1. FusionInsight HD 中，Oozie 在提交作业前需要先上传本作业所依赖的配置文件和 jar 包到 HDFS

对

1. FusionInsight HD 中，Hive 客户端以下哪些参数是用来控制 reduce 并发度的？（多选）

hive.exec.reducers.bytes.per.reducer hive.exec.reducers.max

hive.exec.max.dynamic.partitions hive.exec.max.dynamic.partitions.pernode AB

1. FusionInsight HD 系统中，Solr 示例工程 solr-examplelconfsolr-example.properties 中 ZK URL 和 ZOOKEEPER DEFAULT SERVER PRINCIPAL 的值可以从下面哪个选项获取？

从 Solr Admin UI 界面获取

在 solr-examplelconf solr-example.properties 中，已经默认带有这两个参数从 FusionInsight Manager Solr 服务配置界面获取

A

1. Fusionlnsight HD 中，属于 Streaming 的数据分发方式有哪性？（多选）

Shuffle Grouping Field Grouping Laal Gg

Direct Grouping ABD

1. 一个应用需要同时和两个 FusionInsight 集群交互：既需要访问集群 1 的 HBase 服务，X 需要切集郡 2 的

Hive 服务，那么如下哪儿项操作是必须的？多选）

-在集群 1 上创建本应用所需的账号，并设置此账号能访问本集群的 HBase 服务

-在集群 2 上创建 A 选项中的同名账号，并设置此账号能访问本集 e 的 Hive 服务

-A 选项和 B 选项所创建的账号都必须属于 supr。集群 1 和集群 2 完成互信操作

BC

1. 以下哪些措施能提高 HBase 的查询性能？（多选） 预分 region，使 region 分布均匀，提高并发

查询频繁属性放在 rowkey 前面部分使用二级索引，适应更多查询场景 同时读取的数据存放在同一个 cell 中ABCD

1. 关于 Redis 集群拓扑信息，下面描述正确的是？客户端缓存有集群的拓扑信息。服务端缓存有集群的拓扑信息。两者都是
2. 在 Streaming 的应用开发中，Bolt 类使用下面哪个接口发送 Tuple？

emit (发出) execute open nextTuple A

1. 在 Spark 中，SparksQL 是独立的模块，可以不依赖于 SparkCore 独立完成 SQL 语句的角行等操作。错
2. FusionInsigt HD 中，一个部署在 Tomcat 上的应用，需访向 HBase 服务，这种场景下推荐使用机机账号。错
3. 关于 Flume 中的数据流向，下列哪一项描述是正确的？

（1）数据从 source 到 channel；（2）sink 从 channel 取数据

（1）数据从 channle 到 source；（2）sink 从 source 取数据。

（1）数据从 sink 到 channle；（2）source 从 channel 取数据 o

（1）数据从 channle 到 source；（2）sink 从 source 取数据

A

1. 存在如下业务场景：用户上网日志文件已经存放在 HDFS 上，日志文件内容格式是：每条上网记录三个字段， 分别是姓名，性别，上网时间，字段之间用“，”分隔；要求打印输出所有上网总时间大于两小时的女性网友。请问如下哪些代码片段能实现上述业务场景？（多选）

A、sc.textFile（“/data/file/path”）.map（.spli（t“，”））.map（p=>FemaleInfo（p（O），p（1），p（2）.trim.toInt））.toDF.registerTemp Table（“FemaleinfoTable”）sqlContext.sq（l“select name,sum（stay Time）as stay Time from FemaleInfoTable where gender= female‘group by name”）.filter（“stay Time >=120”）.collect）.foreach（printin）

B、sc.textFile（“/data/file/path”）.map（.spli（t“，”））.map（p=>FemaleInfo（p（O），p（1），p（2）.trim.toInt））.toDF.registerTemp Table（“FemaleinfoTable”）sqlContext.sq（l“select name,sum（stay Time）as stay Time from FemaleInfoTable where gender= female）.filter（“stay Time >=120”）.collect）.foreach（printin）

C、val text=sc.textFile（“/data/fle/path”）val data=text.filter（\_.contains（“female”））Val fermaleData:RDD(String,Int)

=data.map{line=>val t=line.split(‘,’)(t(0),t(2).toInt)}.reduceByKey(\_+\_)Val result=femaleData.filter(line=>line. 2>120) Rusult.collect.map(x=>x.\_1 + ‘,’ + x.\_2).foreach(println)

D、val text=sc.textFile（“/data/fle/path”）val data=text.filter（\_.contains（“female”））Val fermaleData:RDD(String,Int)

=data.map(line=>val t=line.split( ‘ , ’ )(t(0),t(2).toInt)}Val result=femaleData.filter(line=>line. 2>120)Rusult.collect.map(x=>x.\_1 + ‘,’ + x.\_2).foreach(println)

BC

1. 在 FusionInsight HD 产品的要掌 38-以 5 品+品 4 天期雅 S 容甲开发时，可以通过 Solr Admin UIx 对

Collection 进行一些验证。下面（多选）

点击 Cloud 下的 Tree，查看 Solr Cloud 中 cafrCollections、配置集和 ive nodes 等元数据信息点击 cloud，查看备 calection 下各个 shard 的 Replica 的分布及状态

进入来 Collectionyshard Replicats 意地大你和置信检 1dKepca 的（ore-Oevew 开面，可以查看测该 Relica 索引实际交档条数

solr 用户组、solr、admin 须他和.Rrg 的用产可以 i 的 sr 公的和和山

ACD

1. 某项目需要对某个地区的上网数蕾苯地区的犯重。请简妥南摇教华进需保存，并全检素这些上网记录里是否有敏感信息，用于预最优方案？

在 sor 上建立索引，并保存数据，在全文检素时，将所有数据全部返回在 HBase 上存放数据，通过 HBase 的过滤特性，满足模糊匹配查询。

存放数据时，在 Solr 上建信息，通过关使信是在电的要是缺质莞墨爸拿放完整数据，查询时，通过 sor 全文检索获取记录关键

C

1. Spark 应用的计算逻辑会被解析成 DAG，这个解析操作由以下哪个功能模块完成？ Client

ApplicationMaster Executor 。 Driver

D

1. 在 Spark 中 ， 以 下 哪 项 是 对 DataFrame 取 交 集 的 算 子 ？ filer（conditionExpr:String）：DataFrame 根据过滤条件对数据进行过滤 Join(right:DataFrame,joinExprs:Column）：DataFrame 根据指定的表达式进行 join intersect（other:DataFrame）：DataFrame 对 2 个 dataframe 取交集 dropDuplicates(）：DataFrame 数据去重

C

1. FusionInsight HD 中，关于 Oozie 的作业管理，对于已暂停的作业可以进行以下哪几项操作？（多选） 口 kill

口 run

口 resume（唤醒）

口 submit AC

1. HBase 表的 rowkey 设计是一个很重要的开发设计环节。假设存在如下场景，最频繁的查询场景是基于手机号查询每个月、每半年的历史通话记录，以下哪个 rowkey 设计是最优的？

。姓名十手机号

。日期十手机号

。手机号+日期

。手机号+姓名

C

1. FusionInsight HD 系统中，关于 Hive 的 JDBC 接口类型，以下哪种描述是正确的？ 组网隔离环境下，ZooKeeper 的 IP 既可以配置成业务 IP，也可以配置成管理 IP。hive 与传统数据库不同，因此 JDBC 建立的连接不需要关闭。

安全集群环境下，建立与 Hive 的连接前，需要使用统一认证方式进行登录。使用 JDBC 提交 SQL 时，可以在 SQL 语句后面加上分号“;”

