The Cluster Monitoring System of IHEP

Qingbao Hu

huqb@ihep.ac.cn

Computing Center, Institute of High Energy Physics,
Chinese Academy of Sciences

International Symposium on Grids and Clouds (ISGC) 2016

Content

- Overview of IHEP's Monitoring System
- Optimization of the monitoring tools
- Logger-analysis Monitoring
- Summary

Status of IHEP Cluster

- ~ 1,122 work nodes
  - ~ 13,500 CPU cores
- ~ 5PB disk storage
  - Lustre, gLuster, openAFS, etc.
- ~ 5PB tape storage
  - Two IBM 3584 tape libraries, LTO4 tape
  - Modified CERN CASTOR 1.7

Cluster built with blades

Tape libraries

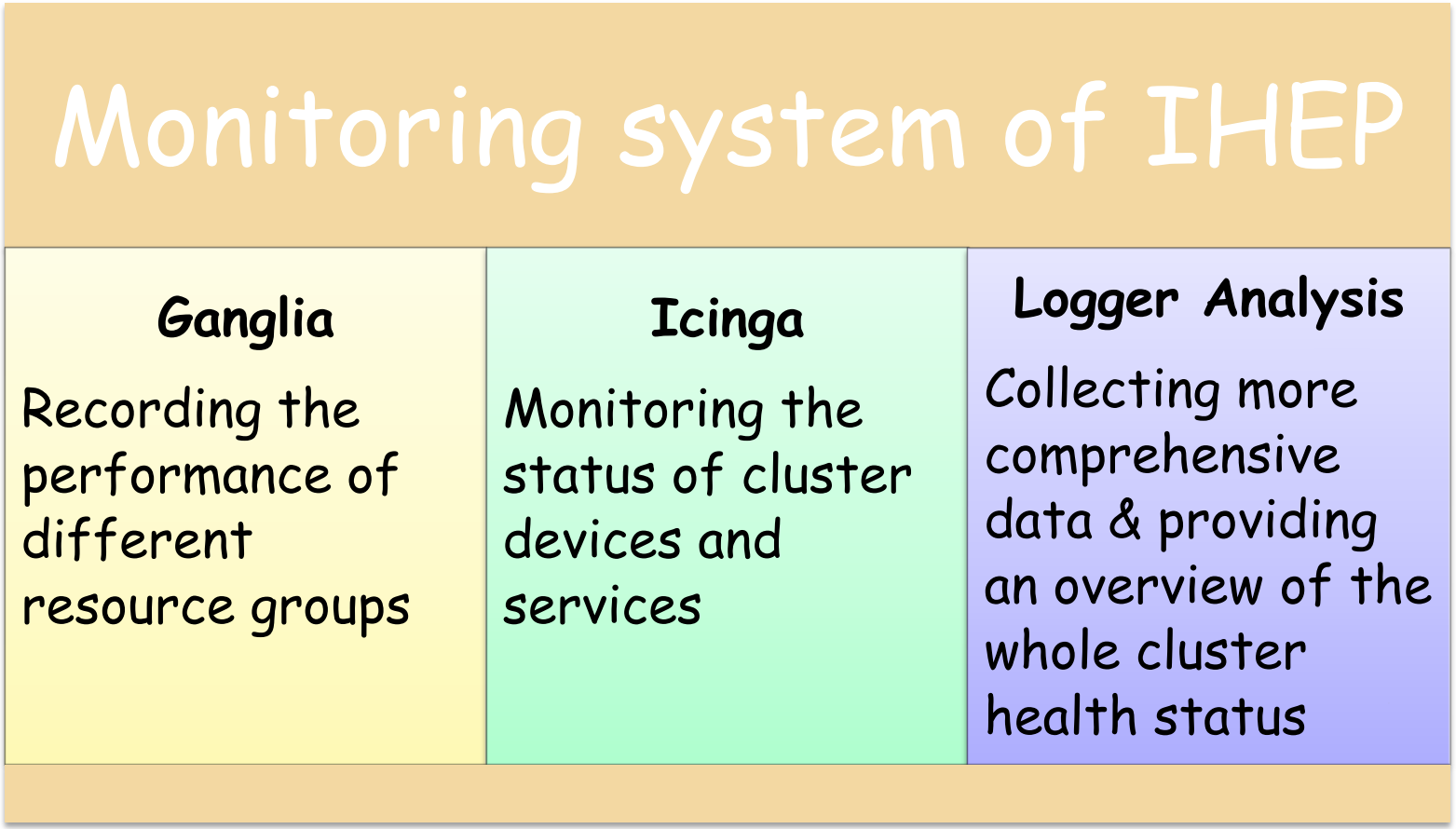
Monitoring requirements

- A large number of hardware and software resources
- Cooperated in complex ways

  - Large Scale (> 2,000 nodes)
  - Heterogeneous device resources
  - Good Scalability
  - Real-time information display and alarm
  - Combination of active detection service and passive information receiving.
  - Auto recovery of failed services.

