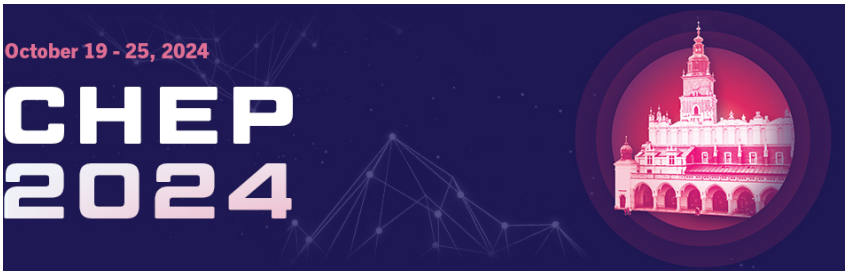

Contribution ID: **110** Type: **Talk**

ML-based Adaptive Prefetching and Data Placement for US HEP systems

Wednesday 23 October 2024 16:51 (18 minutes)

Although caching-based efforts [1] have been in place in the LHC infrastructure in the US, we show that integrating intelligent prefetching and targeted dataset placement into the underlying caching strategy can improve job efficiency further. Newer experiments and experiment upgrades such as HL-LHC and DUNE are expected to produce 10x the amount of data than currently being produced. This means that the available network, storage and compute resources must be utilized efficiently. Additionally, newer generations of DAQs are moving towards streaming readout systems, navigating away from the traditional triggered systems. These newer DAQ systems offer continuous/real-time data calibration, reconstruction and storage by offloading to remote sites results [2, 3]. These observations imply that the available network, storage and compute resources must be used efficiently.
The benefits in CPU efficiency and job turnaround times from colocating the datasets near computation are well-known, especially using dedicated or opportunistic cache storages using XCache or dCache systems [4]. Our prior work using the data transfer logs collected from the OSG dashboard revealed two major observations. First, there is a correlation between transfer time and the choice of storage site. The choice of source site, in case of storage redundancy, was found to be more important in transfer time than the actual file size (the dataset consists of many files). Second, there is not only a popularity skew in the remote files accessed by the jobs, but also files that are part of the same dataset are read more than others in that dataset.

<a href="https://lh3.googleusercontent.com/drive-viewer/AKGpihZWfXLIP$_k$KNZzPjOgiAymVeXWtwcXtXo0dK2Faxfmvc
33uDYMsbR5PJvJhL−8mAVoHFuu3SkMk−6JTpLRBDlDCWLXuf8Zyg = s2560?source =
screenshot.guru" >< imgsrc = "https : //lh3.googleusercontent.com/drive−viewer/AKGpihZWfXLIP$_k$KNZzPjOgi
33uDYMsbR5PJvJhL − 8mAVoHFuu3SkMk − 6JTpLRBDlDCWLXuf8Zyg = s2560"/ ><
/a >

File size vs. Transfer Time for data chunks with markers colored by the Data Source. The data in the figures are collected over 24-hour period in 2019. There are two clearly demarcated groups of transfer time values. Mid-Top left in figures show Group 1, a group of transfer with low transfer times. For this group, the files are served by a single data source (identified by dark yellow color). The second group, Group 2 (see the bottom group of data points) consists of smaller file sizes, and they show a wide distribution of transfer times.

<a href="https://lh3.googleusercontent.com/drive-viewer/AKGpihbP$_i$a71bJWtxq3TIuqafrALmtXmyrK64kAzCJzml
ELSzvubEFZTRtkU = s2560?source = screenshot.guru" >< imgsrc = "https : //lh3.googleusercontent.com/d
viewer/AKGpihbP$_i$a71bJWtxq3TIuqafrALmtXmyrK64kAzCJzmbOcwU77nb$_W$n98UFCWgxnfJk69zVgBX
ELSzvubEFZTRtkU = s2560"/ >< /a >

File size vs. Time of the Day plot of a single day from March of 2020 on the OSG. Left-side plot spans a single day in the March of 2020 and the right-side is a 2.5-hour snapshot of the same day. Each star represents a single file transfer. The color of the star represents a unique source from where the file was transferred from.

Using our analysis of the experiment and data pipelines, and the data access patterns in the US HEP environment, we present intelligent Machine Learning (ML)-based approaches to reduce the latency and improve job efficiency. We incorporate our prefetching and data placement techniques into a simulator by extending the WRENCH [5] simulator to reflect the experiment and data workflows typical in the US HEP infrastructure. The simulator is designed to reflect the US HEP environment (data, storage and computation workflows, and infrastructure). Simulators have long had an established history in planning and development of HEP infrastructure like WLCG (MONARC simulator) [6]. Simulations make it feasible to compare different network, storage and compute settings without building real testbeds for each setting. We present the proof-of-concept of our simulator with tight agreements to the real-world behavior of the experiments in US, specifically those belonging to LHC (CMS etc.) and DUNE.

REFERENCES

1. Fajardo, Edgar, Derek Weitzel, Mats Rynge, Marian Zvada, John Hicks, Mat Selmeci, Brian Lin et al. "Creating a content delivery network for general science on the internet backbone using XCaches." In EPJ Web of Conferences, vol. 245, p. 04041. EDP Sciences, 2020.

2. Lawrence D. Streaming Readout and Remote Compute. Thomas Jefferson National Accelerator Facility (TJNAF), Newport News, VA (United States); 2023 May 1.

3. Suiu, Alice-Florenţa. "EPN2EOS Data Transfer System." PhD diss., University POLITEHNICA of Bucharest, 2023.

4. Acosta-Silva C, Casals J, Peris AD, Molina JF, Hernández JM, Pérez CM, Dengra CP, Yzquierdo AP, Rodríguez FJ. A case study of content delivery networks for the CMS ex-periment.

5. Casanova, Henri, Suraj Pandey, James Oeth, Ryan Tanaka, Frédéric Suter, and Rafael Ferreira Da Silva. "Wrench: A framework for simulating workflow management systems." In 2018 IEEE/ACM Workflows in Support of Large-Scale Science (WORKS), pp. 74-85. IEEE, 2018.

6. Iosif C. Legrand and Harvey B. Newman. "The MONARC Toolset for Simulating Large Network-Distributed Processing Systems". Proceedings of the 32nd Conference on Winter Simulation. WSC '00. Orlando, Florida: Society for Computer Simulation International, 2000, pp. 1794–1801. ISBN: 0780365828.

Primary author: LAMBA KARANAM, Venkat Sai Suman (University of Nebraska-Lincoln)

Co-authors: Dr RAMAMURTHY, Byrav (University of Nebraska-Lincoln); WEITZEL, Derek (University of Nebraska Lincoln (US)); Mr BARLA, Sarat Sasank (University of Nebraska-Lincoln)

Presenter: Dr RAMAMURTHY, Byrav (University of Nebraska-Lincoln)

Session Classification: Parallel (Track 1)

Track Classification: Track 1 - Data and Metadata Organization, Management and Access