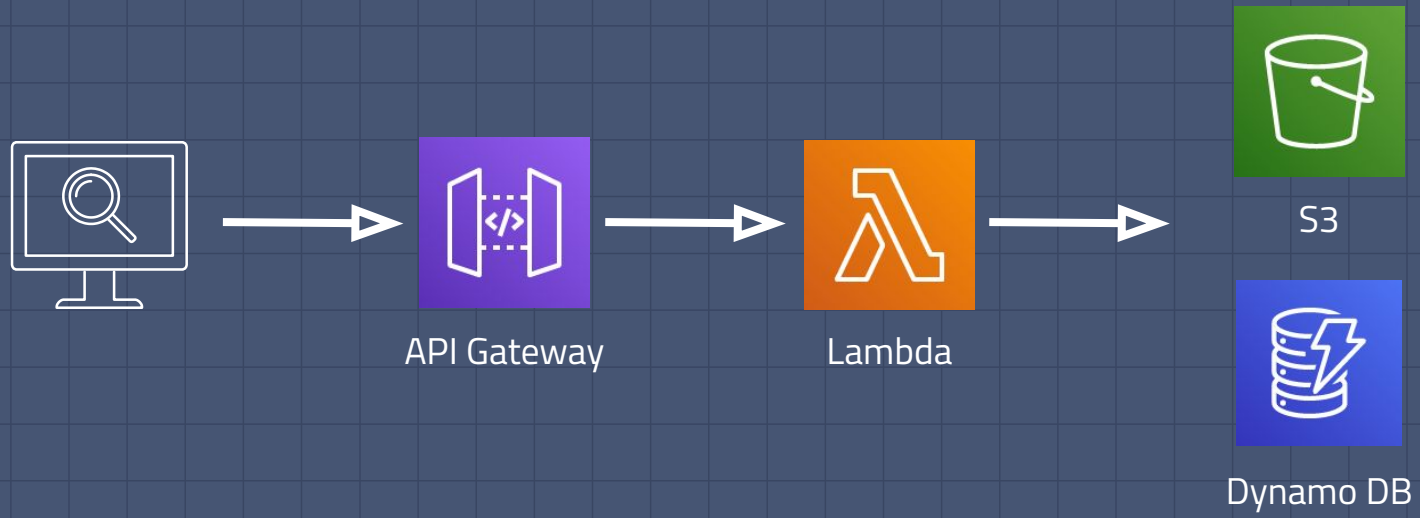
Build a Modern API with AWS

Nathaniel Beckstead



scriptingis.life/glimpseid

$whoami

Nathaniel Beckstead

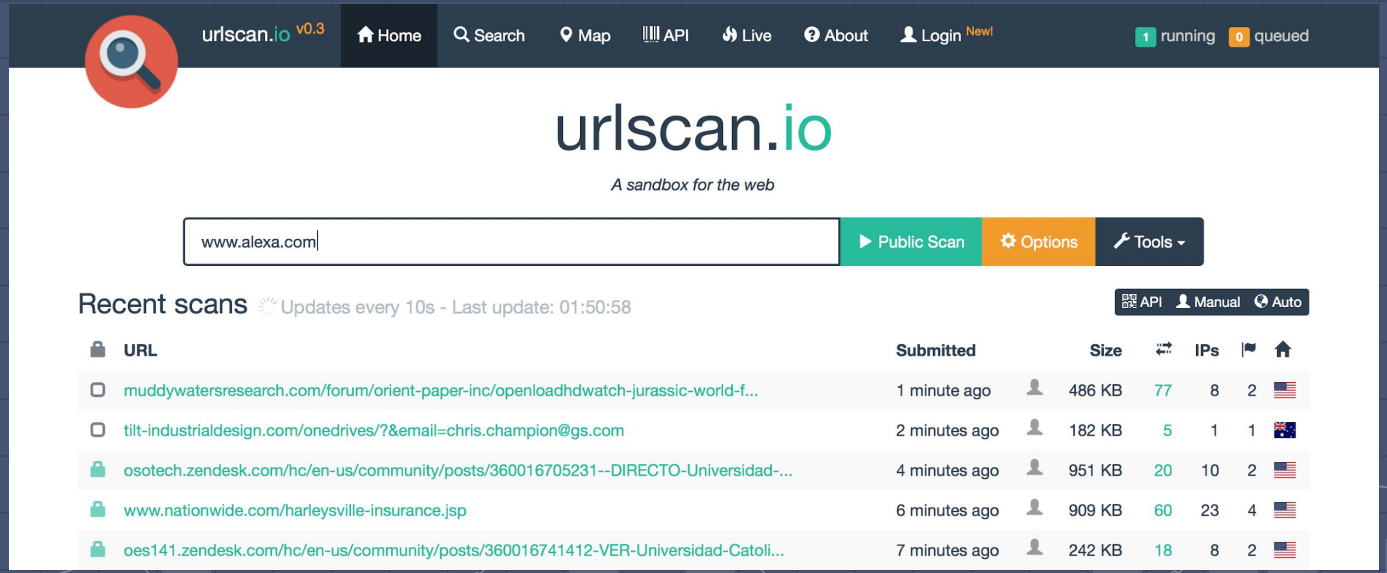
- *CLOUD*
- *DEVOPS*
- *CYBER*

scriptingis.life

Background

- Interned in KeyBank SOC
- Automated parts of phishing response
- URLScan.io
  - Screenshot
  - HTTP Requests
  - IPs/ASNs contacted

Background


urlscan.io v0.3
Home
Search
Map
API
Live
About
Login New!
1 running
0 queued
urlscan.io
A sandbox for the web
www.alexa.com
Public Scan
Options
Tools
Recent scans
Updates every 10s - Last update: 01:50:58
API
Manual
Auto
URL
Submitted
Size
IPs
muddywatersresearch.com/forum/orient-paper-inc/openloadhdwatch-jurassic-world-f...
1 minute ago
486 KB
77
8
2
tilt-industrialdesign.com/onedrives/?&email=chris.champion@gs.com
2 minutes ago
182 KB
5
1
1
osotech.zendesk.com/hc/en-us/community/posts/360016705231--DIRECTO-Universidad-...
4 minutes ago
951 KB
20
10
2
www.nationwide.com/harleysville-insurance.jsp
6 minutes ago
909 KB
60
23
4
oes141.zendesk.com/hc/en-us/community/posts/360016741412-VER-Universidad-Catoli...
7 minutes ago
242 KB
18
8
2

urlscan.io Home Search API Live About Login Sponsored by SecurityTrails 2 running

portailweb.it.cspq.gouv.qc.ca

Lookup Go To Report Rescan

142.213.185.151

Submitted URL: https://Portailweb.it.cspq.gouv.qc.ca
