

# Data Engineering HW2 - Logs Analytics

## Introduction

This report explores the process of processing and analyzing log files using a combination of powerful Big Data technologies. The main technologies involved in this process are Kafka (producer and consumer), Spark Streaming, and HDFS. The primary goal is to establish a flow that starts with a Kafka producer, followed by a Kafka consumer that performs transformations using Spark. The transformed data is then converted into the Parquet file format and stored in HDFS.

## Aim

1. Log Analytics: Log analytics is used for various purposes like Monitoring and Troubleshooting, Security and Threat Detection, etc. Due to the need for Real-time Analysis and Monitoring and Volume and Velocity of logs, we need to use tools like Kafka and Spark Streaming for efficient processing.
2. Data Processing
  - a. Create a system for handling and analyzing log files by using technologies like Kafka, Spark Streaming, Parquet files, and HDFS.
  - b. Build a component in Kafka that will gather log file data and feed it into the system.
  - c. Utilize Spark to transform the log data and perform exploratory data analysis (EDA) on it.
  - d. Convert the analysis results into Parquet file format.
  - e. Store the Parquet files in HDFS, which is a reliable and scalable data storage solution.

## Tools Used

1. Kafka: Kafka is a distributed streaming platform that enables real-time data ingestion and processing. It provides a reliable and scalable solution for handling large volumes of data streams from multiple sources. With Kafka, data can be seamlessly published to topics and consumed by multiple consumers, facilitating efficient data processing and analysis.
2. Spark Streaming: Spark Streaming is a powerful framework that allows processing and analyzing real-time streaming data. It helps to handle data streams continuously, making it suitable for applications that require immediate insights and responses. With Spark Streaming, data can be ingested in small batches, enabling near real-time analysis and facilitating the extraction of valuable insights from streaming data sources.

3. HDFS: HDFS (Hadoop Distributed File System) is a distributed file system designed to store and manage large volumes of data across a cluster of machines. It provides fault tolerance, high scalability, and efficient data storage capabilities
4. Parquet Files: Parquet files are a columnar storage format optimized for Big Data processing. They offer efficient compression, encoding, and retrieval of data, making them ideal for handling large datasets. Parquet files enable faster query execution and reduce storage costs, making them a preferred choice for streaming data processing and analysis in Big Data environments.

## **Steps Followed**

To process and analyze log files, we follow these steps:

Step 1: Injecting Data into Kafka Producer - We start by feeding the log file data into a Kafka producer. This involves injecting the data line by line into the producer.

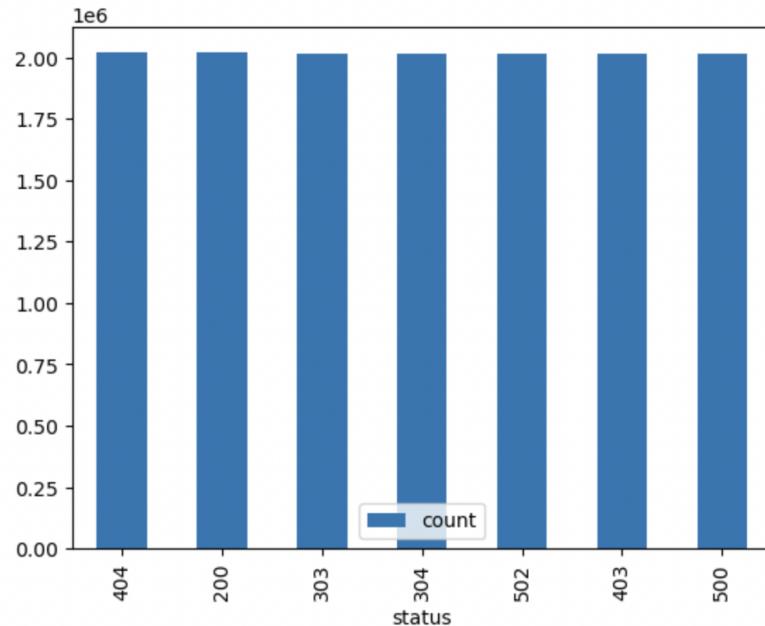
Step 2: Kafka Consumer and Spark Transformation - Next, we set up a Kafka consumer that subscribes to the same topic as the producer. The consumer regularly checks for new data. When new data is received, we use Spark Streaming to perform transformations and conduct exploratory data analysis on it.

Step 3: Creating Parquet Files - The results obtained from the Spark analysis, are converted into the Parquet file format. Parquet is chosen because it provides efficient compression and encoding, which are important for processing large amounts of data. We store the Parquet files in HDFS as it offers fault tolerance and high throughput.

