

Consolidação de máquinas virtuais com alta disponibilidade: uma revisão bibliográfica sistemática

Daniel S. Camargo, Maurício A. Pillon, Charles C. Miers, Marcelo da S. Hounsell

¹Programa de Pós-Graduação em Computação Aplicada (LabP2D/PPGCA – DCC)
Universidade do Estado de Santa Catarina (UDESC) – Joinville/SC – Brasil
daniel@colmeia.udesc.br
{marcelo.hounsell, mauricio.pillon, charles.miers}@udesc.br

Resumo. *A consolidação de máquinas virtuais é uma estratégia utilizada para reduzir o consumo de energia ao concentrar as máquinas virtuais no menor número de máquinas físicas. Porém, a consolidação ocasiona violações no Service Level Agreement, pois afeta a principalmente a disponibilidade de serviços devido as migrações de máquinas virtuais. Mecanismos de alta-disponibilidade possibilitam atingir níveis de disponibilidade maiores do que os estabelecidos em Service Level Agreement. Neste sentido, a presente revisão bibliográfica sistemática realiza um levantamento de trabalhos que relacionam a consolidação de máquinas virtuais com mecanismos de alta disponibilidade, que possam ser utilizados para solucionar o trade-off entre energia-SLA. Foram analisados um total de quarenta e um trabalhos, os quais foram reduzidos para quatro conforme os critérios de exclusão. Como resultados, a presente revisão relaciona os trabalhos encontrados com as suas características e aponta alguns desafios na área para o direcionamento de trabalhos futuros.*

1. Introdução

O paradigma de computação em nuvem evidencia-se por fornecer recursos computacionais sob demanda com garantias de disponibilidade [Nanduri et al. 2014]. Estas garantias são estabelecidas em *Service Level Agreement* (SLA), como um acordo de nível de serviço estabelecido entre o inquilino e o provedor. Violar um dos termos do SLA pode causar penalidades financeiras ao provedor [Beloglazov 2013]. Enquanto a disponibilidade refere-se a uma probabilidade de acesso aos recursos, a alta disponibilidade está relacionada com a tolerância à falhas, *i.e.*, usando técnicas de replicação e/ou redundância dos recursos relacionados.

Um ambiente de execução de nuvem (tipicamente Data Centers (DCs)) deve manter diversas máquinas físicas (MFs) ativas para garantir esta disponibilidade, que muitas vezes permanecem subutilizadas [Chowdhury et al. 2015], [Zahedi Fard et al. 2017]. Ou seja, mesmo com os benefícios da virtualização [Morabito 2015], há um desperdício de recursos e um elevado consumo de energia elétrica em toda a infraestrutura física do DC. Para reduzir este impacto do consumo de energia, uma das principais estratégias utilizadas na literatura é a consolidação de máquinas virtuais (MVs) [Mittal 2014], [Teng et al. 2017]. Esta estratégia visa alocar o maior número de MVs no menor número viável de MFs, possibilitando que as MFs ociosas sejam desativadas (normalmente em hibernação). A consolidação de MVs é uma aplicação prática do problema do empacotamento (*Bin Packing*), de complexidade NP-Difícil [Song et al. 2014]. Aplicar uma abordagem de

consolidação mais rígida, *i.e.*, o mais próximo da a configuração ótima, implica em diversas violações do SLA [Silva and Fonseca 2016]. Portanto, na consolidação de MVs há um dilema entre a redução¹ do consumo de energia e as violações do SLA, denominado de *trade-off energia-SLA*.

Sistemas de alta disponibilidade possuem componentes redundantes, capazes criar réplicas e alternar em caso de falha [Parziale et al. 2014]. No escopo de MVs, é necessário um conjunto de recursos compartilhados que cooperam para assegurar a execução de serviços críticos, os quais que devem ser explicitamente indicados pelo inquilino [Cully et al. 2008, Bin et al. 2011]. Em computação em nuvem, as MVs são replicadas com base em sistemas de arquivos distribuídos, permitindo uma sincronização mais eficiente e segura. Todavia, sistemas de alta disponibilidade necessitam usar mais recursos do que os comuns [Cully et al. 2008], o que resulta em um acréscimo no consumo de energia.

