Pepito, Alyssa Mae & Soleño, Keziah Antonette | Group 23 BM1

CS172

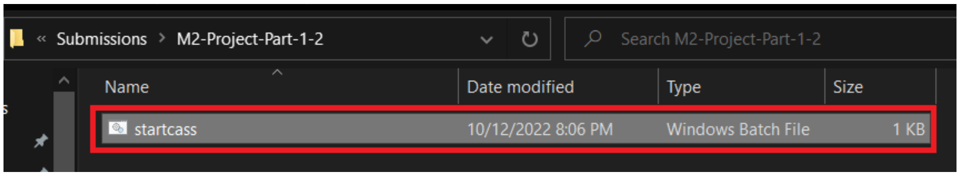
Module 2 Project (Part 1)

I. Initializing Cassandra

Due to the previous "Module 1 Project" activity, we already have Cassandra set up on our local computer. To interact with the Cassandra database, we first must run the Cassandra executable before running the command-line **cqlsh** to interact with Cassandra using CQL. In this project, we automated the process by creating a .bat file that contains the following:

```
start C:\apache-cassandra-3.11.13\bin\cassandra

echo Wait until cassandra is done initializing before continuing

pause

start C:\apache-cassandra-3.11.13\bin\cqlsh
```

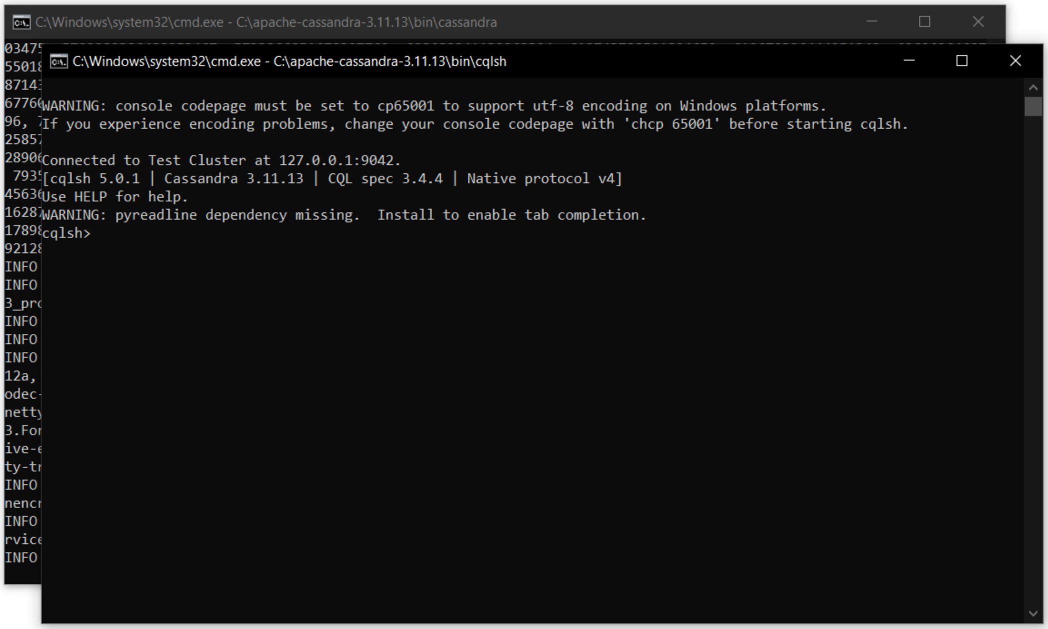
We created this file using a notepad and saved it as **startcass.bat**. Next, we executed this file.



« Submissions › M2-Project-Part-1-2
Search M2-Project-Part-1-2

Name	Date modified	Type	Size
startcass	10/12/2022 8:06 PM	Windows Batch File	1 KB

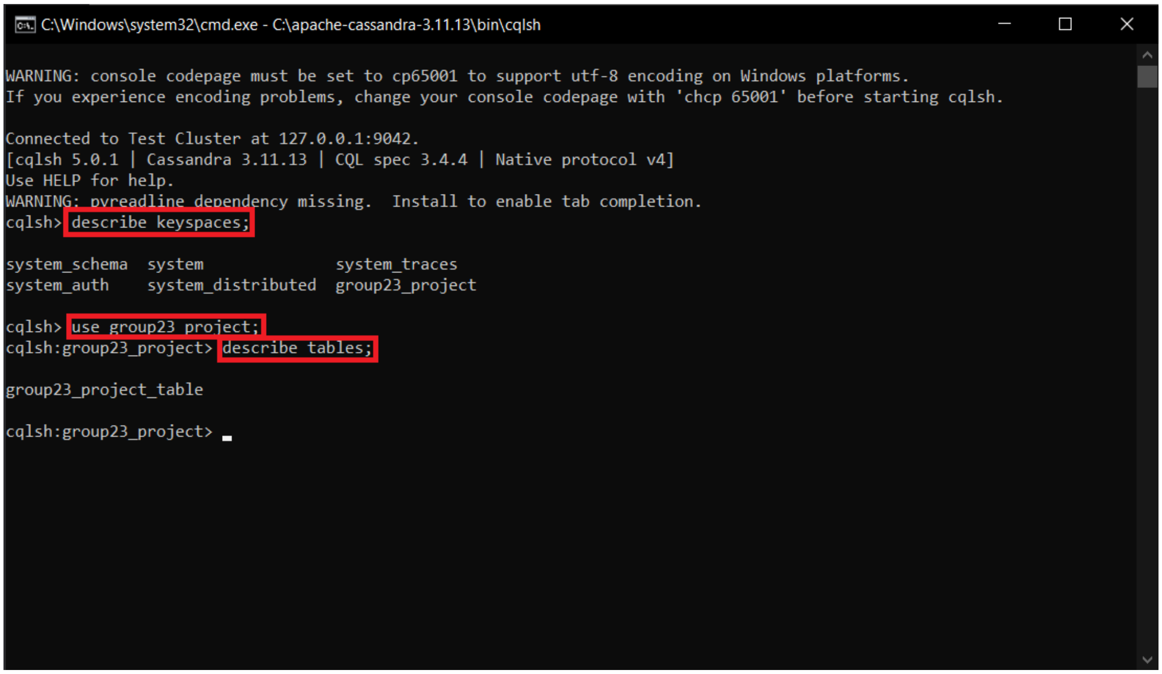
After execution, we had two terminals running in the background: one for the Cassandra database and the other for the CQL shell.



C:\Windows\system32\cmd.exe - C:\apache-cassandra-3.11.13\bin\cassandra
C:\Windows\system32\cmd.exe - C:\apache-cassandra-3.11.13\bin\cqlsh

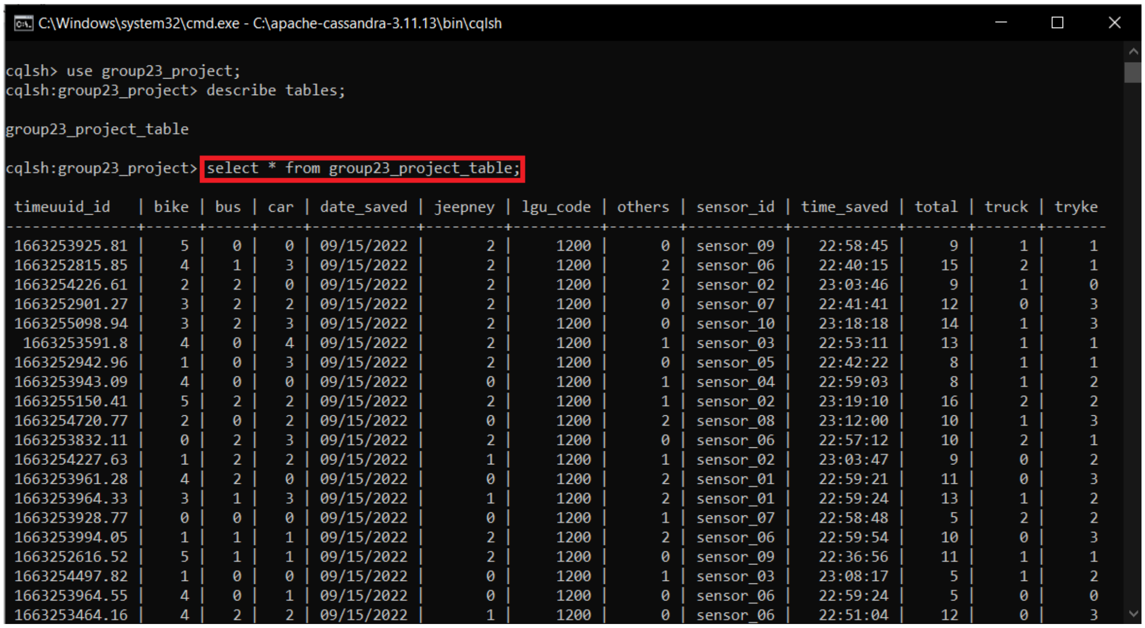
```
WARNING: console codepage must be set to cp65001 to support utf-8 encoding on Windows platforms.
If you experience encoding problems, change your console codepage with 'chcp 65001' before starting cqlsh.

Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.13 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
WARNING: pyreadline dependency missing.  Install to enable tab completion.
cqlsh>
```

In the terminal where the CQL shell was initialized, we ran the following command, "`describe keyspaces;`", to access the keyspace we made previously in the "Module 1 Project" activity. After verifying that our keyspace is still available, we used the command "`use group23_project;`" to use the keyspace. Next, we ran the command "`describe tables;`" to check that the table containing our CCTV counts still exists.


