Understanding Data Popularity and Optimizing Access for Analysis

Research conducted by
Avi Kaufman

What's the Problem?

- Data produced by the LHC have reach the exabyte level of storage requirements.
- The ATLAS database has over 1.5 thousand users submitted over 10 million tasks each year
- This shows that storage capacity and computing resources are critical issues for the HL-LHC run (Run 4)
- With such a large database, access time becomes an issue for users requesting datasets.

What is Data Popularity?

- Dataset/file popularity is a metric showing how often a dataset/file was used as an input for physics analysis within a certain time interval
- Data Popularity metrics allow us to determine which datasets/files should be readily available for analysis
- A good popularity metric can help predict when Datasets/files will be used next, thus increasing efficiency of the Worldwide LHC Computing Grid (WLCG) computing infrastructure
    - This is referred to as “reuse time” and calculating the reuse time is critical to this project

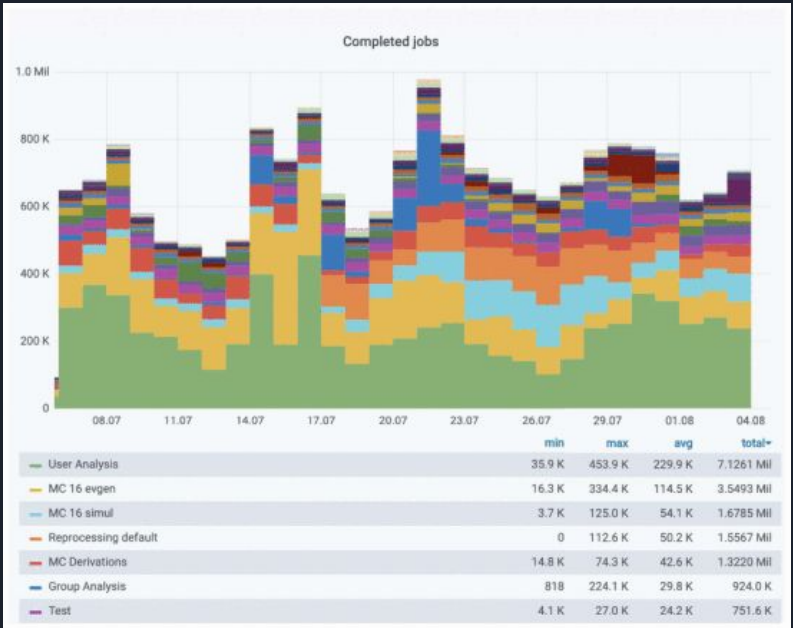
Dcache

- The dCache Storage management system is used as a disk cache for large high-energy physics datasets primarily from the ATLAS experiment
- Files and Datasets present in the dCache have faster access time than files not stored in the dCache
- Storage space is considerably smaller on the dCache than the full data collection
- A primary goal is developing a policy or algorithm to determine which files need to stay in the cache in which files need to be evicted
- In order to do this calculating the reuse time of datasets is crucial

How do we calculate data popularity?

- First we need to collect data on the datasets themselves. ATLAS_PANDA provides data for the WMS PanDa system and we can collect data from tables within ATLAS_PANDA
- Four Tables of use are:
  - ATLAS_PANDA.JEDI_DATASETS - metadata about datasets used in analysis tasks
  - ATLAS_PANDA.JEDI_TASKS - provides the information about all tasks
  - ATLAS_PANDA.JOBSARCHIVED4 - metadata about completed PanDA jobs for the previous 4 days
  - ATLAS_PANDAARCH.JOBSARCHIVED - archived metadata about completed PanDA jobs for last 3 years

