THOMAS MORE

Cloud & IoT Monitoring and Data Analysis with Elasticsearch and Grafana

Realization document

Birate Kabanza Arthur
Student Bachelor Applied Computer Science

2024 - 2025

Table of Contents

1.Introduction

This document describes the realization phase of my internship at Van Genechten Packaging (VGP), which I completed as part of my bachelor's program in Applied Computer Science at Thomas More during the 2024-2025 academic year.

It expands on the previously outlined project plan, which focused on creating a Cloud and IoT Monitoring and Data Analytics Framework. The primary goal was to improve system observability, assure business continuity, and enable data-driven decision-making using tools like Grafana, Elasticsearch, and Kibana.

This document is structured as follows:

Introduction and Internship Context

An overview of Van Genechten Packaging, the internship's background and the importance of the cloud & and IoT monitoring and data analysis framework within the company.

Tools and Technologies

Description of the main tools and technologies used during the internship, such as Elasticsearch, Grafana, Kibana and Logstash.

Analysis

Explanation of the preliminary analysis conducted by the company on possible monitoring and analytics tools, highlighting the reasons for selecting the current stack over alternatives like Power BI.

Implementation and Realization

Describes the practical steps taken to implement the cloud and IoT monitoring and data analysis framework at Van Genechten Packaging.

Data Flow

Overview of how data moves through the system from IoT devices and infrastructure components through Kafka to Elasticsearch and visualization layers.

Dashboards Overview

Presentation of the key dashboards created in Grafana and Kibana, their purposes, features and how they contributed to monitoring and business insights.

Results and Benefits

Summary of the improvement and advantages brought by the monitoring framework including increased visibility, proactive reporting and business continuity.

2. Internship Context

Van Genechten Packaging (VGP) is a leading European packaging company specializing in innovative, sustainable, and high-quality packaging solutions. As part of its digital transformation, VGP has invested in modernizing its IT infrastructure by adopting a private cloud environment built on OpenShift and Open Kubernetes Distribution, This Cloud native platform enables the development of deployment of containerized applications at scale.

In this context, my internship was situated within the IT and data infrastructure team, where the main challenge was to ensure business continuity by establishing a comprehensive monitoring and reporting framework for both cloud infrastructure and IoT systems. The goal was to ensure that all components across the technology stack from hardware to microservices were continuously monitored to prevent downtime, detect anomalies and improve operational efficiency.

To achieve this, various technologies were applied:

Elasticsearch and **Kafka** were used to collect and store monitoring data.

Grafana and **Kibana** were used to create dashboards for visualizing metrics and logs.

Reporting mechanisms were configured in Elasticsearch to notify developers and engineers of activities and system performance.

Elasticsearch's machine learning capabilities to detect anomalies and identify trends within operational data. This allowed a deeper understanding of the system's behavior.

In addition to infrastructure monitoring, the internship also involved setting up monitoring for application and IoT data.

A significant part of the assignment also included hands on experience with Logstash is part of the ELK stack a powerful tool for building real time data pipelines. This allowed scalable ingestion and monitoring of application and IoT data.

This context formed the foundation for the practical implementation of dashboards, reporting systems and analytical tools which are detailed in the following sections of this document.

3.Tools and Technologies

During the internship, a variety of tools and technologies were utilized to build an effective cloud and IoT monitoring and data analysis framework. These technologies were chosen to ensure scalability, real time data processing and visualization enabling quick insights.

Key Tools Used

Elasticsearch:

A distributed search and analytics engine used for storing, searching and analyzing large volumes of monitoring and application data in near real time.

Kibana

A data visualization platform that enables interactive exploration and dashboard creation for data stored in Elasticsearch.

Logstash

A data processing pipeline that ingests, transforms and forwards log and event data to Elasticsearch for indexing and analysis.

Grafana

A visualization platform used to create interactive and customizable dashboards, providing clear insights into cloud infrastructure and IoT data.

Kafka

A distributed event streaming platform that enables the ingestion and processing of real time data pipelines facilitating efficient data flow between systems.

OpenShift / Open Kubernetes Distribution

The private cloud platform used for deploying and managing containerized applications and microservices within VGP's infrastructure.

