

Manifestação técnica complementar sobre a estimativa dos quantitativos

Processo administrativo nº 23432.000601.2026-85

Pregão eletrônico nº 34/2026

1 Objeto e metodologia

Em atendimento ao apontamento da Procuradoria, apresentamos a metodologia e a memória de cálculo dos quantitativos destinados à infraestrutura local e compartilhada de pesquisa em Inteligência Artificial do IFSP Campus Catanduva. O dimensionamento combina previsão de demanda, execução simultânea de projetos, capacidade de armazenamento e quantidade de conexões necessária à arquitetura.

A estimativa é prospectiva. O atendimento futuro a até 10 grupos constitui capacidade desejada de utilização ao longo do tempo, mediante compartilhamento e agendamento. Não se pressupõe que os 10 grupos estejam atualmente constituídos ou utilizem a infraestrutura simultaneamente. A previsão operacional informada pela coordenação é de execução de três a quatro projetos simultâneos, conforme o tamanho das cargas e os modelos utilizados.

A metodologia adota: a) até quatro alocações simultâneas de um nó por projeto, quando suficientes; b) alocação de múltiplos nós para cargas maiores, reduzindo a simultaneidade; c) estimativa inicial de 6,5 TB para dados, modelos, checkpoints e resultados; d) orçamento de armazenamento de backup com versões; e) dimensionamento de enlaces, portas, cabos e reservas a partir da topologia.

2 Documentos e informações de suporte

Foram utilizados os documentos “02.01. TR - V3.docx”, especialmente a tabela do item 1.1, as especificações dos itens 1 a 8, a topologia e a seção 3.4, e “02. Edital.docx”, quanto ao objeto e ao julgamento por item. O TR V3 estabelece a integração final e a configuração distribuída pelo IFSP/CEPIA².

Complementam esses documentos as informações de planejamento prestadas pela coordenação do CEPIA² e registradas nesta manifestação: capacidade futura desejada de até 10 grupos; execução de mais de dois e até quatro projetos simultâneos; uso de LLMs com elevada demanda de memória; datasets de grande volume e checkpoints; e destinação dos HDDs ao backup dos grupos atendidos. A política de retenção e o espelhamento descritos adiante são premissas propostas nesta complementação e aceitas pela coordenação.

Não se utiliza histórico de contratação ou série de utilização como fundamento, pois esses registros não foram apresentados. Os 6,5 TB são uma estimativa inicial de planejamento, sem medição ou decomposição por dataset. A manifestação não atribui aos documentos originais informações que foram definidas nesta complementação.

3 Dimensionamento dos quatro nós

O quantitativo de quatro nós decorre do cenário de até quatro projetos simultâneos, com uma unidade por projeto quando tecnicamente suficiente. Quantidade menor reduz essa capacidade de atendimento. Para aplicações com modelos de linguagem de maior demanda, os recursos poderão ser alocados a um único projeto em múltiplos nós, conforme suporte do software e características da carga.

Memória de cálculo: 4 projetos × 1 nó por projeto = 4 nós no cenário de cargas independentes. Capacidade mínima prevista: 4 × 128 GB = 512 GB de memória utilizável pelos aceleradores, agregada e distribuída; armazenamento local: 4 × 2 TB = 8 TB agregados. Esses cenários são alternativos e não devem ser somados: um projeto distribuído pode reduzir o número de projetos concorrentes.

A soma de memória não representa automaticamente um único espaço de memória acessível por qualquer aplicação. A referência do TR a modelos da ordem de 70 bilhões de parâmetros é indicativa do tipo de pesquisa e não garante execução de qualquer treinamento ou configuração desse porte.

4 Memória de cálculo por item

Item e componente	Qtd.	Composição e finalidade
1 Nós de IA	4	Até quatro projetos simultâneos com um nó por projeto, ou alocação de múltiplos nós a cargas maiores.
2 Switch de alta velocidade	1	Interligação central dos quatro nós com quatro enlaces de 200 GbE.
3 Switch auxiliar 10 GbE	1	Rede compartilhada com quatro conexões dos nós e duas interfaces de dados do NAS.
4 Cabos breakout	3	4 enlaces ÷ 2 enlaces por cabo = 2 cabos em operação + 1 reserva.
5 Patch cords Cat6A	7	4 conexões dos nós + 2 conexões do NAS + 1 reserva.
6 NAS	1	Repositório central compartilhado para os nós e grupos atendidos.
7 SSDs de 4 TB	5	4 unidades instaladas + 1 reserva técnica fria.
8 HDDs de 20 TB	2	Duas unidades para backup em RAID 1; uma mantém o espelho da outra.

Os dois cabos breakout em operação utilizam duas portas de 400 GbE e produzem quatro enlaces de 200 GbE. O TR exige quatro portas físicas de 400 GbE no switch; a justificativa das portas adicionais deve permanecer coerente com o planejamento de expansão. O switch auxiliar exige oito portas RJ-45, das quais seis correspondem às conexões de dados explicitamente calculadas, além das interfaces SFP+ previstas para integração institucional.

Uma reserva de cada componente substituível evita manter um conjunto completo sobressalente. O cabo breakout de reserva pode substituir a conexão de dois nós; o patch cord pode substituir uma conexão dos nós ou do NAS; o SSD de reserva fria permite substituição local de uma unidade do arranjo. A reserva fria não integra a capacidade em operação nem constitui substituição automática.

