国家重点基础研究发展规划(973)项目 下一代互联网信息存储的组织模式和核心技术研究

项目编号: 2004CB318200





[2005] 02 号 总第 3 期

项目办公室 编

2005年06月28日

"下一代互联网信息存储的组织模式和核心技术研究"学术交流会 会议纪要

2005年6月16-17日,国家973计划项目"下一代互联网信息存储的组织模式和核心技术研究"(项目编号:2004CB318200)在北京湖北大厦会议厅召开了学术交流会。

与会人员有:本项目咨询专家北京邮电大学陈俊亮院士、天津大学张 以谟教授,项目专家组西安微电子研究所沈绪榜院士、武汉大学何炎祥教 授,科技部基础司 973 办公室钱小勇博士,教育部科技司基础处李渝红处 长,以及项目组下属六个课题组负责人和主要学术骨干,项目依托单位华 中科技大学科研主管人员等,共计 38 人。

会议由"项目实施组工作会议"和"项目工作汇报和学术交流"两天会议组成。

2005年6月16日召开了"项目实施组工作会议",与会人员包括各课题组负责人和技术骨干。会议围绕项目总体目标之一——构建PB级网络存储系统及其上的GIS应用示范,重点讨论课题之间的接口,以及如何组织好项目下一步的研究工作。经过研讨,达成以下共识:

- 1、课题围绕项目总体目标开展研究,各课题的软硬件成果通过应用示范体现。
- 2、项目实施过程中,课题间应多进行技术交流,进一步细化相互间接口。
- 3、争取在项目中期检查前,初步建立一个反映各课题研究进展、可演示的存储平台,并在其上有小规模的 GIS 应用示范。
- 2005年6月17日召开"项目工作汇报和学术交流",会议由项目首席科学家冯丹教授主持。首先冯丹教授围绕本项目拟解决的两个关键科学问题: 网络存储的组织模式和 PB 级存储系统的组建理论与方法,从项目研究内容、考核指标、取得的成绩和存在的问题、以及建议措施等方面介绍了项目的工作现状。之后由各个课题组汇报工作进展及下一步工作计划。在此基础上,项目咨询组专家、项目专家组专家对项目的工作提出了有益的建议。
- 01 课题:通过对历史用户 I/O TRACE 分析、建模和预测,研究复杂网络环境下的数据存储模型。从理论上研究了互连网络环境下的动态、自适应的网络存储体系结构,以用于构建海量存储系统。研究了对象存储控制器的构成方法,包括对象属性、对象操作、对象封装等。提出了基于对象存储的一种自适应策略触发机制,以支持不同应用。提出了对象存储系统中的一种存储资源规划和负载均衡方法,以提高系统整体性能。为验证以上阶段性理论成果,从对象存储系统、对象存储控制器、缩短 I/O 路径的对象控制器硬件方法等多个层面进行深入研究,包括其软硬件构成和实现方法等。
- 02 课题:结合对象网络存储的特征,提出消除存储节点带宽瓶颈、扩展局域网通道带宽的新途径。结合广域网的特征,提出降低广域网和局域网带宽落差的新方法。由于 Infiniband (IB) 互连网络灵活可扩展,最高单通道带宽可达 30Gbps,拟采用 IB 通过协议转换构建高速通道,通过可减少拷贝次数的 Zero-Copy、分布式存储管理、负载平衡等提高通道性能。

- 03 课题: 研究海量存储系统存取过程中的动力学模型并取得阶段性进展。数据在系统中的分布很大程度上决定了各种 I/0 流的路径和强度; 当前以及预期负载特性和数据分布又决定了新数据在系统中的分布。系统中数据和 I/0 流的分布受限于分配策略,而分配策略又反过来受限于数据、I/0 流在时空域中的布局。通过研究这种复杂系统中动力学特性,以建立最佳系统。建立进化存储模型并实现了原型仿真系统。完成统一存储网(USN)工作,即实现 SAN 和 NAS 的统一存取,实现三方通讯机制。构造了 5TB,十几个存储结点的网络存储原型系统,通过研究分布式文件系统和相关测试工具,寻求大系统的构建方法。
- 04 课题: 用 8 台服务器搭建了一个 P2P 存储系统,对系统的可靠性和可用性进行了研究。改进和优化了基于 Cauchy 矩阵的 Erasure code 实现算法,在现有系统中用 java 实现基于 Erasure Code 的容错机制,定量地分析了 Erasure code 的可靠性和时间开销等问题,提出一种发现低错误相关性的算法。对网络特征及用户行为进行分析,在对等网络行为分析中首次采用聚类的方法,对用户进行了科学的分类。在数据的组织与索引机制研究方面,已经建立了论文资源、用户、用户评论的模型,收集了约 60 万条论文资源描述信息,并结合 ACM 学科分类体系,形成了中心的知识库,提出了 P2P 系统快速定位算法。
- 05 课题: 在存储按需部署研究方面,实现了一个基于存储与计算分离的思想以及自适应网络存储动态部署模型和操作系统存储对象模型的系统。该系统将系统数据进行集中管理,系统数据与计算资源通过网络相联,实现了数据为中心的计算环境可重构。在存储服务质量保证方面,建立了一个磁头位置预测模型和磁盘请求服务时间预测模型。在数据分布方面,设计实现了虚拟化系统的元数据结构,使之适用于随机分布的空间布局;设计实现了目标器端 SCSI 中间层,加入后台数据迁移线程,并通过不同优先级的命令调度来减少性能影响;实现了基于磁盘容量和(容量*带宽)分配方式的虚拟盘数据迁移,并进行了部分性能测试。在存储安全保障方面,

构建了一个系统软硬件平台,设计了基于权限标识的分布式认证机制。实现了一套可运行的网络存储服务管理原型系统。

06 课题:设计了GIS 系统总体方案,并开始研究 GIS 数据分布式管理的中间件。进行二维空间数据表现的原型实验,包括空间数据库的调度、层的封装、符号表现等;研究基于对象存储的新一代 GIS 空间数据存储模式。设计遥感数据分发与服务系统框架,研究基于海量存储系统的遥感数据在线分发与访问模型。收集整理 2002-2004 年 MODIS 卫星数据,达 5TB 遥感数据信息。

各课题组汇报完后,与会人员对相关问题进行了深入的讨论。项目组 咨询专家陈俊亮院士和张以谟教授,项目专家组专家沈绪榜院士和何炎祥 教授等,对项目研究工作提出了指导性意见,主要包括:

- 1、肯定了项目进展,认为项目进展状态良好。
- 2、探讨了目前存在的一些问题,主要包括:在理论研究方面应进一步加强,通过理论指导系统的实现;进一步积极参与相关的国际标准制定。
- 3、下一阶段研究中要考虑做好中期评估准备工作,以利于项目的顺利滚动和进行。

此纪要呈报:

科技部基础司 973 计划办公室 教育部科技司基础处 湖北省科技厅基础处 各课题组所在单位科技处

项目组咨询专家: 张以谟教授、陈俊亮院士、金怡濂院士项目专家组成员。

973 项目 (2004CB318200) 办公室 2005 年 06 月 28 日