QO report

Task1

```
--Q1
select name from Drinker where name = "Eve";
--Q2
select name from Drinker;
--Q3
select price from Beer cross join Serves where name = beer;
--Q4
select distinct price from Beer cross join Serves;
--Q5
select price from Beer b, Serves s where b.name = s.beer;
--Q6
select beer from Likes union select name from Beer;
--Q7
select beer from Likes intersect select name from Beer;
--Q8
select count(*) from Likes;
```

Task2

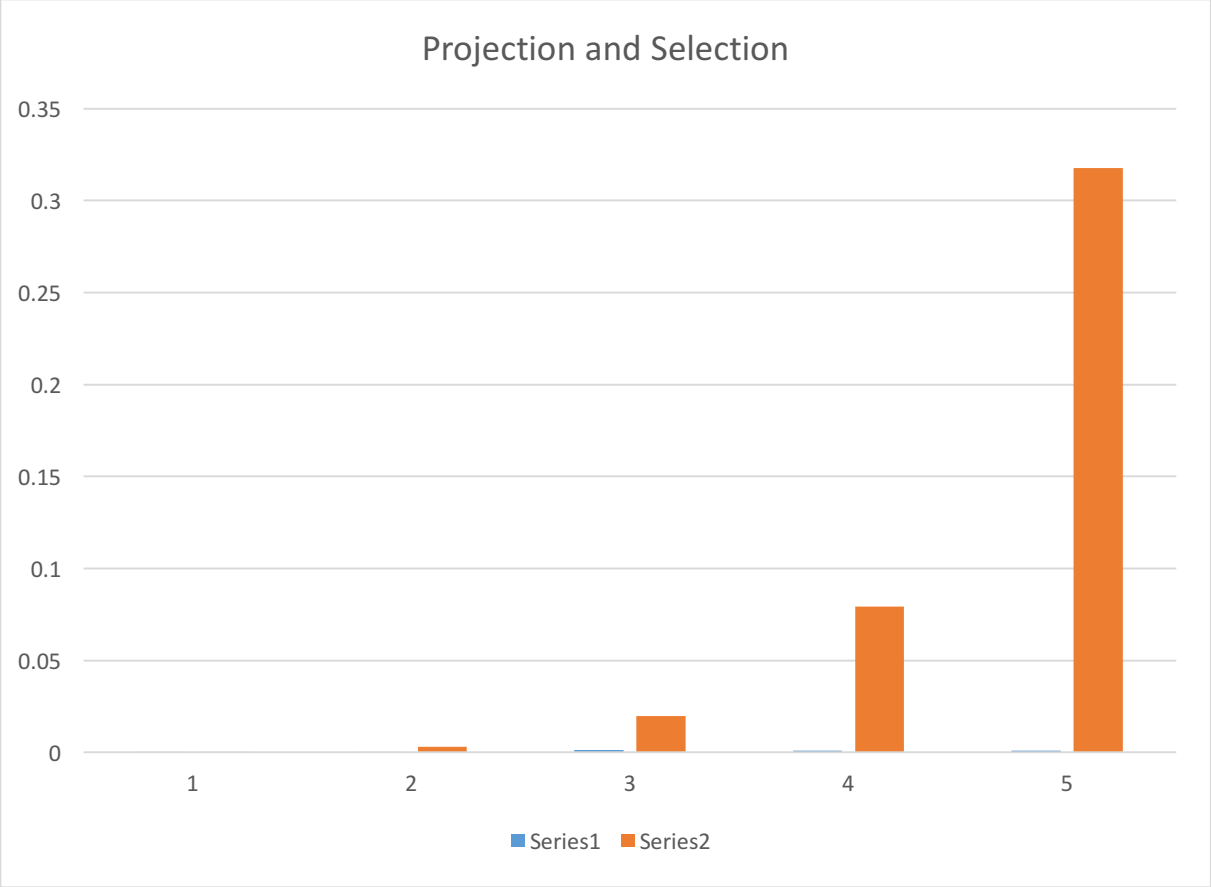
Taking into consideration of query optimization and database size, some hypotheses are made. After investigation, Drinker and Likes are larger table and Serves is the smallest table in test databases. The results would be significant if large table is used. However, for some operations it is not necessary like join, union and intersection. And that is why small tables are used in those corresponding queries.

