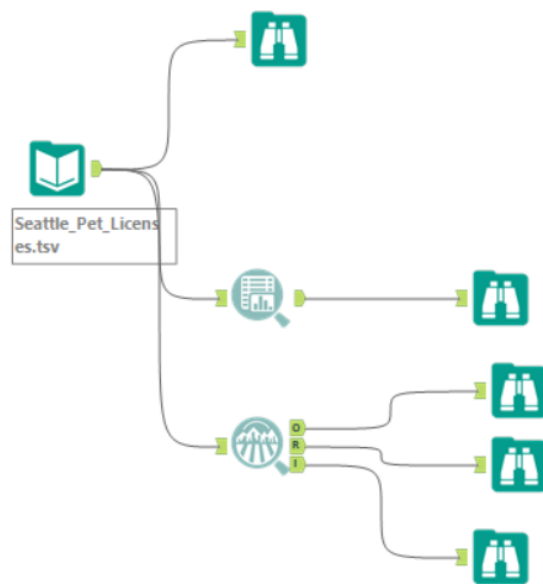
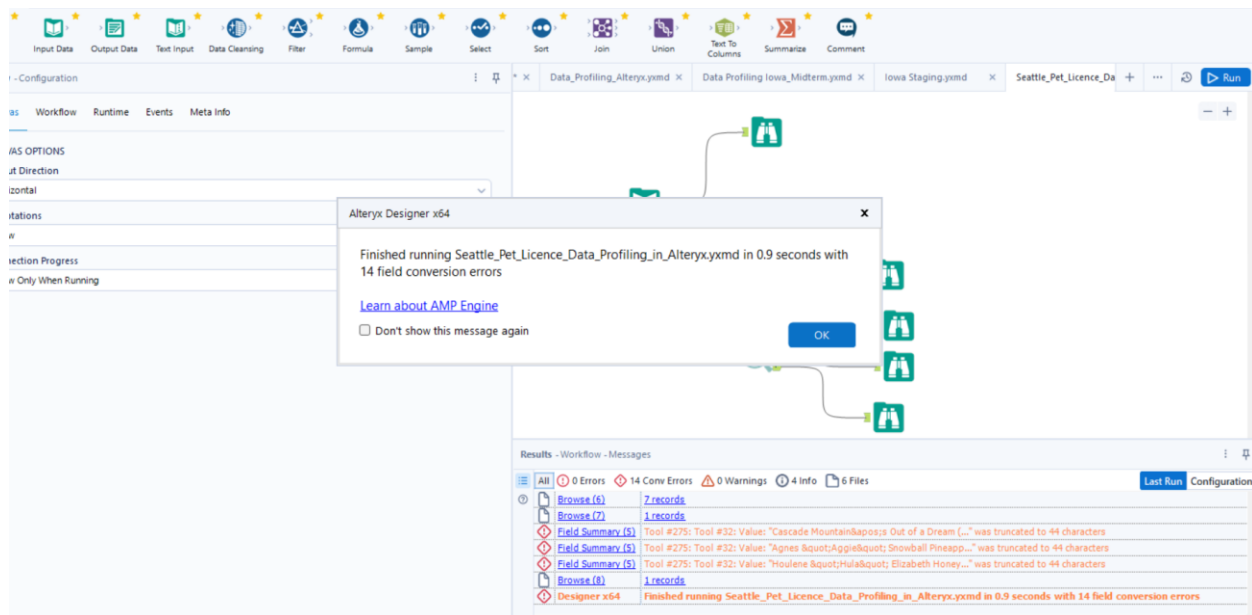


Seattle Pet License dimensional modeling and Loading via Talend - Workshop implementation

--By AMEY PARANGE

Alteryx Profiling workflow screenshot:





Observation:

Name	Datatype	% missing	Unique values	Null values	Observation	Min Value	Max Value
License Issue Date	V_String	0	931	0.00%	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	11	17
License Number	V_String	0.0%	42192	0	Some values of this field have a small number of	4	15

					value counts. If Appropriate, consider combining some value levels together.		
Animal's Name	V_String	0.07%	12696	0	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	1	80
Species	V_String	0.0%	4	0	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	3	4
Primary Breed	V_String	0.0%	237	0	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	3	46
Secondary Breed	V_String	36.0%	259	0	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate,	3	46

					consider combining some value levels together.		
Zip Code	V_String	0.3%	208	0	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	3	5

Talend Implementation

wf_seattlepetsrc_to_stage_load

The image shows the Talend Studio interface for a job named `wf_seattlepetsrc_to_stage_load`. The job design is as follows:

- src_seattlepetlicense** (Source) → **tMap_1** (Transformer) → **tgt_seattlepetlicensestg** (Target) and **tgt_seattlepetlicensestg_azure** (Target).
- Statistics for **tMap_1**: 43086 rows in 5.07s (Main order:1), 8491.53 rows/s.
- Statistics for **tgt_seattlepetlicensestg**: 43086 rows in 6.6s (Main order:1), 6527.18 rows/s.
- Statistics for **tgt_seattlepetlicensestg_azure**: 43086 rows in 5.7s (Main order:2), 7556.3 rows/s.

