SPARK SUMMIT

Next CERN Accelerator Logging Service A road to Big Data

Jakub Wozniak, CERN

#SparkSummit

Geneva, Switzerland / France

Research

Technology Education

22 member states

CERN

Physics Lab
since 1954

Detectors

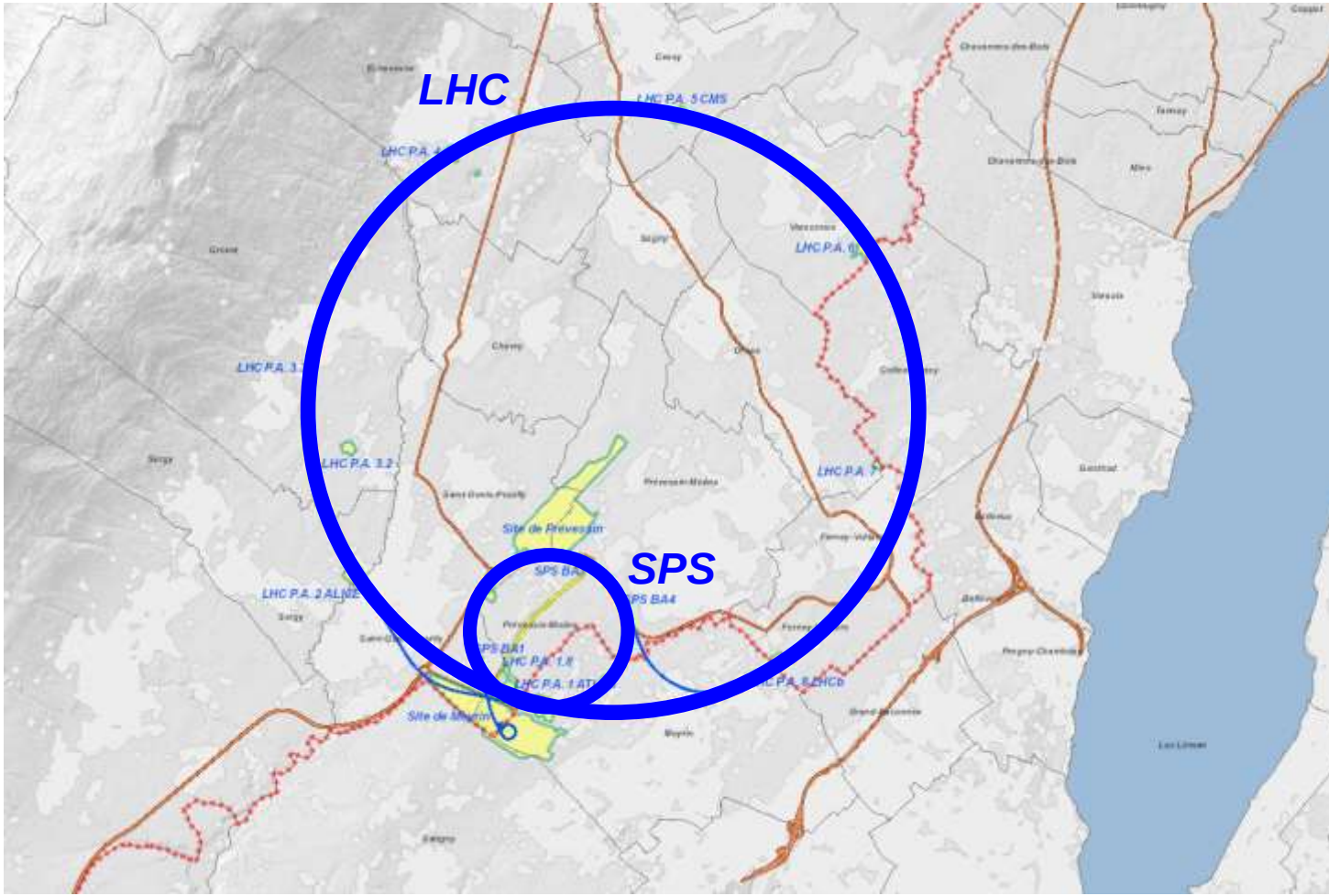
Atlas, CMS, Alice, LHCb

2,500 employees

Large Hadron Collider

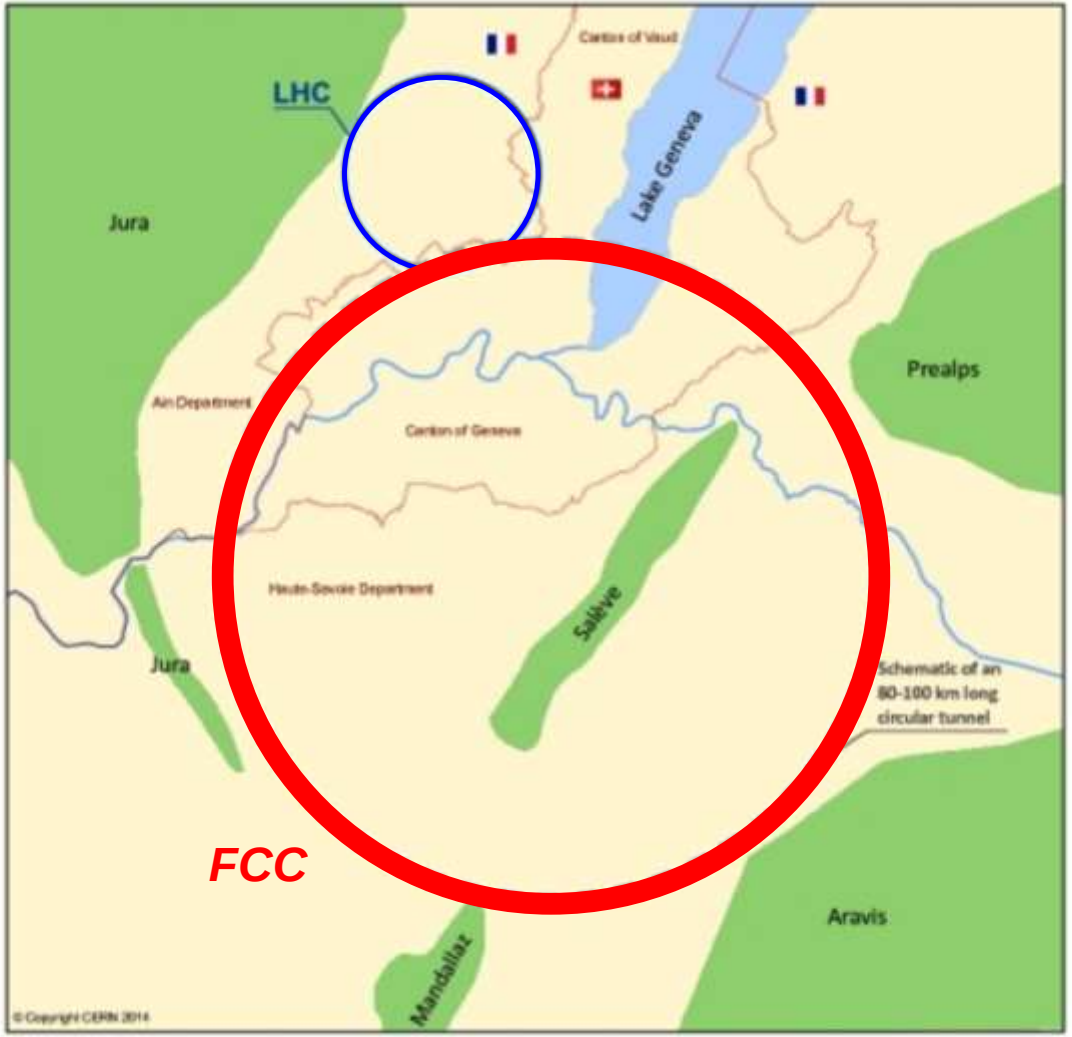
Over 10,000 users

CERN Complex On Map

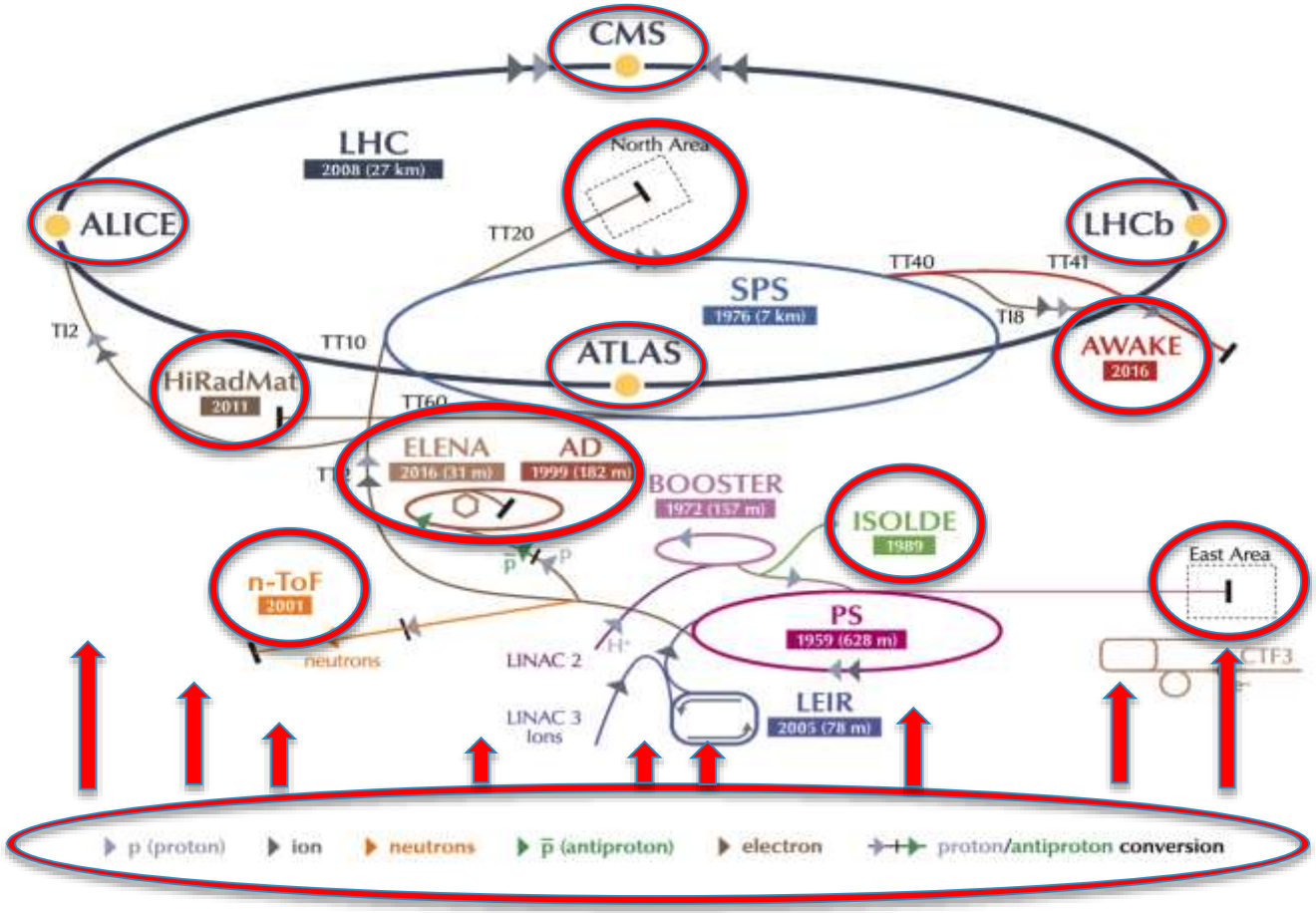


Future Circular Collider (FCC)

