

Uso de nuvem pública para computação de alto desempenho



**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Agricultura Digital
Ministério da Agricultura, Pecuária e Abastecimento**

DOCUMENTOS 188

Uso de nuvem pública para computação de alto desempenho

*Leandro Carrijo Cintra
Cristian Alfonso Latapiat Gonzalez*

Embrapa Agricultura Digital
Av. Dr. André Tosello, nº 209 - Campus da Unicamp,
Barão Geraldo - Campinas, SP
CEP. 13083-886 - Fone: +55 (19) 3211-5700

www.embrapa.br/agricultura-digital
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações
da Unidade Responsável

Presidente
Carla Geovana do Nascimento Macário

Secretário-Executivo
Maria Fernanda Moura

Membros
*Adriana Farah Gonzalez, Alexandre de Castro,
Carla Cristiane Osawa, Debora Pignatari Drucker,
Ivan Mazoni, João Camargo Neto, João Francisco
Gonçalves Antunes, Magda Cruciol*

Revisão de texto
Adriana Farah Gonzalez

Normalização bibliográfica
Carla Cristiane Osawa

Projeto gráfico da coleção
Carlos Eduardo Felice Barbeiro

Editoração eletrônica
Magda Cruciol e Letícia Campos

Imagem da capa
stock.adobe.com

1ª edição
Publicação digital (2022): PDF

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Agricultura Digital

Cintra, Leandro Carrijo.

Uso de nuvem pública para computação de alto desempenho / Leandro
Carrijo Cintra. - Campinas : Embrapa Agricultura Digital, 2022.

PDF (14 p.) : il. color. - (Documentos / Embrapa Agricultura Digital, ISSN
2764-2488 ; 188).

1. Computação em nuvem. 2. IaaS. 3. Computação de alto desempenho. 4.
High Performance Computing. I. Gonzalez, Cristiano Alfonso Latapiat. Título. II.
Embrapa Agricultura Digital. III. Série.

CDD (21. ed.) 004.36

Autores

Leandro Carrijo Cintra

Cientista da Computação, doutor em Bioinformática, analista da Embrapa Agricultura Digital, Campinas, SP

Cristian Alfonso Latapiat Gonzalez

Analista Senior em Nuvem da Rede Nacional de Ensino e Pesquisa (RNP), Campinas, SP

Apresentação

A computação em nuvem evoluiu consideravelmente nos últimos tempos e já é utilizada amplamente no mercado para o atendimento de demandas da área de Tecnologia da Informação relacionadas com as atividades corporativas.

Em relação às atividades na área de pesquisa e inovação, as opções de serviços disponíveis nas plataformas de computação em nuvem ainda não são amplas. Porém, pode-se perceber claramente que este panorama tem se modificado e, novamente, já há um movimento levando a adoção de soluções em nuvem envolvendo demandas de pesquisa e inovação, seja por parte de institutos de pesquisa públicos ou pelos departamentos de pesquisa de empresas privadas ao redor do globo.

Este trabalho fez uma avaliação da usabilidade e do custo-benefício de se utilizar infraestrutura como serviço (IaaS) em nuvem pública para o atendimento de demandas de computação de alto desempenho *High Performance Computing* (HPC), tanto nos aspectos envolvendo o processamento como o armazenamento de dados. Sua conclusão permite afirmar que já é possível o atendimento de algumas demandas na área com um bom custo-benefício.

As informações apresentadas são de interesse de estudantes ou profissionais que estejam atuando em P&D e busquem subsídios para decidir sobre o uso de computação em nuvem para projetos que demandem considerável poder computacional ou capacidade de armazenamento.

