PROPERTY ANALYSIS -DOCUMENTATION

**Initial Data Loading with SSIS**

**A screenshot of a computer

AI-generated content may be incorrect.**

1. AUS\_SubCityDistrictState\_Data data set is loaded to load\_Aus\_SubCityDistrictState.

*AUS\_SubCityDistrictState\_Data Table design.*

A screenshot of a computer

AI-generated content may be incorrect.

The sequence container Australia Location deals with loading this table.

Truncate the load\_ table before loading to avoid data duplication. – Execute SQL task is used.

ETL Load Geography control flow deals with loading the excel file.

OLE DB Destination is load\_Aus\_SubCityDistrictState table

A screenshot of a computer

AI-generated content may be incorrect.

1. *NSW\_PropertyMedainValue* is loaded into table load\_NSW\_PropertyMedianValue

Table Design

A screenshot of a computer

AI-generated content may be incorrect.

The sequence container NSW Property value deals with loading this table.

Truncate the load\_ table before loading to avoid data duplication. – Execute SQL task is used. A screenshot of a computer

AI-generated content may be incorrect.

ETL Load NSW Property Median Value control flow deals with loading the excel file to destination Load\_NSW Property Median Value table.

1. NSW-Public-Schools-Master-Dataset-07032017 .csv file is loaded to load\_ Load\_Nsw\_Public\_Schools\_Master\_Dataset table

Table design

A screenshot of a computer

AI-generated content may be incorrect.

The sequence container NSW Public Schools deals with loading this table.

Truncate the load\_ table before loading to avoid data duplication. – Execute SQL task is used.

A screenshot of a computer

AI-generated content may be incorrect.

ETL Load NSW PublicSchoolMasterDataset control flow deals with loading the excel file to destination Load\_Nsw\_Public\_Schools\_Master\_Dataset.

Some value in town\_suburb column has data with trailing spaces. Derived column is used to remove those spaces.

A screenshot of a computer

AI-generated content may be incorrect.

Data conversion is used to convert the non-Unicode string type to Unicode string type.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

.

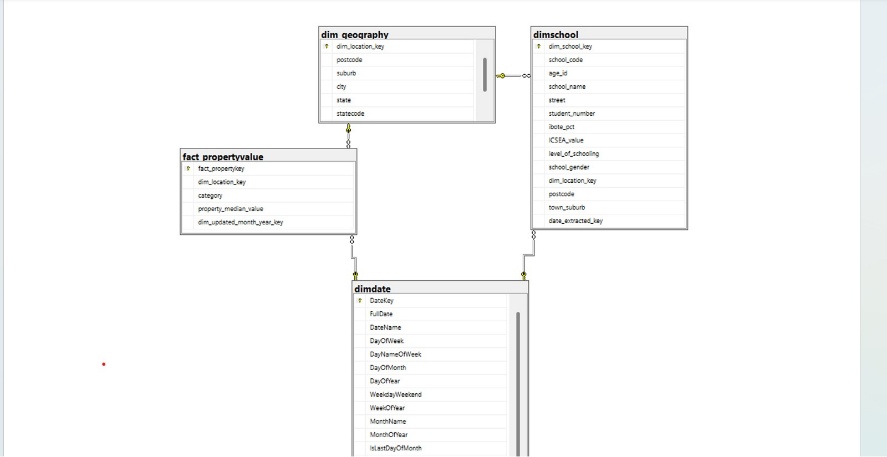
**Dimensional Modeling**

Bus matrix

A screenshot of a computer

AI-generated content may be incorrect.

Snowflake schema is designed.



