

文本复制检测报告单(全文对照)

№:ADBD2018R_2017031021371720180109082018401512674802

检测时间:2018-01-09 08:20:18

检测文献: 1_赵炳_基于分布式图计算的大规模网络分析系统的研究

作者: 赵炳

检测范围: 中国学术期刊网络出版总库

中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库

中国重要会议论文全文数据库

中国重要报纸全文数据库

中国专利全文数据库

互联网资源(包含贴吧等论坛资源)

英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)

港澳台学术文献库

优先出版文献库

互联网文档资源

图书资源

CNKI大成编客-原创作品库

学术论文联合比对库

个人比对库

时间范围: 1900-01-01至2018-01-09

检测结果

总文字复制比: 2.2%

跨语言检测结果: 0%

去除引用文献复制比: 2.2%

去除本人已发表文献复制比: 2.2%

单篇最大文字复制比: 0.7% (基于Hadoop平台的广告检测系统研究与实现)

重复字数: [831]

总段落数: [7]

总字数: [37626]

疑似段落数: [3]

单篇最大重复字数: [257]

前部重合字数: [261]

疑似段落最大重合字数: [570]

后部重合字数: [570]

疑似段落最小重合字数: [85]



指标: ☐ 疑似剽窃观点 ☒ 疑似剽窃文字表述 ☐ 疑似自我剽窃 ☐ 疑似整体剽窃 ☐ 过度引用

表格: 1 脚注与尾注: 0

0% (0) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第1部分 (总745字)

3.8% (176) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第2部分 (总4596字)

1.1% (85) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第3部分 (总8051字)

5.2% (570) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第4部分 (总10952字)

0% (0) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第5部分 (总7566字)

0% (0) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第6部分 (总4907字)

0% (0) 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第7部分 (总809字)

(注释: 无问题部分 文字复制比部分 引用部分)

1. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第1部分

总字数: 745

相似文献列表 文字复制比: 0%(0) 疑似剽窃观点: (0)

2. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第2部分

总字数: 4596

相似文献列表 文字复制比: 3.8%(176) 疑似剽窃观点: (0)

1 面向云服务的超算中心资源综合调度关键技术研究

1.5% (70)

| | | |
|---|--|-----------------------|
| | 王春光(导师：吴泉源;吴庆波) - 《国防科学技术大学博士论文》 - 2015-01-01 | 是否引证：否 |
| 2 | 论农民工的社会保障权利 张会娟(导师：苗连营) - 《郑州大学硕士论文》 - 2012-05-01 | 1.3% (59) 是否引证：否 |
| 3 | 锂离子电池正极材料LiCoO ₂ 的改性及其薄膜制备研究 戴新义(导师：李晶泽) - 《电子科技大学博士论文》 - 2016-03-15 | 1.0% (45) 是否引证：否 |
| 4 | 计算机组卷与考试系统 吴焱焱(导师：姚望舒) - 《苏州大学硕士论文》 - 2016-11-01 | 1.0% (45) 是否引证：否 |

