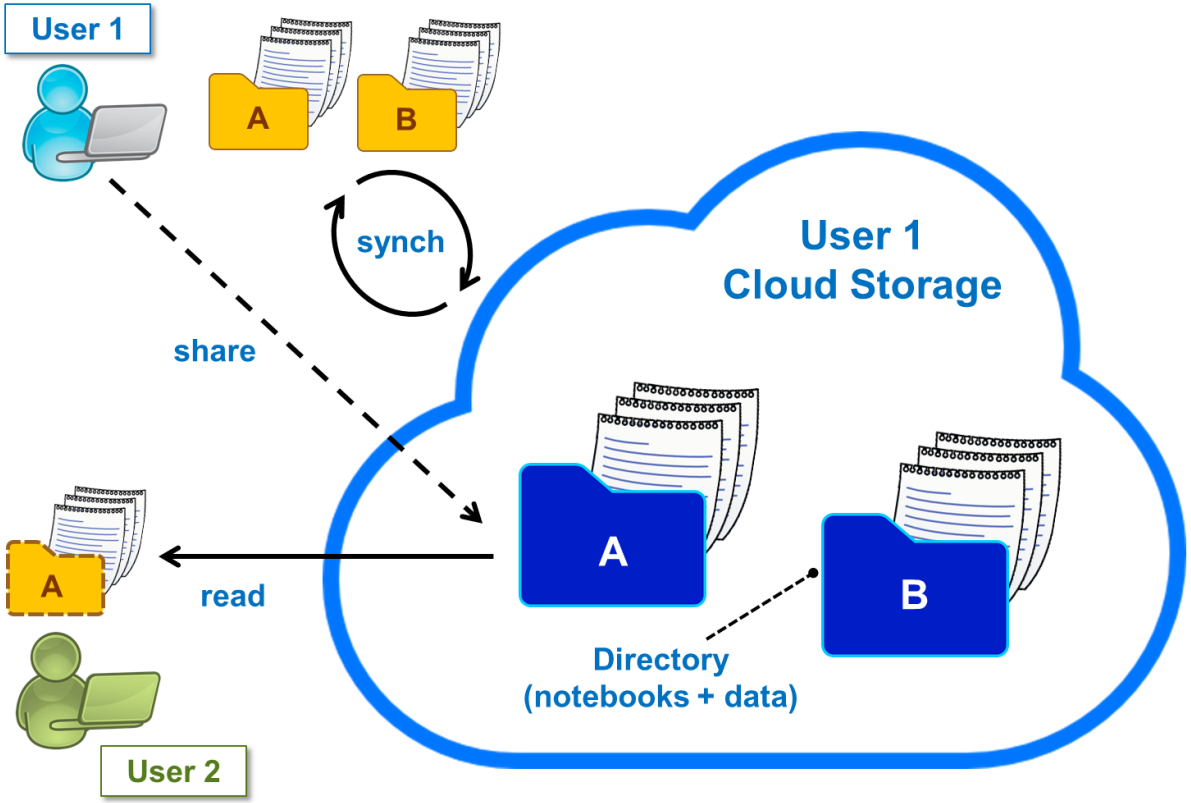
- Strategy to configure the software environment:
  - Docker: single thin image, not managed by the user!
  - CVMFS: configurable environment via "views"
  - CERNBox: custom user environment



