对云环境中改进Mongodb自动分片技术的研究

摘要

随着网际网路2.0技术的快速发展，大规模分布式服务的需求，给存储在云计算带来的巨大的挑战，动摇了传统的的关系型数据库。NoSQL数据库从RDBMS的桎梏中解放出来成为人们关注的焦点。在这篇文章中，首先本着自动的原则和实施机制，提出了MongoDB数据库分片技术，而后为了解决分布不均的问题提出一个在数据操作的频率的基础上的改进的数据自动分片算法. 改进的均衡策略能有效平衡之间的数据碎片，提高集群的并发读写性能。

关键词 NoSQL，MongoDB，自动分片，平衡策略

绪论

近年来，随着互联网数据量的快速增长和互联网的Web 2.0技术的发展，如何有效地存储，处理和提取的大量数据，成为一个迫切需要解决的问题，云计算在这种情境下应运而生。云计算通过网络（通常是因特网）[1]以计算而不是一个产品的形式作为服务的交付，从而共享资源，软件和信息提供给计算机和其他作为计量的服务设备。许多大学，供应商和政府组织各地投资于发展云计算的项目，例如：亚马逊推出简单存储服务（S3）和弹性计算云（EC2），谷歌推动的BigTable和MapReduce，GFS，已经全部成功地应用在生产环境中。分布式文件

系统可以组织大量的数据在云端，并且也能够有效地读取云数据，但我们为更好的结构化数据的管理，仍然需要专门的数据管理工具。