1.Hadoop有哪些部署方式？（    ）

集群模式√

仅主机模式

单例模式√

伪分布模式√

2.下面哪些是hadoop生态圈的应用？（    ）

Flume√

Hive√

Sqoop√

pig√

3.下列哪个属性是mapred-site.xml中的配置？（  ）

fs.defaultFS

mapreduce.framework.name√

yarn.resourcemanager.address

dfs.replication

4.关于Hadoop SSH免密码登录说法错误的是（  ）

在SSH安全协议的原理中， 是一种非对称加密与对称加密算法的结合

生成公钥和私钥的只要输入~/.ssh/ssh-genkey –t rsa命令，四个回车即可

在本机搭建好SSH免密码登陆，就可以随意的访问其它节点服务√

生成SSH密钥之后，只需要输入ssh-copy-id localhost即可免密码登陆

5.关于服务启动正确的是（  ）

sbin/hadoop-daemon.sh start-yarn.sh

sbin/hadoop-daemon.sh start resourcemanager√

sbin/hadoop-daemon.sh namenode start

sbin/start-hdfs.sh

6.以下关于Hadoop说法正确的是（   ）

A、Hadoop支持数据的随机读写

B、因为HDFS中采用了备份的机制， 所以不存在单点故障

C、Hadoop默认调度器策略为FIFO

D、NameNode负责管理metadata，client端每次读写请求， NameNode都需要从磁盘中读取或者写入metadata信息

A

B

C√

D

7.关于HDFS服务进程的描述不正确的是（  ）

A、datanode之间都是独立的，相互之间不会有通信

B、存储客户端上传的数据的数据块

C、一个datanode上存储的所有数据块可以有相同的

D、HDFS的namenode保存了一个文件包括哪些数据块，分布在哪些数据节点上，这些信息也存储在硬盘上

B

C

A

D√

8.下列关于SecondaryNameNode不正确的是（  ）

SecondaryNameNode应与NameNode部署到一个节点

它是NameNode的热备

Secondary namenode就是namenode出现问题时的备用节点

它的目的是帮助NameNode合并编制日志，减少NameNode启动时间√

9.关于Hadoop下列说法错误的是（   ）

HDFS采用了Master/Slave的架构模型

Datanode执行比如打开、关闭、重命名文件操作√

Namenode负责维护文件系统的命名空间

HDFS暴露了文件系统的命名空间、允许用户以文件的形式在上面存储数据

10.关于HDFS Shell操作描述错误的是（    ）

A、Hadoop fs -cat /iflytek/text.txt | tail -5 返回前几行的样本数据

B、Hadoop fs -cat /iflytek/text.txt | shuf -n 5 随机返回指定行数的样本数据

C、[-moveFromLocal <localsrc> ... <dst>] 与put相同，但是文件上传之后会从本地文件系统中移除

D、[-copyFromLocal <localsrc> ... <dst>] 与put相同，从本地文件系统上传文件到HDFS中

C

D

B

A√

11.关于Hadoop Shell操作不正确的是（   ）

bin/Hadoop jar [-target.jar] 表示打包运行MapReduce程序

bin/Hadoop fs –put [-txt] / 表示从本地上传文件到hdfs上

bin/hdfs dfs –put [-txt] / 表示从本地上传文件到hdfs上

bin/hdfs dfs –ls –R / 表示只查看hdfs上根目录下的目录与文件√

12.关于HDFS API的说法错误的是（   ）

FileSystem实现了Closeable接口

FileSystem仅适用于分布式文件系统√

FileSystem是一个实现了文件系统的抽象类

Hadoop通过使用FileSystem类来提供文件系统操作

13.下列关于Hadoop API 说法错误的是（   ）

FileStatus对象存储文件和目录的元数据

Configuration 类的默认实例化方法是以HDFS系统的资源配置为基础

FSDataInputStream是java.io.DataInputStream的子类

Hadoop的文件API不是通用的，只用于HDFS文件系统√

14.关于MapReduce过程各个角色的作用说法不正确的是（   ）

客户端提交一个MR的jar包给JobTracker，再由JobTracker分给TaskTracke

TaskTracker负责执行任务，负责启动、跟踪任务执行、访问任务状态和日志

JobClient负责提交作业

JobTracker负责接收用户提交的作业，负责启动、跟踪任务执行√

15.关于MapReduce执行过程说法错误的是（   ）

MapReduce执行过程分为Mapper过程与Reducer过程

合并、排序、Shuffle是在Mapper输出之后，Reducer输出之前完成的

正常执行MapReduce任务时，Map函数和reduce函数都会被调用执行√

Mapper过程与Reducer过程之间，会有shuffle过程

16.有关MapReduce的输入输出，说法错误的是（    ）

A、链接多个MapReduce作业时，序列文件是首选格式

B、FileInputFormat中实现的getSplits()可以把输入数据划分为分片，分片数目和大小任意定义

C、想完全禁止输出，可以使用NullOutputFormat

D、每个reduce需将它的输出写入自己的文件中，输出无需分

B√

C

A

D

17.默认的MapReduce输入格式是（    ）

TextInputFormat√

KeyValueTextInputFormat

NLineInputFormat

SequenceFileInputFormat

18.关于Writable接口的说法错误的是（   ）

A、 Writable接口是根据 DataInput 和 DataOutput 实现的简单、有效的序列化对象

B、 MapReduce的任意key必须实现WritableComparable接口

C、 MapReduce的任意Value不必实现Writable接口

D、 Hadoop使用自己的序列化格式Writables接口，具有紧凑，快速的特点

D

A

C√

B

19.以下关于Partititoner编程的说法错误的是（    ）

A、 Partitioner<k,v>是partitioner的基类，如果需要定制partitioner也需要继承该类。

B、 Partritioner的作用是对Reduce产生的结果进行分片。

C、 Partitioner组件是负责给mapTask节点输出的数据进行分发。

D、 Partitioner用来控制Map输出的中间结果键值对的划分，分区总数与作业的Reduce任务的数量是一样的。

B√

D

C

A

20.关于Hive的数据存储说法不正确的是（）？

HiveSQL并不支持Having子句和子查询√

Hive 将元数据存储在数据库中(metastore)，目前只支持MySQL、derby

生成的查询计划存储在 HDFS 中，并在随后有 MapReduce 调用执行

Hive 的数据存储在 HDFS 中，大部分的查询由 MapReduce 完成

21.关于Hive与Hadoop之间说法正确的是（）？

select \* from table，有 MR 任务

Hive所有的数据都是存储在 Hadoop 中，查询计划被转化为 MapReduce 任务√

select count（\*）from table，没有 MR 任务

Hive构建在 Hadoop 之上， HQL 中对查询语句的解释、优化、生成查询计划是由 Hive√

22.关于Hive搭建说法正确的是（）？

目前Hive主要是支持Oracle和MySQL数据库

如果使用自带的数据库，需要进行数据库的相关配置√

安装Hive时，首先确保Hadoop已经安装完毕并且能正确使用√

hive-evn.sh文件名是固定的，不得随意修改√

23.关于元数据信息metastore说法正确的是（多选题）？

metastore是hive元数据的集中存放地√

使用MySQL作为外置存储引擎，多用户同时访问√

用内嵌的数据库，一次只能打开一个hive会话√

在做hive操作命令时，先把Hadoop服务启动起来，否则，操作dml语句时，报错√

24.关于常见查询语句表达不正确的是（）？

select \* from stu limit 3：查询所有数据的前三条

select \* from sales order by amount desc limit N ：

Hive中的limit可以有多个参数√

select \* from stu where name like '%ang%'：模糊查询

25.下列关于内部表说法正确的是（）？

复制一张表结构，但不导入数据:create table carcp like car;√

删除表的时候，对跟随数据也随着丢失:drop table db\_name;√

创建表的时候指定格式:create table t1(id int)row format delim√

表的修改:alter table target\_tab add columns (id int);√

26.下列关于外部表说法正确的是（）？

外部表和内部表在元数据的组织上是相同的，而实际数据的存储则有较大的差异√

当删除外部表的时候，metastore中元数据信息会丢失，但是外部目录中指向数据不会丢失√

metastore中的描述信息可以作为外部数据信息的一个引用√

当删除一个外部表 时，仅删除该链接√

27.下列关于分区表说法正确的是（）？

分区表和内部表在元数据的组织上是相同的，而实际数据的存储则有较大的差异

当删除一个分区表时，不会删除实际数据

分区表跟内部表删除数据原理相似，即删除元数据信息metastore时，实际数据也会随着丢失√

Partition 对应于数据库的 Partition 列的密集索引√

28.下列关于Hive桶表说法正确的是（）？

Hive桶表是对数据进行哈希取值，然后放到不同文件中存储√

Hive桶表可以进行分桶抽样√

数据加载到桶表时，会对字段取哈希值，然后与桶的数量取模√

Hive桶表主要用于小样本的取样

29.下列关于分布式数据库HBase说法正确的是（）？（多选题）

A、HBase 不是一个关系型数据库，其设计目标是用来解决关系型数据库在处理海量数据时的理论和实现上的局限性

B、HBase是在海量半结构化、非结构化的背景下所产生的

C、HBase的计算只能是基于MapReduce计算框架的

D、HBase是开源的并基于Google Big Table白皮书

C√

B√

D√

A√

30.HBase依靠( )存储底层数据？（单选题）

Memory

NoSQL

MySQL

HDFS√

31.下列关于HBase特点说法错误的是？（单选题）

对于为null的列并不占用存储空间

支持多条件查询√

毫秒级查询

高扩展性

32.下列关于行键（Row Key）说法错误的是（）？（单选题）

存储数据时，数据按照行键的字典排序存储，经常一起读取的行存储放到一起。

行键可以是任意字符串，保存的是字节数组。

访问HBase表中的行，可以通过单个行键或全表扫描来访问

行键是用来检索记录的主键，与MySQL数据库一样。√

33.下列关于HBase表中单元格（Cell）说法错误的是（）？（单选题）

HBase中通过row和columns确定的为一个存贮单元称为cell

cell中的数据是没有类型的，全部是字节码形式存贮

在一个cell中同一数据的不同版本的顺序是按照时间的顺序排序的√

由{row key, column, version} 唯一确定的单元cell

34.下列关于列族说法正确的是（）？（多选题）

访问控制、磁盘和内存的使用统计都是在列族层面进行的√

列族在创建表时声明，一个列族可以包含多个列√

列中的数据都是以二进制形式存在的，没有数据类型√

一个列族中的所有的列成员可能有着不同的前缀

35.下列关于时间戳说法错误的是（）？（单选题）

每个 cell都保存着同一份数据的多个版本，版本通过时间戳来索引

HBase提供了两种数据版本回收方式:一是保存数据的最后n个版本，二是保存最近一段时间内的版本

每个cell中，不同版本的数据按照时间倒序排序，即最新的数据排在最后面√

时间戳可以由HBase (在数据写入时自动 )赋值，此时时间戳是精确到毫秒的当前系统时间

36.根据下表说法正确的是（）？（多选题）



Row Key表示行键，则Key代表唯一的行键值√

T1:aaa,T2:bbb 表示是由row key1和column-family-column1唯一√

Column-family表示列族，则Column1表示列族中的第一列√

T1、T2表示时间戳，但它们的值是一样的

37.下列关于HBase存储错误的是（）？（单选题）

每个列族存储在HDFS上的一个单独文件夹中

空值不会被存储，不会占据空间

存储时，只要row key（行键）与column-family（列族）就可以确定锁定唯一值√

Key和version number会在每个列族中存储一份

38.下列关于Region与HRegion说法正确的是（）？（多选题）

HRegion不仅是分布式存储的最小单元，也是存储的最小单元

HRegion由一个或者多个Store组成，每个store保存一个columns family√

Region按大小分割的，每个表一开始只有一个Region，随着数据不断插入表，Region不断增大√

HRegion是HBase中分布式存储和负载均衡的最小单元√

39.下列关于Store与HFile说法错误的是（）？（单选题）

每个Strore又由一个memStore和0至多个StoreFile组成

File Info 段–HFile的元信息，不被压缩，用户也可以在这一部分添加自己的元信息。

Meta Block 段-保存表中的数据，这部分可以被压缩√

Trailer–这一段是定长的，保存了每一段的偏移量

40.下列关于HBase架构体系中Zookeeper说法正确的是（）？（多选题）

存贮所有HBase的寻址入口√

存储HBase的schema,包括有哪些table，每个table 有哪些column family

保证任何时候，集群中只有一个running master√

实时监控RegionServer 的状态，将RegionServer的上线和下线信息，实时通知给Ma√

41.关于HBase架构体系中Master说法错误的是（）？（单选题）

Master为Region server 分配region

Master负责region server 的负载均衡

