CS586 Grad Project Milestone 2: 20 Questions
Ted Herring
February 6, 2020

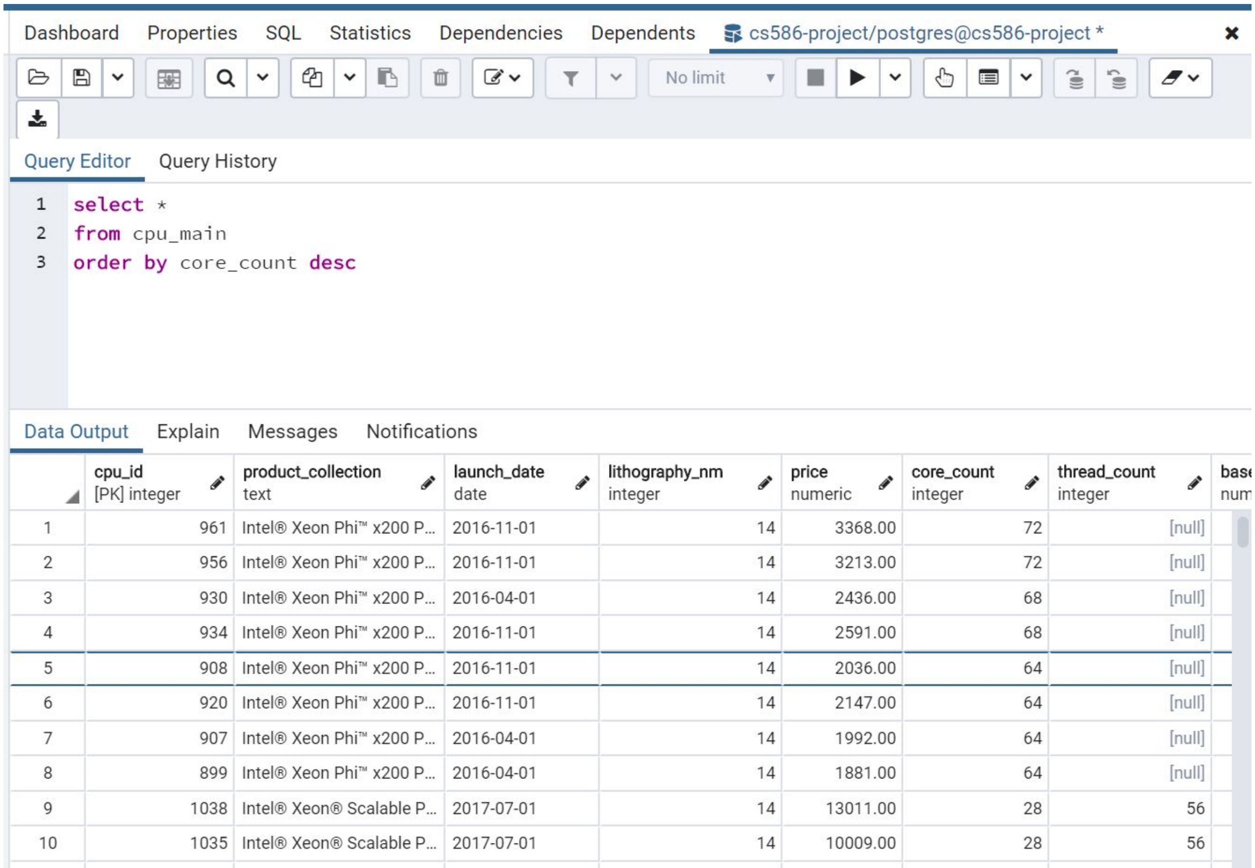
This is the final database project for cs586, Database Management Systems. For this project I took a dataset from Kaggle.com. My dataset was made up of a variety of data for CPUs and GPUs.

Ultimately I think this project went well, getting the data, cleaning it up and getting it into tables went really well. I hosted this db on google cloud, and that worked well, except when I deleted my vm by mistake and had to remake the database.

I do think I had high hopes for this dataset yielding some fairly interesting and meaningful information. Though once I got my data cleaned up and tables built I started to see the limitations on the dataset. This idea persisted as I got further into my queries.

Really, I see the planning and construction of a database is very important. I feel having a coherent shared naming convention among tables is very important for query writing, also keeping attribute names brief and clear is also important. Also, date formatting is something I really had to work with to get something usable, and still I wish I would have used separate year month day(?) attributes, could have been much easier to work with the data.

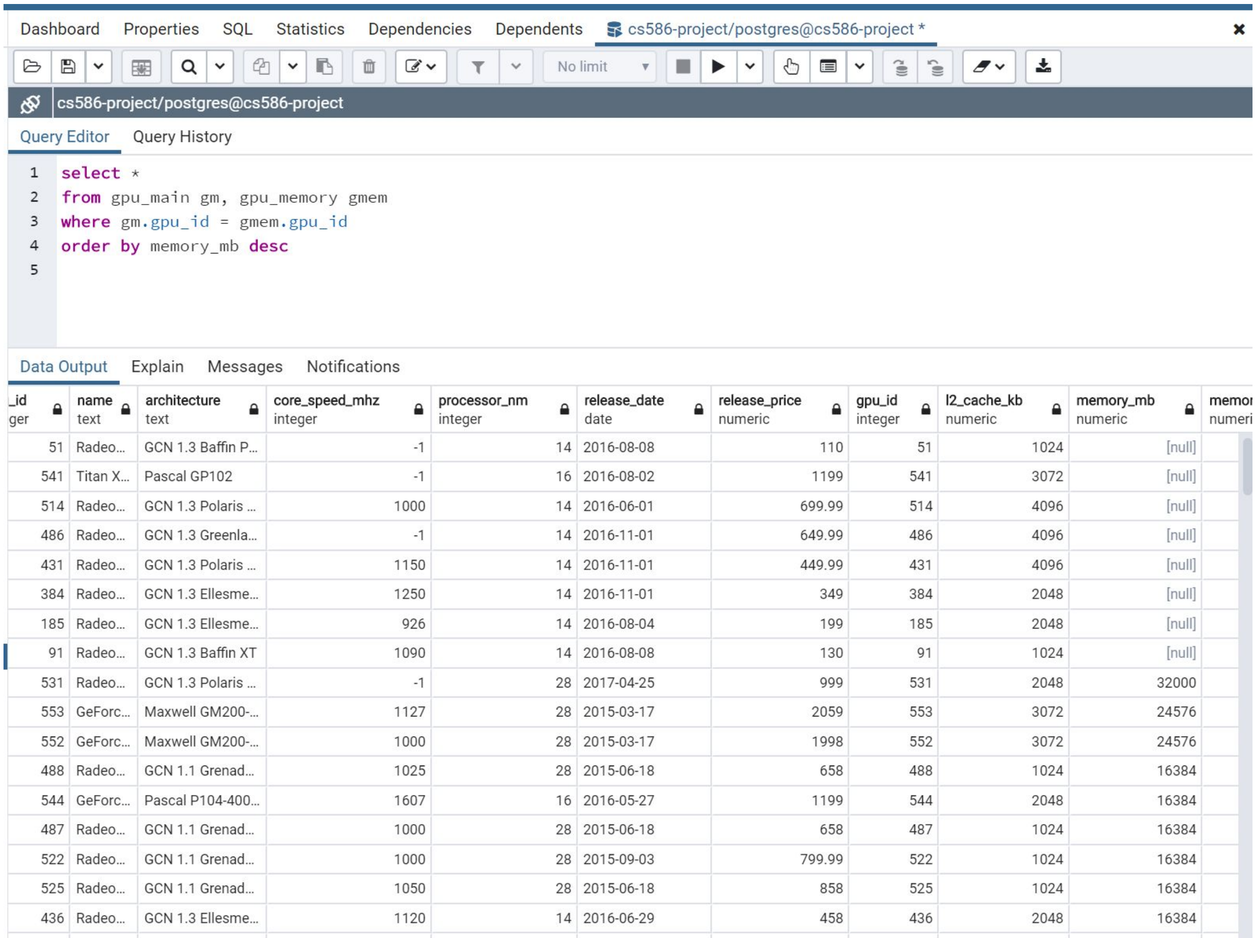
1. What is the highest number of cores available on intel cpus?



```
select *
from cpu_main
order by core_count desc
```

	cpu_id [PK] integer	product_collection text	launch_date date	lithography_nm integer	price numeric	core_count integer	thread_count integer	base num
1	961	Intel® Xeon Phi™ x200 P...	2016-11-01	14	3368.00	72	[null]	
2	956	Intel® Xeon Phi™ x200 P...	2016-11-01	14	3213.00	72	[null]	
3	930	Intel® Xeon Phi™ x200 P...	2016-04-01	14	2436.00	68	[null]	
4	934	Intel® Xeon Phi™ x200 P...	2016-11-01	14	2591.00	68	[null]	
5	908	Intel® Xeon Phi™ x200 P...	2016-11-01	14	2036.00	64	[null]	
6	920	Intel® Xeon Phi™ x200 P...	2016-11-01	14	2147.00	64	[null]	
7	907	Intel® Xeon Phi™ x200 P...	2016-04-01	14	1992.00	64	[null]	
8	899	Intel® Xeon Phi™ x200 P...	2016-04-01	14	1881.00	64	[null]	
9	1038	Intel® Xeon® Scalable P...	2017-07-01	14	13011.00	28	56	
10	1035	Intel® Xeon® Scalable P...	2017-07-01	14	10009.00	28	56	

