

Database Issues

Business Data Management and Analytics

**Contributions by Arthur Adamopoulos, Vince Bruno,
Bill Davey and Patrick Poppins**

Data Administration

- Manage the data needs of the entire organization so all data is available.
- Database Planning
- Data Analysis, Design and Implementation
- Data Maintenance and Protection
- Education and Training
- Support to end users
- High level management role

Database Administration

- Manage physical database servers.
- More of a technical, hands on role.
- Installation and configuration of DBMS
- Data Analysis, Design and Implementation
- Performance Tuning
- Managing Security and Policies
- Managing backups and recovery

Changing Roles

- In larger companies, these two roles may be performed by distinct people or groups.
- In smaller organisations or departments, one person will often perform both roles.
- DBA roles are becoming specialized:
 - Product specific (eg: Oracle DBA)
 - Overlap with developers (server programming)
 - Database types (eg: data warehouses)
 - Specific App Software (eg: SAP or PeopleSoft)

LOST UPDATE PROBLEM

- Sharing data among many users using a database
- How do we solve this problem?
 - When a user/process wants to use a resource, LOCK IT.
 - No other user/process may change it until it is UNLOCKED.

Level Of Locking

1. **Database** - backup
2. **Table** - batch work (updates)
3. **Block/Page** - most common
4. **Record** - 1 record, most commonly required

Lock Types

- Shared Locks
 - other users/processes can read but not update; i.e. allowing querying;
- Exclusive Locks
 - other users can do NOTHING.

Other Locking issues

- Business transactions
- Impact of locking on Transactions
- Deadlocks arise from more than one process holding a resource and requesting other resources which are in turn held by other processes. Resolved by
 - Pessimistic locking scheme
 - Optimistic locking scheme (Versioning)
 - Deadlock Watchdog

Backups

- If your computer room has a fire and everything is lost, how long would it take to replace:
 - Hardware
 - Operating System
 - DBMS (eg: Oracle software)
 - Data in the database
- Which is the most valuable ???

Database Recovery - Backup

- Full Backup
 - everything is backed up on some sort of media.
- Partial Backup
 - backup changes made since last FULL Backup was made.
 - Incremental? Differential?

Database Recovery - Backup

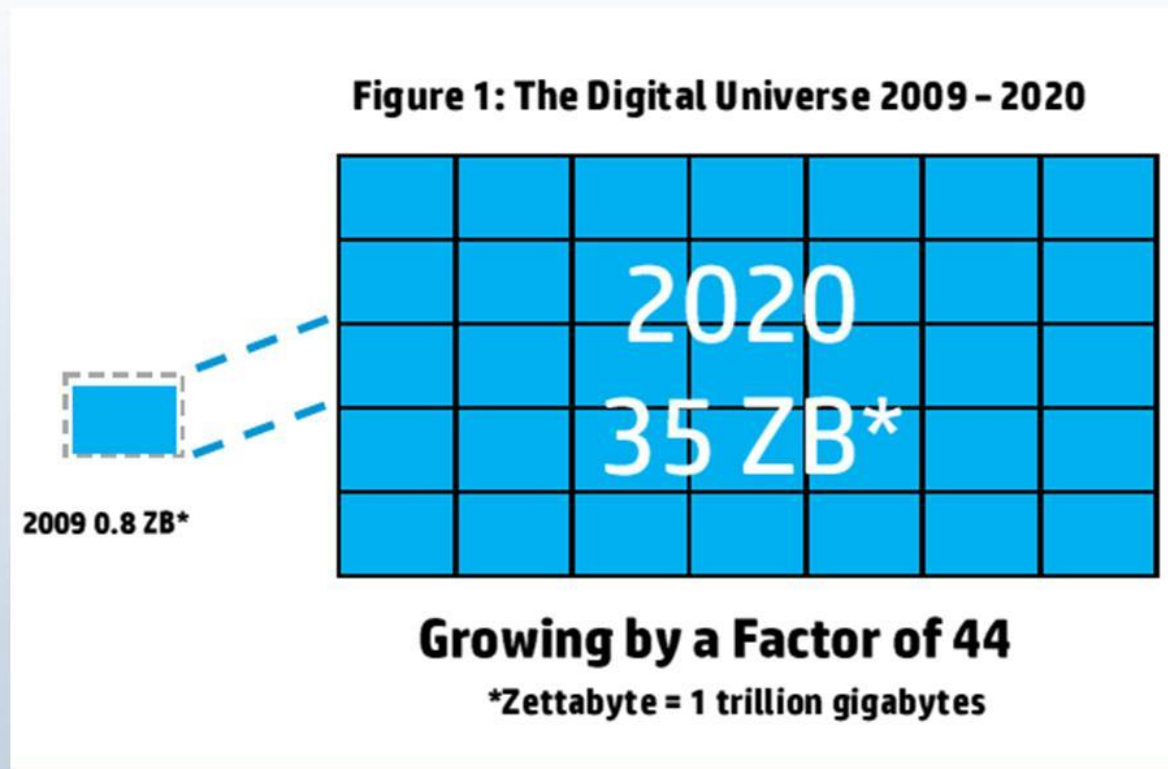
- Recommended backup philosophy
 - FULL Backup each week
 - PARTIAL backups each day
 - ROTATING the media used
 - OFFSITE media storage

