

**CENTRE OF EXCELLENCE (CoE) PROJECT TRACKER
AUTOMATION**

A PROJECT REPORT

submitted by

CB.EN.U4ECE16160

Varsha B

under the guidance of

Mr. Peeyush.K.P.

*in partial fulfilment of the requirements for the award of the
degree of*

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING



AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE 641112

May 2020

AMRITA VISHWA VIDYAPEETHAM
AMRITA SCHOOL OF ENGINEERING, COIMBATORE, 641112



BONAFIDE CERTIFICATE

This is to certify that the project report entitled “**Centre of Excellence Project Tracker Automation**”, submitted by

CB.EN.U4ECE16160

B VARSHA

in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **ELECTRONICS AND COMMUNICATION ENGINEERING** is a Bonafide record of the work carried out under our guidance and supervision at the Amrita School of Engineering, Coimbatore and *JLL Property Consultants India Pvt. Ltd. at Bengaluru.*

Internal Project Advisor
Mr. Peeyush.K.P.
Assistant Professor(Sr.Gr.)

External Project Advisor
Mr.Elanchezhian S
Senior Director
JLL India

Project Coordinator
Mr. Sabarish Narayanan B
Assistant Professor(Sr.Gr.)

Dr. M. Jayakumar
Chairman,
Department of Electronics and Communication Engineering

The project was evaluated by us on:

Internal Examine

External Examiner



May 29, 2020

BONAFIDE CERTIFICATE

This is to certify that the project report entitled: **CoE Project Tracker Automation**, submitted by

NAME: B VARSHA

In partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Electronic and Communication Engineering** is a Bonafide record of the work carried out under guidance and supervision at **Jones Lang LaSalle Property consultants** by **Mr. Elanchezhian S.**

Yours sincerely,

A handwritten signature in black ink that reads "Phokri". The signature is written in a cursive style with a horizontal line underneath.

Level 16 Tower C Epsilon Building No 5
DLF Cyber City Phase III
Gurgaon 122002
Delhi 110001

Registered Office:
Jones Lang LaSalle Property Consultants (India) Pvt Ltd
No 1110 11th Floor Ashoka Estate 34 Barakhamba Road New
T +91 11 3314 1000 F +91 11 3314 1001
Company CIN No. U74140DL1997PTC091209

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and respect to the department of Electronics and Communication, Amrita School of Engineering for providing me an opportunity to improve my skills upon doing this project.

I express my sincere thanks to my class advisors Mr. Karthi B (Asst. Professor) and Mrs. Prabha G (Asst. Professor). I would also like to thank my panel members Mr. Sabarish Narayanan B (Asst Professor) and Dr. Ramanathan R (Asst. Professor) for fairly evaluating us and letting us know our strengths and faults on the run. I would also like to thank my internal guide Mr. Peeyush KP who has supported us on the run. I would also like to thank Dr. M.Jayakumar, Chairperson, Department of ECE.

I feel to acknowledge my indebtedness and deep sense of gratitude to my Manager Mr. Elanchezhian S whose valuable guidance and kind supervision given to me throughout the course which shaped the present work as it shows. Finally, I thank JLL Property Consultants (India) pvt ltd., for allowing me to do the project and submit this report.

ABSTRACT

CoE executes a number of software development projects for the business through multiple teams. The data related to the Project, Team, Weekly Status etc., are captured in multiple excel files / trackers. The management team faces challenges in consolidating and reporting this information on a regular basis.

Hence, to solve this issue, an integrated solution that automatically consolidates the data from various trackers into a central database on a defined schedule.

This model picks up data from all these files / trackers, transforms the data as per the company requirement and loads data into well designed tables which are modelled as per the Data Modelling techniques. The data from these tables are then picked up using a visualisation tool to display few major key parameters in the form of a visual which is interlinked to form a dashboard. This solution is automated such that it automatically updates the data given the time interval.

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION	PAGE NO
TDIM	Technology, Data and Information Management	2
SQL	Structured Query Language	2
SSIS	SQL Server Integration Services	2
ETL	Extract, Transform and Load	2
E-R	Entity Relationship	5
UML	Unified Modelling Language	5
OLE-DB	Object Linking and Embedding Database	10
KPI	Key Performance Indicator	11
SSMS	SQL Server Management Studio	11
NF	Normal Form	12
SSDT	SQL Server Data Tools	20
DB	Database	23
STG	Stage	34
RAG.	Red, Amber , Green	38
EMEA	Europe, Middle East, America	42
APAC	Asia Pacific Countries	43
SPOC	Specific Person of Contact	45

LIST OF FIGURES

FIG NO	TITLE	PAGE NO
Fig.2.1.	Snow Flake Schema	6
Fig 2.2.	ETL Process	7
Fig.2.3.	Dashboard Sample	8
Fig.3.1.	ETL Data Flow Task	9
Fig.3.3	Visualization Example	11
Fig.4.1.	Kinds of Normalization	13
Fig.6.1.1.	General Outline Model	22
Fig.6.1.2	Data Flow Model for Project Tracker	23
Fig.6.1.3	Flow chart for Project Tracker	24
Fig.6.1.4	Scrum cycle – sprint 1	25
Fig.6.1.5	Scrum cycle – sprint 2	26
Fig.6.1.6	Scrum cycle – sprint 3	26
Fig.6.1.7	Scrum cycle – Sprint 4	27
Fig.8.1.	Snow flake schema-based data model	32
Fig.8.2.	Project Tracker master list input	33
Fig.8.3.	Project and people master list	33
Fig 8.4.	Control flow task for Project master	34
Fig.8.5.	Data flow task to load the Project master	34
Fig.8.6.	Control flow task for Person master list	35
Fig.8.7.	Data flow task for Person master list	35
Fig.8.8.	Control Flow task - Tracker master list	36
Fig.8.9.	Data flow task to load Project data	36
Fig.8.10	Data flow task to load Team data	37
Fig. 8.11	Data flow task to load RAG data	37

FIG NO	TITLE	PAGE NO
Fig.8.12	Variable declaration	38
Fig-8.13	Stored Procedure	39
Fig.9.1.	Views	42
Fig.9.2.	Report – Page 1	43
Fig.9.3.	Report – Page 2	44
Fig.9.4.	Report – Page 3	44
Fig.9.5.	Report – Page 4	45
Fig.9.6.	Report – Page 5	45
Fig.9.7.	Report – Page 6	45
Fig.9.7.	Quarter view	46
Fig.9.8	Month view	46
Fig.9.9.	Report – full view.	46
Fig.9.10.	Dashboard – Full view	47
Fig.9.11.	Dashboard – normal view	48

LIST OF TABLES

TABLE NO	TITLE	PAGE NO
Table.4.2.1	Before 1st NF normalization	14
Table.4.2.2.	After 1st NF normalization	14
Table.4.3.1	Before 2 NF normalization	15
Table.4.3.2	First table after 2NF	15
Table .4.3.3.	Second table after 2NF	15
Table .4.4.1	Table before 3NF normalization	16
Table.4.4.2.	First table after 3NF	16
Table .4.4.3	Second table after 3NF	17
Table.4.5.1	Table before B-CNF normalization	17
Table 4.5.2	First table after B-C NF	18
Table 4.5.3.	Second table after B-C NF	18
Fig.4.5.4.	Third table after B-C NF	18
Table .8.1.	Staging table after loading	38
Table .8.2.	Data in one of the CoE data modelled main table after the stored procedure run	40

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	LIST OF ABBREVIATIONS	ii
	LIST OF FIGURES	iv
	LIST OF TABLES	v
1	CHAPTER 1 INTRODUCTION	1
	1.1 Introduction	2
	1.2 Motivation	3
	1.3 Objectives	3
2	CHAPTER 2 PROJECT DESCRIPTION	
	1.1 Description	4
	1.2 Methodology	5
	1.2.1 Data Modelling	5
	1.2.2 Snowflake Schema	6
	1.2.3 ETL Process	7
	1.2.4 Dashboards and Reporting	8
3	CHAPTER 3 PROCESS BREAKDOWN	
	3.1 ETL Load	10
	3.2 SSMS and Querying	11
	3.3 Visualizations	11
4	CHAPTER 4 NORMALISATION RULES	
	4.1 Normalization	12
	4.2 1 st Normal Form	13
	4.3 2 nd Normal Form	14
	4.4 3 rd Normal Form	15
	4.5 B-C Normal Form	17
5	CHAPTER 5 TECHNICAL SPECIFICATION	
	5 Software Specification	19

CHAPTER	TITLE	PAGE
6	CHAPTER 6 DESIGN APPROACH	
	6 Design Approach and Details	21
7	CHAPTER 7 SCHEDULES, TASKS AND MILESTONES	
	7 Schedules, Tasks and milestones	27
8	CHAPTER 8 PROJECT DEMONSTRATION	
	8 Project Demonstration	29
9	CHAPTER 9 RESULTS AND DISCUSSIONS	
	9 Results and Discussions	39
A1	REFERENCES	

CHAPTER 1

INTRODUCTION

INTRODUCTION

1.1. EXECUTIVE SUMMARY

JLL's TDIM CoE has numerous projects tied up under various verticals. Usually the team maintains the data pertaining to the project and updates its weekly status accordingly.

The overall information pertaining to all these projects are maintained in excel sheets. The excel sheets have all details pertaining to the projects belonging to a vertical in a huge list. An excel sheet with some 15 columns and 200+ data rows are not which is preferable as it leads to a heavy data loss and maintenance will become a tougher job when new projects gets added to the CoE. Appending would no longer be easier as it is hard to append every attribute of the project if it gets changed at any point of time. Moreover, there will be an update frequently done every now and then where it consumes a lot of space as you need keep adding new excel sheets when new bulk data comes in, say, a new vertical is being added to CoE. Visualisation is yet another issue in excel as adding a new charts or graphs would again consume a lot of space and yet again picking up fields for analysing becomes harder.

As a solution to all these, there comes the necessity to develop a system which can take in all the data which is coming and segregate them into readable various forms which enables one to easily maintain the data and the huge storage available would enable large heaps of data to flow in. Hence a tracker is to be developed to maintain the data from these excel sheets and stores it in the database in a format where the data could be easily analysed as well as maintained. Visualisations could be done effectively if this system is put forth.

1.2. MOTIVATION

According to TDIM CoE, the information pertaining to all the projects are all kept track in excel sheets where a minimum of 5 sheets are required to handle one vertical's data. There are around 11 verticals which has varying number of projects under it. So minimum of $5 \times 11 = 55$ sheets are required to store data pertaining to one particular vertical. And storage is another issue as data is most likely to get mixed up between the verticals as we enter the rows for similar fields. Visualisation is most likely a cause to

worry as right from the sheets what we have as we will not be able to combine and aggregate columns.

Hence, this has given the necessity to develop the tracker, where the data is loaded into staging tables via SSIS ETL operations. Now that tracker is being split into fact and dimensional tables, where data will be loaded from the Staging tables. And then the data is being segregated and filled into respective dimension tables. Finally, the tables are being picked up by the Power BI tool for visualisation purpose.

1.3. OBJECTIVES

The main objective of this model is such that

- An integrated solution that automatically consolidates the data from various trackers into a central database on a defined schedule.
- Automated reports and dashboards through Power BI
- Effective cost reduction by reducing the storage
- Maintenance of data

CHAPTER – 2

PROJECT DESCRIPTION

2. PROJECT DESCRIPTION

2.1. DESCRIPTION

- An integrated model is being developed which picks up the data present automatically when it is being triggered.
- A system which can schedule itself to run over a period of time to compare the previous data and the new ones and load the ones into the Staging tables.
- A snowflake schema-based data model which extracts data from the staging tables and loads it into the CoE tables.
- Automated dashboard and reports being created to visualise the data present in the CoE tables.

2.2. METHODOLOGY

2.2.1. DATA MODELLING

Data modelling is the process of creating models / containers which can hold data. This data model is the conceptual representation of data objects, the association between various data objects and the rules.

There are 2 kinds of data modelling techniques:

- Entity Relationship (E-R) Model
- Unified Modelling Language (UML)

There are 3 basic tenants of the data model. They are:

- Entity which is a real-world entity
- Attribute which are characteristics or properties of an entity
- Relationship which is the dependency or association between two entities

2.2.2. SNOWFLAKE SCHEMA

Snowflake schema is a logical arrangement of tables in a multidimensional database such that the Entity Relationship diagram resembles a snowflake shape. The snowflake schema is represented by centralised fact tables which are connected to multiple dimensions.

