2025-11-07 Post-Morten indisponibilidade Web

Relatório Pós-Incidente (Post-Mortem): Indisponibilidade de Serviços Web (07/11/2025)

1. Introdução

Na noite de **07 de Novembro de 2025**, um evento meteorológico extremo — incluindo tornados no Paraná, entre eles um Tornado escala F3 — gerou um interesse público sem precedentes pelos serviços do Simepar. O tráfego ao site público (Simepar) atingiu um pico histórico de mais de 150.000 usuários ativos ao longo daquele dia,conforme dados do Google Analytics (Anexo A). É importante esclarecer que esta métrica representa o número total de visitantes únicos que acessaram o site durante o período de 24 horas do dia 07/11. A plataforma Google Analytics não armazena o histórico de *usuários simultâneos*, tornando impossível recuperar a contagem de picos concorrentes de dias passados.

Esse volume massivo de acessos foi direcionado inicialmente ao balanceador de carga (nginx01 e nginx02), que registrou picos superiores a 5.200 conexões simultâneas (Anexo B), sobrecarregando as máquinas virtuais responsáveis pelo atendimento ao site público.

A análise técnica demonstra que a infraestrutura de virtualização (Host KVM) não foi a causa da falha. O equipamento físico kvm-dell-01 permaneceu estável, com uso reduzido de recursos e grande disponibilidade de memória (GB livres) durante todo o evento (Anexo C).

A falha ocorreu no nível da aplicação e das VMs. A sobrecarga extrema de requisições esgotou os processos da linguagem PHP, chamada PHP-FPM (Anexo D), que atingiram o limite de workers e filas internas. Isso levou ao consumo total da memória interna das VMs (vmprognozweb01/02/03) passou vmapache4. As primeiras foram isoladas da restante o mais breve possível (resultando no a falta de acesso ao site público mais prolongado. Esse colapso no S.O, ocasionou corrupção do disco virtual, exigindo reparo manual (fsck) e gerando:

- Indisponibilidade completa do site público,

CLIMATOS

- Quando o Redis atingiu o limite máximo de memória, ele deixou de armazenar novos dados em cache, fazendo com que as requisições passassem a ser direcionadas diretamente ao banco de dados da vmprognoz. Esse comportamento resultou em uma sobrecarga do banco, evidenciada pelo grande número de conexões ativas oriundas do IP da vmprognozweb01, o que acabou provocando indisponibilidade para os clientes.
- Para resolver o problema, foi necessário interromper temporariamente o serviço do Nginx, o que impactou o site público e alguns produtos que dependiam desse serviço para funcionar.
- Após a identificação da interrupção do serviço do Nginx também afetou outros produtos de clientes, o serviço foi restabelecido gradualmente para esses produtos e mantido isolado apenas para o site público.

É fundamental ressaltar, no entanto, que os sistemas operacionais críticos internos, utilizados pela equipe de meteorologistas para monitoramento, geração de alertas e comunicação com a Defesa Civil, não foram afetados e permaneceram 100% operacionais durante todo o evento.O acesso alternativo foi enviado para a equipe de meteorologistas

A equipe de TI atuou imediatamente em duas frentes simultâneas:

- Isolamento

Os serviços críticos dos clientes foram rapidamente isolados para impedir que a falha do site público comprometesse sistemas corporativos. A rota de atendimento foi ajustada para garantir estabilidade e continuidade.

- Contenção

Foi implementado um mecanismo emergencial de controle de tráfego (Rate Limiting) no Nginx, reduzindo o volume de requisições recebidas e permitindo que as VMs recuperadas permanecessem operacionais sem entrar novamente em colapso.

A partir da normalização do ambiente, foi iniciado um plano de ação corretivo e preventivo, com foco em:

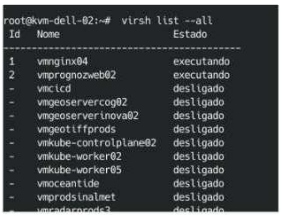
- **Implementação de CDN (Content Delivery Network):** para absorver picos de tráfego e reduzir a carga sobre as VMs internas;
- **Revisão dos limites de PHP-FPM e da arquitetura do front-end**, visando aumentar a resiliência do stack de aplicação;

- **Aprimoramento da infraestrutura**, especialmente no que se refere ao provisionamento de memória das VMs, para suportar picos futuros.

Esse conjunto de medidas visa assegurar robustez, disponibilidade e capacidade de resposta em cenários de grande demanda pública, especialmente durante eventos meteorológicos de alta criticidade.

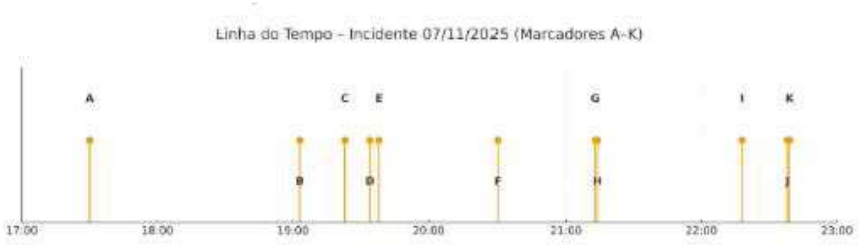
2. Linha do Tempo do Incidente (07/11/2025)

Horário (BRT)	Evento	Impacto / Ação
~17:30(A)	Gatilho	Início de pico de tráfego anômalo (Google Analytics) devido às tempestades no Paraná.
19:03(B)	Alerta	Site público (Simepar) fica indisponível (Erro 502).
19:23(C)	Diagnóstico	Equipe de TI identifica VMs (vmprognozweb01) travadas e sobrecarga no banco de dados. A figura abaixo ilustra o estado das vms na kvm02- que sofreu reboot na tentativa de recuperar a vmprognozweb02 que exibia em “execução” mas sem conexão.

		 Algumas vms precisaram subir uma a uma
19:34(D)	Mitigação	Ação 1 (Isolamento):Serviços de clientes (exemplo CopelGeT, InfoHidro, VFogo) são isolados para protegê-los da falha. Internamente, como visualização do Radvis e StormBoard continuaram funcionando.
19:38(E)	Diagnóstico	VMs vmprognozweb entram em ciclo de falha (sobem e caem imediatamente).
20:30(F)	Diagnóstico	Serviços dependentes (radar/geotiff) param devido à falha das VMs de processamento.
21:13(G)	Recuperação	Causa raiz do travamento das VMs é identificada: corrupção do sistema de arquivos

		devido ao colapso interno.
21:14(H)	Recuperação	Início do reparo manual (fsck) nas máquinas virtuais afetadas.
22:18(I)	Recuperação	VMs (vmprognozweb01/02/03) são reparadas e voltam ao ar. Exemplo vmprognozweb01 em execução: 
22:38(J)	Mitigação	Ação 2 (Contenção): Implementado controle de tráfego (limit_req) no Nginx para proteger a aplicação.
22:39(K)	Estabilização	Site público volta ao ar. O sistema é estabilizado, rejeitando o tráfego excedente, mas mantendo o serviço operacional.

Em formato de linha do tempo:



A — Início do pico de tráfego anômalo

Tornado categoria F3 gera acesso massivo (+150k usuários), disparando a sobrecarga ao Nginx.

B — Site público indisponível (Erro 502)

Primeira queda detectada pelo monitoramento e por usuários externos.

C — Identificação de VMs travadas e DB sobrecarregado

Análise inicial aponta travamento da vmprognozweb01 e queries presas.

D — Mitigação 1: Isolamento dos serviços de clientes

Rotas são ajustadas para impedir que o problema do site público afete clientes.

E — Falhas consecutivas no cluster

VMs reiniciam e caem em loop devido à corrupção interna e esgotamento de PHP-FPM.

F — Paralisação de serviços dependentes

Radar/geotiff e outros fluxos internos param por dependência das VMs afetadas.

G — Causa raiz identificada

Corrupção de sistema de arquivos dentro das VMs causada por colapso interno.

H — Início do reparo manual (fsck)

Equipe executa varredura e correção de disco virtual em todas as VMs.

I — VMs restauradas

vmprognozweb01, 02 e 03 retornam e o cluster volta a responder.

J — Mitigação 2: Contenção aplicada

Rate limiting no Nginx impede nova sobrecarga e estabiliza o ambiente.

CLIMATOS

K — Estabilização final

Site público retorna ao ar; tráfego excedente é rejeitado de forma controlada.

3. Análise da Causa Raiz (Diagnóstico Técnico)

1. A falha ocorreu em cascata, iniciando pela camada de aplicação e evoluindo até comprometer a estabilidade das VMs envolvidas, embora a infraestrutura física tenha continuado operando normalmente durante o evento.
2. **Gatilho (Tráfego):** O volume de mais 150 mil usuários (Anexo A) atingiu o balanceador nginx01 (Anexo B), que repassou a carga para as VMs de aplicação.
3. **Gargalo (Aplicação PHP):** As VMs (vmapache4, vmprognozweb02) receberam um fluxo de requisições muito acima do limite configurado para o PHP-FPM. Os logs indicam saturação progressiva: *[pool www] seems busy*. O sistema aumentou o número de processos até atingir a capacidade de memória disponível dentro da VM (Anexo D).
4. **Colapso (Sistema Operacional da VM):** Sem recursos de memória suficientes, o sistema operacional passou a utilizar intensamente o disco virtual (swap), resultando em elevação brusca de *context switches* e I/O interno (Anexo E). Isso levou ao congelamento das VMs, que deixaram de responder inclusive via SSH.
5. **Falha (Corrupção de Disco):** Diante do congelamento, foi necessário utilizar *virsh destroy* para interrupção forçada. Isso provocou inconsistências no sistema de arquivos das VMs, exigindo execução de *fsck* para reparo. O gráfico de memória da vmprognozweb02 (Anexo F) destaca o momento exato da interrupção.
6. **Infraestrutura (Host KVM):** Os indicadores do host kvm-dell-01 (Anexo C) permaneceram estáveis durante o incidente, sem registrar saturação de recursos, o que permitiu que o foco da análise se mantivesse na camada de aplicação e nas VMs afetadas.

4. Plano de Ação (Ações Corretivas e Preventivas)

Para garantir a estabilidade futura dos serviços e a capacidade de resposta a eventos de alto interesse público, as seguintes medidas estão sendo implementadas:

CLIMATOS

A. Curto Prazo (Contenção Imediata)

1. **Implementação de CDN (Content Delivery Network):**
   - **Justificativa:** É a ação mais crítica. Uma CDN servirá o conteúdo do site público a partir de servidores distribuídos globalmente, absorvendo 99% do tráfego de picos como este, antes que ele atinja nossa infraestrutura.
   - **Status:** Em cotação e implementação emergencial.
2. **Manutenção do Controle de Tráfego (limit_req):**
   - **Justificativa:** Servirá como uma segunda camada de proteção para garantir que, mesmo que o tráfego passe pela CDN, ele não sobrecarregue a aplicação.
3. Replicação base de dados do prognoze adição camada de balanceamento de conexões.

- **Justificativa:** Atualmente, o ambiente opera com um único servidor PostgreSQL (versão 9.5), sem nós de réplica e sem camada de distribuição de conexões. Em cenários de aumento abrupto de acessos, essa topologia concentra toda a carga em um único ponto, limitando a capacidade de resposta

B. Médio Prazo (Resiliência da Aplicação)

1. Dimensionamento de Recursos das VMs:
   - Justificativa: As VMs de aplicação (vmapache4, prognozweb) precisam de mais RAM e CPU alocados para que o S.O. tenha margem para picos, evitando o colapso por I/O Wait interno.
   - Status: Em planejamento.
2. Otimização do Pool PHP-FPM:
   - Justificativa: Continuar o *tuning* do pm = dynamic para encontrar o balanço ideal entre uso de memória e capacidade de resposta a picos.
   - Status: Em andamento.

C. Longo Prazo (Revisão de Arquitetura)

1. **Isolamento Físico de Cargas de Trabalho:**
   - **Justificativa:** O site público (alto tráfego, baixa criticidade) não deve operar no mesmo hardware físico que os serviços de clientes (tráfego estável, alta criticidade). Uma falha no site público não pode, sob nenhuma circunstância, impactar os serviços contratados.

CLIMATOS

Conclusões

O incidente ocorrido em 07 de novembro resultou de um evento excepcional, impulsionado por um pico de tráfego sem precedentes — mais de 150 mil usuários ativos simultâneos — motivado pelo forte interesse público diante das tempestades e do Tornado F3 que atingiu o Estado.

A análise técnica realizada pela equipe de TI indica que a origem da falha esteve concentrada na camada de aplicação, diante de um volume de requisições muito acima do projetado. Esse aumento repentino saturou os processos de atendimento (PHP-FPM) nas máquinas virtuais responsáveis pelo site público, levando ao esgotamento dos recursos internos dessas VMs e, consequentemente, ao congelamento de seus sistemas operacionais. Esse cenário exigiu intervenções manuais de reparo (por exemplo *fsck*) para sua recuperação.

A infraestrutura física de virtualização permaneceu operacional durante todo o evento, permitindo que a atuação da equipe se concentrasse nas VMs afetadas e na recuperação dos serviços.

A resposta operacional foi imediata. A equipe de TI executou duas ações críticas de contenção para estabilizar o ambiente:

- **Isolamento dos serviços dos clientes**, garantindo a continuidade dos sistemas corporativos;
- **Aplicação de controle de tráfego (Rate Limiting)** no Nginx, reduzindo o impacto das novas requisições sobre as VMs restauradas.

Com base nos aprendizados deste evento extremo, um plano de evolução da arquitetura está em andamento. Como prioridade, destaca-se a implementação de uma CDN (Content Delivery Network), que atuará como primeira camada de absorção de tráfego, reduzindo drasticamente o impacto de picos futuros sobre o ambiente interno. Além disso, o dimensionamento e a configuração das VMs de aplicação serão revisados para ampliar a tolerância a cargas excepcionais.

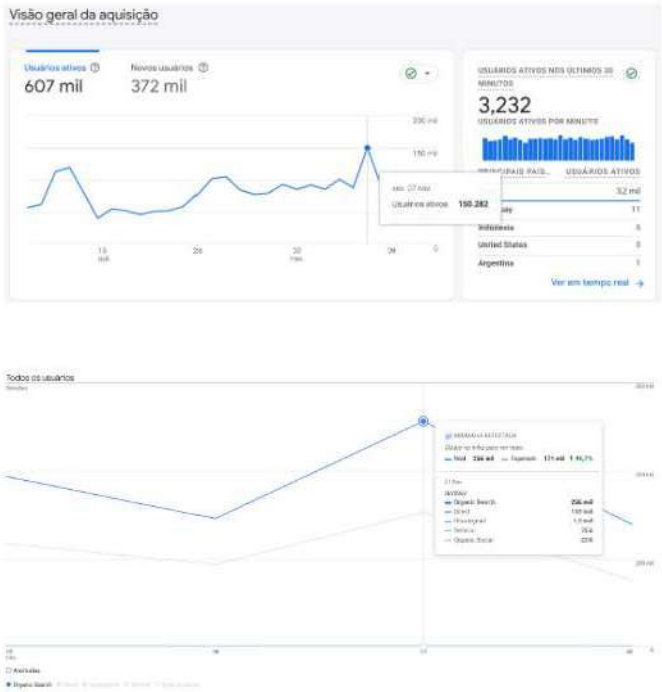
O plano de ação estabelecido busca não apenas correção, mas evolução da infraestrutura, garantindo resiliência, escalabilidade e segurança operacional frente a novos eventos meteorológicos de grande impacto.

CLIMATOS

Anexos (Gráficos de Evidência)

Anexo A: Pico de Tráfego (Google Analytics)

Usuários ativos nos últimos 7 dias, com pico dia 07/11/2025:

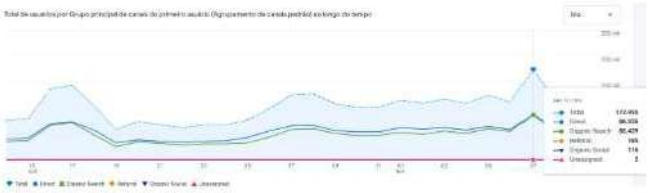


Alerta de anomalia detectada

Demonstra o volume de usuários que acessaram o site- o gatilho do incidente. Mais abaixo detalhamento (real e esperado):

CLIMATOS

Mais uma evidência do alto número de usuários acessando o site:

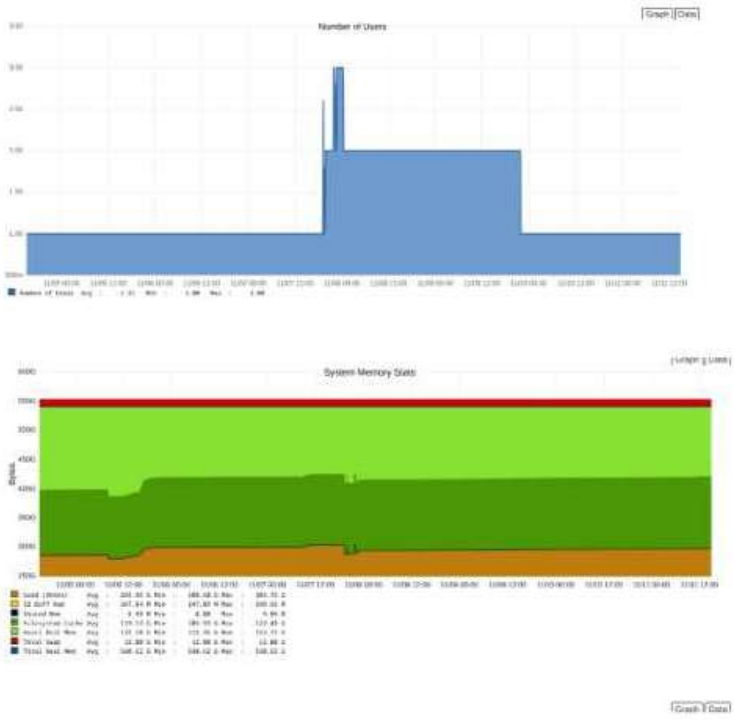


Anexo B: Pico de Conexões (Load Balancer nginx01) *Mostra o impacto na "porta de entrada", com o Nginx recebendo mais de 5.200 conexões.*

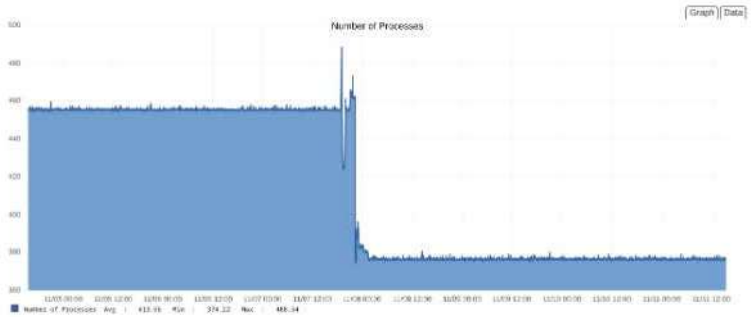
CLIMATOS

Anexo C: Saúde do Host Físico (kvm-dell-01) *Prova de que a infraestrutura (host) estava saudável, com 548G de RAM e recursos de memória (verde) sobrando durante o incidente.*

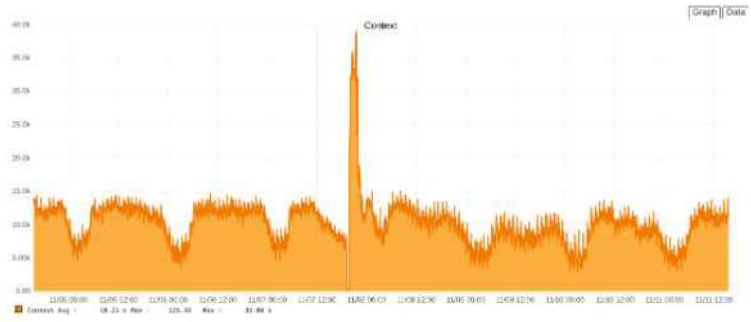
CLIMATOS



Anexo D: Sobrecarga da Aplicação (vmapache4 - Processos) Aqui mostra *o PHP-FPM* tentando criar mais processos (trabalhadores) para lidar com o tráfego, atingindo o limite da VM.Obs: esta máquina onde o site dos clientes foram direcionados, teve configurações do PHP e REDIS alteradas.



Anexo E: Colapso da VM (vmapache4 - Context Switch) Evidência do colapso interno da *VM.* Mostra a CPU gastando tempo trocando de tarefas (39.8k) em vez de trabalhar.



Anexo F: Falha da VM (vmprognozweb02 - Memória) *Mostra a linha de dados da VM sendo interrompida (a lacuna) durante o colapso, provando o período de indisponibilidade.*

CLIMATOS

Documento revisado por Ana,Daniel,Areco

CLIMATOS