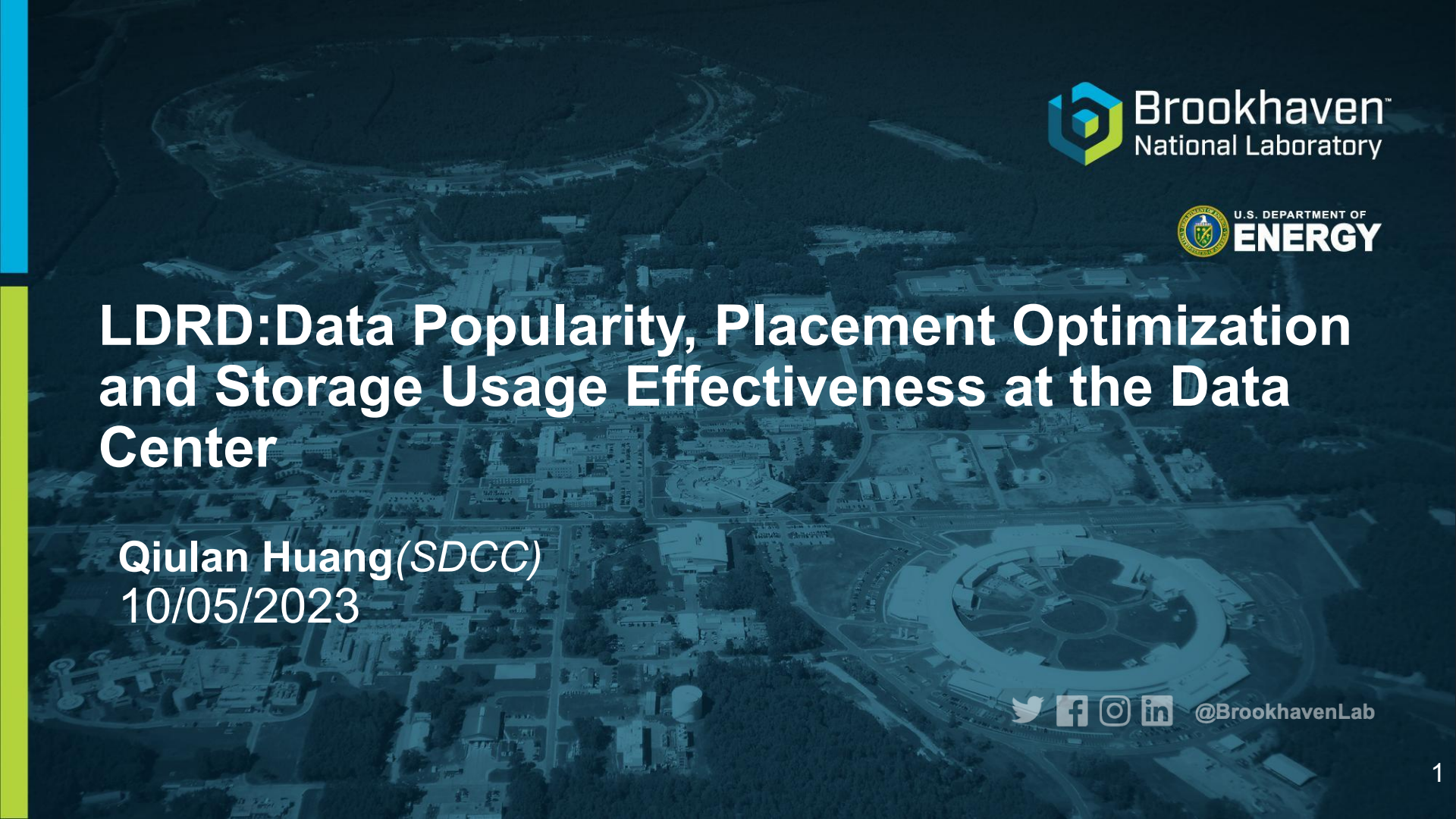


LDRD:Data Popularity, Placement Optimization and Storage Usage Effectiveness at the Data Center

