HCIA-Big Data V3.5 Change Instructions

Version	Date	Instruction
V3.0	2021/1/29	Enable V3.0 training materials, lab guidebooks and exam questions
V3.5	2022/12/16	Enable V3.5 training materials, lab guidebooks and exam questions V3.5 was released on December 16, 2022, and V3.0 will become unavailable from June 16, 2023. The old and new versions coexist for 6 months.

This document is applicable to the version refreshed certification product. It is used to compare the new and old version of the training materials, lab manuals and detailed changes of the test. It is necessary to specify the contents of the addition, deletion or optimization of the knowledge points and their proportions.

1. Training material change instructions

In the PPT unit, each training material needs to Added a page of training material change instructions to introduce the material change point (optimization, addition or deletion) relative to the old material.

Chapter	V3.0	V3.5	Change Instructions
1. Big Data Development Trend and Kunpeng Big Data Solution	5%	5%	Optimized the big data development trend and Kunpeng big data, accounting for 5%.
2. HDFS — Hadoop Distributed File System & ZooKeeper	10%	10%	Optimized the HDFS distributed file system and ZooKeeper, accounting for 10%.
3. HBase — Distributed Database & Hive — Distributed Data Warehouse	20%	20%	Combined HBase technical principles and Hive distributed data warehouse into one chapter, and optimized HBase distributed database and Hive distributed data warehouse, accounting for 20%.
4. ClickHouse — Online Analytical Processing Database Management System	0%	10%	Added the ClickHouse online analytical database management system, accounting for 10%.
5. MapReduce and YARN Technical Principles	10%	10%	Optimized MapReduce and YARN technical principles, accounting for 10%.
6. Spark — In-memory Distributed Computing Engine & Flink — Stream and Batch Processing in a Single Engine	20%	20%	Combined Spark's in-memory distributed computing engine and Flink's stream and batch processing in a single engine into one chapter, and optimized their content, accounting for 20%.

7. Flume's Massive Log Aggregation & Kafka's Distributed Messaging System	10%	10%	Combined Flume's massive log aggregation and Kafka's distributed messaging system into one chapter, and optimized their content, accounting for 10%.
8. Elasticsearch — Distributed Search Engine	5%	5%	Optimized the ElasticSearch distributed full-text search service, accounting for 5%.
9. MRS Huawei's Big Data Platform	0%	5%	Added Huawei's big data platform MRS, accounting for 5%.
10. Huawei DataArts Studio	0%	5%	Added Huawei's data governance center DataArts Studio, accounting for 5%.

2. Lab Guides change instructions

In Word, each lab manual needs to Added a page of experiment manual knowledge point change instructions to introduce the knowledge change points (optimization, addition or deletion) of the experiment manual relative to the old lab manual.

HCIA-Big Data V3.5 Lab Guides

Chapter	V3.0	V3.5	Change Instructions
1.HDFS Practice	10%	10%	Optimized the HDFS distributed file system practice, accounting for 10%.
2.HBase Columnar Database Practice	15%	15%	Optimized HBase column-oriented database practices, accounting for 15%.
3.Hive Data Warehouse Practice	15%	15%	Optimized Hive data warehouse practices, accounting for 15%.
4.ClickHouse Online Analysis Database Practice	0%	5%	Added ClickHouse online analytical processing database management system practices, accounting for 5%.
5.MapReduce Data Processing Practice	10%	10%	Optimized MapReduce data processing practices, accounting for 10%.
6.Spark Memory Computing Practice	10%	10%	Optimized Spark memory computing practices, accounting for 10%.
7.Flink Real-Time Processing System Practice	10%	10%	Optimized Flink real-time processing system practices, accounting for 10%.
8.Flume Data Collection Practice	10%	10%	Optimized Flume data collection practices, accounting for 5%.
9.Kafka Message Subscription Practice	5%	5%	Optimized Kafka message subscription practices, accounting for 10%.
10.Comprehensive Exercise: Hive Data Warehouse	10%	10%	Optimized the comprehensive exercise: Hive data warehouse, accounting for 5%.