Effective URL: https://portailweb.it.cspq.gouv.qc.ca/logon/LogonPoint/tmindex.html
Submission: On March 27 via automatic, source *certstream-suspicious* (March 27th 2019, 9:37:44 pm)

Summary | HTTP 35 | Links 1 | Behaviour | IoCs | Similar 83 | DOM | Content | API

Summary

This website contacted **2 IPs** in **1 countries** across **1 domains** to perform **35 HTTP transactions**.
The main IP is **142.213.185.151**, located in **Québec, Canada** and belongs to BACI - Bell Canada, CA. The main domain is **portailweb.it.cspq.gouv.qc.ca**.
The TLS certificate was issued by *Entrust Certification Authority - L1K* on July 13th 2018 with a validity of 2 years.

This is the first time this domain was scanned on urlscan.io!

83 structurally similar pages on different IPs, domains and ASNs found Show Scans 83

Domain created: October 23rd 2000, 10:43:54 (UTC)
Domain registrar: A.R.C. Informatique Inc.
Current Google Safe Browsing status: Clean

Domain & IP information

IP/ASNs | IP Detail | (Sub)Domains | Domain Tree | Links | Certificates

	IP Address	AS Autonomous System
1 36	142.213.185.151	11489 (BACI - Bell Canada)
35	2	

Screenshot

Live screenshot Full Image

Detected technologies

- Hammer.js (JavaScript Frameworks) Website
- jQuery (JavaScript Frameworks) Website
- Apache (Web Servers) Website

