

MỘT SỐ KHÁI NIỆM CƠ BẢN VÀ KIẾN TRÚC ORACLE SERVER

1. Một số khái niệm cơ bản

1.1. Tại sao phải có cơ sở dữ liệu



✓ Hệ thống các tệp tin cổ điển

○ Ưu điểm:

- Các tệp tin được quản lý riêng từng đơn vị quản lý ít tốn thời gian bởi khối lượng thông tin cần quản lý và khai thác nhỏ.
- Không đòi hỏi về vật chất và chất xám nhiều
- Triển khai ứng dụng nhanh

○ Nhược điểm:

- Thông tin được tổ chức riêng rẽ ở nhiều nơi nên cập nhật dễ làm mất tính nhất quán dữ liệu.
- Hệ thống thông tin được tổ chức thành các hệ thống file riêng lẻ nên thiếu sự chia sẻ thông tin giữa các nơi.

1.2. Cơ sở dữ liệu là gì ?

Cơ sở dữ liệu là tập hợp các dữ liệu có cấu trúc và liên quan với nhau được lưu trữ trên máy tính, được nhiều người sử dụng và được tổ chức theo một mô hình.

Ví dụ:

Quản lý thông tin sinh viên là một ví dụ về cơ sở dữ liệu

- Là các thông tin có ý nghĩa
- Là tập hợp các thông tin có cấu trúc.
- Các thông tin này có liên quan với nhau và có thể hệ thống được.



Được nhiều người sử dụng

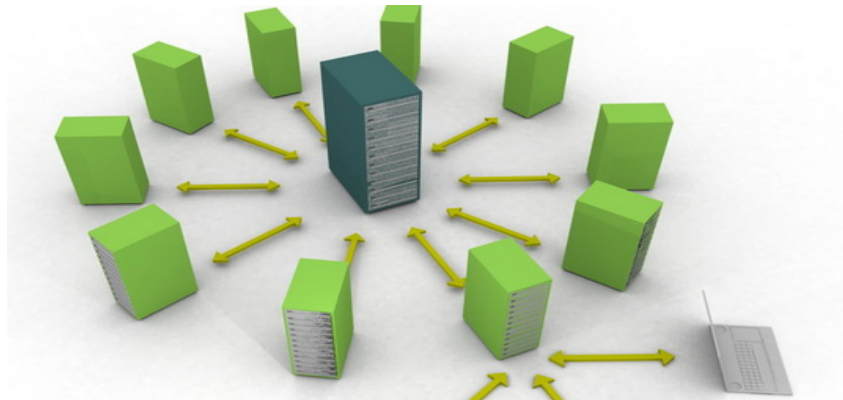
✓ Ưu điểm:

- Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó đảm bảo được tính nhất quán và toàn vẹn dữ liệu. Phần định nghĩa cấu trúc này gọi là meta-data.
- Đảm bảo sự độc lập giữa dữ liệu và chương trình ứng dụng.
- **Trừu tượng hoá dữ liệu (Data Abstraction):** Mô hình dữ liệu được sử dụng để làm ẩn lưu trữ vật lý chi tiết của dữ liệu.
- **Nhiều khung nhìn (multi-view) cho các đối người dùng khác nhau:** Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác nhau.

- **Đa người dùng (multi-user):** Khả năng chia sẻ thông tin cho nhiều người sử dụng và nhiều ứng dụng khác nhau.
- ✓ **Vấn đề cần giải quyết:**
 - Để đạt được các ưu điểm trên, cơ sở dữ liệu đặt ra những vấn đề cần giải quyết. Đó là:
 - **Tính chủ quyền của dữ liệu:** Do tính chia sẻ của CSDL nên chủ quyền của CSDL dễ bị xâm phạm.
 - **Tính bảo mật và quyền khai thác thông tin của người sử dụng:** Do có nhiều người được phép khai thác CSDL nên cần thiết phải có một cơ chế bảo mật và phân quyền hạn khai thác CSDL.
 - **Tranh chấp dữ liệu:** Nhiều người được phép cùng truy cập vào CSDL với những mục đích khác nhau: Xem, thêm, xóa hoặc sửa dữ liệu. Cần phải có cơ chế ưu tiên truy cập dữ liệu hoặc giải quyết tình trạng xung đột trong quá trình khai thác cạnh tranh. Cơ chế ưu tiên có thể được thực hiện bằng việc cấp quyền (hay mức độ) ưu tiên cho từng người khai thác.
 - **Đảm bảo dữ liệu khi có sự cố:** Việc quản lý dữ liệu tập trung có thể làm tăng nguy cơ mất mát hoặc sai lệch thông tin khi có sự cố mất điện đột xuất hoặc đĩa lưu trữ bị hỏng. Một số hệ điều hành mạng có cung cấp dịch vụ sao lưu ảnh đĩa cứng (cơ chế sử dụng đĩa cứng dự phòng - RAID), tự động kiểm tra và khắc phục lỗi khi có sự cố. Tuy nhiên, bên cạnh dịch vụ của hệ điều hành, để đảm bảo an toàn cho CSDL, nhất thiết phải có một cơ chế khôi phục dữ liệu khi có sự cố xảy ra.

