SQL 和 NoSQL 数据库软件架构性能分析和评估

王　晖
哈尔滨信息工程学院　黑龙江哈尔滨　150000

摘　要：合格的软件架构在 SQL 和 NoSQL 数据库的大数据处理这一艰巨任务中起着至关重要的作用。创 建 SQL 数据库是为了组织数据并允许水平扩展。另一方面，NoSQL 数据库支持水平可伸缩性，可以高 效地处理大量非结构化数据。组织需求决定了哪种范式是合适的，然而选择最佳选项并不总是容易 的。数据库设计的差异是 SQL 和 NoSQL 数据库的不同之处。每种 NoSQL 数据库类型也始终采用混合模 型方法。因此，云用户在不同的云存储服务（CSP）之间传输数据具有挑战性。各种云平台 （IaaS、PaaS、SaaS 和 DBaaS）正在监控几种不同的范例。这篇 SLR 的目的是研究解决云数据可移 植性和互操作性的文章，以及 SQL 和 NoSQL 数据库的软件架构。作为最新技术的一部分，介绍了许 多比较数据库的 SQL 和 NoSQL 功能的研究，特别是 Oracle RDBMS 和 NoSQL 文档数据库（MongoDB）在 规模、性能、可用性、一致性和分片方面的研究。研究表明，具有专门定制结构的 NoSQL 数据库可 能是大数据分析的最佳选择，而 SQL 数据库最适合在线事务处理（OLTP）目的。

关键词： 大数据；SQL 和 NoSQL 数据库；MapReduce 聚合；数据库即服务

1. 介绍

特定软件应用程序的架构处理非功能性特征，如可靠性、可用性、可伸缩性、性能、 互操作性、可移植性、适应性和数据分片。在一组质量属性之间总是有权衡，软件架构师 面临着平衡它们的困难任务。大数据系统本质上是分布式的。数据可用性和一致性困 难是由庞大数据系统中的数据分片和复制造成的。由于数据应用的日益扩展，数据库技术 经历了巨大的变化。在十多年的过程中，NoSQL 数据库呈指数级增长，尽管传统的数据库 自动化一直存在。传统模型强加了一个僵化的模式结构，这导致伸缩模糊不清，并抑制了 跨集群的数据修改。相比之下，NoSQL 数据库支持简单的原型。NoSQL 数据库设计的主要特性包括：无模式结构允许数据表示有效且动态地增长通过数据复制收集和分片在大规模集群上进行水平展。

近年来，许多组织积累了大量数据，关系数据库无法有效处理这些数据。在过去的四十年里关系数据库的使用大幅增加。它们遵循 ACID（可用性、一致性、隔离性和持久性）属性，并且是为结构化数据设计的。虽然“大数据”包括管理任何规模的海量数据的工具和技术，但这些工具和技术是可扩展的。大数据包括 5v（容量、速度、多样性、准确性和价 值）以及大量具有不同性质的非结构化数据。大量框架用于大型数据处理，包括 Hadoop/ MapReduce、Spark、Flink 和 Samza。 企业、生产、并行数据库和大数据的 SQL 查询性能和优化主题近年来受到越来越多的关注。无效和未优化的查询可能会消耗系统和服务器资源，导致数据库锁定和数据 丢失问题。信息挖掘需要从原始数据集中提取事实和逻辑关联结构，而不是信息本身。查 询优化是指以最小的成本和系统资源消耗选择最佳的查询执行策略。数据挖掘算法进行深 入和广泛的数据库查询，以从全面的数据中提取模式和知识。XML 和对象数据库等替代 方法从未像 RDBMS 技术那样普及。 在过去的十年中，科学和在线供应商一直质疑数据商店技术的“一刀切”性质。这种思路导致了一种新的替代数据库系统的开发，该系统被称为 NoSQL，它代表“不仅是 SQL”NoSQL 描述了 web 开发人员对非关系数据库的使用。1998 年，NoSQL 一词第一次被使用，旧金山的非关系数据库会议引起了更多的关注。