Plans for
80-100 km tunnel
under Geneva Lake and Alps



Experiments

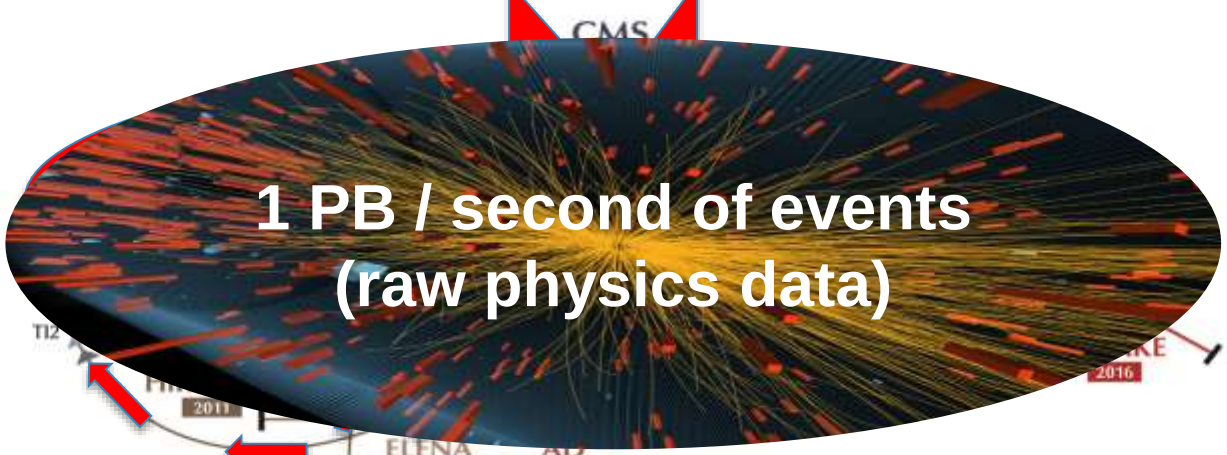


LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron

AD Antiproton Decelerator CTF3 Clic Test Facility AWAKE Advanced WAKefield Experiment ISOLDE Isotope Separator OnLine DEvice

LEIR Low Energy Ion Ring LINAC LINear ACcelerator n-ToF Neutrons Time Of Flight HiRadMat High-Radiation to Materials

