

Tin Tức /

MapReduce Là Gì? Khám Phá Mô Hình Lập Trình MapReduce

MapReduce là gì? Khám phá mô hình lập trình MapReduce

🕒 01/10/2023

➡ Chia sẻ

Để tăng sức mạnh cho Hadoop, MapReduce là một yếu tố quan trọng không thể thiếu. Vậy, MapReduce là gì? Cách hoạt động của chúng ra sao? Nhờ vào khả năng xử lý và phân tích một khối lượng lớn kết quả tìm kiếm, MapReduce đã dần dần trở nên phổ biến hơn và khẳng định vị thế của mình trong thời đại công nghệ số. Trong bài viết này, hãy cùng CMC Cloud tìm hiểu kỹ hơn về khái niệm MapReduce.

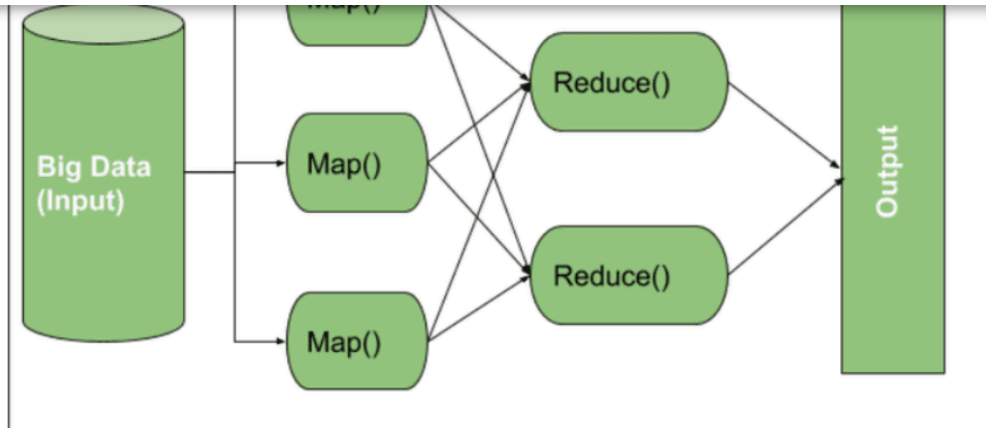
*Khái niệm MapReduce là gì*

MapReduce là gì?

MapReduce là một mô hình được Google phát triển độc quyền, nhằm mục đích xử lý các dữ liệu lớn theo hướng phân tán và song song thuật toán trong một cụm máy tính. Ban đầu, đây là mô hình độc quyền của Google. Nhưng theo thời gian, MapReduce dần trở thành thuật ngữ tổng quát hóa được nhiều người biết đến.

MapReduce bao gồm 2 thủ tục chính:

- **Thủ tục Map:** có vai trò lọc và phân loại dữ liệu
- **Thủ tục Reduce:** tổng hợp dữ liệu



Tìm hiểu khái niệm MapReduce là gì?

Mô hình này hoạt động dựa trên khái niệm biến đổi của bản đồ và thiết lập chức năng lập trình theo đúng hướng chức năng. Chúng ta có thể viết thư viện của thủ tục Map và Reduce bằng nhiều ngôn ngữ khác nhau, tùy theo nhu cầu.

Doanh nghiệp có thể cài đặt 2 thủ tục này miễn phí. Hiện nay, Apache Hadoop là một trong những thủ tục MapReduce phổ biến nhất.

Ưu điểm của MapReduce

MapReduce mang lại một số ưu điểm nổi bật:

- Xử lý tốt các dữ liệu khổng lồ nhờ vào sức mạnh tính toán và phân tích các tác vụ phức tạp một cách nhanh chóng
- Có thể chạy song song trên các máy tính phân tán, mang đến hiệu quả hoạt động cao cho hệ thống
- Có thể được xây dựng trên nhiều ngôn ngữ khác nhau như Ruby, C/C++, Python, Java, Perl,... với hệ thống thư viện hỗ trợ đa dạng
- MapReduce đang được hoàn thiện về mặt phát hiện và xử lý các mã độc, nhằm đảm bảo bảo mật cao nhất.

