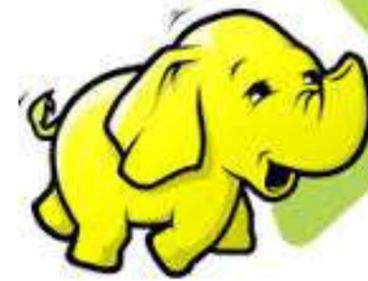




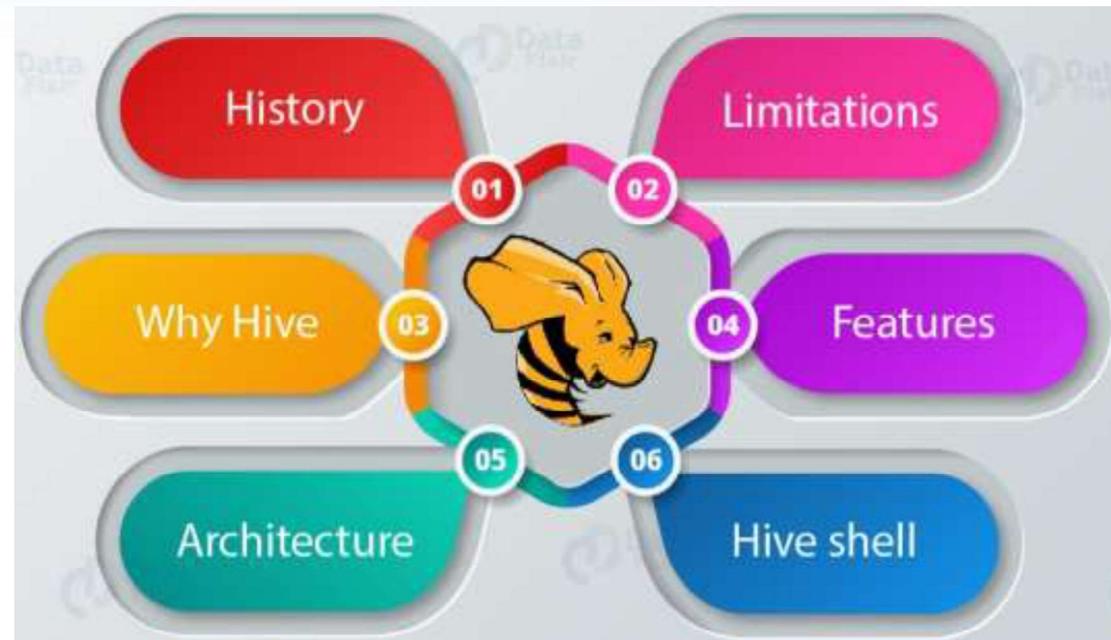
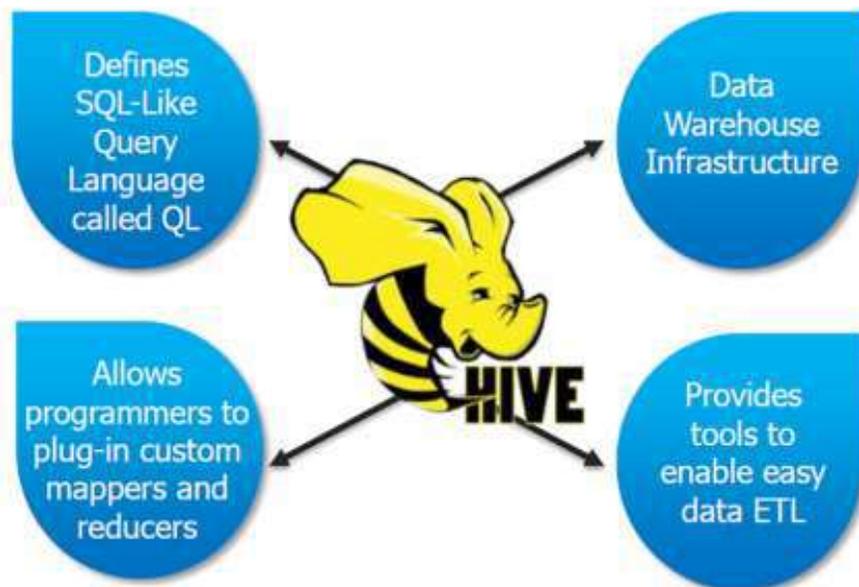
रा.इ.सू.प्रौ.सं
NIELIT