The following hypotheses are justified by timing analysis in task3.

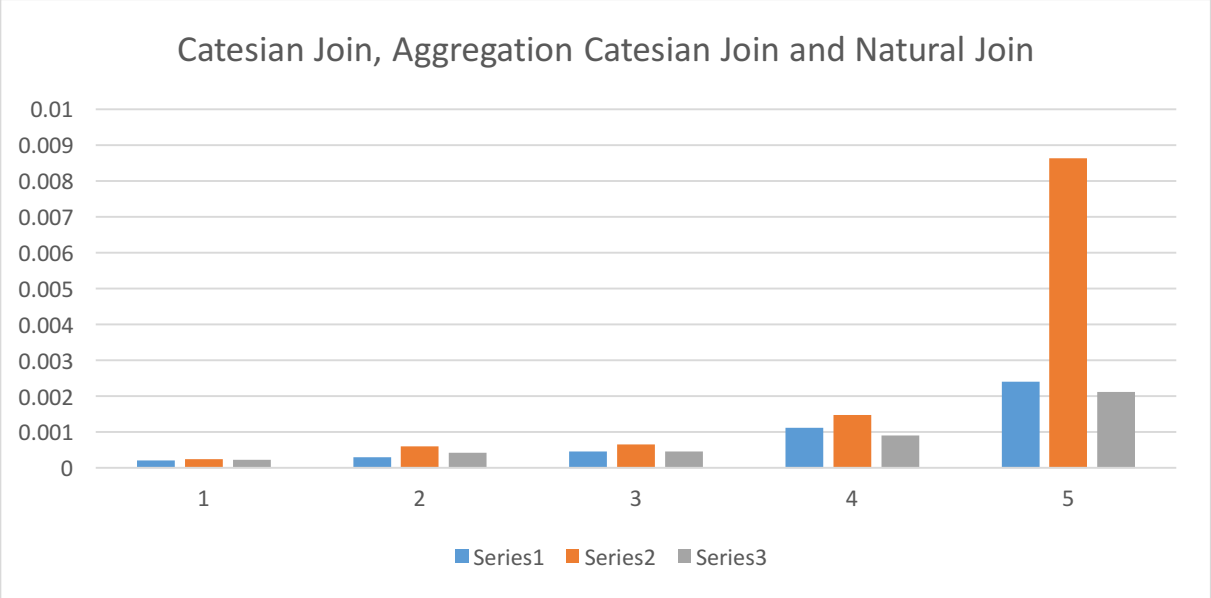
1. Projection is faster than selection. Significant result would be shown accordance to high volume of data. Index issue should be concerned as well.
2. For query 3, 4 and 5, Cartesian product is slower than natural join. The combination of Cartesian join and aggregation function cost most among those tree queries. Also size of table should be considered as well.
3. Union and intersection all remove duplicate record. In this case, union will take probably longer than intersection if the table is large. Because, union of two large table cost significantly.
4. Aggregation function performance depends on size of table.

Task3

All results are generated by query in task1. In order to have persuasive result, it takes average value among 5 consecutive executions. User time is use for justification for all comparison in task3.

