From idea to product with AWS & Python

Lessons learned from a two months hobby project
AWS User Group 23rd of April

Overview

1. Part 1: Commit Stats Product Demo
2. Part 2: Behind the scenes **discussions**
3. Questions?
4. Thank You

Part 1: Commit Stats Product Demo

About Me 1

- Geography
  - Born and raised in Kristiansand, Norway
  - Family in Stockholm, Sweden and siblings in Oslo, Norway
  - 3rd home in Ko Lanta (Kohub), Thailand
  - ~4 years in Gran Canaria
  - Resident and living with Maria in Arinaga
- Activities: Calisthenics, Weightlifting, sports++
- Career:
  - Rocketfarm (consulting, small team)
  - wheel.me (startup)
  - smidige.com (contractor)
  - MongoDB (enterprise)

We are here

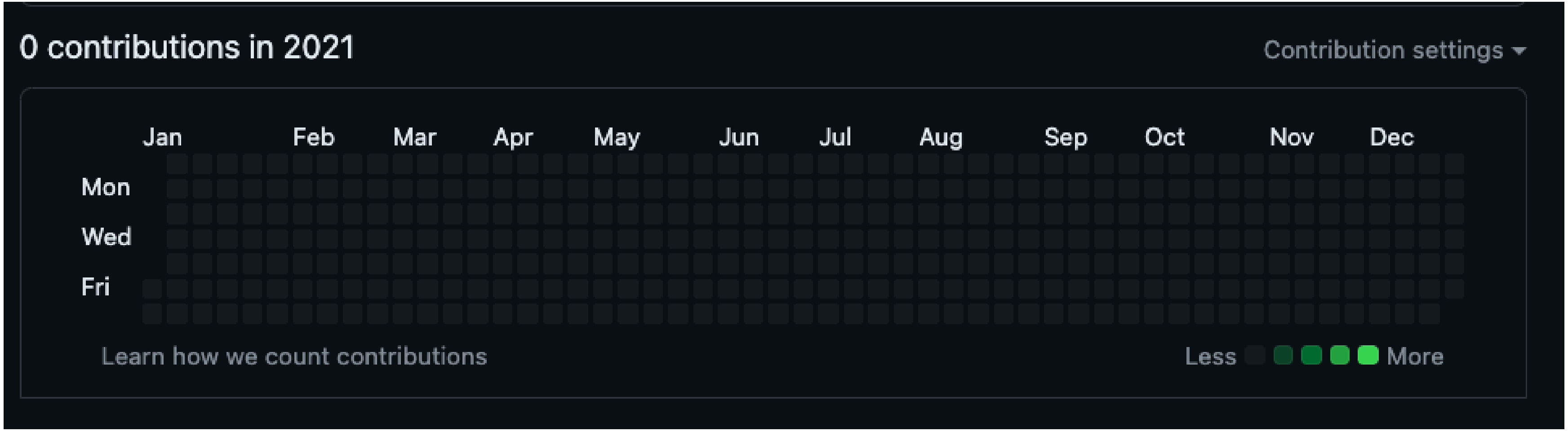
Playa de Sardina del Norte
Gáldar
El Puertillo
Playa de Las Canteras (Las Palmas de Gran Canaria)
Las Palmas de Gran Canaria
GC-2
GC-75
GC-3
Arucas
Barranco de Azuaje
GC-20
GC-43
Agaete
Jardín Botánico Viera y Clavijo
GC-1
Parque Rural de Doramas
GC-200
GC-220
GC-70
GC-21
Playa de la Laja
Tamadaba
Teror
Valleseco
Parque Natural Tamadaba
GC-220
GC-21
Real Club de Golf De Las Palmas
GC-800
Valle de Jinamar
GC-42
Artenara
Vega de San Mateo
GC-80
GC-200
Cruz de Tejeda
Tejeda
GC-810
La Garita
GC-210
GC-600
Valsequillo de Gran Canaria
GC-41
Telde
GC-1
La Aldea de San Nicolás de Tolentino
GC-130
Monumento natural del Roque Nublo
Pico de las Nieves
GC-130
GC-140
Tufia
GC-120
Inagua
GC-120
GC-100
GC-200
Gran Canaria
GC-65
GC-550
Ingenio
Playa del Burrero
Agüimes
GC-1
Mogán
Fataga
Parque Natural de Pilancones
Playa del Cabrón
Vecindario
Arinaga
Palmitos Park
ya de Mogan
Pozo Izquierdo
Castillo del Romeral
Puerto Rico de Gran Canaria
Playa de San Agustín
Patalavaca
Arguineguín
Maspalomas
Pasito Blanco

