

RNCA

REDE NACIONAL DE COMPUTAÇÃO AVANÇADA

Modelo de custos

Relatório de análise e propostas

Grupo de Trabalho constituído no âmbito do Conselho de
Coordenação da RNCA

Dezembro de 2023

Tabela 1 – Relatório

EXT/2023/AIA CA	
DATA	2023-12-11
ELABORADO POR	ELANA ARAÚJO, JOÃO PAGAIME
COLABORAÇÃO	JOÃO BARBOSA, RICARDO VILAÇA, JORGE GOMES E PEDRO INÁCIO

TLP: WHITE

ÍNDICE

Acrónimos 3

1. Introdução	4
2. Termos de referência	6
Missão	7
Objetivos	7
Composição do grupo de trabalho	7
Resultados Previstos	7
Exclusões	7
Condicionantes	7
3. Componentes do Modelo de Custos	7
4. Casos similares	9
4.1. Suporte Europeu para Acesso a Infraestruturas Digitais	9
4.2. Apontamentos sobre a RES - Red Española de Supercomputación	11
4.3. Apontamentos sobre o Sistema Americano XSEDE	12
5. Grandeza das capacidades computacionais	13
6. Propostas e conclusão	14
Figura 1 - Custos do modelo computacional HPC	5
Figura 2 - Custos do modelo computacional Cloud	5
Figura 3 - Custos do modelo computacional HTC	6
Tabela 1 – Relatório	2
Tabela 2 - Custos de Virtual Access	10

ACRÓNIMOS

AI	Inteligência Artificial
CAPEX	Capital Expenditure (Despesas de Capital)
CPU	Central Processing Unit
COVID-19	Doença por Coronavírus 2019
FCT	Fundação para a Ciência e a Tecnologia
FCT-FCCN	Fundação para a Ciência e a Tecnologia - Unidade de Computação Científica Nacional
GPU	Graphics Processing Unit
HPC	High-Performance Computing
HEPSPEC	High Energy Physics SPECification

HTC	High Throughput Computing (Computação de Alto Rendimento)
I. P.	Instituto Público
INCoDe.2030	Iniciativa Nacional Competências Digitais e.2030
NSF	National Science Foundation
OPEX	Operational Expenditure (Despesas Operacionais)
PRACE	Partnership for Advanced Computing in Europe
RICA	Rede Ibérica de Computação Avançada
RNCA	Rede Nacional de Computação Avançada
RNIE	Roteiro Nacional das Infraestruturas de Investigação de Interesse Estratégico
RES	Rede Española de Supercomputación
TCO	Total Cost of Ownership (Custo Total de Propriedade)
US	Unidades de Serviço
VA	Virtual Access
WLCG	Worldwide LHC Computing Grid
XSEDE	Extreme Science and Engineering Discovery Environment

1. INTRODUÇÃO

A Fundação para a Ciência e a Tecnologia I. P. (FCT) desempenha um papel crucial na instalação, manutenção e gestão de meios computacionais avançados, conforme estipulado nas sua Missão e Lei Orgânica¹. Esta responsabilidade inclui instalar, manter, gerir e promover a acessibilidade desses recursos às diversas entidades do Sistema Educativo e do Sistema Científico e Tecnológico Nacional, independentemente de sua natureza pública ou privada. A disponibilização em rede de meios computacionais avançados é essencial para o progresso deste sistema. Tal permite que instituições integradas nestas redes tenham acesso a aplicações informáticas específicas relacionadas com diversos modelos de computação avançada, como HPC (*High Performance Computing*), HTC (*High Throughput Computing*), Computação em Nuvem (*Cloud Computing*), entre outros.

O desenvolvimento da Rede Nacional de Computação Avançada (RNCA) é um objetivo explícito, conforme estabelecido no anexo à Resolução do Conselho de Ministros n.º 26/2018, inserido no âmbito do Eixo 5 do Programa "Iniciativa Nacional Competências Digitais e.2030 - INCoDe.2030".

O Roteiro Nacional das Infraestruturas de Investigação de Interesse Estratégico (RNIE), identificou, desde 2013, infraestruturas digitais, incluindo aquelas de âmbito geral destinadas a servir todas as áreas científicas. A "Rede Nacional de Computação Avançada" foi posteriormente acrescentada a esse roteiro, alinhando-se com a "Rede Ibérica de Computação Avançada - RICA", conforme estabelecido no Memorando de Entendimento assinado entre o Governo da República Portuguesa e o Governo do Reino de Espanha em novembro de 2018.