| | 原文内容 | 相似内容来源 |
|---|---|---|
| 1 | <p>此处有 73 字相似</p> <p>计算结果同步到其他有该节点备份的主机上面。文献[3]提出了一个基于图像处理的开源图计算框架 GraphLab，该框架是一个面向机器学习的流处理并行计算框架。GraphLab将数据抽象成Graph结构，并且将计算过程抽象成Gather、Apply、Setter三个阶段，Gather阶段所有节点从邻接节点接收数据，并且进行计算，Apply阶段将Gather节点计算得到的值发送到Master</p> | <p>面向云服务的超算中心资源综合调度关键技术研究王春光 - 《国防科学技术大学博士论文》 - 2015-01-01 (是否引证：否)</p> <p>1.Lab[90] 是 CMU 的第 23 页国防科学技术大学研究生院博士学位论文Select 实验室推出的面向机器学习的流处理并行框架，Graph Lab 将数据抽象成Graph 结构，将算法的执行过程抽象成 Gather、Apply、Scatter 三个步骤。其并行的核心思想是对顶点的切分，Graph Lab 对图算法的并行化实验所需时间成百倍的减少。Graph</p> |
| 2 | <p>此处有 44 字相似</p> <p>实现了一个完整可用的图计算平台，这个计算平台主要包括存储模块，计算模块，数据导入模块，结果展示模块，状态监控模块等。</p> <p>1.4 论文结构安排</p> <p>本文主要分为六个部分，每个部分按照下列结构进行组织：</p> <p>第一章，</p> <p>绪论。本章主要有研究背景，国内外研究现状，本文研究内容等部分。研究背景主要介绍了对于本文中研究内容的选题来源以及研究的必</p> | <p>锂离子电池正极材料LiCoO₂的改性及其薄膜制备研究 戴新义 - 《电子科技大学博士论文》 - 2016-03-15 (是否引证：否)</p> <p>1.了明显的充放电平台，电池容量为 20 ?Ah cm⁻²m⁻¹，具有较好的循环稳定性。1.6 本论文的结构安排本文主要分为七个部分，章节结构安排如下：第一章为绪论。介绍了本论文的研究背景，特别是锂离子电池正极材料尤其是 Li Co O₂的研究现状。第二章为材</p> <p>计算机组卷与考试系统 吴焱焱 - 《苏州大学硕士论文》 - 2016-11-01 (是否引证：否)</p> <p>1.业务流程，为进入社会做准备，也希望能同时把技术带入学校，也希望在以后的维护中，将系统更加完善。1.3 本文结构安排本文主要分为五个部分，各部分内容结构如下：第一章绪论主要讨论了本次毕业设计课题背景、研究价值。第二章开发环境主要介绍本系统采用的编程环境与架构。第三章系统</p> |
| 3 | <p>此处有 30 字相似</p> <p>本文主要分为六个部分，每个部分按照下列结构进行组织：</p> <p>第一章，绪论。本章主要有研究背景，国内外研究现状，本文研究内容等</p> <p>部分。研究背景主要介绍了对于本文中研究内容的选题来源以及研究</p> <p>的必要性。国内外研究现状主要描述了国内外的学术界和工业界在这个问题上的最新研究，主要包括图存储系统，图计算系统和动态图计</p> | <p>论农民工的社会保障权利 张会娟 - 《郑州大学硕士论文》 - 2012-05-01 (是否引证：否)</p> <p>1.我国农民工社会保障权利的研究以及法律保护奉献自己的绵薄之力。★在论文的结构上,本文做了如下的安排：★第一部分引言,主要简单介绍了本文的选题背景,研究方法以及具体内容的安排。★第二部分,较为详细地介绍了我国目前农民工群体的现状以及我国的社会保障体系,进而引出对农</p> |
| 4 | <p>此处有 29 字相似</p> <p>中的主要研究点以及最后达到的效果，最后的论文组织结构则是对全文的论文组织做了综合概述。</p> <p>第二章，关键技术。关键结束方面</p> <p>主要介绍了本文中用到的一些前沿研究技术，以及这些</p> | <p>论农民工的社会保障权利 张会娟 - 《郑州大学硕士论文》 - 2012-05-01 (是否引证：否)</p> <p>1.社会保障权利的研究以及法律保护奉献自己的绵薄之力。★在论文的结构上,本文做了如下的安排：★第一部分引言,主要简单介绍了本文的选题背景,研究方法以及具体内容的安排。★第二部分,较为详细地介绍了我国目</p> |

| | | |
|--|--|--|
| | <p>技术现在的</p> <p>发展状况。主要包括图存储的实现方式，分布式图存储和单机图存储，以及分布式图存储涉及到的图分割技术。图计算方面主要包括图</p> | <p>前农民工群体的现状以及我国的社会保障体系,进而引出对农民工社会保障问题的探讨。</p> |
|--|--|--|

| 指 标 |
|---|
| 疑似剽窃文字表述 |
| <div>1. GraphLab将数据抽象成Graph结构，并且将计算过程抽象成Gather、Apply、Setter三个阶段，</div> <div>2. 1.4 论文结构安排</div> <div>本文主要分为六个部分，每个部分按照下列结构进行组织：</div> <div>第一章，</div> |

| 3. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第3部分 | 总字数：8051 |
|------------------------------------|-------------|
| 相似文献列表 文字复制比：1.1%(85) 疑似剽窃观点：(0) | |
| 1 黄永建1 | 1.1% (86) |
| - 《学术论文联合比对库》- 2012-05-03 | 是否引证：否 |
| 2 黄永建 | 1.1% (86) |
| - 《学术论文联合比对库》- 2012-05-07 | 是否引证：否 |
| 3 黄永建2 | 1.1% (86) |
| - 《学术论文联合比对库》- 2012-05-08 | 是否引证：否 |
| 4 黄永建 | 0.5% (43) |
| - 《学术论文联合比对库》- 2012-11-02 | 是否引证：否 |