Qiulan Huang*(SDCC)*

10/05/2023

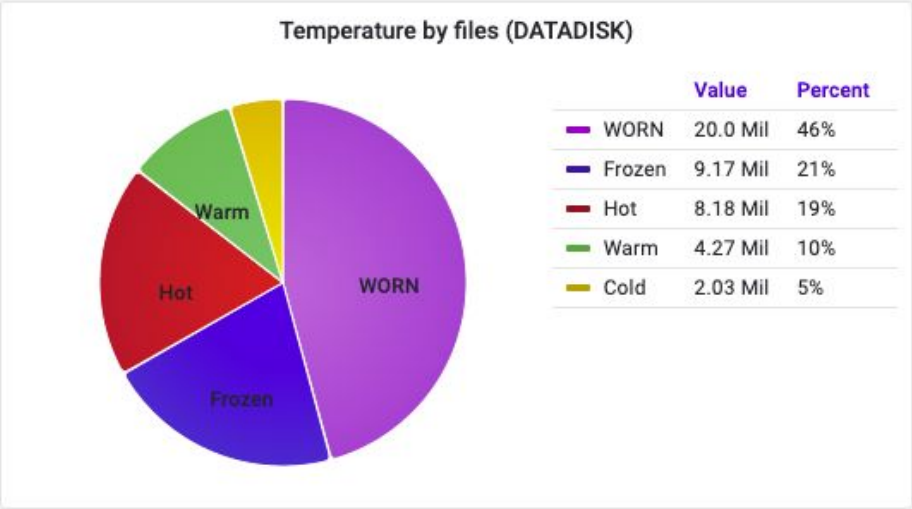
@BrookhavenLab

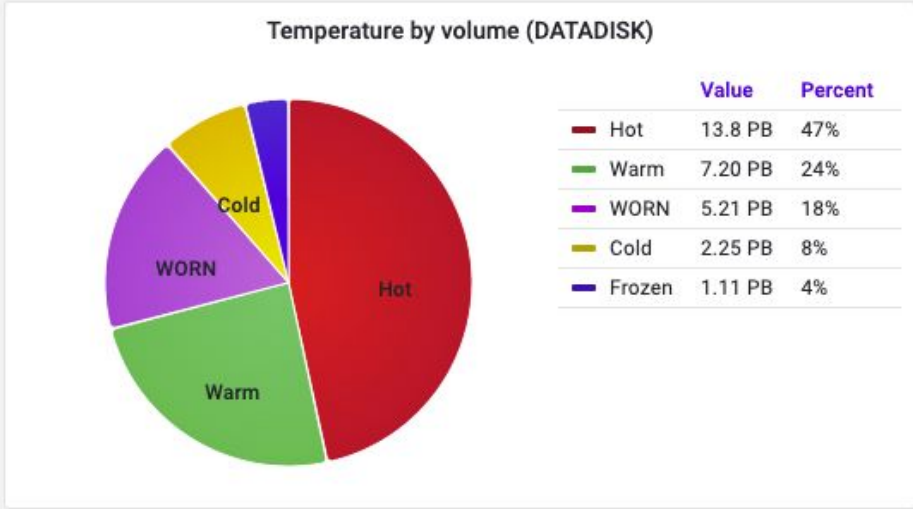
LDRD project

- Project title:Data popularity, placement optimization and storage usage effectiveness at Data Center
- People: SDCC/CSI (a cross-directorate project)
  - **Qiulan Huang (PI**, data collection,analytics, SDCC)
  - Vincent Garonne (Advisor, data mgt, SDCC)
  - Xin Dai (Large-scale data analysis, AI/ML, CSI)
  - Ai Kagawa (AI/ML algorithms, CSI)
- Term: 2022- 2024

Data Temperature

July, 2023: 43.65 M files, 29.57PB





Hot: Last access in the last month

Warm: Last access in the last 6 months

Cold: Last access between 6 months and one year

Frozen: Not accessed in the last year

WORN(Write Once Read Never): No access

Brookhaven National Laboratory

AI/ML For Storage Optimization

Motivation

- In the current multi-tier storage "class" system at the Data Center
  - Unused data is stored on expensive storage
  - Fast IO storage is not currently used effectively

Goals

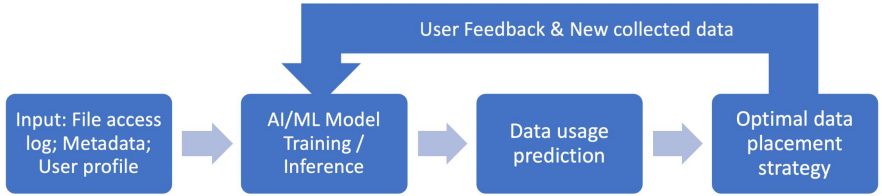
- Design an efficient monitoring platform to collect the relevant information from various distributed data sources
- Develop an optimal data management system for the data center to maximize usable space while minimizing access latency, within budget, hardware, and compliance constraints
  - Heavy use of storage, metadata and data popularity information
  - Develop a precise AI/ML prediction model to possibly forecast the future usage of the data
  - Orchestration of data for optimal movement and placement

