

# A JOURNEY TO THE CENTER OF ROOT'S LOG FILES

**Alexandra-Maria Dobrescu**

# WHO AM I ?

## MAYBE A CURIOUS PARTICLE

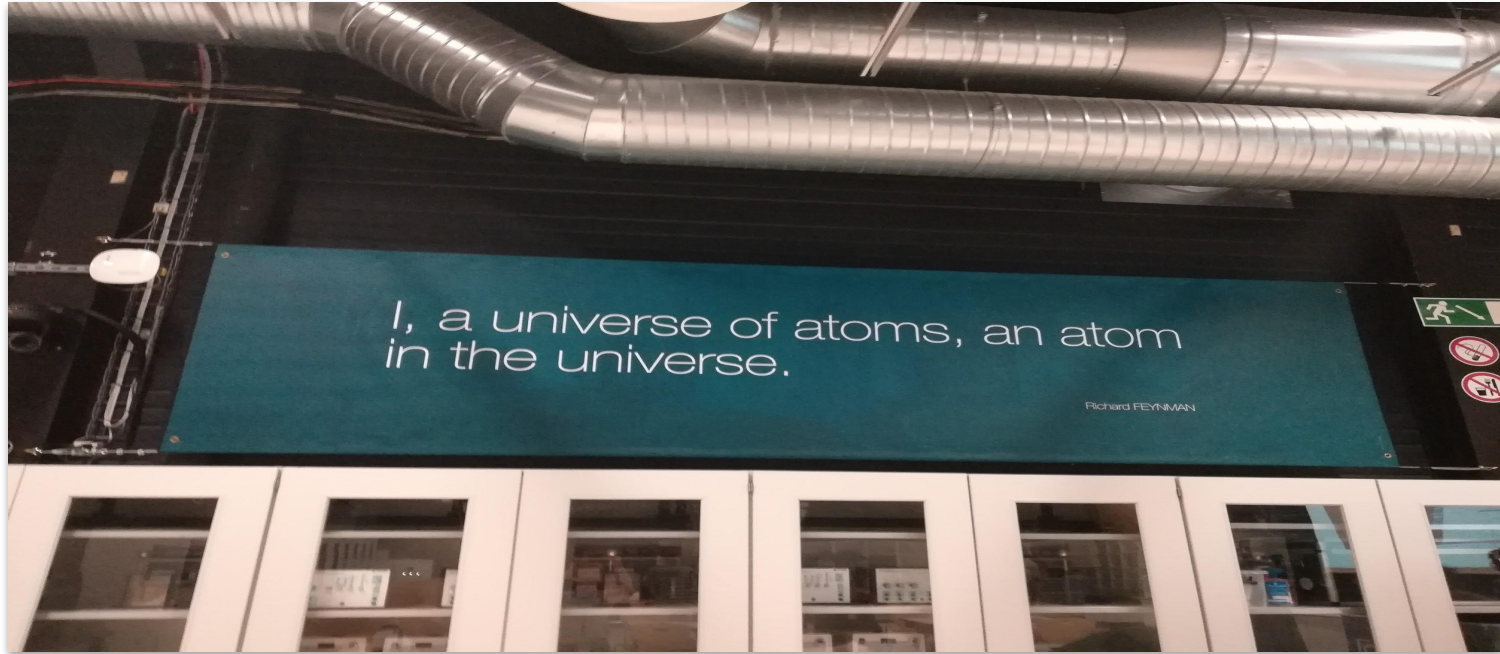
<https://scool.web.cern.ch/content/particle-identities-quiz>

According to your answers, the particle which fits your personality



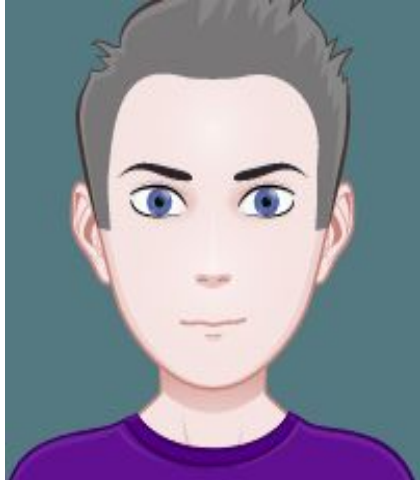
The Higgs boson is the most recently discovered elementary particle. It was predicted by physicists in 1964. However, it took almost 50 years and the most powerful particle accelerator in the world to discover the Higgs Boson in 2012 at CERN.

OR "A UNIVERSE OF ATOMS, AN ATOM IN THE UNIVERSE" ...



