



# Certified Big Data and Apache Hadoop Developer VS-1221

**Vskills Certifications**

Vskills Brochure



## Certified Big Data and Apache Hadoop Developer

### Certification Code VS-1221

Vskills certification for Big Data and Apache Hadoop Developer Certification assesses the knowledge and skills required to become a successful Hadoop Developer, Administrator, Data Scientist Professional etc in the field of Big Data. The certification tests the candidates on various areas in Big Data and Apache Hadoop.

### Why should one take this certification?

This Course is intended for professionals and graduates wanting to excel in their chosen areas. It is also well suited for those who are already working and would like to take certification for further career progression.

Earning Vskills Big Data and Apache Hadoop Developer Certification can help candidate differentiate in today's competitive job market, broaden their employment opportunities by displaying their advanced skills, and result in higher earning potential.

### Who will benefit from taking this certification?

The course is designed for professionals aspiring to make a career in Big Data and Hadoop Framework. Students, Software Professionals, Analytics Professionals, ETL developers, Project Managers, Architects, and Testing Professionals are the key beneficiaries of this course. Other professionals who are looking forward to acquire a solid foundation on Big Data Industry can also opt for this course. This not only improves their skill set but also makes their CV stronger and existing employees looking for a better role can prove their employers the value of their skills through this certification.

### Test Details

- **Duration:** 60 minutes
- **No. of questions:** 50
- **Maximum marks:** 50, Passing marks: 35 (70%)

There is no negative marking in this module.

### Fee Structure

Rs. 3,999/- (Excludes taxes)\*

\*Fees may change without prior notice, please refer <http://www.vskills.in> for updated fees

## Vskills Certified Big Data and Apache Hadoop Developer

### Companies that hire Vskills Big Data and Apache Hadoop Developer

With 1.8 trillion gigabytes of structured and unstructured data in the world, and the volume doubling every two years, the need for big data analysis and business intelligence has never been greater. It adds up to an incredible need for Hadoop professionals who understand how to develop, process and manage half of world's data on Hadoop. Build game-changing Big Data Applications on Hadoop and future-proof your career.

## Table of Contents

### 1. Big Data

- 1.1. Big Data Definition
- 1.2. Big Data Types
- 1.3. Big Data Source
- 1.4. Big Data Challenges
- 1.5. Big Data Benefits
- 1.6. Big Data Applications
- 1.7. Netflix Application

### 2. Apache Hadoop

- 2.1. Introduction
- 2.2. Advantages & Disadvantages
- 2.3. History of Hadoop Project
- 2.4. Need for Hadoop
- 2.5. Hadoop Architecture
- 2.6. RDBMS vs Hadoop
- 2.7. Vendor Comparison
- 2.8. Hardware Recommendations
- 2.9. Hadoop Installation

### 3. HDFS

- 3.1. Basics (Blocks, Namenodes and Datanodes)
- 3.2. HDFS Architecture
- 3.3. Data Read and Write Process
- 3.4. HDFS Permissions
- 3.5. Data Replication
- 3.6. HDFS Accessibility
- 3.7. HDFS Filesystem Operations
- 3.8. HDFS Interfaces
- 3.9. Heartbeats
- 3.10. Rack Awareness
- 3.11. distcp

# Vskills Certified Big Data and Apache Hadoop Developer

---

## 4. MapReduce

- 4.1. MapReduce Basics
- 4.2. MapReduce Work Flow
- 4.3. MapReduce Framework
- 4.4. Hadoop Data Types
- 4.5. MapReduce Internals
- 4.6. Job Formats
- 4.7. Debugging and Profiling
- 4.8. Distributed Cache
- 4.9. Combiner Functions
- 4.10. Streaming
- 4.11. Counters, Sorting and Joins

## 5. YARN

- 5.1. YARN Infrastructure
- 5.2. ResourceManager
- 5.3. ApplicationMaster
- 5.4. NodeManager
- 5.5. Container

## 6. Pig

- 6.1. Pig Architecture
- 6.2. Installation and Modes
- 6.3. Grunt and Pig Script
- 6.4. Pig Latin Commands
- 6.5. UDF and Data Processing Operator

## 7. HBase

- 7.1. HBase Architecture
- 7.2. HBase Installation
- 7.3. HBase Configuration
- 7.4. HBase Schema Design
- 7.5. HBase Commands
- 7.6. MapReduce Integration
- 7.7. HBase Security

## 8. Sqoop and Flume

- 8.1. Sqoop
- 8.2. Flume

## 9. Hive

- 9.1. Hive Architecture
- 9.2. Hive shell
- 9.3. Hive Data types
- 9.4. HiveQL

## 10. Workflow

10.1. Apache Oozie

## 11. Hadoop Cluster Management

11.1. Cluster Planning

11.2. Installation and Configuration

11.3. Testing

11.4. Benchmarking

11.5. Monitoring

## 12. Administration

12.1. dfsadmin, fsck and balancer

12.2. Logging

12.3. Data Backup

12.4. Add and removal of nodes

## 13. Security

13.1. Authentication

13.2. Data Confidentiality

13.3. Configuration

## 14. NextGen Hadoop

14.1. HDFS HA

14.2. HDFS Federation

## Sample Questions

**1. For a MapReduce job, on a cluster running MapReduce v1 (MRv1), what's the relationship between tasks and task templates?**

- A. Once the write stream closes on the DataNode, the DataNode immediately initiates a black report to the NameNode.
- B. The change is written to the NameNode disk.
- C. The metadata in the RAM on the NameNode is flushed to disk.
- D. The metadata in RAM on the NameNode is flushed disk.
- E. The metadata in RAM on the NameNode is updated.
- F. The change is written to the edits file.