Research approach

- Using ELK stack(Elasticsearch+Logstash+Kibana) for data collection and analytics
  - Collect and aggregate log data from a variety of sources using Logstash
  - Transform, process, and enrich log data using Logstash+Kibana
  - Index and search log data using Elasticsearch API
- Develop AI/ML models for the data popularity prediction(Baseline AI/ML Models)
  - XGBoost
  - Gradient Boosting
  - Random Forests
  - Linear Regression
  - Support Vector Machine
  - Deep Neural Network
- Based on the data popularity prediction results, will develop optimal data placement strategy

Competitive Advantages

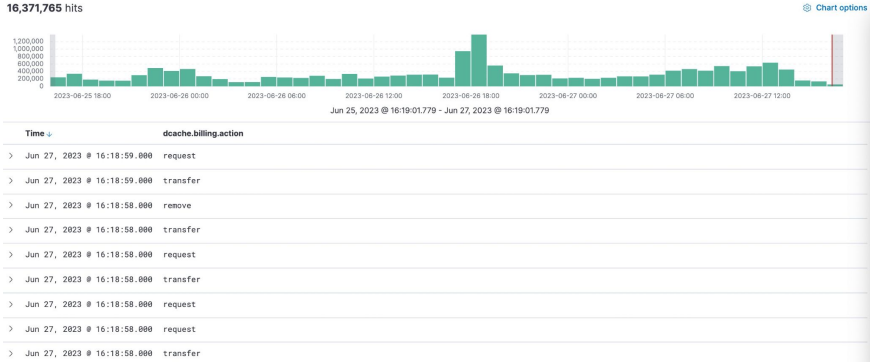
- Big scale storage has large volume of data to provide the big data statistics needed for AI
- Aggregate a variety of data sources to a central data storage
- Standard API to get data and preprocess data
- Utilize comprehensive data and be versatile to the change of storage infrastructure and user feedback



Data collection and Analysis

- Has collected data of the past 2 years
  - Data volume: ~11TB
  - **~10GB** in average per day, **5~8 million events** per day
  - Data source: billing logs, domain logs, etc from various experiments like usatlas, Belle2, etc

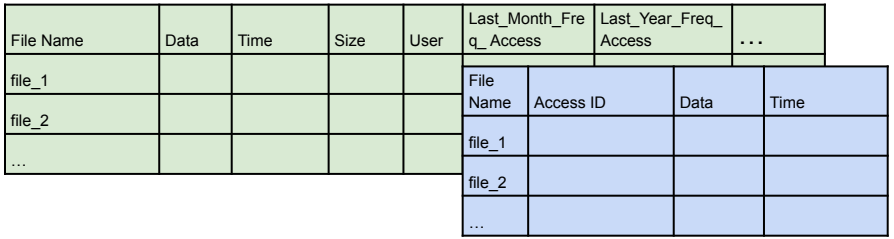
Time: one day	size	records
Raw data	13GB	5,604,498
Preprocessed data	2.7GB	5,604,498



- Maintaining this LDRD's GitHub organization to favor collaboration, expertise sharing and maintain contributions
  - Data collection and analysis.[repo]
  - Prediction algorithm codes,[repo]
  - Statistical Analysis and Visualization repository [repo]

Data preprocessing (1)

- Define the tabular data or comma-separated values (CSV) file format for AI model
  - file id, action (create, transfer, delete), file size, file type, user id, timestamp, file path



client,initiator,isnew,protocol,transfersize,fullsize,storageclass,connectiontime,action,cellname,datestamp,errorcode,errormessage,pnfsid,transaction,p2p,fqan,mappeduid,mappedgid, owner