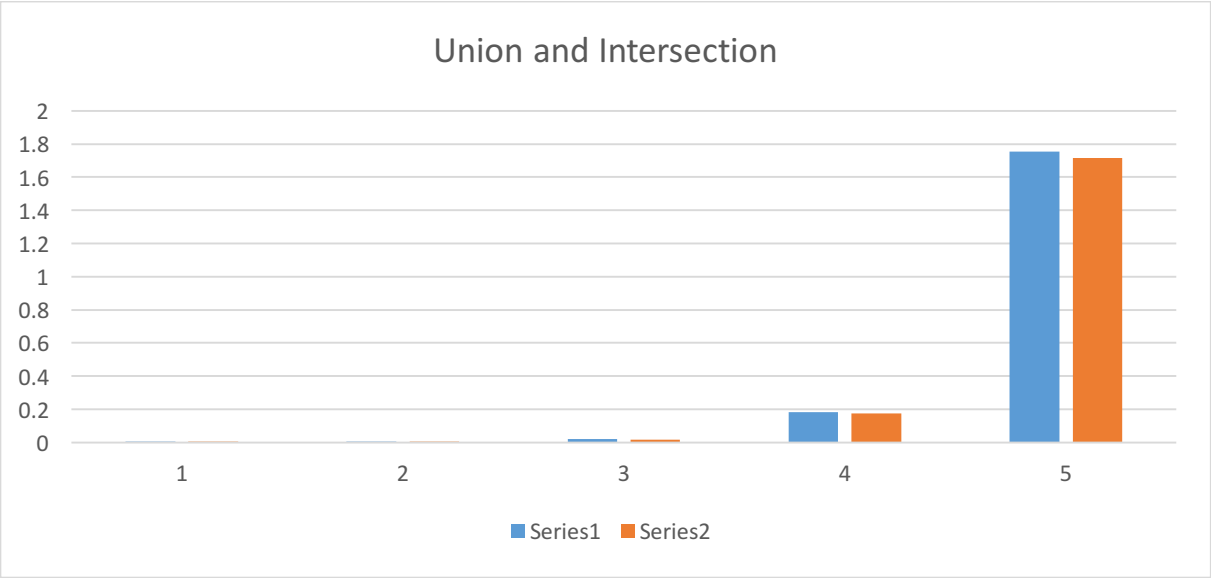


The chart above shows the comparison of projection and selection among 5 databases. According to the chart, projection is faster in small database, but much slower in large database. In this case, name is auto-index in table drinker. Therefore, no matter how data grows, cost is almost the same. As for projection, the time depend on size of table. User time increases sharply in db_10000.

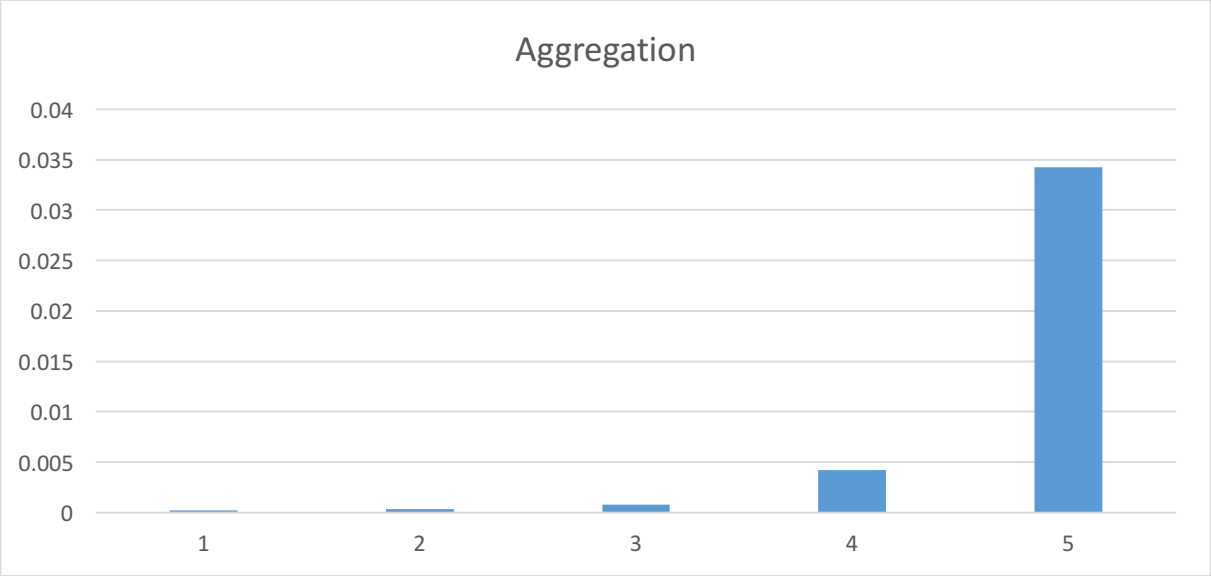


User time increases respectively for those three queries, if size of database increases. The result justify hypothesis in task 2 correctly. Aggregation Cartesian join cost the most among

three and natural join is the fastest. The performance difference is not that significant if the database is small.



Size of database affect most in this comparison case. Almost 9 times slower in db_10000 than in db_1000. And intersection operation is a little bit faster than union operation. Both of them are slow accordance to huge chunk of data.



Performance of aggregation function depends on data size. User time simply increase if size of database increase.