

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Xây dựng hệ thống xử lý dữ liệu nghiệp vụ cho vay
thấu chi phục vụ phát triển báo cáo tuân thủ

MAI THỊ NGỌC

ngoc.mt162947@sis.hust.edu.vn

Ngành Công nghệ thông tin và truyền thông

Giảng viên hướng dẫn: TS. Nguyễn Đức Toàn

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 08/2024

LỜI CAM KẾT

Họ và tên sinh viên: Mai Thị Ngọc
Điện thoại liên lạc: 0358842976
Email: ngoc.mt162947@sis.hust.edu.vn
Lớp: CNTT1.02 - K61
Hệ đào tạo: Cử nhân Kỹ thuật máy tính

Tôi – *Mai Thị Ngọc* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *Tiến sĩ Nguyễn Đức Toàn*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày 20 tháng 08 năm 2024

Tác giả ĐATN

Họ và tên sinh viên

LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời cảm ơn chân thành nhất tới giảng viên hướng dẫn đồ án - TS. Nguyễn Đức Toàn đã thấu hiểu, giao đề tài đúng với chuyên môn, năng lực của em. Nhờ có sự kiên nhẫn và tinh thần hỗ trợ của thầy đã giúp em rất nhiều trong những lúc em gặp khó khăn. Sự hướng dẫn, góp ý tận tâm đó chính là nguồn động viên lớn, giúp em giữ vững tinh thần và tiếp tục phấn đấu trong việc phát triển và hoàn thiện đề tài.

Em cũng xin được gửi lời chân thành cảm ơn tới các thầy cô công tác tại Đại học Bách khoa Hà Nội nói chung và Trường Công nghệ Thông tin và Truyền thông nói riêng, những người đã truyền đạt kiến thức, chỉ bảo tận tình trong suốt quá trình học tập và đã tạo cho em một cơ hội quý giá để có thể khắc phục sai lầm của mình.

Cuối cùng, em xin cảm ơn tới gia đình, bạn bè, những người đã ủng hộ, khích lệ tinh thần cũng như cổ vũ tạo động lực cho em để em có thể hoàn thành nhiệm vụ học tập. Những cuộc trò chuyện, những lời khuyên chân thành và sự quan tâm của mọi người đã giúp em vững tin hơn vào bản thân. Em thật sự biết ơn vì xung quanh luôn có những người yêu thương, đồng hành cùng em trong mỗi bước đường.

Trong quá trình làm đồ án không thể tránh khỏi những thiếu sót, em rất mong nhận được những ý kiến góp ý của thầy cô để đồ án của em được hoàn thiện hơn.

Một lần nữa, em xin chân thành cảm ơn!

TÓM TẮT NỘI DUNG ĐỒ ÁN

Trong bối cảnh hoạt động ngân hàng hiện đại, việc quản lý và phân tích dữ liệu ngày càng trở nên quan trọng. Dữ liệu ngân hàng nói chung và dữ liệu nghiệp vụ cho vay nói riêng không chỉ đơn thuần là thông tin về tài khoản và giao dịch, mà còn bao gồm dữ liệu về khách hàng, tín dụng, quản lý rủi ro, báo cáo tuân thủ, và nhiều loại dữ liệu phức tạp khác. Khối lượng và sự đa dạng của dữ liệu này đặt ra những thách thức lớn đối với các hệ thống quản lý dữ liệu truyền thống như Database, thường là hệ thống OLTP, được thiết kế để xử lý các giao dịch theo thời gian thực, lưu trữ và truy vấn dữ liệu với hiệu năng cao. Tuy nhiên chúng có những hạn chế nhất định khi đối mặt với sự phức tạp dữ liệu ngân hàng như là (i) khả năng mở rộng hạn chế, (ii) khả năng phân tích hạn chế, (iii) Khó xử lý dữ liệu đa dạng, (iv) khó tích hợp dữ liệu...

