

CSC12107 – Information Systems for Business Intelligence

Exercise Session 2: ETL

TA: Nguyen Ngoc Minh Chau

Context Description

SuperMart operates two supermarkets at different locations, each with its own system for managing staff, customers, products, and invoices. The store in the central area uses an MSSQL database to manage its data, while the second store relies on legacy Excel spreadsheets. The SuperMart manager wants a unified view of product sales across both stores.

As a data warehouse engineer, your first task is to perform the ETL process to extract data from these sources into the Dimensional Data Store (DDS). Diagrams illustrating the database schema of the first store, as well as the NDS and DDS schemas, are provided below.

Your task is to use the SSIS tool to implement the ETL process, adhering to the following business rules:

- ProductIDs, CustomerIDs, and StaffIDs are unique across the entire system.
- InvoiceIDs and InvoiceDetailIDs are unique only within each store, to optimize payment processing and reduce bottlenecks when generating these IDs.

Vietnamese version

SuperMart vận hành hai siêu thị tại các địa điểm khác nhau, mỗi địa điểm có hệ thống riêng để quản lý nhân viên, khách hàng, sản phẩm và hóa đơn. Cửa hàng ở khu vực trung tâm sử dụng cơ sở dữ liệu MSSQL để quản lý dữ liệu, trong khi cửa hàng thứ hai dựa vào bảng tính Excel. Quản lý của SuperMart muốn có một cái nhìn thống nhất về doanh số bán hàng của cả hai cửa hàng.

Là một kỹ sư kho dữ liệu, nhiệm vụ đầu tiên của bạn là thực hiện quy trình ETL để trích xuất dữ liệu từ các nguồn này vào Dimensional Data Store (DDS). Các sơ đồ minh họa sơ đồ cơ sở dữ liệu của cửa hàng đầu tiên, cũng như sơ đồ của NDS và DDS, được cung cấp dưới đây.

Nhiệm vụ của bạn là sử dụng công cụ SSIS để triển khai quy trình ETL, tuân theo các quy tắc:

- ProductIDs, CustomerIDs và StaffIDs là duy nhất trong toàn hệ thống.
- InvoiceIDs và InvoiceDetailIDs chỉ duy nhất trong từng cửa hàng, nhằm tối ưu hóa quá trình thanh toán và giảm thiểu tắc nghẽn khi tạo các ID này.

Rules:

- This is an individual assignment. You can discuss (in quite), but this is not allowed to copy from each other.
- Your **submission** includes:
 - **SSIS project and screenshot images of NDS and DDS data: Zip as <StudentID>.zip and submit to “Link nộp project” before 11:00 AM**
 - Video explain your project: Source > Stage > NDS > DDS (1)
 - Screenshot images of your NDS, DDS data (2)
 - Compress all the above video (1) and images (2) into file named **<StudentID>.zip** before **11:59 PM (23:59)**
 - If your file is too large, put the links to the materials into a text file (make sure to share it to me so that I can grade your exercise)

Lab Resources:

Resources	Description
Source_SuperMart_Central.sql	SQL script to create and populate data for the source Store 1 (Central Store).
Source_SuperMart_West.xlsx	Data for the source Store 2 (West Store)
NDS_SuperMart.sql	SQL script to create the NDS
DDS_SuperMart.sql	SQL script to create the DDS

