Smooth Migration of CERN Post Mortem Service to a Horizontally Scalable Service

C. Aguilera-Padilla, S. Boychenko, M. Dragu, M.A. Galilee, J.C. Garnier, M. Koza, K. Krol, R. Orlandi, M.C. Poeschl, T.M. Ribeiro, M. Zerlauth,
CERN, Geneva, Switzerland

ICALEPCS 2015
International Conference on Accelerator & Large Experimental Physics Control Systems
17-23 OCTOBER 2015 MCEC MELBOURNE

Use Case Requirements

Current event-driven PM system use cases:

- LHC Powering Event
- LHC Global Event
- LHC Injection Quality Checks
- LHC Extraction Post Operational Checks

PM system use cases under development:

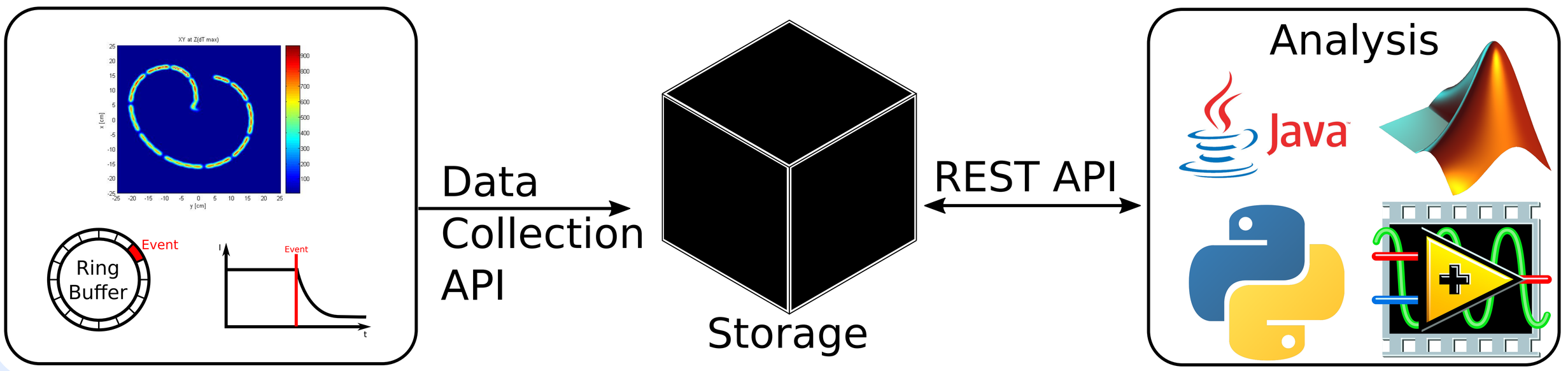
- SPS Quality Checks
- Linac4 Quality Checks