C

319 在使用 Solr 进行全文检索时，可以通过 wt 参数指定查询结果的响应格式。关于 Solr 查询结果的响应格式， 下面说法错误的是？

支持 CSV 和 JSON

支持 CSV、JSON 和 HTML 支持 CSV、JSON 和 XML B

320.向 HBase 集群中增加 RegionServer 主机时，必须先停止原集群，原因是 HBase 不支持动态扩容。对321.FusionInsight HD 中，使用 Streaming 的 Linux 命令行模式提交拓扑时，需要先使用一个拥有 Streaming 提交权限的用户进行 kinit 方式认证 对

322.在 FusionInsight HD 的客户端，执行 kinit{帐号}命令是为了得到 KDC 的哪一项内容？

Krb5 conf TGT

jaas.conf C

1. 关于 FusionInsight HD 平台的 Hive 服务，其 WebHCat 开发接口，以下描述不正确的是？

支持基于 REST 的查询请求

.WebHCat 的返回数据格式是 XML

.WebHCat 基于 HTTP 和 HTTPS 协议对外提供服务

.可以通过 WebHCat 进行建表，查询等操作

B

1. 在 Spark 中，因为 DataFrame 是比 RDD 更上层的接口，调用层次会更多。因此，所有基于 DataFrame 开发的程序的性能会比基于 RDD 开发的程序的性能差。 错
2. 下来哪些 Redis 命令属于原子操作？

.INCR（将储存的数字值增一）

.HINCRBY（为哈希表 key 中的域 field 的值加上增量 increment）

.LPUSH（对链表的 push 操作）

.以上都是D

1. fusioninsight 中，一个应用程序若要充成 Kerberos 安全认证，以下哪些选项是必须用到的？（多选）

krb5.conf 配置文件

在 FusionInsight Manager 上申请的用户账号用户账号的 keytab 文件

jaas.conf 配置文件

ABC

1. FusionInsightHD 中，关于 Hive 的分区（partation）功能，如下描述错误的是？。分区字段要在创建表时定义

分区字段只能有一个，不可以创建多级分区

使用分区，可以减少某些查询的数据扫描范围，进而提高查询效率。分区字段可以作为 where 字句的条件

B

1. 在 Solr 建立索引时，下面哪些代码片段可以显著提升建立索引性能？（多选） 口 IbClient.setRequestWriter（new BinaryRequestWriterO）；

口 if （ updateRequest.getDocumentsMapO.size ） >=10000 ） {cloudSolrClient.request （ updateRequest,collName ）； updateRequest.clear）；

口 if （ updateRequest.getDocuments ） .size （ O>=1000 ） {cloudSolrClient.request （ updateRequest,collName ）； updateRequest.clearO； S8tW9H2xgntdc=new SolrInputDocument）；doc.addField（“id”，i.tostring（）doc.addField

（“name”，“张三 4i）；doc.addField（”features“，”test“+i）：doc.addreldc”price，（float）\*1.oi）；cloudSolrClient.add

（doc）；cloudSolrClient.commitO： AB

1. FusionInsightHD 系统中，Hive 支持的接口类型有哪些？（多选） 口 JDBC 接口

口 ODBC 接口口 Python 接口口 Ruby 接口ABC

1. 通过 HBase 的 createTable 方法创建一张表，必须传入的参数有？

.表名

.表名和列

.表名和列族

.可以为空C

1. 以下哪个不是 MapReduce 的特点？

易于编程 良好扩展性 实时计算 高容错性

C

1. FusionInsight HD 中，属于 Streaming 的数据分发方式有哪些？（多选） 口 Shuffle Grouping 随机分组

口 Field Grouping 字段分组口 Local Grouping 本地分组口 Direct Grouping 直接分组ABCD

1. 在 Kafka 中，Producer 可以通过配置同步参数（producer.type），保证数据按顺序发送。对
2. HDFS 采用的是“一次写入、多次读取”的文件访问模型。所以推荐一个文件经过创建、写入利之后，就不要再去修改。 对
3. 通加能的社 HD 的 e 中，用户自定义的 UDF 可以和 H 机公内置的 UDF 班名，这种情况下，将使用用户自定义的 UDF。 对
4. 下列关于 HDFS 客户端读文件过程正确的是？（多选） 口优先从最近的 DataNode 下载数据。

口文件数据先从 DataNode 返回至 NameNode，然后从 NameNode 下载到客户端。

如果客户端和所连接的 DataNode 在读取时出现故障，那么客户端会放弃这个故障节点，去连接其也副本所在的节点。

口支持多个客户端同时从 DataNode 上读取同一个文件数据。

AC

337.s 国 onnsgtHD 安全版本进行应用开发时，可采用 keytab 文件进行 Kerberos 安全认证 对

338 FusiontnsightHD 中，可以通过哪几种方式查看 Oozie 作业的调试结果？（多选） 口查看 Oozie 的 JAVA API 返回结果是否符合预期

口通过 Hue 的 workflow 仪表板查看程序运行结果

口通过 Hue 的文件浏览器查看是否在 HDFS 指定目录生成了所期望的文件口 Oozie 自带的 Web 界面也可以查看作业执行结果

ABCD

1. solr 是一个高性能，基于 Lucene 的全文检索服务.Solrx 对 Lucene 进行工扩展，损供工果的支持 SolrCloud 模式。对
2. 关于 Streaming 的拓扑（Topology），下面描述错误的是？

一个 Topology 是由一组 Spout 组件和 Bolt 组件通过 Stream Groupings 进行连接的有向无环图（DAG）

Topology 会一直运行，直到它被显式 kill

业务逻辑都被封装进 Topology 中

一个 Topology 只可以指定启动一个 Worker 进程

D

1. 对于多项业务共用 rowkey，导致数据访问矛盾的情况，可以有哪些应对策略？（多选） 口折中法（rowkey 设计兼顾 A 和 B 的业务需求）

口冗余法（用空间换时间/效率） 口创建二级索引

口预分 region ABC

342FusionInsightHD 中，哪些组件对外提供 SQL 或类 SQL 能力？（多选） 口 HBase

口 Hive 口 Spark 口 Redis

BC

343 在 Kafka 中，如下关于 Producer 发送数据的说法错误的是？（多选）

Producer 负责生产数据，Consumer 负责消费数据，Producer 和 Consumer 之间需要建立 Socket 连接

Producer 既可以发送数据到 Broker 上，又可以发送数据到 Consumer 上 Producer 作为消息生产者，可直接将数据写入 ZooKeeper

口 Producer 连接任意一个正常的 Broker 实例就可以生产数据

AB

344.在 MapReduce 的应用程序开发中，下列哪个 JAVA 类负责管理和运行一个计算任务？

.Job

.Context FileSystem Configuration A

345 对于 Spakstreaming 的应用，在一个 JVM 中，同一时间只能有一个 Streamingcontext 处于活跃状态。对

1. HDFS 客户端以 N 副本向 HDFS 写文件时，若其中一个副本写入失败，则所有副本都会返回入失败。错
2. 关于 Flume 采集数据的特点，下列描述正确的是？

。支持分布式和流式采集

。借助 MapReduce 任务并行采集

。只能采集结构化数据

。借助外部定时任务来触发数据采集

B

1. 对于 HBase rowkey 的设计原则，如下描述正确的是？（多选） 口访问权重高的属性值放在 RowKey 前面

口访问权重高的属性值放在 Rowkey 后半部分口离散度好的属性值放在 Rowkey 前半部分口离散度好的属性值放在 Rowkey 后半部分AC

1. Redis 的 LIST 数据结构，适用于下列哪些场景？（多选） 口构建队列系统，例如消息队列

口 uniq 操作，例如获取某段时间所有数据的排重值

口取最新 N 个数据操作：例如对某条微博，获取最新的 10 个评论口模拟栈操作

AC

1. 在 Kafka 中，关于 Consumer 说法正确的是？（多选） 口 Consumer 主要进行数据消费

口 Consumer 消费数据时需要指定 Topic

口多个 Consumer 配置相同的 group.id 时，这些 Consumer 消费的数据不会重复口新 ConsumerAPI 消费数据不用指定 ZooKeeper 地址

