

Shouvik Sharma

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

SUMMARY

Experienced Data Engineer with a track record in driving the development and management of data product development roadmaps aligned with business and data strategy. Skilled in planning and coordinating development of data assets, executing product development within complex categories, and driving collaboration across cross-functional teams. Proficient in product roadmap development, data-driven strategy decisions, and influencing stakeholder consensus.

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.
- Extracted insights from marketing data to create **source attribution funnel** dashboard in Looker, this helped business stakeholders to improve customers' application experience, and increase application rate by 4%.
- Managed data ingestion, validation, and Power BI reporting on Consumer Data Governance processes, ensuring a 95% accuracy rate in reporting metrics.
- Built a refinance product data pipeline for existing loan customers based on their TransUnion credit scores and existing features, contributing to a 10% increase in the overall refinance book.
- Coordinated enhancements to Domain Governance content on SharePoint Online and managed entitlements as SharePoint Online site owner, resulting in a 40% increase in user engagement.
- 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 Management Specialist Intern at CNH Industrial Inc., Racine:

(Mar 2021 – Aug 2021)

- Designed and implemented efficient ETL processes using MS Azure, 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 MS SQL to DynamoDb 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)

- 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 DataFlow.
- Automated hourly status report saving 10 man-hours/week, thus decreasing response time for fixes and campaign failures using DataProc.

Big Data Developer 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 using Dataflow.
- Formulated and implemented data governance policies, resulting in a substantial 30% reduction in data quality issues within Google BigQuery.
- Developed dimensional data models and a Google Query 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](#), 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.