Observa-se um conflito entre as abordagens apresentadas, pois a consolidação de MVs viola a disponibilidade do SLA e um sistema de alta disponibilidade demanda consumo de energia, mas possibilita os requisitos do SLA. Existem diversas revisões de literatura que tratam especificamente de consolidação de MVs em ambientes virtualizados, assim como há trabalhos apenas em sistemas de alta disponibilidade, mas poucos abordam os dois temas simultaneamente. Deste modo, o presente artigo visa realizar uma revisão bibliográfica sistemática sobre os temas de consolidação de MVs e sistemas de alta disponibilidade no contexto de computação em nuvem. Como o argumento de pesquisa dos autores é de encontrar soluções que são efetivamente implementadas em ambientes físicos reais, descarta-se toda e qualquer pesquisa realizada em ambientes de simulação. Neste sentido, a organização do presente trabalho inicia com uma descrição dos trabalhos correlatos de revisão sistemática na Seção 2. A metodologia de pesquisa está sistematizada na Seção 3. A Seção 4 discute a análise dos trabalhos levantados com os critérios de inclusão e exclusão. Por fim, a Seção 5 conclui este artigo com considerações sobre o que foi realizado e direções futuras.

2. Trabalhos correlatos

Em sua grande maioria, os trabalhos correlatos realizam uma revisão sistemática de literatura individualmente nos temas de pesquisa indicados, *i.e.*, apenas em consolidação de MVs ou apenas em sistemas de alta disponibilidade. Portanto, o presente trabalho visa realizar o levantamento de artigos que utilizam em conjunto estas duas abordagens, relatando soluções para computação em nuvem que sejam energeticamente eficientes e seguros frente a disponibilidade. Com base nisso, não foi encontrada até o momento uma revisão sistemática de literatura que relacione os tema de consolidação de MVs com alta disponibilidade.

Há trabalhos específicos para cada tema, sendo organizados em trabalhos correlatos em consolidação e em alta disponibilidade. Destacam-se como trabalhos mais recentes para a consolidação de MVs, o Ahmad 2015 e Hameed et al. 2016; Como os mais recentes para alta disponibilidade, destacam-se Medina and García 2014 e

¹O termo eficiência energética é utilizado apenas quando há a efetiva redução do consumo de energia mantendo-se as mesmas propriedades de desempenho anteriormente observada.

2.1. Consolidação de MVs

Em Ahmad 2015, são revisados os trabalhos relacionados com a consolidação de MVs e a aplicação do mecanismo *Dynamic Voltage and Frequency Scaling* (DVFS) para otimizar o consumo de energia. Como recursos, há um grande foco em otimização da largura de banda de redes e de sistemas de armazenamento distribuído. Ao fim, é proposta uma taxonomia no tema para categorizar a literatura revisada. Os aspectos críticos dos esquemas de migração são investigados, bem como são mostradas questões em aberto neste tema de pesquisa.

No caso de Hameed et al. 2016, é feita uma revisão de literatura e proposta uma taxonomia para as técnicas de alocação de recursos energeticamente eficiente. A consolidação de servidores é citada como uma das principais abordagens para realizar a alocação de máquinas virtuais (que consiste no uso de memória das MFs). Entre os recursos analisados, destaca-se a rede com a largura de banda e tempo de resposta, o processador com a variação da frequência de processamento (DVFS) e o uso de memória de modo eficiente para maximizar o uso dos equipamentos computacionais. Por fim, como produto do trabalho são relacionadas as vantagens e desvantagens dos mecanismos existentes em uma análise abrangente sobre os temas encontrados na taxonomia. Esses temas são: políticas de adaptação de recursos função objetivo, métodos de alocação, operação e interoperabilidade.

2.2. Alta disponibilidade

Já para o tema de alta disponibilidade, o trabalho de Medina and García 2014, que possui um foco nos mecanismos de migração para replicação de MVs e prover alta disponibilidade. Como principais características levantadas, está o levantamento do tipo de hypervisor suportado em cada solução, bem como características mais específicas, como a presença de um processo de migração de memória, verificação em checkpoints e multibackup. Frente aos recursos analisados, está a verificação da migração do armazenamento, migração incremental (memória) e se há algum processo de automatização para isso.

O a revisão sistemática de Endo et al. 2016, é específica para alta disponibilidade em computação em nuvem. Neste sentido, são relacionadas diversas características, mostrando-se um trabalho de alta complexidade e completude. É proposta uma classificação em três camadas: *middleware*, serviços e tecnologias de base. Adicionalmente é feita uma discussão sobre as soluções comerciais existentes, mais focada nos *middlewares*.

