

Vladimir Rodionov, Feb 2017

Summary of work

- The new feature introduces new command-line extensions to the hbase command and, from the client side, is accessible through command-line only
- Operations:
 - Create full backup on a list of tables or backup set
 - Create incremental backup image for table list or backup set
 - Restore list of tables from a given backup image
 - Show current backup progress
 - Delete backup image and all related images
 - Show history of backups
 - Backup set operations: create backup set, add/remove table to/from backup set, etc
- In the current implementation, the feature is already usable, meaning that users can backup tables and restore them using provided command-line tools. Both: full and incremental backups are supported.
- The full list of JIRAs included in this mega patch can be found in three umbrella JIRAs: HBASE-14030 (Phase 1), HBASE-14123 (Phase 2) and HBASE-14414 (Phase 3 - all resolved ones)

What are the remaining work items

All remaining items can be found in Phase 3 umbrella JIRA: [HBASE-14414](#).

They are split into 3 groups: BLOCKER, CRITICAL, MAJOR

Only BLOCKERs and CRITICALs are guaranteed for HBase 2.0 release.

BLOCKER

[HBASE-14417](#) **Incremental backup and bulk loading (Patch available).**

This will allow to run incremental backups on a bulk loaded data

[HBASE-14135](#) **HBase Backup/Restore Phase 3: Merge backup images.**

Merging multiple backup images (full and incrementals) to create a single backup image. Maintenance job, which allows to control number of backup files in a system and improves restore operation time.

[HBASE-14141](#) **HBase Backup/Restore Phase 3: Filter WALs on backup to include only edits from backup tables (Patch available).**

Filtering WAL files on backup is necessary from both: security and space usage point of views. If users backups only subset of tables - it does not make sense to backup WAL files in entirety.

[HBASE-17133](#) **Backup documentation**

HBase book section

[HBASE-15227](#) **Fault tolerance support**

Make system tolerant to backup/restore failures.

CRITICAL

[HBASE-16465](#) **Disable split/merges during backup**

Disabling table split/merges during backup improves overall chances of an operation success. Split or merge during full backup may result in backup failure.

Misc

We have umbrella JIRA (HBASE-14414) to track all the remaining work

All the BLOCKER and CRITICAL JIRAs currently in open state will be implemented by 2.0 release time. Some MAJOR too, but it depends on resource availability

The former development branch (HBASE-7912) is obsolete and will be closed/deleted after the merge.

We want backup to be a GA feature in 2.0

We are going to support full backward compatibility for backup tool in 2.0 and onwards.

Configuration

Backup is disabled, by default. To enable it, the following configuration properties must be added to hbase-site.xml:

```
hbase.backup.enable=true
hbase.master.logcleaner.plugins=YOUR_PLUGINS,org.apache.hadoop.hbase.
backup.master.BackupLogCleaner
hbase.procedure.master.classes=YOUR_CLASSES,org.apache.hadoop.hbase.b
ackup.master.LogRollMasterProcedureManager
hbase.procedure.regionserver.classes=YOUR_CLASSES,org.apache.hadoop.h
base.backup.regionserver.LogRollRegionServerProcedureManager
```

Known Issues

1. Bulk loaded data is not visible for incremental backup (will be resolved by 2.0 release)
2. After first backup sessions, WAL files will be accumulated in old WAL directory until the next backup sessions. Make sure, that you either run backups **regularly** or disable backups in the system.
3. No fault tolerance yet, some backup failures may result in corrupted backup data and will require manual intervention (will be resolved by 2.0 release)