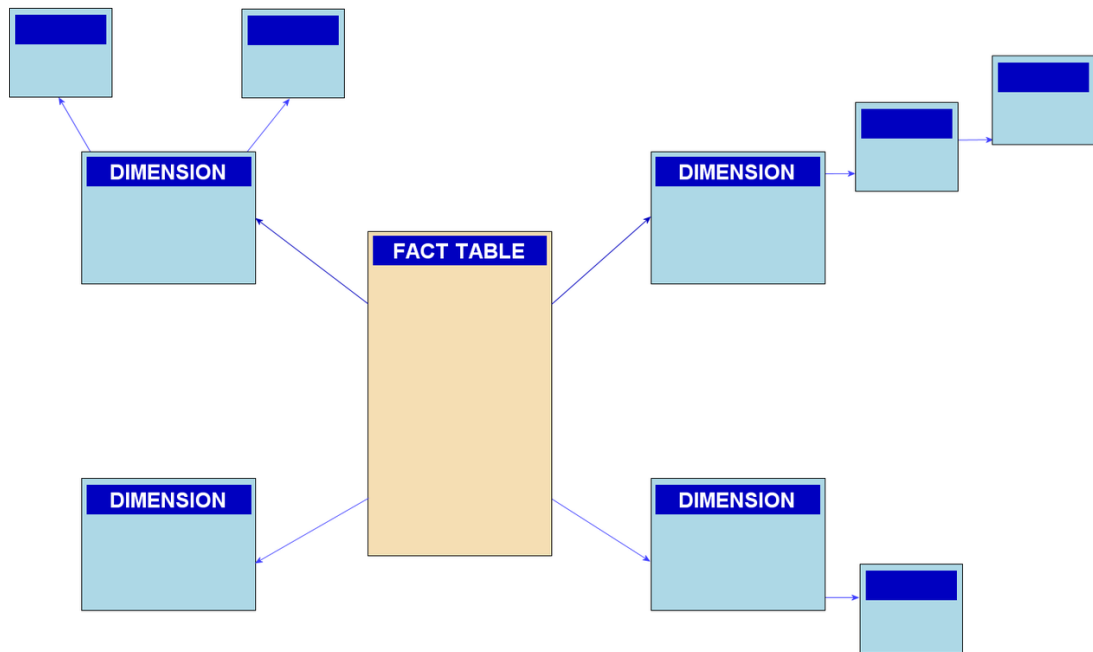


Fig.2.1. Data Modelling Architecture: Snowflake schema type

2.2.3. ETL PROCESS:

ETL abbreviated as Extract, Transform and Load is a process by which data is integrated and transformed from the operational systems into the Datawarehouse environment.

EXTRACT – Extraction is the process where you extract the data from the source files and load it into the ETL tool for further processing.

TRANSFORMATION – Transformation is the process where you transform the extracted raw data into the form which is preferred by you therefore making it efficient to load it into your respective destination.

LOAD – Loading is the process where your ETL tool picks up your transformed data and loads it into the destination where it must be maintained.

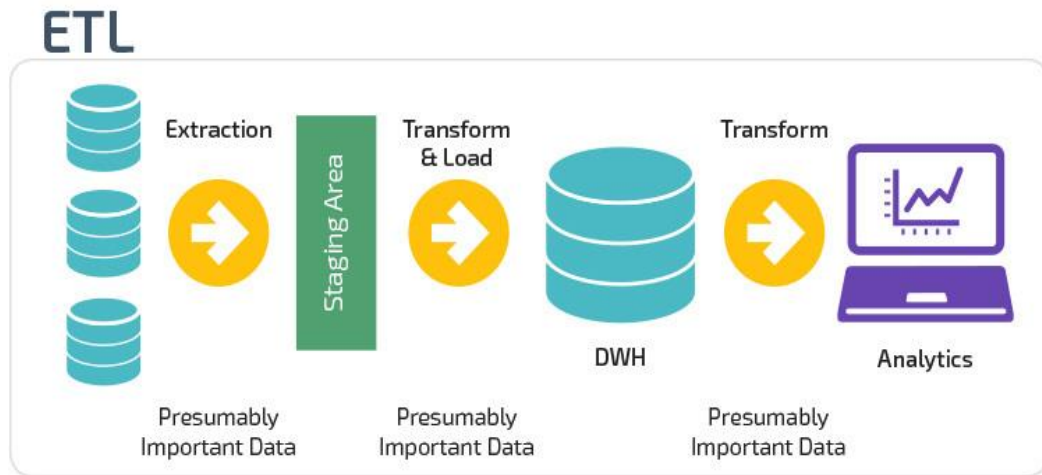


Fig 2.2. ETL Process

2.2.4. DASHBOARDS AND REPORTS

A dashboard is an information management tool that visually tracks, analyses and displays key performance indicators, metrics and key data points to monitor the health of a business, department or a specific process. A report is a multi-perspective view into a dataset, with visuals that represent different findings and insights from that dataset. A report can have single visual or pages full of visuals.



Fig.2.3. Dashboard Sample

CHAPTER – 3

PROCESS BREAKDOWN

3. PROCESS BREAKDOWN

The system basically consists of 3 key modules which are as follows:

3.1. ETL LOAD

This is the stage where data from all the excel sheets are loaded into the Database Staging tables. Here, firstly, the data is being picked up from a specific location, in this case the Excel files are being picked up from the SharePoint linked to my C drive. This is done by setting up a data flow task in SSIS where an excel configuration manager is being added in which the path is fed.

Next is the transformation phase. Generally, when the excel sheet is loaded a few columns are marked NULL in it. Therefore, you pick up the Conditional Split transformation from the SSIS toolbox where you can specify the ISNULL condition. The data which is valid will be transferred and the invalid rows are being removed.

The last stage is the loading stage where the transformed data will be transferred to the desired destination. In this case, it is an OLEDB Destination which is a database connection. Again, for the destination as well, the connection manager has to be configured.

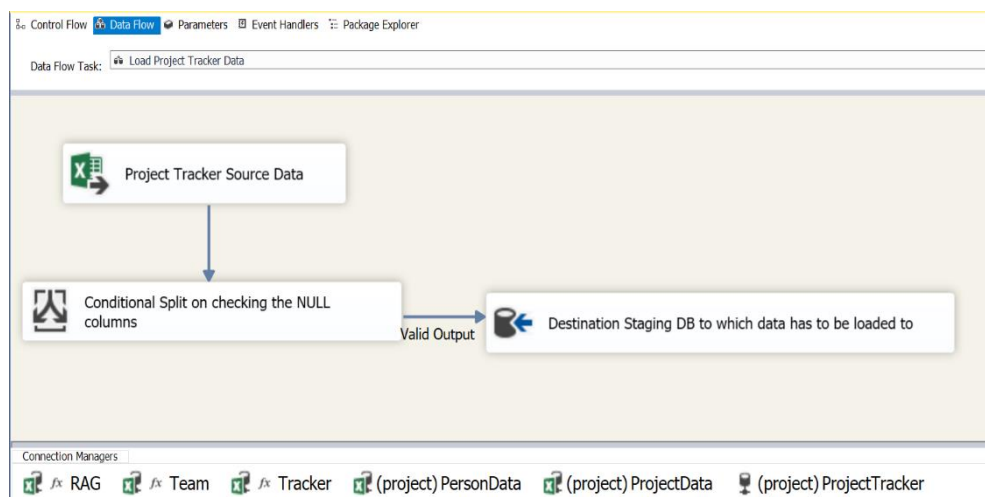


Fig.3.1. ETL Data flow task along with the connection managers

3.2. SSMS AND QUERYING

SQL Server Management Studio (SSMS) is a software application first launched with Microsoft SQL Server 2005 that is used for configuring, managing and administering all components within Microsoft SQL Server.

In this case, when a data model is made, the tables are loaded into the SQL Server via Queries. Now, the data from the Staging tables is being segregated and loaded into the Modelled CoE tables by means of a stored procedure which works on merge and join techniques.

3.3. VISUALISATIONS

The visualisations on the data being loaded is done by the visualising tool which requires views to be created in the SSMS. For every table and view loaded into the visualising tool, one has the privilege to pick upon what kind of visualisation they would like to display and can format anything with respect to the visualisation. Visuals make the data more comprehensible and one can just show the key KPIs which the viewer would be interested in.

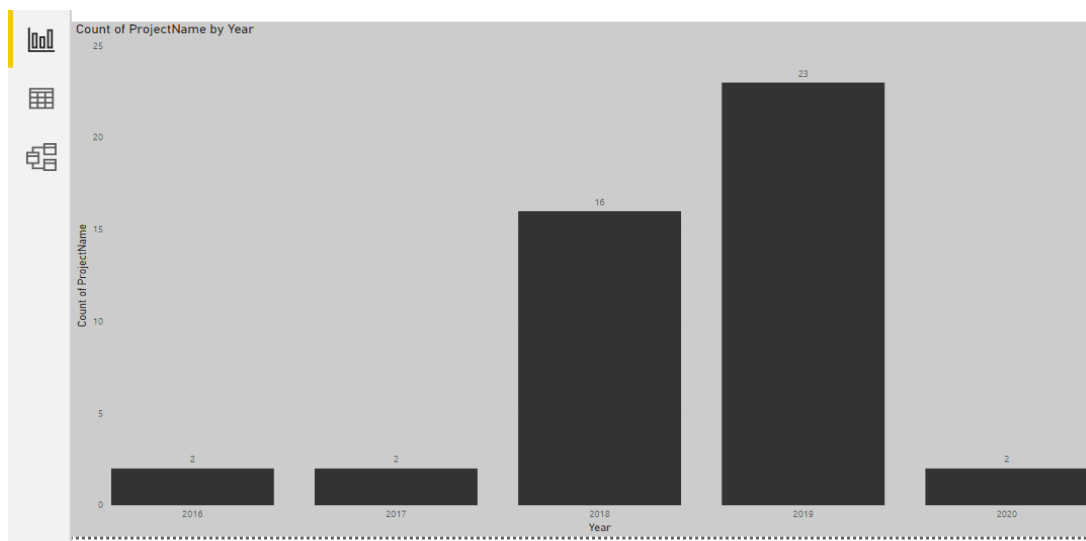


Fig.3.3. Visualisation example

CHAPTER 4

NORMALISATION RULES

4.NORMALISATION RULES

4.1. NORMALISATION

Normalisation is the process of reducing the redundancy of data in the table and improving the data integrity. Normalisation entails organising the columns and tables of a database to ensure that their dependencies are properly enforced by database integrity constraints.

Without Normalisation, we face issues such as

- Insertion anomaly where we cannot insert data to the table without the presence of another attribute.
- Update anomaly where data inconsistency results from data redundancy and a partial update of data.
- Deletion anomaly where certain attributes are lost because of the deletion of other related attributes.

Normalisation comes in 4 forms:

- 1st NF
- 2nd NF
- 3rd NF
- B-C NF

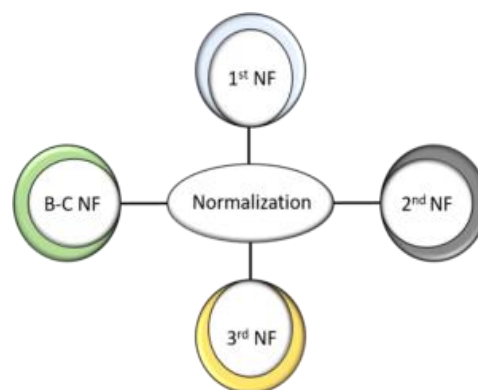


Fig.4.1 – Kinds of Normalisation

4.2. 1st NF

In this case, in a table, we ensure that the cell does not hold more than a single value. Hence, the atomicity of a value at a row and column of a table is always maintained as 1.

Table.4.2.1 Before 1st NF normalisation

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385, 9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389, 8589830302	Punjab

Fig.4.2.2. After 1st NF normalisation

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385	UP
14	John	9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389	Punjab
12	Sam	8589830302	Punjab

4.3. 2nd NF

The table should obey the condition for the 1st NF and the table should not contain any partial dependency.

Table – 4.3.1 Before 2 NF normalisation

TEACHER_ID	SUBJECT	TEACHER_AGE
46	Chemistry	33
46	Biology	33
49	English	29
60	Math	39
60	Computer	39

Table – 4.3.2. First table after 2NF

TEACHER_ID	TEACHER_AGE
46	33
49	29
60	39

Table – 4.3.3. Second table after 2NF

TEACHER_ID	SUBJECT
46	Chemistry
46	Biology
49	English
60	Math
60	Computer

4.4. 3rd NF

A table will be in 3rd NF if it follows all rules of 2nd NF and 1st NF and along with this, it should not contain any transitive partial dependencies.

