

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CURSO DE TECNOLOGIA EM SISTEMAS PARA INTERNET
CÂMPUS GUARAPUAVA**

HYTHAN MATHEUS CORREIA DE OLIVEIRA

**ESTUDO COMPARATIVO DE OTIMIZAÇÃO DE BACKUPS COM O USO DE
RSYNC E RCLONE**

**PROPOSTA DE TRABALHO DE CONCLUSÃO DE CURSO DO CURSO SUPERIOR EM
TECNOLOGIA EM SISTEMAS PARA INTERNET**

**GUARAPUAVA
1º Semestre de 2021**

HYTHAN MATHEUS CORREIA DE OLIVEIRA

**ESTUDO COMPARATIVO DE OTIMIZAÇÃO DE BACKUPS COM O USO DE
RSYNC E RCLONE**

Proposta de Trabalho de Conclusão de Curso de Graduação, apresentado à disciplina de Trabalho de Conclusão de Curso 1, do Curso Superior de Tecnologia em Sistemas para Internet – TSI – da Universidade Tecnológica Federal do Paraná – UTFPR – Campus Guarapuava, como requisito parcial para obtenção do título de Tecnólogo em Sistemas para Internet.

Orientador (a): Prof. Dr. Hermano Pereira

GUARAPUAVA
1º Semestre de 2021

1. PROPOSTA DE TRABALHO DE CONCLUSÃO DE CURSO

1.1. Título

Estudo Comparativo de Otimização de Backups com uso de Rsync e Rclone.

1.2. Modalidade do Trabalho

Pesquisa.

1.3. Área do Trabalho

Administração de rotina de servidores.

1.4. Resumo

Backup tem como definição ser uma cópia de segurança que se faz regularmente para assegurar que um arquivo ou o conjunto de dados não se perca, sendo usado quando há prejuízo ou danos no arquivo original. Dentre os métodos de backups mais comuns está o incremental, que consiste em, na primeira execução, gerar um backup completo e periodicamente gerar backups apenas dos dados que foram alterados desde a última execução. A utilização de um backup incremental, para uma grande quantidade de dados gerados diariamente, aparenta ser uma opção plausível. Pensando nisso, a presente pesquisa fará uma análise quantitativa entre duas ferramentas que possibilitam a execução de tal método de backup, sendo elas Rsync e Rclone. As análises feitas auxiliarão na escolha da ferramenta que apresente o melhor desempenho para um servidor.

2. DESCRIÇÃO DA PROPOSTA

2.1. Introdução

Segundo o dicionário Aurélio (2010) o *backup* tem como definição ser uma cópia de um arquivo que é guardada como reserva para o caso de destruição ou inutilização do arquivo original; cópia de segurança.

O procedimento de salvar cópias, tão usual nos dias de hoje, é uma técnica usada a tempos, por exemplo, antigamente as listas telefônicas continham todos os números de telefone de uma determinada cidade ou região, como não era possível guardar por muito tempo números na memória, as pessoas utilizavam agendas telefônicas ou cadernos, para guardar os números de contatos que consideravam importantes.

Outra forma de *backup* utilizada antes do processo tecnológico era a escrita em cadernos ou diários, como os cadernos de receitas, repassados de geração a geração, onde eram anotadas receitas e dicas para realizar pratos, respeitando a cultura advinda das mais diversas etnias. Fatos como esses, demonstram que as pessoas faziam *backup* antes mesmo de se tornar algo rotineiro. O uso de formas para registrar informações foi sendo ampliado ao longo do tempo para suprir as necessidades das pessoas.

Atualmente, com o avanço tecnológico, as pessoas passaram a salvar seus dados em aplicativos online. Isso possibilitou a liberdade de armazenar seus dados e deixá-los acessíveis a qualquer momento e em qualquer lugar. Um exemplo de ferramenta muito utilizada atualmente é o Google Drive, que possibilita o armazenamento e compartilhamento de arquivos e pastas, que ficam salvos nos servidores da Google, e podem ser acessados de qualquer dispositivo móvel ou computador, desde que possuam o acesso necessário, sendo eles conexão com Internet bem como usuário e senha. Dessa forma, cabe a empresa responsável pela ferramenta armazenar e gerir os *backups* de cada usuário.

Os métodos mais comuns de *backup* utilizados na computação são: backup completo, incremental, diferencial, proteção contínua de dados (CDP), *backup* de metal puro e recuperação instantânea. (IBM, 2019).

