

MariaDB AX Helps IHME Manage Multi-Billion Row Tables

Challenge: Exponential Growth Pushes Storage Engine to the Limit

Commissioned by the World Bank in the 1990s, the Global Burden of Disease (GBD) project, led by the Institute for Health Metrics and Evaluation (IHME), serves as the most comprehensive effort to date to measure disability and death from a multitude of causes worldwide. Today, the study has grown into an international consortium of more than 2,300 researchers from more than 130 countries – and its estimates are being updated annually.

As a result of GBD's continued growth, the data requirements have been scaling rapidly:

- In 2010, there were approximately 2 billion data points
- By 2015, there were nearly 100 billion data points
- By 2016, results were expected to exceed 10 billion per table

GBD results are developed through many internal processes and pipelines that rely on a MySQL-compliant infrastructure. Unfortunately, the choice to adopt MySQL was made at a time when the scope of work was much smaller, and the growth in data to today's size was unimaginable.

In the beginning, IHME invested significant time and resources into performance tuning the legacy database environment to handle the query and data loading traffic from their 16,000 CPU-core high-performance computing cluster. But eventually, issues developed surrounding the platform's ability to handle continued, exponential growth of data.

IHME needed to find a reliable solution that could scale with its skyrocketing data requirements.

Solution: MariaDB AX

Andrew Ernst, IHME's assistant director of infrastructure, explained:

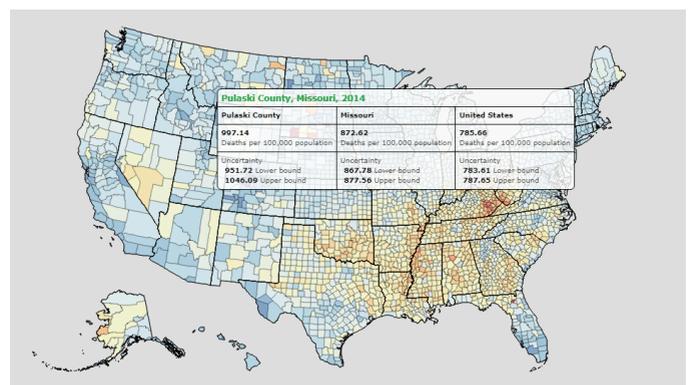


“With MariaDB AX, we were able to improve the performance of our multi-billion row tables.”

– *Andrew Ernst,*
IHME Assistant Director,
Infrastructure

When evaluating solutions, key criteria included cost-effectiveness, a MySQL-compatible interface – and ideally, an open source development offering with community-driven input and contributions. What's more, IHME did not want a solution that would require its team to rewrite vast amounts of research application code.

A couple top contenders were MemSQL and MySQL 5.7, but in the end, neither one made the cut. While MemSQL performed extremely well at scale, it had a number of non-standard constructs for database design and implementation, and lacked definable security mechanisms for authentication and authorization with their non-commercial product. MySQL 5.7 offered higher benchmark speeds for ingest and query optimization, but the platform could not support transformational improvements to scalability and workload efficiency. In particular, IHME



needed a platform that could leverage multi-host sharding with a multi-threaded software. Its database team is small, and while building a sharding infrastructure is reasonably straightforward, IHME knew it wouldn't be realistic to ask the development staff to make the applications shard-aware.

For IHME, the decision to select MariaDB AX – with its powerful open source storage engine – was an easy one. Key benefits include:

- **Higher Performance:** MariaDB AX's columnar storage engine reduces disk I/O, making it much faster than row-based storage for read-intensive analytic workloads on large datasets. Having quick access to results of aggregate functions allowed for better analysis of the data.
- **Improved Scalability and Efficiency:** MariaDB ColumnStore provides advanced data compression, which enabled IHME to compress data to approximately 40% of the original – thereby reducing the hardware needed for its massive data deployments.
- **Better Security:** MariaDB ColumnStore accesses all the same security capabilities delivered in MariaDB Server, including encryption for data in motion, role-based access control, and audit features.
- **Faster Analysis:** MariaDB ColumnStore lets IHME leverage SQL to bring transactional and analytic workloads into a single enterprise-grade system. It simplifies enterprise administration and execution with a standard SQL front end for OLTP and analytics. Therefore, IHME didn't have to change any existing application code that leverages standard SQL.

In summarizing the importance of IHME's new solution, Ernst said:



“High performance, flexible data analytics using MariaDB AX doesn't just make my day-to-day job easier. It will have a profound impact on how the global community can assess disease around the world.”



“Looking to the future, IHME has to design around a future of growing data that allows for regular updates as new data and epidemiological studies are made available. MariaDB AX, with the ColumnStore storage engine, solved both a volume and scale problem within our environment that allows us to seamlessly handle both current and planned increases in workload.”

About IHME

The Institute for Health Metrics and Evaluation (IHME) is an independent population health research center at UW Medicine, part of the University of Washington. In its role as the coordinating center for the Global Burden of Disease (GBD)—a decades-long project to systematically measure the world's health problems—IHME supports teams from around the world who collect and analyze data for the study.

About MariaDB

MariaDB Corporation is the company behind MariaDB, the fastest growing Open Source database. MariaDB is the default in major Linux distributions like Red Hat and SUSE, which in total reaches more than 60 million users. MariaDB can be deployed in a hybrid, public or private cloud with technologies like Docker, Microsoft Azure, Amazon Web Services and OpenStack. Over the past year, the company expanded its product portfolio to include [MariaDB MaxScale](#) and [MariaDB ColumnStore](#), enabling a broader range of use cases across the enterprise. MariaDB, with its commitment to community innovation and customer success, is the leading database preferred by developers and trusted by enterprises.