

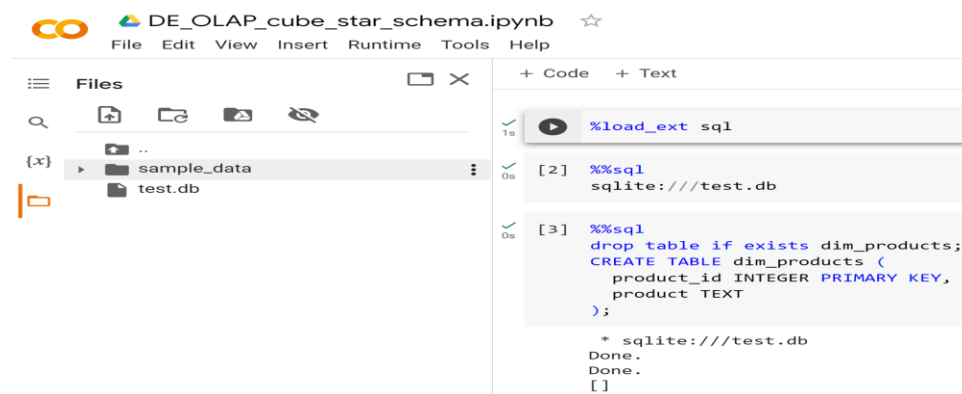
Purpose:

The purpose of this project is to extract data from the 'test.db' SQLite database, comprising the tables 'dim_time,' 'dim_geography,' 'dim_products,' and 'fact_sales.' The extracted data is then transformed and exported into separate CSV files. The primary objective is to enable efficient data analysis and further processing of the information stored within the database.

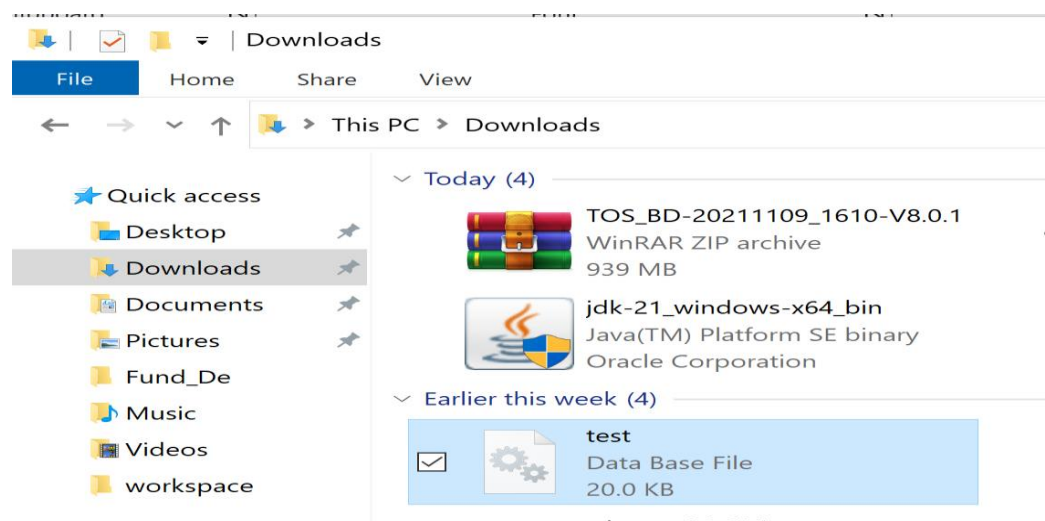
Downloading the Database File for Project Initialization

To access the data source for our project, we will need to run the Colab link provided in the description. This step is crucial as it enables us to download the 'test.db' file from the Collab files. The 'test.db' file contains the essential data that we will be working with in our project. By running the Colab link, we ensure that we have the necessary database file readily available on our local computer, allowing us to establish a connection and retrieve the required data for our analysis and processing tasks in Talend. It serves as the foundation for our subsequent data extraction and transformation efforts within the Talend environment.

Below Image is the screenshot from where we downloaded the test.db file located in collab files,



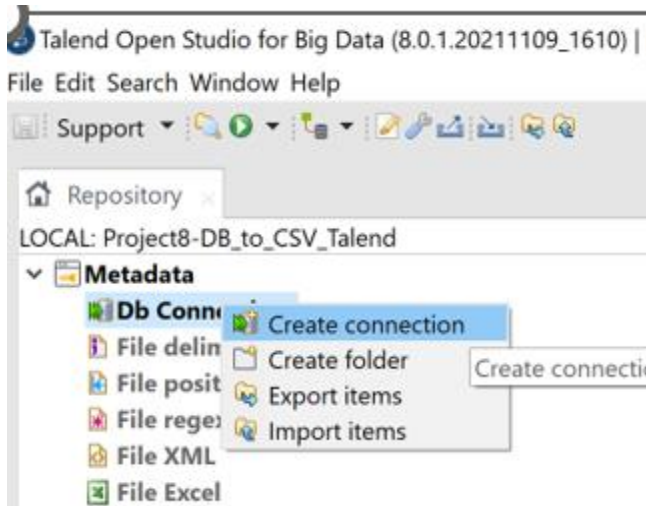
Below image shows the file where test_db stored in my computer,



