

Design Document – Power BI Migration

Prepared For: Ministry of commerce

Document Type: Design document.

Customer: Ministry of commerce

Doc ID: I360- MCPBIM-PH2-DD-V01

Date: 22/02/2024

Author: Aymen ANNOUN

1. Document History

1.1. Change Record

Date	Reviewer	Version	Position
21-Feb-2024	Aymen ANNOUN	1	Data Analytics Specialist

1.2. Reviewers

Date	Reviewer	Version	Position
22-Feb-2024	Wesam Habboub	1	Chief Consultant

2. Confidentiality Agreement

The recipient of this document agrees that the information provided within this document is to be considered confidential and proprietary to Insight 360. Recipient shall not use the confidential information other than for the purposes of its business with Insight 360, and shall disclose it only to its officers, directors, or employees with a specific need to know. The recipient will not disclose, publish, or otherwise reveal any of the confidential information received from Insight 360 to any other party whatsoever except with the specific prior written authorization of Insight 360.

All documents provided by Insight 360 are copyrighted and/or protected from any changes or modifications or any other kind of use outside the scope defined between Insight 360 and the Recipient.

Insight 360 retains all title, ownership and intellectual property rights to the information contained herein, including all supporting documentation. By acceptance of this document, the recipient agrees to be bound by the statement.

Copyright © Insight 360, 2024

3. Table of Contents

1. <i>Document History</i>	2
1.1. Change Record	2
1.2. Reviewers	2
2. <i>Confidentiality Agreement</i>	3
3. <i>Table of Contents</i>	4
4. <i>Executive Summary</i>	7
5. <i>Glossary of Terms</i>	8
6. <i>Solution Architecture</i>	9
6.1. Data Sources and Initial Data Ingestion	10
6.2. Landing, Staging, and Data Warehouse Databases	10
6.3. SQL Server Integration Services (SSIS)	10
6.4. SQL Server Analysis Services (SSAS) Servers	10
6.5. Power BI Report Server (PBIRS) Nodes	11
6.6. End-User Interaction and Report Subscription	11
6.7. SMTP Server Integration for Report Distribution	11
7. <i>Data Warehouse Design (Commercial registries)</i>	12
Multidimensional Design	12
7.1.1. FACT_CR_PARTY_SNAPSHOT:.....	12
7.1.2. FACT_CR_PARTY_AUDIT	13
7.1.3. FACT_CR_ACTIVITY_SNAPSHOT.....	14
7.1.4. FACT_CR_ACTIVITY_AUDIT	14
7.2. Data migration	15
7.3. CRs ETL	16
8. <i>Data warehouse design (other business areas)</i>	17
Multidimensional Design	17
8.1.1. Commercial Operations	17
8.1.2. Partners decisions.....	21
8.1.3. Customers services and branches.....	22
8.1.4. Correspondences.....	26
8.1.5. Committees.....	27

8.1.6.	BASHER and IT.....	28
8.1.7.	MC Performance.....	29
8.2.	Naming Convention	30
8.2.1.	Dimension Table Name.....	30
8.2.2.	Fact Table Name	30
8.2.3.	Column Name	30
8.2.4.	Surrogate Key.....	31
8.2.5.	Business Key.....	31
8.2.6.	Name Column	31
8.2.7.	Translation	31
8.2.8.	Other Fields.....	31
8.2.9.	Role Playing.....	32
8.2.10.	Dates in Fact Tables	32
8.2.11.	Junk Dimension.....	32
8.3.	Design Patterns and Best Practices	33
8.3.1.	Surrogate Keys	33
8.3.2.	Slowly Changing Dimensions	33
8.3.3.	Audit.....	33
8.3.4.	Degenerate Dimensions.....	33
8.3.5.	Snow Flaking and Outrigger Dimensions	34
8.3.6.	Many to Many between Fact-Dimension.....	34
8.3.7.	Many to Many between Dimensions.....	34
8.3.8.	Hierarchy.....	35
8.3.9.	Junk Dimensions	36
8.3.10.	NA Records Management.....	36
9.	Data Integration Design.....	37
Data Sources	37	
9.1.	ETL process	39
9.1.1.	Historical load	40
9.1.2.	Incremental load	40
9.2.	ETL Naming Convention	41
9.2.1.	Master Packages	41
9.2.2.	Landing Packages.....	41
9.2.3.	Staging Packages.....	41

9.2.4.	DWH packages (Dimension/Fact Packages).....	42
9.2.5.	Audit.....	42
9.3.	Landing ETL.....	42
9.4.	Staging ETL.....	43
9.5.	DWH ETL.....	45
10.	<i>ETL Deployment and jobs scheduling.....</i>	48
	ETL Deployment	48
	ETL Scheduling.....	49
11.	<i>Data integration monitoring.....</i>	50
12.	<i>Dashboards Deployment.....</i>	51
13.	<i>Security Access Design</i>	52
	Portal Level Access:	52
	Folder Level Access:.....	52
	Dashboard/Report Level Access:.....	53
	Row Level Security (RLS):	53
8.	<i>Test Plan and Test Cases</i>	55

4. Executive Summary

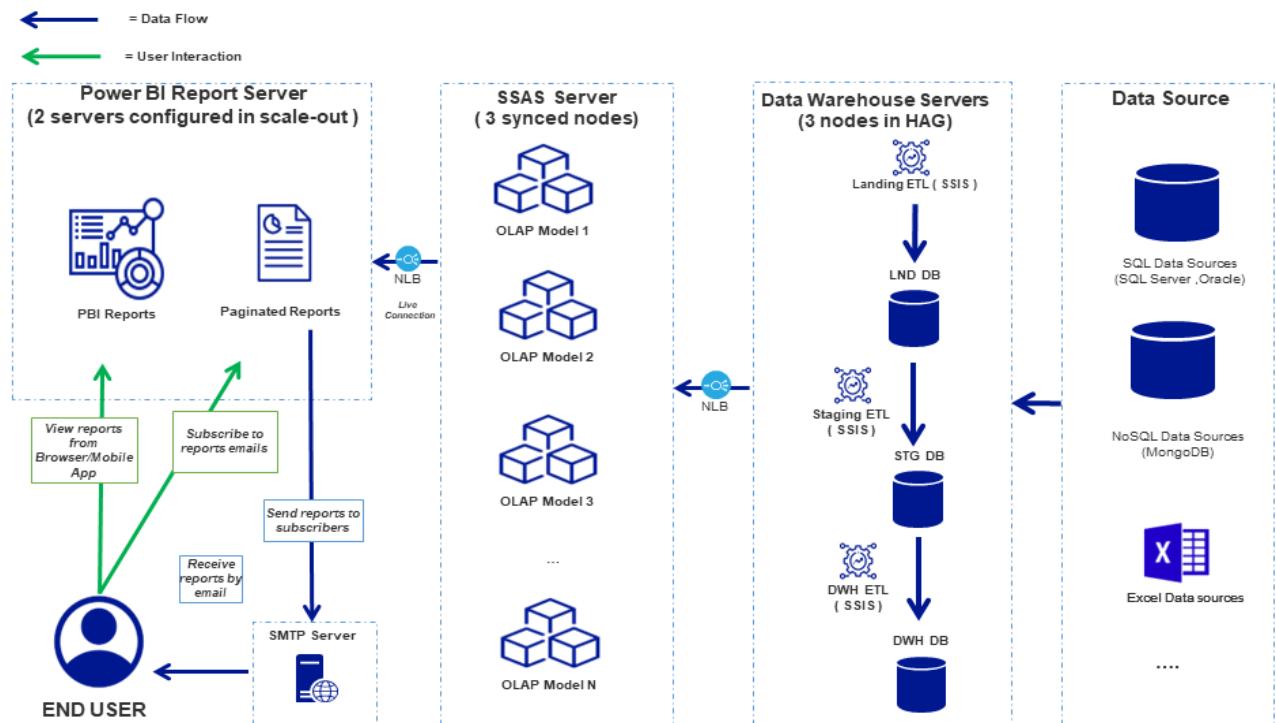
This document outlines the comprehensive design of our Business Intelligence (BI) solution, encompassing ETL processes, OLAP systems, and Power BI reporting. It details the data warehouse design following Kimball's Multidimensional Modeling approach, ensuring efficient data handling and storage. The ETL section highlights strategies for data extraction, transformation, and loading, focusing on performance and integrity. The OLAP portion covers the configuration of SQL Server Analysis Services for advanced data analysis. Finally, the use of Power BI for dynamic reporting and visualization is discussed, emphasizing efficient data access and insightful reporting across business domains. This summary provides an overview of how each BI component integrates to form a cohesive, effective system.

5. Glossary of Terms

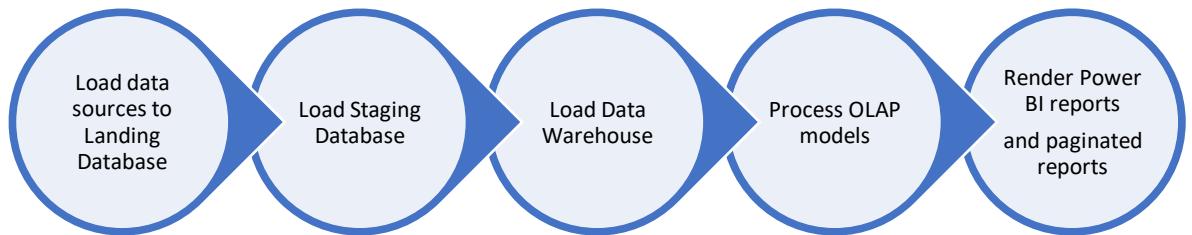
TERM	DEFINITION
BI	BUSINESS INTELLIGENCE
ETL	EXTRACT, TRANSFORM, LOAD
OLAP	ONLINE ANALYTICAL PROCESSING
PBI	POWER BI
DWH	DATA WAREHOUSE
LND	LANDING DATABASE
STG	STAGING DATABASE
SMTP	SIMPLE MAIL TRANSFER PROTOCOL
SSAS	SQL SERVER ANALYSIS SERVICES
PBIRS	POWER BI REPORT SERVER
NLB	NETWORK LOAD BALANCER
PBIRS	Power BI Report Server
SSRS	SQL SERVER REPORTING SERVICES
DR	DISASTER RECOVERY

6. Solution Architecture

The below diagram describes the solution architecture with all components of the solution:



In a simple way, we can present the architecture this way



6.1. Data Sources and Initial Data Ingestion

The Business Intelligence (BI) architecture incorporates a variety of data sources, including SQL Server, Oracle, MongoDB, and Excel and CSV files. These diverse data inputs are essential for feeding the comprehensive data processing system.

6.2. Landing, Staging, and Data Warehouse Databases

Within the three SQL Server Nodes configured for high availability, the first database serves as the landing database for incoming data. This database is critical for the initial collection and aggregation of raw data from various sources. After initial collection, data is moved to the staging database, located within the same SQL Server Nodes. Here, it undergoes essential transformations and quality enhancements. Finally, data is transferred to the Data Warehouse (DWH) database, also within these nodes, where it is structured and optimized for analysis.

6.3. SQL Server Integration Services (SSIS)

SSIS plays a key role in the ETL (Extract, Transform, Load) operations, facilitating the efficient movement and transformation of data. SSIS is utilized for automating data management tasks, integrating data from different sources, and processing complex data transformations before loading the data into the Data Warehouse.

6.4. SQL Server Analysis Services (SSAS) Servers

The system includes three SSAS servers dedicated to handling tabular models. The primary SSAS server processes these models by reading from the DWH Listener, while the other two SSAS servers are synchronized with the primary, ensuring consistent data modeling.

6.5. Power BI Report Server (PBIRS) Nodes

For reporting, two PBIRS nodes are deployed in a scale-out deployment. These nodes manage and deliver Power BI and paginated reports. They are connected to a Network Load Balancer (NLB) that retrieves data from the synchronized SSAS servers, ensuring efficient data access for reporting.

6.6. End-User Interaction and Report Subscription

End-users access reports and dashboards through an NLB configured for PBIRS, providing streamlined access to business intelligence insights.

6.7. SMTP Server Integration for Report Distribution

An SMTP server is integrated into the architecture to enhance the distribution of paginated reports. This server sends paginated reports to PBIRS subscribers, streamlining the process of report distribution. Subscribers can receive reports directly via email, making it easier for users to access critical business intelligence insights. Additionally, this setup allows for the subscription to SSRS reports, further enhancing the accessibility and distribution of vital data.

7. Data Warehouse Design (Commercial registries)

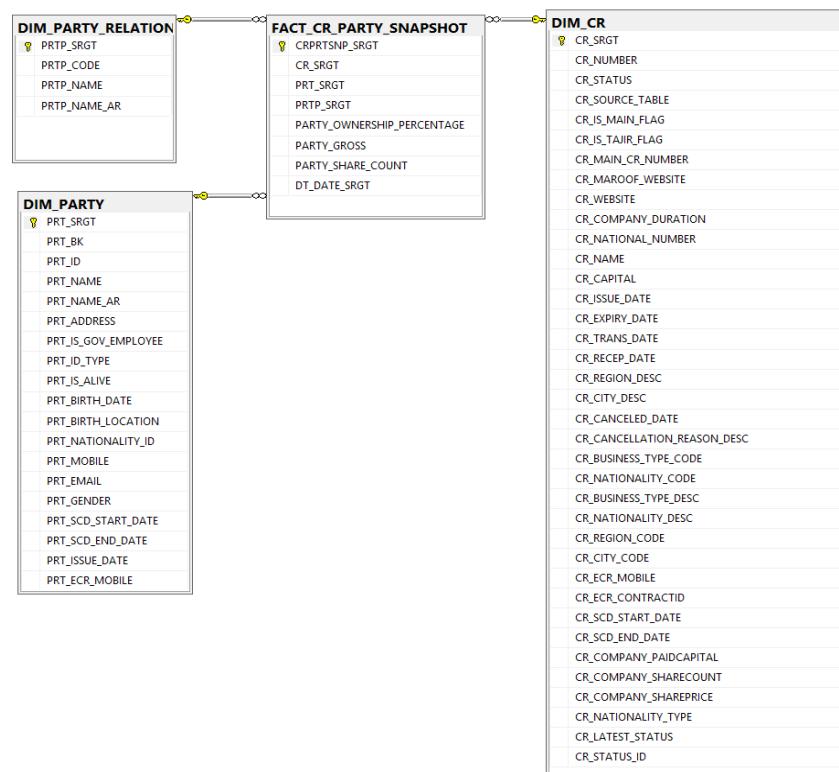
Multidimensional Design

The below will detail the multidimensional design of every star schema that's related to CRs.

The detail of this design especially for the dimension design will be reflected in the meta data excel sheet that will include all the fields, their description, and data lineage with the source systems.

