被誉为全球未来的三大高科技产业除了塑料电子学和仿生人体器官外，还有（）

A、物联网

B、数学

C、经典力学

D、社会学

云计算的基础层是（）

A、IaaS 层

B、PaaS层

C、SaaS层

D、BaaS层

云计算是对（ ）技术的发展与运用。

A、并行计算

B、网格计算

C、分布式计算

D、三个选项都是

云计算的中间层是（）

A、IaaS 层

B、PaaS层

C、SaaS层

D、BaaS层

云计算里面面临的一个很大的问题，就是（ ）

A、服务器

B、存储

C、计算

D、节能

与开源云计算系统HadoopHDFS相对应的商用云计算软件系统是（ ）

A、GoogleGFS

B、GoogleMapReduce

C、GoogleBigtable

D、GoogleChubby

云计算体系结构的（ ）负责资源管理、任务管理用户管理和安全管理等工作。

A、物理资源层

B、资源池层

C、管理中间件层

D、SOA构建层

云计算就是把计算资源都放到( )上。

A、对等网

B、因特网

C、广域网

D、无线网

多选

工业4.0 通过自主控制、联网和计算，使人、机器和信息相互连接，融为一体，包括：（ ）

A、智能工厂

B、智能生产

C、智能物流

D、智能金融

“ 互联网+ ” 行动计划重点是促进以（ ）为代表的新一代信息技术与现代制造业、生产性服务等的融合创新。

A、云计算

B、物联网

C、大数据

D、云存储

大数据具有哪些特征（ ）

A、数据类型繁多

B、价值密度低

C、数据量大

D、处理速度快

下列系统中，属于开源系统的是（）

A、GFS 文件系统

B、HDFS 文件系统

C、HBase 数据管理模块

D、BigTable 数据管理模块

云架构包含以下哪些？

A、基础设施层

B、服务层

C、应用层

D、平台层

云解决方案包括（）

A、云平台开发

B、云咨询

C、云迁移

D、云测试

大数据催生的三元空间世界包括（ ）

A、物理空间

B、社会空间

C、数据空间

D、网络空间

从研究现状上看，下面属于云计算特点的是（ ）

A、超大规模

B、虚拟化

C、私有化

D、高可靠性

第二章 大数据处理架构 Hadoop

单选

在 Hadoop 项目结构中，MapReduce 指的是什么?

