

## **HCIA-Big Data**

Huawei Certified ICT Associate- Big Data

Training and certificating engineers who are capable of using Huawei MRS big data development platform

### **Overview**

(1) Master the technical principles and architectures of common and important big data components, including HDFS, HBase, Hive, ClickHouse, MapReduce, YARN, Spark, Flink, Flume, Kafka, ElasticSearch and ZooKeeper, and capable of using Huawei big data platform MRS

- (2) Able to operate and develop services based on Huawei MRS
- (3) Competent for positions related to big data development engineers

## **Objectives**

After completing the training, you will be able to:

1. Master the technical principles and architectures of common and important big data components, including HDFS, HBase, Hive, ClickHouse

2. MapReduce, YARN, Spark, Flink, Flume, Kafka, ElasticSearch and ZooKeeper, and capable of using Huawei big data platform MRS

- 3. Operate and develop services based on Huawei MRS
- 4. Competent for positions related to big data development engineers

5. Have the knowledge and skills required for big data pre-sales, big data project management, and big data development

#### **Training Content**

- > Big Data Development Trends and the Kunpeng Big Data Solution
  - Big Data Era
    - **Big Data Application Fields**
  - Challenges and Opportunities Faced by Enterprises
  - Huawei Kunpeng Solution
- HDFS Hadoop Distributed File System & ZooKeeper
  - HDFS HDFS Overview HDFS-related Concepts HDFS Architecture HDFS Key Features HDFS Data Read/Write Process ZooKeeper Distributed Coordination Service



ZooKeeper Overview ZooKeeper Architecture

- HBase Distributed Database & Hive Distributed Data Warehouse HBase – Distributed Database HBase Overview and Data Models HBase Architecture HBase Performance Tuning Common Shell Commands of HBase Hive – Distributed Data Warehouse Hive Overview Hive Functions and Architecture Basic Hive Operations
- ClickHouse Online Analytical Processing Database Management System ClickHouse Overview
   ClickHouse Architecture and Basic Features
   Enhanced Features of ClickHouse
- MapReduce and YARN Technical Principles MapReduce and YARN Overview
   Functions and Architectures of MapReduce and YARN Resource Management and Task Scheduling in YARN Enhanced Features of YARN
- Spark In-memory Distributed Computing Engine & Flink Stream and Batch Processing in a Single Engine
  - Spark In-memory Distributed Computing Engine Spark Overview Spark Data Structure Spark Principles and Architecture Flink — Stream and Batch Processing in a Single Engine Flink Principles and Architecture Flink Time and Window Flink Watermark Flink Fault Tolerance Mechanism
- Flume's Massive Log Aggregation & Kafka's Distributed Messaging System Flume: Massive Log Aggregation Overview and Architecture Key Features Applications Kafka: Distributed Messaging System Overview



Architecture and Functions Data Management

- Elasticsearch Distributed Search Engine
  Overview
  System Architecture
  Key Features
- MRS Huawei's Big Data Platform
  Overview of MRS
  MRS Components
  MRS Cloud-Native Data Lake Baseline Solution
- Huawei DataArts Studio
  Data Governance
  Huawei DataArts Studio



# HCIP-Big Data Developer

Huawei Certifed ICT Professional-Big Data Developer

Training and certificating senior engineers with big data application development capabilities

### **Overview**

With HCIP-Big Data Developer V2.0 certification, you will be able to use Huawei MRS or opensource Hadoop platform components to develop applications in different service scenarios, including offline batch processing, real-time retrieval, real-time stream, and converged data warehouse scenarios. In addition, you will have E2E big data solution development capabilities, and are competent for big data application development positions.

## **Objectives**

On completion of this program, the participants will be able to:

Master advanced theories and practice methods of big data development.
 Master advanced usage methods of big data development on HUAWEI CLOUD.
 Master the database routine management methods based on DWS

#### **Training Content**

HCIP-Big Data Developer

- 1. Big Data Scenario-based Solution Offline Batch Processing Theory
  - a. Offline processing solution
  - b. Introduction, technical principles, parameter attributes, and important configurations of the offline batch processing components, including data storage (HDFS), data warehouse (Hive), offline analysis tool (SparkSQL), and data collection tools (Loader, Sqoop, and Kettle)
  - c. Offline batch processing cases
- 2. Big Data Application Development Guide
  - a. Class opening
  - b. Mainstream big data technologies
  - C. Scenario-based big data solution
  - d. Big data application development
- 3. Big Data Scenario-based Solution Real-Time Stream Processing Theory-1
  - a. Real-time stream computing application solution
  - b. Introduction, technical principles, parameter attributes, and important configurations of the real-time stream computing components, including Flume, Kafka, Flink, SparkStreaming, and Redis
- 4. Big Data Scenario-based Solution Real-Time Retrieval Theory-2



