

Professional Summary

Data Engineer with 2+ years of experience building scalable **ETL pipelines**, **data lakes**, and **cloud-native** data architectures on **AWS**, **Azure**, and **GCP** using **Python**, **SQL**, **Apache Spark**, and **Airflow**. Track record of reducing pipeline errors by 40%, accelerating execution by 35%, and delivering analytics-ready data for cross-functional stakeholders.

Professional Experience

Interlinked

Berkeley, California

Data Engineer Intern — Python, PySpark, AWS, Databricks, Airflow

March 2026 – Present

- Engineered real-time wildfire sensor data ingestion pipelines using **Apache Spark** on **Databricks** and **AWS S3**, reducing data availability latency by **40%** and enabling earlier threat detection for emergency response teams.
- Built end-to-end **dbt** transformation workflows on **Azure** converting raw geospatial telemetry into analytics-ready models, cutting data preparation time by **60%** and accelerating risk intelligence dashboard refresh cycles.
- Developed automated data quality monitoring frameworks using **Python** and **SQL** across mission-critical environmental feeds on **AWS**, reducing data integrity incidents by **50%** and ensuring continuous situational awareness.
- Designed and orchestrated feature engineering pipelines using **Apache Airflow** and **PySpark** to deliver validated, clean inputs to predictive wildfire spread models, enabling more precise resource allocation recommendations.
- Implemented complex **PL/SQL** and **dbt** transformation logic across environmental data models on **Azure**, producing documented pipelines that streamlined compliance and reporting workflows for disaster response stakeholders.

Addon Solutions

Gujarat, India

Data Engineer — Python, PySpark, AWS Glue, S3, Airflow

November 2023 – November 2024

- Architected a four-zone **AWS S3 data lake** (Landing, Raw, Trusted, Curated) with **AWS Glue** crawlers and **Athena** tables enabling ad-hoc analytics across all ingested datasets.
- Engineered **ETL pipelines** using **AWS Glue** and **Lambda** to process NACHA banking transaction files; implemented **AWS SES** alerts for ingestion and transformation completion.
- Authored **PySpark** schema standardization scripts for multi-format source files; reduced downstream parsing errors by **40%** across all active pipelines.
- Implemented **data governance** using **AWS Lake Formation** with least-privilege RBAC, lineage tracking, and SOC2/GDPR compliance controls.

N Vision IQ

Gujarat, India

Data Engineer — Airflow, Terraform, Redshift, Python, SQL

November 2022 – October 2023

- Built **Apache Airflow** DAGs for multi-source ingestion (Workday, HubSpot, Zendesk, Kantata, Quickbase) into a unified data warehouse with automated dependency management.
- Supported warehouse migration to **AWS Redshift + S3 + Lake Formation**; provisioned all infrastructure via **Terraform** across dev/staging/production.
- Designed incremental load strategies replacing full-refresh patterns; reduced pipeline execution time by **35%** across all production DAGs.
- Published **Power BI** dashboards from **Snowflake/Azure Synapse marts**, reducing compliance reporting turnaround by 25% for Finance and Operations stakeholders.

Projects

Real-Time SaaS Data Pipeline on AWS

DePaul University

Data Engineer — Kafka, Lambda, Redshift, Terraform, Airflow

August 2025 – October 2025

- Built real-time ingestion pipeline using **Apache Kafka**, **Lambda**, and **S3** feeding into **Redshift**; implemented automated **PySpark** transformations for analytics-ready datasets.
- Monitored pipeline health using **AWS CloudWatch** and **Airflow** alerts; reduced mean issue resolution time by **30%** through proactive alerting and runbooks.

Technical Skills

- **Languages:** Python, SQL, Scala, R, Shell/Bash, Java
- **Data Platforms:** Redshift, BigQuery, Azure Synapse, CosmosDB, PostgreSQL, MySQL, SQL Server
- **ETL / Integration:** Apache Spark, PySpark, Airflow, Kafka, AWS Glue, Azure Data Factory, Hadoop
- **Cloud:** AWS (S3, Glue, Lambda, Redshift, Athena, Lake Formation, CloudWatch), Azure (ADF, Databricks, Synapse), GCP (BigQuery)
- **DevOps:** Terraform, CloudFormation, Jenkins, GitLab CI, Docker, Kubernetes, AWS Fargate
- **Monitoring:** AWS CloudWatch, Prometheus, Grafana, Splunk, DataDog
- **Visualization:** Power BI, Amazon QuickSight, DOMO

Education

- M.S in Computer Science — DePaul University, Chicago, IL January 2025 – December 2026
- B.S in Computer Application — Silver Oak University, Gujarat, India August 2021 – June 2024