Chandini Kalidindi
Assignment 2

In this assignment, I learned how to use aws tools to trigger our scrapers periodically according to a schedule.

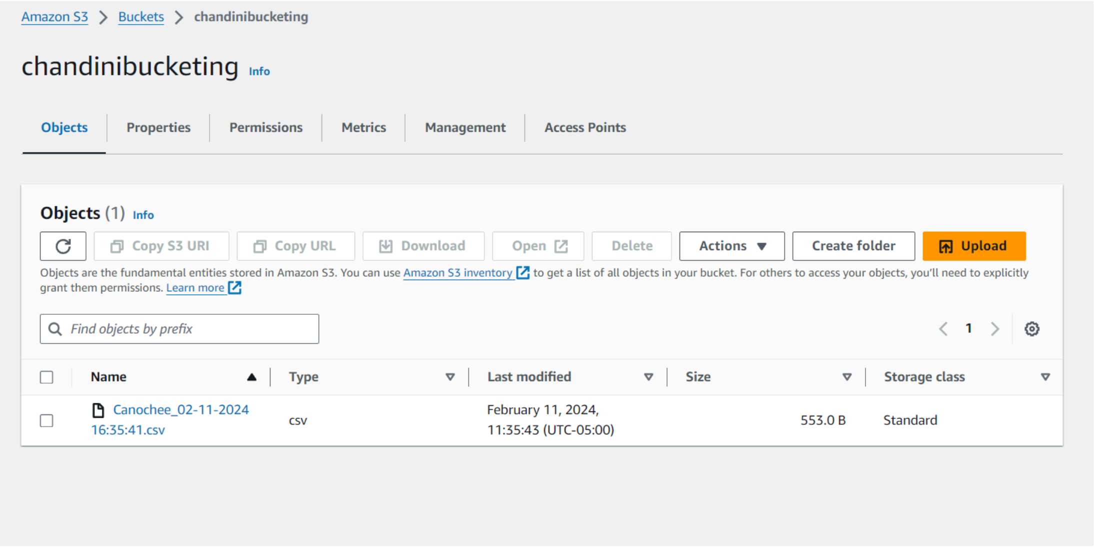
To do this I first created an S3 bucket where my data will be stored. After implementing my scraper into the code given, I also created a ECR repository where I built and pushed my docker image. I then used AWS Lambda to run my code. Overall, this process allowed me to run my code and store my data in the cloud without the need of my own server.

Since I want to periodically collect this outage data, I created an Amazon Event Bridge to trigger my lambda function every 15 mins.

I also had many difficulties trying to push my image to docker. At first, I tried to follow the Windows commands and use AWS tools for powershell but I could not find any download or common that allowed me to use it on my computer. I then went back to using linux commands on AWS CLI. Even though I had AwS CLI downloaded, any complex command kept timing out including docker commands. After restarting my computer, I later used the aws configure command to store my credentials each time I needed authentication. When testing my Lambda function for the first time, I could not figure out what was causing an error. I tried building an image using the example code instead which worked. This means that something was wrong with my code that worked for assignment 1. After testing my scraper code independently, I discovered that python requests was not working specifically with the url I used for my assignment 1 scraper. I was not able to figure out why and it was occurring when I tested it on a separate device too. It seems that since I submitted assignment 1, there is some issue when trying to retrieve that link for security reasons. I chose to use a different link that worked and my lambda function was successful.

Overall, I learned a lot from this assignment about how to utilize AWS Services and also how to troubleshoot a project using multiple tools and services.

