

Procedure v2/Notification-Bus Roadmap

[HBASE-12439](#) Procedure V2

Matteo Bertozzi

(This was a RoadMap for the ProcedureV2 project that Matteo used to keep track kept internally @ his place of employ. Surfaced in HBASE-12439 with some edits by St.Ack. This document is now sealed and was current as of December 2016).

[1 Apache HBase 1.1](#)

[1.1 What went wrong?](#)

[1.2 Pending discussions](#)

[1.3 The good stuff](#)

[2 Apache HBase 1.2](#)

[3 Apache HBase 1.3](#)

[4 Apache HBase 2.0](#)

[5 Apache HBase 2.1+](#)

1 Apache HBase 1.1

In Apache HBase 1.1, we will have *Fault tolerant Master Operations* for the main DDL operations (e.g. create/delete/modify...). The 1.1 client will also have the benefit of the new proc-v2 framework making the sync operation synchronous for real (we are still wire compatible, both ways). One nice benefit from the new proc-v2 rewrite of the handlers is that now we have much more unit-test coverage for the DDL operations.

- Phase 1 - Making DDL Operations fault tolerant
 - [HBASE-13202](#) - core framework (*matteo*)
 - *Umbrella issue committed to branch-1.1. Adds the hbase-procedure module.*
 - HBASE-13529 - WAL Improvements (*matteo*)
 - *Procedure WAL improvements. Committed to branch-1.1.*
 - HBASE-13203 - master create/delete table (*matteo*)
 - *Makes create/delete procedure-based. Master-side. In branch-1.1*

- HBASE-13204 - client create/delete table sync (*matteo*)
 - *Client-side of create/delete.*
- HBASE-13209 - master Add/Modify/Delete Column Family (*stephen+matteo*)
 - *Makes these procedure-based. Server-side.*
- HBASE-13210 - master Modify table (*stephen+matteo*)
 - *Server-side replace of modify table w/ procedure-versions.*
- HBASE-13211 - master Enable/Disable table (*stephen+matteo*)
 - *Enable/Disable procedures on server-side.*
- HBASE-13290 - client enable/disable table sync (*stephen+matteo*)
 - *Client-side.*
- HBASE-13455 - master truncate table (*matteo*)
 - *Master-side procedure version.*
- HBASE-13470 - High level Integration test for master DDL operations (*sophia*)
- HBASE-13551 - Adjust the InterfaceAudience for hbase-procedure (*enis*)
- Phase 1 - branch-1.1 only
 - HBASE-13469 - Make procedure v2 configurable, allow to turn off. (*stephen*)
- Phase 1 - branch-1.2+/master cleanups
 - HBASE-13536 - Cleanup the handlers that are no longer being used (*srikanth*)
 - **HBASE-13823** - Remove Enable/Disable handlers from AM (*matteo/stephen?? - 5day*)
 - *This did not make 1.2. It went into 2.0.*
- Phase 1 - master only cleanups (incompatibilities)
 - HBASE-13537 - Change Admin async operations to return Future (*srikanth*)
 - *Made it into 2.0. Breaks binary compat. Src compat is retained.*
 - **HBASE-13538** - use proc-v2 async for add/delete/modify family (*srikanth???* - 1h)
 - *In 2.0. Finished by Ashish Singhi.*
 - HBASE-13571 - use proc-v2 async for modify table (*srikanth - 30min*)
 - *In 2.0.*
 - HBASE-13572 - use proc-v2 async for truncate table (*ashish*)
 - *In 2.0. Finished by Ashish Singhi.*
 - **HBASE-XXXX** - replace TimeoutIOException with TimeoutException?
 - *Suggestion by Matteo. Not the End of the World.*
- Phase 1 - in-evaluation/in-doubt
 - ⊖ **HBASE-13260** Bootstrap tables as proc-wal replacement (*enis*)
 - ~~*A subset of this patch is to be pushed but it has not been done.*~~

1.1 What went wrong?

The core framework had the first patch posted Nov 6th. Took 5 months to get in.

- No reviews, mostly renames. the final code was essentially the same one posted in the first patch.
- Meaningful discussion about state machine design got posted only few days before merge.
- Simple one-line bugs found only after merge (mainly WALProcedureStore, ProcedureStoreTracker)

The initial create/delete and proc-v2 master side took 3 months to get in.

- In this case stephen was implementing other procedures more or less copy/pasting this work, but he provided feedback and we cleaned up things along the way. (A lot of back and forth to adapt to new changes and cleanup, but good work/interaction here).
- the last procedure to be implemented took only 5 days to get in (mainly the time to get a review), so new procedures with the base stabilized are quite easy and quick to get in.

1.2 Pending discussions

HBASE-13260 *"Bootstrap Tables for fun and profit"* is the main core-changes ~~still in discussion~~.

The idea is to replace the Procedure-WAL using a special Region hosted by the master (skipping assignment & co). The problem with this is that it is slower on write (not good for AM rewrite, since we will write to the Store a lot) and implementing replay tricks and shortcut will not be that easy as it is for the simple Procedure-WAL. In the end, HBASE-13260 was scratched because slow.

TODO: The coprocessor semantic on rollback is still not defined, so when we rollback an operation, e.g. on table creation failure we still don't call the deleteTable coprocessor so the coprocessor state may be "corrupted". It is basically the same bad behavior we have today, but since there were no real discussion around this thing, defining the proper semantic/behavior was moved to the next release.

1.3 The good stuff

- Now we have a proper sync client (the proper async interface with future is 2.0 only). create/delete no longer have to poll for a result and relay on the order of the operations

on the master side. (We are still compatible both ways, of course old client has the same not sync behavior).

- If the master goes down in the middle of a DDL operation (e.g. kill, shutdown or failover), we are able to fix the state with a restart. no hbck needed. Only one master at the time can execute operations (we have a lease on the WAL like the RSs), on master activation the WAL will be replayed completing the operation in progress on the previously active master.
- The client submitting an operation and waiting for the response will survive and get a proper response on master failover.
- we have more testing around DDL operations, and adding more fault injection on them is not that complicated.

2 Apache HBase 1.2

1.2 got few improvement and fixes. the main user-facing thing we got in 1.2 is the ability to monitor the running procedures via shell commands.

- HBASE-13759 - Improve procedure yielding
- HBASE-13415 - Use nonces for double submits from client (stephen)
- HBASE-13832 - try to roll the WAL master on sync failure before aborting (matteo)
- HBASE-13993 - WALProcedureStore fencing is not effective if new WAL rolls (enis)
- HBASE-13476 - Add Replay Order logic for child procedures (~3day)
- HBASE-14432 - Procedure V2 - enforce ACL on procedure admin tasks
- HBASE-14488 - Procedure V2 - shell command to abort a procedure
- HBASE-14487 - Procedure V2 - shell command to list all procedures

3 Apache HBase 1.3

In Apache 1.3 the focus it will probably be improving the supportability of the feature, with more debug adding the procedure tracking in the WebUI.

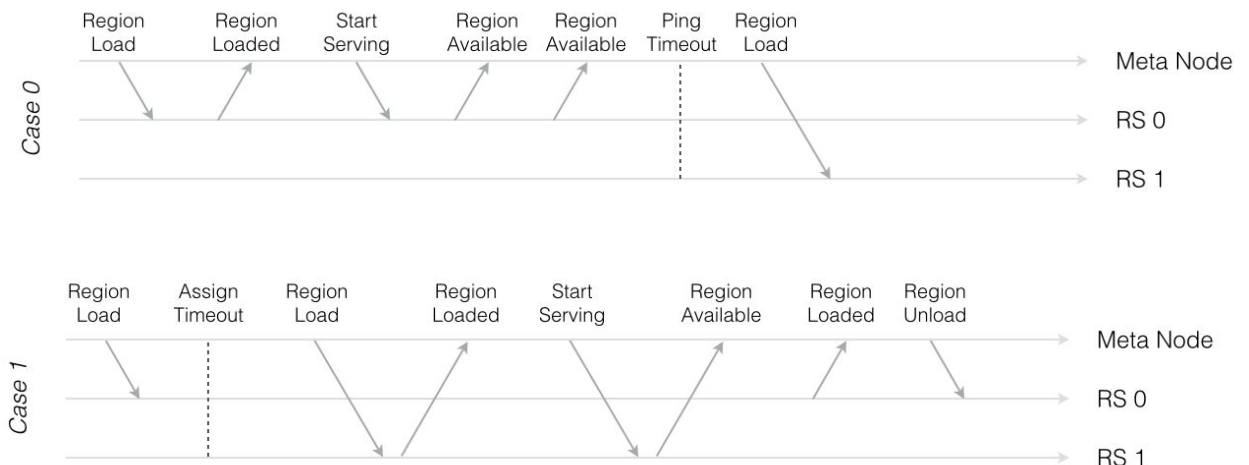
- **HBASE-15141** - Master WebUI (~5days)
 - HBASE-14719 - Add metric for number of MasterProcWALs
 - HBASE-15142 - Procedure v2 - Basic WebUI listing the procedures
 - HBASE-15144 - Procedure v2 - Web UI displaying Store state
 - **HBASE-15143** - Procedure v2 - Web UI displaying queues

