**Table of Content**

[Data Model 2](#_Toc131259170)

[Star Schema 2](#_Toc131259171)

[Tables in SQL database 2](#_Toc131259172)

[Data Flow 3](#_Toc131259173)

[Issues Found 4](#_Toc131259174)

[Result - Estimates business operating hours 5](#_Toc131259175)

# Data Model

I have used star schema in designing the system. Based on the input files, I created the following dimension and fact tables in SQL Server database. Assuming we only want SCD type 1 (If it’s SCD type 2, it will need to change add effective\_from and effective\_to date in the dimension tables).

Dim\_time only contains 15 mins interval, dim\_date only contains year 2017 and 2018 (as the Nmi files only contains 2017 and 2018 data).

## Star Schema

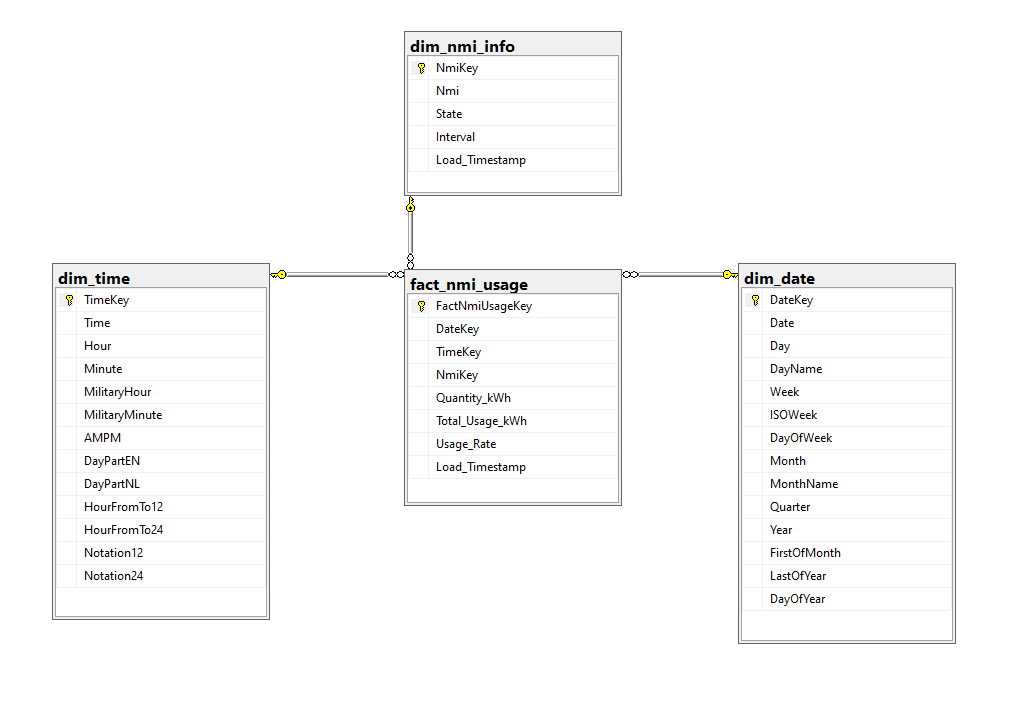
dim\_nmi\_info

fact\_nmi\_usage

dim\_date

dim\_time

## Tables in SQL database



# Data Flow

Run the following steps to create staging, data warehouse, datamart tables and reports.

6). Created 2 reports  
- Total Usage: The total consumption in kWh by Time and Nmi within the date range  
- Daily Average Usage: The average consumption in kWh by Time and Nmi within the date range

Power BI (shell\_nmi\_usage.pbix)

5). Transform and load data to dimension, data warehouse and fact tables  
- transform and load data from staging tables to dimension tables  
- transform and load data from staging tables to dw table  
- transform and load data from dw and dimension tables to fact table dbo.fact\_nmi\_usage (convert quantity in different units to kwh)

3). Extract all nmi source files in folder ConsumptionData based on the nmi names in 1).  
 - perform data wrangling  
 - output the error into a list (should have written to a log file or table)

