Network Data Monitoring Using Third Party Services and Migration of Services to Local Observability Platforms

CS4991 Capstone Report, 2024

Saurav Chanda
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
ufx2ju@virginia.edu

ABSTRACT

My task during my internship with GEICO was migrating its network management devices to Splunk, but I had no way to know what portion of the devices had already been successfully moved. The way to solve this problem was to display a gauge percentage which would compare the number of present devices in Splunk to the ones that already existed in the third-party services, Kentik and Netbrains. This required my team to collect metadata from all service providers about our network devices and conduct rigorous validation to ensure that the correct devices had been moved. After the project was completed, we determined that only about 40 percent of the migration had been completed, but this metric provided lots of insight into additional work needed. From this point in the pipeline, the devices are manually being sent into Splunk, so future work on this project would expand to automate the process and validation of network management.

1. INTRODUCTION

Geico, one of the largest auto insurance companies in the United States, relies on a vast network infrastructure to support its operations and serve its customers effectively. Ensuring the smooth functioning and optimal performance of this network is critical to the company's success. Network monitoring plays a vital role in maintaining the health and reliability of Geico's network, which enables the IT team to identify and resolve issues before they impact business operations.

As part of Geico's ongoing efforts to enhance its network monitoring capabilities, the company initiated a project to migrate its network management devices to Splunk, a powerful data analysis and monitoring platform. This migration aimed to centralize network monitoring, provide better visibility into network performance, and enable more efficient troubleshooting. However, the migration process posed several challenges, including the need to track the progress of the migration, validate the successful transfer of devices, and ensure data consistency across multiple network monitoring tools.

2. RELATED WORKS

The Geico migration project drew inspiration from various studies and industry practices in the field of network monitoring and data analysis.

One influential work is the study by Sandeep, et al. (2018), which explores the use of Splunk for real-time monitoring and analysis of network data. The authors highlight the benefits of using Splunk, such as its ability to ingest and process large volumes of data, its powerful search and visualization capabilities, and its flexibility in integrating with various data sources. They demonstrate how Splunk can be leveraged to identify network anomalies, detect security threats, and optimize network performance. This study provided valuable insights into the potential of Splunk for network monitoring and informed the decision to migrate Geico's network management devices to this platform.

Another relevant work is the research conducted by Liu, et al. (2019), which focuses on the automation of network monitoring and management tasks. The authors propose a framework that combines machine learning techniques with network monitoring tools to enable proactive identification and resolution of network issues. They emphasize the importance of automation in reducing manual effort, improving efficiency, and minimizing downtime. This research aligns with the goals of automating the monitoring process within Geico's existing repositories and pipelines.

3. PROJECT DESIGN

The project aimed to develop a system for monitoring the migration of Geico's network devices to Splunk, a data analysis and monitoring platform. The primary goal was to quantify the progress of the migration and provide a visual representation of this progress on a dashboard.

The first step we conducted was gathering data on the existing network devices from two key sources: Kentik, a network analytics platform, and Netbrains, a network automation tool. Device lists were pulled from both platforms, including crucial information such as IP addresses and management interface details.

To determine which devices had been migrated, a comparison was made between the device lists from Kentik and Netbrains. This comparison primarily relied on matching IP addresses. However, complications arose due to the fact that the same device could have different IP addresses on different platforms, depending on the network configuration. To handle this, the management interfaces were used as a secondary point of comparison, as these tend to remain constant regardless of IP address changes.

A series of checks were implemented to categorize each device:

- If the IP addresses matched between Kentik and Netbrains, the device was considered migrated.
- If the IP addresses did not match, the management interfaces were compared to attempt to identify the device.
- If neither the IP addresses nor the management interfaces matched, the device was marked as not yet migrated to Splunk.

For the second case, where IP addresses differed but the device needed to be identified, we developed a custom Python script. This script analyzed various other device details from the Kentik and Netbrains lists to determine if they represented the same physical device.

With these categorizations in place, it became possible to quantify the overall progress of the migration. The next challenge was to display this progress in a user-friendly way. Grafana, an open-source platform for data visualization, was chosen for this purpose.

However, Grafana requires a data source to populate its visualizations, and the necessary data source didn't exist yet. To solve this, custom log exporters were created using

Python scripts. These scripts handled the device comparisons, calculated the migration percentage, and sent this data to the log exporters.

