

# Exam Advanced Database Systems (211090)

Wednesday 23 January 2008, 13:30 – 17:00 hour (LA 3522)

The exam consists of 6 questions divided into subquestions  
It is NOT permitted to use book or notes

## 1 Indexing (6 points)

Consider a relation  $R$  with attributes  $A$  and  $B$  with the following characteristics:

- 5,000 tuples with 10 tuples per page
- A 2-level B+ tree index on attribute  $A$  with up to 100 index entries per page
- Attribute  $A$  is a candidate key of  $R$
- The values that the attribute  $A$  takes in relation  $R$  are uniformly distributed in the range 1 to 100,000.

Answer the following questions

1. Assuming that the aforesaid index on  $A$  is unclustered, estimate the number of page fetches needed to compute the query  $\sigma_{A>1000 \text{ AND } A<6000}(R)$ . Explain your reasoning.
2. What would be the cost if the above index were clustered? Again, explain your reasoning.

## 2 Query Processing (9 points)

Consider two relations  $R$  and  $S$ .  $R$  is stored in 2000 pages, 20 tuples per page.  $S$  is stored in 5000 pages, 5 tuples per page. The main memory available for processing the queries is 402 pages. There is a clustered B+-tree index on attribute  $B$  of the relation  $S$  with index depth 2. The weight of  $B$  in  $S$  is 1. Answer the following questions (ignore the costs of writing the results to disk).

1. Calculate the cost of  $R \bowtie_{A=B} S$  if we process it using a naïve *nested loop*, that is, using only two pages of main memory. Explain your answer.
2. Answer the same question when using a *block-nested loop*.
3. Answer the same question if an *index-nested loop* is used.

## 3 Object-oriented databases (8 points)

The Maharishi Yogi Community Bank in New Delhi opens bank accounts for groups of persons (each person having at most one account). Each person is permitted to do one of following two transactions on the account: a cash withdrawal and a cash deposit both in

Indian Rupees.

1. Give ODMG ODL class definitions for the classes Bankaccount, Person and Transaction. Explain your answer.
2. Give an example instantiation of the class Bankaccount that contains all data for an account.
3. Give the OQL query that presents the total amount in Rupees of the withdrawls from the bank account with account number '123'.

#### 4 XML (8 points)

Consider the following XML data

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE Features [
  <!ELEMENT Features (Feature+) >
  <!ELEMENT Feature (Description, (Polygon | Point)) >
  <!ATTLIST Feature fid ID #REQUIRED >
  <!ELEMENT Description (#PCDATA) >
  <!ELEMENT Polygon (LineString) >
  <!ATTLIST Polygon srsName CDATA #IMPLIED >
  <!ELEMENT LineString (Coordinates) >
  <!ELEMENT Point (Coordinates) >
  <!ATTLIST LineString srsName CDATA #IMPLIED >
  <!ATTLIST Point srsName CDATA #IMPLIED >
  <!ELEMENT Coordinates (#PCDATA) >
] >
<Features>
  <Feature fid="f:142">
    <Description>Zilverling</Description>
    <Polygon srsName="epsg:27354">
      <LineString srsName="epsg:27354">
        <Coordinates>
          491888.999999459,5458045.99963358 491904.999999458,5458044.99963358
          491908.999999462,5458064.99963358 491924.999999461,5458064.99963358
          491925.999999462,5458079.99963359 491977.999999466,5458120.9996336
          491953.999999466,5458017.99963357 </Coordinates>
        </LineString>
      </Polygon>
    </Feature>
  </Features>
```

1. Is the document *well-formed*? Explain your answer
2. Is the document *valid*? Explain your answer
3. Does the document have *empty elements*? If so, which element(s)?
4. Explain the difference between the attributes `fid` and `srsName`.
5. Does the document contain *namespaces*? If so, what namespace?
6. Suppose you want to list for each feature (for each building) a list of all people working in that building. Extend the DTD in such a way that it supports this.

## 5 Distributed Databases (6 points)

Suppose that we have a relation

```
Employee (SSnum, Name, Salary, Age, Title, Location)
```

which is partitioned as

```
Emp1 (SSnum, Name, Salary)
Emp2 (SSnum, Title, Age, Location)
```

where Emp1 is stored at site B and Emp2 is stored at site C. A query at site A wants the names of all managers in the accounting department whose salary is greater than their age. Answer the following question:

1. Design a multidatabase query plan and a set of SQL statements for this query.

## 6 On-Line Analytical Processing (8 points)

Consider the following schema for an on-line analytical processing (OLAP) application. The database contains for each timestamp the number of sold products of a super market. A company (e.g. Unilever) may have several brands (e.g. Dreft, Dash and Robijntje), which might have several product names (Dreft power gel, Dreft urban black, Dreft heavenly white, etc.)

```
CREATE TABLE products (
    id INTEGER,
    name VARCHAR NOT NULL,
    brand VARCHAR NOT NULL,
    company VARCHAR NOT NULL,
    PRIMARY KEY (id)
);

CREATE TABLE timestamps (
    id INTEGER,
    when CHAR(16),
    PRIMARY KEY (id)
);

CREATE TABLE transactions (
    product_id INTEGER,
    time_id INTEGER,
    number INTEGER NOT NULL,
    FOREIGN KEY (product_id) REFERENCES products(id),
    FOREIGN KEY (time_id) REFERENCES timestamps(id)
);
```

Answer the following questions.

1. Is the schema a so-called star schema? Explain your answer by using the terms “fact table” and “dimension table”.
2. In order for the schema to be suitable for on-line transaction processing (OLTP), it should not allow for any redundant data. Does the schema allow for redundancy? Explain your answer.
3. Give the SQL query that presents for each product the total number of sold items.
4. Give an SQL query that presents per product, per brand, and per company the total number of sold items. If necessary, use the OLAP extensions from the SQL:1999 standard.