FortiSIEM - Elasticsearch Storage Guide

Version 5.2.6

FORTINET DOCUMENT LIBRARY
https://docs.fortinet.com

FORTINET VIDEO GUIDE
https://video.fortinet.com

FORTINET BLOG
https://blog.fortinet.com

CUSTOMER SERVICE & SUPPORT
https://support.fortinet.com

FORTINET TRAINING & CERTIFICATION PROGRAM
https://www.fortinet.com/support-and-training/training.html

NSE INSTITUTE
https://training.fortinet.com

FORTIGUARD CENTER
https://fortiguard.com/

END USER LICENSE AGREEMENT
https://www.fortinet.com/doc/legal/EULA.pdf

FEEDBACK
Email: techdoc@fortinet.com

11/05/2019
FortiSIEM 5.2.6 Elasticsearch Storage Guide

TABLE OF CONTENTS

Change Log

Date	Change Description
03/30/2018	Initial version of Elasticsearch storage guide.
04/11/2018	Revision 2 with updated links under 'Step 1: Download and Install Elasticsearch'.
03/25/2019	Revision 3 with updated information on Hot Data Node and Warm Data Node. Also updates the Elasticsearch versions supported by FortiSIEM. Added instructions for upgrading FortiSIEM to Elasticsearch 6.4.2.
11/05/2019	Revision 4: FortiSIEM supports Elasticsearch 6.8.x.

Setting up Elasticsearch for FortiSIEM Event Storage

Elasticsearch is a distributed database. It can be deployed as an all-in-one node; but more commonly in a cluster setup consisting of a Master Node, Co-ordinating Node and Data Nodes. FortiSIEM currently supports Elasticsearch 6.8.x.

FortiSIEM can work with both Elasticsearch configurations:

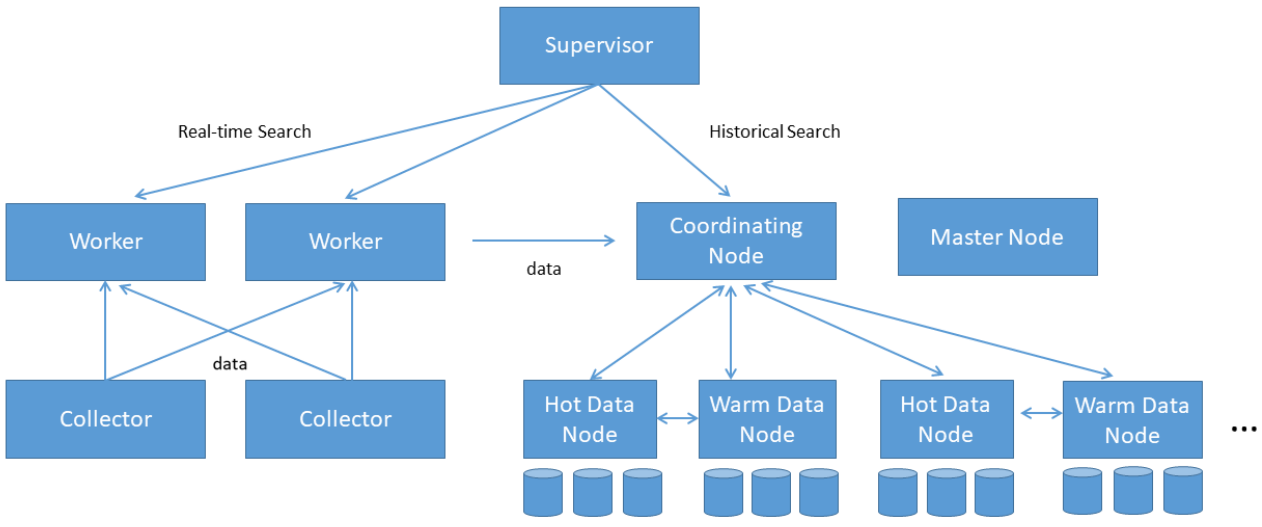
- All-in-One Node
- Cluster

In Full Cluster Deployment Architecture, the Supervisor and Worker nodes perform the real-time operations (Collection, Rules and Inline reports) while the data is indexed and stored in Elasticsearch. Historical search queries are sent from the Supervisor node to the Coordinating node, which communicates with the Hot and Warm Data Nodes to produce search results.