Việc tuân thủ các quy định pháp lý của các ngân hàng là cấp thiết và đòi hỏi sự chính xác, kịp thời. Các ngân hàng cần đảm bảo rằng họ không chỉ tuân thủ các quy định hiện hành mà còn phải chuẩn bị sẵn sàng để đối phó với những thay đổi liên tục trong môi trường pháp lý. Có một số công cụ được sử dụng để xử lý và phân tích các báo cáo tuân thủ như Microsoft Excel, Tableau, Microsoft Power BI,...các công cụ này có ưu điểm là dễ sử dụng, phân tích được các dữ liệu đơn giản tuy nhiên chúng đều gặp phải các hạn chế trong khả năng xử lý dữ liệu lớn và phức tạp, không hỗ trợ tốt cho việc tự động hóa và tích hợp dữ liệu từ nhiều nguồn

Qua việc tìm hiểu nhu cầu sử dụng và những hạn chế xử lý dữ liệu hiện tại, em đã thực hiện đề tài "**Xây dựng hệ thống xử lý dữ liệu nghiệp vụ cho vay thấu chi phục vụ phát triển báo cáo tuân thủ**" làm đồ án tốt nghiệp của mình. Hệ thống giúp tích hợp nhiều nguồn dữ liệu, xử lý được lượng dữ liệu lớn và phức tạp, có khả năng phân tích dữ liệu sâu rộng, đáp ứng các yêu cầu báo cáo tuân thủ và có khả năng mở rộng linh hoạt.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	2
1.3 Định hướng giải pháp.....	2
1.4 Bố cục đồ án	2
CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU.....	4
2.1 Khảo sát hiện trạng	4
2.2 Khảo sát nghiệp vụ cho vay thấu chi.....	5
2.3 Quy trình triển khai	7
2.4 Yêu cầu phi chức năng	7
2.4.1 Tính sử dụng.....	7
2.4.2 Hiệu suất.....	7
2.4.3 An toàn và bảo mật	7
2.4.4 Tính bảo trì.....	8
CHƯƠNG 3. CƠ SỞ LÝ THUYẾT	9
3.1 Tổng quan về Data Warehouse	9
3.1.1 Data Warehouse là gì.....	9
3.1.2 Kiến trúc của Data Warehouse	9
3.1.3 Mô hình dữ liệu Data Warehouse.....	10
3.1.4 Quy trình ETL	11
3.2 Tổng quan về IBM Data Banking Model.....	12
3.2.1 IBM InfoSphere DataStage and QualityStage.....	13
3.2.2 IBM Cognos Business Intelligence	14
3.2.3 Hệ quản trị cơ sở dữ liệu DB2	15

CHƯƠNG 4. THIẾT KẾ, TRIỂN KHAI VÀ ĐÁNH GIÁ HỆ THỐNG	16
4.1 Thiết kế kiến trúc.....	16
4.1.1 Kiến trúc giải pháp logic.....	16
4.1.2 Kiến trúc dữ liệu	17
4.1.3 Mô hình quan hệ thực thể.....	21
4.1.4 Luồng ETL tổng thể.....	22
4.2 Xây dựng luồng ETL chi tiết.....	22
4.2.1 Xây dựng job xử lý dữ liệu nguồn - Staging	22
4.2.2 Xây dựng job xử lý dữ liệu Staging - SoR.....	27
4.2.3 Xây dựng job xử lý dữ liệu SoR - INTF.....	31
4.2.4 Xây dựng job xử lý dữ liệu INTF - DIM	35
4.2.5 Xây dựng job xử lý dữ liệu lên Fact.....	39
4.3 Xây dựng báo cáo Cognos	43
4.3.1 Phân tích yêu cầu nghiệp vụ	43
4.3.2 Kết quả đạt được	44
4.4 Kiểm thử.....	44
CHƯƠNG 5. CÁC GIẢI PHÁP, ĐÓNG GÓP NỔI BẬT VÀ KẾT LUẬN..	46
5.1 Các giải pháp và đóng góp nổi bật.....	46
5.1.1 Xử lý lượng dữ liệu lớn	46
5.1.2 Phát triển các báo cáo tuân thủ tự động	48
5.1.3 Khả năng mở rộng với sản phẩm tín dụng khác	49
5.2 Kết luận	49
5.3 Hướng phát triển.....	50
TÀI LIỆU THAM KHẢO.....	51
PHỤ LỤC.....	53
A. Code SQL một số Job ETL	53

