



# **Relatório de Infraestrutura do Sistema MECRED**

## **Autores**

Richard Fernando Heise Ferreira  
Eduardo Todt

**UNIVERSIDADE FEDERAL DO PARANÁ**  
SETOR DE CIÊNCIAS EXATAS  
DEPARTAMENTO DE INFORMÁTICA  
CENTRO DE COMPUTAÇÃO CIENTÍFICA E  
SOFTWARE LIVRE

# Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Infraestrutura do Ambiente de Hardware</b>	<b>3</b>
2.1	Máquinas . . . . .	3
2.2	Indicadores de Desempenho do Banco de Dados . . . . .	5
2.3	Análise dos indicadores . . . . .	5
<b>3</b>	<b>Propostas para a Infraestrutura Futura</b>	<b>9</b>

# 1. Introdução

Este relatório descreve a infraestrutura de máquinas que sustenta o sistema MECRED. São apresentadas as máquinas dos softwares utilizados nas diferentes instâncias que compõem o projeto e asseguram seu funcionamento, com impacto direto no cotidiano da educação brasileira.

A elaboração deste documento está diretamente vinculada ao Objetivo 2, Meta 4: aprimorar a infraestrutura de rede e servidores, assegurando alta disponibilidade e acesso contínuo ao sistema, mesmo em períodos de pico de uso ou diante de falhas pontuais.; Objetivo 3, Meta 2: Implementar políticas de backup e recuperação de dados para garantir a integridade e a disponibilidade das informações armazenadas na plataforma e o Objetivo 3, Meta 3: Realizar estudos sobre técnicas avançadas de armazenamento de dados que garantam a escalabilidade e disponibilidade de sistemas em nuvem computacional, levando em consideração o aumento do volume de conteúdo e usuários.

Ao detalhar a arquitetura dos serviços e as configurações das máquinas envolvidas, busca-se evidenciar a robustez da solução adotada e embasar decisões técnicas que promovam a evolução consistente da infraestrutura. A análise também contempla a descrição do sistema de backups e da estratégia de redundância de dados atualmente em operação no laboratório. Também são apresentadas estatísticas do banco de dados que comprovam a escalabilidade e o desempenho do projeto ao longo do tempo.

## 2. Infraestrutura do Ambiente de Hardware

O ambiente do sistema MECRED é composto por três máquinas virtuais, cada uma desempenhando um papel específico dentro do ciclo de desenvolvimento, homologação e produção da aplicação. Além delas, ainda há máquinas dos serviços de busca por conteúdo e armazenamento de binários de dados. Por fim, há uma máquina distribuída de gerenciamento de banco de dados que garante a estabilidade, disponibilidade e velocidade do sistema.

### 2.1 Máquinas

#### 1. mecreddev

- **Descrição:** Ambiente de desenvolvimento do MECRED, destinado à implementação e testes de novas funcionalidades em fase experimental. Acessível somente através da rede interna do Departamento de Informática da UFPR.
- **CPU:** AMD EPYC 7401 Processor @ 2.0 GHz — 4 cores / 4 threads
- **Disco:** 100 GB
- **Memória:** 8 GB
- **Sistema:** Debian GNU/Linux 12 (bookworm)
- **Tipo:** Virtual
- **Backup:** Semanal, quintas, 03:00h.

#### 2. mecredhomologa

- **Descrição:** Ambiente de homologação do sistema, utilizado para validação interna das funcionalidades antes da publicação em produção. Pode ser acessada externamente, mas os dados acessados aqui são somente para teste (conceito de *mock data*).
- **CPU:** AMD EPYC 7401 Processor @ 2.0 GHz — 4 cores / 4 threads
- **Disco:** 100 GB
- **Memória:** 8 GB
- **Sistema:** Debian GNU/Linux 12 (bookworm)
- **Tipo:** Virtual
- **Backup:** Diário, 03:00h.

### 3. mecred

- **Descrição:** Ambiente de produção do MECRED, responsável por atender às requisições externas através do endereço <https://mecred.mec.gov.br>.
- **CPU:** AMD EPYC 7401 Processor @ 2.0 GHz — 8 cores / 8 threads
- **Disco:** 300 GB
- **Memória:** 62 GB
- **Sistema:** Debian GNU/Linux 12 (bookworm)
- **Tipo:** Virtual
- **Backup:** Diário, 02:00h. Possui snapshot.

