

ASSIGNMENT #3 – Storage Structures

ISQA 4300/8306

Due date: October 18

Overview

Recently we have been discussing storage structures: database blocks, extents, segments, tablespaces. We have also discussed database objects that make use of that space, tables and indexes in particular. This assignment is designed to give you some hands-on practice with these concepts and to explore their inter-relationships.

Task 1: Choosing tablespaces and sizing data files

There exists in the DIVESHOP schema a collection of tables that constitutes the existing application tables that you have been called upon to administer. The purpose of this is to simulate a database that you've *inherited* as a new database administrator. You'll continue to apply new skills and knowledge to this database.

Figure 1 shows the schema for the database. Oracle datatypes are shown for each column.

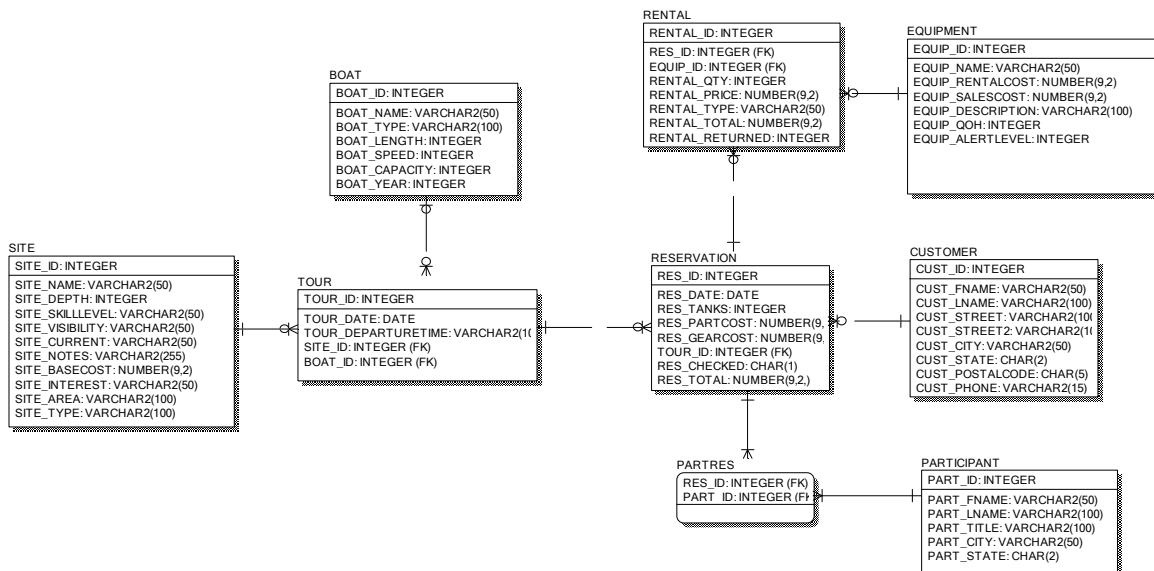


Figure 1 Schema for on-line scuba-diving tour reservation system

Thanks to a properly maintained data model, you have also been given a script that can create the tables found in the DIVESHOP schema. This script is called *diveshop_ddl.sql*. You will want to re-create these tables in your own user schema, but rather than rely on the defaults for the tablespace, the PCTFREE values, and the extent sizes, you will want to specify these explicitly, based on your analysis of the current and projected data needs.

The following statements reflect the amount of data that the company anticipates if the business evolves as planned.