Carla Geovana do Nascimento Macário
Chefia de Pesquisa e Desenvolvimento
Embrapa Agricultura Digital

Sumário

Introdução.....	6
Considerações preliminares.....	7
Tópicos de avaliação e ensaios	8
Relato de resultados	9
Discussão dos resultados	13
Conclusões.....	14
Referências	14

Introdução

A computação de alto desempenho *High Performance Computing (HPC)* é componente primordial no desenvolvimento de uma ampla gama de projetos de pesquisa na Embrapa. Apenas para contextualizar, sem a ambição de listar todas as áreas que se beneficiam desta tecnologia, podemos elencar projetos que envolvem a análise de grandes volumes de dados ômicos, projetos que trabalham com modelagem e simulação de sistemas biológicos, projetos que realizam o zoneamento agrícola no País, e mais recentemente, projetos que se beneficiam do aprendizado de máquina, mais especificamente de técnicas de aprendizado profundo.

Desta forma, é comum a Embrapa empreender esforços para a aquisição e manutenção de equipamentos que possibilitem o armazenamento e processamento intensivo de dados. Tais aquisições podem ocorrer em iniciativas difusas dentro da empresa no âmbito dos projetos de pesquisa, ou de uma forma mais colaborativa, como no caso do Laboratório Multiusuário de Bioinformática da Embrapa (LMB), que mantém uma infraestrutura computacional de alto desempenho utilizada de forma compartilhada por diversas unidades da Empresa. De uma maneira ou outra, estas soluções envolvem a aquisição dos equipamentos computacionais e sua instalação e sustentação *in house*.

A principal vantagem deste modelo consiste no fato de que uma vez adquirido, os equipamentos estarão disponíveis por toda a sua vida útil. É simples ordenar e priorizar o uso internamente para atender a diversos projetos e pesquisadores distintos. Por outro lado, as desvantagens do modelo consistem na rápida obsolescência dos equipamentos dada a intensa evolução da tecnologia de armazenamento e processamento. Outra desvantagem está na ocorrência de subutilização dos recursos computacionais nos casos em que equipamentos específicos são demandados e adquiridos para o atendimento de determinados projetos, e uma vez superadas suas fases que demandam processamento, deixam de necessitar intensivamente dos recursos computacionais em questão. Deve-se ainda considerar que uma boa infraestrutura para receber estes equipamentos não é simples de ser obtida e demanda considerável investimento financeiro.

Estas desvantagens e dificuldades no modelo que envolve a aquisição dos equipamentos computacionais também estão presentes em outras organizações, e estes têm sido os principais fatores que impulsionam o desenvolvimento da computação em nuvem. Nestes novos modelos comerciais, os equipamentos deixam de ser adquiridos em parte ou em sua totalidade, e a organização passa a apoiar a sua TI em infraestrutura computacional, plataformas e software que lhe são disponibilizados como serviços. Esta abordagem já se mostrou extremamente eficaz para atender a um nicho amplo das demandas de TI para as aplicações corporativas (Gartner, 2022). Há, no entanto, dúvidas se a computação em nuvem seria igualmente eficaz para o atendimento de demandas na área de pesquisa e desenvolvimento. E frequentemente, durante a elaboração de projetos que irão demandar processamento de alto desempenho, pesquisadores se questionam qual abordagem apresentaria o melhor custo-benefício para atender às demandas do projeto.

Visando verificar na prática a capacidade da computação em nuvem em atender às demandas de projetos de pesquisa desenvolvidos na Embrapa e também, analisar o quão viável financeiramente seria uma abordagem envolvendo solução em nuvem para atender tais demandas; foi realizada em parceria com a Rede Nacional de Ensino e Pesquisa (RNP) uma prova de conceito (PoC). A descrição dos testes executados nesta PoC e os resultados obtidos encontram-se nas seções seguintes.

Considerações preliminares

A quantidade de aplicações empregadas em computação científica é muito extensa e faz com que a gama de cenários computacionais envolvidos na área seja bastante diversificada. Para aquela categoria de problemas que exigem processamento HPC, no entanto, é possível identificar pontos em comum em relação às demandas computacionais. Os problemas científicos que demandam HPC irão necessitar de um ou mais dos seguintes fatores: alto poder de processamento ou equipamentos com grande quantidade de memória volátil RAM ou alta capacidade de armazenamento de dados.

