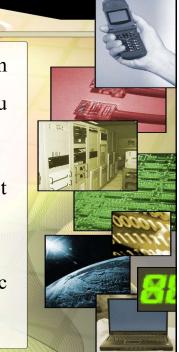


DỮ LIỆU

Dữ liệu là các giá trị phản ánh về sự vật, hiện tượng trong thế giới khách quan và được biểu diễn vật lý dưới nhiều dạng khác nhau.

Ví dụ: một nốt nhạc, một từ trong văn bản, một con số, hình ảnh...

Dữ liệu về đối tượng có thể khác nhau, tùy thuộc vào ngữ cảnh.



CƠ SỞ DỮ LIỆU

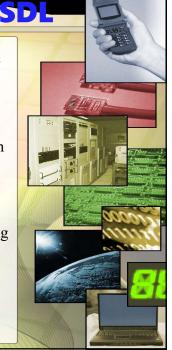
- CSDL (Database) = Tập hợp dữ liệu được tổ chức có cấu trúc liên quan với nhau và được lưu trữ trong máy tính.
- CSDL được thiết kế, xây dựng cho phép người dùng lưu trữ, truy xuất hoặc cập nhật dữ liệu.

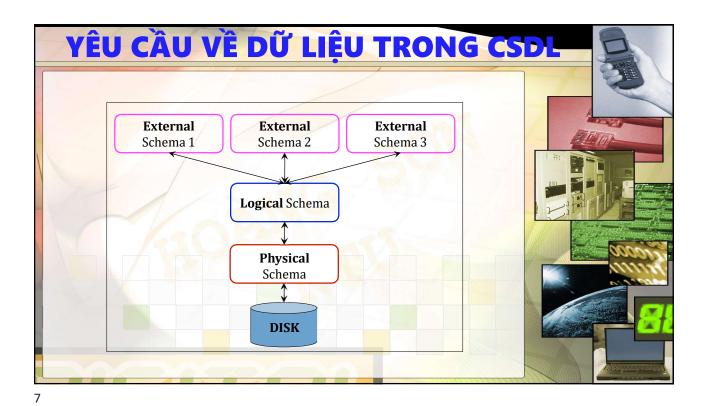


5

YÊU CẦU VỀ DỮ LIỆU TRONG CSDL

- Dữ liệu trong CSDL phải được thể hiện ở các mức độ khác nhau:
 - Mức ngoài (External level)
 - Mô tả một phần của CSDL mà một đối tượng / một nhóm người dùng được quyền tiếp cận
 - Mức luận lý (Logic level)
 - Mô tả những thông tin gì được lưu trữ trong CSDL và những mối quan hệ giữa những thông tin đó
 - Mức vật lý (Physical level)
 - Dữ liệu được lưu trữ như thế nào trên thiết bị lưu trữ.



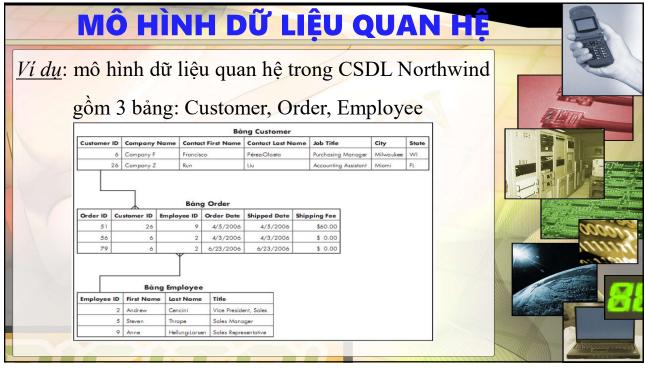


Mô HÌNH DỮ LIỆU
❖ Mô tả cách tổ chức dữ liệu bên trong CSDL.
❖ Mô tả mối quan hệ dữ liệu và các ràng buộc được định nghĩa trên dữ liệu đó.
Mô hình dữ liệu file phẳng (Flat-file data model)
Mô hình dữ liệu mạng (Network model)
Mô hình dữ liệu phân cấp (Hierarchical model)
Mô hình dữ liệu quan hệ (Relational model)
Mô hình dữ liệu hướng đối tượng (Object Oriented model)

MÔ HÌNH DỮ LIỆU QUAN HỆ

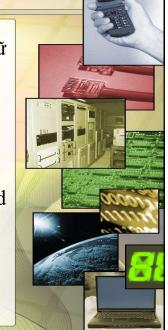
- Dữ liệu được biểu diễn dưới dạng bảng với các hàng và các cột
 - CSDL là tập hợp các bảng (còn gọi là quan hệ)
 - Mỗi hàng là một bản ghi (record), còn được gọi là bộ (tuple)
 - Mỗi cột là một thuộc tính, còn được gọi là trường (field)
- Dữ liệu trong hai bảng liên hệ với nhau thông qua các cột chung.
- Có các toán tử để thao tác trên các hàng của bảng.

