CERN IT Monitoring and Data Analytics

Pedro Andrade (IT-GT)

Openlab Workshop on Data Analytics

16th November 2012

- Monitoring in IT covers a wide range of resources
  - Hardware, OS, applications, files, jobs, etc.
  - Many high level resources are interdependent

- Several application-specific monitoring solutions
  - Similar needs and architecture
    - Publish metric results, aggregate results, alarms, etc.
  - Different technologies and tool-chains
    - Some based on commercial solutions
  - Similar limitations and problems
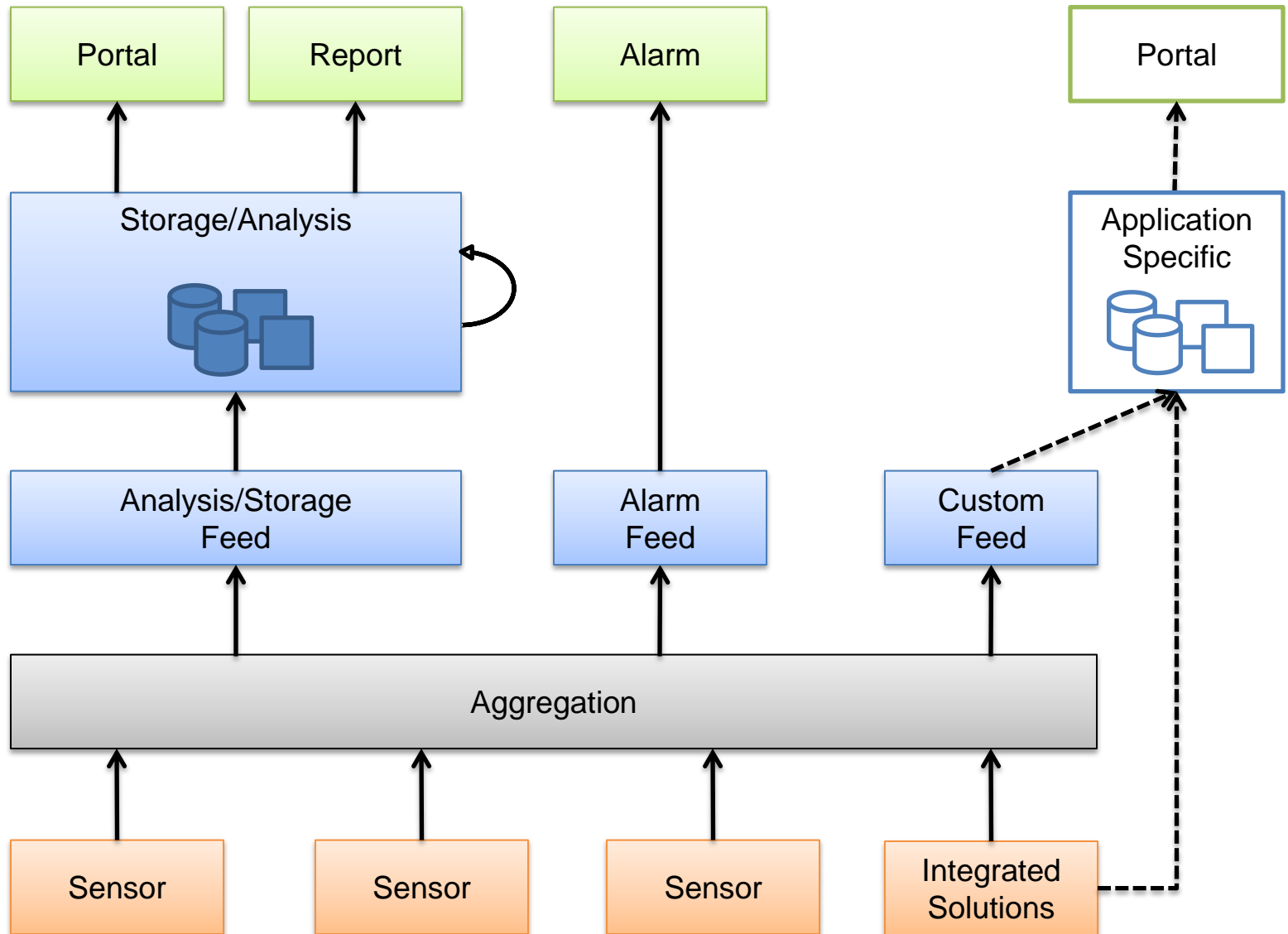   - Limited sharing of monitoring data