| 原文内容 | 相似内容来源 |
|--|--|
| <div>此处有 43 字相似</div> <div>角或者下三角，可以节省存储空间；</div> <div>3. 对于一个无向图来说，邻接矩阵的第i行（或者是第i列）非零元素的个数正好是第i个顶点的度；</div> <div>4. 对于一个有向图来说，邻接矩阵的第i行（或者是第i列）非零元素的个数正好是第i个顶点的出度（入度）；</div> <div>图1-1 无向图的邻接矩阵存储</div> <div>对于一个无向图，如图1-1所示，邻接矩阵的主对角线</div> | <div>黄永建1 - 《学术论文联合比对库》- 2012-05-03 (是否引证：否)</div> <div>1.下三角或者矩阵的上三角的元素。通过对邻接矩阵的访问我们很容易得到图中任意两个顶点之间是否有边或者弧相连，并且可以得到各个顶点的度，对于无向图来说，顶点vi的度是邻接矩阵中第i行或者第i列的元素个数之和。对于有向图来说，则第i行的元素个数之和为顶点vi的出度，第j列的元素个数之和为顶点vi入度。2.3.2邻接表</div> <div>黄永建 - 《学术论文联合比对库》- 2012-05-07 (是否引证：否)</div> <div>1.下三角或者矩阵的上三角的元素。通过对邻接矩阵的访问我们很容易得到图中任意两个顶点之间是否有边或者弧相连，并且可以得到各个顶点的度，对于无向图来说，顶点vi的度是邻接矩阵中第i行或者第i列的元素个数之和。对于有向图来说，则第i行的元素个数之和为顶点vi的出度，第j列的元素个数之和为顶点vi入度。2.2.2邻接表</div> <div>黄永建2 - 《学术论文联合比对库》- 2012-05-08 (是否引证：否)</div> <div>1.下三角或者矩阵的上三角的元素。通过对邻接矩阵的访问我们很容易得到图中任意两个顶点之间是否有边或者弧相连，并且可以得到各个顶点的度，对于无向图来说，顶点vi的度是邻接矩阵中第i行或者第i列的元素个数之和。对于有向图来说，则第i行的元素个数之和为顶点vi的出度，第j列的元素个数之和为顶点vi入度。2.2.2邻接表</div> |
| 1 | |

| | |
|--|--|
| | 黄永建 - 《学术论文联合比对库》 - 2012-11-02 (是否引证 : 否) |
| | 1.下三角或者矩阵的上三角的元素。通过对邻接矩阵的访问我们很容易得到图中任意两个顶点之间是否有边或者弧相连,并且可以得到各个顶点的度,对于无向图来说,顶点 v_i 的度是邻接矩阵中第 i 行或者第 i 列的元素个数之和。对于有向图来说,则第 i 行的元素个数之和为顶点 v_i 的出度,第 j 列的元素个数之和为顶点 v_i 入度。邻接表 (Adjacen |

4. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第4部分 总字数 : 10952

| | | |
|---------------------------------------|--|--------------------------|
| 相似文献列表 文字复制比 : 5.2%(570) 疑似剽窃观点 : (0) | | |
| 1 | 基于Hadoop平台的广告检测系统研究与实现 杨宁(导师 : 李伟) - 《复旦大学硕士论文》 - 2012-09-27 | 2.3% (257) 是否引证 : 否 |
| 2 | 大数据 : Hbase原理、基本概念、基本架构_CDerek - 《网络 (http://blog.sina.com) 》 - 2015 | 1.8% (202) 是否引证 : 否 |
| 3 | Hbase原理、基本概念、基本架构 - 坦GA的博客 - 博客频道 - CSDN.NET - 《网络 (http://blog.csdn.net) 》 - 2017 | 1.8% (202) 是否引证 : 否 |
| 4 | 大数据(五) - HBase - IT十年 - 博客频道 - CSDN.NET - 《网络 (http://blog.csdn.net) 》 - 2017 | 1.4% (157) 是否引证 : 否 |
| 5 | 眼科专科影像云服务平台研发 张旭东(导师 : 宋文爱;王青) - 《中北大学硕士论文》 - 2017-04-11 | 0.8% (93) 是否引证 : 否 |
| 6 | 一种基于Sqoop的数据交换系统 于金良;朱志祥;梁小江; - 《物联网技术》 - 2016-03-20 | 0.8% (86) 是否引证 : 否 |
| 7 | 基于图计算模型的矩阵分解并行化研究 戴世超(导师 : 包晓安) - 《浙江理工大学硕士论文》 - 2016-04-14 | 0.8% (83) 是否引证 : 否 |
| 8 | 一种基于NoSQL数据库的医疗监护大数据存储设计 黎茂林;张鹤;吕德奎; - 《电脑与信息技术》 - 2016-12-15 | 0.5% (53) 是否引证 : 否 |
| 9 | 基于有向图的虚开增值税发票行为检测方法研究 刘丽萍(导师 : 罗晓霞;董富强) - 《西安科技大学硕士论文》 - 2017-06-01 | 0.4% (42) 是否引证 : 否 |
| 10 | 大数据分析技术在风电设备异常预测中的应用 张慧亭;王坚;凌卫青; - 《电脑知识与技术》 - 2017-03-23 1 | 0.3% (36) 是否引证 : 否 |
| 11 | 我国经济政策审计评价研究 王延军(导师 : 张通) - 《财政部财政科学研究所博士论文》 - 2015-05-22 | 0.3% (36) 是否引证 : 否 |
| 12 | 基于云平台的全域匿名算法的研究与实现 胡庆庆(导师 : 刘君强) - 《浙江工商大学硕士论文》 - 2015-01-01 | 0.3% (36) 是否引证 : 否 |
| 13 | 组学大数据的检索系统设计与实现 徐康(导师 : 臧天仪) - 《哈尔滨工业大学硕士论文》 - 2015-06-01 | 0.3% (34) 是否引证 : 否 |