9



QUẢN LÝ DỮ LIỆU

- Quản lý lượng lớn dữ liệu bao gồm cả việc lưu trữ dữ liệu và cơ chế thao tác trên các dữ liệu đó.
- Hai hệ thống quản lý dữ liệu khác nhau:
 - Hệ thống quản lý dựa trên tập tin (File-based systems)
 - Hệ thống CSDL (Database systems)



11

HỆ QUẨN TRỊ CSDL

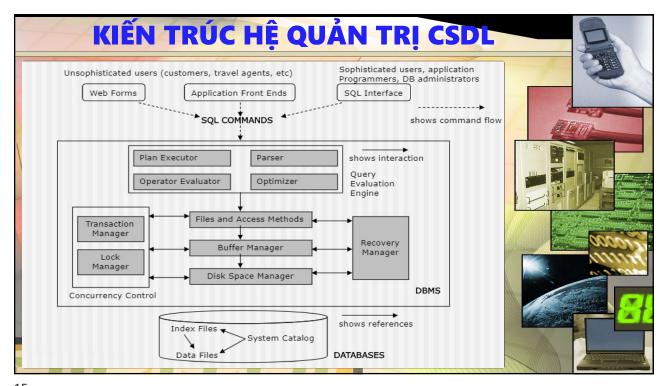
- Là một hệ thống phần mềm cung cấp các công cụ để xây dựng, khai thác và quản lý CSDL.
 - Xây dựng (Sử dụng ngôn ngữ DDL): Định nghĩa cấu trúc CSDL, lưu trữ dữ liệu
 - Khai thác (Sử dụng ngôn ngữ DML): Truy vấn dữ liệu,
 cập nhật dữ liệu
 - Quản lý:
 - Quản lý an toàn và bảo mật
 - Điều khiển truy xuất đồng thời.
 - Khôi phục khi có sự cố.

•





LỊCH SỬ PHÁT TRIỂN CỦA HQTCSDL Decade of RDBMS 1970s 1960s 1980s - 1990s 2000s Mô hình Mô hình mang quan hệ Mô hình No SQL **CODASYL** đối tượng **Database** Mô hình QUEL **SEQUEL** SQL phân cấp Ingres PostgreSQL **dBASE** MongoDB, Oracle **Ingres Corp Sybase SABRE** NoSQL Database, system **Apache MS SQL Server Prototypes** Cassandra, ... for ODBMS System R Non-Stop SQL SQL/DS DB₂ Allbase Oracle





KIẾN TRÚC HQTCSDL - GIAO DIỆN LẬP TRÌNI

- * HQTCSDL cung cấp giao diện lập trình dễ sử dụng và ngôn ngữ tương tác với dữ liệu được lưu trữ bên trong CSDL
 - Giao diện: tương tác dòng lệnh (command line), đồ họa
 (GUI)
 - Ngôn ngữ: SQL, SQL mở rộng (tùy thuộc HQTCSDL)
 <u>Ví dụ</u>: MS SQL Server cung cấp ngôn ngữ Transaction SQL (T-SQL)

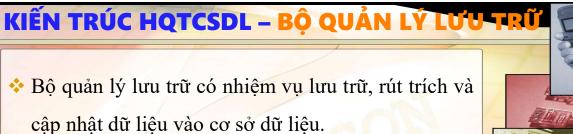


17

KIẾN TRÚC HQTCSDL - GIAO DIỆN LẬP TRÌNH

- Các loại ngôn ngữ truy vấn dữ liệu (SQL)
 - Ngôn ngữ định nghĩa dữ liệu (Data Definition Language DDL): giúp người dùng ra lệnh cho HQTCSDL tạo ra các cấu trúc dữ liệu của CSDL (cách tổ chức dữ liệu và mối liên hệ giữa các đối tượng dữ liệu).
 - Ngôn ngữ xử lý dữ liệu (Data Manipulation Language DML): giúp người dùng thao tác dữ liệu (xem, thêm, xóa, sửa) trong CSDL.
 - Ngôn ngữ điều khiển dữ liệu (Data Control Language DCL): thiết lập quyền truy cập của người dùng trên các đối tượng CSDL.
 - Ngôn ngữ kiểm soát giao dịch (Transaction control languge TCL):
 xử lý các thay đổi có ảnh hưởng đến dữ liệu trong CSDL.

