

原创 hurricane&&storming 于 2021-05-14 23:52:40 发布 4583 收藏 110 版权

分类专栏: 大数据技术原理与应用 文章标签: 大数据 hadoop MapReduce HBase HDFS

 华为云开发者联盟 该内容已被华为云开发者联盟社区收录 [加入社区](#)

 大数据技术原理与应用 专栏收录该内容 4 订阅 2 篇文章 [订阅专栏](#)

大数据原理与应用期末备考 三天速成不挂科

☆ 内容期末大概率考
选择部分直达 → 选择部分

导航

大数据原理与应用期末备考 三天速成不挂科

- 第一章 大数据概述
- 第二章 大数据处理架构 Hadoop
- 第三章 分布式文件系统 HDFS
- 第四章 分布式数据库 HBase
- 第五章 NoSql 数据库
- 第六章 云数据库
- 第七章 MapReduce
- 第八章 Hadoop 再探讨

第一章 大数据概述

1. 试述大数据的四个基本特征。

数据量大：人类进入信息社会后，数据以自然方式增长，数据每两年就会增加一倍多。

数据类型繁多：大数据的数据类型非常丰富，包括结构化数据和非结构化数据，如邮件、音频、视频等，给数据处理和分析技术提出了新的挑战。

处理速度快：由于很多应用都需要基于快速生成的数据给出实时分析结果，因此新兴的大数据分析技术通常采用集群处理和独特的内部设计

价值密度低：有价值的数据分散在海量数据中

 hurricane&&storm... [关注](#)

2. 举例说明大数据的关键技术。

大数据技术层面	功能
数据采集与预处理	利用 ETL 工具将分布在异构数据源中的数据抽到临时中间层后进行清洗、转换和集成后加载到数据仓库中，成为联机分析处理、数据挖掘的基础，也可以利用日志采集工具（如 Flume、Kafka 等）将实时采集的数据作为流计算系统的输入，进行实时处理分析。
数据存储和管理	利用分布式文件系统、NoSQL 数据库等实现对数据的存储和管理。
数据处理与分析	利用分布式并行编程模型和计算框架，结合机器学习和数据挖掘算法，实现对海量数据的处理和分析，并进行可视化呈现。
数据安全和隐私保护	构建数据安全体系和隐私数据保护体系。

3. 详细阐述大数据、云计算和物联网三者之间的区别与联系

区别	联系
大数据侧重于海量数据的存储、处理与分析，从海量数据中发现价值，服务于生产和生活；云计算旨在整合和优化各种 IT 资源并通过网络以服务的方式，廉价地提供给用户；物联网的发展目标是实现“物物相连”，应用创新是物联网的核心。	从整体上看，大数据、云计算和物联网这三者是相辅相成的。大数据根植于云计算，大数据分析的很多技术都来自于云计算，云计算的分布式存储和管理系统提供了海量数据的存储和管理能力，分布式并行处理框架 MapReduce 提供了数据分析能力。没有这些云计算技术作为支撑，大数据分析就无从谈起。物联网的传感器源源不断的产生大量数据，构成了大数据的重要数据来源，物联网需要借助于云计算和大数据技术，实现物联网大数据的存储、分析和处理。

第二章 大数据处理架构 Hadoop

☆☆☆ 1. 试述 Hadoop 具有哪些特性。

1. 高可靠性。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。
2. 高效性。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。



hurricane&&storm...

关注



4. 高容错性。采用冗余数据存储方式，自动保存数据的多个副本，并且能够自动将失败的任务进行重新分配。
5. 成本低。Hadoop 采用廉价的计算机集群，成本比较低，普通用户也很容易用自己的 PC 搭建 Hadoop 运行环境。
6. 运行在 Linux 操作系统上。Hadoop 是基于 Java 开发的，可以较好的运行在 Linux 系统上。
7. 支持多种编程语言。Hadoop 上的应用程序也可以使用其他语言编写，如 C++。

