



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Assignment 1

Submitted to

Sri Lanka Institute of Information Technology

Submitted on

17-May-2022

IT20187828

Y3S1-WD.5.2

Isurika M.D.A



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Step 1 – Data Set Selection

I have chosen a dataset of IBRD bank. This dataset hosted by the World Bank.

The International Bank for Reconstruction and Development (IBRD) is a global development cooperative owned by 189 member countries. As the largest development bank in the world, it supports the World Bank Group's mission by providing loans.

This dataset includes bank loan details which is given to countries to continue their large developing projects. It includes all the loan bank statement details which issues on the last day of every month.

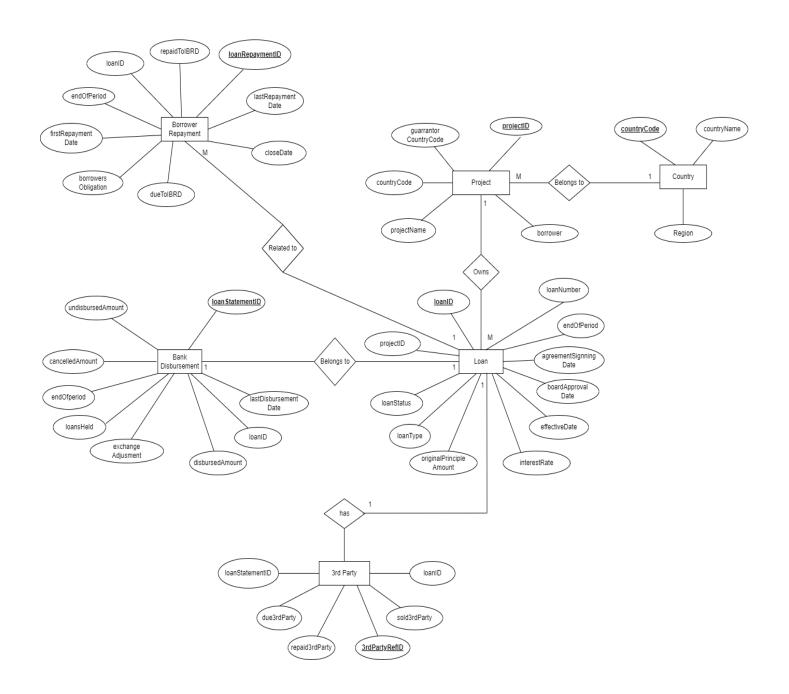
Data Set Link: https://www.kaggle.com/datasets/theworldbank/ibrd-statement-of-loans-data



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

ER diagram Of Data Sources.





IT3021 – Data Warehousing and Business Intelligence Assignment 1

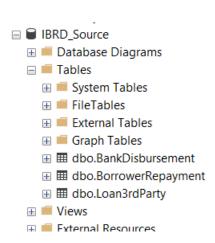
Semester 1,2022

Step 2 – Preparation of data sources

All the data consists in a one large CSV file along with over 900,000 rows. So, I partitioned main large CSV file into seven small sub files. As 3rdParty.CSV, BorrowerRepayment.CSV, Country.CSV, BankDisbursement.CSV, LoanDisbursement.CSV, LoanS.CSV and Projects.CSV.

Then depend on the guidelines I need to reconvert my CSV files into another two different formats as text and database. Here I used some ETL processes to load data into targeted formats.

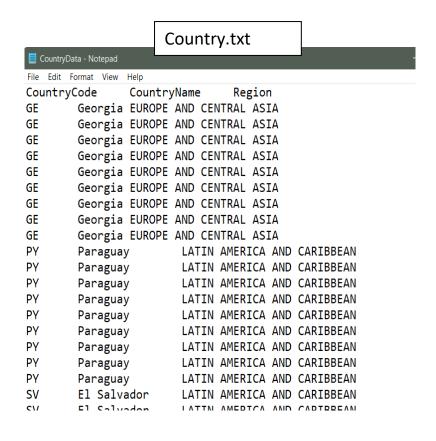
There I convert BorrowerRepayment.CSV, BankDisbursement.CSV and 3rdParty.CSV files into database format. Country.CSV and Projects.CSV into text format. And Loans.CSV keep as it is. So, I can take these as my source files for staging database.



🛢 prjData - Notepad	Project.txt		-
File Edit Format View Help)		
Project ID Pr	oject Name Country Code	e Borrower	
Guarantor			
P169913 Econ. Mana	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	gement and Competitivenes	s DPO GE	
MINISTRY OF FINANC			
	gement and Competitivenes	s DPO GE	
	E - Georgia Georgia		
	Ministerio de Hacie	nda Paraguay	
P168153 PIMA PY	Ministerio de Hacie	• •	
P168153 PIMA PY	Ministerio de Hacie		
P168153 PIMA PY	Ministerio de Hacie		
D1601E3 DTMA DV	M22.22.		



