Property Analysis BI Developer Competition Tasks – Part 1

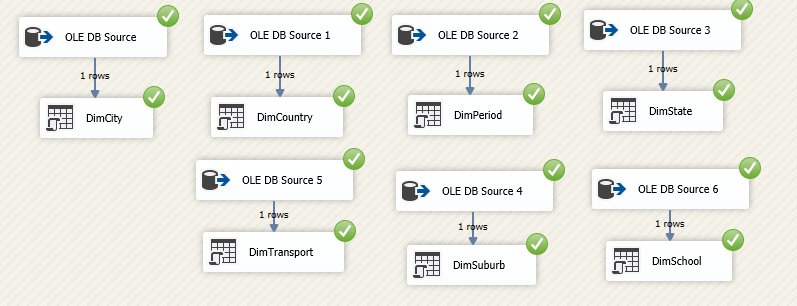
# **Part 1**

# Transform data from multiple data sources, ensure the accuracy and integrity of relevant data sets

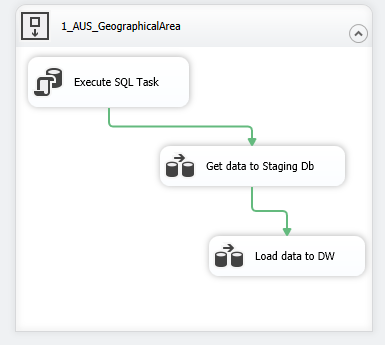
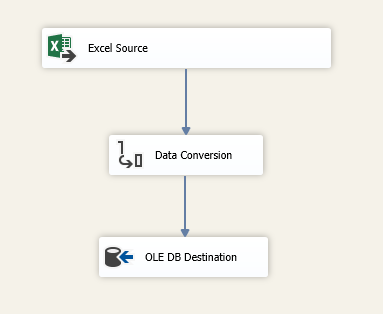
**Solution Approach:**

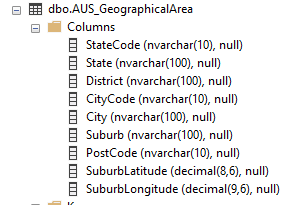
* Do manual simple corrections to the cell values that are catchable to plain sight
* Define staging database structure and warehouse database structure
* Load excel/csv files in to a staging database country-wise
  + Develop SSIS packages to automate the data cleansing and enriching to each country dataset
  + Populate default data for each dimension table on the warehouse database.
  + Load cleansed and enriched data into staging tables of each country
  + Map, combined, transform data from each staging table into standardised warehouse database table.

1. Populate default data for each dimension table on the warehouse database(To avoid NULL in the Foreign keys of the Fact Table)

