

Um Modelo Integrado de Gestão de Recursos para as Nuvens Verdes

Jorge Werner¹ Guilherme Arthur Geronimo²
Carlos Becker Westphall³ Fernando Luis Koch⁴
Rafael Rodrigues de Freitas⁵

*Laboratório de Redes e Gerência
Universidade Federal de Santa Catarina
Florianópolis, Brasil*

Resumo

Green Clouds aim at providing optimized infrastructure that delivery flexibility, quality of services, and reduced energy utilization. To achieve this objective, they must management solutions regulate the internal settings to address the pressing issue of data centre over-provisioning related to the need to match the peak demand. In this context, we propose an integrated solution for resource management based on organisation models of autonomous agents components. These elements view the processing and environmental control elements as service providers (*e.g.* processing servers, load distribution services, temperature reduction services, etc) and coordinate these elements' operations to fulfil the system's overall objectives (*e.g.* SLA agreement, maximum energy utilization, contracted carbon emissions, etc). This work introduces the system management model, analyse the system's behaviour, describe the operation principles, and present an use case scenario.

Keywords: Please list keywords from your paper here, separated by commas.

1 Introdução

Computação Verde visa modelos e técnicas para a gestão integrada de dispositivos de computação e de controle ambiental para proporcionar qualidade de serviço, robustez e eficiência energética.

¹ Email: jorge@lrg.ufsc.br

² Email: arthur@lrg.ufsc.br

³ Email: westphall@lrg.ufsc.br

⁴ Email: koch@lrg.ufsc.br

⁵ Email: freitas@lrg.ufsc.br

O desafio em Nuvens Verdes é a operação com qualidade usando um montante mínimo de recursos de ativos, onde a "qualidade" pode ser definida, num termo de disponibilidade robustez em serviço. O problema pode ser resumido assim: os modelos de previsão de carga em ambientes de *Arquitetura Normal* e de *Computação em Nuvem* são baseados na análise de dados históricos e incrementa a procura dos modelos comerciais, com base nessas informações é possível avaliar a quantidade ideal de recursos a serem (pré-)alocados para uma carga de espera no tempo, no entanto, o que acontece quando um pico repentino de demanda atinge o sistema? Existem três abordagens para o problema:

- (i) deixar uma margem de recursos ativos *on line*;
- (ii) contratar um serviço externo (isto é, Nuvens Federadas) que pode ser comutada para o processo de carga acima da capacidade local, e;
- (iii) ativar recursos ociosos que procecem a carga excedente.

Cada uma dessas abordagens têm seus prós e contras. Por exemplo, (i) deixando uma margem implica a utilização excessiva de energia e custos de capital mais elevados. Na segunda alternativa (ii) implica deslocamento ao longo tempo, o requisito da disponibilidade e robustez dos canais de comunicação, e maiores custos operacionais. A terceira alternativa, (iii) de ligar recursos ociosos, implica num tempo longo de comutação, que geralmente é proibitivo para as operações de qualidade.

O desafio em Computação Verde é otimizar o equilíbrio entre estas diferentes abordagens para tratar a questão premente de centro de dados excesso de provisionamento relacionado com a necessidade de atender à demanda de pico, como descrito por Valancius *et al.* [10]. A solução gira em torno da gestão integrada dos recursos com base no controle sobre a alocação de recursos na operação de servidores (por exemplo, através de técnicas como Máquinas Virtuais), distribuição de carga preditiva, e; operação das unidades de suporte ambiental. Por exemplo, o sistema de gestão deve ser capaz de desligar os servidores não utilizados e reduzir proporcionalmente o apoio do ambiente, desativando as unidades de suporte ambiental, em resposta às características de carga do sistema previsto em um determinado período de tempo.

Nessa abordagem propomos uma solução para o controle integrado de computação e elementos do ambiente nas Nuvens Verdes. Ela se baseia em modelos de organização regulando o comportamento dos agentes, que vêem os elementos ambientais como prestadores de serviços (por exemplo, servidores de processamento, os serviços de distribuição de carga, os processadores de tarefas e os serviços de redução de temperatura). A integração dessas tecnologias com soluções práticas de gestão de rede e computação distribuída leva a um modelo ampliado de sistema de gestão. Por exemplo, os serviços de balanceamento de carga podem se comunicar com as unidades do sistema de

resfriamento sobre um aumento da demanda de processamento de tarefas que requerem conseqüentemente maior capacidade de servidor, o que implica em mais ar-condicionado que está sendo exigido.

Nossa pesquisa avança no estado da arte da seguinte forma:

- Apresenta um Modelo Integrado de Gestão de Recursos para Nuvens Verdes com base nos conceitos de modelos de organização, gerenciamento de redes e computação distribuída.
- Analisa o comportamento do sistema e princípios de funcionamento.
- Valida a proposta demonstrando o valor agregado do sistema em um cenário hipotético de estudo de caso.

Este trabalho está estruturado da seguinte forma. A seção 2 introduz um cenário motivador. Seção 3 descreve o projeto do sistema. Um cenário de caso de uso para validar a proposta é introduzida na seção 4. Concluímos o trabalho na seção 5.