2. 试述 Hadoop 的项目结构以及每个部分的具体功能。

1. HDFS 是 Hadoop 项目的两个核心之一，它是针对谷歌文件系统的开源实现。
2. HBase 是一个提供高可靠性、高性能、可伸缩、实时读写、分布式的列式数据库，一般采用 HDFS 作为其底层数据存储系统。
3. Hadoop MapReduce 是针对谷歌 MapReduce 的开源实现。MapReduce 是一种编程模型，用于大规模数据集（大于 1 TB）的并行运算。
4. Hive 是一个基于 Hadoop 的数据仓库工具，可以用于对 Hadoop 文件中的数据集进行数据整理、特殊查询和分析存储。
5. Pig 是一种数据流语言和运行环境，适合于使用 Hadoop 和 MapReduce 平台上查询大型半结构化数据集。
6. ZooKeeper 是针对谷歌 Chubby 的一个开源实现，是高效和可靠的协同工作系统，提供分布式锁之类的基本服务，用于构建分布式应用，减轻分布式应用程序所承担的协调任务。
7. Sqoop 主要用来在 Hadoop 和关系数据库之间交换数据，可以改进数据的互操作性。

3. 试列举单机模式和伪分布式模式的异同点。

单机模式： Hadoop 只在一台机器上运行，存储采用本地文件系统，没有采用分布式文件系统 **HDFS**。

伪分布式模式： Hadoop 存储采用分布式文件系统 HDFS，但是，HDFS 的名称节点和数据节点都在同一台机器上。

第三章 分布式文件系统 HDFS

1. 试述 HDFS 中的名称节点和数据节点的具体功能。

名称节点 负责管理分布式文件系统系统的命名系统中的每个文件中各个块所在的数据节点。
数据节点 是分布式文件系统 HDFS 的工作节点。



hurricane&&storm...

关注



博客

下载

学习

社区

GitCode

InsCode



2. HDFS 只设置唯一一个名称节点，在简化系统设计的同时也带来了一些明显的局限性，请阐述局限性具体表现在哪些方面。

1. **命名空间的限制。**名称节点是保存在内存中的，因此名称节点能够容纳对象（文件、块）的个数会受到内存空间大小的限制。
2. **性能的瓶颈。**整个分布式文件系统的吞吐量受限于单个名称节点的吞吐量。
3. **隔离问题。**由于集群中只有一个名称节点，只有一个命名空间，因此无法对不同的应用程序进行隔离。
4. **集群的可用性。**一旦这个唯一的名称节点发生故障，会导致整个集群变得不可用。

3. 试述 HDFS 的冗余数据保存策略。

采用了多副本方式多数据进行存储。即先在集群内挑选一个磁盘不太满、CPU 不太忙的数据节点作为第一个副本存放点；选取不同的机架的数据节点作为第二副本存放点；选择与第一副本存放点同机架的不同节点作为第三副本存放点；第四副本存放点从集群中随机挑选节点。

4. 数据复制主要是在数据写入和数据恢复的时候发生，HDFS 数据复制是使用流水线复制的策略，请阐述该策略的细节。

每个块都会向 HDFS 集群中的名称节点发出写请求，名称节点会返回一个数据节点列表给客户端，客户端将数据写入列表中第一个数据节点时，同时把列表传给第一个节点；第一个节点在接收到数据写入本地的同时，会把自己已经接收到的数据传给第二个数据节点，同时第二个数据节点接收到数据时，会在写入的同时将数据发送给第三个节点，以此类推。最后，当文件写完的时候，数据复制也同时完成了。

5. 请阐述 HDFS 在不发生故障的情况下读文件的过程。

- 1) 客户端打开文件，创建输入流；
- 2) 输入流通过远程调用名称节点，获得文件开始部分数据块的保存位置；
- 3) 客户端得到位置后开始读取数据，输入流选择距离客户端最近的数据节点建立连接并读取数据；
- 4) 数据从该数据节点读取至客户端，当该数据块读取完成后，输入流查找下一个数据块；
- 5) 输入流查找下一个数据块；



hurricane&&storm...

