

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与核心技术研究

项目编号：2004CB318200

简 报

[2005] 03 号 总第 4 期

项目办公室 编

2005 年 11 月 30 日

“下一代互联网信息存储的组织模式与核心技术研究”项目年度总结 与学术研讨会会议纪要

2005 年 11 月 21-22 日国家 973 计划项目“下一代互联网信息存储的组织模式与核心技术研究”（编号为 2004CB318200）在北京湖北大厦举行 2005 年度项目年度总结与学术研讨会。

与会人员有 973 项目组咨询专家张以谟教授、金怡濂院士和陈俊亮院士，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，湖北省科技厅王红斌处长，项目专家组专家沈绪榜院士、叶朝晖院士、何炎祥教授、韩承德教授、谢长生教授、代亚非教授，以及项目组下属六个课题组主要学术骨干和项目依托单位华中科技大学科研主管人员等共计 40 人。

会议由“项目年度工作汇报”和“学术研讨”两部分构成。

项目年度工作汇报由项目首席科学家、华中科技大学冯丹教授主持。冯丹教授从项目工作汇报、项目进展情况和下一年研究计划三个部分分别进行介绍，着重从项目研究内容、考核指标、理论研究、系统研制等方面介绍了项目 2005 年度进展状况。之后与会的科技部和教育部分代表以及专家们对项目研究工作发表了意见，主要内容包括：

认为这是一个务实的报告，肯定项目组前阶段的工作。

指出项目年度总结的必要性。年度总结不仅检查项目进展，而且籍此进行学术交流和相互沟通，并能得到专家们指导，少走弯路。项目实施已进行一年二个月，进行年度总结既可以从每个课题，也可以从项目全局来分析进展情况以及自身所处的水平，对项目后续完成很有意义，对项目中期评估很有帮助，可根据目前研究情况作适当的调整，促进项目朝着预期目标前进。

强调项目主要目标应该为国家基础研究做出有分量的贡献，能够在国际上占一席之地，做出一些实实在在的东西。

强调了 973 计划项目的重大意义，希望通过 973 计划项目能解决国家的重大问题或者适用于重大应用，不断提升我国科技实力。

学术研讨会由六个部分组成，分别由六个课题组组长负责，主要内容各课题组对课题开展以来的工作总结和学术报告。

1、华中科技大学等单位承担的 01 课题的报告。分别由华中科技大学王芳副教授和解放军通信指挥学院贾连兴教授对课题 1 研究进展情况进行了汇报。该课题从下一代互联网存储系统的复杂网络应用环境入手，对已有历史用户 I/O 行为的 TRACE 记录进行收集、分析，同时构建分布式 I/O Trace 系统，跟踪实际网络应用系统的数据访问并进行分析，为对象存储系统研究提供依据。在网络存储动态规划与系统结构方面构建了复杂系统离散动力学模型，其结果可适用于离散动力学系统的行为仿真和预测，为复杂性网络动态性能分析提供了一条新的研究途径。在可扩展存储对象组成原理、存储对象控制器构成方法等方面，开展了关于对象存储设备与系统的智能性和自治性的研究，初步对其智能型的部分表征特征如高效组织分布、自我管理、主动服务机制等进行研究分析，具体包括：用户数据在对象存储系统的柔性分布策略和最佳并行度研究，对象在存储设备内部的基于连续扩展块的组织形式研究；综合对象复制和迁移的动态负载均衡机制研究，对象存储系统自我学习的策略触发机制研究；对象存储基于预取扩展属性的自适应预取策略研究；保证用户的对象合法访问以及客户和 OSD 之间命令和数据的完整性的访问控制机制研究。在存储系统性能优化理论

与方法发明，研究了网络存储 I/O 流水调度机制和 I/O 任务的聚散方法等。上述理论研究已经在对象控制器和对象存储系统的实践中加以应用和验证。完成了包括元数据服务器、高速网络、基于对象存储的网络磁盘阵列和客户端的对象存储系统的构建，实现对象扩展属性与操作的接口，基于元数据信息的对象空间与底层物理存储空间的映射，为可扩展对象控制器和缩短 I/O 路径方法的研究工作提供进一步的系统依据和运行平台，也为理论研究提供实验和验证基础平台。实现了基于文件对象和基于存储区域的存储对象控制器的软件控制系统，完成了硬件设计和各项制版准备工作。在缩短 I/O 路径方法的探索中，采用 FPGA 技术形成网络端口到存储设备端口的直接处理通道设计，完成了硬件与固件设计中的部分模块。理论和实践结合已形成对象存储原型系统，并和 GIS 应用示范进行了初步联调。