ABC

1. 在 FusionInsight HD 中，可以在哪里查看 MapReduce 应用程序的运行结果？

.FusionInsight Manager 的 Web 界面

.Loader Web 界面

.Solr Web 界面

.Yarn Web 界面

D

351. HDFS 不适用于下列哪种业务场景？

。海量数据存储

。大量小文件存储

。大文件存储

。流式文件存储

B

353 FusionInsigt HD 中，下列哪些组件可用于数据采集？（多选） 口 Flume

口 Loader

口 HBase

口 MapReduce AB

1. 在 Base 的应用开发中，当一个表的 Rowkey 范围和分布已知，建议预分 region，请问调用如下代码（片段） 对某个表进行预分 region 后，这个表会创建几个 region？

splits【0】=Bytes.toBytes（“A”）；splits【1】=Bytes.toBytes（“H”）；splits【2】=Bytes.toBytes（“O”）；splits【3】=Bytes.toBytes

（“U”）；admin.createTable（htd,splits）；

.3

4

5

6

C

1. 下列哪些原因会导致 HDFS 的 NameNode 进入 safemode（安全只读模式）？（多选） 口主备 NameNode 的元数据所在磁盘空间不足。

口丢失块数超过阀值。口丢失副本超过阀值。口损坏副本超过阀值。ACD

1. 在 Spark 中，关于广播变量以下说法正确的是？（多选） 口广播变量在每个 Task 中保存一份

口广播变量在每个 Executor 中保存一份口广播变量是只读的

口广播变量通过 SparkContext.broadcast）方法生成BCD

1. 当一个 MapRcdue 应用程序被执行时，如下哪一个动作是 map 阶段之前发生的？

split combine partition sort

A

1. 在 flume 中，source 功能模块的主要作用是？

。获取数据，并将原始数据转化成自己处理的数据对象

。缓存数据，根据不同的可靠性策略，将数据保存在内存或文件上

。输出数据到目的地，支持多种输出协议

。拆分数据，根据数据的特性，将数据发送到不同的目的地

A

1. 在 HBase 应用程序运行中，应用可以在创建表的同时写入数据。 错
2. Rcb 的 t0 系统中，关于 e 提供的文件格式，哪个不是列形式是？

ORC（高性能列存储） Parquet 列存储

RCFile 行列混合型存储

TextFile D

1. 当一个任务被提交到 YARN 集群中，下面哪个选项会为这个任务生成 T 个

JOPO

ResourceManager container AppMaster

NodeManager B

1. 高加高共 1H0 中，Q 业客产端的心 A.AI 在运行任务时会调用 Qoziec1et 类的哪个方法？

.Suspend getJobInfo submit run

D

1. FusionInsight HD 中，假设一个拓扑，设道了 spout 并发度为 3，bolt1 并发度为 2，bolt2 并发度为 3，worker

数目为 2，那么 bolt1 的 executor 在 worker 中是如何分布的？

。两个 worker 中的其中一个 worker 分布两个

。两个 worker 各分布一个

。没有固定规律

。两个 worker 都分布两个

A

1. 下面哪个选项是 Yarn 的核心功能？

。存储

。资源管理

。搜索

。数据传输

B

1. 关于 streaming 的容灾能力，下列说法正确的是？（多选）

Supervisor 进程退出后能够被 Nimbus 自动识别并拉起，不影响正运行的业务

Worker 异常退出后能够被 supervisors 自动拉起，无需人工干预

节点故障时，该节点上的任务会被重分配到其他正常的节点，无需人工干预

Nimbus 故障后，备 Nimbus 会自动接管，不影响正在运行的业务

ABCD

1. FusionInsigt HD 中，一个应用程序若要完成 Kerberos 安全认证，以下哪些选项是必须用到的（多选）

Krb5.conf 配置文件

在 fusion insight manager 上申请的用户账号用户账号的 keytab 文件

Jaas.conf 配置文件

ABC

1. fusion insight HD 中，streaming 的特点有哪些？ 流处理

高性能可扩展高容错ABCD

1. Flume 写入 HDFS 时，文件的生成方式有哪些？（多选） 按照源文件名，直接生成文件

设置文件生成时间间隔，达到时间间隔时就关闭老文件，新数据将被写入新文件

设置单个文件大小，，当文件大小达到所设定值时，就关闭老文件，新数据将被写入新文件

设置单个文件存放 event 数量，当文件所存放 event 数量到设定值时，就关闭老文件，新数据将被写入新文件

BCD

369，Spark streaming 作为微批处理的流处理引擎，是将每个时间分片的数据转换为一个 RDD 中的一个分区进行计算 对

370，在 Flume 级联传输的过程中，可以使用 failover 模式传输，这样如果吓一跳的 Flume 节点故障或者数据接受异常时，可以自动切换到另外一路上继续传输 对

371，FusionInsight HD 系统中，hive 支持的接口类型有哪些？

JDBC 接口ODBC 接口Python 接口Ruby 接口ABC

372，将数据块存储大小设置为 128M，HDFS 客户端在写文件时，当写入一个 100M 大小的文件，实际占用存储空间为多大？

128M

100M

64M

50M A

373，Spark 应用的计算逻辑会被解析成 DAG，这个解析操作由以下哪个功能模块完成？？ Client

ApplicationMaster Executor

Driver D

374，Fusion Insight HD 中，Streaming 适用于以下哪些场景？（多选） 流数据监控

网站的实时访问统计离线日志分析

交通流量分析

ABD

375，存在如下场景：每天由线上系统产生新数据 500G，需要对这些数据按天、周、月等维度做统计汇总。请问它适合用 Hive 哪种表来处理？

分桶表，按照天分桶分区表，按天分区

倾斜表，倾斜字段选择日期，倾斜值随机

分区+分桶表，按照时间分区，每个分区 2560 个桶

D

376.Fusioninsight HD 支持通过哪些方式提交 Oozie 作业？（多选）

Shell 命令Java API Hue UI

Rest API ABCD

377,假设有一个应用，需要频繁访问 Oracle 数据库中的用户表，为提高性能，引入 Redis 来缓存用户信息。对此场景，Redis 的最佳数据结构选择是以下哪项？

List Set Hash

Sorted set C

378,HDFS 的应用开发中，下列哪些场是 HDFS 服务支持的接口？（多选）

BufferedOutputStream.write BufferedOutputStream.flush FileSystem.create FileSystem.append

ABCD

1. Fusioninsight HD 中，对 solr 各类资源的创建和读写权限适用，以下说话错误的是？

Solr 用户组用户可以创建配置集，并在创建 collection 是指定该配置集

Solr 用户组的用户可以创建 collection

Collection 创建者、对该 collection 具有读写权限的角色以及 solr admin 角色可以对 collection 进行更新只有 collection 创建者和 solr admin 角色可以删除 collection

D

1. Fusioninsight HD 系统的 v100R002C60 版本中，Hive 仅支持基于 MapReduce 引擎的查询服务，不支持基于

Spark 引擎的查询服务 错

1. Fusioninsight HD 中，适用 streaming 的命令行方式提交 example.jar 中的 om.huawei.example.wordcount 任务， 任务名称为 wcTest，以下执行命令正确的是？

Storm.jar example.jar wcTest om.huawei..example.wordcount Storm.jar example.jar wcTest wordcount

Storm.jar example.jar om.huawei..example.wordcount wcTest Storm.jar example.jar wordcount wcTest

C

1. Hive 支持下列哪些表类型？（多选） 分区表

分桶表倾斜表

分区+分桶表

ABCD

1. 当进行 solr 的某个 collection 设计时，需要设计其 schema，通过配置 schema.xml 文件实现对 schema 的设计，下面关于 schema 说法错误的是？

A、Schema 类似于关系数据库中的表结构，用于定义索引数据的结构主要包含三部分：uniquekye、Field 和 FieldType B、对于不需要存储原始值的字段，其属性 stored 可以设置为 false

