

Rose Duan

roseduan520@gmail.com | github.com/roseduan | roseduan.cn

Education

Chengdu Technological University

2015/09 – 2019/06

Bachelor of Management in International Business

Chengdu, China

- Overall average score: 77/100. Relevant quantitative & computing coursework: Advanced Mathematics, Probability & Statistics, Database Technology (Access), Web Design & Programming, and Data Analysis & Optimization.

Technical Skills

Languages: Go, Rust, Java, C, SQL

Databases & Storage Systems: PostgreSQL internals, MPP architecture, LSM-Tree, B+Tree, Write-Ahead Log (WAL), Bitcask, RocksDB, MySQL, Redis

Distributed Systems & Infrastructure: storage-compute separation, distributed query execution, cluster topology & metadata management, Foreign Data Wrapper (FDW)

Tools: Git, Linux

Work Experience

HashData

2022/06 – Present

Database Kernel Engineer

Beijing, China

- **Storage-Compute Separation.** Implemented support for scalable Warehouses in Apache Cloudberry, including dynamic addition and removal of Segment nodes; adapted the system to retrieve Warehouse topology per transaction, update cluster metadata during scaling operations, and route query execution according to the current topology; implemented a Table Access Method to expose data operations (scan, insert, update) against the cloud-based storage layer.
- **VACUUM Redesign for Cloud Storage.** Led a ground-up redesign of VACUUM and VACUUM FULL to work under Cloudberry's storage-compute-separated, object-storage-based architecture, replacing the local-disk assumptions inherited from PostgreSQL; introduced a new File Cleanup Service responsible for reclaiming obsolete files from remote object storage.
- **Data Lake Integration.** Contributed to Foreign-Data-Wrapper-based access to Apache Iceberg and Apache Hudi, extending support to Parquet, ORC, and Avro formats to make Cloudberry compatible with a unified lakehouse architecture alongside its native MPP architecture.
- **cloudberry_fdw.** Designed and implemented cloudberry_fdw, an MPP-aware foreign data wrapper analogous to postgres_fdw that enables Segment-parallel access to data on a remote Cloudberry cluster; built on postgres_fdw internals and Cloudberry's COPY FROM/TO PROGRAM mechanism, with a custom binary protocol for network-based data exchange between clusters.
- **Time-Series Support.** Implemented a time_series table access method supporting table creation, insertion, and scanning for time-series workloads, with a chunk-based design for time-range partitioning to improve query performance.
- **Parallel Query Processing.** Implemented Parallel Index Scan support in Cloudberry's ORCA query optimizer, enabling parallel query execution plans.
- **Extension Ecosystem.** Ported and adapted multiple PostgreSQL-native extensions — kafka_fdw (added protobuf format and SASL authentication support), pgvector, zombodb, and pgrx — to work correctly under Cloudberry's MPP architecture.

Bilibili

2021/01 – 2022/06

Senior Backend Engineer, Go

Shanghai, China

- Owned development and maintenance of the company's internal distributed key-value storage system, built on brpc with RocksDB as the default storage backend and Redis-compatible data structure semantics.

- Designed and implemented a new embeddable storage engine based on the Bitcask model and integrated it as a pluggable backend, enabling the system to support and switch between multiple storage engines.
- Adopted the Bitcask-based engine for latency-sensitive workloads, achieving lower read/write latency and higher throughput than the existing RocksDB backend for those access patterns.

Qiyuesuo

2019/07 – 2020/12

Java Backend Engineer

Shanghai, China

- Maintained and extended the backend of the company's e-contract signing platform using Java, Spring, MySQL, and Redis.

Manulife

2018/10 – 2019/07

Java Intern

Chengdu, China

- Assisted with backend development tasks using Java and Spring Boot.

Open-Source Contributions

Personal Projects

github.com/roseduan · 1.4k GitHub followers

- **rosetdb** (4.9k stars) — Creator & Maintainer. A lightweight, fast, and reliable embedded key-value storage engine based on the Bitcask model, offering append-only writes, fast crash recovery, and single-disk-seek reads; used in production by multiple real-world systems.
- **mini-bitcask** (466 stars) — Creator & Maintainer. A minimal, from-scratch implementation of the Bitcask storage model, used to teach the design principles underlying rosetdb.
- **lotusdb** (2.3k stars) — Creator & Maintainer. A key-value storage engine combining LSM-tree and B+-tree designs, inspired by the SLM-DB and WiscKey papers, to reduce the read and space amplification typical of pure LSM-tree engines while retaining fast writes.
- **wal** (285 stars) — Creator & Maintainer. A standalone Write-Ahead Log library for LSM-tree- and Bitcask-style storage engines, used as the durability layer underlying rosetdb and lotusdb.
- **diskhash** (30 stars) — Creator & Maintainer. An on-disk hash table index for maintaining persistent, memory-efficient metadata indexes for Write-Ahead-Log-based storage engines.
- **rust-practice** (750 stars) — Creator & Maintainer. A collection of hands-on Rust learning projects and a course for building a SQL database system from scratch, including a Bitcask-based key-value store, an expression evaluator, and an MVCC implementation.

Apache Cloudberry Committer

Since 2026/01

- Contributed across warehouse grammar enhancements and numerous bug fixes; implemented CREATE INDEX CONCURRENTLY and scheduled-task (CREATE TASK) support; maintain the kafka_fdw, pgvector, zombodb, and pgrx extensions for Cloudberry.
- Recognized as a Committer for sustained technical contributions, extension maintenance, and community engagement.

Other Contributions

- **postgresql-interfaces/psqlodbc** — Fixed a hang on *COPY ... FROM STDIN / TO STDOUT* and a bug where *'infinity'/'-infinity'* dates were silently returned as today's date.
- **apache/flink-connector-jdbc** — Fixed an issue with unnecessary unique fields being included in generated PostgreSQL upsert statements.

Selected Talks

- **Deep Dive into Parallel Query Execution Solutions for Apache Cloudberry**, Community Over Code Asia 2026.
- **Building a Unified Lakehouse Solution with Apache Cloudberry**, Community Over Code Asia 2025.