关注

6. 请阐述 HDFS 在不发生故障的情况下写文件的过程。

- 1) 客户端创建文件和输出流；
- 2) HDFS 调用名称节点，在文件系统的命名空间中建一个新的文件，并执行检查；检查通过后，名称节点会构造一个新文件夹，并添加文件信息；
- 3) 客户端通过输出流向 HDFS 的文件写入数据；
- 4) 客户端写入的数据首先会被分成一个个的分包，将分包放入输出流对象的内部队列，并向名称节点申请若干个数据节点，然后通过流水线复制策略打包成数据包发送出去；
- 5) 为保证所有数据节点的数据都是准确的，需要数据节点向发送者发送“确认包”，当客户端收到应答时，将对应的分包从内部队列移除。不断执行 3~5 直至数据写完；
- 6) 客户端关闭输出流，通知名称节点关闭文件。

第四章 分布式数据库 HBase

1. 请阐述 HBase 和传统关系数据库的区别。

主要区别	HBase	传统关系数据库
数据类型	数据模型	关系模型
数据操作	插入、查询、删除、清空，无法实现表与表之间关联	插入、删除、更新、查询、多表连接
存储模式	基于列存储，每个列族都由几个文件保存，不同列族的文件是分离的	基于行模式存储，元组或行会被连续地存储在磁盘中
数据索引	只有一个行键索引	针对不同列构建复杂的多个索引
数据维护	更新操作不会删除数据旧的版本，而是生成一个新的版本	用最新的当前值去替换记录中原来的旧值
可伸缩性	轻易地通过在集群中增加或者减少硬件数量来实现性能的伸缩	很难实现横向扩展，纵向扩展的空间也比较有限

☆☆☆ 2. 试述 HBase 各功能组件及其作用。



hurricane&&storm...

关注

信息和 Region 服务器列表；

(3) 许多个 Region 服务器：Region 服务器是 HBase 中最核心的模块，负责维护分配给自己的 Region，并响应用户的读写请求。

3. 试述 HBase 的三层结构中各层次的名称和作用。

层次	名称	作用
第一层	ZooKeeper 文件	记录了 -ROOT- 表的位置信息
第二层	-ROOT- 表	记录了 .META. 表的 Region 位置信息，-ROOT- 表只能有一个 Region。通过 -ROOT- 表，就可以访问.META.表中的数据
第三层	.META. 表	记录了用户数据表的 Region 位置信息，.META. 表可以有多个 Region，保存了 HBase 中所有用户数据表的 Region 位置信息

4. 试述 HBase 系统基本架构以及每个组成部分的作用。

- (1) 客户端
- 客户端包含访问 HBase 的接口，同时在缓存中维护着已经访问过的 Region 位置信息，用来加快后续数据访问过程。
- (2) ZooKeeper 服务器
- ZooKeeper 可以帮助选举出一个 Master 作为集群的总管，并保证在任何时刻总有唯一一个 Master 在运行，这就避免了 Master 的“单点失效”问题。
- (3) Master 主服务器
- Master 主服务器主要负责表和 Region 的管理工作：管理用户对表的增加、删除、修改、查询等操作；实现不同 Region 服务器之间的负载均衡；在Region分裂或合并后，负责重新调整 Region 的分布；对发生故障失效的 Region 服务器上的 Region 进行迁移。
- (4) Region 服务器
- Region 服务器是 HBase 中最核心的模块，负责维护分配给自己的 Region，并响应用户的读写请求。

第五章 NoSql 数据库

1. 请比较关系数据库和 NoSQL 数据库的优缺点。



关系数据库	以完善的关系代数理论作为基础，有严格的标准，支持事务 ACID 四性，借助索引机制可以实现高效的查询，技术成熟，有专业公司的技术支持	可扩展性较差，无法较好地支持海量数据存储，数据模型过于死板，无法较好支持 Web2.0 应用，事务机制影响了系统的整体性能等
NoSQL 数据库	可以支持超大规模数据存储，灵活的数据模型可以很好支持 Web2.0 应用，具有强大的横向扩展能力等	缺乏数学理论基础，复杂查询性能不高，一般不能实现事务强一致性、很难实现数据完整性，技术尚不成熟，缺乏专业团队的技术支持，维护较困难等

2. 试述 NoSQL 数据库的四大类型。

- 键值数据库：使用一个哈希表，表中有一个特定的 Key 和一个指针指向特定的 Value，Key 可以用来定位 Value，即存储和检索具体的 Value。
- 列族数据库：一般采用列族数据模型，数据库由多个行构成，每行数据包含多个列族，不同的行可以具有不同数量的列族，属于同一列族的数据会被存放在一起。
- 文档数据库：以文档作为最小单位，大都假定文档以某种标准化格式封装并对数据进行加密，同时用多种格式进行解码。
- 图数据库：使用图作为数据模型来存储数据。

