# Shouvik Sharma

2829 S Wells Street, Chicago, IL-60616|Phone: 3124592008 | shouvik19@gmail.com | Linkedin | GitHub | Medium

## SUMMARY

Results-driven Data Engineer with expertise in designing and implementing scalable data solutions. Proficient in Python, SQL, and big data technologies, with a strong background in ETL/ELT processes, data pipelines, and data quality management. Passionate about building reliable and trustworthy web-scale data products and continuously improving data systems. Seeking to leverage my skills in data engineering to contribute to Netflix's high-impact projects and drive data-driven decision-making.

## WORK EXPERIENCE

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

- Engineered efficient and scalable data pipelines to process structured and unstructured data, improving data processing efficiency by 20%.
- 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.
- Developed and optimized complex SQL queries for ETL pipelines, improving query performance by 25%.
- Implemented Apache Airflow for workflow orchestration, resulting in a 30% increase in pipeline reliability.
- Utilized Azure Databricks and Spark for large-scale data processing, building and optimizing Spark-based data pipelines, and implementing machine learning models for predictive analytics.
- Built data quality systems using best-in-class techniques and technologies, enhancing data reliability and enabling teams to deliver more business value.
- Developed Tableau dashboards for marketing attribution, contributing to a 4% increase in application rates. Extracted insights from marketing data to create source attribution funnel dashboard in Looker, helping business stakeholders to improve customers' application experience.
- Collaborated with cross-functional teams to implement data quality checks using SODA, further utilizing AWS services including Redshift and S3 for large-scale data processing and storage.

**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.
- Collaborated with data scientists to define data requirements, implemented ETL processes for seamless data integration, optimized SQL queries, and facilitated the development of predictive models, fostering a synergistic environment for data-driven insights.

**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.
- Created Tableau dashboards to explain variation in success Metrics and Time Series Analysis to higher management.
- Automated reporting process using Excel VBA (Macros) and MySQL maintaining accuracy and saving ~ 75% of time, maintained version control Git, Mercurial, SVN.

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

- Involved in designing databases, data marts, E-R model for OLTP and multi-dimensional model for OLAP using SnowSQL.
- Minimize technical debt in managing new data requirements, ensuring scalability and sustainability of data solutions, resulting in a 20% reduction in maintenance time within six months.
- Automated hourly status report saving 10 man-hours/week, thus decreasing response time for fixes and campaign failures.

**Big Data Developer at Cartesian Consulting, Mumbai*:***                                        **(Apr 2018- Jul 2019)**

- Designed and implemented ETL processes using DynamoDB, resulting in a 50% reduction in processing time.
- Developed dimensional data models and a MariaDB-powered data warehouse, supporting the creation of a campaign data-mart and a comprehensive customer one-view for marketing campaigns.
- Developed data ingestion frameworks to integrate data from various sources, improving data availability for analytics.
- Implemented data transformation and aggregation processes, enabling more efficient reporting and analysis.
- Collaborated with business stakeholders to understand data requirements and deliver solutions that met their needs.

## 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, Databricks Certified Associate Data Engineer

## 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**: Azure Data Factory, Azure Blob, NoSQL, Cassandra, MongoDB, Kubernetes, Snowflake, CircleCI, Airflow, Prefect, Google Data Studio, Azure Synapse Analytics, DynamoDB.
- **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.