Shouvik Sharma

2829 S Wells Street, Chicago, IL-60616**|**Phone: 3124592008 **|** [shouvik19@gmail.com](mailto:shouvik19@gmail.com) | [Linkedin](https://www.linkedin.com/in/shouvik-sharma19/) | [GitHub](https://github.com/ShouvikSharma) | [Medium](https://medium.com/%40shouvik19)

# SUMMARY

Experienced Data Engineer specializing in financial data management and automation. Adept at developing data product roadmaps, implementing ETL pipelines, and driving workflow efficiencies. Skilled in cross-team collaboration and data-driven strategy implementation.

# WORK EXPERIENCE

## Data Engineer at Avant LLC, Chicago: (Aug 2021 –Present)

* Implemented robust data quality checks using SODA for real-time validation, developed custom Python scripts to rectify anomalies, and automated anomaly detection, resulting in a 30% reduction in data discrepancies and enhanced overall data reliability.
* Spearheaded reporting and automation initiatives for the Consumer Data Governance team, resulting in a 20% increase in efficiency and a 15% reduction in reporting errors.
* Coordinated development of data assets, ensuring integration into roadmap, based on short and long-term needs, leading to a 20% increase in efficiency.
* Utilized Snowflake Kafka Connector to seamlessly integrate Kafka Connect with Snowflake, facilitating data ingestion from Azure Event Hub.
* Proficient in writing complex SQL queries for ETL pipelines, providing recommendations for optimization to the DBA team, resulting in a 20% improvement in query performance. Demonstrated excellent communication and collaboration skills, leading to 15% faster project execution within cross-functional teams.
* Supported debt collection practices by developing mission-critical pipelines using PySpark and orchestration tools like Airflow, resulting in a 20% increase in debt collection efficiency and a 10% reduction in delinquency rates.

## Data Engineer Intern at CNH Industrial Inc., Racine: (Mar 2021 – Aug 2021)

* Designed and implemented efficient ETL processes using Microsoft Access, optimizing SQL queries, collaborating with cross- functional teams, automating reporting systems, and conducting performance tuning, resulting in enhanced data management and accessibility.
* Implemented and optimized data engineering solutions on Microsoft Azure, leveraging services such as Azure Databricks, Azure Data Factory, and Azure SQL Database.
* Developed and maintained data pipelines using Azure services, resulting in a 40% increase in data processing speed.

## Data Engineer at Daten Solutions Inc., Chicago*:* (May 2020 – Mar 2021)

* Developed and automated data migration pipeline from SQL Server to Snowflake using SnowSQL and SnowPipe, and further enhanced data quality by performing dimensional modeling on the migrated data.
* Developed and maintained data pipelines using Azure services resulting in a 40% increase in data processing speed.
* Automated ETL processes using Prefect (Python), enhancing data wrangling capabilities and achieving a 40% reduction in time through large-scale data conversions. Facilitated the seamless transfer of BAAN data into standardized formats for integration into Snowflake.
* Orchestrated the development of a data pipeline to stream Azure Event Hub data to Snowflake, achieving a 3-second latency for raw data ingestion and an 11-second end-to-end latency with transformation.
* Implemented Snowpark Container Services, reducing infrastructure management complexity and enabling the deployment of Kafka Connect containers with ease.

## Data Engineer – Practicum Student at Labelmaster, Chicago: (May 2020 – Dec 2020)

* Participated in cross-functional groups to develop companywide data governance reporting capabilities, contributing to a 50% reduction in reporting turnaround time.
* Optimized complex SQL scripts for quality checking of projects and populating output tables for deployment using Azure Pipelines.
* Automated hourly status report using Azure Logic Apps, saving 10 man-hours/week, thus decreasing response time for fixes and campaign failures.

## Data Engineer at Cartesian Consulting, Mumbai*:* (Apr 2018- Jul 2019)

* Designed and implemented ETL processes using Azure Data Factory, resulting in a 50% reduction in processing time.
* Formulated and implemented data governance policies, resulting in a substantial 30% reduction in data quality issues within the MariaDB environment.
* Developed dimensional data models and a MariaDB-powered data warehouse, strictly adhering to integrity and normalization rules. This infrastructure supported the creation of a campaign data-mart and a comprehensive customer one-view for marketing campaigns.

# EDUCATION

* + **MS in Computer Science and Mathematics**, Illinois Institute of Technology, **GPA: 3.8 (Aug 2019 - May2021) Related Courses**: Big Data Technologies, Applied Statistics, Database Management, Data Preparation and Analysis.
  + **MS in Statistics**, NMIMS University, **GPA: 3.35 (Jul 2016 - Apr 2018)**
  + Certifications**:** [Snowflake Pro Certification,](https://www.youracclaim.com/badges/f03d4251-13bd-4fd0-9f0d-45ff17bd718f) SAS Certified Base Programmer for SAS 9, SAS Certified Predictive Modeler

# SKILLS

* + **Programming:** Python, SQL, Scala, Java, HTML, Excel VBA (Macros).
  + **Big Data Ecosystem**: Spark, Hadoop, MapReduce, Hive, Pig, Kafka, Flume, Hbase, Microsoft Azure.
  + **Distributed Data/Computing Tools**: MapReduce, Hive, Spark
  + **Cloud Technologies**: AWS (S3, EC2, Lambda, Athena, RDS, Redshift, EMR), NoSQL, Cassandra, MongoDB, Kubernetes, Snowflake, CircleCI, Airflow, Prefect, Google Data Studio, SharePoint.
  + **Tools:** Tableau, Power BI, Azure ML, RStudio, Jupyter Notebook, DBT, Databricks, IBM-Unica, SSIS, MS Office, JIRA, Looker.
  + **Libraries:** Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras, Nltk, Gensim, Scipy, Beautiful Soup.