

What is Hadoop?

Apache Hadoop is a new way for enterprises to store and analyze data.

[Hadoop](#) is an open-source project administered by the Apache Software Foundation. Hadoop's [contributors](#) work for some of the world's biggest technology companies. That diverse, motivated community has produced a genuinely innovative platform for consolidating, combining and understanding large-scale data in order to better comprehend the data deluge.

Enterprises today collect and generate more data than ever before. Relational and data warehouse products excel at OLAP and OLTP workloads over structured data. Hadoop, however, was designed to solve a different problem: the fast, reliable analysis of both structured data *and* complex data. As a result, many enterprises deploy Hadoop alongside their legacy IT systems, which allows them to combine old data and new data sets in powerful new ways.

Technically, Hadoop consists of two key services: reliable data storage using the Hadoop Distributed File System (HDFS) and high-performance parallel data processing using a technique called MapReduce.

Hadoop runs on a collection of commodity, shared-nothing servers. You can add or remove servers in a Hadoop cluster at will; the system detects and compensates for hardware or system problems on any server. Hadoop, in other words, is self-healing. It can deliver data — and can run large-scale, high-performance processing jobs — in spite of system changes or failures.

Originally developed and employed by dominant Web companies like Yahoo and Facebook, Hadoop is now widely used in finance, technology, telecom, media and entertainment, government, research institutions and other markets with significant data. With Hadoop, enterprises can easily explore complex data using custom analyses tailored to their information and questions.

[Cloudera](#) is an active contributor to the Hadoop project and provides an enterprise-ready, commercial Distribution for Hadoop. [Cloudera's Distribution](#) bundles the innovative work of a global open-source community; this includes critical bug fixes and important new features from the public development repository and applies all this to a stable version of the source code. In short, [Cloudera](#) integrates the most popular projects related to Hadoop into a single package, which is run through a suite of rigorous tests to ensure reliability during production.

You can learn more about Hadoop by reading our [Hadoop Overview](#). Also, learn more about [what Hadoop can do for you](#).

Cloudera Apache Hadoop Certified Integrated Hardware Vendor (IHV)