Uma Proposta de Sistema de Arquivos Distribuídos em Nuvem Privada Multi-região para o IFSC

RESUMO EXPANDIDO - Disciplina de TCC290009

Matuzalem Müller dos Santos
Estudante do Curso de Engenharia de Telecomunicações

Ederson Torresini, Me
Professor orientador

Semestre 2018-1

Resumo- *No decorrer da última década a utilização de nuvens públicas por empresas aumentou consideravelmente. Entretanto, nem todos podem migrar ou oferecer serviços de computação em nuvens públicas, como é o caso atual de diversos setores de órgãos governamentais brasileiros. Neste contexto, é preciso projetar sistemas que operem em nuvens privadas e ofereçam serviços com boa disponibilidade e redundância. Para tal, técnicas de virtualização e sistemas distribuídos são empregados nestas nuvens, garantindo que o sistema possa atender às necessidades dos órgãos. O armazenamento de arquivos é um dos pontos mais críticos destes sistemas, visto que as informações armazenadas são extremamente sensíveis e falhas no sistema de armazenamento podem ser catastróficas para serviços governamentais e privacidade das pessoas. Este trabalho tem por objetivo estudar tecnologias de armazenamento distribuído multi-região utilizando containers e propor uma solução para a nova infraestrutura de serviços e servidores sendo planejada pela CTIC do IFSC câmpus São José.*

Palavras-chave: Nuvem Privada. *IaaS*. Armazenamento Distribuído.

1 Introdução

Segundo Mell e Grance (2011), computação em nuvem é um modelo que possibilita acesso através da rede à recursos (por exemplo, redes, servidores, armazenamento, aplicações e serviços) de forma universal, conveniente e sob demanda, que podem ser rapidamente fornecidos e escaláveis com baixo esforço e interação com o provedor do

serviço. De acordo com as necessidades e limitações do sistema, nuvens podem ser implantadas de forma privada, pública, híbrida e comunitária.

No decorrer da última década notou-se um grande aumento na utilização de serviços em nuvem. Segundo Kavis (2014), em 2013 a maior parte das *start-ups* e empresas de médio porte estavam desenvolvendo suas aplicações para prover serviços em nuvem com suas infraestruturas localizadas em nuvens públicas. A utilização de nuvens públicas têm se tornado cada vez mais comum por conta dos diversos benefícios ofertados. Dentre estes benefícios, estão o modelo de utilização *pay-as-you-go*, onde a empresa paga apenas pelos recursos que utiliza da nuvem, ao invés de comprar inúmeros equipamentos e ter que arcar com a manutenção destes equipamentos mesmo quando não utilizados ou ociosos; a facilidade de implantação de serviços; a contratação de um serviço especializado em *data centers* e infraestruturas que provavelmente oferecerá uma infraestrutura muito mais confiável e a prova de falhas do que a de uma nuvem privada da empresa em questão. Esta última vantagem também permite que desenvolvedores e funcionários de Tecnologia da Informação (TI) se concentrem no serviço ofertado pela empresa, e não em manter a infraestrutura (em nível de *hardware*) do sistema.

Por estes motivos, grandes empresas também estão migrando suas infraestruturas para nuvens públicas com o intuito de melhorar a eficiência de seus serviços e diminuir gastos de manutenção e expansão. Também de acordo com Kavis (2014), uma grande empresa que realizou a migração de seus serviços de uma nuvem privada para uma nuvem pública foi a *NETFLIX*. Em 2009, 100% do tráfego de usuários circulava dentro do próprio *data center* da empresa. Com o aumento do número de usuários do serviço de *streaming*, a empresa precisaria reestruturar sua arquitetura interna para adequar-se à elevada carga de tráfego gerada pelo *streaming* de mídia. Por este motivo, em 2008 a empresa iniciou a migração de sua infraestrutura para a *Amazon Web Services (AWS)*, que seria capaz de prover uma infraestrutura mais confiável e escalável. Segundo Izrailevsky (2016), em janeiro de 2016 esta migração foi finalizada, e atualmente toda a infraestrutura da *NETFLIX* está localizada na nuvem pública da *Amazon*.

