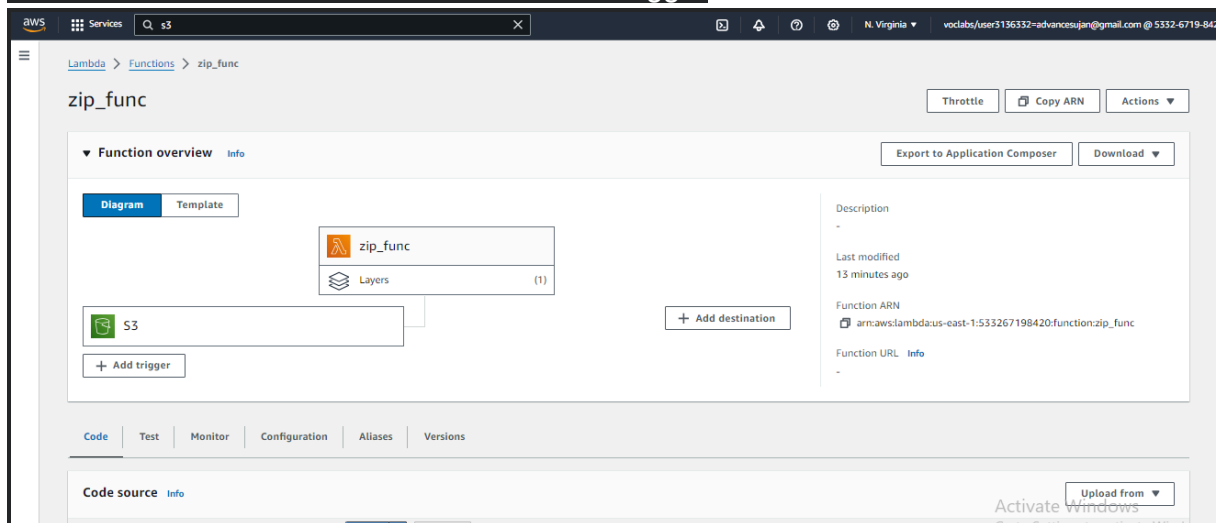


Task 2 : Python and Lambda

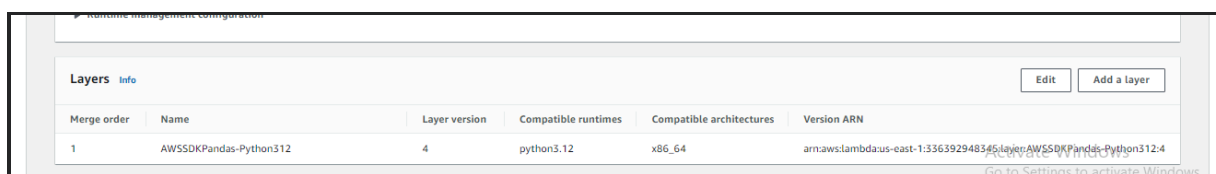
Work with RXNORM file,

1. Scrap the latest RXNORM file from NLM webpage
2. Download the latest RXNORM file with api_key
3. Create a log file for the downloaded file
4. Add header into each rff from RXNORM.xlsx
5. Add CODE_SET & VERSION_MONTH column with default values RxNorm and version month from downloaded filename
6. Convert dates into YYYY-MM-DD7.
7. Save files as txt delimited by comma(,)
8. Validate row_count between original and converted files

1. Create the lambda function with s3 bucket trigger



2. Add necessary configuration like memory, timeout and layers for lambda



3. Upload zip files and header files to the bucket

bucketzipp info

Objects (2) Info

Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

Name	Type	Last modified	Size	Storage class
RxNorm_full_02052024.zip	zip	March 11, 2024, 10:53:40 (UTC+05:45)	238.5 MB	Standard
RxNorm_Header.xlsx	xlsx	March 11, 2024, 10:53:43 (UTC+05:45)	24.7 KB	Standard

4. Write lambda function

```

1 import boto3
2 import zipfile
3 import pandas as pd
4 import io
5 import os
6 from datetime import datetime
7
8 s3 = boto3.client('s3')
9
10 def lambda_handler(event, context):
11     source_bucket = "bucketzipp"
12     source_key = "RxNorm_full_02052024.zip"
13     excel_key = "RxNorm_Header.xlsx"
14
15     try:
16         # Extract the data part from the source key
17         date_str = source_key.split('.')[0].split('.')[0] # Extract the date part without the extension
18         date_obj = datetime.strptime(date_str, '%m%d%Y')
19         date_formatted = date_obj.strftime('%Y-%m-%d')
20     except ValueError as e:
21         return {
22             'statusCode': 400,
23             'body': f'Error: {str(e)}; Source Key: {source_key}'
24         }
25
26     # Define the directory where the extracted rrf folder will be stored in S3
27     extracted_folder_key = f'extracted_rrf/'
28
29     # Retrieve the zip file object directly from S3
30     zip_obj = s3.get_object(Bucket=source_bucket, Key=source_key)['Body'].read()
31
32     print(f"Unzipping the zip file {source_key}.....")
33
34     # Read the zip file object using ZipFile from the io module
35     with zipfile.ZipFile(io.BytesIO(zip_obj)) as zip_ref:
36         for file_info in zip_ref.infolist():
37             if file_info.filename.startswith('rrf/'):