C、对于同一条 document 的某个字段有多少个值，其都需要保留，那么该字段的 multiValued 必须设置为 true

D、copyField 可以进行如下定义：<field name=”cat” type=”string” indexed=”true” stored=”true”docValues=” true” multiValued=”false”/><field name=”dog”type=”string” indexed=”true” stored=”true” docValues=” true”multiValued=”false”/><field name=”text”type=”text\_general” indexed=”true” stored=”false” multiValued=” false”/><copyField source=”cat” dest=”text”maxChars=”30000”/> <copyField source=”dog” dest=”text”maxChars=” 30000”/>

D

384,在通过 Fusionindight Manager 界面创建用户时，想要该用户对所有 Topic 具有读写权限，但不具有创建 Topic

的权限，应该加入 kafka 服务的哪个组？

Kafkaadmin 组Kakfa 组kafkasuperuser 组C

385.Fusioninsight HD 的 hive 应用中，存在如下场景：对数据存放文件有较高的压缩效率，且大部分查询只涉及文件的一部分字段。这种场景适合采用列式文件（ORC File）存储 对

386,某项目需要对某个地区的上网数据进行保存，并全文检索这些上网记录里是否有敏感的信息，用于预防本地区的犯罪，请问这种场景下，如下哪种方案最优方案？

在 solr 上建立索引，并保存数据，在全文检索时，将所有数据全部返回

在 HBase 上存放数据，通过 HBase 的过滤特性，满足模糊匹配查询

存放数据时，在 solr 上建立索引，在 HBase 存放完整数据：查询时通过 solr 全文检索获取记录关键信息，通过关键信息在 HBase 上获取完整记录

CD

1. HDFS 客户端以 N 副本向 HDFS 写文件时，若其中一个副本写入失败，则所有副本都会返回写入失败错
2. 关于 Redis 特点，下面描述正确的是？ 低时延

丰富的数据结构支持数据持久化

支持 SQL 查询接口

ABC

389，在 Spark 中，假设 lines 是一个 DStream 对象，以下哪个语句可以周期统计这个流上的单词数？ lines.flatMap（\_.split（“ “））.map（word =>（word，1））.reduce（\_+\_).print（） lines.flatMap（\_.split（“ “））.map（word =>（word，word.length()））.reduceByKey（\_+\_).print（） lines.flatMap（\_.split（“ “））.map（word =>（word，1））.reduceByKey（\_+\_).print（） lines.flatMap（\_.split（“ “））.flatmap（word =>（word，1））.groupByKey（\_+\_).print（）

C

1. 在 HBase 写数据的业务逻辑实现过程中，如下哪个接口或类是不需要涉及的？

A.Put B.HTable C.HBaseAdmin D.PutList Answer: C

1. 在 MapReduce 的应用程序开发中，下列哪个 JAVA 类负责管理和运行一个计算任务？

A.lob B.Context C.FileSystem D.Configuration Answer: A

1. Fusionlnsight HD 中，Oozie 在提交作业前需要先上传本作业所依赖的配置文件和 jar 包到 HDFS A.正确 B.错误

Answer: A

1. 以下哪些方法可以生成 DStream 对象？

A.KafkaUtils.createStream（…） B.KafkaUtils.createDirectStream（…） C.StreamingContext.socketStream D.StreamingContext.fileStream（…） Answer: ABCD

1. Fusionlnsight HD 系统中，关于 Hive 提供的文件格式，哪个不是列式文件？

A.CRC B.Parquet C.RCFile D.TextFile Answer: D

1. Fusionlnsight HD 中，对 Solr 各类资源的创建和读写权限使用，以下说法错误的是？
2. Solr 用户组用户可以创建配置集，并在创建 Collection 时指定该配置集
3. Solr 用户组的用户可以创建 Collection
4. Collection 创建者、对该 Collection 具有读写权限的角色以及 Solr admin 角色可以对 Collection 进行更新
5. 只有 Collection 创建者和 Solr admin 角色可以删除 Collection Answer: A
6. 在 Spark 中，以下哪项是对 DataFrame 取交集的算子？

A. filter (condifionExpr: String): DataFrame B. join (right: DataFrame, join Exprs: Column): DataFrame

C. intersect (other: Data Fra me): Data Fra me D. dropDuplicates (): DataFrame Answer: C

1. 通过 HBase 的 createTable 方法创建一张表，必须传入的参数有？

A.表名 B.表名和列 C.表名和列族 D.可以为空Answer: C

1. HDFS 客户端所在节点的系统时间与 Fusionlnsight HD 集群的系统时间要保持一致，若有时间差，那么时间差应小于几分钟？

A.5 B.10 C.20 D.30

Answer: A

1. 在 Flume 级联传输过程中，可以使用 fail over 模式传输，这样如果下一跳的 Flume 节点故障或者数据接受异常时，可以自动切换到另外一路上继续传输。

A.正确 B.错误Answer: A

1. Fusionlnsight HD 中，属于 Streaming 服务的角色有哪些？（多选）

A.Nimbus B.Supervisor C.Broker D.quorumpeer Answer: AB

1. 关于 Streaming 的拓扑（Topology），下面描述错误的是？
2. 一个 Topology 是由一组 Spout 组件和 Bolt 组件通过 Stream Groupings 进行连接的有向无环图（DAG）
3. Topology 会一直运行，直到它被显式 kill
4. 业务逻辑都被封装 Topology 中
5. 一个 Topology 只可以指定启动一个 Worker 进程

Answer: D

1. HDFS 客户端以 N 副本向 HDFS 写文件时，若其中一个副本写入失败，则所有副本都会返回写入失败。

A.正确 B.错误Answer: B

1. Fusionlnsight HD 中，使用 Streaming 的命令？方式提交 example.jar 中的 om huawei example WrodCount 任务，任务名称为 wcTeat，以下执行命令正确的是？

A. storm jar example jar wcTest om huawei example WordCount B. storm jar example jar wcTest WrodCount

C. storm jar example jar om huawei example WordCount wcTest D. storm jar example jar WorkCount wcTest Answer: C

1. 对于运行在 MapReduce 平台上的应用程序，此应用程序所依赖的 jar 包会被放到哪里？

A.HIVE B.HBASE C.HDFS D.DB

Answer: C

1. RDD 作为 Spark 最核心的对象，具有以下哪些特点？（多选）

A.只读 B.分区 C.容错 D.高效Answer: ABCD

1. Fusionlnsight HD 中，在实时处理场景中，可使用的计算框架有哪些？（多选）

A. Spark Streaming B.Streaming C.MapReduce D.HDFS Answer: AB

407Fusionlnsiht HD 中，Oozie 客户端的 JAVA API 在运行任务时会调用 OozieClient 类的哪个方法？

A. Suspend B.getJoblnfo C.submit D.run Answer: D

408.将数据块存储大小设置为 128M，HDFS 客户端在写文件时，当写入一个 100M 大小的文件，实际占用储存空间为多大？

A.128M B.100M C.64M D.50M

Answer: B

409.下列哪些原因会导致 HDFS 的 NameNode 进入 safemode（安全只读模式）？（多选） A.主备 NameNode 的元数据所在磁盘空间不足。 B.丢失块数超过阈值。

C.丢失副本超过阈值。 D.损坏副本超过阈值。Answer: AB

410.对于 Fusionlnsight HD 平台的 HBase 组件，增加二级索引需要定义二级索引的哪些属性？（多选）

A.索引名 B.索引列 C.索引列类型 D.索引列所归属的列族名称411.当一个 MapReduce 应用程序被执行时，如下哪一个动作是 map 阶段之前发生的？ A.split B.combine C.partition D.sort

