

UNIVERSITY OF OTAGO EXAMINATIONS 2008

Information Science
Paper INFO 321
DATABASE SYSTEMS

(TIME ALLOWED: 3 HOURS)

This examination paper comprises 6 pages.

Candidates should answer questions as follows:

There are four (4) sections to this examination.
Answer ALL questions in ALL sections.
Questions in this examination total 100 marks.
Suggested time allocations for each question are provided in **[brackets]**
(totalling 180 minutes). Note that this includes a 10 minute reading time
allocation for the scenario in Section A.

The following material is provided:

Use of calculators:

No calculators are permitted.

Candidates are permitted copies of:

Other Instructions:

TURN OVER

Section A

ANSWER ALL QUESTIONS.

Questions in this section (total 29 marks) relate to database administration.

**The following scenario is used in questions 1–3.
Please read this material before answering the questions.
[Allow 10 minutes reading time.]**

Chirp is a small social networking startup company. Registered users build lists of friends and send instant messages to each other, from which the back end automatically builds conversation threads. Threads are mined for key words and phrases that are used to link to other, possibly related conversations. Users rate the relevance of these links, thus improving the accuracy of the mining algorithms.

The entire operation is housed upstairs in a slightly run down 19th century building in Dunedin's wharf district. They chose this location because it was out of the way, and because of the large triple-brick safe in the back of the office, which they use as the server room. The ground floor is occupied by a warehouse full of ex-demolition materials (timber, windows, appliances, etc.), which is frequently visited by heavy trucks. The power lines are rather old and there have been several power cuts recently.

Chirp has four employees, who have complete access to all systems, but only one particular employee works with the database on a regular basis. The database is implemented in Oracle10g running under Linux, and the web front end is built using the Apache web server and PHP (running on the same server as the database). Demand is rapidly increasing and has resulted in some service outages due to overloading.

User data currently comprise various personal information and profile images. Messages and conversations are also stored in the database. Chirp plan to add rich text formatting in multiple languages and multimedia messaging (i.e., pictures, sound and video). The current database is already quite large, with over 500 000 messages and over 200 000 links between conversations. There are currently about 25 000 users, increasing by about 500 per week. Figure 1 on page 3 shows a Beynon-Davies transaction usage map for the transaction "list messages from friends since last login".

Unique B-tree indexes have been defined on all primary keys in the database.

1. [The suggested time allocation for answering this question is 17 minutes.]

Identify *at least two* (2) specific hazards in this scenario that Chirp should plan for. For each of these hazards:

- briefly describe the hazard and explain *why* it is a hazard;
- briefly describe the possible consequences of the hazard; and
- briefly discuss options for how Chirp could avoid and/or recover from the effect(s) of the hazard.

(12 marks)

TURN OVER

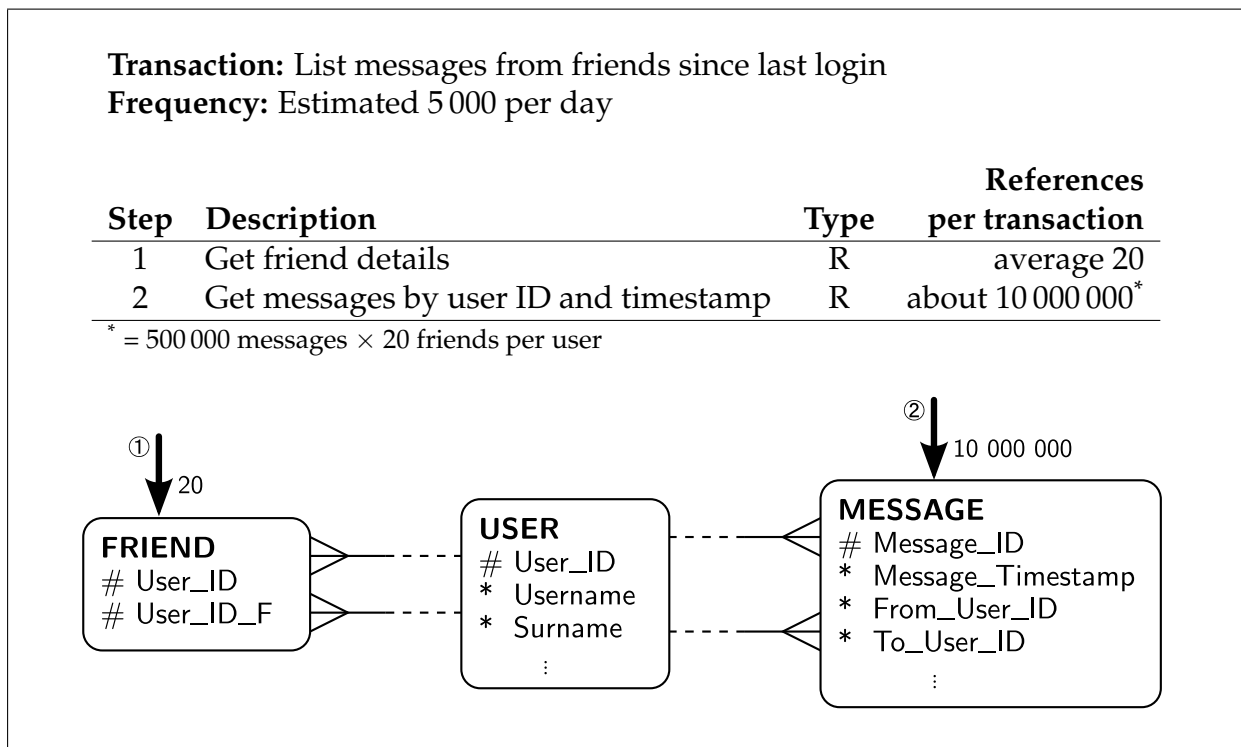


Figure 1: Beynon-Davies transaction usage map for use in questions 2 and 3.