A、流计算框架

B、分布式并行编程模型

C、Hadoop上的工作流管理系统

D、提供分布式协调一致性服务

假设设置的副本数（即参宿 dfs.replication ）是 3，现在系统中有 5 个副本，那么系统会删除几个副本。

A、4

B、5

C、1

D、2

哪个是第三次信息化浪潮的标志。

A、个人计算机

B、互联网

C、信息处理

D、云计算，物联网和大数据

二维码目前不能表示的数据类型是（）。

A、文字

B、数字

C、二进制

D、视频

Hadoop 框架中最核心的设计是什么？

A、为海量数据提供存储的 HDFS 和对数据进行计算的 MapReduce

B、提供整个 HDFS 文件系统的 NameSpace (命名空间)管理、块管理等所有服务

C、Hadoop不仅可以运行在企业内部的集群中，也可以运行在云计算环境中

D、Hadoop被视为事实上的大数据处理标准

（ ）是负责对物联网收集到的信息进行处理、管理、决策的后台计算处理平台。

A、感知层

B、网络层

C、云计算平台

D、物理层

权威的物联网的概念是（）年发布的《物联网报告》中所提出的定义。

A、1998

B、1999

C、2000

D、2005

下列那个不是常见的关系数据库？（）

A、Oracle

B、SQL Server

C、MySQL

D、Qcreator

为数据中心选取地址时，一般不会优先考虑下面哪个因素

A、是否有廉价的电能

B、是否靠近大型水源

C、是否有较大面积的空地

D、是否有良好的空气质量

下列哪个不属于 Hadoop 的特性

A、成本高

B、高可靠性

C、高容错性

D、运行在Linux平台上

物联网的概念，最早是由美国的麻省理工学院在（）年提出来的。

A、1998

B、1999

C、2000

D、2002

当前大数据技术的基础是由（ ）首先提出的。

A、微软

B、百度

C、谷歌

D、阿里巴巴

配置Hadoop时，JAVA\_HOME 包含在哪一个配置文件中。

A． hadoop-default.xml

B． hadoop-env.sh

C． hadoop-site.xml

D． configuration.xsl

世界上第一个原始的搜索引擎为（）。

A、W3Catalog

B、Aliweb

C、Google

D、Baidu

利用 RFID、传感器、二维码等随时随地获取物体的信息，指的是（）。

A、可靠传递

B、全面感知

C、智能处理

D、互联网

在云计算平台中，（）软件即服务。

A、IaaS

B、PaaS

C、SaaS

D、QaaS

关于Hadoop单机模式和伪分布式模式的说法，正确的是

A． 两者都起守护进程，且守护进程运行在一台机器上

B． 单机模式不使用 HDFS，但加载守护进程

C． 两者都不与守护进程交互，避免复杂性

D． 后者比前者增加了 HDFS 输入输出以及可检查内存使用情况

运用云计算、数据挖掘以及模糊识别等人工智能技术，对海量的数据和信息进行分析和处理，对物体实施智能化的控制，指的是（）。

A、可靠传递

B、全面感知

C、智能处理

D、互联网

第三章 分布式文件系统 HDFS

单选

分布式文件系统指的是什么?

A、把文件分布存储到多个计算机节点上，成千上万的计算机节点构成计算机集群

B、用于在 Hadoop 与传统数据库之间进行数据传递

C、一个高可用的,高可靠的，分布式的海量日志采集、聚合和传输的系统

D、一种高吞吐量的分布式发布订阅消息系统，可以处理消费者规模的网站中的所有动作流数据

下面哪一项不属于计算机集群中的节点?

A、主节点(Master Node)

B、源节点(SourceNode)

C、名称结点(NameNode)

D、从节点(Slave Node)

在HDFS2.X中，默认一个块多大?

A、64MB

B、32MB

C、128MB

D、16MB

下列哪一项不属于 HDFS 采用抽象的块概念带来的好处?

A、简化系统设计

B、支持大规模文件存储

C、强大的跨平台兼容性

D、适合数据备份

在 HDFS 中，NameNode 的主要功能是什么?

A、维护了 block id 到 datanode 本地文件的映射关系

B、存储文件内容

C、文件内存保存在磁盘中

D、存储元数据

下面对 FsImage 的描述，哪个是错误的?

A、FsImage 文件没有记录每个块存储在哪个数据节点

B、FsImage 文件包含文件系统中所有目录和文件 inode 的序列化形式

C、FsImage 用于维护文件系统树以及文件树中所有的文件和文件夹的元数据

D、FsImage 文件记录了每个块具体被存储在哪个数据节点

下面对 SecondaryNameNode 第二名称节点的描述，哪个是错误的?

A、SecondaryNameNode 一般是并行运行在多台机器上

B、它是用来保存名称节点中对 HDFS 元数据信息的备份，并减少名称节点重启的时间

C、SecondaryNameNode 通过 HTTPGET 方式从 NameNode 上获取到 FsImage 和 EditLog 文件,并下载到本地的相应目录下

D、SecondaryNameNode 是 HDFS 架构中的一个组成部分

HDFS采用了什么模型?

A、分层模型

B、主从结构模型

C、管道-过滤器模型

D、点对点模型

在 Hadoop 项目结构中，HDFS 指的是什么?

A、分布式文件系统

B、流数据读写

C、资源管理和调度器

D、Hadoop上的数据仓库

下列关于HDFS的描述，哪个不正确?

A、HDFS还采用了相应的数据存放、数据读取和数据复制策略，来提升系统整体读写响应性能

B、HDFS采用了主从(Master/Slave)结构模型

C、HDFS采用了冗余数据存储，增强了数据可靠性

D、HDFS采用块的概念，使得系统的设计变得更加复杂

多选

HDFS 要实现以下哪几个目标?