2 Motivação

Um cenário prático na nossa Universidade motivou essa pesquisa. No passado (não tão distante), foi aplicada a *Arquitetura Normal* composta de diversos *clusters* de processamento configurado para processar diferentes serviços. Enfrentamos os problemas habituais dos grandes centro de dados na época: a falta de espaço em *rack* que a flexibilidade e escalabilidade impactados; um número excessivo de (geralmente desatualizado) servidores que impactaram os custos de operação, a necessidade de um sistema de refrigeração caro, e, uma ineficaz fonte de alimentação ininterrupta (UPS, do inglês, *Uninterruptible Power Supply*), que era escalavelmente problemático devido ao número de servidores envolvidos.

Com o advento da *Computação em Nuvem*, vários *racks* têm sido consolidados em poucos *racks* de servidores usando máquinas virtuais. Os resultados imediatos foram muito positivos: redução de utilização de espaço em *rack*, menor emissão de calor, com conseqüente otimização da infra-estrutura de refrigeração, e, uma solução rápida para o sistemas UPS problemáticos (ou seja, menos servidores significa a menos UPS necessários).

Nosso próximo passo é para **otimizar a utilização da energia e reduzir emissões de carbono** como parte de uma iniciativa institucional da sustentabilidade e preocupação ecológica. Procuramos soluções para *Nuvens Verdes* para alcançar este objetivo.

No entanto, como recursos consolidados usando técnicas de computação em nuvem, percebemos que havia espaço para melhorias para a esquerda. Por exemplo, houve períodos no tempo em que as *Máquinas Virtuais* (MV) estavam ociosas e os servidores foram subutilizados. Com base nos princípios

estabelecidos por Buyya *et al.* [1], o nosso objetivo era promover a gestão energética eficiente e buscar alternativas seguras para desligar os servidores não utilizados. A abordagem intuitiva era concentrar as aplicações em execução (MVs) em alguns servidores e reciclar a capacidade do servidor.

Apesar de atraente, essa abordagem levou a uma questão importante: indisponibilidade do serviço! Este problema obviamente reflete na taxa de erro de SLA. Uma rápida análise concluiu que era relacionado com o tempo necessário para ativar os recursos desligados. A explicação é simples:

- O dimensionamento é baseado no histórico de um ano, analisando a demanda dos serviços, isto é, com base na análise do dia anterior acrescido de uma margem a procura de uma estimativa de crescimento do negócio pode estimar a quantidade de recursos necessários para um serviço em um período em tempo.
- No entanto, quando se tratar de serviços com cargas de trabalho altamente variáveis, essa previsão se torna complexas e muitas vezes inatura. Além do mais, fatores externos podem levar a picos inesperados de demanda. Para isso, deixou uma margem de segurança dos recursos disponíveis (por exemplo, aceitar o erro de estimativa de 20%). No entanto, a demanda superou a limiar em situações específicas.
- Como uma solução, precisamos trazer recursos ociosos *on line*. O lapso de tempo entre a detecção da situação e do momento em que os recursos de processamento disponíveis causou a indisponibilidade do serviço.

Concluiu-se que temos três alternativas para superar esse problema:

- (i) aumentar a margem dos recursos alocados; esta abordagem diminui os benefícios da consolidação do ambiente de Computação em Nuvem, além disso, iria contra o princípio da utilização da energia e redução de emissões de carbono que buscamos em nossa instituição.
- (ii) contratar recursos externos sob demanda (isto é, Nuvens Federadas) que podem ser comutadas para o processo de carga superior; Embora esta seja a tendência na pesquisa de Computação em Nuvem, concluímos que não é o ideal, pois leva a custos operacionais adicionais relacionadas com ambos (a) contratação do provedor de recursos externos e (b) ampliação da rede de comunicação para assegurar um curto tempo de comutação , e;
- (iii) implementar uma solução integrada de gestão do sistema que liga os recursos ociosos de forma mais eficiente.

Optou-se pela terceira opção como analisamos uma dupla oportunidade, a saber: (i) para criar uma base de dados mais confiável de processamento do ambiente perfeitamente capaz de lidar com picos inesperados de demanda, e, (ii) para melhorar nossa utilização da energia e das emissões de pegadas de

carbono, integrando a gestão de controle ambiental na equação.

No que se segue, explicamos a solução proposta demonstrando como uma solução elaborada para gestão de Computação em Nuvem pode proporcionar a confiabilidade e a sustentabilidade.

Este problema pode ser superado quando se tratar de serviços que apresentam o mesmo padrão de demanda a cada dia. Pode-se prever quantas máquinas virtuais são necessários para processar o volume de trabalho. No entanto, quando se tratar de serviços com cargas de trabalho altamente variáveis, essa previsão se torna complexo, tornando o tempo de ativação de recursos em um fator chave. Pelo que pudemos experimentar algumas combinações na configuração da Nuvem.

A **Tabela 1** refere (1) a três combinações possíveis entre as MV e PM (S01, S02 e S03), com (2) o atraso médio de ativação (tempo em segundos), e (3) as chances dos serviços não ser um processo (de risco), apresenta também a energia consumida (baseado na obra de Assunção *et al.* [5]) de acordo com cada cenário. Definimos abaixo alguns escopos para avaliação e comparação

Situação	Estado MF	Estado MV	Tempo	Risco	Energia	Consumo
S01	Desligada	Desligada	30s	Alto	0 Ws	Nenhum
S02	Ligada	Desligada	10s	Médio	200 Ws	Médio
S03	Ligada	Ligada	0s	Nenhum	215 Ws	Alto

Tabela 1
Relação entre Situação & Risco & Atraso de Ativação& Consumo

entre as três diferentes estruturas citadas anteriormente:

Flexibilidade: Refere-se a alocação de recursos e serviços com a possibilidade de variação da configuração das máquinas que hospedam os serviços e dos serviços providos pela infraestrutura;

Disponibilidade: Refere-se ao comportamento da estrutura para com a distribuição da carga do serviços, assim como aos mecanismos da estrutura para garantir a operação dos serviços sem interrupções;

Custo: Refere-se aos gastos de capital (do inglês, CAPEX) para a existência da estrutura e gastos operacionais mensais (do inglês, OPEX) para que a estrutura persista;

Sustentabilidade: Refere-se aos esforços para redução de consumo e ao poder que a estrutura tem sob o ambiente á sua volta.

3 Proposta

Como visto na **Tabela 1**, o *grau de liberdade*, fator da Nuvem, se torna estreito, dados alguns elementos (MF e MV), apresentado pelas situações de controle. Propomos que a ruptura do Gerenciamento de Serviços Centralizados em que vários serviços de gerenciamento pequeno, dá-nos os elementos necessários para aumentar o *grau de liberdade* da Nuvem, criando possibilidade de alcançar uma situação de equilíbrio entre Risco e Consumo.

Mas com o gerenciamento de serviços de vários Nuvens introduzimos um novo problema: a gestão desses serviços se torna uma tarefa complexa. Para isso, usamos os princípios da Teoria das Organizações, para organizar e classificar tais serviços, tornando-os mais fáceis de controlar.

A gestão da Nuvem através dos princípios da Teoria da Organização dá a possibilidade de configuração automática do Sistema de Gerenciamento, uma vez que a adição de um novo elemento (por exemplo, MV, MF, UPS, Ar Condicionado) é apenas uma questão de incluir um novo serviço no Grupo de Gestão, retratado na **Figura 1**.

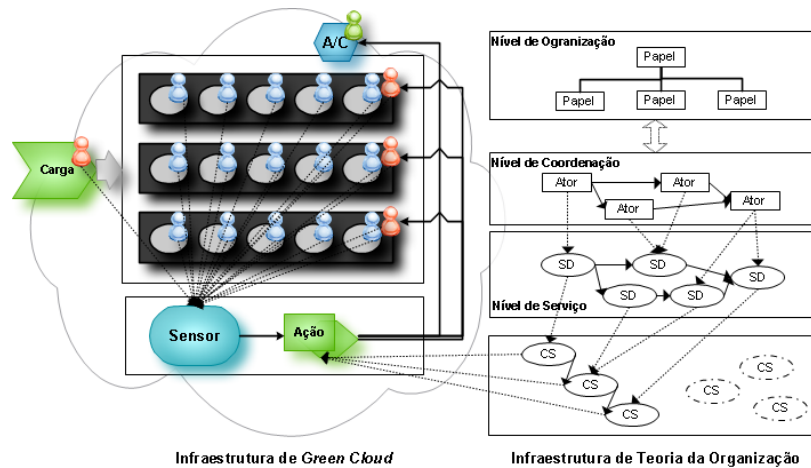


Figura 1. Gerenciamento de serviços distribuídos

Propomos um modelo de gerencia pro-ativa de Nuvens baseado na distribuição de responsabilidades por papéis, visualizado na Figura 2. Onde a responsabilidade da gerencia dos elementos da nuvem é distribuída entre vários agentes, cada qual em uma área de atuação. Estes controlarão individualmente os elementos da nuvem que lhe competem. Agindo de forma orquestrada visando o cumprimento de normas (metas). Tal orquestração baseando-se no fato (1) do conhecimento sobre o estado da nuvem (como um todo) ser compartilhado por todos, (2) de existirem Regras de Planejamento para guiar as ações dos agentes e (3) ao fato de existirem crenças, constantemente revistas, sobre o funcionamento interno da mesma.