Table – 4.4.1 table before 3NF normalisation

EMP_ID	EMP_NAME	EMP_ZIP	EMP_STATE	EMP_CITY
123	Harry	201010	UP	Noida
234	David	02228	US	Boston
345	Jim	60007	US	Chicago
456	Jerry	06389	UK	Norwich
567	John	462007	MP	Bhopal

Table -4.4.2. First table after 3NF

EMP_ID	EMP_NAME	EMP_ZIP
123	Harry	201010
234	Stephan	02228
345	Lan	60007
456	Katharine	06389
567	John	462007

Table – 4.4.3 Second table after 3NF

EMP_ZIP	EMP_STATE	EMP_CITY
201010	UP	Noida
02228	US	Boston
60007	US	Chicago
06389	UK	Norwich
462007	MP	Bhopal

4.5. BC NF

It is also known as the 3.5 NF. A table belonging to this form has to satisfy all of the above 3 forms. Apart from that, this also has the condition that if every functional dependency $A \rightarrow B$, then A has to be the super key of that table.

Table.4.5.1 Table before B-CNF normalisation

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

Table 4.5.2 First table after B-C NF

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

Table 4.5.3. Second table after B-C NF

EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
Designing	D394	283
Testing	D394	300
Stores	D283	232
Developing	D283	549

Table.4.5.4. Third table after B-C NF

EMP_ID	EMP_DEPT
D394	283
D394	300
D283	232
D283	549

CHAPTER 5

TECHNICAL SPECIFICATION

5. TECHNICAL SPECIFICATION

CHAPTER 5.1 SOFTWARE SPECIFICATION

- SqlDBm – Used for creating a data model for the Project Tracker model and to enable the associations between the tables
- SQL Server 2019 – Is a relational database management system which is used to manage and store information.
- SQL Server Management studio 2019 – SQL Server Management Studio is an integrated environment for managing any SQL infrastructure. SSMS can be used to access, configure, manage, administer and develop all components of SQL Server, Azure SQL Database and SQL Datawarehouse used for storing and querying all Project Tracker Data.
- SSDT 2017- SSDT 2017 is a development tool which is primarily used for working on the SSIS ETL process. This enables one to ETL the data into the Project Tracker Staging tables.
- SSIS – SSIS is a component of the SSDT to perform long data migration tasks.
- Microsoft Power BI – Power BI is a cloud-based business analytics tool and data visualiser that enables anyone to visualise and analyse data with high speed, efficiency and understanding.

CHAPTER – 6

DESIGN APPROACH

6. DESIGN APPROACH

6.1 DESIGN APPROACH AND DETAILS

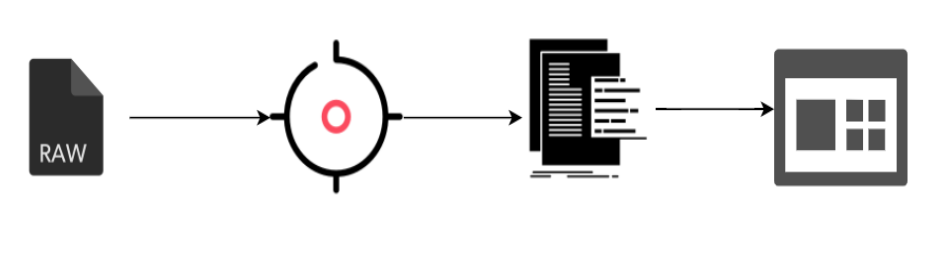


Fig.6.1.1. General outline model of the Project Tracker

Figure 6.1.1 depict the general outline of the ProjectTracker deployment model. The input is some raw data coming from an external source. The data is then fed into the Project Tracker which does all necessary operations within itself and stores the available data. It is an automated system which takes care of all the loading operations. The data from the project tracker is then loaded into the visualising tool to build reports. Finally, the reports are aggregated together to form visual dashboards. This is the entire design of the CoE Project Governance Tracker.

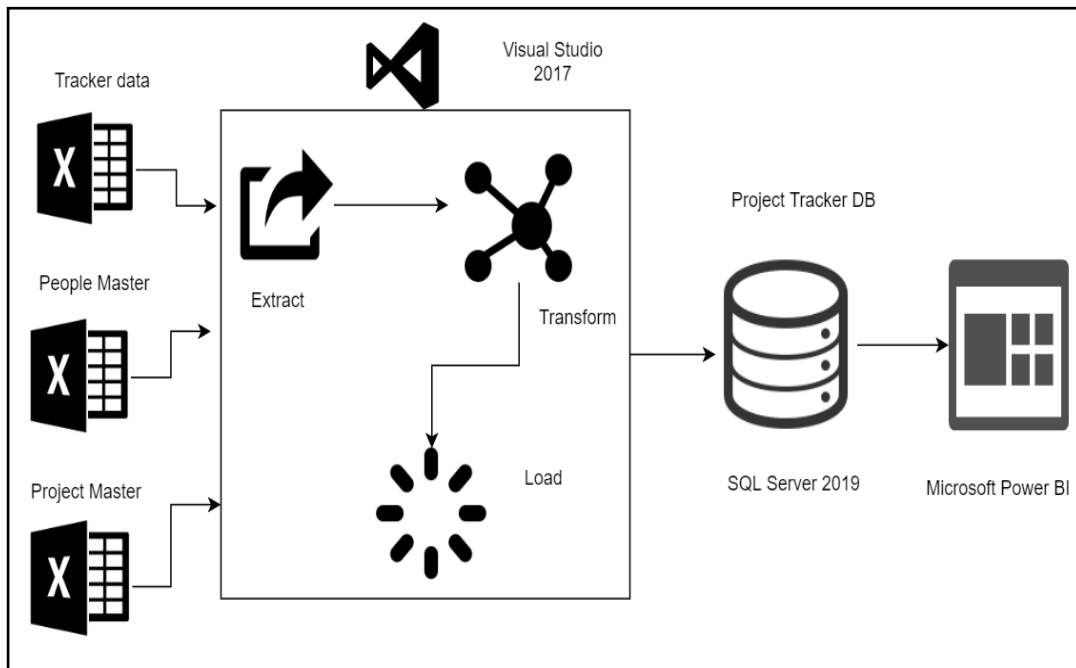


Fig.6.1.2 Data flow model for Project Tracker

Fig.6.1.2 depicts the Data flow diagram for Project Tracker. In this model, the data in the excel files, namely, Tracker data, People Master and the Project Master are loaded into the ETL platform which is SSDT's SSIS platform. There, the data is extracted from the excel files, transformed for the removal of invalid values. The transformed data is then loaded into the ProjectTracker database. There with the help of merge and join technique, a stored procedure created would segregate the data from the corresponding staging tables and loads it into the CoE's data modelled databases. Everything of this is stored into the ProjectTracker DB. There, few views are created to load the data into the visualiser where it will display a few KPI's as required by the viewer. Thus, finally, a dashboard is created in PowerBI, where when you pick each tile, it navigates to a further detailed report related to it.

DATA FLOW DIAGRAM:

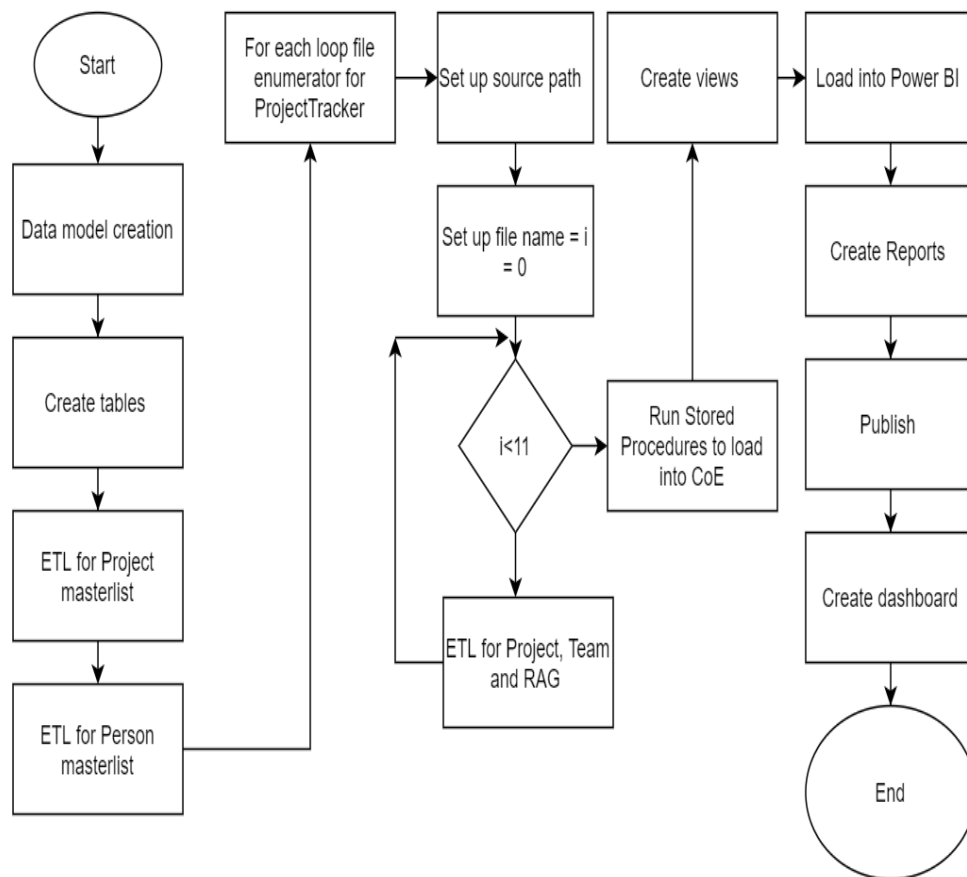


Fig.6.1.3. Flow chart for Project Tracker

Fig.6.1.3 depicts the flow of the Project Tracker module. This starts from the creation of the data model and creating those tables in the Project Tracker database. Project master list and person master list are loaded initially into the SSIS. They are transformed to load only valid data (checking for NULLs) and then they are loaded into the staging tables in the Project Tracker DB. Now the initiation for third master list, the Tracker master list is done. Correspondingly, a for each loop container is being set up and it is configured. So, when executed, it iterates throughout the folder, and loads data in all the excel sheets present within the folder. Once the loading is done, stored procedures are executed. After this step, we can see data in all the CoE tables which we created earlier from our data model. From the data we have segregated, we create views to combine important columns that hold data. Those views are loaded into the visualiser, i.e., Power BI to create reports. Once, the reports are created, we publish

them to the Power BI website. Over there, we pick in the important tiles that will highlight all of the major key values from our model. And hence we create a dashboard.

SCRUM CYCLE

Scrum lifecycle is a number of consecutive steps and iterative stages that should be performed during the realization of any scrum project. The iterative approach is the main principle of Scrum lifecycle. The work on scrum project is subdivided into segments called sprints. The project develops from one sprint to another until the final product is ready. Each sprint is subdivided into several consecutive stages that it must pass from the beginning till the end.

For this project, there are four sprints according to which the project runs. The sprints are as follows.

SCRUM SPRINT-1 FOR DATA MODEL

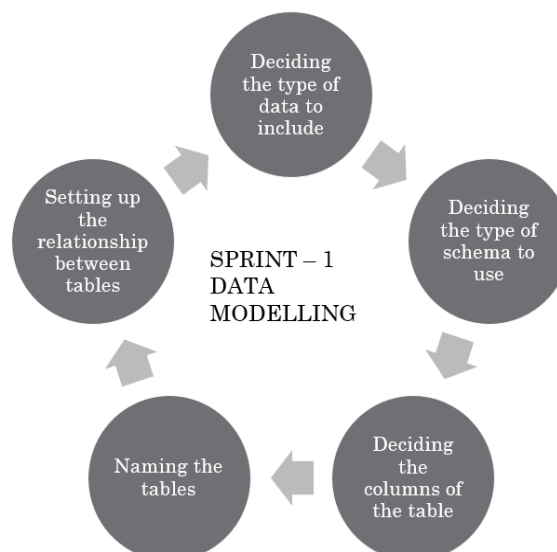


Fig 6.1.4 Scrum cycle – Sprint 1

The first sprint of the scrum cycle depicts how the data model has been constructed. The flow goes as it is mentioned above in Fig.6.1.4.

SCRUM FOR SSIS ETL LOAD

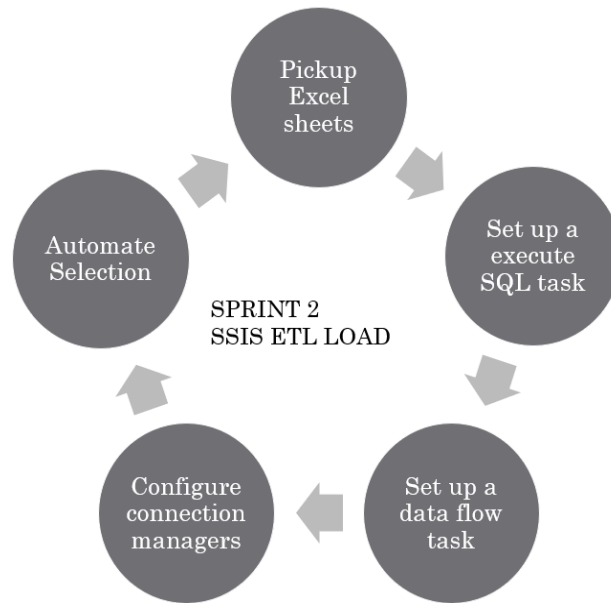


Fig.6.1.5. Scrum cycle – sprint 2

Fig.6.1.5 depicts the scrum cycle for the second stage. i.e., data loading. Here the cycle runs from picking up the data, till loading it into the ProjectTracker database. Automation in data selection is taken case in this sprint.