3. Metodologia da pesquisa sistemática

O método utilizado para realizar a presente revisão sistemática consiste nos fatores relacionados com as pesquisas de Kitchenham 2004, Petersen et al. 2008 e Conforto et al. 2011. De acordo com estes autores, a metodologia de pesquisa de uma revisão sistemática deve conter as questões de pesquisa a serem respondidas, as palavras-chave, critérios objetivos e de inclusão. As estratégias de busca dos trabalhos mais significantes incluem a formação das *strings* de busca, a escolha das bases de busca acadêmicas e a aplicação de um conjunto de critérios para a seleção dos trabalhos existentes.

3.1. Fontes dos trabalhos existentes

Com o objetivo de encontrar trabalhos de fontes diversificadas, adota-se como critério de seleção dos mecanismos de busca o trabalho de Buchinger et al. 2014, que lista cinco as bases mais relevantes, conforme critérios levantados por Oppenheim et al. 2000. Neste sentido, adota-se os quatro mecanismos de buscas: ACM Digital Library, Engineering Village, IEEE Explore e Scopus, conforme relacionados na Tabela 1.

MB _#	Base de busca	Endereço eletrônico
MB ₁	ACM DL	http://dl.acm.org
MB ₂	IEEE Explore	http://ieeexplore.ieee.org
MB ₃	Engineering Village	http://www.engineeringvillage.com
MB ₄	Scopus	https://www.scopus.com

Tabela 1. Mecanismos de busca utilizados como fontes dos trabalhos existentes.

Estas bases também foram selecionadas pelo critério de acesso, sendo que as duas primeiras bases (MB₁ e MB₂) indexam artigos de conferências da própria organização, os demais (MB₃ e MB₄) indexam artigos de diversas organizações. Deste modo espera-se encontrar os resultados de MB₁ e MB₂ contido nas bases de MB₃ e MB₄, a serem corretamente comparados. Assim, como os autores possuem acesso às duas primeiras bases, serão descartados as duplicatas de MB₃ e MB₄.

3.2. Questões de pesquisa

As questões de pesquisa auxiliam na definição das palavras-chave, permitindo consolidar a estratégia de busca. Esta etapa do processo, também é utilizada para o refinamento da seleção dos trabalhos encontrados. Na Tabela 2 estão relacionadas cinco questões de

QP _#	Questão	Motivação
QP ₁	Quais as principais métricas utilizadas para analisar os resultados de uma consolidação?	Métricas relacionadas às violações do SLA, ao consumo de energia e de recursos computacionais.
QP ₂	Quais parâmetros devem ser analisados ao realizar o processo de consolidação?	A carga de recursos das MFs e das MVs, histórico de uso destes recursos por entidade e localização das réplicas.
QP ₃	Quais são as estratégias, políticas e algoritmos utilizados para realizar a consolidação e alta disponibilidade?	Para consolidação: migração e posicionamento. Para alta disponibilidade: replicação ativa ou passiva
QP ₄	Como prover um sistema de alta disponibilidade para MVs em computação em nuvem?	Redundância de recursos e auto-recuperação.
QP ₅	Quais são os principais recursos utilizados por uma solução de alta disponibilidade?	Exige-se o uso de todos os recursos básicos, como o armazenamento, processamento, memória e rede.

Tabela 2. Questões de pesquisa e motivações.

pesquisa, sendo as três primeiras relativas a consolidação de MVs e as três últimas relacionadas com alta disponibilidade. Embora sejam de simples formulação, elas definem o escopo necessário de cada um dos temas.

3.3. Estratégia de busca

Os estudos realizados no presente artigo tiveram início no dia primeiro de junho de 2017, sendo finalizado dia 31 de julho de 2017. Com o objetivo de encontrar as abordagens mais recentes, adota-se um período base de cinco anos de idade de sua publicação, considerando o ano de início da presente pesquisa, *i.e.*, de 01/01/2012 até 31/05/2017. Com base nas questões de pesquisa propostas na Tabela 2, e nos trabalhos correlatos discutidos na Seção 2, é selecionado um conjunto de palavras-chave que irão compor a *string* de buscas. Ao utilizar os operadores lógicos AND (\wedge), OR (\vee), NOT (\neg) e o uso de parêntesis para denotar precedência, forma-se a seguinte *string* de busca:

$$(\text{"Consolidation"} \wedge \text{"High Availability"} \wedge \text{"Availability"} \wedge \text{"Cloud Computing"} \wedge \text{"Virtual Machine"}) \wedge \neg (\text{Simulat*} \vee \text{numeric*}).$$

Devido a ocorrência de diversos pré-testes retornarem muitos trabalhos cujo ambiente de testes eram simulações com resultados numéricos, decidiu-se adicionar o operador NOT (\neg) nestas palavras-chave para excluir esta classe de trabalhos.