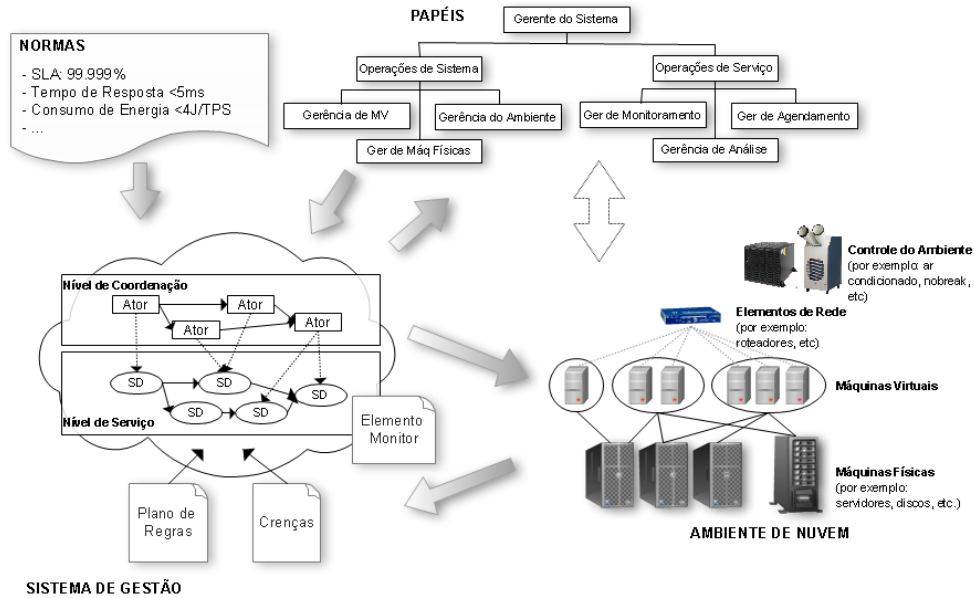


Figura 2. Modelo Proposto

Dado que toda a estrutura do centro de dados é dimensionada e utilizada para prover serviços, esta continua sendo apenas uma ferramenta para disponibilizar tais serviços. Geralmente acordos (SLA) são estabelecidos visando esclarecer a responsabilidades de cada parte (cliente/provedor). Frisamos que estes acordos devam manter-se ao seu nível (o de serviço), tornando-se puramente regras comportamentais (retardo, tolerância a falhas, etc) do serviço, excluindo exigências estruturais e físicas. Com a saída da configuração do ambiente de dentro do acordo, a nuvem pode torna-se flexível.

Com a liberdade de alteração de configuração da estrutura, a mesma pode-se tornar dinâmica e extensível. Permitindo abranger fatores externos aos acordos estabelecidos, porém ainda assim vitais a infraestrutura do centro de dados (por exemplo, consumo de energia, desgaste de hardware, entre outros).

Assim como nós vivemos sob as leis da física, a nuvem também deve existir sob normas bem definidas. Estas normas expressam (1) as regras de comportamentos estabelecidos no SLA e (2) os objetivos (de interesse) internos da nuvem que devem ser levados em conta.

Para que os vários elementos da nuvem trabalhem de forma eficiente, buscando a efetivação destas normas, eles devem ser coordenados por agentes externos aos serviços por eles auditados. Gerenciando, por exemplo, a ativação, desativação e configuração de máquinas virtuais e de máquinas físicas.

Visto a vasta gama de elementos a serem gerenciados a complexidade crescerá proporcionalmente com o tamanho da nuvem. Para evitar tal complexidade inferimos uma hierarquia aos agentes externos. Podemos fazer uma

analogia a uma empresa de grande porte, onde há uma hierarquia a ser seguida e responsabilidades sendo delegadas. Da mesma maneira como em uma empresa, deve existir um *Gerente do Sistema* (o chefe da empresa) que manda em todo o ambiente, e seguindo a hierarquia temos os Coordenadores que dividem as operações entre suas equipes [6], de maneira a facilitar a divisão de tarefas e responsabilidades entre suas equipes.

Dependendo da situação, as decisões gerarão *Operações de Sistema* ou *Operações de Serviço*, ou ambas. As *Operações de Sistema* podem ser divididas em *Gerência de Máquinas Virtuais*, *Gerência de Máquinas Físicas* e *Gerência do Ambiente*. As *Operações de Serviço* podem ser divididas em *Gerência de Monitoração*, *Gerência de Agendamento* e *Gerência de Análise*.

A ação de cada papel reflete diretamente na configuração da estrutura como um todo.

As *Operações de Sistema* agirão na estrutura a qual os serviços estão sendo processados e no ambiente que esta estrutura se localiza. As *Operações de Serviço* agirão na camada do serviços, adquirindo informações sobre ou alterando a configuração a qual o serviços esta sendo executado.

Os três papéis que as *Operações de Sistema* podem ser classificadas, são :

Gerência de Máquinas Virtuais: Responsável pela ações inferidas as máquinas virtuais. Tem papel de interface das máquinas virtuais ao modelo. Como por exemplo, criar ou destruir uma MV, alterar suas configurações e até move-la de um servidor à servidor (seja entre *clusters* locais ou remoto).

Gerência de Máquinas Físicas: Responsável pelas ações inferidas as máquinas físicas. Tem papel de interface das máquinas físicas ao modelo. Como por exemplo desligar e ligar as máquinas físicas, alterar configurações do sistema operacional hospedeiro (por exemplo, como BIOS - do inglês, *Basic Input/Output System*, SMART - do inglês, *Self-Monitoring, Analysis, and Reporting Technology*) e configurações de *hardware* (por exemplo, *cooler*, acelerômetro), assim como equipamentos de *backend* (como *Storages*, *Switchs* e *Site Backups*).

Gerência do Ambiente: Responsável pela ações fora da estrutura. Tem papel de interface do ambiente ao modelo. Como por exemplo, controle de temperatura do centro de dados, controle sob os sistemas de *backup* de energia (*NoBreak* e Geradores), controle sobre a acessibilidade do centro de dados.

Os 3 papéis que as *Operações de Serviço* podem ser classificadas, são:

Gerência de Monitoração: Responsável pela coleta de informações da estrutura, de modo geral, e sua compreensão. Tem papel de manter o modelo ciente do estado da nuvem. Monitora os servidores, MV, trafego de rede e etc. Baseado em parâmetros específicos previamente configurados pelo Gerente do Sistema, tal como (1) a utilização de um recurso e seu limiar (*threshold*) de notificação, (2) a disponibilidade de enlace de rede (dado binário) ou (3) a disponibilidade (*idleness*) de algum elemento da estrutura.

