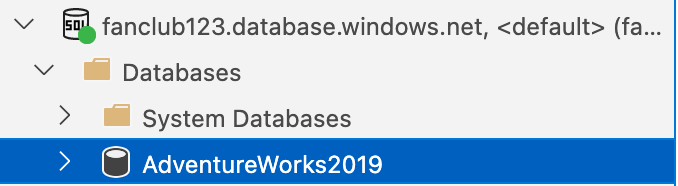
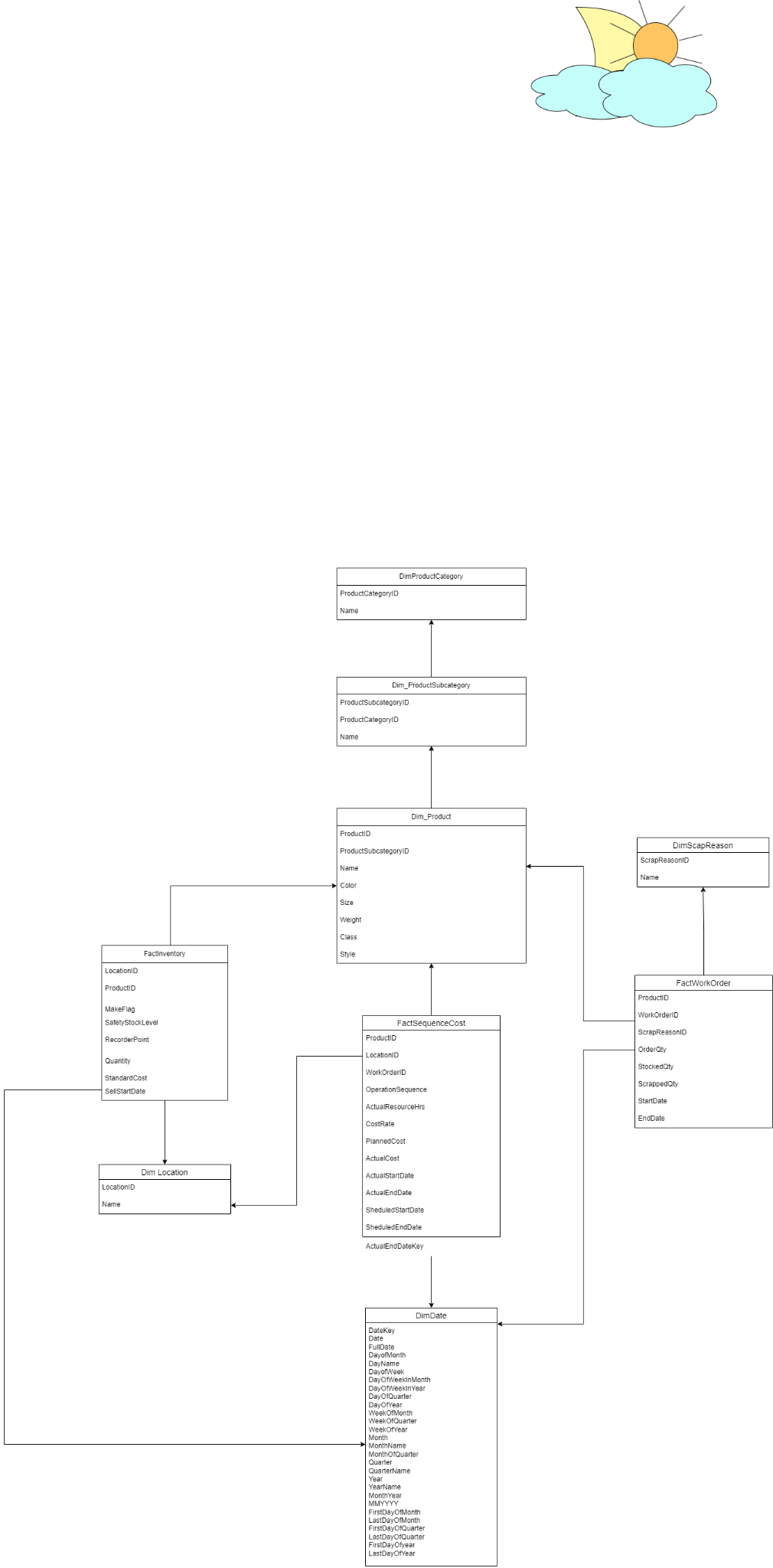
1. ***Preparation***

*1.1 Tạo ra database lưu trữ dữ liệu vận hành như là tầng Staging:*

Nhóm tác giả thực hiện tải toàn bộ database AdventureWorks2019 như hình x mà nhóm sử dụng cho môn học này lên Cloud và sử dụng như là tầng Staging. Điều này có ý nghĩa rằng các hoạt động hàng ngày của doanh nghiệp (hay định ky tuỳ theo doanh nghiệp quyết định) dù ở nhiều nguồn hay nhiều nơi đều sẽ được tải lên và nằm ở database này.