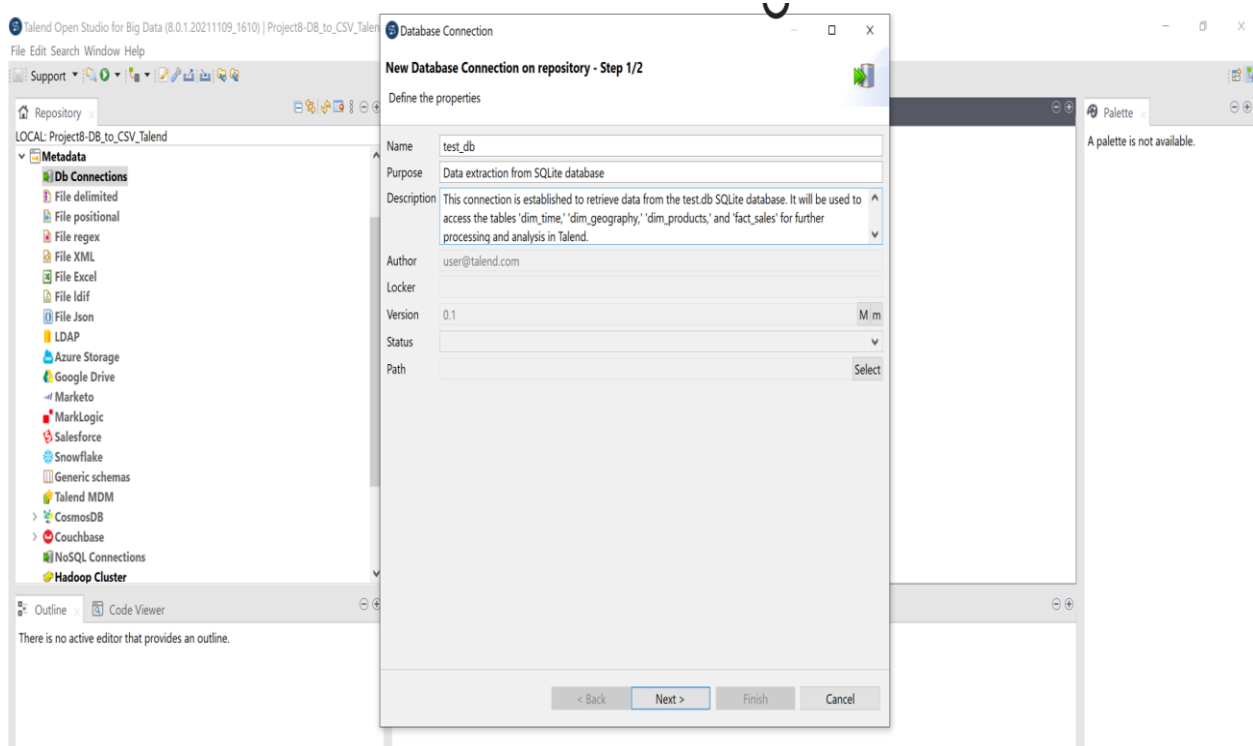
Meta Data Setup:

We Create a new DB connection named 'test_db' in the Metadata menu to establish a connection with the 'test.db' SQLite database. This connection allows access to the database and its tables.

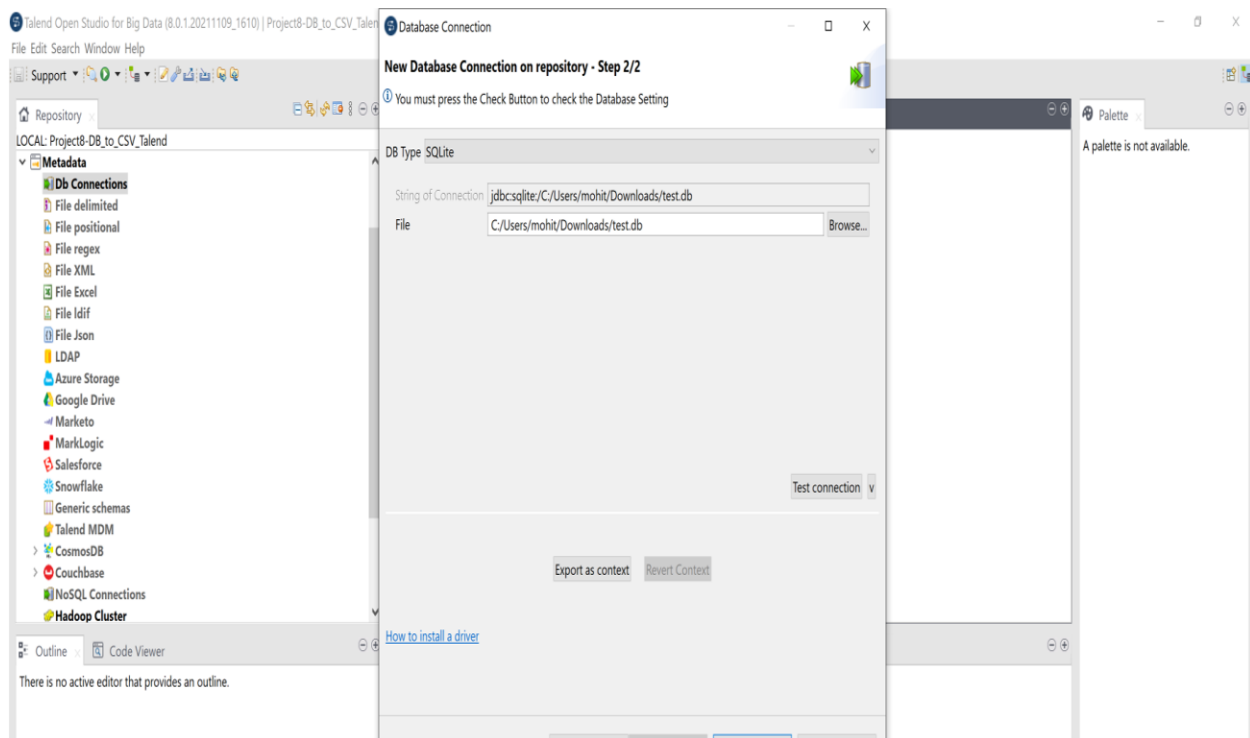
Below image shows the first step in creating a Db connection,



