بسم الله الرحمن الرحيم

DWH Final Project report

(Library Management System DWH)

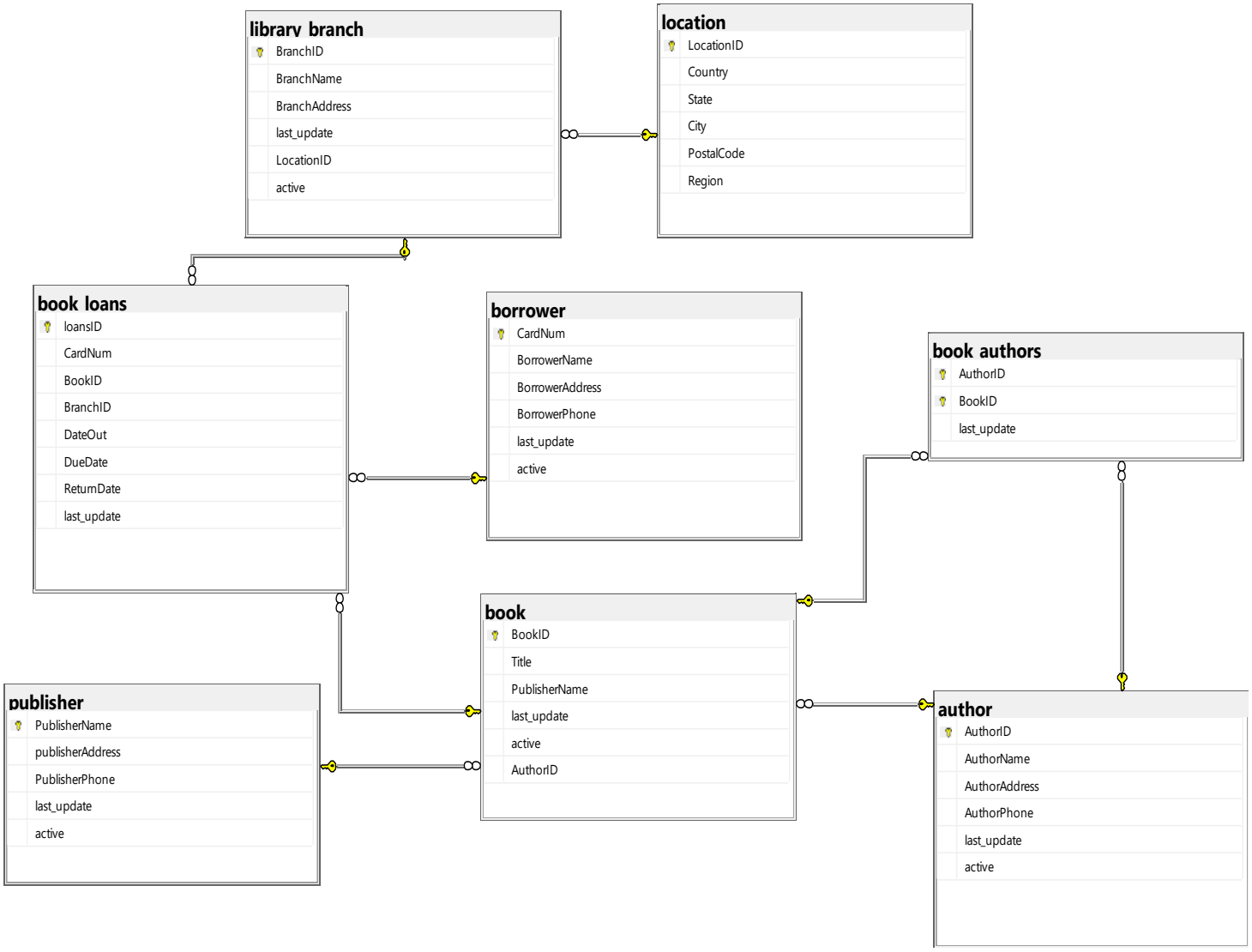
TA: Samaa Mohamed Mahmoud

Group: all

Done by:

Ahmed Mohamed Abass	**20220035**
Mohamed Hussien Aly	**20220286**
Ahmed Yasser Ahmed	**20220046**
Nourhan Mohamed Ahmed	**20220368**
Sarah Ibrahim abd el sattar	**20220150**

2.Physical model of the source system:



3. Dimensional model

a. Define the business processes that you will model.

1. Number of books borrowed analysis
    - Calculate the total number of books borrowed by a borrower from a specific branch.
2. Number of borrowers analysis
    - Calculate the total number of borrowers in each branch.
3. Number of loans in branch and location analysis
    - Calculate the total number of loans in each branch and location.
4. Number of loans per book and author analysis
    - Calculate the total number of loans per book and author.

Entities: Books, Borrower, Library Branches, Location, Author, Publisher, Book Loans.

Measures: Number of books borrowed, Number of borrowers, Total loans per branch and location, Total loans per book and author.

Dimensions: Time, Book, Borrower, Branch, Location, Author, Publisher, Book Loan.

b. Declare the grain of each fact table

- **F_NumBooksBorrower:** Total number of books borrowed by a borrower from a specific branch on a particular date.
- **F_NumBorrowers:** Total number of borrowers in a specific branch on a particular date.
- **F_loanDuration:** Total Number of loans per Branch and location on a particular date.
- **F_loans_book_author:** Total Number of loans per Book and Author on a particular date.

c. Define the type of each fact table.

- **F_NumBooksBorrower:** Transactional Fact
- **F_NumBorrowers:** Transactional Fact
- **F_loanDuration:** Transactional Fact
- **F_loans_book_author:** Transactional Fact

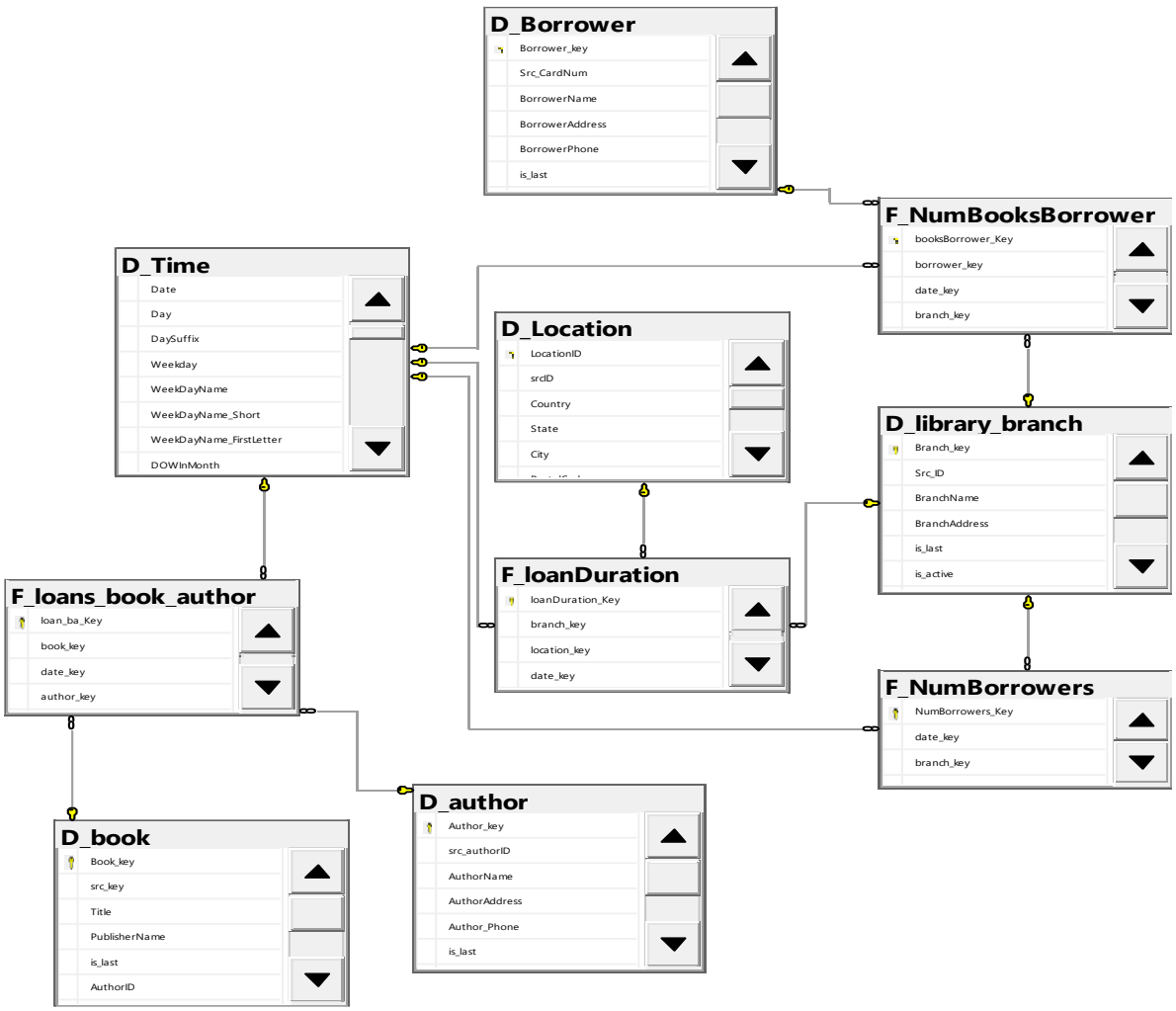
d. Define the dimensions and the type of each one.

- Borrower —> SCD Type 2 Dimension
- Publisher —> SCD Type 2 Dimension
- Author —> SCD Type 2 Dimension
- Book —> SCD Type 2 Dimension
- Library_Branch —> SCD Type 2 and Conformed Dimension
- Location —> Static (Fixed) Dimension
- Time —> Conformed Dimension

e. Define the measures that will appear in the fact tables and the type of each one.

