



Sao lưu & phục hồi (Backup & Restore)

Bởi:

Ths. Phạm Hoàng Nhung

Giới thiệu

Những nguyên nhân gây ra mất dữ liệu:

- Đĩa cứng hư
- Vô ý hay cố ý sửa đổi dữ liệu như xóa hay thay đổi dữ liệu.
- Trộm cắp
- Virus

Để tránh việc mất dữ liệu, chúng ta nên thường xuyên sao lưu cơ sở dữ liệu. Nếu như dữ liệu hay cơ sở dữ liệu bị hư thì ta có thể dùng bản sao lưu (backup) này để khôi phục lại cơ sở dữ liệu bị mất.

Sao lưu cơ sở dữ liệu

Sao lưu-backup một cơ sở dữ liệu (CSDL) là tạo một bản sao CSDL, ta có thể dùng bản sao để khôi phục lại CSDL nếu CSDL bị mất. Bản sao gồm tất cả những file có trong CSDL kể cả transaction log.

Transaction log (hay log file) chứa những dữ liệu thay đổi trong CSDL (Ví dụ như khi ta thực hiện các lệnh INSERT, UPDATE, DELETE). Transaction log được sử dụng trong suốt quá trình khôi phục để roll forward những transaction hoàn thành và roll back những transaction chưa hoàn thành.

Roll back là hủy bỏ giao dịch chưa hoàn thành khi hệ thống xảy ra sự cố,... (hoặc trong trường hợp sao lưu, khi đã thực hiện xong việc sao lưu mà giao dịch chưa hoàn thành) (xem chi tiết ở phần Transaction).

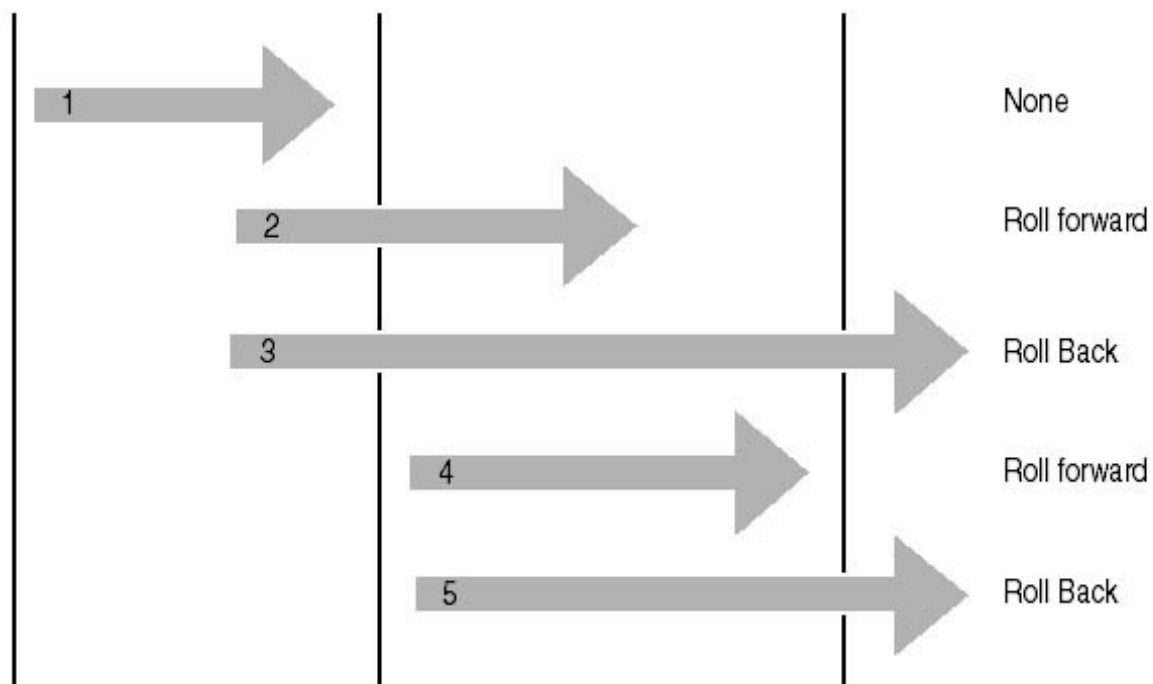
Roll forward là khôi phục tất cả giao dịch đã hoàn thành khi hệ thống xảy ra sự cố,... (hoặc trong trường hợp sao lưu, những giao dịch đã hoàn thành khi đã thực hiện xong việc sao lưu) (xem chi tiết ở phần Transaction).

Checkpoint là thời điểm ghi lại tất cả những trang dữ liệu thay đổi lên đĩa.

Ví dụ minh họa roll back và roll forward:

- Giao dịch 1 commit trước khi checkpoint, không làm gì cả vì dữ liệu đã được thay đổi trong CSDL (ứng với số 1 trong hình).
- Giao dịch 2 và 4 commit sau khi checkpoint nhưng trước khi hệ thống xảy ra sự cố, do đó những giao dịch này được tạo lại từ log file. Điều này gọi là roll forward (ứng với số 2 và 4 trong hình).
- Giao dịch 3 và 5 chưa commit khi hệ thống xảy ra sự cố, do đó những giao dịch này không được thực hiện và trả về CSDL khi chưa xảy ra giao dịch. Điều này gọi là roll back (ứng với số 3 và 5 trong hình).

Hình minh họa quá trình khôi phục giao dịch:



Khôi phục giao dịch

Checkpoint Hệ thống xảy ra sự cố

Sao lưu một transaction log là chỉ sao lưu những thay đổi xảy ra trong transaction log kể từ lần sao lưu transaction log cuối cùng.

Sao lưu & phục hồi (Backup & Restore)

Sao lưu một CSDL ghi lại toàn bộ trạng thái của dữ liệu tại thời điểm thực hiện xong sao lưu.

Trong thời gian sao lưu, SQL Server 2000 cho phép thực hiện việc giao dịch (transaction).