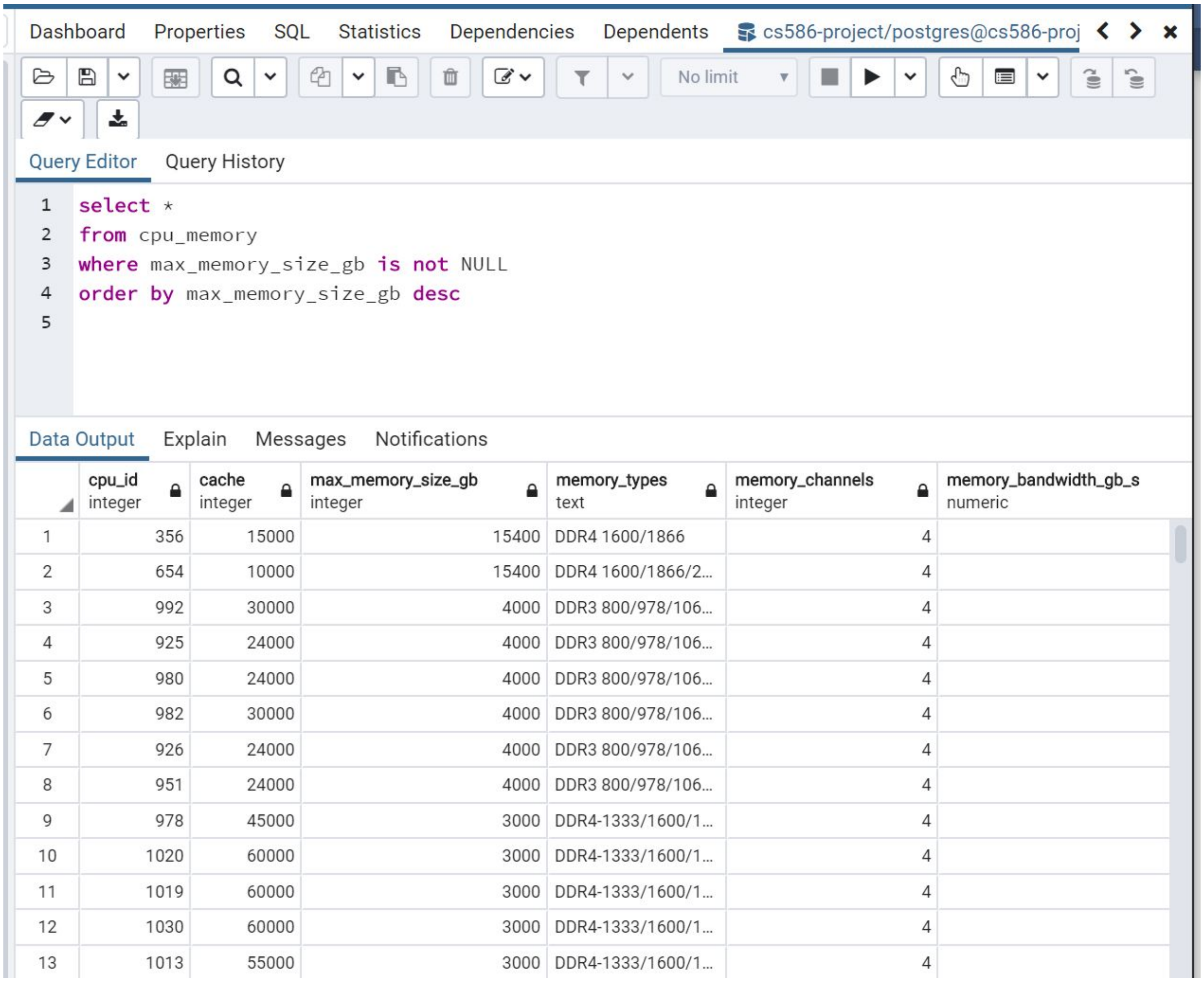
2. What is the largest amount of ram available on the gpus



```
select *
from gpu_main gm, gpu_memory gmem
where gm.gpu_id = gmem.gpu_id
order by memory_mb desc
```

_id ger	name text	architecture text	core_speed_mhz integer	processor_nm integer	release_date date	release_price numeric	gpu_id integer	l2_cache_kb numeric	memory_mb numeric	memor numeri
51	Radeo...	GCN 1.3 Baffin P...	-1	14	2016-08-08	110	51	1024	[null]	
541	Titan X...	Pascal GP102	-1	16	2016-08-02	1199	541	3072	[null]	
514	Radeo...	GCN 1.3 Polaris ...	1000	14	2016-06-01	699.99	514	4096	[null]	
486	Radeo...	GCN 1.3 Greenla...	-1	14	2016-11-01	649.99	486	4096	[null]	
431	Radeo...	GCN 1.3 Polaris ...	1150	14	2016-11-01	449.99	431	4096	[null]	
384	Radeo...	GCN 1.3 Ellesme...	1250	14	2016-11-01	349	384	2048	[null]	
185	Radeo...	GCN 1.3 Ellesme...	926	14	2016-08-04	199	185	2048	[null]	
91	Radeo...	GCN 1.3 Baffin XT	1090	14	2016-08-08	130	91	1024	[null]	
531	Radeo...	GCN 1.3 Polaris ...	-1	28	2017-04-25	999	531	2048	32000	
553	GeForc...	Maxwell GM200-...	1127	28	2015-03-17	2059	553	3072	24576	
552	GeForc...	Maxwell GM200-...	1000	28	2015-03-17	1998	552	3072	24576	
488	Radeo...	GCN 1.1 Grenad...	1025	28	2015-06-18	658	488	1024	16384	
544	GeForc...	Pascal P104-400...	1607	16	2016-05-27	1199	544	2048	16384	
487	Radeo...	GCN 1.1 Grenad...	1000	28	2015-06-18	658	487	1024	16384	
522	Radeo...	GCN 1.1 Grenad...	1000	28	2015-09-03	799.99	522	1024	16384	
525	Radeo...	GCN 1.1 Grenad...	1050	28	2015-06-18	858	525	1024	16384	
436	Radeo...	GCN 1.3 Ellesme...	1120	14	2016-06-29	458	436	2048	16384	

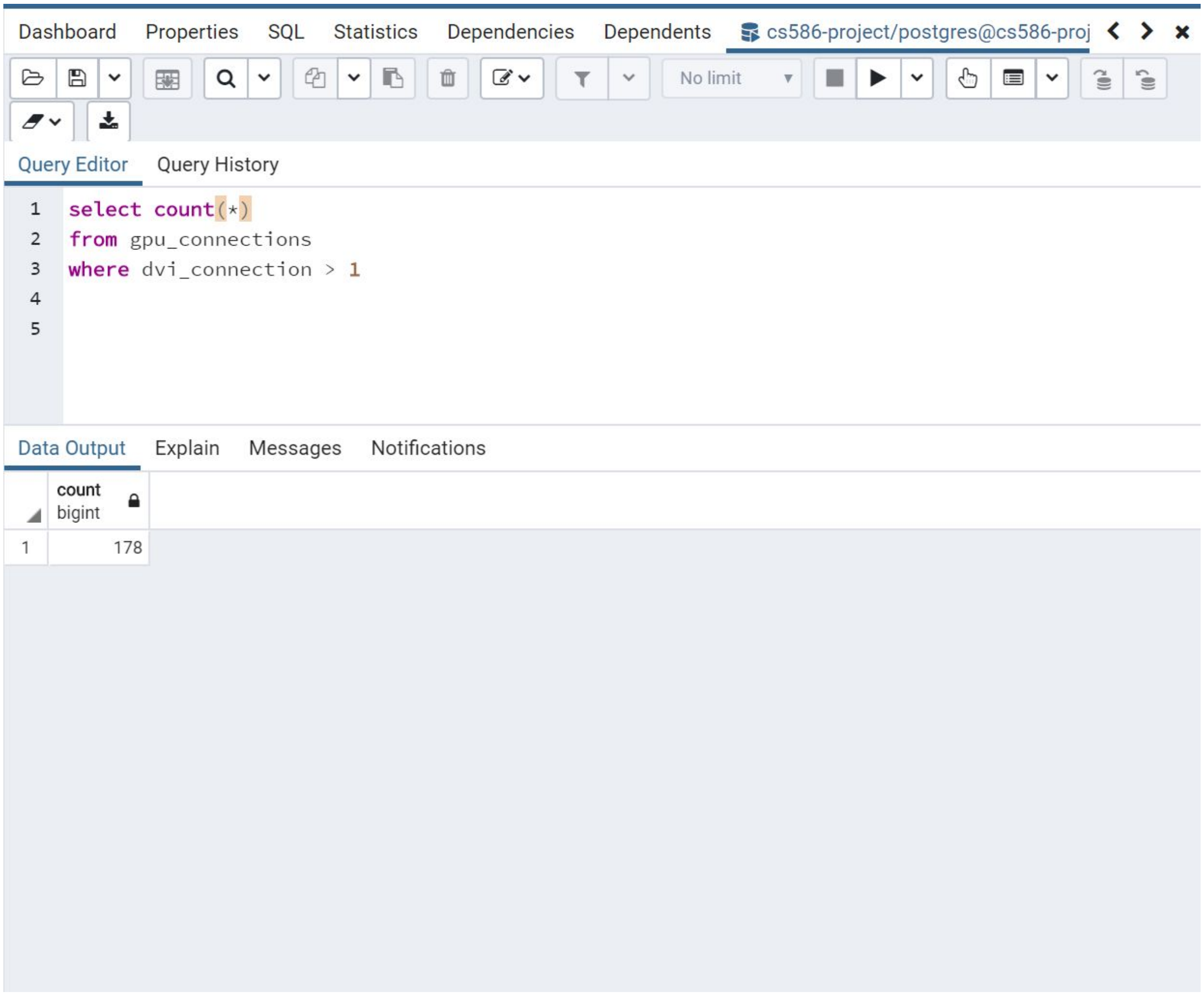
3. What is the largest Max Memory Size supported by the intel cpus



```
select *
from cpu_memory
where max_memory_size_gb is not NULL
order by max_memory_size_gb desc
```

	cpu_id integer	cache integer	max_memory_size_gb integer	memory_types text	memory_channels integer	memory_bandwidth_gb_s numeric
1	356	15000	15400	DDR4 1600/1866	4	
2	654	10000	15400	DDR4 1600/1866/2...	4	
3	992	30000	4000	DDR3 800/978/106...	4	
4	925	24000	4000	DDR3 800/978/106...	4	
5	980	24000	4000	DDR3 800/978/106...	4	
6	982	30000	4000	DDR3 800/978/106...	4	
7	926	24000	4000	DDR3 800/978/106...	4	
8	951	24000	4000	DDR3 800/978/106...	4	
9	978	45000	3000	DDR4-1333/1600/1...	4	
10	1020	60000	3000	DDR4-1333/1600/1...	4	
11	1019	60000	3000	DDR4-1333/1600/1...	4	
12	1030	60000	3000	DDR4-1333/1600/1...	4	
13	1013	55000	3000	DDR4-1333/1600/1...	4	

