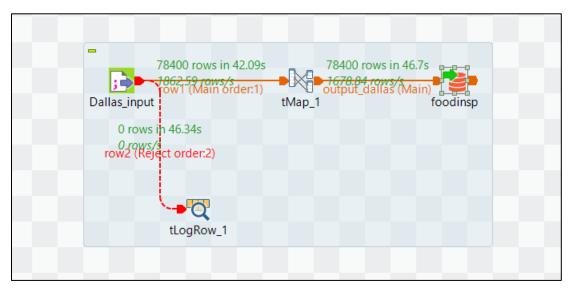
Staging Dallas Food Inspection in SQL Server

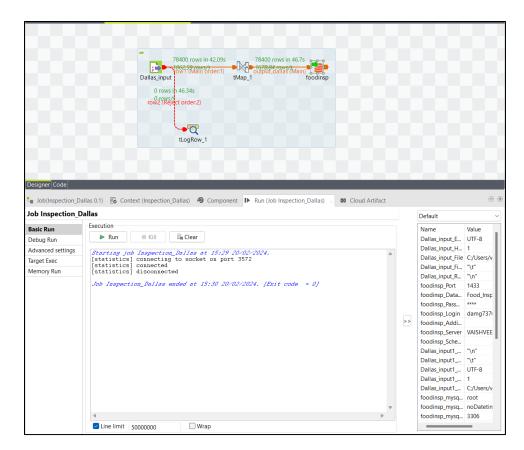
Talend Workflow Explanation:



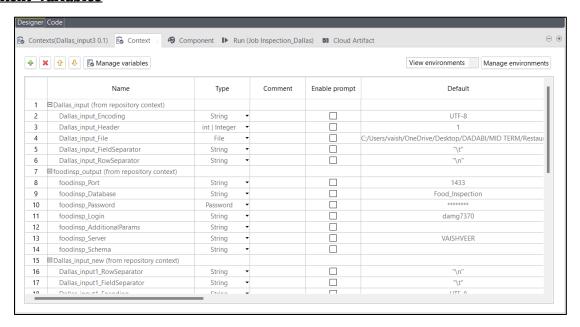
This workflow is reading data from a CSV file, transforms it through tMap_1, sending the main body of transformed data to both an output component which is a SQL Server, and has provisions for handling rejected data and logging it. But there are no rejected rows.

Talend Workflow Components:

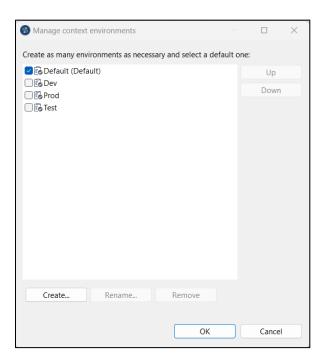
- Dallas_input: This is the input component that reads data from source, which is a CSV file. This component has processed 78,400 rows of data in 42.09 seconds.
- tMap_1: The tMap component is used to transform data. Here, it takes the input data (from Dallas_input) and processes it and the same number of rows 78400 rows in 46.7s is passed.
- output_dallas (Main): This is the main workflow receiving the processed data from the tMap_1 component.
- foodinsp: This is the output component, connected to the output_dallas used to stage the processed data in the SQL Server.
- row2 (Reject Order:2): row2 is a reject flow where rows that do not meet certain criteria in the tMap are sent. No rows are rejected as indicated by 0 rows.
- tLogRow_1: This component is used to log data rows that are rejected and outputs to the console.