**2. How does HDFS Federation help HDFS Scale horizontally?**

- A. HDFS Federation improves the resiliency of HDFS in the face of network issues by removing the NameNode as a single-point-of-failure.
- B. HDFS Federation allows the Standby NameNode to automatically resume the services of an active NameNode.
- C. HDFS Federation provides cross-data center (non-local) support for HDFS, allowing a cluster administrator to split the Block Storage outside the local cluster.
- D. HDFS Federation reduces the load on any single NameNode by using the multiple, independent NameNode to manage individual parts of the filesystem namespace

**3. What is the recommended disk configuration for slave nodes in your Hadoop cluster with 6 x 2 TB hard drives?**

- A. RAID 10
- B. JBOD
- C. RAID 5
- D. RAID 1+0

**4. Your developers request that you enable them to use Hive on your Hadoop cluster. What do install and/or configure?**

- A. Install the Hive interpreter on the client machines only, and configure a shared remote Hive Metastore.
- B. Install the Hive Interpreter on the client machines and all the slave nodes, and configure a shared remote Hive Metastore.
- C. Install the Hive interpreter on the master node running the JobTracker, and configure a shared remote Hive Metastore.
- D. Install the Hive interpreter on the client machines and all nodes on the cluster

## Vskills Certified Big Data and Apache Hadoop Developer

---

5. Which command does Hadoop offer to discover missing or corrupt HDFS data?

- A. The map-only checksum utility,
- B. Fsck
- C. Du
- D. Dskchk
- E. Hadoop does not provide any tools to discover missing or corrupt data; there is no need because three replicas are kept for each data block.

**Answers:** 1 (A), 2 (D), 3 (B), 4 (A), 5 (B)



## Certifications

### ➤ Accounting, Banking and Finance

- Certified AML-KYC Compliance Officer
- Certified Business Accountant
- Certified Commercial Banker
- Certified Foreign Exchange Professional
- Certified GAAP Accounting Standards Professional
- Certified Financial Risk Management Professional
- Certified Merger and Acquisition Analyst
- Certified Tally 9.0 Professional
- Certified Treasury Market Professional
- Certified Wealth Manager

### ➤ Big Data

- Certified Hadoop and Mapreduce Professional

### ➤ Cloud Computing

- Certified Cloud Computing Professional

### ➤ Design

- Certified Interior Designer

### ➤ Digital Media

- Certified Social Media Marketing Professional
- Certified Inbound Marketing Professional
- Certified Digital Marketing Master

### ➤ Foreign Trade

- Certified Export Import (Foreign Trade) Professional

### ➤ Health, Nutrition and Well Being

- Certified Fitness Instructor

### ➤ Hospitality

- Certified Restaurant Team Member (Hospitality)

### ➤ Human Resources

- Certified HR Compensation Manager
- Certified HR Staffing Manager
- Certified Human Resources Manager
- Certified Performance Appraisal Manager

### ➤ Office Skills

- Certified Data Entry Operator
- Certified Office Administrator

### ➤ Project Management

- Certified Project Management Professional

### ➤ Real Estate

- Certified Real Estate Consultant

### ➤ Marketing

- Certified Marketing Manager

### ➤ Quality

- Certified Six Sigma Green Belt Professional
- Certified Six Sigma Black Belt Professional
- Certified TQM Professional

### ➤ Logistics & Supply Chain Management

- Certified International Logistics Professional
- Certified Logistics & SCM Professional
- Certified Purchase Manager
- Certified Supply Chain Management Professional

### ➤ Legal

- Certified IPR & Legal Manager
- Certified Labour Law Analyst
- Certified Business Law Analyst
- Certified Corporate Law Analyst

### ➤ Information Technology

- Certified ASP.NET Programmer
- Certified Basic Network Support Professional
- Certified Business Intelligence Professional
- Certified Core Java Developer
- Certified E-commerce Professional
- Certified IT Support Professional
- Certified PHP Professional
- Certified Selenium Professional
- Certified SEO Professional
- Certified Software Quality Assurance Professional

### ➤ Mobile Application Development

- Certified Android Apps Developer
- Certified iPhone Apps Developer

### ➤ Security

- Certified Ethical Hacking and Security Professional
- Certified Network Security Professional

### ➤ Management

- Certified Corporate Governance Professional
- Certified Corporate Social Responsibility Professional

### ➤ Life Skills

- Certified Business Communication Specialist
- Certified Public Relations Officer

### ➤ Media

- Certified Advertising Manager
- Certified Advertising Sales Professional

### ➤ Sales, BPO

- Certified Sales Manager
- Certified Telesales Executive

**& many more job related certifications**

Contact us at :

**Vskills**

011-473 44 723 or [info@vskills.in](mailto:info@vskills.in)

[www.vskills.com](http://www.vskills.com)