4. How many gpu cards have more than 1 dvi display port connections

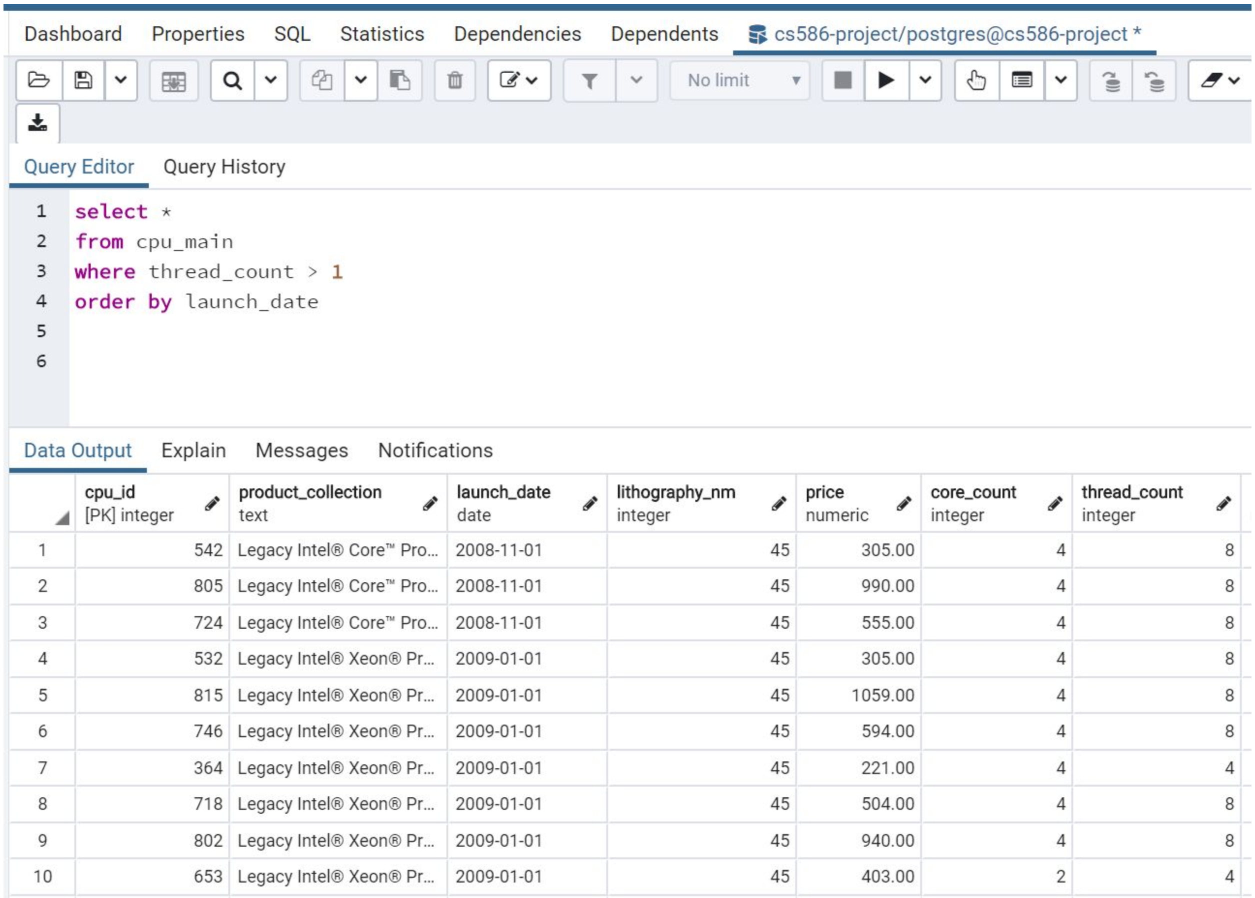


```
select count(*)
from gpu_connections
where dvi_connection > 1
```

```
  count
  bigint
1    178
```

More involved

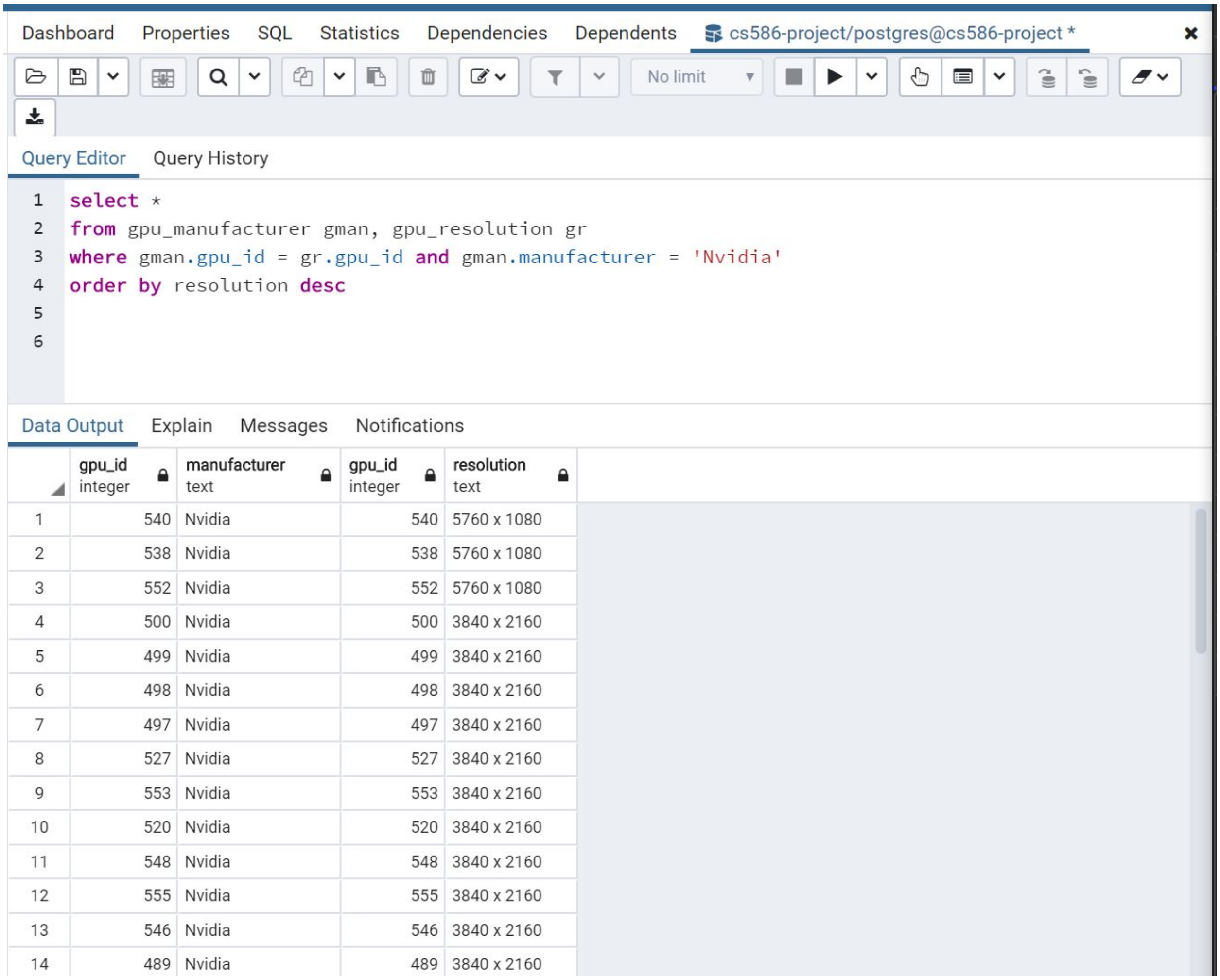
5. What is the earliest launch date of a multi threaded cpu



```
select *
from cpu_main
where thread_count > 1
order by launch_date
```

	cpu_id	product_collection	launch_date	lithography_nm	price	core_count	thread_count
	[PK] integer	text	date	integer	numeric	integer	integer
1	542	Legacy Intel® Core™ Pro...	2008-11-01	45	305.00	4	8
2	805	Legacy Intel® Core™ Pro...	2008-11-01	45	990.00	4	8
3	724	Legacy Intel® Core™ Pro...	2008-11-01	45	555.00	4	8
4	532	Legacy Intel® Xeon® Pr...	2009-01-01	45	305.00	4	8
5	815	Legacy Intel® Xeon® Pr...	2009-01-01	45	1059.00	4	8
6	746	Legacy Intel® Xeon® Pr...	2009-01-01	45	594.00	4	8
7	364	Legacy Intel® Xeon® Pr...	2009-01-01	45	221.00	4	4
8	718	Legacy Intel® Xeon® Pr...	2009-01-01	45	504.00	4	8
9	802	Legacy Intel® Xeon® Pr...	2009-01-01	45	940.00	4	8
10	653	Legacy Intel® Xeon® Pr...	2009-01-01	45	403.00	2	4

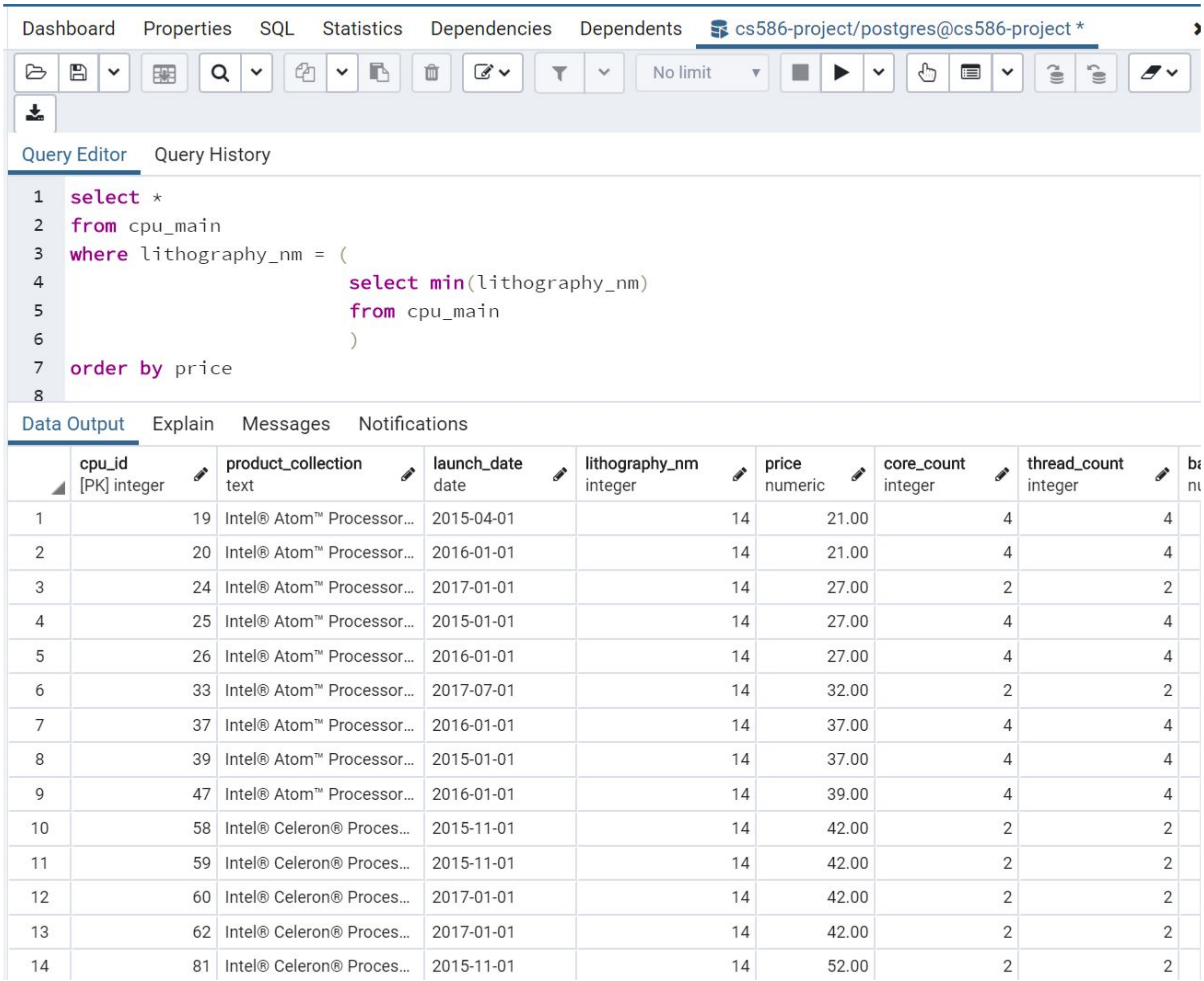
6. For Nvidia cards what is the best resolution offered



```sql
select *
from gpu_manufacturer gman, gpu_resolution gr
where gman.gpu_id = gr.gpu_id and gman.manufacturer = 'Nvidia'
order by resolution desc
```

	gpu_id integer	manufacturer text	gpu_id integer	resolution text
1	540	Nvidia	540	5760 x 1080
2	538	Nvidia	538	5760 x 1080
3	552	Nvidia	552	5760 x 1080
4	500	Nvidia	500	3840 x 2160
5	499	Nvidia	499	3840 x 2160
6	498	Nvidia	498	3840 x 2160
7	497	Nvidia	497	3840 x 2160
8	527	Nvidia	527	3840 x 2160
9	553	Nvidia	553	3840 x 2160
10	520	Nvidia	520	3840 x 2160
11	548	Nvidia	548	3840 x 2160
12	555	Nvidia	555	3840 x 2160
13	546	Nvidia	546	3840 x 2160
14	489	Nvidia	489	3840 x 2160