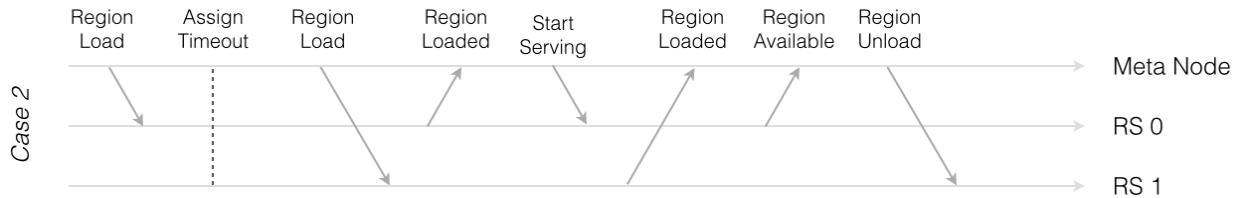
- **HBASE-16102** - Add WAL pretty print tool (and print load graph)
- **HBASE-XXXX** - Add Perf Test for Wal and Sched
- **HBASE-XXXX** - Add proc stats, how many restarts/yield/retriable-failures (~1days)
- HBASE-13212 - master Create/Modify/Delete namespace (stephen)
- **HBASE-XXXX** - Add compression/crypto to procedure WAL
- HBASE-15113 - Procedure v2 - Speedup eviction of sys operation results
- Queues
 - HBASE-14837 - Procedure V2 - Procedure Queue Improvement
 - HBASE-15105 - Procedure V2 - Procedure Queue with Namespaces
 - HBASE-15107 - Procedure V2 - Procedure Queue with Regions

4 Apache HBase 2.0

Defining a better semantic for rollbacks and coprocessors and improving the client side. We still have some handlers not converted in procedures and the main new core implementation here will be the notification-bus that will be used for all the coordinated operations (e.g. snapshot, assignment, ...)

- **HBASE-XXXX** - define a semantic for coprocessor rollback, and fix DDL procedures. (ref: [HBASE-6992](#), HBASE-15575)
- Assignment Manager
 - **HBASE-14350** - Replace Assignment Manager (~3 months) -- *Below diagram is now obsolete. To be replaced!!!*





- HBASE-XXXX - ServerManager rewrite
 - Balancer DS fixup
- HBASE-XXXX - AM rewrite
- Notification BUS - Phase 1 (*still compatible with Procedure-v1*)
 - HBASE-XXXX - Implement proc-v2 OnePhaseProcedure (matteo ~2weeks)
 - HBASE-XXXX - Rewrite Snapshot coordination with OnePhaseProc (matteo, ~3days)
 - HBASE-15521 - Rewrite Restore/Clone SnapshotHandlers (stephen, ~1day)
 - Need another rewrite when we do the new FS
 - HBASE-XXXX - Rewrite Flush coordination with OnePhaseProc (~1day)
- master only cleanups (incompatibilities)
 - HBASE-XXXX - use proc-v2 async namespace operations (~1h)
- Notification BUS - Phase 2
 - HBASE-XXXX - Implement proc-v2 TwoPhaseProcedure (matteo, ~2weeks)
 - HBASE-13687 - Rewrite ACL cache updates (matteo, ~1day)
 - HBASE-13688 - Rewrite Visibility cache updates (anoop? ~1h)
 - HBASE-XXXX - Rewrite Quotas cache updates (~1h)
-
- HBASE-XXXX - Cleanup One/TwoPhaseProcedure using new AM sched/notifications

5 Apache HBase 2.1+

- HBASE-XXXX - Replace Distributed Log Replay (? months)
- Future
 - HBASE-XXXX - Master ops Replication (e.g. grant/revoke)
 - HBASE-XXXX - Master Coordinated compactions