

Análise de Sistemas de Armazenamento Distribuído para Nuvens Computacionais OpenStack

Renato Tanaka

Maurício A. Pillon



Sumário

- Introdução
 - Motivação
 - Problemática
- Objetivo
- Fundamentação Teórica
 - Trabalhos Relacionados
- Proposta
- Considerações Finais
- Cronograma



Introdução

- Nuvens Computacionais provisionam seus recursos de acordo com seu modelo de serviço implementado.
- De acordo com o NIST, modelos de serviço são IaaS, PaaS e SaaS. [1]
- *Cloud Operating Systems* (COS) administram recursos de nuvem e os virtualizam para maior elasticidade. [2]
- OpenStack é uma COS IaaS de código aberto.



Introdução - Motivação


- O desempenho de armazenamento persistente em nuvens possui grande relevância, pois impacta diretamente em aplicações de usuários finais.
- Com a expansão de áreas como Big Data, o desempenho de armazenamento em larga escala se torna de grande relevância. [3]
- Devido a sua larga escala, pequenas otimizações em arquiteturas de nuvem resultam em grandes benefícios.



Introdução - Problemática

- Armazenamento persistente de Máquinas Virtuais (MVs) é realizado por diferentes provedores deste serviço.
- Qual o contexto ótimo de cada provedor de armazenamento?
- Como determinar o melhor provedor, no quesito de desempenho?

Objetivo

- 
- Realizar análise qualitativa e quantitativa de provedores de serviço de armazenamento em bloco da plataforma OpenStack.
 - Realização de testes de estresse (*benchmarks*) com diferentes provedores de serviço, avaliando:
 - Escalabilidade
 - Contexto de Uso.
 - Desempenho.



