Overview of Big Data Solutions and Services at CERN

Luca Canali, IT-DB Hadoop and Spark Service
CERN Knowledge Transfer Forum meeting
CERN, September 29th, 2017

Data at scale @CERN

- Physics data – we use WLCG to handle it
    - Optimised for physics analysis and concurrent access
    - ROOT framework - custom software and data format

- Infrastructure data and metadata
    - Accelerators and detector controllers
    - Data catalogues (collisions, files etc)
    - Monitoring of the WLCG and CERN data centres
    - Systems logs

Modern Distributed Systems for Data Processing

- Tools from industry and open source
  - "Big Data"
  - Distributed systems for data processing
    - Can operate a scale
    - Typically on clusters of commodity-type servers/cloud
    - Many solutions target data analytics and data warehousing
    - Can do much more: data ingestion, streaming, machine learning

Declarative Interfaces for Parallelism

- Young technology but already evolved
  - It is not about SQL vs. no SQL, Map-Reduce
  - SQL is still strong (+ not only SQL, functional languages, etc)
- Systems for data analytics deploy declarative interfaces
  - Tell the system what you want to do
  - Processing is transformed into graph (DAG) and optimized
  - Execution has to be fault-tolerant and distributed

Data Engines on Hadoop Ecosystem

- Several solutions available
  - Pick your data engines and storage formats
- Data-analytics and data warehouse
  - Hadoop / “Big Data Platforms” are often the preferred solution
  - Cost/performance and scalability are very good
- Online systems
  - Competition still open with RDBMS and new in-memory DBs
  - Added value: build platforms to do both online + analytics

Hadoop Ecosystem – The Technology

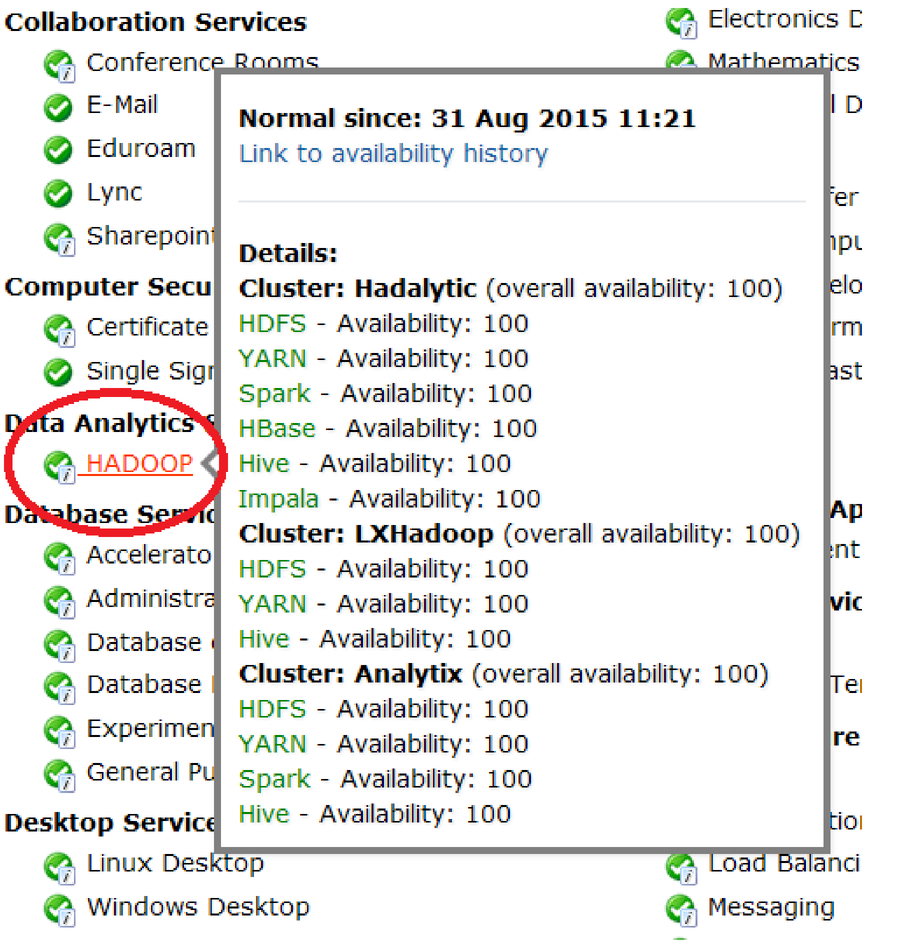
- Hadoop clusters: YARN and HDFS
- Notable components in the ecosystem
  - Spark, HBase, Map Reduce
  - Next generation: Kudu
- Data ingestion pipelines
  - Kafka, Spark streaming

Managed Services for Data Engineering

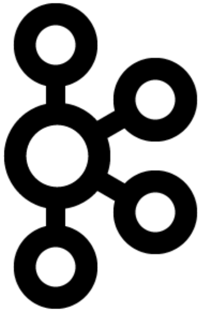
- Platform
  - Capacity planning and configuration
  - Define, configure and support components
- Running central services
  - Build a team with domain expertise
  - Share experience
  - Economy of scale