Desta forma, ao se analisar como as soluções em nuvem se comportam em relação a cada um destes três fatores, tem-se uma compreensão ampla das possibilidades do uso desta nova tecnologia aplicada para atender às demandas em computação científica, atuais e futuras, da área de P&DI.

Nesta PoC, a RNP identificou e contratou o provedor de serviços em nuvem com o melhor custo-benefício, sendo a Huawei a empresa a apresentar a melhor proposta. A equipe da Embrapa, por sua vez, avaliou a efetividade do uso do ambiente computacional em nuvem para a solução de problemas científicos. Para tanto, foram executados testes envolvendo a ferramenta GROMACS (Pronk et al., 2013), que realiza simulação de sistemas moleculares e tem relevância para alguns grupos de pesquisa da Empresa.

Estes testes com problemas reais no GROMACS permitiram que a usabilidade do ambiente fosse avaliada e também permitiram aferir o quão precisa é a plataforma ao estimar o custo dos recursos computacionais que ela disponibiliza. Ao se constatar que as estimativas preliminares correspondiam fielmente aos valores *realmente* gastos ao se utilizar o ambiente para os cálculos computacionais, pode-se adquirir confiança em abstrair e passar a calcular os custos baseando-se nas demandas computacionais de cada problema, sem a necessidade de realmente executar as análises.

Assim, neste trabalho, além dos resultados obtidos com os ensaios de simulações computacionais de sistemas moleculares no GROMACS, são apresentadas discussões relacionadas a análises de dados que já foram executadas no cluster do LMB e existe um histórico de consumo de recursos computacionais. Com tais informações e considerando a precisão das estimativas de custos dos recursos computacionais na plataforma de nuvem da Huawei, é possível que se estime com precisão e se avalie adequadamente o quanto de recursos financeiros seriam necessários para executar as mesmas análises na nuvem.

Por fim, seriam necessários alguns esclarecimentos em relação aos três fatores relevantes para a compreensão dos casos em que as soluções em nuvem são aplicáveis a demandas da computação científica.

Em relação à capacidade de manipular grandes volumes de dados, é preciso avaliar inicialmente se a plataforma permitiria o transporte e o armazenamento adequado desses dados, que são averiguações simples e não exigem maiores explicações.

Em relação a problemas que exigem equipamentos com grande capacidade de memória RAM, também se trata de uma demanda de fácil averiguação. É preciso analisar se a plataforma possibilita instanciar máquinas virtuais com uma quantidade de memória significativa, tais como 512G, 1T ou mais.

Em relação ao fator capacidade de processamento, segue uma discussão um pouco mais aprofundada para que fiquem claros os pontos que seriam necessários identificar na plataforma para se ter certeza de que ela possa ser utilizada no âmbito de soluções HPC.

A alta capacidade de processamento de ambientes computacionais científicos é obtida utilizando-se de diversas técnicas. A primeira abordagem para se obter alto desempenho é desenvolver equipamentos que tenham dezenas ou centenas de processadores. Assim, o primeiro ponto a se avaliar é se a plataforma em nuvem possibilita a instanciação de máquinas virtuais com muitas vCPUs. Ainda, é comum em problemas científicos situações em que apenas uma máquina não é suficiente para tratar em tempo hábil um problema. Desta forma, é preciso avaliar se o ambiente em nuvem permite a instanciação de *clusters* computacionais. E finalmente, é sabido que uma ampla gama de problemas em computação científica atualmente podem se beneficiar da existência de GPUs nas máquinas. Sendo assim, é preciso verificar se há a possibilidade de instanciar máquinas virtuais com GPUs.

Tópicos de avaliação e ensaios

Um ponto importante para ser avaliado está relacionado com a usabilidade do ambiente de computação em nuvem para a solução de problemas em computação científica. Neste caso, a preocupação é única e exclusivamente em garantir que as formas e a performance de acesso à infraestrutura computacional sejam adequadas. Visando responder a esta indagação, nesta PoC foi executado um ensaio envolvendo a simulação de sistemas moleculares no GROMACS. Os pesquisadores que têm esta demanda receberam acesso a uma máquina virtual já pré-configurada, e puderam avaliar o quão adequado eram as formas e a performance de acesso a ela.