The execution console shows the following log:

```
Starting job wf_seattlepetsrc_to_stage_load at 02:53 24/10/2023.
[statistics] connecting to socket on port 3917
[statistics] connected
[statistics] disconnected
Job wf_seattlepetsrc_to_stage_load ended at 02:54 24/10/2023. [Exit code = 0]
```

On the right, a list of variables is shown:

Name
src_seattle_pet...
src_seattle_pet...
src_seattle_pet...
src_seattle_pet...
src_seattle_pet...
seattlepetDBC...
seattlepetDBC...
seattlepetDBC...

Azure output:

The image shows the Azure Data Studio interface with a SQL query executed against the `dadabi` database:

```
select * from seattlepetlicense.seattlepetlicensestg;
```

The results are displayed in a table with 10 rows and 10 columns:

	License_Issue_Date	License_Number	Animal_s_Name	Species	Primary_Breed	Secondary_Breed	ZIP_Code	ProcessID	DI_created_date
1	2015-12-18	5107948	Zen	Cat	Domestic Longhair	Mix	98117	11vWEk	2023-10-24 02:53:5
2	2016-06-14	5116503	Misty	Cat	Siberian		98117	11vWEk	2023-10-24 02:53:5
3	2016-08-04	5119301	Lyna	Cat	Mix		98121	11vWEk	2023-10-24 02:53:5
4	2019-01-27	8005097	Jolene	Cat	Maine Coon	Mix	98133	11vWEk	2023-10-24 02:53:5
5	2019-02-13	962273	Veronica	Cat	Domestic Longhair		98107	11vWEk	2023-10-24 02:53:5
6	2019-06-01	208746	Sweetheart	Cat	Domestic Medium Hair		98116	11vWEk	2023-10-24 02:53:5
7	2019-06-06	79347	Mr Darcy	Cat	Domestic Shorthair		98103	11vWEk	2023-10-24 02:53:5
8	2019-06-25	8007918	Kali	Cat	Domestic Shorthair		98133	11vWEk	2023-10-24 02:53:5
9	2019-07-04	209285	Daisy	Cat	Domestic Shorthair		98117	11vWEk	2023-10-24 02:53:5
10	2019-07-30	5132137	Ada	Cat	American Curl	Mix	98115	11vWEk	2023-10-24 02:53:5

```
2
3 select COUNT(*) from seattlepetlicense.seattlepetlicensestg;
4
```

Results	Messages
	(No column name) ▼
1	43086

MySQL

```

1 • select * from seattlepetlicensestg;
2 • select COUNT(*) from seattlepetlicensestg;
3

```

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

Fetch rows:

	License_Issue_Date	License_Number	Animal_s_Name	Species	Primary_Breed	Secondary_Breed	ZIP_Code	ProcessID	DI_created_date
▶	2015-12-18	S107948	Zen	Cat	Domestic Longhair	Mix	98117	1ivWEK	2023-10-24 02:53:58
	2016-06-14	S116503	Misty	Cat	Siberian		98117	1ivWEK	2023-10-24 02:53:58
	2016-08-04	S119301	Lyra	Cat	Mix		98121	1ivWEK	2023-10-24 02:53:58
	2019-01-27	8005097	Jolene	Cat	Maine Coon	Mix	98133	1ivWEK	2023-10-24 02:53:58
	2019-02-13	962273	Veronica	Cat	Domestic Longhair		98107	1ivWEK	2023-10-24 02:53:58
	2019-06-01	208746	Sweetheart	Cat	Domestic Medium Hair		98116	1ivWEK	2023-10-24 02:53:58
	2019-06-06	79347	Mr Darcy	Cat	Domestic Shorthair		98103	1ivWEK	2023-10-24 02:53:58
	2019-06-25	8007918	Kali	Cat	Domestic Shorthair		98133	1ivWEK	2023-10-24 02:53:58
	2019-07-04	209285	Daisy	Cat	Domestic Shorthair		98117	1ivWEK	2023-10-24 02:53:58
	2019-07-30	S132137	Ada	Cat	American Curl	Mix	98115	1ivWEK	2023-10-24 02:53:58
	2019-08-10	S133113	Spider	Cat	LaPerm		98115	1ivWEK	2023-10-24 02:53:58
	2019-08-16	S130442	Fog	Cat	Domestic Medium Hair	Mix	98106	1ivWEK	2023-10-24 02:53:58

seattlepetlicensestg 7 ×

```
1 select * from seattlepetlicensetstg;
2 select COUNT(*) from seattlepetlicensetstg;
3
```

