

Tiffany Wang

Objective: Data Engineer - West Monroe

Focus Areas: AWS Glue⁺ | Spark/Databricks⁺ | Agile Data Pipelines | Client Collaboration

Phone: (626)-223-6123 | Location: Redondo Beach, CA (Remote Eligible)

LinkedIn: [linkedin.com/in/tiffanywangengineer](https://www.linkedin.com/in/tiffanywangengineer) | Email: tiffany.wang.engineer@gmail.com

SUMMARY

AWS-Certified Data Engineer with 8+ years of experience building **enterprise-scale data solutions** for cross-functional teams. Expert in SQL⁺ (complex query optimization), Spark⁺, and AWS Glue⁺, delivering 40% faster ETL pipelines and \$1.2M+ cloud cost savings. Proven track record in implementing data quality frameworks and collaborating with 20+ client stakeholders.

SKILLS

Core Stack:

SQL⁺ (Window Functions, CTEs), Spark⁺ (Databricks⁺, PySpark), AWS Glue⁺ (CDC Pipelines), Kafka, Redshift⁺

ETL Tools: dbt, Informatica, AWS Lambda, Airflow

Data Quality: Great Expectations, Data Validation Rules, Error Logging

Client Delivery: Agile Scrum⁺, Technical Documentation, Power BI Integration

EXPERIENCE

Solution Engineer II

Capital Group | Irvine, CA | 2019–2023

Multi-Source Healthcare Data Integration (AWS Glue⁺)

- Designed **10+ data enrichment workflows** (lookups, type conversions) using AWS Glue 4.0, reducing manual mapping efforts by **70%**
- **Technical Choice:** Implemented AWS Glue over Talend for serverless scalability, cutting pipeline runtime from 6hr → 45min
- Built unified patient records by mapping 8+ EHR sources into Snowflake⁺, improving BI report accuracy by **35%**

Financial Data Quality Framework

- Developed 50+ data validation rules with Great Expectations, reducing production incidents by **40%**
- **Hook:** Automated error message triage using Python⁺-based log parsing, resolving 90% pipeline failures within 2hr

Freelance Data Architect

Online AI & ML Program | Remote | 2023–Present

Retail Analytics Platform (Spark/Databricks⁺)

- Optimized Spark SQL⁺ jobs for complex joins across 5TB+ datasets, improving query performance by **55%**
- Integrated Power BI⁺ semantic layer with Databricks SQL, enabling self-service analytics for 200+ business users
- **Client Impact:** Reduced dashboard load times from 12s → 3s through query plan optimization

IoT Data Pipeline (Kafka→Redshift⁺)

- Architected real-time streaming solution with **<100ms latency**, handling 500K events/sec for manufacturing clients
- Implemented schema evolution patterns using Avro, ensuring backward compatibility across 15+ device types

EDUCATION

M.S. Software Engineering | Embry-Riddle Aeronautical University | 2015–2017

B.S. Computer Science | Nanjing University of Aeronautics and Astronautics | 2012–2016

Client Impact:

- 70% Faster ETL Development | \$1.2M Cost Savings | 40% Incident Reduction

† JD-Aligned Keywords

AWS Glue | Spark/Databricks | SQL | Data Quality | Power BI