1.3. Hệ quản trị cơ sở dữ liệu

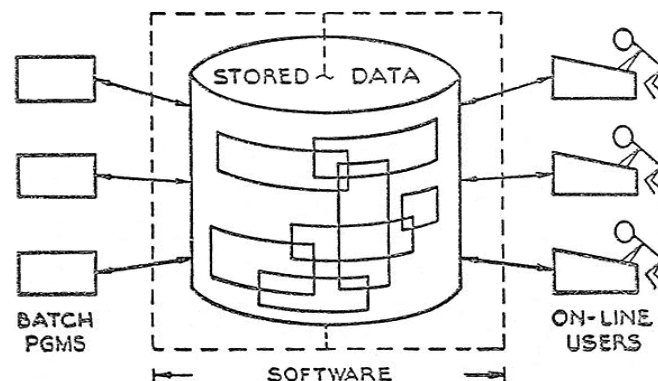
Hệ quản trị cơ sở dữ liệu là tập hợp các phần mềm cho phép định nghĩa các cấu trúc để lưu trữ thông tin trên máy, nhập dữ liệu, thao tác trên các dữ liệu đảm bảo sự an toàn và bí mật của dữ liệu.

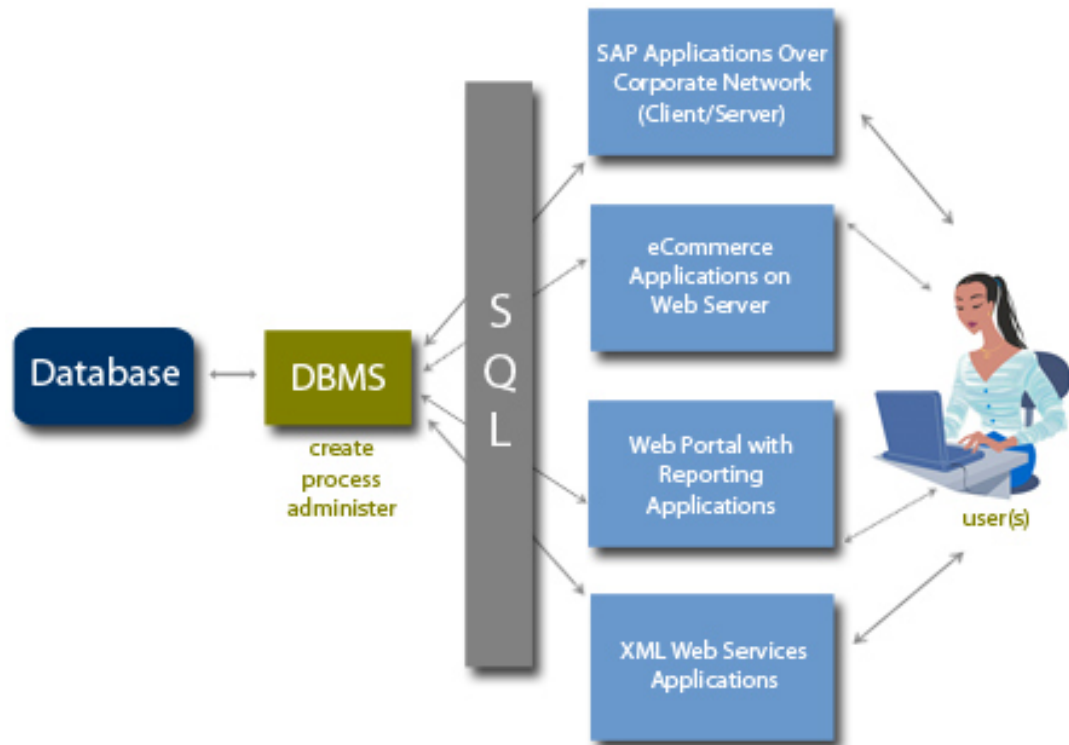


✓ Hệ quản trị cơ sở dữ liệu

- **Định nghĩa cấu trúc:** Là việc xác định kiểu dữ liệu, cấu trúc và những ràng buộc cho dữ liệu được lưu trữ trong CSDL.
- **Nhập dữ liệu:** Là việc lưu trữ dữ liệu vào các thiết bị lưu trữ trung gian được điều khiển bằng Hệ quản trị cơ sở dữ liệu.
- **Thao tác dữ liệu:** Thao tác trên CSDL bao gồm những chức năng như truy xuất cơ sở dữ liệu để tìm kiếm thông tin cần thiết, cập nhật cơ sở dữ liệu và tổng hợp những báo cáo từ dữ liệu.

