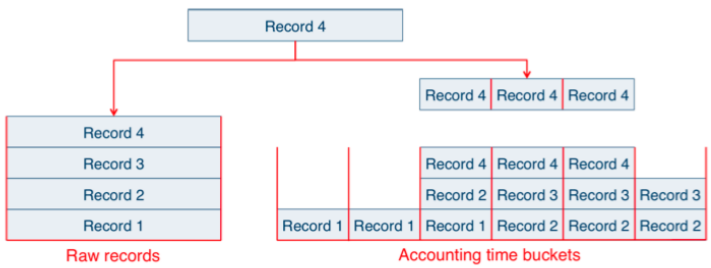
Accounting and Monitoring

Zoltan Mathe

Tuesday 22nd May, 2018

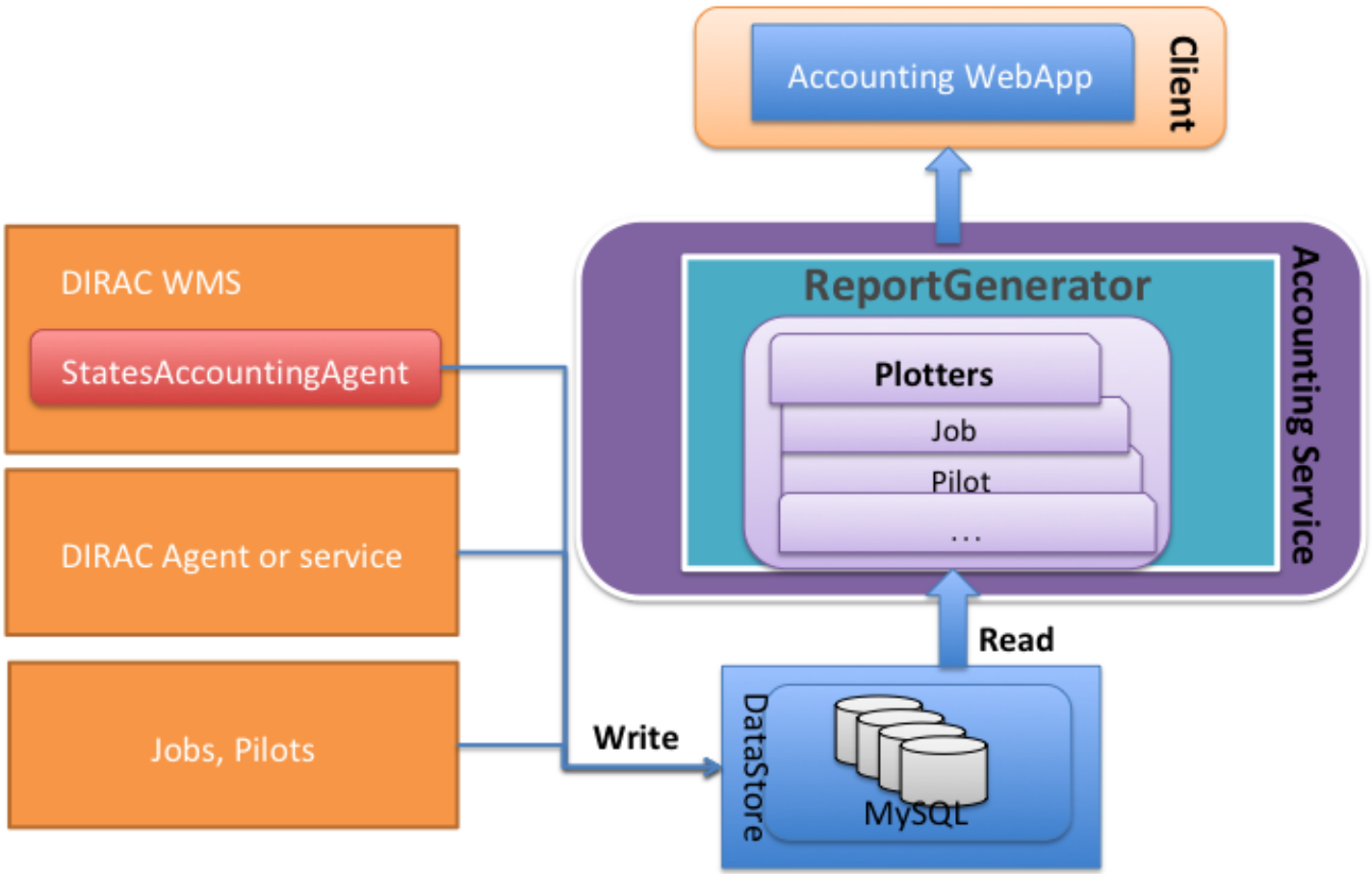
- it is based on the DIRAC framework and designed for:
  - for collecting, storing and visualize time-binned averages and aggregates
  - records are stored in the AccountingDB, in "two" different formats:raw records, time buckets [1]



