Introduction and Issues in File Processing System

21CSC205P – DATABASE MANAGEMENT SYSTEMS

Database Management System (DBMS)

DBMS:

- **Definition** - a collection of interrelated data and a set of programs to access those data.
- goal - store and retrieve the data, conveniently and efficiently.
- designed to manage large bodies of information.
- Database - collection of data.
- Management - defining structures for the storage of information and providing mechanisms for the manipulation of information
- ensure the safety of the information stored.

Database Management System (DBMS)

- **Data:**
  - Raw, unprocessed facts.
  - unorganized
  - Example: 35, Vinod, Ghaziabad.
- **Information:**
  - Processed Data
  - organized
  - Example: The age of Vinod is 35 residing in Ghaziabad.
- **Metadata:**
  - Data about other data
  - Example: Author, date created, date modified, file size etc;

APPLICATIONS

Enterprise Information :

- Sales:
  - For customer, product, and purchase information.
- Accounting:
  - For payments, receipts, account balances, assets and other accounting information.
- Human resources:
  - For information about employees, salaries, payroll taxes, and benefits, and for generation of paychecks.
- Manufacturing:
  - For management of the supply chain and for tracking production of items in factories, inventories of items in warehouses and stores, and orders for items.

APPLICATIONS

Enterprise Information :

- Online retailers:
  - For sales data, online order tracking, generation of recommendation lists, and maintenance of online product evaluations.

APPLICATIONS

Banking and Finance:

- Banking:
  - For customer information, accounts, loans, and banking transactions.
- Credit card transactions:
  - For purchases on credit cards and generation of monthly statements.
- Finance:
  - For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds; also for storing real-time market data to enable online trading by customers and automated trading by the firm.

APPLICATIONS

Universities:

- For student information, course registrations, and grades (in addition to standard enterprise information such as human resources and accounting).

Airlines:

- For reservations and schedule information. Airlines were among the first to use databases in a geographically distributed manner.

Telecommunication:

- For keeping records of calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.

Issues in File Processing System

Issues in File Processing System

- Data redundancy and inconsistency
- Difficulty in accessing data
- Data isolation
- Integrity problems
- Atomicity problems
- Concurrent-access anomalies
- Security problems

Data redundancy and inconsistency

- different programmers
- different files
- different structures
- different programming languages
- duplication – redundancy leads to higher storage and access cost
- data inconsistency

Data redundancy and inconsistency

Employee ID	Employee Name	Department	Salary
602090	Rohit	Human Resource	50000
602091	Rahul	QA	60000
602092	Virat	Accounts	40000
602093	Hardik	Sales	30000
602094	Dhoni	Human Resource	60000
602095	Surya	QA	70000
602096	Ishan	Accounts	45000
602097	Green	Sales	35000
	.		
	.		
	.		

Data redundancy

Employee ID	Employee Name	Department	Salary
602090	Rohit	Human Resource	50000
602090	Rohit	Human Resource	60000
602091	Rahul	QA	60000
602092	Virat	Accounts	40000
602093	Hardik	Sales	30000
602094	Dhoni	Human Resource	60000
602095	Surya	QA	70000
602096	Ishan	Accounts	45000
602097	Green	Sales	35000
	.		
	.		

Data inconsistency

Employee ID	Employee Name	Department	Salary
602090	Rohit	HR	50000
602091	Rahul	QA	60000
602092	Virat	Accounts	40000
602093	Hardik	Sales	30000
602094	Dhoni	Human Resource	60000
602095	Surya	QA	70000
602096	Ishan	Accounts	45000
602097	Green	Sales	35000
	.		
	.		

Difficulty in accessing data

- find out the names of all students who live within a particular city.
- Two options
  - either obtain the list of all students and extract the needed information manually
  - or ask a programmer to write the necessary application program.
- If such a program is written and after several days we need data on those students who have taken at least 60 credit hours
- such a list cannot be extracted and again we need to go for two options
  - Manual
  - Ask programmer
- Hence file systems do not allow needed data to be retrieved conveniently and efficiently.

Data isolation

- different programmers - different files - different structures – different formats – different location
- Data’s are scattered
- So data isolation is difficult.

Integrity problems

- Can’t enforce **consistency constraints**
- Example: Account balance should not be zero
- Programmer enforce this constraint in the system by adding appropriate code.
- new constraints are added, it is difficult to enforce.
- The problem is compounded when constraints involve several data items from different files.

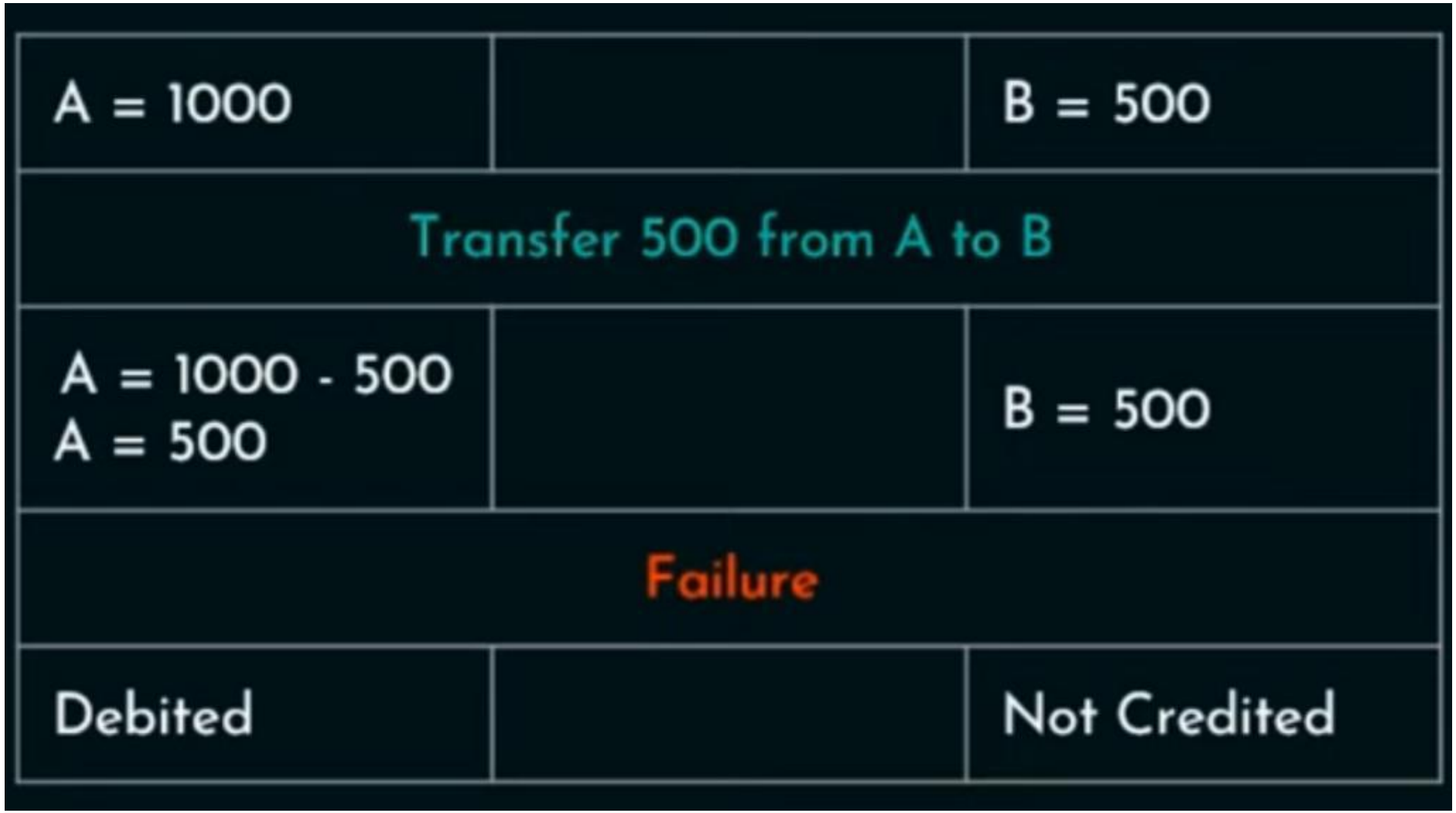
Integrity problems

Employee ID	Employee Name	Department	Salary
602090	Rohit	Human Resource	0
602091	Rahul	QA	60000
602092	Virat	Accounts	40000
602093	Hardik	Sales	30000
602094	Dhoni	Human Resource	60000
602095	Surya	QA	70000
602096	Ishan	Accounts	45000
ABCDEF	Green	Sales	35000
	.		
	.		
	.		

Atomicity problems

- difficult to ensure atomicity in a conventional file-processing system.
- Any system is subject to failure
- if a failure occurs, the data should be restored to the consistent state that existed before the failure.
- Example: Fund transfer
- the fund transfer must be atomic
- Atomic – ALL or NONE

Atomicity problems



Concurrent-access anomalies

- multiple users should be allowed to update the data simultaneously
- In file system, it leads to inconsistency.

Concurrent-access anomalies

$1000		
Clerk A Read $1000	Concurrent	Clerk B Read $1000
Debit $200 Write $800		
		Debit $300 Write $700

Security problems

- Authentication and Authorization in the file system are very difficult.
- example, in a university, account personnel need to see only that part of the database that has financial information. They do not need access to information about academic records.

THANK YOU