



**INSTITUTO FEDERAL**

Goiás

Câmpus Goiânia

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS

PRÓ-REITORIA DE ENSINO

CÂMPUS GOIÂNIA

BACHARELADO EM SISTEMAS DE INFORMAÇÃO

GABRIEL DIAS DOS SANTOS

WADSON ROCHA LEMOS

# **Uma ferramenta para análise da latência de comunicação em plataformas de nuvem**

Goiânia, 2022.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS

PRÓ-REITORIA DE ENSINO

CÂMPUS GOIÂNIA

BACHARELADO EM SISTEMAS DE INFORMAÇÃO

GABRIEL DIAS DOS SANTOS

WADSON ROCHA LEMOS

# **Uma ferramenta para análise da latência de comunicação em plataformas de nuvem**

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

**Orientador:** Prof. Dr. Raphael de Aquino Gomes

Goiânia, 2022.

S237f Santos, Gabriel Dias dos.

Uma ferramenta para análise da latência de comunicação em plataformas em nuvem / Gabriel Dias dos Santos; Wadson Rocha Lemos. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, 2022.  
42f.

Orientação: Prof. Dr. Raphael de Aquino Gomes.

TCC (Trabalho de Conclusão de Curso) – Curso Bacharelado em Sistemas de Informação, Instituto Federal de Educação, Ciência e Tecnologia de Goiás.  
Inclui apêndice.

1. Computação em nuvem. I. Lemos, Wadson Rocha (coautoria). II. Gomes, Raphael de Aquino (orientação). III. Instituto Federal de Educação, Ciência e Tecnologia de Goiás. IV. Título.

CDD 004.678 2

Ficha catalográfica elaborada pela bibliotecária Lana Cristina Dias Oliveira CRB1/ 2.631  
Biblioteca Professor Jorge Félix de Souza,  
Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.



**INSTITUTO FEDERAL**  
Goiás

**MINISTÉRIO DA EDUCAÇÃO**  
**SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA**  
**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**SISTEMA INTEGRADO DE BIBLIOTECAS**

**TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO  
NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

**Identificação da Produção Técnico-Científica**

- |  |   |
|--|---|
| <input type="checkbox"/> Tese  | <input type="checkbox"/> Artigo Científico              |
| <input type="checkbox"/> Dissertação                                 | <input type="checkbox"/> Capítulo de Livro              |
| <input type="checkbox"/> Monografia – Especialização                 | <input type="checkbox"/> Livro                          |
| <input checked="" type="checkbox"/> TCC - Graduação                  | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ |   |

Nome Completo do Autor: **Gabriel Dias dos Santos**

Matrícula: **20162011090010**

Título do Trabalho: Uma ferramenta para análise da latência de comunicação em plataformas de nuvem

**Autorização - Marque uma das opções**

1. ☒ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ☐ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data \_\_\_\_/\_\_\_\_/\_\_\_\_ (Embargo);
3. ☐ Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

- ☐ O documento está sujeito a registro de patente.  
☐ O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.  
☐ Outra justificativa: \_\_\_\_\_

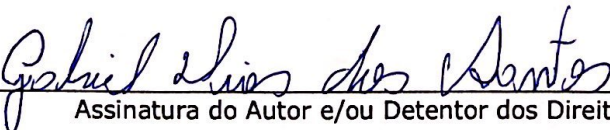
**DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA**

O/A referido/a autor/a declara que:

- i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Goiânia-GO  
Local

12/07/2022.  
Data

  
Assinatura do Autor e/ou Detentor dos Direitos Autorais



**INSTITUTO FEDERAL**  
Goiás

MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
SISTEMA INTEGRADO DE BIBLIOTECAS

### **TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

#### **Identificação da Produção Técnico-Científica**

- |  |   |
|--|---|
| <input type="checkbox"/> Tese  | <input type="checkbox"/> Artigo Científico              |
| <input type="checkbox"/> Dissertação                                 | <input type="checkbox"/> Capítulo de Livro              |
| <input type="checkbox"/> Monografia – Especialização                 | <input type="checkbox"/> Livro                          |
| <input checked="" type="checkbox"/> TCC - Graduação                  | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ |   |

Nome Completo do Autor: Wadson Rocha Lemos

Matrícula: 20192011090054

Título do Trabalho: Uma ferramenta para análise da latência de comunicação em plataformas de nuvem

#### **Autorização - Marque uma das opções**

1. ☒ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ☐ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data \_\_\_\_/\_\_\_\_/\_\_\_\_ (Embargo);
3. ☐ Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

- ☐ O documento está sujeito a registro de patente.  
☐ O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.  
☐ Outra justificativa: \_\_\_\_\_

#### **DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA**

O/A referido/a autor/a declara que:

- i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Goiânia, 22 / 07 / 2022.  
Local Data

Assinatura do Autor e/ou Detentor dos Direitos Autorais





**INSTITUTO FEDERAL**  
Goiás

MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS  
CÂMPUS GOIÂNIA

GABRIEL DIAS DOS SANTOS

WADSON ROCHA LEMOS

## Uma Ferramenta para Análise da Latência de Comunicação em Plataformas de Nuvem

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação do Instituto Federal de Educação, Ciência e Tecnologia de Goiás, como parte das exigências para obtenção do Título de Bacharel em Sistemas de Informação.

Goiânia, 08 de julho de 2022.

### BANCA EXAMINADORA

**Prof. Dr. Raphael de Aquino Gomes**

Orientador - IFG/Câmpus Goiânia

**Profª. Me. Luciana Nishi**

Membro Externo da Banca Examinadora - Universidade de Brasília

**Prof. Me. Carlos Augusto da Silva Cabral**

Membro Interno da Banca Examinadora - IFG/Câmpus Goiânia

Documento assinado eletronicamente por:

- Carlos Augusto da Silva Cabral, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 14/07/2022 08:54:34.
- Luciana Nishi, Luciana Nishi - Outros - Unb (00038174000143), em 11/07/2022 11:36:30.
- Raphael de Aquino Gomes, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 11/07/2022 09:22:48.

Este documento foi emitido pelo SUAP em 11/07/2022. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifg.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 302148

Código de Autenticação: 87efdf7f98



# Dedicatória

Eu **Gabriel Dias dos Santos**, dedico este trabalho a Deus por nortear a minha vida e sempre estar comigo nas dificuldades, aos meus pais pelo exemplo de força e apoio, à minha noiva pelo incentivo, carinho e amor, à minha avó pelo cuidado e carinho comigo.

# Agradecimentos

Eu, **Gabriel Dias dos Santos**, agradeço primeiramente a Deus por me ajudar a ultrapassar todos os obstáculos encontrados ao longo do curso e deste TCC. Aos meus pais Wagner e Viviane por todo apoio e incentivo nos momentos difíceis, à minha irmã Débora e minha noiva Isabella pelo apoio e compreensão, ao meu orientador Prof. Dr. Raphael de Aquino Gomes por todos os ensinamentos e a todos os professores desta instituição que de alguma forma contribuíram para o meu ensinamento e crescimento.

Eu, **Wadson Rocha Lemos**, agradeço primeiramente a Deus por me ajudar a ultrapassar todos os obstáculos encontrados ao longo do curso e deste TCC. À minha mãe e à minha família.



# Resumo

**Título:** Uma ferramenta para análise da latência de comunicação em plataformas de nuvem

**Autores:** Gabriel Dias dos Santos e Wadson Rocha Lemos

**Orientador:** Dr. Raphael de Aquino Gomes

Dada a diversidade de opções ao escolher uma região de implantação de serviços em nuvem surge a necessidade de ferramentas que permitam analisar as variáveis envolvidas e apoiar a tomada de decisão. Diante disso, o presente trabalho visa apresentar uma ferramenta capaz de relatar estatisticamente o tempo de latência nas regiões de provedores de nuvem. Para demonstrar a proposta foi desenvolvido um protótipo que constitui um sistema web distribuído capaz de exibir resultados confiáveis e a coleta de latência nas regiões da AWS. Também é apresentada uma avaliação sobre como os dados obtidos com a ferramenta podem auxiliar na tomada de decisão para o cenário considerado.

## Palavras-chave

Computação em Nuvem, Latência, Tomada de decisões

# Abstract

**Title:** A tool for analyzing communication latency on cloud platforms

**Authors:** Gabriel Dias dos Santos and Wadson Rocha Lemos

**Advisor:** Dr. Raphael de Aquino Gomes

Given the diversity of options when choosing a region to deploy cloud services, tools need to analyze the variables involved and support decision-making. Therefore, the present work aims to present a tool capable of presenting statistics on latency time in the regions of cloud providers. A prototype was developed aiming to demonstrate the proposal. It constitutes a distributed web system capable of displaying reliable results and collecting latency in AWS regions. The work also evaluates how the data obtained with the tool can help in decision-making for the considered scenario.

