УДК 004.891.3
Малаховський О.Ю. –ст. гр. СНс-61
Тернопільський національний технічний університет імені Івана Пулюя

ЗБІР ТА АНАЛІЗ ЛОГІВ В КУБЕРНЕТІС

Науковий керівник к.т.н доц. Баран І.О.

Malakhovskyi O.Y.
Ternopil Ivan Puluj National Technical University

COLLECTING AND ANALYSING LOGS ON KUBERNETES

Supervisor: Ph.D., Assoc. Prof. Baran O.I.
Keywords: elasticsearch, kubernetes, logs

Logs are a critical part of any software development process, they give a deep insight about an application and application lifecycle, what happens in a system and what exactly caused the error when something incorrect happens. A centralized log management and analysis strategy are critical mission for each system, enabling organizations to understand the complex relationship between operational, security, and application events. Building enterprise level application, a system goes to multiple hosts and many servers, managing the logs across that system can be complicated and slow. Furthermore debugging the error in the application across thousands of log files on hundreds of servers can be very complicated and sometimes time-consuming.

The first stage of solving the problem of centralized logging in a Kubernetes environment is a logs collection from different sources. Containers and container management platforms like Openshift produce logs in different ways, for example through syslogs and other logs directly in files. Log files can also be different formats, like JSON or plain text.

The second stage is processing and filtering collected data. Log files can consist of thousands line of text and numbers, furthermore many fields are not important for logging system. Logs in text format are unindexed and sometimes don't have any structure. In contradistinction to text logs, logs in JSON format have more features. For example, all fields are already in "key=value" format, that means we can quickly find and select the requested field. Almost every system and programming language support JSON natively. Usually, a log entry includes such information as: the date and time the event occurred, the container the event occurred on. There are many ways available to transport log data. One way is directly plug input sources and framework can start collecting logs and another way is to send log data via REST API, application code is written to log directly to these sources it reduces latency and improves reliability. A common approach in container world is a syslog driver for transporting. Syslog is a standard driver for message logging in Docker, that means a container can send logs per TCP or UDP independently of any frameworks or log rotation tools. Logging system should filter and parse data, after that find and save a required data. Log analytics occurs by organizing data via processing text data, tagging and storing as indexed text. There are many different search and analytics engines. One of the most popular is elasticsearch. Elasticsearch is a highly scalable free and open-source full-text search and analytics engine.

Typically all application and system logs are collected in centralized logging systems and exposed via APIs and Web UIs. These systems are deployed in clusters and receive logs from either the applications themselves or agent processes running on the host. Centralized logging tools are responsible for parsing, indexing, and analyzing log data to produce on demand insights for their consumers. These insights can range from providing search for error debugging to building reports tracking monthly business performance metrics.

With the rise in popularity of containers, companies are looking at how they can migrate their existing workflows into Docker and onto Kubernetes. Part of that migration is preserving the existing frameworks in place for running software including log collection and analytics. We are going to walk through a few examples of how you can collect and ship your Docker Container logs running on Kubernetes to centralized logging solution.