

高能所数据存储系统现状及规划

Saturday, 29 June 2024 16:00 (15 minutes)

高能所数据存储系统是一个庞大的分布式存储系统，旨在满足高能所各个领域的数据存储需求。这些领域包括粒子物理和天体物理、空间项目、多学科交叉（CSNS/HEPS）、理论物理和加速器设计等领域。为了支持这些需求，我们运行了多个分布式存储系统，其中包括主要的物理数据存储系统 Lustre 和 EOS，以及通用文件系统、专用 HPC 集群、对象存储系统、内容分发文件系统和用户的云盘存储。目前，这些系统的总容量已超过 100PB。

在过去的十多年里，Lustre 和 EOS 作为高能所主要的物理数据存储系统，提供了 100GB/s 的数据访问带宽。为了适应变化的需求，我们引入了最新的磁带库管理软件 CTA，已全面替代现有的 Castor 系统。我们还采用 Restic 作为数据备份系统，以实现对关键数据的高效备份。为了满足用户对 HOME 目录的需求，我们正在探索新的商业文件系统解决方案。为了满足公共云服务平台的存储需求，我们提供了 CEPH 存储集群。同时，我们通过 CVMFS 为高能所的所有实验提供软件存储库。

目前，高能所存储系统采用了“一平台多中心”架构，以北京和东莞两大数据中心为主，辅以稻城、江门等多个分数据中心，形成了分布式存储架构。本文将重点介绍去年高能所数据存储的经验和服务情况，并讨论即将发生的变化、未来的改进和面临的挑战。

Summary

Primary author: LI, Haibo (高能所)

Presenter: LI, Haibo (高能所)

Session Classification: 科学数据存储与管理