Tiffany Wang

Objective: Software Engineer, Data Platform - Moveworks
Focus Areas: AWS EMR† | Real-time Data Pipelines | ML Data Systems | Data Governance

Phone: (626)-223-6123 | Location: Redondo Beach, CA 90277
LinkedIn: linkedin.com/in/tiffanywangengineer | Email: tiffany.wang.engineer@gmail.com

SUMMARY

Cloud Data Architect with 8+ years of expertise in **real-time ML data infrastructure**, building petabyte-scale pipelines for AI copilot systems. Reduced ML training data latency by 90% through Spark/Kafka optimizations. Certified AWS Solutions Architect with proven success in enterprise data governance (SOC 2/GDPR).

SKILLS

Core Technologies: AWS EMR†, Kafka†, Snowflake†, Airflow†, Python† (PySpark)
MLOps: Feature Store Design, Model Training Pipelines, Data Versioning
Data Governance: Column-level Encryption, Audit Logging, GDPR Deletion Workflows
APIs: RESTful Data Services, Lambda†-based Microservices, Hudi† Delta Streaming

EXPERIENCE

Senior Cloud Data Engineer

Capital Group | Irvine, CA | 2019–2023

Key Achievements:

- **Real-time ML Feature Pipeline**
    - Built Kafka†-Spark Streaming system on AWS EMR† processing 2M+ events/sec, reducing feature latency from 15min → 9s for NLP models

    - **Technical Decision**: Chose Hudi† over Parquet for delta streaming, enabling 60% faster time-travel queries
    - Implemented Airflow†-Temporal hybrid orchestrator, improving pipeline reliability to 99.99% uptime

- **Enterprise Data Governance Platform**
    - Developed column masking APIs with Snowflake† Dynamic Data Masking, achieving GDPR compliance in 3 months
    - **Hook**: Built automated lineage tracking using OpenLineage, adopted by 20+ engineering teams

AI Data Infrastructure Consultant

FinTech Innovation Lab | Remote | 2023–Present

Key Projects:

- **Multi-Cloud Feature Store**
    - Architected AWS S3 + Snowflake† hybrid solution, reducing feature serving latency by 75% for 50M+ user copilot
    - Created Python† SDK for feature access, adopted by 150+ data scientists

- **Anomaly Detection Data Mesh**
    - Designed Kafka†-Flink streaming platform detecting 99.7% SLA breaches in real-time
    - **Technical Hook**: Implemented CDC patterns using Debezium for Oracle→Kafka† ingestion

EDUCATION

MS Software Engineering | Carnegie Mellon University | 2015–2017
BS Computer Science | Tsinghua University | 2011–2015

Technical Impact: 90% latency reduction | 99.99% uptime | 150+ developer adoption