4. Analysis

Van Genechten Packaging had performed a comprehensive evaluation of different monitoring and analytics tools to determine the best fit for their cloud and IoT monitoring framework. The decision to use Elasticsearch, Kibana and Grafana over alternatives such as Power BI was based on the following key factors:

- Superior support for real time monitoring and reporting, crucial for maintaining business continuity of their cloud infrastructure and microservices.
- Seamless integration with their private cloud environment (OpenShift/Kubernetes) supporting metrics collection from containers, VMs and IoT devices.
- High scalability and flexibility to handle a large volumes of time series and log data across multiple architectural layers
- The open-source nature and ability to deploy on-premises, ensuring full control over data security and infrastructure.
- Advanced customization and visualization capabilities that provide actionable insights through tailored dashboards

Tool Comparison Summary

Feature / Tool	Elasticsearch + Kibana/Grafana	Power BI
Real time monitoring	Excellent, designed for streaming data, alerting and reporting.	Limited real time support
Integration with cloud and IoT	Native and seamless, supporting container and VM metrics	Limited integration more focused on business data sources
Scalability	Highly scalable for big data and logs	Scalable but with higher costs and infrastructure needs
Alerting and Anomaly detection	Built-in flexible alerting and anomaly detection	Available but less tailored for infrastructure monitoring
Deployment	On-premises or cloud, full control over environment	Primarily cloud-based, limited on-premises options.
Cost	Mostly open-source, cost-effective	Licensing costs can be high at scale.
Dashboard Customization	Highly customizable and extensible	User-friendly but less flexible

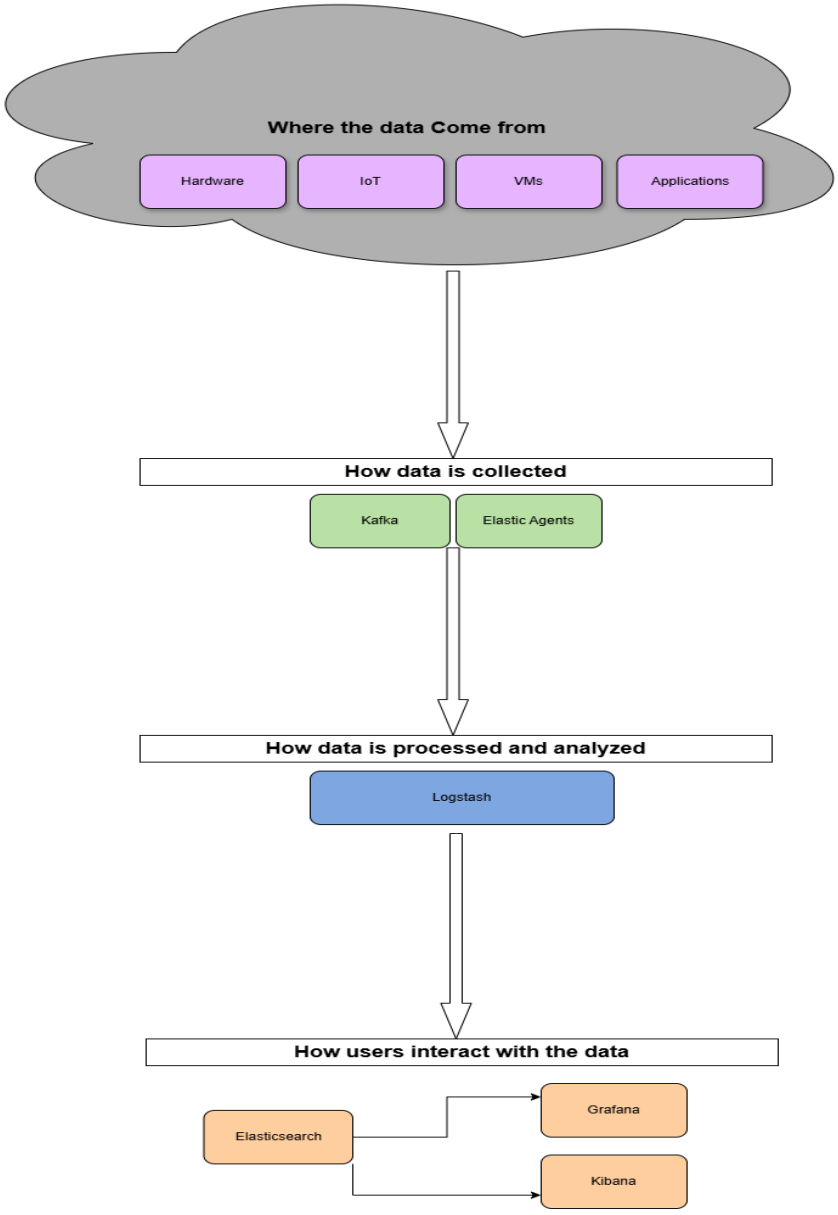
This evaluation led to the adoption of Elasticsearch, Kibana and Grafana as the core tools for VGP's monitoring and data analysis needs, providing a powerful, scalable and secure platform for operational insight.

5.Implementation and Realization

This section describes the practical steps taken to implement the cloud and IoT monitoring and data analysis framework at Van Genechten Packaging. It details how the chosen tools and technologies were applied to build an effective monitoring system that meets the company's business continuity and data insight goals.

