

Eixo Temático: Inovação e Sustentabilidade em Diferentes Setores

**UM MODELO ESTATÍSTICO PARA GERENCIAMENTO DE AMBIENTES DE
COMPUTAÇÃO EM NUVEM VERDE**

**A STATISTICAL MODEL TO MANAGEMENT OF GREEN CLOUD
COMPUTING ENVIRONMENTS**

Tatiana Fernanda Mousquer dos Santos, Carlos Oberdan Rolim e Adriano Mendonça Souza

RESUMO

A sociedade como um todo está em constante crescimento e desenvolvimento nas mais diversas áreas. Uma das preocupações atuais tanto de empresas e governos, é em como proporcionar que isso ocorra com sustentabilidade. Com isso, na área computacional surge o conceito de computação verde, que busca reduzir o desperdício e aumentar a eficiência dos processos, fenômenos e equipamentos envolvidos em uma estrutura computacional. Uma das formas de computação verde adotada pelos grandes centros de dados tem sido o emprego da computação em nuvem. Esse paradigma computacional visa agregar recursos através da consolidação da estrutura física existente em poucas máquinas, prestando serviços com menos custos estruturais e operacionais. Nesse contexto o presente trabalho têm por objetivo utilizar a técnica estatística de séries temporais na definição de um modelo que auxilie no provisionamento de máquinas virtuais em um ambiente de computação em nuvem verde. Os resultados demonstram que o modelo ARIMA (1, 1, 0) é o mais adequado na previsão do uso de recursos computacionais no ambiente estudado.

Palavras-chave: Computacional, Computação Verde, estatística, sustentabilidade.

ABSTRACT

The whole society is in constant growth and development in several areas. One of the current concerns of both companies and governments is how to provide it in a sustainable way. Therefore, in the computer area arises the concept of green computing, which seeks to reduce wastage and increase efficiency of processes, phenomena and devices involved in a computational framework. One way adopted by large green computing data centers has been the use of cloud computing. This computational paradigm aims to aggregate resources through consolidation of existing physical infrastructure in a few machines, providing services with fewer structural and operational costs. In this context, this paper aims to apply the time series technique in the definition of a model to assist in provisioning of virtual machines in a green cloud environment. The results demonstrates that the ARIMA (1, 1, 0) is the most suitable model to predict the future use of computational resources in the studied environment.

Keywords: Computing, Green Computing, statistics, sustainability.

1. Introdução

Com o passar dos tempos, empresas, governos e sociedade de uma maneira geral, incluíram mais um item em sua lista de preocupações: as questões ambientais. A necessidade de se utilizar medidas ecologicamente corretas visando o crescimento sustentável que possibilitem a busca do equilíbrio do ecossistema do planeta é motivo de atenção em diversas áreas.

Na área computacional, essa preocupação acabou cunhando o termo computação verde (*Green Computing*) ou TI Verde (*Green IT*). Os seus princípios se baseiam em fazer o uso eficiente dos recursos computacionais disponíveis, bem como proporcionar o equilíbrio nos pilares econômico, social e ambiental, para o desenvolvimento sustentável da tecnologia da informação (Shulz, 2009). Através dela pode-se dizer que no futuro as ações tecnológicas irão prejudicar o mínimo possível o meio ambiente e serão mais sustentáveis.

Quando discutido sobre computação verde o consumo de energia sempre vêm em primeiro lugar, pois é o fator que mais gera despesas dentro dos abordados em ambientes computacionais. Logo a preocupação é focada em achar técnicas de se fazer as mesmas coisas gastando menos energia.

Com a crescente demanda por poder de processamento não se conseguiu ainda uma “fórmula milagrosa” que seja capaz de fornecer muito processamento com um custo energético baixo em um único dispositivo. A solução mais utilizada para contornar essa situação é o uso de computação em nuvem (*Cloud Computing*).

A utilização de computação em nuvem está crescendo rapidamente, por sua inerente condição em agregar recursos e fazer com que estes sejam transparentes ao usuário. Os serviços da nuvem são prestados da mesma maneira que os de telefonia, ou de energia, onde o usuário faz o uso dos recursos e serviços prestados e ao final do mês paga a conta, sem se preocupar com o que está envolvido na prestação ou quais equipamentos estão sendo utilizados para isso.

Além do benefício da transparência de prestação do serviço, outro atrativo para um ambiente de nuvem é o alto poder computacional proporcionado pelo ambiente o qual pode crescer, ou diminuir, de acordo com as necessidades. Essa característica de “elasticidade” é possível através do uso de máquinas virtuais que são executadas sobre máquinas físicas e gerenciadas por uma camada de software.

