SPRINT 4

Requirements:

Data vendor→ SPOTIFY

Extract → Spotify API (Spotify Developers, user access) loads data to the pipeline.

Transform → pipeline based on Airflow 2 with 6 DAGs (find location user, find the exact song played, find the local time respect Greenwich time, users' interaction while song is played, type of browser used or the Spotify App, temperature of the location while playing song from BBC weather channel).

Load → Kibana. But before use Elasticsearch to refine all the transformed and processed information in the pipeline. Use database to store all the output from the pipeline and use PostgreSQL to retrieve needed information.

Machine learning and deep learning techniques whit transformed data could be applied to find patterns and behaviours from users. Discovery mode of the project!

Customer and data consumer → Spotify to offer related songs to user. Roehampton university research tea to find if weather affects the desire of listening music, and the type of music if previous answer is positive. Expectations from machine learning and deep learning about users; and this can be useful for authorities like government, doctors, and psychologists. Tableau or Power BI is the destination of all findings and discoveries from pre-process data to show to Nontechnical users (children, NO-STEM users, old people for example).

Advertisement and marketing business can buy the output and outcome from the Load and Transform stages to offer relevant advertisement to the listener, even if the listener is in the same building located.

All the project implemented on Google Cloud with 50 USD budget. Try to use free resources and similar free datasets or data engineering projects.

Management:

6 DAGs in python programming language. Please refer to GitHub.

Teamwork kanban board on GitHub link (readme file).

Outline of the project overall:

Overall, the project has a length of 5 months (January 2023 to May 2023). Agile style is used in each step and in each progress. The product owner is the same as Scrum Master and Team Leader. Kanban board is maintained by the same person.

Some useful references for the DATA ENGINEERING project

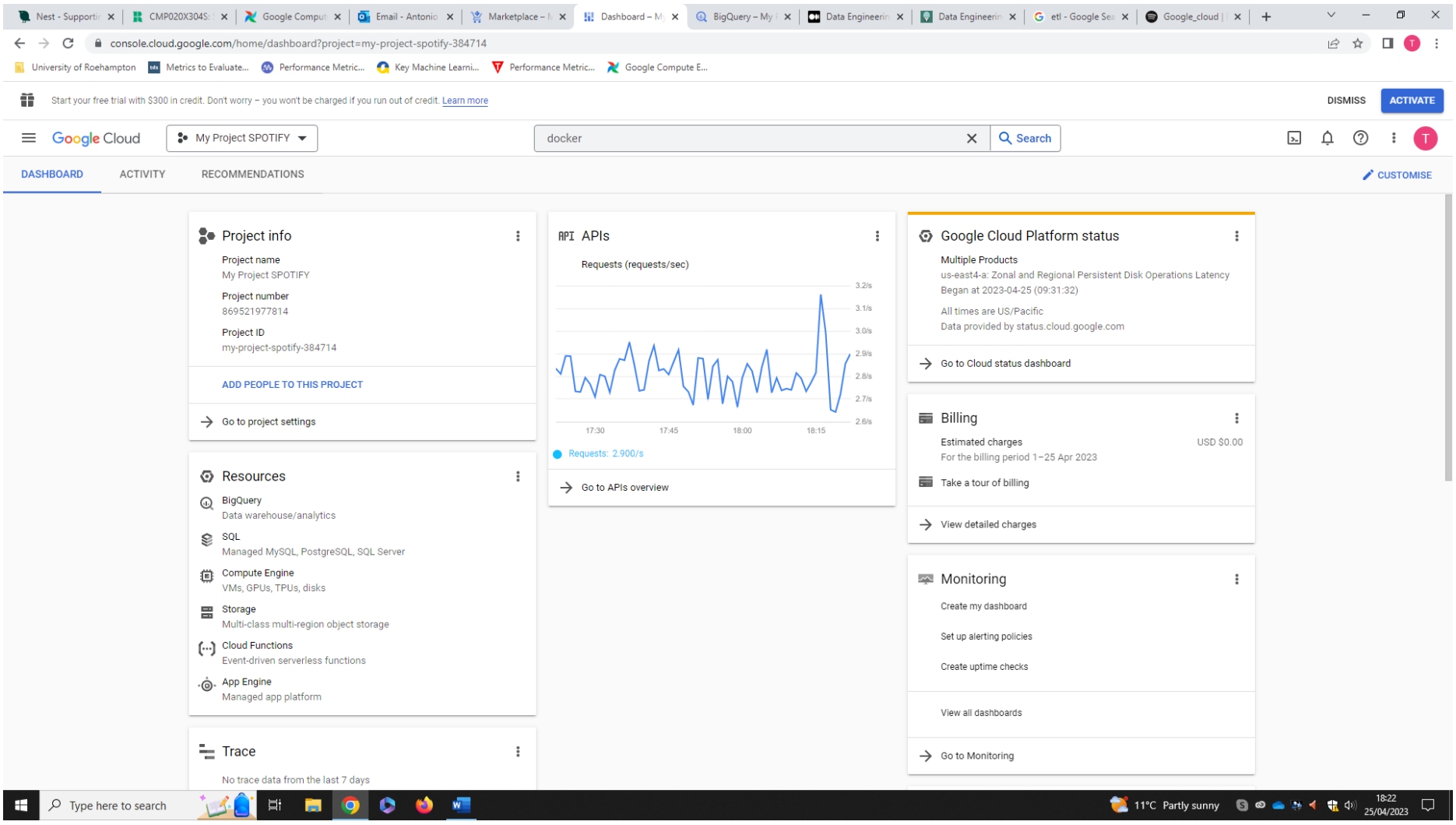
https://github.com/sidharth1805/Spotify_etl, https://github.com/culpgrant/Spotify_ETL, https://github.com/nekoemperor/spotify-ETL-job

