

考试类型： 闭卷

Hadoop 大数据开发试卷(A)

适用专业： 软件工程 考试日期： _____ 时间： 120 分钟 共 4 页

题 号	一	二	三	四	五	总 分
得 分						

一、选择题（本大题共 15 道小题，每小题 2 分，共 30 分）

题	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
答															

- 1、以下（）项不属于 Hadoop 可以运行的模式。
- A. 单机（本地）模式
- B. 互联模式
- C. 伪分布式模式
- D. 分布式模式
- 2、在 Hadoop2. x 版本中，HDFS 默认 Block Size 的大小是（）。
- A. 32MB
- B. 64MB
- C. 128MB
- D. 256M
- 3、下列关于 MapReduce 说法不正确的是（）。
- A. MapReduce 是一种计算框架
- B. MapReduce 隐藏了并行计算的细节，方便使用
- C. MapReduce 来源于 google 的学术论文
- D. MapReduce 程序只能用 java 语言编写
- 4、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（）。
- 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

- 5、HBase 依靠（）存储底层数据。
- A. HDFS
- B. Hadoop
- C. Memory
- D. MapReduce
- 6、Hive 依赖（）提供强大的计算能力。
- A. Zookeeper
- B. Chubby
- C. RPC
- D. MapReduce
- 7、关于 SecondaryNameNode 下面哪项是正确的（）。
- 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
- 8、配置 Hadoop 时，JAVA_HOME 包含在哪一个配置文件中（）。
- A. hadoop-default.xml
- B. hadoop-env.sh
- C. hadoop-site.xml
- D. configuration.xsl
- 9、关于 Hadoop 单机模式和伪分布式模式的说法，正确的是（）。
- A. 两者都起守护进程，且守护进程运行在一台机器上
- B. 单机模式不使用 HDFS，但加载守护进程
- C. 两者都不与守护进程交互，避免复杂性
- D. 后者比前者增加了 HDFS 输入输出以及可检查内存使用情况
- 10、HDFS 的 NameNode 负责管理文件系统的命名空间，将所有的文件和文件夹的元数据保存在一个文件系统树中，这些信息也会在硬盘上保存成以下文件：
- A. Edits 文件
- B. FSImage 文件
- C. A 和 B 两者都是

D. 以上都不对

11、HDFS 无法高效存储大量小文件，想让它能处理好小文件，比较可行的改进策略不包括（）。

- A. 利用 SequenceFile、MapFile、Har 等方式归档小文件
- B. 多 Master 设计
- C. Block 大小适当调小
- D. 调大 namenode 内存或将文件系统元数据存到硬盘里

12、以下四个 Hadoop 预定义的 Mapper 实现类的描述错误的是（）。

- A. IdentityMapper<K, V>实现 Mapper<K, V, K, V>，将输入直接映射到输出
- B. InverseMapper<K, V>实现 Mapper<K, V, K, V>，反转键/值对
- C. RegexMapper<K>实现 Mapper<K, Text, Text, LongWritable>，为每个常规表达式的匹配项生成一个 (match, 1)对。
- D. TokenCountMapper<K>实现 Mapper<K, Text, Text, LongWritable>，当输入的值为分词时，生成 (token, 1)对。

13、Namenode 在启动时自动进入安全模式，在安全模式阶段，说法错误的是（）。

- A. 安全模式目的是在系统启动时检查各个 DataNode 上数据块的有效性。
- B. 根据策略对数据块进行必要的复制或删除。
- C. 当数据块最小百分比数满足的最小副本数条件时，会自动退出安全模式。
- D. 文件系统允许有修改。

14、关于 HDFS 的文件写入，正确的是（）。

- A. 支持多用户对同一文件的写操作。
- B. 用户可以在文件任意位置进行修改。
- C. 默认将文件块复制成三份存放。
- D. 复制的文件块默认都存在同一机架上。