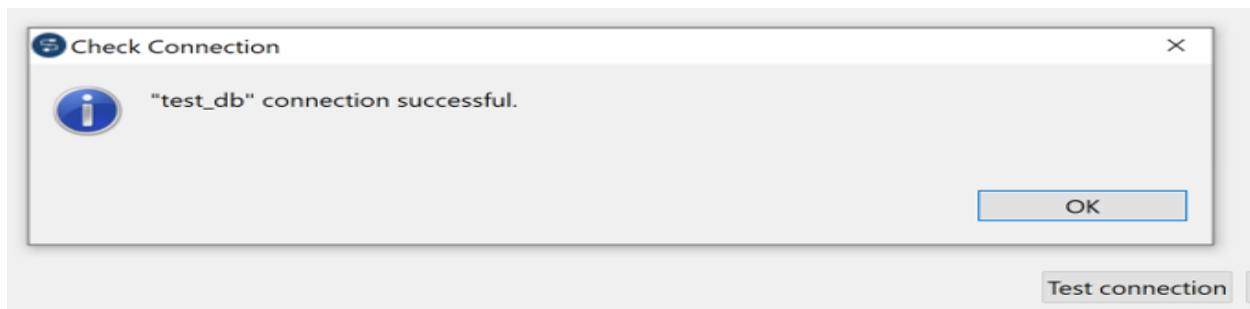
Below image indicates on how we setup the new db connection(status was selected as development),



Below image indicates additional information in initiating the connection such as selecting db type and mentioning the path for the local test.db file,