2、中国科学院计算所承担的 02 课题的报告。由韩冀中副研究员报告了课题 2 研究进展情况。该课题结合下一代网络存储系统的数据交互特征，提出一种新的理论框架，描述网络存储系统中数据通道的组织模式，分别为总线链路层、协议传输层、对象会话层、存储表示层以及服务引擎层。总线链路层由 I/O 总线与网络硬件组成；协议传输层实现数据传输协议；对象会话层赋予数据新的网络属性；存储表示层为应用提供数据组织，服务引擎层支持面向网络的应用服务。重点研究对象会话层和服务引擎层的理论模型和组织架构，以及它们与其它层的交互机制。为了支持基于对象存储这种新的网络存储理论，本课题在协议传输层和存储表示层之间建立了一个基于对象的会话层。通过赋予对象一组与网络特征相关的属性，在会话交互过程中，可以利用网络自动识别并实现对象的传输策略，从而使对象的交互方式变被动为主动，减轻了上层应用的负担。对象会话层支持存储表示层中的文件、对象、数据库等多种组织模式，通过中间层适配协议，可实现与传统应用的无缝连接。为了支持下一代互联网的应用，本课题提出建立服务引擎层，利用可互操作的访问协议，支持在分布、异构系统上部署应用服务。并利用引擎虚拟化管理，实现高速的数据交互机制，能够更好地适应于下一代网络上面向服务应用的推广和部署。

3、华中科技大学承担的 03 课题的报告。由曹强博士报告了课题 3 研究进展情况。其主要工作是设计并实现三个原型系统，构造三个理论研究平台，并在网络存储领域展开广泛的理论研究。三个平台是支持零拷贝、多协议、多通道的存储结点控制器，100TB 的网络存储平台（物理硬件平台）和海量存储软件系统（软件平台）。其中支持零拷贝、多协议、多通道的存储结点控制器已经完成，其主要工作是对大量块设备进行管理，并提出优化方法。而后两个平台主要用于研究海量存储系统的软硬件架构及其核心算法，并为今后的理论验证工作提供良好的平台。本年度理论研究包括以下多个方面：初步分析海量存储系统运行过程中数据存取过程的动力学特性；分析并研究基于存储对象的分布式文件系统；吸收现有几种典型分布式文件系统的优点，建立具有多层次、多操作接口存储管理系统；对数据存储系统和数据分布进化进行研究；磁盘卷和磁盘级的优化调度研究；完成了统一存储网（USN）工作；基于三方架构的安全机制研究；分散式（非集中式）存储系统研究；支持零拷贝、多协议、多通道的存储对象控制器；相关性能测试和仿真方法研究。

4、北京大学承担的 04 课题的报告。由代亚非教授报告了课题 4 研究进展情况。课题从两个研究方向进行工作：(1)从大规模存储系统可靠性分析模型研究而言，本项目目标是建立下一代互联网上的大规模存储系统，其可靠性问题是存储系统中至关重要的基础性研究问题。由于历史及计算复杂性等原因，目前对存储系统的可靠性评估一直沿用 MTTF（平均故障时间）这个标量指标和失效的指数分布假设。然而失效的指数分布假设在大量存储节点参与的系统中是否正确并没有得到回答，因此 MTTF 这一衡量方法将不一定适用于本项目设计的系统。我们提出了对大规模节点参与的存储系统可靠性建模的方法，通过计算给出了实际失效的分布与指数分布的差距，并给出何种条件下可以使用 MTTF 衡量系统可靠性。(2)在开放式对等网络用户行为激励模型研究方面，该课题研究存储的 P2P 组织模式，为了促进用户参与到开放性的对等网络中来，需要提供一种激励机制来鼓励用户参与以及限制不合作的用戶行为。激励机制可以这样分为三个主要的类

别：基于私有的历史记录，客观的基于共享的历史记录，以及主观的基于共享的历史记录。基于私有历史的方法指用户提供更好服务给曾很好服务自己的用户，而后两种则基于全局共享的积分记录。根据我们在实际系统中的测量，发现基于私有信息和公有信息的两种激励都存在问题。提出了一种 Multi-Trust 激励机制的理论模型，以提高基于私有历史纪录的用户信任矩阵的覆盖面，进而达到激励用户贡献和防止系统中合作作弊行为的作用。

