



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise de Serviços de Armazenamento por Objetos em Provedores de Nuvens Públicas

Herbert Vargas de Araujo
Paulo Henrique A. Florentino

Monografia apresentada como requisito parcial
para conclusão do Curso de Computação — Licenciatura

Orientadora
Prof. ^a Dr. ^a Aletéia Patrícia Favacho de Araújo

Brasília
2021

Dedicatória

Eu, Herbert Vargas de Araujo, dedico este trabalho a Deus, a minha filha Manuela, a minha namorada Fernanda e a todos meus amigos e familiares que me apoiaram em todo o processo de graduação, em especial aos meus pais Valter e Vera. Eu, Paulo Henrique A. Florentino, dedico este trabalho principalmente a Deus, a minhas filhas Isabela e Manuela, ao meu pai Carlos Florentino, e a todos os amigos que me acompanharam nesta trajetória.

Agradecimentos

Agradecemos aos nossos pais, que sempre nos apoiaram em toda essa jornada. Somos gratos a todos os colaboradores da Universidade de Brasília por nos proporcionar uma exímia formação acadêmica, especialmente à Carolina Alves Okimoto, por toda ajuda e atenção, e à Prof. ^a Dr. ^a Aletéia Patrícia Favacho de Araújo, por ter nos orientado em todo o processo de execução deste trabalho.

Resumo

À medida que a computação em nuvem se torna cada vez mais popular entre as empresas, os desenvolvedores e as organizações, deve-se considerar o problema da dificuldade em escolher entre o número crescente de provedores e serviços. Sob essa perspectiva, esta pesquisa objetiva comparar os serviços de armazenamento por objeto ofertados pelos principais provedores de nuvem pública a fim de constatar se há uma disparidade relevante entre eles. Para isso, utilizou-se uma ferramenta de *benchmarking* capaz de analisar o desempenho dos serviços de armazenamento por objetos, abrangendo as principais métricas necessárias. Assim, realizou-se testes com diversos parâmetros e analisou-se o comportamento de cada métrica, com variação mínima entre os provedores, entregando um desempenho similar para ambientes com poucos recursos computacionais. Contudo, os resultados mostraram que os grandes provedores apresentam um desempenho similar na maioria dos casos.

Palavras-chave: Computação em Nuvem, Armazenamento por Objeto, Google, AWS, Azure.

Abstract

As cloud computing becomes more and more popular among businesses, developers and organizations, the problem of choosing between the growing number of providers and services must be considered. From this perspective, this research aims to compare the storage services by object offered by the main public cloud providers in order to see if there is a relevant disparity between them. For this, we used a benchmarking tool capable of analyzing the performance of storage services by objects, covering the main metrics needed. Thus, tests were carried out with several parameters and the behavior of each metric was analyzed, with minimal variation between providers, delivering a similar performance for environments with few computational resources. However, the results showed that large providers have a similar performance in most cases.

Keywords: Cloud Computing, Object Storage, Google, AWS, Azure.

Sumário

1	Introdução	1
1.1	Objetivos	2
1.2	Método	2
1.3	Estrutura do Trabalho	3
2	Nuvem Computacional	4
2.1	Definição	4
2.2	Características Essenciais	5
2.3	Modelos de Serviço	5
2.3.1	Software como Serviço (SaaS)	6
2.3.2	Plataforma como Serviço (PaaS)	6
2.3.3	Infraestrutura como Serviço (IaaS)	7
2.3.4	Tudo como um Serviço (XaaS)	7
2.4	Modelos de Implantação	8
2.4.1	Nuvem Pública	8
2.4.2	Nuvem Privada	8
2.4.3	Nuvem Comunitária	9
2.4.4	Nuvem Híbrida	9
2.5	Principais Provedores de Nuvem Pública	10
2.5.1	Amazon Web Services	11
2.5.2	Google Cloud Platform	13
2.5.3	Microsoft Azure	14
2.6	Considerações Finais	16
3	Armazenamento em Nuvem	17
3.1	Visão Geral do Armazenamento em Nuvem	17
3.2	Tipos de Armazenamento	18
3.2.1	Armazenamento por Objetos	18
3.2.2	Armazenamento por Bloco	18

3.2.3 Armazenamento por Arquivos	19
3.3 Serviços de Armazenamento das Principais Nuvens Públicas	19
3.3.1 Serviços de Armazenamento por Objetos	20
3.3.2 Serviços de Armazenamento por Blocos	23
3.3.3 Serviços de Armazenamento por Arquivos	25
3.4 Considerações Finais	28
4 Análise dos Resultados	29
4.1 Benchmark	29
4.1.1 Métricas	30
4.2 Testes	31
4.3 Resultados	32
4.3.1 Taxa de Transferência	32
4.3.2 Tempo de Resposta	34
4.3.3 Largura de Banda	36
4.3.4 Taxa de Sucesso	37
4.4 Considerações Finais	39
5 Conclusão	40
Referências	42

Lista de Figuras

2.1	Quadrante Mágico do Gartner para Serviços de Infraestrutura e Plataforma em Nuvem.	10
2.2	Dados do Relatório da Synergy Research Group.	11
2.3	Infraestrutura Global do AWS.	12
2.4	Infraestrutura Global do GCP.	14
2.5	Infraestrutura Global do Microsoft Azure.	15
4.1	Arquitetura do CosBench.	30
4.2	Principais métricas.	31
4.3	Taxa de transferência para objetos de 64KB.	33
4.4	Taxa de transferência para objetos de 1MB.	33
4.5	Taxa de transferência para objetos de 16MB.	34
4.6	Tempo de resposta para objetos de 64KB.	35
4.7	Tempo de resposta para objetos de 1MB.	35
4.8	Tempo de resposta para objetos de 16MB.	35
4.9	Largura de banda para objetos de 64KB.	36
4.10	Largura de banda para objetos de 1MB.	37
4.11	Largura de banda para objetos de 16MB.	37
4.12	Taxa de sucesso para objetos de 64KB.	38
4.13	Taxa de sucesso para objetos de 1MB.	38
4.14	Taxa de sucesso para objetos de 16MB.	39

Lista de Tabelas

3.1	Comparativo de Armazenamento S3.	21
3.2	Comparativo de Armazenamento Cloud Storage.	22
3.3	Comparativo de Armazenamento Blob Storage.	23
3.4	Comparativo de Armazenamento EBS.	24
3.5	Comparativo de Armazenamento Persistent Disks.	25
3.6	Comparativo de Armazenamento Managed Disk.	25
4.1	Escopo de testes.	32

Capítulo 1

Introdução

Apesar do nome sugerir algo abstrato ou imaterial, o termo nuvem é uma metáfora para serviços e/ou arquiteturas computacionais que são processados em algum lugar remoto, mas não necessariamente distantes e acessíveis por meio de conexão com a internet.

A computação em nuvem é uma evolução dos serviços e produtos da tecnologia da informação. Ela tem como maior foco prover serviços de maneira centralizada para todos, desde usuários finais, que hospedam somente documentos e arquivos pessoais, a até mesmo empresas de grande porte, que podem migrar suas aplicações e dados críticos para a nuvem e de qualquer local de uma forma simples, utilizar recursos de software e hardware de alto nível e com alto poder de processamento de maneira transparente, sob demanda, altamente flexível e escalável, e cobrados por utilização.

Esse conceito de pagar apenas pelo o que usa é antigo, ele nasceu e tem se aprimorado cada vez mais, e pode ser visto em outros serviços básicos e necessários para a vida humana, como por exemplo energia elétrica, água, gás encanado e etc.

Essa flexibilidade e simplicidade que é entregue pelas nuvens computacionais vem motivando cada vez mais clientes a migrar seus recursos e aplicações para plataformas de nuvens públicas, isso inclui organizações, governos, escolas, empresas e etc.

Seguindo a tendência, cada vez mais empresas estão se tornando provedores e lançando suas próprias plataformas de computação em nuvem. Embora a variedade de provedores estimule a competição saudável, isso torna a escolha da plataforma um desafio para gestores de TI.

Identificar os provedores que oferecem os recursos necessários para suportar ambientes é fácil porque a maioria deles possuem um similar conjunto de oferta de serviços. No entanto, dizer qual nuvem oferece o melhor desempenho de processamento é difícil, principalmente porque existem poucos comparativos abrangentes de desempenho entre provedores de nuvem.

Com os serviços de armazenamento por objetos cada vez mais aceitos como substitutos dos sistemas tradicionais de arquivos ou blocos, é importante medir com eficácia o desempenho desses serviços. Assim, as pessoas podem comparar diferentes soluções ou ajustar seus sistemas para um melhor desempenho.

Diante do exposto, este trabalho propõe um modelo sistemático para determinar quais métricas melhor caracterizam a atuação da nuvem em serviços de armazenamento por objetos. A aplicação adequada dessas métricas podem ajudar os usuários a escolher de maneira mais acertiva qual provedor é o ideal para atender suas necessidades.

Assim, considerando a complexidade do assunto e o objetivo de orientar a implementação, é proposto uma ferramenta comparativa para classificar diferentes cargas de trabalho.

1.1 Objetivos

O objetivo geral deste trabalho é comparar os serviços de armazenamento por objeto no ambiente de computação em nuvem. Para atingir o objetivo geral, faz-se necessário atingir os seguintes objetivos específicos:

- Descrever as funcionalidades existentes de diferentes ferramentas de armazenamento;
- Apresentar as características exclusivas do armazenamento por objetos;
- Apresentar e implementar uma ferramenta para realizar tarefas de *benchmark*;
- Analisar e comparar a performance na infraestrutura proposta.

1.2 Método

Para o desenvolvimento deste trabalho foi realizada uma pesquisa bibliográfica sobre computação em nuvem e armazenamento em nuvem, para a fundamentação teórica e contextualização do leitor.

Para a análise dos serviços de armazenamento, foram escolhidos provedores de nuvem pública que representassem as diferentes características gerais que lidam com o armazenamento por objeto.

A análise dos serviços foi feita a partir utilização prática dos mesmos. Os aspectos considerados foram a taxa de transferência, o tempo de resposta, a largura de banda e a taxa de sucesso, e levaram em consideração a geração de diversas cargas de trabalho.

1.3 Estrutura do Trabalho

Este trabalho está dividido em cinco capítulos. O Capítulo 1 é responsável pela introdução do trabalho, descrevendo quais foram os métodos utilizados e também sua estrutura.

No Capítulo 2 serão apresentados conceitos e características de computação em nuvem, abordando os principais provedores de nuvem pública.

O Capítulo 3 proporciona ao leitor uma base teórica sobre armazenamento em nuvem, com estudos de grandes pesquisadores da área, abordando os tipos de armazenamento e principais serviços disponíveis no mercado.

No Capítulo 4 serão apresentadas as análises dos resultados obtidos neste trabalho. Para isso, será apresentada a visão geral da ferramenta de *benchmark*, a arquitetura e metodologia utilizada nos testes propostos.

O capítulo final será a conclusão sobre as reflexões desenvolvidas nos capítulos anteriores e sugestões de trabalhos futuros.

Capítulo 2

Nuvem Computacional

Este capítulo é composto por cinco seções, nas quais serão abordadas as definições, as arquiteturas e os principais conceitos da Computação em Nuvem. A Seção 2.1 é voltada para a definição de nuvem. A Seção 2.2 aborda as características essenciais do ambiente de nuvem. As Seções 2.3 e 2.4 descrevem os modelos de serviço e implantação, respectivamente. E por fim, a Seção 2.5 descreve os três principais provedores públicos e os seus principais serviços.

2.1 Definição

Há na literatura diferentes definições de computação em nuvem, e diversas citações das características que esta nova plataforma deve apresentar. Para Buyya et al.[1], nuvem computacional é um sistema computacional paralelo e distribuído que consiste na coleção de computadores virtualizados interconectados que são supridos dinamicamente e apresentados como um ou mais recursos computacionais unificados baseado em acordo de nível de serviço, estabelecido através de negociação entre provedor e os consumidores. Vaquero et al.[2] caracteriza nuvem como grandes esferas de recursos virtuais úteis, fáceis e acessíveis (tais quais hardware, plataformas e/ou serviços de desenvolvimento). Esses recursos podem ser dinamicamente reconfigurados para ajustar a demanda de carga variável, permitindo também uma utilização otimizada dos recursos. Esse campo de recursos é tipicamente explorado por um modelo de pagamento por uso, no qual garantias são oferecidas pelo provedor da infraestrutura por meio de Acordo de Nível de Serviço (SLA - Service Level Agreement) [3][4] customizável.

Um conceito de computação em nuvem bastante aceito pela comunidade é o do NIST (National Institute of Standards and Technology) [5] que define a computação em nuvem como um modelo para habilitar o acesso por rede ubíquo, conveniente e sob demanda a um conjunto compartilhado de recursos de computação (como redes, servidores, armaze-

namento, aplicações e serviços), que possam ser rapidamente provisionados e liberados com o mínimo de esforço de gerenciamento ou interação com o provedor de serviços [5]. Assim, diante dessas diferentes definições, é possível notar que todas elas apresentam em comum as características essenciais, as quais são as responsáveis por tornar esse modelo único, diferenciando de outras plataformas distribuídas, tais como *clusters* [6] e *grids* [7].