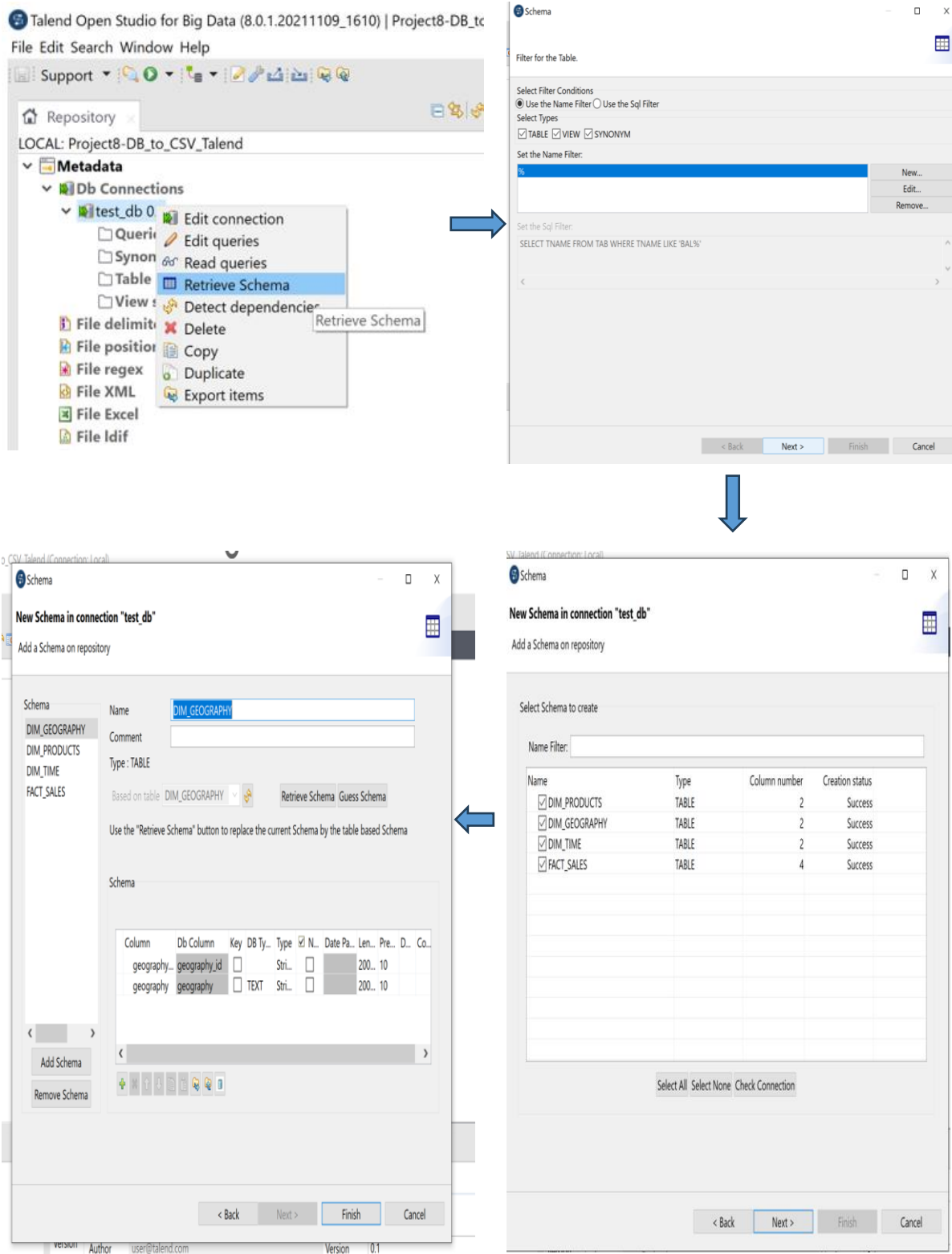


Finally we test whether the Db connection is successful or not,



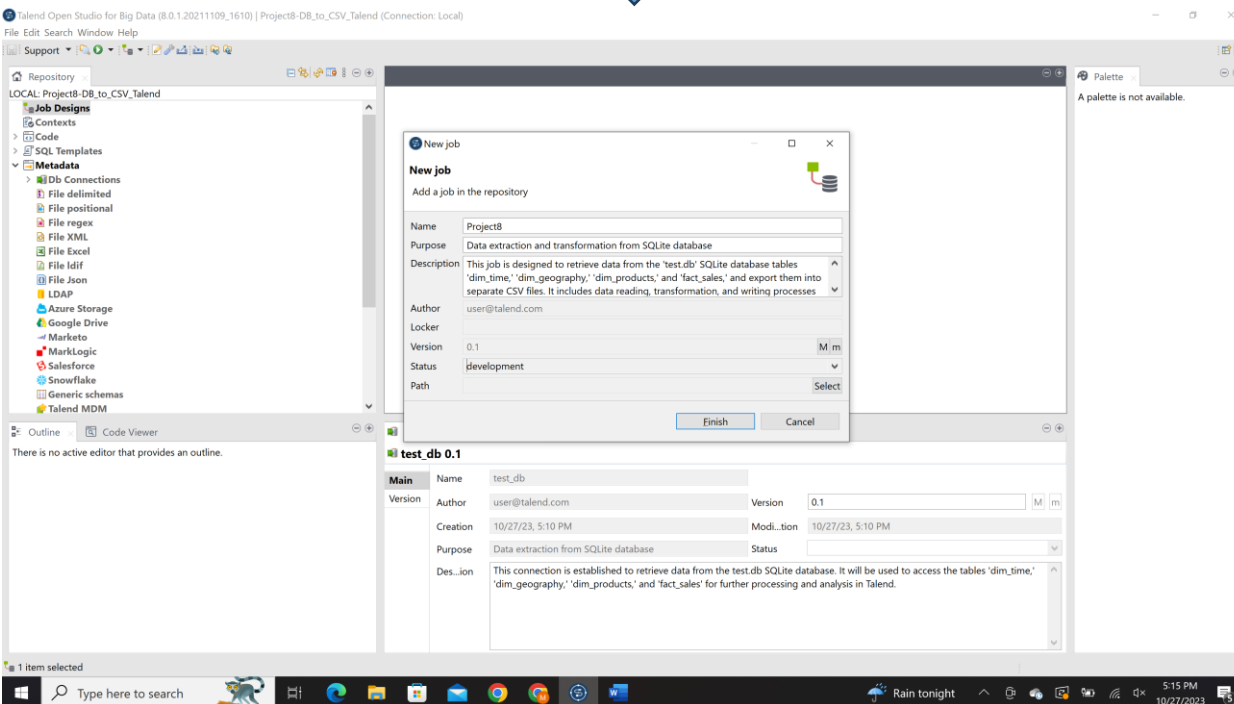
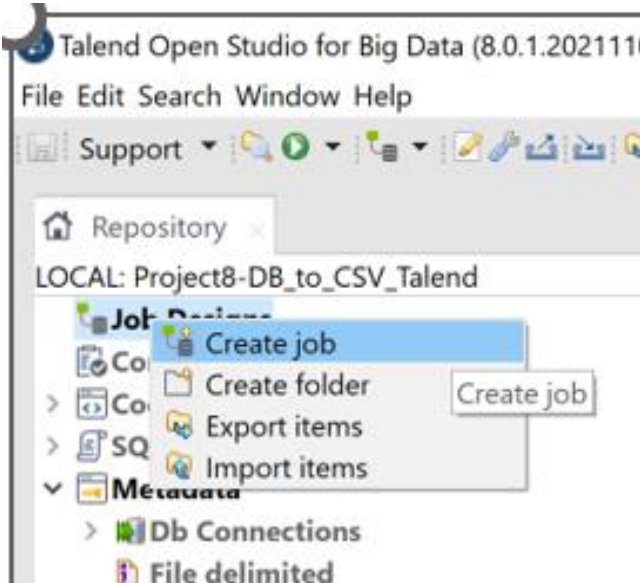
The above image is a reference to show that our Db connection was successful.

Once the connection indicates that its successful, we then retrieve all schema of the test_db as shown in images below also note that the arrows indicate the flow between all the steps that were done,



Job Creation:

We create a new job called 'Project8' under the Job designs menu in Talend to initiate the data extraction and transformation processes.

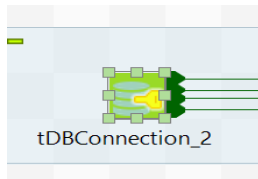


Job Design Overview:

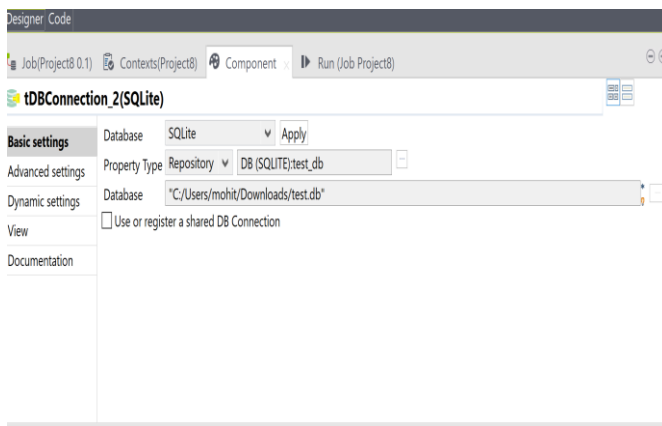
The job 'Project8' is designed to extract data from the 'test.db' SQLite database and export it into separate CSV files for further analysis. It comprises four main components, each serving a specific purpose within the data extraction and transformation process.

tDBConnection Component:

The tDBConnection component establishes and maintains the connection with the database, enabling smooth data interaction and retrieval for the job.



