HBASECON ASIA2019

THE COMMUNITY EVENT FOR APACHE HBASE™

HBCK2: Concepts, trends and recipes for fixing issues within HBase 2

Wellington Chevreuil

HBase Committer Cloudera HBase SW Engineer

HBCK (1) - Little bit of history

- Main tool for general inconsistencies in **hbase-1.x**
  - The *Swiss Knife* for operators
- Packaged together with **hbase** main project
- Provides both *diagnosing* and *fixing* commands
  - Some reports may be misleading, e.g., *"holes in the region chain"*
  - Some options can cause damages if not well understood, e.g., *"-sidelineBigOverlaps", "-removeParents"*
- Commands often work independent of Master
  - Can introduce conflicts on meta information maintained by Master
- Lack of implementation details on documentation/help guide

HBCK1 Commands user guide:

```
Usage: fsck [opts] {only tables}
 where [opts] are:
   -help Display help options (this)
   -details Display full report of all regions.
   -timelag <timeInSeconds>  Process only regions that  have not experienced any metadata updates in the last  <timeInSeconds> seconds.
   -sleepBeforeRerun <timeInSeconds> Sleep this many seconds before checking if the fix worked if run with -fix
   -summary Print only summary of the tables and status.
   -metaonly Only check the state of the hbase:meta table.
   -sidelineDir <hdfs://> HDFS path to backup existing meta.
   -boundaries Verify that regions boundaries are the same between META and store files.
   -exclusive Abort if another hbck is exclusive or fixing.

  Metadata Repair options: (expert features, use with caution!)
   -fix              Try to fix region assignments.  This is for backwards compatiblity
   -fixAssignments   Try to fix region assignments.  Replaces the old -fix
   -fixMeta          Try to fix meta problems.  This assumes HDFS region info is good.
   -noHdfsChecking   Don't load/check region info from HDFS. Assumes hbase:meta region info is good. Won't check/fix any HDFS issue, e.g. hole, orphan, or overlap
   -fixHdfsHoles     Try to fix region holes in hdfs.
   -fixHdfsOrphans   Try to fix region dirs with no .regioninfo file in hdfs
   -fixTableOrphans  Try to fix table dirs with no .tableinfo file in hdfs (online mode only)
   -fixHdfsOverlaps  Try to fix region overlaps in hdfs.
   -fixVersionFile   Try to fix missing hbase.version file in hdfs.
   -maxMerge <n>     When fixing region overlaps, allow at most <n> regions to merge. (n=5 by default)
   -sidelineBigOverlaps  When fixing region overlaps, allow to sideline big overlaps
   -maxOverlapsToSideline <n>  When fixing region overlaps, allow at most <n> regions to sideline per group. (n=2 by default)
   -fixSplitParents  Try to force offline split parents to be online.
   -removeParents    Try to offline and sideline lingering parents and keep daughter regions.
   -ignorePreCheckPermission  ignore filesystem permission pre-check
   -fixReferenceFiles  Try to offline lingering reference store files
   -fixEmptyMetaCells  Try to fix hbase:meta entries not referencing any region (empty REGIONINFO_QUALIFIER rows)

  Datafile Repair options: (expert features, use with caution!)
   -checkCorruptHFiles     Check all Hfiles by opening them to make sure they are valid
   -sidelineCorruptHFiles  Quarantine corrupted HFiles.  implies -checkCorruptHFiles

  Metadata Repair shortcuts
   -repair           Shortcut for -fixAssignments -fixMeta -fixHdfsHoles -fixHdfsOrphans -fixHdfsOverlaps -fixVersionFile -sidelineBigOverlaps -fixReferenceFiles -fixTableLocks -fixOrphanedTableZnodes
   -repairHoles      Shortcut for -fixAssignments -fixMeta -fixHdfsHoles

  Table lock options
   -fixTableLocks    Deletes table locks held for a long time (hbase.table.lock.expire.ms, 10min by default)

  Table Znode options
   -fixOrphanedTableZnodes    Set table state in ZNode to disabled if table does not exists

 Replication options
   -fixReplication   Deletes replication queues for removed peers
```

HBCK2 in a nutshell

- Simpler tool
  - Less fix commands
  - No diagnosis command
  - Requires deeper HBase internal workings from operators
- Shipped independently from hbase
  - Packaged with hbase-operators-tool project
    - **https://github.com/apache/hbase-operator-tools**
  - Can evolve on its own pace
  - New versions can be run without needing whole hbase upgrade
- Master oriented (more later)
- More detailed documentation about each command
- Still a WIP
  - By the time of this presentation, there's still no official release for HBCK2

HBCK2 Concepts

- **AMv2** compliant
  - HBCK1 does not work with HBase 2 AssignmentManager re-implementation
- **Thinner**, but more interactive commands
  - No such thing as hbck1 -fix command
  - Operators required to fix an issue at a time
- **Master** oriented
  - Master must be online
  - Commands implementation should use Master HbckService as much as possible
  - However, new commands may initially require a client side implementation, then get ported to Master's HbckService facade
- **Fix only**, requires other tools for issue diagnosing
- Available only for **2.0.3** onwards, and **2.1.1** onwards