¹ https://www.fct.pt/documentos/Lei_Organica_2013.pdf

O financiamento do RNIE, atribuído na perspetiva de despesas de investimento, encontra complemento no regulamento da RNCA, que aprofunda e especifica as despesas de funcionamento associadas.

Destaca-se que as infraestruturas digitais de computação avançada, de âmbito nacional, exigem suporte centralizado², especialmente em termos de consumo de energia elétrica e de pessoal técnico especializado dedicado à operação.

A relevância da valorização das capacidades computacionais da RNCA foi ressaltada em 2020, quando foi necessária no contexto do concurso "AI 4 COVID-19: Ciência dos Dados e Inteligência Artificial na Administração Pública para reforçar o combate à COVID-19 e futuras pandemias – 2020"³. Esta iniciativa refletiu a importância estratégica da RNCA no suporte às ações de combate à pandemia, destacando o seu papel essencial na aplicação de técnicas avançadas, como Ciência dos Dados e Inteligência Artificial. E não apenas sublinhou a relevância imediata da RNCA no combate a desafios críticos, como a pandemia da COVID-19, mas também destacou o seu potencial para contribuir para futuras situações e desafios científicos emergentes.

A seguir, apresentam-se quadros (figuras 1, 2 e 3) que documentam as diretrizes e a valorização específica definida para a RNCA no âmbito desse concurso, solidificando o compromisso com a promoção da excelência na computação avançada em contextos de importância nacional e global:

Tipo de recurso	HPC - <i>High Performance Computing</i>
Valores unitários	<ul style="list-style-type: none"> • 0,0123 EUR/core.hora • 0,18 EUR/GPU.hora, que acresce ao valor de CPU onde o GPU se encontra agregado. Modelos de referência: NVidia T4 e NVidia V100 • 6,5 EUR/TB.mês para armazenamento de dados até um máximo de 2TB, podendo ser considerados volumes superiores em função da disponibilidade da plataforma

Figura 1 - Custos do modelo computacional HPC.

Tipo de recurso	Cloud (servidores virtuais)
Valores unitários	<ul style="list-style-type: none"> • 0,0123 EUR/vCPU.hora, incluindo 2GB de RAM por cada vCPU • Armazenamento/disco: 15,0 EUR/TB.mês

Figura 2 - Custos do modelo computacional Cloud⁴.

² Regulamento n.º 1049/2020, de 25 de novembro, <https://dre.pt/dre/detalhe/regulamento/1049-2020-149532837>

³ <https://www.fct.pt/apoios/projectos/concursos/datascience/index.phtml.pt>

⁴ O custo de armazenamento da INCD na cloud é 12.46 €/TB/mês devido à triplicação da informação no sistema de armazenamento Ceph;

Tipo de recurso	HTC- <i>High Throughput Computing</i>
Valores unitários	<ul style="list-style-type: none"> • 0,0123 EUR/core.hora • 0,18 EUR/GPU.hora, que acresce ao valor de CPU onde o GPU se encontra agregado. Modelos de referência: NVidia T4 e Nvidia V100 • 6,5 EUR/TB.mês para armazenamento de dados até um máximo de 2TB, podendo ser considerados volumes superiores em função da disponibilidade da plataforma

Figura 3 - Custos do modelo computacional HTC.

Em suma, a valorização das capacidades computacionais da RNCA desempenha um papel fundamental em várias frentes. Destacamos:

- Atendimento a Pedidos e Informação aos Utilizadores -- a valorização permite uma abordagem transparente e informada dos pedidos de potenciais utilizadores da rede. Isto não apenas fornece esclarecimentos aos utilizadores sobre os recursos que estão a solicitar, como também contribui para a compensação dos centros operacionais pelos apoios prestados;⁵
- Estabelecimento de Acordos e Parcerias -- facilita a formação de acordos de prestação de serviços de computação avançada com entidades do setor privado. Essas parcerias estratégicas não só impulsionam o desenvolvimento da RNCA, mas também fortalecem a colaboração entre o setor público e privado em iniciativas tecnológicas avançadas;
- Incentivo a Projetos de Investimento -- A valorização é um impulsionador essencial na formação de projetos de investimento destinados ao desenvolvimento contínuo das plataformas tecnológicas dos centros operacionais. Esses investimentos visam manter a RNCA na vanguarda da inovação, promovendo avanços nas plataformas tecnológicas para atender à crescente procura de recursos computacionais.

