

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1

Cho dù bạn đang lập kế hoạch di chuyển dữ liệu một lần từ MySQL sang PostgreSQL hay muốn liên tục sao chép dữ liệu giữa hai cơ sở dữ liệu này, thì vẫn có nhiều điều cần cân nhắc và một số cách để thực hiện nhiệm vụ.

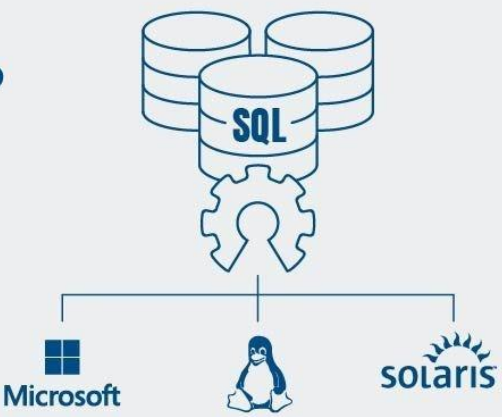
Trong bài viết này, chúng tôi sẽ hướng dẫn từng bước cách di chuyển MySQL sang PostgreSQL bằng Estuary Flow, cho phép bạn di chuyển dữ liệu lịch sử và liên tục thu thập dữ liệu mới bằng tính năng ghi lại dữ liệu thay đổi (CDC). Chúng tôi cũng sẽ khám phá thêm các phương pháp phổ biến khác để chuyển đổi MySQL sang PostgreSQL.

Nhưng trước tiên, hãy cùng tìm hiểu một số điều kiện tiên quyết thiết yếu. Để đảm bảo quá trình di chuyển thành công, điều quan trọng là phải hiểu những khác biệt chính giữa MySQL và PostgreSQL, đồng thời xác định mục tiêu di chuyển của bạn. Hãy bắt đầu bằng phần giới thiệu (lại) ngắn gọn về MySQL.

1. MySQL là gì?


What is MySQL?

MySQL is an open source, relational database management system (RDBMS) based on structured query language (SQL). MySQL is available on all major operating systems, including Windows, Linux and Solaris.




Free to Use

It is free to use for individuals and non-production environments under the GNU General Public License; however, if used commercially, a commercial licence is required.



Individuals



Non-Production Environments

MySQL là một hệ thống quản lý cơ sở dữ liệu quan hệ (RDBMS) được sử dụng rộng rãi dựa trên SQL. Thuộc sở hữu của Oracle, MySQL là một thành phần

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1

quan trọng trong nhiều nền tảng công nghệ, cho phép các nhóm xây dựng và duy trì các dịch vụ và ứng dụng dựa trên dữ liệu.

1.1 Ưu điểm

- Mã nguồn mở và miễn phí: MySQL là mã nguồn mở, cung cấp hỗ trợ cấp doanh nghiệp và triển khai ngay lập tức mà không mất phí.
- Hỗ trợ đa nền tảng: Hỗ trợ các hệ điều hành phổ biến như Windows, Linux và Solaris.
- Khả năng mở rộng: MySQL có thể xử lý hơn 50 triệu hàng và có khả năng mở rộng để quản lý các tập dữ liệu lớn.
- Bảo mật: Có lớp bảo mật dữ liệu mạnh mẽ, giúp xử lý thông tin nhạy cảm một cách an toàn.

1.2 Nhược điểm

- Hiệu suất theo quy mô: Hiệu quả của MySQL giảm dần khi dữ liệu mở rộng, đặc biệt là trong các tình huống truy vấn phức tạp hoặc ghi nhiều.
- Công cụ hạn chế: So với các cơ sở dữ liệu khác, MySQL có ít công cụ phát triển và gỡ lỗi nâng cao hơn.

2. PostgreSQL (Postgres) là gì?



PostgreSQL còn được gọi là Postgres, là một hệ thống quản lý cơ sở dữ liệu quan hệ đối tượng (ORDBMS) mã nguồn mở, cấp doanh nghiệp. Nó hỗ trợ cả truy vấn SQL (quan hệ) và JSON (phi quan hệ), mang lại khả năng mở rộng và thích ứng cao hơn MySQL.

