ASSIGNMENT 5

1. a) StudentID in the entity STUDENT should be used to speed up the query as StudentID is the primary key/ identifier. Also, StudentID is used in both entities STUDENT and REGRISTRATION.
   GPA is a non key attribute that is used to select the required records.
   StudentName is a non key attribute that is used to sort the records.

   b) Create unique INDEX STD on STUDENT (StudentID);
   Create INDEX StdReg ON REGISTRATION (StudentID);

   Create INDEX clust_index

   ON STUDENT (GPA)

   CLUSTER;

   Create INDEX Name ON STUDENT (StudentName);

   Create UNIQUE INDEX StdReg

   ON REGISTRATION (StudentID, CourseID);

8. Recommendations for denormalization:

DEPARTMENT

(Dept ID, Manager ID, SalesGoal, Store ID, Region, Manager ID, SquareFeet)

The reference data denormalization isn't recommended since the table STORE is related to table MANAGER and they are probably more than few departments in each STORE.

EMPLOYEE SCHEDULE

(DepartmentID, EmployeeID, WhereWork, EmployeeName, EmployeeAddress, Date)

A many-to-many relationship (associative entity) with nonkey attributes: Rather than joining three files to extract data from the two basic entities in the relationship, it is better to combine attributes from one of the entities into the record representing the relation in the many-to-many relationship, thus avoiding one join in many data access modules.

9. Recommendations for denormalization: Including SpecialtyDescription attribute would be a possible way for denormalization in the PLAYER relation. PlayerSpecialtyCode refers to SpecialtyCode in SPECIALTY relation. Addition of SpecialtyDescription data to PLAYER relation will lead to redundancy of data which would take up the storage space.

There might be opportunities for further denormalization in these relations, but the following questions need to be asked before proceeding with any further denormalization:

- What is the meaning and the contents of the TeamLocation attribute in the TEAM relation? Is this a code used to reference the LocationID in the LOCATION relation? Or is this a character field storing different information about the Team's location? Can a TEAM instance have more than one location?
- What is the meaning and the contents of the ManagerTeam attribute in the MANAGER relation? Does this attribute refer to the TeamID attribute in the TEAM relation? Is a Manager instance also a Player instance?

13. A row selection qualification clause will be used:

WHERE (Major = 'MIS' or Major = "Computer Science") And Age > 25 And MaritalStatus = "single") Or (Major = "Computer Engineering" And MaritalStatus = "single" And HomeZipcode = 45462).

Indexes on these fields can be used to considerable advantage in this situation. Assume that each index qualification (e.g., Major = "MIS") produces a list of record numbers for the records satisfying that qualification. Lists can be merged to process OR operators, and lists can be intersected to process AND operators. Indexes may be scanned in main memory, and the list operations also done without accessing secondary memory, thus composing the list of qualified records very quickly. Only then does secondary memory need to be accessed for only those records that satisfy the whole query.

16. A cluster is defined by the tables and the column or columns by which the tables are usually joined. The tables are joined by the foreign key and they should have same value in the two tables for the adjacent records. If the tables are populated with data before clustering occurs, this is much harder to achieve. Hence, tables are assigned to a cluster at the time of their creation.

17. Parallel query processing means same query runs on multiple processors but each processor can access in parallel a different subset of the database. In the clause the response of the parallel query processing is based upon the number of processors available. In the general structure, each one set of conditions inside brackets is known as conjunction and it could be assigned to separate processor.

There are two approaches to perform parallel processing:

- During scanning of the table ensure that the subsequent scans of the table are performed parallel with not less than 3 processors
- Another choice might be to give the database management system a clue inside the query.