Dessa forma, a RNCA não apenas se destaca como uma infraestrutura de referência em computação avançada, mas também se posiciona como um catalisador essencial para o avanço científico, tecnológico e colaborativo no panorama nacional.

O presente estudo tem um carácter meramente consultivo e não condiciona decisões futuras da FCT ou de qualquer outro membro da RNCA.

2. TERMOS DE REFERÊNCIA

O grupo de trabalho foi estabelecido em 29 de julho de 2022, com a responsabilidade de abordar os seguintes termos de referência:

⁵ Principalmente na perspetiva das despesas de funcionamento como energia elétrica.

Missão

Propor à FCT uma eventual revisão ao modelo de custos.

Objetivos

- a) Identificar Componentes do Modelo de Custos:
 - Despesas CAPEX (Investimento em equipamento, etc.);
 - Despesas OPEX (Despesas operacionais como energia, recursos humanos, entre outros);
- b) Identificar Casos Semelhantes;
- c) Ajustar Grandezas Existentes aos Pedidos;
- d) Propor Melhorias ao Modelo Existente.

Composição do grupo de trabalho

- Coordenação: FCT-FCCN - Elana Araújo e João Pagaiame;
- Membros:
 - Centro Operacional Deucalion – UMinho: João Barbosa e Ricardo Vilaça;
 - Centro Operacional INCD: Jorge Gomes;
 - Centro de Competências e Visualização UBI: Pedro Inácio.

Resultados Previstos

1. Elaboração de um relatório abrangente pelo Grupo de Trabalho;
2. Apresentação do relatório em uma reunião do Conselho de Coordenação da RNCA;
3. Informar o Conselho Diretivo da FCT, I.P.

Exclusões

Análise detalhada dos custos específicos dos centros operacionais.

Condicionantes

Os resultados gerados têm caráter consultivo e não condicionarão decisões futuras da FCT.

Este conjunto claro de termos de referência guiará o trabalho do grupo, assegurando uma análise abrangente e relevante para a proposta de revisão do modelo de custos da RNCA.

3. COMPONENTES DO MODELO DE CUSTOS

No contexto dos objetivos deste relatório, a identificação dos componentes do modelo de custos é uma etapa crucial que visa desdobrar e categorizar os diversos elementos financeiros associados à operação de um centro operacional dentro da RNCA. Esse processo estratégico tem como objetivo oferecer uma visão dos fatores que contribuem para os custos envolvidos na manutenção e funcionamento eficiente da infraestrutura de supercomputação.

Ao desempenhar esse papel, o grupo de trabalho analisou e categorizou os diversos custos associados à operação do centro operacional. Esses componentes de custos abrangem diversas áreas, desde os investimentos iniciais em equipamentos (CAPEX) até os custos operacionais

contínuos (OPEX), como despesas com energia, recursos humanos e outros fatores essenciais para a operação diária.

3.1. Despesas CAPEX

Despesas de *Capital Expenditure* (CAPEX) na supercomputação referem-se aos investimentos necessários para adquirir e estabelecer a infraestrutura de computação avançada. Como investimentos de impacto significativo a longo prazo na capacidade e eficiência da operação devemos considerar:

- a) Equipamento Computacional e Prazo de Amortização -- Inclui a aquisição de hardware, como servidores, sistemas de armazenamento e outros componentes para a infraestrutura. No contexto de projetos de I&D&I, onde a amortização é imputada à duração dos projetos, dependendo da taxa de evolução tecnológica, o prazo de amortização é de cerca de 4 anos;
- b) Taxa de Substituição de Equipamentos -- a taxa de substituição refere-se à frequência com que os equipamentos são atualizados ou substituídos. No caso da supercomputação, uma taxa de substituição de aproximadamente 25% ao ano pode ser considerada. Essa abordagem leva em conta a garantia típica dos equipamentos e busca manter a infraestrutura atualizada com tecnologias mais recentes para garantir desempenho e eficiência;
- c) Licenciamento de *Software* Científico e de Suporte Tecnológico -- além do hardware, as despesas também podem incluir o licenciamento de *software* científico e ferramentas de suporte às plataformas tecnológicas. Isso abrange desde compiladores otimizados até softwares especializados necessários para a execução eficiente de aplicações científicas na supercomputação.

