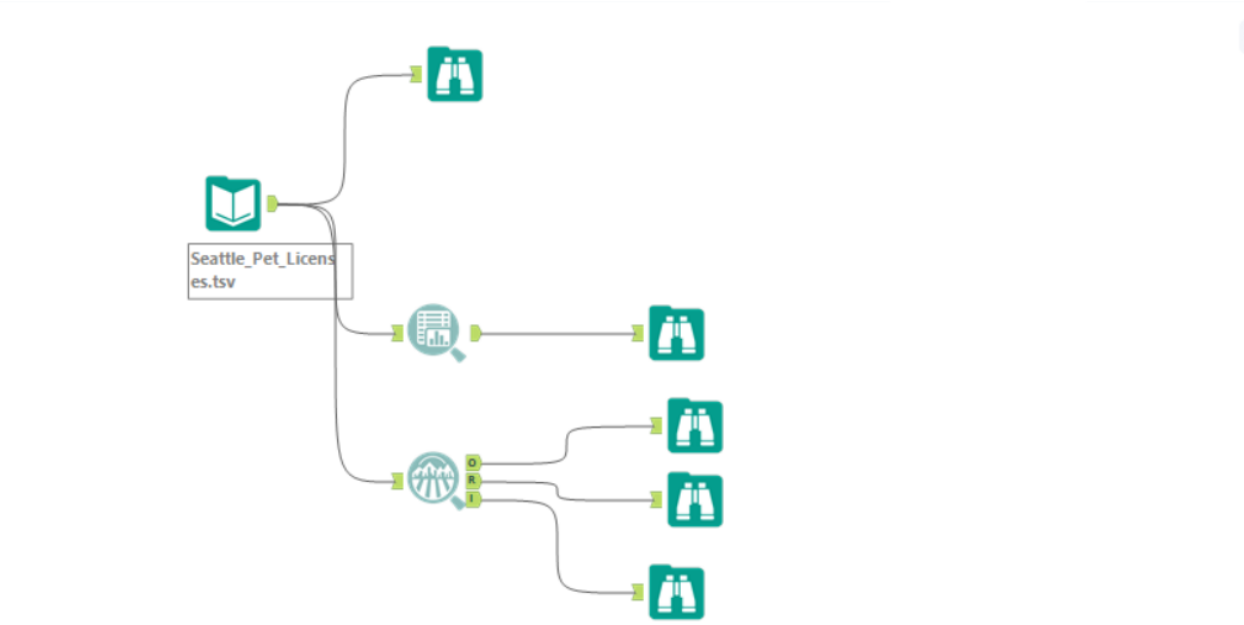
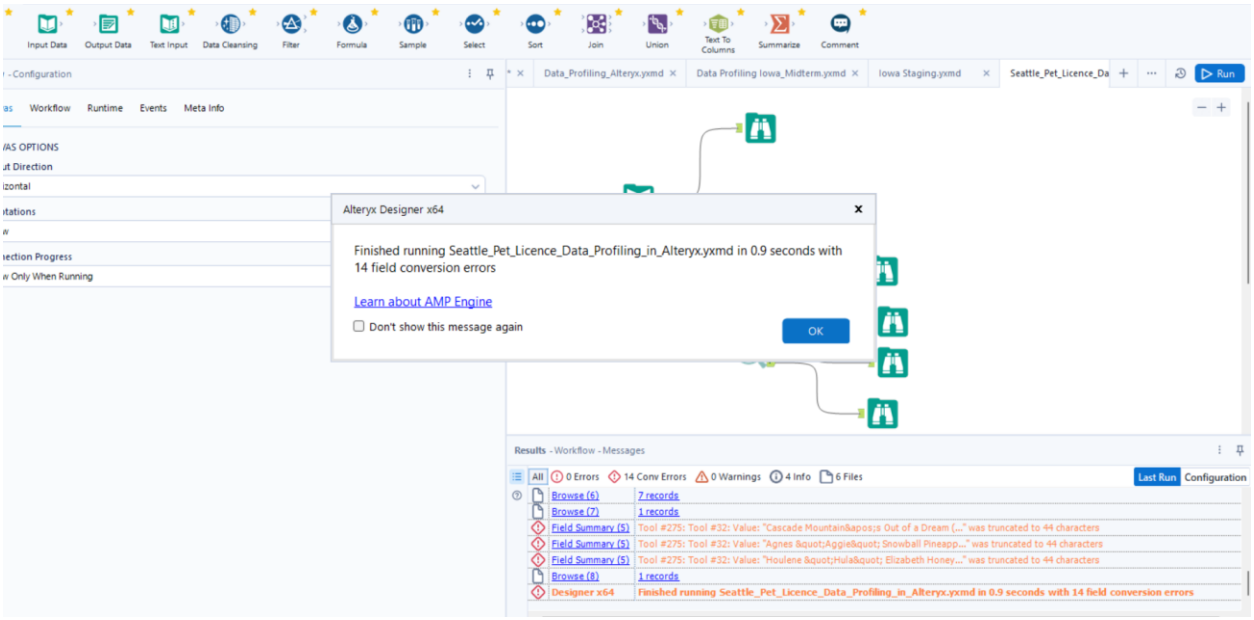
Seattle Pet License dimensional modeling and Loading via Talend - Workshop implementation

--By AMEY PARANGE

Alteryx Profiling workflow screenshot:





Observation:

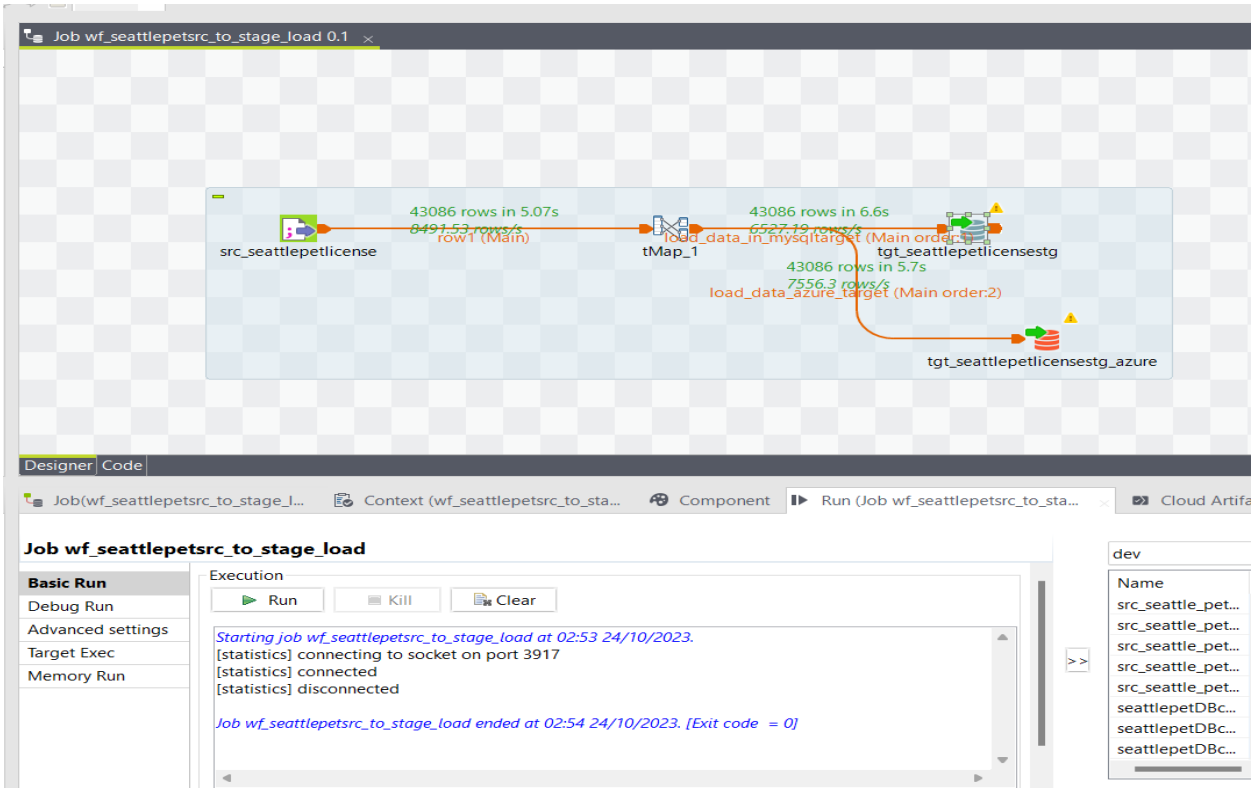
Name	Datatype	% missing	Unique values	Null values	Observation	Min Value	Max Value
License Issue Date	V_String	0	931	0.00%	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	11	17
License Number	V_String	0.0%	42192	0	Some values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

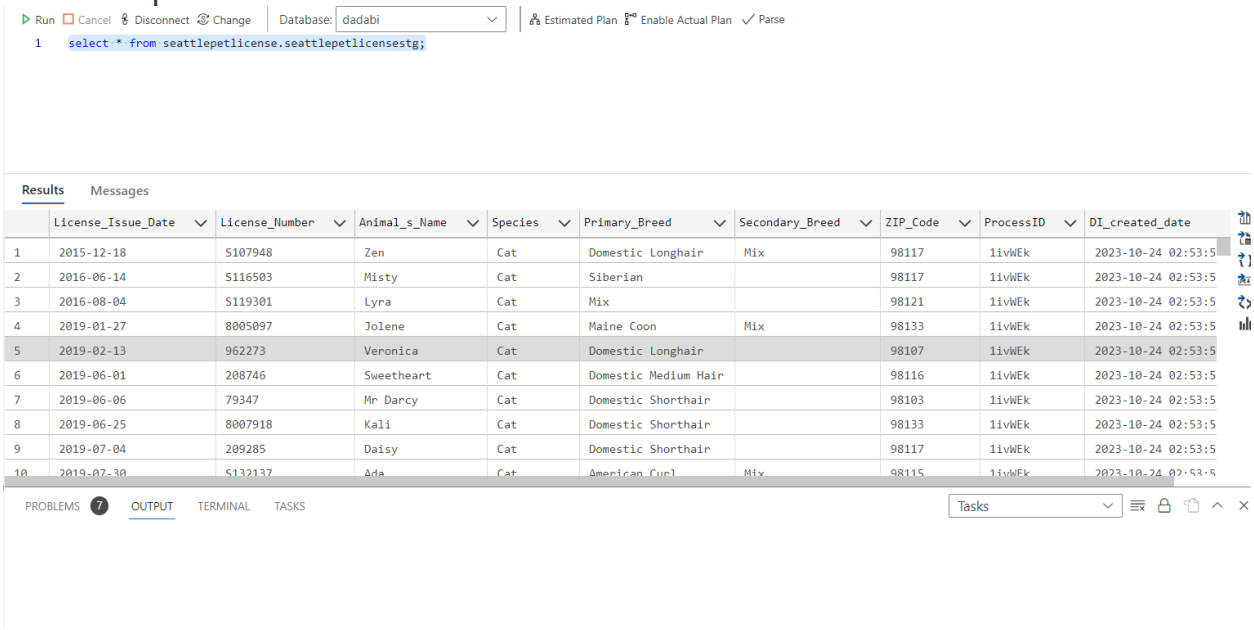
					consider combining some value levels together.		
Zip Code	V_String	0.3%	208	0	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.	3	5

Talend Implementation

wf_seattlepetsrc_to_stage_load



Azure output:



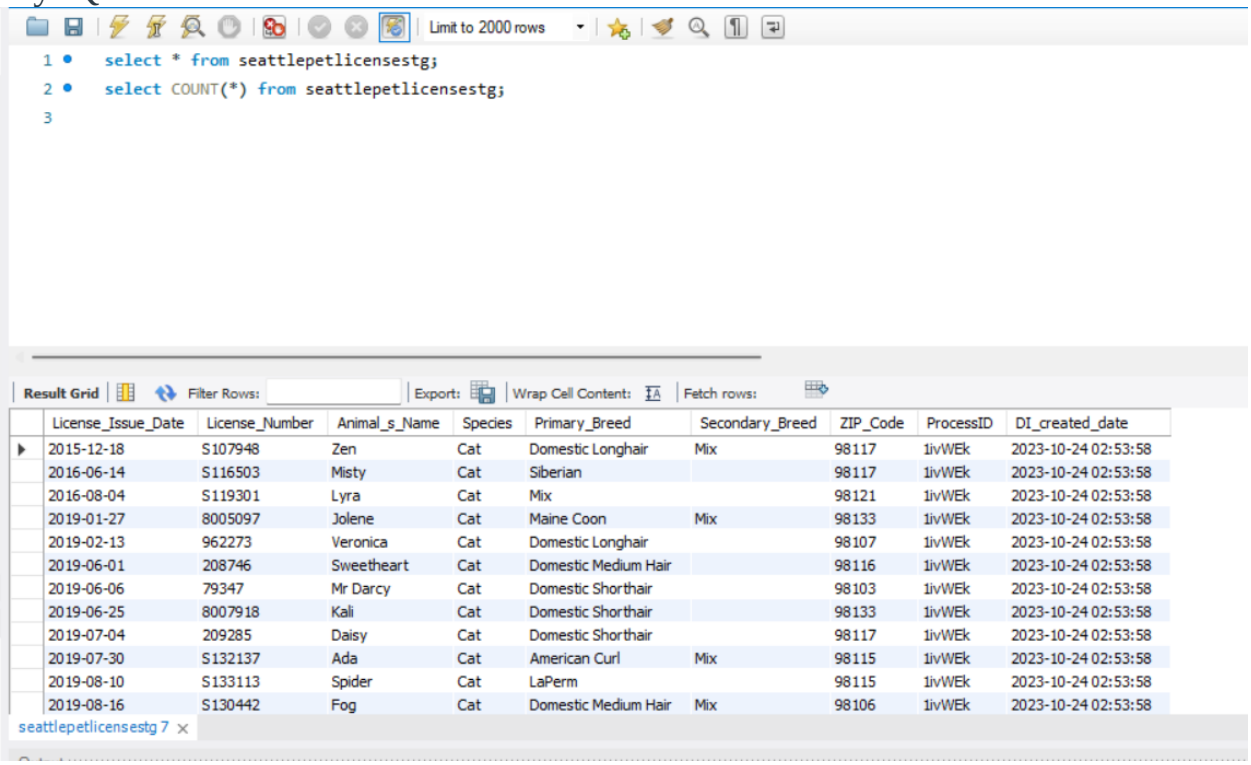
	License_Issue_Date	License_Number	Animal_s_Name	Species	Primary_Breed	Secondary_Breed	ZIP_Code	ProcessID	DI_created_date
1	2015-12-18	S107948	Zen	Cat	Domestic Longhair	Mix	98117	1ivWEk	2023-10-24 02:53:5
2	2016-06-14	S116503	Misty	Cat	Siberian		98117	1ivWEk	2023-10-24 02:53:5
3	2016-08-04	S119301	Lyra	Cat	Mix		98121	1ivWEk	2023-10-24 02:53:5
4	2019-01-27	8005097	Jolene	Cat	Maine Coon	Mix	98133	1ivWEk	2023-10-24 02:53:5
5	2019-02-13	962273	Veronica	Cat	Domestic Longhair		98107	1ivWEk	2023-10-24 02:53:5
6	2019-06-01	208746	Sweetheart	Cat	Domestic Medium Hair		98116	1ivWEk	2023-10-24 02:53:5
7	2019-06-06	79347	Mr Darcy	Cat	Domestic Shorthair		98103	1ivWEk	2023-10-24 02:53:5
8	2019-06-25	8007918	Kali	Cat	Domestic Shorthair		98133	1ivWEk	2023-10-24 02:53:5
9	2019-07-04	209285	Daisy	Cat	Domestic Shorthair		98117	1ivWEk	2023-10-24 02:53:5
10	2019-07-30	S132137	Ada	Cat	American Curl	Mix	98115	1ivWEk	2023-10-24 02:53:5

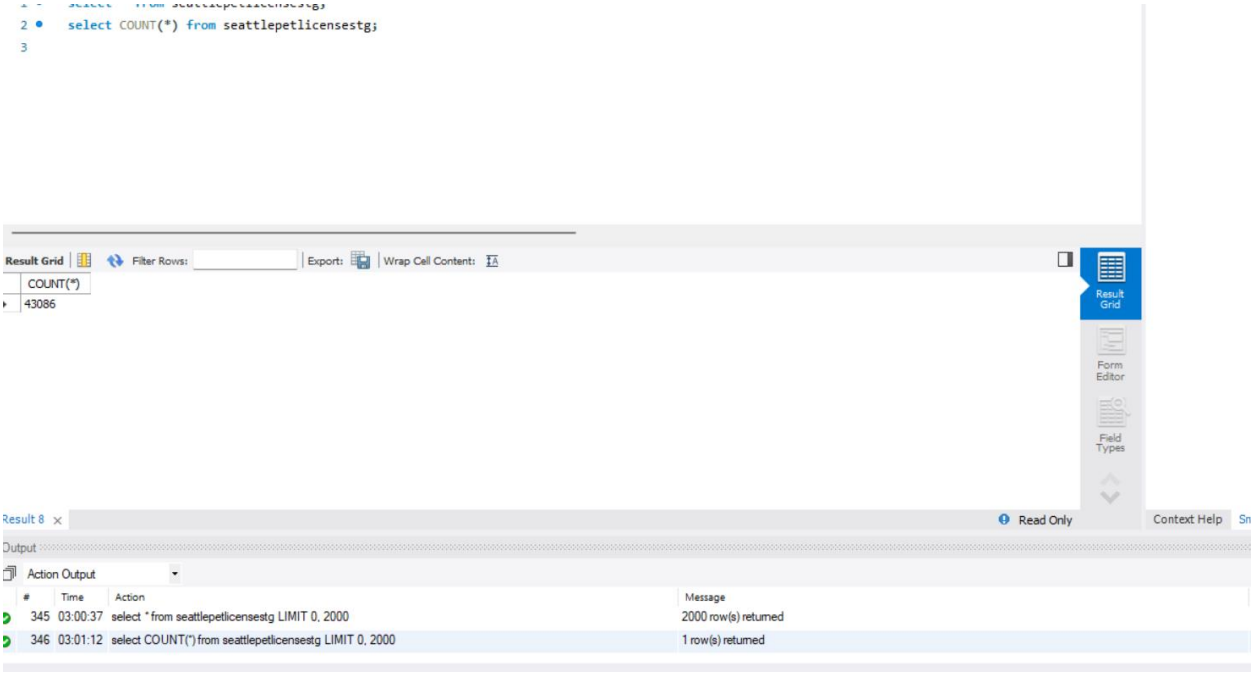
```
2
3   select COUNT(*) from seattlepetlicense.seattlepetlicensestg;
4
```

Results | Messages

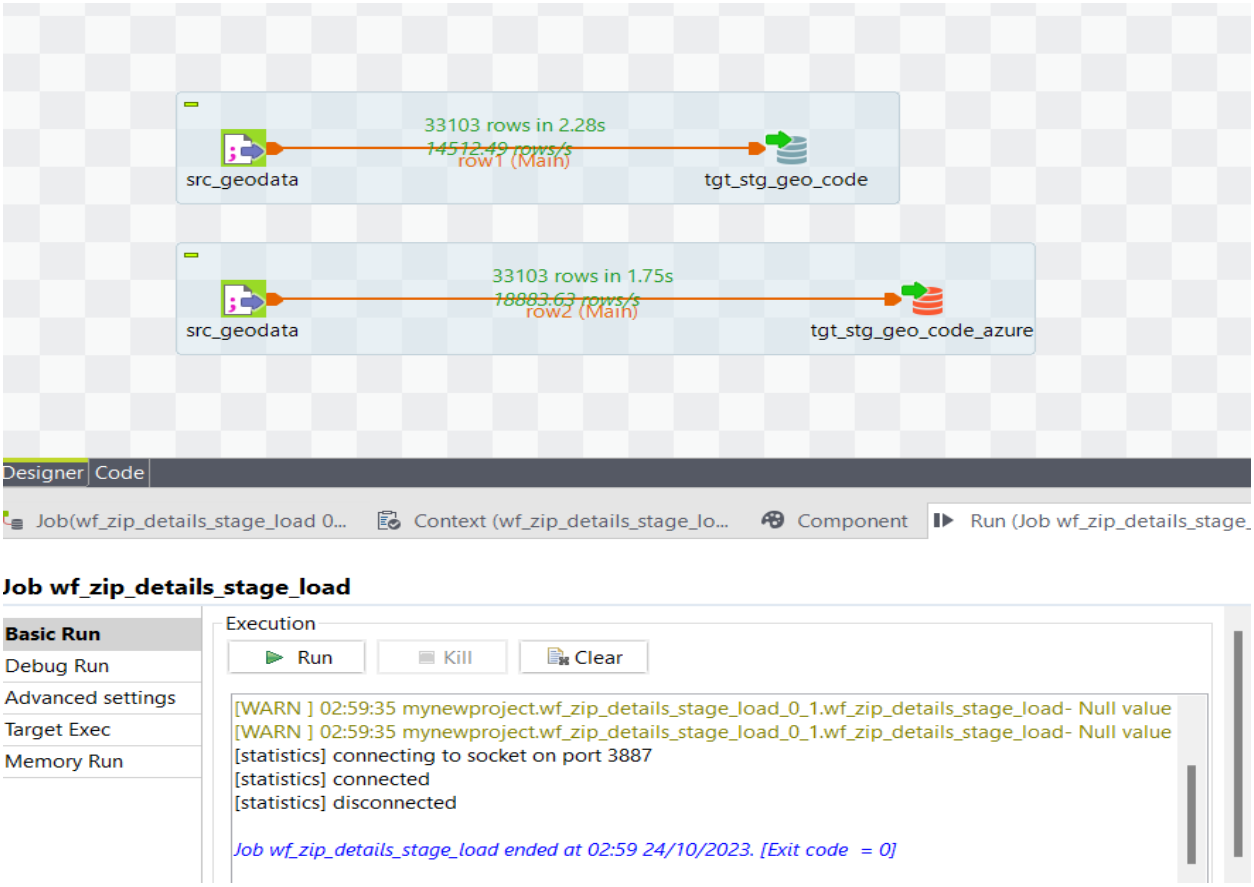
	(No column name)
1	43086

MySQL

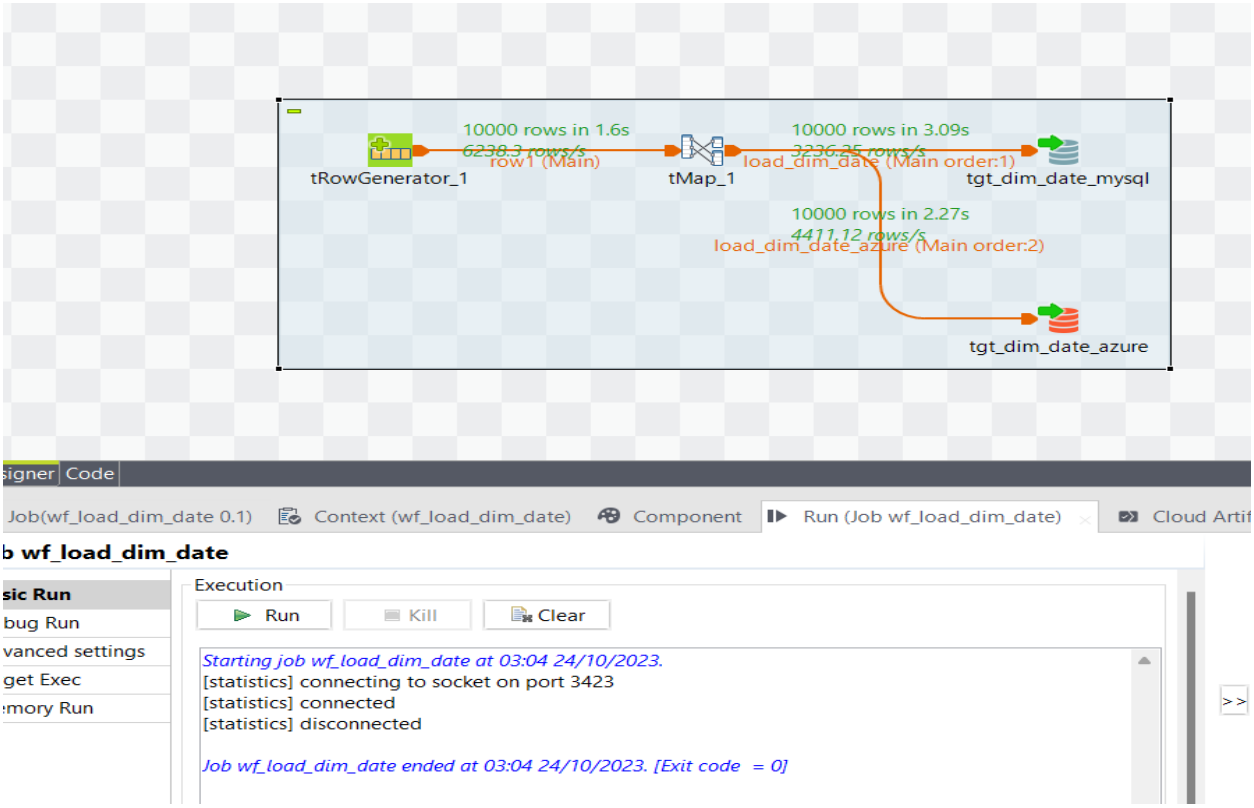




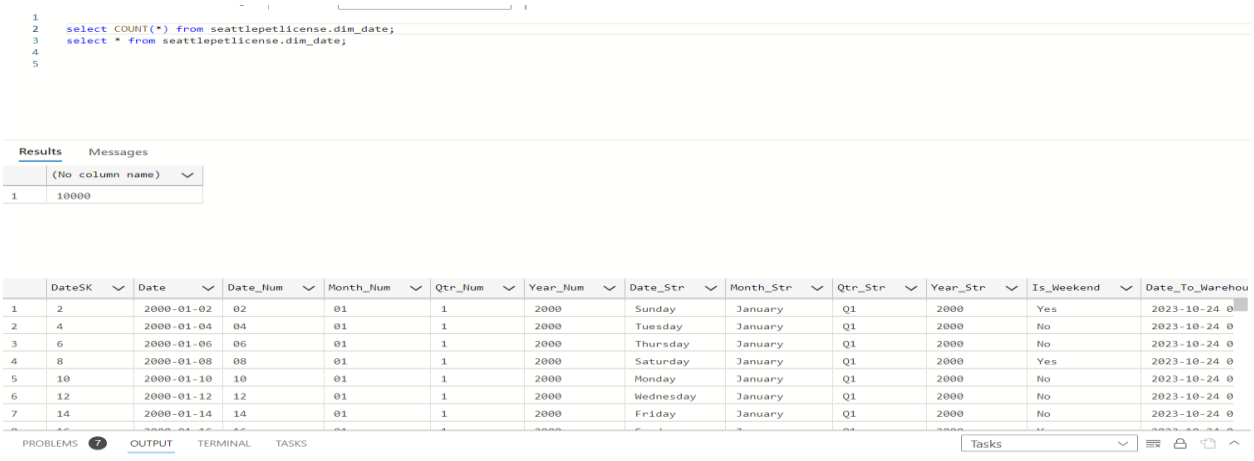
wf_zip_details_stage_load



wf_load_dim_date



Azure:



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

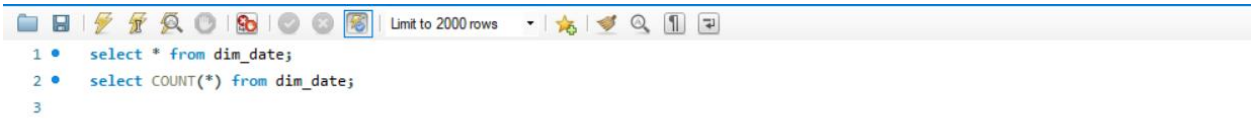
Results Messages

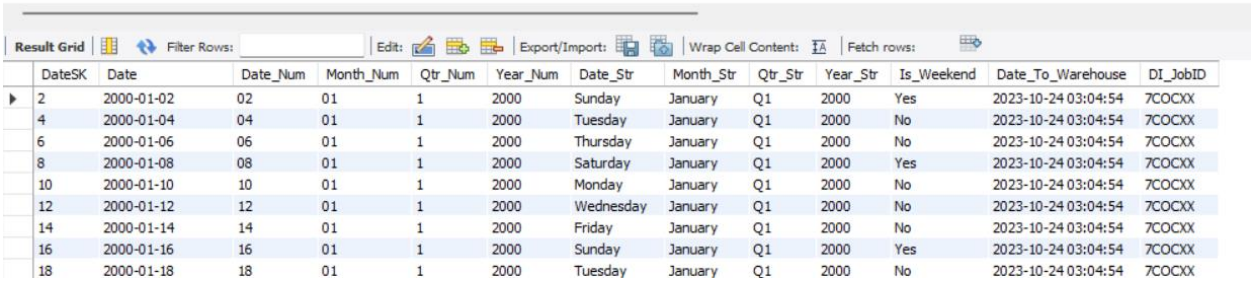
	(No column name)
1	10000

	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

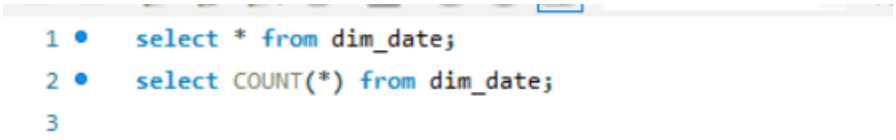
MySQL

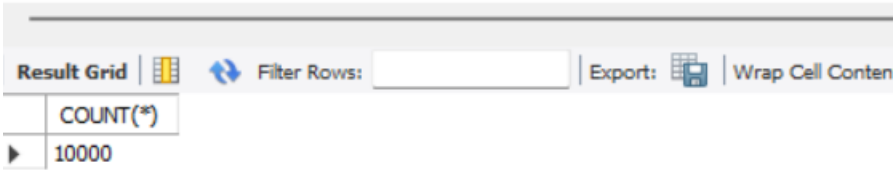




Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rows:

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



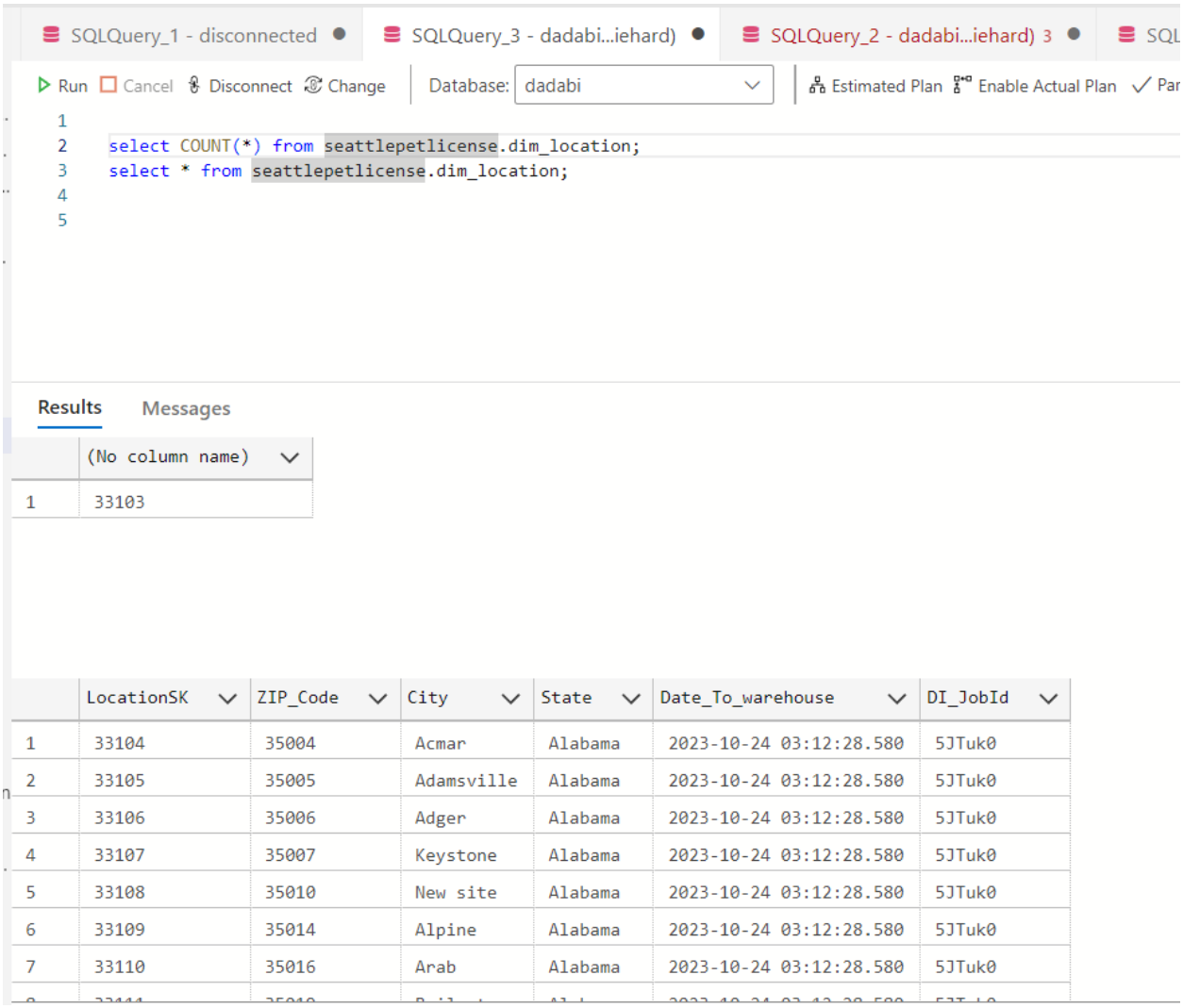


Result Grid | Filter Rows: | Export: | Wrap Cell Conten

COUNT(*)
10000

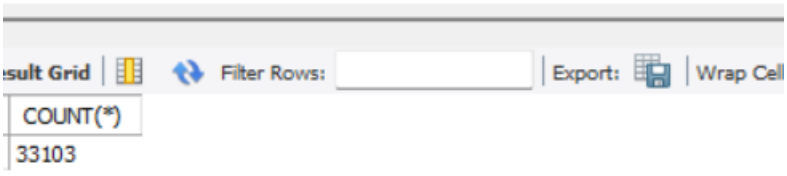
wf_load_dim_location

Azure


SQLQuery_1 - disconnected
SQLQuery_3 - dadabi...iehard)
SQLQuery_2 - dadabi...iehard) 3
Run
Cancel
Disconnect
Change
Database: dadabi
Estimated Plan
Enable Actual Plan
1
2 select COUNT(*) from seattlepetlicense.dim_location;
3 select * from seattlepetlicense.dim_location;
4
5
Results
Messages
(No column name)
1 33103
LocationSK
ZIP_Code
City
State
Date_To_warehouse
DI_JobId
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

MySQL

```
2 •  select COUNT(*) from dim_location;
3
```




```
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

```sql
select COUNT(*) from seattlepetlicense.dim_breed;
select * from seattlepetlicense.dim_breed;
```

Results Messages

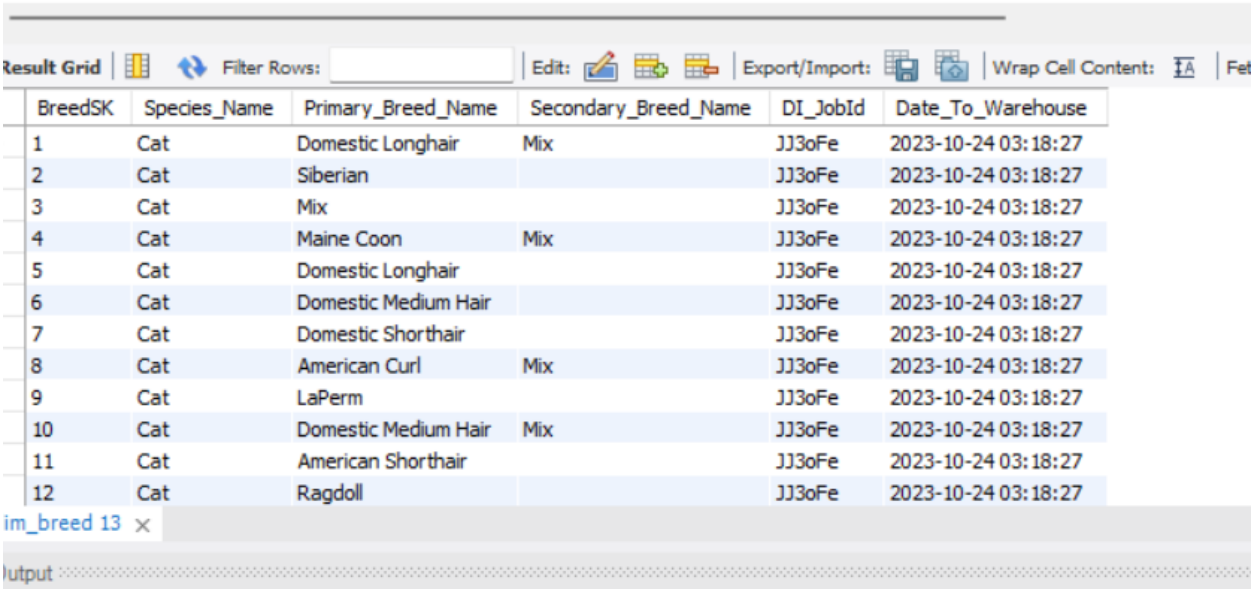
(No column name)
3351

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

PROBLEMS 7 OUTPUT TERMINAL TASKS

Tasks

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



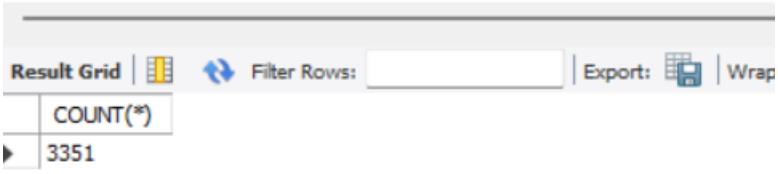
Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fet

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 ×

utput

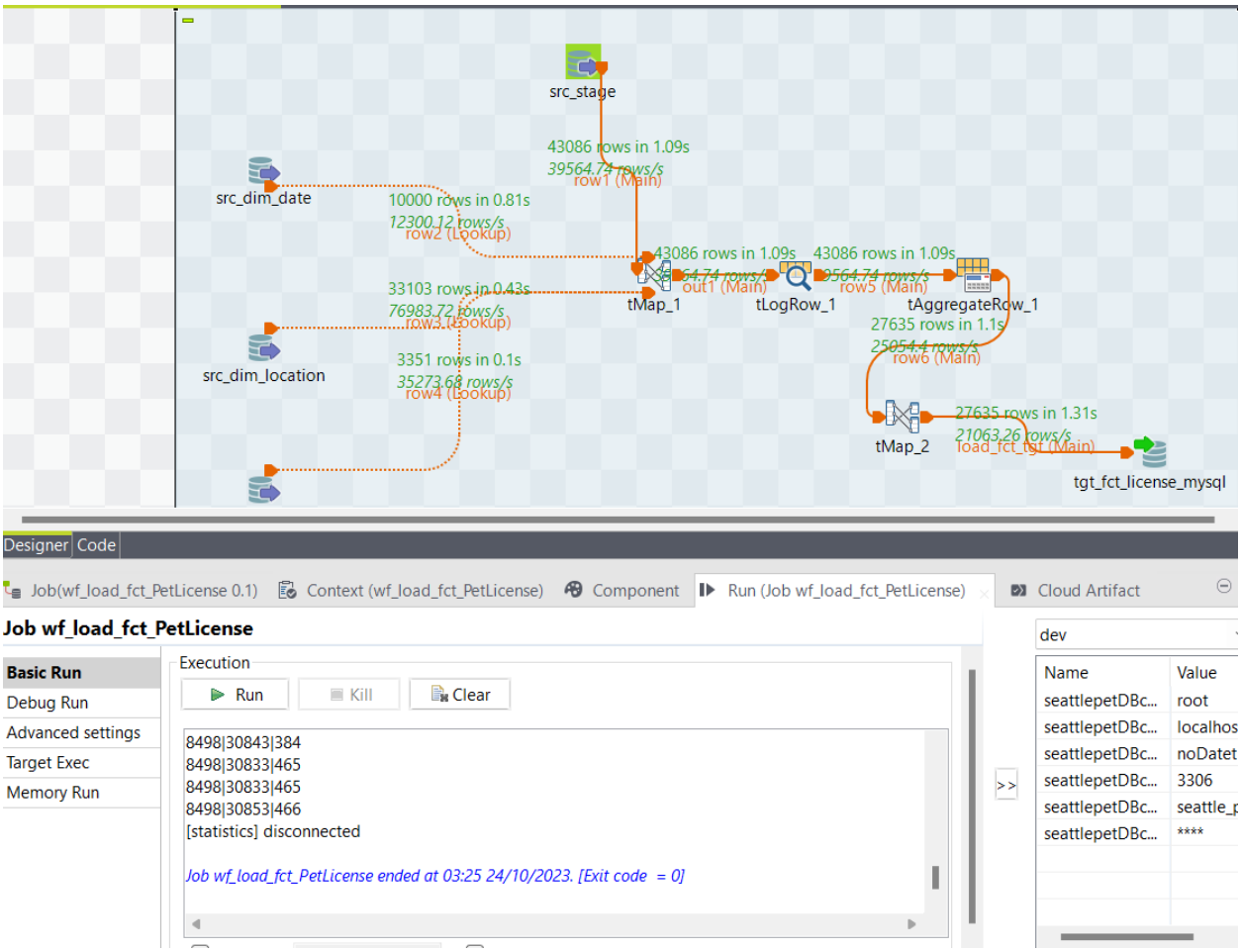
```
2 •  select COUNT(*) from dim_breed;
3
```



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
```

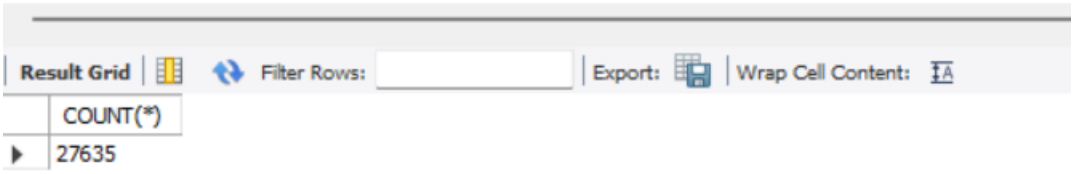
Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell

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

fct_license 15

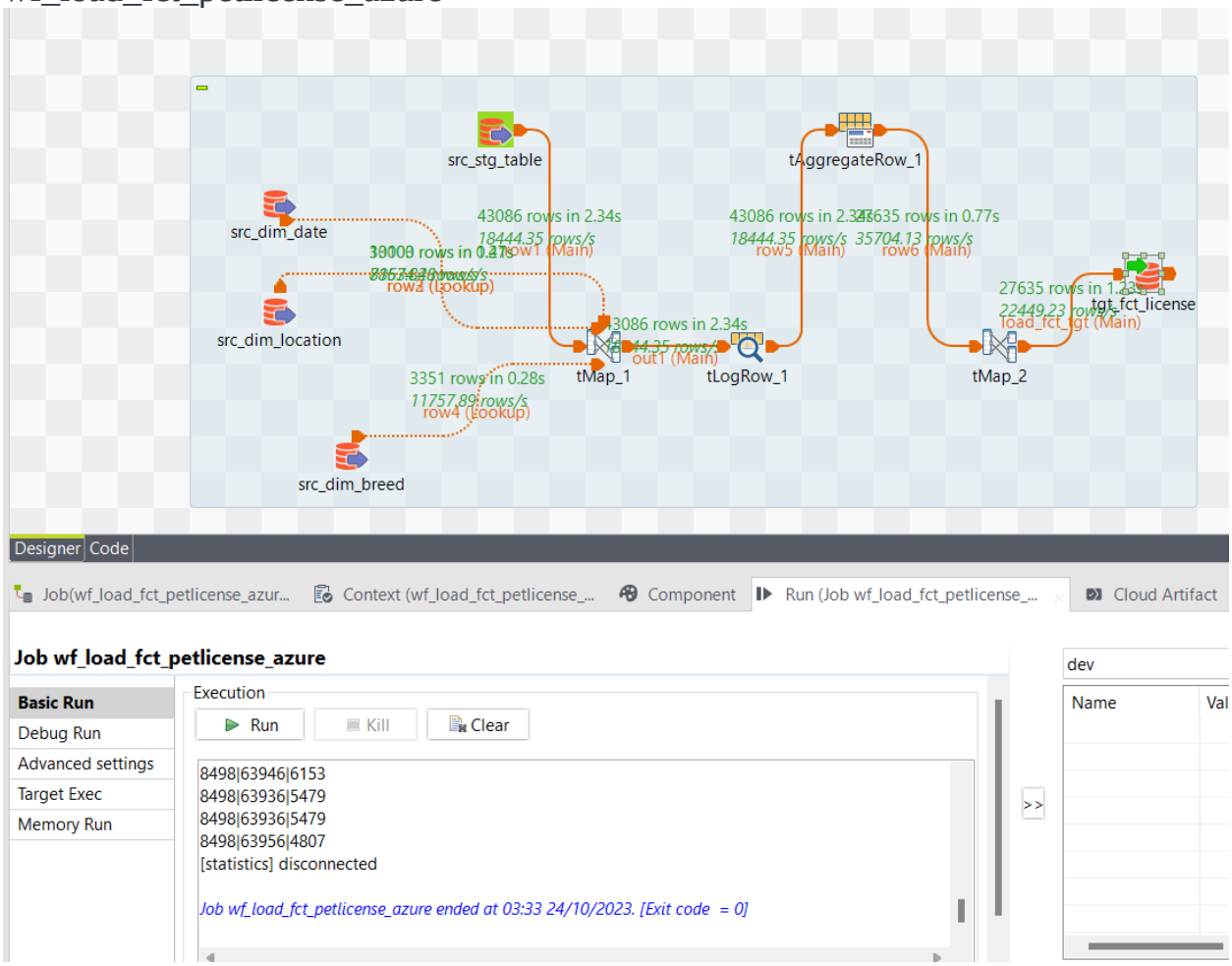
```sql
select * from fct_license;
select COUNT(*) from fct_license;

```



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	3bOftw
2	2	63964	4940	8130	1	2023-10-24 03:33:10.377	3bOftw
3	3	63945	3982	8268	1	2023-10-24 03:33:10.377	3bOftw
4	4	63945	4549	8472	1	2023-10-24 03:33:10.377	3bOftw
5	5	63940	4942	7858	1	2023-10-24 03:33:10.377	3bOftw
6	6	63940	4940	7858	2	2023-10-24 03:33:10.377	3bOftw
7	7	63950	5573	8198	1	2023-10-24 03:33:10.377	3bOftw

PROBLEMS 7 OUTPUT TERMINAL TASKS