IT3021 – Data Warehousing and Business Intelligence Assignment 1



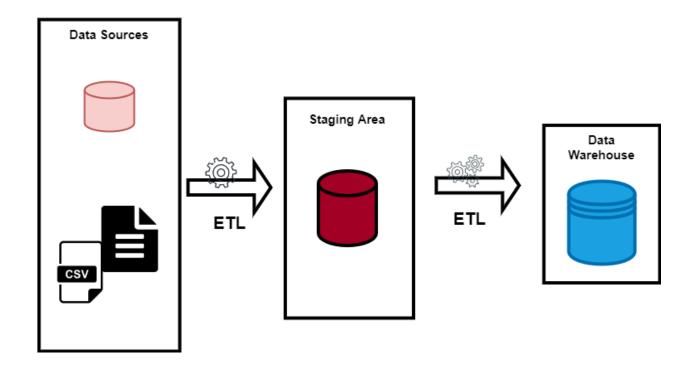
Loan.CSV									
_ A	В	С	D	E	F	G	Н	1	J
1 Loan Nur	m Project IE	Loan Type	Loan Status	Original Principal Amount	Interest Rate	End of Period	Agreement Signing Date	Board Approval Date	Effective Date (Most Recent
2 IBRD906	3(P169913	FSL	Approved	49600000		0 2020-03-31T00:00:00.000		2020-03-26T00:00:00.000	
3 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-04-30T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
4 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-05-31T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
5 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-06-30T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
6 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-07-31T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
7 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-08-31T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
8 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-09-30T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
9 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-10-31T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
10 IBRD906	3(P169913	FSL	Disbursing	49600000		0 2020-11-30T00:00:00.000	2020-03-30T00:00:00.000	2020-03-26T00:00:00.000	2020-04-30T00:00:00.000
11 IBRD906	4(P168153	FSL	Approved	100000000		0 2020-03-31T00:00:00.000		2020-03-19T00:00:00.000	
12 IBRD906	4(P168153	FSL	Signed	100000000		0 2020-04-30T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-10-11T00:00:00.000
13 IBRD9064	4(P168153	FSL	Signed	100000000		0 2020-05-31T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-10-11T00:00:00.000
14 IBRD906	4(P168153	FSL	Signed	100000000		0 2020-06-30T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-10-11T00:00:00.000
15 IBRD906	4(P168153	FSL	Disbursing	100000000		0 2020-07-31T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-07-06T00:00:00.000
16 IBRD906	4(P168153	FSL	Disbursing	100000000		0 2020-08-31T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-07-06T00:00:00.000
17 IBRD906	4(P168153	FSL	Disbursing	100000000		0 2020-09-30T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-07-06T00:00:00.000
18 IBRD906	4(P168153	FSL	Disbursing	100000000		0 2020-10-31T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-07-06T00:00:00.000
19 IBRD906	4(P168153	FSL	Disbursing	100000000		0 2020-11-30T00:00:00.000	2020-04-14T00:00:00.000	2020-03-19T00:00:00.000	2020-07-06T00:00:00.000
10 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-03-31T00:00:00.000		2020-03-19T00:00:00.000	
?1 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-04-30T00:00:00.000		2020-03-19T00:00:00.000	
22 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-05-31T00:00:00.000		2020-03-19T00:00:00.000	
23 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-06-30T00:00:00.000		2020-03-19T00:00:00.000	
24 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-07-31T00:00:00.000		2020-03-19T00:00:00.000	
25 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-08-31T00:00:00.000		2020-03-19T00:00:00.000	
?6 IBRD906	5(P169677	FSL	Approved	250000000		0 2020-09-30T00:00:00.000		2020-03-19T00:00:00.000	
17 10000000	ELDACOCTA.	CC1		250000000		0 2020 40 24700 00 00 00		2020 02 40700 00 00 000	



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

<u>Step 3 – Solution Architecture</u>



Data Sources: As data sources in this BI solution, I used database files and flat files sources as txt and CSV.

Staging Area: inside the staging area I stored all the source files as database files. When upload to the staging area from source database I used some transformations as sorting data by remove duplicate values, remove null values etc. therefor I have used some ETL processes when extracting from Source to Staging.



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Data Warehouse: When upload to the data warehouse from staging database I used to manage historical attributes of slowly changing dimensions. And keep track of changing attributes and fixed attributes. Here also I have used some ETL processes when extracting from Staging to warehouse.

Step 4 - Data warehouse design & development

I have designed the data warehouse schema using Snowflake Schema.

When loading data to fact table I have combined Bank Disbursement Staging, Borrower Repayment Staging and 3rdParty Staging into one fact Table. Therefore, data warehouse includes four dimension tables including date dimension and one fact table.

