

TASK - Python and Lambda

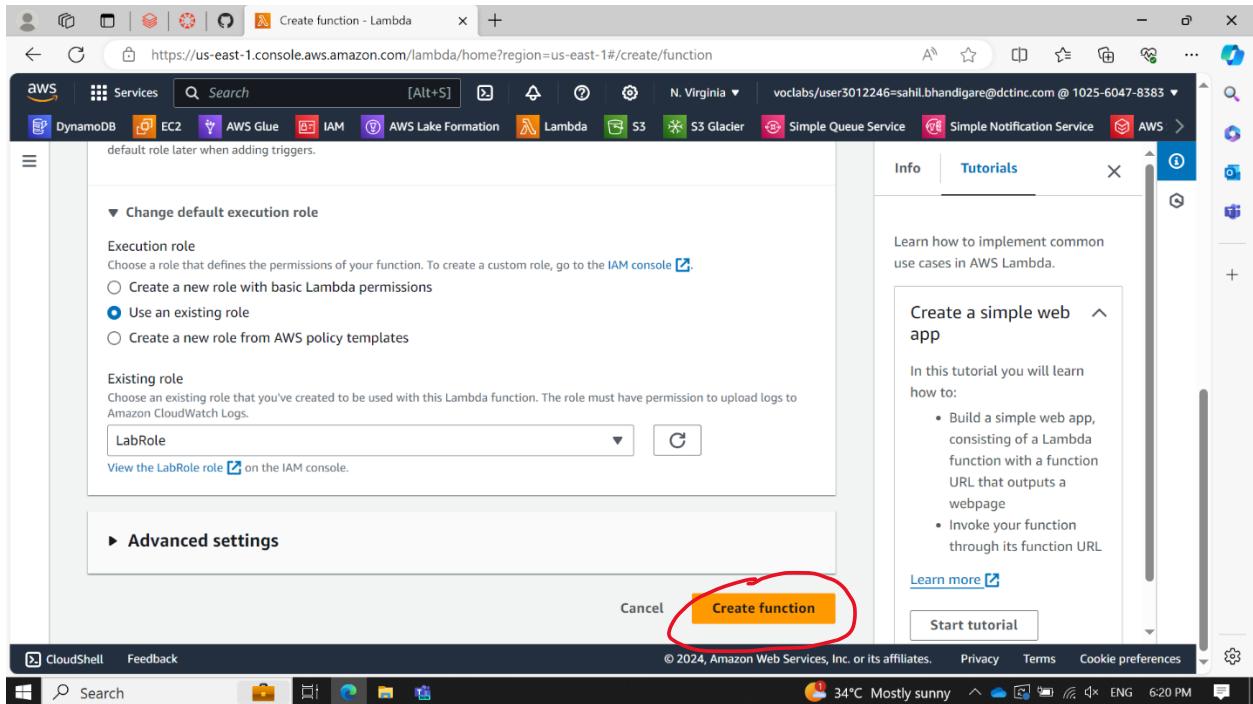
Work with RXNORM file,

1. Add header into each rff from RXNORM.xlsx
2. Add CODE_SET & VERSION_MONTH column with default values RxNorm and version month from downloaded filename
3. Convert dates into YYYY-MM-DD
4. Save files as txt delimited by comma(,)
5. Validate row_count between original and converted files

