**Ashraf Ali Ansari** ashraf.shiksha001@gmail.com |  [LinkedIn](https://www.linkedin.com/in/ashraf-ansari-5b7640132) | (+44) 7756757296

### **Summary**

Results-driven Data Engineer with expertise in cloud-based data solutions, ETL/ELT pipeline development, and big data processing. Proficient in Python, SQL, Spark, and GCP, with hands-on experience designing scalable, cost-effective data architectures and optimizing real-time data ingestion. Skilled in DevOps, CI/CD, and data governance, with a passion for building high-performance data solutions to drive business insights.

### **Skills & Expertise**

| SQL | ETL/ELT | Cloud Computing | Data Processing & Analytics | Databricks | BigQuery |
| --- | --- | --- | --- | --- | --- |
| Azure DevOps | Microsoft Azure | Data Pipelines | Data Integration | Azure Data Lake | Data Governance |
| Data Modeling | Python | REST APIs | Data Pipelines | GitHub | Snowflake |
| Spark/ Pyspark | GCP | Azure Data Factory | PowerShell | Dataflow | Microsoft Azure |

**Soft-Skills:**

| ●Passionate | ● Hard Working | ● Team Player | ● Leadership |
| --- | --- | --- | --- |

### **Experience**

**Smallcap.ai – Backend & Machine Learning Intern** *(June 2024 – Sept 2024)*

* Developed a model to predict the prices of the Web3 tokens, performance of about 25% better than General purpose AI assistants e.g ChatGPT,Gemini
* Part of the backend team for react and flask to develop and test API’s of main application

**Tata Consultancy Services – Big Data Developer** *(Sept 2022 – Sept 2023)*

* Diagnosed and Optimised the gremlin graph query for minimising the runtime for data fetching.
* Inspected and fixed bugs for the System Data team to be consumed by the GSSP API team.
* Administered on changing the post processing scala file to improve ingestion of additional data points to HBASE

**LTI Mindtree – Senior Software Engineer** *(June 2021 – June 2022)*

* Built specialised python ingestion scripts for clients for automatic movement of data into BigQuery Tables from files.
* Supervised in development of scripts for orchestration of the data-pipelines in Apache-Airflow using Airflow DAG's
* Orchestrated to run data Pipelines using Google Dataflow.

**KPMG Global Services – Associate Consultant** *(July 2018 – June 2021)*

* Developed the Django with rest framework application for cloud data migration.
* Built a system of data pipelines in PySpark to ingest data coming from various sources(FTP,API,Zip) to the lake
* Developed an email scraper using beautiful Soup for collecting data about the sensitive emails, inline with Data Protection Policy

### **Education**

MSc, Advanced Computer Science – University of Essex *(Sept 2023 – Oct 2024)*B.Tech, Computer Science & Engineering – MAKAUT *(Aug 2014 – July 2018)*

### **Certifications & Awards**

GCP Certified Professional Data Engineer   
8th place in ACM B.Tech Awards 2018 (India East & North East)