

EDIÇÃO 01 · SETEMBRO 2026

O preço unitário caiu. A conta subiu.

O estado da computação em setembro de 2026: a hora de GPU, a memória, o hypervisor, a inferência, o código que os agentes escrevem, o Redata e a IA fora do inventário.

EM RESUMO

Sete leituras de um mês em que tudo ficou mais barato por unidade e mais caro no total.

Cada leitura traz o mecanismo, o número com quem mediu e quando, e o que muda quando a infraestrutura é medida a partir da máquina.

- 01 **A hora de GPU virou commodity, com volatilidade de commodity.** A mediana da H100 está em US\$ 3,38, contra mais de US\$ 7 em 2024; o contrato de um ano subiu 40% entre outubro e março; para a mesma placa, a diferença entre blocos chega a seis vezes.
LEITURA 01
-
- 02 **A memória reprecificou o servidor.** DRAM convencional de 58% a 63% mais cara em um trimestre e DRAM de servidor de 13% a 18% no seguinte. Os contratos de longo prazo protegem os grandes compradores; o reajuste vai para quem compra sem contrato.
LEITURA 02
-
- 03 **Renovar hypervisor deixou de ser carimbo.** 93% dos 302 líderes de TI ouvidos mudaram a estratégia de VMware desde a aquisição; 86% esperam novo reajuste; a Gartner projeta 35% das cargas de trabalho migrando até 2028.
LEITURA 03
-
- 04 **O token vai custar 90% menos, e a conta de IA já subiu.** Só 6% das 1.719 empresas ouvidas pela McKinsey veem impacto significativo da IA no lucro operacional; 20% já se dizem limitadas pelo custo de operar IA.
LEITURA 04
-
- 05 **O código gerado por IA custa duas vezes.** Três a quatro vezes mais commits, dez vezes mais achados de segurança, 22,7% dos problemas nunca corrigidos e 45% do código com vulnerabilidade conhecida.
LEITURA 05
-
- 06 **O Redata passou no Senado; a fatura de quem opera não muda.** R\$ 5,2 bilhões de renúncia em 2026, com 10% da capacidade reservada ao mercado interno.
LEITURA 06
-
- 07 **A violação de dados de 2026 tem IA fora do inventário.** Incidentes com shadow AI passaram de 20% para 43% das violações estudadas pela IBM; custo médio de US\$ 5 milhões, alta de 12%.
LEITURA 07
-

OS NÚMEROS DO MÊS

Seis números que resumem o mês.

Nenhum é medição nossa. Cada um traz quem mediu e quando.

US\$ 3,38

Mediana do preço da hora de H100 sob demanda, em 69 provedores.

AIMultiple, Cloud GPU Rental Price Index · 25/08/2026

58% a 63%

Alta projetada no preço de contrato da DRAM convencional só no segundo trimestre de 2026.

TrendForce · 31/03/2026

93%

Líderes de TI de grandes empresas que mudaram a estratégia de VMware desde a aquisição pela Broadcom.

CloudBolt, 302 respondentes · janeiro de 2026

6%

Empresas que atribuem impacto significativo da IA ao lucro operacional, entre 1.719 executivos ouvidos.

McKinsey, The State of AI in 2026 · 25/08/2026

22,7%

Problemas introduzidos por código de IA que continuam presentes na versão mais recente do repositório.

arXiv 2603.28592, 302,6 mil commits · 26/04/2026

R\$ 5,2 bi

Renúncia fiscal estimada pelo governo com o Redata em 2026, aprovado pelo Senado.

Senado Federal · 01/09/2026

A mesma GPU custa até seis vezes mais conforme o balcão.

O preço caiu pela metade em dois anos e ficou volátil no caminho. Quem contratou como se fosse tarifa fixa está pagando a diferença.