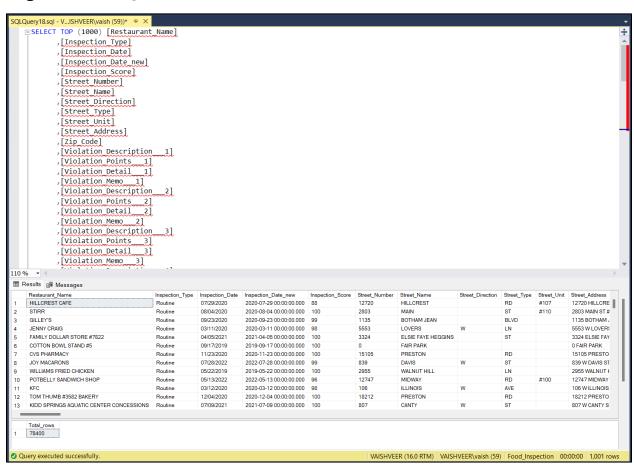
Context Variables



Environments

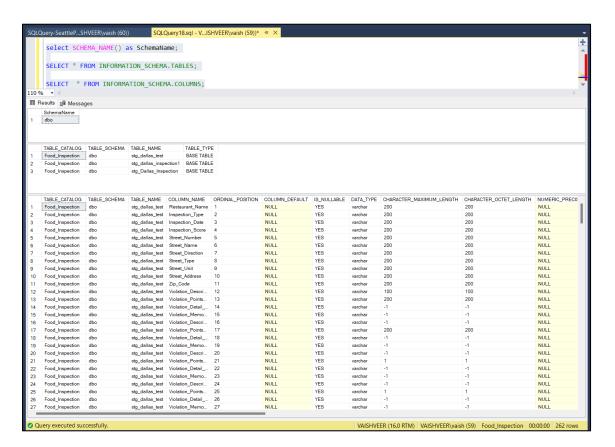


Staged Table In SQL Server



```
SQLQuery-SeattleP...SHVEER\vaish (60)) SQLQuery18.sql - V...ISHVEER\vaish (59))* → ×
      SELECT
            t.name AS TableName,
            c.name AS ColumnName,
            ty.name AS DataType,
            c.max_length AS Size
            c.precision AS Precision,
c.scale AS Scale,
            c.is_nullable AS IsNullable,
            {\tt c.is\_identity} \  \, {\tt AS} \  \, {\tt IsIdentity}
       sys.tables t
            sys.columns c ON t.object_id = c.object_id
       INNER JOIN
            sys.types ty ON c.user_type_id = ty.user_type_id
       WHERE
            t.name = 'stg_Dallas_Inspection'
       ORDER BY
       t.name, c.column_id;
110 % -
DataType Size Precision
                                                 varchar
varchar
varchar
datetime
                                                           200
                                                           200
200
8
      stg Dallas Inspection Inspection Score
                                                  varchar
                                                           200
      stg_Dallas_Inspection Street_Number
                                                           200
      stg_Dallas_Inspection Street_Number stg_Dallas_Inspection Street_Direction stg_Dallas_Inspection Street_Type
                                                  varchar
varchar
varchar
varchar
                                                           200
200
200
200
      stg Dallas Inspection Street Unit
                                                  varchar
                                                           200
      varchar
varchar
varchar
varchar
                                                           200
200
100
      stg_Dallas_Inspection Violation_Points___1
                                                           200
      stq Dallas Inspection Violation Detail 1
                                                  varchar
      stg_Dallas_Inspection Violation_Description_
stg_Dallas_Inspection Violation_Description_
stg_Dallas_Inspection Violation_Points__2
      stg_Dallas_Inspection Violation_Detail___2
                                                  varchar
      varchar
                                                                                                                            VAISHVEER (16.0 RTM) | VAISHVEER\vaish (59) | Food_Inspection | 00:00:00 | 117 rows

    Query executed successfully.
```



