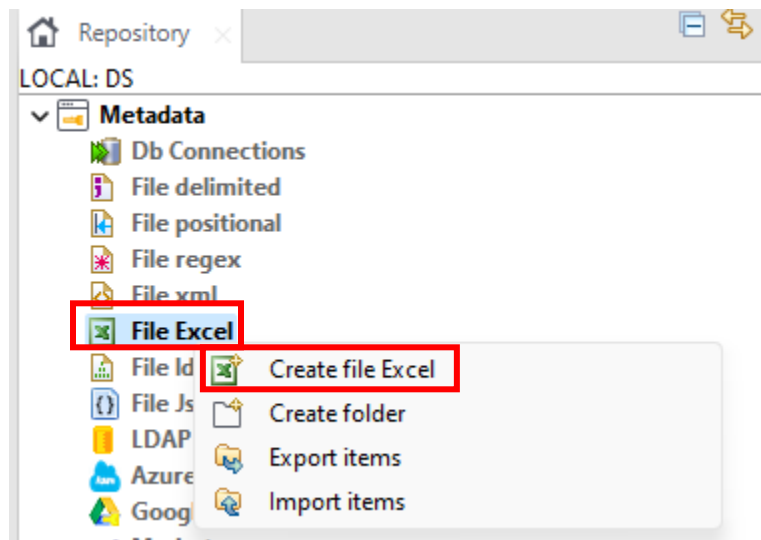
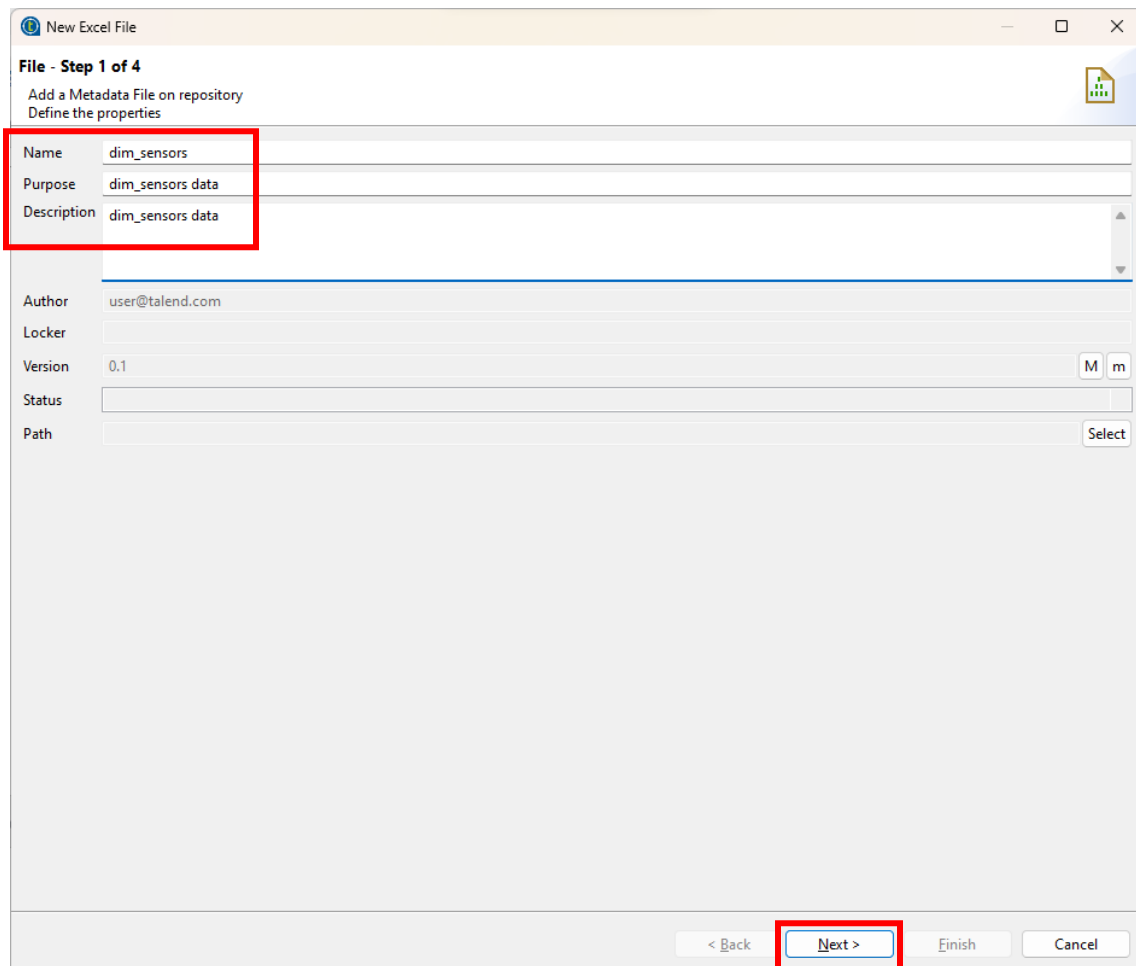