About Me 2

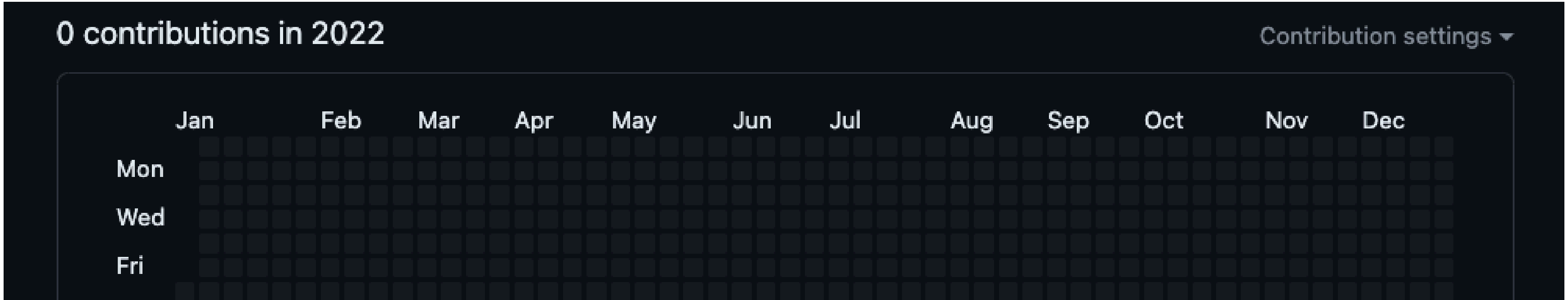
- Tech choices:
  - Python (since 3.6)
  - Terraform (~ 2 years)
  - Pycharm -> VsCode (Not worth doing the VsCode yet?)
  - Golang (Last few months)
  - MongoDB, PostgreSQL, InfluxDB, and DynamoDB
  - Kubernetess+Istio+Helm (~ 4 years)
  - Markdown <3
- Influenced by:
  - Cal Newport (productivity & life philosophy)
  - Podcasts (Latest tech news) & Books (Going deeper)
  - Tim Ferriss+Lex Friedman+ Guy Raz (Exposure to different thinkers)
  - Health: Crossfit+ATG (exercise)
- More info at my Github page

Problem Statement

How do you show you can code if you have only worked in private git repos?



0 contributions in 2021



0 contributions in 2022

Project Background (also found at the botton of project landing page)

I was applying for developer jobs and I couldn’t show my coding experience. The code repositories I had been working on were in different git providers (Gitlab & Azure DevOps) and the code was **private**.

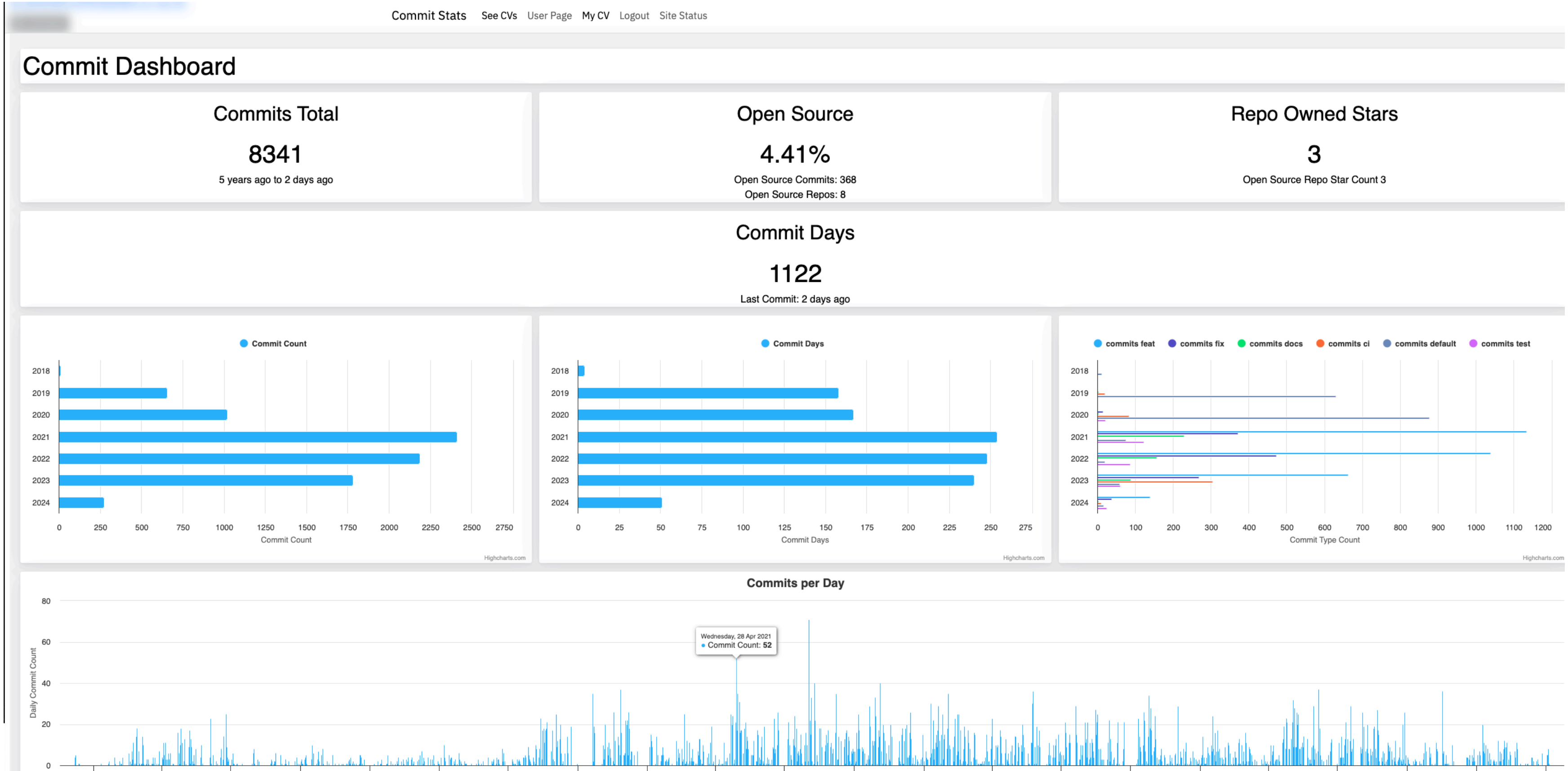
As I am a bit of a data nerd 🤓 I like to track different metrics. Not only commits, PRs, but also LoC (Lines of Code), open source contribution, 3rd-party packages, etc. Therefore, the two main goals of the project are:

> 1. **Support creating a "CV" page of your commit stats that you can use when applying for jobs**
> 2. **(Future) Support weekly feedback on your code and reflect on your developer journey**

I hope you will find it useful and fun 😎 Happy coding!