C:\Windows\system32\cmd.exe - C:\apache-cassandra-3.11.13\bin\cqlsh

```
WARNING: console codepage must be set to cp65001 to support utf-8 encoding on Windows platforms.
If you experience encoding problems, change your console codepage with 'chcp 65001' before starting cqlsh.

Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.13 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
WARNING: pyreadline dependency missing.  Install to enable tab completion.
cqlsh> describe keyspaces;

system_schema  system              system_traces
system_auth    system_distributed  group23_project

cqlsh> use group23_project;
cqlsh:group23_project> describe tables;

group23_project_table

cqlsh:group23_project> _
```

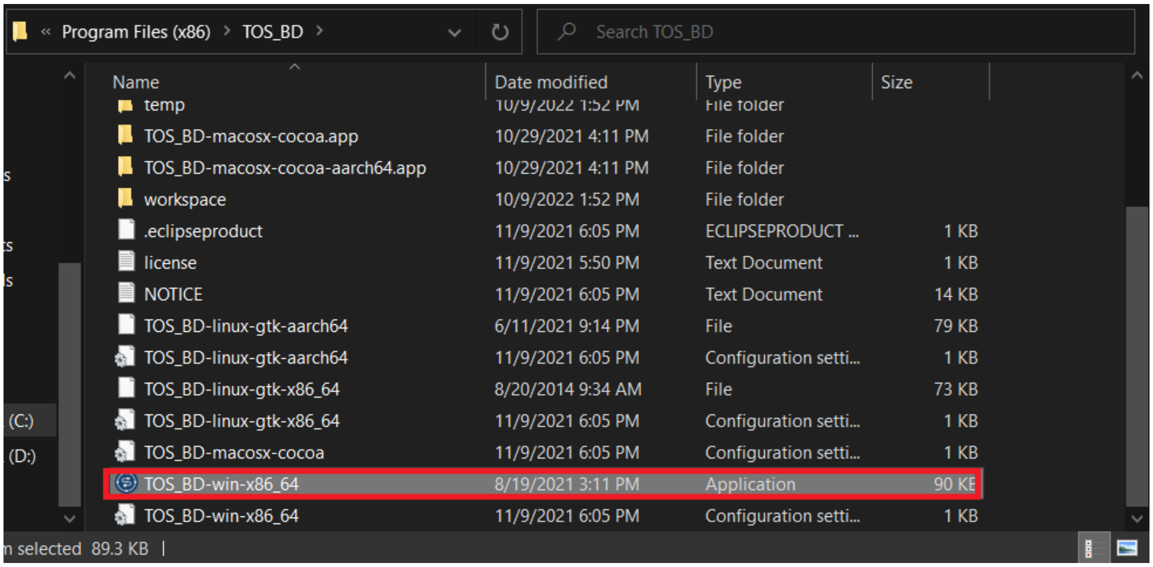
Our table exists however we need to verify if the data inside is complete. By running this command, "`select * from group23_project_table;`", we can see that our data is ready for use in Talend.


C:\Windows\system32\cmd.exe - C:\apache-cassandra-3.11.13\bin\cqlsh

```
cqlsh> use group23_project;
cqlsh:group23_project> describe tables;

group23_project_table

cqlsh:group23_project> select * from group23_project_table;

 timeuuid_id   | bike | bus | car | date_saved | jeepney | lgu_code | others | sensor_id | time_saved | total | truck | tryke
---------------+------+-----+-----+------------+---------+----------+--------+-----------+------------+-------+-------+-------
 1663253925.81 |    5 |   0 |   0 | 09/15/2022 |       2 |     1200 |      0 | sensor_09 |   22:58:45 |     9 |     1 |     1
 1663252815.85 |    4 |   1 |   3 | 09/15/2022 |       2 |     1200 |      2 | sensor_06 |   22:40:15 |    15 |     2 |     1
 1663254226.61 |    2 |   2 |   0 | 09/15/2022 |       2 |     1200 |      2 | sensor_02 |   23:03:46 |     9 |     1 |     0
 1663252901.27 |    3 |   2 |   2 | 09/15/2022 |       2 |     1200 |      0 | sensor_07 |   22:41:41 |    12 |     0 |     3
 1663255098.94 |    3 |   2 |   3 | 09/15/2022 |       2 |     1200 |      0 | sensor_10 |   23:18:18 |    14 |     1 |     3
  1663253591.8 |    4 |   0 |   4 | 09/15/2022 |       2 |     1200 |      1 | sensor_03 |   22:53:11 |    13 |     1 |     1
 1663252942.96 |    1 |   0 |   3 | 09/15/2022 |       2 |     1200 |      0 | sensor_05 |   22:42:22 |     8 |     1 |     1
 1663253943.09 |    4 |   0 |   0 | 09/15/2022 |       0 |     1200 |      1 | sensor_04 |   22:59:03 |     8 |     1 |     2
 1663255150.41 |    5 |   2 |   2 | 09/15/2022 |       2 |     1200 |      1 | sensor_02 |   23:19:10 |    16 |     2 |     2
 1663254720.77 |    2 |   0 |   2 | 09/15/2022 |       0 |     1200 |      2 | sensor_08 |   23:12:00 |    10 |     1 |     3
 1663253832.11 |    0 |   2 |   3 | 09/15/2022 |       2 |     1200 |      0 | sensor_06 |   22:57:12 |    10 |     2 |     1
 1663254227.63 |    1 |   2 |   2 | 09/15/2022 |       1 |     1200 |      1 | sensor_02 |   23:03:47 |     9 |     0 |     2
 1663253961.28 |    4 |   2 |   0 | 09/15/2022 |       0 |     1200 |      2 | sensor_01 |   22:59:21 |    11 |     0 |     3
 1663253964.33 |    3 |   1 |   3 | 09/15/2022 |       1 |     1200 |      2 | sensor_01 |   22:59:24 |    13 |     1 |     2
 1663253928.77 |    0 |   0 |   0 | 09/15/2022 |       0 |     1200 |      1 | sensor_07 |   22:58:48 |     5 |     2 |     2
 1663253994.05 |    1 |   1 |   1 | 09/15/2022 |       2 |     1200 |      2 | sensor_06 |   22:59:54 |    10 |     0 |     3
 1663252616.52 |    5 |   1 |   1 | 09/15/2022 |       2 |     1200 |      0 | sensor_09 |   22:36:56 |    11 |     1 |     1
 1663254497.82 |    1 |   0 |   0 | 09/15/2022 |       0 |     1200 |      1 | sensor_03 |   23:08:17 |     5 |     1 |     2
 1663253964.55 |    4 |   0 |   1 | 09/15/2022 |       0 |     1200 |      0 | sensor_06 |   22:59:24 |     5 |     0 |     0
 1663253464.16 |    4 |   2 |   2 | 09/15/2022 |       1 |     1200 |      0 | sensor_06 |   22:51:04 |    12 |     0 |     3
```

II. Opening Talend for Big Data

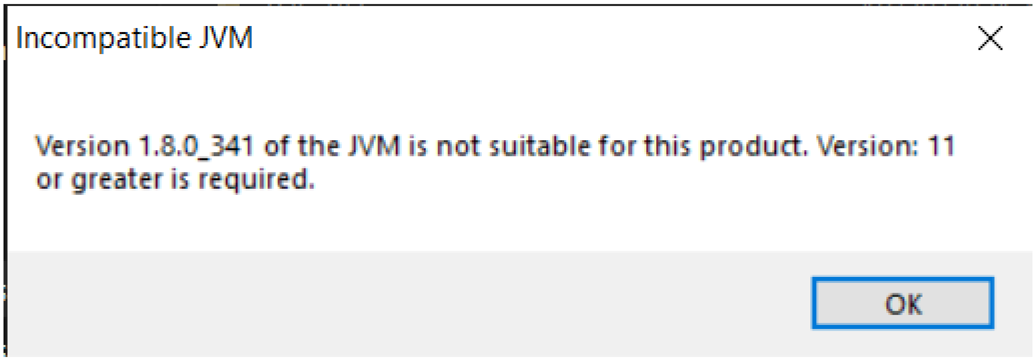
Talend was also set up on our computer due to the recent “Module 2 Assignment 2 and 3” activities. So, we just need to go to the directory folder where our Talend for Big Data is located to open it.



Name	Date modified	Type	Size
temp	10/9/2022 1:52 PM	File folder	
TOS_BD-macosx-cocoa.app	10/29/2021 4:11 PM	File folder	
TOS_BD-macosx-cocoa-aarch64.app	10/29/2021 4:11 PM	File folder	
workspace	10/9/2022 1:52 PM	File folder	
.eclipseproduct	11/9/2021 6:05 PM	ECLIPSEPRODUCT ...	1 KB
license	11/9/2021 5:50 PM	Text Document	1 KB
NOTICE	11/9/2021 6:05 PM	Text Document	14 KB
TOS_BD-linux-gtk-aarch64	6/11/2021 9:14 PM	File	79 KB
TOS_BD-linux-gtk-aarch64	11/9/2021 6:05 PM	Configuration setti...	1 KB
TOS_BD-linux-gtk-x86_64	8/20/2014 9:34 AM	File	73 KB
TOS_BD-linux-gtk-x86_64	11/9/2021 6:05 PM	Configuration setti...	1 KB
TOS_BD-macosx-cocoa	11/9/2021 6:05 PM	Configuration setti...	1 KB
TOS_BD-win-x86_64	8/19/2021 3:11 PM	Application	90 KE
TOS_BD-win-x86_64	11/9/2021 6:05 PM	Configuration setti...	1 KB

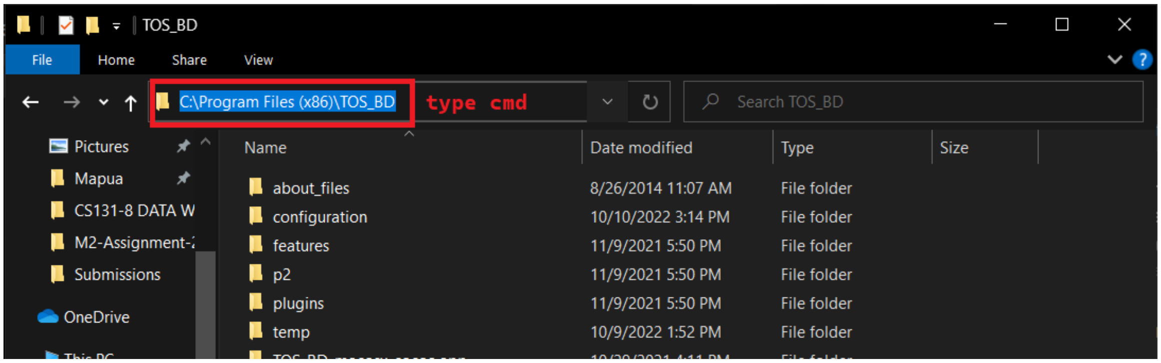
n selected 89.3 KB |

Note: If you are still encountering problems such as “Incompatible JVM” then you may need to update the environment variables of your system. Otherwise, proceed here.