2. [The suggested time allocation for answering this question is 17 minutes.]

Every time a user logs in, Chirp displays messages sent from friends since the user last logged in (a typical user has about twenty friends). Users are complaining that this is very slow.

- (a) Explain why this transaction is performing so poorly. (4 marks)
- (b) Discuss *two* (2) possible solutions to this problem. Consider how each of the solutions will help solve the problem and evaluate potential advantages and disadvantages of both solutions in the context of the scenario. Clearly state any assumptions that you make. (You do not need to write any code.) (8 marks)

3. [The suggested time allocation for answering this question is 9 minutes.]

When adding a physical access method, such as an index, to the Username column in the User table (see Figure 1 above), the choice of access method depends on the column's *selectivity*. Identify Username's level of selectivity, explain what this means and the practical implications for queries, and discuss which access method(s) would be most appropriate for this column. (5 marks)

[SECTION A TOTAL 29 MARKS]

TURN OVER

Section B

ANSWER ALL QUESTIONS.

Questions in this section (total 13 marks) relate to object database management systems (ODBMS).

4. [The suggested time allocation for answering this question is 14 minutes.]

- (a) Explain the connection between *relationships* and *traversal paths* in the Object Data Management Group (ODMG) object model. (4 marks)
- (b) Identify the relational model concept that is the nearest equivalent to an ODMG relationship, and explain how the two concepts differ. (4 marks)

5. [The suggested time allocation for answering this question is 9 minutes.]

“ODBMSs allow a wide range of data types in the same database.” The usual implication of this statement is that relational DBMSs do not do this. Discuss the main misconception that underlies this statement and why it might arise, then explain why this is not a major issue with modern relational database management systems. (5 marks)

[SECTION B TOTAL 13 MARKS]

TURN OVER

Section C

ANSWER ALL QUESTIONS.

Questions in this section (total 40 marks) relate to management of distributed data.

6. **[The suggested time allocation for answering this question is 7 minutes.]**
Ajax is a collection of technologies for developing Web applications. Explain how Ajax works and briefly discuss the advantage(s) of Ajax for database-driven dynamic Web applications. (4 marks)
7. **[The suggested time allocation for answering this question is 18 minutes.]**
Compare in detail the architecture of a networked database system with that of a federated database system, with respect to how the database schema is defined and how applications interact with the system. Use diagrams or examples to illustrate your answer. (10 marks)
8. **[The suggested time allocation for answering this question is 14 minutes.]**
Suppose a customer browses an online retailer's web site for the first time (so they do not have an account yet). They add three items to their shopping cart, then open a second web browser window to the same site to compare two similar products side by side. The first browser window lists three items in their shopping cart, while the second browser window lists no items in their shopping cart.
- (a) What has happened, and why? (4 marks)
- (b) Describe a common technique that can be used to resolve this problem, and briefly discuss any advantages and/or disadvantages of this technique. (4 marks)
9. **[The suggested time allocation for answering this question is 14 minutes.]**
Two techniques for storing XML (extensible markup language) data in a relational database are "store as native data type" (e.g., XMLType in Oracle10g) and "shred to generic schema". Briefly describe both techniques, then compare the relative advantages and disadvantages of each technique. (8 marks)
10. **[The suggested time allocation for answering this question is 18 minutes.]**
Two of Date's objectives for a "true" distributed database system are *location independence* and *distributed query processing*. For each of these objectives:
- (a) describe the goal of the objective; and (4 marks)
- (b) discuss *two* (2) key technical features that are required in order to achieve this objective. Your answer should include examples of how these technical features are implemented in Oracle10g. (6 marks)

[SECTION C TOTAL 40 MARKS]

TURN OVER

Section D

ANSWER ALL QUESTIONS.

Questions in this section (total 18 marks) relate to decision support systems (DSS).

11. **[The suggested time allocation for answering this question is 14 minutes.]**
Decision-making in a business typically occurs at one of three different levels: operational, tactical or strategic. For each of the following four queries, classify it as either an operational, tactical or strategic style of query, and explain why.
- (a) A fire has shut down the Christchurch factory for several weeks. Which other factories have sufficient spare capacity to take over for the missing factory? (2 marks)
 - (b) Has the Wellington factory met its production quota for the week? (2 marks)
 - (c) Rising oil prices have significantly increased our transport costs. Should we move our factories closer to our primary markets? (2 marks)
 - (d) Do we need to order more red LEDs? (2 marks)
12. **[The suggested time allocation for answering this question is 12 minutes.]**
Compare the characteristics of operational data and decision support data with respect to the following three areas:
- the types of decisions they are used to support;
 - their timespan; and
 - their level of detail.
- (6 marks)
13. **[The suggested time allocation for answering this question is 7 minutes.]**
Briefly describe the structure and components of a *star schema*. Use a diagram to illustrate your answer. (4 marks)

[SECTION D TOTAL 18 MARKS]

END