University
of Victoria

University of Victoria
Final Examination
August 2014

Course Name & No.	CSC 370 (Database Systems)
Sections(s)	A01
CRN	30161
Instructor	Michael Zastre
Duration	Three (3) hours

Name	
UVic ID	

- This exam has six (6) questions on fourteen (14) pages including this cover page and the last page.
- Students must count the number of pages and report any discrepancy immediately to the invigilator.
- This exam is to be answered on the paper.
- No materials or aids of any kind are permitted. This is a closed-book exam.
- Ensure all cellphones are switched off. You must obtain permission from an invigilator to temporarily leave the examination room.
- **The total number of marks in this exam is 200.**

Question 1: SQL (40 marks)

Consider the following set of relations:

- `Emp(eno, ename, jobtitle, city)`
- `Proj(pno, pname, budget, city)`
- `Works(pno, eno, responsibility, duration)`
- `Pay(jobtitle, salary)`

For parts (a) through (e) write an SQL statement satisfying the stated goal. State all assumptions. (Even if the SQL is incorrect, some marks may be given for any analysis you perform using the relational algebra.)

(a) Create the table for the `Emp` relation.

(b) Create the table for the `Works` relation.

(c) For each city, find out how many projects are located in that city and what the total budget is over all projects in the city.

(d) For each project, find out what fraction of the budget is spent (in total) on salaries for the people working on that project, and present that answer sorted by the ascending value of the budget.

(e) List all projects in Victoria and include for each project the number of persons working on the project.

(f) Formulate the following query into the relational algebra:

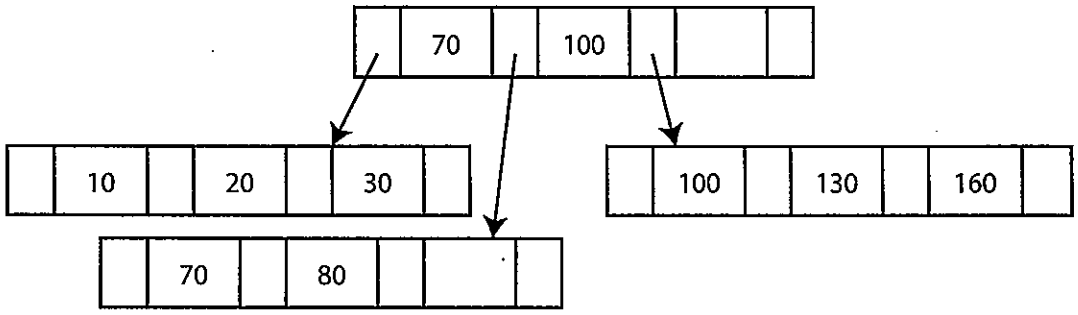
```
SELECT pname, budget
FROM   Proj, Works, Emp
WHERE  jobtitle = 'Programmer'
  AND  Works.eno = Emp.eno
  AND  Works.pno = Proj.pno;
```

Question 2: Application Tier (40 marks)

Given the schema from the previous question and the query in part (f) of that question, write a complete Python script that uses `psycopg2` to implement the query and output the results on the console. You may assume the script will connect to a server named `server10.citibooks.ca` with a user/password combination of `(remoteuser, 'GameOfThronesRulz!')`. You may also assume the database for the user already contains all four tables populated with data.

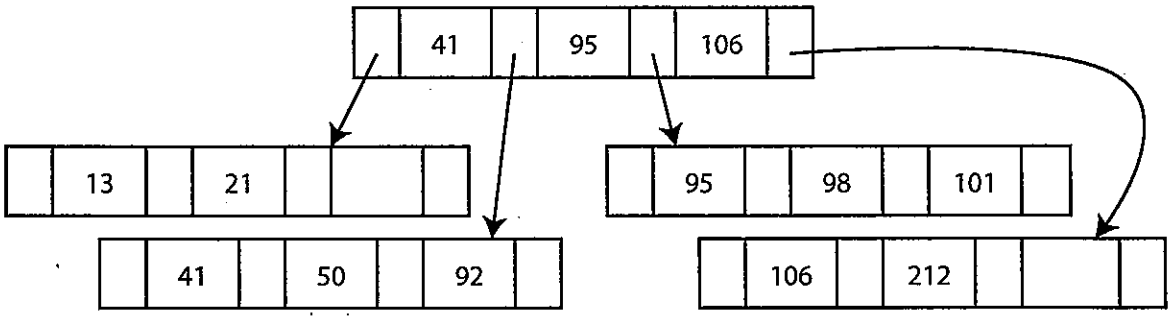
Question 3: Index internals (40 marks)

(a) Consider the B+ tree shown below:



Draw the tree that results from inserting key 65 followed by key 90. (You need draw only the final tree.) Clearly explain in your answer where the values were inserted, what nodes were split (if any), what nodes were created (if any), what pointers re-assigned, etc.