  - efficient for storing historical data, for real time monitoring is not recommended
  - the data can be accessible through the DIRAC web framework using the Accounting web application.

[1] figure taken: The LHCb Experience on the Grid from the DIRAC Accounting Data

- Pilot: for pilots jobs historical report for a given site, CE, etc.
- Job: for jobs historical report for a given site, etc.
- Data operation: data transfers, replication, etc.
- WMS History: for monitoring the DIRAC Workload Management System. It is "deprecated", replaced by WMS monitoring
- Network: for monitoring of network activities on the network layer


Client
Accounting WebApp
DIRAC WMS
StatesAccountingAgent
DIRAC Agent or service
Jobs, Pilots
ReportGenerator
Plotters
Job
Pilot
...
Accounting Service
Read
Write
DataStore
MySQL

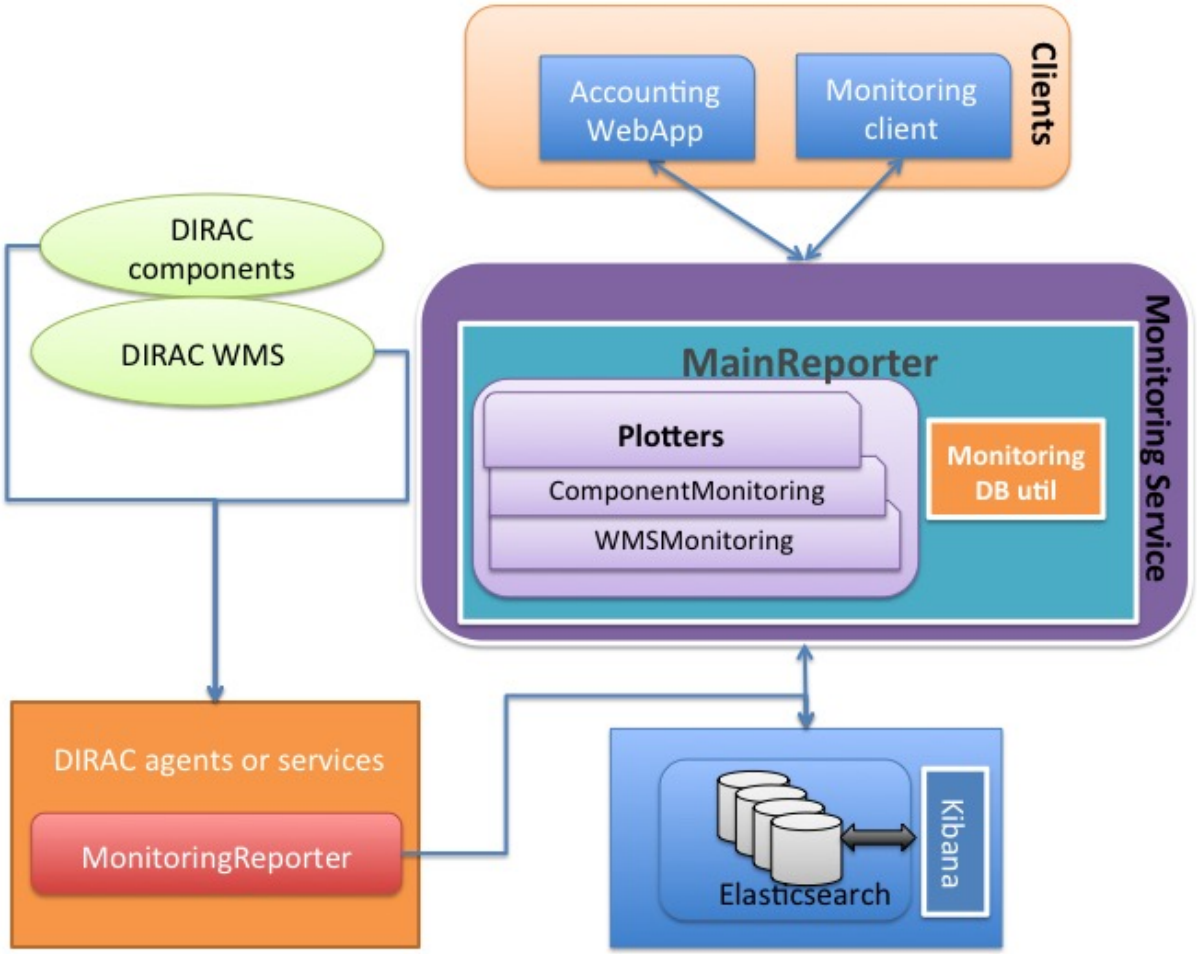
- using dirac-admin-sysadmin-cli -H hostname, more details in [2]
  - install db AccountingDB;
  - install service Accounting DataStore;
  - install service Accounting ReportGenerator;
- for WMS history: install agent WorkloadManagement StateAccountingAgent
- DIRAC web framework[3] for visualize the data
- helper (DataStoreHelper) services can be installed (more details in the administrative guide) in case of high load

