DEVELOPMENT OF STATUS ANALYSYS SYSTEM BASED ON ELK STACK AT J-PARC MLF

[1]Kentaro Moriyama, [2]Takeshi Nakatani, [3]Yoshiji Yasu, [4]Hidetoshi Ohshita and [4]Tomohiro Seya
[1]Neutron Science and Technology Center, Comprehensive Research Organization for Science and Society (CROSS)
[2]J-PARC Center, Japan Atomic Energy Agency (JAEA)
[3]J-PARC Center, High Energy Accelerator Research Organization (KEK)

Introduction

In recent neutron scattering experiments, large quantities of various types of data, including raw data, metadata, logs and metrics have been generated by the system, apparatus and devices. At J-PARC MLF, it is possible to conduct many experiments under various conditions within short time by using high-intensity neutron beams, high-performance neutron instruments, and various sample environments. In this experimental environment, it is essential to perform efficient and effective data analysis. Additionally, since it has been almost nine year from the start of operation in MLF, the rate of occurrence of failures is rising due to ageing of devices. Given that such failure can lead to loss of precious beam time, failure or its signs should be detected early. The MLF status analysis system based on Elasticsearh, Logstash, and Kibana (ELK) Stack, which is one of the web-based framework that is being rapidly adopted for big data analysis, collects various data from neutron instruments. It offers insight to decision-makers in terms of data analysis and experimentation as well as instrument maintenance, by facilitating flexible user-based analysis and visualization.

J-PARC MLF

- J-PARC MLF is an experimental facility for neutron scattering, providing high intensity beam in the world since 2008.
- About a thousand users annually perform several hundreds experiment from a wide variety of research fields using 21 neutron instruments.
- Each neutron instrument is equipped with a large-area neutron detectors and a wide variety of purpose-build sample environment devices.

Status analysis with log

- Neutron instruments generate large amount of various logs that contains useful information from the viewpoint of instrument operation and data analysis.
- There is a need for as tool that can provide useful insight into action required for instrument operation and data analysis.
- However, there is not much opportunity to positively utilize these logs and no convenient tool to easily analyze logs.

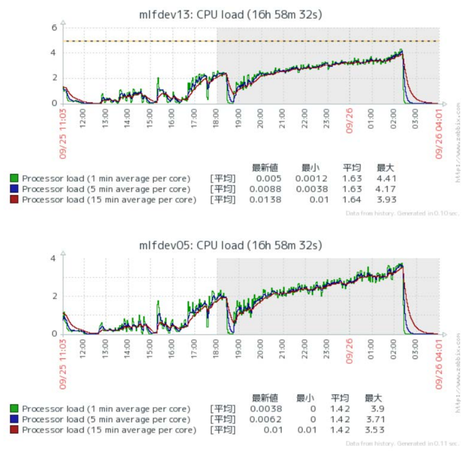
Status analysis system

System structure

- Each main component of the ELK Stack runs on different physical servers.
- Elasticsearch has a cluster configuration composed of 2 physical servers because indexing and search process can be memory- and CPU-resource intensive.
- In Elasticsearch cluster loads and documents are adequately distributed by the master node.
- Logstash and Kibana run on a single server.
- Beat runs on the server generating logs to ship log events to Logstash.

Server	Specification	Unit
Elasticsearch	DELL PowerEdge R410 OS: Scientific Linux 6.9 CPU: Intel Xeon E5620, 2.4GHz Memory: 12GB HDD: 1TB SATA 7.2k rpm x4 Elasticsearch: 5.6.1	2
Logstash, Kibana	DELL PowerEdge R320 OS: Scientific Linux 6.9 CPU: Intel Xeon E5-2420 v2 2.2GHz Memory: 16GB HDD: 1TB SATA 7.2k rpm x4 Elasticsearch: 5.6.1 Kibana: 5.6.1	Each 1
Filebeat	Filebeat: 5.6.1	

Server Specifications



CPU load of cluster node in Elasticsearch

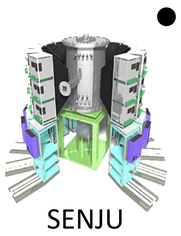
Test for log collection

We applied the status analysis system to SENJU, a single-crystal neutron diffractometer installed BL18 in MLF.

- Collected three types of logs generated by IROHA2, which is a instrument control software integrating device control and DAQ.

Log type	Contents	Frequency of update
Surveillance	timestamp, log level, and physical values	fixed time
Measurement	timestamp, physical values	fixed time
Operation	Timestamp, log level, and message	each time any operation

- Took the simplest indexing condition for testing and brief performance evaluation.