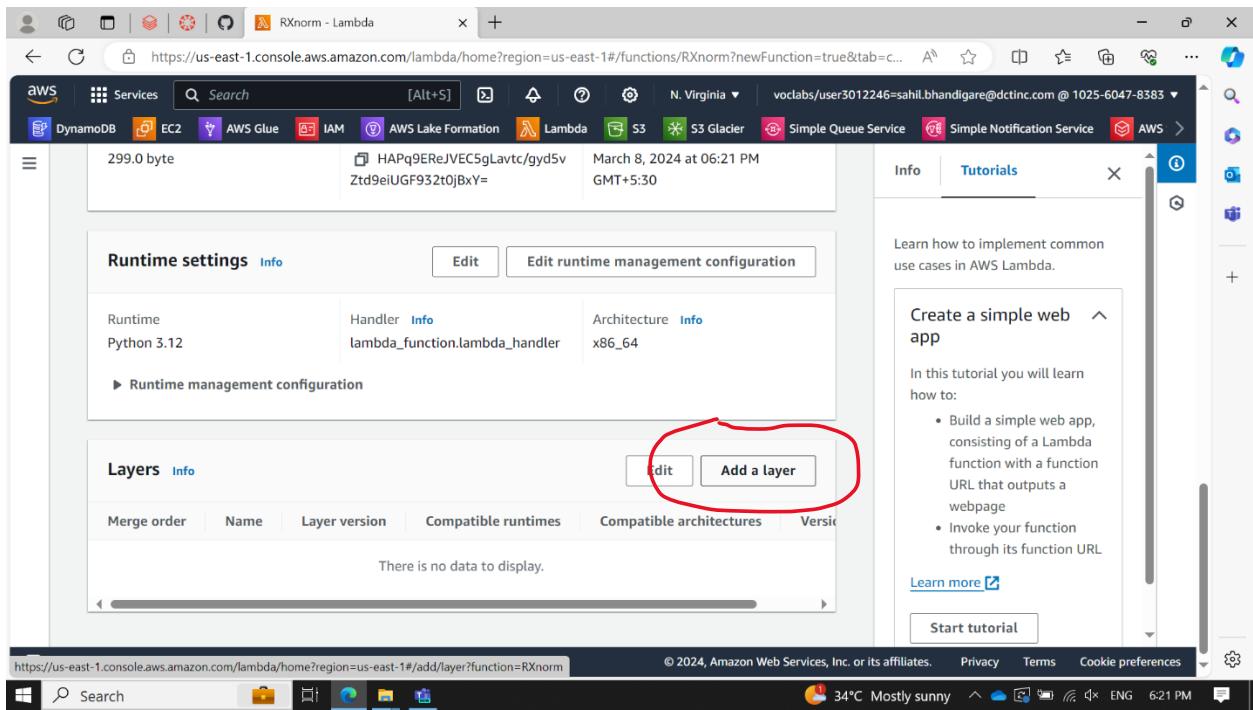
CREATING ONE FUNCTION (RXnorm) IN LAMBDA

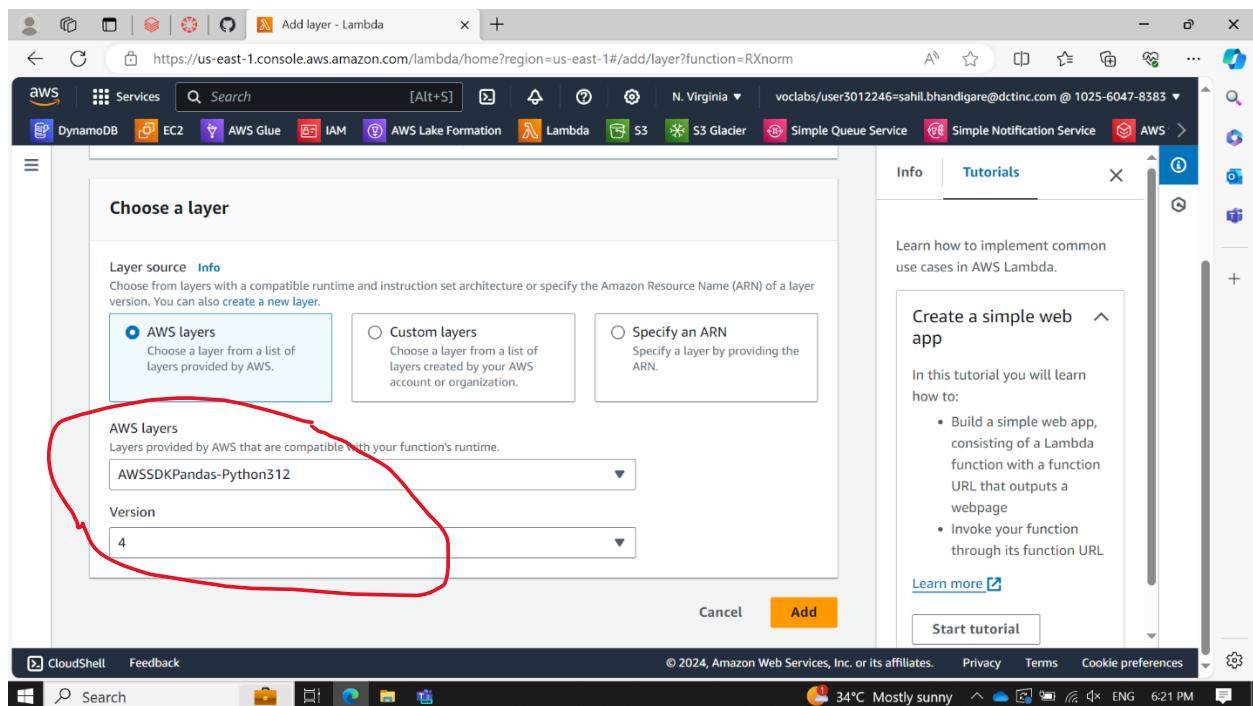
The screenshot shows the AWS Lambda Functions page. On the left, there's a sidebar with options like Dashboard, Applications, Functions (which is selected), Additional resources, and Related AWS resources. The main area displays a table of existing functions: **RoleCreationFunction** and **MainMonitoringFunction**. Below the table is a search bar and a 'Create function' button, which is circled in red. A sidebar on the right provides a tutorial titled 'Create a simple web app'.

