

Selenge Tulga

selengetu@gmail.com | [Portfolio](#) | [Linkedin](#) | [Github](#)

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 - UPMC Motion Lab (Rochester, NY)

May 2024 -Sep 2024

- Automated data pipeline with Python scripting, saving the team over 10 hours of manual work per week
- Built and maintained scalable data pipelines to process **100,000+ patient records**, ensuring **data accuracy** and regulatory compliance for analytics and reporting.

Data Engineer - Zevtronic LLC

Jan 2020 -July 2023

- Built scalable **ETL pipelines** to integrate and transform healthcare data from multiple sources (EHR systems, APIs, relational databases), improving **data processing speed by 70%**.
- Implemented **Apache Kafka** for real-time data streaming and **Apache Spark** for distributed data processing to enable high-performance workflows.
- Designed robust backend solutions using **Python** and optimized **cloud storage** (Amazon S3, Redshift) to enhance data availability and reporting efficiency.

Data Engineer - Ulaanbaatar Railway JVC

Jan 2020-July 2023

- Automated cloud infrastructure provisioning using **Terraform**, reducing AWS deployment time by **34%** and improving system reliability.
- Designed and implemented **real-time ETL workflows** with **Apache Kafka** and **Apache Flink**, enabling seamless processing of railway signal data and operational logs.
- Developed and optimized **CI/CD pipelines** using **GitHub Actions**, automating testing, deployment, and monitoring, increasing development efficiency by **34%**.

Software engineer - Ulaanbaatar Railway JVC

July 2017-Dec 2019

- Ensured **99.9% system uptime** by managing and maintaining datasets from **40+ locomotive signal systems** with rigorous testing protocols.
- Increased revenue by **300%** by developing scalable backend systems for locomotive repair scheduling and passenger booking applications using **Java, PHP, and SQL**.
- Collaborated with teams to implement secure, high-performance data storage solutions using **PostgreSQL** and **AWS RDS**.

Independent Projects

Tweet Sentiment Analysis using Databricks Spark

- Transformed **200K 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

[Databricks Certified Data Engineer Associate](#)

[AWS Certified Data Engineer Associate](#)

[Databricks Certified Data Engineer Professional](#)

[AWS Certified Machine Learning Specialty](#)

Skills

Programming Languages: Python, Java, SQL, PHP, R

Databases: PostgreSQL, MySQL, MongoDB (NoSQL), Redshift, DynamoDB, Oracle, PL/SQL, Presto

Cloud Platforms: AWS (EC2, S3, Lambda, Glue, Athena, Step Functions), Azure, GCP

Data Engineering Tools: Apache Spark, Flink, Airflow, Terraform

Data Visualization: Tableau, Power BI, Amazon QuickSight

Other: ETL/ELT, Data Governance, Data Quality, REST APIs, Github, Jenkins, LLMs