2. 相关问题

为什么 NoSQL 数据库遵循 base 属性而不是 SQL 数据库 ACID 属性？

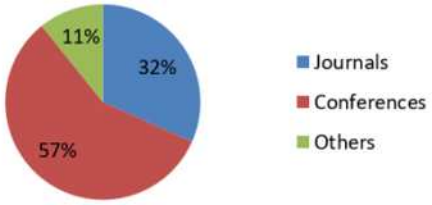
DBaaS 在各种 NoSQL 数据库中有效地解决了数据互操作性和可移植性问题吗？

在第一阶段，我们导出了一组相关的搜索字符串。我

们使用派生的字符串集来查找相 关论文。选择最大的数据库来查找相关文章

3. 结果

在这一节中，我们根据出版年份、论文类型和从特定数字图书馆中选择的研究数量总 结了我们选择的研究（在表中完整列出 A1（附录 A). 根据选择程序和标准，选择了大多数实 证研究文章。根据研究文献，研究人员利用这两种类型的数据库进行推荐的方法和研究。除了 实际调查之外，我们还发现了关于 SQL 和 NoSQL 数据库的调查文章。按照相关的审查选择程 序，我们将选定的研究和出版物分为三个主要类别。如图展示了病历报告的类别饼图。作为实现模型的数学工具，选择了属性元图。元图包含并协调了系统的两个主要特性：统一性（一组相互关联的元素）和可分割性（系统的每个元素也是一个系统）。在这方面，可以将子系统与系统区分开来。这允许在需要时将重点放在系统或其子系统上。



对数据库建模有助于我们预测将存储在数据库中的数据类型以及它们的存储方式。 NoSQL，“不仅是 SQL”的首字母缩写而是一种数据库管理方法，擅长处理大量非结构化数据和大数据分析。各种不同的查询语言都可以用于这些数据库，并且它们不遵循严格的、预先确定的模式结构。然而，在过去的几十年中，关系数据库一直使用行业标准的 SQL 语言。面向文档的数据库是 NoSQL 数据库的一个子集。专注于存储和检索文档的数据库包括 MongoDB 和 CouchDB。这种类型的数据库用于存储和管理主要基于文档的数据。JSON、BSON、XML 和 PDF 等复杂数据格式用于在面向文档的数据库中存储信 息。MongoDB 和 CouchDB 都是免费开源的；然而，MongoDB 更适合分布式设置和 JSON。在研究了几种不同的 NoSQL 数据库特性。MongoDB 是一个专门针对 JSON 构建 的流行数据库，它使用 C++ 编程语言。MongoDB 使用动态模式结构而不是预定义的静态文档。由于查询处理、索引支持和内存聚合方面的改进，数据分析和检 索变得快速而准确。除了提供完全的安全性之外，它还提供恢复和备份实用程序。

数据库以及管理在现代信息技术时代，既有效又高效是关键和重要的 考虑因素。数据库系统的基本特征是永久数据存储保证数据独立于底层物理存储介质，以及通过声明性查询（DBS）实现的查询处理能力。在数据库领域，已经观察到各种领域的各 种方法。包括 ACID、OOM、XML 和数据仓库在内的多个 DMS 都是基于关系数据模型的。然而，基地属性在 NoSQL 数据库中受支持，该数据库主要用于处理大量数据。 新云服务云服务提供商正在以低成本和高效率向客户提供各种功能。然而，多家云服务提供商使用不同的实现和用户界面提供相同的功能，这不可避免地会导致互操 作性问题，不兼容，以及便携性。这些是云服务提供商在采用和推广云技 术时面临的难题 . 在云服务领域内，IaaS、PaaS 和 SaaS 互操作性都是具有不同含义和应用的不同术语 . 出于以下原因，云用户应该从一个 CSP 转移到另一个 CSP: 更 高的停机或故障率、合同终止、公司计划变更、更好的低成本替代方案以及法律难题。使 用多个云提供商的客户无法轻松地在他们之间传输数据。云服务模型努力控制客户的能 力，因为它们的关键架构缺乏互操作性。这个问题通过诉诸供应商锁定来解决，这给基于 云的模型带来了严重的安全问题。数据可移植性将得到改善，因为越来越多的云服务提供商（CSP）采用开放标准来解决互操作性问题。

