

# Replication factor - Splunk Documentation

5-6 minutes

---

1. [Documentation](#)
2. [Splunk® Enterprise](#)
3. [Managing Indexers and Clusters of Indexers](#)
4. [Replication factor](#)



[Download topic as PDF](#)

As part of setting up an indexer cluster, you specify the number of copies of data that you want the cluster to maintain. Peer nodes store incoming data in [buckets](#), and the cluster maintains multiple copies of each bucket. The cluster stores each bucket copy on a separate peer node. The number of copies of each bucket that the cluster maintains is the [replication factor](#).

## Replication factor and cluster resiliency

The cluster can tolerate a failure of (replication factor - 1) peer nodes. For example, to ensure that your system can tolerate a failure of two peers, you must configure a replication factor of 3, which means that the cluster stores three identical copies of each bucket on separate nodes. With a replication factor of 3, you can be certain that all your data will be available if no more than two peer nodes in the cluster fail. With two nodes down, you still have one

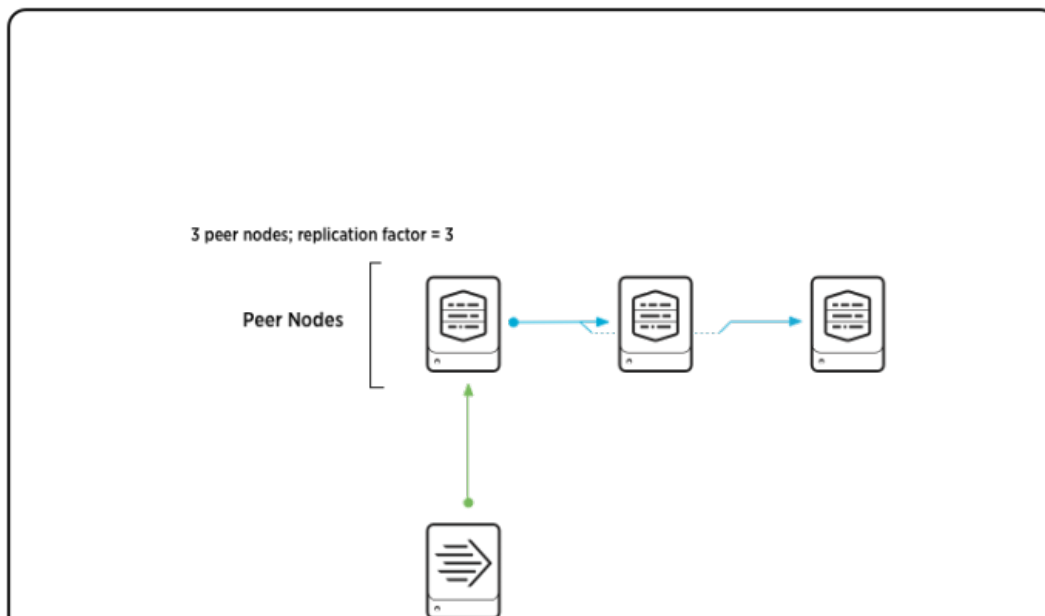
complete copy of data available on the remaining peers.

By increasing the replication factor, you can tolerate more peer node failures. With a replication factor of 2, you can tolerate just one node failure; with a replication factor of 3, you can tolerate two concurrent failures; and so on.

The trade-off is that you need to store and process all those copies of data. Although the replicating activity doesn't consume much processing power, still, as the replication factor increases, you need to run more indexers and provision more storage for the indexed data. On the other hand, since data replication itself requires little processing power, you can take advantage of the multiple indexers in a cluster to ingest and index more data. Each indexer in the cluster can function as both originating indexer ("source peer") and replication target ("target peer"). It can index incoming data and also store copies of data from other indexers in the cluster.

## Example: Replication factor in action

In the following diagram, one peer is receiving data from a forwarder, which it processes and then streams to two other peers. The cluster will contain three complete copies of the peer's data, one copy on each peer.





**Note:** This diagram represents a highly simplified version of peer replication, where all data is entering the system through a single peer. There are a few issues that add complexity to a real-life scenario:

- In most clusters, each of the peer nodes would be functioning as both source and target peer, receiving external data from a forwarder, as well as replicated data from other peers.
- To accommodate horizontal scaling, a cluster with a replication factor of 3 could consist of many more peers than three. At any given time, each source peer would be streaming copies of its data to two target peers, but each time it started a new hot bucket, its set of target peers could potentially change.

Later topics in this chapter describe in detail how clusters process data.

## Replication factor in multisite clusters

A multisite cluster uses a special version of the replication factor, the site replication factor. This determines not only the number of copies that the entire cluster maintains but also the number of copies that each site maintains. For information on the site replication factor, see [Configure the site replication factor](#).

PREVIOUS <a href="#">Basic indexer cluster concepts for advanced users</a>	NEXT <a href="#">Search factor</a>
---	---------------------------------------

This documentation applies to the following versions of Splunk<sup>®</sup> Enterprise: 6.5.7, 7.0.0, 7.0.1, 7.0.2, 7.0.3, 7.0.4, 7.0.5, 7.0.6, 7.0.7,