

HMSProposal

HMS Proposal

Abstract

HMS is monitoring, administration and lifecycle management project for Apache Hadoop clusters.

Proposal

HMS will simplify the process of deployment, configuration, management and monitoring of the collection of Hadoop services and applications that compose a Hadoop cluster. The collection of services (Hadoop Stack) will include at least HDFS, MapReduce, HBase, Hive, HCatalog, Pig and Zookeeper. HMS will be easily configurable to add additional services and applications to the stack. Our plan is to support the Hadoop stack as a unit of deployment and configuration where only certain pre-tested versions of software components are supported to be part of Hadoop stack. Administrators can always enable/disable the individual software components from the Hadoop stack per their deployment needs.

The main use cases that HMS is trying to address are the following:

- Hadoop stack deployment and upgrades
- Hadoop services configuration & management
- Administration of Hadoop services
 - Includes starting and stopping services
 - Hadoop system maintenance tasks, such as fsck, format, re-balance, and compaction
- User access & quota management on Hadoop clusters
- Easily check and be alerted to failures in Hadoop servers
- Automated discovery of new machines that become available
- Expanding and contracting Hadoop clusters
- Automatic resynchronization to 'desired' state (of Hadoop stack) to handle faulty nodes
- Handle node burn-ins (stress test nodes using Hadoop before deploying them for production use)
- Simple monitoring and management UI
- Dynamic configuration - Hadoop configuration deduced from machine attributes (e.g., RAM, CPU, Disk)
- Operational HBase-based (inspired by OpenTSDB) monitoring for Hadoop clusters
- Make it possible for administrators to deploy other Hadoop related services and client applications

HMS is targeted to administrators responsible for managing Hadoop clusters. HMS leverages existing data center management and monitoring infrastructure - Nagios, LDAP, Kerberos, etc. All HMS functionality and data will be accessible via RESTFUL APIs and command line tools to facilitate its integration with existing data center management suites.

For the bare metal provisioning, the cluster admins continue to use their existing infrastructure. Provisioning a machine from scratch is not in the scope of the current roadmap.

Background

Hadoop's ecosystem includes many projects (HDFS, MapReduce, Pig, HBase, etc.). In many cases, users and operators typically want to deploy a combination of some projects as a stack. It takes a significant amount of time to get a properly configured Hadoop cluster up and running. HMS has been designed to solve that problem. HMS automates the whole process of deploying a stack.

HMS is being developed by developers employed with Yahoo!, Hortonworks and IBM. Such a tool would have a large number of users and increase the adoption of Apache Hadoop's ecosystem. We are therefore proposing to make HMS Apache open source.

Rationale

Hadoop clusters are complicated and difficult to deploy and manage. The HMS project aims to improve the usability of Apache Hadoop. Doing so will democratize Apache Hadoop, growing its community and increasing the places Hadoop can be used and the problems it can solve. By developing HMS in Apache we hope to gather a diverse community of contributors, helping to make sure that HMS is deployable in as many different situations as possible. members of the Hadoop development community will be able to influence HMS's roadmap, and contribute to it. We believe having HMS as part of the Apache Hadoop ecosystem will be a great benefit to all of Hadoop's users.

Current Status

Prototype available, developed by the list of initial committers.

Meritocracy

Our intent with this incubator proposal is to start building a diverse developer community around HMS following the Apache meritocracy model. We have wanted to make the project open source and encourage contributors from multiple organizations from the start. We plan to provide plenty of support to new developers and to quickly recruit those who make solid contributions to committer status.

Community

We are happy to report that multiple organizations are already represented by initial team. We hope to extend the user and developer base further in the future and build a solid open source community around HMS.

Core Developers

HMS is currently being developed by four engineers from Hortonworks - Eric Yang, Owen O'Malley, Vitthal (a.k.a Suhas) Gogate and Devaraj Das. In addition, a Yahoo! employee, Jagane Sundar, and an IBM employee, Kan Zhang, are also involved. Eric, Jagane and Kan are the original developers. All the engineers have deep expertise in Hadoop and are quite familiar with the Hadoop Ecosystem.

Alignment

The ASF is a natural host for HMS given that it is already the home of Hadoop, Pig, HBase, Cassandra, and other emerging cloud software projects. HMS has been designed to solve the deployment, management and configuration problems of the Hadoop ecosystem family of products. HMS fills the gap that Hadoop ecosystem has been lacking in the areas of configuration, deployment and manageability.

Known Risks

Orphaned products & Reliance on Salaried Developers

The core developers plan to work full time on the project. There is very little risk of HMS getting orphaned. HMS is in use by companies we work for so the companies have an interest in its continued vitality.

Inexperience with Open Source

All of the core developers are active users and followers of open source. Eric Yang is a committer on Apache Chukwa. Owen O'Malley is the lead of the Apache Hadoop project. Devaraj Das is an Apache Hadoop committer and Apache Hadoop PMC member. Vitthal (Suhas) Gogate has contributed extensively to the Hadoop Vaidya project (part of Apache Hadoop). Jagane Sundar has been contributing, in terms of ideas, to the Hadoop project. Kan Zhang is a Hadoop Committer.

Homogeneous Developers

The current core developers are from Hortonworks, IBM, and, Yahoo!. However, we hope to establish a developer community that includes contributors from several corporations.

Reliance on Salaried Developers

Currently, the developers are paid to do work on HMS. However, once the project has a community built around it, we expect to get committers and developers from outside the current core developers.

Relationships with Other Apache Products

HMS is going to be used by the users of Hadoop and the Hadoop ecosystem in general.

A Excessive Fascination with the Apache Brand

While we respect the reputation of the Apache brand and have no doubts that it will attract contributors and users, our interest is primarily to give HMS a solid home as an open source project following an established development model. We have also given reasons in the Rationale and Alignment sections.

Documentation

There is documentation in Hortonworks's internal repositories.

Initial Source

The source is currently in Hortonworks's internal repositories.

Source and Intellectual Property Submission Plan

The complete HMS code is under Apache Software License 2.

External Dependencies

The dependencies all have Apache compatible licenses. These include BSD, MIT licensed dependencies.

Cryptography

None

Required Resources

Mailing lists

- `hms-dev AT incubator DOT apache DOT org`
- `hms-commits AT incubator DOT apache DOT org`
- `hms-user AT hms incubator apache DOT org`
- `hms-private AT incubator DOT apache DOT org`

Subversion Directory

<https://svn.apache.org/repos/asf/incubator/hms>

Issue Tracking

JIRA HMS

Initial Committers

- Devaraj Das (`ddas AT apache DOT org`)
- Bernd Fondermann (`berndf AT apache DOT org`)
- Vitthal Suhas Gogate (`gogate AT apache DOT org`)
- Owen O'Malley (`omalley AT apache DOT org`)
- Jagane Sunder (`jagane AT sundar DOT org`)
- Eric Yang (`eyang AT apache DOT org`)
- Kan Zhang (`kzhang AT apache DOT org`)

Affiliations

- Devaraj Das (Hortonworks)
- Bernd Fondermann (brainlounge)
- Vitthal Suhas Gogate (Hortonworks)
- Owen O'Malley (Hortonworks)
- Jagane Sunder (Yahoo)
- Eric Yang (Hortonworks)
- Kan Zhang (IBM)
- Chris Douglas (Yahoo)
- Arun C Murthy (Hortonworks)

Sponsors

Champion

- Owen O'Malley

Nominated Mentors

- Owen O'Malley
- Arun C Murthy
- Chris Douglas

Sponsoring Entity

Incubator PMC