Schema Diagrams

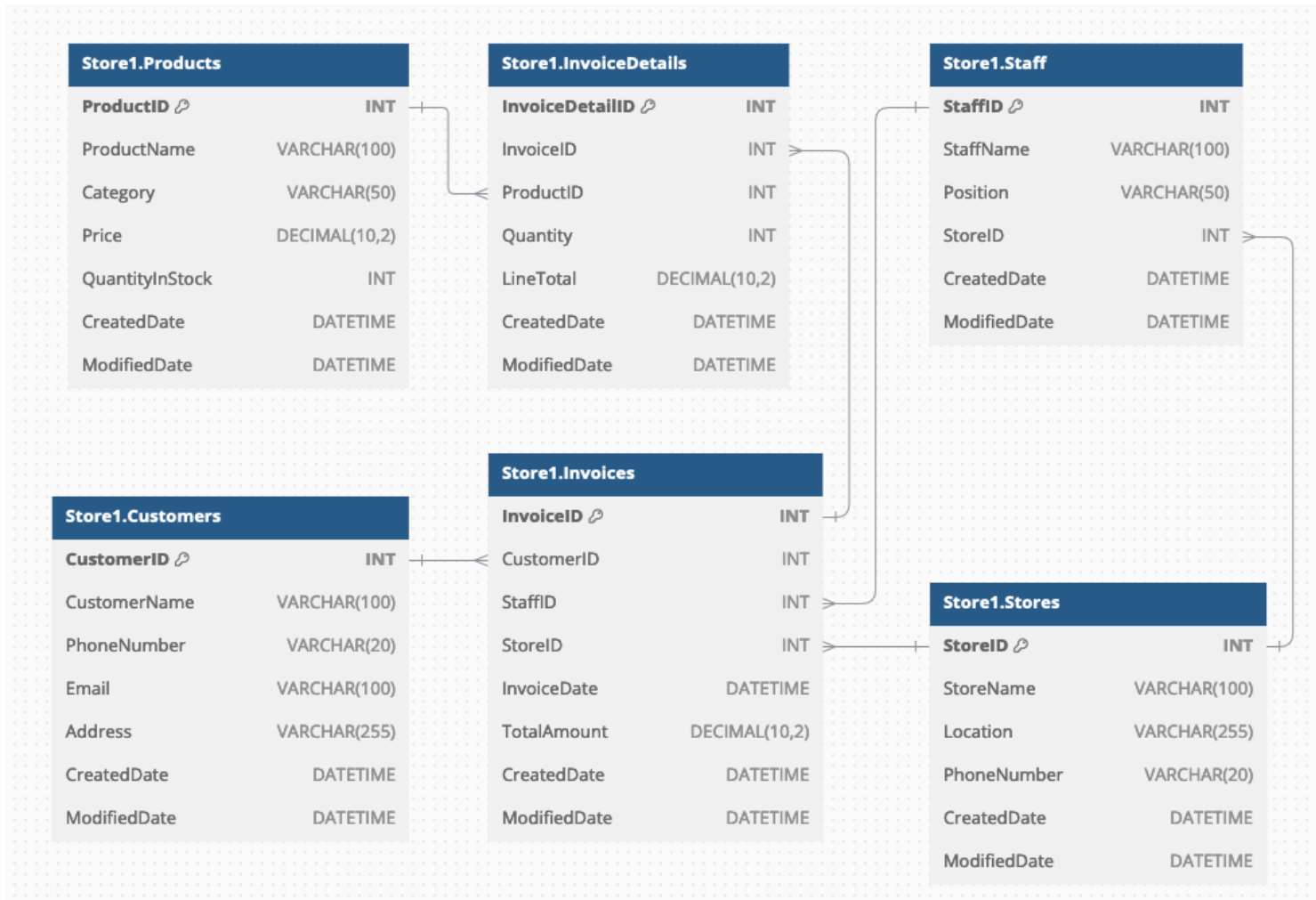


Fig1. Store1 (Central Store) schema diagram

