SystemTestingConfCallSynopsis

This page provides a synopsis of the conference call held to discuss HADOOP-6332

Attendees

Konstantin Boudnik (Yahoo), Alex Loddengaard (Cloudera), Steve Loughran (HP), Steve Watt (IBM), Stuart Hood (Rackspace)

Definition

Cluster QA - Provide Quality Assurance by verifying with runtime testing that a Hadoop Cluster is setup and working correctly. At present there is no way to do this. There are existing tests like TeraSort but they only test certain aspects of the system.

Requirements

Each attendee described their requirements and what they would like to see from the end product.

General Consensus : There is a user need for regression testing (* on what specifically?*)

Steve Loughran :

- Dynamic cluster testing
- Cluster QA

Steve Watt :

- Verify that Hadoop works on IBM Java and IBM distributions such as RHEL and SLES.
- Cluster QA
- Reduce QA time (The existing functional tests take too long and distributed execution might reduce the overall time)

Alex Loddengaard:

- Cluster QA

Kostantin Boudnik:

- Cluster QA

Initial Proposal

- Deployment is a separate concern from Testing. Deployment is outside of the scope of this JIRA. Users will be responsible for setting up their cluster.
- Deployment information is provided as a set of parameters (possibly environment variables) to the testing framework.
- Create a 'Cluster QA' set of JUNIT based tests. Start by selecting several of the existing functional tests and port them to run over the cluster.
- Contain the tests inside a hadoop-version-test.jar (or alternatively separate jars for each project split, i.e. mapred-test, hdfs-test etc.)
- Design the framework such that is able to take the parameters specified and run the tests in the test jar specified against the cluster identified.
- Once the cluster is running the tests could be run against the default cluster with something like : "bin/hadoop -verify"
- Start collecting data on how long system tests (such as TeraSort) take to run on certain cluster configurations so we can provide users a benchmark they can use to validate the health of their cluster.

Next Steps

- Iterate in public (publish this Wiki Link and the Meeting Wiki Page to the mailing lists to solicit feedback)
- Get community engagement and consensus on approach
- Provide first drop of code

Available Resources

- Stephen Watt (70 % of his time)
- Konstantin Boudnik (100 % of his time)