

# Architecture Guide

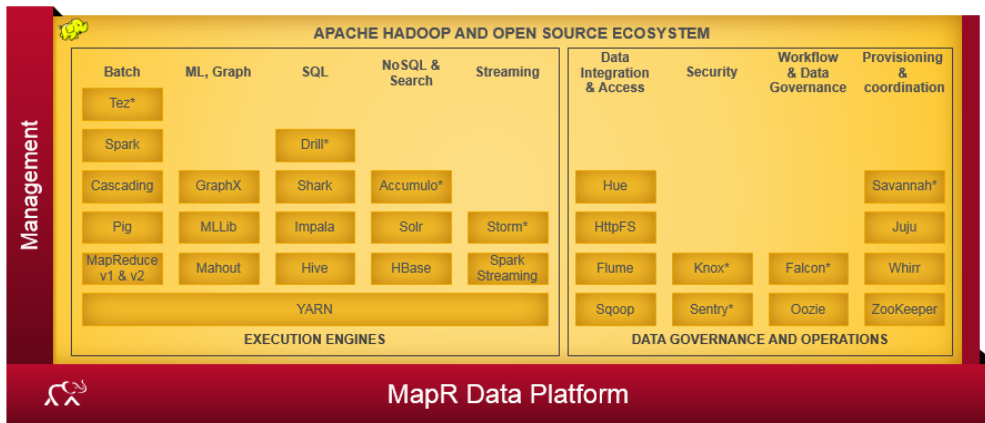
MapR offers a complete, industry-standard Hadoop distribution with key improvements. MapR Distribution for Hadoop includes the full family of Hadoop ecosystem components, which have been tested together on specific platforms. MapR supports the Hadoop FS abstraction interface and improves the performance and robustness of the distributed file system, eliminating the Namenode. The MapR Distribution for Hadoop supports continuous read/write access, improving data load, and unload processes.

MapR Distribution for Hadoop Version 4.0.1 introduced the Hadoop 2.x architecture and YARN (Yet Another Resource Negotiator). Hadoop 2.x and YARN make up a resource management and scheduling framework that distributes resource management and job management duties.

Hadoop 2.x was designed to solve two main problems present in the Hadoop 1.x architecture:

- Centralization of job scheduling, resulting in scheduler bottlenecks
- Separating resource management from application programming concerns

The following image represents the MapR Distribution for Hadoop:



This guide contains architectural details about the components that run on the MapR Data Platform, how the components assemble into a cluster, and the relationships between the components:

- MapR Data Platform
- Cluster Management
- MapR-DB Tables
- Security on MapR
- MapReduce Version 1
- YARN