

doi:10.1631/FITEE.1500441

题目：大数据存储技术综述

概要：对于容量快速增长、日趋多元化的大数据，业界亟需开发可行性更好的存储工具。为满足大数据存储需求，存储机制已经形成从传统数据管理系统到 NoSQL 技术的结构化转移。然而，目前可用的大数据存储技术无法为持续增长的异构数据提供一致、可扩展和可用的解决方案。在科学实验、医疗保健、社交网络和电子商务等实际应用中，存储是大数据分析的第一步。截至目前，亚马逊、谷歌和阿帕奇等公司形成了大数据存储方案的行业标准，但尚未有关于大数据存储技术性能和容量提升的深入调查和文献报告。本文旨在对目前可用于大数据的最先进的存储技术进行全面调查，提供了一个明确的大数据存储技术分类方法，以帮助数据分析师和研究人员了解和选择更适合其需求的存储机制。我们使用布鲁尔的 CAP 定理比较和分析了现有存储方法，评估了不同存储架构的性能，讨论了存储技术的意义、应用及其对其他类别数据的支持。为了加快部署可靠和可扩展的存储系统，文中还突出了未来研究面临的几个挑战。

关键词：大数据；大数据存储；NoSQL 数据库；分布式数据库；CAP 定理；可扩展性；一致性-分区弹性；可用性-分区弹性