

Overview of Hadoop Ecosystem

Hadoop Ecosystem Components

- 1 **HDFS**: Hadoop Distributed File System. Stores data files as close to original form as possible.
- 2 **HBase**: Hadoop's database; supports structured data storage for large tables (like RDBMS).
- 3 **Hive**: Enables analysis of large datasets using SQL-like language.
- 4 **Pig**: Simple data flow language for large dataset analysis. Converts scripts to MapReduce jobs.
- 5 **ZooKeeper**: Coordination service for distributed applications.

Hadoop Ecosystem Components (Contd.)

- ⑥ **Oozie**: Workflow scheduler to manage Hadoop jobs.
- ⑦ **Mahout**: Scalable machine learning and data mining library.
- ⑧ **Chukwa**: Data collection system for managing large distributed systems.
- ⑨ **Sqoop**: Transfers bulk data between Hadoop and relational databases.
- ⑩ **Ambari**: Web-based tool for provisioning, managing, and monitoring Hadoop clusters.