Cornerstones of SWAN



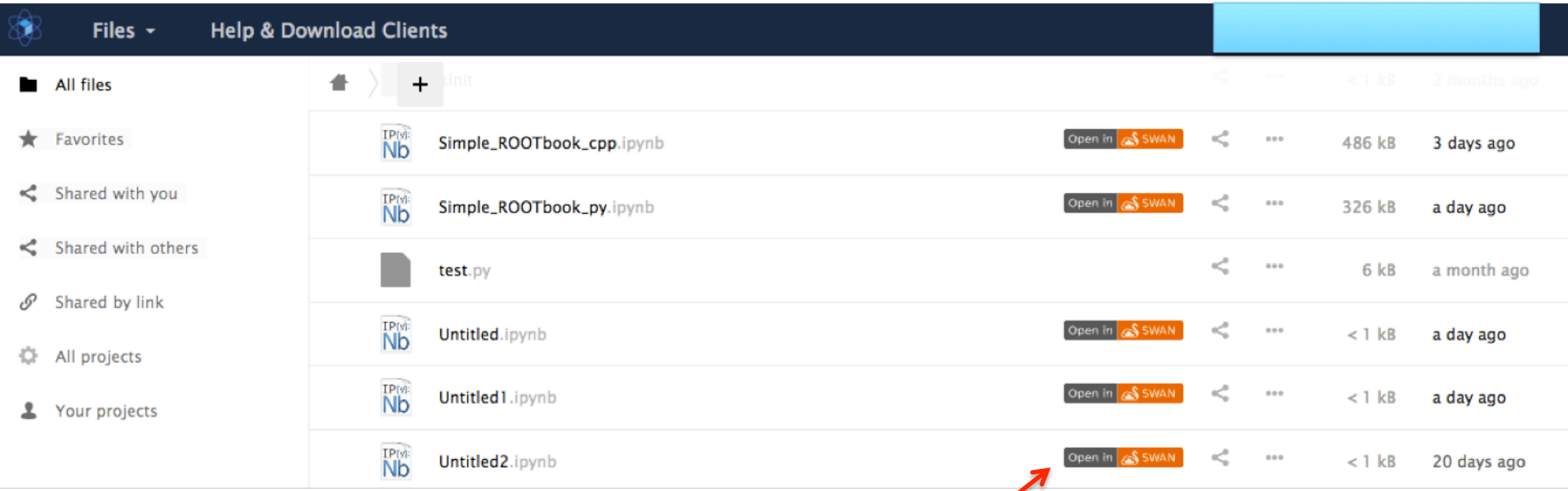
CERNBox: Added Value of Sync & Share



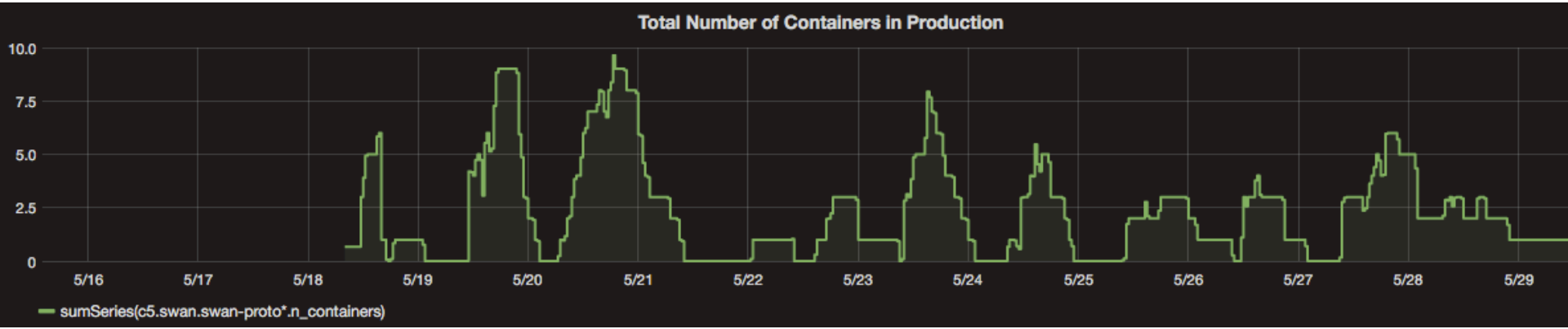
The fundamental shareable unit is the directory: notebook + code + data

Thinking aloud: share your analysis with CDS associated with your paper and open it in SWAN?