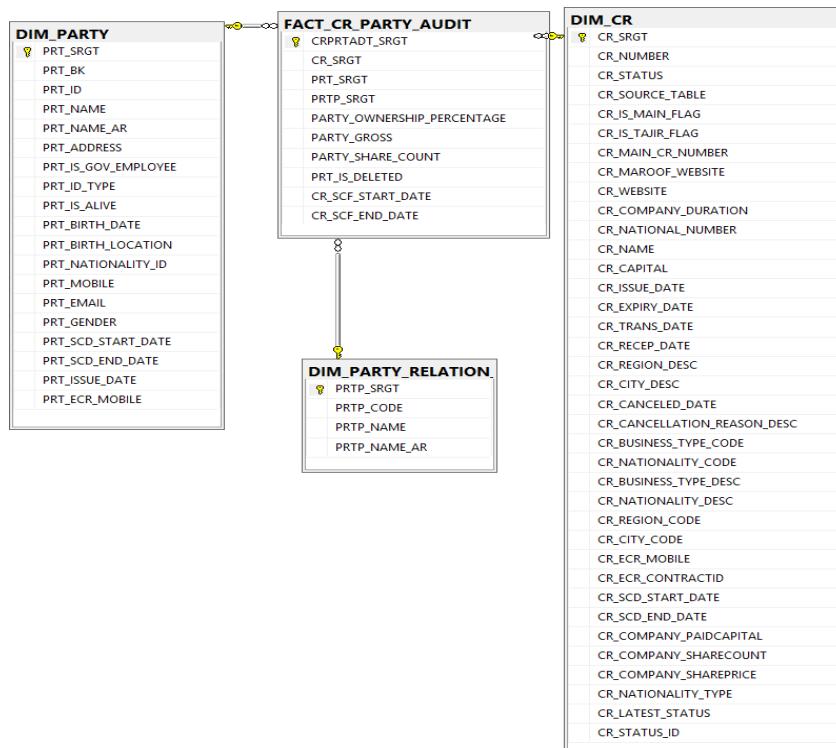
7.1.1. FACT_CR_PARTY_SNAPSHOT:

A Monthly Snapshot table that is about the Contributing Parties and their relationship with the different CRs.



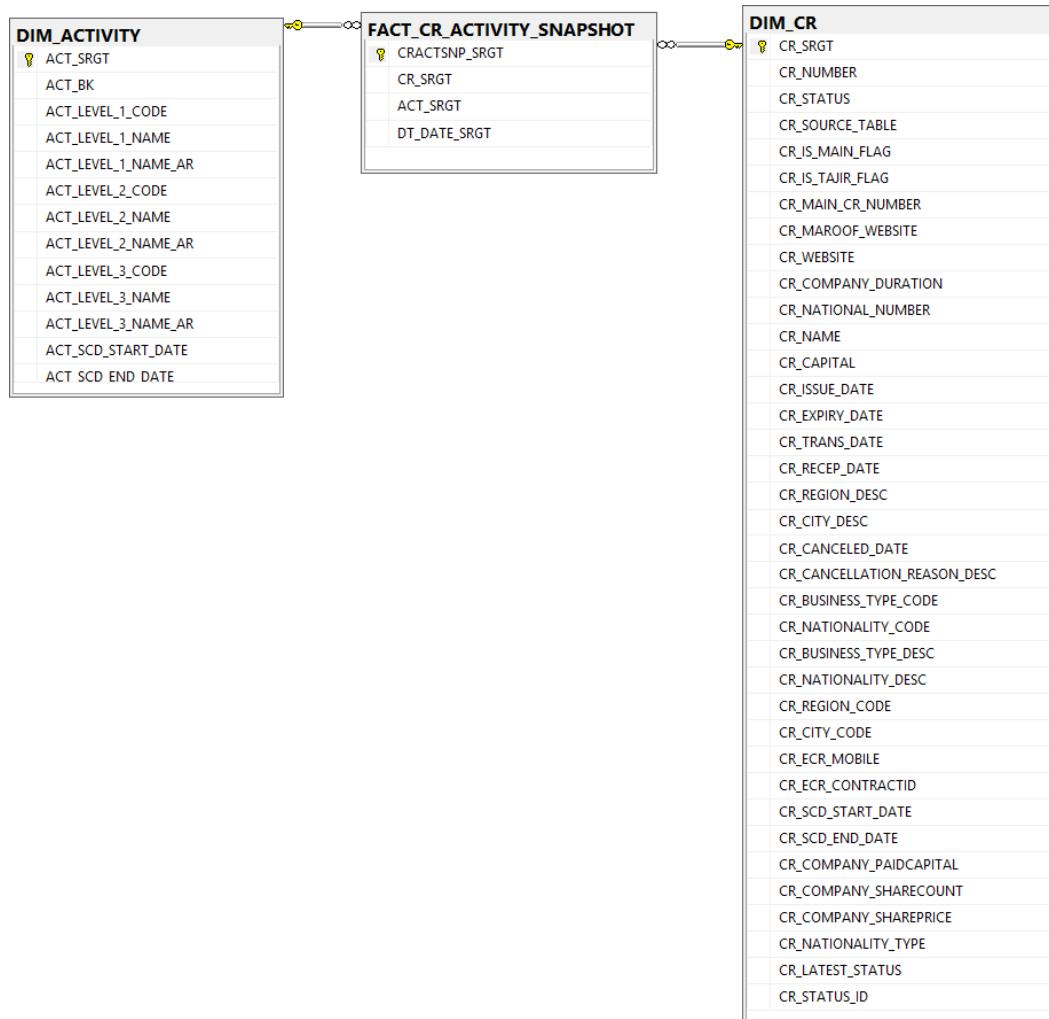
7.1.2. FACT_CR_PARTY_AUDIT

A historical tracker for information about related parties and their relationships with different CRs, it is a type 2 slowly changing fact. It has two date columns, DT_DATE_SCF_START_DATE and DT_DATE_SCF_END_DATE, for tracking history. Additionally, a binary flag column (IS_DELETED) has been introduced to indicate whether a relationship between a Party and a CR is active or not for the period delimited by the date columns for the row in question.



7.1.3. FACT_CR_ACTIVITY_SNAPSHOT

A Monthly Snapshot table that captures the information about the activities of the CRs.



7.1.4. FACT_CR_ACTIVITY_AUDIT

This table contains information about the activities of CRs. It is a type 2 slowly changing fact that includes two date columns, DT_DATE_SCF_START_DATE and DT_DATE_SCF_END_DATE, for historical tracking. Additionally, a binary flag column (IS_DELETED) has been introduced to indicate whether an Activity-CR relationship is active or deleted for the period defined by the date columns for the specific row.



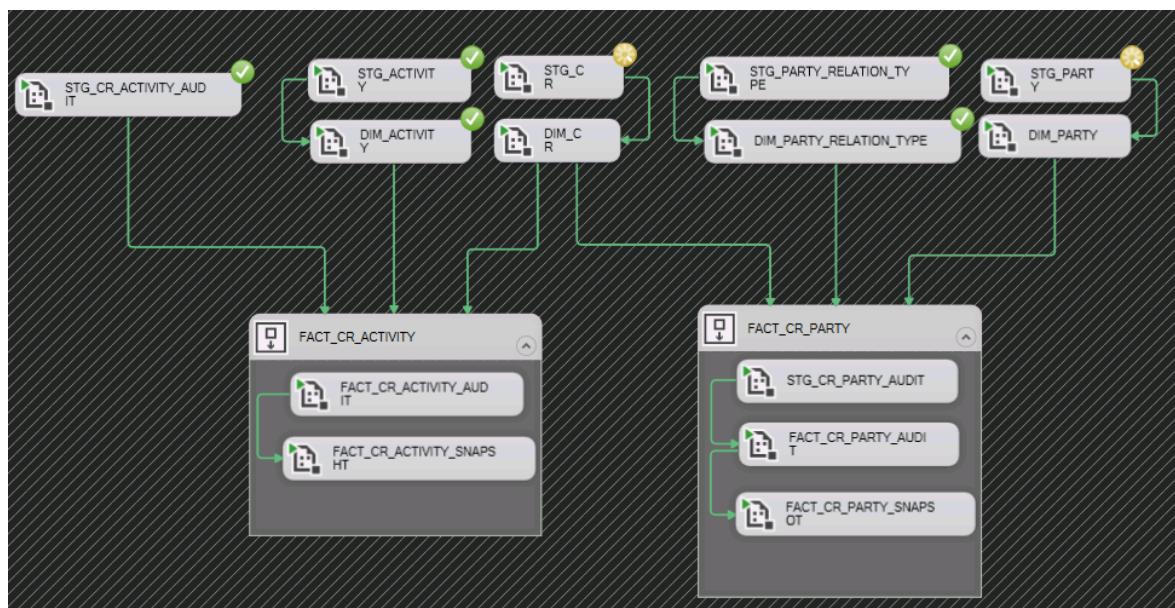
7.2. Data migration

Our primary objective was to ensure the comprehensive migration of all data from the old Data Warehouse to the new system. However, we faced significant challenges, particularly with the migration of parties and activities data. These challenges were largely due to the different design of the new Data Warehouse and the data quality associated with parties' information. Despite these obstacles, we placed a special emphasis on the most crucial component of this migration: the DIM_CR (Dimension Commercial Registry) which has been successfully migrated.

7.3. CRs ETL

The ETL process for the Commercial Registries (CRs) has been developed as an independent project, highlighting its critical significance and unique design requirements, including slowly changing dimensions, slowly changing facts, tracking deletions, and taking snapshots. This process begins by loading data from the Landing Database to the Staging Database, followed by sequentially loading the dimensions and then the fact tables.

The Master package oversees orchestrating the data flow from the Landing Database to Staging and subsequently to the Data Warehouse. This orchestration has been meticulously designed, as outlined below, to ensure that the dependencies among the various packages are carefully managed and respected, facilitating a smooth and efficient data loading process.



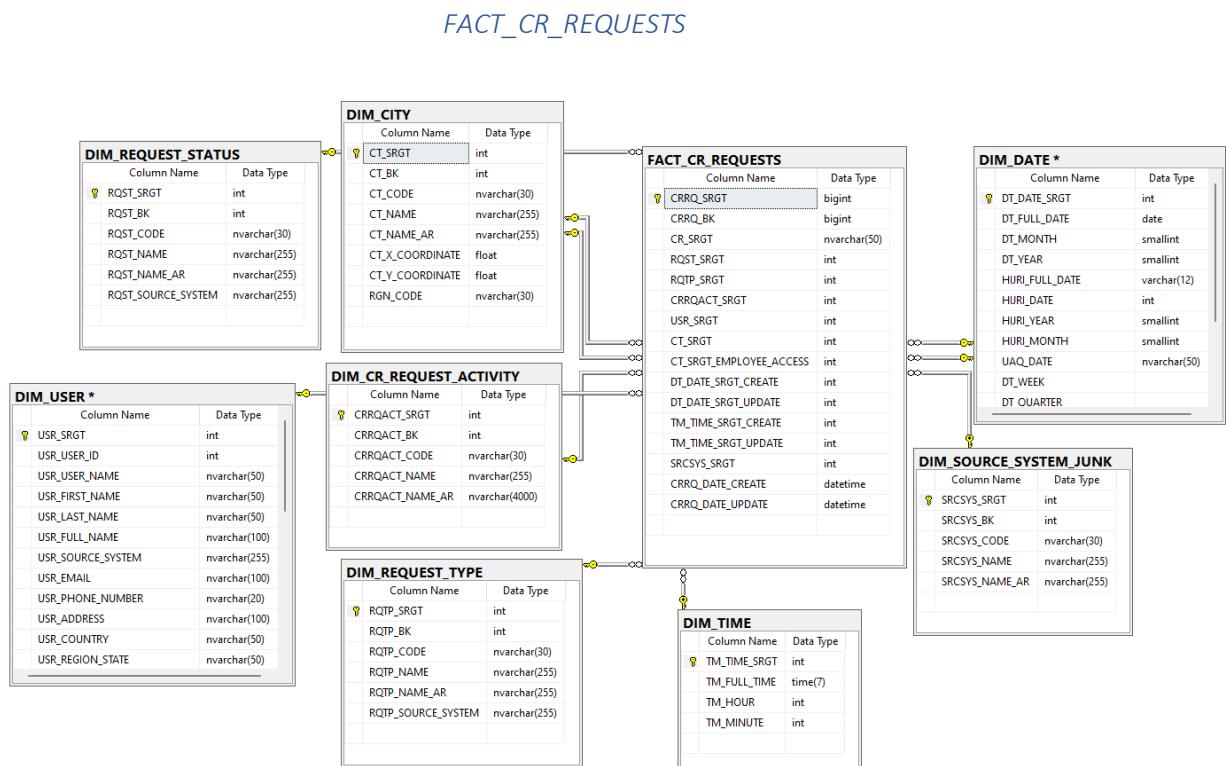
8. Data warehouse design (other business areas)

Multidimensional Design

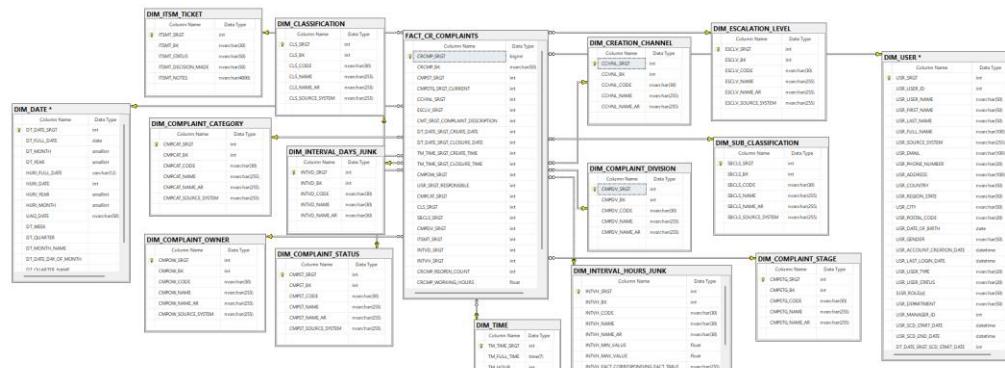
The below will detail the multidimensional design of every star schema in the data warehouse.

The details of this design, especially for the dimension design, will be reflected in the meta data excel sheet that will include all the fields, their descriptions, and their data lineage with the source systems.

