

Table 1

Tools Names and Usages

Tools	Usage	URL
Hadoop Distributed File System (HDFS)	Distributed file processing system for large data sets	https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
Apache Pig	MapReduce data scripting language for extracting, transforming, and loading data into data stores	https://pig.apache.org/
Apache Hive	SQL query language; used with HDFS and HBase data warehouses	https://hive.apache.org/
Apache HBase	Nonrelational (NoSQL) database storage used on HDFS; modeled after Google's Big Table	https://hbase.apache.org/
MapReduce	Hadoop programming model distributes task across servers	https://hadoop.apache.org/docs/stable/index.html
IBM Apache Hadoop Zookeeper	Synchronizes data delivery to servers	https://www.ibm.com/analytcs/hadoop/zookeeper
Apache Cassandra	Nonrelational (NoSQL) database storage	http://cassandra.apache.org/