Stats

Requests	Ad-blocked	Malicious	HTTPS	IPv6
35	0	0	0%	0%

Domains	Subdomains	IPs	Countries	Transfer
1	1	2	1	1,392kB

1,379kB 0

https://urlscan.io/result/26b3b261-23d2-493f-9d36-2be3f4c20c41/dom/

35 HTTP transactions

1 data transactions

Method Protocol	Status	Resource Path		Size x-fer	Time Latency	Type MIME-Type
GET H/1.1	200 OK	tmindex.html /logon/LogonPoint Redirect Chain ▪ https://portailweb.it.cspq.gouv.qc.ca/ ▪ **https://portailweb.it.cspq.gouv.qc.ca/logon/LogonPoint/tmindex.html**	Show response	49 KB 49 KB	552ms 117ms	Document text/html
GET H/1.1	200 OK	wspinner@2x.gif /logon/LogonPoint/receiver/images/common		2 KB 3 KB	666ms 115ms	Image image/gif
GET H/1.1	200 OK	ctxs.large-ui.min.css /logon/LogonPoint/receiver/css		106 KB 107 KB	222ms 113ms	Stylesheet text/css

HTTP Info

- Runs in a Docker container
- Selenium
  - Drive the browser
  - Screenshot
- Browsermob Proxy
  - Record HTTP requests and responses
  - Export to CSV

HTTP Info

becksteadn Update README.md — Latest

.gitignore	Initial commit
Dockerfile	csv and hashes
README.md	Update README.md
getinfo.py	csv and hashes

README.md

HTTP-Info

Homebrewed urlscan.io in a docker container. Screenshot a website and log we

Features

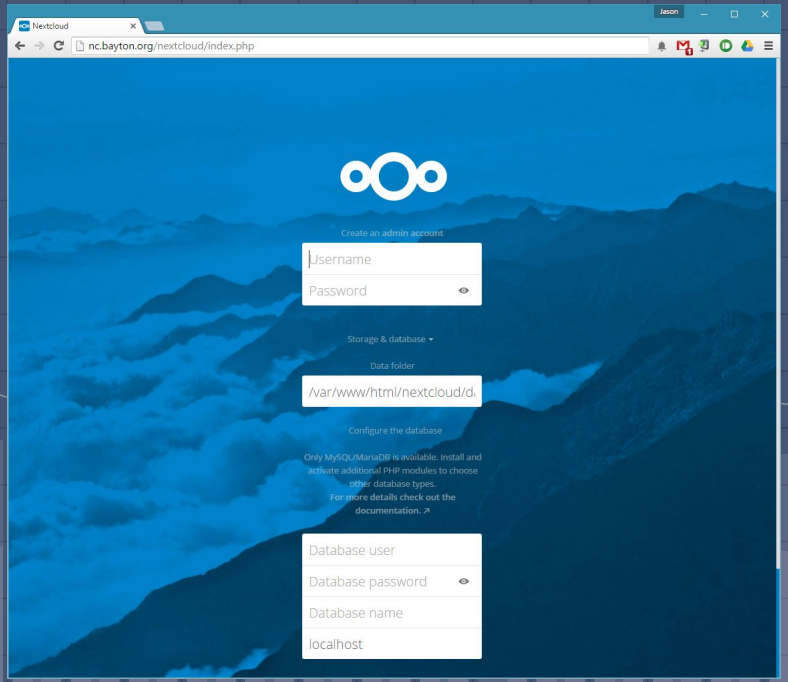
- Screenshots the homepage
- Records URL, method, status code, MIME type, and content size of every H
- Calculates SHA-256 hash of response bodies

