

CCP Configuring Indexing 2.0.1

Runbook Configuring Indexing

Date of publish: 2017-11-06

CLOUDERA

<https://docs.cloudera.com/>

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

Configuring Indexing.....	4
Default Configuration.....	4
Specify Index Parameters.....	5
Turn off HDFS Writer.....	7

Configuring Indexing

The indexing topology is a topology dedicated to taking the data from a topology that has been enriched and storing the data in one or more supported indices. More specifically, the enriched data is ingested into Kafka, written in an indexing batch or bolt with a specified size, and sent to one or more specified indices. The configuration is intended to configure the indexing used for a given sensor type (for example, snort).

Currently, Cloudera Cybersecurity Platform (CCP) supports the following indices:

- Elasticsearch
- Solr
- HDFS under /apps/metron/enrichment/indexed

Depending on how you start the indexing topology, it can have HDFS and either elasticsearch or SOLR writers running.

Just like the Global Configuration file, the Indexing Configuration file format is a JSON file stored in ZooKeeper and on disk at \$METRON_HOME/config/zookeeper/ indexing.

Within the sensor-specific configuration, you can configure the individual writers. The parameters currently supported are:

index	The name of the index to write to (defaulted to the name of the sensor).
batchSize	The size of the batch that is written to the indices at once (defaulted to 1).
enabled	Whether the index or writer is enabled (default true).

Default Configuration

If you do not configure the individual writers, the sensor-specific configuration will use the default values.

You can choose to use this default configuration by either not creating the Indexing Configuration file or by entering the following in the file. You can name the file anything you like, for example index_config.json, but it must be located at \$METRON_HOME/config/zookeeper/indexing.

```
{
}
```

If a writer configuration is unspecified, then a warning is indicated in the Storm console. For example, WARNING: Default and (likely) unoptimized writer config used for hdfs writer and sensor squid. You can ignore this warning message if you intend to use the default configuration.

This default configuration uses the following configuration:

- elasticsearch writer
 - index name the same as the sensor
 - batch size of 1
 - enabled
- hdfs writer
 - index name the same as the sensor
 - batch size of 1
 - enabled

Specify Index Parameters

You can to specify the parameters for the writers rather than using the default values using the CCP Management user interface.

Procedure

1.

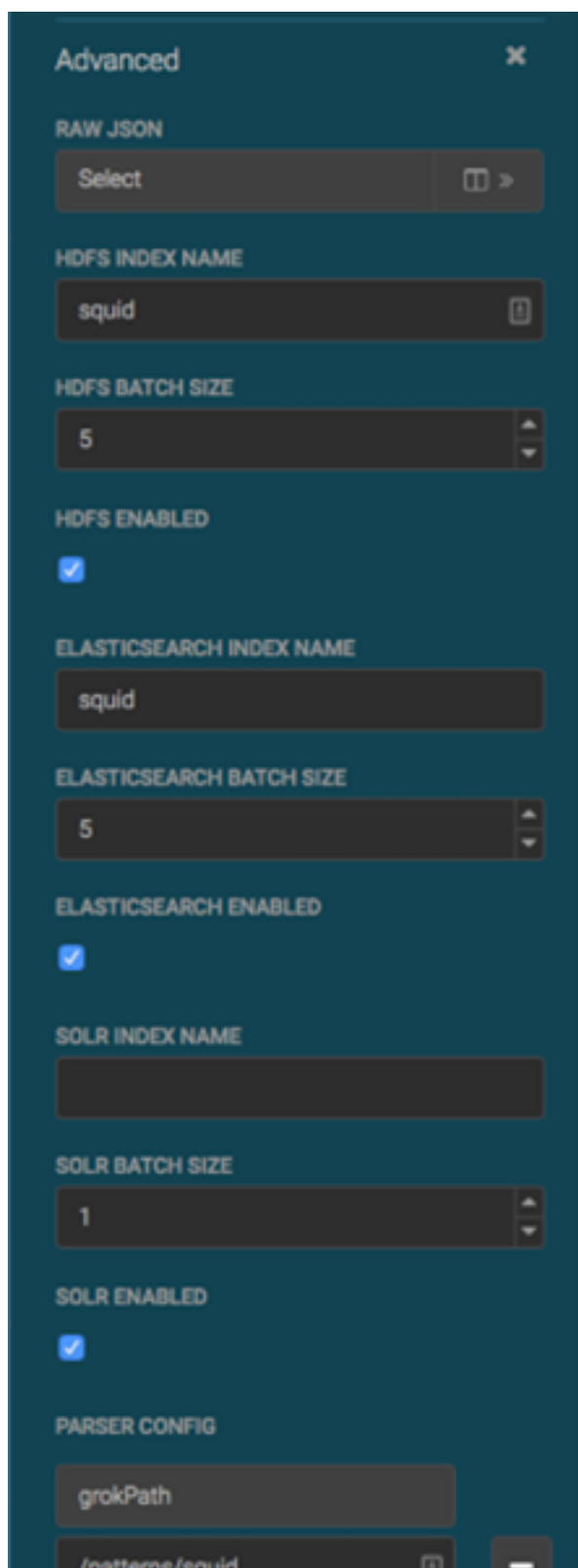


Edit your sensor by clicking (the edit button) next your sensor in the Management Module.

2. Click the **Advanced** button next to **Save** and **Cancel**.

The Management Module expands the panel to display the Advanced fields.

Management Module Advanced Panel



The image shows a screenshot of an 'Advanced' configuration window. It contains several sections for configuring indexing:

- RAW JSON**: A dropdown menu with 'Select' and a right arrow icon.
- HDFS INDEX NAME**: A text input field containing 'squid'.
- HDFS BATCH SIZE**: A numeric input field with a value of 5.
- HDFS ENABLED**: A checkbox that is checked.
- ELASTICSEARCH INDEX NAME**: A text input field containing 'squid'.
- ELASTICSEARCH BATCH SIZE**: A numeric input field with a value of 5.
- ELASTICSEARCH ENABLED**: A checkbox that is checked.
- SOLR INDEX NAME**: An empty text input field.
- SOLR BATCH SIZE**: A numeric input field with a value of 1.
- SOLR ENABLED**: A checkbox that is checked.
- PARSER CONFIG**: A text input field containing 'grokPath'.

3. Enter index configuration information for your sensor.
4. Click the **Raw JSON** field and set the alert field to "type": "nested":

```
},  
"alert": {
```

```
"type": "nested"
}
```

If this field is not set, Elasticsearch can throw an error and the field will not be queryable.

5. Click **Save** to save your changes and push your configuration to ZooKeeper.

Turn off HDFS Writer

You can turn off the HDFS writer when you are configuring and testing your system.

Procedure

Turn off the HDFS index or writer using the following syntax in the index.json file.

```
{
  "elasticsearch": {
    "index": "foo",
    "enabled" : true
  },
  "hdfs": {
    "index": "foo",
    "batchSize": 100,
    "enabled" : false
  }
}
```