Análise de uma estratégia para alta disponibilidade de *clusters* CEPH

Gustavo V. Mitraud[1], Calebe P. Bianchini[1,2]

[1]Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP — Brasil

[2]CESAR – Centro de Estudos e Sistemas Avançados do Recife
Recife, PE – Brasil

gustavo.mitraud@mackenzista.com.br, calebe.bianchini@mackenzie.br,

cpb@cesar.org.br

Resumo. *Ambientes de Computação de Alto Desempenho (CAD) demandam sistemas de armazenamento com disponibilidade ininterrupta e alta resiliência a falhas. Embora sistemas distribuídos apresentem recursos de redundância, a aplicação prática e a validação experimental de replicações em múltiplos locais geográficos ou lógicos (multisite) para recuperação de desastres ainda representam uma lacuna na literatura. Para endereçar esse cenário, este trabalho propõe a implementação e validação experimental de uma estratégia de alta disponibilidade baseada na replicação multisite de clusters CEPH por meio de seu portal de acesso aos dados, denominada RAGOS Gateway Multi-Site Replication. A pesquisa adota abordagem mista, exploratória e aplicada, combinando métodos qualitativos e quantitativos em um estudo de caso com três nós Ubuntu Server 24.04 LTS executando instâncias CEPH completas. Estes foram organizados em zonas primária, secundária e terciária sob um balanceador de carga com verificações proativas de integridade (health checks ativos). Testes preliminares foram conduzidos por 24 horas com operações de leitura e escrita contínuas, além de desligamentos aleatórios de nós, produzindo disponibilidade de 99,9985%, Tempo Médio de Recuperação (MTTR) de 2,33 segundos e Tempo Médio Entre Falhas (MTBF) de 4,80 horas. Os resultados validam a eficácia da arquitetura proposta para ambientes CAD que demandam tolerância a falhas em armazenamento de objetos.*

1. Introdução

A crescente dependência de ambientes de Computação de Alto Desempenho (CAD) elevou os requisitos de disponibilidade e resiliência dos sistemas de armazenamento. Como interrupções podem gerar perdas operacionais severas, a adoção de estratégias robustas de Disaster Recovery (DR) deixou de ser opcional e passou a ser um requisito fundamental para infraestruturas críticas [Tamimi et al. 2019].

O armazenamento distribuído, como o CEPH (software-defined storage de código aberto), tornou-se a solução predominante devido ao seu custo-benefício e capacidade de escalonamento. Embora o sistema ofereça redundância nativa, muitas implantações convencionais operam em um único cluster, o que cria um ponto único

de falha em cenários de desastre de grande escala, como a perda de um data center inteiro.

Esse cenário evidencia uma lacuna crítica: apesar do CEPH suportar replicação multisite (entre diferentes locais) via RADOS Gateway (RGW), a validação experimental dessa funcionalidade em estratégias de DR para CAD ainda é escassa. Autores como Enrico et al. [Bocchi, Enrico et al. 2024] descrevem a aplicação bem-sucedida dessa tecnologia no CERN, garantindo disponibilidade contínua mesmo durante falhas catastróficas, mas a literatura carece de testes sistemáticos adicionais em ambientes controlados.

Diante disso, este trabalho propõe a implementação e validação experimental de uma estratégia de alta disponibilidade baseada na replicação multisite de clusters CEPH por meio do RGW, denominada RAGOS Gateway Multi-Site Replication.

2. Fundamentação Teórica

Disaster Recovery (DR) é o conjunto de políticas, ferramentas e procedimentos que permitem a retomada de sistemas e dados após uma interrupção não planejada. No contexto de infraestruturas de TI críticas, dois indicadores são centrais para mensurar a eficácia de uma estratégia de DR: o *Recovery Time Objective* (RTO), que define o tempo máximo aceitável para restauração de um serviço após uma falha, e o *Recovery Point Objective* (RPO), que determina a quantidade máxima de dados que pode ser perdida em função do tempo decorrido desde o último estado consistente. Shubhashis et al. [Sengupta and Annervaz 2014] destacam a importância desses parâmetros na definição de políticas de replicação entre múltiplos centros de dados, propondo um modelo de planejamento que considera RTO e RPO como variáveis centrais na tomada de decisão arquitetural.

Já o CEPH é um sistema de armazenamento projetado para oferecer desempenho, confiabilidade e escalabilidade sem pontos únicos de falha. Sua arquitetura é estruturada sobre o RADOS (*Reliable Autonomic Distributed Object Store*), o serviço de base responsável por operações de replicação, recuperação e balanceamento de dados de forma autônoma [Weil et al. 2007]. Weil et al. [Weil et al. 2006] destacam que o algoritmo CRUSH (*Controlled Replication Under Scalable Hashing*) permite uma distribuição eficiente de dados sem necessidade de metadados centralizados, conferindo ao CEPH a escalabilidade horizontal e a resiliência estrutural sobre as quais os demais componentes operam.

3. Metodologia, Implementação e Experimentação

O ambiente experimental é composto por três nós rodando Ubuntu Server 24.04 LTS, cada um executando uma instância completa do CEPH com os seguintes serviços devidamente configurados: gerência do mapa do sistema (MON - *Monitor*), armazenamento físico dos dados (OSD - *Object Storage Daemon*), gerenciamento de métricas (MGR - *Manager*) e portal de acesso (RGW). Todos os nós foram implementados como máquinas virtuais sobre o Proxmox VE, uma plataforma corporativa de virtualização de código aberto, visando o controle e reprodutibilidade do experimento.

