

MapReduce章节测试

题量: 32 满分: 100

智能分析

作答时间: 11-24 12:44 至 11-28 12:44

100分

一. 单选题 (共10题, 31分)

1. (单选题)有关MapReduce, 下面哪个说法是正确的 ()

- A. 它提供了资源管理能力
- B. 它是开源数据仓库系统, 用于查询和分析存储在Hadoop中的大型数据集
- C. 它是Hadoop数据处理层
- D. 它提供了进程管理能力

我的答案: C 正确答案: C



3.1 分

2. (单选题)下列传统并行计算框架, 说法错误的是哪一项?

- A. 刀片服务器、高速网、SAN, 价格贵, 扩展性差上
- B. 共享式(共享内存/共享存储), 容错性好
- C. 实时、细粒度计算
- D. 计算密集型

我的答案: B 正确答案: B



3.1 分

3. (单选题)关于MapReduce中的键值对, 下面陈述正确的是 ()

- A. Key类必须实现Writable
- B. Key类必须实现WritableComparable
- C. Value类必须实现WritableComparable
- D. Value类必须继承WritableComparable

一. 单选题 (31分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 多选题 (27.9分)

- 11
- 12
- 13
- 16
- 17
- 18

三. 填空题 (15.5分)

- 20
- 21
- 22

四. 判断题 (25.6分)

- 25
- 26
- 27
- 30
- 31
- 32

作业详情

一. 单选题 (31分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 多选题 (27.9分)

- 11
- 12
- 13
- 16
- 17
- 18

三. 填空题 (15.5分)

- 20
- 21
- 22

四. 判断题 (25.6分)

- 25
- 26
- 27
- 30
- 31
- 32

4. (单选题)在HDFS中，put命令用于（）
- A. 将文件从本地文件系统拷贝到HDFS
- B. 将文件或者目录从本地文件系统拷贝到HDFS
- C. 将文件从HDFS拷贝到本地文件系统

- D. 将文件或者目录从HDFS拷贝到本地文件系统

我的答案: B

正确答案: B

✓

3.1 分

5. (单选题)在MapReduce中，下面哪个数据流的顺序是正确的（）
- a. InputFormat b. Mapper c. Combiner d. Reducer e. Partitioner f. OutputFormat

- A. abcdfe
- B. abcedf
- C. acdefb
- D. abcdef

我的答案: B

正确答案: B

✓

3.1 分

6. (单选题)下面哪个不属于Reducer阶段（）
- A. Shuffle
- B. Sort
- C. Map
- D.

我的答案: C

正确答案: C

✓

3.1 分

7. (单选题)
shuffle和sort输出的key实现了下面哪个接口 ()

- A. Writable
- B. WritableComparable
- C. Configurable
- D. ComparableWritable

我的答案: B

正确答案: B

✓

3.1 分

8. (单选题)下列关于MapReduce模型描述，错误的是哪一项？

- A. MapReduce采用“分而治之”策略
- B. MapReduce设计的一个理念就是“计算向数据靠拢”
- C. MapReduce框架采用了Master/Slave架构
- D. MapReduce应用程序只用Java来写

我的答案: D

正确答案: D

✓

3.1 分

9. (单选题)MapReduce的体系结构中，JobTracker的主要任务是什么？

- A. 负责资源监控和作业调度，监控所有TaskTracker与Job的健康状况
- B. 使用“slot”等量划分本节点上的资源量（CPU、内存等）
- C. 会周期性地通过“心跳”将本节点上资源的使用情况和任务的运行进度汇报给TaskTracker
- D. 会跟踪任务的执行进度、资源使用量等信息，并将这些信息告诉任务（Task）

我的答案: A

正确答案: A

✓

3.1 分

一. 单选题 (31分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 多选题 (27.9分)

- 11
- 12
- 13
- 16
- 17
- 18

三. 填空题 (15.5分)

- 20
- 21
- 22

四. 判断题 (25.6分)

