Connor Brown

CS44800 Homework 4

**Question 1. (0.50 points)**

1. What are the main software modules of a DDBMS (Distributed Database System)? Discuss the main function of each of these modules in the context of the client/server architecture.
   1. Main software modules are data management and the user interface. The data management can include further functions such as query processing, optimization, transaction management, consistency checking, etc. Data management is mainly done within the server, however the client can also include some data management functions. The client also manages the user interface.
2. What are the main reasons for and potential advantages of distributed databases?
   1. The main reasons for the use of a distributed database is shared memory and transparent distribution for the user.
   2. The potential advantages of distributed databases is increased reliability and availability, improved performance, scalability, and different levels of transparency.

**Question 2. (1.00 point)**

Compare relational databases and the MapReduce/Hadoop/BigData Systems. Use the following criteria in your discussion:

1. Structured and unstructured data support
   1. Relational databases work well with structured data.
   2. Big data systems work well with structured, semi-structured, and unstructured data.
2. Cost of deployment
   1. Relational databases require a license to use
   2. Big data systems are free to use.
3. Fault-tolerance
   1. Relational databases have low fault-tolerance.
   2. Big data systems have high fault-tolerance.
4. Query expressiveness
   1. Relational databases use SQL.
   2. Big Data systems use HQL.
5. Transaction processing support
   1. Relational databases have static schema, fast reads, low throughput, and vertical scalability.
   2. Big data systems have dynamic schema, fast reads and writes, high throughput, and horizontal scalability.

**Question 3. (0.50 points)**

Consider the following four schedules for transactions T1, T2 and T3:

1. r1(X); r3(X); w1(X); r2(X); w3(X);
2. r1(X); r3(X); w3(X); w1(X); r2(X);
3. r3(X); r2(X); w3(X); r1(X); w1(X);
4. r3(X); r2(X); r1(X); w3(X); w1(X);

(a) For each schedule, specify whether or not the schedule is conflict serializable and why.

(b) Then, for each serializable schedule, give the equivalent serial schedule.

1. Not serializable because there is a schedule conflict between T1 and T3
2. Not serializable because there is a schedule conflict between T1 and T3
3. Serializable
   1. r2(X); r3(X); w3(X); r1(X); w1(X);
4. Not serializable because there is a schedule conflict between T1 and T3

**Question 4. (0.50 points)**

Consider a database with objects X and Y and assume that there are two transactions T1 and T2. Transaction T1 reads objects X and Y and then writes object X. Transaction T2 reads objects X and Y and then writes objects X and Y. Give two example histories over transactions T1 and T2 that result in a read-write conflict and a write-write conflict respectively. Note: two examples for each type of conflict (i.e., four examples in total).

1. Read-Write Conflict
   1. T2:r(X), T2:r(Y), T2:w(X), T1:r(X) results in a dirty read
   2. T2:r(X), T2:r(Y), T1:r(X), T1:r(Y), T1:w(X) results in an unpredictable read
2. Write-Write Conflict
   1. T2:r(X), T2:r(Y), T1:r(X), T1:r(Y), T1:w(X), T2:w(X) results in an overwrite of uncommitted data
   2. T1:w(X), T2:w(Y), T1:w(Y), T2:w(X) results in a blind write.

**Question 5. (0.50 points)**

1. Discuss the different types of transaction failures.
   1. Computer failure (system crash) - A hardware, software, or network error that causes the computer to fail or hang during an ongoing transaction.
   2. Transaction error - caused by a user aborting the transaction, when there is a logical programming error, or an operation dividing by zero.
   3. Concurrency control enforcement - a forced abort by the system when a transaction causes a deadlock or violates serializability.
   4. Disk failure - some disk blocks lose their data because of a read or write malfunction or because of a read/write head crash.
   5. External causes/ exception conditions - a transaction may be failed if unnecessary data is not found, or an exception condition is not met.
2. What is meant by catastrophic failure?
   1. A catastrophic failure is when a stable, secondary storage device gets corrupt. This causes all the data stored inside to be lost.
3. What is the system log used for?
   1. The system log s used to keep the history of all action executed by the database.
4. What are checkpoints and why are they important?
   1. A checkpoint is a mechanism where all the previous logs are removed from the system and stored permanently in a storage disk. They are important because they are used in the recovery process in case of a crash. The system works back from the crash towards the last checkpoint and reads the logs.