Hadoop Service at CERN IT

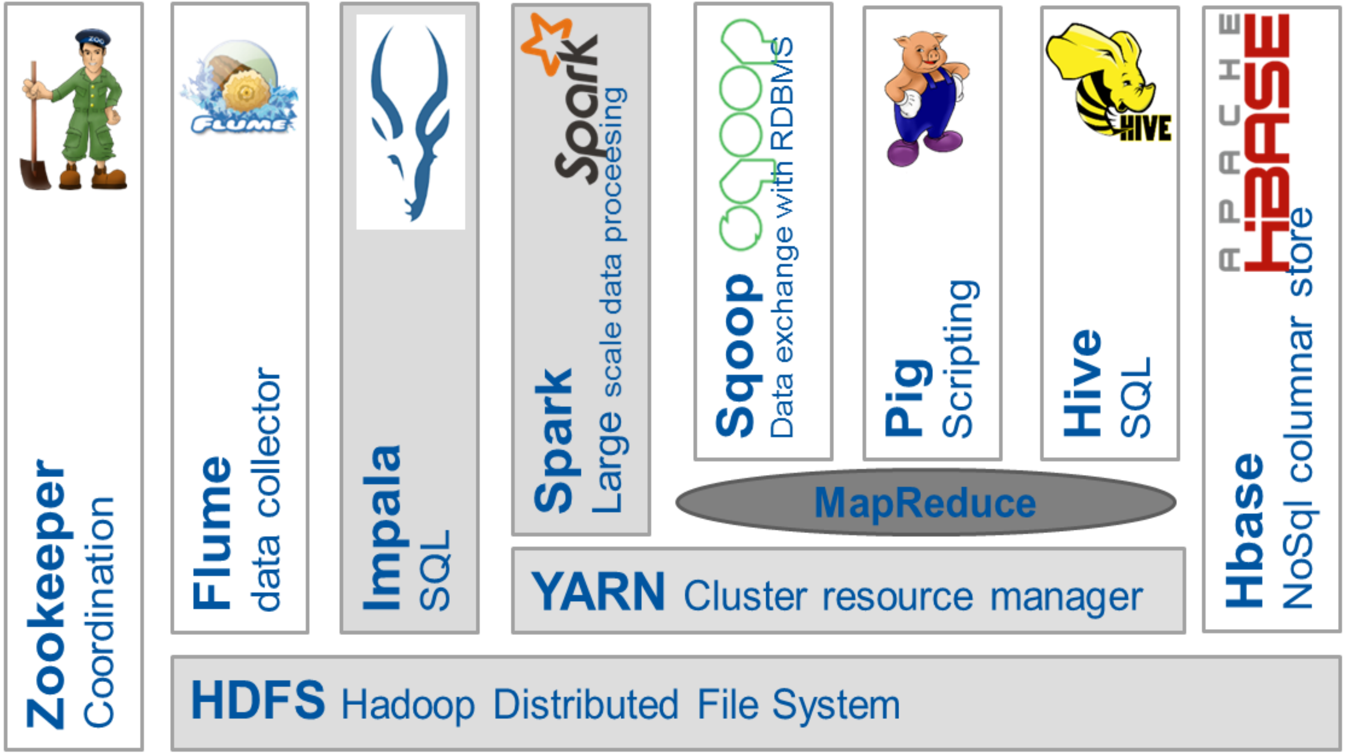
- Setup and run the infrastructure
- Provide consultancy
- Support user community
- Running for more than 2 years



Overview of Available Components



Kafka:
streaming
and ingestion



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

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

Data volume (from backup stats July2017)

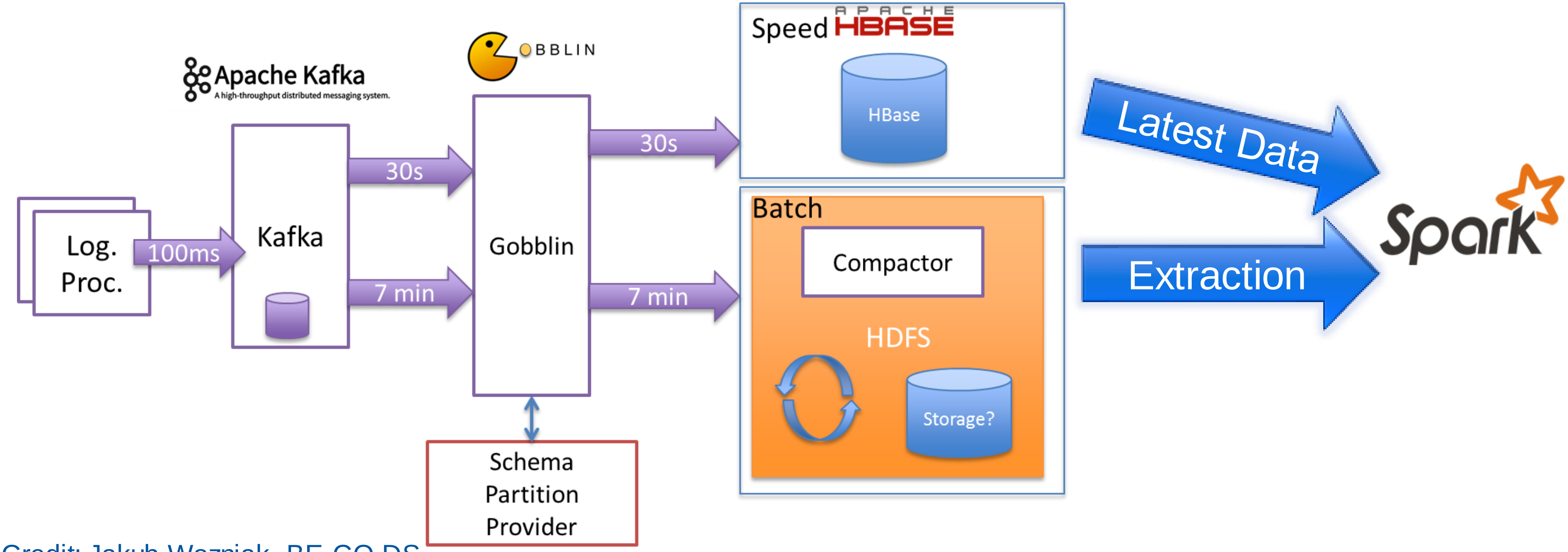
Application	Current Size	Daily Growth
IT Monitoring	420.5 TB	140 GB
IT Security	125.0 TB	2048 GB
NxCALS	10.0 TB	500 GB
ATLAS Rucio	125.0 TB	~200 GB
AWG	90.0 TB	~10 GB
CASTOR Logs	163.1 TB	~50 GB
WinCC OA	10.0 TB	25 GB
ATLAS EventIndex	250.0 TB	200 GB
USER HOME	150.0 TB	20 GB
Total	**1.5 PB**	**4 TB**