A、兼容廉价的硬件设备

B、流数据读写

C、大数据集

D、复杂的文件模型

HDFS 特殊的设计，在实现优良特性的同时，也使得自身具有一些应用局限性，主要包括以下哪几个方面?

A、较差的跨平台兼容性

B、无法高效存储大量小文件

C、不支持多用户写入及任意修改文件

D、不适合低延迟数据访问

HDFS 采用抽象的块概念可以带来以下哪几个明显的好处?

A、支持大规模文件存储

B、支持小规模文件存储

C、适合数据备份

D、简化系统设计

在 HDFS 中，名称节点主要保存了哪些核心的数据结构?

A、FsImage

B、DN8

C、Block

D、EditLog

数据节点的主要功能包括哪些?

A、负责数据的存储和读取

B、根据客户端或者是名称节点的调度来进行数据的存储和检索

C、向名称节点定期发送自己所存储的块的列表

D、用来保存名称节点中对 HDFS 元数据信息的备份，并减少名称节点重启的时间

HDFS 的命名空间包含什么?

A、磁盘

B、文件

C、块

D、目录

下列对于客服端的描述，哪些是正确的?

A、客户端是用户操作 HDFS 最常用的方式，HDFS 在部署时都提供了客户端

B、HDFS 客户端是一个库，暴露了 HDFS 文件系统接口

C、严格来说，客户端并不算是 HDFS 的一部分

D、客户端可以支持打开、读取、写入等常见的操作

HDFS 只设置唯一一个名称节点,这样做虽然大大简化了系统设计，但也带来了哪些明显的局限性?

A、命名空间的限制

B、性能的瓶颈

C、隔离问题

D、集群的可用性

HDFS 数据块多副本存储具备以下哪些优点?

A、加快数据传输速度

B、容易检查数据错误

C、保证数据可靠性

D、适合多平台上运行

HDFS 具有较高的容错性，设计了哪些相应的机制检测数据错误和进行自动恢复?

A、数据源太大

B、数据节点出错

C、数据出错

D、名称节点出错

第四章 MapReduce

单选

下列传统并行计算框架，说法错误的是哪一项?

A、刀片服务器、高速网、SAN,价格贵,扩展性差上

B、共享式(共享内存/共享存储),容错性好

C、编程难度高

D、实时、细粒度计算、计算密集型

下列关于 MapReduce 模型的描述，错误的是哪一项?

A、MapReduce 采用 “ 分而治之 ” 策略

B、MapReduce 设计的一个理念就是 “ 计算向数据靠拢 ”

C、MapReduce 框架采用了 Master/Slave 架构

D、MapReduce应用程序只能用Java来写

MapReduce1.0 的体系结构中，JobTracker 是主要任务是什么?

A、负责资源监控和作业调度，监控所有 TaskTracker 与 Job 的健康状况

B、使用 “ slot ” 等量划分本节点上的资源量(CPU、内存等)

C、会周期性地通过 “ 心跳 ” 将本节点上资源的使用情况和任务的运行进度汇报给 TaskTracker

D、会跟踪任务的执行进度、资源使用量等信息,并将这些信息告诉任务(Task)

下列关于 MapReduce 工作流程,哪个描述是正确的?

A、所有的数据交换都是通过 MapReduce 框架自身去实现的

B、不同的Map任务之间会进行通信

C、不同的Reduce任务之间可以发生信息交换

D、用户可以显式地从一台机器向另一台机器发送消息

下列关于 MapReduce 的说法，哪个描述是错误的?

A、MapReduce具有广泛的应用，比如关系代数运算、分组与聚合运算等

B、MapReduce将复杂的、运行于大规模集群上的并行计算过程高度地抽象到了两个函数

C、编程人员在不会分布式并行编程的情况下，也可以很容易将自己的程序运行在分布式系统上，完成海量数据集的计算

D、不同的 Map 任务之间可以进行通信

下列关于 Map 和 Reduce 函数的描述，哪个是错误的?