3.4. Procedimento de seleção dos trabalhos

A metodologia usada para selecionar ou excluir os trabalhos encontrados pela estratégia de busca ocorre com a definição de alguns critérios. Os resultados iniciais da busca normalmente retornam uma quantidade considerável de trabalhos, sendo necessário reduzir esse número para alcançar os trabalhos mais consistentes. Deste modo, a Tabela 3 mostra os critérios de inclusão e exclusão usados para este filtro.

CI _# : Critérios de Inclusão	CE _# : Critérios de Exclusão
CI ₁ : Implementa em infraestrutura real	CE ₁ : Estudo considera apenas simulação
CI ₂ : Considera simultaneamente consolidação de MVs e alta disponibilidade na pesquisa	CE ₂ : Considera apenas consolidação de MVs ou apenas alta disponibilidade
CI ₃ : Publicado com revisão aos pares	CE ₃ : Publicado em locais não verificáveis
CI ₄ : O trabalho deve estar acessível	CE ₃ : Publicação indisponível em um dos mecanismos de busca
CI ₅ : Os trabalhos devem ser únicos	CE ₅ : Trabalhos duplicados ou similares entre os MBs estão descartados, permanecendo apenas um

Tabela 3. Critérios de inclusão e exclusão dos artigos selecionados.

4. Análise dos trabalhos levantados

A presente Seção consiste na aplicação da sistemática apresentada, para posteriormente ser realizada a correlação entre os trabalhos selecionados. Mostrar a execução desta sistemática com detalhamento é importante para fornecer a reprodutibilidade da presente pesquisa, além de possibilitar alguma análise sobre os temas de pesquisa.

4.1. Resultados da busca e aplicação dos critérios de exclusão

Aplicar as palavras-chave nos quatro mecanismos de busca forneceu um total de quarenta e um trabalhos. Estas palavras foram aplicadas apenas nos títulos, *keywords* e resumo. Considera-se baixa a quantidade de artigos encontrados, quando comparada com a relativa maturidade dos dois temas de pesquisa. Os mecanismos de busca utilizados permitem adicionar um filtro de data de publicação, sendo este critério já validado durante esta fase inicial. Após a coleta e tabelamento dos dados, todos os artigos passaram por uma separação de unicidade (critério de exclusão CE₅), sendo removidos um total de doze trabalhos duplicados, encontrados em mais de um mecanismo de busca. Foi realizada uma leitura e avaliação dos vinte e nove trabalhos restantes, sendo vinte e cinco descartados por satisfazerem apenas aos critérios de exclusão, conforme Tabela 4.

	MB ₁	MB ₂	MB ₃	MB ₄	Total
Total encontrado	1	23	9	8	41
CE ₁	0	-15	-1	0	-16
CE ₂	0	-4	-3	-1	-8
CE ₃	0	0	0	0	0
CE ₄	0	0	-1	0	-1
CE ₅	-1	0	-4	-7	-12
Total disponível	0	4	0	0	4

Tabela 4. Resultados da aplicação dos critérios de exclusão.

Para verificar a consistência entre os temas de pesquisa (consolidação de MVs e alta disponibilidade) com os critérios de exclusão, foi realizada uma leitura do resumo, e quando necessário verificou-se o corpo destes trabalhos. Mesmo tendo aplicado na busca uma restrição de soluções que baseiam-se em simulação, houve um total de dezesseis trabalhos removidos por este motivo (CE₁). Oito trabalhos removidos por não considerarem as abordagens de consolidação e alta disponibilidade (CE₂) e apenas um trabalho que estava indisponível durante a pesquisa (CE₄). Por fim, restou um total de quatro trabalhos mais consistentes para realizar uma análise mais aprofundada no corpo dos artigos.