Tomando como exemplo, uma estrutura cliente-servidor, em que um servidor recebe uma requisição e retorna uma resposta, e sabendo que o mesmo precisa de alta disponibilidade para tais requisições, o ideal não é sobrecarregá-lo com processos muito demorados, neste caso com um processo de *backup*. Além disso, um servidor que estiver nessas condições de sobrecarga, não conseguirá processar corretamente suas aplicações, o que prejudicará

diretamente a experiência do usuário, como por exemplo, sites que demoram muito tempo para carregar. Pensando nisso, a utilização de um *backup* incremental, para uma grande quantidade de dados gerados diariamente, aparenta ser uma opção plausível uma vez que consiste em, na primeira execução, gerar um *backup* completo e periodicamente gerar *backups* apenas dos dados que foram alterados desde a última execução, que, em teoria, diminui o tempo de execução do processo de *backup*.

Tal processo de *backup* necessita do uso de procedimentos e/ou ferramentas para sua execução. Por esse motivo, este trabalho propõe analisar o impacto do uso das ferramentas Rsync (TRIDGELL e MACKERRAS, 1996) e Rclone (CRAIG-WOOD, 2014) no processo de *backup* incremental. O Rsync é uma ferramenta de cópia de arquivos extremamente rápida e versátil, pela qual é possível copiar arquivos entre ou para outra máquina. Possui um algoritmo de transferência delta, que envia apenas os dados diferentes entre a origem e o destino. (TRIDGELL e MACKERRAS, 1996). Inspirado no Rsync, o Rclone, desenvolvido por Nick Craig-Wood, consiste em ser uma ferramenta de linha de comando para gerenciar arquivos no armazenamento em nuvem. Assim como o Rsync, o Rclone permite copiar arquivos de uma origem para um destino, desconsiderando os já copiados. (CRAIG-WOOD, 2014).

De acordo com o resultado da análise realizada e os parâmetros utilizados, uma empresa que busque optar pelo uso de uma dessas ferramentas, Rsync ou Rclone, possuirá subsídios para tomar a decisão.

Ao analisar duas ferramentas e produzir um parâmetro entre ambas, pode-se utilizar a que mais se encaixa nos anseios da empresa, sendo este o objetivo básico, ou seja, auxiliar de forma rápida no momento de escolha entre o Rsync e Rclone.

2.2. Objetivos

2.2.1. Objetivo Geral

Análise e comparação de desempenho de um servidor durante a execução de backup, com as ferramentas Rclone e Rsync.

2.2.2. Objetivos Específicos

- Identificar um cenário adequado para testes de *backup* que seja semelhante ao aplicado usualmente em empresas de pequeno/médio porte;
- Extrair desse cenário informações sobre desempenho: processamento, armazenamento e tempo de execução;

- Analisar e discutir sobre o resultado da aplicação das ferramentas de *backup* Rsync e Rclone sobre esse cenário de testes;

2.3. Procedimentos Metodológicos

Com o intuito de verificar qual ferramenta é mais eficaz no processo de *backup* de dados de um servidor, as ferramentas Rsync e Rclone serão comparadas nas rotinas de *backup*.

Para realização dos testes será necessário configurar um servidor Linux Debian (MURDOCK. 1993) em uma VirtualBox (ORACLE, 2009), o qual será alimentado com arquivos, como por exemplo, imagens, arquivos de texto, etc. A VirtualBox será configurada com 6gb de memória RAM, alocando 20g de espaço em disco para os dados e 4 núcleos do processador (Intel(R) Core i5-3330 CPU 3.00GHz).

Serão instaladas e configuradas as ferramentas necessárias, sendo elas Rsync, Rclone e Htop. Além disso, será escrito um *script*¹ com os comandos necessários de cada ferramenta.

Ambas as ferramentas possuem configurações distintas e para realização desse trabalho será utilizada a versão 1.55.1 do Rclone e a versão 3.2.3 do Rsync, estas as versões correntes das ferramentas até o presente dia da execução desse trabalho.

Para a realização dos *backups* será necessária uma organização em dias a fim de preparar o ambiente para execução dos processos.

Durante sete dias consecutivos, aproximadamente no mesmo horário, em torno das 19 horas será iniciado os processos para que desta forma haja a simulação de algo próximo da rotina de um servidor. Vale ressaltar que os dados serão gravados após a finalização do processo de *backup* com a ferramenta Rsync e na sequencia inicia-se o mesmo processo com a ferramenta Rclone.

