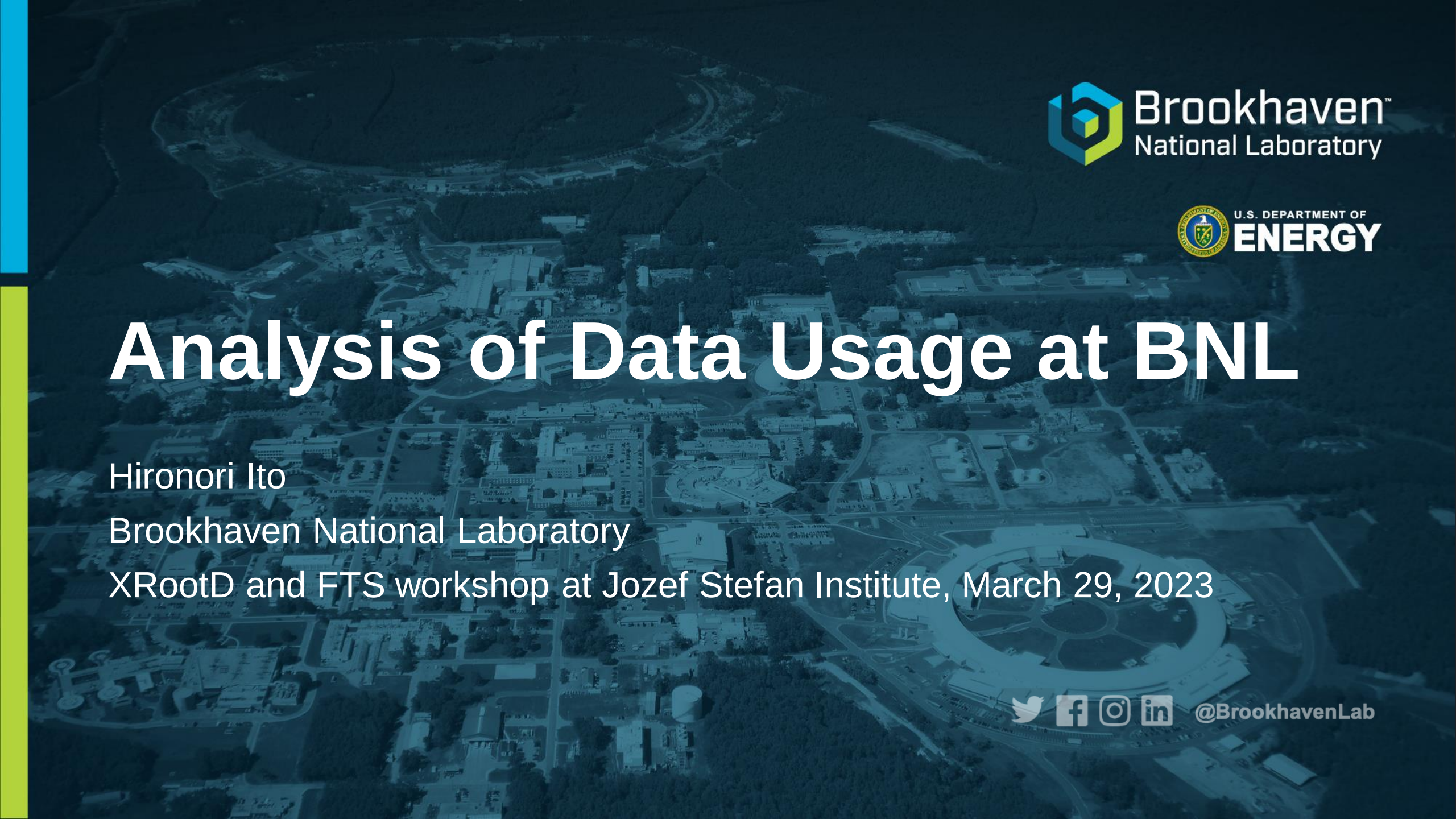


Analysis of Data Usage at BNL

Hironori Ito
Brookhaven National Laboratory

XRootD and FTS workshop at Jozef Stefan Institute, March 29, 2023

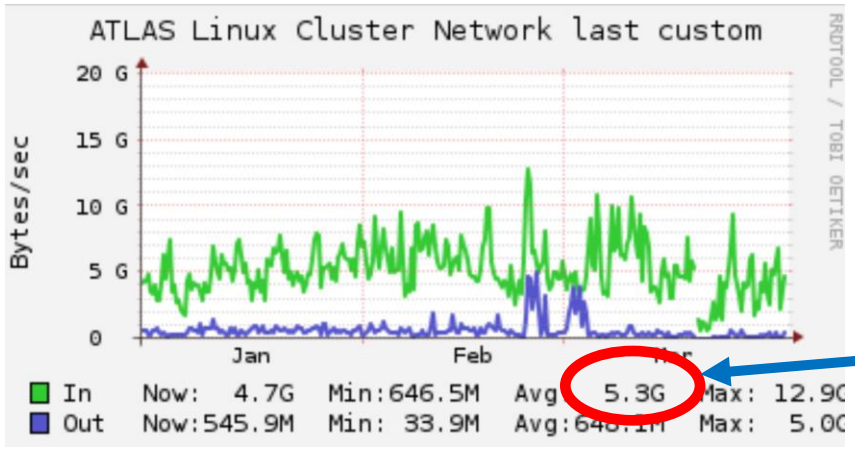
Brookhaven National Laboratory
U.S. DEPARTMENT OF ENERGY
@BrookhavenLab

Analysis of data usage from storage logs

- There are more interests to see how much of the data are actively being used.
  - More attention has been given to how much data we can move between sites. But, once the data gets there, small attention is being given if the data is actively used.
- Experiments are coping the disk usage rate for now. But, that might not be easily attainable if no attention is being paid to the data usage on the storage.