### 4. mecdb4

- **Descrição:** Máquina hospedeira do DSpace.
- **CPU:** AMD EPYC 7401 Processor @ 2.0 GHz — 8 cores / 8 threads
- **Disco:** 2.5 TB
- **Memória:** 62 GB
- **Sistema:** Debian GNU/Linux 9 (stretch)
- **Tipo:** Virtual
- **Backup:** Diário, 03:00h.

### 5. elasticsearchupgrade

- **Descrição:** Máquina hospedeira do motor de busca Elasticsearch.
- **CPU:** AMD EPYC 7401 Processor @ 2.0 GHz — 16 cores / 16 threads
- **Disco:** 250 GB
- **Memória:** 32 GB
- **Sistema:** Debian GNU/Linux 11 (bullseye)
- **Tipo:** Virtual
- **Backup:** Semanal, quintas, 03:00h.

### 6. postgres - CASE

- **Descrição:** Máquina física de banco de dados.
- **CPU:** Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz - 20 cores / 20 threads
- **Disco:** 2 NVME 1,5TB - 4 Discos Sata 6TB
- **Memória:** 252 GB
- **Tipo:** Física
- **Sistema:** Debian GNU/Linux 12 (bookworm)

### 7. postgres - TARS (réplica)

- **Descrição:** Máquina física de banco de dados, réplica da CASE.

- **CPU:** Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz - 20 cores / 20 threads
- **Disco:** 2 NVME 1,5TB - 4 Discos Sata 6TB
- **Memória:** 252 GB
- **Tipo:** Física
- **Sistema:** Debian GNU/Linux 12 (bookworm)

O sistema de backup das máquinas é realizado em dois formatos distintos e conta ainda com replicação off-site, ou seja, em local físico diferente, a fim de garantir maior redundância e segurança dos dados. As máquinas virtuais que são preservadas por meio de snapshots permitem a rápida recuperação de falhas pontuais. Já os backups completos são destinados a cenários mais críticos.

É importante destacar que, ao longo dos 10 anos de operação do projeto MECRED, os backups precisaram ser acionados em apenas duas ocasiões, ambas decorrentes de erro humano. Em ambos os casos, a restauração do sistema foi realizada com sucesso e sem qualquer perda de dados. De forma geral, o sistema mantém uma disponibilidade anual média de aproximadamente 99,5%, o que corresponde a cerca de 1,83 dias de indisponibilidade por ano. Esse resultado representa um excelente indicador de alta disponibilidade e confiabilidade operacional, em termos de infraestrutura.

## 2.2 Indicadores de Desempenho do Banco de Dados

A seguir, as imagens apresentadas constataam dados relevantes referentes as máquinas de banco de dados com o PostgreSQL. Os resultados foram coletados ao longo do último mês. O nome do banco de produção é "portalmec" por razões legadas.

## 2.3 Análise dos indicadores

As Figuras 2.1 e 2.2 apresentam, respectivamente, as taxas de entrada e saída de rede da máquina Postgres - CASE, em megabits por segundo. Observa-se que não há indícios de gargalo de rede para os bancos de dados hospedados diretamente nessa máquina, uma vez que os valores registrados estão significativamente abaixo do limite da infraestrutura de rede do Departamento, que possui conexão interna de 20 Gb/s.