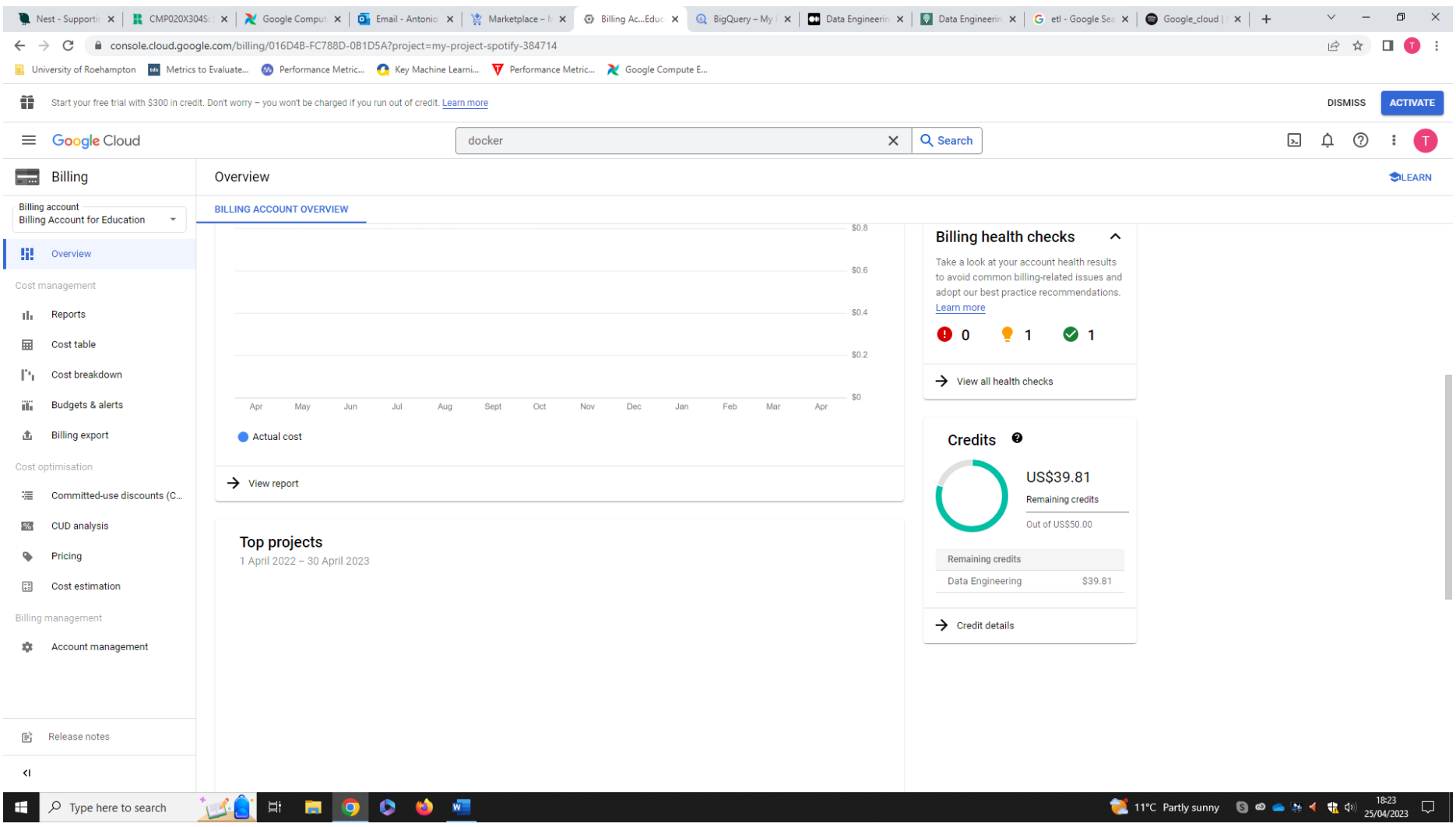
https://youtu.be/i25ttd32-eo, https://youtu.be/jvad7eF0qqY

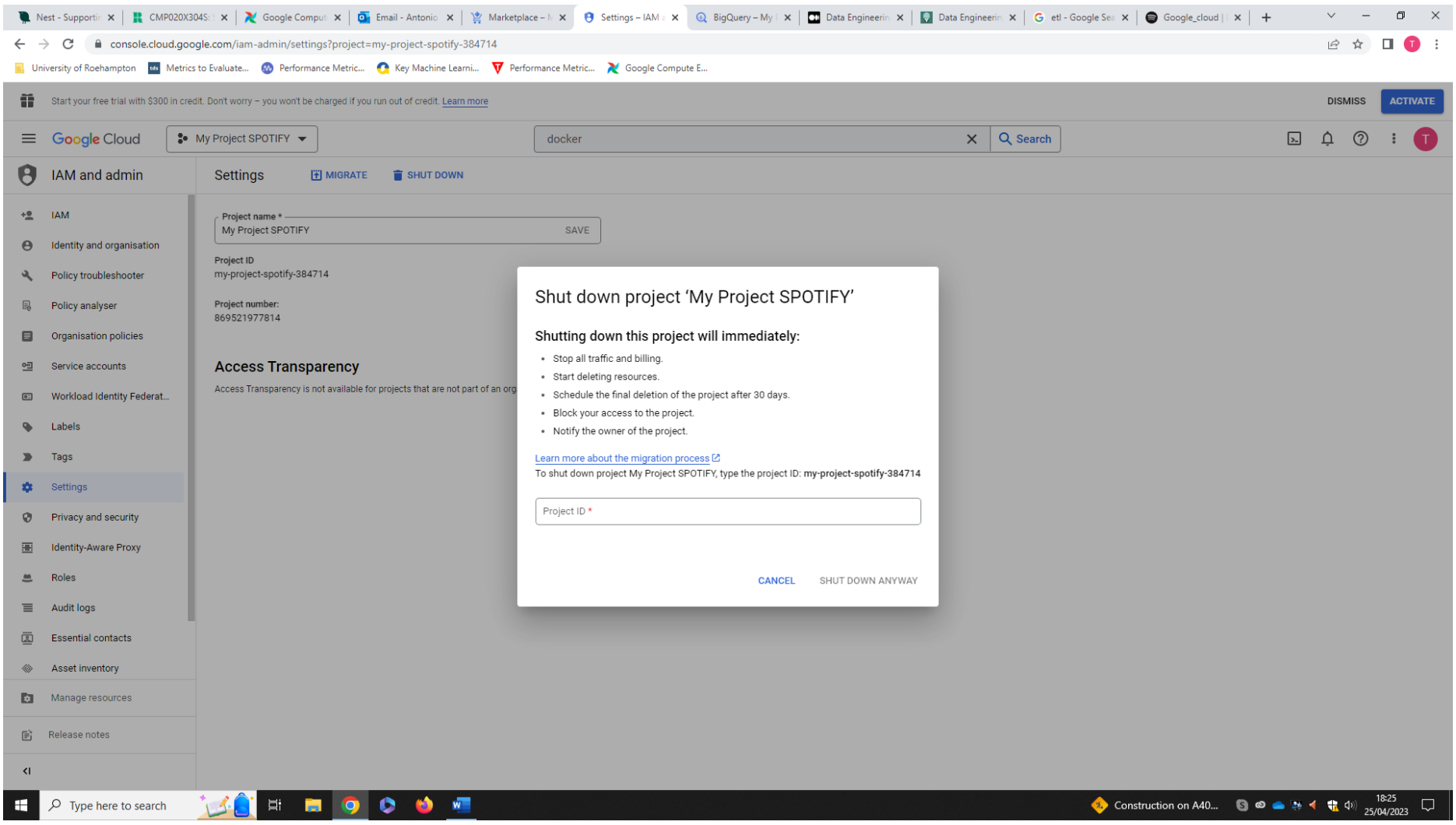
Spanish language and like first link from YouTube: https://youtu.be/eg8t2-E69ew

