1. 下列选项中无法通过大数据技术实现的是?(A)[单选题]

A商业模式发现

B.信用评估

C.商品推荐

D.运营分析

2. Fusioninsight Manager的主要功能有以下哪些？(BCD)[多选题

A.数据集成

B.系统管理

C.安全管理

D.服务治理

3.假设每个用户最低资源保障设置为yarn，scheduler，capacity,root,

QueueAminimum-user-limit-percent=24，则以下说法错误的是？(D)[单选

A第3个用户提交任务时，每个用户最多获得33.33%的资源

B.第2个用户提交住务时,每个用户最多获得50%的资源

C第4个用户提交任务时,每个用户最多获得25%的资源

D第5个用户提交任务时,每个用户最多获得20%的资源

4.华为大解决方案中平台架构包括以下哪些组成部分?(ABD)[多选题

A. Hadoop层

B. Fusiolnght Manager

C. GaussDB 200

D. Datafram

5. Spark自带的资源管理框架是?(B)[单选题

A Standal one

B Mesos

C YARN

D Docker

6.关于RDD,下列说法错误的是?(B)「[单选题

A.RDD具有血统机制( Lineage)

B.RDD默认存储在磁盘

C.RD是一个只读的,可分区的分布式数据集

D.RD是 Spark对基础数据的抽象

7.关于 Spark SQL&Hive区别与联系,下列说法正确的是?(BCD)[多选题]

A. Spark SQL依赖HiVe的元数据

B. Spark SQL的执行引擎为 Spark core.,Hive默认执行引擎为 Mapreduce

C. Spark SQL不可以使用Hive的自定义函数

D. Spark SQL兼容绝大部分Hive的语法和函数

8.在 Fusioninsight集群中, Spark主要与以下哪些组件进行交互?(ABD)[多选题]

A HDFS

B YARN

C Hive

D Zookeeper

9.关于Hive在 Fusioninsight HD中的架构描述错误的是?(A)[单选题]

A.只要有一个 Hiveserver不可用,整个Hive集群便不可用。

B. Motastore用于提供元数据服务，依赖于 DBService

C.在同一时间点,HiveServer只要一个处于 Active状态,另一个则处于 Standby状态

D. Hiveserver负责接收客户端请求，解析，执行HQL命令并返回查询结果

10.通常情况下,Hive以文本文件存储的表会以回车作为其行分隔符,在华为 Fusioninsight

Hive中,可以指定表数据的输入和输出格式处理。(A)[单选题]

A. TRUE

B FAL

11.以下关于Hbase中HMaster的功能描述哪些是正确的?(ABCD)[多选题]

A. Region负载均衡, Region分裂以及分裂后的 Region分配

B.负责建表/修改表/删除表

C.负责 Region Server的负载均衡

D. Regionserver失效后的 Region迁移

12. Nimbus在 Streaming架构中的功能包括?(ABD)[多选题

A.监控任务执行状态

B.任务调度

C.启动/关闭工作进程

D.资源分配

13. Topology在任务完成后会自动结束运行。(B[单选题

A. TRUE

B FALSE

14下面关于flink窗口的描述错误的是(AC) 多选

A.滚动窗口在时间上是重叠的

B.滚动窗口在时间上是不重叠的

C.滑动窗口之间时间点不存在重叠

D.滑动窗口之间时间点存在重叠

15.在 Fusioninsiehtaanarer界面中,对 Loader的操作不包括下列哪个选项?(A)[单选

题

A.切换 Loader主备节点

B.启动 Loader实例

C.配置 Loader参数

D.查看 Loader服务状态

16.创建 Loader作业中,可以在以下哪个步骤中设置过滤器类型?(A)[单选题

A.输入设置

B.转换

C.基本信息

D.输出

17.以下哪些数据源可以通过Loader实现与FusionInght HD的数据交换?(ABCD)[多选题]

A. NoSQL

B FTP Server

C SFTP Server

D MySQL

18. Kafka日志的清理方式有两种： delete和 compact。默认值是 delete。(A)[单选题

A. TRUE

B.FALSE

19kafka-cluster mirroring工具可以实现以下那些功能?(A)[单选题

A kafka集群数据同步方案

B. kafka单集群内数据备份

C. kafka单集群内数据恢复

D.以全部不对

20. fusioninsight HD产品中,一个典型的 kafka集群包含若干个 producer若干个

consummer和一个 zookeeper集群?(B)[单选题]

A. true

B. false

21.ZKFC进程部署在hdfs中的以下那个节点上？(AB)【多选题

A. active namenode

B standby namenode

C datanode

D.以上全部不对

22.以下关于 Kafka Partition偏移量的描述不正确的是?(D)[单选题]

A.每条消息在文件中的位置称为offset(偏移量)

B.消费者通过( offset/.partition. topic)跟踪记录

C.唯一标记一条消息

D. Offset是一个 String型字符串

23. Channel的作用类似队列,用于临时缓存进来的数据,当Sink成功地将数据发送到下

一跳的 Channel或最终目的,数据会从 Channe1移除(A)[单选题

A.TRUE

B FALSE

24. Zookeeper中的数据节点 znode分为哪几种类型?(BD)「[多选题]

A semi-persistent

B. ephemeral

C temporary

D persistent

25.HDFS联邦环境下, Name Spake(命名空间)包含以下哪些内容?(ABC)[多选题

A.目录

B.文件

C.块

D.以上全不正确

26.HDFS存储数据时，关键数据根据实际业务需要保存在具有高度可靠性的节点中,通过修改Danodo的存储策略，系统可以将数据强制保存在指定的节点组中。(A)[单选题

A.TRUE

B.FALSE

27.RDD有Transformation和 Action算子,下列属于 Action算子的是?(B)[单选题

A.map

B. saveASTexFile

C Filter

D. reducebykey

28. Spark on Yarn-client适合用于生产环境是因为可以更快的看到APP的输出。(B)[单

选题

A. TRUE

B FALSE

29.以下关于 Hive SQL基本操作描述正确的是?(D)单选题

A.创建外部表必须要指定Location信息

B.创建外部表使用 external关键字,创建普通表需要指定 internal关键字

C.加教数据到Hive时源数据必列是HDFS的一个路径

D.创建表时可以指定列分割符

30.在 Streaming中,消息可靠性级别中精确一次是通过ACK机制来实现的。(B)[单选题

A TRUE

B FALSE

31.F1ume适用于以下哪些场景的数据收集?(ABC)[多选题

A. Thrift,Avro, Syslog, Kafka等数据源上收集数据

B.本地文件数据采集

C.应用系统产生的日志采集

D.大量数据的实时数据采集

32.在 Zookeeper和Yarn的协同工作中,当 Active Resourcemanager产生故障时, Standby

Resourcemanager会从以下哪些目录中获取 Application相关信息?(B)[单选题]

A metastore

B. Statestore

C Storeage

D Warehouse

33.Zookeeper所有节点都可以处理读请求。(A[单选题

A. TRUE

B FALSE

34.HDFS的副本放置策略中，同一机架不同的服务器之间的距离是(B)[单选题

A.3

B.2

C.1

D.4

1. 下列哪些措施是为了保障数据的完整性(ABCD)[多选题]