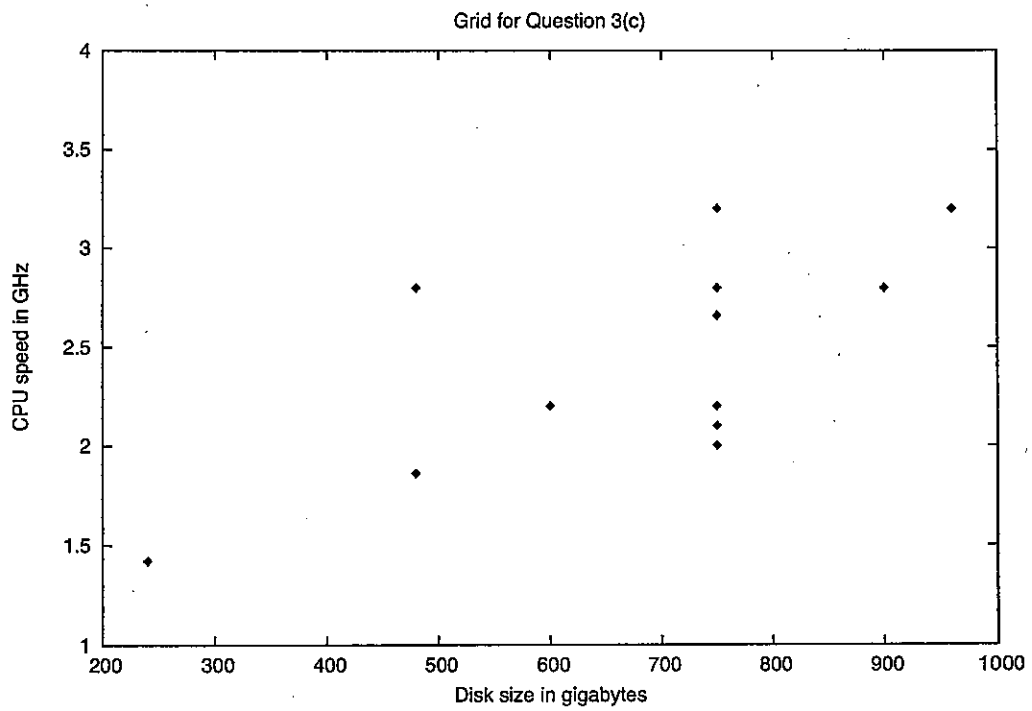
(b) Consider the B+ tree shown below:



Draw the tree that results from inserting key 70. Clearly explain in your answer where the value was inserted, what nodes were split (if any), what nodes were created (if any), what pointers were re-assigned, etc.

(c) Consider the data shown below of some PCs produced by the "Sequential Numbered PCs-are-Us" computer manufacturer. The data is presented in tabular form; the graph presents data from two of the columns.

model	*speed*	*ram*	*hd*
1001	2.66	4096	750
1002	2.10	2048	750
1003	1.42	2048	240
1004	2.80	4096	750
1005	3.20	2048	750
1006	3.20	4096	960
1007	2.20	4096	600
1008	2.20	8192	750
1009	2.00	4096	750
1010	2.80	8192	900
1011	1.86	8192	480
1012	2.80	4096	480



To answer this part of question 3, draw five grid lines (i.e., the total number of grid lines for all dimensions combined) such that there are no more than two points in any bucket.

(d) Draw a diagram showing the grid file with buckets for your answer to (c) above.

(e) Describe *extensible hash tables*. As part of your answer, provide a small example.

Question 4: Data Modelling (30 marks)

(a) Provide an E/R diagram for a database storing information involving *teams*, *players* and their *fans*. Some additional information you must include in your design:

- For each team – its name, its players, its team captain (who is one of its players), and the colours of its uniform.
- For each player – their name.
- For each fan – their name, favorite teams, favorite players, and favorite colour.

Note that a set of colours is not a suitable attribute type for teams. You must find a way around this restriction.

(b) Convert your design for part (a) of this question into the *Object Definition Language*. Make sure you include the keys as you deem them appropriate. The original problem statement introduced a complication involving team colours. Explain why this is not an issue for ODL, and do so by providing a small example.

Question 5: Normal forms (25 marks)

Consider a relation `Equities(B,O,I,S,Q,D)`. The attributes may be thought of as (roughly) a stock broker, the broker's office, a stock investor, the stock itself, number of shares of the stock (quantity), and the yearly dividend produced by the stock. The set of FDs for this relation are $S \rightarrow D$, $I \rightarrow B$, $IS \rightarrow Q$, and $B \rightarrow O$.

(a) What are all the keys for `Equities`?

(b) Verify that the given FDs are their own *minimal basis*.

(c) Use the 3NF synthesis algorithm to find a lossless-join, dependency-preserving decomposition of `Equities` into 3NF relations. Are any of the resulting relations not in BCNF?

Question 6: Relational Algebra (25 marks)

Consider the following schema:

```
Suppliers(sid, sname, address)
Parts(pid, pname, colour)
Catalog(sid, pid, cost)
```

The key for `Suppliers` is `sid`, for `Parts` is `pid`, and for `Catalog` is `sid` and `pid`. The `Catalog` relation associates prices charged for parts by suppliers.

Write the following queries using relational algebra. For items (a) through (e), use the "sequences of assignments" form. For item (f) use the "expression tree" form. List all assumptions. (Some marks will be given for the quality of your answers.)

(a) Find the names of suppliers who supply some blue part.

(b) Find the ids of suppliers who supply some blue or magenta part.

(c) Find the ids of suppliers who supply every magenta part.

(d) Find the ids of suppliers who supply every part.

(e) Find the ids of parts supplied by at least two different suppliers.

(f) Find the id of the most expensive part supplied by company named "Bolts-'R'-Us". (If more than one part shares the same most-expensive cost, then all such parts can be part of the answer set.) Remember to write your answer using the expression-tree form.

END OF EXAM

This page for the exclusive use by markers.

Question 1	
Question 2	
Question 3	
Question 4	
Question 5	
Question 6	
Total	