DATA ENGINEER SENIOR

The PNC Financial Services Group, Inc. seeks a Data Engineer Senior in Pittsburgh, PA. The position allows for the ability to work from home with appropriate telecommuting systems for up to three days per week, with a minimum of two days per week in the office. Within the Retail Lending Data Fabric Solution department, be responsible for performing analytical and engineering tasks on vast amounts of structured and unstructured data to extract actionable business insights. Specific duties include: (i) ingesting large volumes of data from various external sources into Hadoop and building data pipelines (ii) developing IT based Hadoop strategies for business data implementation, date acquisition, execution, reporting, and archive recovery; (iii) advising IT, business, applications, and operations counterparts to ensure data integrity and availability; (iv) developing the PoCs for business problems leveraging Hadoop eco system; and (v) educating business community users with Hadoop best practices and latest techniques to optimize the code.

Bachelor's degree in Engineering, Quantitative Finance, Mathematics, Information Technology, or Information Systems Management plus 5 years of experience in application design, development, and testing using Agile Methodology in or for the Retail Banking environment is required. Must have experience with: (i) extracting data from SQL Server, Teradata and Oracle databases and transforming and loading it on Central repository using ETL tools Hive, Impala, Python, and Spark; (ii) implementing operations including data processing, access, storage, governance, and security using Hive, Impala, Cloudera manager, and Cloudera Navigator; (iii) utilizing data query tools including HQL, SQL, R, and Python to manipulate, analyze and interpret data; (iv) data governance involving Alation data platform; (v) creating data pipelines using Toad for Oracle, adding transformations and filters to the pipeline to clean and manipulate the data; (vi) building data pipelines using pySpark to prepare the consumable datasets for data analytics and reporting; (vii) optimizing existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frames and RDDs; (viii) data analysis, data mapping techniques, and creating data visualizations for end users using Python Plotly library and Tableau; (ix) performing data wrangling, data mining and anomaly detection using Jupyter Lab and Toad Data Point; (x) data modeling and data schema design; and (xi) customer life cycle in banking industry including Payments, Origination, Servicing and Default Management, and Charge off.