- The analysis is done from the storage logs of ATLAS dCache at BNL using TimescaleDB.
  - At BNL, almost all “Read” access by jobs are by XRootD protocol while site-to-site data transfers are by WebDAV.

Checking logs

- dCache records all access information in logs.
- The recorded information includes path, file size, read size, access time, etc…
- Is it really accurate?



Network IO seen from the worker nodes. Values given by NIC on the hosts.

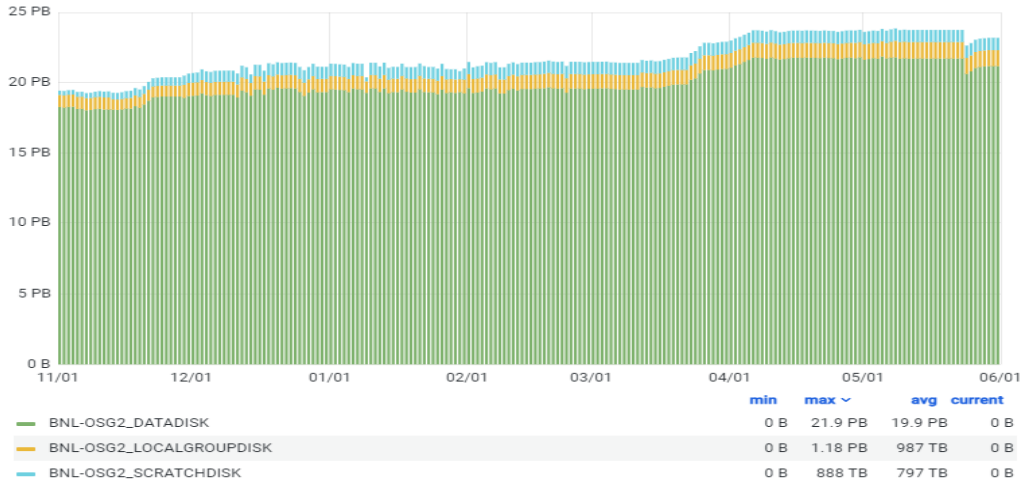
From dCache access logs and summing over all XRootD access during the same period, the average access rate is **5.16** GB/s

