针对语义保障，以下说法错误的是：

A. Storm可以做到at least once

B. Spark Streaming可以做到exactly once

C. Flink可以做到exactly once/at least once

D. Spark Streaming可以做到exactly once/at least once

答案：D

分数：2

下列关于Zookeeper命令行操作，哪项说法错误？

A. Zookeeper集群的启动，需要在每个Zookeeper的节点上执行zkServer.sh start。

B. Zookeeper集群的停止，需要在每个Zookeeper的节点上执行zkServer.sh stop。

C. Zookeeper集群的状态查看，需要在每个Zookeeper的节点上执行zkServer.sh status。

D. Zookeeper集群的进程，是不能通过jps指令查看到。

答案：D

分数：2

下列关于Kafka集群的说法错误的是？

A. Zookeeper启动的进程名为QuorumPeerMain

B. Kafka启动的进程名为Kafka

C. Kafka集群的启动，需要在每个Kafka节点上执行bin/kafka-server-start.sh -daemon config/server.properties

D. Kafka集群的停止，需要在每个Kafka节点上执行bin/kafka-server-stop.sh

答案：D

分数：2

下列关于Kafka架构的说法错误的是？

A. Kafka的主题可以有多个分区，每个分区可以分布在不同的机器。

B. Kafka的分区可以有多个副本因子，每个副本可以分布在不同的机器。

C. Kafka的多个分区划分为一个leader和多个follower。

D. Kafka的架构是点对点的模式。

答案：D

分数：2

下列关于Storm与Hadoop的区别的说法错误的是？

A. Storm用于实时计算，Hadoop用于离线计算。

B. Storm的数据保存在内存中，源源不断；Hadoop处理的数据保存在文件系统中，一批一批处理。

C. Storm的数据通过网络传输进来；Hadoop的数据保存在磁盘中。

D. Storm与Hadoop的编程模型相似。

答案：A

分数：2

下列关于Storm Bolt的程序编写的说法错误的是？

A. 编写的Bolt类通常继承于BaseRichBolt

B. OutputFieldsDeclarer声明发射的属性只能是一个。

C. OutputCollector一次发送的字段值可以是一个或多个，需要与OutputFieldsDeclarer声明发射的属性个数相对应。

D. prepare方法通常完成初始化功能。

答案：B

分数：2

在Storm中，下面哪个是传输数据的基本单位？

A. Tuple

B. Stream

C. Component

D. Task

答案：A

分数：2

Spark是一个什么类型的计算引擎？

A. MapReduce

B. In-Memory

C. Distributed

D. All of the above

答案：D

分数：2

在Spark Streaming中，数据被切分成什么时间窗口进行处理？

A. 微批次

B. 实时流

C. 批处理

D. 分布式计算

答案：A

分数：2

Flink是一种什么类型的分布式计算系统？

A. Hadoop

B. Spark

C. Stream Processing

D. Batch Processing

答案：C

分数：2

下列计算框架中哪个延迟最高？

A. Storm

B. Spark Streaming

C. Flink

D. Hadoop

答案：D

分数：2

下列不是Flume Agent的主要组成部分的是？

A. Source

B. Channel

C. Sink

D. Cloud

答案：D

分数：2

关于bin/kafka-topics.sh --zookeeper master:2181,slave1:2181 --create --replication-factor 3 --partitions 1 --topic first的说法，哪项错误？