Apesar dos diversos benefícios ofertados por nuvens públicas, estas nem sempre são a melhor solução para armazenamento e processamento de dados. Órgãos governamentais possuem grandes quantidades de dados, mas ainda há barreiras na utilização de nuvens públicas por estas entidades. Com foco neste debate, foi publicado o "Manual de Boas práticas, orientações e vedações para contratação de Serviços de Computação em Nuvem"pelo Departamento de Infraestrutura e Serviços de Tecnologia da Informação (2016). Este manual veda a utilização de nuvens públicas em alguns cenários, como em casos em que os serviços de Tecnologia da Informação e Comunicação (TIC) possam comprometer a segurança nacional, por exemplo. Uma das preocupações na contratação de serviços em nuvem por órgãos governamentais expressas no manual reside no controle sobre os dados armazenados por terceiros, os quais devem residir exclusivamente em território nacional, incluindo replicação e cópias de segurança (*backups*), de modo que o contratante disponha de todas as garantias da

legislação brasileira enquanto tomador do serviço e responsável pela guarda das informações armazenadas em nuvem (Departamento de Infraestrutura e Serviços de Tecnologia da Informação, 2016). Portanto, é aconselhável que duas condições sejam satisfeitas na contratação de serviços em nuvens públicas por órgãos governamentais: a primeira é de que o serviço contratado atenda à todos os requisitos de geolocalização, privacidade, segurança e qualidade de serviço determinados pelo contratante; enquanto a segunda condição refere-se à qualidade e continuidade na gestão sobre o serviço contratado, ou seja, é necessário que seja realizado um planejamento a longo prazo de como o serviço em nuvem será integrado e mantido com o sistema atual. Este planejamento deve ser realizado e seguido de acordo com as necessidades dos órgãos, e não ser alterado por mudanças de gestão ou pessoal, o que dificultaria a implantação e integração do serviço em nuvem com o sistema. Já em relação ao fornecimento de serviços em nuvem por terceiros, algumas empresas têm buscado certificações e ofertado serviços específicos para adequar-se à demandas específicas. Um bom exemplo de como provedores de serviços em nuvem estão começando a prover serviços específicos para governos é a *Amazon GovCloud*, a qual foi projetada para oferecer serviços em nuvem para o governo americano (DIEZ; SILVA, 2013). No Brasil, a Rede Nacional de Ensino e Pesquisa (RNP) está trabalhando nos Centros de Dados Compartilhados (CDC) para fornecimento de serviços de computação em nuvem para instituições de ensino e pesquisa no Brasil (Rede Nacional de Pesquisa, 2014). Entretanto, este projeto ainda não atende à todas as instituições de ensino e pesquisa do país, o que torna a utilização de nuvens privadas comum dentro destas instituições e de outros órgãos governamentais.

Além de poder oferecer um maior nível de controle sobre os dados armazenados, nuvens privadas também possuem outras vantagens. Como os administradores dos *data centers* possuem total controle sobre os serviços executados e dados armazenados, é possível rodar serviços específicos em servidores específicos, que possuam o *hardware* dedicado ou que melhor se encaixem no perfil do serviço utilizado ou ofertado. Além disso, nuvens públicas geralmente realizam o gerenciamento de arquivos através de objetos, sendo serviços transparentes ao usuário. Já em nuvens privadas, o gerenciamento de arquivos é construído pelos próprios administradores da nuvem, podendo oferecer diversos tipos de armazenamento (em objetos, arquivos ou blocos), o que permite melhor otimização do sistema. Este controle de desempenho no acesso a dados em nuvens privadas também favorece a utilização de nuvens híbridas, onde parte da infraestrutura do sistema está localizada em uma nuvem privada e outra parte localizada em uma nuvem pública. O armazenamento de dados pode, portanto, ser realizado em uma rede privada otimizada, e servido a aplicações que operem em uma nuvem pública, por exemplo.

Em todos os cenários exemplificados acima, a virtualização pode ser utilizada para melhorar a eficiência do sistema. Seja em nível de sistema operacional, com utilização de máquinas virtuais para aumentar o grau de portabilidade, flexibilidade e segurança do sistema (TANENBAUM; STEEN, 1994), ou de rede, com implementação de técnicas de tunelamento e *Virtual Local Area Networks* (VLANs) para melhor ge-

renciamento da rede (GöRANSSON; BLACK, 2014), o conceito de virtualização está presente na maior parte dos *data centers*. Há também uma crescente utilização de *containers* como substitutos a máquinas virtuais, devido à seu baixo *overhead* na implantação de serviços (MEDEL et al., 2016).
Visando a melhoria do sistema atual e evolução de serviços ofertados à comunidade acadêmica, a Coordenadoria de Tecnologia da Informação e Comunicação (CTIC) do câmpus São José do Instituto Federal de Santa Catarina (IFSC) têm estudado técnicas e implementações de computação em nuvem privada dentro do câmpus. A CTIC propõe a integração entre infraestruturas dos câmpus do IFSC, possibilitando assim uma maior oferta de serviços para toda a comunidade acadêmica. A integração entre infraestruturas também possibilitaria que serviços que demandam mais capacidade computacional fossem implantados, haja vista que um número maior de nodos (servidores) estaria disponível para realizar o processamento destes dados e serviços. Este estudo é complexo e possui diversos desafios, tanto locais (dentro dos próprios câmpus) quanto regionais (na rede do IFSC). Um destes desafios é o armazenamento distribuído de arquivos, foco deste trabalho. A seguir serão apresentadas a motivação, justificativa e objetivos do trabalho, seguidos da fundamentação teórica, estado da arte, propostas de sistemas de arquivos, implementações dos sistemas e conclusões do trabalho.

2 Metodologia

A metodologia do trabalho consiste em 4 etapas que permitirão obter um melhor conhecimento sobre a infraestrutura do câmpus, tecnologias a serem utilizas e implantação e testes dos sistemas de arquivos estudados.

2.1 Estudo da Infraestrutura Atual do IFSC

Será realizado um estudo para entender a infraestrutura atual do câmpus São José do IFSC e quais os serviços ofertados à comunidade acadêmica e interna. Também será necessário estudar e melhor conhecer as ferramentas a serem utilizadas (*Kubernetes, Ceph, GlusterFS e Storage OS*), possibilitando assim que o sistema de arquivos proposto seja compatível e eficiente com a infraestrutura e serviços propostos pela CTIC do câmpus.

2.2 Levantamento de Pré-Requisitos para o Sistema de Arquivos Distribuído

Na etapa de levantamento de requisitos será realizada uma consulta à CTIC do câmpus São José do IFSC para que sejam definidos pré-requisitos e métricas de desempenho para o sistema de arquivos distribuídos proposto. Dentre estas métricas, podem ser destacadas a capacidade de criar *backups* automatizados e velocidade de acesso aos arquivos.

2.3 Implantação do Sistema

Após o levantamento de requisitos será realizada a implantação dos sistemas de arquivos *Ceph, GlusterFS e Storage OS*. Estes sistemas de arquivos poderão ser descartados e novas alternativas consideradas caso eles não atendam aos pré-requisitos levantados juntamente com a CTIC do câmpus.

2.4 Testes dos Sistemas de Arquivos Distribuídos

Após a implantação dos sistemas de arquivos distribuídos serão realizados testes de desempenho e compatibilidade para garantir que o sistema de arquivos atenda às necessidades do câmpus. Estes testes irão consistir na utilização do sistema de arquivos para atender a um dos futuros serviços propostos pela CTIC do câmpus. No momento, planeja-se testar o desempenho dos sistemas de arquivos com o aplicativo *WordPress*, mas outros serviços poderão ser utilizados.

Referências

Departamento de Infraestrutura e Serviços de Tecnologia da Informação. *Boas práticas, orientações e vedações para contratação de Serviços de Computação em Nuvem*. [S.l.], 2016. Disponível em: <https://goo.gl/T2W8t6>. Acesso em: 12 de maio 2018.

DIEZ, O.; SILVA, A. Govcloud: Using cloud computing in public organizations. *IEEE Technology and Society Magazine*, v. 32, n. 1, p. 66–72, 2013. ISSN 0278-0097.

GöRANSSON, P.; BLACK, C. *Software Defined Networks*: A comprehensive approach. Waltham, Estados Unidos: Morgan Kaufmann, 2014.

IZRAILEVSKY, Y. *Completing the Netflix Cloud Migration*. 2016. Disponível em: <https://media.netflix.com/en/company-blog/completing-the-netflix-cloud-migration>. Acesso em: 12 de maio 2018.

KAVIS, M. J. *Architecting The Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS and Iaas)*. New Jersey, Estados Unidos: John Wiley & Sons, 2014.

MEDEL, V. et al. Modelling Performance & Resource Management in Kubernetes. In: *2016 IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC 2016)*. IEEE, 2016. v. 9. Disponível em: <https://books.google.com.br/books?id=gKnNAQAACAAJ>. Acesso em: 21 de maio 2018.

MELL, P.; GRANCE, T. *The NIST Definition of Cloud Computing*: Recommendations of the national institute of standards and technology. [S.l.], 2011. Disponível em: <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>.

Rede Nacional de Pesquisa. *Programa Centros de Dados Compartilhados é inaugurado em Recife (PE)*. 2014. Disponível em: <https://www.rnp.br/destaques/programa-centros-dados-compartilhados-e-inaugurado-recife-pe>. Acesso em: 26 de maio de 2018.

TANENBAUM, A. S.; STEEN, M. V. *Sistemas Distribuídos*: princípios e paradigmas. 2. ed. São Paulo, Brasil: Pearson Education do Brasil, 1994.