```

5. bucket after lambda function is triggered

Amazon S3 > Buckets > bucketzipp

bucketzipp info

Objects (4) Info

Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

Name	Type	Last modified	Size	Storage class
extracted_rrf/	Folder	-	-	-
result/	Folder	-	-	-
RxNorm_full_02052024.zip	zip	March 11, 2024, 11:49:34 (UTC+05:45)	238.5 MB	Standard
RxNorm_Header.xlsx	xlsx	March 11, 2024, 11:49:37 (UTC+05:45)	24.7 KB	Standard

Objects (18) Info

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<

1

>

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	RXNCONSO.csv	csv	March 11, 2024, 20:12:27 (UTC+05:45)	138.5 MB	Standard
<input type="checkbox"/>	RXNCONSO.txt	txt	March 11, 2024, 21:16:23 (UTC+05:45)	138.5 MB	Standard
<input type="checkbox"/>	RXNCUL.csv	csv	March 11, 2024, 20:12:28 (UTC+05:45)	2.1 MB	Standard
<input type="checkbox"/>	RXNCUL.txt	txt	March 11, 2024, 21:16:24 (UTC+05:45)	2.1 MB	Standard
<input type="checkbox"/>	RXNCUICHANGES.csv	csv	March 11, 2024, 20:12:28 (UTC+05:45)	17.9 KB	Standard
<input type="checkbox"/>	RXNCUICHANGES.txt	txt	March 11, 2024, 21:16:24 (UTC+05:45)	17.9 KB	Standard
<input type="checkbox"/>	RXNDOC.csv	csv	March 11, 2024, 20:12:28 (UTC+05:45)	271.7 KB	Standard
<input type="checkbox"/>	RXNDOC.txt	txt	March 11, 2024, 21:16:24 (UTC+05:45)	271.7 KB	Standard
<input type="checkbox"/>	RXNREL.csv	csv	March 11, 2024, 20:13:30 (UTC+05:45)	645.2 MB	Standard
<input type="checkbox"/>	RXNREL.txt	txt	March 11, 2024, 21:17:24 (UTC+05:45)	645.2 MB	Standard
<input type="checkbox"/>	RXNSAB.csv	csv	March 11, 2024, 20:13:37 (UTC+05:45)	10.3 KB	Standard
<input type="checkbox"/>	RXNSAB.txt	txt	March 11, 2024, 21:17:30 (UTC+05:45)	10.3 KB	Standard
<input type="checkbox"/>	RXNSAT.csv	csv	March 11, 2024, 20:14:22 (UTC+05:45)	622.4 MB	Standard
<input type="checkbox"/>	RXNSAT.txt	txt	March 11, 2024, 21:18:15 (UTC+05:45)	622.4 MB	Standard
<input type="checkbox"/>	RXNSTY.csv	csv	March 11, 2024, 20:14:31 (UTC+05:45)	26.1 MB	Standard
<input type="checkbox"/>	RXNSTY.txt	txt	March 11, 2024, 21:18:24 (UTC+05:45)	26.1 MB	Standard

Activate Windows

Go to Settings to activate Windows.

```
1 /* To create reference point for writing SQL queries, you can display the first 5 records of input data by running the following SQL query: SELECT * FROM s3object s LIMIT 5 */
2 SELECT * FROM s3object s LIMIT 5
```

SQLLn 1, Col 1Errors: 0Warnings: 0

Query results

Query results are not available after you choose **Close** or navigate away. Choose **Download results** to download a copy of the following query results.