☆☆☆ 3. 试述键值数据库、列族数据库、文档数据库和图数据库的适用场合和优缺点。（主要是优缺点，相应产品了解即可）

数据库类型	优点	缺点	场合	产品
键值数据库	扩展性好、灵活性好、大量写操作是性能高	无法存储结构化	内容缓存	Redis、SimpleDB
列族数据库	查找速度快、可扩展性强、容易进行分布式扩展、复杂性低	功能较少大都不支持强事务一致性	分布式数据存储与管理	BigTable、HBase、HadoopDB
文档数据库	性能好、灵活性高、复杂性低、数据结构灵活	缺乏统一的查询语法	存储、索引并管理面向文档的数据或者类似的半结构化数据	MongoDB、SisoDB
图数据库	灵活性高、支持复杂的图算法、可用于构建复杂的关系图谱	复杂性高、只能支持一定的数据规模	应用于大量复杂、互连接、低结构化的图结构场	Neo4J、OrientDB



CSDN

博客

下载

学习

社区

GitCode

InsCode

GPU

C (Consistency) : **一致性**。在分布式环境中, 多个数据是一致的。

A (Availability) : **可用性**。指能够快速获取数据, 且在确定的时间内返回操作结果。

P (Tolerance of Network Partition) : **分区容忍性**, 指当出现网络分区的情况时, 分离的系统也能正常运行。

5. 述数据库的 ACID 四性的含义。

A (Atomicity) : **原子性**。指事务必须是原子工作单元, 对于其数据修改, 要么全都执行, 要么全都不执行。

C (Consistency) : **一致性**。指事务在完成时, 必须使所有的数据都保持一致状态。

I (Isolation) : **隔离性**。指并发事务所做的修改必须与其他并发事务所做的修改隔离。

D (Durability) : **持久性**。指事务完成之后, 它对于系统的影响是永久性的, 该修改即使出现致命的系统故障也将一直保持。

第六章 云数据库

1. 云数据库有哪些特性。

动态可扩展、高可用性、较低的使用代价、易用性、高性能、免维护、安全。

2. 试述 UMP 系统的功能。

UMP 系统构建在一个大的集群之上的, 通过多个组件的协同作业, 整个系统实现了对用户透明的 容灾、读写分离、分库分表、资源管理、资源调度、资源隔离和数据安全等功能。

1. 容灾

云数据库必须向用户提供一直可用的数据库连接, 当 MySQL 实例发生故障时, 系统必须自动执行故障恢复, 所有故障处理过程对于用户而言是透明的, 用户不会感知到后台发生的一切。

为了实现容灾, UMP 系统会为每个用户创建两个 MySQL 实例, 一个是主库, 一个是从库, 而且, 这两个 MySQL 实例之间互相把对方设置为备份机, 任意一个 MySQL 实例上面发生的更新都会复制到对方。同时, Proxy 服务器可以保证只向主库写入数据。

2. 读写分离

由于每个用户都有两个 MySQL 实例, 即主库和从库, 因此 UMP 系统可以充分利用主从库实现用户读写操作的分离, 实现负载均衡。UMP 系统实现了对于用户透明的读写分离功能, 当整个功能被开启时, 负责向用户提供访问MySQL数据库服务的 Proxy 服务器, 就会对用户发起的 SQL 语句进行解析, 如果是读操作, 就会被均衡地发送到主库和从库。

3. 分库分表



hurricane&&storm...

关注



建多组 MySQL 实例。除此以外，用户还需要自己设定分库分表规则，如需要确定分区字段，也就是根据哪个字段进行分库分表，还要确定分区字段里的值如何映射到不同的 MySQL 实例上。

4. 资源管理

UMP 系统采用资源池机制来管理数据库服务器上的 CPU、内存、磁盘等计算资源，所有的计算资源都放在资源池内进行统一分配，资源池是为 MySQL 实例分配资源的基本单位。整个集群中的所有服务器会根据其机型、所在机房等因素被划分为多个资源池，每台服务器会被加入到相应的资源池。在资源池划分的基础上，UMP 还在每台服务器内部采用 Cgroup 将资源进一步地细化，从而可以限制每个进程组使用资源的上限，同时保证进程组之间相互隔离。