A possibilidade de alocar novas máquinas virtuais somente quando forem necessárias acaba gerando uma grande economia de consumo energético e isso contribuiu para que esse modelo de prestação de serviços fosse massivamente adotado por Centros de Dados de TI (*Data Centers*) no mundo todo. Porém, à medida que mais e mais máquinas físicas são necessárias para dar suporte às máquinas virtuais o consumo de energia acaba novamente aumentando.

O problema de consumo de energia elétrica para dar suporte à crescente demanda computacional ainda existe. Segundo (Shulz, 2009), em 2006 os centros de dados de TI consumiam cerca de 61 bilhões de quilowatts-hora (kWh), ou 61 bilhões vezes 1000 watts-hora de eletricidade, com um custo aproximado de cerca de US\$ 4,5 bilhões de dólares.

Assim, as necessidades de redução no consumo de energia, em conjunto com a computação em nuvem motivou a criação de um novo conceito chamado de computação em nuvem verde (*Green Cloud Computing*), o qual possui diversos e interessantes desafios relacionados aos diferentes problemas de gestão dos recursos, de mecanismos e políticas para a alocação dinâmica de carga de trabalhos como utilização de memória, processador e

armazenamento em disco, entre os dispositivos existentes, preocupando-se no uso racional da energia elétrica que será consumida.

Por ser uma área nova, a comunidade acadêmica ainda busca subsídios para responder a seguinte questão: Como gerenciar um ambiente de computação em nuvem verde de forma a melhorar o aproveitamento dos recursos computacionais e diminuir o consumo energético?

Por esse motivo a presente pesquisa lança mão da técnica estatística de séries temporais para auxiliar na definição de um modelo estatístico que auxilie no gerenciamento de um ambiente de computação em nuvem verde, proporcionando um melhor provisionamento de máquinas virtuais, o que acaba ocasionando em economia de energia contribuindo para sustentabilidade.

O artigo está estruturado da seguinte forma: na seção 2 são apresentados os trabalhos relacionados na revisão de literatura; na seção 3 são apresentados alguns conceitos envolvendo séries temporais; na seção 4 são demonstradas as análises dos resultados; e por fim na seção 5 são apresentadas as conclusões e pesquisas futuras.

2. Revisão de Literatura

A sociedade como um todo está em constante crescimento e desenvolvimento nas mais diversas áreas. Uma das preocupações atuais é em como proporcionar que o crescimento ocorra com sustentabilidade.

De acordo com a ONU, sustentabilidade “é o atendimento das necessidades das gerações atuais, sem comprometer a possibilidade de satisfação das necessidades das gerações futuras”. Segundo (Cairncross, 1992) o desenvolvimento sustentável tem por premissa a idéia de que o crescimento econômico e a proteção ambiental podem ser compatíveis. Atrelar a consciência ecológica, sem perder a lucratividade de forma a obter a redução nos gastos existentes tem sido a meta de diversos setores da sociedade.

Por ser considerada fundamental para o desenvolvimento, a área computacional buscou formas de proporcionar a sustentabilidade. Com isso surge o conceito de computação verde, que busca reduzir o desperdício e aumentar a eficiência de todos os processos, fenômenos e aparelhos envolvidos em uma estrutura computacional. Além disso, ela engloba o conjunto de práticas para tornar mais sustentável e menos prejudicial o uso da computação, diminuindo a degradação ambiental, colaborando com o desenvolvimento das novas tecnologias e, o mais importante, promovendo um futuro sustentável (Agenda, 2013).

Uma das formas de computação verde adotada pelos grandes centros de dados tem sido o emprego da computação em nuvem. Esse modelo computacional visa agregar recursos através da consolidação da estrutura física existente em poucas máquinas, prestando serviços de uma forma mais otimizada, porém com menos custos estruturais e operacionais.

A seguir serão apresentadas maiores informações e alguns conceitos sobre computação em nuvem e computação verde.

2.1 Computação em Nuvem

A computação em nuvem tornou-se a concretização de um objetivo buscado há muito tempo pela indústria de TI, da computação como uma utilidade. Este novo modelo

tem a capacidade de mudar toda a forma de produção e implantação de software, além de guiar os rumos de como o hardware é desenvolvido e adquirido (Armbrust, 2010).

A computação em nuvem aponta para um futuro onde não se irá operar e processar dados em computadores locais, mas sim em pontos centralizados, terceirizando o processamento e armazenamento de dados. Este conceito não é novo. Já em 1961, John McCarthy, um pioneiro da computação, predisse que no futuro a computação seria organizada como uma utilidade pública (Foster, 2008).

