**Hardware Requirements**

**Minimum Requirements (For Development/Small Scale Testing)**

* **CPU**: Dual-core processor
* **RAM**: 8 GB
* **Storage**: 100 GB
* **Network**: Gigabit Ethernet

**Recommended Requirements (For Production/Cluster Deployment)**

* **Master Node (NameNode, ResourceManager)**
  + **CPU**: Quad-core processor or better
  + **RAM**: 16-64 GB (depends on cluster size)
  + **Storage**: SSDs recommended for faster access, 1 TB or more
  + **Network**: High-speed network (Gigabit Ethernet or higher)
* **Slave Nodes (DataNodes, NodeManagers)**
  + **CPU**: Quad-core processor or better
  + **RAM**: 16-64 GB (depends on workload)
  + **Storage**: Multiple high-capacity disks (HDDs or SSDs), 1 TB or more
  + **Network**: High-speed network (Gigabit Ethernet or higher)

**Software Requirements**

**Operating System**

* **Linux** (Preferred distributions: CentOS, Ubuntu, Debian, Red Hat)
* **Windows** (Supported but less common in production environments)

**Java**

* **Java Development Kit (JDK)**: Version 8 or higher (Hadoop is primarily developed in Java)
  + Make sure JAVA\_HOME environment variable is set correctly.

**SSH**

* **Secure Shell (SSH)**: Password-less SSH access must be configured between nodes in the cluster.

**Hadoop Distribution**

* **Hadoop Version**: Ensure compatibility with other software components.
  + Download the appropriate Hadoop version from [Apache Hadoop](https://hadoop.apache.org/).

**Configuration Recommendations**

* **Memory Allocation**: Allocate enough memory to HDFS and YARN based on the available RAM and expected workload.
* **Storage Configuration**: Use RAID configurations for data redundancy and performance improvements. SSDs are preferred for critical components like the NameNode.
* **Network Setup**: Ensure a high-speed, reliable network setup. In a large cluster, consider using multiple network interfaces for better performance.

**Example Deployment Sizes**

**Small Deployment**

* **Nodes**: 1-10
* **RAM per Node**: 16 GB
* **Storage per Node**: 1 TB
* **Use Case**: Development, testing, small-scale data processing

**Medium Deployment**

* **Nodes**: 10-50
* **RAM per Node**: 32 GB
* **Storage per Node**: 2-4 TB
* **Use Case**: Medium-scale data processing, business analytics

**Large Deployment**

* **Nodes**: 50-100+
* **RAM per Node**: 64 GB or more
* **Storage per Node**: 4-10 TB
* **Use Case**: Large-scale data processing, enterprise-level data analytics, real-time data processing