Result Grid

COUNT(*)
43086

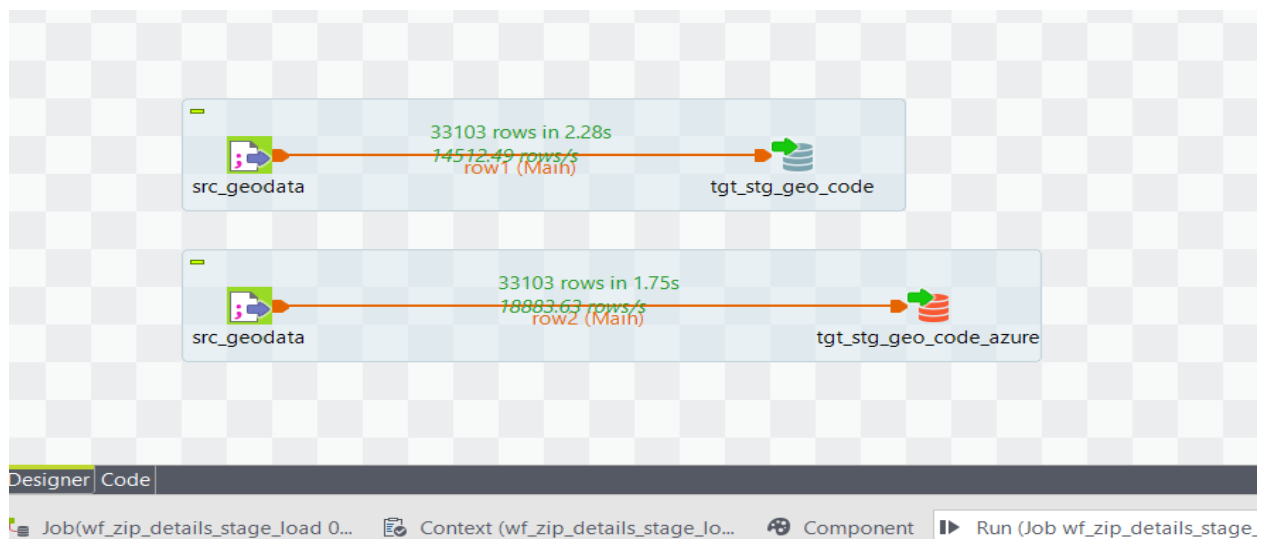
Result 8 x Read Only Context Help Sn

Output

Action Output

#	Time	Action	Message
345	03:00:37	select * from seattlepetlicensetstg LIMIT 0, 2000	2000 row(s) returned
346	03:01:12	select COUNT(*) from seattlepetlicensetstg LIMIT 0, 2000	1 row(s) returned

wf_zip_details_stage_load



Job wf_zip_details_stage_load

Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run

Execution

Run Kill Clear

```
[WARN ] 02:59:35 mynewproject.wf_zip_details_stage_load_0_1.wf_zip_details_stage_load- Null value
[WARN ] 02:59:35 mynewproject.wf_zip_details_stage_load_0_1.wf_zip_details_stage_load- Null value
[statistics] connecting to socket on port 3887
[statistics] connected
[statistics] disconnected

Job wf_zip_details_stage_load ended at 02:59 24/10/2023. [Exit code = 0]
```

wf_load_dim_date

The workflow diagram shows the following components and data flow:

- tRowGenerator_1**: Generates 10000 rows in 1.6s at 6238.3 rows/s.
- tMap_1**: Splits the data into two paths.
- load_dim_date (Main order:1)**: Loads 10000 rows in 3.09s at 3236.25 rows/s to **tgt_dim_date_mysql**.
- load_dim_date_azure (Main order:2)**: Loads 10000 rows in 2.27s at 4411.12 rows/s to **tgt_dim_date_azure**.

