

Hive Notes



[More Detailed Info](#)

hBrunch



È un software di Datawarehouse che facilita lettura, scrittura e gestione di ampi dataset in DFS usando una sintax SQL.

Hive è costituito su Hadoop e fornisce le seguenti features:

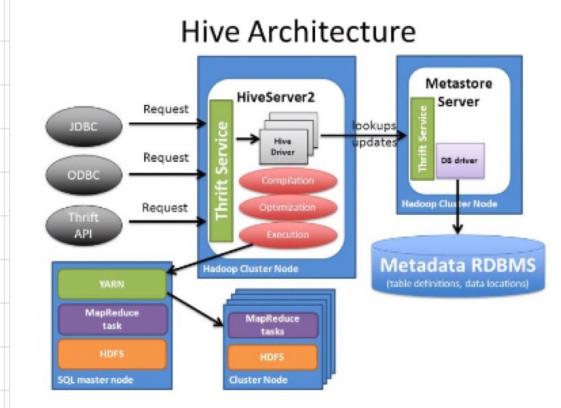
- Tools per accedere ai dati via SQL per ETL, reporting e analisi
- Un meccanismo per imposta una struttura ad una varietà di dati formati
- Accesso ai dati diretti su HDFS o HBase
- Esecuzione di query tramite SPARK, MAPREDUCE o TBZ
- Linguaggio Procedurale con HPL-SQL
- Solo second query retinendo

Hive fornisce SQL fin dal 2006. Può anche essere esteso tramite UDF, UDAF e UDTF. Legge vari formati.

quali Parquet, ORC ecc... Hive non è ideale per OLTP ma ottimo per OLAP. Maximizza scalabilità, performance, estendibilità, fault tolerance con il suo input format. Include:

- **HCATALOG**: una tabella di storage management per Hadoop che permette all'utente di leggere e scrivere più facilmente
- **WEBUICAT**: servizio per eseguire l'applicazione attraverso SSOs.

Scopo principale di Hive lungo è quello di dare una struttura ad un insieme di dati non strutturati per poter poi usare SQL su di essi, senza dover conoscere SPARK, MAPREDUCE o linguaggi di programmazione.



TIPI DI TABELLE - MANAGED

Sia i dati che gli schemi sono gestiti da Hive. I dati sono in una cartella in un file su HDFS. La location può essere configurata dall'utente all'installazione di Hive, quella di default è /user/hive/warehouse.

Se una tabella managed viene eliminata Hive cancella schema e dati.

TIPI DI TABELLE - EXTERNAL

Sono tabelle in cui solo lo schema è gestito da Hive, mentre i dati restano indipendenti.

Managed vs. External Table - What's the Difference?

Managed Table

Hive assumes that it owns the data for managed tables.

External Table

Hive assumes that it does not manage the data.

If a managed table or partition is dropped, the data and metadata associated with that table or partition are deleted.

For managed tables, Hive stores data into its warehouse directory.

Managed table provides ACID/transactional action support.

Statements: ARCHIVE, UNARCHIVE, TRUNCATE, MERGE, CONCATENATE supported

Query Result Caching supported (saves the results of an executed Hive query for reuse)

External Table

Hive assumes that it does not manage the data.

Dropping the table does not delete the data, although the metadata for the table will be deleted.

For external tables, Hive stores the data in the LOCATION specified during creation of the table (typically not in warehouse directory).

External Table does not provide ACID/transactional action support.

Not supported.

External Table does not provide ACID/transactional action support.

Not supported.

External Table does not provide ACID/transactional action support.

Syntax to Create Managed Table

```
CREATE TABLE IF NOT EXISTS stocks (exchange STRING,  
symbol STRING,  
price_open FLOAT,  
price_high FLOAT,  
price_low FLOAT,  
price_adj_close FLOAT)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
```

Syntax to Create External Table

```
CREATE EXTERNAL TABLE IF NOT EXISTS stocks (exchange STRING,  
symbol STRING,  
price_open FLOAT,  
price_high FLOAT,  
price_low FLOAT,  
price_adj_close FLOAT)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
LOCATION '/data/stocks';
```