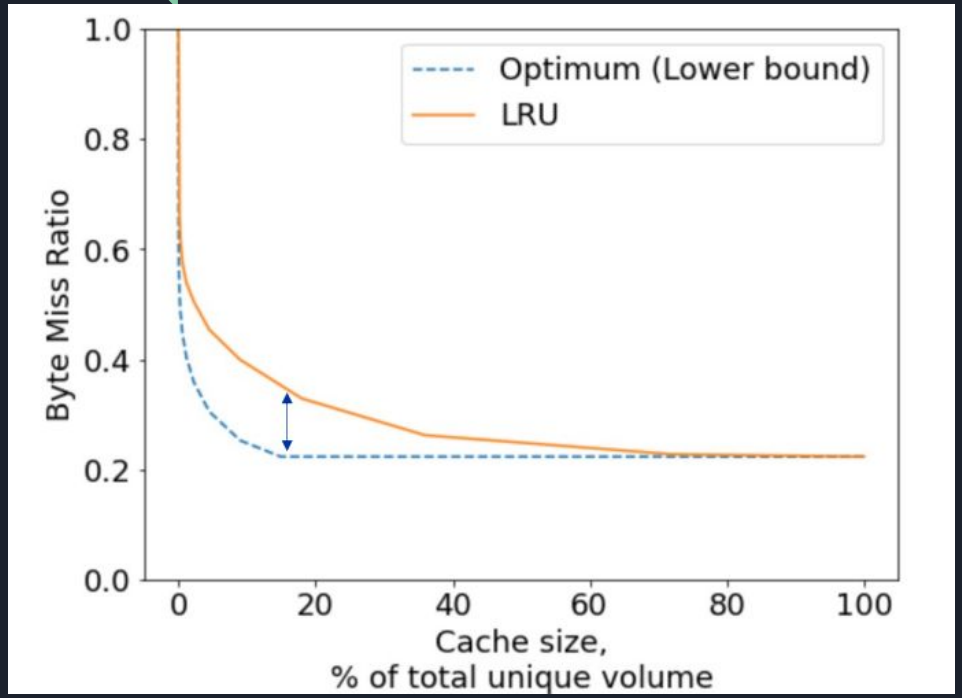
User analysis vs Central Production



	min	max	avg	total
User Analysis	35.9 K	453.9 K	229.9 K	7.1261 Mil
MC 16 evgen	16.3 K	334.4 K	114.5 K	3.5493 Mil
MC 16 simul	3.7 K	125.0 K	54.1 K	1.6785 Mil
Reprocessing default	0	112.6 K	50.2 K	1.5567 Mil
MC Derivations	14.8 K	74.3 K	42.6 K	1.3220 Mil
Group Analysis	818	224.1 K	29.8 K	924.0 K
Test	4.1 K	27.0 K	24.2 K	751.6 K

- User Analysis jobs are among the most common of all activities
- In comparison to Central Production (All other jobs) User Analysis jobs are unplanned and harder to predict
- Central Production jobs are planned and thus the required resources can be estimated in advance
- Developing data popularity metrics for user analysis tasks is of high priority

Increasing hit rates in the dCache

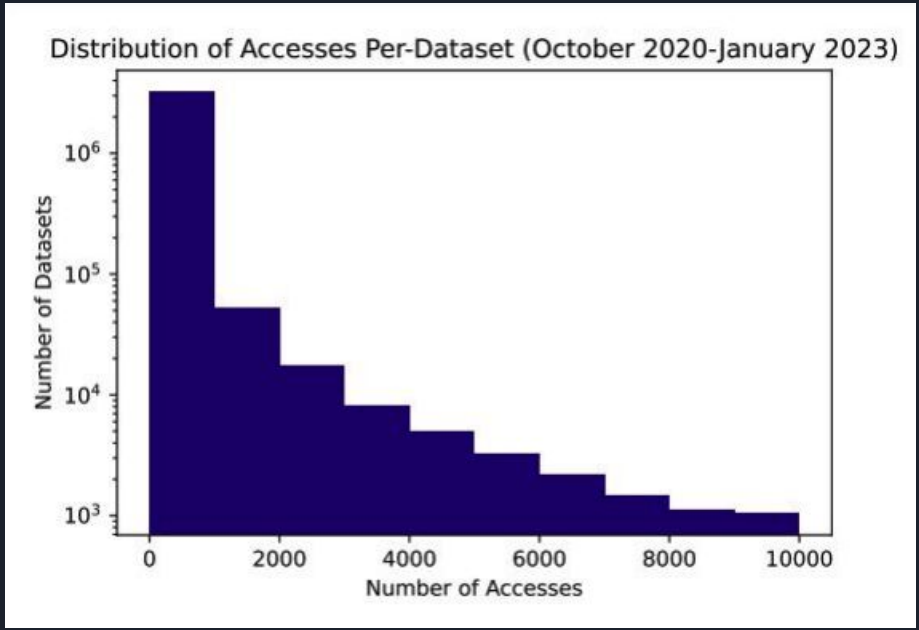


- A hit is when a file or dataset being requested is in the dCache
- A miss is when the file or dataset is not in the dCache
- LRU (Least Recently Used) is a caching algorithm that removes the least recently accessed items from a cache when space is limited.
- LRU serves as a benchmark for future success