Database Recovery – Volume

- Data volume is doubling every two years

- 35 hours of video are uploaded to YouTube every minute; videos are viewed 2 billion times per day
- 5 billion mobile phones in use in 2010
- 30 billion pieces of content shared every month on Facebook

Database Recovery – Volume



Database Recovery - Equipment

- <http://www.oracle.com/us/products/serve-storage/storage/tape-storage/sl8500-modular-library-system/resources/index.html>
- Performance 553 TB/Hr
- Capacity 1,000,000 terabytes

RAID

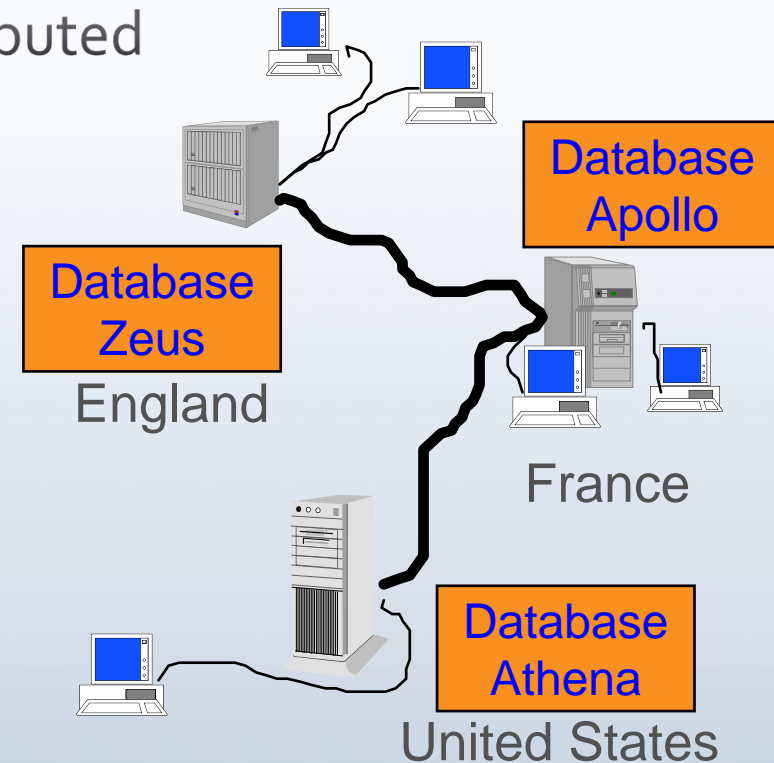
- Redundant Array of Independent Disks
- Instead of using a single disk drive, a RAID controller combines at least 3 drives to act as if they were one drive.
- RAID 5 provides striping and parity.
- Striping means that data writing is spread out over the physical disks as stripes.
- Parity means that redundant data is written on to each drive about the other drives.

RAID – Absolutely Essential

- The parity information is the key.
- If a single drive in the RAID set fails,
THE SERVER WILL KEEP RUNNING !!.
- When the faulty drive is removed and a new one inserted, the new drive will automatically be rebuilt using the parity information from the other drives.
- RAID allows maximum server uptime and is a must for even the smallest server.
- You will learn more about RAID in 'Networking in Business'.

Distributed Databases: Definition

- A single logical business database distributed over multiple physical locations.
- Multiple independent database servers at each location
- Connected by a network.
- Might be different environments
 - Hardware
 - Operating System
 - DBMS Software
- Primary concept is that a user should be able to work with all data from all the servers, as if they were on one machine.



Distribution Mechanisms - Replication

- Replication
 - Keep multiple copies of the same data on each physical server and update at regular intervals.
 - Must accept that not all data will always be exactly up-to-date
 - Currently the most supported and popular form of distribution