Data Flow

This section explains how data moves through the monitoring system and how it's processed and used.



Data sources:

Monitoring data is collected from various layers of the infrastructure, including hardware devices, IoT sensors, Virtual Machines and applications.

Data ingestion:

Elasticsearch agents continuously capture metrics and logs, which are streamed in real time through Kafka serving as a robust message broker that ensures reliable, scalable and fault tolerant data transportation.

Data processing and transformation

A Logstash pipeline is used to collect and transform raw data from various sources such as Kafka topics or Elastic Agents before forwarding to Elasticsearch. The pipeline enables advanced filtering ensuring incoming data is properly structured.

Data Storage and indexing

The data is ingested into Elasticsearch where it is indexed and stored efficiently to enable faster querying and analysis.

Visualization

Processed data is visualized on customizable dashboards in Grafana and Kibana, providing actionable insights through real time graphs and charts.

Setting Up the Monitoring Infrastructure

This section focuses on the practical implementation steps and technologies configured.

Tool deployment

Tools such as Elasticsearch, Grafana, Kibana and Logstash were installed and configured in a secure containerized cloud environment.

Agent setup

Elastic Agents were configured across the cloud stack to gather metrics from VMs, physical hardware and IoT devices.

Data Pipeline Integration

A Kafka pipeline was implemented to handle high throughput, real time data ingestion and ensure message delivery to Logstash and Elasticsearch.

Dashboard design:

Dashboard were built in Grafana and Kibana, designed for technical and operational users improving visibility inro system status and supporting decision-making.

6. Results and Benefits

The implementation of the cloud and IoT monitoring and data analysis framework at van Genechten Packaging delivered significant improvements and value across multiple areas:

Improved Operational Visibility:

Real time monitoring of hardware, storage, IoT devices and applications provides a comprehensive overview of the entire cloud infrastructure.

Proactive Issue Detection:

Early detection of anomalies and system fault via Elasticsearch machine learning reduces downtime and supports faster incident response ensuring business continuity.

Scalable and Flexible Monitoring Platform:

The use of Logstash for data pipelines and Elasticsearch for storage ensures the system can grow with increasing data volume without compromising performance.

Improved collaboration:

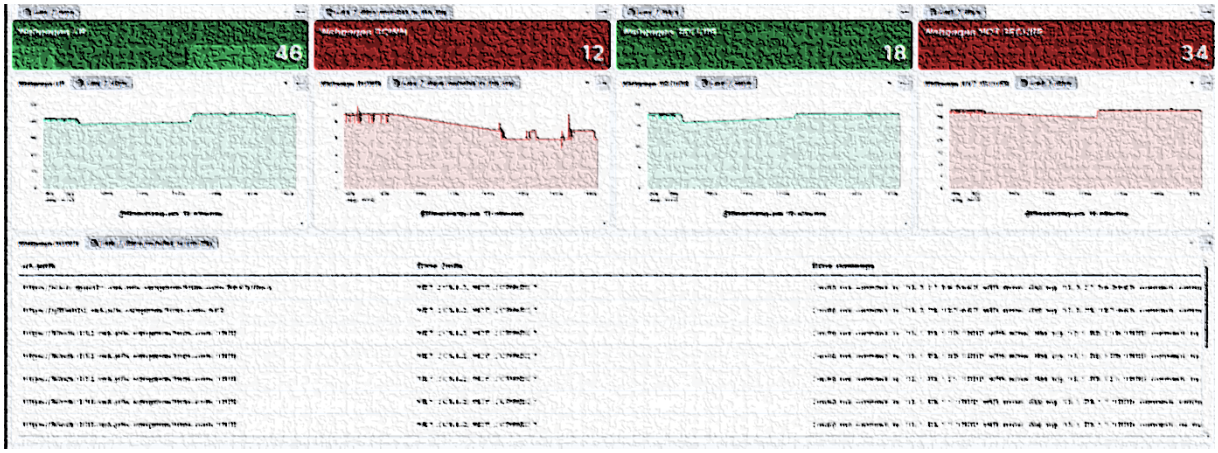
Centralized dashboards and reports improve communication and coordination among IT, operations and management teams.

7.Visualization Layer: Grafana & Kibana Dashboards

Several dashboards using Grafana and Kibana were developed to support the monitoring and analysis of cloud infrastructure and IoT devices. These dashboards provide valuable insights into system health, performance and security.

Kibana Dashboards

1.Webpages status Dashboard



This dashboard shows a real time overview of the health and security status of web pages in infrastructure. It includes:

- Total counts of web pages categorized by security level (secure via HTTPS and Insecure Via HTTP)
- Availability status, indicating how many pages are currently Up and Down
- Error diagnostics, listing the HTTP error type and associated error messages for webpages that are down.