Gerência de Agendamento: Responsável pela agenda da nuvem. Tem um papel pró-ativo dentro do modelo, planejando as ações a serem tomadas diante dos eventos agendados. Como por exemplo, manutenções planejadas, gera um protocolo a ser seguido para a efetivação do mesmo.

Gerência de Análise: Responsável pelos testes comportamentais dos serviços e sua análise. Tem o papel de auditoria do serviço prestado pela estrutura e a compreensão do mesmo. Certifica-se se o serviço prestado esta de acordo com a normas a serem seguidas, inferindo limiares pré estabelecidos e alertando o Gerente do Sistema. Monitora a Qualidade de Serviço que é processado, e cruza com as variações da estrutura, tendo achar padrões entre a Performance obtida e os elementos (variantes/que variaram).

As Regras de Planejamento: são as bases de conhecimento teóricos as quais relacionam *Contextos* com os *Objetivos*. São utilizadas em momentos que decisões devem ser tomadas, durante o planejamento das ações. São conhecimentos primitivos inferidos pela experiência dos administradores. Como por exemplo, se a máquina física apresenta alta carga, para diminuir a carga, mova a MV com maior processamento para outra MF.

As Crenças: são um conjunto de conhecimentos empíricos, utilizados para aprimorar as decisões a serem tomadas. Nesta temos a compreensão empírica do funcionamento da Nuvem. Dos relacionamentos entre as variáveis dentro da estrutura. E o resultado de ações prévias. As crenças expressam a junção do conhecimento prático (das premissas), vindo das normas e o conhecimento empírico, originado dos históricos e experiências passadas. As crenças devem ser revistas freqüentemente por todos os elementos do modelo, assim como o compartilhamento destas revisões. Como por exemplo, a ativação de um servidor tipo X representa o aumento de Y graus em Z minutos.

4 Estudo de Caso

Nesta seção vamos mostrar três exemplos de casos de uso do nosso modelo e vamos comparar as estruturas descritas na **Seção 2**.

4.1 Cenário 1

Nos exemplos que faremos, consideraremos que nosso modelo foi implementado utilizando um Sistema de (Multi) Agentes. Apesar de poder ser implementado em diversos outros sistemas. Considere dois agente, um responsável (1) pela monitoração (aKa AMon) e outro (2) responsável pelo controle do ar condicionado (aKa AAC). Em um sistema reativo (como *Green Computing*), se o agente AMon detectasse uma aumento da carga no sistema, ele iria avisar o Controlador Central, o mesmo ativaria algumas MV para atender a demanda. Tempos depois, o agente AMon informaria ao Controlador Central

que a temperatura elevou-se acima do permitido, o mesmo acionaria o AAC e este aumentaria o sistema de Ar Condicionado. Em um sistema pro-ativo, quando o agente AMon detecta o aumento na demanda, o agente AAC já toma conhecimento do ocorrido e baseado em suas *crenças* chega a conclusão que estatisticamente (teoricamente ou empiricamente) a temperatura irá aumentar e age proativamente aumentando o ar condicionado.

4.2 Cenário 2

Neste caso, propor técnicas para detectar automaticamente a criação de centros de dados, encontrando máquinas ociosas e sobrecarregadas, a fim de determinar um ajuste no equilíbrio de carga, repartição e distribuição de recursos para proporcionar um trabalho mais eficiente.

Criamos um modelo do sistema usando o *Normas* (NM), *Crenças* (BL) e *Plano Regras* (PR), inferindo que precisaríamos (NM) para reduzir o consumo de energia, reduzir os custos de a nuvem e manter uma estrutura minimalista, com base em um mínimo (PR) de violações de SLA e redução de mudanças no ambiente, sem esquecer os parâmetros (BL) do tempo de provisionamento de máquinas virtuais. Com base nessas definições e responsabilidades, os sensores dos agentes atendem da melhor forma para equilibrar o meio ambiente. Vamos considerar três serviços (ou seja, *Web Service*, *backup*, de inicialização remota) rodando em simultâneo e cuja distribuição de carga parece ser complementares. Seus picos elevados (ou seja, variação carga de trabalho) acontecem em momentos diferentes. Com base em inferências a partir de NM, BL e PR os agentes vão monitorar o sistema e determinar as ações de forma dinâmica. Nesta proposta, os agentes têm duas soluções para a adequação dos servidores e máquinas virtuais: uma hora antes do pico, migrar a máquina virtual para um servidor mais robusto ou fora, dependendo do servidor redimensionar a máquina virtual, ajustando a sua utilização. Assim, o sistema funcionaria de forma mais dinâmica, autônomo, de acordo com os requisitos pré-definidos. Nosso ambiente é simplesmente todas as variações de carga de trabalho (de entrada), repartição e distribuição de serviços (movendo/relocando) para redução do uso de recursos (na saída do sistema), buscando a sustentabilidade ambiental.

4.3 Comparação

Nesta seção, consideramos as máquinas físicas como sendo servidores modelo Dell Power Edge T300, com memória RAM de 12GB, com processador de 4 x 2.8GHz (Xeon X3363) e sistema de refrigeração de 740 BTU/h. O servidores provem os serviços correio eletrônico, DNS, antivírus, antispam, servidor *web* e banco de dados.