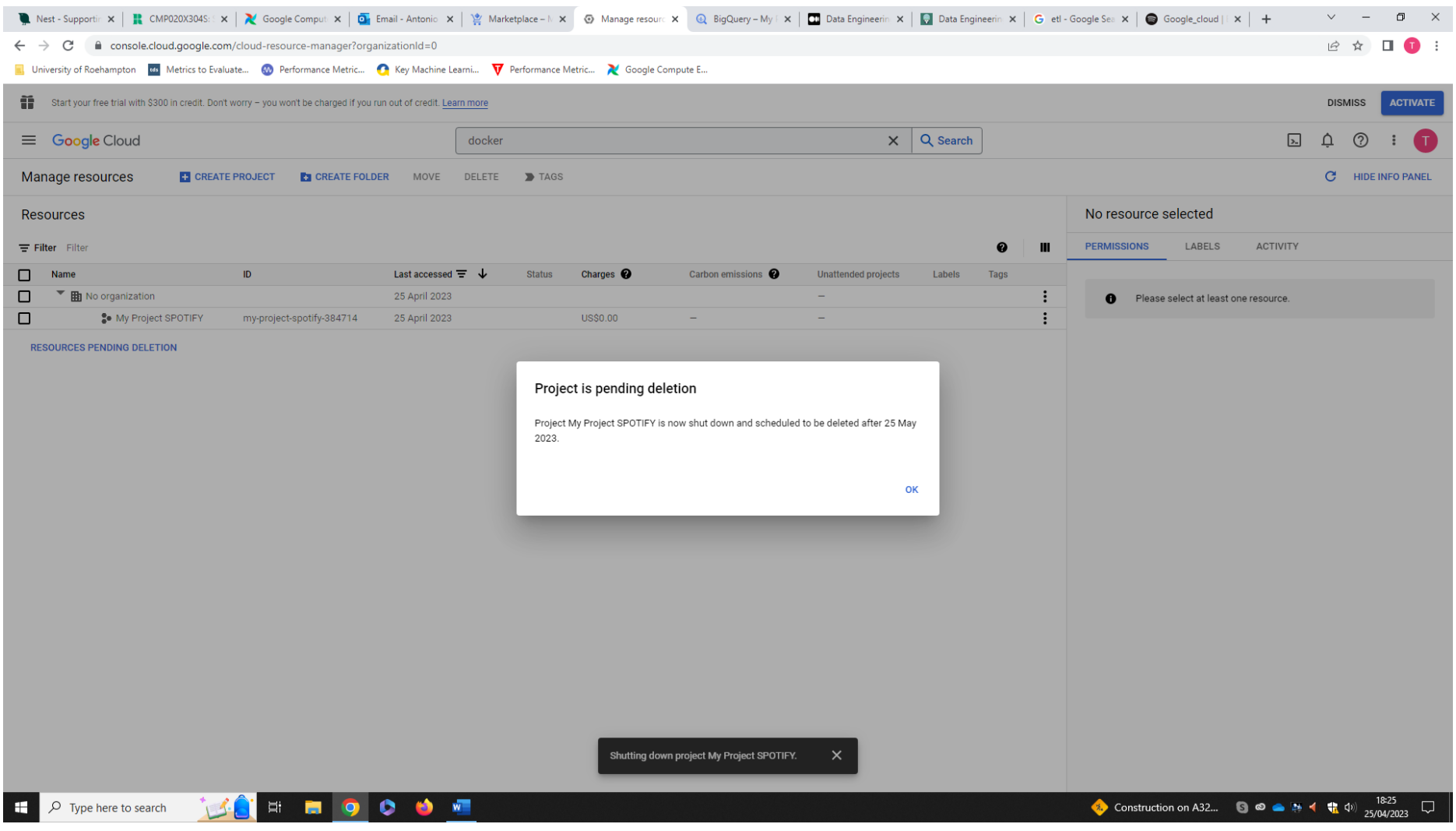
About Spotify and Spotify Engineers: https://youtu.be/L14enOBwi1A