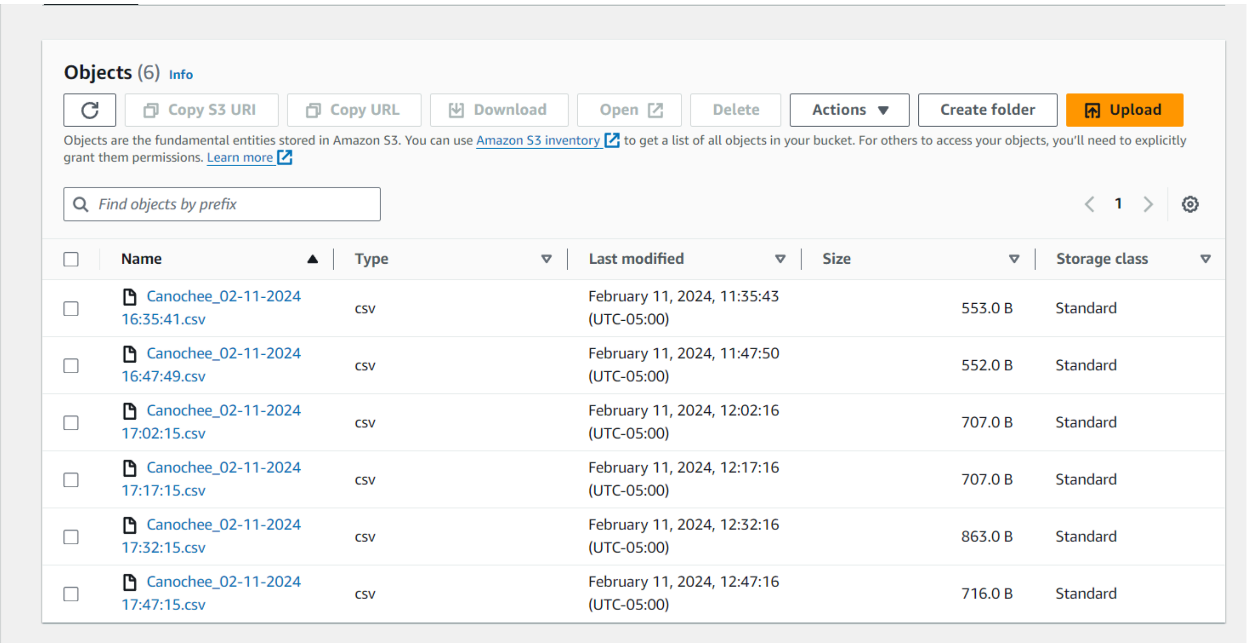
GitHub Repo: [chankal/scraper2 (github.com)]
S3 Bucket: chandinibucketing



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

□	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
□	Canochee_02-11-2024 16:35:41.csv	csv	February 11, 2024, 11:35:43 (UTC-05:00)	553.0 B	Standard

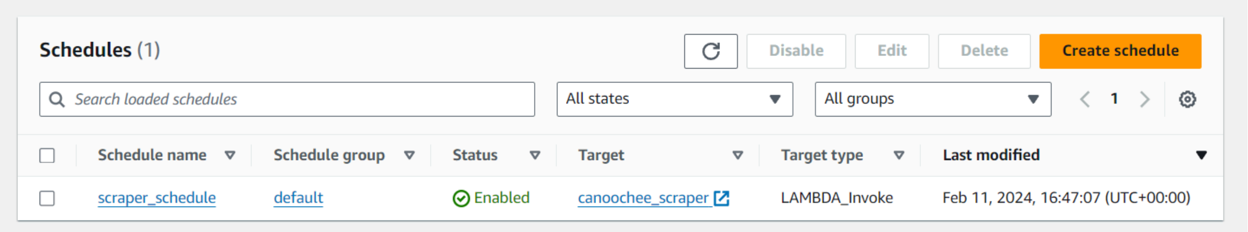
After using Event Bridge to trigger every 15 mins:



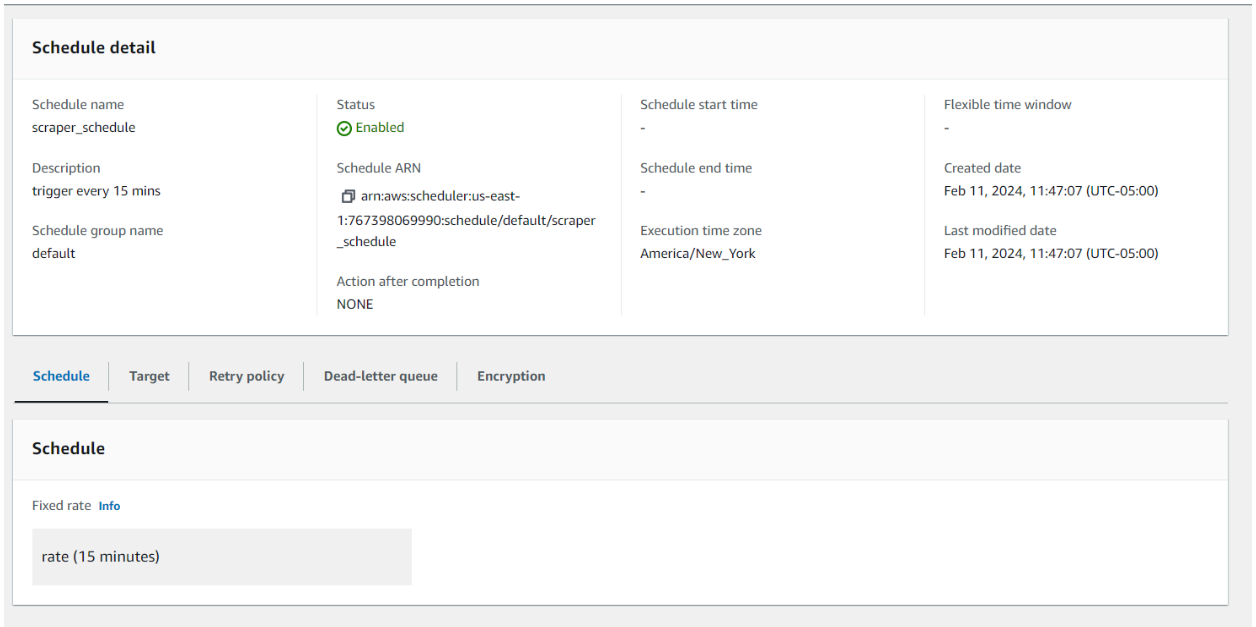
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