Monitoring System Overview

- System overview



Ganglia

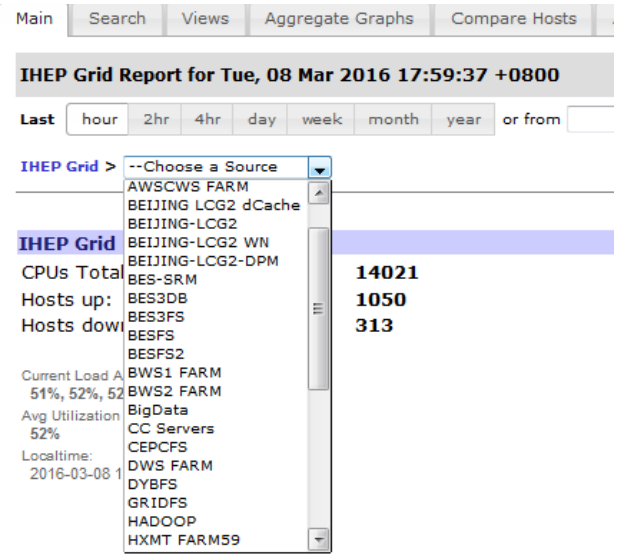
- **Monitoring the health of the cluster**
  - System load
  - CPU utilization
  - Network bandwidth and traffic
  - Memory usage
- **Usage**
  - Records history status of the cluster
  - Helps system manager to fix problem

Ganglia of IHEP

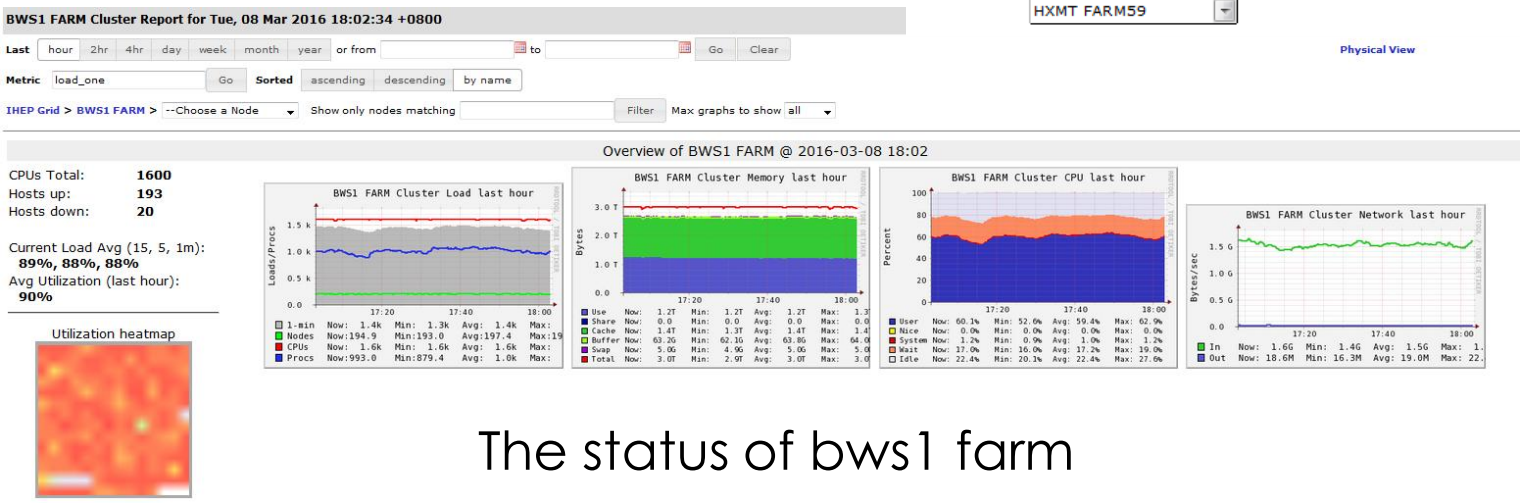
- **The bottleneck of Ganglia**
  - High frequency: Collect 20 metrics from each monitored node every 15 seconds.
  - Pool scalability: Large number of nodes cause a large amount of metrics data, which pulls up the server's peak iowait and slows down the monitoring service.
- **Workaround:**
  - Create a ram disk on the Ganglia server to save the RRDs data.
  - Improves the IO performance of the server disk