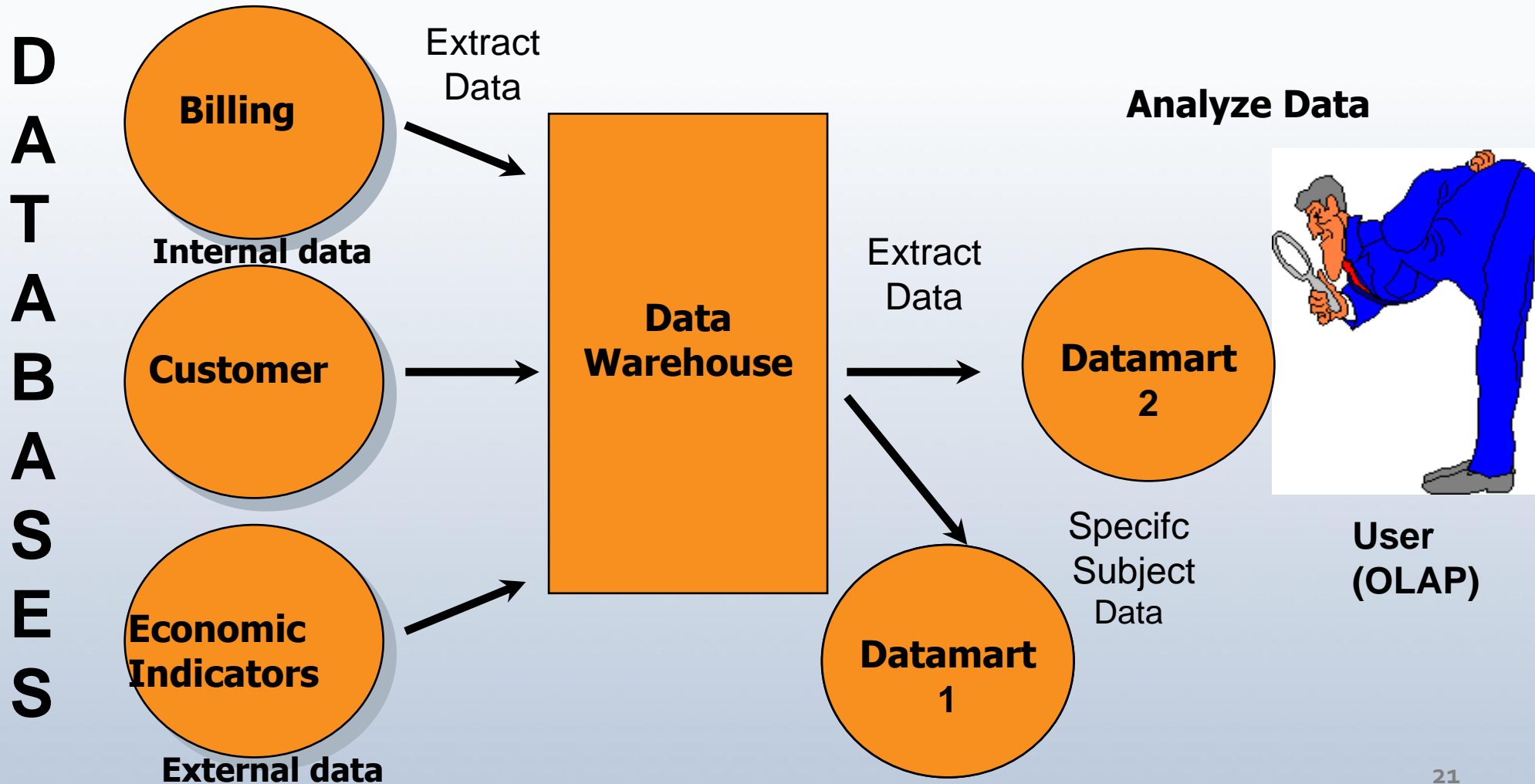
Distribution Mechanisms - Replication

- Replication Types
 - Snapshot
 - take periodic copies of a table or parts
 - Real Time updates
 - as each transaction is made, send transaction details to other sites to update as well.
 - Frequency would depend on needs and comms speed
 - Push/Pull
 - Can be controlled by the originating server (push) or by the receiving server (pull)

Datawarehouses

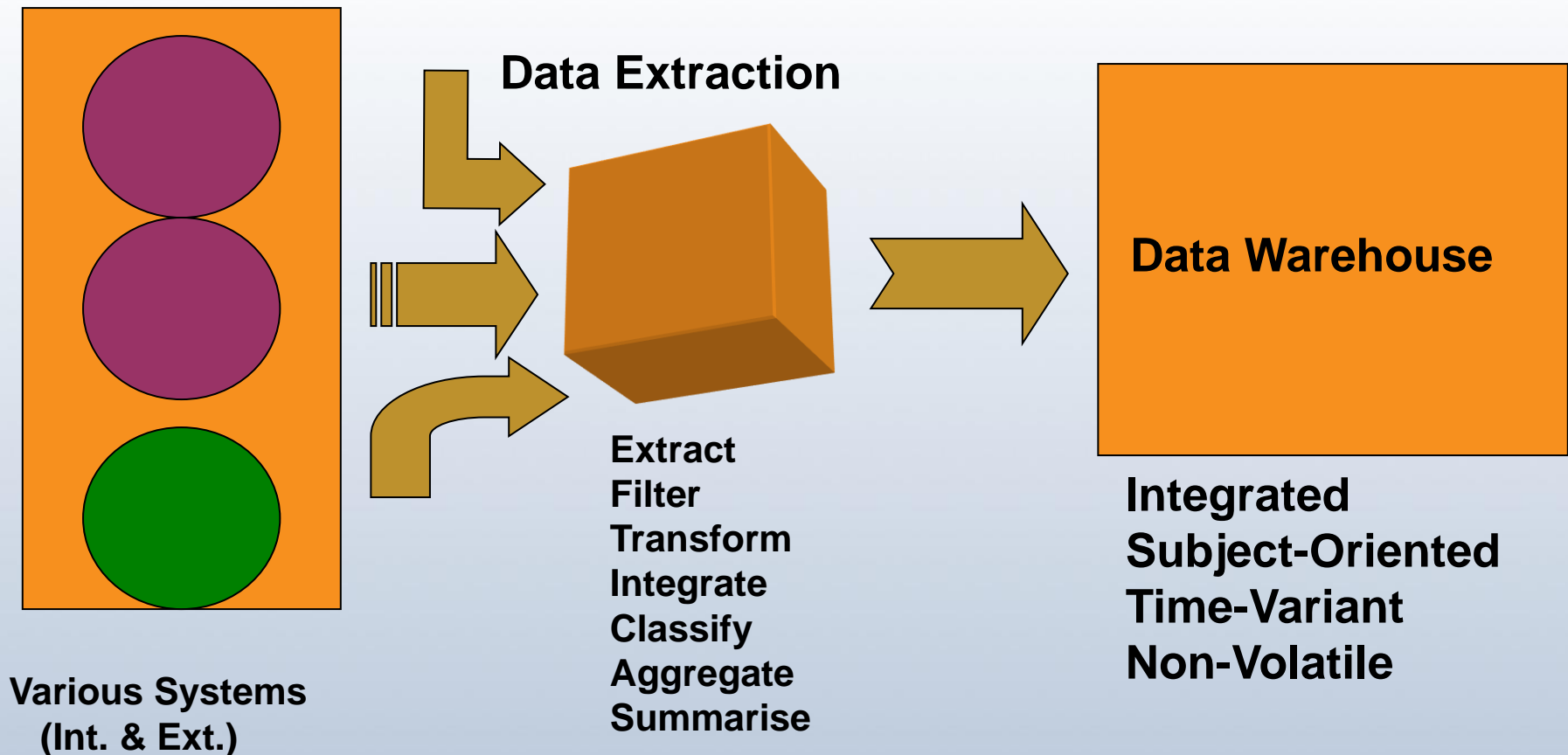
- Operational Databases
 - benefits clerical work
 - transaction processing
 - short term, day to day operations
 - functional system
- Decision Support System
 - information extraction, derived from operational data
 - help decision making process
 - snapshot of operational data
 - “knowledge” worker - management

DataWarehouses



Creating & Maintaining a Data Warehouse

- Operational Data



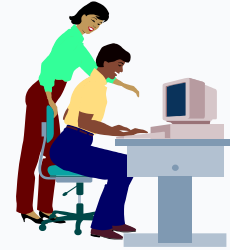
Datawarehouses

- growing demand for decision making tools has created a need for this type of database.
- a new generation of DSS
- OLAP - On-Line Analytical Processing
 - front ends to datawarehouses
 - gives a GUI frontend allow drill down to the operational data level.
- consist of four main characteristics (in relation to data stored)

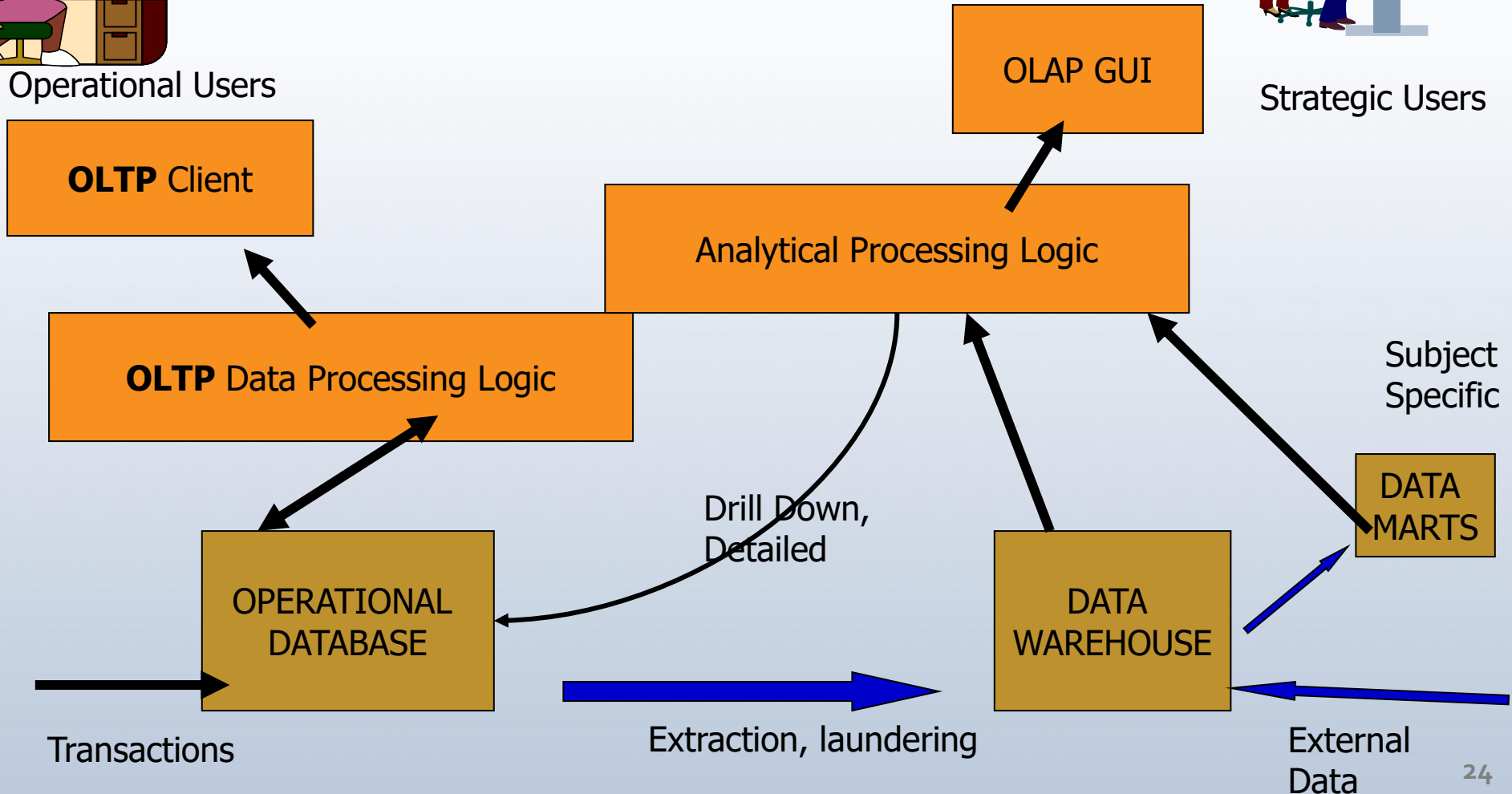
On-Line Analytical Processing SYSTEM



Operational Users



Strategic Users



Datawarehouse: Characteristics

1. Integrated

- centralized?
- consolidates data derived from the entire organisation.
- data can be from multiple data sources, and varying data formats.
- consistent view created of organisation data.

2. Subject Orientated

- data is stored by functional areas of the organisation
- is also organised and aggregated (summarised) by topic; eg. sales, finance, marketing, insurance..etc.
- each topic may have subjects of interest; eg. products, customers, department, regions, classes of insurance...etc.

Datawarehouse: characteristics

3. Time Variant

- operational data focuses on current transaction
- datawarehouses shows data over time
- base on a daily, weekling, monthly..etc basis.

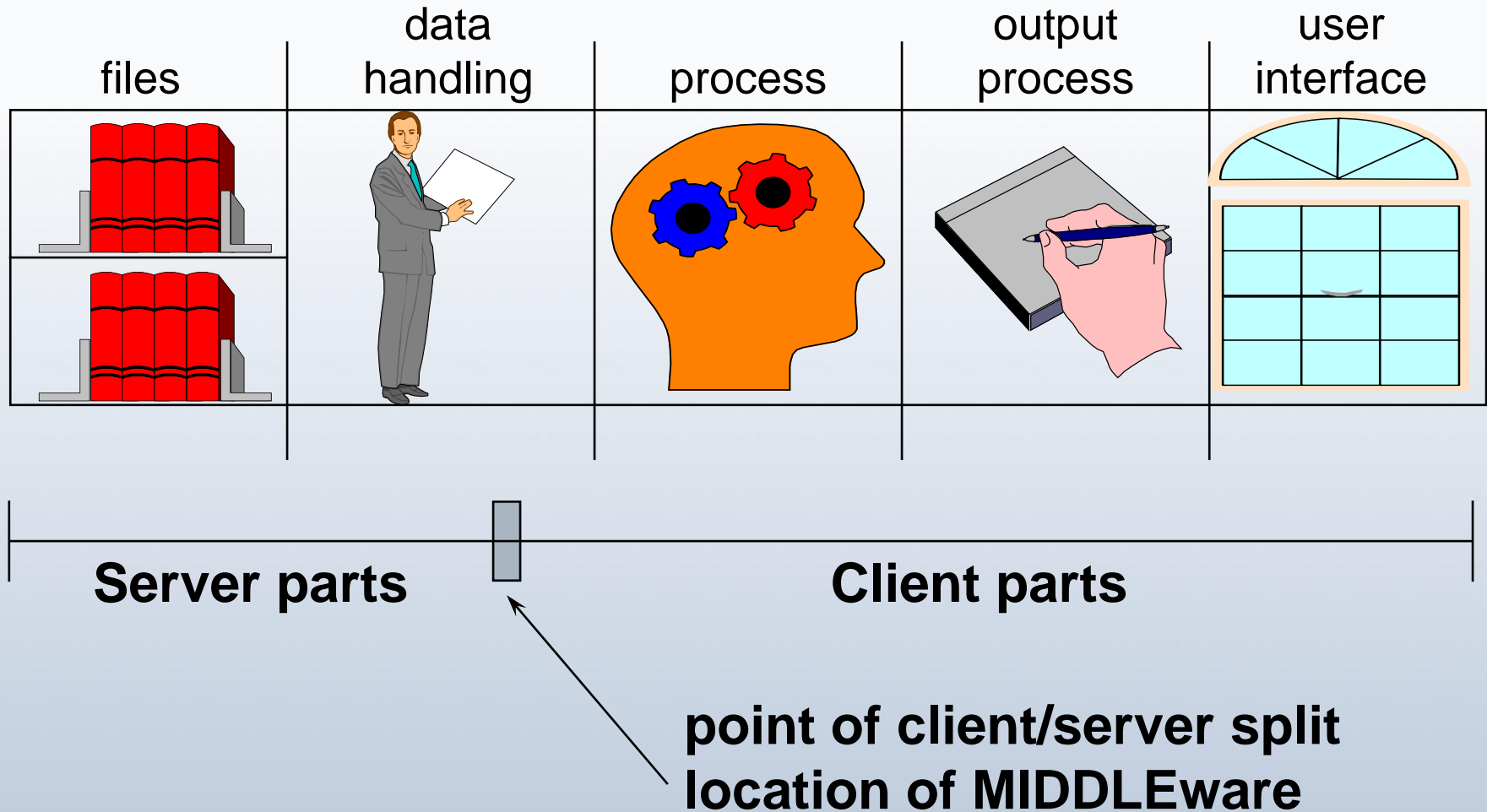
4. Non-Volatile

- data placed in a datawarehouse is never removed
- history
- leads to BIG databases

Datawarehouse - definition

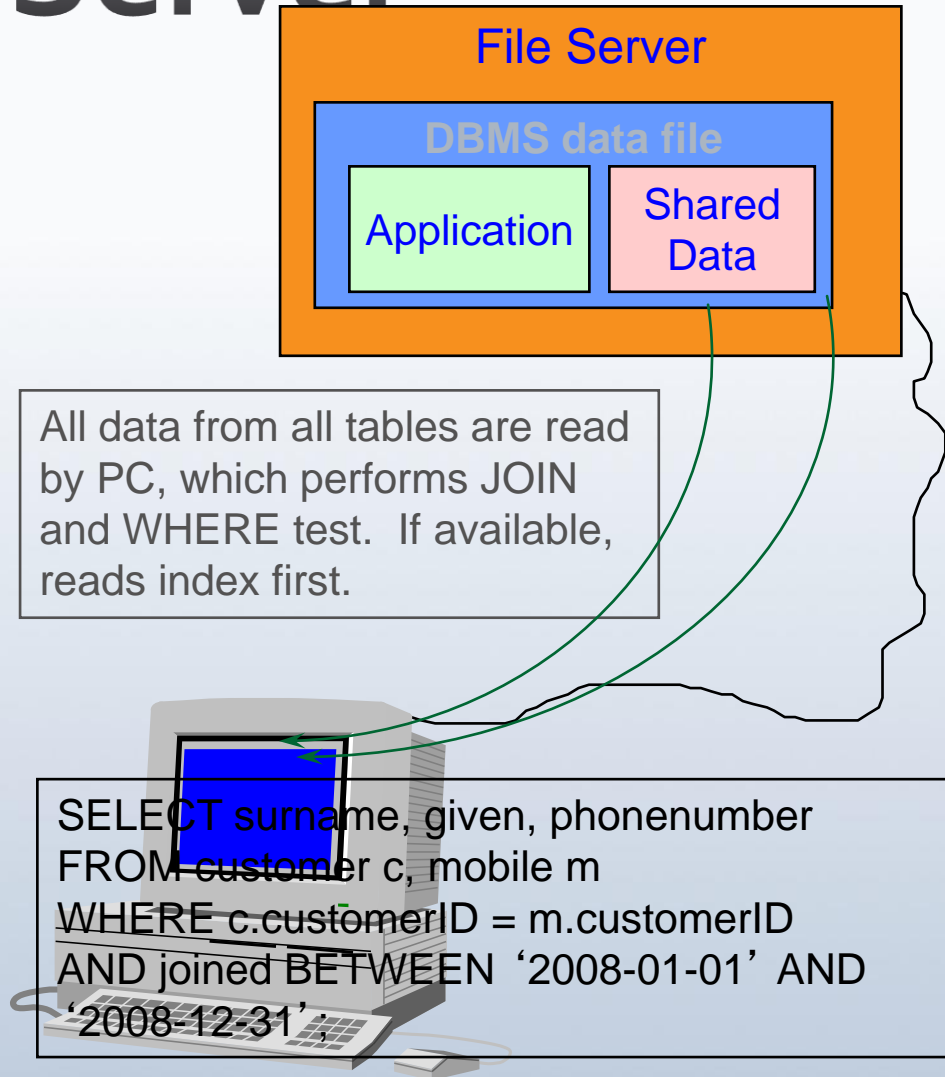
- “The datawarehouse is usually a read-only database optimized for data analysis and query processing. Typically, data are extracted from various sources and are then transformed and integrated (filtered) before loading into the datawarehouse”

Client/Server (2 tier) - application parts

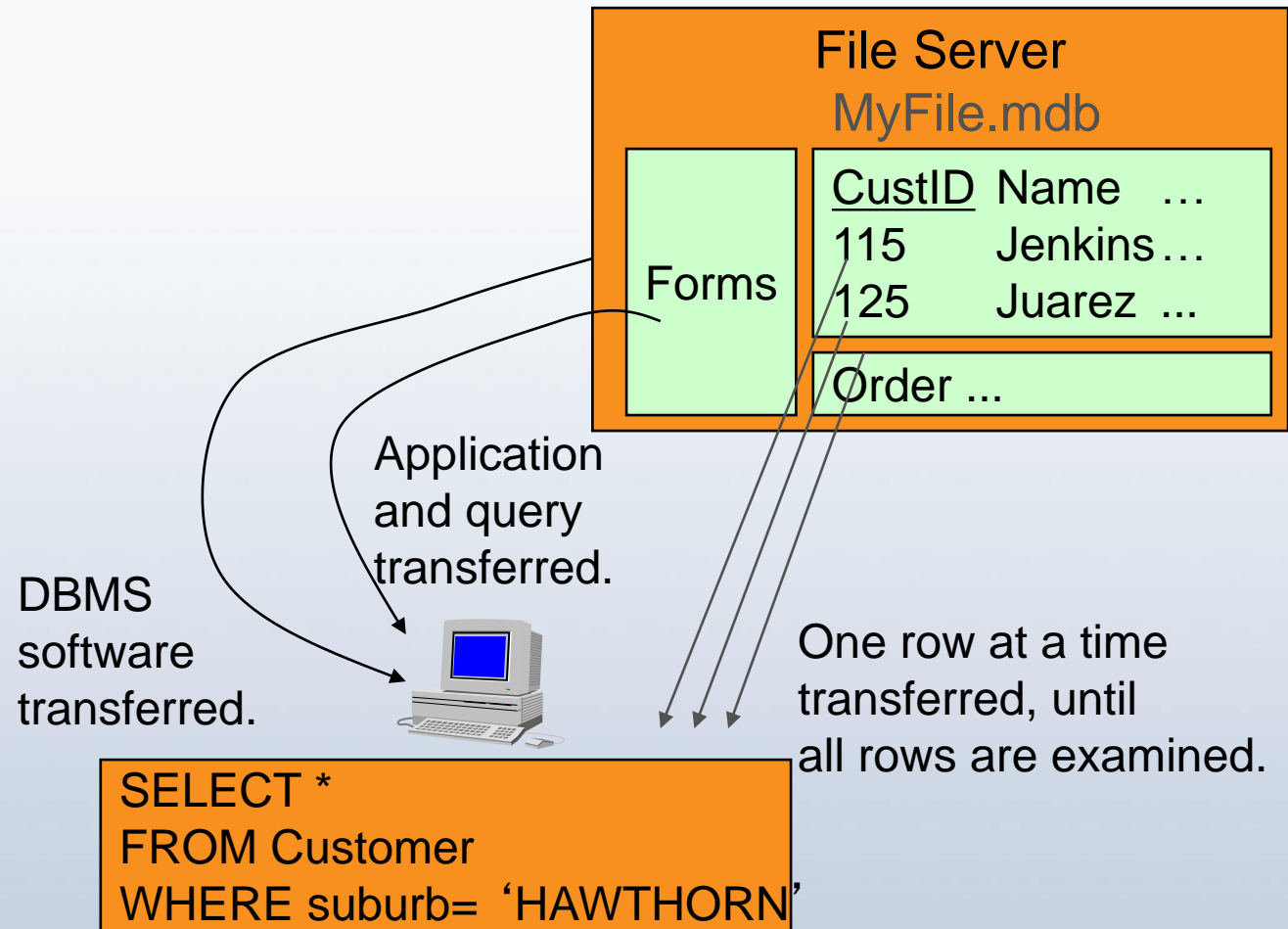


LAN File Server

- **Not** a distributed database.
 - Data file stored on server.
 - Server is passive, appears as giant disk drive to PC.
 - PC processes all data.
 - Retrieves all needed data across the network.
- Performance improvements.
 - Indexes are crucial.
 - Store some data on each PC (replication).
 - Store applications on PC (graphics & forms).
 - Convert to SQL-Server