| 原文内容 | | 相似内容来源 |
|------|--|--|
| 1 | <p>此处有 81 字相似</p> <p>某一时刻快照,为此我们想的是给图中的节点和边数据都加上时间(timestamp)属性,用来代表数据的版本。在这里我们采用</p> <p>Hbase作为存储后端。HBase是一个构建在HDFS上的分布式存储系统,是基于Google Bigtable模型开发的典型的key value存储系统。HBas</p> <p>e设计的初衷就是实现HDFS上的海量数据的存储,因此设计的理念也不同于一般的数据库,他是一个很适合存储非结构化海量数据存</p> | <p>大数据 : Hbase原理、基本概念、基本架构_CDerek - 《网络 (http://blog.sina.com) 》 - (是否引证 : 否)</p> <p>1.要处理,因此会Replay HLog中的数据到MemStore中,然后flush到StoreFiles,完成数据恢复。HBase是一个构建在HDFS上的分布式列存储系统; HBase是基于Google BigTable模型开发的,典型的key/value系统; HBase是Apache Hadoop生态系统中的重要一员,主要用于海量结构化数据存储; 从逻辑上讲, HBase将数据按照表、行</p> <p>Hbase原理、基本概念、基本架构 - 坦GA的博客 - 博客频道 - CSDN.NET - 《网络 (http://blog.csdn.net) 》 - (是否引证 : 否)</p> <p>1.HBase是一个构建在HDFS上的分布式列存储系统; HBase是基于Google BigTable模型开发的,典型的</p> |

| | | |
|---|---|--|
| | | <p>key/value系统；HBase是Apache Hadoop生态系统中的重要一员，主要用于海量结构化数据存储；从逻辑上讲，HBase将数据按照表、行</p> <p>大数据(五) - HBase - IT十年 - 博客频道 - CSDN.NET - 《网络 (http://blog.csdn.net) 》 - (是否引证：否)</p> <p>1.HBase是一个构建在HDFS上的分布式列存储系统，主要用于海量结构化数据存储，从逻辑上讲，HBase将数据按照表、行和列进行存储。HDFS适合批处理场景 不支持数据随机</p> <p>大数据分析技术在风电设备异常预测中的应用 张慧亭;王坚;凌卫青; - 《电脑知识与技术》 - 2017-03-23 1 (是否引证：否)</p> <p>1.行标准化的调整,减少了人为的序列/反序列化操作。 2)存储层。本文主要采用HBase,Hive等分布式数据库作为存储介质,HBase是一个构建在HDFS上的分布式列存储系统,它具有高可靠、高性能以及可伸缩等特点,可以方便地在服务器上搭建起大规模结构化存储集群。 Hive是基于Hadoop的一个数</p> <p>我国经济政策审计评价研究 王延军 - 《财政部财政科学研究所博士论文》 - 2015-05-22 (是否引证：否)</p> <p>1.的数据文件映射为一张数据库表,并提供简单的SQL查询功能(HSQL),可将SQL语句转换为MapReduce任务进行执行。HBase是一个构建在HDFS上的分布式列存储系统,数据存储自动能水平扩展,提供高并发读写操作,主要用于海量结构化数据存储和查询;缺点是不能支持复杂条件查询,只支持按照R</p> |
| 2 | <p>此处有 58 字相似</p> <p>le模型开发的典型的key value存储系统。HBase设计的初衷就是实现HDFS上的海量数据的存储，因此设计的理念也</p> <p>不同于一般的数据库，他是一个很适合存储非结构化海量数据存储的数据库，主要特点是它是基于列的存储，而不是基于行的存储，</p> <p>这样的存储结构可以方便读取海量数据的内容。</p> <p>HBase的基本存储结构就是类似于哈希表的概念，但又不仅仅是简单的一个key对</p> | <p>眼科专科影像云服务平台研发 张旭东 - 《中北大学硕士论文》 - 2017-04-11 (是否引证：否)</p> <p>1.ble 的能力。Hbase 是 Apache24中北大学学位论文的 Hadoop 项目的子项目。Hbase 不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。另一个不同的是 Hbase 基于列的而不是基于行的模式。使用 Hbase 来存储眼科数据具有以下优点：(1) 可存储大量的眼科数据：一个表可以存储上亿行、上百万列的眼科数据</p> <p>基于Hadoop平台的广告检测系统研究与实现 杨宁 - 《复旦大学硕士论文》 - 2012-09-27 (是否引证：否)</p> <p>1.存放Web文档及其以URL作为关键字的属性宽泛的表。HBase不同于传统的关系数据库，它是适合于非结构化数据存储的数据库。所谓非结构化数据存储就是说HBase是基于列的而不是基于行的模式。这里简要介绍一下列存储。基于列存储的数据库中的数据表的每列单独存放；数据即使索引；查询时只访问涉及的列，大量</p> <p>一种基于Sqoop的数据交换系统 于金良;朱志祥;梁小江; - 《物联网技术》 - 2016-03-20 (是否引证：否)</p> <p>1.看这些数据,还可以通过Hadoop分布式计算框架Map Reduce对HBase进行操作,如查询、插入数据等。 HBase不同于一般的关系型数据库,它是一个适合于非结构化数据存储的数据库。所谓非结构化数据存储即HBase是基于列存储而不是基于行存储的。且其没有关系型数据库表中那么多的关联关系,采用稀疏的存储</p> |