A. 该指令表示要创建要在Kafka中创建名称为first的主题。

B. 该指定表示使用Zookeeper连接Kafka。

C. 该指令表示要创建的分区数量为3。

D. 该指令表示使用Zookeeper连接Kafka，先使用master:2181的Zookeeper连接，当连接不成功时，再使用slave1:2181的Zookeeper连接。

答案：C

分数：2

下列哪项不是Storm的应用场景？

A. 每年双十一时，淘宝和京东实时订单销售额和产品数量大屏展示。

B. 通过工业数据分析设备运行状况和稳健程度，展示工件完成情况，运行情况。

C. 集群监控：一般的大型集群和平台，都需要对其进行监控，监控的需求。

D. 批量分析一个月的销售额数据。

答案：D

分数：2

下列关于Storm特点的说法错误的是？

A. Storm可以适用实时处理消息、更新数据库、持续计算等场景。

B. Storm无法保证所有的数据都被处理。

C. 在消息处理过程中出现异常，Storm会进行重试。

D. Storm集群非常容易管理，轮流重启节点不影响应用。

答案：B

分数：2

下列哪项是Storm Spout向下游发射数据的方法？

A. open()

B. declareOutputFields()

C. nextTuple()

D. execute()

答案：C

分数：2

Spark是一个什么类型的计算引擎？

A. MapReduce

B. In-Memory

C. Distributed

D. All of the above

答案：D

分数：2

在Spark Streaming中，哪个算子将DStream中的元素映射到新的DStream中？

A. map()

B. filter()

C. reduceByKey()

D. join()

答案：A

分数：2

Spark Structured Streaming支持什么类型的数据源？

A. 半结构化

B. 结构化

C. 非结构化

D. 半非结构化

答案：B

分数：2

下列哪个不是DataStream API中的时间概念？

A. event time

B. processing time

C. ingestion time

D. execution time

答案：D

分数：2

下列说法错误的是

A. Flume是Cloudera 提供的一个高可用的，高可靠的，分布式系统

B. Flume 基于流式架构，灵活简单

C. 海量日志采集、聚合和传输的系统

D. 业务数据库采集分析和可视化的系统

答案:D

分数:2

下列说法错误的是

A. 从延迟来说Storm可以达到毫秒级

B. 从延迟来说Spark Streaming可以达到秒级

C. 从延迟来说Flink可以达到毫秒级

D. 从延迟来说Spark Streaming可以达到秒级和毫秒级

答案:D

分数:2

关于批处理下列说法错误的是

A. 属于静态数据

B. 批大小无界

C. 访问全部数据

D. 高效、易于管理

答案:B

分数:2

关于实时处理下列说法错误的是

A. 属于动态数据

B. 数据以流的形式存在，是无界的

C. 数据在时间、窗口或者微批级别上处理

D. 实时分析系统相对于批处理较为健壮

答案:D

分数:2

在过去的几年中，下列哪一项不是大数据应用程序的发展过程

A. 仅批处理应用程序（早期的Hadoop实现）

B. 仅流处理（早期的Storm实现）

C. 仅流处理（Spark实现）

D. Lambda架构

答案:C

分数:2

下列关于Zookeeper命令行操作说法错误的是

A. Zookeeper集群的启动，需要在每个Zookeeper的节点上执行zkServer.sh start。

B. Zookeeper集群的停止，需要在每个Zookeeper的节点上执行zkServer.sh stop。

C. Zookeeper集群的状态查看，需要在每个Zookeeper的节点上执行zkServer.sh status。

D. Zookeeper集群的进程，是不能通过jps指令查看到。

答案:D

分数:2

下列关于Kafka的架构说法错误的是

A. Kafka的主题可以有多个分区，每个分区可以分布在不同的机器。

B. Kafka的分区可以有多个副本因子，每个副本可以分布在不同的机器。

C. Kafka的多个分区划分为一个leader和多个follower。

D. Kafka的架构是点对点的模式。

答案:D

分数:2

下列关于Bolt的程序编写说法错误的是

A. 编写的Bolt类通常继承于BaseRichBolt

B. OutputFieldsDeclarer声明发射的属性只能是一个。

C. OutputCollector一次发送的字段值可以是一个或多个，需要与OutputFieldsDeclarer声明发射的属性个数相对应。

D. prepare方法通常完成初始化功能。

答案:B

分数:2

Spark中的RDD指的是

A. Random Data Distribution

B. Resilient Distributed Dataset

C. Recent Data Distribution

D. None of the above

答案:B

分数:2

Spark Streaming中，数据被切分成哪个时间窗口进行处理

A. 微批次

B. 实时流

C. 批处理

D. 分布式计算