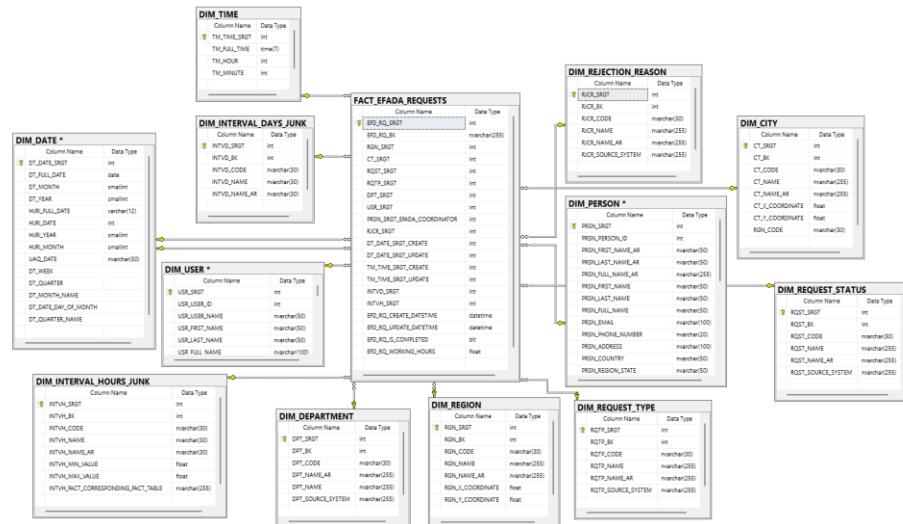
8.1.1. Commercial Operations



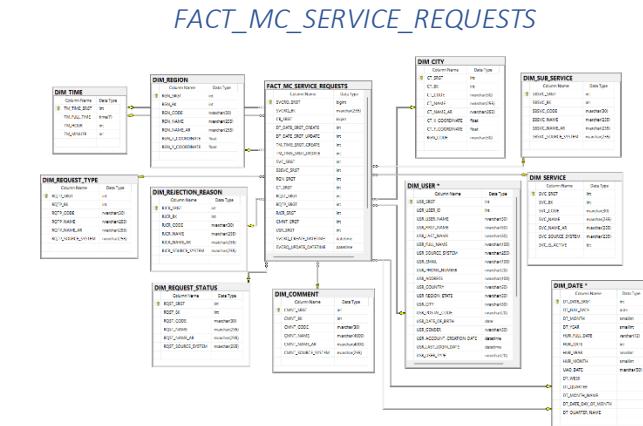
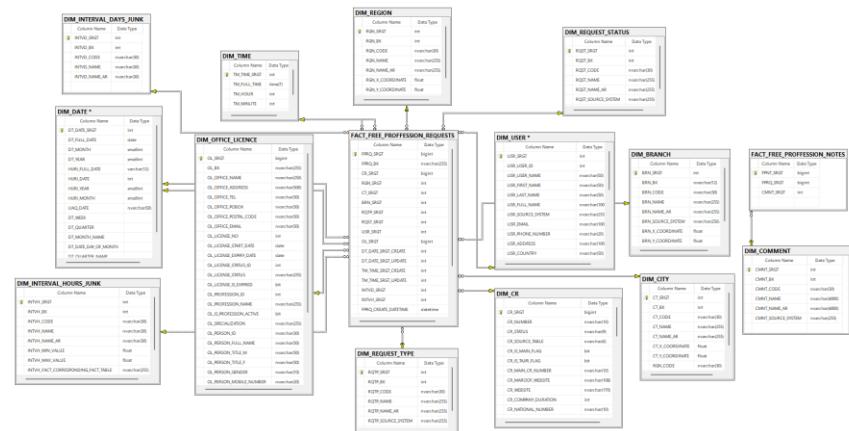
FACT_CR_COMPLAINTS



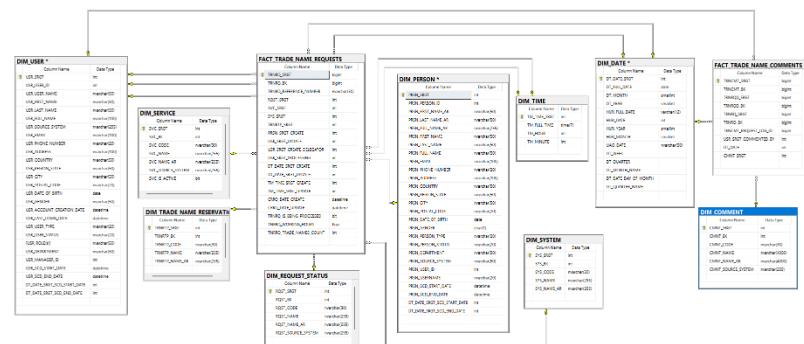
FACT_EFADA_REQUESTS



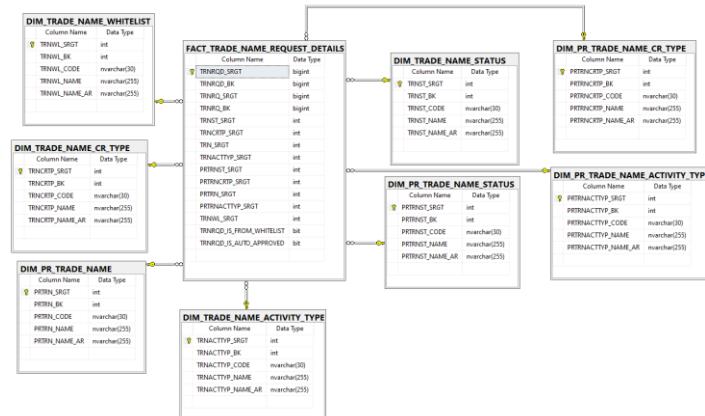
FACT_FREE_PROFESSION_REQUESTS and FACT_FREE_PROFESSION_NOTES



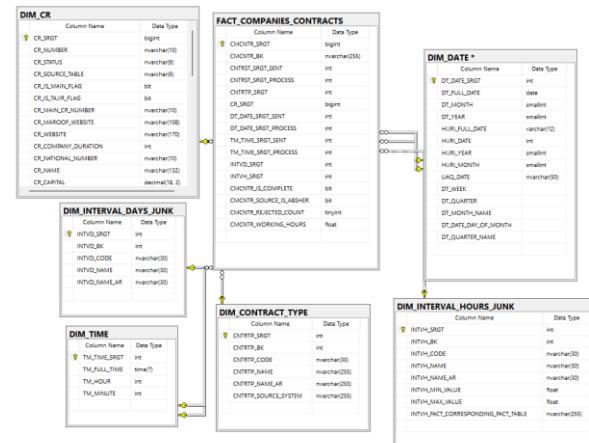
FACT_TRADE_NAME_REQUESTS and FACT_TRADE_NAME_COMMENTS



