ematiq career

WE ARE LOOKING FOR

Data Engineer

At Ematiq, we develop high-performance, real-time systems that support trading across numerous global platforms. These trading systems generate terabytes of data, which are continuously fed into our advanced data warehouse. We are looking for an engineer who is ready to master our domain and take ownership of the infrastructure that collects, stores and serves this data.

The Role & Our Team

As a Data Engineer, you will make sure data from our distributed trading systems reaches the warehouse completely, correctly and on time. You will keep our Athena warehouse fast and cost-efficient, build CDC and replication from our Postgres databases, and design how their data moves from hot to cold storage. You will work with the engineering teams that build our trading systems, and provide the infrastructure our Quant and Trading departments rely on.

You will join a Data team consisting of a Data Engineer and three Data Analysts. We value a supportive environment built on code reviews and active knowledge sharing. Because of the complexity and high responsibility tied to trading data, we value sharp attention to detail, technical precision, and an innate curiosity to dive deep into intricate systems.

We endorse the pragmatic use of Agentic LLMs - we take the meaningful application of AI seriously rather than treating it as a buzzword.

Daily Responsibilities

- Warehouse Performance: Optimize queries, table layouts and partitioning in our Apache Iceberg warehouse on AWS Athena, and keep query costs under control.
- Data Ingestion: Build and maintain reliable ingestion from our distributed systems: Kafka streams, change data capture with Debezium, and Postgres logical replication.
- Database Operations: Look after our PostgreSQL databases, including replication and hot/cold storage of historical data.
- Infrastructure as Code: Manage our AWS data infrastructure (S3, Glue, Athena) with Terraform.
- Orchestrate & Monitor: Orchestrate the data pipelines, investigate data gaps and delays, and make sure backfills and reprocessing produce correct results.

Core stack

We don't expect you to know every technology on this list. Understanding the concepts above, such as collecting data from distributed systems, query performance, replication and storage design, matters more to us than experience with any specific tool. You will learn the rest of the stack with us.

- CAWS (Athena, Glue, S3)
- Apache Iceberg
- Kafka/Kafka Connect
- PostgreSQL
- Debezium
- **Spark**
- **Python**
- Terraform
- Dagster

Send your CV to **jobs@ematiq.com**