

# Data Popularity

BNL-UMISS

# Trace:

- Rucio uses a tracer system that receives a **trace every time a file is either used in grid jobs or downloaded**
- Following file metadata is updated by a rucio demon according to trace:
  - **accessed at** : timestamp when the file is accessed
  - **access\_cnt** : number of times the file is accessed

```
$ rucio get-metadata /belle/user/soumen/Kp11_v1_charged-unskimmedB4/sub00/Kp11_v1_charged-unskimmedB4_00255_job201442422_01.root
accessed_at:      2021-08-30 15:26:31
```

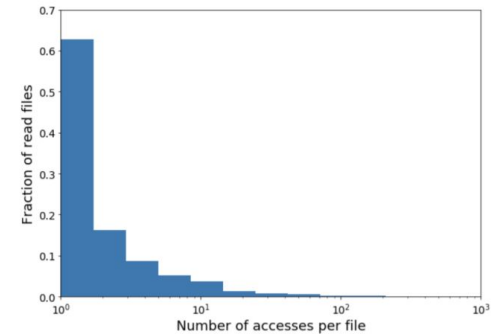
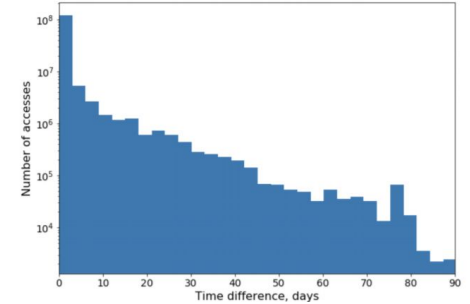
# Trace:

- Trace can send other info:
  - **Scope**
  - **LFN**
  - **Clientstate**: what state the trace is set. E.g. DONE, FAIL\_VALIDATE ,e.t.c
  - **Account** : which account accessed the file..
  - **Remotesite**: site where data is accessed to.
  - **Localsite** : site where data is accessed from.
  - **EventType**: what event type is used in file access. E.g. DOWNLOAD e.t.c

What info to send ?

# Data Popularity:

- Based on the trace, we can get lot of info.
  - Number of times file is accessed.
  - From where a file is accessed.
- This info can be used to understand,
  - Popularity of files
  - Location of accessed place.
  - Dynamic data placement e.t.c



## News:

- Infrastructure for sending trace to Rucio is in place.
- Thanks to BNL team.

Work should start from BelleDIRAC side.

# Send trace from ?

## 1. Grid jobs:

- Once the Worker node download files to run basf2 job -> send trace.
- For both user and prod jobs ?
- For **user Jobs**, download of files to worker node is done by :
  - [wms/Client/BelleDownloadInputData.py](#)
  - Q. **Only those successful download should send the trace ?**
- For prod Jobs:
  - Need to talk to Miyake-san.

# Send trace from ?

## 2. gb2\_ds\_get:

- After the successful download ?
- Plan to move to rucio download, which sends trace automatically.

## 3. ANY other ?

Backup



# Test:

```
[apanta@ccw07 ~]$ gb2_ds_list /belle/user/anil123/xi2sigpipi_procllab_fresh/sub00/lambda2sigpipi_00259_job184626651_00.root
/belle/user/anil123/xi2sigpipi_procllab_fresh/sub00/lambda2sigpipi_00259_job184626651_00.root
[apanta@ccw07 ~]$ gb2_ds_list /belle/user/anil123/xi2sigpipi_procllab_fresh/sub00/lambda2sigpipi_00259_job184626651_00.root -lg
KEK-DISK-TMP-SE: (1/1), 110.1 KB
    /belle/user/anil123/xi2sigpipi_procllab_fresh/sub00/lambda2sigpipi_00259_job184626651_00.root
total 1 good files, 110.1 KB
[apanta@ccw07 ~]$ python
Python 2.7.13 (default, Feb 24 2021, 11:28:05)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-23)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from json import dumps
>>>
>>> import requests
>>> trace = {'scope': 'user.anil123', 'filename': '/belle/user/anil123/xi2sigpipi_procllab_fresh/sub00/lambda2sigpipi_00259_job184626651_00.root', 'clientState': 'DONE', 'account': 'anil123', 'localSite': 'KEK-DISK-TMP-SE', 'remoteSite': 'KEK-DIST-TMP-SE', 'eventType': 'download', 'clientState': 'DONE', 'eventVersion': '1'}
>>> r = requests.post('https://blruciotracer.sdcc.bnl.gov/traces', dumps(trace))
>>> print(r)
<Response [201]>
>>> exit()
```

```
[apanta@ccw07 ~]$ rucio get-metadata /belle/user/anil123/xi2sigpipi_procllab_fresh/sub00/lambda2sigpipi_00259_job184626651_00.root | grep accessed_at:
accessed_at:      None
```

# Trace:

Attribute	Description
UUID	Unique identifier shared by traces originate from the same operation
Event Type	Application type initiating download, e.g. an analysis job
Tool Version	Release version of grid application
Remote Site	Site where data is being downloaded/uploaded to
Local Site	Site where data is being downloaded/uploaded from
Time Start	Time operation started
Time End	Time operation ended
DUID	Unique dataset ataset identifier
Version	Version of Dataset
Client State	Final state of client; either done or error code
File Size	Size of file
GUID	Unique identifier for file
Catalog Time	Time stamp when file catalogs were queried
Transfer Time	Time stamp when first byte arrives
Validation Start	Time stamp when validation began
Application ID	Operation ID, assigned by the application
Hostname	Name of Host
IP	Host's IP address

Table: Attributes and descriptions for the DQ2 Tracer