

Pipeline and Tools



11/11 得分 (100%)

测验通过！

继续课程

返回第 #4 周课程



1 / 1 分

1。

What is data-parallelism as defined in lecture?



Running the same function simultaneously for the partitions of a data set on multiple cores.



正确



Simultaneously processing input data from multiple cores.



Having multiple multiple data pipelines at the same time.



At each step of the data pipeline, process values simultaneously by using multiple cores.



1 / 1 分

2。

Of the following, which procedure best generalizes big data procedures such as (but not limited to) the map reduce process?

- ☐ split->sort->merge
- ☐ split ->shuffle and sort->map->reduce
- ☐ split->map->shuffle and sort->reduce
- ☒ split->do->merge

正确



1 / 1 分

3。

What are the three layers for the Hadoop Ecosystem? (Choose 3)



Data Integration and Processing

正确



Coordination and Workflow Management

正确



Data Management and Storage

正确



Data Creation and Storage

未选择的是正确的



Data Manipulation and Integration

未选择的是正确的



1 / 1 分

4。

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



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



正确



Coordination, Latency, Productivity, Speed, Fault Tolerance



Execution model, Speed, Scalability, Flexibility, Fault Tolerance



Coordination, Latency, Productivity, Flexibility, Fault Tolerance



1 / 1 分

5。

What is the lambda architecture as shown in lecture?



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



A type of hybrid data processing architecture.



正确



An architecture that natively supports lambda calculus.



A type of swappable data processing layer.



1 / 1 分

6.

Which of the following scenarios is **NOT** an aggregation operation?

- ☐ Counting the total number of data.
- ☐ Averaging the total number of data per type.
- ☐ Counting the total number of data per type.
- ☒ Removing undefined values.



正确



1 / 1 分

7.

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

- ☒ Data becomes smaller.



正确

- ☐ Data become organized.
- ☐ Data becomes faster to process.
- ☐ Data becomes personalized.



1 / 1 分

8.

What is K-means clustering?



- ☐ Classify data by k actions.
- ☐ Classify data by k decisions.
- ☐ Divide samples using k lines.
- ☒ Group samples into k clusters.

正确



1 / 1 分

9.

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

- ☒ Bottleneck using HDFS.

正确

- ☒ Map and Reduce Based Computation.

正确

- ☐ Too massive.

未选择的是正确的

- ☒ Java support only.

正确

- ☐ Unable to support machine learning.

未选择的是正确的

☒ No interactive shell and streaming.



正确

☐ Requires nodes and multiple machines.



未选择的是正确的



1 / 1 分

10。

What are the layers (parts) of Spark? (Choose 5)

☐ Spark RDD



未选择的是正确的

☒ MLlib



正确

☒ Spark Streaming



正确

☐ Worker Node



未选择的是正确的

☒ Spark Core



正确

☒ SparkSQL



正确

☒ Graphx



正确

☐ Spark Graph



未选择的是正确的



1 / 1 分

11。

What is in-memory processing?

- ☐ Having the pipeline completely in disk.
- ☒ Writing data to memory between pipeline steps.



正确

- ☐ Having the input completely in disk.
- ☐ Writing data to disk between pipeline steps.
- ☐ Having the input completely in memory.
- ☐ Having the pipeline completely in memory.

