

Shubham Dubey

(973) 979 – 4949 | Shbhamdbey@gmail.com | <https://www.linkedin.com/in/dubshubham/> | East Newark, NJ

Experience

Covetrus, Inc

Data Engineer

Oct 2021- Present

- Managed and optimized 135 Snowflake pipelines, processing 10+ billion rows daily, leading to a 15% cost reduction and enhanced data reliability.
- Streamlined Kafka data into actionable insights for orders, finances, and marketing, improving decision-making.
- Modernized ETL processes by transitioning to ELT methodologies, integrating real-time updates from SQL and NoSQL databases with Debezium connectors, improving data availability and accuracy.
- Engineered OAuth2.0 security protocols for NetSuite API, safeguarding 60,000+ daily calls across multiple endpoints, enhancing data integrity and fostering seamless system integration.

Pfizer

Data Engineer

Sep 2020 - Jan 2021

- Refactored ETL processes, with multiple data sources, enhancing data readiness for machine learning models by 5%.
- Optimized cloud storage solutions for high-quality data ingestion from multiple REST APIs, enhancing data availability.
- Collaborated with cross-functional teams to develop Tableau dashboards with interactive maps, reducing data latency up to 63%.

New Jersey Institute of technology

Data Engineer

Mar 2020 - May 2021

- Saved 600+ hours annually by creating SQL scripts for data accuracy and generating data extracts, automating daily reporting with Appworx scheduler.
- Developed complex SQL queries and stored procedures to map multiple database schemas across 40+ departmental tables, ensuring data validation through visual reports.
- Improved data precision by 75% by tuning SQL performance using Oracle hints, enhancing ETL processing and storage architecture.
- Collaborated with cross-functional teams to evaluate data quality issues and optimized end-to-end operational efficiency by 15% using unit testing.

Projects

Equity Data Pipeline [Python API, MongoDB]

- Developed ETL pipelines for ad-hoc data ingestion from the Y-Finance API, leveraging batch data processing and streaming.
- Automated structural data transformation with Python scripts, reducing manual workload by 80%.
- Enhanced data accuracy through statistical and distance-based cleaning techniques, processing over 150+ data batches.

Annual Financial Trend Analysis [Snowflake, Dimension modelling]

- Designed and implemented a data mart for Volvo Car USA's financial reports, enabling efficient business intelligence reporting and analysis.
- Utilized Slowly Changing Dimension (SCD) methodology to manage historical data and track changes in financial KPIs over the past 15+ years.
- Performed Level of Detail (LOD) analysis on monthly sales trends, identifying seasonal patterns and providing actionable insights for targeted marketing strategies.

Pipeline Guardian [Splunk, CloudWatch]

- Monitored metrics: data volume, latency, quality, pipeline status, duration, and errors for robust data processing.
- Created Splunk dashboard and CloudWatch Alarms to identify issues, compare, and optimize performance.

Skill

Java, Python, SQL, Kafka, Spark, Snowflake, AWS, Azure (Data Factory, Synapse), CI/CD

Education

New Jersey Institute of Technology

Master of Science in Computer Science

Newark, NJ

May 2021

- Coursework:** Cloud Computing, Data Mining, Data Structures and Algorithms, Data Management, System Design
- Award:** Won Hackathon organized by Meta, secured 1st prize (June 2021)