Chapter 7

Introduction to Structured Query Language (SQL)

Answers to Review Questions

1. **In a SELECT query, what is the difference between a WHERE clause and a HAVING clause?**

   Both a WHERE clause and a HAVING clause can be used to eliminate rows from the results of a query. The differences are 1) the WHERE clause eliminates rows before any grouping for aggregate functions occurs while the HAVING clause eliminates groups after the grouping has been done, and 2) the WHERE clause cannot contain an aggregate function but the HAVING clause can.

2. **Explain why the following command would create an error, and what changes could be made to fix the error.**
   SELECT V_CODE, SUM(P_QOH) FROM PRODUCT;

   The command would generate an error because an aggregate function is applied to the P_QOH attribute but V_CODE is neither in an aggregate function nor in a GROUP BY clause. This can be fixed by either 1) placing V_CODE in an appropriate aggregate function based on the data that is being requested by the user, 2) adding a GROUP BY clause to group by values of V_CODE (i.e. GROUP BY V_CODE), 3) removing the V_CODE attribute from the SELECT clause, or 4) removing the Sum aggregate function from P_QOH. Which of these solutions is most appropriate depends on the question that the query was intended to answer.

3. **What type of integrity is enforced when a primary key is declared?**

   Creating a primary key constraint enforces **entity integrity** (i.e. no part of the primary key can contain a null and the primary key values must be unique).

4. **Explain why it might be more appropriate to declare an attribute that contains only digits as a character data type instead of a numeric data type.**

   An attribute that contains only digits may be properly defined as character data when the values are nominal; that is, the values do not have numerical significance but serve only as labels such as ZIP codes and telephone numbers. One easy test is to consider whether or not a leading zero should be retained. For the ZIP code 03133, the leading zero should be retained; therefore, it is appropriate to define it as character data. For the quantity on hand of 120, we would not expect to retain a leading zero

such as 0120; therefore, it is appropriate to define the quantity on hand as a numeric data type.

5. **What is the difference between a column constraint and a table constraint?**

   A column constraint can refer to only the attribute with which it is specified. A table constraint can refer to any attributes in the table.

6. **What are "referential constraint actions"?**

   Referential constraint actions, such as ON DELETE CASCADE, are default actions that the DBMS should take when a DML command would result in a referential integrity constraint violation. Without referential constraint actions, DML commands that would result in a violation of referential integrity will fail with an error indicating that the referential integrity constrain cannot be violated. Referential constraint actions can allow the DML command to successfully complete while making the designated changes to the related records to maintain referential integrity.

7. **Rewrite the following WHERE clause without the use of the IN special operator.**
   WHERE V_STATE IN ('TN', 'FL', 'GA')

   WHERE V_STATE = 'TN' OR V_STATE = 'FL' OR V_STATE = 'GA'

   Notice that each criteria must be complete (i.e. attribute-operator-value).

8. **Explain the difference between an ORDER BY clause and a GROUP BY clause.**

   An ORDER BY clause has no impact on which rows are returned by the query, it simply sorts those rows into the specified order. A GROUP BY clause does impact the rows that are returned by the query. A GROUP BY clause gathers rows into collections that can be acted on by aggregate functions.

9. **Explain why the two following commands produce different results.**
   SELECT DISTINCT COUNT (V_CODE) FROM PRODUCT;
   SELECT COUNT (DISTINCT V_CODE) FROM PRODUCT;

   The difference is in the order of operations. The first command executes the Count function to count the number of values in V_CODE (say the count returns "14" for example) including duplicate values, and then the Distinct keyword only allows one count of that value to be displayed (only one row with the value "14" appears as the result). The second command applies the Distinct keyword to the V_CODEs before the count is taken so only unique values are counted.

10. **What is the difference between the COUNT aggregate function and the SUM aggregate function?**

    COUNT returns the number of values without regard to what the values are. SUM adds the values together and can only be applied to numeric values.

11. **Explain why it would be preferable to use a DATE data type to store date data instead of a character data type.**

    The DATE data type uses numeric values based on the Julian calendar to store dates. This makes date arithmetic such as adding and subtracting days or fractions of days possible (as well as numerous special date-oriented functions discussed in the next chapter!).

12. **What is a recursive join?**

    A recursive join is a join in which a table is joined to itself.