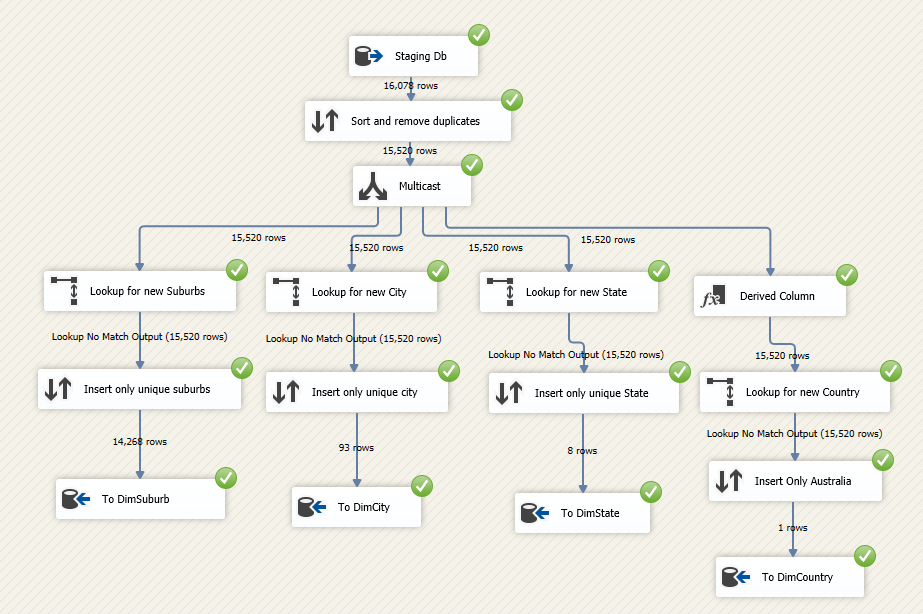


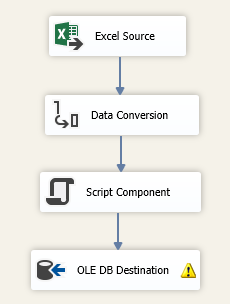
1. Australia Geographical Areas

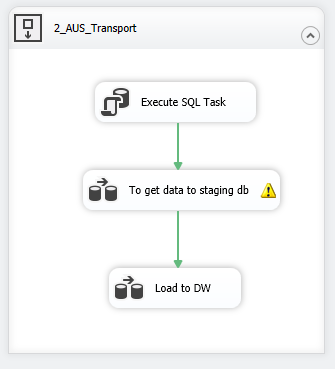
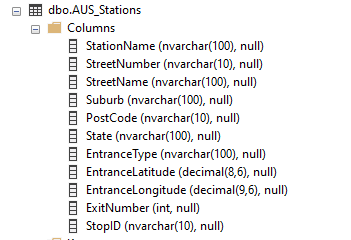




<Staging Table>



1. Australia Transport



<Staging Table>

// Script to retrieve address breakdown from Google Maps API by using latitude and longitude values

try

{

HttpResponseMessage response = client.GetAsync($"https://maps.googleapis.com/maps/api/geocode/json?latlng={Row.Latitude.Trim()},{Row.Longitude.Trim()}&key=xxxxx").Result;

if (response.IsSuccessStatusCode)

{

var jsonString = response.Content.ReadAsStringAsync().Result;

Rootobject rootObject = JsonConvert.DeserializeObject<Rootobject>(jsonString);

var result = rootObject.results[1];

if (result != null)

{

Row.StreetNumber = result.address\_components.Where(c => c.types.Contains("street\_number")).FirstOrDefault()?.short\_name;

Row.StreetName = result.address\_components.Where(c => c.types.Contains("route")).FirstOrDefault()?.long\_name;

Row.Suburb = result.address\_components.Where(c => c.types.Contains("locality")).FirstOrDefault()?.long\_name;

Row.PostCode = result.address\_components.Where(c => c.types.Contains("postal\_code")).FirstOrDefault()?.long\_name;

Row.State = result.address\_components.Where(c => c.types.Contains("administrative\_area\_level\_1")).FirstOrDefault()?.long\_name;

Row.Country = result.address\_components.Where(c => c.types.Contains("country")).FirstOrDefault()?.long\_name;

}

}

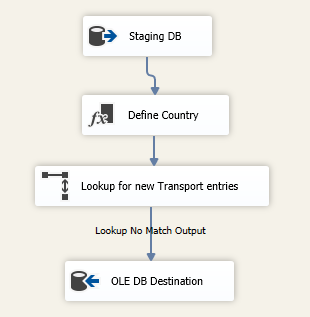
}

catch (Exception ex)

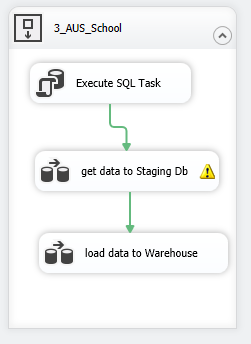
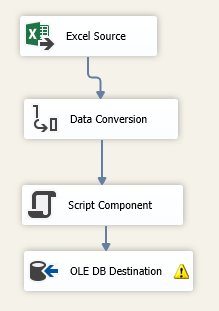
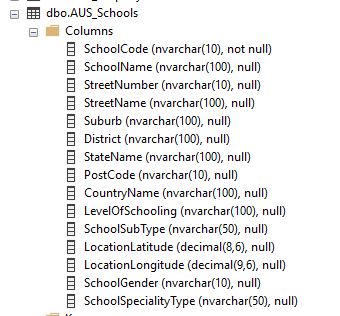
{