Além dessa questão mais subjetiva, mas muito importante, da usabilidade, foram avaliados os fatores capacidade de processamento, disponibilidade de memória RAM e possibilidade de armazenamento de grandes volumes de dados, visto que, estes são fatores importantes para o HPC. Compreendendo-se como se comportam os sistemas computacionais em nuvem relativamente a esses pontos, é possível analisar de uma forma muito efetiva como irá se comportar o sistema para um problema genérico, cujas soluções sejam executadas na nuvem.

O ensaio de simulação computacional com o GROMACS, também forneceu dados para a avaliação do desempenho de processamento envolvendo GPUs. As informações obtidas levaram à conclusão que os custos estimados dos diversos recursos computacionais presentes na plataforma são muito precisos e se concretizam de fato quando se instancia e se utiliza os recursos no ambiente computacional.

Com tais conhecimentos, foi possível realizar duas outras análises do ambiente em nuvem, sem a necessidade de realmente executar todos os procedimentos computacionais. As análises em questão envolveram situações de interesse que já foram resolvidas no sistema de processamento de alto desempenho do LMB e com isso, tem-se à disposição os detalhes das demandas de recursos computacionais necessários.

A primeira análise de interesse envolveu a avaliação do ambiente para o processamento de alto desempenho envolvendo CPUs. Para tanto, considerou-se uma análise já realizada no âmbito do ambiente de processamento do LMB envolvendo técnicas de aprendizado de máquina para a geração de mapas de solos. De fato, não se executou totalmente a análise em nuvem, apenas se usou as informações detalhadas de demanda computacional que estavam registradas nos *logs* do LMB para estimar o custo de se executar a mesma análise em nuvem. A solução em questão foi desenvolvida no âmbito do projeto MDS-BR, e envolve o uso do pacote estatístico R.

Na análise, trabalha-se com informações de solos de todo o território nacional, e como se trata de uma quantidade muito significativa de dados, o *dataset* foi dividido em 556 partes para maximizar o paralelismo e diminuir a demanda por memória RAM. O problema envolve a análise de 10 caracte-

rísticas do solo de interesse, cada qual em 6 profundidades diferentes. Para os efeitos dessa avaliação, basta saber que são necessárias a execução de $10 \times 6 \times 556 = 33360$ tarefas computacionais (*jobs*). Ainda, o projeto avaliou três versões da solução, com diferenciação nos parâmetros de cada uma. Assim, ao final, foram executados 100.080 *jobs*. Este é um problema típico em que *clusters* de computadores são adequados e diminuem significativamente o tempo necessário para a obtenção dos resultados (Vetter, 2015).

A outra análise se deu em torno das necessidades de armazenamento de grandes volumes de dados. Basicamente, questionou-se quanto à possibilidade e qual seria o custo para se manter em nuvem os dados dos projetos que hoje estão presentes no sistema HPC do LMB. A capacidade de armazenamento necessária seria da ordem de 100TB.

Relato de resultados

As simulações envolvendo a ferramenta GROMACS geraram as informações presentes na Tabela 1. Note-se que se tratam de dados reais de uso e bilhetagem, pois de fato a análise foi executada na nuvem. Estes valores foram comparados com as estimativas de custos geradas pela plataforma de nuvem e pode-se constatar que estavam completamente de acordo com as estimativas, permitindo-se assim, usar as estimativas para compreender o custo de outras análises sem a necessidade de realmente executá-las. Os valores estimados fornecidos pela plataforma são apresentados nas Tabelas 2, 3 e 4.

Tabela 1. Informações referentes a valores gastos em recursos que realmente foram instanciados na nuvem. Nestes casos, a modalidade utilizada foi o *pay-per-use*.

Recursos	Unidade	Preço unitário (\$)	Período de bilhetagem	Uso (h)	Amount (USD\$)
High IO 40GB	Dollar/1Hour*GB	0,0002	"Jan 2021"	0	0
			"Fev 2021"	0	0
			"Mar 2021"	324,31	2,60
			"Abr 2021"	720	5,76
			"Mai 2021"	744	5,95
			"Jun 2021"	605	4,84
High IO 1000GB	Dollar/1Hour*GB	0,0002	Jan 2021	31,83	6,37
High IO 10000GB	Dollar/1Hour*GB	0,0002	May 2021	237,93	475,87
Computing-accelerated pi 2.2xlarge.4 8vCPU 32GB linux	Dollar/1Hour	1,158	"Jan 2021"	12,63	14,62
			"Fev 2021"	213,68	247,44
			"Mar 2021"	118,53	137,25
			"Abr 2021"	0.00	0,00
			"Mai 2021"	237,77	275,34

continua...

dynamic BGP fixed bandwidth 5Mbps	Dollar/1Hour	0,065	"Jan 2021"	408,30	26,54
			"Fev 2021"	672	43,68
			"Mar 2021"	744	48,36
			"Abr 2021"	720	46,80
			"Mai 2021"	744	48,36

Tabela 2. Valores estimados para a instanciação de Elastic Cloud Servers nas categorias Pay-per-use, Monthly e Spot Price

		4GB	8GB	16GB	32GB	64GB	128GB	256GB	512GB
ECS - Pay-per-use	2 vCPUs	\$0,101 USD/h	\$0.122 USD/h						
	4 vCPUs		\$0.194 USD/h	\$0.236 USD/h					
	8 vCPUs			\$0.38 USD/h	\$0.464 USD/h				
	16 vCPUs				\$0.752 USD/h	\$0.92 USD/h			
	32 vCPUs					\$1.668 USD/h	\$1.832 USD/h		
	64 vCPUs						\$2.988 USD/h	\$3.656 USD/h	\$5.118 USD/h
ECS - Monthly	2 vCPUs	\$62.08 USD	\$70.98 USD						
	4 vCPUs		\$120.58 USD	\$138.28 USD					
	8 vCPUs			\$242.38 USD	\$279.38 USD				
	16 vCPUs				\$481.08 USD	\$555.08 USD			
	32 vCPUs					\$885.48 USD	\$976.08 USD	\$1,185.98 USD	
	64 vCPUs							\$1,882.28 USD	\$2,368.28 USD
ECS - Spot Price	2 vCPUs	\$0.026 USD/h	\$0.032 USD/h						
	4 vCPUs		\$0.046 USD/h	\$0.058 USD/h					
	8 vCPUs			\$0.084 USD/h	\$0.108 USD/h				
	16 vCPUs				\$0.16 USD/h	\$0.21 USD/h			
	32 vCPUs					\$0.314 USD/h		\$2.568 USD/h	
	64 vCPUs						\$0.62 USD/h		

Tabela 3. Valores estimados para a comunicação.

Recurso	Preço (USD\$)
dynamic BGP fixed bandwidth 5Mbps ⁽¹⁾	0,065
dynamic BGP fixed bandwidth 10Mbps ⁽¹⁾	0,265
dynamic BGP fixed bandwidth 20Mbps ⁽¹⁾	0,665
dynamic BGP fixed bandwidth 50Mbps ⁽¹⁾	1,865
Dynamic BGP traffic ⁽²⁾	0,153

⁽¹⁾Nesta modalidade o custo é estabelecido em função do tempo de uso (por hora), independentemente da quantidade de dados transferida. Neste caso, o custo será maior quanto mais largura de banda se utilize.

⁽²⁾Nesta modalidade o custo de uso é estabelecido em função da quantidade de dados transferida, em GB, independentemente da largura de banda utilizada.

Tabela 4. Valores estimados para o armazenamento de dados. A formação de custos se dá considerando o tempo de uso e a capacidade de armazenamento usada.

