

Assignment 5 Suggested Solutions

Task 1

1. Assume that the record header overhead is 28B, that 63% of the cars have no comments (i.e., comment IS NULL), and that field pointers are 1B except for the non-null comments where the field pointer is 2B. What would the average record size be? Explain your answer.

$28 \text{ (header)} + 10.27 \text{ (field pointers)} + 4 \text{ (id)} + 7.7 \text{ (make)} + 4.9 \text{ (model)} + 1 \text{ (model_year)} + 4 \text{ (mileage)} + 6.2 \text{ (fuel)} + 7.8 \text{ (type)} + 4 \text{ (price)} + 4 \text{ (dealer_id)} + 0.27 * 151.7 \text{ (comment)} = 122.8 \text{ B}$

2. Assume that the database index page size is 16,384B and that the page overhead data is 420B. On average, how many records would fit on one page assuming that the primary index leaf nodes are 75% full?

Record data size on a 75% full node: $(16,384 - 420) * 0.75 = 11,973$

Average number of records per page: $11,973 / 122.8 = 97.5$

3. Assume that the database consists of 25,000 used cars for sale. How many leaf nodes would we need for the database if we assume that leaf nodes are 75% full?

$25,000 / 97.5 = 257$

4. Assume that each index node can hold 389 primary key values when full. How many levels would be in the primary index if we assume that index nodes (except the root node) are also 75% full?

We need only two levels - the leaf nodes and the root node since the root node can hold 389 values but there are only 257 leaf nodes.

5. Assuming that the root node is always buffered in the DBMS, how many disk accesses are needed to retrieve any car record based on the primary key when using the primary index if the record is not in the database buffer already?

1 - the leaf node only.

6. Assume that we want to create a secondary, non-unique index on car make and model. Would you suggest making the index on (make, model) or (model, make)? Explain your answer

(make, model) because this can be used also for queries on make only (WHERE make = ..., ORDER BY make, GROUP BY make, COUNT(DISTINCT make), ...). It is unlikely that similar queries will be requested on model only.

7. What would the SQL statement look like for creating the index, if we want to name the index model_idx?

CREATE INDEX model_idx ON car (make, model)

Task 2

- #: The record number – is referred to by the pPtr and nPtr*
Tid: The transaction Id of the transaction, if any, causing the event
Time: Timestamp of the event
Operation: The database operation invoked
Object: The database object (row) being operated on, if any
Before image: The value of the object before being operated on, if any
After image: The value of the object after being operated on, if any
pPtr: Reference to the previous operation invoked by the same transaction to create a linked list for backwards chaining of operations done by one transaction
nPtr: Reference to the next operation invoked by the same transaction to create a linked list for backwards chaining of operations done by one transaction
- Types of events being logged:*
Transactional events: Start, commit, rollback. These are included because the recovery manager needs to decide whether to UNDO or REDO a transaction depending on the whether and how the transaction completed.
Data operation events: Insert, update, delete. These are included to enable the recovery manager to UNDO or REDO each of the operations.
Checkpoint events These are included so that the recovery manager may decide how far back into the log file it needs to seek to find transactions to recover.