In storing mode, the current data written to the Hot Data Node. When this node is full, data is moved to the Warm Data Node. When the Warm Data Node is full, the data is purged.

In querying mode, Elasticsearch manages the query. When the Coordinating Node is queried, it distributes the query to the Hot and Warm Data Nodes. The Hot and Warm Data Nodes return the response to the Coordinating node.

The following diagram shows a Full Cluster Deployment Architecture:



Pre-Install considerations

Operating System

While Elasticsearch runs on many Operating Systems, FortiSIEM can operate on CentOS 6.8 and RHEL 7.2. *For details, see here.*

Assuming that Elasticsearch runs under 'admin' user profile, set the following parameters for this user. Do this for all nodes.

1. Allocate sufficient file descriptors per process in `/etc/security/limits.conf`
   ```
   admin soft nofile 65536
   admin hard nofile 65536
   ```
2. Allocate sufficient threads per user in `/etc/security/limits.conf`
   ```
   admin soft nproc 4096
   admin hard nproc 4096
   ```
3. Verify the allocations by running `ulimit -a`

CPU

- Scale-out Architecture – more cores are better than single large CPUs.
- The Master Node is light weight and needs 2-4 cores.
- All other nodes need more cores as EPS and Search requirements go up. Typically, 8-16 cores are sufficient. *See 'FortiSIEM Sizing Guide' here for details.*

Memory

- Coordinator and Data Nodes need more memory (16-64 GB). Elastic JVM needs half of this memory. *See 'FortiSIEM Sizing Guide' here for details.*
- For all nodes, disable swapping using the command `swapoff -a` or comment `swap` on `/etc/fstab`
- In the file `/etc/sysctl.conf`, set `vm.max_map_count` to '262144' and check `sysctl vm.max_map_count`.

Network

- Low latency network with 1 Gbps - 10 Gbps would be better but not necessary.
- Avoid clusters spanning geographical distances.
- TCP Port usage:
  - TCP/9200 for HTTP communication between FortiSIEM Supervisor node and Coordinating node.
  - TCP/9300 for communicating between Super and Coordinating (FortiSIEM querying) and Elastic internal. Ports 9200 and 9300 can be configured by the user. For example, they can be 9201, 9301, etc.
- TCP/5601 for Kibana, if needed.

Disk Size

Local Disks with high IOPS (SSD or RAID) for Data nodes are critical since FortiSIEM is a high Read and high Write environment.

Note: Avoid Network Attached Storage for Data nodes.

Elasticsearch needs sufficient disk space to store events since it computes lots of indices. It needs 32TB to store one year of logs at constant 1K EPS with no replication and 64 TB with 1 replication. If that is divided across 5 Data Nodes, the disk space required is 7TB/Data Node with no replication and 14TB/Data Node with 1 replication.

Setting up Elasticsearch

FortiSIEM currently supports Elasticsearch version 6.8.x. Follow the steps below to setup Elasticsearch for FortiSIEM Event Storage.

Step 1: Download and Install Elasticsearch

Follow the steps below to download and install Elasticsearch:

1. Download Elasticsearch using the URLs:
   https://www.elastic.co/downloads/past-releases
   https://www.elastic.co/guide/en/elastic-stack/6.8/index.html
2. Install Elasticsearch using the URL:
   https://www.elastic.co/guide/en/elastic-stack/6.8/index.html

Step 2: Configure Elasticsearch

The basic configuration steps are available *here*. The configurations are suggested only. Your environment might have different requirements.

- Configuration for All-in-One Node
- Configuration for Cluster
- Configure JVM heap size

Configure Elasticsearch All-in-One Node

The configurations for the Data node are defined in the `elasticsearch/config/elasticsearch.yml` file.

Note: In Elasticsearch 6.8.x, X-Pack is installed and Machine Learning (ML) is enabled by default. To avoid unnecessary resource usage, disbale this by adding the following line in the `elasticsearch.yml` file:

```
xpack.ml.enabled: false
```

To configure Elasticsearch All-in-One Node, change the following parameters in the file:

- node.name: data
- network.host: <IP-Address>
- search.remote.connect: false

Configure Elasticsearch Cluster

The configurations for Coordinator node, Master node and each Data node are defined in the elasticsearch/config/elasticsearch.yml file.