- **NumBorrowers :** Non-additive Measure
- **NumBooksBorrowed :** Semi-Additive Measure
- **Total_Loans(2):** Fully Additive Measure

f. Physical model of your Star Schema Model

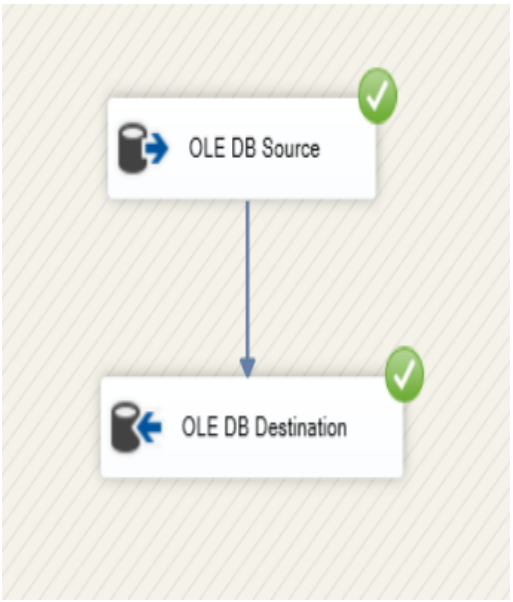


4. Screenshots of the data flow and control flow used for building the DWH.

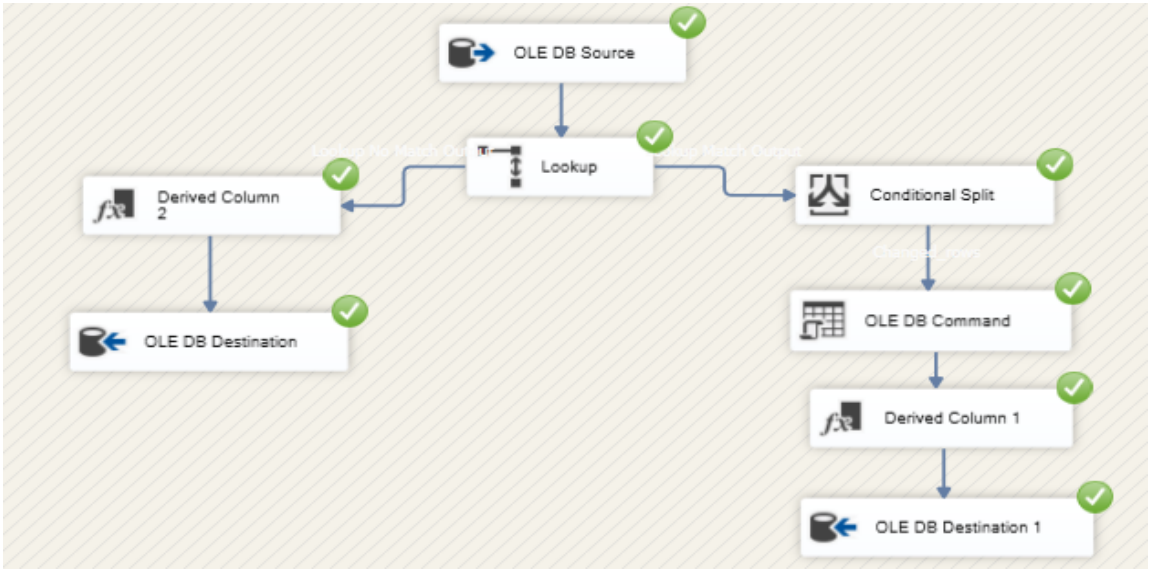
1-Book dimension:-

Control flow

dataflow from source to stage_area

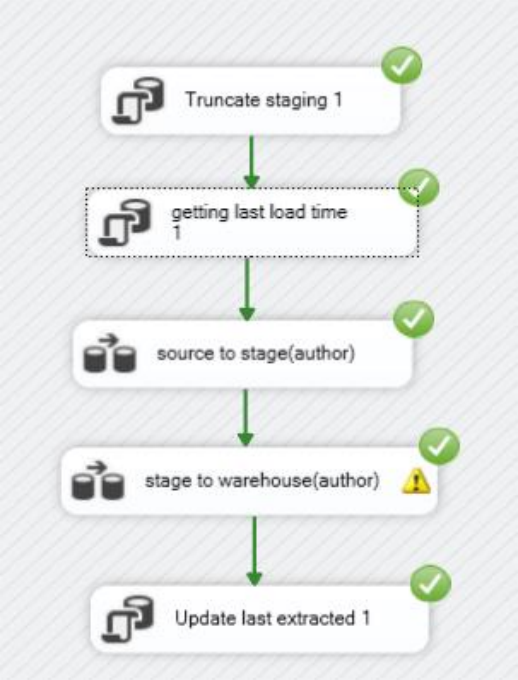


Data flow from stage_area to warehouse(destination) d_book

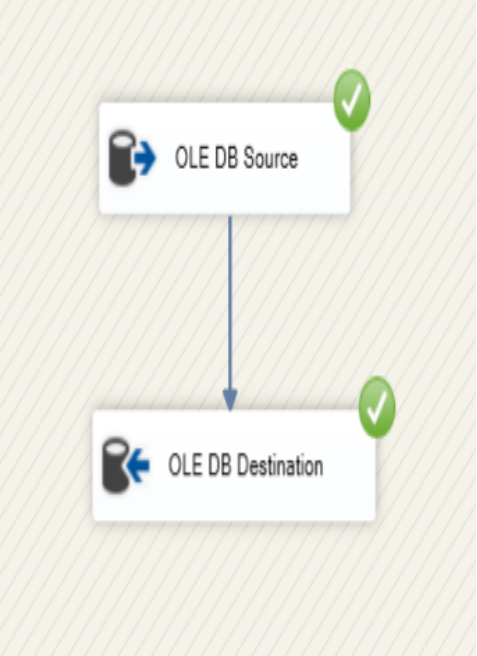


2-Author dimension:-

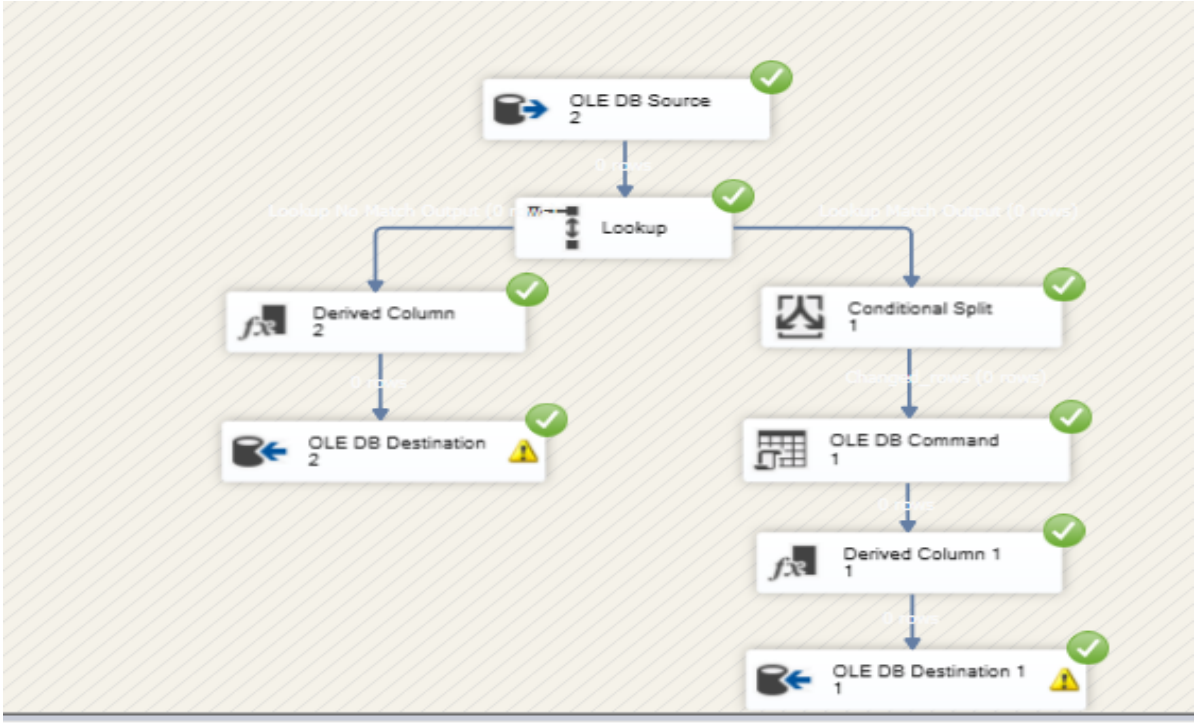
Control flow



dataflow from source to stage_area

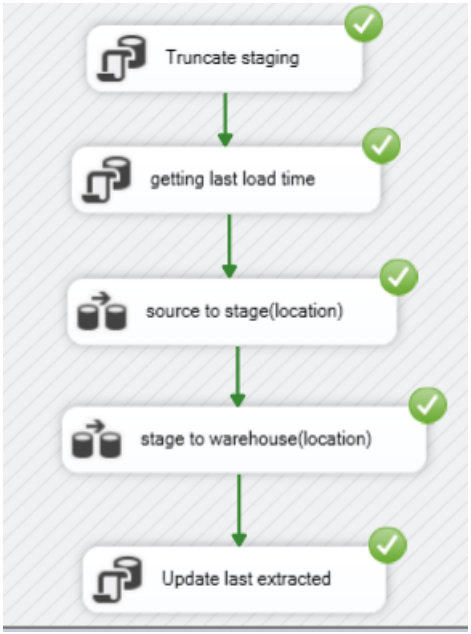


Data flow from stage_area to warehouse(destination) d_author

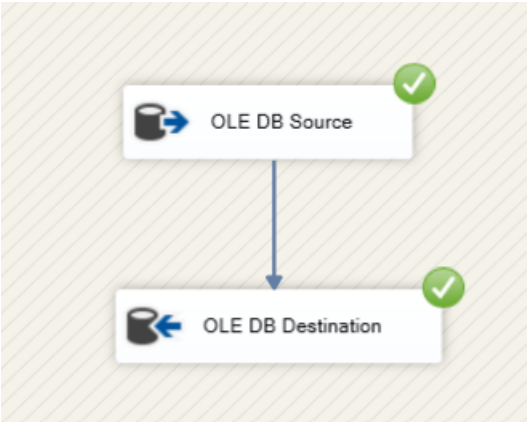


3-Location dimension:-

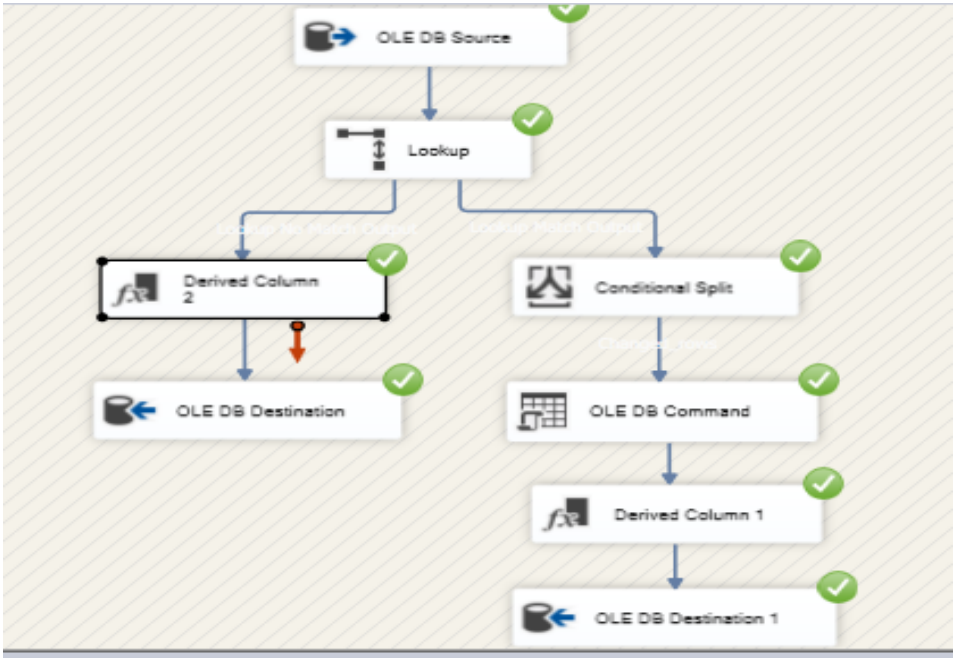