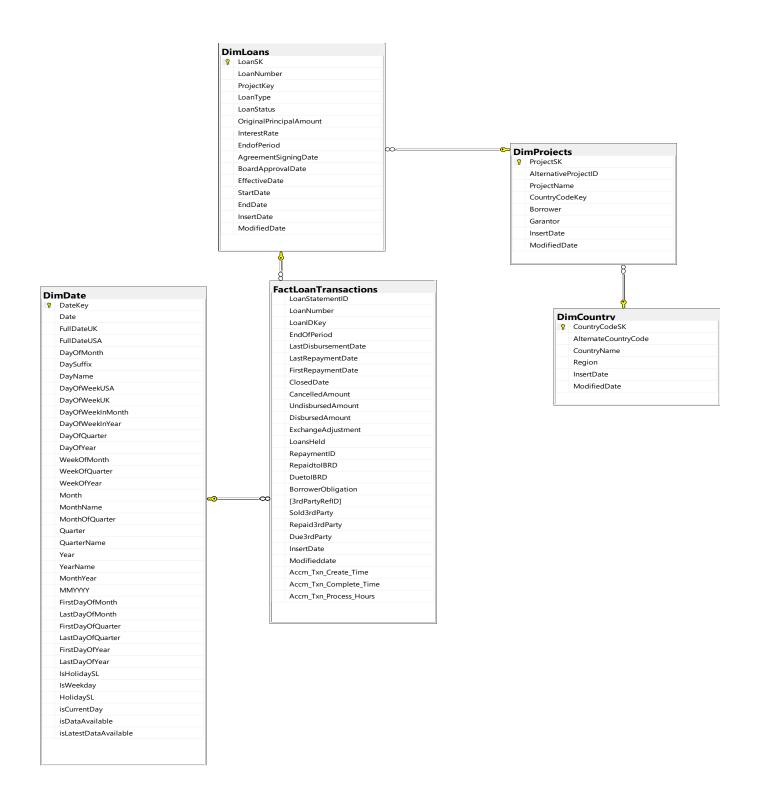
Fact table have relationship with Loan dimension and Date dimension. Then Loan dimension is connected to project dimension. And Project dimension is connected to Country dimension.

Assumptions

I have taken Dim Loan Data as slowly changing dimension, because **loan status** and **interest rate** can be changed time to time, and we need to keep track of their historical data.



IT3021 – Data Warehousing and Business Intelligence Assignment 1

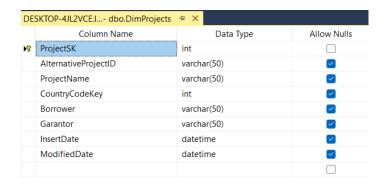




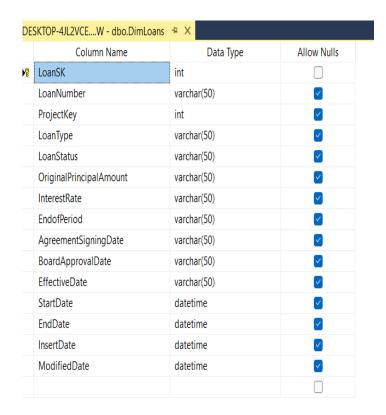
IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Project Dimension



Loan Dimension

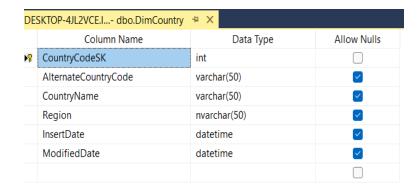




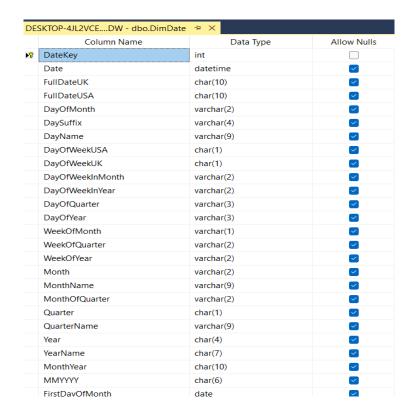
IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Country dimension



Date dimension





IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Fact Loan Transactions

Column Name	Data Type	Allow Null
LoanStatementID	int	<u>~</u>
LoanNumber	varchar(50)	<u>~</u>
LoanIDKey	int	<u>~</u>
EndOfPeriod	int	~
LastDisbursementDate	int	$\overline{\mathbf{C}}$
LastRepaymentDate	int	<u>~</u>
FirstRepaymentDate	int	\checkmark
ClosedDate	int	\checkmark
CancelledAmount	money	$\overline{\mathbf{v}}$
UndisbursedAmount	money	\checkmark
DisbursedAmount	money	\checkmark
Exchange Adjustment	float	$\overline{\checkmark}$
LoansHeld	money	\checkmark
RepaymentID	int	\checkmark
RepaidtoIBRD	money	\checkmark
DuetoIBRD	money	\checkmark
BorrowerObligation	money	\checkmark
[3rdPartyRefID]	int	\checkmark
Sold3rdParty	money	\checkmark
Repaid3rdParty	money	\checkmark
Due3rdParty	money	\checkmark
InsertDate	datetime	\checkmark
Modifieddate	datetime	\checkmark
Accm_Txn_Create_Time	datetime	$\overline{\mathbf{v}}$
Accm_Txn_Complete_Time	datetime	✓
Accm_Txn_Process_Hours	int	$\overline{\mathbf{v}}$

