Role: <mark>Senior Big Data Engineer</mark>

Job Brief:

Looking for a Senior Data Engineer to work within a thriving Data Services department that offers a motivating, flexible and fast-moving environment. The candidate will be responsible for building and optimizing data pipelines and datasets that would serve different teams within the enterprise varying from business functions, technical teams and fellow data analysts and data scientists. The candidate will help in identifying data quality issues and put preventive measures to ensure higher confidence in data. The Data Engineer will help in improving the data platform that will support internal and external needs. The ideal candidate is very passionate about Data Engineering and is self-directed and comfortable to work with ambiguity of requirements to purpose and implement optimal solutions to support the needs of internal teams, systems, and external products and partners.

Key Responsibilities:
a. Work with different business stakeholders and technical teams to gather requirements.
b. Architecture and implement optimal data pipelines.
c. Redesign and improve current processes to eliminate manual processes and ensure timely delivery of data.
d. Prepare scalability plans for smoother expansion when needed.
e. Build data models that ensure wholesome view of the company's needed analytics.
f. Work with Data Analysts to prepare Ad-Hoc or permanent batch/real-time data pipelines to enable further analysis or machine learning requirements.
g. Implement quality checks and monitoring schemes to ensure data quality and successful completion of data pipelines.
h. Design and build data platforms.

Key Requirements:
a. 4+ years or more of experience working as a Business Intelligence / Data Engineering role.
b. Strong problem solving and root cause analysis skills.
c. Advanced knowledge of variety Relational Database Engines of a large scale.
d. Ability to write advanced SQL queries.
e. Enterprise-level knowledge of designing and implementing data pipelines.
f. Ability to work with scripting languages such as Python for data manipulation and data pipelines orchestration.
g. Ability to build complex data models for Reporting or Functional purposes.
h. Solve performance issues in terms of data injection or data retrieval.
i. Candidate should have experience and familiarity working with the following tools:
    i. Relational Database Management Systems, such as: SQL Server, Postgresql.
    ii. Real-time OLAP data stores, such as: Kudu, ClickHouse, Apache Pinot
    iii. Distributed File Systems, Object stores, such as: Hadoop, minio
    iv. Distributed Data Processing Frameworks, such as: Apache Spark
    v. Distributed query engines, such as: Presto
    vi. Stream processing tools, such as: Apache Kafka, Apache Flink, Apache Druid.
    vii. NoSQL databases.
    viii. Dataflow management tools, such as: Apache Nifi, Alteryx, SSIS.
    ix. API integration with external sources.
    x. Object-oriented/Functional programming in: Python, Java, Scala, etc.
    xi. Data Visualization tools, such as: Power Bi, Metabase, Apache Superset.