Other popular caching algorithms

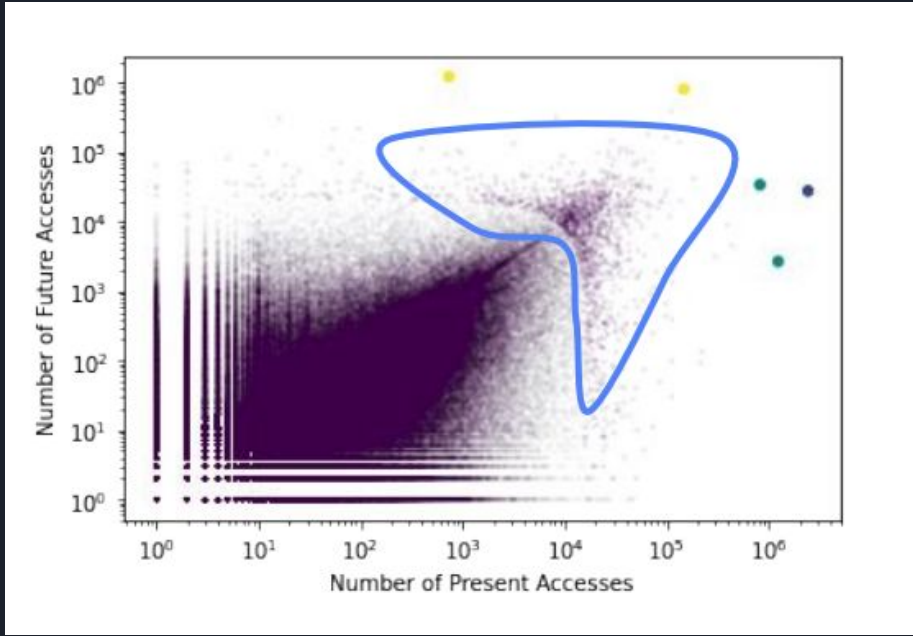
- Least Frequently Used (LFU)
- Most Recently Used (MRU)
- Random Replacement
- Adaptive Replacement Cache (ARC)
- Low Inter-reference Recency Set (LIRS)
- Two-Queue (2Q)
- Clock

Skewed Distribution of dataset popularity



- Majority of data sets have less than 1000 accesses
- A select few have significantly more activity
- This skewed distribution can be taken into account to optimize the dCache storage

Grouping datasets by total Accessing using K-means clustering

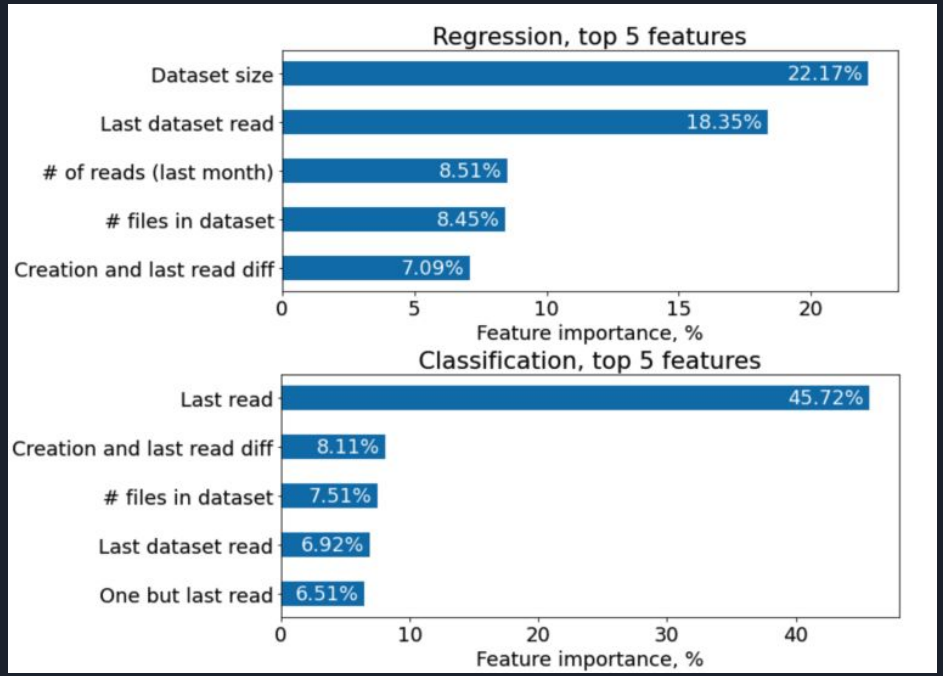


- Jullian Bellavita and her colleagues performed a clustering algorithm on the datasets in ATLAS
- There are two distinct groups
  - Less than 10^3 present and future accesses
  - More than 10^3 present and future accesses
- This information can be used to ping datasets in the more than 10^3 cluster and keep them in the dCache in order to optimize access time

Estimating reuse time using Random Forests

- Olga Chuchuk and Markus Schulz trained a random forest algorithm on 75 days of training data from the ATLAS database and tested their model on 15 days of data
- Advantages of random forest:
  - Classification and regression
  - No need for feature normalization
  - Parallelizable
  - Interpretable
- Results:
  - Regression - Predicts the logarithm of reuse distance
    - RMSE: 0.34
  - Classification - Predicts if reread in the next 15 days
    - RMSE: 0.12
- Does not outperform LRU

Important Parameters/Features



- Unsurprisingly, last read appears to have a high correlation with predicting future reads
- Dataset size is #1 for the regression model but not for classification
- Last dataset read and # of files in dataset are among the top 5 for both