The screenshot shows the 'Create function' wizard. In the 'Basic information' step, the 'Author from scratch' option is selected. The 'Function name' field contains 'RXnorm'. The 'Runtime' dropdown is set to 'Python 3.12'. A sidebar on the right provides a tutorial titled 'Create a simple web app'.

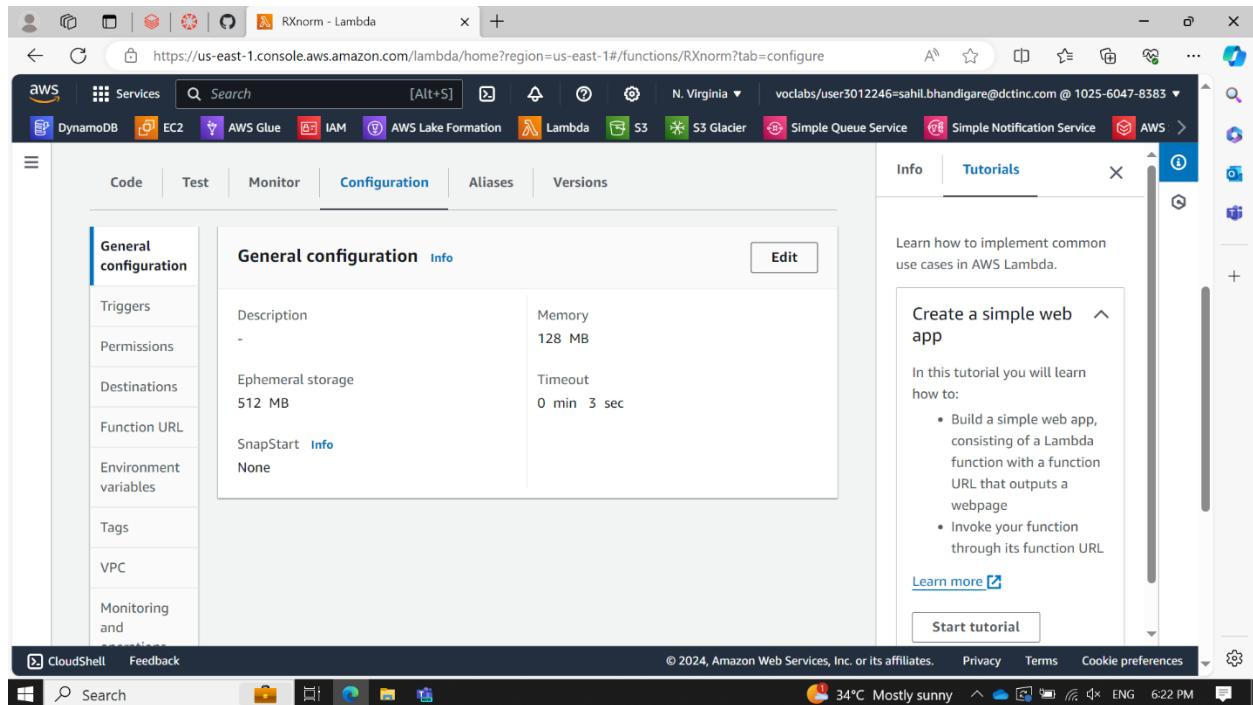


ADDING (PANDAS VERSION 4) LAYER IN IT





CHANGING CONFIGURATION OF OUR FUNCTION



The screenshot shows the AWS Lambda 'Edit basic settings' page for a function named 'RXnorm'. The 'Memory' section is set to 5120 MB. The 'Ephemeral storage' section is set to 512 MB. Under 'SnapStart', the 'None' option is selected. The 'Timeout' is set to 10 minutes and 3 seconds. The 'Execution role' dropdown shows 'Use an existing role' selected. The page includes a sidebar with a tutorial for creating a simple web app.

CREATE 2 S3 BUCKET (1 FOR INPUT AND 1 FOR OUTPUT)

The screenshot shows the 'Create bucket' page in the AWS S3 console. The 'General configuration' section is displayed. Under 'AWS Region', 'US East (N. Virginia) us-east-1' is selected. Under 'Bucket type', 'General purpose' is selected. A bucket name 'mybucketversion1' is entered in the 'Bucket name' field. The page includes a sidebar with a tutorial for creating a simple web app.