Note: In Elasticsearch 6.8.x, X-Pack is installed and Machine Learning (ML) is enabled by default. To avoid unnecessary resource usage, disable this by adding the following line in the elasticsearch.yml file:

```
xpack.ml.enabled: false
```

To configure Elasticsearch Cluster, make specific parameter changes in the configuration file. For example, see the parameters to change for an Elasticsearch Cluster with 1 dedicated Coordinator, 1 dedicated Master and 3 Data nodes.

a) Coordinator Node

- node.name: coordinator
- network.host: <CoordinatorIP-Address>
- discovery.zen.minimum_master_nodes: 2
- discovery.zen.ping.unicast.hosts: ["<DataNode-1-IP>", "<DataNode-2-IP>", "<DataNode-3-IP>", "<MasterNode-IP>"]
- node.data: false
- node.ingest: false
- search.remote.connect: false

b) Master Node

- node.name: master
- network.host: <MasterIP-Address>
- discovery.zen.minimum_master_nodes: 2
- discovery.zen.ping.unicast.hosts: ["<DataNode-1-IP>", "<DataNode-2-IP>", "<DataNode-3-IP>", "<CoordinatorNode-IP>"]
- node.master: true
- node.data: false
- node.ingest: false
- search.remote.connect: false

c) Each Hot Data Node

- node.name: data
- network.host: <DataIP-Address>
- discovery.zen.minimum_master_nodes: 2
- discovery.zen.ping.unicast.hosts: ["<DataNode-1-IP>", "<DataNode-2-IP>", "<DataNode-3-IP>", "<CoordinatorNode-IP>", "<MasterNode-IP>"]]
- node.master: false
- node.data: true

- `node.ingest: false`
- `search.remote.connect: false`
- `node.attr.box_type: hot`

d) Each Warm Data Node

- `node.name: data`
- `network.host: <DataIP-Address>`
- `discovery.zen.minimum_master_nodes: 2`
- `discovery.zen.ping.unicast.hosts: ["<DataNode-1-IP>", "<DataNode-2-IP>", "<DataNode-3-IP>", "<CoordinatorNode-IP>", "<MasterNode-IP>"]]`
- `node.master: false`
- `node.data: true`
- `node.ingest: false`
- `search.remote.connect: false`
- `node.attr.box_type: warm`

  Note: One Hot Data node in the Cluster should be Master eligible. For this node, the parameter `node.master: true`.

Configure JVM heap size

- Based on memory size of the node, change the parameters in `jvm.xml`. It is recommended to provide half of the node's memory size but not more than 30 GB. For example, if the node has 64 GB memory, change the parameters:
  - `-Xms30g`
  - `-Xmx30g`

Step 3: Configure Elasticsearch in FortiSIEM

Once you have chosen the Elasticsearch configuration and set up the cluster according to the performance matrix:

1. Go to FortiSIEM > **ADMIN** > **Setup** > **Storage** > select **Elasticsearch**.
2. Enter the following:
   a. **Cluster Name** - Name of the Elasticsearch Cluster
   b. **Cluster IP/Host** - Coordinating node IP
   c. **Shards** - Number of Shards. Adding or moving shards is easy, but splitting them is not possible. Plan ahead for shard sizing.
   d. **Replicas** - Number of Replicas

   For Shards and Replicas, refer to the 'FortiSIEM Sizing Guide' here.

Upgrading to Elasticsearch 6.8.x

FortiSIEM 5.1.2 supports Elasticsearch 5.6.2. FortiSIEM 5.2.6 supports Elasticsearch 6.8.x, 6.4.2, and 5.6.2. If you are running FortiSIEM 5.1.2 and want to upgrade to FortiSIEM 5.2.6 and Elasticsearch 6.8.x, follow these steps:

1. Upgrade FortiSIEM to 5.2.6. See FortiSIEM Upgrade Guide *here*.
2. Upgrade Elasticsearch to 6.8.x. See Upgrade Elasticsearch *here* and *here*.
3. Set Elasticsearch Data nodes as Hot nodes. **Note:** At least one Hot Data node is required if there is no Replication and at least two Data nodes for one Replication.
4. Login to FortiSIEM and go to **ADMIN > Setup > Storage**. Click **Test** and **Save** to force Elasticsearch to use the new event template.
5. Reboot FortiSIEM.
6. (Optional) You may want to add Warm nodes to utilize Elasticsearch Hot/Warm architecture. FortiSIEM will manage the data movement from Hot to Warm nodes. Configure data movement thresholds under **ADMIN > Settings > Database > Archive**.