Below the diagram is the **Execution** tab with the following log:

```
Starting job wf_load_dim_date at 03:04 24/10/2023.  
[statistics] connecting to socket on port 3423  
[statistics] connected  
[statistics] disconnected  
  
Job wf_load_dim_date ended at 03:04 24/10/2023. [Exit code = 0]
```

Azure:

```
1  
2 select COUNT(*) from seattlepetlicense.dim_date;  
3 select * from seattlepetlicense.dim_date;  
4  
5
```

Results Messages

	(No column name)
1	10000

	DateSK	Date	Date_Num	Month_Num	Qtr_Num	Year_Num	Date_Str	Month_Str	Qtr_Str	Year_Str	Is_Weekend	Date_To_Warehou
1	2	2000-01-02	02	01	1	2000	Sunday	January	Q1	2000	Yes	2023-10-24 0
2	4	2000-01-04	04	01	1	2000	Tuesday	January	Q1	2000	No	2023-10-24 0
3	6	2000-01-06	06	01	1	2000	Thursday	January	Q1	2000	No	2023-10-24 0
4	8	2000-01-08	08	01	1	2000	Saturday	January	Q1	2000	Yes	2023-10-24 0
5	10	2000-01-10	10	01	1	2000	Monday	January	Q1	2000	No	2023-10-24 0
6	12	2000-01-12	12	01	1	2000	Wednesday	January	Q1	2000	No	2023-10-24 0
7	14	2000-01-14	14	01	1	2000	Friday	January	Q1	2000	No	2023-10-24 0

PROBLEMS 7 OUTPUT TERMINAL TASKS Tasks

MySQL

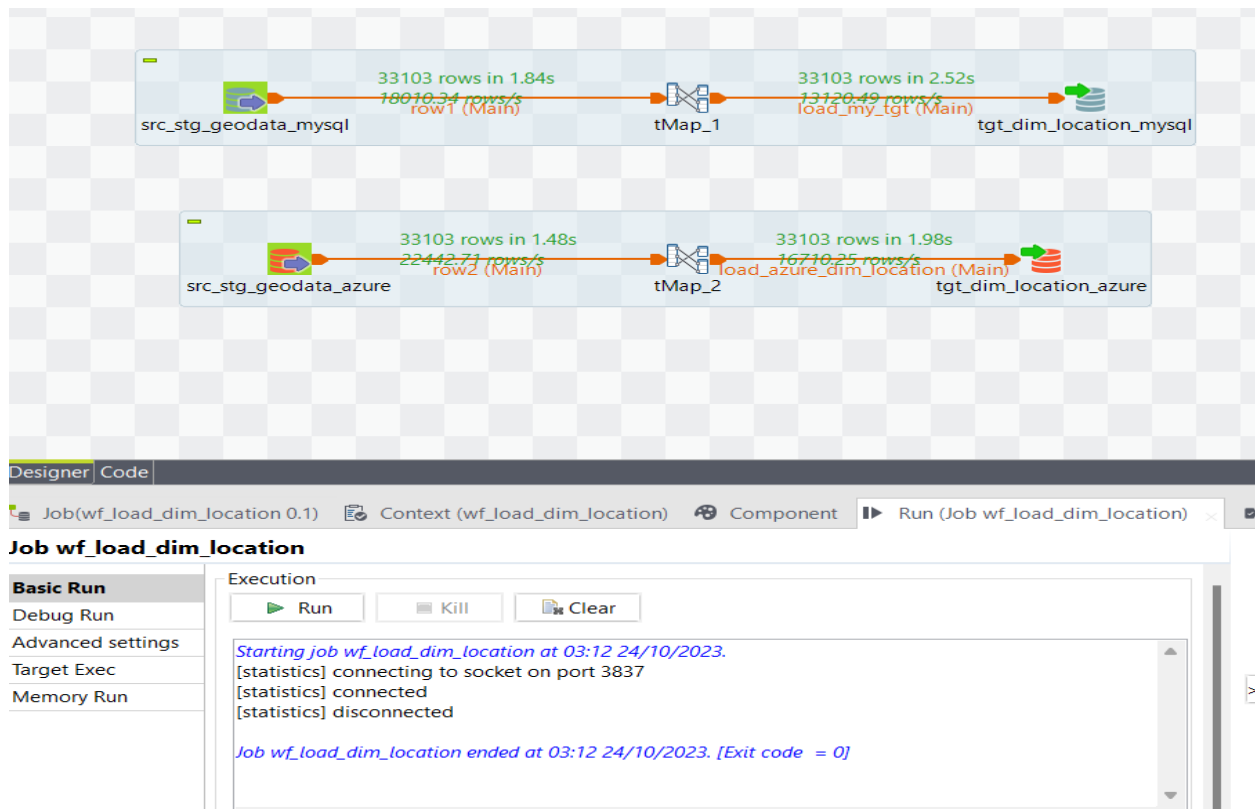
```
1 • select * from dim_date;
2 • select COUNT(*) from dim_date;
3
```

