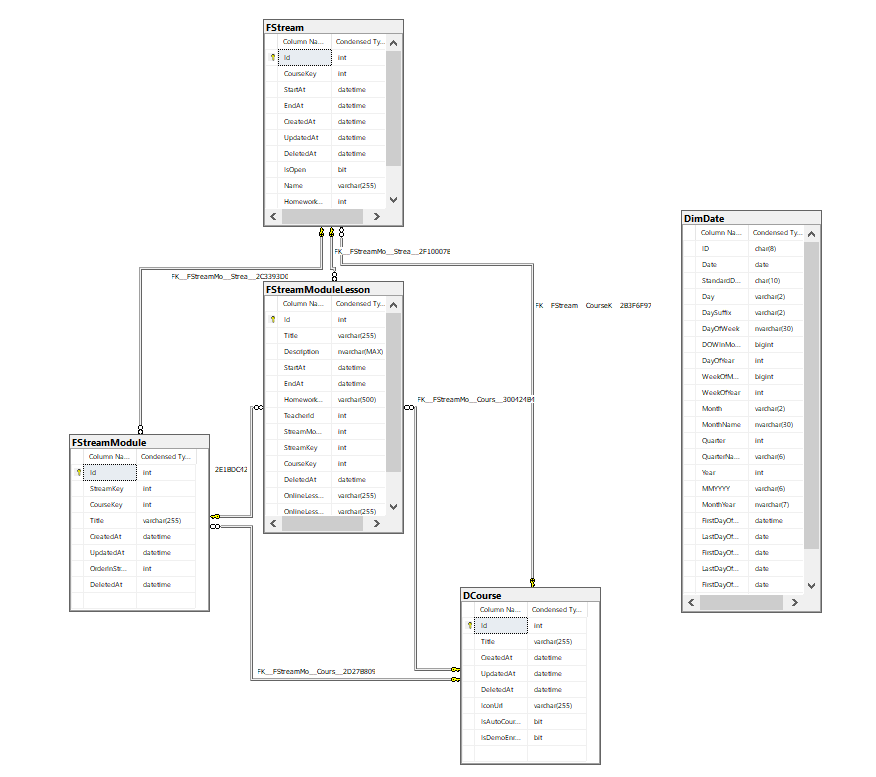
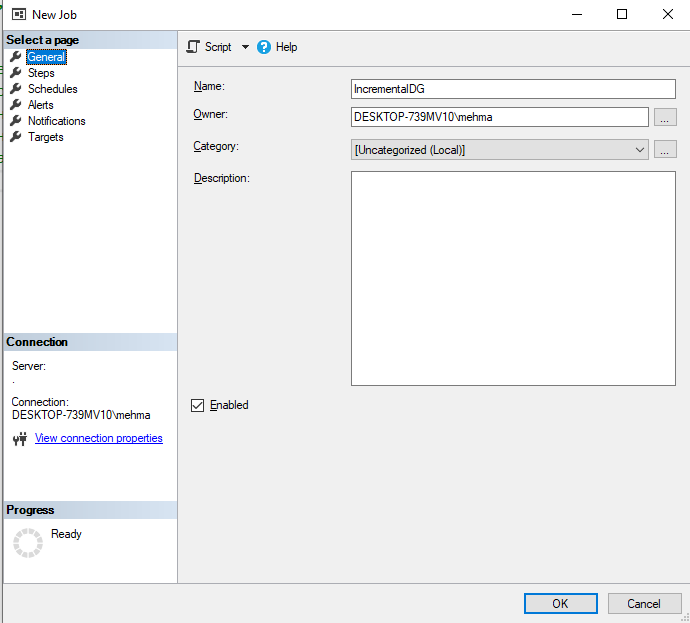
Есть две основные схемы которые я знаю. Это STAR и SNOWLAKE. На самом деле обе моделей хороши. Но основной элемент это память. Если есть ресурс памяти то лучше использовать STAR так как оно более понятное для бизнеса один чуть сложный момент это при миграции нужно из 3 нормальной формы понизить во вторую. Насчет SNOWLAKE оно хорошо тем что занимает меньше памяти но сложен чуть для бизнеса но легче для инженера так как оно сидит в 3 нормальной форме почти как в OLTP. Я нарисую Схему гибридную(STAR и SNOWLAKE) а Data Model будет Логическая только с добавкой типов данных.