Para fins de comparação, consideraremos os seguintes dados apresentados

na **Tabela 2** que foram retirados do centro de dados da universidade, no ambiente de produção.

	Uso de Memória	Carga Média em 5 minutos
Média	0,89 GB	0,09
Máxima	1,10 GB	0,56

Tabela 2
Uso de Memória e Carga

Carga	Consumo	Tempo por Distribuição
abaixo de 20%	0.13 kWh	74% do tempo
entre 20% e 50%	0.14 kWh	18% do tempo
acima de 50%	0.15 kWh	5% do tempo

Tabela 3
Consumo de Energia e Tempo por Distribuição de Carga

Baseados na métricas de Carga e Consumo de Memória, consideraremos que uma máquina física comporte entre 3 (em carga máxima) e 10 (em carga média) máquinas virtuais com configurações suficientes para atender o serviço em questão.

Com os dados de consumo energético listado na **Tabela 3** retirados do site do fabricante [4], calculamos o consumo anual necessário para cada estrutura prover 30 instâncias deste serviço. Note consideramos o conjunto dos serviços como um serviço só.

A monitoração dos serviços mostrou uma utilização média de memória e carga mostrada na **Figura 3**. Com isto, calculamos a distribuição da carga pelo tempo, apresentados na Tabela 3.

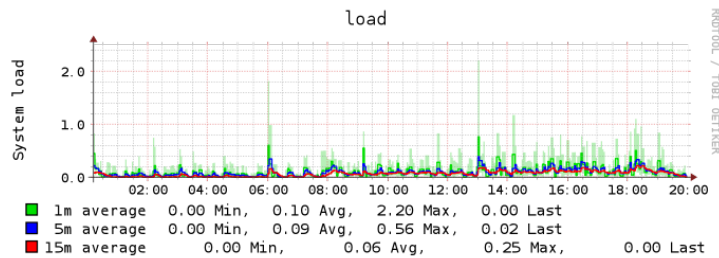


Figura 3. Carga Média do Sistema

Visto que cada estrutura demanda números diferentes de máquinas físicas ligadas, concluímos através da análise que a quantidade de máquinas necessárias para prover os serviços, num ambiente de Arquitetura Normal é de

30 MF em 100% do tempo, num ambiente de Computação em Nuvem esse número cai para 10 MF em 100% do tempo, porém num ambiente de computação em Nuvem Verde teríamos ligadas 3 MF em 10% do tempo, 6 MF em 40% do tempo e 10 MF em 50% do tempo.

Com isto, para a estrutura de Arquitetura Normal, obtemos um consumo aproximado de 37.7 MW por ano mostradas na **Tabela 4**. Para a estrutura de Computação em Nuvem, obtemos um consumo aproximado de 13 MW por ano, uma economia de aproximadamente 63% em comparação com a primeira estrutura mostradas na **Tabela 5**. Na estrutura de computação em Nuvem Verde, obtemos um consumo aproximado de 6.5 MW por ano, uma economia de aproximadamente 82% em comparação com a primeira estrutura.

Carga	Arquitetura Normal	Computação em Nuvem	Nuvem Verde
Baixa	17082 kW _y	6570 kW _y	1971 kW _y
Média	14718 kW _y	5256 kW _y	3154 kW _y
Alta	3942 kW _y	1314 kW _y	1314 kW _y

Tabela 4
Consumo anual usando Arquitetura Normal, Computação em Nuvem e Nuvem Verde

Estrutura	Consumo	Emissão de Calor
Arquitetura Normal	35742 kW _y	194472000 BTU/y
Computação em Nuvem	13140 kW _y	64824000 BTU/y
Nuvem Verde	6439 kW _y	64824000 BTU/y

Tabela 5
Comparativo anual Consumo X Estrutura X Emissão de Calor

Dado que o calor emitido por cada servidor é 740 BTU por hora aproximadamente, na **Tabela 5** mostramos a emissão anual de calor que o sistema de arrefecimento deve ser preparado de suportar, para cada estrutura.

Note que não há diferença de emissão de calor entre Computação em Nuvem e a Nuvem Verde. Visto que ambas estruturas não estendem sua gerência ao ambiente, no caso o sistema de refrigeração deve ficar sempre configurado para o pior caso possível, neste caso 10 (dez) máquinas físicas ligadas.

4.4 Gerenciamento Integrado em Nuvem Verde

Gerenciamento Integrado em Nuvem Verde, como chamamos, é uma estrutura hipotética que vemos como uma tendência desta área [7] é nosso objetivo de pesquisa. Estes aspectos que são descritos abaixo são a referência para que

o nosso modelo tem de cumprir. No comparação da Nuvem Verde, podemos inferir a responsabilidade de consumir menos energia, o termo garante os acordos pré-acordados no SLA, e apresenta os dados:

Na orquestração física é onde se tem conhecimento sobre o estado de todos os equipamentos de TI. Agindo quando for necessário, não quando é preciso. Planejando suas ações com base nas informações da nuvem. Sendo capaz de prever e executar as mudanças necessárias no *hardware* de acordo com a demanda da nuvem. Como por exemplo diminuído ciclos de relógio da CPU quando em superaquecimento, desligando máquinas de acordo com a carga prevista, ou ativando um *backup* remoto em caso de incêndio. O ambiente é capaz ainda de interagir automaticamente com nuvens públicas [2], migrando ou realocando recursos em tempo real, em nuvens remotas. Proporcionando um melhor suporte para os picos de carga de trabalho ou, ocasionalmente, os ataques DoS (do inglês, *Denial of Service*).