A DATABASE SYSTEM





1.4. Mô hình dữ liệu (Data Model)

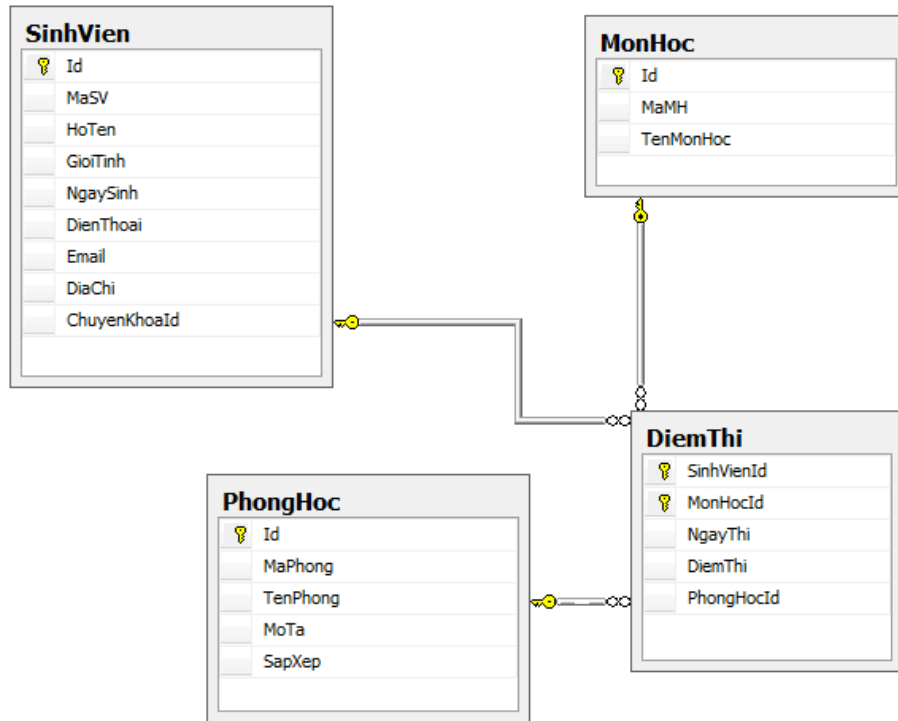
Là một tập những khái niệm dùng để biểu diễn cấu trúc của cơ sở dữ liệu - cung cấp những điều kiện cần thiết để đạt được mức độ trừu tượng dữ liệu.

Cấu trúc dữ liệu bao gồm:

- ✓ Kiểu dữ liệu (data types)
- ✓ Mối quan hệ giữa các dữ liệu (relationships)
- ✓ Những ràng buộc (constraints) mà cơ sở dữ liệu phải tuân theo.

1.5. Lược đồ cơ sở dữ liệu (Database Schema)

Là biểu diễn của cơ sở dữ liệu, bao gồm cấu trúc cơ sở dữ liệu và những ràng buộc trên dữ liệu.



1.6. Ngôn ngữ của HQTCSDL

HQTCSDL phục vụ có nhiều đối tượng người sử dụng khác nhau nên nó phải hỗ trợ ngôn ngữ để người sử dụng tương tác với nó.

- ✓ **Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL):** Là ngôn ngữ được các nhà quản trị cơ sở dữ liệu (DBA) và các nhà thiết kế cơ sở dữ liệu dùng để xây dựng lược đồ khái niệm của cơ sở dữ liệu.
- ✓ **Ngôn ngữ thực hiện dữ liệu (Data Manipulation Language -DML):** Là ngôn ngữ được sử dụng để thao tác với dữ liệu bao gồm việc truy cập đến bản ghi và cập nhật dữ liệu cho bản ghi (thêm, sửa, xoá).

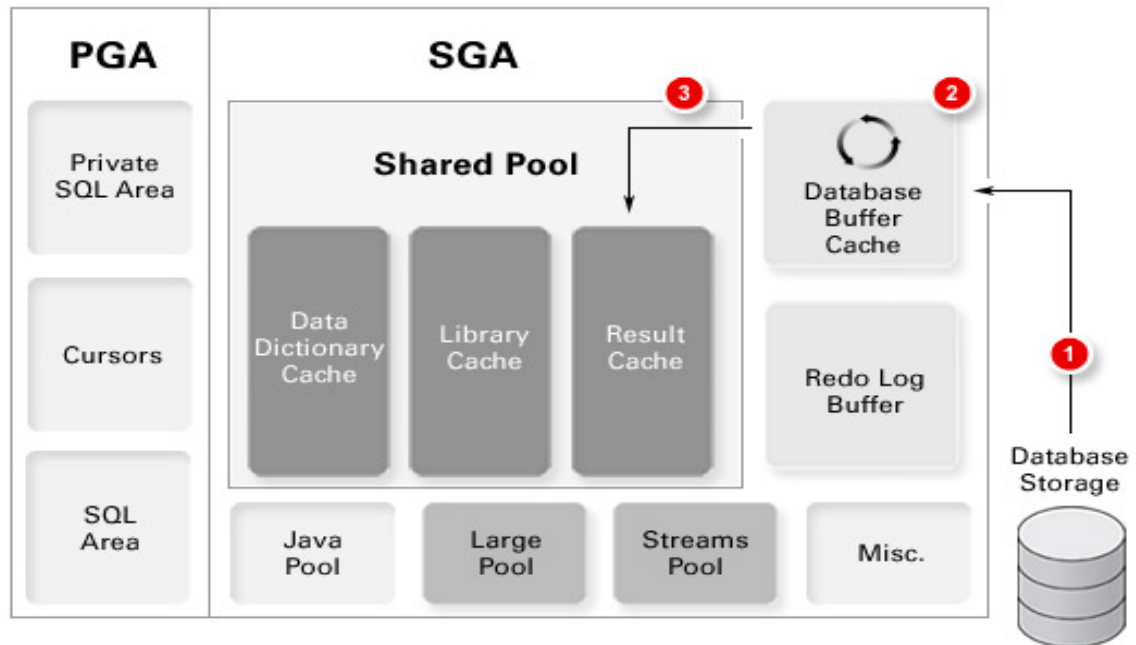
Các lệnh DML có thể được nhúng trong ngôn ngữ lập trình hoặc thực hiện độc lập (ngôn ngữ truy vấn).

