Selenge Tulga

selengetu@gmail.com | Portfolio | Linkedin | Github | **Austin TX | Willing to relocate**

Data Engineer & Machine Learning Specialist with 6+ years of experience building scalable ETL pipelines, cloud-native solutions, and predictive ML models. Expertise in big data processing, automation, and cloud engineering (AWS, GCP, Azure).

Technical Skills

Programming: Python, SQL (PL/SQL, T-SQL), Java
Data Engineering: Apache Spark, Kafka, Airflow, dbt, Snowflake, BigQuery, Redshift, Iceberg,Saas,Jira
Databases: Oracle, PostgreSQL, MySQL, MongoDB, DynamoDB, Trino
Cloud & DevOps: AWS (Glue, Lambda, Kinesis, EMR, SageMaker, S3, Redshift), GCP, Azure, Docker, CI/CD, Jenkins
BI & Visualization: Power BI, Tableau, QuickSight, Grafana
ML & Analytics: MLflow, TensorFlow, Scikit-Learn, SageMaker, LLMs

Work experience

Data Engineer Intern- URMC Motion Lab, Rochester NY   **May 2024 - Sep 2024**
- Automated data pipelines, saving 10+ hours/week and reducing errors by 30%.
- Developed Power BI/Tableau dashboards with DAX queries for healthcare analytics.
- Built scalable pipelines to process 100,000+ patient records daily.

Data Engineer - Zevtronic LLC , Ulaanbaatar, Mongolia   **Jan 2020 - July 2023**
- Designed **30+ ETL pipelines**, integrating data from 10+ sources, increasing processing speed by 70%.
- Applied star & snowflake schemas to structure data marts, improving query performance by 40%
- Led knowledge-sharing sessions on SQL tuning & Airflow DAG optimization, mentoring 5+ junior engineers.

Data Engineer - Ulaanbaatar Railway JVC , Ulaanbaatar, Mongolia   **Dec 2019 - July 2023**
- Optimized SQL queries for 500,000+ railway records, reducing latency by 50%.
- Enhanced batch data updates, improving Trino query accuracy to 98%.
- Developed star schema-based data models for railway signal data, reducing query latency by 50%.

Software engineer - Ulaanbaatar Railway JVC   **Jul 2017 - Dec 2019**
- Developed scalable backend systems for locomotive repair scheduling, increasing revenue by **300%.**
- Built data models ensuring 99% accuracy and 50,000+ transactions daily.
- Optimized SQL query indexing, reducing query execution time by 40%.

Projects

LLM-Powered Flight and Train Delay Prediction (ongoing)
- Developed a data pipeline to compare flight and train delays across the U.S., integrating real-time and historical +15 data sources.
- Building an ML pipeline integrating real-time flight data (FlightAware API) & weather conditions
- Designing interactive 15 dashboards using Tableau and Power BI to visualize insights for stakeholders.

Tweet Sentiment Analysis (Databricks Spark)
- Built a real-time Spark Streaming pipeline, processing 200K+ tweets for sentiment analysis.
- Designed a Delta Lake architecture for optimized storage and performance.
- Integrated a Hugging Face NLP model via MLflow, improving sentiment classification accuracy by 65%.

Butler Till – Capstone Project
- Explored synthetic data techniques (CTGAN, Prophet, Gaussian Copula) to generate high-quality synthetic datasets for analytics use cases.
- Applied SARIMAX & MMM models to evaluate synthetic datasets, improving forecasting accuracy by 12%.

Education

Master of Data Science - University of Rochester   **Aug 2023 - Dec 2024**
Bachelor of Computer Science - Mongolian University of Science and Technology   **Aug 2013 - Jun 2017**

Certification

AWS	Data Engineer – Associate	Machine Learning Associate	Machine Learning Specialty
Databrick:	Data Engineer – Professional	Data Engineer – Associate	

DE: Zach Wilson's DataExpert Data engineering Bootcamp (2025) (Top 0.62% Achiever)