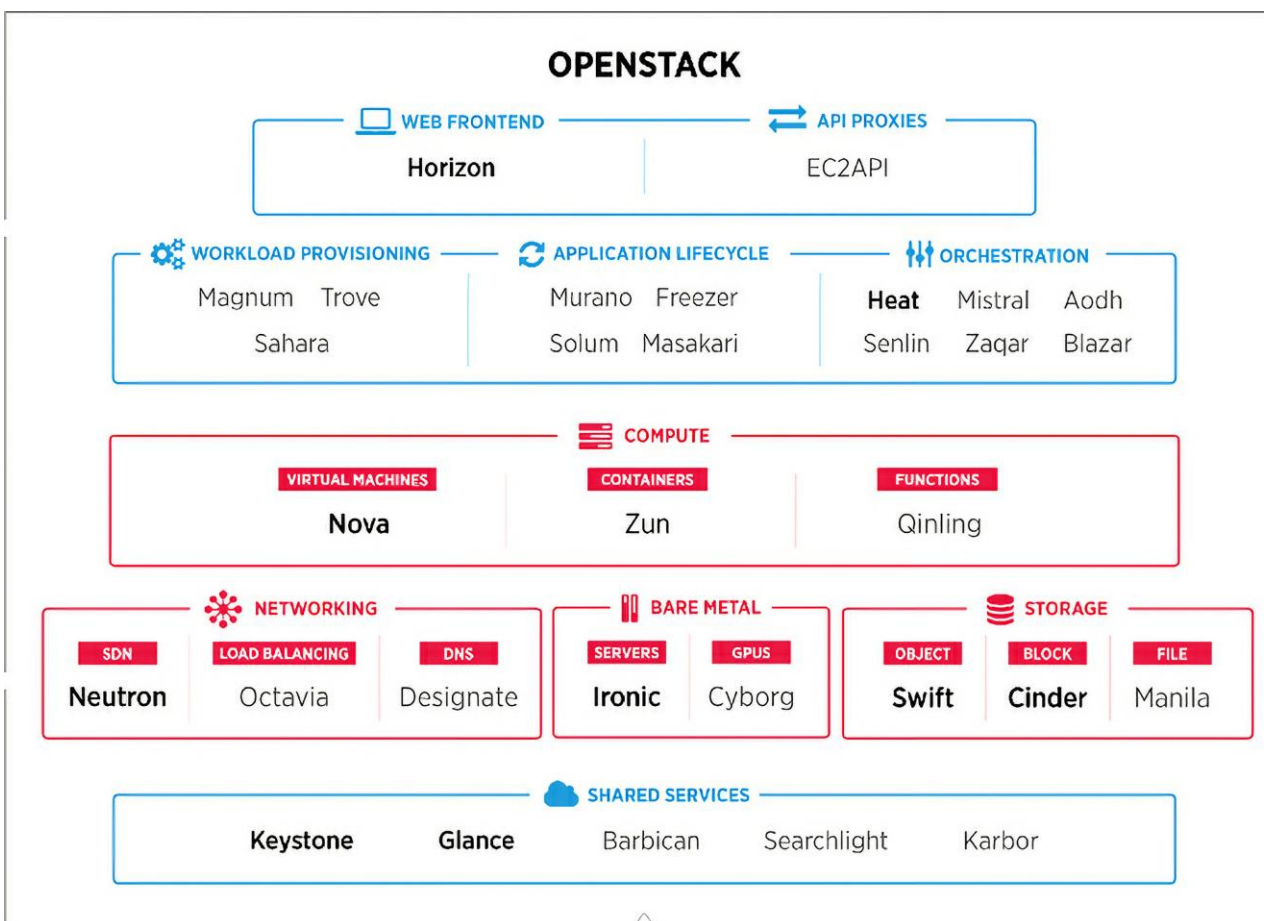
Fundamentação Teórica

- OpenStack possui característica modular.
 - Cada módulo administra um recurso ou funcionalidade da arquitetura de nuvem [5].
- Armazenamento na plataforma é realizado por três módulos distintos, cada um com uma representação diferente dos dados.
 - Bloco;
 - Arquivos;
 - Objetos;



Fundamentação Teórica

Figura 1: Representação dos diversos serviços do OpenStack



Fonte: Retirado de [5]