[2] `http://dirac.readthedocs.io/en/integration/AdministratorGuide/Systems/Accounting/index.html`

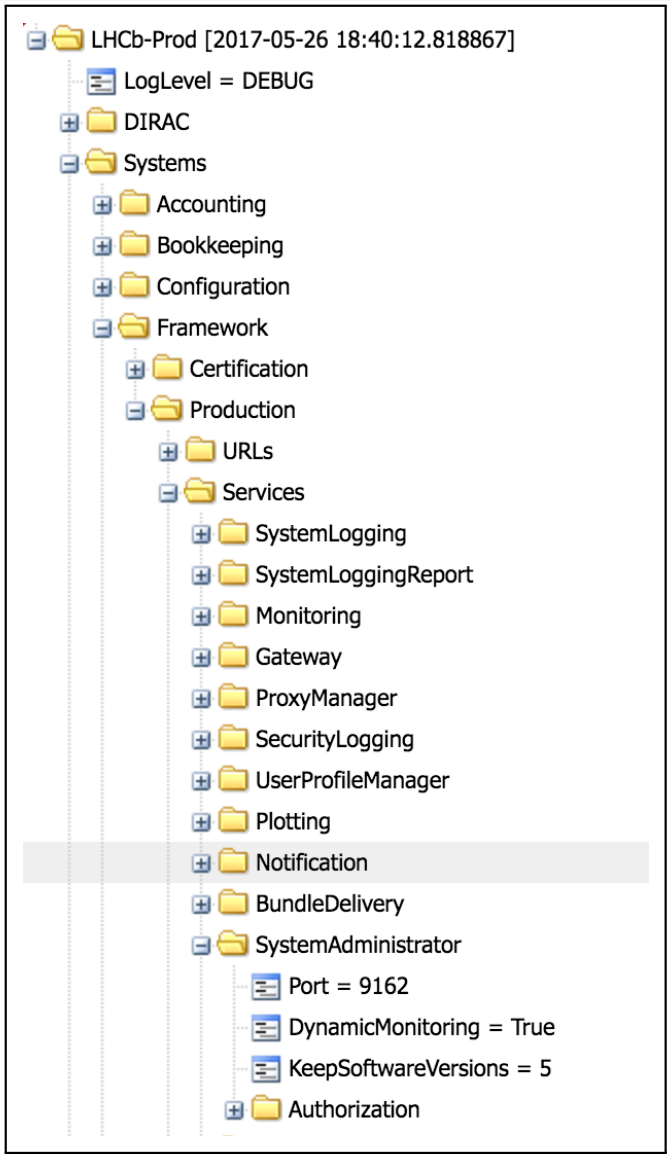
[3] Web `http://dirac.readthedocs.io/en/integration/AdministratorGuide/InstallingWebAppDIRAC/index.html`

