ISSN: 2755-0168

Journal of Biotechnology & Bioinformatics Research

SCIENTIFIC Research and Community

Review Article

Open Access

System Event Metrics Monitoring With Elk Stack: A Generic Solution Case Study

Rakesh Rikhi* and Deepika Rikhi

Austin,Texas, USA

ABSTRACT

Systems (Servers and Desktops) are essential lifelines for any organization that has automated its processes via a modern IT framework. The health of these systems is measured and derived via key factors (aka KPIs) such as uptime, threshold, CPU, and memory utilization. The volume of these metrics can become very large in a short amount of time, which can become a challenge in the retrieval, analysis, and reporting of the information derived from raw data. With the availability of generic software libraries for faster storage, retrieval, and presentation of the data in the ELK stack (Elastic Search, Logstash, and Kibana), the system metrics can be handled effectively and efficiently, which can then be used for faster reporting.

***Corresponding author**

Rakesh Rikhi, Austin,Texas, USA

Received: August 03, 2022; **Accepted:** August 10, 2022; **Published:** August 17, 2022

Keywords: ELK Stack, Elastic Search, Logstash, Kibana, System metrics, KPI

Introduction

In the early 2000s, with the considerable increase of blogs and user groups, there was a sudden jump in the amount of text content created and stored on the internet. Apache Lucene came to the open-source software market and became popular with its features to quickly search text data.

Elastic Search is an enhanced version of Lucene software for storage, distributed search, and analytics. It is designed to handle large volumes of data efficiently. Elastic Search provides full-text search, real-time indexing, scalability, and distributed architecture. Beyond search, Elastic search also provides string aggregation features for analytics, which allow for the summary and analysis of huge datasets.

Elastic also created presentation-layer software libraries with a configuration-based methodology for building data visualization and exploration tools that can work with Elastic Search. Kibana's key features include data visualization, dashboards, search and query, data exploration, monitoring and alerts, reporting, machine learning, and security.

Figure 1: IT systems monitoring is only possible in the modern world with a proper data collection, storage, and virtualization framework.

Logstash is also an open-source software library for data ingestion, which allows data collection from varied sources to be transformed and sent to Elastic search for storage. There are more than 200 prebuilt open-source plugins based on Logstash that can help for indexing the data. Because of its tight integration with Elastic Search, it's a popular choice for loading data into Elastic Search.

These three software libraries can work together to create data pipelines for metrics collection, storage, analysis, presentation, and reporting. Based on the analytics, real-time alerts can be generated, which can be used to troubleshoot and take necessary actions to solve a system event.

Logstash will ingest, transform, and push data to the central elastic search storage. Elastic Search will perform indexing, storage, and analytics. Kibana will visualize the analytics outcome.

In January 2021, Elastic announced that it would change its licensing strategy and not release a new version of the ELK stack libraries under the permissive Apache License. Instead, the latest versions of the libraries will be distributed under the Elastic License, with the source code under the Elastic license or SSPL. These licenses are not open source. To cater to the open-source community, AWS introduced the OpenSearch project, which provides a secure, high-quality, fully open-source search and analytics suite.

The OpenSearch suite has a search engine, OpenSearch, and a visualization tool called OpenSearch Dashboards.

System Metrics

What

System Metrics are numerical representations of system performance data that can be used to determine a system's overall behavior over time. Event Metrics are derived on top of the system performance data to highlight a spike or an issue in the system's health.

Why and How

As modern systems keep running, operating, and catering to business needs 24x7, it becomes important to monitor the system performance over time to catch and fix any issue in the system degradation.

Each system has its own metrics and ways to gather raw system metrics. For example-

Database Systems

The main job of the database systems is to store data so that it can be retrieved and analyzed by the users. The operations are to write and read the data. The performance metrics fall into these classes-

- **Memory Usage:** Memory Capacity, Cache hit ratio, Page Life Expectancy, Resource Usage, Row Counts
- **System Performance:** Database File I/O, CPU Utilization, Lock Waits, Blocking, Indices

Web Servers

These IT systems interact with the outside world over the internet. These are mainly used to respond to user input via web pages. The main performance metrics of these systems are-

- **Server Performance:** Requests Per Second (RPS), Error Rates, Uptime, Thread Count, Average Response Time (ART), Peak Response Time(PRT)
- **Server Availability:** Uptime, Downtime, Network Speed

Network Monitoring

The network that connects various IT systems is crucial in the whole process. The metrics are as follows-

a. Traffic Statistics
b. Device Statistics
c. Bandwidth Utilization

Process Monitoring

Any process can be monitored by creating custom events and metrics for the business processes-

- Incoming traffic statistics
- Process Efficiency- Number of requests catered per hour
- Phase Completion Times

Operation

High-Level Flow diagram is shown below-

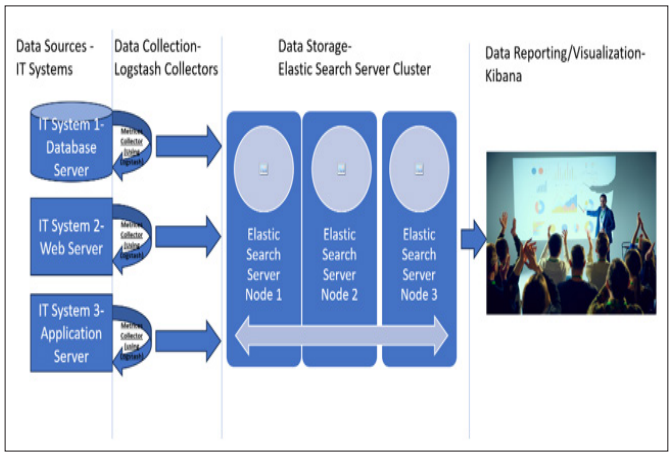

Figure 2: IT systems monitoring- Various Phases using ELK stack components

1. Logstash provides various input plugins that enable a specific event source to be read.
2. The collected data is stored in an Elastic Search Cluster with Load Balancing to make the data collection and storage process efficient.
3. Kibana provides various data visualization data tools (graphs, pie charts, process flows)

Challenges

Because the IT systems landscape is continuously evolving with new technologies like Machine Learning, AI, and Neural Networks, new Logstash collectors need to be created and integrated into the monitoring framework.

All the effort is pivoted on the fact that timely action is taken on the events and reports generated in the data visualization and reporting phase. With the advent of virtual servers and containers, the performance data must be processed in real-time because these servers and containers get created and destroyed within minutes and hours. Eliminating False Positives and Outliers can be a real challenge, so proper data cleanup is required before the analytics.

Futuristic View

Data Science, Machine Learning, and AI should be utilized for anomaly detection in minimal time.

Apart from overutilization and event monitoring, underutilization of the systems and processes should also taken on the radar for performance monitoring.

References

1. The Rise of Elastic Stack 2016. https://www.researchgate.net/publication/309732494_The_Rise_of_Elastic_Stack
2. An Unsupervised Condition Monitoring System for Electrode Milling Problems in the Resistance Welding Process 2022. https://www.researchgate.net/publication/361286396_An_Unsupervised_Condition_Monitoring_System_for_Electrode_Milling_Problems_in_the_Resistance_Welding_Proces
3. Visualizing Web Server Logs Insights with Elastic Stack- A Case study of UMmail's access logs 2018. https://www.researchgate.net/publication/326172565_VISUALIZING_WEB_SERVER_LOGS_INSIGHTS_WITH_ELASTIC_STACK-A_CASE_STUDY_OF_UMMAIL'S_ACCESS_LOGS