// TODO :; Find how to write into error output

}

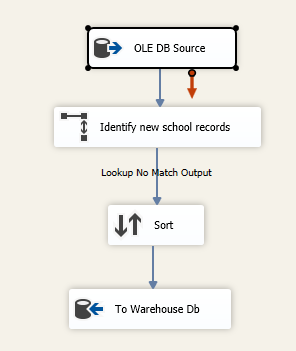


1. Australia Schools

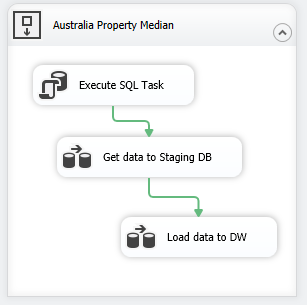
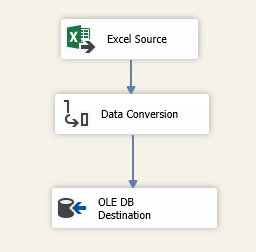
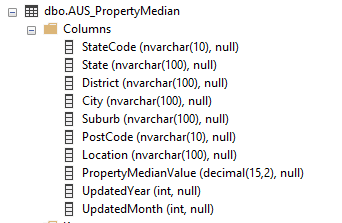


<Staging Table>

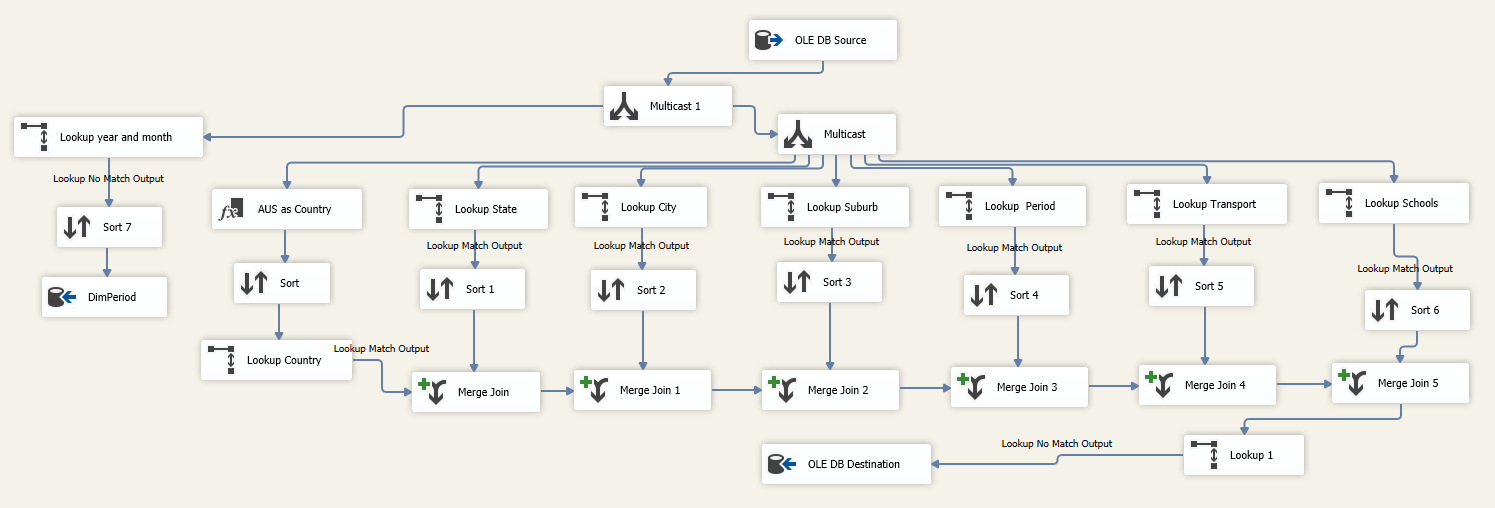
1. sdaf

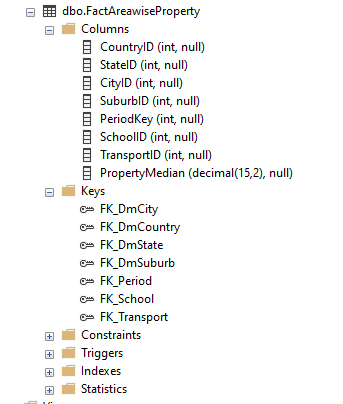


1. Australia Property Median (Loading to DW and Fact Table)



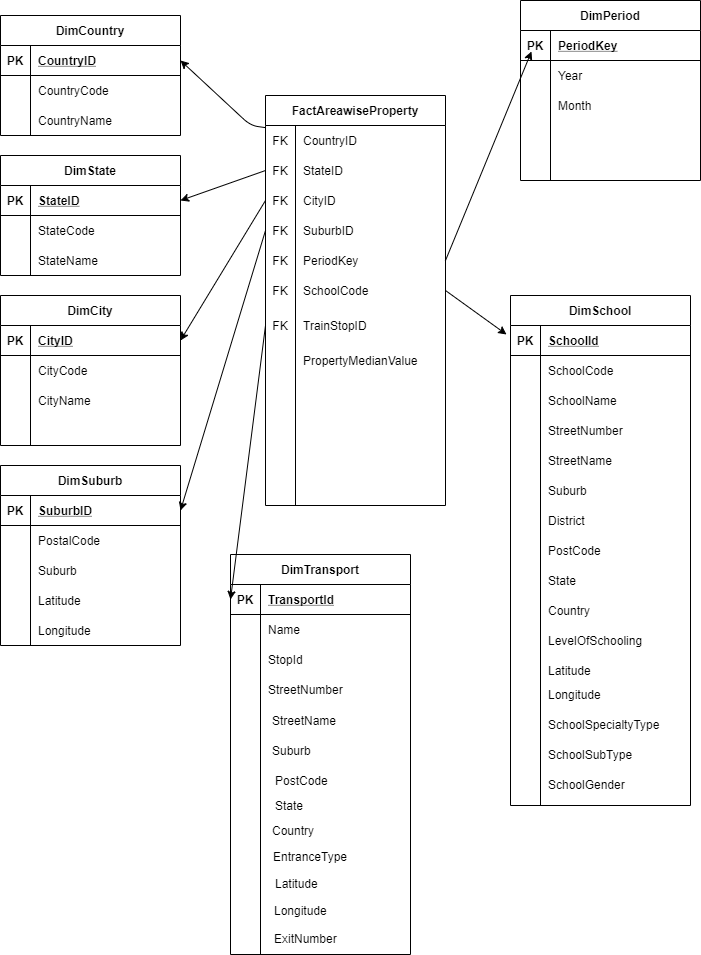
<Staging Table>





**Fact Table** - FacAreawiseProperty

# Design, build and test of a data warehouse solution.



# Generate and submit a copy of SQL script (\*.txt or \*.sql) after database and tables are created in your SQL Server.

USE [master]

GO

/\*\*\*\*\*\* Object: Database [Task1\_Warehouse] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

CREATE DATABASE [Task1\_Warehouse]

CONTAINMENT = NONE

ON PRIMARY

( NAME = N'Task1\_Warehouse', FILENAME = N'C:\Program Files\Microsoft SQL Server\MSSQL14.MSSQLSERVER\MSSQL\DATA\Task1\_Warehouse.mdf' , SIZE = 73728KB , MAXSIZE = UNLIMITED, FILEGROWTH = 65536KB )

LOG ON

( NAME = N'Task1\_Warehouse\_log', FILENAME = N'C:\Program Files\Microsoft SQL Server\MSSQL14.MSSQLSERVER\MSSQL\DATA\Task1\_Warehouse\_log.ldf' , SIZE = 73728KB , MAXSIZE = 2048GB , FILEGROWTH = 65536KB )