2.2 Características Essenciais

Com base em diversos estudos [8][9][10][11], e de acordo com a definição do NIST [5], as características essenciais do ambiente de nuvem computacional são:

1. Autoatendimento sob Demanda: provisionamento automático no qual o consumidor pode usar os serviços da nuvem sem precisar de interação humana com o provedor;
2. Acesso Amplo à Rede: os recursos fornecidos pelos provedores devem ser acessíveis a partir de qualquer dispositivo em rede;
3. Agrupamento de Recursos: os recursos computacionais são agrupados para atenderem a múltiplos consumidores em um modelo *multitenant*, com diferentes recursos físicos e virtuais, provisionados dinamicamente de acordo com a demanda do consumidor;
4. Elasticidade Rápida: capacidade de alocar mais ou menos recursos no momento em que for necessário, com agilidade. Estes provisionamentos são solicitados pelos consumidores, mas em alguns casos, podem ser executados de forma automática pelos provedores;
5. Serviços Mensuráveis: o provedor de nuvem mede ou monitora o fornecimento de serviços por vários motivos, incluindo faturamento, uso eficaz de recursos ou planejamento preditivo geral.

2.3 Modelos de Serviço

A computação em nuvem fornece recursos computacionais distribuídos para os seus clientes, e os modelos de serviço que definem como esses recursos serão providos. O modelo conceitual, que é encontrado com maior frequência na literatura é composto por três modelos principais de fornecimento de serviço em nuvem, conhecidos como [5]: SaaS (*Software as a Service*), PaaS (*Platform as a Service*) e IaaS (*Infrastructure as a Service*).

2.3.1 Software como Serviço (SaaS)

Software como serviço (do inglês, *Software as a Service* — *SaaS*) é a capacidade provida ao consumidor de possibilitar a execução de aplicações do provedor na infraestrutura da nuvem. As aplicações são acessíveis por vários dispositivos do cliente por meio de uma interface de *thin client*, tal como uma *web browser* ou por uma interface de programa. O usuário não controla a infraestrutura básica da nuvem, incluindo a rede, os servidores, os sistemas operacionais, o armazenamento [5].

O SaaS consiste no serviço prestado na camada de aplicação das nuvens. Assim, o SaaS fornece aplicações para os usuários finais, sob responsabilidade do fornecedor, que providencia toda a infraestrutura necessária para o funcionamento. Além disso, o SaaS tem a responsabilidade de oferecer serviços como monitoramento, atualizações e segurança, dispensando custos de aquisição e reduzindo espaço em *datacenters* locais.

Este modelo de serviço comumente é disponibilizado pelos prestadores em uma arquitetura *multitenant*, o que significa dizer que uma única versão da aplicação, com uma única configuração de recursos, é disponibilizada para que seja utilizada por múltiplos clientes diferentes, na forma de locação ou subscrição [12]. Alguns exemplos desse tipo de serviço são: Microsoft Office 365 e Gmail.

2.3.2 Plataforma como Serviço (PaaS)

Plataforma como serviço (do inglês, *Platform as a Service* — *PaaS*) é a capacidade fornecida ao consumidor de implementar, na infraestrutura de nuvem, aplicações, utilizando linguagem de programação, bibliotecas, serviços e ferramentas fornecidas pelo provedor. O usuário não gerencia ou controla a infraestrutura básica da nuvem, incluindo a rede, os servidores, os sistemas operacionais ou o armazenamento, mas exerce controle sobre as aplicações implementadas, e definições de configurações possíveis para o ambiente de hospedagem das aplicações [5].

Os fornecedores do modelo PaaS entregam uma plataforma que tem como principal objetivo agilizar e facilitar o desenvolvimento e a hospedagem de aplicações. Os desenvolvedores deste serviço podem obter todos os sistemas e ambientes necessários para o ciclo de vida do software, seja desenvolvimento, teste, implantação e hospedagem de aplicativos da web [13].

As plataformas ofertadas possuem sistema operacional, linguagens de programação, ferramentas de desenvolvimento e bibliotecas para auxiliar e acelerar o cliente em todas as etapas do desenvolvimento de aplicativos, e prover um ambiente totalmente flexível e escalável.

Neste modelo de serviço, assim como no SaaS, o provedor de serviço é responsável por toda a infraestrutura de servidores, armazenamento e conexões de rede, o que difere é que a responsabilidade de gerenciar e monitorar os aplicativos hospedados é de responsabilidade do contratante.

2.3.3 Infraestrutura como Serviço (IaaS)

Infraestrutura como serviço (do inglês, *Infrastructure as a Service* — *IaaS*) é a capacidade fornecida ao consumidor de prover o processamento, o armazenamento, a rede e os outros recursos fundamentais de computação, no qual este consumidor é capaz de implementar e executar programas arbitrários, que podem incluir sistemas operacionais e aplicativos. O consumidor não gerência ou controla a infraestrutura básica da nuvem, mas tem controle sobre os sistemas operacionais, o armazenamento e a implementação de aplicativos [5].

No modelo IaaS, o cliente tem a maior responsabilidade na manutenção da infraestrutura, pois, o provedor entrega recursos computacionais em nível de infraestrutura, mas todo o sistema operacional, os dados e as aplicações devem ser gerenciados e mantidos pelo contratante.

Um provedor de serviços IaaS investe em infraestrutura, implementa e a mantém para oferecer hardware virtual (por exemplo máquinas virtuais, infraestrutura de rede, etc.), para eliminar a necessidade de um investimento inicial feito por seus assinantes [14].

Com isso, quem contrata serviços no modelo IaaS, reduz gastos com garantias de equipamentos e até mesmo energia elétrica por não precisar se preocupar em abrigar e manter em funcionamento recursos computacionais. Assim sendo, o provedor é o responsável pela segurança e disponibilidade desses equipamentos que oferta.

Outro benefício da utilização do modelo IaaS é sua flexibilidade. Isso ocorre porque a maioria dos fornecedores utiliza a virtualização, assim os contratantes conseguem aumentar e diminuir essas infraestruturas de acordo com suas demandas, e serem cobrados apenas pelo que realmente utilizam, não precisando fazer aquisições super provisionadas.

2.3.4 Tudo como um Serviço (XaaS)

O termo XaaS significa tudo como serviço, ou, qualquer coisa como serviço. Este termo passou a ser utilizado devido a crescente tendência de entregar qualquer tipo de aplicação no modelo "as a Service (aaS)", como por exemplo *Database as a Service* (DaaS) [15] e *Containers as a Service* (CaaS) [15].

Contudo, embora muito utilizado, existem diversos estudos inconclusivos sobre a taxonomia dos modelos como serviço, nos quais identificaram a falta de uma visão unificada para apoiar um entendimento comum de "aaS" [16].

Simmon [17] considerou o frequente uso pelo *marketing* do sufixo informal "aaS", como confuso e ofuscante para a distinção arquitetonicamente bem fundamentada dos modelos IaaS, PaaS e SaaS.

2.4 Modelos de Implantação

Um modelo de implantação refere-se ao acesso e a disponibilidade de ambientes de computação em nuvem. É algo muito flexível devido a um grande leque de opções que se adequam a necessidade de cada usuário [11].

Os modelos de implantação da computação em nuvem podem ser divididos e diferenciados pela localização onde os recursos estão sendo executados, e pelo nível de acesso e controle à infraestrutura.

Atualmente, as plataformas de nuvem computacional são classificadas, segundo NIST [5], em quatro tipos, os quais são: Nuvem Pública, Nuvem Privada, Nuvem Comunitária e Nuvem Híbrida, que serão descritos nas próximas seções.

2.4.1 Nuvem Pública

Este modelo de implantação fornece infraestrutura de recursos e serviços para uso aberto pelo público em geral, por uma empresa terceira, também conhecida como provedora de nuvem [18].

Uma nuvem pública se refere a um modelo de entrega de serviço em nuvem no qual um provedor de serviços disponibiliza recursos de TI altamente escaláveis, como capacidade de CPU e armazenamento ou aplicativos de software, para o público em geral pela Internet.

Neste modelo, o ambiente computacional do provedor é compartilhado para hospedar serviços e aplicações de diversos clientes, que acessam de forma simultânea por meio da Internet. Entretanto, isolamentos por meio de softwares garantem que os clientes desses provedores não enxerguem e nem alterem dados de outros clientes, tendo visibilidade e acesso apenas aos dados e aplicações que os pertence [19].

2.4.2 Nuvem Privada

A chamada nuvem privada é caracterizada quando a infraestrutura é hospedada dentro da empresa, de propriedade da própria instituição ou alugada para este fim. Neste modelo de implantação, as empresas (geralmente, grandes corporações com presença em vários locais) oferecem serviços de nuvem sobre a rede corporativa (pode ser uma rede privada virtual) para seus próprios usuários internos, por um ambiente protegido por *firewall* [20].

Em comparação com a nuvem pública, no qual todos os recursos e aplicativos são gerenciados pelo provedor de serviços, em nuvem privada esses serviços são agrupados e disponibilizados para os usuários em nível organizacional [13]. Assim, os sistemas podem ser altamente modulares, a administração de TI cria opções para usar ferramentas e plataformas internas e de terceiros. Dessa maneira, o usuário final recebe a capacidade de recorrer a essas ferramentas e plataformas como módulos, conforme necessário.

Embora tenha maiores custos iniciais para aquisição e instalação, usualmente, é a preferida para armazenar dados estratégicos e que precisem de tempo de resposta rápido, pois a infraestrutura está dentro da empresa, garantindo uma baixa latência de rede e garantindo maior segurança, dado que ninguém fora do perímetro da empresa terá acesso sem passar pelo *firewall*.

2.4.3 Nuvem Comunitária

Quando organizações constroem e compartilham em conjunto uma infraestrutura de nuvem, isso é chamado de nuvem comunitária. Geralmente, essas organizações trabalham com o mesmo nicho de mercado e partilham interesses como missão, requisitos de segurança, políticas, entre outros [21].

Esse modelo de implementação é adequado quando uma infraestrutura de nuvem é necessária para algumas organizações colaborativas para armazenar e compartilhar seus dados e funcionalidade. As nuvens comunitárias podem ser hospedadas em uma parte isolada de uma nuvem pública ou em um *datacenter* de uma empresa terceirizada [22].

2.4.4 Nuvem Híbrida

A nuvem híbrida é usada para as implementações nas quais o cliente acessa uma nuvem privada e uma nuvem pública, ou qualquer outra combinação de diferentes modelos de implantação [23], vistos anteriormente nesta seção.

Nesse caso, tem-se uma composição de duas ou mais nuvens, sejam elas privadas, públicas ou comunitárias. Essas permanecem como entidades únicas, mas estão unidas pela tecnologia padronizada ou proprietária que permite a portabilidade de dados e aplicativos [5]. Isso permite que a empresa usufrua de quaisquer modelos existentes em um único ambiente.

As principais vantagens deste modelo estão no reaproveitamento de investimentos já feitos pelas empresas, que não possuem necessidade de migrar todo seu ambiente *on premise* para um provedor de nuvem, e também por empresas que prezam ou necessitam que dados críticos sejam mantidos localmente.

2.5 Principais Provedores de Nuvem Pública

Os provedores de nuvem pública virtualizam sua própria infraestrutura, plataforma ou aplicações a partir de hardware proprietário, e alugam seus serviços e recursos para vários clientes simultâneos por meio da Internet [24].

À medida que a computação em nuvem se torna cada vez mais popular entre as empresas, os desenvolvedores e as organizações, existe um número crescente de provedores de serviço surgindo. Contudo, existem três principais provedores que se destacam dentre os demais, os quais são: Google Cloud Platform (GCP) [25], Amazon Web Service (AWS) [26] e Microsoft Azure [27], como pode ser visto no relatório de pesquisa do Gartner, apresentada na Figura 2.1. Além disso, de acordo com o Synergy Research Group [28], que realizou uma pesquisa de mercado dentro de um período de quatro anos, a Amazon mantém a liderança quando se trata de *Market Share*, permanecendo na faixa de 32-34% da fatia do mercado mundial. Enquanto, Microsoft e Google também conseguem ganhar espaço, com respectivamente 20% e 9%, tornando-as então as três maiores empresas prestadoras de serviço em nuvem do mundo, como pode ser notado na Figura 2.2.

Figure 1. Magic Quadrant for Cloud Infrastructure and Platform Services



Figura 2.1: Quadrante Mágico do Gartner para Serviços de Infraestrutura e Plataforma em Nuvem (Fonte [29]).

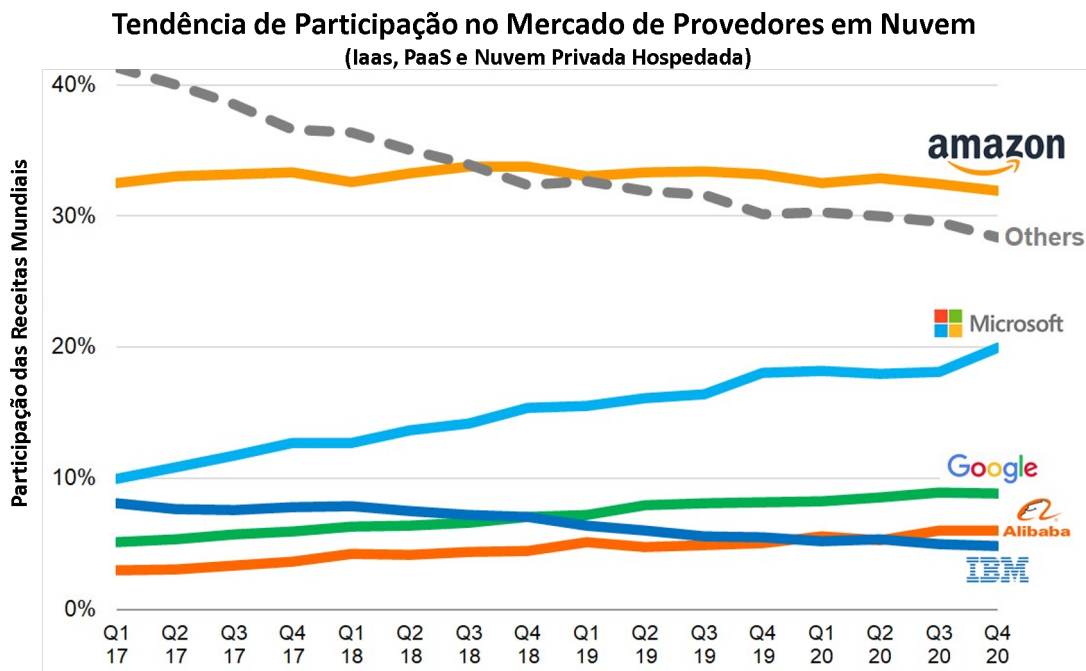


Figura 2.2: Dados do Relatório da Synergy Research Group (Fonte Adaptada [28]).