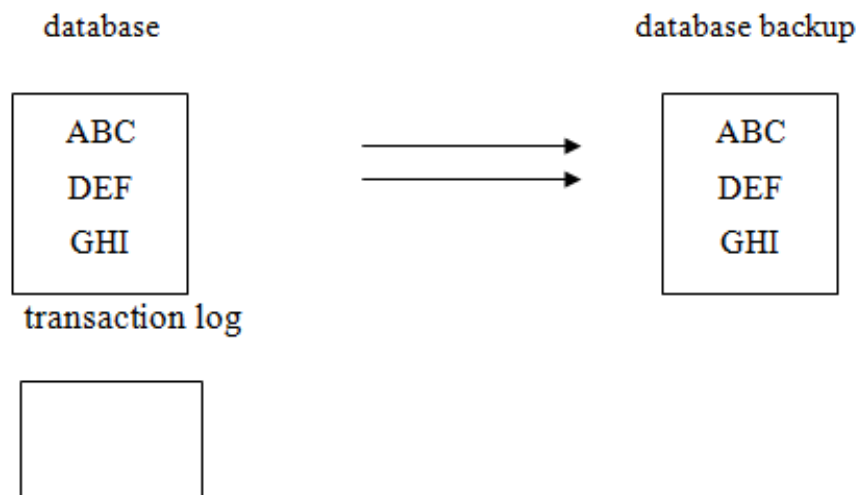
Phục hồi cơ sở dữ liệu

Việc khôi phục một bản sao lưu CSDL sẽ trả về CSDL cùng trạng thái của CSDL khi ta thực hiện việc sao lưu. Giao dịch (transaction) nào không hoàn thành trong khi sao lưu (backup) CSDL được roll back để đảm bảo tính nhất quán CSDL.

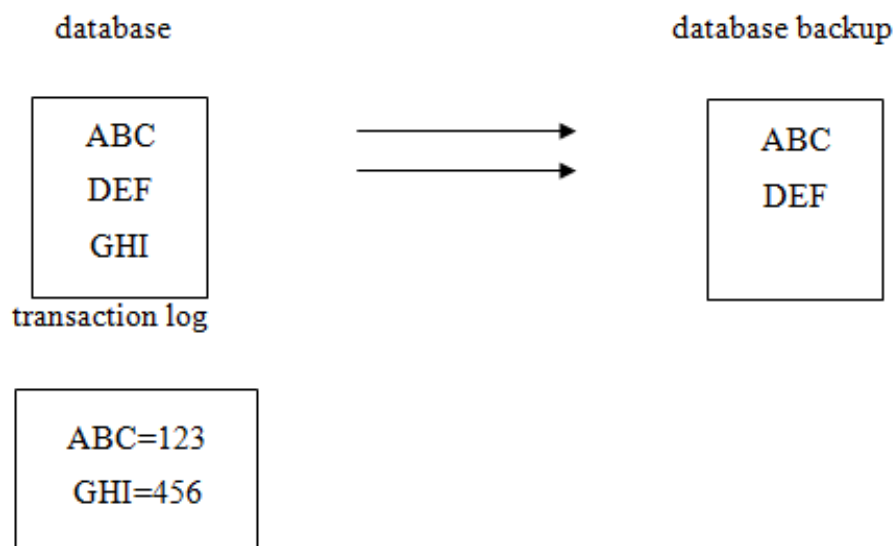
Khôi phục một bản sao lưu transaction log là áp dụng lại tất cả giao dịch (transaction) hoàn thành trong transaction log đối với CSDL. Khi áp dụng bản sao lưu transaction log, SQL Server đọc trước transaction log, roll forward tất cả các transaction. Khi đến cuối bản sao lưu transaction log, SQL Server roll back tất cả transaction mà không hoàn thành khi ta bắt đầu thực hiện sao lưu, tạo lại trạng thái chính xác của CSDL tại thời điểm bắt đầu thực hiện sao lưu.

Ví dụ minh họa sao lưu (backup) và khôi phục (restore) một CSDL có xảy ra giao dịch (transaction) khi thực hiện sao lưu:

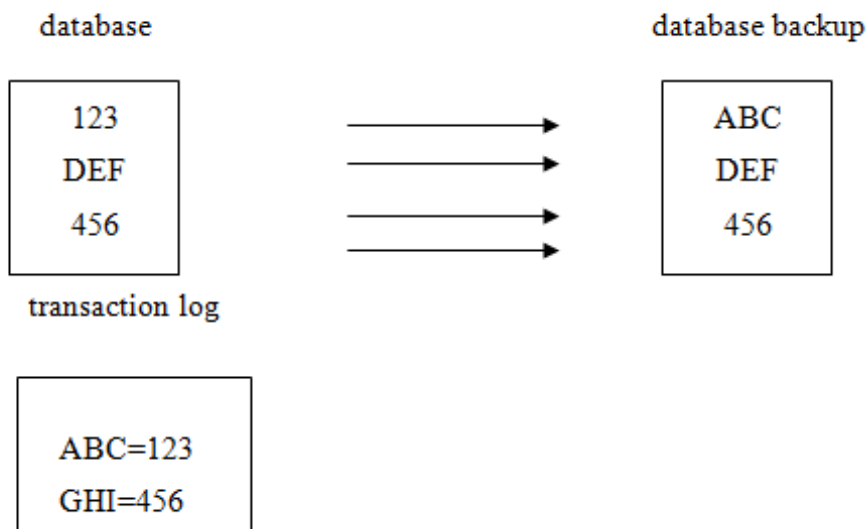
1. Bắt đầu backup: Giả sử CSDL gồm có các dữ liệu ABC, DEF, GHI, JKL, transaction log file không có dữ liệu vì không có giao dịch nào xảy ra. Khi đang thực hiện sao lưu (backup) được một phần dữ liệu thì xảy ra giao dịch, SQL Server 2000 sẽ ưu tiên cho việc giao dịch trước, việc sao lưu (backup) tạm thời dừng lại.