5. 资源调度

UMP 系统中有 3 种规格的用户，分别是数据量和流量比较小的用户、中等规模用户以及需要分库分表的用户。多个小规模用户可以共享同一个 MySQL 实例。对于中等规模的用户，每个用户独占个 MySQL 实例。用户可以根据自己的需求来调整内存空间和磁盘空间，如果用户需要更多的资源，就可以迁移到资源有空闲或者具有更高配置的服务器上对于分库分表的用户，会占有多个独立的 MySQL 实例，这些实例既可以共存在同一台物理机上，也可以每个实例独占一台物理机。

UMP 通过 MySQL 实例的迁移来实现资源调度。借助于阿里集团中间件团队开发的愚公系统，UMP 可以实现在不停机的情况下动态扩容、缩容和迁移。

6. 资源隔离

当多个用户共享同一个 MySQL 实例或者多个 MySQL 实例共存在同一个物理机上时，为了保护用户应用和数据的安全，必须实现资源隔离，否则，某个用户过多消耗系统资源会严重影响到其他用户的操作性能。

7. 数据安全

数据安全性是让用户放心使用云数据库产品的关键，尤其是企业用户，数据库中存放了很多业务数据，有些属于商业机密，一旦泄露，会给企业造成损失。UMP 系统设计了多种机制来保证数据安全。

- 1.SSL 数据库连接。
- 2.数据访问 IP 白名单。
- 3.记录用户操作日志。
- 4.SQL 拦截。

3. 试述 UMP 系统的组件及其具体作用。

1. **Controller 服务器**：向 UMP 集群提供各种管理服务，实现集群成员管理、元数据存储、MySQL 实例管理、故障恢复、备份、迁移、扩容等功能。
2. **Web 控制台**：向用户提供系统管理界面。
3. **Proxy 服务器**：向用户提供访问 MySQL 数据库的服务。除了数据路由的基本功能外，Proxy 服务器中还实现了屏蔽 MySQL 实例故障、读写分离、分库分表、资源隔离、记录用户访问日志等功能。
4. **Agent 服务器**：管理每台物理机上的 MySQL 实例，执行主从切换、创建、删除、备份、迁移等操作，同时还负责收集和分析 MySQL 进程的统计信息、慢查询日志 (Slow Query Log) 和 bin-log。
5. **日志分析服务器**：存储和分析 Proxy 服务器一段时间内的慢日志和统计报表。





CSDN

博客

下载

学习

社区

GitCode

InsCode

GPU

7. 愚公系统：是一个进行增量复制的工具，它结合了全量复制和 bin-log 分析，可以实现在不停机的情况下动态扩容、缩容和迁移。

第七章 MapReduce

1. 试述 Map 函数和 Reduce 函数的输入、输出以及处理过程。

Map 函数的输入为分布式文件的文件块，这些文件块的格式是任意的。Map 函数将输入的元素转换成 `<key, value>` 形式的键值对，键和值的类型也是任意的。

Reduce 函数的输入是 Map 函数输出的结果即中间结果，其任务是将输入的一系列具有相同键的键值对以某种方式组合起来，输出处理后的键值对，输出结果会合并成一个文件。

2. 试述 MapReduce 的工作流程。（需包括提交任务、Map、Shuffle、Reduce 的过程）

- 1) MapReduce 框架使用 InputFormat 模块做 Map 前的预处理，然后将输入文件切分为逻辑上的多个 InputSplit。
- 2) 通过 RecordReader 根据 InputSplit 中的信息来处理 InputSplit 中的具体记录，加载数据并转换为适合 Map 任务读取的键值对，输入给 Map 任务。
- 3) Map 任务会根据用户自定义的映射规则，输出一系列的 `<key, value>` 作为中间结果。
- 4) Shuffle：对 Map 任务输出结果进行分区、排序、合并、归并等操作，得到 `<key, value-list>` 形式的中间结果，再交给对应的 Reduce 进行处理。
- 5) Reduce 以一系列 `<key, value-list>` 中间结果作为输入，执行用户定义的逻辑，输出结果给 OutputFormat 模块。
- 6) OutputFormat 模块会验证输出目录是否存在以及输出结果类型是否符合配置文件中的配置类型，如果都满足，就输出 Reduce 的结果到分布式文件系统。

☆☆☆ 3. Shuffle 过程是 MapReduce 过程的核心，也被称为奇迹发生的地方，试分析 Shuffle 过程的作用。（答案不全，看书 P135）

对 Map 任务输出结果进行分区、排序、合并、归并等处理并交给 Reduce 的过程，减少磁盘 I/O 的读写次数，并减小从 Map 到 Reduce 之间的数据传递量。

4. 早期版本的 HDFS，其默认块（Block）大小为何？用较大的块有什么影响和优缺点。



hurricane&&storm...