Create S3 bucket | S3 | Global

https://s3.console.aws.amazon.com/s3/bucket/create?region=us-east-1&bucketType=general

aws Services Search [Alt+S] Global vocabs/user3012246=sahil.bhandigare@dctinc.com @ 1025-6047-8383 AWS

DynamoDB EC2 AWS Glue IAM AWS Lake Formation Lambda S3 S3 Glacier Simple Queue Service Simple Notification Service AWS

Object Ownership Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

⚠ We recommend disabling ACLs, unless you need to control access for each object individually or to have the object writer own the data they upload. Using a bucket policy instead of ACLs to share data with users outside of your account simplifies permissions management and auditing.

Object Ownership

Bucket owner preferred
If new objects written to this bucket specify the bucket-owner-full-control canned ACL, they are owned by the bucket owner. Otherwise, they are owned by the object writer.

Object writer
The object writer remains the object owner.

CloudShell Feedback Search 34°C Mostly sunny Privacy Terms Cookie preferences

Create S3 bucket | S3 | Global

https://s3.console.aws.amazon.com/s3/bucket/create?region=us-east-1&bucketType=general

aws Services Search [Alt+S] Global vocabs/user3012246=sahil.bhandigare@dctinc.com @ 1025-6047-8383 AWS

DynamoDB EC2 AWS Glue IAM AWS Lake Formation Lambda S3 S3 Glacier Simple Queue Service Simple Notification Service AWS

Block all public access

Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

Block public access to buckets and objects granted through new access control lists (ACLs)
S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.

Block public access to buckets and objects granted through any access control lists (ACLs)
S3 will ignore all ACLs that grant public access to buckets and objects.

Block public access to buckets and objects granted through new public bucket or access point policies
S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.

Block public and cross-account access to buckets and objects through any public bucket or access point policies
S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

⚠ Turning off block all public access might result in this bucket and the objects within becoming public
AWS recommends that you turn on block all public access, unless public access is required for specific and verified use cases such as static website hosting.

I acknowledge that the current settings might result in this bucket and the objects within becoming public.

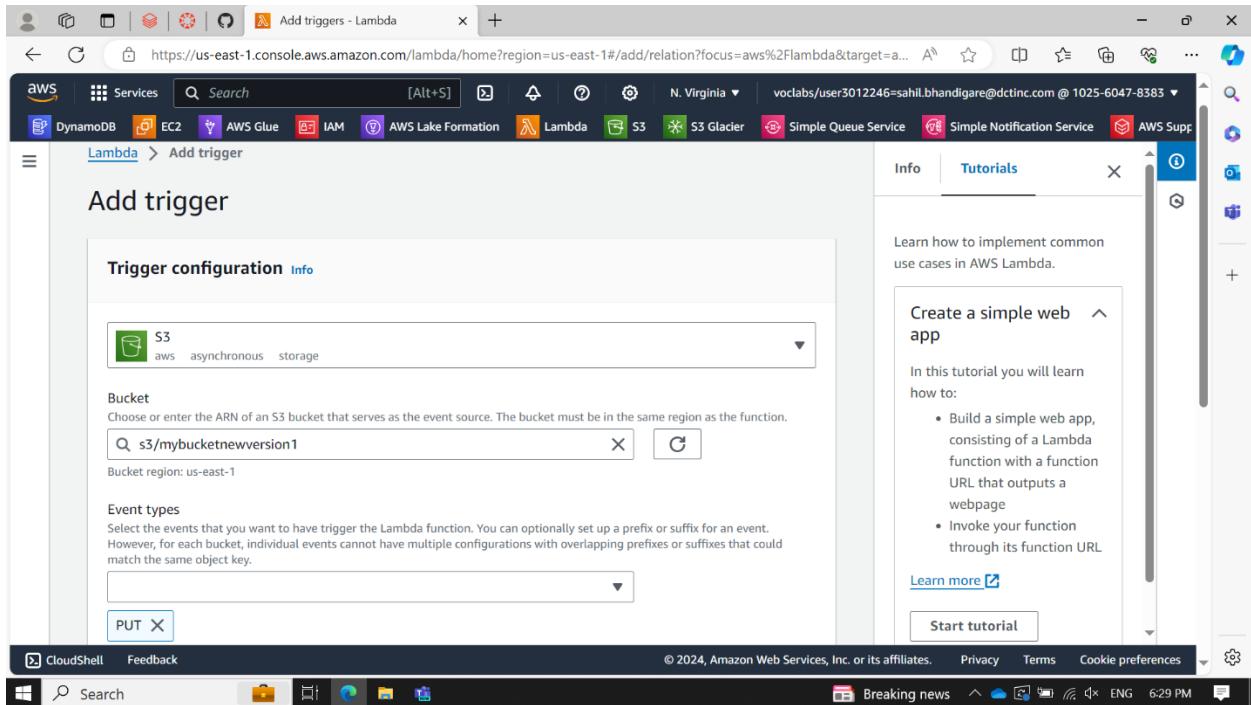
CloudShell Feedback Search 34°C Mostly sunny Privacy Terms Cookie preferences

The screenshot shows the AWS S3 buckets page. At the top, there are three metrics: Total storage (Pending), Object count (Pending), and Average object size (Pending). A note says you can enable advanced metrics in the "default-account-dashboard" configuration. Below this, there are tabs for General purpose buckets and Directory buckets, with General purpose buckets selected. It shows one bucket named "mybucketnewversion1" created on March 8, 2024, at 18:27:22 (UTC+05:30). The bucket is located in the US East (N. Virginia) region and has "Objects can be public" access.