Control flow



data flow from source to stage



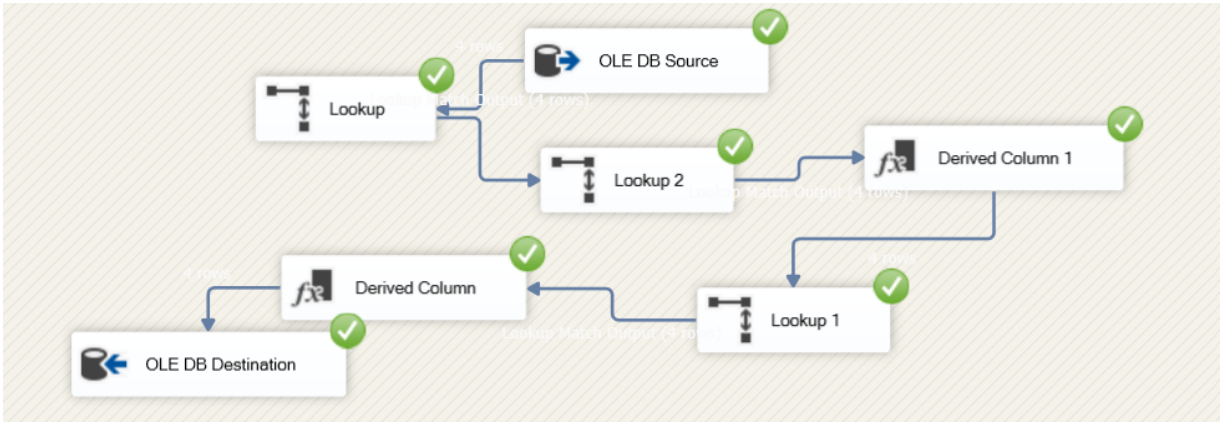
data flow from stage_area to warehouse for d_author



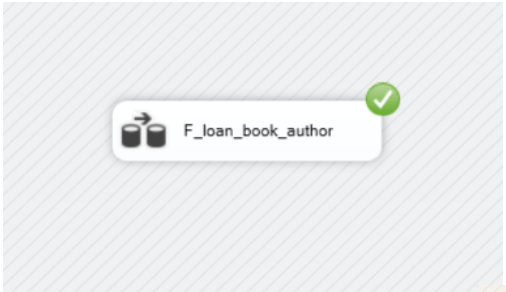
4-F_loans_book_author:-

Control flow

data flow from dimensions(D_time – D_location – D_library_branch) to fact table (F_loans_author)

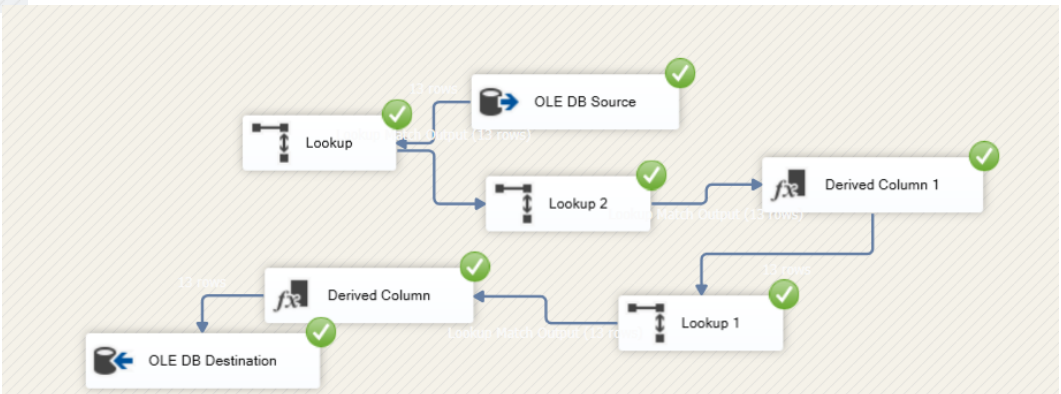


5- F_loans_book_author fact table:-



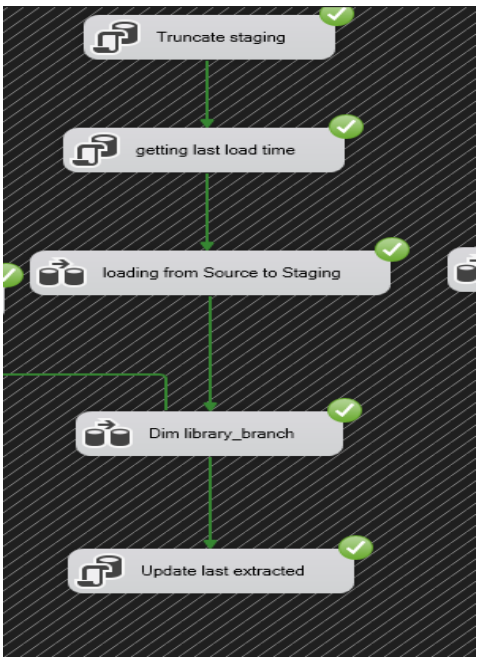
Control flow

dataflow from dimensions (D_time – D_author – D_book) to fact table(F_loan_book_author)

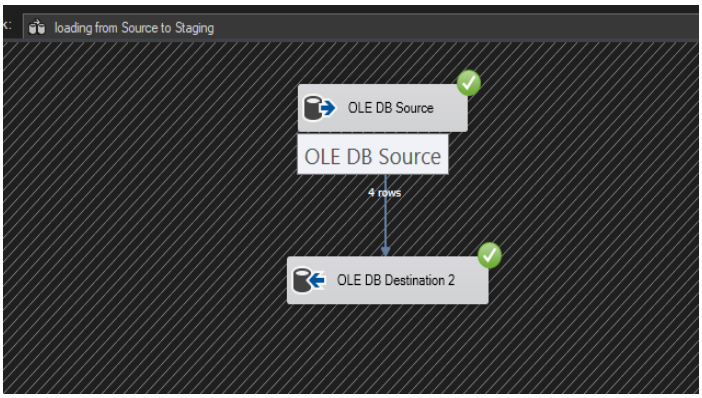


6- library_branch dimension:-

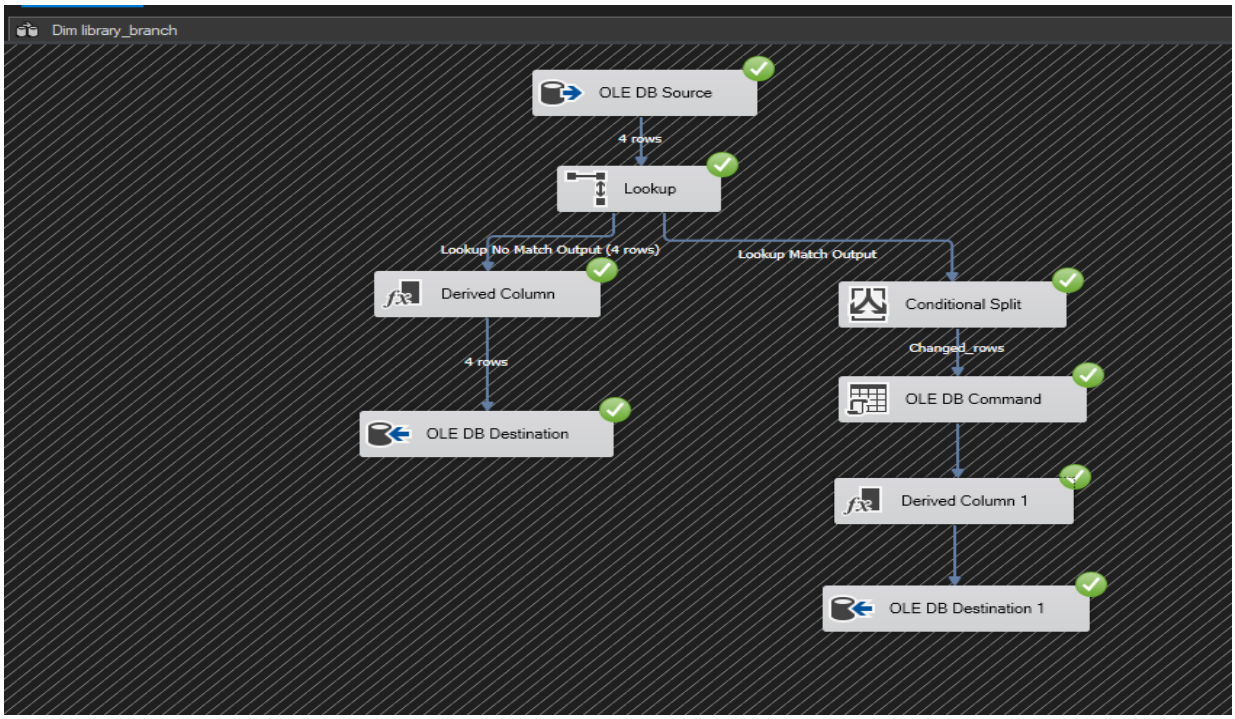
Control flow



Data Flow (Source to Stage)



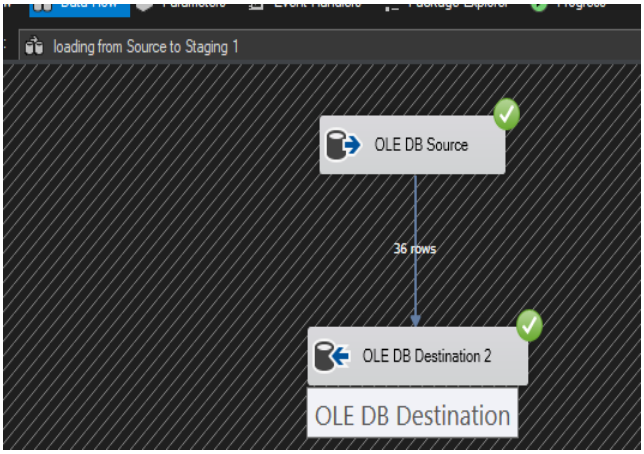
Data flow Dim library_branch(from staging to dimension library branch in datawarehouse applying scd type2):-