SCRUM FOR ETL TASKS

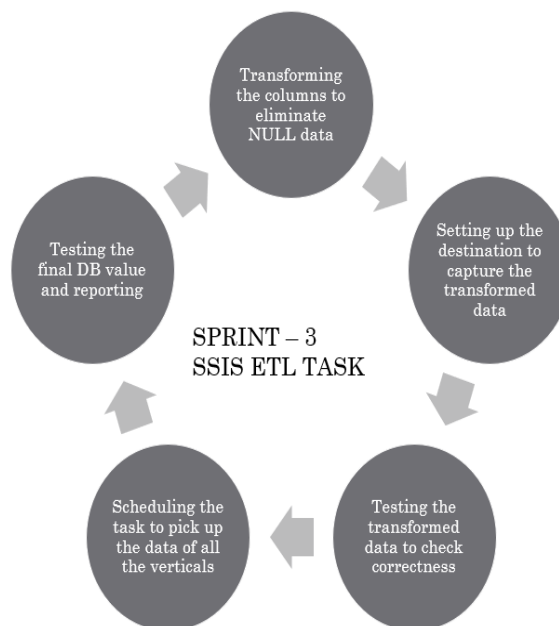


Fig.6.1.6 Scrum cycle – sprint 3

Fig.8.16 depicts the scrum cycle for the transformation and load tasks. The transformation task to eliminate NULL values is initiated. Then it is loaded into DB by means of configuring an OLE-DB destination.

SCRUM FOR REPORTING

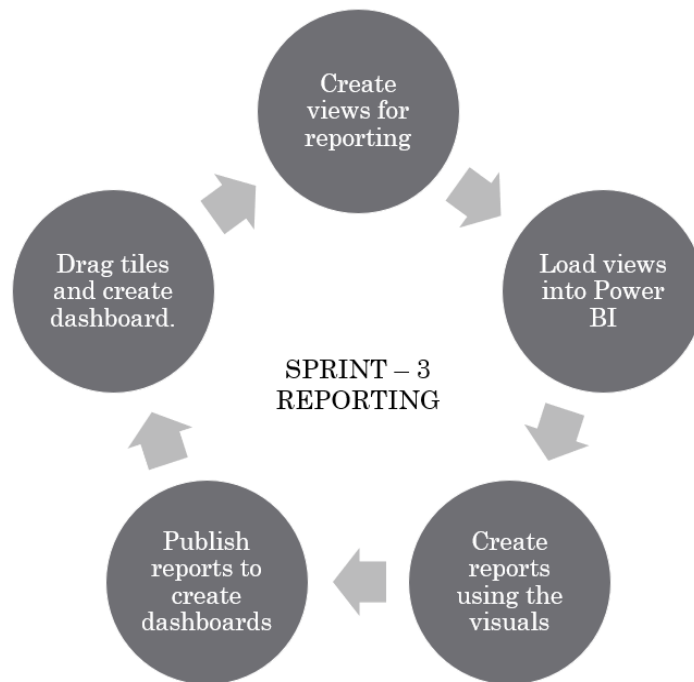


Fig.6.1.7 Scrum cycle – Sprint 4

Fig.6.1.7 depicts the scrum cycle for the last sprint, the reporting one. Views are created and loaded into Power BI. Data columns are picked up and reports are made. Those reports are published to the Power BI online platform. This allows one to edit, rearrange and navigate through the reports. One can pick up tiles to add it to the dashboard and hence the dashboard is being created.

CHAPTER – 7

SCHEDULE, TASKS AND MILESTONES

7. SCHEDULES, TASKS AND MILESTONES

These are the following major tasks that needs to be checked for successful compilation of the Project Tracker module:

- A data model to be created and modelled such that it gives the necessary list of tables to be created, the relationship between the tables and the
- ETL Task has to be scheduled through which two of the major master tables, Project master list and the People master list are loaded into the Project Tracker Database.
- An automation of the file selection must be done while loading the third master list, the tracker master list.
- A transformation must be done which automatically filters all invalid columns and initiates for the loading of valid columns into the Database.
- An automated stored procedure which segregates data coming from staging tables, compares the data and decides whether it has to be loaded, modified or discarded.
- Views for visuals.
- Reports to analyse the KPI's in the loaded data and realise them using various charts and diagrams.
- Dashboard to showcase tiles which shows a gist and when selected, gets directed to the report to give a better understanding of the situation.

CHAPTER – 8

PROJECT DEMONSTRATION

8. PROJECT DEMONSTRATION

Following are some screenshots which explain and demonstrate how Project Tracker module works.

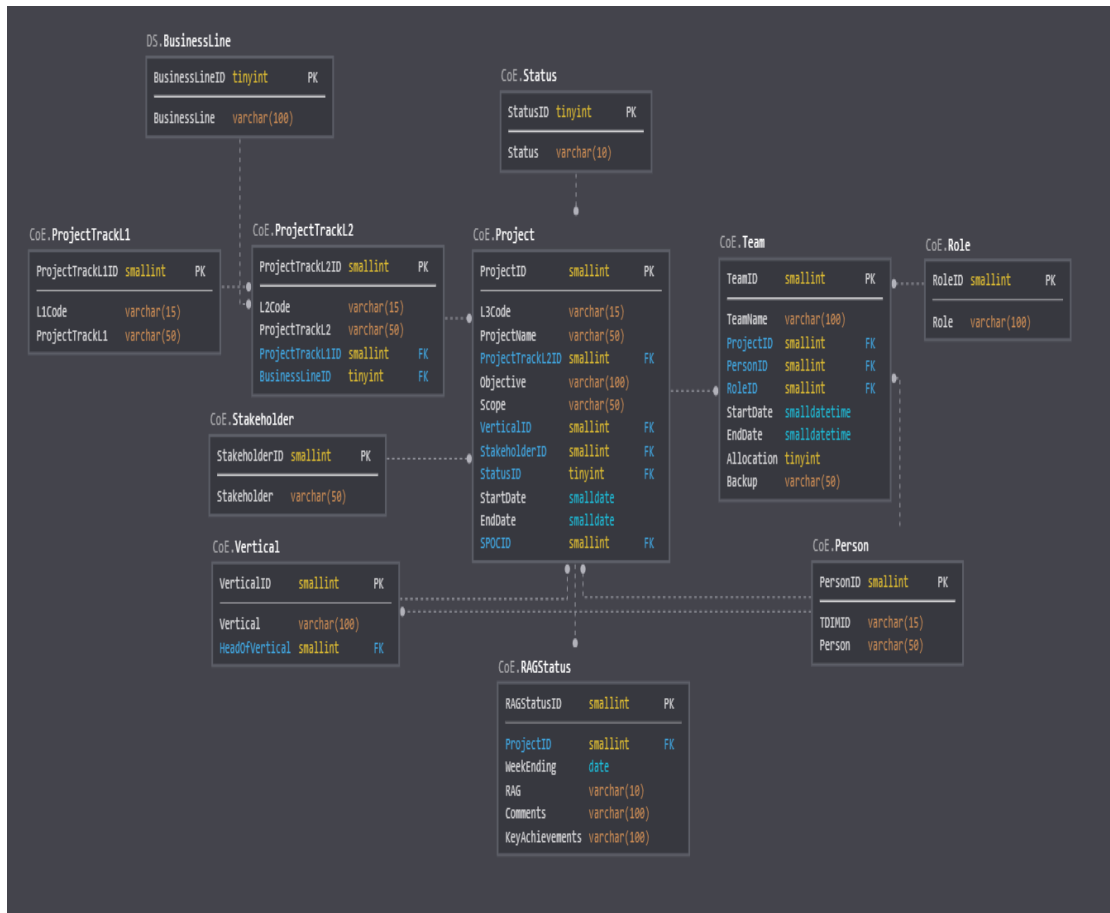


Fig.8.1. Snowflake schema-based data model for Project Tracker

Fig.8.1 depicts the data model created by analysing the data in the excel source files and these are the tables that are to be created to the SQL Server to hold the segregated data.

 Name ▾	Modified ▾	Modified By ▾	+ Add column ▾
 BIProjectTracker.xls	7 minutes ago	Bala, Varsha	
 DEVApp.xls	5 days ago	Bala, Varsha	
 DIProjectTracker.xls	6 days ago	Bala, Varsha	
 DSProjectTracker.xls	5 days ago	Bala, Varsha	
 EAProjectTracker.xls	About an hour ago	Bala, Varsha	
 ESSProject Tracker.xls	5 days ago	Bala, Varsha	

Fig.8.2. Project Tracker master list input

Fig.8.2. depicts the Project Tracker master list available at SharePoint which is being picked up as the third input in the Project Tracker module.




 Name ▾	Modified ▾	Modified By ▾	+ Add column ▾
 2020 04 17_Data in Revamped Project Track...	6 days ago	Bala, Varsha	
 2020 04 17_Projects Hierarchy MASTER v1.xls	April 28	Bala, Varsha	

Fig.8.3. Project and people master list

Fig.8.3. depicts the Project and the People master list present in the SharePoint which is being picked up as the first and second input in the project Tracker module.



Fig8.4. Control flow task to load the Project master data list

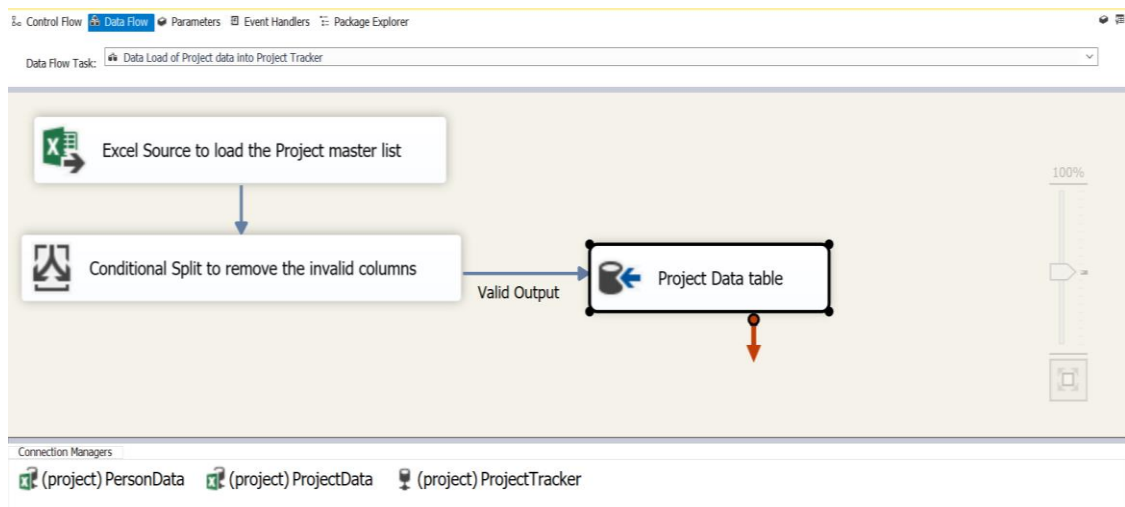


Fig.8.5. Data flow task to load the Project master list

Fig.8.4 and Fig.8.5 depicts the ETL task for the Project masterlist. These load data from the source, transform to include only the valid data and load the data into the database [STG].ProjectData. The excel's ProjectData connection manager and OLE DB'S Project Tracker connection manager are configured for this purpose.