A.元数据可靠性保证

B.重建失效数据盘的副本数据

C.安全模式

D.集群数据均衡

36. Zookeeper的 Scheme认证方式不包括以下哪项?(B)[单选题

A.digest

B,sasl

C. auth

D. world

37. Mapreduce过程中,以下属于 Shuffle机制的是?(AD)[多选题

A Copy

B Partition

C.combine

D Sort/Merge

38. Fusioninsight Manager可以对哪些项目进行健康检查?(ABCD)[多选题]

A.角色

B.主机

C.服务

D.实例

39. Fusioninsight Manager与外部管理平台对接时,支持哪些接口?(BCD)[多选题

A. VPN

B. Syslog

C. FIP

D SNMP

40.下列选项中适合 Mapreduce的场景(D)单选题

A.实时交互计算

B.迭代计算

C.流式计算

D.离线计算

41.以下关于 FusionInsight中carborData说法正确的有?(ABCD)[多选题]

A.使用 Carbon的目的是对大数据即席查询提供超快速响应。

B. Carbon使用轻量级压缩和重量级压缩的组合压缩算法压缩数据,可以减少60%-80%数据存储空间,大大节省硬件存储成本。

C.Cabonr是一种新型 Apache Hadoop本地文件格式,使用先进的列式存储，索引,压缩和

编码技术，以提高计算效率,有助于加速超过PB数量级的数据查询,可用于更换的交互查询。

D.Carbon也是一种将数据源与 Spark集成的高性能分析引擎

42.HDFS联邦机制下,各 Namenode间元数据是不共享的。(A)[单选题]

A. TRUE

B FALSE

43. Applicationmaster采用轮询的方式通过RPC协议向 Resource Manager申请和领取资源(A)[单选题

A TRUE

B FALSE

44.在YARN的任务调度中,一旦 Applicationmlaster 申请到资源后,使与对应的

Resourcemanager通信,要求它启动任务(B)[单选题

A. TRUE

B FALSE

1. 如图所示,下列哪些选项属于宽依赖(AD)[多选题

A groupby：A---B

B map：C---D

C union: D&E---F

D.join：B&F---G

46.HDFS的出现解决了需要在 HIDFS中存储大量的小文件(10B以下)。同时也要存储一些大文件（10MB以上)的混合的场景(B)[单选题

A. TRU

B FALSE

47.下列哪个命令是从HDFS下载日录/文件到本地的?(C)[单选题

A dfs -put

B dfs -cat

C dfs -get

D dfs -mkdir

48.以下选项中,属于HDFS架构关键特性的是?(ABCD)[多选题了

A.HA高可靠性

B.健壮机制

C.元数据持久化机制

D.多方式访问机制

49. Hbase的主 Master是如何选举的?C)[单选题]

A.由 Regionserver进行裁决

B. Master为双主模式,不需要进行裁决

C.通过 Zookeeper进行裁决

D.随机选举

50.关于Hive与 Hadoop其他组件的关系。以下描述错误的是?(D)[单选题]

A.Hive最终将数据存储在 HDFS中

B.Hive是 Hadoop平台的数据仓库工具

C.HQL可以通过 Mapreduce执行任务

D.Hive对 Hbase有强依赖

51.Hive不支持超时重试机制。(B)[单选题

A. TRUE

B FALSE

52. Fusioninsight中, Hiveserver将语句进行编译,解析成对应的Yarn

任务. Spark任务或者HDFS操作,从而完成数转换,分析。(B)[单选题

A. true

B False

53. Hbase的Region是由哪个服务进程来管理的?(A)[单选题]

A. HRegionserver

B Zookeeper

1. HMaster
2. DataNode

54.Streaming的处理节点Bolt中,可以完成以下哪些操作?(ABCD)[多选题

A.连接运算

B.过滤（ Filter）

C.连接数据库

D.业务处理

55.F1ume传输的可靠性是指 Flume在传输数据过程中,当下一跳的 Flume节点故障或者数

据接收异常时,可以自动切换到另外一路上继续传输?(A)[单选题

A. True

B False

56. Fusioninsight HD集群中包含了多种服务,每种服务又由若干角色组成,下面哪些是

服务的角色?(BC)[多选题

A HDFS

B. Namenod

C Datanode

D Hbase

57. Fusioninsight HD系统中使用 Streaming客户端Shell命令提交了拓扑之后,使用 Storm

UI查看发现该拓长时间没有处理数据,可能原因有?(ABCD)[多选题

A.查看客户端异常堆栈,判断是否客户端使用问题

B.查看主 Nimbus的运行日志,判断是否 Nimbus服务端异常

C.查看 Supervisor运行日志,判断是否 Supervisor异常

D.查看 Worker运行日志

58.假设HDFS在写入数据是只存2份,那么在写入过程中, HDFS Client先将数据写入

Datanode1,再将数据写入 Datanode2。(A)[单选题

A.True

B. False

59.容量调度器在进行资源分配,现有同级的2个队列Q1和2,他们的容量均为30,其中Q1已使用8，Q2已使用14,则会优先将资源分配给Q1。(A)[单选题

A. True

B. False

60. Hbase中包含的一些典型的Filter有哪些?(ABCD)[多选题

A. SingleColumValueFilter

B. FilterList

C.RowFilter

D. KeyOnlyFilter

61.Flink流式处理的数据源类型包括?(ABCD)[多选题]

A.Collections

B.JDBC

C.Socket streams

D. Files

62. server. channels.chl. transactioncapacity表示事务大小,即当前 channe支持事务

处理的事件个数,可和 source的 batchsize设置为同样大小,但不能小于 batchsize。)[单

选题

A.Trule

B. False

63.F1ume架构中,一个sink可以连接多个 channel。(B)[单选题

A.True

B. False

64.以下关于 Flink关键特性描述不正确的是?(A[单选题]

A. Sparkstreaming与 Flink相比,时延更低

B.F1ink流式处理引擎能够同时提供支持流处理和批处理应用的功能

C.与 Fusioninght HD中的 Streaming相比,FIink具有更高的吞吐量

D. checkpoint实现了 Flink的容错

65. Fusioninsight是华为企业级大数据存储,查询,分析的统一平台,能够帮助企业快速构

建海量数据信息处理系统，通过对海量信息数据实时与非实时的分析挖掘,发现全新价值点和企业商机。(A)[单选题]

A.True

B False

66. Kafka Cluster Mirroring工具可以实现以下哪项功能?(A)[单选题]

A. Kafka跨集群数据同步方式

B. Kafka单集群内数据备份

C. Kafka单集群内数据恢复

D.以上全不正确

67. Fusioninsight Tool是为技术支持工程师和维护工程师提供的一套健康检查工具,能够检

查集群相关节点,服务的健康状态中潜在的问题，并生成健康检査报告。方便技术支持工程和维护工程师快速了解系统的健康状况(A)[单题

A. TRUE

B FALSE

68. Fusioninsight产品中,关乎Kafka说法不正确的是?(C)[单选题

A. Kafka强依赖Zook eper

B. Kafka的服务端可以产生消息

CKafka的部署的实例个教不得小于2

D.Consumer作为kafka的客户端角色专门进行消息的消费

69.为了提高kafka的容错性, Kafka支持 Partition 的复制策略,以下关于 Leader Partition

和Follow Partition的描述错误的是(C)[单选题]