...FROM WHAT I'VE SEEN AT THE CLOUD CHAMBER WORKSHOP

# BEING MORE SERIOUS:



SUPERVISOR : AXEL NAUMANN, OLIVIER COUET



# DATA COLLECTION

First trip:

Extracting the data  
from the Apache logs

---

APACHE LOGS  
=  
DATA STORAGE

PYTHON  
→  
SQL  
STATEMENT

PY OBJECTS:

- IP ADDRESS;
- TIME;
- USER AGENT;
- VERSION OF ROOT;
- PLATFORM;
- FILE TYPE.

IT FILTERS ROBOTS AND RESOLVES GEOIP.

HOW CAN IT GET  
MORE INTERESTING?

Second step:

From PY Objects to  
columns.

---

## PY OBJECTS:

- IP ADDRESS;
- TIME;
- USER AGENT;
- VERSION OF ROOT;
- PLATFORM;
- FILE TYPE.

PYTHON

SQL  
STATEMENT

COLUMNS

TABLE

SQLITE3 DATABASE

IT CREATES A SQLITE3 FILE.



# WHO WANTS TO SEE A TERMINAL?

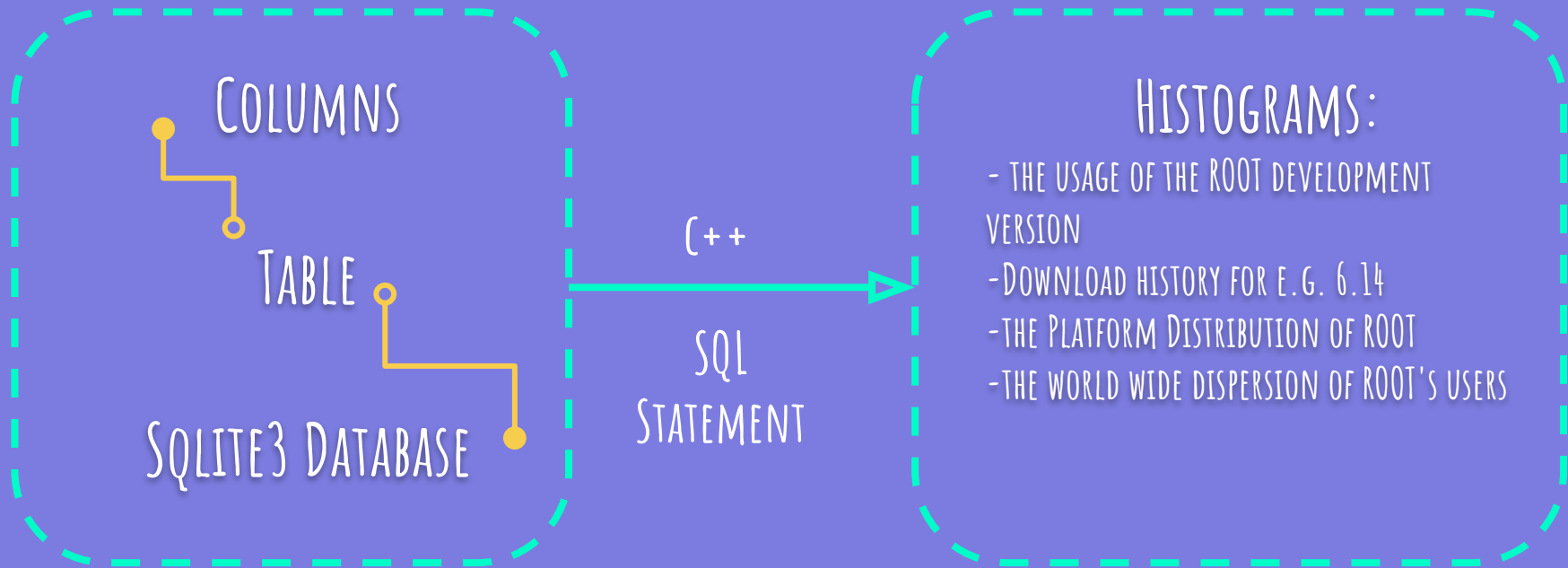
```
adobresc@pcphsft78: ~/Documents
File Edit View Search Terminal Help
IPv4|CN|35.0|105.0|2018-07-12 03:11:43|6.14.00|macosx64-10.13-clang91|tar.gz
IPv4|DE|49.5888|11.0098|2018-07-12 03:19:00|5.34.36|Linux-slc6-x86_64-gcc4.4|tar.gz
IPv4|CN|35.0|105.0|2018-07-12 03:12:10|6.14.00|macosx64-10.13-clang91|tar.gz
IPv4|DE|49.5888|11.0098|2018-07-12 03:26:52|5.34.36|Linux-slc6-x86_64-gcc4.4|tar.gz
IPv4|US|33.6403|-117.7694|2018-07-12 03:26:06|6.14.00|macosx64-10.13-clang91|dmg
IPv4|KR|37.57|126.98|2018-07-12 03:37:09|5.34.36|source|tar.gz
IPv4|KR|37.57|126.98|2018-07-12 03:37:52|5.34.36|source|tar.gz
IPv4|CA|45.5148|-73.5739|2018-07-12 03:40:48|6.14.00|win32.vc15|exe
IPv4|US|37.419200000000004|-122.0574|2018-07-12 03:44:59|5.34.36|win32.vc10.debug|exe
IPv4|US|37.419200000000004|-122.0574|2018-07-12 03:45:45|5.34.36|win32.vc10.debug|exe
IPv6|CN|35.0|105.0|2018-07-12 03:47:26|6.14.00|Linux-ubuntu14-x86_64-gcc4.8|tar.gz
IPv4|CA|45.5148|-73.5739|2018-07-12 04:06:26|6.14.00|win32.vc15|zip
IPv4|CA|45.5321|-73.894|2018-07-12 04:06:29|6.14.00|win32.vc15|zip
IPv4|JP|35.69|139.69|2018-07-12 04:16:28|6.14.00|macosx64-10.13-clang91|dmg