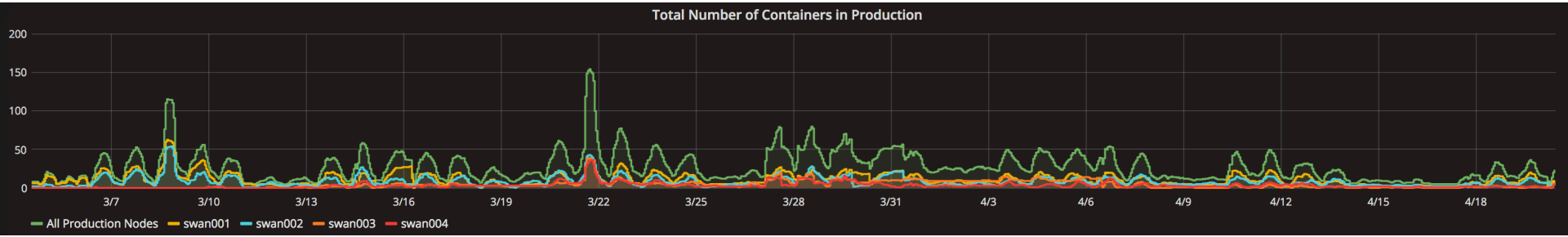
CernBOX Integration



Usage May 2016



Usage April 2017



1 year of SWAN:

- 50 sessions per day
- about 100 notebooks
- peaks of O(100) sessions for trainings
- 4x bigger infrastructure

The Relevance of Producing Content

Focus must be on content: matches well the "copy & share paradigm"

- Collect example notebooks, use cases, software packages (e.g. PyTimber, pieces of Python ecosystem)
- Not only from HEP, e.g. from high school classes

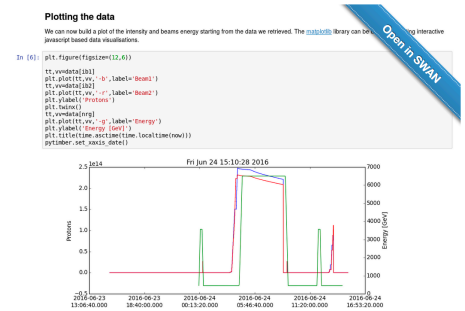
From the SWAN Galleries

Accelerator Complex

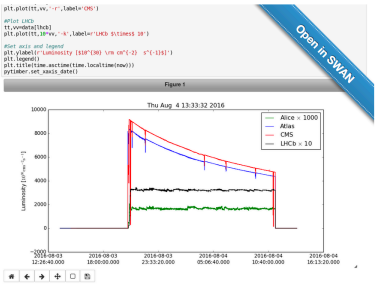
This gallery shows examples of machine studies relative to the CERN accelerators' complex.

Open in SWAN

LHC Page1



Experiments' Luminosities



Particle open data teaching (Hiukkasfysiikan avoin data opetuksessa)

SWAN, the CERN Service for Web based ANalysis, is not only made for analysis of scientific data but also the ideal platform for outreach. Paavo Rikkilä (CMS) put together an introductory course about experimental HEP for future high school teachers. The result is great: check it out in SWAN!