2.5.1 Amazon Web Services

Amazon Web Services (AWS) é um dos mais antigos e experientes provedores do mercado de nuvem. Ela foi lançada publicamente em 2006, e com isso estabeleceu uma base maior de usuários, bem como maiores fatores de confiabilidade em sua plataforma de computação em nuvem.

Atualmente, este provedor mantém várias regiões geográficas, incluindo regiões na América do Norte, África do Sul, América do Sul, Europa, China, Ásia-Pacífico e Oriente Médio, fornecendo mais de 200 tipos de serviços em nuvem para 245 países e territórios [30].

A infraestrutura da AWS é composta por regiões e zonas de disponibilidades. O conceito de zona de disponibilidade (do inglês, Availability Zone — AZ) é a composição de um ou mais *datacenters* distintos com energia, rede e conectividade redundantes em uma região da AWS. As AZs proporcionam aos clientes a capacidade de operar plataformas distribuídas com alta disponibilidade, tolerância a falhas e escalabilidade [31].

O provedor AWS opera em 80 zonas de disponibilidade, distribuídas em 25 regiões geográficas espalhadas pelo Mundo, e anunciou planos de lançamento de mais 5 regiões, conforme apresentado na Figura 2.3. Como citado anteriormente, existe uma quantidade considerável de serviços ofertados pelo provedor AWS. Os principais são [32]:



Figura 2.3: Infraestrutura Global do AWS (Fonte [30]).

1. **Elastic Compute Cloud (EC2):** serviço *web* que fornece capacidade de computação redimensionável, basicamente prove máquinas virtuais escaláveis sob demanda;
2. **Serviços de Armazenamento:** os serviços de armazenamento disponíveis pelo provedor serão detalhados no Capítulo 3, pois é o foco deste trabalho;
3. **Elastic Beanstalk:** plataforma destinada para *upload* de códigos desenvolvidos em Java, .NET, PHP, Node.js, Python, Ruby, Go e Docker. Este serviço automaticamente se encarrega da implementação, desde o provisionamento de capacidade, o balanceamento de carga e a escalabilidade automática, até o monitoramento da saúde do aplicativo;
4. **AWS Lambda:** proporciona a possibilidade de executar códigos sem que seja necessário gerenciar ou provisionar servidores. Dessa forma, o usuário só paga pelo tempo de computação que for utilizado, é a chamada computação “*serverless*” [33];
5. **Amazon RDS:** permite a fácil configuração, operação e escalabilidade de bancos de dados relacionais em nuvem, tais como MySQL, PostgreSQL, Oracle, SQL Server e MariaDB;
6. **Amazon Aurora:** banco de dados relacional de alta performance, proprietário da Amazon;
7. **Amazon DynamoDB:** banco de dados não relacional (NoSQL), proprietário da Amazon;

8. **Amazon Redshift:** *data warehouse* ágil para análise de dados. Esse serviço realiza suas funções por meio de ferramentas como SQL padrão, além de recursos de BI (*Business Intelligence*);
9. **Virtual Private Cloud (VPC):** serviço que permite o provisionamento de uma seção da nuvem AWS isolada, de modo lógico. Com esse serviço, os recursos da AWS só podem ser executados em uma rede virtual que o próprio usuário define;
10. **Amazon CloudFront:** rede de entrega de conteúdo global, que oferece dados de conteúdo aos usuários de maneira segura, em alta velocidade de transferência e com baixa latência.

2.5.2 Google Cloud Platform

O Google Cloud Platform (GCP) começou sua jornada em 2008 e, conseguiu criar uma boa presença no setor de nuvem. A intenção inicial do Google Cloud era fortalecer os próprios produtos do Google, como o mecanismo de pesquisa do Google e o YouTube. Atualmente, eles também introduziram seus serviços empresariais para que qualquer pessoa possa usar o GCP, que compartilha a mesma infraestrutura de Pesquisa Google ou do YouTube.

Os serviços do GCP estão disponíveis na América do Norte, América do Sul, Europa, Ásia e Austrália. Esses locais estão divididos em regiões e zonas. É possível escolher onde localizar os aplicativos para atender aos requisitos de latência, disponibilidade e durabilidade [34]. Atualmente, o Google Cloud Platform está disponível em 24 regiões e 73 zonas, fornecendo serviços em nuvem para um pouco mais de 200 países e territórios [34], e possuem previsão de implantação em mais 9 regiões, como pode ser visto na Figura 2.4.

A lista completa de serviços que compõem o portfólio do GCP oferece mais de 100 produtos, dentre eles os serviços que mais se destacam são [35]:

1. **Compute Engine:** componente que ofereceu serviço de Infraestrutura como Serviço do GPC, o qual permite que os usuários iniciem máquinas virtuais sob demanda;
2. **Serviços de Armazenamento:** os serviços de armazenamento disponíveis pelo provedor serão detalhados no Capítulo 3, pois é o foco deste trabalho;
3. **App Engine:** uma plataforma de computação em nuvem para desenvolver e hospedar aplicações *web* na infraestrutura do Google;
4. **Google Cloud Functions:** funções como serviço (FaaS) escalonáveis e de pagamento por utilização para executar seu código, sem a necessidade de gerenciar servidores;

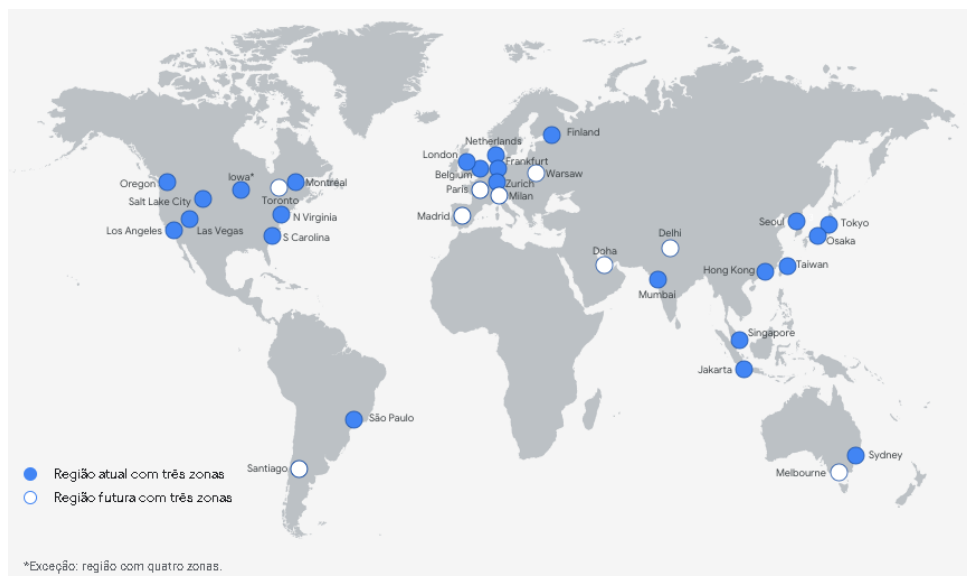


Figura 2.4: Infraestrutura Global do GCP (Fonte [34]).

5. **Cloud SQL:** serviço de banco de dados relacional, totalmente gerenciado, para MySQL, PostgreSQL e SQL Server;
6. **Big Query:** serviço de *data warehouse*, sem necessidade de servidor, que oferece análises em escala de petabytes para facilitar consultas SQL;
7. **Nuvem Privada Virtual (VPC):** serviço de rede gerenciada, uma versão virtual de uma rede física, implementada dentro da rede de produção do Google;
8. **Cloud CDN:** serviço que utiliza balanceamento de cargas e a rede de borda global do Google para exibir conteúdo mais relevante aos usuários, o que acelera seus sites e aplicativos;
9. **Cloud Run:** permite o desenvolvimento e a implantação de aplicativos em contêineres altamente escalonáveis em uma plataforma totalmente gerenciada e sem servidor.

2.5.3 Microsoft Azure

O Microsoft Azure foi lançado em 2010 com a intenção de fornecer uma plataforma de computação em nuvem competente para empresas, e desde então tem obtido sucesso, tornando-se, atualmente, a maior concorrente da gigante AWS.

Assim como o GCP e a AWS, a sua infraestrutura também é dividida em regiões e zonas de disponibilidade [36]. Ao todo são 56 regiões, mas com planos anunciados de expansão de mais 17 como pode ser visto na Figura 2.5. Isso tornará a Microsoft Azure

a maior infraestrutura de nuvem computacional do Mundo, com mais regiões globais do que qualquer outro provedor em nuvem.



* As regiões de acesso reservado não são mostradas

Figura 2.5: Infraestrutura Global do Microsoft Azure (Fonte [36]).

A plataforma de nuvem do Microsoft Azure consiste em mais de 200 produtos e serviços, dentre estes, destacam-se [37]:

1. **Máquinas Virtuais:** serviço de provisionamento de máquinas virtuais (VMs) Linux e Windows, escalável sob demanda;
2. **Serviços de Armazenamento:** os serviços de armazenamento disponíveis pelo provedor serão detalhados no Capítulo 3, pois é o foco deste trabalho;
3. **Azure SQL:** banco de dados SQL em nuvem, gerenciado e fornecido como parte do Microsoft Azure;
4. **Azure Cosmos DB:** um serviço de banco de dados NoSQL totalmente gerenciado para o desenvolvimento de aplicativos modernos;
5. **Azure Kubernetes Service (AKS):** implantação e gerenciamento de aplicativos em contêineres, que fornece Kubernetes sem servidor;
6. **Azure Functions:** uma plataforma de computação sem servidor orientada a eventos que também pode resolver problemas complexos de orquestração;

7. **Azure Cognitive:** uma coleção de APIs hospedadas na nuvem que permite aos desenvolvedores adicionarem facilmente recursos de inteligência artificial para visão, fala, linguagem, conhecimento e pesquisa em aplicativos, em dispositivos e plataformas, como iOS, Android e Windows;
8. **App Service:** uma plataforma totalmente gerenciada para construir, implantar e dimensionar seus aplicativos da *web*;
9. **Rede Virtual:** serviço de rede privada na nuvem.

2.6 Considerações Finais

Neste capítulo foram abordados os conceitos de computação em nuvem, suas características essenciais, seus modelos de serviço, bem como os principais provedores de nuvem pública.

Ademais, como visto no decorrer deste capítulo, cada provedor de nuvem oferece diferentes serviços para seus clientes. Logo, é importante que cada usuário conheça as suas reais demandas e as ofertas de cada provedor de nuvem, pois assim é possível que as escolhas certas sejam feitas. No próximo capítulo, serão apresentados os principais serviços de armazenamento de cada um destes provedores citados.

Capítulo 3

Armazenamento em Nuvem

Neste capítulo serão apresentados os principais conceitos dos serviços de armazenamento em nuvem. O capítulo está dividido em três seções, sendo a primeira Seção 3.1 responsável por apresentar uma visão geral e as características que definem o armazenamento em nuvem. A Seção 3.2 ressalta os mais populares tipos de armazenamento e, por fim, a Seção 3.3 descreve em detalhes os serviços de armazenamento oferecidos atualmente pelos principais provedores.

3.1 Visão Geral do Armazenamento em Nuvem

O armazenamento em nuvem é um dos principais serviços providos em nuvens computacionais, no qual os provedores de serviço fornecem aos clientes capacidade de armazenamento para vários tipos de dados por meio da Internet, enquanto isso, os clientes não precisam conhecer os detalhes das estruturas e dos mecanismos físicos [38].

O armazenamento em nuvem é ofertado por meio de técnicas de virtualização de recursos pelos provedores de serviço em nuvem. Virtualizar o armazenamento significa abstrair o armazenamento lógico do armazenamento físico, ao consolidar todos os dispositivos de armazenamento disponíveis em um *datacenter*, permitindo a criação de discos virtuais independentes do dispositivo e da localização [39].

Estes discos virtuais, comumente pertencem e são gerenciados por provedores, os quais são responsáveis por manter os dados disponíveis e acessíveis, bem como o ambiente físico protegido e funcionando. Assim os clientes podem aproveitar a capacidade de armazenamento para guardar dados estruturados, semi-estruturados e/ou não-estruturados [40]. Para isso, há uma grande diversidade na forma como esses dados são gerenciados, mantidos e organizados pelos provedores. Dessa forma, os provedores oferecem um amplo leque para os serviços de armazenamento. Este capítulo apresentará os principais tipos nas próximas seções.

3.2 Tipos de Armazenamento

Os tipos de armazenamento em nuvem variam das diferentes maneiras que guardam, organizam e apresentam os dados. Cada um desses tipos, trabalham com diferentes formatos de dados e possuem um conjunto exclusivo de características, com capacidades e limitações que se adequam a diferentes casos de uso [41]. Nesta seção serão abordadas as opções de armazenamento mais populares no mercado de nuvem.