4.2. Análise dos trabalhos encontrados

Os quatro trabalhos consistentes Simonin et al. 2013, Jin et al. 2013, Wang et al. 2014 e Li et al. 2016 possuem uma avaliação com maior profundidade, considerando principalmente introdução, proposta e considerações finais. Nesta análise, são verificados os principais critérios, métricas e abordagens que permita correlacioná-los e responder as questões de pesquisa.

4.2.1. [Simonin et al. 2013]

O presente trabalho consiste em uma solução de código aberto para nuvens privadas, possui foco em escalabilidade, eficiência energética e tolerância a falhas. Permite aos usuários desenvolverem infraestruturas virtuais, e ao mesmo tempo controlar o ciclo de vida das suas MVs. A solução permite também a auto-cura (*self-healing*) em caso de

falhas nas MVs como um mecanismo de alta disponibilidade. A consolidação é feita através de um gerenciamento para a distribuição otimizada de MVs (como o mecanismo de posicionamento), utilizado para o gerenciamento de energia.

Esta mostra-se uma solução que idealmente une os critérios de alta disponibilidade com a consolidação de MV. Embora o objetivo do trabalho não seja o de demonstrar efetivamente a consolidação, sua solução utiliza a consolidação de MVs, com os mecanismos de migração e posicionamento, detectando quando há sobre/subutilização de recursos físicos. A arquitetura física dos componentes computacionais é similar a uma árvore *Big-Tree*, utilizada para providenciar a redundância de equipamentos. São utilizados os mecanismos de verificação da saúde das MVs (*health-check*), como *heart-beat* e *self-healing* e *autorecovery* para a alta disponibilidade. Como métricas utiliza-se o uso de tráfego de rede e o tempo de lançamento das MVs.

4.2.2. [Jin et al. 2013]

Neste trabalho, os autores possuem foco na otimização do sistema de armazenamento distribuído, que possibilita reduzir o tempo de migração de uma MV e reduzir a E/S de disco. Sua estratégia é otimizar o armazenamento pré alocando os volumes nos nós para aliviar a migração de MVs entre as MFs. Embora a solução não trate efetivamente de alta disponibilidade, uma solução de sistema de arquivos distribuídos é primordial para que, tanto na migração quanto na replicação de MVs, ocorram com mais efetividade, com menos erros e menos utilização de recursos. São abordadas três áreas de aplicação da solução proposta: em computação de alto desempenho (HPC), em armazenamento compartilhado para realizar a consolidação de MVs e balanceamento de carga e para manutenção e em casos de emergência, com a utilização de replicação e backups de dados. São utilizadas duas métricas para avaliar o método de migração em uma solução de armazenamento distribuído: tempo de migração e desempenho de E/S durante a migração. Estas métricas são relevantes tanto para a consolidação quanto para a alta disponibilidade por replicação. Porém esta solução mostrou uma sobrecarga no uso da rede, tendo em vista que o objetivo é paralelizar a migração de dados com o objetivo de minimizar o tempo de execução.

4.2.3. [Wang et al. 2014]

Os autores propõem uma solução de alta disponibilidade para soluções que executam por longos períodos de tempo, como é o caso do MapReduce. São avaliadas métricas de desempenho para estimar o custo da sua solução frente a um cenário anteriormente observado. Estas métricas consideram o uso dos recursos físicos como a memória RAM. A solução utiliza a migração de MVs para assegurar a alta disponibilidade (*snapshots*) e evitar uma potencial degradação de desempenho. Para isso é proposto um novo protocolo para o problema de posicionamento da MV, para que seja orientado à necessidade da aplicação. Os resultados mostram uma sobrecarga de 10% de memória ao utilizar a solução. As métricas estão mais relacionadas com a aplicação MapReduce, como o tempo de pronto dos *jobs* e a latência de interação entre os nós.

4.2.4. [Li et al. 2016]

Este trabalho propõe uma solução de alta disponibilidade baseada em uma heurística de uso de MVs com tempo compartilhado, baseada no problema do bandido multi-armado (teoria da probabilidade). Esta proposta objetiva aumentar a utilização dos recursos das MVs de backups (memória e armazenamento) enquanto fornece alta disponibilidade. Os experimentos mostram que a tradicional relação de 1:1 para o provisionamento de recursos pode estender-se para 1:M ($M_i 1$). Uma das principais métricas é o tempo de utilização de cada MV de backup e relações de utilização destas MVs de modo compartilhado.