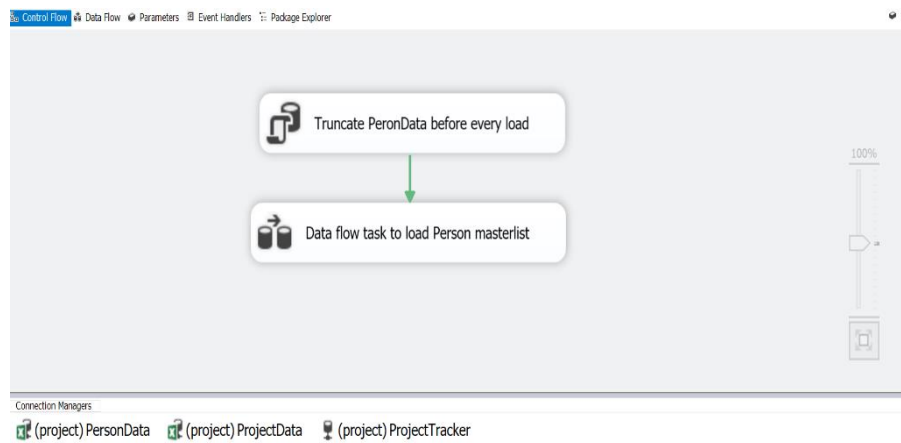


Fig.8.6. Control flow task for Person master list

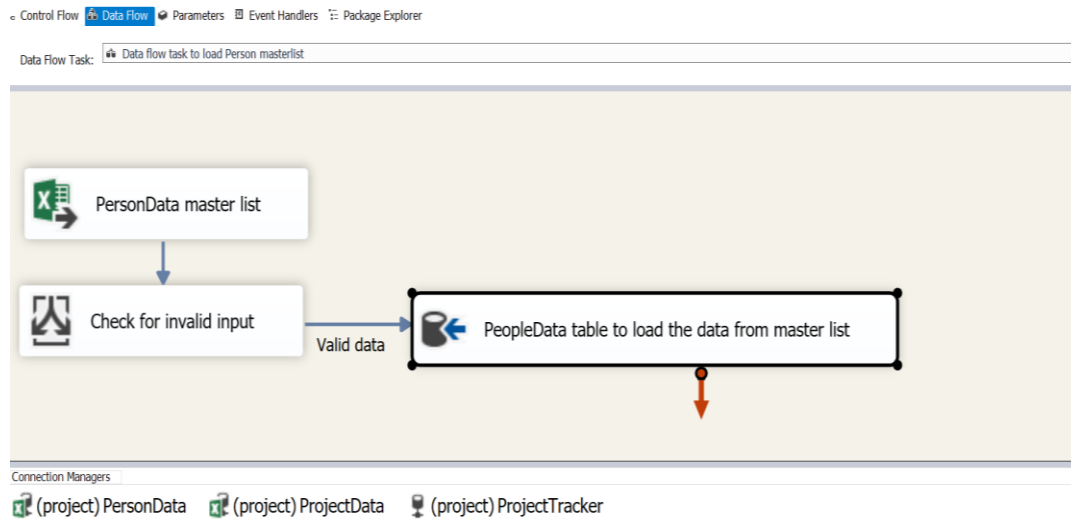


Fig.8.7. Data flow task for Person master list

Fig.8.6 and Fig.8.7 depict the Control flow and Data flow tasks for the Person master list. The PersonData excel connection manager and the ProjectTracker OLE-DB connection manager holds information about the source and the destination.

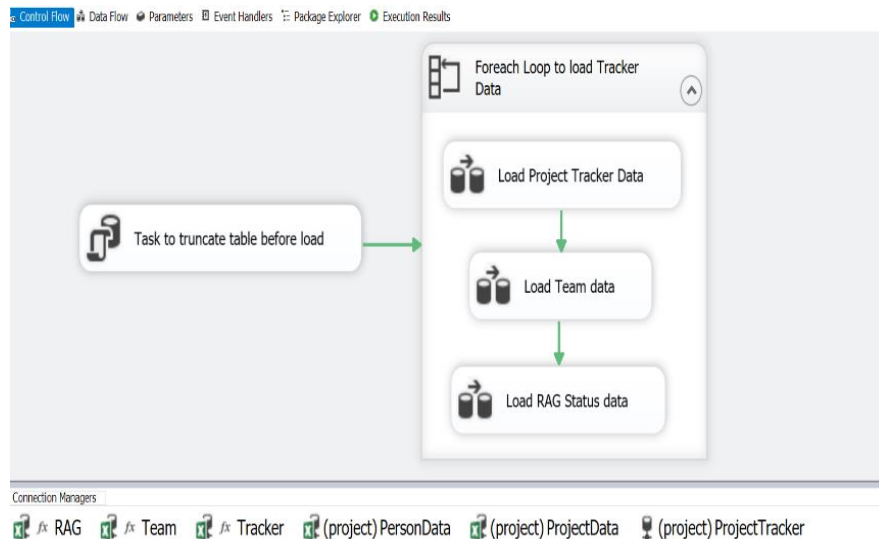


Fig.8.8. Control Flow task for Project Tracker master list

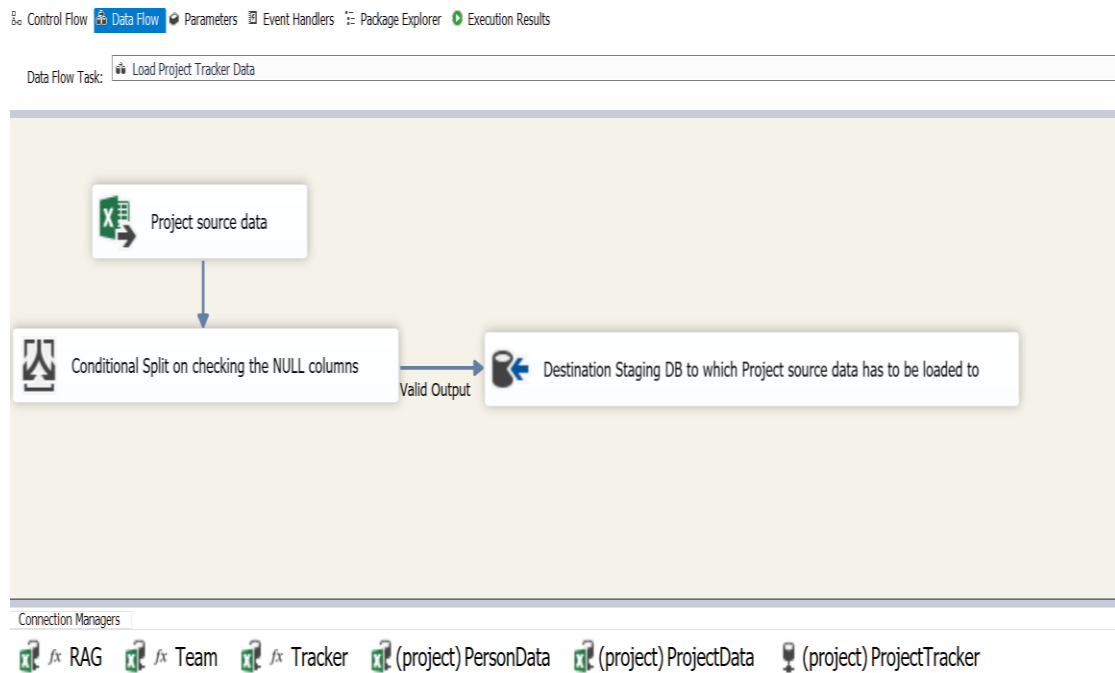


Fig.8.9. Data flow task to load Project data

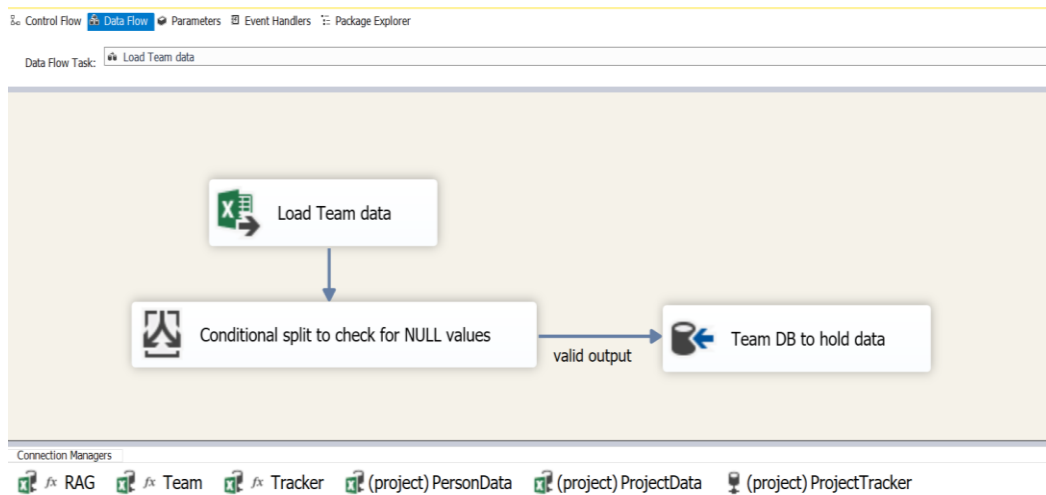


Fig.8.10 Data flow task to load Team data

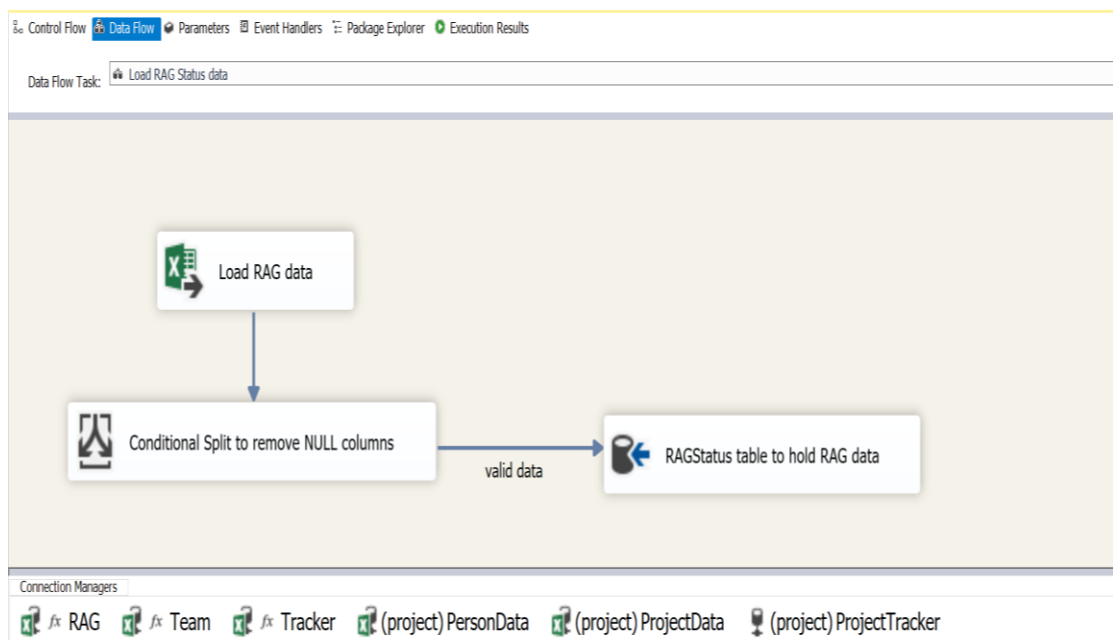
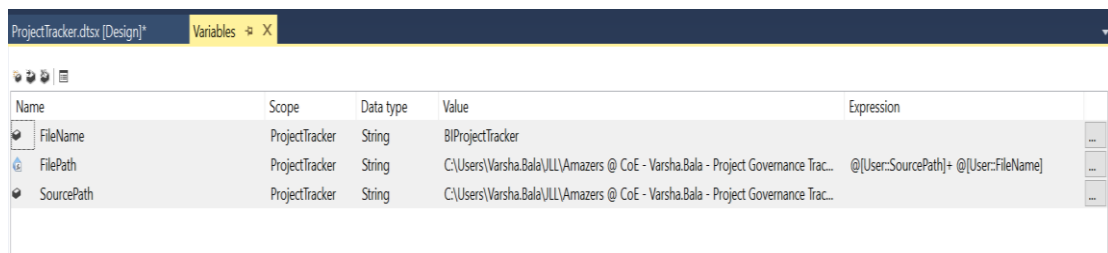


