Shouvik Sharma

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

SUMMARY

Experienced Data Engineer with extensive expertise in building and optimizing data pipelines and production data models using technologies such as Big Query, GCP, AWS, dbt, and Python. Expertise in building resilient, performant code using Java, Scala, and Python. Skilled in developing ETL pipelines, optimizing SQL queries, and implementing big data technologies such as Hadoop, Spark, and Kafka.

WORK EXPERIENCE

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

- Engineered efficient and scalable data pipelines using Databricks, dbt and Apache Spark, to process structured and unstructured data, improving data processing efficiency by 20% over a six-month period.
- Implemented robust data quality checks using SODA and custom Python scripts for real-time validation and anomaly detection, resulting in a 30% reduction in data discrepancies within three months.
- Developed and optimized complex SQL queries in DBT for ETL pipelines, improving query performance by 25% and reducing runtime from 4 hours to 3 hours on average.
- Implemented Apache Airflow for workflow orchestration, resulting in a 30% increase in pipeline reliability and reducing system downtime by 50% within the first quarter of implementation.
- Utilized Airflow, Apache Spark, EMR and Amazon SQS to build and optimize data pipelines for large-scale data processing, enabling the implementation of machine learning models in PySpark that improved predictive accuracy by 15%. SQS facilitated reliable message queuing between pipeline stages, enhancing fault tolerance and scalability.
- Developed Tableau dashboards for marketing attribution, contributing to a 4% increase in application rates over a three-month period. Created source attribution funnel dashboards in Looker, helping business stakeholders improve customer application experience and increase conversion rates by 7%.
- 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, resulting in a 35% improvement in data accuracy and a 20% reduction in processing time.

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

- Designed and implemented ETL processes using Microsoft Access and SQL that reduced the data processing time by 30%.
- Collaborated with data scientists to define data requirements, implemented ETL processes using Databricks, optimized SQL queries, and developed predictive models using Python and Databricks, enhancing data-driven insights and improving model accuracy by 10%.

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.
- 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)**

- Designed databases and data marts, developed E-R models for OLTP, and created multi-dimensional models for OLAP using SnowSQL, leading to a 20% reduction in query processing time.
- 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 and SQL, reducing data processing time by 50%.
- Developed dimensional data models and a data warehouse powered by MariaDB, supporting the creation of a campaign data-mart and a comprehensive customer one-view, which improved marketing campaign efficiency by 20%.
- Implemented data transformation and aggregation processes using R and MariaDB, improving reporting and analysis efficiency by 25%.

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, Big Query.
- **Distributed Data/Computing Tools**: MapReduce, Hive, Spark.
- **Cloud Technologies**: GCP (BigQuery, Dataflow, DataProc, Cloud Composer, Cloud Storage, GKE), AWS (EMR, Redshift, S3), Snowflake, CircleCI, Airflow, Prefect, Google Data Studio, Azure Synapse Analytics, DynamoDB, EMR, Terraform.
- **Tools:** Tableau, Power BI, Azure ML, RStudio, Jupyter Notebook, DBT, Databricks, IBM-Unica, SSIS, MS Office, JIRA, Looker.