最早的“大数据”概念是美国著名未来学家阿尔温.托夫勒在他的（ ）一书中提出的。

正确答案: D

A. 《第四次浪潮》 B. 《大数据时代》

C. 《第四种范式》 D. 《第三次浪潮》

企业内部的经营交易信息是（ ）数据，通过关系数据库进行管理和访问的静态、历史数据。

正确答案: A

A. 结构化B. 非结构化C. 个性化D. 一般化

以纸牌屋为例，Netflix可以通过大数据分析电影题材、挑选演员、播放形式，再根据这些内容拍摄用户感兴趣的电影。

正确答案: 对

随着Web 2.0时代的到来，数据将从非结构化到半结构化到结构化。

正确答案: 错

大数据对思维方式的影响体现在( )

正确答案: AB

A. 全样而非抽样

B. 效率而非精确

C. 因果而非相关

科学研究经历了几种范式，包括（ ）

正确答案: ABCD

A. 实验科学

B. 理论科学

C. 计算科学

D. 数据科学

大数据的基本特征包括（ ）

正确答案: ABC

A. 数据量大

B. 数据种类繁多

C. 处理速度快

D. 价值密度高

大数据是指数据量超过一定的大小，导致常规软件无法在一个可接受的时间范围内完成对其进行抓取、管理和处理工作的数据。

正确答案: 对

大数据技术是一系列使用非传统的工具来对大量的结构化、半结构化和非结构化数据进行处理，从而获得分析和预测结果的数据处理技术。大数据涉及的关键技术有（ ）。

正确答案: ABCDE

A. .数据采集 B. 数据分析和处理技术C. 数据存储及管理技术

D. 数据展示技术 E. 数据隐私和安全

将基础设施作为服务出租属于（ ）

正确答案: C

A. SaaSB. PaaSC. IaaS

在廉价PC服务器上搭建起大规模存储集群，属于哪项云计算的关键技术（ ）

正确答案: B

A. 虚拟化B. 分布式存储C. 分布式计算D. 多租户

智能手环的应用开发，体现了（ ）的数据采集技术的应用。

正确答案: C

A. 网络爬虫B. API接口C. 传感器D. 统计报表

Hadoop的作者是下面哪一位？

正确答案: B

A. Martin Fowler B. Doug Cutting C. Kent Beck D. Grace Hopper

Hadoop体系结构中，完成分布式并行计算的模块是（ ）

正确答案: B

A. HDFSB. MapReduceC. HBase D. Mahout

Google云计算组件GFS和Hadoop系统的（ ）组件有对应关系。

正确答案: A

A. HDFSB. HBaseC. HiveD. Pig

分布式集群资源调度和管理框架是（ ）

正确答案: C

A. HDFSB. MapReduceC. YarnD. Zookeeper

采用冗余数据存储方式，即使一个副本发生故障，其他副本也可以保证正常对外提供服务，体现了Hadoop的哪个特性（ ）

正确答案: A

A. 高可靠性B. 高效性C. 高可扩展性D. 成本低

以下哪一项不属于Hadoop可以运行的模式（ ）

正确答案: C

A. 单机（本地）模式 B. 伪分布式模式 C. 互联模式 D. 分布式模式

作为并行分布式计算平台，Hadoop采用分布式存储和分布式处理两大核心技术，能够高效地处理PB级数据，体现了Hadoop的哪个特性（ ）

正确答案: B

A. 高可靠性B. 高效性C. 高可扩展性D. 高容错性

HDFS的NameNode负责管理文件系统的命名空间，将所有的文件和文件夹的元数据保存在一个文件系统树中，这些信息也会在硬盘上保存为以下文件（ ）。

正确答案: C

A. 日志 B. 命名空间镜像 C. 两者都是

SecondaryNameNode就是NameNode出现问题时的备用节点。

正确答案: 错

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。

正确答案: 错

FsImage文件中没有记录每个块存储在哪个数据节点

正确答案: 对

哪种场景适用于HDFS？（ ）

正确答案: A

A. 大规模文件存储B. 实时读取 C. 随机写文件D. 高效存储大量小文件

HDFS的设计中没有考虑以下哪个特性（ ）

正确答案: C

A. 超大文件 B. 流式的数据访问C. 低数据延迟

D. HDFS基于商用硬件，对于大集群而言，发生节点故障的几率较高

HDFS中当前block大小为128M，如果当前要上传到HDFS中的文件大小为300M，那么在存储时会分配（ ）个block进行存储

正确答案: C C

A. 1B. 2C. 3

如果NameNode意外终止，SecondaryNameNode会接替它使集群继续工作。( )