National Institute of Electronics and Information Technology

Big Data and Hadoop

Module 5-Features of Hive



History

1. When Facebook started gathering data and ingesting it into Hadoop in 2006, the data was incoming at the rate of 10s of GBs/day and in few years increased to around 15 TBs/day.
2. Apache Hive was co-created by **Joydeep Sen Sarma** and **Ashish Thusoo** during their stint at **Facebook**.
3. Both of them realized that in order to make the best use of Hadoop, they would have to write some fairly complex Java Map-Reduce jobs.
4. They realized that they would not be able to teach their fast growing engineering and analyst teams the skill set needed to be able to exploit Hadoop across the organization.
5. SQL was the interface used widely by the engineers and analysts. While SQL could cater to most analytics requirements, the creators also wanted to bring in the programmability that Hadoop provides.
6. Apache Hive was born out of these dual goals - an SQL-based declarative language that also allowed engineers to be able to plug in their own scripts and programs when SQL did not suffice.

History

7. To facilitate the creation of data driven organizations, it was also built to store centralized metadata(Hadoop based) about all the datasets in the organization.
8. Apache Hive helps developers get away with writing complex MapReduce tasks. Hadoop Hive is extremely fast, scalable, and extensible.
9. Additionally, the Hive is capable of decreasing the complexity of MapReduce by providing an interface wherein a user can submit various SQL queries. So, technically, you don't need to learn Java for working with Apache Hive.



Joydeep Sen Sarma

Ashish Thusoo

Apache Hive	
Original author(s)	Facebook
Initial release	October 1, 2010
Stable release	2.3.7 / August 26, 2019
Repository	github.com/apache/hive
Written in	Java
Available in	SQL
Type	Data warehouse
License	Apache License 2.0
Website	hive.apache.org

Why Hive?

1. Don't you think writing MapReduce jobs is tedious work? The Hive, you can just go ahead and submit SQL queries and perform MapReduce jobs for easy execution and processing of extremely large volumes of data.
2. The Hadoop Hive classifies data into tables providing a method for attaching the structure to data stores in HDFS. Processing structured and semi-structured data can be done by using Hive.
3. Apache Hive is an open-source data warehouse system that has been built on top of Hadoop. Hive has a distinct advantage of deploying high-speed data reads and writes within data warehouses while managing large datasets distributed across multiple locations, all thanks to its SQL-like features.
4. Hive provides a structure to the data that is already stored in the database. Users are able to connect with Hive using a command-line tool and a JDBC driver.
5. Similar to SQL the Hive has its own query language, called HiveQL (HQL). HiveQL also supports MapReduce scripts that can be plugged into the queries.
6. Hive increases schema design flexibility and also data serialization and deserialization.
7. Apart from Facebook now, many companies, such as IBM, Amazon and Yahoo! are also using and developing Hive.

Features

1. Facility for data summarization, analysis and querying a large data systems.
2. Supports external tables, making it feasible to process data without having to store it into HDFS.
3. Hadoop can process external data using Hive.
4. Supports partitioning of data at the data level for better performance.
5. A rule-based optimizer for optimizing logical plans.

Limitations of Apache Hive

1. Doesn't offer any real-time queries.
2. Online transaction processing is not well-supported.
3. There can be a delay while performing Hive queries.

Supported FileSystems

1. Flat files or text files
2. Sequence files consisting of binary key–value pairs

Criteria	Hive
Query language	HiveQL (SQL-like)
Used for	Creating reports
Area of deployment	Server side
Supported data type	Structured
Integration	JDBC and BI tools