# Wang Zhiwei

☎ *+86 199 8888 8888*
✉ *nocoding188@gmail.com*
*https://zhiweio.me* • *@zhiweio*

## Basics

- Senior Data Engineer with 7 years of experience specializing in designing and deploying distributed big data systems on AWS and Azure cloud environments.
- Expert in architecting scalable data platforms using Apache Spark, Flink, and EMR, with deep expertise in cloud migration, CI/CD automation, and infrastructure as code.
- Proven track record of steering cross-functional teams and external partners to deliver mission-critical projects, including Enterprise Data Warehouse migrations and Customer Data Platforms.

## Education

Sep 2015–Jul 2019 **Bachelor**, **Computer Science and Technology**, *Suzhou University of Science and Technology*

- Awarded Third Prize in the 4th National University Cloud Computing Application Innovation Competition (2018).
- Project: ”Implementation of Large-scale Distributed Functional Dependency Mining Algorithm Based on Spark”.

## Work Experience

Dec 2022–Present **Senior Data Engineer**, *Cognizant Technology Solutions Shanghai Co. Ltd.*

- Led the comprehensive migration of a 10TB+ legacy data warehouse to AWS, utilizing AWS Glue and DMS to reduce query latency by 40% and infrastructure costs by 25%.
- Architected a unified Customer Data Platform (CDP) integrating data from Salesforce, Databricks, and legacy systems, achieving a 360-degree customer view and improving campaign targeting by 30%.
- Refactored core data processing workflows in Databricks using advanced partitioning and caching, reducing daily processing time from 6 hours to 2.5 hours.
- Designed a custom Master Data Management (MDM) solution on Azure, implementing automated data quality checks that reduced discrepancies by 95%.

**Keywords**: AWS, Azure, Cloud Migration, Spark, Databricks, Python

Jun 2021–Nov 2022 **Data Engineer**, *Patsnap Information Technology (Suzhou) Co. Ltd.*

- Architected a real-time wide table system using Flink, Kafka, and TiCDC to provide up-to-the-minute insights on enterprise financing, solving critical compatibility issues in the open-source ecosystem.
- Designed a custom Spark Diff engine to synchronize over 500 million daily records to TiDB, achieving minute-level processing and ensuring efficient incremental updates.
- Built a scalable data lake solution on AWS S3 and Athena to handle 1.8 billion global patent records, supporting TB-scale data processing for financial risk analysis.
- Developed ’Guardian’, a custom continuous delivery tool integrated with GitLab CI/CD, automating deployments and improving reliability across 10+ localized client projects.

**Keywords**: Flink, Spark, TiDB, AWS, Kafka, CI/CD

Jun 2019–Jun 2021 **Data Engineer**, *Intsig Information Co., Ltd.*

- Engineered a high-throughput real-time data pipeline using Redis and Kafka, processing over 100 billion data points to ensure sub-second freshness for 230 million entities.
- Architected a Python-based ETL optimization library that standardized incremental update logic across 1000+ dimensions, reducing boilerplate code by 90%.
- Led the DevOps transformation by migrating to GitLab and implementing CI/CD pipelines, reducing release times from hours to minutes.

**Keywords**: Python, Redis, Kafka, ETL, DevOps

## Skills

Cloud & Infrastructure Expert **Keywords**: AWS (Redshift, Glue, EMR), Azure (Synapse, Cosmos DB), Docker, Terraform (IaC), CI/CD (GitLab, GitHub Actions)

Big Data & Streaming Expert **Keywords**: Apache Spark, Apache Flink, Kafka, Databricks, Airflow & DolphinScheduler

| | | |
|---|---|---|
| Databases & Storage | Advanced | **Keywords**: PostgreSQL & MySQL, TiDB (Distributed SQL), MongoDB & DynamoDB, Redis, Data Warehousing (Redshift) |
| Languages & Development | Advanced | **Keywords**: Python, SQL, Shell Scripting, Java/Scala (Spark/Flink) |

## Certificates

Jan 2024 **AWS**, *AWS Certified Data Analytics –Specialty*

Sep 2024 **AWS**, *AWS Certified Developer - Associate*

Jun 2024 **Microsoft**, *Microsoft Certified Azure Data Engineer Associate*

Mar 2023 **Databricks**, *Databricks Certified Data Engineer Associate*

Mar 2022 **PingCAP**, *PingCAP Certified TiDB Professional*

## Projects

Dec 2022–Present **Cloud-native ecosystem unifying fragmented customer data across the enterprise.**, *Bayer Customer Data Platform (CDP)*

- Established a robust infrastructure foundation using Terraform and GitHub Actions, reducing deployment time by 60% and ensuring 99.9% uptime.
- Designed a serverless orchestration workflow using AWS Step Functions and EventBridge to coordinate ETL jobs across heterogeneous data sources.
- Implemented a cost-effective, auto-scaling data processing pipeline using AWS Glue and Lambda, optimizing compute costs by 40%.

**Keywords**: AWS, Terraform, Glue, Lambda, Step Functions

Nov 2022–May 2022 **Strategic modernization of a legacy 10TB+ Enterprise Data Warehouse to Amazon Redshift.**, *Bayer Enterprise Data Warehouse Migration to AWS*

- Architected a resilient, self-healing migration pipeline using AWS Step Functions, reducing engineering overhead by 80%.
- Developed a custom Python migration utility (dm4) to automate DDL translation for over 2,000 tables with 99% accuracy.
- Engineered a memory-resident data transfer pipeline using UNIX streams, accelerating 10TB data transfer by 300% compared to traditional methods.

**Keywords**: AWS Redshift, Python, Migration, UNIX Shell

May 2025–Dec 2025 **Comprehensive MDM system leveraging Azure and Power Platform.**, *Retail Master Data Management Platform*

- Designed a billion-scale data architecture using Azure Cosmos DB and Synapse Analytics to manage massive historical data retention.
- Established a robust data bridge between Azure operational systems and AWS analytical warehouses using Synapse Link and custom adapters.
- Optimized data ingestion pipelines using Azure Blob Storage and parallelized Dataflows, achieving a 30x performance gain.

**Keywords**: Azure, Cosmos DB, Synapse Analytics, Data Integration