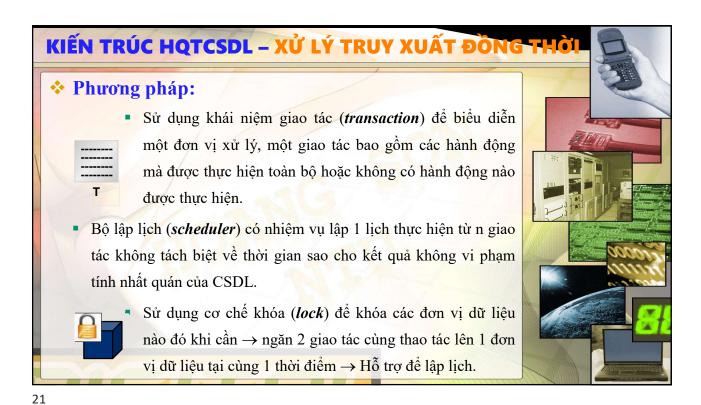


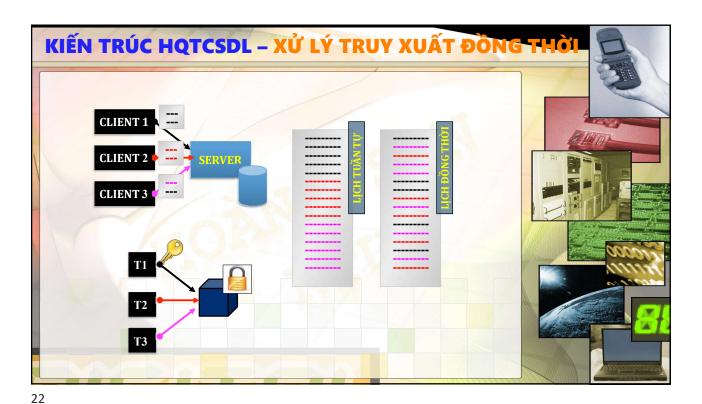


- Điều khiển tương tranh xử lý truy xuất đồng thời.
- Quản lý file và bộ nhớ lưu trữ
- Quản lý khôi phục sau sự cố









KIẾN TRÚC HQTCSDL - XỬ LÝ TRUY XUẤT ĐỒNG THỜI

Vấn đề deadlock

- Do sử dụng cơ chế khóa nên các giao tác sẽ phải chờ khi cần truy xuất 1 đơn vị dữ liệu đang bị khóa.
- Tình huống chờ vĩnh viễn mà vẫn không được truy xuất đơn vị dữ liệu bị khóa gọi là Deadlock (khoá chết)
 - Các giao tác chờ đợi lẫn nhau để được cấp phát tài nguyên và không giao tác nào có thể hoàn tất.
- Thành phần quản lý giao tác sẽ phải can thiệp vào:
 - Hoặc hủy bỏ một trong các giao tác gây deadlock
 - Hoặc ngăn chặn từ trước để không bao giờ sảy ra deadlock



23

KIẾN TRÚC HQTCSDL - AN NINH VÀ BẢO-MẬT

- ❖ Bảo mật dữ liệu: HQTCSDL hỗ trợ các tính năng về chứng thực, phân quyền giúp kiểm soát tốt những người dùng hợp pháp của hệ thống.
- An ninh dữ liệu: HQTCSDL hỗ trợ các phương pháp mã hóa dữ liệu để ngăn chặn các tấn công của những đối tượng tin tặc (đánh cắp thông tin trên đường truyền, đánh cắp nội dung CSDL).





- Files and access methods: tổ chức dữ liệu, hỗ trợ truy cập nhanh đến tập dữ liệu mong muốn.
- ❖ Buffer Manager: đọc dữ liệu vào bộ nhớ để xử lý, và viết vào đĩa để lưu trữ lâu dài theo yêu cầu của chương trình.
- Disk Space Manager: tính toán và quản lý khoảng trống trên đĩa.

KIẾN TRÚC HQTCSDL - QUẨN LÝ FILE & BỘ NHÓ Các khái niệm Tập tin dữ liệu Lưu trữ các metadata (siêu dữ liệu) về cấu trúc của CSDL, đặc biệt là lược đồ của CSDL Chỉ mục Giúp cho việc tìm kiếm dữ liệu được nhanh chóng

KIẾN TRÚC HQTCSDL - KHÔI PHỤC SAU SỰ CÓ

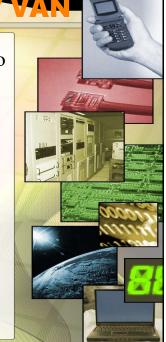
- Mục tiêu: đảm bảo sự tổn thất, sai sót về mặt dữ liệu là ít nhất có thể.
- * Cách tiếp cận: mọi thay đổi trong CSDL phải được ghi nhận lại trong nhật ký (Log)
- Các thành phần hỗ trợ quá trình khôi phục sau sự cố:
 - Quản lý nhật ký (Log manager): đảm bảo ghi nhận đầy đủ và chính xác mọi thay đổi trên CSDL vào nhật ký.
 - Quản lý khôi phục sự cố (Recovery Manager): dựa vào nhật ký để phục hồi lại CSDL về trạng thái nhất quán trước đó (trạng thái thỏa mãn tất cả ràng buộc toàn vẹn của CSDL)

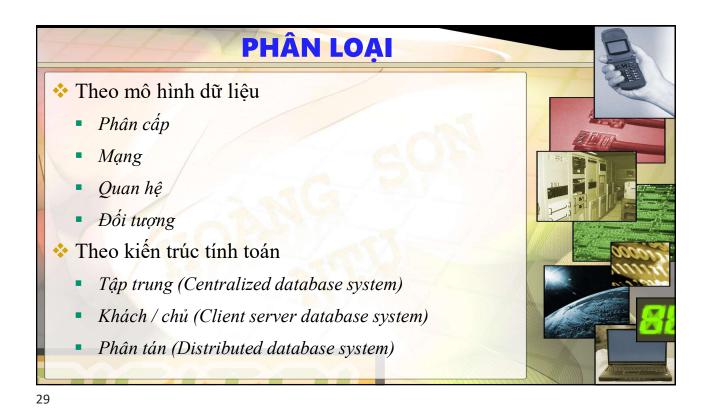


27

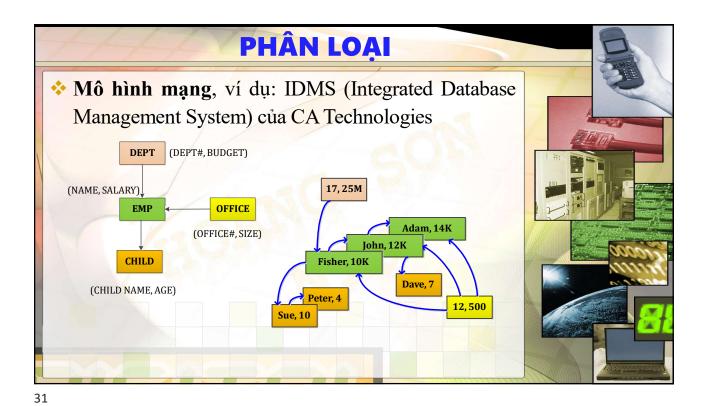
KIẾN TRÚC HQTCSDL - XỬ LÝ TRUY VẪN

- ❖ Biểu diễn câu truy vấn ở dạng ngôn ngữ cấp cao (SQL) và thực hiện câu truy vấn có hiệu quả.
- Các giai đoạn xử lý truy vấn
 - Phân tích (parser)
 - Tối ưu hóa câu hỏi (query optimizer)
 - Lập kế hoạch thực hiện (plan executor)
 - Thực hiện tính toán (operator evaluator)

