PostgreSQL Queries

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

Query:

```
SELECT * FROM rna

Where len > 12 AND timestamp > '2008-12-31 23:59:59.00'

ORDER BY timestamp

Limit 1000;
```

Output:

	id bigint	upi [PK] character varying (30)	timestamp timestamp without time zone	userstamp character varying (60)	crc64 character	len integer	seq_short character varying (4000)
1	975	URS00000003CF	2014-05-29 13:51:05	RNACEN	97F5ACDDA3BE62D9	1471	GAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTA
2	1900	URS000000076C	2014-05-29 13:51:05	RNACEN	3E75483FFBD37CEA	482	ACGGGGGCCCGCACAAGCGGTGGAGCACGTTGGTTTAATTCC
3	3465	URS0000000D89	2014-05-29 13:51:05	RNACEN	7CEA12DE0529F9AC	316	AAAGGCCGGCGATACAGGGTGACAGCCCCGTACACGAAGGCA
4	970	URS00000003CA	2014-05-29 13:51:05	RNACEN	5CB0C9C6E7D299F0	68	AAAGTAAGGTCAGCTAAATAAGCTATCGGGCCCATACCCCGA
5	1459	URS00000005B3	2014-05-29 13:51:05	RNACEN	89325FA51B31E344	273	ATCGTTGCCCCCGTGCCTTGGCCACGTGCTAGGCACCAAGCG
6	1899	URS000000076B	2014-05-29 13:51:05	RNACEN	EB82AB69E57E275B	748	GATCCTGGCTCAGAATCAACGCTGGCGGCGTGCCTAACACAT
7	258	URS0000000102	2014-05-29 13:51:05	RNACEN	74555176E4C44530	673	AAAGTCGTAACAAGGTACCGGTAGGTGAACCTGCCGGTGGAT
8	498	URS00000001F2	2014-05-29 13:51:05	RNACEN	73EE0B12A9C43635	1090	GGATCATTAAAGAGTTCTATAACTCCCAAACCCATGTGAACAT
9	891	URS000000037B	2014-05-29 13:51:05	RNACEN	6EB932C8052187DE	350	ATTGAACGTTGGCGGCAGGCCTAAAACATGCAAGTCGAACGG
10	969	URS00000003C9	2014-05-29 13:51:05	RNACEN	3D510D83F6545042	391	CCGAATCGGTCCTGTGCTCCCCCGGGAGTGCATGGCCAACTA
11	1206	URS00000004B6	2014-05-29 13:51:05	RNACEN	6C31C045F7B99BF7	1150	CGAAAGGCCTAATAATACCGGATAAGCTCACGGACCCGCATG
12	1458	URS00000005B2	2014-05-29 13:51:05	RNACEN	BA146498A2306283	68	TGGTAGTTAGTTTAAACTAAAATAAATGATTTCGACTCATTAG
13	1705	URS00000006A9	2014-05-29 13:51:05	RNACEN	B5091B78E753B090	545	AAGGATCATTGGAACGGAGTTGATTCGAGCTCCGGCTCGACT
14	1897	URS0000000769	2014-05-29 13:51:05	RNACEN	67AC07AB28D28200	1321	GCACGAACGCTTGTAGGCACGCTTAACACATGCAAGTCGAAC
15	117	URS0000000075	2014-05-29 13:51:05	RNACEN	CFF5703FFFBE3596	155	TAAGGAATATTGGTCATGGACGCAAGTCTGAAACCAGCCATG
16	257	URS0000000101	2014-05-29 13:51:05	RNACEN	A7FC90D76847AF92	486	GATGAACGCTAGCGGCAGGCTTAATACATGCAAGTCGAACGG
17	364	URS000000016C	2014-05-29 13:51:05	RNACEN	E58217ABAAF59011	263	CAGCCCTAAACCGATGCGCGTTAGGTGTATCGGTGACCACGA
18	497	URS00000001F1	2014-05-29 13:51:05	RNACEN	CE89DBC6FD1A50E0	946	TTTCCGGTTGATCCTGCCGGACCCCACTGCTATCGGGATAGGA
19	616	URS0000000268	2014-05-29 13:51:05	RNACEN	DB63FAC99001B47C	74	GGCGCTATAGCCAAGTGGTAAGGCATAGGTCTGCAACACCTT
20	729	URS00000002D9	2014-05-29 13:51:05	RNACEN	6AEE84E09761179A	509	AGGGTTTGCTGGCGTCTCGGTTAATAGCAGTAACGGCAGCCG
21	849	URS0000000351	2014-05-29 13:51:05	RNACEN	500031529228EB8D	386	CTGCCATGTTGTGAGTTGCCCTGTGGAGAGGCCCATGTGGAA

seq_long text	md5 character varying (64)
[null]	1febcae280b6c5bbc0bbdbad2cbca823
[null]	03215b62685bc8ad5f4925bc909b0856
[null]	6bdf01cfde1a509de6cd4afa6389b9ce
[null]	7bbb6a4f4a8b2a1077a541f3672e3223
[null]	1ff039e3bf23f9a15017d8e8dc1ccf76
[null]	6bd03c6afd88e866bfe0cb426c840742
[null]	7bb34dfaf604c2f7e1b782f85f82dfca
[null]	6bbfc37889e9f5ad0d862b08239546c1
[null]	0315a2c104510f56a53a668e9b409036
[null]	6bc4100ea68e8be9c8c6d11a02d255...
[null]	6bc6b5bd63b62007466af17bf1073d03
[null]	7bbfcafe0fffad36b5cc66e070d44c8c
[null]	1ff224fce0309cb5499f0e18840f1d26
[null]	7bc34b1f5a6decc21166a0a4497b3226
[null]	030da5c15de5b0161ac319adda5462...
[null]	030f03de618dae51b4f67d5960e12c30
[null]	1fe652d78164472c6792fe5cee36eba2
[null]	7bb5d26e87d22cf1a432b9e49d941438
[null]	03129469d78432a8941fffc9bf3b7386
[null]	0313bf1685011280874e47e4714b98f9

And continue to till 1000 records.