2. Kiến trúc Oracle Server

Như chúng ta đã biết Oracle server là một hệ thống quản trị cơ sở dữ liệu đối tượng -quan hệ cho phép quản lý thông tin một cách toàn diện, nó thích hợp và là lựa chọn số 1 cho bài toán số lớn cần lưu trữ dữ liệu với số lượng lớn. Oracle server bao gồm hai thành phần chính là **Oracle instance** và **Oracle database**.

2.1. Oracle Instance

Oracle instance bao gồm một cấu trúc bộ nhớ **System Global Area (SGA)** và các **background processes** (tiến trình nền) được sử dụng để quản trị cơ sở dữ liệu. **Oracle instance** được xác định qua tham số môi trường **ORACLE_SID** của hệ điều hành.



System Global Area – SGA:

SGA là vùng bộ nhớ chia sẻ được sử dụng để lưu trữ dữ liệu và các thông tin điều khiển của Oracle server. SGA được cấp phát (allocated) trong bộ nhớ của máy tính mà Oracle server đang hoạt động trên đó. Các User kết nối tới Oracle sẽ chia sẻ các dữ liệu có trong SGA, việc mở rộng không gian bộ nhớ cho SGA sẽ làm nâng cao hiệu suất của hệ thống, lưu trữ được nhiều dữ liệu trong hệ thống hơn đồng thời giảm thiểu các thao tác truy xuất đĩa (disk I/O).

SGA bao gồm một vài cấu trúc bộ nhớ chính:

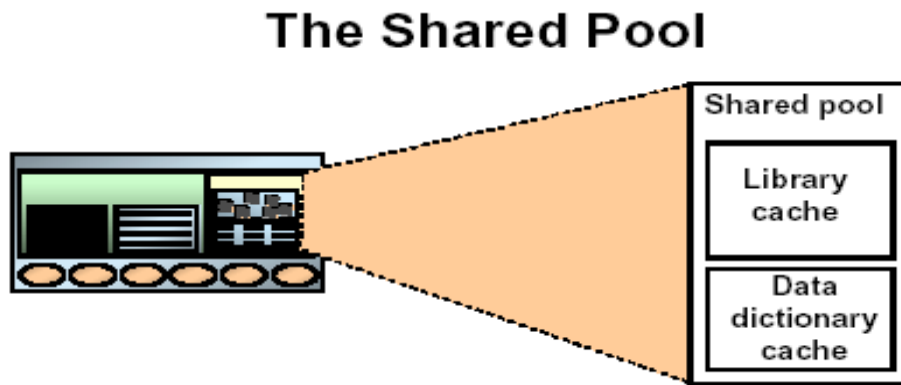
- ✓ **Shared pool:** Là một phần của SGA lưu các cấu trúc bộ nhớ chia sẻ.
- ✓ **Database buffer cache:** Lưu trữ các dữ liệu được sử dụng gần nhất.

- ✓ **Redo log buffer:** Được sử dụng cho việc dò tìm lại các thay đổi trong cơ sở dữ liệu và được thực hiện bởi các background process.

❖ Share Pool

Shared pool là một phần trong SGA và được sử dụng khi thực hiện phân tích câu lệnh (parse phase). Kích thước của Shared pool được xác định bởi tham số SHARED_POOL_SIZE có trong parameter file (file tham số).

Các thành phần của Shared pool gồm có: **Library cache** và **Data dictionary cache**.



Library Cache

Library cache lưu trữ thông tin về các câu lệnh SQL được sử dụng gần nhất bao gồm:

- ✓ *Nội dung của câu lệnh dạng text (văn bản).*
- ✓ *Parse tree (cây phân tích) được xây dựng tùy thuộc vào câu lệnh.*
- ✓ *Execution plan (sơ đồ thực hiện lệnh) gồm các bước thực hiện và tối ưu lệnh.*

Do các thông tin trên đã được lưu trữ trong Library cache nên khi thực hiện lại một câu lệnh truy vấn, trước khi thực hiện câu lệnh, Server process sẽ lấy lại các thông tin đã được phân tích mà không phải phân tích lại câu lệnh. Do vậy, Library cache có thể giúp nâng cao hiệu suất thực hiện lệnh.

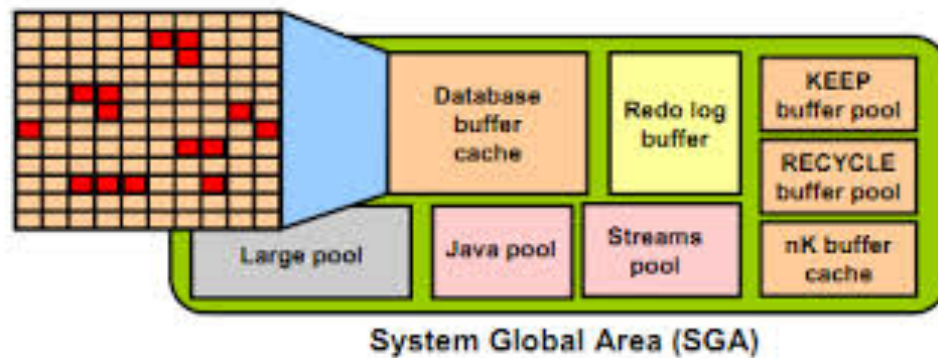
Data Dictionary Cache

Data dictionary cache là một thành phần của Shared pool lưu trữ thông tin của dictionary cache được sử dụng gần nhất như các định nghĩa các bảng, định nghĩa các cột, usernames, passwords, và các privileges (quyền).