- a. Real-time retrieval cases
- 5. Big Data Scenario-based Solution Real-Time Retrieval-practice-1
  - a. Project practice of the scenario-based solution real-time retrieval
- 6. Big Data Scenario-based Solution Real-Time Stream Processing Theory-2
  - a. Real-time stream processing configuration solution and success cases
- 7. Big Data Scenario-based Solution Converged Data Warehouse Theory-2
  - a. Converged data warehouse cases
- 8. Big Data Scenario-based Solution Converged Data Warehouse-practice-1
  - a. Project practice of the scenario-based solution converged data warehouse
- 9. Big Data Scenario-based Solution Real-Time Retrieval-practice-2
  - a. Project practice of the scenario-based solution real-time retrieval
- 10. Big Data Scenario-based Solution Real-Time Stream Processing practice-2
  - a. Project practice of the scenario-based solution real-time streaming processing
- 11. Big Data Scenario-based Solution Real-Time Stream Processing practice-1
  - a. Project practice of the scenario-based solution real-time streaming processing
- 12. Big Data Scenario-based Solution Converged Data Warehouse-practice-2
  - a. Project practice of the scenario-based solution converged data warehouse
  - b. Class closing
- 13. Big Data Scenario-based Solution Converged Data Warehouse Theory-1
  - a. Background of data warehouse
  - b. Introduction to DWS
- 14. Big Data Scenario-based Solution Real-Time Retrieval Theory-1
  - a. Real-time retrieval solution
  - b. Introduction, technical principles, parameter attributes, and important configurations of distributed databases, including HBase, Elasticsearch, and GraphBase
- 15. Big Data Scenario-based Solution Offline Batch Processing-practice-2
  - a. Project practice of the scenario-based solution offline processing
- 16. Big Data Scenario-based Solution Offline Batch Processing-practice-1
  - a. Project practice of the scenario-based solution offline processing



# HCIE-Big Data Developer

Huawei Certified ICT Expert-Big Data-Data Mining

Training and certification Experts with the ability to conduct end-to-end modeling of big data through the Huawei Cloud MRS platform and open source technology platform to solve practical problems related to data mining business

### **Overview**

Master the commonly used big data mining process (including data preprocessing, feature engineering, modeling, model evaluation and optimization); be familiar with the process of using PySpark for big data mining; master the basic concepts of data lakes and the operation skills of data entering the lake; master the data governance methodology and the characteristics of DataArts Studio, Huawei's data governance center

## **Objectives**

After completing this program training, you will be able to:

- > Master common big data mining processes (including data preprocessing, feature
- > Engineering, modeling, model evaluation and optimization).
- > Familiar with the process of using PySpark for big data mining.
- Master the basic concepts of data lakes and the operation skills of data entry into the lake.
- Master the methodology of data governance and the features of DataArts Studio, Huawei's data governance center.

#### **Training Content**

HCIE-Big Data-Data Mining V3.0 Training

- 1. Introduction to Data Mining
  - a. Data Mining Overview
  - b. Process of Data Mining
- 2. Data Preprocessing and Feature Engineering
  - a. Data Preprocessing
  - b. Feature Engineering
- 3. Regression and Classification
  - a. Classification Algorithms
  - b. Regression Algorithms
  - C. Ensemble Algorithms
- 4. Clustering and Dimensionality Reduction
  - a. Clustering Algorithms



- b. Dimensionality Reduction Algorithms
- 5. Association Analysis and Recommendation
  - a. Association Algorithms
  - b. Recommendation Algorithms
- 6. Model Evaluation and Optimization
  - a. Prerequisites for Model Evaluation and Optimization
  - b. Optimization Models
  - C. Model Evaluation and Selection
- 7. Python Data Mining Case Analysis
  - a. Data Reading
  - b. Feature Understanding and Analysis
  - C. Data Preprocessing
  - d. Modeling
- 8. PySpark MLlib
  - a. PySpark MLlib Basics
  - b. Basic Statistical Analysis of PySpark MLlib
  - C. Feature Extraction and Transformation of PySpark MLlib
  - d. Classification and Regression of PySpark MLlib
  - e. Clustering and Dimensionality Reduction of PySpark MLlib
  - f. Association Rules and Recommendation Algorithms of PySpark MLlib
  - g. Evaluation Matrix of PySpark MLlib
- 9. Huawei Big Data Platform MRS
  - a. MRS Cloud-Native Data Lake Baseline Solution
  - b. Components
- 10. Huawei DataArts Studio
  - a. Data Governance Methodology