



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

MARISTELLA RIBAS

UM MODELO DE DECISÃO PARA ADOÇÃO DE SERVIÇOS EM NUVEM
USANDO REDES DE PETRI

FORTALEZA

2015

MARISTELLA RIBAS

**UM MODELO DE DECISÃO PARA ADOÇÃO DE SERVIÇOS EM NUVEM
USANDO REDES DE PETRI**

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Doutora em Engenharia de Teleinformática. Área de concentração: Sinais e Sistemas.

Orientador: Prof. Dr. José Neuman de Souza
Coorientador: Prof. Dr. Alberto Sampaio Lima

FORTALEZA

2015

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Pós-Graduação em Engenharia - BPGE

R366m Ribas, Maristella.

Um Modelo de decisão para adoção de serviços em nuvem usando redes de Petri /
Maristella Ribas. – 2015.
121 f. : il. color. , enc. ; 30 cm.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia,
Departamento de Engenharia de Teleinformática, Programa de Pós-Graduação em
Engenharia de Teleinformática, Fortaleza, 2015.
Área de concentração: Sinais e Sistemas.
Orientação: Prof. Dr. José Neuman de Souza.
Coorientação: Prof. Dr. Alberto Sampaio Lima.

1. Teleinformática. 2. Computação em nuvem. 3. Petri, Redes de. I. Título.

CDD 621.38

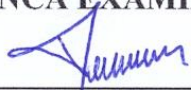
MARISTELLA RIBAS

UM MODELO DE DECISÃO PARA ADOÇÃO DE SERVIÇOS EM NUVEM USANDO REDES DE PETRI

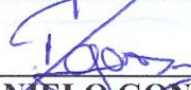
Tese submetida à Coordenação do Curso de Pós-Graduação em Engenharia de Teleinformática, da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Doutor em Engenharia de Teleinformática, área de concentração Sinais e Sistemas.

Aprovada em 24/09/2015.

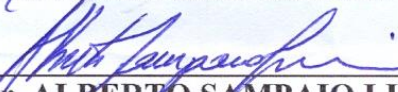
BANCA EXAMINADORA



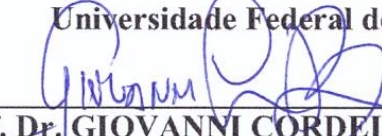
Prof. Dr. JOSE NEUMAN DE SOUZA (Orientador)
Universidade Federal do Ceará




Prof. Dr. DANIELO GONCALVES GOMES
Universidade Federal do Ceará



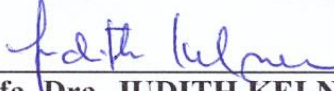
Prof. Dr. ALBERTO SAMPAIO LIMA
Universidade Federal do Ceará



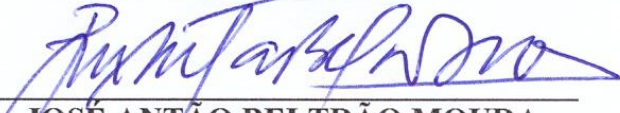
Prof. Dr. GIOVANNI CORDEIRO BARROSO
Universidade Federal do Ceará




Prof. Dr. FLÁVIO RUBENS DE CARVALHO SOUSA
Universidade Federal do Ceará



Profa. Dra. JUDITH KELNER
Universidade Federal de Pernambuco



Prof. Dr. JOSÉ ANTÃO BELTRÃO MOURA
Universidade Federal de Campina Grande



Prof. Dr. NABOR DAS CHAGAS MENDONÇA
Universidade de Fortaleza

Aos meus pais, Fernanda e Helio, eternos apoiadores.

AGRADECIMENTOS

A Deus, fonte de toda inspiração e energia.

Ao meu companheiro Aulins, minhas filhas Helena e Luciana, pela compreensão pelos longos períodos de ausência.

Às minhas companheiras de viagem: minha irmã Fernanda, minha mãe, e ao Prof. Antão Moura, pelo apoio ao pontapé inicial que deu origem a este trabalho final, o artigo apresentado em Cracóvia, no congresso BDIM de 2014.

A toda a minha família, pelo suporte e incentivo.

Ao Prof. Dr. José Neuman de Sousa, pela excelente orientação, suporte e incentivo.

Ao Prof. Dr. Alberto Sampaio Lima, pela orientação, incentivo e direcionamento.

Aos professores participantes da banca examinadora, Giovanni Cordeiro Barroso, Flávio C. S. Sousa, Antão Moura, Nabor das Chagas Mendonça, Judith Kelner e Danielo Gonçalves Gomes, pelo tempo e pelas valiosas colaborações e sugestões.

Ao colega Corneli Furtado Junior, pelo apoio no estudo e aprofundamento da ferramenta utilizada neste trabalho.

À Techne, pelo apoio financeiro recebido.

Aos colegas da Techne, pelas reflexões, críticas e sugestões recebidas.

All models are wrong, but some are useful.

George Edward Pelham Box

RESUMO

Atualmente, os serviços em nuvem são amplamente utilizados, principalmente em infraestrutura como serviço (IaaS), com grandes fornecedores oferecendo várias opções de compra e expandindo quase diariamente a gama de serviços oferecidos. A redução de custos é o principal fator que promove a adoção de serviços em nuvem. No entanto, é preciso avaliar também fatores qualitativos, o que torna o processo de decisão de adoção de serviços em nuvem uma tarefa pouco trivial para os gestores. Este trabalho propõe um modelo para tomada de decisão multicritério (MDMC) utilizando redes de Petri para avaliar um serviço de nuvem comparado com um serviço disponibilizado localmente (*on-premises*), nas dependências do usuário. A avaliação das duas opções considera questões qualitativas e de custo através de um método novo e simples, que incorpora as melhores práticas de especialistas da academia e de tecnologia da informação (TI). Além disso, o uso de redes de Petri permite extensões poderosas para realizar análises mais profundas de fatores específicos, conforme a necessidade de cada cenário. O modelo pode ser útil para apoiar gestores de TI na decisão entre as duas opções e pode ser usado para qualquer tipo de serviço de nuvem (IaaS, SaaS, PaaS). Como o custo é um dos fatores mais importantes para a adoção da nuvem, procedemos a uma análise mais profunda de um fator de custo importante. É apresentada uma extensão ao modelo, também construído com redes de Petri, para simular economias de custo usando uma determinada opção de compra de serviços em nuvens públicas, as instâncias *spot*. Através de extensas simulações em vários cenários, o trabalho conclui que a utilização de instâncias *spot* pode gerar uma grande economia no processo de escalonamento automático, mesmo em aplicações relativamente simples, utilizando apenas alguns servidores. Explorar diferentes opções de compra para os serviços em nuvem faz uma enorme diferença nos custos e pode ter grande influência no processo de tomada de decisão.

Palavras-chave: serviços em nuvem; redes de Petri; *Business Driven IT Management* (BDIM); *Analytical Hierarchical Process* (AHP); instâncias *spot*.

ABSTRACT

Cloud services are now widely used, especially in Infrastructure as a Service (IaaS), with big players offering several purchasing options, and expanding almost daily the range of offered services. Cost reduction is a major factor promoting cloud services adoption. However, qualitative factors need to be evaluated as well, making the decision process of cloud services adoption a non-trivial task for managers. In this work, we propose a Petri net-based multi-criteria decision-making (MCDM) framework, in order to evaluate a cloud service in relation to a similar on-premises offer. The evaluation of both options considers cost and qualitative issues in a novel and simple method that incorporates best practices from academy and IT specialists. Furthermore, the use of Petri net models allows powerful extensions to perform deeper analysis of specific factors as needed. The framework can be helpful for IT managers to decide between the two options, and can be used for any type of cloud service (IaaS, SaaS, PaaS). Since cost is one of the most important factors promoting cloud adoption, we proceed with a deeper analysis of one important cost factor. We propose a Petri net to model cost savings using public clouds spot Instances purchasing option. Through extensive simulations in several scenarios we conclude that spot Instances can be a very interesting option for savings in auto scaling process, even in simple business applications using only a few servers. Exploring different purchasing options for cloud services can make the difference in the decision making process.

Keywords: cloud computing; Petri nets; Business Driven IT Management (BDIM); Analytical Hierarchical Process (AHP); spot Instances.

LISTA DE ILUSTRAÇÕES

FIGURA 1 - ARQUITETURA DA COMPUTAÇÃO EM NUVEM.....	8
FIGURA 2 - MERCADO DE REVENDA DE INSTÂNCIAS RESERVADAS	16
FIGURA 3 – ILUSTRAÇÃO DE UMA RP	22
FIGURA 4 - REPRESENTAÇÃO DO DISPARO DA TRANSIÇÃO T1. (A) ESTADO INICIAL DA RP ANTES DO DISPARO, (B) ESTADO DA RP APÓS O DISPARO	23
FIGURA 5 - EXEMPLO DA ESTRUTURA DE UMA REDE DE PETRI	23
FIGURA 6 - GRAFO EQUIVALENTE À ESTRUTURA DA FIGURA 5, COM MARCAÇÕES.....	24
FIGURA 7 - REPRESENTAÇÃO DE UMA REDE DE PETRI COLORIDA (RPC)	27
FIGURA 8 - HIERARQUIA DE CRITÉRIOS PARA COMPARAÇÃO DE SERVIÇOS EM NUVEM.....	31
FIGURA 9 - MODELO DE FILAS G/G/N EM APLICAÇÃO MULTICAMADAS.....	34
FIGURA 10 - HISTÓRICO DE PREÇOS DE INSTÂNCIAS SPOT	36
FIGURA 11 - FUNÇÕES DE DENSIDADE DE PROBABILIDADE DE PREÇOS DE INSTÂNCIAS SPOT	37
FIGURA 12 - RESULTADOS OBTIDOS EM PESQUISA COM EXECUTIVOS DE TI	38
FIGURA 13 - MODELO DE DECISÃO PROPOSTO.....	47
FIGURA 14 - FUNÇÃO DE AVALIAÇÃO FINAL DO MODELO	52
FIGURA 15 - AVALIAÇÃO DE BENEFÍCIOS E RISCOS	54
FIGURA 16 - EVOLUÇÃO DOS CUSTOS UTILIZANDO TCO	57
FIGURA 17 - EVOLUÇÃO DOS CUSTOS USANDO NPV	58
FIGURA 18 - COMPARAÇÃO DE CUSTOS EM PERCENTUAL NO MODELO SAAS E ON-PREMISES	58
FIGURA 19 - REDE MODIFICADA COM TRANSIÇÃO DE SUBSTITUIÇÃO PARA CÁLCULOS DE TCO.....	60
FIGURA 20 - TRANSIÇÃO DE SUBSTITUIÇÃO QUE EXPLÍCITA CÁLCULOS TCO	61
FIGURA 21 - VISUALIZAÇÃO DA REDE APÓS EXECUÇÃO DO PASSO 1	62
FIGURA 22 - EVOLUÇÃO DOS CUSTOS NO EXEMPLO PAAS USANDO TCO	63
FIGURA 23 - EVOLUÇÃO DOS CUSTOS NO EXEMPLO PAAS USANDO NPV	63
FIGURA 24 - COMPARAÇÃO DE CUSTOS EM PERCENTUAL NO MODELO PAAS	64
FIGURA 25 – VISUALIZAÇÃO DA REDE DE EXEMPLO PAAS APÓS EXECUÇÃO DO PASSO 2	64
FIGURA 26 - CÁLCULO FINAL DO MÉTODO AHP PARA O EXEMPLO PAAS.....	65
FIGURA 27 - TRANSIÇÃO DE SUBSTITUIÇÃO “CÁLCULO E NORMALIZAÇÃO” PARA O EXEMPLO PAAS - ESTADO INICIAL.....	65
FIGURA 28 - TRANSIÇÃO DE SUBSTITUIÇÃO “CÁLCULO E NORMALIZAÇÃO” PARA O EXEMPLO PAAS - ESTADO INTERMEDIÁRIO	66
FIGURA 29 - TRANSIÇÃO DE SUBSTITUIÇÃO “CÁLCULO E NORMALIZAÇÃO” PARA O EXEMPLO PAAS - FINAL DA SIMULAÇÃO	67
FIGURA 30 - COMPARAÇÃO DO CUSTO TOTAL DE PROPRIEDADE	68
FIGURA 31 - INÍCIO DA SIMULAÇÃO DO EXEMPLO IAAS	70
FIGURA 32 – REDE DE EXEMPLO IAAS APÓS A EXECUÇÃO DO PASSO 1	71

FIGURA 33 - EVOLUÇÃO DOS CUSTOS NO EXEMPLO IAAS USANDO TCO	72
FIGURA 34 - EVOLUÇÃO DOS CUSTOS NO EXEMPLO IAAS USANDO NPV	72
FIGURA 35 – COMPARAÇÃO DE CUSTOS EM PERCENTUAL NO MODELO PAAS	72
FIGURA 36 - CÁLCULO FINAL DO MÉTODO AHP PARA O EXEMPLO IAAS.....	73
FIGURA 37 - VISUALIZAÇÃO DA REDE DE EXEMPLO IAAS APÓS EXECUÇÃO DO PASSO 2	73
FIGURA 38 - TRANSIÇÃO DE SUBSTITUIÇÃO “CÁLCULO E NORMALIZAÇÃO” PARA O EXEMPLO IAAS - ESTADO INICIAL.....	74
FIGURA 39 - TRANSIÇÃO DE SUBSTITUIÇÃO “CÁLCULO E NORMALIZAÇÃO” PARA O EXEMPLO IAAS - ESTADO INTERMEDIÁRIO	75
FIGURA 40 - TRANSIÇÃO DE SUBSTITUIÇÃO “CÁLCULO E NORMALIZAÇÃO” PARA O EXEMPLO IAAS - FINAL DA SIMULAÇÃO	75
FIGURA 41 - REDE PRINCIPAL DO MODELO - SIMULAÇÃO DE USO DE INSTÂNCIAS	78
FIGURA 42 – EXEMPLO DE FUNÇÃO DEMAND().....	79
FIGURA 43 - SUB-REDE PARA MONITORAR ESCALONAMENTO: MONITOR AUTO SCALING	82
FIGURA 44 – SUB-REDE SCALE UP	83
FIGURA 45 – SUB-REDE SCALE DOWN	85
FIGURA 46 - SUB-REDE SPOT TERMINATION	87
FIGURA 47 - HISTÓRICO DE PREÇOS DE INSTÂNCIAS SPOT	89
FIGURA 48 – RESULTADOS DAS SIMULAÇÕES: ECONOMIA OBTIDA EM TODAS AS REGIÕES PARA TODOS OS FATORES (EM %)	91
FIGURA 49 - ECONOMIA MENSAL EM DÓLARES	92
FIGURA 50 - ECONOMIA MÉDIA EM TODAS AS REGIÕES.....	93
FIGURA 51 - IMPACTO DOS FATORES	93
FIGURA 52 - COMPARATIVO DE CUSTOS DE HOSPEDAGEM LOCAL E NA NUVEM.....	94
FIGURA 53 - RESPOSTAS DOS GESTORES EM RELAÇÃO À ACURÁCIA DO MODELO	97
FIGURA 54 - RESPOSTAS DOS GESTORES EM RELAÇÃO À EFETIVIDADE DO MODELO.....	98
FIGURA 55 - COMPARATIVO ENTRE AS HIPÓTESES AVALIADAS.....	100
FIGURA 56 - REDE PARA CÁLCULO DE CUSTOS ON-PREMISES NO MODELO SAAS	111
FIGURA 57 - REDE PARA CÁLCULO DE CUSTOS NUVEM NO MODELO SAAS	112
FIGURA 58 - REDE PARA CÁLCULO DE AVALIAÇÃO QUALITATIVA NO MODELO SAAS.....	113
FIGURA 59 - REDE PARA AVALIAÇÃO FINAL DAS SOLUÇÕES	113
FIGURA 60 - AVALIAÇÃO QUALITATIVA DO CRITÉRIO QUALIDADE	114
FIGURA 61 - AVALIAÇÃO QUALITATIVA DO CRITÉRIO ESTRATÉGIA	115
FIGURA 62 - AVALIAÇÃO QUALITATIVA DO CRITÉRIO RISCO	116
FIGURA 63 - COMPARAÇÃO ENTRE CRITÉRIOS ESTRATÉGIA E QUALIDADE.....	116
FIGURA 64 - COMPARAÇÃO ENTRE CRITÉRIOS QUALIDADE E RISCOS	117
FIGURA 65 - COMPARAÇÃO ENTRE CRITÉRIOS ESTRATÉGIA E RISCOS	118

LISTA DE TABELAS

TABELA 1 - PREÇOS DE INSTÂNCIAS RESERVADAS EM 2015 (TAMANHO M3.MEDIUM)	14
TABELA 2 - CUSTOS DE INSTÂNCIAS RESERVADAS EM 2015 (USO ESTIMADO 50% PARA M3.MEDIUM COM LINUX)	15
TABELA 3 - ELEMENTOS DE CUSTO EM NUVENS HÍBRIDAS	35
TABELA 4 - ANÁLISE SWOT DOS SERVIÇOS EM NUVEM	39
TABELA 5 - ÁREAS CRÍTICAS DE RISCO DE SERVIÇOS DE NUVEM E VIRTUALIZAÇÃO	40
TABELA 6 - QUADRO COMPARATIVO DE TRABALHOS RELACIONADOS	41
TABELA 7 - ELEMENTOS DE CUSTO: PARÂMETROS INICIAIS DO MODELO SAAS	54
TABELA 8 – PESOS DE ALTERNATIVAS (PARÂMETRO DE ENTRADA)	55
TABELA 9 - PESOS DE CRITÉRIOS (PARÂMETRO DE ENTRADA)	55
TABELA 10 - CÁLCULO DA AVALIAÇÃO FINAL DAS ALTERNATIVAS (VALORES EM MILHARES DE REAIS)	56
TABELA 11 - ELEMENTOS DE CUSTO PAAS	59
TABELA 12 - PESOS DE ALTERNATIVAS (PARÂMETROS DE ENTRADA)	60
TABELA 13 - PESOS DE CRITÉRIOS (PARÂMETRO DE ENTRADA)	60
TABELA 14 - ELEMENTOS DE CUSTO: PARÂMETROS INICIAIS DO MODELO IAAS	69
TABELA 15 - PESOS DE ALTERNATIVAS (PARÂMETROS DE ENTRADA)	69
TABELA 16 - PESOS DE CRITÉRIOS (PARÂMETROS DE ENTRADA).....	69
TABELA 17 - PARÂMETROS DE ENTRADA E SAÍDA DO MODELO DE REDUÇÃO DE CUSTOS.....	79
TABELA 18 - PARÂMETRO SIMULAÇÃO PARA OBTENÇÃO DE PREÇO SPOT.....	84
TABELA 19 – PARÂMETRO TEMPO E SUA CORRESPONDÊNCIA COM HORA E DIA DA SIMULAÇÃO	84
TABELA 20 - FATORES E NÍVEIS DO EXPERIMENTO	89
TABELA 21 – PREÇOS POR HORA DE USO DAS INSTÂNCIAS UTILIZADAS NAS SIMULAÇÕES	90
TABELA 22 - RELATÓRIO DE DESEMPENHO OBTIDO PARA A SIMULAÇÃO 30.....	91

Lista de abreviaturas e siglas

AHP	<i>Analytical Hierarquical Process</i> (processo de análise hierárquica)
API	<i>Application Programming Interface</i> (interface para programação de aplicações)
AWS	<i>Amazon Web Services</i>
BDIM	<i>Business Driven IT Management</i> (gerenciamento de TI orientado a negócio)
CAPEX	<i>Capital Expense</i> (despesas de capital)
CPU	<i>Central Processing Unit</i> (unidade de processamento central)
CRM	<i>Customer Relation Management</i> (gestão de relacionamento com o cliente)
EC2	<i>Elastic Cloud Computing</i> (máquinas virtuais da Amazon)
ECU	<i>EC2 Computing Unit</i> (1 ECU equivale à capacidade de uma CPU do processador 1.0-1.2 GHz Opteron ou Xeon de 2007)
ELECTRE	<i>Elimination et Choix Traduisant la Réalité</i>
GB	Gigabyte
IAAS	<i>Infrastructure as a Service</i> (infraestrutura como serviço)
MAUT	<i>Muti-Attribute Utility Theory</i> (teoria da utilidade multi-atributos)
NIST	<i>National Institute of Standards and Technology</i>
NPV	<i>Net Present Value</i> (valor líquido atual)
OPEX	<i>Operational Expense</i> (despesa operacional)
OWA	<i>Ordered Weighted Averaging</i> (operador de média ponderada ordenada)
PAAS	<i>Platform as a Service</i> (plataforma como serviço)
PDA	<i>Personal Digital Assistant</i> (assistente pessoal digital)
PLS	<i>Partial Least Squares</i> (mínimos quadrados parciais)
QoS	<i>Quality of Service</i> (qualidade de serviço)
RAM	<i>Random Access Memory</i> (memória de acesso randômico)
SAAS	<i>Software as a Service</i> (<i>software</i> como serviço)
SLA	<i>Service Level Agreement</i> (acordo de nível de serviço)
SLO	<i>Service Level Objective</i> (objetivo de nível de serviço)
SMI	<i>Service Measurement Index</i> (índice de medida do serviço)
SPM	Solicitações por minuto
SSD	<i>Solid State Drive</i> (unidade de armazenamento)
TCO	<i>Total Cost of Ownership</i> (custo total de propriedade)
TI	Tecnologia da Informação

TOPSIS	<i>Technique for Order of Preference by Similarity to Ideal Solution</i>
vCPU	<i>Virtual CPU</i>
VM	<i>Virtual Machine</i> (máquina virtual)
WPM	<i>Weighted Product Model</i> (modelo de produto ponderado)
WSM	<i>Weighted Sum Model</i> (modelo de soma ponderada)

SUMÁRIO

Capítulo 1 INTRODUÇÃO	1
1.1 Resumo da problemática.....	2
1.2 Objetivos da Pesquisa.....	5
Capítulo 2 FUNDAMENTAÇÃO TEÓRICA	7
2.1 Computação em Nuvem.....	7
2.2 Redes de Petri.....	20
2.3 Métodos de Decisão MultiCritério.....	28
Capítulo 3 TRABALHOS RELACIONADOS	32
3.1 Modelos de Custo para Serviços em Nuvem.....	32
3.2 Avaliação Qualitativa de Serviços em Nuvem.....	37
3.3 Quadro Comparativo.....	40
Capítulo 4 O MODELO PROPOSTO	46
4.1 Passo 1: Estimativa de custos.....	48
4.2 Passo 2: Avaliação qualitativa.....	49
4.3 Passo 3: Calcular a razão custo/benefício.....	51
4.4 Exemplos de uso do modelo.....	52
Capítulo 5 REDUÇÃO DE CUSTOS USANDO INSTÂNCIAS SPOT	77
5.1 Rede Principal: Simulação do uso de instâncias.....	78
5.2 Sub-rede para monitorar escalonamento: <i>Monitor Autoscaling</i>	81
5.3 Sub-rede para Escalonar para cima: <i>Scale Up</i>	82
5.4 Sub-rede para escalonamento para baixo: <i>Scale Down</i>	84
5.5 Sub-rede para simular término de instâncias Spot: <i>Spot Termination</i>	86
5.6 Resultados da Simulação.....	88
Capítulo 6 VALIDAÇÃO DO MODELO	96
6.1 Hipótese 1 - Preferência.....	96
6.2 Hipótese 2 - Utilidade.....	96
6.3 Hipótese 3 - Acurácia.....	96
6.4 Hipótese 4 – Efetividade.....	97
6.5 Hipótese 5 - Aprimoramento.....	98
6.6 Dificuldades enfrentadas e necessidades adicionais.....	98
6.7 Sugestões e comentários.....	98
6.8 Consolidação dos Resultados.....	99
Capítulo 7 CONCLUSÃO E TRABALHOS FUTUROS	101
7.1 Trabalhos Futuros.....	102
REFERÊNCIAS	104
APÊNDICE A – Rede preliminar apresentada no congresso BDIM 2014	111
APÊNDICE B - Pesquisa qualitativa utilizada nos exemplos de uso do modelo	114
APÊNDICE C – Questionário para validação de aparência	120

Capítulo 1

INTRODUÇÃO

Computação em nuvem é uma tendência recente de tecnologia cujo objetivo é proporcionar serviços sob demanda com pagamento baseado no uso. Nesse ambiente, as empresas podem alugar capacidade de processamento e armazenamento, reduzindo seus custos. A computação em nuvem surge da necessidade de se construírem infraestruturas de tecnologia da informação (TI) complexas, para as quais os usuários têm que realizar instalação, configuração e atualização de sistemas de *software*. Além disso, recursos de computação e *hardware* são propensos a ficar obsoletos rapidamente. Assim, a utilização de plataformas computacionais de terceiros é uma solução inteligente para os usuários lidarem com infraestrutura de TI. Na computação em nuvem, os recursos de TI são fornecidos como um serviço, permitindo que os usuários os acessem sem a necessidade de conhecimento sobre a tecnologia utilizada. Assim, os usuários e empresas passaram a acessar os serviços sob demanda e independentemente de localização, o que aumentou a quantidade de serviços disponíveis.