Version 1.8.0_341 of the JVM is not suitable for this product. Version: 11 or greater is required.

However, you may also opt to do the following: Type “cmd” in the address bar of your Talend directory folder.

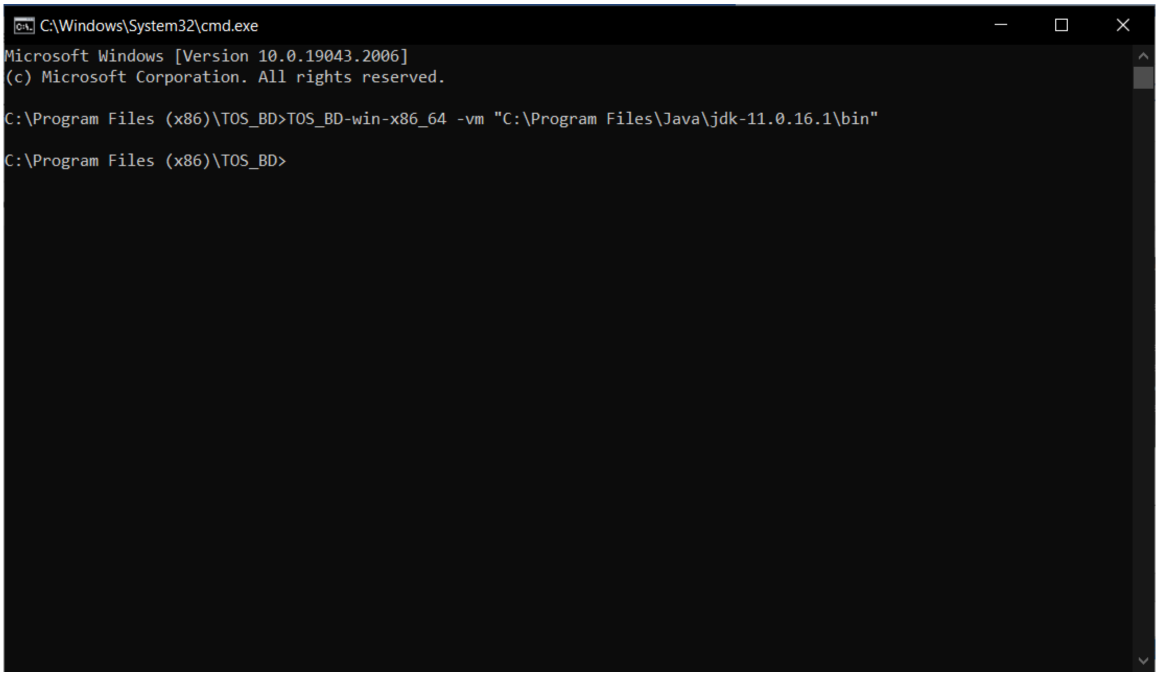


C:\Program Files (x86)\TOS_BD

type cmd

Name	Date modified	Type	Size
about_files	8/26/2014 11:07 AM	File folder	
configuration	10/10/2022 3:14 PM	File folder	
features	11/9/2021 5:50 PM	File folder	
p2	11/9/2021 5:50 PM	File folder	
plugins	11/9/2021 5:50 PM	File folder	
temp	10/9/2022 1:52 PM	File folder	
TOS_BD-macosx-cocoa.app	10/29/2021 4:11 PM	File folder	

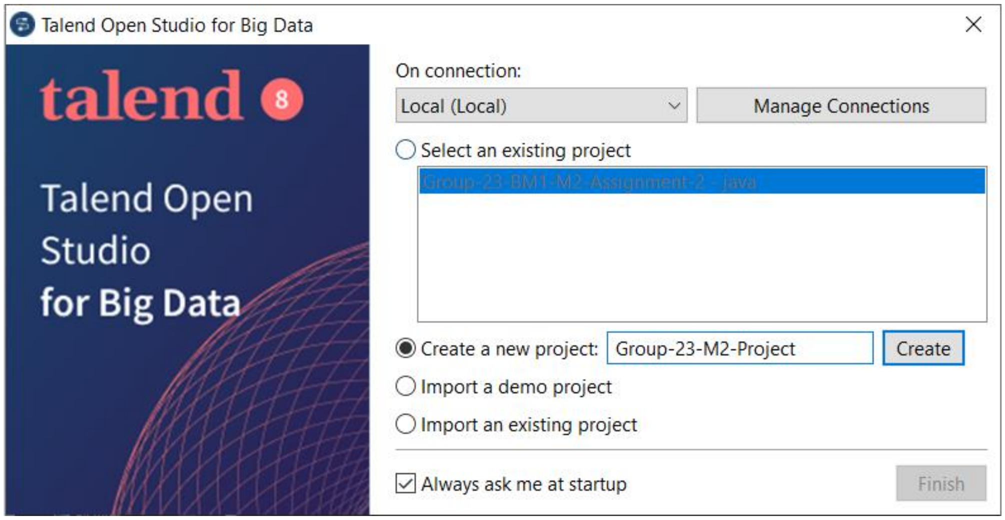
In the command prompt, we typed “`TOS_BD-win-x86_64 -vm "C:\Program Files\Java\jdk-11.0.16.1\bin"`”. Please take note that your Java folder may be different from this example.



C:\Windows\System32\cmd.exe

```
Microsoft Windows [Version 10.0.19043.2006]
(c) Microsoft Corporation. All rights reserved.

C:\Program Files (x86)\TOS_BD>TOS_BD-win-x86_64 -vm "C:\Program Files\Java\jdk-11.0.16.1\bin"

C:\Program Files (x86)\TOS_BD>
```

Once your Talend for Big Data is open, proceed to creating a new project. In this instance, we named our project “Group-23-M2-Project”. Then proceed to select the newly made project and open it in Talend.



Talend Open Studio for Big Data

On connection: Local (Local)