因此，开发人员很难确保不同云服务之间的数据和应用程序保持一致。因此很多人提出了许 多 倡 议 ， 如 MOSAIC，MODACLOUDS 和 Clous4SOA，旨在解决 PaaS 层的语义互操 作性挑战。为每个 PaaS 提供多个 API 意味着这些计划无法独自解决互操作性问题。 正如预期的那样，CIMI 标准与 IaaS API 兼容，有助于降低云用户及其基础设施服务 提供商所需的互操作性程度。通过使用 OCCI，可以降低 CSP 之间的互操作性，同时仍能产 生足够的结果。由于这些标准没有将互操作性特性纳入其底层架构，因此没有被各种 CSP 广泛使用。此外，标准化接口和 API 的可用性也存在问题。 设计模式、云中间基础设施、服务交付云平台（SDCP）和迁移工具都是由其他学者开 发的以促进不同云之间的数据移动。尽管节省了用户的时间，但这些方法 并没有解决不同 CSP 之间的可移植性问题，这是由于必须学习和实现几个 API 而引起的。亚马逊网络服务（AWS）、

微软 Azure 和谷歌应用引擎（GAE）都是云服务提供商帮助消费 者构建和推出基于云的应用的例子。此外，他们还提供数据库即服务（DBaaS）云平台来支 持数据库开发人员。DBaaS 中的数据移植，可能发生在 SQL 和 NoSQL 数据库 以及每种类型的 NoSQL 数据库内部，由于互操作性问题可能会出现问题。许多类型的 NoSQL 数据库遵循不兼容的存储格式和数据范式。需要标准化的 API 来规范各种云存储平台之间的数据移动。 为了托管他们的数据和应用程序，软件开发人员可以利用 DBaaS，一种高度可 扩展和可用的后端云服务。因此，DBaaS 是目前对云客户最有吸引力的选择。数据库 即服务（DBaaS）是由云提供商（CSP）提供的云服务，有助于从传统数据库架构过 渡 到 云 数 据 库 架 构 。SaaS 便 于 远 程 访 问 计 算 机 程 序 及 其 特 性 和 功 能 。PNUTS 、 HBase 、 SimpleDB 和 Google BigTable 是 其 他 一 些 基 于 云 的 数 据 库 服 务。DBaaS 框架有可能很好地处理传统数据库。但是，由于许多数据库遵循不同的方 法，DBaaS 在数据一致性、机密性、完整性、可用性和缺乏安全性等方面的性能会下降。 如果数据隐私和安全性得到保证，被称为 DBaaS 的外包模式可能会取得财务上的成功。研 究中提出了一个针对私有表数据库云的专家虚拟化顾问。云计算允许将众多应用程序 集成到服务本身的框架中。云计算的可扩展性和适应性的价格比以往任何时候都低。 为了在数据移植过程中提供更多的数据互操作性、可移植性和安全性，论文研究了用于 SQL 和 NoSQL 数据库的两个统一 API 框架 CDPort 和 Se-cloudDB。 在交给第三方之前，他们提供的 API 确保了用户关键信息（CSP）的隐私。作为其计划的 MCTool 的一部分，请求被转换为其相应的模型，然后传送到适当的数据库，考虑它可以处理的模型。元数据加密 / 解密密钥确保只有授权用户和数据库管理员拥有可以访问和更改存储在各种云中的数据。他们提出的框架支持加密和解密。