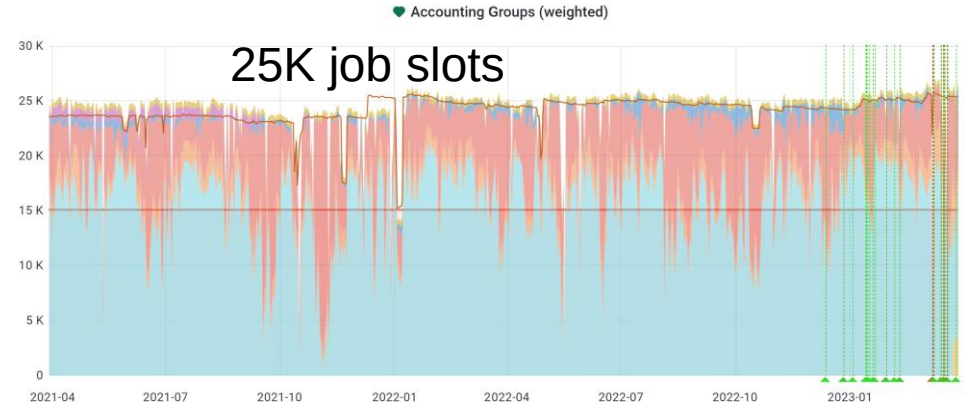
Data volume received in the worker nodes are very similar to the data volume accessed in dCache, indicating the reasonable consistency between the two.

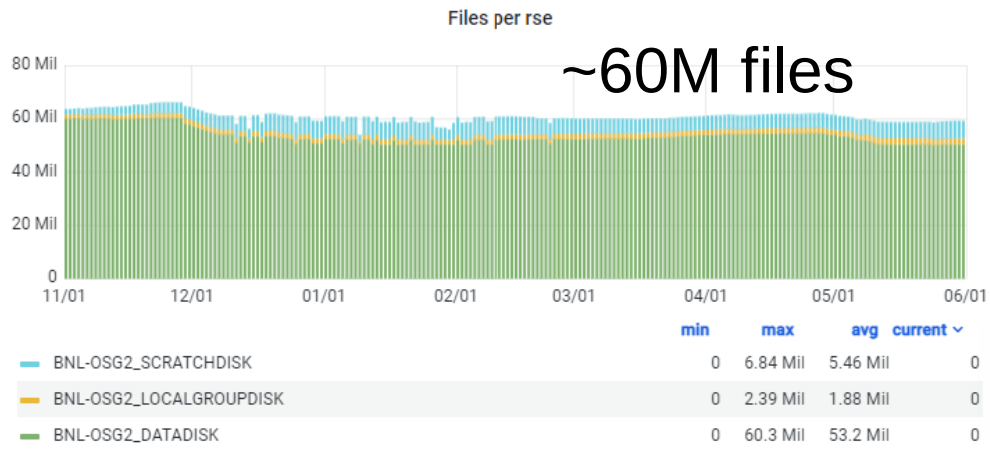
Size of storage at BNL for ATLAS

- BNL dCache storage contains about 60M files corresponds to about 21PB in size during the period of Nov 2021 to June 2022
- BNL has ~25K job slots for ATLAS.

