

Shouvik Sharma

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

SUMMARY

Experienced Data Engineer with over 5 years of comprehensive work experience in Data Engineering, Marketing Analytics and Business Intelligence in banking and retail domains. Ability to solve complex business problems using ETL, Data Mining, Machine Learning & Data Warehousing concepts.

WORK EXPERIENCE

Data Engineer at Avant LLC, Chicago:

(Aug 2021 –Present)

- Designed and developed a scalable data warehouse using Azure Blob Storage, Data Lake, and Synapse to store and manage large volumes of data.
- Extracted insights from marketing data to create **source attribution funnel**, this helped business stakeholders to improve customer application experience, and increase application rate by 4%.
- Migrated marketing campaign pipelines from conventional marketing tools like Responsys to Segment using API supported python libraries.
- Created Databricks job workflows which extracts data from dremio database and uploaded the files to sftp using pyspark and python.
- Build out refinance product data pipeline for existing loan customers based on their transunion credit scores and existing features, this increased the overall refinance book by 10%.
- Developed integration solutions using Logic Apps and APIs to connect disparate systems and automate data flows.
- Understand and write complex sql queries as a source of ETL pipelines, providing required recommendations to the DBA team such as adding an index on the frequently used tables, ultimately to improve query optimization.
- Understand complex SQLs running as part of ETL Loads and make recommendations to the DBA for index creation and partition changes to improve performance.
- Supported debt collection practices of Avant LLC., by developing mission-critical pipelines that delivered hourly customer information to external collection agencies like Livevox, by designing the data pipelines that implemented data cleaning, data processing, and data delivery using Pyspark and orchestration tools like Airflow.

Data Science Intern at CNH Industrial Inc., Racine:

(Mar 2021 – Aug 2021)

- Developed predictive analytics and statistical forecasting models such as LSTM, ARIMA, neural networks etc. to predict quarterly sales, achieved an accuracy of 73% for the back testing of the year 2020.
- Optimized pipeline implementation and maintenance work using Databricks workspace configuration, cluster and notebook optimization.
- Development of tools to allow process automation, analysis & corrective action implementation by the business.

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), making it easier to wrangle data sets and reducing time by as much as 40% by performing large-scale data conversions, and transferring 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.
- Optimized complex SQL scripts for quality checking of projects and populating output tables for deployment using Azure Pipelines.
- 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 Azure Data Factory, resulting in a 50% reduction in processing time.
- Implemented data governance policies resulting in a 30% reduction in data quality issues.
- Applied K-means clustering in Python for segmentation of customers, comparing it with RFM (Recency, Frequency and Monetary Value) analysis for improved campaign targeting.
- Developed dimensional data models and data warehouse adhering to integrity and normalization rules to support campaign data mart and customer one view for marketing campaigns. Wrote complex SQL queries (multiple joins, CTE's, subqueries).

EDUCATION

- MS in Computer Science and Mathematics**, Illinois Institute of Technology, **GPA: 3.8**

(Aug 2019 - May 2021)

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:** SQL, Python, R, SAS, Pyspark, HTML, Excel VBA (Macros), Agile Methodology, PostgreSQL, MySQL, Shell.
- Big Data Ecosystem:** Spark, Hadoop, MapReduce, Hive, Pig, Kafka, Flume, Hbase, Microsoft Azure.
- Cloud Technologies:** AWS (S3, EC2, Lambda, Athena, RDS, Redshift, EMR), NoSQL, Cassandra, MongoDB, Kubernetes, Snowflake, CircleCI, Airflow, Prefect, Google Data Studio, Azure Synapse Analytics.
- Tools:** Tableau, Power BI, Azure ML, RStudio, Jupyter Notebook, SAS E-Miner, SAS CI, IBM-Unica, SSIS, MS Office, JIRA, Looker.
- Libraries:** Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras, Nltk, Gensim, Scipy, Beautiful Soup.
- Datasets:** HTTP, HTML, XML, JSON