

Chinmay Kulkarni  
002209448

### Assignment 3 - Talend Seattle Pet License (Staging Data)

Q1. SQL create query for Stage table

Ans →

The screenshot shows the Talend Studio interface with the 'DDL for seattle\_pet\_license.stg\_spl\_output' tab selected. The DDL code is as follows:

```
1 CREATE TABLE `stg_spl_output` (  
2   `License_Issue_Date` datetime DEFAULT NULL,  
3   `License_Number` varchar(10) DEFAULT NULL,  
4   `Animal_s_Name` varchar(80) DEFAULT NULL,  
5   `Species` varchar(20) DEFAULT NULL,  
6   `Primary_Breed` varchar(80) DEFAULT NULL,  
7   `Secondary_Breed` varchar(80) DEFAULT NULL,  
8   `ZIP_Code` varchar(20) DEFAULT NULL,  
9   `DI_Create_Date` datetime DEFAULT NULL,  
10  `pid` varchar(20) DEFAULT NULL  
11 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

The left sidebar shows the 'seattle\_pet\_license' schema with the 'stg\_spl\_output' table selected. The bottom status bar indicates 'Query Completed'.

Q2. Screenshot of Stage table Schema (Schema Inspector/Table Inspector)

Ans →

### Schema Inspector (Database) –

The screenshot shows the Talend Studio interface with the 'Schema Inspector (Database)' tab selected. The table 'stg\_spl\_output' is selected, and the schema details are displayed in the main pane.

Table	Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra	Comm
stg_spl_output	License_Issue_Date	datetime		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	License_Number	varchar(10)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	Animal_s_Name	varchar(80)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	Species	varchar(20)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	Primary_Breed	varchar(80)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	Secondary_Breed	varchar(80)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	ZIP_Code	varchar(20)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	DI_Create_Date	datetime		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
stg_spl_output	pid	varchar(20)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		

The bottom status bar shows 'Count: 9' and a 'Refresh' button.

Chinmay Kulkarni  
002209448

### Table Inspector (Table) –

The screenshot shows the SQL Server Enterprise Edition Table Inspector for the 'seattle\_pet\_license' table. The left pane shows the 'seattle\_pet\_license' schema with tables 'stg\_spl\_output' and 'seattle\_pet\_license'. The main pane shows the table's columns and their properties.

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra	Comments
License_Issue_Date	datetime		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
License_Number	varchar(10)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
Animal_s_Name	varchar(80)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
Species	varchar(20)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
Primary_Breed	varchar(80)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
Secondary_Breed	varchar(80)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
ZIP_Code	varchar(20)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
DI_Create_Date	datetime		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		
pid	varchar(20)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references		

### Script Created in SSMS for schema –

The screenshot shows the SQL Server Enterprise Edition SSMS interface. The left pane shows the 'SPL\_DB' database with tables 'stg\_spl\_demo' and 'stg\_spl\_output'. The main pane shows the script to create the 'stg\_spl\_demo' table.

```
USE [SPL_DB]
GO

/***** Object: Table [dbo].[stg_spl_demo]    Script Date: 2/12/2024 4:24:39 PM *****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[stg_spl_demo](
    [License_Issue_Date] [datetime] NULL,
    [License_Number] [varchar](10) NULL,
    [Animal_s_Name] [varchar](80) NULL,
    [Species] [varchar](20) NULL,
    [Primary_Breed] [varchar](80) NULL,
    [Secondary_Breed] [varchar](80) NULL,
    [ZIP_Code] [varchar](20) NULL,
    [DI_CreateDate] [datetime] NULL,
    [DI_JobPID] [varchar](20) NULL
) ON [PRIMARY]
GO
```

Chinmay Kulkarni

002209448

Q3 Screenshot of Loaded data in stage table

Ans →

**SSMS –**

Object Explorer: Connect to DESKTOP-F691175 (SQL Server 16.0.1000.6) -> Databases -> SPL\_DB -> Tables -> dbo.stg\_spl\_demo

SQLQuery4.sql - D...DB (damg7370 (70))