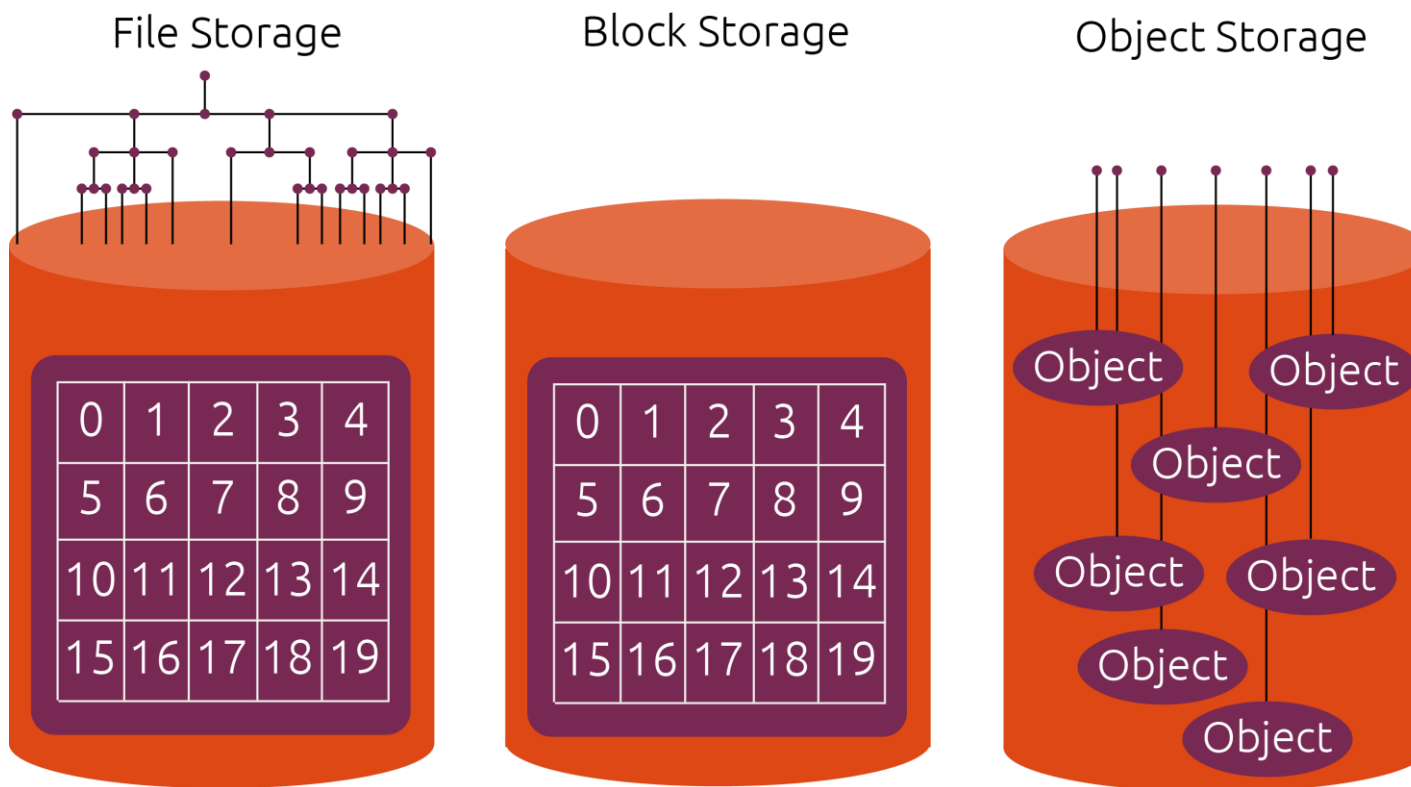
Fundamentação Teórica

- Cada tipo de armazenamento é gerenciado por um módulo diferente:
 - Bloco → Cinder
 - Objeto → Swift
 - Arquivo → Manila
- O módulo Cinder gerencia o armazenamento em bloco através da instanciação de **volumes de armazenamento**.
- Estes volumes realizam a comunicação entre soluções de provedores de armazenamento em bloco (*backends*) e o OpenStack.
- MVs de usuários finais interagem com volumes.



Fundamentação Teórica

Figura 2: Ilustração de tipos de armazenamento de dados.



Fonte: Retirado de [6]



