**TÌM HIỂU HIVE VÀ SO SÁNH VỚI SQL**

HIVE: Sử dụng những câu lệnh như SQL và biên dịch những câu lệnh này thành tập hợp các tác vụ MapReduce.

|  |  |  |
| --- | --- | --- |
| **Tiêu chí so sánh** | **HIVE** | **SQL** |
| Phát triển bởi | Facebook | ISO/IEC (1986) |
| Mục đích sử dụng | Báo cáo, truy xuất nhanh dữ liệu, trả lời những câu hỏi cấp thiết, gần gũi với SQL | Quản lý dữ liệu quy mô vừa và nhỏ |
| Ngôn ngữ | HiveQL (SQL query) | SQL-92 standard |
| Kiểu dữ liệu | Có thêm hỗ trợ array, map và struct | Int, float, text, string, boolean, datetime... |
| JOIN | Hỗ trợ outer join và Equi-join. | Hỗ trợ đầy đủ inner join, outer join, cross join, seft join. |
| Lược đồ quan hệ (Schema) | Tường minh (Explicit) | Có |
| Chức năng CRUD | Ban đầu chỉ có Insert, không có Update, Delete. Sau này bổ sung đầy đủ chức năng. | Đầy đủ |
| Hỗ trợ transaction | Có (Không) | Có |
| Hỗ trợ stored procedure | Không | Có |
| Đánh chỉ số Index | Có (Không) | Có |
| Kích thước dữ liệu | Petabyte (1,000,000 Gigabyte) | Terabyte (1,000 Gigabyte) |
| Kích thước mỗi câu truy vấn | Petabyte (1,000,000 Gigabyte) | Gigabyte |
| Sử dụng bởi | Analysts | Programmers |
| Làm việc trên | Server side (Thrift) trên một cluster | Client side trên một cluster |
| User defined function | Java | SQL User-Defined Functions |
| MapReduce | Có | Không |
| Web interface | Có | Có |
| Cấu trúc dữ liệu | Có | Có |
| JDBC/ODBC | Giới hạn | Có (đầy đủ) |
| Trường hợp sử dụng | Dùng Big Data xử lý và phân tích dữ liệu nhanh chóng. Là lựa chọn tốt cho truy vấn, phân tích tập dữ liệu lịch sử | cung cấp khả năng truy vấn, cập nhật và bảo đảm tính toàn vẹn của dữ liệu một cách tối đa, phù hợp cho doanh nghiệp vừa và nhỏ. |

* Store Procedure: dù Hive không hỗ trợ viết stored trực tiếp trên Hive nhưng ta có thể viết ứng dụng ngôn ngữ khác để chạy stored procedure.
* Lịch sử của Hive:
* Không hỗ trợ kiểu dữ liệu data, timestamp và các hàm thao tác ngày giờ.
* Thao tác trên 1 bảng
* Không dử dụng IN, CONCAT, UNION, INTERSECT,
* Null là null và empty là \n.
* Dù HQL được xem là tương tự SQL nhưng Hive được xem là phần mở rộng của ETL hơn là một kho dữ liệu