# Property Analysis BI Developer Competition Tasks – Part 1

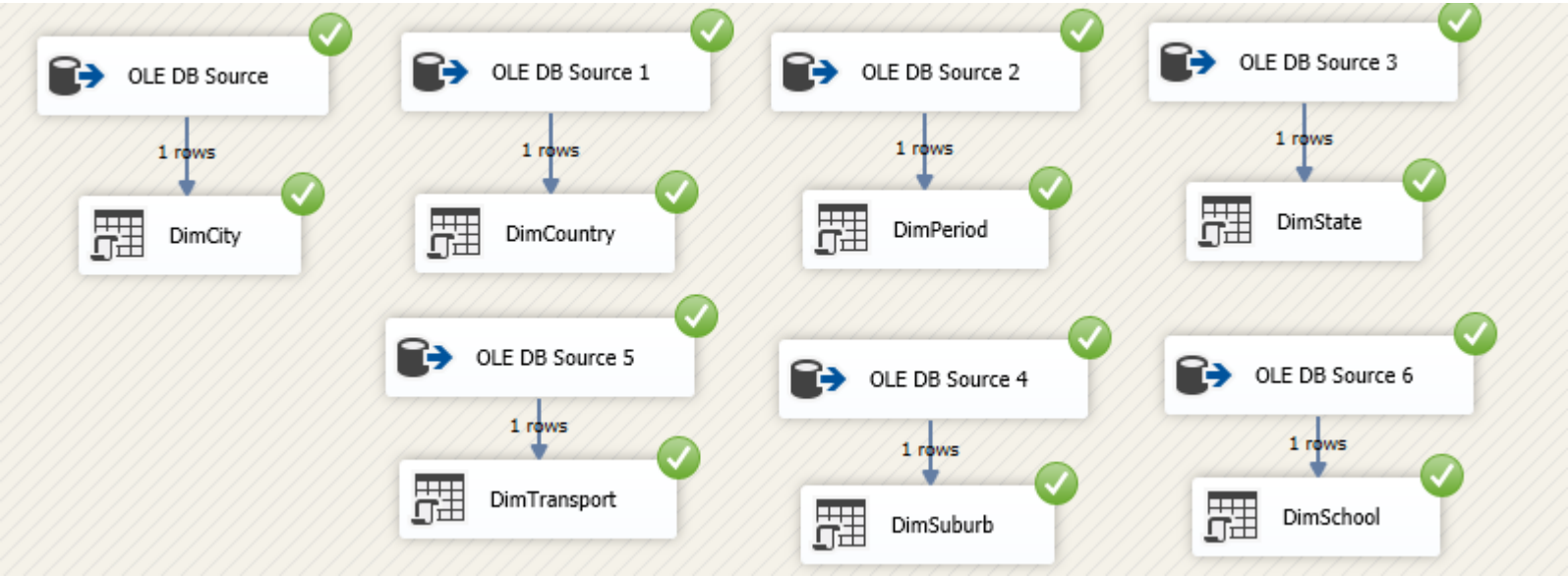
## Part 1

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

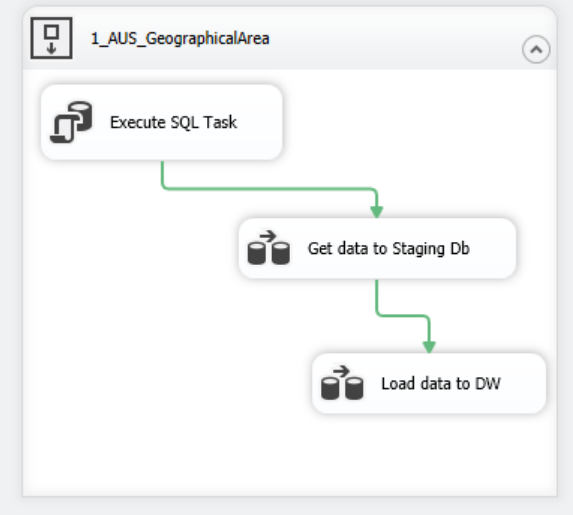
**Solution Approach:**

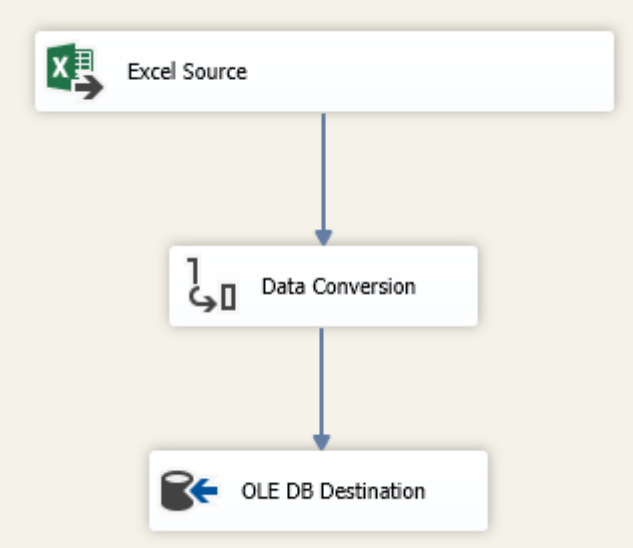
- Do manual simple corrections to the cell values that are catchable to plain sight
- Define staging database structure and warehouse database structure
- Load excel/csv files in to a staging database country-wise
  - Develop SSIS packages to automate the data cleansing and enriching to each country 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.

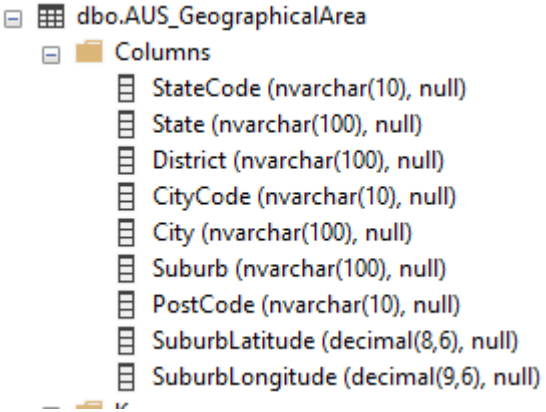
1. Populate default data for each dimension table on the warehouse database(To avoid NULL in the Foreign keys of the Fact Table)



