Database Management System – cs422 DE

Assignment 10 – Week 14

This assignment is based on lecture 12 (chapter 24 – Distributed Databases)

- Submit your *own work* on time. No credit will be given if the assignment is submitted after the due date.
- Note that the completed assignment should be submitted in .doc, .docx, .rtf or .pdf format only.
- In MCQs, if you think that your answer needs more explanation to get credit then please write it down.
- You are encouraged to discuss these questions in the Sakai forum.

---

(1) Compare and contrast a DDBMS with distributed processing. Under what circumstances would you choose a DDBMS over distributed processing?

ANS: Distributed Database management system is a centralized database that can be accessed over a computer network. This can be used by other users that are connected to network. Data can be transport transparently to users.

We don't consider DDBMS as a simple distributed processing because in DDBMS consist of data that is physically distributed across several sites in the network and user need to access it via network.

(2) Compare and contrast a DDBMS with a parallel DBMS. Under what circumstances would you choose a DDBMS over a parallel DBMS?

ANS: As its name Parallel DBMS is running across multiple processors and disk that is designed to do parallel processing to improve performance.

We only consider DDBMS over Parallel only if we don't need high performance with minimum resource.

(3) Discuss the advantages and disadvantages of a DDBMS.

ANS: Advantages

- Improve Reliability
- Improved performance
- Improved Reliability
- Integration
- Economics
- Remaining Competitive

Dis-Advantages

- Complexity
- High Cost
- High chance of Security Break

- Lack of standards
- Lack of experience
- DB design more complex

(4) What is the difference between a homogeneous and heterogeneous DDBMS? Under what circumstances would such systems generally arise?

ANS: In Homogeneous system use the same DBMS but in heterogeneous system, sites may run different DBMS product, which need not be based on the same underlying data model. Homogeneous systems are much easier to design and manage. This approach provides incremental growth and allows increased performance by exploiting the parallel processing. Heterogeneous system usually results when individual sites have implemented their own databases, as well translation are required to allow communication between different DBMS.

(5) What functionality do you expect in a distributed DBMS?

ANS:

- Extended communication services to provide access to remove sites and allows the transfer of queries and amoung the sites using a network.
- Extended system catalog to store data distribution details.
- Distributed query processing, including query optimization and remove data access.
- Extended security control to maintain appropriate authorization/access privileges to the distributed data.
- Extended concurrency control to maintain consistency of distributed and possibly replicated data.

(6) One problem area with DDBMSs is that of distributed database design. Discuss the issues that have to be addressed with distributed database design. Discuss how these issues apply to the global system catalog.

ANS: The issues that have to addressed with distribution database design.

1. Data Fragmentation: To Increase performance data need to be distributed across multipe sites.
2. Data Replication: By distrubting data to multiple sites, data will highly be available for fault tolerant.
3. Concurrency: Distributed database design can handle concurrent multiple users.

The global system design catalog has the same functionality as the system catalog of a centralized system. This holds information specific to the distributed nature of the system. A fully replicated global system catalog compromises site autonomy as every modification to it which will be communicated with other sites.

(7) What are the strategic objectives for the definition and allocation of fragments?

ANS: Below are the main objectives of fragments.

1. Efficiency: Data is stored close to where it is most frequently used. In addition, data that is not needed by local applications is not stored.
2. Parallelism: A transaction can be divided into several subqueries that operates on fragments
3. Security: Data not required by local applications is not stored and consequently not available to authorized users.

(8) Describe alternative schemes for fragmenting a global relation. State how you would check for correctness to ensure that the database does not undergo semantic change during fragmentation.

ANS: There are 2 types of fragmentations horizontal and vertical.

The three rules that must be followed during fragmentations:

1. Completeness
2. Reconstructions
3. Disjointness

(9) What layers of transparency should be provided with a DDBMS? Give examples to illustrate your answer. Justify your answer.

ANS: Transparency hides implementation details from the user. DDBMS provides various levels of transparency. Given below,

1. Distribution Transparency: This allows user to perceive the database as a single, logical entity. The user does not need to know the data is fragmented or the location of data items.
2. Transaction Transparency: This maintains the distributed database's integration and consistency. Distributed transaction can access data stored at more than one location. Each transaction divided into a number on sub transaction.
3. Performance Transparency: This requires a DDBMS to perform as if it were a centralized DBMS. The system should not suffer any performance degrading due to the distributed architecture.
4. DBMS Transparency: This hides the knowledge that the local DBMSs may be different and is therefore applicable only to heterogeneous DDBMSs. It is one of the most difficult transparencies to provide as a generalization.

(10) A DDBMS must ensure that no two sites create a database object with the same name. One solution to this problem is to create a central name server. What are the disadvantages with this approach? Propose an alternative approach that overcomes these disadvantages.

ANS: The disadvantages of the central name server are:

1. If there is a bottleneck in the central site, then three is a performance issue.
2. Loss of autonomy.
3. The sites can't create any new database If the central site fails.

Alternately, To prefix the object with the id of the site that was created.