```
SELECT TOP (1000) [License_Issue_Date],
[License_Number],
[Animal_s_Name],
[Species],
[Primary_Breed],
[Secondary_Breed],
[ZIP_Code],
[DL_CreateDate],
[DL_JobPID]
FROM [SPL_DB].[dbo].[stg_spl_demo]
```

Results: 98 %

License_Issue_Date	License_Number	Animal_s_Name	Species	Primary_Breed	Secondary_Breed	ZIP_Code	DL_CreateDate	DL_JobPID
2023-11-26 00:00:00.000	8163025	Ous	Cat	Domestic Medium Hair	Mix	98109	2024-02-12 14:04:00.333	URpfje
2023-11-26 00:00:00.000	8012391	Nimbus	Cat	Domestic Shorthair	Mix	98121	2024-02-12 14:04:00.333	URpfje
2023-11-26 00:00:00.000	8050611	Butlers	Cat	Colourpoint Shorthair	Mix	98122	2024-02-12 14:04:00.333	URpfje
2023-11-26 00:00:00.000	8050613	Lucius	Cat	Balinese	Mix	98122	2024-02-12 14:04:00.333	URpfje
2023-11-26 00:00:00.000	8034351	Wallace	Cat	Domestic Medium Hair	Mix	98122	2024-02-12 14:04:00.333	URpfje
2023-11-26 00:00:00.000	8050612	Mina	Cat	Domestic Shorthair	Mix	98118	2024-02-12 14:04:00.337	URpfje
2023-11-26 00:00:00.000	8144557	Veronica	Cat	Domestic Shorthair	Mix	98118	2024-02-12 14:04:00.337	URpfje
2023-11-26 00:00:00.000	8050617	Honu	Cat	Domestic Medium Hair	Mix	98115	2024-02-12 14:04:00.337	URpfje
2023-11-26 00:00:00.000	8106968	Trout	Cat	Domestic Medium Hair	Mix	98102	2024-02-12 14:04:00.337	URpfje
2023-11-26 00:00:00.000	8106969	Bartleby	Cat	Domestic Shorthair	Mix	98102	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8050618	Boo	Cat	Domestic Medium Hair	Mix	98118	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8050619	Sym	Cat	Domestic Shorthair	Mix	98118	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8034246	Whiskey	Cat	Domestic Shorthair	Mix	98101	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8034247	Tango	Cat	Domestic Shorthair	Mix	98101	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8050621	Alice	Cat	American Shorthair	Mix	98144	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8042795	