

BWT Task-02 Exercise

Submitted By: ABRAR SAEED

Q1

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

The screenshot shows a database query interface with a script editor at the top and a results grid below. The script editor contains the following SQL query:

```
select * from rnacen.rna where ("timestamp" > '2008-12-31 00:00:00') AND (rna.len > 12) limit 500;
```

The results grid displays 18 rows of data. The columns are: id, upi, timestamp, userstamp, and crc64. The first row is highlighted.

	id	upi	timestamp	userstamp	crc64
1	31,087,201	URS0001DA5A61	2020-08-19 23:19:29.729	rnacen	4962DBF642EE53FA
2	31,087,202	URS0001DA5A62	2020-08-19 23:19:29.729	rnacen	EBD76A8ED6AADC24
3	31,087,203	URS0001DA5A63	2020-08-19 23:19:29.729	rnacen	298F34F2B83CAD7B
4	31,087,204	URS0001DA5A64	2020-08-19 23:19:29.729	rnacen	4CC0C8920FDA6467
5	31,087,205	URS0001DA5A65	2020-08-19 23:19:29.729	rnacen	BE242ABECC4DEFCE
6	31,087,206	URS0001DA5A66	2020-08-19 23:19:29.729	rnacen	B966D5D723EACCD3
7	31,087,207	URS0001DA5A67	2020-08-19 23:19:29.729	rnacen	A8489F9D9613B823
8	31,087,208	URS0001DA5A68	2020-08-19 23:19:29.729	rnacen	3159EE80E79368DD
9	31,087,209	URS0001DA5A69	2020-08-19 23:19:29.729	rnacen	40D64E49EE106006
10	31,087,210	URS0001DA5A6A	2020-08-19 23:19:29.729	rnacen	EA352B049BE52A6
11	31,087,211	URS0001DA5A6B	2020-08-19 23:19:29.729	rnacen	66F893DD36E89FF6
12	31,087,212	URS0001DA5A6C	2020-08-19 23:19:29.729	rnacen	F992788FE9F06734
13	31,087,213	URS0001DA5A6D	2020-08-19 23:19:29.729	rnacen	F09139BD2A0BF717
14	31,087,214	URS0001DA5A6E	2020-08-19 23:19:29.729	rnacen	D95C047D9B759494
15	31,087,215	URS0001DA5A6F	2020-08-19 23:19:29.730	rnacen	72A57A7CA2DDACF3
16	31,087,216	URS0001DA5A70	2020-08-19 23:19:29.730	rnacen	2101F160FCA8CA72
17	31,087,217	URS0001DA5A71	2020-08-19 23:19:29.730	rnacen	0E54B40FF6FB17BE
18	31,087,218	URS0001DA5A72	2020-08-19 23:19:29.730	rnacen	4A2CD9E620BCB369

The interface includes a search bar with the text "Enter a SQL expression to filter results (use Ctrl+Space)". At the bottom, there are buttons for "Refresh", "Save", "Cancel", and "Export data". The "Export data" button is set to "2000" rows and "500" columns. A "Value" field shows "31087201".

The condition of limit 500 has been added because of large result set achieved after executing query. There are more rows returned instead.

Q2

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

```
SELECT count(distinct rrp.id) FROM rnacen.rnc_rna_precomputed rrp
WHERE rrp.update_date < '2022-01-01' AND (rrp.is_active = true )
```

rnc_rna_precomputed update_date < '2022-01-01' AND (is_active = true)						
	pe	update_date	has_coordinates	databases	is_active	
1		2020-08-03	[]	SILVA	[v]	
2		2020-08-07	[]	SILVA	[v]	
3		2020-08-03	[]	SILVA	[v]	
4		2020-08-07	[]	SILVA	[v]	
5		2020-08-03	[]	SILVA	[v]	
6		2020-08-07	[]	SILVA	[v]	
7		2020-08-03	[]	SILVA	[v]	
8		2020-08-07	[]	SILVA	[v]	
9		2020-08-03	[]	SILVA	[v]	
10		2020-08-07	[]	SILVA	[v]	
11		2020-08-03	[]	SILVA	[v]	
12		2020-08-07	[]	SILVA	[v]	
13		2020-08-03	[]	SILVA	[v]	
14		2020-08-07	[]	SILVA	[v]	
15		2020-08-03	[]	SILVA	[v]	
16		2020-08-07	[]	SILVA	[v]	
17		2020-08-03	[]	SILVA	[v]	
18		2020-08-07	[]	SILVA	[v]	
19		2020-08-03	[]	SILVA	[v]	
20		2020-08-07	[]	SILVA	[v]	
21		2020-08-03	[]	SILVA	[v]	
22		2020-08-07	[]	SILVA	[v]	
23		2020-08-03	[]	SILVA	[v]	

