

SUMMARY

Experienced Data Engineer with 4+ years of expertise designing scalable data pipelines and analytics solutions across finance and tech sectors. Skilled in Python, SQL, Spark, Kafka, and cloud platforms (AWS, Azure) to enable real-time and batch processing. Proven ability to optimize data architecture, improve data quality, and support advanced analytics and ML initiatives through robust engineering practices and cross-functional collaboration.

SKILLS

Programming & Scripting: Python, Java, Scala, SQL, JavaScript, Bash/Shell Scripting, DAX
Big Data Technologies: Hadoop, Apache Spark, PySpark, Hive, Kafka, Flume, HBase
Cloud Platforms & Services: AWS (S3, Redshift, Lambda, EC2, Glue), Azure (Data Lake Storage, Synapse, Functions), GCP (Bigquery)
Databases & Storage: MySQL, PostgreSQL, MongoDB, Cassandra, Oracle DB, Relational Databases
Data Warehousing & Modeling: Snowflake, Amazon Redshift, Star Schema, Snowflake Schema, Dimensional Modeling
Data Processing Frameworks: Apache Spark, Apache Flink, PySpark
Data Visualization & Reporting: Tableau, PowerBI, Tabular Models, Excel (Pivot Tables, Macros, VLOOKUP)
ETL & Data Integration Tools: Apache Airflow, Informatica, Talend, dbt
Data Quality & Governance: Data Validation, Data Cleansing, Databricks, Data Catalogs, IAM (AWS/GCP), GDPR/CCPA Compliance
Version Control & Collaboration: Git, GitHub, Bitbucket, Jira, Confluence
Containerization & DevOps: Docker, Kubernetes (EKS/GKE), Terraform, GitLab CI/CD, Jenkins, Travis CI
Monitoring & Logging: Prometheus, Grafana, Splunk, ELK Stack, ELK Stack (Elasticsearch)
Methodologies: Agile, Scrum, CI/CD, Continuous Delivery, Stakeholder Collaboration, Process Improvement
Soft Skills: Problem Solving, Attention to Details, Best Practices, Troubleshooting, Code Review, Technical Documentation

EXPERIENCE

- Data Engineer, Charles Schwab

Dec 2024 – Present | USA

 - Built 12+ real-time data pipelines using Apache Kafka and Spark Structured Streaming to support high-frequency trading, reducing market data ingestion latency to under 2 seconds across equities and options processing systems.
 - Migrated over 30TB of legacy trading and transaction data to AWS S3 and Redshift, achieving a 40% reduction in query latency and cutting monthly infrastructure costs.
 - Created automated ETL workflows in Apache Airflow for risk and compliance reporting, improving on-time report delivery across 10 regulatory submission processes.
 - Engineered dimensional models for investment performance reporting in Power BI, supporting 100+ financial advisors with near real-time updates and reducing ad-hoc data requests.
 - Integrated Spark-based enrichment logic into batch pipelines to validate trading transactions, increasing data accuracy by 35% while maintaining pipeline execution times under SLA thresholds.
 - Enforced access policies with AWS IAM and audit logging for internal and third-party data consumers, fully aligning with SOX, GDPR, and internal data governance standards.
 - Reduced data duplication across environments by 70% through implementation of data cataloging and version control, allowing teams to reuse trusted datasets across research and portfolio analysis teams.
- Data Engineer, Adons Softech

Jan 2022 – Jul 2023 | USA

 - Designed and implemented ETL pipelines in PySpark and Talend across 5 client environments, integrating 40+ diverse data sources including Salesforce, Oracle DB, and S3 into a centralized Snowflake warehouse.
 - Re-engineered existing batch jobs to leverage Spark parallelism, cutting daily data processing time by 60% while handling workloads exceeding 200 million records per execution cycle.
 - Created modular dbt models with test coverage and version control, improving code reuse across teams and reducing onboarding and documentation time for new engineers.
 - Provisioned AWS infrastructure using Terraform with CI/CD workflows via GitLab, reducing manual deployment errors by 95% and ensuring consistent build environments across dev and production.
 - Designed role-based access controls across Redshift and Snowflake environments, aligning with ISO 27001 policies and improving audit traceability across data lineage paths.
 - Led weekly client syncs to define data SLAs, set transformation logic, and finalize 20+ KPIs for executive dashboards used by healthcare, fintech, and logistics clients.
 - Conducted system performance tuning and implemented incremental loads, improving data freshness in analytics dashboards from 6-hour lag to under 30 minutes for real-time reporting scenarios.
- Data Analyst, Cybage Software

Jan 2020 – Dec 2022 | India

 - Analyzed clickstream data from over 15 e-commerce platforms using SQL and Python, identifying user drop-off patterns and supporting marketing decisions that led to a 12% increase in average conversion rates.
 - Developed interactive Tableau dashboards that visualized cross-regional customer behavior metrics for marketing teams, directly impacting segmentation strategies across 6 APAC and North American business units.
 - Automated Excel-based workflows using Python and Power BI, cutting weekly manual reporting time by 8 hours and improving report consistency across 3 internal departments.
 - Built DAX-based data models in Power BI to track user churn and retention, enabling product teams to test and validate feature changes with real-time data insights.
 - Deployed data validation scripts to catch upstream anomalies in customer orders, reducing transactional errors over a 3-month monitoring period.
 - Defined and implemented data quality rules and anomaly checks using SQL for customer and inventory datasets, resulting in a 40% decrease in reporting discrepancies.
 - Collaborated with engineering and QA teams to document reporting logic and metrics, increasing stakeholder clarity and decreasing back-and-forth clarification requests.

EDUCATION

Master in Advanced Data Analytics, University of North Texas | Denton, TX

Aug 2023 – May 2025

Bachelor in Mechanical Engineering Honors, Lovely Professional University | Punjab

Aug 2017 – Aug 2021