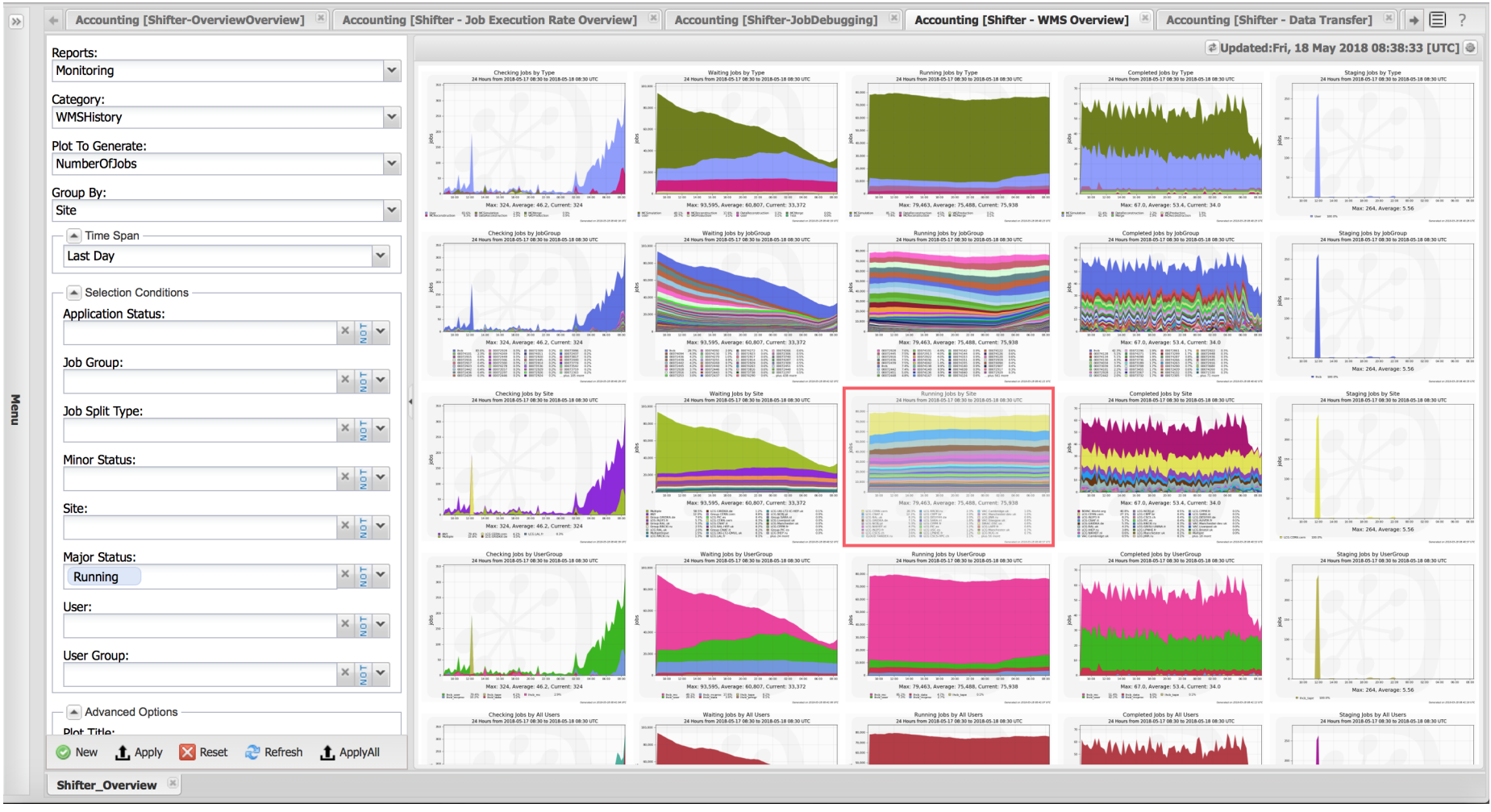
Monitoring System

- It is based on DIRAC framework and designed for:
  - real time monitoring (WMS jobs, DIRAC components, etc.)
  - managing semi-structured data (JSON)
  - efficient data storage, data analysis and retrieval
  - provide high quality reports