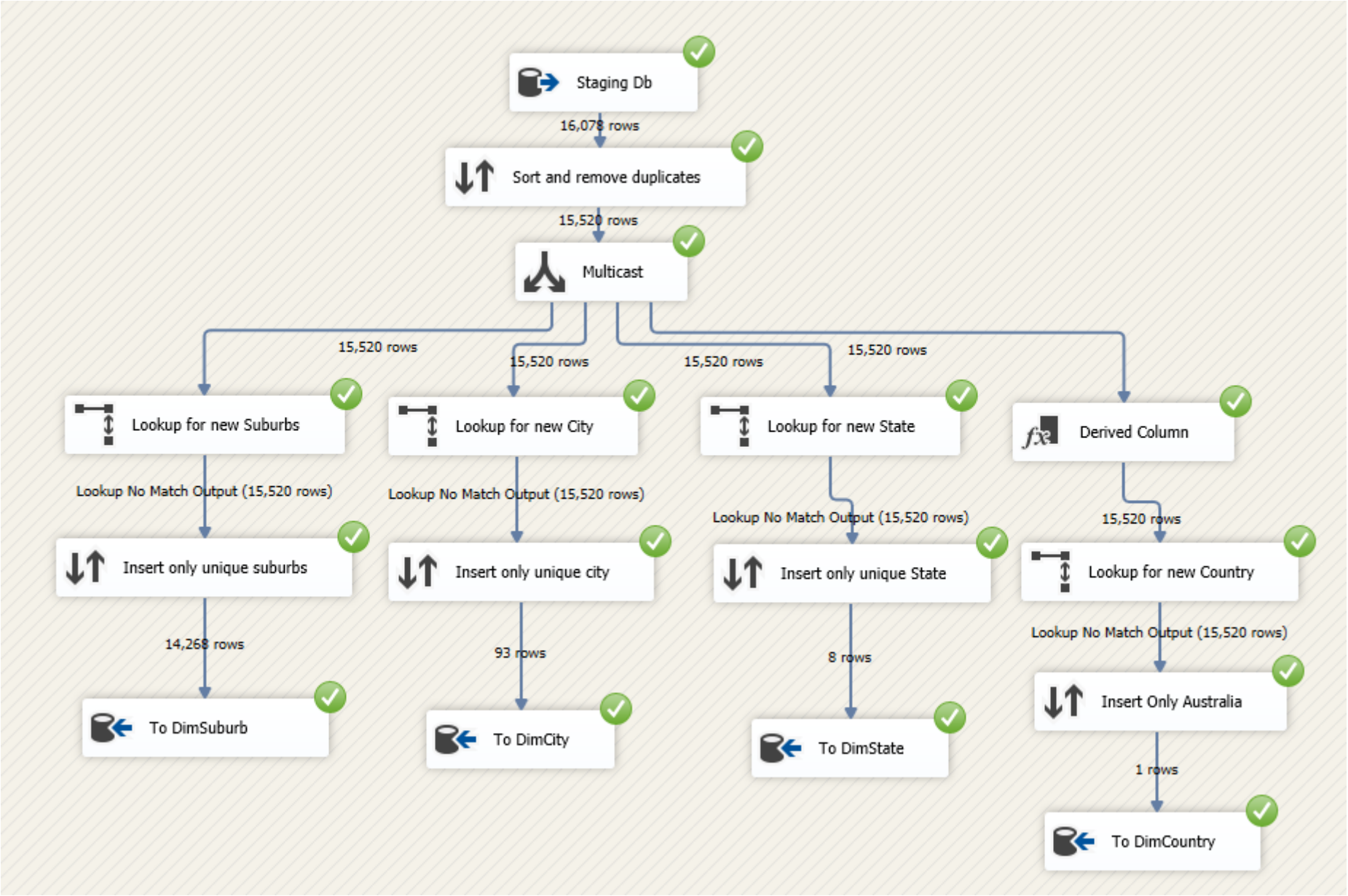
2. Australia Geographical Areas



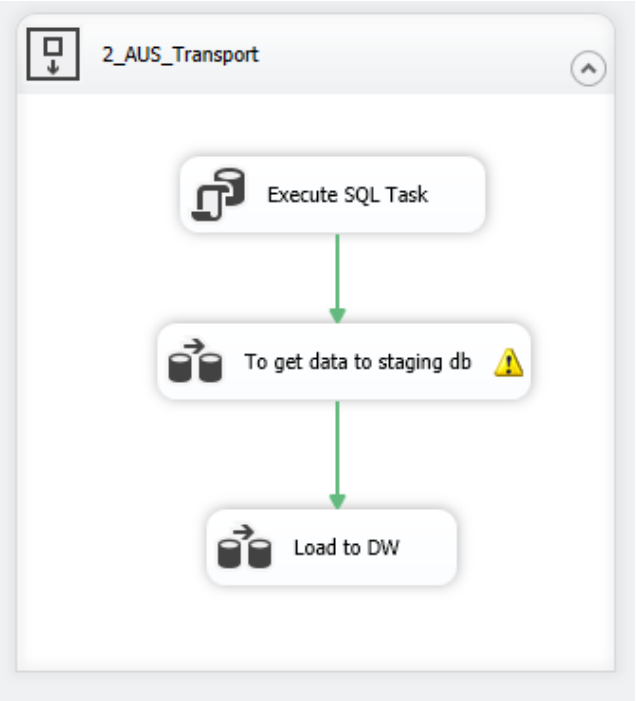



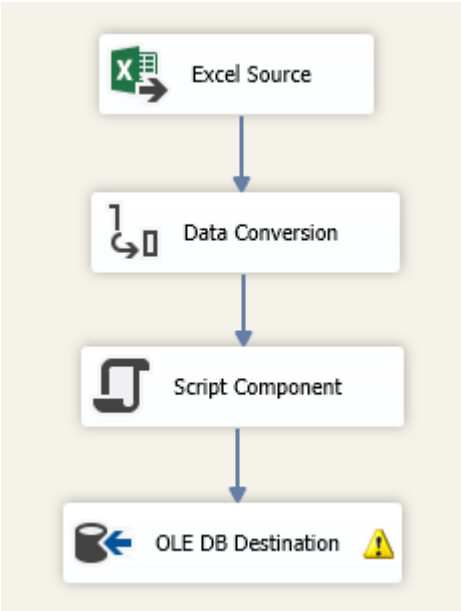



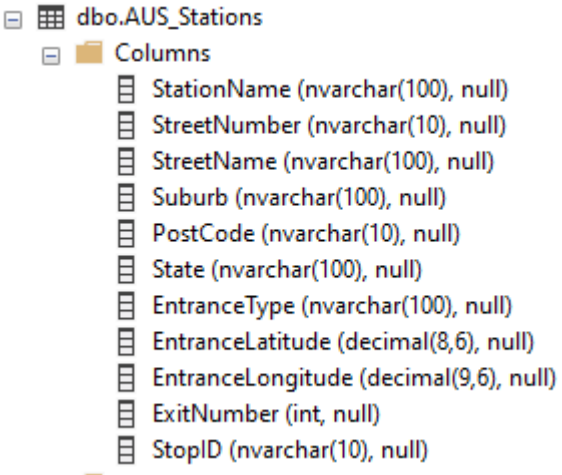

*<Staging Table>*


Staging Db
16,078 rows
Sort and remove duplicates
15,520 rows
Multicast
15,520 rows
15,520 rows
15,520 rows
15,520 rows
Lookup for new Suburbs
Lookup for new City
Lookup for new State
Derived Column
Lookup No Match Output (15,520 rows)
Lookup No Match Output (15,520 rows)
Lookup No Match Output (15,520 rows)
15,520 rows
Insert only unique suburbs
Insert only unique city
Insert only unique State
Lookup for new Country
14,268 rows
93 rows
8 rows
Lookup No Match Output (15,520 rows)
To DimSuburb
To DimCity
To DimState
Insert Only Australia
1 rows
To DimCountry