Ganglia of IHEP

- IOwait < 1%



Different clusters



The status of bws1 farm

Icinga

- **Created as a fork of the Nagios**
  - Plug-in design
  - Active check of the service
  - Flexible configuration by NagiosQL

- **Usage**
  - Hardware (CPU load, disk usage, etc.)
  - Network connectivity (HTTP, POP3, ping, etc.)
  - Computing services on work nodes
  - Distributed file system services …

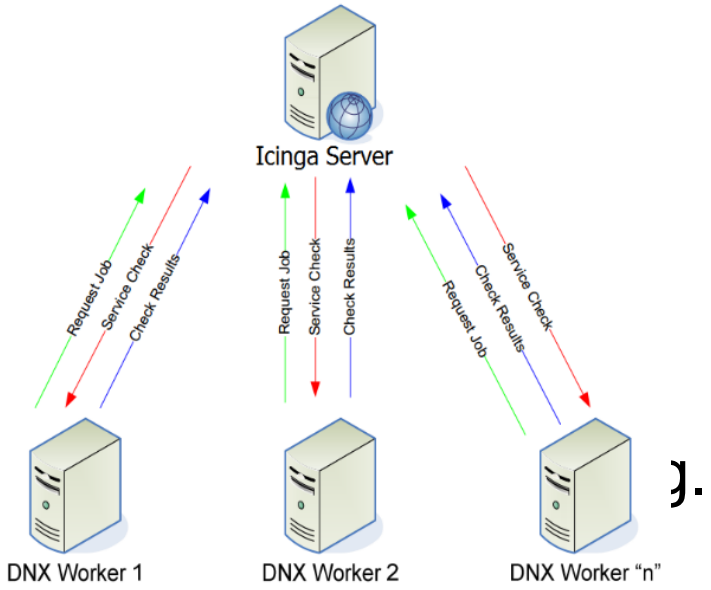
Icinga of IHEP

- **Polling agents we developed**
  - More services monitored
  - Some crashed service faults can be recovered automatically
  - Critical errors are alarmed to system manager via both email and SMS

bws0382	PBS-Client	正常	2016年03月18日 01:28:20	2天 6时55分47秒	1/2	PBS_MOM OK: Daemon is running. Host is listening.
	PBS-zombie	正常	2016年03月18日 01:28:08	11天17时37分54秒	1/2	there is no zombie, OK
	SSH	正常	2016年03月18日 01:28:28	125天 8时44分13秒	1/2	SSH OK - OpenSSH_5.3 (protocol 2.0)
	check_afsfile	正常	2016年03月18日 01:28:08	100天 7时48分14秒	1/2	afsfile afscache are OK
	check_automount	正常	2016年03月18日 01:27:36	125天 8时46分50秒	1/2	Automount OK: Daemon is running. Host is listening.
	check_disk_local	正常	2016年03月18日 01:28:08	29天 1时15分32秒	1/2	DISK OK - free space: / 47480 MB (82% inode=91%):
	check_diskwrite	正常	2016年03月18日 01:28:08	125天 8时51分54秒	1/2	local disk and scratch can write
	check_mem_hardware	正常	2016年03月18日 01:27:45	7天 8时50分54秒	1/2	Memory is OK
	check_ntp_time	正常	2016年03月18日 01:28:12	0天22时33分 9秒	1/2	NTP OK: Offset 0.00288772583 secs
	check_ping	正常	2016年03月18日 01:27:46	1天20时52分39秒	1/2	PING OK - Packet loss = 0%, RTA = 0.93 ms
	check_puppet_agent	正常	2016年03月18日 01:14:10	0天13时54分26秒	1/2	OK: Puppet agent "3.8.1" running catalogversion 0.055582
	glustre_mount	正常	2016年03月18日 01:27:52	22天 9时23分 4秒	1/2	/besfs2 size is OK
	lustre_mount	正常	2016年03月18日 01:28:29	66天10时30分24秒	1/2	besfs bes3fs publicfs dybfs workfs scratchfs cefs are OK