Rucio and WMS PanDa

- Data handling and processing for ATLAS is currently handled by two systems:
  - Rucio
  - Workload Management System (WMS) PanDa
- Rucio
  - A system that keeps track of all ATLAS data stored on different sites. It helps organize and manage the data, ensuring it is available and can be transferred efficiently between storage resources. Rucio also controls who can access the data and sets limits on usage
- PanDa
  - A system that schedules and manages computational jobs for the ATLAS experiment. It assigns jobs to available computing resources based on their availability and requirements. PanDA keeps track of job progress, handles failures, and ensures that the computing tasks are executed effectively and reliably

Current Work

Research conducted by
Avi Kaufman

Using py scripts to parse large datasets

- Input file is rawdata-0425.csv:
  - Structure:
    - 69 thousand rows
    - 100 JSON objects per row
    - 6.9 million JSON objects
    - 15.1 GB of data
- Output file:
  - Structure:
    - file id, action (create, transfer, delete), file size, file type, user id, timestamp, file path

Problems

- Large files can’t be opened all at once
- Each cell is its own JSON object
- Many JSON objects have different formats
- Testing whole file takes a long time

JSON format examples

First Row JSON

- {'_index': 'dcache-billing-usatlas-000055', '_type': '_doc', '_id': '8viGu4cBOBpu8Q1fxLdS', '_score': 1.0, '_ignored': ['message.keyword'], '_source': {'@version': '1', 'log_type': 'dcache-billing', 'ecs': {'version': '8.0.0'}, 'event': {'action': 'remove', 'dataset': 'dcache.billing', 'provider': 'dcache', 'created': '2023-04-26T03:06:22.882Z', 'category': 'network', 'kind': 'event', 'module': 'dcache'}, 'dcache': {'billing': {'client': {'address': 'unknown', 'user': {'dn': 'usatlas1', 'id': '6435', 'group': {'id': '31152'}}}, 'action': 'remove', 'storage': {'class': 'Unknown'}, 'queued': {'time': 0}, 'p2p': False, 'transaction': {'time': 0}, 'error': {'code': '0', 'message': ''}, 'ts': '2023.04.25 23:06:21', 'cell': {'name': 'WebDAV-dcdoor13-external@webdav-dcdoor13_httpsDomain', 'type': 'door'}, 'file': {'size': 23621136071}, 'pnfsid': '00002FEEA4A0E44E4533B842359CEF5CBA28', 'protocol': {'path': '/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc21_13p6tev/80/27/rdo.32721912._006682.pool.root.1'}}}, 'client': {'address': 'unknown', 'user': {'dn': 'usatlas1', 'id': '6435', 'group': {'id': '31152'}}}, 'host': {'name': 'dcache-billing'}, 'log': {'file': {'path': '/var/lib/dcache/billing/2023/04/billing-2023.04.25'}, 'offset': 2380207878}, 'service': {'name': 'billing', 'type': 'dcache', 'version': '7.2.16'}, 'labels': {'env': 'prod', 'type': 'logs', 'tier1': {'category': 'storage'}}, 'type': 'dcache-billing-logs', 'message': '04.25 23:06:21 [door:WebDAV-dcdoor13-external@webdav-dcdoor13_httpsDomain:remove] ["usatlas1":6435:31152:unknown] [00002FEEA4A0E44E4533B842359CEF5CBA28,23621136071] [/pnfs/usatlas.bnl.gov/BNLT0D1/rucio/mc21_13p6TeV/80/27/RDO.32721912._006682.pool.root.1] <Unknown> 0 0 {0:""}', 'error': {'message': '', 'code': '0'}, 'agent': {'ephemeral_id': 'ffb48cf8-8532-4118-a8be-779f5abbe684', 'name': 'dcache-billing', 'id': 'c11208ac-ba21-49c0-bfca-3f45ab204beb', 'type': 'filebeat', 'version': '8.6.0'}, 'input': {'type': 'filestream'}, 'fileset': {'name': 'log'}, 'tags': ['dcache-billing', 'dcache', 'billing', 'logs', 'file-based', 'beats_input_codec_plain_applied'], '@timestamp': '2023-04-26T03:06:21.000Z'}}