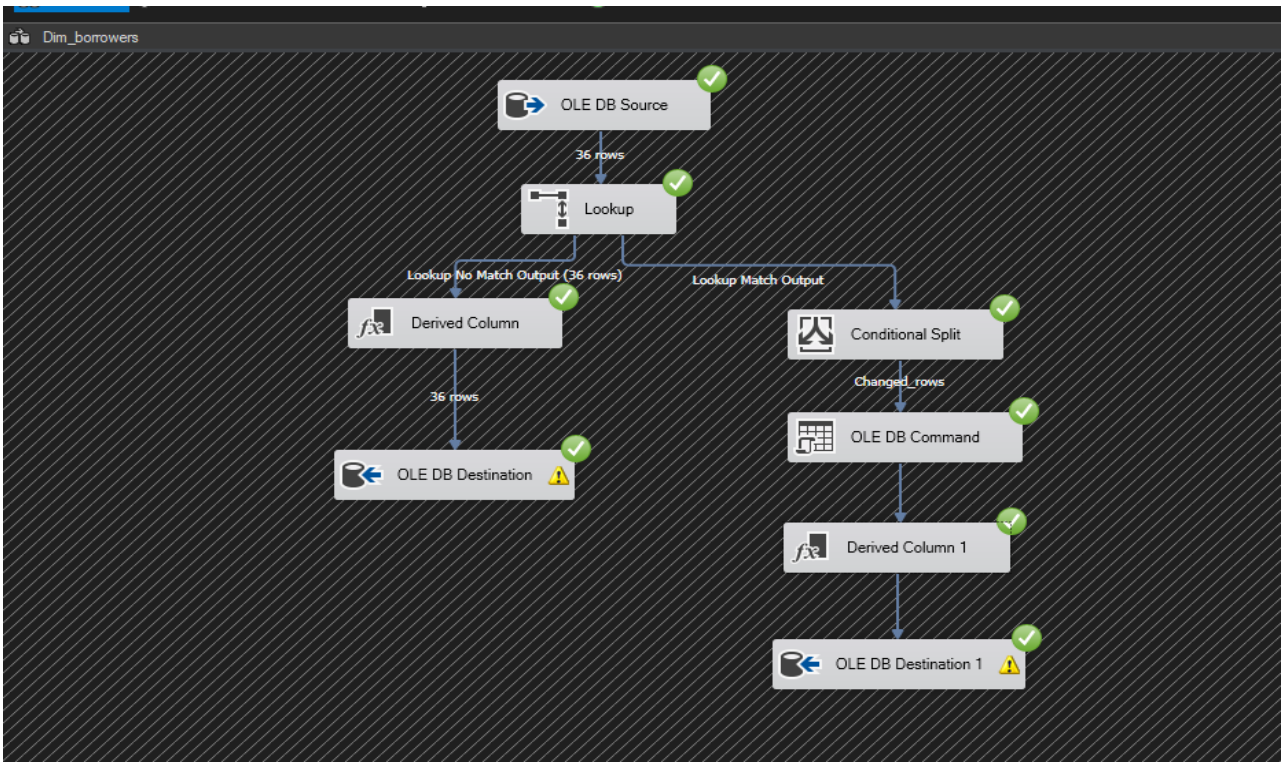
7-Borrower dimension:-

Control flow

Data flow(From source to stage)

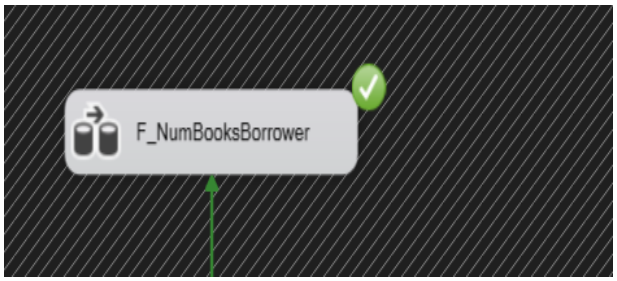


Data flow Dim Borrower(from staging to dimension borrower in datawarehouse applying scd type2):-

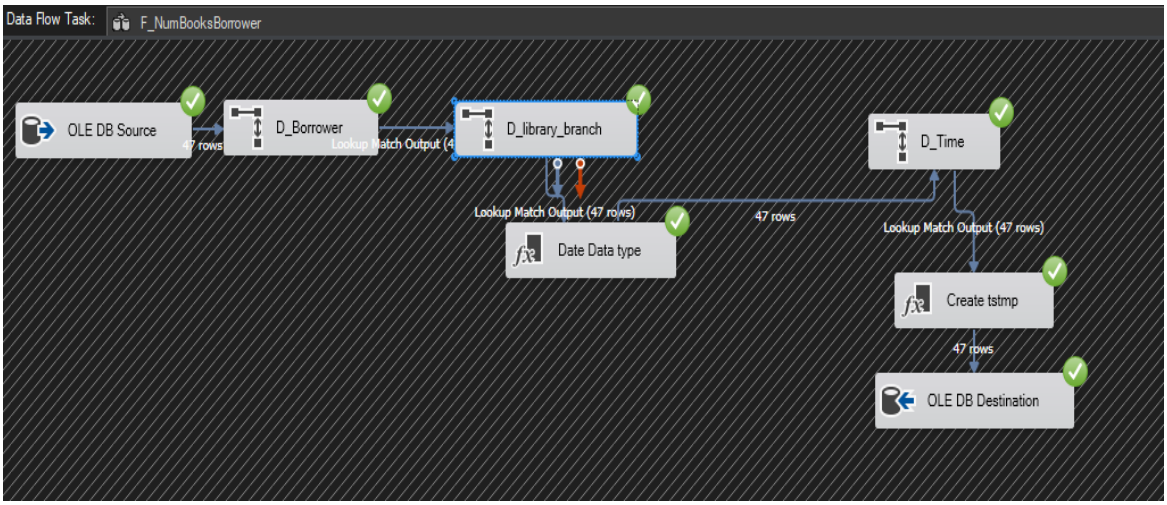


8-F_NumBooksBorrower

Control Flow

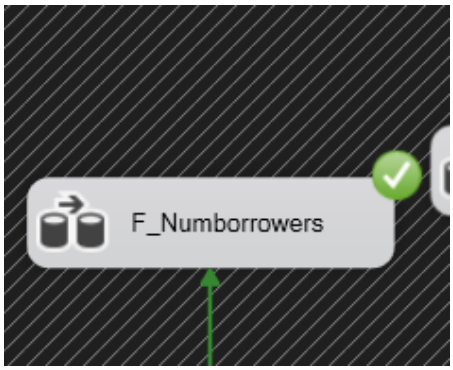


Data flow from dimensions (D_library_branch – D_borrower and D_time) to fact table (F_NumBooksBorrower):-

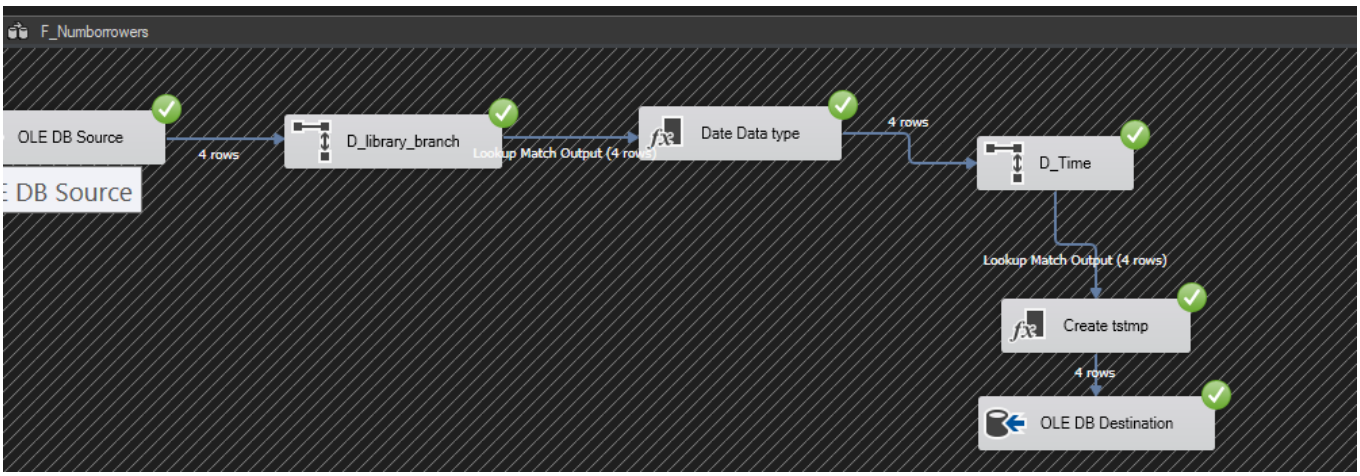


9-F_NumBorrowers

Control flow



Data flow from dimensions (D_library_branch - D_time) to fact table (F_NumBorrowers):-



5. Queries on each fact table and a screenshot of the result set of each query.

1A-F_loans_book_author fact table (first query)

Explanation: this query get all the author's name that there books have been loaned with the summation of the total loans and display it in decreasing order.

```
SELECT
a.AuthorName,
SUM(f.total_loans) AS TotalLoans
FROM
    F_loans_book_author f
JOIN
    D_author a ON f.author_key = a.Author_key
GROUP BY
    a.AuthorName
ORDER BY
    TotalLoans DESC;
```

124 %

Results | Messages

	AuthorName	TotalLoans
1	Patrick Rothfuss	10
2	Frank Herbert	9
3	J.R.R. Tolkien	9
4	last222222	8
5	Stephen King	7
6	nour	7
7	Chuck Palahniuk	7
8	Douglas Adams	7
9	last again22222222222222	6
10	Mark Lee	6
11	Aldous Huxley	5
12	last	5
13	George R.R. Martin	5

1B-F_loans_book_author fact table (second query)

Explanation: this query get all the book's name that have been loaned at a specific month and specific year and return the summation of it.

```
SELECT
t.Year,
t.Month,
b.Title AS BookTitle,
SUM(f.total_loans) AS TotalLoans
FROM
    F_loans_book_author f
JOIN
    D_book b ON f.book_key = b.Book_key
JOIN
    D_Time t ON f.date_key = t.DateKey
GROUP BY
    t.Year, t.Month, b.Title
ORDER BY
    t.Year, t.Month, TotalLoans DESC;
```