Existem diversos conceitos sobre computação em nuvem, para (Armbrust, 2009), a computação em nuvem é um conjunto de serviços de rede ativados, proporcionando escalabilidade, qualidade de serviço, infra-estrutura barata de computação sob demanda e que pode ser acessada de uma forma simples e pervasiva.

Entre tantas definições, outro conceito utilizado é o de (Foster, 2008), onde computação, em nuvem é definida como sendo um paradigma computacional que faz uso de um sistema distribuído de larga escala, guiado por economia de escala, onde um conjunto de recursos de armazenamento, plataforma, poder de processamento e serviços são entregues de forma abstrata, virtualizada, escalável dinamicamente e gerenciável, a clientes externos na Internet. A figura 1 apresenta as camadas do modelo de computação em nuvem.



Figura 1 - Camadas do modelo de Computação em Nuvem

O nível mais baixo do modelo é caracterizado pelos recursos físicos que constituem a fundação da nuvem. Estes recursos podem ser de natureza diferente: clusters, Centros de Dados e computadores pessoais. Acima destes, a infra-estrutura de TI é implementada e gerenciada. Soluções comerciais de computação em nuvem são constituídas por centros de dados hospedando centenas ou milhares de máquinas, enquanto nuvens privadas podem prover um ambiente mais heterogêneo onde até mesmo os ciclos de CPU ociosos poderão ser usados para nivelar a carga de trabalho computacional.

A infra-estrutura física é gerenciada pela camada intermediária de núcleo (core middleware) cujos objetivos são prover um ambiente apropriado para as aplicações e utilizar os recursos físicos da melhor forma. Tecnologias de virtualização provem recursos como isolamento de aplicação, qualidade de serviços e sandboxing (do original inglês, caixa de areia, um dispositivo de segurança que isola o aplicativo em um local específico).

As tecnologias de virtualização auxiliam na criação de um ambiente onde serviços profissionais e comerciais são integrados. Estes incluem: negociação de qualidade de serviço, controle de admissão, gerenciamento de execução e monitoramento, contabilização e faturamento.

A camada intermediária de núcleo (core middleware) e a infra-estrutura física, representam a plataforma onde as aplicações são implantadas. Esta plataforma é disponibilizada através de um middleware (camada de software intermediária) em nível de usuário que provê ambientes e ferramentas visando simplificar o desenvolvimento e implantação de aplicações na nuvem.

No topo do modelo, diferentes tipos de aplicações tiram proveito dos recursos oferecidos pela nuvem. Desenvolvedores de software independentes podem confiar na nuvem para desenvolver e gerenciar novas aplicações e fornecer seus serviços aos usuários que os utilizam sem saber nem mesmo onde estes estão sendo executados.

Computação em nuvem é um campo fértil para o desenvolvimento de pesquisas e trabalhos. Sua franca expansão e desenvolvimento têm em criado diversas oportunidades bem como desafios e isto torna seu estudo uma tarefa importante para o amadurecimento e consolidação da tecnologia.

2.2 Computação Verde

Na onda da discussão sobre o aquecimento global, desenvolvimento sustentável e preservação do ambiente a área de TI cunhou o termo computação verde (também chamado de TI Verde). Computação verde é um conjunto de práticas que visa tornar mais sustentável e menos prejudicial o uso da computação.

A preocupação com economia de energia e corte de gastos sempre existiu em relação ao emprego da tecnologia. No caso da TI, a preocupação é ainda maior, já que os Centros de Dados costumam ser os maiores sugadores da energia elétrica de uma companhia. Só para se ter uma idéia, em um banco, por exemplo, a energia que a TI utiliza pode chegar a quase metade de todo consumo da instituição.

Além da redução de consumo energético, outra preocupação relacionada à área de tecnologia refere-se à produção e consumo de produtos eletrônicos. No ano passado foram vendidos cerca de 6 milhões de desktops no Brasil, segundo a IDC. O País ocupa, hoje, a terceira posição no mundo em unidades vendidas. Grandes corporações, pequenas empresas e usuários domésticos começaram a comprar máquinas ultramodernas, com processadores e placas de última geração. Se nada mudar, uma parte significativa desses equipamentos vai estar em aterros, dividindo espaço com monitores aposentados, tocadores de MP3 ou celulares. Eles farão companhia aos produtos eletrônicos produzido no início da década de noventa, formando toneladas do chamado “*e-waste*”, o lixo eletrônico.

Se jogado sem controle de volta ao ambiente, o lixo eletrônico não só leva milhares de anos para se decompor, como também é um problema ambiental e de saúde pública por conta das substâncias tóxicas utilizadas em sua fabricação, como chumbo e mercúrio, que podem contaminar o solo ou os lençóis freáticos e causar doenças como câncer, por exemplo, ou mutações em pessoas cujas moradias são próximas aos lixões onde as máquinas foram jogadas sem cuidado.