5 Armazenamento compartilhado

A estimativa inicial de 6,5 TB considera datasets de grande volume, modelos, checkpoints de treinamento e resultados. Não constitui volume medido. Sua evolução será acompanhada conforme a entrada de projetos e a política de retenção adotada, evitando duplicação desnecessária de datasets compartilhados.

Quatro SSDs de 4 TB oferecem 16 TB brutos. Em RAID 5, a capacidade nominal após paridade é $(4 - 1) \times 4 \text{ TB} = 12 \text{ TB}$, antes de formatação e reservas operacionais. A quinta unidade permanece como reserva fria. O mínimo de 8 TB úteis exigido pelo TR é um piso de capacidade disponibilizada, e não o resultado nominal desse arranjo.

Na capacidade útil mínima, $8 - 6,5 = 1,5$ TB de diferença para crescimento e operação, equivalente a 18,75% de 8 TB. Esse cálculo é distinto do limite de ocupação de 80% proposto para os HDDs de backup. O armazenamento local dos nós possui finalidade operacional e não deve ser contado como capacidade do repositório compartilhado.

6 Política de retenção

Categoria	Critério inicial adotado
Datasets originais e códigos	Preservar durante o projeto, com identificação de versões e evitando duplicação de datasets compartilhados.
Checkpoints	Reter os dois mais recentes, o melhor segundo a métrica do projeto e o modelo final, sem duplicar arquivos coincidentes.
Modelos e resultados finais	Preservar por cinco anos após o encerramento, observadas exigências específicas do projeto e de tratamento dos dados.
Versões de backup	Versões diárias por sete dias e semanais por quatro semanas, com cópias incrementais quando suportadas.
Ocupação dos HDDs	Limite planejado de 80% da capacidade útil efetivamente disponibilizada, com revisão da retenção e da capacidade quando necessário.

As versões de backup e os arquivos finais têm horizontes distintos. Os arquivos finais permanecem no conjunto protegido durante seu prazo de preservação; a janela de versões do backup não autoriza sua eliminação. A revisão da política deverá observar os compromissos de cada pesquisa e as condições de tratamento dos dados.

7 Dimensionamento do backup em dois HDDs

Os dois HDDs de 20 TB destinam-se ao backup dos grupos atendidos. Adota-se RAID 1 para manter uma cópia espelhada entre os discos: $2 \times 20 \text{ TB} = 40 \text{ TB}$ brutos e 20 TB nominais após espelhamento, antes de formatação. O arranjo não oferece 40 TB úteis.

Para o planejamento inicial, reserva-se orçamento equivalente a duas vezes o conjunto de dados protegido: uma parcela para a cópia corrente e outra para alterações e versões. Trata-se de premissa de capacidade, não de garantia de que toda a política de versões ocupará exatamente esse volume. O consumo dependerá da taxa de alteração, dos checkpoints e do mecanismo de backup.

Parâmetro	Cálculo	Resultado nominal
Dados protegidos estimados	Estimativa inicial do TR	6,5 TB
Orçamento para cópia e versões	$6,5 \text{ TB} \times 2$	13 TB
Capacidade após espelhamento	Dois HDDs de 20 TB em RAID 1	20 TB
Limite planejado de ocupação	$20 \text{ TB} \times 80\%$	16 TB
Margem dentro do limite	$16 \text{ TB} - 13 \text{ TB}$	3 TB
Parcela fora do limite planejado	$20 \text{ TB} \times 20\%$	4 TB

Os resultados são nominais. Após a instalação, o limite de 80% deverá ser recalculado sobre a capacidade útil efetivamente disponibilizada. A política de versões deverá ser ajustada à ocupação observada, preservando os dados prioritários e sem pressupor crescimento anual que não foi medido.

O espelhamento protege contra falha de uma unidade, mas não substitui uma cópia independente contra perda do equipamento ou incidente no laboratório. Para dados insubstituíveis, deverá ser definida cópia em outro equipamento ou local, mediante avaliação dos recursos disponíveis; essa providência não constitui item adicional desta aquisição.

8 Compatibilização e conclusão

A implementação requer comprovação de que o equipamento de armazenamento acomoda simultaneamente os quatro SSDs e os dois HDDs, com interfaces e arranjos compatíveis. A reserva SSD é fria e ficará fora do arranjo. O arranjo RAID 1 dos HDDs e sua finalidade de backup deverão ser expressamente refletidos no TR.

Devem ser harmonizadas as referências a PDU, nobreak e uplinks com o objeto efetivamente contratado. Também deve ser esclarecido que os SSDs do item 7 compõem o armazenamento do item 6, sem duplicidade de fornecimento. Essas observações não alteram os oito quantitativos discriminados nesta manifestação.

A estimativa dos quantitativos resulta, portanto, do cenário prospectivo de até quatro projetos simultâneos, da alocação flexível para modelos maiores, da estimativa inicial de armazenamento e do orçamento de backup com retenção de versões. Os componentes de rede e cabeamento decorrem das conexões da arquitetura. As premissas de demanda aqui registradas complementam o TR V3 e deverão integrar os autos mediante validação da equipe técnica.

Catanduva, 8 de outubro de 2026.

Documento assinado digitalmente
 **MARCIO ANDREY TEIXEIRA**
Data: 08/10/2026 14:18:05-0300
Verifique em <https://validar.itl.gov.br>

Márcio Andrey Teixeira
Coordenação do CEPIA² — IFSP Campus Catanduva