2). Extract source file nmi\_info.csv

1). Run DDL scripts to create staging, data warehouse, dimension and fact tables

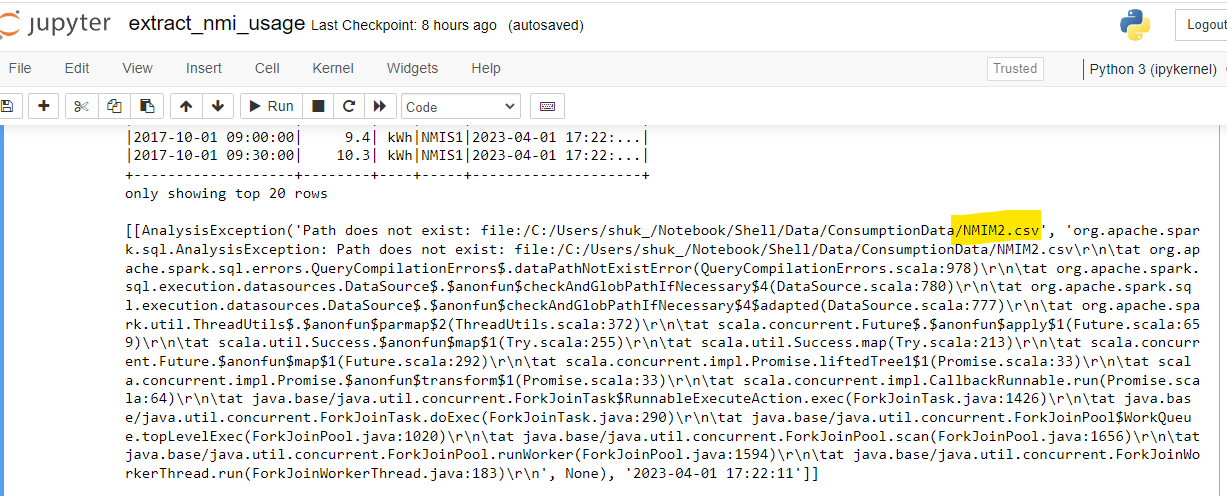
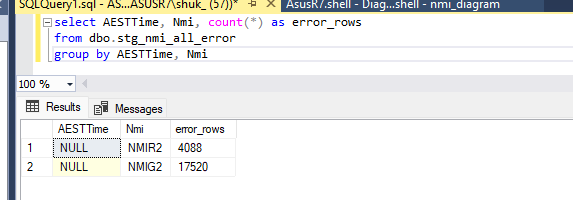
4). Load the data to staging tables in SQL Server database  
- load data to dbo.stg\_nmi\_info, any error route the problem row to a reject/error table dbo.stg\_nmi\_info\_error  
- load data to dbo.stg\_nmi\_all, any error route the problem row to reject/error table dbo. stg\_nmi\_all\_error

Jupyter Notebook (2\_extract\_nmi\_usage.ipynb)