A alternativa mais rápida para eliminar esses produtos, a incineração, também não é sustentável. Além de colocar diversos gases poluentes na atmosfera, como os altamente tóxicos e cancerígenos PAH (hidrocarbonetos policíclicos aromáticos), a prática também culmina com problemas relacionados ao combustível gasto para a realização da queima.

O passivo ambiental em TI envolve mais do que a disposição dos produtos. A fabricação é um problema. Para se ter uma idéia, a produção de uma estação de trabalho com monitor CRT de 17 polegadas demandou, em 2004, 240 quilos em combustíveis fósseis, utilizou 22 quilos de produtos químicos e cerca de 1,4 mil litros de água. Isso acaba aumentando a emissão de dióxido de carbono o que causa sérios impactos ao meio ambiente. As informações são do livro *Computers and the Environment: Understanding and Managing their impacts* (Computadores e o Meio-ambiente: entendendo e gerenciando seus impactos) (Kuehr, 2003), lançado em 2003 pela Universidade da Organização das Nações Unidas (ONU).

A ONU, aliás, é uma das entidades que tem debatido bastante a questão do lixo eletrônico. A organização lançou o StEP (www.step-initiative.org), projeto que une entidades como o MIT (Instituto de Tecnologia de Massachusetts), a Academia Chinesa de Ciências e empresas privadas de TI, como Dell, Microsoft, Phillips e Cisco, com objetivo de encontrar novas formas para diminuir a quantidade de lixo eletrônico e tratá-lo melhor.

Os governos também começaram a se movimentar. Regulamentações foram criadas para atacar o problema, demandando a proibição do uso de substâncias tóxicas. A União Européia, uma das administrações mais avançadas neste sentido, criou duas diretrizes chamadas WEEE (Waste Electrical and Electronic Equipment, que significa Resíduos de Equipamentos Eletroeletrônicos) e RoHS (Restriction of Hazardous Substances, que significa Restrição a substâncias perigosas). Ambas buscam garantir que lançamento dos resíduos químicos oriundos de eletrônicos, especialmente o chumbo e mercúrio, sejam menos agressivos ao meio ambiente.

Como se percebe a computação verde é um assunto que está merecendo atenção. A sua adoção, precisa ser iniciada nos projetos e na estruturação das diversas áreas computacionais para que sejam analisados novos métodos, técnicas e materiais que tragam benefícios não somente para a natureza, mas também para toda a humanidade.

Diante deste contexto se tem alguns autores interessados nesta problemática como (Silva et al, 2010), o qual afirma que os computadores fazem parte de um grupo que é considerado um dos grandes consumidores de energia elétrica, tanto na sua produção, tempo de vida de uso e no seu descarte, com isso apresenta propostas para implantação de procedimentos com vista a reduzir o consumo de energia por parte desses equipamentos de informática nas Instituições de Ensino Superior (IES), e ainda apresenta as vantagens das soluções de computação em nuvem, *grid* computacional e virtualização.

Em (Werner, 2011) se propõe uma solução para o controle integrado de computação e os elementos do ambiente em nuvens verdes. A abordagem funciona com base em modelos de organização que regulam o comportamento dos componentes autônomos (agentes), que veem os elementos ambientais como prestadores de serviços, por exemplo, servidores de processamento, carga de serviço de distribuição, processador de tarefa, serviço de redução de temperatura, entre outros. O trabalho introduz um modelo de gestão do sistema, e um modelo de alocação e distribuição de máquinas virtuais, analisando o comportamento do sistema, são descritos os princípios de funcionamento, e por fim é apresentado um cenário de caso de uso.

Ainda (Wanders, 2011), apresenta o conceito de TI Verde, bem como identifica as melhores práticas que empresas podem adotar para se adequarem a esse novo paradigma da área de tecnologia da informação, conseguindo desse modo utilizar a TI de forma inteligente e consciente, além de reduzir o consumo de energia e o impacto ambiental, para assim colaborar com a sustentabilidade do planeta. Discorre também sobre as melhores estratégias utilizadas com o objetivo de transformar o data center “verde”.

Assim se percebe que nenhuma dessas iniciativas é capaz de proporcionar através de um modelo estatístico, indicativos de provisionamentos de recursos em servidores virtuais.

Por isso a busca por técnicas ou ferramentas que melhorem a relação energia/eficiência é desafiadora e importante para que o crescimento das nuvens aconteça de uma forma sustentável.