## Keywords

Cloud Computing, Latency, Decision Making

# Lista de Figuras

|     |  |    |
|-----|--|----|
| 2.1 | Principais atrasos no envio de dados em rede. . . . .            | 21 |
| 2.2 | Regiões disponibilizadas pela AWS. . . . .                       | 22 |
| 2.3 | Latência versus preços nas regiões da AWS. . . . .               | 23 |
| 3.1 | Modelagem de Processo. . . . .                                   | 25 |
| 3.2 | Tela inicial da ferramenta. . . . .                              | 26 |
| 3.3 | Tela mostrando as regiões consideradas. . . . .                  | 26 |
| 3.4 | Parte superior da tela de <i>Dashboard</i> . . . . .             | 27 |
| 3.5 | Parte inferior da tela de <i>Dashboard</i> . . . . .             | 27 |
| 3.6 | Coefficiente de variação para dados de 27/06/2021. . . . .       | 29 |
| 3.7 | Média de latência por região ofertada pela AWS. . . . .          | 29 |
| 3.8 | Modelagem da base de dados. . . . .                              | 32 |
| 4.1 | Custo mensal para cada uma das empresas. . . . .                 | 33 |
| 4.2 | Volume de dados transmitidos para cada uma das empresas. . . . . | 35 |

# Lista de Tabelas

|     |   |    |
|-----|---|----|
| A.1 | Instâncias utilizadas pela empresa A. . . . . | 40 |
| A.2 | Instâncias utilizadas pela empresa B. . . . . | 41 |
| A.3 | Instâncias utilizadas pela empresa C. . . . . | 41 |
| A.4 | Instâncias utilizadas pela empresa D. . . . . | 42 |

# Lista de Códigos de Programas

|     |                               |           |    |
|-----|-------------------------------|-----------|----|
| 3.1 | funcao_intervaloConfianca()   | . . . . . | 30 |
| 3.2 | funcao_coeficienteVariancia() | . . . . . | 30 |
| 3.3 | funcao_mediaSimples()         | . . . . . | 30 |
| 3.4 | funcao_mediaSimplesHora()     | . . . . . | 31 |

# Lista de Abreviaturas e Siglas

|       |  |
|-------|--|
| API   | <i>Application Programming Interface</i>   |
| AWS   | Amazon Web Services                        |
| CSS   | <i>Cascading Style Sheets</i>              |
| EC2   | <i>Elastic Cloud Computing</i>             |
| EUA   | Estados Unidos da América                  |
| HTML  | <i>HyperText Markup Language</i>           |
| HTTPS | <i>Hyper Text Transfer Protocol Secure</i> |
| ICMP  | <i>Internet Control Message Protocol</i>   |
| IP    | <i>Internet Protocol</i>                   |
| JSON  | <i>JavaScript Object Notation</i>          |
| RTT   | <i>Round Trip Time</i>                     |
| SGBD  | Sistema Gerenciador de Bancos de Dados     |
| TI    | Tecnologia da Informação                   |
| vCPU  | <i>virtual Computing Processing Unit</i>   |

# Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>   | <b>15</b> |
| 1.1      | Objetivos   | 17        |
| 1.2      | Organização do Trabalho   | 17        |
| <b>2</b> | <b>REFERENCIAL TEÓRICO</b>                                      | <b>18</b> |
| 2.1      | Computação em Nuvem   | 18        |
| 2.1.1    | Modelo de cobrança  | 20        |
| 2.2      | Latência de comunicação   | 21        |
| 2.2.1    | O impacto da latência na seleção de serviços em nuvem           | 22        |
| <b>3</b> | <b>SOLUÇÃO PROPOSTA</b>   | <b>24</b> |
| 3.1      | Projeto da ferramenta   | 24        |
| 3.2      | Implementação   | 28        |
| 3.2.1    | Persistência de dados   | 31        |
| <b>4</b> | <b>AVALIAÇÃO</b>  | <b>33</b> |
| <b>5</b> | <b>CONSIDERAÇÕES FINAIS</b>                                     | <b>36</b> |
| 5.1      | Trabalhos futuros   | 36        |
|          | <b>REFERÊNCIAS BIBLIOGRÁFICAS</b>                               | <b>37</b> |
|          | <b>APÊNDICE A – MÁQUINAS VIRTUAIS CONSIDERADAS NA AVALIAÇÃO</b> | <b>40</b> |



## Introdução

---

Computação em nuvem (em inglês, *Cloud Computing*) é um termo utilizado para representar a disponibilidade sob demanda e medida de recursos computacionais. Especificamente, o armazenamento de dados e capacidade de computação, dentre outros serviços, sem o gerenciamento ativo do utilizador (MELL; GRANCE *et al.*, 2011).

Uma das características da computação em nuvem é a possibilidade de acesso facilitado aos serviços, através da internet. Isso permite ao usuário acessar uma grande quantidade de serviços e de várias aplicações em qualquer lugar e independente da plataforma.

De acordo com Convergência Digital (2019), o mercado de nuvem tem potencial de crescer cerca de 55% até o fim deste ano, com uma projeção de US \$266,4 bilhões. Porém, apesar de todo crescimento e de toda evolução tecnológica um dos principais desafios enfrentados é a latência<sup>1</sup> de acesso aos serviços, sobretudo no caso em que a seleção da região do provedor é realizada tendo como foco a otimização de custos. Isso acontece porque os preços de cada serviço em nuvem variam de acordo com cada provedor e a região em que são alocados.

A latência de comunicação é uma variável que tem que ser considerada nas decisões de uso de serviços em nuvem, uma vez que impacta diretamente no negócio do empreendedor. Como por exemplo, em uma plataforma da bolsa de valores, onde são negociados milhões de dólares por dia, cada segundo de atraso impactaria diretamente na venda de uma ação, assumindo a latência um papel extremamente importante (SILVA, 2018).

Segundo o estudo publicado pela Universidade Estadual do Oeste do Parana (UNIOESTE) (2022), o modelo de alocação em zonas de disponibilidade pode elevar significativamente a latência média entre o cliente e o servidor. Isto acontece porque há uma disponibilidade grande de opções que devem ser consideradas. Como exemplo, a Amazon Web Services (AWS) (Amazon Web Services, 2022a) que possui atualmente 26

---

<sup>1</sup>O tempo que um dado leva para ser enviado pela rede de um computador de origem até seu destino final é chamado de latência (MINHACONEXÃO, 2020).

regiões espalhadas pelo globo em diferentes pontos estratégicos, com planos de expansão de seus serviços para novas regiões.

No contexto nacional este aspecto é ainda mais crítico visto que apesar do fato do Brasil possuir regiões dos principais provedores de nuvem, a alocação de recursos nessas regiões não é recomendada em alguns cenários uma vez que devido aos altos custos tributários o país possui o preço mais elevado dentre as regiões disponíveis (ISLAM; ISLAM, 2018).

Apesar da relevância, as ferramentas existentes não permitem analisar de forma adequada as informações sobre latência no acesso a serviços em nuvem. O projeto Cloud Ping (2022) foi desenvolvido para coletar informações das regiões da AWS, executando pings<sup>2</sup> nas regiões. O objetivo é identificar qual a região com menor latência através de medições periódicas. Contudo, não demonstram se fazem uso de métricas que validam a confiabilidade das informações.

O projeto gcping (2022) apresenta métricas de latência referente ao Google Cloud Platform (2022a). A ferramenta faz solicitações *Hyper Text Transfer Protocol Secure* (HTTPS) através do navegador para as regiões, determinando valores relativos de latência. Contudo, não apresenta o histórico das medições anteriores e não demonstra nenhum indício de métricas que comprovam a confiabilidade dos dados. Por utilizarem requisições HTTPS, podem ocorrer exceções no roteamento devido a alguma má implementação no código.

O AWS PING Github - ekalinin (2022) é um outro projeto que analisa dados de latência em regiões da AWS visando identificar a região mais adequada levando em consideração a baixa latência. Assim como o Cloud Ping, utiliza requisições HTTPS para extrair métricas.

Apesar da existência destas ferramentas, não há uma solução que não seja específica de um provedor de nuvem. A ausência de informações estatísticas sobre as medições realizadas também é outro problema apresentado. Diante disso, este trabalho apresenta uma proposta de ferramenta capaz de mensurar e calcular o tempo de latência entre regiões de plataformas de nuvem e indicar antes mesmo do usuário efetuar a contratação do serviço qual região pode ser utilizada usando como parâmetro a baixa latência. O objetivo principal da ferramenta desenvolvida é auxiliar na tomada de decisões de pequenos e grandes projetos.

---

<sup>2</sup>ping é um utilitário que usa o protocolo ICMP para testar a conectividade entre equipamentos.

## 1.1 Objetivos

O objetivo geral deste trabalho de conclusão de curso é propor uma ferramenta que permita oferecer métricas e dados que auxiliem na tomada de decisão para escolha de regiões em plataformas de nuvem com base na baixa latência.

Este objetivo geral será atingido através dos objetivos específicos a seguir:

- Coletar e processar medidas de latência em regiões de plataformas de nuvem.
- Identificar regiões com latência tolerável.
- Disponibilizar métricas que validem a confiabilidade da informação apontada.
- Disponibilizar dados históricos de latências coletadas anteriormente.
- Fazer um balanço entre a latência e o custo dos serviços para cada região.

## 1.2 Organização do Trabalho

A seguir é apresentada uma visão geral do restante do texto.

O Capítulo 2 apresenta os principais conceitos relacionados a latência e computação em nuvem. Posteriormente é apresentado o estado da arte de pesquisas sobre o tema, destacando, os pontos falhos de cada uma delas e o que proposto aqui como solução para solucionar essas lacunas.

O Capítulo 3 visa demonstrar os resultados alcançados, descrevendo a ferramenta implementada e sua arquitetura.

O Capítulo 4 apresenta uma avaliação visando ilustrar como os resultados exibidos pela ferramenta podem auxiliar na tomada de decisão.

Por fim, no Capítulo 5 são apresentadas as considerações finais, com uma descrição dos trabalhos futuros.

## Referencial Teórico

---

Neste capítulo são abordados os temas e conceitos utilizados para o desenvolvimento da ferramenta proposta nesse trabalho. O objetivo é tornar claro o assunto tratado e discutir conceitos relevantes.

### 2.1 Computação em Nuvem

Para um bom entendimento sobre computação em nuvem, de acordo com a Amazon Web Services (2022b): Computação em nuvem é a entrega de recursos de Tecnologia da Informação (TI) sob demanda por meio da Internet com definição de preço de pagamento conforme o uso. Em vez do usuário comprar e manter *datacenters* e servidores físicos, é possível ter acesso a serviços que representam capacidade computacional, armazenamento expansível e outros recursos conforme a necessidade, usando os provedores de nuvem como Huawei Cloud (2022), Oracle Cloud (2022), Azure Cloud (2022), Amazon Web Services (2022b), Google Cloud Platform (2022b), dentre outros.

Existem grandes benefícios em utilizar computação em nuvem. Alguns desses benefícios listados pela Eco IT Segurança Digital (2022) são:

- **Custo e benefícios:** Quando se fala de computação em nuvem é preciso pensar nos gastos que serão evitados. O valor que se é investido é apenas referente ao que for usado. Adotando serviços em nuvem é possível eliminar gastos de capital com a compra de hardware e alguns softwares, e instalação e execução de *datacenters* locais.
- **Redução de tempo:** Com apenas alguns cliques de configuração é possível o cliente adquirir os recursos necessários para montar seu ambiente tecnológico, aliviando a tensão de planejar a capacidade que a equipe de gestão tecnológica possui. Com isso, o tempo necessário para disponibilizar recursos à equipe de desenvolvimento é reduzido de semanas para horas (em alguns casos até mesmo em minutos), resultando em um aumento considerável na produtividade de uma organização.

Não apenas o custo é poupado, mas o tempo necessário para desenvolvimento é drasticamente reduzido.

- **Ação em escala global:** um dos principais benefícios que a computação em nuvem proporciona é a capacidade de dimensionamento elástico. Ou seja, é possível adquirir a capacidade de fornecer os recursos exatos que serão utilizados pela companhia. Para um melhor entendimento, é como se os recursos computacionais como processador, memória RAM, armazenamento e largura de banda pudessem ser aumentados ou diminuídos conforme necessário, não importando a origem geográfica adotada. É possível facilmente implantar um aplicativo em várias regiões no mundo todo utilizando dos conceitos de sistemas distribuídos e computação em nuvem.
- **Aumento na produtividade:** Os *datacenters* locais exigem vários equipamentos e implementações, para configuração de hardware e software e muitas outras diversas tarefas de gerenciamento da TI que consomem tempo e produtividade. Utilizando de computação em nuvem é possível remover muitas dessas necessidades.
- **Desempenho:** Os grandes serviços de computação em nuvem são executados em uma rede mundial de *datacenters* seguros, a qual passa por melhorias o tempo todo com hardwares que possuem grande velocidade e desempenho. São diversos os benefícios através de um único *datacenter* corporativo oriundo da computação em nuvem, incluindo baixa latência para aplicativos e proporção de mais economia. Com isso, é possível eliminar as suposições para determinar a necessidade da capacidade de uma infraestrutura.
- **Segurança:** A maioria dos provedores de computação em nuvem oferecem um grande conjunto de políticas, tecnologias e controles que ajudam a fortalecer sua postura geral de segurança, sendo possível favorecer a proteção dos dados trafegados pela rede, aplicativos e a infraestrutura contra ameaças que possam comprometer o seu negócio.

A computação em nuvem é dividida em quatro categorias:

- **Nuvem Privada:** infraestrutura de nuvem é provisionada para uso exclusivo por uma empresa ou organização e é de total responsabilidade desta os custos que envolvem.
- **Nuvem Comunitária:** é uma infraestrutura de nuvem para uso exclusivo de uma comunidade específica de consumidores de organizações ou empresas.
- **Nuvem Pública:** a mais popular entre as infraestruturas de nuvem pois é provisionada para uso aberto pelo público em geral. De acordo com Bala *et al.* (2020), os principais provedores de serviços na nuvem pública são Amazon Web Services (2022a), Azure (2019), Google Cloud Platform (2022b).

- Nuvem Híbrida: a infraestrutura em nuvem é uma composição de duas ou mais categorias distintas de nuvem (privada, comunitária ou pública). Geralmente utilizam de serviços na nuvem privada para armazenamento de dados e os serviços de nuvem pública para outras funcionalidades.

### 2.1.1 Modelo de cobrança

Como os serviços de nuvem são praticamente iguais entre os principais fornecedores de nuvem, a cobrança também é semelhante. A cobrança por serviços geralmente é efetuada por tempo de uso, região que hospeda o serviço e tipo de licença de software associado.

Como exemplo, considerando o serviço *Elastic Cloud Computing* (EC2) da AWS, uma instância de máquina virtual da família T3a.medium, com 2 núcleos de *virtual Computing Processing Unit* (vCPU), 4 GB de memória RAM, 30 GB de SSD, sistema operacional Linux, na região da Virgínia - Estados Unidos da América (EUA), custa aproximadamente 20,23 dólares mensais. A mesma instância com sistema operacional Windows Server passa a custar aproximadamente 33,66 dólares mensais.

De acordo com a Amazon Web Services (2022c) usando as instâncias sob demanda, o cliente paga pela capacidade computacional por hora ou por segundo, dependendo das instâncias executadas. Não são necessários compromissos de longo prazo nem pagamentos antecipados. O cliente pode aumentar ou diminuir a capacidade computacional dependendo da necessidade do aplicativo e pagar apenas as taxas por hora especificadas para a instância utilizada.

A cobrança sob demanda é recomendada para aqueles usuários que preferem o custo baixo e a flexibilidade da nuvem sem nenhum pagamento adiantado ou compromisso de longo prazo. É bastante eficaz em aplicações com cargas de trabalho breves, com picos de utilização ou cargas imprevisíveis e que não podem ser interrompidas, como aplicativos sendo desenvolvidos ou testados pela primeira vez.

As instâncias *spot* representam outra forma de cobrança que permite solicitar capacidade computacional ociosa com desconto de até 90% em relação ao preço das instâncias sob demanda. As instâncias *spot* são recomendadas para aplicativos que têm períodos de início e de término flexíveis, pois são viáveis somente aos usuários com necessidades computacionais.

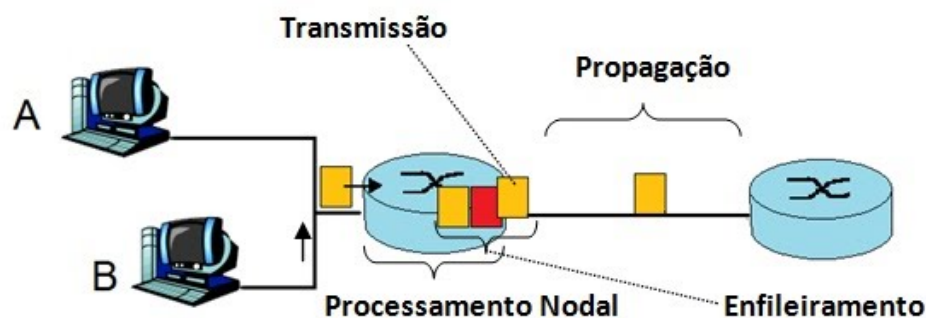
A reserva é um modelo de precificação flexível que oferece preços baixos no uso dos serviços em troca de um compromisso com uma quantidade consistente de uso (medido em dólar/hora) por um período de 1 a 5 anos.

## 2.2 Latência de comunicação

Latência é o termo usado para medir o tempo de resposta ao acessar um site ou o tempo de acesso a uma determinada aplicação. Esse termo tem ligação direta com a área de telecomunicações, onde a latência é definida como o tempo total necessário para um sinal viajar de um ponto a outro, geralmente de um transmissor através de uma rede para um receptor (DIGITAL, 2013).

A latência de comunicação é formada por um conjunto de ações que geram atraso (DIGITAL, 2013): fila, processamento, transmissão e propagação, conforme explicado a seguir e ilustrado na Figura 2.1:

- **fila:** é o tempo que um pacote fica esperando na rede para ser enviado.
- **processamento:** o tempo necessário para um equipamento processar os dados e decidir para onde enviá-los.
- **transmissão:** o tempo necessário para transportar os dados para o canal.
- **propagação:** o tempo necessário para que os dados se propaguem pelo canal de comunicação.



**Figura 2.1:** Principais atrasos no envio de dados em rede.

Fonte: (Google Sites, 2020)

O crescente interesse na medição de desempenho de rede fez com que surgisse uma série de infraestruturas de medição em larga escala, como o RIPE Atlas (2022), RIPE Test (2022), Project BISmark (2022) e Measurement Lab (2022).