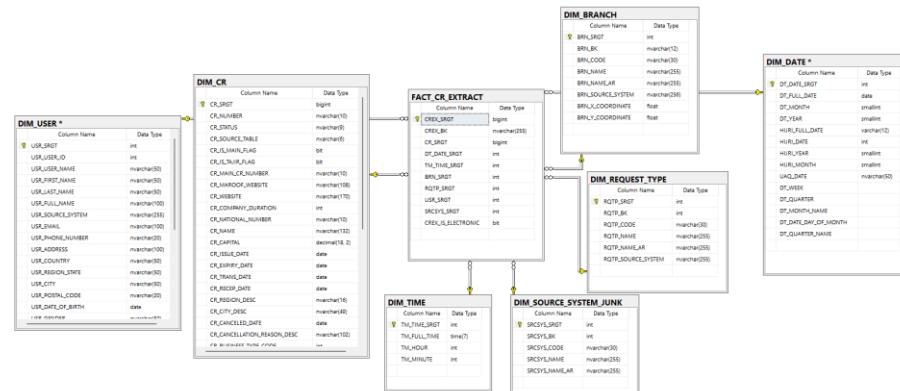
FACT_TRADE_NAME_REQUEST_DETAILS



FACT_COMPANIES_CONTRACTS

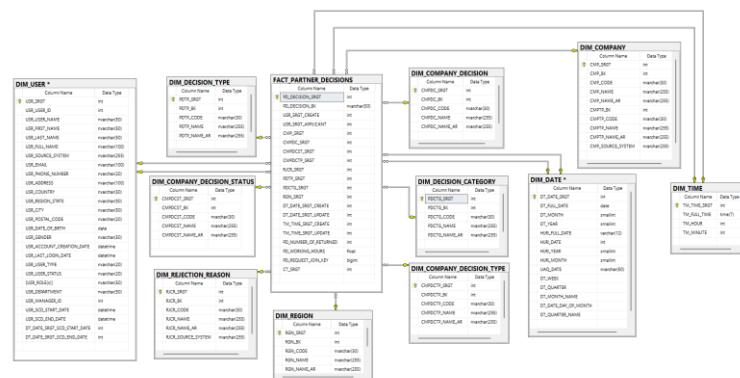


FACT_CR_EXTRACT

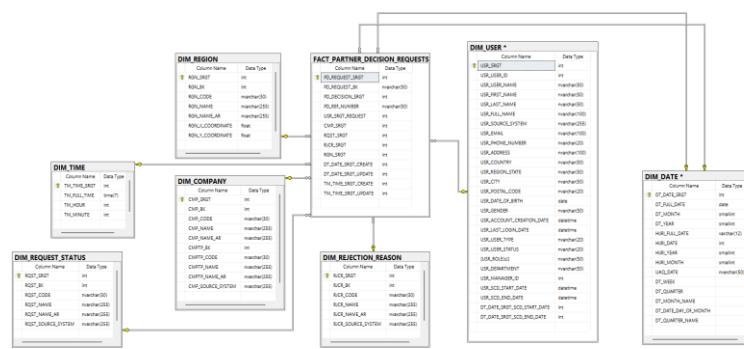


8.1.2. Partners decisions

FACT_PARTNER_DECISIONS

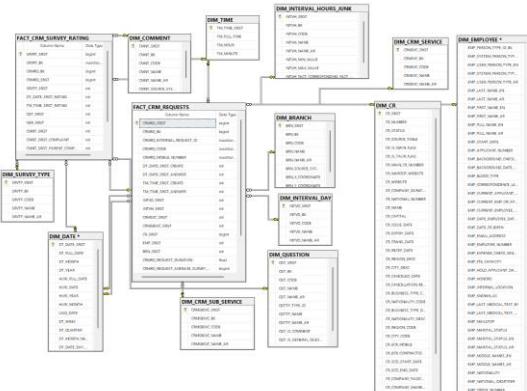


FACT_PARTNER_DECISION_REQUESTS

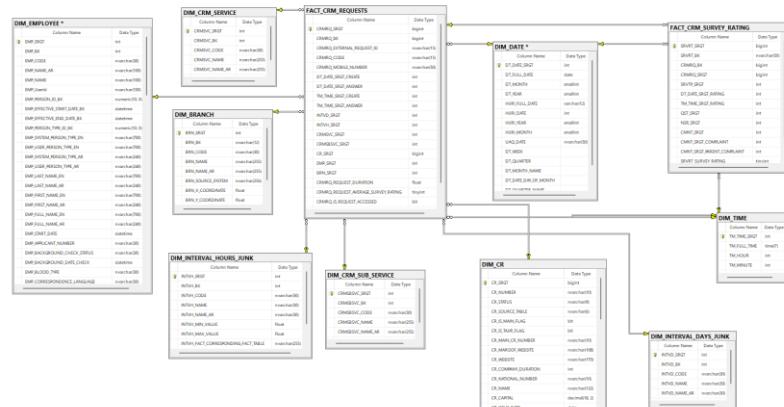


8.1.3. Customers services and branches

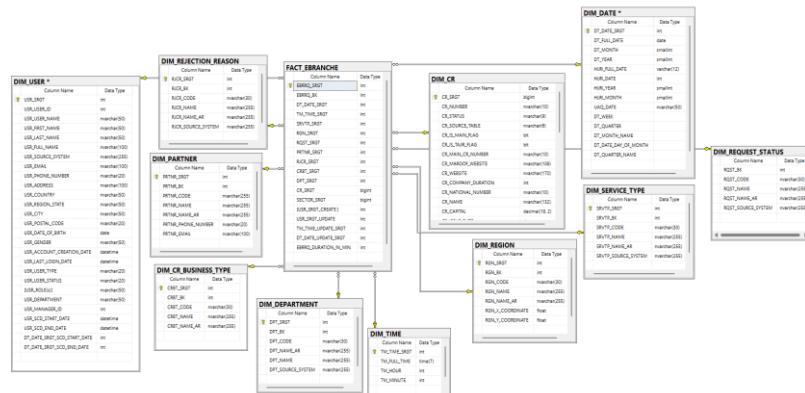
FACT_CRM_REQUESTS



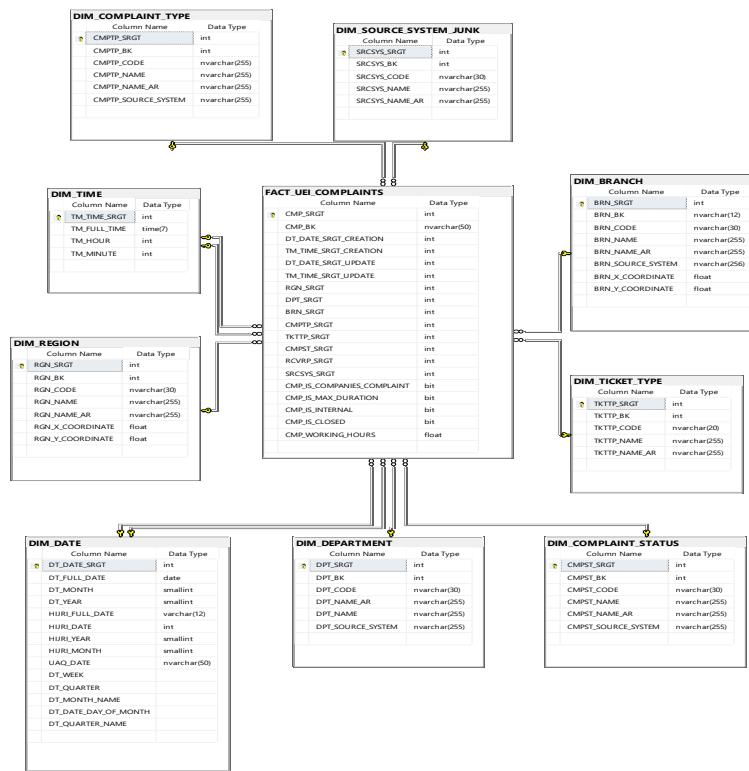
FACT_CRM_SURVEY_RATING



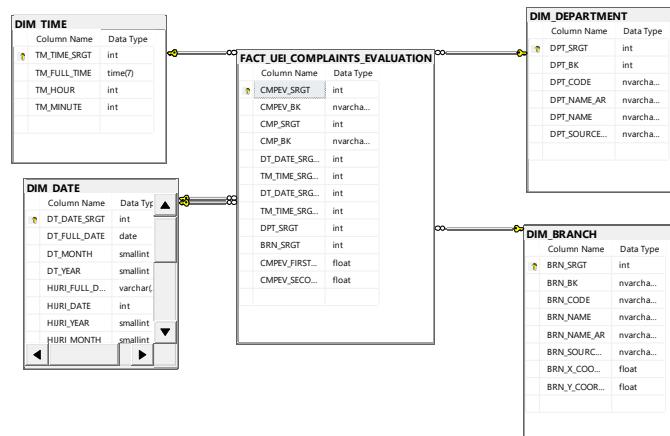
FACT_EBRANCHE



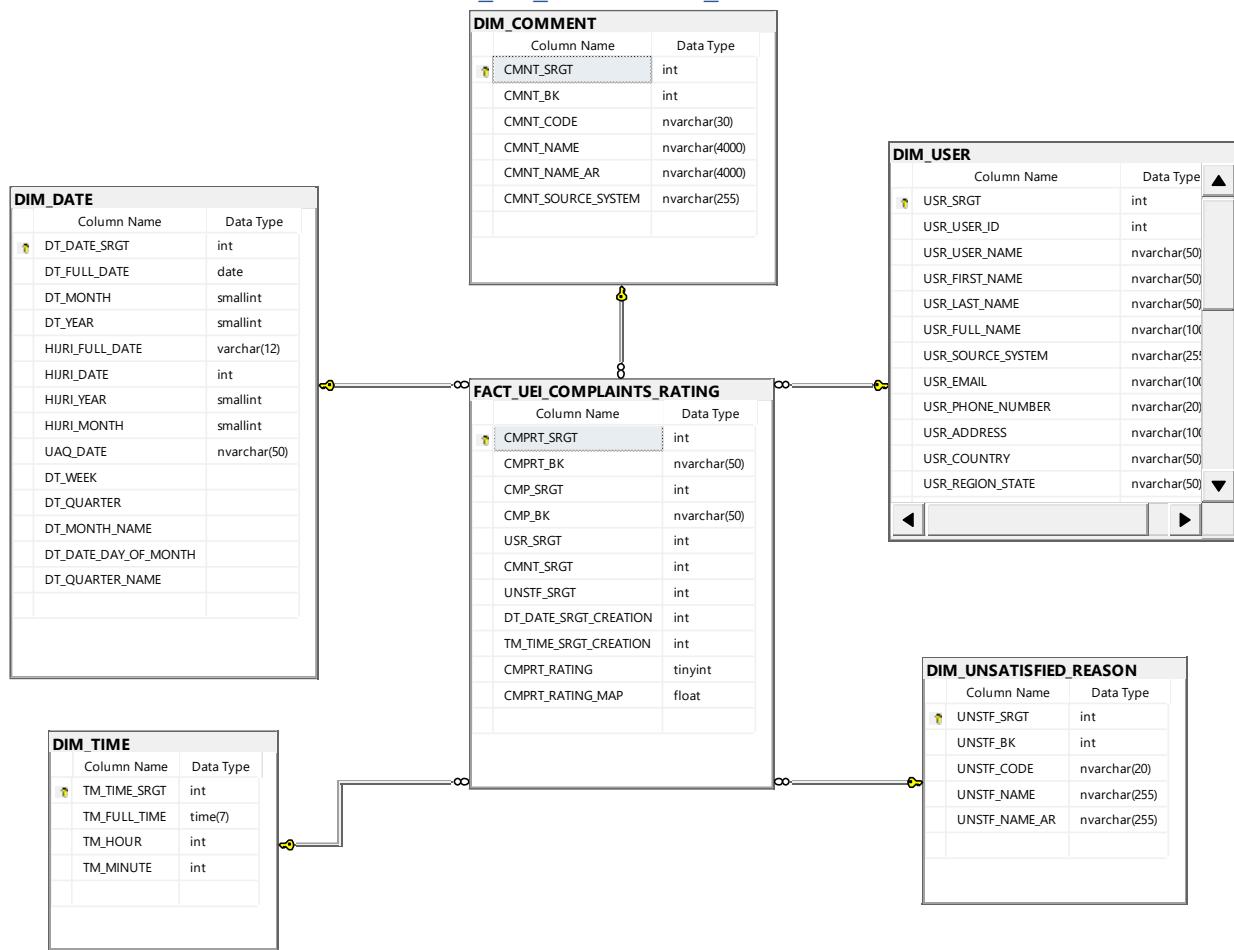
FACT_UEI_COMPLAINTS



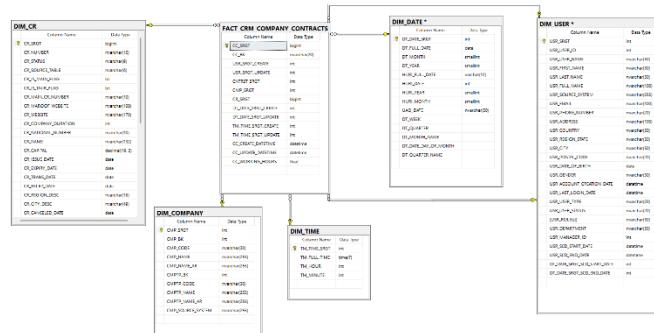
FACT_UEI_COMPLAINTS_EVALUATION



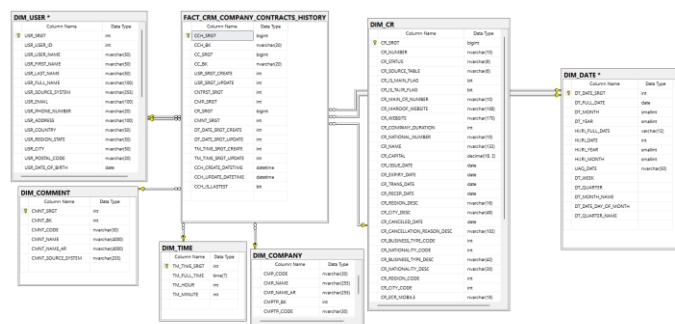
FACT_UEI_COMPLAINTS_RATING



FACT_CRM_COMPANY_CONTRACTS



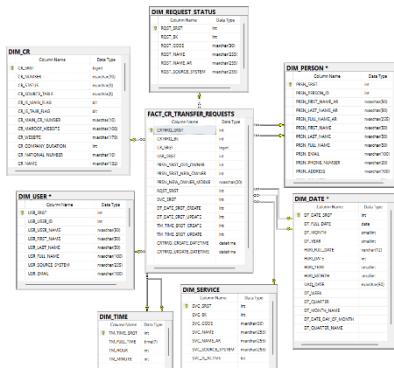
FACT_CRM_COMPANY_CONTRACTS_HISTORY



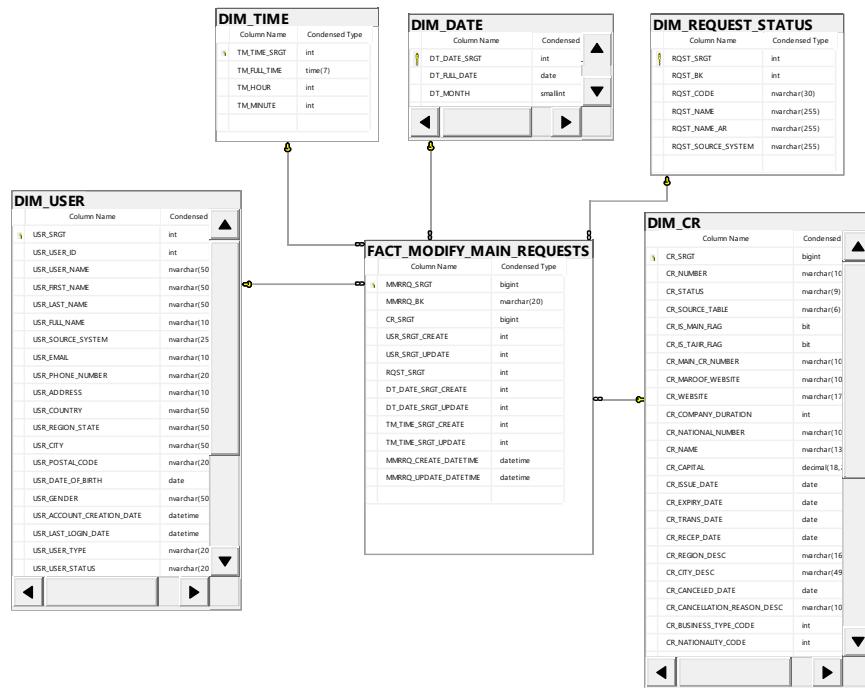
FACT_CONVERT_FIRM_CONTRACTS



FACT_CR_TRANSFER_REQUESTS

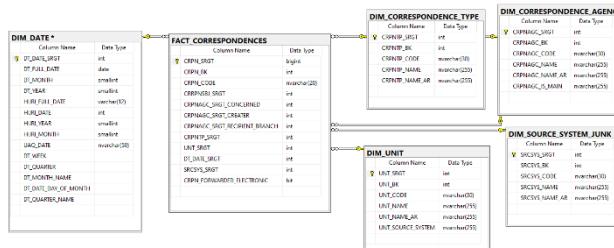


FACT MODIFY MAIN REQUESTS



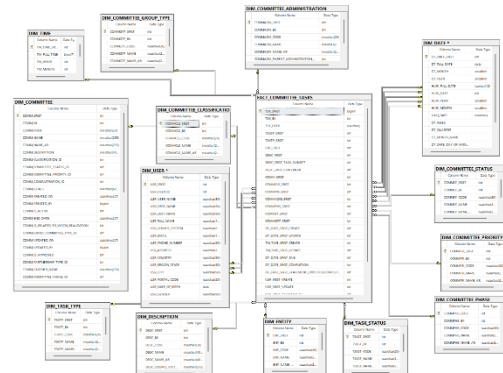
8.1.4. Correspondences

FACT_CORRESPONDENCES

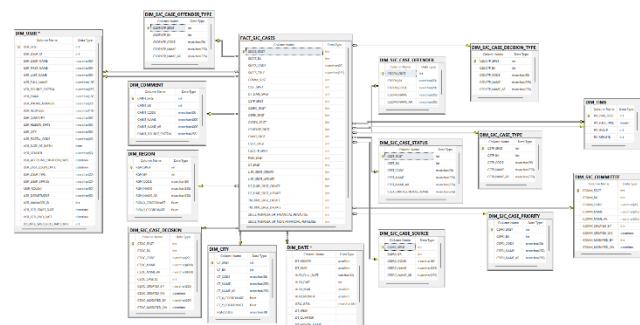


8.1.5. Committees

FACT_COMMITTEE_TASKS

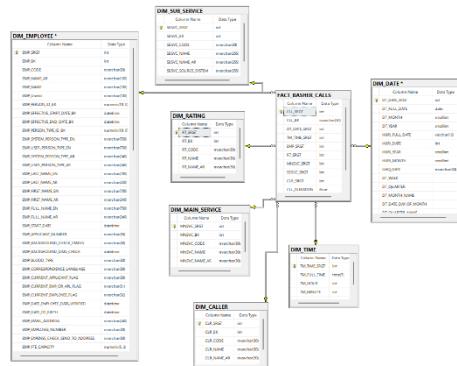


FACT_SJC_CASES

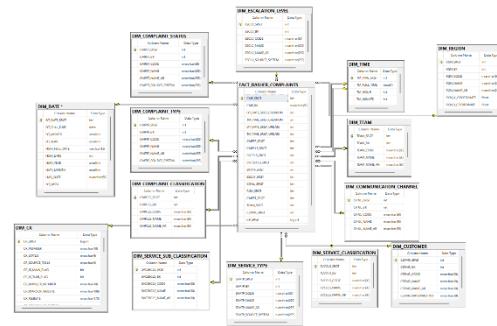


8.1.6. BASHER and IT

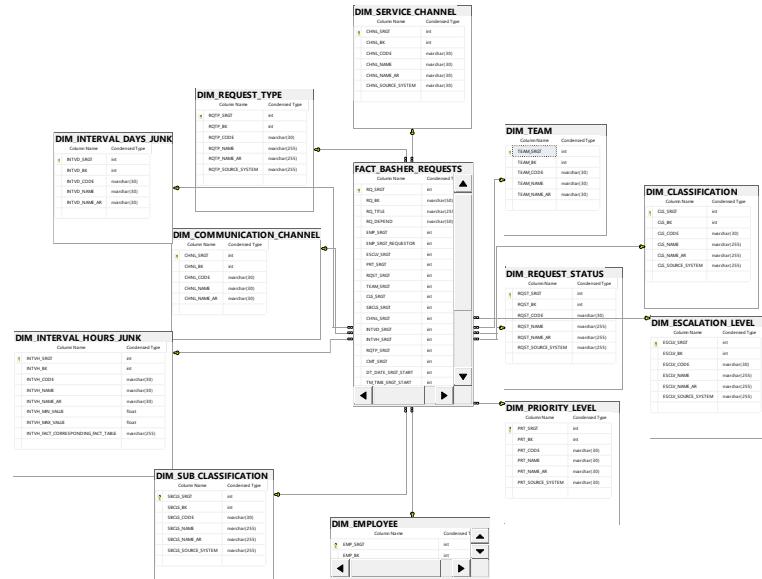
FACT_BASHER_CALLS



FACT_BASHER_COMPLAINTS

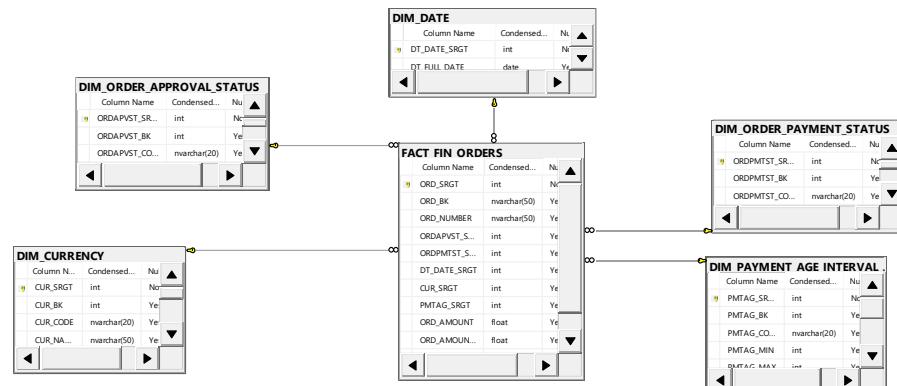


FACT_BASHER_REQUESTS

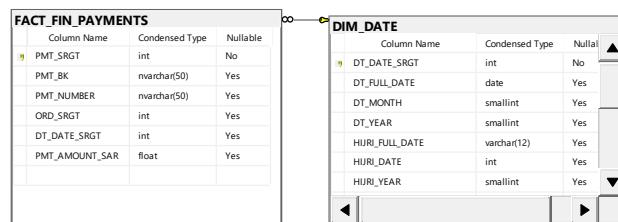


8.1.7. MC Performance

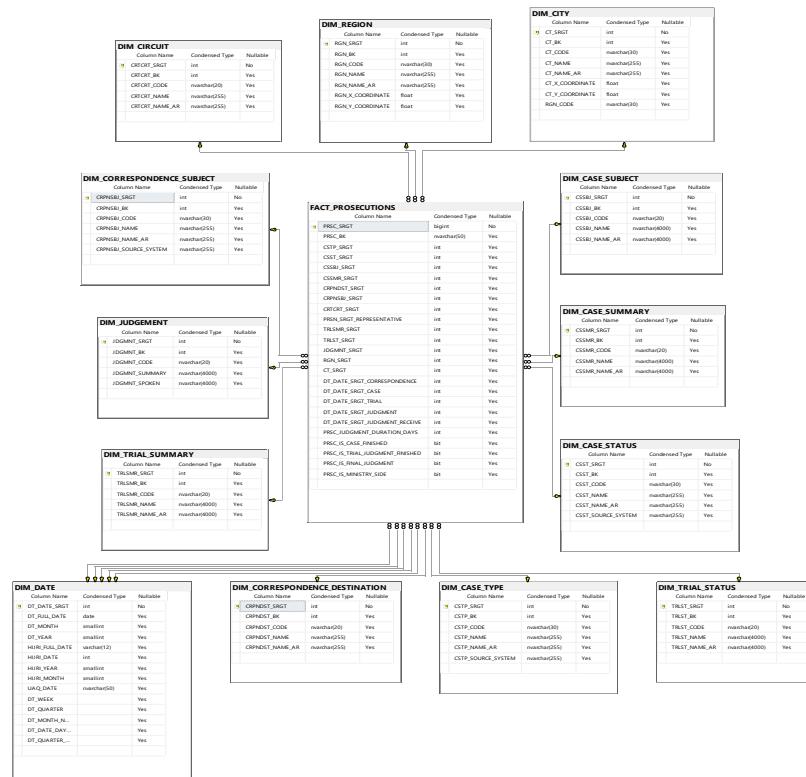
FACT_FIN_ORDERS



FACT_FIN_PAYMENTS



FACT_PROSECUTIONS



8.2. Naming Convention

8.2.1. Dimension Table Name

- Start with “DIM_”.
- No space will be used.
- Must be in Capital letter.
- Words will be separated by _.
- The Name of the table will always be in single form.
- Example: DIM_PRODUCT

8.2.2. Fact Table Name

- Start with “FACT_”.
- No space will be used.
- Must be in Capital letter.
- Words will be separated by _.
- The Name of the table will always be in single form.
- Example: FACT_PRODUCT

8.2.3. Column Name

- No Space will be used.
- Must be in Capital letter.
- Words will be separated by _.
- The column name will not be in plural.
- The Column Name will start with the abbreviation of the Table.
- Example: PRD_NAME_EN

8.2.4. Surrogate Key

- Dimension Table Name Abbreviation (without Dim/Fact) + _SRGT.
- Example: PRD_SRGT
- The surrogate keys will have the same name between all the tables (Fact and dimensions)

8.2.5. Business Key

- Table Name abbreviation + COLUMN_NAME +_BK
- Example: PRD_NO_BK

8.2.6. Name Column

- The name column of the dimension records should start always with the abbreviation of the table name.
- Example: PRD_NAME

8.2.7. Translation

- Add “_AR” for Arabic at the end of the filed name.
- Example: PRD_NAME_AR

8.2.8. Other Fields

- Each field should be preceded by the abbreviation of table name + “_” + same name as in the source database.

