# MapReduce练习题

**单选题**

1. 有关MapReduce，下面哪个说法是正确的（C ）

A.它提供了资源管理能力（yarn）

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(Hive)

C.它是Hadoop数据处理层

2. 在MapReduce中，如果将reducer数设置为0会发生怎样的情形（C ）

A.仅有Reduce作业发生

B.仅有Map作业发生

C.Reducer输出会成为最终输出(Mapper输出是最终输出)

3. 在MapReduce中，下面哪个会将输入键值对处理成中间键值对（A ）

A.Mapper

B.Reducer

C.Mapper和Reducer

4. 在MapReduce中，Map数取决于（A ）的总量

A.任务数

B.输入数据

C.输出数据

5. 在Hadoop的分区阶段，默认的Partitioner是（C ）

A.HashPar

B.Partitioner

C.HashPartitioner

6. 下面哪个是一种编程模型，它将大规模的数据处理工作拆分成互相独立的任务然后并行处理（A ）

A.MapReduce

B.HDFS

C.Pig

7. 在Hadoop中，下面哪个是默认的InputFormat类型，它将每行内容作为新值，而将字节偏移量作为key（B ）

A.FileInputFormat

B.TextInputFormat

C.KeyValueTextInputFormat

8. 在MapReduce中，对于map输出的中间结果，负责按key进行分区的是（B ）

A. RecordReader

B. Combiner

C. Partitioner

9. 在MapReduce中，下面哪个阶段是可以同时（并行）进行的（C ）

A. Shuffle和Map

B. Shuffle和Sort

C. Reduce和Sort

10. 关于MapReduce中的键值对，下面陈述正确的是（B ）

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

11. 在HDFS中，put命令用于（B ）

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

12. 在Hadoop中，下面哪一个是默认的OutputFormat（D ）

A. SequenceFileOutputFormat

B. LazyOutputFormat

C. DBOutputFormat

D. TextOutputFormat

13. 下面哪个不是Hadoop的输入格式（A ）

A. ByteInputFormat

B. TextInputFormat

C. SequenceFileInputFormat

D. KeyValueInputFormat

14. 在MapReduce中，下面哪个数据流的顺序是正确的（B ）

a. InputFormat

b. Mapper

c. Combiner

d. Reducer

e. Partitioner

f. OutputFormat

A. abcdfe B. abecdf C. acdefb D. abcdef

必须是先分区，后合并

15. 下面哪个不属于Reducer阶段 （C ）

A. Shuffle

B. Sort

C. Map

D. Reduce

16. Mapper排序后的输出将作为下面哪个的输入（C ）

A. Reducer

B. Mapper

C. Shuffle

17. 如何禁用reduce阶段（B ）

A. 设置conf.setNumreduceTasks(0)

B. 设置job.setNumreduceTasks(0)

C 设置job.setNumreduceTasks()=0

18. shuffle和sort输出的key实现了下面哪个接口（B ）

A. Writable

B. WritableComparable

C. Configurable

D. ComparableWritable

19. 在MapReduce编程时，如下阶段的顺序是（B ）

a. Partitioner b. Mapper c. Combiner d. Shuffle/Sort

A. Mapper Partitioner Shuffle/Sort Combiner

B. Mapper Partitioner Combiner Shuffle/Sort

C. Mapper Shuffle/Sort Combiner Partitioner

D. Mapper Combiner Partitioner Shuffle/Sort

20. 下面生成中间键值对的是（B ）

A. Reducer B. Mapper C. Combiner D. Partitioner

21. 以下哪一项不属于Hadoop可以运行的模式\_\_C\_\_\_\_。

 A. 单机（本地）模式  B. 伪分布式模式  C. 互联模式  D. 分布式模式

22. Hadoop的作者是下面哪一位\_\_B\_\_\_\_。

 A. Martin Fowler  B. Doug cutting  C. Kent Beck  D. Grace Hopper

23. 下列关于MapReduce说法不正确的是(C )

A. MapReduce是一种计算框架

B. MapReduce来源于google的学术论文

C. MapReduce程序只能用java语言编写

D.MapReduce隐藏了并行计算的细节，方便使用。

24. HDFS是基于流数据模式访问和处理超大文件的需求而开发的，具有高容错、高可靠性、高可扩展性、高吞吐率等特征，适合的读写任务是（D ）

A. 一次写入，少次读

B. 多次写入，少次读

C. 多次写入，多次读

D. 一次写入，多次读

25. HBase依靠（A ）存储底层数据

A. HDFS

B. Hadoop

C. Memory

D. MapReduce

26. 下面与HDFS类似的框架是( C )

A. NTFS

B. FAT32

C. GFS

D.EXT3

27. 关于SecondaryNameNode下面哪项是正确的（C ）

A. 它是NameNode的热备

B. 它对内存没有要求

C. 它的目的是帮助NameNode合并编辑日志，减少NameNode启动时间

D. SecondaryNameNode应与NameNode部署到一个节点。

28. 大数据的特点不包括下面哪一项（D ）

A. 巨大的数据量

B. 多结构化数据

C. 增长速度快

D. 价值密度高