No entanto, o fato da computação em nuvem ser uma tendência atual, e de existirem inúmeras vantagens em seu uso, não implica que todos os serviços de TI devam ser, necessariamente, implantados em nuvem. Existem situações em que a hospedagem local dos serviços de TI é vantajosa, ou, em alguns casos, até obrigatória. Por hospedagem local entende-se o modelo em que os serviços de TI são implantados em recursos computacionais próprios, adquiridos e mantidos pela instituição que os utilizará. Em geral, na literatura, esse modelo de utilização é denominado *on-premises* ou *in-house*.

Uma das muitas decisões difíceis para o gestor de TI, com as novas possibilidades disponibilizadas pela computação em nuvem, é: adotar um novo serviço de TI em nuvem ou prover o mesmo serviço instalado no modelo convencional, *on-premises*? O gestor precisa avaliar inúmeros fatores, sendo possivelmente o fator custo o mais importante. Alguns fatores podem ser erroneamente avaliados, seja por desconhecimento do gestor, por dificuldade na precisão do cenário de uso, ou por fatores externos imprevisíveis (como, por exemplo, variações cambiais). Os serviços em nuvem são atraentes devido ao baixo investimento inicial – no entanto, o departamento de TI já não terá qualquer controle sobre a infraestrutura, e vai depender completamente do cumprimento das cláusulas contratuais de acordo de nível de serviço (SLA) para garantir a qualidade do serviço.

Este trabalho tem como objetivo analisar o problema do negócio de identificar fatores e parâmetros para avaliar vantagens e desvantagens de optar por serviços em nuvem em detrimento do uso de soluções equivalentes de hospedagem local. É introduzido um modelo que permite a combinação de diversos fatores relacionados a esse tipo de decisão para auxiliar na avaliação de ambas as opções. O modelo propõe a avaliação separada dos aspectos qualitativos e do fator custo, para evitar a supervalorização do fator custo em relação aos demais. Através de abrangente pesquisa, são enumerados vários dos fatores mais relevantes, de acordo com a literatura atual.

A estrutura do modelo foi projetada utilizando redes de Petri, a fim de se beneficiar de sua descrição formal, que fornece uma interface visual simples, mas poderosa o suficiente para realizar simulações sofisticadas. O modelo foi concebido de modo a ser simples o suficiente para ser facilmente utilizável por gestores de TI, além de flexível e adaptável para diversos cenários. O uso das redes de Petri coloridas garante a flexibilidade do modelo e também permite a criação de extensões para simulações mais elaboradas, como a detalhada no capítulo 5, onde se estuda em profundidade um dos fatores de custo muito relevantes na avaliação, que é o custo de uso de máquinas virtuais.

A maioria dos artigos relacionados à avaliação dos serviços em nuvem observa um ponto estritamente técnico, o que se diferencia desse trabalho, que segue a linha de pesquisa da gerência dos serviços de TI do ponto de vista do negócio, o chamado Business Driven IT Management (BDIM). Essa linha de pesquisa busca a “aplicação de um conjunto de modelos, práticas, técnicas e ferramentas para mapear e avaliar quantitativamente dependências entre serviços de TI e desempenho dos negócios, utilizando a avaliação quantificada para melhorar a qualidade dos serviços de TI e os resultados dos negócios relacionados” (Sauvé, J. et al., 2006).

1.1 Resumo da problemática

Como visto anteriormente, a computação em nuvem representa uma opção adicional dentre as possibilidades para se obter um serviço de TI. No entanto, a variedade de serviços oferecidos na nuvem e a diferença de arquitetura entre os inúmeros serviços disponíveis dificulta muito a escolha do serviço. Como exemplo, considere uma escola privada com 2000 estudantes que precisa melhorar o seu *software* de controle acadêmico. O *software* atual foi comprado quando a escola era muito menor e, com as novas tecnologias móveis, tornou-se obsoleto. O *software* atual é executado no departamento de TI da escola, usando dois servidores, um para o servidor web e outro para o banco de dados. No entanto, esses dois servidores também precisam ser

substituídos, uma vez que já apresentam problemas técnicos. O gerente de TI, depois de cuidadosa pesquisa, decide comprar um novo *software*, que tem duas opções de implantação:

1) opção local (*on-premises*): licença permanente, pela qual a escola terá que pagar uma taxa de licenciamento inicial de R\$ 145.000 (valor para até 2500 alunos). Há também uma taxa anual (opcional) de manutenção de R\$ 29.000 para suporte e atualizações de *softwares* técnicos. Será necessário providenciar a infraestrutura de TI para instalar e executar o *software*.

2) opção de *software* como serviço (SaaS): assinatura anual de R\$ 99.000 (valor para até 2500 alunos). A escola vai acessar o *software* usando um navegador através da internet, e o fornecedor vai assinar um contrato SLA para garantir certo nível de serviço (disponibilidade, performance e segurança, por exemplo).

Em qualquer das opções acima, o vendedor recomenda o pacote de treinamento, para aprender a personalizar o *software* para as necessidades da escola (a personalização é feita através de parâmetros, sem codificação), e para garantir que os usuários possam utilizá-lo da forma mais adequada.

Essa pode ser uma decisão muito difícil para o gestor de TI. Ele tem que avaliar diferentes critérios, sendo o custo, possivelmente, o mais importante. A opção 2 (SaaS) é atraente pelo baixo investimento inicial, no entanto, o departamento de TI não terá mais controle sobre a infraestrutura, e vai depender completamente do cumprimento das cláusulas do contrato de SLA para fornecer um serviço essencial para professores e alunos.

O objetivo deste trabalho é introduzir um modelo formal para apoiar essa difícil decisão, investigando os elementos necessários para fundamentá-la da melhor maneira possível. O modelo deverá ser formal, porém simples e claro o suficiente para ser utilizado por gestores de TI com conhecimento e experiência variados de computação em nuvem. O modelo deverá ser aplicável a qualquer serviço de nuvem.

Segundo Bazerman, M. H. & Moore, D. (2004), a anatomia de uma decisão pode ser dividida em seis etapas:

- 1) Definir o problema, o que segundo os autores é uma tarefa muito suscetível a erro, pois muitas vezes o tomador de decisão (a) define o problema em termos de uma solução proposta, (b) deixa de notar um problema maior, ou (c) diagnostica o problema em termos de sintomas. O objetivo da tomada de decisão ótima é resolver o problema, não apenas eliminar os sintomas temporários;
- 2) Identificar os critérios relevantes na tomada de decisão;
- 3) Ponderar os critérios;
- 4) Gerar alternativas, identificando possíveis custos de ação;

- 5) Classificar cada alternativa segundo cada critério, o que muitas vezes é a etapa mais difícil do processo de decisão;
- 6) Identificar a solução ótima, o que teoricamente seria multiplicar as classificações da etapa 5 pelo peso de cada critério, somar as classificações ponderadas de todos os critérios para cada alternativa e escolher a solução cuja soma das classificações ponderadas seja a maior.

Os autores ressaltam que, em geral, as etapas 1 a 6 descritas não são seguidas de modo estritamente racional, pois existem diversos vieses que dificultam o processo. Por exemplo, é um comportamento frequente reter a informação mais recente ou de maior impacto e atribuir a essa informação um peso muitas vezes exagerado em uma decisão. No entanto, a análise das restrições de racionalidade no processo decisório foge do escopo deste trabalho, apesar de ser levada em consideração em alguns elementos do modelo proposto.

Este trabalho pretende fornecer elementos de apoio para a decisão, da seguinte forma:

- **Fase 1 (definição do problema):** Dada a necessidade de aquisição de um novo serviço de TI, adquiri-lo na nuvem é a melhor opção? Ou é melhor adquiri-lo para implantação local (no exemplo, seria melhor adquirir o *software* na opção SaaS ou adquirir o mesmo *software*, bem como sua infraestrutura, para instalação local)?
- **Fase 2 (identificação dos critérios):** Quais são os critérios mais relevantes? Como avaliá-los?
- **Fase 3 (ponderação de critérios):** Como ponderar os critérios?
- **Fase 4 (alternativas):** Serão estudadas duas alternativas: implantação em nuvem e implantação local.
- **Fase 5 (classificação das alternativas):** Como classificar as alternativas?
- **Fase 6 (solução):** Qual é a solução ótima?

Este trabalho introduz um modelo formal, utilizando redes de Petri, para apoiar essa decisão. O modelo será descrito detalhadamente no capítulo 4, e em linha gerais é dividido em três passos:

- 1) Estimar a vantagem (ou desvantagem) de custo do serviço de nuvem em relação ao serviço hospedado localmente (*on-premises*);
- 2) Avaliar benefícios e riscos de modo qualitativo usando um método multicritério para tomada de decisão, como o AHP;
- 3) Calcular a razão custo/benefício. A opção com a menor taxa é a melhor.

Cada passo possui suas dificuldades próprias. Em particular, a estimativa de custos na nuvem pode ser uma tarefa extremamente complexa, e por esse motivo foi destacada como um passo separado dos demais. Ao longo deste trabalho, são investigados os esforços da comunidade científica para propor soluções para as dificuldades encontradas na execução de cada um desses passos, nos mais variados cenários. A partir desses estudos, o modelo proposto

foi construído, buscando sumarizar os fatores básicos a serem avaliados e fornecer um apoio concreto aos tomadores de decisão.

1.2 Objetivos da pesquisa

Este trabalho estuda o problema de identificar fatores e parâmetros para avaliar vantagens e desvantagens do uso de serviços de nuvem em comparação com as soluções locais para aplicações empresariais. O objetivo principal é apoiar a decisão entre a adoção do serviço em modelo nuvem, em geral nuvem pública, ou partir para a construção do mesmo serviço em seu ambiente local, adquirindo infraestrutura própria e arcando com os custos técnicos para implantação e manutenção do serviço.

O trabalho não aborda os aspectos relacionados à migração de aplicações para a nuvem, quando o serviço de TI já está funcionando no ambiente local e se deseja implantá-lo na nuvem. O problema estudado refere-se à comparação de novos serviços de TI a serem implantados, no sentido de investigar qual a melhor opção, nuvem ou *on-premises*. Para tanto, será necessário avaliar cuidadosamente todos os fatores envolvidos.

Como o fator custo é um dos mais relevantes nesse tipo de decisão, também é um objetivo dessa pesquisa conhecer e detalhar os modelos de custo mais comumente utilizados para calcular o custo do serviço na nuvem, e propor políticas de redução de custo, para executar a avaliação da forma mais correta possível. As diferentes possibilidades de uso de serviços em nuvem podem trazer significativa diferença no fator custo. Dessa forma, um correto entendimento dos modelos existentes e a aplicação das políticas mais adequadas podem fazer muita diferença na avaliação final das duas opções.

Os custos internos para a montagem da infraestrutura local são muito específicos, dessa forma as possibilidades de otimização nessa área fogem do escopo dessa pesquisa. No entanto, assume-se que o gestor tomador de decisão possua sua metodologia própria para fornecer, como dado de entrada, os custos internos de sua própria infraestrutura.

Além do fator custo, os fatores qualitativos são importantes na tomada de decisão e também devem ser avaliados adequadamente. Por exemplo, o desempenho do serviço possui requisitos especiais a serem avaliados? O serviço pode ficar indisponível quando a conectividade com a Internet for interrompida? Dessa forma, o estudo desses fatores e como combiná-los para nortear a decisão final também é fundamental nesse tipo de decisão.

Para validação do modelo proposto, serão apresentados, no capítulo 6, os resultados da verificação realizada a respeito do modelo – se o mesmo aparenta ser útil, prático, pertinente

e adequado para o seu propósito. Utilizando a metodologia da validação de aparência (Kitchenham & Pfieeger, 2002), o modelo foi apresentado a gerentes da área de sistemas de informação, que responderam a um questionário indicando, em sua visão, se o modelo parece ser adequado. O questionário avalia as hipóteses de preferência, utilidade, acurácia, efetividade e aprimoramento, da seguinte forma:

- **Hipótese 1 (Preferência):** os gestores preferem o modelo apresentado ao método atual de decisão?
- **Hipótese 2 (Utilidade):** os gestores consideram o modelo útil?
- **Hipótese 3 (Acurácia):** os gestores consideram que o modelo apresentado pode produzir resultados considerados precisos o bastante para nortear a tomada de decisão?
- **Hipótese 4 (Efetividade):** os gestores consideram o modelo consistente e objetivo o bastante para nortear a tomada de decisão?
- **Hipótese 5 (Aprimoramento):** os gestores consideram que o modelo apresentado melhora o processo de tomada de decisão?

O restante da tese está organizado da forma que segue, No capítulo 2, são apresentados os fundamentos teóricos utilizados na construção do modelo. No capítulo 3, são investigados trabalhos da comunidade científica relacionados com o problema, tanto na área de custos de serviços em nuvem quanto na área de análise qualitativa. No capítulo 4, o modelo proposto é formalmente detalhado, e são apresentados exemplos de uso para serviços de nuvem do tipo IaaS, PaaS e SaaS. No capítulo 5, é apresentada uma extensão do modelo que permite a redução do custo de serviços de nuvem usando instâncias *spot*. No capítulo 6, são apresentados os resultados da validação do modelo. Finalmente, o capítulo 7 apresenta as conclusões e os trabalhos futuros.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

Este trabalho utiliza diversos conceitos da computação em nuvem e faz uso de ferramentas clássicas, como redes de Petri e métodos de decisão multicritério (MDMC), para propor um novo modelo de decisão para a adoção de serviços de nuvem. Assim, este capítulo discorre sobre os conceitos e fundamentos utilizados na nova metodologia proposta.

2.1 Computação em nuvem

A nuvem é uma metáfora para a Internet ou infraestrutura de comunicação entre os componentes arquiteturais, baseada em uma abstração que oculta a complexidade da infraestrutura. Cada parte dessa infraestrutura é provida como serviços e estes são normalmente alocados em centros de dados, utilizando *hardware* compartilhado para computação e armazenamento.

Os analistas de mercado visualizam a nuvem como elemento central de uma nova arquitetura na qual inúmeros dispositivos eletrônicos centralizam suas informações, em um modelo denominado cliente-nuvem (em uma analogia ao já conhecido modelo cliente-servidor). A nuvem será o ponto de controle e o local para registro das informações. Os aplicativos estarão espalhados em múltiplos dispositivos do cliente (até mesmo em dispositivos inusitados, como uma raquete de tênis inteligente que armazena dados de sua utilização, por exemplo). Os aplicativos devem ser sensíveis ao que está acontecendo naquele momento e podem obter dados de vários dispositivos ao mesmo tempo (Gartner Group, 2013).

2.1.1 Arquitetura da Computação em nuvem

A arquitetura de computação em nuvem é baseada em camadas, sendo que cada uma delas trata de uma particularidade na disponibilização de recursos para as aplicações (Buyya, R. et al., 2009). Uma camada é uma divisão lógica de componentes de *hardware* e *software*. Alguns desses recursos computacionais podem ser agrupados e organizados para realizar uma determinada tarefa do sistema como um todo. Cada camada pode ter seu gerenciamento ou monitoramento de forma independente das outras camadas, melhorando a flexibilidade, reuso e escalabilidade no tocante à substituição ou adição de recursos computacionais sem afetar as outras camadas. A Figura 1 exibe essas camadas e suas respectivas associações.

Figura 1 - Arquitetura da Computação em nuvem.



Adaptado de Buyya, R. et al., 2009.

A camada de mais baixo nível é a de infraestrutura física, que contém centros de dados, *clusters*, *desktops* e outros recursos de *hardware*, podendo se compor de recursos heterogêneos. Com isso, fornece certa flexibilidade e facilidade de agregação de novos recursos à medida que se tornem necessários. Uma camada de *middleware* é responsável por gerenciar a infraestrutura física e tem por objetivo fornecer um núcleo lógico de uma nuvem. Esses serviços contêm negociações de qualidade de serviço (QoS), gerenciamento dos SLAs, serviços de cobrança, serviços para verificar aceitação de requisições baseado no QoS e preço, serviços para cálculo, serviços de gerenciamento de virtualização, entre outros.

Um nível acima da camada de *middleware* encontra-se a camada responsável por prover suporte para a construção de aplicações, e que contém ferramentas ou ambientes de desenvolvimento. Esses ambientes possuem interfaces Web 2.0, *mash-ups*, componentes, recursos de programação concorrente e distribuída, suporte a *workflows*, bibliotecas de programação e linguagens de programação. Essa camada de desenvolvimento não é utilizada pelos usuários finais e, sim, pelos usuários mais experientes, aqueles que desenvolvem as soluções para computação em nuvem. Por fim, encontra-se a camada das aplicações de computação em nuvem. Essa camada é de interesse dos usuários finais, pois é por meio dela que utilizam os aplicativos.

As camadas abaixo das aplicações são responsáveis pelas características de escalabilidade, disponibilidade, ilusão de recursos infinitos e alto desempenho. Algumas soluções de arquitetura podem incluir uma camada de gerenciamento de adaptações,

responsável por fornecer ajustes às soluções. Essas adaptações ocorrem de forma automática ou semiautomática e, com isso, diminuem-se os esforços humanos para gerenciar arquiteturas de computação em nuvem (Lenk, A. et al., 2009).

A infraestrutura do ambiente de computação em nuvem normalmente é composta por um grande número – centenas ou milhares – de máquinas físicas ou nós físicos de baixo custo, conectados por meio de uma rede. Cada máquina física tem as mesmas configurações de *software*, mas pode ter variação na capacidade de *hardware* em termos de CPU, memória e armazenamento em disco. Dentro de cada máquina física existe um número variável de máquinas virtuais (VM), ou nós virtuais em execução, de acordo com a capacidade do *hardware* disponível na máquina física. Os dados são persistidos, geralmente, em sistemas de armazenamento distribuídos.

Uma definição clássica da computação em nuvem é apresentada pelo *National Institute of Standards and Technology* (NIST):

Computação em nuvem é um modelo que possibilita acesso, de modo conveniente e sob demanda, a um conjunto de recursos computacionais configuráveis (por exemplo, redes, servidores, armazenamento, aplicações e serviços) que podem ser rapidamente adquiridos e liberados com mínimo esforço gerencial ou interação com o provedor de serviços. (NIST, 2009)

A computação em nuvem pode ser descrita por cinco características essenciais, três modelos de serviço e quatro modelos de implantação, detalhados a seguir (Sousa, F. et al., 2010).

2.1.2 Características essenciais

As características essenciais são vantagens que as soluções de computação em nuvem oferecem. Algumas dessas características, em conjunto, definem exclusivamente a computação em nuvem e fazem a distinção com outros paradigmas. Por exemplo, a elasticidade rápida de recursos e o amplo acesso e medição de serviço são características básicas para compor uma solução de computação em nuvem.

2.1.2.1 Self-service sob demanda

O usuário pode adquirir unilateralmente recurso computacional, como tempo de processamento no servidor ou armazenamento na rede, à medida que necessite e sem precisar de interação humana com os provedores de cada serviço. O *hardware* e o *software* dentro de uma nuvem podem ser automaticamente reconfigurados e orquestrados, e essas modificações são apresentadas de forma transparente para os usuários, que possuem perfis diferentes e assim

podem personalizar os seus ambientes computacionais – por exemplo, com a instalação de *software* e configuração de rede para a definição de determinados privilégios.

2.1.2.2 *Ampla acesso*

Recursos são disponibilizados por meio da rede e acessados através de mecanismos padronizados que possibilitam o uso por plataformas do tipo *thin*, tais como celulares, *laptops* e *Personal Digital Assistants* (PDAs). A interface de acesso à nuvem não obriga os usuários a mudar suas condições e ambientes de trabalho, como, por exemplo, linguagens de programação e sistema operacional. Os sistemas de *software* instalados localmente para o acesso à nuvem são leves como um navegador de Internet.

2.1.2.3 *Pooling de recursos*

Os recursos computacionais do provedor são organizados em um *pool* para servir a múltiplos usuários usando um modelo multi-inquilino (*multi-tenant*), com diferentes recursos físicos e virtuais, dinamicamente atribuídos e ajustados de acordo com a demanda dos usuários. Esses usuários não precisam ter conhecimento da localização física dos recursos computacionais, podendo somente especificar a localização em um nível mais alto de abstração, tais como o país, estado ou *datacenter*.

2.1.2.4 *Elasticidade rápida*

Recursos podem ser adquiridos de forma rápida e elástica, em alguns casos automaticamente, caso haja a necessidade de escalar com o aumento da demanda, e liberados, na retração dessa demanda. Para os usuários, os recursos disponíveis para uso parecem ser ilimitados e podem ser adquiridos em qualquer quantidade e a qualquer momento.

Entende-se por elasticidade a “capacidade de adicionar e remover recursos de forma automática de acordo com a carga de trabalho sem interrupções e utilizando os recursos de forma otimizada” (Coutinho, E. et al., 2013). Para implementar a elasticidade, as principais formas utilizadas são técnicas de replicação e redimensionamento de recursos.

2.1.2.4.1 *Replicação de recursos*

Utilizada para programar a elasticidade por meio do ajuste automático da quantidade de número de réplicas para a carga de trabalho atual, ou seja, a adição de novas réplicas se o sistema não consegue lidar com a carga de trabalho atual, ou remoção de réplicas desnecessárias.

2.1.2.4.2 Redimensionamento de recursos

O redimensionamento permite ajustar a quantidade de recursos de acordo com a carga de trabalho. Por exemplo, pode-se incrementar ou decrementar a quantidade de CPU nas máquinas virtuais de forma automática, garantindo a qualidade e reduzindo os custos.

2.1.2.5 Serviço medido

Sistemas em nuvem automaticamente controlam e aperfeiçoam o uso de recursos por meio de uma capacidade de medição. A automação é realizada em algum nível de abstração apropriado para o tipo de serviço, tais como armazenamento, processamento, largura de banda e contas de usuário ativas. O uso de recursos pode ser monitorado e controlado, possibilitando transparência para o provedor e o usuário do serviço utilizado.

Para garantir a qualidade do serviço, pode-se utilizar a abordagem baseada em SLA. O acordo de nível de serviço fornece informações sobre os níveis de disponibilidade, funcionalidade, desempenho ou outros atributos do serviço, como o faturamento e até mesmo penalidades em caso de violação desses níveis.

2.1.3 Modelos de serviços

O ambiente de computação em nuvem é composto de três modelos de serviços. Esses modelos são importantes, pois definem um padrão arquitetural para soluções de computação em nuvem.

2.1.3.1 Software como um Serviço (SaaS)

O modelo de SaaS proporciona sistemas de *software* com propósitos específicos que são disponíveis para os usuários por meio da Internet e acessíveis a partir de vários dispositivos do usuário por meio de uma interface *thin client* como um navegador Web. Como exemplos de SaaS podem-se destacar os serviços de *Customer Relationship Management* (CRM) da *Salesforce* e o *Google Docs*.

2.1.3.2 Plataforma como um Serviço (PaaS)

O modelo de PaaS fornece ferramentas de desenvolvimento para aplicações, auxiliando a implementação de sistemas de *software*. Nesse modelo, o usuário tem controle sobre as aplicações implantadas, incluindo sistemas de banco de dados e servidores web, assim como as

configurações de aplicações hospedadas em uma infraestrutura. *Google App Engine* e *Force.com* são exemplos de PaaS.

2.1.3.3 Infraestrutura como um Serviço (IaaS)

A IaaS fornece recursos, tais como máquinas virtuais, servidores, rede, armazenamento e outros recursos de computação fundamentais, para construir um ambiente de aplicação sob demanda, que pode incluir sistemas operacionais e aplicativos. O usuário tem controle sobre os sistemas operacionais, armazenamento, aplicativos implantados e, eventualmente, seleciona componentes de rede, tais como *firewalls*. O *Amazon Elastic Cloud Computing (EC2)*, assim como as máquinas virtuais da *Microsoft Azure* são exemplos de IaaS.

2.1.4 Modelos de implantação

Quanto ao acesso e à disponibilidade, há diferentes tipos de modelos de implantação para os ambientes de computação em nuvem. A restrição ou abertura de acesso depende do processo de negócios, do tipo de informação e do nível de visão desejado. Pode-se perceber que determinadas empresas não desejam que todos os usuários possam acessar e utilizar determinados recursos no seu ambiente de computação em nuvem. Nesse sentido, surge a necessidade de ambientes mais restritos, em que somente alguns usuários devidamente autorizados possam utilizar os serviços providos. Os modelos de implantação da computação em nuvem podem ser divididos em nuvem pública, privada, comunidade e híbrida.

2.1.4.1 Nuvem privada

A infraestrutura de nuvem é utilizada exclusivamente por uma organização, sendo essa nuvem local ou remota e administrada pela própria empresa ou por terceiros.

2.1.4.2 Nuvem pública

A infraestrutura de nuvem é disponibilizada para o público em geral, sendo acessada por qualquer usuário que conheça a localização do serviço.

2.1.4.3 Nuvem comunidade

Fornecer uma infraestrutura compartilhada por uma comunidade de organizações com interesses em comum.

2.1.4.4 Nuvem híbrida

A infraestrutura é uma composição de duas ou mais nuvens, que podem ser do tipo privada, pública ou comunidade e que continuam a ser entidades únicas, mas conectadas por meio de tecnologia proprietária ou padronizada que permite a portabilidade de dados e aplicações.

Cada modelo de implantação deve tratar de questões relativas à segurança no ambiente de computação em nuvem. De forma geral, técnicas de criptografia de dados são utilizadas para garantir a privacidade dos dados contra acessos não autorizados. No entanto, as técnicas de criptografia têm implicações significativas no acesso aos dados, visto que existe uma complexidade computacional associada para as operações de criptografar e descriptografar os dados. Dessa forma, técnicas de criptografia de dados e seus respectivos custos e benefícios devem ser investigados.