A hora de GPU é hoje o insumo de computação com o preço mais público e menos comparado do mercado. O índice da AIMultiple, atualizado em 25 de agosto, acompanha 69 provedores e 17 modelos de GPU. A mediana da H100 sob demanda está em torno de US\$ 3,38 por hora, contra mais de US\$ 7 no início de 2024. Para a mesma placa, o preço de tabela dos grandes provedores de nuvem costuma ficar de três a seis vezes acima da menor oferta dos neoclouds.

A queda não veio em linha reta. A SemiAnalysis, que pesquisa mensalmente mais de 100 participantes do mercado, registrou a alta do contrato de um ano: de US\$ 1,70 por hora em outubro de 2025 para US\$ 2,35 em março de 2026, quase 40% em seis meses. Em abril, o contrato de um mês aparecia esgotado no painel. Barato e sem vaga, ao mesmo tempo.

Do lado de quem vende, a capacidade é financiada com dívida. A CoreWeave, o maior neocloud de capital aberto, fechou junho com US\$ 35 bilhões de dívida total, US\$ 640 milhões de juros em um único trimestre, mais do que o dobro do registrado um ano antes, e um plano de US\$ 35 a 39 bilhões de investimento para 2026.

A resposta racional de quem compra é travar preço: uma reserva de três anos, para tirar a volatilidade do orçamento. É essa resposta que produz a perda. A reserva fixa o preço de 2024 sobre uma hora que o mercado vende hoje pela metade, e fixa também a quantidade, que ninguém confrontou com o uso real.

EXIBIÇÃO 01

A hora de H100 caiu pela metade e, no contrato de um ano, voltou a subir 40%.

Preço por GPU-hora, em dólares. Cada linha compara dois momentos medidos pela mesma fonte.



Fontes: AIMultiple, Cloud GPU Rental Price Index, 25/08/2026 (mediana de 69 provedores); SemiAnalysis, H100 1-Year Rental Price Index, 02/04/2026.

NA PRÁTICA

Uma reserva de três anos assinada em 2024 paga hoje perto do dobro da mediana de mercado pela mesma hora, e ninguém no comitê sabe dizer quanto dela está em uso.

A memória reprecificou o servidor, e o reajuste vai para quem não tem contrato.

Os fabricantes deslocaram capacidade para HBM e para servidor de IA. Quem compra servidor comum em 2026 disputa wafer com a IA e perde.

A TrendForce projetou, em 31 de março, alta de 58% a 63% no preço de contrato da DRAM convencional só no segundo trimestre de 2026, com os fabricantes deslocando capacidade para HBM e para memória de servidor. Em 9 de julho, a mesma casa projetou mais 13% a 18% para a DRAM de servidor no terceiro trimestre e descreveu o mecanismo: os contratos de longo prazo dos grandes provedores de nuvem impedem o fornecedor de reajustar para esses clientes, e a alta passa a vir de quem compra sem contrato e do volume vendido fora deles.

Os fabricantes de servidor foram os primeiros a repassar. Em dezembro, Dell, Lenovo, HP e HPE preparavam reajustes de cerca de 15% em servidores, e o COO da Dell resumiu na teleconferência de resultados: a demanda está muito à frente da oferta. A consultoria Context, citada pela The Register em janeiro, mediu alta média de 63% nos preços dos fabricantes de DRAM entre setembro e dezembro de 2025, e de 28% na RAM enterprise vendida pelos fabricantes de servidor.

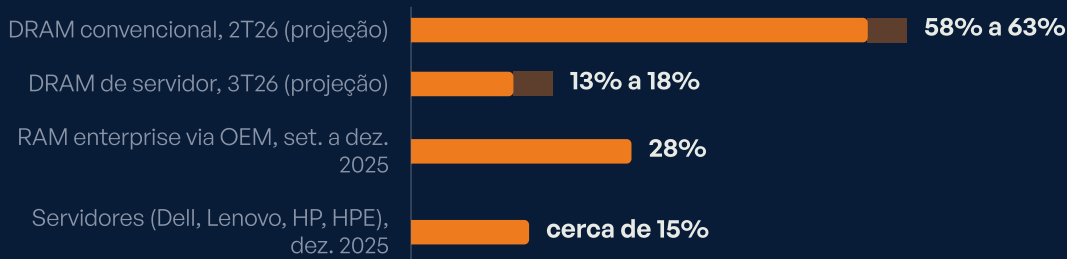
Não é uma alta de ciclo: é uma realocação. A TrendForce registra em julho que o consumidor chegou ao limite do que consegue pagar e que a DRAM de servidor segue subofertada.

A resposta racional é antecipar a compra. É a resposta que enche o rack de memória reservada e não usada, no exato momento em que memória é o componente mais disputado do mercado.

EXIBIÇÃO 02

Cada elo da cadeia repassou uma alta diferente, e a maior ficou com quem compra sem contrato.

Variação de preço no período indicado. Cada barra tem a própria base e o próprio período; elas não se somam.



Fontes: TrendForce, 31/03/2026 e 09/07/2026; Context via The Register, 14/01/2026; The Register, 04/12/2025.

NA PRÁTICA

O mesmo pedido de compra, cotado de novo em setembro, custa mais e demora mais, enquanto os racks já pagos seguem com memória reservada e ninguém sabe dizer quanto dela está em uso.

A renovação de VMware obrigou a abrir o inventário.

A pesquisa mostra o susto. O que ela não mostra é o que o time encontra quando a proposta chega.

Desde que a Broadcom comprou a VMware e levou o portfólio para assinatura por core, renovar o hypervisor deixou de ser um carimbo. A CloudBolt ouviu 302 líderes de TI de grandes empresas em janeiro de 2026: 93% mudaram a estratégia de VMware ao menos uma vez desde a aquisição, 63% mudaram duas ou mais vezes, 86% seguem preocupados com o próximo reajuste e 41% relatam mais pressão da diretoria. A Gartner projetou, no Symposium de setembro de 2025, que a VMware perde 35% das cargas de trabalho até 2028.

O que a pesquisa não mostra é o que acontece quando a proposta chega. O time abre o inventário para saber o que roda em cada host e descobre que o inventário tem dez anos e nunca correspondeu ao que roda na máquina: máquinas virtuais órfãs, snapshots de 2019, clusters dimensionados para sistemas já desligados. A renovação é calculada por core e cobra também pelos cores que só existem na planilha.

A resposta racional é negociar desconto. É a resposta que renova o mesmo desperdício com preço novo, porque o desconto incide sobre uma contagem que ninguém conferiu.

O reajuste fez o que nenhuma auditoria interna tinha conseguido: tornou impossível renovar no piloto automático. Quem aproveitou o susto para medir a própria infraestrutura a partir da máquina saiu da mesa com um número. Quem só negociou saiu com um desconto sobre um número errado.

EXIBIÇÃO 03

Nove em dez líderes de TI já mudaram de estratégia, e quase todos esperam o próximo reajuste.

Percentual de 302 líderes de TI de grandes empresas, pesquisa de janeiro de 2026.



Fonte: CloudBolt, pesquisa com 302 líderes de TI, janeiro de 2026.

NA PRÁTICA

A proposta de renovação chega com a contagem de cores tirada de um inventário de dez anos. A leitura feita a partir da máquina encontra os hosts que só existem na planilha, e eles saem da conta antes da assinatura.

O token vai custar 90% menos. A conta de IA já subiu.

As duas coisas são verdade ao mesmo tempo, e a pesquisa de agosto mostra quem está pagando a diferença.

A Gartner projetou em 25 de março que, até 2030, rodar inferência em um modelo de 1 trilhão de parâmetros vai custar aos provedores mais de 90% menos do que custou em 2025. O analista que assina a projeção avisa na mesma nota que o barateamento libera aplicações mais caras, principalmente agentes, que gastam mais tokens por tarefa: «those applications are going to be more expensive, not less» (essas aplicações vão custar mais, não menos).

A McKinsey mediu o resultado. O State of AI de 2026 ouviu 1.719 executivos: 80% dizem que a produtividade individual melhorou; 40% das empresas com receita acima de US\$ 1 bilhão já escalam agentes, contra 27% no ano anterior; 37% atribuem algum impacto no lucro operacional à IA, a mesma fatia de 2025; 6% conseguem chamar esse impacto de significativo. Outros 20% afirmam que o custo de operar IA já limita o quanto usam.

Um em cada cinco está limitado pelo custo, com o token mais barato da história. É o mecanismo da nuvem, dez anos depois: o preço unitário cai, o consumo explode e quase ninguém atribui o custo a uma tarefa concluída. A Gartner estima, desde junho de 2025, que mais de 40% dos projetos de agentes serão cancelados até o fim de 2027, por custo crescente, valor pouco claro ou controle de risco insuficiente.

A licença do modelo é a parte pequena da conta. A parte grande é a computação que os agentes consomem em inferência, orquestração e chamadas repetidas, diluída em linhas que ninguém liga a um resultado.

EXIBIÇÃO 04

80% relatam ganho de produtividade individual. Só 6% veem impacto significativo no lucro.

Percentual de respondentes, The State of AI in 2026, 1.719 executivos. A barra em destaque é a que o comitê de orçamento vai cobrar.



Fonte: McKinsey, The State of AI in 2026, via The Register, 25/08/2026.

NA PRÁTICA

O piloto do agente custa por mês; o comitê pergunta quanto custa por tarefa concluída. A diferença entre as duas contas mora na computação que ninguém atribuiu.

O código gerado por IA chega quatro vezes mais rápido e fica.

Três estudos mediram o custo que aparece nos relatórios de segurança. Nenhum mediu o custo de rodar esse código todo dia.

A Apiiro analisou dezenas de milhares de repositórios de empresas da Fortune 50 entre dezembro de 2024 e junho de 2025: desenvolvedores assistidos por IA entregam commits em um ritmo de três a quatro vezes maior, e os achados de segurança mensais saltaram de cerca de 1.000 para mais de 10.000, com alta de 322% em caminhos de escalonamento de privilégios e de 153% em falhas de arquitetura.

Um estudo acadêmico publicado em abril (arXiv 2603.28592) seguiu 302,6 mil commits de autoria verificada de IA em 6.299 repositórios, de cinco assistentes diferentes. Mais de 15% dos commits de cada assistente introduzem ao menos um problema, e 22,7% dos problemas continuam presentes na versão mais recente do repositório. Ninguém voltou para consertar.

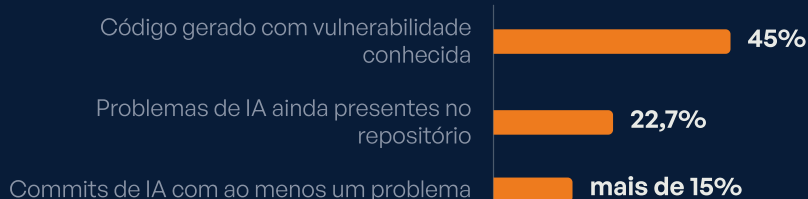
A Veracode testou mais de 150 modelos em 80 tarefas e quatro linguagens: 45% do código gerado carrega uma vulnerabilidade conhecida, a mesma proporção de dois anos atrás, e o tamanho do modelo tem efeito mínimo. O rastreador da Georgia Tech contou 6, 15 e 35 CVEs atribuídos a ferramentas de código em janeiro, fevereiro e março de 2026.

Esse é o custo que aparece nos relatórios. O que não aparece é o custo de rodar: a função gerada às pressas, com a mesma query repetida e a alocação que ninguém revisou, passa na revisão e executa milhões de vezes por dia. A conta chega ao orçamento de infraestrutura, longe de quem escreveu.

EXIBIÇÃO 05

Quase metade do código gerado sai com vulnerabilidade, e um quinto dos problemas nunca é corrigido.

Percentuais de três medições independentes, cada uma com a própria base.



Fontes: Veracode, 24/03/2026 (150+ modelos, 80 tarefas); Liu et al., arXiv 2603.28592, 26/04/2026 (302,6 mil commits, 6.299 repositórios).

NA PRÁTICA

Uma função gerada em minutos, com a mesma consulta ao banco dentro de um loop, passa no teste e vai para produção. O custo dela aparece como um servidor a mais no trimestre seguinte, sem nome de função na fatura.

O Brasil aprovou o Redata. A fatura de quem opera não muda.

O incentivo desonera quem constrói o data center. O serviço mal dimensionado continua mal dimensionado, só que mais perto.

Em 1º de setembro, o Senado aprovou o PL 278/2026, que cria o Regime Especial de Tributação para Serviços de Datacenter, o Redata: suspensão de Imposto de Importação, IPI, PIS/Cofins e PIS/Cofins-Importação para quem montar data center no país, com foco em nuvem e inteligência artificial. O governo estima renúncia em torno de R\$ 5,2 bilhões em 2026 e de R\$ 1 bilhão em cada um dos dois anos seguintes. O texto foi enviado à sanção presidencial; a MP 1.318/2025, que criava o regime, havia caducado sem votação.

As contrapartidas estão no texto: energia renovável ou de baixa emissão, eficiência hídrica de até 0,05 litro por kWh no resfriamento, investimento no país de 2% do valor dos equipamentos e ao menos 10% da capacidade de processamento reservada ao mercado interno (8% e 1,6% no Norte, Nordeste e Centro-Oeste).

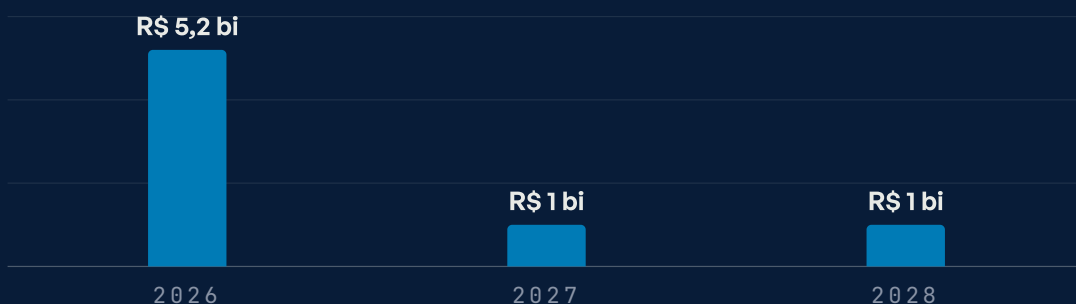
Dez por cento. O incentivo desonera quem constrói, e a demanda que esses prédios vão atender pode vir, em até 90%, de fora do país. A fatura de nuvem de uma empresa brasileira não cai porque um galpão novo passou no Congresso.

Capacidade nova favorece quem sabe do que o próprio serviço precisa e consegue levá-lo para onde estiver mais barato: nuvem pública, nuvem privada ou edge. Quem não sabe vai pagar o mesmo preço por um servidor mais perto. O Redata é política industrial, e política industrial não corrige serviço mal dimensionado. Só oferece um lugar novo para dimensionar mal.

EXIBIÇÃO 06

A renúncia concentra-se em 2026 e cai a um quinto nos dois anos seguintes.

Renúncia fiscal estimada pelo governo com o Redata, em bilhões de reais por ano.



Texto aprovado pelo Senado em 1º de setembro e enviado à sanção. Contrapartidas e prazos podem mudar com veto.

Fonte: Senado Federal, Senado Notícias, 01/09/2026; Agência Brasil, 01/09/2026.

NA PRÁTICA

O serviço superdimensionado que hoje roda em uma região estrangeira passa a rodar superdimensionado a 40 km da sede, com a mesma folga de capacidade paga todo mês.

LEITURA 07 • A IA FORA DO INVENTÁRIO

A violação de dados de 2026 tem um processo de IA que ninguém inventariou.

Duas políticas corretas produzem o buraco entre elas. A parte que mais cresce da infraestrutura mora nesse buraco.

A IBM publicou em 29 de julho o Cost of a Data Breach Report 2026, que estudou 602 organizações que sofreram violações entre março de 2025 e fevereiro de 2026, em 17 setores e 16 países. O custo médio global da violação chegou a US\$ 5 milhões, alta de 12% sobre a edição anterior.

O que mudou de um ano para o outro foi a origem. Os incidentes envolvendo shadow AI, ferramentas de IA adotadas sem conhecimento da TI ou da segurança, mais do que dobraram: de 20% para 43% das violações estudadas. Mais de dois terços das organizações não tinham processo de governança para limitar esse uso; entre as atacadas por meio dos próprios modelos, 92% não controlavam o acesso a essas ferramentas; só quatro em dez limitam quem acessa os sistemas de IA. Ataques de inversão de modelo e injeção de prompt custaram, em média, cerca de US\$ 6 milhões.

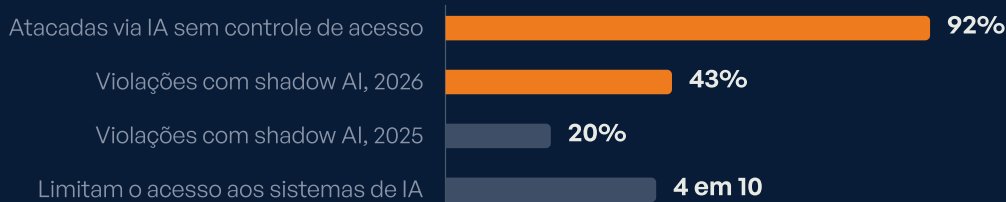
Nada disso chega pelo procurement. A pilha de IA chega como um processo em uma máquina que a empresa já paga, iniciada por uma equipe de dados que precisava testar uma ferramenta sem esperar seis semanas de compras, e recebe uma credencial de produção no caminho. A varredura de segurança, corretamente limitada ao inventário, não a vê. A fatura, corretamente atribuída ao centro de custo, a vê e não a nomeia.

A leitura que contém o não declarado começa na computação e sobe: o que roda em cada nó, o que ele alcança e quanto custa, declarado ou não.

EXIBIÇÃO 07

A shadow AI dobrou de peso nas violações, e nove em dez organizações atacadas via IA não controlavam o acesso.

Percentual das 602 organizações violadas estudadas pela IBM entre março de 2025 e fevereiro de 2026.



Fonte: IBM, Cost of a Data Breach Report 2026, 29/07/2026, via Cybersecurity Dive.

NA PRÁTICA

Um nó de GPU aprovado em abril para previsão de demanda roda também um construtor de agentes que alguém iniciou em março, com credencial de banco de dados viva e fora de qualquer lista de varredura.

O QUE FAZER COM ISSO

Três listas curtas, por quem lê.

Nenhuma pede uma compra. Todas pedem um número que hoje não existe na fatura.

SE VOCÊ ASSINA A FATURA

- 01 **Peça a utilização real** da capacidade reservada de GPU e de memória, medida na máquina, antes de renovar qualquer reserva de três anos. Sem esse número, negocia-se desconto, não quantidade.
- 02 **Exija o custo por tarefa concluída** de cada piloto de agente, e não o custo mensal. É a única base em que escalar e cancelar viram decisões comparáveis.
- 03 **Trate o Redata como opção de localização**, não como redução de custo. Ele só muda a conta de quem já sabe do que o próprio serviço precisa.

SE VOCÊ OPERA A INFRAESTRUTURA

- 01 **Refaça o inventário a partir da máquina** antes da renovação de hypervisor: hosts, cores em uso, snapshots, máquinas virtuais sem dono. A contagem de cores é a base do preço.
- 02 **Meça o consumo de memória por serviço** e reduza reservas acima do uso sob carga. Cada gigabyte liberado é um gigabyte que ninguém precisa comprar a preço de escassez.
- 03 **Leia a computação de baixo para cima**: processos por nó, o que cada um alcança, quanto custa, declarado ou não. É a única leitura que contém a IA fora do inventário.

SE VOCÊ COMPRA OU RENOVA NESTE TRIMESTRE

- 01 **Cote a hora de GPU em mais de um balcão** e compare com a mediana pública, US\$ 3,38 em agosto, antes de assinar. Para a mesma placa, a diferença entre provedores chega a seis vezes.
- 02 **Confira quanto da memória já paga está em uso** antes de aprovar servidor novo. O pedido de compra encolhe ou desaparece.
- 03 **Meça o consumo de CPU por função** nos releases com código gerado por IA, em comparação com o release anterior. O custo de rodar não aparece em nenhum relatório de segurança.



METODOLOGIA E FONTES

Como este relatório é feito.

Todo número deste relatório vem de uma fonte publicada, com quem mediu e quando, e nenhum é medição nossa. Lemos a fonte primária sempre que ela estava acessível. Quando um site bloqueou a leitura automática (gartner.com, mckinsey.com, ibm.com), o texto foi conferido em duas coberturas independentes, e a fonte primária está citada mesmo assim. Projeções aparecem como projeções, com a casa que as fez. Nenhuma ressalva foi cortada para caber.

Números lidos e deixados de fora, por virem de partes interessadas ou não terem fonte primária localizável, ficam no registro de produção da edição, junto com os arquivos de origem.

01	Senado Federal, Senado Notícias, 01/09/2026, «Incentivo para instalação de data centers no Brasil é aprovado pelo Senado». PL 278/2026. https://www12.senado.leg.br/noticias/materias/2026/09/01/incentivo-para-instalacao-de-data-centers-no-brasil-e-aprovado-pelo-senado
02	Agência Brasil, 01/09/2026, «Senado aprova projeto que cria incentivos para data centers no Brasil». https://agenciabrasil.ebc.com.br/politica/noticia/2026-09/senado-aprova-projeto-que-cria-incentivos-para-data-centers-no-brasil
03	AIMultiple, Cloud GPU Rental Price Index, atualizado em 25/08/2026. 69 provedores, 17 modelos de GPU, mediana de medianas. https://aimultiple.com/gpu-index
04	SemiAnalysis, 02/04/2026, «The Great GPU Shortage: Rental Capacity», e o painel público H100 Daily & Contract Pricing Index (abril de 2026). https://newsletter.semianalysis.com/p/the-great-gpu-shortage-rental-capacity
05	Capacity, 12/08/2026, «CoreWeave's debt hits \$35bn: what it means for the neocloud refinancing wall». Balanço de 30/06/2026. https://capacityglobal.com/news/coreweaves-debt-hits-35bn/
06	TrendForce, 31/03/2026, «AI Server Demand to Drive Memory Contract Price Increases in 2Q26 as CSPs Secure Supply via Long-Term Agreements». https://www.trendforce.com/presscenter/news/20260331-12995.html
07	TrendForce, 09/07/2026, «Long-Term Agreements Cap Price Increases; Server DRAM Contract Prices Expected to Rise 13-18% QoQ in 3Q26». https://www.trendforce.com/presscenter/news/20260709-13140.html
08	TrendForce, 03/07/2026, «AI Server Demand Continues to Support Memory Prices in 3Q26, but Gains Moderate as Consumer Demand Weakens». https://www.trendforce.com/presscenter/news/20260703-13134.html
09	The Register, 04/12/2025, «Server prices set to jump 15% on memory cost spike». Fala de Jeff Clarke, COO da Dell, em teleconferência de resultados. https://www.theregister.com/2025/12/04/server_prices_15_percent_jump_memory_costs/
10	The Register, 14/01/2026, «DRAM price hike to hit server and infrastructure costs». Dados da consultoria Context (Gurvan Meyer). https://www.theregister.com/on-prem/2026/01/14/dram-price-hike-to-hit-server-and-infrastructure-costs/5166984
11	CloudBolt, janeiro de 2026, pesquisa com 302 líderes de TI de grandes empresas, «Broadcom VMware Renewal 2026: What to Do in the 12 Months Before It Hits». https://www.cloudbolt.io/blogs/broadcom-vmware-renewal-timeline/



METODOLOGIA E FONTES • CONTINUAÇÃO

- 12 SDxCentral, 11/09/2025, «VMware to lose third of workloads to hyperscalers: Gartner». Julia Palmer, Gartner Symposium.
<https://www.sdxcentral.com/news/vmware-to-lose-third-of-workloads-to-hyperscalers-gartner/>

- 13 Gartner, 25/03/2026, «Gartner Predicts That by 2030, Performing Inference on an LLM With 1 Trillion Parameters Will Cost GenAI Providers Over 90% Less Than in 2025». Texto conferido em CIO Dive, 25/03/2026.
<https://www.ciodive.com/news/ai-inference-costs-drop-2030-gartner/815725/>

- 14 Gartner, 25/06/2025, «Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027».
<https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>

- 15 McKinsey, The State of AI in 2026, 1.719 respondentes. Cobertura de The Register, 25/08/2026, «McKinsey says enterprise AI is finally 'on the road to ROI'».
<https://www.theregister.com/ai-and-ml/2026/08/25/mckinsey-says-enterprise-ai-is-finally-on-the-road-to-roi/5292388>

- 16 Apiiro, «4x Velocity, 10x Vulnerabilities: AI Coding Assistants Are Shipping More Risks», e The Register, 05/09/2025.
<https://apiiro.com/blog/4x-velocity-10x-vulnerabilities-ai-coding-assistants-are-shipping-more-risks/>

- 17 Liu, Widyasari, Zhao, Irsan, Chen e Lo, «Debt Behind the AI Boom: A Large-Scale Empirical Study of AI-Generated Code in the Wild», arXiv 2603.28592, v2 de 26/04/2026.
<https://arxiv.org/abs/2603.28592>

- 18 Veracode, 24/03/2026, «Spring 2026 GenAI Code Security Update».
<https://www.veracode.com/blog/spring-2026-genai-code-security/>

- 19 Cloud Security Alliance, 04/04/2026, «Vibe Coding's Security Debt: The AI-Generated CVE Surge». Rastreador Vibe Security Radar, Georgia Tech.
<https://labs.cloudsecurityalliance.org/research/csa-research-note-ai-generated-code-vulnerability-surge-2026/>

- 20 IBM, Cost of a Data Breach Report 2026, 29/07/2026. Cobertura de Cybersecurity Dive, «As data breaches grow costlier, ungoverned AI creates new risks».
<https://www.cybersecuritydive.com/news/data-breach-costs-ai-governance-ibm/826463/>

A ÚNICA AÇÃO DESTE RELATÓRIO

Avaliar leva 48 horas e não muda uma linha de código.

Medimos a infraestrutura inteira a partir da máquina: o que cada recurso custa, o que roda nele, onde está exposto. Nuvem pública, nuvem privada e edge, em uma medida só, com um número de economia pelo qual você pode nos cobrar.

Não é um piloto: é uma entrega fechada.

thebunnylab.com/enterprise/br

O The Bunny Lab for Enterprise torna legível a computação de grandes organizações e age sobre o que encontra. Nosso produto é o FlexSuite, uma plataforma de compute intelligence em cinco estágios: Avaliar, Modernizar, Direcionar, Prever, Implantar. Nada é arrancado: nos integramos à pilha que o cliente já roda. Preço por nó, atrelado a resultados.

The Compute Report é publicado todo mês. Edição 02: outubro de 2026.