Đồng thời, ở bước này nhóm cũng thực hiện viết ra các Script cho các bảng Dim và Fact mà nhóm đã thiết kế:



**DimProductCategory**

SELECT ProductCategoryID, Name

FROM Production.ProductCategory

CREATE TABLE DimProductCategory (

ProductCategoryID int PRIMARY KEY,

Name nvarchar(50)

)

**DimProductSubcategory**

SELECT ProductSubcategoryID, ProductCategoryID, Name

FROM Production.ProductSubcategory

CREATE TABLE DimProductSubcategory (

ProductSubcategoryID int PRIMARY KEY,

ProductCategoryID int,

Name nvarchar(50)

)

**DimProduct**

SELECT ProductID, ProductSubcategoryID, Name, Color, Size, Weight, Class, Style

FROM Production.Product

CREATE TABLE DimProduct(

ProductID INT PRIMARY KEY,

ProductSubcategoryID INT,

Name NVARCHAR(50),

Color NVARCHAR(15),

Size NVARCHAR(5),

Weight DECIMAL(8,2),

Class NVARCHAR(2),

Style NVARCHAR(2)

)

**DimLocation**

SELECT [LocationID]

,[Name]

FROM [Production].[Location]

CREATE TABLE [DimLocation] (

[LocationID] INT PRIMARY KEY,

[Name] NVARCHAR(50)

)

**DimScrapReason**

SELECT ScrapReasonID

, Name

FROM [Production].[ScrapReason]

CREATE TABLE [DimScrapReason] (

[ScrapReasonID] INT PRIMARY KEY,

[Name] NVARCHAR(50)

)

**FactWorkOrder**

SELECT ProductID,

ScrapReasonID,

WorkOrderID,

OrderQty,

StockedQty,

ScrappedQty,

StartDate,

convert(varchar(20), EndDate,112) as EndDate

FROM Production.WorkOrder

CREATE TABLE FactWorkOrder(

ProductID INT,

WorkOrderID INT,

ScrapReasonID INT,

OrderQty INT,

StockedQty INT,

ScrappedQty INT,

StartDate DATETIME,

EndDate varchar(20)

)

**FactSequenceCost**

SELECT ProductID, work.LocationID, WorkOrderID, OperationSequence, ActualResourceHrs, lo.CostRate, PlannedCost, ActualCost, ActualEndDate

convert(varchar(20), ActualEndDate,112) as ActualEndDate

FROM Production.WorkOrderRouting work

JOIN Production.Location lo

ON work.LocationID = lo.LocationID

CREATE TABLE FactSequenceCost (

ProductID INT,

LocationID INT,

WorkOrderID INT,

OperationSequence INT,

ActualResourceHrs DECIMAL(9, 4),

CostRate money,

PlannedCost money,

ActualCost money,

ActualEndDate varchar(20)

);

**FactInventory**

SELECT inv.LocationID, pr.ProductID, ProductSubcategoryID, MakeFlag, SafetyStockLevel, ReorderPoint, StandardCost,Quantity, convert(varchar(20), SellStartDate,112) as SellStartDate

