

## Job Description

<b>Company</b>	Course5 Intelligence
<b>Position Title</b>	IC
<b>Reporting to</b>	Head of Manufacturing
<b>Schedule</b>	Full-Time
<b>Work Location</b>	Mumbai / Bangalore / Gurgaon

### OVERVIEW

#### Course5 Intelligence

We enable organizations to make the most effective strategic and tactical moves relating to their customers, markets, and competition at the rapid pace that the digital business world demands. Founded in 2000, our business areas include Market Intelligence, Big Data Analytics, Digital Transformation, Artificial Intelligence, and Analytics. Rapid advances in Artificial Intelligence and Machine Learning technology have enabled us to create disruptive technologies and accelerators under our Course5 Intelligence suites that combine analytics, digital, and research solutions to provide significant and long-term value to our clients. More information can be found at [www.course5i.com](http://www.course5i.com)

#### Global Offices

United States | United Kingdom | United Arab of Emirates | India | Singapore

#### Job 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.

### Requirements & Qualifications:

- 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.
- Strong 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.