DDL Script for stage table

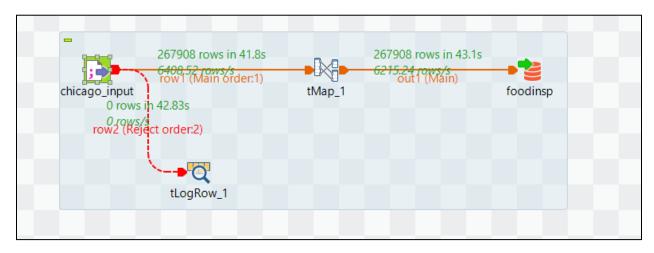
```
USE [Food_Inspection]
GO
SET ANSI NULLS ON
GO
SET QUOTED IDENTIFIER ON
GO
CREATE TABLE [dbo].[stg Dallas Inspection](
       [Restaurant_Name] [varchar](200) NULL,
       [Inspection_Type] [varchar](200) NULL,
       [Inspection_Date] [varchar](200) NULL,
       [Inspection_Date_new] [datetime] NULL,
       [Inspection Score] [varchar](200) NULL,
       [Street Number] [varchar](200) NULL,
       [Street_Name] [varchar](200) NULL,
       [Street_Direction] [varchar](200) NULL,
       [Street_Type] [varchar](200) NULL,
       [Street_Unit] [varchar](200) NULL,
       [Street_Address] [varchar](200) NULL,
       [Zip Code] [varchar](200) NULL,
       [Violation_Description___1] [varchar](100) NULL,
       [Violation_Points___1] [varchar](200) NULL,
       [Violation Detail 1] [varchar](max) NULL,
       [Violation_Memo___1] [varchar](max) NULL,
       [Violation_Description___2] [varchar](max) NULL,
       [Violation Points 2] [varchar](200) NULL,
       [Violation Detail 2] [varchar](max) NULL,
       [Violation Memo 2] [varchar](max) NULL,
       [Violation_Description___3] [varchar](max) NULL,
       [Violation_Points___3] [varchar](1) NULL,
       [Violation_Detail___3] [varchar](max) NULL,
       [Violation Memo 3] [varchar](max) NULL,
       [Violation Description 4] [varchar](max) NULL,
       [Violation_Points___4] [varchar](1) NULL,
       [Violation_Detail___4] [varchar](max) NULL,
       [Violation_Memo___4] [varchar](max) NULL,
       [Violation_Description___5] [varchar](max) NULL,
       [Violation_Points___5] [varchar](max) NULL,
       [Violation Detail 5] [varchar](max) NULL,
       [Violation_Memo___5] [varchar](max) NULL,
        [Violation Description 6] [varchar](max) NULL,
        [Violation_Points___6] [varchar](max) NULL,
       [Violation_Detail___6] [varchar](max) NULL,
       [Violation_Memo___6] [varchar](max) NULL,
       [Violation_Description___7] [varchar](max) NULL,
       [Violation_Points___7] [varchar](max) NULL,
       [Violation_Detail___7] [varchar](max) NULL,
       [Violation_Memo___7] [varchar](max) NULL,
       [Violation_Description___8] [varchar](max) NULL,
       [Violation_Points___8] [varchar](max) NULL,
       [Violation Detail 8] [varchar](max) NULL,
```

```
[Violation_Memo___8] [varchar](max) NULL,
[Violation_Description___9] [varchar](max) NULL,
[Violation_Points___9] [varchar](max) NULL,
[Violation_Detail___9] [varchar](max) NULL,
[Violation Memo 9] [varchar](max) NULL,
[Violation Description 10] [varchar](max) NULL,
[Violation_Points___10] [varchar](max) NULL,
[Violation_Detail___10] [varchar](max) NULL,
[Violation_Memo___10] [varchar](max) NULL,
[Violation_Description___11] [varchar](max) NULL,
[Violation_Points___11] [varchar](max) NULL,
[Violation_Detail___11] [varchar](max) NULL,
[Violation_Memo___11] [varchar](max) NULL,
[Violation Description 12] [varchar] (max) NULL,
[Violation_Points___12] [varchar](max) NULL,
[Violation_Detail___12] [varchar](max) NULL,
[Violation_Memo___12] [varchar](max) NULL,
[Violation Description 13] [varchar](max) NULL,
[Violation Points 13] [varchar](100) NULL,
[Violation Detail 13] [varchar](1000) NULL,
[Violation_Memo___13] [varchar](max) NULL,
[Violation_Description___14] [varchar](max) NULL,
[Violation_Points___14] [varchar](100) NULL,
[Violation Detail 14] [varchar](max) NULL,
[Violation_Memo___14] [varchar](max) NULL,
[Violation_Description___15] [varchar](max) NULL,
[Violation_Points___15] [varchar](1001) NULL, [Violation_Detail___15] [varchar](max) NULL,
[Violation_Memo___15] [varchar](max) NULL,
[Violation Description 16] [varchar](max) NULL,
[Violation_Points___16] [varchar](100) NULL,
[Violation_Detail___16] [varchar](max) NULL,
[Violation Memo 16] [varchar](max) NULL,
[Violation_Description___17] [varchar](max) NULL,
[Violation_Points___17] [varchar](100) NULL,
[Violation_Detail___17] [varchar](max) NULL,
[Violation_Memo___17] [varchar](max) NULL,
[Violation_Description___18] [varchar](max) NULL,
[Violation_Points___18] [varchar](100) NULL,
[Violation_Detail___18] [varchar](max) NULL,
[Violation_Memo___18] [varchar](max) NULL,
[Violation Description 19] [varchar](max) NULL,
[Violation Points 19] [varchar](100) NULL,
[Violation Detail 19] [varchar](max) NULL,
[Violation_Memo___19] [varchar](max) NULL,
[Violation_Description___20] [varchar](max) NULL,
[Violation_Points___20] [varchar](100) NULL,
[Violation_Detail___20] [varchar](max) NULL,
[Violation Memo 20] [varchar](max) NULL,
[Violation_Description___21] [varchar](max) NULL,
[Violation_Points___21] [varchar](100) NULL,
[Violation Detail 21] [varchar](max) NULL,
[Violation_Memo___21] [varchar](max) NULL,
[Violation_Description___22] [varchar](max) NULL,
[Violation_Points___22] [varchar](1000) NULL,
[Violation_Detail___22] [varchar](max) NULL,
```

```
[Violation_Memo___22] [varchar](max) NULL,
        [Violation_Description___23] [varchar](max) NULL,
        [Violation_Points___23] [varchar](100) NULL,
        [Violation_Detail___23] [varchar](max) NULL,
        [Violation Memo 23] [varchar](max) NULL,
        [Violation Description 24] [varchar] (max) NULL,
        [Violation_Points___24] [varchar](100) NULL,
        [Violation_Detail___24] [varchar](max) NULL,
        [Violation_Memo___24] [varchar](max) NULL,
        [Violation_Description___25] [varchar](max) NULL,
        [Violation_Points___25] [varchar](100) NULL,
        [Violation_Detail___25] [varchar](max) NULL,
        [Violation_Memo___25] [varchar](max) NULL,
        [Inspection Month] [varchar](8) NULL,
        [Inspection Year] [varchar](20) NULL,
        [Lat_Long_Location] [varchar](1000) NULL,
        [DI_CreateDate] [datetime] NULL,
        [DI WorkflowName] [varchar](25) NULL
) ON [PRIMARY] TEXTIMAGE_ON [PRIMARY]
```

