Request:

The service will get a request with a repository name and will reach out to download the repository from Intuit GitHub servers and scan that repository.

** 1000s of requests a day

Suggestion:

I'm suggesting to mirror GitHub repositories to Google Cloud Storage. Then to create Git triggers and invoke automatically repository builds on push / pull requests.

On demand, File-Scanner microservice will receive last modified files read channel from Google Cloud Storage and scan them to detect a secret.

** We can add an option to scan full repository or just last modified files.

Limitation:
Default API request hourly rate limit: 1000 requests

Configuration Service

Call to google cloud if there is no valid scan from last hour (configured time), otherwise retun last completed job results

Note:

Cloud Storage is a highly scalable service that uses auto-scaling technology to achieve very high request rates.

** Approximately 1000 object write requests per second, which includes uploading, updating, and deleting objects.

** Approximately 5000 object read requests per second, which includes listing objects, reading object data, and reading object metadata.

Files scan validity time / GCP credentials etc

Scan Repository Request (**private / inner api**)

Secret Detect Response

File-Scanner Micro-Service

Files Read Channel

Google Cloud Storage

Authomatically invoke repository builds

Git triggers on Push / Pull Requests

Persist repository scan job details and it's results

SQL Database

Intuit GitHub

Repository Scan Job

- id: uuid (PK)
- repository: string (githab repository)
- type: enum (secret detect)
- status: enum (initiated / complited / failed)
- created on: instant
- modified on: instant

- **unique constraint:** repository name + status initiated

1   *

Secret Detect

- id: uuid (PK)
- scan job id: reference (FK)
- file name: string
- access key line: int
- secret key line: int
- created on: instant
- modified on: instant