Почему я выбрал такую модель это вопрос я думаю взгляда. Если компания слишком большая наверняка не стоит выходить за пределы определенной схемы, но так как тут данных мало мне показалось так будет понятнее. DimDate я не включил во внешние ключи по простой причине так как на практике было такое что из за этого у меня остановился ЕТЛ. Я конечно расширил бы все это изменил бы некоторые моменты но тут думаю много импровизировать не так уж хорошо будет.



Для ETL я буду использовать SSIS(SQL Server Integration Services).

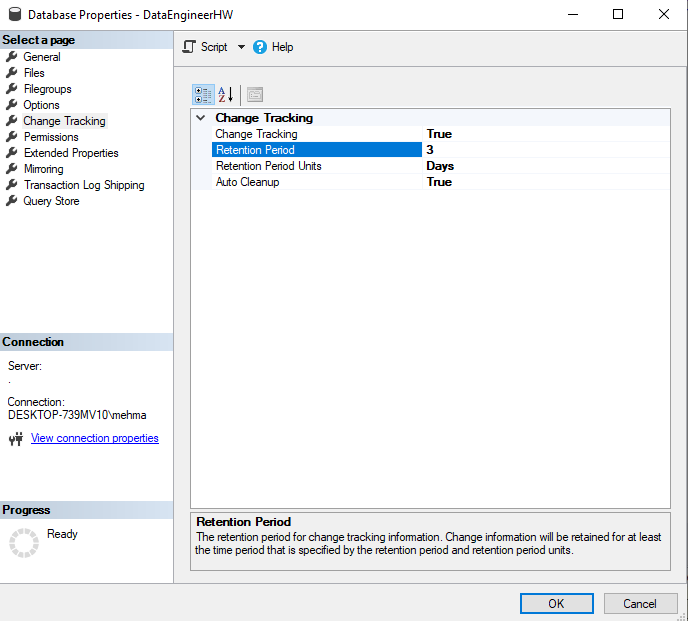
Чтоб работало по расписянию создаю Джоб через Агент



Для инкремента таблиц можно использовать несколько способов, но я выбрал ChangeTracking.

ChangeTracking я поставлю на 3 дня. Возможно появится вопрос почему я не использовал обычные Тригерры для инкремента или же CDC(Change Data Capture) ответ сново взгляд. Если бы мы тянули не из самого сервера можно было бы использовать другие методы.

SSIS Проект я добавлю как проект тоже. Я там использовал максимально понятную модель ЕТЛ точнее ЕЛТ использовал SSIS только как оркестратор так как обе части у меня SQLServer я использовал это. В добавок я бы вывел эти Merg-и в процедуры. Это будет намного лучше чем внутри пакета. Самые простые пары причин это то-что во первых процедуры хранят план , а во вторых если будук какие то ощибки или же доработки нам не нужно будет делать каждый раз деплой пакета можно просто через CI/CD в Git процедурку.



Это для того чтобы если как то что то несработает можно два дня не сливать данные фулл. Точнее если как то два дня ETL не сработает нам не придется запускать фулл.

Насчет витрины данных я построю простую витрину для анализа данных. Но на самом деле я бы использовал тут (SSAS) И в Tabular – ной модели передал бы бизнесу. Это бы намного увеличило скорость работы а еще такие Tool ы как PowerBi и Excel очень хорошо работают с SSAS. В витрину DimDate Я не добавлю это лучще по нужде давать бизнесу или же если ужсильно надо приджойнить.