8Cairo University
Faculty of Engineering
4th year Computer Engineering Department

Fall 2018

Advanced Database Systems
Project Requirements

Objectives

After this project, the student should be able to
- Understand the concepts of database tuning.
- Get through the phases of database tuning.
- Use different database tuning techniques
    - o Optimizing the schema
    - o Optimizing the memory requirements
    - o Optimizing queries.
- Use different database optimization tools.

Note:
- Security and Access Control, Recovery, transaction and concurrency control are BONUS.

Note:
- One day late makes you lose 1/4 of the grade.
- Two days late makes you lose 1/2 of the grade.
- Three days late makes you lose 3/4 of the grade.

Requirements

It is required to apply different database tuning methods on an existing database system. Students working in (2-4 students) team should pass the following project phases:

Phase 1: Get a working database project & fill it

You should get a working database project and understand it well, since you are going to modify it to enhance its performance, you can take your 2nd year database project.

You should fill your database with large volume of data (in the order of 1,000,000 to 10,000,000), you can use a database filling program which fills the database with random data.

> Example:

Table Name	Row Count	Main Key	Indexes	FK	Identity Column	Max Row Size (Bytes)
Category	16	Yes	3		Yes	59
Category_Keyword	409	Yes	2	2		8

> Note: This report can be generated by the DBMS you use.

Phase 2: Query Processing

Based on your system, choose **the most critical queries** (min. 4 queries). This group of queries should contain data insertion and retrieval, and then you should use a query analyzer tool to estimate the cost of the query, the execution plan, memory and cache usage, degree of parallelism, actual running time and result analysis.

Deliverables from this phase

1. A report containing:
    a. Selected queries
    b. Execution plan for each query (Query tree).
    c. Server traces information.
    d. Memory and cache usage report.
    e. Parallel query processing report in case of multiprocessor devices.
    f. Comments on the performance.

Phase 3: Optimization and Validation

Optimization

In this phase you should use different optimization techniques to enhance your database performance, which reflect on the total system performance, in this phase you will pass through the following steps:

1. **Schema optimization**: your schema may need to be modified.
2. **Memory and cache optimization**: you can use stored procedures or any other techniques like changing the block size or any other parameter.
3. **Index tuning**: you can add or remove indexes as needed.
4. **Query optimization**: rewrite the query to enhance the performance.

   Note: any modifications need to be justified.

Validation

The final step is to validate your optimization. This will be done by running the queries on a large volume of data (in order of 100,000). Validate your work by running the system on your database **before** and **after** optimization and compare the results.

Finally report the effect of the **database size** on the system performance. This will be done by running the system on different database sizes (in the order of: 10,000 – 100,000 – 1,000,000 – 10,000,000), you can also try to change the hardware (run on another hardware device) and observe the change in the performance.

Deliverables of this phase

A report describing the following:

1. After Optimization Query Statistics
   a. Execution plan for each query (Query tree)
   b. Parallel query processing report in case of multiprocessor devices "Theoretically from the execution plan even it's not implemented by the DBMs".

2. Optimization Details
   a. The new database statistics after modification "As in phase 1".
   b. The enhancement in the schema.
   c. The enhancement in the memory management.
   d. The modification in the indexes.
   e. For each query describe your modifications on the query statement.

3. Validation Details *(These are Graphs and/or Tabularized Comparisons ONLY)*
   a. For each query describe: "You will make this once on all the data you have in DB .. compare the memory and time used before and after optimization"
      i. The effect of all the previous modifications on the query performance (present this as a comparison).
      ii. The percentage of enhancement.
   b. A graph explains the effect of the database size on performance (use the same mix above but with different database sizes).
   c. The effect of changing the hardware specification. (if applicable)

      d. Conclusion on the pervious phases "Put your comments and observations on the performance before and after the optimization and how it changed".

Useful Links:

1. Query Optimizer, Wikipedia,
   http://en.wikipedia.org/wiki/Query_optimizer

2. Tips, Tricks, and Advice from the SQL Server Query Optimization Team,
   http://blogs.msdn.com/queryoptteam/

3. Analyzing queries, msdn SQL Server Developer Center,
   http://msdn.microsoft.com/en-us/library/aa217001(SQL.80).aspx

4. SQL Server Optimization, msdn SQL Server Developer Center,
   http://msdn.microsoft.com/en-us/library/aa964133(SQL.90).aspx

5. Optimizing SQL Server Query Performance, TachNet magazine,
   http://207.46.16.252/en-us/magazine/2007.11.sqlquery.aspx

6. MSSQL Tip, query optimization,
   http://www.mssqltips.com/category.asp?catid=37

7. Are Your SQL Server Application Queries Wasting Memory, MSSQL Tips,
   http://www.mssqltips.com/tip.asp?tip=1632

8. Lengthy SQL Server Queries Will Consume Your CPU, MSSQL Tips,
   http://www.mssqltips.com/tip.asp?tip=1500

9. Maximum Capacity Specifications for SQL Server, msdn SQL Server Developer Center,
   http://technet.microsoft.com/en-us/library/ms143432.aspx

10. Parallel Query Processing, msdn SQL Server Developer Center,
    http://technet.microsoft.com/en-us/library/ms178065.aspx

11. Execution Plan Caching and Reuse, msdn SQL Server Developer Center,
    http://technet.microsoft.com/en-us/library/ms181055.aspx

12. Specifying Max Degree of Parallelism in SQL Server for a Query, MSSQL Tips,
    http://www.mssqltips.com/tip.asp?tip=1047