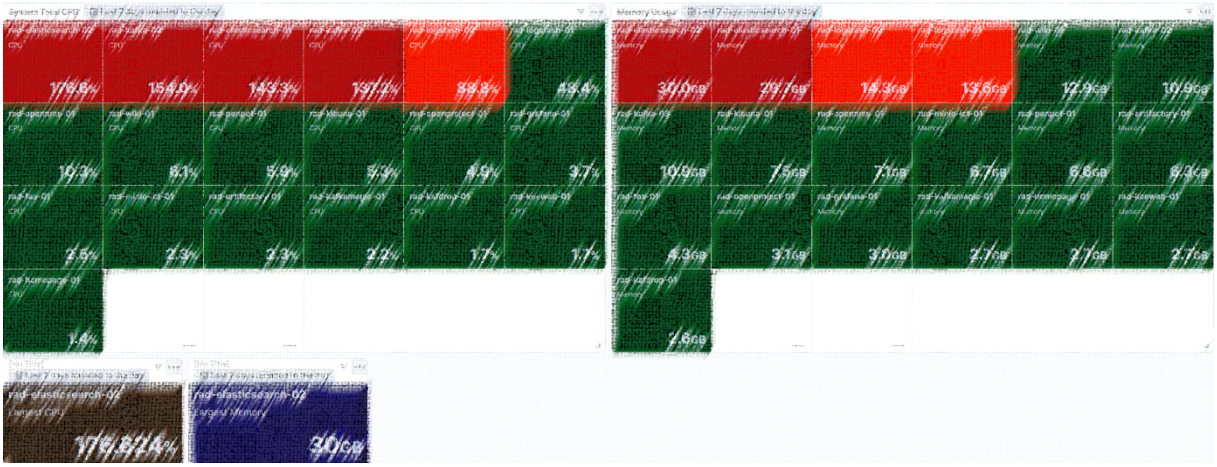
This visualization allows teams to monitor page reliability, identify insecure endpoints and act quickly on web related incidents.

2. Backup Monitoring Dashboard

This dashboard tracks the status and performance of system backups it includes:

- Total number of backups performed over the last 7 days and last 24hours
- Breakdown of backup outcomes: number of successful and failed backups
- Highlight of the largest backup in terms of memory usage and longest completion time
- Identification of the smallest backup with the shortest completion time

3. Virtual Machines CPU and Memory usage Monitoring dashboard

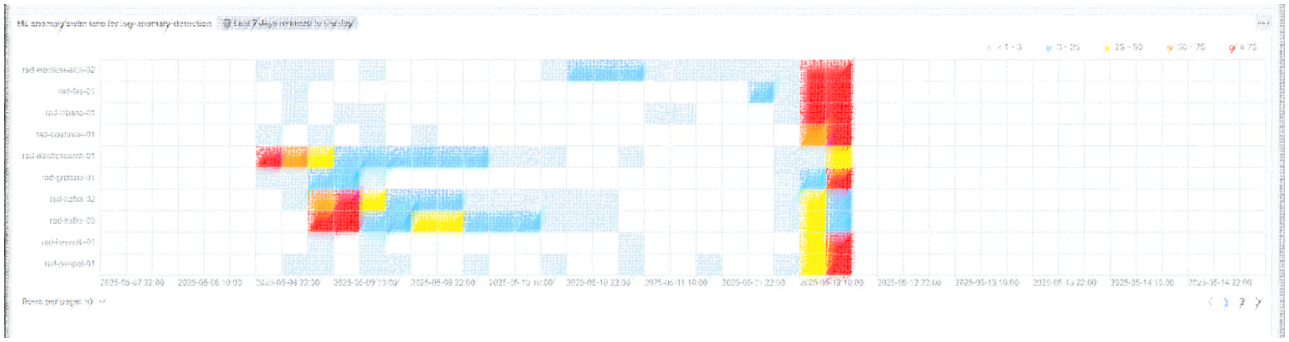


This dashboard provides detailed insights into the resource utilization of virtual machines in the environment it includes:

- Real time CPU and memory usage across all monitored VMs
- Highlight of the VM with the highest CPU consumption
- Highlight of the VM with the highest memory consumption

This dashboard enables proactive resource management, helps detect performance bottlenecks and support infrastructure scaling decisions.

