# Quiz1

| | |
|---|---|
| Deadline | Friday, 28 February 2020 at 11:59PM |
| Latest Submission | Monday, 24 February 2020 at 2:49PM |
| Raw Mark | 3.00/3.00 (100.00%) |
| Late Penalty | N/A |
| Final Mark | 3.00/3.00 (100.00%) |

**Question 1 (1 mark)**

Consider a relational schema with two tables *R(x,y,z)* and *S(w,x)* and an SQL query on this schema:

```
select * from R, S where R.x=S.x and R.y = 2 and S.w > 4;
```

which of the following relational algebra expressions will most likely produce the most efficient evaluation of the query.

Note that *Sel[c]R* is relation selection, *Proj[a,b]R* is relational projection, and *(R Join S)* is relational join (natural join).

You can assume that 10% of tuples in *R* have an attribute *y* with value 2, and 33% of tuples in *S* have an attribute *w* with value larger than 4.

| | |
|---|---|
| (a) ○ | Sel[y=2 and w>4](R Join S) |
| (b) ○ | Sel[y=2](Sel[w>4](R Join S)) |
| (c) ○ | Sel[y=2](R Join (Sel[w>4](S)) |
| (d) ○ | Sel[w>4]((Sel[y=2](R)) Join S) |
| (e) ◉ | (Sel[y=2](R)) Join (Sel[w>4](S)) |

✔ Your response was correct.
Mark: 1.00

Applying filters (select) before joins typically produces a much more efficient join. And often the filters themselves can make use of indexes.

**Question 2 (1 mark)**

Which of the following SQL queries will give a list of the file paths of the data files for all of the tables in the public schema? Names should be given relative to the PGDATA directory.

| | |
|---|---|
| (a) ○ | None of the queries solves the problem. |
| (b) ◉ | `select c.relname, pg_relation_filepath(c.oid) as relfile`<br>`from pg_class c join pg_namespace n on (c.relnamespace = n.oid)`<br>`where c.relkind='r' and n.nspname = 'public';` |
| (c) ○ | `select c.relname, pg_relation_filepath(c.oid) as relfile`<br>`from pg_class c join pg_namespace n on (c.relnamespace = n.oid)`<br>`where c.relkind='t' and n.nspname = 'public';` |
| (d) ○ | `select c.relname, 'data/'\|\|n.oid\|\|'/'\|\|c.oid as relfile`<br>`from pg_class c join pg_namespace n on (c.relnamespace = n.oid)`<br>`where c.relkind='t' and n.nspname = 'public';` |
| (e) ○ | `select c.relname, 'data/'\|\|c.oid\|\|'/'\|\|n.oid as relfile`<br>`from pg_class c join pg_namespace n on (c.relnamespace = n.oid)`<br>`where c.relkind='r' and n.nspname = 'public';` |

✔ Your response was correct.
Mark: 1.00

Using pg_relation_filepath() is guaranteed to produce a valid path. Tables have pg_class.relkind='r'.

**Question 3 (1 mark)**

Under the PGDATA directory are two subdirectories `base` and `global`. Under `base` are a further collection of subdirectories. The subdirectories under `base` and the `global` directory itself contain files which hold table data. What is the difference between the tables held under `base` and those held in `global`?

| | |
|---|---|
| (a) ○ | All user tables are located under **base**; all catalog tables are located in **global**. |
| (b) ○ | All catalog tables are located under **base**; all user tables are located under **global**. |
| (c) ○ | All user tables and catalog tables are located under **base**; **global** contains log tables. |
| (d) ○ | User tables and catalog tables are distributed between **base** and **global** to minimise file access costs. |
| (e) ◉ | All user tables plus some catalog tables are located under **base**; **global** contains catalog tables shared by all databases. |

✔ Your response was correct.
Mark: 1.00

The catalog tables under **base** are those containing data local to each database (e.g. pg_class, and any user-defined tables).