HBASECON ASIA2019

HBCK2 Commands user guide:

```
Commands:
 assigns [OPTIONS] <ENCODED_REGIONNAME>...
   Options:
    -o,--override  override ownership by another procedure
   A 'raw' assign that can be used even during Master initialization
   (if the -skip flag is specified). Skirts Coprocessors. Pass one
   or more encoded region names. 1588230740 is the hard-coded name
   for the hbase:meta region and de00010733901a05f5a2a3a382e27dd4 is
   an example of what a user-space encoded region name looks like.
   For example:
     $ HBCK2 assign 1588230740 de00010733901a05f5a2a3a382e27dd4
   Returns the pid(s) of the created AssignProcedure(s) or -1 if none.

 bypass [OPTIONS] <PID>...
   Options:
    -o,--override   override if procedure is running/stuck
    -r,--recursive  bypass parent and its children. SLOW! EXPENSIVE!
    -w,--lockWait   milliseconds to wait before giving up; default=1
   Pass one (or more) procedure 'pid's to skip to procedure finish.
   Parent of bypassed procedure will also be skipped to the finish.
   Entities will be left in an inconsistent state and will require
   manual fixup. May need Master restart to clear locks still held.
   Bypass fails if procedure has children. Add 'recursive' if all
   you have is a parent pid to finish parent and children. This
   is SLOW, and dangerous so use selectively. Does not always work.

 unassigns <ENCODED_REGIONNAME>...
   Options:
    -o,--override  override ownership by another procedure
   A 'raw' unassign that can be used even during Master initialization
   (if the -skip flag is specified). Skirts Coprocessors. Pass one or
   more encoded region names. 1588230740 is the hard-coded name for
   the hbase:meta region and de00010733901a05f5a2a3a382e27dd4 is an
   example of what a userspace encoded region name looks like.
   For example:
     $ HBCK2 unassign 1588230740 de00010733901a05f5a2a3a382e27dd4
   Returns the pid(s) of the created UnassignProcedure(s) or -1 if none.

 setTableState <TABLENAME> <STATE>
   Possible table states: ENABLED, DISABLED, DISABLING, ENABLING
   To read current table state, in the hbase shell run:
     hbase> get 'hbase:meta', '<TABLENAME>', 'table:state'
   A value of \x08\x00 == ENABLED, \x08\x01 == DISABLED, etc.
   Can also run a 'describe "<TABLENAME>"' at the shell prompt.
   An example making table name 'user' ENABLED:
     $ HBCK2 setTableState users ENABLED
   Returns whatever the previous table state was.

 setRegionState <ENCODED_REGIONNAME> <STATE>
   Possible region states:
      OFFLINE, OPENING, OPEN, CLOSING, CLOSED, SPLITTING, SPLIT,
FAILED_OPEN, FAILED_CLOSE, MERGING, MERGED, SPLITTING_NEW, MERGING_NEW,
ABNORMALLY_CLOSED
   WARNING: This is a very risky option intended for use as last resort.
   Example scenarios include unassigns/assigns that can't move forward
   because region is in an inconsistent state in 'hbase:meta'. For
   example, the 'unassigns' command can only proceed if passed a region
   in one of the following states: SPLITTING|SPLIT|MERGING|OPEN|CLOSING
   Before manually setting a region state with this command, please
   certify that this region is not being handled by a running procedure,
   such as 'assign' or 'split'. You can get a view of running procedures
   in the hbase shell using the 'list_procedures' command. An example
   setting region 'de00010733901a05f5a2a3a382e27dd4' to CLOSING:
     $ HBCK2 setRegionState de00010733901a05f5a2a3a382e27dd4 CLOSING
   Returns "0" if region state changed and "1" otherwise.
```

HBCK2 Usage trends

- **Master not completing initialisation**
  - Meta/Namespace table "NOT online" issues
- **Table RIT issues**
- **Procedures stuck**
- **Table in wrong state**
- **Missing regions in META**
  - User induced via incompatible *OfflineMetaRepair* tool

HBCK2 for Operators: How do I get and run it?

- **Not released so far, requires local build**
- **Requirements**
  - JDK 1.8 or higher
  - Git
  - Maven
- **Checkout related apache github repository:**
  - $ git clone https://github.com/apache/hbase-operator-tools.git
- **Build HBCK2 upon desired hbase version:**
  - $ mvn -Dhbase.version=2.1.5 clean install
  - Above command will produce HBCK2 jar file under ./hbase-hbck2/target/, named hbase-hbck2-1.0.0-SNAPSNOT.jar (assuming current version is 1.0.0-SNAPSHOT)
- **Upload generated jar to the given hbase cluster and run it as below:**
  - $ hbase hbck -j ../hbase-hbck2-1.0.0-SNAPSHOT.jar