| | | |
|---|---|--|
| 3 | <p>此处有 147 字相似</p> <p>存储，而不是基于行的存储，这样的存储结构可以方便读取海量数据的内容。HBase的基本存储结构就是类似于哈希表的概念，但又不仅仅是简单的一个key对应一个value。每个key对应的可能是多个属性的数据结构，但是又没有传统数据库中那么多的关联关系，也可以叫做是松散数据。简单的说，在HBase中存储的数据可以看做是一张巨大的表，这个表的属性是可以根据我们的需求动态增加的，在HBase中没有表与表之间的关联查询。在存储数据的时候我们只要制定了要存储数据的column和family即可，不用关心数据的类型，在HBase中所有的数据都</p> | <p>基于Hadoop平台的广告检测系统研究与实现 杨宁 - 《复旦大学硕士论文》 - 2012-09-27 (是否引证：否)</p> <p>1.并发处理。HBase是介于Map Entry(key & value)和DB Row之间的一种数据存储方式。不仅是一个key对应一个value，还可能需要存储多个属性的数据结构，但没有传统数据库表中那么多的关联关系，即所谓松散数据。简单来说，HBase中创建的表的属性可以根据需求去动态增加，在HBase中没有表与表之间关联查询。Apache HBase的一个数据行拥有一个可选择的键和任意数量的列。表是疏松存储的，因此用户可以给行定义各种不同</p> <p>一种基于NoSQL数据库的医疗监护大数据存储设计 黎茂林;张鹤;吕德奎; - 《电脑与信息技术》 - 2016-12-15 (是否引证：否)</p> <p>1.据库HBase进行大数据存储。4医疗实时监护大数据存储设计4.1实现过程概述根据HBase特性,在HBase中的表创建的可以看成是一张很大的表,这个表的属性可以根据需求去动态增加,在HBase中没有表与表之间关联查询。只需要指定数据存储到HBase的Column Family、Column就可以,Column也不需要指定它的具体类型:c</p> <p>一种基于Sqoop的数据交换系统 于金良;朱志祥;梁小江; - 《物联网技术》 - 2016-03-20 (是否引证：否)</p> <p>1.e是基于列存储而不是基于行存储的。且其没有关系型数据库表中那么多的关联关系,采用稀疏的存储结构。即HBase中的表创建的可以看做是一张很大的表,这个表的属性可以根据需求去动态增加,在HBase中没有表与表之间的关联查询,只需要告诉用户其数据存储在哪一个列族就可以了。向HBase中导入数据使用命令:sudo-u hdfs sqoop impo</p> <p>组学大数据的检索系统设计与实现 徐康 - 《哈尔滨工业大学硕士论文》 - 2015-06-01 (是否引证：否)</p> <p>1.需求,当达到最大数目的时候,数据被持久化到HDFS中[25]。用户可以在HBase里创建一个很大的表格,并且这个表格的列可以动态的实现增加,在HBase里没有表与表之间的关联查询。只需要制定要保存到的列的列族的名称,不用特别的指定要存储的具体类型。HBase一行由一个rowkey键及无数的列组成,是</p> |
| 4 | <p>此处有 58 字相似</p> <p>Base存在着如下的特点： HBase可以存储的数据规模非常大，一个表可以由数十亿行和上百万列； 每行的数据 都有一个可排序的主键和任意多的列，并且列的数量是可以根据需求动态增加的，同一张表中的不同数据可以有完全不同的列。</p> <p>面向列的存储和控制，面向列的独立检索； 空列并不会占用存储单元，这样可以让我们的表设计的非常稀疏 每个单</p> | <p>大数据：Hbase原理、基本概念、基本架构 CDerek - 《网络 (http://blog.sina.com) 》 - (是否引证：否)</p> <p>1.增加廉价的商用服务器，来增加计算和存储能力。 Hbase表的特点 大：一个表可以有数十亿行，上百万列； 无模式：每行都有一个可排序的主键和任意多的列，列可以根据需要动态的增加，同一张表中不同的行可以有截然不同的列； 面向列：面向列（族）的存储和权限控制，列（族）独立检索； 稀疏：空（null）列并不占用存储空间，表可以设计的非常稀疏</p> <p>Hbase原理、基本概念、基本架构 - 坦GA的博客 - 博客频道 - CSDN.NET - 《网络 (http://blog.csdn.net) 》 - (是否引证：否)</p> <p>1.增加廉价的商用服务器，来增加计算和存储能力。 Hbase表的特点 大：一个表可以有数十亿行，上百万列</p> |