2.1.5 Modelos de precificação e modalidades de aquisição

Cada provedor de serviços de nuvem tem um modelo próprio para cobrança pelo uso do serviço e, dependendo do modelo (IaaS, PaaS, SaaS), o uso é medido em termos de infraestrutura, plataforma ou *software*. Por exemplo, no modelo SaaS, é muito comum a cobrança pela quantidade de usuários que acessam o *software*. Já no modelo IaaS, pode-se cobrar pela quantidade de disco (Gb) utilizado, ou pelo tempo (horas ou minutos) que uma máquina virtual permanece ligada.

Ainda sobre a cobrança de uso de máquinas virtuais, existem variações sobre a forma de contratação do serviço. Por exemplo, a *Amazon Web Services (AWS)*, que é um dos líderes de mercado no fornecimento de IaaS, atualmente oferece três opções de compra, a seguir descritas.

2.1.5.1 Por demanda (on-demand)

Nessa opção, a cobrança é por cada hora que a máquina virtual (também chamada de instância) permanece ligada. Não há investimento inicial nem compromisso de uso. É a forma mais simples de usar e pagar, mas em geral tem o custo por hora mais alto.

2.1.5.2 Reservada

Desde dezembro de 2014, essa opção propicia que os clientes paguem pelo período reservado, ao invés de por horas de uso. Existem duas opções de períodos para aquisição junto à AWS: anual ou três anos. Podem existir também outras opções de período, oferecidas por clientes que

compraram da AWS o período padrão e revendem uma parte dele. As opções de pagamento são: totalmente adiantado, parcialmente adiantado ou nada adiantado – neste último caso, os pagamentos são mensais.

A tabela 1 ilustra os preços para a região leste americana (US-East) e para instâncias de tamanho *m3.medium*. AWS apresenta um custo estimado da hora, para facilitar a comparação com a opção *on-demand*, no entanto, o pagamento nessa opção é sempre para o período completo (um ou três anos), podendo ser feito adiantado ou em parcelas mensais. Dessa forma, quando se adquire uma instância reservada por um ano, não faz diferença em termos de custo, se ela estiver ligada ou não. O custo das instâncias reservadas pode ser equivalente à operação local de servidores, como apresentado em estudo sobre custos (E-fiscal, 2013).

Tabela 1 - Preços de Instâncias reservadas em 2015 (tamanho *m3.medium*)

PERÍODO DE 1 ANO					
Opção de pagamento	Inicial	Mensal*	Efetiva por hora**	Economia em relação ao <i>on-demand</i>	<i>On-demand</i> por hora
Sem taxas iniciais	\$0	\$35.04	\$0.038	28%	\$0.067 por hora
Adiantado parcial	\$211	\$12.41	\$0.0411	39%	
Adiantado integral	\$353	\$0.00	\$0.0403	40%	
PERÍODO DE 3 ANOS					
Opção de pagamento	Inicial	Mensal*	Efetiva por hora**	Economia em relação ao <i>on-demand</i>	<i>On-demand</i> por hora
Adiantado parcial	\$337	\$10.95	\$0.0278	60%	\$0.067 por hora
Adiantado integral	\$687	\$0.00	\$0.0261	63%	

Fonte: Amazon Web Services (AWS - Instâncias Reservadas, 2015)

Por exemplo, os custos por hora de uma instância média (*m3.medium*) com sistema operacional Linux na *Amazon* variam de US\$ 0.026 centavos até US\$ 0.067 centavos, dependendo do tipo de reserva, conforme ilustrado na tabela 2. No exemplo, foi utilizada a calculadora da *Amazon* que indica o melhor tipo de contratação de acordo com a estimativa de uso. Foi estimado uso de 50% da instância, ou seja, 12 horas por dia. As opções apresentadas foram:

Por demanda (sem contrato): nesse caso, o custo seria US\$ 0.067 por hora. Considerando 12 horas por dia, em média de 30,5 dias por mês, o valor mensal seria de 366 horas x US\$ 0.067 = US\$ 24.53. Multiplicando-se esse resultado por 12 meses ou 36 meses, têm-se os valores estimados de US\$ 294.36 para um ano e US\$ 883.08 para três anos, respectivamente.

Tabela 2 - Custos de instâncias reservadas em 2015 (uso estimado 50% para m3.medium com Linux)

Modalidade de compra	Investimento inicial	Mensal fixo	Custo efetivo da hora	Custo efetivo mensal	Custo por 1 ano	Custo por 3 anos
Por demanda (sem contrato)	---	---	0.067	24.53	294.36	883.08
Reserva por 1 ano sem pagamento adiantado	0.00	35.04	0.048	35.04	420.48	1,261.44
Reserva por 1 ano com pagamento adiantado parcial	211.00	12.41	0.041	30.00	359.92	1,079.76
Reserva por 1 ano com pagamento adiantado integral	353.00	---	0.040	29.42	353.00	1,059.00
Reserva por 3 anos com pagamento adiantado parcial	337.00	10.95	0.028	20.32	---	731.20
Reserva por 3 anos com pagamento adiantado integral	687.00	---	0.026	19.09	---	687.00

Fonte: (Amazon Web Services - Calculator, 2015)

Reserva por um ano sem pagamento adiantado: nesse caso, o valor mensal (fixo) seria de US\$ 35, podendo ser utilizado por todo o mês (média de 732 horas no mês). O valor da hora fica em US\$ $35/732$ horas = US\$ 0.048. O valor total por um ano fica em US\$ 420.48, e para três anos US\$ 1,261.44.

Reserva por um ano com pagamento adiantado parcial: nesse caso, o custo mensal (fixo) seria de US\$ 12.41, porém existe um investimento inicial de US\$ 211.00. O valor mensal fica em $(US\$ 211/12 + US\$ 12.41) = US\$ 30$ e o valor por hora, $US\$ 30/732$ horas = US\$ 0.041. O valor total por um ano fica em US\$ 359.92, e para 3 anos US\$ 1,079.76.

Reserva por um ano com pagamento adiantado integral: nesse caso, não há custo mensal fixo, porém existe um investimento inicial de US\$ 353.00. O valor mensal fica em $(US\$ 353/12) = US\$ 29.42$ e o valor por hora, $US\$ 29.42/732$ horas = US\$ 0.040. O valor total por um ano fica em US\$ 353, e para três anos US\$ 1,059.

Reserva por três anos com pagamento adiantado parcial: nesse caso, o custo mensal (fixo) seria de US\$ 10.95, porém existe um investimento inicial de US\$ 337. O valor mensal fica em $(US\$ 337/36 + US\$ 10.95) = US\$ 20.32$ e o valor por hora, $US\$ 20.32/732$ horas = US\$ 0.028. O valor total por três anos é de US\$ 731.2.

Reserva por três anos com pagamento adiantado integral: nesse caso, não há custo mensal fixo, porém existe um investimento inicial de US\$ 687. O valor mensal fica em $(US\$ 687/36) = US\$ 19.09$ e o valor por hora, $US\$ 19.09/732$ horas = US\$ 0.026. O valor total por três anos é de US\$ 687.

Note que a modalidade com reserva oferece custos menores, porém existe um comprometimento de 1 a 3 anos. Nesse período, mesmo que o cliente desista do uso da máquina, ele terá que continuar pagando o valor mensal ou perderá o investimento inicial. Para amenizar

esse problema, a AWS permite que os clientes repassem suas máquinas reservadas para outros clientes, negociando o valor da reserva. Por exemplo, a figura 2 apresenta dois exemplos em que terceiros (*3rd party*) ofertam instâncias reservadas:

- Caso 1: por 20 meses, com investimento inicial de US\$ 60 e custo por hora US\$ 0.050. Esse era o modelo de venda de reserva da AWS até novembro de 2014. Nesse modelo, existia um investimento inicial e também um custo por hora, em geral menor do que o custo sem reserva.
- Caso 2: por 8 meses, com investimento inicial de US\$ 235. Possivelmente, esse cliente adquiriu uma instância na modalidade “Reserva por 1 ano com pagamento adiantado integral”, utilizou durante 4 meses e não pretende utilizar os restantes 8 meses do período.

Figura 2 - Mercado de revenda de instâncias reservadas

Purchase Reserved Instances										
Platform		Linux/UNIX		Availability Zone		Any		Tenancy		Default
Instance Type		m3.medium		Term		Any		Offering		Any
Seller	Term	Effective Rate	Upfront Price	Hourly Rate	Availability Zone	Offering	Quantity Available	Desired Quantity		
AWS	12 months	\$0.048	\$0.00	\$0.048	us-east-1d	No Upfront	Unlimited	1	Add to Cart	
AWS	12 months	\$0.048	\$0.00	\$0.048	us-east-1b	No Upfront	Unlimited	1	Add to Cart	
AWS	12 months	\$0.048	\$0.00	\$0.048	us-east-1e	No Upfront	Unlimited	1	Add to Cart	
AWS	12 months	\$0.048	\$0.00	\$0.048	us-east-1a	No Upfront	Unlimited	1	Add to Cart	
3rd Party	20 months	\$0.054	\$60.00	\$0.050	us-east-1a	Light Utilization	1	1	Add to Cart	
AWS	12 months	\$0.041	\$211.00	\$0.017	us-east-1d	Partial Upfront	Unlimited	1	Add to Cart	
AWS	12 months	\$0.041	\$211.00	\$0.017	us-east-1b	Partial Upfront	Unlimited	1	Add to Cart	
AWS	12 months	\$0.041	\$211.00	\$0.017	us-east-1e	Partial Upfront	Unlimited	1	Add to Cart	
AWS	12 months	\$0.041	\$211.00	\$0.017	us-east-1a	Partial Upfront	Unlimited	1	Add to Cart	
3rd Party	8 months	\$0.041	\$235.00	\$0.000	us-east-1a	All Upfront	1	1	Add to Cart	
AWS	36 months	\$0.028	\$337.00	\$0.015	us-east-1d	Partial Upfront	Unlimited	1	Add to Cart	

Fonte: (AWS - Console, 2015)

2.1.5.3 Spot

Nessa opção, a cobrança é por horas de uso, de modo similar à opção *on-demand*, no entanto, o valor da hora não é fixo. Os clientes registram lances do valor que eles concordam em pagar pela hora de uso. A AWS define dinamicamente o preço *spot*, que varia em tempo real baseado em oferta e demanda. Se o valor do lance do cliente é maior do que o preço *spot* atual, a instância é ligada. Se o preço *spot* é alterado, e fica acima do valor do lance, a instância é desligada automaticamente por AWS. A última hora de uso, quando incompleta, não é cobrada do cliente quando a instância é desligada pela AWS devido à mudança do preço *spot*.

2.1.5 Adoção de serviços em nuvem

O modelo de computação em nuvem permite às empresas uma nova opção em seu planejamento de disponibilização de novos serviços de TI, internos ou externos. Não é mais necessário adquirir a infraestrutura física de computadores, redes, armazenamento de dados para hospedar o seu serviço, a menos que se pretenda criar uma nuvem privada. É possível agora simplesmente contratar serviços em nuvem que fornecem infraestruturas similares para que os técnicos de TI da empresa hospedem os aplicativos (usando modelo IaaS), ou ainda mais simples, contratar o próprio serviço, já pronto para uso direto pelos usuários finais, sem necessidade de administração interna do serviço pela TI (modelo SaaS).

No entanto, existem algumas barreiras que dificultam a adoção dos serviços pelas empresas, entre elas: incerteza quanto à segurança e privacidade dos dados, bem como quanto a mudanças nos modelos econômicos adotados pelo fornecedor. Por exemplo, a *Amazon* anunciou em 2014 a 42ª redução de preços de seus serviços, porém, no final do ano, alterou significativamente o modelo de precificação da opção de compra reservada de máquinas virtuais. Em 2015, com a alta do dólar, os contratos celebrados nessa moeda sofreram reajustes automáticos que tornaram os serviços bem menos vantajosos economicamente. Outra barreira importante a ser considerada é a conectividade existente no local. Serviços em nuvem demandam uma conexão à internet de baixa latência, que nem sempre está disponível em locais muito remotos.

Os serviços em nuvem possuem alguns diferenciais importantes (Avram, 2014), como a seguir descritos.

2.1.5.1 Redução de custos

A utilização dos serviços em nuvem reduz drasticamente o custo de entrada para as pequenas empresas, que podem agora se beneficiar de aplicações de computação intensiva anteriormente disponíveis apenas em corporações de grande porte. Esses exercícios computacionais tipicamente envolvem grandes quantidades de poder computacional para períodos relativamente curtos de tempo. A computação em nuvem também representa uma enorme oportunidade para muitos países do terceiro mundo que até agora têm sido deixados para trás na revolução de TI – alguns provedores de serviços disponibilizam serviços em países que normalmente não possuem recursos para a implantação generalizada dos serviços de TI.

2.1.5.2 Baixo ou nenhum investimento inicial

O modelo em nuvem pode fornecer um acesso quase imediato a recursos de *hardware*, sem investimento de capital inicial para os usuários, permitindo uma entrada mais rápida no mercado para muitas empresas. Tratando-o como uma despesa operacional (na indústria, *Opex* em oposição a um modelo *Capex*), isso também ajuda a reduzir drasticamente os custos iniciais em computação corporativa. A nuvem torna-se uma infraestrutura adaptável que pode ser compartilhada por diferentes usuários finais, e cada um dos quais pode usá-la de maneira muito diferente.

2.1.5.3 Inovação

A computação em nuvem pode reduzir as barreiras de TI para inovação, desde aplicações *online* onipresentes, como *Facebook* e *Youtube*, até as aplicações mais centradas como *TripIT.com* (para o gerenciamento de viagens) ou *Mint* (para gerenciar finanças pessoais).

2.1.5.4 Escalabilidade

A computação em nuvem torna mais fácil para as empresas expandir os seus serviços – cada vez mais dependentes de informações precisas – de acordo com a demanda do cliente. Uma vez que os recursos computacionais são geridos através de *software*, esses podem ser provisionados rapidamente de acordo com a necessidade. Na verdade, um dos objetivos da computação em nuvem é escalar os recursos para cima ou para baixo de forma dinâmica através de *APIs* de *software*, dependendo da carga de utilização do cliente, com interação mínima com o provedor.

2.1.5.5 Viabilidade de novas classes de serviços

A computação em nuvem também possibilita novas classes de aplicações que não eram possíveis antes. Exemplos incluem: (a) aplicações móveis interativas sensíveis a contexto, localização e ambiente, que respondem em tempo real às informações fornecidas pelos usuários humanos e sensores não-humanos (por exemplo, sensores de tensão e umidade dentro de um contêiner de transporte); (b) processamento em lote paralelo, que permite aos usuários tirar vantagem de enormes quantidades de processamento para analisar terabytes de dados em relativamente curtos períodos de tempo, como os serviços que implementam o modelo de programação paralela *MapReduce*; (c) análise de negócios que podem utilizar uma vasta quantidade de recursos computacionais para entender os clientes, hábitos de compra, cadeias de abastecimento; e (d) extensões dos aplicativos de *desktop* de computação intensiva que

podem enviar dados para processamento na nuvem, deixando apenas a apresentação dos dados processados no cliente.

2.1.6 Migração de serviços para a nuvem

A migração dos serviços existentes na infraestrutura local da empresa é um problema um pouco diferente da adoção. Apesar de possuírem vários elementos em comum (já que, para migrar um serviço, inicialmente é necessário existir a intenção de adotar o modelo em nuvem), a migração apresenta outras características a ser analisadas, por exemplo, o serviço não é novo, e os usuários já estão acostumados com um determinado nível de serviço. A nuvem, apesar de fornecer inúmeros benefícios, pode não ser adequada a todo tipo de serviço de TI – por exemplo, a aplicativos pequenos que precisam estar disponíveis 100% do tempo, em locais onde a conectividade da internet é pouco confiável.

Discernir entre aplicações legadas adequadas para migração para a nuvem não é um problema simples. Fatores como custo, desempenho, segurança e preocupações legais precisam ser avaliados. Limitações técnicas que podem impedir a adoção da nuvem pela organização também podem existir, por exemplo, quando os aplicativos legados violam as restrições técnicas ambientais impostas pelo provedor de nuvem. Existem *frameworks* específicos para apoiar esse tipo de decisão, como o descrito em Beserra, P. et al. (2012), em que os autores propõem alguns passos necessários para avaliação: 1) definir o perfil da organização; 2) avaliar as restrições da organização; 3) definir o perfil da aplicação a ser migrada; 4) definir o perfil do provedor de serviços de nuvem; 5) avaliar restrições técnicas e/ou financeiras; e, quando necessário, alterar o provedor de serviços de nuvem para encontrar o que seja mais adequado à aplicação em estudo.

Existem vários tipos de migração de serviços para nuvem, conforme descritos em Jamshidi, Ahmad, & Pahl (2013) e a seguir comentados.

2.1.6.1 Tipo I: Substituição

Substituir cada camada do aplicativo com serviços de nuvem equivalentes é o tipo de migração menos invasivo. As camadas de dados e/ou regras de negócios são migradas para a pilha de nuvem. Esse tipo de migração é considerado mais como uma melhoria na aplicação legada, ao invés da habilitação do serviço na nuvem propriamente dito. Esse tipo de migração requer uma série de reconfigurações para ajustar incompatibilidades em cada camada migrada.

2.1.6.2 Tipo II: Migração parcial

Nesse caso, somente alguns componentes do *software* são migrados para a nuvem. Como um exemplo, a migração da funcionalidade de auditoria de um sistema de saúde para a nuvem pode ser classificada como um tipo II de migração.

2.1.6.3. Tipo III: Migração de toda a pilha de aplicativos

É o caminho mais fácil de migração, em que toda a aplicação é monoliticamente encapsulada em uma ou mais máquinas virtuais rodando na nuvem. Esse tipo de migração não precisa de adaptação, assumindo que a pilha da aplicação pode ser portada "como está" em uma máquina virtual.

2.1.6.4 Tipo IV: Cloudify

É o tipo de migração mais completo, em que a aplicação é convertida para um sistema completo formatado para nuvem, através da composição de vários serviços de nuvem especializados. Os tipos IV e II são semelhantes, mas o âmbito da migração do tipo II está limitado a algumas funções. O tipo de migração IV é o mais invasivo, pois precisa de adaptação ao nível de composição dos processos de negócios e de suas infraestruturas correspondentes.

O problema de decidir entre como e quando migrar uma aplicação existente para a nuvem não é trivial. O trabalho de Saripalli & Pingali (2011) propõe um sistema de apoio à decisão para avaliar os diversos fatores envolvidos nesse tipo de decisão, tais como:

- 1) Adequabilidade: funcionalidades, compatibilidade com legado, qualidade da rede – latência e largura de banda;
- 2) Valor econômico: custo do capital, custo de operação, retorno sobre o investimento – ROI;
- 3) Controle: facilidade de integração, gerenciamento – facilidade de monitoramento, autonomia e adaptabilidade das *APIs*;
- 4) Usabilidade: tempo para inicialização da aplicação, agilidade na exibição gráfica, simplicidade;
- 5) Confiabilidade: elasticidade, alta disponibilidade, recuperação de desastres, observância de requisitos, confiança;
- 6) Segurança: confidencialidade, integridade, disponibilidade, facilidade de auditoria, isolamento de multi-inquilinos.

2.2 Redes de Petri

As redes de Petri (RP) foram originalmente descritas por Carl Adam Petri em sua tese de doutorado intitulada **Comunicação entre Autômatos**, defendida em 1962 (Petri, 1962). A partir de então, consideráveis trabalhos teóricos e aplicações práticas com RP têm sido

realizados, principalmente nas áreas de modelagem (*software* e *hardware*), redes de computadores, comunicações, sistemas distribuídos, protocolos de comunicação, sistemas operacionais, sistemas de controle de produção, automação industrial, modelagem de controladores lógicos, análise de fluxo de tarefas, *chips* VLSI, modelagem de sistemas a eventos discretos (Jensen, 2009).

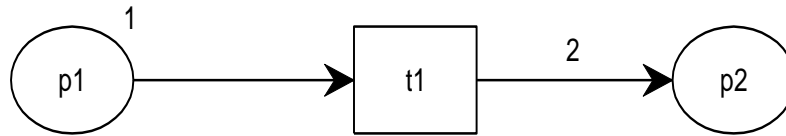
A rede de Petri é uma ferramenta gráfica e matemática que se adapta bem a um grande número de aplicações em que as noções de eventos e evoluções simultâneas são importantes (Cardoso & Valette, 1997).

Uma RP é um tipo de grafo bipartido e direcionado, em que os arcos nunca ligam dois nós do mesmo tipo. A Figura 3 apresenta os elementos gráficos que compõem uma rede de Petri. Nela, p_1 e p_2 são os lugares e t_1 é a transição. Nesse caso, p_1 é um lugar de entrada de t_1 e p_2 é um lugar de saída de t_1 . O arco que liga p_1 a t_1 possui peso 1. O arco que liga t_1 a p_2 possui peso 2. O lugar p_1 possui uma ficha, representada na figura pelo valor “1” próximo ao lugar.

Nas RPs, a ocorrência de um evento está associada ao disparo de uma transição e os lugares de entrada e saída da transição representam, respectivamente, as pré-condições e pós-condições associadas à ocorrência do evento. Os arcos de entrada de uma transição t_j têm origem em um ou mais lugares de entrada p_i de t_j e terminam na transição t_j ; os arcos de saída têm origem na transição t_j e terminam em um ou mais lugares de saída p_i de t_j . Conforme a Figura 3, o lugar p_1 é entrada de t_1 , visto que um arco se origina em p_1 e termina em t_1 . O lugar p_2 é saída de t_1 , visto que um arco de peso 2 se origina em t_1 e termina em p_2 .

As fichas são usadas nas RPs para simular a dinâmica e as atividades concorrentes do sistema. O estado de uma RP é representado por um número k_i de fichas contidas em cada lugar p_i , chamada marcação, conforme apresentado na Figura 3. O estado do sistema é dado pela distribuição de fichas nos lugares da RP e cada lugar representa um estado parcial do sistema. A mudança de estado é representada pelo movimento de fichas na RP, que acontece quando ocorre o disparo de transições. Cada evento que ocorre no sistema é associado ao disparo de uma transição no modelo RP. O disparo de uma transição significa que o seu evento correspondente ocorreu.

Figura 3 – Ilustração de uma RP



Fonte: (Barroso, [s.n])

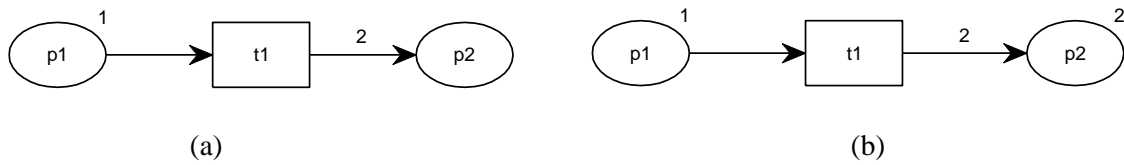
Uma transição é dita habilitada se cada lugar de entrada da transição contém um número de fichas maior ou igual ao peso do arco que o conecta à transição. Uma transição habilitada pode ou não disparar. Quando ocorre o disparo de uma transição, fichas são removidas dos lugares de entrada da transição e fichas são adicionadas aos lugares de saída. A quantidade de fichas removidas e acrescentadas depende do peso do arco. A nova marcação resultante do disparo da transição representa o novo estado do sistema. A marcação inicial M_0 representa o estado inicial da RP.

O comportamento dinâmico das RPs obedece à regra de disparo de transições, a saber:

- uma transição é dita habilitada se cada lugar de entrada p de t é marcado com pelo menos $w(p, t)$ fichas, em que $w(p, t)$ é o peso do arco de p para t ;
- uma transição habilitada pode ou não disparar;
- o disparo de uma transição t remove $w(p, t)$ fichas de cada lugar de entrada p de t , e adiciona $w(t, p)$ fichas a cada lugar de saída p de t , em que $w(t, p)$ é o peso do arco direcionado de t para p .

A RP apresentada na Figura 4 ilustra a regra de disparo através da modelagem do comportamento dinâmico de um sistema e sua evolução. A Figura 4(a) apresenta o estado inicial do sistema. A mudança de estado, que acontece através do disparo da transição t_1 , é apresentada na Figura 4(b). Na Figura 4(a), existem duas fichas no lugar de entrada p_1 e nenhuma ficha no lugar de saída p_2 . A marcação da rede é $M_0 = (2, 0)$. Nesta marcação a transição t_1 está habilitada e pode disparar. Conforme apresentado na Figura 4(b), no disparo da transição t_1 , fichas são removidas do lugar de entrada p_1 e fichas são adicionadas ao lugar de saída p_2 , originando uma nova marcação ou estado do sistema $M_1 = (1, 2)$. Como pode ser observado nesse exemplo, a quantidade de fichas removidas do lugar de entrada e adicionada ao lugar de saída depende do peso dos arcos.

Figura 4 - Representação do disparo da transição t1. (a) estado inicial da RP antes do disparo, (b) estado da RP após o disparo



Fonte: (Barroso, [s.n])

Formalmente, a estrutura de uma rede de Petri C é uma 4-tupla $C = (P, T, I, O)$, onde:

$P = \{p_1, p_2, \dots, p_n\}$ é um conjunto finito de lugares, $n \geq 0$.