SA2 DOCUMENTATION

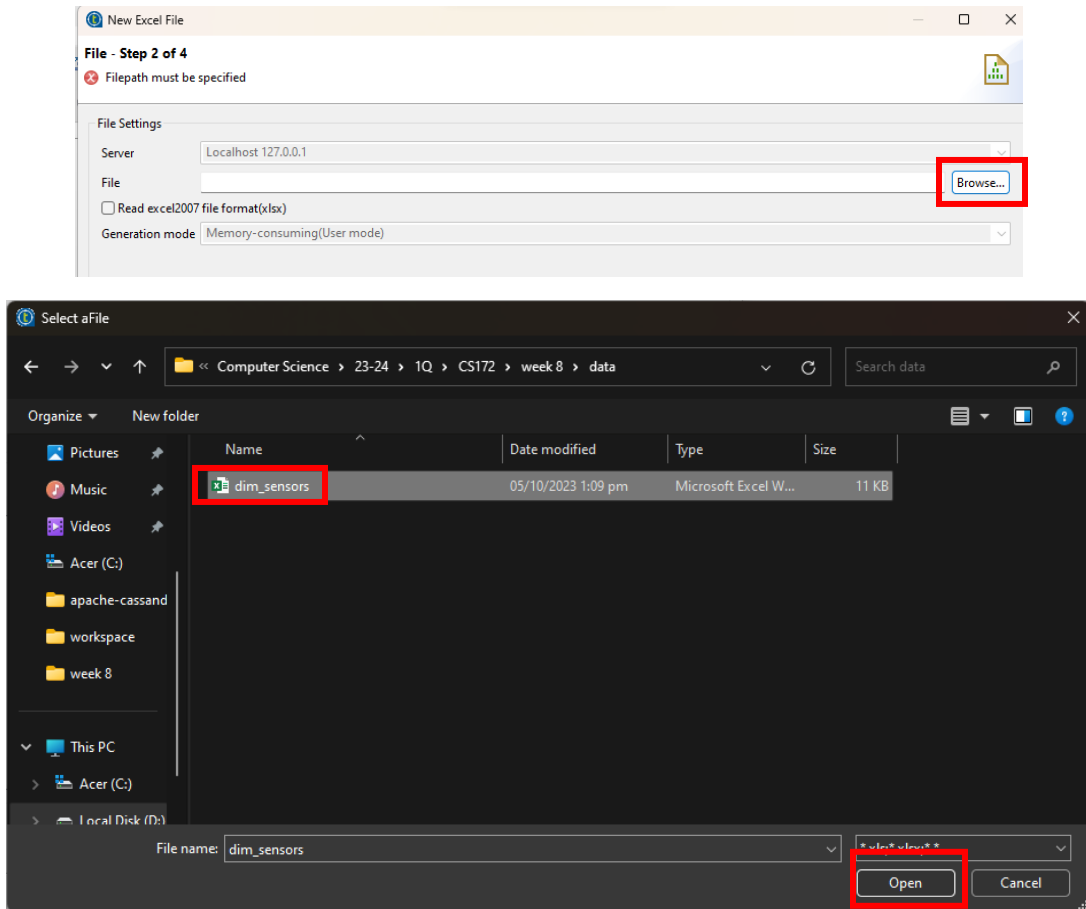
We will import data into a Cassandra table from an Excel file using Talend. First, we need to import the Excel file into Talend. Right-click on file Excel and click Create file Excel.



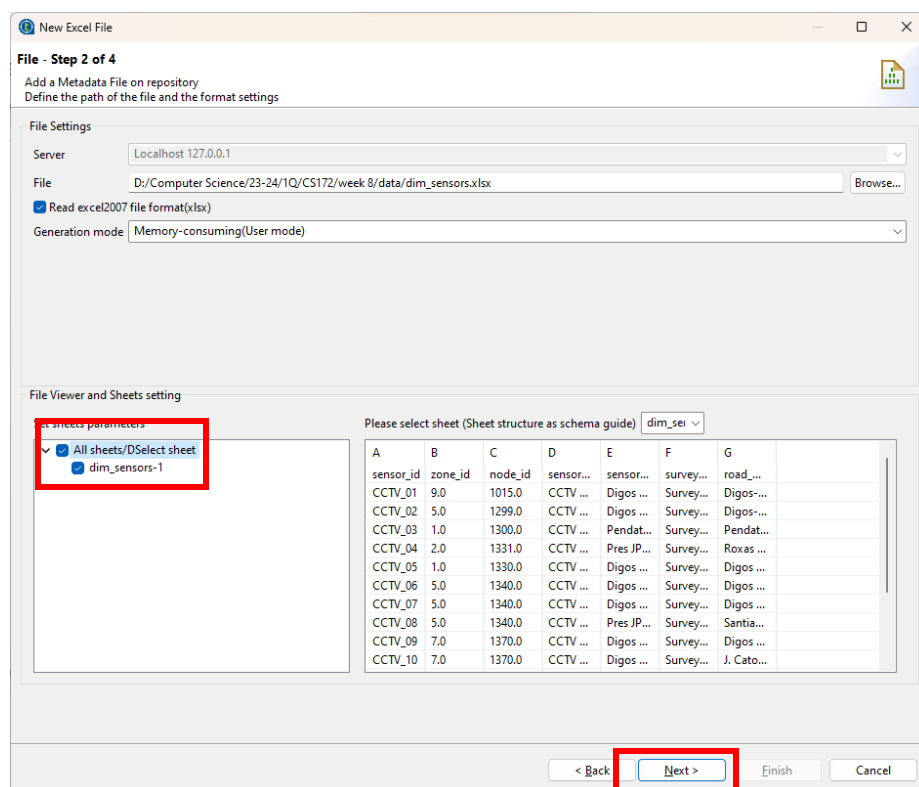
We will enter dim_sensors as the name, and dim_sensors data for purpose and description. And then we click next.



Now we click on Browse and click on the Excel file that contains our data. The file is dim_sensors. Next, we click open.



Now we can see the data from the Excel file. We will click on dim_sensors-1 to specify the data we need. Then we click next.



Now we see a preview of the data. We will click on set heading row as column names. Afterwards, we click next.

New Excel File

File - Step 3 of 4

Add a Metadata File on repository
Define the setting of the parse job

File Settings

Encoding: UTF-8

☐ Advanced separator(for number)

Thousands separator: ,

Decimal separator: .

Metadata column setting

First column: 1

Last column:

Rows To Skip

If any rows must be ignored, specify the following parameters

Header ☒ 1

Footer ☐

Limit Of Rows

If the number of lines must be limited, specify this number

Limit ☐

Preview | **Output**

☒ Set heading row as column names [Refresh Preview](#)

sensor_id	zone_id	node_id	sensor_type	sensor_location	survey_point	road_monitored
CCTV_01	9	1015	CCTV camera	Digos - Makar Rd - E. Bulaong Ave Intersection	Survey Point 1	Digos-Makar Rd
CCTV_02	5	1299	CCTV camera	Digos - Makar Rd - Pendatun Ave Intersection	Survey Point 2	Digos-Makar Rd
CCTV_03	1	1300	CCTV camera	Pendatun Ave - Pres JP Laurel Ave Intersection	Survey Point 3	Pendatun Ave.
CCTV_04	2	1331	CCTV camera	Pres JP Laurel Ave - Roxas Ave Intersection	Survey Point 4	Roxas Ave.
CCTV_05	1	1330	CCTV camera	Digos - Makar Rd - E. Roxas Ave Intersection	Survey Point 5	Digos Makar Rd
CCTV_06	5	1340	CCTV camera	Digos - Makar Rd - Santiago Blvd Intersection	Survey Point 6a	Digos Makar Rd
CCTV_07	5	1340	CCTV camera	Digos - Makar Rd - Santiago Blvd Intersection	Survey Point 6b	Digos Makar Rd

[Export as context](#) [Revert Context](#)

[< Back](#) [Next >](#) [Finish](#) [Cancel](#)

Now we can see the schema. We set the name of the schema as dim_sensors_metadata. Then we click finish to create the Excel file in Talend.

New Excel File

File - Step 4 of 4

Add a Schema on repository
Define the Schema

Name dim_sensors_metadata

Comment

Schema

[Click to update schema preview](#) [Guess](#)

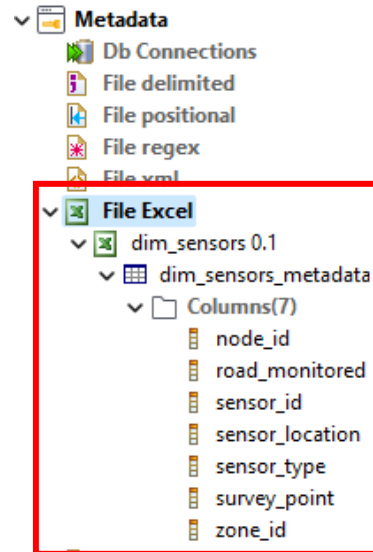
Description of the Schema

Column	Key	Type	<input checked="" type="checkbox"/> N..	Date Pattern (Ctrl+Spa...	Length	Precision	Default	Comment
sensor_id	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		7	0		
zone_id	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		1	0		
node_id	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
sensor_type	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		11	0		
sensor_location	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		49	0		
survey_point	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		15	0		
road_monitored	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		16	0		

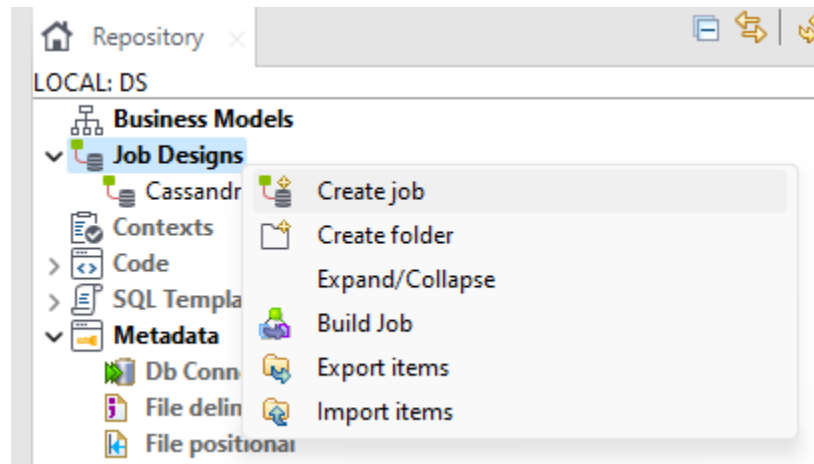
[+](#) [-](#) [↑](#) [↓](#) [📄](#) [📁](#) [🔍](#) [🔗](#)

[< Back](#) [Next >](#) [Finish](#) [Cancel](#)

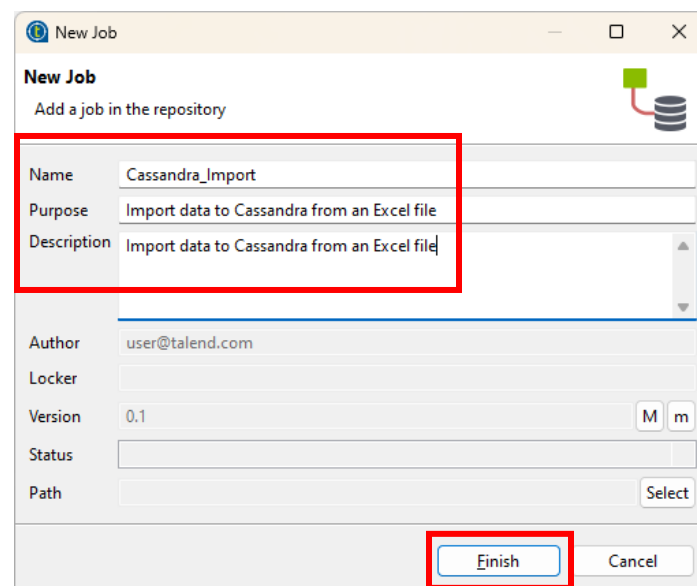
The dim_sensors Excel file is now created. We can verify this since we can now see the dim_sensors, dim_sensors_metadata, and the columns within it under the File Excel tab.



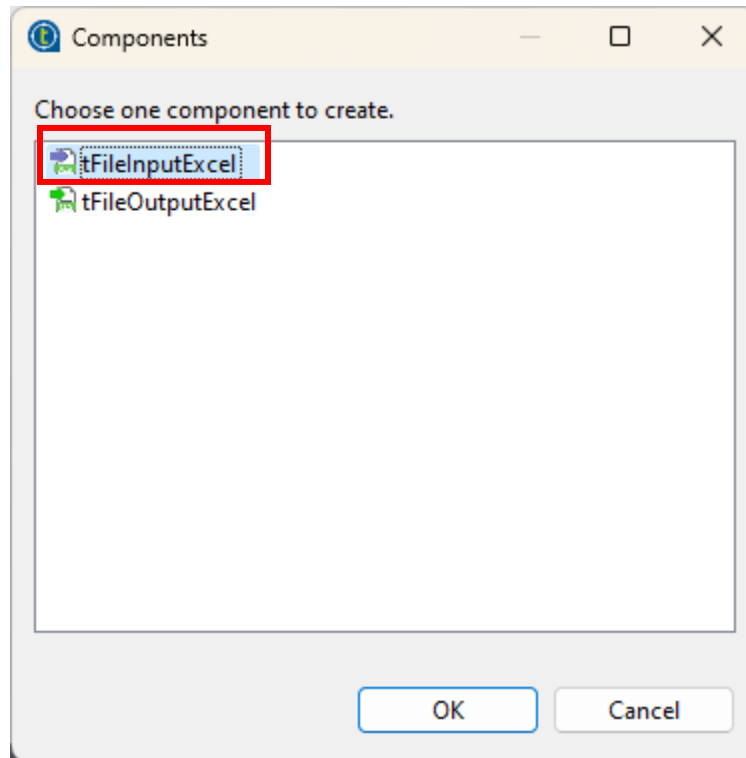
Now we need to create a job. We will right-click on Job Designs and click Create Job.



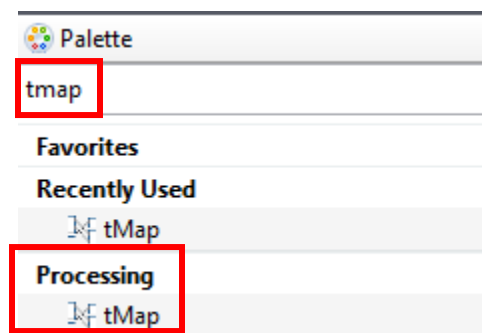
We set the name to Cassandra_Import and the purpose and description to Import data to Cassandra from an Excel file. Then we click finish.



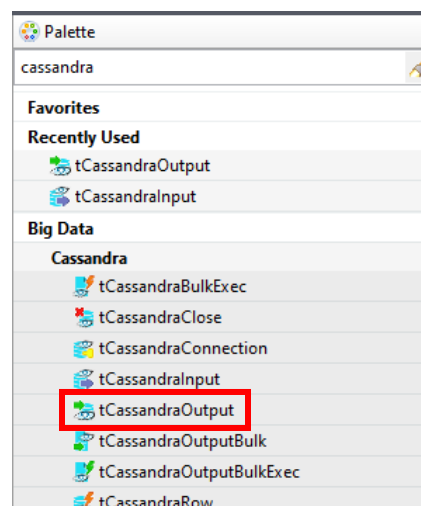
Now that we have our job, we will drag the dim_sensors Excel file we created earlier into the designer. We will click on tFileInputExcel as component.



Next we will search for tmap in the palette. Then we drag the component to the designer.



