What is Apache Hadoop?

Apache Hadoop is a Big Data framework that is part of the Apache Software Foundation. Hadoop is an open source softwareproject that is extensively used by some of the biggest organizations in the world for distributed storage and processing of data on a level that is just enormous in terms of volume. That's the reason the Apache Hadoop runs its processing on large computer clusters built on commodity hardware. Some of the features of the Hadoop platform are that it can be efficiently used for data storage, processing, access, analysis, governance, security, operations and deployment.

Hadoop is a top level project that is being built and used by a diverse group of developers, users and contributors cutting across nationalities under the auspices of the Apache Foundation. Hadoop is currently governed under the Apache License 2.0.

Hadoop operates on thousands of nodes that involve huge amounts of data and hence during such a scenario the failure of a node is a high probability. So the Hadoop platform is resilient in the sense that The Hadoop distributed file systems immediately upon sensing of a node failure divert the data among other nodes thus allowing the whole platform to operate without any interruptions.
The idea of Apache Hadoop was actually born out of a Google project called the MapReduce, which is a framework for breaking down an application into smaller chunks that can then be parsed on a much smaller and granular level. Each of the smaller blocks is individually operated on nodes which are then connected to the main cluster. The present Hadoop framework consists of its major components namely MapReduce, the Hadoop Kernel, and the HDFS (Hadoop Distributed File System). Then there are other related projects like the Apache HBase, Sqoop, Hive, Pig, and so on.