Trong giai đoạn phân tích lệnh (parse phase), Server process sẽ tìm các thông tin trong dictionary cache để xác định các đối tượng trong câu lệnh SQL và để xác định các mức quyền tương ứng. Trong trường hợp cần thiết, Server process có thể khởi tạo và nạp các thông tin từ các file dữ liệu.

Data buffer cache

Khi thực hiện một truy vấn, **Server process** sẽ tìm các blocks cần thiết trong database buffer cache. Nếu không tìm thấy block trong database buffer cache, Server process mới đọc các block từ data file và tạo luôn một bản sao của block đó vào trong vùng nhớ đệm (buffer cache). Như vậy, với các lần truy xuất tới block đó sau này sẽ không cần thiết phải truy xuất vào datafile nữa.

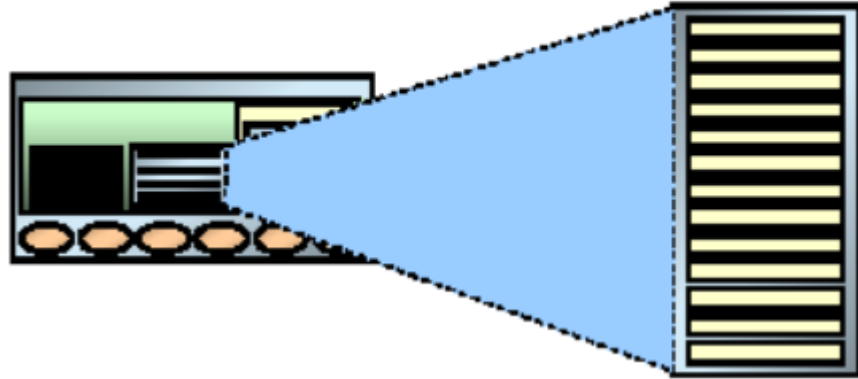


Database buffer cache là vùng nhớ trong SGA sử dụng để lưu trữ các block dữ liệu được sử dụng gần nhất.

Redo log buffer

Server process ghi lại các thay đổi của một instance vào redo log buffer, đây cũng là một phần bộ nhớ SGA.

Redo Log Buffer



Một số đặc điểm cần chú ý của Redo log buffer:

- ✓ Kích thước được xác định bởi tham số `LOG_BUFFER`.
- ✓ Lưu trữ các redo records (bản ghi hồi phục) mỗi khi có thay đổi dữ liệu.
- ✓ Redo log buffer được sử dụng một cách thường xuyên và các thay đổi bởi một transaction có thể nằm đan xen với các thay đổi của các transactions khác.
- ✓ Bộ đệm được tổ chức theo kiểu circular buffer (bộ đệm nối vòng) tức là dữ liệu thay đổi sẽ tiếp tục được nạp lên đầu sau khi vùng đệm đã được sử dụng hết.

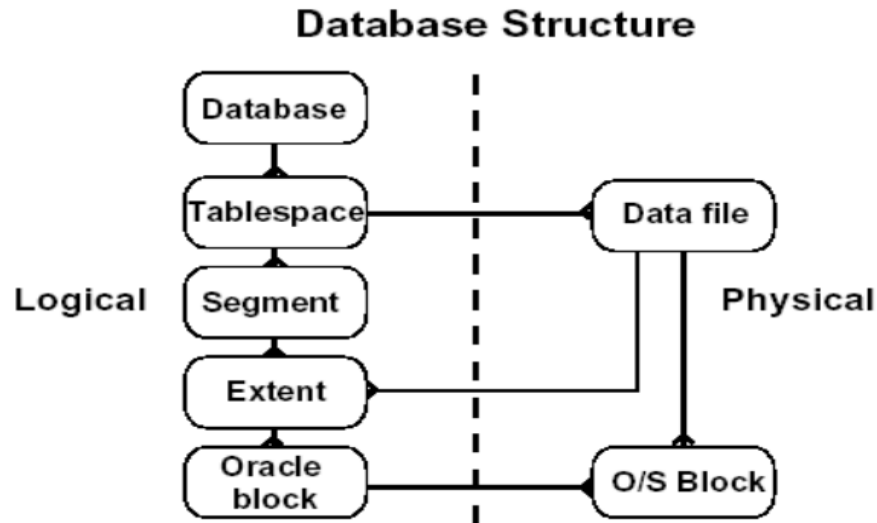
Background process

Background process (các tiến trình nền) thực hiện các chức năng thay cho lời gọi tiến trình xử lý tương ứng. Nó điều khiển vào ra, cung cấp các cơ chế xử lý song song nâng cao hiệu quả và độ tin cậy. Tùy theo từng cấu hình mà Oracle instance có các Background process như:

- ✓ **Database Writer (DBW0):** Ghi lại các thay đổi trong data buffer cache ra các file dữ liệu.
- ✓ **Log Writer (LGWR):** Ghi lại các thay đổi được đăng ký trong redo log buffer vào các redo log files.
- ✓ **System Monitor (SMON):** Kiểm tra sự nhất quán trong database.
- ✓ **Process Monitor (PMON):** Dọn dẹp lại tài nguyên khi các tiến trình của Oracle gặp lỗi.
- ✓ **Checkpoint Process (CKPT):** Cập nhật lại trạng thái của thông tin trong file điều khiển và file dữ liệu mỗi khi có thay đổi trong buffer cache.

