

## DATA ENGINEER

Established in 1963, Tokyo Electron (TEL) is a leading global company of innovative semiconductor and flat panel display production equipment. We are committed to empowering the newest advancements in technology and to the progress of our future equipment, providing innovative tools and always supporting our global customers as a Trusted Partner.

In Europe, since 1994 we have expanded operations and we now operate in 9 Countries with more than 500 employees. Our Business is focused on Sales & Service supporting and developing Customers' needs across Europe. Above all, our employees remain the most important element for our continued growth and success.

### About the role:

A new Fixed Term position has recently become available for a Data Engineer within TELMS operations in Santry, Co. Dublin. This position will provide a high quality developer service and support across the business with a particular emphasis on data engineering activities such as data collection, migration and cleansing; in addition to database design, performance tuning and distribution of applications for data engineering and analyses.

### General requirements & Core Skills;

The core requirements of this position is having a deep understanding of database architecture and the working principles of supporting server architectures and its various components. The ideal candidate will have a keen interest in this aspect of development. The following core skills are a requirement of this position;

- Python(Anaconda)
- SQLAlchemy
- Read/Load Data
- Jupyter Notebooks
- Pandas
- Data Filtering
- Pivot Tables
- Import/Export Formats
- Chunking
- Importing from files
- Exporting to files
- JSON
- Read/Save/Transmit/Transform/Create
- Database
- PostgreSQL
- SQL/PSQL
- Database/SQL Query optimizations
- Performance Tuning
- Backup/Restore
- Replication
- Views/Materialized Views
- MS SQL Server
- Read
- MySQL etc
- Read
- Data Engineering
- ETL - Extract, Transform, Load
- Git
- Create a new repository
- Clone
- Branch/Merge
- Pull/Push/Update
- Webserver
- NGinX/Apache2

- Linux
- Ubuntu/Debian Server
- Shell Scripting
- Cron Scheduler
- Services
- LXC/LXD

**Other desirable skills but not essential skills include;**

- Machine Learning
- Scikit learn
- Tensorflow
- Keras
- Keras GPU - CUDA, cuDNN
- Statistical Analyses

**Python (Anaconda)**

A lot of the work revolves around developing python scripts to parse raw data, extract metrics for analytics and scheduling these scripts to run at regular intervals to augment existing datasets with derived analytics. It is essential to be proficient in the crafting of re-usable code in a collaborate environment as part of a data science team.

A well-rounded proficiency in Python is required with a preference for the Anaconda distribution for data science. Experience using SQLAlchemy with demonstrated abilities to read and load datasets to and from databases using SQLAlchemy libraries. In conjunction with SQLAlchemy and working knowledge of Pandas is required for data manipulation, cleaning, sorting and filtering and creating of Pivot Tables. Working knowledge of the various Import/Export Formats for data frames and the use of chunking for reading, leading and saving huge datasets. Demonstrated capabilities for the importing from files and exporting to propriety file formats to retain structured intermediate datasets including JSON formats for Read/Save/Transmit/Transform/Create etc. A thorough working knowledge of Jupyter Notebooks and formats for documentation of data exploration and algorithm development.

**Databases**

Second to the ability to ETL (Extract, Transform, and Load) data, is the ability to design and create meaningful relational tables with a balance in abstraction and query performance for huge datasets. It is essential to understand table relations and the impact of joining tables and performance trade-offs to load and generate metrics from a dataset with hundreds of millions of records. Knowing and understanding the use of view or materialized views. A proper working understanding of relational and time series databases is required with a preference for experience in PostgreSQL using SQL/PSQL. A working knowledge of Database/SQL Query optimizations, performance Tuning and database Backup/Restore procedures is a must. In addition knowledge of replication, MS SQL Server and MySQL would be advantageous.

**Data Engineering**

Demonstrated abilities on ETL (Extract, Transform, and Load). Be prepared to solve a sample use case presented during the screening process. Create your algorithm in plain English.

**GIT**

- A high level of git proficiency is required. You should be able to perform all of the following task from a command shell
- Create a new repository
- Clone from an existing repository on a git server
- Create Branches and Merge branches
- Create tags
- Pull Update and Push
- You must be able to clearly explain what all of the above steps do and achieve

**Webserver**

The dashboard application is built of a web server on Linux and a working knowledge of web servers is required including Nginx, Apache2 and the basics of HTTP or HTTPS, style sheets and web pages. Knowledge of the inner workings of Grafana will be advantageous.

**Linux**

All of the above is built around the latest LTS Ubuntu distributions. A working knowledge of Ubuntu/Debian Servers would be a huge advantage. A working knowledge of basic Shell Scripting is required and the use of a Cron Scheduler to launch Python scripts with is also useful to have a basic understanding of Linux Services and Linux containers LXC and LXD.

### **Machine Learning**

On top of the above range of skills, the ideal candidate should be able to use Keras and Tensorflow, construct a Neural network tailored to solve specific problems and run the neural network in and interpret the results of the training data and make adjustments as required to improve the models. Once a good practical model is achieved, this model needs to be implemented in a production environment to process real time data and generate prediction of events and failures.

### **Qualifications:**

- Degree in data analytics, data engineering, software engineering, or equivalent with 3+ or more years of practical experience
- Strong analytical & problem-solving skills.
- Pro-active with an ability to work with minimal supervision.
- Excellent written and verbal communication skills.
- Ability to work as part of a high performing team.
- High level of attention to detail with demonstrated capability of following procedures closely.
- Good planning & organisation skills.

All employees are expected to demonstrate flexibility and adaptability to Tokyo Electron Europe (Irish Branch) business requirements and carry out other work that falls within the scope of their abilities as requested by TEE Limited. This is likely to involve visits to, or periods working at, any TEE office.

A positive Team-working approach and adherence to Safety Policies and Procedures is essential. A professional telephone manner, good level of interpersonal and self-presentation skills are pre-requisites.

In support of our mission statement to deliver QUALITY, CUSTOMER SATISFACTION AND STAFF SATISFACTION, employees will be expected to support colleagues and management as required and to contribute ideas and suggestions in our search for CONTINUOUS IMPROVEMENT.

Tokyo Electron offers an excellent, comprehensive benefits and compensation package. For immediate consideration please apply via [tee.hr@europe.tel](mailto:tee.hr@europe.tel)

For more information about TEL please click here [www.tel.com](http://www.tel.com)