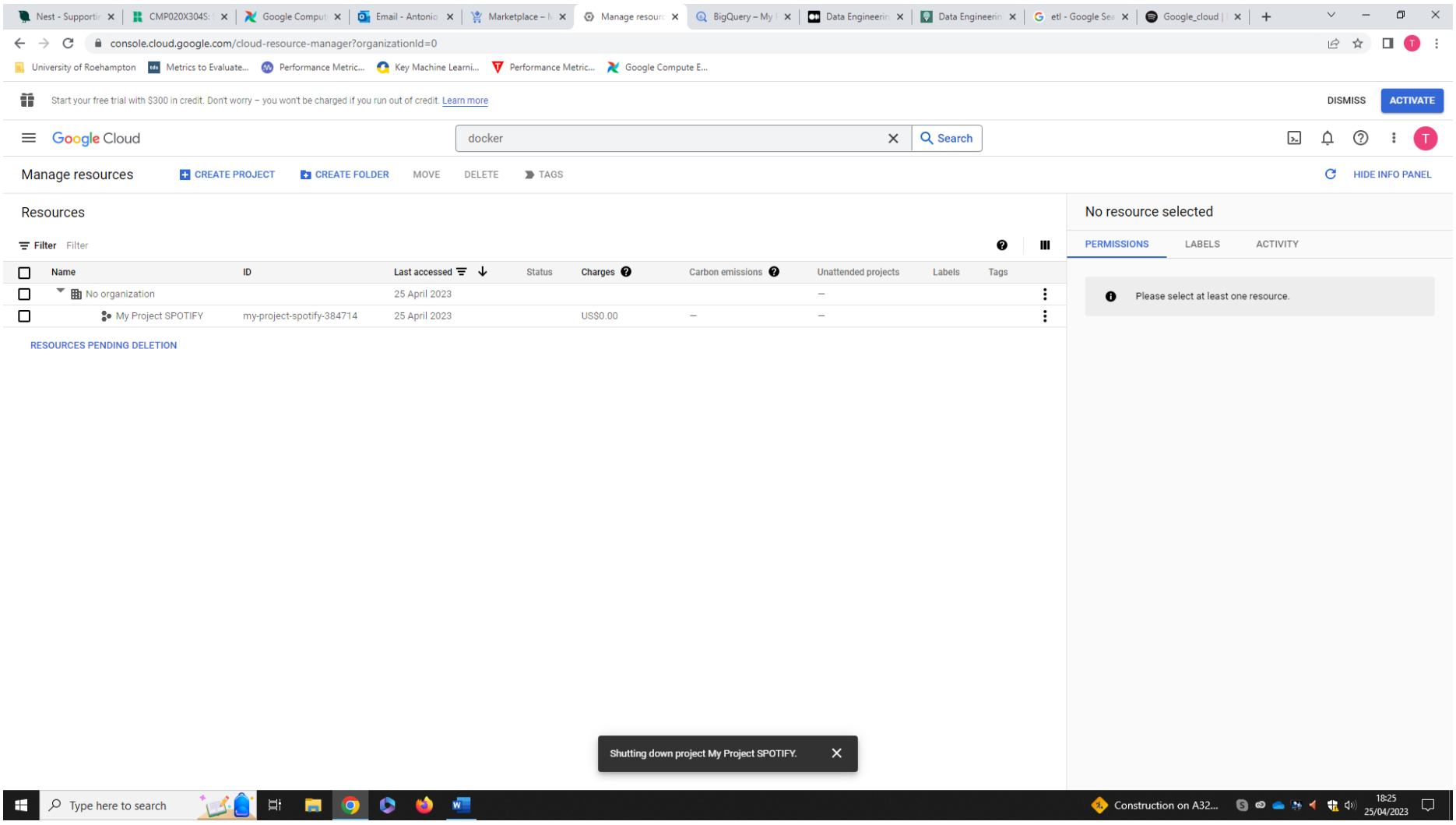
Notes and considerations: Netflix is like Spotify but uses video, not only music. This project can be applied to Netflix data producer.

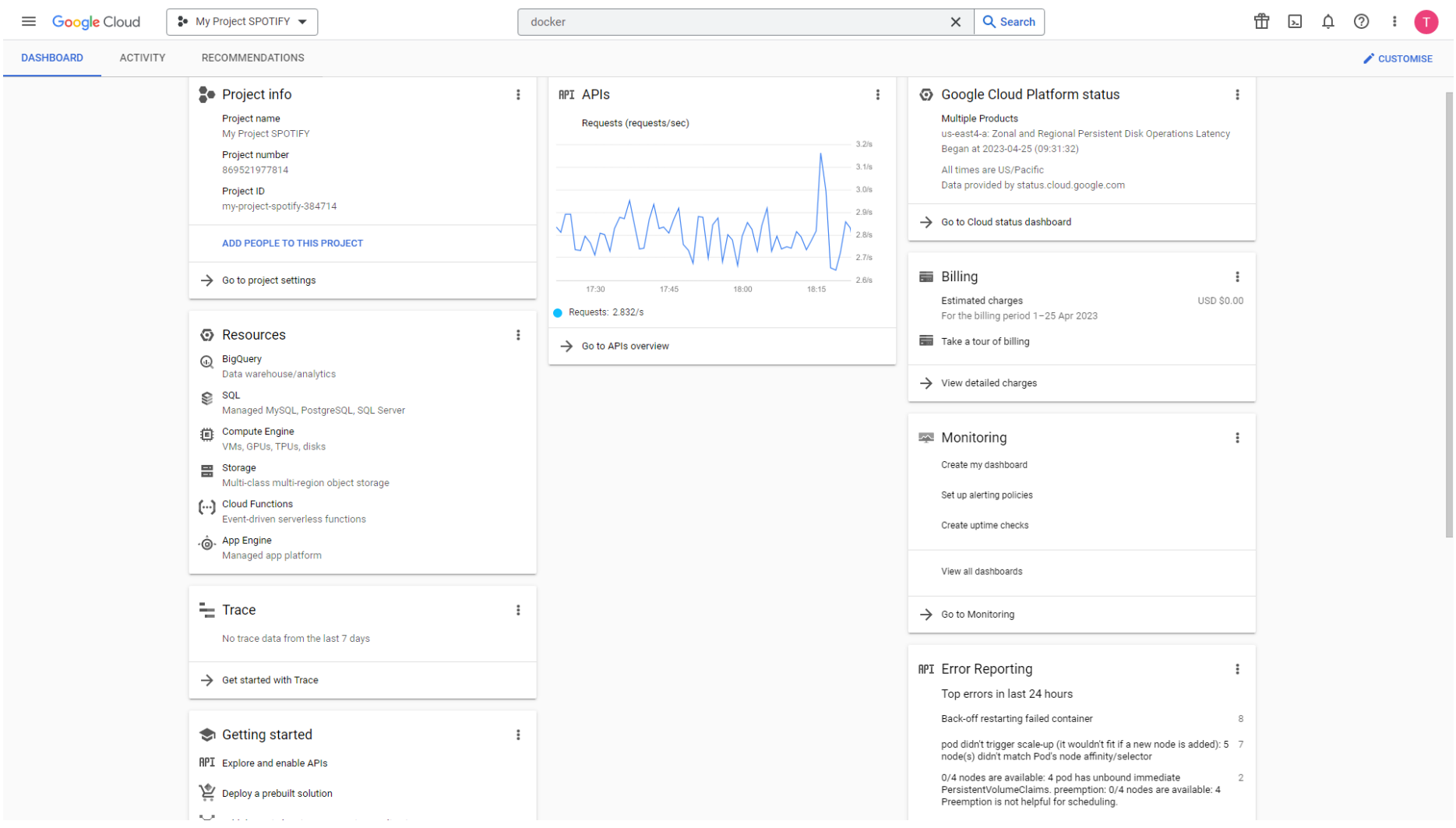

console.cloud.google.com/home/dashboard?project=my-project-spotify-384714
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
ACTIVATE
Google Cloud
My Project SPOTIFY
docker
Search
DASHBOARD
ACTIVITY
RECOMMENDATIONS
CUSTOMISE
Project info
Project name
My Project SPOTIFY
Project number
869521977814
Project ID
my-project-spotify-384714
ADD PEOPLE TO THIS PROJECT
Go to project settings
Resources
BigQuery
Data warehouse/analytics
SQL
Managed MySQL, PostgreSQL, SQL Server
Compute Engine
VMs, GPUs, TPUs, disks
Storage
Multi-class multi-region object storage
Cloud Functions
Event-driven serverless functions
App Engine
Managed app platform
Trace
No trace data from the last 7 days
APIs
Requests (requests/sec)
3.2/s
3.1/s
3.0/s
2.9/s
2.8/s
2.7/s
2.6/s
17:30
17:45
18:00
18:15
Requests: 2.900/s
Go to APIs overview
Google Cloud Platform status
Multiple Products
us-east4-a: Zonal and Regional Persistent Disk Operations Latency
Began at 2023-04-25 (09:31:32)
All times are US/Pacific
Data provided by status.cloud.google.com
Go to Cloud status dashboard
Billing
Estimated charges
USD $0.00
For the billing period 1–25 Apr 2023
Take a tour of billing
View detailed charges
Monitoring
Create my dashboard
Set up alerting policies
Create uptime checks
View all dashboards
Go to Monitoring
Type here to search
11°C Partly sunny
18:22
25/04/2023

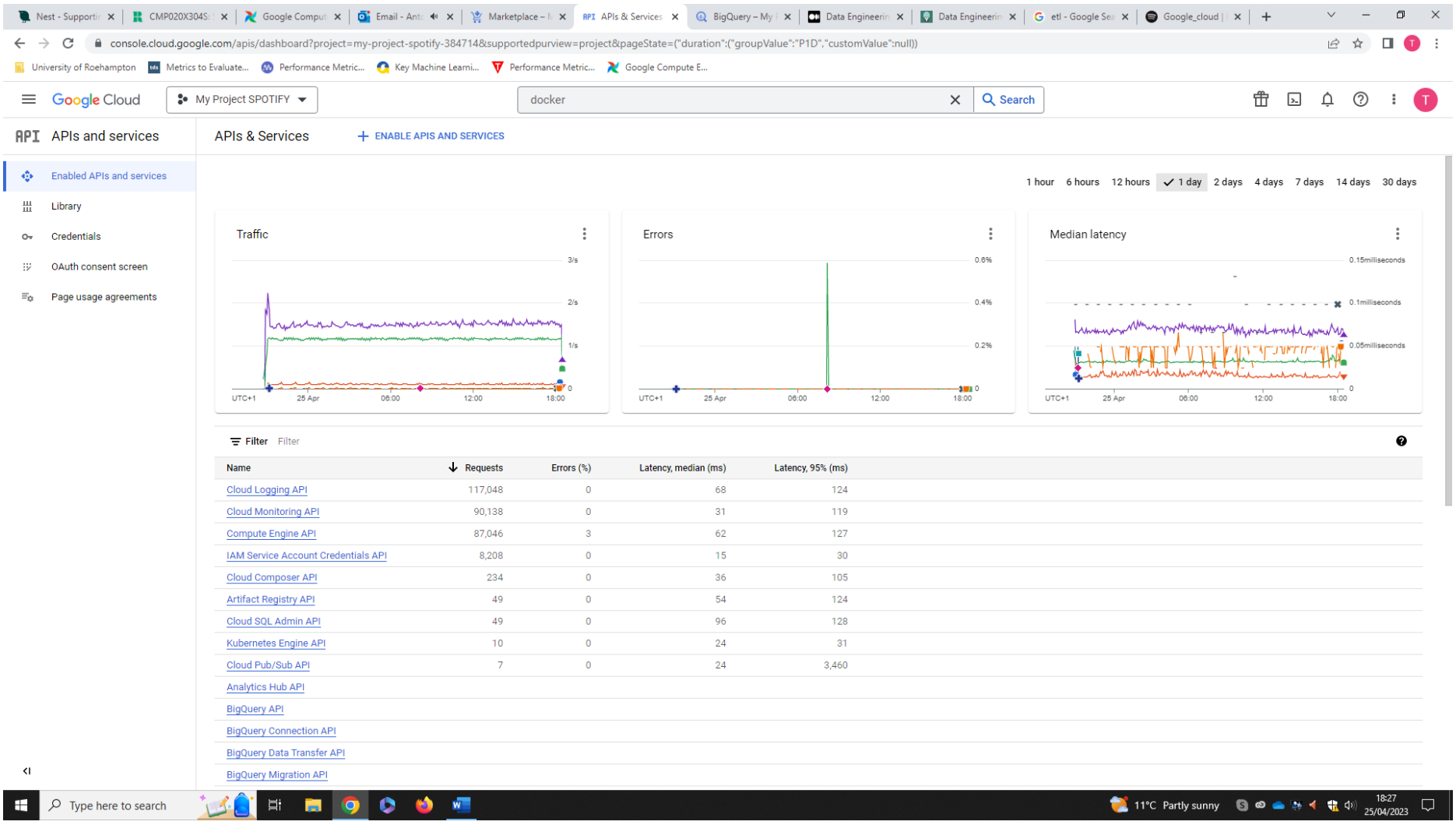

console.cloud.google.com/billing/016D4B-FC788D-0B1D5A?project=my-project-spotify-384714
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
ACTIVATE
Google Cloud
docker
Search
Billing
Overview
LEARN
Billing account
Billing Account for Education
BILLING ACCOUNT OVERVIEW
Overview
Cost management
Reports
Cost table
Cost breakdown
Budgets & alerts
Billing export
Cost optimisation
Committed-use discounts (C...
CUD analysis
Pricing
Cost estimation
Billing management
Account management
Release notes
$0.8
$0.6
$0.4
$0.2
$0
Apr
May
Jun
Jul
Aug
Sept
Oct
Nov
Dec
Jan
Feb
Mar
Apr
Actual cost
View report
Billing health checks
Take a look at your account health results to avoid common billing-related issues and adopt our best practice recommendations.
Learn more
0
1
1
View all health checks
Credits
US$39.81
Remaining credits
Out of US$50.00
Remaining credits
Data Engineering
$39.81
Credit details
Top projects
1 April 2022 – 30 April 2023
Type here to search
11°C Partly sunny
18:23
25/04/2023

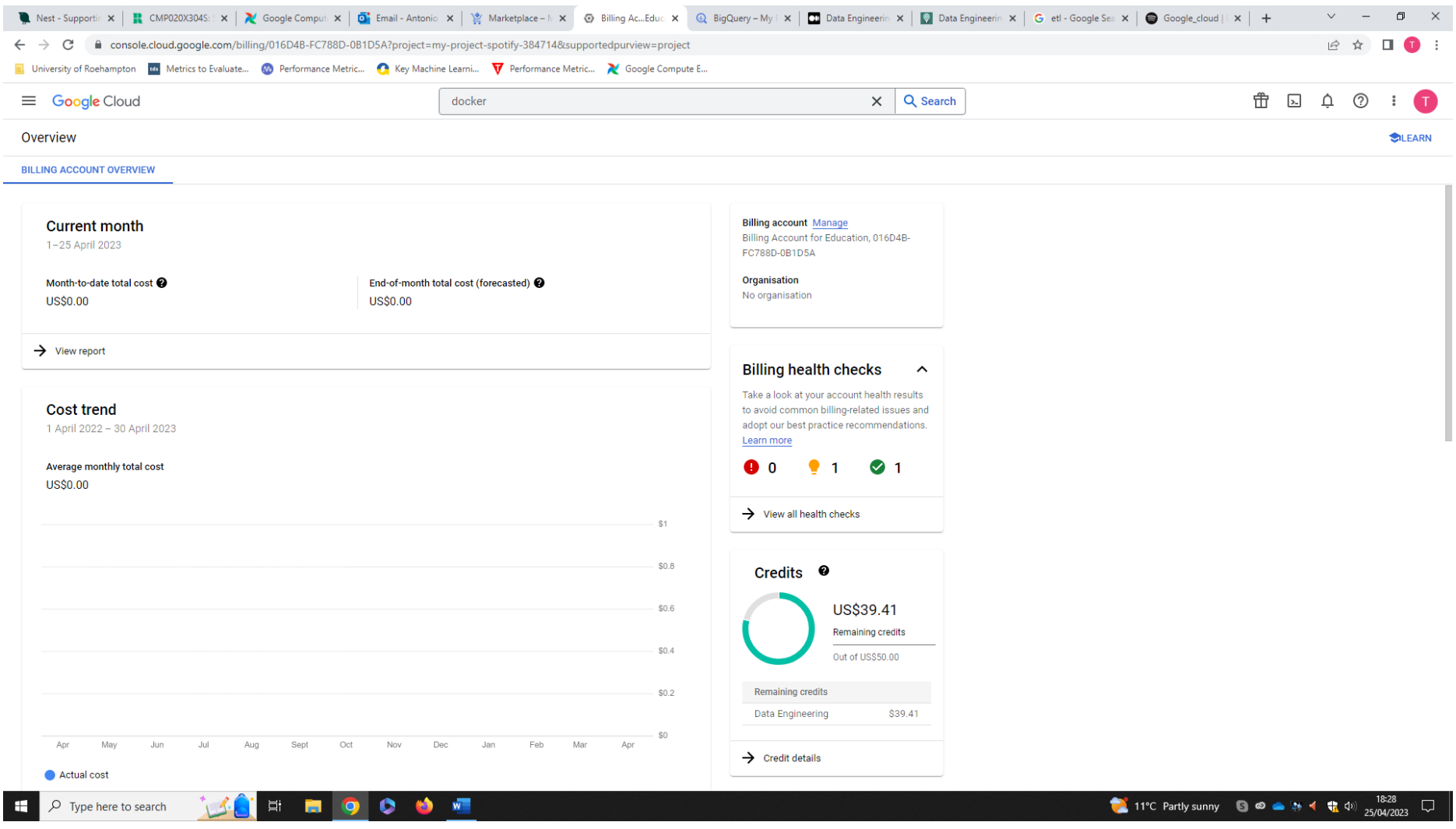

console.cloud.google.com/iam-admin/settings?project=my-project-spotify-384714
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
ACTIVATE
Google Cloud
My Project SPOTIFY
docker
Search
IAM and admin
Settings
MIGRATE
SHUT DOWN
IAM
Identity and organisation
Policy troubleshooter
Policy analyser
Organisation policies
Service accounts
Workload Identity Federat...
Labels
Tags
Settings
Privacy and security
Identity-Aware Proxy
Roles
Audit logs
Essential contacts
Asset inventory
Manage resources
Release notes
Project name *
My Project SPOTIFY
SAVE
Project ID
my-project-spotify-384714
Project number:
869521977814
Access Transparency
Shut down project 'My Project SPOTIFY'
Shutting down this project will immediately:
Stop all traffic and billing.
Start deleting resources.
Schedule the final deletion of the project after 30 days.
Block your access to the project.
Notify the owner of the project.
Learn more about the migration process
To shut down project My Project SPOTIFY, type the project ID: my-project-spotify-384714
Project ID *
CANCEL
SHUT DOWN ANYWAY
Type here to search
Construction on A40...
18:25
25/04/2023

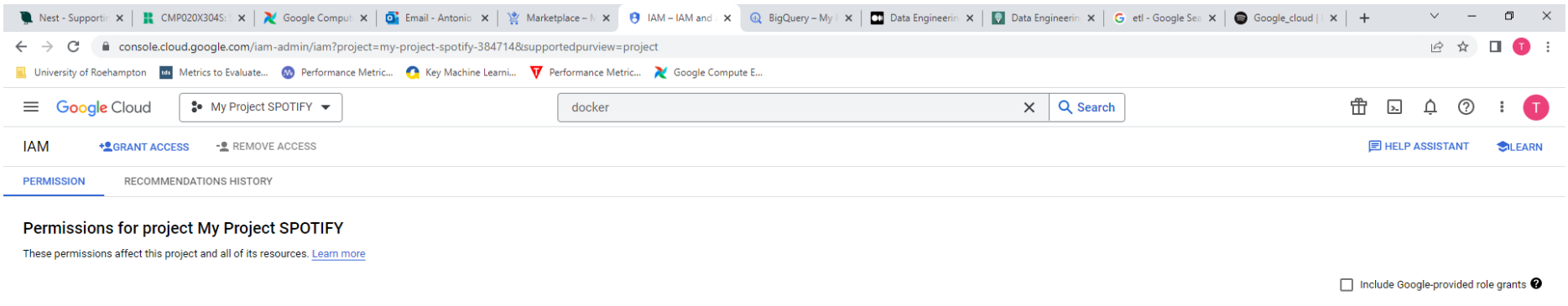

console.cloud.google.com/cloud-resource-manager?organizationId=0
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
ACTIVATE
Google Cloud
docker
Search
Manage resources
CREATE PROJECT
CREATE FOLDER
MOVE
DELETE
TAGS
HIDE INFO PANEL
Resources
Filter
Name
ID
Last accessed
Status
Charges
Carbon emissions
Unattended projects
Labels
Tags
No organization
25 April 2023
My Project SPOTIFY
my-project-spotify-384714
25 April 2023
US$0.00
RESOURCES PENDING DELETION
No resource selected
PERMISSIONS
LABELS
ACTIVITY
Please select at least one resource.
Project is pending deletion
Project My Project SPOTIFY is now shut down and scheduled to be deleted after 25 May 2023.
OK
Shutting down project My Project SPOTIFY.
Type here to search
Construction on A32...
18:25
25/04/2023

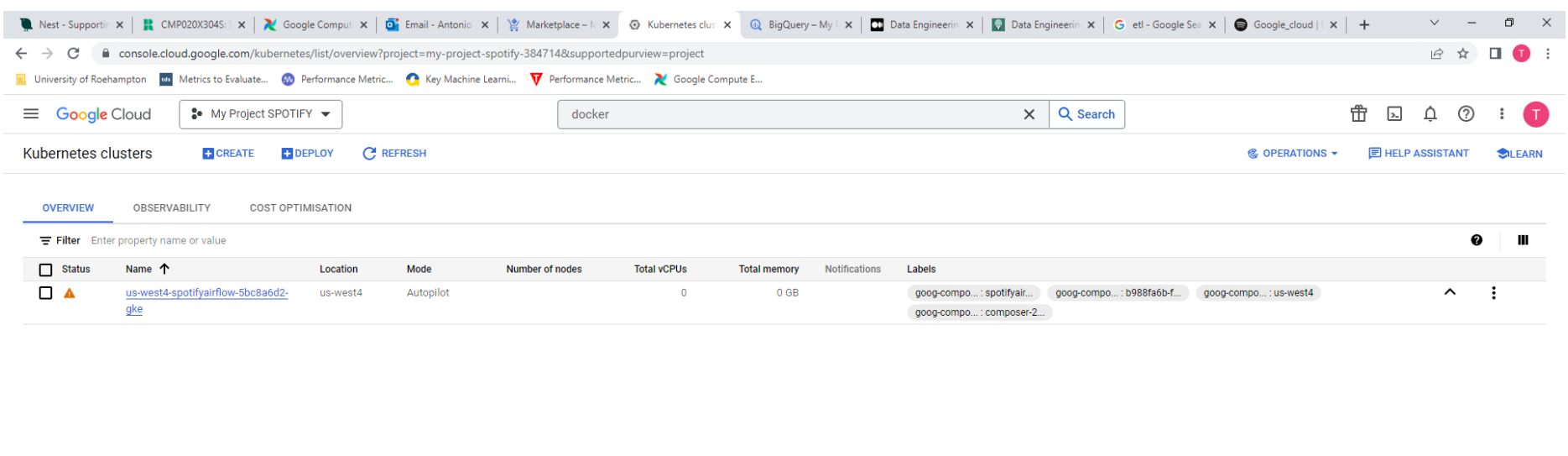

console.cloud.google.com/cloud-resource-manager?organizationId=0
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
ACTIVATE
Google Cloud
docker
Search
Manage resources
CREATE PROJECT
CREATE FOLDER
MOVE
DELETE
TAGS
HIDE INFO PANEL
Resources
Filter
Name
ID
Last accessed
Status
Charges
Carbon emissions
Unattended projects
Labels
Tags
No organization
25 April 2023
My Project SPOTIFY
my-project-spotify-384714
25 April 2023
US$0.00
RESOURCES PENDING DELETION
No resource selected
PERMISSIONS
LABELS
ACTIVITY
Please select at least one resource.
Shutting down project My Project SPOTIFY.
Type here to search
Construction on A32...
18-25
25/04/2023

