Jobs Monitoring and Troubleshooting with Flume/Elasticsearch/Kibana

Qiulan Huang/OPOS Group

11-18-2015

Outline

- Introduction
- Design and Implementation
- Running the monitoring
- Plans and Future work

Overview

- Monitoring with Flume/Elasticsearch/Kibana
  - Monitor and troubleshooting jobs running at Fermilab Grid clusters and OSG Facility
  - It's a common tool to all experiments in Fermilab
  - Visualized by Kibana/Grafana
- Where is the data from?
  - Users' condor event logs
- What information include?
  - ClusterId, JobId, Job start time, Job end time, Job exit code, Job site, disk usage, memory usage etc.

Why and What?

- Condor logs contain the life history of the jobs
  - When and Where job submits, runs, checkpoints, finishes and exit conditions
  - Details for troubleshooting
  - Accounting information including CPUtime, Systime, CPU cores, Disk and Memory
- What can this tool do?
  - Debug jobs and diagnose issues
  - Monitoring
  - Accounting

Design and Implementation

- Data fields in Elasticsearch

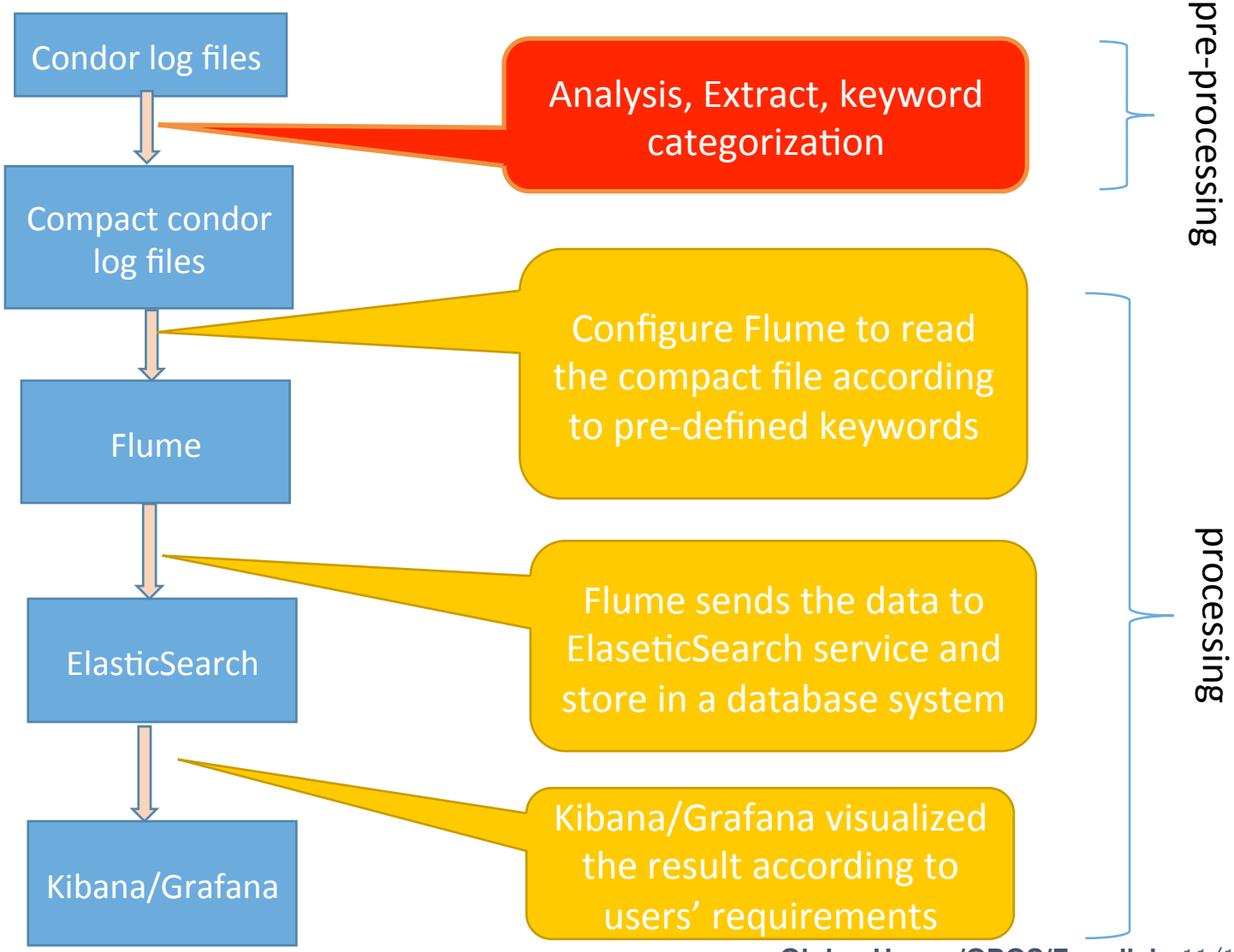
t @fields.job_memory_usage	
t @fields.cluster_id	
t @fields.experiment	
t @fields.job_cpu_allocated	
t @fields.job_disk_uasge	
◷ @fields.job_end_time	
t @fields.job_id	
t @fields.job_return_value	
t @fields.job_site	
◷ @fields.job_start_time	
t @fields.job_status	
t @fields.job_sys_time	
t @fields.job_usr_time	
t @fields.job_workslot	

