电子科技大学

**研究生课程综合考核报告**

**课程名称：数据库新技术**

**教师姓名：胡旺**

**学生姓名：**

**学 号：**

**成 绩：**

**学 期：2018年下学期**

**《数据库新技术》课程综合考核要求**

综合报告要求：请从《数据库新技术及应用》教材所涵盖的面向对象数据库、XML数据库或NoSQL、分布式数据库/区块链技术、数据仓库、数据挖掘、空间数据库、多媒体数据库、移动数据库、时态数据库、主动数据库、工程数据库、实时数据库、内存数据库、数据库安全等内容中任选一个主题，**较全面综述其应用需求产生的背景、国内外研究进展、技术实现思路及体系结构、存在的主要问题及发展趋势。综述报告应该包含中英文摘要、中英文关键词、正文、结论、参考文献等组成部分。综述内容具有较大广度和较高深度，篇幅不少于4000字（6页以上A4版面）。**

提交时间：2018年10月24日（第8周星期三）19:30—20:00；

提交地点：立人楼（第二教学楼）B402；

提交形式：书面报告（**正反双面打印**）;

**特别申明：课程综合考核报告必须由学生本人独立完成，否则，雷同及下载的报告不给予课程成绩。**

**背景**

随着互联网的发展，互联网应用对高并发、低延时的需求已经放在了首位，如果使用数据库纵向扩展的方式，则需要很大的服务器，采购和维护成本很高，并且单台服务器配置总有极限，这就要求数据库能适用于集群环境，能横向扩展。并且，从业务角度看，数据一致性可以不像原先那样放在首位，可以容忍数据在短时间内不一致。针对这些需求，传统的关系型数据库不那么适用了，NoSQL数据库孕育而生。现在应用的整合多是依靠web services等技术使其松耦合，无需直接访问数据库，使一个团队能在应用级别控制并发访问、数据一致性，这都为使用NoSQL创造了条件。

“下一代的数据库产品应该具备这几个特点：非关系型的、分布式的、开源的、可以线性扩展的。”这类数据库最初的目的在于提供现代网站可扩展的数据库解决方案，它开始于2009年年初，目前正在快速发展。这种类型的数据库具有以下特点：自由的Schema，数据多处备份，简单的编程API，数据的最终一致性等。所以，将这种类型的数据库称为NoSQL（不仅仅是SQL，全称为“not only sql”）数据库。

由于我们需要处理的数据集越来越大，其存储量已经远远超过了单机的容量，数据处理的需求也远远超过了单机CPU的运算能力。所以我们需要分布式的解决方案。

我们对数据提供速度的要求越来越高，在80年代，可能很多运算都需要跑一整晚。但是这种事情放在现在就变得不可接受了。对于复杂的统计分析我们可以忍受，但是对于网站应用来说，快速响应是必须的。

我们需要提供7×24的服务。如果你的网站只有一个静态页，那估计问题不大，只需要做好WebServer的容错性就行了。而如果你是一个背后有数据库，有缓存的动态网站，那你就必须做好数据层的容错及自动故障迁移。

很多应用场景需要数据层提供更高的写性能和数据吞吐。比如日志型应用，对写性能的要求可能非常高，当写性能成为瓶颈时，通常我们很难难过升级单机配置来解决。所以分布式的需求在这里变得也很重要。

我们对非结构化数据的存储和处理需求日增，在这个变化的世界，互联网领域的应用可能越来越难像软件开发一样，去预先写义各种数据结构。

我们的应用场景对一致性，隔离性以及其它一些事务特性的需求可能越来越低，相反的，我们对性能，对扩展性的需求可能越来越高。于是在新的需求下，我们必须做出抉择，放弃一些我们习惯了的优秀功能，去获取一些我们需要的新的特性。

**国内外研究进展**

**发展现状**

计算机体系结构在数据存储方面要求具备庞大的水平扩展性，而NoSQL致力于改变这一现状。Google的 BigTable 和Amazon 的Dynamo使用的就是NoSQL型数据库。

NoSQL项目的名字上看不出什么相同之处，但是，它们通常在某些方面相同：它们可以处理超大量的数据。

这场革命仍然需要等待。的确，NoSQL对大型企业来说还不是主流，但是，一两年之后很可能就会变个样子。在NoSQL运动的最新一次聚会中，来自世界各地的150人挤满了CBS Interactive的一间会议室。分享他们如何推翻缓慢而昂贵的关系数据库的暴政的经验，怎样使用更有效和更便宜的方法来管理数据。

“关系型数据库给你强加了太多东西。它们要你强行修改对象数据，以满足RDBMS （relational database management system，关系型数据库管理系统）的需要，”在NoSQL拥护者们看来，基于NoSQL的替代方案“只是给你所需要的”。

水平扩展性(horizontal scalability)指能够连接多个软硬件的特性,这样可以将多个服务器从逻辑上看成一个实体。

**挑战**

尽管大多数NoSQL数据存储系统都已被部署于实际应用中，但归纳其研究现状，还有许多挑战性问题。  
1、已有key-value数据库产品大多是面向特定应用自治构建的，缺乏通用性；  
2、已有产品支持的功能有限（不支持事务特性），导致其应用具有一定的局限性；  
3、已有一些研究成果和改进的NoSQL数据存储系统，但它们都是针对不同应用需求而提出的相应解决方案，如支持组内事务特性、弹性事务等，很少从全局考虑系统的通用性，也没有形成系列化的研究成果；  
4、缺乏类似关系数据库所具有的强有力的理论(如armstrong公理系统)、技术（如成熟的基于启发式的优化策略、两段封锁协议等）、标准规范（如SQL语言）的支持。  
5、目前，HBase数据库是安全特性最完善的NoSQL数据库产品之一，而其他的NoSQL数据库多数没有提供内建的安全机制，但随着NoSQL的发展，越来越多的人开始意识到安全的重要，部分NoSQL产品逐渐开始提供一些安全方面的支持。

　　随着云计算、互联网等技术的发展，大数据广泛存在，同时也呈现出了许多云环境下的新型应用，如社交网络网、移动服务、协作编辑等。这些新型应用对海量数据管理或称云数据管理系统也提出了新的需求，如事务的支持、系统的弹性等。同时云计算时代海量数据管理系统的设计目标为可扩展性、弹性、容错性、自管理性和“强一致性”。目前，已有系统通过支持可随意增减节点来满足可扩展性；通过副本策略保证系统的容错性；基于监测的状态消息协调实现系统的自管理性。“弹性”的目标是满足Pay-per-use 模型，以提高系统资源的利用率。该特性是已有典型NoSQL数据库系统所不完善的，但却是云系统应具有的典型特点；“强一致性”主要是新应用的需求。

**技术实现思路及体系结构**

CAP原理

分布式数据系统的CAP原理三要素：

一致性

可用性

分区容忍性

NoSQL运动的两个核心理论基础

Google的BigTable

将各列数据进行排序存储。数据值按范围分布在多台机器，数据更新操作有严格的一致性保证。

Amazon的Dynamo

将数据按key进行hash存储，其数据分片模型有比较强的容灭性，因此它实现的是相对松散的弱一致性：最终一致性。

基于key值存储的NoSQL数据模型