Icinga of IHEP

- **The bottleneck of Icinga**
  - Single collector node.
  - Vast amounts of the service c
    server load, which reduces th g.
  - Many detection results are de



- **Workarounds:**
  - Distributed Nagios eXecutor. (DNX)+ Icinga Sever
    - » A modular extension of Nagios
    - » DNX Worker requests jobs from the Icinga (Scheduling) Server
    - » DNX Worker executes the plug-in agents and return the results to Icinga server.
  - Balance the load of servers via distribution
  - Decrease the latency of the polling

Icinga in IHEP

	scale of Monitoring hosts	scale of Monitoring service	The average host delay	The average service delay
No DNX	1257	9796	251.588sec	256.930sec
No DNX	1265	12222	789.429sec	789.000sec
Use DNX	1343	13841	0.365sec	0.644sec

Logger-analysis Monitoring

- **Monitoring based on the logger Analysis**

  Log : records relating to activities occurring on system.

  - The reliability of the hardware
  - The stability of the service
  - The availability of the system
- **logger-analysis requirements**
  - Large Scale & Scalability
  - Real-time information display and alarm
  - Convenient query
  - Flexible configuration
- **Provides a novel monitoring based on log analysis**

Logger-analysis

- Log data store & search
- Elasticsearch: Search & Analyze Data in Real Time
  - Distributed, scalable, and highly available
  - Real-time search and analytics capabilities
  - RESTful API

elasticsearch.

Real-time Log Collection

- Flume
  - Distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data.
  - Simple and flexible architecture based on streaming data flows.

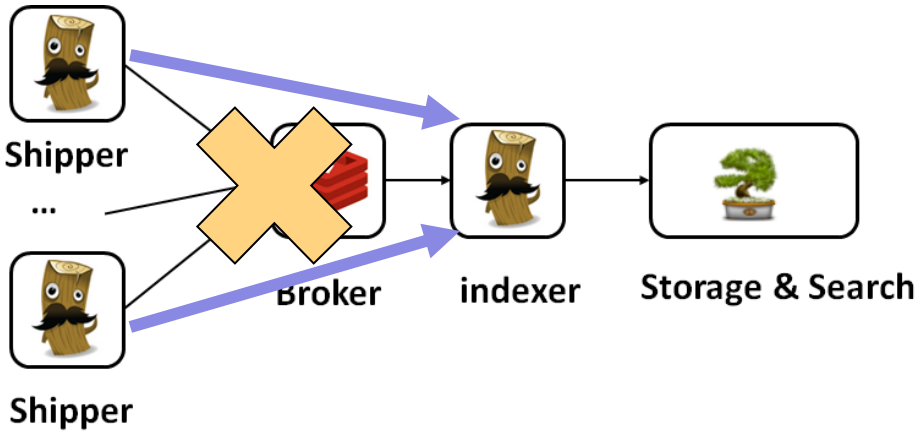
Flume

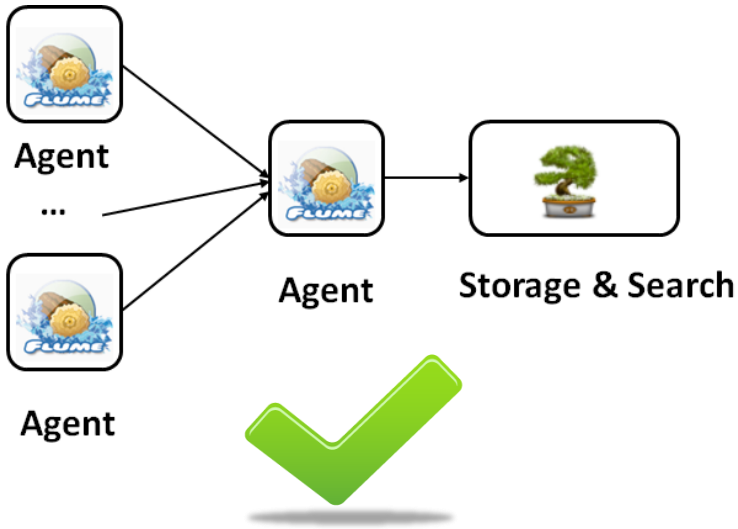
- Logstash : Process Any Data, From Any Source
  - Centralize data processing of all types
  - Normalize varying schema and formats
  - Quickly extend to custom log formats
  - Easily add plugins for custom data sources

