CryoDrill—Near-Data Processing in Deep and Cold Storage Hierarchies

Marcus Paradies
German Aerospace Center, Institute of Data Science
marcus.paradies@dlr.de

MOTIVATION

Nowadays, modern high-performance, high-resolution observational instruments and complex models of the earth system and of physical, chemical, and biological processes generate multiple hundreds of petabytes of scientific data per year contributing significantly to the infamous data deluge. Important application domains, which generate large amounts of scientific data, include earth observation, high-energy physics, radio astronomy, and weather forecasting, among others.

Increasingly, *digital data archives* store such scientific data in private cloud infrastructures for further investigation and long-term preservation, and disseminate them through *data platforms* via order-based catalogs. To reduce the total cost of ownership, such data platforms employ a hierarchical storage management system with large, disk-based caches and robotic tape libraries [2].

Often, data is stored in coarse-grained, value-added data products, whose granularity is determined by observational instead of data access characteristics. Fine-granular data tiling of such data products is often only applied late in the data analysis workflow. Prefetching all the data from a slower storage layer in advance is typically not possible due to the ad-hoc nature of scientific analysis tasks and the sheer size of the required data working set to achieve satisfactory results for long-term trend analysis and prediction.

With the proliferation of scientific data for general analysis by the academic community, wasteful data transfers across storage layers and unacceptable high data access latencies will likely become a major research barrier in the near future.

NEAR-DATA PROCESSING

Data movement is one of the most time- and energy-consuming tasks for data-intensive, scientific workflows. To overcome this challenge, near-data processing (NDP) has been advocated to reduce the amount of data to be transferred as early as possible [4]. Recent work proposed to leverage the processing power of modern SSD devices to offload computations directly into storage [1, 3]. Unfortunately, for large-scale scientific data, this only benefits the upper layers of the storage hierarchy, where data access is already quite fast (compared to slow tape data access). Further, NDP is only applied on a single storage layer, ignoring the fact that multiple NDP opportunities exist on the data path in the storage hierarchy. In a deep storage hierarchy, as it is common for active data archives, NDP can be by far more beneficial if pushed further down the storage hierarchy. For example, modern tape drive controllers are already quite powerful and natively support data encryption and compression and could be used as NDP devices as well.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

SYSTEM DESIGN

We propose CryoDrill (cryo- : "icy cold"), an NDP framework, which pushes selected parts of the computation of a complex scientific data analysis workflow down the storage hierarchy to enable processing close to the data while minimizing wasteful data movements up the storage hierarchy. CryoDrill specifically targets complex data analysis tasks on large amounts of scientific data residing in cold storage devices, such as *archival disks*, *massive-array-of-idle-disks* systems, and robotic tape libraries. In particular for petabyte-scale scientific data, multiple layers of disk-based caches aim at hiding the data access latency by introducing a deep storage hierarchy, which data has to traverse up to be finally processed.

To the best of our knowledge, CryoDrill is the first framework of its kind to consider NDP across multiple storage layers by hiding the labor-intensive task of manually selecting how to place parts of the processing onto heterogeneous computing resources. Since the overall goal is to reduce unnecessary data movement by applying data reduction close to where the data resides, we use a cost-based model, which takes intermediate result sharing opportunities within a storage layer and hardware characteristics into account, such as processing capabilities and available bandwidth resources. CryoDrill can use *in-storage* processing resources by extending storage controllers to run simple computation tasks, e.g., filtering, data tiling and tile selection, and aggregation, directly within the storage device or exploit *near-storage* processing capabilities through FPGAs. We propose a compiler-based approach, which can directly generate the target-specific operator variant on the specific hardware device, and an orchestration layer, which has full control over resource allocation & scheduling and manages data transfers across multiple processing devices.

REFERENCES

[1] Jaeyoung Do, Yang-Suk Kee, Jignesh M. Patel, Chanik Park, Kwanghyun Park, and David J. DeWitt. 2013. Query Processing on Smart SSDs: Opportunities and Challenges. In *SIGMOD'13*. 1221–1230.

[2] Matthias Grawinkel, Lars Nagel, Markus Mäsker, Federico Padua, André Brinkmann, and Lennart Sorth. 2015. Analysis of the ECMWF Storage Landscape. In *FAST'15*. 15–27.

[3] Boncheol Gu, Andre S. Yoon, Duck-Ho Bae, Insoon Jo, Jinyoung Lee, Jonghyun Yoon, Jeong-Uk Kang, Moonsang Kwon, Chanho Yoon, Sangyeun Cho, Jaeheon Jeong, and Duckhyun Chang. 2016. Biscuit: A Framework for Near-data Processing of Big Data Workloads. In *ISCA'16*. 153–165.

[4] Louis Woods, Zsolt István, and Gustavo Alonso. 2014. Ibex: An Intelligent Storage Engine with Support for Advanced SQL Offloading. *Proc. VLDB Endow.* 7, 11 (July 2014), 963–974.