IPv4|KR|37.57|126.98|2018-07-12 04:19:45|5.34.36|macosx64-10.11-clang70|tar.gz
IPv4|CN|31.8639|117.2808|2018-07-12 04:28:09|6.14.00|source|tar.gz
IPv4|CN|31.0456|121.3997|2018-07-12 04:28:44|5.99.06|source|tar.gz
IPv4|CN|31.0456|121.3997|2018-07-12 04:29:38|5.34.36|Linux-centos7-x86_64-gcc4.8|tar.gz
IPv4|JP|34.6617|133.935|2018-07-12 04:30:19|5.34.36|macosx64-10.11-clang70|tar.gz
IPv4|JP|36.0833|140.1167|2018-07-12 04:32:20|6.14.00|Linux-ubuntu18-x86_64-gcc7.3|tar.gz
IPv4|JP|36.0833|140.1167|2018-07-12 04:38:26|6.14.00|Linux-ubuntu18-x86_64-gcc7.3|tar.gz
IPv4|CA|45.5148|-73.5739|2018-07-12 04:51:58|5.34.36|win32.vc10|exe
IPv6|CN|39.9289|116.3883|2018-07-12 03:53:44|6.14.00|Linux-fedora27-x86_64-gcc7.2|tar.gz
IPv6|CN|30.5801|114.2734|2018-07-12 03:59:03|6.14.00|source|tar.gz
IPv4|AU|-27.4846|153.0201|2018-07-12 05:23:17|6.14.00|macosx64-10.13-clang91|tar.gz
IPv6|US|39.76|-98.5|2018-07-12 06:09:34|6.14.00|source|tar.gz
IPv4|AU|-27.0|133.0|2018-07-12 06:03:58|6.14.00|Linux-ubuntu16-x86_64-gcc5.4|tar.gz
IPv4|JP|35.7894|139.6233|2018-07-12 06:25:26|6.14.00|win32.vc15|exe
IPv4|JP|36.0833|140.1167|2018-07-12 06:32:50|5.34.36|source|tar.gz
IPv4|JP|35.69|139.69|2018-07-12 06:33:05|5.34.36|source|tar.gz
IPv4|JP|35.685|139.7514|2018-07-12 06:43:04|6.14.00|source|tar.gz
IPv4|BR|-23.5477|-46.6358|2018-07-12 07:00:42|6.14.00|source|tar.gz
IPv4|CN|39.8897|115.275|2018-07-12 06:58:09|6.14.00|source|tar.gz
IPv4|JP|36.0833|140.1167|2018-07-12 07:23:03|6.14.00|source|tar.gz
IPv4|TW|25.0392|121.525|2018-07-12 07:25:32|5.34.10|source|tar.gz
IPv4|US|37.419200000000004|-122.0574|2018-07-12 07:31:40|6.14.00|win32.vc15.debug|exe
IPv4|US|37.419200000000004|-122.0574|2018-07-12 07:32:34|6.14.00|win32.vc15.debug|exe
IPv4|KR|37.57|126.98|2018-07-12 07:35:03|6.12.06|macosx64-10.13-clang90|tar.gz
solite>
```

NO ONE. :(

Third trip:

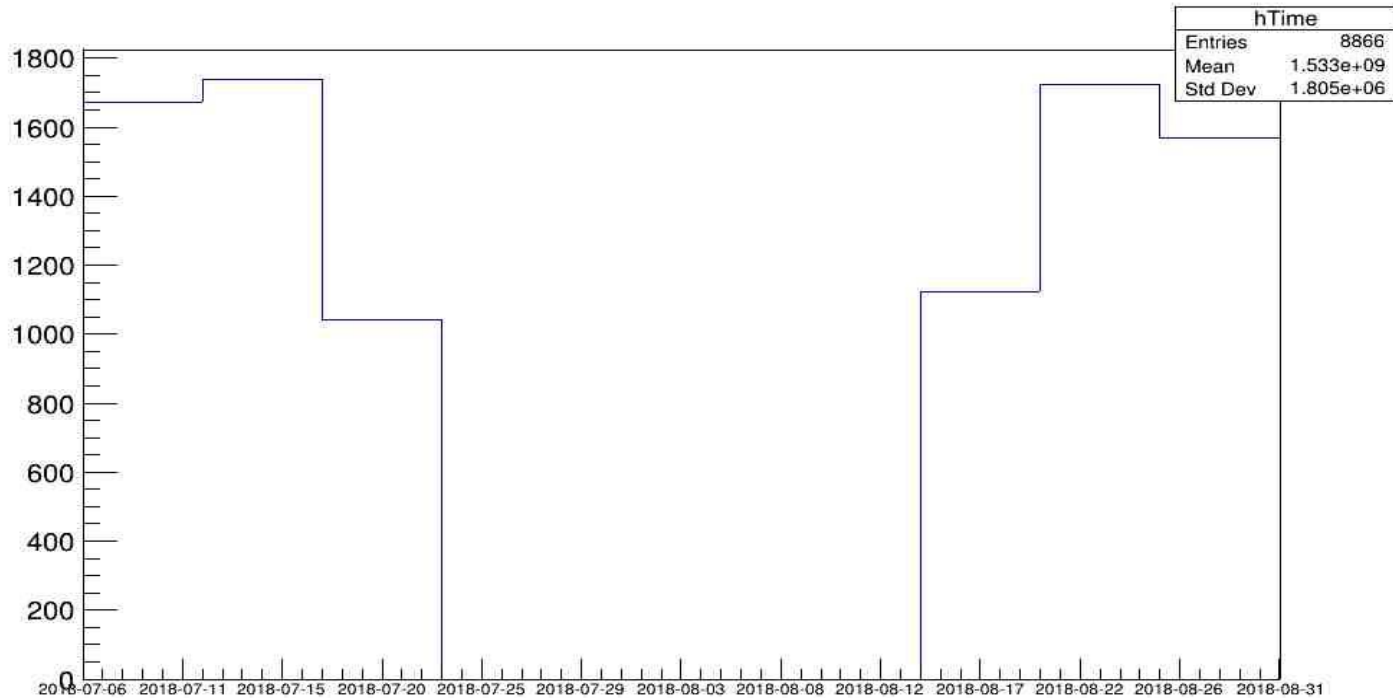
From SQLiteDB to  
Histograms.

