Evaluating Elasticsearch-Kibana-Grafana for CMS storage space monitoring

Natalia Ratnikova
CERN IT Monitoring meeting
19 August 2016
Goal

• Provide a central view of space occupied by all CMS data stored in heterogeneous storage systems deployed at CMS sites.
CMS Space Monitoring

- Storage dumps produced locally by the sites include user areas and CMS production data not accounted in the CMS central data catalogues.
- Central infrastructure has been deployed as CMS web service at CERN with Oracle DB backend and data service returning space usage records in xml, perl, or json format.
- By now collected more than a year wealth of monitoring data, asynchronously pushed by the sites, generally at weekly intervals.
- Now looking for a visualization solution.
Visualization options: initial thoughts

Original design proposed by Eddie Karavakis, similar to ATLAS DDM storage accounting based on Rucio storage summaries:
Visualization options: current status

• Growing popularity of Kibana in CMS:
  – CMS:CRAB
    https://meter.cern.ch/public/_plugin/kibana/#/dashboard/temp/CMS::CRAB
  – CMS Castor & EOS
    https://meter.cern.ch/public/_plugin/kibana/#/dashboard/temp/castormeter::globalrates
  – Local deployments at CMS sites (Brian)
• Expertise with Grafana based monitoring at Fermilab and CERN (FIFE, EOS)
• Consolidation of CERN IT monitoring:
  http://itmon.web.cern.ch/itmon/index.html
• Both Kibana and Grafana now work with Elasticsearch

=> we are evaluating Elasticsearch + Kibana and Grafana
Evaluating Kibana vs Grafana (1):

• General comparison:
  – Feature set
  – Users base
  – User support model
  – Available documentation
  – Extensibility
  – Server site requirements:
    • dependencies
    • OS support
    • Hardware configuration
  – Client-side support (browsers, APIs)
  – Scalability and performance considerations
Evaluating Kibana vs Grafana (2):

• Suitability for CMS Space Mon use cases/requirements:
  – Ability to show space usage information for multiple sites
  – Ability to select the groups of sites
  – Ability to represent the storage namespace (a tree-like structure)
  – Ability to select time range
  – Ability to group space usage by different data types
  – Historical plots for asynchronously provided space usage records
  – Effort required for configuration changes:
    • new/decommissioned sites
    • new data types.
  – Access by the collaborators around the world
  – Ease of navigation (usability)
  – Performance: response time, availability, reliability
Data organization in the Elasticsearch

• Is filtering and aggregation provided in Kibana (Grafana) sufficient
  – or some external analytics could be better, e.g. on the oracle data service level or some additional layer

• Currently we record:
  – Timestamp, DirPath, DirSize, SiteName

More fields may be added, such number of files, available space. The contents and format are discussed in collaboration with ATLAS and LHCb within WLCG Storage Accounting project

minutes of last meeting: https://indico.cern.ch/event/561433/
Comments ...

Suggestions ...