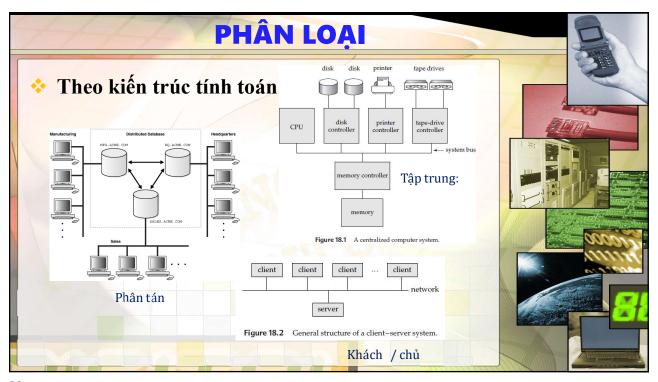




PHÂN LOẠI * Mô hình phân cấp, ví dụ: IMS (Information Management System) của IBM (DEPT#, BUDGET) DEPT (NAME, SALARY) 17, 25M **CHILD OFFICE** Adam, 14K John, 12K (CHILD NAME, AGE) (OFFICE#, SIZE) Fisher, 10K 12,500 Dave, 7 Peter, 4 12,500 Sue, 10 **12, 500**



PHÂN LOẠI * Mô hình quan hệ, ví dụ: SQL Server của Microsoft OCCUPIED DEPT 12 Fisher 17 25M **DEPT** (DEPT #, BUDGET) John 12 **OFFICE OFFICE** (OFFICE #, SIZE) 12 Adam 12 500 EMP (NAME, SALARY) CHILD (CHILD NAME, AGE) CHILD **EMP** 10 10K Sue Fisher WORKS (DEPT #, NAME) 4 John 12K Peter **OFFSPRING (NAME, CHILD NAME)** 7 Dave Adam 14K **OCCUPIED** (NAME, OFFICE #) WORKS OFF SPRING Fish er Sue 17 Fisher 17 John Fisher Peter Adam 17 Jone Dave



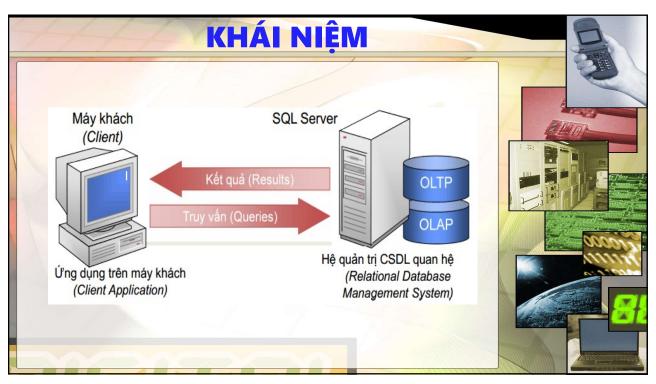


KHÁI NIỆM

- SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System RDBMS) sử dụng câu lệnh SQL để trao đổi dữ liệu giữa máy Client và máy cài SQL Server.
- SQL Server cung cấp ngôn ngữ lập trình database hướng thủ tục Transact-SQL được thiết kế để mở rộng khả năng của SQL



35



KHÁI NIỆM

- OLTP (online transaction processing): hệ thống xử lý giao dịch trực tuyến.
 - Đặc trưng bởi một số lượng lớn giao dịch (insert, update, delete) trong một thời gian ngắn.
 - Truy vấn OLTP đơn giản và ngắn hơn và do đó đòi hỏi ít thời gian hơn trong quá trình xử lý và cũng cần ít không gian hơn.
 - Được thiết kế nhằm mục đích thu thập dữ liệu, phải duy trì ràng buộc toàn vẹn dữ liệu.



37

KHÁI NIỆM

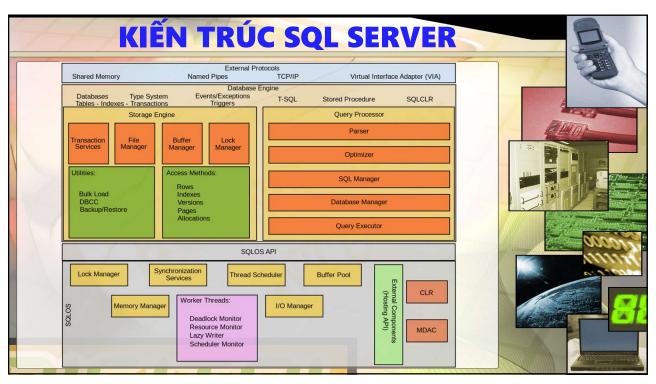
- OLAP (online analytical processing): hệ thống xử lý phân tích trực tuyến.
 - Tổ chức và thống kê khối lượng lớn dữ liệu được dùng để phân tích, báo cáo, hỗ trợ ra quyết định.
 - Cho phép người dùng thực hiện các truy vấn phức tạp để trích xuất dữ liệu đa chiều.
 - Giao dịch trong OLAP dài và do đó mất nhiều thời gian hơn để xử lý và yêu cầu không gian lớn.
 - Các giao dịch trong OLAP ít thường xuyên hơn so với OLTP.



