

**Hadoop** A platform for storage and processing of large data sets. Hadoop consists of a distributed filesystem, Java interfaces, and services for managing distribution of tasks. Core code is maintained by the Apache Software Foundation.

**MapR distribution for Hadoop** MapR provides an enterprise-ready distribution for Hadoop including the core open source components as well as a high-performance distributed filesystem, high-availability services, and easy-to-use management interface.

**MapReduce** A programming model for parallel processing, introduced in Google's 2004 paper<sup>1</sup>. Hadoop provides a framework for running MapReduce jobs written in Java.

**Pig** A high level data flow language for processing data. Pig programs describe steps to be executed on sets of tuples, and are executed by running one or more MapReduce jobs.

**Hive** A framework for executing SQL-like queries on large data sets. The queries are executed as MapReduce jobs.

---

<sup>1</sup><http://research.google.com/archive/mapreduce.html>