- Several improvements and challenges
  - Combine and correlate monitoring data
  - Better understand infrastructure/services performance
  - Move to a virtualized and dynamic infrastructure
  - Optimize effort allocated to monitoring tasks

- Need to implement a common monitoring strategy
  - Following a tool-chain approach
  - Adopt existing tools, avoid home grown solutions
  - Aggregate and correlate all monitoring data
  - Make monitoring data easy to access

Monitoring Strategy

- Agile Infrastructure (AI) project

- Architecture Design and Support
  - Design, technology watch, testing
  - Support to individual monitoring teams
- Data Processing and Visualization
  - Data analysis and correlation
  - Alarms, notifications, dashboards
- Core Services and Tools
  - Common services operation
  - Common tools/libraries development

Core Monitoring
(Lemon, SAM)

Service Monitoring
(Castor, Security, etc.)

Monitoring Strategy

- ## Data Storage and Analysis
  - – Store all monitoring data in a common location
    - • Easy sharing of monitoring data and analysis tools
  - – Feed the system with processed data
  - – Use one single common data format (JSON)
  - – Permanent storage for historical data

- ## Data Visualization and Alarms
  - – Provide powerful easy-to-use portals and dashboards
    - • Over raw data and processed data
  - – Provide an efficient delivery of notifications
    - • Notifications directly sent to correct consumer targets

Data Storage and Analysis

- High-level use cases collected from all IT groups
  - Classified under three categories

- Fast and Furious (FF)
  - Real time queries on particular attributes
- Digging Deep (DD)
  - Complex analysis jobs over entire data sets
- Correlate and Combine (CC)
  - Correlation of data across different data sets

Fast and Furious

- ## Find, filter, join, alarm
  - Hardware, nodes, exceptions, web servers, router, network, connections, urls, jobs, data transfers, site and services, status, users, batch jobs, etc.

- ## Examples
  - Get metrics values for hardware and selected services
  - Filter metrics per different types (role, cluster, etc)
  - Aggregate exceptions and errors
 - Raise alarms according to appropriate thresholds

Digging Deep

- Reorder, statistics, reporting
  - Historical data, traffic, configurations, service status and availability, CPU usage, disk usage, etc.

- Examples
  - Curation of hardware and network historical data
  - Analysis and statistics on batch job and network data

Data Storage and Analysis

CERN IT Department

Correlate and Combine

- ## Correlate
  - Raw data, alarms, metrics, hardware, users, services, traffic, load, problems, connections, IPs, site status, job metadata, data transfer metadata, etc.

- ## Examples
  - Correlation between alarms, hardware, and services
  - Correlation between usage, hardware, and services
  - Correlation between job status and grid status

Storage & Analysis
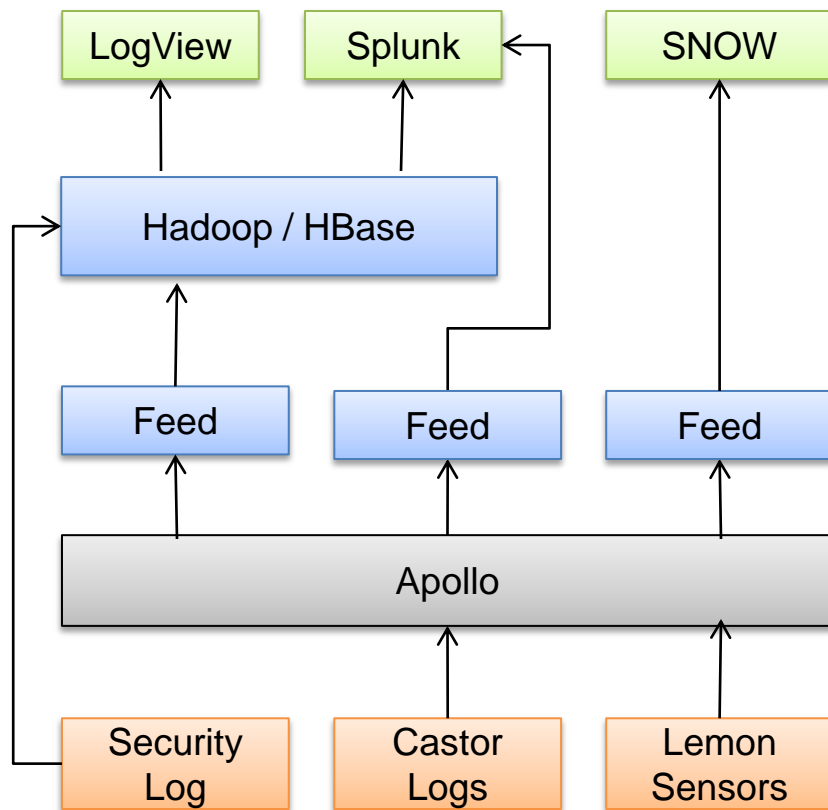
- **Small Hadoop cluster running latest CDH4**
  - Security net log data added to HDFS/HBase
  - Castor log data added to HDFS/HBase, queried by LogViewer interface
  - Lemon data will be imported via messaging
  - MR for ad-hoc queries under development

