WHITEPAPER

SPLUNK OPTIMIZATION

Optimizing Splunk Workload Pricing

Strategies to unlock efficiency & cost savings using a security and telemetry data pipeline

DATABAHN

Introduction

In log analytics, whether security or observability, understand and optimizing your Splunk workload pricing can lead to significant cost savings and improved performance. The Splunk Workload Pricing Model, a cornerstone for managing your Splunk expenses, is driven by two primary components: Splunk Virtual Compute (SVC) and Storage Blocks.

The model was announced in 2021 as an alternative to volume-based pricing. It was an attempt to give their customer more flexibility. However, this flexibility didn't translate into immediate or obvious cost-savings – or even allow for greater ingestion of data.

In this paper, we will discuss strategies to optimize your usage and lower your costs based on your usage patterns, as well as the use of a Security Data Pipeline platform to optimize ingestion.

SVC and Storage Blocks

Splunk Virtual Compute (SVC)

DEFINITION

SVC encompasses the compute, I/O, and memory resources required to monitor and analyze your data sources in Splunk. It is analogous to virtual CPUs (cVPUs) in cloud environments, serving as the measure of computational power used in Splunk Cloud.

IMPACT FACTORS

The primary factors influencing SVC consumption include the volume of data ingested, the frequency and complexity of ad hoc queries, and the scheduling of queries to populate dashboards and trigger alerts. High volumes of data, complex queries, and frequent dashboard updates can lead to increased SVC requirements.

Storage Blocks

DEFINITION

Storage Blocks refer to the amount of storage needed to adhere to your data retention policies within Splunk. This involves maintaining and storing historical data, determined by organizational policies and compliance needs.

IMPACT FACTORS

Storage Blocks are essential for ensuring that your data is available for historical analysis and compliance, impacting the overall cost of data retention.

Navigating SVC usage and Managing Costs

When you engage with Splunk, the team assigns a set amount of SVC unit credits based on the aggregate data sources you use and your data interaction patterns. These evaluations create a segregation between two different types of datasets depending on how you use them:

- **High Ingestion-to-SVC Ratio:** Datasets that are infrequently queried, such as compliance storage or data lakes, tend to have higher ingestion-to-SVC ration. These datasets are ingested into Splunk, requiring relatively less compute power for their analysis.

- **Low Ingestion-to-SVC Ratio:** Datasets monitored with scheduled queries and real-time dashboards, such as security or service monitoring data, have a lower ingestion-to-SVC ratio. These datasets require more compute resources for frequent and real-time processing.

If your organization exhausts its allotted SVC credits, you will encounter overages. This can lead to increased and unexpected costs, as well as operational challenges including the inability to initiate new queries until existing ones are completed.

Lowering Ingestion-to-SVC

Streamlining Data Processing

DataBahn's Data Fabric enables you to meaningfully reduce the volume ingested within Splunk through its native volume reduction and data transformation frameworks. In addition to this, the platform natively offers volume reduction content and AI-generated recommendations and insights to continually optimize the data volume sent to Splunk. By performing real-time data analysis, DataBahn uncovers valuable insights, statistics, and aggregates that are then streamed to Splunk. By reducing the amount of raw data ingested, DataBahn lowers the compute requirements (SVC) and enhances query performance. For example, raw data queries that previously took five minutes can be optimized to complete in just a matter of a few seconds.

Generating Summary Insights

DataBahn's Summary Insights Framework provides a powerful tool for optimizing data management and reducing computational overhead. This framework enables the extraction of essential metadata from each log source, with the added capability to extend and aggregate this metadata across multiple log sources. The metadata captured includes critical observation details such as the first and last-seen timestamps, the frequency of occurrences, and other pertinent contextual information.

By leveraging this framework, you can perform searches against these summarized data indexes rather than the raw logs. This approach significantly reduces the overall utilization of Splunk Virtual Compute (SVC) resources. The summarized data indexes are designed to be 99% less voluminous compared to the original logs. As a result, you can store these indexes for extended periods – ranging from months to years – without the need to retain the full, raw, logs. This extended retention capability allows for comprehensive historical analysis and reporting while dramatically lowering storage and compute costs associated with managing large volumes of raw log data.

Context-aware and Granular Volume Control

DataBahn's context-aware volume controller framework offers a sophisticated approach to managing and optimizing your data streams, specifically by focusing on "hot detection" data. This framework evaluates your Splunk security content, including the rules and searches you have configured. By leveraging this evaluation, DataBahn enables you to send only the data that is critical for your search and analytics use cases within Splunk. This targeted approach helps in reducing the amount of data processed which, in turn, minimizes the compute units required for processing and analyzing your data.

In addition to this streamlined data forwarding capability, the volume controller framework includes various granular volume control options. These options encompass:

- **Data Aggregation:** This method allows you to consolidate data from multiple sources or events into a summarized form. By aggregating data, you can reduce the overall volume of data without sacrificing the essential security context needed for effective analysis and deduction.

- **Data Suppression:** Through data suppression, you can selectively filter out or omit certain data elements that are deemed less relevant or redundant. This ensures that only the most pertinent information is retained and forwarded, helping to further refine the data stream and reduce unnecessary processing overhead.

