**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 |

SCD1:-

select \* into source from [Northwind].[Customers]

select \* into target from source

truncate table target

insert into target

select a.[CustomerID], a.[CompanyName], a.[ContactName],

a.[ContactTitle], a.[Address],a.[City], a.[Region],

a.[PostalCode],a.[Country], a.[Phone], a.[Fax] from source a left join target b ON a.CustomerID=b.CustomerID

where b.CustomerID IS NULL;

update source

SET City='Hyderabad'

WHERE CustomerID='ALFKI';

UPDATE target

set target.City=a.city

from source a inner join target b on a.CustomerID=b.CustomerID;

SCD2 DateType:-

select \* into sourcesuppliers from [Northwind].[Suppliers]

select \* into targetsuppliers from sourcesuppliers

truncate table targetsuppliers

alter table targetsuppliers

add startdate date,enddate date;

insert into targetsuppliers select a.[SupplierID],a.[CompanyName],

a.[ContactName],a.[ContactTitle], a.[Address], a.[City],

a.[Region], a.[PostalCode], a.[Country], a.[Phone],

a.[Fax], a.[HomePage],getdate() as startdate,b.enddate from sourcesuppliers a left join targetsuppliers b

on a.SupplierID=b.SupplierID

where b.SupplierID IS NULL;

update sourcesuppliers

set contactname='priya'

where SupplierID=1;

update targetsuppliers

set targetsuppliers.enddate=getdate()

from sourcesuppliers a inner join targetsuppliers b

on a.SupplierID=b.SupplierID

where a.contactname<>b.contactname;

SCD2 Flagtype:-

select \* into sourcecategories from [Northwind].[Categories]

alter table sourcecategories

add flag char(2);

select \* into targetcategories from sourcecategories

truncate table targetcategories

insert into targetcategories select a.[CategoryID],

a.[CategoryName], a.[Description],a.[Picture],

'Y' as flag from sourcecategories a

left join targetcategories b on a.CategoryID=b.CategoryID

where b.CategoryID IS NULL;

update sourcecategories

set CategoryName='Products'

where CategoryID=1;

update targetcategories

set targetcategories.flag='N'

from sourcecategories a

inner join targetcategories b

on a.CategoryID=b.CategoryID

where a.CategoryName in ('Products','Condiments','Produce');

SCD2 versiontype:-

select \* into sourceproducts from [Northwind].[Products];

alter table sourceproducts

add version int;

select \* into targetproducts from sourceproducts

TRUNCATE TABLE targetproducts

select \* from targetproducts

insert into targetproducts

select a.[ProductID],a.[ProductName],a.[SupplierID],

a.[CategoryID],a.[QuantityPerUnit],a.[UnitPrice],

a.[UnitsInStock],a.[UnitsOnOrder],a.[ReorderLevel],

a.[Discontinued],1 as 'version' from

sourceproducts a left join targetproducts b

on a.ProductID=b.ProductID

where b.ProductID IS NULL;

update sourceproducts

set ProductName='Kurta'

where ProductID=1;

update targetproducts

set targetproducts.version=2

from sourceproducts a join targetproducts b

on a.ProductID=b.ProductID

where b.ProductID in (2,4,6,8);

SCD3 history:-

select \* into sourceemployee from [Northwind].[Employees]

select \* from sourceemployee

select \* from targetemployee

alter table sourceemployee

add previous\_name varchar(20),current\_name varchar(20)

select \* into targetemployee from sourceemployee

truncate table targetemployee

insert into targetemployee select a.[EmployeeID],

a.[LastName], a.[FirstName], a.[Title], a.[TitleOfCourtesy],

a.[BirthDate], a.[HireDate], a.[Address], a.[City], a.[Region],

a.[PostalCode], a.[Country], a.[HomePhone],

a.[Extension], a.[Photo],a.[Notes],

a.[ReportsTo], a.[PhotoPath], a.previous\_name,a.current\_name

from sourceemployee a

left join targetemployee b on a.employeeid=b.employeeid

where b.employeeid IS NULL;

drop table targetemployee

update targetemployee

set FirstName='Priya'

WHERE EmployeeID=1;

update targetemployee

set targetemployee.current\_name=sourceemployee.FirstName,

targetemployee.previous\_name=targetemployee.current\_name

from sourceemployee left join

targetemployee on sourceemployee.employeeid=targetemployee.employeeid

where targetemployee.FirstName<>sourceemployee.FirstName;

1. Load the date Dimension using SQL.

create table dates(

id int,

dates datetime,

year\_id int

)

create procedure date\_list 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 date\_list

select \* from dates

# years

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)

select \* from years

# quarters

create table quarters

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

year\_id int not null,

quarter\_number int not null,

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

insert into quarters

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

select \* from quarters

# months

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,

quarter\_id int not null,

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

foreign key(quarter\_id) references quarters(quarter\_id));

insert into months

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

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

order by y.year\_id

select \* from months

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

