Impact Factor: 9.2 ISSN-L: 2544-980X

Analysis of Log-Files of Technological Devices

Urinov Nasillo Fayzilloevich[1], Abdullaeva Dilnavoz Khusniddinovna[2]

Annotation. The article covers a method of using specialized stack of software solutions for the collection, storage and centralized control of log files of process equipment. Research on methods of analyzing the process equipment data were carried out based on open data set provided by the NASA Ames Research Center.

Key words: data collection, technological equipment, log files, ELK stack, data visualization, analysis of technological data.

Introduction

Data collection from technological equipment is one of the important tasks in industrial enterprises, as well as in conducting of research work. Tracking a variety of diagnostic data and timely receipt of information about the operation or failure of equipment allows production immediately responding to emergencies, take actions to reduce downtime and extend equipment life [1, 2].

In order to ensure the performance of systems, engineers analyze various types of data generated by control systems during their operation [3]. Files about events created by CNC machines - log files, which are formed in the form of text, can serve as a source of data on the operation of technological equipment used for further analysis. Based on the process information provided in the log files, monitoring of the system parameters, identifying irregular situations, or obtaining other information (for example, equipment uptime, machine status data) that can be useful in analyzing the causes of failures can be carried out. However, in its raw form, working with log files becomes a labor-intensive task due to the lack of the ability to search and filter data, which is unstructured. As a result, the creation of a system for aggregation, control and display of log files to present technological information to higher levels of enterprise management is an urgent task [4]. To solve this task, the article proposes a way to use a set of specialized software tools that allows collecting, storing and centrally controlling the log files of technological equipment.

Methods of research

Toolkit for analyzing unstructured technological data

Collection, storage, processing of unstructured data at the production site is an expensive task for most enterprises [5]. In this regard, when analyzing existing tools for processing log files, in addition to technical characteristics, it is important to take into account the cost of software components. In this regard, the ELK software stack was chosen to build a system for analyzing log files, which is also distributed as open source solution. ELK is an abbreviation of the names of three software products: Elasticsearch, Logstash, Kibana, developed and supported by Elastic (USA).

The core of the stack is a component of Elasticsearch, which is a database with a full-text search and analysis system based on the specialized Apache Lucene search technology, which makes Elasticsearch different from relational databases or NoSQL systems. In relational databases, there are

[1] *candidate of technical sciences, associated professor*
[2] *doctoral student, Bukhara Engineering-Technological Institute*

concepts such as rows, columns, tables, and schemas. Elasticsearch and similar repositories work differently. The basic unit of information stored in Elasticsearch is a json-document, which is a text format for data exchange between a client and a server. As shown in Fig. 1, documents are stored inside types, and types are stored inside indexes. An index can contain one or more types, and each type can contain a huge number of documents.

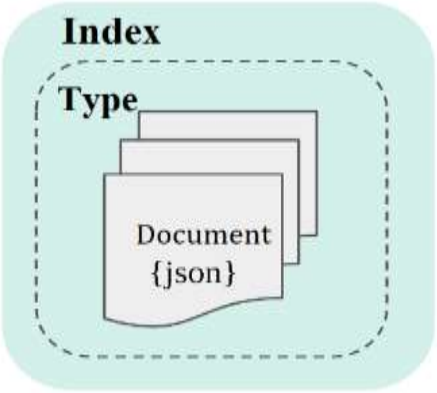

Fig. 1. Index structure in Elasticsearch.

Index structure in Elasticsearch can be compared to the database structure in relational databases. Continuing the analogy, the type in Elasticsearch corresponds to the table, and the document corresponds to the records in the table (Table 1).

Relational database	Database	Table	Record in table	Field in table
Elasticsearch	Index	Type	Document (json)	Field in document

Indexes are divided into data segments and distributed among cluster nodes. A node is a single server in the system, which may be part of a large cluster of nodes. The cluster consists of several nodes, each of which is responsible for storing and managing its part of the data (Fig. 2). Elasticsearch is a distributed system that is designed to work even if the hardware it runs on fails. For this, copies - replicas of the main index segments are provided. In presence of replicas, if the first node fails, then the segment from this node will still be available in the other two nodes. In order to access the distributed system of main shards, a coordinating node receives search queries and then sends reformulated queries to the cluster nodes.