Product Demo 1

- "CV/Dashboard" of your commits across languages



Commit Dashboard

Commits Total

8341

5 years ago to 2 days ago

Open Source

4.41%

Open Source Commits: 368

Open Source Repos: 8

Repo Owned Stars

3

Open Source Repo Star Count 3

Commit Days

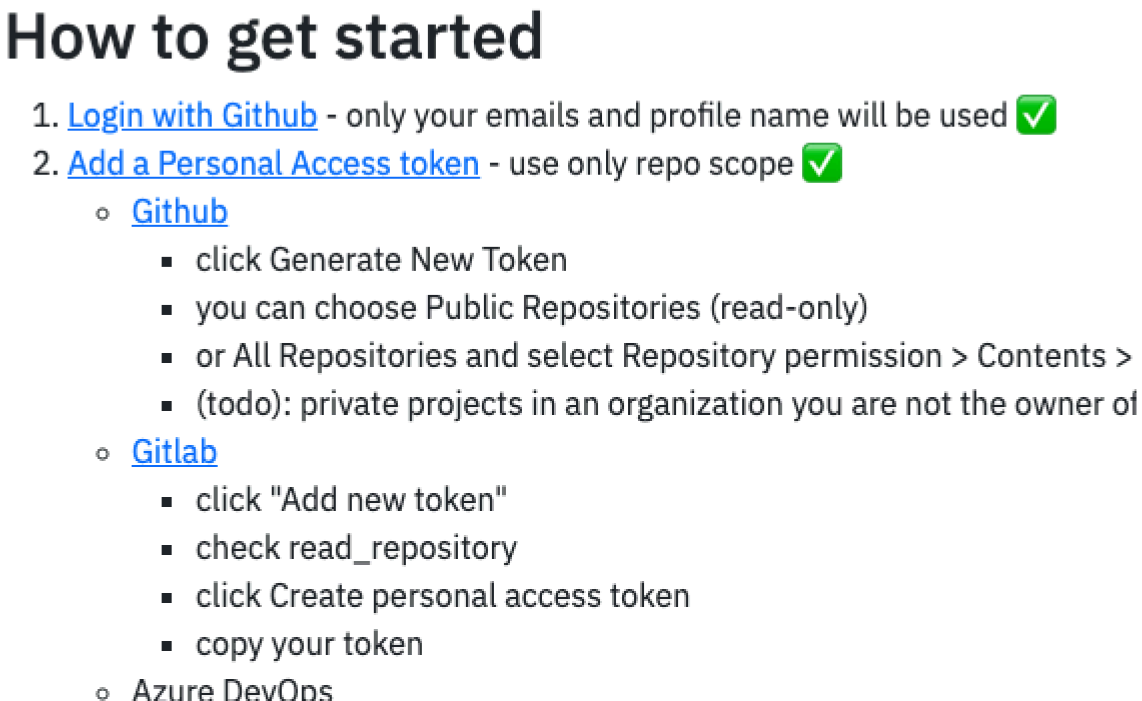
1122

Last Commit: 2 days ago

Product Demo 2

- How to collect the commits:

1. Login with Github
2. Add your token

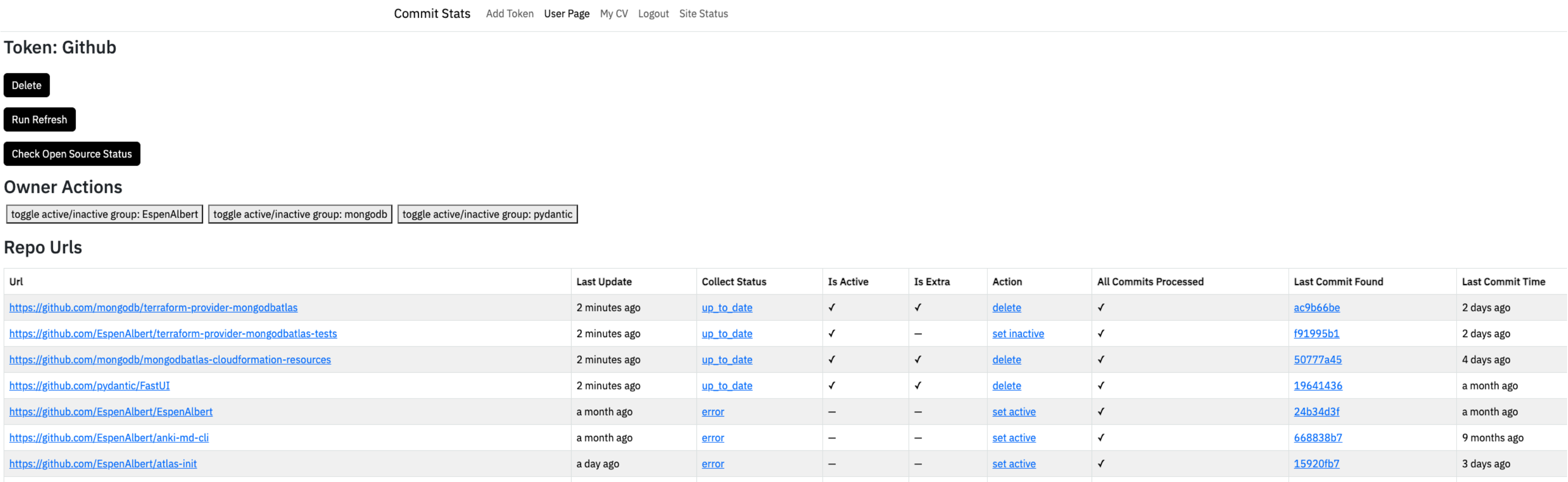


How to get started

1. Login with Github - only your emails and profile name will be used ✅
2. Add a Personal Access token - use only repo scope ✅
   - Github
     - click Generate New Token
     - you can choose Public Repositories (read-only)
     - or All Repositories and select Repository permission > Contents >
     - (todo): private projects in an organization you are not the owner of
   - Gitlab
     - click "Add new token"
     - check read_repository
     - click Create personal access token
     - copy your token
   - Azure DevOps