As três instâncias estão organizadas segundo a hierarquia *multisite* do RGW: uma zona primária (instância principal que processa primariamente as operações de escrita), uma zona secundária e uma zona terciária. Todas pertencem ao mesmo *Realm* (domínio global que define os limites da replicação) e *Zone Group* (agrupamento lógico de zonas que compartilham configurações e metadados), garantindo a sincronização automática de dados entre os três *clusters* [Bocchi, Enrico et al. 2024].

O teste de carga e resiliência foi conduzido ao longo de um período contínuo de 24 horas. Durante toda a execução, um cliente realizou operações constantes baseadas no protocolo padrão de armazenamento em nuvem S3 sobre um compartimento lógico compartilhado, denominado *bucket*. Tais ações incluíram requisições de escrita, leitura, exclusão e listagem de objetos, simulando uma carga de trabalho representativa de um ambiente de produção em operação contínua. Nesse período, de forma aleatória, cada um dos três nós foi desligado por um período de 30 minutos. Após o desligamento e a posterior reativação de cada nó, aguardou-se um intervalo de uma hora antes que o próximo desligamento fosse realizado.

Esse protocolo foi desenhado para simular falhas reais e não planejadas de nós de armazenamento, avaliando a capacidade do sistema de manter a continuidade das operações S3 sem intervenção manual, bem como a velocidade de ressincronização dos dados após a reintegração de cada nó. As métricas coletadas foram registradas de forma automatizada ao longo de toda a janela de 24 horas, garantindo a rastreabilidade e a reprodutibilidade dos resultados. A Tabela 1 consolida os principais indicadores obtidos durante a execução do experimento.

Table 1. Indicadores obtidos durante os testes

Métrica	**Valor**
Disponibilidade	99,9985%
MTBF (*Mean Time Between Failures* - Tempo Médio Entre Falhas)	4,80 horas
MTTR (*Mean Time To Recovery* - Tempo Médio de Recuperação)	2,33 segundos
Total de Falhas Detectadas	5
Tempo Total de Indisponibilidade	11,65 segundos

A disponibilidade de 99,9985% ao longo de 24 horas de operação contínua posiciona a solução dentro da categoria conhecida como *five nines* (99,999%), referência de excelência para sistemas de missão crítica. O MTTR de 2,33 segundos indica que, em média, o balanceador de carga detectou a falha de cada nó e redirecionou o tráfego para as zonas disponíveis em menos de 3 segundos, sem necessidade de intervenção manual. O MTBF de 4,80 horas reflete o intervalo médio entre os eventos de falha simulados, compatível com o protocolo de desligamentos aleatórios adotado.

3.1. Considerações Finais

Este trabalho demonstrou que a estratégia RAGOS Gateway Multi-Site Replication é capaz de garantir alta disponibilidade, resiliência e tolerância a falhas em *clusters* CEPH com replicação em múltiplos locais geográficos ou lógicos. Os

resultados experimentais, em especial a disponibilidade de 99,9985% e o MTTR de 2,33 segundos ao longo de 24 horas, validam a eficácia da arquitetura proposta frente a cenários de interrupção não planejada. Esses achados corroboram a aplicabilidade da solução em ambientes de produção críticos e expandem a base empírica iniciada por estudos como o de Enrico et al. [Bocchi, Enrico et al. 2024] para um contexto metodológico controlado e reprodutível.

Como trabalhos futuros, propõe-se a replicação do experimento em *clusters* geograficamente distribuídos para avaliar o impacto da latência de rede na sincronização entre zonas, dimensão explorada por Uehara et al. [Uehara et al. 2018] em seu modelo de replicação com consciência geográfica.

Agradecimentos

Os autores agradecem o apoio da MackCloud, Laboratório Multidisciplinar de Computação Científica e Nuvem[1]; e do projeto SPRACE – Processo nº 2018/25225-9, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Este trabalho foi financiado em parte pelo Fundo Mackenzie de Pesquisa e Inovação (MackPesquisa) – Projetos nº 231009 e 251005.

References

Bocchi, Enrico, Lekshmanan, Abhishek, Valverde, Roberto, and Goggin, Zachary (2024). Enabling storage business continuity and disaster recovery with ceph distributed storage. *EPJ Web of Conf.*, 295:01021.

Sengupta, S. and Annervaz, K. (2014). Multi-site data distribution for disaster recovery—a planning framework. *Future Generation Computer Systems*, 41:53–64.

Tamimi, A. A., Dawood, R., and Sadaqa, L. (2019). Disaster recovery techniques in cloud computing. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pages 845–850.

Uehara, K., Chen, Y.-F. R., Hiltunen, M., Joshi, K., and Schlichting, R. (2018). Feasibility study of location-conscious multi-site erasure-coded ceph storage for disaster recovery. In *2018 IEEE International Conference on Cloud Engineering (IC2E)*, pages 204–210.

Weil, S. A., Brandt, S. A., Miller, E. L., Long, D. D. E., and Maltzahn, C. (2006). Ceph: a scalable, high-performance distributed file system. In *Proceedings of the 7th Symposium on Operating Systems Design and Implementation*, OSDI '06, page 307–320, USA. USENIX Association.

Weil, S. A., Leung, A. W., Brandt, S. A., and Maltzahn, C. (2007). Rados: a scalable, reliable storage service for petabyte-scale storage clusters. In *Proceedings of the 2nd International Workshop on Petascale Data Storage: Held in Conjunction with Supercomputing '07*, PDSW '07, page 35–44, New York, NY, USA. Association for Computing Machinery.

[1] `https://mackcloud.mackenzie.br`