4. Log Anomaly Detection Dashboard



This dashboard leverages machine learning in Elasticsearch to identify unusual patterns and potential issues in log data it includes:

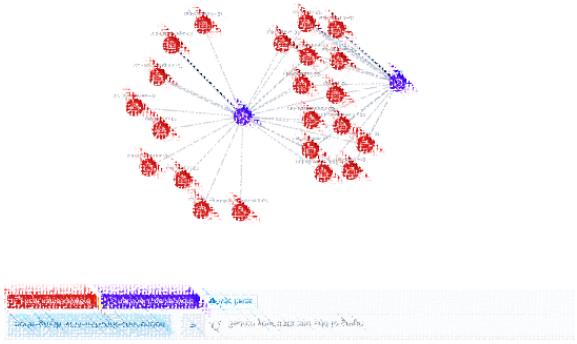
- Real time detection of anomalies based on log frequency and message content
- Highlight of unusual events such as unexpected spikes in errors or rare log entries
- Drill down capabilities to inspect affected services, timestamps and log details.

This dashboard enhances proactive monitoring by surfacing hidden issues early supporting faster incident response and root cause analysis.

5. Cloud infrastructure Node graph

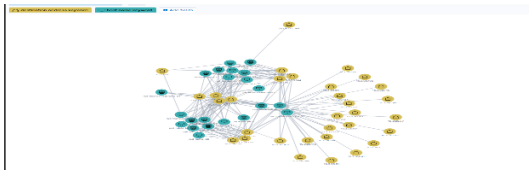
This node graph visualizes the interconnections within the cloud infrastructure helping to understand dependencies and improve observability. It includes the following relationships:

Virtual Machines ⟶ Network Type (IPV4 or IPV6)



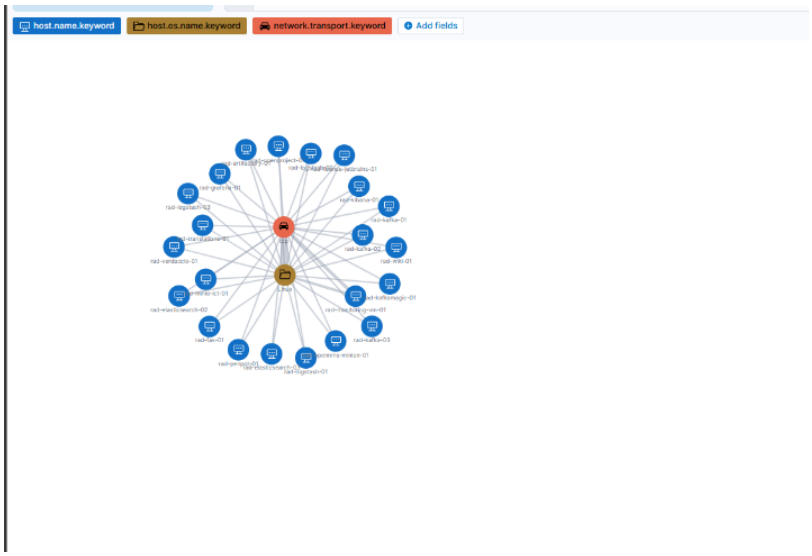
Displays which virtual machines are using IPV4 , IPV6 or both aiding in network configuration and transition planning

Virtual Machine ⟶ IP Address



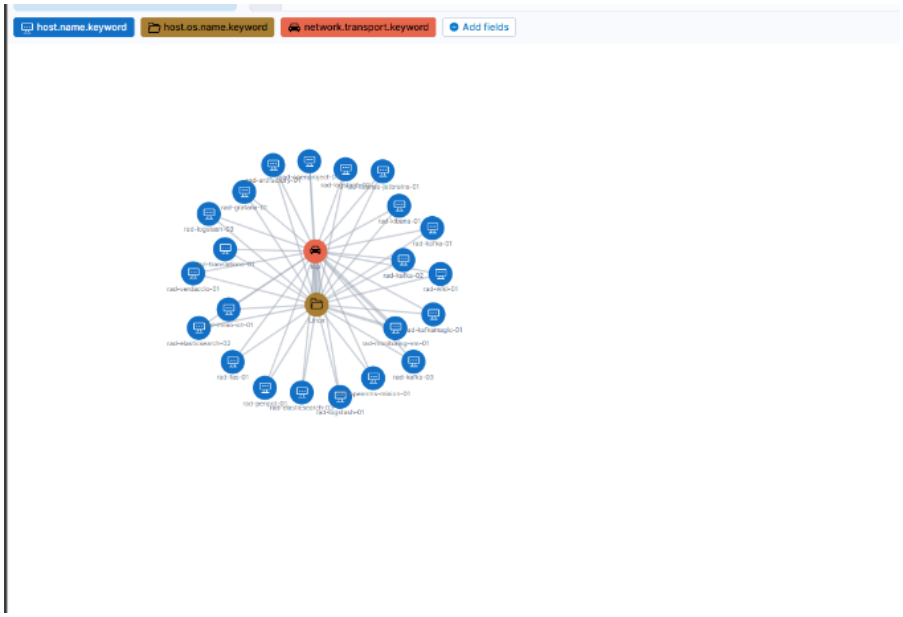
Shows each virtual machine's assigned IP address, helping track connectivity issues and resource allocation.