Lidando com o contexto de grupo automaticamente, é capaz de migrar esses grupos ou elementos de nuvens públicas. Abrange um novo nível de alta disponibilidade, sendo capaz de estendê-lo automaticamente a nuvens públicas. Permitindo a criação de nuvens espelho.

Com a comunicação dentro da nuvem, podemos adotar uma configuração minimalista [9] e [1], garantindo o processamento local para a maioria do seu trabalho, deixando os picos de carga de trabalho para uma nuvem externa. Por ter um gerenciamento automatizado, baseado em experiências anteriores e os resultados, pode-se gerir com interferência humana mínima. E com um sistema de gestão 24/7, com o objetivo de proporcionar uma melhor utilização dos recursos, ele vai ampliar a vida útil de equipamentos, diminuir o tempo de inatividade a partir de erros humanos e reduzir as despesas através da adoção de estratégias inteligentes para a utilização dos recursos.

A estrutura tem a capacidade de adotar metas, objetivos de SLA (*e.g.* 99,999%), metas para o consumo de energia (kWh X média por dia) ou metas para a emissão de calor (média de Y BTU por dia).

A estrutura deve ainda reagir com os eventos do ambiente a fim de cumprir as metas pré-definidas. Eventos como o estado da UPS para baixo, sensores de temperatura acusando altos graus ou alarmes de incêndio. Em paralelo, se adapta ao ambiente dinâmico, a fim de cumprir as metas. Adaptações, como diminuição do sistema de arrefecimento, quando é interessante ativar a UPS, ou bloquear o acesso ao centro de dados, quando necessário.

O desafio é reduzir, para reduzir o número de máquinas físicas necessárias, reduzindo os custos de CAPEX. Neste novo cenário com poucas máquinas físicas, o desafio é ter robustez de equipamentos, especificamente recursos de E/S e de rede. Com isso, torna-se necessário a existência de profissionais capazes de fazê-lo. No entanto, pode-se reduzir a necessidade de refrigeração. O número de máquinas físicas reduzidas e facilidade de gerenciamento deman-

dam menos homens/hora, porém capacitados.

A redução na quantidade de máquinas físicas [3] e [2] impacta na redução de equipamentos de refrigeração e reduz duplamente o consumo de energia, pelos servidores e pelo sistema de refrigeração, consumindo menos BTU/hora.

4.5 Análises & Tendências

Notamos que os tópicos analisados aperfeiçoam-se a cada evolução das estruturas, conforme demonstrado na **Tabela 6**. Apresentando uma maior modularização das áreas e expansão da funcionalidades.

Funcionalidade	Normal	Nuvem	Nuvem Verde
Flexibilidade	Baixa	Média	Média
Disponibilidade	Baixa	Média	Média
Custo	Alto	Média	Alto
Sustentabilidade	Baixa	Média	Alta

Tabela 6
Comparação entre as tecnologias

Na orquestração física a tendência é a separação da configuração dos servidores físicos, da configuração das MV, responsáveis pelos serviços. Permitindo crescimento horizontal da nuvem e uma livre configuração de suas MV.

Para a orquestração de serviços, a reorganização dinâmica dos recursos dentro da nuvem mostrou-se imprescindível para a otimização da utilização dos recursos, melhorando o desempenho de suas MV e serviços.

O balanceamento de carga mostrou-se extremamente dependente da mobilidade da estrutura. Este, primordialmente preso pela estaticidade da estrutura física, mostra agora tendências de adaptação contextual, variando seu local de processamento de acordo com o fluxo inferido pelo ambiente.

A alta disponibilidade mostrou-se uma funcionalidade em expansão vertical e horizontal. Partindo inicialmente do nível de serviços, utilizando redundância de serviços, esta estendeu-se ao nível das MV, utilizando redundância e adaptações de MV, e ao nível das máquinas físicas, utilizando estratégias sustentáveis como "*wake-on-fail*".

O quesito CAPEX demonstrou amadurecimento, trocando quantidade por qualidade. Quanto aos equipamentos, mostra tendências a utilização de menos equipamentos, porém mais robustos e especializados, como *storages*, redes *infiniband* e processadores gráficos. Quanto a utilização dos mesmos, começa a prezar mais pela saúde do mesmos, almejando aumentar sua vida útil com uma estratégia parcimoniosa de uso.

Os OPEX visam o limiar da subsistência. Procurando cada vez mais utilizar o mínimo (recurso energético) e demandar o mínimo (de calor). Paralelamente prevê a automatização da nuvem, e uma interferência humana é cada vez menor, mesmo exigindo uma mão-de-obra mais capacitada.