Answer: A

1. Fusionlnsight HD 中，关于 Hive UFD 二次开发，如下描述正确的是？（多选）
2. 用户自定义的 UDF 在被使用之前，需要先在 Hive 系统中创建此 UD 只
3. 用户自定义的 UDF，不允许加入汇总和状态等信息。
4. 用户自定义的 UDF，可以根据实际情况添加 deterministic 和 statefull 等注解。
5. 安全集群下，用户自定义的 UDF，建议在每次使用前都创建一次。Answer: AC
6. 在 Fusionlnsight HD 产品中，Kafka 服务的角色名称是？

A.Producer B.Broker C.Consumer D.ZooKeeper Answer: B

1. Spark 应用程序有哪几种参数设置方法？（多选） A.在应用的配置文件 spark-defaults.conf 中进行配置B.在提交应用时，通过—conf 进行设置
2. 在应用的代码里面，通过 SparkContext 的 setProperty 方法进行设置
3. 在应用的代码里面，通过 SparkConf 对象进行设置

Answer: ABD

1. 在 Fusionlnsight HD 集群中，Flume 不支持将采集数据写入到集群中那个服务？

A.HDFS B.HBase C.Kafks D.Redis Answer: D

1. 在 Spark 中，假设 lines 是一个 DStream 对象，以下哪个语句可以周期统计这个流上的单词数？

A. lines.flatMap(\_.split? ?)).map(word=>(word,1)).reduce(\_+\_).print ( ) B.lines.flatMap(\_.split??)).map(word=>(word,word.length())).reduceByKey(\_+\_).print()

1. lines.flatMap(\_.split? ?)).map(word=>(word,1)).reduceByKey(\_+\_).print ( )
2. lines.flatMap(\_.split? ?)).flatmap(word=>(word,1)).groupByKey(\_+\_).print ( ) Answer: C
3. HBase 过滤器能将列名或列值设置为过滤条件，同时支持多个过滤器配合使用。

A.正确 B.错误Answer: A

1. 关于 Streaming 的容灾能力，下列说法正确的是？（多选）
2. Supervisor 进程退出后能够被 Nimbus 自动识别并拉起，不影响正在运行的业务。
3. Worker 异常退出后能够被 Supervisor 自动拉起，无需人工干预。
4. 节点故障时，该节点上的任务会被重分配到其他正常的节点，无需人工干预。D.Nimbus 故障后，备 Nimbus 会自动接管，不影响正在运行的业务。

Answer: ABCD

1. HBase 表的 rowkey 设计是一个很重要的开发设计环节。

A.姓名+手机号 B.日期+手机号 C.手机号+日期 D.手机号+姓名Answer: C

1. Fusionlnsigt HD 中，下列哪些组件可用于数据采集？（多选）

A.Flume B.Loader C.HBase D.MapReduce Answer: AB

421 关于 Solr 查询结果的响应格式，下面说法错误的是？

A.支持 CSV 和 JSON B.支持 CSV、JSON 和 HTML C.支持 CSV、JSON 和 XML

Answer: B

1. 在 Fusionlnsight HD 产品的 Solr 应用开发时，可以通过 Solr Admin UI 对 Collection 进行一些验证。下面关于

Solr Admin UI 说法正确的有？（多选）

1. 点击 Cloud 下的 Tree，查看 Solr Cloud 中 Collections、配置集和 live\_nodes 等元数据信息
2. 点击 Cloud，查看各 Collection 下各个 ShardReplica 的分布及状态
3. 进入某 Collection 的 Shard Replica 的 Core Overview 界面，可以查看到该 Replica 索引实际文档条数、存储大小和位置信息
4. Solr 用户组、Solr admin 角色和 Super group 的用户可以访问 Solr Admin UI Answer: ABCD
5. 某项目需要对某个地区的上网？？进行保存，并全文检索这些上网记录里是否有？？信息，用于预防本地区的犯罪。请问这种场景下，如下哪种方案最优方案？
6. 在 Solr 上建立索引，并保存数据，在全文检索时，将所有数据全部返回。
7. 在 HBase 上存放数据，通过 HBase 的过滤特性，满足模糊匹配查询。
8. 存放数据时，在 Solr 上建立索引，在 HBase 存放完整数据？？时，通过 Solr 全文检索获取和记录关键信息， 通过关键信息在 HBase 上获取完整记录。

Answer: C

1. Fusionlnsight HD 的 Hive 中，用户自定义的 UDF 可以和 Hive 内置的 UDF 重名，这种情况下，将使用用户自定义的 UDF。

A.正确 B.错误Answer: B

1. 假设有一个应用，有 10 个表，每个表记录数为千万级别，字段数约 20 个。现使用 Redis 来缓存这 10 个表的数据，对其数据结构的设计，以下哪一项是最佳设计？
2. 采用 hash 结构，且一个表使用一个 hash KEY，表中一行记录对于 hash KEY 的一个 field。
3. 采用 hash 结构，每个表的每一行记录使用一个 hash KEY，hash KEY 的 field 对应表记录的字段，且设计 KEY 时各个表添加不同的前缀以区分。
4. 采用 string 结构，每个表的每一行记录的每个字段使用一个 KEY。
5. 采用 string 结构，每个表的每一行记录使用一个 KEY，value 为表中一行记录的所有字段？？后的值。

Answer: B

1. Fusionlnsight HD 的 Hive 应用中，存在如下场景：？？？存放文件有较高的？？效率，且大部分？？只涉及文件的一部分字母，这种场景适合采用列式文件（ORC F??）存储。

A.正确 B.错误Answer: A

1. 在 Fusionlnsight HD 中，Flume 支持以下哪些 source 类型？（多选）

A.Avro B.HBase C.Kafka D.HTTP Answer: ACD

1. Fusionlnsight HD 中，下列哪几项属于 Oozie 的 MapReduce Action 的配置项？（多选） A.name-node B.source C.mapred.mapper.class D.job-tracker Answer: ACD
2. 在 Kafka 中，Producer 可以通过配置同步参数（producer.type），保证数据按顺序发送。

A.正确 B.错误Answer: A

1. 在 Fusionlnsight HD 的客户端，执行 kinit{账号}命令是为了得到 KDC 的哪一项内容？

A.Krb5.conf B.TGT C.ST D.jaas.conf Answer: B

1. Hive 支持下列哪些表类型？（多选）

A.分区表 B.分桶表 C.倾斜表 D.分区+分桶表Answer: ABCD

1. 在 Spark 中，累加器可以实现高速并行的计数器和变量求和；在 Spark 的应用程序开发过程中，只有在 Driver

上获取此计数器的值。A. 正 确 B. 错 误A

1. Redis 的 LIST 数据结构，适用于下列哪些场景？（多选）

A.构建队列系统，例如消息队列 B.uniq 操作，例如获取某段时间所有数据的排重值C.取最新 N 个数据操作：例如对某条微博，获取最新的 10 个评论 D.模拟栈操作Answer: ACD

1. Fusionlnsight HD 中，在使用 Streaming 的 ACK 机制时，下列哪些说法是正确的？（多选）
2. 启用了 Acker 之后，Streaming 会识别发送失败的 Tuple 并自动重发，无需人为干预。
3. Acker 会标识处理超时或者处理失败的消息为 fail。
4. 从 spout 开始，形成的 Tuple 树中任何一个环节失败都会标记整棵树失败。
5. 应用需要在 spout 的 fail（）接口方法中实现消息重发逻辑。Answer: BCD

