

Development Contributors Meetings

Hive Contributors Minutes 110425

Meeting date: April 25, 2011

Location: Facebook Palo Alto

Attendees: <http://www.meetup.com/Hive-Contributors-Group/events/17272914/>

The 0.7 release is out, and Carl proposed an 0.7.1 release for items such as PostgreSQL metastore upgrade scripts and Maven artifact publication. Rules for a point release were discussed: no metastore changes, and no changes to API's such as Thrift and extension interfaces. Everyone was fine with this; Carl will manage the release.

Plans for 0.8 were also discussed; optimistic target is two months from now.

There were some user-oriented questions about join optimization, which is an area of much confusion; besides better documentation, it would be useful to start having User Meetups again, as well as to review existing config parameter defaults as part of each release.

Alan gave an update on HCatalog (formerly Howl):

- incubation has been approved and is under way (svn, JIRA and mailing list are available; working on publishing website)
- for svn, Hive code inclusion is via external reference (not a separate copy)
- currently only works with secure Hadoop
- 0.1 branch will be cut in a few weeks
- 0.2 is targeted for end of June, with support for Hadoop Streaming as well as some kind of blob support

Alan is going to work on publishing the HCatalog roadmap.

John opened a discussion on Hive Contributor Day (June 30, 10am - 6pm) as part of the Yahoo Hadoop Summit in Sunnyvale. The rough agenda is (1) presentations in the morning, followed by (2) hands-on UDF hacking in the afternoon. We'll be soliciting proposals for presentations. For the afternoon, we would like to work on a "Hive SDK" which would make it easier for users to develop extensions such as UDF's without having to set up a full Hive build environment. Facebook data science already has some useful bits built up in this area, so we're going to work on open-sourcing that on GitHub. In addition, they also have a library of private UDF's which should be open sourced, so we're going to try to farm out this effort and use it as a learning experience for people who want to develop their own UDF's.

John gave a status update on the indexing support being worked on by the Harvey Mudd Clinic students; bitmap indexing has been committed, and automatic index usage is very close, but it will still be in a very "experts-only" state, so a lot of followup work remains.

Carl brought up the need to improve the maintainability of Hive's test suite, and John mentioned a number of other maintenance projects needed such as checkstyle enforcement and removing Thrift-generated code. The challenge with these is that the patches tend to be very large, and by the time they're ready, they need to be rebased, ad infinitum. The proposed solution is to (1) get an assignee and committer paired up and dedicated (2) vote on a "quiet period" via the dev mailing list and (3) carry out the work during that period. We can give that a try and see if we can make progress.

John mentioned some summer intern projects being considered at Facebook (list to be added to wiki), and asked for further suggestions.

There was discussion around using some Yahoo QA machines (or the OSUOSL cluster) for automatic patch validation; Carl got some contact info and is going to look into that more.

We ended with some review and discussion of HIVE-2038 (metastore listener). It was decided that generic events are going to be dealt with in a followup JIRA.