Master发现失效的region server 并重新分配其上的region

Master只能够启动一个HMaster√

42.下列关于HBase架构体系中RegionServer与HRegionServer说法错误的是（）？（单选题）

A、HRegionServer主要负责响应用户I/O请求，向HDFS文件系统中读写数据，是HBase中最核心的模块

B、维护Master 分配给它的region，处理对这些region 的IO 请求

C、不同的HRegion不仅可以分布在不同的HRegionServer上，而且一个HRegion可以拆分到多个HRegionServer上的

D、负责切分在运行过程中变得过大的region

C√

A

D

B

43.下列关于HBaseAPI提供了哪些操作（）？（多选题）

A、API提供了计数器操作，计数器能够在一个操作中完成读和修改

B、API提供了接口，使得系统通过提供包装器集成了MapReduce框架，该包装器能够将表转换成MapReduce作业的输入源和输出目标

C、API提供了建表、删表、增加列族和删除列族操作，同时还提供了修改表和列族元数据的功能，如压缩和设置块大小

D、API还提供了协处理器操作，用于实现轻量级批处理作业，或者使用表达式基于各种操作来分析或汇总数据

E、API通过设置过滤器可以匹配返回的列，通过设置起始和终止的时间范围可以选择查询的版本

F、API提供了高效遍历某个范围的行的功能，同时可以限定返回哪些列或者返回的版本数

E√

A√

F√

B√

C√

D√

44.启动HBase的进程，除了使用start-hbase.sh ，还可以通过hbase-deamon.sh单独启动HBase所示需要的守护进程，那么下列关于单独启动HBase进程表示正确的是（）？（单选题）

hbase-deamon.sh zookeeper

hbase-deamon.sh regionserver start

start hbase-deamon.sh master

hbase-deamon.sh start master√

45.除了通过jps命令查看守护进程验证HBase是否正常启动外，还可以WebUI上通过哪个端口号来验证HMaster已经启动（）?（单选题）

8080

8088

60010√

40010

46.hbase-site.xml配置文件中的属性hbase-rootdir的作用( )？（单选题）

hbase的安装目录

配置hbase表数据存放的路径目录√

hbase启动产生的临时文件

没有任何作用

47.下列关于HBase Shell命令创建表正确的是（）？（多选题）

create ’表名’

create ’表名1’，’表名2’

create ’表名’,’列族1’，’列族2’√

create ’表名’,’列族’√

48.下列关于HBase向表中添加数据Shell操作正确的是（）？（单选题）

put ’表名’,’行名’,’值’

put ’表名’,’列族:列名’,’行名’,’值’

put ’表名’,’列族:列名’,’值’

put ’表名’,’行名’,’列族:列名’,’值’√

49.下列关于HBase Shell操作正确的是（）？（多选题）

A、删除一张表，drop ’表名’

B、删除一张表，先要屏蔽该表，才能对该表进行删除，第一步 disable ’表名称’；第二步drop ’表名

C、删除记录，delete ’表名’,’行名’ ,’列族:列名’

D、查看表中记录总数， count ‘表名‘

C√

A

B√

D√

50.关于HBase put构造方法写法正确的是（）？（多选题）

Put (byte[] row,long ts,RowLock rowlock)√

Put (byte[] row,RowLock rowlock)√

Put（byte[] row）√

Put (byte[] row,long ts)√

51.关于HBase Get的构造函数写法正确的是()？（多选题）

Get (byte[] row,long ts,RowLock rowlock)

Get(byte[] row,RowLock rowlock)√

Get(byte[] row,long ts)

Get(byte[] row)√

52.关于HBase Delete()方法说法正确的是（）？（多选题）

A、Delete deleteColumns(byte[] family,byte[] qualifier)，此方法从给定列中删除与给定时间戳相等或更旧的版本

B、Void setTimestamp(long timestamp) ，此方法从所有列族的所有列中删除与给定时间戳相同或更旧的版本

C、Delete deleteFamily(byte[] family,long timestamp),此方法删除给定列族中所有的列（包括所有版本）

D、Delete deleteColumn(byte[] family,byte[] qualifier,long timestamp) ，此方法只删除与时间戳匹配的给定列的指定版本，如果不存在，则不删除

