

## “1+X”大数据平台运维职业技能等级证书（初级） - 机试练习题库

### 一、单选题 30 道

#### 大数据平台安装 5 道

- 1、从 Vi 编辑模式下保存配置返回到 Centos7 系统使用（ ）  
A.按 “shift”+“:” 组合  
B.按 “Esc” 键，然后按 “shift” + “:” 组合  
C.按 “Esc”键，然后按 “shift”+“:”，输入 wq 回车  
D.直接输入 “exit”
- 2、在 Centos7 系统中，查看和修改 IP 信息需要用到（ ）文件  
A. /etc/sysconfig/network/ifcfg-ens0  
B. /etc/sysconfig/network-script/ifcfg-ens0  
C. /etc/sysconfig /ifcfg-ens0  
D. /etc/resolv.conf
- 3、tar 命令用于对文件进行打包压缩或解压，-t 参数含义（ ）  
A. 创建压缩文件  
B. 解开压缩文件  
C. 查看压缩包内有哪些文件  
D.向压缩归档末尾追加文件
- 4、Hadoop 的安装部署的模式属于本地模式（ ）  
A. 默认的模式，无需运行任何守护进程（daemon），所有程序都在单个 JVM 上执行。  
B. 在一台主机模拟多主机。即，Hadoop 的守护程序在本地计算机上运行，模拟集群环境，并且是相互独立的 Java 进程。  
C. 完全分布模式的守护进程运行在由多台主机搭建的集群上，是真正的生产环境。  
D.高容错全分布模式的守护进程运行在多台主机搭建的集群上
- 5、Hadoop 的安装部署的模式属于伪分布模式（ ）  
A. 默认的模式，无需运行任何守护进程（daemon），所有程序都在单个 JVM 上执行。  
B. 在一台主机模拟多主机。即，Hadoop 的守护程序在本地计算机上运行，模拟集群环境，并且是相互独立的 Java 进程。  
C. 完全分布模式的守护进程运行在由多台主机搭建的集群上，是真正的生产环境。  
D.高容错全分布模式的守护进程运行在多台主机搭建的集群上

#### 大数据平台配置 5 道

- 6、HDFS 默认 Block Size 的大小是（ ）。  
A.32MB B.64MB C.128MB D.256M
- 7、下列关于 MapReduce 说法不正确的是( )  
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
- 8、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

9、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

10、大数据的特点不包括下面哪一项（ ）。

- 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

#### 大数据平台组件安装配置 5 道

11、HBase 依赖\_( )存储底层数据

- A. HDFS B. Hadoop C. Memory D. MapReduce

12、HBase 依赖（ ）提供消息通信机制

- A Zookeeper
- B Chubby
- C RPC
- D Socket

13、在 hive 中已知表 t

est(name)的记录如下，

tom

tom\_green

tomly

lily

代码 select \* from test where name rlike 'tom.\*'的结果有（ ）条记录

- A 1
- B 2
- C 3
- D 0

14、flume 有哪个组件用于采集数据（ ）

- A. source
- B. channel
- C. sink
- D. web server

15、flume 中那种 source 类型支持 Unix 的命令标准在生产数据（ ）

- A. Avro Source
- B. Thrift Source
- C. Exec Source
- D. JMS Source

#### 大数据平台实施 4 道

16、大数据正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析，从中发现新知识、创造新价值、提升新能力的（ ）

- A. 新一代技术平台
- B. 新一代信息技术和服务业态
- C. 新一代服务业态
- D. 新一代信息技术

17、整个大数据平台按其职能划分为五个模块层次，运行环境层是（ ）

- A. 运行环境层为基础设施层提供运行时环境，它由 2 部分构成，即操作系统和运行时环境。
- B. 基础设施层由 2 部分组成：Zookeeper 集群和 Hadoop 集群。它为基础平台层提供基础设施服务
- C. 基础平台层由 3 部分组成：任务调度控制台、HBase 和 Hive。它为用户网关层提供基础服务调用接口。
- D. 用户网关层用于为终端客户提供个性化的调用接口以及用户的身份认证，是用户唯一可见的大数据平台操作入口。

18、大数据平台架构设计( )的思想

- A. 分层模块化设计
- B. 高集合化设计
- C. 统一结构化设计
- D. 全分布设计

19、大数据系统出现“java.net.NoRouteToHostException: No route to host ”解决方案

( )

- A. zookeeper 的 IP 要配对
- B. 关掉防火墙 service iptables stop
- C. master 和 slave 配置成同一个 IP 导致的，要配成不同 IP
- D. SSH 配置错误导致，主机名一定要严格匹配，重新配置 ssh 免密码登录