Virtual Machine → Operating System



Links Virtual machines to their operating system (Linux, Windows) supporting patch management and compliance checks.

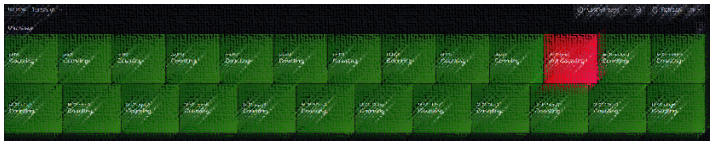
Virtual Machine → Network transport Layer

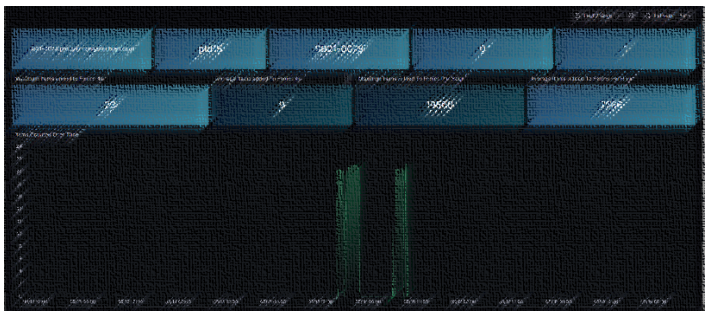


Visualizes the type of network transport protocol used by the Virtual Machine (TCP, UDP) enabling deeper understanding of communication patterns and potential security gaps.

Grafana Dashboards

1.Smart box Counter Monitoring (For Developers)



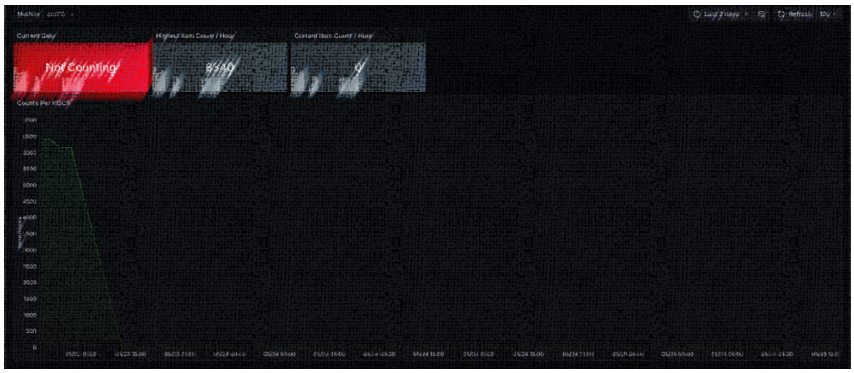


This dashboard is designed to provide developers with real time visibility into operational state of the counting machines it includes:

- A status panel that visually displays machines in green (actively counting) and red (not counting/ inactive)
- Interactive navigation allowing users to click on a machine to access detailed information.
- Detailed views show item count trends over time helping developers monitor machine performance identify downtime and anomalies in counting behaviour

This dashboard supports effective monitoring, debugging, and maintenance of smart box counters in production environments.

