Kiểm tra giữa kì môn Big Data Analysis – HK2

1. C
2. B
3. C
4. A
5. C
6. C
7. B
8. A
9. A
10. A
11. B
12. C
13. A
14. A
15. B
16. B
17. D
18. A
19. B
20. A
21. A
22. B
23. B
24. A
25. B
26. C
27. B
28. B
29. A
30. D
31. A
32. D
33. **So sánh pig và hive**

**1. Pig :**

- Pig được sử dụng để phân tích một lượng lớn dữ liệu. Và được sử dụng để thực hiện tất cả các loại thao tác xử lý dữ liệu trong Hadoop. Nó cung cấp ngôn ngữ Pig-Latin để viết mã.

- Apache Pig có hai phần là Pig-Latin và Pig-Engine. Pig Engine được sử dụng để chuyển đổi tất cả các tập lệnh này thành MAP mọt cách cụ thể và giảm bớt các nhiệm vụ. Apache Pig chứa ít dòng mã hơn so với MapReduce.

+ Để lưu trữ dữ liệu trong pig không cần tạo lược đồ.

+ Trình biên dịch của Pig sẽ dịch Pig Latin thành các chuỗi chương trình MapReduce.

**2. Hive :**

- Hive được xây dựng phía trên trong kiến trúc của Hadoop và được sử dụng để xử lý dữ liệu có cấu trúc trong Hadoop.

- Nó cung cấp nhiều loại ngôn ngữ truy vấn khác nhau thường được gọi là Ngôn ngữ truy vấn Hive. Apache Hive là một kho dữ liệu và cung cấp giao diện giống SQL giữa người dùng và hệ thống tệp phân tán Hadoop (HDFS).

+ Nó là một cơ sở hạ tầng kho dữ liệu

+ Hive sử dụng một ngôn ngữ gọi là HQL và nó khá giống với SQL.

+ Để dễ dàng trích xuất, chuyển đổi và tải dữ liệu, nó cung cấp một số công cụ.

+ Trong Hive, chúng ta có thể sử dụng và xác định mapper và reducer tùy chỉnh.

+ Đối với công việc liên quan đến phân tích dữ liệu và báo cáo, nó được ưu tiên hơn.

**Sự khác biệt giữa Pig và Hive:**

|  |  |  |
| --- | --- | --- |
| STT | Pig | Hive |
| 1. | Pig hoạt động ở phía client của một cụm. | Hive hoạt động trên phía máy chủ của một cụm. |
| 2. | Pig sử dụng ngôn ngữ pig-latin. | Hive sử dụng ngôn ngữ HiveQL. |
| 3. | Pig là một ngôn ngữ luồng dữ liệu thủ tục. | Hive là một ngôn ngữ SQLish so sánh. |
| 4. | Nó được phát triển bởi Yahoo. | Nó được phát triển bởi Facebook. |
| 5. | Nó được sử dụng để xử lý dữ liệu có cấu trúc và bán cấu trúc. | Nó chủ yếu được sử dụng để xử lý dữ liệu có cấu trúc. |
| 6. | Nó được sử dụng để lập trình. | Nó được sử dụng để tạo báo cáo. |
| 7. | Tập lệnh Pig kết thúc bằng đuôi .pig | Trong Hive, tất cả các phần mở rộng đều được hỗ trợ. |
| 8. | Không hỗ trợ phân vùng. | Hỗ trợ phân vùng. |
| 9. | Tải dữ liệu nhanh chóng. | Tải dữ liệu chậm. |
| 11. | Không hỗ trợ JDBC | Hỗ trợ JDBC |
| 12. | Không hỗ trợ ODBC | Hỗ trợ ODBC |
| 13. | Pig không có cơ sở dữ liệu siêu dữ liệu chuyên dụng. | Hive sử dụng biến thể chính xác của ngôn ngữ SQL-DDL chuyên dụng bằng cách xác định trước các bảng. |
| 14. | Nó hỗ trợ định dạng tệp Avro. | Nó không hỗ trợ định dạng tệp Avro. |
| 15. | Pig thích hợp cho các cấu trúc dữ liệu phức tạp và lồng nhau. | Hive thích hợp cho các hệ thống OLAP xử lý hàng loạt . |
| 16. | Pig không hỗ trợ lược đồ để lưu trữ dữ liệu. | Hive hỗ trợ lược đồ để chèn dữ liệu trong bảng. |

Tham khảo

<https://www.geeksforgeeks.org/difference-between-pig-and-hive/>

<https://data-flair.training/blogs/pig-vs-hive/>