3.2. Despesas OPEX

Despesas operacionais (OPEX) referem-se, neste caso, aos custos contínuos associados à operação, manutenção e suporte de uma infraestrutura. Sendo custos essenciais para garantir o funcionamento eficiente:

- a) Energia Elétrica e Despesas de Alojamento -- incluem os custos relacionados ao consumo de energia elétrica para alimentar os equipamentos computacionais, bem como despesas associadas ao espaço físico para alojar a infraestrutura. Isso abrange também os custos de refrigeração para manter os equipamentos operando eficientemente;
- b) Recursos Humanos -- salários e despesas associadas ao pessoal envolvido no suporte diário, incluindo administradores de sistema, técnicos de suporte e outros profissionais necessários para manter a operação contínua;
- c) Seguros e Contratos de Suporte -- incluem os custos associados a seguros específicos para a infraestrutura de supercomputação e contratos de suporte que garantem a manutenção e o bom funcionamento dos equipamentos;
- d) Formação do Pessoal -- abrange despesas relacionadas à formação contínua do pessoal, garantindo que esteja atualizado com as mais recentes tecnologias e práticas;

- e) Missões/Deslocações -- incluem custos associados a viagens e deslocações para participação de eventos, conferências ou realização de eventos e atividades específicas;
- f) Divulgação -- despesas associadas a atividades de divulgação, como a participação em conferências, workshops, e outras iniciativas de disseminação e promoção;
- g) Desenvolvimento e Manutenção de Sistema de Informação Específico -- custos de desenvolvimento e manutenção de sistemas de informação específicos para a infraestrutura, incluindo softwares dedicados para operações;
- h) Implementação e Operação de Serviços da Infraestrutura -- custos relacionados à implementação e operação contínua dos serviços essenciais, como serviços de armazenamento, clusters de computadores e outros elementos. Esses serviços visam garantir a funcionalidade e eficiência da infraestrutura no suporte à procura de recursos de computação avançada.

Procurámos atribuir às componentes do modelo de custos uma natureza genérica e abrangente, conforme delineado no regulamento da RNCA. No entanto, é importante destacar que a intenção é *"aprofundar e complementar o RNIE na ótica das despesas de funcionamento"*. Isso sugere que, normalmente, as despesas de CAPEX são consideradas externas ao modelo de custos da RNCA. Essa distinção é relevante, uma vez que as componentes CAPEX costumam ser suportadas por projetos de investimento que recorrem a financiamento europeu, conforme especificado nas diretrizes do regulamento. Essa compreensão é essencial para direcionar estrategicamente os recursos no âmbito da RNCA, considerando as fontes de financiamento.

Ao identificar os componentes do modelo de custos, espera-se facilitar a gestão eficaz dos recursos financeiros, permitindo otimizações para redução de despesas ou realocação estratégica de recursos, mas também fortalecer a transparência na comunicação dos custos associados à operação da RNCA. Ao oferecer informações claras e fundamentadas, atendemos às necessidades dos *stakeholders*, incluindo a FCT e outros parceiros envolvidos, estabelecendo uma base sólida que sustenta as análises e recomendações deste relatório.

4. CASOS SIMILARES

Um dos objetivos deste documento é a identificação de casos similares, uma vez que as questões levantadas na RNCA sobre a valorização dos apoios em computação avançada não são exclusivas do contexto nacional. Dessa forma, é essencial examinar casos internacionais análogos que possam enriquecer e informar o presente processo. Este estudo abrange uma análise de diversos cenários europeus, e instituições norte-americanas. Além disso, contempla uma exploração das práticas adotadas por outros grandes centros e redes de supercomputação que também abordam de maneira destacada a valorização dos apoios. A abordagem comparativa destas referências internacionais fornece informações e linhas orientadoras para o desenvolvimento da RNCA, proporcionando uma perspetiva global para os desafios específicos enfrentados no âmbito nacional.

