Apache Hadoop® is an open source software framework that provides highly reliable distributed processing of large data sets using simple programming models. Hadoop, known for its scalability, is built on clusters of commodity computers, providing a cost-effective solution for storing and processing massive amounts of structured, semi-structured and unstructured data with no format requirements.

A [data lake architecture](https://www.ibm.com/analytics/data-lake) including Hadoop can offer a flexible data management solution for your [big data analytics](https://www.ibm.com/analytics/hadoop/big-data-analytics) initiatives. Because Hadoop is an open source software project and follows a distributed computing model, it can offer a lower total cost of ownership for a big data software and storage solution.

Hadoop can also be installed on cloud servers to better manage the compute and storage resources required for big data. Leading cloud vendors such as Amazon Web Services (AWS) and Microsoft Azure offer solutions. Cloudera supports Hadoop workloads both on-premises and in the cloud, including options for one or more public cloud environments from multiple vendors.