Now we will search for Cassandra in the palette. Then we drag the tCassandraOutput component to the designer since we will export data into Cassandra.



Before connecting, we need to make a Cassandra table that will store the data we have. In the Cassandra CQLSH we will create a table named dim_sensors. We can see that the table is created.

```
C:\Windows\System32\cmd.e X + v
cqlsh:keyspacedatasci> CREATE TABLE dim_sensors (sensor_id TEXT PRIMARY KEY, zone_id INT, node_id INT, sensor_type TEXT, sensor_location TEXT, survey_point TEXT, road_monitored TEXT);
cqlsh:keyspacedatasci> SELECT * FROM dim_sensors;

 sensor_id | node_id | road_monitored | sensor_location | sensor_type | survey_point | zone_id
-----+-----+-----+-----+-----+-----+-----
(0 rows)
cqlsh:keyspacedatasci> |
```

Then in the tCassandraOutput_1 component basic settings, we will set the host to 127.0.0.1. We then set the keyspace to keyspacedatasci since our table is within that keyspace. Next, we set the column family to dim_sensors since that is the name of the table we created.

tCassandraOutput_1

Basic settings

PROPERTY: Built-In

☐ Use existing connection

DB Version: Cassandra 3.0.x

Host: 127.0.0.1 * Port: 9042

☐ Required authentication

☐ Use SSL

Keyspace configuration

Keyspace: keyspacedatasci * Action on keyspace

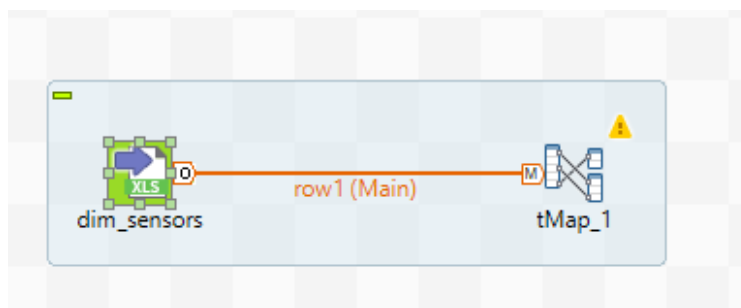
Column family configuration

Column family: dim_sensors * Action on column

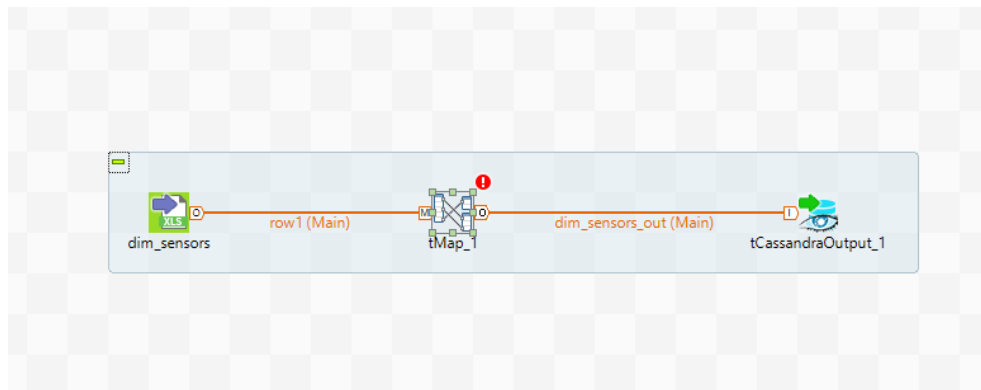
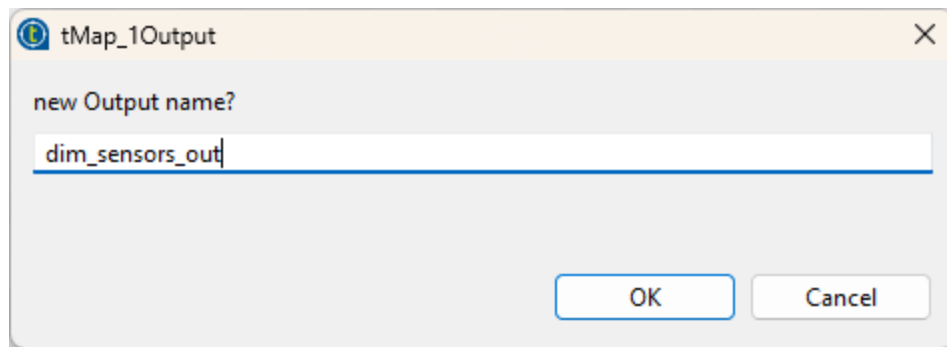
Action on data: Insert

Schema: Built-In Edit schema ...

