Selenge Tulga

selengetu@gmail.com | Portfolio | Linkedin| Github | **Austin TX | Remote Work/Relocation: Open**

Education

Master of Data Science - University of Rochester   **December 2024**
Bachelor of Computer Science - Mongolian University of Science and Technology   **June 2017**

Professional experience

Data Science Intern - URMC Motion Lab   (Rochester, NY)   **May 2024 -Sep 2024**
- Automated data pipeline with Python scripting, saving the team over **10 hours per week** and reducing manual errors by **30%**.
- Built and maintained scalable data pipelines using **AWS Glue** and Python, processing **100,000+ patient records daily**, ensuring **100% regulatory compliance** and achieving **zero data inconsistencies**.

Data Engineer - Zevtronic LLC   **Jan 2020 -July 2023**
- Designed and implemented scalable ETL pipelines using **AWS Glue and Lambda**, automating data integration and transformation from **10+ sources**, increasing data processing speed by **70%**.
- Utilized **Redshift Spectrum** to enable efficient querying of datasets exceeding **1TB**, optimizing reporting performance by **40%** and reducing query run time by **20 minutes**.
- Developed robust backend solutions using Python, orchestrating workflows and optimizing cloud storage across **S3, Redshift, and Snowflake**, ensuring **99.9% uptime** for analytics systems.
- Modeled data with dimensional schemas and star schemas using SQL, standardizing analytics across **5+ teams** and reducing reporting inconsistencies by **15%**.

Data Engineer - Ulaanbaatar Railway JVC   **Jan 2020-July 2023**
- Designed and implemented real-time ETL workflows using **Apache Kafka and Flink**, enabling seamless processing of **500,000+ daily railway signal records**, reducing data latency by **50%**.
- Automated cloud infrastructure provisioning using **Terraform** and integrated CI/CD pipelines via **Jenkins**, cutting deployment time by **34%** and minimizing manual intervention.
- Built pipelines leveraging **Apache Iceberg** for batch data updates, improving data update performance by **25%**, and ensured compatibility with **Trino** for querying datasets exceeding **500GB**, achieving query accuracy of **98%**.

Software engineer - Ulaanbaatar Railway JVC   **July 2017-Dec 2019**
- Increased revenue by **300%** by developing scalable backend systems for locomotive repair scheduling and passenger booking applications using **Java**, **Oracle**, and **PHP**, reducing downtime by **40%**.
- Collaborated on building data models for operational workflows, ensuring data accuracy of **99%** and scalability to support **50,000+ transactions daily**.
- Maintained source code repositories using **Git**, facilitating efficient tracking of feature development.

Independent Projects

Tweet Sentiment Analysis using Databricks Spark
- Transformed **200000 tweets** using **PySpark** and **Delta Lake** architecture for real-time analysis.
- Designed a **real-time sentiment analysis pipeline** using **Spark Streaming** and Medallion architecture.
- Deployed a pre-trained **HuggingFace model** via **MLflow** for scalable sentiment classification.

Certification

AWS: Certified Data Engineer – Associate, Machine Learning Associate, Machine Learning Specialty
Databricks: Certified Data Engineer – Professional, Data Engineer – Associate

Skills

Programming Languages:SQL, Python, Java, PHP, R
Databases: PostgreSQL, MySQL, MongoDB (NoSQL), Redshift, DynamoDB, Oracle, PL/SQL, Presto
Data Engineering Tools: Apache Spark, Flink, Airflow, Iceberg, Trino
AWS Tools: Glue, Lambda, Kinesis, Redshift, S3, EMR, Athena, Sagemaker, QuickSight