A

B√

C

D√

53.关于扫描说法正确的是（）？（多选题）

A、当遇到了与设置的终止行相同或大于终止行的行键时，扫描也会停止。如果没有指定终止行键，会扫描到表尾

B、用户可以通过setTimestamp（）设置的时间戳，或者通过setTimeRange（）设置时间范围，进一步对结果进行限制

C、一旦设置好了Scan实例，就可以调用HTable的getScanner（）方法，获得用于检索数据的ResultScanner实例

D、由于扫描操作的工作方式类似于迭代器，所以用户无需调用scan（）方法创建实例，只需调用HTable的getScanner（）方法

C

B√

A

D√

54.关于ResultScanner类说法正确的是（）？（多选题）

A、当用户扫描到尾或到终止行时，由于没有足够的行来填充数据，返回的结果数组可能会小于既定长度

B、ResultScanner把扫描操作转换为类似的get操作，它将每一行数据封装成一个Result实例，并将所有的Result实例放入一个迭代器中

C、扫描操作通过一次RPC请求返回所有匹配的行，而是以行为单位进行返回

D、调用close（）方法释放所有由扫描控制的资源

A√

D√

B√

C

55.关于HBase 缓存和批量处理说法正确的是（）？（多选题）

A、缓存是面向行一级的操作，而批量则是面向列一级的操作

B、每一个next（）调用都会为每行数据生成一个单独的RPC请求

C、setScannerCaching()可以设置缓存大小，getScannerCaching()可以返回当前缓存大小的值

D、在扫描设置setBatch（2），则一次next（）返回的Result实例会包括10列

C√

D

A√

B√

56.关于HTable使用方法说法正确的是（）？（多选题）

A、HTableDescriptor getTableDescriptor(),这个方法，在每个表都使用一个HTableDescriptor实例来定义自己的表结构

B、所有客户端缓冲的写操作：close()方法会隐式调用flushCache（）方法

C、Configuration getConfiguration(), 这个方法允许用户访问HTable实例中使用的配置

D、byte[] getTableName(),这是一个获取表名称的快捷方法

D√

B√

C√

A√

57.关于Bytes类说法正确的是（）?（多选题）

A、Static long toLong(byte[] bytes,int offset,int length), 此方法表示用户可以输入一个字节数组、一个偏移值和一个长度值

B、Static long toLong（byte[] bytes）, 此方法表示用户可以输入一个字节数组

C、Static long toLong(byte[],int offset), 此方法表示用户可以输入一个字节数组再加一个偏移值

D、Bytes类支持原生Java类型到字节数组的互转

D√

C√

A√

B√

58.关于HBase 过滤器说法错误的是（）?（单选题）

A、Get和Scan两个类都支持过滤器，但这些类对象提供的基本API不能对行键、列名或列值进行过滤，但是通过过滤器可以达到这个目标

B、过滤器最基本的接口叫Filter，除此之外，还有一些由HBase提供的无需编程就可以直接使用的类

C、用户不仅可以使用HBase中预定好的过滤器，而且可以实现自定义的过滤器

D、HBase中两种主要的数据读取函数是get()和scan()，Scan类支持过滤器，然而Get类不支持

B

C

A

D√

59.关于CompareFilter说法正确的是（）？（单选题）

A、CompareFilter包含compare（）方法与构造器（comparator）

B、比较过滤器往往筛选数据的一段子集或特定数据

C、比较运算符可以决定什么被包含，什么被排除

D、比较器都继承自WritableArrayComparable，WritableArrayComparable实现了Writeable和Comparable接口

B

A

C

D√

60.比较操作的过程传入参数时，关于其中参数可用值说法正确的是（）？（多选题）

NO\_OP => 排除所有√

LESS => 小于√

LESS\_OR\_EQUAL => 小于等于√

GREATER\_OR\_EQUAL => 大于等于√

61.关于专业过滤器说法正确的是（）？（多选题）

A、前缀过滤器（PrefixFilter）所有的row的实例匹配prefix的时候返回结果集合

B、单列值过滤器（SingleColumnValueFilter）选定列簇和某一列，然后与列的value相比，正确的返回全部的row

C、分页过滤器（PageFilter）页过滤，通过设置pagesize参数可以返回每一页page的数量

D、行键过滤器（KeyOnlyFilter）因为一些应用只想获取data数据，而不是真实的val，可以使用这个过滤器

D√

B√

C√

A√

62.关于自定义过滤器说法正确的是（）？（多选题）

A、filterRow（List<KeyValue> kvs）一旦所有行和列经过之前两个方法的检查之后，这个方法就会被调用

B、reset() 在迭代扫描中为每个新行重置过滤器

C、filterRowKey(byte[] buffer,int offset,int length) 该方法使用Filter的实现方法检查行键，用户可以跳出列数据以避免以后的处理