Dimension Tables - dim\_geography,dimdate,dimschool

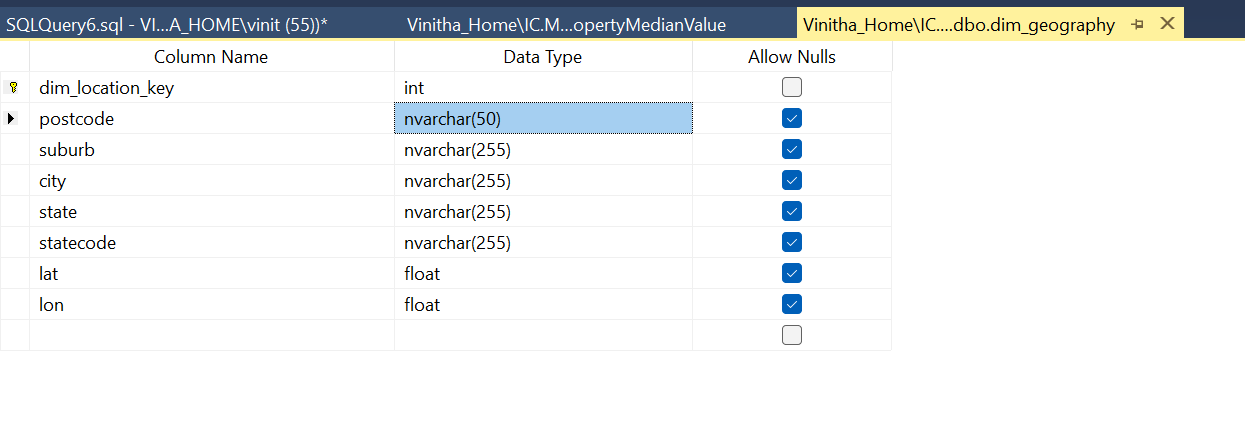
Fact Table - fact\_propertyvalue

**Data Transformation Pipeline.**

Second SSIS package is created to load data to dim and fact tables.

Table Design:

dim\_geography



A screenshot of a computer

AI-generated content may be incorrect.

Identity specification is set to yes for surrogate key.

Dimdate:

A screenshot of a computer

AI-generated content may be incorrect.

dimschool

A screenshot of a computer

AI-generated content may be incorrect.

For surrogate key identity specification is set to yes.

Also dim\_location\_key is created that refers to Dim\_geography table.

And data\_extracted\_key is created that refers to Dim\_date table.

Data Flows:

A screenshot of a computer

AI-generated content may be incorrect.

*Geography Dimension* – To load date from [load\_Aus\_SubCityDistrictState] to dim\_geography table.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Slowly changing Dimension Type 1 used.

A screenshot of a computer

AI-generated content may be incorrect.

*School Dimension – To load data from Load\_Nsw\_Public\_Schools\_Master\_Dataset to DimSchool table.*

A screenshot of a computer

AI-generated content may be incorrect.

All the values are fetched from staging table.

A screenshot of a computer

AI-generated content may be incorrect.

Use derived column to change the format and type of date\_extracted column.

Next Data Conversion is used to convert the data type of age\_id from string to int data type.

Lookup location key step is to map suburb from school data set to dim\_geography dataset and setup location key in dim\_school table.

A screenshot of a computer

AI-generated content may be incorrect.

Lookup data extracted key step is to map the data extracted from public school data set to dimdate table and setup dimdate key in dimschool table.

A screenshot of a computer

AI-generated content may be incorrect.

Next Slowing changing Dimension Type 1 is used.

A screenshot of a computer

AI-generated content may be incorrect.

*Property Fact Table -To load data to fact table.*

A screenshot of a computer

AI-generated content may be incorrect.

Derived column Data Transformation in state and city is used to

* 1. Create new column category in fact table.
  2. To clean data in state column. REPLACE(state,"NEW SOUTH WALES\_NSW","New South Wales").

Some rows has state values as NEW SOUTH WALES\_NSW which is same as New South Wales.

A screenshot of a computer

AI-generated content may be incorrect.

DimLocationKeyLookup -Is used to map suburb and postcode to DimGeography table and get Dimlocation key . A screenshot of a computer

AI-generated content may be incorrect.

Data conversion is used to change the datatype of updated\_year and updated\_month similar to the datatype in dimdate table.

Lookup in Dimdate for updated month and year is used to map to dimdate table and to fetch the date key.

A screenshot of a computer

AI-generated content may be incorrect.

Finaly all the rows are loaded to fact table.