A. Kafka针对 Partition的复制需要选出一个 Leader。由该 Leader负责 Partition的读

写操作。其他的副本节点只是负责数据同步

B.由于 Leader Server承載了全部的请求压力。因此从集群的整体考虑, Kafka会将 Leader.均衡的分散在每个实例上,来确保数据均衡

C.一个 Kafka集群各个节点间不可能互为 Leader和 Flower

D.如果 Leader失效。那么将会有其他 fol lower来接管(成为新的 Leader)

70.Kata是ー个高吞吐.分布式,基于发布订阅的消息系统,利用kafka技术可在廉价FC

Server上搭建起大规模消息系统。(A)[单选题

A TRUE

B FALSE

71.flink是ー个批处理和流处理结合的统一计算框架,其核心是一个数据分发以及并行计

算的流数据处理引擎(A)[单选题

A True

B False

72. Hbase集群定时执行 Compaction的目的是什么?(ACD)[多选题

A.减少同一个 Regior 同个 Columnfamily下的文件数目

B.减少同一个 Region的文件数目

C.提升数据读取牲能

D.提升数据写入能力

73.下列关于 Flink barrier描述错误的是?(C)[单选题

A.一个 barrier将本周期快照的数据与下ー个周期快照的数据分隔开来

B. barrier是F1ink快照的核心

C.在插入 barrier的时候,会暂时阻断数据流

D. barrier周期性插入到数据流中,并作为数湉流的一部分随之流动

74.Flume进程级联时,以下哪些Sink

上一跳F1ume发送过来的数据?(BD)

多选题

A. Null Sink

B. thrift sink

C Hive sink

D avro sink

75. Flink可ッ车ラ以些口进行统计?(ABCD)[多选题]

A.时间窗口

B.滑动

76.关

on Insight HD Streaming的 Supervisor描述正确的是?(B)[单选题]

A. Supervisor是在 Topology中接受数据然后执行处理的组件

B. Supervisor负责接受 Nimbus分配的任务,启动和停止属于自己管理的 Worker进程

C. Supervisor负责资源分配和任务调度

D. Supervisor是运行具体处理逻辑的过程

77 Hadoop tnr yarn scheduler capacity. root. Queueafinim. m-user-limit-percent

设置为50,下面说法错误的是?(D)[单选题]

A.一个用户提交任务,可以使用 Queue的100%的资源。

B.如果 Queue中已经有2个用户的任务运行,这时第3个用户提交的任务需要等待释放资

源。

C. Queue中必须保障毎个用户至少得到50%的资源

D. ueuea中的每个用户最多只能获得50%的资

78.关于 Flume,下列说法错误的是?CD)[多选题

A. Flume级联节点之间的数据传输支持加密

B.F1ume支持多级联和多路复用

C. Source到 Channel到Sink等进程内部有加密的必要

D. Flume级联节点之间的数据传输不支持压缩

79.以下关于 Kafka logs中 segment file的说法正确的是?(ABD)[多选题]

A.通过索引信息可以快速定位 message

B.稀疏存储,即将原来的完整数据,只间隔的选择多条进行存储

C.是 index元数据全部映射到 memory,可以避免 segment file的 index数据I0磁盘操作

D.索引文件稀疏存储,可以大幅度降低 index文件元数据占用空间大小

80. Streaming主要通过 zookeeper提供以下的哪项实现事件侦听?(B)[单选题

A.分布式锁机制

B Watcher

C Checkpoint

81. Zookeeper在分布式应用

作用

哪些选项?(C)[单选题

A.选举 Master节点

B.保证各节点上数据的

C.分配集群资源

D.存储及群中

82. Hadoop集群模

据的分布情况会非常关键,用户需要根据数据分布情况,决定

集群是扩容,数据

要做负载均衡等, fusioninsight资源分布监控说法正确的

)[多选题

户快速聚焦在最关键的资源消耗上。

主业的资源分布查看界面,看到关键的资源分布情况。

C.通过

ard界面,可以查看到主机资源分不清哪个矿,例如内存占用率在50-75%的

主机列表,并提供链接跳转。

D.帮助用户迅速找到资源消耗最高的点,采取适当的措施。

83.以下关于华为大数据解决方案说法正确的是(ABCD)[多选题]

A. GAUSSDB是一款开源数据库产品

B. Farmer是数据服务框架

C. Fusion Insight Manager是一个分布式系统管理框架,管理员可以通过多接入点操控分布

式集群

D. Fusioninsight HD是基于开源大数据软件 Hadoop的增强版本

84.下图展示了文件A,B,C的存放位置。其中文件A和D具有关联性,它们的存储位置符合

Collocation同分布策略(B)[单选题]

NN

File A

A B

B C

B

C D

File B

C

D

File C

DN2

DN3

DN4

DN5

DN6 File D

A TRUE

B FALS

85.HDFS中 Namenode的主备仲裁,是由哪个组件控制的(D)[单选题

A HDFS Client

B. Nodemanager

C Resourcemanager

D Zookeeper Failover Controller

86. Spark Streaming计算基于 Stre

解成一系列短小的批处理作业(A)[单

选题

A. TRUE

B FAL

LSE

87.导入数据到Hiy表时

据合法性,只会在读取数据时候检查。(A)[单选题

A. TRUE

B FALSE

5的处理逻都在Bolt中(A)[单选题]

89.安装 sioninsight HD的 Streaming组件是, Nimbus角色要求安装几个节点?(C)[单

选题

90. Flume的 propertles. protertlses配置文件中可以配置多个 channel来传输数据(A[单

选题

A. True

B False

91.驱动型 source是f1ume周期性主动去获取数据(B)「单选题

B. False

92. Fusioninsight HD数据高可靠性体现在以下哪些方面?(ABCD)[多选题

A.硬盘热插拔

B.关键数据掉电保护

C.跨数据中心容灾

D.第三方备份系统集成

93. Fusioninsight H系统审计日志不可以记录下面哪些操作?(C)[单选题]

A.手动清除告警

B.启停服务实例

C.查询历史监控

D.除服务实例

94.下图展示了HDFS的标签存储策略,观察图, Hbase的数据会被存储到哪些数据节点上

(AB)?[多选题

Namenode

/Hbase

/Hive

TE

/Spark

2

/Flume T3

Datanode A

Datanode B

Datanode C

T2

Datanode D

Datanode E

Datanode F

T2

A Datanode A

B Datanode B

C Datanode E

D Datanode F

95. Hbase的数据文件File中一个 Keyvalue格式包含Key, yale, Timestamp, Keytype

等内容(A)[单选题]

A. TRUE

B FALSE

96.在有N个节点 Fusioninsight HD集群中部署Bse时,推荐部署()个 Master进程

()个 Regionserver进程(C)?[单选题

A.3,N

B N

D.2,2

97. Flink的数据换操作在以下哪些环节中完成(B)?【单选题

A channel

B Transformation

C sink

D source

98. Kafka消息传输保障通常有以下哪几种(ABD)?[多选题

A.仅有一次( Exactly Once)

B.最少一次( At Least Once)

C.最多三次( At Most Three Times)

D.最多一次( At Most Once)

99 Kafka Logs segment文件命名规刚

局的第一个 segment从0开始,后续每

个 segment文件名为上一个全局 partion

fset(偏移 message数)。(A)[单选

题]

A. TRUE

B. FALSE

100. Hbase的

哪些?“(D)[多选题]

A.面向列

B.高性能C

101.YAk中,“从”节点负责以下哪些工作?(BC)[多选题]

