**ETL – Talend Workflow**

**CHICAGO:**

A screenshot of a computer

Description automatically generated

* Added the Chicago dataset in CSV format.
* Processed 952816 rows in 94.94 seconds.
* Used T-normalize to normalize data in Violations column. There was more than 1 violation found in the Violations cell, and hence we split it into 1 violation per record.

A screenshot of a computer

Description automatically generated

* Two T-map components have been used
* The first T-map: Added three functions (DI\_Process\_ID, DI\_CurrentDate, DI\_WorkflowFileName)- As shown in the below screenshots

A close up of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

* Second T-map has been used to split the Violations into multiple columns (Violations, Violation Code, Violation Description, Violation Comment) – As shown in the below screenshot

A screenshot of a computer

Description automatically generated

Regex for Violation\_Count, Risk\_Score, Result\_Score, Total\_Violation\_points.

Violation\_Count

A screenshot of a computer

Description automatically generated

Risk\_Score

A screenshot of a computer

Description automatically generated

Result\_Score

A screenshot of a computer program

Description automatically generated

Total\_Violation\_Points

A screenshot of a computer

Description automatically generated

**DALLAS:**

A screenshot of a computer

Description automatically generated

* Added the Dallas dataset in CSV format.
* Processed 1960000 rows in 160.14 seconds
* No reject rows have been found.
* Two T-map components have been used
* First T-map: All the violations are found in a split state in the CSV file (Violation Memo, Violation Detail, Violation Description, Violation Comment). Combines all of them into a single column (Violations All). Added a separate column (Violation Points) for calculations. – As shown in the below screenshot

A screenshot of a computer

Description automatically generated

* Used the t-normalize component to Split the Violations in a cleaner manner, using item separator symbol ("\\|").

A screenshot of a computer

Description automatically generated

* Second T-map: The Violations All column has been split into four columns (Violation Code, Violation Description, Violation Comment, Violation Points) – As shown in the screenshot below.

A screenshot of a computer

Description automatically generated

* Two functions (DI\_CurrentDate and DI\_WorkflowFileName) have been added.
* Created two columns Risk and Facility\_Type and gave them a default value.
* Finally, everything has been mapped to the output server – Microsoft SQL Server.

**Loading Dimensional Model - Talend Workflow**

1. **Dim\_Business\_License**

A screenshot of a computer

Description automatically generated

T-map 1 of Dim\_Business\_License

A screenshot of a computer

Description automatically generated

T-map 2 of Dim\_Business\_License

A screenshot of a computer

Description automatically generated

T-Map 3 of Dim\_Business\_License

A screenshot of a computer

Description automatically generated

T-UniqRow of Dim\_Business\_License

A screenshot of a computer

Description automatically generated

1. **Dim\_Date**

A screenshot of a computer

Description automatically generated

T-Map of Dim\_Date

A screenshot of a computer

Description automatically generated

1. **Dim\_Facility\_Type**

A screenshot of a computer

Description automatically generated

T-Map 1 of Dim\_Facility\_Type

A screenshot of a computer

Description automatically generated

T-Map 2 of Dim\_Facility\_Type

A screenshot of a computer

Description automatically generated

T-Map 3 of Dim\_Facility\_Type

A screenshot of a computer

Description automatically generated

T-UniqRow of Dim\_Facility\_Type

A screenshot of a computer

Description automatically generated

1. **Dim\_Food\_Inspections\_Results**

A screenshot of a computer

Description automatically generated

T-Map 1 of Dim\_Food\_Inspections\_Results

A screenshot of a computer

Description automatically generated

T-Map 2 of Dim\_Food\_Inspections\_Results

A screenshot of a computer

Description automatically generated

T-Map 3 of Dim\_Food\_Inspections\_Results

A screenshot of a computer

Description automatically generated

T-UniqRow of Dim\_Food\_Inspections\_Results

A screenshot of a computer

Description automatically generated

1. **Dim\_InspectionType**

A screenshot of a computer

Description automatically generated

T-Map 1 of Dim\_InspectionType

A screenshot of a computer

Description automatically generated

T-Map 2 of Dim\_InspectionType

A screenshot of a computer

Description automatically generated

T-Map 3 of Dim\_InspectionType

A screenshot of a computer

Description automatically generated

T-UniqRow of Dim\_InspectionType

A screenshot of a computer

Description automatically generated

1. **Dim\_Risk\_Category**A screenshot of a computer

   Description automatically generatedT-Map 1 of Dim\_Risk\_Category A screenshot of a computer

