**Data Warehousing Exercise 1**

1. Create CDM, PDM, LDM, Bus Matrix for the below databases:
   1. Hr
   2. Northwind or Airports
2. Perform the below SCD’s based on the conditions below,

Note: Create Target tables with required additional columns

|  |  |  |
| --- | --- | --- |
| SCD Type | Source Table | SCD column |
| SCD 1 | northwind.customers | City |
| SCD 3 (2 History) | northwind.employees | FirstName |
| SCD 2 Date type | northwind.suppliers | Contact Name |
| SCD 2 Flag type | northwind.categories | CategoryName |
| SCD 2 Version type | northwind.products | Productname |

SCD 1:

SELECT \* FROM [Northwind].[Customers]

SELECT \* INTO SCD1\_SR FROM [Northwind].[Customers]

SELECT \* FROM SCD1\_SR

SELECT \* INTO SCD1\_TR FROM SCD1\_SR

INSERT INTO SCD1\_TR

SELECT DISTINCT S.CustomerID,S.CompanyName,S.ContactName,S.ContactTitle,S.Address,S.City,S.Region,S.PostalCode,S.Country,

S.Phone,S.Fax

FROM SCD1\_SR S LEFT JOIN SCD1\_TR T ON T.CustomerID = S.CustomerID

WHERE T.CustomerID IS NULL

UPDATE SCD1\_TR

SET City = S.City

FROM SCD1\_TR T JOIN SCD1\_SR S

ON S.CustomerID = T.CustomerID

WHERE S.City <> T.City

SCD 2 (DATE TYPE):

SELECT \* INTO SCD\_2 FROM [Northwind].[Suppliers]

SELECT \* INTO SCD\_TR FROM [Northwind].[Suppliers]

ALTER TABLE SCD\_TR

ADD S\_ID INT IDENTITY(1,1), START\_DATE DATETIME,END\_DATE DATETIME

INSERT INTO SCD\_TR

SELECT DISTINCT S.SupplierID,S.CompanyName,S.ContactName,S.ContactTitle,S.Address,S.City,S.Region,S.PostalCode,S.Country,S.Phone,

S.Fax,GETDATE() as START\_DATE, NULL as END\_DATE

FROM SCD\_2 S LEFT JOIN SCD\_TR T ON S.SupplierID = T.SupplierID

where (T.END\_DATE IS NULL and T.ContactName != S.ContactName) or T.SupplierID is null

UPDATE SCD\_TARGET

SET END\_DATE = GETDATE()

FROM SCD\_TR T JOIN SCD\_2 S

ON T.SupplierID = S.SupplierID

WHERE T.ContactName != S.ContactName and T.END\_DATE is NULL

UPDATE SCD\_2

SET ContactName = 'santosh'

WHERE SupplierID = 1

alter table SCD\_TR

DROP COLUMN HomePage

SCD 2 (FLAG):

SELECT \* FROM [Northwind].[Categories]

SELECT \* INTO SCD\_SOURCE FROM [Northwind].[Categories]

SELECT \* INTO CAT\_TARGET FROM [Northwind].[Categories]

ALTER TABLE CAT\_TARGET

ADD S\_ID INT IDENTITY(1,1) Primary key,FLAG VARCHAR(2)

UPDATE SCD\_SOURCE

SET CategoryName = 'rice'

WHERE CategoryID = 6

INSERT INTO CAT\_TARGET

SELECT S.CategoryID,S.CategoryName,S.Description,S.Picture,'Y' as FLAG

FROM SCD\_SOURCE S LEFT JOIN CAT\_TARGET T ON S.CategoryID = T.CategoryID

WHERE T.CategoryID IS NULL or (T.CategoryName != S.CategoryName and T.FLAG = 'Y')

UPDATE CAT\_TARGET

SET FLAG = 'N'

FROM SCD\_SOURCE S JOIN CAT\_TARGET T

ON S.CategoryID = T.CategoryID

WHERE S.CategoryName != T.CategoryName and T.FLAG = 'Y'

UPDATE SCD\_SOURCE

SET CategoryName = 'phone'

WHERE CategoryID = 6

SCD 3 :

SELECT \* FROM [Northwind].[Employees]

SELECT \* INTO EMP\_SR FROM [Northwind].[Employees]

SELECT \* INTO EMP\_tr FROM [Northwind].[Employees]

ALTER TABLE EMP\_tr

ADD S\_KEY INT IDENTITY(1,1)

INSERT INTO EMP\_tr

SELECT S.EmployeeID,S.LastName,S.FirstName,S.Title,S.TitleOfCourtesy,S.BirthDate,S.HireDate,S.Address,S.City,

S.Region,S.PostalCode,S.Country,S.HomePhone,S.Extension,S.Photo,S.Notes,S.ReportsTo,S.PhotoPath,

NULL as FIR\_HIS,NULL AS SEC\_HIS

FROM EMP\_SR S LEFT JOIN EMP\_tr T ON S.EmployeeID = T.EmployeeID

WHERE T.EmployeeID IS NULL

UPDATE EMP\_TR

SET FirstName = S.FirstName, FIR\_HIS = T.FirstName, SEC\_HIS = T.FIR\_HIS

FROM EMP\_SR S JOIN EMP\_tr T ON T.EmployeeID = S.EmployeeID

WHERE T.FirstName != S.FirstName

UPDATE EMP\_SR

SET FirstName = 'santosh'

where EmployeeID = 1

1. Load the date Dimension using SQL.

create table Dates1

(id int primary key not null,

Dates date,

year\_id int

)

create procedure form\_dates

as

declare @start\_date date = '01-01-1992'

declare @end\_date date = '12-31-2024'

declare @count int =0

declare @year int =1992

declare @year\_id int = 1

while @start\_date <= @end\_date

begin

insert into Dates values(@count+1,@start\_date,@year\_id)

set @count = @count+1

set @start\_date = dateadd(day,1,@start\_date)

if @year <> year(@start\_date)

set @year\_id = @year\_id+1

set @year = year(@start\_date)

end

exec form\_dates

create table Years

(

year\_id int not null identity(1,1) primary key,

years int not null

)

insert into Years

select distinct(year(Dates)) from Dates

order by year(Dates)

create table quaters

(quater\_id int identity(1,1) primary key,

year\_id int not null,

quater\_number int not null,

foreign key (year\_id) references years(year\_id))

insert into quaters

select distinct(y.year\_id),datename(qq,d.dates)

from years y join dates d on y.year\_id = d.year\_id

order by year\_id

create table months

(month\_id int not null primary key identity(1,1),

month\_number int ,

month\_name varchar(20),

year\_id int not null,

quater\_id int not null,

foreign key (year\_id) references years(year\_id),

foreign key (quater\_id) references quaters(quater\_id)

);

insert into months

select distinct(month(d.dates)),datename(month,d.dates),d.year\_id,datename(qq,d.dates)

from dates d join years y on y.year\_id = d.year\_id

order by d.year\_id

1. Load the date Dimension using python.
2. Perform Source system analysis for HR database. Template attached below with two sample tables.