LHC Beam



6.5TeV

Proton Energy

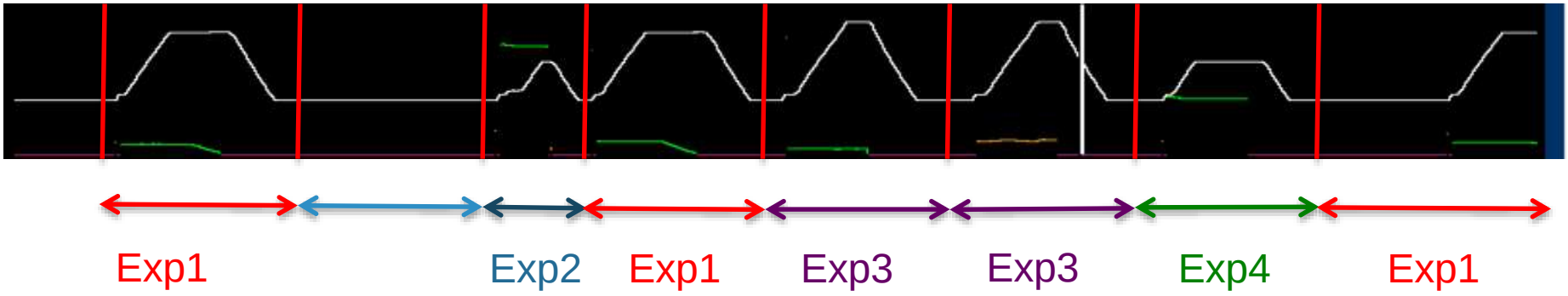
50MeV

Total stored beam energy
116e09 p x 2800 bunches

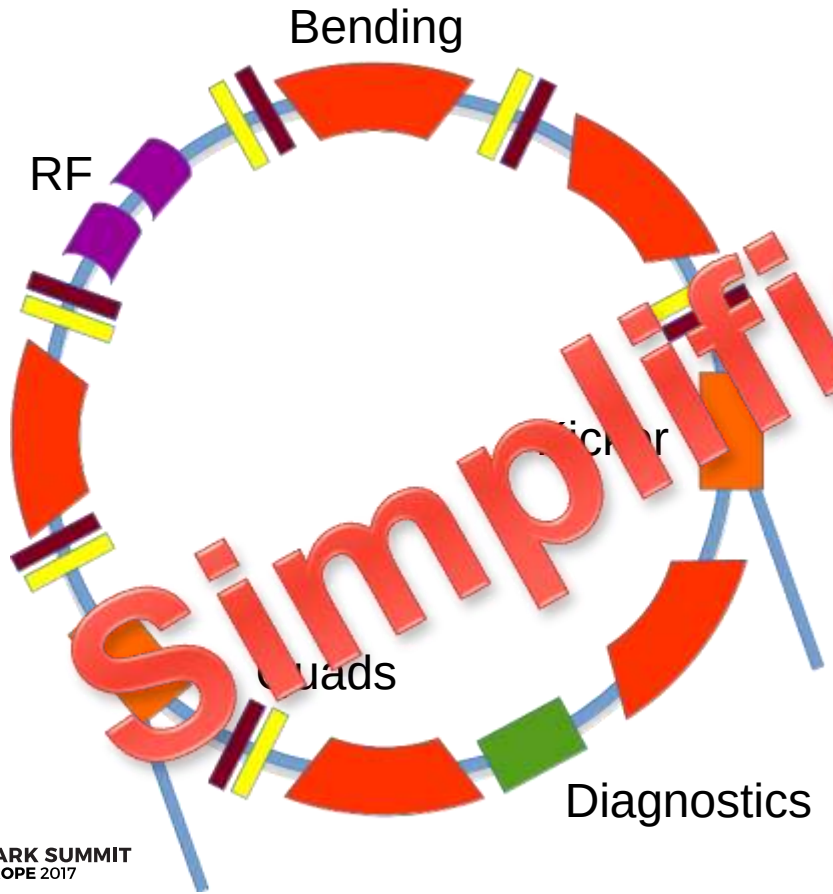
LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron
AD Antiproton Decelerator CTF3 Clic Test Facility AWAKE Advanced WAKefield Experiment ISOLDE Isotope Separator OnLine DEvice
LEIR Low Energy Ion Ring LINAC LINear ACcelerator n-ToF Neutrons Time Of Flight HiRadMat High-Radiation to Materials

Accelerator Timing

- **Beam time shared** per **experiment** in **cycles**
  - of 1.2 sec, 2.4 sec, 3.6 sec, ...



Accelerator Hardware



Simplification!

- **Magnets**
  - Dipoles - bending
  - Quadrupoles - focusing
  - Kickers - inject/extract
- **Radio-Freq. cavities** - to accelerate
- **Beam diagnostics** (intensity, position)

- Timing & synchronization systems
- Orbit feedback & steering systems
- Radiation monitors
- Security systems
- Vacuum, cooling & ventilation
- High voltage electrical systems