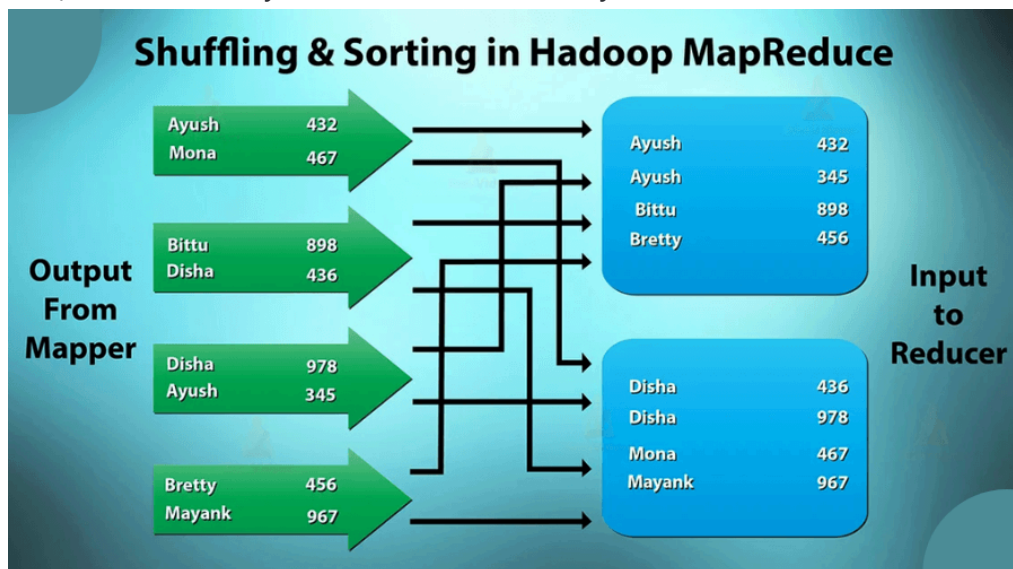
Các hàm của MapReduce

MapReduce sử dụng 2 hàm chính là Map và Reduce. Cả hai hàm này đều được định nghĩa từ người dùng và chúng nối tiếp nhau trong quy trình xử lý dữ liệu, cụ thể:

- **Map:** Xử lý một cặp khóa (key, value) để tạo ra một cặp khóa mới (keyl, valuel) có vai trò trung gian. Lúc này, doanh nghiệp chỉ cần tải

Reduce: Hợp nhất các khóa trung gian (key), value) và giá trị tương ứng của nó để tạo thành nhiều khóa khác nhau thông qua việc ghép nối chúng lại. Các cặp giá trị và khóa này sẽ được đưa vào Reduce thông qua con trỏ vị trí, giúp cho nhân viên IT dễ dàng quản lý được danh sách với một số lượng lớn và phân bổ giá trị hợp lý trên bộ nhớ của hệ thống.

- **Shuffle:** Ngoài 2 hàm chính trên, trong MapReduce còn có 1 hàm trung gian tên là Shuffle. Sau khi Map hoàn thành vai trò của nó, Shuffle sẽ thu thập và tổng hợp các cặp từ khóa và dữ liệu do Map tạo ra và di chuyển đến Reduce để xử lý.

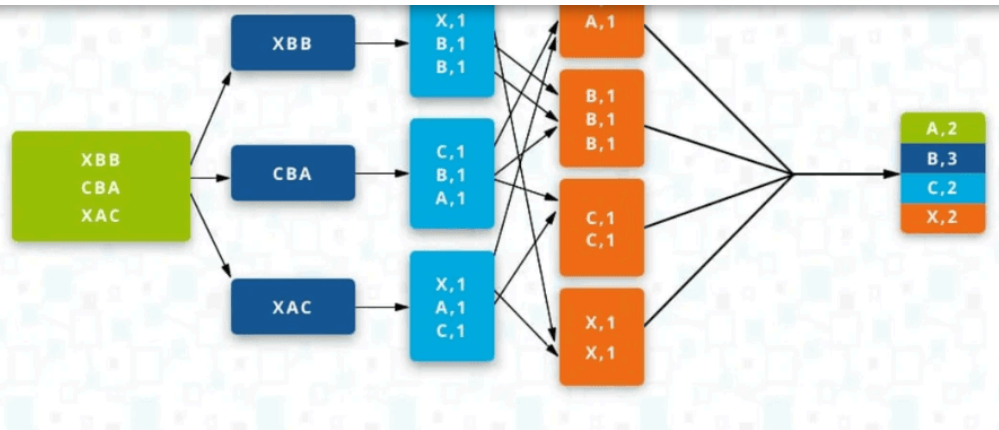


Các hàm trong MapReduce là gì?

Cách MapReduce hoạt động

MapReduce hoạt động thông qua 2 giai đoạn chính:

- **Map:** Lấy tập dữ liệu và chuyển đổi thành tập dữ liệu khác, mỗi một phần tử đều được tách nhỏ thành cặp khóa - giá trị (key-value)
- **Reduce:** Lấy thông tin từ Map làm thông tin đầu vào (input) và kết hợp các bộ khóa - giá trị thành các bộ giá trị nhỏ hơn



Cách hoạt động của MapReduce

Tác vụ Reduce luôn diễn ra sau tác vụ Map. Các cặp khóa - giá trị có thể khác nhau, tùy theo từng nhu cầu sử dụng. Tất cả các giá trị đầu vào và đầu ra (input và output) đều được lưu trữ trong HDFS.