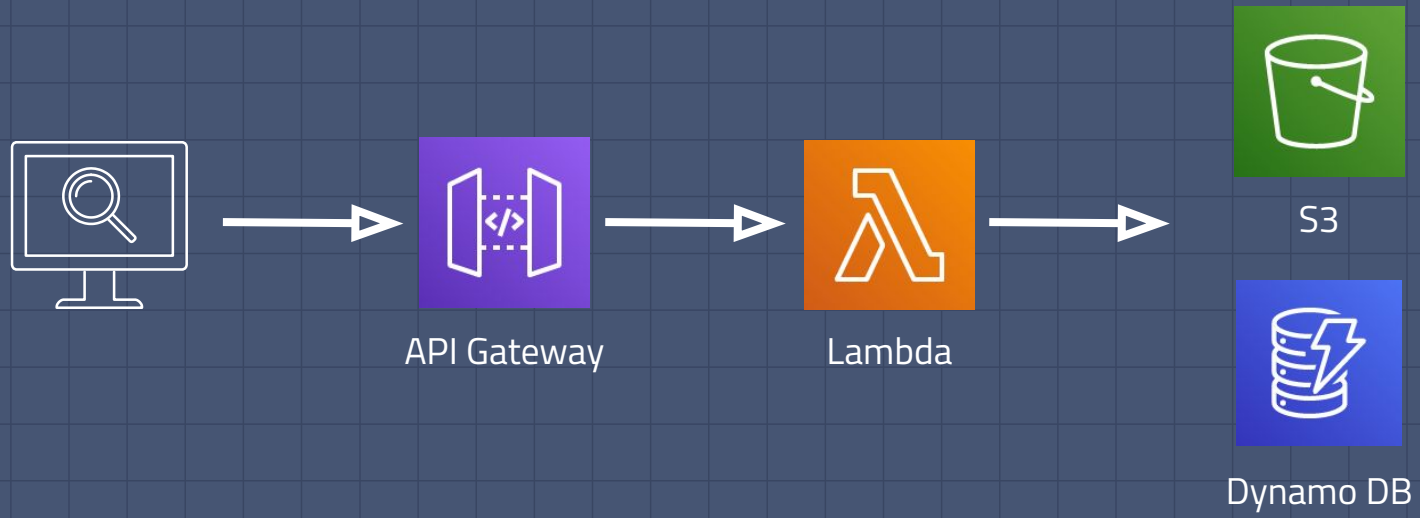
```
Request URL: https://scriptingis.life/
Request Method: GET
Response Status: 200
Reponse Size: 5386
Content Type: text/html; charset=utf-8
SHA256 Sum: eac29b75557e9198eba1e15da4f0d995c62e0a9c62ecc5f156cd5b52c5462bf7


Request URL: https://scriptingis.life/style.css
Request Method: GET
Response Status: 200
Reponse Size: 21725
Content Type: text/css; charset=utf-8
SHA256 Sum: 578fdcff3f6f4b4a0d0cb6535a293197b2e24948927a70051d58b456f3badf9f


Request URL: https://scriptingis.life/images/profile.png
Request Method: GET
Response Status: 200
Reponse Size: 641573
Content Type: image/png
SHA256 Sum: 14d375d76f26f03037f98558934a9276051c4b34e7d12e65530c2a22246531b7
```

Selenium

- Web browser automation primarily designed for testing
- Render a page and interact with elements

```
driver.get('https://www.w3.org/')
for a in driver.find_elements_by_xpath('.//a'):
    print(a.get_attribute('href'))
```




API Gateway
Lambda
S3
Dynamo DB

Lambda

Serverless Computing

1

Lambda

- Serverless Computing
- Only charged for execution time and resources used
- Run when triggered by
  - AWS IoT
  - DynamoDB, S3
  - API Gateway
  - Time

Limitations

- Need to include all resources in upload
  - ZIP - 50MB
  - S3 - 250MB
- Small compute power
  - 128MB - 3GB memory
  - CPU power scales with memory limit
- No root access
  - Limited OS privileges

Lambda Cost

- 1,000,000 requests free
- 400,000 GB-seconds of compute resources free

- 800,000 seconds of runtime with 512MB memory
  - ~40,000 20-second scans

Lambda

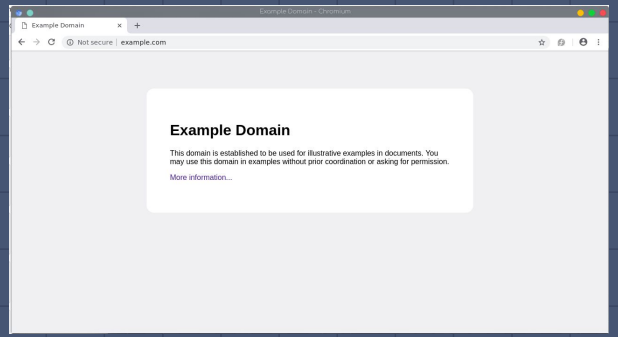
- Python script with Selenium driving Chromium