3.2.1 Armazenamento por Objetos

O armazenamento por objetos é uma estratégia que gerencia e manipula o armazenamento de dados como unidades distintas, chamadas de objetos. Estes objetos são compostos por três partes [42]:

- Dados: o dado do objeto armazenado, que pode ser de qualquer tipo, como um vídeo, uma foto, um arquivo de texto e etc;
- Metadados: é a informação contextual dos dados armazenados, tais como sua utilidade e sua confidencialidade, ou qualquer outra característica que seja relevante para a maneira como os dados são usados;
- Identificador: refere-se a um identificador global exclusivo permitindo que o objeto seja encontrado em um sistema distribuído de armazenamento.

A motivação para se utilizar este tipo de armazenamento é a necessidade de ter sistemas de armazenamento capazes de realizar mais trabalho computacional de E/S, aliviando assim, os *hosts* para que possam focar em outros trabalhos de processamento [43].

Em nuvens computacionais esses objetos são armazenados em repósitorios comumente chamados de *buckets*, que são acessíveis por protocolo HTTP. Os *buckets* são como diretórios e cada um deles pode ter um ou mais objetos [44]

3.2.2 Armazenamento por Bloco

Neste tipo de armazenamento os dados são guardados em volumes, nos quais os arquivos são divididos em blocos de tamanhos iguais. Cada bloco possui seu próprio endereço, mas ao contrário dos objetos, eles não possuem metadados. Este armazenamento em bloco pode ser acessado pelo sistema operacional e montado como um *drive*.

Em nuvens computacionais sistemas de armazenamento em bloco são usados para hospedar bancos de dados, dar suporte a operações aleatórias de leitura/gravação e manter arquivos de sistema das máquinas virtuais em execução [45].

O sistema de armazenamento em nuvem em nível de bloco fornece aos usuários recursos de armazenamento em nível de bloco por meio de SCSI (Small Computer System Interface) ou FC (*Fibre Channel*) [46].

Geralmente este tipo de armazenamento é utilizado por meio do protocolo iSCSI (*Internet Small Computer System Interface*), que encapsula comandos SCSI em pacotes TCP/IP e transfere os dados por meio de rede Ethernet. O iSCSI permite que aplicativos executados em uma única máquina cliente compartilhem dados remotos, mas não é diretamente adequado para compartilhar dados entre várias máquinas [47].

3.2.3 Armazenamento por Arquivos

O armazenamento por arquivos permite o acesso a arquivos de múltiplos *hosts* compartilhando por meio de uma rede de computadores e, portanto, possibilitam que diversos usuários compartilhem arquivos e recursos de armazenamento [43].

Este tipo de armazenamento implementa um sistema hierárquico que oferece acesso compartilhado a dados de arquivos. Os usuários podem criar, excluir, modificar, ler e gravar arquivos, bem como organizá-los logicamente em árvores de diretório para oferecer um acesso intuitivo [48].

O sistema de armazenamento em nível de arquivo refere-se principalmente ao sistema de arquivo distribuído. Ele fornece serviços de acesso ao armazenamento em nível de arquivo por meio do protocolo Network File System (NSF) [46].

Um NAS é um servidor de arquivos que normalmente apresenta uma interface de arquivo para a rede empregando CIFS (*Common Internet Files System*) ou NFS (*Network File System*). Um NAS típico se conecta a uma rede de área local e fornece serviços de compartilhamento de arquivos para muitos clientes [47].

3.3 Serviços de Armazenamento das Principais Nuvens Públicas

As principais nuvens públicas possuem serviços disponíveis para atender os mais comuns tipos de armazenamento que foram citados neste capítulo. Todavia, embora lidem com o mesmos formatos de dados, possuem particularidades que as diferenciam na entrega destes serviços.

3.3.1 Serviços de Armazenamento por Objetos

Os serviços de armazenamento por objeto são utilizados para entregar recursos para armazenar e proteger qualquer volume de dados em vários casos de uso. Os principais provedores fornecem estes serviços em diferentes classes, as quais se diferenciam pela frequência de acesso aos dados armazenados.

Amazon Web Service

O Amazon Simple Storage Service, popularmente conhecido como Amazon S3, é um dos serviços de armazenamento por objeto pioneiros da computação em nuvem. Esse serviço é dividido em seis classes [49]:

1. **Amazon S3 Standard Storage:** oferece um armazenamento por objetos com altos níveis de resiliência, disponibilidade e performance para dados acessados com frequência. Como fornece baixa latência e alto *throughput*, o S3 Standard é adequado para uma grande variedade de casos de uso, como aplicativos na nuvem, sites dinâmicos, aplicativos móveis, *big data* e etc;
2. **Amazon S3 Intelligent-Tiering:** foi projetada para otimizar os custos movendo automaticamente os dados para o nível de acesso mais econômico, sem sobrecarga operacional;
3. **Amazon S3 Standard-Infrequent Access:** indicado para dados acessados com menos frequência, mas que exigem acesso rápido quando necessários. A combinação de baixo custo e alta performance tornam esta classe ideal para armazenamento de longa duração, *backups* e *datastores* para arquivos de recuperação de desastres;
4. **Amazon S3 One Zone-Infrequent Access:** indicado para dados acessados com menos frequência, mas que exigem acesso rápido quando necessários. Ao contrário de outras classes de armazenamento S3, que armazenam dados em, no mínimo, três zonas de disponibilidade (AZs), a S3 One Zone-IA armazena dados em uma única AZ, com um custo 20% inferior ao S3 Standard-IA;
5. **Amazon S3 Glacier:** classe de armazenamento em nuvem do Amazon S3 seguro, resiliente e de custo extremamente baixo para arquivamento de dados e *backups* de longa duração;
6. **Amazon S3 Glacier Deep Archive:** classe de armazenamento mais barata do Amazon S3, oferece suporte à retenção e preservação digital de longo prazo para dados que podem ser acessados uma ou duas vezes por ano. Esta classe é projetada para clientes que mantêm conjuntos de dados por 7 à 10 anos ou mais, para

cumprir requisitos de conformidade normativa, especialmente, em setores altamente regulados como serviços financeiros, saúde e setores públicos.

Diante dos serviços descritos, a Tabela 3.1 apresenta as principais diferenças entre essas classes de armazenamento do Amazon S3.

Tabela 3.1: Comparativo de Armazenamento S3 (Fonte Adaptada [49]).

Classe	S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier	S3 Glacier Deep Archive
Disponibilidade	99,99%	99,9%	99,9%	99,5%	99,99%	99,99%
Custo por repouso	por GB recuperado	por GB recuperado	por GB recuperado	por GB recuperado	por GB recuperado	por GB recuperado
Duração mínima de armazenamento	N/D	30 dias	30 dias	30 dias	90 dias	180 dias
Taxa de recuperação	N/D	N/D	por GB recuperado	por GB recuperado	por GB recuperado	por GB recuperado
Latência (tempo até o primeiro byte)	milissegundos	milissegundos	milissegundos	milissegundos	minutos ou horas	horas

Google Cloud Platform

O GCP [50] oferece o serviço de armazenamento por objetos chamado de Cloud Storage, que assim como a Amazon, concede classes diferentes, e para todas elas pode ser definido se os dados serão replicados regionalmente ou em múltiplas regiões [51]. Essas são as quatro classes disponíveis pela GCP:

1. **Cloud Storage Standard:** recomendado para dados que precisam ser acessados com muita frequência, incluindo *sites*, *streaming* de vídeos e APPs para dispositivos móveis;
2. **Nearline Storage:** recomendado para dados que podem ser armazenados por, no mínimo 30 dias, incluindo *backup* de dados e conteúdo multimídia infrequente;
3. **Coldline Storage:** recomendado para dados que podem ser armazenados por, no mínimo 90 dias, incluindo recuperação de desastres;
4. **Archive Storage:** recomendado para dados que podem ser armazenados por, no mínimo 365 dias, incluindo arquivos regulatórios.

Assim, no comparativo da Tabela 3.2 são apresentadas as principais diferenças entre as classes de armazenamento por objeto do Google Cloud Platform.

Tabela 3.2: Comparativo de Armazenamento Cloud Storage (Fonte Adaptada [52]).

Classe	Cloud Storage Standard	Nearline Storage	Coldline Storage	Archive Storage
Disponibilidade	99,99%	99.9%	99.9%	99.9%
Custo por repouso	por GB mensal	por GB mensal	por GB mensal	por GB mensal
Duração mínima de armazenamento	N/D	30 dias	90 dias	365 dias
Taxa de recuperação	N/D	por GB recuperado	por GB recuperado	por GB recuperado
Latência (tempo até o primeiro byte)	Milissegundos	Milissegundos	Milissegundos	Milissegundos

Microsoft Azure

A Microsoft, por meio de sua nuvem Azure, disponibiliza o serviço Blob Storage. Neste serviço é possível definir quatro opções de redundância de dados: redundância local, redundância por zona de disponibilidade, redundância geográfica e redundância geográfica apenas para leitura [53]. O armazenamento do Azure oferece diferentes níveis de acesso, permitindo que você armazene dados de objeto de maneira mais econômica. As camadas de acesso disponíveis incluem:

1. **Premium Performance:** ideal para cargas de trabalho que exigem tempos de resposta rápidos e consistentes. Os dados são armazenados em unidades de estado sólido (SSDs), que são otimizadas para baixa latência;
2. **Hot Tier:** otimizado para armazenar dados que são acessados com frequência, baseados em cargas de trabalho com uso intensivo de E/S que exigem latência de armazenamento baixa e uniforme;
3. **Cool Tier:** armazena dados mais antigos não utilizados com frequência, mas que devem estar disponíveis imediatamente quando acessados. Esta camada é destinada a dados que permanecerão por, pelo menos, 30 dias. Exemplos de cenários neste caso incluem *backups* de curto prazo e recuperação de desastres;
4. **Archive Tier:** otimizado para armazenar dados raramente acessados e armazenados por, pelo menos, 180 dias, com requisitos de latência flexíveis, na ordem de horas. Esse serviço é indicado para *backups* de longo prazo, *backups* secundários e conformidade.

Na Tabela 3.3 é possível observar as principais diferenças de cada modelo de armazenamento por objeto do provedor Azure.

Tabela 3.3: Comparativo de Armazenamento Blob Storage (Fonte Adaptada [54]).

Classe	Premium Performance	Hot Tier	Cool Tier	Archive Tier
Disponibilidade	99.9%	99.9%	99%	Offline
Custo por repouso	por GB mensal	por GB mensal	por GB mensal	por GB mensal
Duração mínima de armazenamento	N/D	N/D	30 dias	180 dias
Taxa de recuperação	N/D	N/D	por GB recuperado	por GB recuperado
Latência (tempo até o primeiro byte)	Milissegundos de um dígito	Milissegundos	Milissegundos	Horas

3.3.2 Serviços de Armazenamento por Blocos

Os serviços de armazenamento em blocos comumente são utilizados para entregar recursos de armazenamento de maneira elástica para máquinas virtuais hospedadas em nuvem. Os principais provedores fornecem estes serviços em diferentes tipos, com o intuito de atender demandas de cargas de trabalho específicas.

Amazon Web Services

O Elastic Block Store (EBS) [55], serviço disponibilizado pela Amazon, fornece volumes persistentes de armazenamento em nível de bloco para uso com instâncias do Amazon EC2. Cada volume do Amazon EBS é replicado automaticamente dentro da sua zona de disponibilidade, para protegê-lo contra falhas de componentes, oferecendo alta disponibilidade e durabilidade. Os volumes do Amazon EBS estão disponíveis em tipos que diferem nas características de desempenho e no preço:

1. **Cold HDD:** baseados em unidades de disco rígido, podem ser utilizados para cargas de trabalho com conjuntos de dados menores, em que os dados são acessados com pouca frequência ou quando a uniformidade da performance não é de fundamental importância;
2. **Throughput Optimized HDD:** baseados em unidades de disco rígido, de baixo custo projetado para cargas de trabalho com alto rendimento e acessadas com frequência;
3. **General Purpose SSD:** oferecem armazenamento econômico, ideal para uma gama de cargas de trabalho. Eles oferecem alto desempenho a um preço moderado, adequado para uma ampla variedade de cargas de trabalho;
4. **Provisioned IOPS SSD:** projetados para atender às necessidades de cargas de trabalho intensivas de E/S, particularmente cargas de trabalho de banco de dados sensíveis ao desempenho de armazenamento e consistência na taxa de transferência de E/S de acesso aleatório.

Tabela 3.4: Comparativo de Armazenamento EBS (Fonte Adaptada [56]).

Tipo de volume	Cold HDD	Throughput Optimized HDD	General Purpose SSD	Provisioned IOPS SSD
Tipo de disco	HDD	HDD	SSD	SDD
Descrição	Adequados para grandes cargas de trabalho sequenciais de dados frios	Cargas de trabalho frequentemente acessadas e com alto consumo da taxa de transferência	Uso geral, para uma ampla variedade de cargas de trabalho transacionais	Cargas de trabalho transacionais sensíveis à latência e essenciais para os negócios
Capacidade máxima por disco	16 TB	16 TB	16 TB	64 TB
Taxa de transferência máxima	250 MB/s	500 MB/s	1.000 MB/s	4.000 MB/s
IOPS máximos	250	500	16.000	256.000

Google Cloud Platform

O Persistent Disk [57] fornece às instâncias virtuais do Google Cloud (Compute Engines), recursos de armazenamento de que essas instâncias precisam para operar. O nível de flexibilidade e as opções de configuração disponíveis tornam o disco permanente um serviço de armazenamento em nuvem muito acessível para clientes novos na computação em nuvem. O usuário pode começar escolhendo uma região geográfica que melhor atenda aos requisitos da carga de trabalho. A seguir serão mostrados os três tipos de discos permanentes:

1. **Standard Persistent Disks:** adequados para grandes cargas de trabalho de processamento de dados que usam principalmente E/S sequenciais;
2. **Balanced Persistent Disks:** uma alternativa aos discos permanentes SSD que equilibram desempenho e custo. Com o mesmo IOPS máximo que os discos permanentes SSD e IOPS mais baixo por GB, o Persistent Disks balanceado oferece níveis de desempenho adequados para a maioria dos aplicativos de uso geral, a um preço entre o dos discos permanentes padrão e SSD;
3. **SSD Persistent Disks:** apropriados para aplicativos corporativos e necessidades de banco de dados de alto desempenho que exigem menor latência e mais IOPS do que os discos permanentes padrão fornecem. Os discos permanentes SSD são projetados para latências de milissegundos de um dígito, e a latência observada é específica do aplicativo.

