Runbook Configuring Indexing

Date of Publish: 2018-11-15

http://docs.hortonworks.com

Contents

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, Hortonworks Cybersecurity Platform (HCP) 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 HCP Management Module.

Procedure

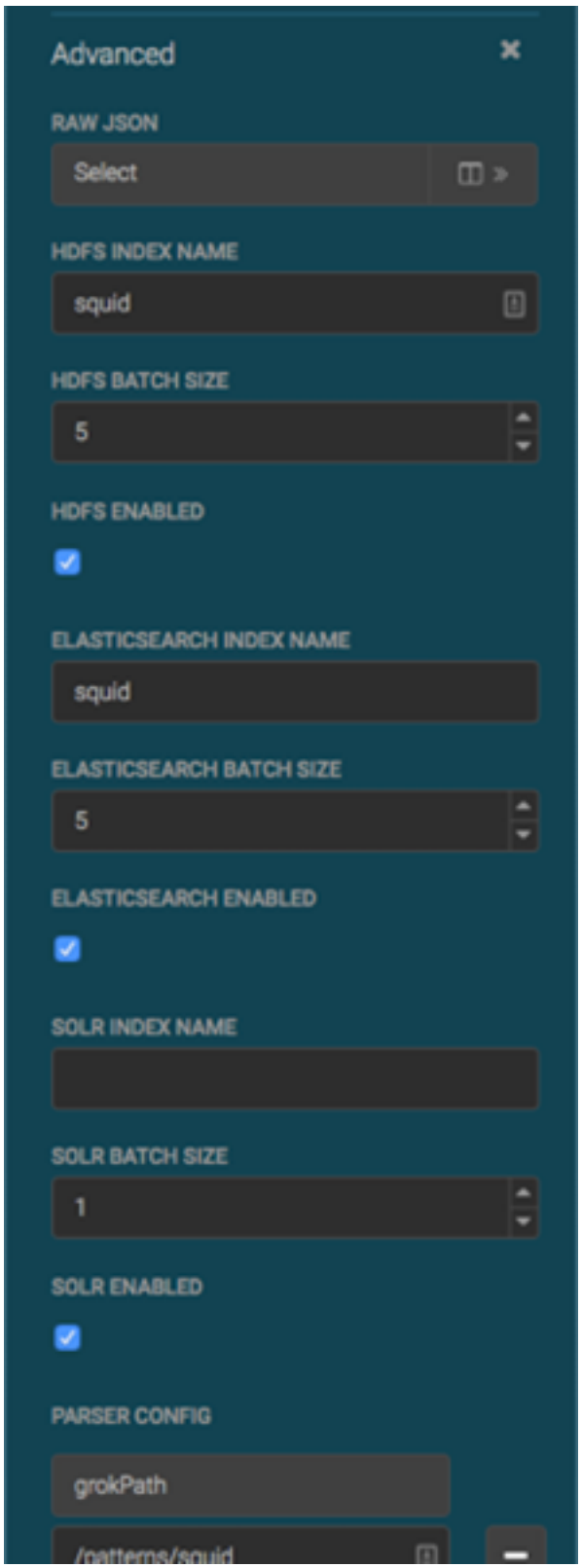
1. Edit your sensor by clicking

   (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



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
 }
}
```