

SAHIL MOKASHI

✉ sahilmokashi6.1@gmail.com · ☎ (+91) 8971767006 · in sahil-mokashi

🎓 PROFILE SUMMARY

Data Engineer with a strong background in Google Cloud technologies and Python, with a track record of designing, optimizing data pipelines and leveraging Google Cloud services like Google Cloud Storage, BigQuery, Pub/Sub, Dataflow.

- <2 year experience working in Data Engineering
- Proficient with working on Cloud Technologies.
- Proficient in Python, SQL and worked with handling data, visualization, and exploratory data analysis.
- Experienced with working on PySpark (GCP - Dataproc) for data processing and ETL.
- Experience with working on most public cloud technologies (GCP / AZURE / AWS) with expertise in GCP.
- 2x Google Cloud certified.
- 1x Azure Cloud certified.

🎓 EDUCATION

S.G. Balekundri Institute of Technology, Belgaum , India

2018 – 2022

Bachelors in Computer Science and Engineering (CSE)

cgpa - 8.05

👤 EXPERIENCE

Persistent Systems, Pune, India

2022 – Present

Software Engineer

Jul. 2022 – Present

Data Engineering - Python, GCP, PySpark, SQL, ETL

POC : Amazon Reviews Analysis(GCP) - This project scrapes reviews from amazon for specific options using tools with selenium , Data is processed, analysed and Data is stored in Bigquery with reports. Python-Flask based app that is used to trigger the features.

- Worked on creating pipelines using Apache Beam (DataFlow) for data ingestion to GCS.
- Used cloud functions as triggers for the processing jobs.
- Used PySpark (DataProc) for data processing from GCS bucket.
- Loaded the data into data warehouse (Bigquery) with analysis reports using Dataflow pipelines.
- Worked with airflow dags to schedule the jobs.

WingMate - This is a GENAI tool that combines use of multiple generative AI via the use of API's , this adds features like code translations from python to ruby , testcase generation ,database gen , Git automation ,etc. Built on Python , Django and leverages cloud services(AWS).

- Worked on data processing and ETL using PySpark .
- worked with creating a docker image for the app
- worked with github actions to automate the process of creating the Docker image and deploying the image to ECR repository , and deploying it to AWS Apprunner.

ETL - ETL tasks on cloud using PYTHON.

- Trained and worked Used beam pipelines , spark jobs , airflow
- tasks revolving around data engineering(ETL) over cloud
- worked with Python and Sql extensively
- worked on data migrations
- use of Dataflow, cloud Storage, Dataproc, PySpark composer to accomplish tasks in optimised and efficient way

Software Engineering Intern

Feb. 2022 – Jul. 2022

Python, Java, Spring/SpringBoot

- worked using core java
- Worked on creating microservices using technologies spring / springboot
- worked with mysql database, along with rest api's.

Uilatech, Belgaum, India

Sept. 2021 – Oct. 2021

Intern

Python, Image Processing, OpenCV

Academic internship trained and worked on various machine learning concepts focusing on OpenCv and Image Processing

- worked on multiple mini projects for image processing and OpenCv.
- used pretrained models for various general purpose classifications.
- used mediapipe for movement detection in various ML projects.

⚙️ SKILLS

- Programming Languages: Python , Java
- Data: PySpark , SQL , Apache Beam , Apache Airflow ,ETL
- Platform: windows
- Cloud : GCP (proficient) ,Azure and AWS
- Operations: GIT
- other skills beginner: Flask , Django, Rest, spring/springboot.

♥️ HONORS AND AWARDS

WINNER of Persistent AI Technathon

Jul. 2023

Secured first place in the AI Hackathon , Built a java/spring based microservice that serves as a common notification platform for mail/sms/push notification. Use of inbuilt java mail sender and AWS (Ses,Sns,Push Notification) services. Successfully able to send mail and sms notifications using Rest API

Top performer of Interns batch 2022 at Persistent

Jul. 2022

Awarded for being among the top performers of Interns batch 2022(GEMS FY23) at Persistent Systems.

Winner of Project Exhibition at S.G.B.I.T

Mar. 2019

Secured 1st place at project exhibition held in my college(2nd year) for "IOT based Pollution monitoring system". Arduino Uno based Pollution (smoke) monitoring system that uses smoke detector to detect pollution.It sends the readings to 'thingspeak' which can be accessed via internet. This system triggers a sound alarm at the device and the user can monitor the pollution levels via their phone/web.

🎓 CERTIFICATIONS

- Google Cloud Certified : Associate Cloud Engineer Oct. 2022
- Google Cloud Certified : Professional Data Engineer Mar. 2023
- Microsoft : Azure Fundamentals AZ-900 Jul. 2022

📌 MISCELLANEOUS

- GitHub: <https://github.com/Sahil-A-Mokashi>
- Languages: English - Fluent, Hindi - Fluent , Kannada - Fluent , Marathi - limited working proficiency
- Leetcode: <https://leetcode.com/sahil-mokashi/>