2620:0:210:9800:0:0:0:58,door:WebDAV2-dcdoor32-externalipv6@webdav2-dcdoor32_httpsDomain:AAYE5RL/IaA:1694231773007000,f,Http-1.1,5241995048,5241995048,bnlt0d1:BNLT0D1@osm, 227423,transfer,dc247_6@dc247sixDomain,2023-09-09
00:00:00.471-04,0,"",00002E5FB57E25ED4ABB9EC77C50A91506F8,pool:dc247_6@dc247sixDomain:1694232000471-30895,f,,6435,31152,usatlas1
130.199.156.212,door:Xrootd-dcdoor26-internal@xrootd-dcdoor26Domain:AAYE4Wx9LwA:1694216094645000,f,Xrootd-5.0,1518066459,6277020158,bnlt0d1:BNLT0D1@osm,15906439,transfer,dc 217_17@dc217seventeenDomain,2023-09-09
00:00:01.264-04,0,"",0000C63A1287366948949756 73B942BA62B2,pool:dc217_17@dc217seventeenDomain:1694232001264-27578,f,/atlas,6439,69907,/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=663551/CN=Robot: ATLAS Pilot2
130.199.156.172,door:Xrootd-dcdoor36-internal@xrootd-dcdoor36Domain:AAYE5RwlLCA:1694231967445000,f,Xrootd-5.0,3640735851,3640735851,bnlt0d1:BNLT0D1@osm,33764,transfer,dc220_ 5@dc220fiveDomain,2023-09-09
00:00:01.314-04,0,"",0000805023C18CA54D9F8CDC96E10CB82EF0,pool:dc220_5@dc220fiveDomain:1694232001314-31426,f,/atlas/Role=production,6435,31152,/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo1/CN=614260/CN=Robot: ATLAS Pilot1

Data Preprocessing(2)

Datasets

- Our datasets consist of metadata records on the last access for several million files stored on the SDCC.
- Initial batches of records consisted of data labeled by **the datetime of last access, access type, file path, date, time, and other details**.
- Later batches were further preprocessed and have labels such as **client, initiator, protocol, transfer size, full file size, storage class, connection time, file ID, and file owner.**

Data Preprocessing:

Two datasets

1. “Reduced set” of 2300 hot data and 1900 cold data records
2. “Large set” of 15000 hot data and 15000 cold data records

- The dataset used in this study is first labeled as "hot" or "cold" based on the timestamp information.
- Inputs of the model: Categorical columns (Action, File Path, and details) are converted to **one-hot encoding** as inputs for our model training.

Model and Results

This work was accepted for the undergraduate student poster at SC23

Model Architecture:

- Input of the model: one-hot encoding of the Categorical columns
- A deep neural network model consisting of three fully connected layers
- Each is followed by a ReLU activation function.
- Output of the model: hot-or-cold classification

Model Training:

- The preprocessed dataset is randomly split 80/20 into training and testing sets.
- The model is trained on the training set using the Adam optimizer and using the Cross Entropy Loss function, with a hidden size of 64 neurons and learning rate of 0.001.
- The training loop iterates for 10 epochs, and the model's parameters are updated to minimize the loss.

Model Evaluation:

- After training, the model's performance is evaluated on the testing set to assess its predictive accuracy, precision, and recall.

	Reduced Set	Large Set (Hot)	Large Set (Cold)
Accuracy	82.15%	90.53%	--
Recall	76.36%	81.03%	99.87%
Precision	94.27%	99.67%.	84.37%

Distributed Data Parallel (DDP)

- **Distributed Data Parallel (DDP)**, allowing our model to utilize multiple GPUs in multiple nodes using the BNL's Institutional Cluster (IC), further accelerated our analysis.
- Utilizing the PyTorch's "Torch Distributed Elastic" feature.
  - Worker failures are handled gracefully by restarting all workers.
  - Worker `RANK` and `WORLD_SIZE` are assigned automatically.
  - Number of nodes is allowed to change between minimum and maximum sizes (elasticity).
- Preliminary Scaling Results using BNL IC AI cluster
  - NVIDIA V100 GPUs.

# of GPUs	# of Nodes	Average Run Time per Epoch (sec.)	# of batches processed per GPU in each epoch	Batch Size
1	1	82.5	188	32
4	2	44.9	47	32

Summary

- The project is progressing steadily according to the plan
  - Various data sources have been collected, aggregated and being processed
  - Evaluate the baseline model and train data
  - **Completed the LDRD Mid-year review**
- Supported a SULI student's work, which has been accepted for undergraduate student poster at SC23
- A potential new program "Tiering storage at a data center" will grow from the work
  - Will introduce the work like data popularity prediction model to decide data placement among various storage classes
- Next step
  - Complete the develop the state-of-the-art data usage prediction model
  - Quality Assurance & Control of training data & performance monitoring
  - Implement the policy engine and data placement tool
    - Put hot data on good pools and cold data on the old generation pools
  - A student intern will join