CCP Understanding Profiler 2.0.1

Understanding Profiling

Date of publish: 2017-11-06

https://docs.cloudera.com/

Legal Notice

Contents

Understanding Profiling

A profile describes the behavior of an entity on a network. This feature is typically used by a data scientist and you should coordinate with the data scientist to determine if they need your assistance with customizing the Profiler values.

CCP installs the Profiler which runs as an independent Apache Storm topology. The configuration for the Profiler topology is stored in Apache ZooKeeper at /metron/topology/profiler. These properties also exist in the default installation of CCP at $METRON_HOME/config/zookeeper/profiler.json. You can change the values on disk and then upload them to ZooKeeper using $METRON_HOME/bin/zk_load_configs.sh.

Note:

The Profiler can persist any serializable object, not just numeric values.

CCP supports the following profiler properties:

profiler.workers	The number of worker processes to create for the topology.
profiler.executors	The number of executors to spawn per component.
profiler.input.topic	The name of the Kafka topic from which to consume data.
profiler.output.topic	The name of the Kafka topic to which profile data is written. Only used with profiles that use the triage` result field](#result).
profiler.period.duration	The duration of each profile period. Define this value along with profiler.period.duration.units.
profiler.period.duration.units	The units used to specify the profile period duration. Define this value along with profiler.period.duration.
profiler.ttl	If a message has not been applied to a Profile in this period of time, the Profile is forgotten and its resources cleaned up. Define this value along with profiler.ttl.units.
profiler.ttl.units	The units used to specify profiler.ttl.
profiler.hbase.salt.divisor	A salt is prepended to the row key to help prevent hotspotting. This constant is used to generate the sale. Ideally, this constant should be roughly equal to the number of nodes in the Apache HBase cluster.
profiler.hbase.table	The name of the HBase table that profiles are written to.
profiler.hbase.column.family	The column family used to store profiles.
profiler.hbase.batch	The number of puts written in a single batch.

profiler.hbase.flush.interval.seconds	The maximum number of seconds between batch writes to HBase.