2.1 Ưu điểm của việc sử dụng PostgreSQL

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1

- Hiệu suất: PostgreSQL xử lý các truy vấn phức tạp và dữ liệu quy mô lớn với hiệu quả cao hơn.
- Khả năng mở rộng: Hỗ trợ các kiểu do người dùng định nghĩa và nhiều ngôn ngữ thủ tục khác nhau.
- Tuân thủ ACID: PostgreSQL đảm bảo độ tin cậy cao và khả năng chịu lỗi với tính năng ghi nhật ký trước.
- Khả năng không gian địa lý: Hỗ trợ các đối tượng địa lý, lý tưởng cho các dịch vụ và ứng dụng dựa trên vị trí.

2.2 Nhược điểm của PostgreSQL

- Đường cong học tập: PostgreSQL có thể khó học và cấu hình hơn so với MySQL.
- Hiệu suất ở quy mô nhỏ hơn: Có thể chậm hơn đối với các hoạt động đơn giản, quy mô nhỏ.

3. Tại sao nên di chuyển từ MySQL sang PostgreSQL

Di chuyển cơ sở dữ liệu là một quá trình quan trọng mà các doanh nghiệp cần thực hiện để đáp ứng nhu cầu dữ liệu đang thay đổi, cải thiện hiệu suất hoặc giảm chi phí. Dưới đây là một số lý do tại sao bạn nên cân nhắc di chuyển từ MySQL sang PostgreSQL:

- Hiệu suất tốt hơn: Nếu cơ sở dữ liệu MySQL hiện tại của bạn gặp khó khăn với các truy vấn phức tạp hoặc tập dữ liệu lớn, PostgreSQL có thể cung cấp khả năng tăng cường hiệu suất mà bạn cần.
- Tính năng nâng cao: Hỗ trợ của PostgreSQL cho các kiểu dữ liệu nâng cao và khả năng không gian địa lý có thể mang lại sự linh hoạt hơn trong việc xử lý dữ liệu.
- Hiệu quả về chi phí: Di chuyển sang PostgreSQL có thể giảm chi phí lưu trữ và vận hành, đặc biệt nếu MySQL không được tối ưu hóa cho các trường hợp sử dụng của bạn.

3.1 Di chuyển dữ liệu so với sao chép dữ liệu

Điều quan trọng là phải phân biệt giữa di chuyển và sao chép cơ sở dữ liệu:

- Di chuyển cơ sở dữ liệu thường là một quá trình một lần trong đó dữ liệu được di chuyển từ MySQL sang PostgreSQL với mục đích chuyển hoàn toàn sang PostgreSQL.
- Sao chép cơ sở dữ liệu bao gồm việc sao chép dữ liệu liên tục giữa MySQL và PostgreSQL, cho phép bạn duy trì đồng bộ cả hai cơ sở dữ liệu.

Tùy thuộc vào nhu cầu, bạn có thể chọn giải pháp này hay giải pháp kia. Ví dụ, nếu ứng dụng của bạn phụ thuộc nhiều vào các truy vấn phức tạp, việc chuyển đổi hoàn toàn sang PostgreSQL có thể mang lại lợi ích. Tuy nhiên, nếu bạn cần sự

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1

đơn giản của MySQL cho một số tác vụ nhất định, thì sao chép liên tục có thể là giải pháp tốt hơn.

4. Phương pháp Di chuyển dữ liệu từ MySQL sang PostgreSQL

4.1 Phương pháp 1: Kết nối MySQL với PostgreSQL bằng Estuary

Estuary Flow là một nền tảng tích hợp dữ liệu thời gian thực mạnh mẽ được thiết kế để đơn giản hóa và tối ưu hóa quy trình sao chép giữa MySQL và PostgreSQL. Với Estuary, bạn có thể dễ dàng thiết lập nhiều tích hợp dữ liệu khác nhau, bao gồm MySQL với BigQuery và MySQL với Snowflake, cùng nhiều kết nối khác.

Tại sao nên chọn Estuary Flow để di chuyển từ MySQL sang PostgreSQL?