PHIÊN BẢN

- Standard: phiên bản chuẩn, tính năng cơ bản đầy đủ, chạy tốt trên hệ thống lên tới 4 CPU và 2 GB RAM
- * Developer: dành cho nhà phát triển, chứa đầy đủ tính năng nhưng chỉ phân quyền cho một người duy nhất.
- Enterprise: phiên bản cao cấp, có đầy đủ các tính năng
- Web: dành cho các ứng dụng web
- Express: là bản cơ bản, sử dụng tối đa 1 CPU và 1GB RAM, dung lượng lưu trữ tối đa là 10GB.

39



KIẾN TRÚC SQL SERVER

- Protocols: có 4 giao thức
 - Shared memory: Dành cho các kết nối cục bộ và mục đích khắc phục sự cố.
 - Named pipes: Dành cho những kết nối trong mạng LAN.
 - TCP/IP: Dành cho các kết nối với mạng WAN.
 - VIA-Virtual Interface Adapter: Yêu cầu phần cứng đặc biệt được thiết lập bởi nhà cung cấp và cũng không được hỗ trợ từ bản SQL 2012.



41

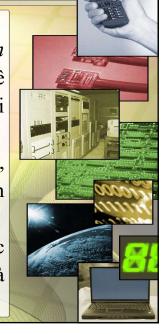
KIẾN TRÚC SQL SERVER

- Database Engine: thành phần chính của SQL Server quản lý các đối tượng vật lý và logic của CSDL.
- Query Processor (Relational Engine): chứa bộ phân tích Query, tối ưu hóa Query và bộ thực thi Query.
 - Query Parser (Command Parser) và Compiler (Translator): chịu trách nhiệm kiểm tra cú pháp của truy vấn và chuyển đổi truy vấn sang ngôn ngữ của máy.



KIẾN TRÚC SQL SERVER

- Query Processor (Relational Engine) tt
 - Query Optimizer: chuẩn bị output là Execution Plan bằng cách lấy input là truy vấn, các thống kê và cây Algebrizer (tiến trình xác minh các đối tượng CSDL).
 - Execution Plan: tương tự như bản đồ chỉ đường, chứa thứ tự các bước thực hiện như là một phần của việc thực hiện các truy vấn.
 - Query Executor: thực hiện truy vấn từng bước một dưới sự chỉ dẫn của Execution Plan và cũng là nơi Storage Engine sẽ được liên lạc.



43

KIẾN TRÚC SQL SERVER

- Storage Engine: Chịu trách nhiệm lưu trữ, truy xuất dữ liệu trong hệ thống lưu trữ, thao tác dữ liệu, khóa và quản lý các transaction.
- ❖ SQL OS: Cung cấp các dịch vụ hệ điều hành khác nhau, ví dụ như hoạt động quản lý bộ nhớ với buffer pool, log buffer, phát hiện deadlock (khóa chết) bằng cách sử dụng cấu trúc block và lock.

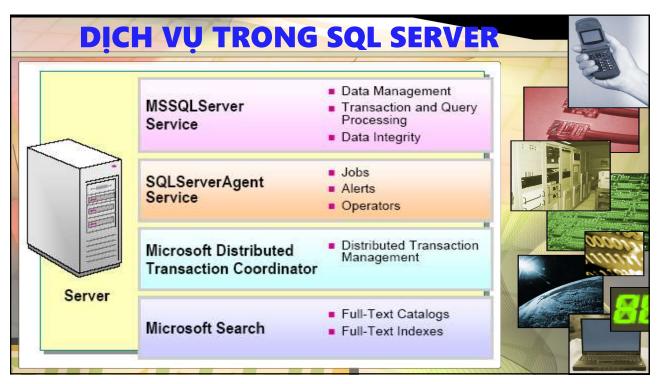


KIẾN TRÚC SQL SERVER

SQL OS

- Checkpoint: là một tiến trình nội bộ, ghi tất cả các trang đã sửa đổi (Dirty Page) từ Buffer Cache vào ổ đĩa vật lý. Ngoài ra, nó cũng ghi các bản log từ Log Buffer vào file vật lý.
- Lazy Writer: đẩy các Dirty Page và ổ cứng để giải phóng bộ nhớ trong Buffer Pool. Điều này xảy ra khi SQL Server đang bị thiếu bộ nhớ.