关注



CSDN

博客

下载

学习

社区

GitCode

InsCode

GPU

采用较小的块，说明 Map 任务较多，而创建多个 Map 任务进程需要耗费大量时间。块的大小设置主要从以下考虑：减少磁盘寻道时间、减少 NameNode 内存消耗、Map 崩溃问题、监管时间问题、问题分解问题、约束 Map 输出。

第八章 Hadoop 再探讨

1. 请描述 HDFS HA 架构组成组件及其具体功能。

在一个典型的 HA 集群中，一般设置两个名称节点，其中一个名称节点处于“活跃”状态，另一个处于“待命”状态。处于活跃状态的名称节点负责对外处理所有客户端的请求，处于待命状态的名称节点则作为备用节点，保存足够多的系统元数据，当名称节点出现故障时提供快速恢复能力。也就是说，在 HDFS HA 中，处于待命状态的名称节点提供了“热备份”，一旦活跃名称节点出现故障，就可以立即切换到待命名称节点，不会影响到系统的正常对外服务。

2. 请分析 HDFS HA 架构中数据节点如何和名称节点保持通信。

在 HDFS 联邦中，所有名称节点会共享底层的数据节点存储资源。每个数据节点要向集群中所有的名称节点注册，并周期性地向名称节点发送“心跳”和块信息，报告自己的状态，同时也会处理来自名称节点的指令。

3. 请阐述 MapReduce 1.0 体系结构中存在的问题。

- 1) 存在单点故障问题。
- 2) JobTracker “大包大揽”导致任务过重。
- 3) 容易出现内存溢出。
- 4) 资源划分不合理。

《大数据技术原理与应用》课程标准.pdf

06-18

《大数据技术原理与应用》课程标准.pdf 《大数据技术原理与应用》课程标准.pdf 《大数据技术原理与应用》...

大数据技术原理-----期末复习-基础知识

m0_46493223的博客 3251

一、大数据技术 1、试述数据产生方式经历的几个阶段。运营式系统阶段：在这个阶段，数据的产生方式是...

11 条评论



DavidZyy123

热评

总结的非常到位



hurricane&&storm...

关注



CSDN

博客

下载

学习

社区

GitCode

InsCode

GPU

!《PLC原理及应用》试卷-新.pdf

04-15

《PLC原理及应用》试卷-新 适合高等教育大学生学习的专业课复习**备考**，期末考试，考研复试的资料。

数据科学与大数据技术-专业导论.pptx

12-24

2021 数据科学与**大数据** 专业培养 / 专业前景 / 远期目标 汇报人： 制作时间：2021.11.7 数据211 数据科学与...

大数据技术原理与应用答案 林子雨 第二版 热门推荐

qq_43395427的博客 2万+

第一章 1.试述信息技术发展史上的3次信息化浪潮及具体内容。 信息化浪潮 发生时间 标志 解决问题 代表公...

大数据的特征 (4V+1O)

1万+

数据量大 (Volume)。第一个特征是数据量大，包括采集、存储和计算的量都非常大。**大数据**的起始计量单...

大数据技术原理与应用 (最后三天备考了!!!)

嘟嘟的博客 7795

大数据原理与应用期末**备考** 三天速成不挂科 导航**大数据原理与应用**期末**备考** 三天速成不挂科第一章 **大数据**...

大数据技术原理与运用习题总结

qq_59641085的博客 6410

第一章 **大数据**概述 1. 【单选题】**大数据**的4V特性不包括_____。 A、数据量大 B、数据类型繁多 C、数...

《大数据技术原理与应用》第三版 林子雨编著 第一章 大数据概述

xuyar123的博客 6022

《**大数据技术原理与应用**》第三版 林子雨编著 第一章 **大数据**概述 三次信息化浪潮P3。 信息化浪潮 发生时间...