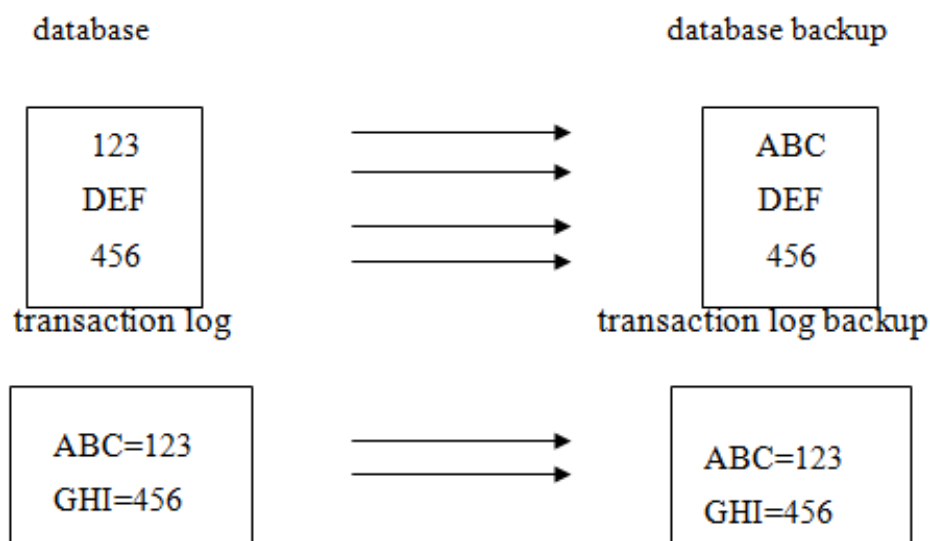
2. Xảy ra giao dịch (transaction), dữ liệu ABC được thay bằng 123, GHI được thay bằng 456.



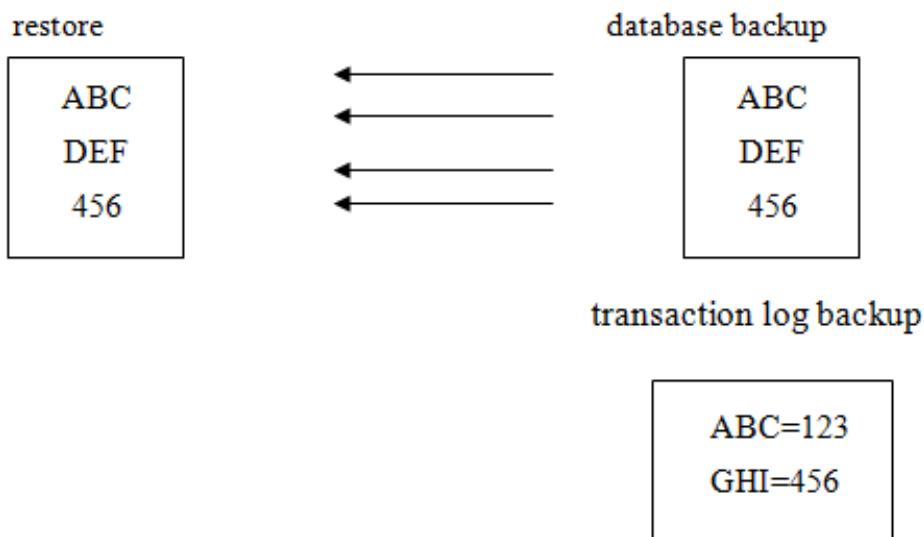
3. Khi thực hiện giao dịch (transaction) xong, SQL Server thực hiện tiếp việc sao lưu (backup), sẽ chép phần còn lại của dữ liệu nhưng dữ liệu đã thay đổi do xảy ra giao dịch.



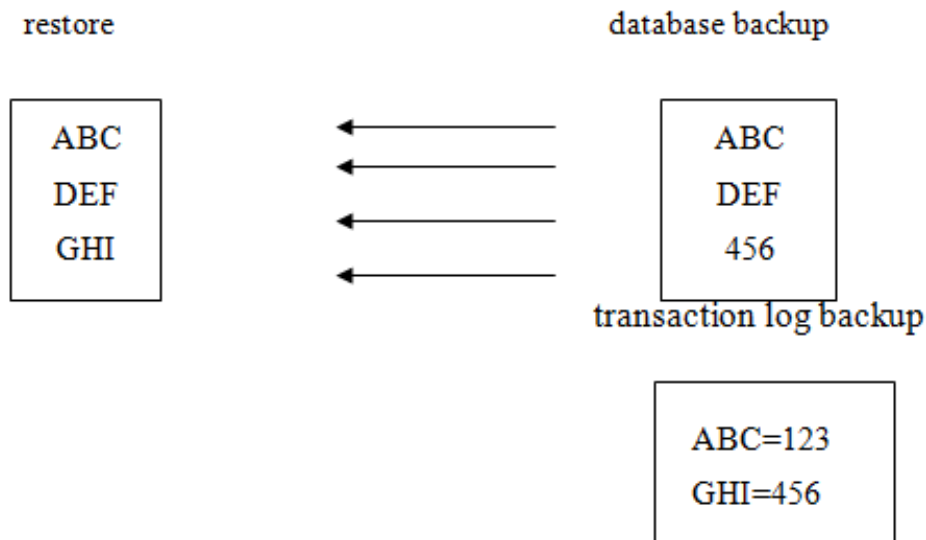
4. Khi sao lưu xong phần dữ liệu thì sẽ chép tiếp phần transaction log.



5. Khi có yêu cầu khôi phục (restore) CSDL , CSDL được khôi phục trước, chép lại toàn bộ CSDL của bản sao lưu CSDL đó .



6. Sau đó SQL Server sẽ khôi phục tiếp phần transaction log. Trước tiên sẽ roll forward nhưng khi đọc đến dữ liệu thứ ba thì nó thấy dữ liệu này đã được thay đổi rồi do đó nó sẽ roll back (trả về dữ liệu ban đầu khi chưa thực hiện giao dịch) để nhất quán dữ liệu.



Các loại Backup và Restore

Các loại sao lưu-Backups

Full Database Backups: Copy tất cả data files, user data và database objects như system tables, indexes, user-defined tables trong một database.

Differential Database Backups: Copy những thay đổi trong tất cả data files kể từ lần full backup gần nhất.

File or File Group Backups : Copy một data file đơn hay một file group.