- Giao diện thân thiện với người dùng: Ứng dụng web trực quan của Flow chỉ yêu cầu chuyên môn kỹ thuật tối thiểu, giúp thiết lập và quản lý đường ống dữ liệu dễ dàng hơn so với các công cụ dòng lệnh hoặc công cụ có tập lệnh.
- Đồng bộ hóa dữ liệu theo thời gian thực: Estuary Flow thực hiện việc sao lưu toàn bộ dữ liệu MySQL của bạn vào PostgreSQL, sau đó truyền dữ liệu theo thời gian thực. Điều này đảm bảo cả hai cơ sở dữ liệu luôn đồng bộ trong vòng vài mili giây, giảm thiểu độ trễ dữ liệu.
- Sao chép hiệu quả về chi phí: Sử dụng tính năng ghi lại dữ liệu thay đổi (CDC) cho MySQL, Flow chỉ ghi lại và chuyển những thay đổi được thực hiện trên dữ liệu của bạn. Phương pháp này ít tốn tài nguyên tính toán hơn và giảm chi phí vận hành so với các công cụ di chuyển dữ liệu truyền thống.
- Độ tin cậy cao: Flow đảm bảo khả năng phục hồi dữ liệu bằng cách lưu trữ dữ liệu MySQL đã thu thập trong các bộ sưu tập được hỗ trợ bởi đám mây trước khi ghi vào PostgreSQL. Cơ chế này đảm bảo tính toàn vẹn của dữ liệu và cung cấp khả năng lưu trữ ngẫu nhiên chính xác một lần, bảo vệ dữ liệu khỏi nguy cơ mất mát.

Hướng dẫn từng bước: Di chuyển MySQL sang PostgreSQL bằng Estuary Flow:

- Bước 1: Đăng ký miễn phí trên Estuary: Bắt đầu bằng cách đăng ký tài khoản miễn phí trên nền tảng Estuary. Đối với nhu cầu sản xuất quy mô lớn, bạn có thể liên hệ với Estuary để đăng ký tài khoản tổ chức.
- Bước 2: Chuẩn bị cơ sở dữ liệu MySQL của bạn

Đăng nhập vào ứng dụng web Estuary và đảm bảo rằng cơ sở dữ liệu MySQL của bạn đáp ứng các điều kiện tiên quyết.

- Bước 3: Thu thập dữ liệu MySQL theo thời gian thực

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1


Điều hướng đến tab Ảnh chụp và chọn Ảnh chụp mới.

Chọn ô MySQL và điền các thông tin bắt buộc, bao gồm địa chỉ máy chủ MySQL, tên người dùng cơ sở dữ liệu và mật khẩu.

Chọn các bảng bạn muốn ghi lại và xuất bản cài đặt của mình.

Capture Details

Provide a unique name and specify a source system for your capture. Fill in the required details and click "Discover Endpoint" to test the connection.

Connector Search*
 MySQL

Name*
estuary/mysqlcapture

Details

Choose the external system you're connecting to.

Select a prefix from the drop-down and add a unique name. (ex: acmeCo/marketing_data)

Describe your changes or why you're changing things.

Endpoint Config

CONNECTOR HELP

Server Address*
host:port

The host or host:port at which the database can be reached.

Login Username*
flow_capture

The database user to authenticate as.

Login Password*
.....

Password for the specified database user.

• Bước 4: Chuyển sang Postgres

Chọn ô PostgreSQL trong phần Materialize.

Cung cấp thông tin đăng nhập cơ sở dữ liệu PostgreSQL cần thiết, bao gồm địa chỉ cơ sở dữ liệu, tên người dùng và mật khẩu.

Ảnh xạ dữ liệu MySQL đã thu thập vào các bảng mới trong PostgreSQL và xuất bản dữ liệu.

Estuary Flow hiện sẽ xử lý việc sao chép dữ liệu theo thời gian thực, đảm bảo rằng mọi dữ liệu mới trong MySQL đều được tự động thu thập và sao chép sang PostgreSQL.