- There will be 100 sites to which tours will be conducted.
- On average, there will be 100 tours to each site per year.
- Fifty boats will be used for the tours, with one boat being used on a given tour.
- On average, there will be 10 reservations associated with a single tour.
- Each reservation is booked by one customer.
- The average number of participants per reservation is 3.
- On average, each participant is listed on 2 reservations.
- The list of past customers is being imported from a legacy system, containing approximately 14,000 customers.
- For the customers, only the street address and postal code is recorded for the address in the legacy system (ok, this is a bit contrived, but assume it's true.) The customer city will be updated by looking up the city in a postal service master address list. The average size (not maximum size) of a row will expand from 80 to 93 bytes.
- Each year, 70% of those placing reservation are new customers, and 30% are customers who have gone on tours in previous years.
- As part of a reservation, a customer can rent diving equipment. There are 25 types of equipment rented, including goggles, airtanks, fins, underwater camera, spotlight, etc. Rental_Qty indicates the number of units of each type of equipment rented on a given reservation.
- On average, 5 different types of equipment are rented per reservation. The rental record currently has null values for the RENTAL_TYPE column. The company anticipates that this column will be updated in the future to store a value among the following: {Rental; Sale; Replacement}. The average size of the value stored in this column will be 5 bytes.

Subtask 1:

Using information provided above, as well as any information you can glean from the existing tables in the DIVESHOP schema, determine for *each* table :

1. the appropriate value of PCTFREE. You don't need to justify a particular value (e.g. why 25 rather than 30), but you do need to justify why you chose a large (>20), medium (10-20), or small (0-9) value of PCTFREE for each table (i.e. what you think gets updated in each table). Assume that the initial row size is the average length of the rows in the current tables in the DIVESHOP schema, since these will be initially imported into your tables. You may assume that the tables in DIVESHOP reflect the characteristics of data that will be entered into the database in the future, too. The exception are the CUSTOMER and RENTAL tables, as described above. Otherwise, you may assume that variable character strings may fluctuate some in size, but the long-term average length will be approximately what it is in the current DIVESHOP tables.
2. the approximate amount of storage needed to store a year's worth of activity. Use the guidelines found on slide #10 of the lecture "Tables and Indexes". You may assume that all numbers (number(), integer, decimal) occupy 4 bytes and all date or time values occupy 7 bytes each. A 10% fudge factor may be used in all cases.
3. Reasonable values for the INITIAL and NEXT extent sizes.

Subtask 2:

Using the guidelines discussed in the lectures, decide which tablespaces should be created for this system and which tables will go in each. Name each tablespace using the convention `xx_tablespaceName`, where `xx` = your initials. For example, if I were to create a tablespace called 'theater' I would name it 'pw_theater'.

Subtask 3:

For each of the tablespaces from Subtask 2, indicate what is the smallest size file that would hold one year's worth of transactions? Remember, the size required for a tablespace would be the total of the size of all the tables that will be placed in that tablespace.

Task 2: Table implementation

Subtask 1:

Create each of the tablespaces that you identified above in Task 1 subtask 3, placing each data file in `c:\app\oracle\oradata\dbadb\assignment3`. **USE A FILESIZE OF 2MB FOR EACH FILE.** DO NOT USE THE SIZE YOU CALCULATED IN SUBTASK 3 of Task 1. The reason for this is that I need to have some control over the storage space used on our Oracle server. Be explicit about your choices for EXTENT MANAGEMENT and SEGMENT SPACE MANAGEMENT.

Subtask 2:

Modify the script `diveshop_ddl.sql` to make the CREATE TABLE statements more precise. That is, modify them to reflect specific tablespace, PCTFREE, INITIAL and NEXT values, according to your analysis in Task 1.

Subtask 3:

Run the modified script `diveshop_ddl.sql` to create your tables.

Subtask 4:

Populate your tables, using either the data that currently exists in the DIVESHOP tables, or with the script `diveshop_data.sql`.

Subtask 5:

It seems that there is additional legacy customer data available. Insert into the CUSTOMER table data from WOLCOTT4300.CUSTOMER_LEGACY. Note: the columns of the two tables do not match exactly.

(Hint: There is a variant of the INSERT command that looks like this:

```
INSERT INTO table1 (column1, column2, ..., columnx)
SELECT columna, columnb, ..., columnn
FROM schema.table2;
```

Q1: Did you encounter any problems in any of the subtasks? Indicate what they were and how you fixed them.

Q2: Examine the extents being used for the CUSTOMER table's segment (find the answer in DBA_EXTENTS or USER_EXTENTS). How many extents are used? Does this answer surprise you? Explain the logic behind Oracle's allocation of extents for this table.