Product Demo 3

- Select your repos
- Full Collection Status & Open Source Repos


Commit Stats
Add Token
User Page
My CV
Logout
Site Status

Token: Github

Delete
Run Refresh
Check Open Source Status

Owner Actions

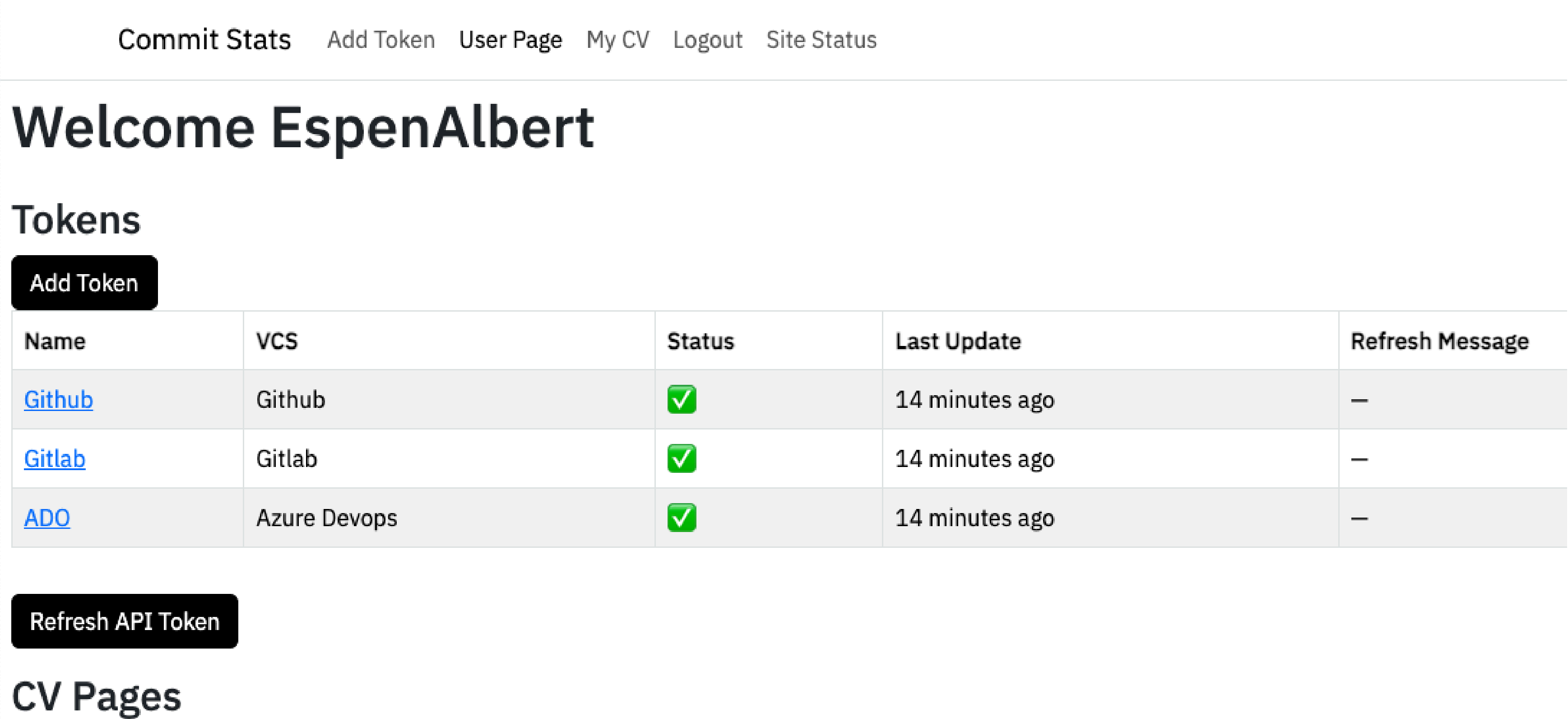
toggle active/inactive group: EspenAlbert
toggle active/inactive group: mongodb
toggle active/inactive group: pydantic

Repo Urls

Url	Last Update	Collect Status	Is Active	Is Extra	Action	All Commits Processed	Last Commit Found	Last Commit Time
https://github.com/mongodb/terraform-provider-mongodbatlas	2 minutes ago	up_to_date	✓	✓	delete	✓	ac9b66be	2 days ago
https://github.com/EspenAlbert/terraform-provider-mongodbatlas-tests	2 minutes ago	up_to_date	✓	–	set inactive	✓	f91995b1	2 days ago
https://github.com/mongodb/mongodbatlas-cloudformation-resources	2 minutes ago	up_to_date	✓	✓	delete	✓	50777a45	4 days ago
https://github.com/pydantic/FastUI	2 minutes ago	up_to_date	✓	✓	delete	✓	19641436	a month ago
https://github.com/EspenAlbert/EspenAlbert	a month ago	error	–	–	set active	✓	24b34d3f	a month ago
https://github.com/EspenAlbert/anki-md-cli	a month ago	error	–	–	set active	✓	668838b7	9 months ago
https://github.com/EspenAlbert/atlas-init	a day ago	error	–	–	set active	✓	15920fb7	3 days ago

