Data Engineer

Description

As a data engineer at **[add company name]**, you will be responsible for maintaining, expanding, and optimizing our data warehouse, data pipeline, and data products.
The data engineer will support multiple stakeholders, including software developers, database architectures, data analysts, and data scientists to ensure an optimal data delivery architecture. **[Add a thorough description of the different types of projects and data sources, and outline the specific deliverables a candidate is expected to deliver.]**
The ideal candidate possesses strong technical abilities to solve complex problems with data.

Responsibilities

- Design, build and maintain batch or real-time data pipelines in production.
- Maintain and optimize the data infrastructure required for accurate extraction, transformation, and loading of data from a wide variety of data sources.
- Automate data workflows such as data ingestion, aggregation, and ETL processing.
- Prepare raw data in Data Warehouses into a consumable dataset for both technical and non-technical stakeholders.
- Partner with data scientists, functional leaders in sales, marketing, and product to deploy machine learning models in production.
- Build, maintain, and deploy data products for analytics and data science teams on cloud platforms *(e.g. AWS, Azure, GCP)*.
- Leverage data controls to maintain data privacy, security, compliance, and quality for allocated areas of ownership.
- **[If necessary, add additional job requirements]**

Experience

- Bachelor's degree in Computer Science or equivalent practical experience.
- *[X+ years of relevant working experience.]*

Minimum Qualifications

- Advanced SQL skills and experience with relational databases and database design.
- Experience working with cloud Data Warehouse solutions (e.g., Snowflake, Redshift, BigQuery, Azure, etc.).
- Experience working with data ingestion tools such as Fivetran, stitch, or Matillion.
- Working knowledge of Cloud-based solutions (e.g. AWS, Azure, GCP).
- Experience building and deploying machine learning models in production.
- Strong proficiency in object-oriented languages: Python, Java, C++, Scala.
- Strong proficiency in scripting languages like Bash.
- Strong proficiency in data pipeline and workflow management tools (e.g., Airflow, Azkaban).
- *[Make sure to mention any other technologies relevant to your project.]*
- Strong project management and organizational skills.

What will make you stand out

- Good understanding of NoSQL databases like Redis, Cassandra, MongoDB, or Neo4j.
- Experience with working on large data sets and distributed computing (e.g. Hive/Hadoop/Spark/Presto/MapReduce).
- *[If necessary, add a preferred qualification]*