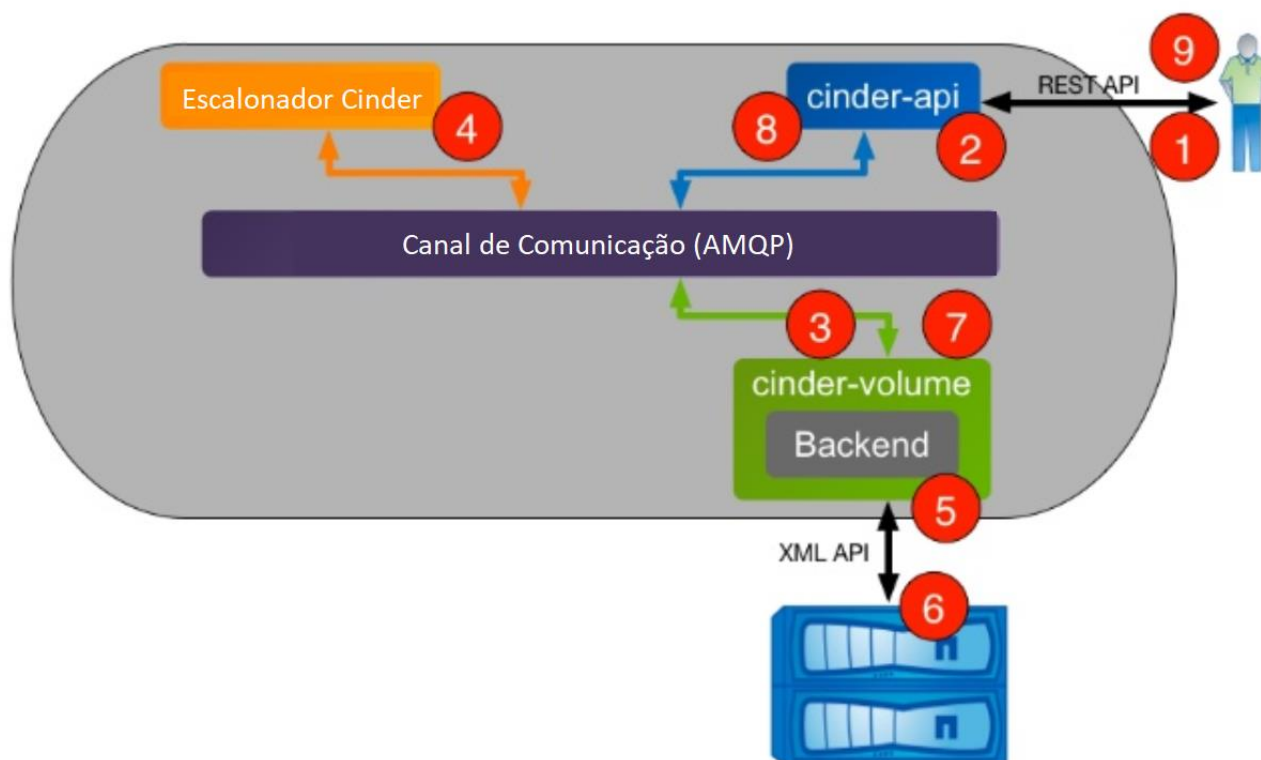
Fundamentação Teórica

- Cada tipo de armazenamento possui uma aplicação e contexto de uso.[5]
- Armazenamento em bloco é utilizado para prover armazenamento persistente em Máquinas Virtuais (MVs).
- Máquinas virtuais são gerenciadas pelo módulo Nova, e armazenamento em bloco é administrado pelo Cinder.



Fundamentação Teórica

Figura 3: Exemplo de requisição de criação de volume.