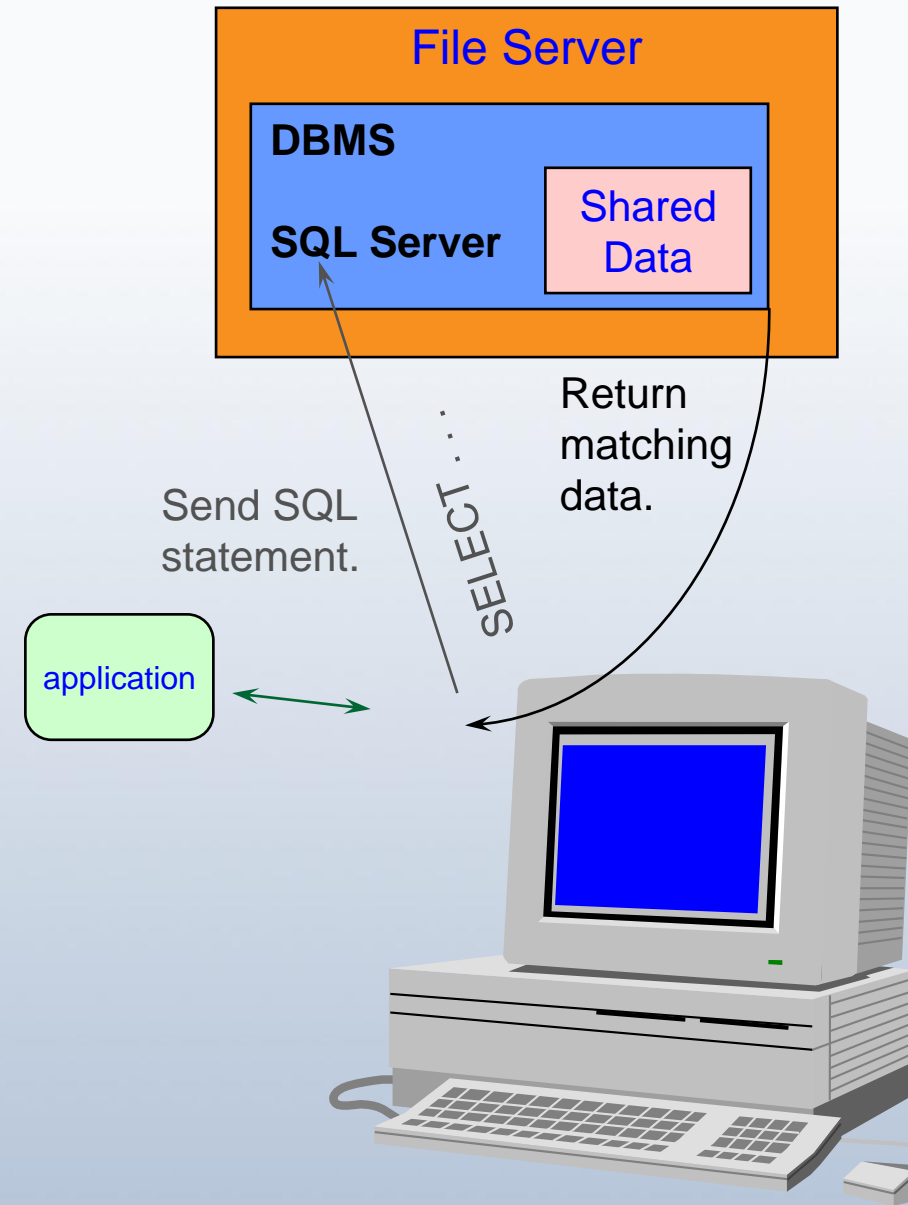


LAN File Server: Slow

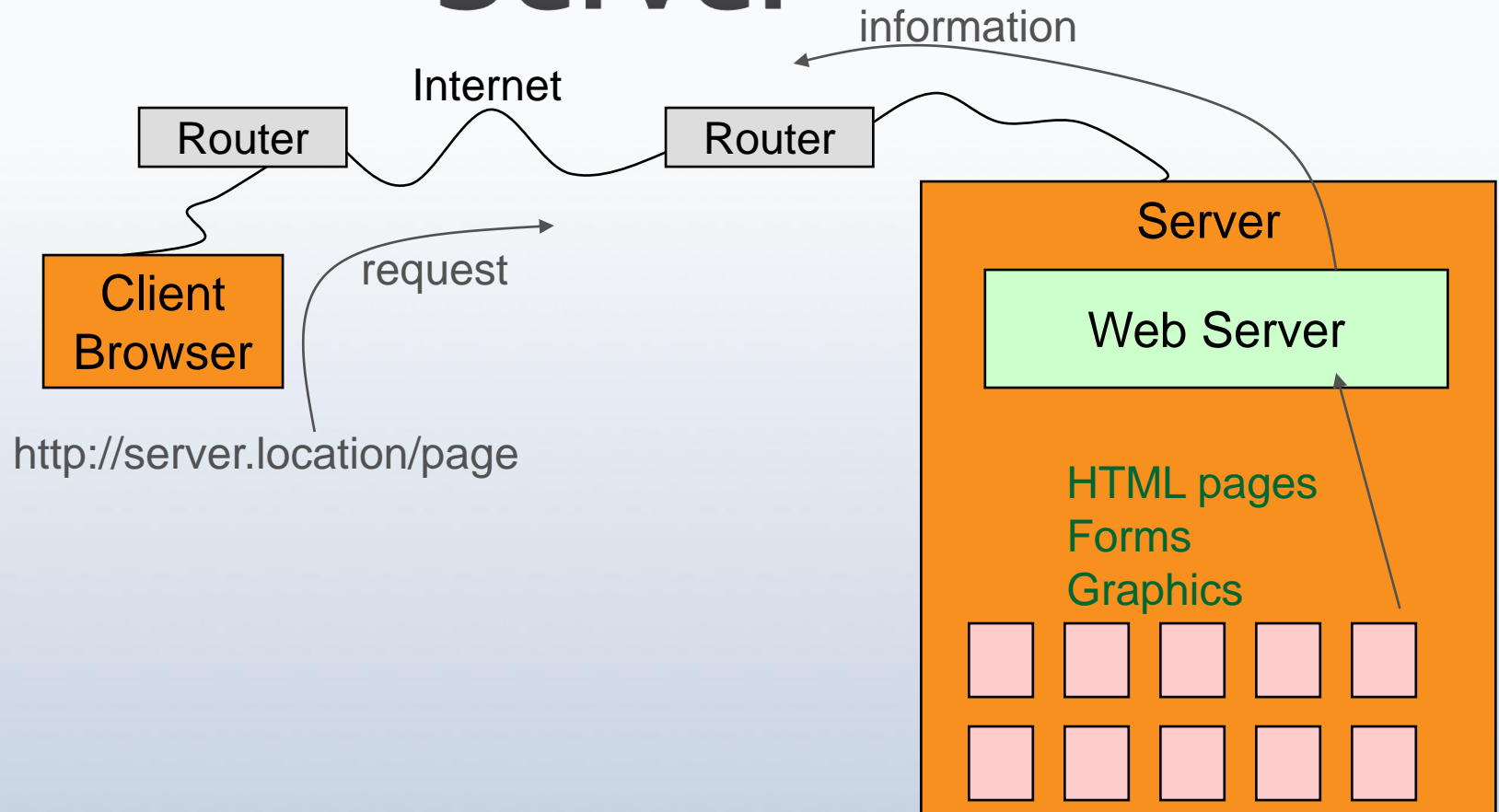


Client-Server Databases

- One machine machine is dominant (server) and handles data for many clients.
- Client machines handle front-end tasks and small data tables that are not shared.



The Internet as Client-Server



Web Server Database Fundamentals

① Request Server/Form.html

