Shouvik Sharma

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

SUMMARY

Results-driven Data Engineer with a proven track record in designing and implementing scalable data solutions. Specialized in leveraging AWS, Ab Initio, and Snowflake for ETL development, data warehousing, and cloud-based storage. Seeking to apply my expertise in data engineering to contribute to innovative projects at Avant, LLC.

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.
- Led end-to-end data transformation workflows with dbt (data build tool), achieving a 20% improvement in analytics process efficiency and ensuring reproducibility.
- 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%.
- Migrated marketing campaign pipelines from conventional marketing tools like Responsys to Segment using API supported python libraries.
- 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.
- 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.
- Proficient in writing complex SQL queries as a source for ETL pipelines, providing recommendations to the DBA team for optimization, including index creation and partition changes using Teradata SQL Assistant.
- Supported debt collection practices by developing mission-critical pipelines using PySpark and orchestration tools like Airflow.

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

- Designed and implemented efficient ETL processes using Ab Initio, 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.
- Developed UNIX shell scripts for data dependency analysis and data lineage utilizing Ab Initio on cloud & AWS data storage using S3/Redshift.

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.
- Developed UNIX shell scripts for data dependency analysis and data lineage utilizing Ab Initio on cloud & AWS data storage using S3/Redshift.

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.

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

- Designed and implemented ETL processes using AWS services including S3, EC2, and RDS.
- Collaborate with project team members (Product Managers, Architects, Analysts, Developers, Project Managers, etc.) to ensure development and implementation of new data driven business solutions.
- Formulated and implemented data governance policies, resulting in a substantial 30% reduction in data quality issues within the MariaDB environment.

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:** Python, SQL, Scala, Java, HTML, Excel VBA (Macros).
- **Big Data Ecosystem**: Spark, Hadoop, Kafka, EMR, Snowflake, Ab Initio
- **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, Azure Synapse Analytics.
- **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.
- **Datasets:** HTTP, HTML, XML, JSON