Result Grid												
Filter Rows:												
Edit: Export/Import: Wrap Cell Content: Fetch rows:												
DateSK	Date	Date_Num	Month_Num	Qtr_Num	Year_Num	Date_Str	Month_Str	Qtr_Str	Year_Str	Is_Weekend	Date_To_Warehouse	DI_JobID
2	2000-01-02	02	01	1	2000	Sunday	January	Q1	2000	Yes	2023-10-24 03:04:54	7COCXX
4	2000-01-04	04	01	1	2000	Tuesday	January	Q1	2000	No	2023-10-24 03:04:54	7COCXX
6	2000-01-06	06	01	1	2000	Thursday	January	Q1	2000	No	2023-10-24 03:04:54	7COCXX
8	2000-01-08	08	01	1	2000	Saturday	January	Q1	2000	Yes	2023-10-24 03:04:54	7COCXX
10	2000-01-10	10	01	1	2000	Monday	January	Q1	2000	No	2023-10-24 03:04:54	7COCXX
12	2000-01-12	12	01	1	2000	Wednesday	January	Q1	2000	No	2023-10-24 03:04:54	7COCXX
14	2000-01-14	14	01	1	2000	Friday	January	Q1	2000	No	2023-10-24 03:04:54	7COCXX
16	2000-01-16	16	01	1	2000	Sunday	January	Q1	2000	Yes	2023-10-24 03:04:54	7COCXX
18	2000-01-18	18	01	1	2000	Tuesday	January	Q1	2000	No	2023-10-24 03:04:54	7COCXX

```
1 • select * from dim_date;
2 • select COUNT(*) from dim_date;
3
```

Result Grid	
Filter Rows:	
Export: Wrap Cell Content:	
COUNT(*)	
10000	

wf_load_dim_location



Azure

SQLQuery_1 - disconnected

SQLQuery_3 - dadabi...iehard)

SQLQuery_2 - dadabi...iehard) 3

SQL

Run Cancel Disconnect Change Database: dadabi Estimated Plan Enable Actual Plan Par

1

2 select COUNT(*) from seattlepetlicense.dim_location;

3 select * from seattlepetlicense.dim_location;

4

5

Results Messages

(No column name) v

1 33103

	LocationSK v	ZIP_Code v	City v	State v	Date_To_warehouse v	DI_JobId v
1	33104	35004	Acmar	Alabama	2023-10-24 03:12:28.580	5JTuk0
2	33105	35005	Adamsville	Alabama	2023-10-24 03:12:28.580	5JTuk0
3	33106	35006	Adger	Alabama	2023-10-24 03:12:28.580	5JTuk0
4	33107	35007	Keystone	Alabama	2023-10-24 03:12:28.580	5JTuk0
5	33108	35010	New site	Alabama	2023-10-24 03:12:28.580	5JTuk0
6	33109	35014	Alpine	Alabama	2023-10-24 03:12:28.580	5JTuk0
7	33110	35016	Arab	Alabama	2023-10-24 03:12:28.580	5JTuk0
8	33111	35018	Birmingham	Alabama	2023-10-24 03:12:28.580	5JTuk0

MySQL

2 • `select COUNT(*) from dim_location;`

3

Result Grid			Filter Rows:	<input type="text"/>	Export:		Wrap Cell
COUNT(*)							
33103							

1 • `select * from dim_location;`

2 • `select COUNT(*) from dim_location;`

3

Result Grid

Filter Rows:

Edit:

Export/Import:

Wrap Cell Content:

Fetch rows:

	LocationSK	ZIP_Code	City	State	Date_To_warehouse	DI_JobId
▶	1	35004	Acmar	Alabama	2023-10-24 03:12:26	5JTuk0
	2	35005	Adamsville	Alabama	2023-10-24 03:12:26	5JTuk0
	3	35006	Adger	Alabama	2023-10-24 03:12:26	5JTuk0
	4	35007	Keystone	Alabama	2023-10-24 03:12:26	5JTuk0
	5	35010	New site	Alabama	2023-10-24 03:12:26	5JTuk0
	6	35014	Alpine	Alabama	2023-10-24 03:12:26	5JTuk0
	7	35016	Arab	Alabama	2023-10-24 03:12:26	5JTuk0
	8	35019	Baileyton	Alabama	2023-10-24 03:12:26	5JTuk0
	9	35020	Bessemer	Alabama	2023-10-24 03:12:26	5JTuk0
	10	35022	Zcta 35022	Alabama	2023-10-24 03:12:26	5JTuk0
	11	35023	Hueytown	Alabama	2023-10-24 03:12:26	5JTuk0
	12	35031	Blountsville	Alabama	2023-10-24 03:12:26	5JTuk0

dim_location 12

×

Output

wf_load_dim_breed



```
1
2 select COUNT(*) from seattlepetlicense.dim_breed;
3 select * from seattlepetlicense.dim_breed;
4
5
```