- **Specialized controls electronics**

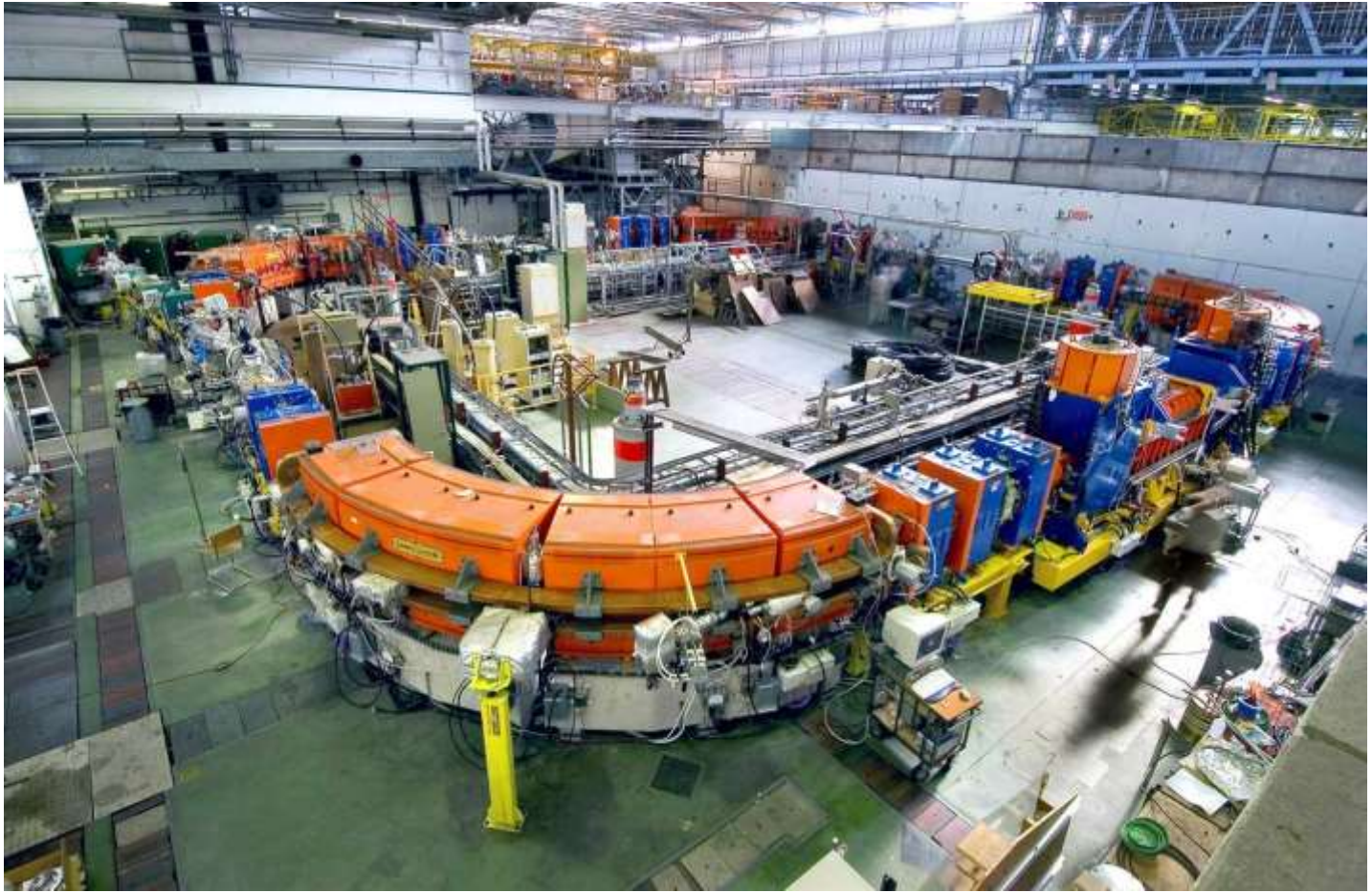
Hardware Controls

In practice

- **Thousands** of different devices
- **Hundreds** of properties each

LEIR

Low Energy Ion Ring
@CERN



Controls Data Logging

- **Equipment readings** (acquisitions or settings)
  - **Not physics data** from experiments!
  - Needed to **operate the accelerators**

- **Online monitoring & ad-hoc queries**
  - Alarms, device/system failures
  - Beam degradation or dumps

- **Offline analytics & studies**
  - Beam quality improvements
  - New beam types
  - Future experiments or machines

CERN Accelerator Logging Service

- **Old system** (**CALS**) based on **Oracle (2 DBs)**
  - **~20,000 devices (**from **~120,000** devices**)**
  - **1,500,000** signals
  - **5,000,000** extractions per day
  - **71,000,000,000** records per day
    - **2 TB / day (unfiltered data, 2 DBs)**
  - **1 PB of total data storage (heavily filtered in 95%)**

Current Issues With CALS

- **Performance / scalability** problems
  - Difficult to scale horizontally
  - *“… to extract 24h of data takes 12h”*

- **CALS & tools not ready for Big Data!**
  - *Have to extract subsets of data to do analysis!*

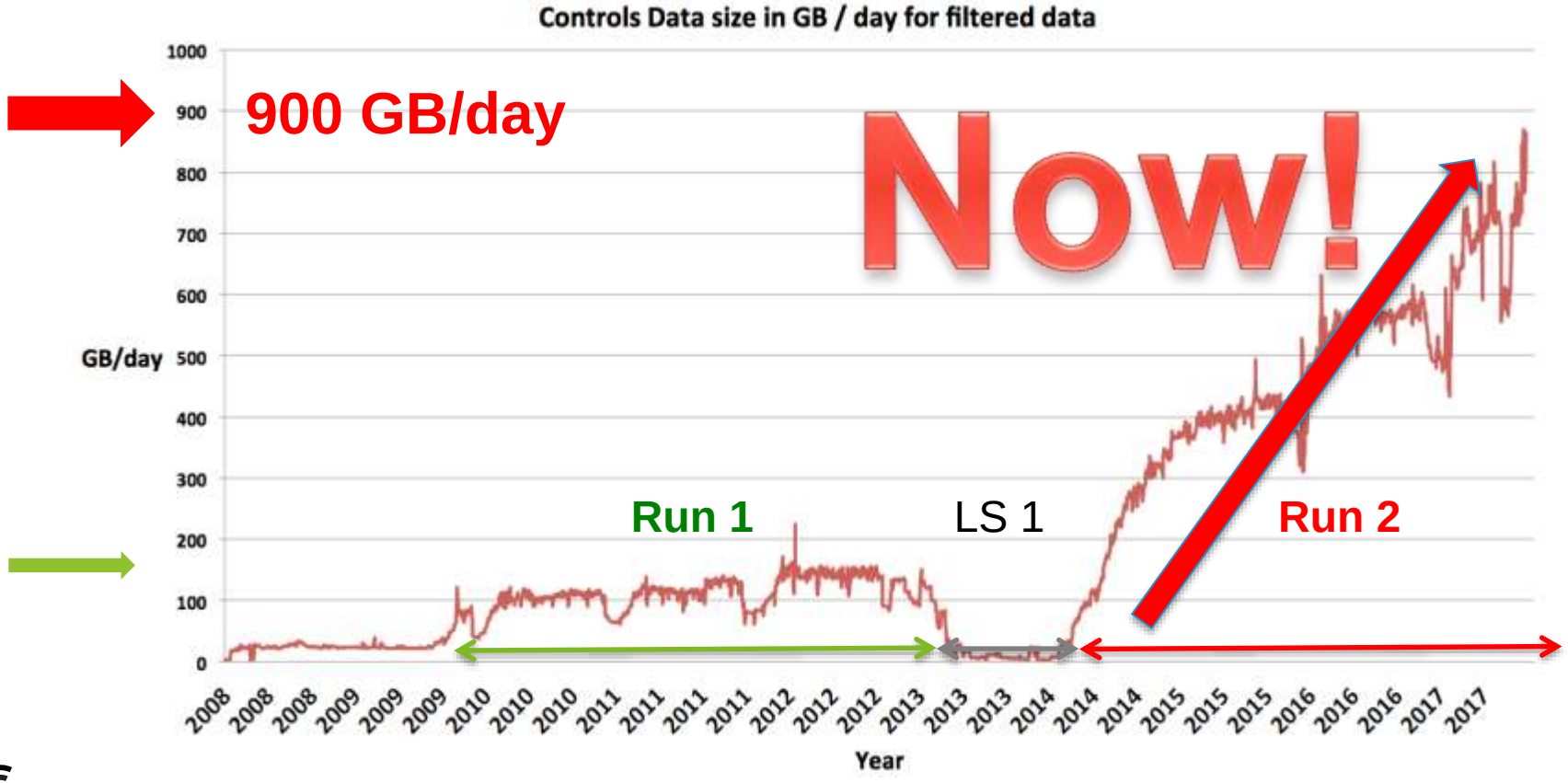
Upcoming Challenges

- **Future big scientific projects**
  - High Luminosity LHC (**HL-LHC**)
  - Future Circular Collider **(FCC)**

