

Welcome to Hadoop!

Table of contents

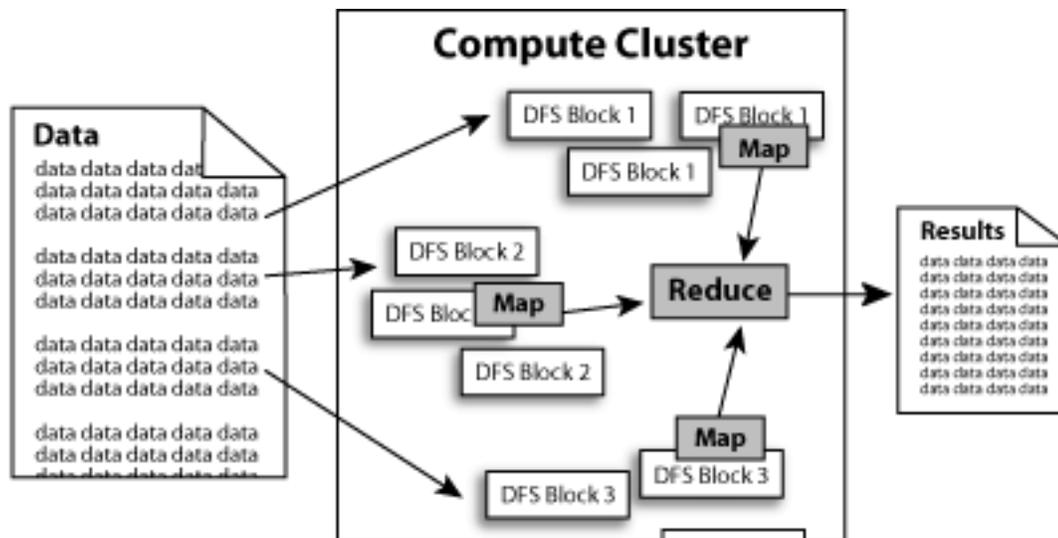
1 Getting Started	2
2 Getting Involved	3

Hadoop is a software platform lets one easily write and run applications that process vast amounts of data.

Here's what makes Hadoop especially useful:

- **Scalable:** Hadoop can reliably store and process petabytes.
- **Economical:** It distributes the data and processing across clusters of commonly available computers. These clusters can number into the thousands of nodes.
- **Efficient:** By distributing the data, Hadoop can process it in parallel on the nodes where the data is located. This makes it extremely rapid.
- **Reliable:** Hadoop automatically maintains multiple copies of data and automatically redeploys computing tasks based on failures.

Hadoop implements [MapReduce](#), using the Hadoop Distributed File System ([HDFS](#)) (see figure below.) MapReduce divides applications into many small blocks of work. HDFS creates multiple replicas of data blocks for reliability, placing them on compute nodes around the cluster. MapReduce can then process the data where it is located.



1. Getting Started

The Hadoop project plans to scale Hadoop up to handling thousands of computers. However, to begin with you can start by installing in on a single machine or a very small cluster.

1. [Learn about](#) Hadoop by reading the documentation.
2. [Download](#) Hadoop from the release page.
3. [Install and configure](#) Hadoop. Scroll down the page.
4. [Discuss it](#) on the mailing list.

Welcome to Hadoop!

2. Getting Involved

Hadoop is an open source volunteer project under the Apache Software Foundation. We encourage you to learn about the project and contribute your expertise. Here are some starter links:

1. See our [How to Contribute to Hadoop](#) page.
2. Give us [feedback](#): What can we do better?
3. Join the [mailing list](#): Meet the community.