| | | |
|---|---|---|
| | | <p>；无模式：每行都有一个可排序的主键和任意多的列，列可以根据需要动态的增加，同一张表中不同的行可以有截然不同的列；面向列：面向列（族）的存储和权限控制，列（族）独立检索；稀疏：空（null）列并不占用存储空间，表可以设计的非常稀疏</p> <p>大数据(五) - HBase - IT十年 - 博客频道 - CSDN.NET - 《网络（http://blog.csdn.net）》 - （是否引证：否）</p> <p>1.上千次操作，并且读写访问均是非常简单的操作。 Hbase表的特点 大：一个表可以有数十亿行，上百万列；无模式：每行都有一个可排序的主键和任意多的列，列可以根据需要动态的增加，同一张表中不同的行可以有截然不同的列；面向列：面向列（族）的存储和权限控制，列（族）独立检索；稀疏：对于空（null）的列，并不占用存储空间，表可以设计的</p> <p>基于云平台的全域匿名算法的研究与实现 胡庆庆 - 《浙江工商大学硕士论文》 - 2015-01-01（是否引证：否）</p> <p>1.B级别）的高速读写而设计，所以HBase的一张数据表的容量可以达到很高的程度，可以有数十亿行，上百列。数据表的每一行都有一个可以排序的主键和任意多的列，列可以根据需要动态地增加，而且同一张数据表中，不同的行可以有截然不同的列。从上可以得知，HBase是面向列（族）的存储，而不是面向行</p> |
| 5 | <p>此处有 70 字相似</p> <p>以根据需求动态增加的，同一张表中的不同数据可以有完全不同的列。</p> <p>面向列的存储和控制，面向列的独立检索；空列</p> <p>并不会占用存储单元，这样可以让我们的表设计的非常稀疏</p> <p>每个单元中可以存储多版本的数据，在默认情况下版本号是自动分配的，是插入的时间戳，并且可以由用户定义。</p> <p>数据类型单一，HBase中存储的数据都是字符串的形式，并不区分他们的具体类型。</p> <p>因为Hba</p> | <p>大数据：Hbase原理、基本概念、基本架构 CDerek - 《网络（http://blog.sina.com）》 - （是否引证：否）</p> <p>1.同一张表中不同的行可以有截然不同的列；面向列：面向列（族）的存储和权限控制，列（族）独立检索；稀疏：空（null）列并不占用存储空间，表可以设计的非常稀疏；数据多版本：每个单元中的数据可以有多个版本，默认情况下版本号自动分配，是单元格插入时的时间戳；数据类型单一：Hbase中的数据都是字符串，没有类型。Hbase数据模型 Hbase逻辑视图 注意上图中的英文说明</p> <p>Hbase原理、基本概念、基本架构 - 坦GA的博客 - 博客频道 - CSDN.NET - 《网络（http://blog.csdn.net）》 - （是否引证：否）</p> <p>1.同一张表中不同的行可以有截然不同的列；面向列：面向列（族）的存储和权限控制，列（族）独立检索；稀疏：空（null）列并不占用存储空间，表可以设计的非常稀疏；数据多版本：每个单元中的数据可以有多个版本，默认情况下版本号自动分配，是单元插入时的时间戳；数据类型单一：Hbase中的数据都是字符串，没有类型。Hbase逻辑视图 注意上图中的英文说明 Hbase基本概念</p> <p>大数据(五) - HBase - IT十年 - 博客频道 - CSDN.NET - 《网络（http://blog.csdn.net）》 - （是否引证：否）</p> <p>1.中不同的行可以有截然不同的列；面向列：面向列（族）的存储和权限控制，列（族）独立检索；稀疏：对于空（null）的列，并不占用存储空间，表可以设计的非常稀疏；数据多版本：每个单元中的数据可以有多个版本，默认情况下版本号自动分配，是单元插入时的时间戳；数据类型单一：Hbase中的数据都是字符串</p> |