Refresh Save Cancel Export data 2000 55,930,772

	id	taxid	description	upi	rna_ty
1	URS0001B43F79_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F79	rRNA
2	URS0001B43F7A	[NULL]	rRNA from 1 species	URS0001B43F7A	rRNA
3	URS0001B43F7A_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7A	rRNA
4	URS0001B43F7B	[NULL]	rRNA from 1 species	URS0001B43F7B	rRNA
5	URS0001B43F7B_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7B	rRNA
6	URS0001B43F7C	[NULL]	rRNA from 1 species	URS0001B43F7C	rRNA
7	URS0001B43F7C_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7C	rRNA
8	URS0001B43F7D	[NULL]	rRNA from 1 species	URS0001B43F7D	rRNA
9	URS0001B43F7D_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7D	rRNA
10	URS0001B43F7E	[NULL]	rRNA from 1 species	URS0001B43F7E	rRNA
11	URS0001B43F7E_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7E	rRNA
12	URS0001B43F7F	[NULL]	rRNA from 1 species	URS0001B43F7F	rRNA
13	URS0001B43F7F_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7F	rRNA
14	URS0001B43F80	[NULL]	rRNA from 1 species	URS0001B43F80	rRNA
15	URS0001B43F80_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F80	rRNA
16	URS0001B43F81	[NULL]	rRNA from 1 species	URS0001B43F81	rRNA
17	URS0001B43F81_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F81	rRNA
18	URS0001B43F82	[NULL]	rRNA from 1 species	URS0001B43F82	rRNA
19	URS0001B43F82_4498	4,498	Avena sativa SSU rRNA	URS0001B43F82	rRNA
20	URS0001B43F83	[NULL]	rRNA from 1 species	URS0001B43F83	rRNA
21	URS0001B43F83_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F83	rRNA
22	URS0001B43F84	[NULL]	rRNA from 1 species	URS0001B43F84	rRNA
23	URS0001B43F84_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F84	rRNA

Refresh Save Cancel Export data 2000 55,930,772

The highlighted number at the bottom right corner (55,930,772) tells the number of rows returned based on the conditions of query that are all those pre-computed RNA that are active and have last update before 2022.

Q3

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

The screenshot shows a database query editor with a script window and a results grid. The script window contains the following SQL query:

```
SELECT count(rrp.id)
FROM rnacen.rnc_rna_precomputed rrp
WHERE rrp.rna_type IN ('snoRNA', 'tRNA')
AND EXTRACT(YEAR FROM rrp.update_date) IN (2011, 2014, 2016, 2020)
```

The results grid shows the following data:

Grid	count
1	915,377

The results grid is highlighted with a yellow box. The query editor also shows a search bar with the text "SELECT count(rrp.id) FROM rnacen.rnc_" and a placeholder text "Enter a SQL expression to filter results (use Ctrl+Space)".

Q4

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

```
select distinct rrp.databases from rnacen.rnc_rna_precomputed rrp join rnacen.rna
on rrp.upi = rna.upi where rna.len not in (100,200,300,400,15)
```

The screenshot shows a database query interface with a script editor and a results grid. The script editor contains the following SQL query:

```
select distinct rrp.databases from rnacen.rnc_rna_precomputed rrp join rnacen.rna
on rrp.upi = rna.upi where rna.len not in (100,200,300,400,15)
```

The results grid displays the following data:

Grid	databases
1	[NULL]
2	MGI, NONCODE
3	ENA, Ensembl, GeneCards, HGNC, RefSeq, snoDB
4	Ensembl, Ensembl/GENCODE, EVIncRNAs, GeneCards, LncBook, LNCipedia, MalaCards, NONCODE
5	ENA, miRBase, MirGeneDB, PirBase, RefSeq
6	Ensembl Plants, Rfam, SRPDB
7	ENA, Greengenes, Modomics, RefSeq, Rfam, SILVA
8	ENA, Rfam, WormBase
9	ENA, Ensembl, Ensembl/GENCODE, GeneCards, LncBook, LNCipedia, RefSeq
10	ENA, Rfam, SRPDB
11	Ensembl, PirBase
12	Ensembl, GeneCards, HGNC, miRBase, RefSeq
13	ENA, PDBe, PirBase
14	Ensembl, GeneCards, MalaCards, NONCODE
15	ENA, Ensembl Metazoa, FlyBase, miRBase, RefSeq, Rfam
16	5S rRNAdb, CRW, ENA, RefSeq, Rfam, WormBase
17	ENA, NONCODE, RefSeq

The interface also shows a status bar indicating "1586 row(s) fetched" and a timestamp "8m 17s, on 2024-06-16 at 13:11:02".