As tendências de responsabilidade social vem agregando valores as empresas que adotam soluções sustentáveis em seu modo de trabalhar. Assim a questão de sustentabilidade vem, junto com as questão do OPEX e do CAPEX, formar um triângulo o qual exige equilíbrio da parte gerencial. Apesar de geralmente a queda das OPEX e CAPEX serem reflexo da adoção de soluções sustentáveis (como *desligamento e hibernação* e estruturas minimalistas), outras vezes a adoção da mesma tem conseqüências inversas, visto que produtos ecológicos (os quais agridem menos a natureza) tendem a ser mais caros. A tendência, então, é que o equilíbrio entre estes três pontos seja controlado pela nuvem de forma automatizada. Como por exemplo sistemas de nobreaks inteligentes que recarregam-se em horários de baixo custo e provem sua energia em horários de alto custo [11].

A interação entre a nuvem e seu ambiente foi o ponto mais fraco encontrado. Talvez pela dificuldade, ou falta de padrões, de intercomunicação entre os elementos não computacionais, com sistemas de refrigeração (não reativos), *nobreaks* e alarmes. E até mesmo de elementos subjugados, como equipamentos de interconexão de redes. Porém, incentivados por algumas pesquisas [8], esta atitude esta mudando. Tendendo para nuvens que residem em ambientes que mudam para melhor comporta-la, utilizando sistemas de refrigeração pró-ativos (baseados na utilização dos serviços), sistemas de *nobreaks* inteligentes e segurança de dados baseado no acesso físico da nuvem.

5 Conclusão

Neste trabalho, propusemos um modelo integrado de gestão de recursos para a nuvem verde com base nos conceitos de modelos de organização, gerenciamento de redes e computação distribuída. Nós demonstramos que a solução proposta oferece confiabilidade e sustentabilidade, contribuindo para o nosso objetivo de **otimizar a utilização de energia e reduzir as emissões de carbono**. Obtivemos os seguintes resultados no ambiente analisado:

- Dinâmica *Orquestração Física* e *Orquestração de Serviços* levou a uma **economia de energia de 20%**, quando comparado a outros métodos estáticos.
- Melhoria dos esquemas de *Balanceamento de Carga* e *Alta Disponibilidade* fornecem até **90% de diminuição de erro de SLA**, isto devido ao fato de isolar os grupos de serviços nas nuvens, em diferentes cenários de carga inesperados (como ataques DoS), para prevenir o impacto de SLA sobre outros serviços.

- Redução da margem de utilização de recurso que fornece redução de **33% nos investimentos - Capex** e adicionalmente **poupança de energia de 30%**.
- Finalmente, a gestão do sistema de refrigeração fornecidos pelo módulo de gestão ambiental fornece cerca de **50% de economia de emissões de calor** e conseqüente reduz as emissões de carbono.

Como trabalho futuro, a intenção de simular estratégias, para obter um *feedback* mais preciso do modelo. Testando diferentes abordagens de *Crenças* e *Plano Regras*. Além disso, gostaríamos de explorar algumas técnicas de inteligências artificiais, tais como redes *bayesianas*, no recálculo da crença.

Referências

- [1] Buyya, R., A. Beloglazov and J. H. Abawajy, *Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges*, Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA2010), Las Vegas, USA, July 12-15, 2010. **abs/1006.0308** (2010).
- [2] Buyya, R., R. Ranjan and R. Calheiros, *Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services*, in: C.-H. Hsu, L. Yang, J. Park and S.-S. Yeo, editors, *Algorithms and Architectures for Parallel Processing*, Lecture Notes in Computer Science **6081**, Springer Berlin / Heidelberg, 2010 pp. 13–31, 10.1007/978-3-642-13119-62. URL <http://dx.doi.org/10.1007/978-3-642-13119-62>
- [3] Calheiros, R. N., R. Ranjan, A. Beloglazov, C. A. F. De Rose and R. Buyya, *Cloudsim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms*, Software: Practice and Experience (2010), pp. 37–48. URL <http://dx.doi.org/10.1002/spe.995>
- [4] Dell, C., *Dell energy smart solution advisor* (2010). URL <http://solutions.dell.com/DellStarOnline/DCCP.aspx>
- [5] Dias de Assuncao, M., L. Lefèvre and A.-C. Orgerie, *Investigating the energy consumption of virtual machines* (2010), poster in E-Energy 2010 conference : The First International Conference on Energy-Efficient Computing and Networking, Passau, Germany.
- [6] Dignum, F., V. Dignum, J. Padget and J. Vázquez-Salceda, *Organizing web services to develop dynamic, flexible, distributed systems*, in: *iiWAS '09: Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (2009), pp. 225–234.
- [7] Durkee, D., *Why cloud computing will never be free*, Queue **8** (2010), pp. 20–29.
- [8] Google, C., *Google data center power usage efficiency* (2010). URL <http://www.google.com/corporate/datacenters/measuring.html>
- [9] Liu, L., H. Wang, X. Liu, X. Jin, W. B. He, Q. B. Wang and Y. Chen, *Greencloud: a new architecture for green data center*, in: *ICAC-INDST '09: Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session* (2009), pp. 29–38.
- [10] Valancius, V., N. Laoutaris, L. Massoulié, C. Diot and P. Rodriguez, *Greening the internet with nano data centers*, in: *CoNEXT '09: Proceedings of the 5th international conference on Emerging networking experiments and technologies* (2009), pp. 37–48.
- [11] Vytelingum, P., T. D. Voice, S. D. Ramchurn, A. Rogers and N. R. Jennings, *Agent-based micro-storage management for the smart grid*, in: *AAMAS '10: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems* (2010), pp. 39–46.