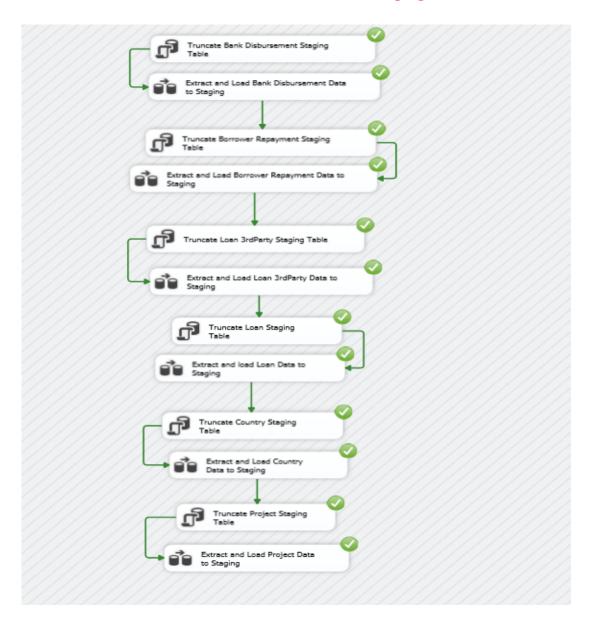


IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Step 5 - ETL Development

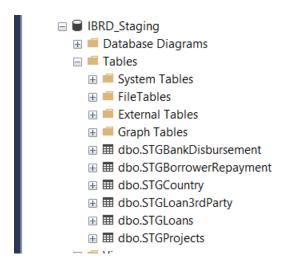
From Data Sources to Staging





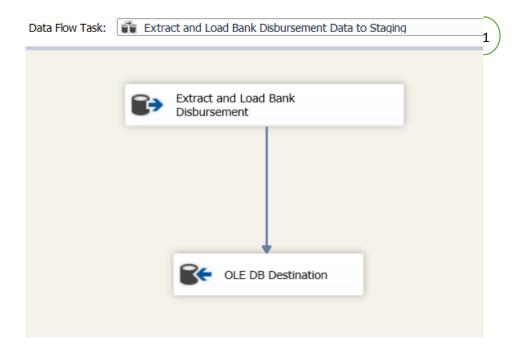
IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022



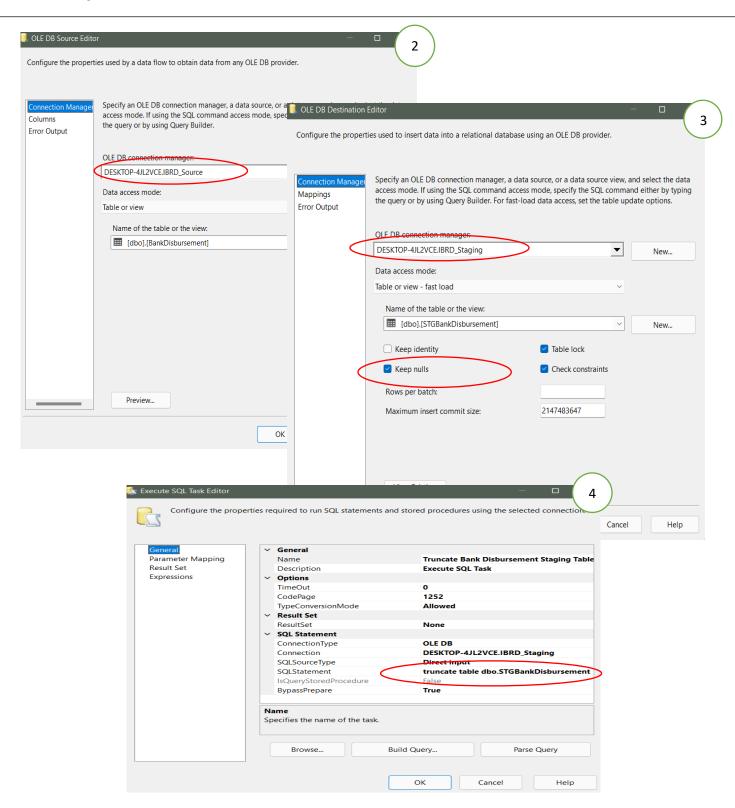
Load Data to Staging from IBRD Source database

- ** As mentioned in step 2 I have three tables in IBRD Source database. Therefore, I used same procedure to load all the three tables to Staging.
- ** here I only demonstrating Extract and load procedure of Bank Disbursement Table





