8 USF's student table

Use the following tables when answering the questions. The tables below consistent of information from USF's undergraduate student system.

1. All columns with the same name (such as CID, SID, etc.) can be assumed to represent the same thing and join easily.
   - **SID** is an integer and is unique for each student.
   - **CID** is an integer and is unique for each class.
   - **PID** is an integer and is unique to each professor.
2. Student Table: Each row is a unique student currently enrolled at USF.
   - **BirthDate** is a date while **ParentCity** and **ParentState** represent the city and state where their parent's live while **CurrentCity** and **CurrentState** are where the student currently resides. Both are varchar
3. Classes Table: Maps currently enrolled students to what classes they enrolled in over the course of their studies.
   - **EnrollDate** is the date they they enrolled in the class.
   - **Semester** is the a varchar(10) which is equal to "Summer", "Spring" or "Fall." while **Yr** is the year, in integer form.
   - **Grade** is the grade they received. **If the student is currently taking the class or if the student withdraws the course, then grade is null.**
4. Catalog Table: Represents the list of classes available at USF for all currently enrolled undergraduates to take.
   - **Department**, **ClassName** and **Units** represent the department, name and number of units of each class.
5. WithDraw Table: Represents information about currently enrolled students who withdraw from a class.
   - **DropDate** is the date the student withdrew from the class.
6. Some other assumptions:
   - A student cannot repeat a class and a student can only withdraw from a class a single time. In other words, students get one shot at taking each class.
   - **All currently enrolled students are in the Class table.**

Figure D.8: *Student* Table, 4,525 Rows

SID	BirthDate	ParentCity	ParentState	CurrentCity	CurrentState	major
1	01-01-99	Oakland	CA	San Francisco	CA	Math
3	01-21-99	Fremont	CA	San Francisco	CA	Undeclared
2	11-08-98	Philadelphia	PA	Burlingame	CA	Comp. Sci.

Figure D.9: *Classes* Table, 72,485 Rows

SID	CID	EnrollDate	Semester	Yr	Grade
12	101	01-11-2015	Spring	2015	3.7
12	109	01-11-2015	Spring	2015	
12	800	01-10-2016	Spring	2016	0.0
12	1923	01-10-2016	Spring	2016	4.0
12	111	08-15-2016	Fall	2016	
12	546	08-15-2016	Fall	2016	
12	999	08-15-2016	Fall	2016	

Figure D.10: *Catalog* Table, 1,288 Rows

CID	PID	Department	ClassName	units
800	12	Math	Calculus I	4
801	22	Math	Calculus I	4
1118	102	English	Intro. to Shakespeare	2
45	888	Physics	Freshmen Seminar	2

Figure D.11: *Withdraw* Table, 14,888 Rows

SID	CID	DropDate
12	109	03-11-2015
1114	888	03-18-2015
765	2345	10-22-2015
9022	891	05-21-2015

1. Draw a picture showing the four tables and how they connect. Make sure to pay attention to which columns match with which table.
2. How many students are there currently enrolled?

```
select
        count(distinct sid)
from
        students;
```

3. What are the top 10 currently enrolled students (SID only) in terms of number of classes ever enrolled? E.g. which SID's enrolled in the most classes.

```
select
        sid
from
        classes
group by sid
order by count(1) desc
limit 10;
```

4. For each student (SID only), report the number of distinct departments that they have ever taken classes from.

```
select
        sid, count(distinct department) as numDepts
from
        students
left join
        catalog
using(cid)
group by 1;
```

5. What are the top five currently enrolled students (SID only) in terms of number of classes withdrawn?

```
select
        sid
from
        withdraw
group by sid
order by count(1) desc
limit 5;
```

6. Which department has the most popular major?

```
select
        major
from
        student
group by 1
order by count(1) desc
limit  1;
```

7. Which department has the largest number of currently enrolled students withdrawing from their classes?

```
select
        department
from
        withdraw
left join
        catalog
using(cid)
group by department
order by count(1) desc
limit 1;
```

8. For each student (SID only), report both how many classes they are enrolled in and how many they have withdrawn from.

```
select
        sid, count(classes.sid) as numEnrolled, count(withdraw.cid) as numWithDrawn
from
        classes
left join
        withdraw
using(sid, cid)
group by 1;
```

9. How many currently enrolled students have never withdrawn from a class?

```
select
        count( distinct sid)
from
        students
where sid not in
        (select distinct sid from withdraw);
```

We could also use a JOIN to answer this question:

```
select
        students.sid
from
        students
left join
        withdraw
on students.sid = withdraw.sid
group by 1
having count(withdraw.sid) = 0;
```

10. Which currently enrolled student (SID only) has the highest percentage of their classes withdrawn?

```
select
        sid
from
        classes
left join
        withdraw
using(sid, cid)
group by 1
order by count( withdraw.dropdate)::float / count( classes.sid) desc
limit 1;
```

11. Which department (department name) had the highest average time between date enrolled and date withdrawn (in number of days), for those currently enrolled students who withdrew from a class in that department?

```
select
        department
from
        classes left join withdraw using( sid, cid)
        left join catalog using( cid )
where dropdate is not null
group by 1
order by avg( dropdate - enrollate)  desc
limit 1;
```

12. Report, for each student the number of classes that they have taken in each department. Make sure to include rows (with a count of zero) for those departments from which a student has never taken a class.

```
select
        lhs.sid
        , rhs.department
        , count(classes.sid) as numclasses
from
        (select distinct sid from classes) as lhs
cross join
        (select distinct department from catalog) as rhs
left join
        classes
        on lhs.sid = classes.sid and rhs.sid = classes.sid
group by 1;
```

13. Of those currently enrolled students who withdraw from at least one class, what is the average number of classes that they withdrew from?

```
select avg( ct )
from
        (select count(1) as ct
        from withdraws group by sid)
        as innerQ;
```

14. Which professor (PID only) had the average highest percentage of their students withdraw from a class? This should be an average over classes as professors can teach multiple courses.

```
select pid, avg( pct) as apct
from (
        select pid, cid, sum(coalesce( wd, 0))::float / count(1) as pct
        from
                (select sid, cid from classes ) as lhs
        left join
                (select sid, cid, 1 as wd from withdraw) as rhs
                using(sid, cid)
        left join
                catalog
        using(cid)
        ) as innerQ
group by 1
order by 2 desc
limit 1;
```

15. Calculate each currently enrolled student's GPA, making sure to weigh it by the number of units.

```
select
        SID, sum( units * grade ) / sum( units) as GPA
from
        classes
left join
        catalog
using( cid )
where grade is not null
group by 1;
```

16. Of all classes with more than 15 currently enrolled students, which ones have an average grade given of more than 3.5? Make sure to not include withdraws.

```
select cid
from
classes
where grade is not null
group by 1
having count(1) > 15 and avg(grade) > 3.5;
```

17. Write a single query which calculates the average GPA of currently enrolled students who (1) withdraw from more than 10% of their classes and (2) withdraw from less than 10% of their classes.

```
select avg( GPA) as avgGPA, more_than10
from
(select
        sid
        , sum( units * grade ) / sum( units) as GPA
        , case when wd::float/ccount > .1 then 1 else 0 end as more_than10
from
        (select sid, sum( withdraw.dropdate) as wd, sum(1) as ccount
                from classes left join withdraw using( sid, cid)
                where grade is not null or dropdate is not null
                group by 1 ) as studentInfo
        left join classes using(sid)
        left join catalog using(cid)
        group by 1)
as innerQ
group by more_than10;
```