Recurso	Preço (UDS\$)
High IO	0,0002
General Purpose SSD	0,00025
Ultra-high I/O	0,0004

Preço cobrado em dólar por GB a cada hora de armazenamento.

Na análise envolvendo o desempenho do ambiente em nuvem para o problema de mapeamento de solos, considerou-se três cenários distintos: 1) cenário envolvendo máquinas virtuais menores que poderiam processar apenas um *job* por vez; 2) cenário envolvendo máquinas virtuais maiores que poderiam processar dezenas de *jobs* simultaneamente; e 3) envolvendo máquinas virtuais alocadas segundo a modalidade “ECS Spot Price”, em que os custos são bem menores, mas não se tem a garantia de que os recursos computacionais continuarão instanciados por um longo período. Caso a plataforma necessite dos recursos para atender a outros clientes, é possível que o processamento seja interrompido. Para este problema em particular, isto não seria uma desvantagem significativa e, por esta razão, se analisou a questão de custos nesta modalidade.

A classificação de solos exige a execução de 100.080 *jobs* que precisam de um tempo médio de processamento de 0,74 h cada um. Individualmente os *jobs* utilizam apenas um processador e não há vantagem em uma máquina *multicore*. Em relação à memória, a demanda em média é da ordem de 8GB, sendo que a variação é ampla, indo de 1GB até um máximo de 15GB por *job*.

Assim, considerando as opções disponíveis na plataforma, o cenário 1 exigiria máquinas virtuais com 16GB de memória e 4 vCPUs. Há uma subutilização do recurso, pois as 4 vCPUs não seriam necessárias, mas no cenário com menos memória não é possível evitar este fator.

Para o cenário 2, escolheu-se máquinas virtuais com 512GB de memória RAM (as maiores disponíveis) e 64 vCPUs. Nestas máquinas virtuais (VMs) é possível a execução de 40 *jobs* simultaneamente.

Para o cenário 3, considerou-se máquinas virtuais com 128GB de memória e 64 vCPUs. Isto porque no cenário 3 se trabalhou com a modalidade de alocação por “spot price” e neste caso, não havia disponibilidade para uso das máquinas virtuais com mais memória.

Por fim, é preciso esclarecer que para se obter processamento de alto desempenho é necessário o uso de *clusters*. A plataforma em nuvem não oferece uma opção pré-configurada, mas possibilita a instanciação de diversas máquinas virtuais similares com facilidade. A instalação de um software

gerenciador de *workflow*, tais como *Sun Grid Engine (SGE)*, *Portable Batch System (PBS)*, ou Torque possibilitaria que dezenas dessas máquinas virtuais trabalhassem em conjunto na solução.

Com base nestas informações e considerando os custos estimados apresentados na Tabela 2, é possível calcular o tempo e os custos para se resolver o problema na plataforma em nuvem. Como exemplo, considere a instanciação de 20 VMs no cenário 1. Neste caso, o cluster executaria, em média, $20/0.74=27.02$ jobs por hora. Para a análise completa, com os 100.080 jobs, seriam necessárias **3702,96 h (154,29 dias) de trabalho do cluster**. O custo total ficaria em $20*3702,96*0,236=US\$17.477,97$. A Tabela 5 apresenta o panorama completo da análise de processamento, com o tempo e o custo dos três cenários com diferentes quantidades de Vms.

Tabela 5. Custos estimados para o processamento da análise de classificação de solos em diferentes cenários de alocação de recursos na nuvem.

	vCPUs	Mem (GB)	VMs ⁽¹⁾	Jobs/VM	Preço unitário por hora (USD\$)	Tempo	Custo (USD\$)
Cenário 1	4	16	20	1	0,236	3702,96 h 154,29 dias	17477,97
	4	16	50	1	0,236	1481,18 h 61,72 dias	17477,97
	4	16	100	1	0,236	740,59 h 30,85 dias	17477,97
	4	16	1000	1	0,236	74,06 h 3,09 dias	17477,97
Cenário 2	64	512	20	60	5,118	61,71 h 2,75 dias	6317,25
	64	512	50	60	5,118	24,69 h 1,03 dias	6317,25
	64	512	100	60	5,118	12,34 h 0,52 dias	6317,25
Cenário 3	64	128	20	15	0,62	246,86 h 10,29 dias	3061,11
	64	128	50	15	0,62	98,75 h 4,12 dias	3061,11
	64	128	100	15	0,62	49,37 h 2,05 dias	3061,11
	64	128	1000	15	0,62	4,95 h 0,20 dias	3061,11