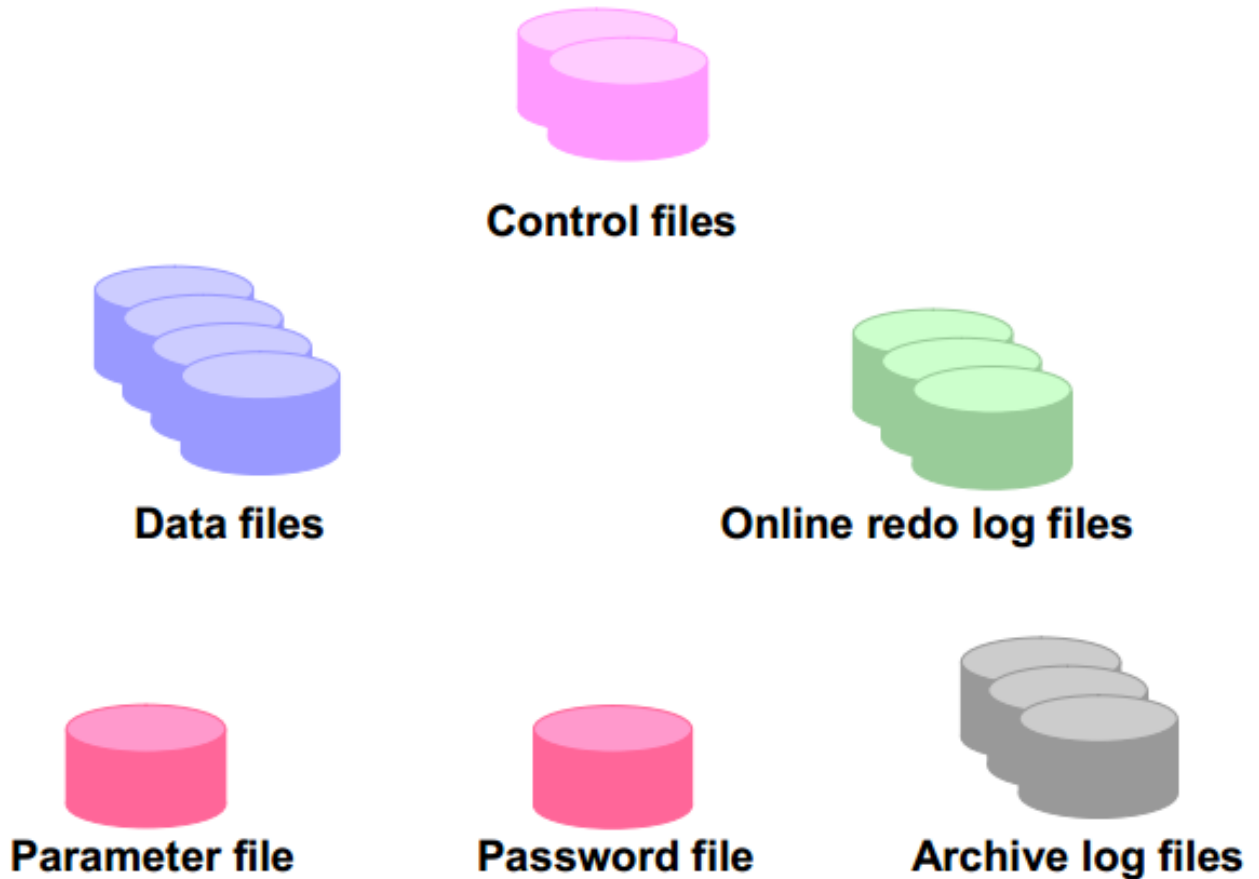
2.2. Oracle Database

Oracle database là tập hợp các dữ liệu được xem như một đơn vị thành phần (Unit). Database có nhiệm vụ lưu trữ và trả về các thông tin liên quan. Database được xem xét dưới hai góc độ **cấu trúc logic** và **cấu trúc vật lý**.



Tuy vậy, hai cấu trúc dữ liệu này vẫn tồn tại tách biệt nhau, việc quản lý dữ liệu theo cấu trúc lưu trữ vật lý không gây ảnh hưởng tới cấu trúc logic trong Oracle database được xác định bởi tên một tên duy nhất và được quy định trong tham số DB_NAME của parameter file.