4.1. Suporte Europeu para Acesso a Infraestruturas Digitais

Com o objetivo de identificar modelos de custos vinculados à disponibilização de meios computacionais e de armazenamento, direcionamos a nossa atenção para projetos europeus que fazem uso do mecanismo de *Virtual Access* (VA). Esse mecanismo, estabelecido pela Comissão Europeia, facilita o acesso a infraestruturas digitais sem custo direto para o utilizador final. Essas infraestruturas abrangem uma variedade de serviços, como computadores de grande capacidade, redes, GRIDs, repositórios, bases de dados e participação em comunidades virtuais. No âmbito do VA, cada infraestrutura determina os custos de acesso aos seus serviços com base exclusivamente nos custos associados, fundamentados em informações contabilísticas. Vale a pena ressaltar que o mecanismo de VA utiliza dados referentes ao ano anterior à submissão da proposta, abrangendo apenas os custos incorridos, excluindo a margem de lucro.

O VA oferece três opções de cálculo da unidade de imputação, cada uma com suas características distintas:

- *Unit costs* – que identificam todos os custos envolvidos na prestação do serviço, incluindo a amortização dos equipamentos e custos operacionais, como recursos humanos. O total desses custos é dividido pela capacidade total para se obter um custo unitário;
- *Actual costs* – que se baseia na estimativa dos custos adicionais para tornar o serviço acessível externamente, desconsiderando os custos de capital da infraestrutura. Essa opção é preferencial quando os custos de capital e operacionais já são cobertos por outros financiamentos;
- Combinação das duas opções anteriores -- utilizada quando existem custos adicionais além dos associados à operação e ao capital, especialmente para abrir o serviço a novos utilizadores.

Os valores médios apresentados na tabela seguinte foram calculados com base nas informações de custos de VA provenientes de 21 fornecedores europeus, dos quais 19 são académicos e dois são comerciais:

Unidade	Serviços amostrados	Desvio padrão	Preço unitário
vCPU core/hora	23	0.0093	0.0174 €
GPU /hora	3	0.5668	0.7868 €
TB /mês (total)	18	5.38	14.61 €
TB /mês (RAID)	5	2.52	7.93 €
TB /mês (replicação)	13	3.63	17.18 €

Tabela 2 - Custos de *Virtual Access*

Os valores individuais utilizados no cálculo das médias foram provenientes de duas propostas de projetos europeus em curso até novembro de 2022, referentes aos anos de 2019 e 2020. A análise abrangeu instalações em diversos países, incluindo República Checa, Espanha, Polónia, Alemanha, Eslováquia, Itália, França, Turquia, Portugal, Grécia e Holanda. Os custos apresentados na tabela foram ajustados para facilitar a comparação, sendo subtraídos os gastos gerais em 25%. A não divulgação dos valores individuais visa preservar a privacidade dos fornecedores.

Contudo, o mecanismo de VA apresenta desafios significativos, como a dependência de valores reportados no ano anterior à submissão do projeto, que são utilizados para sustentar a prestação de serviços ao longo de dois ou três anos sem atualizações. Além disso, a inclusão de

certos custos pode ser complexa caso não tenham sido estimados corretamente, sejam incorridos por outras entidades ou estejam fora do período contabilístico de amortização. Estas limitações decorrem da necessidade de ancorar os custos em informações contabilísticas, o que, em alguns casos, pode resultar em valores abaixo do custo efetivo. Por outro lado, a principal vantagem reside na capacidade de imputação com base no *Total Cost of Ownership* (TCO) dos serviços, proporcionando uma média que deverá refletir razoavelmente os custos reais.

Em relação à normalização das unidades de utilização, destaca-se o uso da métrica HEPSPÉC no *Worldwide LHC Computing Grid* (WLCG⁶), originária do grupo de *benchmarking* HePIX. Apesar de inicialmente focada em física de altas energias, essa métrica foi adotada pela infraestrutura europeia EGI para a contabilização padronizada do tempo de utilização. Utilizando as métricas de tempo de relógio normalizado (*wall-clock time*) e tempo de CPU normalizado (CPU time), essa abordagem contribui para uma avaliação mais precisa e comparável do acesso a meios de computação de alto débito e de computação em nuvem.