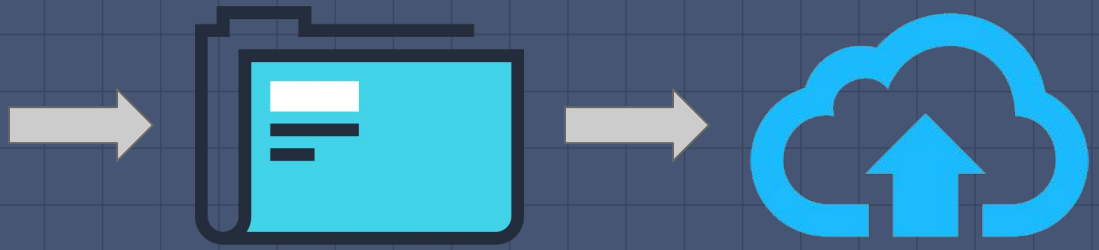
1. Load webpage
2. Screenshot
3. Upload image to S3

21Buttons/pychromeless

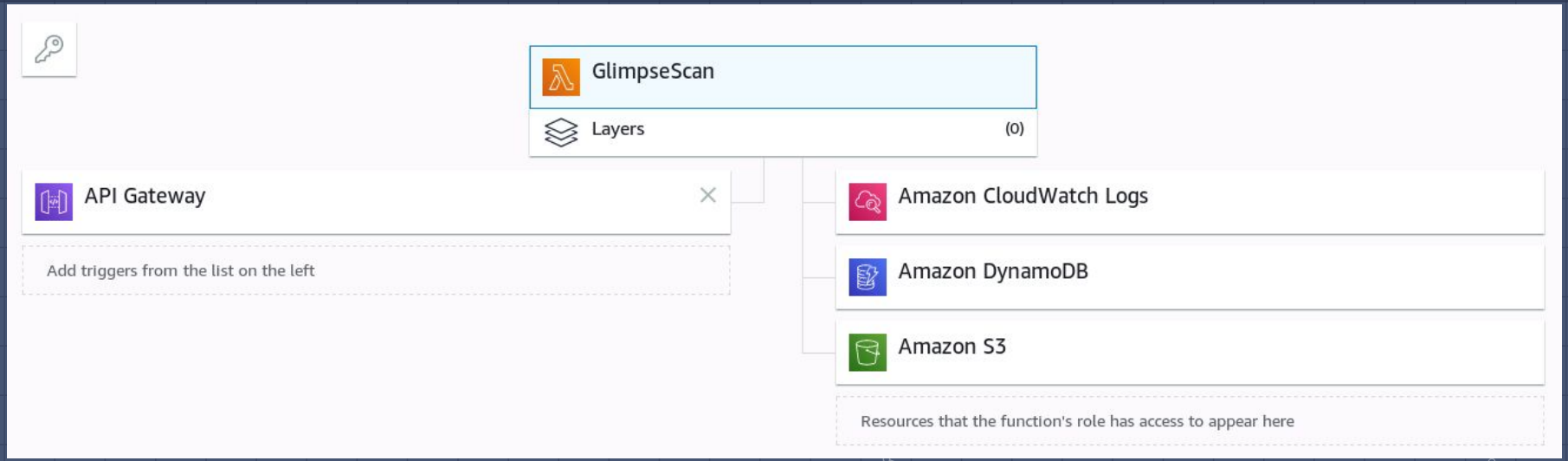
Lambda

- Python script with Selenium driving Chromium





21Buttons/pychromeless


GlimpseScan
Layers
(0)
API Gateway
Add triggers from the list on the left
Amazon CloudWatch Logs
Amazon DynamoDB
Amazon S3
Resources that the function's role has access to appear here

Lambda

- Lambda invokes a function in your code
- Parameters passed as a dictionary

Handler Info

lambda_function.lambda_handler

```
def lambda_handler(event, context):
```

Lambda Deployment

- Makefile
- AWS CLI

```
pack: clean fetch-dependencies
    mkdir build
    cp -r src build/.
    cp -r bin build/.
    cp -r lib build/.
    pip install -r requirements.txt -t build/lib/.
    cd build; zip -9qr build.zip .
    cp build/build.zip .
    rm -rf build

deploy: pack
    aws s3 cp ./build.zip s3://${S3_BUCKET}/${S3_KEY} --profile ${AWS_USER}
    aws lambda update-function-code --function-name ${FUNCTION_NAME} --s3-bucket ${S3_BUCKET} --s3-key ${S3_KEY} --profile ${AWS_USER}
```