A.集群中所有资源的统一管理和分配

B.监督 container的生命周期管理

C.监控每个 Container的资源使用(内存、CPU等)情况

D.管理日志和不同应用程序用到的附属服务

102. Spark有哪些特点?(ABCD)[多选题]

A.快

B.灵

轻

103.Hive中“ Group by”指的是通过一定的规将每一个数据集划分成若干个小的数据集

然后针对若干个小的数据集进行数据分组处理(A[单选题]

B False

104. Flink适用于高并发处理数据、毫秒级时延的应用(A)[单选题

B False

105.与开源 scoop相比, loader具有哪些増强特性?(ABCD)[多选题

A.高可靠性

B.高性能

C.安全性

D.图形化

106. Fusioninsight HD loader可以将HDFS数据导出到以下哪些目标端?(ABCD)[多选题

A. orale数据库

B.SFTP服务器

C.BD2数据库

D.FTP服务器

107. Fusioninsight HD

在创建作业时,连接器有以下哪些作业?(ABCD)【多选

题]

A.配置作业与内部数源连

B.配置作业与

源连接

C.确定有转换

D.提供优化参数

导出性能

eeper集群的节点数为5节点时,请问集群的容灾能力和多少节点是等价的?(B

B.4

D.以上全不正确

109. Fusioninsight家族包含下列哪些子产品?(ABCD)「[多选题

A. Fusioninsight Farmer

B. Fusioninsight Miner

C. Fusioninsight HD

D. GAUSSDB 200

11. Fusioninsight Manager用户权限管理不支持哪个配置?(D(单选题

A.给用户配置角色

B.给用户组配置角色

C.给角色配置权限D.

给用户组配置权限

1.rn调度器分配资源的顺序,下面哪一个描述的正确?(C)[单选题]

A.任意机器一一》本地资源一一同

B.任意机器一一>同机架二一本地资源C

本地资源一一>同架>任意机器D

同机架一一>任意机器∠一>本地资源

11.以下哪个不属于 Hadoop中 Mapreduce组件的特点?(C)[单选题

A.高容错

B.良好的扩展性

C.实时计算

D.易于编程

113. Hbase的最小存储单元是 Region。(A)单选题

A. TRUE

B FALSE

114. Hbase的某张表的 Rowke

litke)9.E.a.2.请问表里面有几个 Region?(C)[单

选题]

115

据流如何

Transformation之间传输数据,数据流可以分为那些类型?

Datastream Trans forma t ion

2

로

对一流

B. distributing流

对多流

D redistributing流

11.为了保障流应用的快照存储的可靠性,快照主要存储在哪里?(OD)[单选题]

A. jobmanager的内存中

B.可靠性高的单机数据库中

C.本地文件系统中

D.hdfs中

11在 fusionlnsigh产品中,关于 kafka的 I topic.以下描述不正确的是?(B)[单选题]

A. topic的 partition数量可以创建时配置

B.每个 topic只能被分成个 partition区

C.每条发布到 kafka的消息都有一个类别,这个类别被称为 topic.也可以理解为一个存储

消息的队列

D.每个 partition在存储层面对应一个1og文件,10g文件中记录了所有的消息数据

118.Kafa集群在运行期间,直接依赖于下面哪些组件?(B)[单选题

A spark

B zookeeper

C hdfs

D hbase

11.关于 fusioninsight matasて界面hi白志收集的描述中,哪个不对?(C)[单选题]

A.可指定实例进行日志收

水定收集 metastore的日志

B可指定节点ip进行

不载某个ip的日志

C.可指定特定用户进行日

例如仅下载user用户的日志

D可指定向文少ん集26ト12010的日き

120. fusionis

于管理操作,下列错误的是?(C)[单选题

A.可对条进行

们和卸载服

常用服务隐藏或显示

务的当前状态

121.为减少写入磁盘的数据量, mapreduce支持对MOF( mapoutfile)进行压缩后再写入。

(A)[单选题]

122. Resourcemanager采用高可用方案,当 Active resourcemanager发现故障时,只能通

过内置的 zookeeper来启动 standby的 resourcemanager,将其状态切换为 active。B)[单

选题]

4. True

B False

12.Hive是一种数据仓库处理工具,使用关M的HvL语言实现数据查询功能,所有

Hive的数据都存储在 HDFS中。(A)[单选题

A. True

alse

124.在 Webhcat架构中,用户能够通过安全的HIPS协议执行以下哪些操作?D)[单选题]

A.执行 Hive DDL操作

B.运行 Mapreduce任务

C.运行 Hive HOL任务

D.以上全都正确

125. Flink的兼容性体现在以下哪些方面?(ABCD)[多选题

A.YARN能够作为F1ink集群的资源调度管理器

B.能够与 Hadoop原有的 Mappers和 Reducers混合使用

C.能够使用 Hadoop的格式化输入和输出口

D.能够从本地获取数据

126. Kafka Consumer写数据总体流程是, rconsumer连接指定 Topicpartition所在的

Leaderbroker.用主动获取方式从Kafl

消息。(A)[单选题

A. true

B false

127.图形化的健康巡检

ght Too.由 Fusioncare和 Syschecker两部分组成。

(A)[单选题

A. true

B false

128. Fusionis

界面上,当收到 Kafka磁盘容量不足告警,且该告警的原因

已经找更硬件故降)象管理员需要考虑扩容解决此问题。()[单选题

129

如到示,Fik流式理接口 I Datastream API支持的话言包括?(C)【多选题

1다

ava

B.C语言

C Scala

D Python

130.以下关于 Hbase的文件存储模块( Hbase Filestream,简称HIFS)描述正确的有?

(ABCD)[多选题

A.应用在 Fusioninsight H的上层应用

B.HFS封装了 Hbase与HDFS的接口

C.为上层应用提供文件的存储,读取,除等功能

D.HFS是 Hbase的独立模块

13若在消息处理过程中允并部分信息丢失,关闭消息可靠性处理机制的方式有以下哪些?

ACD)[多选题

A.将参数 Config. Topology\_ ACKKRS设置为0

B.将参数 Config. Topology\_ ACKKRS设置为 False

C. Spout发送消息时,使用不指定消息 messageid的接口进行发送

D.Bolt发送消息时使用 Unanchor方式发送

2.在 Flink技术架构中,以下哪项是流处尹和批处理的计算引擎?B)[单选题

A Standalone

B Runtime

C Flink Core

D. Datasteam

133.下图为F1ume数据

”号处的组件是?(C)[单选题]

Cha

Channel

Sink

A Interceptor

B Channel Processor

C Channel Selector

D.以上全不正确

134.采用 Flume传输数据过程中,为了防止因F1me进程重启而丢失数据,可以使用以下哪种

Channel类型?(BC)[多选题

A Memory Channel

B JDBC Channel

C File Channel

D HDFS Channel

135.华为 Fusioninsight HD-是国内首家符合国家金融等级保护的大数据平台,其安全性现

在以下哪些方面?(D「单选题

A.系统安全

B.权限认证

C.数据安全

D.以上全都正确

136.在华为大数据解决方案中, hadoop底层包含以下那些组件?(ABD)「[多选题

A fink

B hive

C miner

D spark

137. spark的核心模块是

题

A spark streaming

B spark core

mapreduce

D spark sql

se存储型的描述正确的是?(ABCD)[多选题]

值相同, /qualifier也相同的多个 keyvalue.也可能有多个,此时使用时间戳

值可以关联多个 value

期有时间戳,类型等关键信息

D.每一个 keyvalue都有一个 qualifier标识

139. hbase的底层数据以()的形式存在的?(A)[单选题

A keyvalue

B.列存储C.