#### 大数据平台监控 5 道

20、hadoop 平台下检查 h3cu 目录的健康状态采用（ ）命令

- A. hdfs fsck /h3cu
- B. find /h3cu
- C. list /h3cu
- D. hadoop fs -lsr /h3cu

21、查看当前 yarn 运行任务列表采用（ ）命令

- A. yarn application -log
- B. HDFS application -list

C. yarn application -kill

D. yarn application -list

22、报告 hdfs 使用情况采用（ ）命令

A. hdfs dfsadmin -setSpaceQuota 500g /path/xxx/

B. hdfs dfsadmin -report

C. hdfs dfsadmin -clrSpaceQuota /user/testname

D. hdfs dfs -getfacl /user/testname

23、map/reduce 参数监控 metrics，shuffle 成功的次数（ ）

A. mapred.ShuffleMetrics.ShuffleConnections

B. mapred.ShuffleMetrics.ShuffleOutputBytes

C. mapred.ShuffleMetrics.ShuffleOutputsFailed

D. mapred.ShuffleMetrics.ShuffleOutputsOK

24、启动 Hadoop 和 HBase 之后，执行 jps 命令，已经看到有 HMaster 的进程，但是进入到监控界面查看，没有节点信息：

Backup Master `hadoop1`

Current Active Master: `hadoop1`

Tasks

Show All Monitored Tasks Show non-RPC Tasks Show All RPC Handler Tasks Show Active RPC Calls Show Client Operations

View as JSON

Start Time	Description	State	Status
Mon Dec 30 15:01:23 CST 2019	Master startup	RUNNING (since 6mins, 34sec ago)	Initializing Master file system (since 6mins, 34sec ago)

Software Attributes

Attribute Name	Value	Description
HBase Version	2.0.0, revision=7483b11e4da77adbfc8062b3b22cbe7c2cb91c1	HBase version and revision

最有可能原因（ ）

A. HADOOP 服务未启动

B. HBase 服务未启动

C. HBase 处于安全模式

D. HBase 服务未安装

大数据平台综合 6 道

25、HBase 分布式模式最好需要（ ）个节点

A 1 B 2 C 3 D 4

26、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

27、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
- 28、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
- 29、关于 SecondaryNameNode 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. 它是 NameNode 的热备,对内存有严格要求
- 30、关于 Hadoop 的运行机制过程顺序说法正确的是（ ）
- A. 作业提交--作业初始化--任务分配--任务的执行--任务进度和状态的更新--任务结束
  - B. 作业初始化--作业提交--任务分配--任务的执行--任务进度和状态的更新--任务结束
  - C. 任务分配--作业提交--作业初始化--任务的执行--任务进度和状态的更新--任务结束
  - D. 任务执行-作业提交--作业初始化--任务的分配--任务进度和状态的更新--任务结束

## 二、不定项题 10 道

- 1、下面描述正确的是( )
- A. HDFS 将数量众多的廉价机器使得硬件错误成为常态
  - B. HDFS 支持多种软硬件平台中的可移植性
  - C. HDFS 上的一个文件大小是 G 或 T 数量级的，支持的文件数量达到千万数量级
  - D. HDFS 上的文件可以任意修改
- 2、YARN 服务组件包括( )
- A. NameManager
  - B. ApplicationMaster
  - C. Container
  - D. ResourceManager
- 3、对 NodeManager 服务描述正确的是( )
- A. 整个集群有多个 NodeManager，NodeManager 负责单节点资源管理和使用
  - B. NodeManager 管理当前节点的 Container 资源抽象
  - C. 通过心跳机制定时地向 ResourceManager 汇报本节点上的资源使用情况
  - D. NodeManager 服务器跟 ResourceManager 毫无关系
- 4、下列属于 hadoop 的发行版本的是（ ）
- A. Apache
  - B. Cloudera
  - C. Redhat
  - D. Centos
- 5、下面的命令描述正确的是( )
- A. 创建一个名称为 test 组的命令：groupadd test
  - B. 改用户 Tom 设置密码命令：passwd Tom

- C. 删除用户名为 Tom 的命令：userdel Tom  
D. 切换当前用户帐户为 admin 的命令：su admin

6、MapReduce 不合适对哪些场景的使用（ ）

- A. MapReduce 不适合做低延迟数据访问场景的使用  
B. MapReduce 不适合存储大量小文件  
C. MapReduce 不支持多用户写入及任意修改文件  
D. MapReduce 不支持大文件存储

7、Hive 有哪些方式保存元数据（ ）

- A. single user mode  
B. multiuser mode  
C. remote user mode  
D. telnet user mode

8、HDFS 保障可靠性的措施（ ）

- A. 数据冗余机制  
B. 数据节点心跳包  
C. 数据节点块报告  
D. 数据完整性检测

9、根据 Grafana 仪表盘展现内容，下列描述正确的是（ ）



- A. Hdfs 健康状态资源使用率过高  
B. 集群磁盘读写失败节点有坏节点  
C. 存储空间占用比例最高的是 hb\_xdata-datanode 节点

D. 集群中 `lf-cdn` 内存使用最少

10、以下是 HTools 工具特点 ( )

- A. 友善的向导式操作流程
- B. 无需安装 HTools 客户端
- C. 支持同时管理多个 Hadoop 集群和节点
- D. 支持系统配置文件的推送和同步