

第2章 大数据处理架构Hadoop

1. 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

2. Hadoop的作者是下面哪一位__B__。

A. Martin Fowler

B. Doug cutting

C. Kent Beck

D. Grace Hopper

3. 下列哪个程序通常与 NameNode 在同一个节点启动__D__。

A. TaskTracker

B. DataNode

C. SecondaryNameNode

D. Jobtracker

4. 下列哪项通常是集群的最主要瓶颈__C__。

A. CPU

B. 网络

C. 磁盘IO

D. 内存

肃5、Doug Cutting所创立的项目的名称都受到其家人的启发，以下项目不是由他创立的项目是

宸A. Hadoop

薄B. Nutch

蒂C. Lucene

奠D. Solr

腿答案：D

薄6、配置Hadoop时，JAVA_HOME包含在哪一个配置文件中

袈A. hadoop-default.xml

芫B. hadoop-env.sh

袈C. hadoop-site.xml

羈D. configuration.xml

芫答案：B

7 Hadoop配置文件中，hadoop-site.xml显示覆盖hadoop-default.xml里的内容。在版本0.20中，hadoop-site.xml被分离成三个XML文件，不包括

肇A. conf-site.xml

舫B. mapred-site.xml

蒂C. core-site.xml

蚤D. hdfs-site.xml

膈答案：A

8、关于Hadoop单机模式和伪分布式模式的说法，正确的是

申A. 两者都起守护进程，且守护进程运行在一台机器上

芳B. 单机模式不使用HDFS，但加载守护进程

蒯C. 两者都不与守护进程交互，避免复杂性

繇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

螂答案：D

999、下列关于Hadoop API的说法错误的是

- 荏A. Hadoop的文件API不是通用的，只用于HDFS文件系统
- 芳B. Configuration类的默认实例化方法是以HDFS系统的资源配置为基础的
- 螭C. FileStatus对象存储文件和目录的元数据

薇答案：A

菖10、Hadoop fs中的-get和-put命令操作对象是

- 熊A. 文件
- 袂B. 目录
- 置C. 两者都是

虬答案：C

11、以下说法不正确的是

- 蒯A. Hadoop Streaming使用Unix中的流与程序交互
- 菟B. Hadoop Streaming允许我们使用任何可执行脚本语言处理数据流
- 腰C. 采用脚本语言时必须遵从UNIX的标准输入STDIN，并输出到STDOUT
- 螂D. Reduce没有设定，上述命令运行会出现问题

蒂答案：D

螭12. Hadoop 作者 答案C Doug cutting

蒿a)Martin Fowler

杓b)Kent Beck

芳c)Doug cutting

蛭13. 下列哪项可以作为集群的管理？ 答案：ABD

薹a)Puppet

蛭b)Pdsh

置c)Cloudera Manager

蜎d)Zookeeper

羈14. 配置机架感知的下面哪项正确:

殓A如果一个机架出问题, 不会影响数据读写

緡B写入数据的时候会写到不同机架的 DataNode 中

CCC MapReduce 会根据机架获取离自己比较近的网络数据

答案ABC

寇15. 下列哪个是 Hadoop 运行的模式: 答案ABC

节a)单机版

殓b)伪分布式

羈c)分布式

二、判断题

1. Ganglia 不仅可以进行监控, 也可以进行告警。(正确)

蒞2. Block Size 是不可以修改的。(错误)

羈3. Nagios 不可以监控 Hadoop 集群, 因为它不提供 Hadoop 支持。(错误)

瞽4. 如果 NameNode 意外终止, SecondaryNameNode 会接替它使集群继续工作。(错误)

萁5. Cloudera CDH 是需要付费使用的。(错误)

逯6. Hadoop 是 Java 开发的, 所以 MapReduce 只支持 Java 语言编写。(错误)

肇7. Hadoop 支持数据的随机读写。(错)

8、Hadoop 自身具有严格的权限管理和安全措施保障集群正常运行。(错误)

9、hadoop dfsadmin -report 命令用于检测 HDFS 损坏块。(错误)

瞽10. Hadoop 默认调度器策略为 FIFO (正确)

膝11. 集群内每个节点都应该配 RAID, 这样避免单磁盘损坏, 影响整个节点运行。(错误)

12. Hadoop 环境变量中的 HADOOP_HEAPSIZE 用于设置所有 Hadoop 守护线程的内存。它默认是 200 GB。(错误)