

Fetching data from APIs(GitHub) using Django and GraphQL without hitting the rate limits

Manaswini Das

Department of Computer Science and Engineering
College of Engineering and Technology, Bhubaneswar, Odisha, India



Who am I?

- Student
- Processing Foundation fellow 2019
- Girlsript Summer of Code mentor 2019
- Outreachy 2018 alumna

GitHub API

- Data is already sorted and classified.
- Excellent source for visualization
- meticulous information regarding pull requests, issues, activity feeds, repositories, organizations

What are rate limits?

- Controls the amount of outgoing and incoming traffic to and from a network
- Ensures better flow of data
- Increases security
- Prevents user from making an attempt to loads tons of information

Challenges

- GitHub API uses rate limits, which need to be respected and going over the rate limit would not yield more data but just errors
- Fetching data from GitHub takes time, not because of rate limits, but it can be a lot of data
- We want to regularly update data and take into account data we already did upload to Open Humans.

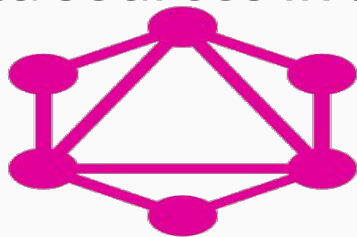
Tools for rate limiting- Celery and requests_respectful

- Celery is a simple, flexible, and reliable distributed system to process vast amounts of requests
requests-respectful:
- Is a minimalist wrapper on top of *requests* to work within rate limits of any amount of services simultaneously
- Can scale out of a single thread, single process or even a single machine
- Proxies *requests* HTTP verb methods (for minimal code changes)
- Works with both Python 2 and 3 and is fully tested

Check it out: <https://github.com/SerpentAI/requests-respectful>

GraphQL

- Open sourced, server-side runtime specification developed by Facebook under MIT license
- Query language to query, fetch and mutate data from any data sources in a declarative way.