Fig. 8.11 Data flow task to load RAG data

Table .8.1. Staging table after loading

	L1Code	L2Code	Vertical	Role	BusinessLine
1	OneComp	OneComp	Application Development	Full Stack .Net Developer	Engineering - IT - Enterprise Systems
2	Global Property Solutions	Global Property Solutions (UI)	Application Development	React Web Application Developer	Engineering - Product Engineer - Data
3	Markets Tech APAC	QLOE App Development	Application Development	.Net Web App. Developer	Engineering - Product Engineer - Occupier
4	Markets Tech APAC	MarketLink Product Development	Application Development	Senior Quality Assurance Analyst	Engineering - Product Engineer - Occupier
5	Digital Marketing	Adobe Experience Manager (AEM)	Application Development	AEM Back-end Engineer	Product - Core
6	PDS Support	PDS 360	Application Development	VR 360 Developer	Engineering - Product Engineer - Occupier
7	CS OVA	CS OVA (App)	Application Development	.Net Web App. Developer	Client Success - Services & Implementation
8	Markets Tech APAC	Virtual Tour 360	Application Development	VR Developer	Engineering - Product Engineer - Occupier
9	CS OVA	CS OVA (App)	Application Development	.Net Web App. Developer	Client Success - Services & Implementation
10	Global Property Solutions	Global Property Solutions (UI)	Application Development	React Web Application Developer	Engineering - Product Engineer - Data
11	Global Property Solutions	Global Property Solutions (UI)	Application Development	API Developer	Engineering - Product Engineer - Data
12	Enterprise Application Architecture	Enterprise Application Architecture	Application Development	Full Stack Java Developer	Engineering - IT - Enterprise Systems
13	OLM	OLM APAC	Application Development	Sr. Drupal Developer	Engineering - Product Engineer - Occupier
14	PowerSearch Quality Testing	PowerSearch Quality Testing	Application Development	Sr. Quality Assurance - App	Engineering - Product Engineer - Occupier
15	OLM	OLM APAC	Application Development	Sr. Drupal Developer	Engineering - Product Engineer - Occupier
16	Global Property Solutions	Global Property Solutions (UI)	Application Development	Full Stack .Net Developer	Engineering - Product Engineer - Data
17	Markets Technical Operations Gr...	Markets Technical Operations Gr...	Application Development	DevOps Engineer	Engineering - Product Engineer - Occupier
18	Spotlight	Spotlight	Application Development	.Net Developer	Client Success - Services & Implementation
19	Spotlight	Spotlight	Application Development	.Net Developer	Client Success - Services & Implementation
20	PDS Clarizen	PDS Clarizen	Application Development	PDS Release Manager / QA	Engineering - Product Engineer - Occupier
21	PowerSearch Quality Testing	PowerSearch Quality Testing	Application Development	Lead Quality Assurance - App	Engineering - Product Engineer - Occupier
22	Markets Technical Operations Gr...	Markets Technical Operations Gr...	Application Development	Lead Full Stack Software Engin...	Engineering - Product Engineer - Occupier
23	OneComp	OneComp	Application Development	Full Stack Java Developer	Engineering - IT - Enterprise Systems
24	PDS Support	PDS Smart Design	Application Development	.Net Web App. Developer	Engineering - Product Engineer - Occupier
25	Digital Marketing	Adobe Experience Manager (AEM)	Application Development	AEM Senior Lead Engineer	Product - Core
26	Markets Technical Operations Gr...	Markets Technical Operations Gr...	Application Development	Project Manager	Engineering - Product Engineer - Occupier



Name	Scope	Data type	Value	Expression
FileName	ProjectTracker	String	BlProjectTracker	
FilePath	ProjectTracker	String	C:\Users\Varsha.Bala\LL\Amazers @ CoE - Varsha.Bala - Project Governance Trac...	@[User:SourcePath]+ @[User:FileName]
SourcePath	ProjectTracker	String	C:\Users\Varsha.Bala\LL\Amazers @ CoE - Varsha.Bala - Project Governance Trac...	

Fig.8.12 Variable declaration

Fig.8.8 till 8.12 and Table 8.1 depicts the load of the ProjectTracker master list. Here, data is being loaded from the ProjectTracker folder. There are around 11 excel files, hence source path and file name of the initial one is being picked up. The for loop container iterates through all the files and loads it into the staging tables Team , Tracker and RAG. The file path is generated dynamically by concatenating the source path and file name. The for each loop file enumerator takes care of picking up all the files with the specified extension.

```

CREATE PROCEDURE [CoE].[ProjectLoad]
AS
BEGIN
    MERGE [CoE].[Project] AS Tgt
    USING(
        select
DISTINCT(PD.L3Code)           AS L3Code,
PD.ProjectTrackL3             AS ProjectName,
PTL2.ProjectTrackL2ID         AS ProjectTrackL2ID,
PRA.Scope                     AS Scope,
V.VerticalID                  AS VerticalID,
ST.StakeholderID              AS StakeholderID,
S.StatusID                    AS StatusID,
PRA.StartDate                 AS StartDate,
PRA.EndDate                   AS EndDate,
P.PERSONID                    AS SPOCID

FROM      STG.ProjectData     AS PD
INNER JOIN CoE.ProjectTrackL2 AS PTL2 ON (PD.L2Code = PTL2.L2Code)
INNER JOIN STG.ProjectAdd     AS PRA ON (PD.L3Code = PRA.ProjectID)
INNER JOIN STG.PersonData     AS PERD ON (PD.ProjectTrackL2 = PERD.L2Code)
INNER JOIN CoE.Vertical       AS V   ON (PERD.Vertical = V.Vertical)
INNER JOIN CoE.Stakeholder    AS ST   ON (ST.Stakeholder = PERD.Stakeholder)
INNER JOIN CoE.Status         AS S    ON (S.Status = PRA.Status)
INNER JOIN CoE.Person         AS P    ON (PRA.SPOC = P.TDIMID)

        ) AS Src ON (
        Src.[ProjectName]      = Tgt.[ProjectName]
    AND Src.[L3Code]           = Tgt.[L3Code]
    AND Src.[ProjectTrackL2ID] = Tgt.[ProjectTrackL2ID]
    AND Src.[Scope]           = Tgt.[Scope]
    AND Src.[VerticalID]      = Tgt.[VerticalID]
    AND Src.[StakeholderID]   = Tgt.[StakeholderID]
    AND Src.[StatusID]        = Tgt.[StatusID]
    AND Src.[StartDate]       = Tgt.[StartDate]
    AND Src.[EndDate]         = Tgt.[EndDate]
        )

    WHEN NOT MATCHED BY TARGET THEN
        INSERT(ProjectName,L3Code,ProjectTrackL2ID,Scope,VerticalID,StakeholderID,StatusID,StartDate,EndDate,SPOCID)
        VALUES(Src.[ProjectName],Src.[L3Code],Src.[ProjectTrackL2ID],Src.[Scope],Src.[VerticalID],Src.[StakeholderID],Src.[StatusID],Src.[StartDate],Src.[EndDate],Src.[SPOCID])
    WHEN MATCHED THEN
        UPDATE
            SET ProjectName      = Src.[ProjectName],
                L3Code           = Src.[L3Code],
                ProjectTrackL2ID = Src.[ProjectTrackL2ID],
                Scope             = Src.[Scope],
                VerticalID        = Src.[VerticalID],
                StakeholderID     = Src.[StakeholderID],
                StatusID          = Src.[StatusID],
                StartDate         = Src.[StartDate],
                EndDate           = Src.[EndDate],
                SPOCID            = Src.[SPOCID];
END

```

Fig-8.13 Stored Procedure to load data from Staging table into CoE tables

The stored procedure mentioned above follows the merge and join condition. In this procedure, it checks for three cases.

- If the data is not present in the CoE table, then it loads the data into the CoE table from STG table.
- If data is present in the CoE table which is in the STG table, then it compares the data. It updates the columns which have been modified.
- If the data is present in the CoE table and if it exactly matches that of the STG table, then the load is discarded.

Table .8.2. Data in one of the CoE data modelled main table after the stored procedure run

	BusinessLineID	BusinessLine
1	2	Client Success - Services & Implementation
2	3	Engineering - IT - Brand/Comms Mktg
3	4	Engineering - IT - COE Management
4	5	Engineering - IT - Data Infrastructure
5	6	Engineering - IT - EMEA / APAC
6	7	Engineering - IT - Enterprise Systems
7	8	Engineering - IT - IT Infrastructure
8	9	Engineering - IT - IT Operations & Governance
9	10	Engineering - Product Engineer - Data
10	11	Engineering - Product Engineer - FM/PM
11	12	Engineering - Product Engineer - Investor
12	13	Engineering - Product Engineer - Occupier
13	14	Product - Core

Table.8.2. depicts the BusinessLine table after the data load through stored procedure. Similarly, 10 other tables are loaded in the same way.

CHAPTER –9
RESULTS AND DISCUSSIONS

9. RESULTS

At the end of this project, we were able to develop a tool which can address the problems of data maintenance.

100 %

Results Messages

	ProjectName	ProjectID	Vertical	Businessline
1	Acumen Support	1	Managed Services & Production Support	Engineering - Product Engineer - FM/PM
2	Capital Markets - Salesforce Support	2	Enterprise System Support	Engineering - Product Engineer - Investor
3	Capital Markets - Salesforce Support	3	Enterprise System Support	Engineering - Product Engineer - Investor
4	Capital Markets - Salesforce Support	4	Program Management Office	Engineering - Product Engineer - Investor
5	APAC Cloud Service Management	5	Enterprise IT Infrastructure	Engineering - IT - IT Infrastructure
6	Corporate Solution APAC - BI reporting	6	Business Intelligence & Reporting	Client Success - Services & Implementation
7	CS OVA	7	Data & Information Infrastructure	Client Success - Services & Implementation
8	Cyber Security Support (EMEA)	8	Enterprise IT Services & Security	Engineering - IT - IT Operations & Governance
9	Cyber Security Support (APAC)	9	Enterprise IT Services & Security	Engineering - IT - IT Operations & Governance
10	Data Stewards - CapForce	10	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
11	Data Stewards - CF/HFF	11	Data Governance & Stewardship	Engineering - Product Engineer - Investor
12	Data Stewards - Hotel Data	12	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
13	Data Stewards - MDM	13	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
14	Data Stewards - Sub Party Data	14	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
15	Data Stewards - Agent Insights & CF	15	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
16	Data Stewards - CRM integration	16	Data Governance & Stewardship	Client Success - Services & Implementation
17	Data Stewards - MarketSphere (US)	17	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
18	Data Stewards - Project Symphony	18	Data Governance & Stewardship	Engineering - IT - Data Infrastructure
19	Global CRM Dev & Support	19	Enterprise System Support	Engineering - IT - Enterprise Systems
20	EMEA Capforce - Data Migration	20	Data & Information Infrastructure	Engineering - Product Engineer - Data
21	EMEA Data Engineering	21	Data & Information Infrastructure	Engineering - Product Engineer - Data
22	EMEA Data Engineering	22	Data & Information Infrastructure	Engineering - Product Engineer - Data
23	CS Analytics	23	Enterprise Analytics	Product - Core
24	People Analytics	24	Enterprise Analytics	Product - Core
25	People Analytics	25	Enterprise Analytics	Product - Core
26	Enterprise Data Warehouse (EDW)	26	Data & Information Infrastructure	Engineering - Product Engineer - Data

Fig.9.1. View created to load data into Power BI to make reports

Fig.9.1 is one of the views, which contains only the key columns which we would like to display in our dashboard. This view picks up the ProjectID, ProjectName column from the Project table, Businessline from the Businessline table and Vertical from the Vertical table upon the join condition.

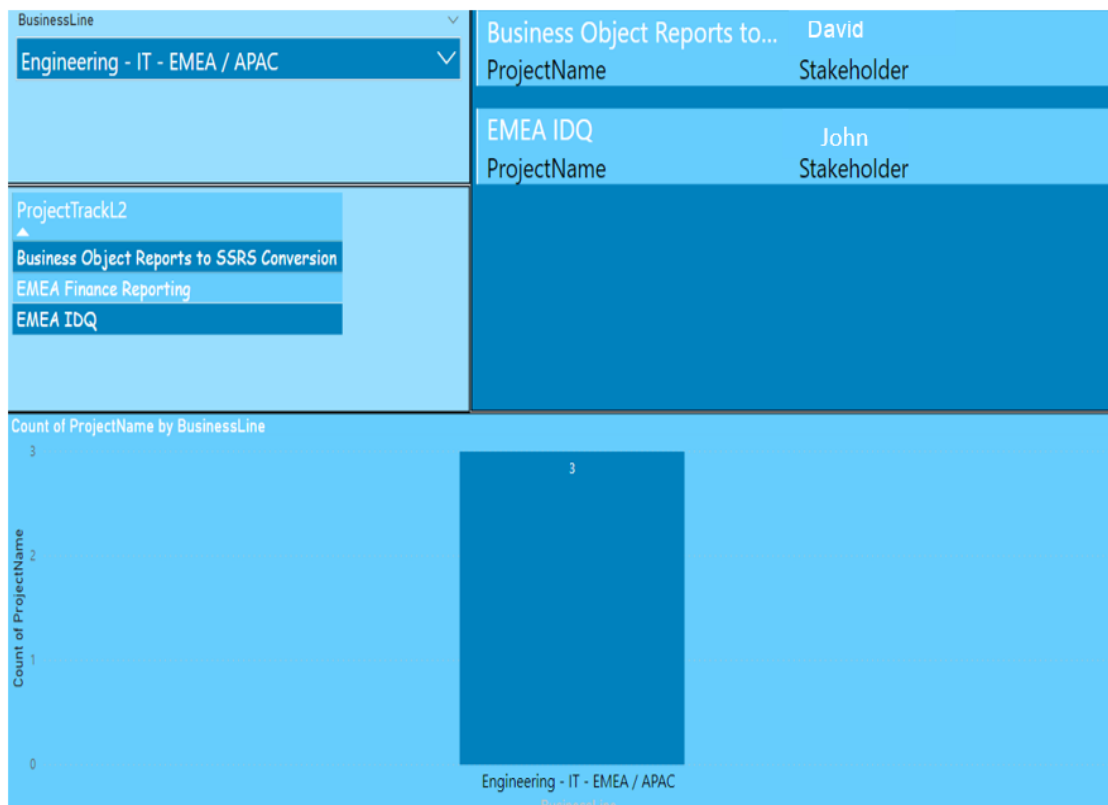


Fig.9.2. Report – Page 1

Fig.9.2 represents the first page of the report which we create using Power BI. This picks up the Businessline data and provides it as an option to choose between various Businesslines. The leftmost tile is the selection tile which enables us to do the above mentioned option. The second tile which is directly present below the selection tile is the one which displays all the ProjectTrackL2s which fall below the chosen vertical. The tile present to the right corner depicts the Stakeholder's name for a particular project under a particular Businessline. The final bar chart represents the number of projects under chosen Businessline.

