

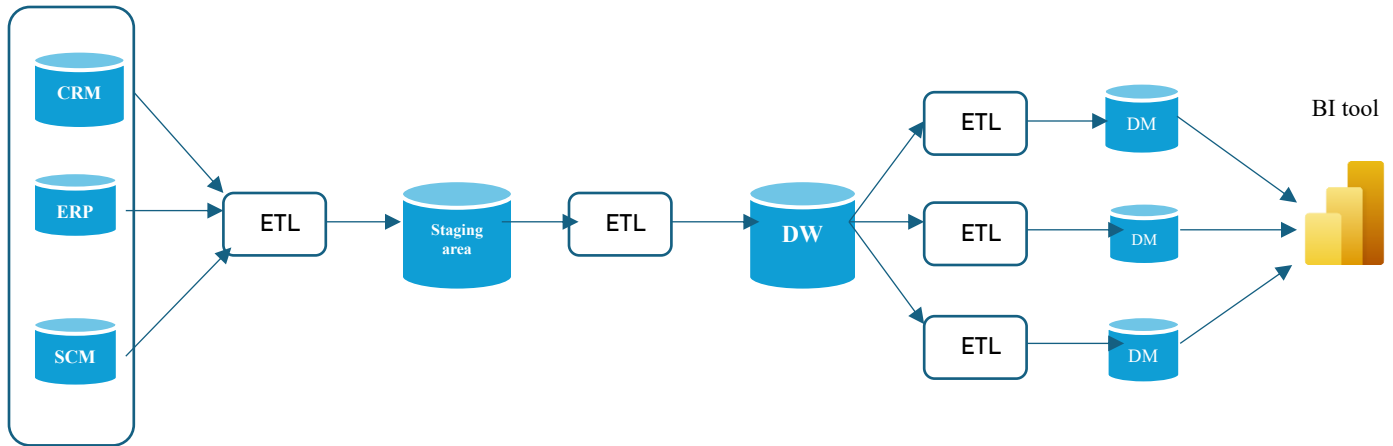
Mục lục

1. Kiến trúc hệ thống	1
2. Thiết kế dữ liệu & Data Dictionary	2
2.1. Cấu trúc bảng nguồn (Source).....	2
2.2. Cấu trúc bảng Staging.....	2
2.3. Cấu trúc bảng Data Warehouse (DWH).....	2
3. Quy trình ETL chi tiết	3
3.1. Chuẩn bị môi trường	3
3.2. Source → Staging.....	5
3.3. Staging → Data Warehouse	6
3.4. DWH → Data Mart	10
3.5. Xây dựng luồng tổng thể (Load Plan) & Lập lịch.....	11
3.6. Kết nối và xây dựng báo cáo BI với Power BI.....	13
3.7. Xử lý SCD Type 2 (Lưu vết thay đổi lịch sử).....	14

1. Kiến trúc hệ thống

Sơ đồ tổng thể

- **Nguồn dữ liệu:** Hệ thống bán hàng (CRM, ERP, SCM)
- **Staging Area:** nơi chứa bản sao tạm thời của dữ liệu gốc, được trích xuất từ nguồn sang để xử lý trước khi đẩy vào bảng đích
- **Data Warehouse (DWH):** Kho dữ liệu doanh nghiệp
- **Data Mart:** Vùng dữ liệu tính toán cho từng nghiệp vụ/phòng ban
- **BI Tool:** Power BI để tạo báo cáo



Hình 1: Sơ đồ ETL và các vùng dữ liệu

2. Thiết kế dữ liệu & Data Dictionary

2.1. Cấu trúc bảng nguồn (Source)

Tên bảng	Thuộc tính	Ý nghĩa
SRC_REGION	REGION_ID, COUNTRY, REGION	Vùng miền, quốc gia
SRC_CITY	CITY_ID, CITY, POPULATION, REGION_ID	Thành phố
SRC_CUSTOMER	CUST_ID, DEAR, LAST_NAME, FIRST_NAME, ADDRESS, PHONE, AGE, SALES_PERS_ID	Khách hàng
SRC_PRODUCT	PRODUCT_ID, PRODUCT, PRICE, FAMILY_NAME	Sản phẩm
SRC_ORDERS	ORDER_ID, ORDER_DATE, CUST_ID, CITY_ID	Đơn hàng
SRC_ORDER_LINES	ORDER_ID, PRODUCT_ID, QTY	Chi tiết đơn hàng
SRC_SALE_PERSON	SALES_PERS_ID, FIRST_NAME, LAST_NAME, HIRE_DATE	Nhân viên bán hàng

2.2. Cấu trúc bảng Staging

Các bảng giống hệt bảng nguồn

2.3. Cấu trúc bảng Data Warehouse (DWH)

Dimension Table (DIM)

- **DIM_DATE:** Ngày tháng chuẩn hóa (DATE_ID, DATE_KEY, MONTH, QUARTER, YEAR, DAY_OF_MONTH, DAY_OF_WEEK, MONTH_ID, DAY_NAME)
- **DIM_CUSTOMER:** Thông tin khách hàng (CUSTOMER_ID, DEAR, LAST_NAME, FIRST_NAME, ADDRESS, PHONE, AGE)
- **DIM_PRODUCT:** Thông tin sản phẩm (PRODUCT_ID, PRODUCT_NAME, PRICE, CATEGORY)
- **DIM_SALES_PERSON:** Nhân viên bán hàng (SALES_PERSON_ID, FIRST_NAME, LAST_NAME, HIRE_DATE)
- **DIM_CITY:** Thành phố/khu vực (CITY_ID, CITY_NAME, POPULATION, COUNTRY_NAME, REGION_NAME)

Fact Table (FACT)

- **FACT_ORDERS:** Thông tin giao dịch đơn hàng (ORDER_DATE, CITY_ID, CUST_ID, ORDER_ID, SALES_PERSON_ID, PRODUCT_ID, QUANTITY, UNIT_PRICE, GROSS_SALES)

