Contribution ID: **37** Type: **Oral Presentation**

Long-term data archive of IHEP: From CASTOR to EOSCTA

Friday, 25 March 2022 12:00 (30 minutes)

CASTOR is the primary tape storage system of CERN and has been used for over fifteen years at IHEP. By 2021, the data volume from experiments has reached 12PB. Two replicas are saved in tape for most raw data, as a result, the capacity of CASTOR has exceeded 20PB. However, numerous factors hinder the performance of CASTOR. For example, new experiments such as JUNO and HEPS, ask for long-term storage and quick increasement of data. To satisfy these requirements, we plan to replace tape storage system from CASTOR to EOSCTA. New data of LHAASO has been saved gradually in EOSCTA since late 2021. Moreover, BESIII online data and JUNO raw data will be saved directly in EOSCTA from 2022.

In this paper, we describe the current infrastructure of EOSCTA at IHEP, whose components comprise a CTA tape system as back-end, an EOS filesystem as disk buffer, a Ceph as queue manager and a PostgreSQL database. We set up two EOSCTA instances which serve for four experiments, and the front end is consisted with multiple disk file systems (LUSTRE and EOS). According to the difference of data generation, different workflows are designed to receive data from remote experimental stations or local disk arrays to EOSCTA. In order to ensure the effectiveness and reliability of EOSCTA, we have adopted several tools.

CASTOR will be replaced by EOSCTA and all the existed data of CASTOR will be migrated to EOSCTA. To achieve this target, we also have to upgrade tape drives from LTO4 to LTO7, which includes five CASTOR instance and two types of tape library. It is planned to complete most of the migration by 2023. The paper reports the migration plan, the steps and methods of data migration, and the inspection to ensure the data integrity. To elucidate the process of data migration, this paper takes the data of BESIII as an example.

Based on the previous experience of EOSCTA, we present the outlook in the requirements of experiments at IHEP, discuss a possible way to use EOSCTA as massive tape data storage for multiple data sources, and also design a unified workflow which is more suitable for local data.

Primary author: YAO, qiuling (IHEP)

Presenter: YAO, qiuling (IHEP)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data