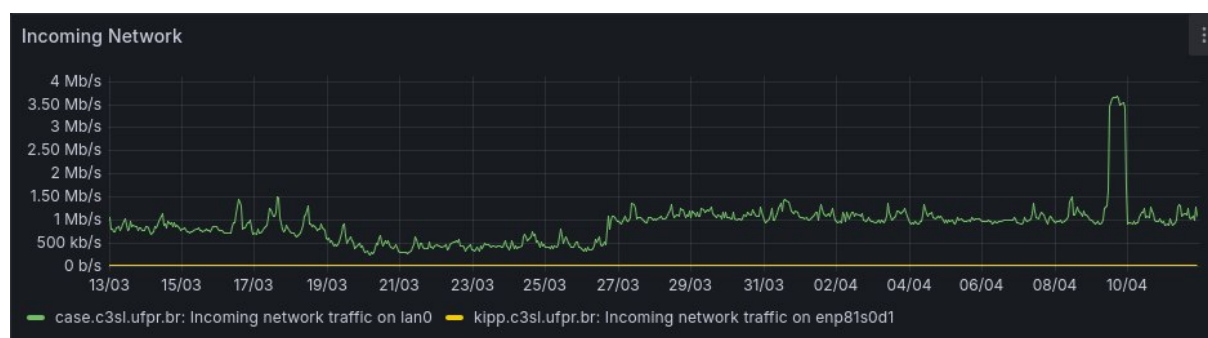


Figura 2.1: Taxa de entrada de rede na máquina Postgres - CASE.

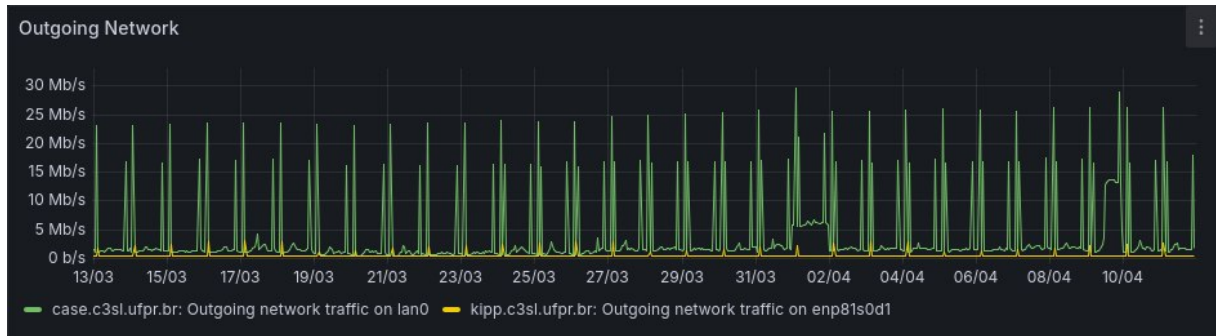


Figura 2.2: Taxa de saída de rede na máquina Postgres - CASE.

As Figuras 2.3, 2.4 e 2.5 fornecem uma visão geral dos recursos da máquina Postgres - CASE. O uso de CPU raramente ultrapassa 20%, com picos concentrados nos períodos de execução dos backups noturnos dos bancos com cópia diária. Em relação ao armazenamento, dos 27 terabytes disponíveis, apenas cerca de 6,24 gigabytes estão em uso, o que representa uma utilização extremamente baixa frente à capacidade total da máquina. Quanto à memória RAM, o uso aparente de aproximadamente 80% pode parecer elevado, mas, na prática, indica estabilidade e eficiência no gerenciamento dos recursos. Esse comportamento consistente ao longo do tempo, independentemente da carga de transações, é resultado de otimizações aplicadas, como o ajuste fino do PostgreSQL e melhorias no sistema de cache – o mesmo comportamento pode ser visto na figura 2.6, que revela como o uso da cache é excelente para o banco de dados do MECRED. A queda inicial no uso de memória observada no gráfico decorre justamente dessas melhorias, refletindo um uso mais racional e estável da RAM.

