

Jafar Sharif Shaik

513-680-5451 | jafarsharif6@gmail.com | [linkedin.com/in/jafar-sharif/](https://www.linkedin.com/in/jafar-sharif/) | <http://github.com/shaik-jafar6>

EDUCATION

University of Cincinnati- Master of Science in Information Systems

Cincinnati, Ohio | Apr 2023

Course work: Data Modelling, Database Design, Data Visualization, Big Data Integration, Data Warehouse and Business Intelligence

TECHNICAL SKILLS

- **Programming Languages:** Python, Pandas, Numpy, Postgres, MS SQL Server, DynamoDB
- **Big Data:** HDFS, Spark, Hive, Sqoop, AWS Glue, Kafka, Kinesis, AWS EMR, Apache Hudi, AWS Lambda, S3, Big Query
- **Tools:** Airflow, MS Excel, Snowflake, Tableau, DBT, Git, Jenkins, Confluence, CI/CD

PROFESSIONAL EXPERIENCE

Kraft Heinz – Senior Data Engineer

Chicago, Illinois | Jan 2024 – Present

- Designed and Implemented ELT Pipelines using DBT and orchestrated using Dagster ensuring automated data movement
- Developed and operated modern data architecture approaches to meet key business objectives and provide end-to-end data solutions
- Developed custom handlers using python to automate data ingestion from different source systems to snowflake
- Process and analyzed the business requirements to determine project scope and derive modeling diagrams using design principles
- Designed and implemented **generic macros** to promote reusability and standardization across dbt projects, optimizing development workflows
- Mentored and trained data engineers to effectively adopt and integrate **advanced DBT and snowflake features** into their workflows, fostering team proficiency in tools
- Acted as a data owner and functional data subject matter expert for Kraft Heinz NA sales area

Wolters Kluwer – Data Engineer

Chicago, Illinois | Jun 2023 – Dec 2023

- Led the successful migration of oracle database to snowflake, resulting in a **50%** reduction in query response time
- Developed a **Python** script to automate conversion of **DDL's** from **Oracle** to **Snowflake**, reducing manual effort by **80%**
- Designed and automated a data loading workflow using **Python, Snowpipe, SQS, S3** and **Airflow** that reduced data loading time by 30% and enabled the efficient transfer of data across more than **50** schemas from **Oracle** to **Snowflake**
- Developed **DBT**-based **ELT** pipelines to transform Snowflake data, powering **30+** business dashboards
- Implemented techniques for **Optimizing** costs and queries resulting in a **30%** reduction in overall credits consumption
- Served as SME with knowledge of supported platforms and tools **Airflow, Python, AWS, Snowflake** and **DBT**

University of Cincinnati – Graduate Research Analyst

Cincinnati, Ohio | Jan 2023 - May 2023

- Spearheaded the development of healthcare system models by conducting data profiling and analysis using **SQL**
- Designed and built **Python**-based data ingestion and cleansing processes from diverse sources and third-party API's to **MS SQL Server**, boosting efficiency and reducing errors by **20%**
- Developed interactive **Tableau** dashboards and **KPIs** for clients, facilitating faster decision-making processes and improving efficiency by **30%**

Novus Platform OFC Private Ltd – Senior Data Engineer

Hyderabad, India | Mar 2020 - Aug 2022

- Designed and Implemented the RWD commercial alerts system for FDA-approved therapies that automatically transmit weekly provider-level data to pharma partners, resulting in an annual revenue generation of **\$7M USD**
- Automated **ETL** pipelines using **Airflow, AWS Glue, Glue Crawler, DynamoDB** and **Pyspark** resulting in a **40%** decrease in processing time for bringing data from various sources including **LIMS, Salesforce**, and **XIFIN** to the **Data Lake**
- Created a Python-based **NDC** lookup library with **Airflow** to streamline the Cancer Research team's data workflow
- Designed and Integrated reusable frameworks to support common services like Orchestration, Data Quality, Data Catalog and Data lineage enabling data discovery and Data Governance
- Architected Click Stream Data Ingestion using AWS Kinesis to queue and AWS Firehose to dump the data in S3
- Standardized project intake process, incident management, change management and release management using **JIRA**

Neudesic Technologies Private Limited – Data EngineerHyderabad, India | **Oct 2019 – Feb 2020**

- Led the migration of over **100** pipelines from **Control-M** and **Informatica PowerCenter** to **Airflow** on **Google Cloud**
- Collaborated with the business and leadership to create plans, milestones and roadmap for project implementation
- Instituted the use of lifecycle policies to save up to **\$500/month** in **Google Cloud Storage** spend

Virtusa Consulting Services Private Limited – Data EngineerHyderabad, India | **May 2016 - Oct 2019**

- Achieved a **20 %** reduction in runtime for daily batch jobs by optimizing **SQL** queries and **PL/SQL** functions
- Utilized **Informatica PowerCenter**, **SQL** and **Python** to implement end-to-end **ETL** pipeline for standardizing data from diverse sources (SAP, Oracle, Salesforce) comprising **Multi Terabytes** of datasets
- Developed a Tableau dashboard to monitor KPI'S such as operational activities of medical equipment delivery and related services, directly contributing to a savings of **\$ 820K USD** in operational costs
- Managed team's GitHub repositories, implemented version and process control, and conducted peer code reviews

CERTIFICATIONS

- Snowflake SnowPro Core Certification
- Databricks Certified Associate Developer for Apache Spark 3.0
- AWS Certified Data Engineer – Associate
- AWS Certified Solutions Architect – Associate
- Astronomer Certification for Apache Airflow Fundamentals