Splunk® Enterprise Managing Indexers and Clusters of Indexers 7.3.1

Indexer cluster deployment overview

Generated: 9/20/2019 10:43 pm
Indexer cluster deployment overview

This topic describes the main steps to deploying indexer clusters. Subsequent topics describe these steps in detail.

Before you attempt to deploy a cluster, you must be familiar with several areas of Splunk Enterprise administration:

- **How to configure indexers.** In particular, see "How the indexer stores indexes", along with the other topics in this manual that describe managing indexes.
- **What a search head does.** For an introduction to distributed search, see "About distributed search" in the Distributed Search manual. Note, however, that you configure search heads for indexer clusters somewhat differently from other search heads. For information on the differences see "Search head configuration overview" in this manual.
- **How to use a forwarder to get data into an indexer.** See "Use forwarders" in the Getting Data In manual.

**Important:** This chapter assumes that you are deploying independent search heads in the indexer cluster. For information on how to incorporate search heads that are members of a search head cluster, see "Integrate the search head cluster with an indexer cluster" in the Distributed Search manual.

**Migrating from a non-clustered Splunk Enterprise deployment?**

Clustered indexers (peers) have several different requirements from non-clustered indexers. It is important that you be aware of these issues before you migrate your indexers. See "Key differences between clustered and non-clustered deployments of indexers". After you read that material, go to "Migrate non-clustered indexers to a clustered environment" for the actual migration process.

**Important:** Before migrating an indexer from non-clustered to clustered, be certain of your needs. The process goes in one direction only. There is no supported procedure for converting an indexer from clustered to non-clustered.

**Deploying a multisite cluster?**

Multisite indexer clusters are considerably more complex than single-site clusters. Deploying them requires that you consider some additional issues and
perform an entirely different set of configurations. If you are deploying a multisite cluster, read this topic first and then read "Multisite indexer cluster deployment overview".

Deploy a cluster

When you deploy a cluster, you enable and configure the cluster master and the peer nodes that perform the indexing. You also enable a search head to search data in the cluster. In addition, you usually set up forwarders to send data to the cluster. Here is a diagram of a small cluster, showing the various nodes that you deploy:
These are the key steps in deploying clusters:

1. Identify your requirements:
   
a. Understand your data availability and failover needs. See "About indexer clusters".
   
b. Determine whether you will be deploying a basic, single-site cluster or a multisite cluster. Multisite clusters offer strong disaster recovery capabilities because they allow you to distribute copies of your data across multiple locations. They also enable search affinity, which reduces network traffic by limiting searches to local data. For more information, read "Multisite indexer clusters".
   
c. Decide what replication factor you want to implement. The replication factor is the number of copies of raw data that the cluster maintains. Your optimal replication factor depends on factors specific to your environment, but essentially involves a trade-off between failure tolerance and storage capacity. A higher replication factor means that more copies of the data will reside on more peer nodes, so your cluster can tolerate more node failures without loss of data availability. But it also means that you will need more nodes and more storage to handle the additional data. For multisite clusters, you also need to decide how many copies to put on each site. For more information, see "Replication factor".
   
   **Warning:** Make sure you start by choosing the right replication factor for your needs. It is inadvisable to increase the replication factor after the cluster contains a significant amount of data. The cluster would need to perform a large amount of bucket copying to match the increased replication factor, slowing significantly the overall performance of your cluster while the copying is occurring.
   
d. Decide what search factor you want to implement. The search factor tells the cluster how many searchable copies of indexed data to maintain. This helps determine the speed with which a cluster can recover from a downed node. A higher search factor allows the cluster to recover more quickly, but it also requires more storage space and processing power. For most single-site deployments, the default search factor value of 2 represents the right trade-off, allowing searches usually to continue with little interruption when a node goes down. For multisite clusters, you also need to decide how many searchable copies to put on each site. For more information, see "Search factor".
   
   **Warning:** Make sure you start by choosing the right search factor for your needs. It is inadvisable to increase the search factor after the cluster contains a significant amount of data. The cluster would need to perform a large amount of
processing (transforming non-searchable bucket copies into searchable copies) to match the increased search factor, and this will have an adverse effect on the overall performance of your cluster while the processing is occurring.

e. Identify other factors that also determine the size of your cluster; for example, the quantity of data you will be indexing. It usually makes sense to keep all your indexers in a single cluster, so for horizontal scaling, you will need to add peer nodes beyond those required by the replication factor. Similarly, depending on the anticipated search load, you might need to add more than one search head.

f. Study the topic "System requirements and other deployment considerations for indexer clusters" for information on other key issues.

2. **Install the Splunk Enterprise cluster instances on your network.** At a minimum, you will need (replication factor + 2) instances:

   - You need at least the replication factor number of **peer nodes**, but you might want to add more peers to boost indexing capacity, as mentioned in step 1e.
   - You also need two more instances, one for the **master node** and the other for the **search head**.

For multisite clusters, you must also take into account the search head and peer node requirements of each site, as determined by your search affinity and disaster recovery needs. See "Multisite indexer cluster deployment overview".

For information on how to install Splunk Enterprise, read the *Installation Manual*.

3. **Enable clustering on the instances:**

   a. Enable the master node. See "Enable the master node".

   b. Enable the peer nodes. See "Enable the peer nodes".

   c. Enable the search head. See "Enable the search head".

**Important:** For multisite clusters, the process of enabling cluster nodes is different. See "Multisite indexer cluster deployment overview".

4. **Complete the peer node configuration:**

   a. Configure the peers' index settings. This step is necessary only if you need to augment the set of default indexes and apps. In general, all the peers must use
the same set of indexes, so if you add indexes (or apps that define indexes) to one peer, you must add them to all peers, using a cluster-specific distribution method. There might also be other configurations that you need to coordinate across the set of peers. See "Prepare the peers for index replication" for information on how to do this.

b. Configure the peers' data inputs. For most purposes, it is best to use forwarders to send data to the peers, as discussed in "Ways to get data into an indexer cluster". As that topic states, you will usually want to deploy load-balancing forwarders with indexer acknowledgment enabled.

After you enable the nodes and set up data inputs for the peers, the cluster automatically begins indexing and replicating the data.

5. **Configure the master node to forward its data to the peer nodes.** This best practice provides several advantages. See "Best practice: Forward master node data to the indexer layer".

**Other deployment scenarios**

This manual also provides guidance on a few other cluster deployment scenarios:

- Add indexers with existing data to a cluster. See "Migrate non-clustered indexers to a clustered environment".
- Deploy **SmartStore** indexes on a new indexer cluster. See "Deploy SmartStore on a new indexer cluster".
- Migrate data currently on an indexer cluster to SmartStore. See "Migrate existing data on an indexer cluster to SmartStore".
- Bootstrap SmartStore indexes onto an indexer cluster. See "Bootstrap SmartStore indexes".
- Migrate a single-site cluster to multisite. See "Migrate an indexer cluster from single-site to multisite".
- Employ clusters purely for index scalability, where index replication is not a requirement. See "Use indexer clusters to scale indexing".