Splunk® Enterprise
Managing Indexers and Clusters of Indexers 8.0.5

System requirements and other deployment considerations for indexer clusters

Generated: 7/21/2020 11:19 am
System requirements and other deployment considerations for indexer clusters

Indexer clusters are groups of Splunk Enterprise indexers, so, for the most part, you just need to adhere to the system requirements for indexers. For detailed software and hardware requirements for indexers, read "System requirements" in the Installation Manual. The current topic notes additional requirements for clusters.

**Note:** System requirements vary somewhat for clusters with SmartStore indexes. See SmartStore system requirements.

**Summary of key requirements**

These are the main issues to note:

- Each cluster node (master, peer, or search head) must reside on a separate Splunk Enterprise instance.
- Each node instance must run on a separate machine or virtual machine, and each machine must be running the same operating system and version.
- All nodes must be connected over a network.
- There are strict version compatibility requirements between cluster nodes.

For example, to deploy a cluster consisting of three peers, one master, and one search head, you need five Splunk Enterprise instances running on five machines connected over a network. And all machines must be running the same operating system and version.

These are some additional issues to be aware of:

- Compared to a non-clustered deployment, clusters require more storage, to accommodate the multiple copies of data.
- Index replication, in and of itself, does not increase your licensing needs.
- You cannot use a deployment server to distribute updates to peers.

See the remainder of this topic for details.

**Required Splunk Enterprise instances**

Each cluster node must reside on its own Splunk Enterprise instance. Therefore, the cluster must consist of at least (replication factor + 2) instances: a minimum of replication factor number of peer nodes, plus one master node and one or more search heads. For example, if you want to deploy a cluster with a replication factor of 3, you must set up at least five instances: three peers, one master, and one search head. To learn more about the replication factor, see “Replication factor” in this manual.

The size of your cluster depends on other factors besides the replication factor, such as the amount of data you need to index. See "Indexer cluster deployment overview”.

Although the master has search capabilities, you should only use those capabilities for debugging purposes. The resources of the master must be dedicated to fulfilling its critical role of coordinating cluster activities. Under no circumstances should the master be employed as a production search head. See "Additional roles for the master node".
Splunk Enterprise version compatibility

Interoperability between the various types of cluster nodes is subject to strict compatibility requirements. In brief:

- The master node must run the same or a later version from the peer nodes and search heads.
- The search heads must run the same or a later version from the peer nodes.
- The peer nodes must all run exactly the same version, down to the maintenance level.

Compatibility between the master and the peer nodes and search heads

Peer nodes and search heads can run different versions from the master, subject to these restrictions:

- The master node must run the same or a later version than the peer nodes and search heads.
- The master node can run at most three minor version levels later than the peer nodes. For example, a 7.0 master node can run against 6.6, 6.5, and 6.4 peer nodes, but not 6.3 peer nodes.
- The master node must run version 6.2 or later.
- The peer nodes and search heads must run version 6.1 or later.
- For master nodes running version 6.1 or earlier, mixed-version clusters are not available. The peer nodes and search heads must run the same version as the master.

Compatibility between the master and 6.1 peer nodes

To run a 6.2 or later master against 6.1 peer nodes, you must set the attribute use_batch_mask_changes in the master's server.conf file to false:

```
splunk edit cluster-config -use_batch_mask_changes false
```
You do not need to restart the master if you set this attribute with the CLI.

Caution: After upgrading all peer nodes to 6.2 or later, you must revert use_batch_mask_changes to true.

Compatibility between peer nodes

All peer nodes must run the same version of Splunk Enterprise, down to the maintenance level. You must update all peer nodes to a new release at the same time. You cannot, for example, run an indexer cluster with some peer nodes at 6.n.2 and others at 6.n.1.

Compatibility between peer nodes and search heads

Starting with 6.3, the peer nodes and search heads can run different versions from each other. The search heads must run the same or a later version from the peer nodes.

Search head clusters participating in an indexer cluster have the same compatibility requirements as individual search heads. For information on other search head cluster version requirements, see "System requirements and other deployment considerations for search head clusters" in the Distributed Search manual.

Machine requirements

Each node of the cluster (master node, peer nodes, and search heads) must run on its own, separate machine or virtual machine. Other than that, the hardware requirements, aside from storage, are basically the same as for any Splunk Enterprise instance. See "Reference hardware" in the Capacity Planning Manual.
The main difference is in the storage requirements for peer nodes, discussed below.

**Note:** The storage needs of the master node are significantly lower than those specified in the "Reference hardware" topic, since the master does not index external data.

### Operating system requirements

Indexer clustering is available on all operating systems supported for Splunk Enterprise. For a list of supported operating systems, see System requirements in the *Installation Manual*.

All indexer cluster nodes (master node, peer nodes, and search heads) must run on the same operating system and version.

If the indexer cluster is integrated with a search head cluster, then the search head cluster instances, including the deployer, must run on the same operating system and version as the indexer cluster nodes.

### Synchronization of system clocks across the cluster

It is important that you synchronize the system clocks on all machines, virtual or physical, that are running Splunk Enterprise instances participating in the cluster. Specifically, this means your master node, peer nodes, and search heads. Otherwise, various issues can arise, such as timing problems between the master and peer nodes, search failures, or premature expiration of search artifacts.

The synchronization method you use depends on your specific set of machines. Consult the system documentation for the particular machines and operating systems on which you are running Splunk Enterprise. For most environments, Network Time Protocol (NTP) is the best approach.

### Storage considerations

When determining storage requirements for your clustered indexes, you need to consider the increased capacity, across the set of peer nodes, necessary to handle the multiple copies of data.

It is strongly recommended that you provision all peer nodes to use the same amount of disk storage.

Clusters use the usual settings for managing index storage, as described in "Configure index storage".