Dessa forma na próxima seção irá se falar sobre uma técnica estatística chamada séries temporais, que pode auxiliar no provisionamento de recursos computacionais como o consumo em disco contribuindo para a sustentabilidade.

3. Análise estatística: séries temporais

Com a evolução das tecnologias se percebe uma maior dependência da TI (Tecnologia de Informação) nas empresas. Este trabalho apresenta a utilização de uma técnica estatística chamada de séries temporais, cujo objetivo é verificar se através desta técnica é possível fazer o gerenciamento de utilização de processador de um servidor em um ambiente de computação em nuvem verde.

De acordo com Morettin e Toloí (2004), uma série temporal é qualquer conjunto de observações ordenadas no tempo. Sendo que os objetivos de uma série temporal são investigar o mecanismo gerador da série temporal, ou seja, como foi gerada a série, fazer previsões de valores futuros da série, analisar seu comportamento ao longo do tempo para assim analisar informações relevantes contidas nos dados da série. Uma metodologia bastante utilizada na análise de uma série temporal é Box e Jenkins, esta consiste em ajustar modelos auto-regressivos (AR) integrados de médias móveis (MA).

Conforme (Tápia, 2000), a utilização de séries temporais pelo método Box e Jenkins, é representada pelo conjunto de processos estocásticos ARIMA (do inglês Auto Regressive Integradet Moving Avarage) representado pelas letras (p, d, q), em que p é a parte autoregressiva, d representa o número de diferenças efetuadas na série para que se possa tornar estacionária e q representa as médias móveis.

A construção da modelagem ARIMA, parte do pressuposto de que as séries temporais envolvidas na análise são geradas por um processo estocástico estacionário, sendo representada a partir de um modelo matemático. Dessa forma se Z_t não é estacionária, mas $Z_t = \Delta Z_t = Z_t - Z_{t-1}$ é estacionária, então Z_t é dita integrada de ordem (1). Caso seja necessário efetuar duas diferenças para tornar a série estacionária, então Z_t é denominada de ordem (2), assim $Z_t = \Delta^2 Z_t = \Delta(\Delta Z_t) = \Delta(Z_t - Z_{t-1})$ se torna estacionária.

A construção do modelo usando séries temporais segue um ciclo iterativo da metodologia de Box e Jenkins, é composto pelas quatro etapas, conforme (Gujarati, 2000): (i) Identificação: descobrir os valores apropriados para os parâmetros. Para determinar suas ordens e valores, a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) auxiliam nessa tarefa; (ii) Estimação: estimar os parâmetros dos termos autorregressivos; (iii) Verificação de diagnóstico: nessa etapa procura-se atestar se o modelo identificado e estimado é adequado, ou seja, se ele descreve adequadamente a série de dados. A forma de verificação comumente utilizada é a realização da análise dos resíduos do modelo; (iv) Previsão: consiste em realizar a previsão, mas é importante verificar a potencialidade de previsão do modelo.

Ao se modelar uma série temporal se pressupõe a utilização de uma série estacionária. Desse modo há duas características para os modelos, uma é quando a série já

se encontra estacionária, representada pelo modelo ARMA, e a outra é quando se faz necessário estacionarizar a série, para após aplicar a modelagem, este se denomina ARIMA, ambos os modelos são designados genericamente por ARIMA (p, d, q). A seguir serão apresentados os resultados as análises apresentam que houve necessidade de se tornar a série estacionária aplicando uma diferença (d).

4. Análise dos resultados

Para se efetuar o gerenciamento de um ambiente de computação em nuvem, é necessário ter uma ferramenta que seja capaz de prever o comportamento dos recursos computacionais e assim fazer um melhor provisionamento de recursos computacionais, como utilização de disco, de memória e processador e com isso ocasionar uma maior economia de energia. Esse é um problema enfrentado em *data centers*.

Não existem muitos trabalhos na literatura, que indiquem métodos ou modelos estatísticos, que sejam capazes de gerar indicativos de provisionamentos de recursos em servidores virtuais, que busquem ferramentas capazes de melhorar a relação energia/eficiência gerando a sustentabilidade de energia na computação em nuvem. Por esse motivo o presente trabalho visa utilizar o método estatístico: séries temporais.

Os dados foram obtidos de um cluster de servidores virtuais de uma empresa de hospedagem de sites, foram coletados durante um período de 30 dias, sendo que a coleta era feita cada trinta segundos. Ao final desse período foi montado um *dataset* contendo o consumo de disco desse período. Consumo de disco é o consumo em megabytes das requisições sobre o espaço de armazenamento do sistema (Shubert et al, 2011). Os dados foram manipulados através do *software Statistica*.