124 %

Results | Messages

	Year	Month	BookTitle	TotalLoans
1	2018	4	The Hitchhikers Guide to the Galaxy	8
2	2018	4	Brave New World	7
3	2018	4	Nourhan	7
4	2018	4	Nourhannnnnnnnnnnn	6
5	2018	4	Holes	6
6	2018	4	The Giving Tree	5
7	2018	4	New again	5
8	2018	4	Nourhannnnnnnnnnnn111111111111	5
9	2018	5	The Name of the Wind	10
10	2018	5	The Green Mile	9
11	2018	5	Dune	9
12	2018	5	It	7
13	2018	5	The Hobbit	7

2A-F_loanDuration fact table (first query)

Explanation: this query get all the library_branch's name and the summation of the total loans in it then sort data desc.

```
--                  queryyyyyyyyyyyyyyy      on       fact       tables

SELECT
    b.BranchName,
    SUM(f.total_loans) AS TotalLoans
FROM
    F_loanDuration f
JOIN
    D_library_branch b ON f.branch_key = b.Branch_Key
GROUP BY
    b.BranchName
ORDER BY
    TotalLoans DESC;
```

124 %

Results | Messages

	BranchName	TotalLoans
1	Ann Arbor	46
2	Saline	32
3	Central	30
4	Sharpstown	22

2B-F_loanDuration fact table (second query)

Explanation: this query get the summation of the total loans that was in specific month in specific year.

```
SELECT
t.Year,
t.Month,
SUM(f.total_loans) AS TotalLoans
FROM
    F_loanDuration f
JOIN
    D_Time t ON f.date_key = t.DateKey
GROUP BY
    t.Year, t.Month
ORDER BY
    t.Year, t.Month;
```

124 %

Results | Messages

	Year	Month	TotalLoans
1	2018	4	84
2	2018	5	46

3A- F_NumBooksBorrower fact table (first query)

Explanation: this query gets top borrowers by total books borrowed

```
SELECT
    d_borrower.BorrowerName,
    SUM(f.NumBooksBorrowed) AS TotalBooksBorrowed
FROM F_NumBooksBorrower AS f
JOIN D_Borrower AS d_borrower ON f.borrower_key = d_borrower.Borrower_key
GROUP BY d_borrower.BorrowerName
ORDER BY TotalBooksBorrowed DESC;
```

Results Messages

BorrowerName	TotalBooksBorrowed
Pierre Rosenfrucht	14
Anya Munehay	11
Pam Carson	8
Gregoire Clapp	7
Fleurette Kleeborn	6
Boony Trevaskis	6
Mattheus Hovard	6
Licha Inglese	6
Eugenia Worrall	5
Damara Irvine	5
Jacintha Blackly	3
Brady Curtiss	3
Claudell Dipple	3
Baird Diggin	3
Bernetta Rubino...	3
Malena Fawbert	3
Kath Griffen	3
Richmound Falcus	3
Rowen McGuane	3
Storm Edgeller	3
Sybyl Deem	3
Tressa Champe	3
Ulysses Landman	2
Nessy Waterfield	2

3B-F_NumBooksBorrower fact table (second query)

Explanation: this query gets total number of books borrowed by branch and date

```
SELECT
    d_branch.BranchName,
    d_time.Month,
    SUM(f.NumBooksBorrowed) AS TotalBooksBorrowed
FROM F_NumBooksBorrower AS f
JOIN D_library_branch AS d_branch ON f.branch_key = d_branch.Branch_Key
JOIN D_Time AS d_time ON f.date_key = d_time.DateKey
GROUP BY
    d_branch.BranchName,
    d_time.Month
ORDER BY d_time.month;
```

150 %

Results Messages

	BranchName	Month	TotalBooksBorrowed
1	Ann Arbor	1	20
2	Central	1	5
3	Saline	1	10
4	Sharpstown	1	10
5	Central	2	6
6	Ann Arbor	3	1
7	Central	3	10
8	Saline	3	20
9	Sharpstown	3	9
10	Ann Arbor	4	21
11	Central	4	9
12	Saline	4	3
13	Sharpstown	4	6
14	Ann Arbor	5	5

3c- F_NumBooksBorrower fact table (third query)

Explanation: this query gets total number of borrowed books per specific branch and date

```
-- Total Books Borrowed in a Specific Branch and Month (F_NumBooksBorrowed)
SELECT
    d_branch.BranchName,
    d_time.MonthName,
    SUM(f.NumBooksBorrowed) AS TotalBooks
FROM F_NumBooksBorrower AS f
JOIN D_library_branch AS d_branch ON f.branch_key = d_branch.Branch_Key
JOIN D_Time AS d_time ON f.date_key = d_time.DateKey
WHERE d_branch.BranchName = 'Ann Arbor' AND d_time.MonthName = 'May'
GROUP BY d_branch.BranchName, d_time.MonthName;
```

Messages

nchName	MonthName	TotalBooks
n Arbor	May	5

4- F_NumBorrowers fact table

Explanation: this query gets total number of borrowers per branch per year

```
--Total Number of Borrowers Per Branch
SELECT
    d_branch.BranchName,
    d_time.Year,
    SUM(f.NumBorrowers) AS TotalBorrowers
FROM F_NumBorrowers AS f
JOIN D_library_branch AS d_branch ON f.branch_key = d_branch.Branch_Key
JOIN D_Time AS d_time ON f.date_key = d_time.DateKey
GROUP BY d_branch.BranchName, d_time.Year
ORDER BY d_time.Year;
```

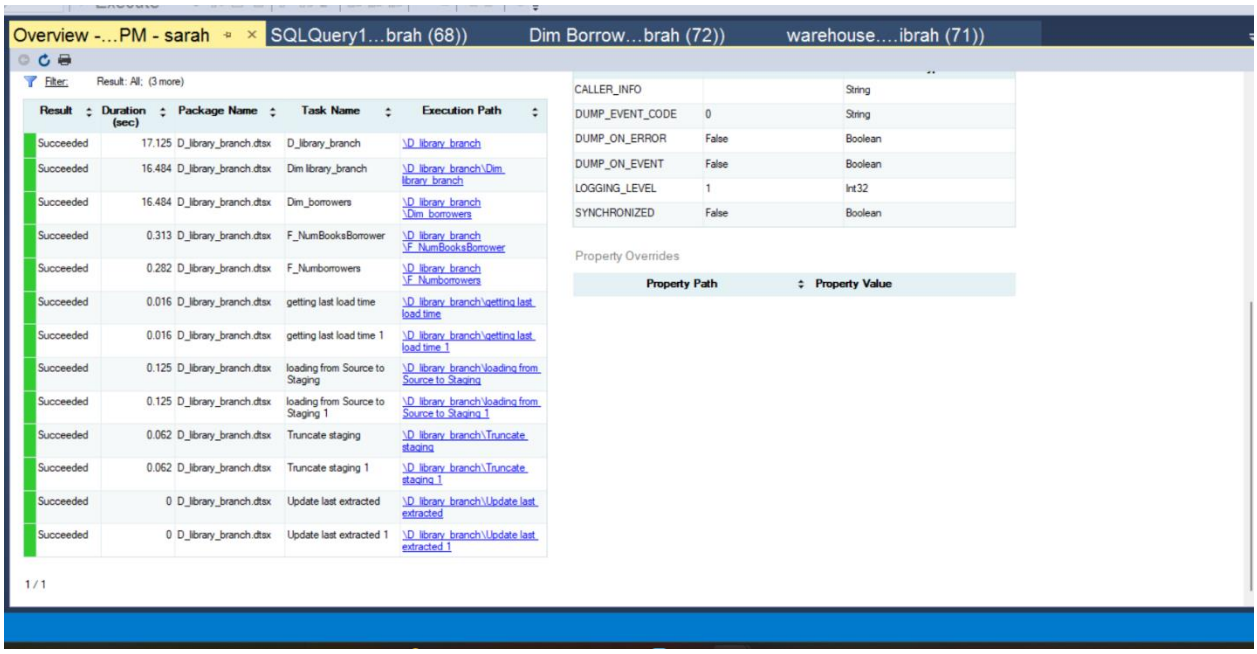
150 %

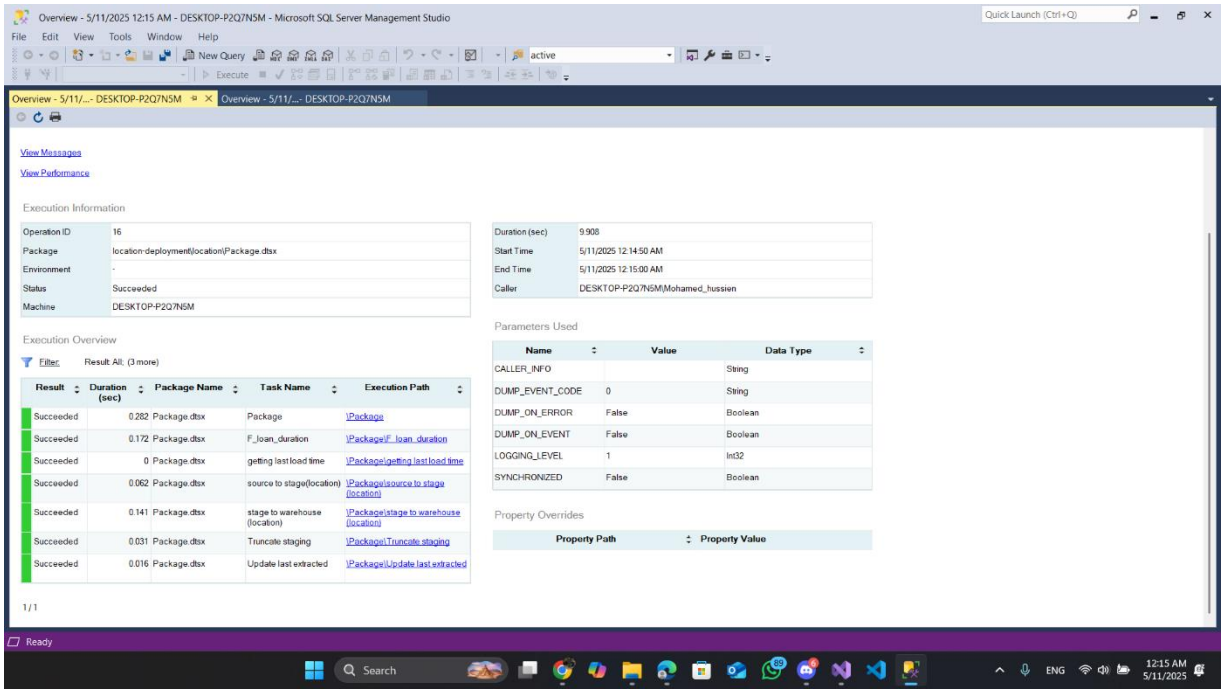
Results Messages

	BranchName	Year	TotalBorrowers
1	Ann Arbor	2018	15
2	Central	2018	11
3	Saline	2018	13
4	Sharpstown	2018	8