Logstash

Real-time Log Collection

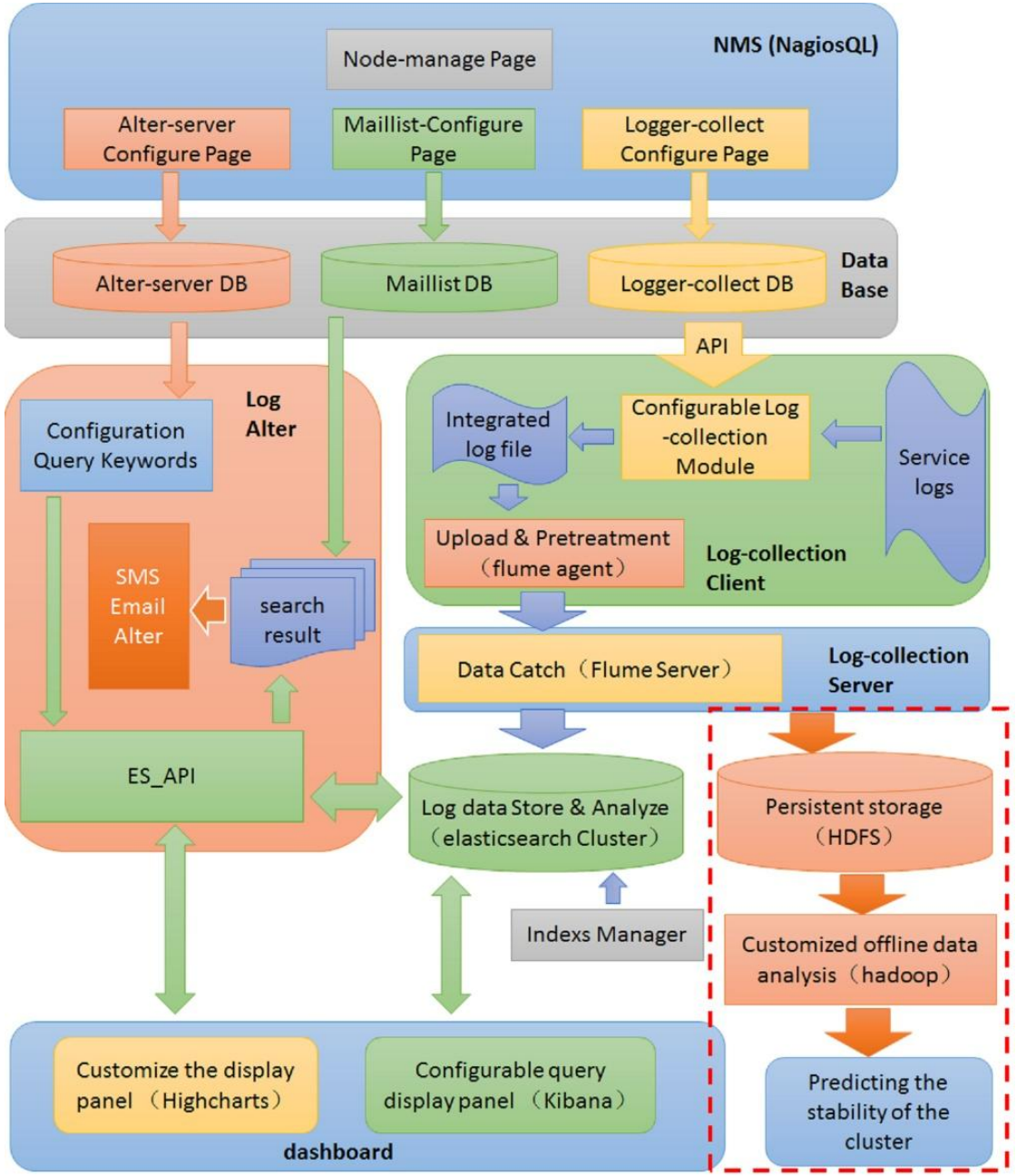
- Three models (throughput)
  - 1.Logstash & Redis & Elasticsearch (low)
  - 2.Logstash & Elasticsearch (middle)
  - 3.Flume & Elasticsearch (high)