~21PB





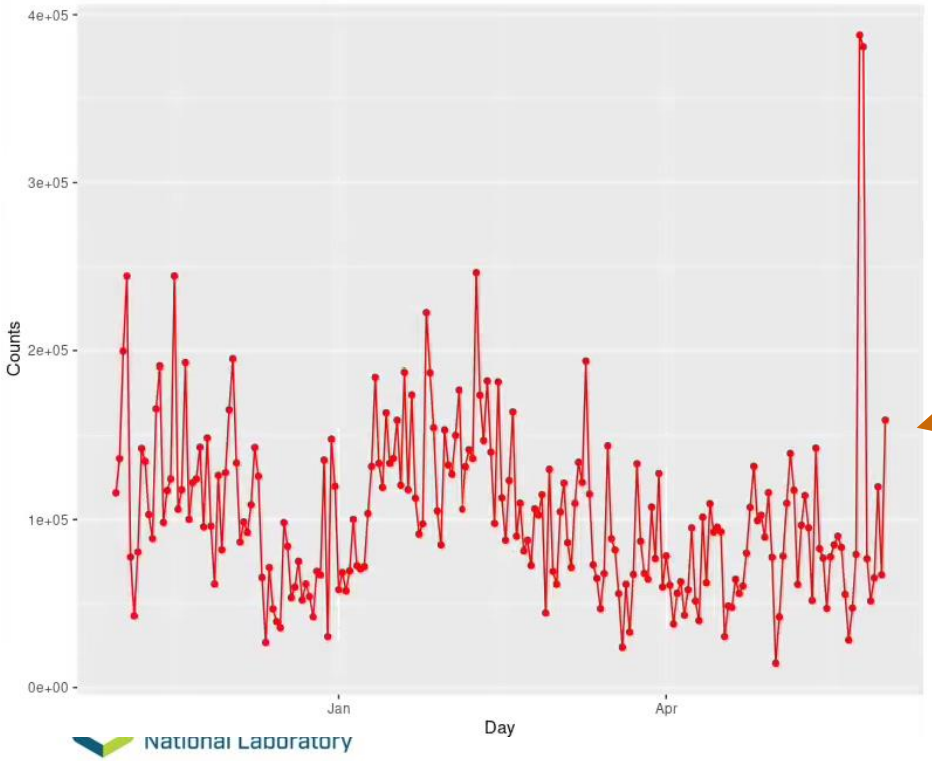


Brookhaven National Laboratory

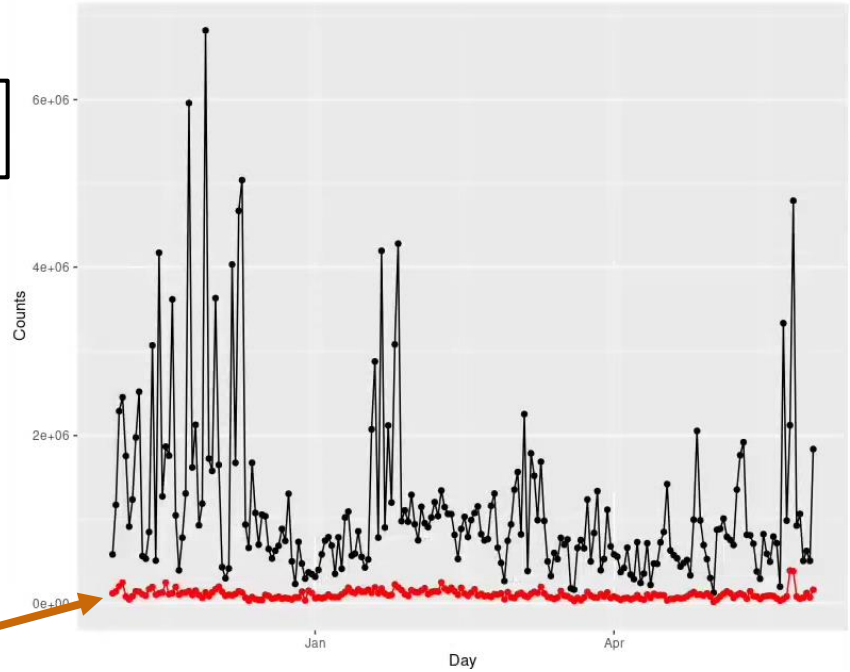
Daily Usage of files by jobs at BNL

~ 0.2% of total number of files are used per day

Total number of **unique** files used per day ~ 100K



BNL has
~60M files
~21PB



Total number of files used per day ~ 1M

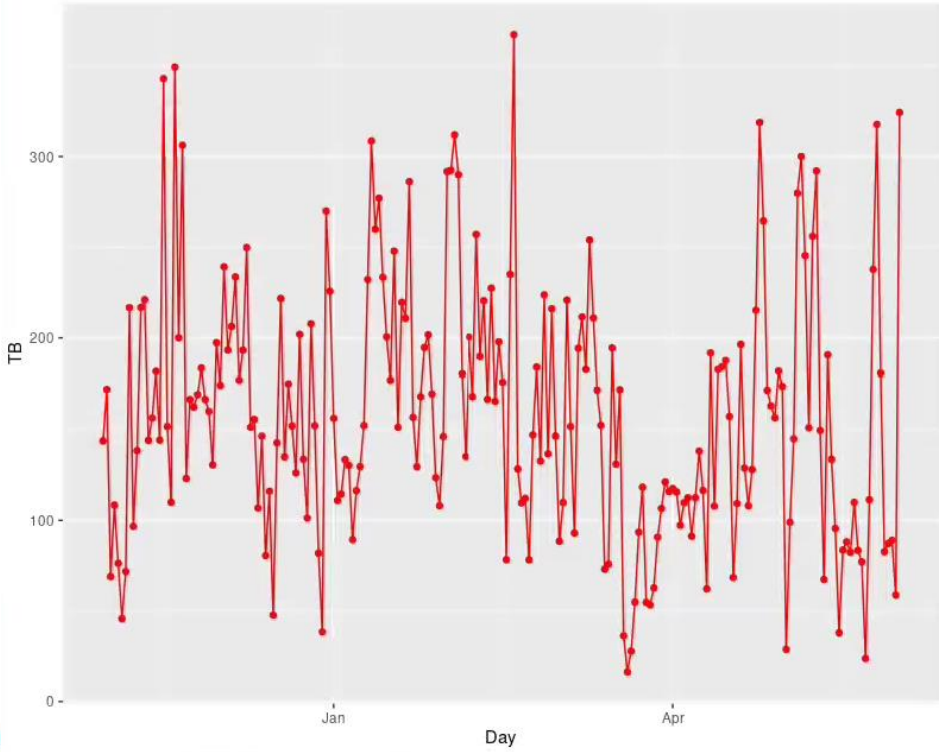