答案:A

分数:2

下列关于Flume说法错误的是

A. Flume是Cloudera 提供的一个高可用的，高可靠的，分布式系统

B. Flume 基于流式架构，灵活简单

C. 海量日志采集、聚合和传输的系统

D. 业务数据库采集分析和可视化的系统

答案:D

分数:2

下列关于语义保障说法错误的是。

A. 从语义保障来说，Storm可以做到at least once

B. 从语义保障来说，Spark Streaming可以做到exactly once

C. 从语义保障来说，Flink可以做到exactly once/at least once

D. 从语义保障来说，Spark Streaming可以做到exactly once/at least once

答案:D

分数:2

下列说法错误的是 。

A. 每月广告的活跃用户相关度分析属于批处理，时间大于1h

B. 广告印象数标签趋势属于在线处理，时间大于1s

C. 确定性的工作流程如查找推文属于批处理，时间小于500ms

D. 金融交易属于近实时处理，时效小于1ms

答案:C

分数:2

35. 关于批处理下列说法错误的是。

A. 属于静态数据

B. 批大小无界

C. 访问全部数据

D. 高效、易于管理

答案:B

分数:2

36. 关于实时处理下列说法错误的是 。

A. 属于动态数据

B. 数据以流的形式存在，是无界的

C. 数据在时间、窗口或者微批级别上处理

D. 实时分析系统相对于批处理较为健壮

答案:D

分数:2

37. 从近实时解决方案在较高层面看，不是基本分析过程的是 。

A. 流数据的生成

B. 分布式流数据的高性能计算

C. 以可查询消耗层的形式探索和可视化

D. 仪表板的形式探索和可视化

答案:A

分数:2

38. 下列哪一项不是以流的形式处理数据 。

A. 计算

B. 存储

C. 转换

D. 处理和增强

答案:B

分数:2

39. 在过去的几年中，下列哪一项不是大数据应用程序的发展过程 。

A. 仅批处理应用程序（早期的Hadoop实现）

B. 仅流处理（早期的Storm实现）

C. 仅流处理（Spark实现）

D. Lambda架构

答案:C

分数:2

40. 关于流处理和批处理下列说法错误的是 。

A. Hadoop可以构建几乎处理任何数据量的应用程序

B. Storm可扩展性强，且具有轻量级快速处理能力

C. Hadoop满足了对数据量的需求，Storm在速度方面非常出色

D. Storm在整个数据集的短时窗口上变现的很好，并且在以后的某个时间有修正数

据的机制。

答案:D

分数:2

41.下列计算框架延迟最高的是

A. Storm

B. Spark Streaming

C. Flink

D. Hadoop

答案:D

分数:2

42.下列不是Flume Agent 的主要组成部分的是

A. Source

B. Channel

C. Sink

D. Cloud

答案:D

分数:2

43.下列下列说法错误的是 。

A. 从语义保障来说，Storm可以做到at least once

B. 从语义保障来说，Spark Streaming可以做到exactly once

C. 从语义保障来说，Flink可以做到exactly once/at least once

D. 从语义保障来说，Spark Streaming可以做到exactly once/at least once

答案:D

分数:2

44.下列说法错误的是 。

A. 从延迟来说Storm可以达到毫秒级

B. 从延迟来说Spark Streaming可以达到秒级

C. 从延迟来说Flink可以达到毫秒级

D. 从延迟来说Spark Streaming可以达到秒级和毫秒级

答案:D

分数:2

45.下列关于Kafka的架构说法错误的是 。

A. Kafka的主题可以有多个分区，每个分区可以分布在不同的机器。

B. Kafka的分区可以有多个副本因子，每个副本可以分布在不同的机器。

C. Kafka的多个分区划分为一个leader和多个follower。

D. Kafka的架构是点对点的模式。

答案:D

分数:2

46. Apache Spark支持哪种语言 。

A. Java B. Python

C. Scala D. All of the above

答案:D

分数:2

47. Spark中的RDD指的是 。

A. Random Data Distribution

B. Resilient Distributed Dataset

C. Recent Data Distribution

D. None of the above

答案:B

分数:2

48. Spark Streaming支持哪些类型的数据源 。

A. Hadoop Distributed File System（HDFS）

B. Kafka

C. Flume

D. 所有以上的数据源

答案:D

分数:2

49. Spark Streaming中，数据被切分成哪个时间窗口进行处理 。

A. 微批次

B. 实时流

C. 批处理

D. 分布式计算

答案:A

分数:2

50. 下列哪个不是Flink流处理框架的核心特性 ？

A. 支持高吞吐低延迟

B. 支持有状态流处理

C. 支持Exactly-Once语义

D. 支持广播流

答案:D

分数:2