A análise dos valores médios provenientes do mecanismo de VA revela desafios e vantagens significativas, destacando a importância de considerações contabilísticas e a necessidade de abordagens transparentes na prestação de serviços de infraestrutura digital. A preservação da privacidade dos fornecedores é uma prática essencial, ao mesmo tempo em que a imputação com base no TCO oferece uma visão mais realista dos custos associados. A normalização das unidades de utilização, exemplificada pelo HEPSPÉC, evidencia a busca por métricas consistentes em ambientes federados, contribuindo para uma análise abrangente e comparativa.

4.2. Apontamentos sobre a RES - Red Española de Supercomputación

Em reuniões realizadas com Oriol Pineda Martínez⁷ - coordenador da política de acesso à infraestrutura – em que foi discutido este tema, como parte de uma investigação mais ampla sobre o modelo de custos adotado por uma das congéneres da RNCA, a Red Española de Supercomputación (RES), foi compreendido sobre os elementos-chave do modelo de custos da RES e explorado os princípios fundamentais que norteiam a tomada de decisões no que diz respeito aos investimentos em infraestrutura e operações.

Foi também realizada uma pesquisa adicional com informações públicas disponíveis sobre os custos associados à RES, a fim de obter uma visão mais holística. Um aspecto notável desse modelo é a consideração dos investimentos, divididos entre infraestrutura de investigação e

⁶ Infraestrutura de computação distribuída utilizada para processar e analisar os dados gerados pelos experimentos no Grande Colisor de Hádrons (LHC) no CERN (Organização Europeia para Investigação Nuclear): <https://cds.cern.ch/record/2798712/files/document.pdf>

⁷ <https://www.bsc.es/pineda-martinez-oriol>

suporte, incluindo redes e energia elétrica. A RES, alinhada com as melhores práticas, avalia periodicamente a eficiência e os custos operacionais para otimizar o desempenho geral.

Ao analisarmos a métrica preferida pela RES, notamos uma inclinação em direção ao número de *Flops* como uma medida universal, independente do hardware. Essa escolha estratégica facilita comparações entre diferentes supercomputadores, contribuindo para uma gestão mais eficaz e transparente dos recursos.

O relacionamento entre a RES e o *Partnership for Advanced Computing in Europe* (PRACE) é importante para entender o financiamento e a dinâmica operacional. A RES, como parte integrante do PRACE, beneficia-se de uma rede de colaboração europeia que promove a partilha de recursos e conhecimentos. O PRACE desempenha um papel significativo no apoio à pesquisa computacional de ponta em toda a Europa, consolidando esforços para otimizar a utilização de supercomputadores.

Quando se trata de atribuição de projetos e subsídio de centros operacionais, a RES mantém uma abordagem pragmática. Priorizando a maximização da utilização de recursos, a rede intervém apenas em casos específicos de queixas relacionadas à lentidão no processamento. Essa flexibilidade é uma estratégia proativa para garantir a eficácia operacional e a satisfação dos utilizadores finais.

Em relação ao custo de venda de recursos computacionais, a RES adota uma abordagem transparente, solicitando ofertas de mercado para garantir uma estimativa precisa. Essa prática reflete o compromisso da RES em fornecer serviços de supercomputação com preços competitivos e alinhados com as práticas do mercado.

No contexto do PRACE, são notáveis a contribuição para a padronização de métricas e a promoção da transparência em relação aos custos operacionais. A RES, ao integrar-se nesse ecossistema europeu de computação avançada, não beneficia apenas da colaboração internacional, como também contribui para o desenvolvimento contínuo de modelos de custos eficientes e adaptáveis.

4.3. Apontamentos sobre o Sistema Americano XSEDE

Colaborando com centros de computação avançada como o *Texas Advanced Computing Center (TACC)*, o *Extreme Science and Engineering Discovery Environment (XSEDE)* opera sob uma filosofia de acesso aberto, permitindo à comunidade académica e de investigação aproveitar os seus recursos avançados de computação, armazenamento e rede de forma geralmente gratuita. O XSEDE não é apenas uma iniciativa na promoção da investigação científica nos Estados Unidos, é também um pioneiro no desenvolvimento de um modelo de custos.