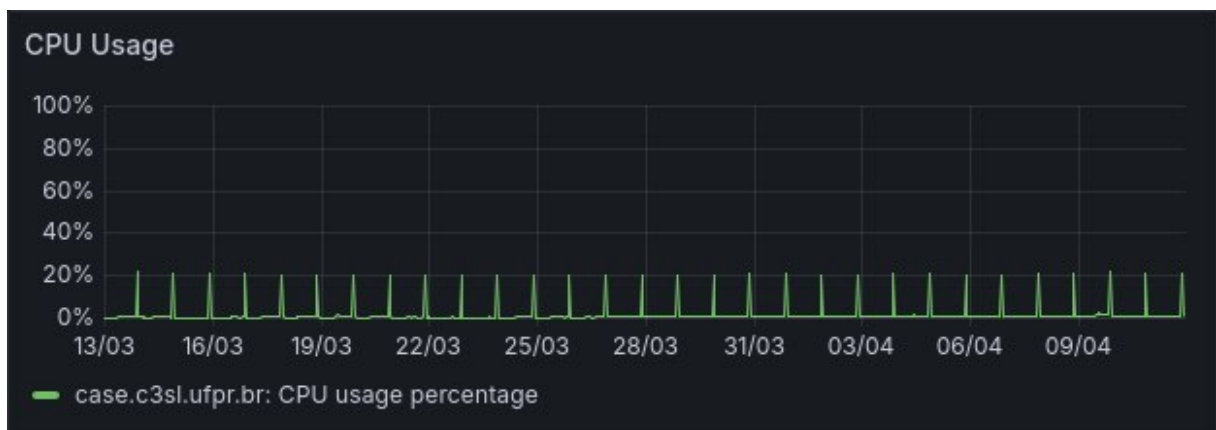


Figura 2.3: Uso de CPU na máquina Postgres - CASE.



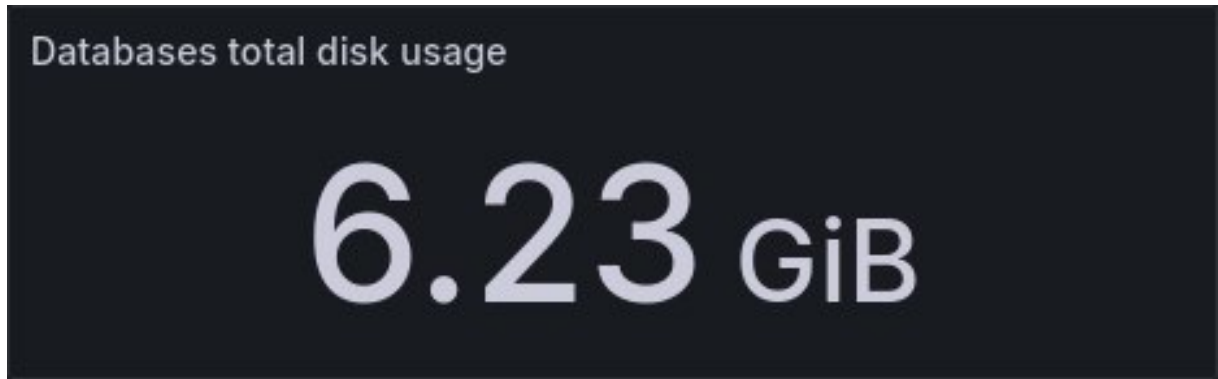


Figura 2.4: Uso de disco na máquina Postgres - CASE.



Figura 2.5: Uso de memória RAM na máquina Postgres - CASE.



Figura 2.6: Uso de cache na máquina Postgres - CASE e no banco portalmec.

A Figura 2.7 demonstra que o número de transações ao longo do mês manteve-se estável. Mesmo durante o pico observado, os componentes do banco de dados e do sistema responderam adequadamente, sem ocorrência de indisponibilidade, o que evidencia a resiliência da infraestrutura. Já a Figura 2.8 exibe as consultas mais lentas registradas. A mais demorada apresentou tempo de resposta de 722 milissegundos, seguida por 450 ms e 94 ms. Considerando que o tempo máximo aceitável para uma consulta é de 2 segundos, conclui-se que o sistema opera com folga nesse aspecto, não apresentando gargalos relevantes relacionados à latência de consultas.

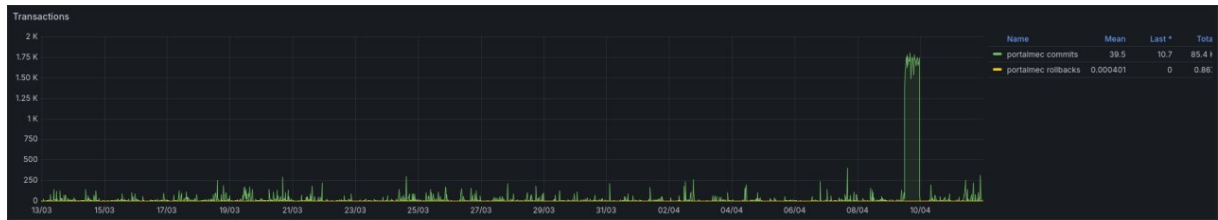


Figura 2.7: Número de transações por segundo no banco portalmec.