---

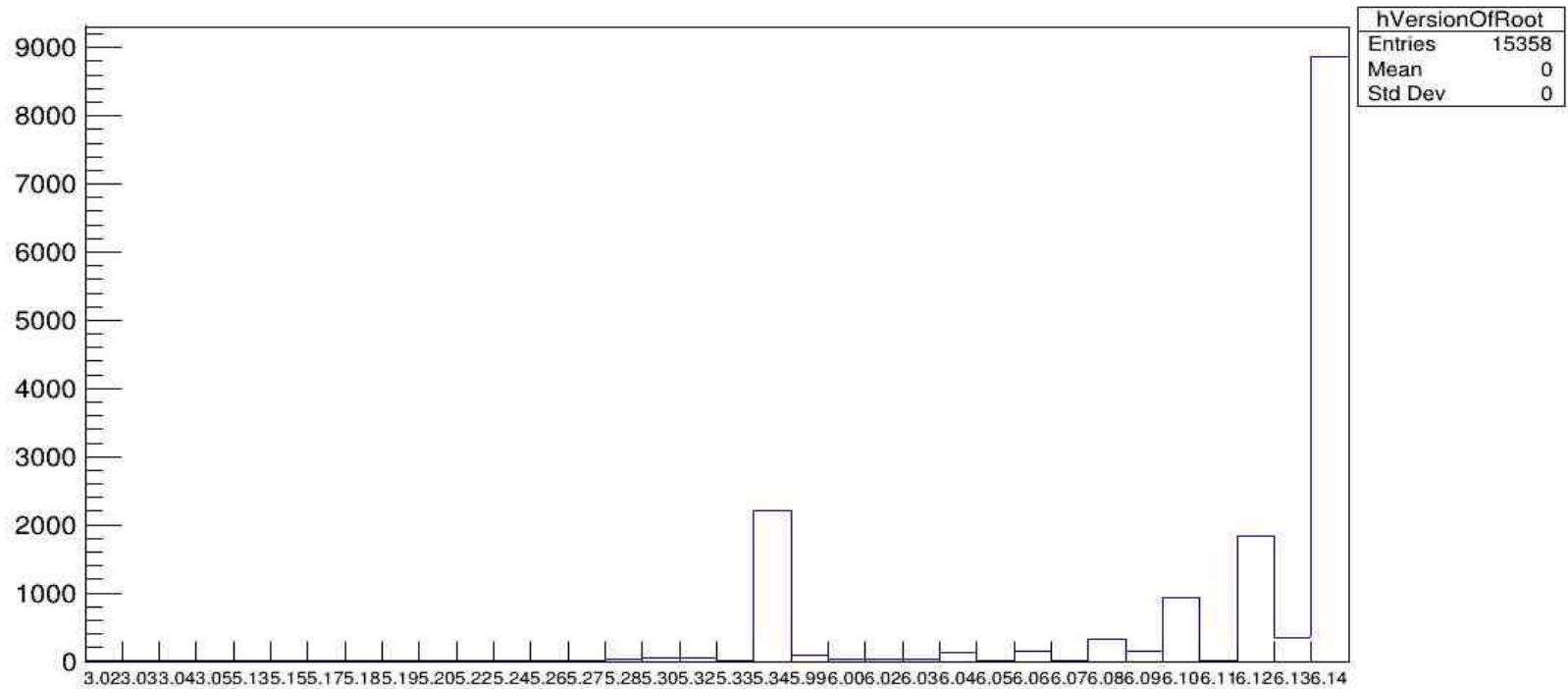


# CREATE SQLITE TUTORIALS FOR LOG STATISTICS

## 6.14 DOWNLOAD HISTORY

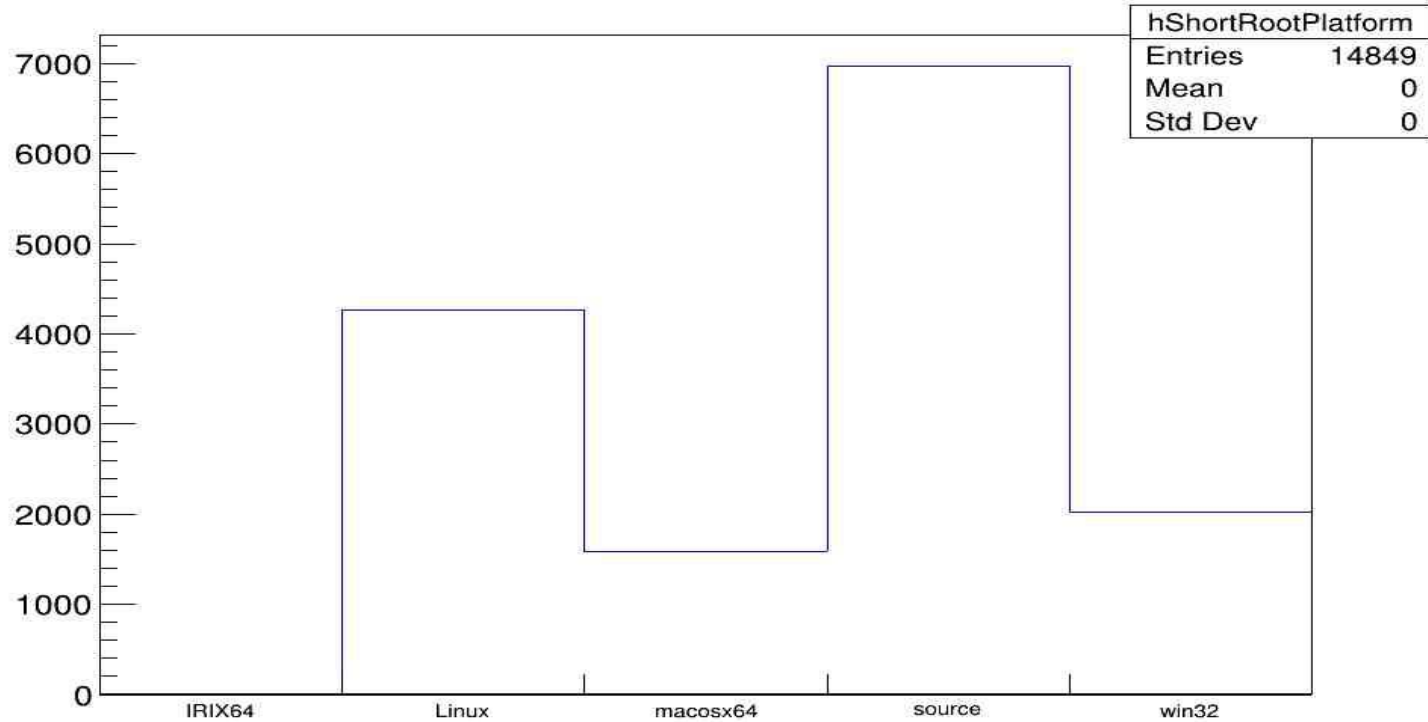


# THE USAGE OF PAST ROOT VERSIONS



