Job Description

Company	Course5 Intelligence
Position Title	Senior Data Engineer
Reporting to	-
Schedule	Full-Time
Work Location	Bangalore / Mumbai

About Team

Course5i's Advanced Analytics & AI Practise team requires a self-learning Data Engineer, that can lead a team of Data engineer analysts, to solve complex data engineering problems for our global clients in the field of Automotive, Technology, Manufacturing and Industry 4.0.

Our team of 60+ Data Engineers + Data scientists already use multiple available open-source and licensed tools in the market to solve complex problems and drive business decisions. The team has experience in usage of tools and technologies like, SQL, Python, etc. to extract, transform, and load large-scale datasets from various data source systems. The team employs appropriate algorithms to discover patterns, test hypotheses, and build actionable models to optimize business processes.

Solves Data Sciences, AI & Analytical problems, and effectively communicate methodologies and results. Draw relevant inferences and insights from data including identification of trends and anomalies. Work closely with internal/external stakeholders such as business teams, product managers, engineering teams, and partner teams and align them with respect to your focus area.

Responsibilities:

- Designing, building, and deploying a scalable cloud agnostic data ingestion architecture and pipelines to process both Stream and Batch Data using open-source tools and technologies.
- You should be experienced in creating and deploying complete ETL pipeline for cloud, on Prem and hybrid model from scratch.
- You are responsible for Ingesting data from a variety of data sources like csv's/pdf's/docs/storage containers, streams, and databases, process it to get meaningful insights.
- Writing clean, fully tested and well-documented code in Python 3.5+ with pandas, NumPy, Dask, TensorFlow, scikit-learn, and Django
- Design, develop, test, deploy, support, enhance data integration solutions seamlessly to connect and integrate enterprise systems in an Enterprise Data Platform
- Creating Rest API's development for sharing data
- Working directly with clients to identify pain points and opportunities in pre-existing data pipelines and build or improve clients' analytics processes
- Advising clients on the usage of different distributed storage and computing technologies from the plethora of options available in the ecosystem

What you should have

- 5 to 8 years of overall industry experience specifically in data engineering
- Out of which at least 2 years should be on implementing the ETL pipelines in Azure environment
- Strong experience in building data pipelines and analysis tools using Python and PySpark
- Design and develop data ingestion programs to process large data sets in Batch mode using python, Databricks, and Spark.
- Good understanding of underlying Azure Architectural concepts and distributed computing paradigms
- Hands on experience of Azure Data bricks / spark, Spark streaming and its various core components like GraphX, MLLib
- Hands on Experience in working on Microsoft Azure Services like ADLS, Event Hubs, scale sets, Load Balancers, Azure Functions, Logic Apps, Azure Data Factory etc
- Stong experience in Apache Kafka streaming
- Experience working with SQL (mySQL/Postgres/SQL Server)
- Exposure to NoSQL (cassandra/Mongo/Cosmos DB) & Graph databases (Neo4j)
- Experience on visualization tools like Grafana / Power BI
- Understanding of Linux commands, shell scripting & monitoring Linux based production machines

Qualification:

- Bachelors or master's degree in computer science or equivalent