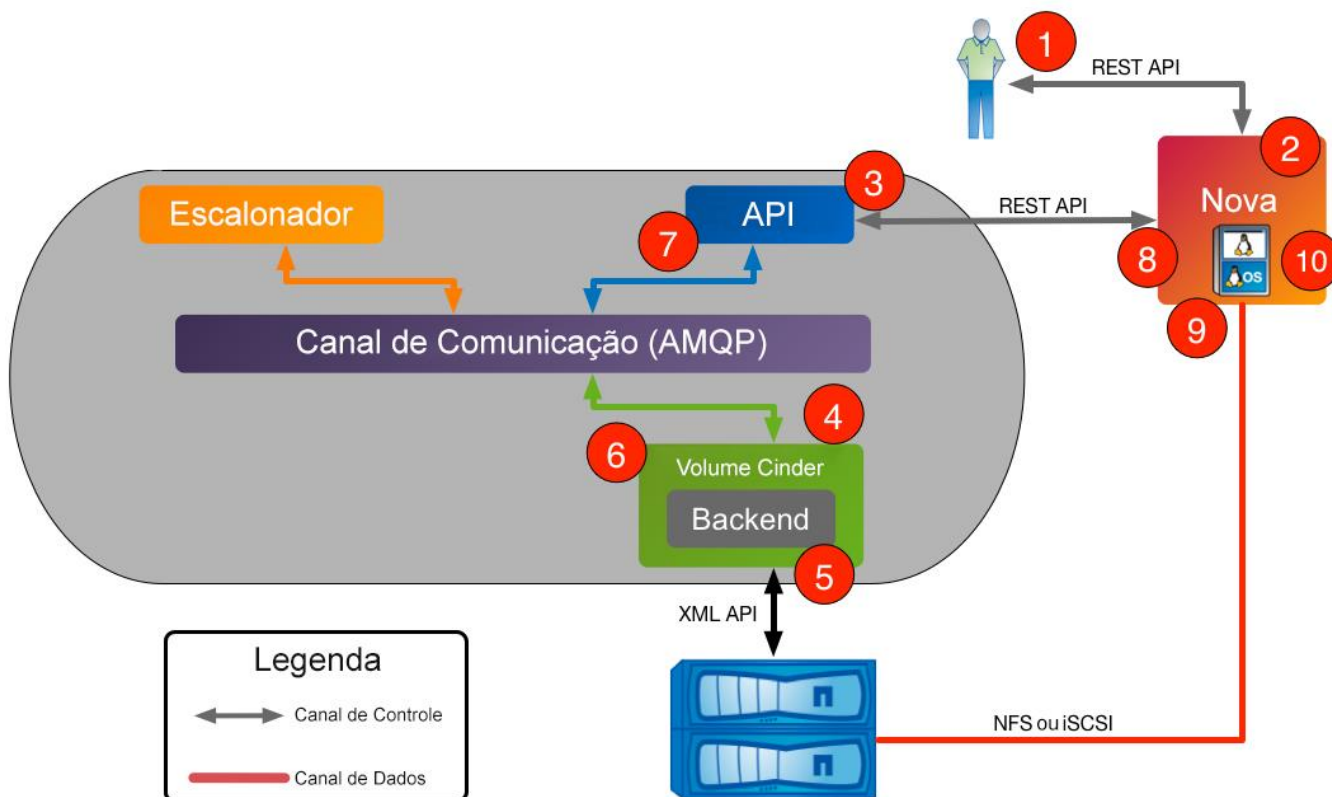


Fonte: Adaptado de [7]



Fundamentação Teórica

Figura 4: Ilustração de interação entre volume Cinder e MV no Nova



Fonte: Adaptado de [7]



Trabalhos Relacionados

Tabela 1: Trabalhos relacionados

Autores	Sistemas de Armazenamento	Ferramentas de <i>Benchmark</i>
Acquaviva, L. et al., 2018 [8]	HDFS, Ceph, GlusterFS, XtremeFS	Ganglia, Inotify
Gudu; Hardt; Streit, 2014 [9]	Ceph	Netcat, IPERF, DD, OSD Tell, RADOS Bench, fio
WANG et al., 2013 [10]	Ceph	RADOS bench
Zhang;Gaddam;Chronopoulos., 2015 [11]	Ceph	Bonnie ++, DD, Rados Bench, OSD Tell, IPERF and Netcat