A chave para a acessibilidade proporcionada pelo XSEDE reside no financiamento e no apoio de agências governamentais, como a *National Science Foundation (NSF)*. Essas colaborações sustentam o acesso aberto, eliminando barreiras financeiras para investigadores académicos e permitindo que explorem todo o potencial dos recursos de computação avançada disponíveis. Embora o objetivo principal do XSEDE seja servir à comunidade académica, a iniciativa também estende sua mão à indústria, participando em projetos específicos. Nesses casos, acordos

personalizados e taxas adaptáveis são implementados para atender à procura e circunstâncias particulares.

O sistema de alocação de recursos do XSEDE é baseado em **Unidades de Serviço (US)**, uma métrica que avalia a complexidade e intensidade do uso dos recursos. A atribuição de valores financeiros diretos a essas USs pode ser complexa devido às nuances envolvidas nos custos reais de fornecimento. No entanto, uma estimativa aproximada sugere que cada US pode ter um valor mínimo de 0,05 dólares. Isso implica que uma alocação de um milhão de USs teria um valor estimado de pelo menos 50.000 dólares, oferecendo uma perspectiva sobre a escala e o impacto dos recursos disponíveis.

Esse modelo flexível não só permite que a investigação acadêmica floresça em um ambiente de acesso aberto, como também facilita colaborações específicas com a indústria, promovendo uma dinâmica entre academia e setor privado. O compromisso do XSEDE com o equilíbrio entre acessibilidade à computação avançada e sustentabilidade financeira exemplifica um novo paradigma na condução de pesquisas científicas e impulsiona a inovação nos Estados Unidos. O XSEDE não é apenas uma infraestrutura tecnológica avançada, é também um catalisador que para uma nova era de descobertas científicas e colaborações interdisciplinares.

5. GRANDEZA DAS CAPACIDADES COMPUTACIONAIS

No processo de formulação de pedidos computacionais na RNCA, os potenciais utilizadores desempenham um papel crucial ao identificar e comunicar suas necessidades para alcançar os objetivos de seus projetos científicos ou de inovação. Essas necessidades são expressas por meio de grandezas computacionais estrategicamente definidas, visando uma alocação eficiente de recursos. Essas grandezas, baseadas nas necessidades dos utilizadores, estão em constante evolução para acompanhar a procura da comunidade científica.

- 1) Capacidade de Processamento Baseada em CPU:
 - Descrição: Representa o número de unidades de processamento (CPU core) necessária. Essa métrica é ajustada dinamicamente com base na duração do projeto, seguindo princípios de otimização de recursos;
 - Adaptação Temporal: A quantidade de unidades de processamento é ajustada com base no tempo especificado, reduzindo para períodos mais longos e aumentando para períodos mais curtos;
 - Alinhamento Científico⁸: Estudos recentes destacam a importância de correlacionar a capacidade de processamento com a carga computacional específica de diferentes tipos de simulações científicas;
 - Exemplo: 500.000 CPU core. horas, equivalente a ~680 unidades de processamento concorrentes durante 1 mês (730 horas).

⁸ Fonte: "Correlating Computational Workload Characteristics with HPC Resource Usage Patterns": https://link.springer.com/chapter/10.1007/978-3-031-32041-5_16

2) Capacidade de Processamento com Base em Placa de GPU:

- Descrição: Similar à capacidade de CPU, mas aplicada a placas de GPU, cada uma contendo milhares de unidades de cálculo em vírgula flutuante;
- Ajuste Dinâmico: Assim como na CPU, a adaptação temporal é aplicada, otimizando o número de GPUs conforme a duração do projeto;
- Exploração de Tecnologia: A capacidade de GPU, essencial para tarefas intensivas em cálculos em vírgula flutuante, é mapeada considerando a evolução das arquiteturas de GPU. A adaptação temporal é apoiada por inovações contínuas em hardware de GPU;
- Exemplo: Um projeto de pesquisa em inteligência artificial pode requerer 10.000 horas de GPU para treinar modelos complexos de aprendizado profundo.

3) Armazenamento de Dados Tipo HPC:

- Descrição: Indica o número de *TeraBytes* de espaço de armazenamento em sistemas HPC e a duração em meses necessária;
- Características: O armazenamento HPC, embora ofereça alto desempenho, pode apresentar desafios em termos de fiabilidade. Estratégias avançadas de redundância de equipamentos são empregues para garantir a integridade dos dados;
- Exemplo: Um projeto de sequenciamento genómico pode requerer 20 *TeraBytes* de armazenamento HPC para armazenar e analisar grandes conjuntos de dados genéticos.