- 25
- 26
- 27
- 30
- 31
- 32

作业详情

- A.
Map将小数据集进一步解析成一批<key,value>对，输入Map函数中进行处理
- B.
Map每一个输入的<k 1,v 1>会输出一批<k 2,v 2>。<k 2,v 2>是计算的中间结果
- C.
Reduce输入的中间结果<k 2,List(v 2)>中的List(v 2)表示是一批属于不同一个k 2的value
- D. Reduce输入的中间结果<k 2,List(v 2)>中的List(v 2)表示是一批属于同一个k 2的value

我的答案: C 正确答案: C  3.1 分

二. 多选题 (共9题, 27.9分)

11. (多选题)MapReduce相较于传统的并行计算框架有什么优势?

- A.
非共享式，容错性好
- B.
普通PC机，便宜，扩展性好
- C.
简单
- D. 批处理、非实时、数据密集型

我的答案: ABCD 正确答案: ABCD  3.1 分

12. (多选题)下列关于MapReduce的体系结构的描述，说法正确的有？

- A.
用户编写的MapReduce程序通过Client提交到JobTracker端
- B.
JobTracker负责资源监控和作业调度
- C.
TaskTracker监控所有TaskTracker与Job的健康状况
- D. TaskTracker 使用“slot”等量划分本节点上的资源量（CPU、内存等）

我的答案: ABD 正确答案: ABD  3.1 分

一. 单选题 (31分)

- 1 2 3
- 6 7 8

二. 多选题 (27.9分)

- 11 12 13
- 16 17 18

三. 填空题 (15.5分)


- 20 21 22

四. 判断题 (25.6分)

- 25 26 27
- 30 31 32

作业详情

- A. 从磁盘或从网络读取数据，即IO密集工作
- B. 计算数据，即CPU密集工作
- C. 针对不同的工作节点选择合适硬件类型
- D. 负责协调集群中的数据存储

我的答案: AB 正确答案: AB  3.1 分

14. (多选题)对于MapReduce 而言，其处理单位是split。split 是一个逻辑概念，它包含哪些元数据信息？

- A. 数据起始位置
- B. 数据长度
- C. 数据所在节点
- D. 数据大小

我的答案: ABCD 正确答案: ABCD  3.1 分

15. (多选题)下列关于Map 端的Shuffle的描述，哪些是正确的？

- A. MapReduce默认1000MB缓存
- B. 多个溢写文件归并成一个或多个大文件，文件中的键值对是排序的
- C. 当数据很少时，不需要溢写到磁盘，直接在缓存中归并，然后输出给Reduce
- D. 每个Map任务分配多个缓存，使得任务运行更有效率

我的答案: BC 正确答案: BC  3.1 分

16. (多选题)MapReduce的具体应用包括哪些？

- A.

一. 单选题 (31分)

- 1 2 3
- 6 7 8

二. 多选题 (27.9分)

- 11 12 13
- 16 17 18

三. 填空题 (15.5分)

- 20 21 22

四. 判断题 (25.6分)

- 25 26 27
- 30 31 32

作业详情

B.
分组与聚合运算

C.
矩阵-向量乘法

D. 矩阵乘法

我的答案: ABCD

正确答案: ABCD

✓

3.1分

17. (多选题)MapReduce执行的全过程包括以下哪几个主要阶段？

A.
从分布式文件系统读入数据

B.
执行Map任务输出中间结果

C.
通过 Shuffle阶段把中间结果分区排序整理后发送给Reduce任务

D. 执行Reduce任务得到最终结果并写入分布式文件系统

我的答案: ABCD

正确答案: ABCD

✓

3.1分

18. (多选题)
下列关于分布式并行编程的描述，哪些是正确的？

A.
“摩尔定律”，CPU性能大约每隔18个月翻一番

B.
分布式程序运行在大规模计算机集群上

C.
谷歌公司最先提出了分布式并行编程模型MapReduce

D. MapReduce是Hadoop的开源实现

我的答案: ABC

正确答案: ABC

✓

3.1分

19. (多选题)下列说法正确的是？

A.
MapReduce体系结构主要由四个部分组成，分别是：Client、JobTracker、TaskTracker以及Task

一. 单选题 (31分)

