

Let's Rebalance Data Across An Indexer Cluster In 15 Minutes

Bharath Aleti

Sr. Product Manager, Splunk

.conf2016

splunk >

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Agenda

- Factors leading to uneven data distribution
- What is Data Rebalance
- How to trigger Data Rebalance
- Benefits
- Limitations

Uneven Data Distribution

Contributing Factors

- Addition of new indexers
- Forwarders sending incoming traffic to select indexers
- Multiple forwarders randomly choosing select indexers
- Unbalanced Multisite configuration
- Random node selection by indexer replication
- Node failure or going offline

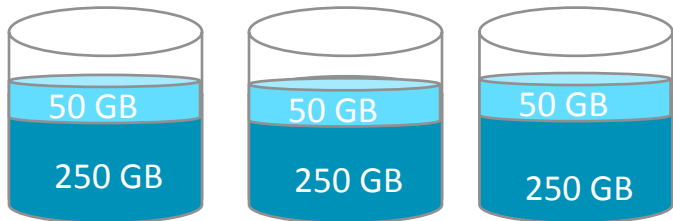
Ramifications

- Higher load on existing indexers
- Poor utilization of new indexers
- Node detention, on reaching max available storage on a single indexer
- Incomplete searches due to nodes in auto-detention state
- Higher per node storage requirement

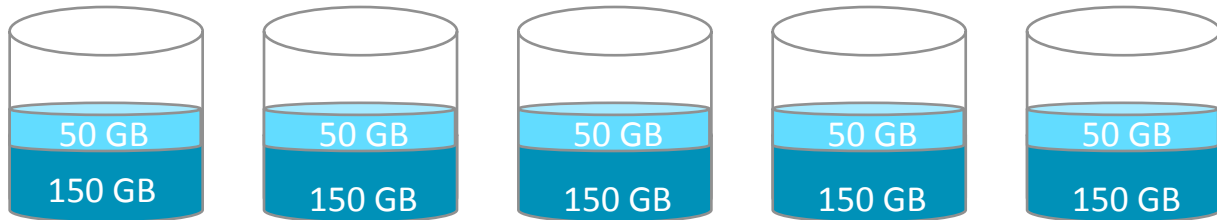
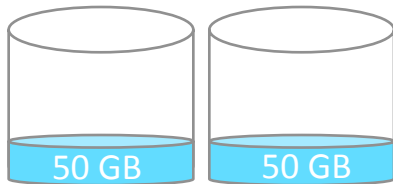
Data Rebalance

Balanced Data Distribution between New and Existing indexers

Existing Indexers



New Indexers



Problem

- *Uneven data distribution after adding new indexers*
- *Higher load on existing indexers*
- *Poor utilization of new indexers*

Even data and search load distribution

- *Lower storage requirement per node*
- *Improved search performance*

Data Rebalance

- Redistributes bucket copies so that each peer has approximately the same number of buckets (within a given threshold)
- Rebalances all (non-searchable, searchable and primary) buckets
- Supports multi-site cluster configurations
 - Data is rebalanced within a site as well as across sites
- Option to monitor the status of data rebalance
- Configurable to tune desired rebalance threshold for optimal storage utilization
- Optional timer to stop rebalance, within a given time period.

Data Rebalance CLI and REST calls

- `splunk rebalance cluster-data -action start [-index index_name] [-max_runtime 100]`
- `splunk rebalance cluster-data -action stop`
- `splunk rebalance cluster-data -action status`

REST Calls

```
curl -k -u admin:changeme https://master:mgmt/services/cluster/master/control/control/rebalance-data -d action=[start/stop/status]
```

```
curl -k -u admin:changeme https://master:mgmt/services/cluster/config/config -d rebalance_threshold=0.9
```

Data Rebalance UI

splunk> Apps ▾ Administrator ▾ 1 Messages ▾ Settings ▾ Activity ▾ Help ▾ Find

Indexer Clustering: Master Node

Edit ▾ More Info ▾ Documentation ↗

No Peers Configured

To learn how to configure peer nodes, refer to the documentation. [Learn More ↗](#)

- Node Type
- Master Node Configuration
- Distribute Configuration Bundle
- Data Rebalance
- Disable Indexer Clustering

Search Heads (1)