○ Select an existing project

Group-23-BM1-M2-Assignment-2 - java

● Create a new project: Group-23-M2-Project

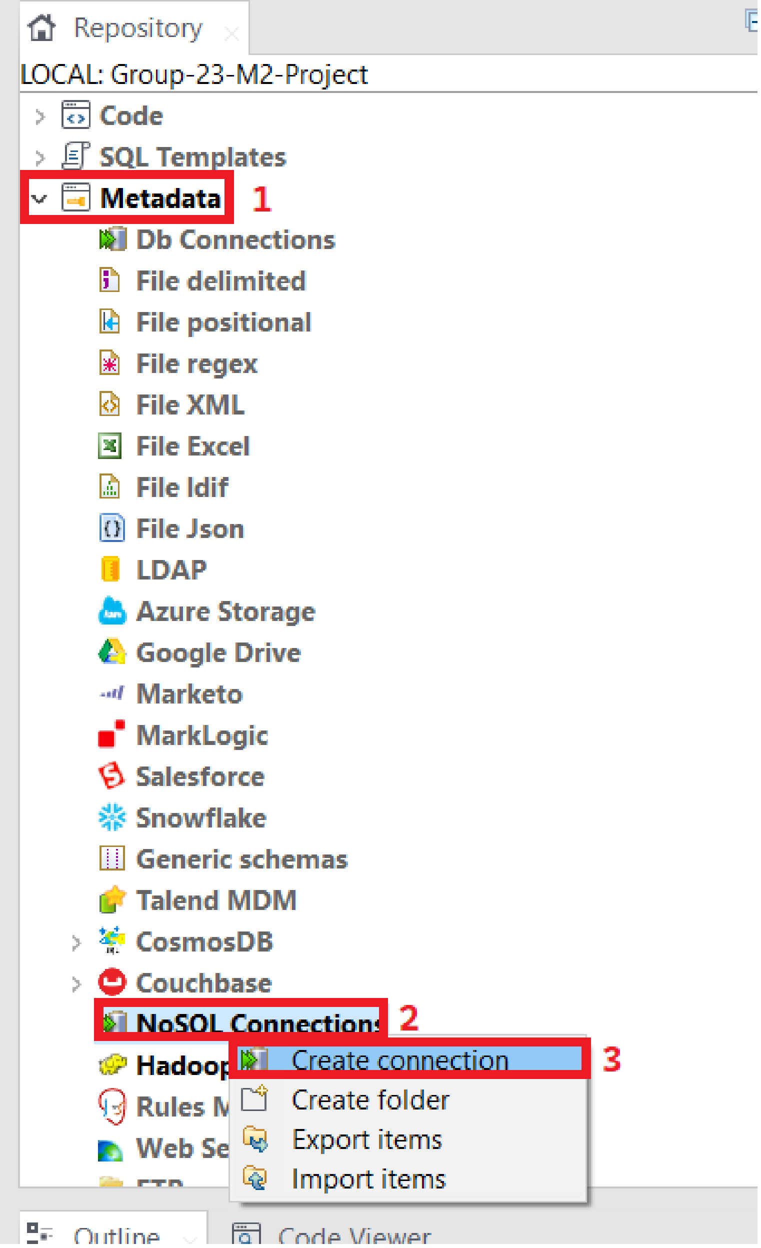
○ Import a demo project

○ Import an existing project

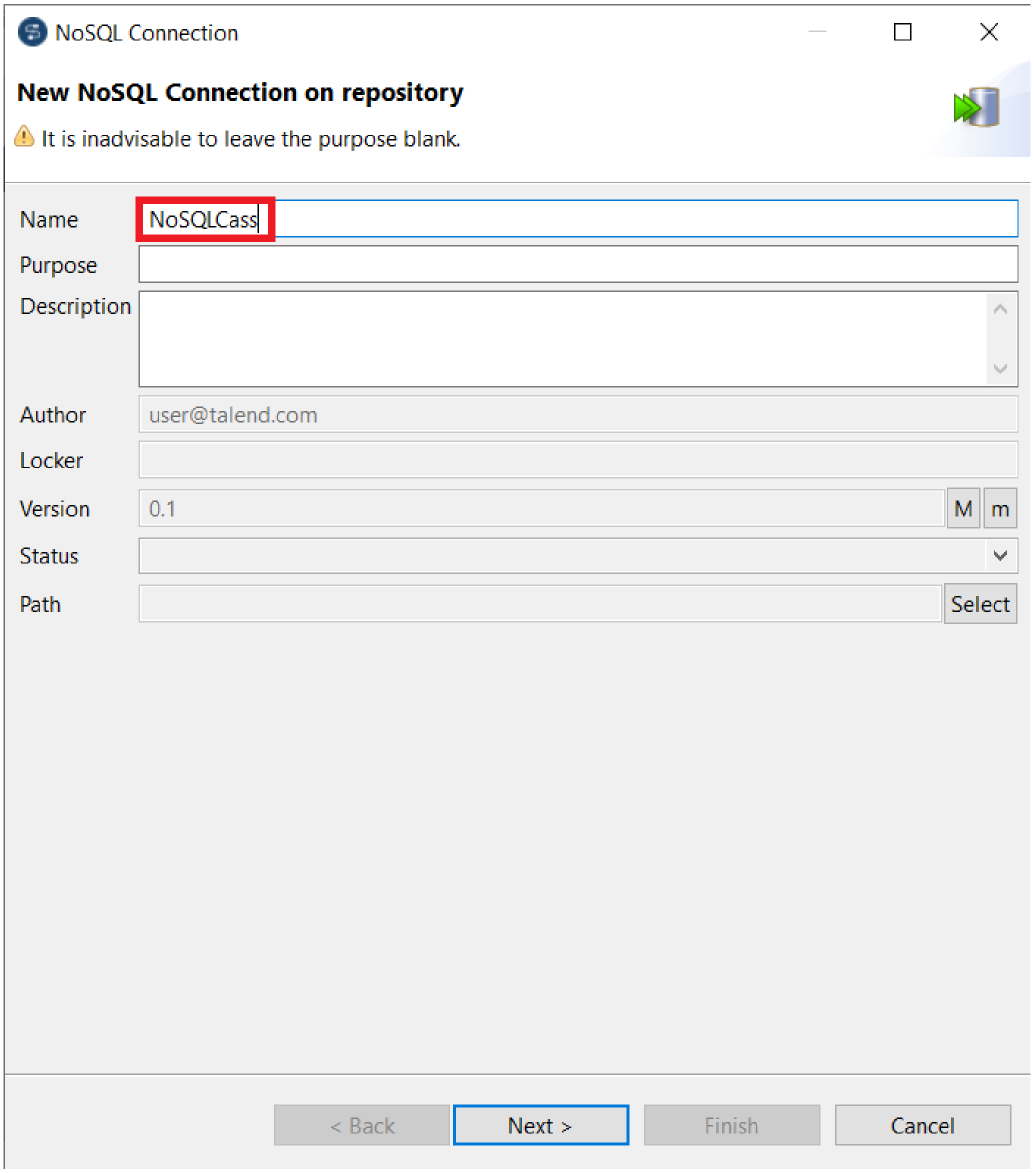
☑ Always ask me at startup

III. Talend Cassandra NoSQL Connection and Job

After creating a new project, the Talend window should now appear. To establish the connection between the Cassandra database and Talend, select *Metadata*. Then, right click *NoSQL Connections > Create connection*.



The new NoSQL Connection window should appear, where the user will be prompted to specify a name, purpose, and description for the new NoSQL connection. Once all required fields are populated, click *Next >*. In this assignment, we named the connection as *NoSQLCass* and left the remaining text field and area blank.



NoSQL Connection

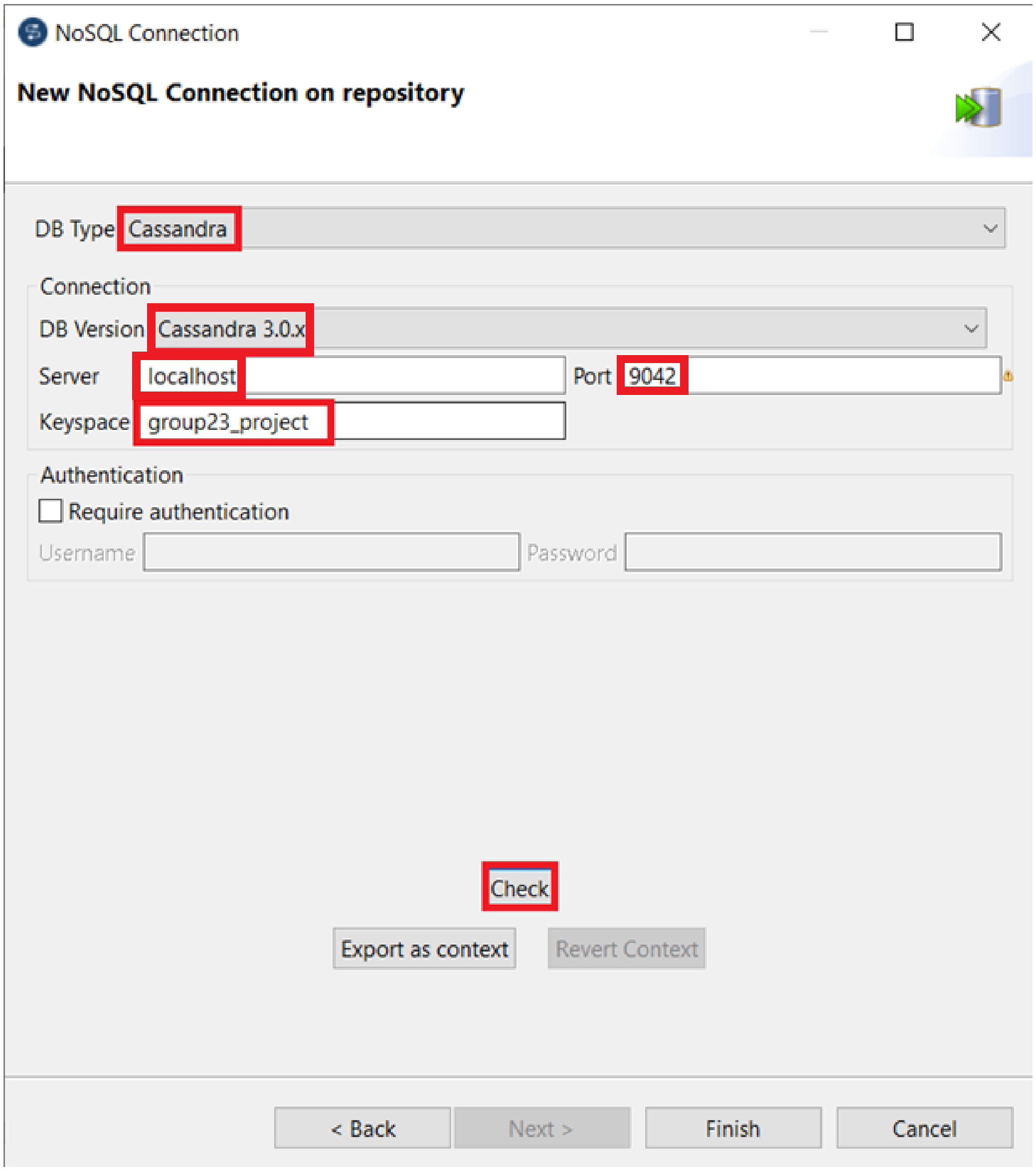
New NoSQL Connection on repository

⚠ It is inadvisable to leave the purpose blank.

Name	NoSQLCass
Purpose	
Description	
Author	user@talend.com
Locker	
Version	0.1
Status	
Path	

After specifying the name (and purpose and description) of the connection, select the database (DB Type) that you wish to establish a Talend connection with. Since one requirement of this assignment involves Cassandra, select *Cassandra* as the **DB Type** and *Cassandra 3.0.x* as the **DB Version**. Then, specify *localhost* or *127.0.0.1* as the **Server** with *9042* as its **Port**.

From a previous assignment, we created a keyspace called *group23_project*, where we stored the CCTV_counts table (*group23_project_table*). Since the objective of this project is to export that table into a flat file, we specified the **Keyspace** *group23_project* in this connection. Select "Check" to verify all fields.