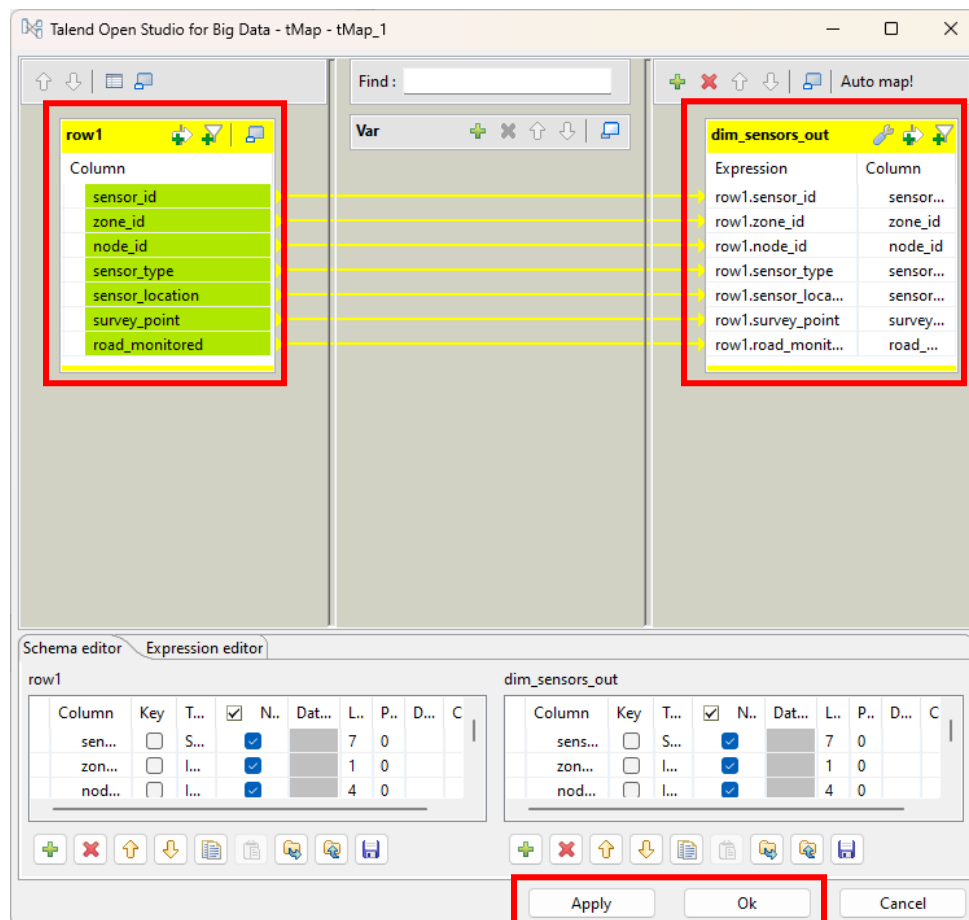
Then we will connect dim_sensors component to tMap_1 component. We can see it below.



Next, we connect the tMap_1 component to tCassandraOutput_1 component. We use dim_sensors_out as the output name and click ok.



Then in the tMap_1 component settings, we will select all columns in row1 and drag them to dim_sensors_out. Then we click on apply and ok.



Everything is now configured. We will just click on Run to run the job. The data from Excel file should then be imported to Cassandra.

The screenshot shows a data integration workflow in a tool like Talend. The workflow consists of four components connected in a sequence: **dim_sensors** (input), **tMap_1** (transform), **dim_sensors_out (Main)** (output), and **tCassandraOutput_1** (target). Performance metrics are displayed above the components: **dim_sensors** processes 15 rows in 0.88s at 16.97 rows/s; **tMap_1** processes 15 rows in 4.89s at 3.07 rows/s. The job is titled **Job Cassandra_Import** and is currently in the **Execution** phase. The logs show the job starting at 21:43 on 06/10/2023, connecting to a socket on port 3465, and ending at the same time with an exit code of 0.