⁽¹⁾VMs = quantidade de máquinas virtuais

Por fim, é necessário considerar o armazenamento e o transporte dos dados gerados na análise para uma avaliação completa. O problema de geração de mapas do solo tem um input de 65GB, uma geração de aproximadamente 5TB de dados intermediários e um output de algumas dezenas de imagens (mapas de solo) que podem ser desconsideradas para efeitos de estimativa de custos. Assim, uma boa estimativa dos custos envolvidos será conseguida considerando-se o transporte de 65GB de dados e o armazenamento de uma massa de 5TB de dados. O transporte de 65GB de dados teria um valor muito baixo, USD\$9.95 e demandaria um tempo de 3 h. Um espaço de armazenamento de 5TB custaria USD\$24.00 por dia. Assim, a depender do cenário escolhido, o custo de armazenamento variaria. Por exemplo, na opção 1 do cenário 1 são 154 dias de processamento e o custo seria da ordem de USD\$3696.00 com o armazenamento. Já na opção 4 do cenário 3, o tempo de processamento seria de menos de 1 dia, de maneira que o custo com o armazenamento de dados intermediários se tornaria irrisório.

Com relação ao armazenamento de longo prazo de grandes volumes de dados, considerou-se o caso do LMB para um estudo. Apesar de este montante variar no tempo, pode-se dizer que uma mé-

dia de 100TB armazenados constantemente é uma boa estimativa para a demanda. A Tabela 4 informa que o custo seria de USD\$ 0,0002 dólar/GB a cada hora de armazenamento. Para manter-se 100TB armazenados por ano, tem-se um **custo de $100 \times 1024 \times 0,0002 \times 24 \times 365 = \text{USD\$}179404.80$** . Para se avaliar se este valor estimado faz sentido, fez-se uso de um espaço de armazenamento de 10TB durante 10 dias, e, como se pode ver no Tabela 1, de fato o valor estimado condiz com os valores efetivamente utilizados.

Discussão dos resultados

O processamento envolvendo GPUs na nuvem tem um custo viável e se apresenta como uma excelente solução. As GPUs mais modernas e com um poder computacional mais elevado são equipamentos de custo elevado. Além desse fator, no geral, usuários que demandam tal recurso utilizarão o equipamento por um período e posteriormente transcorrerá um grande hiato até a próxima ocasião de uso. Definitivamente, a aquisição de GPUs profissionais que não sejam compartilhadas não se mostra um investimento adequado para a maioria dos casos de uso.

O uso em nuvem tem a vantagem de permitir o acesso a GPUs sempre modernas, além de possibilitar o escalonamento em ocasiões que demandam mais de uma GPU. Caso haja um problema que demande considerável poder computacional, e o algoritmo e ou o *dataset* possibilitem o trabalho de diversas GPUs; no ambiente em nuvem isto se torna economicamente viável, pois o gasto ocorrerá efetivamente durante o tempo de uso.

Sendo assim, pode-se considerar que os resultados obtidos nos ensaios de simulação com o GROMACS indicam fortemente o uso de GPUs em nuvem.

Já o ensaio envolvendo o mapeamento de solos que testou o processamento em CPUs trouxe algumas revelações importantes que não permitem a mesma conclusão para todo e qualquer uso de CPUs. Pode-se constatar uma variação significativa nos custos de processamento em função da arquitetura escolhida. O fator mais relevante para explicar esta ocorrência está relacionado com o mal aproveitamento dos recursos de processamento na primeira configuração. Assim, quando se trabalhou com uma máquina mais apropriada para o tipo de problema com o qual se estava lidando, os custos financeiros melhoraram significativamente. E é justamente este um ponto de bastante preocupação, pois diferentemente de aplicações da TI tradicional, que permitem uma identificação bastante precisa dos recursos computacionais necessários, no caso de aplicações científicas é bem mais difícil fazer um mapeamento a priori.