Product Demo 4

- Wait...
- Updated every 15 min



Welcome EspenAlbert

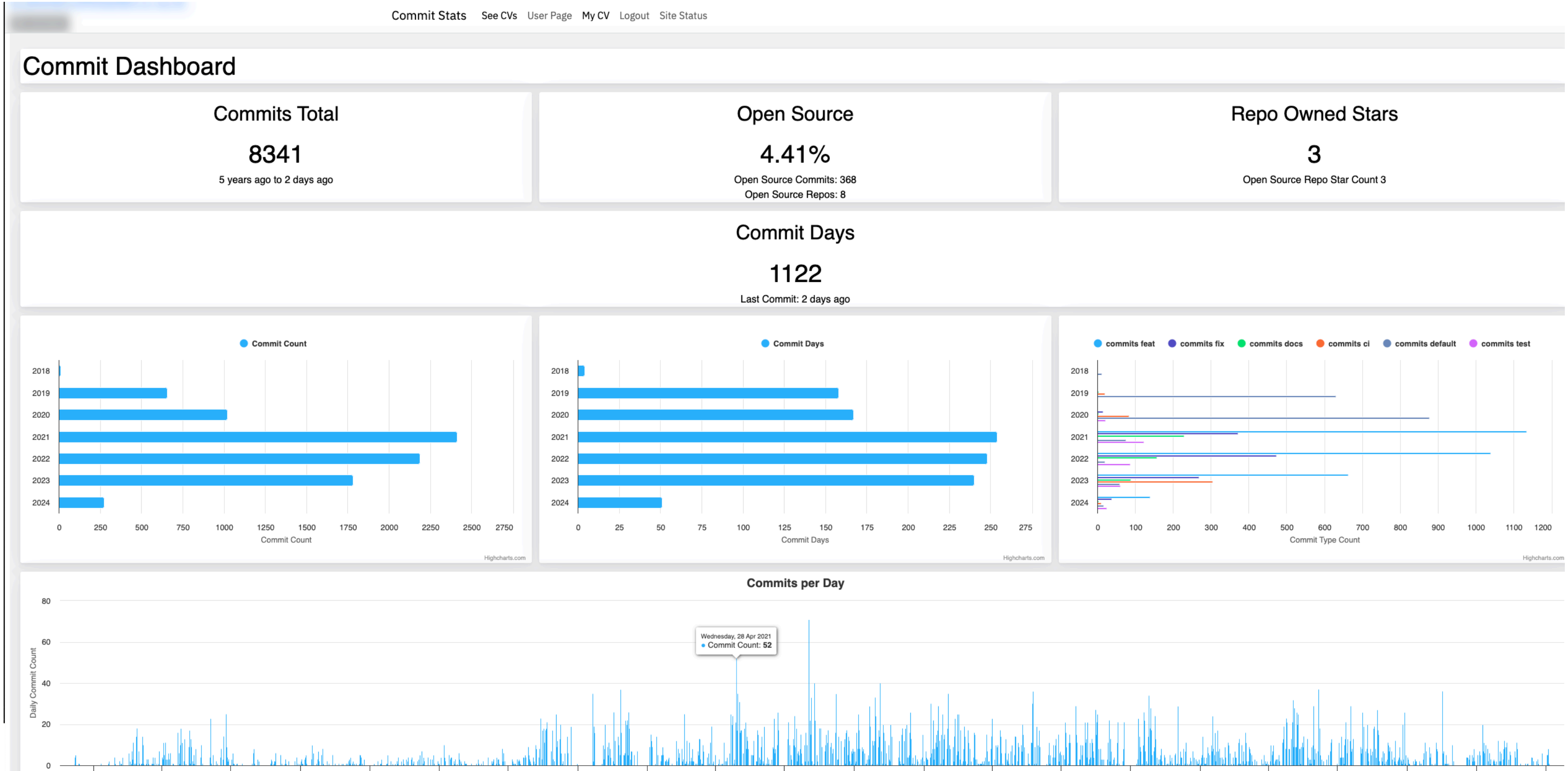
Tokens

Name	VCS	Status	Last Update	Refresh Message
Github	Github	✅	14 minutes ago	—
Gitlab	Gitlab	✅	14 minutes ago	—
ADO	Azure Devops	✅	14 minutes ago	—