D、filterKeyValue(KeyValue v) 如果本行数据没有被之前的方法过滤掉，那么执行框架会调用这个方法检索这一行中每个KeyValue实例

D√

B√

C

A√

63. 关于单计数器说法正确的是（）？（多选题）

用负值时不会减少当前计数器的值

使用0时可以得到当前计数器的值√

用负值时可以减少当前计数器的值√

使用正值是增加计数器的值√

64.关于多计数器说法正确的是（）？（多选题）

A、可以一次更新多个计数器，但这些计数器必须位于同一行

B、单计数器一次操作一个计数器，由htable的incrementColumnValue方法执行，多计数器由Increment类和htable的increment方法完成

C、hbase提供了计数器功能，可以方便修改并提供原子性保证

D、通常hbase更新操作步骤：获取rowlock，查询数据，在客户端修改数据，执行put等操作，释放rowlock

A√

D√

C√

B√

65.关于协处理器说法正确的是（）？（多选题）

A、使用协处理器可以建立辅助索引，或维护数据引用完整性

B、协处理器不允许用户在服务器端执行自己的代码

C、协处理器用来在region服务器端做一些额外的计算，如监听事件和简单统计任务等，为客户端减少压力

D、协处理器分两类：Observer和EndPoint

A√

D√

B

C√

66.关于Observer说法不正确的是（）？（单选题）

A、负责处理master服务器的所有回调函数，对应的操作为DDL比如createTable，deleteTable，addCoumn等

B、Observer类似触发器，在特定事件发生时执行，分为RegionObserver，MasterObserver，WALObserver

C、除了region本身的生命周期时间，regionObserver还处理客户端调用事件，还包括所有DML，如get，put，delete等

D、管理region级别的操作，操作分为：region生命周期变化和客户端API调用

A

B

D

C√

67.关于Endpoint说法不正确的是（）？（单选题）

A、endpoint扩展了RPC，客户端利用endpoint可以在单个或多个region执行一些统计等操作，然后将各自结果返回给客户端

B、endPoint扩展了hbase的RPC协议，类似RDBMS中的存储过程

C、实现endPoint需要两个步骤：1）扩展CoprocessorProtocol接口，设定客户端和服务器端的RPC协议；2）扩展BaseEndPointCoprocessor

D、如果是单个region，则用CoprocessorExec方法。如果是涉及多个region则用htable.coprocessorProxy方法

D√

B

A

C

68.关于表的说法错误的是（）？（单选题）

A、用户可以用一个Java String或byte[] 指定表的名称

B、HBase的面向列的存储格式允许你存储很多细节到相同的表中，而在关系型数据库模型中，将会被分成很多单独的表

C、在HBase中数据最终会存储在一张表或多张表中，使用表的主要原因是控制表中的所有行以达到共享表内的某些特性的目的

D、用户可以用一个名称或已经存在的描述符(descriptor)创建一个表

B

A

D

C√

69.关于表中列族说法正确的是（）？（多选题）

A、列是用列族(column family)名和列限定符(column qualifier)的组合来称呼的，两部分用一个冒号分开：family:qualifier

B、列族(column family)名称必须由可打印字符组成：限定符(qualifier)可以由任意二进制字符组成

C、Java中的类有一些误拼。一个更合适的名称是HColumnFamilyDescriptor，它指出它的目的是定义列族(column family)参数而不是实际的列(column)

D、创建一个列族(column family)时，你可以指定各种参数来控制它的所有特性

D√

B√

A√

C√

70.关于HBaseAdmin的基本接口描述正确的是（）？（多选题）

A、close():关闭HBaseAdmin实例的所有资源，包括与远程服务器的连接

B、Boolean isMasterRunning()：通过这个接口检查master是否正在运行。

C、HConnection getConnection()：返回连接实例

D、Configuration getConfiguration()：访问创建HBaseAdmin实例时使用到的配置实例，用户可以通过修改这个实例中的变量达到改变HBaseAdminAPI调用时依赖的配置的目的

C√

D√

A√

B√

71.关于API建表的相关方法正确的是（）？（多选题）

A、void createTable(HTableDescriptor desc,byte[][] splitKeys)

B、void createTable(HTableDescriptor desc,byte[] startKey,byte[] endKey,int numRegions)

C、void createTableAsync(HTableDescriptor desc,byte[][] splitKeys)

D、void createTable(HTableDescriptor desc)

