Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Pipeline and Tools

LATEST SUBMISSION GRADE
100%

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

- ○ Simultaneously processing input data from multiple cores.
- ○ At each step of the data pipeline, process values simultaneously by using 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->sort->merge
- ○ split ->shuffle and sort->map->reduce
- ○ split->map->shuffle and sort->reduce
- ◉ split->do->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 Creation and Storage
- ☑ Data Integration and Processing

✓ **Correct**

- ☑ Data Management and Storage

✓ **Correct**

4. — 1 / 1 point

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

- ○

Coordination, Latency, Productivity, Flexibility, Fault Tolerance

Execution model, Speed, Scalability, Flexibility, Fault Tolerance

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

Coordination, Latency, Productivity, Speed, Fault Tolerance

✓ **Correct**

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

An architecture that natively supports lambda calculus.

A type of swappable data processing layer.

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

A type of hybrid data processing architecture.

✓ **Correct**

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

Counting the total number of data per type.

Averaging the total number of data per type.

Counting the total number of data.

Removing undefined values.

✓ **Correct**

7. **1 / 1 point**

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

Data become organized.

Data becomes smaller.

Data becomes personalized.

Data becomes faster to process.

✓ **Correct**

8. **1 / 1 point**

What is K-means clustering?

Classify data by k actions.

- [ ] Classify data by k decisions.
- [x] Group samples into k clusters.
- [ ] Divide samples using k lines.

✓ **Correct**

9. **1 / 1 point**

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

- [ ] Too massive.
- [x] Java support only.

✓ **Correct**

- [x] Bottleneck using HDFS.

✓ **Correct**

- [x] No interactive shell and streaming.

✓ **Correct**

- [ ] Requires nodes and multiple machines.
- [x] Map and Reduce Based Computation.

✓ **Correct**

- [ ] Unable to support machine learning.

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

- [x] Graphx

✓ **Correct**

- [ ] Worker Node
- [x] Spark Streaming

✓ **Correct**

- [x] Spark Core

✓ **Correct**

- [x]

SparkSQL

✓ **Correct**

☐ Spark RDD

☑ MLlib

✓ **Correct**

☐ Spark Graph

11. **1 / 1 point**

What is in-memory processing?

○ Having the pipeline completely in disk.

○ Having the input completely in disk.

○ Writing data to disk between pipeline steps.

○ Having the input completely in memory.

◉ Writing data to memory between pipeline steps.

○ Having the pipeline completely in memory.

✓ **Correct**