2.4. Cấu trúc bảng Data Mart

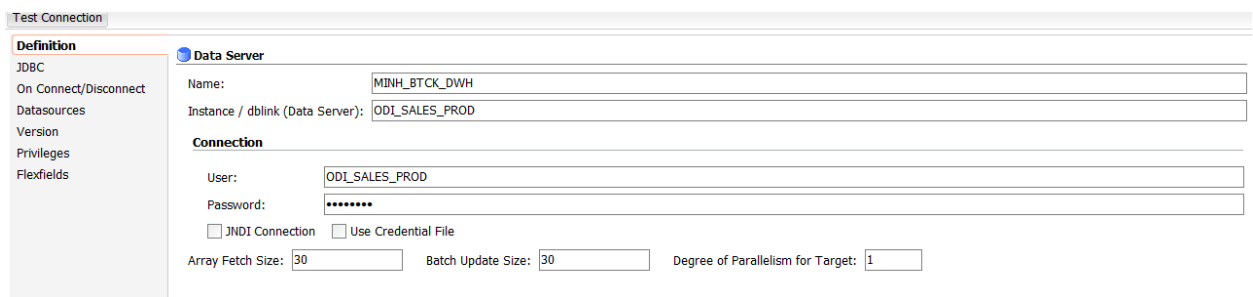
- **RPT_SALE_MONTH:** Báo cáo doanh thu & số lượng theo tháng, từng sản phẩm
- **RPT_REGION_MONTH:** Báo cáo doanh thu & số lượng theo tháng, từng vùng

3. Quy trình ETL chi tiết

3.1. Chuẩn bị môi trường

- Cài đặt Oracle Database, Oracle ODI Studio, Power BI Desktop
- Cấu hình kết nối DB (Host, Port, Service Name, User) trong ODI với các hệ thống bán hàng, Staging area, Data Warehouse, Data Mart trong topology bao gồm Physical Architecture, Contexts, Logical Architecture

Bước 1: Cấu hình kết nối với Data Server



The screenshot shows the 'Test Connection' dialog in Oracle ODI Studio. The 'Definition' tab is selected. The 'Data Server' section contains the following information:

- Name:** MINH_BTCK_DWH
- Instance / dblink (Data Server):** ODI_SALES_PROD
- User:** ODI_SALES_PROD
- Password:** (masked with dots)
- ☐ JNDI Connection
- ☐ Use Credential File
- Array Fetch Size:** 30
- Batch Update Size:** 30
- Degree of Parallelism for Target:** 1

Bước 2: Cấu hình JDBC

Test Connection

Definition

JDBC

On Connect/Disconnect
Datatypes
Version
Privileges
Flexfields

JDBC Driver: oracle.jdbc.OracleDriver

JDBC URL: jdbc:oracle:thin:@192.168.1.200:1525/orclpdb

Properties

Key	Value

Bước 3: Chọn Schema và Work Schema

Definition

Physical Schema [Data Server: MINH_BTCK_DWH]

Context

Version

Privileges

Flexfields

Name: MINH_BTCK_DWH.ODI_SALES_PROD

Schema (Schema): ODI_SALES_PROD

Schema (Work Schema): ODI_SALES_PROD

☒ Default

Work Tables Prefix

Errors: E\$_ Loading: C\$_ Integration: I\$_ Temporary Indexes: DX\$_

Journalizing elements prefixes

Datastores: J\$_ Views: JV\$_ Triggers: T\$_

Naming Rules

Character Encoding: BASE38

Local Object Mask: %SCHEMA.%OBJECT

Remote Object Mask: %SCHEMA.%OBJECT@%DSERVER

Partition Mask: %SCHEMA.%OBJECT PARTITION(%PARTITION)

Sub-Partition Mask: %SCHEMA.%OBJECT SUBPARTITION(%PARTITION)

Local Sequence Mask: %SCHEMA.%OBJECT.nextval

Remote Sequence Mask: %SCHEMA.%OBJECT.nextval@%DSERVER

Local Sequence Current Value Mask: %SCHEMA.%OBJECT.currval

Remote Sequence Current Value Mask: %SCHEMA.%OBJECT.currval@%DSERVER

Bước 4: Tạo Logical Schema và Context

Definition

Logical Schema

Privileges

Flexfields

Name: MINH_BTCK_DWH

Context	Physical Schemas
Development	MINH_BTCK_DWH.ODI_SALES_PROD
Global	MINH_BTCK_DWH.ODI_SALES_PROD
Production	MINH_BTCK_DWH.ODI_SALES_PROD
Test	MINH_BTCK_DWH.ODI_SALES_PROD

Bước 5: Chuyển sang tab Designer vào phần model để Reverse- Engineering lấy schema của database

MINH_BTCK_ORDERS

Reverse Engineer | Check Model | Generate and Deploy Data Services...

Definition

Reverse Engineer

Selective Reverse-Engineering

Control

Journalizing

Journalized Tables

Services

Markers

Memo

Version

Privileges

Flexfields

Model [Model Folder: MINH]

Name: MINH_BTCK_ORDERS

Code: MINH_BTCK_ORDERS

Technology: Oracle

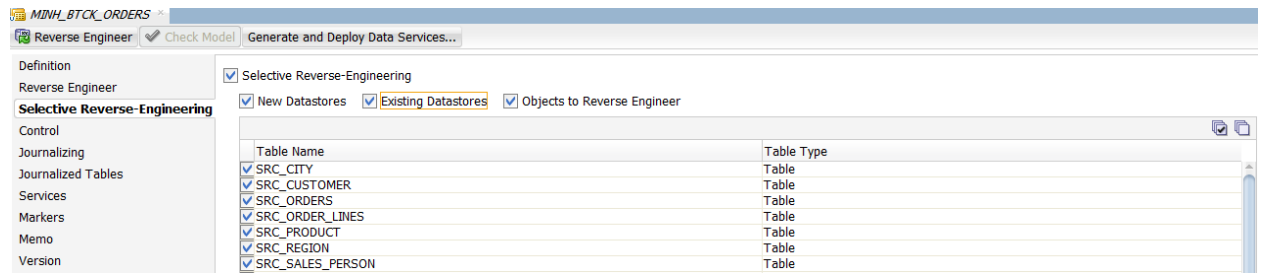
Logical Schema: MINH_BTCK_ORDERS

Action Group: <Generic Action>

Default Folder:

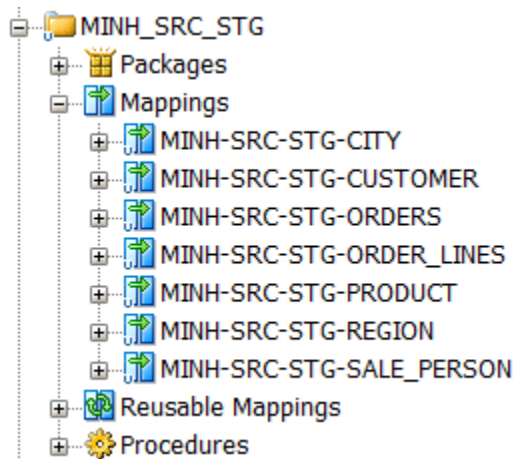
☐ Display the Metadata changes in the Model tree

Description:



3.2. Source → Staging

Tạo folder để mapping: 1:1 từ Source sang Staging (giữ nguyên dữ liệu, không biến đổi)



MINH-STG-ORDER_LINES

100%

Overview Logical Physical

MINH_STG_ORDER_LINES - Properties

Find

Attributes

Attributes:

Name	Data Type	Le...	Sc...	D...	Expression	Ex
ORDER_ID	NUMBER	10	0		SRC_ORDER_LINES.ORDER_ID	N.
LORDER_ID	NUMBER	10	0		SRC_ORDER_LINES.LORDER_ID	N.
PRODUCT_ID	NUMBER	10	0		SRC_ORDER_LINES.PRODUCT_ID	N.
QTY	NUMBER	10	0		SRC_ORDER_LINES.QTY	N.
AMOUNT	NUMBER	10	0		SRC_ORDER_LINES.AMOUNT	N.

General

Target

Integration Type: Control Append

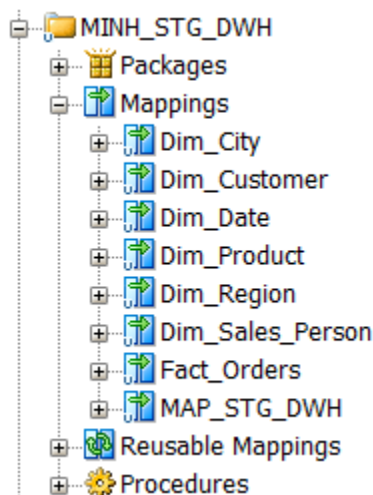
Update Key:

Maximum Number of Errors Allowed:

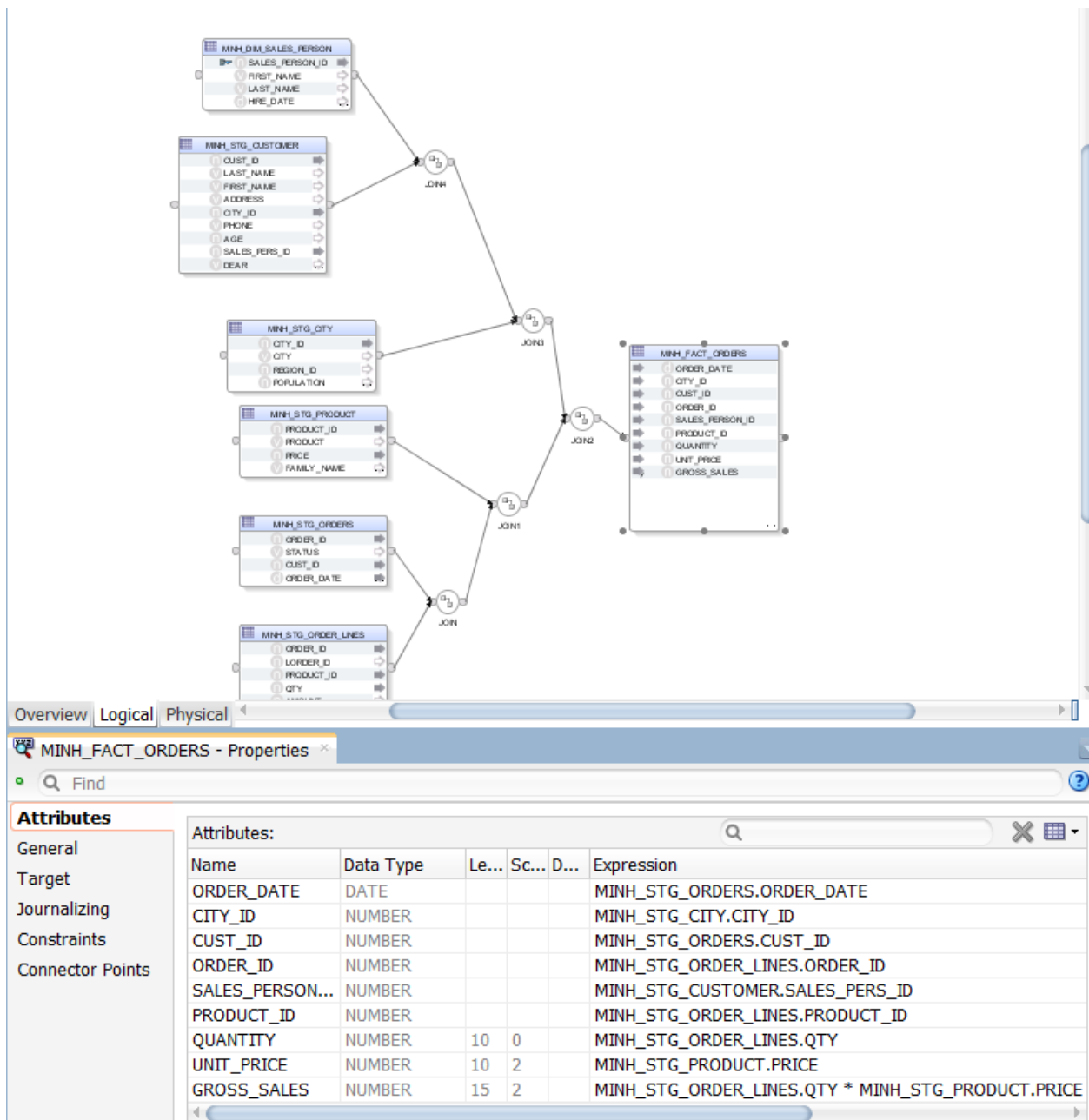
Integration Errors as Percentage: ☐

3.3. Staging → Data Warehouse

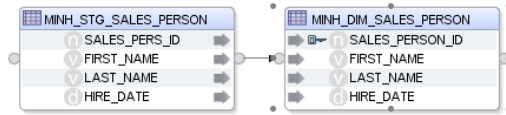
Tạo fodder để ETL dữ liệu từ Staging area sang Data Warehouse



Bảng Fact



Bảng Dim_Sales_Person



Overview Logical Physical

MINH_DIM_SALES_PERSON - Properties

Find

Attributes

General

Target

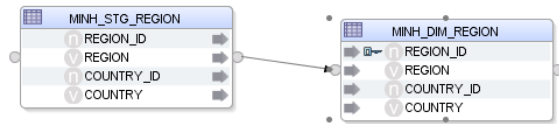
Journalizing

Constraints

Connector Points

Name	Data Type	Le...	Sc...	D...	Expression	Ex...	Fl...	Key	In...	U...	C...
SALES_PERSON...	NUMBER				MINH_STG_SALES_PERSON.SALES_PERS_ID	N...		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
FIRST_NAME	VARCHAR2	50			MINH_STG_SALES_PERSON.FIRST_NAME	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
LAST_NAME	VARCHAR2	50			MINH_STG_SALES_PERSON.LAST_NAME	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
HIRE_DATE	DATE				MINH_STG_SALES_PERSON.HIRE_DATE	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Bảng Dim_region