A、Map将小数据集进一步解析成一批<key,value>对,输入Map函数中进行处理

B、Map每一个输入的<k 1 ,v 1 >会输出一批<k 2 ,v 2 >。<k 2 ,v 2 >是计算的中间结果

C、Reduce输入的中间结果<k 2 ,List(v 2 )>中的List(v 2 )表示是一批属于不同k 2 的value

D、Reduce输入的中间结果<k 2 ,List(v 2 )>中的List(v 2 )表示是一批属于同一个k 2 的value

下面哪一项不是 MapReduce 体系结构主要部分?

A、Client

B、JobTracker

C、TaskTracker以及Task

D、Job

关于 MapReduce1.0 的体系结构的描述，下列说法错误的是?

A、Task 分为 Map Task 和Reduce Task 两种,分别由JobTracker 和TaskTracker 启动

B、slot 分为Map slot 和Reduce slot 两种,分别供MapTask 和Reduce Task 使用

C、TaskTracker 使用“slot”等量划分本节点上的资源量(CPU、内存等)

D、TaskTracker 会周期性接收JobTracker 发送过来的命令并执行相应的操作(如启动新任务、杀死任务等)

下列说法错误的是?

A、Hadoop MapReduce 是 MapReduce 的开源实现，后者比前者使用门槛高很多

B、MapReduce 采用非共享式架构，容错性好

C、MapReduce主要用于批处理、实时、计算密集型应用

D、MapReduce 采用 “ 分而治之 ” 策略

多选

MapReduce 相较于传统的并行计算框架有什么优势?

A、非共享式，容错性好

B、普通PC机，便宜，扩展性好

C、编程简单，只要告诉 MapReduce 做什么即可

D、批处理、非实时、数据密集型

MapReduce体系结构主要由以下那几个部分构成?

A、Client

B、JobTracker

C、TaskTracker

D、Task

下列关于 MapReduce 的体系结构的描述，说法正确的有?

A、用户编写的 MapReduce 程序通过 Client 提交到 JobTracker 端

B、JobTracker 负责资源监控和作业调度

C、TaskTracker 监控所有 TaskTracker 与Job 的健康状况

D、TaskTracker 使用 “ slot ” 等量划分本节点上的资源量(CPU、内存等)

MapReduce 的作业主要包括什么?

A、从磁盘或从网络读取数据，即 I/O 密集工作

B、计算数据，即 CPU 密集工作

C、针对不同的工作节点选择合适硬件类型

D、负责协调集群中的数据存储

对于MapReduce 而言，其处理单位是 split。split 是一个逻辑概念，它包含哪些元数据信息?

A、数据起始位置

B、数据长度

C、数据所在节点

D、数据大小

下列关于 Map 端的 Shuffle 的描述，哪些是正确的?

A、MapReduce 默认为每个 Map 任务分配 1000MB 缓存

B、多个溢写文件归并成一个或多个大文件，文件中的键值对是排序的

C、当数据很少时，不需要溢写到磁盘，直接在缓存中归并，然后输出给Reduce

D、每个 Map 任务分配多个缓存，使得任务运行更有效率

MapReduce 的具体应用包括哪些？

A、关系代数运算（选择、投影、并、交、差、连接）

B、分组与聚合运算

C、矩阵-向量乘法

D、矩阵乘法

MapReduce 执行的全过程包括以下哪几个主要阶段?

A、从分布式文件系统读入数据

B、执行 Map 任务输出中间结果

C、通过 Shuffle 阶段把中间结果分区排序整理后发送给 Reduce 任务

D、执行 Reduce 任务得到最终结果并写入分布式文件系统

下列说法正确的是?

A、MapReduce 体系结构主要由四个部分组成,分别是：Client、JobTracker、TaskTracker以及Task

B、Task 分为 Map Task 和 Reduce Task 两种，均由 TaskTracker 启动

C、在 MapReduce 工作流程中，所有的数据交换都是通过 MapReduce 框架自身去实现的

D、在 MapReduce 工作流程中，用户不能显式地从一台机器向另一台机器发送消息

————————————————

版权声明：本文为CSDN博主「hurricane&amp;&amp;storming」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/weixin\_48033173/article/details/116784319