$T = \{t_1, t_2, \dots, t_m\}$ é um conjunto finito de transições, $m \geq 0$.

$I: T \rightarrow P^\infty$ é a função entrada, um mapeamento de transições para multiconjuntos de lugares (multiconjuntos são conjuntos que permitem várias instâncias de um mesmo elemento)

$O: T \rightarrow P^\infty$ é a função saída, outro mapeamento de transições para multiconjuntos de lugares

A Figura 5 apresenta um exemplo da estrutura formal de uma rede de Petri.

Figura 5 - Exemplo da estrutura de uma rede de Petri

$$\begin{aligned}
 C &= (P, T, I, O) \\
 P &= \{p_1, p_2, p_3, p_4, p_5\} \\
 T &= \{t_1, t_2, t_3, t_4\} \\
 I(t_1) &= \{p_1\} & O(t_1) &= \{p_2, p_3, p_5\} \\
 I(t_2) &= \{p_2, p_3, p_5\} & O(t_2) &= \{p_5\} \\
 I(t_3) &= \{p_3\} & O(t_3) &= \{p_4\} \\
 I(t_4) &= \{p_4\} & O(t_4) &= \{p_2, p_3\}
 \end{aligned}$$

Fonte: (Peterson, 1981)

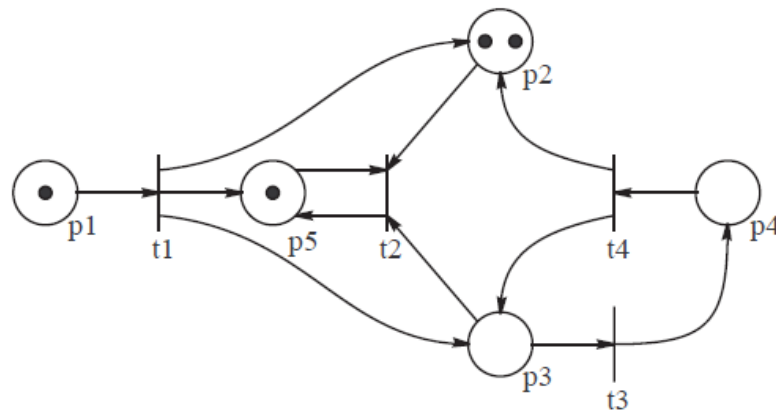
O conjunto de lugares e o conjunto de transições são disjuntos, ou seja, $P \cap T = \emptyset$. A representação gráfica da estrutura de uma rede de Petri é muito útil para ilustrar os conceitos da teoria de redes de Petri. Um grafo de rede de Petri é um multigrafo bipartido direcionado $G = (V, A)$, onde

$V = \{v_1, v_2, \dots, v_s\}$ é um conjunto de vértices e

$A = \{a_1, a_2, \dots, a_r\}$ é um multiconjunto de arcos direcionados $a_i = (v_j, v_k)$, com $v_j, v_k \in V$.

O conjunto V pode ser particionado em dois subconjuntos distintos P e T tais que $V = P \cup T$, $P \cap T = \emptyset$, e para cada arco direcionado $a_i \in A$, se $a_i = (v_j, v_k)$, então ou $v_j \in P$ e $v_k \in T$ ou $v_j \in T$ e $v_k \in P$. Uma marcação μ de uma rede de Petri $C = (P, T, I, O)$ é uma função $\mu : P \rightarrow \mathbb{N}$ do conjunto de lugares P para os inteiros não negativos \mathbb{N} . (Peterson, 1981). A Figura 6 apresenta o grafo correspondente à estrutura da Figura 5, com a marcação $(1,2,0,0,1)$.

Figura 6 - Grafo equivalente à estrutura da Figura 5, com marcações



Fonte: (Peterson, 1981)

2.2.1 Propriedades e métodos de análise de redes de Petri

A aplicação de RPs na modelagem de sistemas tem a vantagem de permitir verificar as propriedades dos modelos construídos através dos métodos formais de análise, a saber: Árvore (Grafo) de Alcançabilidade ou Cobertura, Matriz de Incidência e Equação de Estado e Técnicas de Redução e Decomposição (Murata, 1989); (Jensen & Kristensen, 2009). A análise das propriedades das RPs com esses métodos pode revelar informações importantes sobre a estrutura e comportamento do sistema modelado, permitindo ao projetista realizar modificações e as correções antes da implementação. Algumas dessas propriedades são:

- Vivacidade (*liveness*): permite saber se um sistema não possui bloqueio (*deadlock*) e se todos os elementos do sistema estão ativos;
- Reversibilidade ou reinicialização (*reversibility*): permite saber se o sistema é capaz de sempre retornar ao estado inicial após uma tarefa realizada;
- Limitação (*boundedness*): permite verificar a consistência de um sistema quanto aos limites de sua capacidade de armazenamento e de realização de tarefas.

Em geral, os sistemas do mundo real são complexos e possuem vários processos com características similares, mas não idênticas. As RPs possuem apenas um tipo de ficha, que

pode ser inteiro ou booleano. O fato das RPs não manipularem tipos de dados diferentes dificulta a modelagem de sistemas reais e complexos (Jensen & Kristensen, 2009).

Para modelar sistemas reais, muitas vezes é necessário construir várias sub-redes independentes com estruturas basicamente idênticas para processos similares. Isso pode tornar o modelo RP extremamente grande, dificultando o desenvolvimento do projeto e a visualização dos modelos na sua totalidade. Além disso, pode ser difícil observar similaridades e diferenças entre as redes individuais que representam as partes similares. Outro fato é que as RPs não tratam de restrições de tempo, algo inerente aos sistemas reais.

Para contornar esses problemas foram desenvolvidas extensões de RP capazes de descrever sistemas mais complexos de forma mais compacta, entre as quais estão as Redes de Petri Coloridas (RPCs), que são redes de Petri de alto nível, e as redes de Petri com restrições de tempo (Jensen & Kristensen, 2009).

Este trabalho utiliza uma classe especial de redes de Petri denominada Redes de Petri Coloridas Generalizadas (RPCs). As redes de Petri coloridas receberam esta denominação porque, quando foram propostas, as mesmas utilizavam diferentes tipos de marcação especificados por números naturais e representados visualmente por cores: 1-vermelha, 2-azul, 3-verde, etc. As redes de Petri coloridas generalizadas utilizam como marcação um tipo de dado abstrato, como em uma linguagem de programação. O termo “colorida” é usado ainda por razões históricas.

2.2.2 Redes de Petri Coloridas

As redes de Petri Coloridas (RPCs) são uma linguagem gráfica para a construção de modelos de sistemas a eventos discretos e análise de suas propriedades. RPCs são uma linguagem de modelagem que combina as capacidades das redes de Petri com os recursos de uma linguagem de programação de alto nível (Jensen & Kristensen, 2009)

A linguagem de programação das RPC é a CPN ML, que se baseia na linguagem de programação funcional Standard ML (Ullman, 1998). As RPCs são destinadas ao uso prático, em especial porque elas permitem a construção de modelos compactos e paramétricos. A vantagem das RPCs sobre as outras redes de Petri é a capacidade de modelar sistemas complexos e fornecer modelos com um alto nível de abstração e de representação gráfica (Jensen & Kristensen, 2009).

O tempo desempenha um papel significativo numa vasta gama de sistemas concorrentes. O funcionamento correto de alguns sistemas depende crucialmente do tempo

tomado por certas atividades, e diferentes decisões de projeto podem ter um impacto significativo sobre o desempenho de um sistema. As RPCs incluem um conceito de tempo que faz com que seja possível capturar o tempo tomado por eventos no sistema. Isso significa que as RPCs podem ser aplicadas para a análise de desempenho baseada em simulação para investigar métricas, tais como atrasos, taxa de transferência e tamanho de fila, e para a modelagem e validação de sistemas de tempo real.

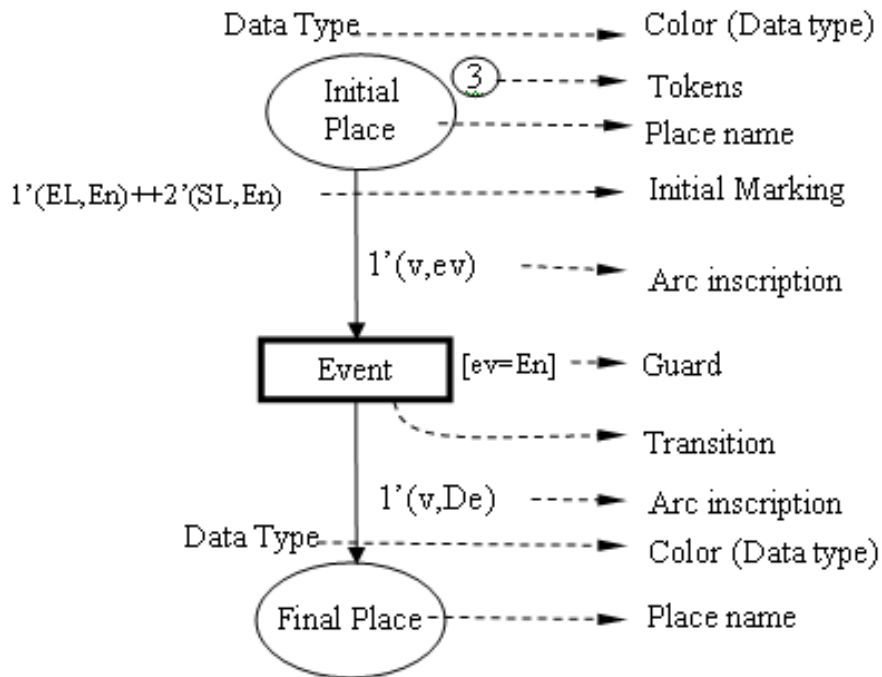
Modelos RPC podem conter diferentes níveis de abstração. Podem ser construídos em um conjunto de módulos, que são importantes quando lidamos com modelos de sistemas de grande porte. Os módulos interagem uns com os outros através de um conjunto de interfaces bem definidas, de uma forma semelhante às linguagens de programação. O conceito de módulos em RPC baseia-se num mecanismo de estruturação hierárquica, o que permite que um módulo tenha submódulos que podem ser reutilizados em diferentes partes do modelo.

RPC são ferramentas adequadas para a modelagem de jogos, pois fornecem recursos importantes, como a modularidade, facilidade de manutenção e capacidade de expansão. Essas características permitem a adição de novos processos ao modelo original, tornando possível agregar novas funções ou novos processos ao sistema.

O CPN Tools (CPN Group, 2015) é uma ferramenta para edição, simulação, análise de espaço de estado e análise de desempenho de modelos RPC. O usuário do CPN Tools trabalha diretamente com a representação gráfica do modelo RPC (Jensen & Kristensen, 2009). A ferramenta foi originalmente desenvolvida pelo grupo CPN na Universidade de Aarhus de 2000 a 2010, e atualmente é mantida pelo grupo AIS da Universidade de Tecnologia de Eindhoven, na Holanda.

A Figura 7 mostra a estrutura principal de um modelo RPC, constituído por lugares, transições e arcos. Com RPC, é possível usar os tipos de dados e manipulação de dados complexos.

Figura 7 - Representação de uma rede de Petri colorida (RPC)



Fonte: (Barroso, [s.n])

Os lugares representam o estado do sistema modelado. Os nomes dos lugares são escritos dentro das elipses e não têm nenhum significado oficial – eles têm, contudo, enorme importância prática para a legibilidade de um modelo RPC. Cada lugar pode ser marcado com uma ou mais fichas, e cada ficha tem um valor de dados ligado a ele. Esse valor de dados é chamado cor da ficha. É o número de fichas e as cores das fichas nos lugares individuais que, em conjunto, representam o estado do sistema. Isso é chamado de uma marcação do modelo RPC. As fichas em um lugar específico constituem a marcação daquele lugar.

As transições representam os eventos que podem ocorrer no sistema. Tal como acontece com os lugares, os nomes das transições são escritos dentro dos retângulos. Quando a transição ocorre, ele remove fichas de seus lugares de entrada (aqueles lugares que têm um arco que conduz à transição) e adiciona fichas aos seus lugares de saída (os lugares que têm um arco proveniente da transição). Quando ocorre uma transição, as fichas que são removidas dos lugares de entrada e adicionadas aos lugares de saída são determinadas por meio das expressões de arco, que são as inscrições textuais posicionadas ao lado dos arcos individuais. As expressões de arco são escritas na linguagem de programação CPN ML e são construídas a partir de variáveis, constantes, operadores e funções. Quando todas as variáveis em uma expressão são ligadas a valores do tipo correto, a expressão pode ser avaliada.

As transições possuem uma guarda, que é uma expressão booleana. Quando uma guarda está presente, ela deve ser avaliada como verdadeira para que a transição possa ser

habilitada, caso contrário, a transição estará desabilitada e não pode ocorrer. Assim, uma guarda coloca uma restrição adicional sobre a habilitação da transição.

Com um modelo RPC temporizado, medidas de desempenho, tais como comprimento máximo da fila e média de tempo de espera, podem ser calculadas. Podemos verificar, também, se a operação de um sistema em tempo real obedece aos prazos exigidos.

Em um modelo de RPC temporizado, as fichas podem transportar um selo de tempo, além de sua cor. Isso significa que a marcação de um lugar em que as fichas possuem selo de tempo é agora um multiconjunto temporizado. Além disso, o modelo RPC possui um relógio global que representa o tempo de modelo. A distribuição das fichas nos lugares, em conjunto com as suas temporizações e o valor do relógio global, é chamada de marcação temporizada. Mesmo num modelo RPC hierárquico temporizado há um único relógio global. O selo de tempo especifica o tempo em que a ficha está pronta para ser utilizada, ou removida por uma transição que ocorre.

Retardos de tempo podem ser implementados nos arcos e nas transições. Se houver um tempo de retardo em uma transição, a sua ocorrência adiciona o valor do retardo a todas as fichas de saída que carregam um selo de tempo. Se há um tempo de retardo associado a um arco de saída de uma transição, o retardo é adicionado apenas para as fichas temporizadas que são colocadas no lugar de saída associado a esse arco.

Neste trabalho, utilizamos a ferramenta CPN Tools pelos seguintes motivos:

- O modelo construído (apresentado no capítulo 4) possui uma formalização matemática;
- O modelo construído (apresentado no capítulo 4) tem uma visualização gráfica, facilitando sua apresentação aos gestores;
- A extensão do modelo construído (apresentada no capítulo 5) permite a simulação de eventos para validação das políticas de redução de custo propostas.

2.3 Métodos de Decisão Multicritério (MDMC)

Existem muitas abordagens para o problema de tomada de decisões utilizando diversos critérios, gerando uma diversidade de métodos. Tais métodos podem ser agrupados da seguinte forma (Triantaphyllou, 2000):

- Métodos baseados em *Muti-Attribute Utility Theory (Maut)*
 - soma ponderada (WSM)
 - produto ponderado (WPM)
 - processo de análise hierárquica (AHP)
- Métodos baseados em sobreposição (*outranking*).
 - *ELECTRE*
 - *TOPSIS*

2.3.1 Métodos baseados em Multi-Attribute Utility Theory (Maut)

Nessa categoria encontram-se o modelo de soma ponderada (WSM), o modelo de produto ponderado (WPM) e o processo de análise hierárquica (AHP). O método WSM é o mais comumente utilizado devido à sua simplicidade. No entanto, ele só deve ser usado quando se aplica a suposição de utilidade aditiva: não se deve somar maçãs e laranjas; somente critérios que utilizam as mesmas unidades e escala podem ser combinados. Isso limita severamente o método, embora muitos autores utilizem WSM sem considerar a restrição.

O método WPM combina critérios multiplicando proporções de métricas para várias alternativas que estão sendo comparadas. Essa divisão de métricas com a mesma unidade anula a unidade e, por essa razão, é chamada de análise adimensional.

A AHP decompõe o problema de decisão em uma hierarquia de critérios, alternativas e comparações entre pares para expressar a importância relativa de um critério em detrimento de outro. Usando essas comparações, é possível a construção de matrizes de pares e calcular o autovalor para calcular os critérios de classificação. A AHP pode combinar critérios qualitativos e quantitativos, e é largamente utilizada. O método é semelhante ao WPM na medida em que utiliza proporções de métricas. É, portanto, fácil de combinar critérios que utilizam diferentes unidades ou escalas, pois apenas os valores relativos são usados para comparar duas alternativas de acordo com um determinado critério. No entanto, difere da WPM em vários aspectos importantes: 1) trata facilmente hierarquias de critérios; 2) emprega uma escala de 9 pontos de fácil utilização e compreensão pelos tomadores de decisão; e 3) permite uma verificação a ser realizada para identificar comparações de pares inconsistentes entre alternativas.

2.3.2 Métodos baseados em sobreposição (outranking)

Nessa categoria destacam-se os métodos *ELECTRE* e *TOPSIS*. A sobreposição funciona da seguinte forma: alternativa A sobrepõe a alternativa B quando A é, no mínimo, tão boa quanto B em um grande número de critérios (condição de concordância), e o seu pior desempenho ainda é aceitável sobre os outros critérios (não condicionamento da discordância). Após a determinação, para cada par de alternativas, se uma alternativa é superior à outra, essas avaliações de pares são combinadas em um *ranking* parcial ou completo de alternativas. Uma alternativa é dita dominada se houver alternativa superior em um ou mais critérios e equivalente nos demais critérios. O método *ELECTRE* usa sobreposição. Com sobreposição, mesmo quando a alternativa A não domina a alternativa B quantitativamente, o tomador de decisão pode ainda optar por A como quase certamente melhor do que B. Com *TOPSIS*, o conceito

básico é que a alternativa escolhida deve ter a menor distância de uma solução ideal e a maior distância de uma solução negativa ideal. Em geral, utiliza-se distância euclidiana.

2.3.3 *Variações dos métodos clássicos*

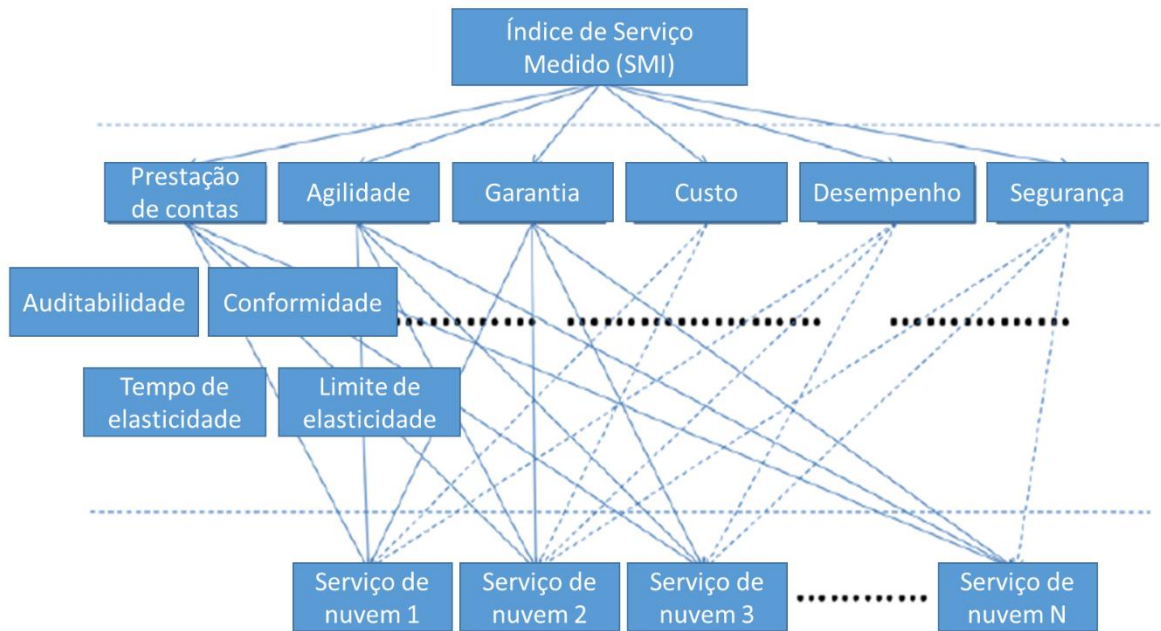
Em (Merigó & Gil-Lafuente, 2010), os autores desenvolveram uma nova abordagem que usa o operador de média ponderada ordenada (OWA) na seleção de produtos financeiros. Os operadores de agregação apresentados são úteis para problemas de tomada de decisão, pois estabelecem uma comparação entre uma alternativa ideal e as opções disponíveis, a fim de encontrar a melhor escolha. O trabalho de (Yuen, 2012) apresenta uma versão modificada do AHP, utilizada para selecionar uma oferta de SaaS de uma lista de fornecedores de *software* equivalente. Nesse trabalho, o autor propõe uma modificação na comparação entre pares para reduzir efeitos indesejáveis de subvalorização ou sobrevalorização de uma alternativa em relação a outra.

2.3.4 *A escolha do método AHP*

A revisão de literatura apresentada por (Ho, Xu, & Dey, 2010) para avaliação de fornecedores fornece evidências de que os métodos multicritério permitem decisões melhores do que a abordagem tradicional baseada somente em custo. O método AHP foi considerado a abordagem integrada mais popular. A escolha da metodologia multicritério para apoiar o processo de decisão depende diretamente do assunto em questão. O trabalho de (Tamanini, Pinheiro, & Santos, 2012) menciona que a escolha de um método deve ser o resultado de uma avaliação dos parâmetros escolhidos, o tipo e a precisão dos dados, forma do tomador de decisão de pensar, e seu conhecimento sobre o problema.

O método AHP é largamente utilizado para diversas aplicações, inclusive na comparação de serviços de nuvem. O trabalho de (Garg, Versteeg, & Buyya, 2013) descreve uma hierarquia de critérios para comparação de serviços em nuvem, criando um índice para medição do serviço (*Service Measurement Index – SMI*), conforme ilustrado na Figura 8. Os autores destacam que a primeira camada representa os objetivos de análise, que visam encontrar um índice de gestão de serviços que satisfaça os requisitos essenciais do usuário. A segunda camada contém as hierarquias de atributos de QoS essenciais e não-essenciais. A camada inferior contém os valores de todos os serviços em nuvem para todos os atributos nos níveis QoS mais baixos da hierarquia apresentadas na segunda camada.

Figura 8 - Hierarquia de critérios para comparação de serviços em nuvem



Fonte: (Garg, Versteeg, & Buyya, 2013)

O artigo de (Yuen, 2012) lista diversas situações onde o método AHP foi empregado com sucesso em decisões relacionadas a serviços de nuvem:

- Seleção de *softwares* modelo SaaS para gestão de relacionamento com o cliente;
- Concepção de rede de *data center*;
- Serviço bancário baseado em nuvem;
- Segurança de informação de computação em nuvem;
- *Software* para hospitais.

No modelo proposto neste trabalho, o método AHP foi utilizado para construir o fator moderador qualitativo de comparação entre as opções de *software* nos modelos SaaS e *on-premises*. O método foi escolhido devido a: 1) facilidade de tratamento de hierarquias de critérios, tornando-o genérico o suficiente para as diversas características da computação em nuvem; 2) facilidade de utilização e compreensão pelos tomadores de decisão, através do uso de uma escala de nove pontos que facilita a comparação entre critérios.

Capítulo 3

TRABALHOS RELACIONADOS

Durante a revisão de literatura, foram identificados diversos trabalhos relacionados aos fatores que afetam as decisões de adoção de serviços em nuvem. Alguns trabalhos relacionam-se à análise de custos, e nessa categoria vários se dedicam a técnicas para minimizar os custos, usando métodos como programação linear e teoria das filas. A maioria dos trabalhos analisados se refere à avaliação qualitativa, pois a variedade de aspectos qualitativos a serem analisados é muito grande. Dentre os trabalhos de avaliação qualitativa, o método AHP (Haas & Meixner, 2013) destaca-se como um dos mais populares. Poucos trabalhos procuram avaliar os dois aspectos simultaneamente. Somente um trabalho utiliza redes de Petri para simulação de resultados, como este trabalho propõe. A análise comparativa entre utilização de serviços de nuvem e hospedagem *on-premises* é feita por poucos autores, e em geral somente para um cenário (SaaS, IaaS). Na pesquisa realizada, não foi encontrado nenhum trabalho com a mesma proposta deste, ou seja, propor um *framework* para apoio ao processo decisório que possibilite análise de custos em nuvem e *on-premises* para os diversos cenários da nuvem, e que permita a utilização de método multicritério para abranger a análise qualitativa.

3.1 Modelos de custo para serviços em nuvem

Diversos autores escreveram sobre temas relacionados aos modelos de custo na computação em nuvem, apesar de ser uma área relativamente recente. De fato, a maioria das publicações tem menos de dois anos, sendo algumas ainda em versão preliminar. No entanto, a grande maioria dos trabalhos refere-se à computação científica, onde o problema encontrado é a execução de diversas tarefas, sequenciais, que necessitam de alto poder computacional, e que muitas vezes não manipulam grandes quantidades de dados.

É o caso, por exemplo, do trabalho de (McGough, A.S. et al., 2014) onde os autores comparam custos para execução de tarefas tipo HTC (*High-Throughput Computing*) em dois ambientes: a) Nuvem pública (*Amazon*) e 2) *Cluster* formado por *desktops* de laboratórios. O modelo de custos na nuvem é baseado nas horas de uso do servidor somado ao tempo para *upload* e *download* de dados. O artigo ressalta também a diferença entre horas efetivas de uso e horas de relógio de uso (p. ex. em horas de relógio de uso, quando o uso se inicia às 7:59, a cobrança é a mesma de início às 7:00, ou seja, é mais interessante iniciar o uso de um servidor no início da hora). O modelo de custos no *cluster* interno considera o custo de pessoal com

manutenção, *hardware*, energia e emissões de carbono, fatores que devem ser considerados quando se avaliam custos de nuvens privadas.

