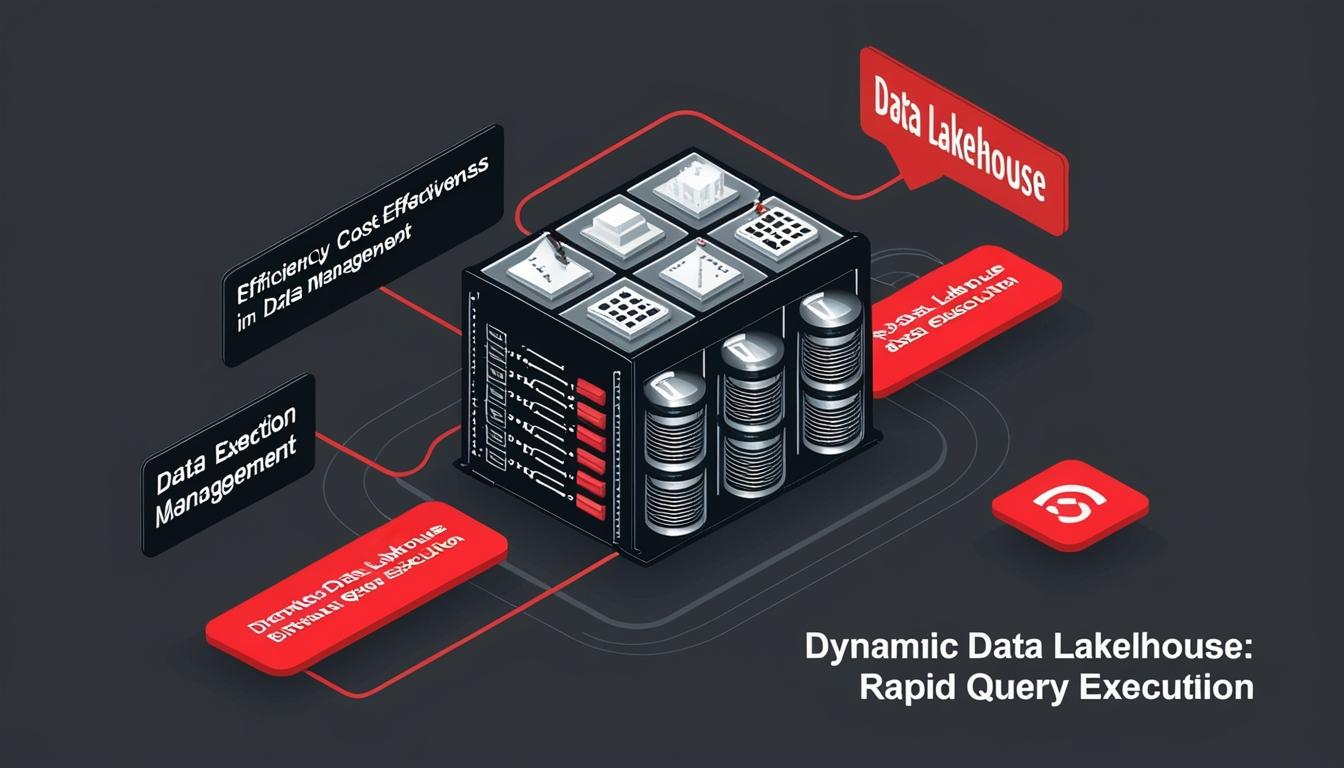
# Onehouse launches Compute Runtime to boost data lakehouse performance



Organisations are increasingly utilising data lakehouses to manage and analyse vast amounts of information; however, as the volume of stored data grows, so do the challenges associated with slow and costly queries. Onehouse, a technology vendor specialising in data lakehouse solutions and a prominent contributor to open source data formats such as Apache Hudi and Apache XTable, is addressing this issue with the launch of its new Onehouse Compute Runtime (OCR). This innovative system promises to enhance query performance by up to 30 times and potentially reduce costs by up to 80%.

The implementation of various open data table formats—including Apache Hudi, Apache Iceberg, and Delta Lake—is common in today’s data lakehouse environments. Onehouse is leading the way with the Apache XTable project, designed to promote interoperability among these formats. With the introduction of the OCR, the company aims to simplify the querying process across popular data platforms such as Amazon Redshift, Databricks, Google BigQuery, and Snowflake, enabling enterprises to operate more efficiently with their data.