435 某大型生产企业，计划将内部的物流数据和销售数据进行大数据转型，设计部给出了数据存储方案的分析，以下哪些描述是正确的？

A. HBase 存储海量数据，并支持动态扩展，能完成物流和销售数据的存储需求。 B. Hbase 支持高校随即读取，经过合理设计能完成对物流情况的实时分析与指挥。 C. 物流数据非常稀疏，HBase 能高效地处理稀疏数据。

D. HBase 可以搭建在普通商用服务器上，构建成本相对较低。

Answer: ABCD

436.HDFS 集群跟目录下存在一个文件 text.txt，下列哪个命令可以查到此文件所存储的 DatNode 节点信息？

A. hdfs fsck /test.txt –files B. hdfs fsck /text.txt –locations

C. hdfs fsck /test.txt –blocks D. hdfs fsck /test.txt –list –corruptfileblocks Answer: B

437.HDFS 客户端以 N 副本写文件时，下列关于写过程正确的是？（多选）

A.每个 DataNode 最多存储 1 个副本。 B.支持多用户同时对同一文件的写操作。

C.数据块的第一个副本优先放在写入数据块的客户端所在的节点。 D.复制的文件块默认都存在同一机架上。Answer: AC

438.A. JDBC 接口 B. ODBC 接口 C. Python 接口 D. Ruby 接口

Answer: ABC

1. 在 Fusionlnsight HD 中，可以在哪里查看 MapReduce 应用程序的运行结果？

A. Fusionlnsight Manager 的 web 界面 B. Loader Web 界面

C. Solr web 界面 D. Yarn Web 界面

Answer: D

1. MapReduce 任务最终是在下列哪一项中被执行的？ A.NodeManager B.container C.ResourceManager D.AppMaster Answer: B
2. Yarn 中哪个角色是管理单个节点资源（CPU/Memory）的？ A.NodeManager B.Resource Manager C.Data Node D.NameNode Answer: A
3. 下来哪些 Redis 命令属于原子操作？

A.INCR（将储存的数字值增一） B.HINCRBY（为哈希表 key 中的域 field 的值加上增量 increment） C.LPUSH（对链表的 push 操作） D.以上都是

Answer: D

1. Fusionlnsigt HD 中，一个应用程序若要完成 Kerberos 安全认证，以下哪些选项是必须用到的？（多选）

A. krb5.conf 配置文件 B.在 Fusionlnsight Manager 上申请的用户账号

C.用户账号的 keytab 文件 D.jaas.conf 配置文件

Answer: ABC

1. 对 HBase 的列族，可以设置如下哪些属性？（多选） A.Versions B.TTL C.Coprocessor D.Compression Answer: ABD

445Fusionlnsight HD 中，哪些组件对外提供 SQL 或类 SQL 能力？（多选）

A.HBase B.Hive C.Spark D.Redis Answer: BC

446.在线日志查询方案中，采用了？处理来完成计算工作，整个计算过程中，需要将中间计算结果暂时存放，下列哪些组件适合存储中间计算结果？（多选）

A.HDfS B.HBase C.Kafka D.Redis Answer: CD

447.存在如下场景：每天由线上系统产生新数据 500G，需要对这些数据按天、周、月等维度做统计汇总。请问它适合用 Hive 哪种表来处理？

A.分桶表，按照天分桶 B.分区表，按天分区

C.倾斜表，倾斜字段选择日期，倾斜值随机 D.分区+分桶表，按照时间分区，每个分区 2560 个桶

Answer: B

448 在 Spark 中，假设 lines 是一个 DStream 对象，filter 语句可以过滤掉 80%的数据，针对以下两个语句说法正确的是：X:lines.filter（…）.groupByKey（…）Y:lines.groupByKey（…）.filter（…）

A. X 比 Y 的性能更高 B. X 比 Y 的性能更低 C. X 和 Y 和性能一样 D. 无法确性 X 和 Y 的性能差异

Answer: A

1. 关于 Redis 集群拓扑信息，下面描述正确的是？

A.客户端缓存有集群的拓扑信息 B.服务端缓存有集群的拓扑信息C.两者都是 D.两者多不是

Answer: C

1. 在 Kafka 中，Consumer 可以同时支持离线消息和在线消息消费，但是消费数据时都是顺序消费的。

A.正确 B.错误Answer: A

1. Fusionlnsight HD 系统的 V100R002C60 版本中，Hive 仅支持基于 MapReduce 引擎的查询服务，不支持基于

Spark 引擎的查询服务。A.正确 B. 错 误 Answer: A

1. 在 Fusionlnsight HD 的 HBase 中，如下哪个场景不会触发 Flush 操作？

A.客户端发起扫描请求时将 scan caching 设置过大 B.通过客户端指定方式发起 Flush 作

C.RegionServer 总内存超过阈值 D.Region 的 MemStore 设置超过阈值

Answer: A

1. 在 Kafka 中，下列哪条命令可以查看某个 Topic 具有多少个分区？

A.Kafka-topics.sh –describe B.Kafka-topics.sh –list C.kafka-topics.sh –alter D.kafka-topics.sh –create Answer: A

1. 在 Fusionlnsight HD 中，Flume 一个配置文件中，若存在多个 source，则 source 名称不能重复。

A.正确 B.错误Answer: A

1. HDFS 不适用于下列哪种业务场景？

A.海量数据存储 B.大量小文件存储 C.大文件存储 D.流式文件存储Answer: B

1. 一个应用需要同时和两个 Fusionlnsight 集群交互：既需要访问集群 1 的 HBase 服务，又需要访问集群 2 的

Hive 服务；那么如下哪几项操作是必须的？（多选）

1. 在集群 1 上创建本应用所需的账号，并设置此账号能访问本集群的 HBase 服务
2. 在集群 2 上创建 A 选项中的同名账号，并设置此账号能访问本集群的 Hive 服务

C.A 选项和 B 选项所创建的账号都必须属于 supergroup 组

D.集群 1 和集群 2 完成互信操作

Answer: ABD

457Fusionlnsight HD V100R002C60 中，关于 Hive 的 Python 接口类型，以下描述不正确的是？

1. 要求 Python 版本不低于 2.6.6，不高于 2.7.0
2. 客户端必须装有 setuptools，且版本不低于 5.0
3. hive\_python\_client 脚本提供了直接执行 SQL 的功能，如 hive\_python\_client ‘show tables’ D.Python 二次开发不支持安全集群

Answer: D

1. Fusionlnsight HD 中，Streaming 适用于以下哪些场景？（多选）

A.流数据监控 B.网站的实时访问统计 C.离线日志分析 D.交通流量分析Answer: ABD

1. 关于 Flume 中的数据流向，下列哪一项描述是正确的？ A.（1）数据从 source 到 channel；（2）sink 从 channel 取数据 B.（1）数据从 channle 到 source；（2）sink 从 source 取数据 C.（1）数据从 sink 到 channle；（2）source 从 channel 取数据 D.（1）数据从 channle 到 source；（2）sink 从 source 取数据 Answer: A
2. 关于 Flume 采集数据的特点，下列描述正确的是？

A.支持分布式和流式采集 B.借助 MapReduce 任务并行采集C.只能采集结构化数据 D.借助外部定时任务来触发数据采集Answer: A

1. 以下哪些措施能提高 HBase 的查询性能？（多选）

A.预分 region，使 region 分布均匀，提高并发 B.查询频繁属性放在 rowkey 前面部分

C.使用二级索引，适应更多查询场景 D.同时读取的数据存放在同一个 cell 中

1. 在 Fusionlnsight HD 安全版本进行应用开发时，可采用 keytab 文件进行安全认证。

A.正确 B.错误Answer: A

463 当一个任务被提交到 YARN 集群中，下面哪个选项会为这个任务生成一个 job id？

A.ResourceManager B.contaioner C.AppMaster D.NodeManager Answer: A

1. 在 Fusionlnsight HD 集群中，关于 kinit 操作命令，如下哪些说法是错误的？（多选）

A.只能使用人机账号。 B.只能使用机机账号。 C.一个客户端不支持多个账号同时使用。D.执行此命令得到的票据在 24 小时候会超时，需再次执行 kinit 命令去重新登录。

Answer: ABC

1. 在一个 MapReduce 应用程序中，map 函数的输出经由 MapReduce？？处理后，发送到 reduce 函数。这个处理过程是属于？？对进行排序和分组。

