



shreyas13

[< Finish Later](#)

## Submitting project to Spârkâthon by NMIMS Navi Mumbai

Describe your project

### Project Name

WHAT ARE YOU CALLING IT?

DataSec.AI - Data Anonymization Microservice

### Tagline

WRITE A SHORT, SHARP AND ON POINT DESCRIPTION OF YOUR PROJECT

DataSec.AI aims to tackle data privacy issues of the 21st century by leveraging cutting-edge technologies integrated with state-of-the-art Artificial Intelligence Algorithms.

### The problem it solves

DESCRIBE WHAT CAN PEOPLE USE IT FOR, OR HOW IT MAKES EXISTING TASKS EASIER/SAFER E.T.C  
(MARKDOWN SUPPORTED)

Our project aims to mask all the sensitive Personally Identifiable Information (PII) on the web. This masking logic will

Our project aims to mask all the sensitive personally identifiable information (PII) on the web. This masking logic will work in real-time and can connect to the company VPN and intercept all the traffic passing through the network. The masking logic can be configured by our clients, once their accounts are authorized by the admin. Several types of masks will be provided to ensure that our software covers all types of PII, especially in the pharmaceutical industry. The software can be deployed as both Cloud and On-Premise setup based upon our client company's desire. Containerized deployment on Google Kubernetes Engine helps speed up the anonymization process, auto-scaling, auto-healing in case of errors, regular health checks, and periodic report generation. The CI/CD pipeline helps to push and deploy new code modifications with great ease. We have employed Istio's Service Mesh Architecture to deploy our Project on Google Kubernetes Engine. Squid Proxy acts as a Reverse Proxy capable of intercepting all the traffic on a given network. Squid Proxy acts as a sidecar to the Python ICAP Server which Masks/Unmasks PII Data from the intercepted traffic. Redis is used for the purpose of in-memory caching of Masking logic, Request Configurations, Response Configurations, and User ID Management. Flask framework is used to develop the Configuration Software. PostgreSQL Database is used for the purpose of RDBMS. SpaCy's Presidio Analyzer Engine is leveraged to detect and anonymized the sensitive PII data from requests and responses.

## Challenges we ran into

TELL US ABOUT ANY SPECIFIC BUG OR HURDLE YOU RAN INTO WHILE BUILDING THIS PROJECT. HOW DID YOU GET OVER IT? (MARKDOWN SUPPORTED)

We ran into a couple of challenges while developing our project:

1. Squid Proxy & ICAP plugin is not well documented.
2. ICAP Server's Python library does not support all the anonymization features.
3. Deployment of Squid Proxy and ICAP Server satisfying Istio's Service Mesh Architecture required DevOps expertise.
4. Integrating critical functionality with Flask-backend required some research.
5. Integrated SpaCy NLP Engine (Presidio Analyzer) with ICAP Server.
6. Implementing end-to-end CI/CD Pipeline (Cloud Build) on Google Cloud Platform.

## Technologies we used

WRITE A COMMA SEPARATED LIST OF TECHNOLOGIES YOU USED IN BUILDING THE PROJECT.

k Heroku PostgreSQL Redis Google Cloud Platform (GCP) Squid proxy  
ICAP server Presidio Analyzer Spacy NLP Engine

## Links

ADD LINKS TO GITHUB, WEBSITE, APP STORE E.T.C OR WHEREVER THE PROJECT CAN BE TESTED LIVE

<https://github.com/Shreyas-I/Sparkathon> ✕

Paste or type a link (e.g. <https://devfolio.github.io/>) and then press 'Enter'.

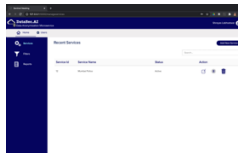
## Video Demo

ADD LINK TO VIDEO DEMOING THE FUNCTIONING OF THE PROJECT

<https://youtu.be/UiyhenmmKfE>

## Pictures

UPLOAD A MAXIMUM OF 5 PICTURES (SIZE: MAX 1MB EACH) SHOWCASING YOUR PROJECT (THE FIRST IMAGE WILL BE USED AS A PREVIEW ON SOCIAL MEDIA).



### Applicable tracks

Please only select the sponsor tracks or prizes that are related to your submission. A complete list can be found on [the hackathon microsite](#).

☐ Best Dapp Hack (Portis)

☐ Best Dapp Hack (Tezos)

☐ Best Ethereum/Matic Hack

☐ Cogno AI Internship Interview Prize