7. What is the least expensive cpu with the lowest Lithography

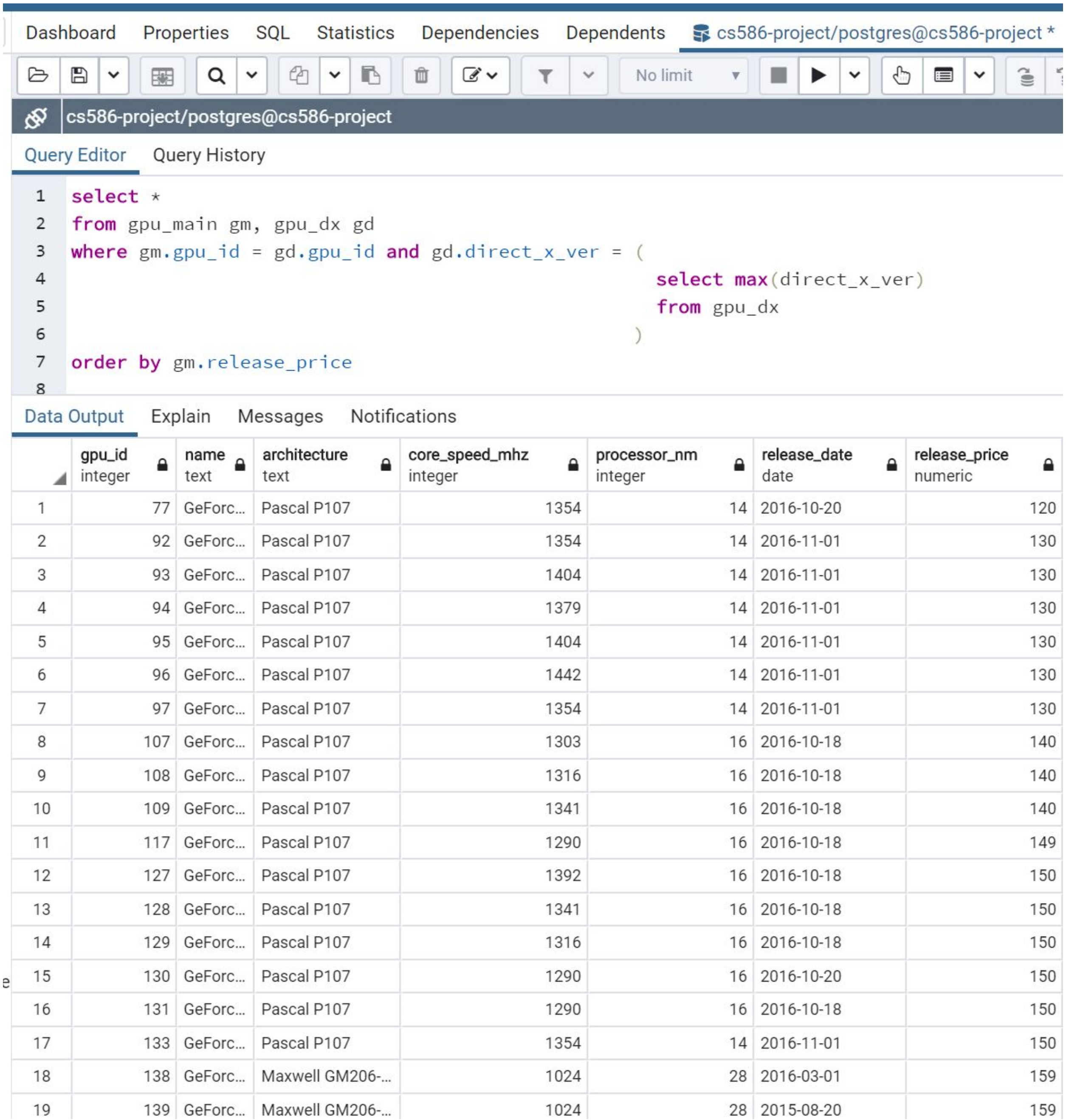


```
select *
from cpu_main
where lithography_nm = (
                        select min(lithography_nm)
                        from cpu_main
                        )
order by price

```

	cpu_id [PK] integer	product_collection text	launch_date date	lithography_nm integer	price numeric	core_count integer	thread_count integer	ba nu
1	19	Intel® Atom™ Processor...	2015-04-01	14	21.00	4	4	
2	20	Intel® Atom™ Processor...	2016-01-01	14	21.00	4	4	
3	24	Intel® Atom™ Processor...	2017-01-01	14	27.00	2	2	
4	25	Intel® Atom™ Processor...	2015-01-01	14	27.00	4	4	
5	26	Intel® Atom™ Processor...	2016-01-01	14	27.00	4	4	
6	33	Intel® Atom™ Processor...	2017-07-01	14	32.00	2	2	
7	37	Intel® Atom™ Processor...	2016-01-01	14	37.00	4	4	
8	39	Intel® Atom™ Processor...	2015-01-01	14	37.00	4	4	
9	47	Intel® Atom™ Processor...	2016-01-01	14	39.00	4	4	
10	58	Intel® Celeron® Proces...	2015-11-01	14	42.00	2	2	
11	59	Intel® Celeron® Proces...	2015-11-01	14	42.00	2	2	
12	60	Intel® Celeron® Proces...	2017-01-01	14	42.00	2	2	
13	62	Intel® Celeron® Proces...	2017-01-01	14	42.00	2	2	
14	81	Intel® Celeron® Proces...	2015-11-01	14	52.00	2	2	

8. What is the least expensive gpu with the highest direct x version



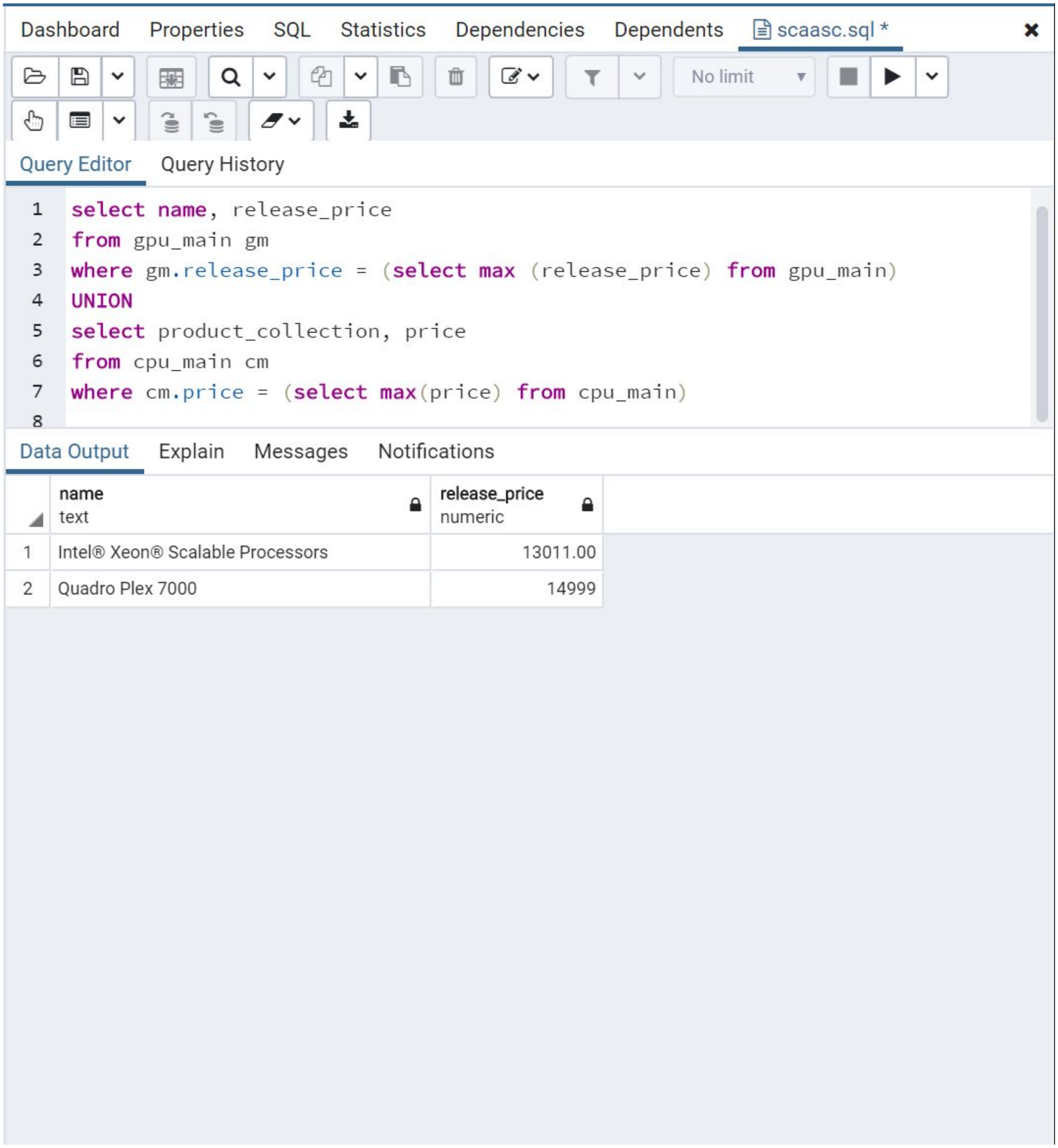
```
select *
from gpu_main gm, gpu_dx gd
where gm.gpu_id = gd.gpu_id and gd.direct_x_ver = (
                                                    select max(direct_x_ver)
                                                    from gpu_dx
                                                  )
order by gm.release_price