Prometheus, an open-source systems monitoring and alerting toolkit, was employed as the data source for both Grafana and the log exporters. Prometheus provides a powerful query language called PromQL, which was used to scrape the relevant data from the logs and feed it into a gauge visualization on the Grafana dashboard, allowing employees to see a tangible piece of information which would assist in their work

The final step was to automate this process. To ensure the migration progress was regularly updated without continuous manual intervention, the entire tool chain was packaged into a Docker container. Docker allows the application to run in a consistent environment, independent of the underlying infrastructure. We then setup a cron job to run this Docker container every day at 7 AM Pacific Time, allowing for real-time updates.

In summary, the project involved data gathering from Kentik and Netbrains, data comparison and categorization using Python scripts, progress quantification, data visualization with Grafana and Prometheus, and automation using Docker and cron. The result was a robust, automated system for monitoring the progress of Geico's network device migration to Splunk.

4. RESULTS

The immediate outcome of this project was a functional, automated tool for monitoring the progress of Geico's network device migration to Splunk. The tool provided real-time, quantitative insights into the status of the migration, enabling the network team to track their progress, identify issues, and optimize their approach. The tool's dashboard, built using Grafana, offered a clear, visual representation of the migration's progress. This allowed important stakeholders across the organization, from the network team to senior leadership, to easily understand and track the status of this crucial migration.

By automating the data collection and analysis process using Python scripts, Prometheus, and Docker, the tool ensured that the migration progress was consistently and accurately reported. This eliminated the potential for manual errors and provided a reliable, always-up-to-date view of the migration's status. The project's approach—using data from Kentik and Netbrains, processing it with custom Python scripts, and visualizing it with Grafana—could serve as a model for similar monitoring initiatives within Geico. The architecture is flexible and scalable, allowing it to be adapted to monitor other technical processes or systems.

5. CONCLUSION

This project addressed a critical need within Geico: the ability to monitor and track the progress of a complex, large-scale network migration. By providing clear, real-time insights into the migration's status, the tool empowered the network team to manage the process more effectively and efficiently. The project's key features, including automated data collection, processing, and visualization, ensured that the tool provided reliable, up-to-date information. The use of industry-standard tools like Grafana, Prometheus, and Docker made the system robust, scalable, and maintainable.

The tool's benefits extend beyond the immediate context of the Splunk migration. It has demonstrated the importance of data-driven, automated monitoring in managing complex technical processes. This approach can help Geico to proactively identify and address issues, optimize resource allocation, and make informed, data-backed decisions, ultimately helping the overall goal of becoming a more efficient tech company. For Geico's tech teams and leadership, this project has provided valuable insights into methods to leverage data and automation to manage large-scale technical initiatives. It has shown the importance of having real-time, quantitative insights into critical processes, and how these insights can drive more effective decision-making and problem-solving.

Moving forward, the knowledge gained from this project can be easily utilized as Geico continues to grow its technical capabilities. The project has laid the foundation for a new approach to tech management within the organization. A more proactive, data oriented, and automated company is the ultimate goal. As Geico comes across future technical challenges, the processes of this project can serve as a guide.

6. FUTURE WORK

The successful implementation of this network migration monitoring tool opens up a range of possibilities for future enhancements and applications. One immediate next step would be to expand the tool's capabilities to not just monitor, but actively manage the migration process. This could involve integrating the tool with the systems responsible for executing the migration, allowing it to automatically trigger device transfers based on predefined criteria or schedules. Such an expansion would further streamline the migration process and reduce the need for manual interventions.

Another potential avenue for future work is to generalize the tool's architecture to monitor and manage other types of large-scale technical initiatives beyond network migrations. The core components of the tool—data ingestion, processing, visualization, and automation—are highly adaptable and could be applied to a wide range of use cases. For example, the tool could be modified to monitor the rollout of new software across the organization, tracking the progress of installations and identifying any compatibility issues. It could also be used to monitor the performance of critical business applications, providing real-time insights into key metrics and alerting relevant teams when issues arise. By extending the tool's capabilities and scope, Geico could develop a centralized, data-driven platform for managing its diverse technical operations.

REFERENCES

Sandeep, K., Goyal, A., & Sharma, S. (2018). Real-time network monitoring and analysis using Splunk. Proceedings of the International Conference on Advanced Computing and Communication Systems (ICACCS), 516-520. https://doi.org/10.1109/ICACCS.2018.8441741

Liu, Y., Zhao, H., & Zhang, W. (2019). An automated framework for network monitoring and management using machine learning. Journal of Network and Systems Management, 27(3), 701-720. https://doi.org/10.1007/s10922-018-9482-x