Contudo, soluções complementares mais simples podem ser utilizadas através do próprio sistema operacional, como é o caso do programa `ping`. Este comando testa a conectividade de uma rede utilizando o protocolo *Internet Control Message Protocol* (ICMP). A palavra PING é a abreviação do termo em inglês “*Packet Internet Network Grouper*”, que significa algo como “Agrupador de Pacotes da Internet”. Esse comando basicamente envia dados para um destino e coleta métricas relativas à resposta.



A análise da latência de comunicação é necessária porque o provedor disponibiliza seus serviços em diferentes regiões, como ilustrado na Figura 2.2 para a AWS. Para aplicações sensíveis ao atraso cada milissegundo pode representar perdas e impactar o negócio.



**Figura 2.2:** *Regiões disponibilizadas pela AWS.*

Fonte: (Amazon Web Services, 2022a)

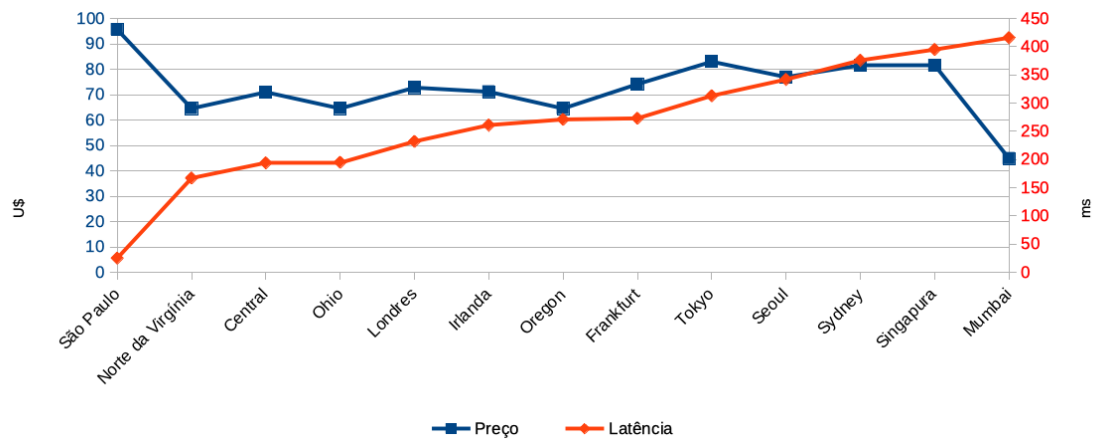
Além do comando `ping`, outras ferramentas podem ser utilizadas especificamente para este propósito como o Cloud Ping (2022), usada para medir latência de regiões da AWS. A ferramenta registra latência entre regiões utilizando o protocolo TCP, armazenando o RTT a cada 6 horas. Contudo, ele não permite análises estatísticas sobre as medidas coletadas.

### 2.2.1 O impacto da latência na seleção de serviços em nuvem

A latência é um dos fatores decisivos na hora de escolher a região de um provedor para utilizar seus serviços de nuvem, dado as múltiplas opções disponíveis com variação de latência e preços.

É sabido que, quanto mais longe um *host* está de seu servidor de acesso, maior é o tempo de propagação na rede, o que tende a tornar a latência maior. Como exemplo, para hospedar um jogo online que deve suportar vários jogadores, a baixa latência é de suma importância, sendo que uma região que proporcione esse benefício é o mais indicado.

Para ilustrar a importância da análise da latência em conjunto com o preço, a seguir é apresentada uma análise utilizando a ferramenta Cloud Ping. Foram coletadas informações de latência das regiões da AWS e foi utilizada a calculadora de preços da AWS para estimar o custo mensal do serviço EC2. Nesta análise foi considerada uma instância Linux com 4GB de RAM, 4 vCPU. Os resultados são apresentados na Figura 2.3.



**Figura 2.3:** Latência versus preços nas regiões da AWS.

Fonte: Elaborada pelos autores

Conforme pode ser observado na figura, utilizar somente um dos dois parâmetros (preço ou latência) não constitui uma boa alternativa, visto que eles são de certa forma inversamente proporcionais. A região de São Paulo possui a menor latência, por estar mais próxima, mas é a região mais cara para se alocar o serviço. Por outro lado, a região de Mumbai possui o preço mais acessível apesar de oferecer a maior latência. Dessa forma, cabe ao usuário encontrar um meio termo entre os dois parâmetros, o que motiva a existência de ferramentas que o auxiliem nesta tarefa.

---

## Solução Proposta

---

Diante da problemática tratada nos capítulos anteriores, foi proposta uma solução cuja arquitetura e desenvolvimento são descritos neste capítulo. A ferramenta desenvolvida permite coletar dados referentes à latência de regiões de provedores de nuvem. A partir deste dados são apresentadas métricas que permite avaliar a confiabilidade dos indicadores e informar ao usuário qual a região mais adequada para hospedagem de aplicações, levando em consideração a baixa latência. Com isso é possível realizar uma comparação entre custo e latência, possibilitando economia enquanto mantém o desempenho da aplicação

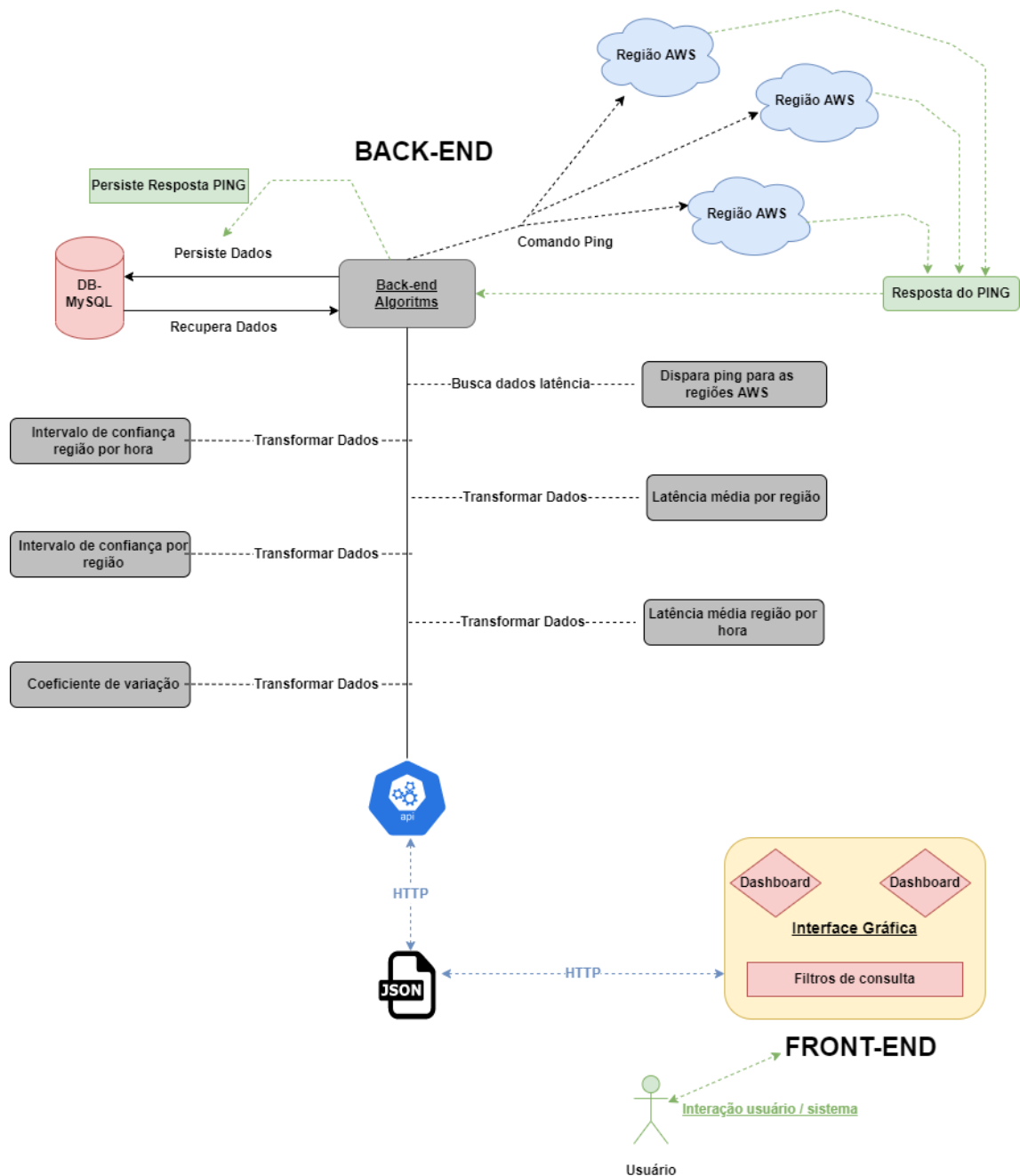
Apesar da ferramenta ter sido pensada inicialmente para suportar um conjunto amplo de provedores de nuvem, a implementação constitui um protótipo que considera apenas as regiões da AWS disponíveis em novembro de 2021. A escolha da AWS se deu por ser o principal provedor do mercado (BALA *et al.*, 2020), atendendo grande parte do mercado mundial.

### 3.1 Projeto da ferramenta

A ferramenta proposta possui uma arquitetura distribuída, composta por duas partes: *front-end* e *back-end*, conforme ilustrado na Figura 3.1.