NoSQL Connection

New NoSQL Connection on repository

DB Type: Cassandra

Connection

DB Version: Cassandra 3.0.x

Server: localhost Port: 9042

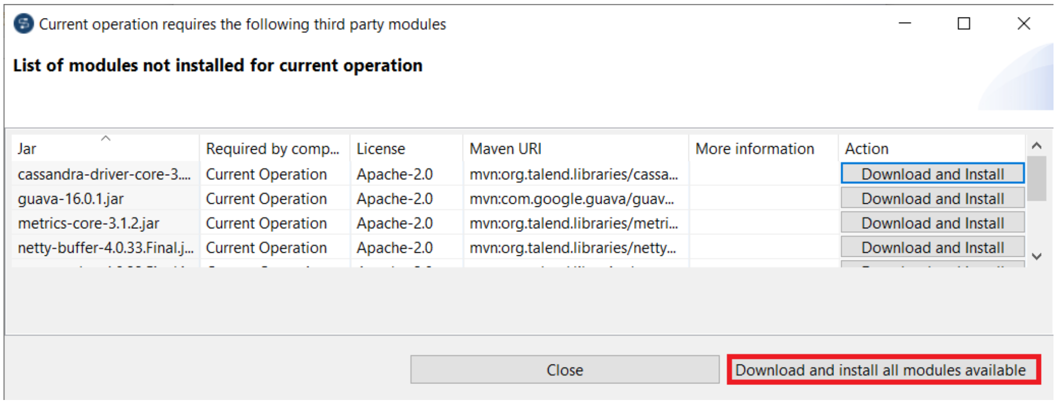
Keyspace: group23_project

Authentication

□ Require authentication

Username: Password:

After clicking “Check” to verify the contents of the fields, the following window will appear to prompt you to download and install the required modules. Click *Download and install all modules available*.

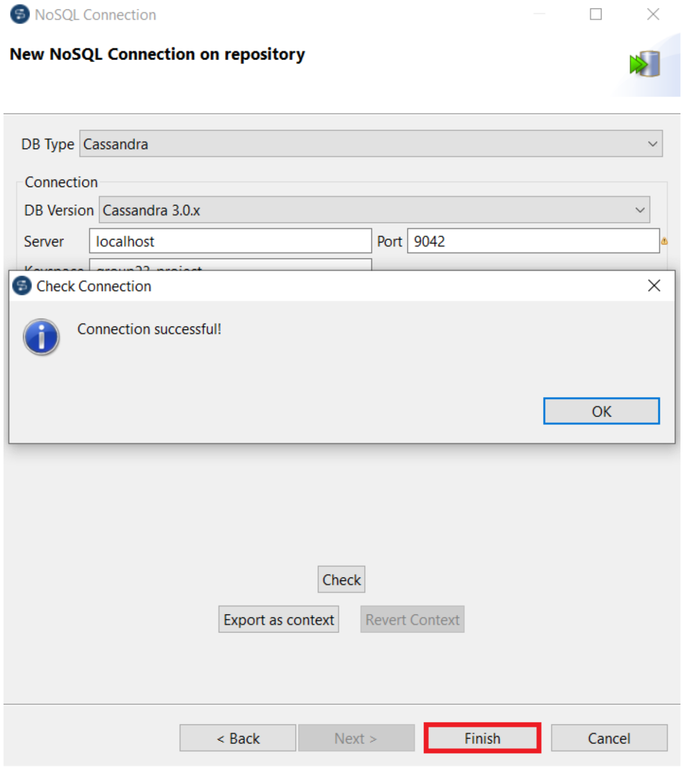


Current operation requires the following third party modules

List of modules not installed for current operation

Jar	Required by comp...	License	Maven URI	More information	Action
cassandra-driver-core-3....	Current Operation	Apache-2.0	mvn:org.talend.libraries/cassa...		Download and Install
guava-16.0.1.jar	Current Operation	Apache-2.0	mvn:com.google.guava/guav...		Download and Install
metrics-core-3.1.2.jar	Current Operation	Apache-2.0	mvn:org.talend.libraries/metri...		Download and Install
netty-buffer-4.0.33.Final.j...	Current Operation	Apache-2.0	mvn:org.talend.libraries/netty...		Download and Install

You will be returned to the previous window. The following dialog box should appear to indicate that Connection creation was successful. Click *OK* to close this dialog box. Finally, click *Finish.*



NoSQL Connection

New NoSQL Connection on repository

DB Type: Cassandra

Connection

DB Version: Cassandra 3.0.x

Server: localhost

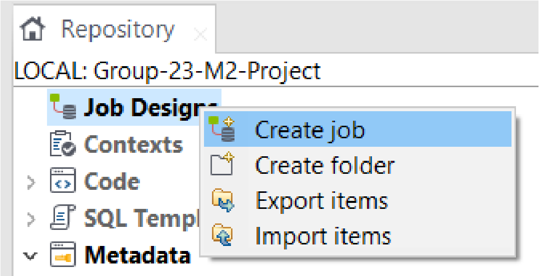
Port: 9042

Check Connection

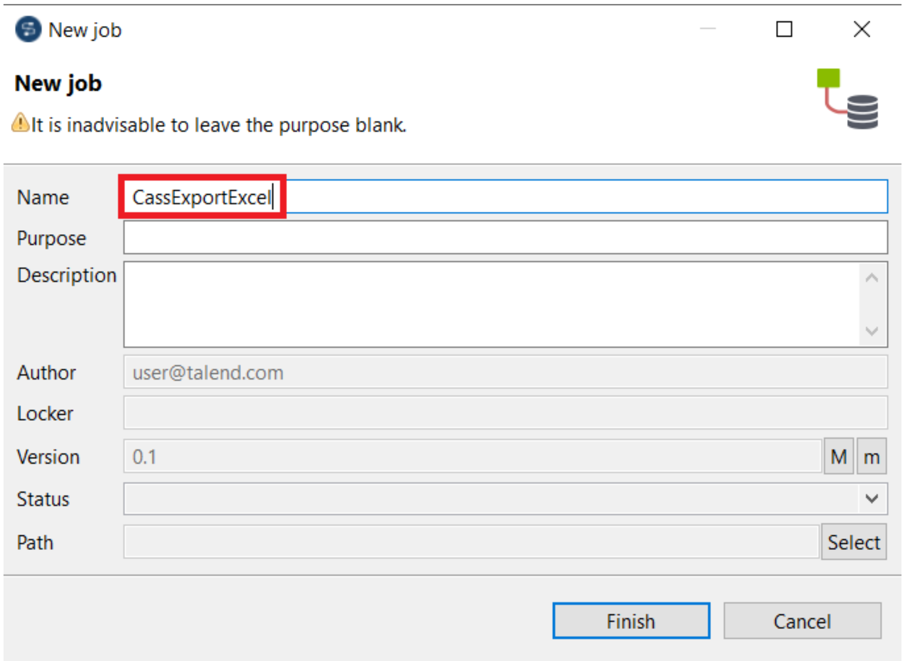
Connection successful!

IV. Exporting CCTV_Counts from Cassandra to an Excel File

To begin exporting the Cassandra table into a flat file, select *Create job* after right-clicking on **Job Designs** under the Repository.



Then, name the new job as *CassExportExcel*.

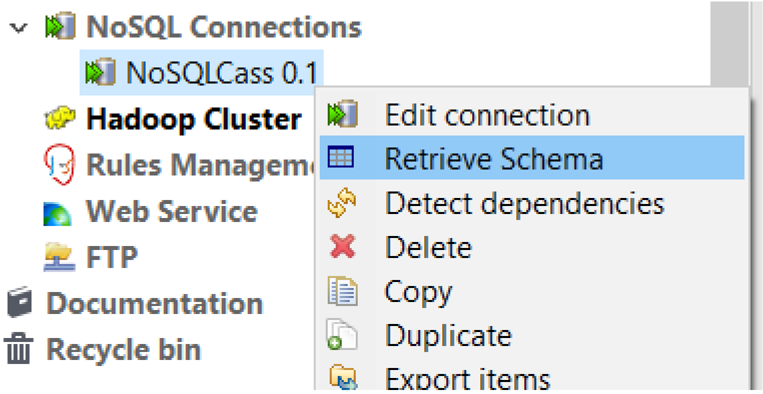


New job

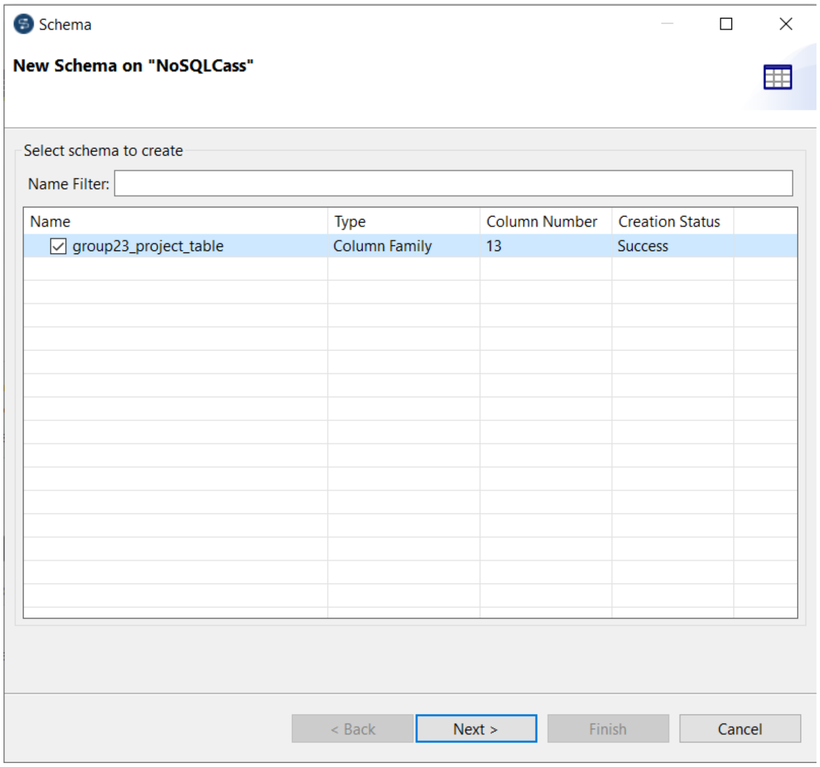
⚠It is inadvisable to leave the purpose blank.