| | | |
|---|--|---|
| | | <p>，没有类型。行存储和列存储的比较 传统行式数据库数据是按行存储的 没有索</p> <p>眼科专科影像云服务平台研发 张旭东 - 《中北大学硕士论文》 - 2017-04-11 (是否引证：否)</p> <p>1.族) 独立检索；(4) 空不占空间：空 (null) 列并不占用存储空间，表可以设计的非常稀疏；(5) 数据可以有多个版本：每个单元中的数据可以有多个版本，默认情况下版本号自动分配，是单元格插入时的时间戳；(6) 数据类型单一：Hbase 数据库没有类型，所有的数据均是字符串[22]。3.1.3 分布式眼科数据平</p> <p>基于Hadoop平台的广告检测系统研究与实现 杨宁 - 《复旦大学硕士论文》 - 2012-09-27 (是否引证：否)</p> <p>1.，上西刀列2面向列：面向列(族)的存储和权限控制，列(族；)独立检索。3稀疏：对于为空(null)的列，并+占用存储空间，因此，表可以设计的非常稀疏。 HBase表中存储的行记录有三个基本类型的定义：Row Key, Time Stamp,30基于Hadoop平台的广告检测</p> |
| 6 | <p>此处有 31 字相似</p> <p>在查询的时候我们需要指定数据的Version Number。Version可以由用户自己定义，有两种定义version的方式，第一种是保存数据的n个版本，二是保存近一段时间内的版本。</p> <p>对于我们的图模型来说，每增加或者修改节点或者边，相应的修改节点和边的版本号为最新版本。每删除一个节点或者边就不更新版本号</p> | <p>基于Hadoop平台的广告检测系统研究与实现 杨宁 - 《复旦大学硕士论文》 - 2012-09-27 (是否引证：否)</p> <p>1.据排在最前面。为了避免数据存在过多版本造成的的管理 (包括存5^和索引)负担，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 (比如最近七天)。用户可以针对每个列族进行设置。31基于Hadoop平台的广告检测系统 由{row k</p> |
| 7 | <p>此处有 42 字相似</p> <p>还存储了顶点到边RDD的路由表。路由表是顶点RDD中的一个特殊数据结构，它记录了顶点RDD中所有的顶点和边RDD的对应该</p> <p>关系。在计算过程中边RDD需要顶点的数据的时候，顶点RDD会根据路由表将顶点数据发送</p> <p>到边的RDD分区。顶点RDD和边RDD的分布如下图所示。</p> <p>在大部分的图计算过程中，边的计算都需要两端顶点的数据，即形成</p> | <p>基于有向图的虚开增值税发票行为检测方法研究 刘丽萍 - 《西安科技大学硕士论文》 - 2017-06-01 (是否引证：否)</p> <p>1.n Strategy) 进行分区，将边数据以多分区形式分布在集群上。路由表记录分区内顶点跟所有边 RDD分区的关系。在边 RDD 需要顶点数据时，顶点 RDD 会根据路由表把顶点数据发送至边RDD 分区。</p> <p>24651133Part.1Part.2Prop</p> |
| 8 | <p>此处有 83 字相似</p> <p>起来，形成三元组。在形成三元组的过程中，只有根据顶点RDD形成的重复顶点视图需要在不同的边分区之间移动，merge操作不</p> <p>需要移动顶点数据和边数据。而且在大部分图中顶点数据都是远远的小于边数据的，随着迭代次数的增加，需要进行更新的边数据也越来越少，这样可以大大的减少数据的移动量，从而加快</p> <p>整个计算过程。</p> <p>Spark在顶点和边RDD的存储中采用数组的方式存储顶点数据和边数据，这样做可以减少访问性能的下降。S</p> | <p>基于图计算模型的矩阵分解并行化研究 戴世超 - 《浙江理工大学硕士论文》 - 2016-04-14 (是否引证：否)</p> <p>1.n 操作不需要对数据进行 shuffle操作。在整个运算过程中，只有在创建重复顶点视图时需要移动顶点数据，边数据不需要移动。由于顶点数据一般比边数据要少得多，而且随着迭代次数的增加，需要更新的顶点数目也越来越少，这样就可以大大减少数据的移动量，从而加快执行速度。浙江理工大学硕士学位论文 基于图计算模型的矩阵分解并行化研究27重复顶点视图在创建之后即被装载进内</p> |