- 123
- 678

二. 多选题 (27.9分)

- 111213
- 161718

三. 填空题 (15.5分)

- 202122

四. 判断题 (25.6分)

- 252627
- 303132

作业详情

B.
Task 分为Map Task 和Reduce Task 两种，均由TaskTracker 启动

C.
在MapReduce工作流程中，所有的数据交换都是通过MapReduce框架自身去实现的

D. 在MapReduce工作流程中，用户不能显式地从一台机器向另一台机器发送消息

我的答案: ABCD

正确答案: ABCD

✓

3.1 分

三. 填空题 (共5题, 15.5分)

20. (填空题)MapReduce最早是由 () 开发的。

我的答案:

(1) 谷歌

✓

3.1 分

正确答案:

(1) Google Lab; 谷歌实验室; 谷歌; Google

21. (填空题)
MapReduce计算模型的核心是Map和Reduce两个函数，Map的意思是 () ， Reduce的意思是归约 。

我的答案:

(1) 映射

✓

3.1 分

正确答案:

(1) 映射

22. (填空题)
MapReduce借助于 () 程序设计语言（如Lisp等）的设计思想，提供了一种简便的并行程序设计方法。

我的答案:

(1) 函数式

✓

3.1 分

正确答案:

(1) 函数式

23. (填空题)
MapReduce体系结构由 () 、 JobTracker 、 TaskTracker 以及Task scheduler四个部分组成。

我的答案:

(1) client

✓

3.1 分

一. 单选题 (31分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 多选题 (27.9分)

- 11
- 12
- 13
- 16
- 17
- 18

三. 填空题 (15.5分)

- 20
- 21
- 22

四. 判断题 (25.6分)

- 25
- 26
- 27
- 30
- 31
- 32

一. 单选题 (31分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 多选题 (27.9分)

- 11
- 12
- 13
- 16
- 17
- 18

三. 填空题 (15.5分)

- 20
- 21
- 22

四. 判断题 (25.6分)

- 25
- 26
- 27
- 30
- 31
- 32

24. (填空题)
() 被称为MapReduce的“心脏”，是所谓“奇迹发生的地方”，它描述着数据从map task输出，到reduce task输入的这段过程。

我的答案: (1) Shuffle 3.1分

正确答案: (1) Shuffle

四. 判断题 (共8题, 25.6分)

25. (判断题)Hadoop框架是用Java语言实现的，因此，MapReduce应用程序则一定也要用Java来编写。

A. 对

B. 错

我的答案: 错 正确答案: 错 3.1分

26. (判断题)MapReduce编程模型的其中一个设计理念，就是“计算向数据靠拢”，而不是“数据向计算靠拢”。

A. 对

B. 错

我的答案: 对 正确答案: 对 3.1分

27. (判断题)MapReduce适宜计算密集型应用。

A. 对

B. 错

我的答案: 错 正确答案: 错 3.1分

28. (判断题)Mapreduce的input split就是一个block。

A. 对

B. 错

我的答案: 错 正确答案: 错 3.1分

29. (判断题)每个map槽就是一个线程。

A. 对

B. 错

作业详情

30. (判断题)不同的Map任务之间不会进行通信。

- A. 对
- B. 错

我的答案: 对

正确答案: 对

✓

3.1 分

31. (判断题)不同的Reduce任务之间，则可能会发生信息交换。

- A. 对
- B. 错

我的答案: 错

正确答案: 错

✓

3.1 分

32. (判断题)Input Split是MapReduce对文件进行处理和运算的输入单位，需要对文件进行物理切分。

- A. 对
- B. 错

我的答案: 错

正确答案: 错

✓

3.9 分

一. 单选题 (31分)

- 1

2

3
- 6

7

8

二. 多选题 (27.9分)

- 11

12

13
- 16

17

18

三. 填空题 (15.5分)

- 20

21

22

四. 判断题 (25.6分)

- 25

26

27
- 30

31

32