

# Big Data Analytics for Future Circular Collider Reliability and Availability Studies

V. Begy <sup>1</sup>, A. Apollonio <sup>2</sup>, J. Gutleber <sup>2</sup>, M. Martin-Marquez <sup>2</sup>, A. Niemi <sup>3</sup>,  
J.-P. Penttinen <sup>4</sup>, E. Rogova <sup>5</sup>, A. Romero-Marin <sup>2</sup>, P. Sollander <sup>2</sup>



University of Vienna <sup>1</sup>      CERN <sup>2</sup>  
Tampere University of Technology <sup>3</sup>      Ramentor Oy <sup>4</sup>  
Delft University of Technology <sup>5</sup>

# Future Circular Collider Study [1]

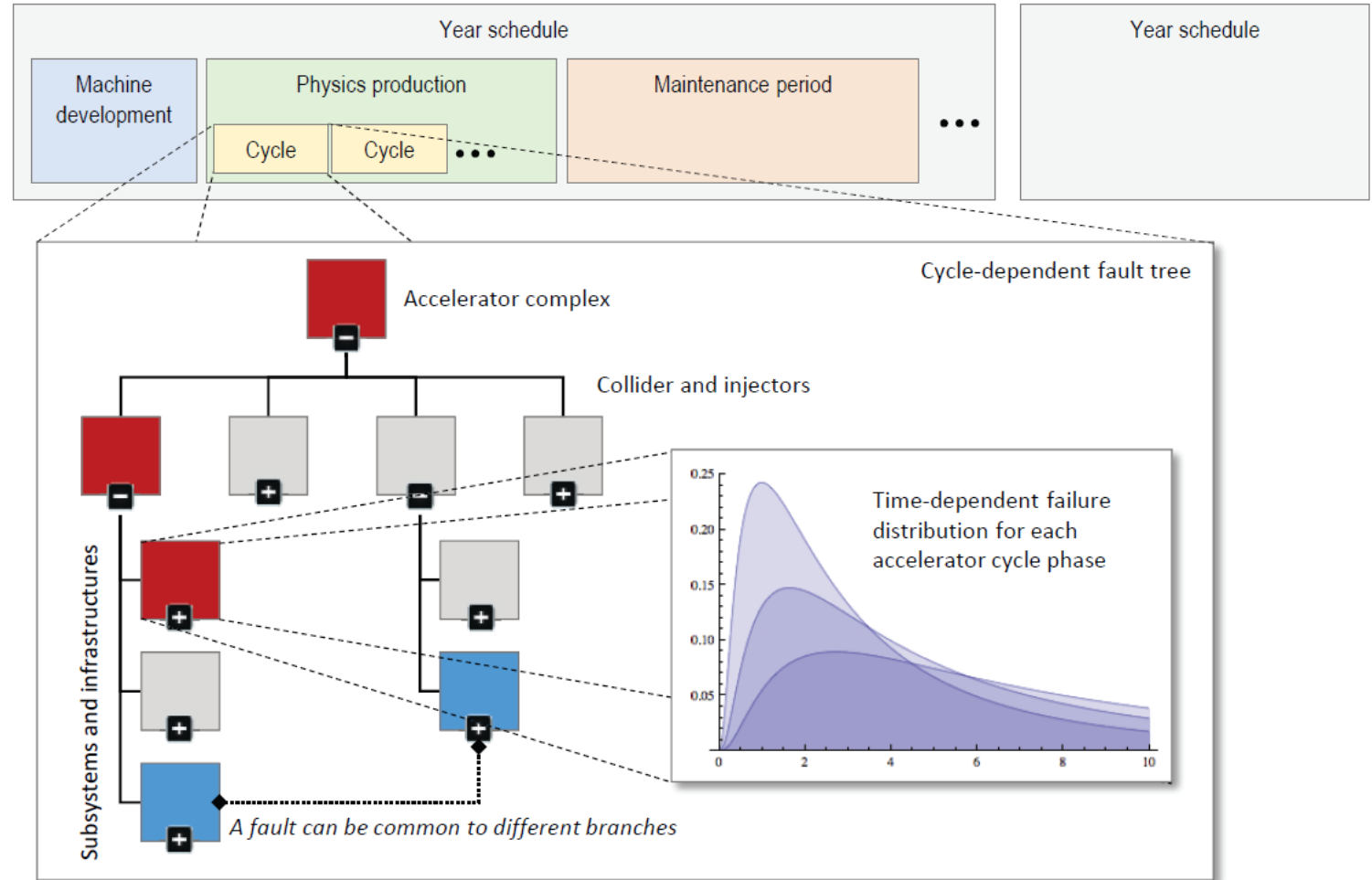
- International collaboration hosted by CERN
- Launched 2014
- Response to a request stated in 2013 Update of the European Strategy for Particle Physics
- With a goal to conceive **conceptual designs** for energy and intensity frontier circular colliders for the post-LHC era