O *back-end* é um microsserviço que foi desenvolvido em Java utilizando o *framework* Spring (2022), utilizando a base de dados MySQL (2022). O microsserviço é responsável pela coleta de dados das regiões da AWS e persistência na base de dados. Outra funcionalidade implementada é o tratamento de dados e o cálculo de medidas estatísticas sobre os mesmos, mostradas na Figura 3.1. A partir dos dados tratados é disponibilizado uma *Application Programming Interface* (API) que é utilizada pelo *front-end* através de requisições HTTP no formato *JavaScript Object Notation* (JSON).

O *front-end* representa a interface com a qual o usuário irá interagir. Este componente recebe os dados processados no *back-end*, apresentando-os de forma organizada e de fácil entendimento ao usuário. Ele foi desenvolvido utilizando JavaScript, TypeScript,



**Figura 3.1:** Modelagem de Processo.

Fonte: Elaborada pelos autores

*HyperText Markup Language* (HTML), *Cascading Style Sheets* (CSS) e a biblioteca React (2022).

O *front-end* é composto de três telas, como explicado a seguir.

A primeira tela, **Página inicial**, traz uma breve descrição sobre a ferramenta. Como mostrado na Figura 3.2, possui dois botões de direcionamento para as outras áreas da ferramenta, sendo elas o Dashboard e Regiões.

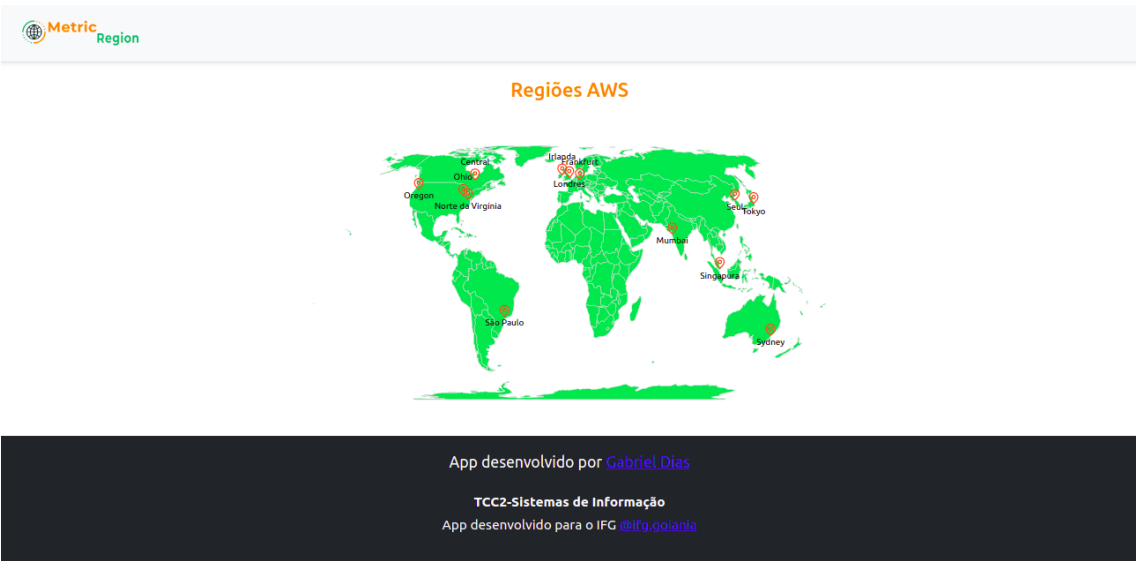
A segunda tela, **Regiões**, renderiza um mapa global identificando os pontos das



**Figura 3.2:** Tela inicial da ferramenta.

Fonte: Elaborada pelos autores

regiões considerados na ferramenta, conforme mostrado na Figura 3.3.



**Figura 3.3:** Tela mostrando as regiões consideradas.

Fonte: Elaborada pelos autores

Por fim, a terceira tela, ***Dashboard***, constitui a principal parte da ferramenta pois nela são apresentadas as informações sobre a latência. Na parte superior, mostrada na Figura 3.4, são apresentados gráficos relativos às métricas consideradas.

O gráfico de barras, à esquerda, traz informações sobre a média geral das regiões. Estes valores são calculados considerando a média aritmética sobre todas as medições realizadas. O gráfico de pizza, à direita, mostra o intervalo de confiança referente à variação de latência geral de cada região.

A parte inferior da tela de *Dashboard*, mostrada na Figura 3.5, apresenta uma tabela com as mesmas métricas consideradas na parte superior, mas considerando valores para cada hora. Os valores apresentados são customizados de acordo com o filtro aplicado na tela, sendo possível selecionar um período específico e retornar métricas deste período específico.



Figura 3.4: Parte superior da tela de Dashboard.

Fonte: Elaborada pelos autores

| Tabela de Métricas |            |                           |                         |                     |                       |
|--------------------|------------|---------------------------|-------------------------|---------------------|-----------------------|
| Selecione a data.  |            | Selecione a hora inicial. | Selecione a hora final. | Consultar           |                       |
| Cidade             | Data       | Horário                   | Média Latência          | Intervalo Confiança | Coefficiente Variação |
| São Paulo          | 2021-10-27 | 00:00:00-01:00:00         | 25.61                   | 1.83                | 27.01%                |
| Norte da Virgínia  | 2021-10-27 | 00:00:00-01:00:00         | 173.00                  | 6.20                | 13.51%                |
| Ohio               | 2021-10-27 | 00:00:00-01:00:00         | 190.52                  | 7.79                | 15.41%                |
| Central            | 2021-10-27 | 00:00:00-01:00:00         | 192.07                  | 9.05                | 17.77%                |
| Frankfurt          | 2021-10-27 | 00:00:00-01:00:00         | 243.59                  | 8.24                | 12.75%                |
| Oregon             | 2021-10-27 | 00:00:00-01:00:00         | 245.74                  | 5.68                | 8.72%                 |
| Londres            | 2021-10-27 | 00:00:00-01:00:00         | 251.68                  | 4.85                | 7.33%                 |
| Irlanda            | 2021-10-27 | 00:00:00-01:00:00         | 269.08                  | 10.70               | 15.12%                |
| Tokyo              | 2021-10-27 | 00:00:00-01:00:00         | 338.64                  | 9.27                | 10.32%                |
| Seul               | 2021-10-27 | 00:00:00-01:00:00         | 346.41                  | 4.32                | 4.71%                 |
| Sydney             | 2021-10-27 | 00:00:00-01:00:00         | 371.83                  | 8.82                | 8.94%                 |
| Mumbai             | 2021-10-27 | 00:00:00-01:00:00         | 372.14                  | 6.18                | 6.26%                 |
| Singapura          | 2021-10-27 | 00:00:00-01:00:00         | 402.84                  | 6.72                | 6.29%                 |

Figura 3.5: Parte inferior da tela de Dashboard.

Fonte: Elaborada pelos autores

## 3.2 Implementação

As medidas de latência foram coletadas utilizando o comando `ping`, através do pacote `net` integrante da biblioteca nativa de Java. Uma questão que precisou ser respondida é a frequência com que os dados precisariam ser coletados de forma a permitir uma análise confiável sobre a representatividade da média. Para determinar este parâmetro foi realizada uma análise inicial com medidas coletadas a cada 2 segundos durante 52 horas entre os dias 27 e 30 de junho de 2021. Inicialmente foi investigada a dispersão considerando dados de um mesmo dia (24 horas de medição). Para análise foi utilizada o coeficiente de variação, que permite a comparação de distribuições. O uso desta medida de dispersão foi considerado porque a comparação da dispersão de pelo simples confronto de seus desvios padrão nem sempre é suficiente, pois os valores de média das amostras ou populações podem estar bastante afastados (LAPPONI, 2004). A Equação 3-1 descreve seu cálculo:

$$CV = \frac{\sigma}{\mu} \times 100 \quad (3-1)$$

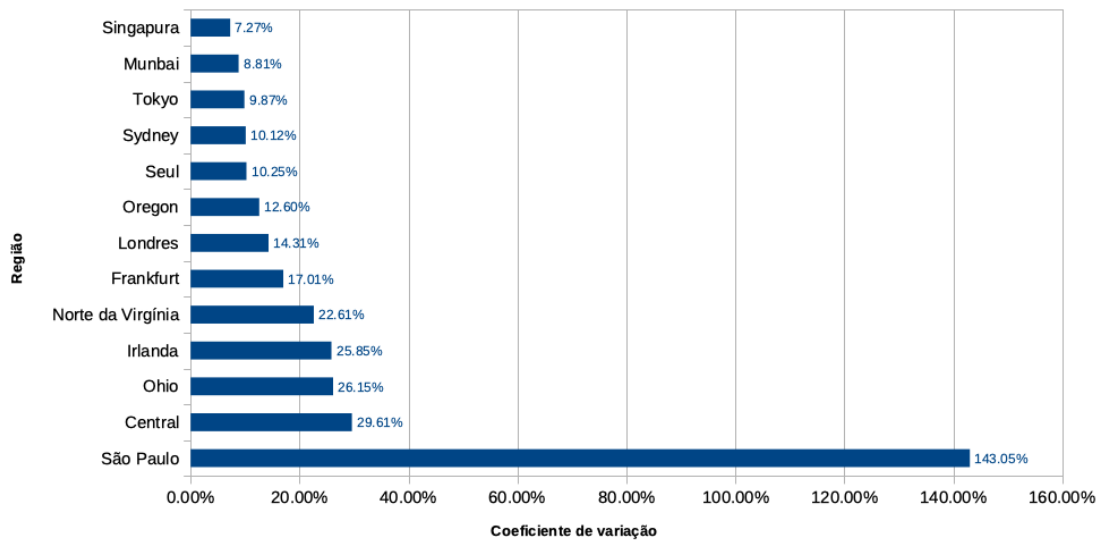
onde:

- $CV$ : coeficiente de variação, dado em porcentagem.
- $\sigma$ : desvio padrão.
- $\mu$ : média.

