

Introducing Apache Hive 4.0 GA: Revolutionizing Data Warehousing

The Apache Software Foundation (ASF) proudly announces the release of Apache® Hive 4.0, marking a significant milestone in the realm of distributed data warehousing. For over a decade, Apache Hive has been the cornerstone of data warehouse and data lake architectures, empowering Companies and organisations worldwide to perform analytics at an unprecedented scale while seamlessly managing vast amounts of data through SQL queries.

A Legacy of Innovation

Since its inception in 2010, Apache Hive has continuously evolved to meet the ever-growing demands of modern data management. At its core, Hive offers a distributed, fault-tolerant data warehouse system known for its scalability and reliability. With support for Kerberos authentication and seamless integration with Apache Ranger and Apache Atlas for enhanced security and observability, Hive has become one of the best go-to solutions for enterprises seeking robust data management solutions.

Empowering the Data Ecosystem

At the heart of Apache Hive lies the Hive Metastore (HMS), a centralized repository of metadata that serves as a fundamental building block for data lakes leveraging a myriad of open-source technologies, including Apache Spark, PrestoDB and TrinoDB. The Hive Metastore facilitates seamless access to metadata for various clients, including Hive, Impala, and Spark, making it a vital component of the modern data ecosystem.

What's New in Apache Hive 4.0

The latest release, Apache Hive 4.0, represents a culmination of efforts from a vibrant community of developers, featuring over 5000 commits aimed at introducing new features, bug fixes, and performance enhancements. You can find a comprehensive list of major changes [here](#). Among the key highlights of Apache Hive 4.0 are:

- **Hive Iceberg Integration:** Streamlining data management with seamless integration of Iceberg tables.
- **Improved Transaction and Locking Capability:** Enhancing the ACID compliance of Hive with improved transaction handling and locking mechanisms.
- **Table Maintenance:** Introducing compaction mechanisms for both Hive ACID and Iceberg tables to optimize storage and performance.
- **Hive Docker Support:** Simplifying deployment with official Apache Hive Docker images for easier setup and configuration. Explore the Docker images on [DockerHub](#) for seamless deployment.
- **Compiler Improvements:** Anti-join support, branch pruning, column histogram statistics, HPL/SQL support, scheduled queries, new and improved cost-based optimization (CBO) rules leading to better query plans.
- **Materialized Views Support:** Enabling the creation and management of materialized views for accelerated query processing.
- **Runtime Optimizations:** Enhancing query performance with optimizations in Tez and LLAP, ensuring faster data processing.
- **Hive Replication:** Introducing improved replication features both for external and ACID tables for efficient data distribution and disaster recovery.
- **Support for Apache Ozone:** Apache Hive 4.0 introduces support for Apache Ozone, enabling seamless integration with Ozone-based object stores for scalable and efficient storage solutions.

Additional Changes

- **Hive CLI Deprecated:** The Hive command-line interface (CLI) is deprecated in favor of Beeline, the new JDBC client for Hive.
- **Reduced Supported Execution Frameworks:** Hive on MapReduce (MR) is deprecated and Hive on Spark is removed in favor of Hive on Tez. This shift reflects the community's commitment to leveraging modern execution frameworks for improved performance and scalability.
- **Support for Aarch64 (ARM) Chips:** Expanded compatibility to a broader range of hardware architectures, including ARM-based systems.
- **JWT Authentication:** Added support for JSON Web Token (JWT) authentication in HiveServer2 and HiveMetaStore, enhancing security for Thrift over HTTP interactions. This addition ensures robust authentication and authorization, fortifying data transactions within the Hive ecosystem.
- **New UDFs:** Several new User-Defined Functions (UDFs) to extend the functionality of Hive. For a comprehensive list of all UDFs, including the new ones introduced in Apache Hive 4.0, refer to the [Hive UDFs page](#).

Join the Apache Hive Community

Apache Hive 4.0 represents one of the most significant releases to date, offering unparalleled capabilities for managing and analyzing data at scale. Whether you're a data engineer, analyst, or architect, Apache Hive empowers you to unlock the full potential of your data infrastructure.

Get started with Apache Hive today and join the thriving community of users and contributors shaping the future of data warehousing and analytics. Visit the [Apache Hive website](#) to download the latest release and join the conversation on our [mailing](#)

[lists](#) and [GitHub repository](#). Follow us on [Twitter](#) for the latest updates and announcements.