Component configuration for tDBConnection,

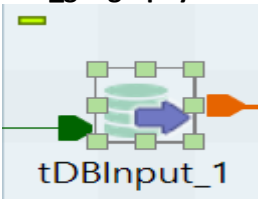


tSQLiteInput Component:

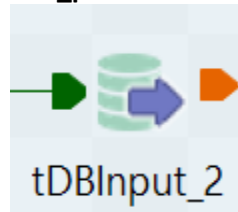
This component establishes the connection with the 'test.db' database and retrieves data from the tables.

Below are the images of tables and their tSQLiteInput Component used for the job,

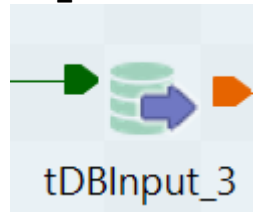
dim_geography



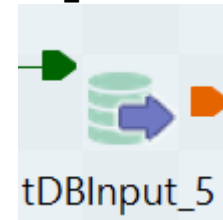
dim_products



dim_time



fact_sales



Below images shows all four tSQLiteInput Component basic settings,

Job(Project8 0.1) Contexts(Project8) Component x Run (Job Project8)

tDBInput_1(SQLite)

Basic settings	Database	SQLite	Apply
Advanced settings	Property Type	Repository	DB (SQLite):test_db
Dynamic settings	<input type="checkbox"/> Use an existing connection		
View	Database	"C:/Users/mohit/Downloads/test.db"	
Documentation	Schema	Repository	DB (SQLite):test_db - DIM_GEOGRA Edit schema
	Table Name	"DIM_GEOGRAPHY"	
	Query Type	Built-In	Guess Query Guess schema
	Query	<pre>"SELECT DIM_GEOGRAPHY.geography_id, DIM_GEOGRAPHY.geography FROM DIM_GEOGRAPHY"</pre>	

Job(Project8 0.1) Contexts(Project8) Component x Run (Job Project8)

tDBInput_2(SQLite)

Basic settings	Database	SQLite	Apply
Advanced settings	Property Type	Repository	DB (SQLite):test_db
Dynamic settings	<input type="checkbox"/> Use an existing connection		
View	Database	"C:/Users/mohit/Downloads/test.db"	
Documentation	Schema	Repository	DB (SQLite):test_db - DIM_PRODUC Edit schema
	Table Name	"DIM_PRODUCTS"	
	Query Type	Built-In	Guess Query Guess schema
	Query	<pre>"SELECT DIM_PRODUCTS.product_id, DIM_PRODUCTS.product FROM DIM_PRODUCTS"</pre>	

Job(Project8 0.1) Contexts(Project8) Component x Run (Job Project8)

tDBInput_3(SQLite)

Basic settings	Database	SQLite	Apply
Advanced settings	Property Type	Repository	DB (SQLite):test_db
Dynamic settings	<input type="checkbox"/> Use an existing connection		
View	Database	"C:/Users/mohit/Downloads/test.db"	
Documentation	Schema	Repository	DB (SQLite):test_db - DIM_TIME Edit schema
	Table Name	"DIM_TIME"	
	Query Type	Built-In	Guess Query Guess schema
	Query	<pre>"SELECT DIM_TIME.time_id, DIM_TIME.time_quarter FROM DIM_TIME"</pre>	

tDBInput_5(SQLite)

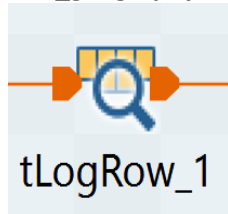
Basic settings	Database	SQLite	Apply
Advanced settings	Property Type	Repository	DB (SQLite):test_db
Dynamic settings	<input type="checkbox"/> Use an existing connection		
View	Database	"C:/Users/mohit/Downloads/test.db"	
Documentation	Schema	Repository	DB (SQLite):test_db - FACT_SALES
	Table Name	"FACT_SALES"	
	Query Type	Built-In	Guess Query Guess schema
	Query	<pre>"SELECT FACT_SALES.product_id, FACT_SALES.geography_id, FACT_SALES.time_id, FACT_SALES.SalesAmt FROM FACT_SALES"</pre>	

tLogRow Component:

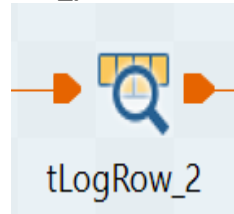
The tLogRow component displays the data contents in the console, allowing us to verify the accuracy of the retrieved data from the database tables.

Below are the images of tables and their tLogRow Component used for the job,

dim_geography



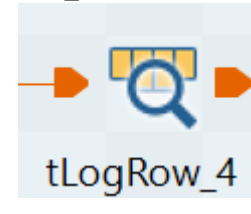
dim_products



dim_time



fact_sales

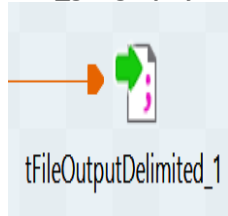


tFileOutputDelimited Component:

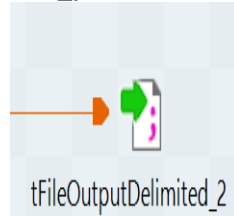
The tFileOutputDelimited component writes the extracted data into separate CSV files, ensuring that each table's data is stored in a dedicated file for further processing and analysis.

