

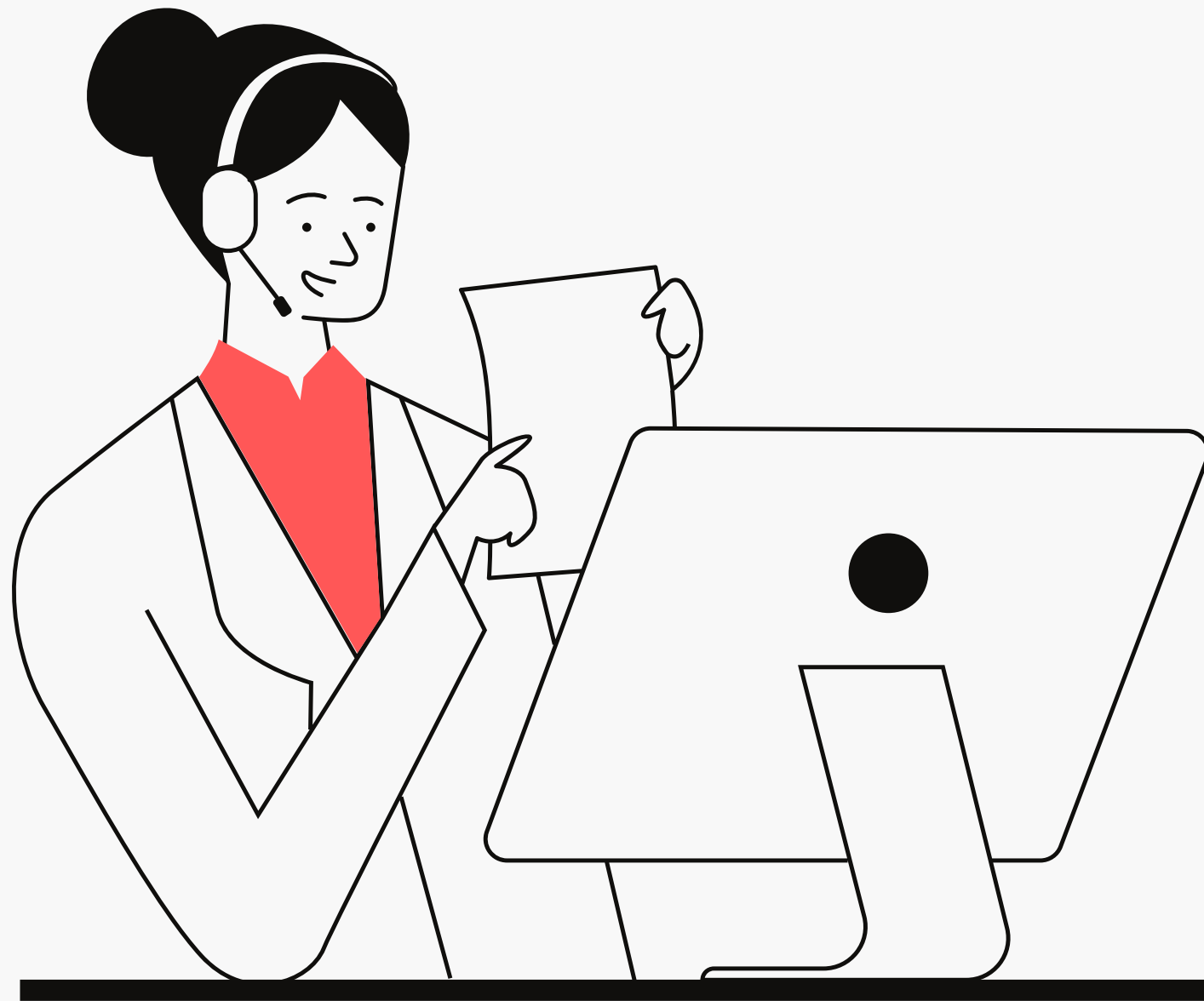
Monitoring ETL process

19HTTT1.BI01

01



Thành viên nhóm



Trần Ngọc Lam - 19127040

Nguyễn Phương Vy - 19127088

Nguyễn Kim Thị Tố Nga - 19127219

Nội dung

01

Tóm tắt

02

**Các ý tưởng
monitoring**

03

**Kết hợp
monitoring với AI**

1. Tóm tắt

ETL là một qui trình phức tạp

phải xử lý các định dạng dữ liệu đầu vào khác nhau, khối lượng dữ liệu khác nhau,...