Name	CassExportExcel
Purpose	
Description	
Author	user@talend.com
Locker	
Version	0.1
Status	
Path	

Next, go to the Repository tab and under *NoSQL Connections*, right click on the newly created NoSql Connection to Cassandra and click on “Retrieve Schema”.



In the Schema window prompt, check mark the box beside the name of the table that exists in your Cassandra database. Click on “Next >” to proceed.



Schema

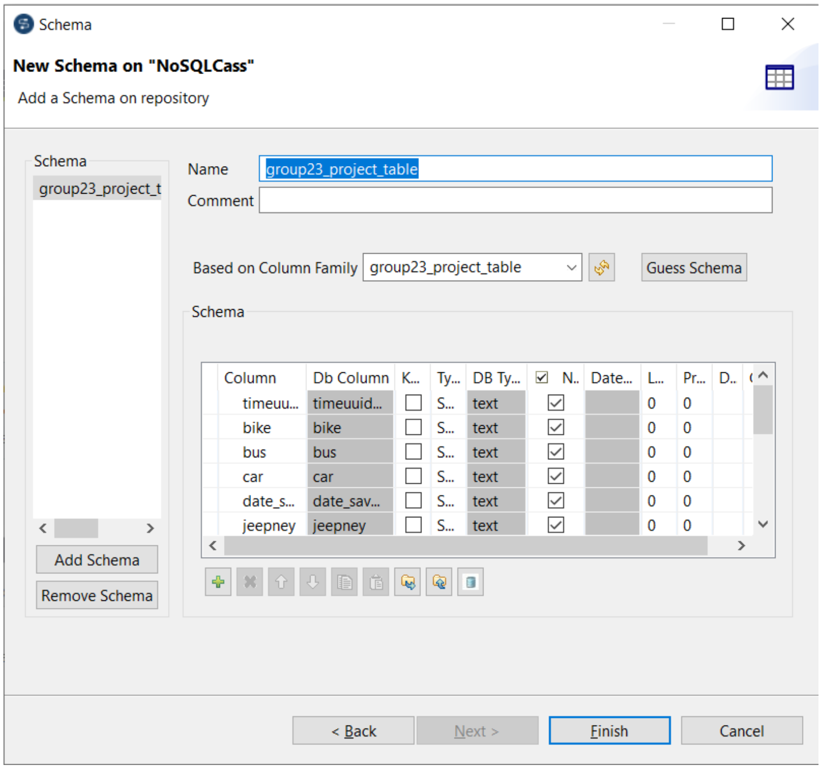
New Schema on "NoSQLCass"

Select schema to create

Name Filter:

Name	Type	Column Number	Creation Status
☑ group23_project_table	Column Family	13	Success

There is nothing to change here so click on “Finish”.



Schema

New Schema on "NoSQLCass"

Add a Schema on repository

Schema: group23_project_t

Name: group23_project_table

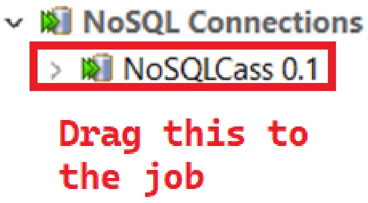
Comment:

Based on Column Family: group23_project_table

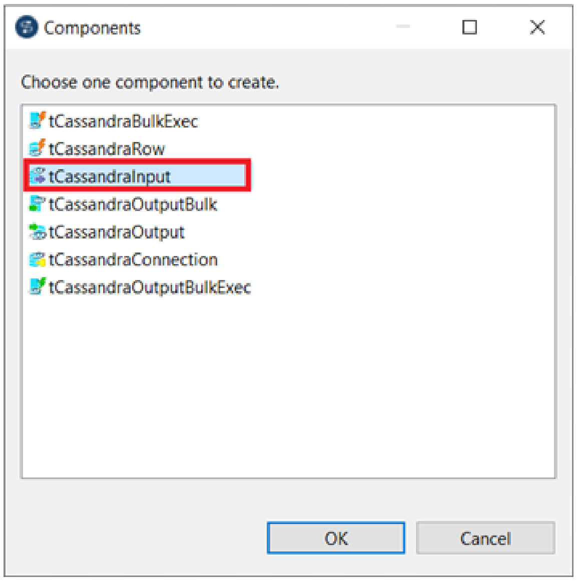
Schema

Column	Db Column	K...	Ty...	DB Ty...	☑ N..	Date...	L...	Pr...	D..
timeuu...	timeuuid...	☐	S...	text	☑		0	0	
bike	bike	☐	S...	text	☑		0	0	
bus	bus	☐	S...	text	☑		0	0	
car	car	☐	S...	text	☑		0	0	
date_s...	date_sav...	☐	S...	text	☑		0	0	
jeepney	jeepney	☐	S...	text	☑		0	0	

Drag the NoSQL Connection (*NoSQLCass*) into the job and choose *tCassandraInput*. Click "OK" for it to be placed.

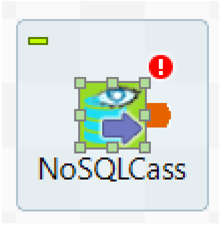


Drag this to the job

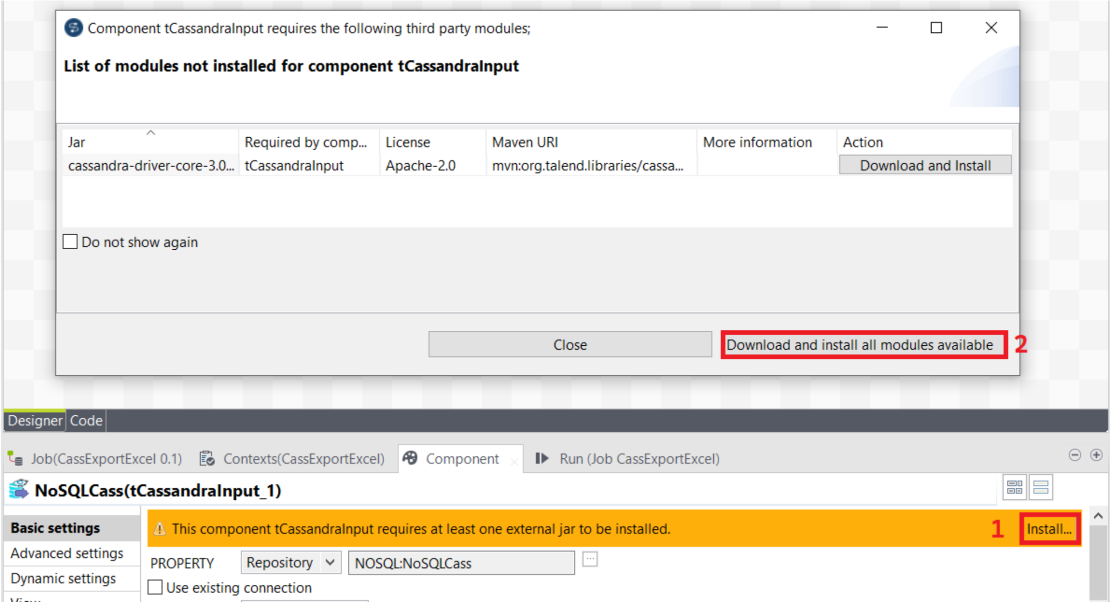


Choose one component to create.

The newly made *tCassandraInput* component should look like this:



Double click the tCassandraInput. Since the **tCassandraInput** subjob requires an external .jar file to be installed (1), select *Download and install all modules available* (2) to do so.



Component tCassandraInput requires the following third party modules;

List of modules not installed for component tCassandraInput

Jar	Required by comp...	License	Maven URI	More information	Action
cassandra-driver-core-3.0...	tCassandraInput	Apache-2.0	mvn:org.talend.libraries/cassa...		Download and Install

□ Do not show again

Close
Download and install all modules available
2
Designer
Code
Job(CassExportExcel 0.1)
Contexts(CassExportExcel)
Component
Run (Job CassExportExcel)
NoSQLCass(tCassandraInput_1)
Basic settings
Advanced settings
Dynamic settings
1
Install...

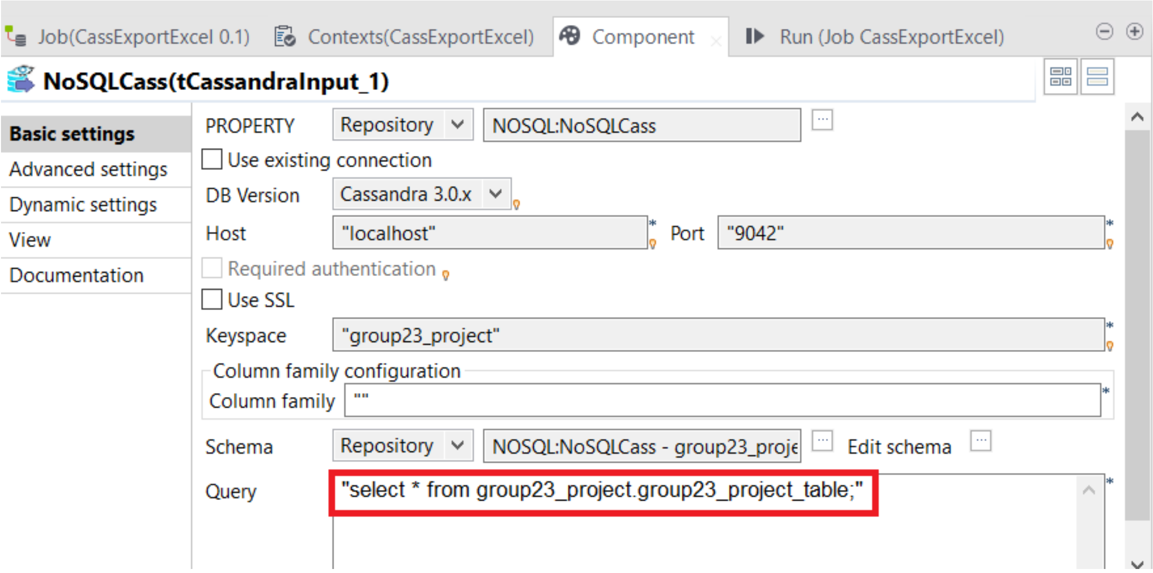
This component tCassandraInput requires at least one external jar to be installed.