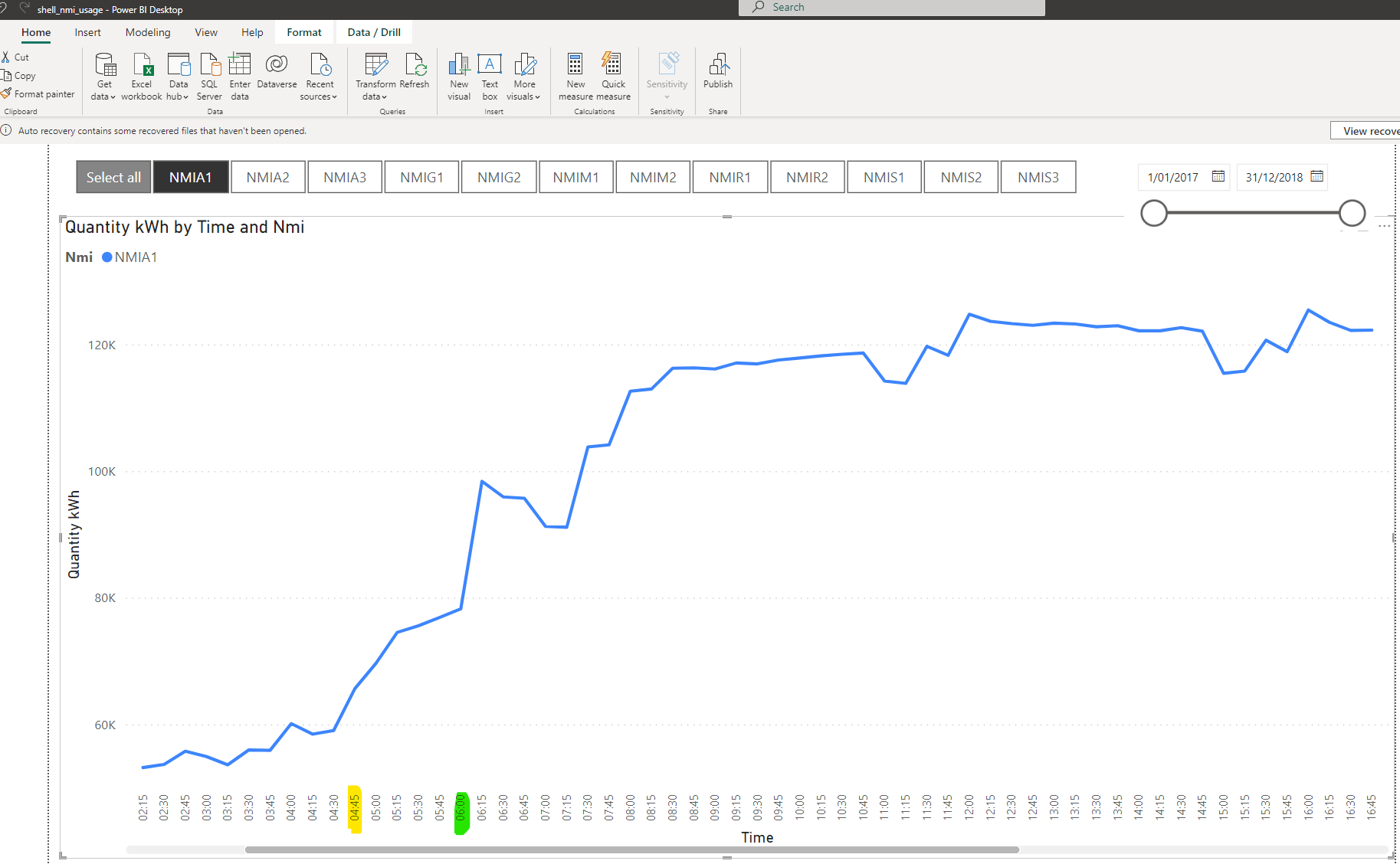
Microsoft SQL Server Developer (1\_create\_tables\_ddl.sql)

Microsoft SQL Server Developer (scritps =3\_transform\_load\_tables.sql, backup file =shell.bak)

# Issues Found

1. Couldn’t find the Nmi csv file named “NMIG2” and “NMIM2”, the names are existed in nmi\_info.csv. Caught the following exception while extracting the nmi csv files.  
   
2. A lot of NULL values of AESTTime are found in csv file “NMIR2” and “NMIG2”.  
   

# Result - Estimates business operating hours

The Power BI reports show each Nmi’s total/average consumption in kWh, it allows the user to filter the date range. By just looking at a particular site i.e. NMIA1, it looks like after 04:45 am, the energy consumption starts to increase. And by 06:00am, it has a sharp rise. My assumption is, staff started to arrive the site around 04:45am and the site officially opens at 6am.

It drops dramatically at 20:30 and not much consumption from 22:30 onward. My guess is the site closes at 20:30. There may have a few staff stay to finish up the work so the consumption is relative low between 20:30 and 22:30. Graphical user interface, chart, application, line chart

Description automatically generated

# Screenshot of the Power BI Reports

Graphical user interface, chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated