13. Recovery

Most of content here is borrowed from following sources-
>> Book Elmasri/Navathe

Recovery System - log

To recover from failures that affect transactions, the system maintains a log to keep track of all transaction operations that affect the values of database items.

Recovery manager keeps track of the following operations of transactions-

- Begin Transaction
- read or write
- End Transaction operation by the transaction: i.e. Commit or Rollback

Log is kept on disk and backed up on some other storage to against any catastrophic failures

Typically (dependents on recovery technique), following entries called log records, are placed in log-

[start_transaction, T]: indicates start of transaction T, where T is a unique ID for the transaction

[write, T, X, OldValue, NewValue]
T has changed X from old-value to new value

[read, T, X]

[commit, T] or [abort, T]

Recovery Process

- One of the common approach to enable recovery is, all transaction updates occur in buffer, and does not write the buffer to disk till transaction commits
    - However, memory may not be large enough to accommodate all such data, and in practice some "stealing" is done
- Also, buffers may not be flushed to the disk, every time a transaction commits – because of efficiency reasons it is done by some buffer/cache management techniques
- Recovery is done by redoing and undoing some of the operations such that
    - Database consistency is maintained
    - Concurrency control is enforced

- Typically, we **undo** effects of transaction that were uncommitted transactions at the time of crash; and
  redo effects of transaction that were committed at the time of crash.

Recovery using log files

- To facilitate recovery, recovery subsystem may also need to-
  - log entries forcefully written to disk before actually committing a transaction– Write Ahead Logging Protocol
  - A checkpoint when memory buffer is flushed to the disk
  - A list of active transactions at the time of check point

Check-pointing

- Another type of entry in the log is called a checkpoint.
- A [Checkpoint] record is written into the log periodically at the point when the system writes out all DBMS buffers that have been modified to the database on disk.
- As a consequence of this, all transactions having [commit, T] before this point need not be redone in case of system crash, since all the updates of these transactions were already recorded in the database during check-pointing.
- The recovery manager of a DBMS must decide at what interval to take a checkpoint – it may be fixed time interval or a number of transactions committed.

Recovery Algorithms

- Write Ahead Logging (WAL) protocol
- ARIES Recovery Algorithm