   Description automatically generated

T-Map 2 of Dim\_Risk\_Category

A screenshot of a computer

Description automatically generated

T-Map 3 of Dim\_Risk\_Category

A screenshot of a computer

Description automatically generated

T-UniqRow of Dim\_Risk\_Category

A screenshot of a computer

Description automatically generated

1. **Dim\_Violation**

A screenshot of a computer

Description automatically generated

T-Map 1 of Dim\_ViolationA screenshot of a computer

Description automatically generated

T-Map 2 of Dim\_Violation

A screenshot of a computer

Description automatically generated

T-Map 3 of Dim\_Violation

A screenshot of a computer

Description automatically generated

T-UniqRow of Dim\_Violation

A screenshot of a computer

Description automatically generated

1. **Fact\_Inspection**

A computer screen shot of a computer

Description automatically generated

T-Map 1 of Fact\_Inspection

A screenshot of a computer

Description automatically generated

T-Map 2 of Fact\_Inspection

A screenshot of a computer

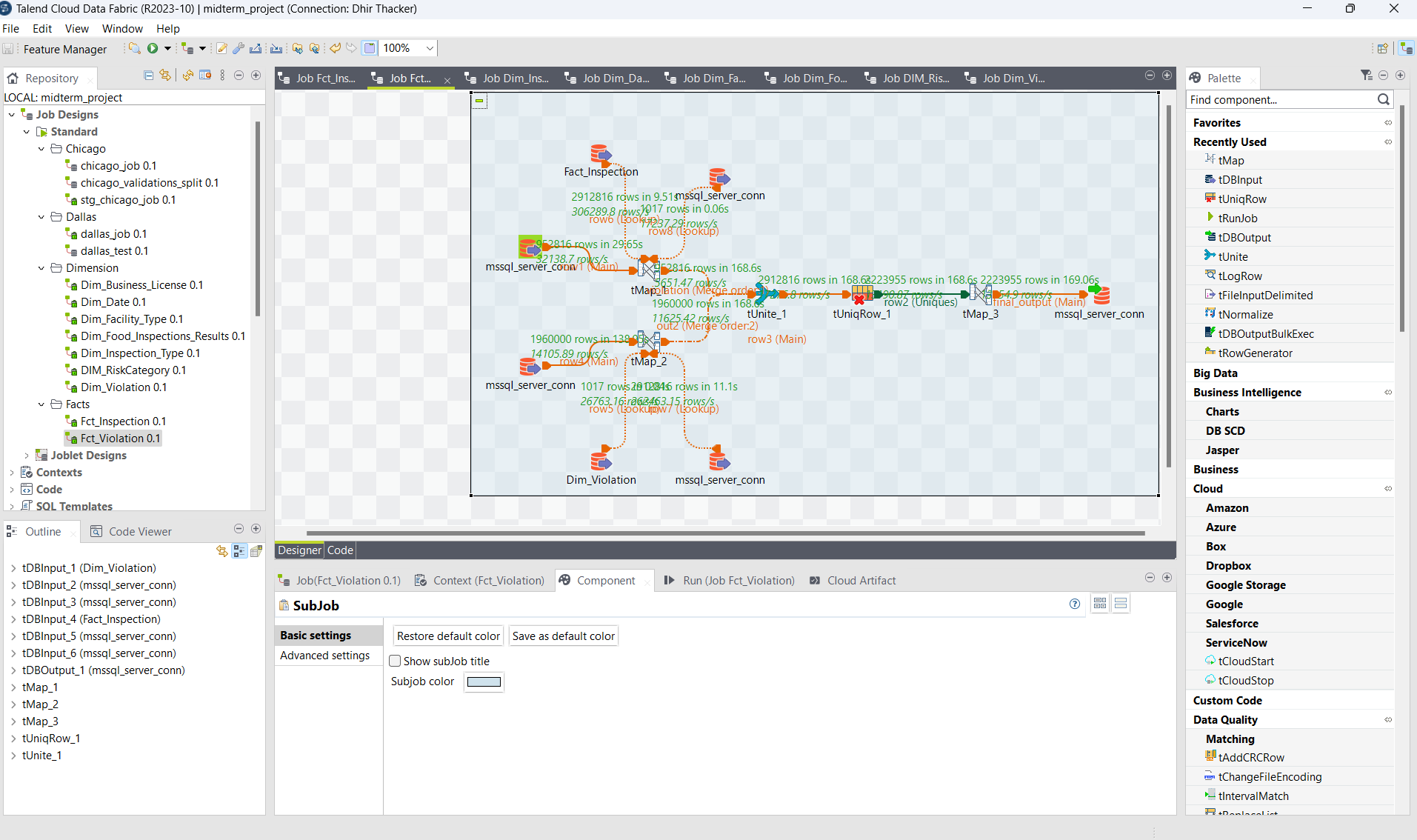
Description automatically generated

T-Map 3 of Fact\_Inspection

A screenshot of a computer

Description automatically generated

1. **Fact\_Violation**



T-Map 1 of Fact\_Violation

A screenshot of a computer

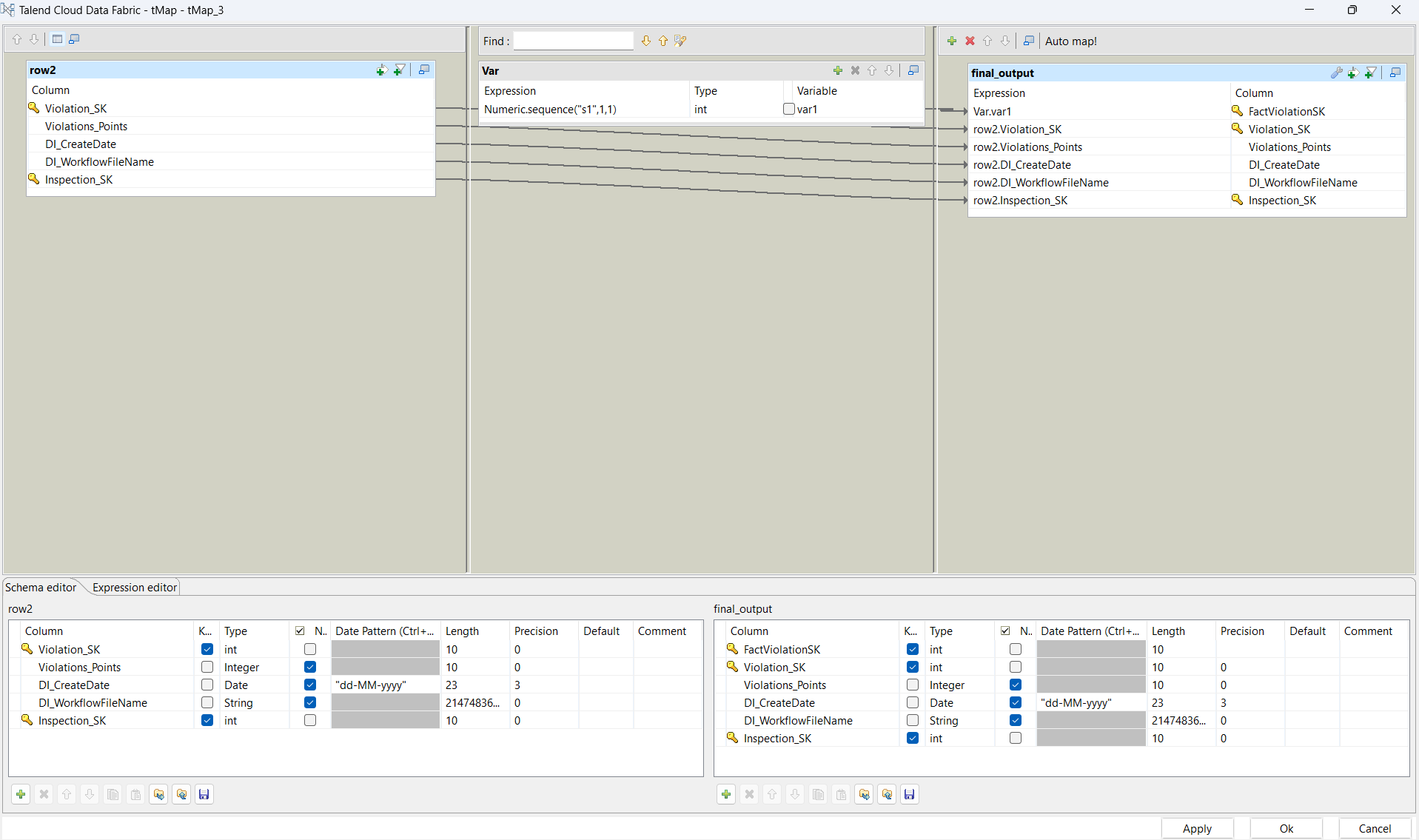
Description automatically generated

T-Map 2 of Fact\_Violation

A screenshot of a computer

Description automatically generated

T-Map 3 of Fact\_Violation



T-UniqRow of Fact\_Violation

