1. 当前大数据技术的基础是由( B)首先提出的。
2. 微软
3. 谷歌
4. 百度
5. 阿里巴巴
6. 以下关于大数据的4V特征说法错误的是（C）
7. 数据规模大
8. 数据类型多样
9. 数据处理速度慢
10. 数据价值密度低
11. 以下关于大数据计算模式的说法错误的是（D）
12. 常见的四种大数据计算模式分别是批处理计算、流计算、图计算与查询分析计算
13. 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
14. 图计算针对图结构数据处理，代表产品有Graphx、Giraph等
15. 批处理计算解决流数据的实时计算问题
16. 大数据的技术框架自底向上为（A）
17. 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
18. 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
19. 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
20. 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
21. 以下关于Hadoop与Spark开源大数据技术栈说法错误的是（C）
22. 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。
23. 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。
24. 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。
25. 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。
26. 以下关于Sqoop1和Sqoop2说法错误的是（D）
27. 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
28. 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
29. 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
30. Sqoop1的用户可以显式为作业指定Connector，避免误用
31. 以下不属于CDC系统应用的是（C）
32. 异地机房同步
33. 数据库实时备份
34. 流式数据收集系统
35. 业务Cache刷新
36. 以下关于Otter阶段模型说法错误的是（B）
37. 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
38. Select 和 Load 一般部署在原机房，而 Extract和 Transform则部署在目标机房
39. 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
40. Extract阶段负责数据的连接和过滤
41. 下列关于Flume说法错误的是（A）
42. 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
43. 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
44. 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
45. 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
46. 下列关于Flume拓扑架构说法错误的是（B）
47. 常见的Flume拓扑架构有两种：多路合并和多路复用
48. 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
49. 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
50. 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
51. 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 。