Ví dụ về cách hoạt động của MapReduce

Để hiểu rõ hơn về cách hoạt động của MapReduce, hãy xem qua ví dụ sau. Twitter nhận khoảng 500 triệu tweet (bài đăng) mỗi ngày, chia ra là khoảng 3000 tweet mỗi một giây. Hệ thống MapReduce sẽ hoạt động giống như trong sơ đồ ở mục trên, gồm các hành động:

- Tokenize: Mã hóa từng tweet thành bản đồ token và viết lại chúng thành các cặp khóa - giá trị (key-value)
- Filter: Lọc các thông tin không cần thiết khỏi bản đồ token và viết lại bản đồ đã lọc thành các cặp khóa - giá trị
- Count: Có chức năng đếm token của mỗi một từ
- Aggregate Counters: Tổng hợp giá trị của những bộ đếm tương đồng nhau thành những đơn vị nhỏ để quản lý dễ hơn

Ứng dụng và tính năng của MapReduce

Ngày nay, MapReduce được ứng dụng trong nhiều mục đích khác nhau, ví dụ như:

Câu hỏi thông kê số lượng bài viết thì các từ nội bộ chứa các từ khóa này

- Thống kê các câu lệnh bên trong tài liệu, gồm lệnh Pattern hoặc Match
- Thống kê số lượng đường link (URLs) xuất hiện trong 1 Webpages
- Thống kê số lượng truy cập của người dùng vào từng địa chỉ URL khác nhau trên Website
- Thống kê các từ khóa trên trang Web,...

Trên đây là tất cả các thông tin về MapReduce là gì và các thông tin hữu ích khác về mô hình MapReduce, giúp hỗ trợ phần nào cho doanh nghiệp trong việc xử lý dữ liệu hiệu quả hơn. Đừng quên theo dõi Website CMC Cloud để cập nhật thêm nhiều thông tin công nghệ, lập trình mới nhất.

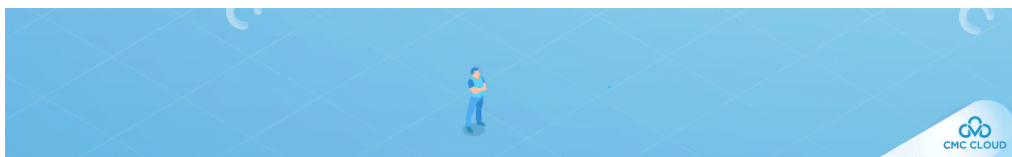
CMC Cloud - Giải pháp Điện toán đám mây Toàn diện & Linh hoạt nhất. Cho phép tùy biến sử dụng và quản trị 25+ dịch vụ

- Website: <https://cmccloud.vn>
- Facebook: <https://facebook.com/cmccloud.vn>
- Hotline: [1900.2010](tel:1900.2010)

CLOUD COMPUTING

CMC CLOUD

 Chia sẻ



Bài viết liên quan

Tối ưu hóa quản trị và bảo mật hệ thống với CLOUD IAM & CLOUD TRACE SERVICE



Thông tin về CMC Cloud IAM và Cloud Trace Service

Nguy cơ an ninh mạng mới và giải pháp bảo mật đám mây toàn diện từ CMC CLOUD



Nguy cơ an ninh mạng mới và giải pháp bảo mật đám mây toàn diện từ CMC Cloud

Giải pháp toàn diện cho quản lý dữ liệu: CLOUD BACKUP & ELASTIC VOLUME



Giải pháp toàn diện cho quản lý dữ liệu: Cloud Backup và Elastic Volume từ CMC Cloud

TĂNG TỐC ĐỘ VÀ HIỆU SUẤT: Khám phá sức mạnh của CMC Cloud trong hạ tầng mạng



Tăng tốc độ và hiệu suất: Khám phá sức mạnh của CMC Cloud trong hạ tầng mạng

Hiệu suất tính toán cao và tăng tốc xử lý tác vụ lớn với CMC CLOUD ELASTIC GPU



Hiệu suất tính toán cao và tăng tốc xử lý tác vụ lớn với CMC Cloud Elastic GPU

Cùng chuyên mục



Bảo mật hệ
thống với...



Môi và giải
pháp bảo...

Đăng ký nhận thông tin

Bạn muốn cập nhật thông tin mới nhất từ CMC Cloud?! Hãy để lại địa chỉ email của bạn ngay để nhận những bản tin bổ ích

Về CMC Cloud

[Giới thiệu](#)

[Tin tức](#)

[Tuyển dụng](#)

[Liên hệ](#)

[CMC Telecom Website](#)

Dịch vụ

[Compute](#)

[Storage](#)

[Networking](#)

[Database](#)

[Container](#)

[Management](#)

Dịch vụ nổi bật

Sao lưu, khôi phục dữ liệu

Điều khoản & Chính sách

Chính sách bảo mật

Cam kết bảo vệ dữ liệu cá nhân



© 2023, CMC Cloud. All rights reserved.

Hotline: **1900 2010**

Email: **support@cmccloud.vn**



Giấy ĐKKD: 0102900049. Nơi cấp: Sở Kế hoạch & Đầu tư Thành phố Hà Nội
Chịu trách nhiệm nội dung: Đặng Tùng Sơn