CV Pages

Product Demo 5

- Back to the beginning



Commit Dashboard

Commits Total

8341

5 years ago to 2 days ago

Open Source

4.41%

Open Source Commits: 368
Open Source Repos: 8

Repo Owned Stars

3

Open Source Repo Star Count 3

Commit Days

1122

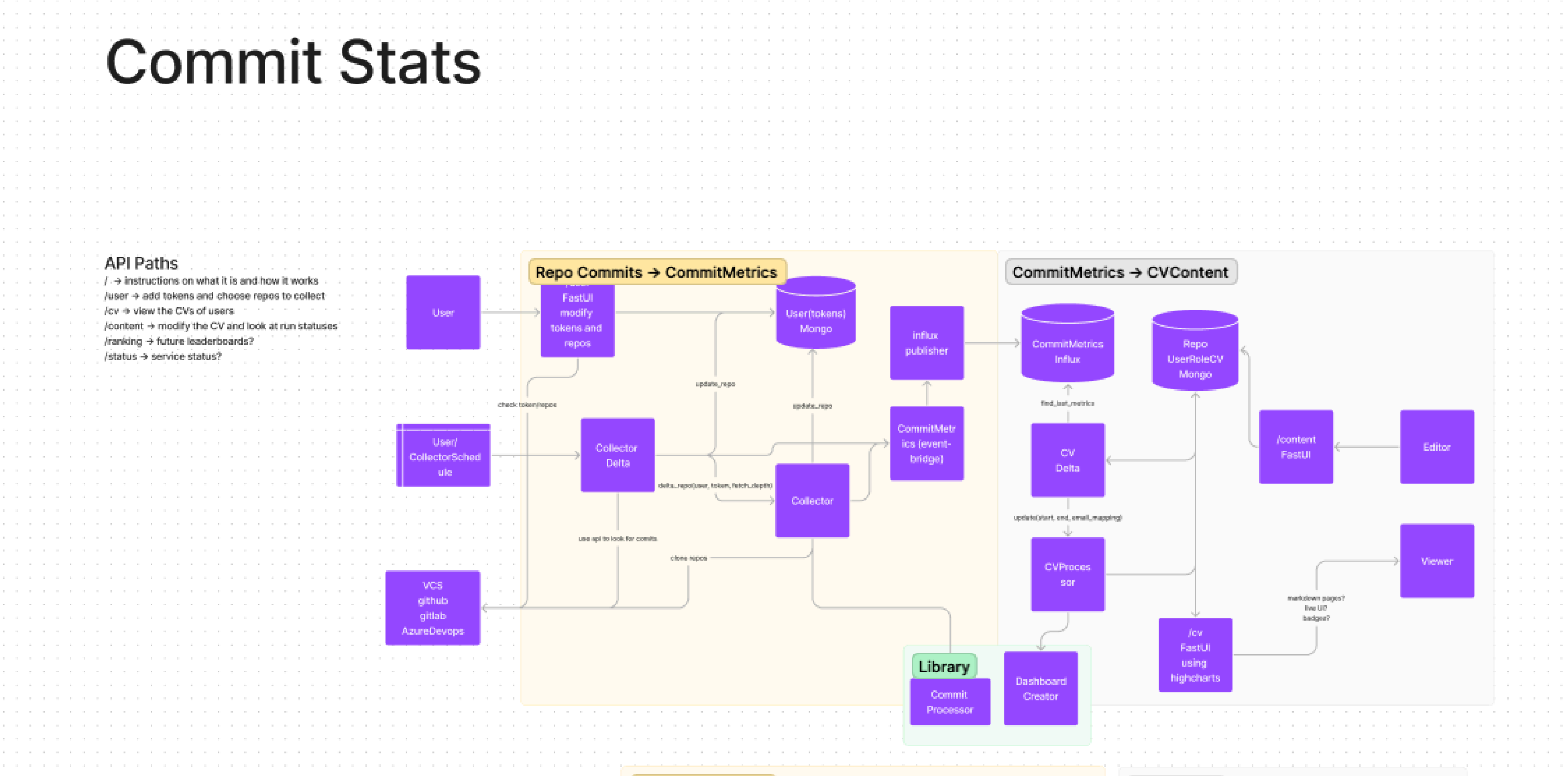
Last Commit: 2 days ago

Part 2: Behind the scenes **discussions**

1. Architecture 😱
2. Tech stack decision
3. How to deploy the artifacts? (Lambda/ECS/EC2?)
4. How do you do the routing? (Staticfiles, CDN, APIs, CORS, and SPA)
5. SaaS products?
6. AWS Organization?
7. Cost of the solution?
8. Multiple environments? CI workflows?
9. Quality checks & Local testing
10. Future of the Project

Architecture 😱

- Figma



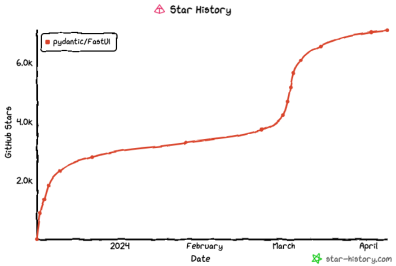
Commit Stats

API Paths

/ → instructions on what it is and how it works
/user → add tokens and choose repos to collect
/cv → view the CVs of users
/content → modify the CV and look at run statuses
/ranking → future leaderboards?
/status → service status?

Tech stack decision

- What stack would you choose?
- I’m a Python fan and inspired by Samuel Colvin starting the pydantic company
  - Pydantic Services Inc. emerges from stealth today with $4.7 million in seed funding led by Sequoia
- Wanted to use FastUI (FastAPI + React)
  - 7.1k stars in a few months
- Highcharts



Artifacts Produced

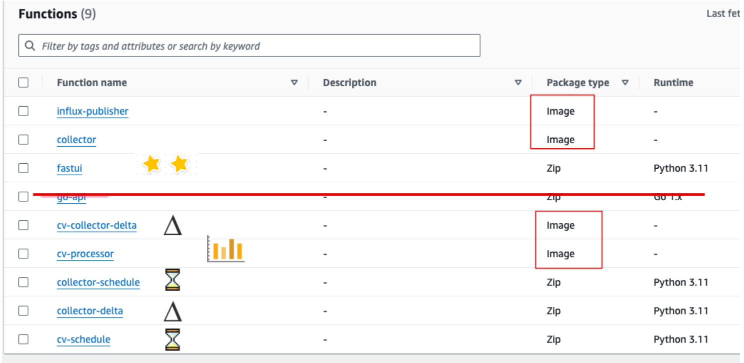
- classic static files
  - `index.css`
  - `index.html`
  - `index.js`
  - +++
- zip files for the lambda functions
- or a docker image

Tech candidates

- Plotly Dash
- Your Choice?

