Design and construction of High Energy Photon Source (HEPS) scientific data storage system

Yaosong Cheng, Qingbao Hu, Qiuling Yao

chengys@ihep.ac.cn

Institute of High Energy Physics, Chinese Academy of Science

0. Introduction

China's High Energy Photon Source is about to become one of the world's brightest fourth-generation synchrotron radiation facilities, is being under intensive construction in Beijing's Huairou District, and will be completed at the end of 2025. The 14 beamlines for the phase I of HEPS will produce more than 300PB/year of raw data. Efficiently storing, managing, and accessing this massive amount of data is a significant challenge faced by HEPS.

In accordance with HEPS data policy, the storage of massive data requires the implementation of long-term data preservation and ensuring efficient data access. In order to balance the cost-effectiveness of storage devices and realize the high reliability of data storage, a three-tier storage is designed for storing data, including beamline storage, central storage, and tape storage. Accordingly, there is a storage policy for data preservation (see Fig. 1), the raw data and processed data are stored on the beamline storage for a maximum of 7 days, on the central storage for a maximum of 90 days, and only the raw data are archived to tape for long-term storage with two copies. Of course, this data storage policy could be adjusted according to the actual data volume and funding situation of HEPS.

1. Beamline storage

- Architecture
  - Utilizes NVMe SSD arrays
  - Fully symmetric distributed storage system
  - Achieves high data input/output speeds
  - Total Storage Capacity: 1.8PB
- Performance Enhancement
  - Employs high-performance private client DPC by Oceanstor (Fig. 2)
  - Single-process read/write speed: 5GB/s
  - Aggregated read/write bandwidth capacity: 60GB/s

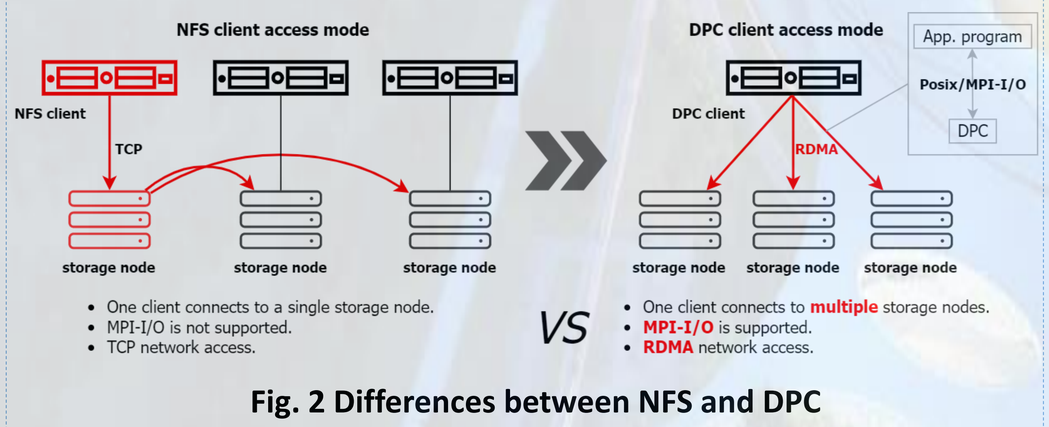

Fig. 2 Differences between NFS and DPC

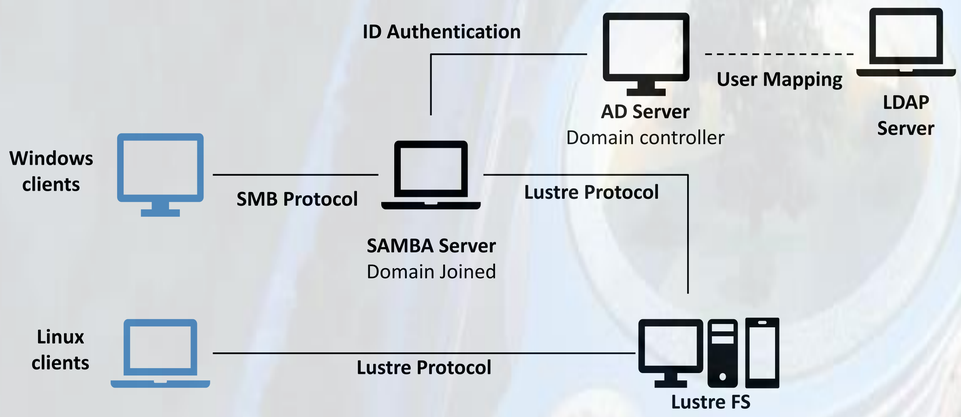

Fig. 4 Cross-platform user and permission synchronization solutions

3. Tape storage

- Architecture
  - Compliant with the LTO9 standard
  - Total Storage Capacity: 50PB
  - Utilizes EOSCTA for tape management (Fig. 5)
- Advantages of EOSCTA
  - Universal open-source tape management software
  - Provides a user-friendly interface and efficient management methods
  - Widely used in high-energy physics
- Developed the data transmission system