#### Determine your storage requirements

It is important to ensure you have enough disk space to accommodate the volume of data your peer nodes will be processing. For a general discussion of Splunk Enterprise data volume and how to estimate your storage needs, refer to "Estimating your storage requirements" in the *Capacity Planning Manual*. That topic provides information on how to estimate storage for non-clustered indexers, so you need to supplement its guidelines to account for the extra copies of data that a cluster stores.

With a cluster, in addition to considering the volume of incoming data, you must consider the replication factor and search factor to arrive at your total storage requirements across the set of peer nodes. With a replication factor of 3, you are storing three copies of your data. You will need extra storage space to accommodate these copies, but you will not need three times as much storage. Replicated copies of non-searchable data are smaller than copies of searchable data, because they include only the data and not the associated index files. So, for example, if your replication factor is 3 and your search factor is 2, you will need more than two, but less than three, times the storage capacity compared to storing the same data on non-clustered indexers.
Exactly how much less storage your non-searchable copies require takes some investigation on your part. The index files excluded by non-searchable copies can vary greatly in size, depending on factors described in "Estimating your storage requirements" in the Capacity Planning Manual.

**Important:** A master is not aware of the amount of storage on individual peer nodes, and therefore it does not take available storage into account when it makes decisions about which peer node should receive a particular set of replicated data. It also makes arbitrary decisions about which peer should make some set of replicated data searchable (in cases where the search factor is 2 or greater). Therefore, you must ensure that each peer node has sufficient storage not only for the data originating on that peer, but also for any replicated copies of data that might get streamed to it from other peers. You should continue to monitor storage usage throughout the life of the cluster.

**Storage requirement examples**

As a ballpark figure, incoming syslog data, after it has been compressed and indexed, occupies approximately 50% of its original size:

- 15% for the rawdata file.
- 35% for associated index files.

In practice, this estimate can vary substantially, based on the factors described in "Estimating your storage requirements" in the Capacity Planning Manual.

Assume you have 100GB of syslog data coming into Splunk Enterprise. In the case of a non-clustered indexer, that data would occupy approximately 50GB (50% of 100GB) of storage on the indexer. However, in the case of clusters, storage calculations must factor in the replication factor and search factor to arrive at total storage requirements across all the cluster peers. (As mentioned earlier, you cannot easily predict exactly how much storage will be required on any specific peer.)

Here are two examples of estimating cluster storage requirements, both assuming 100GB of incoming syslog data, resulting in 15GB for each set of rawdata and 35GB for each set of index files:

- **3 peer nodes, with replication factor = 3; search factor = 2:** This requires a total of 115GB across all peer nodes (averaging 38GB/peer), calculated as follows:
  - Total rawdata = (15GB \* 3) = 45GB.
  - Total index files = (35GB \* 2) = 70GB.
- **5 peer nodes, with replication factor = 5; search factor = 3:** This requires a total of 180GB across all peer nodes (averaging 36GB/peer), calculated as follows:
  - Total rawdata = (15GB \* 5) = 75GB.
  - Total index files = (35GB \* 3) = 105GB.

**Storage hardware**

In pre-6.0 versions of Splunk Enterprise, replicated copies of cluster buckets always resided in the colddb directory, even if they were hot or warm buckets. Starting with 6.0, hot and warm replicated copies reside in the db directory, the same as for non-replicated copies. This eliminates any need to consider faster storage for colddb for clustered indexes, compared to non-clustered indexes.

**Licensing information**

As with any Splunk Enterprise deployment, your licensing requirements are driven by the volume of data your indexers process. Contact your Splunk sales representative to purchase additional license volume. Refer to "How licensing works"
in the *Admin Manual* for more information about Splunk Enterprise licensing.

There are just a few license issues that are specific to index replication:

- All cluster nodes, including masters, peers, and search heads, need to be in an Enterprise **license pool**, even if they're not expected to index any data.
- Cluster nodes must share the same licensing configuration.
- Only incoming data counts against the license; replicated data does not.
- You cannot use index replication with a free license.

**Ports that the cluster nodes use**

These ports must be available to cluster nodes:

- **On the master:**
  - The management port (by default, 8089) must be available to all other cluster nodes.

- **On each peer:**
  - The management port must be available to all other cluster nodes.
  - The replication port must be available to all other peer nodes.
  - The receiving port must be available to all forwarders sending data to that peer.

- **On each search head:**
  - The management port must be available to all other nodes.
  - The http port (by default, 8000) must be available to any browsers accessing data from the search head.

**Deployment server and clusters**

Do not use deployment server with cluster peers.

The deployment server is not supported as a means to distribute configurations or apps to cluster peers. To distribute configurations across the set of cluster peers, instead use the **configuration bundle** method outlined in the topic "Update common peer configurations".

For information on how to migrate app distribution from deployment server to the configuration bundle method, see "Migrate apps to a cluster".

**Additional roles for the master node**

As a general rule, you should dedicate the Splunk Enterprise instance running the master node to that single purpose. Constrain use of the master's built-in search head to debugging only.

Under limited circumstances, however, you might be able to colocate one or more of these lightweight functions on the master instance:

- **License master**
- **Monitoring console**
- **Search head cluster deployer**

To use the master instance for any of these additional roles, the master's cluster must remain below the following limits:

- 30 indexers
- 100,000 buckets
• 10 indexes
• 10 search heads

Do not colocate a deployment server on the master under any circumstances.

A master node and a deployment server both consume significant system resources while performing their tasks. The master node needs reliable and continuous access to resources to perform the ongoing management of the cluster, and the deployment server can easily overwhelm those resources while deploying updates to its deployment clients.

For a general discussion of management component colocation, see Components that help to manage your deployment in the Distributed Deployment Manual.