Staging Chicago Food Inspection in SQL Server

Talend Workflow Explanation:



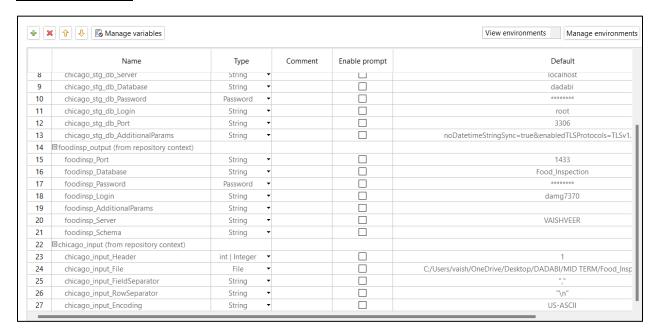
This workflow is reading data from a CSV file, transforms it through tMap_1, sending the main body of transformed data to both an output component which is a SQL Server, and has provisions for handling rejected data and logging it. But there are no rejected rows.

Talend Workflow Components:

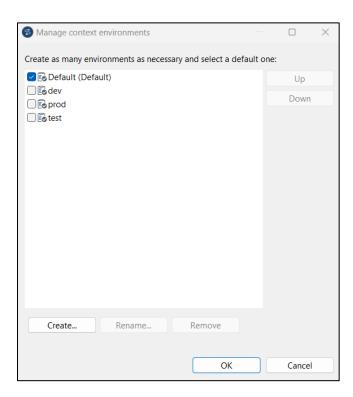
- chicago_input: This is the input component that reads data from source, which is a CSV file. This component has processed 267908 rows of data in 6408.25 seconds.
- tMap_1: The tMap component is used to transform data. Here, it takes the input data (from Dallas_input) and processes it and the same number of rows in 6215.24 is passed.
- out1 (Main): This is the main workflow receiving the processed data from the tMap_1 component.
- foodinsp: This is the output component, connected to the out1 used to stage the processed data in the SQL Server.

- row2 (Reject Order:2): row2 is a reject flow where rows that do not meet certain criteria in the tMap are sent. No rows are rejected as indicated by 0 rows.
- tLogRow_1: This component is used to log data rows that are rejected and outputs to the console.

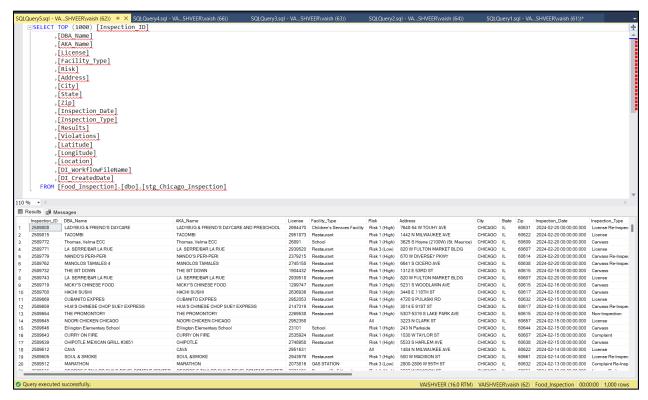
Context Variable

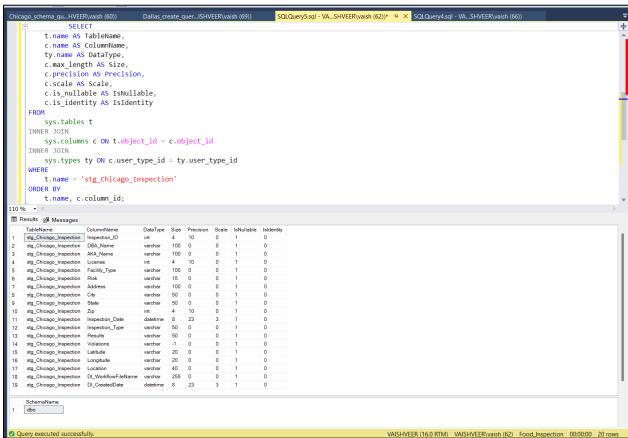


Environment



Stagged Table in SOL Server





```
select count (*) from stg Chicago Inspection

110 %

Results Messages

(No column name)

1 267908
```

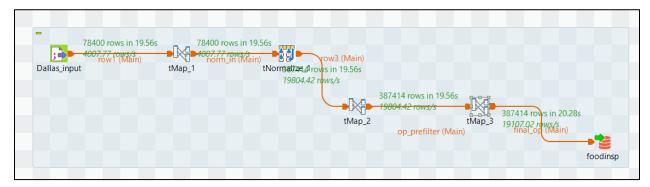
DDL Script for Stage Table:

```
USE [Food_Inspection]
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[stg_Chicago_Inspection](
       [Inspection ID] [int] NULL,
       [DBA_Name] [varchar](100) NULL,
       [AKA_Name] [varchar](100) NULL,
       [License] [int] NULL,
       [Facility_Type] [varchar](100) NULL,
       [Risk] [varchar](15) NULL,
       [Address] [varchar](100) NULL,
       [City] [varchar](50) NULL,
       [State] [varchar](50) NULL,
       [Zip] [int] NULL,
       [Inspection_Date] [datetime] NULL,
       [Inspection_Type] [varchar](50) NULL,
       [Results] [varchar](50) NULL,
       [Violations] [varchar](max) NULL,
       [Latitude] [varchar](20) NULL,
       [Longitude] [varchar](20) NULL,
       [Location] [varchar](40) NULL,
       [DI_WorkflowFileName] [varchar](255) NULL,
       [DI_CreatedDate] [datetime] NULL
ON [PRIMARY] TEXTIMAGE_ON [PRIMARY]
GO
```

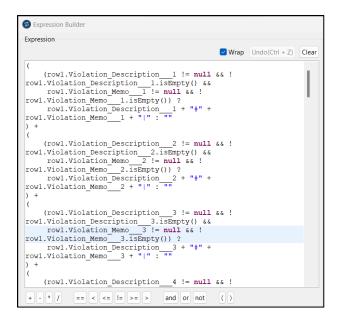
```
⇒ × SQLQuery5.sql - VA...SHVEER\vaish (62))
     USE [Food Inspection]
    GO
     SET ANSI_NULLS ON
     SET QUOTED_IDENTIFIER ON
    CREATE TABLE [dbo].[stg_Chicago_Inspection](
         [Inspection_ID] [int] NULL,
[DBA_Name] [varchar](100) NULL,
         [AKA_Name] [varchar](100) NULL,
         [License] [int] NULL
         [Facility Type] [varchar](100) NULL,
         [Risk] [varchar](15) NULL
         [Address] [varchar](100) NULL,
         [City] [varchar](50) NULL,
         [State] [varchar](50) NULL,
         [Zip] [int] NULL,
         [Inspection_Date] [datetime] NULL,
         [Inspection_Type] [varchar](50) NULL, [Results] [varchar](50) NULL,
         [Violations] [varchar](max) NULL,
         [Latitude] [varchar](20) NULL,
[Longitude] [varchar](20) NULL,
         [Location] [varchar](40) NULL,
         [DI_WorkflowFileName] [varchar](255) NULL,
         [DI_CreatedDate] [datetime] NUL
      ON [PRIMARY] TEXTIMAGE_ON [PRIMARY]
110 % ▼ ◀
                                                                                                   VAISHVEER (16.0 RTM) | VAISHVEER\vaish (59) | Food_Inspection | 00:00:00 | 0 rows
```

Cleaned, Normalized, Datatype Converted Dallas Dataset (stagged in SQL Server)

Talend Workflow Explanation:



In this workflow we have taken the input file in csv format and performed transformations in tMap1. In this tmap new columns are created to represent the cleaned data, ie Inspection Date is checked for null values and converted to Date format. If any null values were encountered, we replaced it with '01/01/1990'. Similarly for any column of type string we have handled the null values by replacing with 'NA' and for any numeric column with null values we have replaced it with 0000. Furthermore, in tMap1 we have added a condition to merge the data from Violation Description and Violation Memo into a single row (as below) so that it can be given to tNormalize to perform multiple row transformations.



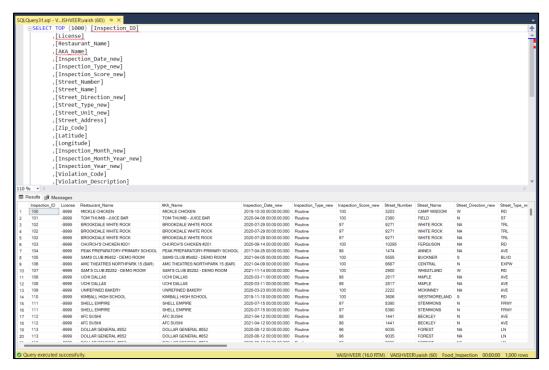
This transformed data from the tmap is given as input to tNormalize 1 where it separates the merged row into multiple rows based on delimeter '|'.

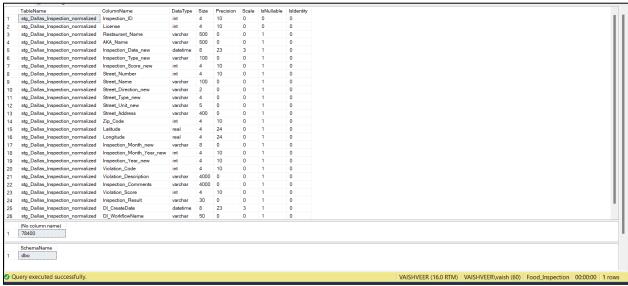


Now, the output of the tNormalize1 is given to a tMap2 to further split the multiple rows into respective columns-Violation_Code, Violation_Description and Inspection_Comments(Violation_Memo as reference). Furthermore, a new column is added where the respective Violation_Points are added together to get Violation_Score. In tMap3, another column is added to get the Inspection_Results based off the calculations of Violation_Score.

We have also added a column AKA_Name where the restaurant name has been duplicated to fill in. We have also added a column for risk category which has been populated with the value 'Invalid' The final output is a stagged table in SQL Server.