HBCK2 for Operators: Recipes

- **Meta/Namespace table regions "NOT online"**
  - Due to corruption or manual deletion of /hbase/MasterProcWALs files
  - Meta may miss info about RS assignment
  - Master logs show regions assigned to an old RS start code

```
WARN org.apache.hadoop.hbase.master.HMaster: hbase:meta,,1.1588230740 is NOT online; state={1588230740 state=OPENING, ts=1550754721289, server=regionserver01.example.com,16020,1550676598448}; ServerCrashProcedures=true. Master startup cannot progress, in holding-pattern until region onlined.
```

- **Run HBCK2 assigns command for META region 1588230740:**
  - **$ hbase hbck -j ../hbase-hbck2-1.0.0-SNAPSHOT.jar assigns 1588230740**
- **Similar issue may affect namespace and user tables regions**
  - Affected regions names would be mentioned on log messages similar to above

HBCK2 for Operators: Recipes

- **Table RIT issues**
  - **Usually, due several RSes crashes/slowness while regions are transitioning**

```
WARN org.apache.hadoop.hbase.master.assignment.AssignmentManager: STUCK Region-In-Transition rit=OPENING,
location=regionserver01.example.com,16020,1542314816394, table=hbase:acl, region=11bf6b18ddacdd864728e6cf1199b2a7
...
WARN org.apache.hadoop.hbase.ipc.RpcServer: Dropping timed out call: callId: 702 service: ClientService methodName: Mutate size: 272
connection: 1.1.1.1:56492 deadline: 1542316740911 param: region= hbase:meta,,1,
row=hbase:acl,,1404406671604.11bf6b18ddacdd864728e6cf1199b2a7. connection: 1.1.1.1:56492
```

- **Run HBCK2 assigns command for the given region encoded name 11bf6b18ddacdd864728e6cf1199b2a7:**
  - **$ hbase hbck -j ../hbase-hbck2-1.0.0-SNAPSHOT.jar assigns 11bf6b18ddacdd864728e6cf1199b2a7**

HBCK2 for Operators: Recipes

- **Procedures stuck**
  - While troubleshooting causes for RITs, check for procedures attempting to transition regions states:
  - **$ echo "list_procedures" | hbase shell**
- **Output for list_procedures shows WAITING_TIMEOUT and/or procedures running for days**

```
PID Name State Submitted Last_Update Parameters
6 org.apache.hadoop.hbase.master.assignment.UnassignProcedure WAITING_TIMEOUT 2019-03-29 11:15:06  2019-04-08 06:33:35  ...
7 org.apache.hadoop.hbase.master.procedure.DeleteTableProcedure RUNNABLE 2019-03-29 11:24:39 2019-03-29 11:24:39 ...
```

- **Other procedures fail to acquire lock owned by one of the stuck procedures:**

```
ERROR: org.apache.hadoop.hbase.procedure2.ProcedureAbortedException: f7910bfc9c9... owned by pid=6, CANNOT run 'this' (pid=347).
```

- **Run HBCK2 bypass command to get rid of stuck procedures:**
  - **$ hbase hbck -j ../hbase-hbck2-1.0.0-SNAPSHOT.jar bypass 6 7**

HBCK2 for Operators: Recipes

- **Table in wrong state**
  - Can happen after hanging enable/disable table procedures, or related sub-procedures
  - Bypassing procedures can lead to this as well
  - Table indefinitely in temporary states ENABLING/DISABLING
    - **scan 'hbase:meta', {COLUMN => "table:state"}**

```
usertable column=table:state, timestamp=1555406568751, value=\x08\x03.
```

    - **enable 'usertable'**

```
ERROR: Table tableName=usertable, state=ENABLING should be disabled!
```

- **Run HBCK2 setTableState to manually bring table state to one of the final ones ENABLED/DISABLED:**
  - **$ hbase hbck -j ../hbase-hbck2-1.0.0-SNAPSHOT.jar setTableState usertable DISABLED**

HBCK2 for Operators: Recipes

- **Missing regions in META**
  - Operator induced when running incompatible tool *OfflineMetaRepair* (HBASE-21665)
  - Typically manifests as holes on the region chain, or in the case of namespace region missing, master fails initialisation
    - scan 'hbase:meta', {COLUMN => "table:state", ROWPREFIXFILTER => 'hbase:namespace'}

```
ROW                                    COLUMN+CELL
0 row(s)
```

- **Still under development through HBASE-22567, HBCK2 addMissingRegionsInMeta can be used to re-add missing regions:**
  - $ hbase hbck -j ../hbase-hbck2-1.0.0-SNAPSHOT.jar addMissingRegionInMeta hbase:namespace
- **Still WIP, so syntax might change.**
- **Check HBASE-22567 for latest developments**

HBCK2 for Contributors

- **Apache github repository: https://github.com/apache/hbase-operator-tools**
  - **HBCK2 defined as sub-module hbase-hbck2 of hbase-operator-tools**
- **HBASE-21745**
  - **Umbrella jira for tracking potential new HBCK2 features**
- **Faced a new issue in HBase 2? Have a new idea for HBCK2 command?**
  - **Great! Contributions are welcome!**
  - **Start a [DISCUSS] mail thread on dev@hbase.apache.org**
  - **Post a comment on HBASE-21745 describing your idea**

Thanks !