Welcome to Hama project

Table of contents

1. Downloadable User Guide

- Apache Hama (v0.2) : User Guide [pdf]

2. Getting Started

Hama is a distributed computing framework based on BSP (Bulk Synchronous Parallel) computing techniques for massive scientific computations. Requirements

- hadoop-0.20.x for HDFS
- Sun Java JDK 1.6.x or higher version

Startup script The $HAMA_HOME/bin directory contains some script used to start up the Hama daemons.

- start-bspd.sh - Starts all Hama daemons, the BSPMaster, GroomServers and Zookeeper.

Configuration files The $HAMA_HOME/conf directory contains some configuration files for Hama. These are:

- hama-env.sh - This file contains some environment variable settings used by Hama. You can use these to affect some aspects of Hama daemon behavior, such as where log files are stored, the maximum amount of heap used etc. The only variable you should need to change in this file is JAVA_HOME, which specifies the path to the Java 1.5.x installation used by Hama.
- groomservers - This file lists the hosts, one per line, where the GroomServer daemons will run. By default this contains the single entry localhost.
- hama-default.xml - This file contains generic default settings for Hama daemons. **Do not modify this file**.
- hama-site.xml - This file contains site specific settings for all Hama daemons and BSP jobs. This file is empty by default. Settings in this file override those in hama-default.xml. This file should contain settings that must be respected by all servers and clients in a Hama installation.

Setting up Hama This section describes how to get started by setting up a Hama cluster.

- **BSPMaster and Zookeeper settings** - Figure out where to run your HDFS namenode and BSPMaster. Set the variable bsp.master.address to the BSPMaster's intended host:port. Set the variable fs.default.name to the HDFS Namenode's intended host:port.

An example of a hama-site.xml file: <!--?xml version="1.0"?--> <!--?xml-stylesheet type="text/xsl" href="configuration.xsl"?--> <configuration> <property> <name>bsp.master.address</name> <value>mydomain.com:40000</value> <description>The address of the bsp master server. Either the literal string "local" or a host:port for distributed mode </description> </property> <property>

<name>fs.default.name</name> <value>hdfs://mydomain.com:9000/</value> <description> The name of the default file system. Either the literal string "local" or a host:port for HDFS. </description> </property> <property> <name>hama.zookeeper.quorum</name> <value>mydomain.com</value> <description>Comma separated list of servers in the ZooKeeper Quorum. For example, "host1.mydomain.com,host2.mydomain.com,host3.mydomain.com". By default this is set to localhost for local and pseudo-distributed modes of operation. For a fully-distributed setup, this should be set to a full list of ZooKeeper quorum servers. If HAMA_MANAGES_ZK is set in hama-env.sh this is the list of servers which we will start/stop zookeeper on. </description> </property> </configuration> Starting a Hama cluster Run the command: # $HAMA_HOME/bin/start-bspd.sh This will startup a BSPMaster, GroomServers and Zookeeper on your machine. Stopping a Hama cluster Run the command: # $HAMA_HOME/bin/stop-bspd.sh to stop all the daemons running on your cluster. Run the BSP Examples Run the command: # $HAMA_HOME/bin/hama jar hama-0.2.0-examples.jar pi or test