<pfmegrnargs> Script ×

rnc_rna_precomputed

```
select distinct rrp.databases from rncen.rnc_rna_precomputed rrp join rncen.rna
on rrp.upi = rna.upi where rna.len not in (100,200,300,400,15)
```

rnc_rna_precomputed 1 ×

select distinct rrp.databases from rncen

Enter a SQL expression to filter results (use Ctrl+Space)

Grid

Value ×

1571 databases

1572 NONCODE,RefSeq

1573 GeneCards,LncBook,MalaCards,RefSeq

1574 GeneCards,HGNC,LncBook

1575 ENA,Ensembl,GeneCards,Rfam,snOPY

1576 GeneCards,Rfam,RiboCentre

1577 Ensembl,Ensembl Plants

1578 ENA,Ensembl,Ensembl/GENCODE,GeneCards,HGNC,LncBook,LNCipedia,LncRNAdb,MalaCards,RefSeq

1579 Ensembl,GeneCards,LncBook,LNCipedia,MalaCards,Rfam

1580 Ensembl Metazoa,SILVA

1581 GeneCards,LncBase,MalaCards,miRBase,MirGeneDB,RefSeq,TarBase

1582 MGI,NONCODE,RefSeq

1583 Ensembl,MGI,miRBase,MirGeneDB,RefSeq,Rfam

1584 ENA,Expression Atlas,GeneCards,LNCipedia,MalaCards,NONCODE

1585 ENA,NONCODE

1586 miRBase,MirGeneDB,PirBase

1587 ENA,GeneCards,HGNC,RefSeq,snoDB,snOPY

Refresh Save Cancel

Export data 2000 1,586

1586 row(s) fetched - 8m 17s, on 2024-06-16 at 13:11:02

Activate Windows
Go to Settings to activate Windows.

Q5

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?

Script × rnc_rna_precomputed rnc_sequence_regions rnc_sequence_exons

```
select rsra.id, rsra.chromosome, rsra.exon_count, rsra.region_name as myregions
from rnacen.rnc_sequence_regions_active rsra limit 500
```

rnc_sequence_regions_active 1 ×

select rsra.id, rsra.chromosome, rsra.exon_count, rsra.region_name as myregions

Enter a SQL expression to filter results (use Ctrl+Space)

Grid	123 id	ABC chromosome	123 exon_count	ABC myregions
1	181,644	CACQ02003602	1	URS000000086E_80884@CACQ02003602/5181-
2	181,645	CACQ02008496	1	URS000000086E_80884@CACQ02008496/758-8
3	181,646	CACQ02008999	1	URS000000086E_80884@CACQ02008999/276-3
4	182,352	supercont1.68	1	URS00000001578_403677@supercont1.68/74803
5	183,208	10	3	URS0000000224D_5855@10/12361-13959:-
6	183,209	2	1	URS0000000224D_5855@2/146582-148180:-
7	183,210	3	3	URS0000000224D_5855@3/806003-807601:+
8	183,710	CACQ02003602	2	URS0000000246A_80884@CACQ02003602/5251-
9	183,711	CACQ02008496	1	URS0000000246A_80884@CACQ02008496/828-
10	183,712	CACQ02009340	1	URS0000000246A_80884@CACQ02009340/233-
11	183,786	supercont1.28	1	URS0000000259C_403677@supercont1.28/14388
12	185,619	12	1	URS00000002F4A_5823@12/1587712-1587794:-
13	187,569	I	4	URS00000003D30_4896@I/3754985-3755801:-
14	187,570	I	5	URS00000003D30_4896@I/3786530-3787370:+
15	187,571	II	4	URS00000003D30_4896@II/1611692-1612508:-
16	187,572	II	4	URS00000003D30_4896@II/1635374-1636190:+
17	187,573	III	5	URS00000003D30_4896@III/1089166-1089994:-
18	187,574	III	5	URS00000003D30_4896@III/1108951-1109779:+
19	187,622	I	1	URS00000003DC0_4896@I/149658-149776:+
20	187,623	I	1	URS00000003DC0_4896@I/1563475-1563593:+

Refresh Save Cancel Export data 2000 500

Writable Smart Insert 2 · 55 · 136 Sel: 0 | 0

