



### 每日测试

#### 【正确】 题目1:

下面对agent描述正确是? [多选题]

- A、Agent是一个JVM进程
- B、以事件的形式将数据从源头送至目的
- C、主要有3个部分组成, Source、Channel、Sink
- D、可以有多个agent连接使用

【参考答案】: ABCD

【您的答案】: ABCD

#### 【正确】 题目2:

下面source可以实现监控文件夹变化的是? [单选题]

- A、exec
- B、avro
- C、spooling directory
- D、taildir

【参考答案】: C

【您的答案】: C

#### 【正确】 题目3:

下列的sink类型中可以连接source的是? [单选题]

- A、hdfs
- B、logger
- C、file
- D、avro

【参考答案】: D

【您的答案】: D

#### 【正确】 题目4:

下列对几种channel类型描述正确的是? [多选题]

- A、Memory Channel是内存中的队列
- B、File Channel将所有事件写到磁盘
- C、kafka channel可以将数据发给kafka
- D、File Channel较Memory channel更容易导致数据丢失

【参考答案】: ABC

【您的答案】: ABC

#### 【正确】 题目5:

下列关于增加topic的分区数的描述正确的是? [单选题]

- A、可以增加, 使用命令add进行添加。

- B、可以增加，使用命令alter进行修改。
- C、不可以增加，因为是一旦创建就不能改变的。
- D、不可以增加，会导致数据丢失。

【参考答案】：B

【您的答案】：B

---

**【正确】** 题目6：

下列选项中哪个是Kafka消费者分区分配的策略？[单选题]

- A、range
- B、hash
- C、range和roundrobin
- D、roundrobin和hash

【参考答案】：C

【您的答案】：C

---

**【错误】** 题目7：

下列属于Kafka的高效读写数据原因的是？[多选题]

- A、顺序写磁盘。
- B、建立分区。
- C、零复制技术。
- D、建立索引。

【参考答案】：ABCD

【您的答案】：ACD

---

**【正确】** 题目8：

下列关于Exactly Once语义中描述错误的是？[单选题]

- A、当ack=-1，可以保证数据不会丢失。
- B、当ack=0，可以保证数据不会重复。
- C、当ack=1，即可以保证数据不会重复，也可以保证数据不会丢失。
- D、幂等性无法保证跨分区跨会话的Exactly Once。

【参考答案】：C

【您的答案】：C

---

**【正确】** 题目9：

下列关于channel 的特征描述错误的是？[单选题]

- A、Channel是位于Source和Sink之间的缓冲区。
- B、Channel允许Source和Sink运作在不同的速率上。
- C、Channel是线程不安全的。
- D、可以同时处理几个Source的写入操作和几个Sink的读取操作。

【参考答案】：C

【您的答案】：C

---

**【错误】** 题目10：

下列最不适用于memory channel 传输数据的情景是？[单选题]

- A、用户交易数据
- B、用户个人信息
- C、用户行为数据
- D、集群状态监控日志数据

【参考答案】：A

【您的答案】：D

---

【正确】 题目11:

下列关于flume 的配置中能够体现出多路复用特征的是? [单选题]

- A、一个source 对应多个channel (selector.type为multiplexing)
- B、一个source 对应多个channel (selector.type为replicating )
- C、多个source 对应一个channel (selector.type为multiplexing)
- D、多个source 对应一个channel (selector.type为replicating )

【参考答案】：A

【您的答案】：A

---

【错误】 题目12:

下列关于flume 数据丢失描述正确的是? [多选题]

- A、框架本身的不稳定性。
- B、agent宕机导致数据丢失。
- C、channel存储数据已满，source无法写入导致未写入数据丢失。
- D、只有在使用memory channel 的情况下数据才会可能丢失。

【参考答案】：BC

【您的答案】：CD

---

【正确】 题目13:

下列常用source 中描述正确的是? [多选题]