Below are the images of tables and their tFileOutputDelimited Component used for the job,

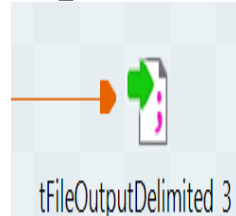
dim_geography



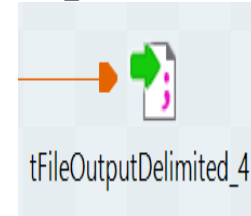
dim_products



dim_time



fact_sales



Below images shows all four tFileOutputDelimited Component basic settings,

tFileOutputDelimited_1

Basic settings	Property Type Built-In	
Advanced settings	<input type="checkbox"/> Use Output Stream	
Dynamic settings	File Name "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/DIM_GEOGRAPHY.csv"	
View	Row Separator "\n"	Field Separator ","
Documentation	<input type="checkbox"/> Append <input checked="" type="checkbox"/> Include Header <input type="checkbox"/> Compress as zip file	
	Schema Built-In	Edit schema Sync columns

tFileOutputDelimited_2

Basic settings	Property Type Built-In	
Advanced settings	<input type="checkbox"/> Use Output Stream	
Dynamic settings	File Name "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/DIM_PRODUCTS.csv"	
View	Row Separator "\n"	Field Separator ","
Documentation	<input type="checkbox"/> Append <input checked="" type="checkbox"/> Include Header <input type="checkbox"/> Compress as zip file	
	Schema Built-In	Edit schema Sync columns

Job(Project8 0.1) Contexts(Project8) Component x Run (Job Project8)

tFileOutputDelimited_3

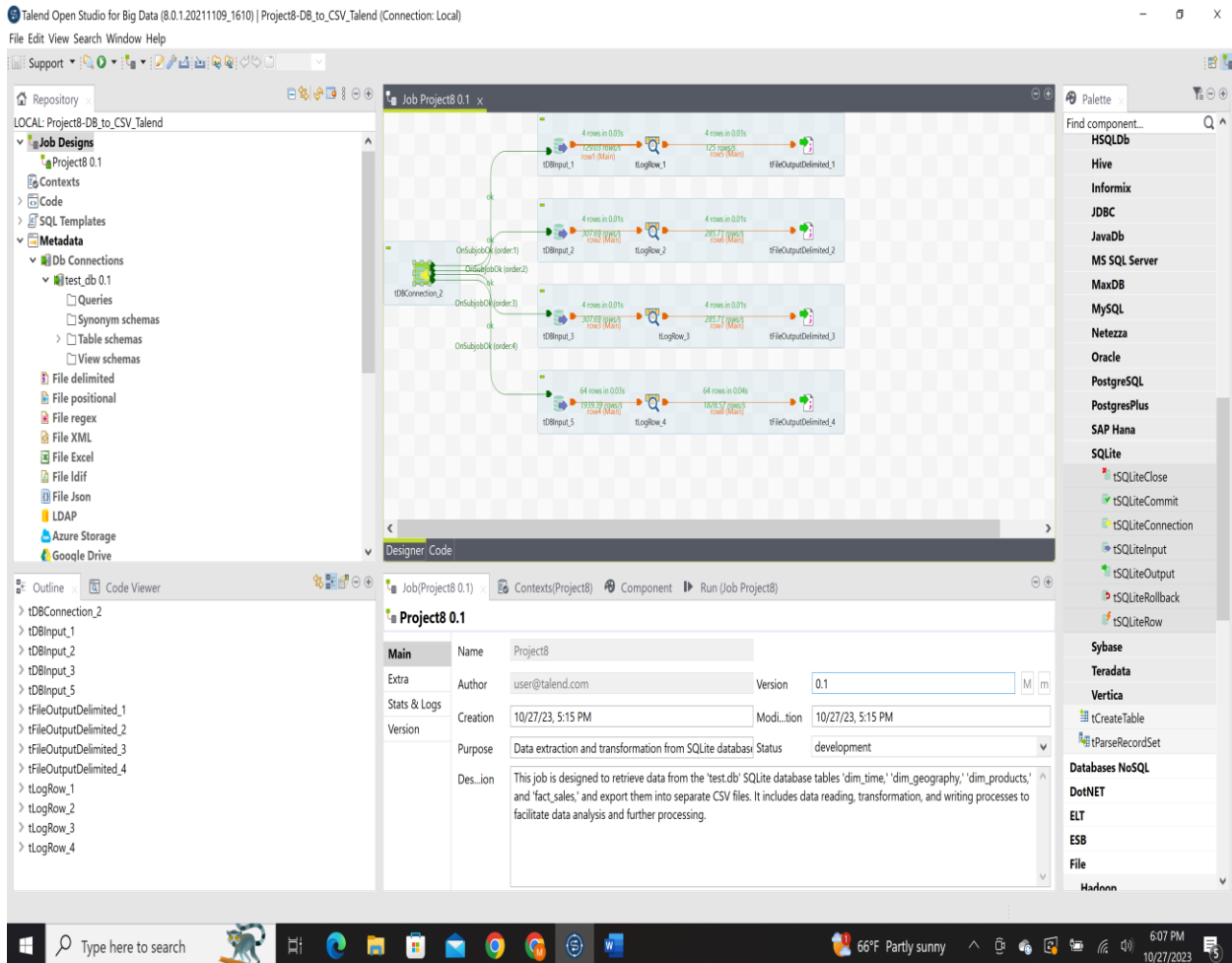
Basic settings	Property Type Built-In	
Advanced settings	<input type="checkbox"/> Use Output Stream	
Dynamic settings	File Name "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/DIM_TIME.csv"	
View	Row Separator "\n"	Field Separator ","
Documentation	<input type="checkbox"/> Append <input checked="" type="checkbox"/> Include Header <input type="checkbox"/> Compress as zip file	
	Schema Built-In	Edit schema Sync columns

