Data Engineer

Job Purpose

Analyzing, designing, developing and managing the infrastructure and the data that feeds Data Science models. The Data Engineer is expected to be in charge of the whole lifecycle of the datasets, including updates, backups, synchronization, and policy access.

Job Responsibilities (List the major responsibilities/outcomes expected of this role)

- **Managing the lifecycle (from data collection to archive) of ML/DL datasets and ensure their usability for Nielsen's Data Scientists.**
- **Design, build and integrate data from various sources.**
- Design ETL pipelines with scripted components.
- Optimize data workflows, choosing the most cost-efficient approach.
- Automate the management of recurrent task in the pipeline.
- Perform feasibility studies/analysis with a critical point of view.
- Support and maintain (troubleshoot issues with data and applications).
- Develop technical documentation for applications, including diagrams and manuals.
- Work on many different software challenges always ensuring a combination of simplicity and maintainability within the code.
- Contribute to architectural designs of large complexity and size, potentially involving several distinct software components.
- Working closely with data scientists and a variety of end-users (across different cultures) to ensure technical compatibility and user satisfaction.
- Work as a member of a team, encouraging team building, motivation and cultivate effective team relations.

Role Requirements

E=essential, P=preferred.
E - Bachelor's degree in computer engineering.
P - Master's degree in data engineering or related.
E - Demonstrated experience and knowledge in **Big Data and NoSQL databases.**
E - Demonstrated experience and knowledge in **Object-Oriented Programming**.
E - Demonstrated experience and knowledge in **distributed systems**.
E - Proficient in programming languages: **Python.**
E - Experience designing and implementing **data warehouses**.
E - Experience developing **ETL** pipelines.
E - Experience working with distributed storage systems in the cloud (Azure, GCP or AWS).
P - Experience managing deep learning datasets.
P - Experience managing **Cassandra**.
P - Experience working with **Spark**.
P - Experience implementing CICD pipelines for automation.
E - Experience in the use of collaborative developing tools such as Git, Confluence, Jira, etc.
E - Problem-solving capabilities.
E - Strong ability to analyze and synthesize. (Good analytical and logical thinking capability)
E - Proactive attitude, resolutive, used to work in a team and manage deadlines.
E - Ability to learn quickly.
E - Agile methodologies development (SCRUM/KANBAN).
E - Minimal work experience of 3-4 years with evidence.
E - Ability to keep fluid communication written and oral in English, both written and spoken.