Contribution ID: **68**   Type: **Talk**

Comparing Cache Utilization Trends for Regional Data Caches

Thursday 24 October 2024 13:30 (18 minutes)

The surge in data volumes from large scientific collaborations, like the Large Hadron Collider (LHC), poses challenges and opportunities for High Energy Physics (HEP). With annual data projected to grow thirty-fold by 2028, efficient data management is paramount. The HEP community heavily relies on wide-area networks for global data distribution, often resulting in redundant long-distance transfers. This work studies how regional data caches [1] mitigate network congestion and enhance application performance, using millions of access records from regional caches in Southern California, Chicago, and Boston, serving the LHC's CMS experiment [2]. Our analysis demonstrates the potential of in-network caching to revolutionize large-scale scientific data dissemination.

We analyzed the cache utilization trends, see examples in Figure 1, and identified distinct patterns from the three deployments. Our exploration further employed machine learning models for predicting data access patterns within regional data caches. These findings offer insights crucial for resource usage estimation, including storage cache and network requirements, and can inform improvements in application workflow performance within regional data cache systems. Figure 2 shows the sample prediction result of the hourly cache utilization for SoCal Cache. While the predictions capture volume spikes, they may not precisely match the heights.
<img src="https://sdm.lbl.gov/students/chep24/chep24-size$_r$atio$_a$ll − socal.png" >
$Fig.1(a).SoCalCache(capacity2PB)fromJune2020toDec.2023,with29.48milliondataaccessrequestswhere23.8PBwererea$
$< imgsrc = "https : //sdm.lbl.gov/students/chep24/chep24 − size_ratio_all − chic.png"width =$
$"500"height = "100" >$
$Fig.1(b).ChicagoCache(capacity340TB)fromOct.2022toDec.2023,with5.7milliondataaccessrequestswhere11.0PBwerere$
$< imgsrc = "https : //sdm.lbl.gov/students/chep24/chep24 − size_ratio_all − bost.png"width =$
$"500"height = "100" >$
$Fig.1(c).BostonCache(capacity170TB)fromAug.2023toDec.2023,with27.7milliondataaccessrequestswhere5.0PBwererea$

Fig. 1. Fraction (by data volume) of cache hits (blue) and cache misses (orange) every day for three different in-network cache installations

<img src="https://sdm.lbl.gov/students/chep24/chep24-hourly-socal-hits$_s$ize.png"width = "500"height =$
$"100" >$
$Fig.2.HourlycachehitvolumefromSoCalCache, showingactualmeasurements(bluebars), trainingpredictions(or$

In this study, we present a detailed analysis of cache utilization trends across the SoCal, Chicago, and Boston regional caches serving the LHC CMS experiment. Our investigation reveals distinct patterns in cache usage unique to each region, highlighting the need for tailored approaches in cache design. Our exploration of neural network models were demonstrated to provide accurate predictions of cache usage trends, these models could be used anticipate future needs for storage and network resources. By understanding and leveraging these insights, we could significantly enhance the efficiency of resource allocation and optimize application workflow performance.

REFERENCES
[1] E. Fajardo, D. Weitzel, M. Rynge, M. Zvada, J. Hicks, M. Selmeci, B. Lin, P. Paschos, B. Bockelman, A. Hanushevsky, F. Wˆurthwein, and I. Sfiligoi, "Creating a content delivery network for general science

on the internet backbone using XCaches," EPJ Web of Conferences, vol. 245, p. 04041, 2020. [Online]. Available: https://doi.org/10.1051/epjconf/202024504041

[2] C. Sim, K. Wu, A. Sim, I. Monga, C. Guok, D. Hazen, F. Wurthwein, D. Davila, H. Newman, and J. Balcas, "Predicting resource utilization trends with southern california petabyte scale cache," in 26th International Conference on Computing in High Energy & Nuclear Physics (CHEP2023), 2023.

Primary author: SIM, Alex

Co-authors: GUOK, Chin; DAVILA FOYO, Diego (Univ. of California San Diego (US)); WURTHWEIN, Frank (UCSD); NEWMAN, Harvey (California Institute of Technology (US)); MONGA, INDER (LAWRENCE BERKELEY NATIONAL L); WU, John (LAWRENCE BERKELEY NATIONAL LABORATORY); BALCAS, Justas (California Institute of Technology (US)); MONGA, Ronak (Indiana University)

Presenter: WU, John (LAWRENCE BERKELEY NATIONAL LABORATORY)

Session Classification: Parallel (Track 1)

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