15、下面哪个程序负责 HDFS 数据存储（）。

- A. NameNode
- B. Jobtracker
- C. Datanode
- D. SecondaryNameNode

得分	评卷人

二、判断题（本大题共10道小题，每题2分，共20分）

- 1、Block Size 是不可以修改的。（ ）
- 2、如果 NameNode 意外终止，SecondaryNameNode 会接替它使集群继续工作。（ ）
- 3、Hadoop 是 Java 开发的，所以 MapReduce 只支持 Java 语言编写。（ ）
- 4、因为 HDFS 有多个副本，所以 NameNode 是不存在单点问题的。（ ）
- 5、NameNode 本地磁盘保存了 Block 的位置信息。（ ）
- 6、HDFS 的设计灵感是来自于谷歌发表的论文 GFS。（ ）
- 7、MapReduce 是 Hadoop 专属的，其他框架没有。（ ）
- 8、HBase 的设计是基于雅虎发表的论文 BigTable。（ ）
- 9、ZooKeeper 集群是主从式结构的集群，所以不存在单点故障问题。（ ）
- 10、使用类 SQL 语句的 HQL 操作 Hive，可以进行 CRUD 操作。（ ）

得分	评卷人

三、简答题（本大题共 4 道小题，每题 5 分，共 20 分）

1. 详细描述 HDFS 集群的读写过程。

2. 简述 Hadoop 1. x 到 Hadoop2. x 发生了哪些变化。

3. 详述 MapReduce 程序的运行原理。

4. 详述实现二次排序的过程。

得分	评卷人

四、阅读程序然后按要求填空（本大题共 2 道小题，每题 5 分，共 10 分）

1、请在下面程序的下划线中补充完整程序（共10处）。

```
public class WordCount {  
    publicstatic class TokenizerMapper extends  
        Mapper<_____, _____, _____, _____> {  
        private final static IntWritable one = newIntWritable(1);  
        private Text word = new Text();  
        public void map(LongWritable key, Text value,Context context){  
            StringTokenizeritr = new StringTokenizer(value.toString());  
            while (itr.hasMoreTokens()) {  
                word.set(itr.nextToken());
```

```
                context.write(word, one);  
            }  
        }  
    }  
    public static class IntSumReducer extends  
        Reducer<_____, _____, Text, IntWritable> {  
        private IntWritable result = newIntWritable();  
        public void reduce(_____ key, Iterable<_____> values, Context context) {  
            _____ sum = 0;  
            for (IntWritable val : values) {  
                sum += val.get();  
            }  
            result._____(sum);  
            context.write(key, result);  
        }  
    }  
    public static void main(String[] args) throws Exception {  
        略……  
    }  
}
```

2. 宾馆里有100个房间，从1-100进行编号，第一个服务员将所有的房间门都打开，第二个服务员把所有编号是2的倍数的房间“相反处理”，第三个服务员将所有编号是3的倍数的房间再作“相反处理”……，以后每个服务员都是如此操作，当第100个服务员来过，后，请编程计算哪几个房间的门是打开的？（所谓“相反处理”是指原来开着的门关上，原来关上的门打开）请将程序填写完整（共5处）。

```
public class HotelDoor {  
    public static void main(String[] args){  
        boolean[] a=new _____[101];  
        final int N=101;  
        int i,j;  
        for(i=1;i<N;i++)_____; //第1个服务员将所有房间设置为打开状态  
        for(i=2;i<N;i++)  
            for(_____; j<N; j++)  
                if(j%i==0)_____; //执行相反处理
```

```
for(i=1;i<N;i++)
    if(a[i]==true) System._____.print(i+" "); //显示打开状态的房间编号
}}
```

得分	评卷人

五．程序设计题（本大题共3道小题，第一题5分，第二题5分，第三题10分，共20分）

1、按以下要求编写程序

有一个样本数据如下，请使用你最熟悉的语言编写 MapReduce 程序，该样本数据中每个元素出现的次数（注

意：如若篇幅有限，则不需要写出作业配置相关的代码）。

样本数据如下：a ， b ， c ， d

b ， b ， f ， e

a ， a ， c ， f

2、按以下要求编写程序

对天气数据进行二次排序，先按照年份进行排序，再按照气温进行排序，结果是三列数据（year，temp，sid），请写出除 Map 任务和 Reduce 任务之外所必需分区器类，分组比较器类和排序比较器类（假设复合键类已经定义，自然键是 year，自然值是 temp）。

3、按以下要求编写程序

有两个文本文件，文件中的数据按行存放，请编写 MapReduce 程序，找到两个文件中彼此不相同的行。