



ETL

Incremental load of the fact table

Assumptions

Initial load of dimensional data done

Initial load of facts $\leq 31.12.1997$ done

Product deleted , product added, product changed (in the source database NorthWindDB)

CDC for dimensional data done for data $> 31.12.1997$

1. Create (or empty) the temporary fact table

- All fields from fact table + all business keys fields
- Allow NULLs
- No primary key or foreign keys
 - (that's ok – it is only a temporary table)

```
CREATE TABLE [dbo].[temp_fact_sales](  
[C_ID] [int] NULL,  
[E_ID] [int] NULL,  
[P_ID] [int] NULL,  
[D_ID] [int] NULL,  
[CustomerID] [nvarchar](5) NULL,  
[ProductID] [int] NULL,  
[EmployeeID] [int] NULL,  
[OrderDate] [datetime] NULL,  
[Quantity] [bigint] NULL,  
[SalesAmount] [decimal](18, 2) NULL
```

2. Populate the temporary fact table

Insert data in all fields
EXCEPT the surrogate
key fields

```
use NorthWindDB
go
insert into
[DM_Northwind].[dbo].[temp_fact_sales](
Quantity
,SalesAmount
,ProductID
,CustomerID
,EmployeeID
,OrderDate
)
```

```
select
Quantity
,[Order Details].UnitPrice * Quantity
,Products.ProductID
,Customers.CustomerID
,Employees.EmployeeID
,Orders.OrderDate
from [Order Details]
join Orders on
Orders.OrderID = [Order Details].OrderID
join Customers on
Orders.CustomerID = Customers.CustomerID
join Employees on
Employees.EmployeeID = Orders.EmployeeID
join Products on
Products.ProductID = [Order Details].ProductID
where Orders.OrderDate >
(select lastUpdate from
DM_Northwind.dbo.LastUpdate)
```

3. Look up keys in dimension tables (and update the temporary fact table)

Update each surrogate key field by looking up the value in the corresponding dimension table

```
UPDATE [temp_fact_sales]
SET P_ID = (select P_ID from dim_product
where dim_product.ProductID =
temp_fact_sales.ProductID
AND ValidTo = '9999-12-31'
)
```

Repeat for all
dimensions

4. Populate the final fact table (add rows)

Insert only surrogate keys and facts from the temporary table to the final fact table

```
INSERT INTO F_Sales(  
  D_ID  
  ,C_ID  
  ,P_ID  
  ,E_ID  
  ,Quantity  
  ,SalesAmount  
)
```

```
SELECT  
  d_id  
  ,c_id  
  ,p_id  
  ,e_id  
  ,Quantity  
  ,SalesAmount  
FROM temp_fact_sales
```

Exercise

1)

Write the SQL code for the incremental load of your fact table

2)

Test your solution

Write some queries "across" the changes, like this:



Save your results - they will be part of hand in #4

```
Select  p.name
        , p.category
        , sum(f.itemsSold) as
i..
from    factsale f
join    dimProduct p
on      (p.ID = f.productID)
where   p.name = 'Ketchup'
group by
        rollup( p.name,
p.category)
```

Name	Category	itemsSold
-----	-----	-----
Ketchup	condiment	3
Ketchup	vegetable	4
Ketchup		7