O trabalho desenvolvido por (Alfonso, C. et al, 2013) apresenta diversos fatores de custos a serem analisados, especialmente na montagem das nuvens privadas, dentre eles:

- custos de aquisição e configuração de equipamentos;
- custos do espaço físico para colocação do equipamento;
- custos de aquisição da refrigeração;
- custos administrativos (contratos, financiamentos, documentação, etc.);
- custos de licenças de *software*;
- custos de manutenção;
- custos de operação (pessoal);
- custos de energia;
- consumo dos componentes principais;
- consumo dos nós inativos;
- energia consumida pelo sistema de refrigeração.

Além disso, o artigo destaca a importância de se estimar corretamente o uso das instâncias reservadas das nuvens públicas, pois esse mecanismo permite uma redução significativa nos custos. Como vimos no capítulo 1, as instâncias reservadas permitem adquirir o direito de pagar menos por hora de uso, mediante um pagamento inicial.

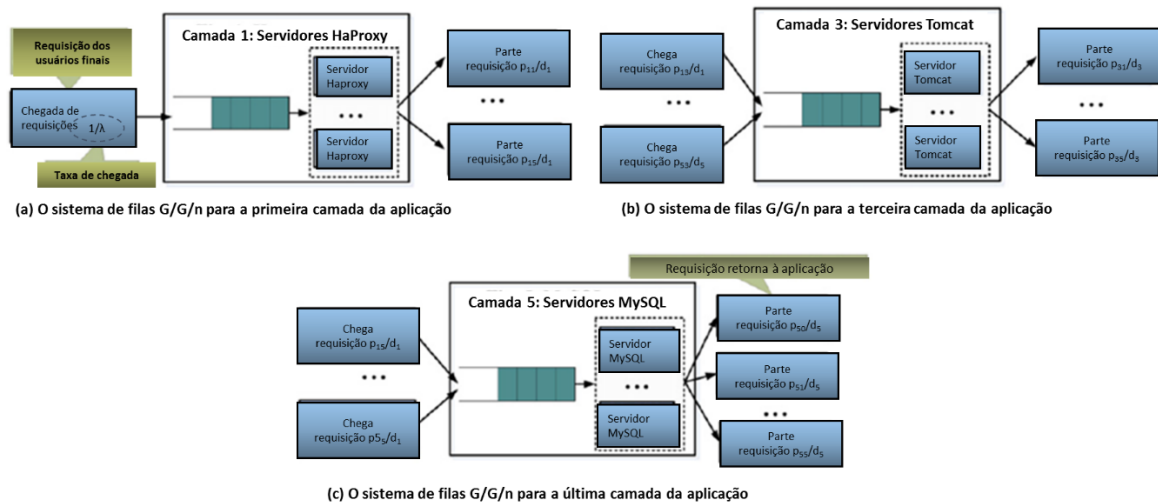
O trabalho em (Han, R. et al., 2014) traz conceitos interessantes, principalmente por analisar um problema similar aos comuns em aplicações empresariais: aplicações de nuvem em mais de uma camada. O caso de uso analisado (*e-commerce*) é bem similar a tais situações.

No artigo, os autores propõem um algoritmo para elasticidade de aplicações de várias camadas que procura reduzir o custo. O algoritmo utiliza monitores para detectar alterações na carga de trabalho em cada uma das camadas, para verificar se está aumentando a demanda, e em qual camada isso ocorre. Utilizando teoria das filas, o algoritmo constrói um modelo analítico com as filas entre as camadas para calcular tempos de resposta e verificar onde está o gargalo da aplicação, conforme ilustrado na Figura 9. Usando os resultados da análise da teoria das filas, o algoritmo é projetado para medir o custo gasto na adição de um servidor dividido pela redução no tempo de resposta por causa dessa adição.

A maioria das pesquisas recentes sobre computação em nuvem tem sido focada em aspectos técnicos. Ainda não existem informações suficientes na literatura sobre os aspectos gerenciais da computação em nuvem (Chun & Choi, 2014). Os provedores de serviços para desenvolvimento de *software* na nuvem (PaaS) têm enfrentado uma série de desafios gerenciais, os quais incluem a necessidade de um monitoramento efetivo do ambiente de infraestrutura

como serviço (IaaS) e o controle das aplicações em produção utilizadas por seus clientes ou usuários.

Figura 9 - Modelo de filas G/G/n em aplicação multicamadas



Fonte: (Han, R. et al., 2014)

Uma técnica bastante utilizada para otimização de custos em serviços de nuvem é a programação linear, conforme apresenta o trabalho de (Malawski, M., Figiela, K., & Nabrzyski, J., 2013). A formulação do problema é elegante e simples, e os autores apenas mencionam os algoritmos utilizados e apresentam os resultados. O problema estudado é a alocação de tarefas de diversos tipos para execução em máquinas virtuais.

De acordo com (Baars, T. et al, 2014), as áreas usuárias de serviços de TI nas empresas devem acompanhar os custos dos serviços que utilizam, permitindo um controle maior dos custos de TI, melhorando o processo de tomada de decisão e o alinhamento com as necessidades do negócio, facilitando, assim, uma utilização efetiva da TI. Os autores citam a necessidade de um modelo para acompanhamento desses custos (*chargeback*) em serviços da nuvem.

O trabalho apresentado em (Van den Bossche, R. et al., 2013) formula uma série de algoritmos para organizar tarefas a serem executadas em máquinas virtuais, analisar a viabilidade da execução de tais tarefas na nuvem privada, dado um prazo a ser cumprido, e executá-las na nuvem se o prazo exigir e o orçamento permitir. Nesse trabalho, os autores não consideram o custo da nuvem privada, somente o da nuvem pública é levado em conta. Na verdade, o custo de operação da nuvem privada será um fator importante na decisão de aquisição da mesma ou operação exclusiva em nuvem pública.

O trabalho de (Altmann & Kashef, 2014) apresenta um modelo de custo genérico para nuvens híbridas federadas, ou seja, nuvens que englobam a nuvem privada instalada dentro da empresa e também nuvens públicas que permitem a migração de máquinas virtuais entre as mesmas. O modelo de custo é proposto a partir de uma vasta revisão de literatura, e é composto de elementos de custo fixos e variáveis. A lista de elementos de custo apresentada é bem abrangente e pode ser usada como ponto de partida para a construção do custo total de qualquer serviço em nuvem. A Tabela 3 apresenta os elementos de custo listados no trabalho, organizados por tipo.

Tabela 3 - Elementos de custo em nuvens híbridas

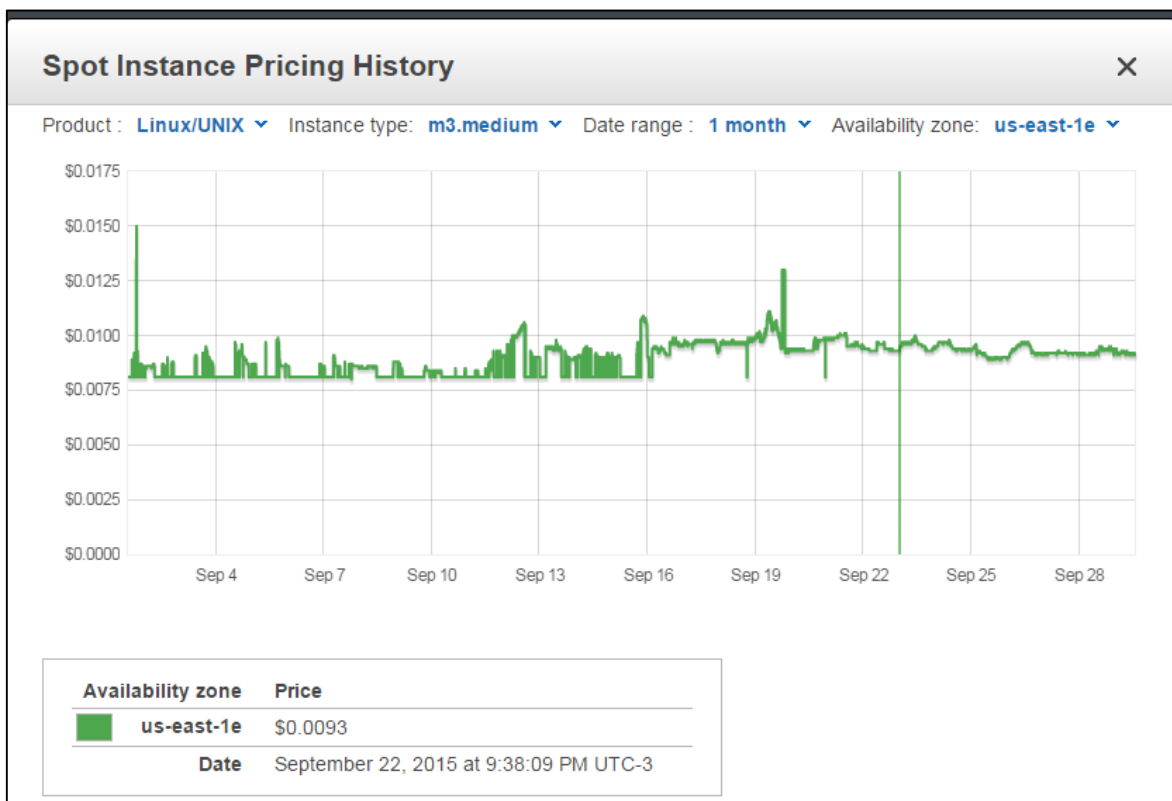
Tipo de Custo	Elemento de custo
(a) Eletricidade	(a1) Refrigeração (a2) Dispositivos eletrônicos (inativos) (a3) Dispositivos eletrônicos (em uso)
(b) Hardware	(b1) Servidor (b2) Dispositivo de rede
(c) Software	(c1) Licenças de <i>software</i> básico para o servidor (c2) Licenças de <i>software</i> “ <i>middleware</i> ” (c3) Licenças de <i>software</i> de aplicativos
(d) Pessoal	(d1) Manutenção de <i>software</i> (d2) Manutenção de <i>hardware</i> (d3) Suporte adicional
(e) Espaço físico	(e1) <i>Rack</i> de servidores, ar condicionado (e2) Cabeamento (e3) Dependências
(f) Serviços de nuvem	(f1) Conectividade (Internet) (f2) Uso de servidores na nuvem (f3) Transferência de dados para a nuvem (f4) Transferência de dados da nuvem (f5) Uso de armazenamento na nuvem (f6) Transferência de dados entre nuvens
(g) Implantação (<i>deployment</i>)	(g1) Número de implantações

Fonte: adaptado de (Altmann & Kashef, 2014)

O trabalho de (Tang, S. et al., 2014) apresenta uma estratégia para utilizar o recurso de *spot Prices* que a *Amazon* disponibiliza para alugar sua capacidade não utilizada no momento. A Figura 10 apresenta o histórico dos preços *spot* no mês de setembro de 2015. Note que, no dia 22 de setembro, o preço por hora de uma instância tipo *m3.medium* é de US\$ 0.0093

por hora, bem inferior ao preço da hora na modalidade *on-demand* (US\$ 0.067). Na modalidade *spot*, é possível registrar um lance de preço, e caso o preço da instância *Amazon EC2* por hora caia abaixo do valor registrado, a instância é alocada para o usuário pelo preço estipulado no lance, enquanto o preço de mercado da *Amazon* continuar abaixo do lance. Quando o preço subir, a instância é automaticamente desligada. Esse tipo de utilização é adequado para a execução de tarefas que possam ser interrompidas e continuadas em um segundo momento. A estratégia apresentada pelos autores é construída através da formulação do problema como um Processo de Decisão de Markov Restrito. Com base nesse modelo, é possível obter uma estratégia de lances randomizada através de programação linear. No exemplo da Figura 10, o preço de US\$ 0.0093 do dia 22 de setembro é inferior também ao menor preço por hora praticado utilizando-se uma instância reservada de 3 anos, que é de US\$ 0.0261 (vide Tabela 1). Nessas condições, é mais vantajoso utilizar a instância *spot*.

Figura 10 - Histórico de Preços de Instâncias Spot

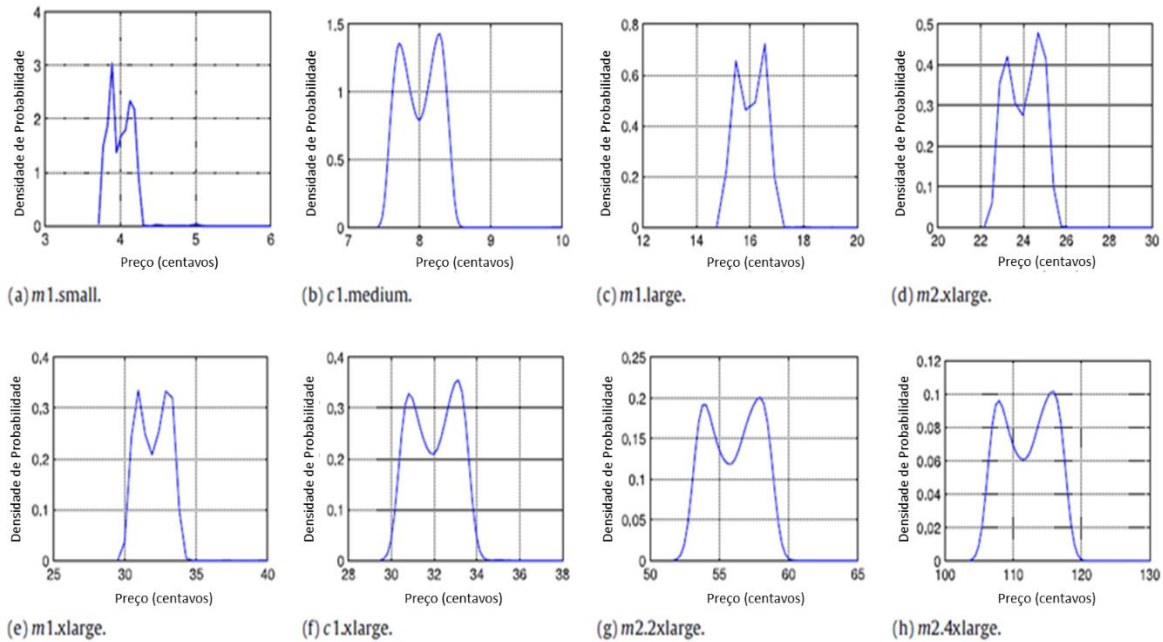


Fonte: (Amazon Web Services - Spot Instances, 2015)

Continuando com a investigação das possibilidades de uso das instâncias *spot*, o trabalho de (Javadi, Thulasiram, & Buyya, 2013) apresenta uma caracterização estatística no preço desse tipo de instância. Pode-se observar, nos gráficos que representam as densidades de probabilidade de preços, bimodalidade nas funções densidade de probabilidade, em geral

simétrica (a exceção é a instância *m1.small*, possivelmente por causa de diversos padrões de uso desse tipo de instância, por ser o recurso mais barato oferecido). A Figura 11 apresenta os gráficos produzidos no trabalho.

Figura 11 - Funções de Densidade de Probabilidade de Preços de Instâncias Spot



Fonte: (Javadi, Thulasiram, & Buyya, 2013)

A metodologia proposta em (Sripanidkulchai & Sujichantararat, 2012) é útil para a comparação de infraestrutura como serviço (IaaS) versus a opção de *on-premises*, e apresenta alguns exemplos onde o custo do IaaS não é tão competitivo, principalmente em mercados emergentes onde o salário dos profissionais de TI é menor.

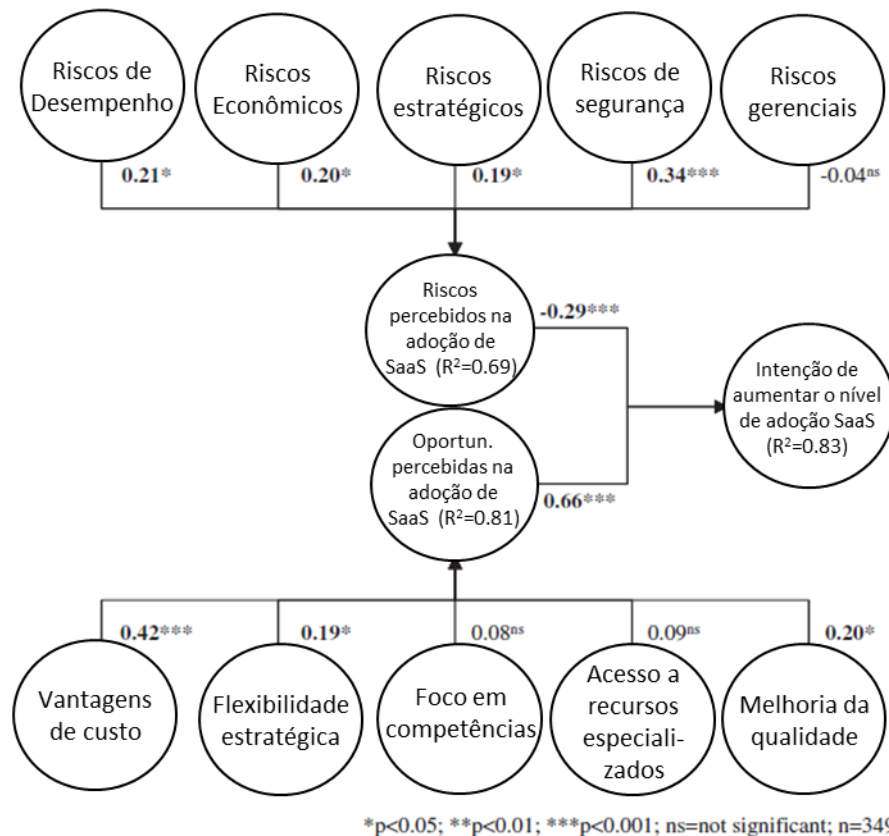
A análise dos custos *on-premises* em relação às soluções de SaaS foi amplamente estudada em (Bibi, Katsaros, & Bozanis, 2012). Esse estudo propõe uma visão diferente dos fatores de custo, mais adequada ao modelo SaaS, mas com algumas semelhanças com o modelo apresentado em (Altmann & Kashef, 2014).

3.2 Avaliação qualitativa de serviços em nuvem

A literatura sobre avaliação qualitativa dos serviços de nuvem vem crescendo junto com a oferta de tais serviços. Muitos fornecedores de *software* no modelo tradicional buscam o modelo SaaS para ampliar sua gama de clientes. No entanto, existem diferenças significativas entre os dois modelos. Segundo (Gagnon, S. et al, 2011), “o modelo SaaS precisa superar diversos obstáculos para se mostrar um modelo de negócio de sucesso. Licenciar aplicativos na nuvem traz riscos incalculáveis”.

A pesquisa realizada por (Benlian & Hess, 2011) revelou que as vantagens relacionadas a custos representam a maior oportunidade, enquanto segurança, desempenho e riscos econômicos representam os fatores de risco dominantes na análise de adoção do modelo SaaS, de acordo com a visão dos executivos entrevistados na pesquisa. A Figura 12 ilustra os resultados numéricos obtidos segundo a metodologia adotada, de regressão parcial de mínimos quadrados (*Partial Least Squares – PLS*).

Figura 12 - Resultados obtidos em pesquisa com executivos de TI



Fonte: (Benlian & Hess, 2011)

O estudo realizado por (Lee, Park, & Lim, 2013) apresenta os resultados de um levantamento na Coreia, a fim de avaliar a adoção de SaaS e seus benefícios relacionados ao negócio, através de quatro medidas – aprendizado e crescimento, processos internos do negócio, satisfação de clientes e desempenho financeiro, seguindo o modelo *Balanced Score Cards – BSC* (Kaplan & Norton, 1996). Os resultados indicaram que esses quatro elementos-chave para o sucesso de SaaS estão inter-relacionados, confirmando assim a principal premissa do BSC. Os autores também utilizam regressão de mínimos quadrados.

O trabalho de Zardari e Bahsoon (2011) propõe um processo com base na Engenharia de Requisitos Orientada a Metas (Gore) visando o fornecimento de uma abordagem sistemática para avaliar um provedor de nuvem.

O trabalho de (Wu, Lan, & Lee, 2011) apresenta um *framework* para avaliação de fatores importantes na decisão de adoção de computação em nuvem, baseados no método Dematel. Nessa abordagem, riscos e benefícios são avaliados para montar um diagrama a fim de visualizar causa e efeito. Os autores também apresentam os resultados de um estudo de caso realizado. Neste, foram avaliados oito benefícios (PB1 a PB8) e sete riscos (PR1 a PR7): PB1 (pagamento por uso), PB2 (fácil e rápida distribuição para usuários finais), PB3 (pagamentos mensais), PB4 (utilização de sistemas padrão), PB5 (menor custo de pessoal de TI), PB6 (oferece sempre a funcionalidade mais atual), PB7 (simplifica o compartilhamento de sistemas com parceiros), PB8 (aparenta ser a tendência do futuro), PR1 (localização dos dados e segurança), PR2 (segurança de rede e aplicação), PR3 (integração de dados e segregação), PR4 (autenticação e autorização), PR5 (vulnerabilidades da virtualização), PR6 (acesso a dados e *backup*), e PR7 (gerenciamento de identidade e processo de *login*).

O trabalho de (Sousa, E. et al., 2015) apresenta um *framework* onde custo e confiabilidade são avaliados na decisão de adoção de serviços de nuvem. Nesse trabalho, os autores utilizam as redes de Petri estocásticas como ferramenta para modelagem e análise do espaço de estados possíveis.

Em (Marston, S. et al, 2011), os autores identificam os pontos fortes, fracos, oportunidades e ameaças (Swot) para a indústria de computação em nuvem, bem como as várias questões que afetam as diferentes partes interessadas. A Tabela 4 ilustra os principais pontos abordados nessa análise Swot.

Tabela 4 - Análise Swot dos serviços em nuvem

Forças	Fraquezas	Oportunidades	Ameaças
<ul style="list-style-type: none"> ❖ Escalabilidade rápida ❖ Redução de custos de infraestrutura ❖ Redução de custos de gerenciamento e atualizações ❖ Economia de energia 	<ul style="list-style-type: none"> ❖ Falta de controle sobre dados na nuvem ❖ Impossibilidade de obter garantias de qualidade de serviço em casos de missão crítica 	<ul style="list-style-type: none"> ❖ Mercados emergentes - avanço como foi a telefonia móvel ❖ Pequenos negócios ❖ Combinações de serviços na nuvem (<i>mash-ups</i>) ❖ Redução de emissão de carbono 	<ul style="list-style-type: none"> ❖ Aspectos culturais dos departamentos de TI não favoráveis à inovação ❖ Segurança ❖ Desempenho e confiabilidade ❖ Falta de padrões ❖ Regulamentações locais

Fonte: adaptado de (Marston, S. et al, 2011)

Os fatores qualitativos que afetam a contabilidade e o retorno no uso de recursos de nuvem dentro dos setores da empresa são objetos do estudo em (Baars, T. et al, 2014). Destaca-se a importância de fatores como previsibilidade, transparência, responsabilidade, mensurabilidade, acurácia, facilidade de controle, comparabilidade e o custo da contabilidade.

O trabalho de (Mouratidis, H. et al., 2013) apresenta um framework para seleção de provedores de nuvem com base em requisitos de segurança e de privacidade. Os autores discorrem sobre os fatores de segurança em serviços em nuvem tais como: integridade de dados, segregação de dados, disponibilidade de dados e de rede, proteção de dados, políticas da organização, observância de aspectos legais, ameaças à segurança e privacidade, perda do controle físico e vulnerabilidades na virtualização. Além disso, apresentam uma metodologia para listar requisitos de segurança e avaliar os provedores de nuvem com base nos requisitos listados.

O trabalho de (Boampong & Wahsheh, 2012) discorre sobre os principais riscos em serviços de nuvem e apresenta estratégias de mitigação. Os resultados de entrevistas apontam a segurança da informação como o risco mais crítico dentre as diversas áreas. A Tabela 5 apresenta as áreas de risco estudadas e os resultados apresentados.

Tabela 5 - Áreas críticas de risco de serviços de nuvem e virtualização

Área de Risco	Crítico	Relevante	Irrelevante
Segurança da informação	91.7%	8.3%	0.0%
Gerenciamento de operações	41.7%	58.3%	0.0%
Gerenciamento de mudanças	41.7%	50.0%	8.3%
Recuperação de desastres/planos de continuidade do negócio	66.7%	33.3%	0.0%
Gerenciamento do nível de serviço de terceiros	41.7%	41.7%	16.7%
Gerenciamento de interface	8.3%	50.0%	41.7%
Regulamentação e legislação	33.3%	41.7%	25.0%

Traduzido de (Boampong & Wahsheh, 2012)

A segurança é o principal fator de risco que afeta a adoção dos serviços em nuvem para a maioria dos autores, ao passo que a redução de custos é o principal benefício esperado. (Benlian & Hess, 2011); (Wu, Lan, & Lee, 2011); (Bayrak, 2013).