3. Australia Transport



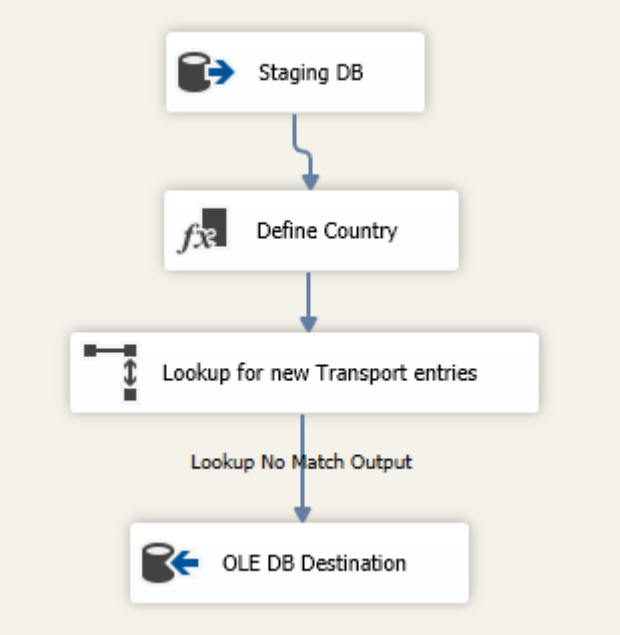



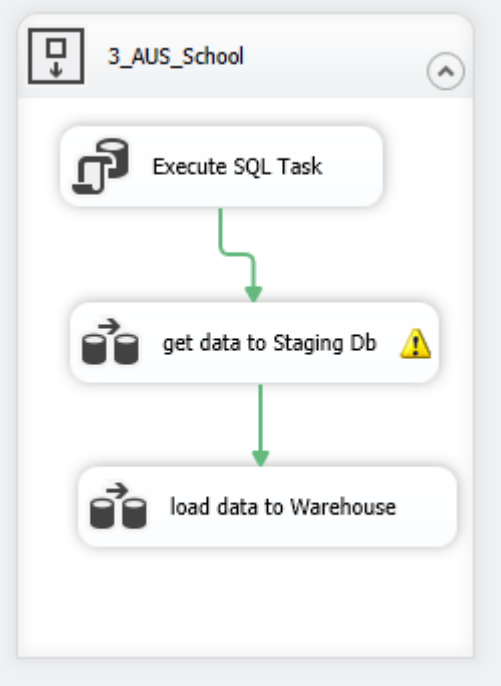

*<Staging Table>*

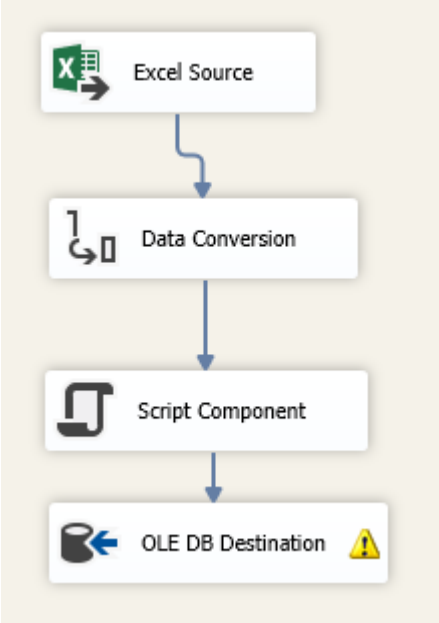
```
// Script to retrieve address breakdown from Google Maps API by using latitude and longitude values
try
{

   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;
```

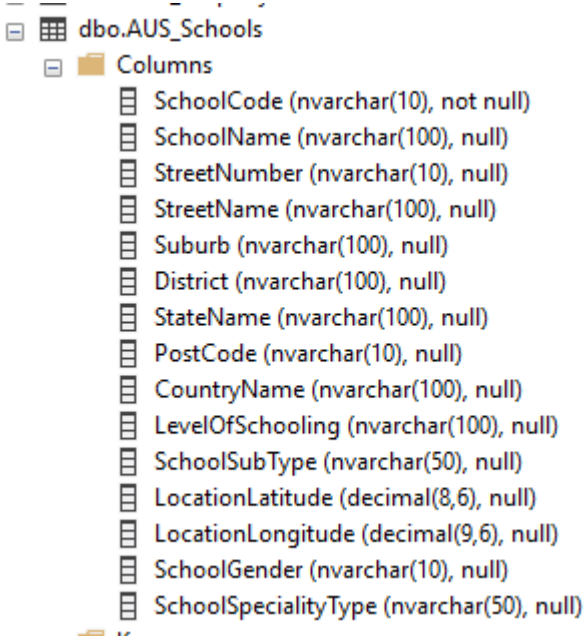
```
            Row.State = result.address_components.Where(c => c.types.Contains("administrative_area_level_1")).FirstOrDefault()?.long_name;
             Row.Country = result.address_components.Where(c => c.types.Contains("country")).FirstOrDefault()?.long_name;
          }
        }
}
catch (Exception ex)
{
   // TODO :; Find how to write into error output
}
```



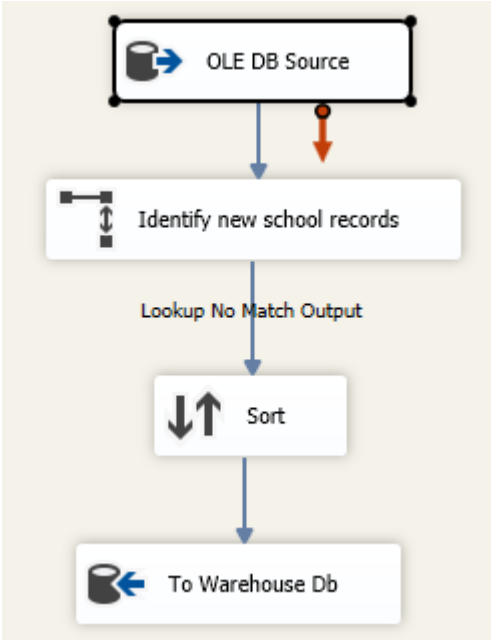
4. Australia Schools



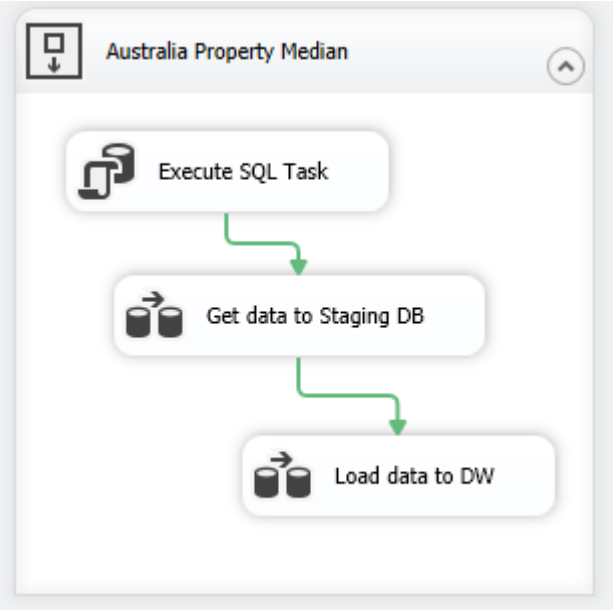



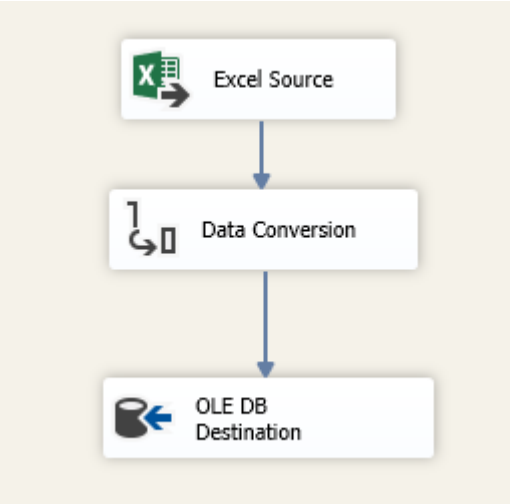