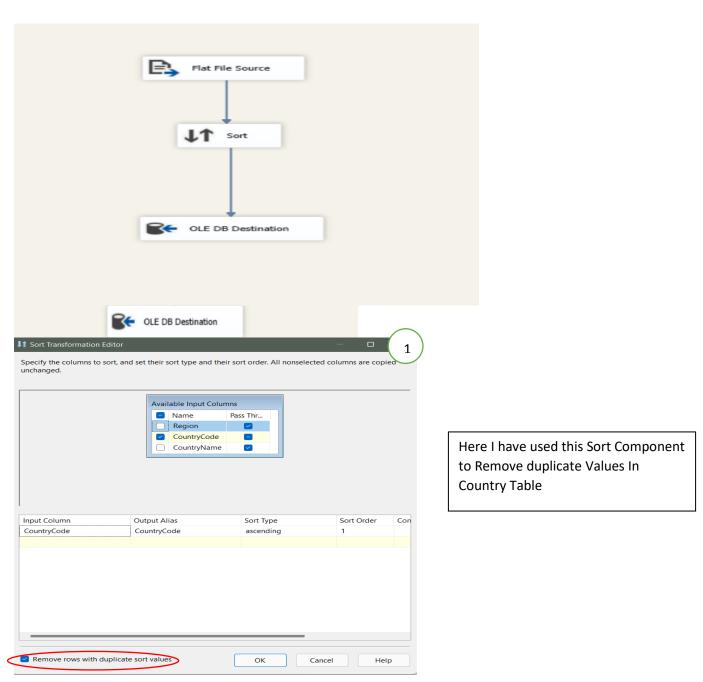
IT3021 – Data Warehousing and Business Intelligence Assignment 1





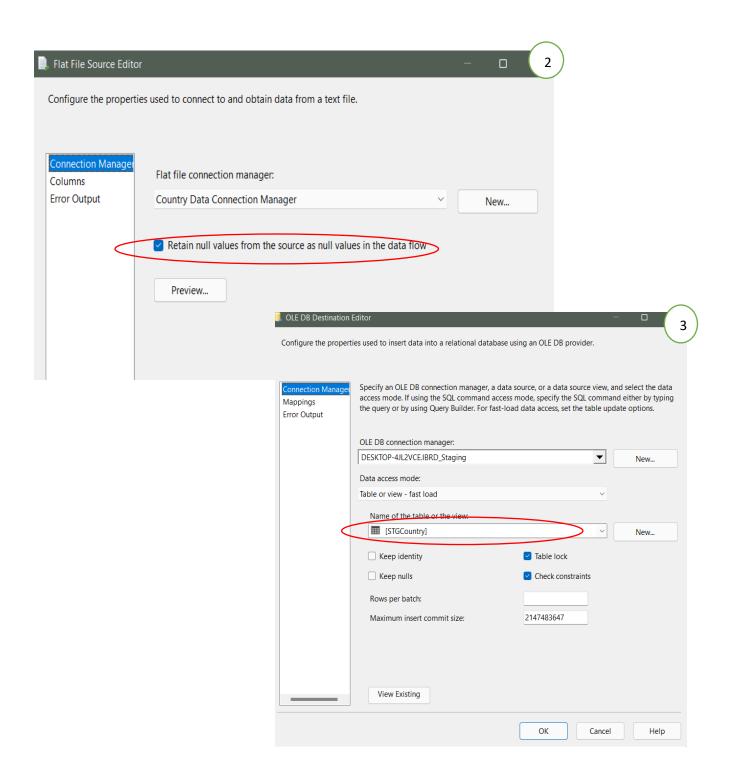
IT3021 – Data Warehousing and Business Intelligence Assignment 1

- ** As mentioned in step 2 I have two test files as Country.txt and Project.txt and one Loans.CSV file. Therefore, I used same procedure to load all the three files to Staging by using FLAT FILE SOURCE COMPONENT.
- ** here I only demonstrating Extract and load procedure of Country Staging Table