Open in SWAN

Github repository: https://github.com/cmsopendata-finland/kurssimateriaali

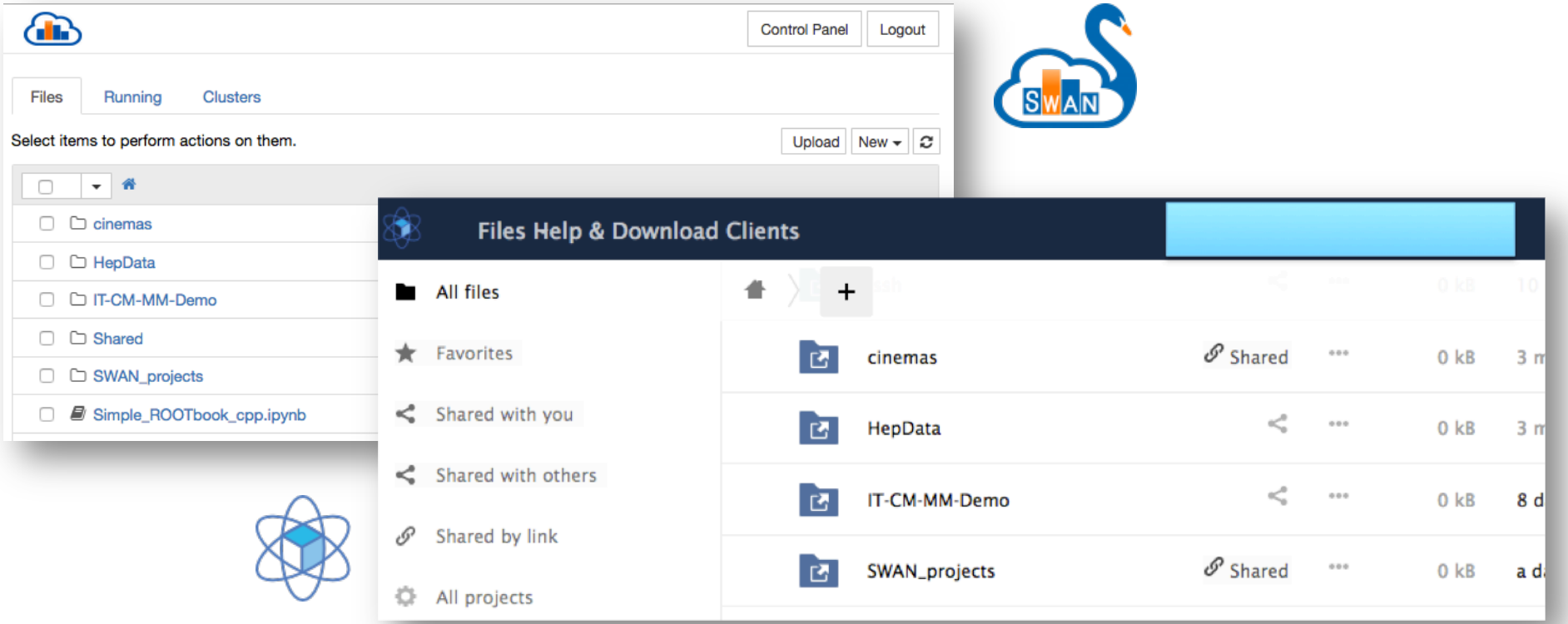
Bottomline

- SWAN is the service for web based analysis at CERN: interface is jupyter notebooks
- Initial goals: integration of services, easy access and reproducibility
  - Is analysis preservation a long term version of reproducibility?
- SWAN building blocks helping reproducibility:
  - Notebooks: no justification not to document!
  - Frozen software stacks: LCG releases
  - Containers for abstracting from the platform
  - Integration with sync and share: share and checkpoint directories (fundamental shareable unit)
  - Integration with mass storage


SWAN

Backup

CERNBox is Your Home



Usage

- Peak: 243 simultaneous users, 503 notebooks, zero slowdowns
- Users from EP, IT, BE, EN
  - BE users are quite active: beam physics, AWAKE, Access to Spark clusters

Some official events powered by SWAN:

- Hadoop Tutorial (Cern Training Catalogue) http://cern.ch/go/vL7d
- ROOT Summer Students' workshop http://cern.ch/go/QxF8
- GridKA School of Computing http://cern.ch/go/7kgF
- CERN School of Computing http://cern.ch/go/k6nq
- "Practical Statistics" Academic Training http://cern.ch/go/Q9ZG

People Counting on It!

urgent: figures do not open any more

Hello,

the figures suddenly stopped working. If I do a simple thing:

import matplotlib.pyplot as plt

plt.figure()

Did you change anything in the last 10 min? I obtain the error message below. Could you please fix it asap as I have an analysis to finish tonight.

Thanks,

Note: in the end it was not a SWAN issue but rather matplotlib trying to interact with a non-existing X server.

People Counting on It!