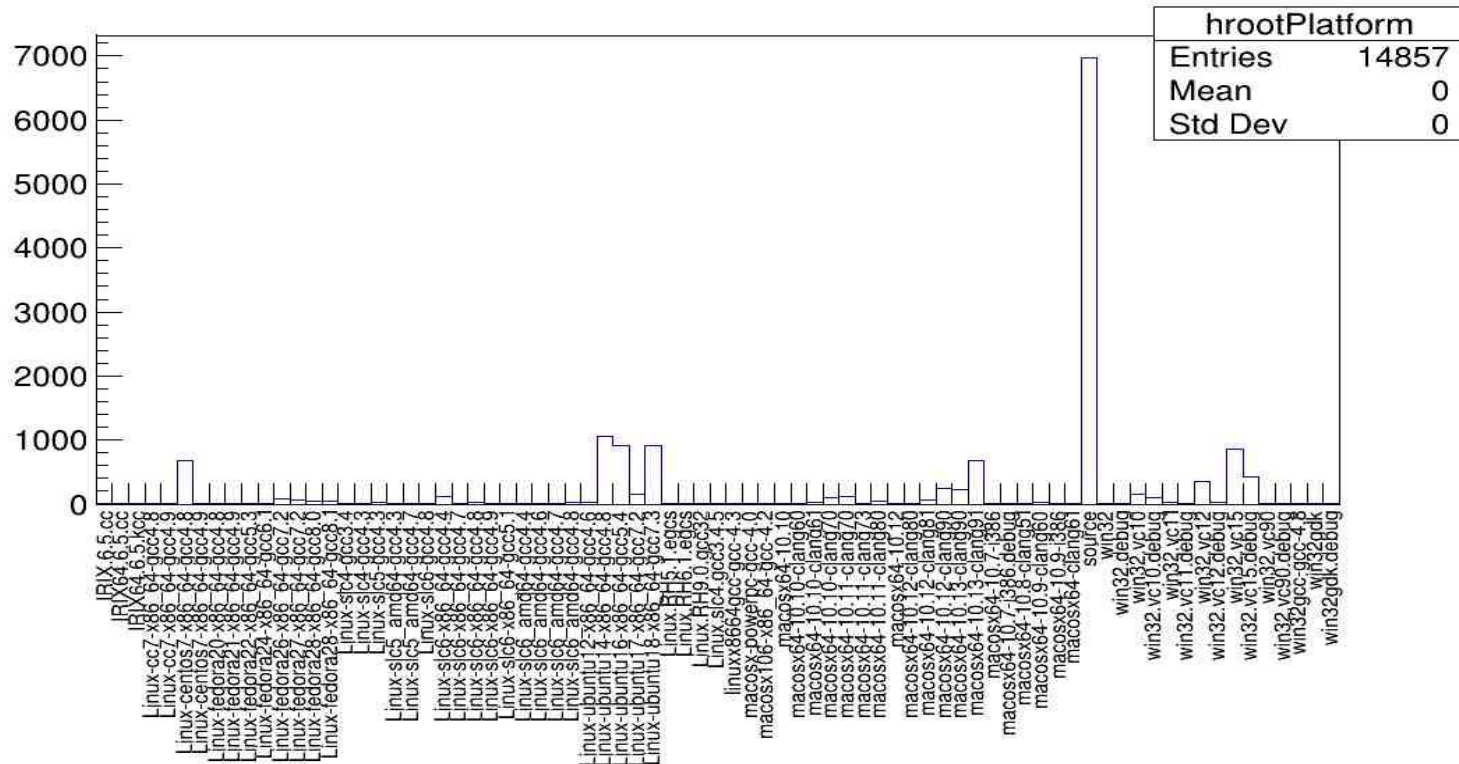
# THE PLATFORM DISTRIBUTION OF ROOT

Short Platform Distribution



## THE PLATFORM DISTRIBUTION OF ROOT

## Platform Distribution



IS THE JOURNEY  
ENDING?