Visualization & Alarms
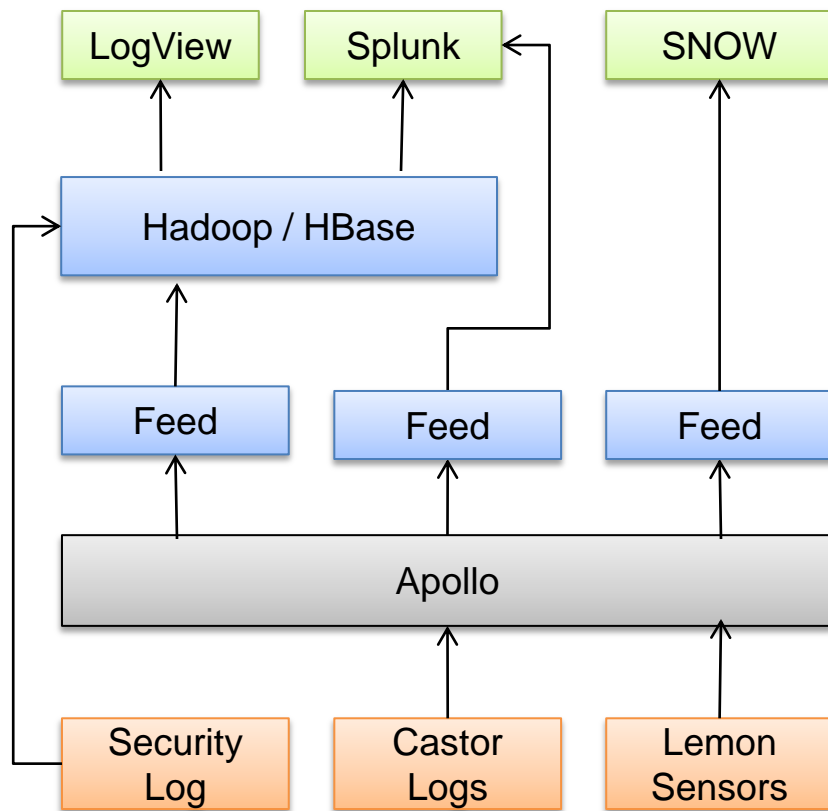
- Splunk dashboards
  - Different dashboards tested with lemon monitoring data
- LogViewer interface
  - Operations tool based on data in HBase
- SNOW integration
  - Lemon notifications delivered as SNOW tickets via messaging

- Consolidate monitoring of new Agile Infrastructure nodes and services and move to production

- Increase the exploitation of the central data storage with more complex data analysis and data correlation activities

- Integrate more monitoring applications and data sets under the common monitoring architecture