Trong Oracle Database sẽ bao gồm các thành phần như sau:



a. Cấu trúc vật lý:

Cấu trúc vật lý bao gồm tập hợp các control file, online redo log file và các datafile:

❖ Datafiles

Mỗi một Oracle database đều có thể có một hay nhiều datafiles. Các database datafiles chứa toàn bộ dữ liệu trong database. **Các dữ liệu thuộc cấu trúc logic của database như tables hay indexes đều được lưu trữ dưới dạng vật lý trong các datafiles của database.**

Một số tính chất của datafiles:

- ✓ *Mỗi datafile chỉ có thể được sử dụng trong một database.*
- ✓ *Bên cạnh đó, datafiles cũng còn có một số tính chất cho phép tự động mở rộng kích thước mỗi khi database hết chỗ lưu trữ dữ liệu.*

- ✓ Một hay nhiều datafiles tạo nên một đơn vị lưu trữ logic của database gọi là *tablespace*.
- ✓ Một datafile chỉ thuộc về một *tablespace*.

Dữ liệu trong một datafile có thể đọc ra và lưu vào vùng nhớ bộ đệm của Oracle. Ví dụ: khi một user muốn truy cập dữ liệu trong một table thuộc database. Trong trường hợp thông tin yêu cầu không có trong cache memory hiện thời, nó sẽ được đọc trực tiếp từ các datafiles ra và lưu trữ vào trong bộ nhớ. Tuy nhiên, việc bổ sung hay thêm mới dữ liệu vào database không nhất thiết phải ghi ngay vào các datafile. Các dữ liệu có thể tạm thời ghi vào bộ nhớ để giảm thiểu việc truy xuất tới bộ nhớ ngoài (ổ đĩa) làm tăng hiệu năng sử dụng hệ thống. Công việc ghi dữ liệu này được thực hiện bởi DBWn background process.

❖ Redo Log Files

Mỗi Oracle database đều có một tập hợp từ 02 redo log files trở lên. Các redo log files trong database thường được gọi là database's redo log. Một redo log được tạo thành từ nhiều redo entries (gọi là các redo records).

Chức năng chính của redo log là ghi lại tất cả các thay đổi đối với dữ liệu trong database.

Redo log files được sử dụng để bảo vệ database khỏi những hỏng hóc do sự cố. Oracle cho phép sử dụng cùng một lúc nhiều redo log gọi là multiplexed redo log để cùng lưu trữ các bản sao của redo log trên các ổ đĩa khác nhau.

Các thông tin trong redo log file chỉ được sử dụng để khôi phục lại database trong trường hợp hệ thống gặp sự cố và không cho phép viết trực tiếp dữ liệu trong database lên các datafiles trong database. Ví dụ: khi có sự cố xảy ra như mất điện bất chợt chẳng hạn, các dữ liệu trong bộ nhớ không thể ghi trực tiếp lên các datafiles và gây ra hiện tượng mất dữ liệu.

Tuy nhiên, tất cả các dữ liệu bị mất này đều có thể khôi phục lại ngay khi database được mở trở lại. Việc này có thể thực hiện được thông qua việc sử dụng ngay chính các thông tin mới nhất có trong các redo log files thuộc datafiles. Oracle sẽ khôi phục lại database cho đến thời điểm trước khi xảy ra sự cố.

Công việc khôi phục dữ liệu từ các redo log được gọi là rolling forward.

❖ Control Files:

Mỗi Oracle database đều có ít nhất một control file. Control file chứa các mục thông tin quy định cấu trúc vật lý của database như:

- *Tên của database.*
- *Tên và nơi lưu trữ các datafiles hay redo log files.*
- *Time stamp (mốc thời gian) tạo lập database, ...*

Mỗi khi nào một instance của Oracle database được mở, control file của nó sẽ được sử dụng để xác định data files và các redo log files đi kèm. Khi các thành phần vật lý của database bị thay đổi (ví dụ như, tạo mới datafile hay redo log file), Control file sẽ được tự động thay đổi tương ứng bởi Oracle.

Control file cũng được sử dụng đến khi thực hiện khôi phục lại dữ liệu.

b. Cấu trúc logic:

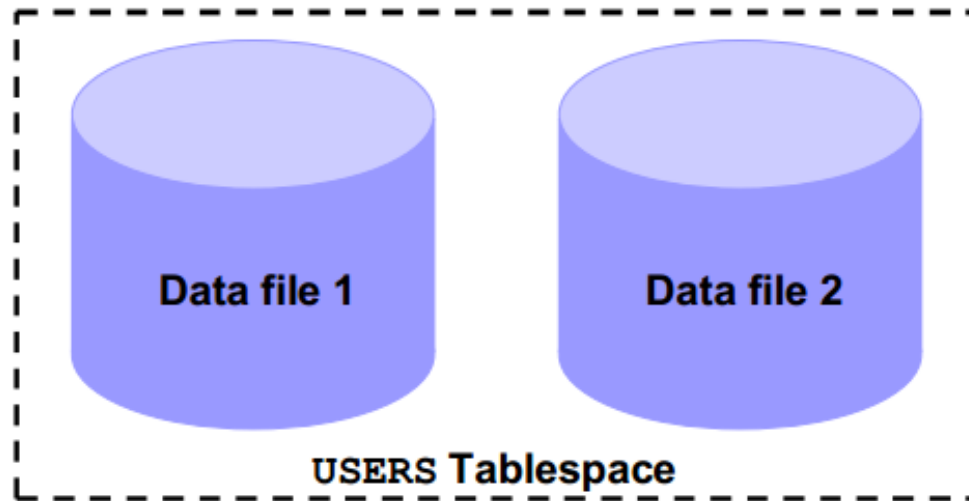
Cấu trúc logic của Oracle database bao gồm các đối tượng tablespaces, schema objects, data blocks, extents và segments.

