**2020年上海市高等学校信息技术水平考试试卷**

**四级 大数据与云计算（A 场）**

（本试卷考试时间 150 分钟）

**一、单选题 ( 本大题 18 道小题 ，每小题 1 分，共 18 分），从下面题目给出的A、B、C、D四个可供选择的答案中选择一个正确答案。**

1.Sqoop底层运行的任务是\_\_\_\_。

A.map task和reduce task

B.map task

C.reduce task

D.sink task

2.设森林F对应的二叉树为B，它有m个结点，B的根为p，p的右子树结点个数为n,森林F中第一棵子树的结点个数是\_\_\_\_。

A.m-n

B.m-n-1

C.n+1

D.条件不足，无法确定

3.Docker容器中解决资源（CPU、内存）可配额的技术是\_\_\_\_。

A.cgroups

B.aufs

C.kernel namespace

D.device mapper

4.已知表emp存在字段name1,以下语句能够正常运行的是\_\_\_\_。

A.alter table emp rename to emp1

B.alter table emp rename emp to emp1

C.alter table emp change column name1 to name2

D.alter table emp change name name12

5.如果用含有隐变量的概率模型来建模问题，那么该问题的最有可能的解法是\_\_\_\_。

A.EM算法

B.Logistic回归

C.协同过滤

D.高斯过滤

6.对于分布式缓存数据库Redis，说法错误的是\_\_\_\_。

A.Redis支持多线程处理

B.Redis可以持久化数据

C.Redis支持的数据类型包括String、List、Set

D.Redis集群之间是异步复制的

7.虚拟化的主要特征不包括\_\_\_\_。

A.实现简单

B.高扩展性

C.高实用性

D.高安全性

8.与云计算相比，\_\_\_\_不是云计算的特征。

A.适合紧耦合科学计算

B.资源高度共享

C.支持虚拟机

D.适合商业领域应用

9.下列说法中有误的是\_\_\_\_。

A.虚拟化产品不算云计算平台的核心，但其选型正确与否非常关键。

B.虚拟化平台一定要提供完善的高可用、安全性、扩展性与自动化管理等功能。

C.虚拟化平台具备容量、性能、配置、安全、可靠性等内置在虚拟化内部的自动化解决方案。

D.虚拟化所实现的计算资源池化、动态调配、自动化管理、高可用等功能是云计算平台所依赖的必不可少的基础特征。

10.对象存储是为互联网应用提供非结构化文件云存储的服务， 不适合\_\_\_\_的场景。

A.关系型数据库存储

B.大量日志文件存储

C.直播录制存储

D.游戏APK存储

11.数据库表的两种相关约束是\_\_\_\_。

A.主键和外键

B.外键和索引

C.主键和索引

D.内键和外键

12.Linux返还上一级目录使用的命令是\_\_\_\_。

A.cd ..

B.cd

C.cd .

D.cd ...

13.\_\_\_\_适合适用MapReduce处理的。

A.离线计算

B.迭代计算

C.实时交互计算

D.流式计算

14.MongoDB数据库中基本单位为\_\_\_\_。

A.doc

B.table

C.set

D.elevation

15.分布式文件系统存储三种类型的数据：定长块、大文件和\_\_\_\_。

A.blob对象

B.block对象

C.circle对象

D.obstacle对象

16.当数据写入时产生的最大的一个StoreFile大小超过一定阈值，region发生split。

A.正确

B.错误

C.

D.

17.Docker能够以直接在容器（Container）中修改内容的方式来制作镜像。

A.正确

B.错误

C.

D.

18.HDFS支持两个用户同时写入。

A.正确

B.错误

C.

D.

**二、多选题 ( 本大题 3 道小题 ，每小题 .5 分，共 1.5 分），从下面题目给出的A、B、C、D四个可供选择的答案中选择所有正确答案。**

1.在OpenStack中以下\_\_\_\_服务是在控制节点运行的。

A.KeyStone

B.Nova

C.Glance

D.Neutron

2.云计算包含\_\_\_\_技术。

A.并行计算

B.资源虚拟化

C.嵌入式

D.分布式计算

3.尽管人们对云计算寄予很大期望，云计算在实际发展中还有许多障碍，主要是缺乏统一的技术标准和运营标准、缺乏相应的法律法规保障和\_\_\_\_。

A.进行必要的数据交换和同步

B.软件许可证问题

C.云服务平台故障

D.数据安全性

**三、填空题 ( 本大题 3 道小题 ，每空 1 分，共 4 分）。**

1.\_\_\_\_\_\_和\_\_\_\_\_\_可以用来确定K-Means算法中的k值。

2.Redis回收使用的是\_\_\_\_\_算法。

3.OpenStack计算节点上虚拟机的默认路径是 \_\_\_\_\_\_。

**四、操作题**

**素材、样张、KS目录均在zip文件中，可双击此图标打开**

**以下第（一）题（简答题）、第（二）题（综合题）题目请在文件"C:\KS\大数据与云计算-答题纸.docx"中作答！**

**请在答题纸作答！此处答题一律无效！**

**（一）、简答题（本大题包括5道小题，每小题4分，共20分）**

1、 请简述云计算中有哪些不同的层？

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！  |

2、 Spark中常用的数据类型是哪三种？三者之间的区别是什么？

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！  |

3、 请基于下述数据，使用K-means算法完成以下要求。

a. 实现肘部法则，画出肘部法则曲线；

b. 设置k=2，并调用库函数并进行模型训练；

c. 输出聚类簇中心的坐标、各点所属簇中心的位置及代价；

d. 画出样本数据及聚类中心的位置。

数据如下：

[[2,6],[4,7],[3,2],[6,5],[7,3],[8,5],[2,4],[3,2],[5,8],[6,10],[12,17],[10,12],[15,20],[16,13],[11,16],[10,15],[19,12],[17,15],[16,12],[13,20]]

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！  |

**（二）、综合题（20分）**

某电影公司希望根据近5年已上映电影的票房收入，建立模型预测电影的票房收益。

1. 在不考虑数据来源的情况下，你认为有哪些因素会影响电影票房，请至少给出5项。

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！  |

2. 在建模之前你认为需要做哪些数据预处理工作，请至少给出3项。

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！  |

3. 你认为应该采用什么算法建模，并简述该算法的实现过程。

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！  |

4. 某部新电影上映之后票房不理想，公司希望根据某影评网站的用户评论分析票房失败的原因，请用正则表达式筛选出包含"不满意""烂片""不好"等文字的评论

|  |
| --- |
| 请在答题纸作答！此处答题一律无效！ |