t @fields.cluster_id	3251170
t @fields.experiment	mu2e
t @fields.job_cpu_allocated	1
t @fields.job_disk_uasge	16734
◷ @fields.job_end_time	October 22nd 2015, 07:05:31.000
t @fields.job_id	3251170.4603
t @fields.job_memory_usage	955
t @fields.job_return_value	0
t @fields.job_site	Omaha
◷ @fields.job_start_time	October 22nd 2015, 05:00:27.000
t @fields.job_status	terminated
t @fields.job_sys_time	38
t @fields.job_usr_time	7412
t @fields.job_workslot	None

Data schema

Implementation

- I used Flume+Elasticsearch+Kibana/Grafana



Some Details

- Deployed Flume+Elasticsearch+Kibana platform
- Updated the data in Elasticsearch twice per day automatically
  - In order not to duplicate download the condor logs of finished clusters, I define the different status to mark them
  - N---New, R---Run, VF--- Just Finished, F--- Finished
  - Status change: N→R, R→VF,VF→F
- Mu2epro and novapro jobs information are available now
- Easy to extend to other experiments
- Defined and generated plots and dashboards to get metrics like:
  - Number of finished/disconnect/evicted jobs per facility over time
  - Job pre-emption rate per facility ordered by time
  - Categorize the types of job's exit code by site
  - Monitoring job CPU time, Systime, memory usage and disk usage by site
  - ……

How to use this tool?

- Access to the URL
  - http://osgmonitor.fnal.gov:5601
- Troubleshooting by Query and search
  - For example: How many disconnect jobs of MU2E in Fermigrid?
  - Query in Kibana with the condition: Fermigrid AND mu2e AND disconnect

Top 0 @fields.experiment	Top 0 @fields.job_site	Top 0 @fields.cluster_id	Count
mu2e	Fermigrid	3369590	21

```
@fields.job_status: disconnect @fields.experiment: mu2e @message: Experiment=mu2e ClusterId=3369590 JobId=3369590.1228 StartTime=2015-11
-10 10:18:14 EndTime=2015-11-11 16:58:10 Site= Fermigrid Host= slot32@fnpc7017.fnal.gov Status=disconnect ReturnValue=None SystemTime=None Us
erTime=None DiskUsage=None MemoryUsage=3847 CPUAllocated=None @fields.job_sys_time: None @fields.job_disk_uasge: None @fields.job_end_tim
e: November 11th 2015, 10:58:10.000 @fields.job_cpu_allocated: None @fields.job_start_time: November 10th 2015, 04:18:14.000 @fields.job_
workslot: slot32@fnpc7017.fnal.gov @fields.job_usr_time: None @fields.cluster_id: 3369590 @fields.job_site: Fermigrid @fields.job_return
```