正确答案: 错

HDFS是基于流数据模式访问和处理超大文件的需求而开发的，默认的最基本的存储单位是64MB，具有高容错、高可靠性、高可扩展性、高吞吐率等特征，适合的读写任务是（ ）。

正确答案: C

A. 一次写入，少次读取B. 多次写入，少次读取

C. 一次写入，多次读取D. 多次写入，多次读取

在HDFS中，Block Size 是不可以修改的。

正确答案: 错

FileSystem类是一个（ ）。

正确答案: B

A. 接口B. 抽象类 C. 普通类 D. 内部类

HDFS默认的当前工作目录是/user/$USER，fs.defaultFS的值需要在哪个配置文件中说明？（ ）

正确答案: B

A. mapred-site.xmlB. core-site.xmlC. hdfs-site.xml D. 以上都不是

HDFS中的Block默认保存（ ）份。

正确答案: C C

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

SecondaryNameNode的作用是（ ）。

正确答案: C C

A.监控NameNodeB.管理DataNodeC.合并FsImage和EditLogD.支持NameNode HA

（ ）负责文件和目录的创建、删除和重命名等

正确答案: B B

A. DataNodeB. NameNodeC. Client

分布式文件系统所采用的计算机集群，都是由专用硬件构成的

正确答案: 错 错

下面哪个程序负责HDFS的数据存储？

正确答案: C C

A. NameNode B. JobTracker C. DataNodeD. SecondaryNameNode

因为HDFS有多个副本，所以NameNode是不存在单点问题的。

正确答案: 错 错

Client端上传文件的时候，正确的选项是（ ）。

正确答案: B B

A. 数据经过NameNode传递给DataNode

B. Client端将文件切分为Block，依次上传

C. Client只上传数据到一台DataNode，然后由NameNode负责Block复制工作

关于HDFS的文件写入，正确的是（ ）。

正确答案: C C

A. 支持多用户对同一文件的写操作 B. 用户可以在文件的任意位置进行修改

C. 默认将文件块复制成三份存放 D. 复制的文件块默认都存放在同一机架上

DistributedFileSystem调用create方法后的返回类型是（ ）。

正确答案: A A

A. FSDataOutputStreamB. DataOutputStream

C. DFSOutputStreamD. FSDataInputStream

HBase来源于哪篇论文？

正确答案: C C

A. The Google File System B. MapReduce C. BigTable D. Chubby

下面对HBase的描述哪些是正确的？

正确答案: BCD BCD

A. 不是开源的B. 是面向列的 C. 是分布式的D. 是一种NoSQL数据库

HBase依靠（ ）存储底层数据

正确答案: A A

A. HDFSB. HadoopC. Memory D. MapReduce

HBase依赖（ ）提供消息通信机制。

正确答案: A A

A. ZookeeperB. ChubbyC. RPCD. Socket

HBase依赖（ ）提供强大的计算能力

正确答案: D D

A. ZookeeperB. ChubbyC. RPCD. MapReduce

下面与Zookeeper类似的框架是？

正确答案: D D

A. ProtobufB. JavaC. KafkaD. Chubby

下面哪些选项正确描述了HBase的特性？

正确答案: ABCD ABCD

A. 高可靠性B. 高性能C. 面向列D. 可伸缩

HBase表中每个cell的多版本（ ）是通过什么表示的？

正确答案: A A

A. timestampB. rowkeyC. block\_idD. cell\_id

HBase的物理存储单元是（ ）

正确答案: B B

A. RegionB. ColumnFamilyC. ColumnD. Row

HBASE表的索引构成包括（ ）

正确答案: ABCD ABCD

A. 行键B. 列族C. 列限定符D. 时间戳

HBASE表中单元格中的值是（ ）

正确答案: A C

A. 未经解释的字符串

B. 字符串

C. 任意类型的数据

D.

实现定义好类型的数据

HBASE表中一个单元格可以保存一个数据的多个版本，每个数据版本对应相同的时间戳。（ ）

正确答案: 错 错

对

错

客户端通过访问Master服务器来获得Region位置信息（ ）

正确答案: 错 错

对

错

传统关系型数据库数据定位是二维坐标，Hbase是四维坐标定位数据。（ ）

正确答案: 对 对

对

错

META.表记录所有元数据的具体位置, .META.表只有唯一一个Region，名字是在程序中被写死的（ ）

正确答案: 错 错

对

错

Client与（ ）通信进行表和Region的管理操作。

正确答案: C C

A.

RegionServer

B.

ZooKeeper

C.

Master

D.

HDFS

关于Hbase中Region描述正确的是（ ）

正确答案: ACD

A.

Region是负载均衡和数据分发的基本单位

B.