- Important **increase of data loads**
  - Data **frequency increase** *(from 1Hz to 100Hz)*
  - Much **bigger vectors**
  - Limited **filtering**

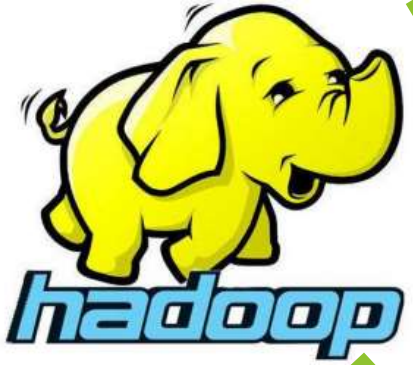
HL-LHC Luminosity Forecast

Current Controls Data Storage



Big Data For Controls?

CALS



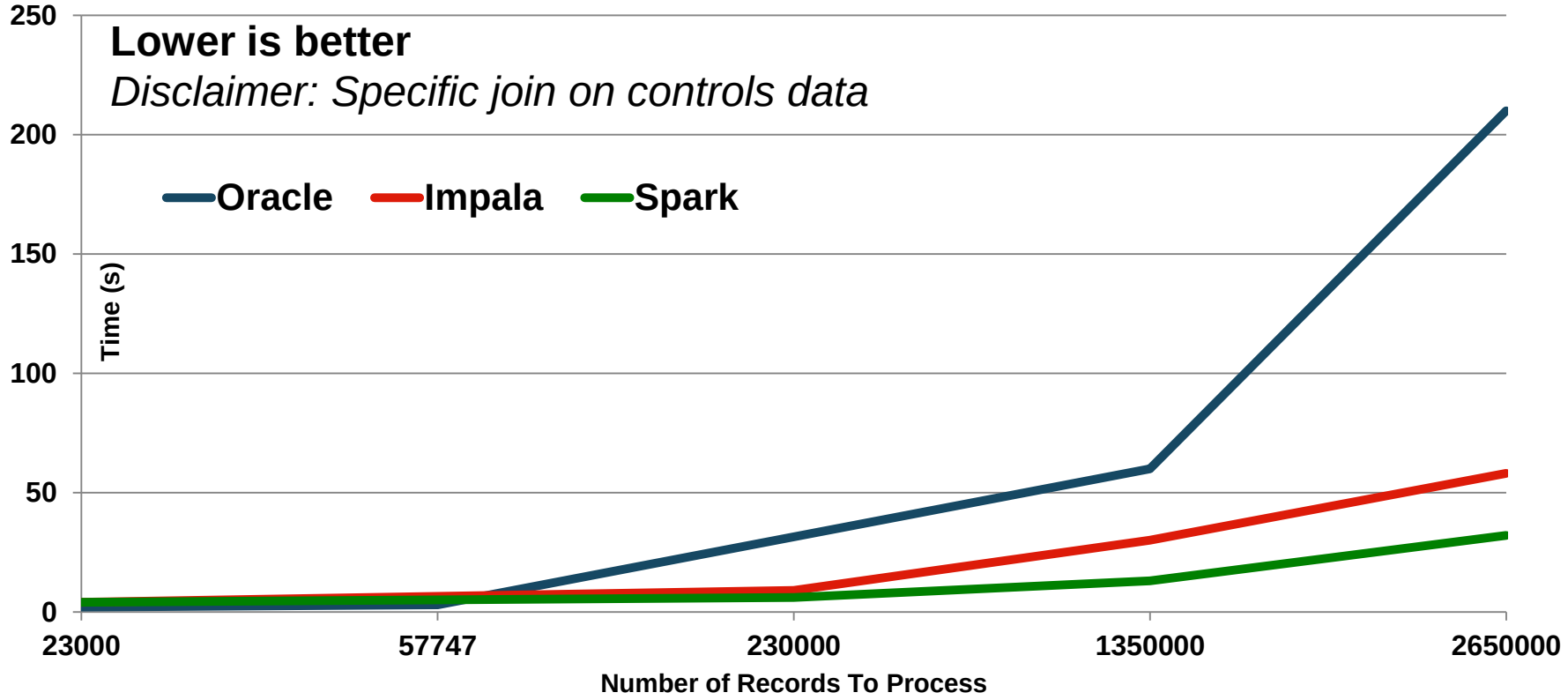
Spark

Impala

Kudu

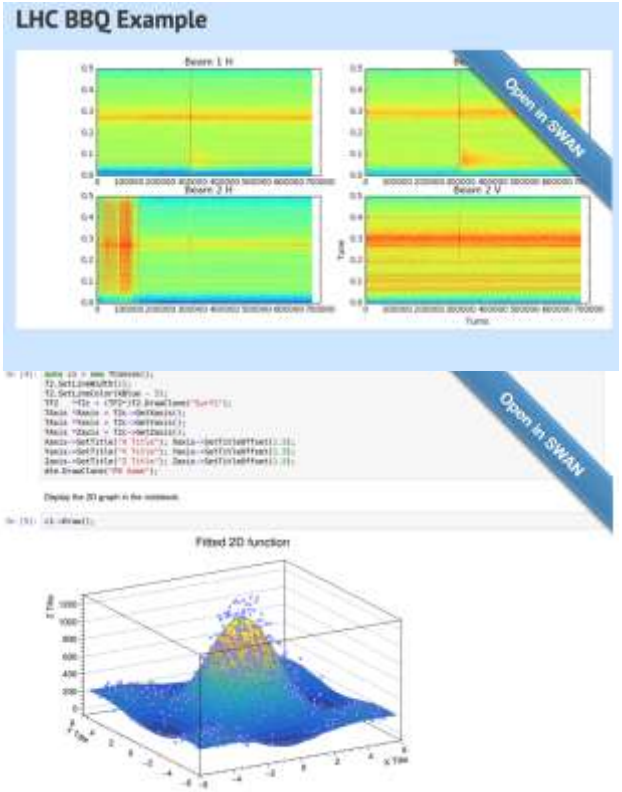
Next CALS (NXCALS)