A√

C√

D√

B√

72.关于HBase与MapReduce集成说法正确的是（）？（多选题）

A、一个MapReduce作业中的源和目标都可以是HBase表，但也有可能在一个作业中同时使用HBase作为输入和输出

B、使用MapReduce作业作为过程的一部分，该作业可以使用HBase进行读取或写入

C、用户可以在存储组合结果的不同表中查询数据。用户从哪里读数据和向哪里写数据是不受限制的

D、通过MapReduce的API实现可以以任何形式访问HBase的代码

B√

A√

C√

D√

73.下列哪项不能用于实时计算（）？（单选题）

Spark Streaming

Puma

Flink

Hadoop√

74.下列关于Storm特性说法正确的是（）？（多选题）

可扩展性√

只支持一种语言

高容错性√

高可靠性√

75.下列关于Storm体系架构说法错误的是（）？（单选题）

A、Nimbus负责在集群分发代码

B、Supervisor监听分配给它的节点，根据Nimbus的委派在必要时启动和关闭工作进程

C、Storm在建模事件流时，把流中的事件抽象为tuple即元组

D、bolt可以消费任意数量的输入流，只要将流方向导向该bolt，但是该bolt不能够发送新的流给其他bolt

C

D√

A

B

76.下列关于Storm集群组件说法错误的是（）?（单选题）

搭建Storm集群之前必须要进行Zookeeper环境部署

一个运行中的Topology由分布在不同工作节点上的多个工作进程组成

Nimbus的作用类似于Hadoop中的Namenode的角色√

Nimbus和Supervisor节点之间所有的协调工作是通过Zookeeper集群来实现的

77.下列关于Storm集群搭建说法错误的是（）？（单选题）

在Storm 1.0版本之后，搭建Storm集群之前，必须要搭建ZeroMQ、JZMQ环境√

Nimbus和Supervisor节点之间所有的协调工作是通过Zookeeper集群来实现的

Nimbus的作用类似于Hadoop中的Namenode的角色

Storm搭建之前必须要完成Python环境的部署

78.下列关于Storm启动进程说法正确的是（）？（单选题）

sbin/storm nimbus

nohup bin/storm nimbus

bin/storm nimbus &

nohup bin/storm nimbus &√

79.下列关于storm.yaml参数解释错误的是（）？（单选题）

A、supervisor.slot.ports: 每一台worker机器，你用这个配置来指定多少workers运行在哪台机器

B、storm.Zookeeper.servers: 这是一个为Storm集群配置的Zookeeper集群的主机列表

C、java.library.path: 这是Storm使用的本地库（ZeroMQ和JZMQ）载入路径

D、storm.local.dir: Nimbus和Supervisor守护程序需要一个本地磁盘目录存储大量状态（像jars，conf，其它），每台机器都创建这些目录，赋可写权限

C

D√

B

A

80.下列关于Storm基本概念说法正确的是（）？（多选题）

A、Topology: Storm实时运行应用包逻辑上成为一个topology，一个Storm的topology相当于MapReduce的job

B、Spout: spout是Topology中数据流的数据入口，充当数据采集器功能

C、Bolt: 在Topology中所有操作都是在bolt中执行的，它可以进行过滤、计算、连接、聚合、数据库读写，以及其他操作

D、Fields Grouping：按字段分组，比如按userid，具有userid的tuple会被分到相同的Bolts。而不同的userid则会被分配到不同的Bolts。

A√

C√

D√

B√

81.下列关于Spout API说法错误的是（）？（单选题）

A、Spout类里面最重要的方法是nextTuple，用于循环发射数据

B、Activate()和deactivate():Spout可以被暂时激活和关闭

C、Ack()：处理失败tuple回调方法

D、Close()：在spout关闭前执行，但是并不能得到保证其一定执行，kill-9时不执行，Storm kill [toponame]时执行

C√

B

A

D

82.下列关于Bolt API说法正确的是（）？（多选题）

A、worker在执行该Bolt时，会先调用prepare方法传入当前执行的上下文

B、cleanup：Bolt关闭的时候被调用,由于是被supervisor使用kill 9干掉的,所以并不保证cleanup会被执行

C、execute：处理输入的单个元组

D、Bolt可以发射多条消息流，使用OutputFiledsDeclarer.declareStream定义Stream，使用OutputCollector.emit选择要发射的Stream

C√

D√

A√

B√

83.下列关于构建Topology说法错误的是（）？（单选题）

A、当TOPOLOGY\_DEBUG(setDebug)被设置成true时，Storm会记录下每个组件发射的每条消息

B、调用submitTopology方法来提交Topology，它需要接受三个参数：要运行的Topology的名字、一个配置对象、以及要运行的Topology本身

