

# Data Engineer

Location: Hybrid / Remote | Experience: 2 - 8 Years

## About the Role

---

We are looking for a skilled Data Engineer to design, build, and maintain scalable data pipelines and infrastructure that power analytics and AI-driven solutions for our clients. You will work closely with data scientists, analysts, and business stakeholders to ensure reliable, high-quality data flows across the organization.

## Key Responsibilities

---

- Design, develop, and maintain robust ETL/ELT pipelines for batch and real-time data processing.
- Build and optimize data warehouse and data lake architectures on cloud platforms (AWS, Azure, GCP).
- Implement data models (dimensional, star/snowflake schemas) to support reporting and analytics needs.
- Develop and manage workflows using orchestration tools such as Apache Airflow, Prefect, or equivalent.
- Ensure data quality, governance, lineage, and cataloging across all data assets.
- Collaborate with data scientists to operationalize machine learning models and integrate AI/ML pipelines.
- Optimize query performance and tune databases including Snowflake, Databricks, Redshift, and BigQuery.
- Implement CI/CD practices for data pipeline deployment and version control.
- Monitor pipeline health, set up alerting, and troubleshoot data issues proactively.
- Stay current with emerging data engineering tools and best practices.

## Required Qualifications

---

- Bachelor's or Master's degree in Computer Science, Information Technology, or a related field.
- 2+ years of hands-on experience in data engineering or a related role.
- Strong proficiency in SQL and Python; experience with Scala or Java is a plus.
- Experience with cloud data platforms: AWS (S3, Glue, Redshift, EMR), Azure (Data Factory, Synapse), or GCP (BigQuery, Dataflow).
- Hands-on experience with Snowflake, Databricks, or similar modern data platforms.
- Working knowledge of Apache Spark, PySpark, and distributed computing frameworks.
- Familiarity with data orchestration tools (Airflow, Prefect, Luigi).
- Understanding of data modeling, data warehousing concepts, and database design.
- Experience with version control (Git) and CI/CD pipelines.

## Preferred Qualifications

---

- Experience with real-time streaming technologies (Kafka, Spark Streaming, Kinesis).
- Knowledge of dbt for data transformation and modeling.
- Experience with containerization (Docker, Kubernetes) and infrastructure-as-code (Terraform).
- Familiarity with AI/ML tools and frameworks, including Claude API for AI-augmented data workflows.
- Cloud certifications (AWS, Azure, GCP) are a plus.
- Experience working in analytics consulting or client-facing data engineering engagements.

## Tools & Technologies

---

Cloud: AWS (S3, Glue, Redshift, EMR, Lambda), Azure (Data Factory, Synapse, ADLS), GCP (BigQuery, Dataflow, Cloud Storage) | Data Platforms: Snowflake, Databricks (Delta Lake, Unity Catalog), Apache Spark, PySpark | Orchestration: Apache Airflow, Prefect | Languages: SQL, Python, Scala | AI/ML: Claude API, LangChain, MLflow | Other: dbt, Git, Docker, Terraform, Kafka, CI/CD tools

## What We Offer

---

- Opportunity to work on cutting-edge data and AI projects across diverse industries.
- Collaborative, learning-oriented work culture with mentorship programs.
- Competitive compensation and benefits package.
- Flexible work arrangements (hybrid/remote options).
- Professional development and cloud certification sponsorship.