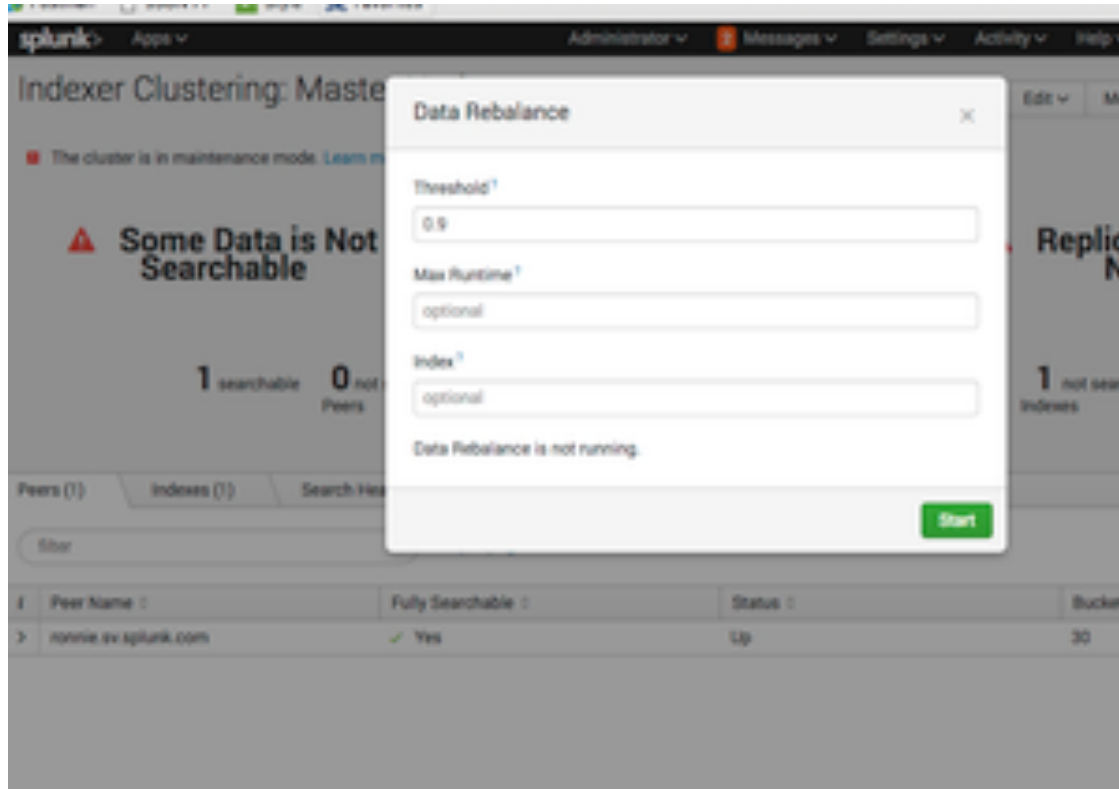
filter 10 per page ▾

i	Search head name ↕	Status ↕
>	dlu-mbp.sv.splunk.com	✓ Up

About Support File a Bug Documentation Privacy Policy

© 2005-2016 Splunk Inc. All rights reserved.

Data Rebalance UI



The screenshot shows the Splunk Data Rebalance configuration interface. A modal window titled "Data Rebalance" is open, displaying the following settings:

- Threshold: 0.9
- Max Runtime: optional
- Index: optional

Below the input fields, it states "Data Rebalance is not running." and includes a green "Start" button.

The background interface shows a warning: "Some Data is Not Searchable" with a red triangle icon. It also displays summary statistics: "1 searchable" and "0 not searchable" Peers. At the bottom, a table lists the cluster peers:

Peer Name	Fully Searchable	Status	Buckets
nonnie.sv.splunk.com	✓ Yes	Up	30

Benefits

- Reduced storage consumption per node, w/ even distribution to all nodes.
- Potential costs savings for cloud/on-prem customers to use instances with smaller storage footprint
- Newly added indexers, with resident primaries, immediately available for new search requests.
- Improved search performance by harnessing I/O throughput across all available indexers

Limitations

- **Optimal but not perfectly balanced**
 - Goal is to achieve a practical balance, not a perfect balance.
- **Based on number of buckets**
 - Data rebalancing balances the number of buckets, not the actual data storage.
A large number of small buckets can skew the actual disk utilization
- **Assumes similar total disk capacity per node**
 - Doesn't account for variable total disk capacity per node at this time
- **Concurrent searches**
 - Because primary bucket transitions can be in-progress, search results are not guaranteed when data rebalance is in progress. Best practice is to run data rebalance during a maintenance window.

THANK YOU

.conf2016