PROPERTY: Repository | NOSQL:NoSQLCass

□ Use existing connection

Now, insert the following **Query** on the appropriate text area to show the group23_project_table (CCTV_counts table) and to prepare it for exporting:

```
“select * from group23_project.group23_project_table;”
```



PROPERTY: Repository | NOSQL:NoSQLCass

□ Use existing connection

DB Version: Cassandra 3.0.x

Host: "localhost" Port: "9042"

□ Required authentication

□ Use SSL

Keyspace: "group23_project"

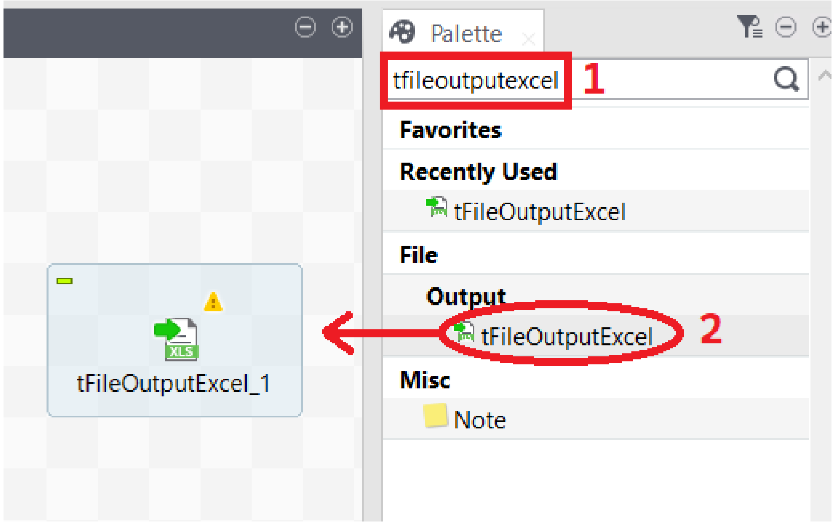
Column family configuration

Column family: ""

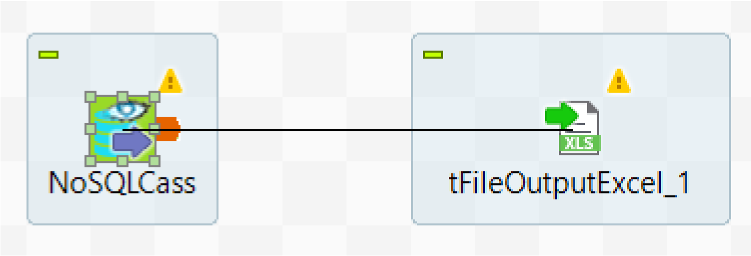
Schema: Repository | NOSQL:NoSQLCass - group23_proje | Edit schema

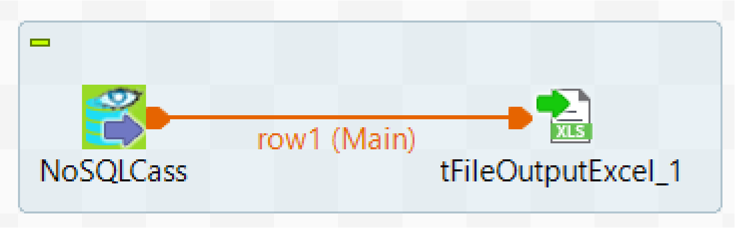
Query: "select * from group23_project.group23_project_table;"

Then, in the Palette tab to your right, search for **tFileOutputExcel**, then select and drag it. This component will perform the flat file export of the Cassandra table. Drag the job to the workspace, as indicated.

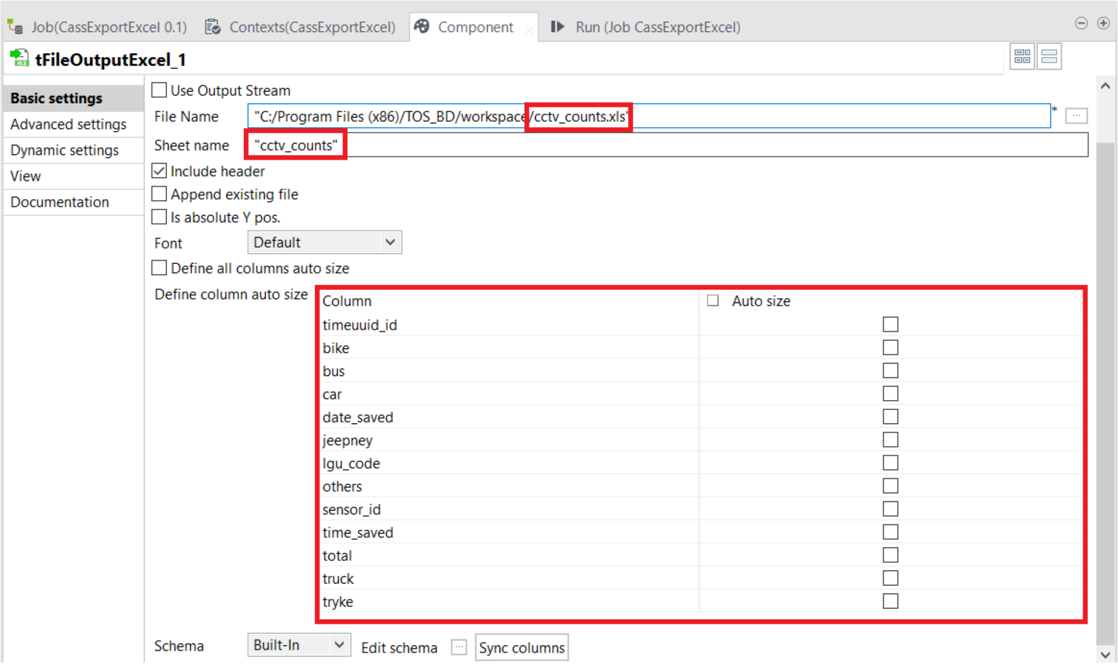


Now, drag an arrow from the *NoSQLCass* component that was created earlier through the *tCassandraInput* component to the *tFileOutputExcel* job.





Then, specify the directory where the exported .xls file will be saved. In this assignment, we saved the file in the directory *C:/Program Files (x86)/TOS_BD/workspace/cctv_counts.xls* with a sheet name of *cctv_counts*. Each column from the Cassandra table will be automatically detected by Talend, as indicated in the *Define column auto size* section.



tFileOutputExcel_1

□ Use Output Stream

File Name "C:/Program Files (x86)/TOS_BD/workspace/cctv_counts.xls"

Sheet name "cctv_counts"

☑ Include header

□ Append existing file

□ Is absolute Y pos.

Font Default

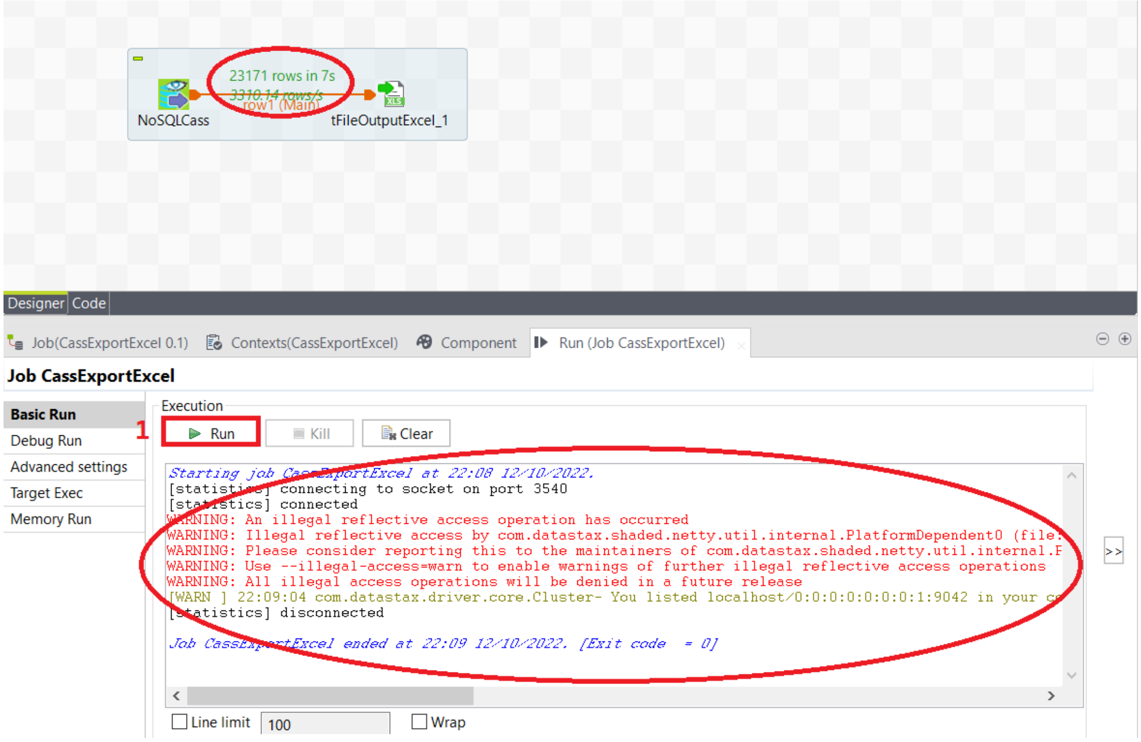
□ Define all columns auto size

Define column auto size

Column	□ Auto size
timeuuid_id	□
bike	□
bus	□
car	□
date_saved	□
jeepney	□
lgu_code	□
others	□
sensor_id	□
time_saved	□
total	□
truck	□
tryke	□