O trabalho de (Garg, R. & Stiller, B., 2014) apresenta uma metodologia baseada em Topsis para avaliação do problema de adoção de infraestrutura na nuvem.

3.3 Quadro comparativo

O quadro comparativo na Tabela 6 destaca as características semelhantes dos trabalhos apresentados na revisão de literatura em termos de metodologia, problema analisado e escopo da análise.

Os trabalhos anotados com ‘X’ na coluna Custos se relacionam à análise de custos de serviços de nuvem. A coluna ‘Qualidade’ indica se o trabalho aborda aspectos qualitativos dos serviços de nuvem. A coluna ‘AHP’ indica se o trabalho utiliza esse método na avaliação

qualitativa. A coluna ‘RP’ indica se o trabalho utiliza a ferramenta redes de Petri. As colunas IaaS, PaaS e SaaS indicam se o escopo da análise abrange, respectivamente, cada um dos modelos. Por fim, a coluna ‘on-premises’ indica se o escopo do trabalho abrange a comparação de serviços em nuvem com serviços hospedados localmente.

Tabela 6 - Quadro comparativo de trabalhos relacionados

Trabalho	Custos	Qualidade	AHP	RP	IaaS	PaaS	SaaS	on-premises
(Alfonso, C. et al, 2013)	X				X			
(Altmann, J. & Kashef, M.M., 2014)	X				X			X
(Avram, M.G., 2014)		X						
(Baars et al, 2014)		X						
(Bayrak, T., 2013)		X						
(Benlian, A. & Hess, T.,2011)		X					X	
(Beserra, P. et al, 2012)		X						
(Bibi, S., Katsaros, D. & Bozanis, P., 2012)	X	X			X		X	X
(Boampong, P. & Wahsheh, L., 2012)		X						
(Gagnon, S. et al, 2011)		X					X	
(Garg, S.K.,Versteeg, S. & Buyya, R., 2013)		X	X		X	X	X	
(Garg, R. & Stiller, B., 2014)	X	X			X			
(Han, R. et al., 2014)	X				X		X	
(Javadi, B., Thulasiram, R. & Buyya, R., 2013)	X				X			
(Lee, S., Park, S. & Lim, G.,2013)		X						
(Malawski, M., Figiela, K. & Nabrzyski, J., 2013)	X				X			
(Marston, S. et al, 2011)		X			X	X	X	
(McGough, A.S. et al, 2014)	X				X			X
(Mouratidis, H. et al, 2013)		X			X	X	X	
(Ribas, M. et al., 2014)	X	X	X	X			X	X
(Saripalli, P. & Pingali, G., 2011)		X					X	
(Sripanidkulchai, K. & Sujichantararat, S., 2012)	X				X			X
(Sousa, E. et al., 2014)	X	X		X	X			
(Tang. S. et al., 2014)	X	X			X			
(Van den Bossche, R. et al., 2013)	X							
(Wu, W.WW, Lan, L. & Lee, Y., 2011)		X					X	
(Yuen, K., 2012)		X	X				X	

Fonte: a autora.

O artigo de (Alfonso, C. et al, 2013) apresenta modelos de custos para IaaS e destaca a complexidade na estimativa desses custos, citando a importância do correto uso das modalidades de compra disponíveis (p. ex. instâncias reservadas). No entanto, o foco do trabalho é na área de custos e economia de energia, não incorporando aspectos qualitativos.

O trabalho de (Altmann & Kashef, 2014) analisa profundamente os fatores de custos a serem considerados para calcular o custo final de serviços de nuvens híbridas, o tipo mais abrangente de implantação, pois envolve infraestrutura local. No entanto, os fatores qualitativos também são pouco explorados.

As principais vantagens e barreiras para adoção de serviços de nuvem são exploradas no trabalho de (Avram, 2014). Redução de custos, baixo ou nenhum investimento inicial, inovação, escalabilidade, viabilidade de novas classes de serviços são vantagens contrapostas a barreiras como segurança e privacidade, conectividade, confiabilidade, interoperabilidade, valor econômico, mudanças organizacionais, questões políticas devido a fronteiras nacionais. O trabalho explora os impactos de tais fatores, mas não propõe nenhum modelo matemático para apoio à decisão. De forma similar, o trabalho de (Bayrak, 2013) discorre sobre os aspectos favoráveis e desfavoráveis da terceirização de aplicações para entidades externas, sem abordar as dificuldades referentes às avaliações dos custos.

A pesquisa realizada por (Benlian, A. & Hess, T., 2011) faz uma análise elaborada sobre a visão dos executivos entrevistados, de que resulta que as vantagens relacionadas a custos são listadas como a maior oportunidade, enquanto segurança, desempenho e riscos econômicos representam os fatores de risco dominantes na análise de adoção do modelo SaaS. Porém, a pesquisa também não aborda em profundidade a questão do custo nos serviços em nuvem, além de ser focada nos serviços de nuvem tipo SaaS.

O passo-a-passo proposto para nortear a migração e adoção de serviços em nuvem descrito em (Beserra, P. et al, 2012) fornece um método prático para tal tipo de decisão, mas, novamente, as questões de custo não são abordadas em profundidade, pois não é o foco do trabalho.

O trabalho de (Bibi, Katsaros, & Bozanis, 2012) inspirou o início desta pesquisa, apresentando fatores qualitativos (análise Swot) e discorrendo especificamente sobre os custos nos modelos IaaS e PaaS. O presente trabalho vai além, no sentido em que generaliza os elementos de custo, tornando-o aplicável a qualquer modelo de nuvem (PaaS, por exemplo), e utiliza o paradigma BDIM, no sentido de criar um modelo para quantificar e atribuir uma nota final a cada solução, incorporando aspectos quantitativos e qualitativos.

O trabalho de (Boampong & Wahsheh, 2012) foca exclusivamente na questão qualitativa – riscos – discorrendo sobre os principais riscos em serviços de nuvem e apresentando estratégias de mitigação.

Em (Gagnon, S. et al, 2011), o foco é a análise da viabilidade do modelo de negócio SaaS para os desenvolvedores de *software*. São apresentados os principais desafios, em uma análise inerentemente qualitativa.

O trabalho em (Garg, Versteeg, & Buyya, 2013) descreve uma hierarquia de critérios para comparação de serviços em nuvem, criando um índice para medição do serviço (SMI) de modo similar a este trabalho, mas ainda assim distinto, pois 1) inclui a análise dos custos juntamente com os demais fatores qualitativos, o que aumenta o viés na decisão, 2) foca em serviços de nuvem, sem a comparação com a solução *on-premises*, e 3) não aborda em profundidade a análise do fator custo.

O artigo em (Han, R. et al., 2014) propõe um algoritmo para elasticidade de aplicações de várias camadas que procura reduzir o custo, exemplificando a complexidade desse tipo de análise, assim como este trabalho introduz o modelo para redução de custos usando instâncias *spot*, no capítulo 5. No entanto, os aspectos qualitativos não fazem parte do escopo da análise.

O trabalho de (Javadi, Thulasiram, & Buyya, 2013) se aprofunda na análise dos custos de serviços de nuvem, em especial das instâncias *spot*, e é um excelente exemplo da complexidade desse tipo de análise, porém não há espaço para análise qualitativa, dada a profundidade do trabalho na área de custos.

O estudo realizado por (Lee, Park, & Lim, 2013) foca no aspecto qualitativo e de certo modo se assemelha a este presente trabalho, no sentido de investigar o alinhamento com o negócio, utilizando o modelo BSC. Porém, não se aprofunda na questão da complexidade dos custos inerentes aos modelos de nuvem.

O trabalho de (Malawski, M., Figiela, K., & Nabrzyski, J., 2013) utiliza programação linear para otimização dos custos de serviços em nuvem, e apresenta um método elegante para o problema. Porém, o método é aplicável para um determinado problema, a alocação de tarefas de diversos tipos para execução em máquinas virtuais, focando assim os serviços IaaS e sem análises mais aprofundadas de aspectos qualitativos ou alinhamento com o negócio.

No trabalho de (McGough, A.S. et al., 2014), os autores exploram diversos aspectos relevantes dos custos de infraestrutura na nuvem e *on-premises*, apresentam políticas para redução de custos, porém exploram pouco os aspectos qualitativos envolvidos.

Aspectos qualitativos, retratados através de uma análise Swot, são apresentados no trabalho de (Marston, S. et al, 2011), porém a questão dos custos também não é abordada em

profundidade no trabalho. Da mesma forma, o *framework* para seleção de provedores de nuvem apresentado por (Mouratidis, H. et al., 2013) também é baseado em requisitos de segurança e de privacidade, com ênfase nesses aspectos qualitativos.

No trabalho que deu origem a este (Ribas, M. et al., 2014), o foco é específico no modelo SaaS, e também não explora em profundidade questões relacionadas a custo, como este trabalho aborda no capítulo 5.

O trabalho de (Saripalli & Pingali, 2011) propõe um sistema de apoio à decisão para avaliar os diversos fatores envolvidos na decisão de adoção de serviços em nuvem, porém aspectos de custo, especialmente relacionados à hospedagem local, não são abordados em profundidade.

O trabalho de (Sousa, E. et al., 2015) apresenta um *framework* onde custo e confiabilidade são avaliados durante o processo de planejamento da infraestrutura. Nesse trabalho, os autores utilizam as redes de Petri estocásticas como ferramentas de modelagem para análise de estados e obtenção de diferentes panoramas de custo e confiabilidade. O trabalho explora os cenários em profundidade, porém o foco é na infraestrutura (IaaS), utilizando um modelo de custo mais adequado para esse tipo de serviços de nuvem.

O trabalho de (Tang, S. et al., 2014) apresenta uma estratégia para definir um lance ótimo para utilizar o recurso de *spot Prices* da *Amazon*, que poderia ser incorporado no modelo que será descrito no capítulo 5, porém não avalia em profundidade os aspectos qualitativos referentes aos modelos SaaS e PaaS.

O trabalho de (Sripanidkulchai & Sujichantararat, 2012) apresenta uma metodologia útil para a comparação de infraestrutura como serviço (IaaS) versus a opção de *on-premises*, porém é restrita ao modelo IaaS e não abrange aspectos qualitativos. Já o trabalho de (Baars, T. et al, 2014) foca nos fatores qualitativos que afetam a contabilidade e o retorno no uso de recursos de nuvem dentro dos setores da empresa, sem aprofundar a questão dos custos. O aspecto qualitativo é também o foco principal do trabalho de (Wu, Lan, & Lee, 2011), onde riscos e benefícios são avaliados para montar um diagrama para visualizar causa e efeito, e os custos têm um destaque limitado. O trabalho de (Yuen, 2012) explora fatores qualitativos através de um novo método proposto para melhorar o método AHP, sendo este novo método o principal foco do trabalho.

O trabalho apresentado em (Van den Bossche, R. et al., 2013) foca na viabilidade de execução de tarefas em máquinas virtuais, utilizando a nuvem pública como uma extensão

da nuvem privada, dado um prazo e um orçamento determinado. Não são analisados a fundo aspectos qualitativos, e o problema analisado se refere a IaaS.

Enfim, dentro da pesquisa realizada, não foi encontrado nenhum trabalho que apresente um modelo de apoio à decisão com as características que este trabalho apresenta:

- Separação clara da análise de custos e fatores qualitativos, com o objetivo de reduzir o viés subjetivo da decisão;
- Utilização de método multicritério genérico que permite a avaliação dos critérios qualitativos mais adequados a cada tipo de serviço de nuvem;
- Proposição de conjunto de critérios qualitativos básicos com base na literatura;
- Fácil extensão para incorporação de novos critérios qualitativos;
- Proposição de conjunto de critérios de custo básicos com base na literatura;
- Fácil extensão para incorporação de novos critérios de custo;
- Uso de ferramenta de modelagem que permite que as extensões, em especial para cálculo de custos, possam ser validadas através de simulações;
- Visualização simples e clara dos fatores envolvidos na decisão.

Dessa forma, a principal contribuição deste trabalho para a pesquisa científica constitui-se no modelo proposto, considerando a rede principal (apresentada no capítulo 4) e sua extensão para redução de custos utilizando instâncias *spot* (apresentada no capítulo 5).

Capítulo 4

O MODELO PROPOSTO

Este trabalho propõe um *framework* inédito que permite a apreciação detalhada dos custos combinada com fatores qualitativos para produzir uma avaliação final sobre duas soluções, a que utiliza serviços de nuvem e a solução que hospeda o serviço internamente, *on-premises*. O objetivo é empregar uma metodologia que seja simples, para que os gerentes possam visualizar e entender o seu funcionamento (e, para tanto, os grafos da rede de Petri são muito adequados), ao mesmo tempo em que possa ser facilmente adaptável para os diversos cenários.

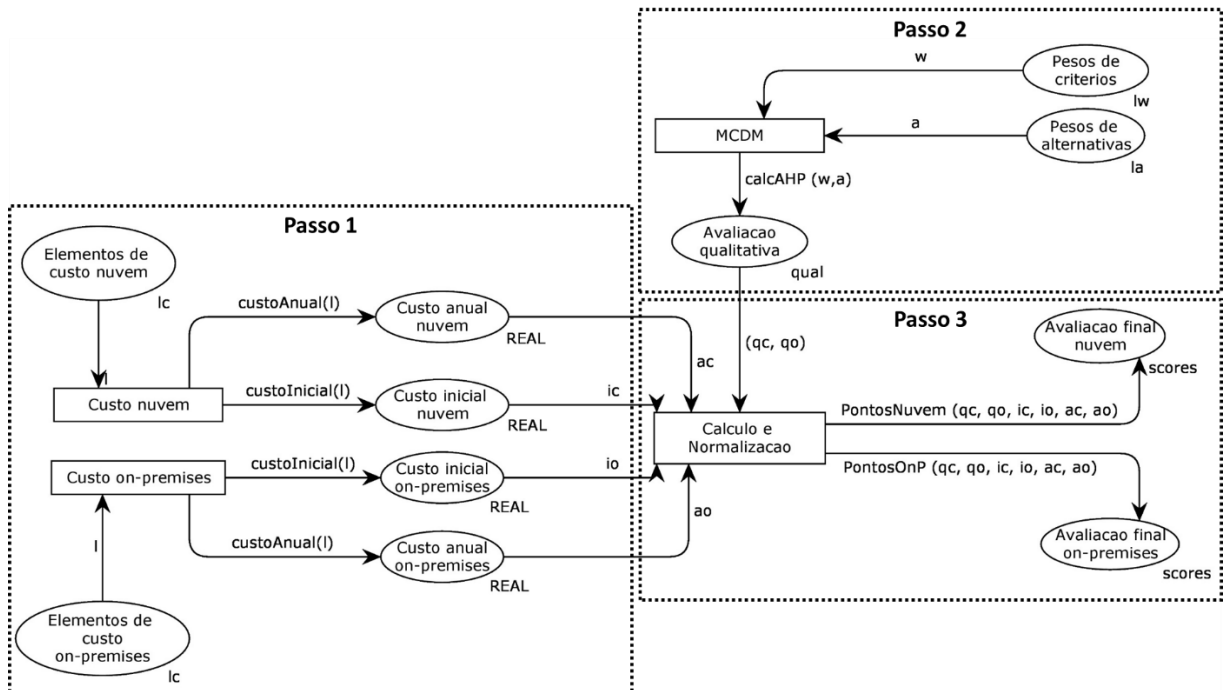
De acordo com o tipo de serviço (IaaS, PaaS e SaaS), muitos aspectos a serem analisados são diferentes. Por exemplo, o fator ‘custo de reserva de instância’ só aparece quando o tipo de serviço de nuvem analisado é IaaS. No entanto, existem aspectos em comum na análise. Por exemplo, o fator ‘consumo de energia’ é um aspecto comum a ser analisado, seja o serviço do tipo IaaS, PaaS ou SaaS. Além disso, a tecnologia de nuvem está evoluindo rapidamente e novos serviços são disponibilizados constantemente, com novos aspectos a serem analisados. Dessa forma, a adaptabilidade do *framework* é fundamental para que o mesmo continue sendo útil com o passar do tempo.

O mecanismo do *framework* é dividido em três passos:

1. Estimar a vantagem (ou desvantagem) de custo do serviço de nuvem em relação ao serviço hospedado localmente (*on-premises*);
2. Avaliar benefícios e riscos de modo qualitativo usando um método multicritério para tomada de decisão, como o AHP;
3. Calcular a razão custo/benefício. A opção com a menor taxa é melhor.

A Figura 13 ilustra o *framework* proposto. A parte esquerda ilustra o passo 1, comparação de custos. A parte superior direita ilustra o passo 2, a avaliação qualitativa. Por fim, a parte inferior direita ilustra o passo 3, a normalização e cálculo do resultado final, que produz a saída do modelo.

Figura 13 - Modelo de decisão proposto



Fonte: a autora

O modelo utiliza redes de Petri coloridas (RPC) para permitir o uso de tipos de dados complexos (denominadas cores) e também aproveitar as possibilidades da linguagem de programação funcional ML. O modelo CPN contém:

Lugares, desenhados como elipses ou círculos;

Transições, desenhadas como caixas retangulares;

Arcos direcionados conectando lugares e transições; e

Inscrições textuais próximas a lugares, transições e arcos.

Dessa forma, o modelo pode ser resumido a um único grafo. O modelo admite quatro parâmetros de entrada:

1. Elementos de custo da nuvem;
2. Elementos de custo local (*on-premises*);
3. Pesos dos critérios qualitativos;
4. Pesos das alternativas em relação aos critérios.

E produz duas saídas:

1. Avaliação final do serviço de nuvem;
2. Avaliação final do serviço local.

No entanto, para compreender o modelo completamente, é necessário conhecer e entender a estrutura construída, ou seja, as cores (tipos) e inscrições (funções). Esses elementos serão discutidos em detalhe nas próximas seções, que apresentarão cada passo do *framework*.

4.1 Passo 1: Estimativa de custos

A estimativa de custos para serviços de TI, tanto operados localmente (*on-premises*) como adquiridos de um provedor de nuvem, variam significativamente dependendo da aplicação, do tamanho da empresa e da complexidade dos processos de negócio envolvidos, conforme destacado no trabalho elaborado por (The Yankee Group, 2013). O trabalho apresentado em (Ribas, et al., 2014), segue a linha utilizada por (Bibi, Katsaros, & Bozanis, 2012). Os custos são divididos em:

- custos iniciais, os investimentos realizados para disponibilizar o serviço;
- custos anuais, custos operacionais e de depreciação dos equipamentos.

4.1.1 Cores

Para capturar elementos de custo genéricos, foram definidas as seguintes cores (tipos de dados):

- **ce** (elemento de custo), formado por uma tupla com três itens de informação básica: tipo de custo (inicial ou anual), nome do elemento de custo (por exemplo: *software*, *hardware*, pessoal), e valor financeiro do elemento de custo;
- **lc**, uma lista de elementos de custo.

Por exemplo, um elemento de custo representando a assinatura de um aplicativo no modelo SaaS que custa R\$ 99.000,00, paga anualmente, seria representado da forma: (Anual, Assinatura, R\$ 99.000,00). A cor **lc** é utilizada nos lugares do modelo “Elementos de custo nuvem” e “Elementos de custo *on-premises*”.

4.1.2 Lugares

No passo 1, serão utilizados os lugares que representam as duas entradas do modelo, e também lugares que armazenam temporariamente o total calculado a partir da soma dos elementos de custo de entrada.

O lugar “Elementos de custo nuvem” contém os elementos de custo identificados para estimar o custo total do serviço de nuvem. Essa lista precisa ser um dos parâmetros de entrada, pois é específica de cada serviço. Um exemplo de lista de custos para operar um serviço no modelo SaaS pode ser:

- (Inicial, Treinamento, R\$ 19.200,00)
- (Anual, Salários, R\$ 60.000,00)

- (Anual, Internet, R\$ 36.000,00)
- (Anual, Assinatura, R\$ 99.000,00)

O lugar “Elementos de custo *on-premises*” contém uma lista de elementos de custo identificados para estimar o custo total do serviço quando operado localmente (*on-premises*). Assim como a lista de elementos de custos para operar o serviço na nuvem, essa lista também é um parâmetro de entrada do modelo, por ser específica de cada serviço. Essa lista é em geral mais extensa e muitas vezes mais difícil de estimar. Dentre os elementos de custo listados na Tabela 3, podem-se identificar vários que se aplicariam facilmente a um serviço operado localmente. Exemplos de elementos de custo dessa lista poderiam ser:

- (Inicial, *Hardware*, R\$ 28.000,00)
- (Inicial, *Software*, R\$ 145.000,00)
- (Anual, Eletricidade, R\$ 3.600,00)
- (Anual, Refrigeração, R\$ 12.000,00)

Os lugares “Custo inicial nuvem”, “Custo anual nuvem”, “Custo inicial *on-premises*” e “Custo anual *on-premises*” são lugares para armazenar temporariamente a soma dos elementos de cada tipo em cada lista, ou seja, para os serviços em nuvem, o total inicial e o anual, e para os serviços locais, também o total inicial e o anual.

4.1.3 Transições, arcos e inscrições

As transições “Custo nuvem” e “Custo *on-premises*” são utilizadas para efetuar o cálculo do total de investimento inicial e de despesas anuais para cada serviço (em nuvem ou local). As inscrições dos arcos representam funções programadas que realizam o cálculo e convertem os tipos de dados:

- Função custoInicial(list): recebe uma lista de elementos de custo (t, n, v) e retorna a soma de todos os elementos de custo do tipo “inicial”. Dessa forma, calcula-se o investimento total necessário para operar o serviço.
- Função custoAnual(list): recebe uma lista de elementos de custo (t, n, v) e retorna a soma de todos os elementos de custo do tipo “anual”. Dessa forma calcula-se a despesa operacional e depreciação necessária para operar o serviço.

4.2 Passo 2: Avaliação qualitativa

A avaliação de riscos e benefícios será realizada utilizando um método multicritério. A proposta básica é utilizar o AHP (Haas & Meixner, 2013), pois é um método bastante popular na análise de serviços de TI e nuvem, conforme constatado na revisão de literatura (Garg, Versteeg, & Buyya, 2013), (Yuen, 2012). Duas alternativas serão comparadas: o serviço em nuvem e o

serviço operado localmente. O método AHP baseia-se na comparação entre pares para obter pesos que são atribuídos aos critérios e também às alternativas referentes a cada critério. Os elementos do modelo utilizados para implementar o passo 2 são a seguir apresentados.

4.2.1 Cores

As seguintes cores (tipos de dados) são utilizadas:

- **weight** (peso): um par de valores (critério, peso), em que o primeiro elemento é o critério qualitativo a ser avaliado e o segundo, o peso atribuído ao critério, que deve ser um valor entre 0 e 1;
- **lw**: uma lista de pesos;
- **alt** (alternativas): uma tupla com três valores (critério, alternativa, peso) onde critério representa o nome do critério qualitativo a ser avaliado (por exemplo, qualidade, desempenho etc.), a alternativa deve ser uma das duas possíveis, nuvem ou local, e o peso é o peso do critério para a alternativa;
- **la**: uma lista de alternativas;
- **qual**: um par de números reais (qc, qo) representando a avaliação qualitativa final obtida após aplicação do método, respectivamente para a alternativa nuvem (qc) e local (qo).

Exemplos de listas de pesos poderiam ser:

- [(estratégia, 0.3), (qualidade, 0.4), (risco, 0.3)].
- [(estratégia, 0.5), (qualidade, 0.2), (risco, 0.3)].

Exemplos de listas de alternativas poderiam ser:

- [(estratégia, nuvem, 0.5), (estratégia, local, 0.5), (qualidade, nuvem, 0.5), (qualidade, local, 0.5), (risco, nuvem, 0.5), (risco, local, 0.5)]

4.3.2 Lugares

No lugar “Pesos de critérios” será colocada a lista de critérios e seus respectivos pesos. Esse é um parâmetro do modelo, e deve ser obtido a partir da aplicação do método multicritério desejado – no caso do AHP, através da comparação de pares.

No lugar “Pesos de alternativas” será colocada a lista de pesos associados aos critérios para cada alternativa, a fim de realizar a comparação final de acordo com o método utilizado. Também é um parâmetro do modelo.

O lugar “Avaliação qualitativa” contém o resultado da avaliação qualitativa das duas alternativas comparadas, de acordo com o cálculo executado segundo a metodologia AHP.

4.3.3 Transições, arcos e inscrições

A transição “MCDM” calcula a avaliação final baseada nos pesos de cada critério utilizando a função (calcAHP) para realizar os cálculos necessários. A função calcAHP (lw, la) recebe a

lista de pesos de critérios e a lista de pesos de alternativas, e calcula a avaliação final de cada alternativa. No modelo, essa função implementa o método AHP, mas a função pode ser customizada para utilizar outro método de decisão multicritério – como, por exemplo, a rede de cognição primitiva (Yuen, 2012) –, contanto que a função retorne um par de números reais (qc, qo) como resultado para representar a avaliação qualitativa das alternativas nuvem e local.

4.3 Passo 3: Calcular a razão custo/benefício

O último passo do modelo normaliza os custos calculados no passo 1 e calcula a razão custo/benefício para avaliação final das alternativas. Para esse passo, o modelo utiliza os elementos a seguir detalhados.