One job can access the same files multiple times in short time. Similar jobs can also access the same files multiple times in short time.

Note: it does not include the data out from BNL to the other sites through WebDAV.

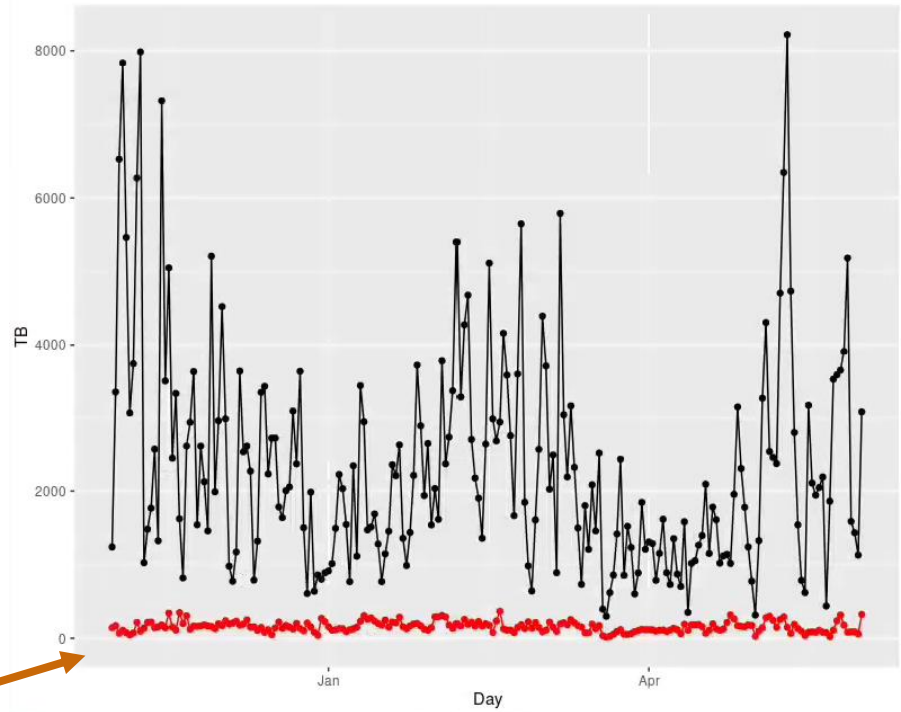
Daily data volume used by jobs at BNL

~ 0.8% of total data volume are used per day

Total size of **unique** used files per day ~ 150TB



BNL has
~60M files
~21PB



Note: it does not include the data out from BNL to the other sites.