IT3021 – Data Warehousing and Business Intelligence Assignment 1

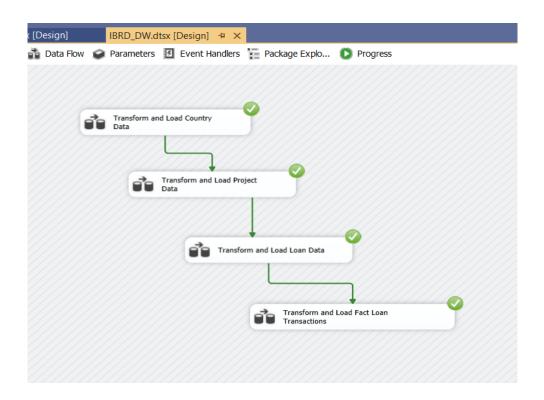


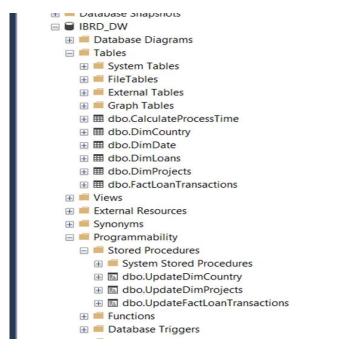


IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

From Data Staging to Data Warehouse



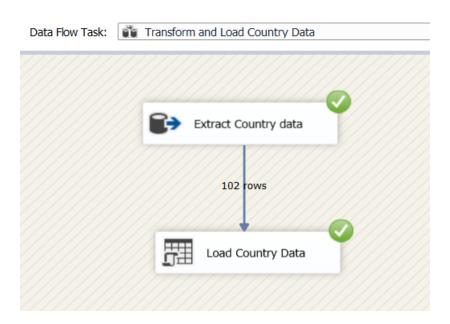


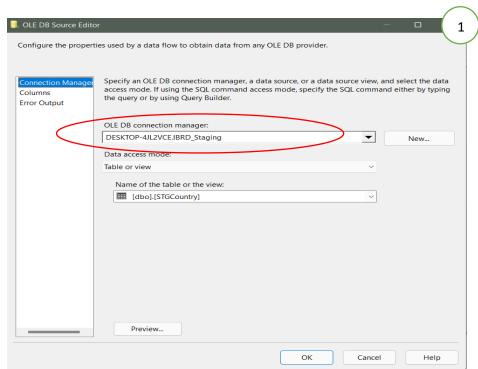


IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

1 step: Transform and Loading from Country Staging Table to Dim Country





First step is to configure OLE DB Source Connection



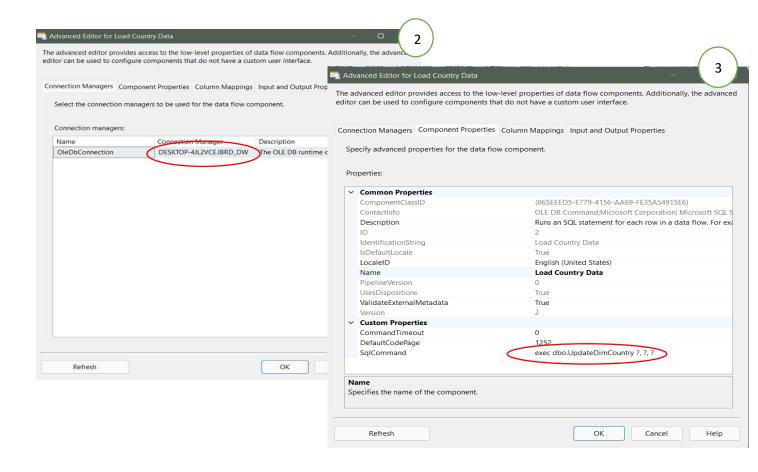
IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Steps of Configuring OLE DB Command Object

When Load to DW I used Stored Procedure Because in Country Dimension I have not any historical data to manage. I only want update rows if already existing in the table or add a new record.

There for I used OLE DB COMMAND COMPONENT to load the data to DW.





IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Stored Procedure to manage Dim Country table data

```
Query1.sql - DE...-4JL2VCE\acer (62))* * X DESKTOP-4JL2VCE.I...ctLoanTransactions
 USE [IBRD_DW]
 /***** Object: StoredProcedure [dbo].[UpdateDimCountry] Script Date: 17-May-22 13:52:38 ******/
 SET ANSI_NULLS ON
 SET QUOTED_IDENTIFIER ON
□ALTER PROCEDURE [dbo].[UpdateDimCountry]
 @CountryCode varchar(50),
 @CountryName varchar(50),
 @Region nvarchar(50)
 AS
⊟BEGIN
if not exists (select CountryCodeSK
 from dbo.DimCountry
 where AlternateCountryCode = @CountryCode)
insert into dbo.DimCountry
 (AlternateCountryCode, CountryName, Region, InsertDate, ModifiedDate)
  (@CountryCode, @CountryName, @Region, GETDATE(), GETDATE())
 END;
⊟if exists (select CountryCodeSK
  from dbo.DimCountry
 where AlternateCountryCode = @CountryCode)
BEGIN
⊟update dbo.DimCountry
 set CountryName = @CountryName,
 Region = @Region,
 ModifiedDate = GETDATE()
 where AlternateCountryCode = @CountryCode
 END;
 END;
```

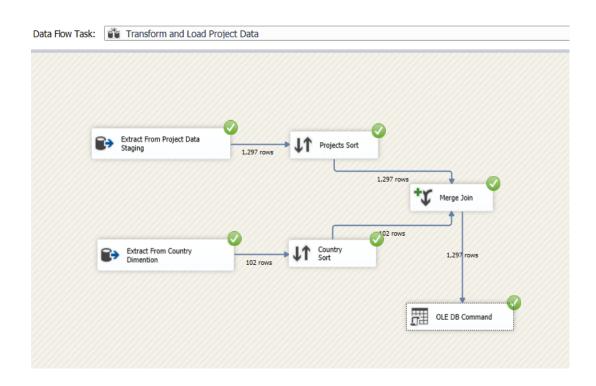


IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

2 step: Transform and Loading from Project Staging Table to Dim Project

** Same procedure used as above-mentioned step 1.



Difference between 1^{st} step and this 2^{nd} step is here I have to use another source connection manager to connect with Country Dimension to get data to 'CountryCodeSK' column .