FROM Production.ProductInventory inv

JOIN Production.Product pr

ON inv.ProductID = pr.ProductID

CREATE TABLE FactInventory (

LocationID INT,

ProductID INT,

ProductSubcategoryID INT,

MakeFlag BIT,

SafetyStockLevel INT,

ReorderPoint INT,

StandardCost money,

Quantity INT,

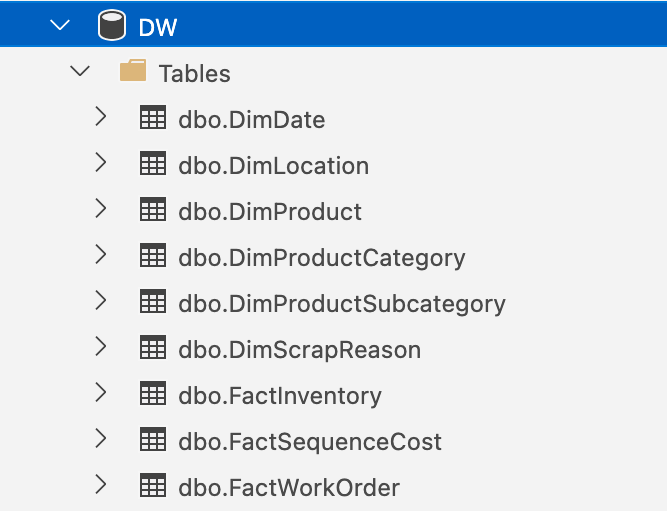
SellStartDate varchar(20)

)

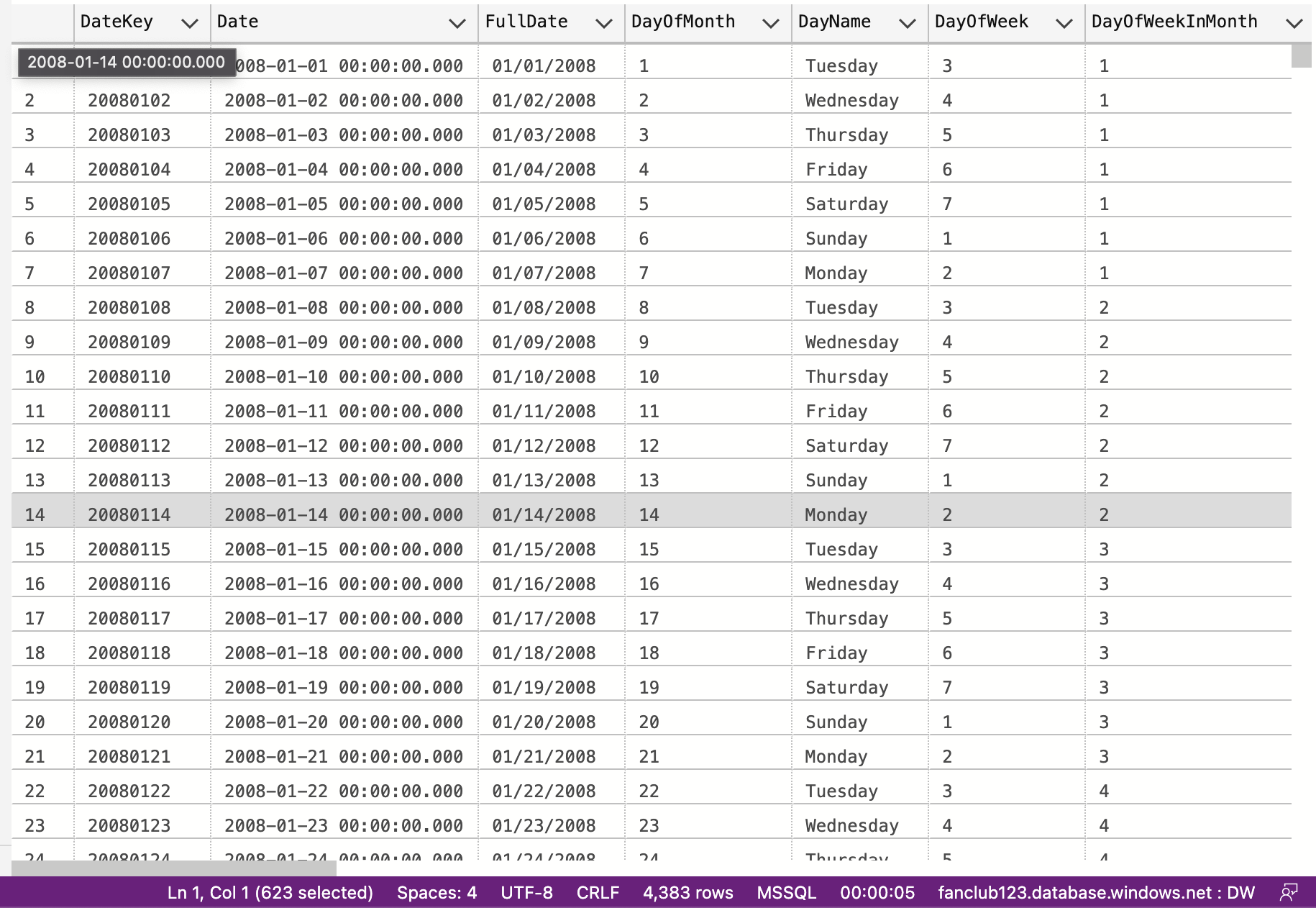
1.2 Tạo ra database mục tiêu DW:

Tiếp đến nhóm tác giả tạo ra 1 database tên là DW như hình x. Database này để lưu trữ các bảng Dim và Fact hoàn chỉnh cho thiết kế Data Warehouse.

Bước đầu, nhóm tiến hành tạo rỗng các bảng Dim và Fact với các kiểu dữ liệu thuộc tính tương thích với dữ liệu nguồn ở tầng Staging.

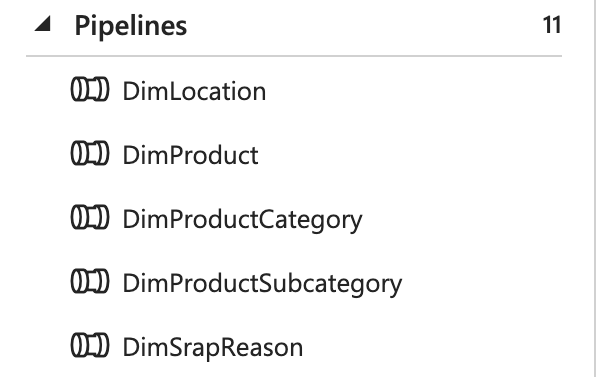


Ngoài các bảng ở trên, nhóm cần bảng Dim Date phù hợp với khoảng thời gian hoạt động dữ liệu doanh nghiệp là từ 2008-2019. Bảng DimDate được tạo sẽ giống như sau:

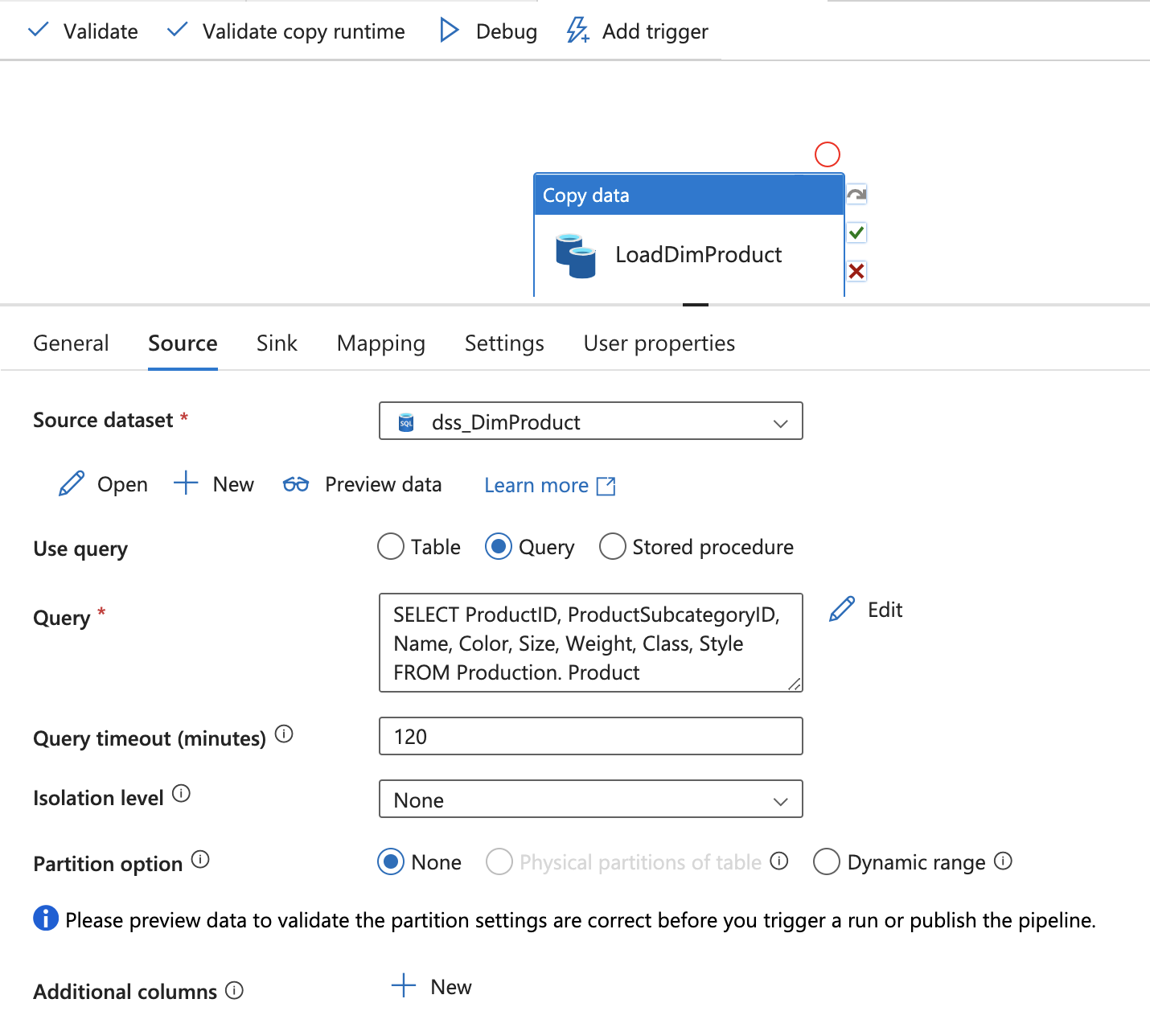


1. ***Xây dựng ADF pipeline cho các bảng Dim***

Nhóm tác giả thực hiện tạo Pipeline cho từng bảng Dim như hình x



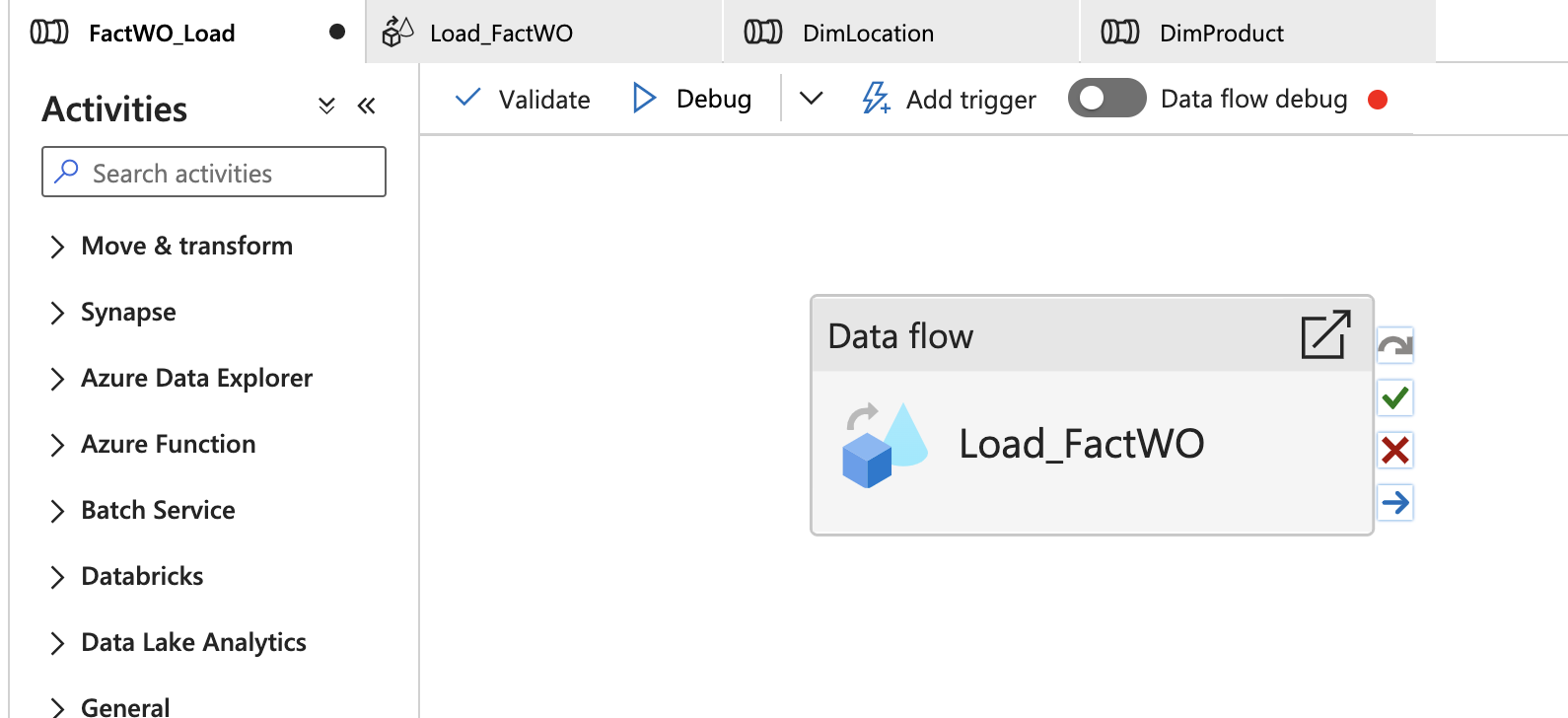
Với mỗi bảng Dim, nhóm tác giả sẽ thực hiện lấy Source bằng query đã chuẩn bị và lấy từ database Staging (AdventureWorks2019) đã tải lên server. Đồng thời cũng thực hiện tạo link service trỏ đến database Staging.



