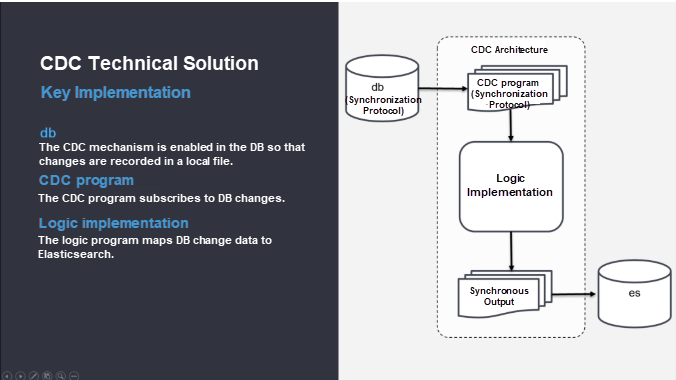
Phát hiện sự thay đổi dữ liệu của DB bằng cơ chế Change Data Capture:

*Dữ liệu thay đổi được ghi lại từ DB, được đẩy sang một chương trình trung gian và được đẩy đồng bộ đến Elaticsearch bằng cách sử dụng logic của chương trình trung gian. Dựa trên cơ chế CDC, dữ liệu chính xác được trả về ở tốc độ cực nhanh để đáp ứng với các truy vấn. Giải pháp này ít ảnh hưởng đến luồng chạy chính của các ứng dụng do đó, nó có thể được trừu tượng hóa và tách biệt với các hệ thống, làm cho nó phù hợp cho việc sử dụng quy mô lớn. Điều này được minh họa trong hình sau.*

**

**Case Study: Synchronize Data from MySQL to Elasticsearch**

Đồng bộ dữ liệu sử dụng cơ chế Binlog Mysql

Binlog là một cơ chế chức năng được cung cấp bởi MySQL. BinLog là một tệp nhị phân trên đĩa, chứa tất cả các sự kiện thay đổi nội dung hoặc cấu trúc của cơ sở dữ liệu MySQL, ví dụ: chèn, cập nhật, xóa. Tất cả các thay đổi sẽ được lưu thêm vào (append only) vào một file log được đánh số thứ tự. Mỗi thay đổi được đặc trưng bởi hai tham số: file log được ghi vào vị trí offset trên file log. MySql Binlog có ba loại:

* Statement Based: lưu trữ các câu lệnh làm thay đổi dữ liệu
* Row based: lưu trữ các dữ liệu thay đổi
* Mixed: lưu trữ cả statement và data thay đổi của từng row

Mysql hỗ trợ cơ chế replicate binary log. MySql sẽ streamming tất cả các thay đổi của database từ master tới slave. Toàn bộ binary log sẽ được replicate realtime tới slave.

Nó ban đầu được thiết kế để đồng bộ hóa giữa cơ sở dữ liệu chính và phụ:

* Cơ sở dữ liệu chính ghi dữ liệu vào tệp binlog khi cơ chế BinLog được bật.
* Cơ sở dữ liệu phụ lấy dữ liệu binlog từ cơ sở dữ liệu chính, phát lại dữ liệu binlog và cập nhật dữ liệu trong cơ sở dữ liệu phụ.
* Khi bạn bật Binlog, lưu ý rằng:

Một cơ sở dữ liệu chính thường được liên kết bởi nhiều cơ sở dữ liệu phụ và phải cập nhật các thay đổi thời gian thực cho hệ thống, dẫn đến tắc nghẽn tài nguyên máy chủ.

Các bảng được đồng bộ hóa phải có các khóa chính.