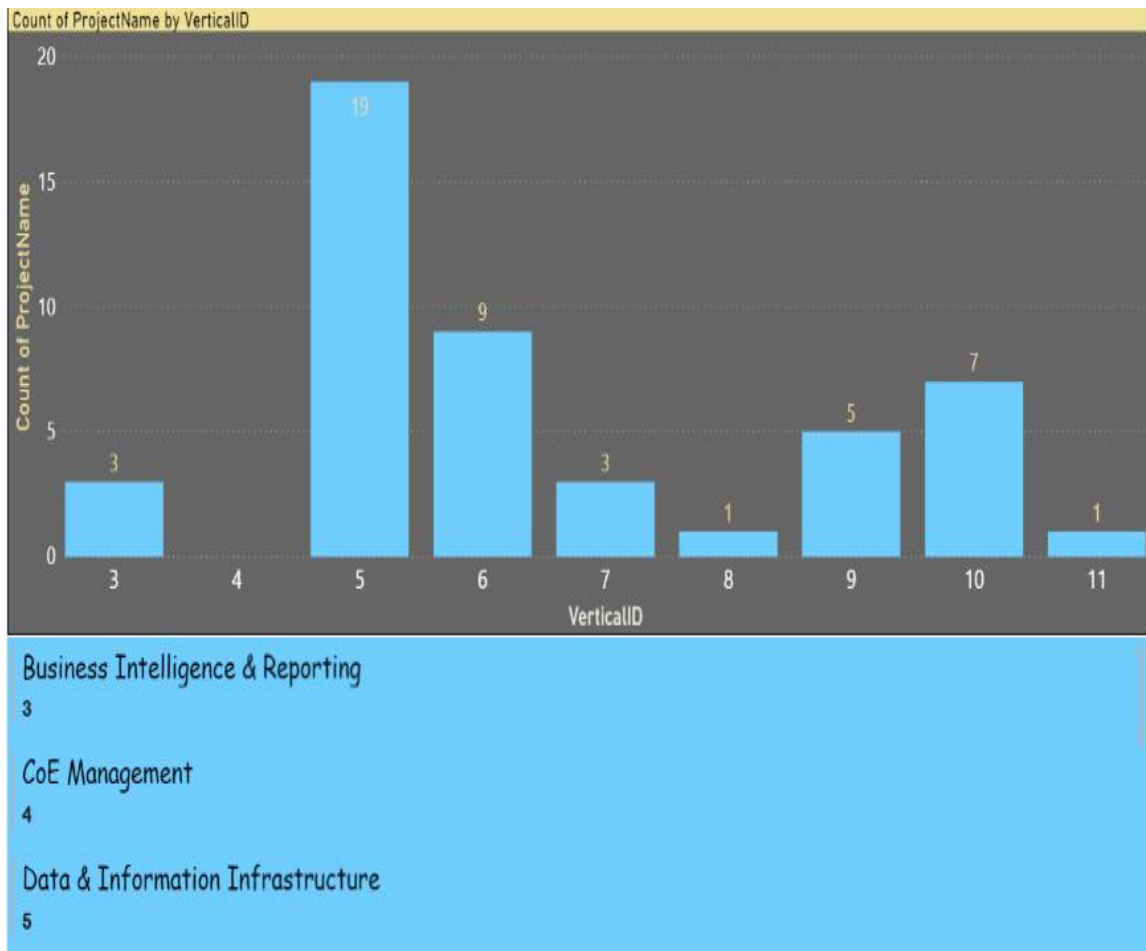


Fig.9.3. Report – Page 2

Fig.9.3 depicts the chart representing the number of Projects under each vertical. Here, the chart is plotted against verticalID, where the vertical name corresponding to each verticalID is mentioned below. One can scroll a little down to check out all the vertical names.

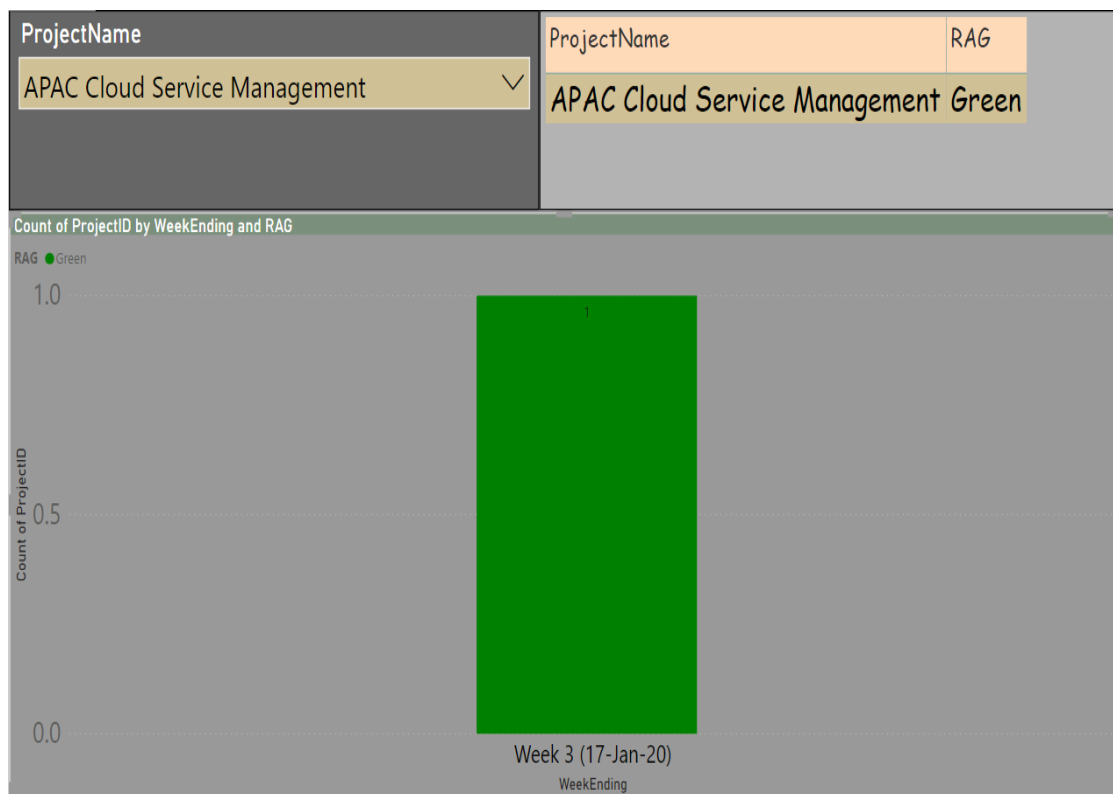


Fig.9.4. Report – Page 3

Fig.9.4 represents the RAG status of each vertical. This also enables selection. Upon the Project name selected, The RAG status is displayed across the ProjectName and a chart is being plotted depicting the status for a particular week. Here, APAC Cloud Service Management has a status Green marked on Week 3 which is on 17th January. The status would be Red, Amber or Green based on the current status of the Project, Previous status and the milestones achieved till date since the beginning of the Project.

ProjectName	SPOC	ProjectName
Acumen Support	Vishakh R	Acumen Support
Full Stack .Net Developer	Ashwin S	Acumen Support
Lead SSIS/SQL Developer	Shruthi Pathak	Acumen Support
MSBI Developer	Venkat Bulusu	Acumen Support
Sr. SSIS/SQL Developer	Swathi S	Acumen Support
SSIS/SQL Developer	Shravan R	Acumen Support
SSIS/SQL Developer	Vinitha R	Acumen Support
SSIS/SSRS Developer	Sriram S	Acumen Support

Fig.9.5. Report – Page 4

Fig.9.5 depicts the SPOC's name, team members and their corresponding roles upon selection of the Project. If the project name is changed, then its corresponding team members and their roles are displayed underneath.

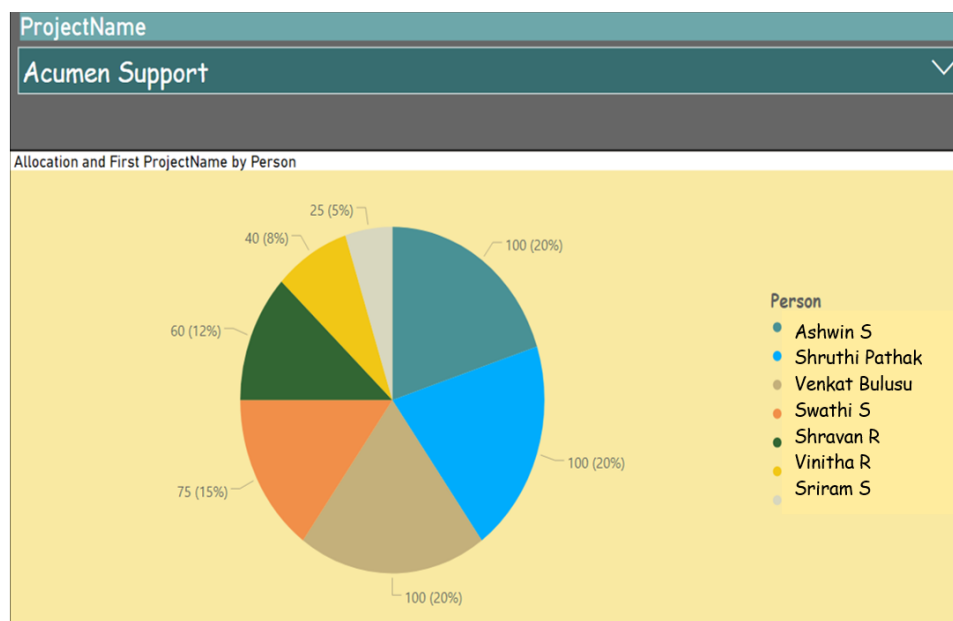


Fig.9.6. Report – Page 5

Fig.9.6 depicts the allocation of each person belonging to a project. Upon selection of the Project Name in the above selection box, the pie chart displays the contribution of each person. Each colour corresponds to each person. Towards the right corner, the colour is marked against the person's name for reference.