Highlights and Use Cases

- Accelerator logging
- Industrial controls
- Streaming, data enrichment, analytics
  - Monitoring team
  - Security team
- Physics
  - Development of “Big Data solutions” for physics
  - Analytics, for experiments computing

Next Gen. Archiver for Accelerator Logs

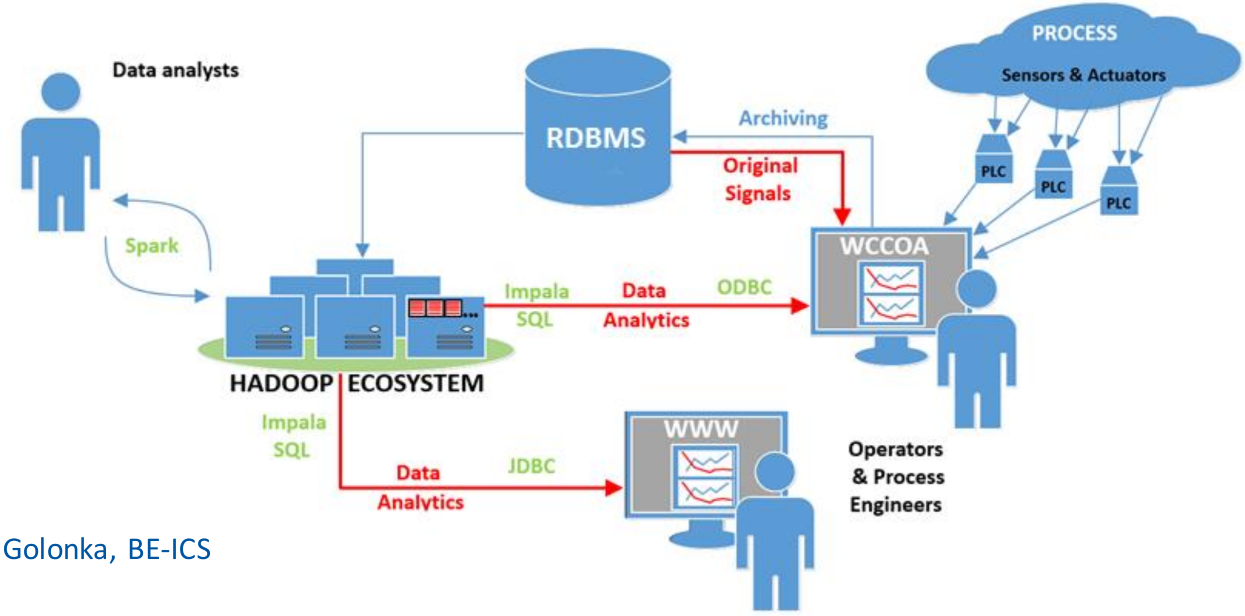
Pilot architecture tested by CERN Accelerator Logging Services
Critical system for running LHC - 700 TB today, growing 200 TB/year
Challenge: service level for critical production



Credit: Jakub Wozniak, BE-CO-DS

Industrial control systems

- Complex monitoring and metric archiving of devices in the LHC tunnel and detectors
  - Current data rates: 250kHz, 500GB/day