- WLCG monitoring architecture similarities
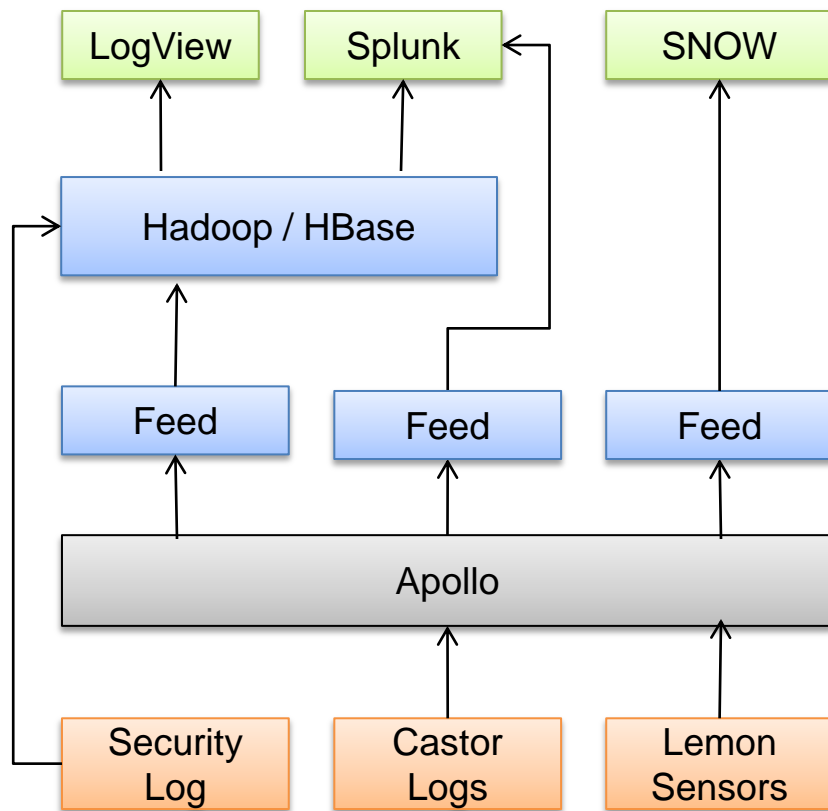
AI Monitoring   *WLCG Monitoring*

Conclusions
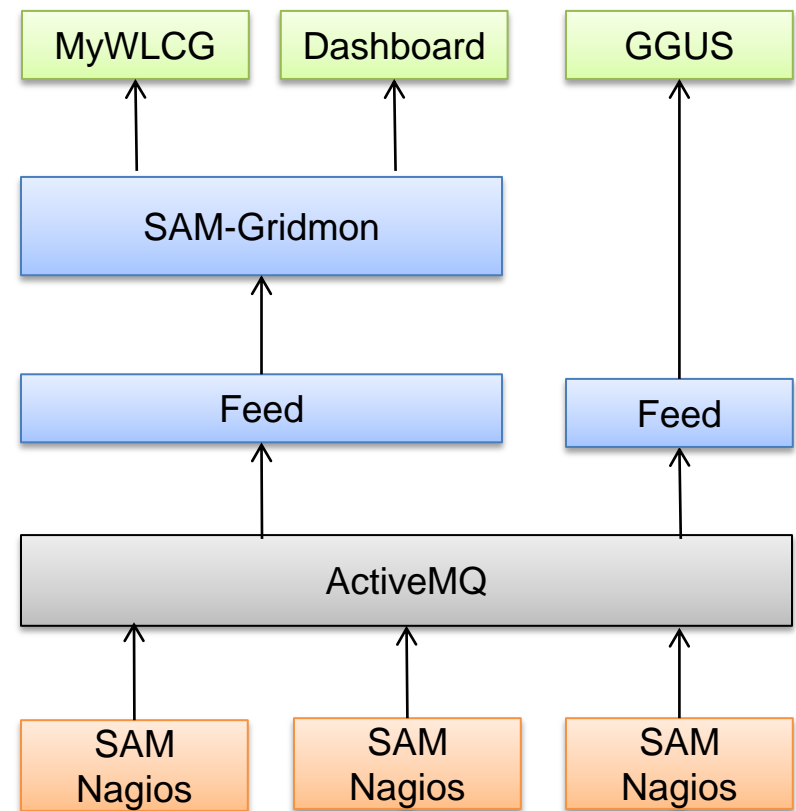
- Monitoring architecture defined and approved
  - Based on well defined monitoring layers
  - Initial set of technologies identified and deployed

- Implementation plan ongoing
  - Moving towards a common system
  - Core components of the architecture in place
  - Enables new Agile Infrastructure nodes to be monitored
  - Assures smooth transition for today's applications

- Simple data analytics tested… but a lot to do !

CERN**IT** Department