# Data Analysis

In this section, we analyze the transformed data obtained from Spark.

## Count of Different Status



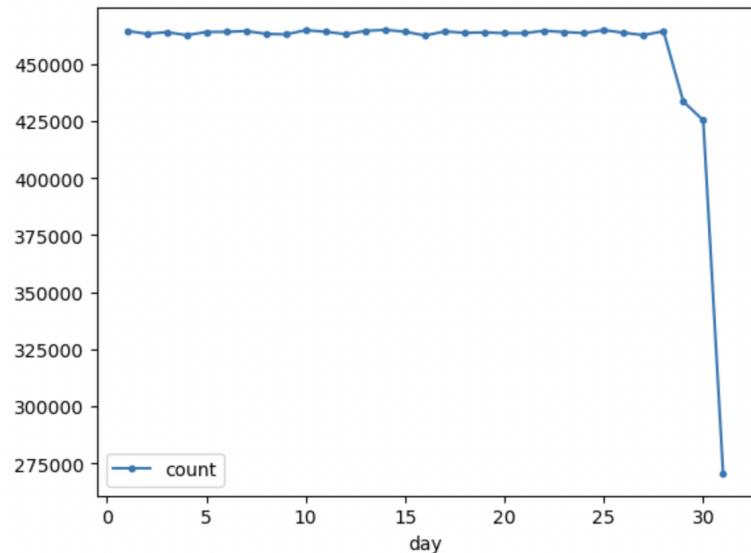
## Top 20 Frequent Endpoints

uri	count
/Archives/edgar/data/0001260125/000114420419015071/tv516461_10k.htm	94919
/Archives/edgar/data/0001675033/000156459021054728/gecc-10q_20210930.htm	94783
/Archives/edgar/data/0001179929/000117992923000043/moh-20230320_g29.jpg	94671
/Archives/edgar/data/0001866295/000110465921080403/tm2117388d5_ex3-32g008.jpg	94659
/Archives/edgar/data/0001468457/000093247112005362/0000932471-12-005362.txt	94650
/Archives/edgar/data/0001267332/000117494721000300/form10k-25692_indexpl20031.htm	94619
/Archives/edgar/data/0001326380/000132638023000019/MetaLinks.json	94616
/Archives/edgar/data/0001608293/000119312514225607/g732497g91g62.jpg	94597
/Archives/edgar/data/0000352825/000035282519000012/0000352825-19-000012.txt	94575
/Archives/edgar/data/0001491487/000114420415051928/0001144204-15-051928.txt	94574
/Archives/edgar/data/0001763499/000121390021017164/s131140_10k.htm	94567
/Archives/edgar/data/19411/0001104659-12-037849.txt	94557
/Archives/edgar/data/0000205007/000114554923005577/primary_doc.xml	94556
/Archives/edgar/data/0000004281/000119312517062657/R94.htm	94548
/Archives/edgar/data/0001795772/000121390021017160/s131128_10k.htm	94536
/Archives/edgar/data/0001260125/000114420419015093/tv516458_10k.htm	94529
/Archives/edgar/data/0001002638/000100263815000016/0001002638-15-000016-index.htm	94526
/Archives/edgar/data/0001326380/000132638023000019/gme-20230128.htm	94520
/Archives/edgar/data/0000205007/000114554923005577/0001145549-23-005577.txt	94496
/Archives/edgar/data/0001056673/000094930398000131/0000949303-98-000131.txt	94482