Com a definição dos dias e horários serão executados os *scripts* para gerar um *backup* dos arquivos salvos e definidos.

O *backup* definido acontecerá com arquivos, imagens e vídeos, iniciando com 1 mega byte, e chegando até 1 giga byte, por execução. O tamanho dos arquivos afeta o processo de *backup*, fazer *backup* de mil arquivos de 1 mega byte é mais custoso do que de apenas um arquivo de 1 giga byte (SOMASUNDARAM, 2009).

Os mesmos arquivos serão utilizados para ambas as ferramentas, para que os dados

¹ Segundo Jargas, Aurélio Marinho (2008, p. 24) *script* é uma lista de comandos para serem executados em sequência (um comando após o outro). Um roteiro predefinido de comandos e parâmetros.

compilados no processo sejam da mesma natureza e tamanho, evitando divergências em tempo devido ao arquivo.

Durante esses processos, serão analisados o consumo de processamento e de memória da máquina, com o auxílio do Htop (MUHAMMAD, 2004), e o tempo para a conclusão do processo com cada ferramenta.

2.4. Considerações Finais

O presente trabalho tem o objetivo de comparar o desempenho das ferramentas Rsync e Rclone, no âmbito de rotinas de backup, possibilitando dessa forma a análise para o melhor aproveitamento do servidor. Em primeiro momento, não se destacam possíveis dificuldades. Durante o estudo imagina-se alcançar êxito no processo avaliativo.

Com a análise dos dados coletados, pretende-se criar subsídios a outros usuários e desta forma otimizar o tempo no momento de escolha das ferramentas para a necessidade e manutenção dos backups.

2.5. Planejamento do Trabalho

Atividades	TCC1			TCC2			
	Jun	Jul	Ago	Set	Out	Nov	Dez
1. Revisão dos apontamentos da banca		x					
2. Revisão bibliográfica	x	x	x	x	x	x	x
3. Escrita da fundamentação teórica	x	x	x				
4. Redação do projeto de TCC	x	x	x				
5. Defesa do projeto de TCC		x					
6. Implementação da metodologia proposta				x	x		
7. Escrita da monografia de TCC					x	x	
8. Elaboração da apresentação final						x	
9. Defesa final do TCC							x

2.6. Recursos Necessários:

- Computador;
- ISO Linux Debian;

- Máquina Virtual (Oracle VM Virtualbox);
- Ferramenta Rsync (gratuita);
- Ferramenta Rclone (gratuita);
- Ferramenta Htop (gratuita);
- Libre Office Calc (para criação da tabela comparativa).

2.7. Horário de Trabalho

Horário	SEG	TER	QUA	QUI	SEX	SAB
18h - 19h			Orientações			
19h - 20h			Orientações	TCC	TCC	TCC
20h - 21h	TCC	TCC	TCC	TCC		
22h - 23h	TCC	TCC	TCC			

3. REFERÊNCIAS

CRAIG-WOOD, N. **Rclone documentation**. 2014. Disponível em: < <https://rclone.org/> >. Acesso em: 26 março de 2021.

Dicionário Aurélio De Língua Portuguesa. Editora Positivo, 5° ed. Brasil, 2010.

GOOGLE. **Google Drive**. 2012. Disponível em: <<https://www.google.com/intl/pt-BR/drive/>>. Acesso em: 25 de março de 2021.

IBM CLOUD EDUCATION. **Backup and Restore**. 2019. Disponível em: <<https://www.ibm.com/cloud/learn/backup-and-restore>>. Acesso em: 12 de março de 2021.

JARGAS, A. M. **Shell Script Professional**. 1. ed. São Paulo, SP, Brasil, p. 24, 2008.

MUHAMMAD, H. **HTOP**. 2004. Disponível em: < <https://htop.dev/> >. Acesso em: 11 de maio de 2021.

MURDOCK, I. **Linux Debian**. 1993. Disponível em: < <https://www.debian.org/> >. Acesso em: 11 de maio de 2021.

SOMASUNDARAM, G. **Armazenamento e gerenciamento de informações: como armazenar, gerenciar e proteger informações**. Tradução: Acauan Pereira Fernandes; revisão técnica: EMC Brasil. – Dados eletrônicos. – Porto Alegre, 2011.

TRIDGELL, A.; MACKERRAS, P. **Rsync manual pages**. 2009. Disponível em: <<https://www.samba.org/ftp/rsync/rsync.html>>. Acesso em: 26 de março de 2021.