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1



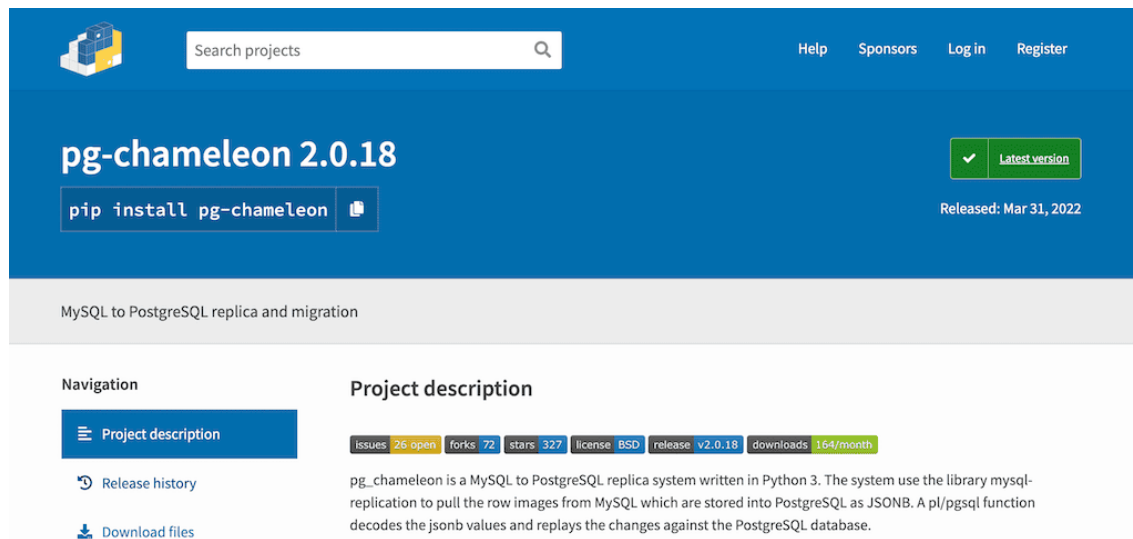
Để biết thêm trợ giúp về phương pháp này, hãy xem tài liệu Estuary Flow trên:

Cách tạo luồng dữ liệu

Trình kết nối MySQL

Trình kết nối PostgreSQL

4.2 Phương pháp 2: Sử dụng pg-chameleon cho MySQL để sao chép liên tục từ PostgreSQL



pg-chameleon là một công cụ Python tận dụng tính năng sao chép gốc của MySQL, chuyển đổi dữ liệu sang PostgreSQL. Phương pháp này lý tưởng cho việc sao chép liên tục, đảm bảo dữ liệu của bạn được đồng bộ hóa liên tục giữa MySQL và PostgreSQL. Tuy nhiên, lưu ý rằng quá trình này có thể gây ra một chút chậm trễ do bản chất của việc sao chép.

4.3 Phương pháp 3: Sử dụng pgloader làm công cụ di chuyển MySQL sang PostgreSQL

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1

pgloader là một công cụ dòng lệnh mạnh mẽ giúp đơn giản hóa quá trình di chuyển dữ liệu từ MySQL sang PostgreSQL. Công cụ này đặc biệt phù hợp cho việc di chuyển một lần và hỗ trợ nhiều nguồn dữ liệu khác nhau. Cách thiết lập như sau:

- Bước 1: Cài đặt pgloader: Thiết lập dễ dàng cho quá trình di chuyển của bạn

pgloader có sẵn thuận tiện thông qua hầu hết các trình quản lý gói. Hãy sử dụng các lệnh sau dựa trên hệ điều hành của bạn:

Debian/Ubuntu: `apt-get install pgloader`

macOS: `brew install pgloader`

Đối với các hệ thống khác, hãy tải xuống tệp nhị phân trực tiếp từ trang web pgloader chính thức.

- Bước 2: Thu thập thông tin xác thực cơ sở dữ liệu

Xác định vị trí thông tin kết nối MySQL và PostgreSQL của bạn trong các tệp cấu hình tương ứng:

MySQL: Thường được tìm thấy trong `my.cnf` (hoặc tương tự).

PostgreSQL: Thường được tìm thấy trong `postgresql.conf`.

Những thông tin chi tiết này thường bao gồm:

Tên máy chủ/Địa chỉ IP: Máy chủ nơi cơ sở dữ liệu đang chạy.

