FIT 100 Organizing Business Data

- Companies, universities, government agencies,etc. have many database applications in common
  - Employee records
  - Payroll records
  - Customers/clients/students records
  - Products and services, etc.
- Databases at the UW described in the book are completely adequate for representing and managing this data
- These databases can be controlled
  - Changes can be made by "authorized" employees
  - Changes can be made periodically, in batches

FIT 100 Maintaining On-Line Databases: Transactions

- Many databases are useful if they are online and can be changed interactively
- Online changes of databases fall under two general categories: transactions and concurrent interactions
- A transaction involves more than one operation (display or change) to one or a few records of a database
  - Cash transfers between accounts
- Many transactions are taking place at once, typically
- Keeping your DB "correct" is a problem

FIT 100 Correct Database?

- Because each transaction is making changes to data in more than one place, an all or nothing policy is in place. Either the whole transaction takes place, or none of it does.
- Consider the following transactions that will transfer $100 from your Savings to your Checking account:
  - Withdraw $100 from Account #232323 Savings and
  - Deposit $100 to Account #232323 Checking
- When the transaction finishes, the *balance* in checking should be increased by $100, and savings decreased by $100
- What if there is only $ 90 in Savings?

FIT 100 Maintaining On-Line Databases: Concurrent Interactions

- Multiple concurrent interactions on the same data source can occur
- Possibility that two computers can be making changes to the same data at the same instant, possibly corrupting it
  - Credit card purchases
  - ATM withdrawal of cash
  - Flight reservation
  - …

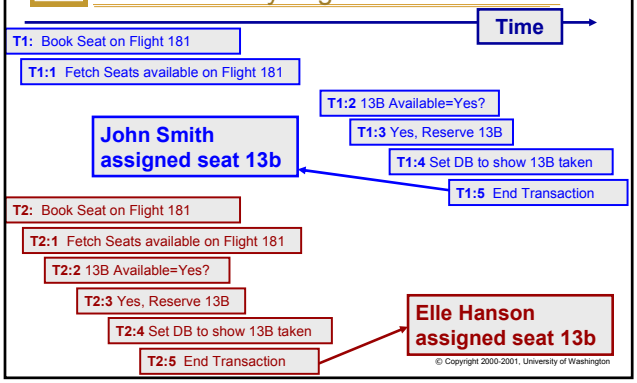
Visa processed 110,086,395 transactions worldwide on December 14, 1998 , a 1-day world's record

- The Web and Internet deal with millions of multiple, simultaneous operations
- Keeping your DB "correct" is a problem. Concurrency control is the fix

FIT 100 Correct Database?

- Two or more computers making changes to the same data, at the same instant could be a problem!
- Example:
  Two interactions by two travel agents in two different cities
  - T1: Reserve John Smith seat 13B on United Flight 181 on 4/2/02
  - T2: Reserve Elle Hanson seat 13B on United Flight 181 on 4/2/02
- When the interactions finish, only one of these two people should have a reservation
- What if the operations take place at the same time?

FIT 100 A Reason Why Flights Overbook???



FIT 100

Someone is NOT going to be happy with this outcome!

FIT 100 Correctness

- The DB System must assure that every DB change happens as if the operations happened one-at-a-time.
- This one-at-a-time protocol solves the "problem" with T1 and T2:
  - T1 is applied first, then T2: John Smith gets the seat
  - T2 is applied first, then T1: Elle Hanson gets the seat
- Online Database systems make sure such concurrency problems do not arise by "locking" the data (only one computer at a time can unlock and change the data)
  - Locks can be done at the database level, the table level, or the record level
- This is also called concurrency control.

FIT 100 Reliability

- What happens to the database when…
  - The power goes out?
  - Someone spills a drink into the disk drives?
  - The computer crashes with all the changes to the DB for the last 3 hours still in RAM (volatile memory)
  - A new employee accidentally deletes the grades before they are sent out to students?
  - A virus cleans off data storage disks?
  - A hacker gets into the business DB and begins transferring funds to a Swiss Account?
  - A plane destroys 2 buildings and, among other things, all the data stores that are part of hundreds of businesses?
  - …
- Any of these situations can happen, plus many more.

FIT 100 Basic Mechanisms

- Several techniques preserve the integrity of the data from inside and OUTSIDE the database
  - Error detection/correction in the hardware
  - Passwords and authentication to verify that the person making changes has that authority
  - Validation … verifying that changes to the DB are "plausible"
  - "Commitment" … keep record in a safe place of all changes made to the database, and then when it has been verified, make the actual change effective
- Backup copies of a database (like the UW's) must be made regularly, and kept off-site

Do you back up your information?

FIT 100 Redundancy Can Be Very, Very Good

- To protect against computer crashes, disk failures, loss of power, etc. duplicate the hardware, disks, power sources, etc.
- The duplicate systems can compare answers as a means of detecting errors
- There are systems specially designed to recover disk failures
- By keeping a snapshot of the database and a record of all the transactions, *it is possible in case of catastrophic disaster to reconstruct the database by applying all of the transactions to the old database*