

Data Engineer

Data

SMB

Enterprise

We are hiring a Data Engineer to design, build, and maintain the data infrastructure that powers analytics, machine learning, and business intelligence across [Company Name]. You will own data pipelines from ingestion to consumption, ensure data quality and reliability, and collaborate with analysts, data scientists, and product teams to make data accessible and trustworthy.

Key Responsibilities

- Design and implement scalable data pipelines that ingest, transform, and load data from diverse sources (APIs, databases, event streams, third-party SaaS tools)
- Build and maintain a modern data warehouse or lakehouse architecture using tools like Snowflake, BigQuery, Databricks, or Redshift
- Develop and enforce data quality checks, monitoring, and alerting to ensure pipeline reliability and data accuracy
- Optimize query performance and storage costs in the data warehouse through partitioning, clustering, and materialization strategies
- Collaborate with data analysts and data scientists to model data for downstream consumption (star schemas, OBT, metrics layers)
- Manage orchestration workflows using tools like Airflow, Dagster, or Prefect to schedule and monitor data jobs
- Implement data governance practices including cataloging, lineage tracking, access controls, and PII handling

Required Skills & Experience

- 3+ years of data engineering experience building production data pipelines
- Strong SQL skills including window functions, CTEs, and query optimization
- Proficiency in Python for data processing, scripting, and pipeline development
- Experience with at least one modern data warehouse (Snowflake, BigQuery, Databricks, or Redshift)
- Hands-on experience with an orchestration tool (Airflow, Dagster, or Prefect)
- Familiarity with data modeling concepts (dimensional modeling, star schema, data vault)
- Understanding of cloud platforms (AWS, GCP, or Azure) and their data-related services (S3, GCS, Glue, Dataflow)
- Experience with version control for data (dbt, Git-based workflows)

Nice-to-Have

- Experience with dbt for data transformation and testing
- Familiarity with streaming data tools (Kafka, Kinesis, Flink, or Spark Streaming)
- Knowledge of data cataloging and lineage tools (Atlas, DataHub, or OpenMetadata)
- Experience with infrastructure-as-code for data infrastructure (Terraform, Pulumi)

- Exposure to analytics engineering or metrics layer tools (MetricFlow, Cube)

Tech Stack

Python

SQL

Snowflake

dbt

Airflow

Spark

Kafka

AWS S3

Terraform

Git

Docker

BigQuery

What We Offer

- Competitive salary and equity at [\[Company Name\]](#)
- Opportunity to build the data platform from the ground up or scale an existing one
- Annual conference and training budget (Data Council, Coalesce, etc.)
- Comprehensive health, dental, and vision insurance
- Flexible remote work with async-first communication culture
- Collaborative team that values data quality and engineering best practices

Interview Process

1. Recruiter phone screen (30 min) — background, experience overview, and logistics
2. Technical screen (45 min) — SQL problem-solving and data modeling discussion
3. Data pipeline design interview (60 min) — design a pipeline for a realistic business scenario, discuss trade-offs
4. Coding interview (45 min) — Python-based data processing problem
5. Hiring manager conversation (45 min) — team dynamics, career growth, and mutual fit
6. Reference checks and offer