GO

ALTER DATABASE [Task1\_Warehouse] SET COMPATIBILITY\_LEVEL = 140

GO

IF (1 = FULLTEXTSERVICEPROPERTY('IsFullTextInstalled'))

begin

EXEC [Task1\_Warehouse].[dbo].[sp\_fulltext\_database] @action = 'enable'

end

GO

ALTER DATABASE [Task1\_Warehouse] SET ANSI\_NULL\_DEFAULT OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET ANSI\_NULLS OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET ANSI\_PADDING OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET ANSI\_WARNINGS OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET ARITHABORT OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET AUTO\_CLOSE OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET AUTO\_SHRINK OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET AUTO\_UPDATE\_STATISTICS ON

GO

ALTER DATABASE [Task1\_Warehouse] SET CURSOR\_CLOSE\_ON\_COMMIT OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET CURSOR\_DEFAULT GLOBAL

GO

ALTER DATABASE [Task1\_Warehouse] SET CONCAT\_NULL\_YIELDS\_NULL OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET NUMERIC\_ROUNDABORT OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET QUOTED\_IDENTIFIER OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET RECURSIVE\_TRIGGERS OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET DISABLE\_BROKER

GO

ALTER DATABASE [Task1\_Warehouse] SET AUTO\_UPDATE\_STATISTICS\_ASYNC OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET DATE\_CORRELATION\_OPTIMIZATION OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET TRUSTWORTHY OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET ALLOW\_SNAPSHOT\_ISOLATION OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET PARAMETERIZATION SIMPLE

GO

ALTER DATABASE [Task1\_Warehouse] SET READ\_COMMITTED\_SNAPSHOT OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET HONOR\_BROKER\_PRIORITY OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET RECOVERY FULL

GO

ALTER DATABASE [Task1\_Warehouse] SET MULTI\_USER

GO

ALTER DATABASE [Task1\_Warehouse] SET PAGE\_VERIFY CHECKSUM

GO

ALTER DATABASE [Task1\_Warehouse] SET DB\_CHAINING OFF

GO

ALTER DATABASE [Task1\_Warehouse] SET FILESTREAM( NON\_TRANSACTED\_ACCESS = OFF )

GO

ALTER DATABASE [Task1\_Warehouse] SET TARGET\_RECOVERY\_TIME = 60 SECONDS

GO

ALTER DATABASE [Task1\_Warehouse] SET DELAYED\_DURABILITY = DISABLED

GO

EXEC sys.sp\_db\_vardecimal\_storage\_format N'Task1\_Warehouse', N'ON'

GO

ALTER DATABASE [Task1\_Warehouse] SET QUERY\_STORE = OFF

GO

USE [Task1\_Warehouse]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimCity] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimCity](

[CityId] [int] IDENTITY(1,1) NOT NULL,

[CityCode] [nvarchar](10) NULL,

[City] [nvarchar](100) NULL,

CONSTRAINT [PK\_Dim\_City] PRIMARY KEY CLUSTERED

(

[CityId] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimCountry] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimCountry](

[CountryID] [int] IDENTITY(1,1) NOT NULL,

[CountryCode] [nvarchar](10) NULL,

[CountryName] [nvarchar](100) NULL,

CONSTRAINT [PK\_DimCountry] PRIMARY KEY CLUSTERED

(

[CountryID] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimPeriod] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimPeriod](

[PeriodKey] [int] IDENTITY(1,1) NOT NULL,

[Year] [int] NULL,

[Month] [int] NULL,

CONSTRAINT [PK\_DimPeriod] PRIMARY KEY CLUSTERED

(

[PeriodKey] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimSchool] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimSchool](