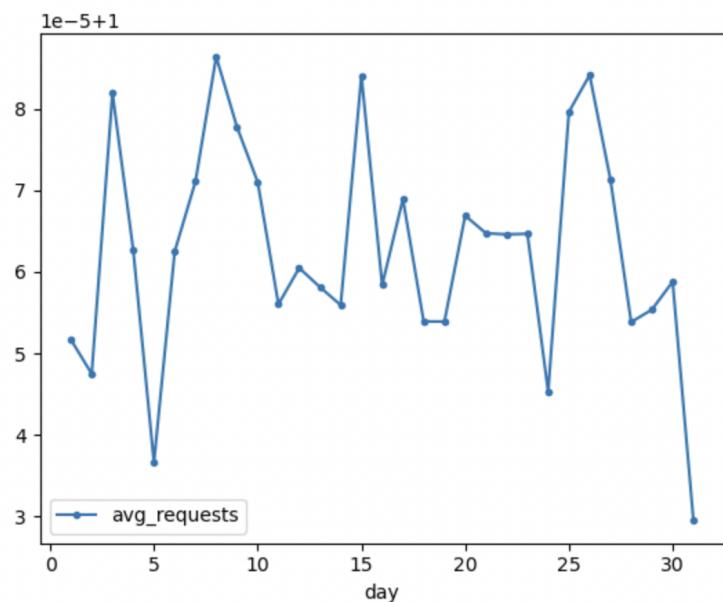
## Top 10 error endpoints

uri	count
/Archives/edgar/data/0001675033/000156459021054728/gecc-10q_20210930.htm	81302
/Archives/edgar/data/0001468457/000093247112005362/0000932471-12-005362.txt	81246
/Archives/edgar/data/19411/0001104659-12-037849.txt	81164
/Archives/edgar/data/0001260125/000114420419015093/tv516458_10k.htm	81134
/Archives/edgar/data/0001866295/000110465921080403/tm2117388d5_ex3-32g008.jpg	81132
/Archives/edgar/data/0001491487/000114420415051928/0001144204-15-051928.txt	81129
/Archives/edgar/data/0001326380/000132638023000019/gme-20230128.htm	81125
/Archives/edgar/data/0001179929/000117992923000043/moh-20230320_g29.jpg	81113
/Archives/edgar/data/0001260125/000114420419015071/tv516461_10k.htm	81079
/Archives/edgar/data/0000081391/000093247112005362/0000932471-12-005362.txt	81073

### ***Count of Unique Daily Hosts***



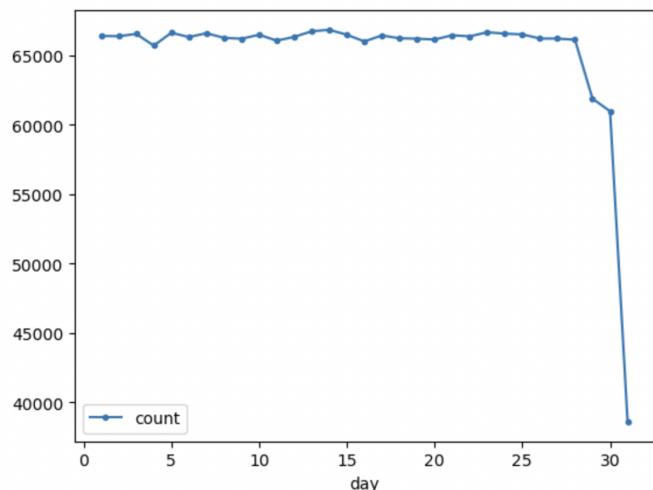
### ***Average Daily Unique Hosts***



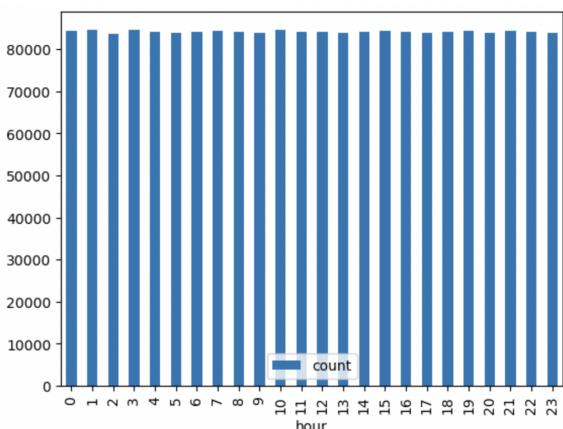
## Top 20 Endpoints 404

uri	count
/Archives/edgar/data/921112/000089542110000396/0000895421-10-000396.txt	13697
/Archives/edgar/data/1935285/0001445546-23-001552-index.htm	13680
/Archives/edgar/data/0001260125/000114420419015062/0001144204-19-015062.txt	13674
/Archives/edgar/data/1027443/000091293802000245/0000912938-02-000245.txt	13664
/Archives/edgar/data/0001608293/000119312514225607/g732497g91g62.jpg	13661
/Archives/edgar/data/0001801390/000121390021017183/s131106_10k.htm	13653
/Archives/edgar/data/0001704304/000092963819000386/0000929638-19-000386.txt	13641
/Archives/edgar/data/19411/0001104659-12-037849.txt	13627
/Archives/edgar/data/0001795772/000121390021017160/s131128_10k.htm	13627
/Archives/edgar/data/0000008670/000120677422002461/adp4094281-def14a.htm	13626
/Archives/edgar/data/0001002638/000100263815000016/0001002638-15-000016-index.htm	13623
/Archives/edgar/data/0001798682/000121390021017167/s131143_10k.htm	13614
/Archives/edgar/data/0001221326/99999999512000938/999999995-12-000938.txt	13609
/Archives/edgar/data/0001056673/000094930398000131/0000949303-98-000131.txt	13608
/Archives/edgar/data/0001671697/000092963819000390/form10k.htm	13608
/Archives/edgar/data/0000004281/000119312517062657/R98.htm	13602
/Archives/edgar/data/0000352825/000035282519000012/fstr-20181231.htm	13601
/Archives/edgar/data/0001675033/000156459021054728/gecc-10q_20210930.htm	13599
/Archives/edgar/data/925548/000092153005000094/0000921530-05-000094.txt	13597
/Archives/edgar/data/0001326380/000132638023000019/gme-20230128_g1.jpg	13592