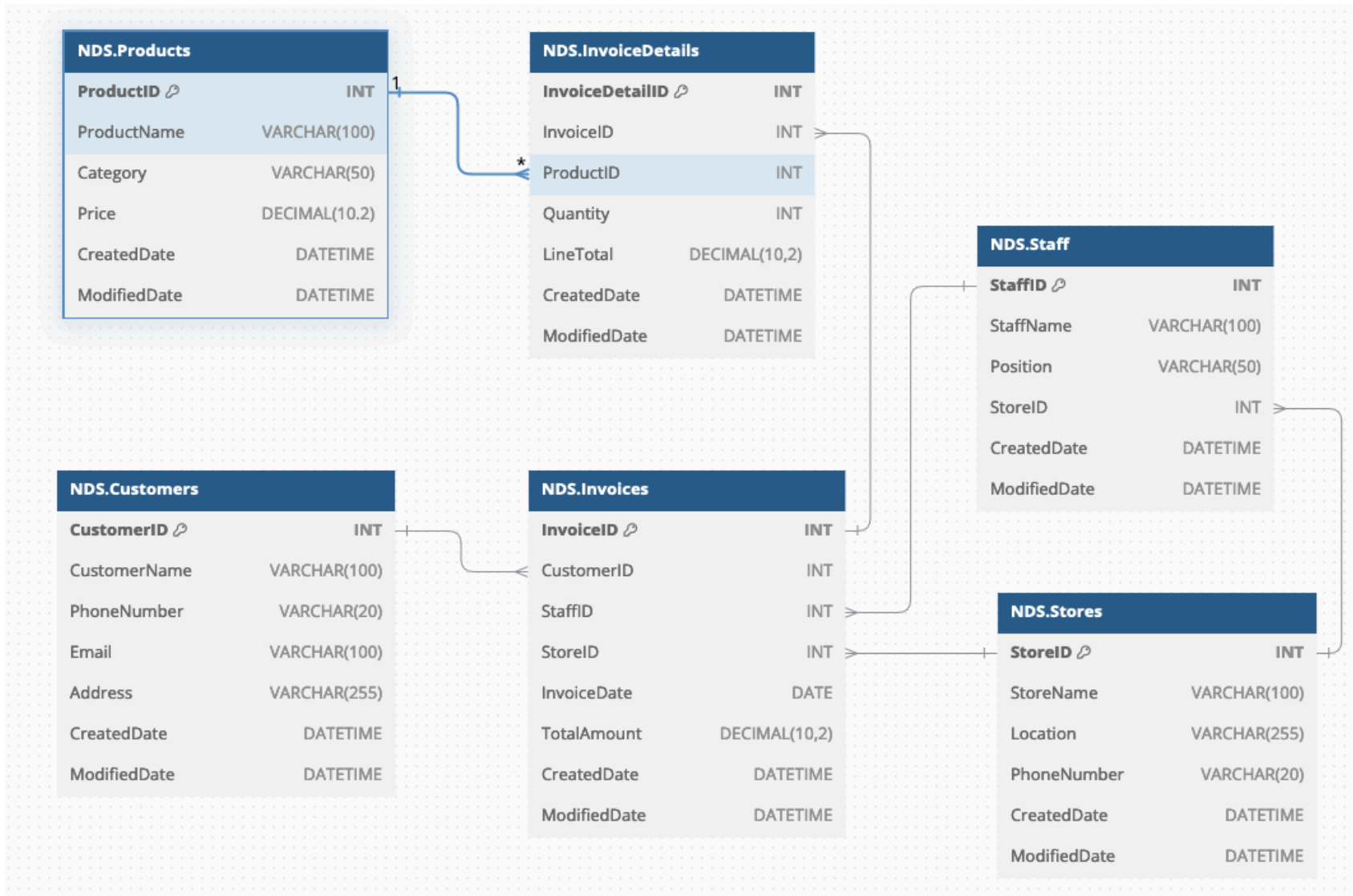


Fig 2. NDS schema diagram

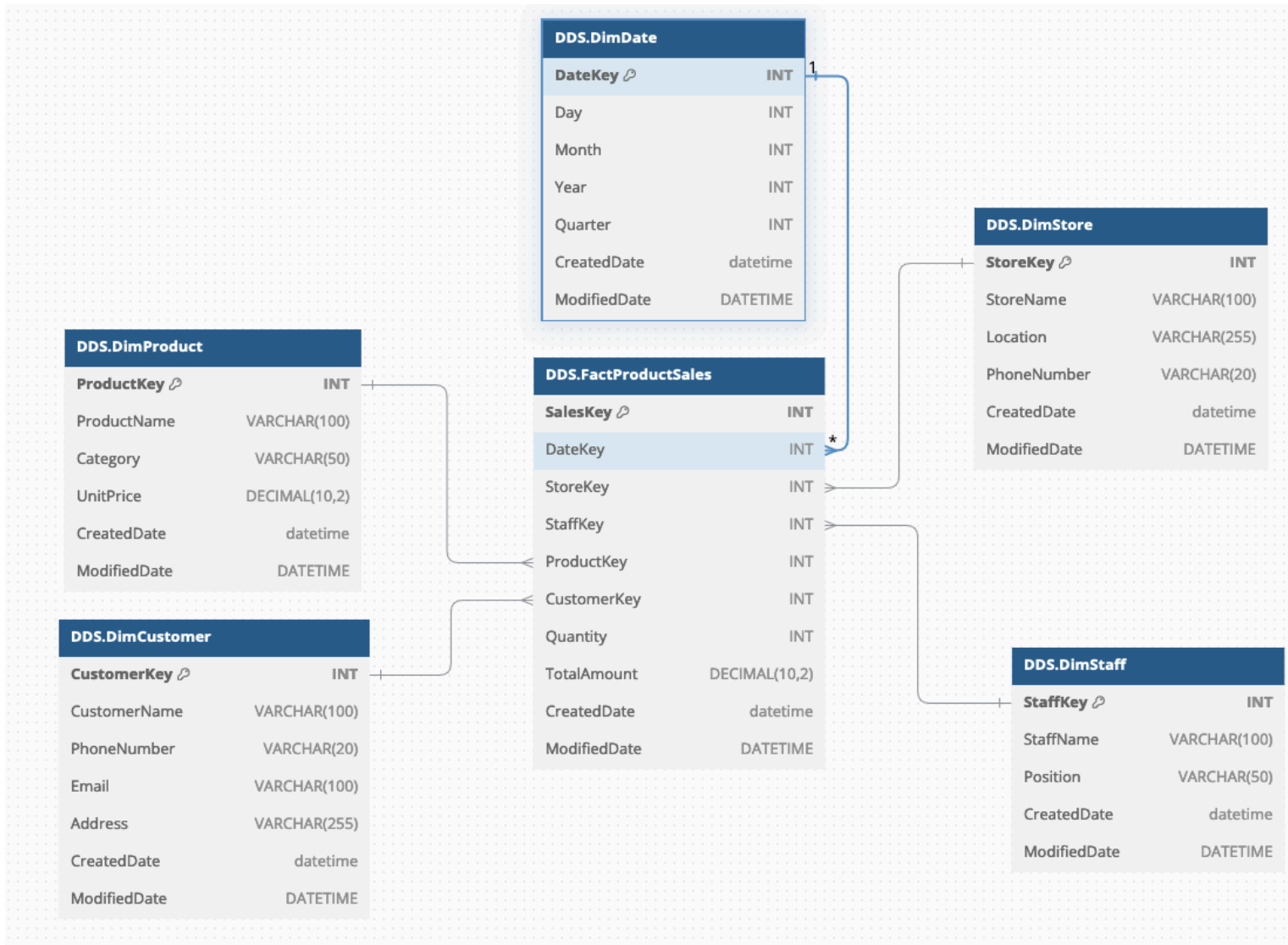


Fig 3: DDS schema diagram