```

	gpu_id integer	name text	architecture text	core_speed_mhz integer	processor_nm integer	release_date date	release_price numeric
1	77	GeForc...	Pascal P107	1354	14	2016-10-20	120
2	92	GeForc...	Pascal P107	1354	14	2016-11-01	130
3	93	GeForc...	Pascal P107	1404	14	2016-11-01	130
4	94	GeForc...	Pascal P107	1379	14	2016-11-01	130
5	95	GeForc...	Pascal P107	1404	14	2016-11-01	130
6	96	GeForc...	Pascal P107	1442	14	2016-11-01	130
7	97	GeForc...	Pascal P107	1354	14	2016-11-01	130
8	107	GeForc...	Pascal P107	1303	16	2016-10-18	140
9	108	GeForc...	Pascal P107	1316	16	2016-10-18	140
10	109	GeForc...	Pascal P107	1341	16	2016-10-18	140
11	117	GeForc...	Pascal P107	1290	16	2016-10-18	149
12	127	GeForc...	Pascal P107	1392	16	2016-10-18	150
13	128	GeForc...	Pascal P107	1341	16	2016-10-18	150
14	129	GeForc...	Pascal P107	1316	16	2016-10-18	150
15	130	GeForc...	Pascal P107	1290	16	2016-10-20	150
16	131	GeForc...	Pascal P107	1290	16	2016-10-18	150
17	133	GeForc...	Pascal P107	1354	14	2016-11-01	150
18	138	GeForc...	Maxwell GM206-...	1024	28	2016-03-01	159
19	139	GeForc...	Maxwell GM206-...	1024	28	2015-08-20	159

Joined Data

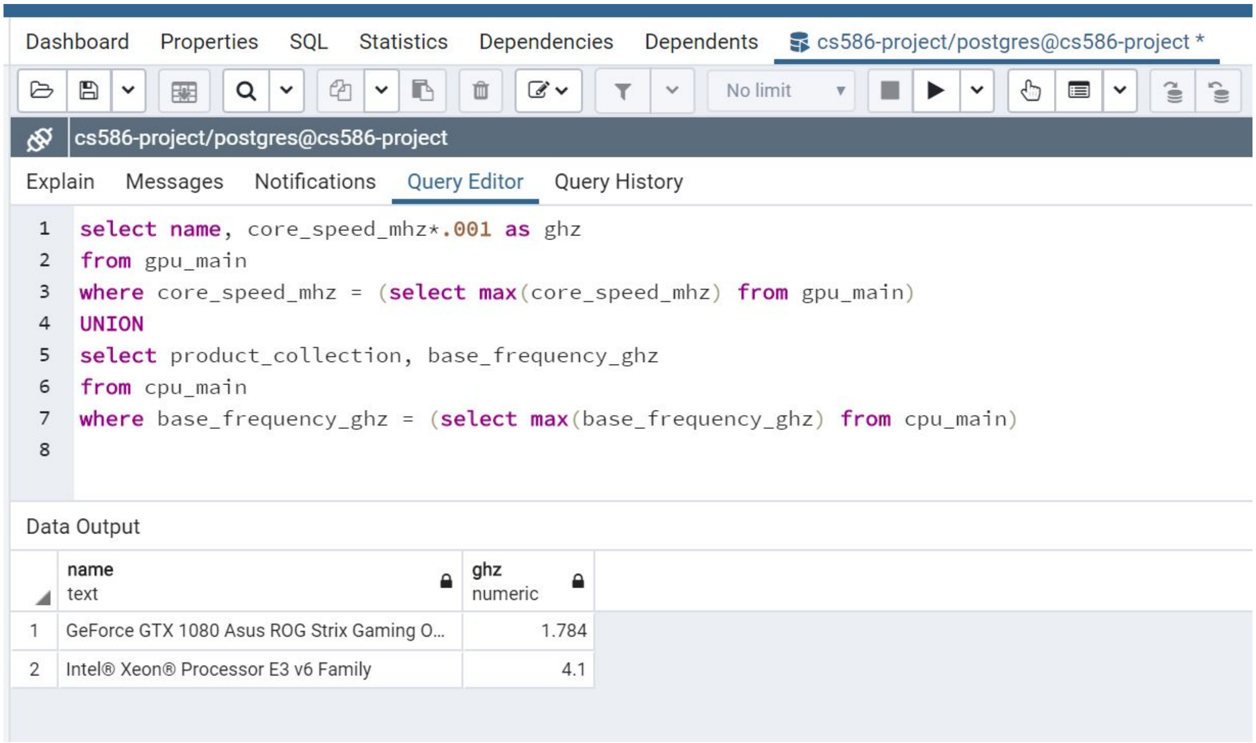
9. What is the most expensive cpu and gpu in this dataset



```sql
select name, release_price
from gpu_main gm
where gm.release_price = (select max (release_price) from gpu_main)
UNION
select product_collection, price
from cpu_main cm
where cm.price = (select max(price) from cpu_main)
```

	name text	release_price numeric
1	Intel® Xeon® Scalable Processors	13011.00
2	Quadro Plex 7000	14999

10. What are the highest core speed cpu and gpus

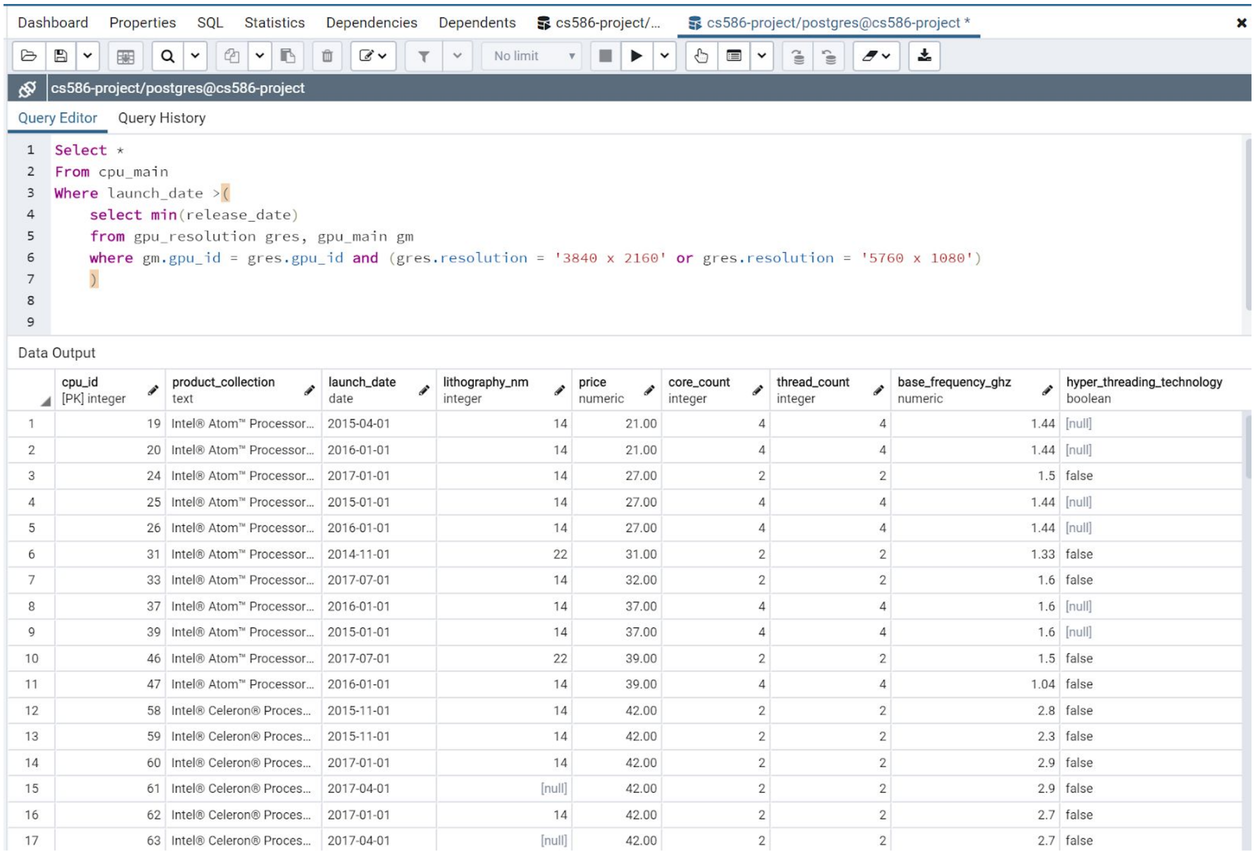


```
select name, core_speed_mhz*.001 as ghz
from gpu_main
where core_speed_mhz = (select max(core_speed_mhz) from gpu_main)
UNION
select product_collection, base_frequency_ghz
from cpu_main
where base_frequency_ghz = (select max(base_frequency_ghz) from cpu_main)

```

	name text	ghz numeric
1	GeForce GTX 1080 Asus ROG Strix Gaming O...	1.784
2	Intel® Xeon® Processor E3 v6 Family	4.1

11. What cpus came since time as gpus with resolutions 4k and above were released

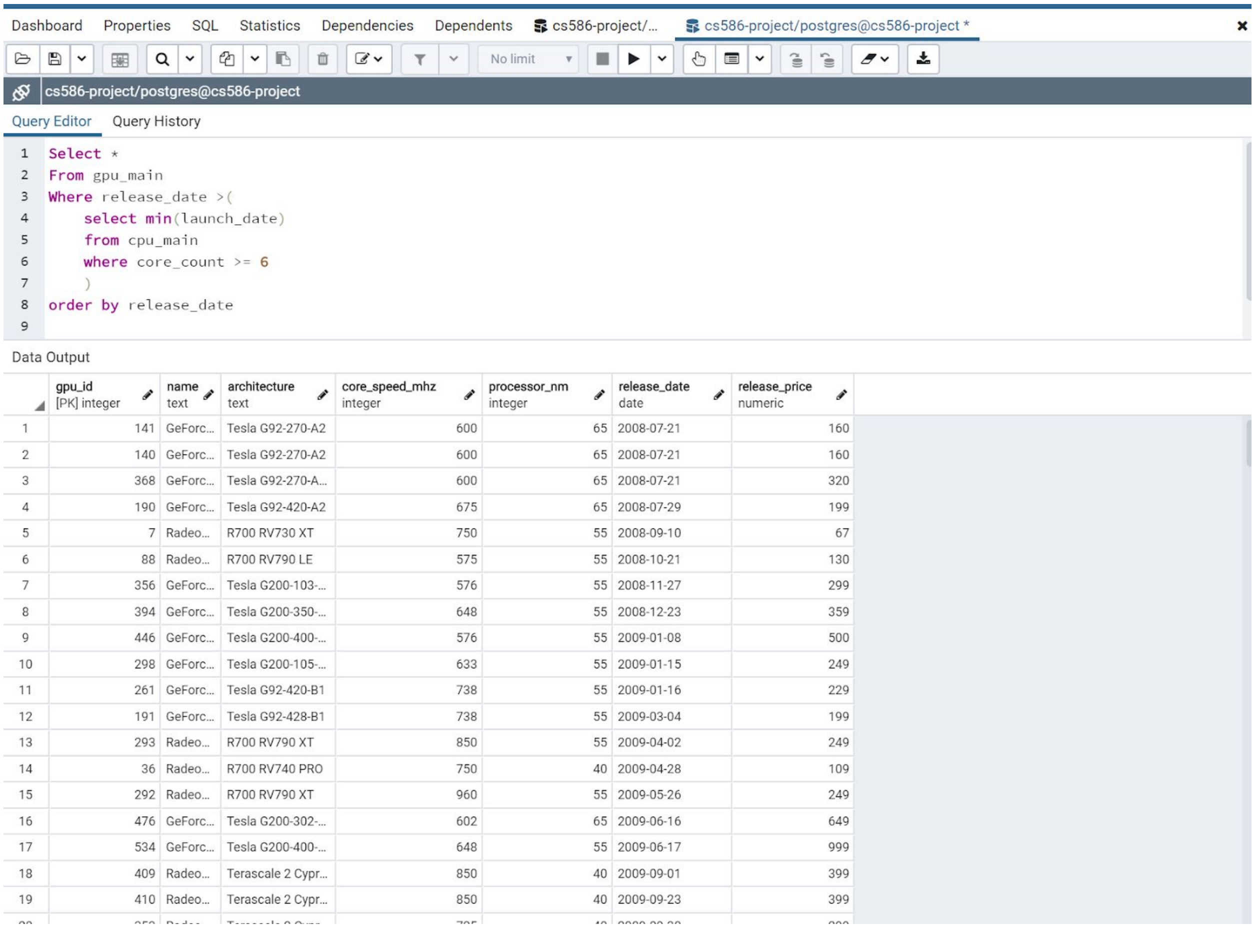


```
Select *
From cpu_main
Where launch_date >(
    select min(release_date)
    from gpu_resolution gres, gpu_main gm
    where gm.gpu_id = gres.gpu_id and (gres.resolution = '3840 x 2160' or gres.resolution = '5760 x 1080')
    )
```