Cổng: Số cổng được sử dụng để kết nối cơ sở dữ liệu.

Tên người dùng: Tên người dùng có quyền truy cập vào cơ sở dữ liệu.

Mật khẩu: Mật khẩu được liên kết với tên người dùng.

Tên cơ sở dữ liệu: Tên của cơ sở dữ liệu cụ thể mà bạn muốn di chuyển.

Ngoài ra, bạn có thể thấy các thông tin xác thực này được đặt dưới dạng biến môi trường.

- Bước 3: Tinh chỉnh quá trình di chuyển của bạn bằng Tệp cấu hình (Tùy chọn)

Để kiểm soát và tùy chỉnh chi tiết hơn, hãy tạo một `.loadtệp`. Tệp này cho phép bạn xác định cách thức xử lý các khía cạnh khác nhau của quá trình di chuyển:

Ánh xạ kiểu dữ liệu: Chỉ định cách chuyển đổi kiểu dữ liệu MySQL sang kiểu dữ liệu PostgreSQL tương đương.

Chọn bảng: Chọn các bảng cụ thể để di chuyển hoặc loại trừ một số bảng nhất định.

Thao tác lược đồ: Kiểm soát việc tạo và sửa đổi lược đồ.

Lập chỉ mục: Quản lý việc tạo chỉ mục trong quá trình di chuyển.

	VIETTEL AI RACE	TD187
	CÁCH DI CHUYỂN MYSQL SANG POSTGRESQL	Lần ban hành: 1

```
LOAD DATABASE
FROM mysql://user:password@localhost/mydb
INTO postgresql://user:password@localhost/mypgdb
WITH include drop, create tables, create indexes
set MySQL variables charset = 'UTF8';
```

- Bước 4: Thực thi pgloader: Khởi chạy quá trình di chuyển MySQL sang PostgreSQL

Sau khi đã chuẩn bị xong thông tin kết nối và tệp cấu hình (tùy chọn), hãy bắt đầu di chuyển bằng lệnh sau:

```
Bash
pgloader mysql://[MySQL connection details] postgresql://[PostgreSQL connection detail]
```

Thay thế chỗ giữ chỗ bằng thông tin xác thực của bạn.

Nếu bạn đã tạo .loadtệp cấu hình, hãy đưa tệp đó vào lệnh:

```
Bash
pgloader --load [filename.load]
```

Để biết danh sách đầy đủ các tùy chọn và thông tin chi tiết về cấu hình, hãy tham khảo tài liệu pgloader chính thức.

5. Chọn công cụ đường ống dữ liệu phù hợp

Khi lựa chọn công cụ di chuyển MySQL sang PostgreSQL, điều quan trọng là phải hiểu rõ cả cơ chế tích hợp và chi phí liên quan. Các công cụ như Estuary Flow vượt trội nhờ sử dụng đồng bộ hóa dữ liệu theo thời gian thực thay vì xử lý hàng loạt truyền thống, giúp tránh tình trạng chậm trễ và nhu cầu quét cơ sở dữ liệu nhiều lần. Phương pháp này không chỉ hiệu quả hơn mà còn tiết kiệm chi phí, đặc biệt là cho nhu cầu sao chép liên tục.

Trước khi đưa ra quyết định, hãy đảm bảo rằng công cụ di chuyển bạn chọn phù hợp với các yêu cầu cụ thể và nằm trong phạm vi ngân sách của bạn.

6. Phần Kết luận

Di chuyển và sao chép cơ sở dữ liệu là một khía cạnh quan trọng trong chiến lược và kiến trúc dữ liệu của mọi doanh nghiệp. Khi các yêu cầu và trường hợp sử dụng thay đổi, việc sở hữu đúng loại cơ sở dữ liệu trở nên quan trọng hơn bao giờ hết.

MySQL và PostgreSQL đều có vị trí và trường hợp sử dụng tối ưu riêng. Cho dù mục tiêu của bạn là chuyển đổi vĩnh viễn từ cái này sang cái kia hay duy trì cả hai song song, chúng tôi hy vọng hướng dẫn này sẽ giúp bạn đơn giản hóa quy trình.