第一次随堂测验

1.当前大数据技术的基础是由（ B ）首先提出的。

A. 微软

B. 谷歌

C. 百度

D. 阿里巴巴

2.以下关于大数据的4V特征说法错误的是（ C ）

A. 数据规模大

B. 数据类型多样

C. 数据处理速度慢

D. 数据价值密度低

3.以下关于大数据计算模式的说法错误的是（ D ）

A．常见的四种大数据计算模式分别是批处理计算、流计算、图计算

B．查询分析计算主要解决大规模数据的存储管理和查询分析问题

C．图计算针对图结构数据处理，代表产品有Graphx、Giraph等

D．批处理计算解决流数据的实时计算问题

4.大数据的技术框架自底向上为（ A ）

A.数据收集层、数据存储层、资源管理与服务协调层、计算引擎层、数据分析层、数据可视化层

B.数据收集层、数据存储层、计算引擎层、资源管理与服务协调层、数据分析层、数据可视化层

C.数据收集层、资源管理与服务协调层、数据存储层、计算引擎层、数据分析层、数据可视化层

D.数据收集层、数据存储层、资源管理与服务协调层、数据分析层、计算引擎层、数据可视化层

5.以下关于Hadoop与Spark开源大数据技术栈说法错误的是（ C ）

A．其数据收集层主要由关系型与非关系型数据收集组件，分布式消息队列构成。

B．HDFS是Hadoop分布式文件系统，具有良好的扩展性与容错性等优点，降低了大数据存储成本。

C．MapReduce是一个批处理计算引擎，Tez是基于MapReduce的通用DAG计算引擎，但其不能高效地实现复杂的数据处理逻辑。

D．Storm/Spark Streaming是分布式流式实时计算引擎，能高效的处理流式数据，允许用户通过简单的API完成实时应用程序的开发工作。

6.以下关于Sqoop1和Sqoop2说法错误的是（ D ）

A.Sqoop1客户端仅支持命令行访问方式(CLI),Sqoop2客户端支持命令行和web两种访问方式

B.Sqoop2客户端需访问Hive、Hbase等数据源，Sqoop2通过服务端访问数据源，客户端只需发送请求

C.Sqoop1是Client-Only架构，所有软件依赖部署到客户端，Sqoop2是Client/Server架构，所有软件依赖部署到服务器端

D.Sqoop1 是一个同时包含Map和Reduce的MapReduce作业

7.以下不属于CDC系统应用的是（ C ）

A.异地机房同步

B.数据库实时备份

C.流式数据收集系统

D.业务Cache刷新

8.以下关于Otter阶段模型说法错误的是（ B ）

A.为了让系统具有良好的扩展性和灵活性，Otter 将整个同步流程抽象为 Select、Extract、Transform、Load 四个阶段

B.Select 和 Load 一般部署在原机房，而 Extract和 Transform则部署在目标机房

C.Select 阶段是与数据源对接的阶段，为解决数据来源的差异性而引入

D.Extract阶段负责数据的连接和过滤

9.下列关于Flume说法错误的是（ A ）

A.Flume将数据流水线中传递的数据称为“Event”，每个Event由字节数组(数据内容)部分构成

B.Flume屏蔽了流式数据源和后端中心化存储系统之间的异构性，使得整个数据流非常容易扩展和演化

C.Flume主要具备以下几个特点：良好的扩展性，高度定制化，声明式动态化配置，语义路由，良好的可靠性

D.Flume Agent允许用户设置其他组件更灵活地控制数据流，包括Interceptor，Channel Selector和Sink Processor等

10.下列关于Flume拓扑架构说法错误的是（ B ）

A.常见的Flume拓扑架构有两种：多路合并和多路复用

B.Flume支持将数据路由到多个目标系统中，这是通过Flume内置的多路合并功能实现的

C.Flume支持多种方式供外部数据源将流式数据发送给Flume，其中包括远程过程调用(RPC)、TCP或UDP、执行命令

D.多路复用拓扑和多路合并拓扑第一层类似，第二层有所不同

11. Flume的数据流水线是通过一系列称为{agent}的组件构成的，其中传递的数据称为{event}。注：如果是字母，全部小写。如果带符号，请用英文符号。