

Ben Zhang

Senior Director, Data Engineering & Platform

bennzhang@gmail.com

626-438-9033

Experienced Data Engineering Leader with over ten years of expertise in creating and managing enterprise-level data analytics, big data, and cloud solutions across various sectors. Proficient in simplifying complex data environments, leading high-performing data product development, and delivering crucial data pipelines. Specializes in constructing top-tier data lakes, data marts, ensuring data governance and data privacy. Skilled in developing and optimizing data infrastructures using modern tools and cloud platforms.

Work Experience

Senior Director, Data Engineering & Platform

Apr 2021 - Present

Beachbody

- Led the design, development, and optimization of scalable data pipelines and centralized data products.
- Spearheaded the migration of the Enterprise Data Warehouse from legacy systems Hive, Redshift, Oracle to Snowflake, enhancing performance and scalability
- Established Single Source of Truth : Developed a comprehensive Enterprise Data Warehouse, ensuring data consistency and reliability across the organization.
- Architected and implemented specialized data marts for finance, marketing, royalties, and customer data, simplifying data access and reporting
- Developed a Customer 360 platform for enhanced customer segmentation and personalization
- Develop and implement marketing models, including LTV, churn, Multi-Touch Attribution MTA, and Marketing Mix Models (MMM) , to optimize budget allocation and maximize ROI.
- Treated Data as a Product : Established SLAs, enforced data quality, applied data governance, and set up job monitoring and alerting systems. Built data catalog, data lineage.
- Cost Monitoring and Optimization
- Technologies : Snowflake, Redshift, Hive, DBT, AWS EC2, S3, Kinesis, ECS, Neptune GraphDB, DynamoDB, Lambda, MWAA Airflow, Event-Bus, Tableau, Amplitude, Adobe Analytics, Google Analytics, Python, Java, Postgres, MySQL, pySpark, Jupyter, HDFS

Principal Data Engineer/Data Architect

Mar 2018 - Apr 2021

Universal Music Group

- Led the architecture and development of two major public-facing data products - Universal Music Artist (UMA) and Universal Music Label (UML) . These highly available, scalable, and high performing applications process hundreds of billions of records daily, providing artists and label managers with comprehensive insights into music consumption and engagement.
- Architected and built the Global Audience Platform (GAP) . GAP integrates CDP, customer segmentation, marketing automation, and campaign performance tracking sub-systems to deliver a unified customer 360 view of UMG's global audiences from multi-channel sources
- Designed and implemented a Petabyte Data Lake and multiple Data Marts, providing scalable storage and analytics capabilities
- Established Data Validation Framework : Managed the development of an e-commerce and financial Data Validation framework to support financial auditing, reporting, and reconciliation processes.
- Implemented Data Quality Framework : Created a Machine Learning-powered Data Quality Framework to enable anomaly detection, ensuring high data integrity.

- Launched Monitoring and Alerting System : Created and launched a comprehensive, company-wide data status monitoring and alerting system to maintain data reliability.
- Led the consolidation of over 1, 000 Google Analytics accounts, improving data consistency and accessibility.
- Technologies : BigQuery, DBT, GCP Cloud Storage, Cloud Functions, Dataflow, App Engine, BigTable, Cloud Composer Airflow, Apache Beam, Python, Java, SQL, NoSQL, ElasticSearch, Scala

Sr. Manager, Software Engineer

Apr 2014 - Feb 2018

Westlake Financial Services

- Productionized Machine Learning Models : Successfully implemented and productionized machine learning models for car loan risk control and pricing, enhancing predictive accuracy and financial decision-making
- Developed Next-Gen Payment Systems : Designed and built advanced text2Pay and payment text reminder systems utilizing Twilio, improving customer payment experience and operational efficiency.
- Automated E-Sign Process with DocuSign : Created a seamless DocuSign E-sign automation process. This included integrating DocuSign Powerform, auto-populating customer information into templates, downloading signed documents, extracting signer information, and updating customer records.
- Optimized ETL and Automation Jobs : Designed, built, and optimized numerous mission-critical ETL and automation jobs, ensuring data integrity and streamlined operations
- Built Highly-Available Microservices : Developed robust, highly-available microservices using Spring Boot/Python Flask Gunicorn, ensuring scalable and efficient service delivery.
- Technologies : Java (Spring MVC, Spring Boot) , Python (Flask, Gunicorn) , MySQL, Oracle, Postgres, BigQuery, GCP (Cloud Storage, Dataflow, Cloud Function) , Pentaho Kettle, ETL, Apache NiFi

Senior Software Engineer

Jun 2006 - Apr 2014

Acxiom Corporation

- Designed and Developed Core Technology Product - Abilitec : Created and advanced Abilitec, a Core Technology product that unifies disparate representations of names, addresses, emails, and social information from various sources and time periods. Abilitec delivers a single, current, and accurate view of each individual. The product offers services such as identity recognition, move updates, and data standardization, benefiting clients including major banks, retailers, dealers, and investment firms

Core Skills

Big Data : BigQuery, Snowflake, Redshift, Glue, Athena, Databricks,

AWS : ECS, EC2, ECR, S3, RDS, DynamoDB, Glue, Athena, Redshift, Lambda, SNS, CloudWatch, IAM, Event-Bus, Kinesis stream, Kinesis firehose, MWAA, Cloud Formation,

GCP : Compute Engine, GKE, GAE, Cloud Storage, PubSub, Cloud Function, Dataflow, Cloud composer,

Data Pipeline : Apache Airflow, DBT, Fivetran, Apache NiFi, Pentaho Kettle ETL, Alteryx,

Databases : Oracle, MySQL, PostgreSQL, BigTable, Hbase, MongoDB, Redis, DuckDB, ElasticSearch,

Programming : Python, Java, SQL, Scala, Go, Shell Script, Jupyter Notebook, API : Restful, GraphQL, SOAP APIs,

Big Data Processing : Apache Beam, Apache Spark, Apache Flink, Kafka, Presto, pySpark,

Product Analytics : Amplitude, Adobe Analytics,

Machine Learning : Pandas, Numpy, Matplotlib, Plotly, Scikit-Learn, Jupyter Notebook

Education

University of Arkansas

Dec 2007

Doctor of Philosophy Computer Science

Shanghai Jiaotong University

Dec 2001

Master of Information