2. Smart counter Monitoring dashboard (For Operators)

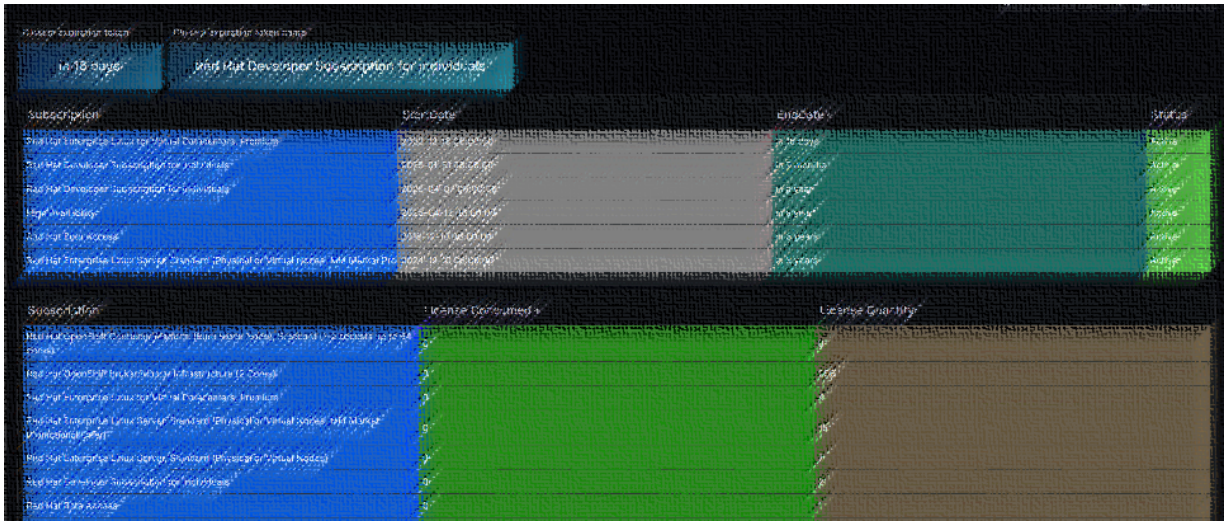


This dashboard is designed for production operators to monitor real time performance it includes:

- A simple time-based visualization showing the number of the number of items counted per hour
- Enables quick checks on production flow consistency without technical details
- Promotes fast reaction to drops or irregularities in item counting.

This dashboard empowers operators to maintain smooth operations and quickly detect production issues.

3. Red Hat License Monitoring Dashboard

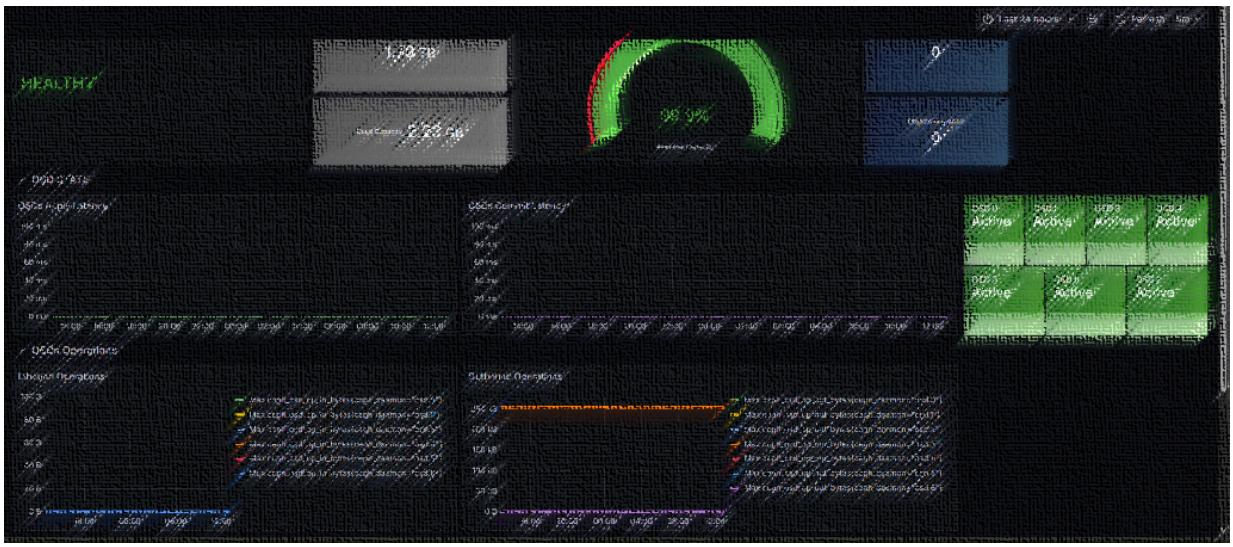


This dashboard provides a clear overview of Red Hat subscription usage and expiry status it includes:

- Expiration tracking, listing all the licenses with their respective expiration dates
- Highlight of the subscription closest to expiration enabling proactive renewal planning.

This dashboard prevents service disruption and supports efficient license management.

4, Ceph Cluster Monitoring Dashboard



This dashboard offers a depth visibility into health and performance of the Ceph storage it includes:

- OSD (Object Storage Daemon) metrics such as apply and commit latency
- Inbound and Outbound I/O operations, helping cluster activity
- OSD status overview (active / inactive) and overall cluster health status

This dashboard helps ensure storage reliability, identify latency issues, and maintain optimal performance of the Ceph cluster.