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

Query:

```
Select count(id) from rnc_rna_precomputed
Where is_active is TRUE AND update_date < '2022-01-01';
```

Output:

	count bigint
1	55930772

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

Query:

```
Select rna_type, count(id)
From rnc_rna_precomputed
Where ((update_date > '2010-12-31' AND update_date < '2012-01-01')
  OR (update_date > '2013-12-31' AND update_date < '2015-01-01')
  OR (update_date > '2015-12-31' AND update_date < '2017-01-01')
  OR (update_date > '2019-12-31' AND update_date < '2021-01-01'))
  AND (rna_type = 'snoRNA' OR rna_type = 'tRNA')
Group By rna_type;
```

Output:

	rna_type character varying (500)	count bigint
1	snoRNA	70872
2	tRNA	844505

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

Query:

Select display_name, min_length from rnc_database

Where min_length NOT IN (15,100,200,300,400);

Output:

	display_name character varying (60)	min_length bigint
1	ENA	10
2	GENCODE	32
3	MGnify	27
4	GeneCards	16
5	RDP	1337
6	snoRNA Database	45
7	Rfam	24
8	TAIR	19
9	PSICQUIC	21
10	WormBase	17
11	FlyBase	18
12	snoDB	25
13	SRPDB	30
14	ZFIN	83
15	tmRNA Website	59
16	Expression Atlas	72
17	Ensembl Fungi	19
18	ZWD	21
19	Ensembl Protists	11
20	GtRNAdb	53
21	5SrRNAdb	95
22	Modomics	54
23	MalaCards	16

	display_name character varying (60)	min_length bigint
24	RiboVision	30
25	Ensembl	20
26	Ensembl Metazoa	18
27	PLncDB	199
28	MGI	21
29	Ensembl/GENCODE	32
30	PDBe	10
31	PomBase	47
32	SGD	58
33	HGNC	33
34	TarBase	16
35	Ribocentre	28
36	NONCODE	201
37	REDIPortal	0
38	CRW	107
39	lncRNAdb	61
40	EVLncRNAs	199
41	snOPY	42
42	MirGeneDB	20
43	RGD	49
44	Greengenes	1253
45	dictyBase	32
46	LncBook	54

47	LncBase	17
48	IntAct	18

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.**

Query:

So basically, this query has two parts.

a. Select r.region_name as myregions From rnc_sequence_regions r
left join rnc_rna_precomputed p on r.urs_taxid=p.id
Where p.is_active is TRUE
Order By r.id
Limit 500;

Output:

	myregions text
1	URS00000003C7_4896@I/2362324-2362771:+
2	URS000000086E_80884@CACQ02003602/5181-5237:+
3	URS000000086E_80884@CACQ02008496/758-814:+
4	URS000000086E_80884@CACQ02008999/276-332:-
5	URS0000001578_403677@supercont1.68/74803-74863:+
6	URS0000001C12_332648@10/2349948-2350037:+
7	URS0000001C12_332648@5/2159942-2160020:-
8	URS0000001C12_332648@8/1272226-1272302:-
9	URS000000224D_5855@10/12361-13959:-
10	URS000000224D_5855@2/146582-148180:-
11	URS000000224D_5855@3/806003-807601:+
12	URS000000246A_80884@CACQ02003602/5251-6018:+
13	URS000000246A_80884@CACQ02008496/828-1389:+
14	URS000000246A_80884@CACQ02009340/233-794:+
15	URS0000002F4A_5823@12/1587712-1587794:-
16	URS0000003142_284591@B/1186533-1186628:-
17	URS0000003D30_4896@I/3754985-3755801:-
18	URS0000003D30_4896@I/3786530-3787370:+
19	URS0000003D30_4896@II/1611692-1612508:-
20	URS0000003D30_4896@II/1635374-1636190:+
21	URS0000003D30_4896@III/1089166-1089994:-
22	URS0000003D30_4896@III/1108951-1109779:+
23	URS0000003DC0_4896@I/149658-149776:+

Continue to till 500 records.

b. Select chromosome, count(exon_count) as ExonCount From rnc_sequence_regions
Where Exists (Select r.region_name as myregions From rnc_sequence_regions r
left join rnc_rna_precomputed p on r.urs_taxid=p.id
Where p.is_active is TRUE
Order By r.id
Limit 500)
Group By Chromosome;

Output:

	chromosome text	exoncount bigint
1	AGKD04104609.1	1
2	VUNX01061108.1	2
3	KE928300.1	1
4	AGKD04301638.1	51
5	NOIK01000067.1	1
6	BHFT01003808.1	3
7	JAHLQT010018911.1	6
8	CAAAFN010000081.1	1
9	GL345627.1	4
10	LFEI02000293.1	1
11	AGKD04425150.1	4
12	ABRO02099192.1	2
13	PEKY01003615.1	3
14	PTFD01000861.1	1
15	AGKD04247669.1	34
16	scaffold_15645	7
17	KI636991.1	1
18	CAJNNU010000162.1	10
19	Scfld02_1185	1
20	scaffold_114618	4
21	NW_026047066.1	23
22	JH583770.1	14
23	NJHO01006013.1	9
24	scaffold_10918	1

Total rows: 1000 of 575770 Query complete 00:07:57.586