From the center of  
ROOT's log files

TO

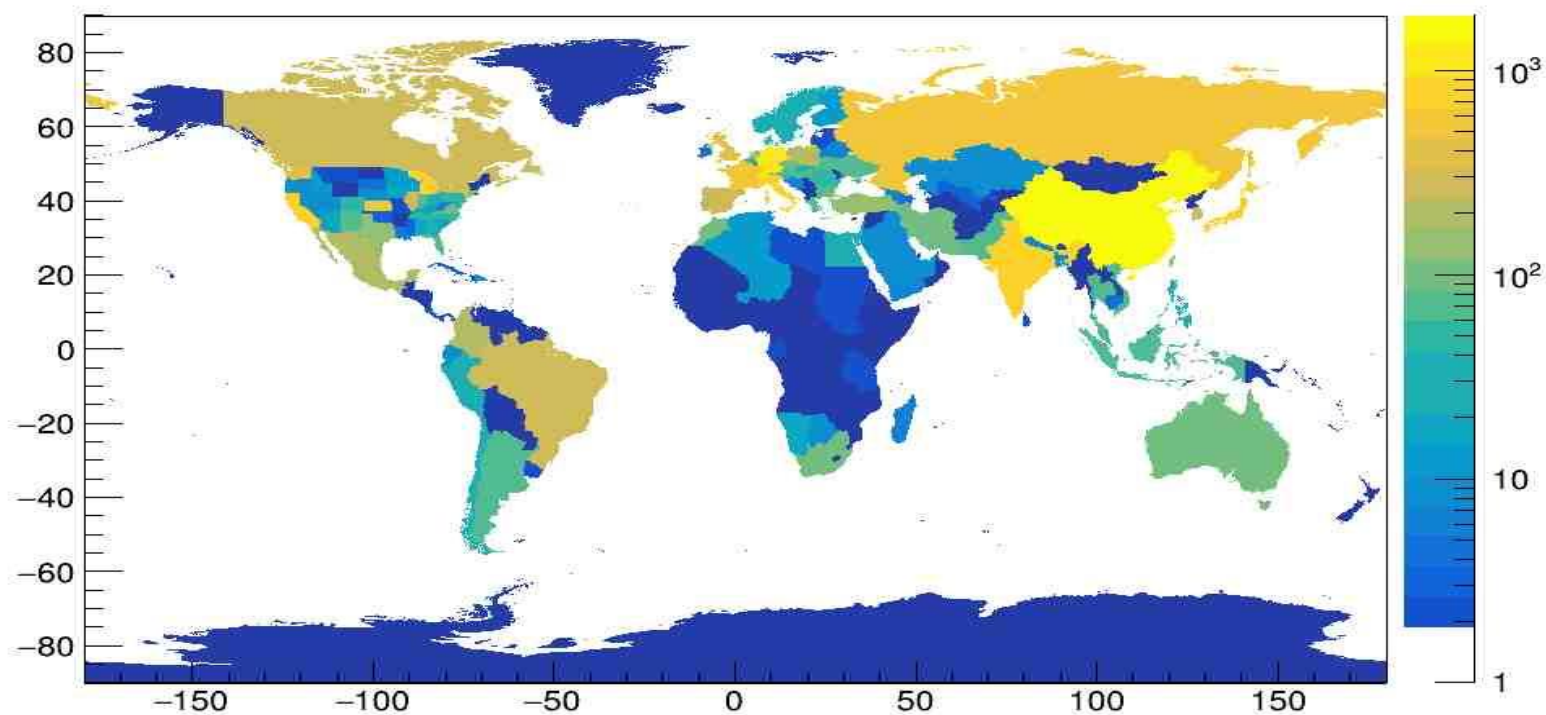
the Wide World.

---



# THE WORLD WIDE DISPERSION OF ROOT'S USERS

World Map



# WHAT ARE THE HISTOGRAMS HIDING?

- 1 set of macros  
using **TSQLServer**
- 1 set of macros  
using **RDataFrame**

---

# TSQLSERVER

Abstract base class defining  
an interface to SQL Server

What I've learned and used:

1. Create macros with ROOT
2. Connect to SQLite3 database
3. Launch Queries
4. Generate a Histogram
5. Fill the Histogram with the specific data stored in the columns.

The inconvenient: the database  
should be downloaded locally.