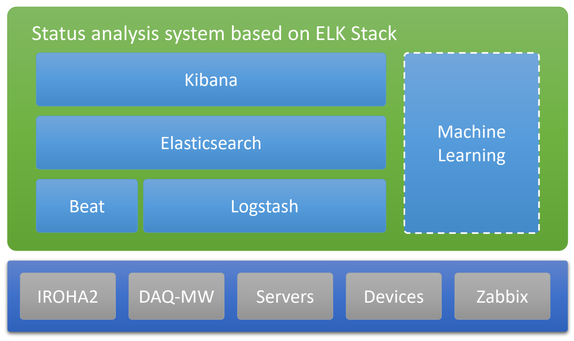
SENJU

Item	Condition
The number of Shard	1
The number of Replica	0
Document schemas	not defined

Big data approach

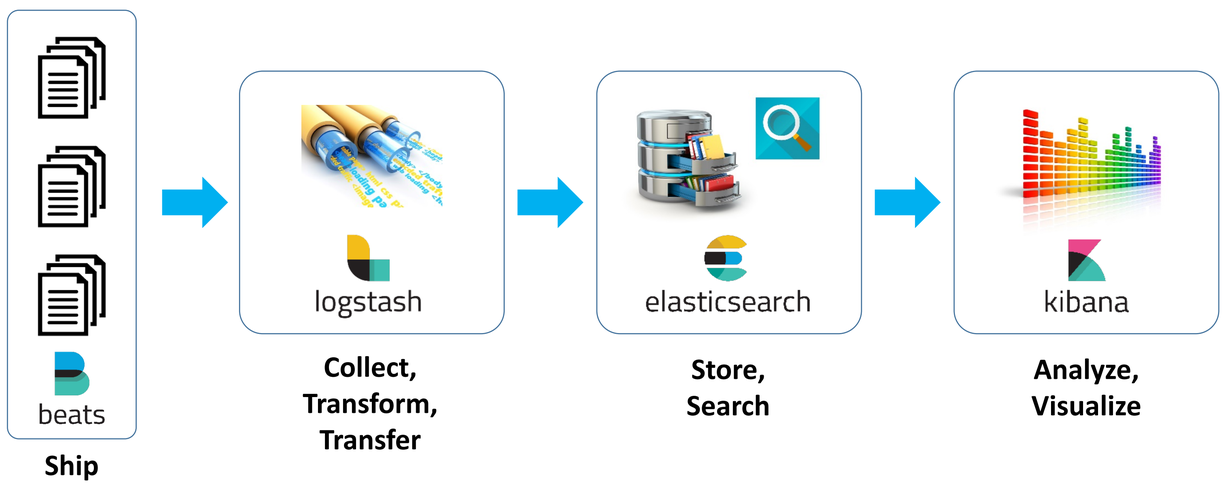
We employed a bid data approach for the status analysis of various log information.

- The system based on the ELK Stack, an open-source software for big data analysis.
- It collects and analyze various log information from neutron instruments.
- We plan to combine the system with a machine learning schema to facilitate anomaly detection and advanced data analysis.



ELK Stack

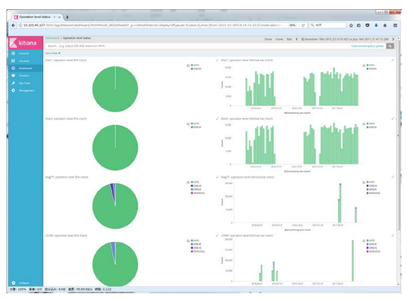
The ELK Stack, developed by Elastic Inc. consists of three main components and a subcomponents. It allows to easily and flexibly analyze log information leading to gain insight to facilitate data analysis as well as instrument operation.

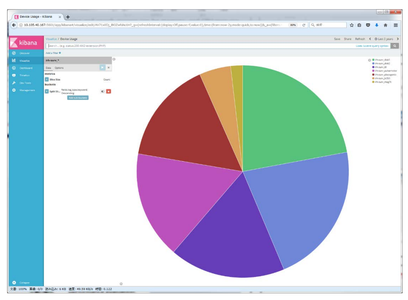


Status Analysis

We prepared a number of graphs and dashboards with Kibana to utilize the status analysis of the IROHA2 log supporting instrument operation and decision-making such as experimental planning and data analysis.

Use case1. analysis for experimental planning	Use case2. analysis for experimental data analysis	Use case3. analysis for instrument operation
• Monitor the experiment progress. • Correlation analysis of sample environmental parameters. • **Support decision-making for experimental planning.**	• Search, analyze, and download sample environmental parameters. • Correlation analysis of sample environmental parameters and neutron detection. • **Support decision-making for experimental data analysis.**	• Monitor the operation status. • Error and warning occurrence. • Device usage condition • Detect failure and its signs. • **Support decision-making for device improvement planning.**





Conclusion and Future Plan

- We have developed a status analysis system based on the ELK Stack to integrally analyze log information generated in neutron scattering experiments.
- The system was applied to the neutron scattering instruments SENJU at J-PARC MLF and the status analysis of IROHA2 log was performed.
- We are planning to introduce an online machine-learning scheme for advanced status analysis.