Results Messages

(No column name) ▾
3351

	BreedSK ▾	Species_Name ▾	Primary_Breed_Name ▾	Secondary_Breed_Name ▾	DI_JobId ▾	Date_To_Warehouse ▾
	3352	Dog	Bulldog, French	Spaniel, Cavalier King Charles	JJ3oFe	2023-10-24 03:18:28.330
	3353	Dog	Spaniel, Tibetan		JJ3oFe	2023-10-24 03:18:28.330
	3354	Dog	Akita	Collie, Smooth	JJ3oFe	2023-10-24 03:18:28.330
	3355	Dog	Lhasa Apso	Poodle, Toy	JJ3oFe	2023-10-24 03:18:28.330
	3356	Dog	Retriever	Poodle, Toy	JJ3oFe	2023-10-24 03:18:28.330
	3357	Dog	Spaniel, Cavalier King Charles	Dachshund, Standard Long Haired	JJ3oFe	2023-10-24 03:18:28.330
	3358	Dog	Anatolian Shepherd	Siberian Husky	JJ3oFe	2023-10-24 03:18:28.330
	3359	Dog	Border Collie	Border Collie	JJ3oFe	2023-10-24 03:18:28.330

- 1 • `select * from dim_breed;`
- 2 • `select COUNT(*) from dim_breed;`
- 3

Result Grid						
Filter Rows:						
Edit:						
Export/Import:						
Wrap Cell Content:						
Fet						
BreedSK	Species_Name	Primary_Breed_Name	Secondary_Breed_Name	DI_JobId	Date_To_Warehouse	
1	Cat	Domestic Longhair	Mix	JJ3oFe	2023-10-24 03:18:27	
2	Cat	Siberian		JJ3oFe	2023-10-24 03:18:27	
3	Cat	Mix		JJ3oFe	2023-10-24 03:18:27	
4	Cat	Maine Coon	Mix	JJ3oFe	2023-10-24 03:18:27	
5	Cat	Domestic Longhair		JJ3oFe	2023-10-24 03:18:27	
6	Cat	Domestic Medium Hair		JJ3oFe	2023-10-24 03:18:27	
7	Cat	Domestic Shorthair		JJ3oFe	2023-10-24 03:18:27	
8	Cat	American Curl	Mix	JJ3oFe	2023-10-24 03:18:27	
9	Cat	LaPerm		JJ3oFe	2023-10-24 03:18:27	
10	Cat	Domestic Medium Hair	Mix	JJ3oFe	2023-10-24 03:18:27	
11	Cat	American Shorthair		JJ3oFe	2023-10-24 03:18:27	
12	Cat	Ragdoll		JJ3oFe	2023-10-24 03:18:27	

