

Arpit Solanki

DATA ENGINEER

☎ +91-8238059340 | ✉ solankiarpit1997@gmail.com | 🏠 arpitsolanki.in | 📞 arpit1997 | 🌐 arpit-solanki | in arpit-solanki

Education

Indian Institute of Information Technology

Vadodara, IN

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE, CGPA: 7.8/10

May 2018

Experience

Atlan (formerly SocialCops)

Delhi, IN

DATA ENGINEER

Jun. 2018 - Present

- Built data pipelines using Airflow for Ministry of Rural Development India's project DISHA which includes data ingestion from third party APIs, transformation using Python/R and data sink Elasticsearch including data versioning to rollback to previous versions.
- Managed on-premise deployments for DISHA project, packaged the entire stack of Vue.js frontend app, Node.js backend and Airflow pipelines into debian package.
- Created a health monitoring service for DISHA deployment using Prometheus and Grafana to alert stakeholders of critical scenarios like failing ETL pipelines, high resource usage, high traffic on dashboard etc.
- Improved cataloging performance of Catalog component of the product by 90% using Presto and Hive. Improved the data versioning by reducing the storage used by over 80%
- Added features like data deduplication, data extraction from SQL file dumps, data ingestion from S3, running SQL queries in the workflow to Workflows component of the product

Codementor

Remote

MENTOR

Jan. 2019 - Present

- Took projects involving Python and delivered them within assigned timelines.
- Took 100+ sessions with average 5 star rating and mentored developers in technologies like Python, Airflow, Spark and assisted them in problem solving

SocialCops

Delhi, IN

DATA ENGINEER INTERN

Jan. 2018 - May 2018

- Built a data integration framework using celery with support for state management and incremental data ingestion.
- Worked with data scientists and Business intelligence team to build ETL pipelines using Airflow to power analytics and dashboards
- Worked on building system native packages (Debian and RPM) of data integration services and ETL pipelines written in Python.

Projects

Human Detection and Tracking

Vadodara, IN

PROJECT LEAD

June 2016

- A command line application built with OpenCV and Python to detect individual humans and track them using their faces
- The application provides an API to users to train their own face datasets, the feature was built to support use cases like surveillance systems. The project has got 600+ stars, available on Github.

PQusic

Vadodara, IN

PROJECT LEAD

October 2016

- PQusic is a web application to listen audio songs, manage playlists and see trending songs
- Implemented the application using Python, Django, HTML, CSS and vanilla JS

Skills

Languages Python, Javascript, Shell scripting

Big Data Tools Presto, Hive, Airflow, Spark, Kafka, Elasticsearch, Parquet

Tools/Platform Flask, Docker, Kubernetes, AWS, Celery, RabbitMQ, Prometheus, Grafana, MongoDB

Others

Stackoverflow - Written 230+ answers, overall top 5% in Python with 5.7K reputation

Blogs - Testing Celery Tasks - <https://medium.com/@solankiarpit1997/testing-celery-tasks-a8a3568d0213>