Second Row JSON

- '{\'_index\': \'dcache-billing-usatlas-000055\', \'_type\': \'_doc\', \'_id\': \'ufiGu4cBOBpu8Q1fxLdE\', \'_score\': 1.0, \'_ignored\': [\'message.keyword\'], \'_source\': {\'@version\': \'1\', \'log_type\': \'dcache-billing\', \'ecs\': {\'version\': \'8.0.0\'}, \'event\': {\'action\': \'request\', \'dataset\': \'dcache.billing\', \'provider\': \'dcache\', \'created\': \'2023-04-26T03:06:22.891Z\', \'category\': \'network\', \'kind\': \'event\', \'module\': \'dcache\'}, \'dcache\': {\'billing\': {\'client\': {\'address\': \'130.199.156.100\', \'user\': {\'dn\': \'/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=663551/CN=Robot: ATLAS Pilot2\', \'id\': \'6439\', \'group\': {\'id\': \'69907\'}}}, \'action\': \'request\', \'storage\': {\'class\': \'GROUPDISK:LOCAL@osm\'}, \'queued\': {\'time\': 0}, \'p2p\': False, \'transaction\': {\'time\': 142}, \'error\': {\'code\': \'0\', \'message\': \'\'}, \'ts\': \'2023.04.25 23:06:22\', \'cell\': {\'name\': \'Xrootd-dcdoor02-internal@xrootd-dcdoor02Domain\', \'type\': \'door\'}, \'file\': {\'size\': 2938178050}, \'pnfsid\': \'0000780C3C62D463485DB15FDF537F24E8C5\', \'protocol\': {\'path\': \'/pnfs/usatlas.bnl.gov/localgroupdisk/rucio/data17_13tev/3a/75/daod_phys.30498649._000215.pool.root.1\'}}}, \'client\': {\'address\': \'130.199.156.100\', \'user\': {\'dn\': \'/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=663551/CN=Robot: ATLAS Pilot2\', \'id\': \'6439\', \'group\': {\'id\': \'69907\'}}}, \'host\': {\'name\': \'dcache-billing\'}, \'log\': {\'file\': {\'path\': \'/var/lib/dcache/billing/2023/04/billing-2023.04.25\'}, \'offset\': 2380249344}, \'service\': {\'name\': \'billing\', \'type\': \'dcache\', \'version\': \'7.2.16\'}, \'labels\': {\'env\': \'prod\', \'type\': \'logs\', \'tier1\': {\'category\': \'storage\'}}, \'type\': \'dcache-billing-logs\', \'message\': \'04.25 23:06:22 [door:Xrootd-dcdoor02-internal@xrootd-dcdoor02Domain:request] ["/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=663551/CN=Robot: ATLAS Pilot2":6439:69907:130.199.156.100] [0000780C3C62D463485DB15FDF537F24E8C5,2938178050] [/pnfs/usatlas.bnl.gov/LOCALGROUPDISK/rucio/data17_13TeV/3a/75/DAOD_PHYS.30498649._000215.pool.root.1] GROUPDISK:LOCAL@osm 142 0 {0:""}\', \'error\': {\'message\': \'\', \'code\': \'0\'}, \'agent\': {\'ephemeral_id\': \'ffb48cf8-8532-4118-a8be-779f5abbe684\', \'name\': \'dcache-billing\', \'id\': \'c11208ac-ba21-49c0-bfca-3f45ab204beb\', \'type\': \'filebeat\', \'version\': \'8.6.0\'}, \'input\': {\'type\': \'filestream\'}, \'tags\': [\'dcache-billing\', \'dcache\', \'billing\', \'logs\', \'file-based\', \'beats_input_codec_plain_applied\'], \'fileset\': {\'name\': \'log\'}, \'@timestamp\': \'2023-04-26T03:06:22.000Z\'}}'

Second Row Second format JSON

- {"_index": "dcache-billing-usatlas-000055", "_type": "_doc", "_id": "BviGu4cBOBpu8Q1fvLNK", "_score": 1.0, "_ignored": ["message.keyword"], "_source": {"@version": "1", "log_type": "dcache-billing", "ecs": {"version": "8.0.0"}, "event": {"action": "transfer", "id": "door:Xrootd-dcdoor05-internal@xrootd-dcdoor05Domain:AAX6NIR6JNA:1682478378942000", "dataset": "dcache.billing", "provider": "dcache", "created": "2023-04-26T03:06:20.842Z", "category": "network", "kind": "event", "module": "dcache"}, "dcache": {"billing": {"initiator": {"name": "door:Xrootd-dcdoor05-internal@xrootd-dcdoor05Domain:AAX6NIR6JNA:1682478378942000"}, "action": "transfer", "storage": {"class": "bnlt0d1:BNLT0D1@osm"}, "p2p": false, "connection": {"time": 71}, "error": {"code": "0", "message": ""}, "ts": "2023.04.25 23:06:19", "cell": {"name": "dc250_4", "type": "pool"}, "file": {"size": 403653271}, "pnfsid": "0000DB7AFA8F61A746EBB79109A93AA6DA68", "isnew": false, "protocol": {"port": 51944, "name": "Xrootd-5.0", "address": "130.199.159.20", "path": "/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/12/f6/evnt.18792047._000471.pool.root.1"}, "transfer": {"size": 10183}}}, "host": {"name": "dcache-billing"}, "log": {"file": {"path": "/var/lib/dcache/billing/2023/04/billing-2023.04.25"}, "offset": 2380093839}, "service": {"name": "billing", "type": "dcache", "version": "7.2.16"}, "labels": {"env": "prod", "type": "logs", "tier1": {"category": "storage"}}, "type": "dcache-billing-logs", "message": "04.25 23:06:19 [pool:dc250_4:transfer] [0000DB7AFA8F61A746EBB79109A93AA6DA68,403653271] [/pnfs/usatlas.bnl.gov/BNLT0D1/rucio/mc16_13TeV/12/f6/EVNT.18792047._000471.pool.root.1] bnlt0d1:BNLT0D1@osm 10183 71 false {Xrootd-5.0:130.199.159.20:51944} [door:Xrootd-dcdoor05-internal@xrootd-dcdoor05Domain:AAX6NIR6JNA:1682478378942000] {0:\"\"}", "error": {"message": "", "code": "0"}, "destination": {"address": "130.199.159.20"}, "agent": {"ephemeral_id": "ffb48cf8-8532-4118-a8be-779f5abbe684", "name": "dcache-billing", "id": "c11208ac-ba21-49c0-bfca-3f45ab204beb", "type": "filebeat", "version": "8.6.0"}, "input": {"type": "filestream"}, "tags": ["dcache-billing", "dcache", "billing", "logs", "file-based", "beats_input_codec_plain_applied"], "fileset": {"name": "log"}, "@timestamp": "2023-04-26T03:06:19.000Z"}}