8.2.9. Role Playing

- The name of the Role-playing dimension will be as described below:
- DT_DATE_SRGT_<Role Name>
- Example: Completion Date → DT_DATE_SRGT_POSTING

8.2.10. Dates in Fact Tables

In some cases, in addition to having the SRGT of the date, it will be better to add the dates as datetime fields in the fact table to facilitate the calculations.

8.2.11. Junk Dimension

The Junk Dimension should be named as: DIM_Abbreviation of table name + '_' + JUNK.

8.3. Design Patterns and Best Practices

8.3.1. Surrogate Keys

We need to create a whole new set of keys in the data warehouse database, separate from the keys in the transaction source systems. We call these keys surrogate keys, although they are also known as meaningless keys, substitute keys, non-natural keys, or artificial keys. A surrogate key is a unique value, usually an integer, assigned to each row in the dimension. This surrogate key becomes the primary key of the dimension table and is used to join the dimension to the associated foreign key field in the fact table.

All the dimensions of the data warehouse must include a surrogate key.

8.3.2. Slowly Changing Dimensions

Type 2 Slowly Changing Dimension will be supported, so each SCD dimension will have by default 2 fields:

Example:

- CR_SCD_START_DATE
- CR_SCD_END_DATE

8.3.3. Audit

An Audit Dimension (DIM_AUDIT) will be maintained, for each pipeline execution a row will be added to this dimension including information about the pipeline, time, and records.

8.3.4. Degenerate Dimensions

If we encounter a field that represents a degenerate dimension like transaction Id, simply nothing is required and no need to create any dimension table related to it.

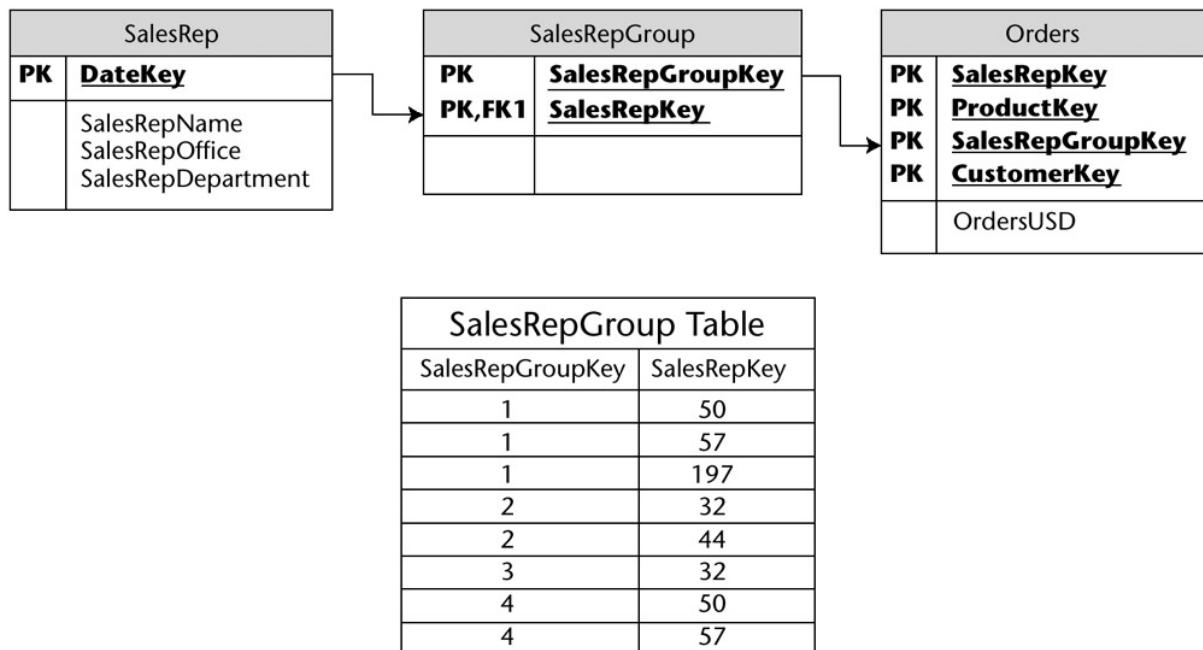
8.3.5. Snow Flaking and Outrigger Dimensions

Snow flaking is allowed for outrigger dimensions, so in a few cases we support the idea of connecting lookup or grouping tables to the dimensions. One of these cases involves rarely used lookups.

Example: Joining the Date table to the DateOfBirth field in the Customer dimension so we can count customers grouped by their month of birth.

8.3.6. Many to Many between Fact-Dimension

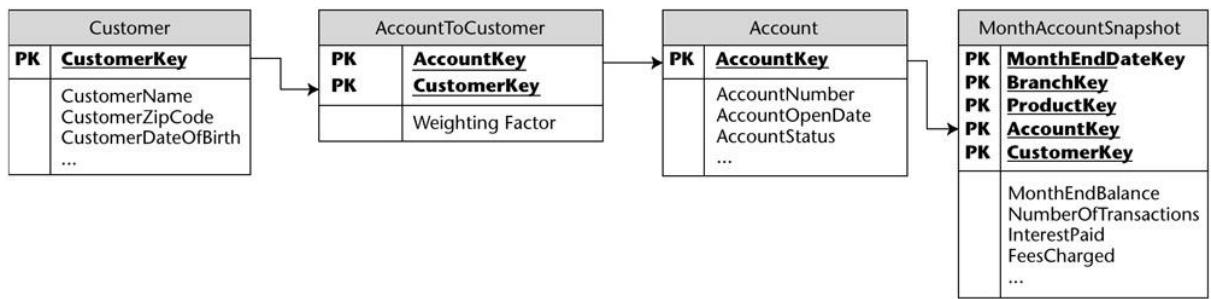
Many to many relations will be handled by adding a bridge table as described in the below example:



In this example each SalesRep might have multiple orders, and one order can be generated by multiple SalesReps.

8.3.7. Many to Many between Dimensions

Many to many relations will be handled by adding a bridge table as described in the below example:



In this example, each customer might have multiple Accounts and one accounts can belong to multiple customers

8.3.8. Hierarchy

Almost all the current BI tools support Parent/Child hierarchy, so no need to have any specific treatment for this pattern. Simply the table will remain as defined in the source system.

8.3.9. Junk Dimensions

Junk are miscellaneous attributes that don't belong to any existing dimension, they are typically flags or indicators that describe or categorize the transaction in some way. They're usually low cardinality, with a dozen distinct values or less. Even though the name is a bit disrespectful, the contents of the junk dimension are often important. We will handle those attributes by adding a Junk Dimension including all the existing combination existing in the fact table (Select Distinct of Junk attributes).

Below is an example of Sales transactions table, we gathered Transaction Type and Payment Type in one dimension including the combination of both attributes.

DimTransactionInfo		
PK	TransactionInfoKey	
	TransactionType	PaymentType
TransactionInfo Dimension		
TransactionInfoKey	TransactionType	PaymentType
1	Regular Sale	Cash
2	Regular Sale	Check
3	Regular Sale	Credit
4	Regular Sale	Debit
5	Refund	Cash
6	Refund	Check
7	Refund	Credit
8	Refund	Debit
9	No Sale	Cash
10	No Sale	Check
11	No Sale	Credit
12	No Sale	Debit

8.3.10. NA Records Management

Each dimension should have the below default records to categorize the absence/data inconsistency issues:

- Null: Missing in Source (-1) غير موجود في المصدر
- Not Applicable (Dimension not used) (-2) غير ملائم (غير ملزم)
- No Lookup Data - (-3) غير معروف في المصدر
- Not Happened yet - (-4) لم يحدَّد بعد

9. Data Integration Design

Data Sources

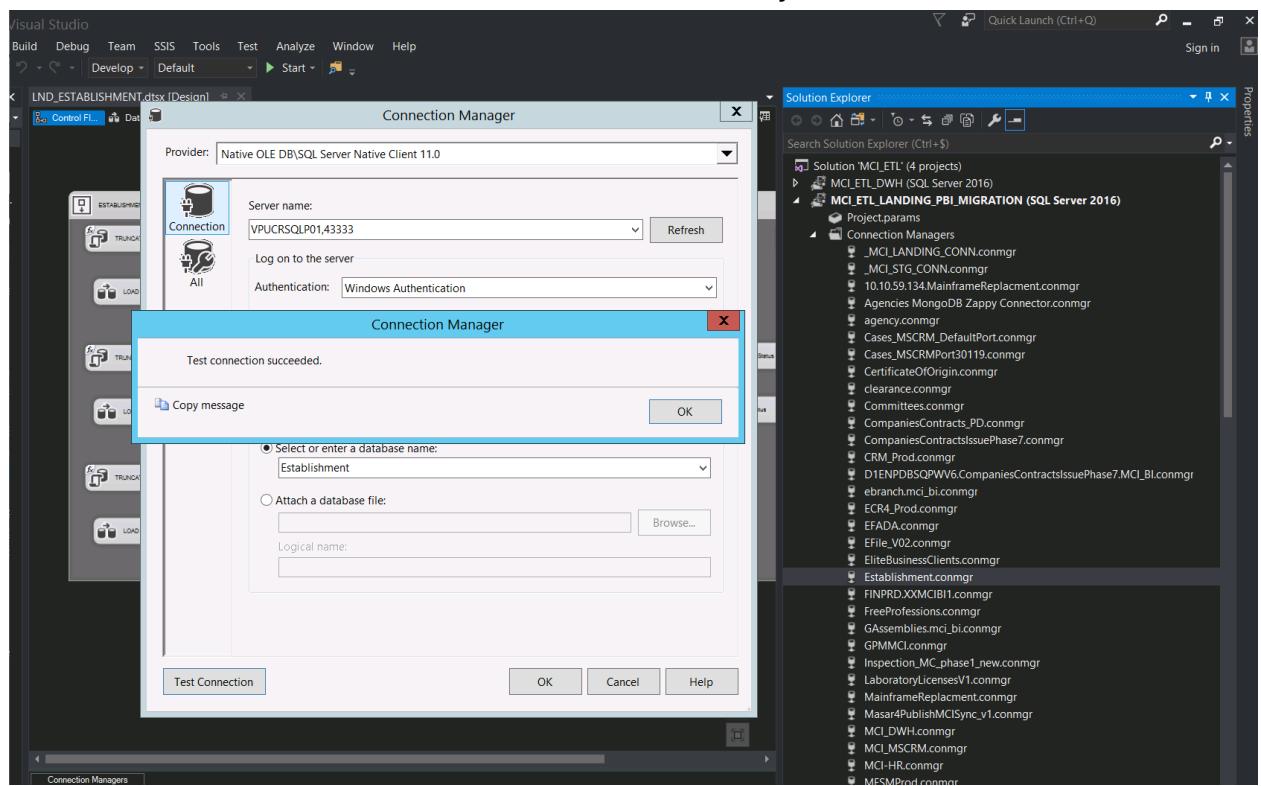
DWH Data comes from multiple sources, including SQL Server, Oracle, MongoDB, and Excel and CSV files. These diverse data inputs are essential for feeding the comprehensive data processing system.

Below is the list of the data sources of the landing tables used by our dashboards:

Data Source	Type
[10.10.51.191].Masar4PublishMCISync_v1	SQL Server Database
[10.10.51.58].Sales	SQL Server Database
[10.10.51.74].ECR4_Prod	SQL Server Database
[10.10.51.74].FreeProfessions	SQL Server Database
[10.10.52.171].clearance	SQL Server Database
[10.10.52.171].EFADA	SQL Server Database
[10.10.52.222].CompaniesContracts_PD	SQL Server Database
[10.10.52.222].CompaniesContractsIssuePhase7	SQL Server Database
[10.10.52.223].CertificateOfOrigin	SQL Server Database
[10.10.52.223].PartnersDecisions_V02	SQL Server Database
[10.10.59.134].Agencies	SQL Server Database
[10.10.59.134].AllGeneralCR	SQL Server Database
[10.10.59.134].Committees	SQL Server Database
[10.10.59.134].GPMMC1	SQL Server Database
[10.10.59.134].LaboratoryLicensesV1	SQL Server Database
[10.10.59.134].MainframeReplacment	SQL Server Database
[10.10.59.134].MCI-HR	SQL Server Database
[10.10.59.134].QviewQuery	SQL Server Database
[10.10.59.134].Survey	SQL Server Database
[10.10.59.134].TicketManagement1	SQL Server Database
[10.10.59.168].[EFile_V02]	SQL Server Database
[10.10.59.63].[TradeNamesCore_Prod]	SQL Server Database
[10.10.59.63].CRM_Prod	SQL Server Database

[10.10.59.63].MainframeReplacement	SQL Server Database
[10.10.59.63].TradeNamesCore_Prod	SQL Server Database
[10.50.16.58,30119].Cases_MSCRM	SQL Server Database
[10.50.16.58,30119].MCI_MSCRM	SQL Server Database
[10.50.16.78].[MCI_LANDING]	SQL Server Database
[10.50.16.78].MCI_DWH	SQL Server Database
[D1ENPDBSQPWV6].CompaniesContractsIssuePhase7	SQL Server Database
[FNMCISQLPROD\DB].ebranch	SQL Server Database
[FNMCISQLPROD\DB].MOAMALAT	SQL Server Database
[lib://AttachedFiles/Saudi Arabia Regions.xlsx]	Excel file
[vpblgsqlp001,32492].UEI_PROD	SQL Server Database
10.10.51.181.SJC	SQL Server Database
10.10.59.134.QviewQuery	SQL Server Database
10.10.59.134.Survey	SQL Server Database
10.10.59.246.MFSMProd	SQL Server Database
10.50.0.94:32309/admin	SQL Server Database
10.10.52.56.ArabicBasher	SQL Server Database
10.10.52.56.call_center2	SQL Server Database
D:\Excel Data Sources\Medium Level Apps\Mapping_CircuitCity.xlsx	Excel File
DWHSERVER\ D:\Excel Data Sources\Tasfiah	Excel file
FINPRD.XXCUSTOM	Oracle database
FNMCISQLPROD\DB.GAssemblies	SQL Server Database
10.200.16.135.Inspection_MC_phase1_new	SQL Server Database
MOCDB.QVIEW	Oracle database
D:\Excel Data Sources\Medium Level Apps\organization location.xlsx	Excel file

The Data sources in the ETL (SSIS) are defined as Project level Data sources,



9.1. ETL process

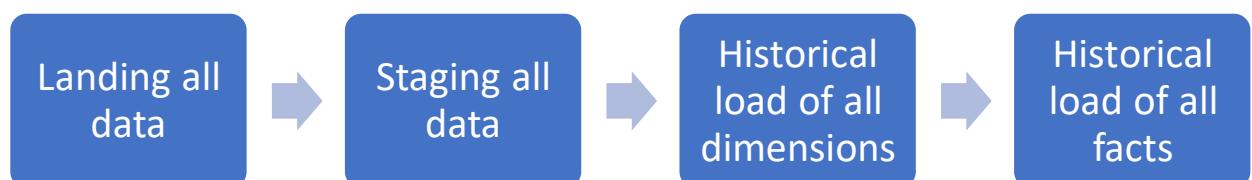
The data flow will pass through by 3 layers:

- Landing from the source systems, the landing area will be a replica from the source tables. The first step of the process is extracting the data from the primary source systems and storing it in a landing area by process of truncate-and-load.
- Staging area which contains the consolidated data from all databases, it is a replica of the data warehouse from structure point of view and is an intermediate area used for data transformation before loading the data in the data warehouse especially for fact tables. The second step will be consolidating the data copied to the landing area and copying it to a staging area.