Após a manipulação e análise do *dataset*, foi então definido um modelo capaz de lidar com o monitoramento e predição de recursos como o consumo de disco. Assim conforme (Box & Jenkins, 1994), o modelo ARIMA (p, d, q) citado anteriormente é adequado para a previsão de séries temporais cujo processo estocástico não é estacionário, como é o caso dos dados analisados, conforme figura 2. Logo, foi necessário fazer uma diferenciação na série original para que essa torna-se estacionária.

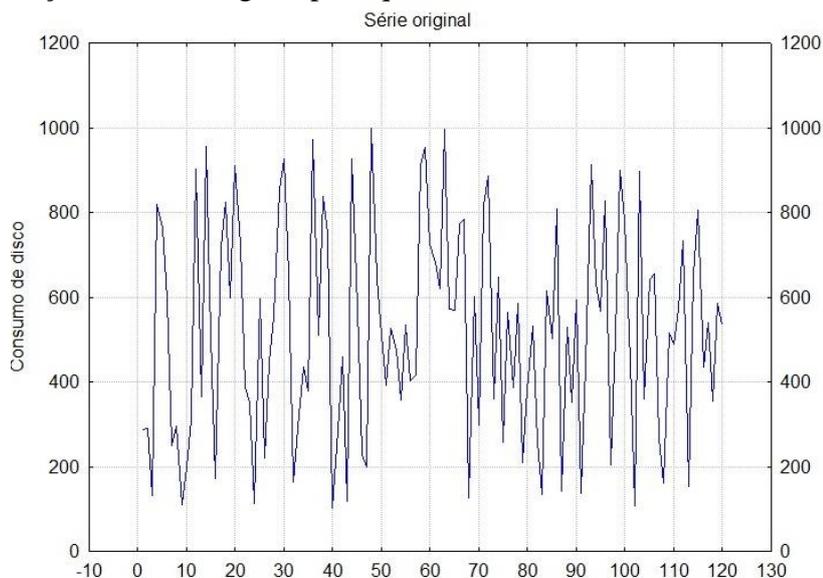


Figura 2 - Série original do consumo em disco de um servidor virtual

Através da utilização do modelo ARIMA, se verificará uma previsão de utilização de disco em servidores virtuais. Assim podem-se diminuir desperdícios na utilização de recursos do sistema e como consequência haver uma diminuição no consumo de energia dos servidores. Para encontrar o melhor modelo foram analisadas 200 observações relacionadas ao consumo de disco. Na tabela 1 estão os melhores modelos ARIMA, estimados para a série.

Variável	Modelo	Coefficiente	p-value	AIC	BIC	Ruído branco
DISCO	ARIMA (1,1,0)	$\Phi = -0,4519$	0,0000	-5,453485	-5,407026	Sim
DISCO	ARIMA (0,1,1)	$\theta = 0,92100$	0,0000	-5,470151	-5,446922	Sim
DISCO	ARIMA (2,1,1)	$\Phi = -0,5888$ $\theta = -0,3037$	0,0000 0,00849	-5,436818	-5,367131	Não

Tabela 1 - Modelos estimados para o consumo de disco em servidores virtuais

Ao analisar a tabela acima, nota-se que os modelos de previsão encontrados foram ARIMA auto-regressivos (p), com uma diferenciação (d), já que houve necessidade de dar uma diferença nas séries para que pudessem se tornar estacionárias. Sendo que os critérios AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion) são usados para comparar modelos diferentes, medindo a verossimilhança. Esses aumentam conforme a soma dos quadrados dos resíduos (SQE) aumenta. Já o p-value indica o quanto as variáveis são significativas, nesse caso ao nível de 5%, que foi a porcentagem utilizada para se modelar as séries.

Assim ao se fazer a análise dos critérios mencionados, deve-se escolher o melhor modelo de previsão. O menor valor relacionado aos critérios AIC e BIC observado na tabela 1, refere-se ao modelo ARIMA (2,1,1). Porém, além de ser o menor valor nos critérios AIC e BIC, também tem que haver a independência dos erros, ou seja, o ruído branco. Por esse motivo o modelo mais parcimonioso, ou seja, com menor número de parâmetros foi ARIMA (1,1,0), conforme mostra figura 3 na análise dos resíduos do modelo.