Clients
Accounting WebApp
Monitoring client
DIRAC components
DIRAC WMS
Monitoring Service
MainReporter
Plotters
ComponentMonitoring
WMSMonitoring
Monitoring DB util
DIRAC agents or services
MonitoringReporter
Elasticsearch
Kibana

- Requirements:
  - Elasticsearch
  - RabbitMQ (optional)
  - WMS Monitoring:
    - StatesMonitoringAgent (WMS)
  - Component Monitoring:
    - DynamicMonitoring = True (SystemAdministrator)
- Accounting web application (WebAppDIRAC)




Accounting [Shifter-OverviewOverview]
Accounting [Shifter - Job Execution Rate Overview]
Accounting [Shifter-JobDebugging]
Accounting [Shifter - WMS Overview]
Accounting [Shifter - Data Transfer]
Updated:Fri, 18 May 2018 08:38:33 [UTC]
Reports:
Monitoring
Category:
WMSHistory
Plot To Generate:
NumberOfJobs
Group By:
Site
Time Span
Last Day
Selection Conditions
Application Status:
Job Group:
Job Split Type:
Minor Status:
Site:
Major Status:
Running
User:
User Group:
Advanced Options
Plot Title:
New
Apply
Reset
Refresh
ApplyAll
Menu
Shifter_Overview


Accounting [Untitled 3]
Reports:
Monitoring
Category:
Component Monitoring
Plot To Generate:
Number of running threads
Group By:
host
Time Span
Last Week
Selection Conditions
Component:
DataManagement_FileCatalog
Host:
PID:
Status:
Advanced Options
Plot Title:
Pin Dates
Do not scale units
New
Apply
Reset
Refresh
ApplyAll
Menu
Updated:Fri, 18 May 2018 08:46:28 [UTC]
Memory usage by host
7 Days from 2018-05-11 to 2018-05-18
MB
Max: 2,873, Min: 870, Average: 2,039, Current: 1,309
Number of running threads by host
7 Days from 2018-05-11 to 2018-05-18
threads
Max: 26.6, Min: 21.5, Average: 23.9, Current: 25.9
Memory usage by host
7 Days from 2018-05-11 to 2018-05-18
MB
Number of running threads by host
7 Days from 2018-05-11 to 2018-05-18
threads
Shifter_Overview
Default

- using dirac-admin-sysadmin-cli -H hostname, more details in [4]
- install db MonitoringDB;
- install service Monitoring Monitoring;
- Elasticsearch follow in `https://www.elastic.co`
- RabbitMQ for failover, follow in `https://www.rabbitmq.com`

[4] `http://dirac.readthedocs.io/en/integration/AdministratorGuide/Systems/MonitoringSystem/index.html`

- Since October 2016 LHCb is using the Monitoring System
- Elasticsearch clusters are provided by CERN (one for production, one for certification)
- Size: 380 GB, 586 indexes, $4 * 10^9$ documents
- 8 nodes cluster: 3 master, 2 search and 3 data nodes (more details in the 'Monitoring performance of a highly distributed and complex computing infrastructure in LHCb' 2016 CHEP paper)

- Accounting and Monitoring:
  - DIRAC Graph library, which is based on Matplotlib and used to create high quality plots
  - DIRAC web framework based on Tornado (see DIRAC Web 4.0 presentation)
- Monitoring:
  - Messaging Queue system, we are using RabbitMQ for high availability
  - Elasticsearch distributed search and analytic engine (see The rise of ElasticSearch presentation)
- Accounting
  - MySQL (AccountingDB)

- DIRAC provides:
  - two different systems for accounting and monitoring
  - a graph library for creating the reports
  - Accounting web application for visualising the data

Thank you! Questions?