第一章 大数据概述

1【单选题】

人类社会的数据产生方式大致经历了三个阶段，不包括:

A、运营式系统阶段

B、用户原创内容阶段

C、互联网应用阶段

D、感知式系统阶段

答案：C

数据产生方式经历了三个阶段：运营式系统阶段、用户原创内容阶段、感知式系统阶段

2【单选题】

以下哪个现象不属于大数据的典型特征：

A、数据包含噪声及缺失值

B、数据量大

C、数据类型多

D、产生速率高

答案：A

大数据的特征:数据量大、数据类型多、处理速度快、价值密度低

3【单选题】

以下哪项不属于大数据思维的涵盖内容：

A、从模型驱动到数据驱动

B、通过采样的手段获取目标群体的统计特性

C、数据就是生产资料

D、全样本分析

答案：B

4【单选题】

在思维方式方面，不是大数据显著影响的是\_\_\_\_\_\_\_\_。

A、全样而非抽样

B、效率而非精确

C、存储而非计算

D、相关而非因果

答案：C

5【单选题】

下列属于批处理计算的是\_\_\_\_\_\_\_\_。

A、Storm

B、Hive 属于查询分析计算

C、Sqoop

D、Spark

答案：D

批处理计算的代表为MapReduce、Spark

6【单选题】

云计算平台层 （PaaS） 指的是什么？

A、操作系统和围绕特定应用的必需的服务

B、将基础设施(计算资源和存储)作为服务出租

C、从一个集中的系统部署软件，使之在一台本地计算机上(或从云中远程地)运行的一个模型

D、提供 硬件、软件、网络等基础设施 以及提供咨询、规划和系统集成服务

答案：D

7【单选题】

就数据的量级而言，1EB=2(20) TB。

8【单选题】

对于数据的数量级来说，1PB= 2(20) GB;

数量级从小到大排列B、KB、MB、GB、TB、PB、EB、ZB 每两个中间都差2的十次方倍（1024）

答案：C

9【多选题】

数据产生方式大致经历了三个阶段，包括\_\_\_\_\_\_\_\_。

A、移动互联网数据阶段

B、运营式系统阶段

C、用户原创内容阶段

D、感知式系统阶段

答案：BCD

10【多选题】

人类社会的数据产生方式大致经历了三个阶段，不包括:

A、运营式系统阶段

B、移动互联网时代

C、感知式系统阶段

D、互联网应用阶段

答案：BD

11【多选题】

大数据的特征包含\_\_\_\_\_\_\_\_。

A、数据量大

B、数据类型繁多

C、处理速度快

D、价值密度低

答案：ABCD

12【多选题】

图灵奖获得者、著名数据库专家Jim Gray博士认为，人类自古以来在科学研究上先后经历了四种范式，下列哪些属于 \_\_\_\_\_\_\_\_这四种范式。

A、实验科学

B、猜想科学

C、理论科学

D、计算科学

答案：ACD

科学研究上先后经历了四种范式分别是:实验科学、理论科学、计算科学、数据密集型科学

13【多选题】

大数据对思维方式的影响包括\_\_\_\_\_\_\_\_。

A、全样而非抽样

B、效率而非精确

C、实践而非理论

D、相关而非因果

答案：ABD

大数据对思维方式的影响:全样而非抽样、效率而非精确、相关而非因果

14【多选题】

大数据的计算模式包括\_\_\_\_\_\_。

A、批处理计算

B、图计算

C、流计算

D、查询分析计算

答案：ABCD

15【多选题】

云计算的典型服务模式包括\_\_\_\_\_\_\_\_。

A、平台即服务

B、物联网即服务

C、基础设施即服务

D、软件即服务

答案：ACD

16【判断题】

物联网与云计算、大数据是相辅相成的关系，物联网就是指无线传感器。

答案：×

17【判断题】

MapReduce是分布式并行计算框架，其计算模式属于流计算，实时性好。

答案：×

MapReduce 属于批处理计算

第二章 大数据处理架构Hadoop

Hadoop 的特性：

高可靠性

高效性

高可扩展性

高容错性

成本低

运行在Linux上

支持多种程序语言

1【单选题】

启动hadoop所有 进程的命令是\_\_\_\_\_\_\_\_

A、start-all.sh

B、start-hdfs.sh

C、start-hadoop.sh

D、start-dfs.sh

答案：A

2【单选题】

以下对Hadoop的说法 错误 的是\_\_\_\_\_\_\_\_

A、Hadoop MapReduce是针对谷歌MapReduce的开源实现，通常用于大规模数据集的并行计算

B、Hadoop2.0增加了NameNode HA和Wire-compatibility两个重大特性

C、Hadoop是基于Java语言开发的，只支持Java语言编程

D、Hadoop的核心是HDFS和MapReduce

答案：C

3【单选题】

以下哪个 不是 hadoop的特性\_\_\_\_\_\_\_\_。

A、成本高

B、高可靠性

C、支持多种编程语言

D、高容错性

答案：A

4【单选题】

以下名词解释不正确的是\_\_\_\_\_\_\_\_。

A、HDFS：分布式文件系统，是Hadoop项目的两大核心之一，是谷歌GFS的开源实现

B、Hive：一个基于Hadoop的数据仓库工具，用于对Hadoop文件中的数据集进行数据整理、特殊查询和分析存储

C、HBase：提供高可靠性、高性能、分布式的行式数据库，是谷歌BigTable的开源实现（列式）

D、Zookeeper：针对谷歌Chubby的一个开源实现，是高效可靠的协同工作系统

答案：C

5【单选题】

Hadoop是\_\_\_\_\_\_\_\_公司旗下的分布式计算平台。

A、Oracle

B、Google

C、Apache

D、Amazon

答案：C

6【单选题】

Hadoop项目结构中，\_\_\_\_\_\_\_\_负责资源管理和调度。

A、YARN

B、HA

C、Hive

D、Storm

答案：A

7【单选题】

下列选项中哪一门技术属于大数据平台\_\_\_\_\_\_\_\_。

A、Tomcat

B、Hadoop

C、ASP.NET

D、Apache

答案：B

8【单选题】

下列不属于Hadoop生态的技术是\_\_\_\_\_\_\_\_。

A、Hive

B、HDFS

C、HBase

D、SQL Server

答案：D

9【单选题】

Hadoop运行在\_\_\_\_\_\_\_\_操作系统之上

A、Windows

B、Linux

C、Unix

D、IOS

答案：B

10【多选题】

以下哪些组件是Hadoop的生态系统的组件\_\_\_\_\_\_\_\_。

A、MapReduce

B、Oracle

C、HBase

D、HDFS

答案：ACD

11【多选题】

Hadoop的核心组件是\_\_\_\_\_\_\_\_和\_\_\_\_\_\_\_\_。

A、GFS

B、HDFS

C、Spark

D、MapReduce

答案：BD

12【多选题】

Hadoop生态系统的优势包含\_\_\_\_\_\_\_\_。

A、高扩展

B、低成本

C、开源工具成熟

D、大型关系数据库系统

答案：ABC

13【多选题】下列\_\_\_\_\_\_\_\_ 不属于Hadoop生态系统的优势。

A、高扩展

B、低成本

C、低容错性

D、大型关系数据库系统

答案：CD

14【多选题】

大数据技术方案为了简化并行分布式计算，采用\_\_\_\_\_\_\_\_软件模块进行处理。

A、Java

B、Map

C、Reduce

D、SQL

答案：BC

15【多选题】

以下\_\_\_\_\_\_\_\_工具属于Hadoop生态系统的开源工具。

A、Hive

B、HBase

C、Mysql

D、Zookeeper

答案：ABD

16【判断题】

Hadoop是IBM公司开发的一款商用大数据软件。

答案：×

17【判断题】

Hadoop是基于Java语言开发的，具有很好的跨平台特性。

答案：√

18【判断题】

Hadoop是跨平台的，安装Hadoop时没必要安装JDK。

答案：×