

# Lead Data Engineer

1. **TVS D&A Team:** TVS Motor Company is the third largest 2-wheeler company in India with a revenue of over 20,000 crore (US\$2.9 billion). It has an annual sale of more than 3 million units and an annual capacity of over 4 million vehicles. TVS Motor is also the 2nd largest exporter in India with exports to over 60 Countries. A member of the TVS Group, it is the largest company of the group in terms of size and turnover.
2. **Connected Mobility & New Product Development (NPD):** The Connected Mobility & NPD team is a group of Data Engineers, ML Engineers & Data Scientists supporting TVS Motor Company & group companies in their Data Analytics journey. Members of this team work on cutting edge problems with the intent to not only solve for today's problem at TVS Group companies, but also for creating competitive advantage. It is organized as a shared advanced capability hub of Deep Learning capabilities.
3. **Position:** Lead Data Engineer reporting the Head of Data Science and Engineering, Connected Mobility & NPD.

## Position : Lead Data Engineer

### Responsibilities:

- Design, build and deploy a scalable Cloud agnostic data ingestion architecture and pipelines to process both Stream and Batch Data using open source tools and technologies.
- Experience in creating and deploying complete ETL pipeline both on Prem and on Cloud platform from scratch.
- 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.
- Develop programs in Python notebooks/.py files & SQL as part of data extraction, cleaning, transformation and processing.
- Rest API's development for sharing data.

### Mandatory Skills:

- Overall 8-12 years of experience into Data Engineering.
- Strong understanding & familiarity with all Azure / Bigdata Ecosystem components.
- In depth knowledge and understanding of Apache Kafka to handle streaming data.
- In depth knowledge and hands on experience of Azure Data bricks / spark, Spark streaming and its various core components like GraphX, MLlib.
- Design and develop data ingestion programs to process large data sets in Batch mode using python, Databricks, and Spark.
- Strong understanding of underlying Azure Architectural concepts and distributed computing paradigms.
- Hands-on programming experience in Python & OOPs Concepts.
- Hands on experience with major components like Airflow / Nifi / ADF.
- Experience working with SQL (mySQL/Postgres/SQL Server), NoSQL (cassandra/Mongo/Cosmos DB) & Graph databases (Neo4j).
- Advanced working SQL knowledge to create complex queries.
- Hands On experience on visualization tools like Grafana / kibana / Power BI.
- Good understanding of Linux commands, shell scripting and monitoring Linux based production machines.
- Experience working on Azure Cloud services IaaS, PaaS.
- 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.
- Should have worked on various Bigdata file formats like Avro, ORC, Parquet, Nested Json.
- Experience of data migration and deployment from On-Prem to Cloud environment and vice-versa.
- Individual Contributor.
- Problem solving skills.

### Nice-to-Have Skills:

- Certification on any cloud platform.
- Experience in creating any knowledge graph.
- Experience in working on any time-series database.

- Knowledge on continuous integration/continuous deployment.
- Hands on experience in creating micro-service architecture using docker containers and orchestration services like docker swarm and Kubernetes.
- Hands on experience in Golang programming.
- Machine Learning / Deep Learning model development Skills.