Inicialmente foi analisado o coeficiente de variação para cada um dos dias considerados. Os resultados para cada dia foram semelhantes, sendo que os valores para o dia 27/06/2021 são apresentados na Figura 3.6. Como pode ser visto na figura os valores foram altos, sendo que para a região de São Paulo chegou acima de 140%. Isso acontece devido ao fato de ser a região mais próxima, fazendo com que as oscilações típicas da rede interfiram muito no valor do coeficiente. O mesmo comportamento não é visualizado nas outras regiões devido ao fato das oscilações representarem pouca proporção em valores mais elevados.

A mesma análise foi realizada considerando os intervalos de seis horas, quatro horas e uma hora. Para seis e quatro horas o comportamento foi equivalente ao de 24 horas. Contudo, ao considerar intervalos de uma hora, com exceção da região de São Paulo, todos os coeficientes de variação neste caso foram abaixo de 20%, o que constitui um valor aceitável (ResearchGate, 2022). Diante disso, é possível afirmar que realizar coletas a cada hora é suficiente para se ter uma representatividade significativa dos dados. Contudo, para inserir o aspecto de aleatoriedade a coleta não é realizada em um momento fixo mas sim em um momento aleatório dentro do intervalo de cada hora. A Figura 3.7

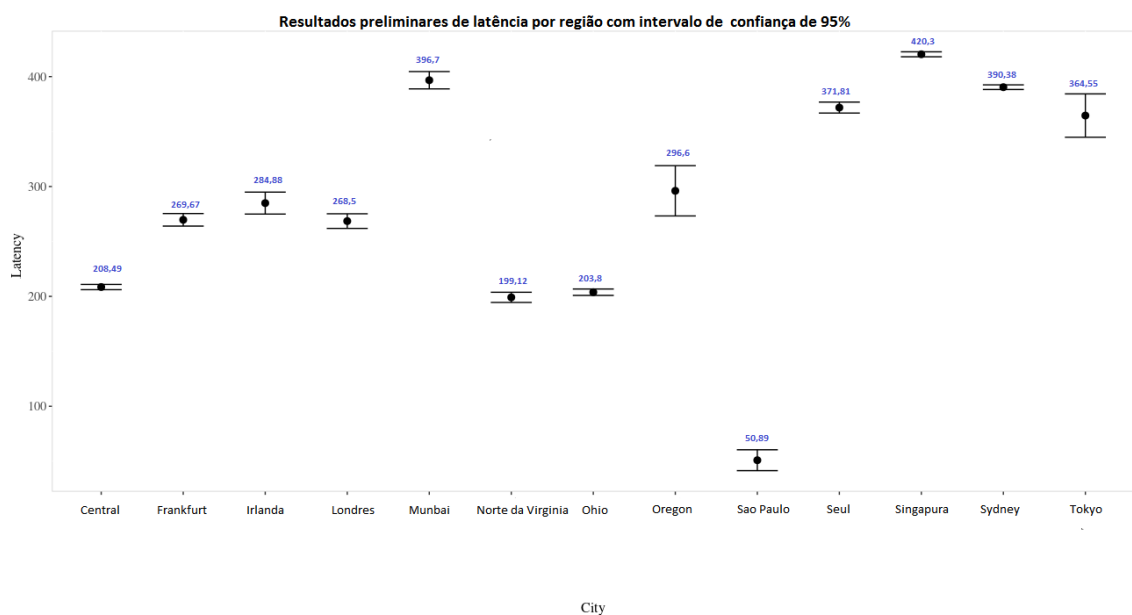




**Figura 3.6:** Coeficiente de variação para dados de 27/06/2021.

Fonte: Elaborada pelos autores

apresenta uma estimativa do valor da latência para as regiões, considerando um intervalo de confiança de 95%.



**Figura 3.7:** Média de latência por região ofertada pela AWS.

Fonte: Elaborada pelos autores

Para descrever melhor a implementação, a seguir são descritas funções importantes para tratamento estatístico no sistema. Inicialmente, no Código 3.1 é apresentada a função responsável por calcular o intervalo de confiança. Ela recebe como parâmetro uma lista de valores do tipo `double`, que representa as medições de latência, e um valor também do tipo `double`, que representa a significância desejada, valor definido na função.

---

**Código 3.1** funcao\_intervaloConfianca()

```
1 public double confidenceInterval(double[] values, double significance) {  
2     DescriptiveStatistics statistics = new DescriptiveStatistics(values);  
3     TDistribution tDist = new TDistribution(statistics.getN() - 1);  
4     double a = tDist.inverseCumulativeProbability(1.0 - significance / 2);  
5     return a * statistics.getStandardDeviation() / Math.sqrt(statistics.getN());  
6 }
```

---

O Código 3.2 apresenta a função responsável por calcular o coeficiente de variação. A função recebe como parâmetro a média de latência de uma região específica e uma lista de valores do tipo `double` que representa os dados de latência colhidos na região.

---

**Código 3.2** funcao\_coeficienteVariancia()

```
1 public Double coefficientVariance(Double media, double[] values) {  
2     double mean = media;  
3     double temp = 0;  
4     for (double a : values) {  
5         temp += (a - mean) * (a - mean);  
6     }  
7     return Math.sqrt(temp / (values.length)) / mean * 100;  
8 }
```

---

No Código 3.3 é apresentada a função responsável por calcular a média de cada região considerando todos os dados de latência disponíveis para cada região. Sendo assim, ela recebe uma lista de latência de cada região e retorna a média. Essa função é utilizada para montar o gráfico de barras na tela de *Dashboard*.

---

**Código 3.3** funcao\_mediaSimples()

```
1 public Double simpleMedia(List<Metric> latencys) {  
2     Double sumLatency = 0.0;  
3     for (int i = 0; i <= LIST_SIZE; i++) {  
4         sumLatency += Double.parseDouble(latencys.get(i).getLatency());  
5     }  
6     return sumLatency / LIST_SIZE;  
7 }
```

---

O Código 3.4, por fim, apresenta a função que calcula a média simples por hora. Esta função recebe uma lista de latência que foi coletada no intervalo de hora especificado e retorna a média nesse intervalo. Esta função é utilizada para gerar dados de latência na tabela da tela de *Dashboard*.

---

**Código 3.4** `funcao_mediaSimplesHora()`

---

```
1 public Double simpleMediaByHour(List<Metric> latencys) {  
2     Double sumLatency = 0.0;  
3     for (int i = 0; i <= latencys.size()-1; i++) {  
4         sumLatency += Double.parseDouble(latencys.get(i).getLatency());  
5     }  
6     return sumLatency / latencys.size();  
7 }
```

---

### 3.2.1 Persistência de dados

Para armazenar os dados que são colhidos pela aplicação foi utilizado o Sistema Gerenciador de Bancos de Dados (SGBD) MySQL por permitir criar e gerenciar um banco de dados relacional que atende às necessidades da aplicação. A base de dados é composta por duas tabelas:

- TB\_REGIAO: armazena os dados que descrevem as regiões da AWS.
- TB\_METRICA: armazena os dados coletados através do comando ping.

A tabela TB\_REGIAO é composta por seis atributos, sendo eles:

- ID\_REGIAO: serve como chave primária da tabela para identificação da região;
- CIDADE: nome da cidade onde está localizada a região;
- ENDPOINT: URL da região, usada no comando ping;
- INTERVALO\_IP: identifica o intervalo de endereços *Internet Protocol* (IP) do *end-point*;
- LOCALIZACAO: identifica em qual continente está localizada a região;
- REGIAO: nome da região.

A tabela TB\_METRICA é composta por nove atributos, sendo eles:

- ID\_METRICA: serve como chave primária da tabela para identificação da métrica;
- ICMP\_SEQUENCE: sequência ICMP do comando ping;
- IP\_MAQUINA: endereço IP da máquina da qual o comando ping obteve resposta;

- LATENCIA: latência medida da região específica;
- NOME\_MAQUINA: nome da máquina com a qual foi obtida a medição;
- TTL: número máximo de pacotes que a medição considera;
- DATA: data em que foi realizada a medição;
- HORA: hora em que foi realizada a medição;
- ID\_CIDADE: identifica a cidade a que se relaciona a medição.

A Figura 3.8 apresenta o diagrama entidade-relacionamento da base de dados utilizada.