Third Row JSON

- {"_index": "dcache-billing-usatlas-000055", "_type": "_doc", "_id": "n_iGu4cBOBpu8Q1f_OpK", "_score": 1.0, "_source": {"@version": "1", "log_type": "dcache-billing", "ecs": {"version": "8.0.0"}, "event": {"action": "request", "dataset": "dcache.billing", "provider": "dcache", "created": "2023-04-26T03:06:37.163Z", "category": "network", "kind": "event", "module": "dcache"}, "dcache": {"billing": {"client": {"address": "unknown", "user": {"dn": "usatlas1", "id": "6435", "group": {"id": "31152"}}}, "action": "request", "ts": "2023.04.25 23:04:55", "cell": {"name": "RemoteTransferManager@srm-dcsrm03Domain", "type": "door"}, "storage": {"class": "bnlt0d1:BNLT0D1@osm"}, "queued": {"time": 32}, "error": {"code": "0", "message": ""}, "p2p": false, "protocol": {"path": "/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/05/79/daod_topq1.33192563._000957.pool.root.1"}, "transaction": {"time": 101144}}}, "client": {"address": "unknown", "user": {"dn": "usatlas1", "id": "6435", "group": {"id": "31152"}}}, "host": {"name": "dcache-billing"}, "log": {"file": {"path": "/var/lib/dcache/billing/2023/04/billing-2023.04.25"}, "offset": 2380841279}, "service": {"name": "billing", "type": "dcache", "version": "7.2.16"}, "labels": {"env": "prod", "type": "logs", "tier1": {"category": "storage"}}, "type": "dcache-billing-logs", "message": "04.25 23:04:55 [door:RemoteTransferManager@srm-dcsrm03Domain:request] [\"usatlas1\":6435:31152:unknown] [,0] [/pnfs/usatlas.bnl.gov/BNLT0D1/rucio/mc16_13TeV/05/79/DAOD_TOPQ1.33192563._000957.pool.root.1] bnlt0d1:BNLT0D1@osm 101144 32 {0:\"\"}", "error": {"message": "", "code": "0"}, "agent": {"ephemeral_id": "ffb48cf8-8532-4118-a8be-779f5abbe684", "name": "dcache-billing", "type": "filebeat", "id": "c11208ac-ba21-49c0-bfca-3f45ab204beb", "version": "8.6.0"}, "input": {"type": "filestream"}, "tags": ["dcache-billing", "dcache", "billing", "logs", "file-based", "beats_input_codec_plain_applied"], "fileset": {"name": "log"}, "@timestamp": "2023-04-26T03:04:55.000Z"}}

First problem and solution

- Problem:
  - JSON objects formatted incorrectly
  - Must include double quotes (“”) not single (‘’)
  - JSON object is currently a string object and not a JSON
- Solution:
  - Import ast library to convert the string into a dictionary
  - Use the JSON library to convert the dictionary into a JSON object

Code

```
        ]
    )

    print("Starting to parse data...")

    for i, row in enumerate(reader):
        print("Parsing row: " + str(i))
        # current_row = i
        # if i == 0:
        #     continue
        if i == 100:  # Use this to limit the number of lines to read.
            break
        else:
            for i in range(len(row)):
                row[i] = json.dumps(ast.literal_eval(row[i]))  # Convert string to dict
                log_entry = row[i]  # Assuming the log entry is in the ith column
                extracted_info = extract_info(log_entry)
                writer.writerow(extracted_info)

print("Done!")
```

Second problem and solution

- Problem:
  - Multiple JSON object formats
  - Not all objects have requested information
- Solution:
  - Use try catch blocks to try each known format
  - If not found put, “unknown”

Code

```
# Function to extract the desired information from the log entry
def extract_info(log_entry):
    log_data = json.loads(log_entry)  # Parse the log entry as JSON
    try:
        file_id = log_data["_source"]["dcache"]["billing"]["pnfsid"]
    except:
        file_id = "unknown"
        # print("File ID not found")
    action = log_data["_source"]["event"]["action"]
    try:
        file_size = log_data["_source"]["dcache"]["billing"]["file"]["size"]
    except:
        file_size = "unknown"
        # print("File size not found")
    file_type = log_data["_source"]["dcache"]["billing"]["storage"]["class"]
    try:
        user_id = log_data["_source"]["dcache"]["billing"]["client"]["user"]["id"]
    except:
        try:
            user_id = log_data["_source"]["dcache"]["billing"]["initiator"]["name"]
        except:
            user_id = "unknown"
            # print("User ID not found")
    timestamp = log_data["_source"]["dcache"]["billing"]["ts"]
    file_path = log_data["_source"]["dcache"]["billing"]["protocol"]["path"]
    return [file_id, action, file_size, file_type, user_id, timestamp, file_path]
```

Third problem and solution

- Problem:
  - Python Script hard coded
  - Needs to allow for command line calls with optional input file
- Solution:
  - Import sys library for command line parameters

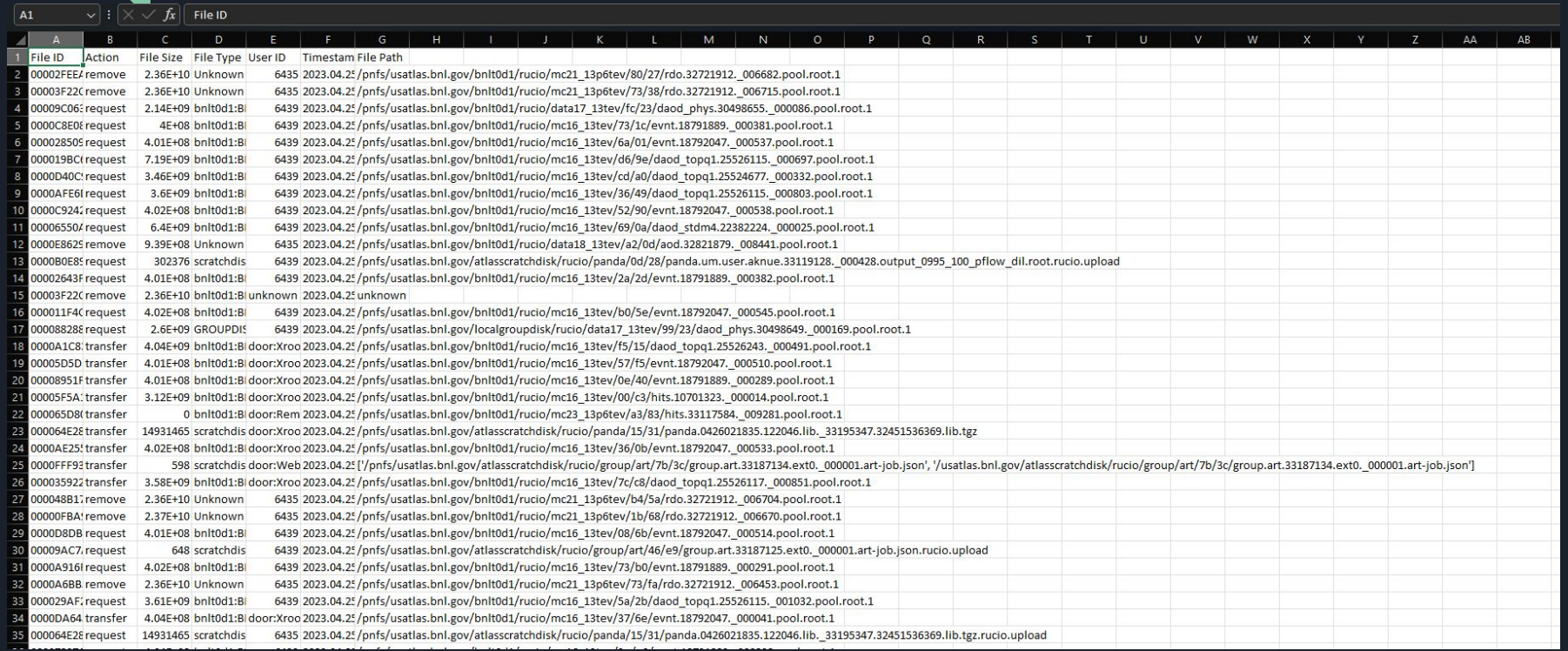
Code

```
if len(sys.argv) < 2:
    print("Please provide the input filename as a command-line argument.")
    sys.exit()

input_filename = sys.argv[1]

with open(input_filename, "r") as infile, open(
    "output.csv", "w", newline=""
```