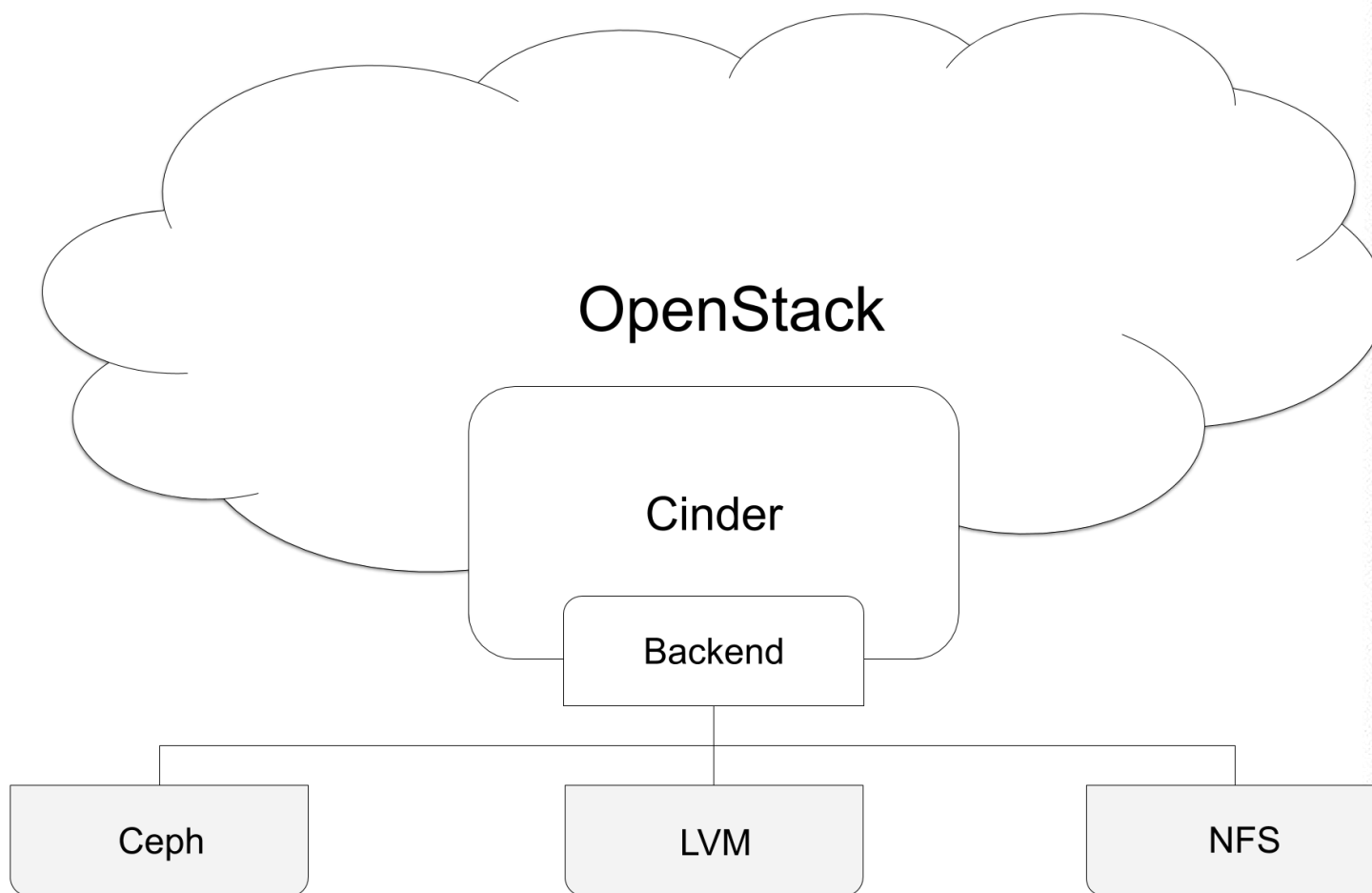
Proposta

- Realizar a aplicação de testes de desempenho (*benchmarks*) em diferentes *backends* do sistema de armazenamento em bloco Cinder.
- Após a aplicação dos testes, efetuar uma análise qualitativa e quantitativa das soluções de diferentes provedores do serviço de armazenamento em bloco.
- Determinar os contextos ótimos de aplicação às soluções analisadas, bem como definir a solução com melhor desempenho.



Proposta

Figura 5: Representação visual do contexto de aplicação do presente trabalho.



Fonte: Elaborado pelo Autor



Proposta

- Ferramentas a serem utilizadas:
 - Bonnie++;
 - Rally;
- Cenários:
 - Replicação tripla do armazenamento;
 - Qualidade de Serviço;
 - Expansão e diminuição da arquitetura;
 - Número de MVs instanciadas;



Proposta

Tabela 2: Descrição das métricas a serem avaliadas nos *benchmarks*.