Task 3: (ISQA 8306 students only)

Check to see whether CUSTOMER has any chained or migrated rows.

Q3: How many does it have (should be 0), and how do you know?

The CUST_CITY column currently has NULL values. Experiment to determine how large the updates to the cust_city need to be to produce row migration or chaining. (There is no requirement here that the cust_city values be realistic or at all related to other data in the table. We are trying to force row migration.)

Q4: Describe your approach to doing this, and what did you discover? How large did the update have to be to force row migration?

Q5: In Task 1, we learned that the updates to the CUST_CITY column would increase the average row length from 80 to 93 bytes. In light of your answer to Q4, how well did your choice of PCTFREE for this table satisfy the requirement stated in Task 1?

Task 4: Reflection

Pick **two** (2) of the bullet points from this list and answer the questions¹. Please indicate which questions you are answering.

- What did I actually achieve with this piece of work? Which were the most difficult parts, and why were they difficult for me? Which were the most straightforward parts, and why did I find these easy?
- How well do I think I achieved the intended learning outcomes for this task? Where could I have improved my achievement? Why didn't I improve it at the time?
- What have I got out of doing this assignment? How have I developed my knowledge and skills? How do I see the payoff from doing this assignment helping me in the longer term?
- What else have I got out of doing this assignment? Have I developed other skills and knowledge, which may be useful elsewhere at another time? If so, what are my own emergent learning outcomes from doing this assignment?
- What was the best thing I did? Why was this the best thing I did? How do I know that this was the best thing I did?
- What worked least well for me? Why did this not work well for me? What have I learned about the topic concerned from this not having worked well for me? What have I learned about myself from this not having worked well for me? What do I plan to do differently in future as a result of my answers to the above questions?
- With hindsight, how would I go about this assignment differently if doing it again from scratch? To what extent will this assignment influence the way I tackle anything similar in future?
- What did I find the greatest challenge in doing this work? Why was this a challenge to me? To what extent do I feel I have met this challenge? What can I do to improve my performance when next meeting this particular sort of challenge?

¹ Questions taken from <http://www.escalate.ac.uk/resources/reflection/09.html>

- What was the most boring or tedious part of doing this assignment for me? Can I see the point of doing these things? If not, how could the assignment have been re-designed to be more stimulating and interesting for me?
- Has it been worth the effort I put in? Do the marks represent a just reward? Should this assignment be worth more or less marks in the overall scheme of things?
- Do I feel that my time on this assignment has been well spent? If not, how could I have used my time more sensibly? Or should the assignment have been designed differently? Which parts of the assignment represent the time best spent? Which parts could be thought of as time wasted?
- How useful do I expect the feedback to be, that I receive on my efforts for this assignment? What sorts of feedback do I really want at this point in time? What sorts of feedback do I really need at this point in time? What are my expectations of getting useful feedback now, based on the feedback (or lack of it) that I've already received on past work?
- Overall, how has this assignment helped (or hindered) my motivation to learn more about this part of my syllabus? Has it encouraged me, or disillusioned me?
- To what extent has this assignment helped me to clarify what I need to learn about this topic? Have I a clearer picture after doing the assignment, or a foggier one? Who can help me gain a clearer picture, if the latter?
- To what extent has this assignment helped me to see where the goalposts stand for future assessments such as exams? Has it given me useful insights into what will be expected of me in future?
- What advice would I give go a friend about to start on the same assignment? How much time would I suggest that it would be worth putting into it? What pitfalls would I advise to be well worth not falling into?
- What are the three most important things that I think I need to do with this topic at this moment in time? Which of these do I think is the most urgent for me to do? When will I aim to start doing this, and what is a sensible deadline for me to have completed it by?

Deliverables

Task 1:

- Answers for subtasks 1-3

Task 2:

- SQL statements used to create tablespaces
- Modified diveshop_ddl.sql script.
- Answers to Q1, Q2
- SQL used to move each index

Task 3 (ISQA 8306 only)

- Answers to Q3-5

Task 4:

- Answers to questions in two bullet points (your pick).