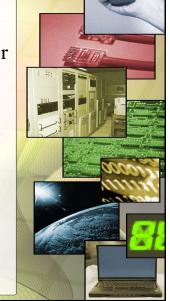
45



DICH VU TRONG SQL SERVER

* MSSQL Server Services

- Cấp phát tài nguyên máy tính cho nhiều user đồng thời.
- Quản lý dữ liệu.
- Xử lý truy vấn và giao dịch
- Bảo đảm toàn ven dữ liệu.



47

DICH VU TRONG SQL SERVER

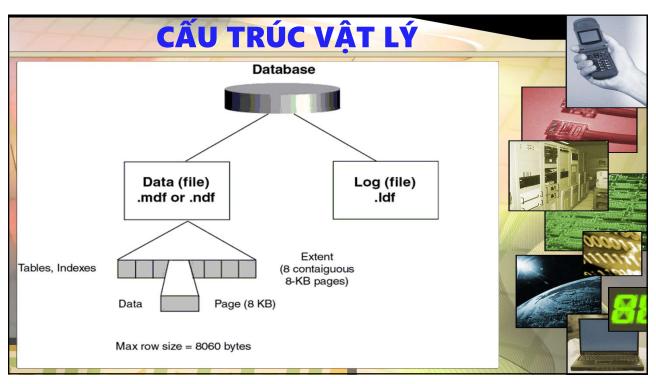
SQLServerAgent Services

- Cảnh báo về trạng thái của một process, ví dụ như một công việc nào đó được hoàn thành hoặc xảy ra lỗi.
- Tạo ra một công việc mới và lập lịch để tự động hoá các nhiệm vụ như Backup dữ liệu, Replication, ...





- Microsoft Distributed Transaction Coordinator (MS DTC)
 - Cho phép các clients làm việc với nhiều nguồn dữ liệu khác nhau trong một transaction.
- * Microsoft Search
 - Là một full-text engine cho phép đánh chỉ mục và truy vấn cho dữ liệu văn bản không cấu trúc



CẤU TRÚC VẬT LÝ

- Primary data file (.mdf): 1 file chính chứa data và những system tables.
- Secondary data file (.ndf): 1 hay nhiều file phụ chỉ sử dụng khi database được phân chia để chứa trên nhiều dĩa.
- Transaction log file (.ldf): 1 file ghi lại tất cả những thay đổi diễn ra trong một database và chứa đầy đủ thông tin để có thể rollback hay rollforward.

51

SYSTEM DATABASE SQL Server bao gồm CSDL hệ thống (system databases) và một hay nhiều user database. Các system databases cơ bản bao gồm: Databases System Databases master model msdb tempdb

SYSTEM DATABASE

- * master: là CSDL kiểm soát tất cả các hoạt động trên SQL Server, chứa thông tin về hệ thống: các tài khoản đăng nhập, cấu hình hệ thống, thông tin về các CSDL đã tạo, các thủ tục hệ thống thực hiện các tác vụ quản trị hệ thống, các thủ tục của người dùng tạo thêm...
- * model: CSDL mẫu để tạo ra các CSDL người dùng
- * msdb: được sử dụng cho SQL Server Agent để lập lịch các công việc và các cảnh báo.
- tempdb: lưu trữ các thông tin tạm thời của các hoạt động trên SQL



53

4//		
Đối tượng	Mô tả	
Table	Đối tượng lưu trữ dữ liệu của CSDL	
Data type	Kiểu dữ liệu	3 75:
Constraint	Các qui tắc để kiểm tra ràng buộc dữ liệu	
Default	Các giá trị mặc định nếu giá trị của một field không được nhập vào	way.
Rule	Các thông tin mà định nghĩa các giá trị hợp lệ trong một field	
Index	Là một cấu trúc lưu trữ nhằm truy xuất nhanh dữ liệu	