- Create visualizations with various charts
- Create dashboards
- Coding with calling Elasticsearch API to introduce a custom metric in Kibana/Grafana to get data and do more complicated analysis

Dashboard

- Number of finished/disconnect/evicted jobs per site over time
  - MU2E

    http://osgmonitor.fnal.gov:5601/#/dashboard/MU2E-Number-of-terminated-slash-evicted-slash-disconnect-jobs-per-facility-over-time?_g=()&_a=(filters:!(),panels:!((col:1,id:mu2e-number-of-terminated-jobs-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization),(col:7,id:mu2e-number-of-evicted-jobs-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization),(col:1,id:mu2e-number-of-diconnect-jobs-per-facility-over-time,row:4,size_x:6,size_y:3,type:visualization),(col:7,id:mu2e-number-of-evicted-and-disconnect-jobs-per-facility-over-time,row:4,size_x:6,size_y:3,type:visualization)),query:(query_string:(analyze_wildcard:!t,query:mu2e)),title:'MU2E%20Number%20of%20finished%2Fevicted%2Fdisconnect%20jobs%20per%20facility%20over%20time')
  - NOvA

    http://osgmonitor.fnal.gov:5601/#/dashboard/NOvA-Number-of-terminated-slash-evicted-slash-disconnect-jobs-per-facility-over-time?_g=()&_a=(filters:!(),panels:!((col:1,id:nova-number-of-terminated-jobs-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization),(col:7,id:nova-number-of-evicted-jobs-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization),(col:1,id:nova-number-of-disconnect-jobs-per-facility-over-time,row:4,size_x:6,size_y:3,type:visualization),(col:7,id:nova-number-of-evicted-and-disconnect-jobs-per-facility-over-time,row:4,size_x:6,size_y:3,type:visualization)),query:(query_string:(analyze_wildcard:!t,query:'*')),title:'NOvA%20Number%20of%20finished%2Fevicted%2Fdisconnect%20jobs%20per%20facility%20over%20time')
- Pre-emption rate per facility over time
  - MU2E

    http://osgmonitor.fnal.gov:5601/#/dashboard/MU2E-pre-emption-rate-per-facility-over-time?_g=()&_a=(filters:!(),panels:!((col:1,id:mu2e-pre-emption-rate-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization),(col:4,id:mu2e-disconnect-rate-per-facility-over-time,row:4,size_x:6,size_y:3,type:visualization),(col:7,id:mu2e-evicted-rate-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization)),query:(query_string:(analyze_wildcard:!t,query:'*')),title:'MU2E%20pre-emption%20rate%20per%20facility%20over%20time')
  - NOvA

    http://osgmonitor.fnal.gov:5601/#/dashboard/NOvA-rate-of-pre-emption-per-facility-over-time?_g=()&_a=(filters:!(),panels:!((col:1,id:nova-pre-emption-rate-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization),(col:4,id:nova-rate-of-disconnect-per-facility-over-time,row:4,size_x:6,size_y:3,type:visualization),(col:7,id:nova-rate-of-evicted-per-facility-over-time,row:1,size_x:6,size_y:3,type:visualization)),query:(query_string:(analyze_wildcard:!t,query:'*')),title:'NOvA%20rate%20of%20pre-emption%20per%20facility%20over%20time')
- All these dashboards will be available in Grafana
  - http://fermicloud079.fnal.gov/monitor/dashboard/db/evicted-and-disconnected-jobs-by-site

Plan and Future work

- I will include jobs information of other experiments in Elasticsearch

Questions?