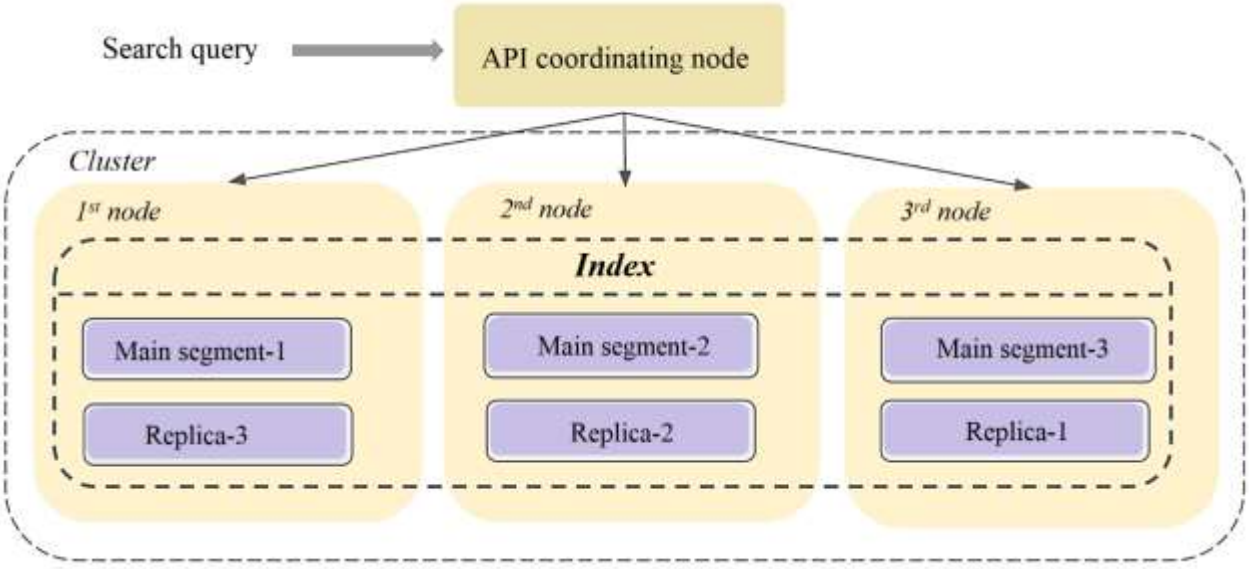

Fig.2. Architecture of Elasticsearch.

The second component of the stack, i.e. Logstash is a log file aggregator that collects data from various input sources, performs the necessary transformations, and then sends them to the database for further processing. The event-processing container in Logstash has three stages: entry, filtering, output (Fig. 3).

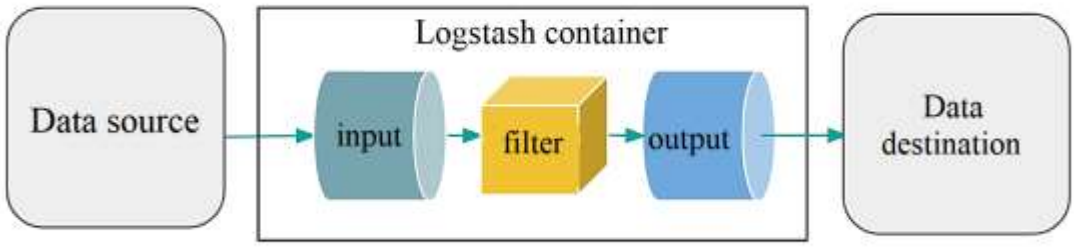

Fig.3. Architecture of Logstash.

Only the input and output stages are required, filtering is an optional part. The input stage creates events, filters modify input events, and outputs send them to the destination. The Logstash container is stored in a configuration file. Configuration file sections, i.e. input{}, filter{}, output{} contain one or more plugin configurations. The input plugin is designed to customize the events passed to Logstash. The filter plugin is used to modify the data. The output plugin is used to send data to the destination.

Conclusions

In the course of the study, the most suitable technologies for storing and analyzing log files of technological equipment were studied and selected, the structural and architectural schemes of the system were developed, and a test bench was deployed. Then using Kibana, a visualization tool, infographics were built based on the data. A practically significant result is the ability to work with data not by reading text files, but by using a tool for analytics and building various data control panels. This makes the process of analyzing initially unstructured data accessible and fast. Subject specialists get access to data and a powerful tool with which they can conduct analysis. One of the advantages of the selected software package is that the data entered into Elasticsearch is stored in a structured form, the structure of the logs is determined at the stage of setting up the configuration file in Logstash. Having the data in a prepared and easy-to-analyze form, specialists can focus directly on the analysis itself to obtain important information from the log data, rather than spend time structuring the data, thereby significantly reducing the time spent in the analysis process.

References:

1. Nikishechkin P., Kovalev I., Nikich A. An approach to building a cross-platform system for the collection and processing of diagnostic information about working technological equipment for industrial enterprises // MATEC Web of Conferences.–2017.–V. 129.–P. 03012. https://doi.org/10.1051/matecconf/201712903012.
2. Kvashnin D.Yu., Kovalev I.A., Nezhmetdinov R.A., Chekryzhov V.V. Aggregation of information on the operation of technological equipment using the Industrial Internet of Things // Automation in Industry, No. 5. 2019. pp. 29-32.
3. Evstafieva S.V., Pushkov R.L., Salamatin E.V. Collection and visualization of operational data from technological equipment // Automation in industry, No. 5. 2019. pp. 26-28.
4. Nikishechkin P.A., Kovalev I.A., Grigoriev A.S., Nikich A.N. Cross-platform system for collecting and processing diagnostic information on the operation of technological equipment // Bulletin of "Stankin" MSTU. - 2017. - No. 1 (40). - pp. 34-56.

5. Kovalev I.A., Nezhmetdinov R.A., Chervonnova N.Yu., Abdulov R.R. Synthesis of systems for remote diagnostics and monitoring of CNC machine tools using Web-components // Automation in industry, No. 5. 2021. pp.12-32