HCP Tuning Guide 1

General Tuning

Date of Publish: 2018-10-15

http://docs.hortonworks.com

Contents

Introduction to Tuning HCP

Tuning your Hortonworks Cybersecurity Platform (HCP) architecture can help maximize the performance of the Apache Metron Storm topologies.

In the simplest terms, HCP powered by Apache Metron is a streaming architecture created on top of Kafka and three main types of Storm topologies: parsers, enrichment, and indexing. Each parser has its own topology and there is also a highly performant, specialized spout-only topology for streaming PCAP data to HDFS.

The HCP architecture can be tuned almost exclusively using a few primary Storm and Kafka parameters along with a few Metron-specific options. You can think of the data flow as being similar to water flowing through a pipe, and the majority of these options assist in tweaking the various pipe widths in the system.

General Tuning Suggestions

Tuning Hortonworks Cybersecurity Platform (HCP) depends in large part on tuning three areas: Kafka, Storm, and indexing.

Indexing is where most of your turning issues are likely to occure because it is the most IO intensive.

The second area that needs tuning is parallelism in both Kafka and Storm. The performance of the Storm topology. and therefore the performance of Metron, degrades when it cannot ingest data fast enough to keep up with the data source. Therefore, much of Metron tuning focuses on adjusting the data throughput of the Storm topologies. For more information on tuning a Storm topology, see Apache Storm Overview.

The third area that requires analysis and tuning is consumer lags on the key Kafka topics: enrichment, indexing, parser.

When tuning your Metron configuration, consider the following:

- Look at Elasticsearch and Solr tuning
- Assign small values for parallelism, and increase values incrementally
- Aim for an even balance across your topologies
- Check your system logs for the following:
    - Empty results - may indicate that your data is broken
    - Kafka - Consumer lags on key Kafka topics
    - Load average or system latency - a high load average might indicate underlying stress on the machine
    - Exceptions - Any exceptions shown in the Storm log or key topologies can indicate possible problems with underlying systems and data
- What topology do I want to tune?
- What is the capacity of Storm topology?

It is also important to consider the growth of your cluster and data flow. You might want to set the number of tasks higher than the number of executors to accommodate for future performance tuning and rebalancing without the need to bring down your topologies.