Função autocorrelação dos resíduos
D(-1); ARIMA (1,0,0) residuals;

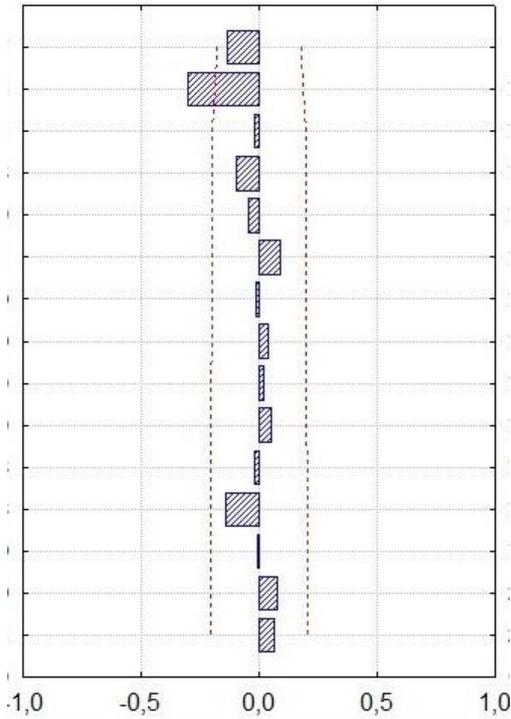


Figura 3 - Função autocorrelação dos resíduos

Outra maneira de verificar se o modelo é adequado ou não, é através da análise dos resíduos através da distribuição normal, conforme figura 4.

Distribuição normal dos resíduos
D(-1); ARIMA (1,0,0) residuals;

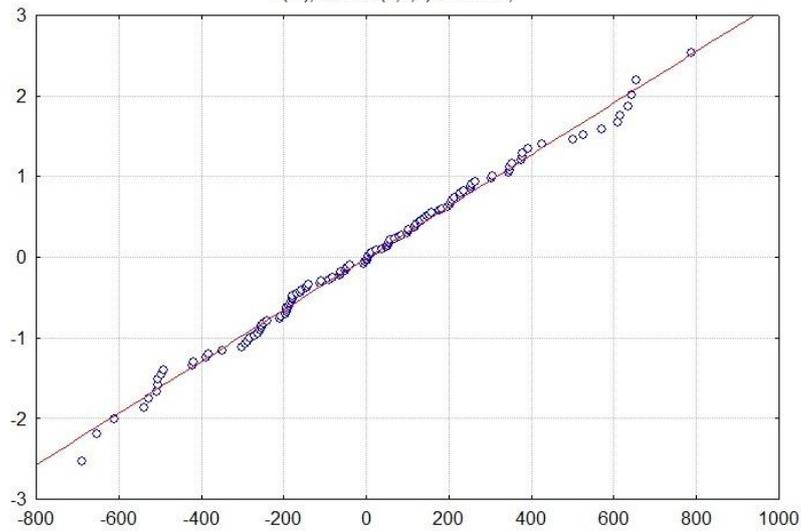


Figura 4 - distribuição normal dos resíduos

Com isso através dos critérios já mencionados, modelo parcimonioso, p valor significativo ($p < 0,05$), existência de ruído branco, observação dos valores dos critérios AIC e BIC, verifica-se que o melhor modelo de previsão dos dados analisados do consumo de disco em servidores virtuais, foi o ARIMA (1,1,0). Após escolher o melhor modelo ARIMA, foi efetuada a previsão dos instantes futuros, que podem ser verificados na tabela 2.

Consumo em disco de um servidor virtual
82.833
78.931
80.654
79.893
80.229
80.081
80.147
80.118
80.130
80.125
80.127
80.126

Tabela 2- Previsões da série analisada

A partir dos critérios assumidos, observa-se na tabela 2, a previsão da variável de uso em disco, a 12 passos à frente, isso significa que conforme mencionado, os dados foram coletados de 30 em 30 segundos o que indica que os 12 passos de previsão apresentam o comportamento da variável de uso em disco ao passo futuro de 6 minutos. Assim ao se fazer essa previsão da para afirmar que é possível saber o quanto o recurso de disco em um servidor virtual vai ser utilizado.

Por fim, a partir das análises nas séries e com base nos resultados obtidos, verifica-se que a proposta abordada mostrou-se válida. A utilização de uma série temporal, a partir do modelo ARIMA, pode ser uma alternativa para gerenciar um ambiente de computação em nuvem verde, com isso melhorar o aproveitamento dos recursos computacionais e assim diminuir o consumo energético.

5. Conclusão

Efetuar o provisionamento de recursos computacionais em um ambiente de computação em nuvem, de modo com que esse provisionamento diminua o consumo de energia de *datacenters* a fim de contribuir com a sustentabilidade é um desafio.

O presente trabalho visou investigar a existência de um método estatístico capaz de apontar de forma preditiva o consumo de recursos computacionais (disco), necessários fazer um melhor provisionamento de máquinas virtuais, ocasionando economia de energia e conseqüentemente contribuir para a sustentabilidade.

