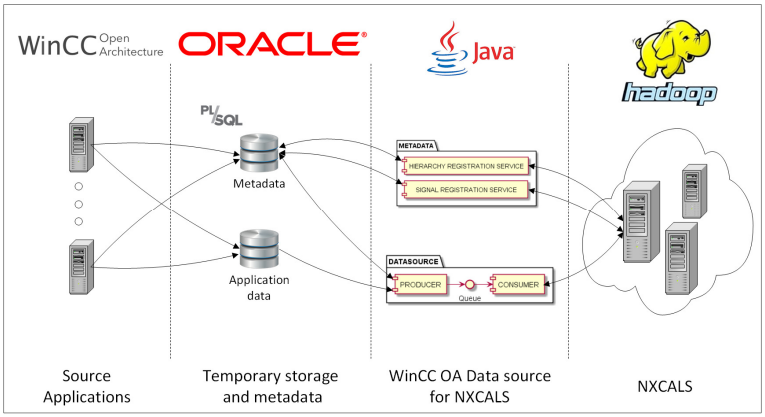
Big Data Archiving From Oracle to Hadoop

I. Prieto Barreiro, M. W. Sobieszek. CERN, Geneva, Switzerland

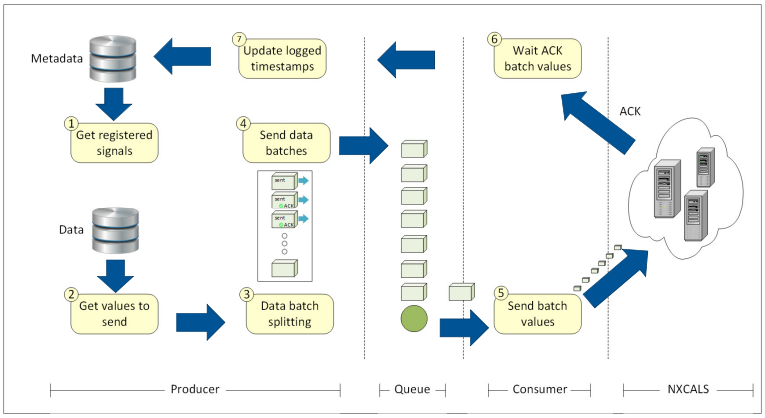
MOPHA117

Context



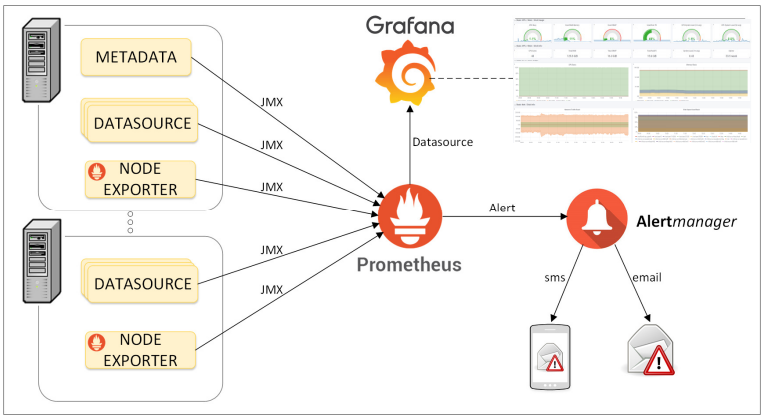
- The Next CERN Accelerator Logging Service (NXCALS), based on Hadoop, will replace the old Oracle based logging service (CALS)
- Many applications using database to database transmission in PL/SQL in the old logging service will be affected
- Over 250 WinCC OA applications registered 1.6M signals in the logging service and generate around 175k values per second
- A new distributed service was implemented to feed NXCALS with the application's data. The service was designed to be extremely robust, scalable and fail-safe with guaranteed data delivery and no data loss

Data Processing



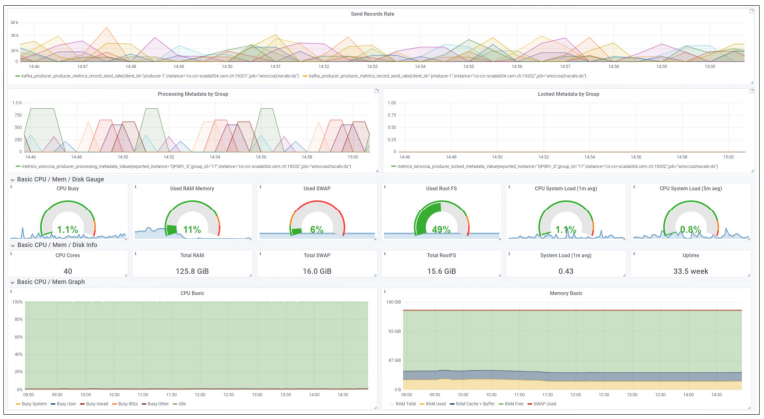
1. Get the list of signals registered in the logging service
2. Get the signal values not yet transmitted to NXCALS
3. Split the bulk signal values into multiple data batches
4. Send the data batches to the communication channel
5. Send the individual batch values to the logging service
6. Wait for the reception acknowledge of all the batch values
7. Update the timestamps of the last logged values

Monitoring and Alerting



- The status of all processes and hosts is constantly monitored by the Prometheus ecosystem
- The Metadata and Datasource processes expose via JMX relevant metrics for monitoring their current state
- The host metrics are exposed by Prometheus' node exporter
- Prometheus compares the metric values against a set of customized alerts and sends notifications to the AlertManager
- The AlertManager groups the alerts by type and sends notifications by email or sms to the relevant receivers, depending on the type of alarm
- Grafana is used to display custom charts and dashboards with the relevant metrics

Grafana Dashboards



- Custom charts and dashboards to monitor the health state of the service and the hosts where it is deployed
- Datasource process: data transfer rates, number of data groups being processed, number of data groups locked due to an exception, number of read & publication errors, etc.
- Metadata process: number of signals registered in CALS and not in NXCALS (i.e. inconsistency between logging systems)
- Host metrics: CPU usage, percentage of used RAM and SWAP, network traffic, disk space used, etc.

CERN Beams Department
Industrial Controls and Safety Systems Group (ICS)
e-mail: iprietob@cern.ch

ICALEPCS 2019