

# SARTHAK JAIN

## Data Engineer

7983452171 sarthakjain31797@gmail.com https://www.linkedin.com/in/xsarthakjain/ Agra, India

### SUMMARY

Data Engineer with 2+ years of experience in the tech industry. Demonstrated ability to design, develop, and implement ETL pipelines. Proficient in Pyspark, Python, cloud technologies AWS,GCP, and databases.

### EXPERIENCE

#### Data Engineer

Tata Consultancy Services 2021 - Present Bangalore, India

##### Medalia(ETL Pipeline) & Other data migration Project

- Implemented an ETL (Extract Transform & Load) Pipeline to handle 80,000 daily data payloads.
- Deployed a batch ETL pipeline and an API using GCP Cloud Function,Django,Flask with a maximum payload size of 10MB.
- Designed Airflow DAG files to schedule tasks in multiple data products.
- Utilized BigQuery to perform analysis on large datasets

##### Created numerous Data Products

Constructed a robust Data Ingestion system and developed multiple Data Products to perform diverse tasks. Utilized SQL, Airflow,Informatica and GCP Cloud to design and implement the solutions. Created data product which can able to handle 100M+ records everyday with using SQL,Python Cloud services,Airflow.

#### Full Stack Developer

IWCN 2021 - 2021 Agra, India

##### ScrapingProject

- It scraps the number of appearances and their metadata of given urls, It handles almost 10,000 request perdays.
- Implemented application using Django and Sqlite.
- I designed a database to store correlated data for URLs.
- Used Beautiful Soup, and Selenium for Scraping and also avoiding 429 error.
- Implement Data Model for entire application.

### PROJECTS

#### Uber Cab Pipeline

23

q https://github.com/XSarthakJain/UberCab

- Created a data ingestion pipeline that takes Kaggle datasets as input on a daily basis.
- Implemented transformations on the dataset to extract Uber CAB status and count for each day, including accepted, rejected, and pending rides.
- Developed a process to store the transformed data in MongoDB for further analysis.
- Utilized Looker Data Studio to perform visualizations and generate meaningful insights from the data.

#### LetsChat Social Media

2020

q https://github.com/XSarthakJain/LetsChat

- This social media platform has various features such as Profile, chatting, login, user feed and many more.
- we used Django, Radix Server, Socket Programming, Machine Learning, HTML,CSS, JS.

### SKILLS

#### Web Development

HTML	CSS Javascript	Bootscript
Skatch	Beautiful Soup	React Ajax
Heroku	Django	

#### Programming

Python	Java	C
--------	------	---

#### Data

Pipelines ETL Pipelines MySQL
-------------------------------

BigQuery PySpark SQL Airflow MongoDB
--------------------------------------

Oracle Looker SQLITE
----------------------

Data modeling, AWS Glue, AWS Redshift, S3,
--

Lambda,Cloud Function
-----------------------

#### Cloud

Google Cloud platform GCP	Amazon Web Service AWS
---------------------------	------------------------

#### Control Systems

GIT	Github	Gitlab
-----	--------	--------

### EDUCATION

#### Master of Computer Application

AKTU University

2018 - 2021 Mathura,India

GPA  
8.36 / 10

#### Bachelor of Computer Applications

DBRAU University

2015 - 2018 Agra, India

### ACHIEVEMENTS



#### Certification and Awards

Google ACE Certification  
Best Team Award  
Learning Achievement Award