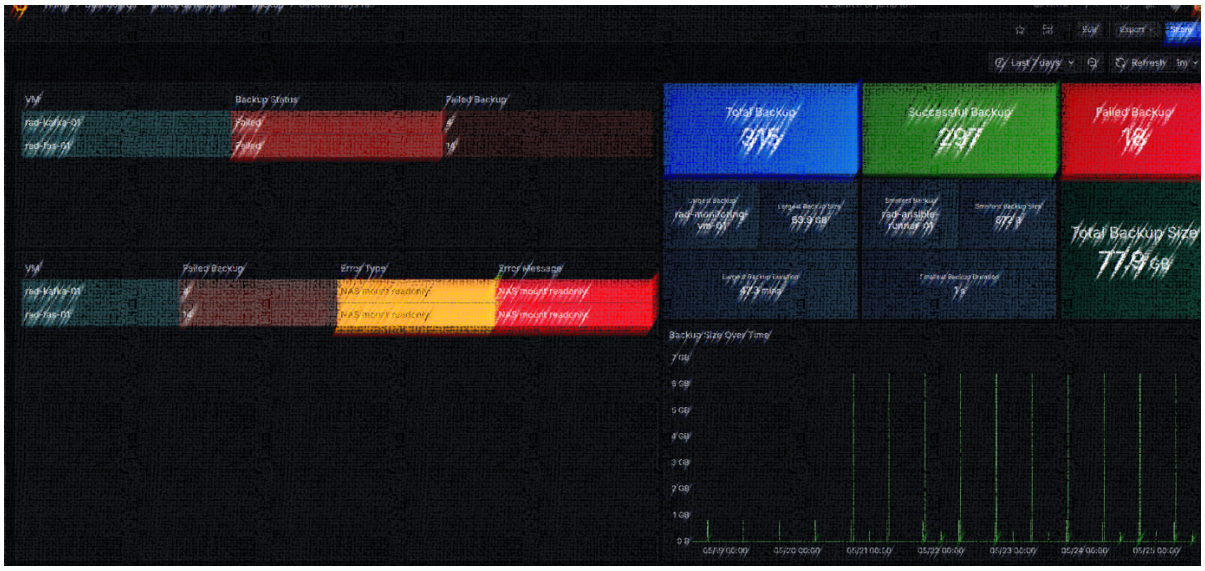
5. IDRAC Servers Monitoring Dashboard (Incomplete)

This dashboard wad intended to monitor critical server hardware metrics through IDRAC (Integrated Dell Remote Access Controller). The goals included:

- Visualizing CPU, memory, disk and power usage of physical servers
- Tracking hardware health and performance trends.

However, due to data ingestion issues, the required metrics were not successfully captured and the dashboard remained incomplete. This highlighted the importance of ensuring data source integration and validation in early project stages.

6, Backup Monitoring Dashboard



This dashboard provides an overview of system backup performance and reliability it includes:

- Total number of backups with counts of successful and failed operations
- Identification of the largest backup (by memory) and the one the one with the longest duration
- Highlight of the smallest backup and the one with the shortest duration

7. Kubernetes Cluster Monitoring Dashboard

This dashboard provides real time monitoring of the Kubernetes cluster, organized by namespace. Key features include:

- Visualization of CPU, memory (RAM) and pod memory usage per namespace
- Interactive panels that allow users click on a specific pod to navigate to a detailed dashboard for in depth analysis
- Helps identify resource bottlenecks, monitor workload distribution and maintain cluster health.

This dashboard improves observability across the Kubernetes environment and supports efficient troubleshooting and scaling decisions.

8. Conclusion

The internship project at Van Genechten Packaging provided a comprehensive opportunity to apply data analytics and monitoring techniques in real world cloud infrastructure, Through the setup and implementation of monitoring tools like Grafana, Kibana, Elasticsearch and Logstash. We achieved the primary goal of improving observability, reporting and actionable insights across various layers of IT landscape, including IoT systems and business critical applications.

Key accomplishments included the creation of dashboards for infrastructure health, backup monitoring, Kubernetes performance and anomaly detection. Additionally, we explored the potential of machine learning in Elasticsearch to improve capabilities to detect abnormal behaviors in logs.

Evaluating against the initial project objectives, the results show a significant improvement in business continuity, data transparency and incident responsiveness.

Looking ahead, I recommend:

- Extending data ingestion to currently incomplete dashboards (e.g., IDRAC server monitoring)
- Scaling anomaly detection with more training data
- Expanding the use of machine learning for predictive maintenance

The internship has not only contributed to VGP's IT monitoring but has also strengthened my skills in data engineering and operational analytics paying the way to future growth in this domain.

Reference list

- Andidog. (2022, APRIL 21). Grafana Dashboards Best Practices – Dashboards as Code. Retrieved from https://andidog.de/blog/2022-04-21-grafana-dashboards-best-practices-dashboards-as-code
- Grafana labs. IDRAC Hosts Stats Dashboards – Grafana. Retrieved from https://grafana.com/grafana/dashboards/12106-idrac-host-stats/
- Toxigon, Integrating Kibana With Grafana. Retrieved from https://toxigon.com/integrating-kibana-with-grafana
- Udemy. AI + ML Search with Opensearch. Retrieved from https://www.udemy.com/course/ai-ml-search-with-opensearch/