# Chukwa

## http://incubator.apache.org/chukwa/images/chukwa\_logo\_small.jpg

Chukwa is a Hadoop subproject devoted to large-scale log collection and analysis. Chukwa is built on top of the Hadoop distributed filesystem (HDFS) and MapReduce framework and inherits Hadoop's scalability and robustness. Chukwa also includes a exible and powerful toolkit for displaying monitoring and analyzing results, in order to make the best use of this collected data.

#### Documentation

- Chukwa\_How\_To\_Contribute How to be part of Chukwa community
- Chukwa\_How\_To\_Release Release process for Chukwa
- FAQ In progress...
- Sending\_information\_to\_Chukwa A tutorial walking through the process of sending a log file to chukwa and how Chukwa parses records from the datasink file.
- Chukwa\_Processes\_and\_Data\_Flow A description of the various processes that operate on Chukwa data and how that data moves through HDFS.
- Anomaly\_Detection\_Framework\_with\_Chukwa A description of Anomaly Detection Framework design for Chukwa 0.2.

# Presentations

- ChukwaPoster.pdf Chukwa Poster
- chukwa\_presentation.pdf An overview of the Chukwa Monitoring System
- chukwa\_presentation\_cca08.pdf A talk presented about Chukwa by Berkeley graduate students at Cloud Computing and its Applications 08 (http://cca08.org) October 2008.

## Download

Chukwa is part of the Hadoop distribution. You can view the source as part of the Hadoop Apache SVN repository here

# Papers

• chukwa\_cca08.pdf - Cloud Computing and its Applications (CCA) 2008

#### Links

- JIRA HADOOP-3719 The original Apache JIRA ticket for contributing Chukwa to Hadoop as a contrib project.
- JIRA HADOOP-4709 A batch update to the JIRA in Hadoop/src/contrib. After this update the Chukwa team will be fully embracing the Apache JIRA development model, as suggested in the comments on this JIRA.