| | | |
|--|--|--|
| | | |
|--|--|--|

| 指 标 |
|--|
| 疑似剽窃文字表述 |
| <ol style="list-style-type: none"> 1. Hbase作为存储后端。HBase是一个构建在HDFS上的分布式存储系统，是基于Google Bigtable模型开发的典型的key value存储系统。HBas 2. 不同于一般的数据库，他是一个很适合存储非结构化海量数据存储的数据库，主要特点是它是基于列的存储，而不是基于行的存储， 3. 不仅仅是简单的一个key对应一个value。每个key对应的可能是多个属性的数据结构，但是又没有传统数据库中那么多的关联关系，也可以叫做是松散数据。 简单的说，在HBase中存储的数据可以看做是一张巨大的表，这个表的属性是可以根据我们的需求动态增加的，在HBase中没有表与表之间的关联查询。 4. 都有一个可排序的主键和任意多的列，并且列的数量是可以根据需求动态增加的，同一张表中的不同数据可以有完全不同的列。 5. 并不会占用存储单元，这样可以让我们的表设计的非常稀疏 每个单元中可以存储多版本的数据，在默认情况下版本号是自动分配的，是插入的时间戳， 6. 关系。在计算过程中边RDD需要顶点的数据的时候，顶点RDD会根据路由表将顶点数据发送 7. 需要移动顶点数据和边数据。而且在大部分图中顶点数据都是远远的小于边数据的，随着迭代次数的增加，需要进行更新的边数据也越来越少，这样可以大大的减少数据的移动量，从而加快 |

| | |
|------------------------------------|----------|
| 5. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第5部分 | 总字数：7566 |
| 相似文献列表 文字复制比：0%(0) 疑似剽窃观点：(0) | |

| | |
|------------------------------------|----------|
| 6. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第6部分 | 总字数：4907 |
| 相似文献列表 文字复制比：0%(0) 疑似剽窃观点：(0) | |

| | |
|------------------------------------|---------|
| 7. 1_赵炳_基于分布式图计算的大规模网络分析系统的研究_第7部分 | 总字数：809 |
| 相似文献列表 文字复制比：0%(0) 疑似剽窃观点：(0) | |

| 表格检测详细结果 |
|----------|
|----------|

| | | |
|-------------------------|---------|---------|
| 原文表格1：未获取到表格标题 共有1个相似表格 | | |
| follow | | |
| id | person1 | person2 |
| 1 | 1 | 2 |
| 2 | 3 | 1 |
| 3 | 4 | 1 |

| | | | | | |
|--|---|---|---|---|---|
| 相似表格1：表2湿气对2种体系附着力的影响 | | | | | |
| 相似度：69.23% | | | | | |
| 来源：一种用于激光固化快速成形的低翘曲光敏树脂的研究-段玉岗,王学让,王素琴,李涤尘,卢秉恒-《西安交通大学学报》-2001-11-20 | | | | | |
| 放置时间/周 | 0 | 1 | 2 | 3 | 4 |
| 阳离子体系/级 | 1 | 2 | 3 | 4 | 5 |
| 混杂体系/级 | 1 | 1 | 1 | 1 | 1 |

| |
|---|
| 说明：1.总文字复制比：被检测论文总重合字数在总字数中所占的比例 |
| 2.去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例 |
| 3.去除本人已发表文献复制比：去除作者本人已发表文献后，计算出来的重合字数在总字数中所占的比例 |
| 4.单篇最大文字复制比：被检测文献与所有相似文献比对后，重合字数占总字数的比例最大的那一篇文献的文字复制比 |
| 5.指标是由系统根据《学术论文不端行为的界定标准》自动生成的 |
| 6.红色文字表示文字复制部分;绿色文字表示引用部分 |
| 7.本报告单仅对您所选择比对资源范围内检测结果负责 |



 <http://check.cnki.net/>

 <http://e.weibo.com/u/3194559873>

CNKI科研诚信管理系统研究中心