Welcome to Hadoop!

Table of contents

1. What Is Hadoop?

The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing, including:

- **Hadoop Core**, our flagship sub-project, provides a distributed filesystem (HDFS) and support for the MapReduce distributed computing metaphor.
- **HBase** builds on Hadoop Core to provide a scalable, distributed database.
- **Pig** is a high-level data-flow language and execution framework for parallel computation. It is built on top of Hadoop Core.
- **ZooKeeper** is a highly available and reliable coordination system. Distributed applications use ZooKeeper to store and mediate updates for critical shared state.
- **Hive** is a data warehouse infrastructure built on Hadoop Core that provides data summarization, adhoc querying and analysis of datasets.

2. Who uses Hadoop?

A wide variety of companies and organizations use Hadoop for both research and production. Users are encouraged to add themselves to the Hadoop users wiki page.

3. News

3.1. 31 October 2008 - Apache Con US is next week

It is hard to believe, but ApacheCon US 2008 is in New Orleans next week. Hadoop will have very good representation including:

- Hadoop Camp with lots of talks about Hadoop
    - Introduction to Hadoop by Owen O'Malley
    - A Tour of Apache Hadoop by Tom White
    - Programming Hadoop Map/Reduce by Arun Murthy
    - Hadoop at Yahoo! by Eric Baldeschwieler
    - Hadoop Futures Panel
    - Using Hadoop for an Intranet Seach Engine by Shivakumar Vaithyanthan
    - Cloud Computing Testbed by Thomas Sandholm
    - Improving Virtualization and Performance Tracing of Hadoop with Open Solaris by George Porter
    - An Insight into Hadoop Usage at Facebook by Dhruba Borthakur
    - Hadoop on Amazon Web Services by Jinesh Varia
    - Hive by Ashish Thusoo
    - Pig by Alan Gates

- • Zookeeper, Coordinating the Distributed Application by Ben Reed
- • Querying JSON Data on Hadoop using Jaql by Kevin Beyer
- • HBase by Michael Stack
- Hadoop training on Practical Problem Solving in Hadoop
- Cloudera is providing a test Hadoop cluster and a Hadoop hacking contest.

There is also a new Hadoop tutorial available.

3.2. 2 July 2008 - Hadoop Wins Terabyte Sort Benchmark

Hadoop Wins Terabyte Sort Benchmark: One of Yahoo's Hadoop clusters sorted 1 terabyte of data in 209 seconds, which beat the previous record of 297 seconds in the annual general purpose (daytona) terabyte sort benchmark. This is the first time that either a Java or an open source program has won.

3.3. 23 January 2008 - Hadoop at ApacheCon Europe

Hadoop will be well represented at ApacheCon Europe in Amsterdam this year. Please join us at one or more of the following sessions:
- April 10: Hadoop BOF
- April 11: A tour of Apache Hadoop by Tom White
- April 11: Programming with Hadoop's MapReduce by Owen O'Malley
- April 11: Deploying Grid Services using Apache Hadoop by Allen Wittenauer

3.4. 23 January 2008 - Hadoop promoted to Top Level Project

The Lucene PMC and the ASF Board have voted to promote Hadoop to be a top level project. Hadoop can now be found at http://hadoop.apache.org/