Congratulations! You passed!

Grade received 98.70% **Latest Submission Grade** 98.70% **To pass** 80% or higher

Go to next item

1. What is data-parallelism as defined in lecture? **1 / 1 point**

- ○ At each step of the data pipeline, process values simultaneously by using multiple cores.
- ○ Simultaneously processing input data from multiple cores.
- ◉ Running the same function simultaneously for the partitions of a data set on multiple cores.
- ○ Having multiple multiple data pipelines at the same time.

Correct

2. Of the following, which procedure best generalizes big data procedures such as (but not limited to) the map reduce process? **1 / 1 point**

- ◉ split->do->merge
- ○ split->map->shuffle and sort->reduce
- ○ split ->shuffle and sort->map->reduce
- ○ split->sort->merge

Correct

3. What are the three layers for the Hadoop Ecosystem? (Choose 3) **1 / 1 point**

- ☑ Coordination and Workflow Management

Correct

- ☐ Data Manipulation and Integration
- ☑ Data Management and Storage

Correct

- ☑ Data Integration and Processing

Correct

- ☐ Data Creation and Storage

4. **1 / 1 point**

What are the 5 key points in order to categorize big data systems?

○

Coordination, Latency, Productivity, Speed, Fault Tolerance

Execution model, Speed, Scalability, Flexibility, Fault Tolerance

Execution model, Latency, Scalability, Programming Language, Fault Tolerance

Coordination, Latency, Productivity, Flexibility, Fault Tolerance

Correct

5. What is the lambda architecture as shown in lecture? **1 / 1 point**

A type of architecture that only contains part of the data processing method.

A type of swappable data processing layer.

An architecture that natively supports lambda calculus.

A type of hybrid data processing architecture.

Correct

6. Which of the following scenarios is **NOT** an aggregation operation? **1 / 1 point**

Averaging the total number of data per type.

Counting the total number of data per type.

Removing undefined values.

Counting the total number of data.

Correct

7. **1 / 1 point**

What usually happens to data when aggregated as mentioned in lecture?

Data becomes personalized.

Data become organized.

Data becomes smaller.

Data becomes faster to process.

Correct

8. 1 / 1 point

What is K-means clustering?

- Classify data by k decisions.
- Divide samples using k lines.
- Classify data by k actions.
- (selected) Group samples into k clusters.

Correct

9. 0.8571428571428571 / 1 point

Why is Hadoop not a good platform for machine learning as mentioned in lecture? (Choose 4)

- (selected) Map and Reduce Based Computation.

Correct

- (selected) No interactive shell and streaming.

Correct

- Requires nodes and multiple machines.
- Too massive.
- Java support only.
- (selected) Bottleneck using HDFS.

Correct

- Unable to support machine learning.

You didn't select all the correct answers

10. What are the layers (parts) of Spark? (Choose 5) 1 / 1 point

- (selected) Spark Streaming

Correct

- (selected)

Spark Core

Correct

Graphx

Correct

MLlib

Correct

Worker Node

Spark Graph

SparkSQL

Correct

Spark RDD

11. 1 / 1 point

What is in-memory processing?

Writing data to disk between pipeline steps.

Having the input completely in disk.

Writing data to memory between pipeline steps.

Having the pipeline completely in memory.

Having the pipeline completely in disk.

Having the input completely in memory.

Correct