Data Output

	cpu_id [PK] integer	product_collection text	launch_date date	lithography_nm integer	price numeric	core_count integer	thread_count integer	base_frequency_ghz numeric	hyper_threading_technology boolean
1	19	Intel® Atom™ Processor...	2015-04-01	14	21.00	4	4	1.44	[null]
2	20	Intel® Atom™ Processor...	2016-01-01	14	21.00	4	4	1.44	[null]
3	24	Intel® Atom™ Processor...	2017-01-01	14	27.00	2	2	1.5	false
4	25	Intel® Atom™ Processor...	2015-01-01	14	27.00	4	4	1.44	[null]
5	26	Intel® Atom™ Processor...	2016-01-01	14	27.00	4	4	1.44	[null]
6	31	Intel® Atom™ Processor...	2014-11-01	22	31.00	2	2	1.33	false
7	33	Intel® Atom™ Processor...	2017-07-01	14	32.00	2	2	1.6	false
8	37	Intel® Atom™ Processor...	2016-01-01	14	37.00	4	4	1.6	[null]
9	39	Intel® Atom™ Processor...	2015-01-01	14	37.00	4	4	1.6	[null]
10	46	Intel® Atom™ Processor...	2017-07-01	22	39.00	2	2	1.5	false
11	47	Intel® Atom™ Processor...	2016-01-01	14	39.00	4	4	1.04	false
12	58	Intel® Celeron® Proces...	2015-11-01	14	42.00	2	2	2.8	false
13	59	Intel® Celeron® Proces...	2015-11-01	14	42.00	2	2	2.3	false
14	60	Intel® Celeron® Proces...	2017-01-01	14	42.00	2	2	2.9	false
15	61	Intel® Celeron® Proces...	2017-04-01	[null]	42.00	2	2	2.9	false
16	62	Intel® Celeron® Proces...	2017-01-01	14	42.00	2	2	2.7	false
17	63	Intel® Celeron® Proces...	2017-04-01	[null]	42.00	2	2	2.7	false

12. What gpus came out since cpus with 6 or more cores arrived



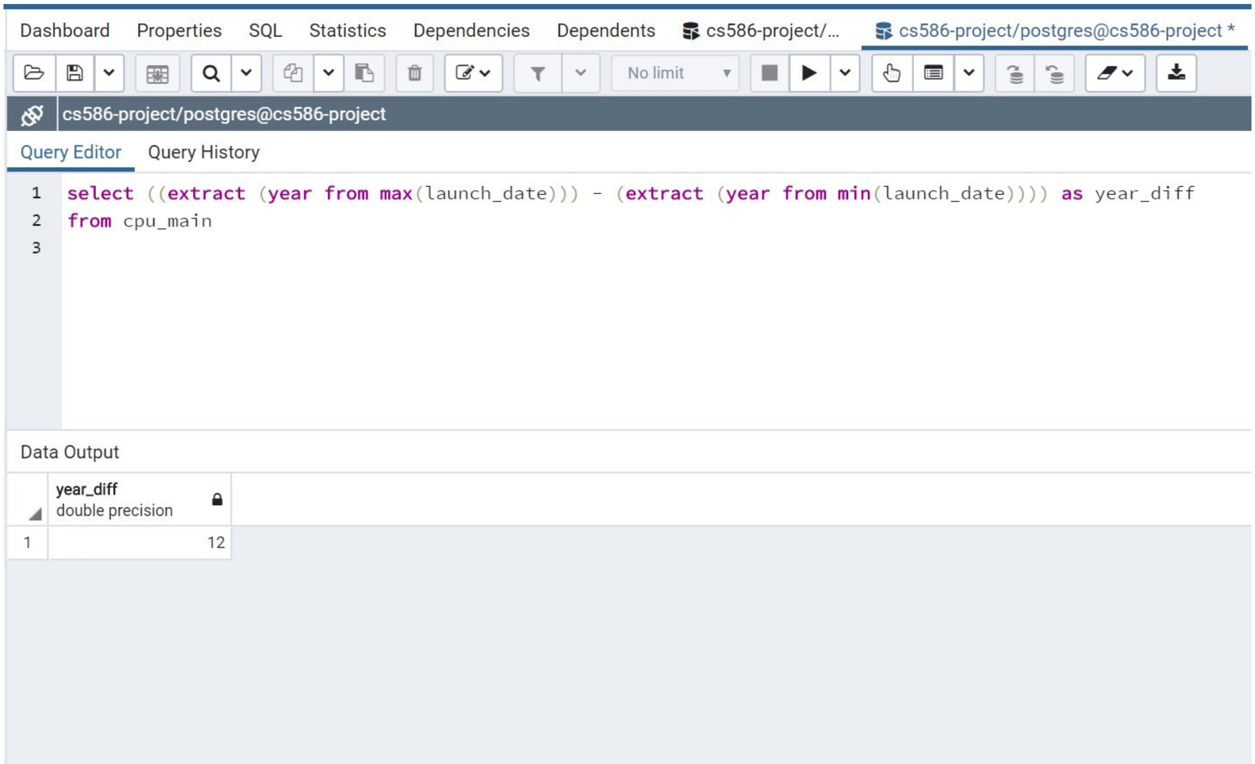
```
Select *
From gpu_main
Where release_date >(
    select min(launch_date)
    from cpu_main
    where core_count >= 6
    )
order by release_date
```

Data Output

	gpu_id [PK] integer	name text	architecture text	core_speed_mhz integer	processor_nm integer	release_date date	release_price numeric
1	141	GeForc...	Tesla G92-270-A2	600	65	2008-07-21	160
2	140	GeForc...	Tesla G92-270-A2	600	65	2008-07-21	160
3	368	GeForc...	Tesla G92-270-A...	600	65	2008-07-21	320
4	190	GeForc...	Tesla G92-420-A2	675	65	2008-07-29	199
5	7	Radeo...	R700 RV730 XT	750	55	2008-09-10	67
6	88	Radeo...	R700 RV790 LE	575	55	2008-10-21	130
7	356	GeForc...	Tesla G200-103-...	576	55	2008-11-27	299
8	394	GeForc...	Tesla G200-350-...	648	55	2008-12-23	359
9	446	GeForc...	Tesla G200-400-...	576	55	2009-01-08	500
10	298	GeForc...	Tesla G200-105-...	633	55	2009-01-15	249
11	261	GeForc...	Tesla G92-420-B1	738	55	2009-01-16	229
12	191	GeForc...	Tesla G92-428-B1	738	55	2009-03-04	199
13	293	Radeo...	R700 RV790 XT	850	55	2009-04-02	249
14	36	Radeo...	R700 RV740 PRO	750	40	2009-04-28	109
15	292	Radeo...	R700 RV790 XT	960	55	2009-05-26	249
16	476	GeForc...	Tesla G200-302-...	602	65	2009-06-16	649
17	534	GeForc...	Tesla G200-400-...	648	55	2009-06-17	999
18	409	Radeo...	Terascale 2 Cypr...	850	40	2009-09-01	399
19	410	Radeo...	Terascale 2 Cypr...	850	40	2009-09-23	399

Change over time

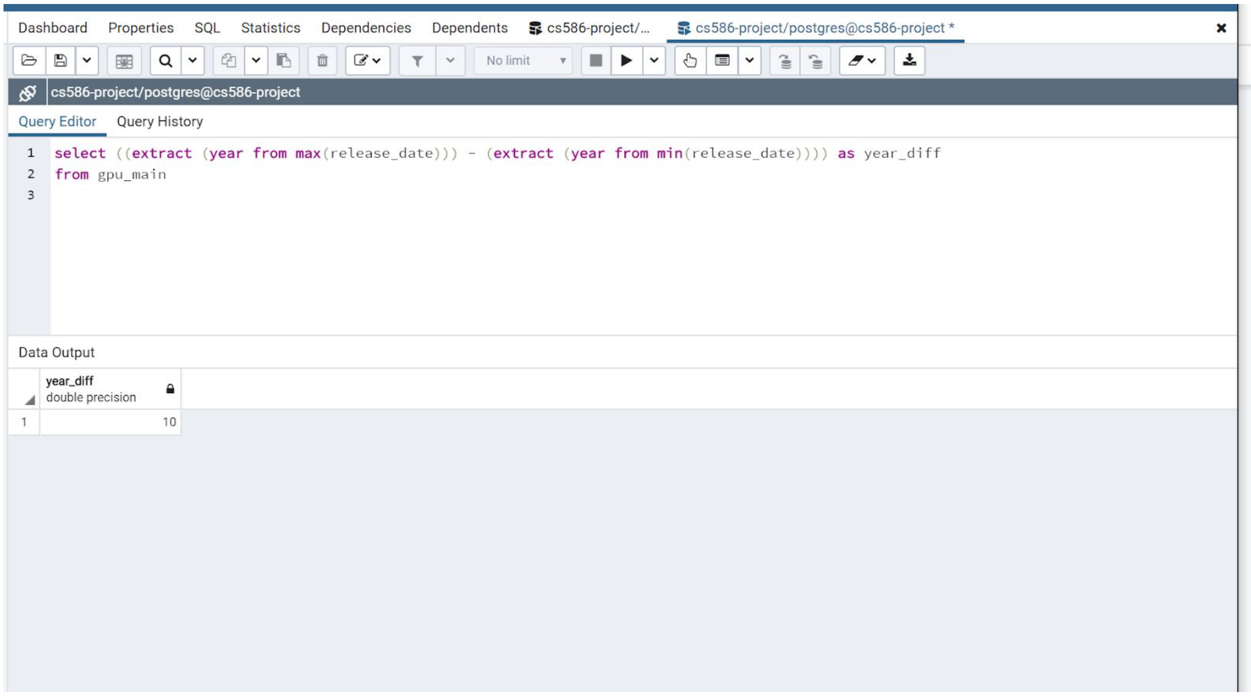
13. What is the difference in dates for the cpu table



```
select ((extract (year from max(launch_date))) - (extract (year from min(launch_date)))) as year_diff
from cpu_main
```

```
   year_diff
   double precision
1              12
```