-------- Forwarded Message --------
Subject: Re: cannot start notebooks
Date: Tue, 13 Dec 2016 13:54:21 +0100

[redacted]

… and died again,can you pls restart once more, I'd need to create just one more plot today for a meeting in 5 minutes ;-)

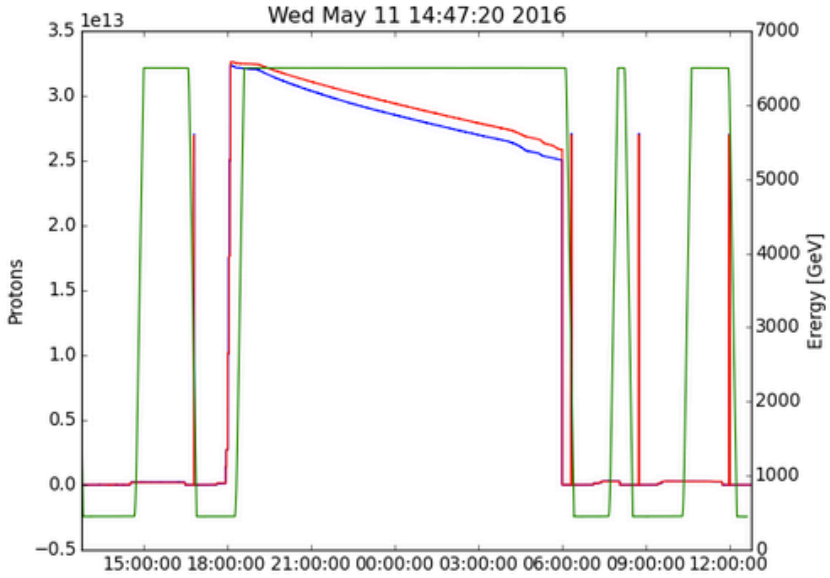
thx

[redacted]

Note: an EOS glitch. A reboot of the Docker daemon solved the issue.

```
In [3]: now=time.time()
        ib1="LHC.BCTDC.A6R4.B1:BEAM_INTENSITY"
        ib2="LHC.BCTDC.A6R4.B2:BEAM_INTENSITY"
        nrg="LHC.BOFSU:OFSU_ENERGY"
        data=db.get([ib1,ib2,nrg],now-3600*24,now)
```

```
In [4]: plt.clf()
        tt,vv=data[ib1]
        plt.plot_date(epoch2num(tt),vv,'-b',label='Beam1')
        tt,vv=data[ib2]
        plt.plot_date(epoch2num(tt),vv,'-r',label='Beam2')
        plt.ylabel('Protons')
        plt.twinx()
        tt,vv=data[nrg]
        plt.plot_date(epoch2num(tt),vv,'-g',label='Energy')
        plt.ylabel('Erergy [GeV]')
        plt.title(time.asctime(time.localtime(now)))
```