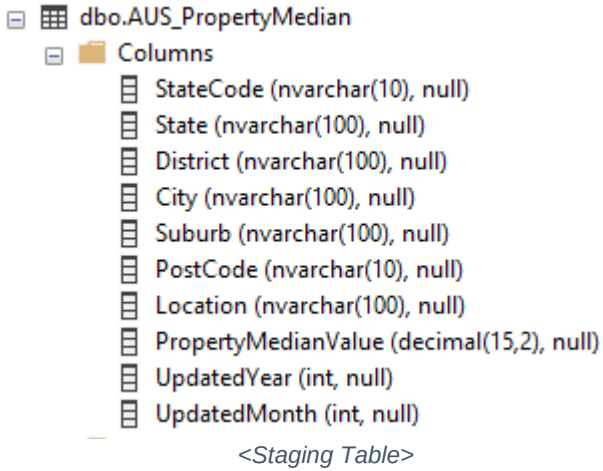

*<Staging Table>*



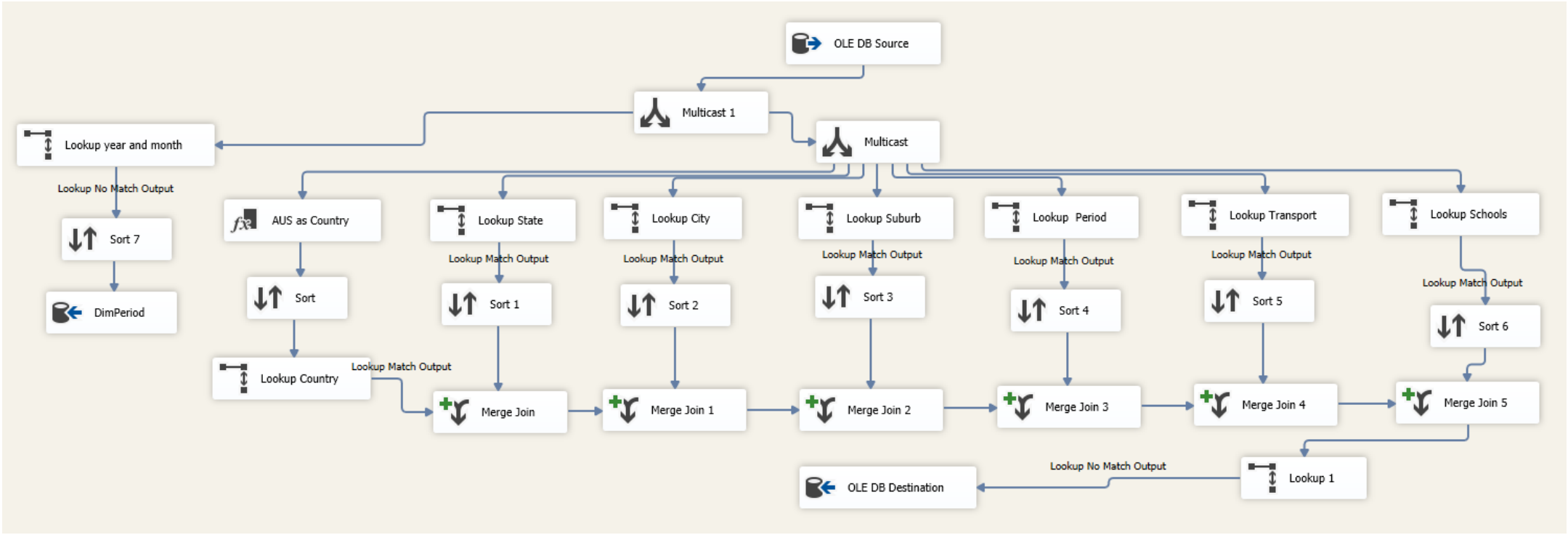
6. Australia Property Median (Loading to DW and Fact Table)

