

HadoopContributorsMeeting20100506

Notes by Arun C. Murthy

- Shared goals
 - Hadoop is HDFS & Map-Reduce in this context of this set of slides
- Priorities
 - Yahoo
 - Correctness
 - Availability: Not the same as high-availability (6 9s. etc.) i.e. SPOFs
 - API Compatibility
 - Scalability
 - Operability
 - Performance
 - Innovation
 - Cloudera
 - Test coverage, api coverage
 - APL Licensed codec (lzo replacement)
 - Security
 - Wire compatibility
 - Cluster-wide resource availability
 - New apis (FileContext, MR Context Objs.), documentation of their advantages
 - HDFS to better support non-MR use-cases
 - Cluster metrics hooks
 - MR modularity (package)
 - Facebook
 - Correctness
 - Availability, High Availability, Failover, Continuous Availability
 - Scalability
- Bar for patches/features keeps going higher as the project matures
 - Build consensus (e.g. Python Enhancement Process, JSR etc.)
 - Run/test on your own to prove the concept/feature or branch and finish
 - Early versions of libraries should be started outside of the project (github etc.) e.g. input-formats, compression-codecs etc.
 - github for all the above
 - Prune contrib
- Maven for packaging
- Tom: Hadoop Common/HDFS/Mapreduce 0.21 release
 - Rebranch
 - Close blockers
 - Definition? Deadline? Please help!
 - <http://bit.ly/common21blockers>
 - <http://bit.ly/hdfs21blockers>
 - <http://bit.ly/mapreduce21blockers>
 - Create builds
 - Post-split considerations: #artifacts, bin, conf, docs
 - Test
 - Community driven, wiki
 - Vote and release
 - Caveats: not stable
- Owen: Release Manager (see [slides](#))
- Agenda for next meeting
 - Eli: Hadoop Enhancement Process (modelled on PEP?)
 - Branching strategies: Development Models