How to deploy the artifacts? (Lambda/ECS/EC2?)

- Artifacts reminder
  - Static files (index.css|html|js ++)
  - aws zip files
  - docker images?
- What would you choose and why?
- I chose Lambda



Functions (9)

Filter by tags and attributes or search by keyword

Last fet

□	Function name	Description	Package type	Runtime
□	influx-publisher	-	Image	-
□	collector	-	Image	-
□	fastui ⭐ ⭐	-	Zip	Python 3.11
□	~~go-api~~	~~-~~	~~Zip~~	~~Go 1.x~~
□	cv-collector-delta Δ	-	Image	-
□	cv-processor 📊	-	Image	-
□	collector-schedule ⌛	-	Zip	Python 3.11
□	collector-delta Δ	-	Zip	Python 3.11
□	cv-schedule ⌛	-	Zip	Python 3.11

- Why images? 250MB limit on Lambda functions
  - Python packages 😱

```
144M python3.11/site-packages//plotly
89M python3.11/site-packages//pyarrow
81M python3.11/site-packages//botocore
71M python3.11/site-packages//pandas
60M python3.11/site-packages//numpy
42M python3.11/site-packages//dash
```

  - ECR Lifecycle rules
- How to choose memory size?
- How to trigger lambdas?
  - EventBridge (CloudWatch Events)
  - Supports schedule: `cron(0/15 ** *?*)` or `rate(1 minute)`
  - Specific event publishing by using `detail-type`, see `events.tf`
- Lambda cold start problem
  - Provisioned concurrency to the rescue? redit thread
  - 10 × 0.000004167 × 0.5 × 86400 = $1.80/day, or $54/month

How do you do the routing? (Staticfiles, CDN, APIs, CORS, and SPA)

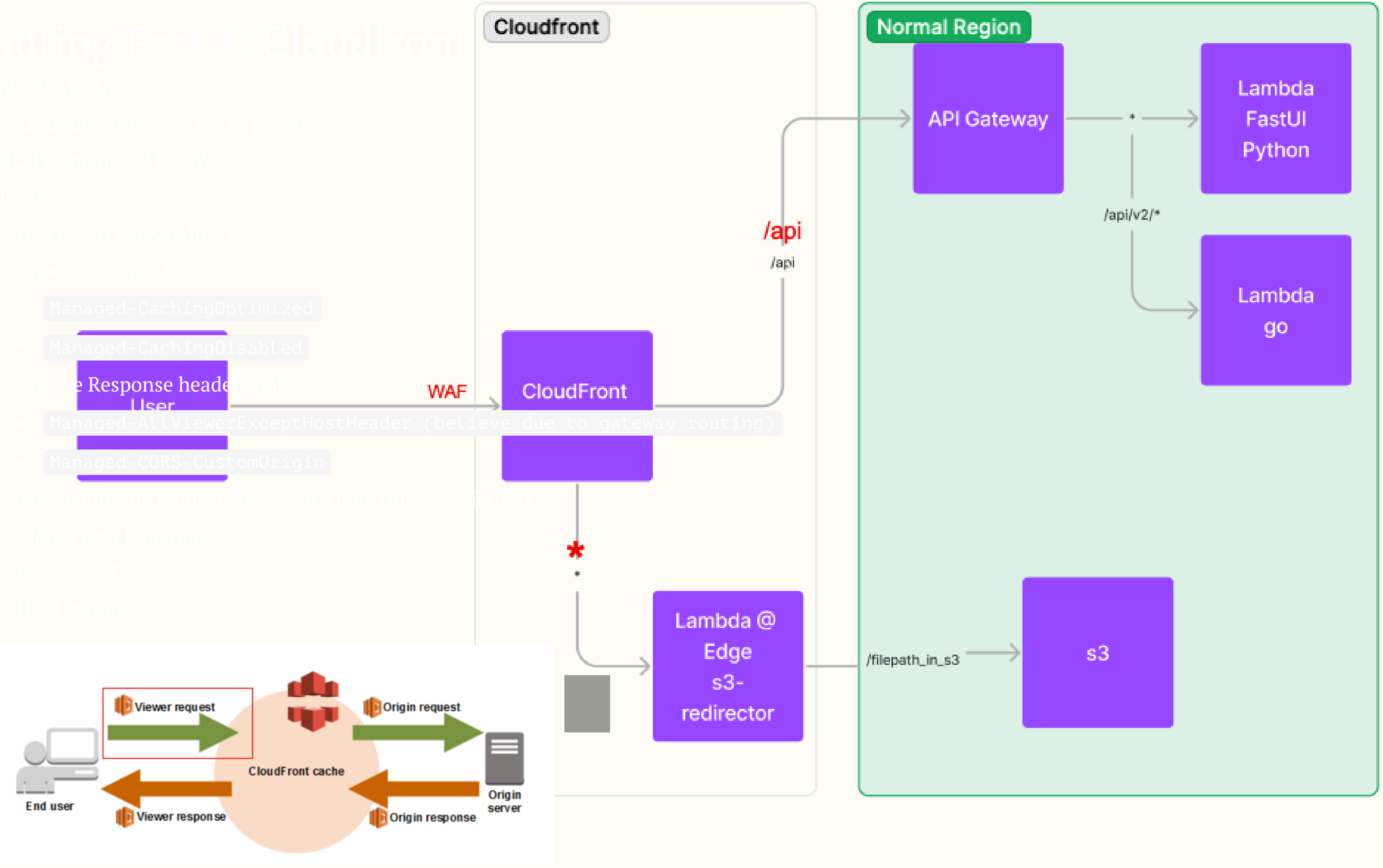
- What AWS Services?