Final Output



A1 | File ID

	A	B	C	D	E	F	G
1	File ID	Action	File Size	File Type	User ID	Timestam	File Path
2	00002FEE4	remove	2.36E+10	Unknown	6435	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc21_13p6tev/80/27/rdo.32721912._006682.pool.root.1
3	00003F22(	remove	2.36E+10	Unknown	6435	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc21_13p6tev/73/38/rdo.32721912._006715.pool.root.1
4	00009C06:	request	2.14E+09	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/data17_13tev/fc/23/daod_phys.30498655._000086.pool.root.1
5	0000C8E08	request	4E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/73/1c/evnt.18791889._000381.pool.root.1
6	000028505	request	4.01E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/6a/01/evnt.18792047._000537.pool.root.1
7	000019BC(	request	7.19E+09	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/d6/9e/daod_topq1.25526115._000697.pool.root.1
8	0000D40C(	request	3.46E+09	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/cd/a0/daod_topq1.25524677._000332.pool.root.1
9	0000AFE6(	request	3.6E+09	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/36/49/daod_topq1.25526115._000803.pool.root.1
10	0000C924:	request	4.02E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/52/90/evnt.18792047._000538.pool.root.1
11	000065504	request	6.4E+09	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/69/0a/daod_stdm4.22382224._000025.pool.root.1
12	0000E8625	remove	9.39E+08	Unknown	6435	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/data18_13tev/a2/0d/aod.32821879._008441.pool.root.1
13	0000B0E85	request	302376	scratchdis	6439	2023.04.2	/pnfs/usatlas.bnl.gov/atlasscratchdisk/rucio/panda/0d/28/panda.um.user.aknue.33119128._000428.output_0995_100_pflow_dil.root.rucio.upload
14	00002643F	request	4.01E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/2a/2d/evnt.18791889._000382.pool.root.1
15	00003F22(	remove	2.36E+10	bnlt0d1:BI	unknown	2023.04.2	unknown
16	000011F4(	request	4.02E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/b0/5e/evnt.18792047._000545.pool.root.1
17	000088288	request	2.6E+09	GROUPDIS	6439	2023.04.2	/pnfs/usatlas.bnl.gov/localgroupdisk/rucio/data17_13tev/99/23/daod_phys.30498649._000169.pool.root.1
18	0000A1C8:	transfer	4.04E+09	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/f5/15/daod_topq1.25526243._000491.pool.root.1
19	00005D5D	transfer	4.01E+08	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/57/f5/evnt.18792047._000510.pool.root.1
20	00008951F	transfer	4.01E+08	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/0e/40/evnt.18791889._000289.pool.root.1
21	00005F5A:	transfer	3.12E+09	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/00/c3/hits.10701323._000014.pool.root.1
22	000065D8(	transfer	0	bnlt0d1:BI	door:Rem	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc23_13p6tev/a3/83/hits.33117584._009281.pool.root.1
23	000064E28	transfer	14931465	scratchdis	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/atlasscratchdisk/rucio/panda/15/31/panda.0426021835.122046.lib._33195347.32451536369.lib.tgz
24	0000AE25!	transfer	4.02E+08	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/36/0b/evnt.18792047._000533.pool.root.1
25	0000FFF93	transfer	598	scratchdis	door:Web	2023.04.2	['/pnfs/usatlas.bnl.gov/atlasscratchdisk/rucio/group/art/7b/3c/group.art.33187134.ext0._000001.art-job.json', '/usatlas.bnl.gov/atlasscratchdisk/rucio/group/art/7b/3c/group.art.33187134.ext0._000001.art-job.json']
26	000035922	transfer	3.58E+09	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/7c/c8/daod_topq1.25526117._000851.pool.root.1
27	000048B1	remove	2.36E+10	Unknown	6435	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc21_13p6tev/b4/5a/rdo.32721912._006704.pool.root.1
28	00000FBA'	remove	2.37E+10	Unknown	6435	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc21_13p6tev/1b/68/rdo.32721912._006670.pool.root.1
29	0000D8DB	request	4.01E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/08/6b/evnt.18792047._000514.pool.root.1
30	00009AC7(	request	648	scratchdis	6439	2023.04.2	/pnfs/usatlas.bnl.gov/atlasscratchdisk/rucio/group/art/46/e9/group.art.33187125.ext0._000001.art-job.json.rucio.upload
31	0000A916I	request	4.02E+08	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/73/b0/evnt.18791889._000291.pool.root.1
32	0000A6BB	remove	2.36E+10	Unknown	6435	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc21_13p6tev/73/fa/rdo.32721912._006453.pool.root.1
33	000029AF:	request	3.61E+09	bnlt0d1:BI	6439	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/5a/2b/daod_topq1.25526115._001032.pool.root.1
34	0000DA64	transfer	4.04E+08	bnlt0d1:BI	door:Xroo	2023.04.2	/pnfs/usatlas.bnl.gov/bnlt0d1/rucio/mc16_13tev/37/6e/evnt.18792047._000041.pool.root.1
35	000064E28	request	14931465	scratchdis	6435	2023.04.2	/pnfs/usatlas.bnl.gov/atlasscratchdisk/rucio/panda/15/31/panda.0426021835.122046.lib._33195347.32451536369.lib.tgz.rucio.upload