PUTTING FIELDS INTO OUR INPUT BUCKET

The screenshot shows the AWS S3 Objects page for the bucket "mybucketnewversion1". It displays two objects: "excel/" and "final_output/". Both are folder objects. The "excel/" folder was last modified on March 8, 2024, at 18:27:22 (UTC+05:30). The "final_output/" folder was last modified on March 8, 2024, at 18:27:22 (UTC+05:30). The "Actions" dropdown menu is open, showing options like Copy S3 URI, Copy URL, Download, Open, Delete, and Create folder. There is also a "Upload" button. A note at the bottom explains that objects are fundamental entities stored in Amazon S3 and provides a link to the Amazon S3 inventory.

ADDING TRIGGER (S3 INPUT BUCKET) TO OUR FUNCTION



CODE :

```
import os

import boto3

import zipfile

import io

import logging

from datetime import datetime

import pandas as pd
logger = logging.getLogger()

logger.setLevel(logging.INFO)
```

```

s3_client = boto3.client('s3')
def default_columns(df, file_name):

#     if 'RXNATOMARCHIVE' in file_name:

#         df['CREATED_TIMESTAMP'] = df['CREATED_TIMESTAMP'].apply(lambda x:
datetime.strptime(x, "%Y-%m-%d %I:%M:%S %p") if pd.notnull(x) else x)

#         df['UPDATED_TIMESTAMP'] = df['UPDATED_TIMESTAMP'].apply(lambda x:
datetime.strptime(x, "%Y-%m-%d %I:%M:%S %p") if pd.notnull(x) else x)

#         df['LAST_RELEASED'] =
df['LAST_RELEASED'].fillna('').astype(str).apply(lambda x: datetime.strptime(x, '%Y-%m-%d').strftime('%Y-%m-%d') if pd.notnull(x) else x)

#     elif 'RXNSAB' in file_name:

#         df['VSTART'] = df['VSTART'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d').strftime('%Y-%m-%d') if pd.notnull(x) else x)

#         df['VEND'] = df['VEND'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d').strftime('%Y-%m-%d') if pd.notnull(x) else x)

df['Code set'] = 'Rxnorm'

df['Version Month'] = "02052024"

def lambda_handler(event, context):

    try:

        logger.info("Received event: %s", event)

        rows_before_conversion = 0

        rows_after_conversion = 0
        # Get the S3 bucket and object key from the event

        bucket = event['Records'][0]['s3']['bucket']['name']

        key = event['Records'][0]['s3']['object']['key']

```

```
# Download the zip file from S3

response = s3_client.get_object(Bucket=bucket, Key=key)

zip_data = response['Body'].read()
# Extract the zip file

with zipfile.ZipFile(io.BytesIO(zip_data)) as zip_ref:

    for file_name in zip_ref.namelist():

        if 'rrf/RX' in file_name:

            # Extract the file

            with zip_ref.open(file_name) as file_data:

                df = pd.read_csv(file_data, delimiter='|')

                rows_before_conversion += len(df)

            # Get header names from the corresponding Excel sheet

            num_cols_df = df.shape[1]

            sheet_name = os.path.splitext(os.path.basename(file_name))[0]

            header_file_key = 'excel/RxNorm_Header.xlsx'

            header_df = read_excel_from_s3(bucket, header_file_key,
sheet_name)

            if header_df is not None:

                df.columns = header_df.iloc[:, 0].tolist()[:num_cols_df]

                default_columns(df, file_name)

            # Convert DataFrame to CSV format

            csv_data = df.to_csv(index=False)

            rows_after_conversion += df.shape[0]
```

```
# Upload the CSV file to another S3 bucket

destination_bucket = 'mybucketdestinationnew1'

destination_key = f"{sheet_name}.csv"

s3_client.put_object(Bucket=destination_bucket,
Key=destination_key, Body=csv_data)

return {

    'statusCode': 200,

    'body': f'Success. Rows before conversion: {rows_before_conversion},\nRows after conversion: {rows_after_conversion}'

}

except Exception as e:

    logger.error("Error: %s", e)

    return {

        'statusCode': 500,

        'body': f'Error: {e}'

    }

def read_excel_from_s3(bucket, key, sheet_name):

    try:

        response = s3_client.get_object(Bucket=bucket, Key=key)

        excel_bytes = response[ 'Body' ].read()

        df = pd.read_excel(io.BytesIO(excel_bytes), sheet_name=sheet_name,
header=None)

        return df
```



```
1 import os
2
3 import boto3
4
5 import zipfile
6
7 import io
8
9 import logging
10
11 from datetime import datetime
12
13 import pandas as pd
14 logger = logging.getLogger()
15
16 logger.setLevel(logging.INFO)
17
18 s3_client = boto3.client('s3')
19 def default_columns(df, file_name):
20     # if 'RXNATOMARCHIVE' in file_name:
```

