

GSoCMonitoringAndWebInterface

GSoC 2010: ZooKeeper Monitoring Recipes and Web-based Administrative Interface

- Student: [Andrei Savu](#) (savu.andrei at gmail dot com)
- Assigned mentor: Patrick Hunt (phunt at apache dot org)

Abstract

ZooKeeper is a complex distributed system. Understanding how well it is running is tremendously important. Patrick Hunt has created a [Django-based dashboard](#) that allows some insight into how ZooKeeper is running. This is the foundation I'm going to build on. This project would capture much more information from ZooKeeper, adding hooks to retrieve it where necessary and visualize it in an appealing and useful way. I'm also going to provide a bunch of monitoring recipes for systems like: Ganglia, Nagios, Cacti.

Committed to trunk

- <https://issues.apache.org/jira/browse/ZOOKEEPER-808>
 - Hue Application: <http://github.com/andreisavu/hue> (branch: zookeeper-browser app: apps/zkui)
- <https://issues.apache.org/jira/browse/ZOOKEEPER-809>
 - will open another JIRA for ACLs (get, set) and per session ZK authentication
- <https://issues.apache.org/jira/browse/ZOOKEEPER-732>
 - added some fixes on the existing patch created by Lei Zhang
- <https://issues.apache.org/jira/browse/ZOOKEEPER-765>
- <https://issues.apache.org/jira/browse/ZOOKEEPER-799>
 - Github Repository: <http://github.com/andreisavu/zookeeper-monitoring>
- <https://issues.apache.org/jira/browse/ZOOKEEPER-744>
- <https://issues.apache.org/jira/browse/ZOOKEEPER-754>

Milestones

Community Bonding (starts: 26 April ends: 24 May)

Activities:

- read mail lists archives - **done**
- read source code- **done**
- discuss with the community members (monitoring and administration requirements, production stories) - **done**
- discuss with the Adobe Hadoop / Hbase team about their specific monitoring requirements - **done**

Expected results:

- understand source code and the known bugs - **done**
- understand how the software is used in production - **done**
 - ZooKeeper is the kind of service that you put in production and forget about it
 - got positive feedback: works as expected "out of the box"
 - monitoring requirements: ensure that it keeps working as expected
- understand monitoring requirements - **done**
- understand debugging requirements - **done**
- setup a development environment - **done**
 - on the local machine running Ubuntu 9.10, java1.6, Eclipse, ant
 - tracking my changes on github: <http://github.com/andreisavu/zookeeper>

Monitoring and Data Collection (starts: 24 May ends: 20 June)

Activities:

- deploy small scale (multinode) cluster for development (virtual machines) - **done**
 - I've used <http://github.com/phunt/zkconf> for this task. I've deployed local "clusters" with 3,5 and 9 nodes
- identify important health signals add hooks (if needed) for realtime data collection - **done**
 - added new 4letterword 'mntr' for monitoring - going to be released in zookeeper 3.4.0
 - important signals: latency, packets sent / received, outstanding requests, znode count, watch count, ephemerals count, followers count, synced followers, pending syncs, open file descriptor count
- create scripts / plugins for cluster monitoring using Cacti, Ganglia, Nagios - **done**
 - <http://github.com/andreisavu/zookeeper-monitoring>
- document script install procedures - **done** (I'm making the assumption the user has previous experience configuring Nagios, Cacti or Ganglia)
- collaborate with the Adobe Hadoop / Hbase team and deploy the monitoring scripts in production - **work in progress**

Expected results:

- production ready scripts / plugins for monitoring - **done**
- easy to understand and follow install guides - **done**

Web Application (starts: 20 June ends: 9 august)

Activities:

- package zkpython bindings (distutils, .deb, .rpm) **done**
 - - already available: apt-get install python-zookeeper
 - <https://wiki.cloudera.com/display/DOC/ZooKeeper+Installation>
- simple authentication and custom authentication backend based on zookeeper
 - - not needed: the web-based application will use the authentication provided by Hue
- view server, environment and connection info: most of the code already works **done**
 - - I've rewrite all the code in the Hue application
 - The code uses 4letter word commands: 'stat' and 'mntr'
- znode hierarchy browser **done**
 - - you can navigate and perform simple CRUD operations on znodes
- deploy on production or development cluster at Adobe (if possible) **work in progress**
 - - this should be pretty easy if Adobe is also using Hue

Expected results:

- packages for zkpython **done**
- working web application **done**

Cleanup and final fixes (starts: 9 august ends: 16 august)

Activities:

- improve tests and documentation **done**

Submit code to code.google.com : 30 August

Related JIRA

- <https://issues.apache.org/jira/browse/ZOOKEEPER-701>