Soha Khalid

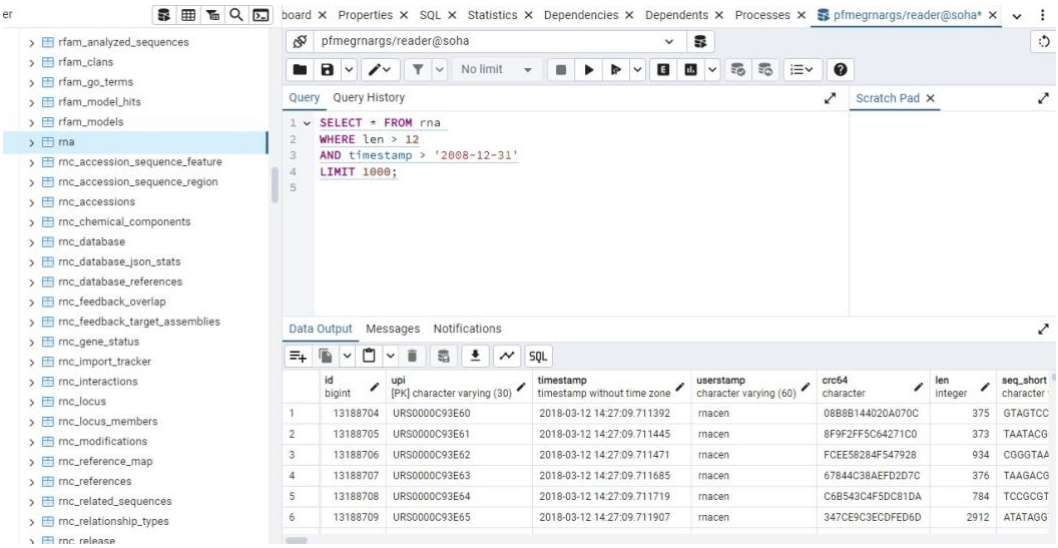
BWT – Data Engineering

Task 02 – Exercise

You will be using Rnacen (RNACentral) schema to find and explore the data available in all tables that will help you to recognise the potential tables you can query to answer the following questions.

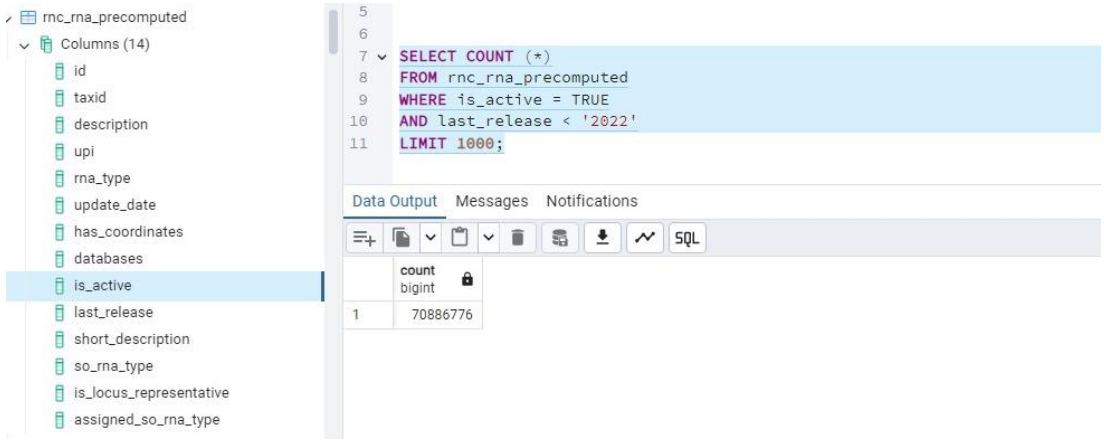
1. **Write a query to get data having length of Rna structures more than 12 with them being added after 2008.**

   SELECT * FROM rna
   WHERE len > 12
   AND timestamp > '2008-12-31'
   LIMIT 1000;



2. **How many pre computed RNA are present that are still active and got their last release update before 2022.**

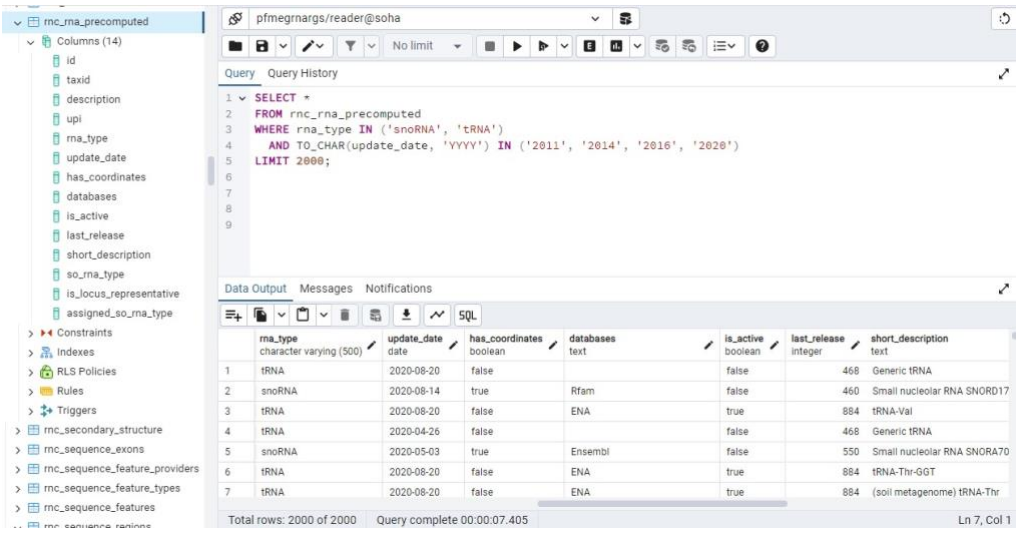
```
SELECT COUNT (*)
FROM rnc_rna_precomputed
WHERE is_active = TRUE
AND last_release < '2022'
LIMIT 1000;
```

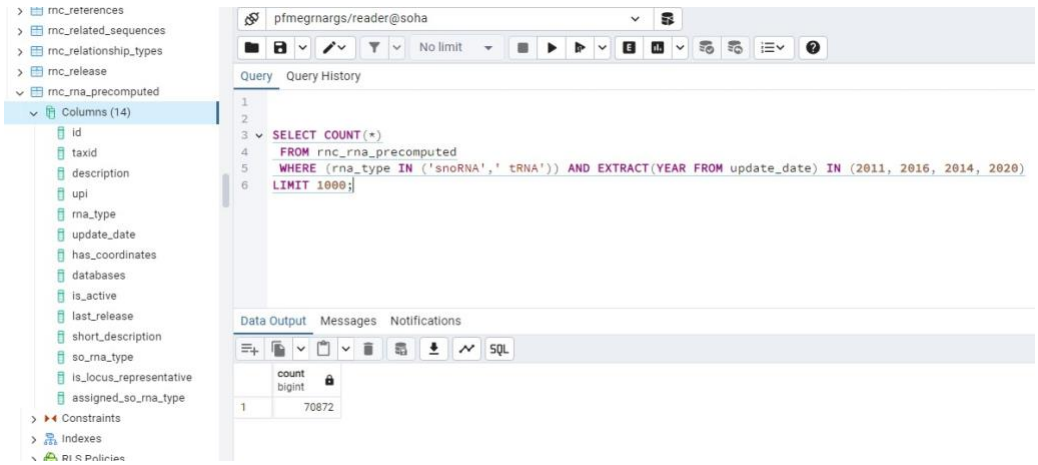


3. **How many total pre computed RNA records for snoRNA and tRNA were recorded in 2011, 2016, 2014, and 2020.**

```
SELECT *
FROM rnc_rna_precomputed
WHERE rna_type IN ('snoRNA', 'tRNA')
  AND TO_CHAR(update_date, 'YYYY') IN ('2011', '2014', '2016', '2020')
LIMIT 1000;
```

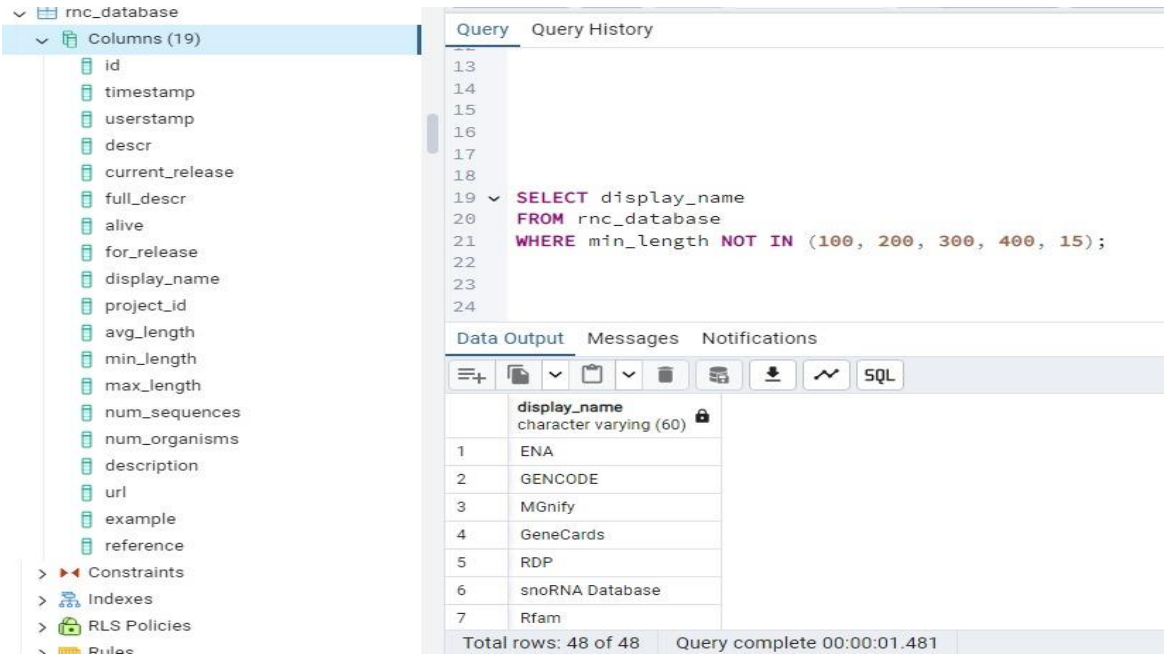
```
SELECT COUNT(*)
 FROM rnc_rna_precomputed
 WHERE (rna_type IN ('snoRNA',' tRNA')) AND EXTRACT(YEAR FROM update_date) IN
(2011, 2016, 2014, 2020)
LIMIT 1000;
```





4. **Can you give me the names of all databases built for RNA with minimum length other than 100, 200, 300, 400, and 15.**

```
SELECT display_name
FROM rnc_database
WHERE min_length NOT IN (100, 200, 300, 400, 15);
```