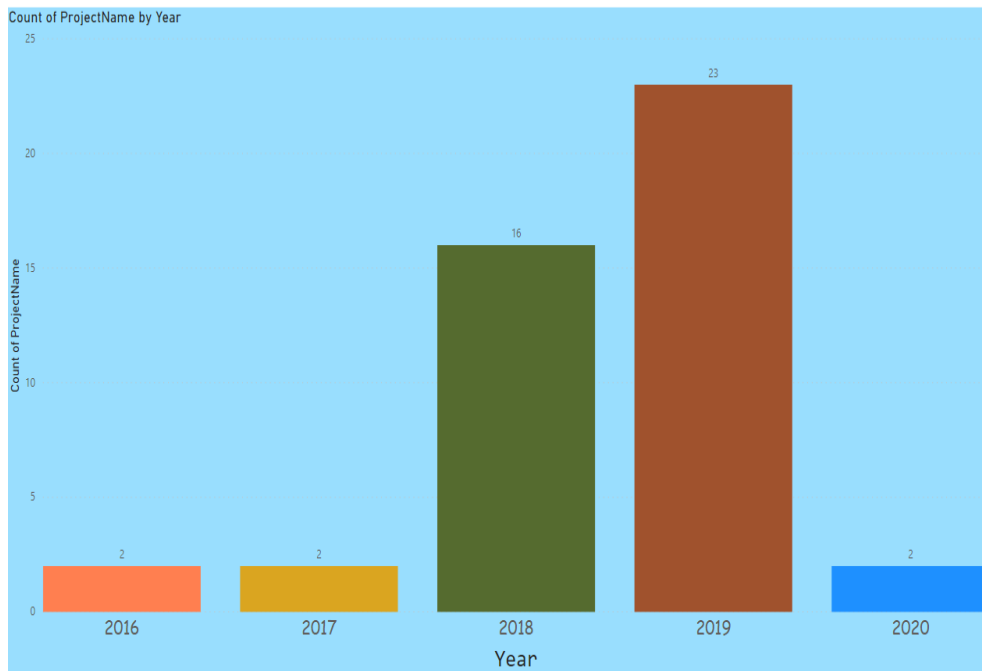


Fig.9.7. Report – Page 6

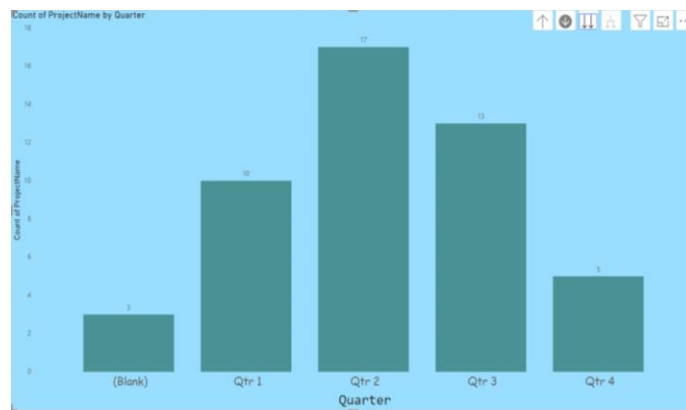


Fig.9.8. Quarter view

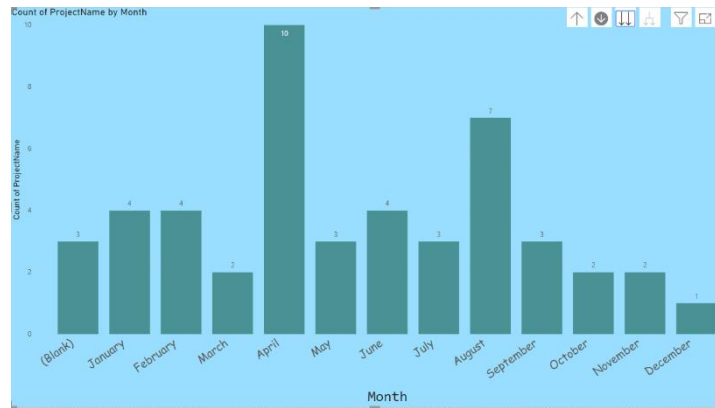


Fig.9.9 Month view

Fig.9.7. depicts the a chart which calculates the number of projects initiated for a year plotted against the year. This chart has a special feature called the drill down feature. When you initially view it, it shows the number of projects plotted against the year. Now when you drill down the data, it shows the number of projects initiated per quarter. When you further drill down, it shows the number of projects initiated per month. When you specifically click on a year in the top most drill, then when you drill down, it displays the data for the projects initiated in that particular year per quarter and per month. This enables the viewer to easily analyse the flow of projects since the beginning till the current date.

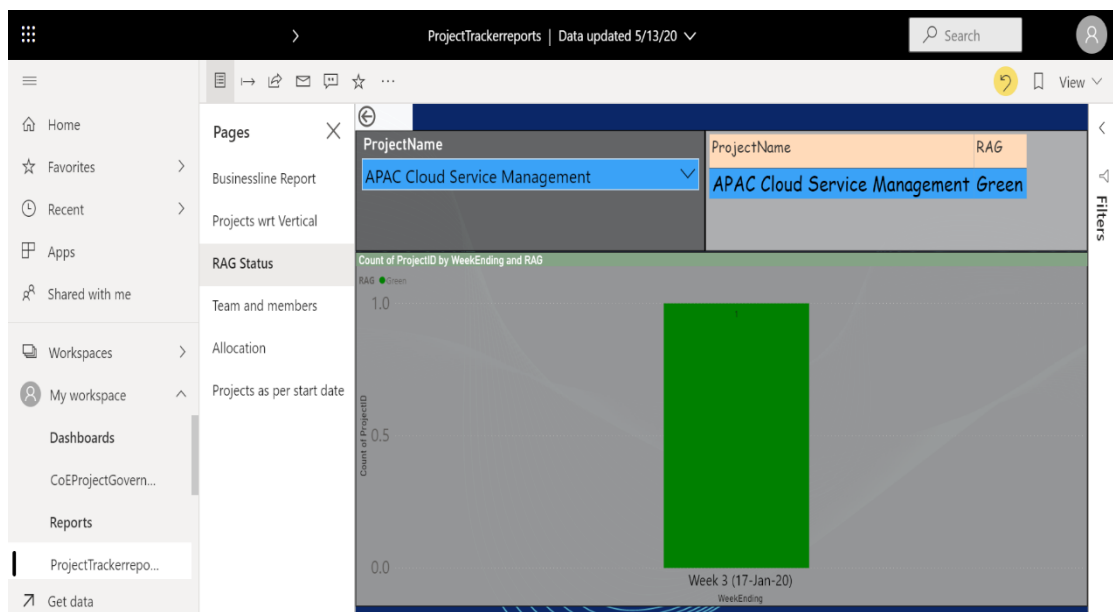


Fig.9.9. Report – full view.

Fig.9.9 displays the complete view of the reports when published on the Power BI platform. The pages pane towards the left corner of your screen allows you to navigate to any page where you would like to visit to analyse the data. Currently, it is in page 3 which displays the RAG status of a particular team. Now when you click against any of the page name, it navigates to that particular page displaying the data upon selection, if present, or just displays it. The main tiles from the reports have to be pulled to the dashboard in order to create a dashboard. Hence, all sort of editing is possible in the report view only. From the report pages, you can decide on the important KPI's which you would like to display in the dashboard.

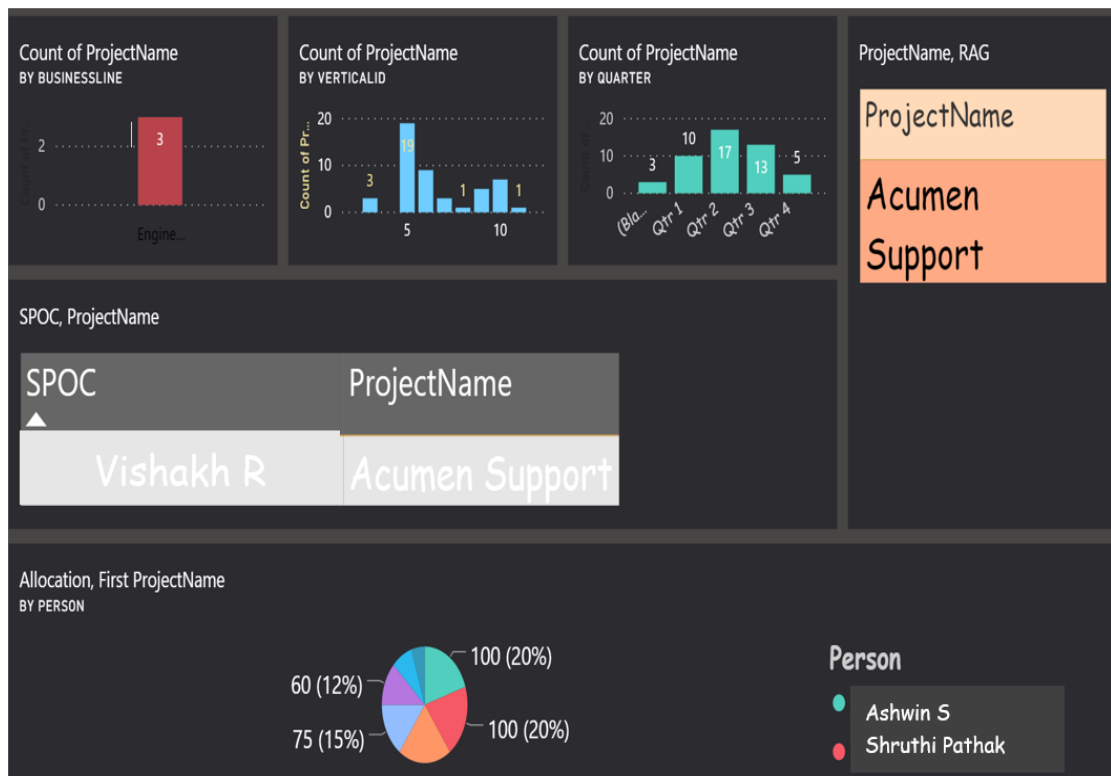


Fig.9.10. Dashboard – Full view

Fig.910. shows the full view of the dashboard. The tiles in the dashboard are the major KPI's which one would like to analyse while looking through the company information. When clicked on any of the tiles, it automatically navigates to the page where information pertaining to it is present.

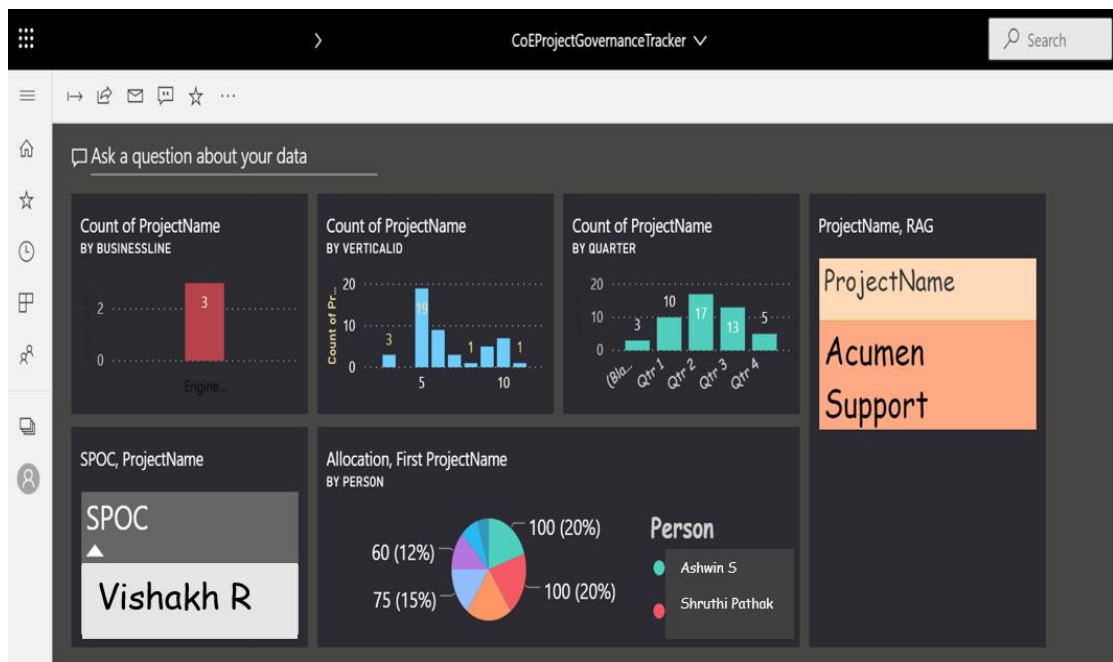


Fig.9.11. Dashboard – normal view

Fig.9.11. depicts the normal view of the dashboard. Here, there's a ask a question about your data space, where when you enter something or pick upon some of the suggested questions, it displays data associated with it. Thus, the entire objective of the project to process and display data through a visual tool is done as mentioned. The tiles present in the dashboard are not fixed ones. One can even change the tiles, reshape the tiles, change the theme of the tiles as desired.

REFERENCES

- 1) R. Somya, D. Manongga and M. A. Ineke Pakereng, "Service-Oriented Business Intelligence (SoBI) for Academic and Financial Data Integration in University," *2018 International Seminar on Application for Technology of Information and Communication*, Semarang, 2018, pp. 1-5
- 2) D. Wang, "The Realization of Integration for Business Intelligence between SQL Server 2005 Reporting Services and Office SharePoint Server 2007," *2010 International Conference on Challenges in Environmental Science and Computer Engineering*, Wuhan, 2010, pp. 274-277
- 3) E. Pourabbas and A. Shoshani, "The Composite Data Model: A Unified Approach for Combining and Querying Multiple Data Models," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 5, pp. 1424-1437, 1 May 2015
- 4) B. Barabino, M. Di Francesco and S. Mozzoni, "An Offline Framework for Handling Automatic Passenger Counting Raw Data," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2443-2456, Dec. 2014
- 5) O. Moscoso-Zea, J. Paredes-Gualtor and S. Luján-Mora, "A Holistic View of Data Warehousing in Education," in *IEEE Access*, vol. 6, pp. 64659-64673, 2018
- 6) D. Petković, "Comparison of Different Solutions for Solving the Optimization Problem of Large Join Queries," *2010 Second International Conference on Advances in Databases, Knowledge, and Data Applications*, Menuires, 2010, pp. 51-55
- 7) Q. Hanlin, J. Xianzhen and Z. Xianrong, "Research on Extract, Transform and Load(ETL) in Land and Resources Star Schema Data Warehouse," *2012 Fifth International Symposium on Computational Intelligence and Design*, Hangzhou, 2012, pp. 120-123