行存储D.

实时存储

14.二级索引为s提供了按照某些列的值进行素引的能力素身先查素引表,再定位

到数据表中的位置,不用全表扫描(A)[单选题]

A. true

B false

14. colocation(同分布)文件级的同分布实现文件的快速访问,避免了因数据迁移带来的大量

网络开销(A)[单选题

142. fusioninstght h集群包含多种服务,每种服务又有若干个角色组成,下面那些是角色?

(AC)[多选题

B hbase

C datanode

143. fusininsight HD平台中,那些组件支持对列表加密?(CD)[多选题

A HDFS

B Flink

C hbase

D hive

14. Hadoop系统中,如果HDFS文件系的备

是3,那么 Mapreduce每次允许task

都是从3个有副本的机器上传输需要处理

[单选题

B False

145.YARN容量调度器的主

?(ABCD)[多选题

A.容量保证

B.动态更新賀で件

C.灵活

D

mster Mirroring工具可以实现以下哪些功能?(A)[单选

数据同步方案

B. Kafka单集群内数据备份

C. Kafka但集群内数据恢复

D.以上全不正确

147.硬件故障被认为是常态,为了解决这个问题,HDFS设计了副本机制。默认情况下,

份文件,HDFS会存(A)份?[单选题]

ABC

48.现有3个机架,有一个文件需要存3分,书中本1和副本2存放在与 client相同的

机架且不同的服务器上。根据HDFS的副本放置策略定要存放在其他机架。(B)[单选题

A. TRUE

B FALSE

149.F1ume架构中,一个 Soirce可以连接多个 Channel。(A)[单选题

B. FALSE

13. Fusionllne6的ae中保存了一张用户信息表tee,Roey为用户id

其中一类为用户呢称,现在按先后顺序往这列写 keyvalue:oo1:Li,00:Mary,0: Lily

请问scan'msg\_ table'( VERSIONS=>2)会返回哪几条数据?(B)[单选题

B.1: Lily C.001:Li,001:

Mary, 001: Lily D. 001: Ma

001: Lily

151. Kafka中 partition replication之间同步数据,从 partition的1 eader复制数据到

fo11ower需要线程( replication Fetche

Follower(一个 follower相当于 consumer

主动从1 eader批量拉取消息的,这极

(A)[单选题]

B False

152.以下哪些选项属于

安全性?(BC)[多选

A.操作系统安全加固

B.组件数据

数据完

D.用户权

ager的内存和CPU的数量,是通过下列哪些选项进行配置?(BCD)[多选题]

hyduler capacity root. Queuea maximum-capacity

anager resource cpu"vcore

C Yarn! ngdenlanager vmem-pmom-ratio

D Yarn. modemanager. resource. memory-mb

154.下列选项中,哪些是 Mapreduce一定会有的过程?(BCD)[多选题

A Combine

B Map

C Reduce

D Partition

15.下列选项中,哪些是 Spark的重要角色(ABD)[多选题]

A. Nodemanager

B. Resourcemanager

C. Datanode

D. Driver

156.以下属于 Hive SQL中DDL(数据定义语言)的是?(ABD)[多选题]

A.修改表B

删除表C

数据导入D.

建表

157.关于HIVE的描述不

单选题

A.Hive最佳使用景是

的北处理作业

B.Hive可以

模数据集上实现低延迟快速的查询

C.Hivp构建在

处理的 Hadoop之上, Hadoop通常有较高的延迟并且在作业提

的时候需腰大量的开销

司菜作过程严格遵循 Hadoop Mapreduce的作用执行模型,Hive将用户的HveL

解释器转换为 Mapreduce Hadoop集群上

158.在 Flink中, checkpoint机制能够保证应用在运行过程中出现失效时,从某一个检查

点恢复,在此过程中,流快照是根据数据流入建立的(A)[单选题

A. TRUE

B. FALSE

153Hive中的“ roup by”指的是通过一定规则将一个数据集划分成若干个小的数据集,

然后针对若干个小的数据集进行数据分组处理A「单选题

A. True

B. False

160.Hive中 UNION ALI操作符用于合并两个或多个 select语句的结果集,结果集中,不

允许有重复的值。(B)[单选题

A. True

B. False

161. Structured/ Streaming在 Output阶段可以定义不同的数据写入方式,包括以下哪些

方式?(BCD)[多选题

A General Mode

B Append Mode

C Complete Mode

D Update Mode

162.Fume数据采集的过程中,下列选项中对数据进行过滤和修饰的是?(D)[单选题

B. Channel

C. Channel selector

D. Interceptor

163.以下关于

ight Cibase的描述不正确的是?(B)[单选题

A. Chase的读

统一封装了行定义的接口,自动进行冷字段的合并和解析,

左应用程序户进行合并和解释

于 HB/se的聚簇表开发框架

了一套 Webui进行元数据定义,提供了只管医用的表设计工具,降低表设

的 java API提供一套 Hbase连接池管理的接口,内部进行连接共享,减少客户

端应用开发难度。

164. Kafka是一个高吞吐,分布式,基于发布订阅的消息系统,利用 Kafka技术可在廉价

PC Server上搭建起大规模消息系统。(A)[单选题

A. TRUE

B. FALSE

165.YARN中设置队列 Queue的最大使用资源量,需要配置哪个参数?B)[单选题

A yarn scheduler capacity. root. Queuea minimum-user-limitmpercent

B yarn. scheduler capacity. root. Queuea maximum-capacity

C yarn. scheduler capacity. root. Queuea, minimum. user-limit-factor

D yarn scheduler capacity. root.Queuea/stat

166如果某些 Containers的物理内存利用率超过了配置的内存阈值,但所有 Containers

的总内存利用率并没有超过设置的 Nodemanager内存值,那么内存使用过多的 containers

仍可以继续运行。(A)[单选题]

B

167.下图为 Spark&MADRE duce的性能对比数据,可以得出结论,与 Mapreduce计算相比, Spark

用\_的资源,获得

倍性能?(D)[单选题

数据大小

Spark

1025TB

B

102

B

72分钟

B

节点

23分钟

2100

234分钟

Cores

206

S040

19

1ATB/min

6080

Rate/node

0.67GB/min

4.27T8/min

4271B/mr

na Gr

20.7GB/m

5时排序基准规

22

A.1/10,4

B.1/8,3

1

D.1/10

下哪利场景会出发F1lush操作?(ABCD)[多选题]

新 Memstore,默认周期为1小时

文件数量达到阈值时

C. Resio中 Memstore的总大小,达到了预设的F1 ush Size國值

D. Memstore占用内存的总量和 Regionserver总内存比值超出了预设的阈值大小

169. Fusioninsight HD系统中, Hbase支持动态扩展列。(A)[单选题

B Fal

170.以下哪些选项属于Hive的数据存储模型?(E)[单选题

A.桶

B.数据库

C.分区D

表

E.以上全都正确

171.Hive中的这条命令” ALTER TABLE employee ADD columns( columnl string).”是

什么含义?(C)[单选题

A.创建表

B.刑除表

C.添加列

D.修改文件格式

172.下列关于 Sparksereaming:和 Streaming比较说法不正确的是?(B)【单选题]