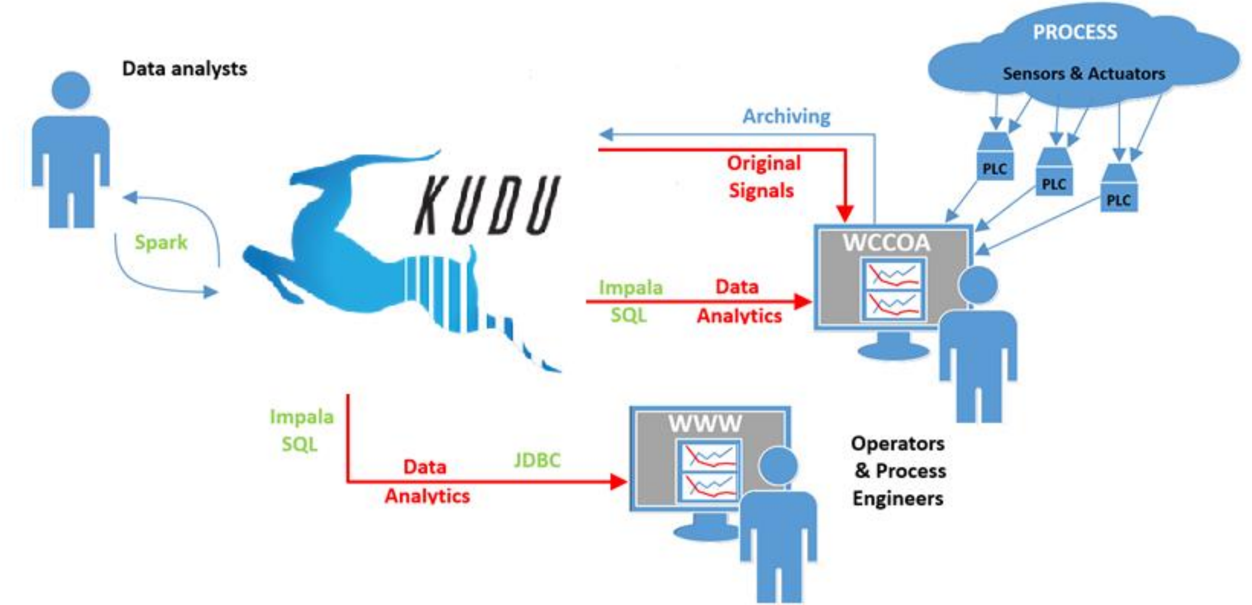


Credit: Piotr Golonka, BE-ICS

Possible Evolution - SCADA controls

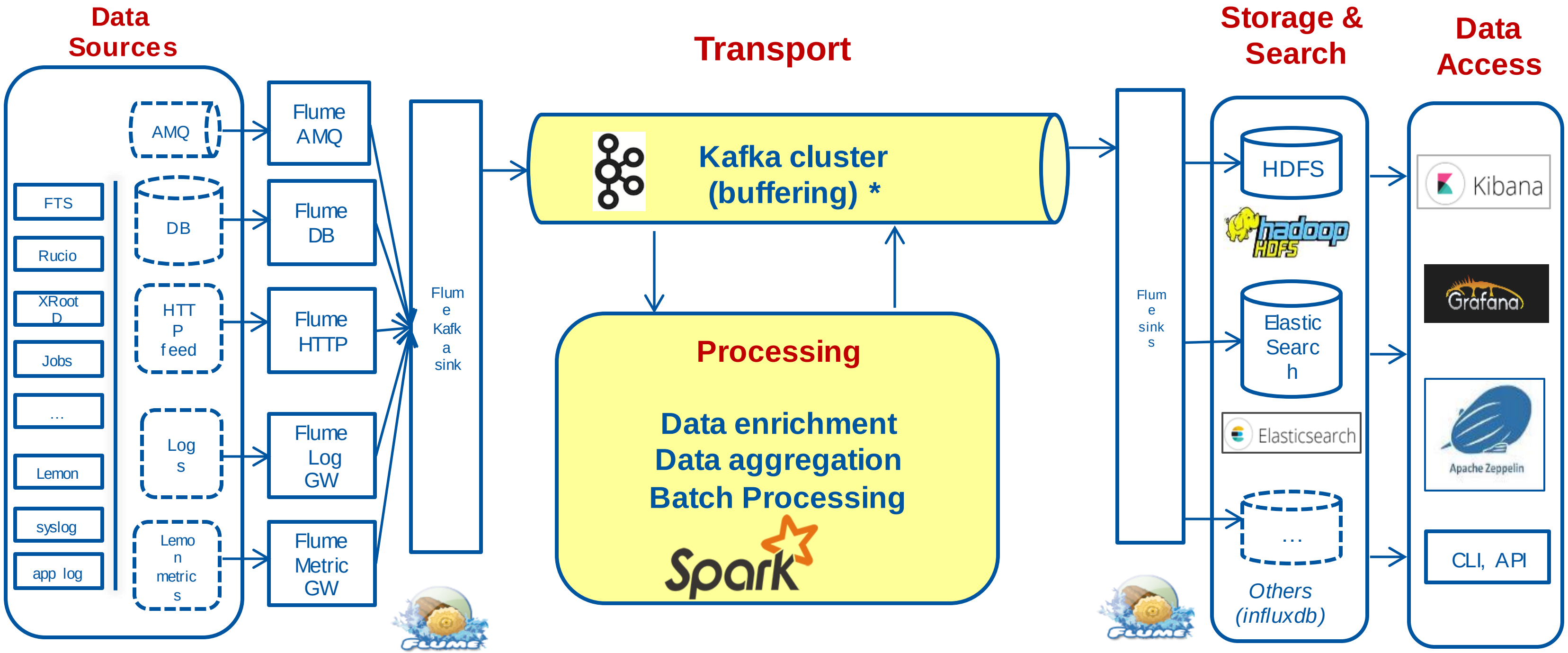
What can we gain with Kudu:

- reduce ingestion latency for analytics
- speed up live data queries and reporting
- simplification of the svstem



