

Vinayak Mehta

DG-3/228, Vikas Puri, New Delhi, India

+91-9868525444; vmehta94@gmail.com; vinayak.io

EXPERIENCE

Grofers, Software Engineer — (Feb 2019 - Nov 2019) -> **Software Engineer II** — (Dec 2019 - May 2020)

- Designed and built a notebook-centric data platform which involved setting up JupyterHub and Airflow on our existing Kubernetes cluster; creating a Python library to automate common tasks like running queries on various databases within the organization, sending email reports, and loading data into Redshift which powered our Tableau dashboards; creating a CLI tool to automate scheduling notebooks as Airflow DAGs; and creating Jenkins jobs on GitHub repos as continuous deployment glue between JupyterHub and Airflow.
- The platform helped break silos between the data teams across the Gurgaon and Bangalore offices, and enabled everyone to schedule email reports, ML model training and ETL jobs, and build other complex workflows.
- Utilized the existing Prometheus and Grafana infrastructure to set up monitoring and alerting, which helped us maintain scheduling SLAs, and reduce our resource usage on the Kubernetes cluster.
- Helped organize the monthly [Grofers Tech Talks meetup](#) in our Bangalore office.

SocialCops, Software Engineer — June 2016 - July 2018

- Designed and built a part of the internal data platform responsible for data ingestion (from PDFs, REST/SOAP web services, Google Sheets) and data transformation (merging/concatenating tables, with support for uploading Python/R scripts for custom transformation rules), which enabled the data team within the organization to create and schedule ETL jobs on their own, and drive projects like the [DISHA](#) and [UN SDG tracking](#) dashboards.

PROJECTS

Camelot – github.com/camelot-dev/camelot

A Python library to extract tabular data from PDFs.

Excalibur – github.com/camelot-dev/excalibur

A web interface to extract tabular data from PDFs.

Present – github.com/vinayak-mehta/present

A terminal-based presentation tool with colors and effects.

TALKS

- NotebookOps: A pattern for building notebook-centric data platforms — [JupyterCon](#) (2020)
- The Hitchhiker's Guide to CLIs in Python — [PyCon US](#), [EuroPython](#), [PyCon AU](#) (2020)
- Extracting tabular data from PDFs with Camelot & Excalibur — [PyCon US](#), [PyCon AU](#), [PyCon India](#) (2019)

ARTICLES

- [The Hitchhiker's Guide to CLIs in Python](#) (2020)
- [Airflow, Meta Data Engineering, and a Data Platform for the World's Largest Democracy](#) (2018)
- [How to Create a Workflow in Apache Airflow to Track Disease Outbreaks in India](#) (2018)

SKILLS

Programming Languages: Python, some experience with C, Rust, JavaScript, HTML and CSS

Frameworks and Tools: Git, Flask, Celery, Airflow, Docker, Kubernetes, some experience with Spark

EDUCATION

Recurse Center – New York, USA

Participant, Aug 2020 - Oct 2020

Bharati Vidyapeeth's College of Engineering – New Delhi, India

Bachelor of Technology, Computer Science and Engineering, May 2016