

Basics

- Senior Data Architect with 6+ years of experience designing and delivering scalable, cloud-native data solutions on AWS and Azure.
- Expert in modernizing legacy data warehouses and building high-throughput ETL/ELT pipelines using Spark, Kafka, and Flink, managing 10TB+ datasets and billions of records.
- Proven track record of leading cross-functional teams to deliver mission-critical projects, including Enterprise Data Warehouse migrations and Customer Data Platforms, driving digital transformation and operational efficiency.

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 (Technical Lead), Cognizant Technology Solutions Shanghai Co. Ltd.**
- Served as Technical Lead for Bayer's digital transformation, architecting a unified Customer Data Platform (CDP) integrating 8+ systems and billions of records to enable 360-degree customer analytics.
 - Led the strategic migration of a 10TB+ legacy Enterprise Data Warehouse to AWS Redshift, designing a serverless migration factory that reduced query latency by 40% and infrastructure costs by 25%.
 - Designed and implemented a Retail Master Data Management (MDM) system using Azure and Power Platform, resolving critical data inconsistencies and reducing discrepancies by 95%.
 - Optimized core data processing workflows in Databricks, reducing daily processing time by 60% and ensuring 99.9% data availability for business reporting.
- Keywords:** AWS, Azure, Data Architecture, Team Leadership, Cloud Migration
- Jun 2021–Nov 2022 **Data Engineer, Patsnap Information Technology (Suzhou) Co. Ltd.**
- Architected a real-time data warehouse solution for the TFFI SaaS product using Flink and TiDB, enabling minute-level data freshness for 180 million global patent records.
 - Designed a scalable data lake architecture on AWS S3 and Athena, orchestrating complex Spark jobs to process TB-scale datasets for financial risk analysis.
 - Developed 'Guardian', an automated continuous delivery system, reducing deployment costs and improving reliability for 10+ localized banking and government projects.
- Keywords:** Real-time Analytics, Flink, TiDB, AWS, Spark
- Jun 2019–Jun 2021 **Data Engineer, Intsig Information Co., Ltd.**
- Spearheaded data engineering for Qixinbao, architecting high-throughput real-time pipelines using Redis and Kafka to process 100 billion+ data points for 230 million entities.
 - Designed complex data models for 1000+ business dimensions and developed a Python-based ETL optimization library, reducing boilerplate code by 90%.
 - Led DevOps transformation by implementing CI/CD pipelines, reducing release times from hours to minutes and improving team efficiency.
- Keywords:** High-throughput Systems, Kafka, Redis, Data Modeling, DevOps

Skills

- Data Architecture & Cloud Expert **Keywords:** AWS (Redshift, Glue, Lambda), Azure (Synapse, Cosmos DB), Data Warehouse & Lakehouse, ETL/ELT Pipeline Design, Master Data Management (MDM)
- Big Data & Streaming Expert **Keywords:** Apache Spark, Apache Flink, Kafka, Databricks, TiDB

Programming & DevOps Advanced **Keywords:** Python, SQL, Terraform (IaC), CI/CD (GitHub Actions, GitLab), Docker

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 data lakehouse unifying customer data across the enterprise.**, *Bayer Customer Data Platform (CDP)*
○ Architected a serverless data lakehouse on AWS, integrating Salesforce, Databricks, and legacy systems to establish a Single Source of Truth for customer data.
○ Designed a modular Terraform architecture for Infrastructure as Code, reducing deployment time by 60% and ensuring 99.9% uptime.
○ Engineered high-throughput integration layers and event-driven orchestration workflows using AWS Step Functions and Lambda, optimizing compute costs by 40%.
Keywords: AWS Lakehouse, Terraform, Data Integration, Serverless

May 2025–
Dec 2025 **Centralized master data governance system on Azure and Power Platform.**, *Retail Master Data Management Platform*
○ Architected a comprehensive MDM solution leveraging Azure Cosmos DB and Power Platform to centralize and govern retail data assets.
○ Designed a high-performance storage layer with Cosmos DB to manage billions of records with sub-second response times.
○ Implemented a metadata-driven 'Zero Code' framework and automated ETL pipelines, significantly reducing manual maintenance overhead.
Keywords: Azure, Master Data Management, Cosmos DB, Power Platform

Nov 2022–
May 2022 **Strategic migration of 10TB+ legacy warehouse to Amazon Redshift.**, *Enterprise Data Warehouse Migration to AWS*
○ Spearheaded the migration of a 10TB+ SQL Server warehouse to Amazon Redshift, resolving scalability bottlenecks.
○ Architected a serverless, event-driven migration pipeline using AWS Step Functions and Lambda, automating complex workflows.
○ Optimized Redshift schemas and data transfer pipelines, achieving a 300% acceleration in data transfer and 5x improvement in query performance.
Keywords: Cloud Migration, Amazon Redshift, Performance Optimization, Python