4.3. Correlação dos trabalhos

Após uma análise no conteúdo dos trabalhos, a Tabela 5 relaciona as principais características encontradas. Estas características consistem nos tipos de recursos analisados, as métricas, quais as abordagens e mecanismos utilizados.

Características	Simonin et al. 2013	Jin et al. 2013	Wang et al. 2014	Li et al. 2016
Recursos	Rede e MVs	Disco e rede	Memória	Memória e armazenamento
Métricas	Tráfego e tempo	Taxa de E/S e tempo de migração	Tempo de pronto e latência	Tempo de utilização de provisionamento
Abordagens	Consolidação e alta disponibilidade	Sistema de arquivos distribuídos	Consolidação e alta disponibilidade	Racionalização de recursos e alta disponibilidade
Mecanismos	Posicionamento e <i>health-check</i>	Migração e sincronização	Migração, posicionamento e degradação	Compartilhamento de recursos

Tabela 5. Resumo e correlação dos trabalhos.

A seleção destas características foram baseadas nas questões de pesquisa, relacionadas na Tabela 2. Observa-se na Tabela 5 que as características dos trabalhos são relativamente heterogêneas, indicando que há uma diversidade de temas relacionados com a consolidação de MVs e alta disponibilidade.

5. Considerações finais

O presente trabalho mostra uma revisão bibliográfica sistemática com os temas de consolidação de MVs e sistemas de alta disponibilidade. É desenvolvida e aplicada uma metodologia que permitiu levantar uma grande quantidade de trabalhos e ao mesmo tempo realizar uma seleção qualitativa acerca do contexto das pesquisas. Esta seleção qualitativa reduziu a quantidade de quarenta e um trabalhos para quatro (cerca de dez por cento), que estavam fortemente alinhados com os critérios de exclusão, previamente definidos. Neste sentido, observa-se que a busca em quatro mecanismos de buscas com uma frase de busca composta por poucas palavras-chave relativamente simples retornou uma quantidade baixa de trabalhos relacionados com as temáticas definidas. Considera-se também

baixa a quantidade de artigos encontrados que efetivamente realizam uma implementação em um ambiente real, pois um dos critérios de exclusão é se o trabalho é uma simulação numérica.

Analisados os quatro trabalhos finalmente selecionados foram consolidados em uma tabela com quatro características. Os resultados desta tabela mostra características relativamente heterogêneas, indicando que há uma diversidade entre métricas, recursos e mecanismos nestas abordagens, relacionadas com a consolidação de MVs e alta disponibilidade. Futuramente, estima-se que os resultados da presente revisão bibliográfica sistemática seja utilizadas para o desenvolvimento de novas pesquisas. O objetivo é possibilitar a união das duas temáticas em uma solução que forneça alta disponibilidade com baixo consumo energético, permitindo eliminar o *trade-off* energia-SLA da consolidação de MVs.

Referências

- [Ahmad 2015] Ahmad, R. W. *et. al.* (2015). A survey on virtual machine migration and server consolidation frameworks for cloud data centers. *JNCA*, 52:11–25.
- [Beloglazov 2013] Beloglazov, A. (2013). Energy-efficient management of virtual machines in data centers for cloud computing. *PHD Thesis, University Of Melbourne, AU*. Phd thesis.
- [Bin et al. 2011] Bin, E., Biran, O., Boni, O., Hadad, E., Kolodner, E. K., Moatti, Y., and Lorenz, D. H. (2011). Guaranteeing High Availability Goals for Virtual Machine Placement. In *Proceedings of the 2011 31st International Conference on Distributed Computing Systems, ICDCS '11*, pages 700–709, Washington, DC, USA. IEEE Computer Society.
- [Buchinger et al. 2014] Buchinger, D., Cavalcanti, G. A. d. S., and Hounsell, M. d. S. (2014). Mecanismos de busca acadêmica: uma análise quantitativa. *Revista Brasileira de Computação Aplicada*, 6(1):108–120.
- [Chowdhury et al. 2015] Chowdhury, M. R., Mahmud, M. R., and Rahman, R. M. (2015). Implementation and performance analysis of various vm placement strategies in cloudsim. *Journal Of Cloud Computing*, 4(1):20.
- [Conforto et al. 2011] Conforto, E. C., Amaral, D. C., and Silva, S. L. (2011). Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos. In *8º Congresso Brasileiro de Gestão de Desenvolvimento de Produto-CBGDP*.
- [Cully et al. 2008] Cully, B., Lefebvre, G., Meyer, D., Feeley, M., Hutchinson, N., and Warfield, A. (2008). Remus: High availability via asynchronous virtual machine replication. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, pages 161–174. San Francisco.
- [Endo et al. 2016] Endo, P. T., Rodrigues, M., Gonçalves, G. E., Kelner, J., Sadok, D. H., and Curescu, C. (2016). High availability in clouds: systematic review and research challenges. *Journal of Cloud Computing*, 5:16.
- [Hameed et al. 2016] Hameed, A., Khoshkbarforousha, A., Ranjan, R., Jayaraman, P. P., Kolodziej, J., Balaji, P., Zeadally, S., Malluhi, Q. M., Tziritas, N., Vishnu, A., Khan,