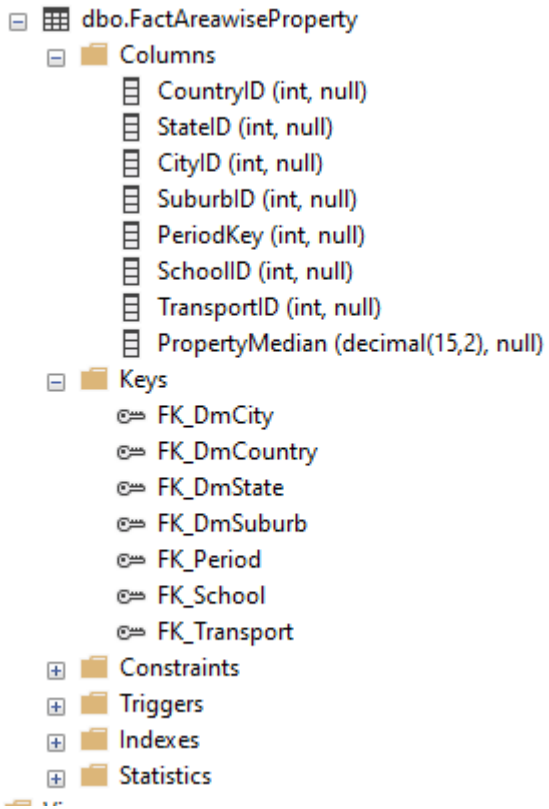


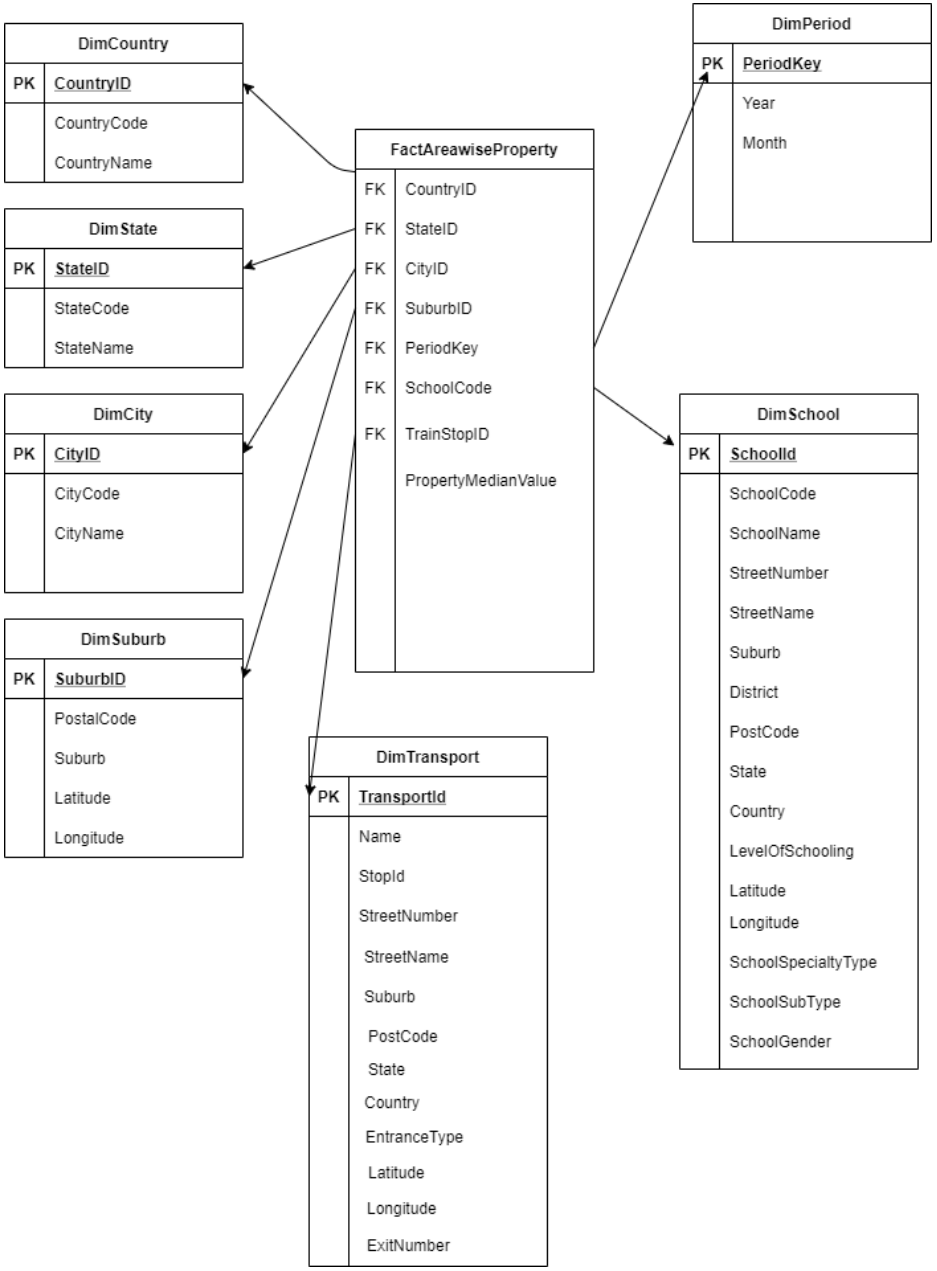

*<Staging Table>*






**Fact Table** - *FacAreawiseProperty*

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




3. 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
```