Dallas Final Staged Table





Create Script

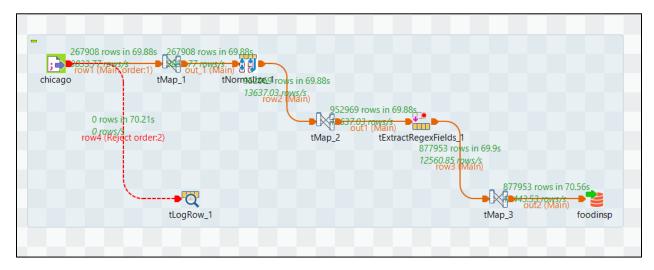
```
SCIDENTYLEY OF TABLE [Mob]. Stg. Dallas_Inspection_normalized] Script Date: 2/24/2024 5:02:14 PM *****/
SST ANSI_NULLS ON

SST QUOTED_IDENTIFIER ON
60

SCREATE TABLE [Mob].[stg. Dallas_Inspection_normalized] (
[Inspection_10] [int] NOT NULL,
[License] [int] NOT NULL,
[License] [int] NOT NULL,
[Inspection_10] [date_ine] [Mot.],
[Inspection_10] [date_ine] [date_ine] NULL,
[Inspection_Type_ne] [date_ine] NULL,
[Inspection_Type_ne] [varchar](100) NULL,
[Inspection_Type_ne] [varchar](100) NULL,
[Street_Name] [Varchar](100) NULL,
[Street_Name] [Varchar](100) NULL,
[Street_Name] [Varchar](100) NULL,
[Street_Type_ned] [varchar](100) NULL,
[Street_Type_ned] [varchar](100) NULL,
[Street_Type_ned] [varchar](100) NULL,
[Inspection_type_ned] [varchar](100) NU
```

Cleaned, Normalized, Datatype Converted Chicago Dataset(stagged in SQL Server)

Talend Workflow Explanation:



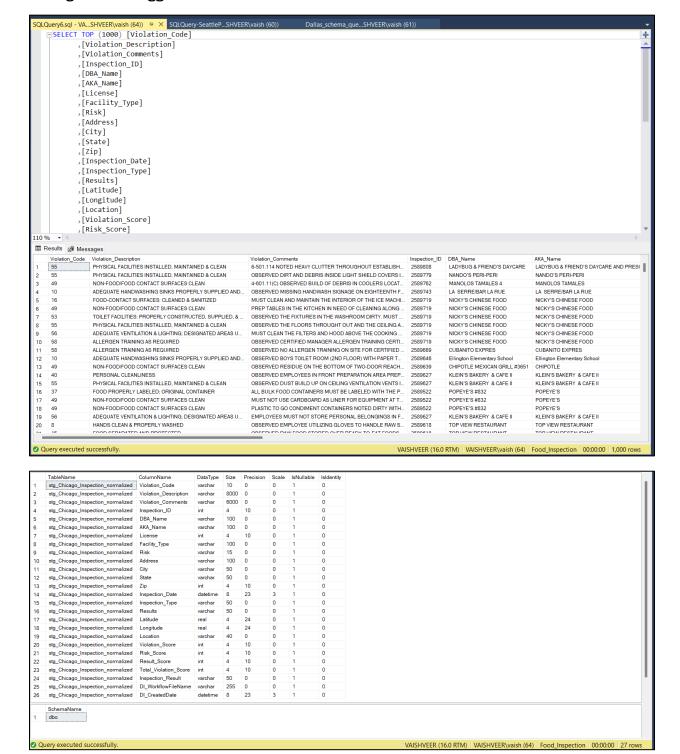
In this workflow we have taken the input file in csv format and performed transformations in tMap1. In this tmap existing columns are loaded to represent the cleaned data, ie for example Inspection Date is checked for null values and converted to Date format. If any null values were encountered, we replaced it with '01/01/1990'. Similarly for any column of type string we have handled the null values by replacing with 'NA' and for any numeric column with null values we have replaced it with 0000.

This transformed data from the tmap is given as input to tNormalize 1 where it separates the merged row for Violations into multiple rows based on delimeter '|'.

Now, the output of the tNormalize1 is given to a tMap2 to further split the multiple rows. This is then given to tExtractRegexFields_1 to split the multiple rows into respective columns for-Violation_Code, Violation_Description and Inspection_Comments. In tMap3, another column is added to get the Violation_Score based on the provided guidelines and then another new column Inspection_Results is created based off the calculations of Violation_Score.

Column to normalize Violations	* Item separator	"\\"

