Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

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

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

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 ->shuffle and sort->map->reduce
- ( ) split->map->shuffle and sort->reduce
- ( ) split->sort->merge

Correct

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

- [x] Data Management and Storage

Correct

- [ ] Data Creation and Storage
- [x] Coordination and Workflow Management

Correct

- [ ] Data Manipulation and Integration
- [x] Data Integration and Processing

Correct

4. What are the 5 key points in order to categorize big data systems? 1 / 1 point

- ( ) 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

- ( ) An architecture that natively supports lambda calculus.
- ( ) A type of architecture that only contains part of the data processing method.
- ( ) A type of swappable data processing layer.
- (●) A type of hybrid data processing architecture.

Correct

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

- (●) Removing undefined values.
- ( ) Counting the total number of data per type.
- ( ) Averaging the total number of data per type.
- ( ) Counting the total number of data.

Correct

7. What usually happens to data when aggregated as mentioned in lecture? 1 / 1 point

- ( ) Data become organized.
- ( ) Data becomes faster to process.
- (●) Data becomes smaller.
- ( ) Data becomes personalized.

Correct

8. What is K-means clustering? 1 / 1 point

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

Correct

9. Why is Hadoop not a good platform for machine learning as mentioned in lecture? (Choose 4) 1 / 1 point

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

Correct

- [x] Java support only.

Correct

- [x] No interactive shell and streaming.

Correct

- [ ] Unable to support machine learning.
- [x] Bottleneck using HDFS.

Correct

- [ ] Too massive.

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

- [x] Graphx

Correct

- [ ] Spark Graph
- [x] Spark Core

Correct

- [ ] Worker Node
- [x] SparkSQL

Correct

- [x] Spark Streaming

Correct

- [x] MLlib

Correct

- [ ] Spark RDD

11. What is in-memory processing? 1 / 1 point

- ( ) Having the input completely in disk.
- ( ) Having the pipeline completely in disk.
- ( ) Having the pipeline completely in memory.
- (●) Writing data to memory between pipeline steps.
- ( ) Writing data to disk between pipeline steps.
- ( ) Having the input completely in memory.

Correct