Lab week 12 (ISYS2120 sem2 2022)

Welcome to week 12's lab. This week covers some questions on indexing and OLAP, as well as time to work on Asst4.

A. Indexing

Consider the table from the Lab08 unidb schema:

```
CREATE TABLE UnitOfStudy (
  uoSCode       CHAR(8),
  deptId        CHAR(3)  NOT NULL,
  uoSName       VARCHAR(40) NOT NULL,
  credits       INTEGER  NOT NULL,
  PRIMARY KEY (uoSCode),
  UNIQUE (deptId, uoSName)
);
```

With this CREATE TABLE statement, a typical system will structure this table with a primary index on uoSCode (the primary key); there are no secondary indices until we declare some.

A(i) With the structure produced by the CREATE TABLE statement, write a query which the system can calculate efficiently (that is, looking at only an amount of data similar in quantity to the output of the query, rather than scanning through the whole table.

A(ii) Explain why the system will not be able to calculate the result of the query below without scanning through the whole table

```
SELECT uoSName
FROM UnitOfStudy
WHERE deptId = 'SIT';
```

A(ii) Declare an index to allow the query from A(ii) to be answered efficiently. Can you declare an index that covers the query?

B. ROLAP

Consider the star schema below, for supermarket data

```
CREATE TABLE Time
(
   time_id    CHAR(2),
   week       VARCHAR(10),
   month      VARCHAR(15),
   quarter    VARCHAR(10),
   CONSTRAINT Time_PK PRIMARY KEY (time_id)
);

CREATE TABLE Market
(
   market_id  CHAR(2),
```

```
   city       VARCHAR(30),
   state      VARCHAR(20),
   region     VARCHAR(5),
   CONSTRAINT Market_PK PRIMARY KEY (market_id)
);

CREATE TABLE Product
(
   product_id CHAR(2),
   name       VARCHAR(30),
   category   VARCHAR(20),
   price      float,
   CONSTRAINT Product_PK PRIMARY KEY (product_id)
);

CREATE TABLE Sales
(
   market_id  CHAR(2) REFERENCES Market,
   product_id CHAR(2) REFERENCES Product,
   time_id    CHAR(2) REFERENCES Time,
   sales_amt  INTEGER,
   CONSTRAINT Sales_PK PRIMARY KEY (market_id, product_id,
time_id)
);
```

A file Supermarket.sql is on Canvas, with a tiny amount of data for use in order to run this on PostgreSQL.

B(i) Consider the following CUBE query:

```
SELECT M.region, P.category, SUM(S.Sales_amt)
FROM Sales S, Market M, Product P
WHERE S.Market_Id = M.Market_Id
and S.Product_Id = P.Product_Id
GROUP BY CUBE(M.region, P.category)
```

Show the result of this query, both in relational table form as output by PostgreSQL, and also sketch how it will look as a data cube representation. In each representation, where do you find the partial sums ("marginals")?

B(ii) How would you describe in English, the question that is answered using the query of B(i)

B(iii) Write a query to slice the data, taking only sales in 'Wk-05-2006', and dicing these by Product_Id and the location's State

C. Work for asst4

Asasst4 is due at the end of this week, you should use the time in lab to make as much progress as possible. In particular, each group member should show their progress (initial answers for the subparts that they are leading) to the lab demonstrator for feedback; also members can help one another with feedback..

Before the end of the week

Before Sunday October 30, you need to finish the assessments that are due (Asst4, Quiz 12).