

**Job Title** : Data Engineer (Pyspark, Databricks, ADF)

**Location** : Pune-Baner

**Experience** :4+ years Mandatory

**Job Type** : Full-time, Work From Office

**About Us:**

PibyThree (πby3) is a Global Cloud Consulting & Services provider founded by tenured IT professionals from the industry. PibyThree's expertise is in Cloud Transformation, Cloud FinOps, IT Automation, Application Modernization and Data & Analytics. The mission is to make businesses successful, leveraging technology for automation and enhanced productivity

**Job Description**

We are seeking a skilled Azure Data Engineer with strong hands-on experience in PySpark, Azure Data Factory (ADF), and Azure Databricks. The candidate will be responsible for building scalable data pipelines, performing complex transformations, and delivering high-quality data solutions on the Azure cloud platform.

---

**Key Responsibilities**

- Develop and maintain ETL/ELT pipelines using PySpark on Azure Databricks.
- Design, orchestrate, and monitor data workflows using Azure Data Factory (ADF).
- Integrate diverse data sources into Azure Data Lake Storage (ADLS) and transform data using Spark-based frameworks.
- Optimize PySpark jobs for performance, reliability, and cost efficiency.
- Implement Delta Lake solutions for structured and incremental data processing.
- Collaborate with data architects, analysts, and business teams to translate requirements into technical data solutions.
- Ensure end-to-end data quality, validation, and error-handling mechanisms.
- Manage CI/CD deployments for ADF and Databricks.
- Develop documentation, pipeline monitoring dashboards, and operational playbooks.

---

**Required Skills**

- 4–8+ years of experience in Data Engineering with strong PySpark expertise.
- Hands-on experience with:
  - Azure Databricks (Spark, Delta Lake, Notebooks)
  - Azure Data Factory (ADF) pipelines, triggers, activities

- Azure Data Lake Storage (ADLS Gen2)
  - Strong SQL and data modeling skills.
  - Experience working with Parquet, Delta, Hive tables.
  - Knowledge of orchestration, scheduling, and monitoring best practices.
  - Experience with Git and CI/CD tools (Azure DevOps preferred).
  - Strong understanding of ETL/ELT principles, data warehousing, and distributed computing.
- 

### **Good to Have**

- Experience with Azure Synapse, Azure Functions, or Event Hub/Kafka.
- Knowledge of Python automation beyond PySpark.
- Experience in DevOps concepts and infrastructure-as-code (Terraform is a plus).
- Exposure to Snowflake or other data platforms.