4) Armazenamento de Dados Tipo Cloud:

- Descrição: Similar ao armazenamento HPC, mas adaptado às nuances da cloud científica;
- Adaptação Dinâmica: O armazenamento em ambientes de cloud é projetado para ser dinâmico, refletindo as características particulares desse ambiente flexível e escalável;
- Exemplo: Um projeto de pesquisa em ciência de dados, utilizando técnicas de aprendizado de máquina, pode necessitar de armazenamento em cloud para lidar com grandes conjuntos de dados distribuídos.

Além dessas grandezas, a RNCA oferece um conjunto fundamental de apoios:

- Suporte Técnico: Inclui a disponibilização de documentação, auxílio no login, resolução de problemas, entre outros, como treinamentos e softwares personalizados quando aplicável;
- Espaço Mínimo de Armazenamento: Em alguns casos, é oferecido um espaço de armazenamento mínimo suficiente para a execução do trabalho do utilizador.

Essa abordagem estratégica visa garantir não apenas uma alocação eficaz de recursos computacionais, mas também um alinhamento estreito com as necessidades individuais dos utilizadores. A capacidade de ajustar as grandezas conforme a natureza e a duração do projeto permite uma flexibilidade crucial para otimizar o uso dos recursos disponíveis na RNCA.

6. PROPOSTAS E CONCLUSÃO

Perante o atual contexto da RNCA, apresentamos a seguinte proposta para otimizar o sistema de custos vigente, com inspiração no modelo XSEDE, propondo uma abordagem na apresentação dos custos e transformando-os em US, i.e., unidades computacionais padronizadas, em vez de valores monetários. As propostas são:

1. Revisão Dinâmica do Modelo de Custos -- recomendamos a implementação de uma revisão dinâmica deste modelo, assegurando sua adaptação contínua às mudanças tecnológicas, procura dos utilizadores e evoluções das boas práticas internacionais;
2. Catálogo de Serviços -- desenvolvimento de um catálogo de serviços, atribuindo a cada um uma quantidade específica de unidades computacionais com base na procura e impacto nos recursos;
3. Otimização da Alocação de Recursos -- otimizar a alocação de recursos, identificando possíveis redundâncias e áreas onde a eficiência pode ser melhorada, como por exemplo, a monitorização dos projetos alocados;
4. Transparência e Acompanhamento -- implementar um sistema de acompanhamento em tempo real, permitindo que os administradores visualizem o consumo de unidades computacionais associadas aos serviços utilizados;
5. Comunicação -- sessões informativas para os utilizadores da RNCA, explicando o novo modelo de unidades computacionais e promovendo a compreensão dos custos associados.

Os benefícios esperados na RNCA são vários. A transparência oferece uma visão do consumo de recursos, possibilitando uma compreensão detalhada de como suas atividades impactam diretamente os custos associados. A facilidade de comparação entre serviços e recursos promoverá escolhas informadas e eficientes, contribuindo para uma alocação mais precisa de recursos.

A tomada de decisão fundamentada será potenciada pela disponibilidade de dados precisos, permitindo que os gestores tenham uma maior compreensão dos recursos e investimentos. A simplificação administrativa resultante da mudança para o modelo de unidades computacionais reduzirá a carga operacional.

Além disso, é importante destacar que esta proposta de modernização não apenas aprimora o sistema de custos, mas também promove a unificação de todos os diferentes centros operacionais. Essa uniformidade contribuirá para a redução de disparidades entre os centros, tornando as operações mais similares e diminuindo as diferenças que podem impactar a eficiência e a gestão integrada da RNCA.

Em síntese, a abordagem proposta promoverá a transparência e eficiência, mas também estabelecerá uma base sólida para a evolução contínua da RNCA, garantindo que a instituição esteja adequadamente posicionada para atender à crescentes procura da comunidade científica.

Com este relatório, este grupo de trabalhos conclui a sua missão e encontra-se à disposição para quaisquer ajustes adicionais ou informações complementares que se julguem necessários.