**Figura 3.8:** Modelagem da base de dados.

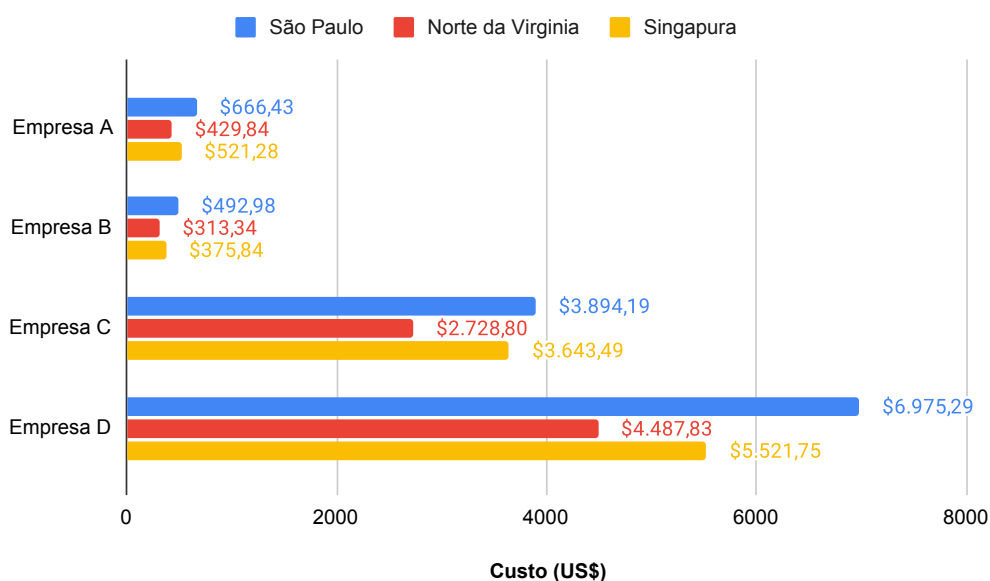
Fonte: Elaborada pelos autores

## Avaliação

Para demonstrar a aplicabilidade da ferramenta, neste capítulo é apresentado um estudo de caso usando dados de uso da plataforma AWS por empresas reais.

A AWS fornece uma ferramenta chamada *Pricing Calculator* (2022) que permite calcular uma estimativa do custo que terá um conjunto de recursos selecionados. A ferramenta permite auxiliar tomadas de decisões relacionadas ao custo mas não oferece qualquer análise sobre a latência.

Considerando o histórico de utilização de máquinas virtuais de cinco empresas reais, aqui nomeadas como “A”, “B”, “C” e “D” por questão de confidencialidade, a ferramenta *Pricing Calculator* (2022) foi utilizada para estimar o gasto mensal de cada empresa, considerando que todas as instâncias foram alocadas no modelo de cobrança sob demanda e ficaram executando no esquema 24x7. O Apêndice A apresenta os tipos e quantidades de instâncias para cada empresa. Os resultados são apresentados na Figura 4.1.



**Figura 4.1:** Custo mensal para cada uma das empresas.

Fonte: Elaborada pelos autores

Para a análise da latência, a ferramenta desenvolvida foi utilizada para identificar a região com a menor e a maior latência, as quais foram São Paulo e Singapura, respectivamente. De maneira adicional, também foi considerada a região Norte da Virgínia por ser a região com os menores custos. Foi avaliado o impacto da latência nas atividades das empresas, usando como métrica o volume de dados que é possível transmitir considerando todas as máquinas virtuais alocadas.

Na avaliação foi considerada uma análise puramente teórica baseada somente no atraso de propagação em rede. Para tal, foi adotado como padrão o tamanho da janela TCP como sendo 64 KB. Além disso, a latência dada pela ferramenta foi usada como sendo o *Round Trip Time* (RTT), consequentemente, a quantidade de pacotes transmitidas em cada segundo é dada pela Equação 4-1.

$$q = \frac{1}{\Delta} \quad (4-1)$$

onde:

- $q$ : quantidade de pacotes transmitidos por segundo;
- $\Delta$ : RTT (medida de latência dada pela ferramenta).

Dadas as suposições apresentadas, o volume de dados transmitidos a cada segundo é dado pela Equação 4-2.

$$v = \frac{q \times J}{10^9} \quad (4-2)$$

onde:

- $v$ : volume de dados transmitido por segundo (em Terabytes);
- $q$ : quantidade de pacotes transmitidos por segundo;
- $J$ : tamanho da janela TCP (em kilobytes).

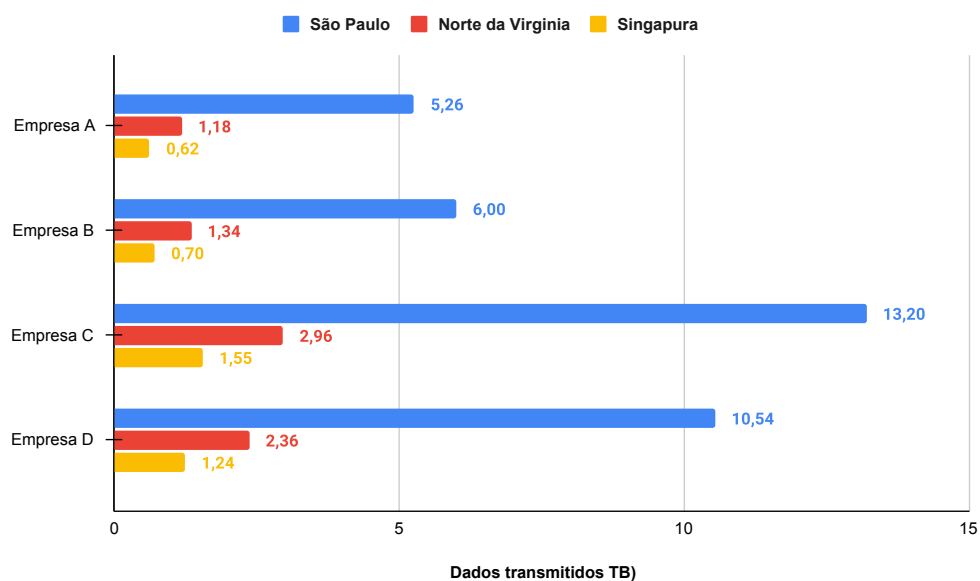
Por fim, o volume de dados transmitidos durante o período de um mês por uma empresa  $i$  é dado pela Equação 4-3.

$$\Theta_i = v_i \times 60 \times 60 \times 24 \times 30 \quad (4-3)$$

onde:

- $\Theta_i$ : volume de dados transmitidos durante o período de um mês pela empresa  $i$  (em Terabytes);
- $v_i$ : volume de dados transmitidos por segundo pela empresa  $i$  (em Terabytes).

Os resultados sobre a estimativa do volume de dados transmitidos é apresentado na Figura 4.2.



**Figura 4.2:** *Volume de dados transmitidos para cada uma das empresas.*

Fonte: Elaborada pelos autores

Com base nestes resultados é possível tomar a decisão sobre qual região é mais adequada para que a empresa faça uma contraposição entre custo e latência. Como exemplo, se a empresa tiver como alvo o Brasil e baixa latência for um requisito essencial a região ideal seria a de São Paulo, mesmo sendo a com maior custo. Por outro lado, se for tolerado certo nível de latência, a região de Norte da Virgínia passa a ser mais interessante por possuir um custo inferior.

Este tipo de análise deve ser considerado na tomada de decisão. Considerando as regiões com maior e menor custo (São Paulo e Norte da Virgínia, respectivamente) os resultados da avaliação permitem afirmar que embora a região do Norte da Virgínia faça com que o volume de dados transmitidos seja em média 77,48% menor, o custo também tem uma redução média de 34,38%, o que representa no estudo de caso uma economia mensal total de US\$ 4.069,08. Cabe então às empresas decidir se o volume inferior de dados trafegados não justificariam a economia.



## Considerações Finais

---

O mercado de nuvem está cada vez mais promissor, ganhando espaço no mundo inteiro com novas tecnologias e praticidades, facilitando o desenvolvimento de quem trabalha com tecnologia e, conseqüentemente, a experiência do usuário final.

Hoje em dia, grande parte das empresas utiliza algum provedor de serviço em nuvem para hospedar suas aplicações, pela facilidade, infraestrutura de qualidade, segurança, praticidade e gerenciamento. Contudo, devido à multiplicidade de regiões disponíveis com custo variado entre elas surge a necessidade de uma solução que permita a análise da latência de comunicação, para que este parâmetro seja usado na tomada de decisões. Diante disso, este trabalho apresentou a proposta e desenvolvimento de uma ferramenta que permite explorar medidas relacionadas à latência. O protótipo inicial considerou as regiões da AWS, apresentando métricas estatísticas, que permitem analisar a confiabilidade dos dados.

Além da ferramenta desenvolvida, também foi apresentada uma avaliação sobre como a ferramenta pode ser utilizada para produzir insumos à tomada de decisão.

### 5.1 Trabalhos futuros

Diante dos resultados obtidos e limitações observadas é proposto os seguintes trabalhos futuros:

- **Inclusão de outros provedores:** a proposta inicial era considerar para a ferramenta um conjunto maior de provedores, além da AWS. Contudo, o tempo disponível não foi suficiente para a implementação. Sugere-se então a inclusão de outras opções como Google Cloud Platform, Oracle Cloud, Alibaba Cloud, Huawei Cloud, Microsoft Azure, dentre outros.
- **Comparação com custo:** a avaliação apresentada no Capítulo 4 é bastante útil mas não é parte integrante da ferramenta. Sugere-se a integração com APIs e serviços de precificação para que seja oferecida como parte dos resultados.