Transaction Log Backups: Ghi nhận một cách thứ tự tất cả các transactions chứa trong transaction log file kể từ lần transaction log backup gần nhất. Loại backup này cho phép ta phục hồi dữ liệu trở ngược lại vào một thời điểm nào đó trong quá khứ mà vẫn đảm bảo tính nhất quán.

Các mô hình khôi phục- Recovery Models

Full Recovery Model: Đây là model cho phép phục hồi dữ liệu với ít rủi ro nhất. Nếu một database ở trong mode này thì tất cả các hoạt động không chỉ insert, update, delete mà kể cả insert bằng **Bulk Insert**, hay **Bcp** đều được log vào transaction log file. Khi có sự cố thì ta có thể phục hồi lại dữ liệu ngược trở lại tới một thời điểm trong quá khứ. Khi data file bị hư nếu ta có thể backup được transaction log file thì ta có thể phục hồi database đến thời điểm transaction gần nhất được committed.

Bulk-Logged Recovery Model : Ở mode này các hoạt động mang tính hàng loạt như Bulk Insert, bcp, Create Index, WriteText, UpdateText chỉ được log minimum vào

transaction log file đủ để cho biết là các hoạt động này có diễn ra mà không log toàn bộ chi tiết như trong Full Recovery Mode. Các hoạt động khác như Insert, Update, Delete vẫn được log đầy đủ để dùng cho việc phục hồi sau này.

Simple Recovery Model : Ở mode này thì Transaction Log File được truncate thường xuyên và không cần backup. Với mode này bạn chỉ có thể phục hồi tới thời điểm backup gần nhất mà không thể phục hồi tới một thời điểm trong quá khứ.

Muốn biết database của bạn đang ở mode nào bạn có thể **Right-click lên một database nào đó** trong SQL Server Enterprise Manager chọn **Properties->Options->Recovery**.

Bảng thông kê các mô hình phục hồi(Recovery Models) và kiểu sao lưu(Backup Type) tương ứng:

Model				
Backup Type				
Database	Database differential	Transactionlog	File or file differential	
Simple	Required	Optional	Not allowed	Not allowed
Full	Required(or file backups)	Optional	Required	Optional
Bulk-Logged	Required(or file backups)	Optional	Required	Optional

Full Database backup

Chép lại toàn bộ CSDL, lược đồ của tất cả các bảng và cấu trúc file tương ứng. Không phải tất cả những trang được chép đến backup set, chỉ những trang chứa dữ liệu thật sự mới được chép đến backup set vì thế database backup thường nhỏ hơn CSDL mà nó sao lưu. Cả những trang dữ liệu và những trang transaction log được chép đến backup set.

Khi tạo full database backup nhưng không có những bản transaction log backup riêng thì khi có yêu cầu khôi phục , ta chỉ có thể khôi phục lại được trạng thái của CSDL khi quá trình sao lưu hoàn thành. Không có cách gì khôi phục lại được CSDL mới nhất.

Ta nên dùng Full database backup nếu hệ thống có những đặc điểm sau:

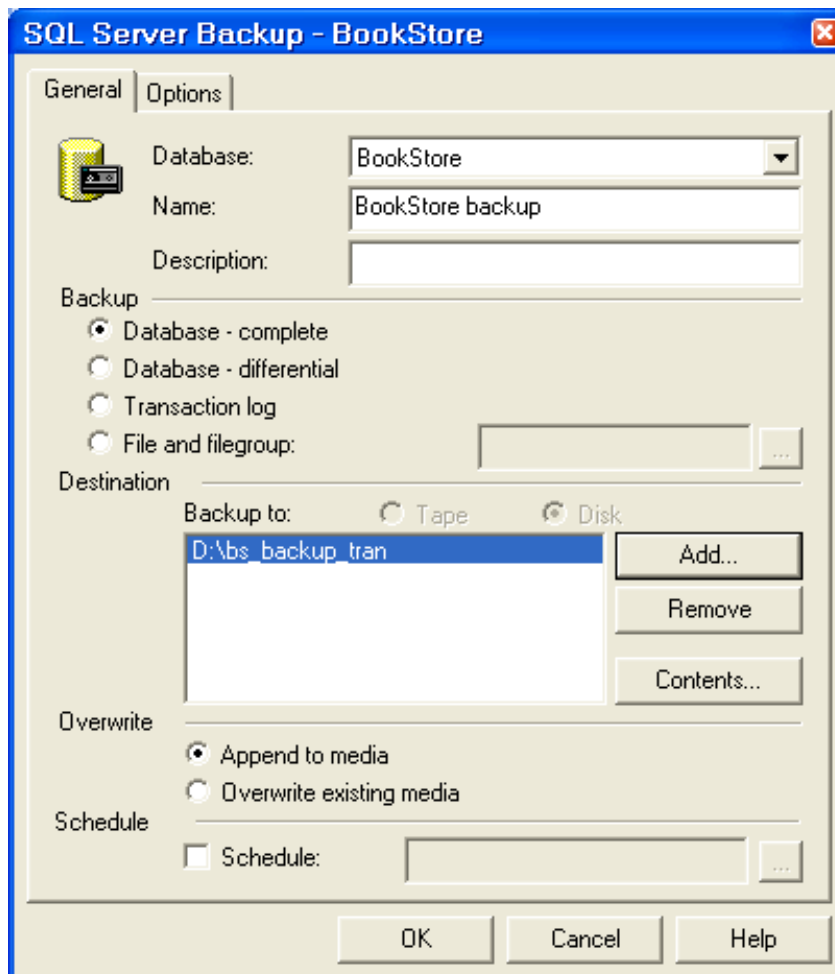
Sao lưu & phục hồi (Backup & Restore)

- Dữ liệu ít quan trọng và những thay đổi của CSDL có thể tạo lại bằng tay tốt hơn là dùng transaction log.
- CSDL ít thay đổi, như CSDL chỉ đọc.
- Sao lưu 1 CSDL là sao lưu toàn bộ CSDL mà không để ý đến nó có thay đổi so với lần sao lưu cuối cùng không. Điều này có nghĩa là sẽ mất nhiều vùng nhớ cho 1 bản sao và tốn nhiều thời gian để thực hiện sao lưu so với việc dùng transaction log backup và differential backup.