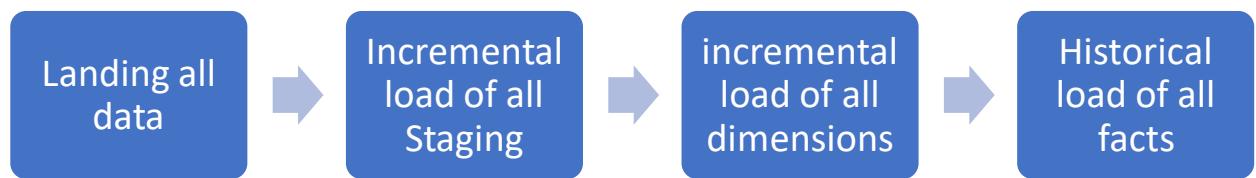
- The data warehouse is the last layer, and it will populate Dimensions and Facts from data in the staging area. It includes only verified and validated data. The third step will run tasks to populate the dimensions, and then run tasks to populate the fact tables.



9.1.1. Historical load



9.1.2. Incremental load



9.2. ETL Naming Convention

9.2.1. Master Packages

- STAGING MASTER: ETL_<Business subject>_STG_Master
 - o Example: ETL_EFD_STG_MASTER for the master package of EFADA
- DWH MASTER DIMENSION: MASTER_DIMENSION
- DWH MASTER FACT: MASTER_FACT

9.2.2. Landing Packages

LND_<Business subject abbreviation>_<DatasetName>

- o Example: LND_CRM_FreeProfessions

“_i” and “_h” will be added to the packages if they have historical and incremental load.

(Incremental load is applicable only for specific dataset table where Incremental load is possible and required) landing tables will always be a copy of the source tables to allow data quality checks to be done on a copy of the source table.

9.2.3. Staging Packages

- o STG_<Business subject>_TableName
 - Example: STG_SJC_COMMITTEE for the staging table corresponding to the dimension DIM_SJC_COMMITTEE
- o STG_<FactName>
 - Example :STG_FACT_TRADE_NAME_REQUESTS

9.2.4. DWH packages (Dimension/Fact Packages)

The name of the package will be the same as the (Dimension/Fact) table.

Example:

- DIM_USER
- FACT_TRADE_NAME_REQUESTS

9.2.5. Audit

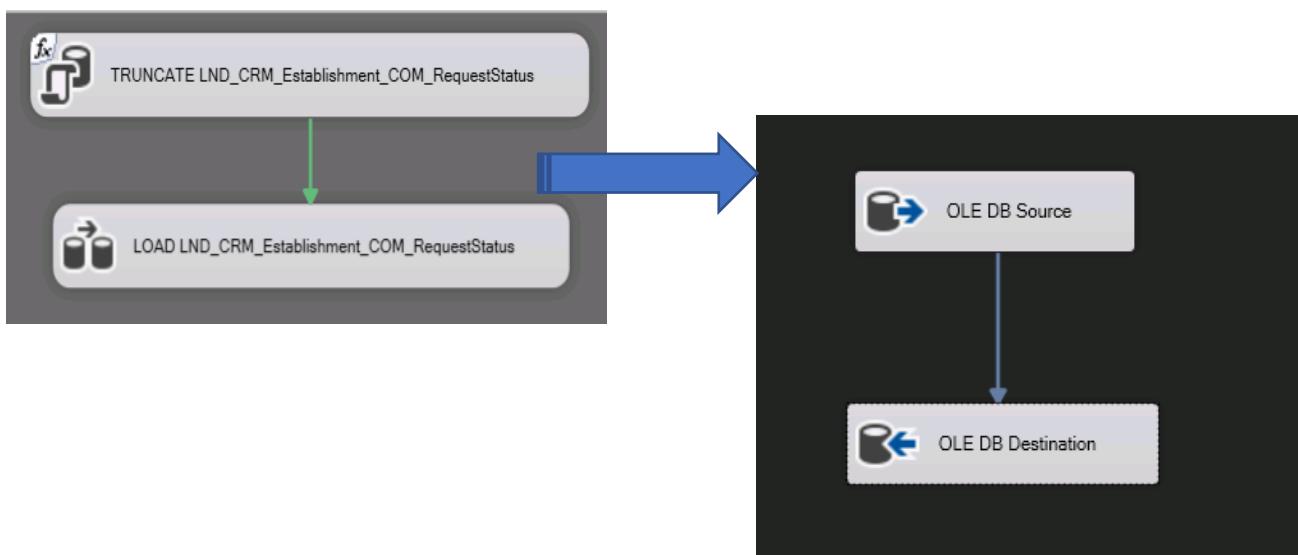
All the packages used for staging or data warehouse should have an audit record (in DIM_AUDIT) that includes the date of load, Package ID, start time, end time, number of records existing in the table before and after the load and the count of errors, if any, during the load.

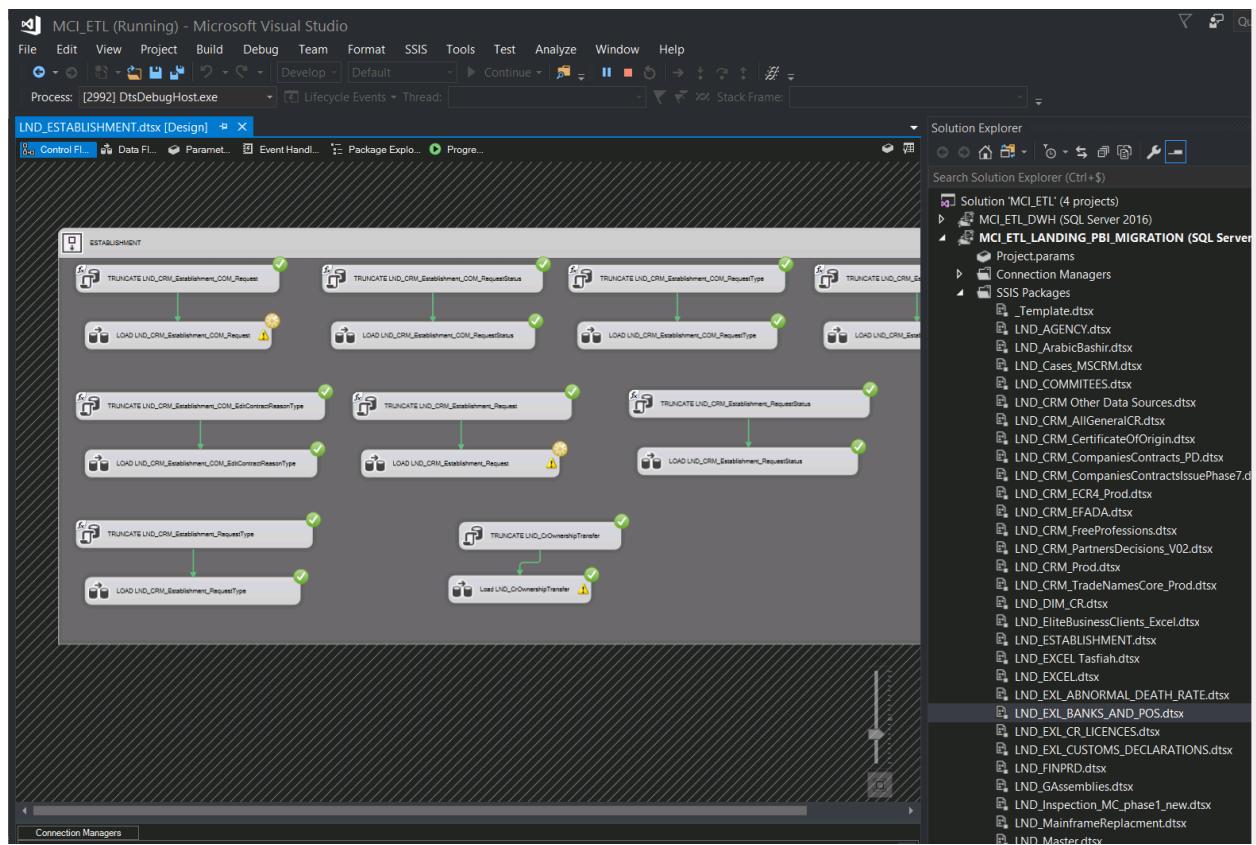
Package run will be keeping the last successful load date.

9.3. Landing ETL

In this step, the data will be copied from the different sources (Excel sheets, DB tables, etc..) and landed to the Landing Database.

This process will happen frequently (daily/monthly), for that a naming convention is required to organize the daily data extracted from the different sources.



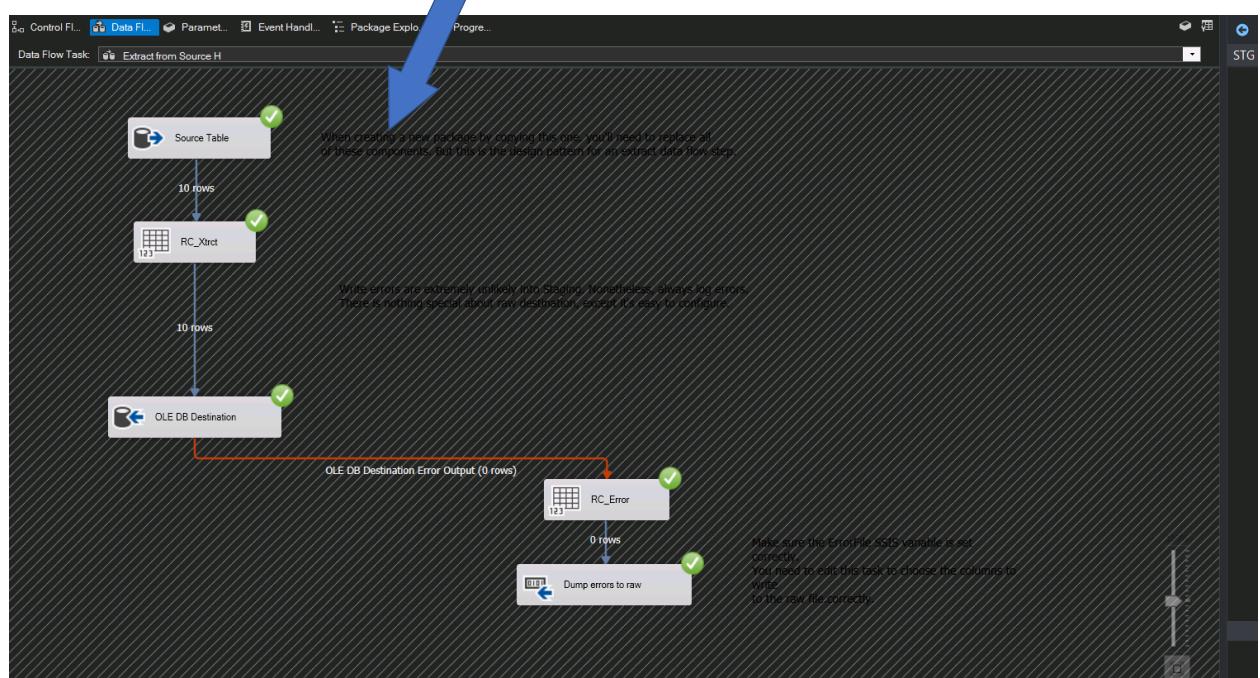
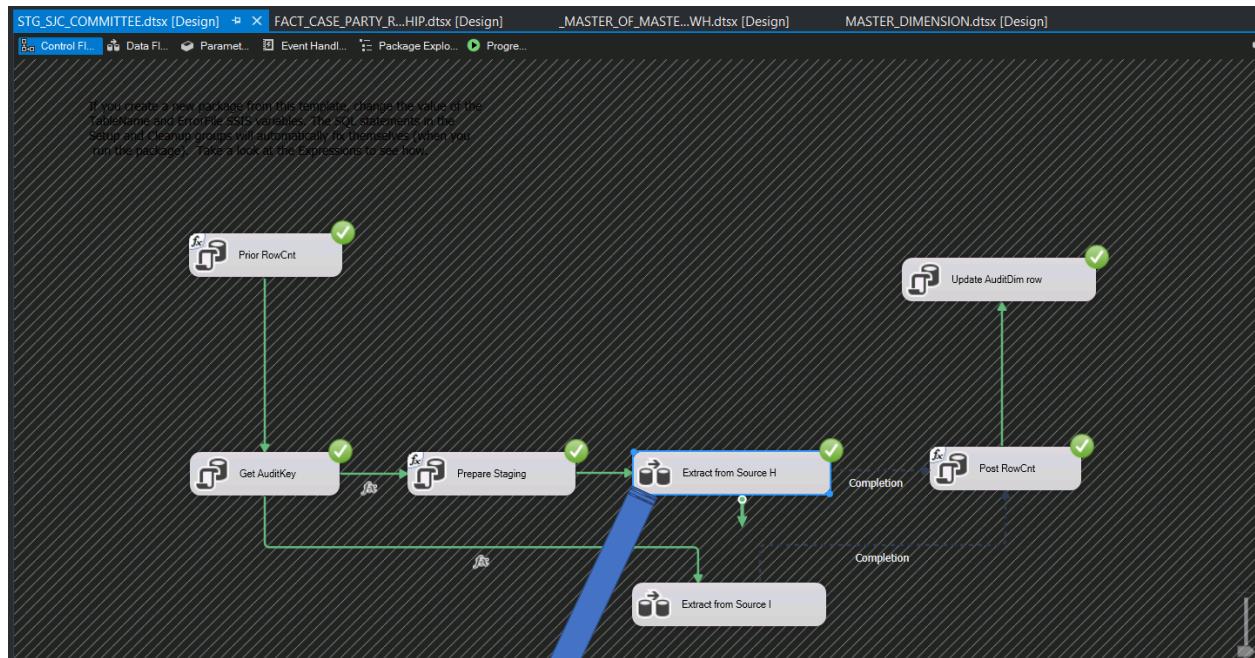


9.4. Staging ETL

After the data is copied to the landing database, another set of packages will execute to consolidate the data where required and move it to the staging database. The staging process will be used for the following tasks:

- **For Dimensions:** will be used for renaming fields as per the DW naming conventions, adding calculated columns, joining multiple tables to construct flattened dimensions.
 - o The staging table structure for dimensions will be like DW dimension tables' structure except for surrogate keys and SCD attributes.
- **For Facts:** will be used for renaming fields as per the DW naming conventions, adding calculated columns, joining multiple tables to construct one fact table, will be used later to join with the dimension tables to get the surrogate keys from dimensions.