Overview Logical Physical

MINH_DIM_REGION - Properties

Find

Attributes

General

Target

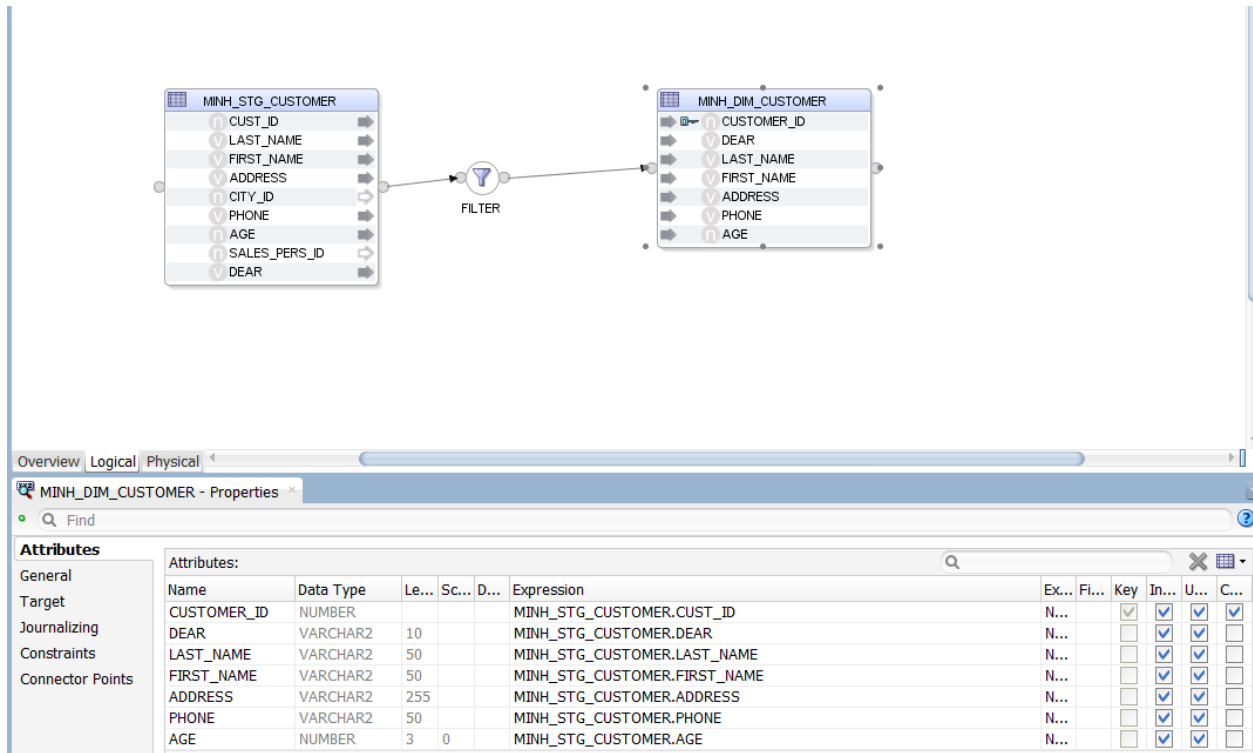
Journalizing

Constraints

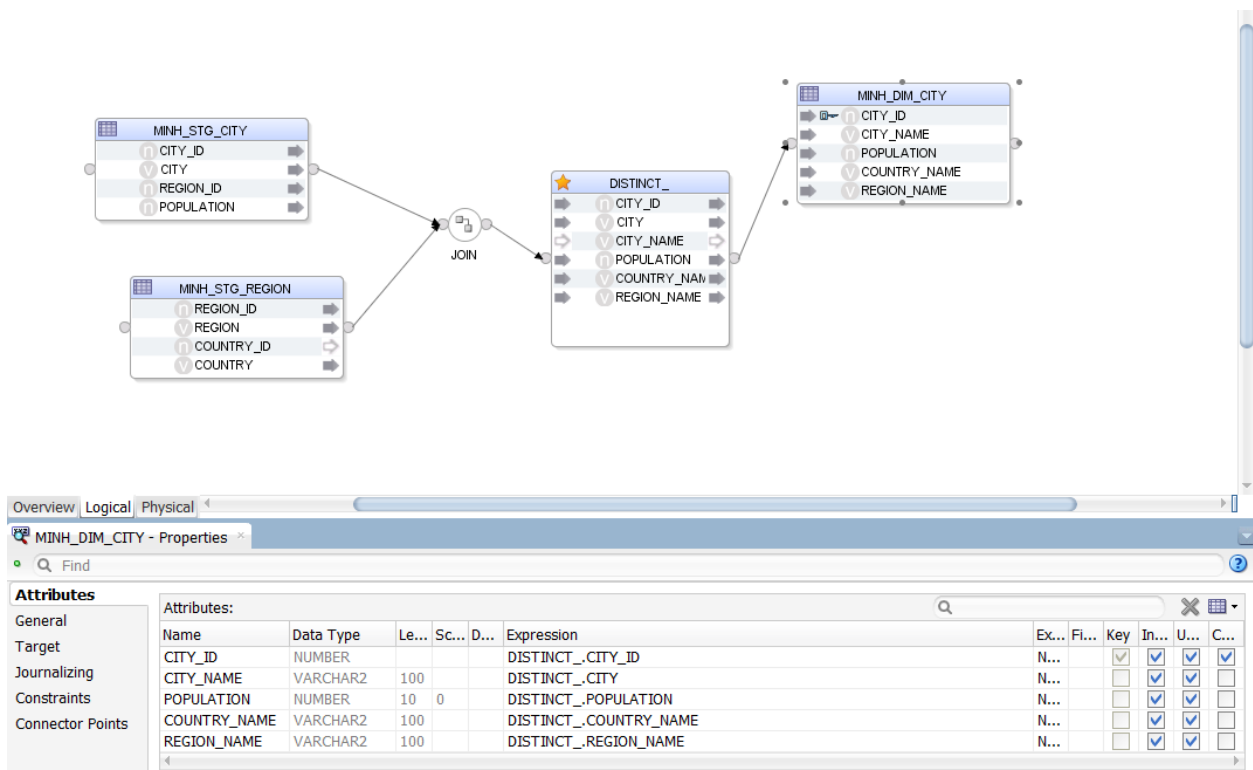
Connector Points

Name	Data Type	Le...	Sc...	D...	Expression	Ex...	Fl...	Key	In...	U...	C...
REGION_ID	NUMBER	10	0		MINH_STG_REGION.REGION_ID	N...		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
REGION	VARCHAR2	50			MINH_STG_REGION.REGION	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
COUNTRY_ID	NUMBER	11	0		MINH_STG_REGION.COUNTRY_ID	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
COUNTRY	VARCHAR2	50			MINH_STG_REGION.COUNTRY	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