5、清华大学和中国科学院计算所共同承担的 05 课题的报告。分别由许鲁研究员和舒继武副教授汇报“下一代互联网存储按需部署模型的研究”和“下一代互联网存储服务质量研究”。“下一代互联网存储按需部署模型的研究”通过对象化技术，提出了一个可以有效支持带内和带外混合管理模式的网络存储设备对象模型。该模型不仅可以有效地支持不同的管理模式，而且可以根据需要动态地进行管理模式的调整。这个模型不仅可以适用于局域环境，而且可以应用于广域环境。通过对象化技术，该模型可以支持数据冗余和部署。并在此基础上，对于操作系统类型的典型存储对象，研究了其部署模型。“下一代互联网存储服务质量研究”从网络存储服务质量、面向海量信息的数据分布以及存储安全三个方面进行。在网络存储服务质量领域，对具有服务质量保证的自主存储管理系统以及对存储系统服务质量控制的性能评价标准进行了研究，完成了存储服务质量纵向体系结构设计及模块设计。面向海量信息的数据分布领域，研究了自动适应系统环境的数据分布策略，在变化中的存储资源配置环境中实现各设备上的均衡负载，并且保证算法的时间和空间复杂度尽量最小，同时数据转移的总量尽可能最少；研究了数据副本的分布与保存策略，利用空闲的存储空间对每块数据保存一定数量的副本，从而有效提高了数据的可靠性，同时利用不同位置的副本来提高存储数据的访问性能。存储安全领域，研究如何建立灵活的对象安全属性以及严谨高效的安全认证方法。

6、武汉大学承担的 06 课题的报告。由喻占武教授报告了课题研究进展情况。06 课题针对目前数字地球、数字海洋、数字城市等大型 GIS 应用

系统存在的 PB 级海量数据组织管理和存储存在的访问性能、扩展性、安全性等诸多问题，基于多层次可扩展对象存储理论，设计了大型多媒体 GIS 系统总体方案；研究多比例尺、多分辨率、多数据源空间数据管理关键技术，进行二维、三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、三维显示建模与显示等；研究基于对象存储技术，在国际上提出了“基于对象存储的新一代 GIS 空间数据存储模型”，并进一步研究空间语义存储对象、多层次空间元数据服务和 GIS 数据分布式管理的中间件，研究了海量卫星遥感数据存储关键技术。在应用示范系统研究开发方面，设计实现了二维 GIS 示范子系统、三维 GIS 示范子系统、多媒体应用示范子系统和海量卫星遥感数据管理子系统原型，收集、处理、分析近三年的中国地区 MODIS 卫星数据，数据量达 10TB。并初步与课题一和课题三模型（系统）进行了联调。

与会人员与报告人员对项目研究相关的问题进行了深入的讨论。项目咨询专家金怡濂院士、陈俊亮院士和张以谟教授，项目专家组专家沈绪榜院士、何炎祥教授等在听取了报告之后，对开展项目研究工作发表了指导性意见并提出中肯的建议，主要内容包括：

1、肯定了项目进展，认为这一年来研究有了很大进展，主要体现在两个方面：一是在理论研究上，每一个课题都进行较深入的理论研究。二是在系统研发上，各个课题都针对自己的目标做了实际系统开发工作。

2、探讨了目前存在的一些问题，主要包括：

项目的闪亮点是什么，指出本项目的最大闪亮点应该是提出的一套先进、完整的理论体系架构并能够为大家所接受、认同，能在实际中发挥效益，同时能在该领域国际一流期刊和 Top 会议上发表论文。

研究目标的聚焦，认为本项目提出的“下一代互联网信息存储的组织模式与核心技术研究”是针对下一代互联网，进行存储组织模式和核心技术的研究。因此，整个研究应该聚焦在这个目标上。

研究平台的统一，各个课题组虽然可以在自己搭建的平台上实验，但整个项目应该是一个平台，每一课题研究成果都应在这个平台上得到体现，

不能独立开发，各做各的，如示范应用与理论研究要紧密结合，GIS应用建立在前五项研究成果上，充分利用这些研究成果。

3、对未来工作提出建议，主要包括：

理顺各课题关系。

加强理论研究。

加强应用，专家认为项目设想非常好，能否满足国家重大应用还需要实际应用来验证。

做好中期评估准备工作，973项目均采用2+3的方式，立项2年后进行中期评估。在中期评估时主要不是看成果，而是看工作状态和研究前景。做好评估准备工作有利于项目的顺利滚动。

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

项目咨询专家：张以谟教授、金怡濂院士、陈俊亮院士

项目专家组专家：沈绪榜院士、叶朝晖院士、何炎祥教授

973项目(2004CB318200)办公室

2005年11月30日