Microsoft Azure

O Managed Disk [59] é projetado para ser utilizado com máquinas virtuais do Azure. Este serviço oferece armazenamento em bloco altamente durável e de alto desempenho para seus aplicativos críticos. A Microsoft oferece quatro tipos de Managed Disks:

Tabela 3.5: Comparativo de Armazenamento Persistent Disks (Fonte Adaptada [58]).

Tipo de volume	Standard Persistent Disks	Balanced Persistent Disks	SSD persistent Disks
Tipo de disco	HDD	SSD	SSD
Descrição	Armazenamento em blocos eficiente e confiável	Armazenamento em bloco econômico e confiável	Armazenamento em bloco rápido e confiável
Capacidade máxima por disco	64 TB	64 TB	64 TB
Taxa de transferência máxima	750 (Leitura) / 1.500 (Escrita)	6000	30000
IOPS máximos	1.200 MB/s	2.800 MB/s	4.800 MB/s

1. **Ultra Disk:** indicado para cargas de trabalho com uso intensivo de E/S, como SAP Hana, bancos de dados de camada superior (por exemplo SQL e Oracle) e outras cargas de trabalho de transações pesadas;
2. **Premium SSD Managed Disks:** ideal para cargas de trabalho sensíveis à produção e ao desempenho;
3. **Standard SSD Managed Disks:** apropriado para servidores *web*, aplicativos empresariais pouco usados e desenvolvimento/teste;
4. **Standard HDD Managed Disks:** conveniente para ambientes de *backup*, não com uma criticidade menor e acesso infrequente.

Na Tabela 3.6 é possível observar as principais diferenças de cada modelo de armazenamento por bloco ofertado pela nuvem da Azure.

Tabela 3.6: Comparativo de Armazenamento Managed Disk (Fonte Adaptada [60]).

Tipo de volume	Ultra disk	Premium SSD	Standard SSD	Standard HDD
Tipo de disco	SSD	SSD	SSD	HDD
Descrição	Cargas de trabalho com uso intensivo de E/S, como SAP Hana ou Bancos de Dados	Cargas de trabalho sensíveis à produção e ao desempenho	Servidores Web, aplicativos empresariais pouco usados e desenvolvimento/teste	Backup, não crítico, acesso não frequente
Capacidade máxima por disco	65.536 GiB	32.767 GiB	32.767 GiB	32.767 GiB
Taxa de transferência máxima	2.000 MB/s	900 MB/s	750 MB/s	500 MB/s
IOPS máximos	160.000	20.000	6.000	2.000

3.3.3 Serviços de Armazenamento por Arquivos

Os serviços de armazenamento por arquivos implementam um sistema hierárquico que oferece acesso a dados de arquivos, possibilitando que diversos usuários compartilhem arquivos e recursos de armazenamento. Os principais provedores fornecem estes serviços em diferentes classes, as quais se diferenciam pela frequência de acesso aos dados armazenados.

Amazon Web Services

O Amazon Elastic File System (Amazon EFS) [61] oferece um sistema de arquivos elástico, por meio de uma interface de serviços *web*, sem servidor (do inglês, *serverless*). Ele foi desenvolvido para escalar sob demanda sem interromper os aplicativos, aumentando e diminuindo automaticamente conforme você adiciona e remove arquivos, eliminando a necessidade de provisionar e gerenciar a capacidade para acomodar o crescimento.

O Amazon EFS oferece suporte ao protocolo Network File System versão 4 (NFSv4.1 e NFSv4.0), portanto, os aplicativos e as ferramentas comumente utilizados em *datacenters* locais funcionam perfeitamente.

Com o Amazon EFS, o usuário paga apenas pelo armazenamento usado por seu sistema de arquivos e não há taxa mínima ou custo de instalação. O Amazon EFS oferece uma variedade de classes de armazenamento projetadas para diferentes casos de uso. Esses incluem [62]:

1. **EFS Standard:** concebida para cargas de trabalho de sistemas de arquivos ativa. Os dados são armazenados regionalmente dentro e entre várias Zonas de Disponibilidade (AZs).
2. **EFS Standard – Infrequent Access (Standard – IA):** tem custos otimizados para arquivos acessados com menos frequência. Os dados armazenados na classe de armazenamento EFS Standard-IA custam menos em comparação aos da classe de armazenamento EFS Standard;
3. **EFS One Zone:** elaborada para cargas de trabalho de sistemas de arquivos ativa. Os dados são armazenados em uma única Zona de disponibilidade. As taxas padrão de transferência de dados são aplicadas no acesso de sistemas de arquivos entre zonas de disponibilidade ou entre regiões;
4. **EFS One Zone – Infrequent Access (EFS One Zone – IA):** os dados armazenados na classe de armazenamento EFS One Zone-IA custam menos em comparação aos da classe de armazenamento One Zone. Neste serviço é paga uma taxa sempre que houver leitura ou gravação em um arquivo. Os dados são armazenados em uma única zona de disponibilidade. As taxas padrão de transferência de dados são aplicadas no acesso de sistemas de arquivos entre zonas de disponibilidade ou entre regiões.

Google Cloud Platform

O Google Cloud Filestore [63] é um serviço de armazenamento de arquivos gerenciado pelo Google. O Filestore oferece compatibilidade nativa para aplicativos empresariais atuais

e dá suporte a qualquer cliente compatível com NFSv3. Aplicativos recebem o benefício de recursos como desempenho de escalonamento horizontal, capacidade de centenas de TB e bloqueio de arquivos, sem a necessidade de instalar ou manter qualquer *plug-in* especializado ou software do lado do cliente. Este serviço pode ser adquirido nas classes:

1. **Básico (HDD):** indicado para uso geral, incluindo teste e desenvolvimento;
2. **Básico (SSD):** oferece alto desempenho, mas com uma capacidade limitada;
3. **Alto Escalonamento (SSD):** o melhor desempenho de categoria, projetado para grandes capacidades.

Microsoft Azure

O serviço Azure Files [64] oferece compartilhamento de arquivos totalmente gerenciados na nuvem que são acessíveis por meio do Protocolo SMB (*Server Message Block*), ou do Protocolo NFS (*Network File System*), padrões do setor.

Os compartilhamentos podem ser montados de maneira simultânea por implantações locais ou na nuvem. É possível acessar os compartilhamentos de Arquivos do Azure do protocolo SMB em clientes Windows, Linux e MacOS.

É possível acessar os compartilhamentos do protocolo NFS em clientes Linux e MacOS. Além disso, os compartilhamentos do protocolo SMB podem ser armazenados em cache nos Windows Servers com o Azure File Sync [65], para acesso rápido perto de onde os dados estão sendo utilizados. A Microsoft oferece o Azure Files em quatro tipos de camadas:

1. **Premium:** permite cargas de trabalho com E/S elevada, com alta taxa de transferência e baixa latência. Os compartilhamentos de arquivos Premium são oferecidos no armazenamento baseado em SSD (unidade de estado sólido) de alto desempenho;
2. **Transação otimizada:** ideal para cargas de trabalho de transação intensas, que não precisam da latência oferecida por compartilhamentos de arquivos Premium. Os compartilhamentos de arquivos com otimização de transação são uma ótima opção para aplicativos que exigem armazenamento de arquivos ou para armazenamento de *back-end*;
3. **Quente:** são otimizados para armazenamento de cenários de uso geral, como compartilhamentos de equipe e a sincronização de arquivos do Azure;
4. **Fria:** são um recurso econômico e otimizado para cenários de armazenamento de arquivos *online*.

3.4 Considerações Finais

Neste capítulo foram abordados os principais conceitos de armazenamento em nuvem, dando ênfase aos três mais relevantes tipos de armazenamento disponíveis no mercado, o armazenamento por objetos, o armazenamento por blocos e o armazenamento por arquivos, demonstrando suas características e particularidades. Em seguida, foram apresentados para cada tipo de armazenamento os serviços ofertados pelos principais provedores de nuvens públicas. No próximo capítulo, serão apresentados os resultados obtidos na comparação dos serviços de armazenamento de objetos, por meio da utilização de uma ferramenta de *benchmark*.

Capítulo 4

Análise dos Resultados

Neste capítulo serão apresentados os resultados obtidos a partir deste trabalho, comparando o armazenamento de objeto entre os principais provedores de nuvens públicas. O capítulo é composto por três seções, sendo a Seção 4.1 responsável por mostrar uma visão geral da ferramenta de *benchmark* utilizada no comparativo. A Seção 4.2 é voltada para a arquitetura e metodologia utilizada nos testes. Por último a Seção 4.3 apresenta uma análise dos resultados obtidos.

4.1 Benchmark

O *benchmarking* escolhido para os testes foi o Cloud Object Storage Benchmarking (COSBench) desenvolvido pela Intel [66]. Esta ferramenta pode ser baixada gratuitamente na plataforma do GitHub¹ e é *open source*, o que significa que possui o código aberto.

O COSBench utiliza uma arquitetura escalável e distribuída, que consiste em dois componentes principais: COSBench Driver e COSBench Controller. O primeiro (COSBench Driver) é o responsável por gerar as cargas de trabalho para simular muitos clientes. Isso será realizado pela criação simultânea de diferentes operações de leitura e escrita em um armazenamento de objeto em nuvem. O COSBench Controller é o responsável por controlar a execução dos *drivers*, selecionando os parâmetros de carga de trabalho, coletando resultados, e gerando estatísticas agregadas [67].

O COSBench Controller e o COSBench Driver podem ser implantados no mesmo nó ou em nós diferentes. O nó pode ser uma máquina física ou uma instância de máquina virtual (VM). Para garantir a facilidade de uso, a ferramenta foi adaptada para ser acessível a partir de navegadores web e utilitários de linha de comando, tornando-a amigável para os usuários e as ferramentas de automação. Os usuários podem usar essas interfaces para

¹<https://github.com/intel-cloud/cosbench>

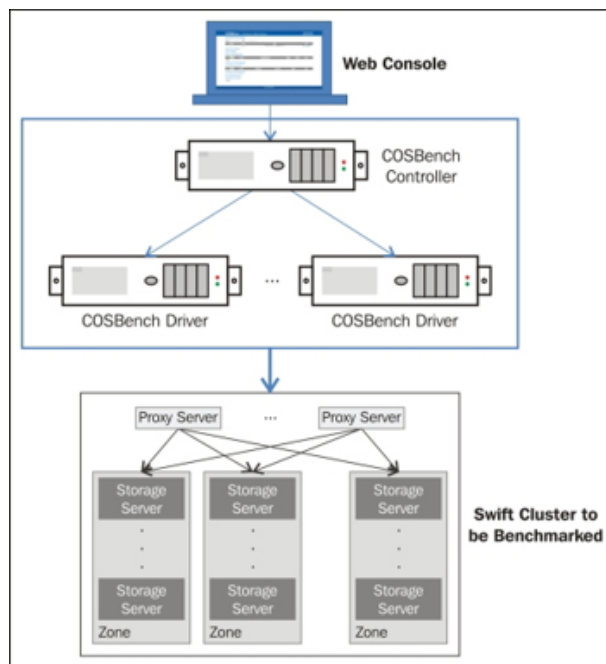


Figura 4.1: Arquitetura do CosBench.

enviar ou cancelar cargas de trabalho, verificar o status do tempo de execução, e examinar ou exportar relatórios finais [38].

Para que o COSbench em nosso modelo de carga de trabalho funcione com a maioria dos sistemas de armazenamento de objetos, definimos nossa interface de armazenamento como um pequeno pacote de operações comuns vistas em vários protocolos específicos. Essas operações são suficientes para assumir tarefas como localização de gargalos e medição de capacidade, tornando essa ferramenta de benchmark bastante prática.

4.1.1 Métricas

Para analisar o desempenho dos principais serviços de armazenamento de objeto em nuvem disponíveis no mercado, os testes foram realizados com diferentes parâmetros e cenários:

1. Taxa de Transferência: pode ser definida como a quantidade de dados movidos com êxito de uma origem para um destino em um determinado período. A taxa de transferência é medida em operações por segundo (OPs), que é definida pela divisão entre o total de solicitações bem-sucedidas e o tempo total de execução;
2. Tempo de Resposta: apresenta a duração entre o início e a conclusão da operação. Essa métrica é demonstrada pela média de todas as solicitações bem-sucedidas;

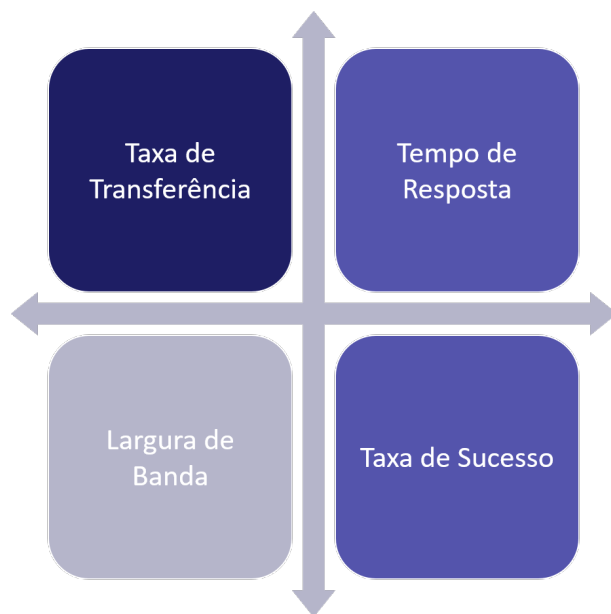


Figura 4.2: Principais métricas.

