The Register

DATABASES

DuckDB promises greater stability with 1.0 release

POST A COMMENT

In-process OLAP database flies the nest

Lindsay Clark   Wed 5 Jun 2024 | 19:00 UTC

DuckDB has become a fully fledged database release with its 1.0 iteration, promising a new data model and greater stability to enhance backwards compatibility.

With more than a million downloads a month, the in-process analytical database management system has attracted a lot of attention since its 0.5.0 iteration, released in September 2022.

Speaking to *The Register*, DuckDB co-developer and founder of support company DuckDB Labs Hannes Mühleisen told *The Register* the 1.0 release was more about stability than new features.

Firstly, there is a new storage format.

"Every time we had a major release of DuckDB… you would have to reload your data into the system because the format would just change in non-compatible ways because we had been making a ton of changes. But now, we guarantee a backward compatibility and also limited forward compatibility with the storage format, which means that if you write data to a DuckDB database format, now, you will still be able to read that file 10 years down the line. That's a big change," he said.

He said the new file format fills a niche in the market as users can create multiple tables in a single file and have transactional updates to those files, which were efficient and compressed, he said.

DuckDB was born in Amsterdam's Centrum Wiskunde & Informatica mathematical and theoretical computing research center, where Mühleisen is a professor. Embedded within a host process, the database requires no DBMS server software to install, update or maintain. For example, the DuckDB Python package can run queries directly on data in Python software library Pandas without importing or copying data. Written in C++, DuckDB is free and open source under the MIT License.

Former Google BigQuery engineer Jordan Tigani pointed out that DuckDB bucks the trend of cloud-based scale-out data warehousing and takes advantage of more powerful laptops. He's such a fan he co-founded MotherDuck, a company which provides backend extension to DuckDB.

Hyoun Park, CEO and chief analyst at Amalgam Insights, on the other hand, said he sees DuckDB as a one trick pony whose trick is high-performance analytics and file transformations with limited resources.

"The file transformation capability is useful for transforming unstructured data in parquet or other unstructured formats into a performant in memory database. And the database is also useful for conducting high-performance analytics in edge environments, or remote environments, which will be increasingly useful for offloading analytic processing," he said.

Park told us DuckDB is also very easy to deploy and support.

Mühleisen said third-party technology firms were adopting DuckDB under the MIT license, as well as an increasing number of single use cases "where the data scientist is on their laptop."

For example, he said, DuckDB Labs was working with FiveTran to help it use the database in its Apache Iceberg table format implementation.

"That's something that didn't exist in the beginning — that people would just grab DuckDB and put it into their into the pipeline as a component — but that's really growing strongly," Mühleisen said.

Matthew Mullins, CTO at collaborative analytics company Coginiti, said: "As a tool builder we're most excited about DuckDB's close partnership with the Apache Arrow community because it's enabling us to build a new generation of highly performant data analysis tools that leverage columnar data formats. This integration not only boosts performance but also simplifies the data exchange process, enabling more efficient and scalable data operations.

MORE CONTEXT

DuckDB shuns VC breadcrumbs so support isn't all it's quacked up to be

VCs lay $52.5M golden egg for MotherDuck's serverless analytics platform

Ex-BigQuery exec and Motherduck CEO: For some users, the answer is to think small

Great Graph Database Debate: Abandoning the relational model is 'reinventing the wheel'

"Since its initial release, DuckDB has delighted data engineers, data scientists, and tool builders by being easy to work with and incredibly performant. Users love its friendlier SQL. It's become a favorite for handling CSV and Parquet files efficiently, no matter where they reside. Additionally, the capability to directly attach to PostgreSQL and MySQL databases for zero-ETL analysis has simplified data workflows and reduced the cost of moving data," Mullins said.

The ability to work on data in PostreSQL and MySQL comes through DuckDB plugins. Mühleisen said the team was hoping to develop a platform where people could upload, share and download DuckDB plugins for connectivity, or support for new scalar functions or index types, he said. ®

Other stories you might like

Silicon startups: Chip innovation poised to be dominated by China
INTERVIEW Coming SiB boom will be a Chinese one while the West is years behind, industry watcher tells us
On-Prem 6 Jun 2024 | 2

7-year-old Oracle WebLogic bug under active exploitation
Experts say Big Red will probably re-release patch in an upcoming cycle
Patches 6 Jun 2024 | 2

BT chief blames regulations for UK lagging in next-gen network rollout
Someone noticed Britain's about to elect a new govt, amirite?
Networks 6 Jun 2024 | 6

Modernizing storage for the age of AI
How Dell is boosting storage performance for AI with PowerScale
Sponsored Feature

Can AI models trained on human speech help us understand dogs?
What's that Lassie? Our nefarious AI overlords are about to take over the world? You must be barking
Science 6 Jun 2024 | 8

Microsoft Research chief scientist has no issue with Windows Recall
As tool emerges to probe OS feature's SQLite-based store of user activities
OSes 6 Jun 2024 | 20

Tokyo takes on Tinder by developing its own dating app it hopes will arrest population decline
We're from the government and we want you to find love and procreate
Public Sector 6 Jun 2024 | 7

Wi-Fi can watch your heart beating, Taiwan's datacenter power outfit Delta finds
COMPUTEX May come in handy for healthcare orgs, or to turn you off private 5G
Networks 6 Jun 2024 | 5

Google takes on virtual desktops with acquisition of app-streamer Cameyo

Legacy apps on Chromebooks is the goal, which won't terrify VDI incumbents like Citrix and Omnissa

Virtualization6 Jun 2024 |

Synology stretches into a backup appliance, with object stores to come

COMPUTEX Would like to vanquish Veeam and Veritas, but tidying up backup rigs is an easier job

Storage6 Jun 2024 | 2

FTX's $24B tax bill written down to just $200M

No, this isn't the weird logic of cryptocurrency infecting the IRS. It's a 'Fight us and everyone loses even more' kinda deal

Personal Tech6 Jun 2024 | 12

Elon Musk confirms 12K H100s ordered for Tesla were instead prioritized for xAI

COMMENT Tycoon's private firms got Nvidia GPUs first because automaker couldn't use them yet

AI + ML5 Jun 2024 | 8