# Referências Bibliográficas

Amazon Web Services. **AWS**. 2022. Disponível em: <https://aws.amazon.com/pt/>. Acesso em: 26 abr 2022. 15, 19, 22

\_\_\_\_\_. \_\_\_\_\_. 2022. Disponível em: <https://aws.amazon.com/pt/what-is-cloud-computing/>. Acesso em: 26 abr 2022. 18

\_\_\_\_\_. **AWS Pricing**. 2022. Disponível em: <https://aws.amazon.com/pt/ec2/pricing/>. Acesso em: 15 jun 2022. 20

AWS PING Github - ekalinin. **AWS PING**. 2022. Disponível em: <https://github.com/ekalinin/awsping>. Acesso em: 18 abr 2022. 16

AZURE. **O que é nuvem?** 2019. Disponível em: <https://azure.microsoft.com/pt-br/overview/what-is-the-cloud/>. Acesso em: 15 ago 2021. 19

Azure Cloud. **Microsoft Cloud**. 2022. Disponível em: <https://azure.microsoft.com/pt-br/>. Acesso em: 26 abr 2022. 18

BALA, R; GILL, B; SMITH, D; WRIGHT, D. **Magic quadrant for cloud Infrastructure as a Service**. 2020. Disponível em: <https://www.gartner.com/doc/reprints?id=1-1CMAPXNO&ct=190709&st=sb>. Acesso em: 13 Jun 2022. 19, 24

Cloud Ping. **Cloud Ping**. 2022. Disponível em: <https://www.cloudping.co/grid>. Acesso em: 10 ago 2021. 16, 22

Convergência Digital. **Mercado global de nuvem vai crescer 55% até 2022**. 2019. Disponível em: <https://www.convergenciadigital.com.br/Cloud-Computing/Mercado-global-de-nuvem-vai-crescer-55%25-ate-2022-52237.html>. Acesso em: 10 ago 2021. 15

DIGITAL, ACM. **Adequação do Ping**. 2013. Disponível em: [https://dl.acm.org/doi/pdf/10.1145/2504730.2504765?casa\\_token=UgwpPHZYt0UAAAAA:waZMI0Knwyooht4Zy9YDiIyKVrv1Z5SszOXxeF\\_yiRulhBkV116fQdaVC2Wb81mC5dXK038Sz4I](https://dl.acm.org/doi/pdf/10.1145/2504730.2504765?casa_token=UgwpPHZYt0UAAAAA:waZMI0Knwyooht4Zy9YDiIyKVrv1Z5SszOXxeF_yiRulhBkV116fQdaVC2Wb81mC5dXK038Sz4I). Acesso em: 23 Abr 2022. 21

Eco IT Segurança Digital. **Eco IT**. 2022. Disponível em: <https://blog.ecoit.com.br/computacao-em-nuvem/>. Acesso em: 26 abr 2022. 18

gcping. 2022. Disponível em: <https://www.gcping.com/>. Acesso em: 12 abr 2022. 16

Google Cloud Platform. **GCP**. 2022. Disponível em: <https://cloud.google.com/>. Acesso em: 12 abr 2022. 16

\_\_\_\_. \_\_\_\_\_. 2022. Disponível em: <https://cloud.google.com/>. Acesso em: 26 abr 2022. 18, 19

Google Sites. **Conceito de Redes**. 2020. Disponível em: <https://sites.google.com/site/conceitosderedes/perda-de-pacotes>. Acesso em: 12 Fev 2020. 21

Huawei Cloud. **Huawei**. 2022. Disponível em: <https://www.huaweicloud.com/intl/pt-br/>. Acesso em: 26 abr 2022. 18

ISLAM, Noman; ISLAM, Zeeshan. An economic perspective on major cloud computing providers. **arXiv preprint arXiv:1810.05088**, 2018. 16

LAPPONI, Juan Carlos. **Estatística usando excel**. [S.l.]: Elsevier Brasil, 2004. 28

Measurement Lab. **Overview and an Invitation to the Research Community. SIG-COMM CCR 2010, 2010**. 2022. Acesso em: 03 mai 2022. 21

MELL, Peter; GRANCE, Tim *et al.* The NIST definition of cloud computing. Computer Security Division, Information Technology Laboratory, National ..., 2011. 15

MINHACONEXÃO. **Entenda o que é a Latência e sua influência na conexão com a internet**. 2020. Disponível em: <https://www.minhaconexao.com.br/blog/entenda-o-que-e-a-latencia-e-sua-influencia-na-conexao-com-a-internet/>. Acesso em: 12 ago 2021. 15

MySQL. **MySQL- Oracle**. 2022. Disponível em: <https://www.mysql.com/>. Acesso em: 26 jun 2022. 24

Oracle Cloud. **Oracle**. 2022. Disponível em: <https://www.oracle.com/br/cloud/>. Acesso em: 26 abr 2022. 18

Pricing Calculator. **AWS**. 2022. Disponível em: <https://calculator.aws/#/addService>. Acesso em: 01 jul 2022. 33

Project BISmark. **Georgia Tech et al**. 2022. Disponível em: <https://atlas.ripe.net>. Acesso em: 03 mai 2022. 21

---

React. **Meta Platforms, Inc.** 2022. Disponível em: <https://pt-br.reactjs.org/>. Acesso em: 26 jun 2022. 25

ResearchGate. **What are the acceptable values for the percentage deviation (DEV%) and the coefficient of variance (CV)?** 2022. Disponível em: [https://www.researchgate.net/post/What\\_are\\_the\\_acceptable\\_values\\_for\\_the\\_percentage\\_deviation\\_DEV\\_and\\_the\\_coefficient\\_of\\_variance\\_CV](https://www.researchgate.net/post/What_are_the_acceptable_values_for_the_percentage_deviation_DEV_and_the_coefficient_of_variance_CV). Acesso em: 26 abr 2022. 28

RIPE Atlas. **RIPE NCC.** 2022. Disponível em: <https://atlas.ripe.net>. Acesso em: 03 mai 2022. 21

RIPE Test. **RIPE NCC.** 2022. Disponível em: <https://atlas.ripe.net>. Acesso em: 03 mai 2022. 21

SILVA, Nilton Neves. **O IMPACTO DA DIMINUIÇÃO DO TEMPO DE LATÊNCIA NA PLATAFORMA DE NEGOCIAÇÃO DA BM&FBOVESPA NA LIQUIDEZ DO MERCADO DE AÇÕES NO BRASIL.** 2018. Disponível em: <http://dspace.insper.edu.br/xmlui/bitstream/handle/11224/2213/Nilton%20Neves%20Silva.pdf?sequence=1>. Acesso em: 08 ago 2021. 15

Spring. **VMware, Inc. or its affiliates.** 2022. Disponível em: <https://spring.io/>. Acesso em: 26 jun 2022. 24

Universidade Estadual do Oeste do Parana (UNIOESTE). **Avaliação de Soluções para Alocação de Aplicações Distribuídas em Ambientes de Nuvem.** 2022. Disponível em: <https://www.seer.ufrgs.br/reic/article/download/87475/50664>. Acesso em: 17 abr 2022. 15

# Apêndice A – Máquinas virtuais consideradas na avaliação

**Tabela A.1:** *Instâncias utilizadas pela empresa A.*

| Empresa | Instância  | Quantidade |
|---------|------------|------------|
| A       | t3a.small  | 1          |
|         | t3.small   | 1          |
|         | c5a.xlarge | 1          |
|         | t3.small   | 1          |
|         | t2.large   | 1          |
|         | t2.micro   | 1          |
|         | t2.micro   | 1          |
|         | t3a.small  | 4          |
|         | t3a.small  | 4          |
|         | t2.small   | 1          |
|         | t3a.small  | 1          |
|         | t3a.small  | 1          |
|         | t3a.small  | 1          |

Fonte: Elaborada pelos autores

**Tabela A.2:** *Instâncias utilizadas pela empresa B.*

| Empresa | Instância  | Quantidade |
|---------|------------|------------|
| B       | c5.xlarge  | 1          |
|         | t2.micro   | 1          |
|         | t3a.small  | 4          |
|         | r5a.large  | 1          |
|         | t2.micro   | 2          |
|         | t3a.medium | 1          |

Fonte: Elaborada pelos autores

**Tabela A.3:** *Instâncias utilizadas pela empresa C.*

| Empresa | Instância  | Quantidade |
|---------|------------|------------|
| C       | t3a.medium | 1          |
|         | t3a.medium | 1          |
|         | t3a.medium | 1          |
|         | t3a.medium | 1          |
|         | t3a.small  | 11         |
|         | t3a.small  | 2          |
|         | t3a.small  | 2          |
|         | t3a.xlarge | 8          |
|         | m5a.xlarge | 6          |

Fonte: Elaborada pelos autores

**Tabela A.4:** *Instâncias utilizadas pela empresa D.*

| Empresa | Instância  | Quantidade |
|---------|------------|------------|
| D       | t3a.xlarge | 4          |
|         | t3.medium  | 3          |
|         | t3.medium  | 3          |
|         | m5a.xlarge | 6          |
|         | t3.medium  | 2          |
|         | t3.medium  | 1          |
|         | t3a.medium | 1          |
|         | t3a.medium | 1          |
|         | m5a.large  | 8          |
|         | r5.large   | 5          |
|         | r5.large   | 2          |
|         | t3a.nano   | 1          |
|         | t3a.medium | 3          |
|         | r4.4xlarge | 1          |

Fonte: Elaborada pelos autores