tFileOutputDelimited_4

Basic settings	Property Type Built-In	
Advanced settings	<input type="checkbox"/> Use Output Stream	
Dynamic settings	File Name "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/FACT_SALES.csv"	
View	Row Separator "\n"	Field Separator ","
Documentation	<input type="checkbox"/> Append <input checked="" type="checkbox"/> Include Header <input type="checkbox"/> Compress as zip file	
	Schema Built-In	Edit schema Sync columns

Overall Job Components:

The overall job design consists of a sequential flow starting from the tSQLiteInput component, which retrieves the data, followed by the tLogRow component for data verification, and concluding with the tFileOutputDelimited component, which exports the data into separate CSV files.



This job design ensures the smooth extraction, verification, and export of data from the SQLite database, providing a streamlined approach to data processing and analysis within the Talend environment.

Below Images show the output that was shown when the job was run,

Job Project8

Execution: Run, Kill, Clear

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Starting Job Project8 at 17:59 27/10/2023.
[statistics] connecting to socket on port 3523
[statistics] connected

tLogRow_1

geography_id	geography
1	USA
2	Asia
3	Australia
4	Europe

tLogRow_2

product_id	product
1	Laptop
2	Mobile
3	TV
4	Tablet

Job Project8

Execution: Run, Kill, Clear

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

tLogRow_3

time_id	time_quarter
1	Q1
2	Q2
3	Q3
4	Q4

tLogRow_4

product_id	geography_id	time_id	SalesAmt
1	1	1	797
1	1	2	54
1	1	3	886
1	1	4	693
1	2	1	673
1	2	2	74
1	2	3	884
1	2	4	268
1	3	1	28
1	3	2	797

Job Project8

Execution: Run, Kill, Clear

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

3	3	1	124
3	3	2	147
3	3	3	840
3	3	4	91
3	4	1	502
3	4	2	675
3	4	3	942
3	4	4	822
4	1	1	394
4	1	2	796
4	1	3	474
4	1	4	190
4	2	1	629
4	2	2	419
4	2	3	437
4	2	4	322
4	3	1	636
4	3	2	650
4	3	3	717
4	3	4	280
4	4	1	995
4	4	2	796
4	4	3	459
4	4	4	758

[statistics] disconnected

Job Project8 ended at 17:59 27/10/2023. [Exit code = 0]

Line limit 100 Wrap

Job Project8

Execution: Run, Kill, Clear

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

1	3	2	797
1	3	3	139
1	3	4	733
1	4	1	258
1	4	2	561
1	4	3	117
1	4	4	564
2	1	1	231
2	1	2	760
2	1	3	677
2	1	4	707
2	2	1	329
2	2	2	366
2	2	3	26
2	2	4	759
2	3	1	443
2	3	2	117
2	3	3	257
2	3	4	960
2	4	1	289
2	4	2	811
2	4	3	176
2	4	4	728
3	1	1	867
3	1	2	727
3	1	3	958
3	1	4	72
3	2	1	576
3	2	2	830
3	2	3	316

The screenshot shows the Microsoft Excel interface with the Home tab selected. The ribbon includes sections for Clipboard, Font, Alignment, and Number. The active cell is A1, which contains the text 'time_id'. The formula bar shows 'time_id'.

FACT_SALES - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

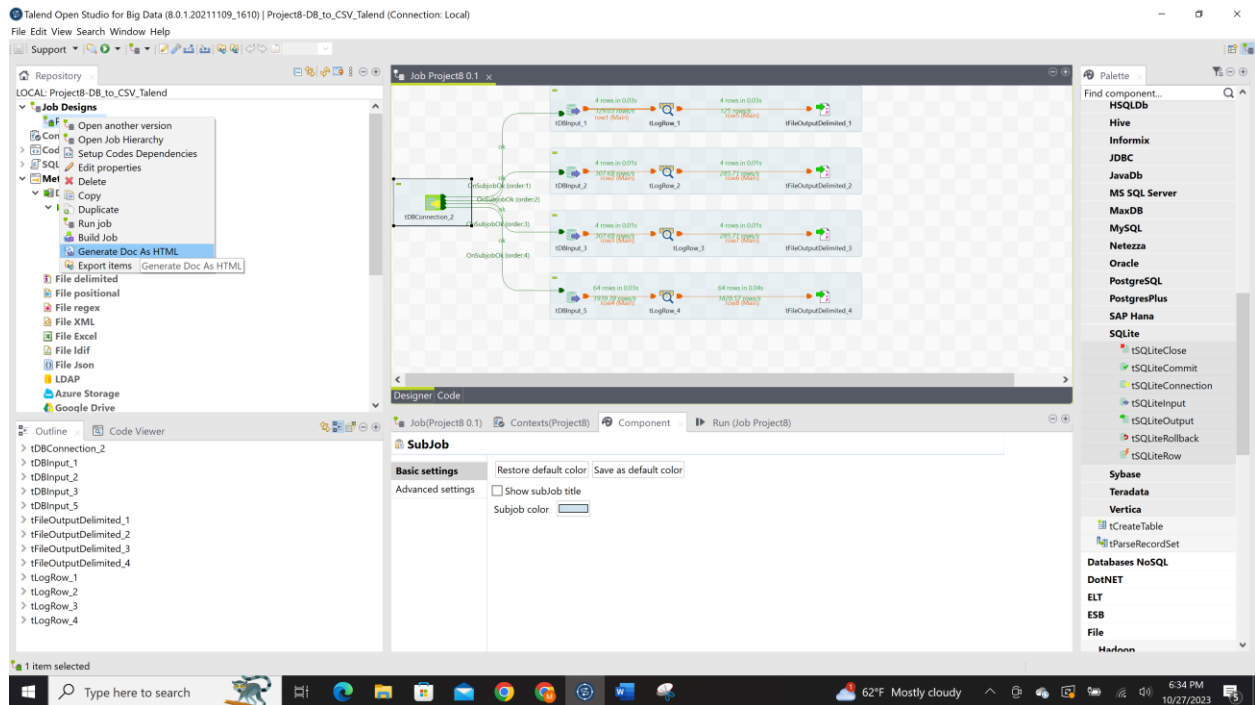
Paste Cut Copy Format Painter Clipboard Font Alignment Number