Vinoth Chandar, the founder and CEO of Onehouse, articulated the need for an advanced runtime optimised specifically for lakehouse workloads. “We feel we need a specialized runtime that is optimized for lakehouse workloads,” Chandar stated in an exclusive interview with VentureBeat. He noted that many current vendors simply adapt their existing engines to operate with open table formats rather than optimising them for specific needs, which he considers a missed opportunity.

Challenges also arise from widely-used data processing frameworks like Apache Spark, which, while capable, lack the specific optimisations necessary for different table formats and lakehouse architectures. Kyle Weller, head of product at Onehouse, elucidated that formats such as Hudi and Iceberg serve as metadata abstractions that describe table constructions, but Spark remains a broad, generic framework that does not inherently facilitate the use of these formats without specialised knowledge and manual tuning.

The design of Onehouse Compute Runtime includes three central components: adaptive workload optimisations, high-performance lakehouse input/output (I/O), and serverless compute management. These features enable the runtime to intelligently tune execution based on historical workload patterns, thus automatically optimising data ingestion and query processes, particularly in cases where data partitioning and organisation have often presented challenges for users.

One of the early adopters of the Onehouse Compute Runtime is Conductor, a digital optimisation vendor. Emil Emilov, principal software engineer at Conductor, shared insights into the benefits his company has observed. With Onehouse acting as their central data store, Conductor has been able to streamline data ingestion and enhance the freshness of insights derived from their analytics. Emilov stated, “Onehouse Compute Runtime also accelerates query performance, which means faster access to those insights,” ultimately leading to improved customer satisfaction.

While the enhancements to query speed and efficiency are significant, the financial implications of these performance improvements are equally noteworthy. Chandar emphasised that “cost and performance are two sides of the same coin” in the context of lakehouse operations. By strategically optimising data organisation, the Onehouse Compute Runtime holds the potential to lower the overall compute costs associated with running extensive workloads, demonstrating its dual impact on both performance and expenditure for organisations seeking to harness data more effectively.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://aws.amazon.com/marketplace/pp/prodview-zw7rtdegdab5g> - This link corroborates the information about Onehouse Managed Lakehouse, its features, and how it enhances data lakehouse management, including continuous data ingestion, automated data management, and support for various query engines.
2. <https://www.youtube.com/watch?v=XCrGB-FVf6Q> - This video explains the concept of a universal data lakehouse, the role of Apache Hudi, and how Onehouse integrates these technologies to provide a unified data storage solution, which supports the claims about Onehouse's capabilities.
3. <https://www.onehouse.ai/solutions/optimize-lakehouse-tables> - This page details the optimization of lakehouse tables using Onehouse, including performance acceleration, cost reduction, and automated maintenance, aligning with the benefits mentioned in the article.
4. <https://aws.amazon.com/marketplace/pp/prodview-zw7rtdegdab5g> - This link also supports the information about Onehouse's support for various open data table formats such as Apache Hudi, Apache Iceberg, and Delta Lake, and its interoperability efforts.
5. <https://www.onehouse.ai/solutions/optimize-lakehouse-tables> - This page further explains the optimization techniques used by Onehouse, including table cleaning, clustering, compaction, and file-sizing, which are crucial for enhancing query performance and reducing costs.
6. <https://www.youtube.com/watch?v=XCrGB-FVf6Q> - The video transcript discusses the challenges of maintaining separate data warehouses and data lakes and how Onehouse addresses these issues by providing a unified, serverless, and auto-optimized solution.
7. <https://aws.amazon.com/marketplace/pp/prodview-zw7rtdegdab5g> - This link provides details on how Onehouse enables data ingestion from various sources and supports multiple query engines, which is essential for the efficient operation of data lakehouses across different platforms.
8. <https://www.onehouse.ai/solutions/optimize-lakehouse-tables> - This page highlights the importance of observability and monitoring in optimizing lakehouse tables, which aligns with the article's mention of tracking key metrics and optimizing storage and performance.
9. <https://www.youtube.com/watch?v=XCrGB-FVf6Q> - The video explains how Onehouse simplifies the entire data pipeline by shifting from separate batch and stream data pipelines to a single incremental workflow, reducing the need for manual tuning and maintenance.
10. <https://aws.amazon.com/marketplace/pp/prodview-zw7rtdegdab5g> - This link supports the claim that Onehouse allows customers to control their data within their own cloud space and avoid vendor lock-in, promoting flexibility and choice in data processing engines and tools.
11. <https://venturebeat.com/data-infrastructure/apache-hudi-creator-onehouse-debuts-compute-runtime-with-up-to-30x-faster-data-lakehouse-queries/> - Please view link - unable to able to access data