SWAN
Service for
Web-based ANalysis

https://swan.web.cern.ch

E. Tejedor, D. Piparo, P. Mató
L. Mascetti, J. Moscicki, M. Lamanna
CERN

CHEP
12/10/2016
**Jupyter Notebook:** A web-based interactive computing interface and platform that combines code, equations, text and visualisations.
Access TTree in Python using PyROOT and fill a histogram

Loop over the TTree called "events" in a file located on the web. The tree is accessed with the dot operator. Same holds for the access to the branches: no need to set them up - they are just accessed by name, again with the dot operator.

In [1]:
```python
import ROOT

T = ROOT.TFile.Open("http://indico.cern.ch/event/395198/material/0/0.root");
h = ROOT.TH1F("TracksPt","Tracks:Pt [GeV/c];#",128,0,64)
for event in T.events:
    for track in event.tracks:
        h.Fill(track.Pt())
c = ROOT.TCanvas()
h.Draw()
c.Draw()
```

<table>
<thead>
<tr>
<th>TracksPt</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Entries</td>
<td>499832</td>
</tr>
<tr>
<td>Mean</td>
<td>12.54</td>
</tr>
<tr>
<td>Std Dev</td>
<td>6.554</td>
</tr>
</tbody>
</table>
**Jupyter Notebook**: A web-based interactive computing interface and platform that combines code, equations, text and visualisations.

ROOT has been fully integrated with the Jupyter technology

- Two flavours: Python and ROOT C++
- **JavaScript** interactive visualisation
SWAN: Data analysis “as a service”

*Interface:* Jupyter Notebooks

*Goals:*

- **Analysis only with a web browser**
  - Platform independent ROOT-based data analysis
  - Calculations, input and results “in the Cloud”
- **Easy sharing** of scientific results: plots, data, code
  - Storage is crucial: mass & synchronised
- **Simplify teaching** of data processing and programming
  - Gallery of analysis examples
- **Integration with other analysis ecosystems:** R, Python, …
SWAN in the CERN Ecosystem

SWAN federates a set of production technologies at CERN:

• Authentication with CERN credentials (SSO)
• Infrastructure: virtual machines in OpenStack Cloud
• Software distribution: CVMFS
  – Centrally distributed software
• Storage access: CERNBox, EOS
  – Experiments’ and users’ data

Plus some external technologies:

• JupyterHub
• Docker
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
CERNBox: Sync & Share

User 1

User 2

11/10, 16:30 – J. Moscicki
CERNBox: The Data Hub for Data Analysis
In collaboration with IT-DB, IT-ST

10/10, 14:45 – D. Piparo
Expressing Parallelism in ROOT
• Pilot Service released in June
https://swan.cern.ch
• Open to everyone with a CERN account
  – And a CERNBox account activated
• Stats of the first 100 days (Jun – Sep):
  – 1800 sessions created
  – 3700 notebooks opened
  – Peak of 100 sessions
• See backup slides for some user examples
  – E.g ATLAS Open Data,
    LHC Page 1 (BE)
ATLAS ROOTbooks Gallery!

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

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.
• Service open outside CERN since end of July
• SWAN pilot service available
  – ROOT integrated with Jupyter
  – CVMFS for software distribution
  – EOS mass storage + CERNBox synchronisation

• Future plans:
  – Incorporate user feedback
  – Facilitate use of experiment software
  – Enrich CVMFS repositories with new packages (e.g. ML libraries)
  – Improve experience with storage: sharing
  – Exploit external resources (e.g. Spark clusters)
Backup
Example notebooks at swan.web.cern.ch

Click on the blue ribbon to open them in SWAN!

Click on the image for a static visualisation
R. De Maria, BE-ABP-HSS
- Read measurements coming from pick-ups in a database
- Plot time series
- Needs also SciPy and to share the notebooks with his colleagues
L. Anderlini
Rare B meson decay in LHCb
- Read data from EOS
- Setup complex fit
- Document and inspect results

Results coming from real data! (published now)
Mano S. (14 years old), K12 student

- Approaches programming for the first time
- Verifies numerically what he learned at school
- Shares results with his supervisor and classmates
G. Lo Presti, M. Lamanna
“Castor data corruption incident”
• Describe incident, data source, analysis and results in a single document