Violet	Cat	Siamese	Mix	98117	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8162128	Missy	Cat	Domestic Longhair	Mix	98122	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8162129	Velma	Cat	American Shorthair	Mix	98115	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8134690	Friday the 13th	Cat	Domestic Shorthair	Mix	98136	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8032730	Patti Coey	Cat	Asian Shorthair	Mix	98105	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8032729	Albus Coey	Cat	Asian Shorthair	Mix	98105	2024-02-12 14:04:00.337	URpfje
2023-11-27 00:00:00.000	8050628	Killua	Cat	American Shorthair	Mix	98122	2024-02-12 14:04:00.340	URpfje
2023-11-27 00:00:00.000	8050631	Buddy	Cat	Domestic Medium Hair	Mix	98188	2024-02-12 14:04:00.340	URpfje

Query executed successfully. DESKTOP-F691175 (16.0 RTM) - damg7370 (70) - SPL\_DB 00:00:00 1,000 rows

**MYSQL –**

Navigator: Filter objects, sakila, seattle\_pet\_license, Tables, stg\_spl\_output, Views, Stored Procedures, Functions, sys, world

Administration: Schemas, Information

Table: stg\_spl\_output

Columns: License\_Issue\_Date, License\_Number, Animal\_s\_Name, Species, Primary\_Breed, Secondary\_Breed

Query: SELECT \* FROM seattle\_pet\_license.stg\_spl\_output;

Result Grid: 20 rows, 9 columns

License_Issue_Date	License_Number	Animal_s_Name	Species	Primary_Breed	Secondary_Breed	ZIP_Code	DL_Create_Date	pid
2015-12-18 00:00:00	5107948	Zen	Cat	Domestic Longhair	Mix	98117	2024-02-12 14:03:58	URpfje
2016-06-14 00:00:00	5116503	Misty	Cat	Siberian	Mix	98117	2024-02-12 14:03:58	URpfje
2016-08-04 00:00:00	5119301	Lyra	Cat	Mix	Mix	98121	2024-02-12 14:03:58	URpfje
2019-02-13 00:00:00	962273	Veronica	Cat	Domestic Longhair	Mix	98107	2024-02-12 14:03:58	URpfje
2019-08-10 00:00:00	5133113	Spider	Cat	LaPerm	Mix	98115	2024-02-12 14:03:58	URpfje