Métrica	Unidade de Medida	Descrição
Tempo de Criação	Segundos	Tempo decorrido na operação de criação de volume.
Tempo de Anexação	Segundos	Tempo decorrido na operação de anexação de volume a uma MV.
Tempo de Destruição	Segundos	Tempo decorrido operação de destruição de volume.
Leitura	Operações / Segundos (IOPS)	Quantidade de operações por segundo durante leitura.
Escrita	Operações / Segundos (IOPS)	Quantidade de operações por segundo durante escrita.
Latência	Milissegundos	Tempo decorrido entre tempo de requisição ao <i>kernel</i> e conclusão da operação.

Considerações

- Desempenho de serviço de armazenamento em bloco no OpenStack possui grande importância em aplicações de usuários finais.
- Análise a ser realizada possui complexidade considerável, devido a fatores e parâmetros sensíveis que podem vir a comprometer a integridade dos testes.
- Após a aplicação dos testes, é esperado que seja possível determinar o melhor contexto para os *backends* analisados.

Cronograma

Etapas	2019					2020						
	Agosto	Setembro	Outubro	Novembro	Dezembro	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho
1												
2												
3												
4												
5												
6												
7												
8												

• 1 - Revisão bibliográfica, fundamentação teórica sobre nuvens, sistemas de armazenamento e OpenStack.

• 2 - Definição da proposta e ferramentas a serem utilizadas.

• 3 - Escrita do TCC - 1

• 4 - Instalação e configuração do ambiente de testes.

• 5 - Aplicação dos testes.

• 6 - Geração de gráficos baseados nos resultados dos testes.

• 7 - Realização da análise quantitativa e qualitativa.

• 8 - Escrita do TCC - 2.

Perguntas?



<renato.tanaka@edu.udesc.br>

<http://www.labp2d.joinville.udesc.br>



UDESC
UNIVERSIDADE
DO ESTADO DE
SANTA CATARINA



Departamento
de Ciência da
COMPUTAÇÃO



Referências

- [1] MELL, Peter et al. The NIST definition of cloud computing. 2011.
- [2] CHEN, Zuo-Ning et al. Evolution of cloud operating system: from technology to ecosystem. **Journal of Computer Science and Technology**, v. 32, n. 2, p. 224-241, 2017.
- [3] HASHEM, Ibrahim Abaker Targio et al. The rise of “big data” on cloud computing: Review and open research issues. **Information systems**, v. 47, p. 98-115, 2015.
- [4] OPENSTACK. What is OpenStack? 2019. Disponível em:<<https://www.openstack.org-/software/>>
- [5] OPENSTACK. Storage concepts. OpenStack, 2018. Disponível em:<https://docs.openstack-.org/arch-design/design-storage/design-storage-concepts.html>
- [6] CANONICAL. What are the different types of storage: block, object and file? Canonical, 2015. Disponível em: <https://ubuntu.com/blog/what-are-the-different-types-of-storage-block-object-and-file>

Referências

- [7] NETAPP.OpenStack Deployment and Operations Guide. NetApp, 2017. Disponível em:<https://netapp.github.io/openstack-deploy-ops-guide/ocata/content/ch_cinder.html>
- [8] ACQUAVIVA, Luca et al. Cloud Distributed File Systems: A Benchmark of HDFS, Ceph, GlusterFS, and XtremeFS. In: **2018 IEEE Global Communications Conference (GLOBECOM)**. IEEE, 2018. p. 1-6.
- [9] GUDU, Diana; HARDT, Marcus; STREIT, Achim. Evaluating the performance and scalability of the ceph distributed storage system. In: **2014 IEEE International Conference on Big Data (Big Data)**. IEEE, 2014. p. 177-182.
- [10] WANG, Feiyi et al. Performance and scalability evaluation of the ceph parallel file system. In: **Proceedings of the 8th Parallel Data Storage Workshop**. ACM, 2013. p. 14-19.
- [11] ZHANG, X.; GADDAM, S.; CHRONOPOULOS, A. T. Ceph distributed file system benchmarks on an openstack cloud. In: **2015 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)**. IEEE, 2015. p. 113-120.