[SchoolId] [int] IDENTITY(1,1) NOT NULL,

[SchoolCode] [nvarchar](10) NULL,

[SchoolName] [nvarchar](100) NULL,

[StreetNumber] [nvarchar](10) NULL,

[StreetName] [nvarchar](100) NULL,

[Suburb] [nvarchar](100) NULL,

[District] [nvarchar](100) NULL,

[PostCode] [nvarchar](10) NULL,

[State] [nvarchar](100) NULL,

[Country] [nvarchar](100) NULL,

[LevelOfSchooling] [nvarchar](100) NULL,

[Latitude] [decimal](8, 6) NULL,

[Longitude] [decimal](9, 6) NULL,

[SchoolSpecialtyType] [nvarchar](100) NULL,

[SchoolSubType] [nvarchar](100) NULL,

[SchoolGender] [nvarchar](50) NULL,

CONSTRAINT [PK\_DimSchool] PRIMARY KEY CLUSTERED

(

[SchoolId] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimState] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimState](

[StateId] [int] IDENTITY(1,1) NOT NULL,

[StateCode] [nvarchar](10) NULL,

[StateName] [nvarchar](100) NULL,

CONSTRAINT [PK\_Dim\_State] PRIMARY KEY CLUSTERED

(

[StateId] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimSuburb] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimSuburb](

[SuburbId] [int] IDENTITY(1,1) NOT NULL,

[PostalCode] [nvarchar](10) NULL,

[Suburb] [nvarchar](100) NULL,

[Latitude] [decimal](8, 6) NULL,

[Longitude] [decimal](9, 6) NULL,

CONSTRAINT [PK\_Dim\_Suburb] PRIMARY KEY CLUSTERED

(

[SuburbId] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[DimTransport] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[DimTransport](

[TransportId] [int] IDENTITY(1,1) NOT NULL,

[Name] [nvarchar](100) NULL,

[StopId] [nvarchar](10) NULL,

[StreetNumber] [nvarchar](10) NULL,

[StreetName] [nvarchar](100) NULL,

[Suburb] [nvarchar](100) NULL,

[PostCode] [nvarchar](10) NULL,

[State] [nvarchar](100) NULL,

[Country] [nvarchar](100) NULL,

[EntranceType] [nvarchar](100) NULL,

[Latitude] [decimal](8, 6) NULL,

[Longitude] [decimal](9, 6) NULL,

[ExitNumber] [int] NULL,

CONSTRAINT [PK\_DimTransport] PRIMARY KEY CLUSTERED

(

[TransportId] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

GO

/\*\*\*\*\*\* Object: Table [dbo].[FactAreawiseProperty] Script Date: 23/12/2020 9:43:39 AM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[FactAreawiseProperty](

[CountryID] [int] NULL,

[StateID] [int] NULL,

[CityID] [int] NULL,

[SuburbID] [int] NULL,

[PeriodKey] [int] NULL,

[SchoolID] [int] NULL,

[TransportID] [int] NULL,

[PropertyMedian] [decimal](15, 2) NULL

) ON [PRIMARY]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_DmCity] FOREIGN KEY([CityID])

REFERENCES [dbo].[DimCity] ([CityId])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_DmCity]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_DmCountry] FOREIGN KEY([CountryID])

REFERENCES [dbo].[DimCountry] ([CountryID])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_DmCountry]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_DmState] FOREIGN KEY([StateID])

REFERENCES [dbo].[DimState] ([StateId])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_DmState]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_DmSuburb] FOREIGN KEY([SuburbID])

REFERENCES [dbo].[DimSuburb] ([SuburbId])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_DmSuburb]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_Period] FOREIGN KEY([PeriodKey])

REFERENCES [dbo].[DimPeriod] ([PeriodKey])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_Period]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_School] FOREIGN KEY([SchoolID])

REFERENCES [dbo].[DimSchool] ([SchoolId])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_School]

GO

ALTER TABLE [dbo].[FactAreawiseProperty] WITH CHECK ADD CONSTRAINT [FK\_Transport] FOREIGN KEY([TransportID])

REFERENCES [dbo].[DimTransport] ([TransportId])

GO

ALTER TABLE [dbo].[FactAreawiseProperty] CHECK CONSTRAINT [FK\_Transport]

GO

USE [master]

GO

ALTER DATABASE [Task1\_Warehouse] SET READ\_WRITE

GO