A. Sparkstreaming

个微批处理框架,事件需要积累到一定量时才进行处理

B. Streaming的执行逻辑是即时启动,运行完后再回收

C. Sparkstreaming的吞吐量大约是 Streaming的2-5倍

D. Spaalstreaming事件处理时延比 Streaming更高

173.以下关于F1ink与其它组件交互说法正确的是?(ABCD)[多选题

A. Flink任务的运行依赖Yarn来进行资源调度管理

B.F1ink的 checkpoint的实现依赖于

C. Flink可以将接收的组件发送给K

D. Flink在HDFS文件系统中读写数

174.创建 Loader作业时

哪个步骤中设置Map数?A)[单选题

A.输出

B.输入设置

C.转换

D.基本信息

客

持事提供较弱的顺序保证,可以连接任何数量的 Source和Sink。A)[单选

176. Kafka集群中, Kafka服务端的角色是?(A)[单选题

A Broker

B Consumer

C Zookeeper

D Producer

7. Kafka Producer读数据总体流程是, Producer连接任意存活的 Broker,请求制定 Topic、

Partition的 Leader元数据信息,然后直接与对应的 Broker直接连接,发布数据。B)[单选

题

rule

1.下图所示的Map阶段,是描述哪个过程?石)[单选题

ap Output

<Hello. 1>

Map Output

<World, 1s

<Bye. 1-

<Hello, 1>

<World. 1>

World.-

-Bye. 1>

<Hello, 1>

<Hadoop. 1>

?

<Hadoop, 1>

<Bye.15

<Hadoop. 1>

<Hello, 1>

<Bye.1>

<Hadoop, 1>

c Hello. 1>

< Bye.1-

≤Had

combine

179. Spark根据RDD的依赖关系来划分 Stage,调度器从DAG图末端出发,逆向遍历整个依

赖关系链,遇到窄依赖就断开,遇到宽依赖就将其加入当前 Stage。(B)[单选题

A. TRUE

B. FALSE

180.关于 Dataset,下列说法不正确的是?(D)[单选题

A. Dataset不需要反序列化就可执行大部分操作

B. Dataset是一个由特定域的对象组成的强类型集合

C. Dataset与RDD高度类似,性能比RDD好

D. Dataset执行sort, filter, shuff1e登操作需要进行反序列化

181. Hbase中 penscanner的过程,会创建两种不同的 Scanner来读取H-File和 Memstore

的数据,HFi1e对应的 Scan

Storefilescanner, Memstore对应的 Scanner为

Memstorescanner。(A)[单选题

A. TRUE

B. FALSE

182.执行 Hbase取数据业务,需要读取哪几部分数据?(BD)[多选题

A. Master

B. Hfile

D. Memstore

183.使用 Hbase客户端批量写入10条数据,某个 Hregionserver节点上包含该表的2个

Region,分别为A和B,10条数据中有6条属于A,4条属于B,请问写入这10条数据需要向

该 Hregion Server发送几次RPC请求?(D)[单选题]

A.10

B.6

184.Hive不适用于以下哪个场

单选

A.非实时分析,例如日

析

B.数据挖掘,例如用户

析,区域展示

C.数据汇总,例如母天,每

击数,点击排行

D.实时在线数

描述正确的是(C)[单选题

不可再修改表名

可再增加新列

需要制定 external关键字

,不可再修改列名

186. Fusioninsight HD系统中Hive支持的存储格式包括?(ABC)[多选题

A. Textfile

B. Sequencefile

C. RCFILE

D. Hfile

187.在 Flink的运行流程中,负责申请资源的角色是?(B)工单选题

A. Resourcemanager

B. Jobmanager

Client

Taskmanager

188.在 Fuisoninsight HD中,创建 Loader作业的进行数据转换的正确步骤是?(A)[单选

A.输入设置,转换,输出

抽取,转换,输出

C.加载,转换,输出

D.加载,转换,抽取

189.F1ume用子收集数据,其传输的数据基本单位是?(C)【单选题

A. Split

B. Block

Event

D. Packet

190. Fusionninsiget HD系统中, Flume数据流在节点内不需要经过哪个组件?(C)[单选

题

A. Sourc

B. Sink

C. Topic

anne

191.传统数据处理的数据

?(D)单选题]

D. GB

哪个命令是除文件的?(C)[单选题]

193.下图为HDFS分级存储的配置,如果 Block的副本数为4,则以下说法错误的是?(C)

[单选题

策略ID

名称

B1ock放置位置」备选存储策略副本的存储策略

(副本书)

LAZY\_PERSITST RAMDISK: 1, DIS DISK

DISK

K:n-1

A.第1个 Block副本将写入 RAM DISK,其余副本写入DISK

B.如果是第1个副本之外的其他副本写入失败,则尝试写入“副本的备选存储策略”指定的

存储策略

C.如果第1个 Block副本写入 RAM DISK失败,则尝试写入”备选存储策略”的存储类型

D.4个 Block都将写入 RAM DISK

194Eado平台中,要查看Yx眼务中ー个 Yapplication的信息,通常需要使用什么命令?

(D)[单选题]

A. container

jar

C. application-attefpt

D. Application

195. Spark适用于以下哪些场景?(ABCD)[多选题]

A.交互式查询

B.实时流处理

C.批处理

D.图计算

196.下图展示了 Structured Streaming的计算模型。通过观察可以得出最终T3的计算结

果是(B)[单选题

Ow eat

3

ne

输入无边

界表数据

「=1的计算结果

T=2的计算结果1

T=3的计算结果

T-1的计算墙果T-2的计算

2

计算结果

dog 3

3

B. Cat 2, dog 4, owl 2

C. Cat 2, dog 3, owl

D Cat 1, cat 1, dog 2, dog 2, owl 2

197. Spark Streaming容错机制是指RDD中任意的 Partition出错,都可以根据其父RDD重

新计算生成,如果父RDD丢失,则需要去磁盘中查找原始数据。(A)[単选题

A. TRUE

B. FALSE

198. Hbase元数据 Meta Region/路面信息保存在哪里?(B)[单选题

A Zookeeper

B.Meta表

C Haster

D.Root表

19. Hbase的B1 6omf liter是用来顾虑数据的?(B)[单选题

A. TRUE

B. FALSE

200.执行 Hbase读数据业务,需要读取哪几部分数据?(CD)[单选题

B. Master

C. Hfile

D. Memstore

201. Fusioninsight中Hive包含2

rver和 Metastore'?(A)[单选题

A. TRUE

B. FALSE

202. Colocation同分布

分下买现文件的快速访问,避免了因数据迁移带来的大

量网络开销(A)[单选题

A. TRUE

B. FALSE

リe与传统据仓库的对比,以下描述错误的是?(D)[单选题

数据存储独位于数据存储之外,从而解耦合元数据和数据,灵活性高,而传统数据

,灵活性低

B.Hiす身HDFS存储,理论上存储量可无限扩展,而传统数据仓库存储量会有上限

C.由子Ave的数据存储在 HDFS中,所以可以保证数据的高容错,高可靠

D.由于Hive基于大数据平台,所以查询效率比传统数据仓库快

204.YARN的基于标准调度,是对下列选项中的哪个进行标签化?(C)[单选题]