Usage per Day, Week and Month

BNL has
~60M files
~21PB

Scale with period → Not too many reuse

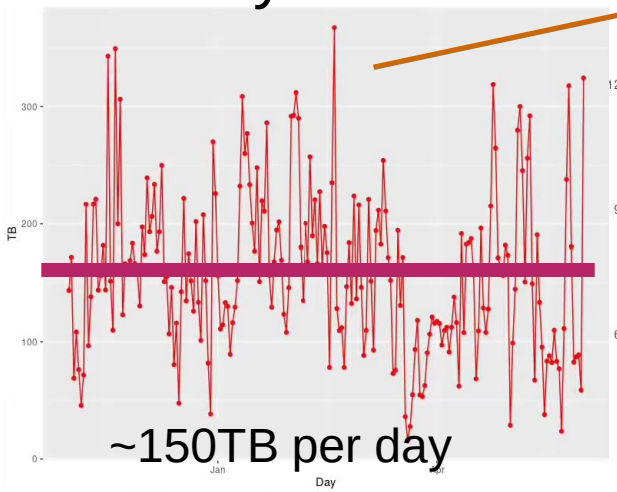
Daily

~7 times

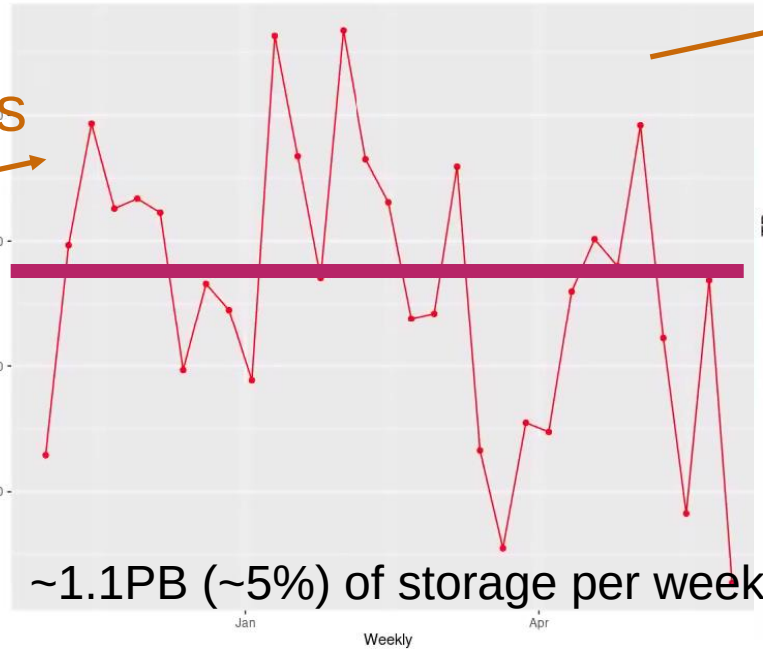
Weekly

~4 times

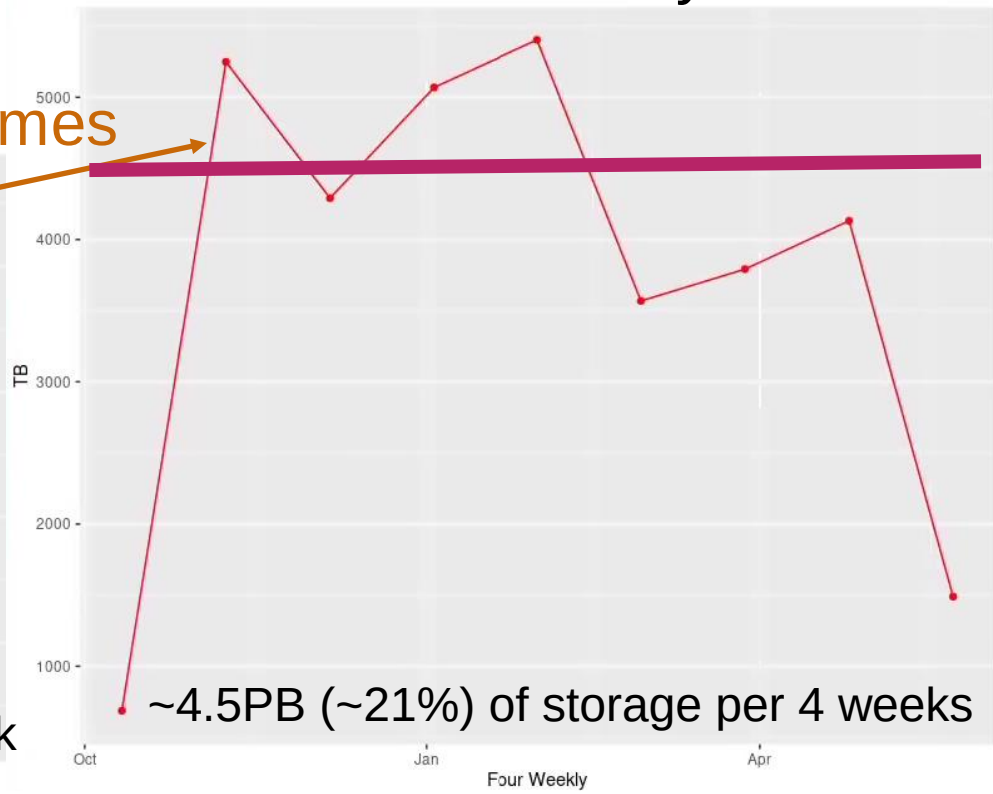
Monthly



~150TB per day



~1.1PB (~5%) of storage per week



~4.5PB (~21%) of storage per 4 weeks

NOTE: Ignore the first and last bins for weekly and monthly plots

Some other stats

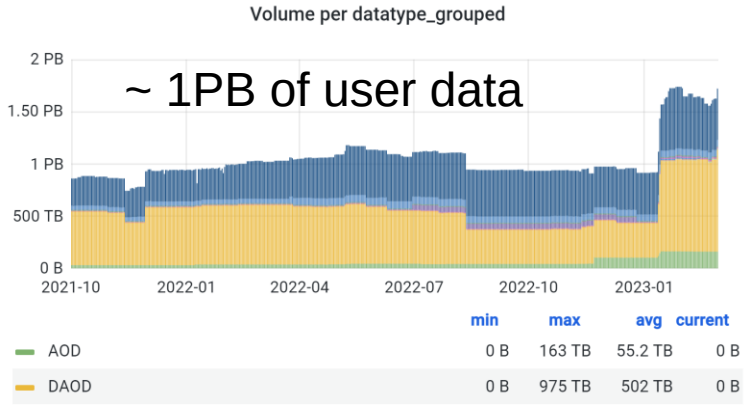
- For the files that are written and read at least once by jobs between Nov 2021 to May 2022, the average time between the file creation time and 1st access time is 6.5 days.
  - $\text{Average}(t_{1st\ atime} - t_{ctime}) = 6.5\text{days}$
- For the files that are written and read at least once by jobs between Nov 2021 and May 2022, and deleted between the same period, the average time between last access time and deletion time is17 days
  - $\text{Average}(t_{delete} - t_{last\ atime}) = 17\text{days}$
- For the files that are written and deleted between Nov 2021 and May 2022, average time between the file creation and deletion time is 29 days
  - $\text{Average}(t_{delete} - t_{ctime}) = 29\text{days}$
- The same analysis is done at different site (AGLT2).
  - The data usage rate at AGLT2 are very similar to that of BNL with scaled by the number of job slots.
  - Proportionality factor with the site HS06 value. ~ 0.4TB/day/KHS06