C、Storm的运行有两种模式：本地模式和分布式模式

D、Topology中的每个节点必须给予一个id，id是其它Bolts用于订阅该节点的输入流

B

C

D√

A

84.下列那个选项不能够用于图计算（）？（单选题）

Infinite Graph

Pregel

Hama

Zookeeper√

85.下列关于Pregel和Hama的说法错误的是（）？（单选题）

A、Hama是Google Pregel的开源实现

B、Hama是在HDFS上实现的BSP(Bulk Synchronous Parallel)计算框架，弥补Hadoop在计算能力上的不足

C、Pregel是一种基于BSP模型实现的并行图处理系统

D、Hama的主要应用领域是：矩阵计算、面向图计算、PageRank、排序计算、BFS、实时计算

A

C

B

D√

86.下列关于Hama结构说法错误的是（）？（单选题）

A、Hama主要有三部分构成：BSPMaster 、GroomServers 和Zookeeper

B、BSPMaster与Groom Server两者之间通讯使用非常简单的LIFO(后进先出)原则对计算的任务进行分配、调度

C、GroomServer在HDFS或者其他文件系统上运行，通常，GroomServer与数据结点在一个物理结点上运行，以保证获得最佳性能

D、Zookeeper用来管理BSPPeer的同步，用于实现Barrier Synchronisation机制

C

A

D

B√

87.下列关于BSP的编程模型说法正确的是（）？（多选题）

A、BSP程序包含一序列的superstep（超步），每一个superstep包含三个步骤

B、每个计算节点进行并行计算, 在communication的阶段进行收发, 将运行结果记录在barrier上

C、Hama提供用户自定义的函数bsf()，通过bsf函数，用户可以编写自己的BSP程序

D、BSP是一种跟MapReduce平行的一种并行计算方法

B√

A√

C√

D√

88.下列关于Hama的安装和使用说法错误的是（）？（单选题）

bsp.master.address即bsp中的BSPMaster的地址和端口

fs.default.name是hadoop中nameNode的地址和端口

安装Hama环境之前，需要完成JDK和Hadoop的环境部署

启动Hama进程的命令是sbin/start-bspd.sh√

89.下面选项中哪个选项不是NoSQL数据库（）？（单选题）

Cassandra

Redis

Hive√

HBase

90.下列哪项不属于NoSQL数据库兴起的原因（）？（单选题）

NoSQL数据库都具有非常高的读写性能，尤其在大数据量下，同样表现优秀

NoSQL的Cache是记录级的，是一种细粒度的Cache

关系数据库已经无法满足Web2.0的需求

NoSQL数据库包括完善的事务机制和高效的查询机制√

91.下列关于NoSQL与关系型数据库的比较错误的是（）？（单选题）

A、RDBMS很难实现横向扩展，纵向扩展的空间也比较有限，性能会随着数据规模的增大而降低，NoSQL可以很容易通过添加更多设备来支持更大规模的数据

B、NoSQL很难实现横向扩展，纵向扩展的空间也比较有限

C、RDBMS已经标准化（SQL），NoSQL还没有行业标准，不同的NoSQL数据库都有自己的查询语言，很难规范应用程序接口

D、RDBMS都可以很容易实现数据完整性，但NoSQL却无法实现

A

B√

D

C

1. 下列关于典型的四大类型NoSQL数据库描述错误的是（）？（单选题）

典型的NoSQL数据库通常包括键值数据库、列族数据库、文档数据库和图形数据库

文档数据库缺乏统一的查询语法

HBase数据库属于文档数据库√

键值数据库无法存储结构化信息，条件查询效率较低

93.下列哪项不是MongoDB的主要特点（）？（单选题）

支持各种编程语言

具有良好的扩展性

不能够设置索引√

支持丰富的查询表达式，可轻易查询文档中内嵌的对象及数组

94.下列关于MongoDB概念解析描述错误的是（）？（单选题）

集合存在于数据库中，有固定的结构√

MongoDB 的文档不需要设置相同的字段，并且相同的字段不需要相同的数据类型

MongoDB的单个实例可以容纳多个独立的数据库

集合就是 MongoDB 文档组，类似于 RDBMS 的Table

95.下列关于MongoDB Shell描述错误的是（）？（单选题）

查询当前数据库下所有集合：show collections

查询所有数据库：show dbs

删除该集合下所有文档：db.<collectionName>.remove({})

查询第一条数据：db.<collectionName>.find()√

96.下列关于MongoDB以$符号开头的特殊键描述错误的是（）？（单选题）

$inc：自增或自减

$pull：删除数组中指定的值

$push：向数组中首部添加一个元素√

$set：更新字段