B. Code SQL báo cáo 014-CSTT	60
---	-----------

DANH MỤC HÌNH VẼ

Hình 3.1	Kiến trúc của Data Warehouse	10
Hình 3.2	Mô hình sao (Star schema)	11
Hình 4.1	Kiến trúc giải pháp logic của hệ thống	16
Hình 4.2	Kiến trúc dữ liệu của hệ thống	17
Hình 4.3	Mô hình quan hệ thực thể các DIM, FACT	21
Hình 4.4	Kiến trúc của Data Warehouse	22
Hình 4.5	Job Parallel xử lý dữ liệu từ file nguồn lên STAG	23
Hình 4.6	Cấu hình bảng nguồn từ file nguồn lên STAG	23
Hình 4.7	Cấu hình bảng đích từ file nguồn lên STAG	24
Hình 4.8	Cấu hình bảng đích từ file nguồn lên STAG	24
Hình 4.9	Cấu hình các tham số từ file nguồn lên STAG	25
Hình 4.10	Biên dịch job parallel từ file nguồn lên STAG	25
Hình 4.11	Job Sequence xử lý dữ liệu từ nguồn lên STAG	26
Hình 4.12	Cấu hình Job Parallel trong Job Sequence xử lý dữ liệu từ nguồn lên STAG	26
Hình 4.13	Biên dịch Job Sequence từ nguồn lên STAG	26
Hình 4.14	Job Parallel xử lý dữ liệu từ STAGING lên SOR	27
Hình 4.15	Cấu hình bảng nguồn từ STAGING-SOR	27
Hình 4.16	Cấu hình bảng đích từ STAGING-SOR	28
Hình 4.17	Cấu hình DDL các trường dữ liệu từ STAGING-SOR	29
Hình 4.18	Cấu hình các tham số từ STAGING - SOR	29
Hình 4.19	Biên dịch Job Parallel từ STAGING-SOR	30
Hình 4.20	Kết quả chạy job parallel từ STAGING - SOR	30
Hình 4.21	Job Sequence xử lý dữ liệu từ STAGING-SOR	30
Hình 4.22	Cấu hình Job Parallel trong Job Sequence xử lý dữ liệu từ STAGING-SOR	31
Hình 4.23	Biên dịch Job Sequence từ STAGING-SOR	31
Hình 4.24	Job Parallel xử lý dữ liệu từ SOR lên INTF	32
Hình 4.25	Cấu hình bảng nguồn từ SOR-INTF	32
Hình 4.26	Cấu hình bảng đích từ SOR-INTF	33
Hình 4.27	Cấu hình DDL các trường dữ liệu từ SOR-INTF	33
Hình 4.28	Cấu hình các tham số từ SOR-INTF	34
Hình 4.29	Biên dịch job parallel từ SOR-INTF	34
Hình 4.30	Kết quả chạy job parallel từ SOR-INTF	34

Hình 4.31	Job Sequence xử lý dữ liệu từ SOR-INTF	35
Hình 4.32	Cấu hình Job Parallel trong Job Sequence xử lý dữ liệu từ SOR-INTF	35
Hình 4.33	Job Parallel xử lý dữ liệu từ INTF - DIM	36
Hình 4.34	Cấu hình bảng nguồn từ INTF - DIM	36
Hình 4.35	Cấu hình bảng nguồn từ INTF - DIM	37
Hình 4.36	Cấu hình bảng đích từ INTF - DIM	37
Hình 4.37	Cấu hình DDL các trường dữ liệu từ INTF - DIM	38
Hình 4.38	Cấu hình các tham số từ INTF - DIM	38
Hình 4.39	Job Sequence xử lý dữ liệu từ INTF - DIM	39
Hình 4.40	Cấu hình Job Parallel trong Job Sequence xử lý dữ liệu từ INTF - DIM	39
Hình 4.41	Job Parallel xử lý dữ liệu từ INTF, DIM - FACT	40
Hình 4.42	Cấu hình bảng nguồn từ INTF, DIM - FACT	40
Hình 4.43	Cấu hình bảng đích từ INTF, DIM - FACT	41
Hình 4.44	Cấu hình DDL các trường dữ liệu từ INTF, DIM - FACT	41
Hình 4.45	Cấu hình các tham số từ INTF, DIM - FACT	42
Hình 4.46	Job Sequence xử lý dữ liệu từ INTF, DIM - FACT	42
Hình 4.47	Cấu hình Job Parallel trong Job Sequence xử lý dữ liệu từ INTF, DIM - FACT	43
Hình 4.48	Mẫu biểu báo cáo 014-CSTT	43
Hình 4.49	Báo cáo dư nợ cho vay phục vụ nhu cầu đời sống và dư nợ tín dụng	44

DANH MỤC BẢNG BIỂU

Bảng 2.1	So sánh giữa cơ sở dữ liệu và kho dữ liệu	4
Bảng 2.2	Quy trình triển khai hệ thống xử lý dữ liệu	7
Bảng 3.1	Chín concept của IBM Banking Data Model	13
Bảng 5.1	Danh sách các job phát triển và số lượng bản ghi	48