

Understanding Profiling

Date of publish: 2017-11-06

CLOUDERA

Legal Notice

© Cloudera Inc. 2019. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 ("ASLv2"), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER'S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Understanding Profiling.....4

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 <code>profiler.period.duration.units</code> .
profiler.period.duration.units	The units used to specify the profile period duration. Define this value along with <code>profiler.period.duration</code> .
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 <code>profiler.ttl.units</code> .
profiler.ttl.units	The units used to specify <code>profiler.ttl</code> .
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.