4. 论述

今天，仅仅依靠结构化数据是不够的，因为天文数据的非结构化性质需要快速的数据分析 和信息提取方法。SQL 数据库功能在有效处理各种形式的大型数据方面受到限制。NoSQL 数据库 的功能允许高效处理海量数据集。NoSQL 数据库在存储容量扩展、模式灵活性、可伸缩性和实 时访问方面表现出色。NoSQL 数据库坚持基本属性。NoSQL 数据库没有优先考虑数据一致性和安 全性，而是将重点放在提高数据读 / 写效率上。应用程序负责确保数据的一致性。因此，它是处理大型数据集的绝佳选择。此外，NoSQL 数据库不像结构化数据库那样在数据、列或表级别应用限制。 这项研究分析了大约 140 项以前的研究，这些研究比较了 SQL 数据库和 NoSQL 数据库的有用性、生产率和可靠性。这项研究考虑了大量的数据。根据我们的调查，NoSQL 数据 库提供了更大的扩展能力 .SQL 数据库更适合事务性系统，并使用更 多资源来确保数据完整性和一致性，而 NoSQL 数据库更适合处理大量不同的数据集，并使 用更少的资源来确保数据完整性和一致性。另一方面，NoSQL 数据库是不同的， 因为它们更重视数据的可访问性。我们的测试结果表明，关系数据库不能被 NoSQL 数据库有效 地取代。因为这两种数据库各有优缺点，所以组织选择使用哪种数据库将取决于该企业特有的 要求。仅举一个例子，NoSQL 数据库可以使用 MapReduce 编程模块，更适合并行计算当它们在集群环境中实现时。 NoSQL 数据库是建立在一个更加灵活和动态的模式上的，与关系数据库相反，关系数 据 库 强 烈 依 赖 于 一 个 被 称 为 “ 模 式 ” （ 表 格 形 式 ） 的 预 设 数 据 结 构。例如，要跟踪学生数据，应该使用 StdRegNo、StdName 和 StdAddress 字段。使用关系数据库时，首先需要构建一个满足所有必要的域和完整性要求的模式。

5. 结论

研究结果表明，没有必要从关系数据库过渡到 NoSQL 数据库。由于这两种类型的数据 库各有优缺点，公司可以选择最符合其要求的 DBMS。如果组织优先考虑数据标准化和一致 性，则可以利用 SQL 数据库。当企业有大量非结构化数据并且对数据可用性有很高要求 时，应该使用 NoSQL。对于小型数据集的聚合，关系数据库可能优于 NoSQL 数据库；对于 大数据分析，关系数据库可能优于关系数据库。由于 MapReduce 的分布式特性，它最适合 用于集群。尽管 MapReduce 在聚合过程中速度较慢，但它在并行计算中更有效，并且被开 发用于处理大量非结构化数据。NoSQL 数据库由于其分散和高度可伸缩的特性，非常适合生成具有不同特性的海量数据的应用程序。当涉及地理空间数据时，关系数据库的可伸缩 性优于 NoSQL 数据库。NoSQL 数据库的数据响应时间比关系数据库更快，尤其是

在处理大量地理空间数据时。

参考文献：

[1] Kumar, L.; Rajawat, S.; Joshi, K. Comparative analysis of nosql (mongodb) with mysql Database. Int. J. Mod. Trends Eng. Res. 2015, 2, 120 – 127.

[2] 张元鸣 . NoSQL 数据库技术 [M]. 北京 : 清华大学出版社 , 2023.02.01.

[3] 柳俊 . 大数据存储——从 SQL 到 NoSQL[M]. 北京 : 清华大学出版社 ,2021:9–11.

[4] Sharma, M.; Sharma, V.D.; Bundele, M.M. Performance Analysis of RDBMS and No SQL Databases: PostgreSQL, MongoDB and Neo4j. In Proceedings of the 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Rajasthan, India, 22 – 25 November 2018; pp. 1 – 5.

[5] McColl, R.C.; Ediger, D.; Poovey, J.; Campbell, D.; Bader, D.A. A performance evaluation of open source graph Databases. In Proceedings of the First Workshop on Parallel Programming for Analytics Applications, Orlando, FL, USA, 16 February 2014; pp. 11 – 18.