## Daily Count of 404 errors



## Hourly Count of 404 errors



## **Conclusion**

To sum it up, this report outlines a method for handling and examining log files by utilizing a mix of Big Data technologies. By making use of tools like Kafka, Spark Streaming, and the HDFS File System, the project allows for effective processing, transformation, and analysis of data. The process involves moving log files to a Kafka producer, then using a Kafka consumer with Spark transformations to extract useful information, creating Parquet files for storage, and finally storing them in the HDFS. The insights gained from analyzing the data can offer valuable details for resolving issues, enhancing performance, and making informed decisions in the future.

## **Future Work**

In the future, we can do clustering on various features like status code, endpoint, and timestamp to offer a quick view of the data and identify any anomalies.

# Appendix

## Snapshot of Folders for each EDA

```
HW2 -- zsh -- 205x51
~ - python3.10 /opt/anaconda3/envs/DistributedComputing/bin/jupyter-lab - java ... ~/Documents/Studies/USF/Summer/Data_Engineering/HW2 -- zsh
(DistributedComputing) manavmiddha@Manavs-MacBook-Air HW2 % hdfs dfs -ls /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp
2023-06-30 23:08:26,957 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 9 items
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 22:03 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/context_size_count3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 20:37 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/daily_host_cnt3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 20:37 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpoints3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:52 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:52 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_day_3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:54 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/host3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 20:37 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/error_endpoints3
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:50 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3
```

## Files created in Status Count EDA

```
HW2 -- zsh -- 205x51
~ - python3.10 /opt/anaconda3/envs/DistributedComputing/bin/jupyter-lab - java ... ~/Documents/Studies/USF/Summer/Data_Engineering/HW2 -- zsh
(DistributedComputing) manavmiddha@Manavs-MacBook-Air HW2 % hdfs dfs -ls /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3
2023-06-30 23:09:33,951 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 287 items
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:50 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/_spark_metadata
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:50 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/commits
drwxr-xr-x  - manavmiddha supergroup          45 2023-06-30 20:37 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/metadata
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:50 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/offsets
drwxr-xr-x  - manavmiddha supergroup          8355 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00000-313c0ca2-f128-4cc9-92fe-25a909b48b61-c000.sn
drwxr-xr-x  - manavmiddha supergroup          543 2023-06-30 21:14 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00000-d3774b4c-3693-421a-9032-f6c8ea571885-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7391 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00001-aeef19bed-6dfb-4360-ad18-6326e04049f6-c000.sn
drwxr-xr-x  - manavmiddha supergroup          8117 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00002-41571ad6-46b6-43c3-9c9e-53f6129f5db0-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7947 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00003-e2b278af-f357-419a-a1dc-089f647e09c7-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7918 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00004-567342cb-f29b-4f42-b25f-3b5f818f787f-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7890 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00005-33c4fb13-faea-447a-98c8-fa4cfb31620e-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7781 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00006-588429be-6477-4613-b126-626047f4b1b5-c000.sn
drwxr-xr-x  - manavmiddha supergroup          8529 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00007-cf86de9e-7d35-4196-b2a5-492563f6c559-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7771 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00008-4f33683d-baa2-4c4c-acaa-84ed914f9c398-c000.sn
drwxr-xr-x  - manavmiddha supergroup          8121 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00009-2e6c3f40-38b8-42f7-a79f-9b5b32f2ecdf-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7967 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00010-93facc20-adae-4d97-94f7-135d3bb8f542-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7334 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00011-8c8b3933-d293-4ab9-a98a-29389cf73d52-c000.sn
drwxr-xr-x  - manavmiddha supergroup          7939 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/status_count3/part-00012-cd23fe78-2ebc-47cc-9dbb-24dc9a46b05e-c000.sn
drwxr-xr-x  - manavmiddha supergroup          86698 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-38a5b557-8475-48f9-9000-9697072d934a-c000.sna
drwxr-xr-x  - manavmiddha supergroup          552 2023-06-30 21:52 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-549236a4-08d8-400f-82db-4cc0d165d480d-c000.sna
drwxr-xr-x  - manavmiddha supergroup          552 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-a1fc2c444-d27d-4fc9-9b4b-6444f7112e69-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85993 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00001-d2e83336-7fa5-442b-8e03-69580c6f2b6f-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85937 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00002-df258765-d08f-4ea-81ff-7fbe62236c6-c000.sna
drwxr-xr-x  - manavmiddha supergroup          84415 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00003-d2e83336-7fa5-442b-8e03-69580c6f2b6f-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85712 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00004-cbd2ce5-1fea-aee8-8fc1-fcceaa582eb64-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85947 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00005-f952023f-4f4e-408c-8cda-064a98eb4de5-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85821 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00006-681e241b-ed06-4f3b-aaed-921e2f30372a-c000.sna
drwxr-xr-x  - manavmiddha supergroup          87476 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00007-574f7088-085f-418d-b0ef-edd034800fd9-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85819 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00008-c4fc4d6c-d9de-4c40-ad0b-dc220e8e68d-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85816 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00009-bf46af3-a099-4e8a-8cef-702039fabb7-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85418 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00010-7b4a23b3-1c64-4480-8760-573f5fd752a4-c000.sna
drwxr-xr-x  - manavmiddha supergroup          84535 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00011-baa3eb05-9aa-451a-a3ff-25537e46aab1-c000.sna
drwxr-xr-x  - manavmiddha supergroup          86426 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00012-3e065f21-47b8-48f2-8725-d018ee2d7a8f-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85804 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00013-e679122-e029-4be4-80e6-c13210f5ba01-c000.sna
```