A.正确 B.错误

1. Fusionlnsight HD 中，可以通过哪几种方式查看 Oozie 作业的调试结果？（多选）
2. 查看 Oozie 的 JAVA API 返回结果是否符合预期
3. 通过 Hue 的 workflow 仪表板查看程序运行结果
4. 通过 Hue 的文件浏览器查看是否在 HDFS 指定目录生成了所期望的文件
5. Oozie 自带的 Web 界面也可以查看作业执行结果

Answer: ABCD

1. Fusionlnsight HD 中，关于 Ooze 的作业管理，对于已暂停的作业可以进行以下哪几项操作？（多选）

A.kill B.run C.resume D.submit Answer: AC

1. Fusionlnsight HD 中，一个完整的 Streaming CQL 应用程序至少包含以下哪些部分？（多选）

A.创建输入流 B.创建输出流 C.业务操作语句（例如 SELECT、INSERT、JION 等）

D.应用提交语句Answer: ABC

1. 在 Streaming 的应用开发中，Bolt 使用下面哪个接口发送 Tuple？ A.emit B.execute C.open D.nextTuple

Answer: B

1. 在编写 MapReduce 应用程序时，通常需要开发者实现哪两个接口？

A. map 与 combine B. reduce 与 combine C. combine 与 sort D. map 与 reduce Answer: D

1. 关于 Fusionlnsight HD 平台的 Hive 服务，其 WebHCat 开发接口，以下描述不正确的是？

A.支持基于 REST 的查询请求 B.WebHCat 的返回数据格式是 SML

C.WebHCat 基于 HTTP 和 HTTPS 协议对外提供服务 D.可以通过 WebHCat 进行建表，查询等操作

Answer: B

1. HDFS 运行时，NameNode 会将文件系统的元数据从磁盘中全部加载到内存，因此文件系统所能存储的文件总数受限于 NameNode 的内存容量。

A.正确 B.错误Answer: A

1. 假设有一个应用，需要频繁访问 Orcle 数据库中的用户表，为提高性能，引入 Redis 来缓存用户信息。对此场景，Redis 的最佳数据结构选择是以下哪项？

A.list B.set C.hash D.sorted set Answer: C

1. 在 Fusionlnsight HD 中，使用 SparkSQL 时，可以通过以下哪些方式（或工具）执行 SQL 语句？（多选）

A.JDBC B.spark-beeline C.spark-sql Answer: ABC

1. 以下哪个不是 MapReduce 的特点？

A.易于编程 B.良好扩展性 C.实时计算 D.高容错性Answer: C

476 下面哪个选项是 Yam 的核心功能？

A.存储 B.资源管理 C.搜索 D.数据传输Answer: B

1. 在 Spark 中，SparkSQL 是独立的模块，可以不依赖于 SparkCore 独立完成 SQL 语句的解析、优化和执行等操作。 A.正确 B.错误

Answer: B

1. 在 MapReduce 开发框架中，InputFormat 类的功能是？
2. 合并 map 阶段的结果中具有相同 key 值的键值对
3. 将所有 map 阶段的结果合并，得到最终的结果
4. 对 map 阶段的结果进行排序、合并
5. 将输入的数据分割成一个个的 split，并将 split 进一步拆分成<key，value>对，然后作为 map 函数的输入

Answer: D

1. HDFS 采用的是“一次写入、多次读取”的文件访问模型。所以推荐一个文件经过创建、写入和关闭之后，就不要再去修改。A.正确 B.错误

Answer: A

1. Fusionlnsight HD 中，关于 Hive 和 Sparksql，下列哪些描述是正确的？（多选）

A.两者都能对外提供 SQL 或类 SQL 能力 B.两者都能运行在 MapReduce 计算框架上

1. 两者都能通过表字段和 HBase 字段映射的方式访问 HBase 上的数据
2. 两者都能运行在 Spark 计算框架上

Answer: ACD

1. 在 HBase 应用程序运行中，应用可以在创建表的同时写入数据。

A.正确 B.错误Answer: B

1. HDFS 的应用开发中，下列哪些是 HDFS 服务支持的接口？（多选） A.BufferedOutputStream.write B.BufferedOutputStream.flush C.FileSystem.create D.FileSystem.append

Answer: ABCD

1. Spark 应用的计算逻辑会被解析或 DNG，这个解析操作由以下哪个功能模块完成？ A.Client B.Application Master C.Executor D.Driver

Answer: D

1. 对于 N（N>1）副本储存的文件，HDFS 客户端发起读取文件请求，如果所读副本节点故障，直接返回失败，不会转到其他副本节点进行读取。A.正确 B.错误

Answer: B

1. Spark Streaming 作为微批处理的流处理引擎，是将每个时间分片的数据转换为一个 RDD 中的一个分区进行计算。A.正确 B.错误

Answer: A

486 下列关于 HDFS 客户端读文件过程正确的是？（多选）

1. 优先从最近的 DataNode 下载数据。
2. 文件数据先从 DataNode 返回至 NameNode，然后从 NameNode 下致到致客户端。
3. 如果客户端和所连接的 DataNode 在读取时出现故障，那么客户端会放弃这个鼓掌节点，去连接其他副本所在的节点。
4. 支持多个客户端同时从 DataNode 上读取同一个文件数据。

Answer: ACD

487 在 Fusionlnsight HD 中，Spark 默认采用 YARN 作为集群资源管理系统。

A.正确 B.错误Answer: A

488 在 Fusionlnsight HD 中，关于 Hive 的分区（partition）功能，如下描述错误的是？

A. 分区字段要在创建表时定义 B. 分区字段只能有一个，不可以创建多级分区

C. 使用分区，可以减少某些查询的数据扫描范围，进而提高查询效率 D. 分区字段可以作为 where 字句的条件

Answer: B

1. 向 HBase 集群中增加 RegionServer 主机时，必须先停止原集群，原因是 HBase 不支持动态扩容。

A.正确 B.错误Answer: B

1. 在 Flume 中，source 功能模块的主要作用是？
2. 获取数据，并将原始数据转化成自己处理的数据对象
3. 缓存数据，根据不同的可靠性策略，将数据保存在内存或文件上C.输出数据到目的地，支持多种输出协议

D.拆分数据，根据数据的特性，将数据发送到不同的目的地Answer: A

1. 使用 Fusionlnsight HD 的 Hive 查询数据，当查询的中间结果数据量很大时，可以选择压缩中间结构数据来获取较好的性能。以下哪些是 Hive 中间结果压缩的相关参数？（多选）

A.hive.exec.compress.intermediate B.hive.intermediate.compression.codec C.hive.exec.compress.output D.hive.intermediate.compression.type Answer: ABD

1. 对于 Spark Streaming 的应用，在一个 JVM 中，同一时间只能有一个 StreamingContext 处于活跃状态。

A.正确 B.错误Answer: A

1. Fusionlnsight HD 中，使用 Streaming 的 Linux 命令行模式提交拓扑时，需要先使用一个拥有 Streaming 提交权限的用户进行 kinit 方式认证

A.正确 B.错误Answer: A

1. Fusionlnsight HD 中，Hive 客户端以下哪些参数是用来控制 reduce 并发度的？（多选） A.hive.exec.reducers.bytes.per.reducer B.hive.exec-ducers.max C.hive.exec.max.dynamic.partitions D.hive.exec.max.dynamic.partitions.pernode

Answer: AB

495 述错误的是？

A.可以从 HDFS 上直接加载到 hive 表 B.可以从 HiveServer 所在节点的本地硬盘直接加载到 hive 表

C.可以从客户端所在节点的本地硬盘直接加载到 hive 表

D.Hive 的数据加载过程了不会解析具体的文件内容，主要是文件转移的过程

Answer: C

496.在集群正常的情况下，Redis 客户端发起一次 get 调用，客户端跟服务端有（）次消息交互？

A.1 B.2 C.3 D.4 Answer: A

497Fusionlnsight HD 中，属于 Streaming 的数据分发方式有哪些？（多选）

A. Shuffle Grouping B. Field Grouping C. Local Grouping D. Direct Grouping Answer: ABCD 498.Oozie 的 workflow.xml 文件是工作流配置文件

