

Chinmay Kulkarni  
002209448

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

Q1. SQL create query for Stage table

Ans →

The screenshot shows the MySQL Workbench interface. In the top navigation bar, the schema is set to 'seattle\_pet\_license'. The 'Tables' tab is selected, and a new table named 'stg\_spl\_output' is being created. The DDL code for the table is displayed in the main pane:

```
CREATE TABLE `stg_spl_output` (
  `License_Issue_Date` datetime DEFAULT NULL,
  `License_Number` varchar(10) DEFAULT NULL,
  `Animal_s_Name` varchar(80) DEFAULT NULL,
  `Species` varchar(20) DEFAULT NULL,
  `Primary_Breed` varchar(80) DEFAULT NULL,
  `Secondary_Breed` varchar(80) DEFAULT NULL,
  `ZIP_Code` varchar(20) DEFAULT NULL,
  `DI_Create_Date` datetime DEFAULT NULL,
  `pid` varchar(20) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

The 'Output' pane at the bottom shows a single log entry indicating the table was successfully created:

#	Time	Action
15	14-08-01	SELECT COUNT(*) FROM stg_spl_output

Message: 1 rows(s) returned.

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

Ans →

**Schema Inspector (Database)**

The screenshot shows the MySQL Workbench Schema Inspector. The schema is set to 'seattle\_pet\_license'. The 'Tables' tab is selected, and the 'stg\_spl\_output' table is shown with its columns and their properties:

Table	Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra	Comment
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		

Chinmay Kulkarni  
002209448

### Table Inspector (Table) –

The screenshot shows the Table Inspector interface for the 'seattle\_pet\_license' schema. The left sidebar lists 'Schemas' (sakila, seattle\_pet\_license, sys, world) and 'Tables' (stg\_spl\_output). The main pane displays the 'Columns' tab for the 'seattle\_pet\_license' table, which has 9 columns:

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 Object Explorer and SQL Query Editor in SSMS. The Object Explorer shows the database structure, including the 'stg\_spl\_demo' table under the 'Tables' node. The SQL Query Editor contains the T-SQL script for creating 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 –

The screenshot shows the Object Explorer on the left with the connection to DESKTOP-F69117S (SQL Server 16.0.1000.6). The Results pane displays the output of the following SQL query:

```
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]
```

The results show 1,000 rows of data from the stg\_spl\_demo table, including columns: License\_Issue\_Date, License\_Number, Animal\_s\_Name, Species, Primary\_Breed, Secondary\_Breed, ZIP\_Code, DL\_CreateDate, and DL\_JobPID. The data includes various animal names like Gus, Nimbus, Butters, Wallace, etc., with their respective breeds and details.

#### MySQL –

The screenshot shows the Navigator on the left with the database seattle\_pet\_license selected. The Results Grid pane displays the output of the following SQL query:

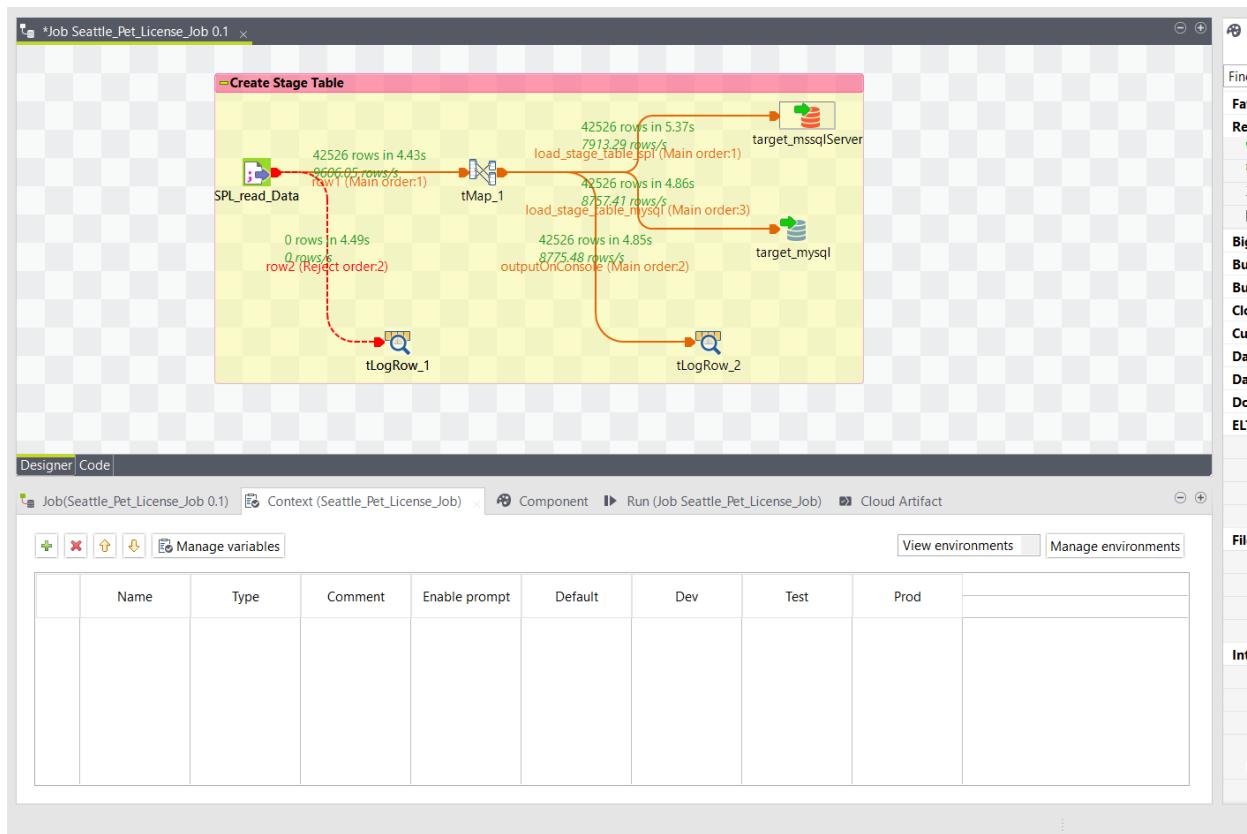
```
1 • SELECT * FROM seattle_pet_license.stg_spl_output;
```

The results show data from the stg\_spl\_output table, including columns: License\_Issue\_Date, License\_Number, Animal\_s\_Name, Species, Primary\_Breed, Secondary\_Breed, ZIP\_Code, DI\_Create\_Date, and pid. The data includes various animal names like Zen, Misty, Lyra, Veronica, Spider, etc., with their respective breeds and details.

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 the MySQL Workbench interface. The Navigator pane displays the schema "seattle\_pet\_license" and its tables, including "stg\_spl\_output". The SQL editor pane contains the following SQL code:

```
-- CREATE DATABASE Seattle_Pet_License;
USE seattle_pet_license;
SELECT COUNT(*) FROM stg_spl_output;
```

The Result Grid pane shows the output of the query:

COUNT(*)
42526

The Output pane shows the history of actions:

#	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 -

The screenshot shows the SSMS interface with the following details:

- Object Explorer:** Shows the database structure under "DESKTOP-F69I17S (SQL Server 16.0.1000.6)". The "SPL\_DB" database is selected.
- SQL Query Editor:** Contains the following T-SQL code:

```
USE SPL_DB;
SELECT * FROM dbo.stg_spl_demo;
SELECT count(*) as RowCount_spl FROM dbo.stg_spl_demo;
```
- Results Grid:** Displays the output of the last query:

RowCount_spl
1
42526
- Status Bar:** Shows "Query executed successfully." and "DESKTOP-F69I17S (16.0 RTM) | damg7370 (54) | SPL\_DB | 00:00:00 | 1 rows".

### **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 →

Execution

8034249|98115

tLogRow_1					
License_Issue_Date	License_Number	Animal_s_Name	Species	Primary_Breed	Secondary_Breed ZIP_Code
Sat May 27 00:00:00 EDT 2023	8046618	Pudding	Dog	Chihuahua, Long Coat	null

[statistics] disconnected

Job Seattle\_Pet\_License\_Job ended at 13:07 12/02/2024. [Exit code = 0]

Line limit 1000000  Wrap

### After resolving the error →

Chinmay Kulkarni  
002209448