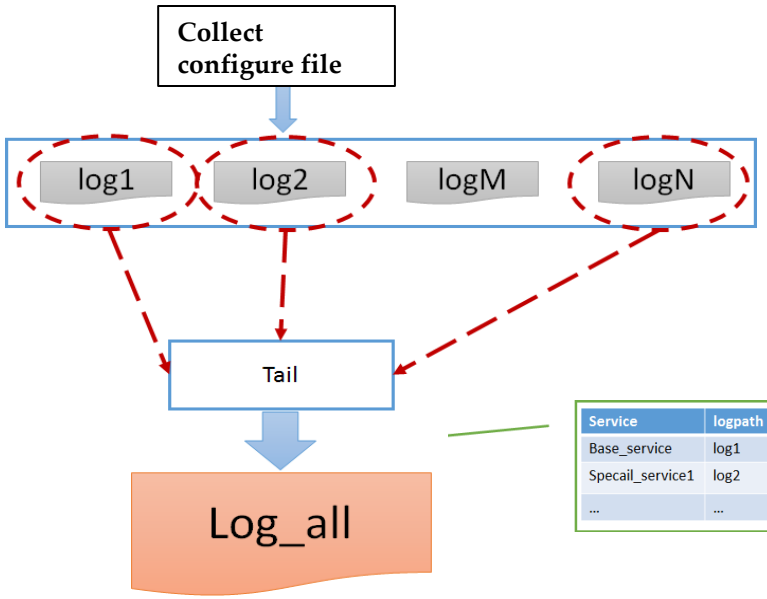


Logger-analysis

- Flexibility
- Scalability
- Real-time



Logger-collect client



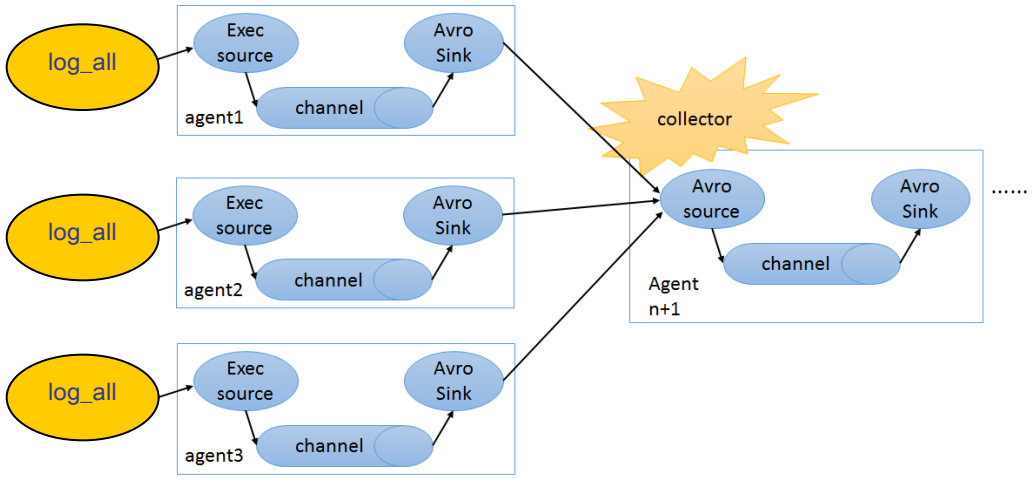
- No log missed
  - logpath + inode + offset
  - Tail + awk

  tmp file record the file inode and the file offset info to guarantee the continuity of the log data collection when collect service crash.

- Logs from various servers can be collected
  - Log format defined by dedicate configure file by administrator

