# SWAN protoservice

Massimo Lamanna IT/ST
Danilo Piparo EP/SFT

- Initiated as a collaboration between EP SFT and IT ST
  - CERNBox and ROOT teams
  - Users in several departments/activities
  - Presently operated jointly by EP SFT and IT ST
  - Initial emphasys on Jupyter notebooks, Python ecosystem and files sync&share
- Recently being extended to reach out the Spark world
  - Collaboration between IT DB and EP SFT
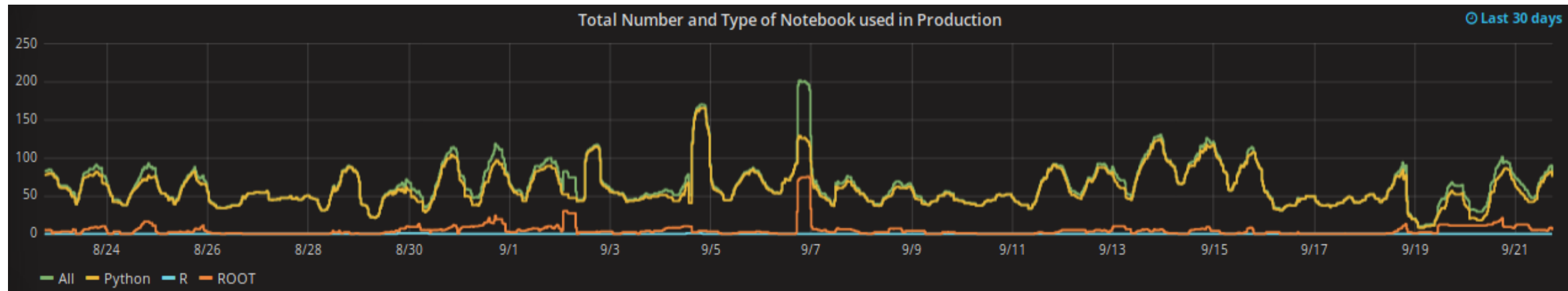
Hadoop Service Commitment to Support NXCALS

Central logging and analytics platform for all the accelerator systems
Critical service for LHC – no data, no beam
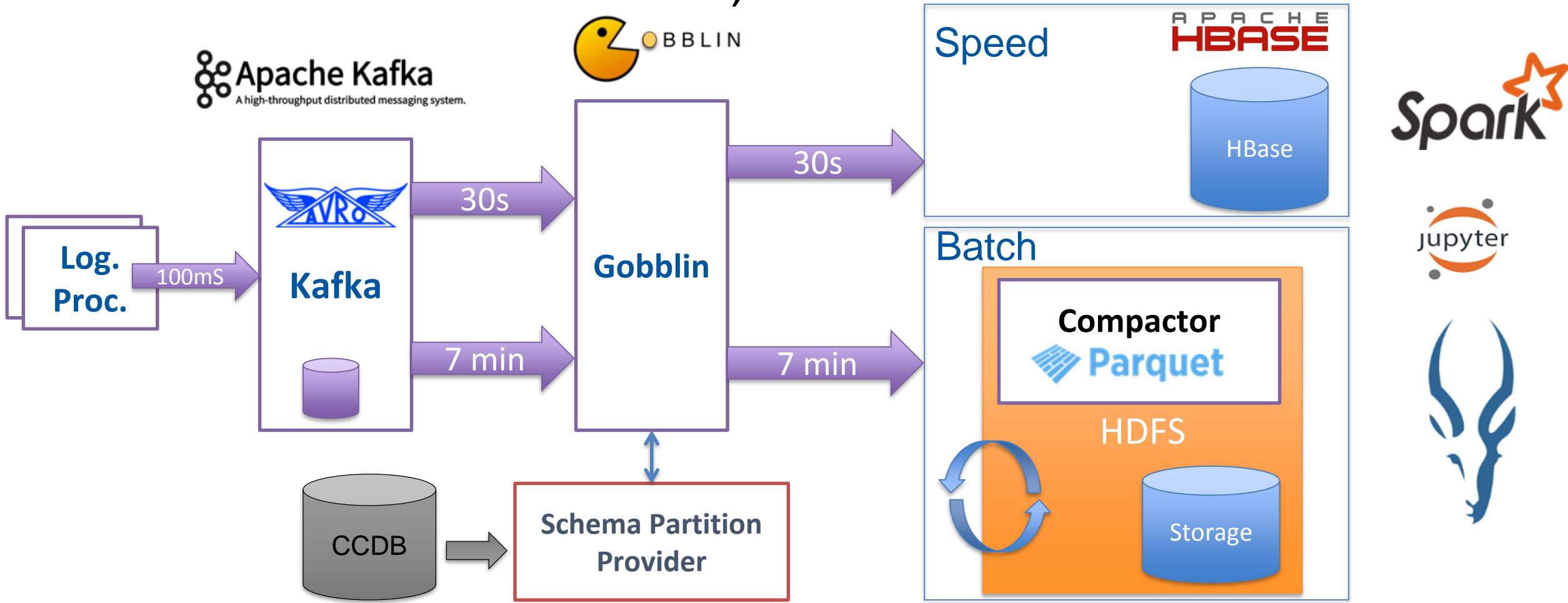
Credit: Jakub Wozniak, BE-CO-DS

- Current throughput:
  - 700 GB/day, 60M events/day

Hadoop Clusters at CERN IT

- 3 current production clusters (+ 1 for QA)
- A new system for BE NXCALs (accelerator logging) platform
  - Coming in Q4 2017 + Critical production service

Cluster Name	Configuration	Primary Usage
lxhadoop	18 nodes (cores – 576, Mem – 1.15TB, Storage – 1.17 PB)	Experiment activities
analytix	36 nodes (cores – 780, Mem – 2.62TB, Storage – 3.6 PB)	General Purpose
hadalytic	12 nodes (cores – 384, Mem – 768GB, Storage – 2.15 PB)	SQL oriented installation
NxCALS	24 nodes (cores – 1152, Mem – 12TB, Storage – 4.6 PB, SSD - 92 TB)	Accelerator Logging Service

SWAN and Hadoop, Spark

- IT-DB + EP-SFT worked on integration of Hadoop and Spark with Jupyter notebooks (SWAN)
  - Configuration and integration with Docker containers
  - Software stack, alignment with cvmfs
  - Security (Kerberos) + encryption (work in progress)
- Spark + notebooks
  - More use cases
  - BE controls, Physics (CMS Big Data, Totem) + potentially many more