多媒体应用设计师备考笔记.docx

11-01

多媒体**应用**设计师**备考**笔记

大数据技术原理与应用：概念、存储、处理、分析与应用 完整高清PDF

06-12

大数据技术原理与应用：概念、存储、处理、分析与**应用** 完整高清PDF

大数据技术原理及应用(20220212184941).pdf

02-13

大数据技术原理及应用(20220212184941).pdf

数据库原理与应用.rar

01-20

里面包含天津大学数据库课程的历年试卷，覆盖比较全面，可以用来**备考**。内容以数据库**原理**方向的试卷为...

软件设计师备考资料!!!! 最新发布

05-08

软考中级软件设计师**备考**资料，用于**备考**!

大数据技术原理与应用 (复习)

qq_52821374的博客 2990

大数据 第一章 **大数据**的四个基本特性 (p8) --4V 数据量大 (Volume) 数据类型繁多 (Variety) 处理速度快...

专业技术人员资格考试时间公布!备考人员请注意!.doc

11-19

专业技术人员资格考试时间公布!**备考**人员请注意!.doc

大数据技术大致包含哪些内容

lmseo5hy的博客 9199

关于**大数据**的概念，指的是无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合。而...

考计算机三级数据库技术怎么备考?

03-02

备考计算机三级数据库技术考试，可以按照以下步骤进行：

“相关推荐” 对你



hurricane&&storm...

关注

关于我们 招贤纳士 商务合作 寻求报道  400-660-0108  kefu@csdn.net  在线客服 工作时间 8:30-22:00

公安备案号11010502030143 京ICP备19004658号 京网文〔2020〕1039-165号 经营性网站备案信息
北京互联网违法和不良信息举报中心 家长监护 网络110报警服务 中国互联网举报中心 Chrome商店下载 账号管理规范 版权与免责声明
版权申诉 出版物许可证 营业执照 ©1999-2023北京创新乐知网络技术有限公司

 hurricane&&storming
码龄3年  高校学生

200

4万+

7337

17万+



原创

周排名

总排名

访问

等级

3569

464

727

254

4643

积分

粉丝

获赞

评论

收藏
































私信

关注

搜博文文章




热门文章

- 计算机网络期末复习考试 (最后三天备考)
 27530
- 学生信息管理系统 图形用户界面 (GUI)
java 实现对数据库的操作 数据库用的mysql
 20878
- 计算机网络期末复习考试 (最后三天备考!!!)
 17170
- SQL Server 触发器 (trigger) ----- 用例详解 (你需要举一反三的触发器实用方法都在这了)
 13999
- 大数据技术原理与应用 (最后三天备考了!!!)
 7786

分类专栏

 hurricane&&storm...

关注

	Linux	3篇
	Git	1篇
	SQL	7篇
		

最新评论

- 学生信息管理系统 图形用户界面 (GUI) ...
阿柒897: 这个要和MySQL链接吗?
- C++ string and string_view (这篇教你怎...
竹子酒: 我不是这个意思; string_view这个设计本身就是只读的, 并且从实现上来讲 ...
- C++ string and string_view (这篇教你怎...
hurricane&&storming: 看最后一个 string_view 修改字符串
- C++ string and string_view (这篇教你怎...
竹子酒: string_view怎么可以修改字符的呢?
- 学生信息管理系统 图形用户界面 (GUI) ...
Z454649: 主键可以重复

您愿意向朋友推荐“博客详情页”吗?











强烈不推荐

不推荐

一般般

推荐

强烈推荐

最新文章

- Sockets 编程 —— 基于Python的简单聊天应用, 包括一个聊天服务器和一个聊天客户端。服务器允许多个客户端连接, 并将来自一个客户端的消息广播给其他所有客户端。
- makefile 编写
- Machine Learning --- 机器学习 (斯坦福大学) 相关材料中英翻译
- 2023年 1篇
- 2022年 3篇
- 2021年 142篇
- 2020年 55篇



hurricane&&storm...

关注



CSDN

博客

下载

学习

社区

GitCode

InsCode



大数据原理与应用期末备考 三天速成不...

导航

第一章 大数据概述

第二章 大数据处理架构 Hadoop

第三章 分布式文件系统 HDFS

第四章 分布式数据库 HBase

第五章 NoSql 数据库

第六章 云数据库

第七章 MapReduce

第八章 Hadoop 再探讨



hurricane&&storm...

关注