□	Name	Type	Last modified	Size	Storage class
□	Canochee_02-11-2024 16:35:41.csv	csv	February 11, 2024, 11:35:43 (UTC-05:00)	553.0 B	Standard
□	Canochee_02-11-2024 16:47:49.csv	csv	February 11, 2024, 11:47:50 (UTC-05:00)	552.0 B	Standard
□	Canochee_02-11-2024 17:02:15.csv	csv	February 11, 2024, 12:02:16 (UTC-05:00)	707.0 B	Standard
□	Canochee_02-11-2024 17:17:15.csv	csv	February 11, 2024, 12:17:16 (UTC-05:00)	707.0 B	Standard
□	Canochee_02-11-2024 17:32:15.csv	csv	February 11, 2024, 12:32:16 (UTC-05:00)	863.0 B	Standard
□	Canochee_02-11-2024 17:47:15.csv	csv	February 11, 2024, 12:47:16 (UTC-05:00)	716.0 B	Standard

Scheduler Details :



□	Schedule name	Schedule group	Status	Target	Target type	Last modified
□	scraper_schedule	default	✓ Enabled	canoochee_scraper	LAMBDA_Invoke	Feb 11, 2024, 16:47:07 (UTC+00:00)



Schedule detail

Schedule name
scraper_schedule

Description
trigger every 15 mins

Schedule group name
default

Status
✓ Enabled

Schedule ARN
arn:aws:scheduler:us-east-1:767398069990:schedule/default/scraper_schedule

Action after completion
NONE

Schedule start time
-

Schedule end time
-

Execution time zone
America/New_York

Flexible time window
-

Created date
Feb 11, 2024, 11:47:07 (UTC-05:00)

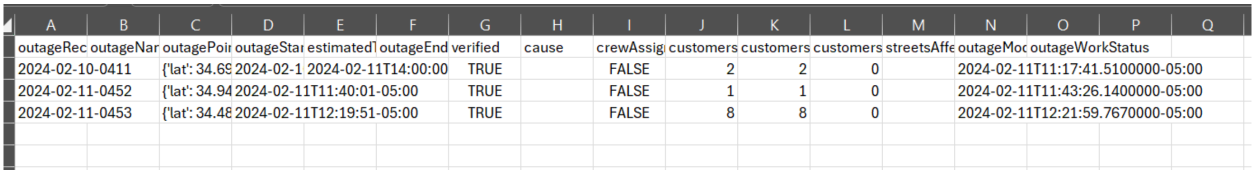
Last modified date
Feb 11, 2024, 11:47:07 (UTC-05:00)

Schedule

Fixed rate Info

rate (15 minutes)

Sample data:



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
outageRec	outageNar	outagePoir	outageStar	estimated	outageEnd	verified	cause	crewAssig	customers	customers	customers	streetsAffe	outageMoc	outageWorkStatus		
2024-02-10-0411		{'lat': 34.69	2024-02-1	2024-02-11T14:00:00		TRUE		FALSE	2	2	0		2024-02-11T11:17:41.5100000-05:00			
2024-02-11-0452		{'lat': 34.94	2024-02-11T11:40:01-05:00			TRUE		FALSE	1	1	0		2024-02-11T11:43:26.1400000-05:00			
2024-02-11-0453		{'lat': 34.48	2024-02-11T12:19:51-05:00			TRUE		FALSE	8	8	0		2024-02-11T12:21:59.7670000-05:00			