Now if we check on the `dim_sensors` table in the Cassandra CQLSH, it should now be populated with our data. Here we can see that the data is now inside the table.

The screenshot shows a terminal window with the CQLSH prompt. A query has been executed: `SELECT * FROM dim_sensors;`. The results are displayed as a table with 7 columns: `sensor_id`, `node_id`, `road_monitored`, `sensor_location`, `sensor_type`, `survey_point`, and `zone_id`. The data is as follows:

sensor_id	node_id	road_monitored	sensor_location	sensor_type	survey_point	zone_id
CCTV_07	1340	Digos Makar Rd	Digos - Makar Rd - Santiago Blvd Intersection	CCTV camera	Survey Point 6b	5
CCTV_08	1340	Santiago Blvd.	Pres JP Laurel Ave - Santiago Blvd Intersection	CCTV camera	Survey Point 7	5
CCTV_09	1370	Digos Makar Rd	Digos Makar Rd - Jose Catolico Ave Intersection	CCTV camera	Survey Point 8a	7
CCTV_10	1370	J. Catolico Ave.	Digos Makar Rd - Jose Catolico Ave Intersection	CCTV camera	Survey Point 8b	7
CCTV_02	1299	Digos-Makar Rd	Digos - Makar Rd - Pendatun Ave Intersection	CCTV camera	Survey Point 2	5
CCTV_11	1370	J. Catolico Ave.	Digos Makar Rd - Jose Catolico Ave Intersection	CCTV camera	Survey Point 8c	7
CCTV_14	1360	Leon Llido St.	Salvani - Leon Llido St Intersection	CCTV camera	Survey Point 11	7
CCTV_12	1364	Digos Makar Rd	Digos Makar Rd - Honorio Arriola St. Intersection	CCTV camera	Survey Point 9	7
CCTV_04	1331	Roxas Ave.	Pres JP Laurel Ave - Roxas Ave Intersection	CCTV camera	Survey Point 4	2
CCTV_15	1396	Digos Makar Rd	Digos Makar Rd. - NLSA Road (Lagao Public Market)	CCTV camera	Survey Point 12	7
CCTV_05	1330	Digos Makar Rd	Digos - Makar Rd - E. Roxas Ave Intersection	CCTV camera	Survey Point 5	1
CCTV_13	1310	Aparente St.	Salvani - Apparante St Intersection	CCTV camera	Survey Point 10	1
CCTV_06	1340	Digos Makar Rd	Digos - Makar Rd - Santiago Blvd Intersection	CCTV camera	Survey Point 6a	5
CCTV_03	1300	Pendatum Ave.	Pendatun Ave - Pres JP Laurel Ave Intersection	CCTV camera	Survey Point 3	1
CCTV_01	1015	Digos-Makar Rd	Digos - Makar Rd - E. Bulaong Ave Intersection	CCTV camera	Survey Point 1	9

The terminal shows 15 rows of data. The prompt `cqlsh:keystacedatasci>` is visible at the bottom.

References

Singh, B. (2021, December 8). Importing Data into Cassandra with CSV - Balvinder Singh - Medium.

Medium. <https://erbalvindersingh.medium.com/importing-data-into-cassandra-with-csv-a006ff5f2904>

Talend. (n.d.) Guide to talend. https://www.tutorialspoint.com/talend/talend_quick_guide.htm

Talend. (2023, June 9). Talend | A complete, scalable data management solution.

<https://www.talend.com/>

Talend Help Center. (n.d.). Excel file (XLSX) import. [https://help.talend.com/r/en-](https://help.talend.com/r/en-US/7.3/tdc_bridges/mirexcelfileimport)

[US/7.3/tdc_bridges/mirexcelfileimport](https://help.talend.com/r/en-US/7.3/tdc_bridges/mirexcelfileimport)

Talend Help Center. (n.d.). Creating a connection to a Cassandra database.

<https://help.talend.com/r/en-US/7.3/studio-user-guide/creating-connection-to-cassandra-database>