Motivation

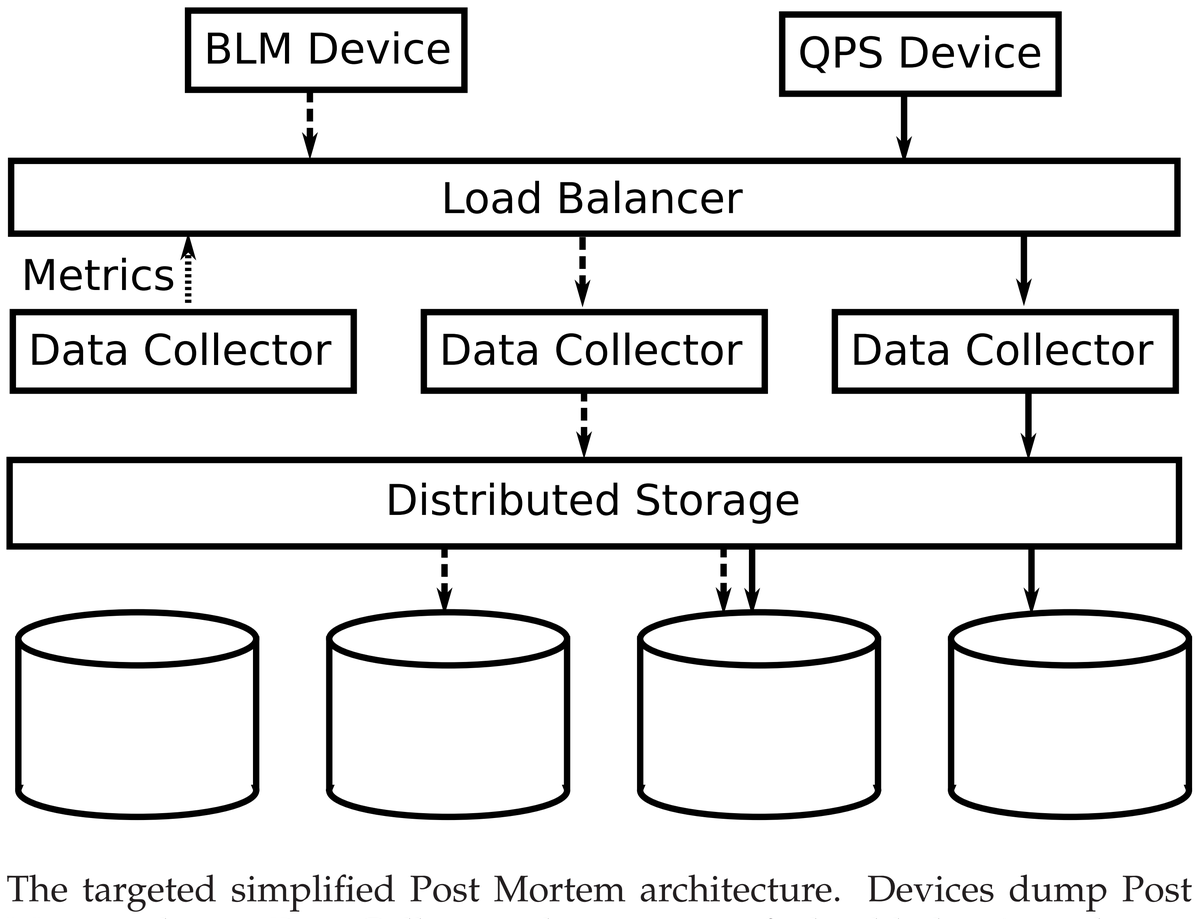
We have identified the following shortcomings in the current PM system scalability:

- Static load distribution
- Lack of data consistency and integrity verification
- Storage architecture and data format exposed to the end-users
- Data duplication
- File system performance limits and back pressure
- Monolithic implementation of data collectors
- Neither delivery guarantee nor ordering guarantee of data dumps

Decouple Data Usage from Data Storage



Future Data Collection and Storage



The targeted simplified Post Mortem architecture. Devices dump Post Mortem data to Data Collectors that are part of a load-balancing scheme. This way the load can be distributed among the Data Collectors of the Post Mortem cluster. The Data Collector that processes the dump writes the data to the distributed storage. The data are automatically replicated by the storage infrastructure.

Future PM System Properties

✓ Horizontal scalability of data collection thanks to load-balancing. Any device will dump transparently to the Post Mortem system without any configuration effort.

✓ Horizontal scalability of storage. Adding storage capacity consists of adding more nodes.

✓ Horizontal scalability of nodes provides computing resources for the API.

✓ Interventions on the PM system will not result in downtime of the service but will be transparent. In the worst case a performance drop can be observed.

✗ Horizontal clusters may bring split brain issues in which parts of the cluster cannot communicate anymore and diverge. Modern technologies implement proper quorum to prevent this from happening: the smallest cluster considers itself inoperable, and the largest cluster continues the operations.

✗ A good load balancing implementation may require specific networking and system administration configuration.

✗ The maintenance of the cluster will rely on multiple expertise: software engineers, system administrators, network engineers, storage engineers.

Steps

1: Provide APIs to prevent direct Storage Access
- Language Agnostic API to access data
- Data dumps from both CMW RDA2 and RDA3 protocols

2: Commission the new data storage and serialization
- Keep the current storage running
- Automatic or semi-automatic data forwarding to new storage
- Run spare analysis tools on new storage

3: Decommission old PM system
- Transfer all data
- Rewire production APIs to new storage
- Switch off old storage