Cách tạo Full database backup bằng EM

Thực hiện các bước sau:

1. Kích vào server group, và kích vào server chứa Database muốn backup.
2. Kích **Databases**, kích phải chuột vào database, trở chuột vào **All Tasks**, sau đó kích **Backup Database**.

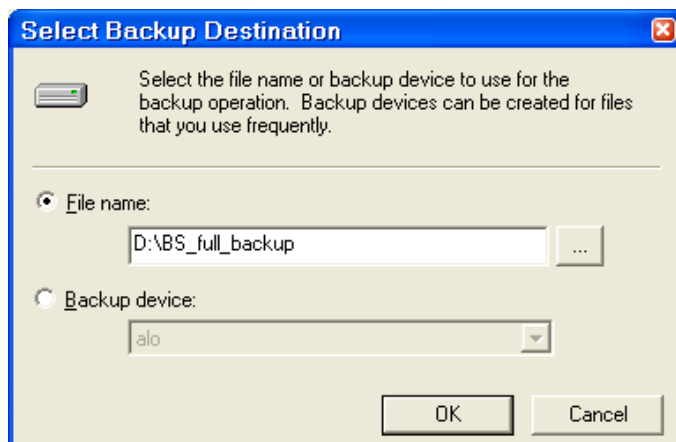


3. Trong **Namebox**, nhập tên của backup set. Trong **Description**, có thể soạn chú thích cho backup set này.

4. Dưới mục Backup, kích **Database - complete**.

5. Dưới mục **Destination**, kích **Tape** hoặc **Disk** (tùy thuộc bạn muốn backup vào loại thiết bị nào), sau đó chỉ ra đường dẫn chứa tệp tin backup.

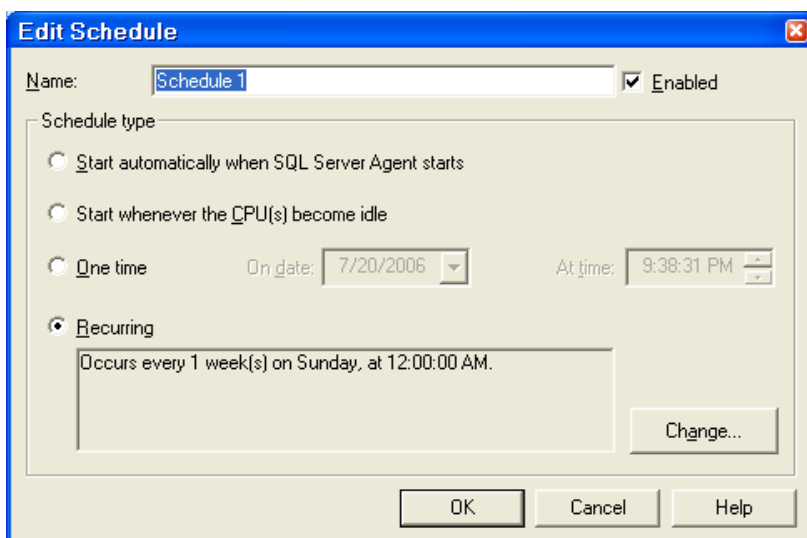
Nếu không có xuất hiện nơi để chọn đường dẫn _backup destinations, kích Add để thêm đường dẫn mới.



6. Dưới Overwrite, thực hiện như sau:

- Kích **Append to media** để thêm một tệp backup mới
- Kích **Overwrite existing media** để ghi đè lên tệp đang tồn tại.

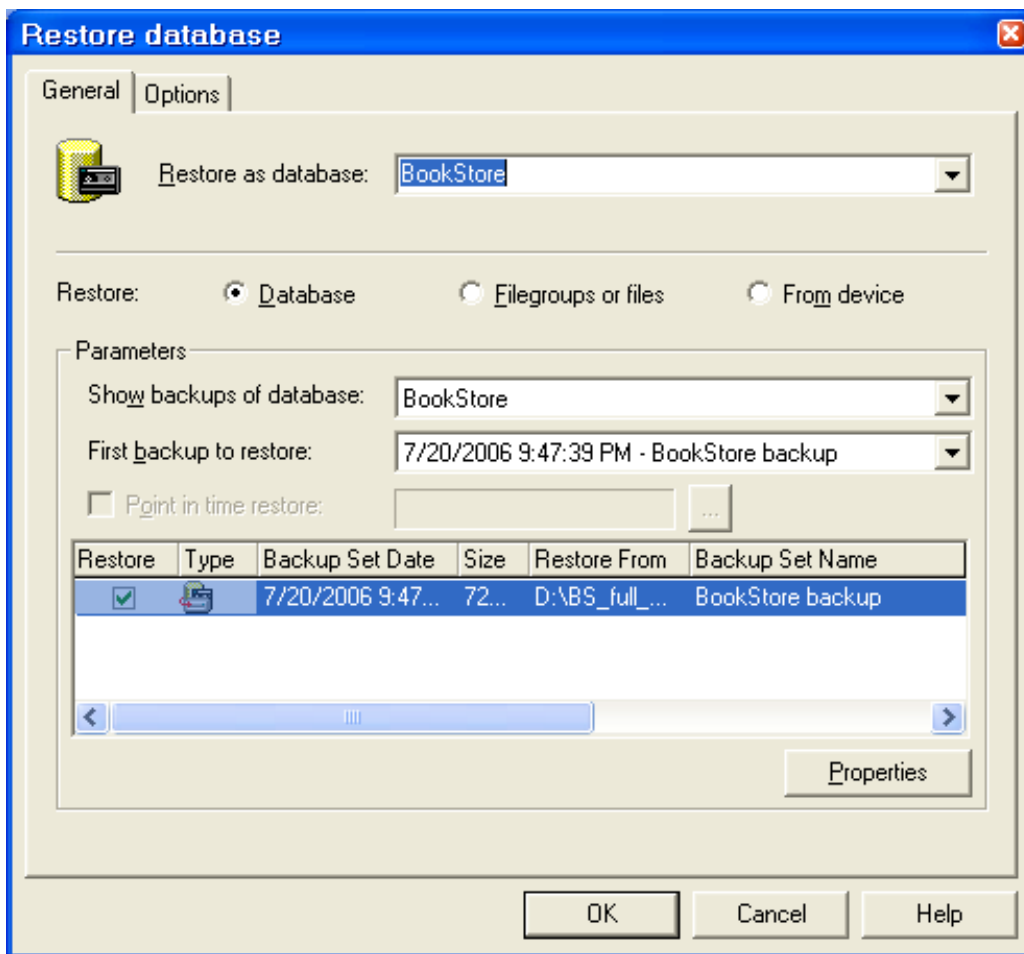
7. [Chức năng không bắt buộc] Chọn **Schedule** check box để xếp lịch cho việc backup (backup operation). Ví dụ bạn muốn tự động backup hàng tuần vào 12h00 ngày chủ nhật, hoặc... thì bạn có thể sử dụng chức năng này.