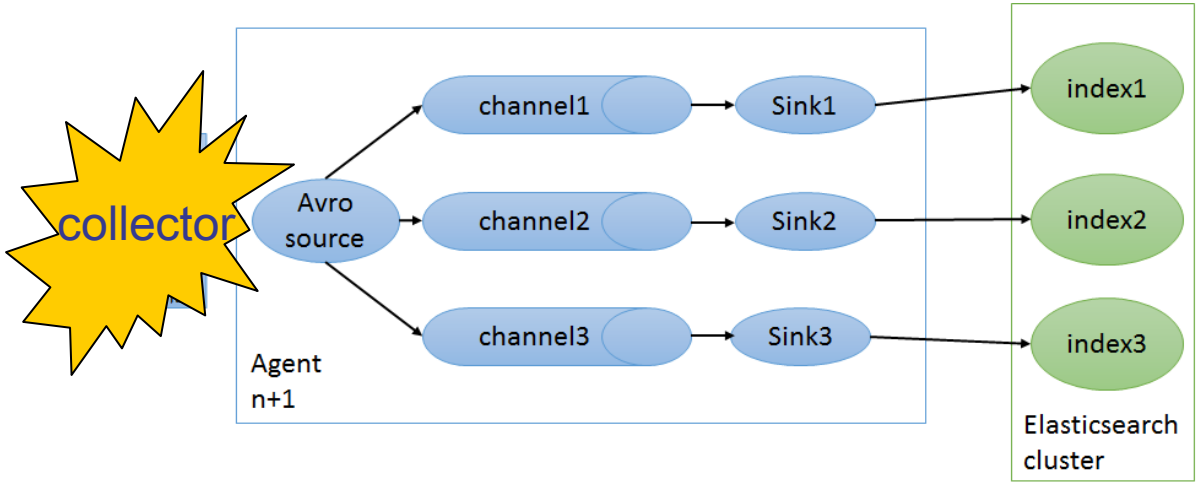
Logger-collect client

- Flume multi-agent fan-in flow model
- Pre-processing log & Upload data real-time



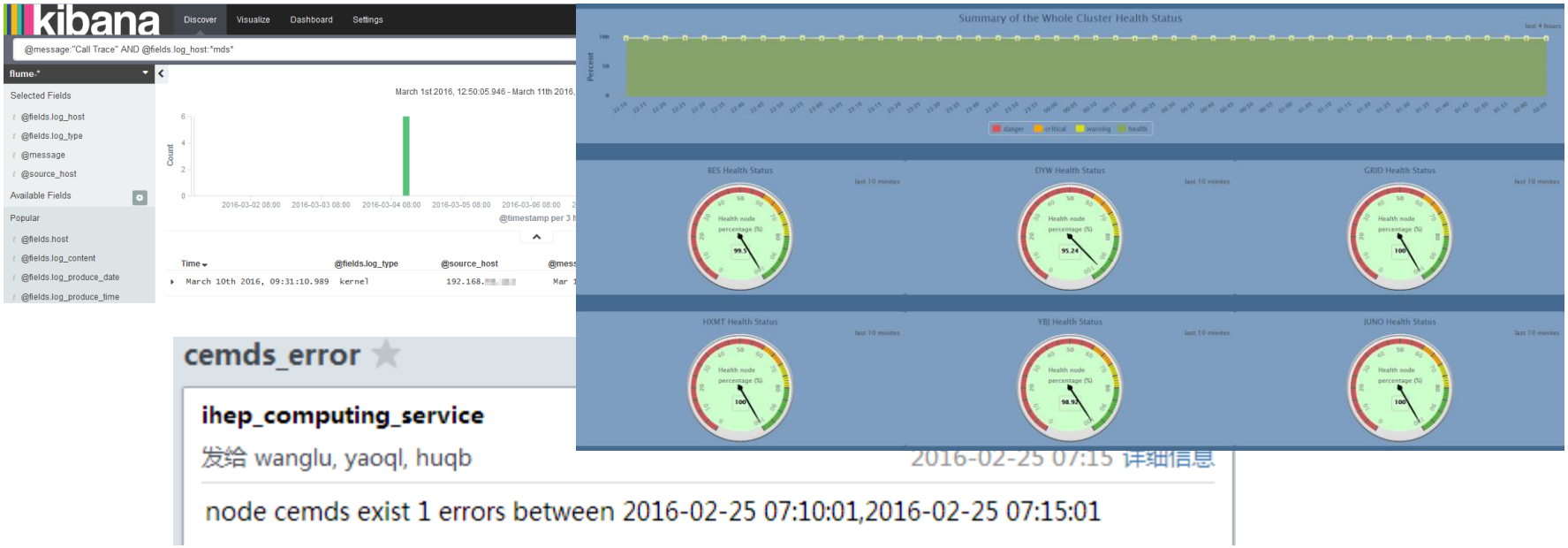
Logger-collect server

- Flume multi-agent fan-out flow model
- Separate different service log data and store in different indexes.