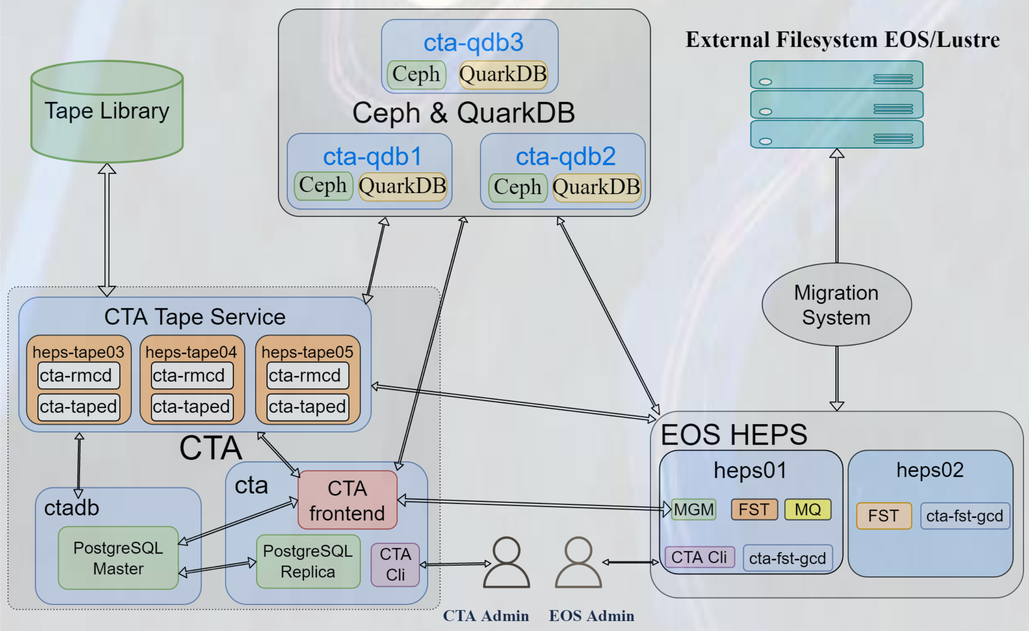

Fig. 5 EOSCTA architecture

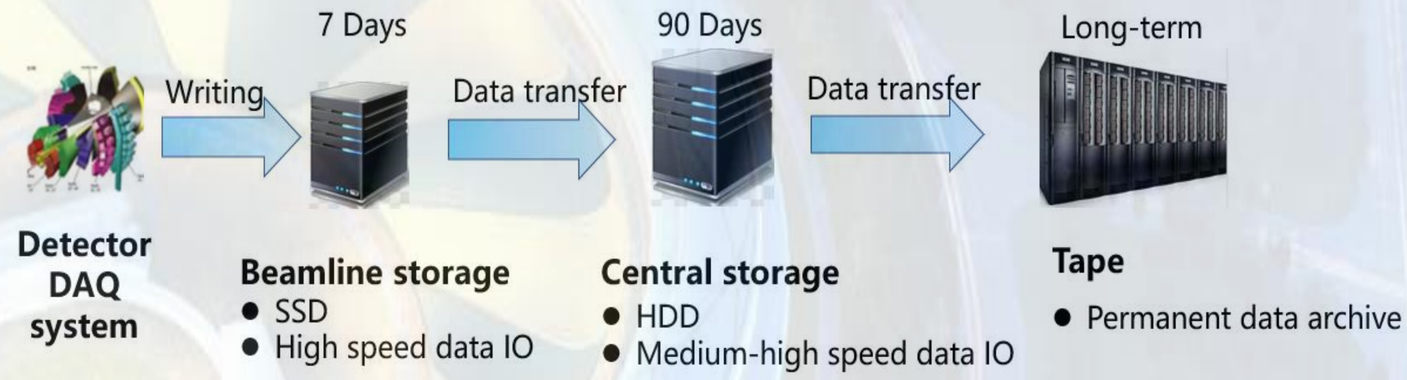

Fig. 1 The HEPS storage policy

2. Central storage

- Architecture
  - Leverages distributed high-density HDD arrays
  - Achieves medium-high speed data I/O
  - Utilizes the open source and mature Lustre storage system (Fig. 3)
  - Multi-MDT Architecture
  - Ensures system stability during large-scale data access
  - Enhances read/write performance
  - Total capacity: 28PB
- Advantages of Lustre
  - Widely used in the high-performance computing field
  - Lustre supports parallel I/O, efficiently handling numerous read and write requests
  - Supports configurations ranging from small systems to thousands of nodes
  - Supports a variety of storage hardware and network protocols
  - We have extensive operational experience and reusable tools
    - User behavior anomaly detection system
    - Cross-platform user and permission synchronization solutions (Fig. 4)
    - Support for NVIDIA-GDS accelerated GPU access to storage
- Operation and maintenance experience
  - Supports 12 scientific applications
  - Comprises 22 file system instances
  - Total storage space exceeds 40PB

