Persona Draft: Data Engineer

Generated: 10/25/2026, 10:57:11 AM | Job Level: Individual Contributor | Company Size: Enterprise ($1B+ ARR) | Industry: Unspecified

First draft only — validate with real customer conversations.

Overview

Primary Job Title:
Data Engineer

Secondary Job Titles:
Senior Data Engineer, Data Engineer III, Data Engineer 4/5, Big Data Engineer, Data Platform Engineer, Cloud Data Engineer

Job Level:
Individual Contributor

Average Years of Experience:
3-6 years (ranging from 3+ at mid-level to 6-8+ for senior/staff tracks within the same IC ladder)

Company Size:
Enterprise ($1B+ ARR)

Industry:
Not specified

Common Tools:
Apache Spark
Databricks
Snowflake
Apache Airflow
AWS (Redshift, Glue, EMR, S3)
Google Cloud Platform (BigQuery, Dataproc)
Microsoft Azure (Data Factory, Synapse, Fabric)
Kafka
dbt
Python
Kubernetes/Docker
Microsoft Purview (data governance)

Sample Profiles

Data Engineer — Booz Allen Hamilton

Data Engineer — Booz Allen Hamilton

Senior Data Engineer — Capital One

Senior Data Engineer — Capital One

Sr. Azure Data Engineer — Capital One

Where They Gather

Databricks Data + AI Summit (Event)
https://www.databricks.com/dataaisummit

Snowflake Summit (Event)
https://www.snowflake.com/summit/

dbt Summit (formerly Coalesce) (Event)
https://www.getdbt.com/coalesce

AWS re:Invent (Event)
https://reinvent.awsevents.com/

r/dataengineering (Forum)
https://www.reddit.com/r/dataengineering/

Data Engineering Weekly (Publication)
https://www.dataengineeringweekly.com/

AWS Big Data Blog (Publication)
https://aws.amazon.com/blogs/big-data/

Data Engineering Central (Blog)
https://dataengineeringcentral.substack.com/

Seattle Data Guy (Ben Rogojan) (Influencer)
https://www.youtube.com/c/SeattleDataGuy

Zach Wilson (DataExpert.io) (Influencer)
https://www.dataexpert.io/

Joe Reis (Influencer)
https://joereis.substack.com/

Mehdi Ouazza (Data with Mehdi) (Influencer)
https://www.datawithmehdi.com/

Persona Draft: Data Engineer (continued)

Organizational Structure

Reports To:

Data Engineering Manager / Manager, Data Engineering (sometimes Head of Data or Senior Manager of Data Platform & Engineering)

This Role:

Data Engineer

Stakeholders:

Data Science / ML teams

Business/Data Analysts and Analytics Engineering

Data Architecture / Data Governance

Product and Engineering (platform/software) teams

Compliance, Legal & Privacy (in regulated industries)

Business stakeholders / domain leaders (e.g., Finance, Ops, Marketing)

Their Immediate Work Priorities

1. Design, build, and maintain scalable ETL/ELT data pipelines (batch and streaming) that move and transform data from multiple source systems into warehouses/lakes

2. Ensure data quality, integrity, and reliability through validation frameworks, monitoring, and automated error-handling (e.g., Great Expectations, Soda)

3. Build and optimize data models (logical/physical, dimensional/OLAP and transactional/OLTP schemas) to support analytics and reporting

4. Migrate legacy/on-prem data warehouses and ETL workloads to modern cloud platforms (e.g., Teradata/Informatica to Databricks, Azure, Snowflake)

5. Implement data governance, security, access control, and regulatory compliance (encryption, masking, lineage, HIPAA/financial regulations)

6. Collaborate cross-functionally with data scientists, analysts, architects, and business stakeholders to translate requirements into technical solutions

7. Support and operationalize ML/AI workflows - feature engineering, feature stores, and embedding models into production pipelines

8. Monitor, troubleshoot, and optimize production pipeline performance, cost, and scalability, including on-call support for incidents

9. Build and maintain CI/CD automation, version control, and testing practices for data pipeline code

10. Design self-service data tools/platforms and documentation to improve accessibility of data assets across the organization

Their Development Priorities

1. Earning cloud platform certifications (AWS Certified Data Engineer, Google Professional Data Engineer, Databricks/Snowflake certifications) to validate and broaden cloud expertise

2. Deepening hands-on expertise in distributed processing frameworks (Spark, PySpark, Flink) to handle larger-scale and real-time data challenges

3. Building a track record of leading migrations (e.g., on-prem to cloud, legacy ETL to modern lakehouse) as a portfolio achievement

4. Gaining experience mentoring junior engineers and leading code reviews/technical design sessions as a step toward senior/staff or lead roles

5. Expanding into AI/ML-adjacent skills (feature stores, LLM/GenAI integration, MLOps) as data+AI convergence reshapes the role

6. Developing data governance and compliance expertise (lineage, cataloging, Purview, regulatory frameworks) valued in regulated enterprises

7. Broadening multi-cloud fluency (AWS, Azure, GCP) rather than specializing in just one platform

8. Strengthening data modeling and architecture skills (Kimball methodology, medallion architecture) to move toward Data Architect tracks

9. Building business-facing communication skills to translate technical work into measurable business value/ROI

10. Contributing to reusable frameworks/accelerators and engineering standards to demonstrate platform-level technical leadership