user	database	query	calls	exec_time	rows	plan_time
portalmecc	portalmecc	SELECT "activities".* FROM "a"	53	722	1590	0
portalmecc	portalmecc	SELECT COUNT(*) FROM "act"	53	450	53	0
portalmecc	portalmecc	SELECT COUNT(*) FROM "leai"	3936	94.4	3936	0
portalmecc	portalmecc	SELECT COUNT(*) FROM "act"	554	94.4	554	0
portalmecc	portalmecc	SELECT "activities".* FROM "a"	554	92.3	16620	0
portalmecc	portalmecc	SELECT COUNT(*) FROM "leai"	3936	86.2	3936	0
portalmecc	portalmecc	SELECT COUNT(*) FROM "dov"	3936	49.4	3936	0
portalmecc	portalmecc	SELECT COUNT(DISTINCT "us"	3936	31.1	3936	0
portalmecc	portalmecc	SELECT \$4 AS one FROM "use"	427	29.2	105	0

Figura 2.8: Consultas lentas registradas no banco portalmec.

Com base nos indicadores apresentados, conclui-se que a infraestrutura atual da máquina Postgres - CASE atende plenamente às demandas operacionais do sistema MECRED. Os gráficos analisados evidenciam uma utilização eficiente dos recursos computacionais, com ampla margem de desempenho tanto em rede, quanto em processamento, armazenamento e memória. As estratégias de otimização adotadas, como os ajustes finos no PostgreSQL e a melhoria do sistema de cache, têm se mostrado eficazes, garantindo estabilidade e desempenho consistente mesmo em momentos de maior carga. Dessa forma, os dados confirmam que a solução implantada é robusta, confiável e preparada para sustentar o crescimento futuro do sistema.

### 3. Propostas para a Infraestrutura Futura

A arquitetura atual já oferece uma base robusta, confiável e altamente disponível. No entanto, foram identificados alguns pontos que podem ser aprimorados. Primeiramente, embora a utilização de máquinas virtuais (VMs) seja suficiente e confiável, essa abordagem não reflete as práticas mais modernas de implantação adotadas nos padrões atuais da indústria. Com o objetivo de aumentar a escalabilidade do sistema e proporcionar uma formação discente alinhada com tecnologias de ponta, propõe-se a substituição das VMs por pods em ambiente Kubernetes — uma plataforma de orquestração de contêineres que automatiza a implantação, o escalonamento e a gestão de aplicações em contêineres. O pod é a menor unidade de implantação no Kubernetes e representa um grupo de um ou mais contêineres que compartilham o mesmo ambiente de execução, incluindo recursos de rede e armazenamento. Esses contêineres são executados em conjunto, o que permite comunicação eficiente e gerenciamento unificado. Os pods serão provisionados a partir de imagens Docker — um formato padronizado e portátil para empacotar aplicações e suas dependências — integradas ao repositório Harbor, um gerenciador de imagens Docker que permite controle de acesso, escaneamento de vulnerabilidades e replicação entre registros.

Outro gargalo identificado refere-se ao uso do DSpace, conforme detalhado no relatório específico sobre limitações da arquitetura. Planeja-se substituir o DSpace pelo serviço Simple Storage Service (S3), voltado ao armazenamento de dados binários (como vídeos, PDFs, entre outros). A implementação ocorrerá sobre o Ceph, uma solução de armazenamento distribuído altamente escalável, utilizando o RADOS Gateway (radosgw), que expõe a interface compatível com S3 sobre o núcleo de objetos do Ceph. Essa mudança tornará o sistema de armazenamento em nuvem mais escalável e resiliente, adotando uma tecnologia amplamente consolidada, bem documentada e compatível com as necessidades do projeto.

Em relação ao Elasticsearch, trata-se de um sistema distribuído que conta com a integração do Kibana para geração de relatórios, monitoramento e visualização de dados. Atualmente, todos os nós encontram-se na mesma máquina virtual. Embora essa arquitetura ainda atenda satisfatoriamente à demanda de busca, ela será revista futuramente, caso novas necessidades de desempenho ou escalabilidade surjam.

Com as mudanças propostas aqui, é possível afirmar que o sistema MECRED estará alinhado ao que há de mais moderno e principal em termos de infraestrutura técnica e disponibilidade de serviço.