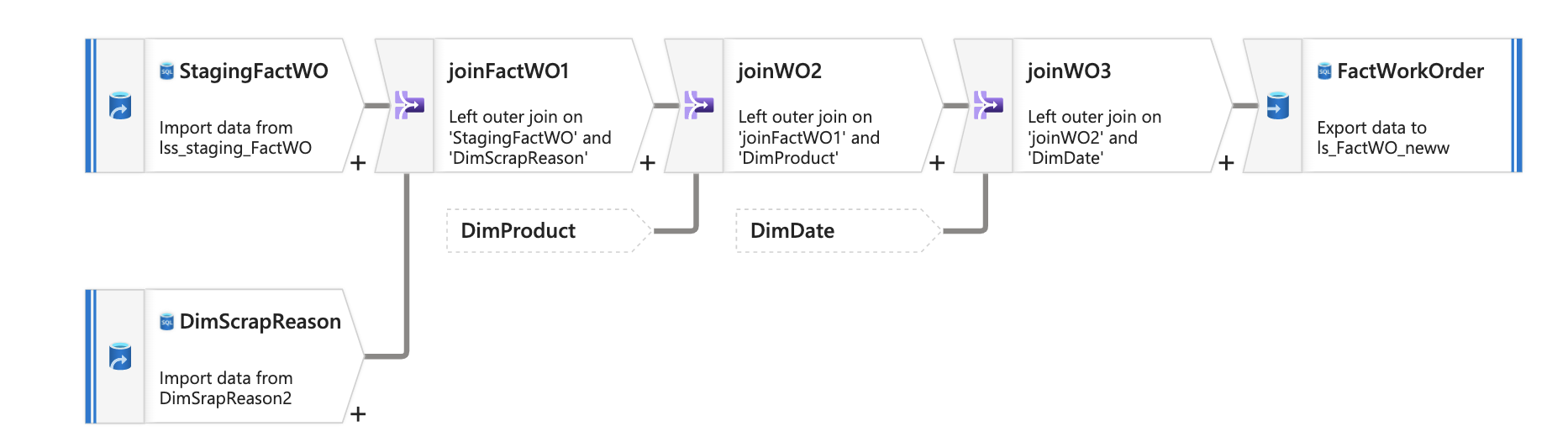
Còn ở tab Sink, nhóm tạo link service tới database DW đã tạo trên server.

Như vậy, tương tự cho các bảng Dim khác.