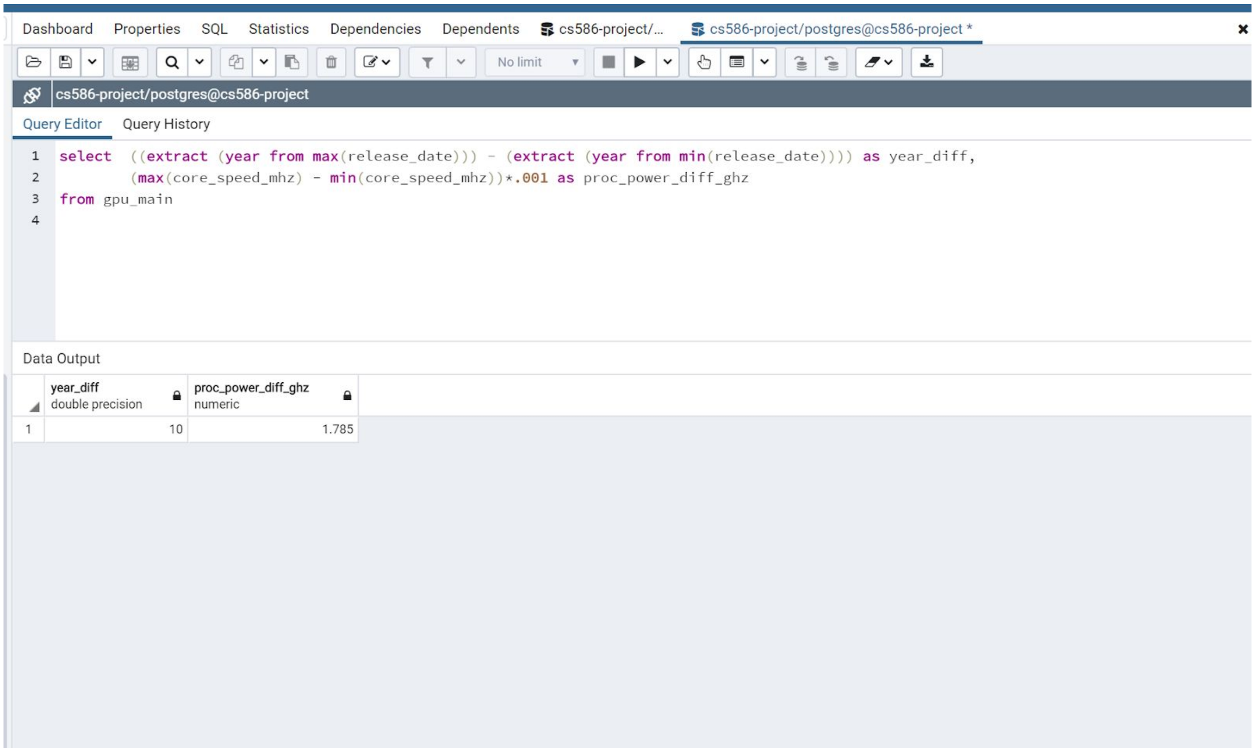
14. What is the difference in dates for the gpu table



```
select ((extract (year from max(release_date))) - (extract (year from min(release_date)))) as year_diff
from gpu_main
```

```
  year_diff
  double precision
1               10
```

15. What is the change of cpu processing power over the dates in the table

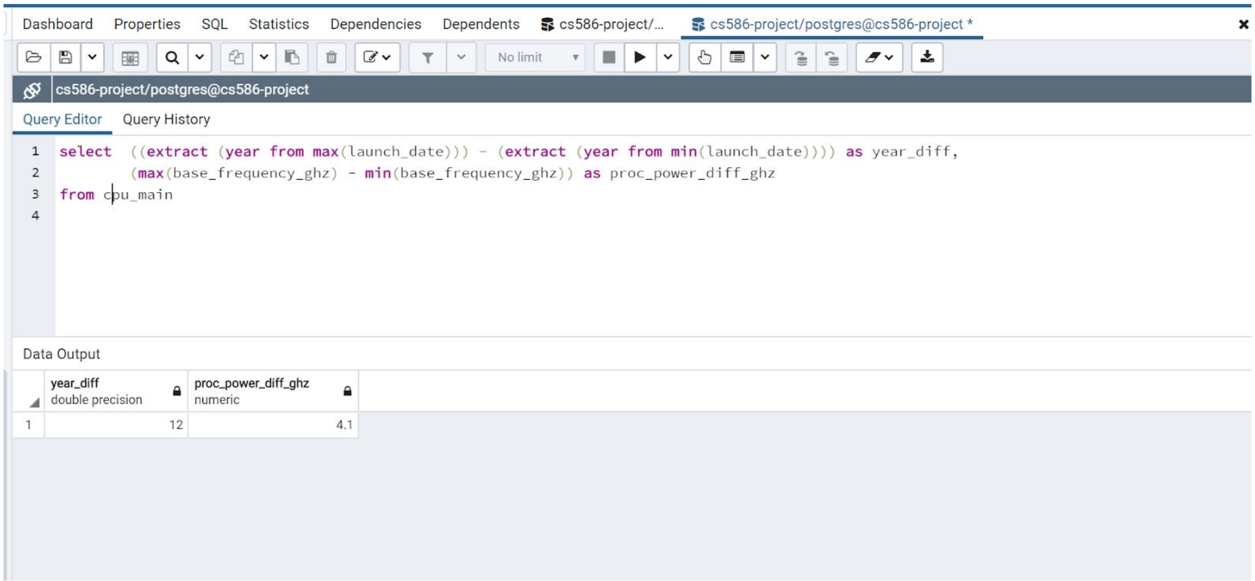


```sql
select  ((extract (year from max(release_date))) - (extract (year from min(release_date)))) as year_diff,
        (max(core_speed_mhz) - min(core_speed_mhz))*.001 as proc_power_diff_ghz
from gpu_main

```

	year_diff double precision	proc_power_diff_ghz numeric
1	10	1.785

16. What is the change of gpu processing power over the dates in the table

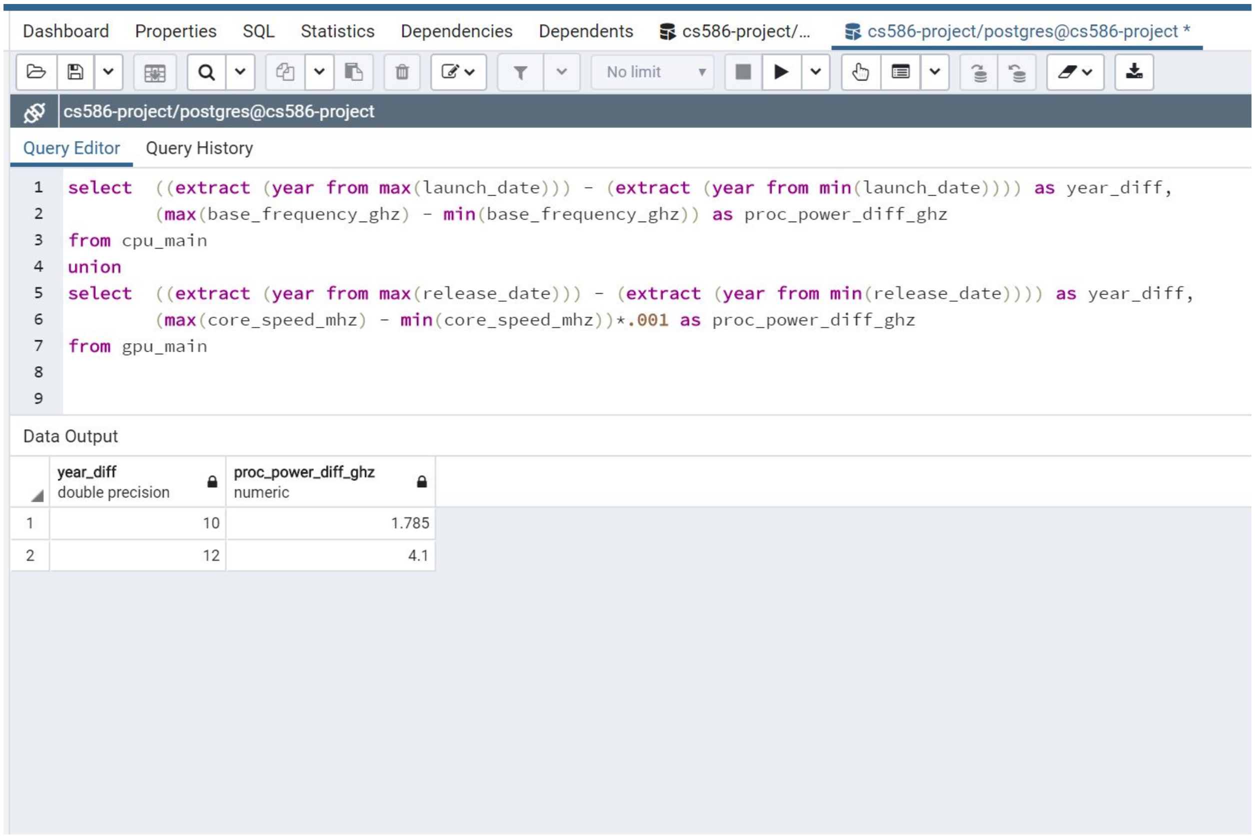


```
select  ((extract (year from max(launch_date))) - (extract (year from min(launch_date)))) as year_diff,
        (max(base_frequency_ghz) - min(base_frequency_ghz)) as proc_power_diff_ghz
from cpu_main
```