Storage

Simple Storage Service (S3) and DynamoDB

2

S3

- Key-Value Storage
- Host publicly accessible images
- Uploading done through **boto** Python module

```
conn = S3Connection(S3_KEY_ID, S3_SECRET_KEY)
bucket = conn.get_bucket('glimpsefiles')
key = Key(bucket, 'screenshots/' + screenshot_filename)
key.set_contents_from_filename(screenshot_path)
```

S3 Cost

- Storage
  - First 50TB - $0.023 per GB
- Access
  - PUT
    - Data added $0.002 per GB
    - $0.005 per 1,000 requests
  - GET
    - Data returned $0.0007 per GB
    - $0.0004 per 1,000 requests

DynamoDB

- NoSQL Database
  - No set structure = No normalizing!
- No setup, maintenance, or clustering

```
db = DynamoDB(DB_TABLE)

exists = False
db_data = db.get({'urlhash': url_hash})
```

- Cost
  - $0.25 per GB of storage
  - $1.25 per million writes
  - $0.25 per million reads
  - Free Tier - 25GB storage, 2.5 million reads, 1GB data transfer out

API Gateway

REST API Development and Management

3

API Gateway

- Visual API development
- Proxy for other AWS services
  - DynamoDB queries
- Convert between HTTP requests and Lambda execution

API Gateway

/screenshot - GET - Method Execution

TEST

Client

Method Request

Auth: NONE

ARN: [redacted]/*/GET/screenshot

Query Strings: url

Integration Request

Type: LAMBDA

Query Strings: url

Region: us-east-1

Lambda glimpseScreenshot

Method Response

HTTP Status: 200

Models: application/json => Empty

Integration Response

HTTP status pattern: -

Output passthrough: Yes

API Gateway Cost

- $3.50 per million API calls
- Caching
  - 0.5GB for $0.020 per hour
    - $15 per month

Website

Frontend is hard :(

4

Github Pages

- Host a static site for free

1. Make a new repository
2. Add an index.html
3. Settings -> GitHub Pages -> master branch
4. Struggle with CSS
5. Profit

Conclusions

DevOps Borat
@DEVOPS_BORAT

To make error is human. To propagate error to all server in automatic way is #devops.

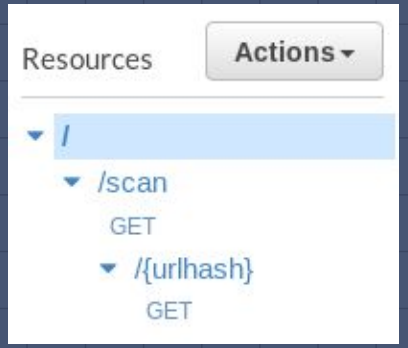
2:55 PM · Feb 26, 2011 · Mobile Web

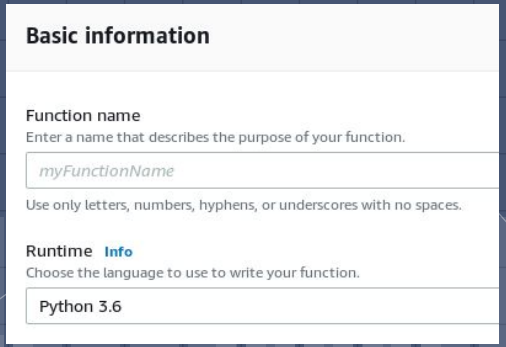
Why Use The Cloud?