- S. U., and Zomaya, A. (2016). A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. *Computing*, 98(7):751–774.
- [Jin et al. 2013] Jin, X., Wang, H., Wang, J., Cheng, S., and Li, J. (2013). A Partners Assisted Virtual Machine Live Storage Migration for Intensive Disk I/O Workloads. In *2013 IEEE 10th International Conference on High Performance Computing and Communications 2013 IEEE International Conference on Embedded and Ubiquitous Computing*, pages 1693–1698.
- [Kitchenham 2004] Kitchenham, B. (2004). *Kitchenham, 2004 Procedures for Performing Systematic Reviews*. Keele University Technical Report TR/SE-0401.
- [Li et al. 2016] Li, X., Qi, Y., Chen, P., and Zhang, X. (2016). Optimizing Backup Resources in the Cloud. In *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*, pages 790–797.
- [Medina and García 2014] Medina, V. and García, J. M. (2014). A Survey of Migration Mechanisms of Virtual Machines. *ACM Comput. Surv.*, 46(3):30:1–30:33.
- [Mittal 2014] Mittal, S. (2014). Power management techniques for data centers: A survey. *Corr.* Arxiv: 1404.6681.
- [Morabito 2015] Morabito, R. (2015). Power consumption of virtualization technologies: An empirical investigation. *IEEE/ACM UCC 2015 SD3C*.
- [Nanduri et al. 2014] Nanduri, R., Kakadia, D., and Varma, V. (2014). Energy and SLA aware VM Scheduling. *arXiv:1411.6114 [cs]*. arXiv: 1411.6114.
- [Oppenheim et al. 2000] Oppenheim, C., Morris, A., McKnight, C., and Lowley, S. (2000). The evaluation of WWW search engines. *Journal of Documentation*, 56(2):190–211.
- [Parziale et al. 2014] Parziale, L., Lasmayous, G., Pattabhiraman, M. S., Reed, K., Zhang, J. X., and Redbooks, I. B. M. (2014). *End-to-End High Availability Solution for System z from a Linux Perspective*. IBM Redbooks. Google-Books-ID: jSoiBQAAQBAJ.
- [Petersen et al. 2008] Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. In *EASE*, volume 8, pages 68–77.
- [Silva and Fonseca 2016] Silva, R. A. C. d. and Fonseca, N. L. S. d. (2016). Topology-Aware Virtual Machine Placement in Data Centers. *Journal of Grid Computing*, 14(1):75–90.
- [Simonin et al. 2013] Simonin, M., Feller, E., Orgerie, A. C., Jégou, Y., and Morin, C. (2013). An Autonomic and Scalable Management System for Private Clouds. In *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, pages 198–199.
- [Song et al. 2014] Song, W., Xiao, Z., Chen, Q., and Luo, H. (2014). Adaptive resource provisioning for the cloud using online bin packing. *IEEE Trans. Comput.*, 63(11):2647–2660.
- [Teng et al. 2017] Teng, F., Yu, L., Li, T., Deng, D., and Magoulès, F. (2017). Energy efficiency of VM consolidation in IaaS clouds. *The Journal of Supercomputing*, 73(2):782–809.

- [Wang et al. 2014] Wang, Y., Yang, R., Wo, T., Jiang, W., and Hu, C. (2014). Improving utilization through dynamic VM resource allocation in hybrid cloud environment. In *2014 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, pages 241–248.
- [Zahedi Fard et al. 2017] Zahedi Fard, S. Y., Ahmadi, M. R., and Adabi, S. (2017). A dynamic VM consolidation technique for QoS and energy consumption in cloud environment. *The Journal of Supercomputing*, pages 1–22.