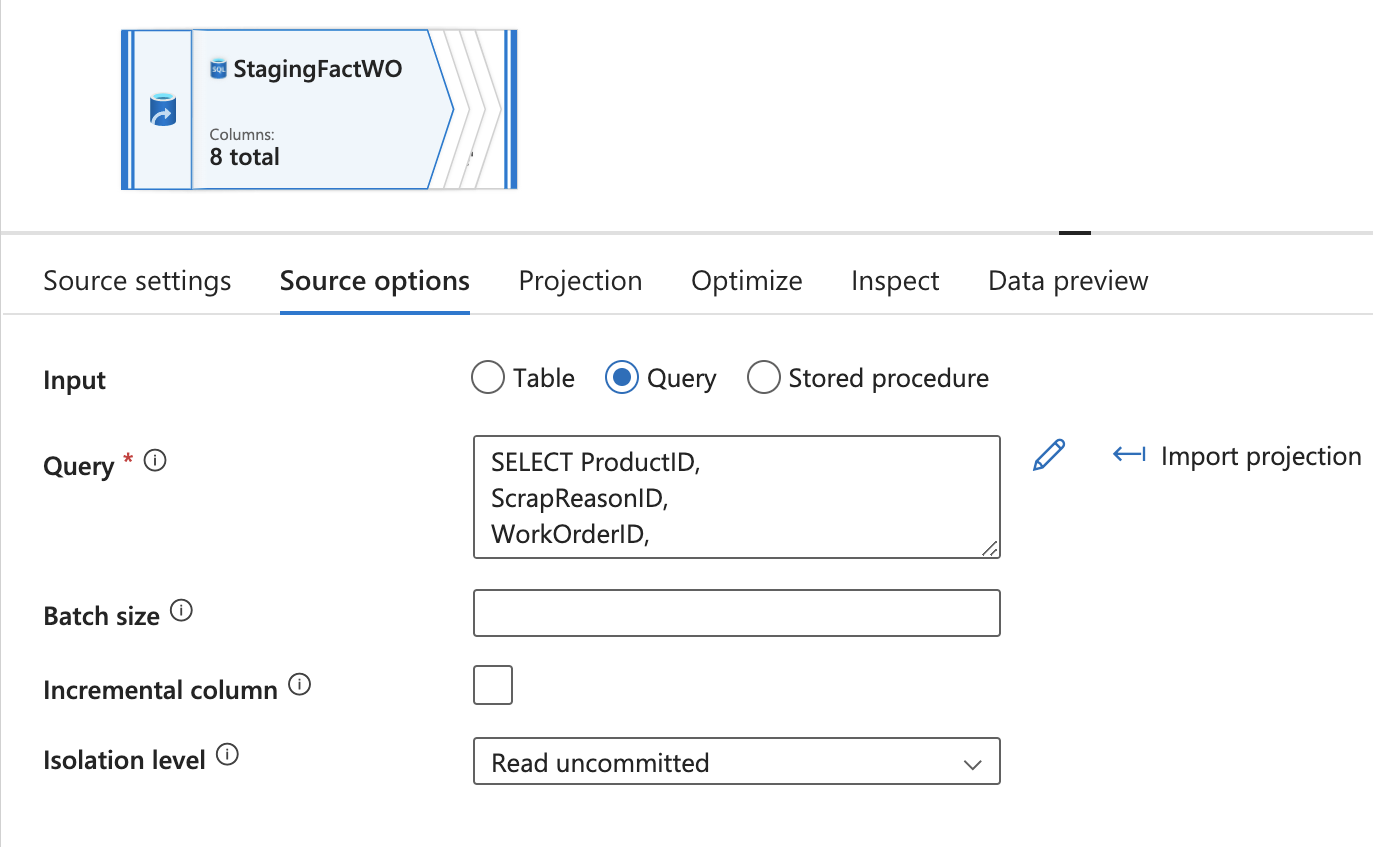
1. ***Xây dựng ADF pipeline cho bảng FACT***

Nhóm thực hiện tạo Pipeline, sau đó kéo sử dụng activity Data Flow như hình x 

Ở bước này, với mỗi bảng Fact tương ứng có mối liên hệ với các Dim riêng nó. Dưới đây là snapshot nhanh về 1 dataflow của bảng FactWorOrder:



Tương tự ở các bảng Dim, nhóm cũng thực hiện query để lấy dữ liệu cho Fact từ tằng staging.



Bên cạnh đó, nhóm tác giả sử dụng tác vụ “JOIN” để join đồng thời giữa dữ liệu của Fact và Dim tương ứng của nó trước khi đổ dữ liệu vào bảng Fact cuối cùng ở database DW.

DW

Như vậy, nhóm đã thực hiện hoàn tất quá trình ETL theo kiểu Full-load, dưới đây là bản thiết kế Data Warehouse hoàn chỉnh của nhóm:

