



Sri Lanka Institute of Information Technology

Data Warehousing and Business Intelligence

Assignment 2

2021

Name: Kulasekara K.M.S.B

Registration No: IT19129068

Table of Contents

1.	Data source for the Assignment 2	5
1.1	Original Data Source	5
1.2	Data Warehouse	6
	Data Preparation	6
	Solution Architecture.....	6
	Data Warehouse Design.....	6
2.	SSAS Cube implementation	8
2.1	Creating the Data Source	8
2.2	Creating a Data Source View	8
2.3	Creating the Cube	9
	Dimension Structures	10
	Dimension Usage	12
	KPIs	12
3.	Demonstration of OLAP operations.....	14
3.1	Roll-up and Drill-down.....	14
	Roll-up	14
	Drill-down.....	14
3.2	Slice and Dice.....	15
	The MDX Query	15
	The original view of data.....	15
	Slice	16
	Dice.....	16
3.3	Pivot.....	17
	Before pivot.....	17
	After Pivot.....	18
4.	SSRS Reports.....	19
4.1	Report with a matrix	19
	Data set	19
	Report	20
4.2	Drill-down report.....	20
	Dataset	20

Report	20
4.3 Drill-through report	22
Dataset	22
Report	23
4.4 Report with two parameters	25
Dataset	25
Report	27

Table of Figures

Figure 1-1ER Diagram	5
Figure 1-2 High Level Architectural diagram	6
Figure 1-3 Dimensional Model.....	7
Figure 2-1 SSAS Project Hierarchy	8
Figure 2-2 Data Source View	9
Figure 2-3 The Cube	10
Figure 2-4 Dimension Structure (Dim Region Details)	10
Figure 2-5 Dimension Structure (Dim Region Coordinates)	11
Figure 2-6 Dimension Structure (Dim Date)	11
Figure 2-7 Dimension Usage	12
Figure 2-8 KPIs (KPI Hydro Difference).....	12
Figure 2-9 KPIs (KPI Nuclear Difference)	13
Figure 2-10 KPIs (KPI Thermal Difference).....	13
Figure 3-1 Roll-up.....	14
Figure 3-2 Drill-down	14
Figure 3-3 MDX Query (Slice, Dice)	15
Figure 3-4 Original View	15
Figure 3-5 Slice.....	16
Figure 3-6 Dice	16
Figure 3-7 Pivot (Before pivot)	17
Figure 3-8 PivotChart Fields (Before pivot)	17
Figure 3-9 Pivot (After pivot).....	18
Figure 3-10 PivotChart Fields (After pivot).....	18
Figure 4-1 Report with matrix (Query)	19
Figure 4-2 Report with a matrix	20
Figure 4-3 Drill-down report (before).....	20
Figure 4-4 Drill-down Report (After)	21
Figure 4-5 Drill-through second report (report data structure).....	22
Figure 4-6 Query for Region_names dataset.....	22
Figure 4-7 Query for DataSet1	23
Figure 4-8 Drill-through (First report)	23
Figure 4-9 Drill-through (Second report).....	24
Figure 4-10 Report with two parameters (Report data structure).....	25
Figure 4-11 Query for Regions dataset	25
Figure 4-12 Query for Years dataset.....	26
Figure 4-13 Query for DataSet1 dataset.....	26
Figure 4-14 Report with two parameters.....	27

1. Data source for the Assignment 2

1.1 Original Data Source

Data Source - Daily Power Generation in India (2017-2020)

Description –

The data set consisted of three tables.

1. Daily Power Generation – Region-wise daily power generation in India. (Hydro, Thermal and Nuclear)
2. Region Details – Data such as Region name, Area, National Share and House Hold Size.
3. Region Coordinates – Geographical data of the regions.

ER Diagram –

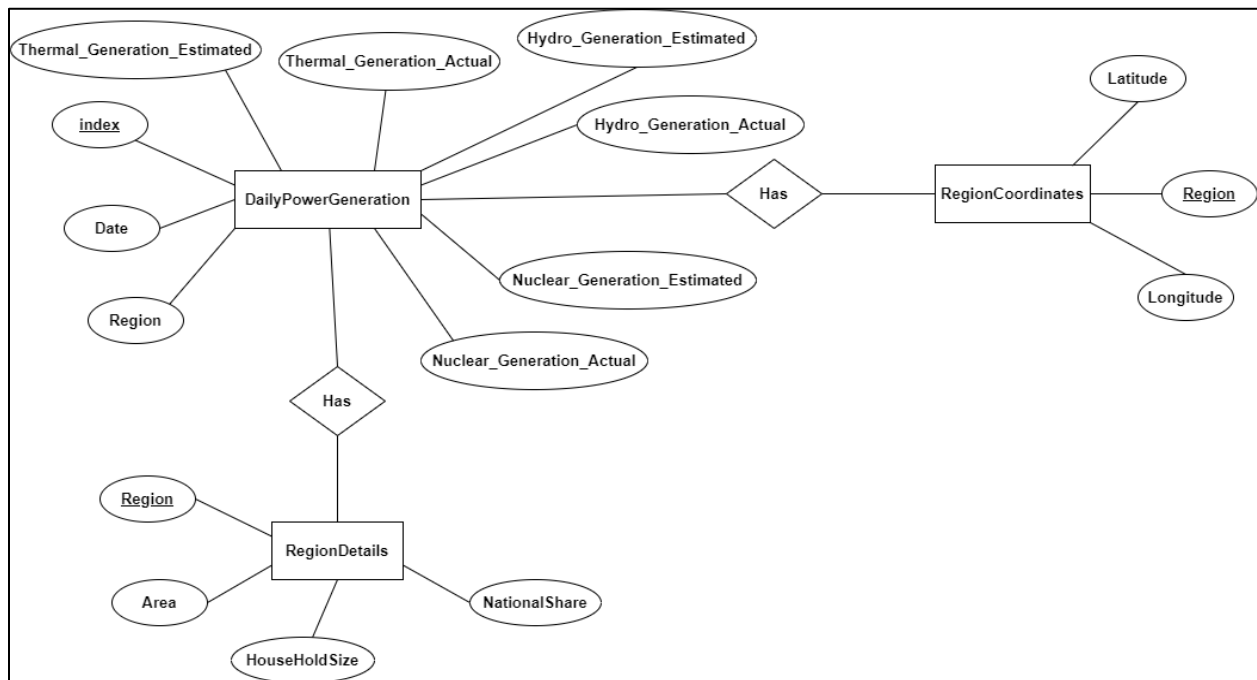


Figure 1-IER Diagram

1.2 Data Warehouse

Data Preparation

The data sources were converted in to three different types.

1. region_coordinates.csv
2. RegionDetails.txt
3. PowerGeneration database file

Solution Architecture

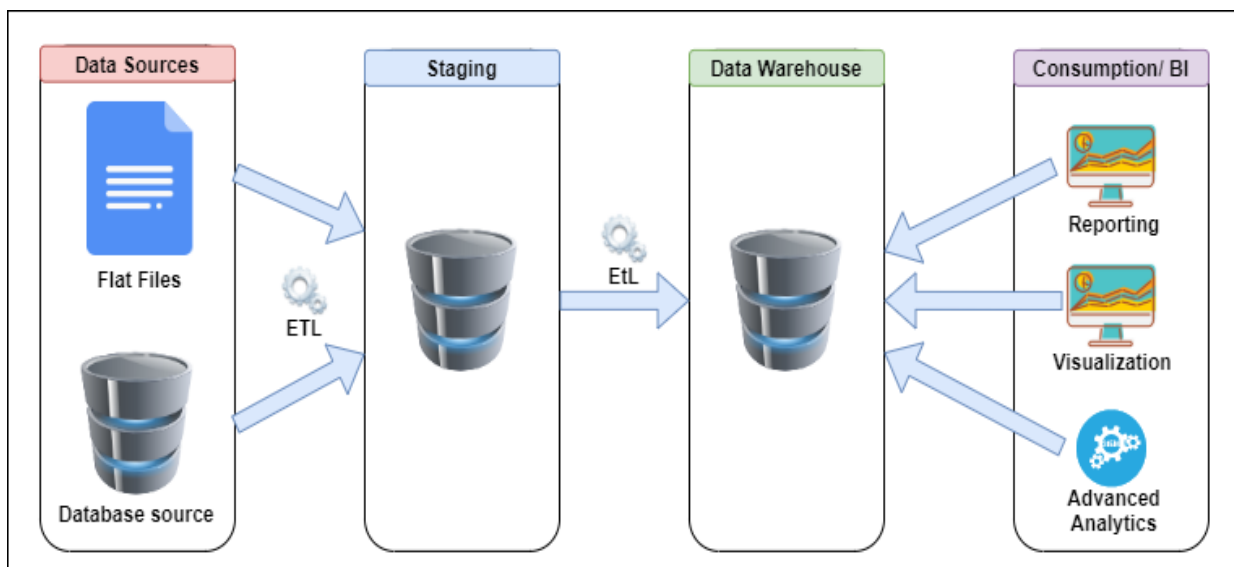


Figure 1-2 High Level Architectural diagram

Data Warehouse Design

The dimensional model is designed based on the concepts of the star schema. There are three dimensions (DimRegionDetails, DimRegionCoordinates, DimDate) and one fact table (FactPowerGeneration) used to design the model.

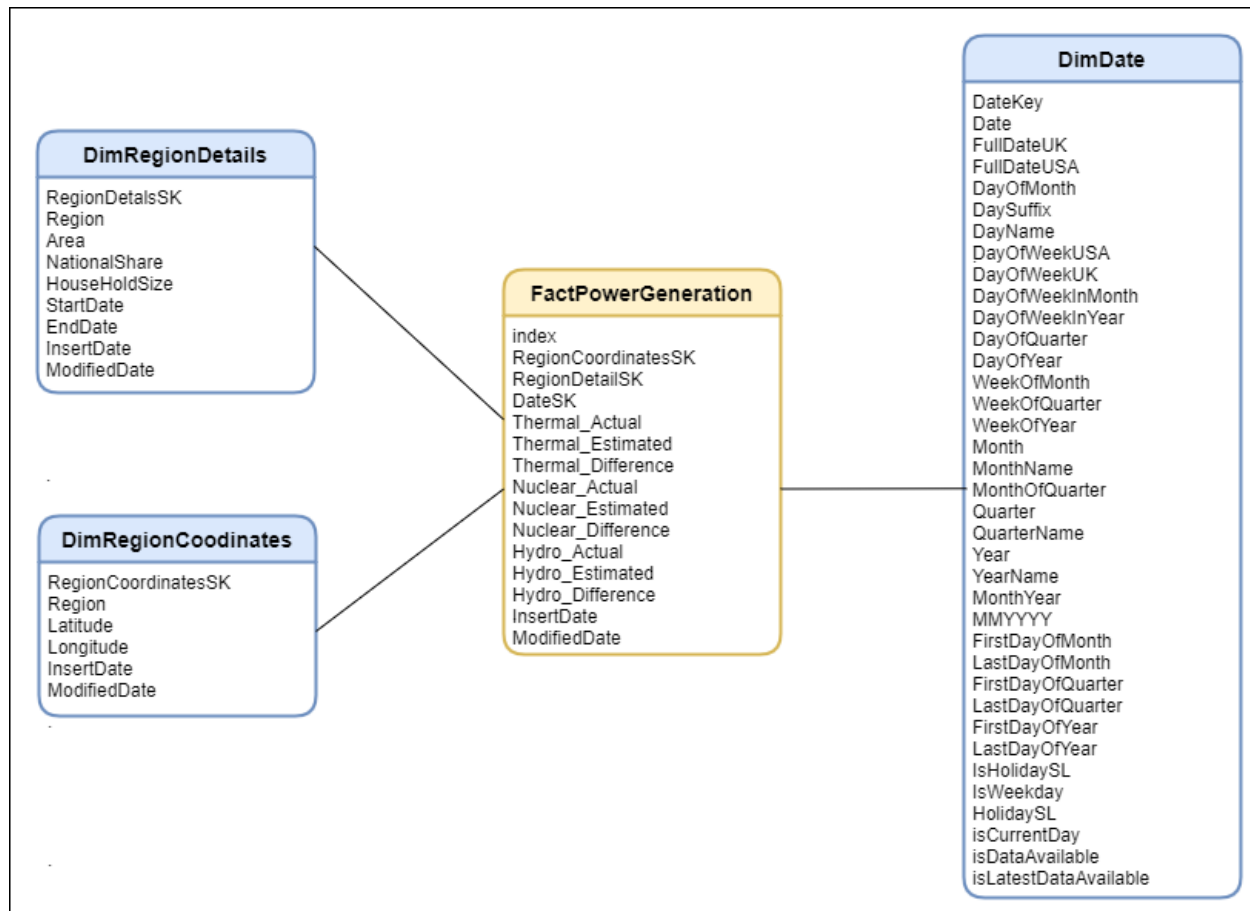


Figure 1-3 Dimensional Model

The DimRegionDetails dimension which is considered as the Slowly Changing Dimension consists of nine attributes with RegionDetailsSK being the Surrogate Key. This has been designed as a Type Two slowly changing dimension.

The DimDate dimension consists of thirty-seven attributes. The DateKey has been used as the Surrogate Key.

The coordinates of each region have been included in the DimRegionCoordinates dimension. The RegionCoordinatesSK has been used as the Surrogate Key.

There are fifteen attributes in the FactPowerGeneration table. This table consists of the daily power generation data in India's three main power generation sources. The Thermal_Difference, Nuclear_Difference and Hydro_Difference are derived attributes that contain the difference between actual and estimated power generation.

2. SSAS Cube implementation

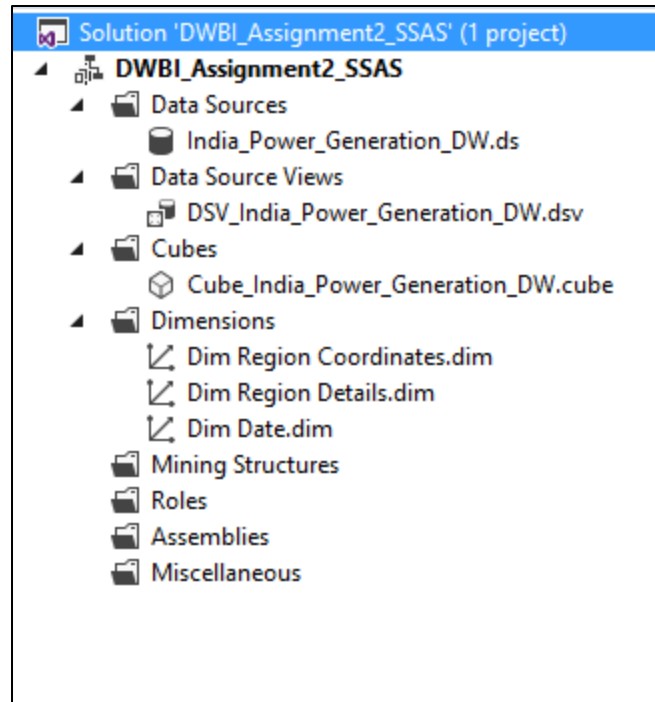


Figure 2-1 SSAS Project Hierarchy

The SSAS project was named as DWBI_Assignment2_SSAS.

2.1 Creating the Data Source

- In the Data Source Wizard, the connection that was created previously in SSIS could be seen and the same was selected.
- In the Impersonation Information page, the first option; **Use a specific Windows User name and Password**, was selected.
- Finally, the data source name was provided as 'India Power Generation DW'.

2.2 Creating a Data Source View

- In the Data Source Wizard, the data source that was created in the previous step was selected.

- In the **Name Matching** page, the **Same name as primary key** option was selected under **Create logical relationships by matching columns**.
- In the next step, the dimension tables and fact table was selected.
- The data source view was named as 'DSV_India_Power_Generation'.

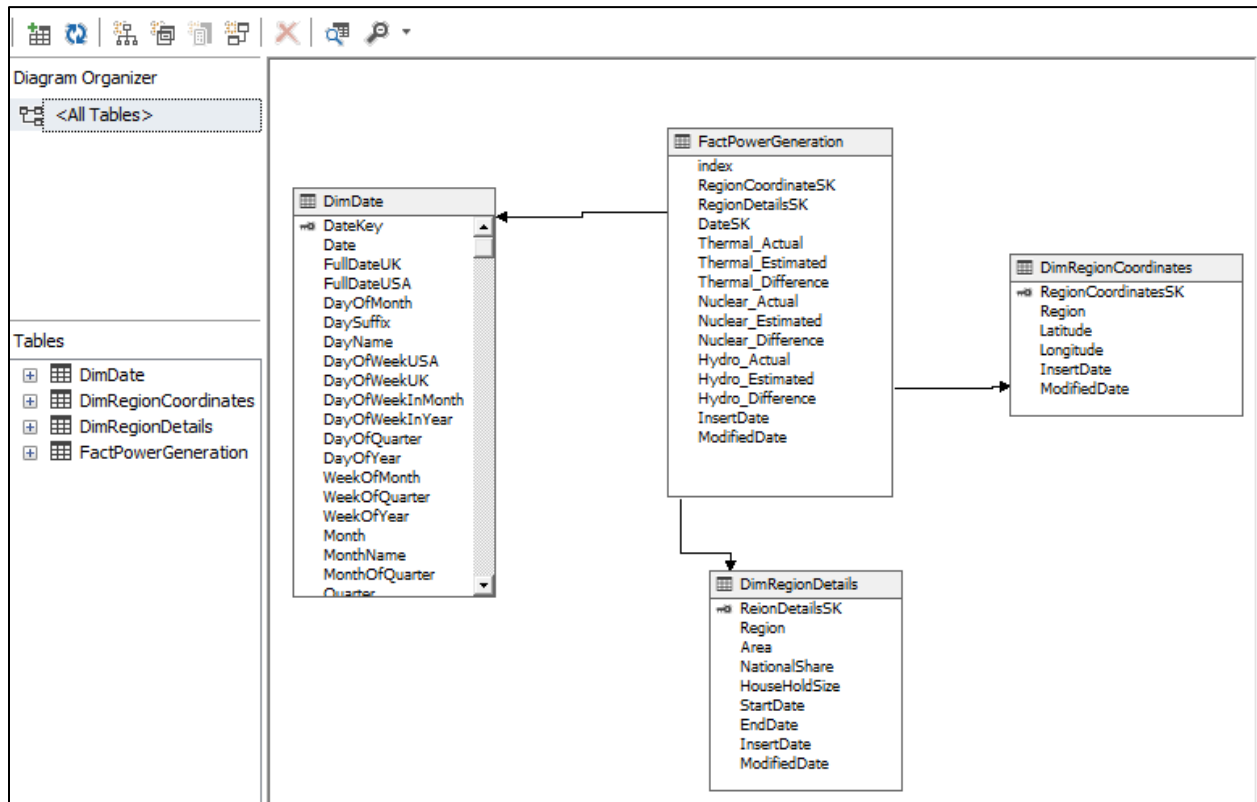


Figure 2-2 Data Source View

- As the final step, the tables were linked accordingly using the surrogate keys.

2.3 Creating the Cube

- In Cube Wizard, **Use existing tables** option was selected in the **Select Creation Method Page**.
- In the **Select Measure Group Tables** page, 'FactPowerGeneration' table was selected as the fact table.
- In the Select Measures page, all the fields were selected.
- The cube was named as 'Cube_India_Power_Generation'.

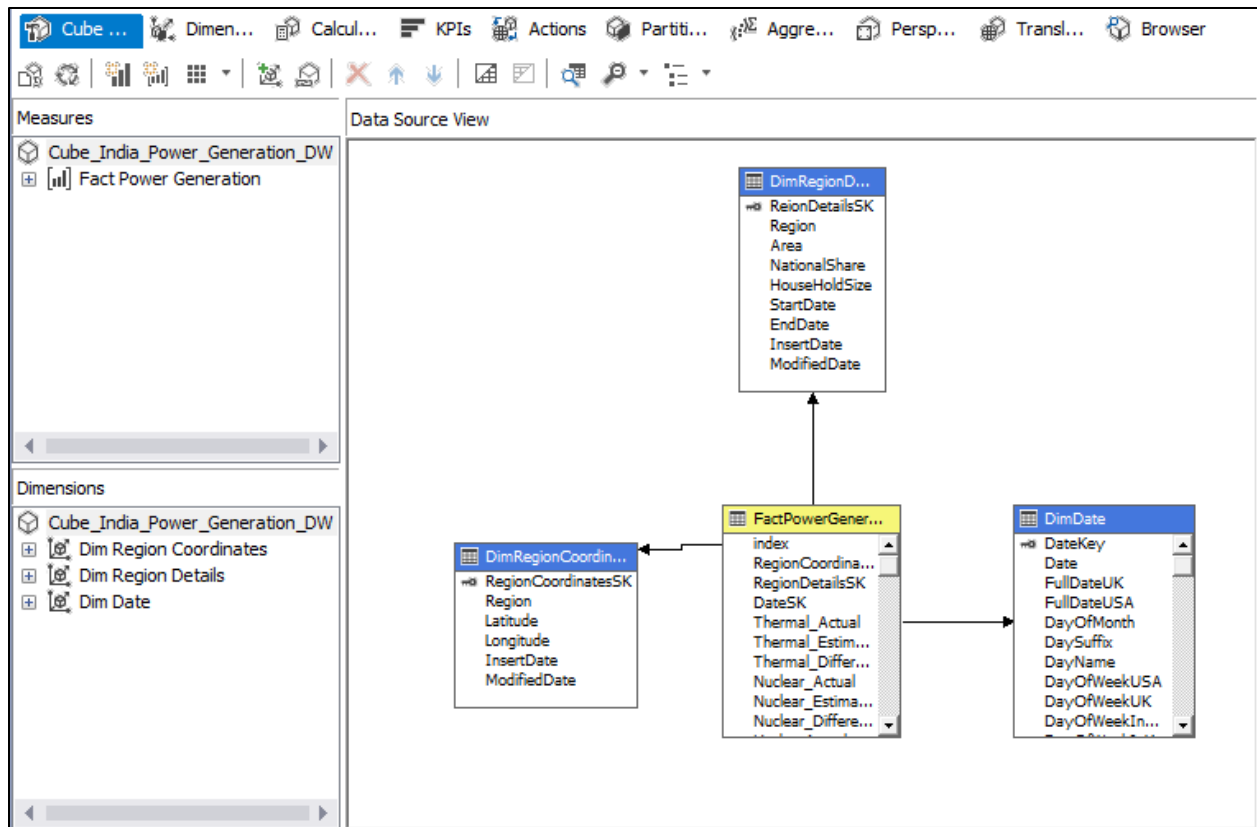


Figure 2-3 The Cube

Dimension Structures

- Dim Region Details –

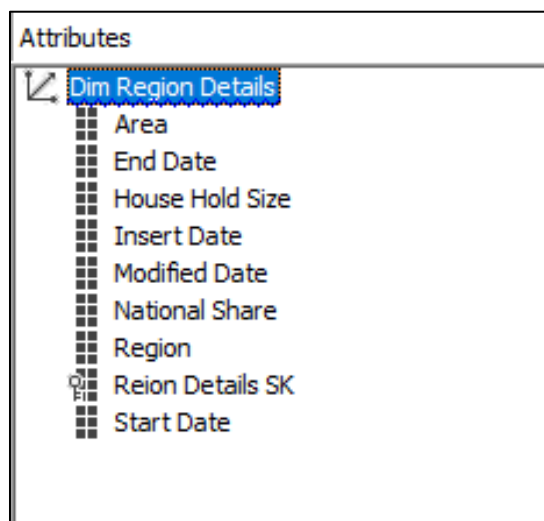


Figure 2-4 Dimension Structure (Dim Region Details)

- Dim Region Coordinates –

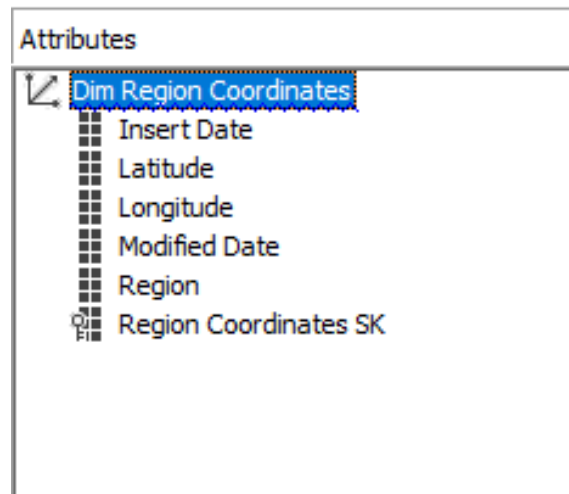


Figure 2-5 Dimension Structure (Dim Region Coordinates)

- Dim Date –

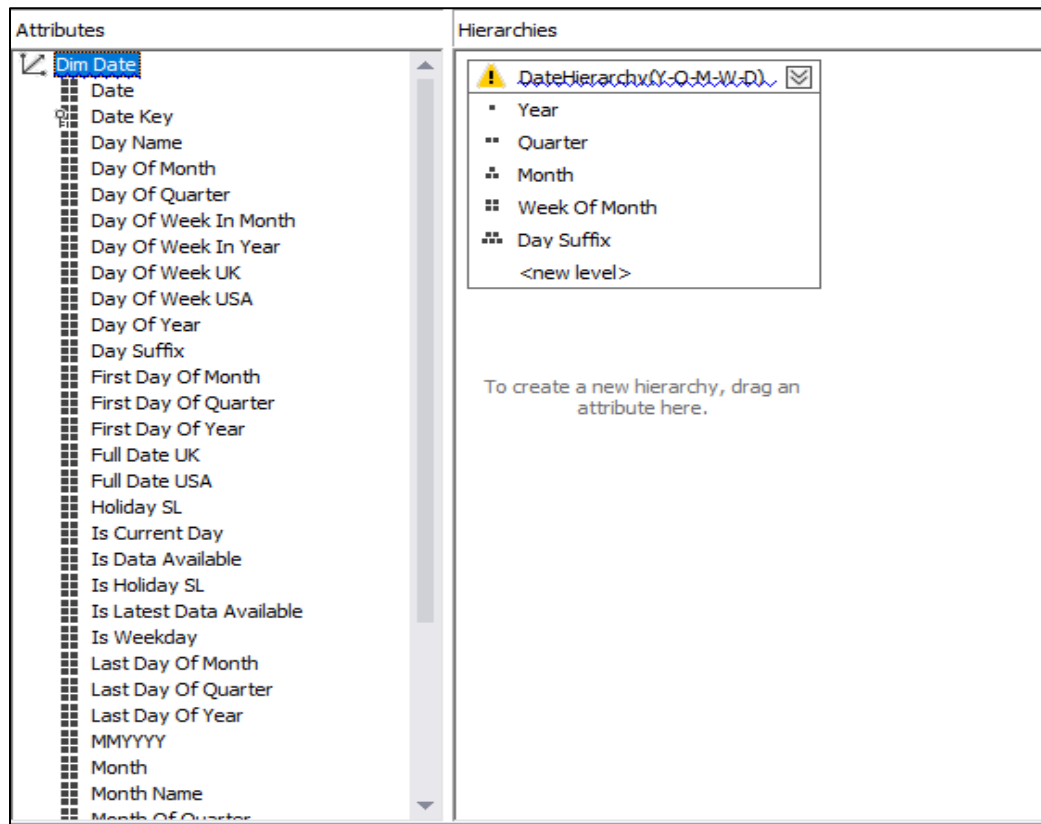


Figure 2-6 Dimension Structure (Dim Date)

The Date Hierarchy: Year → Quarter → Month → Week of Month → Day Suffix

Dimension Usage

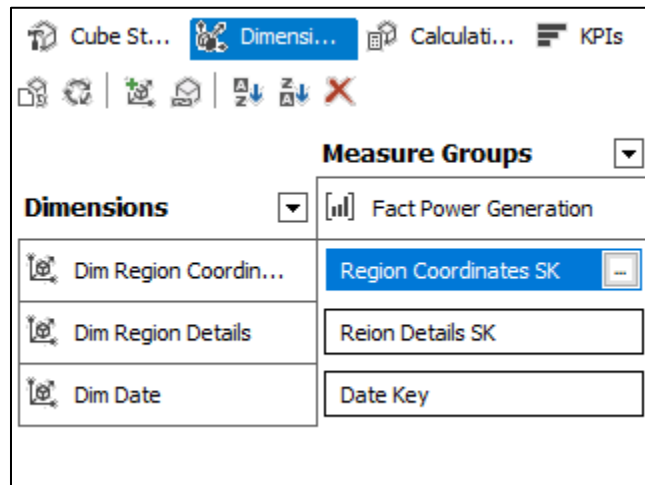


Figure 2-7 Dimension Usage

KPIs

KPI Hydro Difference –

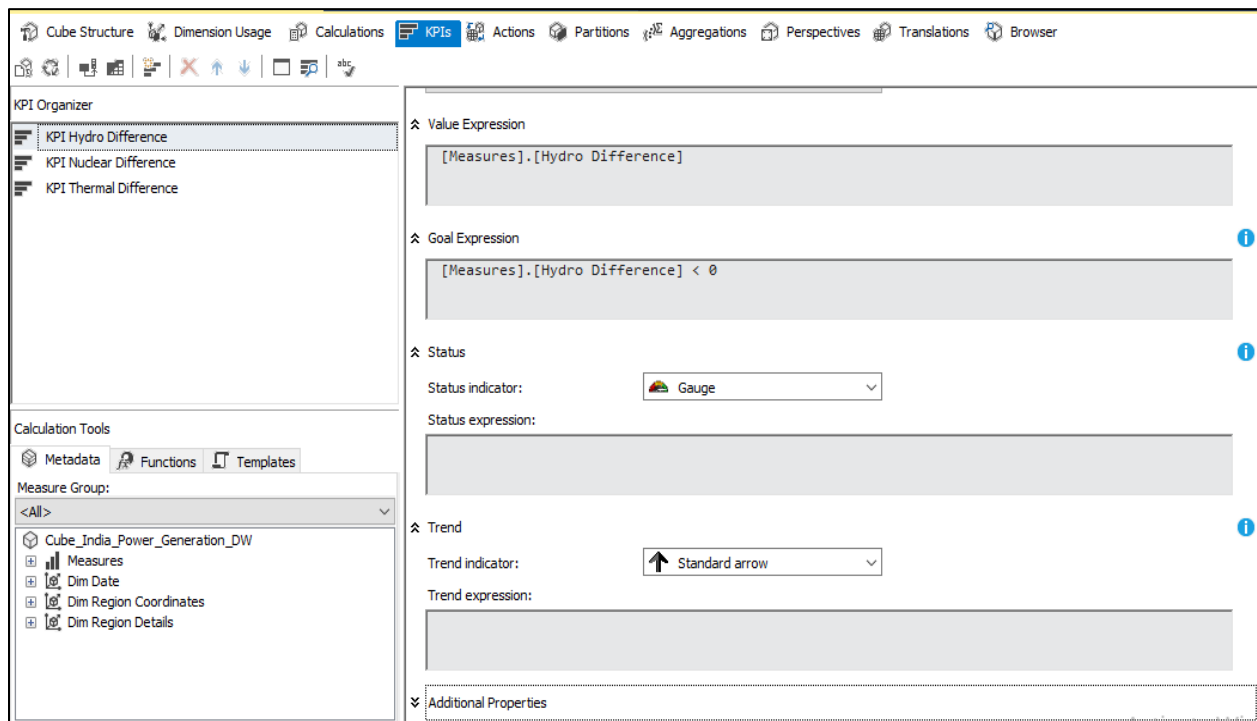


Figure 2-8 KPIs (KPI Hydro Difference)

KPI Nuclear Difference –

The screenshot shows the 'KPIs' configuration window in a software application. The 'KPI Organizer' on the left lists 'KPI Nuclear Difference' as the selected item. The main configuration area on the right is titled 'Name: KPI Nuclear Difference'. It includes an 'Associated measure group' dropdown set to 'Fact Power Generation'. The 'Value Expression' field contains '[Measures].[Nuclear Difference]'. The 'Goal Expression' field contains '[Measures].[Nuclear Difference] < 0'. The 'Status' section has a 'Status indicator' dropdown set to 'Gauge' and an empty 'Status expression' field. The 'Trend' section has a 'Trend indicator' dropdown set to 'Standard arrow' and an empty 'Trend expression' field. Information icons (i) are present next to the Goal Expression, Status, and Trend sections.

Figure 2-9 KPIs (KPI Nuclear Difference)

KPI Thermal Difference –

The screenshot shows the 'KPIs' configuration window for 'KPI Thermal Difference'. The 'KPI Organizer' on the left lists 'KPI Thermal Difference' as the selected item. The main configuration area on the right is titled 'Name: KPI Thermal Difference'. It includes an 'Associated measure group' dropdown set to '<All>'. The 'Value Expression' field contains '[Measures].[Thermal Difference]'. The 'Goal Expression' field contains '[Measures].[Thermal Difference] < 0'. The 'Status' section has a 'Status indicator' dropdown set to 'Gauge' and an empty 'Status expression' field. The 'Trend' section has a 'Trend indicator' dropdown set to 'Standard arrow' and an empty 'Trend expression' field. Information icons (i) are present next to the Goal Expression, Status, and Trend sections.

Figure 2-10 KPIs (KPI Thermal Difference)

3. Demonstration of OLAP operations

3.1 Roll-up and Drill-down

For this, the excel worksheet has been connected to the cube using the Data tab and Hydro Actual, Nuclear Actual and Thermal Actual data have been selected as the Values.

Roll-up

Row Labels	Hydro Actual	Nuclear Actual	Thermal Actual
⊕ Eastern	43271.72	0	435854.26
⊕ NorthEastern	15848.05	0	28902.57
⊕ Northern	166166.16	24702.13	591154.2
⊕ Southern	65360.05	50478.23	556719.91
⊕ Western	33632.11	24064.31	1093065.45
Grand Total	324278.09	99244.67	2705696.39

Figure 3-1 Roll-up

Drill-down

Row Labels	Hydro Actual	Nuclear Actual	Thermal Actual
⊖ Eastern			
⊕ 2017	6199.53	0	56600.55
⊖ 2018			
⊖ 1			
⊖ February			
⊖ 1			
Friday	25.1	0	489.39
Saturday	25.1	0	485.39
Thursday	25.1	0	489.39
⊖ 2			
Friday	25.1	0	488
Monday	25.1	0	485.39
Saturday	25.1	0	488
Sunday	25.1	0	485.39
Thursday	25.1	0	488
Tuesday	25.1	0	488
Wednesday	25.1	0	488
⊖ 3			
Friday	25.1	0	490.85
Monday	25.1	0	488
Saturday	25.1	0	490.85
Sunday	25.1	0	488
Thursday	25.1	0	488
Tuesday	25.1	0	488

Figure 3-2 Drill-down

Drill-down can be done along **Year → Quarter → Month → Week → Day** for a selected Region.

3.2 Slice and Dice

For this, the excel worksheet has been connected to the cube using the Power Pivot mode.

The Region, Date Hierarchy, Hydro Actual, Nuclear Actual and Thermal Actual fields have been selected for this operation.

The MDX Query

```
SELECT NON EMPTY { [Measures].[Hydro Actual], [Measures].[Nuclear Actual], [Measures].[Thermal Actual] } ON COLUMNS, NON  
EMPTY { ([Dim Region Details].[Region].[Region].ALLMEMBERS * [Dim Date].[DateHierarchy(Y-Q-M-W-D)].[Day Suffix].ALLMEMBERS ) }  
DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Cube_India_Power_Generation_DW] CELL  
PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Figure 3-3 MDX Query (Slice, Dice)

The original view of data

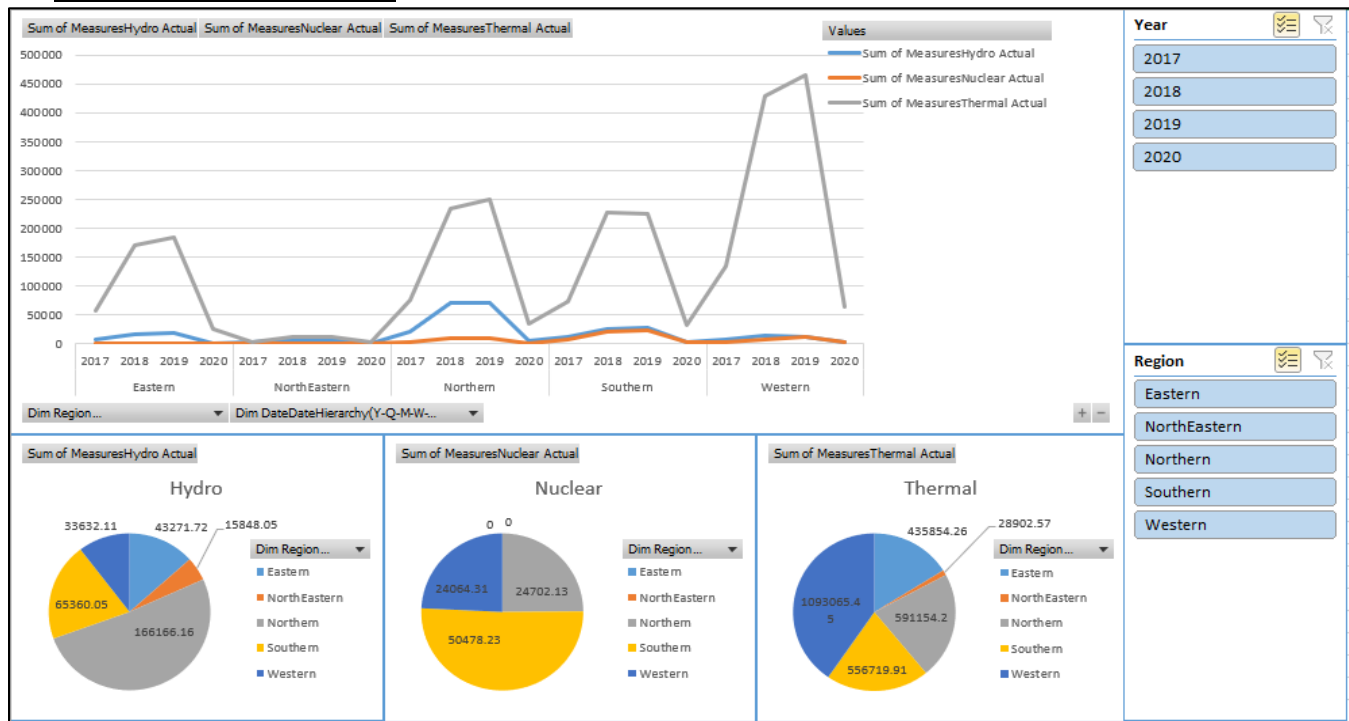


Figure 3-4 Original View

Slice

As shown in the *Figure 3-5*, the cube has been sliced to display data for year 2017.

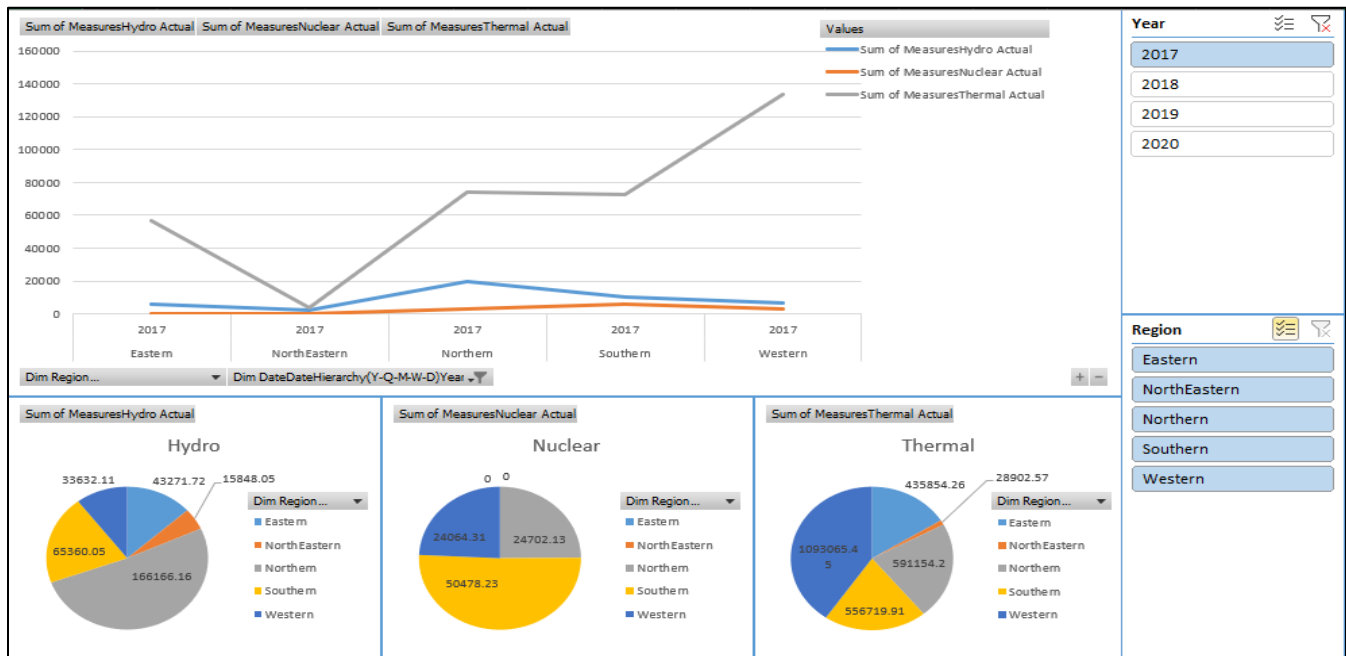


Figure 3-5 Slice

Dice

As shown in the *Figure 3-6*, the cube is diced to get Hydro Difference data for years 2017 and 2018 in Northeastern and Northern regions.

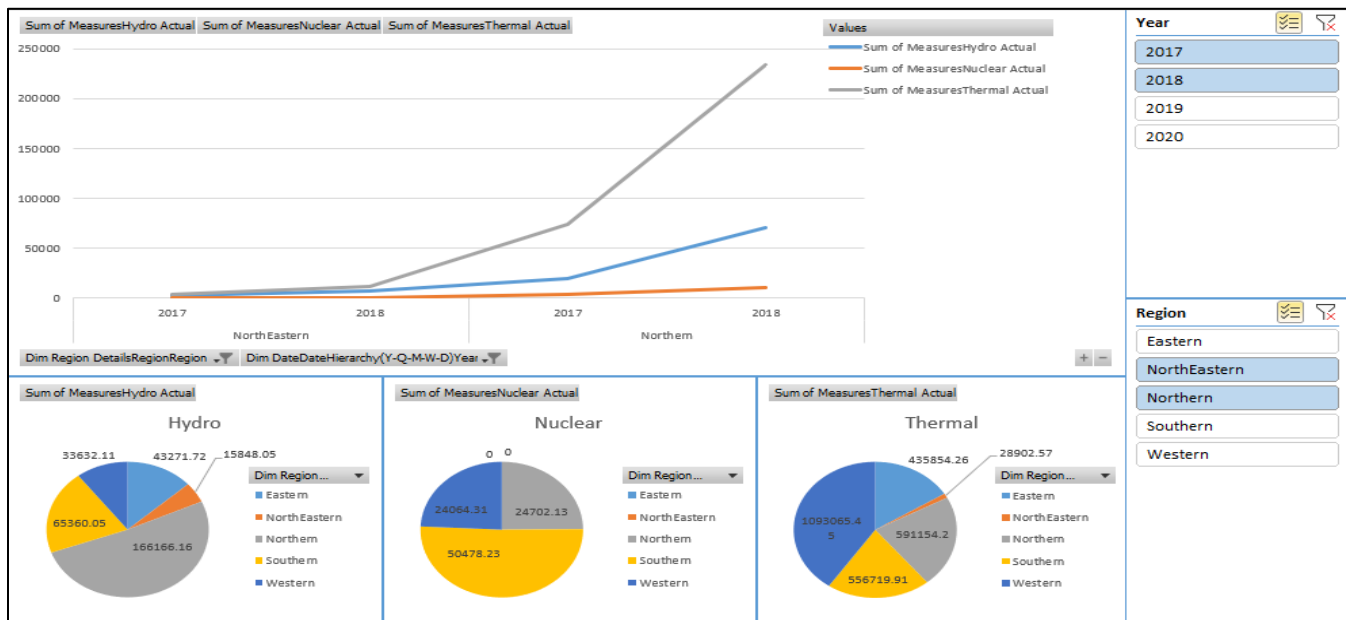


Figure 3-6 Dice

3.3 Pivot

For this, the excel worksheet has been connected to the cube using the Power Pivot mode.

The Region, Year, Hydro Actual, Nuclear Actual and Thermal Actual fields were selected for this operation.

Before pivot

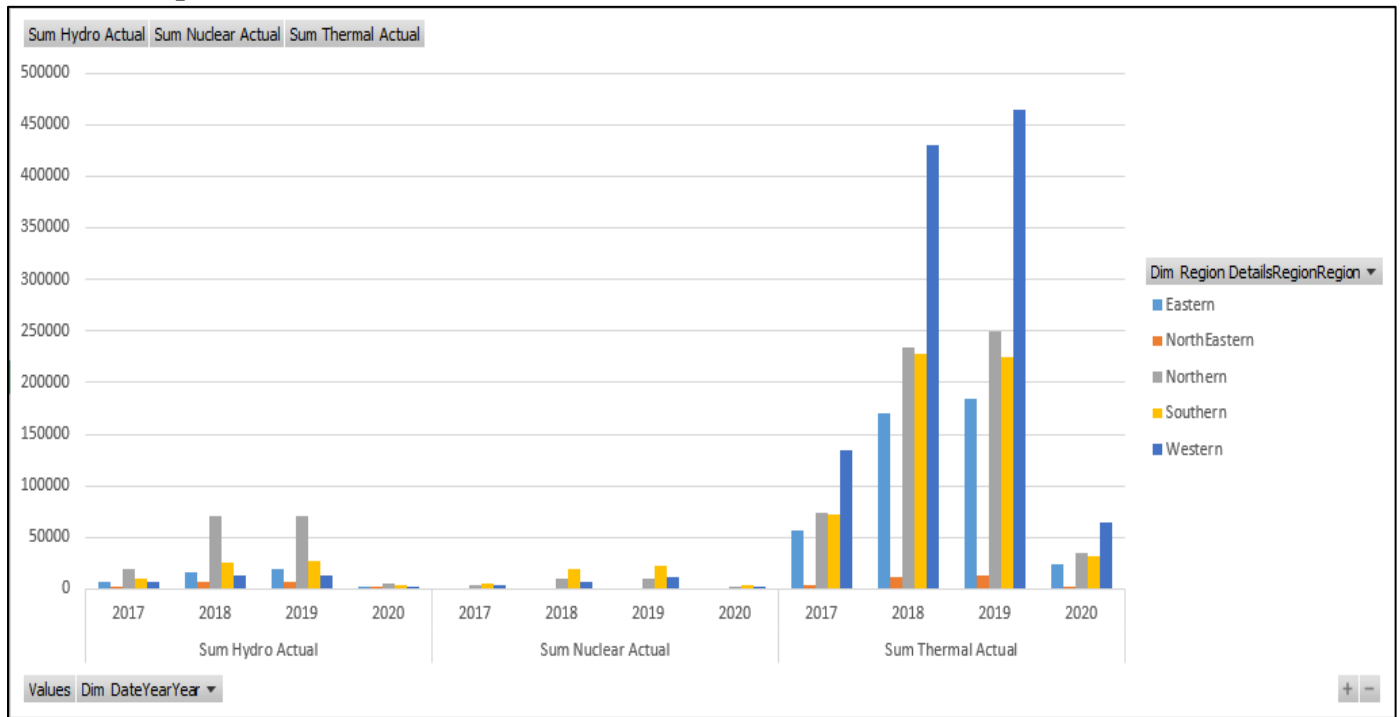


Figure 3-7 Pivot (Before pivot)

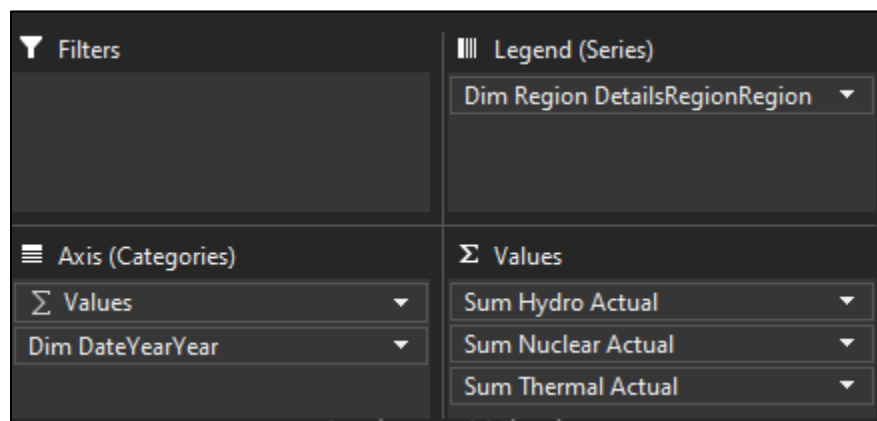


Figure 3-8 PivotChart Fields (Before pivot)

After Pivot

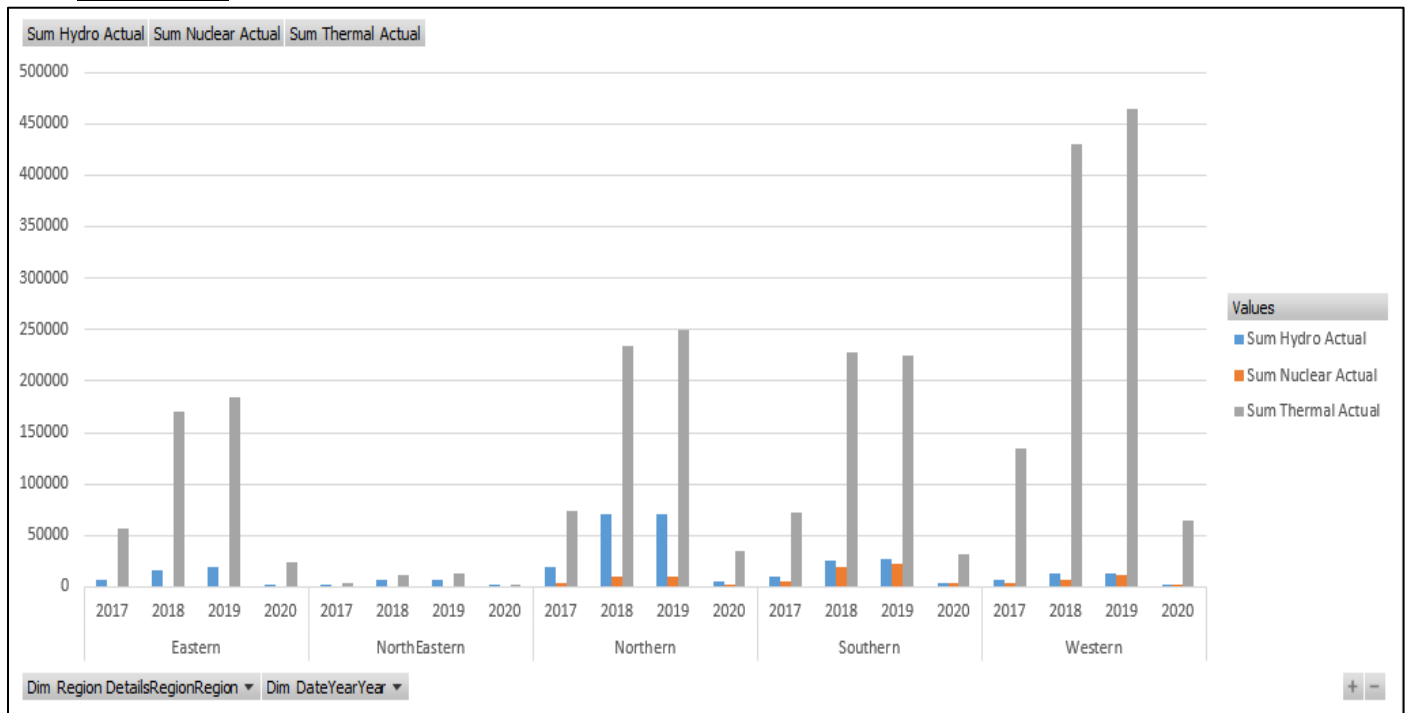


Figure 3-9 Pivot (After pivot)

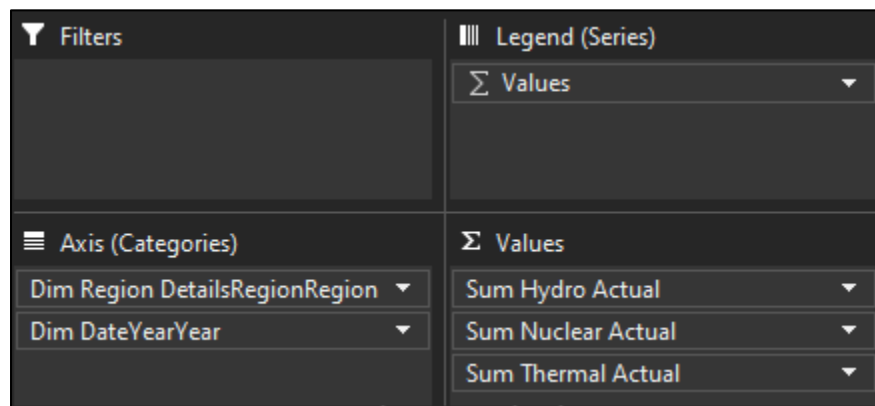


Figure 3-10 PivotChart Fields (After pivot)

4. SSRS Reports

The data warehouse India_Power_Generation_DW has been selected as the data source.

4.1 Report with a matrix

Data set

Query

Dataset name – **DataSet1**

```
SELECT
    DimRegionDetails.Region
    ,DimRegionDetails.Area
    ,DimRegionDetails.NationalShare
    ,DimRegionDetails.HouseHoldSize
    ,DimRegionCoordinates.Latitude
    ,DimRegionCoordinates.Longitude
    ,DimDate.[Year]
    ,DimDate.Quarter
    ,DimDate.[Month]
    ,DimDate.WeekOfMonth
    ,DimDate.DaySuffix
    ,DimDate.DayName
    ,DimDate.[Date]
    ,FactPowerGeneration.Thermal_Actual
    ,FactPowerGeneration.Thermal_Estimated
    ,FactPowerGeneration.Thermal_Difference
    ,FactPowerGeneration.Nuclear_Actual
    ,FactPowerGeneration.Nuclear_Estimated
    ,FactPowerGeneration.Nuclear_Difference
    ,FactPowerGeneration.Hydro_Actual
    ,FactPowerGeneration.Hydro_Estimated
    ,FactPowerGeneration.Hydro_Difference
FROM
    FactPowerGeneration
    INNER JOIN DimRegionDetails
        ON FactPowerGeneration.RegionDetailsSK = DimRegionDetails.ReionDetailsSK
    INNER JOIN DimRegionCoordinates
        ON FactPowerGeneration.RegionCoordinateSK = DimRegionCoordinates.RegionCoordinatesSK
    INNER JOIN DimDate
        ON FactPowerGeneration.DateSK = DimDate.DateKey
```

Figure 4-1 Report with matrix (Query)

Report

Region details Report					
Region	Area	National Share	House Hold Size	Latitude	Longitude
Eastern	418,336.00	13.2459	4.9	22.8962	85.9800
NorthEastern	262,179.00	8.3014	4.9	25.5736	93.2473
Northern	889,881.00	28.1766	5.2	28.6139	77.2090
Southern	636,251.00	20.1458	4.1	12.2602	77.1461
Western	508,042.00	16.0863	4.4	23.9074	72.7097
24/06/2021 12:05:32					

Figure 4-2 Report with a matrix

4.2 Drill-down report

Dataset

The same dataset that was used in the matrix report has been used for this. (Figure 4-1)

Report

Before drill-down –

< < 1 of 1 > > ↺ ↻ 100% ↵ 🖨️ Find Next						
Region-wise(Y-Q-M-W-D) Thermal Difference details Report						
Region	Year	Quarter	Month	Week Of Month	Date	Thermal Difference
▣ Eastern						12,376.07
▣ NorthEastern						-1,364.77
▣ Northern						38,727.05
▣ Southern						53,991.74
▣ Western						1,909.51
24/06/2021 12:10:02						

Figure 4-3 Drill-down report (before)

<div> <div> <div>1 of 1</div> <div>100%</div> <div>Find Next</div> </div> <div> <div>Region-wise(Y-Q-M-W-D) Thermal Difference details Report</div> </div> </div>						
Region	Year	Quarter	Month	Week Of Month	Date	Thermal Difference
Eastern	2017					4,707.64
	2018	1	1			404.97
			2	1	01/02/2018 00:00:00	-17.96
					02/02/2018 00:00:00	-10.49
					03/02/2018 00:00:00	-6.96
				2		-67.26
				3		-71.58
				4		-225.24
				5		-176.39
			3			-823.46
		2				-3,080.17

Drill-down can be done along **Year → Quarter → Month → Week → Day** for a selected Region.

4.3 Drill-through report

Dataset

First report - The same dataset that was used in the matrix report has been used for the first report which displays region wise house hold size. (*Figure 4-1*)

Second report –

- Report Data Structure

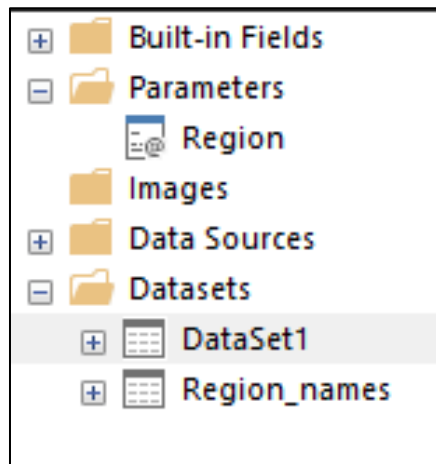


Figure 4-5 Drill-through second report (report data structure)

One parameter has been used for filtering.

- Query for Region_names dataset

```
SELECT
    DimRegionDetails.ReionDetailsSK
    ,DimRegionDetails.Region
FROM
    DimRegionDetails
```

Figure 4-6 Query for Region_names dataset

- Query for DataSet1

```

SELECT
    DimRegionDetails.Region
    ,DimRegionDetails.Area
    ,DimRegionDetails.NationalShare
    ,DimRegionDetails.HouseHoldSize
    ,DimRegionCoordinates.Latitude
    ,DimRegionCoordinates.Longitude
    ,DimDate.[Year]
    ,DimDate.Quarter
    ,DimDate.[Month]
    ,DimDate.WeekOfMonth
    ,DimDate.DaySuffix
    ,DimDate.DayName
    ,DimDate.[Date]
    ,FactPowerGeneration.Thermal_Actual
    ,FactPowerGeneration.Thermal_Estimated
    ,FactPowerGeneration.Thermal_Difference
    ,FactPowerGeneration.Nuclear_Actual
    ,FactPowerGeneration.Nuclear_Estimated
    ,FactPowerGeneration.Nuclear_Difference
    ,FactPowerGeneration.Hydro_Actual
    ,FactPowerGeneration.Hydro_Estimated
    ,FactPowerGeneration.Hydro_Difference
FROM
    FactPowerGeneration
    INNER JOIN DimRegionDetails
        ON FactPowerGeneration.RegionDetailsSK = DimRegionDetails.ReionDetailsSK
    INNER JOIN DimRegionCoordinates
        ON FactPowerGeneration.RegionCoordinateSK = DimRegionCoordinates.RegionCoordinatesSK
    INNER JOIN DimDate
        ON FactPowerGeneration.DateSK = DimDate.DateKey
where DimRegionDetails.Region in (@Region)

```

Figure 4-7 Query for DataSet1

Report

First report –

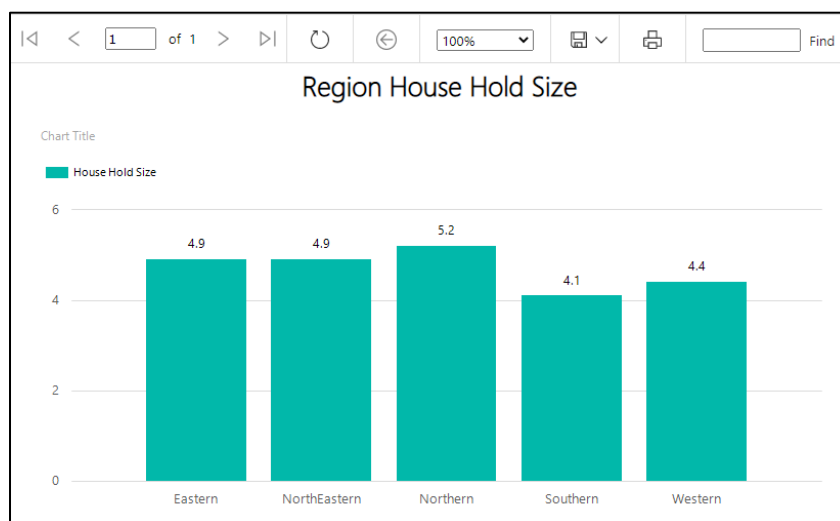


Figure 4-8 Drill-through (First report)

The report displays House Hold Size of each Region. Once we click on one region it will take us to a report where the yearly power generation data are displayed for the selected region.

Second report –

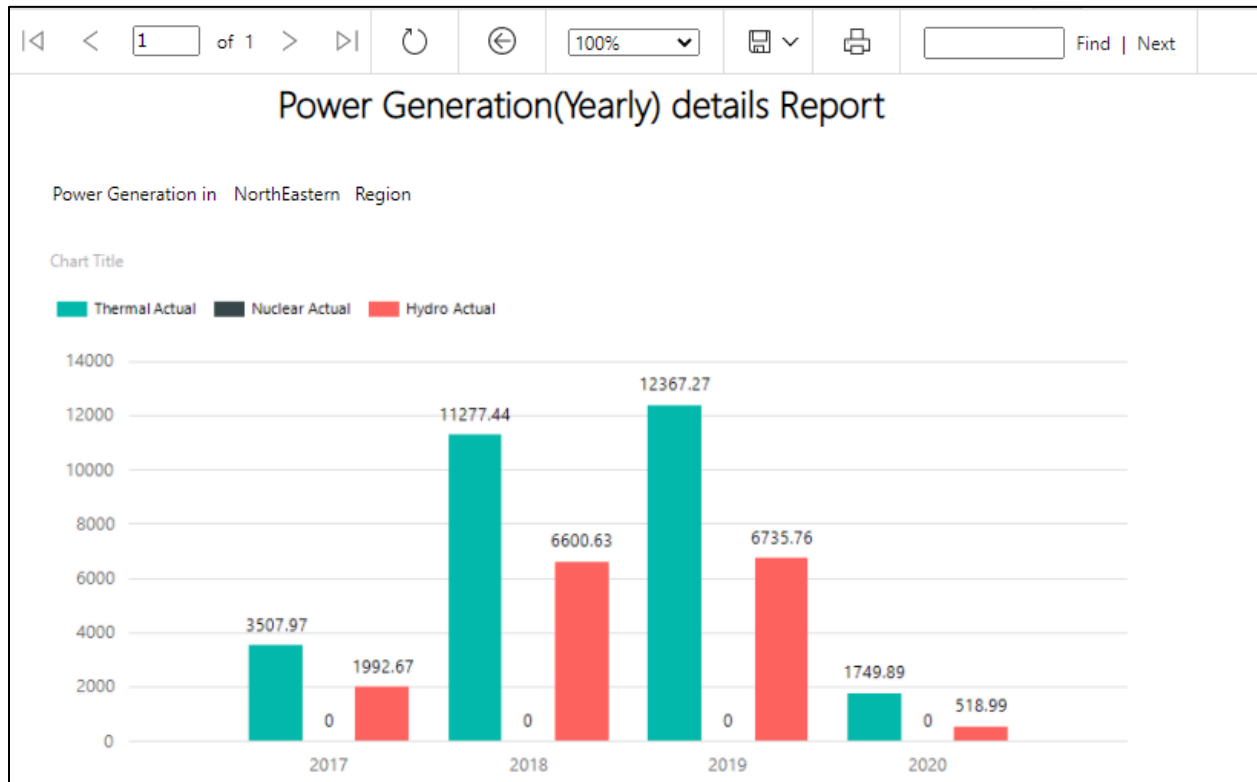


Figure 4-9 Drill-through (Second report)

Since we have selected the Northeastern region in the example, it shows power generation data of all three sources in that region.

4.4 Report with two parameters

Dataset

- Report Data Structure

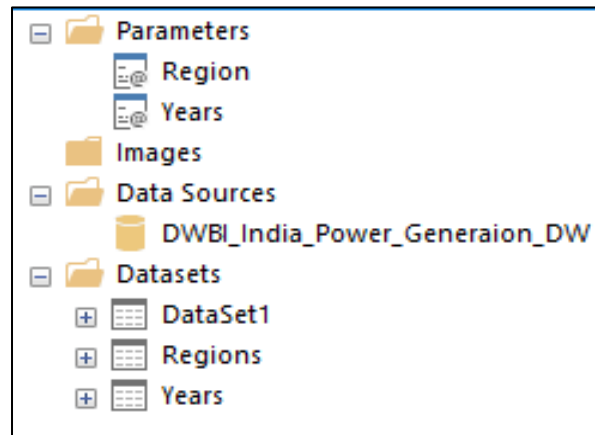


Figure 4-10 Report with two parameters (Report data structure)

As we can see in *Figure 4-10*, there are two parameters called Region and Years. The Region parameter would allow us to select a region. Then it will filter down the years in which we have minus values for Hydro Difference attribute.

$$\text{Hydro Difference} = \text{Hydro Actual} - \text{Hydro Estimated}$$

In other words, it will filter the years in which we have experienced a shortage in hydro power generation.

As shown in *Figure 4-10*, there are three data sets called Regions, Years and DataSet1 that respectively give us region names, years as per the condition and power generation data for the selected year.

- Query for Regions dataset

```
SELECT
    DimRegionDetails.ReionDetailsSK
    ,DimRegionDetails.Region
FROM
    DimRegionDetails
```

Figure 4-11 Query for Regions dataset

- Query for Years dataset

```

SELECT
    DimDate.Year
    ,SUM(FactPowerGeneration.Hydro_Difference)
FROM
    DimDate
    INNER JOIN FactPowerGeneration
        ON DimDate.DateKey = FactPowerGeneration.DateSK
    INNER JOIN DimRegionDetails
        ON DimRegionDetails.ReionDetailsSK = FactPowerGeneration.RegionDetailsSK
WHERE
    DimRegionDetails.Region in (@Region)
GROUP BY
    DimDate.Year
HAVING
    SUM(FactPowerGeneration.Hydro_Difference) < 0|

```

Figure 4-12 Query for Years dataset

- Query for DataSet1 dataset

```

SELECT
    DimRegionDetails.Region
    ,DimRegionDetails.Area
    ,DimRegionDetails.NationalShare
    ,DimRegionDetails.HouseHoldSize
    ,DimRegionCoordinates.Latitude
    ,DimRegionCoordinates.Longitude
    ,DimDate.[Year]
    ,DimDate.Quarter
    ,DimDate.[Month]
    ,DimDate.WeekOfMonth
    ,DimDate.DaySuffix
    ,DimDate.DayName
    ,DimDate.[Date]
    ,FactPowerGeneration.Thermal_Actual
    ,FactPowerGeneration.Thermal_Estimated
    ,FactPowerGeneration.Thermal_Difference
    ,FactPowerGeneration.Nuclear_Actual
    ,FactPowerGeneration.Nuclear_Estimated
    ,FactPowerGeneration.Nuclear_Difference
    ,FactPowerGeneration.Hydro_Actual
    ,FactPowerGeneration.Hydro_Estimated
    ,FactPowerGeneration.Hydro_Difference
FROM
    FactPowerGeneration
    INNER JOIN DimRegionDetails
        ON FactPowerGeneration.RegionDetailsSK = DimRegionDetails.ReionDetailsSK
    INNER JOIN DimRegionCoordinates
        ON FactPowerGeneration.RegionCoordinatesSK = DimRegionCoordinates.RegionCoordinatesSK
    INNER JOIN DimDate
        ON FactPowerGeneration.DateSK = DimDate.DateKey
where DimRegionDetails.Region in (@Region) AND DimDate.[Year] in (@Years)

```

Figure 4-13 Query for DataSet1 dataset

Report

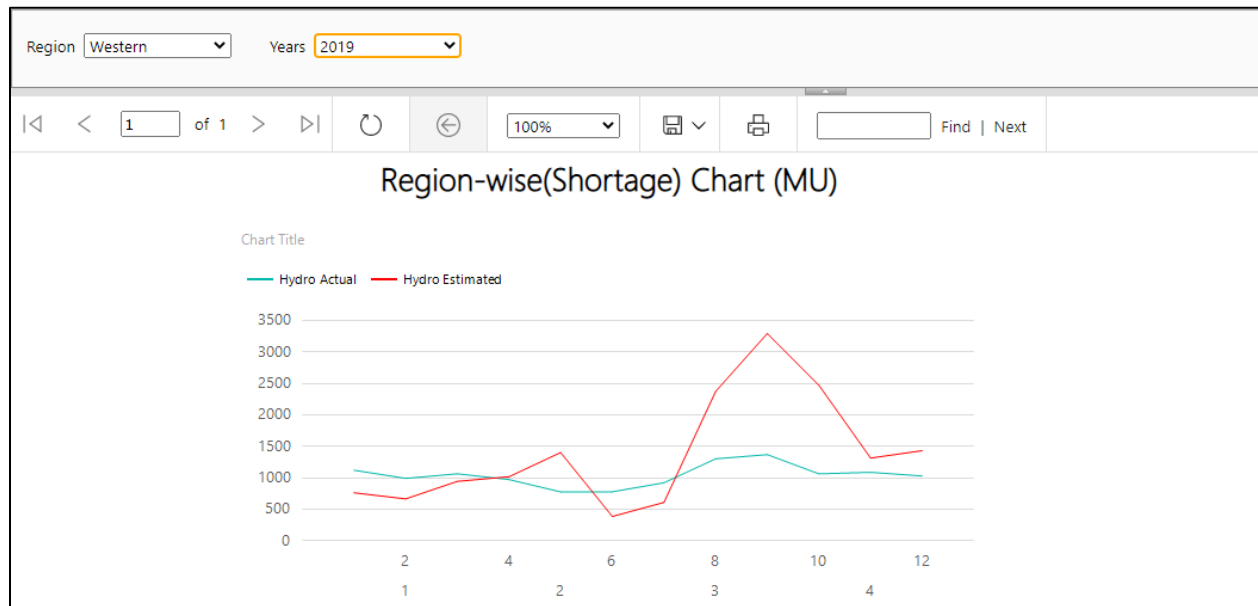


Figure 4-14 Report with two parameters

As shown in *Figure 4-14*, once we select the Western region it filters out years 2019 and 2020. Year 2019 was selected for the demonstration.

5. References

- <https://www.kaggle.com/>
- <https://courseweb.sliit.lk/>