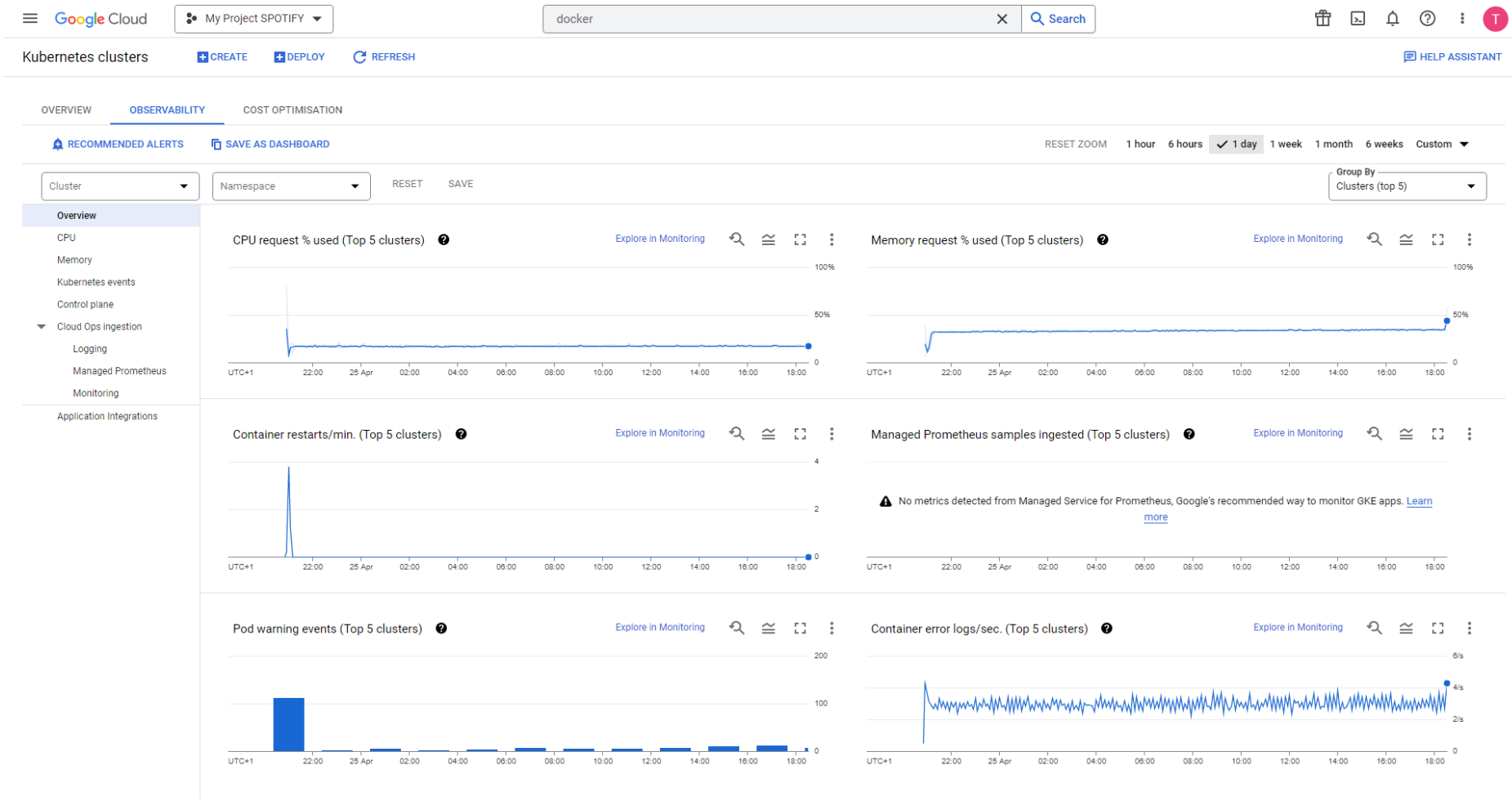

Google Cloud
My Project SPOTIFY
docker
Search
DASHBOARD
ACTIVITY
RECOMMENDATIONS
CUSTOMISE
Project info
Project name
My Project SPOTIFY
Project number
869521977814
Project ID
my-project-spotify-384714
ADD PEOPLE TO THIS PROJECT
Go to project settings
APIs
Requests (requests/sec)
Requests: 2.832/s
Go to APIs overview
Google Cloud Platform status
Multiple Products
us-east4-a: Zonal and Regional Persistent Disk Operations Latency
Began at 2023-04-25 (09:31:32)
All times are US/Pacific
Data provided by status.cloud.google.com
Go to Cloud status dashboard
Billing
Estimated charges
USD $0.00
For the billing period 1–25 Apr 2023
Take a tour of billing
View detailed charges
Resources
BigQuery
Data warehouse/analytics
SQL
Managed MySQL, PostgreSQL, SQL Server
Compute Engine
VMs, GPUs, TPUs, disks
Storage
Multi-class multi-region object storage
Cloud Functions
Event-driven serverless functions
App Engine
Managed app platform
Monitoring
Create my dashboard
Set up alerting policies
Create uptime checks
View all dashboards
Go to Monitoring
Trace
No trace data from the last 7 days
Get started with Trace
Error Reporting
Top errors in last 24 hours
Back-off restarting failed container 8
pod didn't trigger scale-up (it wouldn't fit if a new node is added): 5 node(s) didn't match Pod's node affinity/selector 7
0/4 nodes are available: 4 pod has unbound immediate PersistentVolumeClaims. preemption: 0/4 nodes are available: 4 Preemption is not helpful for scheduling. 2
Getting started
Explore and enable APIs
Deploy a prebuilt solution


console.cloud.google.com/apis/dashboard?project=my-project-spotify-384714&supportedpurview=project&pageState=("duration":("groupValue":"P1D","customValue":null))
Google Cloud
My Project SPOTIFY
docker
Search
APIs and services
APIs & Services
ENABLE APIS AND SERVICES
Enabled APIs and services
Library
Credentials
OAuth consent screen
Page usage agreements
1 hour 6 hours 12 hours 1 day 2 days 4 days 7 days 14 days 30 days
Traffic
Errors
Median latency
Filter
Name Requests Errors (%) Latency, median (ms) Latency, 95% (ms)
Cloud Logging API 117,048 0 68 124
Cloud Monitoring API 90,138 0 31 119
Compute Engine API 87,046 3 62 127
IAM Service Account Credentials API 8,208 0 15 30
Cloud Composer API 234 0 36 105
Artifact Registry API 49 0 54 124
Cloud SQL Admin API 49 0 96 128
Kubernetes Engine API 10 0 24 31
Cloud Pub/Sub API 7 0 24 3,460
Analytics Hub API
BigQuery API
BigQuery Connection API
BigQuery Data Transfer API
BigQuery Migration API
Type here to search
11°C Partly sunny
18:27
25/04/2023


console.cloud.google.com/billing/016D4B-FC788D-0B1D5A?project=my-project-spotify-384714&supportedpurview=project
Google Cloud
docker
Search
Overview
LEARN
BILLING ACCOUNT OVERVIEW
Current month
1–25 April 2023
Month-to-date total cost
US$0.00
End-of-month total cost (forecasted)
US$0.00
View report
Cost trend
1 April 2022 – 30 April 2023
Average monthly total cost
US$0.00
Actual cost
Billing account Manage
Billing Account for Education, 016D4B-FC788D-0B1D5A
Organisation
No organisation
Billing health checks
Take a look at your account health results to avoid common billing-related issues and adopt our best practice recommendations.
Learn more
0 1 1
View all health checks
Credits
US$39.41
Remaining credits
Out of US$50.00
Remaining credits
Data Engineering $39.41
Credit details
Type here to search
11°C Partly sunny
18:28
25/04/2023


console.cloud.google.com/iam-admin/iam?project=my-project-spotify-384714&supportedpurview=project
Google Cloud
My Project SPOTIFY
docker
Search
IAM
GRANT ACCESS
REMOVE ACCESS
HELP ASSISTANT
LEARN
PERMISSION
RECOMMENDATIONS HISTORY
Permissions for project My Project SPOTIFY
These permissions affect this project and all of its resources. Learn more
Include Google-provided role grants
VIEW BY PRINCIPALS
VIEW BY ROLES
Filter Enter property name or value
Type Principal Name Role Security insights Inheritance
869521977814-compute@developer.gserviceaccount.com Compute Engine default service account Editor
869521977814@cloudbuild.gserviceaccount.com Cloud Build Service Account
869521977814@cloudservices.gserviceaccount.com Google APIs Service Agent Editor
ittonycash@gmail.com Tony Cash Owner


console.cloud.google.com/kubernetes/list/overview?project=my-project-spotify-384714&supportedpurview=project
Google Cloud
My Project SPOTIFY
docker
Search
Kubernetes clusters
CREATE
DEPLOY
REFRESH
OPERATIONS
HELP ASSISTANT
LEARN
OVERVIEW
OBSERVABILITY
COST OPTIMISATION
Filter Enter property name or value
Status Name Location Mode Number of nodes Total vCPUs Total memory Notifications Labels
us-west4-spotifyairflow-5bc8a6d2-gke us-west4 Autopilot 0 0 GB
goog-compo... : spotifyair...
goog-compo... : b988fa6b-f...
goog-compo... : us-west4
goog-compo... : composer-2...



END. TUESDAY, 25 APRIL 2023