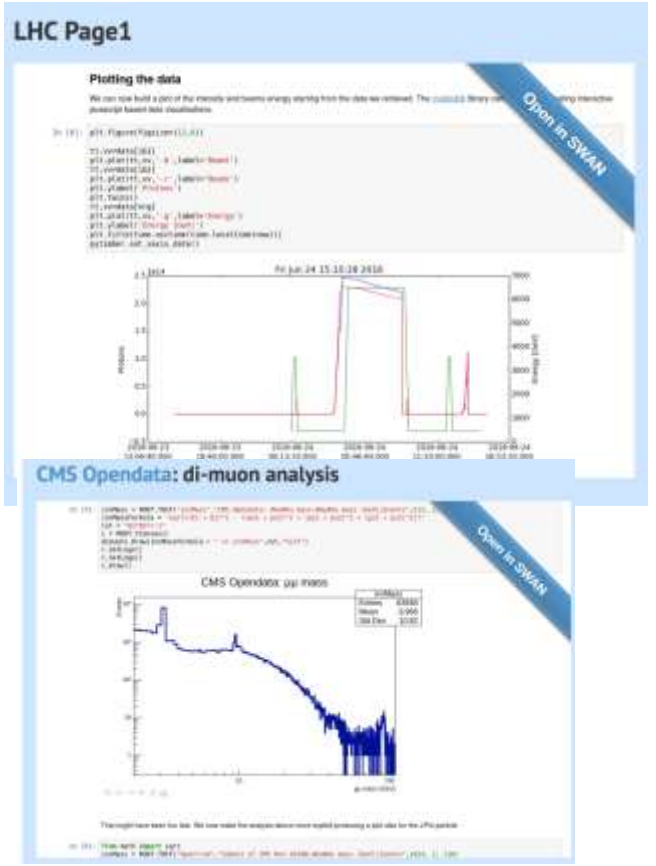
Query Performance Results



Python on Jupyter Platform

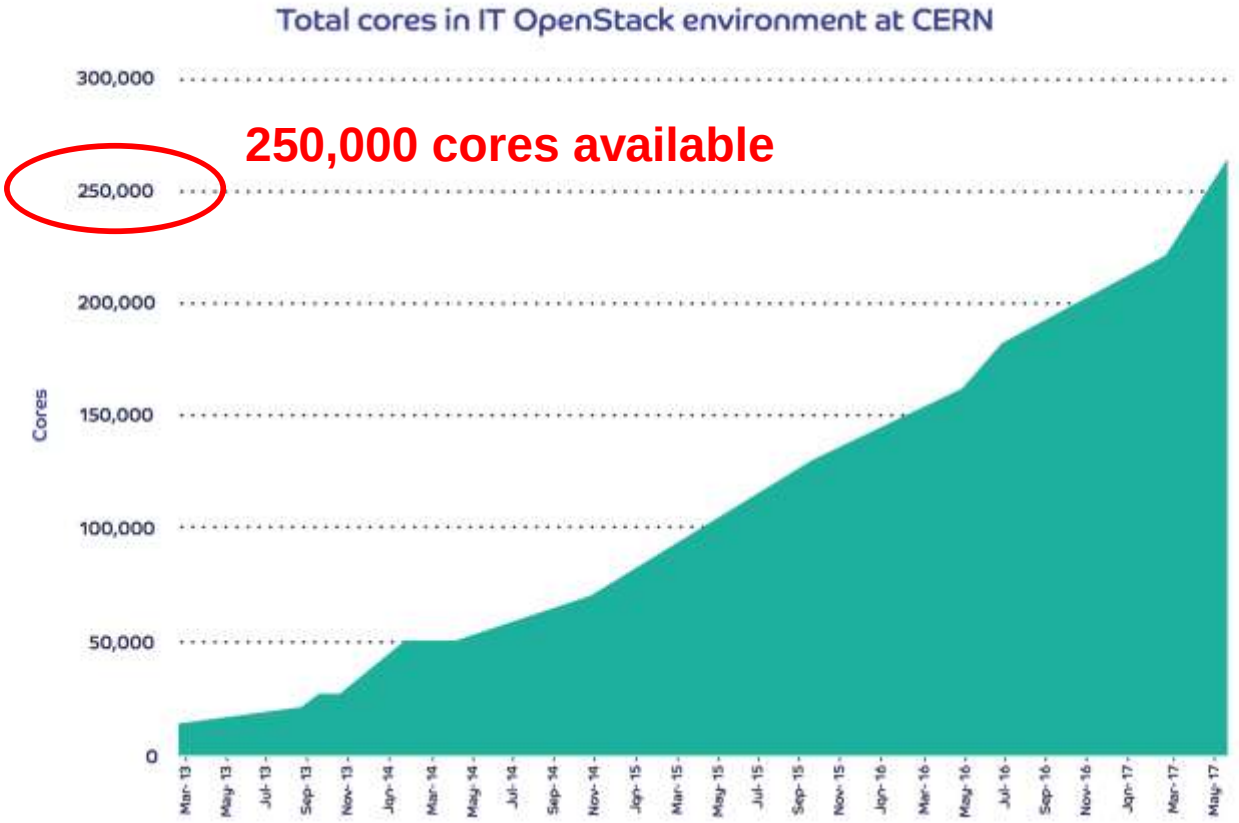
Data Science & Scientific Computing



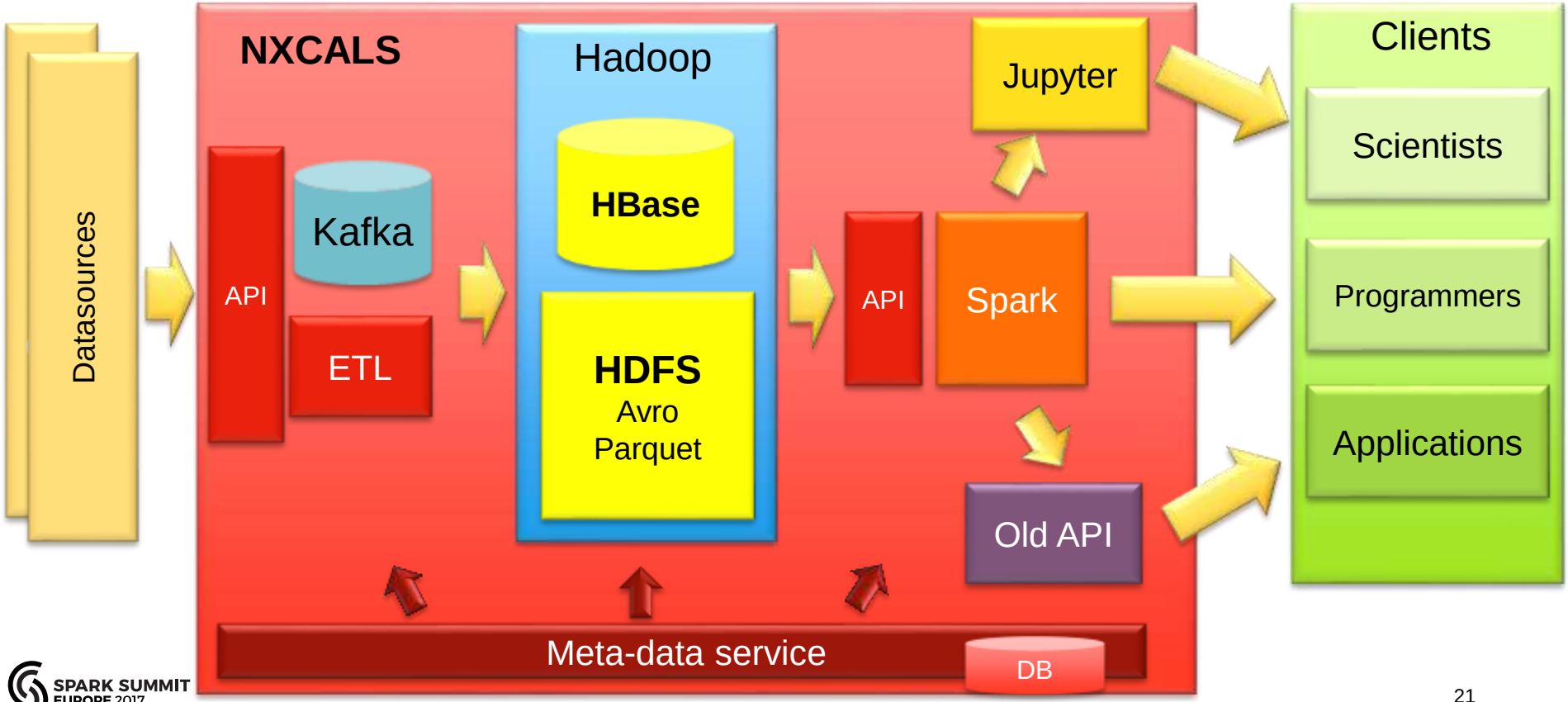


SPARK SUMMIT EUROPE 2017

CERN On-Premise Cloud



NXCALS Architecture



Future Directions

- **Visualization** for **Analytics**
- **S**ervice for **W**eb based **An**alysis (**SWAN**)
  - With **NXCALS** for **Accelerator Controls data…**
  - …and **Spark** as **first class citizen**

- **Unified Software Platform**
  - **Interactive data analysis** in the cloud

Conclusions

- **Technology** has to answer **scientific needs**
- ***Big Data*** is present in **Accelerator Controls**

- CERN started a ***sector-wide joint effort***
  - ***NXCALS*** (Controls Big Data Storage)
  - ***SWAN*** (Analysis & Visualization)

- **To promote and support science**

- **Spark** is ***our choice*** for data extraction & analytics!

Links

- **NXCALS** - https://gitlab.cern.ch/acc-logging-team/nxcals
  - **Please come to the technical talk today afternoon!**

- **SWAN** - https://swan.web.cern.ch/

- Science, accelerators, experiments
  - https://greybook.cern.ch
  - https://home.cern/about/accelerators
  - PS virtual visit: https://goo.gl/1zqSTA
- Lectures from accelerator schools, summer courses
  - http://cas.web.cern.ch/previous-schools
  - https://indico.cern.ch/category/345/
- Beams Department, Controls Group - https://be-dep-co.web.cern.ch/
- Accelerators state - https://op-webtools.web.cern.ch/vistar/vistars.php?usr=SPS1

- Take Part: https://jobs.web.cern.ch/