The screenshot shows the AWS Lambda console interface. In the top navigation bar, the URL is https://us-east-1.console.aws.amazon.com/lambda/home?region=us-east-1#/functions/RXnorm?tab=code. The main area displays the code editor for a Lambda function named 'RXnorm'. The code itself is a Python script named 'lambda_function.py' which handles S3 objects by reading them into pandas DataFrames and logging the results. A sidebar on the right provides a tutorial on creating a simple web app using Lambda.

CREATING OUR SECOND BUCKET

Name	AWS Region	Access	Creation date
mybucketdestinationnew1	US East (N. Virginia) us-east-1	Objects can be public	March 8, 2024, 18:45:55 (UTC+05:30)
mybucketnewversion1	US East (N. Virginia) us-east-1	Objects can be public	March 8, 2024, 18:27:22 (UTC+05:30)

The screenshot shows the AWS S3 console interface. In the top navigation bar, the URL is https://s3.console.aws.amazon.com/s3/buckets?region=us-east-1&bucketType=general. The main area displays a list of buckets under the 'General purpose buckets' tab. There are two buckets listed: 'mybucketdestinationnew1' and 'mybucketnewversion1', both located in the US East (N. Virginia) region. Both buckets have 'Objects can be public' access and were created on March 8, 2024, at different times.

ADDING ZIP FILE INTO INPUT BUCKET

The screenshot shows the AWS S3 console interface. The top navigation bar includes links for Services, AWS Lambda, and other AWS services like DynamoDB, EC2, IAM, and S3. The main content area is titled 'mybucketnewversion1' and shows the 'Objects' tab selected. There are three objects listed:

Name	Type	Last modified	Size	Storage class
excel/	Folder	-	-	-
final_output/	Folder	-	-	-
rrf.zip	zip	March 8, 2024, 18:38:10 (UTC+05:30)	1.8 MB	Standard

The bottom of the screen shows standard browser controls, a search bar, and a system status bar indicating 33°C, Partly sunny, and the time as 6:37 PM.

AFTER DEPLOYING CODE CHECKING LOGS

The screenshot shows the AWS CloudWatch logs console. The left sidebar has a 'CloudWatch' section with 'Logs' expanded, showing 'Log groups', 'Log Anomalies', 'Live Tail', and 'Logs Insights'. The main content area is titled 'Log events' for the log group '\$252Faws\$...'. It displays log entries with columns for 'Timestamp' and 'Message'.

Timestamp	Message
2024-03-08T18:48:02.320+05:30	INIT_START Runtime Version: python:3.12.v20 Runtime Version ARN: arn:aws:lambda:...
2024-03-08T18:48:05.092+05:30	[INFO] 2024-03-08T13:18:05.092Z Found credentials in environment variables.
2024-03-08T18:48:05.232+05:30	START RequestId: 9d55616f-0347-4019-8a7d-40c041883c04 Version: \$LATEST
2024-03-08T18:48:05.233+05:30	[INFO] 2024-03-08T13:18:05.233Z 9d55616f-0347-4019-8a7d-40c041883c04 Received e...
2024-03-08T18:48:08.625+05:30	END RequestId: 9d55616f-0347-4019-8a7d-40c041883c04
2024-03-08T18:48:08.626+05:30	REPORT RequestId: 9d55616f-0347-4019-8a7d-40c041883c04 Duration: 3393.56 ms Bill...

The bottom of the screen shows standard browser controls, a search bar, and a system status bar indicating 33°C, Partly sunny, and the time as 6:49 PM.

OUTPUT FILES INTO OUR DESTINATION BUCKET

The screenshot shows the AWS S3 console interface. The top navigation bar includes tabs for Lambda, CloudWatch, and mybucketdestinationnew1. The main navigation bar lists services like DynamoDB, EC2, IAM, AWS Glue, AWS Lake Formation, Lambda, S3, S3 Glacier, Simple Queue Service, Simple Notification Service, and AWS. The current view is under the S3 service, specifically for the 'mybucketdestinationnew1' bucket. The left sidebar shows the bucket name. The main content area displays a table titled 'Objects (5) Info' containing one item: 'RXNCUI.csv'. The table columns include Name, Type, Last modified, Size, and Storage class. The file 'RXNCUI.csv' is a CSV file, last modified on March 8, 2024, at 18:48:06 (UTC+05:30), with a size of 2.1 MB and a Standard storage class. Below the table are buttons for Actions (Copy S3 URI, Copy URL, Download, Open, Delete, Create folder) and a search bar. The bottom of the screen shows the Windows taskbar with various pinned icons.

This screenshot shows the same AWS S3 console interface as the previous one, but now displaying five objects in the 'mybucketdestinationnew1' bucket. The table in the 'Objects (5) Info' section lists the following files:

Name	Type	Last modified	Size	Storage class
RXNCUI.csv	csv	March 8, 2024, 18:48:06 (UTC+05:30)	2.1 MB	Standard
RXNCUICHANGES.csv	csv	March 8, 2024, 18:48:07 (UTC+05:30)	17.5 KB	Standard
RXNDOC.csv	csv	March 8, 2024, 18:48:07 (UTC+05:30)	265.0 KB	Standard
RXNSAB.csv	csv	March 8, 2024, 18:48:07 (UTC+05:30)	9.7 KB	Standard
RXNSTY.csv	csv	March 8, 2024, 18:48:09 (UTC+05:30)	25.2 MB	Standard

The rest of the interface is identical to the first screenshot, including the top navigation bar, sidebar, and taskbar.

HEADERS APPLIED TO EACH FILES

Office 4 Excel window showing the RXNUI file.

File **Home** **Insert** **Page Layout** **Formulas** **Data** **Review** **View** **Automate** **Help**

AutoSave OFF RXNS... • Saved to this PC Search

Comments Share

Font Alignment Number Styles Cells Editing Add-ins Analyze Data

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

A1 RXNUI

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	RXNUI	TUI	STN	STY	ATUI	CVF	Code set	Version	Month										
2	3 T116		A1.4.1.2.1 Amino Acid, Peptide, or Protein		Rxnorm		02052024												
3	19 T125		A1.4.1.1.3 Hormone		Rxnorm		02052024												
4	19 T109		A1.4.1.2.1 Organic Chemical		Rxnorm		02052024												
5	21 T125		A1.4.1.1.3 Hormone		Rxnorm		02052024												
6	21 T109		A1.4.1.2.1 Organic Chemical		Rxnorm		02052024												
7	22 T121		A1.4.1.1.1 Pharmacologic Substance		Rxnorm		02052024												
8	22 T109		A1.4.1.2.1 Organic Chemical		Rxnorm		02052024												
9	33 T116		A1.4.1.2.1 Amino Acid, Peptide, or Protein		Rxnorm		02052024												
10	33 T126		A1.4.1.1.3 Enzyme		Rxnorm		02052024												
11	36 T123		A1.4.1.1.3 Biologically Active Substance		Rxnorm		02052024												
12	36 T116		A1.4.1.2.1 Amino Acid, Peptide, or Protein		Rxnorm		02052024												
13	38 T109		A1.4.1.2.1 Organic Chemical	4096	Rxnorm		02052024												
14	38 T121		A1.4.1.1.1 Pharmacologic Subst	4096	Rxnorm		02052024												
15	43 T126		A1.4.1.1.3 Enzyme		Rxnorm		02052024												
16	43 T116		A1.4.1.2.1 Amino Acid, Peptide, or Protein		Rxnorm		02052024												
17	44 T109		A1.4.1.2.1 Organic Chemical	4096	Rxnorm		02052024												
18	44 T121		A1.4.1.1.1 Pharmacologic Subst	4096	Rxnorm		02052024												
19	47 T109		A1.4.1.2.1 Organic Chemical		Rxnorm		02052024												
20	47 T131		A1.4.1.1.5 Hazardous or Poisonous Substar		Rxnorm		02052024												

Ready Accessibility: Unavailable 33°C Partly sunny ENG 6:52 PM

Office 4 Excel window showing the VCUI file.