4.3.1 Cores

A cor “*scores*” representa uma lista de números reais para armazenar a avaliação final de cada alternativa. O uso de uma lista permite ao *framework* produzir vários valores para cada alternativa. Por exemplo, pode-se calcular avaliação de cada alternativa analisando o primeiro ano de uso dos serviços e após cinco anos de uso, para cada alternativa. Nesse exemplo, a lista teria dois elementos, o primeiro elemento representando a avaliação para uso no primeiro ano e o segundo elemento representando a avaliação para uso durante cinco anos.

4.3.2 Lugares

O lugar “Avaliação final nuvem” representa a avaliação (ou avaliações) finais obtidas para a alternativa nuvem. Similarmente, o lugar “Avaliação final *on-premises*” representa a avaliação (ou avaliações) finais obtidas para a alternativa local. Esses lugares contêm os resultados que representam a saída do modelo.

4.3.3 Transições, arcos e inscrições

A transição “Cálculo e normalização” recebe as estimativas de custo (inicial e anual) para os serviços de nuvem e para serviços locais, e também as avaliações qualitativas de ambos, e calcula a avaliação final. Para esse fim, são utilizadas duas funções customizáveis:

- Função “PontosNuvem”: calcula a avaliação de acordo com as técnicas definidas pelo usuário. Pode-se, por exemplo, implementar essas funções para calcular o custo total de propriedade (TCO) para um ano e para cinco anos de uso, como foi feito em (Bibi, Katsaros, & Bozani, 2012). Em seguida, os valores são normalizados para obter resultados entre 0 e 1, e calcular a razão custo/benefício da utilização de cada serviço por um ou cinco anos, conforme ilustrado na Figura 14. Essa função pode ser

customizada para usar outros métodos, como, por exemplo, valor líquido atual (NPV), como utilizado no trabalho de (Sripanidkulchai & Sujichantararat, 2012);

- Função “PontosOnP”: similar à “PontosNuvem”, utilizada para calcular a razão custo/benefício para os serviços locais.

Figura 14 - Função de avaliação final do modelo

```

fun PontosNuvem(qc, qo, ic, io, ac, ao:REAL):scores=
  let
    val tco1c = ic + ac
    val tco5c = ic + 5.0*ac
    val tco1o = io + ao
    val tco5o = io + 5.0*ao
  in
    [qc/normalize(tco1c, tco1o), qc/normalize(tco5c, tco5o)]
end

```

Fonte: a autora

4.4 Exemplos de uso do modelo

Para exemplificar o uso do modelo proposto, serão construídos e analisados alguns cenários da decisão.

4.4.1 Exemplo SaaS

O problema analisado no cenário SaaS será o mesmo apresentado no trabalho de (Ribas, et al., 2014). Naquele trabalho, foi apresentada uma versão inicial do modelo, que estava em construção. O Apêndice A detalha a rede construída na época, que serviu como base para a criação do modelo proposto neste trabalho. O exemplo utilizado era uma instituição de ensino superior (IES) com 2.000 alunos que necessitava atualizar seu sistema de gestão acadêmica. O sistema atual tinha sido adquirido quando a IES era muito menor, e ficou obsoleto com o advento das novas tecnologias web e móveis. O sistema atual utilizava dois servidores, um para o servidor web e outro para o gerenciador de bancos de dados. Os dois servidores também precisavam ser substituídos devido a problemas técnicos, pois já tinham alcançado sua vida útil. O gerente de TI, após pesquisar, decidiu comprar um sistema acadêmico novo, denominado *SCHOOL1*, com duas opções de instalação:

1) Opção local: Licença perpétua, onde a IES investiria inicialmente uma taxa de licenciamento de R\$ 145.000 (valor até 2.500 alunos). Existiria também uma taxa de manutenção anual opcional de R\$ 29.000 para acesso ao serviço de suporte técnico e

atualizações com novas versões e funcionalidades do sistema. A IES deveria prover a infraestrutura de TI necessária para instalar e hospedar o sistema.

2) Opção SaaS: assinatura anual de R\$ 99.000,00 (valor até 2.500 alunos), que incluiria suporte técnico, atualizações de versões e hospedagem. Dessa forma, os usuários da IES utilizariam um navegador web para acessar o sistema. Além disso, nessa opção existiria um acordo de nível de serviço para garantir níveis aceitáveis em termos de disponibilidade, desempenho e segurança.

Em ambas as opções, o fornecedor recomendava a aquisição de um pacote de treinamento para habilitar o pessoal de TI e usuários gestores no processo de parametrização do sistema para as necessidades da IES, a fim de garantir a utilização do sistema da forma mais adequada e eficiente. Além disso, eram necessários serviços profissionais de TI para customizações no sistema, com a construção de relatórios específicos e integrações com outros sistemas da IES. Na opção local, era necessário adquirir dois servidores (ou utilizar uma solução de virtualização equivalente) para o banco de dados e o servidor de aplicação. Eram necessárias também licenças de *software* para Windows Server e Microsoft SQL Server. Os salários dos profissionais de TI foram estimados em R\$ 60.000 anuais para cada profissional, sendo necessários dois. Estimaram-se custos de Internet da ordem de R\$ 24.000 e energia R\$ 3.600 anuais. Para a opção SaaS, os custos de treinamento eram idênticos, mas o custo do serviço era 10% menor, por não haver custos de instalação. Os custos de Internet eram mais altos na opção SaaS, estimados em R\$ 36.000, pois era necessária uma conexão com maior velocidade. No entanto, os custos de pessoal foram reduzidos, pois somente um profissional de TI seria necessário. Os custos de depreciação e manutenção dos equipamentos foram estimados em 12% do custo de aquisição.

4.4.1.1 Passo 1: estimativa de custos

A partir da descrição do problema, pode-se montar a Tabela 7 com as listas de elementos de custos estimados, parâmetro de entrada do modelo.

As transições “Custo nuvem” e “Custo *on-premises*” (Figura 13 – passo 1) calculam os investimentos iniciais e despesas anuais para cada opção e colocam nos lugares:

- Custo anual nuvem: R\$ 195.000,00
- Custo inicial nuvem: R\$ 53.760,00
- Custo inicial *on-premises*: R\$ 221.600,00
- Custo anual *on-premises*: R\$ 177.560,00

Tabela 7 - elementos de custo: parâmetros iniciais do modelo SaaS

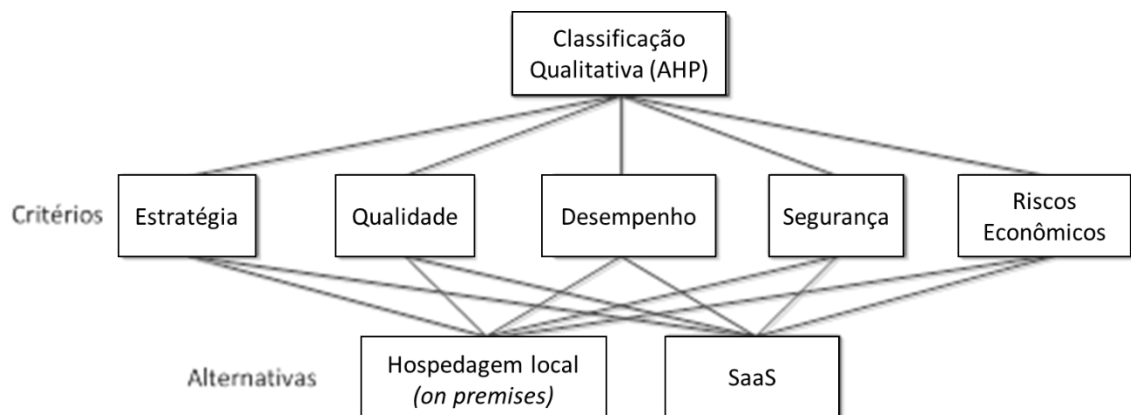
Elementos de custo <i>on-premises</i>			Elementos de custo nuvem		
inicial	treinamento	R\$ 19.200,00	inicial	treinamento	R\$ 19.200,00
inicial	serviços	R\$ 38.400,00	inicial	serviços	R\$ 34.560,00
inicial	servidores	R\$ 8.000,00			
inicial	sistema	R\$ 145.000,00			
inicial	Windows e SQL	R\$ 11.000,00			
anual	salários	R\$ 120.000,00	anual	salários	R\$ 60.000,00
anual	internet	R\$ 24.000,00	anual	internet	R\$ 36.000,00
anual	<i>software</i>	R\$ 29.000,00	anual	assinatura	R\$ 99.000,00
anual	<i>hardware</i>	R\$ 960,00			
anual	eletricidade	R\$ 3.600,00			

Fonte: a autora

4.4.1.2 Passo 2: Avaliação qualitativa

Para a avaliação qualitativa, foi construída uma hierarquia de critérios, para utilizar o método AHP. A Figura 15 ilustra a hierarquia utilizada.

Figura 15 - Avaliação de benefícios e riscos



Fonte: a autora

Existem duas alternativas para a classificação: SaaS e local (*on-premises*). Trabalhos anteriores (Benlian & Hess, 2011), (Wu, Lan, & Lee, 2011) foram utilizados para destacar os critérios básicos propostos para a avaliação qualitativa:

- **Estratégia:** refere-se a questões tais como: 1) flexibilidade para mudar de provedor de TI, 2) escalabilidade, 3) possibilidade de implantação mais rápida do *software*, 4) redução da dependência do fornecedor (*lock-in*) devido a menores custos iniciais e 5) possibilidade de concentrar esforços nas competências específicas da empresa;
- **Qualidade:** refere-se à eficiência e eficácia dos serviços que a aplicação fornecerá para melhorar os processos suportados. Os serviços do fornecedor SaaS podem ter uma qualidade melhor, devido à maior maturidade de seus processos internos e ao aproveitamento dos recursos de escalabilidade da nuvem;

- **Desempenho:** refere-se à possibilidade de que a alternativa possa não entregar o nível de serviço esperado, e é um fator importante a ser considerado quando há deficiências na infraestrutura de rede (internet);
- **Riscos de Segurança:** refere-se à possibilidade de corrupção ou vazamento de dados, erros de autenticação, entre outras ameaças à segurança. Esse é de longe o inibidor mais importante da adoção de SaaS;
- **Riscos econômicos:** refere-se à possibilidade de o cliente ter que pagar mais para atingir o nível esperado de serviço do que o inicialmente previsto – os chamados "custos escondidos".

O custo foi deliberadamente excluído da avaliação qualitativa, para evitar reações políticas e emocionais que poderiam influenciar inadequadamente a atribuição de pesos aos critérios qualitativos, como recomendado em (Haas & Meixner, 2013). Por simplicidade, não foram incluídos critérios considerados pouco significativos nos estudos anteriores, tais como 1) acesso a recursos especializados (humanos e tecnológicos) não disponíveis na instituição, e 2) riscos gerenciais, como efeitos negativos na reputação e na carreira do gerente de TI responsável pela aplicação se o sistema é hospedado em um fornecedor externo (Benlian & Hess, 2011).

Tabela 8 – pesos de alternativas (parâmetro de entrada) Tabela 9 - pesos de critérios (parâmetro de entrada)

Critério	Alternativa	Peso
Estratégia	<i>On-premises</i>	0.357
Estratégia	Nuvem	0.643
Qualidade	<i>On-premises</i>	0.42
Qualidade	Nuvem	0.58
Riscos de desempenho	<i>On-premises</i>	0.643
Riscos de desempenho	Nuvem	0.357
Riscos de segurança	<i>On-premises</i>	0.75
Riscos de segurança	Nuvem	0.25
Riscos econômicos	<i>On-premises</i>	0.9
Riscos econômicos	Nuvem	0.1

Fonte: a autora

Critério	Peso
Estratégia	0.15
Qualidade	0.2182
Riscos de desempenho	0.2417
Riscos de segurança	0.2077
Riscos econômicos	0.1821

Fonte: a autora

Para obter os pesos dos critérios e alternativas, que são parâmetros de entrada do modelo, gerentes experientes foram consultados para realizar a comparação de pares. Os índices de consistência CI e taxa CR foram calculados conforme propõe o método AHP (Haas & Meixner, 2013), resultando $CR < 0.1$, aceitável. As Tabelas 8 e 9 apresentam os pesos calculados. A transição “MCDM” (Figura 13 – passo 2) pode então calcular as avaliações qualitativas, obtendo os valores:

- Avaliação qualitativa nuvem (qc): 0.38

- Avaliação qualitativa local (qo): 0.62

4.4.1.3 Passo 3: Calcular a razão custo/benefício

A transição “Cálculo e Normalização” (Figura 13 – passo 3) pode então calcular o TCO para o primeiro ano e também para cinco anos de uso, a partir dos dados obtidos no passo 1. Em seguida, procede-se com a normalização e por fim o cálculo da razão custo/benefício (fator qualitativo / TCO normalizado), obtendo o resultado final do modelo, a avaliação das duas alternativas, no primeiro ano e após cinco anos de uso, conforme ilustrado na Tabela 10.

Tabela 10 - Cálculo da avaliação final das alternativas (valores em milhares de reais)

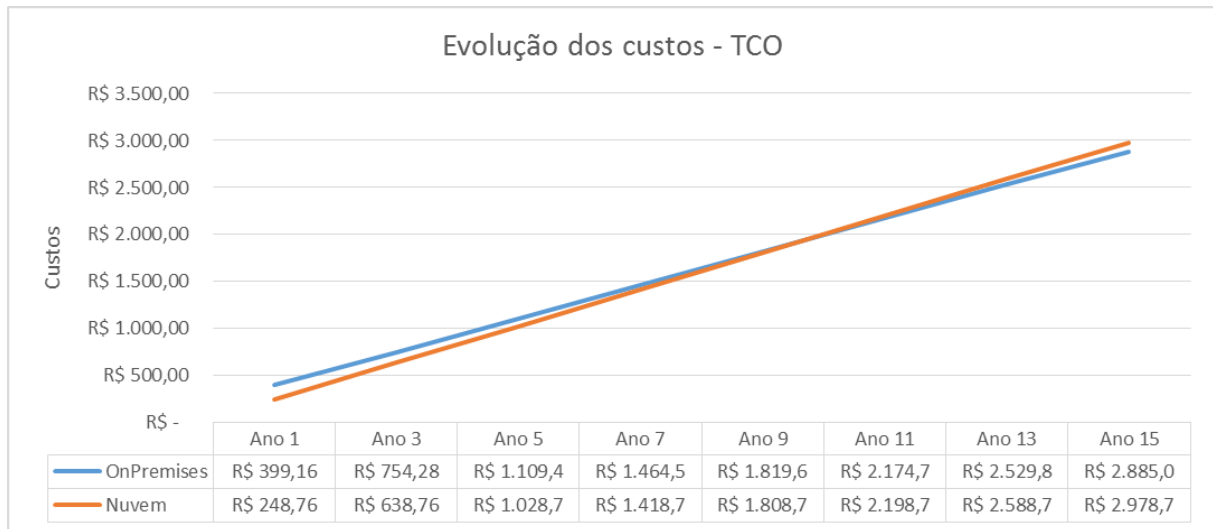
			Ano 1		Após 5 anos		Fator qualitativo	Avaliação Final	
	Inicial	Anual	TCO	Normalizado	TCO	Normalizado		Ano 1	Após 5 anos
<i>On-premises</i>	221,60	177,56	399,16	61,61%	1.109,40	51,89%	62%	1,0064	1,1949
Nuvem	53,76	195,00	248,76	38,39%	1.028,76	48,11%	38%	0,9897	0,7898

Fonte: a autora

A avaliação final da alternativa nuvem é bem similar à da alternativa *on-premises* no primeiro ano devido aos altos custos iniciais da opção *on-premises*. No entanto, na análise após cinco anos, a opção *on-premises* recebe uma avaliação bem maior. Esse exemplo reflete a avaliação de gestores brasileiros na época da realização da pesquisa (2013), pois no Brasil encontram-se altos custos de internet, custo mediano de salários, e uma forte percepção de riscos econômicos e de segurança. De fato, a análise qualitativa foi a maior barreira para a avaliação da alternativa nuvem. A vantagem obtida com menores custos não foi suficiente para reverter essa situação. No exemplo, como o fator qualitativo obtido foi de 62% para a opção *on-premises*, para que a alternativa nuvem tivesse uma avaliação superior à da *on-premises*, seus custos normalizados deveriam ser 61%, ou menos.

Além do resultado final do modelo, a construção do mesmo permite observar a evolução do custo das duas alternativas, conforme ilustrado na Figura 16. Os custos são inicialmente mais baixos para a alternativa nuvem, mas, após cinco anos, os custos se tornam bastante similares. Como no exemplo utilizado, a função custo adotada é o TCO, que é linear, ao longo do tempo, os custos da nuvem serão maiores que os da alternativa local.

Figura 16 - Evolução dos custos utilizando TCO



Fonte: a autora

No entanto, se a função utilizada no *framework* (alterando a função ScoresCloud do *framework*) fosse, por exemplo, valor líquido atual (*net present value* – NPV), a evolução dos custos teria uma avaliação diferente.

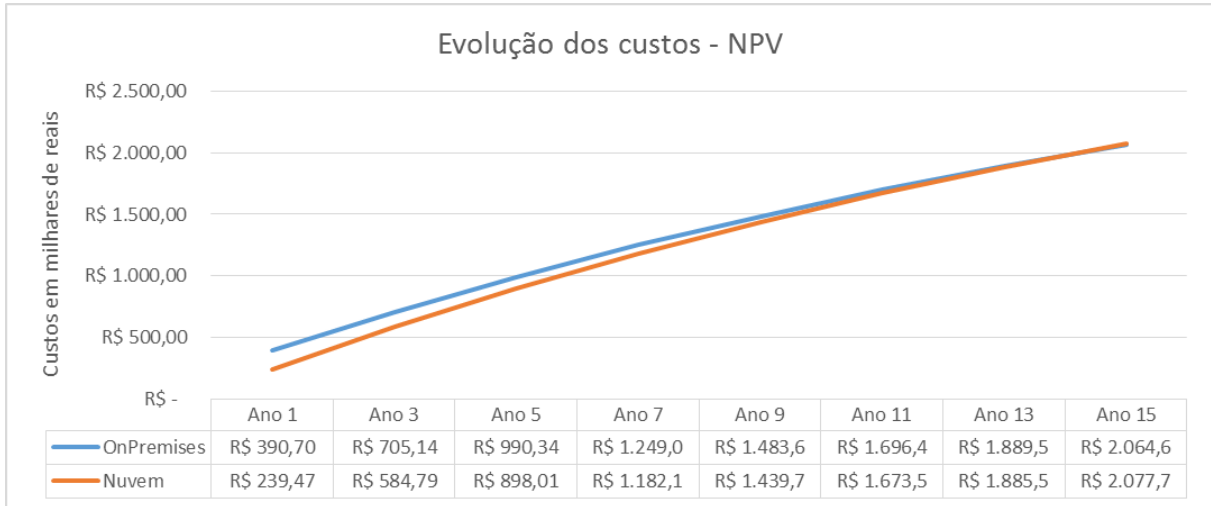
$$NPV = \sum_{t=0}^{Y-1} \frac{c(t)}{(1+k)^t}$$

NPV é definido pela fórmula acima, onde $c(t)$ representa o custo total do investimento no ano t , ao longo de Y anos no futuro. k é o custo anual do capital, dessa forma o capital investido no ano atual é atualizado para o ano seguinte pelo fator $(1+k)$. Usando esse método, que também tem sido utilizado na literatura (Sripanidkulchai & Sujichantararat, 2012) em comparações de custo entre serviços em nuvem e serviços *on-premises*, a evolução dos custos não será linear. No exemplo utilizado, adotando $k=5\%$, tem-se uma evolução onde o custo da nuvem é menor do que o custo da alternativa *on-premises*, conforme ilustrado na Figura 17.

Nesse exemplo, a análise da evolução dos custos usando mais de uma função de cálculo de custos confirma que, de fato, a avaliação qualitativa continua sendo a barreira para a adoção. A mudança de método acentua a vantagem econômica da alternativa nuvem, porém não o suficiente para que a avaliação final seja favorável a ela. A vantagem da nuvem é pequena, especialmente se considerado o panorama de cinco anos de uso. Na Figura 18, nota-se que após

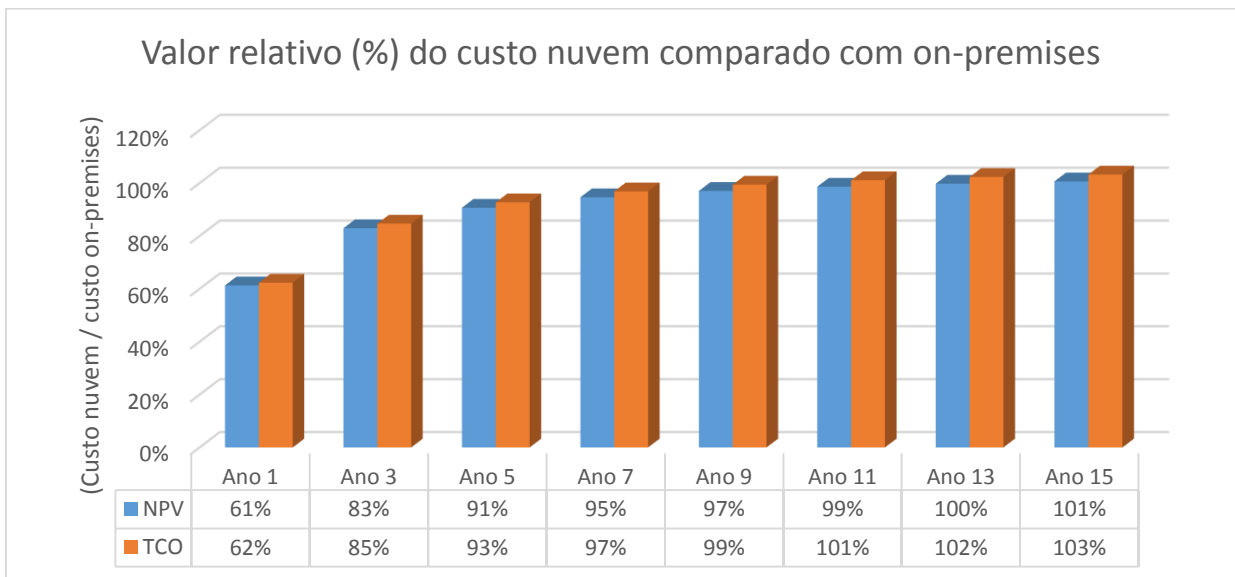
esse período, os custos são bem próximos em qualquer um dos métodos avaliados (TCO ou NPV).

Figura 17 - Evolução dos custos usando NPV



Fonte: a autora

Figura 18 - Comparação de custos em percentual no modelo SaaS e on-premises



Fonte: a autora

4.4.2 Exemplo PaaS

Para o cenário PaaS, serão utilizados os parâmetros de custo apresentados na ferramenta fornecida pelo trabalho de (InfoTech Research Group, 2015). Essa ferramenta é uma planilha excel, que elenca os principais elementos de custo necessários para se avaliar o Custo Total de Propriedade de um banco de dados instalado localmente, comparado com um banco de dados

instalado em nuvem pública. A Tabela 11 contém, a título de exemplo, os valores financeiros de uma comparação realizada entre o banco SQL Server e o seu equivalente na nuvem pública da Microsoft (Windows Azure).

Para a parte qualitativa, foram utilizadas as respostas obtidas na pesquisa adicional de qualidade, que serão detalhadas no Apêndice B – Pesquisa Qualitativa. Nesta pesquisa, os critérios foram resumidos para facilitar a coleta dos dados. Foi selecionado um respondente para exemplificar o uso do *framework* com base nas suas respostas.

Na comparação entre pares, o respondente indicou:

- Estratégia um pouco mais importante que Qualidade (8/5)
- Qualidade tão importante quanto Risco (6/6)
- Risco mais importante que Estratégia (7/4)

Na comparação entre alternativas, o respondente indicou:

- Em relação à estratégia, leve preferência pelo ambiente local (6/7)
- Em relação à qualidade, forte preferência pela nuvem (7/4)
- Em relação ao risco, indiferente (8/8)

Tabela 11 - Elementos de custo PaaS

Elementos de custo local		
Inicial	Aquisição de <i>hardware</i>	US\$ 4,330.00
Inicial	Aquisição de <i>software</i> – S.O	US\$ 40,482.75
Anual	Suporte ao <i>hardware</i> (garantia estendida proporcionalizada)	US\$ 600.00
Anual	Suporte ao <i>software</i>	US\$ 4,389.60
Anual	Custos de energia	US\$ 1,240.00
Anual	DBA	US\$ 76,000.00
Elementos de custo nuvem		
Anual	<i>Databases</i>	US\$ 5,251.34
Anual	Transferência de Dados	US\$ 1,432.80
Anual	Suporte Microsoft	US\$ 12,000.00
Anual	DBA	US\$ 25,000.00
Anual	Conectividade	US\$ 1,200.00

Fonte: (InfoTech Research Group, 2015)

Após montar as matrizes e obter os autovetores, foi possível obter os pesos, conforme apresentados nas Tabelas 12 e 13, que foram utilizados como entrada para o método. Note que o público entrevistado para esse exemplo não apresenta uma preferência forte por nenhuma das opções. Essa diferença pode ser devido ao perfil dos gestores, que não era o mesmo, ou ao intervalo de tempo decorrido entre a pesquisa SaaS, que foi bem anterior às demais.

