Tiffany Wang

Objective: Data Engineer - Zonic
Focus Areas: AWS EMR[†] | Python (Pandas)[†] | Real-time Data Pipelines | API-Driven Solutions

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

SUMMARY

AWS-Certified Data Engineer with 8+ years of experience building **scalable data solutions** for financial and AI systems. Expert in Python[†] (Pandas, PySpark), AWS[†] (EMR, S3, Lambda), and SQL[†], delivering 90% latency reduction and 99.99% uptime for mission-critical pipelines. Proven track record in GDPR-compliant data governance and cross-functional collaboration with 20+ engineering teams.

SKILLS

Core Stack:
Python[†] (Pandas[†], PySpark), AWS[†] (EMR[†], S3[†], Lambda[†], Athena), SQL[†], Airflow[†], Git[†]
Data Engineering: ETL Optimization, Data Modeling[†], Entity Detection[†], Record Linking[†], Parquet/JSON/CSV
APIs & Governance: REST/SOAP APIs[†], Snowflake[†], Kafka, Dynamic Data Masking, OpenLineage

EXPERIENCE

Senior Cloud Data Engineer

Capital Group | Irvine, CA | 2019–2023

AWS EMR[†]-Kafka Real-Time Pipeline

- Built Spark Streaming system on AWS EMR[†] processing **2M+ events/sec**, reducing NLP feature latency by **90%** (15min → 9s)
- **Technical Choice**: Selected Hudi over Parquet for delta streaming, improving time-travel query speed by **60%**
- Orchestrated pipelines via Airflow[†]-Temporal hybrid scheduler, achieving **99.99% uptime**

GDPR Compliance Platform (Snowflake[†])

- Designed column-level encryption APIs with Snowflake[†] Dynamic Data Masking, ensuring GDPR compliance in **3 months**
- **Hook**: Implemented automated data lineage tracking (OpenLineage), adopted by **20+ teams** for audit logging

AI Data Infrastructure Consultant

FinTech Innovation Lab | Remote | 2023–Present

Multi-Cloud Feature Store (AWS S3[†] + Snowflake[†])

- Architected hybrid feature store reducing serving latency by **75%** for a **50M+ user** fintech copilot
- **SDK Development**: Created Python[†] SDK with Pandas[†] integration, adopted by **150+ data scientists**

Anomaly Detection Mesh (Kafka-Debezium)

- Deployed CDC pipeline using Debezium→Kafka for real-time SLA monitoring, detecting **99.7% breaches**
- **Impact**: Reduced incident response time by **40%** via automated alerts

EDUCATION

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

Quantifiable Impact:

- 90% latency reduction | 99.99% uptime | $500K+ cost savings from optimized cloud resources

[†] **JD-Aligned Keywords**

AWS EMR/S3/Lambda | Python/Pandas | Airflow | Snowflake | Entity Detection | GDPR Compliance