Khôi phục Full database backup bằng EM

Thực hiện các bước sau:

1. Kích vào server group, và kích vào server chứa Database muốn backup.
2. Kích **Databases**, kích phải chuột vào database, trở chuột vào **All Tasks**, sau đó kích **Restore Database**.



3. Trong **Restore as database** box, soạn thảo hoặc chọn tên database nếu muốn thay đổi tên Database mặc định không.

4. Kích **Database**.

5. Trong danh sách **First backup to restore**, kích vào bản sao lưu muốn được phục hồi(vì có thể có nhiều bản đã được Backup).

6. Trong danh sách **Restore**, kích vào database muốn được phục hồi.

Mô tả quá trình thực hiện:

Khôi phục lại database backup là trả về trạng thái của CSDL khi lệnh backup được thực thi. SQL Server tạo lại CSDL theo các bước sau:

- Chép tất cả dữ liệu trong bản sao vào CSDL khôi phục.
- Bất kỳ giao dịch nào không hoàn thành trong database backup thì được roll back để bảo đảm tính nhất quán dữ liệu.

Quá trình này bảo đảm CSDL sau khi khôi phục là một bản sao của CSDL khi thực hiện sao lưu, trừ những giao dịch không hoàn thành được roll back. Điều này đảm bảo tính toàn vẹn dữ liệu.

Ngoài ra, để tránh việc cố tình viết đè lên CSDL đã tồn tại, quá trình khôi phục thực hiện kiểm tra an toàn một cách tự động. Quá trình khôi phục không thực hiện nếu:

- Tên CSDL khôi phục đã tồn tại trên server và tên CSDL cần khôi phục không tương ứng với tên CSDL ghi trong backup set.
- Tên CSDL khôi phục đã tồn tại trên server nhưng dữ liệu bên trong không giống với dữ liệu bản sao database backup. Ví dụ: CSDL khôi phục có cùng tên với CSDL đã có trong SQL Server nhưng dữ liệu thì khác ví dụ như có những bảng dữ liệu khác.
- Một hoặc nhiều file yêu cầu tạo tự động bằng thao tác khôi phục (không để ý đến CSDL đó tồn tại hay chưa) nhưng những file này có cùng tên với CSDL đã tồn tại rồi.

Tuy nhiên việc kiểm tra an toàn có thể không có tác dụng nếu có mục đích viết đè.

Transaction log backup

Giới thiệu

Chỉ ghi lại những thay đổi trong transaction log. Transaction log backup chỉ chép lại log file. Nếu chỉ có bản sao log file thì không thể khôi phục lại được CSDL. Nó được sử dụng sau khi CSDL đã được khôi phục lại.

Sao lưu transaction log định kỳ để tạo ra 1 chuỗi transaction log backup cho phép user linh động lựa chọn để khôi phục CSDL. Tạo transaction log backup làm cho CSDL có thể khôi phục đến thời điểm xảy ra sự cố.

Khi tạo transaction log backup, điểm bắt đầu backup là:

- Điểm kết thúc của transaction log backup trước đó(nếu có một transaction log backup tạo ra trước đó).
- Transaction log backup như là một phần cuối của database backup hoặc differential backup mới nhất nếu không có transaction log backup nào được tạo ra trước đó (database backup hoặc differential backup chứa một bản sao vùng tích cực của transaction log).

Ta nên dùng transaction log backup nếu hệ thống có những đặc điểm sau:

Sao lưu & phục hồi (Backup & Restore)

- Tài nguyên để thực hiện database backup giới hạn như thiếu vùng lưu trữ hoặc thời gian thực hiện backup. Ví dụ: CSDL 10 terabyte đòi hỏi nhiều thời gian và vùng lưu trữ để backup.
- Bất kỳ việc mất những thay đổi sau lần database backup cuối cùng là không thể chấp nhận được. Ví dụ : hệ thống CSDL kinh doanh tài chính, nó không thể chấp nhận mất bất kỳ giao dịch nào.
- Mong muốn trả về CSDL tại thời điểm xảy ra sự cố. Ví dụ muốn khôi phục lại CSDL trước khi xảy ra sự cố 10 phút.
- CSDL thay đổi thường xuyên.

Vì transaction log backup thường sử dụng tài nguyên ít hơn nên chúng được backup thường xuyên hơn. Điều này giảm khả năng mất dữ liệu hoàn toàn.

Ít gặp trường hợp transaction log backup lớn hơn database backup. Ví dụ CSDL có tỉ lệ giao dịch cao và những giao dịch ảnh hưởng đến phần lớn CSDL gây ra transaction log tăng nhanh hoặc ít sao lưu transaction log. Trong trường hợp này tạo transaction log backup thường xuyên hơn.

Khôi phục CSDL và áp dụng transaction log backup:

- Sao lưu transaction log hiện hành nếu sự cố xảy ra (trừ khi đĩa chứa file transaction log bị hư).
- Khôi phục database backup mới nhất.
- Áp dụng tất cả các transaction log backup được tạo ra sau khi thực hiện full database backup.
- Áp dụng transaction log backup cuối cùng được tạo ra ở bước 1 để khôi phục lại CSDL đến thời điểm xảy ra sự cố.