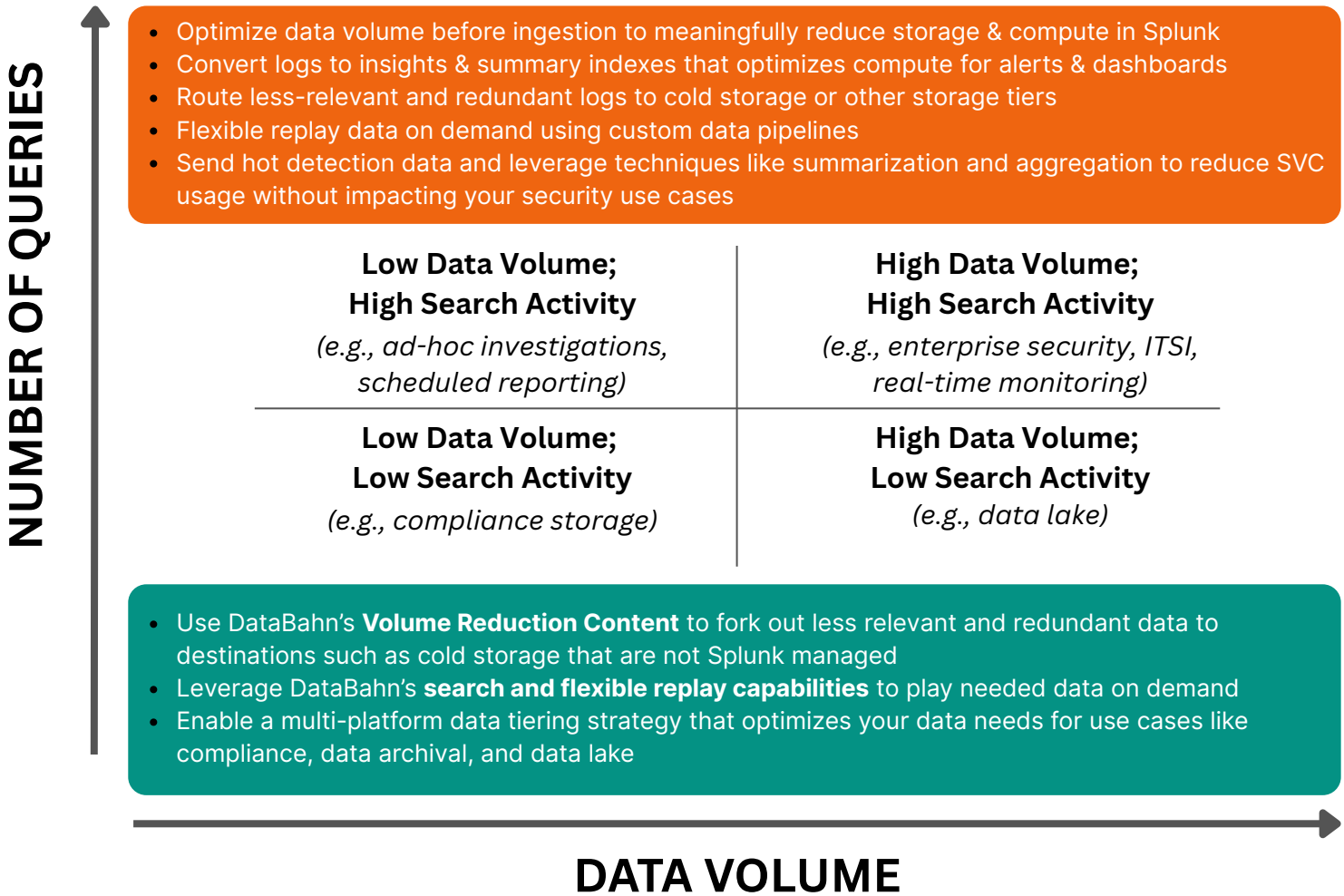
These features collectively enable you to meaningfully summarize your data while preserving the crucial security context, thus enhancing the efficiency of your Splunk deployment and optimizing resource allocation.

Multi-vendor data tiering

The approach of multi-vendor data tiering involves strategically directing various subsets of data to the most appropriate storage solutions based on their specific use cases, thereby optimizing both performance and cost-effectiveness. For

instance, data used for real-time dashboard visualizations should continue to be streamed to Splunk Cloud to ensure immediate updates and interactive capabilities. On the other hand, compliance-related data, which often requires long-term storage for regulatory purposes, should be transferred to more cost-effective archival solutions, such as Amazon S3, Microsoft Azure Blob Storage, or Google Cloud Storage.

Additionally, data that is queried on an ad hoc basis, which may not require constant access, can be stored in affordable log search platforms to manage expenses efficiently. By employing this tiered storage approach, organizations can significantly reduce the need for extensive Storage Blocks while still fulfilling their data management requirements, leading to more effective cost and resource management.



By using DataBahn in tandem with Splunk, you can aim to optimize your SVC consumption by 50% or more. With this level of optimization, customers have been able to ingest data from new sources without procuring more SVC credits. This combined value is why companies at all stages of SVC adoption are considering DataBahn.

Conclusion

By gaining a thorough understanding of how those components affect your Splunk usage and costs, and by applying effective optimization strategies, such as those provided by DataBahn's data fabric, you can significantly enhance your Splunk environment. DataBahn's data fabric offers advanced features designed to streamline data management and optimize resource utilization, resulting in substantial cost reductions. Moreover, it contributes to improved overall performance and efficiency in your data analytics process.

In essence, leveraging DataBahn's data fabric as part of your strategy to navigate the Splunk Workload Pricing Model not only facilitates cost savings but also enhances the performance and effectiveness of your data analytics operations. This comprehensive approach ensures that you maximize the value of your Splunk deployment while maintaining fiscal responsibility and operational excellence.