A.正确 B.错误

Answer: A

1. 对于 HBase rowkey 的设计原则，如下描述正确的是？（多选）

A.访问权重高的属性值放在 RowKey 前面 B.访问权重高的属性值放在 Rowkey 后半部分C.离散度好的属性值放在 Rowkey 前半部分 D.离散度好的属性值放在 Rowkey 后半部分Answer: AC

1. 在 MapReduce 应用程序开发中，setMapOutputCompressorClass 类的作用是什么？

A.设 Smap 输入的压缩方法 B.设置 map 输出的传递方式C.设置 map 输出的压缩方法 D.设置 reduce 输出的压缩方法Answer: C

1. Fusionlnsight HD 支持通过哪些方式提交 Oozie 作业？（多选）

A. Sheell 命令 B.Java API C.Hue UI D.Rest API Answer: ABCD

1. 关于 Redis 特点，下面描述正确的是？（多选）

A.低时延 B.丰富的数据结构 C.支持数据持久化 D.支持 SQL 查询接口

Answer: ABC

1. Fusionlnsight HD 中，Streaming 的特点有哪些？（多选）

A.流处理 B.高性能 C.可扩展 D.高容错Answer: ABCD

1. 以下业务应用中，哪个不是 Hive 适用的场景？

A.实时的实时的在线数据分析 B.数据挖掘（用户行为分析，兴趣分区，区域展示）

C.数据汇总（每天/每周用户点击数，点击排行） D.非实时分析（日志分析，统计分析）

Answer: A

1. Spark Streaming 可以从 Kafka 接收数据并进行计算，计算结果只能存入 HDFS，不能再写回 Kafka。A.正确 B.错误

Answer: B

1. 在 Kafka 中，关于 Comsumer 说法正确的是？（多选）

A. Consumer 主要进行数据消费 B. Consumer 消费数据时需要指定 Topic

1. 多个 Consumer 配置相同的 group.id 时，这些 Consumer 消费的数据不会重复
2. 新 ConsumerAPI 消费数据不用指定 ZooKeeper 地址

Answer: ABCD

1. 在 HBase 表设计中，对于多项业务公用 rowkey，导致数据访问矛盾的情况，可以有哪些应对策略？（多选） A.折中法（rowkey 设计兼顾 A 和 B 的业务需求） B.冗余法（用空间换时间厂效率）

C.创建二级索引 D.预分 region Answer: ABC

1. Fusionlnsigt HD 中，一个部署在 Tomcat 上的应用，需访问 HBase 服务，这种场景下推荐使用机机账号。

A.正确 B.错误Answer: A

1. 在 Spark 中，关于广播变量以下说法正确的是？

A.广播变量在每 Task 中保存一份 B.广播变量在每个 Executor 中保存一份

C.广播变量是只读的 D. 广播变量通过 SparkContext.broadcast（）方法生成Answer: BCD

1. 关于 Kafka 的 Peoducer，如下说法正确的是？（多选）

A. Producer 是消息生产者 B. Producer 生产数据需要指定 Topic C.可以同时起多个 Producer 进程向同一个 Topic 进行数据发送

D.Producer 生产数据时需要先连接 ZooKeeper，而后才连接 Broker Answer: ABC

1. Fusionlnsigt HD 大数据平台提供了哪些分布式计算框架？（多选）

A.HBase B.Spark C.MapReduce D.Streaming Answer: BCD

1. 关于 Fusionlnsight HD 的 Spark，以下哪些编程语言可以用来开发 Spark 应用？（多选）

A.C B.Scala C.Java D.Python Answer: BCD

1. Fusionlnsight HD 系统中，关于 Hive 的 JDBC 接口类型，以下哪种描述是正确的？ A.组网隔离环境下，ZooKeeper 的 IP 既可以配置成业务 IP，也可以配置成管理 IP。B.Hive 与传统数据库不同，因此 JDBC 建立的连接不需要关闭。
2. 安全集群环境下，建立与 Hive 的连接前，需要使用统一认证方式进行登录。
3. 使用 JDBC 提交 SQL 时，可以在 SQL 语句后面加上分号“；”。

Answer: C

1. Flume 写入 HDFS 时，文件的生成方式有哪些？（多选）
2. 按照源文件名，直接生成文件
3. 设置文件生成时间间隔，达到时间间隔时就关闭老文件，新数据将被写入新文件
4. 设置单个文件大小，当文件大小达到所设定值时，就关闭老文件，新数据将被写入新文件
5. 设置单个文件存放 event 数量，当文件所存放 event 数量到设定值时，就关闭老文件，新数据将被写入新文件

Answer: ABCD

1. 由于 Spark 是基于内存的计算引擎，因此，一个 Spars 应用可以处理的数据量不能超过分给这个 Spark 应用的内存总和。

A. 正确 B. 错误

Answer: B

1. A.Spark B.Streaming C.MapReduce D.HDFS Answer: AC

517 关于 Kafka 的特点如下描述正确的是？（多选）

A. Kafka 是一个高吞吐、分布式、基于发布订阅的消息系统 B. Kafka 可以做到消息的持久化

C. Kafka 适用于离线和在线的消息消费场景 D. Kafka 可以保证每个 Partition 内的消息有序

Answer: ABCD

518Redis 不适用于夏眠哪些场景？（多选）

A.复杂事务操作 B.计数器应用：例如统计某个网站的登陆人数

C.一条消息被多个消费者同时消费且保证消息被所有订阅者成功消费的消息队列应用（订阅—通知） D.取最新 N 个数据：例如对某条微博，获取最新的 10 个评论

Answer: AC

1. Fusionlnsight HD 中，关于 Oozie 的作业操作命令，下列哪几项描述是正确的？（多选） A. submit 命令可以让作业的状态变为 PREP B. start 命令可以让 PREP 状态的作业变为 RUNNING C. run 命令可以让作业状态变为RUNNING D. kill 命令可以作用在 PREP/SUSPEND/RUNNING 状态的作业上

Answer: ABCD

1. Fusionlnsight HD 中，Oozie 支持 HDFS Action。

A.正确 B.错误Answer: A

1. Flume 支持实时采集数据的 source 类型是哪项？

A.taildir B.Log4j C.JMS D.Thrift Answer: A

1. Fusionlnsigt HD 中，用户想通过 HBase shell 操作来查询某个 HBase 表中的内容，这种场景下推荐管理员给这个用户分配一个机机账号。

A.正确 B.错误Answer: B

1. 在 Spark 应用开发中，以下哪项的代码能正确进行单词统计？
2. val counts=textFile.map(line=>line.split(“ ‘)).map(-rd=>(word,1)).reduceByKey(\_+\_)
3. val counts=textFile.flatMap(line=>line.split(“ “)).map(word=>(word,1)).reduceByKey(\_+\_)
4. val counts=textFile.map(line=>line.split(“ “)).map(word=>(word,1)).groupByKey()
5. val counts=textFile.flatmap(line=>line.split(“ “)).map(-rd=>(word,1)).reduceByKey() Answer: B
6. Fusionlnsight HD 中，Streaming 打包工具用于将业务代码 jar 包和其他所依赖的 jar 包等，打成一个完整的

Streaming 应用的 jar 包A.正确 B.错误Answer: A

1. 在 Streaming 的应用开发中，Bolt 使用下面哪个接口发送 Tuple？ A.emit B.execute C.open D.nextTuple

Answer: B

1. 在 Kafak 中，如下关于 Producer 发送数据的说法错误的是？（多选）
2. Producer 负责生产数据，Consumer 负责消费数据，Producer 和 Consumer 之间需要建立 Socket 连接
3. Producer 既可以发送数据到 Broker 上，又可以发送数据到 Consumer 上
4. Producer 作为消息生产者，可直接将数据写入 ZooKeeper D. Producer 连接任意一个正常的 Broker 实例就可以生产数据

Answer: ABC