Vì thế, mặc dù sử dụng transaction log backup tăng khả năng khôi phục, nhưng tạo và áp dụng chúng cũng phức tạp hơn dùng Full Database backup. Khôi phục CSDL sử dụng cả full database backup và transaction log backup chỉ khi ta có chuỗi transaction log backup liên tục.

SQL Server 2000 không cho phép lưu transaction log trong cùng file lưu CSDL. Vì nếu file này hư thì ta không thể sử dụng nó để khôi phục tất cả những thay đổi kể từ lần sao lưu full database backup cuối cùng.

Cắt (truncate) transaction log

Khi SQL Server sao lưu xong transaction log, nó cắt phần không tích cực của transaction log. SQL Server sử dụng lại phần cắt này. Phần không tích cực là phần của transaction log không còn sử dụng nữa trong quá trình khôi phục CSDL vì tất cả giao dịch trong phần này đã hoàn tất. Ngược lại, phần tích cực của transaction log chứa những giao dịch đang chạy và chưa hoàn thành.

Điểm kết thúc phần không tích cực của transaction log, điểm cắt, là điểm đầu tiên của những sự kiện sau:

- Checkpoint gần nhất tương ứng với điểm đầu tiên mà tại đó SQL Server sẽ roll forward những giao dịch trong quá trình khôi phục.
- Bắt đầu của giao dịch tích cực cũ nhất; 1 giao dịch chưa commit hoặc roll back. Tương ứng với điểm đầu tiên mà SQL Server roll back giao dịch trong suốt quá trình khôi phục.

Điều kiện transaction log backups

Transaction log không nên sao lưu:

- Nếu CSDL thiết lập **trunc. log on chkpt** (truncate log on checkpoint) là TRUE (thì không thể tạo ra log record dùng để roll forward); tạo database backup hoặc differential backup thay thế.
- Nếu bất kỳ thao tác nonlogged nào xảy ra trong CSDL kể từ khi thực hiện sao lưu full database backup lần cuối cùng; tạo full database backup hoặc differential backup thay thế.
- Cho đến khi thực hiện sao lưu full database backup vì transaction log backup chứa những thay đổi của database backup.
- Nếu transaction log bị cắt, trừ khi database backup hoặc differential backup được tạo ra sau khi cắt transaction log.
- Nếu bất kỳ file nào được thêm vào hay xóa khỏi CSDL; database backup nên tạo ra thay thế ngay lúc đó.

Cách tạo transaction log backup bằng EM

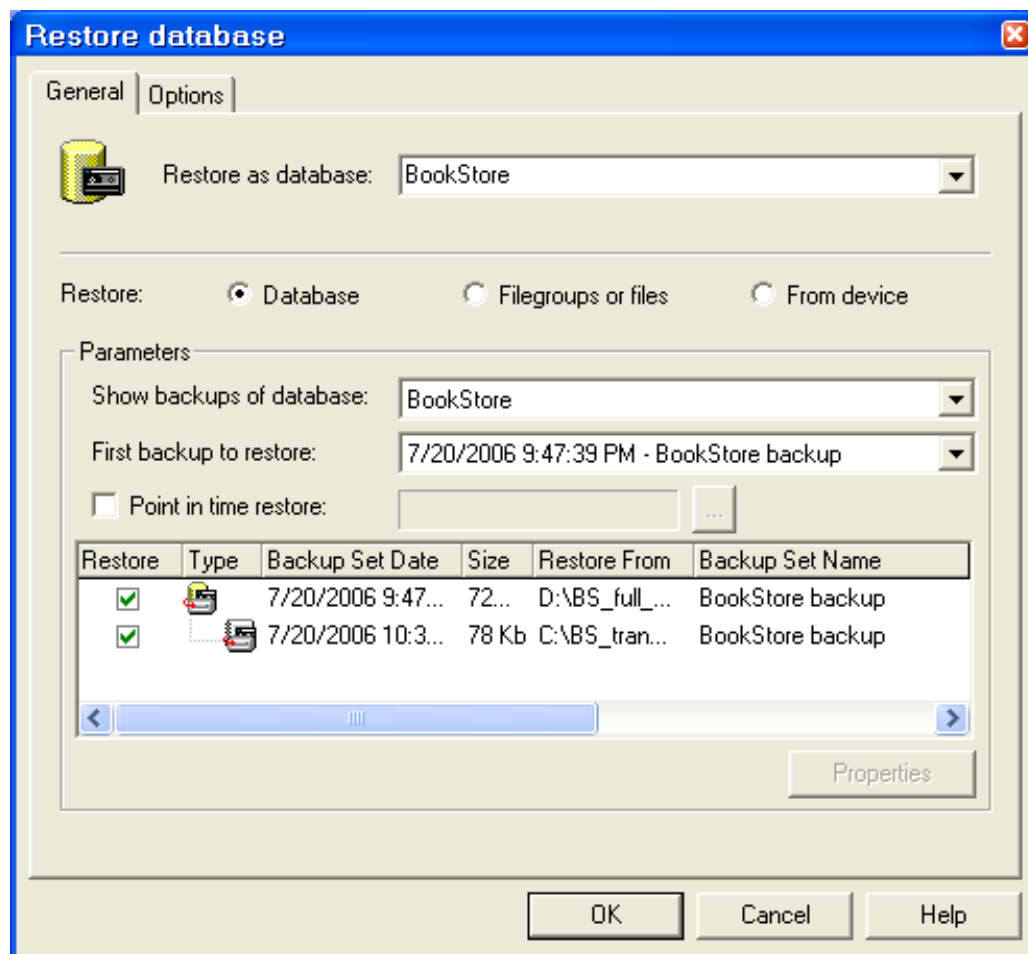
Cách làm tương tự như đối với tạo Full Database backup, tuy nhiên ở bước 4 ta phải chọn **Transaction log**.

Sao lưu & phục hồi (Backup & Restore)

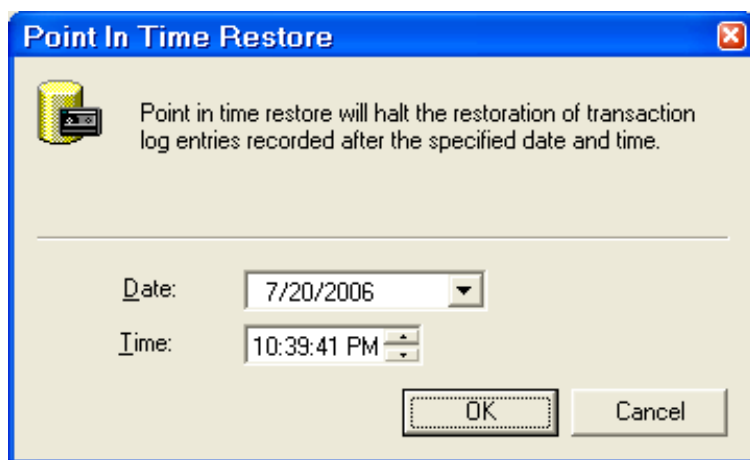
Nếu lựa chọn Transaction Log không được phép thì ta phải kiểm tra lại recovery model để thiết đặt là Full hoặc Bulk-Logged. Bởi vì chỉ những model này mới hỗ trợ transaction log backup.

Khôi phục transaction log backup bằng EM

Thực hiện tương tự như cách **Khôi phục Full database backup**.



Tuy nhiên, ta có thể xác định được thời điểm nào đó muốn khôi phục dữ liệu trong quá khứ.



Differential backup

Differential backup chỉ ghi lại những trang thay đổi ngay sau khi thực hiện sao lưu full database lần cuối cùng. Do đó, sẽ nhanh hơn thực hiện full database backup rất nhiều.

Không như transaction log backup, differential backup không tạo lại CSDL chính xác tại thời điểm xảy ra sự cố, nó cũng như full database backup, chỉ tạo lại CSDL tại thời điểm backup cuối cùng. Vì thế, differential backup thường được bổ sung bằng cách tạo transaction log sau mỗi differential backup. Sử dụng kết hợp database backup, differential backup, và transaction log backup ta có thể giảm tối thiểu khả năng mất dữ liệu và thời gian khôi phục dữ liệu.

Ta nên dùng differential backup nếu hệ thống có những đặc điểm sau:

- Dữ liệu ít quan trọng và những thay đổi của CSDL có thể tạo lại bằng tay tốt hơn là dùng transaction log.
- Tài nguyên để thực hiện database backup giới hạn như thiếu vùng lưu trữ hoặc thời gian thực hiện sao lưu. Ví dụ: CSDL 10 terabyte đòi hỏi nhiều thời gian và vùng lưu trữ để thực hiện sao lưu.
- Tối thiểu hóa thời gian khôi phục và giảm việc mất những giao dịch bằng cách kết hợp differential backup với full database backup và transaction log backup.

Xem xét ví dụ sau:

24:00 (Thứ 3)	Full Database backup (1)
6:00 (Thứ 4)	Differential database backup (2)
12:00 (Thứ 4)	Differential database backup (3)

18:00 (Thứ 4)	Differential database backup (4)
24:00 (Thứ 4)	Full Database backup (5)
6:00 (Thứ 5)	Differential database backup (6)
12:00 (Thứ 5)	Differential database backup (7)

Differential backup tạo vào lúc 6:00 ngày thứ tư (2) chứa tất cả những thay đổi của database backup tạo từ lúc 24:00 ngày thứ ba(1).

Differential backup tạo vào lúc 6:00 ngày thứ năm(6) chứa tất cả những thay đổi của database backup tạo từ lúc 24:00 ngày thứ tư (5).

Nếu có khôi phục CSDL đến trạng thái vào 12:00 ngày thứ năm, ta thực hiện những bước sau:

-Khôi phục database tạo lúc 24:00 ngày thứ tư.

-Khôi phục differential backup tạo lúc 12:00 ngày thứ năm.

Bất kỳ thay đổi nào sau trưa thứ năm đều bị mất trừ khi có khôi phục transaction log backup.

Các tạo Differential backup và khôi phục cũng tương tự như thực hiện với Full database backup.

Sự khác nhau giữa differential backup và transaction log backup:

-Giống nhau: tối thiểu hóa thời gian sao lưu.

-Khác nhau:

- Differential backup: chỉ lưu lần thay đổi cuối cùng
- Transaction log backup: chứa tất cả những thay đổi kể từ lần sao lưu full database backup cuối cùng.

Do differential backup lưu những trang thay đổi, gồm những trang dữ liệu và cả trang transaction log thay đổi. Vì sao lưu differential backup có kích thước lớn hơn sao lưu transaction backup nên ta ít sao lưu differential backup thường xuyên so với sao lưu transaction backup. Do đó ta không thể khôi phục CSDL đến thời điểm xảy ra sự cố khi sao lưu differential backup và không thể khôi phục CSDL đến thời điểm mà ta mong muốn.

File hoặc Filegroup backup

Chỉ sao lưu những file CSDL chỉ định. File hoặc file group backup thường được sử dụng chỉ khi không có đủ thời gian để sao lưu toàn bộ CSDL.

Sử dụng file hoặc file group backup có thể tăng tốc độ khôi phục bằng cách chỉ khôi phục những file hoặc filegroup bị hư. Khi sao lưu file hoặc file group thì SQL Server không có sao lưu file transaction log do đó ta phải tạo transaction log backup sau khi sao lưu file hoặc file group.

Một CSDL có 2 filegroup filegroup_a và filegroup_b nhưng chỉ có đủ thời gian để sao lưu 1 nửa filegroup, do đó:

- Sao lưu filegroup_a vào các ngày thứ hai, tư, sáu.
- Sao lưu transaction log ngay sau khi sao lưu filegroup.
- Sao lưu filegroup_b vào các ngày thứ năm, sáu, bảy.
- Sao lưu transaction log ngay sau khi sao lưu filegroup.

Khôi phục file hoặc filegroup backup:

File hoặc file group có thể được khôi phục từ database backup hoặc file hoặc file group. Ta không cần thiết backup transaction log nếu không có thay đổi từ khi sao lưu file hoặc file group.

Nếu filegroup_b cần khôi phục vì 1 bảng trong filegroup bị hư, ta sẽ:

- Khôi phục filegroup_b backup tạo vào ngày thứ năm.
- Áp dụng transaction log backup của filegroup_b.