2019-11-21 00:00:00	8002549	Maxx	Cat	American Shorthair	Mix	98125	2024-02-12 14:03:58	URpfje
2020-01-12 00:00:00	5151906	Sammy	Cat	Domestic Shorthair	Mix	98116	2024-02-12 14:03:58	URpfje
2020-05-19 00:00:00	8015649	Bradley	Cat	Siamese	Mix	98101	2024-02-12 14:03:58	URpfje
2020-05-24 00:00:00	5142869	Mickey	Cat	Domestic Longhair	Mix	98126	2024-02-12 14:03:58	URpfje
2020-07-03 00:00:00	5112835	Diamond	Cat	Domestic Shorthair	Mix	98103	2024-02-12 14:03:58	URpfje
2020-07-16 00:00:00	83454	Dall	Cat	Domestic Shorthair	Mix	98119	2024-02-12 14:03:58	URpfje
2020-07-21 00:00:00	5131986	Nacho	Cat	Domestic Shorthair	Mix	98126	2024-02-12 14:03:58	URpfje
2020-08-18 00:00:00	8019541	Grade	Cat	Domestic Medium ...	Mix	98133	2024-02-12 14:03:58	URpfje
2020-09-08 00:00:00	5141990	Ramon	Cat	Domestic Shorthair	Mix	98116	2024-02-12 14:03:58	URpfje
2020-09-17 00:00:00	8015827	Luna	Cat	Ragdoll	Mix	98112	2024-02-12 14:03:58	URpfje
2020-11-03 00:00:00	5112996	Pixel	Cat	Domestic Medium ...	Mix	98112	2024-02-12 14:03:58	URpfje
2020-11-17 00:00:00	5146576	C.C.	Cat	Domestic Shorthair	Mix	98117	2024-02-12 14:03:58	URpfje
2020-11-24 00:00:00	77412	Gemma	Cat	Siamese	American Shorthair	98126	2024-02-12 14:03:58	URpfje
2020-12-15 00:00:00	896993	Oliver	Cat	Domestic Longhair	Mix	98199	2024-02-12 14:03:58	URpfje
2020-12-15 00:00:00	832271	Silky Frances	Cat	Domestic Shorthair	Mix	98115	2024-02-12 14:03:58	URpfje
2020-12-20 00:00:00	5114008	Jesse	Cat	Domestic Shorthair	Mix	98103	2024-02-12 14:03:58	URpfje

Output: Action Output

Message: 1 row(s) returned

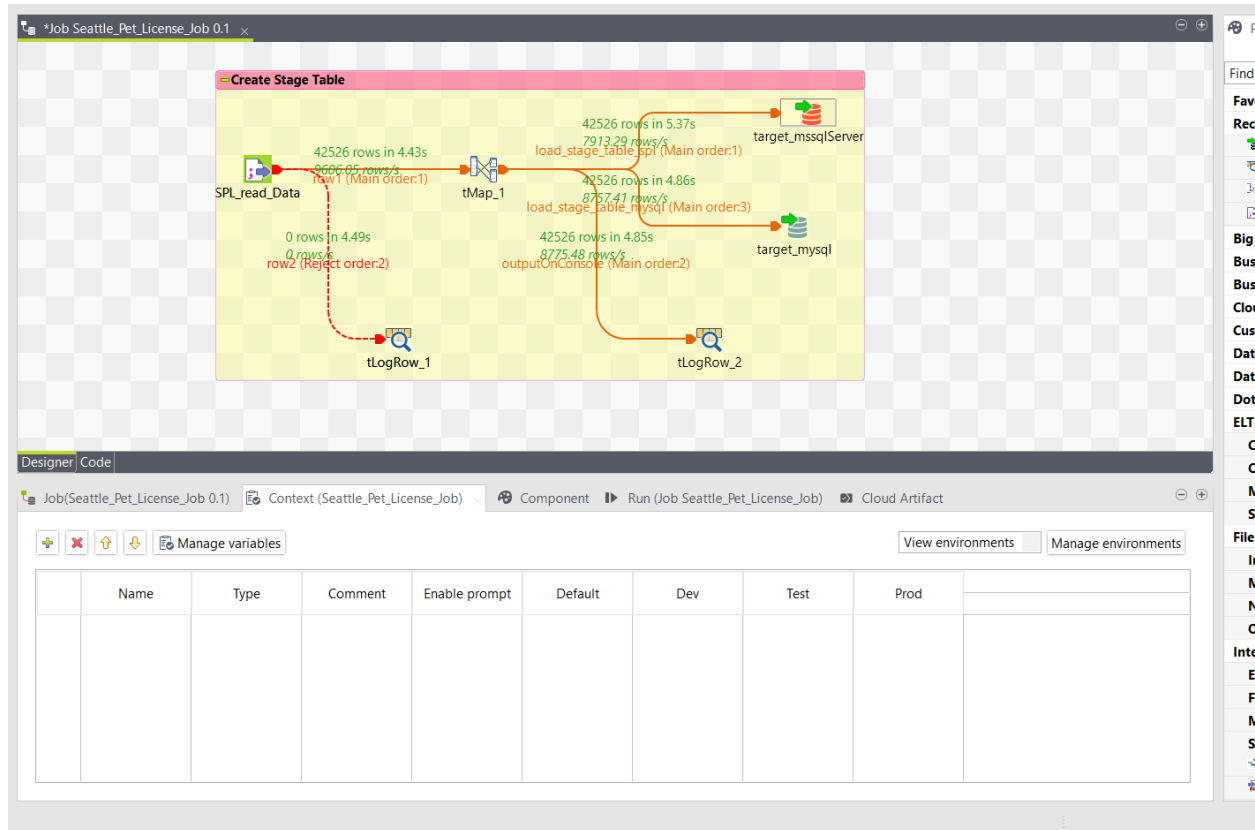
Query Completed

Chinmay Kulkarni

002209448

Q4. Screenshot of Talend workflow/job(After run with row counts)

Ans → **Talend Workflow –**



Q5. Count of Rows in the stage table

Ans → **Row Count →**

**MYSQL →**

The screenshot shows a Talend SQL script in the "SQL Editor" view. The script is as follows:

```
1 -- CREATE DATABASE Seattle_Pet_License;
2 USE seattle_pet_license;
3 SELECT COUNT(*) FROM stg_spl_output;
4
```

The "Result Grid" shows the output of the SQL query:

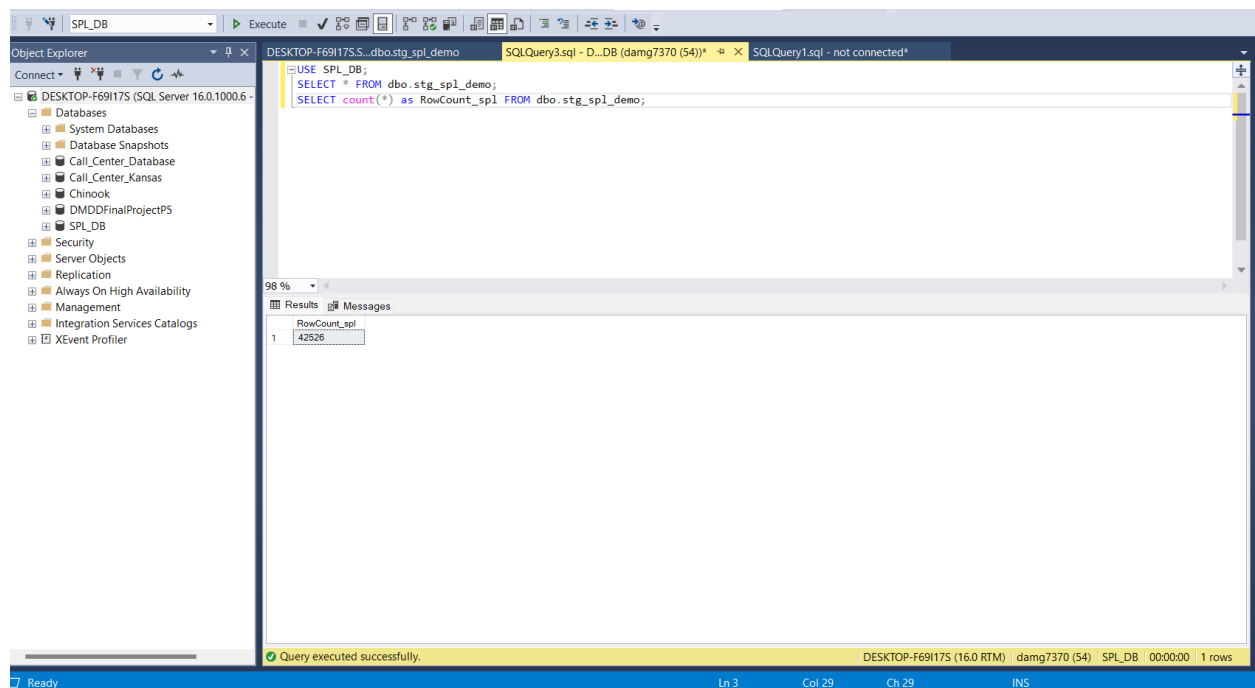
COUNT(*)
42526

The "Output" view shows the execution log:

#	Time	Action	Message
14	14:07:59	USE seattle_pet_license	0 row(s) affected
15	14:08:01	SELECT COUNT(*) FROM stg_spl_output	1 row(s) returned
16	14:08:15	SELECT COUNT(*) FROM stg_spl_output LIMIT 0, 1000	1 row(s) returned
17	14:10:04	ANALYZE TABLE 'seattle_pet_license`.`stg_spl_output'	OK
18	15:03:53	ANALYZE TABLE 'seattle_pet_license`.`stg_spl_output'	OK
19	15:04:02	ANALYZE TABLE 'seattle_pet_license`.`stg_spl_output'	OK

Chinmay Kulkarni  
002209448

### SSMS –

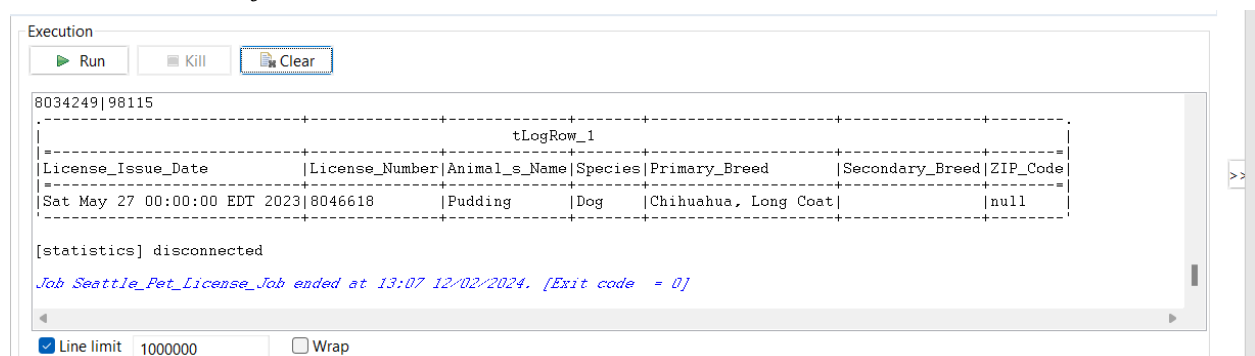


### Handling Reject →

- The message "Couldn't parse value for column Zip\_Code 'V5J1P8'" appears in a reject row.
- This is because the Zip\_Code column's datatype is an integer, and the value not parsed combines integer and string.
- To fix this, go into the Map Tool's properties and change the Zip\_Code column's datatype from integer to string.

Below is a screenshot following the datatype change - (and hence we get 42526 rows else with the reject row we get 42525 rows ).

### A row that is been rejected →



### After resolving the error →

Chinmay Kulkarni  
002209448

