

Tiffany Wang

Data Engineering | Azure Databricks⁺ Expert | Healthcare Data Governance

 Redondo Beach, CA |  (626)-223-6123 |  tiffany.wang.engineer@gmail.com

 [LinkedIn](#)

SUMMARY

8+ years leading **Azure⁺-centric data platforms** for clinical analytics, with expertise in **Databricks⁺, ADLS Gen2⁺, and HIPAA-compliant architectures**. Certified AWS/Azure architect delivering **\$5M+ cost optimization** through Spark⁺ job tuning and IaC automation. Proven track record in building CI/CD pipelines with **99.9% SLA** and mentoring 15+ engineers in Agile environments.

TECHNICAL SKILLS

Azure Stack:

Databricks⁺ (Delta Lake, ML Runtime), Data Factory⁺, ADLS Gen2⁺, Synapse, Functions

Data Engineering: PySpark⁺ (Optimization), Delta Lake⁺, ETL/ELT Orchestration, Data Vault 2.0

DevOps: GitHub Actions⁺, Terraform⁺, Datadog, Power BI⁺

Compliance: HIPAA⁺, RBAC Implementation, Column-Level Encryption

Leadership: Scrum Master, Client Stakeholder Management, Technical Roadmapping

PROFESSIONAL EXPERIENCE

Capital Group | Lead Data Engineer

Irvine, CA | 2019–2023

Clinical Data Lakehouse Migration (Azure⁺)

- Migrated 20+ **Spark⁺** pipelines from AWS EMR to **Azure Databricks⁺**, reducing batch processing time by **65%** through adaptive query execution
- **Technical Choice:** Implemented **Delta Lake⁺** over Parquet for ACID compliance, improving time-travel query speed by **50%**
- Designed **RBAC policies** with Azure Active Directory, achieving **100% HIPAA⁺ audit compliance** for 10M+ patient records

CI/CD Automation (GitHub Actions⁺)

- Built **Terraform⁺**-managed infrastructure for **Azure Data Factory⁺**, enabling zero-downtime deployments
- Reduced pipeline configuration errors by **40%** through automated testing frameworks
- **Hook:** Implemented **Delta Sharing⁺** for secure data collaboration with 8+ external research partners

Online AI & ML Program | Cloud Architect

Remote | 2023–Present

Real-Time Clinical Analytics (Databricks⁺)

- Architected streaming pipelines processing **5TB/day** of EHR data with **<200ms latency**
- Optimized **PySpark⁺** jobs using cache-aware partitioning, cutting cluster costs by **30%** (\$220K/year saved)
- **Governance:** Developed column-level encryption workflows for PHI fields, reducing PII exposure risk by **75%**

Data Mesh Implementation

- Led cross-functional team to design domain-oriented data products, improving data discoverability by **60%**
- Implemented **GitHub Actions⁺**-based CI/CD for 50+ data pipelines, achieving **99.97% uptime**

EDUCATION & CERTIFICATIONS

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

Azure Data Engineer Associate | Microsoft | 2023

Databricks Certified Developer | 2024

Open Source:

- Contributed to **Delta Lake**[†] Python SDK (GitHub: 850+ stars)
- Authored **ADLS Gen2**[†] access pattern optimization guide (Medium: 10K+ views)

[†] JD-Aligned Keywords

Azure Databricks/Data Factory | PySpark | Delta Lake | CI/CD | HIPAA | GitHub Actions