Function developed based on ES API

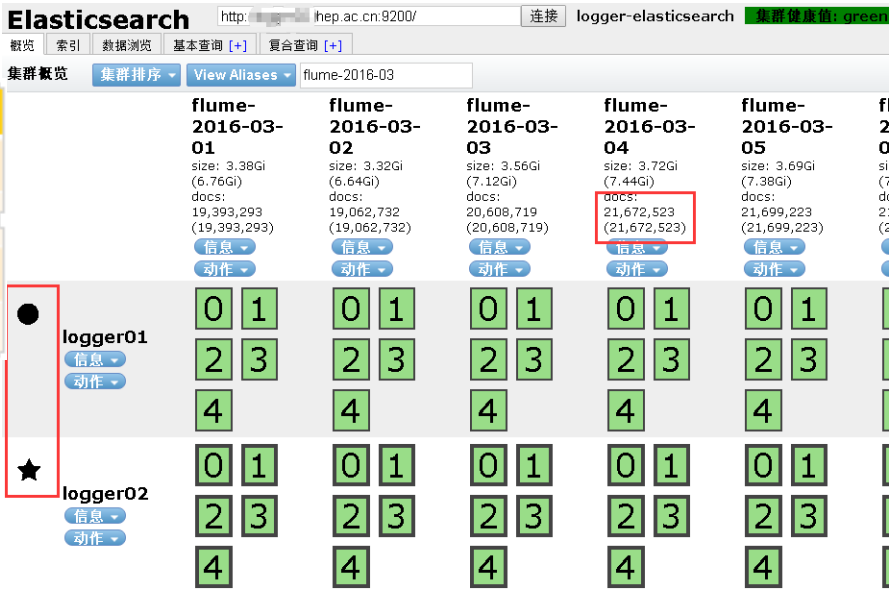
- use keywords to locate the service failure time
- Real-time email alerts
- Display the health status of the whole cluster



Log-analysis deployed at IHEP

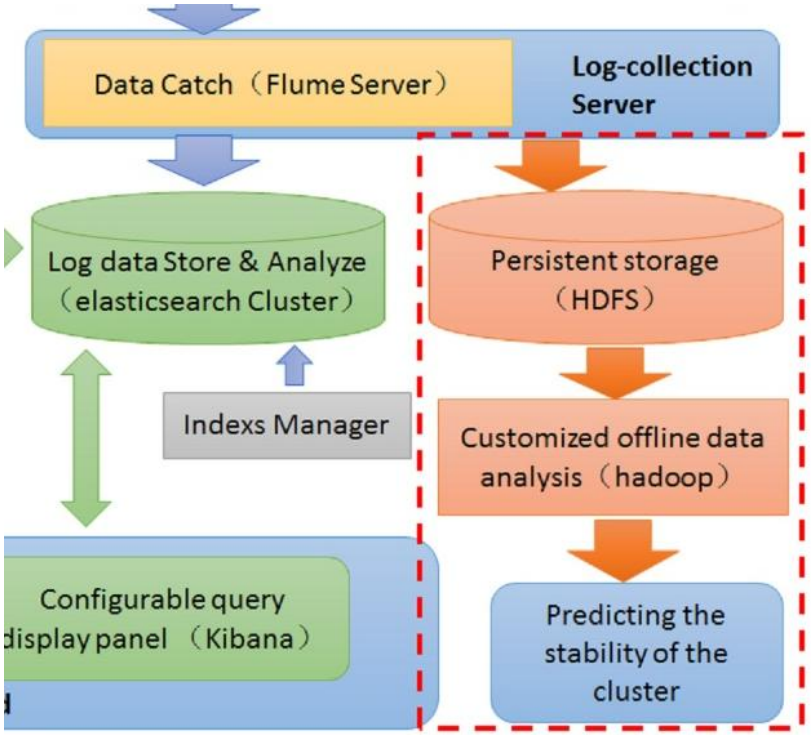
- The number of monitored nodes > 2,000
- The amount of logs collected per day ~ 20M entries
- The interval between the log produced and stored < 40 s
- Maximum throughput reach 20,000 records per second

	Elasticsearch-server	Flume-server
CPU	E5-2640 v3 @ 2.60GHz	E5420 @ 2.50GHz
RAM	64G	16G
System	Scientific Linux 6.2	Scientific Linux 6.2
Number	2	1



Next step

- Regular expression of log format will be supported for more detailed fields
- Archive log data to HDFS
- Offline log mining based on Hadoop or Storm



Summary

- Ganglia and Icinga guarantee the stability of the IHEP cluster.
- Log-analysis provides a novel monitoring.
- Log mining will be done next.

Thank you!
Any Question?