- Easy
  - Heavy lifting done by AWS
  - Graphical interfaces for everything
  - Logging and dashboards built in
- Cheap
  - Free tiers
  - Charge by the millions
- Scalable
  - Duplicate and automate

Easy

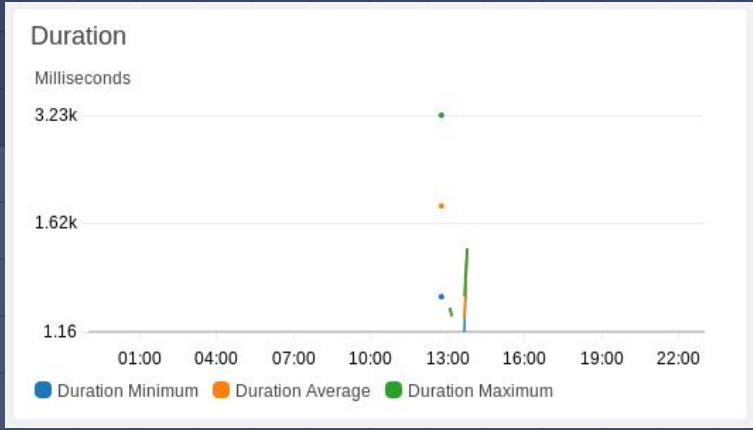

Create DynamoDB table
DynamoDB is a schema-less database that only requires a table name and primary key. The ta
data, and sort data within each partition.
Table name*
Primary key*
Partition key
String
Add sort key


Resources
Actions
/
/scan
GET
/{urlhash}
GET


Basic information
Function name
Enter a name that describes the purpose of your function.
myFunctionName
Use only letters, numbers, hyphens, or underscores with no spaces.
Runtime
Info
Choose the language to use to write your function.
Python 3.6

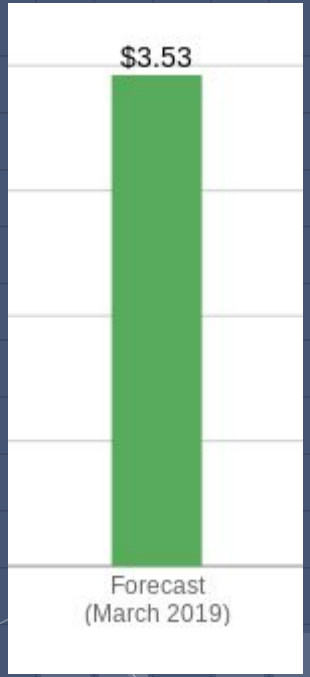
Dashboards


Latency
Get latency (Milliseconds)
25
20
15
10
5
0
3/27 00:00
3/27 08:00
3/27 16:00
Put latency (Milliseconds)
30
20
10
0
3/27 00:00
3/27 08:00
3/27 16:00


Duration
Milliseconds
3.23k
1.62k
1.16
01:00
04:00
07:00
10:00
13:00
16:00
19:00
22:00
Duration Minimum
Duration Average
Duration Maximum


Error count and success rate (%)
Count
No unit
2
1
0
100
66.7
33.3
01:00
04:00
07:00
10:00
13:00
16:00
19:00
22:00
Errors
Success rate (%)

Cheap



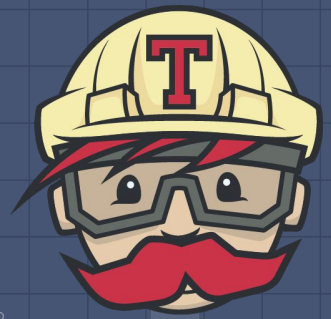
Top Free Tier Services by Usage

View all

Service	Free Tier usage limit	Month-to-date usage
AWS Key Management Service	20,000 free requests per month for AWS Key Management Service	0.42% (84.00/20,000 Requests)
AWS Lambda	400,000 seconds of compute time per month for AWS Lambda	0.13% (533.39/400,000 seconds)
AWS Lambda	1,000,000 free requests per month for AWS Lambda	0.02% (242.00/1,000,000 Requests)
AmazonCloudWatch	1,000,000 API requests for Amazon Cloudwatch	0.01% (58.00/1,000,000 Requests)
AmazonCloudWatch	5 GB of Log Data Ingestion for Amazon Cloudwatch	0.00% (0.00/5 GB)

Next Up

- Continuous Integration
- Network Activity Logging
- Support Multiple Regions, User-Agents
- ...



Questions?

scriptingis.life/glimpseid

Resources

- Boto 3 Documentation
  - Dynamo DB
  - S3
- AWS Blog - Project Ideas!
- Open Guide to AWS
- Using Python on Lambda
- More on urlscan.io
- /r/aws