Pipeline and Tools



11/11 得分 (100%)

测验通过!

继续课程

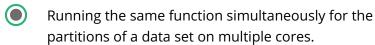
返回第#4周课程



1/1分

1。

What is data-parallelism as defined in lecture?





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.

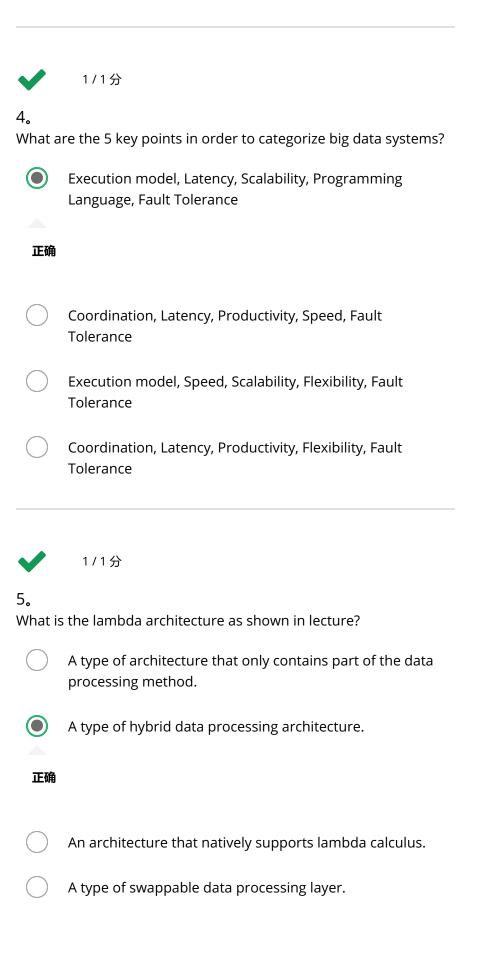


1/1分

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



V	1/1分					
5。 Which	of the following scenarios is NOT an aggregation operation?					
	Counting the total number of data.					
	Averaging the total number of data per type.					
\bigcirc	Counting the total number of data per type.					
	Removing undefined values.					
正确						
~	1/1分					
7。 What u ecture	sually happens to data when aggregated as mentioned in ?					
	Data becomes smaller.					
正确						
	Data become organized.					
	Data becomes faster to process.					
	Data becomes personalized.					

8. What is K-means clustering?

1/1分

	Classify data by k actions.					
\bigcirc	Classify data by k decisions.					
	Divide samples using k lines.					
	Group samples into k clusters.					
正确						
~	1/1分					
	Hadoop not a good platform for machine learning as ned in lecture? (Choose 4)					
<u> </u>	Bottleneck using HDFS.					
正确						
正确	Map and Reduce Based Computation.					
	Too massive.					
未选择	圣的是正确的					
✓ 	Java support only.					
正确						
	Unable to support machine learning.					

未选择的是正确的

正确	No interactive shell and streaming.
未选择	Requires nodes and multiple machines. 译的是正确的
√ 10。 What a	1 / 1 分 re the layers (parts) of Spark? (Choose 5)
	Spark RDD
未选择	军的是正确的
正确	MLlib
正确	Spark Streaming
	Worker Node
未选择	幹的是正确的
✓ _	Spark Core

正确

\checkmark	SparkSQL					
正确						
	Graphx					
正确						
	Spark Graph					
未选择的是正确的						
~	1/1分					
11。 What is	s 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.					