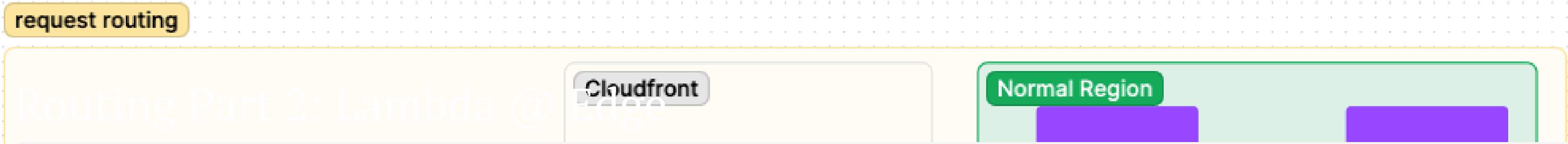
1. Routing Part 1: Cloudfront
2. Routing Part 2: Lambda @ Edge
3. Routing Part 3: Lambda @ Edge
4. Routing Part 4: API Gateway
5. Routing Part 5: Lessons Learned

request routing

Routing Part 1: Cloudfront

- AWS WAF? No
  - Minimum price of 5$ + 1$ per rule
- SSL/TLS managed by AWS
- Origins
  - HTTP -> HTTPS redirect
  - Choose caching behavior
    - `Managed-CachingOptimized`
    - `Managed-CachingDisabled`
  - Choose Response headers policy
    - `Managed-AllViewerExceptHostHeader (believe due to gateway routing)`
    - `Managed-CORS-CustomOrigin`
- Can hook into the request stages with Function associations
  - CloudFront Functions?
  - Lambda@Edge
  - Hook points





Routing Part 2: Lambda @ Edge

```python
ALL_S3_DOCUMENTS = set(['CodeLazy.js', 'CodeLazy.js.map', 'MarkdownLazy.js',
'MarkdownLazy.js.map', 'index.css', 'index.html',
'index.js', 'index.js.map'])

def lambda_handler(event, context) -> dict | None:
    """based on example: https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/lambda-examples.html#la
    request: dict = event["Records"][0]["cf"]["request"]
    path = request["uri"]
    print(f"path is: {path}")
    s3_path = path.lstrip("/")
    if s3_path in ALL_S3_DOCUMENTS:
        print(f"no redirection {s3_path} exist")
        return request
    print("returning html directly")
    return {
        "status": "200",
        "statusDescription": "OK",
        "body": _html,
        "bodyEncoding": "text",
        "headers": {
            "content-type": [
                {
                    "value": "text/html;charset=UTF-8",
                }
            ]
        },
```

Routing Part 3: Lambda @ Edge

- inside of python script: `_html=`

```
<!doctype html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>CommitStats Landing Page</title>
    <script type="module" crossorigin src="/index.js"></script>
    <link rel="stylesheet" crossorigin href="/index.css">
  </head>
  <body>
    <div id="root" class="highcharts-light"></div>
  </body>
</html>
```

Routing Part 4: API Gateway

- AWS API Gateway
- Anyone experience with multiple stages?

Routing Part 5: Lessons Learned

- Tip: Avoid splitting up APIs if you can
    - Fell on my face quite hard
        - CORS (Cross Origin Resource Sharing)
            - Where to add the headers?
            - Do one thing at a time
- Tip: Invalidating cache
    - Anyone knows how to do it with Terraform?

SaaS products?

Database?

1. MongoDB Atlas with Free Forever (up to 512MB)
    1. Use AWS KMS to encrypt the tokens
2. InfluxDB Cloud Serverless ($250 in free credits for 90 days)
3. Alternatives
    1. Elasticsearch
    2. RDS (PostgreSQL)
    3. DynamoDB
    4. DocumentDB?

Message Queue alternatives?

1. EventBridge
2. Kafka (Confluence)
3. Redis?
4. RabbitMQ

AWS Organization?

- Would you use than more than one AWS account?
- Root account
  - Route53
  - IAM Config
  - ECR
- `{Project Name}` Account: CommitStats
  - Billing separate
  - Need access to route53 to add DNS records (Terraform and `assume_role` to the rescue)
- Terraform Basic Needs
  - S3 with replication
  - KMS (Optional -> can be costly)

Cost of the solution?

- Which AWS service do you think is the most expensive?
- KMS
  - Stopped encrypting terraform state to save money
- Cost Explorer
- InfluxDB, most costly for now ($2.49 for February)
  - Data out $0.17 for 1.88GB
  - Data In $1.59 for 635.851 MB
  - Query count $0.64 for 5342 queries (0.012 per 100 query executions)
  - Storage:$0.09 for 43.461 GB-hr

Multiple environments? CI workflows?

- What if you need to host stage/QA environments?
  - Recreate the full terraform with a new environment or re-use infrastructure
- lambda versioning? aliases?
- Time for a live deployment?

Quality checks & Local testing

- Time for a demo?

Future of the Project

- Alternative to InfluxDB for hosting
- Improved on-boarding experience
- More in the `Future ideas` section

Questions?

- Observability tools?
  - Sentry?
  - OTEL - Open Telemetry
  - ELK (Elasticsearch, Logstash, Kibana)
  - Cloudwatch

Thank You

- AWS User Group organizers
  - Andrey
  - Antonio
  - Andras
- Leo for hosting us
- Everyone for joining
- slidev