# RAMS Modelling and Simulation [2]

- Integrated luminosity is a key performance indicator for particle colliders
- Depends directly on the **availability** of the accelerator complex
- LHC and its injectors as a case study to assess the potentials of industrial **reliability** engineering methods in the domain of particle accelerators
- Evaluate the feasibility and operability of large scale collider (100 km perimeter)

# RAMS Modelling and Simulation

- Continuous modelling and simulation of subsystem cycle and operation phase-dependent failure distributions
- Predictive models with high fidelity as output



# Challenges: Data Integration & Data Quality

- Significant manual effort to extract, prepare and analyse the heterogeneous operational and maintenance data required
- Limited amount of available data is suitable for reliability and availability analysis
  - Diverging data quality and integrity

# Data Integration & Data Quality

```
1 SELECT * FROM lhclog_cryo
2 WHERE variable_name = "QURCA_4_CV242.POSRST"
3 AND utc_timestamp BETWEEN "2015-01-14" AND "2015-03-14";
```

	variable_name	utc_timestamp	value
1	QURCA_4_CV242.POSRST	2015-03-05 00:00:08.830000000	57.819892883300781
2	QURCA_4_CV242.POSRST	2015-03-05 00:00:24.390000000	57.711101531982422
3	QURCA_4_CV242.POSRST	2015-03-05 00:00:39.310000000	57.654994964599609
4	QURCA_4_CV242.POSRST	2015-03-05 00:01:02.300000000	57.536369323730469
5	QURCA_4_CV242.POSRST	2015-03-05 00:01:12.880000000	57.52679443359375
6	QURCA_4_CV242.POSRST	2015-03-05 00:01:25.950000000	57.614585876464844
7	QURCA_4_CV242.POSRST	2015-03-05 00:01:43.370000000	57.610916137695312
8	QURCA_4_CV242.POSRST	2015-03-05 00:02:17	57.615604400634766
9	QURCA_4_CV242.POSRST	2015-03-05 00:02:30.070000000	57.721435546875

Structured data

Detail of Event 35048

Date22-APR-2015 00:30

SystemLHCB

Sub equipmentCompresseur

CategoryReporting

SeverityProblem solved

Location / Siteproduction / P6

Title2CV210 doesn't keep its position

FaultInstrumentation failure

Actions2CV210 doesn't keep its position. Amperage doesn't change, so the valve is real open. Probably a failure on the actuator since 09/03/2015.

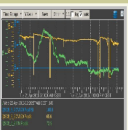
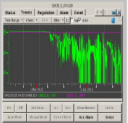
Team P6 will check that.

AuthorSDEMAS

Relation with EVT:

Date	Comments	Author
27-APR-2015 11:10	le WO a ete fait (numero 21863236)	MBARATIE

