

Análise Comparativa de Tempo de Restauração: Backups Incrementais, Diferenciais e Intercalados

Fernanda Ribeiro Martins

Graduanda em Ciência da
Computação

Instituto Federal Catarinense
Campus Blumenau

fernandamartinsrm@gmail.com

Luiz Ricardo Uriarte

Professor do Instituto Federal
Catarinense

Campus Blumenau

luiz.uriarte@ifc.edu.br

Monique Ellen dos Santos

Graduanda em Ciência da
Computação

Instituto Federal Catarinense
Campus Blumenau

moniqueellen1402@gmail.com

ABSTRACT

The use of backups to store copies of an organization's digital data is a fundamental and necessary practice as a corrective measure against data corruption in situations such as natural disasters, human errors, hardware and software failures. From these copies, it is possible to recover lost or incorrectly altered data.

The objective of this research is to compare the restoration time among three approaches: incremental-only backups, the alternation of full and incremental backups, and differential backups.

Through Bacula, it is possible to manage backups and data restoration, as well as obtain information regarding the performed operations. As a result, the restoration time was 52,73% shorter when planning a scheme compared to performing only incremental backups, and 18,75% shorter compared to differential backups. However, this approach also demonstrated the need for a larger storage space for its execution.

Keywords

Backups; Incremental; Full; Differential; Bacula, Restoration.

RESUMO

A utilização de backups para salvar cópias dos dados digitais de uma organização é uma prática fundamental e necessária como medida corretiva contra a corrupção de dados em situações como desastres naturais, erros humanos e falhas de hardware e software. A partir dessas cópias, é possível recuperar os dados perdidos ou alterados incorretamente. O objetivo desta pesquisa foi comparar o tempo de restauração entre três estratégias de backups: apenas incrementais, intercalação de backups completos e incrementais e apenas diferenciais. Por meio do Bacula, foi possível realizar a gestão de backups e restauração de dados, além de obter informações a respeito das operações realizadas. Como resultado, o tempo de restauração foi 52,73% menor ao planejar um esquema de alternância entre os backups se comparado a apenas backups incrementais, e 18,75% menor em relação a backups diferenciais. No entanto, essa abordagem também demonstrou a necessidade de um maior espaço de armazenamento para sua execução.

Palavras-chave

Backup; Incremental; Completo; Diferencial; Bacula; Restauração.

1. INTRODUÇÃO

Em um contexto de crescente migração do armazenamento de dados físicos para os digitais, torna-se indispensável pesquisar e implementar procedimentos que assegurem a manutenção e a recuperação das informações diante de possíveis incidentes. [1] explicam que o backup é uma maneira eficaz de corrigir uma fonte

de dados em caso de corrupção, pois armazenar a cópia dos dados de um dispositivo de armazenamento possibilita a reconstrução do disco ou a recuperação de arquivos individuais que foram comprometidos.

Existem diferentes tipos de backups: o backup completo, que copia todos os dados de uma fonte; o backup incremental, que salva apenas as alterações desde o último backup de qualquer tipo [3]; e o backup diferencial, que copia todos os dados desde o último backup completo [2]. Ambos os métodos apresentam vantagens e desvantagens, impactando principalmente o espaço disponível de armazenamento e o tempo de execução. Assim, em sistemas nos quais o tempo de restauração é um fator crítico, torna-se essencial avaliar qual das estratégias oferece melhor desempenho.

Para a administração de backup e restauração, existem ferramentas que facilitam esse processo. Como demonstra [5], o Bacula oferece facilidade de uso e automação, além de um tempo de execução elevado em relação a outras ferramentas populares, como Rdiff-backup e Zbackup.

Dessa forma, o objetivo da pesquisa é comparar o tempo necessário para a restauração de dados por meio do Bacula partindo de três estratégias de backup: uma sequência de backups incrementais, uma intercalação de backups incrementais e completos e uma sequência de backups diferenciais.

2. REFERENCIAL TEÓRICO

2.1 Tipos de backup

A perda ou corrompimento de dados em dispositivos de armazenamento podem ocorrer por diversos motivos, como erro de usuário, falhas de disco, problemas em hardware e software, desastres naturais e outras causas externas. Esses incidentes podem ocasionar sérias consequências, incluindo perda de clientes e encomendas, danos financeiros, prejuízo de tempo e enfraquecimento da imagem de uma empresa no mercado [3]. Diante dos riscos, é fundamental planejar e executar backups regulares, que agem como medida corretiva contra tais eventos.

2.1.1 Backup completo

É uma cópia de todos os dados de um sistema de arquivos em um dispositivo de backup. Por envolver a leitura e gravação integral dos dados, esse procedimento demanda maior capacidade de armazenamento e tempo de execução [3]. No entanto, o tempo de restauração a partir de um backup completo é mais rápido, uma vez que os dados estão contidos em um único arquivo e o processo se resume à transferência de dados, sem a necessidade de operações complexas de recuperação.

2.1.2 Backup incremental

É uma cópia das alterações realizadas nos dados desde um determinado ponto de referência, que pode ser um ponto temporal ou um backup completo [2].

Essa abordagem reduz o tempo da operação de backup e o espaço de armazenamento necessário. No entanto, a recuperação apresenta maior complexidade, e o problema se agrava quando um dos dados incrementais é perdido ou corrompido [2]. “De acordo com Chervenak, Vellanki, Kurmas [1998] as restaurações serão mais lentas em um sistema de backup incremental, que deve começar com o último backup completo e aplicar as alterações dos backups incrementais subsequentes”.

2.1.3 Backup diferencial

É uma cópia de todos os dados que foram modificados desde o último backup completo, funcionando como um backup incremental cumulativo [2].

Em comparação com os demais tipos de backups, o diferencial é mais rápido que o completo, porém mais lento que o incremental. Em contrapartida, sua operação de recuperação é menos complexa que a do incremental, pois requer apenas dois arquivos de backup, embora ainda seja mais lenta que a do completo [2].

2.2 Comparação entre os tipos de backup

Conforme analisado por [3], a escolha de qual tipo de backup executar e quando fazê-lo depende das necessidades específicas do sistema e do que é viável no contexto. Em determinados sistemas, pode ser importante minimizar o tempo em que as operações de backup ocorrem, enquanto em outros o tempo de restauração requer maior prioridade.

Além disso, o espaço de armazenamento disponível deve ser considerado, especialmente ao lidar com grandes volumes de dados e os custos associados.

O Quadro 1 apresenta uma comparação entre os tipos de backup com base em três aspectos principais: operação, tempo de restauração e uso de armazenamento.

Quadro 1 - Comparação entre tipos de backup

Tipo de Backup	Operação de Escrita/Leitura	Operação de Restauração	Espaço de Armazenamento
Backup Completo	Lenta	Rápida	Maior
Backup Diferencial	Intermediária	Intermediária	Médio
Backup Incremental	Rápida	Lenta	Menor

Considerando os pontos positivos e negativos de ambas as técnicas, pode ser planejado um esquema de intercalação, onde um backup completo é realizado após uma série de backups incrementais ou diferenciais, e o processo se repete.[1]

2.3 Ferramentas de backups e restauração

Ferramentas de backup e restauração são soluções essenciais para automatizar, simplificar e monitorar esses processos. “Segundo Preston [2007], existem dois tipos de automatização. Um tipo permite seus backups completarem um ciclo completo sem requerer de você uma intervenção manual, como injeção e carregamento de novos volumes. [...]. O segundo tipo de autenticação é na verdade muito mais importante. Esse tipo de automatização faz referência a como o backup “pensa”. Seu processo de backup deve saber fazer o backup sem lhe informar.”

2.3.1 Bacula

Bacula é um software open source baseado em uma estrutura Cliente/Servidor que permite a administração de backups, restaurações e verificação de dados. Oferece diversas opções de configuração de armazenamento, como backup em múltiplos dispositivos, sendo escalável desde computadores únicos até uma rede com vários computadores [4].

Com base em [5], a configuração do Bacula é realizada ao definir informações referentes aos principais componentes do sistema: Director, Console, File, Storage e Catalog (Database Server).

2.3.1.1 Director

Programa que controla todas as operações de backup, restauração, verificação e arquivos, sendo responsável principalmente por executar agendamentos (schedules) e autenticação de conexões. [4, 5]

2.3.1.2 Console

Programa que fornece ao administrador uma interface para o gerenciamento de jobs, visualização de mensagens e status de informações, realizando assim a comunicação com o programa Director atual.

A interface pode ter tanto por linha de comando (CLI), que apresenta um maior número de funcionalidades, quanto por interface gráfica (GUI). [4, 5]

2.3.1.3 File Daemon (FD)

Programa responsável por acessar os dados dos arquivos e diretórios do cliente que serão enviados ao Storage Daemon (SD). Faz isso comunicando-se com o programa Director e determinando quais dados do cliente serão utilizados no backup e a forma de armazenamento, transferindo os dados selecionados. [4, 5]

2.3.1.4 Storage Daemon (SD)

É a parte do sistema que faz a comunicação direta com os meios de armazenamento físico a serem usados para backup e recuperação. Esses meios de armazenamento podem ser discos, nuvem, fitas ou outras mídias. [4, 5]

2.3.1.5 Catalog (Database Server)

Funciona como um catálogo que armazena os índices (localização) e volume dos arquivos que foram realizados backups, armazenado essas informações em um banco de dados, que atualmente suporta SQLite, MySQL e PostgreSQL. [4, 5]

3. PROCEDIMENTOS METODOLÓGICOS

3.1 Configuração do Ambiente Virtual

Para a execução dos testes, foi criada uma máquina virtual em um ambiente de simulação (VirtualBox), no qual foram instalados o Ubuntu Server e o software Bacula, permitindo a simulação de casos de backup e restauração de dados. A máquina foi configurada com 40GB de disco e 10 MB de memória, fornecendo espaço suficiente para armazenar os backups e dados utilizados nos testes.

3.2 Instalação e configuração do Bacula

A instalação do Bacula foi realizada em um ambiente único, com o objetivo de simplificar a configuração e execução dos testes. Dessa forma, foi possível simular um ambiente real em menor escala e gerenciar todos os recursos e processos em um mesmo local. Para realização dos testes, foram configurados os principais arquivos do sistema.

No Bacula Director foi definido o cliente, a forma de armazenamento dos backups e recuperação (Storage), bem como os diretórios a serem incluídos ou excluídos durante os procedimentos (Filesets). Além disso, foram criados os jobs de backup incremental, backup completo e restauração, que tem a função de definir e executar tarefas, especificando quais dados serão incluídos, qual cliente será utilizado, onde os dados serão armazenados e quando o job será executado (Schedule).

No Storage Daemon, foram configurados os dispositivos de armazenamento nos quais os backups e restaurações devem ser gravados.

No File Daemon foi especificado o nome do cliente, qual o Bacula Director e a porta que deve ser utilizada para a comunicação entre eles.

E para permitir uma visualização gráfica da execução e listagem dos jobs, foi instalado na máquina o Webmin, um software para administrar remotamente servidores Linux.

3.3 Execução dos testes

Os dados utilizados nos testes foram obtidos a partir de repositórios públicos no GitHub. No total, foram utilizados aproximadamente 5.6 GB de dados, que serviram como base para as diferentes abordagens de teste.

3.3.1 Restauração de backups incrementais

Inicialmente, foi realizado um total de sete backups incrementais sequenciais em uma pasta do cliente após um backup total, que serve como ponto de partida para a restauração. Em seguida, foi simulada uma perda de dados significativa na pasta do cliente, e executada uma operação de restore. Concluída a restauração, os dados retornaram ao local original.

3.3.2 Restauração de backups intercalados

Nesta abordagem, os sete backups incrementais foram realizados de forma intercalada com backups completos, sequencialmente, em uma pasta do cliente. Assim como na abordagem anterior, foi simulada a corrupção dos dados e executada a operação restore, com os dados novamente recuperados na pasta do cliente.

3.3.3 Restauração de backup diferenciais

Da mesma forma que na restauração de backups incrementais, foi realizado um total de sete backups diferenciais subsequentes a um backup completo. Em seguida, foi simulada a perda de dados na pasta do cliente e executada a operação de restore.

3.4 Coleta dos resultados

O Bacula possui uma configuração de mensagens associadas a um job, permitindo definir como essas mensagens serão manipuladas e enviadas, seja por e-mail ao usuário ou destinadas a um arquivo [4].

Assim, informações sobre o job, como data e horário de início e término, identificação e nome do job, quantidade de bytes restaurados, tempo de execução e local direcionado, foram obtidos através desse recurso.

4. RESULTADOS

A partir do primeiro teste de restauração utilizando a abordagem de backup incrementais (ver figura 1), foi obtido um tempo total de execução de 55 segundos. (ver figura 2).

client	name	starttime	joblevel	jobfiles	jobbytes
teste-fd	Incremental1	2024-10-21 22:02:26	I	1074	3217810361
teste-fd	Incremental1	2024-10-21 21:56:59	I	290	1951070220
teste-fd	Incremental1	2024-10-21 21:53:51	I	4694	1024440084
teste-fd	Incremental1	2024-10-21 21:38:15	I	1985	114367948
teste-fd	Incremental1	2024-10-21 21:37:12	I	1548	53774098
teste-fd	Incremental1	2024-10-21 21:36:18	I	99	4598168
teste-fd	Incremental1	2024-10-21 21:35:28	I	329	1297437544

Figura 1 - Lista dos backups executados pela abordagem de backups incrementais

```
Build OS: x86_64-pc-linux-gnu ubuntu 24.04
JobId: 68
Job: Restore.2024-10-21_22.04.58.22
Restore Client: "teste-fd" 13.0.4 (12Feb24) x86_64-pc-linux-gnu,ubuntu,24.04
Where: /
Replace: Always
Start time: 21-Oct-2024 22:05:00
End time: 21-Oct-2024 22:05:55
Elapsed time: 55 secs
Files Expected: 6,056
Files Restored: 6,056
Bytes Restored: 6,193,324,665 (6.193 GB)
Rate: 112605.9 KB/s
FD Errors: 0
FD termination status: OK
SD termination status: OK
Termination: Restore OK
```

Figura 2 - Mensagem gerada após a execução da restauração pela abordagem de backups incrementais

Já no segundo teste de restauração, utilizando a abordagem de backup intercalados (ver figura 3), o tempo total de execução foi de 26 segundos (ver figura 4).

Full	2024-10-23 09:33:10	B	F	1,800	60,310,368	T
Full	2024-10-23 09:37:53	B	F	1,800	60,310,368	T
Incremental	2024-10-23 09:13:25	B	F	3,784	174,730,627	T
Incremental	2024-10-23 09:14:58	B	I	737	554,266,336	T
Full	2024-10-23 09:19:22	B	F	4,809	2,680,067,181	T
Incremental	2024-10-23 09:22:14	B	I	1,114	1,969,701,457	T
Incremental	2024-10-23 09:28:26	B	I	250	3,199,193,469	T

Figura 3 - Lista dos backups executados pela abordagem de backups intercalados

```
Build OS: x86_64-pc-linux-gnu ubuntu 24.04
JobId: 19
Job: Restore.2024-10-23_08.33.07.25
Restore Client: "ubuntu-fd" 13.0.4 (12Feb24) x86_64-pc-linux-gnu,ubuntu,24.04
Where: /
Replace: Always
Start time: 23-Oct-2024 08:33:09
End time: 23-Oct-2024 08:33:35
Elapsed time: 26 secs
Files Expected: 5,882
Files Restored: 5,882
Bytes Restored: 5,897,891,883 (5.897 GB)
Rate: 226842.0 KB/s
FD Errors: 0
FD termination status: OK
SD termination status: OK
Termination: Restore OK
```

Figura 4 - Mensagem gerada após a execução da restauração pela abordagem de backups intercalados

Por fim, no terceiro teste a abordagem usando apenas backup diferenciais (ver figura 5) resultou em um tempo total de 32 segundos (ver figura 6).

Full	2024-10-23 10:42:00	B	F	140	1,873,222	T
Dif	2024-10-23 10:43:08	B	F	238	6,468,648	T
Dif	2024-10-23 10:43:43	B	D	1,555	53,836,922	T
Dif	2024-10-23 10:44:30	B	D	3,539	168,257,181	T
Dif	2024-10-23 10:47:24	B	D	4,275	722,523,511	T
Dif	2024-10-23 10:49:35	B	D	4,564	2,673,593,795	T
Dif	2024-10-23 11:03:58	B	D	5,637	5,091,418,437	T

Figura 5 - Lista dos backups executados pela abordagem de backups diferenciais

```
2024-10-23 10:42:00, backup 1000, bacula backup dir 1000, via fd277.
Build OS: x86_64-pc-linux-gnu ubuntu 24.04
JobId: 153
Job: Restore_2024-10-23_13.05.05_32
Restore Client: 'ubunt-fd' 13.0.4 (12Feb24) x86_64-pc-linux-gnu,ubuntu,24.04
Where: /
Replace: always
Start time: 23-Oct-2024 13:05:08
End time: 23-Oct-2024 13:05:40
Elapsed time: 32 secs
Files Expected: 5,883
Files Restored: 5,883
Bytes Restored: 5,897,891,883 (5.897 GB)
Rate: 184309.1 KB/s
FD Errors: 0
FD termination status: OK
SD termination status: OK
Termination: Restore OK
```

Figura 6 - Mensagem gerada após a execução da restauração pela abordagem de backups diferenciais

Com base nos resultados obtidos, é possível constatar que o tempo de restauração é cerca de 52,73% menor quando se realiza um esquema de intercalação das técnicas de backup em relação a apenas backups incrementais, e 18,75% menor em relação a apenas backups diferenciais.

Entretanto, um problema verificado durante a execução dos testes diz respeito ao espaço de armazenamento. Enquanto que a primeira abordagem utilizou 5.6 GB de dados, a segunda exigiu 8.2 GB, resultando em uma aumento de 46,43% no uso de memória. A terceira abordagem, por sua vez, demandou 9.5 GB. Comparando a terceira com a segunda abordagem, houve uma redução de 15,85%.

Assim, embora a comparação do tempo seja favorável à primeira abordagem, é crucial considerar o espaço de armazenamento necessário, uma vez que a escolha do esquema de backup deve alinhar-se com às necessidades específicas do sistema e aos recursos disponíveis.

5. CONCLUSÃO

O trabalho utilizou o Bacula, uma ferramenta de backup, restauração e verificação de dados, para analisar comparativamente o tempo de restauração de três abordagens de backups: apenas incrementais, intercalação de backups completos e incrementais e apenas diferenciais.

Em um ambiente de virtualização foi possível simular perda e corrupção de dados, realizar a restauração e obter informações relacionadas.

Os resultados obtidos mostraram que a estratégia de realizar esquemas intercalados teve uma vantagem de 52,73% no tempo de execução em relação apenas aos backups incrementais e 18,75% em comparação com apenas os backups diferenciais. Entretanto, indicam diferenças consideráveis em relação ao espaço de disco usado nas operações (ver figura 8), destacando a importância de combinar o planejamento de backups de acordo com as prioridades do sistema.

Comparação entre abordagens

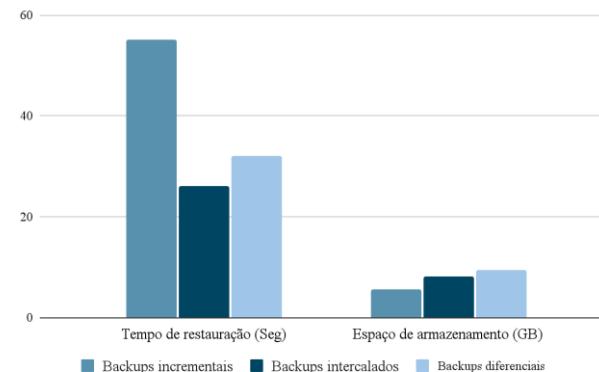


Figura 8 - Comparação entre tipos de backup

REFERÊNCIAS

- Chervenak, A., Vellanki Z. and Kurmas, Z. 1998. Protecting File Systems: A Survey of Backup Techniques. In Proceedings of the Joint NASA and IEEE Mass Storage Conference. https://www.researchgate.net/publication/2431989_Protecting_File_Systems_A_Survey_of_Backup_Techniques. Acesso em 16-10-2024.
- Nadee, P. e Somwang, P. 2021. Efficient incremental data backup of unison synchronize approach. Thailand: Bulletin of Electrical Engineering and Informatics. https://www.researchgate.net/publication/355375926_Efficient_incremental_data_backup_of_unison_synchronize_approach. Acesso em: 16-09-2024.
- Preston, W. C. 2007. Backup & Recovery: Inexpensive Backup Solutions for Open Systems. Sebastopol: O'Reilly Media. Disponível em: <https://www.pdfdrive.com/backup-recovery-inexpensive-backup-solutions-for-open-systems-e162525399.html>. Acesso em: 16-10-2024.
- Sibbald, K. 2024. Bacula Main Reference Manual. Bacula Community Edition v.15.0.2. Disponível em: <https://www.bacula.org/15.0.x-manuals/en/main/main.pdf>. Acesso em: 19-10-2024.
- Reis, D. D. R. A. 2018. Avaliação de ferramentas de backup para implementação em uma organização empresarial. Orientador: Raimundo Viégas Junior. 2018. 61 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Faculdade de Computação, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém. Disponível em: <http://bdm.ufpa.br/jspui/handle/prefix/1434>. Acesso em: 20-10-2024.