File Home Insert Page Layout Formulas Data Review View Automate Help

AutoSave OFF RXNS... • Saved to this PC Search

Comments Share

Font Alignment Number Styles Cells Editing Add-ins Analyze Data

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

A1 VCUI

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	VLCI	RCUI	VSAB	RSAB	SON	SF	SVER	VSTART	VEND	IMETA	RMETA	SLC	SCC	SRL	TFR	CFR	CXTY	TTYL	ATNL	LA
2	C5233830 C1140182 MMX_2024-MMX				Micromedex	MMX	2024_01_02			2020AA		;;Micromedex;6200 S		3				MTH_RXN_DCSA,DDF	EN	
3	C5233828 C1140261 NDDF_2024NDDF				FDB MedK	NDDF	2024_01_03			2020AA		;;First Dat	;;First Dat	3				CDC,CDA,NDC	EN	
4	C5233835 C1140284 RXNORM_RXNORM				RxNorm	VcRXNORM	20AA_240205F			2020AA		RxNorm	Ct RxNorm Ct	0				SCDF,DFG,AMBIGUIT	EN	
5	C3531723 C2720507 SNOMEDC SNOMEDC US Edition				SNOMEDC	US Edition	SNOMEDC2023_06_30			2020AA		National Li	National Li	9				FN,SY,PT,PTGB,SYGB	EN	
6	C5233837 C1140288 VANDF_20VANDF				Veterans F	VANDF	2023_12_29			2020AA		Michael Lir	Michael Lir	0				CD,MTH_F_DCSA,DDF	EN	
7	C5233836 C1812643 MTHSP_2_MTHSP				Metathesa	MTHSP	2024_01_22006_12_21			2020AA		RxNorm	Ct RxNorm Ct	0				DP,SU,MT,ANADA,AN	EN	
8	C5233833 C1876239 GS_2024_GS				Gold Stanc	GS	2024_01_C2007_05_04			2020AA		RxNorm	Ct Kay Morgan	3				CD,IN,MT,AMBIGUIT	EN	
9	C5233831 C4722517 ATC_2024.ATC				Anatomic	ATC	2024_01_19			2020AA		;;WHO Col	;;WHO Col	0				IN,RXN_IN_ATC_LEVEL	EN	
10	C5233832 C3539983 CVX_2024_CVX				Vaccines /	CVX	2024_01_08			2020AA		;;CDC, Nat	;;CDC, Nat	0				AB,PT,RXN_DATE_LAS	EN	
11	C5233829 C38558951 MTHCMFS MTHCMFS				Metathesa	MTHCMFS	2020			2020AA		RxNorm	Ct RxNorm Ct	0				PT	EN	
12	C5233838 C4255544 DRUGBAN DRUGBAN				DrugBank	DrugBank	DRUGBAN 5.0_2024_01_04			2020AA		;;Omrx Pers	;;Omrx Pers	0				SY,IN,FSY FDA_UNII	EN	
13	C5233834 C4554231 USP_2024 USP				USP Comp	USP	2024_01_08			2020AA		Jeffrey Shi	Jeffrey Shi	0				CD,IN MONOGR	EN	

Ready Accessibility: Unavailable 33°C Partly sunny ENG 6:52 PM

AutoSave (Off) X Search office4 X

File Home Insert Page Layout Formulas Data Review View Automate Help Comments Share

Clipboard Font Alignment Number Conditional Formatting Table Styles Cells Editing Add-ins Analyze Data

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As... X

A1 CUI1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	CUI1	VER_STAR	VER_END	Cardinality	CUI2	Code set	Version	Month										
2	91669	RXNORM	_RXNORM_	1	91669	Rxnorm	02052024											
3	91714	RXNORM	_RXNORM_	1	1000992	Rxnorm	02052024											
4	91811	RXNORM	_RXNORM_I	1	857082	Rxnorm	02052024											
5	91812	RXNORM	_RXNORM_I	1	857087	Rxnorm	02052024											
6	91919	RXNORM	_RXNORM_I	1	849580	Rxnorm	02052024											
7	91942	RXNORM	_RXNORM_	1	1293743	Rxnorm	02052024											
8	91960	RXNORM	_RXNORM_	1	991063	Rxnorm	02052024											
9	91961	RXNORM	_RXNORM_	1	991084	Rxnorm	02052024											
10	92286	RXNORM	_RXNORM_I	1	895989	Rxnorm	02052024											
11	92548	RXNORM	_RXNORM_	1	1048284	Rxnorm	02052024											
12	92549	RXNORM	_RXNORM_	1	1048288	Rxnorm	02052024											
13	92643	RXNORM	_RXNORM_I	1	856710	Rxnorm	02052024											
14	92690	RXNORM	_RXNORM_	1	994091	Rxnorm	02052024											
15	92691	RXNORM	_RXNORM_	1	994093	Rxnorm	02052024											
16	92840	RXNORM	_RXNORM_I	1	859873	Rxnorm	02052024											
17	92841	RXNORM	_RXNORM_I	1	859869	Rxnorm	02052024											
18	92887	RXNORM	_RXNORM_I	1	856955	Rxnorm	02052024											
19	93008	RXNORM	_RXNORM_I	1	793871	Rxnorm	02052024											
20	93162	RXNORM	_RXNORM_I	1	857113	Rxnorm	02052024											

Ready Accessibility: Unavailable 33°C Partly sunny 100% 6:52 PM