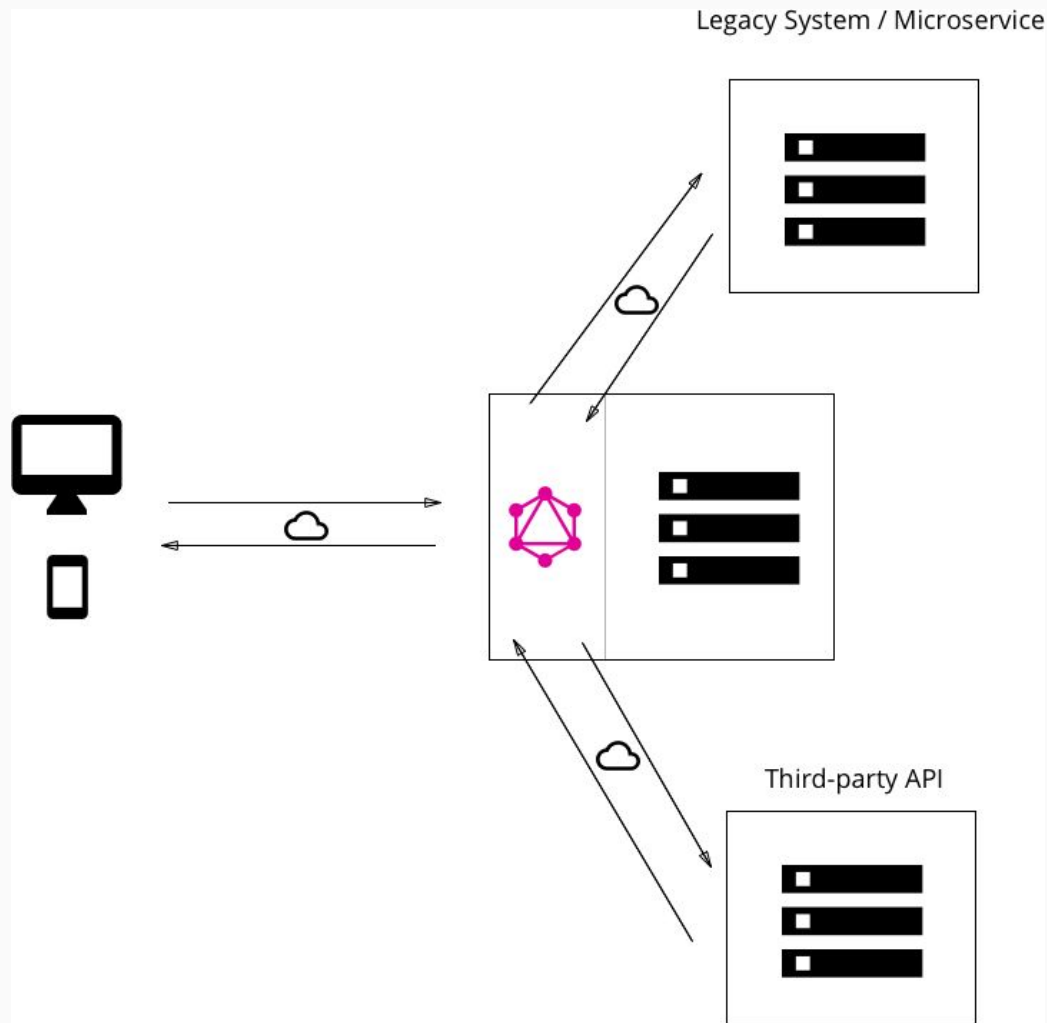


GraphQL

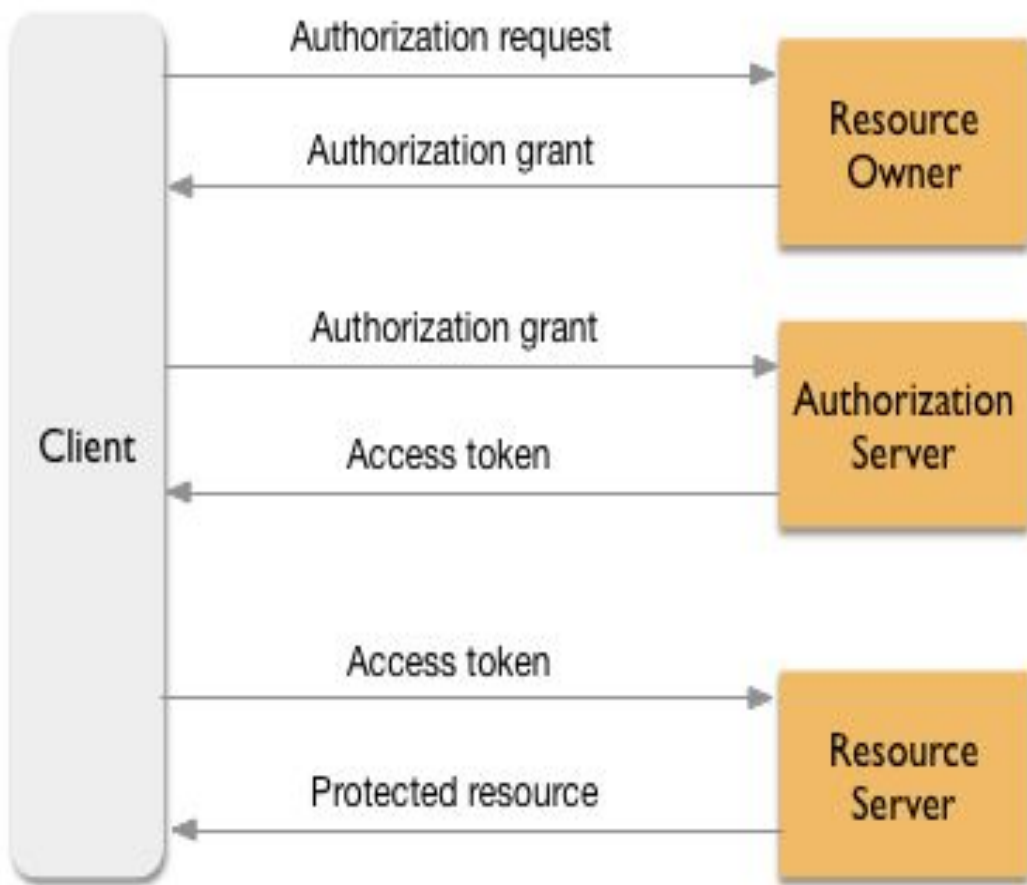
Why GraphQL?