3. Largura de Banda: determina a capacidade de transmissão, analisando a velocidade que os dados trafegam através de uma rede específica. A largura de banda é calculada dividindo o total de *megabytes* transferidos por segundo (MB/s);
4. Taxa de Sucesso: indica a proporção de operações bem-sucedidas. É representada pelo cálculo de dividir o total de solicitações bem-sucedidas pelo número total de solicitações.

4.2 Testes

No ambiente de experimentação, o COSBench foi instalado em sua versão 0.4.2 em uma máquina virtual linux na distribuição Ubuntu 12.04.1 LTS Desktop, em uma infraestrutura local visto que se fosse implementado em alguma nuvem computacional isso poderia interferir nos resultados esperados.

O CosBench foi empregado para submeter a carga de trabalho e coletar métricas de desempenho para os sistemas de armazenamento S3 da AWS, Blob Storage da Microsoft Azure, e Cloud Storage da Google Cloud Platform.

O COSBench mede o desempenho de leitura e escrita. Para isso, foi utilizado um perfil de 50% leitura e 50% escrita, e os resultados exibidos em cada métrica refletem a média entre esses perfis. Neste estudo foram executados testes variando as taxas de leitura e escrita entre 30% e 70%. No entanto, a média do comportamento não variou. Por esse motivo foram considerados perfis balanceados (50%). A medição do desempenho

Tabela 4.1: Escopo de testes.

Teste	Tipo	Tamanho do Objeto	Leitura	Escrita
1	Pequeno	64 KB	50%	50%
2	Pequeno	64 KB	50%	50%
3	Pequeno	64 KB	50%	50%
4	Médio	1 MB	50%	50%
5	Médio	1 MB	50%	50%
6	Médio	1 MB	50%	50%
7	Grande	16 MB	50%	50%
8	Grande	16 MB	50%	50%
9	Grande	16 MB	50%	50%

da rede foi realizada em condições normais, que incluem atividades de leitura, gravação e exclusão. O tamanho do arquivo operado, ao medir o desempenho da rede é dividido em três tipos: pequeno (64 KB), médio (1 MB) e grande (16 MB), conforme proposto em [68] e também porque o tamanho ideal para ser avaliado é totalmente dependente da taxa de transmissão, que no caso do ambiente de testes é limitada.

Os testes foram realizados para cada tamanho de objeto descrito na Tabela 4.1, levando em consideração diferentes tamanhos de objetos. Assim sendo, foram avaliados a largura de banda, o tempo de resposta, a taxa de sucesso da operação e a taxa de transferência. Para cada uma dessas métricas, a quantidade de objetos foi variada entre 10, 100 e 500.

4.3 Resultados

Os resultados apresentados nesta seção foram obtidos com base na observação do comportamento médio a partir de quinze execuções com sucesso de cada um dos testes realizados.

4.3.1 Taxa de Transferência

A taxa de transferência é apresentada nas Figuras 4.3, 4.4 e 4.5. Na Figura 4.3, que apresenta os testes com objetos de 64KB, é possível perceber que essa taxa cai a medida em que a quantidade de objetos aumenta, para objetos dez objetos a média entre os provedores ficou em torno de 84,45 e para cem e quinhentos objetos, 84,37 e 83,55 respectivamente. No entanto, essa variação na quantidade de objetos causou uma leve piora. Assim, o aumento do tamanho dos objetos para 1MB traz uma piora mais expressiva, conforme observado na Figura 4.4, na qual há uma queda de 63% na taxa de transferência comparado com os objetos de 64KB.

Ao avaliar a Figura 4.5, que apresenta os testes com objetos de 16MB, pode-se perceber que o número de operações por segundo cai 84% se comparado a objetos de 1MB, e 94% se

comparado a objetos de 64KB. Esse comportamento foi observado para os três provedores, os quais apresentaram pouca variação entre si.

A diminuição na taxa de transferência se deu porque ao aumentar o tamanho de cada objeto o tempo para leitura e escrita do mesmo também aumenta. Além disso, quanto mais objetos precisam ser gravados e lidos em um mesmo teste, maior é a concorrência pelos recursos da rede, fazendo mais uma vez com que o tempo de execução de cada operação aumente. Esse fato pode ser observado claramente na Seção 4.3.2 que apresenta os resultados de tempo de resposta.

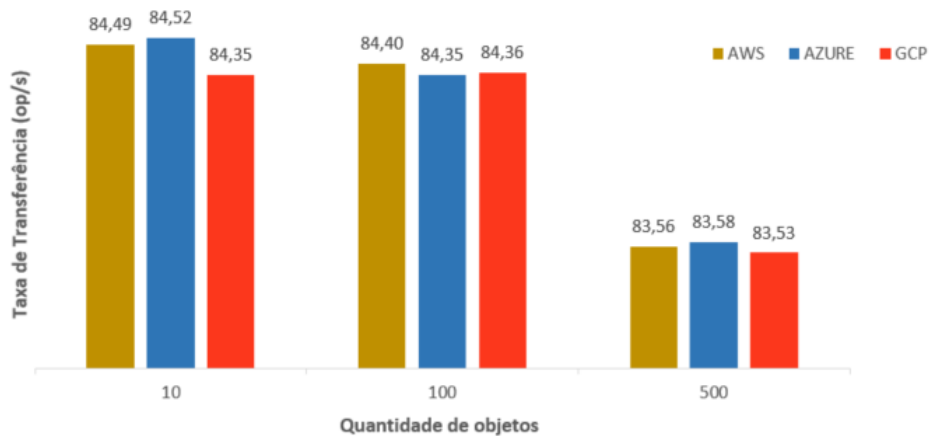


Figura 4.3: Taxa de transferência para objetos de 64KB.

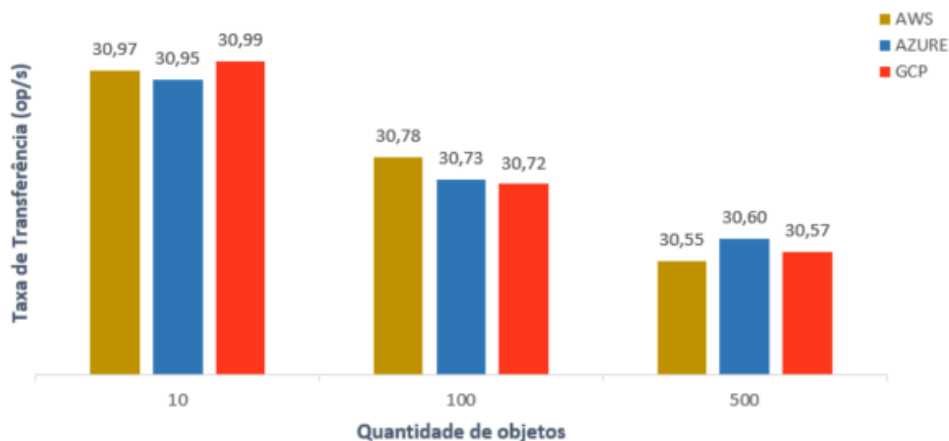


Figura 4.4: Taxa de transferência para objetos de 1MB.

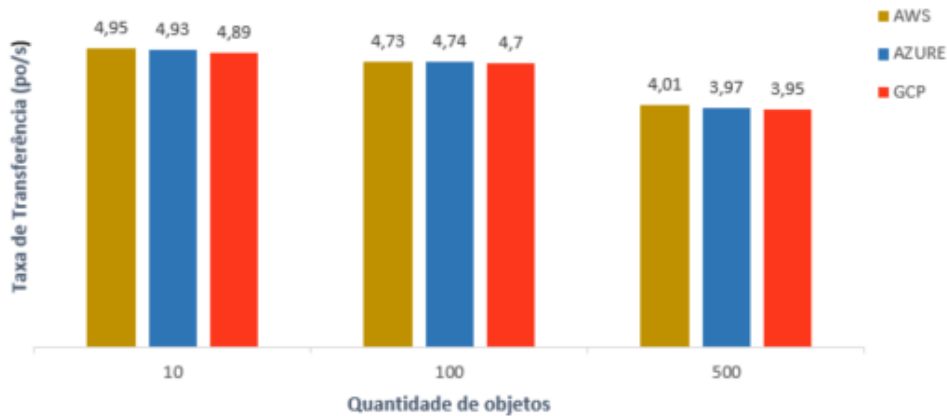


Figura 4.5: Taxa de transferência para objetos de 16MB.

4.3.2 Tempo de Resposta

O tempo de resposta é apresentado no gráfico em segundos e, representa a duração das solicitações bem-sucedidas, as quais são caracterizadas por todas as fases de operações, sendo elas: criação, leitura, escrita e exclusão dos objetos.

Os testes realizados com objetos de 64KB não apresentaram uma piora significativa com o aumento de objetos, para objetos dez objetos a média entre os provedores ficou em torno de 0,565 e para cem e quinhentos objetos, 0,561 e 0,560 respectivamente, conforme pode ser observado na Figura 4.6.

No entanto, para o aumento do tamanho de cada objeto, o tempo de duração das operações aumentou substancialmente, apresentando quase o dobro de segundos para objetos de 1MB, como pode ser observado na Figura 4.7. Além disso, quando observado objetos de 16MB o aumento foi ainda maior, passando da marca de quatro vezes do resultado de 64K e quase o triplo comparado com os resultados para 1MB, conforme apresentado na Figura 4.8.

Esse comportamento já justificado pelo aumento do tamanho dos objetos e da concorrência pelos recursos, também é influenciado pela largura de banda, como será apresentado na Seção 4.3.3.

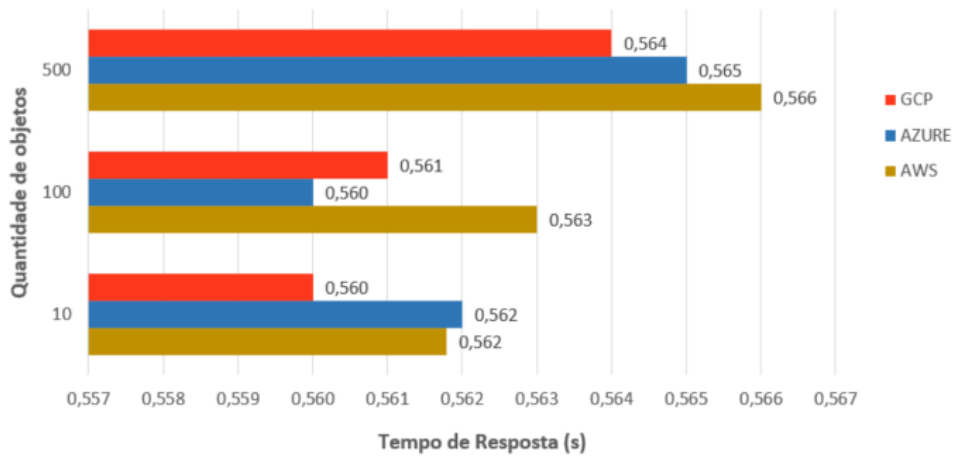


Figura 4.6: Tempo de resposta para objetos de 64KB.

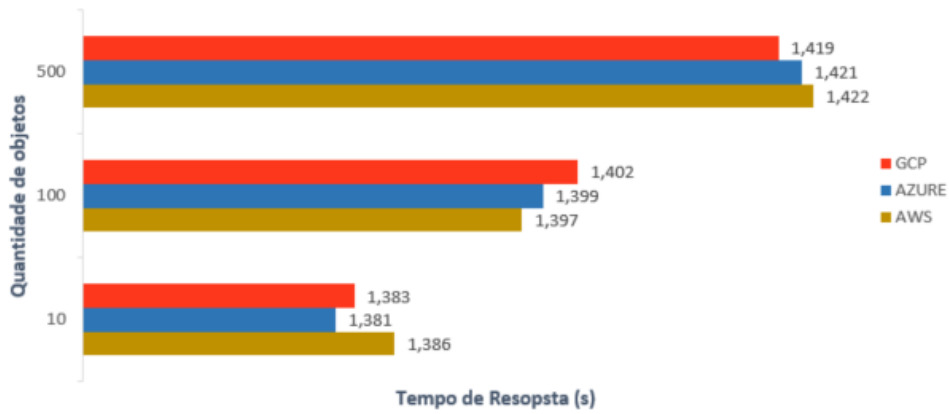


Figura 4.7: Tempo de resposta para objetos de 1MB.

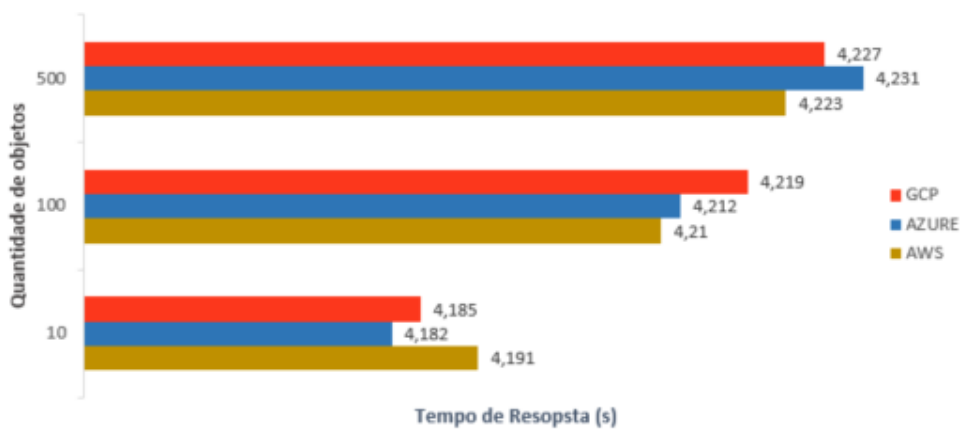


Figura 4.8: Tempo de resposta para objetos de 16MB.