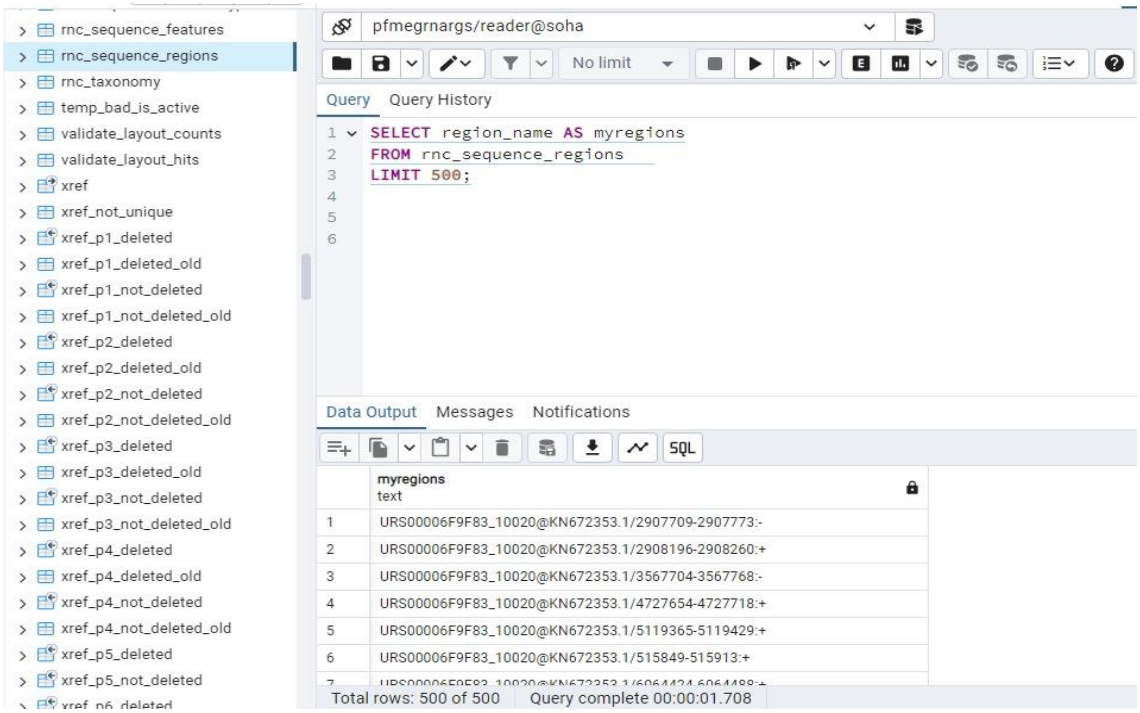
5. **Can you get complete 500 records of sequences for active regions and name your column as myregions in which you are getting the region name column value. Then tell me what different chromosomes with exon_count we have for regions including center, east and north using the name you set for your column.**

   SELECT region_name AS myregions

   FROM rnc_sequence_regions

   LIMIT 500;

   For last part no result can be retrieved from data because no such value named “center, north and east” existed in the given database.


rnc_sequence_features
rnc_sequence_regions
rnc_taxonomy
temp_bad_is_active
validate_layout_counts
validate_layout_hits
xref
xref_not_unique
xref_p1_deleted
xref_p1_deleted_old
xref_p1_not_deleted
xref_p1_not_deleted_old
xref_p2_deleted
xref_p2_deleted_old
xref_p2_not_deleted
xref_p2_not_deleted_old
xref_p3_deleted
xref_p3_deleted_old
xref_p3_not_deleted
xref_p3_not_deleted_old
xref_p4_deleted
xref_p4_deleted_old
xref_p4_not_deleted
xref_p4_not_deleted_old
xref_p5_deleted
xref_p5_not_deleted
pfmegrnargs/reader@soha
No limit
Query
Query History
1 SELECT region_name AS myregions
2 FROM rnc_sequence_regions
3 LIMIT 500;
4
5
6
Data Output
Messages
Notifications
SQL
myregions
text
1 URS00006F9F83_10020@KN672353.1/2907709-2907773:-
2 URS00006F9F83_10020@KN672353.1/2908196-2908260:+
3 URS00006F9F83_10020@KN672353.1/3567704-3567768:-
4 URS00006F9F83_10020@KN672353.1/4727654-4727718:+
5 URS00006F9F83_10020@KN672353.1/5119365-5119429:+
6 URS00006F9F83_10020@KN672353.1/515849-515913:+
Total rows: 500 of 500
Query complete 00:00:01.708