New IT Monitoring

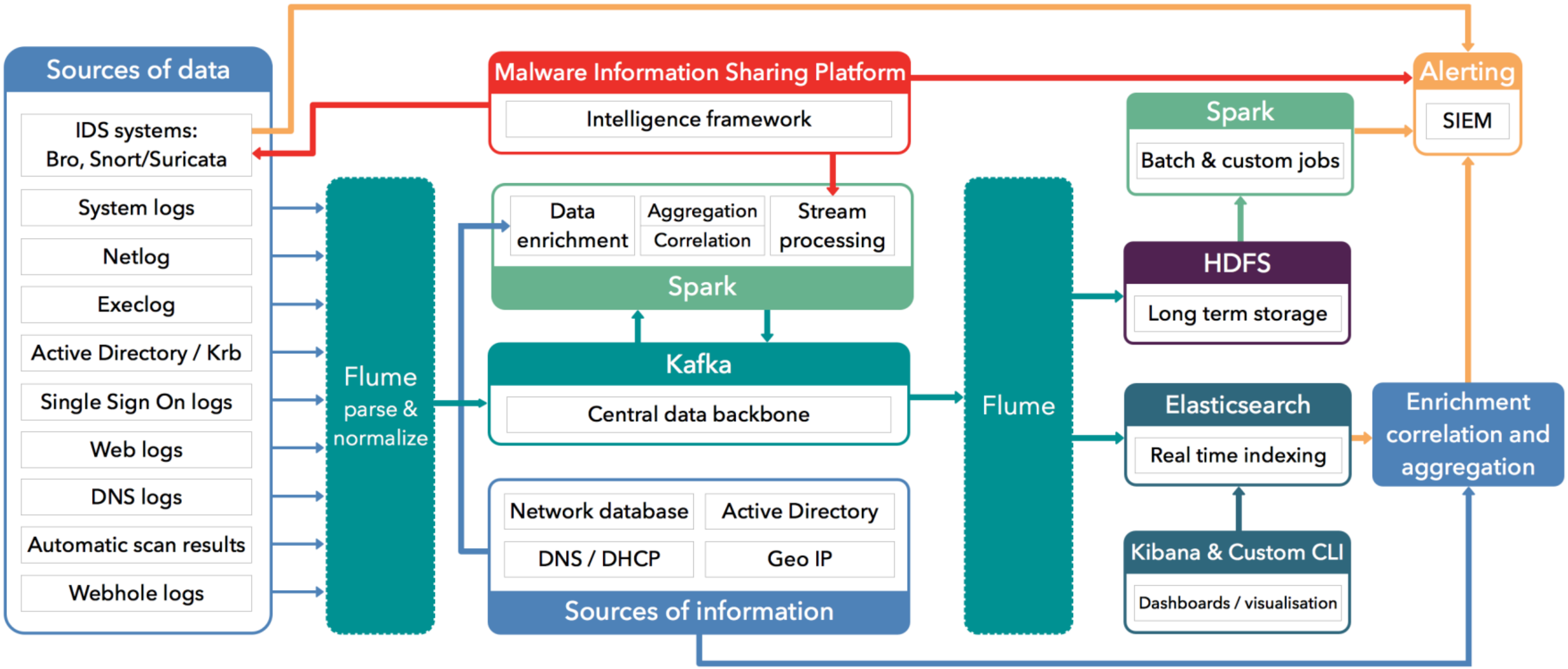
Critical for CC operations and WLCG



- Data now 200 GB/day, 200M events/day
- At scale 500 GB/day
- Proved effective in several occasions

Credits: Alberto Aimar, IT-CM-MM

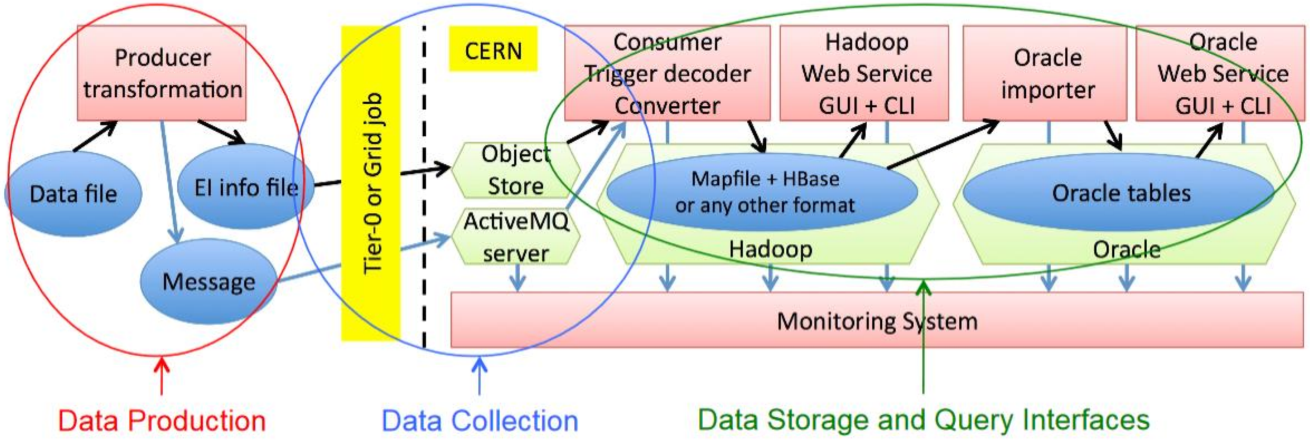
Computer Security intrusion detection use cases



Credits: CERN security team, IT-DI

ATLAS EventIndex

- Searchable catalog of ATLAS events
  - First “Big Data” project in our systems
  - Over 80 billions of records, 140TB of data

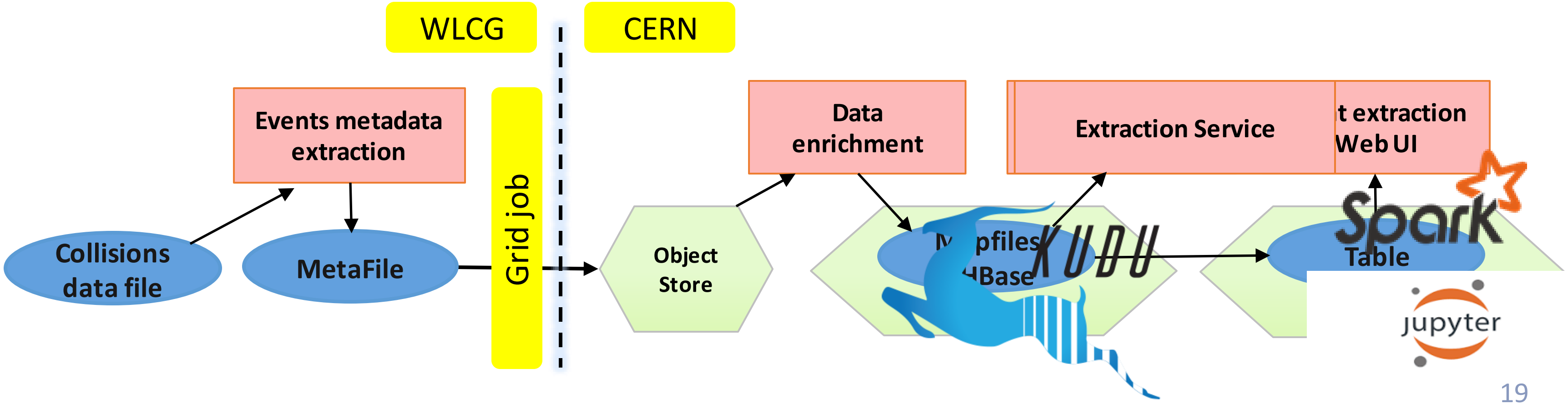


Credits: Dario Barberis, 2017

Possible evolution -> ATLAS EventIndex

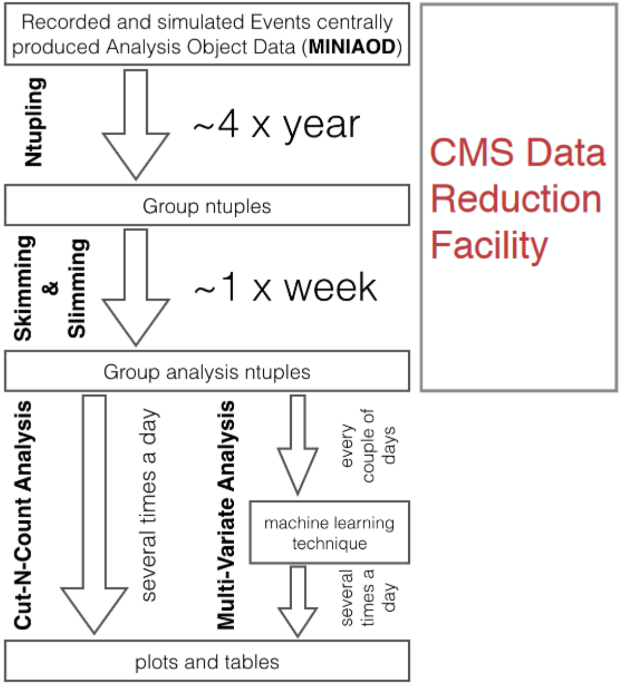
What can we gain with Kudu:

- Reduce ingestion latency by removal of multi-staged data loading into HDFS
- Enable in place data mutation
- Enable common analytic interfaces Spark and Impala
- ...and increase analytic performance



CMS Big Data Project and Openlab

Proposal: CMS Data Reduction Facility



- Demonstration facility optimized to read through petabyte sized storage volumes
  - Produce sample of reduced data based on potentially complicated user queries
  - Time scale of hours and not weeks
- If successful, this type of facility could be a big shift in how effort and time is used in physics analysis
  - Same infrastructure and techniques should be applicable to many sciences

Physics Analysis and "Big Data" ecosystem

- Challenges and goals:
  - Use tools from industry and open source
    - Current status: Physics uses HEP-specific tools
    - Scale of the problem 100s of PB – towards exascale
  - Develop interfaces and tools
    - Already developed first prototype to read ROOT files into Apache Spark
    - Hadoop-XRootD connector -> Spark can read from EOS
  - Challenge: testing at scale

Jupyter Notebooks and Analytics Platforms

- Jupyter notebooks for data analysis
  - System developed at CERN (EP-SFT) based on CERN IT cloud
  - SWAN: Service for Web-based Analysis
  - ROOT and other libraries available

- Integration with Hadoop and Spark service
  - Distributed processing for ROOT analysis
  - Access to EOS and HDFS storage
  - Opens the possibility to do physics analysis on Spark using Jupyter notebooks as interface
  - An example notebook with CERN/LHCb opendata -> https://cernbox.cern.ch/index.php/s/98RK9xIU1s9Lf08

Jupyter Notebooks and Analytics Platforms

Apache Spark

LHCb Opendata

This notebok illustrates a nice interplay between LHCb Opendata, Spark and ROOT - all teaching interesting flavour Physics!

Plotting a feature:

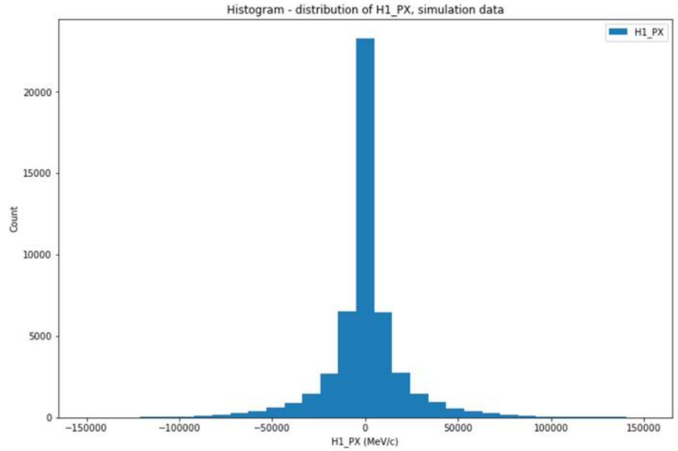
You can plot any feature of the data in a histogram. Choose any suitable binning that allows you to observed the distribution of the variable clearly. [illegible]ing a histogram for the first kaon candidate's momentum x-component (H1_PX):

Open in SWAN

In [8]:
```
# Plot a histogram of the distribution of the H1_PX variable, using Pandas

# This is a basic solution that  moves all the data from the Spark DataFrame
# into a Python Pandas DataFrame. It's OK for small sata sets, but it has scalability issues

h1px_data = sim_data_df.select("H1_PX").toPandas() # select H1_PX data and moves it to Pandas
h1px_data.plot.hist(bins=31, range=[-150000, 150000], title="Histogram - distribution of H1_PX, simulation data")
xlabel('H1_PX (MeV/c)')
ylabel('Count');
```



SWAN Customisation

Specify the parameters that will be used to contextualise the container which is created for you. See the online SWAN guide for more details.

Software stack more...

89

Platform more...

x86_64-slc6-gcc49-opt

Environment script more...

e.g. $CERNBOX_HOME/MySWAN/myscript.sh

Number of cores more...

Number of cores to associate to the container.

1

Spark cluster more...

Hadalytic

Start my Session

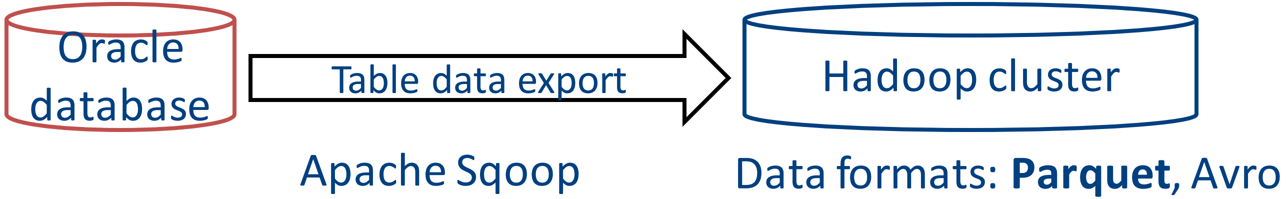
Jupyter Notebooks and Analytics Platforms

jupyter

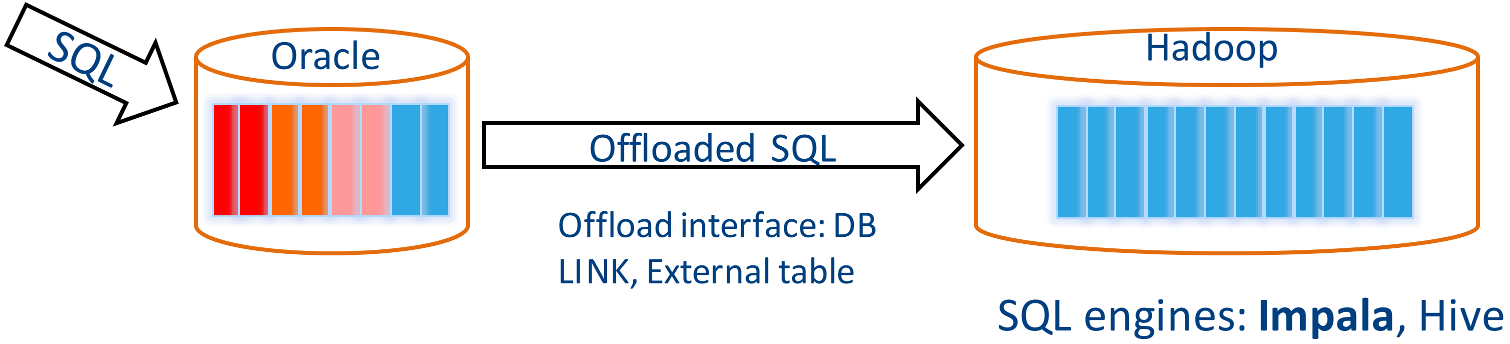
ROOT
Data Analysis Framework

Offloading from Oracle to Hadoop

- Step1: Offload data to Hadoop



- Step2: Offload queries to Hadoop

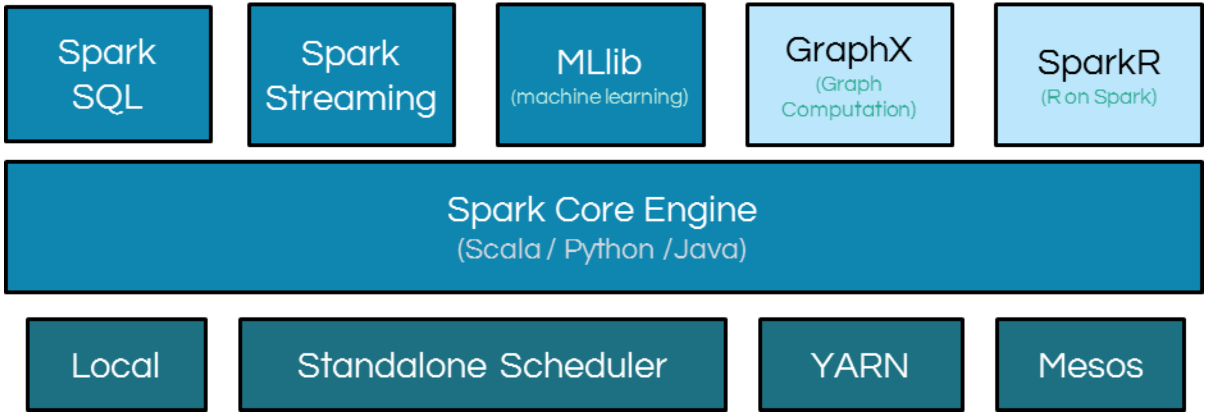


Analytics platform for controls and logging

- Use distributed computing platforms for storing analyzing controls and logging data
  - Scale of the problem 100s of TBs
- Build an analytics platform
  - Technology: focus on Apache Spark
  - Empower users to analyze data beyond what is possible today
  - Opens use cases for ML on controls data

Apache Spark

- Powerful engine, in particular for data science and streaming
  - Aims to be a “unified engine for big data processing”
- At the center of many “Big Data”, Streaming and ML solutions



Engineering Efforts to Enable Effective ML

- From “Hidden Technical Debt in Machine Learning Systems”, D. Sculley at al. (Google), paper at NIPS 2015

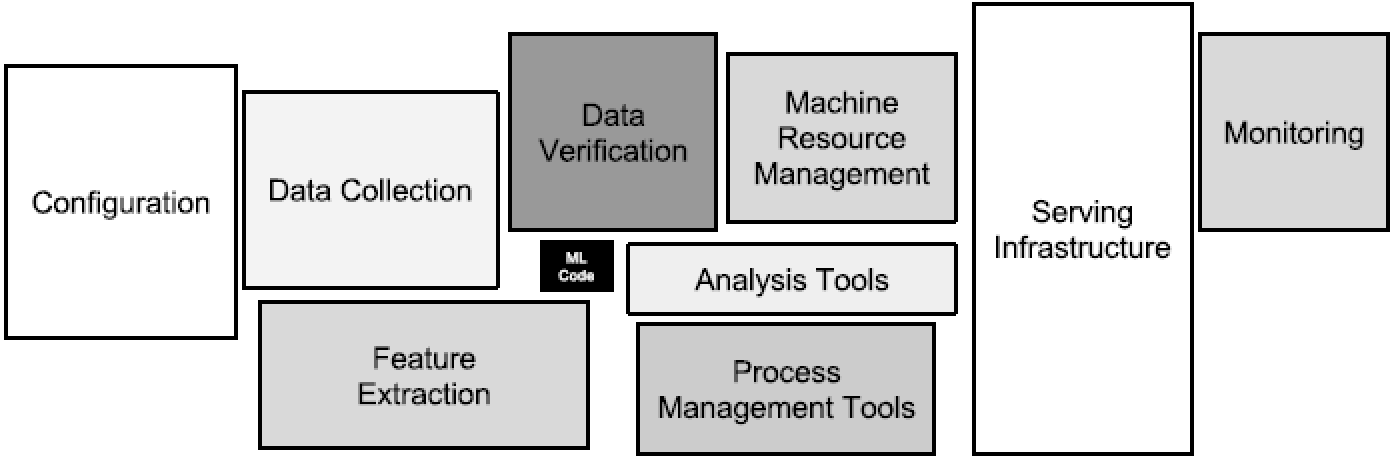
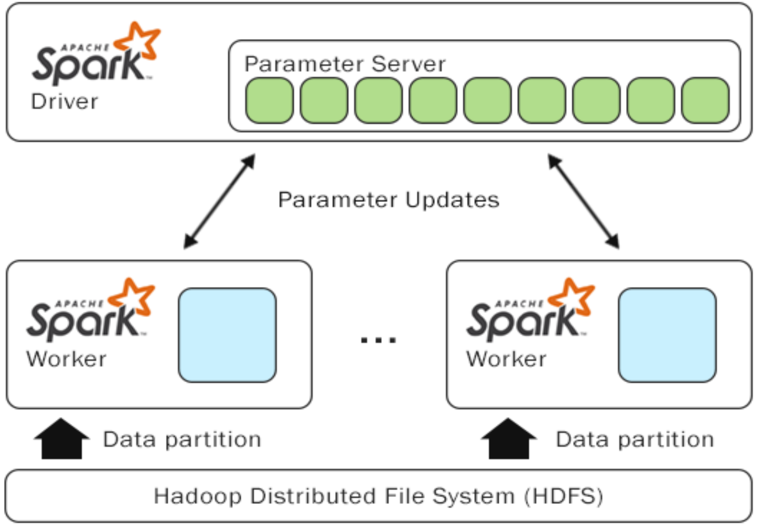


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Machine Learning with Spark

- Spark has tools for machine learning at scale
    - Spark library MLlib
- Distributed deep learning
    - Working on use cases with CMS and ATLAS
    - We have developed an integration of Keras with Spark
- Possible tests and future investigations:
    - Frameworks and tools for distributed deep learning with Spark available on open source:
        - BigDL, TensorFlowOnSpark, DL4j, ..
    - Also of interest HW solutions: for example FPGAs, GPUs etc



https://github.com/cerndb/dist-keras
Main developer: Joeri Hermans (IT-DB)

Spark as a Database Engine

- Spark SQL is now mature
  - Feature-rich, scalable, flexible
  - Combine it with data formats and storage solutions and will act as a relational database (for analytics)

Data Analytics (SQL)

Not Only Spark..

- Other components in the ecosystem for database-like workloads
- Analytics
  - Impala, a SQL engine written in C++
- Fast layer:
  - HBASE and Kudu
  - Streaming solutions

In the following,

Some additional thoughts on challenges and opportunities

R&D: Hadoop and Spark on OpenStack

- Tests of deploying Hadoop/Spark on OpenStack are promising
- Appears a good solution to deploy clusters where local storage locality is not needed
  - Example: possible candidates for Spark clusters for physics data processing reading from EOS (or from remote HDFS)
- Also run tests of Hadoop clusters with local storage
  - Using ad-hoc and “experimental configuration” in particular for the storage mapping, thanks to the collaboration with OpenStack team at CERN
  - Promising results, we plan to further explore

R&D: Architecture and Components Evolution

- Architecture decisions on data locality
  - Currently we deploy Spark on YARN and HDFS

- Investigating: Spark clusters without directly attached storage?
  - Using EOS and/or HDFS accessed remotely?
  - EOS integration currently being developed for Spark
  - Spark clusters "on demand" rather than Yarn clusters?
  - Possibly on containers

Scale Up – from PB to EB in 5-10 years?

- Challenges associated with scaling up the workloads
  - Example from the CMS data reduction challenge: 1 PB and 1000 cores
    - Production for this use case is expected 10x of that.
    - New territory to explore
- HW for tests
  - CERN clusters + external resources, example: testing on Intel Lab equipment (16 nodes) in February 2017

Challenges

- Platform
  - Provide evolution for HW
  - Build robust service for critical platform (NXCALS and more) using open source software solutions in constant evolution
- Service
  - Evolve service configuration and procedures to fulfil users needs
- Knowledge
  - Only 2-3 years experience
  - Technology keeps evolving

Training and Teaching Efforts

- Intro material, delivered by IT-DB
  - "Introduction and overview to Hadoop ecosystem and Spark“, April 2017. Slides and recordings at: https://indico.cern.ch/event/590439/
  - 2016 tutorials: https://indico.cern.ch/event/546000/
- More training sessions:
  - Planned for November 2017, presentations + hands-on
    - Introduction and overview to Hadoop ecosystem and Spark. Subscribe at: https://cta.cern.ch/cta2/f?p=110:9:207485681243790::::X_STATUS,X_COURSE_ID:D,5331
    - See also presentations at the Hadoop Users Forum: https://indico.cern.ch/category/5894/

Community

- Recent activity on configuration
  - Contacted with Hadoop admins at SARA
  - Also contacts with Princeton (via CMS Bigdata project)
- Opportunities to share with industry and “Big Data” communities at large
  - See presentations by CERN Hadoop service at Kafka Summit, Strata Data, Spark Summit, XLDB
- More sites interested in Hadoop and Big Data
- Opportunities to share experience across WLCG sites and with other sciences

Conclusions

- Hadoop, Spark, Kafka services at CERN IT
  - Analytics, streaming, ML, logging/controls
- Our goals: service delivery and working on selected projects with the user community
- We are growing
  - Service (new NXCals platform for accelerator logging)
  - Experience and community