Splunk® Enterprise Securing Splunk Enterprise 7.3.1

Manage data integrity

Generated: 9/23/2019 9:43 pm
Manage data integrity

The Splunk Enterprise data integrity control feature provides a way to verify the integrity of data that is indexed.

When you enable data integrity control for an index, Splunk Enterprise computes hashes (using SHA 256) on every slice of data and stores those hashes so that you can go back later and verify the integrity of your data.

How it works

When you enable data integrity control, Splunk Enterprise computes hashes on every slice of newly indexed raw data and writes it to a l1Hashes file. When the bucket rolls from hot to warm, Splunk Enterprise computes a hash on the contents of the l1Hashes and stores the computed hash in l2Hash. Both hash files are stored in the rawdata directory for that bucket.

Note that data integrity control hashes newly indexed data, data coming from a forwarder should be secured and encrypted with SSL. For more information, see About securing Splunk with SSL.

Check your hashes to validate your data

To check Splunk Enterprise data, run the following CLI command to verify the integrity of an index or bucket:

```bash
./splunk check-integrity -bucketPath [ bucket path ] [ -verbose ]
```

```bash
./splunk check-integrity -index [ index name ] [ -verbose ]
```

Configure data integrity control

To configure Data Integrity Control, edit indexes.conf to enable the enableDataIntegrityControl attribute for each index. The default value for all indexes is false (off).

```bash
enableDataIntegrityControl=true
```

Data Integrity in clustered environments

In a clustered environment, the cluster master and all the peers must run Splunk Enterprise 6.3 or later to enable accurate index replication.
Optionally modify the size of your data slice

By default, data slices are set to 128kb, which means that a data slice is created and hashed every 128KB. You can optionally edit `indexes.conf` to specify the size of each slice.

```plaintext
rawChunkSizeBytes = 131072
```

Store and secure your data hashes

For optimal security, you can optionally store your hashes outside the system where the data is hosted, such as a different server. To avoid naming conflicts, store your secured hashes in separate directories.

Regenerate hashes

If you lose your hashes for a bucket, Use the following CLI command to re-generate hash files on a bucket or index. This command extracts the hashes embedded in the journal:

```plaintext
./splunk generate-hash-files -bucketPath [ bucket path ] [ verbose ]
./splunk generate-hash-files -index [ index name ] [ verbose ]
```