Documents attached

Preview	Comments	Author
	-	SDEMAS
	-	SDEMAS

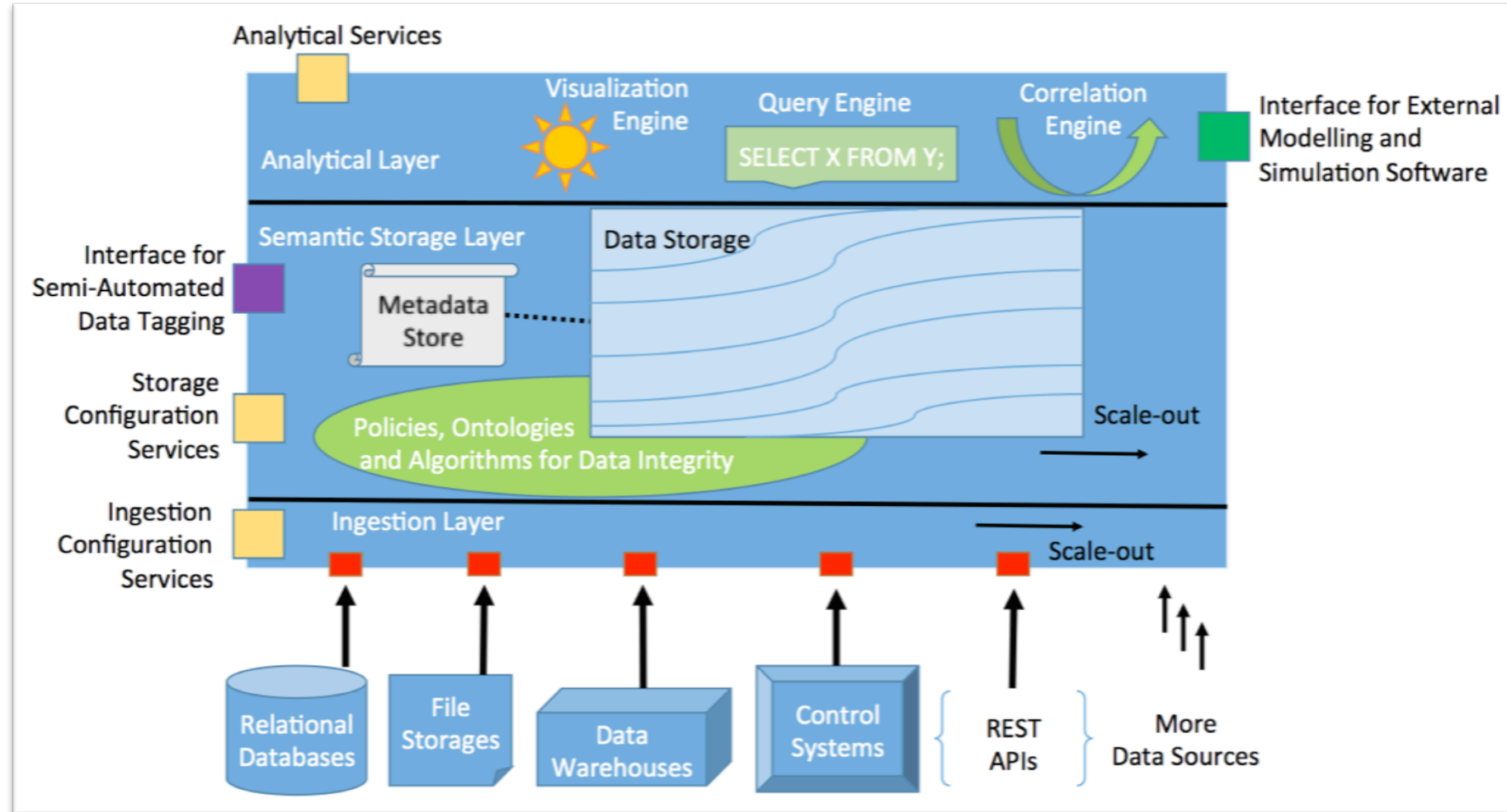
Unstructured data of low quality

# Reliability and Availability Data Analytics Platform

- Data lake approach
- Integration of heterogeneous structured, semi-structured and unstructured data
- Semi-automated data quality annotation



