Richard Clegg

Lead Data Engineer | Greater Chicago Area | https://www.linkedin.com/in/richclegg/ | https://github.com/rclegg

Lead data engineer with 8+ years building data lakes, cloud data warehouses and ETL/ELT pipelines on AWS, with experience across FedRAMP, HIPAA and FERPA data. Now a consultant at an AWS Premier Tier Services Partner and Anthropic partner, designing and building generative AI on AWS, contributing to solution architecture and leading AI assessments for prospective customers. I align business and data strategy, translate complex technical detail for any audience, and have taught college courses in technical communication, data analytics and library science.

SKILLS

Languages: Python, SQL, Bash, Go
AWS: S3, Glue, Athena, Redshift, Step Functions, Lambda, ECS, MWAA, Secrets Manager, EKS, Amazon Quick
Data and orchestration: Airflow, Spark on Kubernetes, BigQuery, SQL Server, PySpark
DevOps: Terraform, CI/CD, Azure DevOps, Kubernetes, Snyk, Renovate
AI and search: Amazon Bedrock, Bedrock Knowledge Bases, Bedrock inference profiles, Amazon OpenSearch

EXPERIENCE

Senior Data Engineer, UTurn Data Solutions

Jan 2025 - Present | Chicago, IL

Project-based consulting | Sep 2026 - Present

- Design and build generative AI on Amazon Bedrock, including knowledge bases over client documents, with pre-processing that absorbs high-volume loads and batch ingestion from S3.
- Lead an AI cost visibility project for a client, using Bedrock inference profiles and cost analytics to track what AI workloads spend.
- Led the full deployment of Amazon Quick, the AI studio: Identity Center integration, group management and custom knowledge bases.
- Lead a rapid AI assessment for a prospective customer as a key member of the pre-sales team.
- Advise on query performance, data migrations, loading Amazon OpenSearch from S3, and scalable patterns for EKS workloads.

Embedded with a healthcare client | Jan 2025 - Aug 2026

- Led the migration from Talend to Apache Airflow on Amazon MWAA, retiring a six-figure annual license and integrating AWS Secrets Manager and S3.
- Served as interim tech lead, setting technical direction, planning sprints and keeping stakeholders aligned.
- Designed and delivered a PII redaction service on ECS, with a Python FastAPI layer on Lambda, for HIPAA-aligned data handling.
- Led SQL Server performance tuning, including indexing, partitioning and execution plan analysis, which improved pipeline throughput and data quality.
- Introduced the team's first Git-based source control practice and owned delivery end to end, with Terraform and CI/CD for every service.

Bushel

2021 - Jan 2025 | Fargo, ND

DevOps Engineer II | Nov 2023 - Jan 2025

- Engineered across GCP, Azure and AWS, automating provisioning and CI/CD with Terraform and Azure DevOps Pipelines.

- Kept running the big data platform I had designed, while expanding into Kubernetes, Linux and Windows Server administration and security tooling.

Data Engineer III, then Data Team Lead | 2021 - Nov 2023

- Led a team of data engineers building a BigQuery and GCS data platform orchestrated with Airflow.
- Set up Spark on Kubernetes for large-scale distributed processing, with all infrastructure in Terraform.
- Added pipeline monitoring and alerting, tuned jobs and queries, and mentored junior engineers.

Data Engineer, Grainbridge

2021 | Omaha, NE

- Designed the data lake and warehouse on AWS S3 serving QuickSight and SageMaker.
- Built PySpark ETL on AWS Glue and ingestion from RDS and DynamoDB.
- Created a serverless app using Step Functions and Lambda to monitor and retry Glue workflows.

Senior Data Engineer, Amplifi Loyalty Solutions

2020 - 2021 | Naperville, IL

- Led architecture and implementation of serverless data pipelines on Glue, S3, Athena and Aurora.
- Wrote Python ETL (pandas, NumPy) for business-critical pipelines and advised on testing and deployment.

Data Analyst, SPR

2019 - 2020 | Chicago, IL

- Automated business processes with Python on Jupyter, built dashboards, and created a Postgres master data management database.
- Reduced risk by fixing bash and PHP scripts in a Salesforce data pipeline.

Director of Data Impact, RPH Innovations, LLC

2018 - 2019 | Chicago, IL

- Led enterprise architecture strategy and ran a data lake on AWS Redshift Spectrum.
- Reduced operating costs by over $100,000 through workflow automation.

Information Services Technical 2, Argonne National Laboratory

2016 - 2017 | Chicago, IL

- Advised researchers on scholarly communication and DOE scientific and technical information compliance.

Web Services Librarian and Computer Science Faculty Member, Dominican University

2013 - 2016 | River Forest, IL

- Led development of a SharePoint portal, supervised a service desk of eight, and provided data governance for an open-source publication database.

EDUCATION

- Master of Library and Information Science (MLIS), Dominican University
- B.A. Public Relations, Advertising, and Applied Communication, Elmhurst University
- Studied at Oxford University

CERTIFICATIONS

- AWS Certified Solutions Architect - Professional (Issued May 2026, valid to May 2029)
- AWS Certified Solutions Architect - Associate (Renewed Jan 2025, valid to May 2029)