## Endpoints with the most 404 errors

```
HW2 -- zsh -- 205x51
~ - python3.10 /opt/anaconda3/envs/DistributedComputing/bin/jupyter-lab - java ... ~/Documents/Studies/USF/Summer/Data_Engineering/HW2 -- zsh
(DistributedComputing) manavmiddha@Manavs-MacBook-Air HW2 % hdfs dfs -ls /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3
2023-06-30 23:15:13,078 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 209 items
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:53 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/_spark_metadata
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:53 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/commits
drwxr-xr-x  - manavmiddha supergroup          45 2023-06-30 20:37 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/metadata
drwxr-xr-x  - manavmiddha supergroup          0 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/offsets
drwxr-xr-x  - manavmiddha supergroup          552 2023-06-30 21:10 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-0ff1a4ba-bf18-4e22-8173-277ac4a59f0e-c000.sna
drwxr-xr-x  - manavmiddha supergroup          86698 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-38a5b557-8475-48f9-9000-9697072d934a-c000.sna
drwxr-xr-x  - manavmiddha supergroup          552 2023-06-30 21:52 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-549236a4-08d8-400f-82db-4cc0d165d480d-c000.sna
drwxr-xr-x  - manavmiddha supergroup          552 2023-06-30 21:49 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00000-a1fc2c444-d27d-4fc9-9b4b-6444f7112e69-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85993 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00001-d2e83336-7fa5-442b-8e03-69580c6f2b6f-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85937 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00002-df258765-d08f-4ea-81ff-7fbe62236c6-c000.sna
drwxr-xr-x  - manavmiddha supergroup          84415 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00003-d2e83336-7fa5-442b-8e03-69580c6f2b6f-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85712 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00004-cbd2ce5-1fea-aee8-8fc1-fcceaa582eb64-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85947 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00005-f952023f-4f4e-408c-8cda-064a98eb4de5-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85821 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00006-681e241b-ed06-4f3b-aaed-921e2f30372a-c000.sna
drwxr-xr-x  - manavmiddha supergroup          87476 2023-06-30 20:55 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00007-574f7088-085f-418d-b0ef-edd034800fd9-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85819 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00008-c4fc4d6c-d9de-4c40-ad0b-dc220e8e68d-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85816 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00009-bf46af3-a099-4e8a-8cef-702039fabb7-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85418 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00010-7b4a23b3-1c64-4480-8760-573f5fd752a4-c000.sna
drwxr-xr-x  - manavmiddha supergroup          84535 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00011-baa3eb05-9aa-451a-a3ff-25537e46aab1-c000.sna
drwxr-xr-x  - manavmiddha supergroup          86426 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00012-3e065f21-47b8-48f2-8725-d018ee2d7a8f-c000.sna
drwxr-xr-x  - manavmiddha supergroup          85804 2023-06-30 20:56 /Users/manavmiddha/Documents/Studies/USF/Summer/Data_Engineering/temp/endpts_404_3/part-00013-e679122-e029-4be4-80e6-c13210f5ba01-c000.sna
```