Wang Zhiwei

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

Basics

- Senior Data Architect with 7 years of experience designing and delivering scalable cloud-native data platforms and enterprise data warehouses on AWS and Azure.
- Expertise in building high-throughput real-time streaming (Flink, Kafka) and batch processing (Spark, Databricks) architectures, optimizing data reliability and query performance by 40%+.
- Proven track record of leading cross-functional teams in digital transformation initiatives, including migrating 10TB+ legacy systems and implementing Master Data Management (MDM) solutions for Fortune 500 clients.

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 (Tech Lead)**, *Cognizant Technology Solutions Shanghai Co. Ltd.*, https://www.cognizant.com

- Led the end-to-end architecture and delivery of a unified Customer Data Platform (CDP) on AWS, integrating 8+ disparate systems (Salesforce, Databricks) to create a "Single Source of Truth" for marketing analytics.
- Spearheaded the strategic migration of a 10TB+ legacy SQL Server data warehouse to Amazon Redshift, reducing query latency by 40% and infrastructure costs by 25% through serverless orchestration.
- Architected a Retail Master Data Management (MDM) system using Azure Power Platform and Cosmos DB, enforcing data governance and reducing master data discrepancies by 95% for enterprise operations.
- Mentored a team of engineers on cloud-native best practices (IaC, CI/CD), fostering engineering excellence and ensuring 99.9% platform uptime.

Keywords: AWS, Azure, Solution Architecture, Team Leadership, Data Governance

Jun 2021–Nov 2022 **Data Engineer (Core Big Data Developer)**, *Patsnap Information Technology (Suzhou) Co. Ltd.*, https://www.patsnap.com

- Architected a real-time data warehouse solution for the TFFI SaaS platform using Flink, Kafka, and TiDB, enabling minute-level insights for 180 million global patent records.
- Designed a custom Spark Diff engine to handle daily synchronization of 500 million analysis records, optimizing incremental updates and reducing database load significantly.
- Led the technical delivery of localized data systems for 10+ major banking clients (e.g., Agricultural Bank of China), automating deployment pipelines to reduce delivery costs.

Keywords: Real-time Data Warehouse, Flink, TiDB, Spark, Client Delivery

Jun 2019–Jun 2021 **Data Engineer**, *Intsig Information Co., Ltd.*, https://www.intsig.com

- Architected high-throughput ETL pipelines processing 100 billion data points for Qixinbao, ensuring sub-second data freshness for 230 million enterprise entities.
- Modernized the DevOps infrastructure by migrating to GitLab CI/CD, automating testing and deployment to reduce release cycles from hours to minutes.
- Developed Python-based productivity tools to standardize ETL logic across 1000+ dimensions, reducing boilerplate code by 90%.

Keywords: ETL Optimization, Python, DevOps, High-throughput Systems

Skills

Data Architecture & Cloud — Expert — **Keywords**: AWS (Redshift, Glue, EMR), Azure (Synapse, Cosmos DB), Data Warehouse (Snowflake, TiDB), Terraform (IaC), Master Data Management (MDM)

Big Data & Streaming	Expert **Keywords**: Apache Spark (Databricks), Apache Flink, Kafka, Hadoop Ecosystem, Real-time Data Processing
Software Engineering & DevOps	Advanced **Keywords**: Python, SQL & NoSQL, CI/CD (GitLab, GitHub Actions), Docker & Kubernetes, Serverless Architecture
Leadership & Soft Skills	Advanced **Keywords**: Solution Architecture Design, Cross-functional Collaboration, Project Management, Stakeholder Communication, Agile Methodologies

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 **Unified cloud-native data platform integrating fragmented customer data for advanced analytics.**, *Bayer Customer Data Platform (CDP)*

- Architected a serverless data lakehouse on AWS using Glue, Lambda, and Step Functions, optimizing compute costs by 40% while handling billions of daily records.
- Engineered a custom high-throughput Salesforce integration layer to overcome API limits, synchronizing 10M+ records daily with improved reliability.
- Implemented comprehensive CI/CD pipelines via GitHub Actions and Terraform (IaC), reducing deployment time by 60% and ensuring zero configuration drift.

Keywords: AWS Architecture, Data Lakehouse, Terraform, Salesforce Integration

May 2025–Dec 2025 **Enterprise MDM solution on Azure to centralize and govern retail data assets.**, *Retail Master Data Management Platform*

- Designed a billion-scale data storage architecture using Azure Cosmos DB and Blob Storage, achieving 30x performance gains in bulk data ingestion.
- Established a cross-cloud data synchronization bridge between Azure operational systems and AWS analytical warehouses using Synapse Link.
- Enforced Zero Trust security models and automated data quality checks, streamlining data stewardship and saving 20+ operational hours weekly.

Keywords: Azure Architecture, Master Data Management, Cosmos DB, Data Governance

Nov 2022–May 2023 **Strategic migration of 10TB+ legacy warehouse to cloud-native Amazon Redshift.**, *Enterprise Data Warehouse Migration to AWS*

- Architected a self-healing, event-driven migration pipeline using AWS Step Functions, automating the transfer of 3,000+ tables with 99% accuracy.
- Developed a proprietary Python migration engine with AST parsing to automate DDL translation, reducing engineering overhead by 80%.
- Optimized Redshift schema design with columnar compression, reducing storage footprint by 40% and improving analytical query performance by 5x.

Keywords: Cloud Migration, Amazon Redshift, Python Automation, Performance Tuning