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*

Skills

Languages	Python, Javascript, Shell scripting
Big Data Tools	Presto, Hive, Airflow, Spark, Kafka, Elasticsearch
Tools/Platform	Flask, Docker, Kubernetes, AWS, Celery, RabbitMQ, Prometheus, Grafana, MongoDB, Parquet

Experience

SocialCops *Delhi, IN*
Data Engineer *Jun. 2018 - Present*

- Led the development and deployment of Ministry of Rural Development India's project DISHA. Developed ETL pipelines using Airflow to power DISHA Dashboard.
- Managed on-premise deployments on NIC's Meghraj Cloud. Deployed the entire stack Vue.js dashboard, ETL pipelines and dashboard modification APIs, ETL pipelines using Debian packages.
- 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, hightraffic 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 Bussiness 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 *Vaodara, 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 *Vaodara, 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

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`