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

Stored Procedure to manage Dim Project table data

```
LQuery2.sql - DE...-4JL2VCE\acer (62)) * X DESKTOP-4JL2VCE.l...ctLoanTransactions
  SET ANSI_NULLS ON
  SET QUOTED_IDENTIFIER ON
 □ALTER PROCEDURE [dbo].[UpdateDimProjects]
  @ProjectID varchar(50),
  @ProjectName varchar(50),
  @CountryCodeKey int,
  @Borrower varchar(50),
  @Garantor varchar(50)
  AS
 ⊟BEGIN
 if not exists (select ProjectSK
  from dbo.DimProjects
  where AlternativeProjectID = @ProjectID)
 insert into dbo.DimProjects
  (AlternativeProjectID , ProjectName, CountryCodeKey, Borrower, Garantor,
  InsertDate, ModifiedDate)
  (@ProjectID, @ProjectName, @CountryCodeKey, @Borrower, @Garantor,
  GETDATE(), GETDATE())
  END;
 ⊑if exists (select ProjectSK
  from dbo.DimProjects
  where AlternativeProjectID = @ProjectID)
 ⊟BEGIN
 ⊨update dbo.DimProjects
  ProjectName = @ProjectName,
  CountryCodeKey = @CountryCodeKey,
  Borrower = @Borrower,
  Garantor = @Garantor,
  ModifiedDate = GETDATE()
  where AlternativeProjectID = @ProjectID
  END;
```

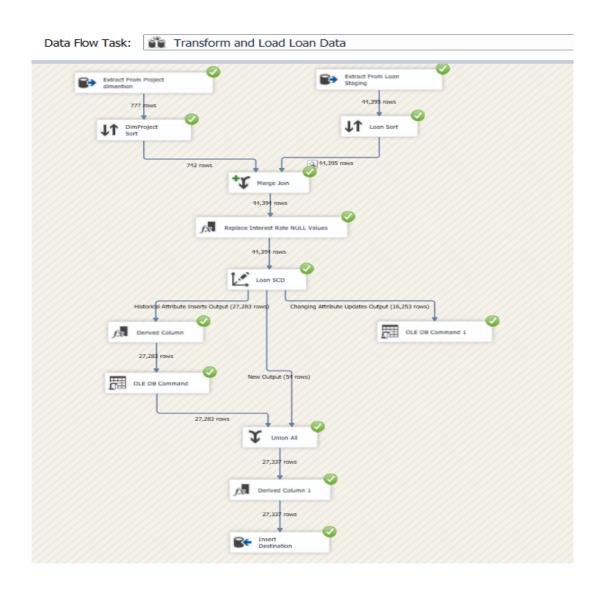


IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

3 step: Transform and Loading from Loan Staging Table to Dim Loan

Here I have loaded the loan dimension as a slowly changing dimension. Loan data needs some history management. When existing data row needs update, we cannot update that data row as it but we have to add that as a new record while updating modify date and end date column of already existing record.



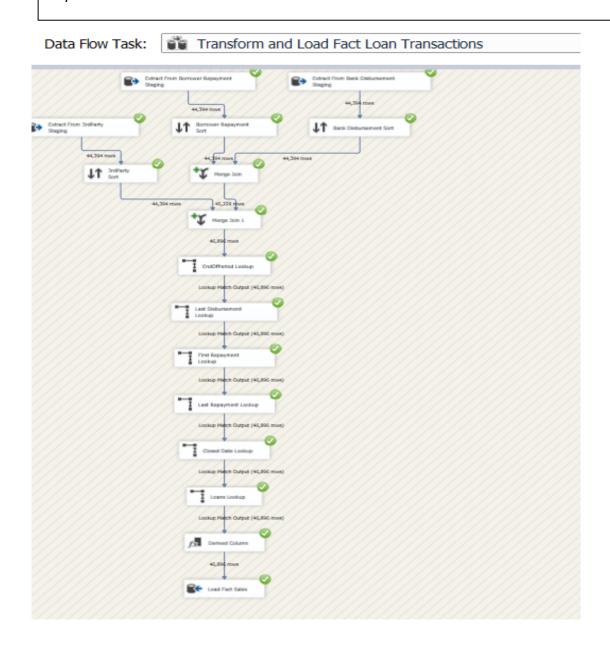


IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

4 step: Transform and Loading Fact Loan Transaction Table

Here I have used three staging tables Borrower Repayment, Bank Disbursement and 3rdparty and then all are loaded to Fact Table. Since fact table have relationship with date dimension and loan dimension, I used LOOKUP COMPONENT to get surrogate keys of them.