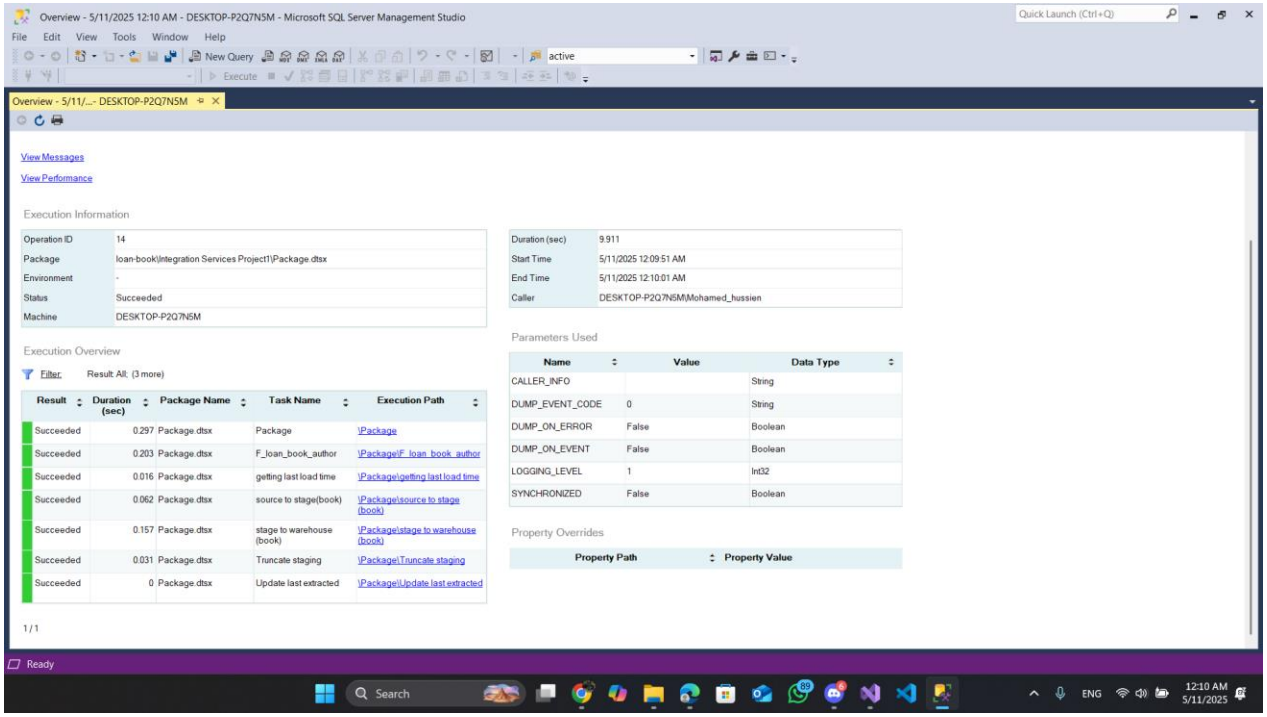
جامعة القاهرة

Report overview of deployment

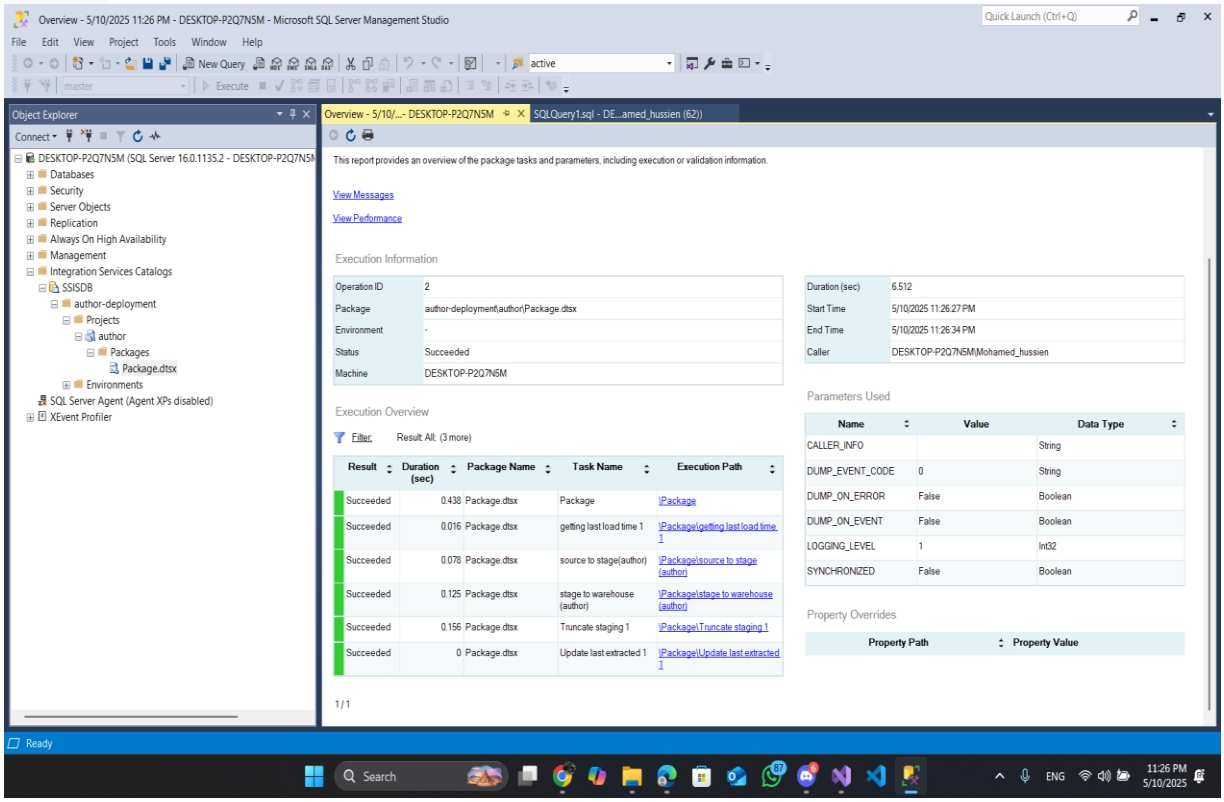


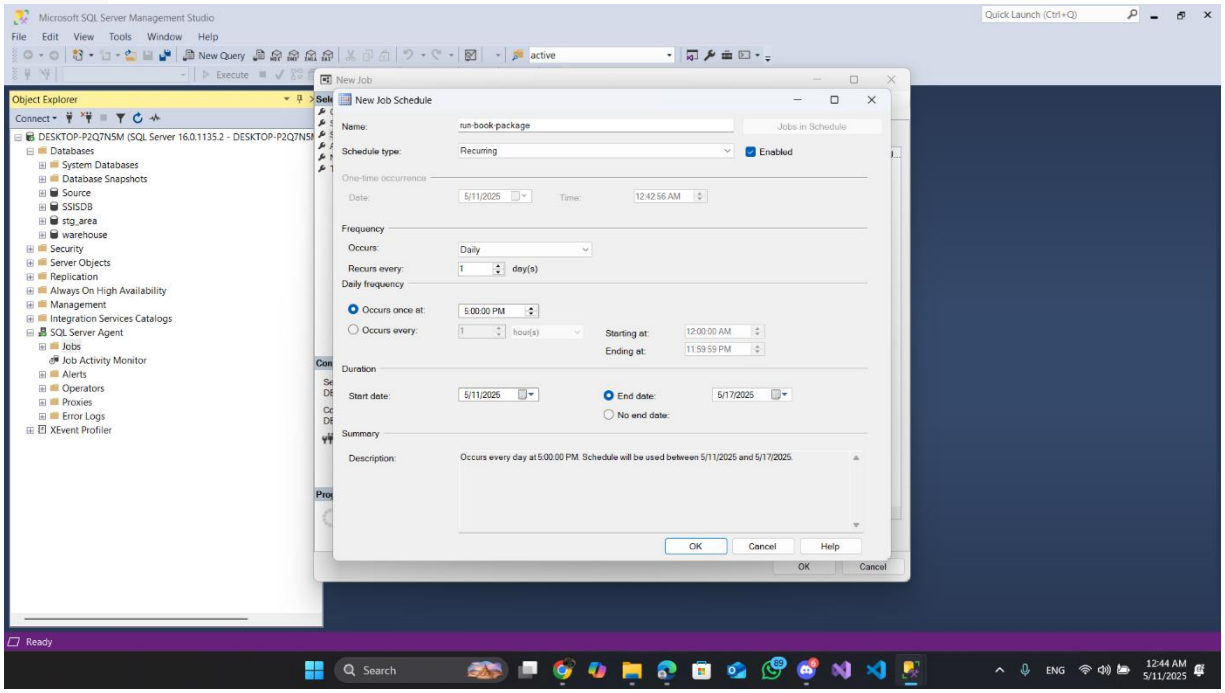


جامعة القاهرة

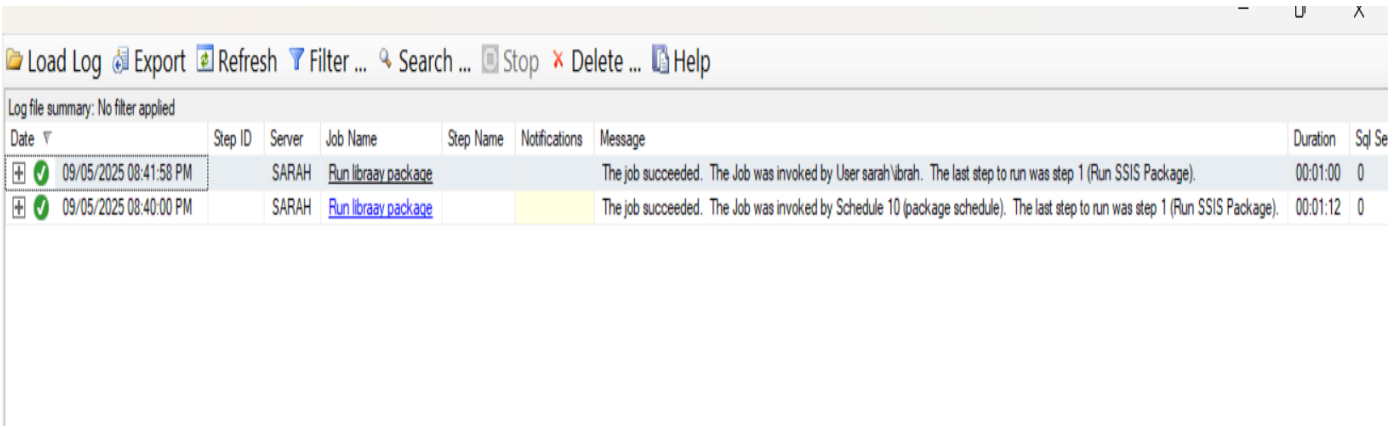


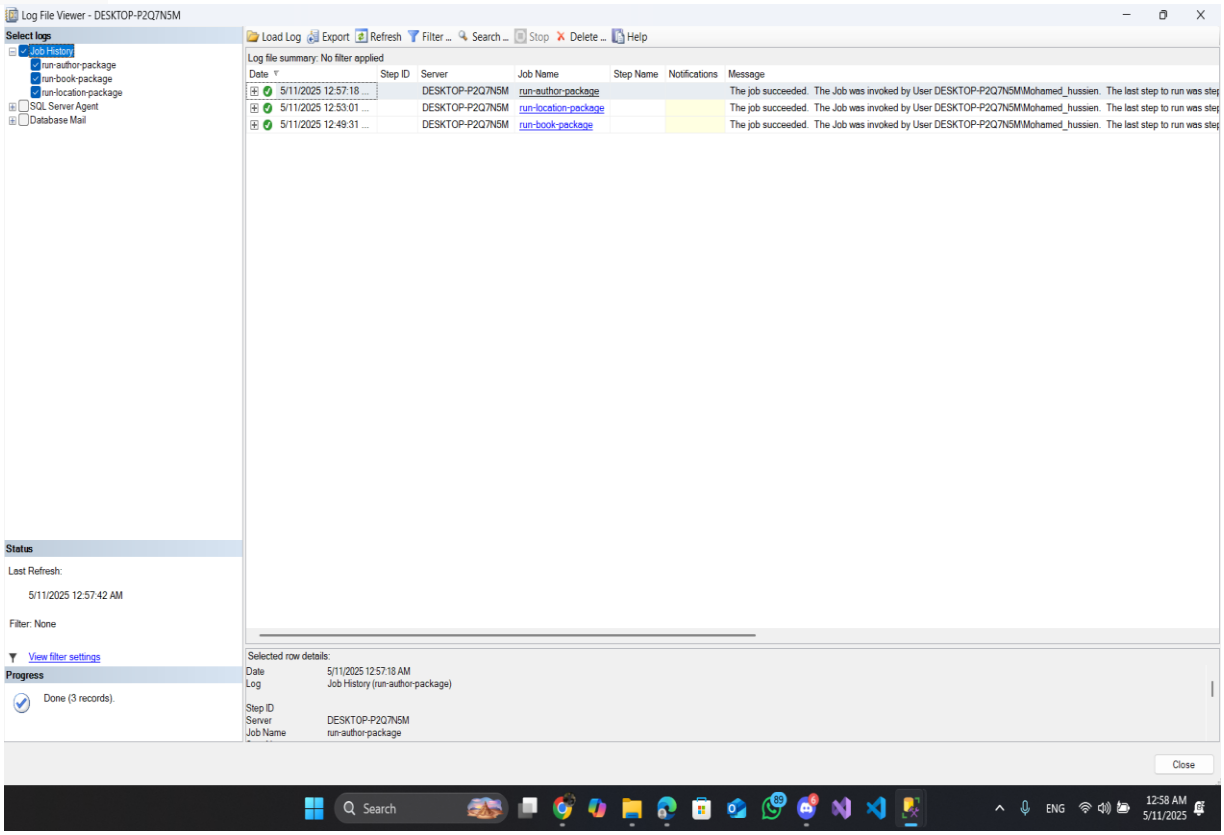
جامعة القاهرة