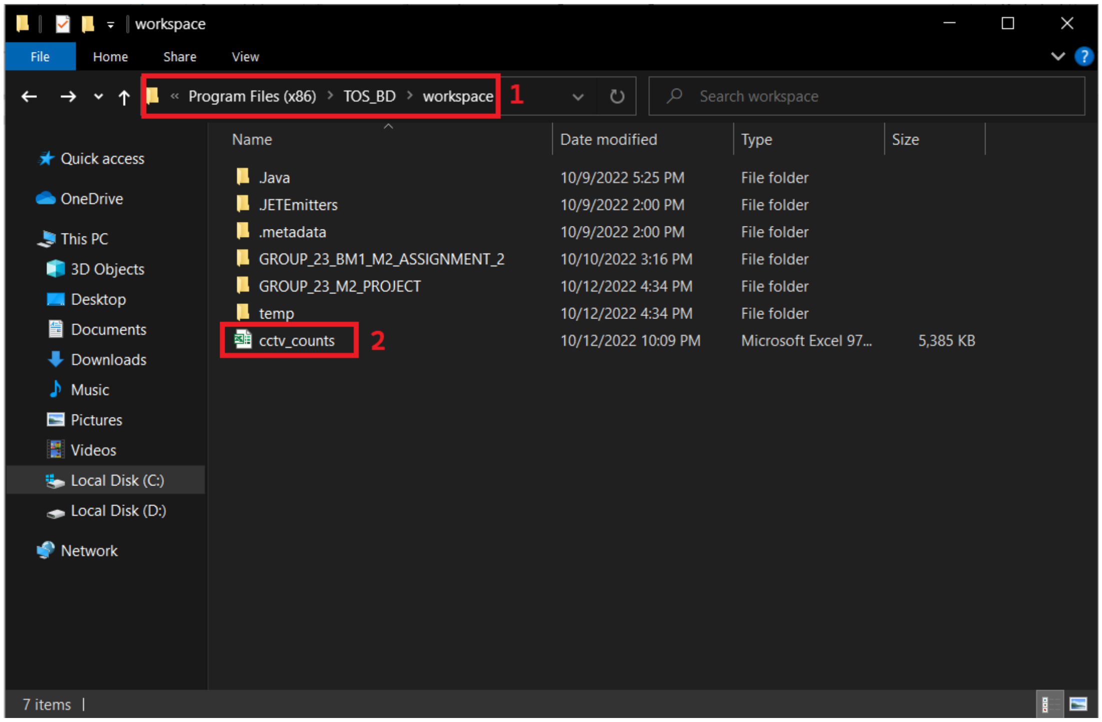
Schema Built-In Edit schema Sync columns

To run the execution, go to the *Run (Job CassExportExcel)* tab and select the *Run* command. The encircled portion of the screenshot should appear, indicating that the exporting is taking place.



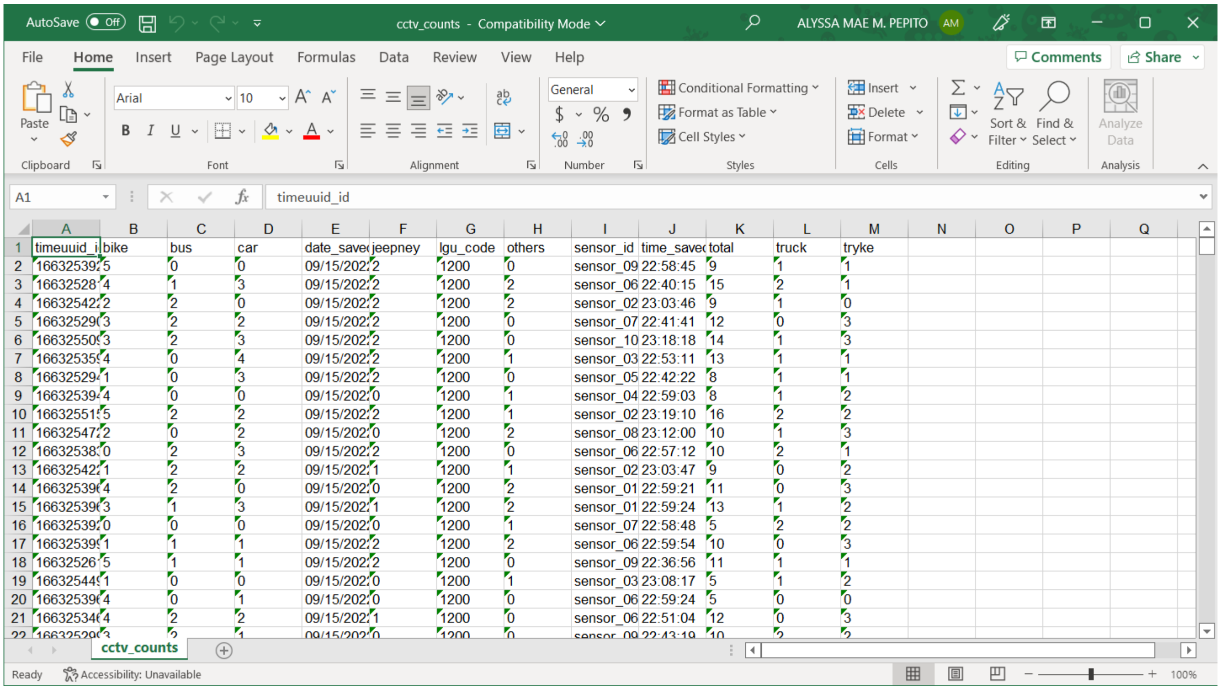
```
Starting job CassExportExcel at 22:08 12/10/2022.
[statistics] connecting to socket on port 3540
[statistics] connected
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by com.datastax.shaded.netty.util.internal.PlatformDependent0 (file:
WARNING: Please consider reporting this to the maintainers of com.datastax.shaded.netty.util.internal.P
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
[WARN ] 22:09:04 com.datastax.driver.core.Cluster- You listed localhost/0:0:0:0:0:0:0:1:9042 in your c
[statistics] disconnected

Job CassExportExcel ended at 22:09 12/10/2022. [Exit code  = 0]
```

Once the execution has finished, go to the specified directory that made earlier and look for a “cctv_counts.” The directory should contain the .xls file, as indicated.



Name	Date modified	Type	Size
Java	10/9/2022 5:25 PM	File folder	
JETEmitters	10/9/2022 2:00 PM	File folder	
.metadata	10/9/2022 2:00 PM	File folder	
GROUP_23_BM1_M2_ASSIGNMENT_2	10/10/2022 3:16 PM	File folder	
GROUP_23_M2_PROJECT	10/12/2022 4:34 PM	File folder	
temp	10/12/2022 4:34 PM	File folder	
cctv_counts	10/12/2022 10:09 PM	Microsoft Excel 97...	5,385 KB

Select the file. The contents of the flat file should be like the screenshot below.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	timeuuid_i	bike	bus	car	date_save	jeepney	lgu_code	others	sensor_id	time_save	total	truck	tryke
2	16632539	5	0	0	09/15/202	2	1200	0	sensor_09	22:58:45	9	1	1
3	16632528	4	1	3	09/15/202	2	1200	2	sensor_06	22:40:15	15	2	1
4	16632542	2	2	0	09/15/202	2	1200	2	sensor_02	23:03:46	9	1	0
5	16632529	3	2	2	09/15/202	2	1200	0	sensor_07	22:41:41	12	0	3
6	16632550	3	2	3	09/15/202	2	1200	0	sensor_10	23:18:18	14	1	3
7	16632535	4	0	4	09/15/202	2	1200	1	sensor_03	22:53:11	13	1	1
8	16632529	1	0	3	09/15/202	2	1200	0	sensor_05	22:42:22	8	1	1
9	16632539	4	0	0	09/15/202	0	1200	1	sensor_04	22:59:03	8	1	2
10	16632551	5	2	2	09/15/202	2	1200	1	sensor_02	23:19:10	16	2	2
11	16632547	2	0	2	09/15/202	0	1200	2	sensor_08	23:12:00	10	1	3
12	16632538	0	2	3	09/15/202	2	1200	0	sensor_06	22:57:12	10	2	1
13	16632542	1	2	2	09/15/202	1	1200	1	sensor_02	23:03:47	9	0	2
14	16632539	4	2	0	09/15/202	0	1200	2	sensor_01	22:59:21	11	0	3
15	16632539	3	1	3	09/15/202	1	1200	2	sensor_01	22:59:24	13	1	2
16	16632539	0	0	0	09/15/202	0	1200	1	sensor_07	22:58:48	5	2	2
17	16632539	1	1	1	09/15/202	2	1200	2	sensor_06	22:59:54	10	0	3
18	16632526	5	1	1	09/15/202	2	1200	0	sensor_09	22:36:56	11	1	1
19	16632544	1	0	0	09/15/202	0	1200	1	sensor_03	23:08:17	5	1	2
20	16632539	4	0	1	09/15/202	0	1200	0	sensor_06	22:59:24	5	0	0
21	16632534	4	2	2	09/15/202	1	1200	0	sensor_06	22:51:04	12	0	3
22	16632529	3	2	1	09/15/202	0	1200	0	sensor_09	22:43:19	10	2	2

REFERENCES

Connect to Cassandra Data and Transfer Data in Talend. (n.d.). CData Software. Retrieved from https://www.cdata.com/kb/tech/cassandra-jdbc-talend.rst

Managing NoSQL metadata. (n.d.). Talend. Retrieved from https://help.talend.com/r/en-US/7.3/studio-user-guide-big-data/managing-nosql-metadata