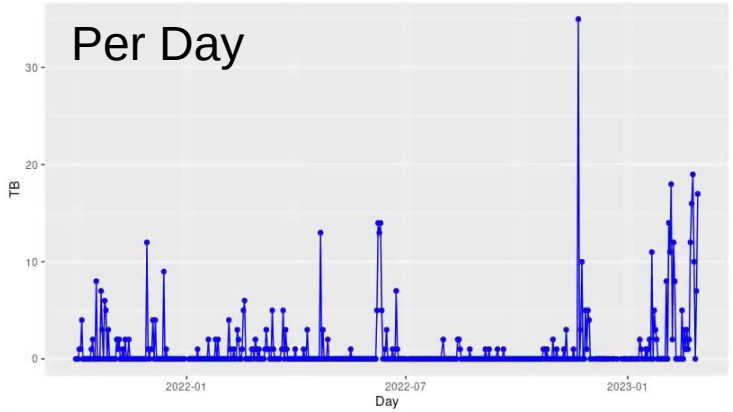
User storage area

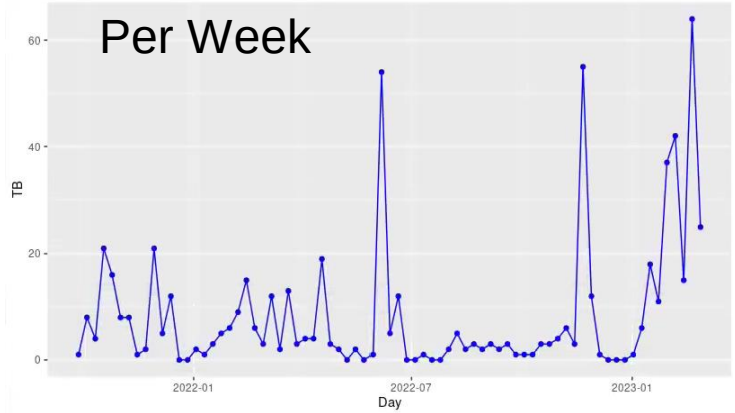
- User data stored in user area (LOCALGROUPDISK)

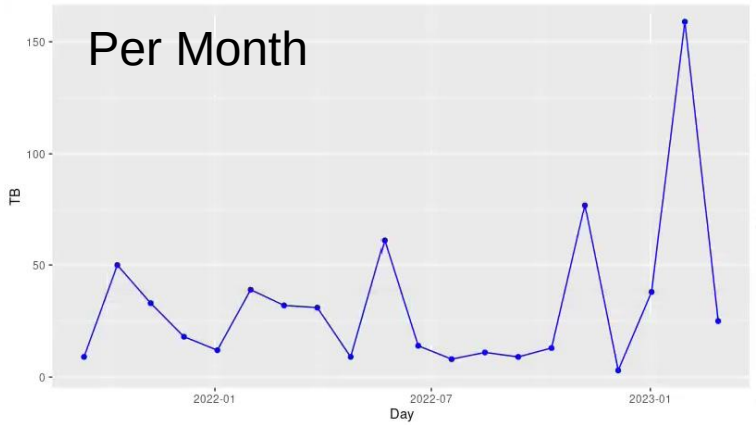
- Unique data read per day, week and month
  - Seems to show more reuse
    - Not linearly scaling with length of the period
  - Seems to show less total usage than the production area
    - ~5% per month vs ~20% per month

Should it be pure “Cache”?









Brookhaven National Laboratory

Conclusion

- The production data see less reuse over the long period.
  - Large fractions of the data are not used frequently by local jobs.
  - Different data type might see different rate of reuse.
- The user data see more reuse.
  - Pure cache might work
  - Is it more efficient in terms of cost and labor?
  - What kind of cache?
    - Tape backed pure XCache?