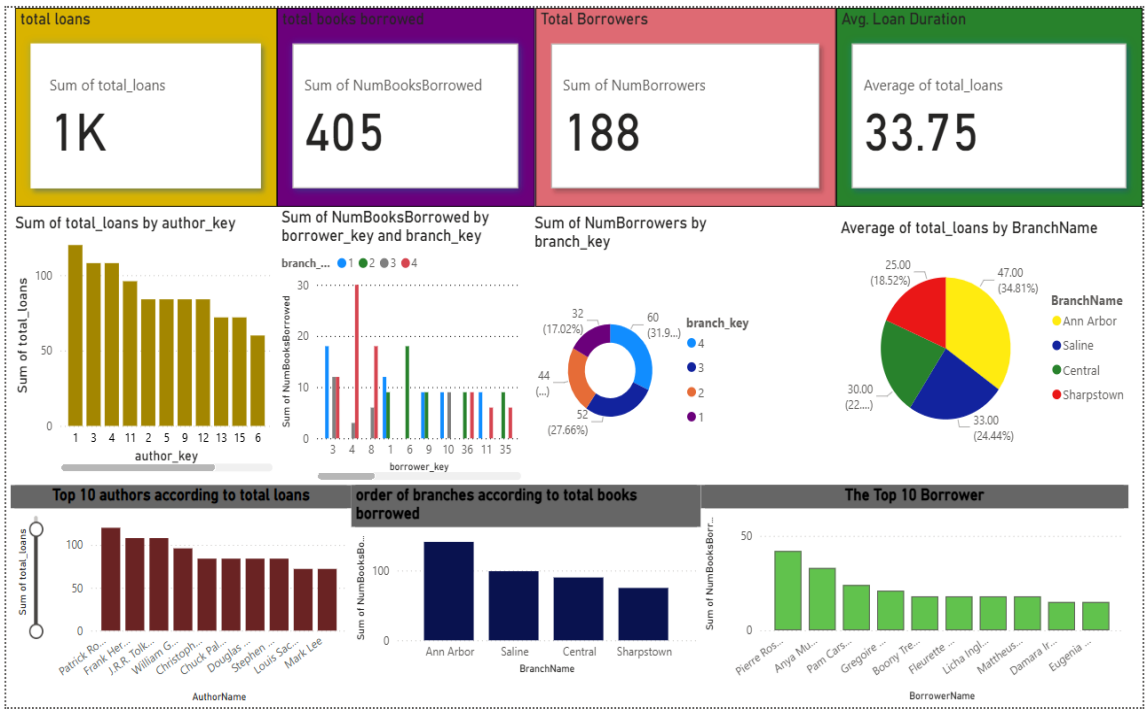


Jobs Scheduling Screenshots:

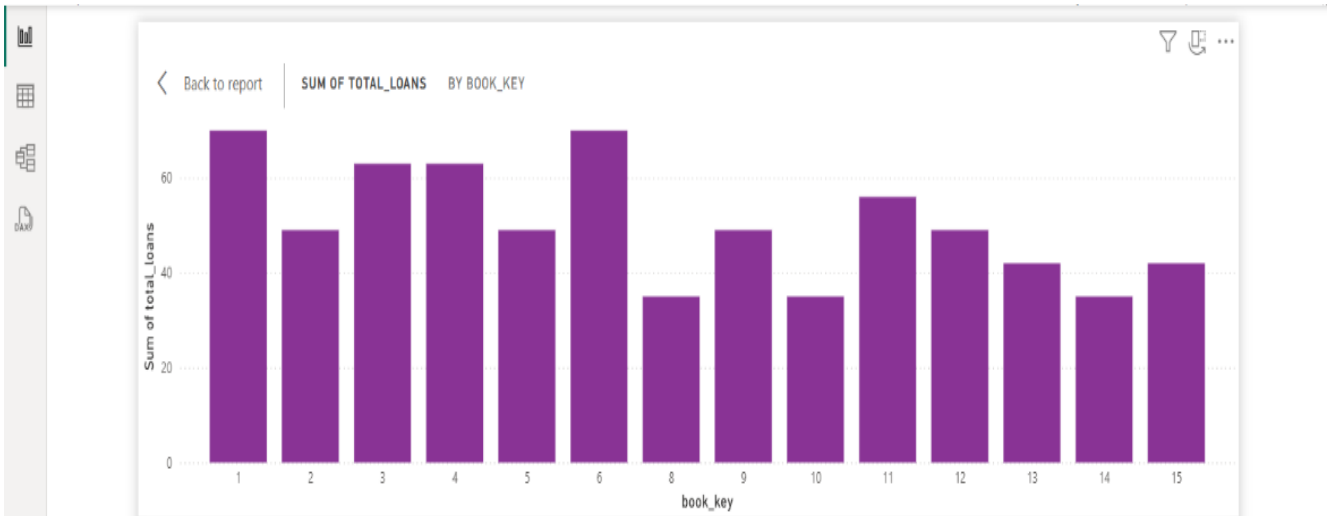




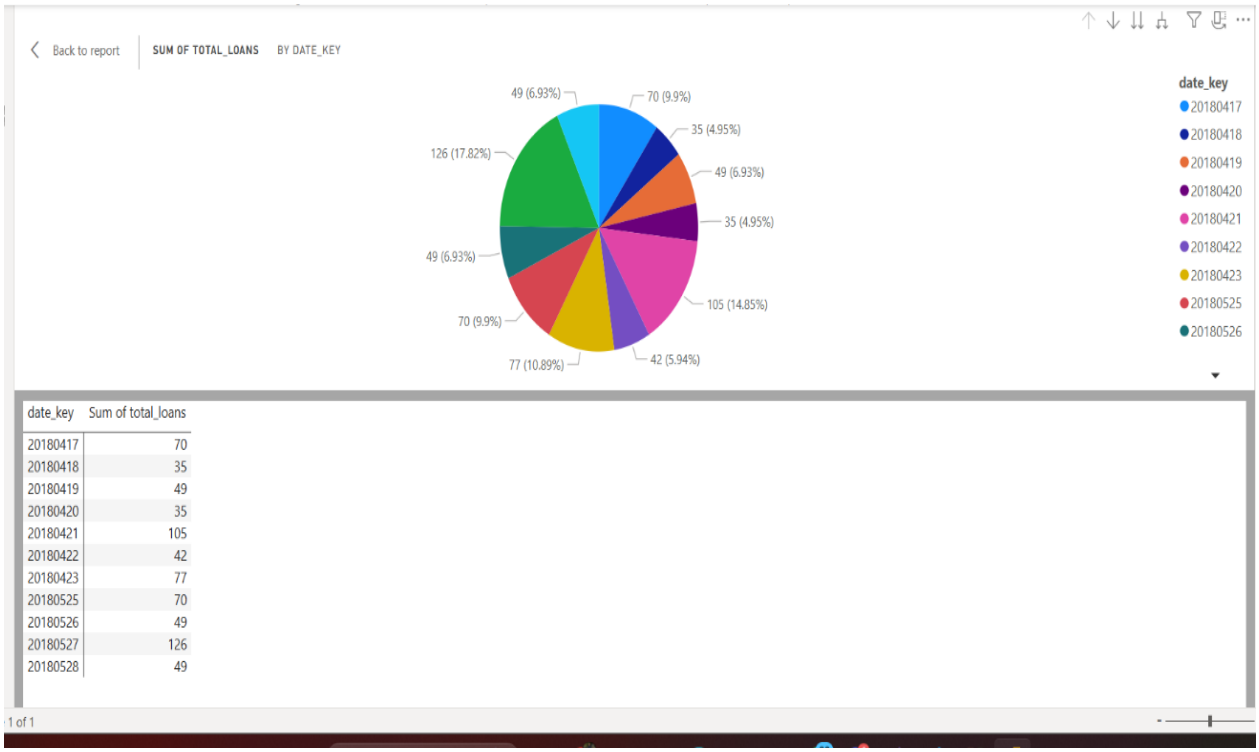
7. [Bonus] Build an interactive dashboard for the DWH using any data visualization tool (Ex: Microsoft Power BI).