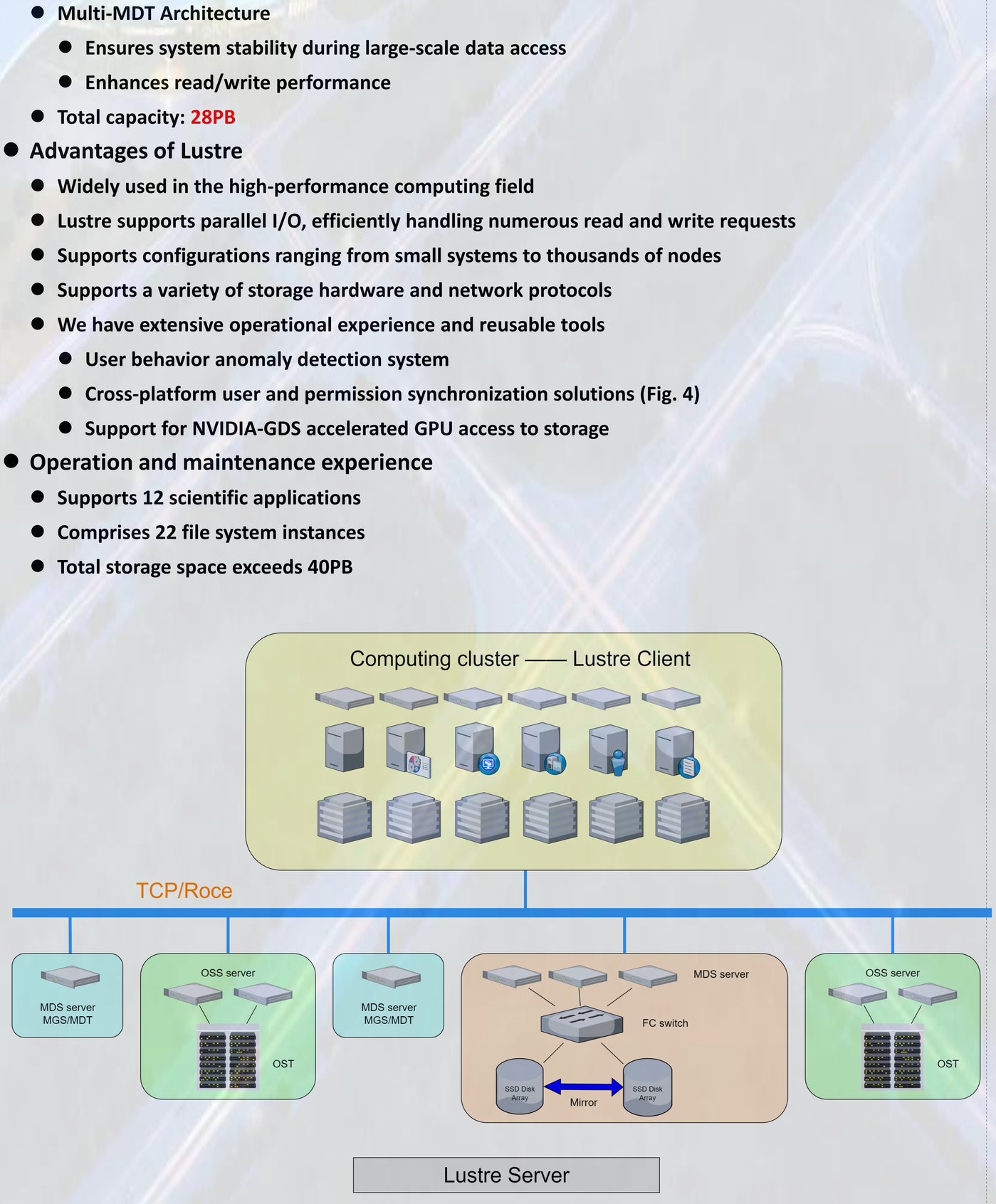

Fig. 3 Lustre architecture in the HEPS central storage