[Download results](#)

Status

Successfully returned 5 records in 2101 ms
Bytes returned: 804 B

RawFormatted

```
1 RXAUT,AUS,STR,ARCHIVE_TIMESTAMP,CREATED_TIMESTAMP,UPDATED_TIMESTAMP,CODE,IS_BRAND,LAT,LAST_RELEASED,SAUI,V5AB,RXCUI,SAB,TTY,MERGED_TO,RXCUI,Code set,Version Month
2 947,A18335755,Mesna,2020-04-27,03/10/2005,02:03:47 PM,04/27/2020,09:04:18 PM,44,,ENG,06-APR-20,,RXNORM_15AB_200406F,44,RXNORM_IN,44,RxNorm,2024-02-05
3 1424,A18334755,Beta-Alanine,2020-04-27,03/10/2005,02:03:47 PM,04/27/2020,09:04:15 PM,63,,ENG,06-APR-20,,RXNORM_15AB_200406F,63,RXNORM_IN,63,RxNorm,2024-02-05
4 1684,A18334529,4-Aminobenzoic Acid,2020-04-27,03/10/2005,02:03:47 PM,04/27/2020,09:04:14 PM,74,,ENG,06-APR-20,,RXNORM_15AB_200406F,74,RXNORM_IN,74,RxNorm,2024-02-05
5 2182,A18751816,Eicosapentaenoic Acid,2020-04-27,03/10/2005,02:03:47 PM,04/27/2020,09:04:18 PM,96,,ENG,06-APR-20,,RXNORM_15AB_200406F,96,RXNORM_P2N,96,RxNorm,2024-02-05
6
```

CSV delimiter

Comma

Tab

Custom

SQL query

Add SQL from templates

Run SQL query

Amazon S3 Select supports only the SELECT SQL command. Using the S3 console, you can extract up to 40 MB of records from an object that is up to 128 MB in size. To work with larger files or more records, use the AWS CLI, AWS SDK, or Amazon S3 REST API. For more complex SQL queries, use [Amazon Athena](#)

1 /* To create reference point for writing SQL queries, you can display the first 5 records of input data by running the following SQL query: SELECT * FROM s3object s LIMIT 5 */
2 SELECT * FROM s3object s LIMIT 5

SQLLn 1, Col 1Errors: 0Warnings: 0

Query results

Download results

Status

Successfully returned 5 records in 2311 ms
Bytes returned: 354 B

RawFormatted

CUI1	VER_START	VER_END	Cardinality	CUI2	Code set	Version Month
89504	RXNORM_17AA_170905F	RXNORM_15AA_221005F	1	89504	RxNorm	2024-02-05
91669	RXNORM_17AA_170905F	RXNORM_15AB_181205F	1	91669	RxNorm	2024-02-05
91714	RXNORM_15AA_100802F	RXNORM_15AA_100802F	1	1000992	RxNorm	2024-02-05
91811	RXNORM_0AAC_050210F	RXNORM_05AA_090706F	1	857082	RxNorm	2024-02-05

Activate Windows

5.Cloudwatch log insights

#	@timestamp	@message	@logStream	@log
1	2024-03-11T21:18:59.78...	...	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
2	2024-03-11T21:18:24.08...	END RequestId: 94d6d380-6ea6-481c-8f75-e1bd87d0fdcl	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
3	2024-03-11T21:18:24.08...	REPORT RequestId: 94d6d380-6ea6-481c-8f75-e1bd87d0fdcl Duration: 218028.7...	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
4	2024-03-11T21:18:23.54...	Number of rows in ROONSTY after processing: 461874	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
5	2024-03-11T21:18:23.54...	Processed .RRF files uploaded back to S3 successfully.	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
6	2024-03-11T21:18:23.66...	Number of rows in ROONSAT after processing: 7222404	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
7	2024-03-11T21:17:29.91...	Number of rows in ROONAB after processing: 13	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
8	2024-03-11T21:17:29.80...	Number of rows in ROONREL after processing: 7154306	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
9	2024-03-11T21:16:23.69...	Number of rows in ROONDOC after processing: 3445	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
10	2024-03-11T21:16:23.57...	Number of rows in ROONUICHANGES after processing: 153	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
11	2024-03-11T21:16:23.44...	Number of rows in ROONCUI after processing: 30046	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
12	2024-03-11T21:16:23.26...	Number of rows in ROONCHSD after processing: 1133065	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
13	2024-03-11T21:16:12.27...	Number of rows in ROONATOMARCHIVE after processing: 371368	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
14	2024-03-11T21:16:07.93...	Number of rows in ROONSTY initially: 461874	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
15	2024-03-11T21:16:07.34...	Number of rows in ROONSAT initially: 7222404	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
16	2024-03-11T21:15:46.89...	Number of rows in ROONAB initially: 13	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
17	2024-03-11T21:15:46.77...	Number of rows in ROONREL initially: 7154306	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
18	2024-03-11T21:15:26.96...	Number of rows in ROONDOC initially: 3445	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
19	2024-03-11T21:15:26.93...	Number of rows in ROONUICHANGES initially: 153	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func
20	2024-03-11T21:15:26.91...	Number of rows in ROONCUI initially: 30046	2024/03/11/[LATEST]4dcb25caf3ac4e4d886ce1clacc4ca39	533267198420:/aws/lambda/zip_func