

Welcome to Apache Hadoop!

Table of contents

1 What Is Hadoop?.....	2
2 Who Uses Hadoop?.....	2
3 News.....	2
3.1 July 2009 - New Hadoop Subprojects.....	2
3.2 March 2009 - ApacheCon EU.....	2
3.3 November 2008 - ApacheCon US.....	3
3.4 July 2008 - Hadoop Wins Terabyte Sort Benchmark.....	3

1. What Is Hadoop?

The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing. Hadoop includes these subprojects:

- [Hadoop Common](#): The common utilities that support the other Hadoop subprojects.
- [Avro](#): A data serialization system that provides dynamic integration with scripting languages.
- [Chukwa](#): A data collection system for managing large distributed systems.
- [HBase](#): A scalable, distributed database that supports structured data storage for large tables.
- [HDFS](#): A distributed file system that provides high throughput access to application data.
- [Hive](#): A data warehouse infrastructure that provides data summarization and ad hoc querying.
- [MapReduce](#): A software framework for distributed processing of large data sets on compute clusters.
- [Pig](#): A high-level data-flow language and execution framework for parallel computation.
- [ZooKeeper](#): A high-performance coordination service for distributed applications.

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 [PoweredBy](#) wiki page.

3. News

3.1. July 2009 - New Hadoop Subprojects

Hadoop is getting bigger!

- Hadoop Core is renamed Hadoop Common.
- MapReduce and the Hadoop Distributed File System (HDFS) are now separate subprojects.
- Avro and Chukwa are new Hadoop subprojects.

See the summary descriptions for all subprojects above. Visit the individual sites for more detailed information.

3.2. March 2009 - ApacheCon EU

Welcome to Apache Hadoop!

In case you missed it... [ApacheCon Europe 2009](#)

3.3. November 2008 - ApacheCon US

In case you missed it... [ApacheCon US 2008](#)

3.4. 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.