- A、Taildir Source：支持断点续传、多目录文件监控。
- B、Avro Source：Avro端口监听并接收来自外部的Avro客户流的事件。
- C、Exec Source：Exec Source的配置就是设定一个Unix(linux)命令，然后通过这个命令不断输出数据。
- D、Spooling Directory Source：监测配置的目录下新增的文件，并将文件中的数据读取出来。

【参考答案】：ABCD

【您的答案】：ABCD

---

【正确】 题目14:

下列常用sink 中描述正确的是? [多选题]

- A、HDFS Sink：当需要将事件消息写入到Hadoop分布式文件系统（HDFS）时，可以使用HDFS Sink。
- B、Avro Sink：和 Avro Source一起工作，用于构建Flume分层收集数据消息结构。
- C、Kafka Sink：通过该Sink可将事件消息数据发布到Kafka topic 上。
- D、Logger Sink：可以将数据输出到控制台上。

【参考答案】：ABCD

【您的答案】：ABCD

---

【正确】 题目15:

下列关于kafka 造成数据丢失的场景描述正确的是? [多选题]

- A、ack=0时，leader故障有可能丢失数据。
- B、ack=1时，在follower同步成功之前leader故障，将会丢失数据。
- C、ack=-1时，在follower同步完成后，broker发送ack之前，leader发生故障，可能会丢失数据。
- D、ack=-1时，在follower同步完成前，broker发送ack之后，leader发生故障，可能会丢失数据。

【参考答案】：AB

【您的答案】：AB

---

【正确】 题目16:

下列关于kafka 的分区和副本描述错误的是? [单选题]

- A、一个topic可以分为多个partition，分布式存储在不同的服务器上。
- B、每一个partition内部一定是一个无序的队列。
- C、一个topic的每个partition都有若干个副本，分布式存储在不同的服务器上。
- D、每个分区及副本都有可能成为leader或者follower。

【参考答案】：B

【您的答案】：B

---

【正确】 题目17：

下列关于kafka 文件存储机制划分描述错误的是？[单选题]

- A、每个topic可以分为多个partition，分布式存储在不同的服务器上。
- B、每个partition命名规则是partition-[orderNumber]，每个partition又分为多个segment。
- C、每个segment对应两个文件：".index"文件和".log"文件。
- D、".index"文件存储大量的索引信息，".log"文件存储大量的数据。索引文件中偏移量等于实际写入该分区的条数减一。

【参考答案】：D

【您的答案】：D

---

【正确】 题目18：

下列关于kafka 生产者分区策略描述正确的是？[多选题]

- A、需要将producer发送的数据封装成一个ProducerRecord对象。
- B、指明 partition 的情况下，直接将指明的值直接作为 partiton 值。
- C、没有指明 partition 值但有 key 的情况下，将 key 的 hash 值与 topic 的 partition 数进行相除取整得到 partition 值。
- D、既没有 partition 值又没有 key 值的情况下， kafka采用Sticky Partition(黏性分区器)，会随机选择一个分区，并尽可能一直使用该分区，待该分区的batch已满或者已完成，kafka再随机一个分区进行使用。

【参考答案】：ABD

【您的答案】：ABD

---

【正确】 题目19：

下列关于kafka生产者的副本数据同步策略描述错误的是？[单选题]

- A、leader维护了一个动态的ISR队列，当ISR中的follower完成数据的同步之后，leader就会给producer发送ack。
- B、如果follower长时间未向leader同步数据，则该follower将被踢出ISR。
- C、leader发生故障之后，就会从ISR中选举新的leader。
- D、当acks=-1时，leader接收到消息在写入磁盘之前就会返回ack。

【参考答案】：D

【您的答案】：D

---

【错误】 题目20：

下列关于kafka 中两个重要索引位置描述错误的是？[多选题]

- A、LEO：指的是每个副本最大的offset。
- B、HW：指的是消费者能见到的最大的offset，ISR队列中最小的offset。
- C、LEO：指的是消费者能见到的最大的offset，ISR队列中最大的offset。
- D、HW：指的是每个副本最大的offset。

【参考答案】：CD

【您的答案】：AB