E por fim, como a modalidade de alocação dos recursos poderia ser a “*Spot Price*”, que não garante que as VMs estarão sempre disponíveis, mas que para o problema de processamento em análise não traria nenhum impacto significativo caso os equipamentos viessem a ser desligados, conseguiu-se reduzir os custos ainda mais.

Desta forma, o uso de uma infraestrutura computacional em nuvem para processamento de alto desempenho exigirá uma ação administrativa mais intensa. É necessário que haja uma interação grande entre os profissionais que administram e alocam os recursos, com os especialistas da área que conhecem o problema e demandam os recursos. Esta seria a única maneira de se configurar ambientes em que o uso dos recursos se desse da forma mais racional possível.

Por fim, a avaliação do armazenamento de grandes volumes de dados no ambiente em nuvem mostrou-se inviável economicamente neste momento. O custo para uso pelo período de um ano

excedeu em muito o investimento em um *storage*. Mesmo considerando as despesas com refrigeração e energia, ainda seria uma vantagem grande ter os dados alocados em infraestrutura própria.

Conclusões

As demandas computacionais em problemas de computação científica são bastante amplas. A infraestrutura em nuvem se mostrou muito adequada para alguns tipos de demandas e inviável para outros. Desta maneira, não seria possível uma situação em que todos os recursos computacionais da Empresa voltados para processamento e armazenamento de dados científico fossem transferidos para uma plataforma em nuvem. Por outro lado, ficou evidente que alguns problemas se beneficiam consideravelmente de uma solução envolvendo a nuvem.

Assim, a indicação é que a Empresa explore uma solução híbrida, mantenha ainda seu parque computacional próprio, mas inicie projetos que envolvam o uso dos recursos computacionais em nuvem naquelas situações em que isso seja vantajoso. Para as demandas por processamento em GPUs esta decisão poderia ser quase automática. No geral, em função da natureza desta demanda, a solução em nuvem será sempre vantajosa.

Para o processamento envolvendo CPUs será necessária uma avaliação caso a caso, e a seleção daqueles que têm vantagens de rodarem na nuvem. Pensando em uma solução mista, o processamento em CPUs na nuvem poderia se dar por meio de transbordamentos, ou seja, naqueles momentos em que houvesse uma demanda superior à capacidade da infraestrutura local.

Em relação ao armazenamento de grandes volumes de dados na nuvem, o presente estudo indicou que esta alternativa não é viável economicamente. A equipe de TI deve se manter atenta a possíveis propostas futuras, mas por ora a melhor solução ainda é a aquisição de equipamentos de armazenamento próprio.

Referências

GARTNER. **Gartner forecasts worldwide public cloud end-user spending to reach nearly \$600 billion in 2023:** inflationary pressures creating a push and pull effect for cloud spending. Stamford, 2022. Disponível em: <https://www.gartner.com/en/newsroom/press-releases/2022-10-31-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-600-billion-in-2023>. Acesso em: 2 nov. 2022.

PRONK, S.; PÁLL, S.; SCHULZ, R.; LARSSON, P.; BJELKMAR, P.; APOSTOLOV, R.; SHIRTS, M. R.; SMITH, J. C.; KASSON, P. M.; VAN DER SPOEL, D.; BERK, H.; LINDAHL, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics**, v, 29, n. 7, p. 845-854, Apr. 2013. DOI: [10.1093/bioinformatics/btt055](https://doi.org/10.1093/bioinformatics/btt055).

VETTER, J. S. (ed.). **Contemporary high performance computing:** from petascale toward exascale. Boca Raton: CRC Press, 2015. v. 2. (Computational science series).

