HACHECHE SARA 5672186

A/ Apache Hive is an open source data warehouse system. It allows querying and analyzing large data sets Big Data stored in Hadoop files , and we call it apache hive because it is developed in software for processing and analyzing big data documents .

B / The importance and role of appache hive is it allows companies to process huge amounts of data in a short time. This is accomplished by facilitating the use of parallel computer processing on a massive scale, and not to mention that it is accessible and easy to learn. It offers modules with many options and it can easily be used on several machines to adapt to almost any size of data set. that's why apache hive is very important these days and it's very economical and it helps us to save a lot of time

C/

1/ **Schema Browsing** this process is intended to make understand its conformation and or its operation plan Diagram of the combustion engine and there is an architecture that allows you to understand much better work plans and facilitate tasks and better understand what to do .

2/  **Detached Query Execution** this method of queries is easy to use and it allows to combine and perform and work on data and open several windows in a database and as it allows to open several requests at the same time it allows the user to go back when he wishes .

3 / **No Local Installation** with hive we can work without even installing stuff on our computers but we can work online on webs that's why hive is very popular it's because it's accessible by everyone and without any problem and easy to manipulate .

4 / **Authentication** the haddop uses some settings to connect so for the first time we come in and it asks us to enter a username and groups etcthis last was added because it is more efficient for installation media .

D / The main idea of ​​partitioning in Hive is to organize data in a table based on the values ​​of one or more specific columns. T It improves query performance by allowing the elimination of entire partitions that are not relevant for a given query, Hive helps optimize query performance, makes it easier to handle large data, and improves join operations.

2/

|  |  |  |
| --- | --- | --- |
| MapReduce | HIVE | PIG |
| 1- MapReduce is a low level programming model for distributed data processing  2- MapReduce divides the data into smaller blochs and performs two main operations: map and reduce.  3- The map operation processes the input data | 1- Hive is a data warehousing infrastructure with a SQL-like query language  2- Hive allows users to write SQL and processing structured data  3- HIVE process and analyze large datasets stored in Hadoop Distributed File System | 1- Pig is a high-level data processing platform with a scripting language.  2- to express data transformations and analysis operations  3- Pig is suitable for processing unstructured and semi-structured dat |

3/

A/ Users prefer Pig Latin for a lot reasons because Pig Latin is a ETL ) data pipelines , which means that users can specify the desired data processing operations without worrying about the details and it allows for easy manipulation of large amounts of data.

B / Pig is very important because it simplifies and speeds up the processing of big data within the framework of the Apache Hadoop framework ,

C /

1 /**Pigs eat anything** is Pig can operate on data with or without metadata, whether relational, nested or unstructured.

2 /**Pigs live anywhere** Pig is not tied to a single parallel framework and can be implemented on different platforms, such as Hadoop.

3 /**Pigs are domestic animals** Pig is designed to be easily controlled and modified by users. It allows the integration of user code and supports user-defined functions,

4 /**Pigs fly** Pig processes data quickly and aims to constantly improve performance without weighing down the system.

D /

**1/ In 2007** Project Pig was initiated as a research project at Yahoo! Research, where scientists from Yahoo! designed Pig and made a first implementation

**2/ In 2008** Pig was featured in a paper at the SIGMOD conference and opened as an open source project through the Apache Incubator.

**3/ In 2009** Pig gained adoption, becoming a sub-project of Apache Hadoop and being used by other companies, including Amazon for its Elastic MapReduce service.

**4/ In 2010** Pig's popularity continued to grow, and it became a standalone project within the Apache Foundation, distinguishing itself as a top-level Apache project.

4/