4.3.3 Largura de Banda

A largura de banda determina a capacidade de transmissão, indicando a velocidade em que os dados trafegam através de uma rede específica. Assim sendo, quanto maior for a largura de banda, maior será a velocidade da conexão, visto que por ela passará mais dados ao mesmo tempo.

A largura de banda aumentou com o passar dos testes, no entanto como o link da infraestrutura utilizada possuía limitação esse aumento também foi limitado. Esse fato ocorreu, porque ao aumentar o tamanho dos objetos, as operações demandam mais recursos, portanto esse é o comportamento esperado e, pode ser observado nos graficos que são apresentados em MB/s.

Na Figura 4.9 é possível perceber que a largura de banda sofreu uma leve melhora quando começaram os testes com 500 objetos, mas o que realmente demonstra um aumento significativo da largura de banda é o aumento do tamanho dos objetos armazenados.

Quando comparamos os resultados da Figura 4.9 com a Figura 4.10 e a Figura 4.11, podemos perceber, respectivamente, um aumento aproximado de 70% e de 209%.

Ainda que a largura de banda tenha aumentado, a piora nas taxas de transferência e tempo de resposta são justificadas devido a limitação da banda, ou seja, o ganho em largura de banda não foi suficiente para manter o mesmo desempenho das operações.

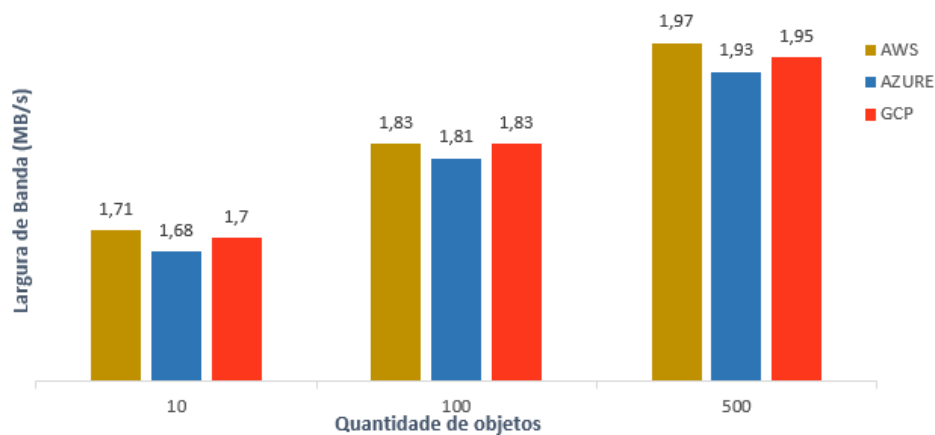


Figura 4.9: Largura de banda para objetos de 64KB.

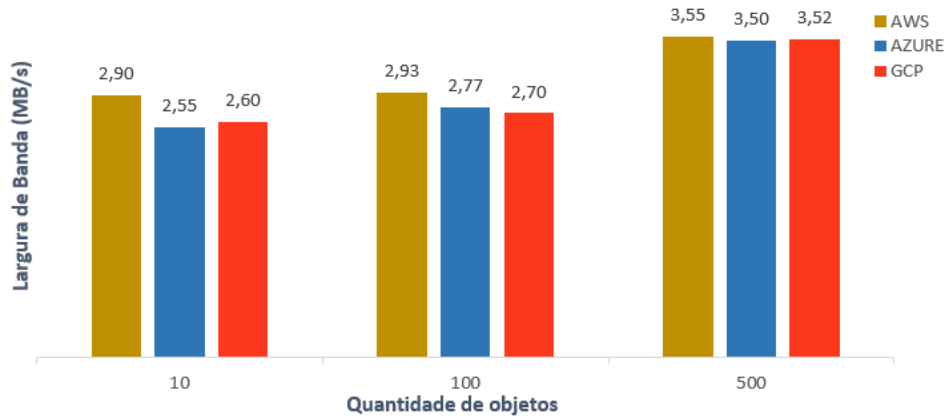


Figura 4.10: Largura de banda para objetos de 1MB.

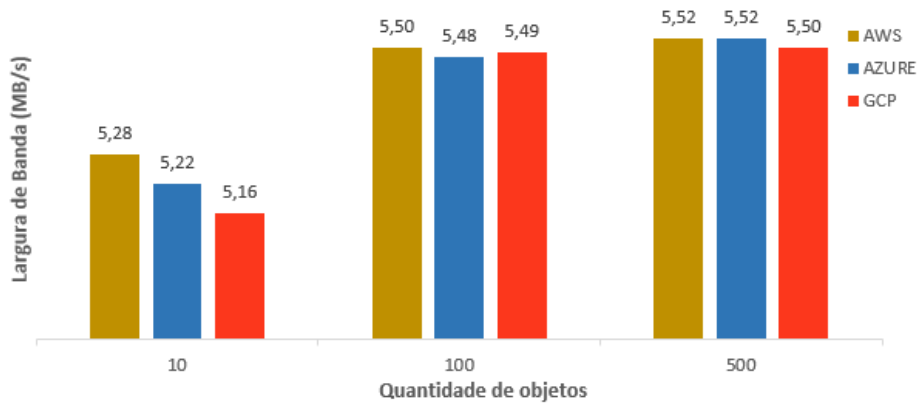


Figura 4.11: Largura de banda para objetos de 16MB.

4.3.4 Taxa de Sucesso

A taxa de sucesso é apresentada nas Figuras 4.12, 4.13 e 4.14 em porcentagem, visto que seu conceito gira em torno da proporção entre o número de solicitações bem-sucedidas pelo número total de solicitações. Assim, uma taxa de 100% significa que todas as operações requisitadas foram executadas com sucesso.

Dessa forma, ao se fazer a análise do armazenamento de objetos usando COSBench, pode-se perceber que os resultados da taxa de sucesso diminuem a medida em que o tamanho dos objetos aumentam.

Como é possível observar nas Figuras 4.12, 4.13 e 4.14, a taxa de sucesso de uma forma geral, em todos os testes realizados, ficou em níveis aceitáveis, não sendo inferiores a 80%, o que não resultou em eventuais falhas.

Esse comportamento condiz com o que foi observado na subseção anterior, visto que, com a limitação da largura de banda alguns pacotes podem ser perdidos, resultando em

falhas nas operações. Isso não quer dizer que o teste falhou, uma vez que o COSBench trata as operações "init", "prepare", "cleanup" e "descarte" como um trabalho especial que deve ser concluído sem erros para resultar no status "concluído"; erros em trabalhos especiais irão encerrar o teste. Por outro lado, o trabalho normal associado à medição de desempenho pode tolerar falhas, que são rastreadas pela "taxa de sucesso".

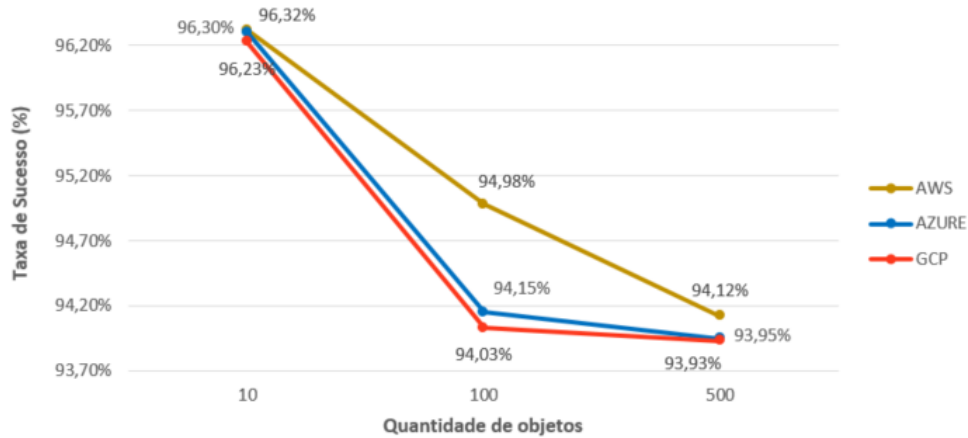


Figura 4.12: Taxa de sucesso para objetos de 64KB.

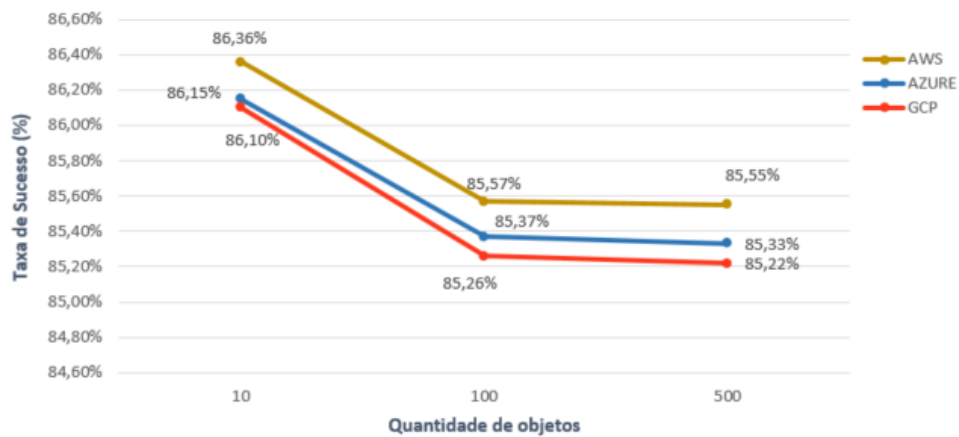


Figura 4.13: Taxa de sucesso para objetos de 1MB.

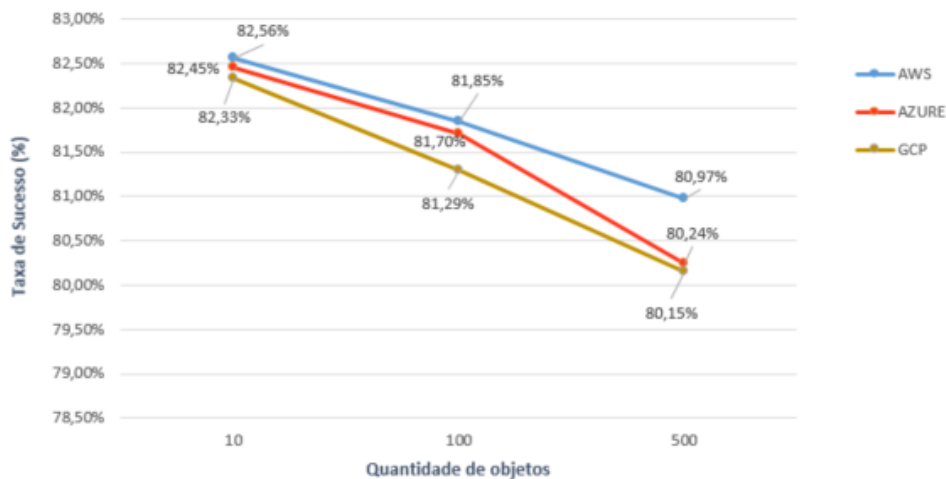


Figura 4.14: Taxa de sucesso para objetos de 16MB.

4.4 Considerações Finais

Neste capítulo foi apresentada a análise dos resultados do comparativo do armazenamento por objeto dos três maiores provedores de nuvem pública mencionados neste trabalho. Para isso, foi detalhada a ferramenta de *benchmark*, suas camadas e funcionalidades, bem como as métricas e o escopo de execução dos testes, com o objetivo de atender as necessidades das organização na implantação deste tipo de solução.

Os resultados dos experimentos comprovaram que a performance dos armazenamentos por objetos de cada provedor é semelhante, uma vez que o desempenho das métricas de taxa de transferência, tempo de resposta e taxa de sucesso tiveram um percentual de queda média similar.

Os testes também mostraram que as operações de arquivos pequenos, médios e grandes foram bem-sucedidos em geral com diferentes larguras de banda. Quanto maior o arquivo operado, maior será a largura de banda. No próximo capítulo serão apresentadas as conclusões sobre este trabalho, assim como indicações de trabalhos futuros.

Capítulo 5

Conclusão

Neste trabalho propôs-se uma arquitetura para aferir métricas de serviço de armazenamento por objetos em nuvens públicas, visando auxiliar futuros usuários a escolher entre os serviços ofertados pelos principais provedores, com o intuito de aumentar a eficiência em ambientes.

Dentro deste contexto, este trabalho fez uma proposta de uso de ferramenta de aferição de desempenho, para funcionar como referência durante a adoção destas tecnologias de armazenamento. A infraestrutura compreende uma máquina virtual que gera cargas de trabalho em repositórios dos provedores, enquanto coleta análises e exporta relatórios.

Apesar de ser possível observar algumas diferenças entre o desempenho de cada provedor analisado, os resultados observados no Capítulo 4 indicam algumas ressalvas. Caso os clientes utilizem objetos maiores, com uma estrutura de rede limitada, as métricas de taxa de transferência, tempo de resposta e taxa de sucesso, podem sofrer um impacto negativo. Porém, o presente trabalho mostra que embora as métricas apresentem piora, o comportamento entre as provedoras é muito similar.