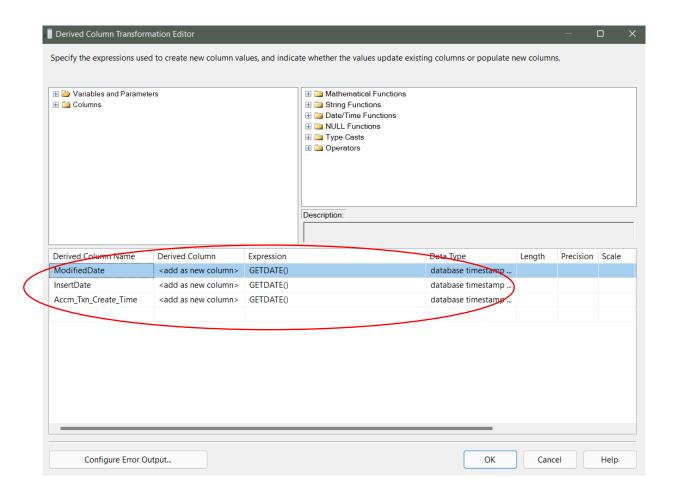




IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

I have used DERIVED COLUMN COMPONENT to add modified data column, insert date column. Additionally, Accm_Txn_Create_Time column which I will have to use in Step 6 of implementation.

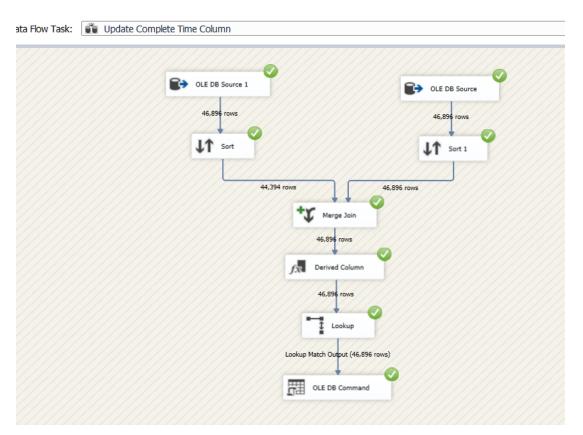




IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

<u>Step 6 - ETL Development – Accumulating Fact Tables</u>



	LoanStatementID	Derived Column.Accm_Txn_Complete_Time
1	24411	2022-05-13 19:20:24.790
2	24412	2022-05-13 19:20:24.790
3	24413	2022-05-13 19:20:24.790
4	24414	2022-05-13 19:20:24.790
5	24415	2022-05-13 19:20:24.790
6	24416	2022-05-13 19:20:24.790
7	24417	2022-05-13 19:20:24.790
8	24418	2022-05-13 19:20:24.790
9	24419	2022-05-13 19:20:24.790
10	24420	2022-05-13 19:20:24.790
11	24421	2022-05-13 19:20:24.790
12	24422	2022-05-13 19:20:24.790
13	24423	2022-05-13 19:20:24.790
14	24424	2022-05-13 19:20:24.790
15	24425	2022-05-13 19:20:24.790
16	24426	2022-05-13 19:20:24.790
17	2//27	2022_05_12 10-20-24 700

This is the table I created to update Txn_Complete_Time.

I generate it by using a excel sheet and then import to DW.



IT3021 – Data Warehousing and Business Intelligence Assignment 1

Semester 1,2022

	Modifieddate	Accm_Txn_Create_Time	Accm_Txn_Complete_Time	Accm_Txn_Process_Hours
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.79 0	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93
:27.463	2022-05-17 16:56:27.463	2022-05-17 16:56:27.463	2022-05-13 19:20:24.790	-93

```
∃ALTER PROCEDURE [dbo].[UpdateFactLoanTransactions]
        @LoanStatementID int
       ,@LoanNumber varchar(50)
       ,@LoanIDKey int
       ,@EndOfPeriod int
       ,@LastDisbursementDate int
       ,@LastRepaymentDate int
       ,@FirstRepaymentDate int
       ,@ClosedDate int
       ,@CancelledAmount varchar(50)
       ,@UndisbursedAmount varchar(50)
       ,@DisbursedAmount varchar(50)
       ,@ExchangeAdjustment varchar(50)
       ,@LoansHeld varchar(50)
       ,@RepaymentID int
       ,@RepaidtoIBRD varchar(50)
       ,@DuetoIBRD varchar(50)
       ,@BorrowerObligation varchar(50)
       ,@3rdPartyRefID int
       ,@Sold3rdParty varchar(50)
       ,@Repaid3rdParty varchar(50)
       ,@Due3rdParty varchar(50)
       ,@InsertDate datetime
       ,@Modifieddate datetime
       ,@Accm_Txn_Create_Time datetime
       ,@Accm_Txn_Complete_Time datetime
       ,@Accm_Txn_Process_Hours int
        update dbo.FactLoanTransactions
        Accm_Txn_Complete_Time = @Accm_Txn_Complete_Time,
        Accm_Txn_Process_Hours = @Accm_Txn_Process_Hours
        where LoanStatementID = @LoanStatementID
```

Stored Procedure I used to truncate Fact Loan Table to avoid duplicating data when reloading.



IT3021 – Data Warehousing and Business Intelligence Assignment 1