Calibri 11 A A Wrap Text General \$ % € 0.00 0.00

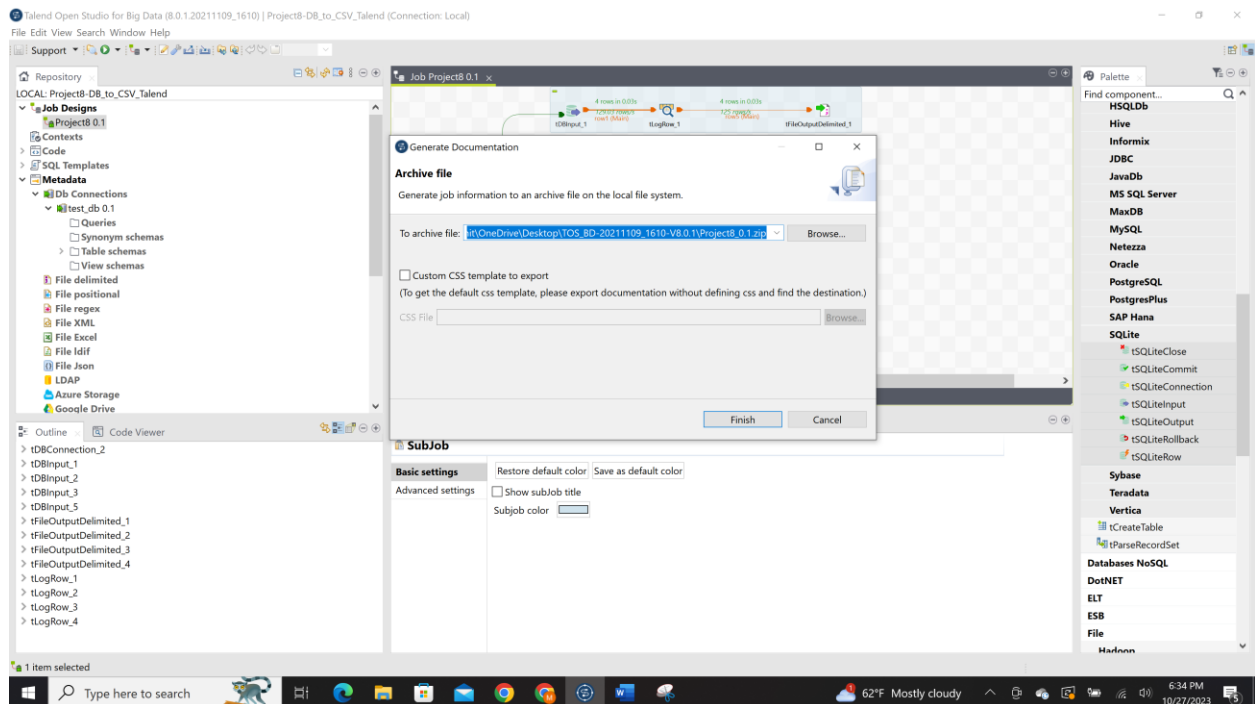
A1 product_id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	product_id	geography	time_id	Salesamt																											
2	1	1	1	797																											
3	1	1	1	54																											
4	1	1	1	886																											
5	1	1	1	693																											
6	1	2	1	973																											
7	1	2	2	74																											
8	1	2	3	884																											
9	1	2	4	388																											
10	1	3	1	28																											
11	1	3	2	797																											
12	1	3	3	139																											
13	1	3	4	733																											
14	1	4	1	358																											
15	1	4	2	561																											
16	1	4	3	117																											
17	1	4	4	564																											
18	2	1	1	231																											
19	2	1	2	760																											
20	2	1	3	677																											
21	2	1	4	707																											
22	2	2	1	329																											
23	2	2	2	366																											
24	2	2	3	26																											
25	2	2	4	759																											
26	2	3	1	443																											
27	2	3	2	117																											

Below images shows the steps to generated doc as html,



This generates an archive in which there is html doc,



After extraction the location of html doc,

Project8

FileHomeShareView

TOS_BD-20211109_1610-V8.0.1Project8_0.1Project8

Name	Date modified	Type	Size
pictures	10/27/2023 9:43 PM	File folder	
default	10/27/2023 9:42 PM	Cascading Style S...	2 KB
Project8_0.1	10/27/2023 9:42 PM	Chrome HTML Do...	213 KB
Project8_0.1	10/27/2023 9:42 PM	XML Document	53 KB

Quick access

DesktopDownloadsDocumentsPicturesFund_De