	year_diff double precision	proc_power_diff_ghz numeric
1	12	4.1

Change over time comparisons

17. How do the above changes compare

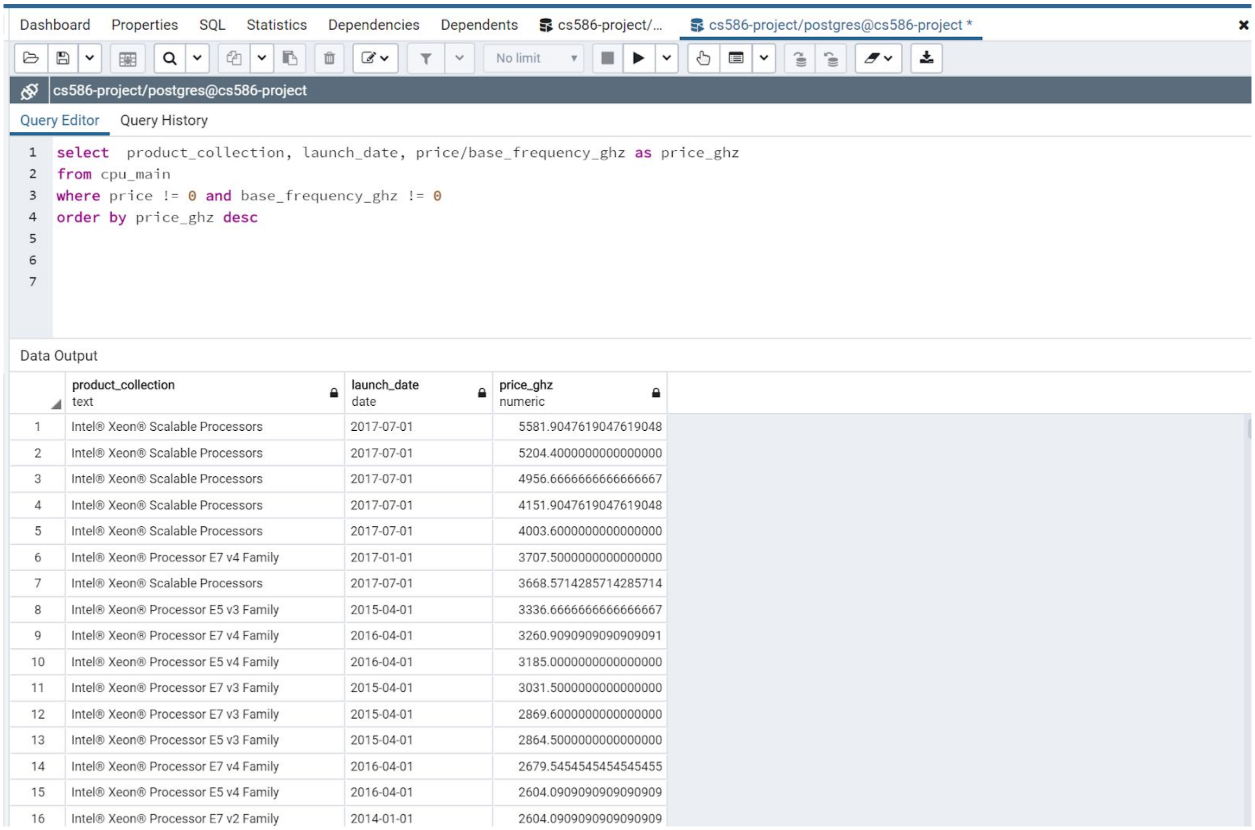


```sql
select  ((extract (year from max(launch_date))) - (extract (year from min(launch_date)))) as year_diff,
        (max(base_frequency_ghz) - min(base_frequency_ghz)) as proc_power_diff_ghz
from cpu_main
union
select  ((extract (year from max(release_date))) - (extract (year from min(release_date)))) as year_diff,
        (max(core_speed_mhz) - min(core_speed_mhz))*.001 as proc_power_diff_ghz
from gpu_main


```

	year_diff double precision	proc_power_diff_ghz numeric
1	10	1.785
2	12	4.1

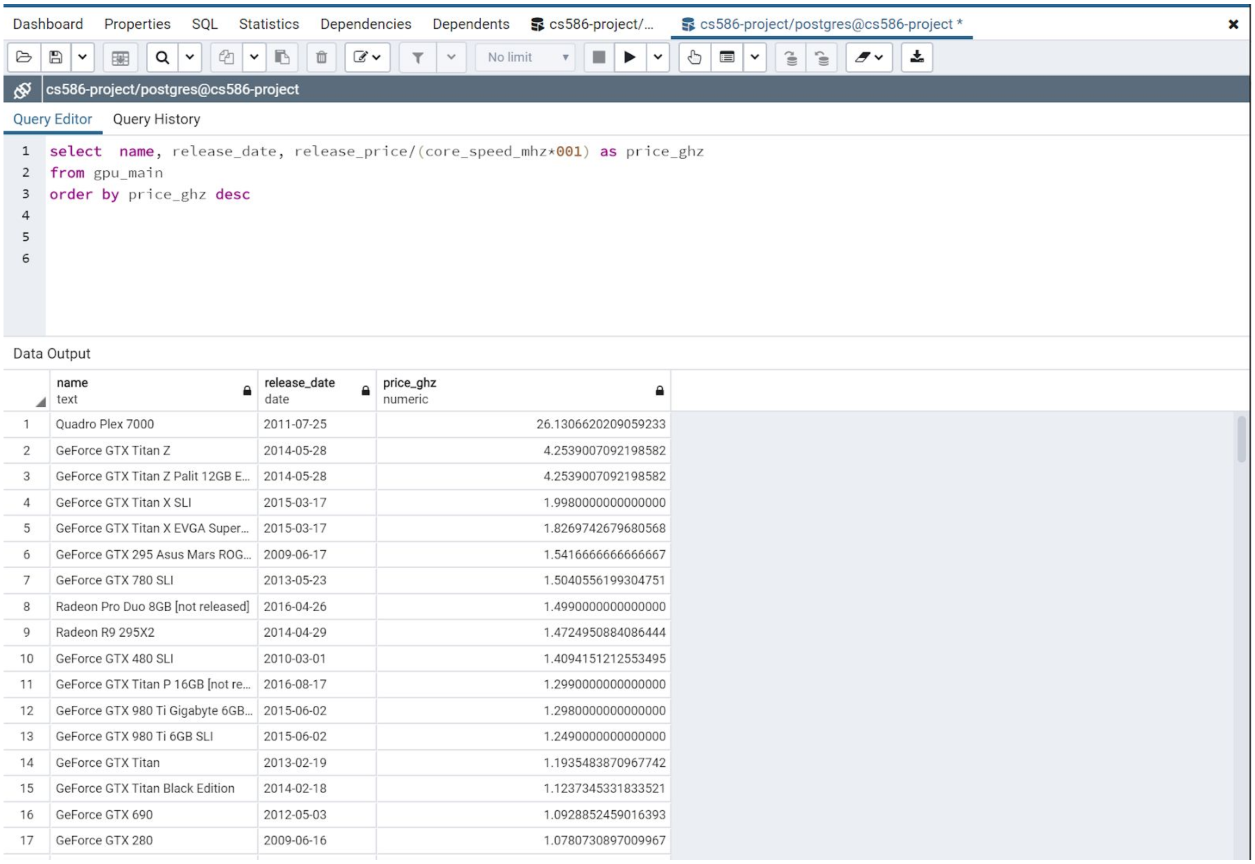
18. How has the cost with relation to cpu processing power changed



```sql
select  product_collection, launch_date, price/base_frequency_ghz as price_ghz
from cpu_main
where price != 0 and base_frequency_ghz != 0
order by price_ghz desc
```

	product_collection text	launch_date date	price_ghz numeric
1	Intel® Xeon® Scalable Processors	2017-07-01	5581.9047619047619048
2	Intel® Xeon® Scalable Processors	2017-07-01	5204.4000000000000000
3	Intel® Xeon® Scalable Processors	2017-07-01	4956.6666666666666667
4	Intel® Xeon® Scalable Processors	2017-07-01	4151.9047619047619048
5	Intel® Xeon® Scalable Processors	2017-07-01	4003.6000000000000000
6	Intel® Xeon® Processor E7 v4 Family	2017-01-01	3707.5000000000000000
7	Intel® Xeon® Scalable Processors	2017-07-01	3668.5714285714285714
8	Intel® Xeon® Processor E5 v3 Family	2015-04-01	3336.6666666666666667
9	Intel® Xeon® Processor E7 v4 Family	2016-04-01	3260.9090909090909091
10	Intel® Xeon® Processor E5 v4 Family	2016-04-01	3185.0000000000000000
11	Intel® Xeon® Processor E7 v3 Family	2015-04-01	3031.5000000000000000
12	Intel® Xeon® Processor E7 v3 Family	2015-04-01	2869.6000000000000000
13	Intel® Xeon® Processor E5 v3 Family	2015-04-01	2864.5000000000000000
14	Intel® Xeon® Processor E7 v4 Family	2016-04-01	2679.5454545454545455
15	Intel® Xeon® Processor E5 v4 Family	2016-04-01	2604.0909090909090909
16	Intel® Xeon® Processor E7 v2 Family	2014-01-01	2604.0909090909090909

19. How has the cost with relation to gpu processing power changed



```sql
select  name, release_date, release_price/(core_speed_mhz*001) as price_ghz
from gpu_main
order by price_ghz desc
```

	name text	release_date date	price_ghz numeric
1	Quadro Plex 7000	2011-07-25	26.1306620209059233
2	GeForce GTX Titan Z	2014-05-28	4.2539007092198582
3	GeForce GTX Titan Z Palit 12GB E...	2014-05-28	4.2539007092198582
4	GeForce GTX Titan X SLI	2015-03-17	1.9980000000000000
5	GeForce GTX Titan X EVGA Super...	2015-03-17	1.8269742679680568
6	GeForce GTX 295 Asus Mars ROG...	2009-06-17	1.5416666666666667
7	GeForce GTX 780 SLI	2013-05-23	1.5040556199304751
8	Radeon Pro Duo 8GB [not released]	2016-04-26	1.4990000000000000
9	Radeon R9 295X2	2014-04-29	1.4724950884086444
10	GeForce GTX 480 SLI	2010-03-01	1.4094151212553495
11	GeForce GTX Titan P 16GB [not re...	2016-08-17	1.2990000000000000
12	GeForce GTX 980 Ti Gigabyte 6GB...	2015-06-02	1.2980000000000000
13	GeForce GTX 980 Ti 6GB SLI	2015-06-02	1.2490000000000000
14	GeForce GTX Titan	2013-02-19	1.1935483870967742
15	GeForce GTX Titan Black Edition	2014-02-18	1.1237345331833521
16	GeForce GTX 690	2012-05-03	1.0928852459016393
17	GeForce GTX 280	2009-06-16	1.0780730897009967

20. With relation to Moore's law, how do these findings seem

Ultimately the answer to this question is very poorly answered by the current database. Looking at the range of the data it is too varied to produce much of an interesting or non misleading answer set. Ultimately I have a mix of different processors and graphics cards from different industry segments and this ends up skewing the data pretty bad.

Overall I find my dataset turned out to be pretty much a hodgepodge of stats, and I really see how hard it is to construct a coherent dataset to get meaningful analysis. I think taking this data from a public source (kaggle.com) does not really insure much reliability, and after cleaning up what I could and setting data to useful types a lot of the potential of the dataset was diminished.