



Certified Big Data and Apache Hadoop Developer VS-1221

V-Skills Certifications

**A Government of India
&
Government of NCT Delhi Initiative**

V-Skills



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. 4,999/- (Includes all taxes)

Please note that completing the Video based course by Whizlabs is mandatory to appear in this certification exam.

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

Module 1: Introduction to Big Data and Hadoop

1. Today's Market
2. Current Situation
3. Introduction to Big Data
4. Sources of Big Data
5. Technical & Business Drivers
6. Big Data Use Cases - Banking, Healthcare, Agriculture
7. Traditional DBMS & their Limitations
8. Introduction to Hadoop
9. Hadoop Usage
10. Real-Time Use Cases - Retail, Farming

Module 2: Getting started with Hadoop

1. Hadoop History
2. Hadoop v/s RDBMS
3. Hadoop Architecture
4. Hadoop Ecosystem components
5. Hadoop Storage - HDFS
6. Hadoop Processor - MapReduce
6. Hadoop Server Roles: NameNode, Secondary NameNode, DataNode
7. Anatomy of File Write and Read

Module 3: Hadoop Distributed File System

1. HDFS Architecture
2. HDFS internals and use cases
3. HDFS Daemons
4. Files and blocks
5. NameNode memory concerns
6. Secondary NameNode
7. HDFS access options

Module 4: MapReduce

1. Use cases of MapReduce
2. MapReduce Architecture
3. Understand the concept of Mappers, Reducers
4. Anatomy of MapReduce Program
5. MapReduce Components - Mapper Class, Reducer Class, Driver code
6. Splits and Blocks
7. Understand Combiner and Partitioner
8. Write your own Partitioner
9. Joins - Map Side, Distributed, Distributed Cache, Reduce Side Join
10. Counters
11. Map Reduce API & Data Types

Vskills Certified Big Data and Apache Hadoop Developer

Module 5: Pig

1. Introduction to Apache Pig
2. Pig Data Types
3. Operators in Pig
4. Pig program structure and execution process
5. Joins & filtering using Pig
6. Group & co-group
7. Schema merging and redefining functions
8. Pig functions

Module 6: Hive

1. Understanding Hive
2. Hive Architecture & Components
3. Using Hive command line interface
4. Data types and file formats
5. Hive DDL & DML operations
6. Hive vs. RDBMS

Module 7: HBase

1. What is HBase
2. HBase architecture
3. HBase in Hadoop Ecosystem
4. HBase vs. HDFS
5. HBase Data model
6. Physical Model in HBase
7. Components of HBase
8. Managing large data sets with HBase
9. Using HBase in Hadoop applications

Module 8: Sqoop

1. Introducing Sqoop
2. The principles of Sqoop Design
3. Connectors and Drivers
4. Importing Data with Sqoop
5. Exporting Data with Sqoop

Module 9: ZooKeeper

1. Overview of Zookeeper
2. How ZooKeeper Works
3. The ZooKeeper CLI
4. Reading and Writing Data
5. Sequential and Ephemeral znodes
6. Watches

Vskills Certified Big Data and Apache Hadoop Developer

7. Versioning and ACLs
8. Zookeeper use cases

Module 10: Flume

1. Flume Overview
2. Channels
3. Sinks and Sink Processors
4. Sources and Channel Selectors
5. Interceptors, ETL, and Routing
6. Monitoring Flume

Module 11: Oozie

1. Introduction to Oozie
2. Oozie - Simple/Complex Flow
3. Oozie - Components
4. Oozie Service/ Scheduler
5. Use Cases - Time and Data triggers
6. Running/Debugging a Coordinator Job
7. Bundle

Module 12: Yarn

1. History of Yarn
2. Core Components
3. YARN Administration
4. Capacity Scheduler
5. YARN Distributed-shell

Module 13: Troubleshooting, Administering and Optimizing Hadoop

1. Planning a Hadoop Cluster
2. Identity, Authentication and Authorization
3. Resource Management
4. Cluster Maintenance
5. Troubleshooting
6. Monitoring
7. Backup and Recovery

Module 14: Real-Time Projects

1. Twitter Data Analysis
2. Stack Exchange Ranking and Percentile data-set
3. Loan Dataset
4. Data-sets by Government
5. Machine Learning Dataset like Badges datasets
6. NYC Data Set
7. Weather Dataset

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 block 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

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 & Finance

- Certified AML- KYC Compliance Officer
- Certified Business Accountant
- Certified Commercial Banker
- Certified Equity Research Analyst
- Certified Foreign Exchange Professional
- Certified Hedge Fund Manager
- Certified Merger and Acquisition Analyst
- Certified Tally 9.0 Professional
- Certified Treasury Markets Professional
- Certified Wealth Manager

► Foreign Trade

- Certified Export Import (Foreign Trade) Professional

► Hospitality

- Certified Restaurant Team Member (Hospitality)

► Human Resources

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

► Logistics & Supply Chain Management

- Certified International Logistics Professional
- Certified Logistics & SCM Professional
- Certified Purchase Manager

► Law

- Certified IPR & Legal Manager

► Life Skills

- Certified Business Communication Specialist
- Certified Public Relations Officer

► Media

- Certified Advertising Manager
- Certified Advertising Sales Professional

► Office Skills

- Certified Data Entry Operator
- Certified Office Administrator

► Project Management

- Certified Project Management Professional

► Real Estate

- Certified Real Estate Consultant

► Information Technology

- Certified Android Apps Developer
- Certified ASP.NET Programmer
- Certified Basic Network Support Professional
- Certified Business Intelligence Professional
- Certified C# Professional
- Certified CAD Professional
- Certified Cloud Computing Professional
- Certified Computer Fundamentals (MS Office) Professional
- Certified Core Java Developer
- Certified CSS Designer
- Certified Data Mining and Warehousing Professional
- Certified DHTML & Javascript Developer
- Certified Django Developer
- Certified DTP operator
- Certified E-commerce Professional
- Certified E-Governance Professional
- Certified Enterprise Applications Integration Specialist (Biztalk)
- Certified Ethical Hacking and Security Professional
- Certified Facebook Apps Developer
- Certified Grid Computing Professional
- Certified Hadoop and Mapreduce Professional
- Certified HTML Designer
- Certified HTML5 Developer
- Certified iPhone Apps Developer
- Certified IT Support Professional
- Certified J2ME Programmer
- Certified Joomla Developer
- Certified Linux Administrator
- Certified Magento Professional
- Certified MySQL DB Administrator
- Certified Network Security Professional
- Certified Open Source CMS (Drupal) Professional
- Certified PHP Professional
- Certified PL/SQL Developer
- Certified Python Professional
- Certified Router Support Professional
- Certified Selenium Professional
- Certified SEO Professional
- Certified Software Quality Assurance Professional
- Certified Software Security Professional
- Certified Software Testing Professional
- Certified SQL Server 2008 Programmer
- Certified WiMax(4G) Professional
- Certified Wordpress Developer
- Certified XML Developer

► Sales, BPO

- Certified Telesales Executive

Contact us at:

V-Skills

011-473 44 723 or info@vskills.in

Intelligent Communication Systems India Limited

DSIIDC Administrative Building,
Okhla Industrial Estate-III, New Delhi-110020

www.vskills.in