- The staging table structure for fact will be like DW fact table structure except for dimensions' surrogate keys.

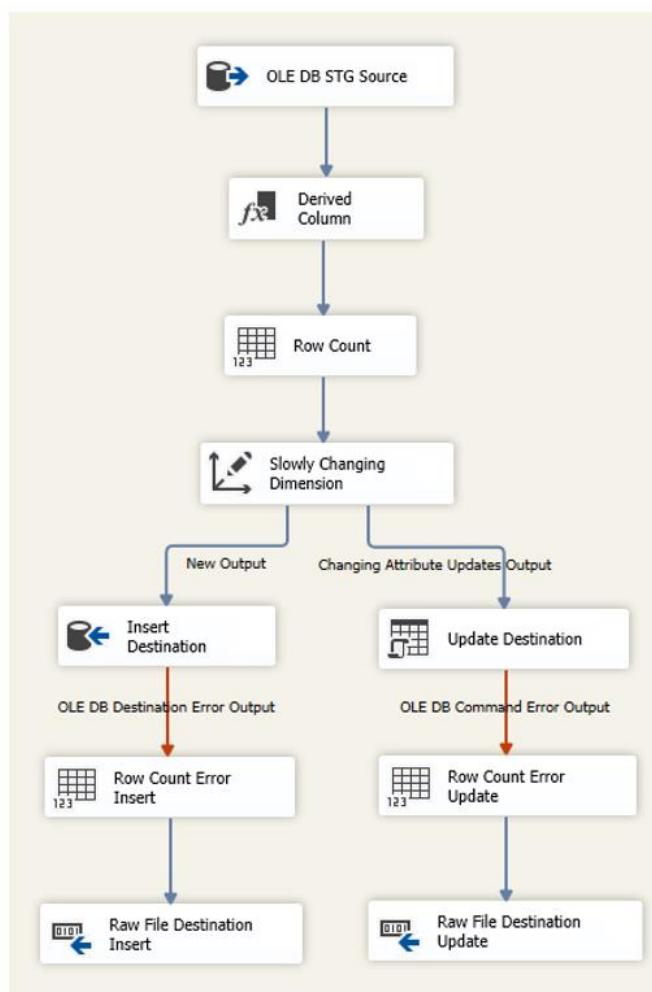


9.5. DWH ETL

The DWH process will contain packages to populate the dimension and fact tables which execute in the following order: dimensions and then facts.

DW Dimension Load package:

- Dimension will be loaded from their corresponding staging tables:
 - o Non-Slowly changing dimension:
 - Simple load between staging and dimension table in the data warehouse
 - o Slowly changing dimension:
 - Type 2 will be applied wherever it is applicable using the below pattern structure to manage SCD2.



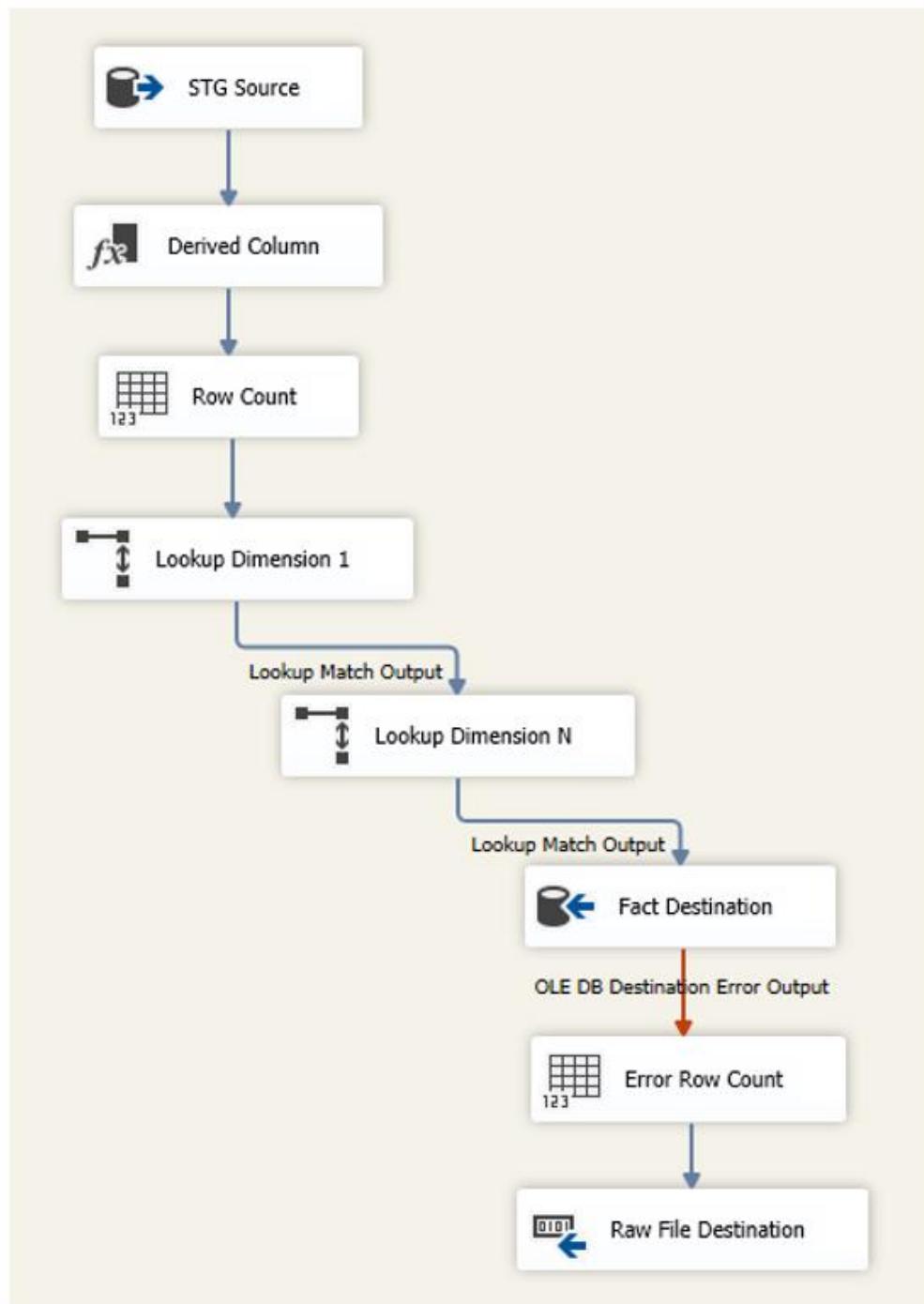
- Junk Dimension Load:

Depending on the case, 2 options are available to handle Junk Dimension loads:

- Populate the dimension based on Cartesian Product between all the flag values of the attributes – this option is recommended when the dimension includes mainly flags and attributes that might have a few numbers of values.
- Populate the dimension using Select Distinct of flag values from the fact table – this option is recommended when the number of values of the flags is not predefined.

DW Fact Load pipeline:

Fact load will be done by joining the corresponding staging fact table with the related dimension (using business keys and start date/end date of the record in case SCD Type 2) to get the corresponding surrogate keys of the dimensions and load the fact table.



10. ETL Deployment and jobs scheduling

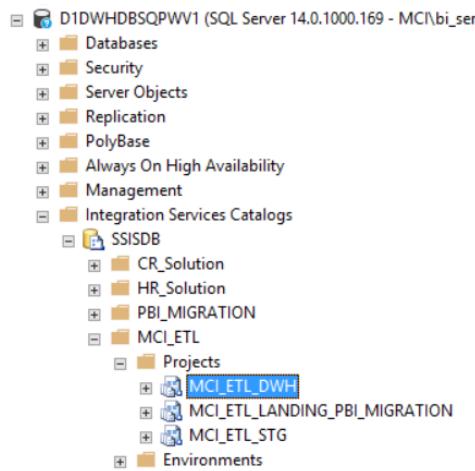
ETL Deployment

After development and testing in Visual Studio, SSIS packages are deployed to the SSISDB catalog on all DWH nodes to ensure high availability for job execution. The streamlined deployment includes:

- **Deployment to SSISDB Catalog:** Making SSIS packages available across all DWH nodes.
- **Creating SQL Server Agent Jobs:** Setting up jobs in SQL Server Agent on each node for package execution.
- **Adding Control Step:** Incorporating a control step within the SQL Server Agent job to check and ensure that only the primary DWH node triggers the execution, thus preventing duplicate processing.

This approach guarantees efficient execution and maximizes resource utilization within the DWH environment.

The screenshots below detail the deployment process within the SSISDB catalog and the configuration steps in SQL Server Agent, providing a visual walkthrough of the setup for efficient job execution.



Job Properties - MCI_ETL_JOB

Select a page:

- General
- Steps
- Schedules
- Alerts
- Notifications
- Targets

Script **Help**

Job step list:

S...	Name	Type	On Success...	On Failure
1	CHECK PIRAMARY REPLICA	Transact...	Go to the...	Quit the job ...
2	LANDING	SQL Ser...	Go to the...	Go to the n...
3	BACKUP AND SHRINK LANDING LOG FILE	SQL Ser...	Go to the...	Go to the n...
4	SHRINK TEMPDB Database	Transact...	Go to the...	Go to the n...
5	STAGING	SQL Ser...	Go to the...	Quit the job ...
6	BACKUP AND SHRINK STAGING LOG FILE	SQL Ser...	Go to the...	Go to the n...
7	Rebuild Indexes	Transact...	Go to the...	Go to the n...
8	DWH	SQL Ser...	Go to the...	Quit the job ...
9	Process OLAP Cubes	SQL Ser...	Go to the...	Quit the job ...
10	Sync SSAS Server 1	Transact...	Go to the...	Go to the n...
11	Sync SSAS Server 2	Transact...	Quit the j...	Go to the n...

Connection:

Server: D1DWHDBSQPWV1
Connection: MCI\bi_service

[View connection properties](#)

ETL Scheduling

The screenshots below illustrate the setup of a daily schedule for job execution within SQL Server Agent, set to run at 6 PM every day,

Job Properties - MCI_ETL_JOB

Select a page:

- General
- Steps
- Schedules
- Alerts
- Notifications
- Targets

Connection:

Server: D1DWHDBSQPWV1
Connection: MCI\bi_service

[View connection properties](#)

Schedule list:

ID	Name	Enabled	Description	Job...
1066	daily refresh	Yes	Occurs every day at 6:30:00 PM. Schedule will be used starting on 3/...	View

Job Schedule Properties - daily refresh

One-time occurrence:

Date: 2/19/2024 Time: 2:14:40 PM

Frequency:

Occurs: Daily Recur every: 1 day(s)

Daily frequency:

Occurs once at: 6:00:00 PM
Occurs every: 1 hour(s) Starting at: 6:30:00 PM Ending at: 11:59:59 PM

Duration:

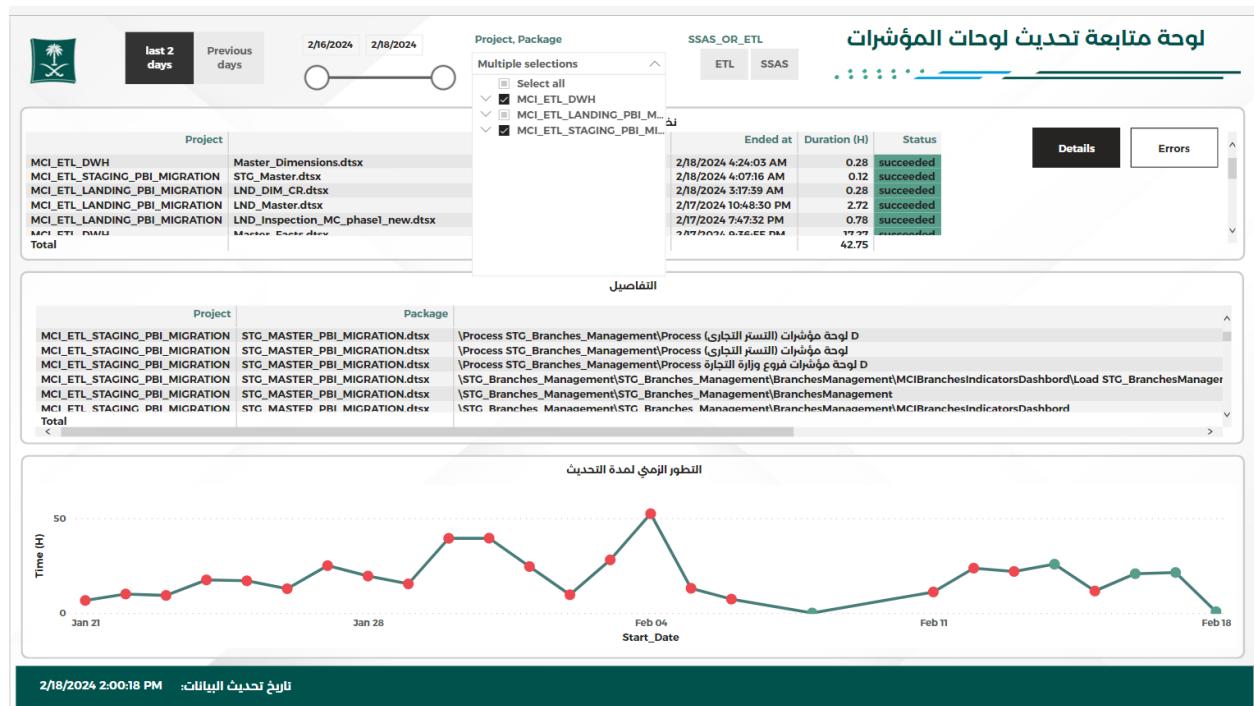
Start date: 3/4/2023 End date: 2/19/2024
 No end date

Summary:

Description: Occurs every day at 6:00:00 PM. Schedule will be used starting on 3/4/2023.

OK **Cancel** **Help**

11. Data integration monitoring



The dashboard in the image provides an overview of ETL process performance within a data warehousing environment:

- **Filters:**
 - **ETL Projects and Packages:** It lists individual ETL tasks, likely representing various stages of data loading such as staging, dimension and facts management.
 - **Task type:** data loading (ETL) or SSAS models processing.
 - **Date filters:**
 - Quick date selection (last two days or previous days)
 - Slider to select multiple dates to analyze.
- **Execution Log:** A table displays key metrics for each ETL job, including end time, duration, and status, essential for immediate operational insights.

- **Performance Chart:** A time series chart plots the execution times, with color coding to quickly highlight successes (green) and failures (red: if at least one failure on any component), useful for spotting trends or delays.
- **Detail Exploration:** Tabs for "Details" and "Errors" suggest functionality for deeper investigation into each job, aiding in troubleshooting and analysis.