# Reliability and Availability Data Analytics Platform: Proposed Architecture

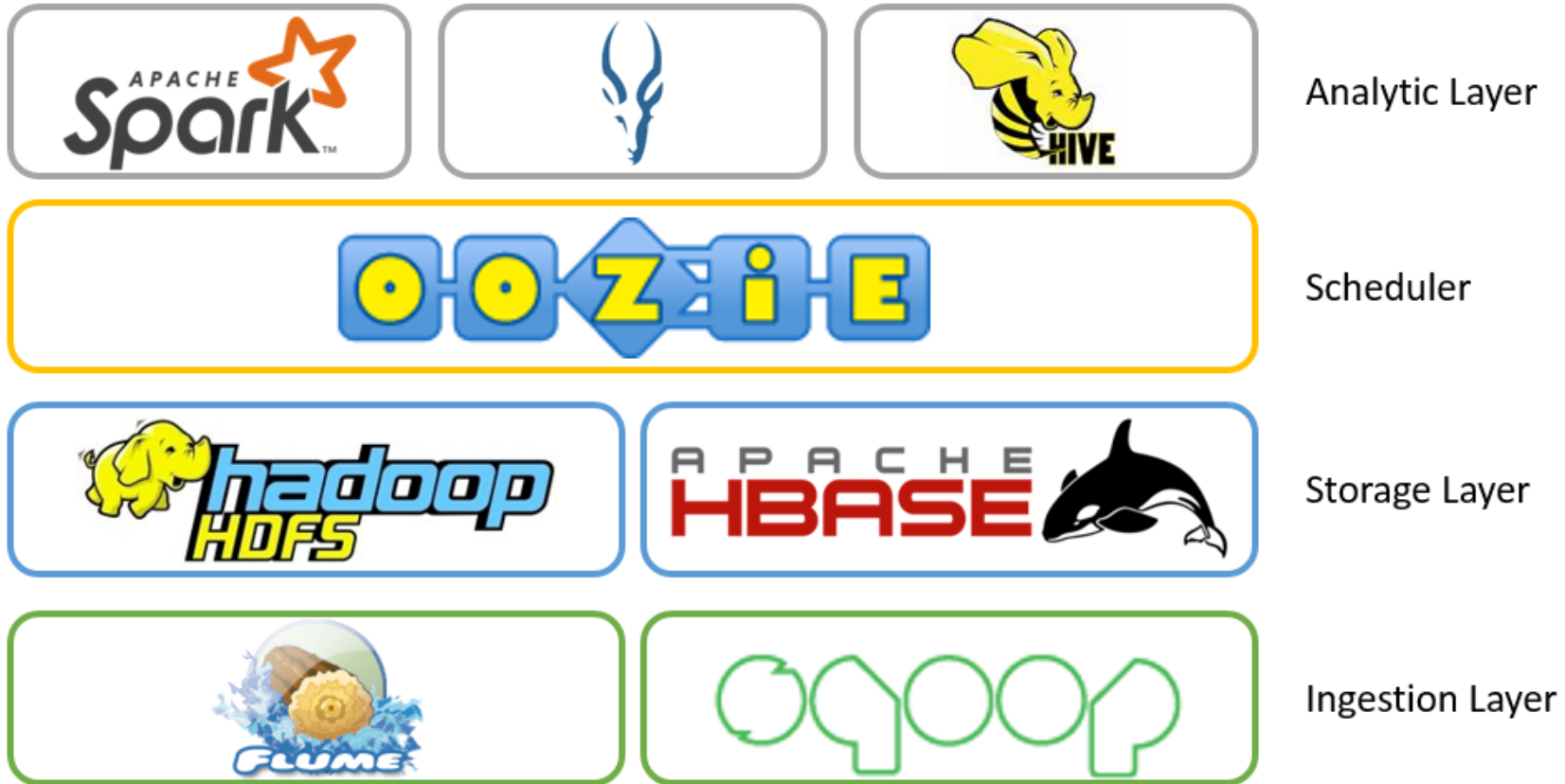


# Big Data Infrastructure

## Deployed Hadoop Cluster

- ✓ Shared-nothing architecture [3]
- ✓ HDFS enables flexible storage layer
- ✓ Large developer community