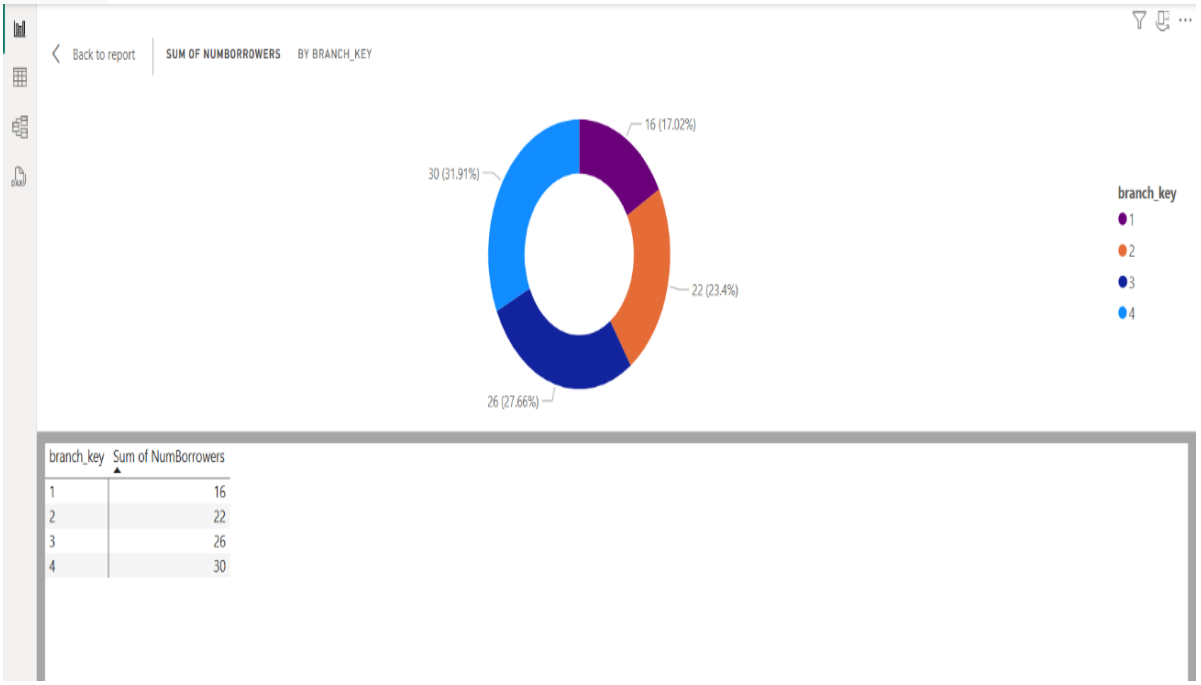
author_key	Sum of total_loans
1	70
2	49
3	63
4	63
5	49
6	35
7	35
8	35
9	49
10	35
11	56
12	49
13	42
14	35



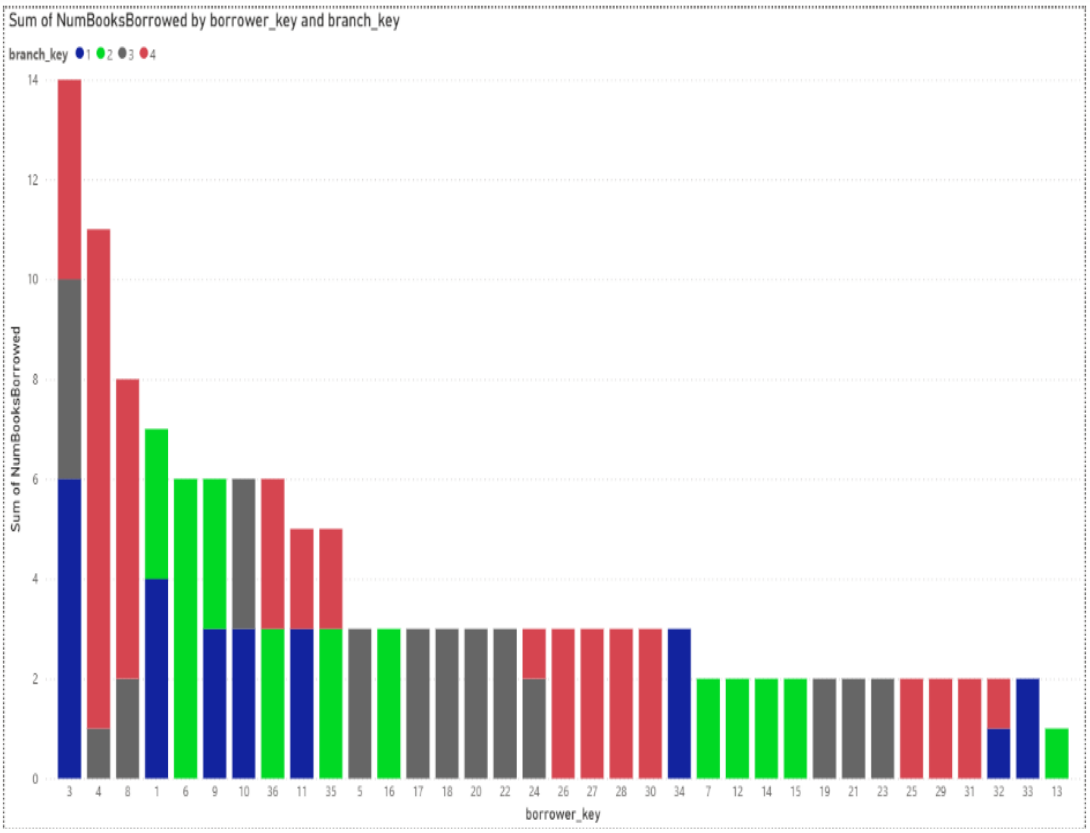
book_key	Sum of total_loans
1	70
2	49
3	63
4	63
5	49
6	70
8	35
9	49
10	35
11	56
12	49
13	42
14	35
15	42

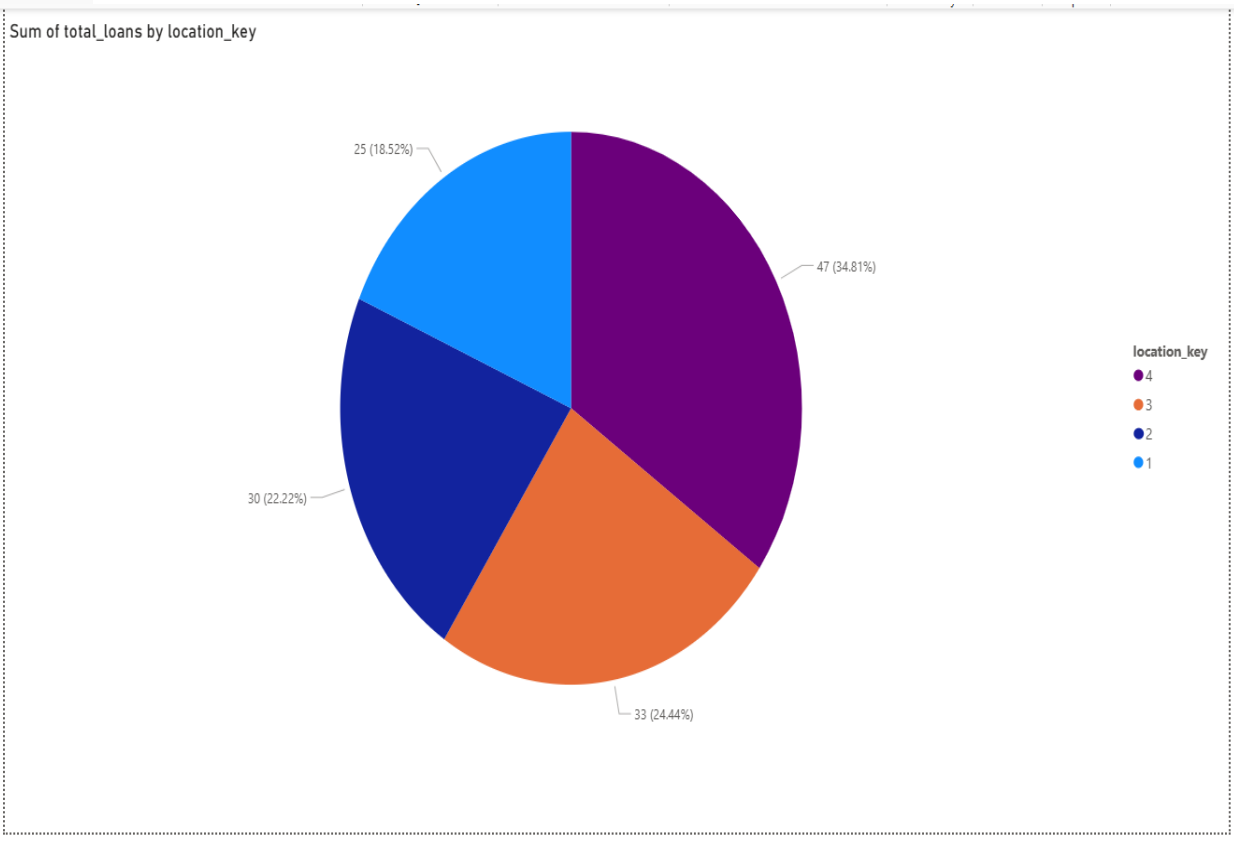


date_key	Sum of total_loans
20180417	70
20180418	35
20180419	49
20180420	35
20180421	105
20180422	42
20180423	77
20180525	70
20180526	49
20180527	126
20180528	49



branch_key	Sum of NumBorrowers
1	16
2	22
3	26
4	30




Sum of total_loans by location_key
25 (18.52%)
47 (34.81%)
30 (22.22%)
33 (24.44%)
location_key
4
3
2
1