A. Appmaster

B. Resourcemanager

C. Nodemanager

D. Container

205.对于Hive中关于普通表和外部表描述不正确的是?(C)て单选题

默认创建普通表

B.除外部表时,只除外部表数据而不型元数据

C.外部实质是将已存在于HDFS上的文件路径跟表关联起来

D.删除普通表时,元数据和数据同时被删除

206.在 Fusioninsight产品中,关于的建 Kafka的 Topic,以下哪些描述是正确的?(C)[单

选题

A.在创建 Kafka的 Topic时,必须设置 Partition个数

B.在创建 Kafka的opir时,必须设置 Partition副本个数

C.设置多副本可以增强 Kafka服务的容灾能力

D.以上全都正确

.2 Zookeeper的增强性包括在审计日志中添加 ephemeral node被刑除的审计日志(A)[单

选题

A. TRUE

B. Falst

208.下列选项中,关于 Zookeeper可靠性含义说法正确的是?(D)[单选题

A.可靠性通过主备部署模式实现

B.可靠性是指更新更新只能成功或

有中间状态

C.可靠性是指无论哪个 server,对

同一个视图

D.可靠性是指一个消息被一个 Serve

所有的 Server接受

209.华为 Fusioninsigh

案例都有哪些?(ABCD)[多选题

A.数字政府

B.智慧园区

C.智慧交通

D.金融

210

ght Hager会定时备份哪些元数据?(ABCD)[多选题]

211.YARN中默认的资源调度器是?(B)[单选题

A.FIFO调度器

B.容量调度器

C.Fair调度器

D.以上全不是

212.假设在写入数据时只存2份,那么写入过程中, HDFSCLIEHT/编将数据写入 Datanodel,

再将数据写入 Datanode2(B)[单选题]

A. TRUE

B. FALS

21 Yarn-client和 Yarn-cluster主要区是 Application Master进程的区别別(A)【单选

题」

A. TRUE

B. FALSE

214.部署 Fusioninsight HD时,同一集群内的 Flume Server节点建议至少部署几个?(D)

[单选题

215.以下哪些是 Kafka实际的应用场景?(ABC)[多选题]

A.网站活性跟踪

B.日志收集

C.聚合统计系统运营数据(监控数据)

D.资源管理

216. Fusioninsight HD系统中,下面

查看到 Loader作业执行的结果(C)[单

选题

A.通过 Loader UI界面查看

B.通过YARN任务管理

C.通过 Nodemanager查

D.通过 Manager的警

217.如图

a逍息消费者 Consumer读取消息的描述,错误的有?(D)[单选

题

么的描。精的有

A. Consumer使用 Offset来记录读取的位置

B.图片中的蓝色框为 Kafka的一个 Topic,即可以理解为一个队列,每个格子代表一条消

C.生产者产生的消息逐条放到 Topic的末尾

D.消费者从右至左顺序读取消息

218.在大数据时代,企业所面临的挑战有以下哪些?(ABCD)[多选题]

企业各部门间数据分散,相同数据在各内部存储格式不一致

数据结构多样化。

C.竞争对手的技术进步。

D.数据存在噪音、缺失、存储类型不规范等问题,需要进行大量的数据预处理工作。

219. Fusioninsight HD中 Hbase默认使用什么组件作为其底层文件存储系统?(C)[单选

File

B Kafka

HDFS

D Memory

220.在很多小文件场景下, Spark会起很多Task,当SL逻辑中存在 Shuffle操作时,会大

大增加hash分桶数,严重影响性能。 Fusioninsight中,针对小文件的场景通常采用()

算子,来对 Table中的小文件生成的 partition进行合并,减少 partition数,从而避免在

shuffle的时候,生成过多的hash分桶,提升性能?(D)[单选题

BCD

coalesce

connect

221.下列关于 Worker(

ector(线程)、Task(任务)说法正确的是?(ABD

多选题

A.每个 Executor(线程人

Task(任务)

B.每个 Worket运行多个小et6r(线程)

C.每个 Workst能为个拓扑运行 Executor(线程)

D.每个 Execu

运行不同组件( Spout或Bolt)的Task(任务)

于 Kafka的描述错误的有?(C)[单选题

a、Java\_语言实现

运营数据处理管道的基础

C.由

Hadoop开发,并于2011年开源

D.具有消息持久化、高吞吐、实时等特性

223.查看 kafka某 Topic的 partition详细信息时,使用如下那个命令?(D)[单选题]

A bin/kakfa-topics sh --create

B bin/kakfa-topics sh --delete

C bin/kakfa-topics. sh--list

D bin/kakfa-topics sh --describe

24.用户权限管理基于角色的访问控制(RBAC),提供可视化的多组统一的集群中用户权限管

理。(A)[单选题

A. true

B false

22对高价值高度聚合的信息和知识的批次处理是大数据行业主要商业诉求(A)[单选题

A true

B. false

26. Mapreduce过程史、默认情况下,广个分片就是一个块,也是一个 maptask..(A)[单选

题

A. true

B false

22. Fusioninsight'家族包含以下哪些子产品(ABCD)?[多选题]

A. GAUSSDB 200

B. Fusioninsight Miner

C. Fusioninsight HD

D. Fusioninsight Farmer

22.以下关于 Zookeeper关键特性中的原子说法正确的是(B)?[单选题

A.客户端发送的更新会按照他们被发的进行应用

B.更新只能全部完成或失败,不会部

C.一条消息被一个 server接收,将截所

D.集群中无论哪台服务器,对外示均是同

229. Flume支持监控并传

,可实现实时数据传输,以下描述的是哪一类

Source(D)?[单选

ttp sourc

C. syslog Sour

D

director source

项不属手创建 Loader作业时必选项(A)?[单选题

B.名称

C.连接

D.类型

231.为了保证流应用的快照存储的可靠性,快照主要存储在(C)?[单选题]A.

本地文件系统中

B. Jobmanager的内存中

C.HDFS中

D.可靠性高的单机数据库中

232. Fusioninsight HD系统使用 Streaming客户端 Shell命令查看拓扑或者提交拓扑失败,

以下哪些定位手段是正确的(ABCD)?[多选题

A.查看 Supervisor运行日志,判断是否 Superisor异

B.查看 Worker运行日志

C.查看主 Nimbus的运行日志,判断是否 Nimbus服务端异常

D.查看客户端异常堆栈,判断是否客户端使用问题

233.如果需要由数据生产者决定数据发送给目标Bolt的某一个确定的Task,应选择以下哪

种消息发布策略(C)?「单选题]

A.局部字段分组

B.广播分组

C.直接分组D

全局分组

234.华为大数据产品 Fusion Insight HD中的 Streaming具有哪些关键特性(ABC)?[多选

A.容灾能力

B.灵活性

C.消息可靠性

D.可伸缩性

235.以下关于 Hive SQ1基本操作描

)?[单选题

A.加载数据到Hve时源数据必须是

B.创建外部表必须要指定10ccig信息

C.创建外部表使用 extel

创建普通表需要指定 internal关键字

D.创建表示可以指定列分階

236. Flink采

oint

用程序运行中的容错性。(A)[单选题

A. TRUE

B. FALSE

支持的时间操作类型包括以下哪些选项?(BD)[多选题

B.处到

C.采

D.事件时间

238. Flink中的()接口用于流数据处理,()接口用于批处理?(C)[单选题]

A. Databatch API, Datastream API

B. Stream API, Batch API

C. Datastream API, Dataset API

D. Batch API, Stream API

239.下列属于 Flume数据监控的指标是?(ACD)[多选题

A.Sink写入数据量

B. Datanode数量

C. Channel缓存数据量

D. Source接受数据量

240.调用 Zookeeper客户端命委中包즘哪些信息?(ABCD)[多选题

A.Ip地址

端口号

C.服务器名称

D.用户名

24.以下关于 Zookeeper的 Leader节点在收到数据变更请求后的读写流程说法正确的是?