3. Test content change instructions

3.1 HCIA-Big Data V3.5 Test content：

Chapter	V3.0	Chapter	V3.5	Change Instructions
1. Big Data Development Trend and Kunpeng Big Data 1.1 Big Data Era 1.2 Big Data Application Fields 1.3 Opportunities and Challenges in the Big Data Era 1.4 Huawei Kunpeng Big Data Ecosystem	3%	1. Big Data Development Trend and Kunpeng Big Data 1.1 Big Data Era 1.2 Big Data Application Fields 1.3 Challenges and Opportunities Faced by Enterprises 1.4 Huawei Kunpeng Solution	3%	Optimized the big data development trend and Kunpeng big data.
2. HDFS Distributed File System and ZooKeeper 2.1 HDFS Overview and Application Scenarios 2.2 HDFS System Architecture 2.3 Key Functions 2.4 ZooKeeper Principles and Deployment Solution	12%	2. HDFS — Hadoop Distributed File System & ZooKeeper 2.1 HDFS 2.1.1 HDFS Overview 2.1.2 HDFS-related Concepts 2.1.3 HDFS Architecture 2.1.4 HDFS Key Features 2.1.5 HDFS Data Read/Write Process 2.2 ZooKeeper Distributed Coordination Service 2.2.1 ZooKeeper Overview 2.2.2 ZooKeeper Architecture	12%	Optimized the HDFS distributed file system and ZooKeeper.
3. Technical Principles of Hive Distributed Data Warehouse and HBase 3.1 Hive Overview and Hive Core Modules 3.2 Hive Functions, Architecture, and Table Types 3.3 Basic Hive Operations and SQL Connection 3.4 HBase Overview 3.5 HBase Functions and Architecture 3.6 HBase Key Processes	21%	3. HBase — Distributed Database & Hive — Distributed Data Warehouse 3.1 HBase — Distributed Database 3.1.1 HBase Overview and Data Models 3.1.2 HBase Architecture 3.1.3 HBase Performance Tuning 3.1.4 Common Shell Commands of HBase 3.2 Hive — Distributed Data Warehouse	20%	Combined HBase technical principles and Hive distributed data warehouse into one chapter, and optimized HBase distributed database and Hive distributed data warehouse.

		3.2.1 Hive Overview 3.2.2 Hive Functions and Architecture 3.2.3 Basic Hive Operations			
4. ClickHouse — Online Analytical Processing Database Management System	0%	4. ClickHouse — Online Analytical Processing Database Management System 4.1 ClickHouse Overview 4.2 ClickHouse Architecture and Basic Features 4.3 Enhanced Features of ClickHouse	8%	Added the ClickHouse online analytical database management system.
5. MapReduce and YARN Principles 5.1 Introduction to MapReduce and YARN 5.2 Functions and Architectures of MapReduce and YARN 5.3 Resource Management and Task Scheduling of YARN 5.4 Enhanced Features of YARN	9%	5. MapReduce and YARN Principles 5.1 MapReduce and YARN Overview 5.2 Functions and Architectures of MapReduce and YARN 5.3 Resource Management and Task Scheduling of YARN 5.4 Enhanced Features of YARN	12%	Optimized MapReduce and YARN technical principles.
6. Spark — In-memory Distributed Computing Engine & Flink — Stream and Batch Processing in a Single Engine 6.1 Spark Overview 6.2 Spark Principle and Architecture, RDD Wide and Narrow Dependency, Lineage Mechanism, Transformation and Action Operators 6.3Spark Streaming, DataFrame, DataSet 6.4Spark SQL 6.5 Flink Overview 6.6 Flink Principles and Technical Architecture 6.7 Flink window, window type, and time mechanism 6.8 Flink status management and Flink status management mechanism	15%	6. Spark — In-memory Distributed Computing Engine & Flink — Stream and Batch Processing in a Single Engine 6.1 Spark — In-memory Distributed Computing Engine 6.1.1 Spark Overview 6.1.2 Spark Data Structure 6.1.3 Spark Principles and Architecture 6.2 Flink — Stream and Batch Processing in a Single Engine 6.2.1 Flink Principles and Architecture 6.2.2 Flink Time and Window 6.2.3 Flink Watermark 6.2.4 Flink Fault	20%	Combined Spark's in-memory distributed computing engine and Flink's stream and batch processing in a single engine into one chapter, and optimized their content.

		Tolerance Mechanism		
7. Flume massive log aggregation and Kafka distributed message subscription system 7.1 Flume Massive Log Aggregation 7.2 Introduction and Architecture of Flume 7.3 Introduction to Key Features of Flume 7.4 Flume Application Example 7.5 Introduction to Kafka 7.6 Kafka Architecture and Functions 7.7 Kafka Data Management	16%	7. Flume's Massive Log Aggregation & Kafka's Distributed Messaging System 7.1 Flume: Massive Log Aggregation 7.1.1 Overview and Architecture 7.1.2 Key Features 7.1.3 Applications 7.2 Kafka: Distributed Message Subscription System 7.2.1 Overview 7.2.2 Architecture and Functions 7.2.3 Data Management	12%	Combined Flume's massive log aggregation and Kafka's distributed messaging system into one chapter, and optimized their content.
8. Elasticsearch Distributed Search Engine 8.1 Introduction to Elasticsearch 8.2 Elasticsearch Architecture and Cache Mechanism 8.3 Elasticsearch Index Principles 8.4 Elasticsearch API Operations	5%	8. Elasticsearch — Distributed Search Engine 8.1 Overview 8.2 System Architecture 8.3 Key Features	5%	Optimized the ElasticSearch distributed full-text search service.
9. MRS Huawei's Big Data Platform	0%	9. MRS Huawei's Big Data Platform 9.1 Overview of MRS 9.2 MRS Components 9.3 MRS Cloud-Native Data Lake Baseline Solution	4%	Added Huawei's big data platform MRS.
10. Huawei DataArts Studio	0%	10. Huawei DataArts Studio 10.1 Data Governance 10.2 Huawei DataArts Studio	4%	Added Huawei's data governance center DataArts Studio.