A análise dos resultados demonstra que é viável a utilização de séries temporais para prever o consumo de disco em servidores virtuais, neste estudo o melhor modelo de predição foi ARIMA (1, 1, 0). Com isso chega-se a conclusão de que a computação em nuvem verde pode se beneficiar da utilização de séries temporais, pois esta possibilitou analisar as informações contidas no *dataset* e com isso fazer previsões sobre o padrão do comportamento destes dados no futuro.

Assim através da capacidade preditiva desta técnica estatística, consegue-se fornecer indicativos necessários ao provimento de recursos computacionais a fim de que o ambiente não tenha tanta ociosidade e desperdício de energia.

Ainda como proposta futura de um novo trabalho pretende-se estudar e verificar a predição de outras variáveis importantes para o desempenho de serviços em ambientes de computação em nuvem, como a verificação do consumo de memória em uma arquitetura de computação em nuvem, além disso, pretende-se fazer um comparativo com outras técnicas preditivas como por exemplo redes neurais, com o intuito de verificar qual é a técnica que tem menor margem de erro para se efetuar uma predição mais exata.

Referências Bibliográficas

ARMBRUST, M., FOX, A., GRIFFITH R., JOSEPH, A. D., **A view of cloud computing.** Commun. ACM, 53(4):50–58, April 2010.

ARMBRUST, M., FOX, A., GRIFFITH, R., JOSEPH, A. D., KATZ, R. H., KONWINSKI, A., LEE, G., PATTERSON, D. A., RABKIN, A., STOICA, I., ZAHARIA, M. **Above the clouds: A berkeley view of cloud computing.** Technical report, EECS Department, University of California, Berkeley. 2009.

AGENDA Sustentável. URL: Disponível em <http://www.agendasustentavel.com.br> - Acesso em Junho 2013.

BOX, G. E. P., JENKINS, G. M.;REINSEL, G. C. **Time Series Analysis:Forecasting and Control (Third ed.).** Englewood Cliffs NJ: Prentice-Hall, 1994.

CAIRNCROSS, F.; **Meio Ambiente: custos e benefícios.** Cid Knipel Moreira.1a edição, São Paulo: Nobel, 1992.

FOSTER,I., ZHAO, Y., RAICU, I., LU, S., **Cloud computing and grid computing 360-Degree compared.** In 2008 Grid Computing Environments Workshop, pages 1–10. IEEE, November 2008.

GARTNET Gartner Identifies the Top 10 Strategic Technologies for 2011 URL: <http://www.gartner.com/it/page.jsp?id=1454221>. Acesso em Junho 2013.

GUJARATI, D. N. **Econometria básica.** São Paulo: Makron Books, 2000.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 1ª edição. Editora Edgard Blucher. 2004.

KUEHR R., WILLIAMS, E., **Computers and the Environment: Understanding and Managing Their Impacts**, Kluwer Academic Publishers, Eco-Efficiency in Industry and Science Series, Dordrecht/NL, Outubro 2003, 300 pages.

SCHUBERT, F.; ROLIM, C.O.; WESTPHALL, C.B.; **Aplicação de algoritmos de provisionamento baseados em contratos de nível de serviço para computação em nuvem**. IX Workshop em Clouds, Grids e Aplicações, 2011.

SILVA, Manoel R. P.; ZANETI, Gislaine B.; ZAGO, Maria G.; SOUZA, André N. **TI Verde – Princípios e Práticas Sustentáveis para Aplicação em Universidades**. Simpósio Brasileiro de Sistemas Elétricos. Belém – PA. 2010

SHULZ, G. **The Green and Virtual Data Center**. New York: CRC Press - Taylor Francis Group, 2009. ISBN 978-1-4200-8666-9.

TÁPIA, Milena. **Redes Neurais Artificiais: Uma Aplicação na Previsão de Preços de Ovos**. Dissertação de mestrado, Universidade Federal de Santa Catarina - UFSC, 2000.

WANDERS, Markus.; **Data Center Verde: como reduzir o impacto ambiental**. Cadernos de Graduação - Ciências Exatas e Tecnológicas. Aracaju - SE. V. 13, n.13, p. 25-36. Jan./Jun. 2011.

WERNER, Jorge. **Uma abordagem para alocação de máquinas virtuais em ambientes de computação em nuvem verde**. Dissertação submetida ao Curso de Pós-Graduação em Ciências da Computação para a obtenção do Grau de Mestre em Ciências da Computação. Universidade Federal de Santa Catarina, Florianópolis – SC. 2011.