---

# RDATAFRAME

The Swiss army knife for  
ROOT Data Analysis

What I've learned and used:

1. Create macros with ROOT
2. Connect to SQLite3 database
3. *Launch a general Query to select the entire table*
4. Generate a Histogram
5. Fill the Histogram with the specific data stored in the columns
6. *Simplify the procedure with a lambda expression*

Advantage: More interesting task to come

---

## OTHER TRAVEL EXPENSES:

- Learn how to use UNIX
- Improve Regular Expressions
- gdb

## LAST BUT NOT LEAST - LET THE OTHERS KNOW ABOUT THE JOURNEY:

- Submit a PR on GitHub

AND MAKE NEW FRIENDS @JAKOB, @ENRICO

99.9 %

One last visit :

Improving the RDataFrame data source class  
for reading **remote** SQLite files over http(s).

# AHA!

**PURPOSE:** TO UPDATE ON ROOT.CERN.CH THE SQLITE3 DATABASE. IN THIS WAY, THE USER CAN BENEFIT OF A REAL TIME EXPERIENCE WITH THE LOG FILES. LIKE A JOURNEY IN REAL TIME.

**MEANING:** THE LOCAL FILE BECOMES A REMOTE FILE.

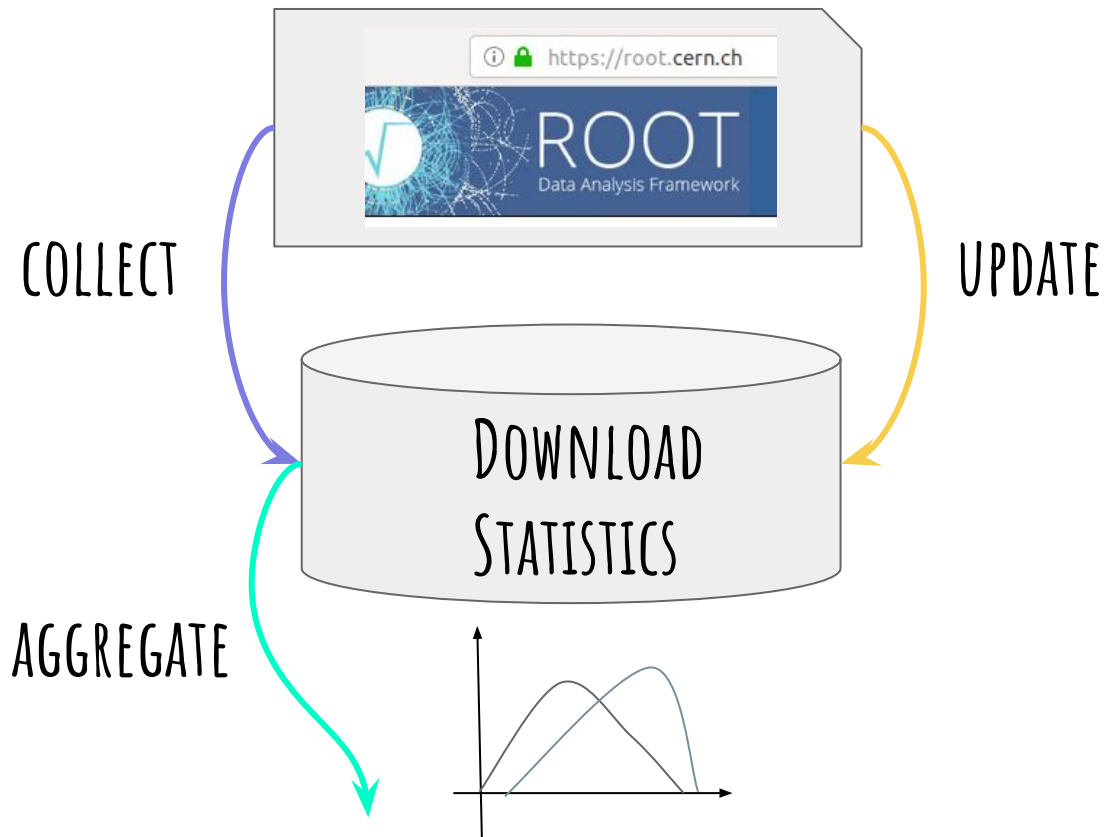
**"THE INGREDIENTS" :**

- A CUSTOMISED SQLITE VFS LAYER;

**"RECIPE SECRET":**

- ACCESS TO THE DAVIX LIBRARY IN JAKOB'S BRAND NEW RSQLITEDS.CXX FILE.

# CONCLUSION



- Displayed by Tutorials
- Using **Remote** SQLite DataSource access



THANK YOU !

