Curating Log files

GOPALIKA SHARMA

PAARTH KOTAK

PRACHI VED

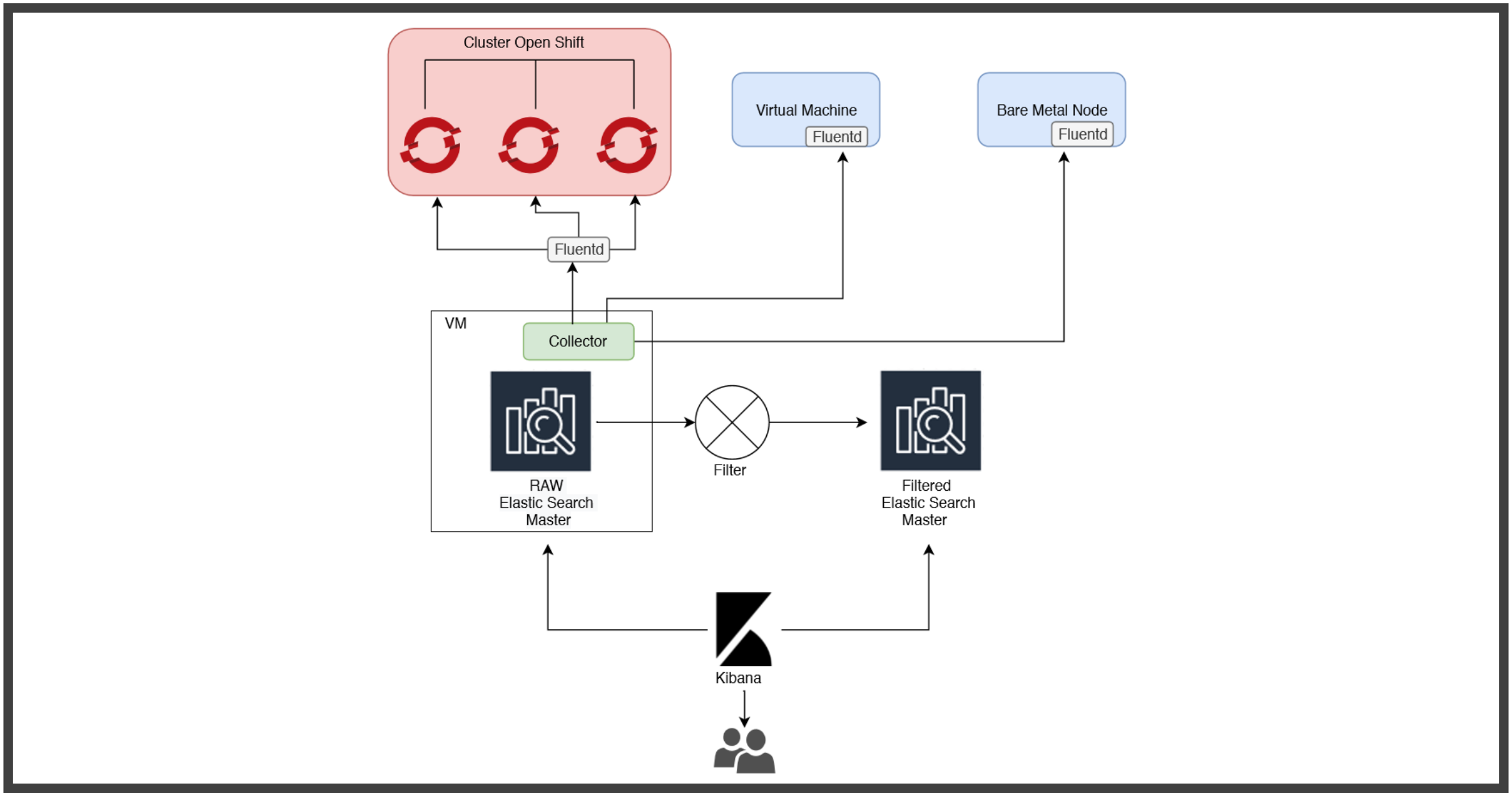
RASHI CHAUHAN

Purpose of our project

To build a system which will start to capture the system logs generated at MOC, automate the process of anonymizing logs and organize logs by a methodology which will help to see if there is a pattern such as demands on different services and how they are correlated.

Technology Overview

1. **OpenShift:** Enterprise-ready Kubernetes container platform with Full–stack automated operations to manage hybrid cloud and multi-cloud deployments.
2. **Elasticsearch:** Search engine which provides a distributed multi-tenant capable full text search engine.
3. **Kibana :** Open source data visualization plugin for Elasticsearch.


Cluster Open Shift
Virtual Machine
Fluentd
Bare Metal Node
Fluentd
Fluentd
VM
Collector
RAW
Elastic Search
Master
Filter
Filtered
Elastic Search
Master
Kibana

Component Description

1. **Sources:** (OpenShift Cluster, VMs and Bare metal nodes) - Fluentd pods consisting of multiple containers will collect all the logs pertaining to that pod and store it in a file named fluentd.log.
2. **Collector:** Script to collect logs from all fluentd.log on all sources and push them in ElasticSearch index.
3. **RAW ElasticSearch Master:** Index containing all raw log files.
4. **Filter Script:** Filter the raw log files using Machine Learning algorithms (After analyzing PII we will decide on a feasible algorithm).
5. **Filtered Elastic Search Master:** Index containing all curated log files.
6. **Kibana:**View the logs.

Lets have a look at the logs..