This dashboard allows for efficient tracking and management of the ETL process, ensuring reliable data flow and system performance.

12. Dashboards Deployment

For the deployment of PBI Reports and SSRS reports, we have organized them into specific folders corresponding to each department. This setup follows a clear structure outlined in the provided table, which lists the names, types of dashboards/reports, and their designated folder paths within BI portal. Here's a succinct overview:

Item Name	Item Path	Item type
الأسماء التجارية	الأعمال التجارية/	Power BI Report
السجل التجاري _ التقرير التفصيلي	الأعمال التجارية/	Paginated report
المهن الاستشارية	الأعمال التجارية/	Power BI Report
الموقع الإلكتروني وبيانات معروفة	الأعمال التجارية/	Power BI Report
الوكالات التجارية	الأعمال التجارية/	Power BI Report
تقرير أسباب الرفض	الأعمال التجارية/	Power BI Report
تقرير عمليات تاجر	الأعمال التجارية/	Power BI Report
عقود الشركات	الأعمال التجارية/	Power BI Report
عمليات خدمات الوزارة	الأعمال التجارية/	Power BI Report
قرارات الشركاء القديم	الأعمال التجارية/	Power BI Report
لوحة مؤشرات السجل التجاري	الأعمال التجارية/	Power BI Report
لوحة مؤشرات السجل التجاري - التقرير التفصيلي	الأعمال التجارية/	Paginated report
لوحة مؤشرات السجل التجاري -بيانات الأشخاص	الأعمال التجارية/	Paginated report
لوحة مؤشرات نظام سهل بلس	الأعمال التجارية/	Power BI Report

مؤشرات الإفادة الإلكترونية	الأعمال التجارية/	Power BI Report
لوحة مؤشرات اللجان	السياسات والأنظمة/	Power BI Report
لوحة مؤشرات اللجان شبه القضائية	السياسات والأنظمة/	Power BI Report
مؤشرات باشر - أعمال	باشر/	Power BI Report
لوحة مؤشرات نظام باشر لخدمات الموظفين	تقنيات المعلومات/	Power BI Report
نظام باشر - مركز الخدمات قضايا التستر	تقنيات المعلومات/	Power BI Report
لوحة مؤشرات نظام إدارة القضايا التقارير الاسبوعية - ادارة العمليات خدمة العملاء	حماية المستهلك/	Power BI Report
الفرع الإلكتروني	خدمة العملاء والفروع/	Power BI Report
لوحة متابعة جودة خدمات الوزارة مؤشرات رضا العملاء	خدمة العملاء والفروع/	Power BI Report
نظام إدارة علاقات العملاء	خدمة العملاء والفروع/	Power BI Report
لوحة مؤشرات (التستر التجاري)	لوحات فروع وزارة التجارة/	Power BI Report
لوحة مؤشرات فروع وزارة التجارة مؤشرات المعاملات الإلكترونية	لوحات فروع وزارة التجارة/	Power BI Report
مؤشرات المعاملات الإلكترونية للمنظومة	معاملات/	Power BI Report
	معاملات/	Power BI Report

13. Security Access Design

For our security access design within the Power BI environment, we've implemented a comprehensive, multi-layered security framework to ensure that sensitive data is accessible only to authorized personnel. This framework consists of four key levels of security:

Portal Level Access: This foundational layer grants entry to all users who have successfully authenticated themselves. It serves as the initial gatekeeper, ensuring that only verified individuals can enter the portal.

Folder Level Access: At this level, access is refined to business area-specific folders. Users are granted access to folders relevant to their operational domain, ensuring they can access data and analytics pertinent to their area of work.

Dashboard/Report Level Access:

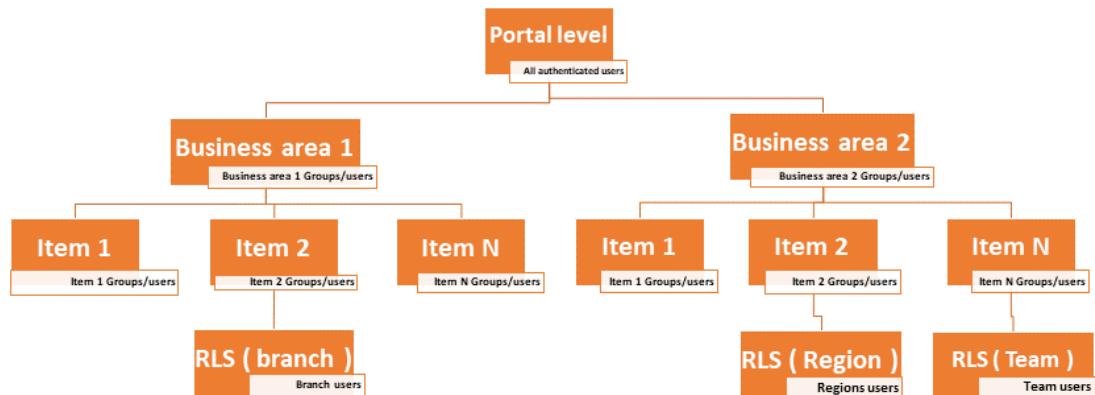
Further narrowing the scope, this layer controls access to individual dashboards and reports. Access permissions are allocated to specific users or groups, based on their need to interact with items. This ensures that data and insights are shared with the right people, enhancing data security and operational efficiency.

Row Level Security (RLS):

The most granular layer, RLS, restricts data visibility within reports and dashboards to the row level. Users can view only the data they are authorized to see, such as information specific to their region, branch, team, or department. This layer is crucial for maintaining strict data confidentiality and compliance with data protection regulations.

This structured approach to security ensures that our Power BI deployment is not only secure but also highly tailored to the specific needs and roles of our users, promoting a secure, efficient, and data-driven working environment.

The below diagram describes the access to Power BI report server folders and dashboards.



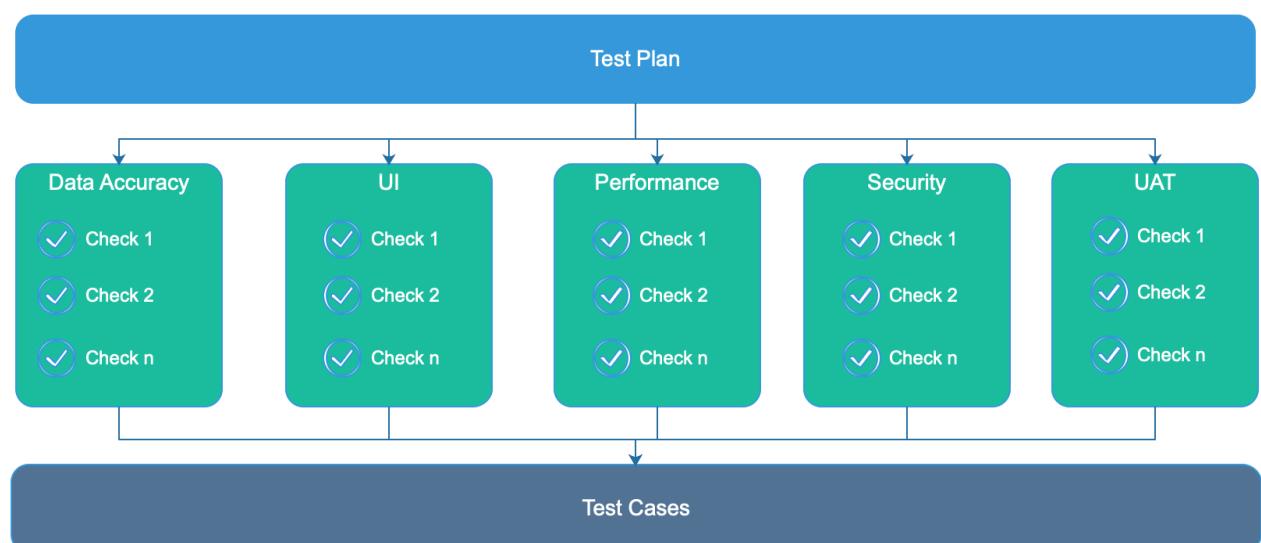
7.1. User Access Matrix template

Folder / dashboard	AD Users/Groups
Folder / dashboard	User 1 User 2 ...
Folder / dashboard	User 1 User 2 ...
..	..

8. Test Plan and Test Cases

8.1. Test Plan

This section will define the test strategy that will be carried out on the solution. Passing all the tests mentioned in this document will be considered the acceptance criteria for the solution. The below diagram summarizes the checkpoint categories in the test plan:



8.2. Data Accuracy Test

All data accuracy scripts (where applicable), and results will be saved in an excel file.

- Landing Database
 - Execute scripts to check the counts of all tables in the source and all tables in the landing, the scripts should return the difference.

- Data Warehouse (must start after landing database data accuracy testing is completed)
 - Fact Tables
 - Execute scripts to check the counts in the data warehouse vs the staging, using the data warehouse logic in the script, the scripts should return the difference.
 - Execute scripts to check the sum of the main numerical values that are used as measures, using the data warehouse logic in the script, the scripts should return the difference.
 - Dimension Tables
 - For slowly changing dimensions, execute scripts to check the counts of the latest records vs the counts in staging, the scripts should return the difference.
 - For the other dimensions, execute scripts to check the counts in the data warehouse vs the staging, the scripts should return the difference.
- Data Model and Dashboards (must start after data warehouse data accuracy testing is completed)
 - Define test cases for every indicator and metric in the dashboard and build scripts for each test case in the dashboards to return the same data from the data warehouse and values provided by the business, and make sure the 3 numbers are matching. (Business, Data warehouse and dashboard object).
 - If building the scripts for the previous 2 points are too complex, the table can be extracted to excel where a pivot table can be created and used to validate the data.

8.3. Dashboards UI Test

- The appropriate chart type should be selected as per the nature of the data (ex: pie charts are recommended to be used when there's 5 or less values in the dimension)
- All charts should be interactive, unless stated otherwise.
- Sorting on the charts should be correct and consistent.
- All dashboards and charts design and UI should be consistent.
- All charts and titles should be aligned.
- All decimal values should have the same precision unless stated otherwise.
- Chart legends should be consistent.
- All charts x axis and y axis should have labels.
- All charts should have data labels where applicable.
- Make sure all titles and labels are spelled correctly.
- All fonts should be consistent.
- Make sure that all the dashboard buttons are working as expected.

8.4. Performance Test

- ETLs should execute in a timely manner.
- Make sure that all the scheduled ETL jobs have run successfully and in a timely manner for 2 consecutive runs at least.
- Power BI Datasets should refresh in a timely manner.
- Dashboards should load in a timely manner.
- Any filtering or interactivity in the dashboards should load in a timely manner.

8.5. Security Test

- Define test cases for every access level:
 - Power BI Folder level
 - Dashboard level
 - Row level.
- The test case will include tests of:
 - Non authorized user to access every level.
 - Authorized user to access the level that he is authorized to access.

8.6. UAT

- Build scripts (where applicable if the data source is a database) for the test cases that the end users will use to validate the dashboards data, the test cases script and source and target results must be saved in an excel file.
- All the Indicators and metrics have at least one value covered by the test cases.
- The source values should be either shared by the end users or generated on the source database using the scripts shared by the end users source results must be pre-approved by the end users before the start of the development.
- The target values should be extracted from the dashboards.

8.7. Test Cases Template

A	B	C	D	E	F	G	H	I	J	
TC ID	Stream Name	Application Name	Sheet Name	Object Name	Object # (if labeled)	Dim	Phase 1 dashboard result	Phase 2 dashboard result	% Difference	
262	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	عدد الطلبات المكتملة حسب النوع		نقل ملكية	136141.00	136141.00	0.00%	
263	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	اجمالي الطلبات			468921.00	468921.00	0.00%	
264	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	عدد الطلبات المرفوضة			24797.00	24797.00	0.00%	
265	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	نسبة عدد الطلبات المرفوضة			5.29%	0.05	0.00%	
266	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	عدد طلبات الاداء المكتملة			408752	408752	0.00%	
267	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	عدد الطلبات المرفوضة عليها			383955	383955	0.00%	
268	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	نسبة عدد الطلبات المرفوضة عليها			0.8188	0.8188	0.00%	
269	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	متوسط ساعات العمل المسنويق			14.07	14.09	0.14%	
270	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	الطلبات المكتملة حسب الادارة		دراة خدمات الشركات	23393	23393	0.00%	
271	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	الطلبات المكتملة حسب هذه المعالجة		إيام	4.2	32671	32671	0.00%
272	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	متوسط ساعات العمل المسنويق للطلبات المكتملة حسب الشهر			44614	16.3	16.3	0.00%
273	الأعمال التجارية	مؤهلات الاداء الالكترونية	الملاخص التفصيلي لخدمة الاداء	عدد الطلبات المكتملة حسب شهر			44614	10710	10710	0.00%
274	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	الطلبات الفائمة حسب النوع		تغير مسمى مؤسسة	25	25	0.00%	
275	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	عدد طلبات الاداء تحت المعالجة			201	201	0.00%	
276	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	عدد أيام الاقليم طلب			1158	1159	0.09%	
277	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	الطلبات الفائمة حسب الادارة		دراة خدمات الشركات	0.0299	0.0299	0.00%	
278	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	الطلبات الفائمة حسب التاريخ			44866	1 for Jun 2023	1 for Jun 2023	0.00%
279	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	الطلبات الفائمة حسب المدة			يوم او اقل	117	117	0.00%
280	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	الطلبات الفائمة لدى الوزارة حسب الحالة			طلب جديد	164	164	0.00%
281	الأعمال التجارية	مؤهلات الاداء الالكترونية	منابع طلبات الاداء الفائلة	الطلبات الفائمة لدى العديل حسب الحاجة			إعادة تعيين	20714	20714	0.00%
282	الأعمال التجارية	مؤهلات الاداء الالكترونية	تحليل البيانات المرفوضة	متوسط عدد الرفض حسب الموقف			85	86	0.00%	
283	الأعمال التجارية	مؤهلات الاداء الالكترونية	تحليل البيانات المرفوضة	عدد الطلبات المرفوضة			24797	24797	0.00%	
284	الأعمال التجارية	مؤهلات الاداء الالكترونية	تحليل البيانات المرفوضة	معدل الطلبات المرفوضة			0.0529	0.0529	0.00%	
285	الأعمال التجارية	مؤهلات الاداء الالكترونية	تحليل البيانات المرفوضة	أسباب الرفض		تاميل رفاق عدد المنابعه وتنزيل للمؤسسه	672	672	0.00%	
286	الأعمال التجارية	مؤهلات الاداء الالكترونية	تحليل البيانات المرفوضة	عدد طلبات الرفض حسب المؤلفين	Haifa M. alomran		645	645	0.00%	
287	الأعمال التجارية	مؤهلات الاداء الالكترونية	تحليل البيانات المرفوضة	عدد طلبات الرفض حسب النوع		نقل ملكية	9410	9410	0.00%	
288	الأعمال التجارية	مؤهلات الاداء الالكترونية	التقارير	الطلبات المرفوضة		نوع الطلب : إثبات ملكية سجل تجاري رقم الطالبة	285823	نوع الطلب : إثبات ملكية سجل تجاري رقم الطالب	0.00%	
289	الأعمال التجارية	مؤهلات الاداء الالكترونية	التقارير	الطلبات الفائمة		نوع الطلب : نقل ملكية	738839	نوع الطلب : نقل ملكية	0.00%	