Com isso, conclui-se que a ferramenta testada, embora eficiente, não deve ser o único recurso utilizado para embasar na decisão da escolha entre os três principais fornecedores de nuvens públicas. Visto que, como resultado do comparativo não obteve uma diferença de performance tão significativa, existem outros pontos que podem influenciar na decisão de escolha, como por exemplo o valor dos serviços.

Para trabalhos futuros, sugere-se aprofundar a pesquisa sobre os tipos básicos de serviços que os provedores ofertam, estabelecendo então um conjunto de métricas relacionadas ao desempenho destes serviços em nuvem e ferramentas criadas para medi-los, com foco em diferentes tipos de público-alvo, como por exemplo, usuários finais, empresas de grande porte, órgãos governamentais e outros.

Também é apontado como um possível trabalho futuro, que sejam propostas novas soluções para medir os desempenhos dos serviços básicos de maneira unificada, com o

intuito de diminuir os estágios de avaliação necessários para implementar ambientes computacionais complexos em nuvem.

Referências

- [1] Buyya, R., C. S. Yeo, S. Venugopal, J. Broberg e I. Brandić: *Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility*. *Future Gener. Comput. Syst.*, 25:599–616, 2009. 4
- [2] Vaquero, Luis M., Luis Rodero-Merino, Juan Caceres e Maik Lindner: *A Break in the Clouds: Towards a Cloud Definition*. *SIGCOMM Comput. Commun. Rev.*, 39(1):50–55, dezembro 2009, ISSN 0146-4833. <https://doi.org/10.1145/1496091.1496100>. 4
- [3] Wu, Linlin e Rajkumar Buyya: *Service level agreement (SLA) in utility computing systems*. Em *Performance and dependability in service computing: Concepts, techniques and research directions*, páginas 1–25. IGI Global, 2012. 4
- [4] Patel, Pankesh, Ajith H Ranabahu e Amit P Sheth: *Service level agreement in cloud computing*. 2009. 4
- [5] Mell, Peter e Timothy Grance: *The NIST Definition of Cloud Computing*. Relatório Técnico 800-145, National Institute of Standards and Technology (NIST), Gaithersburg, MD, September 2011. <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>. 4, 5, 6, 7, 8, 9
- [6] Li, Hui, David Groep e Lex Wolters: *Workload characteristics of a multi-cluster supercomputer*. Em *Workshop on Job Scheduling Strategies for Parallel Processing*, páginas 176–193. Springer, 2004. 5
- [7] Iosup, Alexandru, Catalin Dumitrescu, Dick Epema, Hui Li e Lex Wolters: *How are Real Grids Used? The Analysis of Four Grid Traces and Its Implications*. Em *2006 7th IEEE/ACM International Conference on Grid Computing*, páginas 262–269, 2006. 5
- [8] Stavinoha, Ken E: *What is Cloud Computing and Why Do We Need It*. Retrieved September, 18:2015, 2010. 5
- [9] Srinivas, J, K Venkata Subba Reddy e A MOIZ Qyser: *Cloud computing basics*. *International journal of advanced research in computer and communication engineering*, 1(5):343–347, 2012. 5
- [10] Dillon, Tharam, Chen Wu e Elizabeth Chang: *Cloud computing: issues and challenges*. Em *2010 24th IEEE international conference on advanced information networking and applications*, páginas 27–33. Ieee, 2010. 5

- [11] Diaby, Tinankoria e Babak Bashari Rad: *Cloud computing: a review of the concepts and deployment models*. International Journal of Information Technology and Computer Science, 9(6):50–58, 2017. 5, 8
- [12] Tsai, Wei Tek e Xin Sun: *SaaS multi-tenant application customization*. Em *2013 IEEE seventh international symposium on service-oriented system engineering*, páginas 1–12. IEEE, 2013. 6
- [13] Jadeja, Yashpalsinh e Kirit Modi: *Cloud computing-concepts, architecture and challenges*. Em *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*, páginas 877–880. IEEE, 2012. 6, 9
- [14] Junath, N e G Shanmugarathinam: *A Cloud service and conceptual modeling of IAAS*. 2012. 7
- [15] Duan, Yucong, Yuan Cao e Xiaobing Sun: *Various 'aaS' of everything as a service*. agosto 2015. 7
- [16] Duan, Y., G. Fu, N. Zhou, X. Sun, N. C. Narendra e B. Hu: *Everything as a Service (XaaS) on the Cloud: Origins, Current and Future Trends*. Em *2015 IEEE 8th International Conference on Cloud Computing*, páginas 621–628, 2015. 7
- [17] Simmon, Eric: *Evaluation of cloud computing services based on NIST SP 800-145*. NIST Special Publication, 500:322, 2018. 8
- [18] *O que são provedores de nuvem?* <https://www.redhat.com/pt-br/topics/cloud-computing/what-are-cloud-providers>, acesso em 2021-07-17. 8
- [19] Behl, Akhil e Kanika Behl: *An analysis of cloud computing security issues*. Em *2012 world congress on information and communication technologies*, páginas 109–114. IEEE, 2012. 8
- [20] Furht, Borko e Armando Escalante: *Handbook of Cloud Computing*. Springer Publishing Company, Incorporated, 1st edição, 2010, ISBN 1441965238. 8
- [21] Marinos, Alexandros e Gerard Briscoe: *Community cloud computing*. Em *IEEE international conference on cloud computing*, páginas 472–484. Springer, 2009. 9
- [22] Noushad, Noufia: *A New Review on Cloud Computing Systems*. 2019. 9
- [23] Esteves, Rui: *A taxonomic analysis of cloud computing*. Em *1st Doctoral Workshop in Complexity Sciences ISCTE-IUL/FCUL*, 2011. 9
- [24] *O que são provedores de nuvem?*, 2021. <https://www.redhat.com/pt-br/topics/cloud-computing/what-are-cloud-provider>, acesso em 2021-03-30. 10
- [25] *Google Cloud Platform*, 2021. <https://cloud.google.com/>, acesso em 2021-03-26. 10
- [26] *Amazon Web Services*, 2021. <https://aws.amazon.com/pt/>, acesso em 2021-03-26. 10

- [27] *Microsoft Azure*, 2021. <https://azure.microsoft.com/pt-br/>, acesso em 2021-03-26. 10
- [28] *Cloud Market Ends 2020 on a High while Microsoft Continues to Gain Ground on Amazon*, 2021. <https://www.srgresearch.com/articles/cloud-market-ends-2020-high-while-microsoft-continues-gain-ground-amazon>, acesso em 2021-03-14. 10, 11
- [29] *Positioning technology players within a specific market*. <https://www.gartner.com/en/research/methodologies/magic-quadrants-research>, acesso em 2021-02-14. 10
- [30] *Infraestrutura global*, 2021. <https://aws.amazon.com/pt/about-aws/global-infrastructure/>, acesso em 2021-03-28. 11, 12
- [31] *Regiões e zonas de disponibilidade*, 2021. https://aws.amazon.com/pt/about-aws/global-infrastructure/regions_az/, acesso em 2021-03-28. 11
- [32] *Produtos em nuvem*, 2021. https://aws.amazon.com/pt/products/?nc2=h_q1_prod_fs_f, acesso em 2021-09-19. 11
- [33] Baldini, Ioana, Paul Castro, Kerry Chang, Perry Cheng, Stephen Fink, Vatche Ishakian, Nick Mitchell, Vinod Muthusamy, Rodric Rabbah, Aleksander Slominski *et al.*: *Serverless computing: Current trends and open problems*. Em *Research advances in cloud computing*, páginas 1–20. Springer, 2017. 12
- [34] *Locais do Cloud*, 2021. <https://cloud.google.com/about/locations?hl=pt-br>, acesso em 2021-03-28. 13, 14
- [35] *Google Cloud Platform Services Summary*, 2021. <https://cloud.google.com/terms/services>, acesso em 2021-04-02. 13
- [36] *Geografias do Azure*, 2021. <https://azure.microsoft.com/pt-br/global-infrastructure/geographies/>, acesso em 2021-04-02. 14, 15
- [37] *Produtos do Azure*, 2021. <https://azure.microsoft.com/pt-br/services/>, acesso em 2021-09-19. 15
- [38] Zeng, Wenying, Yuelong Zhao, Kairi Ou e Wei Song: *Research on cloud storage architecture and key technologies*. Em *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, páginas 1044–1048, 2009. 17, 30
- [39] Voorsluys, William, James Broberg e Rajkumar Buyya: *Cloud Computing: Principles and Paradigms*, volume 8, páginas 1 – 41. janeiro 2011, ISBN 9780470940105. 17
- [40] Heuseler, Fábio Martins: *Uma abordagem multifacetada para exploração integrada de dados estruturados e não-estruturados em ambientes OLAP*. 2010. 17
- [41] *Cloud Market Ends 2020 on a High while Microsoft Continues to Gain Ground on Amazon*, 2021. <https://www.redhat.com/pt-br/topics/data-storage/file-block-object-storage>, acesso em 2021-04-03. 18

- [42] *What is Object Storage?*, 2021. <https://www.snia.org/education/what-is-object-storage>, acesso em 2021-10-01. 18
- [43] Kulkarni, G., R. Waghmare, R. Palwe, V. Waykule, H. Bankar e K. Koli: *Cloud storage architecture*. Em *2012 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, páginas 76–81, 2012. 18, 19
- [44] Dewan, Hrishikesh e RC Hansdah: *A survey of cloud storage facilities*. Em *2011 IEEE World Congress on Services*, páginas 224–231. IEEE, 2011. 18
- [45] *AWS EBS and S3: Object Storage Vs. Block Storage in the AWS Cloud*, 2021. <https://cloud.netapp.com/blog/block-storage-vs-object-storage-cloud>, acesso em 2021-03-29. 18
- [46] Wan, Jian, Xun Chen e Xindong You: *Design and Implementation of Multiple Volume Servers Block Level Storage System based on Load-Balanced Strategy*. *International Journal of Database Theory and Application*, 7(4):59–78, 2014. 19
- [47] Deng, Yuhui: *Deconstructing network attached storage systems*. *Journal of Network and Computer Applications*, 32(5):1064–1072, 2009. 19
- [48] *O que é o armazenamento de arquivos na nuvem?*, 2021. <https://aws.amazon.com/pt/what-is-cloud-file-storage/>, acesso em 2021-03-29. 19
- [49] *Categorias de armazenamento do Amazon S3*, 2021. <https://aws.amazon.com/pt/s3/storage-classes/>, acesso em 2021-04-10. 20, 21
- [50] *Cloud Storage*. <https://cloud.google.com/storage>, acesso em 2021-08-24. 21
- [51] *Locais de intervalos*, 2021. <https://cloud.google.com/storage/docs/locations>, acesso em 2021-04-06. 21
- [52] *Classes de armazenamento*, 2021. <https://cloud.google.com/storage/docs/storage-classes#key-concepts>, acesso em 2021-04-11. 22
- [53] *Armazenamento com redundância geográfica*, 2021. <https://docs.microsoft.com/pt-br/azure/storage/common/storage-redundancy#geo-redundant-storage>, acesso em 2021-10-01. 22
- [54] *Comparing block blob storage options*, 2021. <https://docs.microsoft.com/pt-br/azure/storage/blobs/storage-blob-storage-tiers>, acesso em 2021-04-10. 23
- [55] *Amazon Elastic Block Store*, 2021. <https://aws.amazon.com/pt/ebs/?ebs-whats-new.sort-by=item.additionalFields.postDateTime&ebs-whats-new.sort-order=desc>, acesso em 2021-04-18. 23
- [56] *Tipos de volumes do Amazon EBS*, 2021. <https://aws.amazon.com/pt/ebs/features/>, acesso em 2021-04-17. 24
- [57] *Persistent Disk*, 2021. <https://cloud.google.com/persistent-disk>, acesso em 2021-04-18. 24

- [58] *Storage options*, 2021. <https://cloud.google.com/compute/docs/disks>, acesso em 2021-04-17. 25
- [59] *Azure Disk Storage*, 2021. <https://azure.microsoft.com/en-us/services/storage/disks/>, acesso em 2021-04-18. 24
- [60] *Disk comparison*, 2021. <https://docs.microsoft.com/pt-br/azure/virtual-machines/disks-types>, acesso em 2021-04-17. 25
- [61] *Amazon Elastic File System*. <https://aws.amazon.com/pt/efs/>, acesso em 2021-08-17. 26
- [62] *Managing EFS storage classes*, 2021. <https://docs.aws.amazon.com/efs/latest/ug/storage-classes.html>, acesso em 2021-10-01. 26
- [63] *Cloud Filestore*. <https://cloud.google.com/filestore>, acesso em 2021-08-24. 26
- [64] *Arquivos do Azure*. <https://azure.microsoft.com/pt-br/services/storage/files/#overview>, acesso em 2021-08-22. 27
- [65] *O que é a Sincronização de Arquivos do Azure?* <https://docs.microsoft.com/pt-br/azure/storage/file-sync/file-sync-introduction>, acesso em 2021-08-22. 27
- [66] Zheng, Qing, Haopeng Chen, Yaguang Wang, Jiengang Duan e Zhiteng Huang: *COS-Bench: A Benchmark Tool for Cloud Object Storage Services*. Em *2012 IEEE Fifth International Conference on Cloud Computing*, páginas 998–999, 2012. 29
- [67] Foresti, Sara e Fadi El Moussa: *Report on Techniques for Selective and Secure Data Sharing*. 29
- [68] Hariyanto, Agus e Bekt Maryuni Susanto: *Analysis of object-based storage implementation on storage data center*. Em *Proceeding of the 1st International Conference on Food and Agriculture*, 2018. 32