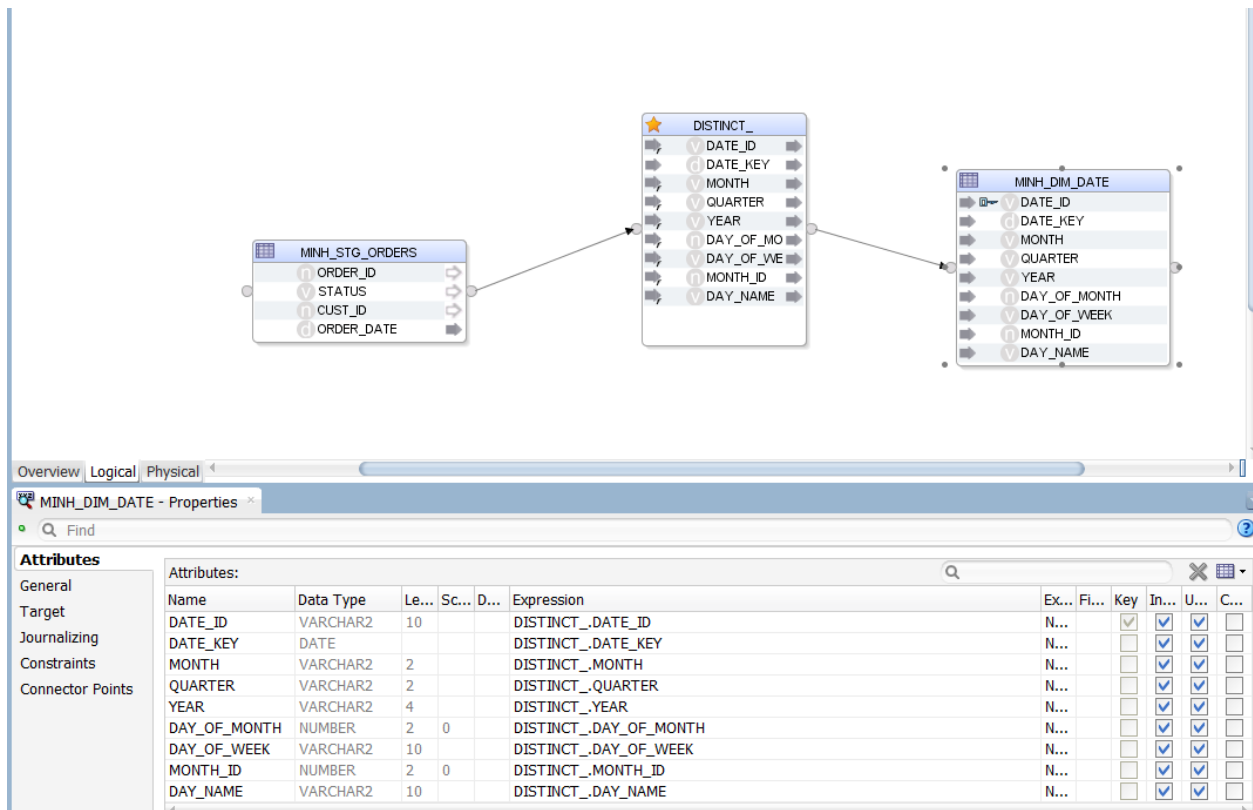
Bảng Dim_Customer



Bảng dim_city

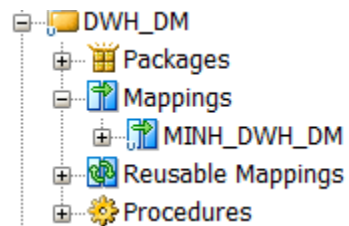


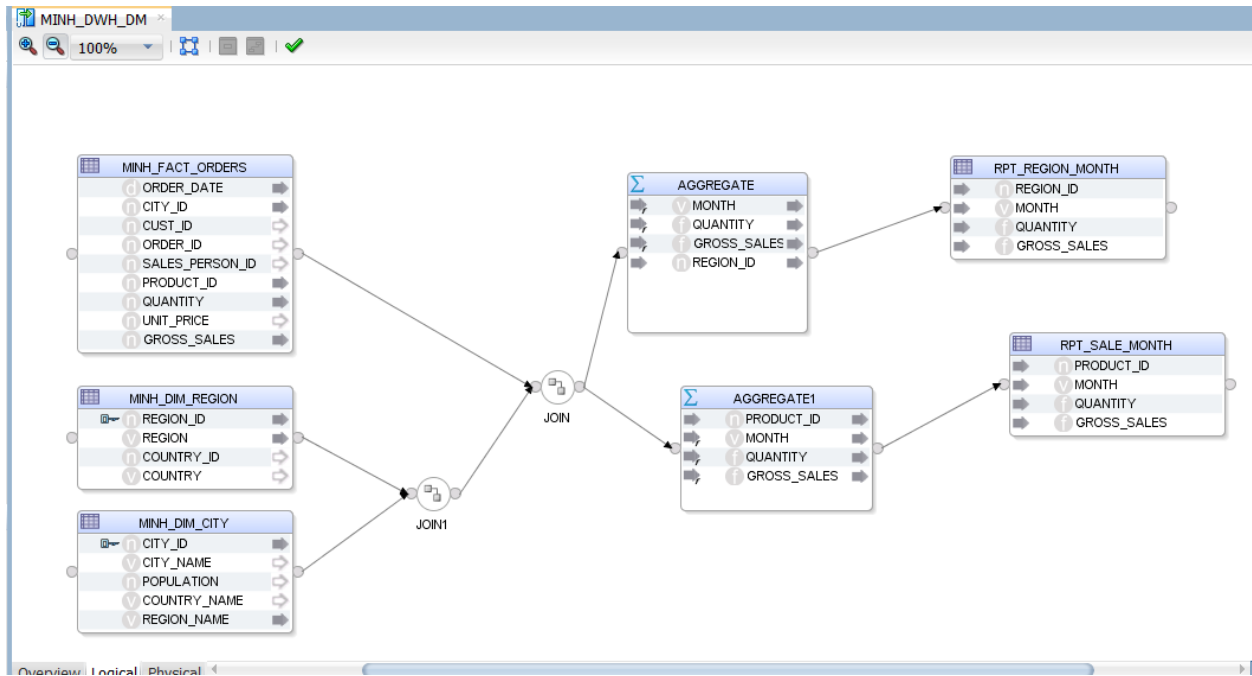
Bảng Dim_date



3.4. DWH → Data Mart

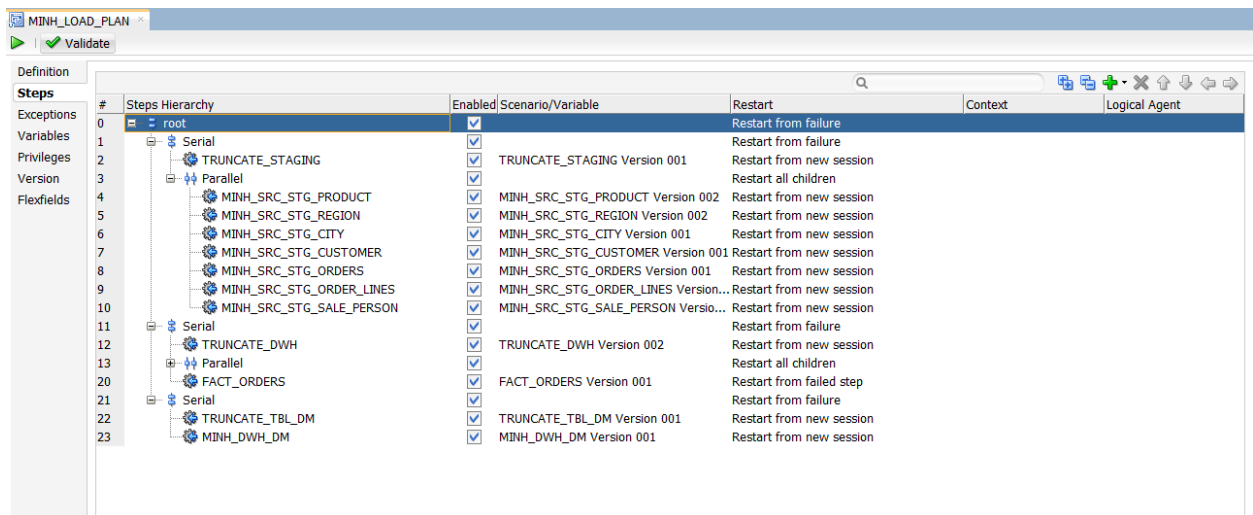
Tạo fodder để ETL dữ liệu từ Data Warehouse vào Data Mart