Chicago Final Stagged Table



Create Script

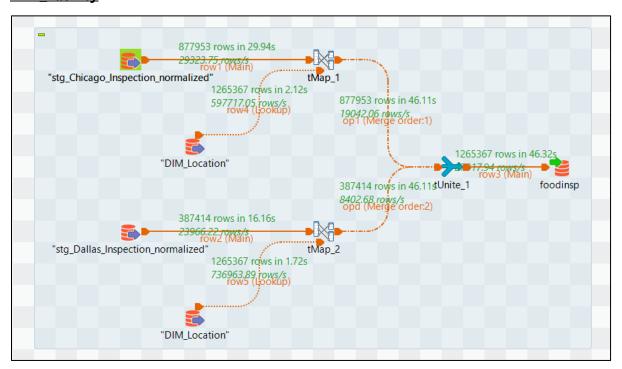
```
USE [Food_Inspection]
        GO
         /****** Object: Table [dbo].[stg_Chicago_Inspection_normalized] Script Date: 2/23/2024 8:57:01 PM ******/
         SET QUOTED_IDENTIFIER ON
       GREATE TABLE [dbo].[stg_Chicago_Inspection_normalized](
   [Violation_Code] [varchar](10) NULL,
                [Violation_Description] [varchar](8000) NULL,
               [Violation_Comments] [varchar](6000) NULL,
               [Inspection_ID] [int] NULL,
[DBA_Name] [varchar](100) NULL,
[AKA_Name] [varchar](100) NULL,
               [License] [int] NULL,

[Facility_Type] [varchar](100) NULL,

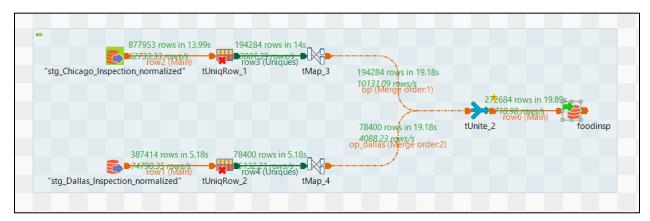
[Risk] [varchar](15) NULL,
               [Address] [varchar](100) NULL,
                [City] [varchar](50) NULL,
               [State] [varchar](50) NULL,
[Zip] [int] NULL,
               [Inspection_Date] [datetime] NULL,
               [Inspection_oter] (varchar](50) NULL,
[Results] (varchar](50) NULL,
[Latitude] [real] NULL,
[Longitude] [real] NULL,
[Location] [varchar](40) NULL,
                [Violation_Score] [int] NULL,
               [Risk_Score] [int] NULL,
[Result_Score] [int] NULL,
[Total_Violation_Score] [int] NULL,
               [Inspection_Result] [varchar](50) NULL,
[DI_WorkflowFileName] [varchar](255) NULL,
[DI_CreatedDate] [datetime] NULL
         ) ON [PRIMARY]
110 % ▼ ◀

₩ Connected. (1/1)
                                                                                                                                                     VAISHVEER (16.0 RTM) | VAISHVEER\vaish (64) | Food_Inspection | 00:00:00 | 0 rows
```

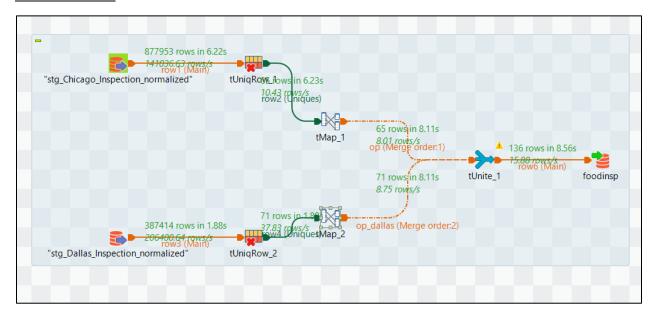
DIM_Facility



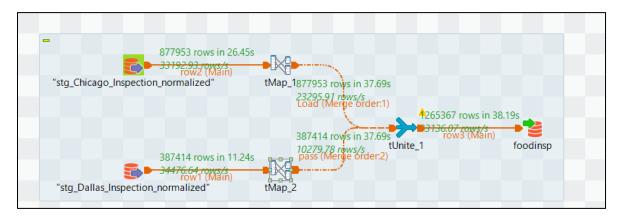
DIM_Inspections



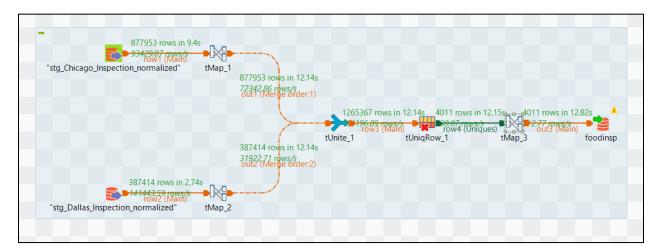
DIM Violation



DIM Location



DIM Date



FACT_Inspection