同一个Region会被拆分到多个Region服务器上的

C.

每个Region服务器负责管理一个Region集合

D.

每个Region的最佳大小取决于单台服务器的有效处理能力

HBase分布式模式最少需要（ ）个节点？

正确答案: C

A.

1

B.

2

C.

3

D.

4

下列哪些选项是安装HBase前所必须安装的？

正确答案: AB

A.

操作系统

B.

JDK

C.

Shell Script

D.

Java Code

Hbase中Zookeeper服务器的功能描述正确的是（ ）

正确答案: ABCD

A.

Zookeeper存储了-ROOT-表地址

B.

Zookeeper提供稳定可靠的协同服务

C.

Hbase中启动多个Master，Zookeeper可以选出一个做集群的总管

D.

Zookeeper实时监控每个Region服务器的状态并通知Master

一个Region服务器内部管理多个HLog文件（ ）

正确答案: 错

对

错

如果对HBase表的添加数据记录，可以使⽤（ ）命令进⾏操作。

正确答案: C

A.

create

B.

get

C.

put

D.

Scan

HBase伪分布式模式需要（ ）个节点？

正确答案: A

A.

1

B.

2

C.

3

D.

4

以下更新HBase表中数据的语法，正确的是（ ）。

正确答案: D

A.

update user. xiaoming’ set info.age=1

B.

update ‘users’,’xiaoming’,’info.age’,’29’

C.

put user. xiaoming’ info.age=1

D.

put ‘users’,’xiaoming’,’info.age’,’29’

在HBase Shell操作中，请选择以下哪个命令⽤于删除整⾏操作（ ）。

正确答案: C C

A.

delete from ‘users’.’xiaoming’

B.

delete table from ‘xiaoming’

C.

deleteall ‘users’,’xiaoming’

D.

deleteall ‘xiaoming’

NewSQL具备而NoSQL不具备的哪些特性（ ）

正确答案: C C

A.

海量数据存储特性

B.

可扩展和高性能特性

C.

ACID和SQL特性

D.

分布式数据管理特性

Base 原理包括哪几项（ ）

正确答案: ABC ABC

A.

最终一致性

B.

基本可用性

C.

软状态

D.

分区容忍性

在分布式系统中，N1 和 N2 节点上存有相同的数据，当用户 A 对 N1 节点的数据进行修改时，正常情况下，N2 节点会同步 N1 节点的数据，但是当出现分区容错时，即 N1 与 N2 节点由于某种原因数据无法同步，这时，用户 B 需要从 N2 节点获取数据，如果系统返回原来的旧数据给用户 B，这个系统满足什么样的 CAP 条件。（ ）

正确答案: A A

A.

选择分区容错性和可用性，放弃一致性

B.

选择分区容错性和一致性，放弃可用性

C.

满足分区容错性，可用性和一致性

D.

选择一致性和可用性，放弃分区容错性

NoSQL是⼀种⾮关系型、分布式、不遵循ACID、不提供SQL功能的数据库。

正确答案: 对 对

对

错

NoSQL数据库允许在一个数据元素里存储不同类型的数据。

正确答案: 对 对

对

错

NoSQL包括完善的事务机制和高效的查询机制。

正确答案: 错 对

对

错

CAP理论的核⼼思想是保证系统的强⼀致性。

正确答案: 错 错

对

错

以下属于列族数据库的是（ ）

正确答案: BC C

A.

Redis

B.

BigTable

C.

HBase

D.

MongoDB

下列数据库中，哪个是最简单的NoSQL数据库？

正确答案: A A

A.

键值数据库

B.

⽂档数据库

C.

列族存储数据库

D.

图数据库

交易型系统，比如银行核心系统对一致性和可用性要求比较高，基本上选择关系型数据库系统进行数据的管理，由于它放弃了分区容错性，因此系统在扩展上存在限制。

正确答案: 对 对

对

错

以下哪些符合 NoSQL 数据库的特点？

正确答案: ABCD ABC

A.

NoSQL 数据满足最终一致性

B.

灵活的数据模式

C.

NoSQL 集群的可扩展性强，可动态添加和删除节点

D.

数据以多副本的方式存放在不同节点上

NoSQL 是指 no SQL，表示用非关系型数据库替代关系型数据库。

正确答案: 错 错

对

错

适合于存储大量复杂、互连接的数据类型，比如人际交往、推荐系统和知识图谱的 NoSQL

数据库有（ ）

正确答案: B B

A.

MongoDB

B.

Neo4j

C.

Redis

D.

HBase

在一个分布式系统中可以同时满足一致性（Consistency）、可用性（Availability）和分区容错

性（Partition tolerance）（ ）

正确答案: 错 错

对

错