3.5. Xây dựng luồng tổng thể (Load Plan) & Lập lịch

- **Tạo Load Plan:** gom toàn bộ job ETL từ Source → Staging → DWH → Data Mart thành luồng tổng gồm các job chạy tuần tự và song song



- **Lập lịch chạy:** bằng ODI Agent

MINH_LOAD_PLAN x Load Plan Scheduling DEVELOPMENT / MINH_OracleDIAgent1

Definition

Execution Cycle
Variables
Privileges
Version

Scheduling [Load Plan: MINH_LOAD_PLAN]

Context: Logical Agent:

Log Level:

Status

☒ Active
☐ Inactive
☐ Active for the period:

☐ Starting: Date: Time:
☐ Ending: Date: Time:
☐ Every day between: from: to:
☐ Except these days of the month
☐ Except these days of the week: ☐ Monday ☐ Tuesday ☐ Wednesday ☐ Thursday
☐ Friday ☐ Saturday ☐ Sunday

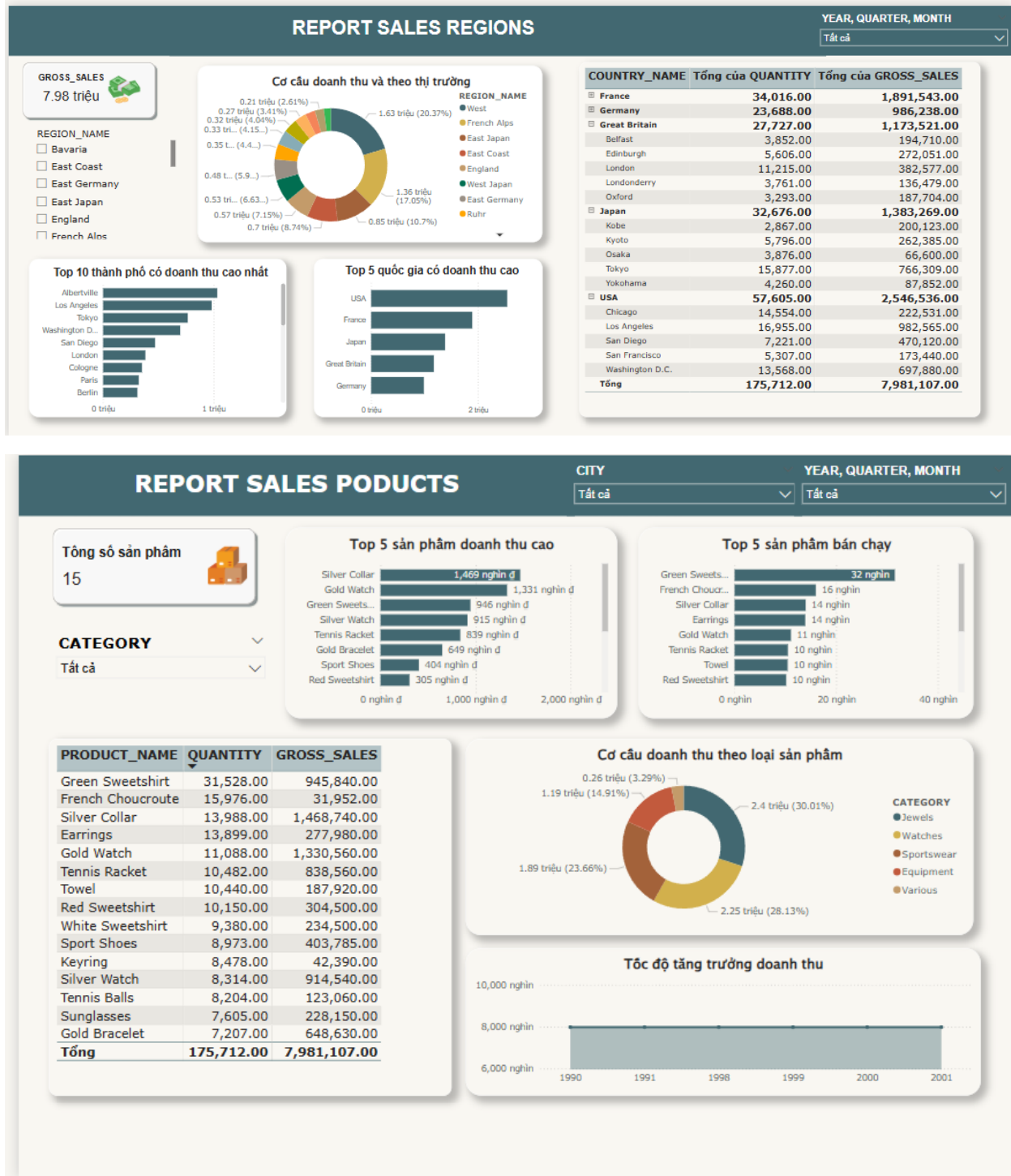
Execution

☐ On startup Date: Time:
☒ Simple
☐ Hourly
☐ Daily
☐ Weekly
☐ Monthly (day of the month)
☐ Monthly (week day)
☐ Yearly