❖ Tablespaces

Một database có thể được phân chia về mặt logic thành các đơn vị gọi là các tablespaces, Tablespaces thường bao gồm một nhóm các thành phần có quan hệ logic với nhau.

Databases, Tablespaces, và Datafiles

Mối quan hệ giữa các databases, tablespaces, và datafiles có thể được minh họa bởi hình vẽ sau:

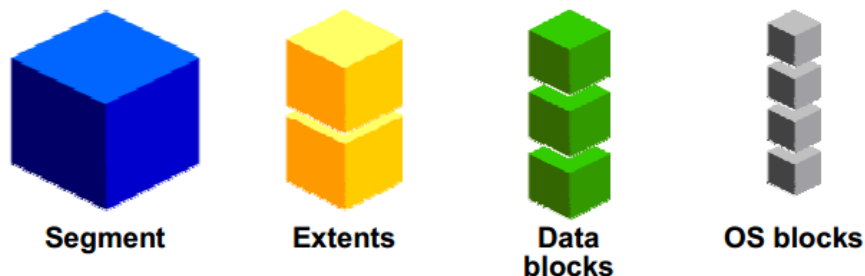


✓ **Một số điểm cần chú ý:**

- *Mỗi database có thể phân chia về mặt logic thành một hay nhiều tablespaces.*
- *Mỗi tablespace có thể được tạo nên, về mặt vật lý, bởi một hoặc nhiều datafiles.*
- *Kích thước của một tablespace bằng tổng kích thước của các datafiles của nó. Ví dụ: trong hình vẽ ở trên USERS tablespace có kích thước là tổng data file 1 và datafile 2 cộng lại.*
- *Kích thước của database cũng có thể xác định được bằng tổng kích thước của các tablespaces của nó.*

❖ **Data Blocks, Extents, and Segments**

Oracle điều khiển không gian lưu trữ trên đĩa cứng theo các cấu trúc logic bao gồm các data blocks, extents, và segments.



Oracle Data Blocks

Là mức phân cấp logic thấp nhất, các dữ liệu của Oracle database được lưu trữ trong các data blocks. Một data block tương ứng với một số lượng nhất định các bytes vật lý của database trong không gian đĩa cứng. Kích thước của một data block được chỉ ra cho mỗi Oracle database ngay khi database được tạo lập. Database sử dụng, cấp phát và giải phóng vùng không gian lưu trữ thông qua các Oracle data blocks.

Extents

Là mức phân chia cao hơn về mặt logic các vùng không gian trong database. Một extent bao gồm một số data blocks liên tiếp nhau, cùng được lưu trữ tại một thiết bị lưu giữ. Extent được sử dụng để lưu trữ các thông tin có cùng kiểu.

Segments

Là mức phân chia cao hơn nữa về mặt logic các vùng không gian trong database. Một segment là một tập hợp các extents được cấp phát cho một cấu trúc logic. Segment có thể được phân chia theo nhiều loại khác nhau:

Kiểu	Mô tả
Data segment	Mỗi một non-clustered table có một data segment. Các dữ liệu trong một table được lưu trữ trong các extents thuộc data segment đó. Với một partitioned table thì mỗi partition lại tương ứng với một data segment. Mỗi Cluster tương ứng với một data segment. Dữ liệu của tất cả các table trong cluster đó đều được lưu trữ trong data segment thuộc Cluster đó.
Index segment	Mỗi một index đều có một index segment lưu trữ các dữ liệu của nó. Trong partitioned index thì mỗi partition cũng lại tương ứng với một index segment.
Rollback segment	Một hoặc nhiều rollback segments của database được tạo lập bởi người quản trị database để lưu trữ các dữ liệu trung gian phục vụ cho việc khôi phục dữ liệu. Các thông tin trong Rollback segment được sử dụng để: <ul style="list-style-type: none">- Tạo sự đồng nhất các thông tin đọc được từ database- Sử dụng trong quá trình khôi phục dữ liệu- Phục hồi lại các giao dịch chưa commit đối với mỗi user

Temporary segment	Temporary segments được tự động tạo bởi Oracle mỗi khi một câu lệnh SQL statement cần đến một vùng nhớ trung gian để thực hiện các công việc của mình như sắp xếp dữ liệu. Khi kết thúc câu lệnh đó, các extent thuộc temporary segment sẽ lại được hoàn trả cho hệ thống.
--------------------------	--

❖ Parameter file:

Parameter file chỉ ra các tham số được sử dụng trong database. Người quản trị database có thể sửa đổi một vài thông tin có trong file này. Các tham số trong parameter file được viết ở dạng văn bản.

- ❖ **Password file:** Xác định quyền của từng user trong database. Cho phép người sử dụng khởi động và tắt một Oracle instance.
- ❖ **Archived redo log files:** Là bản offline của các redo log files chứa các thông tin cần thiết để phục hồi dữ liệu.

❖ Schema và Schema Objects

Schema là tập hợp các đối tượng (objects) có trong database. Schema objects là các cấu trúc logic cho phép tham chiếu trực tiếp tới dữ liệu trong database. Schema objects bao gồm các cấu trúc như tables, views, sequences, stored procedures, synonyms, indexes, clusters và database links.