**What is Hadoop?**

Hadoop is an **open-source distributed computing framework** that allows storage and processing of large datasets across clusters of computers using simple programming models.

**🔹 Properties of Hadoop**

1. **Scalability** – Easily expand by adding more nodes.
2. **Fault Tolerance** – Recovers data automatically using replication.
3. **Distributed Processing** – Data is processed in parallel across nodes.
4. **Cost Effectiveness** – Runs on low-cost commodity hardware.
5. **Open Source** – Freely available and supported by Apache.

**🔹 Hadoop Ecosystem (2-line explanation)**

The Hadoop ecosystem consists of a collection of tools and modules built around Hadoop to enable **efficient data storage, processing, analysis, and management** across distributed environments.

**🔧 Hadoop Ecosystem Components – Concise Descriptions**

**1. HDFS (Hadoop Distributed File System)**

Stores data across multiple machines with redundancy; ensures high availability and fault tolerance.

**2. MapReduce**

A programming model that processes data in parallel using Map (filtering) and Reduce (aggregation) phases.

**3. YARN (Yet Another Resource Negotiator)**

Cluster management technology that allocates system resources and schedules jobs.

**4. Hive**

Data warehouse tool that enables querying and managing large datasets using SQL-like syntax (HiveQL).

**5. Pig**

Scripting platform using Pig Latin for analyzing large data sets, especially suitable for ETL tasks.

**6. Sqoop**

Used to efficiently transfer bulk data between Hadoop and structured data stores like RDBMS.

**7. Oozie**

A scheduler that manages Hadoop jobs and allows chaining of MapReduce, Pig, Hive, and Sqoop tasks in workflows.

**8. HBase**

A distributed NoSQL database that provides real-time read/write access to large datasets, built on top of HDFS.

**9. Mahout**

Machine learning library designed for scalable, distributed machine learning algorithms (e.g., clustering, classification).

**10. Flume**

Data ingestion tool designed for collecting, aggregating, and transporting log data into Hadoop.

**11. ZooKeeper**

A centralized service for maintaining configuration information, naming, and distributed synchronization across clusters.