

VAIBHAV RABBER

[+91-99885-00414](tel:+91-99885-00414) | vrongmeal@gmail.com | [LinkedIn \(/in/vrongmeal\)](https://www.linkedin.com/in/vrongmeal/) | [Website \(vrongmeal.com\)](https://www.vrongmeal.com) | [Github \(@vrongmeal\)](https://github.com/vrongmeal)

SKILLS

Technical: Rust, Go, C++, Apache Arrow, Apache DataFusion, Git, Docker, Lua, Python, Javascript

EXPERIENCE

Software Developer (Founding Engineer), GlareDB

01/2023 - 05/2024

- Founding member of the GlareDB engineering team building a fast and scalable [open-source](#) SQL database engine based on Apache Arrow and DataFusion in Rust for distributed analytical workloads.
- Extended support for scanning tables from multiple datasources dynamically with support for filters and limit pushdown such as Postgres, BigQuery, Snowflake, Azure Blob Storage, MS SQL Server, SQLite, and ClickHouse.
- Enhanced the external datasources by implementing write operations (INSERT and COPY TO) to standalone Parquet, CSV and JSON files in cloud storage (GCS, S3 and Azure), and databases like Postgres, MySQL, Snowflake etc.
- Reverse-engineered the Snowflake Go connector and wrote an implementation in Rust from scratch with support for Arrow IPC.
- Added SSH tunnelling capabilities for connecting to Postgres and MySQL datasources on a private network.
- Worked on a hybrid query execution model enabling GlareDB to process local data on cloud as required, hence minimising the compute overhead on a user's system to approximately nil in 95% of the cases.
- Implemented the Postgres wire protocol codec making GlareDB compatible with most of the existing Postgres clients.
- Crafted a framework to run retry-able background jobs such as compaction and metrics collection based on the given operation with a mechanism to avoid running duplicate jobs for a defined time interval.
- Engineered a system to collect compute metrics based on the number of bytes of data processed in a query with a 100% accuracy.
- Overhauled the SQL Logic Tests runner making it extendable with pre-test and post-test hooks, along-with the ability to run tests in parallel by running each test in an isolated database namespace improving the CI performance 8 folds.
- Integrated various other features such as session-lifetime (temporary) tables, access control on objects disallowing unexpected write operations, and fixed numerous critical bugs in the query planner.
- Forked and updated various open-source dependencies used by GlareDB like [Arrow](#), [DataFusion](#), [Iceberg](#) and [BigQuery](#) client.

Software Developer (Analyst), Goldman Sachs

07/2021 - 01/2023

- Member of the Global Markets Franchise Data team working on the proprietary distributed time-series database providing real-time market data to over 100 teams in the firm for analytics or backend service.
- Incorporated load balancing for market-data streams in the data collection service using the consistent hashing algorithm, reducing the overload on streaming servers to 1/2 times the original load.
- Multiplied the rate limit for requests to S3 storage exponentially and upgraded the storage engine to efficiently save and fetch data with key names larger than the provided limitation.
- Revamped the C++ and Java client interface to query data and introduced features like setting the number of retries for multi-clustered database instances, increasing the query success rate by 200%.
- Expanded the Lua-based query language to support Date types eliminating code smell around the data type by 100%.
- Optimised the query to return curve data structure by adding support for the datatypes natively in C++, increasing the capacity of clients to run thousands of parallel queries in the same amount of time.
- Resolved multiple critical bugs and conceived a testing framework around the data collector, decreasing the number of production issues concerning it by 90%.

Software Development Intern (Summer Analyst), Goldman Sachs

05/2020 - 06/2020

- Developed a server-side processing unit to generate consolidated order books from real-time market data in C++.
- Integrated the processor with various database clients (C++, Python, Java, etc.), enabling developers to query the transformed data with the language of their choice.

Core Contributor and Mentor, Google Summer of Code with Zulip

05/2019 - Present

- Added features like creating conversation transcripts and notifications for messaging events, and re-engineered components like message edit box, drafts, and edit history modals with modern CSS techniques as a [developer in GSoC 2019](#).
- Mentoring since 2020 by guiding developers through their projects, reviewing pull-requests, and participating in design discussions.

PROJECTS

Pinger

- Led the development of an [open-source](#) uptime and status monitoring application with a distributed architecture in Go with extendable interfaces to plug in different metrics storage backends (Influx, Timescale, Prometheus etc.) and check types (HTTP, ICMP, WS, etc.).
- Benchmarked and [improved the performance](#) of in-house key-value store 3 folds for write and 6 folds for read operations.

Gasper

- Steered the development a [platform](#) for around 1000+ college students to automate the deployment of web applications to showcase their projects in hackathons and coursework assignments

Leaf

- Authored a general-purpose reloader in Go to watch for file changes and run a set of commands, with [over 100 stars](#) on [Github](#).

OPEN SOURCE CONTRIBUTIONS

• [Cockroach DB](#) • [Apache Arrow](#) • [Apache DataFusion](#) • [Caddy Server](#)

EDUCATION

B.Tech. in Civil Engineering

07/2017 - 05/2021

IIT Roorkee (CGPA: 8.087/10)