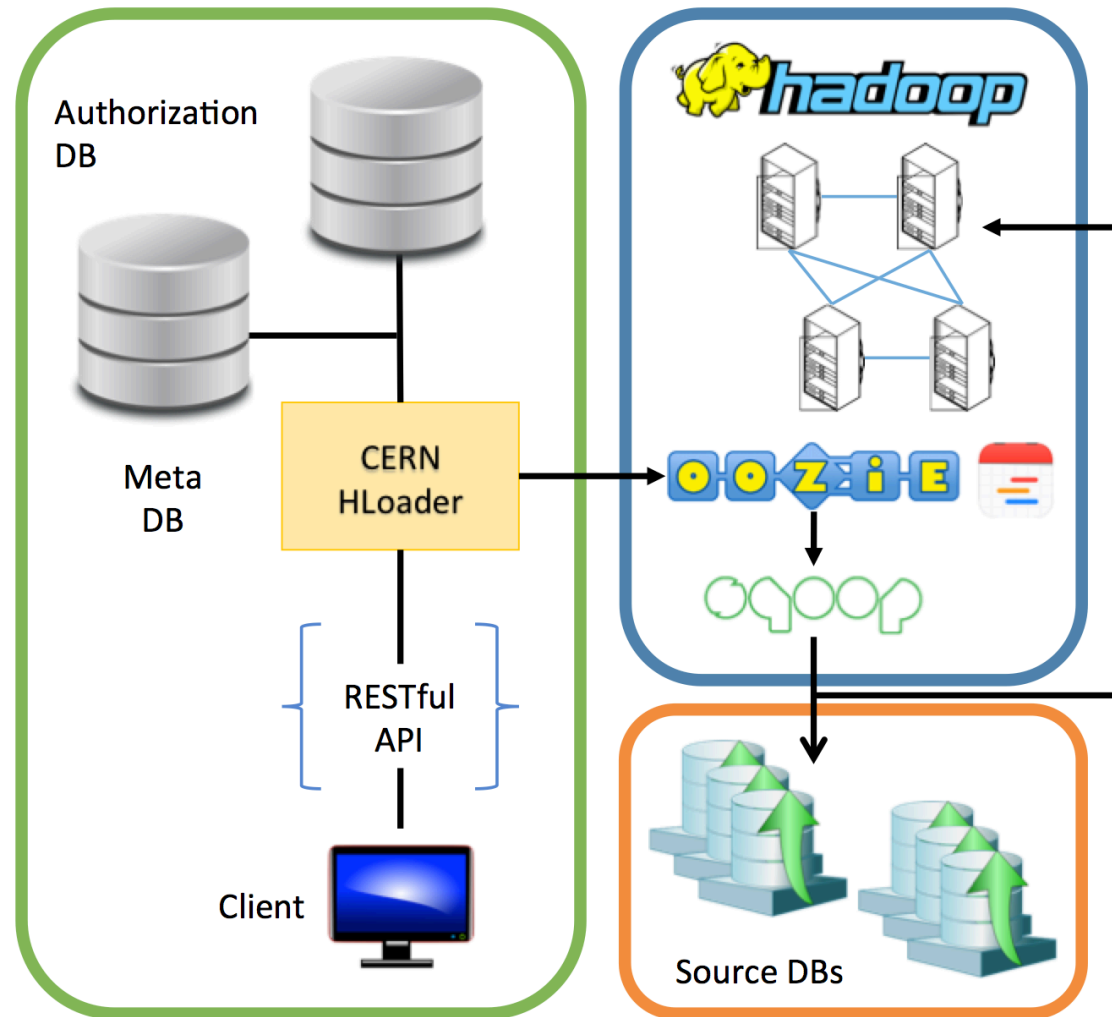
# Big Data Infrastructure: Software Frameworks Stack



# CERN HLoader

- Built around Apache Sqoop & Apache Oozie
- Provides a REST API, exposing metadata on available source / target storage systems to authorized users
- Allows to submit one-off or re-occurring Sqoop jobs using Oozie workflow / coordinator apps

# CERN HLoader: Architecture



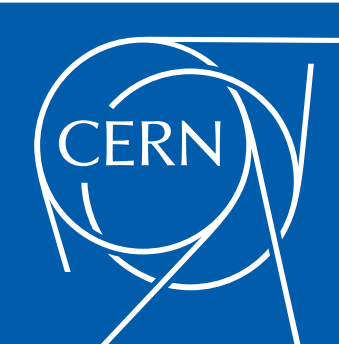
# Selected Use-Case

## Super Proton Synchrotron Beam Quality

- Direct injector to LHC
- Specify QoS for SPS beams
- Simulate injection scenarios and their impacts on LHC beam availability
- SPS BQM, OP eLogbook data

# Conclusions & Future Work

- **Data lake** architecture meets the data management requirements of the FCC RAMS study
- CERN IT Hadoop service provides suitable infrastructure for the realization of the data lake
- Continued incremental development of the Reliability and Availability Data Analytics Platform
- Work on specified use cases (French electricity grid, power converters, ...)





# Big Data Infrastructure: Hardware Specifications

- 14 nodes
- 64 GB of memory per node (total of 896 GB)
- 32 CPU cores (Ivy Bridge-EP 2,6 GHz) per node (total of 448 CPU cores)
- 48 x 4 TB disk space per node (total of 2,69 PB)