

CollectionRebuilding

Collection Rebuilding

Collection rebuilding is creating an index from scratch (not incremental updates). A full rebuild in which a new collection replaces the old collection would be required in cases such as the following:

- When building a new collection with no previous collection data existing.
- When launching something new.
- When a collection has become corrupted for some reason.
- When redefining an existing field typechanging your schema in a way that requires a rebuild. For example, merely adding fields to the schema does not require a rebuild, but changing the type of a field would.

A Procedure for New Index Building with rsync based index replication

Perform the procedure below from the master server to do collection rebuilds in a production environment.

1. Turn off distribution by running **rsyncd-stop**. This prevents the slaves from getting data from the master.
Note: Ensure that a distribution is not running when you run rsyncd-stop.
2. Run the script, **abc** (Atomic Backup post-Commit), to create a snapshot for a safe backup.
3. If you have a separate process that does incremental updating that might come in while you are performing this procedure, you may want to disable it.
4. Remove the index directory, **/solr/data/index/**, on the master server.
5. If you have changes to the schema or any new configurations to be installed, stop the server. Make the changes to the schema/configurations and install them.
6. Restart the server.
7. Re-index all of your documents.
8. Run the script, **optimize**, to optimize the collection.
9. Re-enable index distribution with the **rsyncd-start** script. The new collection data will be pulled by the slaves while still serving requests.

Note: If you have configured Solr to take snapshots only for optimized indices, and have an index builder that only issues optimize commands when the index is completely rebuilt, you can skip steps dealing with disabling distribution.

Alternative Approaches for New Index Building

- Create an "offline" solr port, index from scratch on the offline port, disable snapshot pulling, shut down the master, copy the index from the offline port to the master, enable snapshot pulling.
- Create an "offline" solr port, index from scratch on the offline port, disable snapshot pulling, shut down the master, copy the index from the offline port to the master, disable slave boxes one-at-a-time and copy the index to each manual, enable snapshot pulling. (This last one in particular requires a lot more setup time and thought.)