Nêu vấn đề

Khi ETL xảy ra sai sót

- Dữ liệu không được tải vào cơ sở dữ liệu
- Dữ liệu được tải vào không chính xác
- Dữ liệu được sử dụng để đưa ra report hoặc quyết định không chính xác

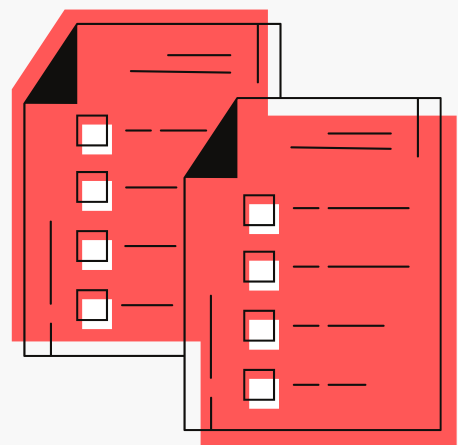


Logging: giúp khắc phục sự cố hoặc hiểu rõ hơn về các quy trình đang chạy.

Text file log provider, SQL Server Profiler log provider, Window Event log provider, XML file log provider

Monitoring

2 cách monitoring quá trình ETL trong SQL Server



Build - in reporting: được tích hợp vào SSIS trong SQL Server.

Audit được tích hợp vào SSIS sẽ nằm trên máy chủ SSIS mà SSIS đang chạy, do đó nếu ta có 10 máy chủ khác nhau, ta phải xem report của từng máy chủ một.

2. Các ý tưởng monitoring

Thông thường việc giám sát quy trình ETL bao gồm: kiểm tra luồng dữ liệu giữa 2 cơ sở dữ liệu hoặc 2 bảng, và thực hiện kiểm tra tính chính xác của dữ liệu. Có thể được thực hiện theo nhiều cách.

Giám sát các tác vụ vận chuyển dữ liệu bên ngoài

Mục tiêu bao gồm:

- Thời gian thực hiện vận chuyển dữ liệu
- Cách hoạt động và hiệu quả của các tác vụ nguồn

Giám sát các tác vụ ETL

Đo thời gian thực hiện và việc tác vụ có chạy hay không ở một thời điểm cụ thể.

Kiểm tra sự bất thường trong dữ liệu

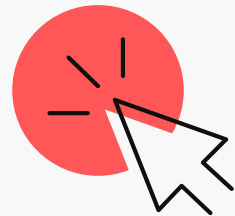
Thông thường, những thay đổi lớn trong các báo cáo không có lời giải thích cho thấy có thể đã xảy ra sự cố với dữ liệu.

Giám sát các thông số của dữ liệu

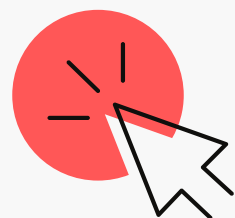
Thời gian phản hồi của cơ sở dữ liệu, khả năng kết nối,... → những chỉ số quan trọng và cho ta cái nhìn toàn diện của vấn đề.

Chủ động giám sát

Dự đoán các vấn đề xảy ra trong tương lai bằng cách dành thời gian quan sát, phân tích kiến trúc và mô hình, cũng như xác định được các dấu hiệu tiềm ẩn như sau:



Khối lượng dữ liệu được xử lý bởi một công việc cụ thể đang tăng lên theo thời gian.



Một công việc luôn mất nhiều thời gian để hoàn thành hơn so với trước đây.



Một cơ sở dữ liệu đang phát triển lớn dần gây ra sự mất tương xứng với các thành phần còn lại của quá trình.

3. Kết hợp monitoring với AI

01 Dữ liệu xấu do lỗi quy trình xảy ra quá nhanh

02 Kết quả của quy trình ETL bị hỏng không khác với dữ liệu chính xác.

03 Không dễ theo dõi nguyên nhân gây lỗi

**Những vấn đề
có thể gặp phải
khi con người
giám sát quá
trình ETL**

Lợi ích khi kết hợp monitoring với AI

- Liên tục theo dõi dữ liệu đi vào một cách tự động.
- Cảnh báo cho các bộ phận cụ thể khi phát hiện bất thường
- Thực hiện giám sát tại chỗ, dữ liệu có thể trở nên chính xác hơn.

01

Phát hiện và cảnh báo thông qua các chỉ số của ETL

Cung cấp khả năng phát hiện bất thường theo thời gian thực và cảnh báo trực tiếp



02

Xác định các tắc nghẽn cụ thể

Cho biết hệ thống đang chậm lại ở đâu và cung cấp cho câu trả lời ngay lập tức,

03

Định lượng các tác động của thay đổi trong việc quản lý dữ liệu

Hệ thống được sửa đổi và nâng cấp có thể ảnh hưởng đến quy trình tạo ra và giải thích dữ liệu.

04

Giảm chi phí vận hành

Thời gian dùng để tìm ra và giải quyết vấn đề có thể gây ra tổn thất lớn trong tiền tệ cũng như thời gian và nhân lực.

4. Demo monitoring ETL

Thanks for listening