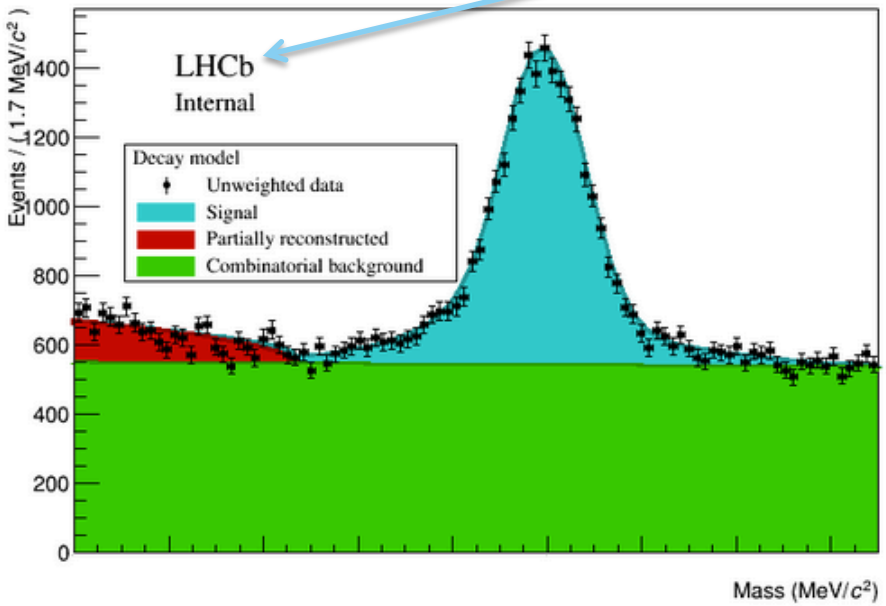
R. De Maria, BE-ABP-HSS

https://github.com/rdemaria/pytimber/blob/master/examples/LHC%20Page1.ipynb

- Read measurements coming from pick-ups in a database
- Plot time series
- Needs also SciPy and to share the notebooks with his colleagues

```
title = { "model": "Signal" ,  "pdfBkg" : "Partially reconstructed" , "cmbBkg": "Combinatorial background"}

for (component, color) in [ ("model",kCyan), ("pdfBkg",kRed), ("cmbBkg",kGreen)]:
    model.plotOn (frame, LineColor(color+2) , DrawOption('L'), Components(component), LineWidth(5))
    model.plotOn (frame, FillColor(color+1) , DrawOption('F'), Components(component), LineWidth(0), Name("P"+component
))
    leg.AddEntry ( frame.findObject ("P"+component), title[component] , "F" )

data.plotOn ( frame, MarkerColor ( ROOT.kBlack ) )
frame.Draw()
Graphics().lhcbMarker(0.2,0.8, "Internal")

leg.Draw()

ROOT.gPad.Draw()
```

Results coming from real data! (published now)



L. Anderlini

Rare B meson decay in LHCb

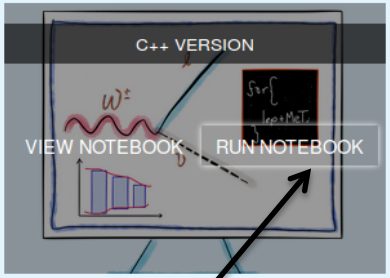
- Read data from EOS
- Setup complex fit
- Document and inspect results

Atlas Opendata and SWAN

ATLAS ROOTbooks Gallery!

How deep can you go?

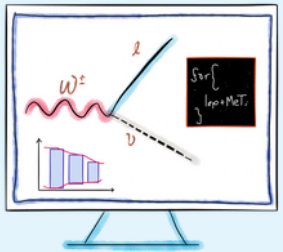
Analysis notebooks at
http://opendata.atlas.cern/webanalysis/ROOTbooks.php

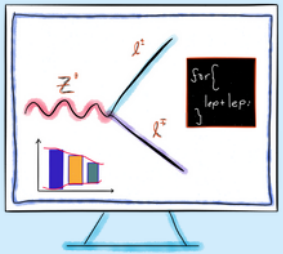


Runnable in SWAN!

The W Analysis ROOTbook

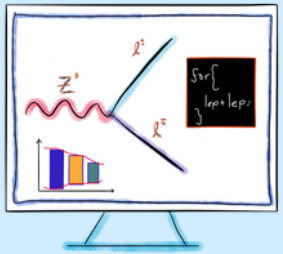
The W boson analysis is intended to provide an example for a high statistics analysis using the ATLAS open data dataset. Furthermore it tests the description of the real data by the simulated W boson data which represents the most extensive dataset in terms of luminosity.





The Z Analysis ROOTbook

Many analyses selecting leptons suffer from Z + jets as a contributing background due to its large production cross section. It is therefore vital to check the correct modelling of this process by the Monte-Carlo simulated data. It is important to measure well known Standard Model particles, to confirm that we understand properly the detector and software. We are then ready to search for new physics.



open data ATLAS

User Support

- Treat every user as a smart person: explain why the system had issues, how we fixed them
- Encourage feedback, attract contributions
- Set up a forum-like mailing lis
- *Never* push difficulties back to users