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.