(D)[单选题

A.仅写入内存

B.同时写入磁盘和内存

C.先写入内存再写入磁盘

D.先写磁盘再写内存

242.以下关于 Zookeeper的 Leader选举说法正确的是?(AB)[多选题

A.当实例n为奇数时,假定n=2x+1,则成为 leader节点需要x+1票

B. Zookeeper选举1 eader时,需要

票数

C.当实例数为8时,则成为1 eader

容灾能力为4

D.当实例数n为奇数时,假定n=2x+1,

er节点需要x票

243.传统数据处理的性能

C)[多远题]

A.数据存储成本高

B.流式数据处理性不

C.扩展能力有

D.批量数据妙失

24罗活用子业场腰?(BD)【多选题

数据访

件存储

与访问

D.随枺

245.HDFS中的主备仲裁,是由哪个组件控制的?(A)[单选题

A. Zooleeper Failover Controller

B. Nodemanager

C. Resourcemanager

HDFS Client

246.在 Mapreduce编程中,代码不仅要描述做什么,还要描述具体怎么做。(B)[单选题

A. TRUE

B. FALSE

247.在 Loader历史作业记录中,可以查看以不哪些内容?(ABCD)[多选题

A.作业状态

B.脏数据链接

C.错误行/文件数量D

作业开始/运行时间

248.关于 Kafka磁盘容量不足的告警,对于可能的原因以下分析不正确的是?(C)[单选题]A.

业务规划不合理,导致数据分配不均,使部分磁盘达到使用率上限

B.数据保存时间配置过长,数据累计达到磁盘使用率上限

C. Broker节点故障导致

D.用于存储 Kafka数据的磁盘配置(如磁盘数目,磁盘大小等),无法满足当前业务数据流

量,导致磁盘使用率达到上限

249. Fusioninsight HD中 Loader从SFTP服务器导入文件时,不需要做编码转换和数据转

换且速度最快的文件类型是以下哪项?(B)[单选题

A. graph-file

lnary-r1

C. text-file

250. Zookeeper可以为 Fusion Insight

提供分布式管理支持?(ACD)[多选

A H

oader

D Spark

较

251.以下关子F

多租户管理的描述正确的有?(ABCD)[多选题]

A.可规化的多级租户里,与企业组织结构相匹配,简化系统资源分配与管理

理和户资源计算资源(CPU/内存/I0)、存储资源(HDFS)、服务资源( Hbase…)

目结构相匹配的多级的租户模型,不同部门对应不同的租户,按需动态増删租户D

基

户资源隔离,为租户SLA保驾护航

252.以下关于 Kafka Partition说法正确的有?(ABCD)[多选题

A.引入 Partition机制,保证了 Kafka的高吞吐能力

B.每个 Partition都是有序且不可变的消息队列

C. Partition数量决定了每个 consumer group中并发消费者的最大数量

D.每个 Partition在存储层面对应一个1og文件

253.hive在1oad是不检索数据是否符合 schema的,hive遵循的是 schema on read(读时

模式)只有在读时模式的时候才检查hive的数据字段, shena./A)[单选题]

A. TRUE

B FALSE

254. Loader可以实现以下哪些转换规则(ABCD)?[多选题]

A.长整型时间转换

B.增量转换

C.拼接转换

D.空值转换

255. Fusioninsight

系统中使用 Streaming客户端shel1命令提交拓扑之后,使用 Storm

U查看发现该拓长时间没有处理数培,可能原因(ABCD)[多选题

A.拓扑业务存在年错误,提交之后无法正常运行

Bsup足oaoy中接收数据然后执行处理的组件

C.拓扑过于复杂或者并发太大,导致 worker启动时间过长,超过 supervisor的等待

时间

D. Supervisor的 slots资源被耗尽,拓扑提交上去后分不到s1ot去启动 worker进程

256. Fusioninsight Manager界面显示Hive服务状态为Bad时,可能的原因有哪些?(ABD)

多选题]

A. Dbservice服务不可用

B.HDFS服务不可用

C. Hbase服务不可用

D. Metastoro实例不可用

257.以下哪些选项属于 Fusionins

忧点?(E)[单选题]

A.稳定和高性能

B.长稳测试

C.SL兼容性

D.支持数据插入,是新和

E.以上全都正

258.下关于F

formation的说法正确的是?(ABD)[多选题]

A.、可叭读 window定时间窗口

操件是对每个元素执行 boolean函数

对文本进行切分

D.ke)源头数据按照key进行分组,以保证同一个key的元数据分到同样的组中

259. Fusioninsight Manager对服务的管理操作,下面说法错误的是?(A)[单选题

A.可设置不常用的服务隐藏显示

B.可以添加和卸载服务

C.可对服务进行启停重启操作

D.可查看服务的当前状态

260.下面关于 Zookeeper特性的描述错误的是?(B)[单选题

A.消息更新只能成功或者失败,没有中间状态

B. Zookeeper节点数必须为奇数个

C.一条消息要被超过半数的 Server接收,它将可以成功写入磁盘

D.客户端所发送的更新会按照它们被发送的顺序进行应用

26Fue的数据流可以根据 headers,的信息发送到不同的 Channel中、()[单选题

B. False

262. Hbase中数据存储的文件格式是什么?(C)[单选题

A. Hlog

B. Textfile

HFII

equencerlle

263. Hbase中的 Region是由哪个服务进程来管理的?(A)[单选题]

A. Hregionserver

B Zookeeper

C Master

D Datanode

264.在 Flink中, checkpoint机制能

明在运行过程中出现失败时,从某一个检查

点恢复,在此过程中,流快照是根据

次创建的。(A)[单选题

A TRUE B FALSE

265.以下关于 Fusionin:

全可靠性,说法正确的有?(ABCD)[多选题

A.阻止外部攻击者通过管

务数据

B. Fusioninsight支网

集群业务平面、集群管理平面和集群外维护网络。彼

此之间实施物理

C.避免业务平

阻赛集群管理通道

D.网络平面

务带宽抢占,相互干扰

ioninsight/产品中,关于 Kafka组件说法正确的是?(ABCD)[多选题

Tpic时,副本数不得大于当前存活的 Broker实例个数,否则创建 Topic将会失

B. Kafka

oducer发送消息时,可以指定该消息被哪个 Consumer消费

C. Kafka会将元数据信息存放到 Zookeeper上

D. Kafka安装完成后就不能再配置数据存放目录

267. Kafka所有消息都会被持久化到硬盘中,同时 Kafka通过对 Topic Partition设置

Replication来保障数据可靠(A)[单选题

A. TRUE

B FALSE

28在 Fusioninsight产品中,关于 Kafka Topic,,以下描不正确的有?(A)「[单选题

A.每个 Topic只能被分成一个 partition(区)

B. Topic的 partition数量可以在创建时配置

C.每个 Partition的存储层面对应一个108文件,18文件中记录了所有的信息数据

D.每条发布到Kaa的消息都有一个类别,这个別被称为Topi也可以理解为一个存储

消息的队列