- MINH_OracleDIAgent1 - 121
- 5553 - MINH_DWH_DM - Jun 15, 2025 11:04:56 AM
 - 5552 - TRUNCATE_TBL_DM - Jun 15, 2025 11:04:50 AM
 - 5551 - FACT_ORDERS - Jun 15, 2025 11:04:35 AM
 - 5550 - DIM_PRODUCT - Jun 15, 2025 11:04:21 AM
 - 5549 - DIM_REGION - Jun 15, 2025 11:04:19 AM
 - 5548 - DIM_CITY - Jun 15, 2025 11:04:17 AM
 - 5547 - DIM_SALES_PERSON - Jun 15, 2025 11:04:14 AM
 - 5546 - DIM_CUSTOMER - Jun 15, 2025 11:04:11 AM
 - 5545 - DIM_DATE - Jun 15, 2025 11:04:09 AM
 - 5544 - TRUNCATE_DWH - Jun 15, 2025 11:04:03 AM
 - 5543 - MINH_SRC_STG_CUSTOMER - Jun 15, 2025 11:03:55 AM
 - 5542 - MINH_SRC_STG_REGION - Jun 15, 2025 11:03:52 AM
 - 5541 - MINH_SRC_STG_CITY - Jun 15, 2025 11:03:50 AM
 - 5540 - MINH_SRC_STG_SALE_PERSON - Jun 15, 2025 11:03:48 AM
 - 5539 - MINH_SRC_STG_ORDER_LINES - Jun 15, 2025 11:03:46 AM
 - 5538 - MINH_SRC_STG_ORDERS - Jun 15, 2025 11:03:43 AM
 - 5537 - MINH_SRC_STG_PRODUCT - Jun 15, 2025 11:03:41 AM
 - 5536 - TRUNCATE_STAGING - Jun 15, 2025 11:03:35 AM
 - 5535 - Dim_City_Physical_SESS - Jun 15, 2025 11:01:14 AM
 - 5534 - Dim_Sales_Person_Physical_SESS - Jun 15, 2025 10:59:56 AM

3.6. Kết nối và xây dựng báo cáo BI với Power BI

Kết nối Oracle với Power BI để tạo báo cáo



REPORT SALES PRODUCTS

CITY: Tất cả YEAR, QUARTER, MONTH: Tất cả

Tổng số sản phẩm: 15

CATEGORY: Tất cả

Top 5 sản phẩm doanh thu cao

Top 5 sản phẩm bán chạy

PRODUCT_NAME	QUANTITY	GROSS_SALES
Green Sweetshirt	31,528.00	945,840.00
French Choucroute	15,976.00	31,952.00
Silver Collar	13,988.00	1,468,740.00
Earrings	13,899.00	277,980.00
Gold Watch	11,088.00	1,330,560.00
Tennis Racket	10,482.00	838,560.00
Towel	10,440.00	187,920.00
Red Sweetshirt	10,150.00	304,500.00
White Sweetshirt	9,380.00	234,500.00
Sport Shoes	8,973.00	403,785.00
Keyring	8,478.00	42,390.00
Silver Watch	8,314.00	914,540.00
Tennis Balls	8,204.00	123,060.00
Sunglasses	7,605.00	228,150.00
Gold Bracelet	7,207.00	648,630.00
Tổng	175,712.00	7,981,107.00

Cơ cấu doanh thu theo loại sản phẩm

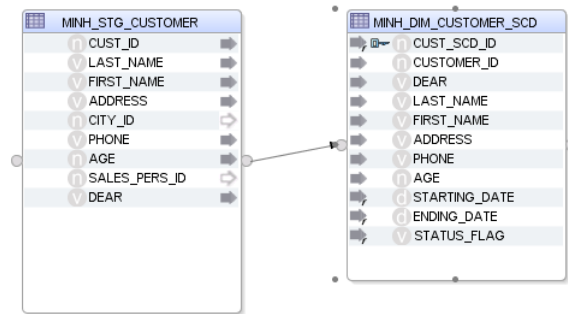
CATEGORY

- Jewels
- Watches
- Sportswear
- Equipment
- Various

Tốc độ tăng trưởng doanh thu

3.7. Xử lý SCD Type 2 (Lưu vết thay đổi lịch sử)

- **Thiết kế bảng:** DIM_CUSTOMER_SCD (lưu lịch sử khách hàng)
- **Mapping nâng cao:** cập nhật khi có thay đổi thông tin khách hàng
- **Cột SCD cần bổ sung:** starting_date, ending_date, status_flag



Overview Logical Physical

MINH_DIM_CUSTOMER_SCD - Properties

Find

Attributes

General Target Journalizing Constraints Connector Points

Name	Data Type	Le...	Sc...	D...	Expression	Ex...	Fl...	Key	In...	U...	C...
CUST_SCD_ID	NUMBER				:MINH_BTCK.SEQ_SCD_NEXTVAL	N...		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
CUSTOMER_ID	NUMBER				MINH_STG_CUSTOMER.CUST_ID	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
DEAR	VARCHAR2	50			MINH_STG_CUSTOMER.DEAR	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
LAST_NAME	VARCHAR2	50			MINH_STG_CUSTOMER.LAST_NAME	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
FIRST_NAME	VARCHAR2	50			MINH_STG_CUSTOMER.FIRST_NAME	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
ADDRESS	VARCHAR2	100			MINH_STG_CUSTOMER.ADDRESS	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
PHONE	VARCHAR2	50			MINH_STG_CUSTOMER.PHONE	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
AGE	NUMBER				MINH_STG_CUSTOMER.AGE	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
STARTING_DATE	DATE				SYSDATE	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
ENDING_DATE	DATE				SYSDATE	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
STATUS_FLAG	VARCHAR2	1			1	N...		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	