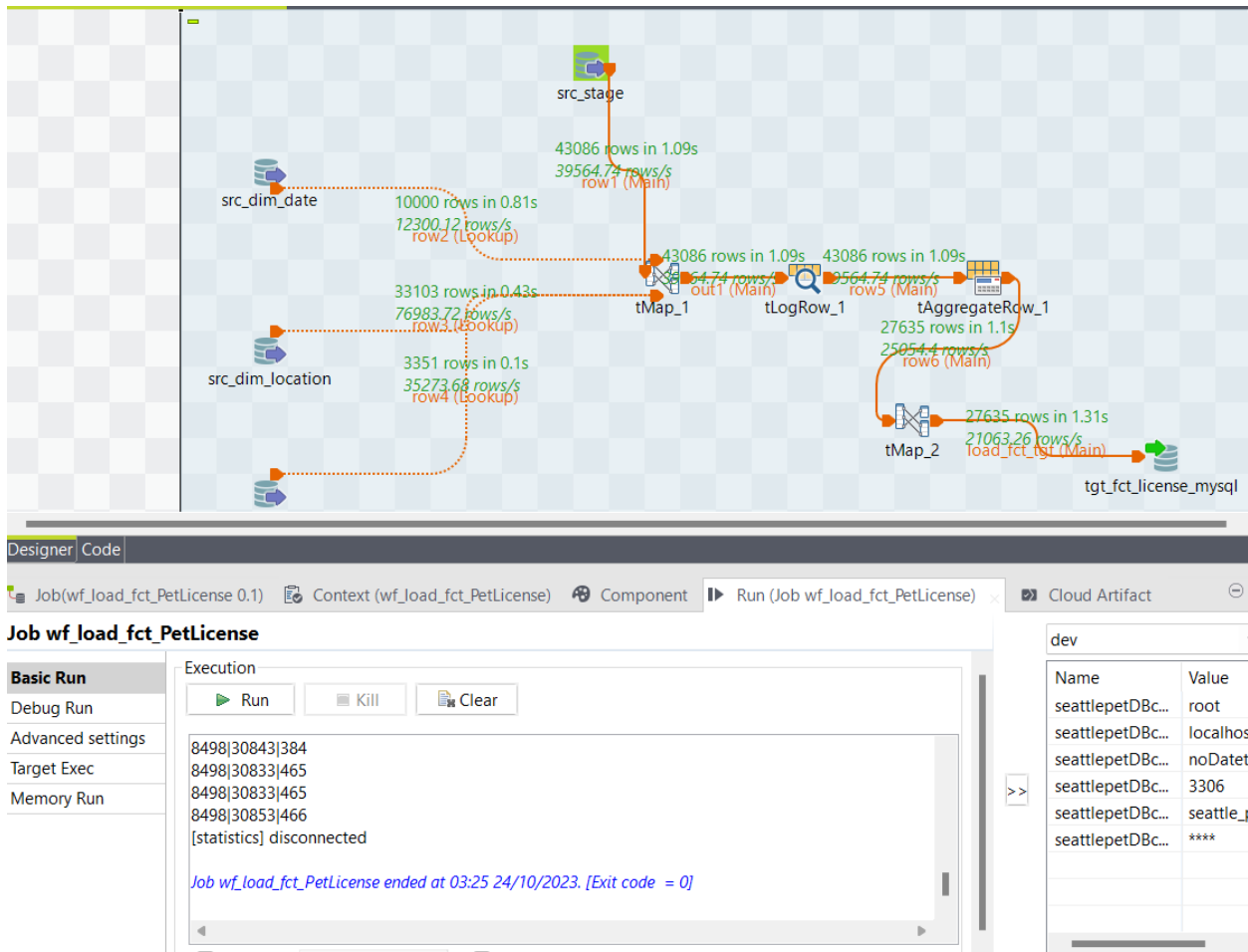
im_breed 13 ×

Output

- 2 • `select COUNT(*) from dim_breed;`
- 3

Result Grid	
Filter Rows:	
Export:	
Wrap	
	COUNT(*)
▶	3351

wf_load_fct_PetLicense







MySQL

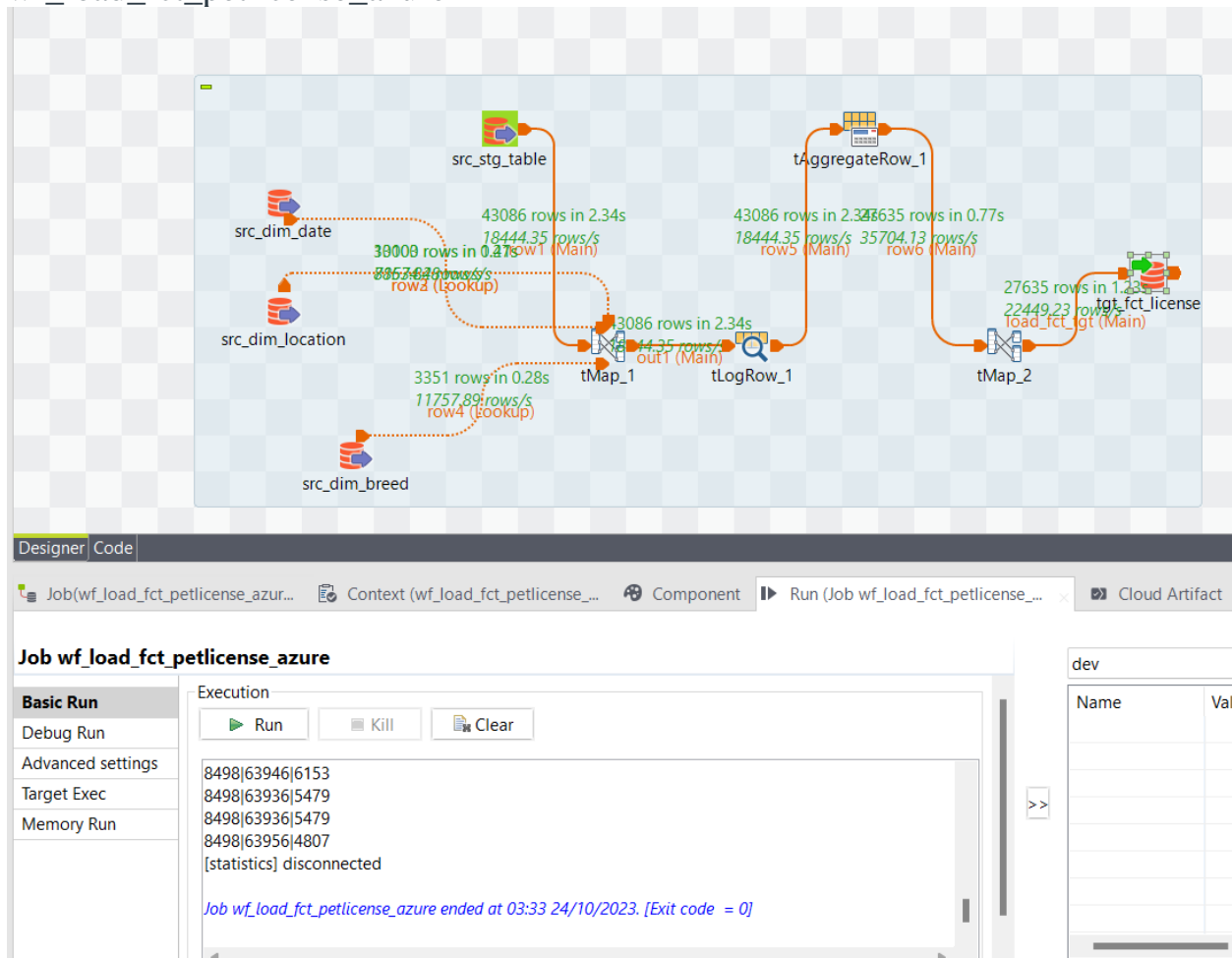
```
1 • select * from fct_license;  
2 • select COUNT(*) from fct_license;  
3
```

Result Grid						
Filter Rows:						
Edit: Export/Import: Wrap Cell						
LicensesSK	LocationSK	BreedSK	DateSK	No_of_License	Date_To_Warehouse	DI_JobID
1	30834	1854	7838	1	2023-10-24 03:25:49	WffFUNs
2	30831	400	7976	1	2023-10-24 03:25:49	WffFUNs
3	30840	312	8044	1	2023-10-24 03:25:49	WffFUNs
4	30842	1020	8316	1	2023-10-24 03:25:49	WffFUNs
5	30843	601	8180	1	2023-10-24 03:25:49	WffFUNs
6	30840	338	8044	1	2023-10-24 03:25:49	WffFUNs
7	30841	3174	8382	1	2023-10-24 03:25:49	WffFUNs
8	30851	7	8044	1	2023-10-24 03:25:49	WffFUNs
9	30831	432	7976	1	2023-10-24 03:25:49	WffFUNs
10	30846	534	8180	1	2023-10-24 03:25:49	WffFUNs
11	30842	670	8180	1	2023-10-24 03:25:49	WffFUNs
12	30833	379	7976	1	2023-10-24 03:25:49	WffFUNs


```
1 • select * from fct_license;  
2 • select COUNT(*) from fct_license;  
3
```

Result Grid			 Filter Rows: <input type="text"/>	Export: 	Wrap Cell Content: 
	COUNT(*)				
▶	27635				

wf_load_fct_petlicense_azure



AzureDB

```
1
2 select COUNT(*) from seattlepetlicense.Fct_License;
3 select * from seattlepetlicense.Fct_License;
4
5
```

Results Messages

	(No column name) ▾
1	27635

	LicensesSK ▾	LocationSK ▾	BreedSK ▾	DateSK ▾	No_of_License ▾	Date_To_Warehouse ▾	DI_JobID ▾
1	1	63952	3580	8200	1	2023-10-24 03:33:10.377	3b0ftw
2	2	63964	4940	8130	1	2023-10-24 03:33:10.377	3b0ftw
3	3	63945	3982	8268	1	2023-10-24 03:33:10.377	3b0ftw
4	4	63945	4549	8472	1	2023-10-24 03:33:10.377	3b0ftw
5	5	63940	4942	7858	1	2023-10-24 03:33:10.377	3b0ftw
6	6	63940	4940	7858	2	2023-10-24 03:33:10.377	3b0ftw
7	7	63950	5573	8198	1	2023-10-24 03:33:10.377	3b0ftw
8	8	63940	4705	8472	1	2023-10-24 03:33:10.377	3b0ftw

PROBLEMS 7 OUTPUT TERMINAL TASKS