- **Versioning:** To avoid multiple versioning of REST API
- **Ask for what you need:** Provision of asking only those fields that user needs
- **Get many APIs' response in single request:** Calls one query to get data from multiple REST APIs

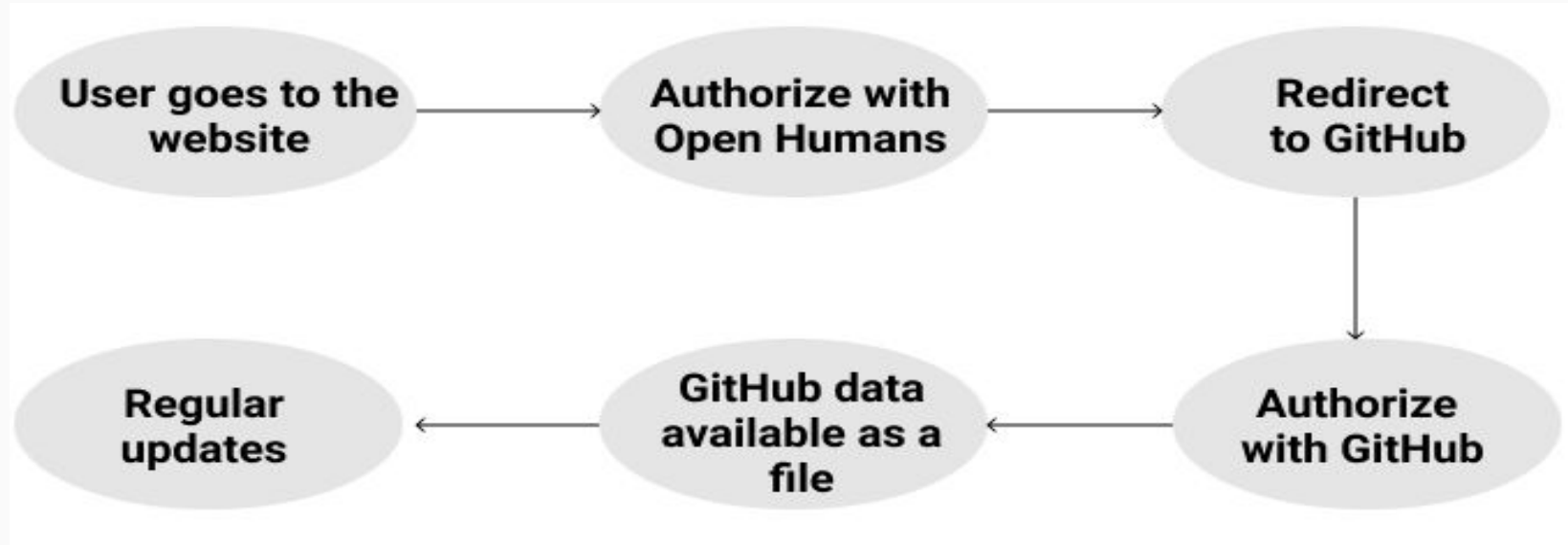
How does GraphQL work?



OAuth



Workflow



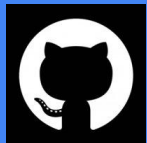
Demo time

<https://github.com/manaswinidas/oh-github-source/>



Thank you

Reach out to me:



/manaswinidas



/ManaswiniDas4



/manaswini-das