Tabela 12 - Pesos de alternativas (parâmetros de entrada)

Critério	Alternativa	Peso
Estratégia	On-premises	0.538
Estratégia	Nuvem	0.462
Qualidade	On-premises	0.364
Qualidade	Nuvem	0.636
Riscos	On-premises	0.5
Riscos	Nuvem	0.5

Fonte: a autora

Tabela 13 - Pesos de critérios (parâmetro de entrada)

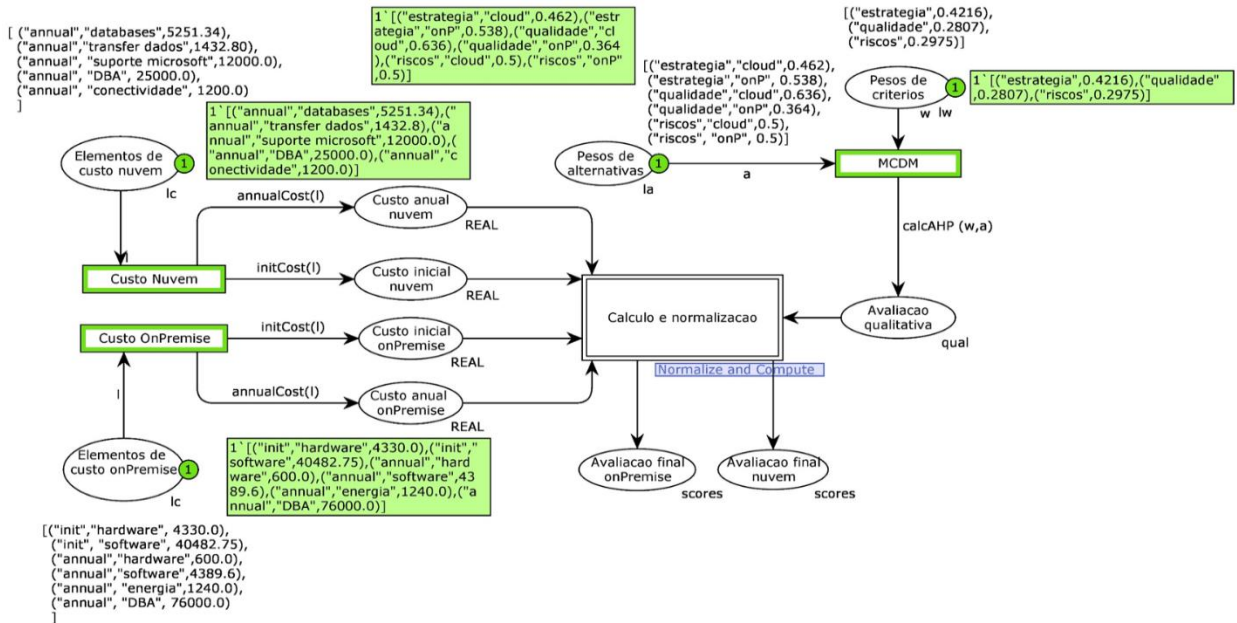
Critério	Peso
Estratégia	0,4216
Qualidade	0,2807
Riscos	0,2975

Fonte: a autora

Para esse exemplo, o modelo proposto será detalhado ainda mais, de modo a explicitar alguns dos cálculos realizados pela transição “Cálculo e Normalização”, para facilitar a visualização dos cálculos executados.

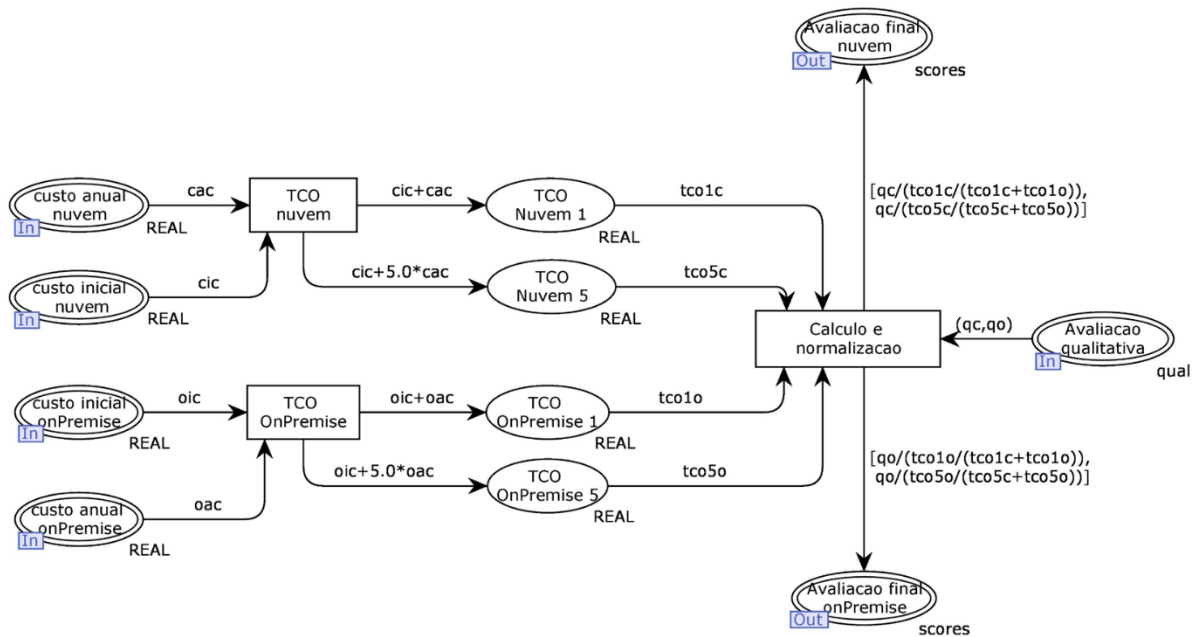
Dessa forma, a transição “Cálculo e Normalização”, responsável pelos cálculos referentes ao Passo 3 do modelo, foi transformada em uma transição de substituição, criando um modelo hierárquico como ilustrado nas Figuras 19 e 20.

Figura 19 - Rede modificada com transição de substituição para cálculos de TCO



Fonte: a autora

Figura 20 - Transição de substituição que explicita cálculos TCO



Fonte: a autora

Nesse modelo, que apenas detalha os cálculos de modo a permitir o acompanhamento visual da simulação, a transição “Cálculo e Normalização” do modelo original foi substituída pela sub-rede “Cálculo e Normalização” descrita na Figura 20. Nessa sub-rede, os cálculos de TCO, que na rede original ficam encapsulados nas funções “PontosNuvem” e “PontosOnP”, ficam explícitos nos lugares e transições:

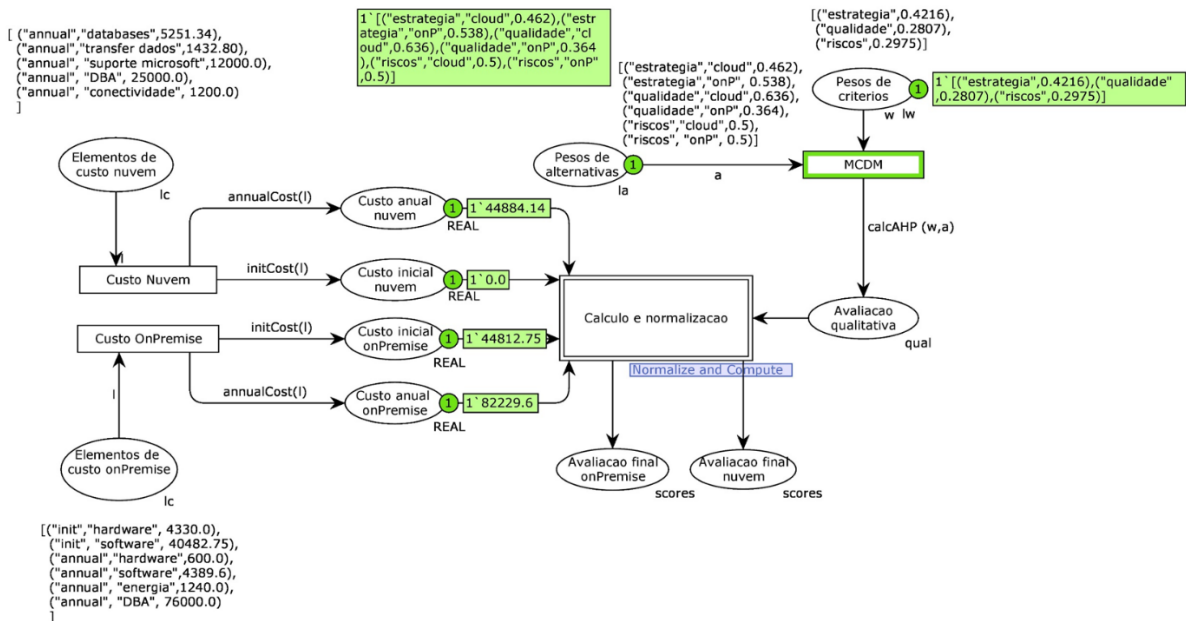
- Transições “TCO Nuvem” e “TCO on-premises”: calculam o TCO para um e cinco anos, somando o custo inicial e o custo anual, multiplicado pelo número adequado de anos;
- Lugares “TCO Nuvem 1” e “TCO Nuvem 5”: armazenam temporariamente o valor de TCO para o Ano 1 e para cinco anos, para a alternativa nuvem;
- Lugares “TCO on-premises 1” e “TCO on-premises 5”: armazenam temporariamente o valor de TCO para o Ano 1 e para cinco anos, para a alternativa on-premises;
- Transição “Cálculo e Normalização”: implementa as fórmulas (descritas nos arcos) para calcular os scores das opções nuvem e on-premises.

Dessa forma, pode-se visualizar a simulação da rede com os parâmetros de entrada que serão utilizados para o cenário PaaS, em cada um dos passos.

4.4.2.1 Passo 1: Estimativa de custos

Esse passo é realizado pelas transições “Custo Nuvem” e “Custo On-premises”, que estão habilitadas conforme indicado na Figura 19. Ao disparar tais transições, a rede fica na configuração ilustrada na Figura 21.

Figura 21 - Visualização da Rede após execução do Passo 1



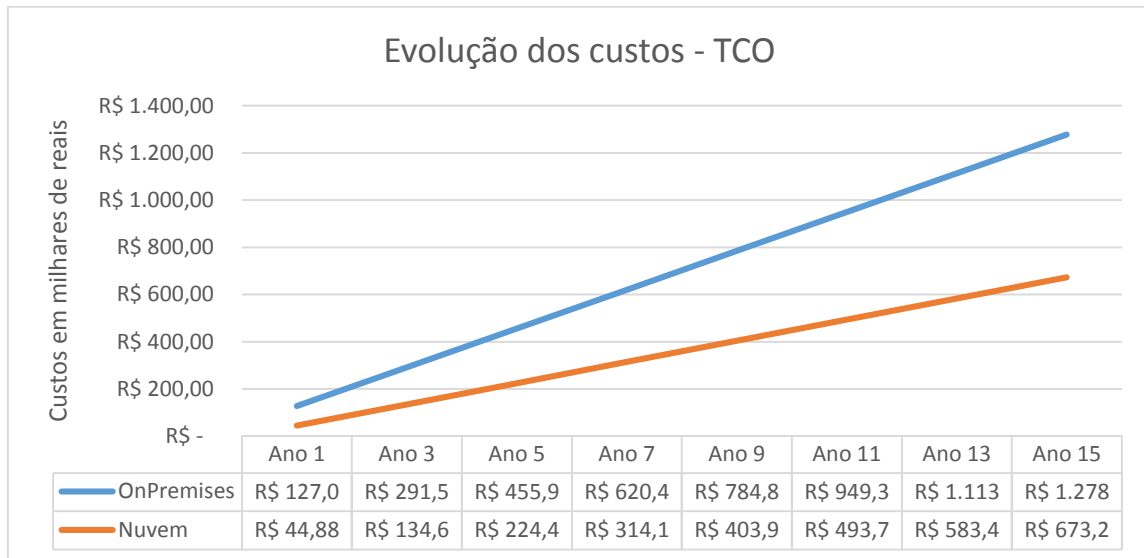
Fonte: a autora

Nesse momento, a rede calculou o custo inicial e anual da utilização dos dois modelos (nuvem e *on-premises*) e armazenou as informações nos locais "Custo anual nuvem", "Custo inicial nuvem", "Custo anual *on-premises*" e "Custo inicial *on-premises*". No nosso exemplo, com base nos elementos de custo listados na Tabela 10, foram calculados os custos para nuvem: inicial zero, anual US\$ 44,884.14, e para *on-premises*: inicial US\$ 44,812.75 e anual US\$ 82,229,60.

Note-se que o alto custo anual da opção *on-premises* é altamente afetado pelo custo da mão de obra do DBA, que representa US\$ 76,000.00 anuais. Alguns autores já destacaram a importância do custo da mão de obra nas comparações de custo nuvem x *on-premises*. No trabalho de (Sripanidkulchai & Sujichantararat, 2012), os cenários de avaliação envolvem os Estados Unidos e Tailândia, e os autores relatam que, devido ao menor custo da mão de obra na Tailândia, o modelo de nuvem passa a ser bem menos competitivo do que nos Estados Unidos, onde o custo da mão de obra mais alta acaba deixando a opção nuvem mais atraente, devido ao compartilhamento dos recursos humanos.

Outra análise interessante que pode ser feita após esses cálculos é a evolução dos custos, conforme realizado no exemplo SaaS, para verificar a evolução dos custos usando a metodologia TCO e também NPV, utilizando como base as informações de custo inicial e anual de cada uma das alternativas. As Figuras 22 e 23 ilustram a evolução dos custos em ambas as metodologias.

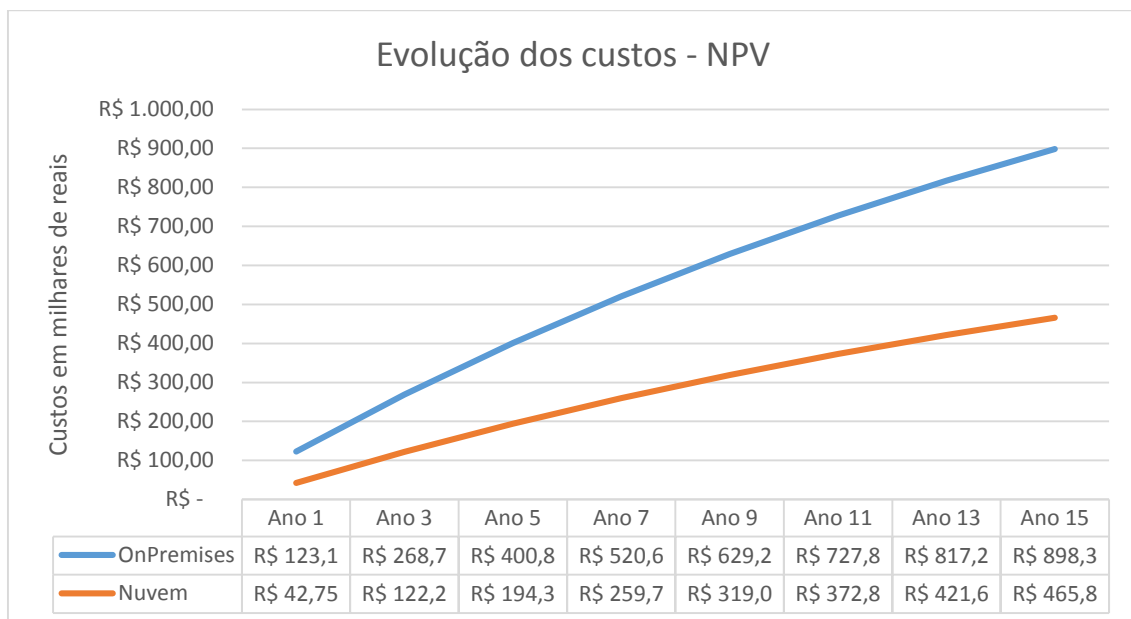
Figura 22 - Evolução dos custos no exemplo PaaS usando TCO



Fonte: a autora

Pode-se visualizar que, nesse exemplo, os custos da nuvem seguem sempre inferiores aos da opção local, qualquer que seja a metodologia financeira utilizada.

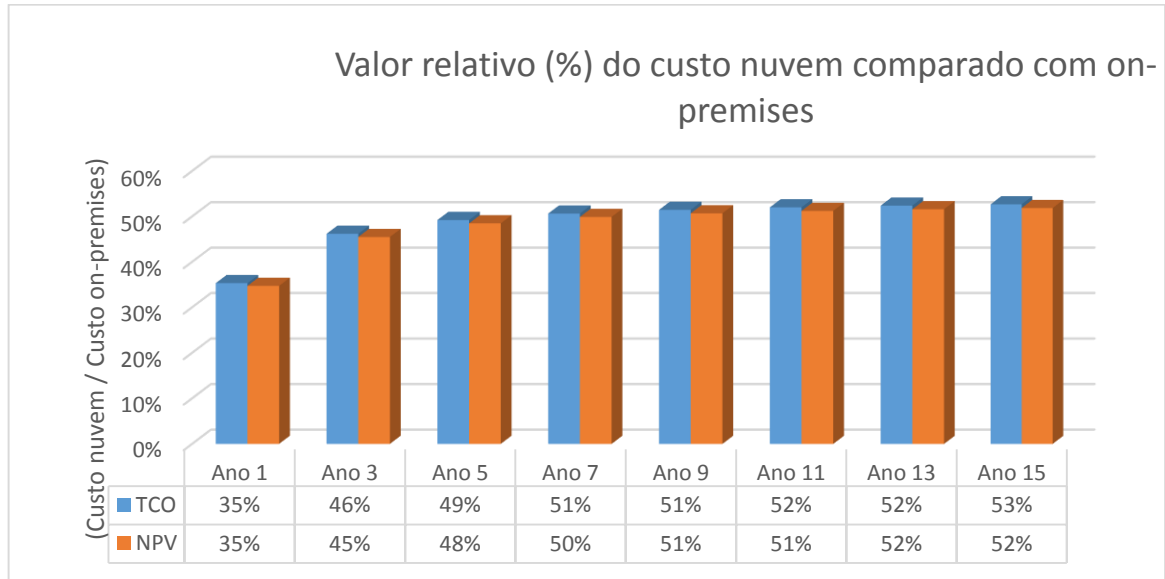
Figura 23 - Evolução dos custos no exemplo PaaS usando NPV



Fonte: a autora

O gráfico comparativo de percentuais (Figura 24) ilustra que o custo da opção nuvem é muito superior no primeiro ano de uso (35%), ficando próximo da metade do custo da opção *on-premises* praticamente por todo o período da análise, em qualquer uma das metodologias adotadas (TCO e NPV).

Figura 24 - Comparação de custos em percentual no modelo PaaS

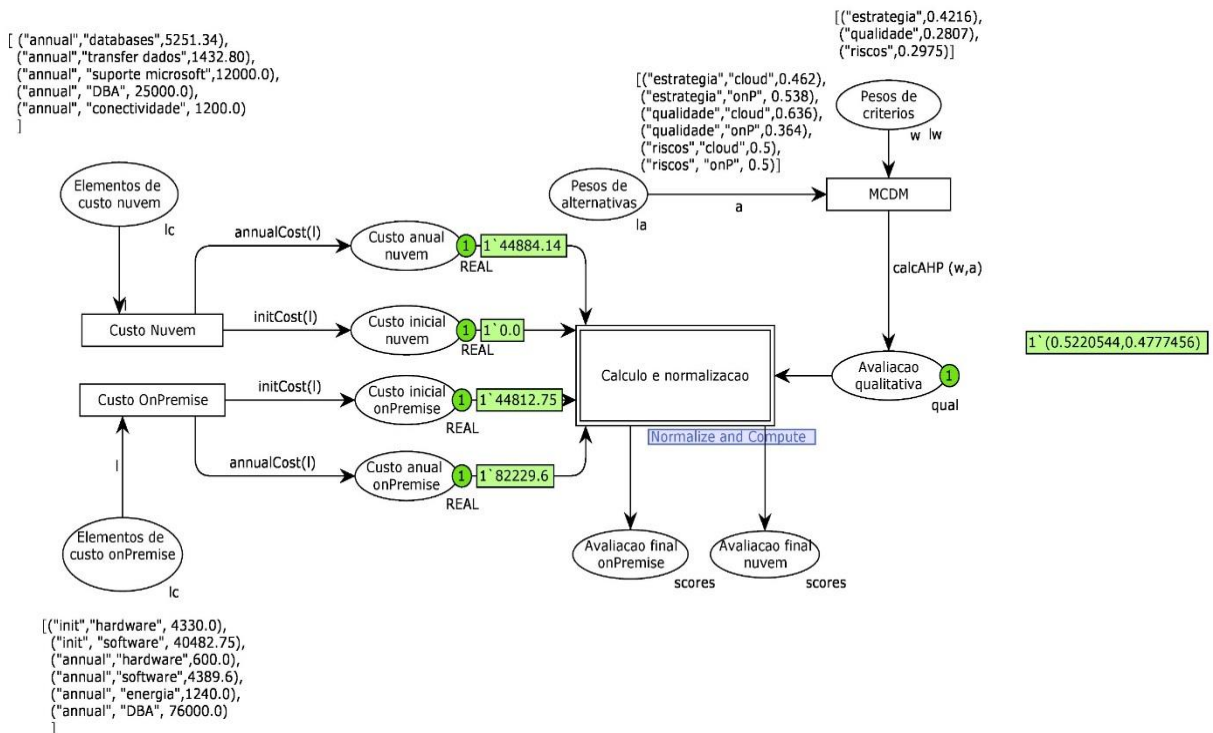


Fonte: a autora

4.4.2.2 Passo 2: Avaliação qualitativa

Esse passo é realizado pela transição “MCDM”, que está habilitada conforme indicado na Figura 22. Ao disparar tal transição, a rede fica no estado ilustrado na Figura 25.

Figura 25 – Visualização da rede de exemplo PaaS após execução do Passo 2



Fonte: a autora

Nesse momento, a rede calculou a avaliação qualitativa de acordo com os parâmetros informados, através da função calcAHP (pesos, alternativas), que basicamente efetua uma operação matricial que multiplica a matriz de pesos de alternativas pelo vetor de pesos de critérios. Nesse exemplo, essa operação matricial é ilustrada na Figura 26.

Figura 26 - Cálculo final do método AHP para o exemplo PaaS

	Estratégia	Qualidade	Risco	X	Estratégia	0.4216	=	Nuvem	0.522
Nuvem	0.462	0.636	0.5		Qualidade	0.2807		On Premises	0.478
On Premises	0.538	0.364	0.5		Risco	0.2975			

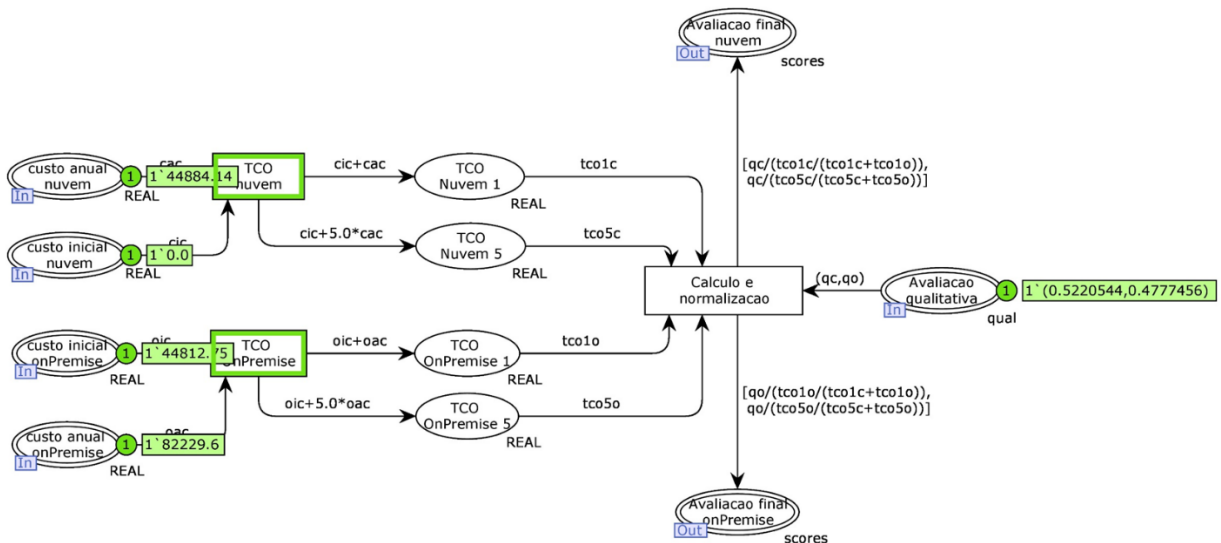
Fonte: a autora

4.4.2.3 Passo 3: Calcular a razão custo/benefício

Esse passo é realizado pela transição de substituição “Cálculo e Normalização” ilustrada na Figura 28. Nesse momento da simulação, as informações de custo já foram processadas e encontram-se nos lugares ‘Custo anual nuvem’, ‘Custo inicial nuvem’, ‘Custo inicial on-premises’, ‘Custo anual on-premises’, conforme ilustrado na Figura 27. Assim, as seguintes transições estão habilitadas:

- TCO nuvem: calcula o TCO para o primeiro ano e para cinco anos de uso para a alternativa nuvem, colocando os valores calculados nos lugares TCO Nuvem 1 e TCO Nuvem 5. As operações de cálculo estão descritas nos arcos de saída.
- TCO on-premises: calcula o TCO para o primeiro ano e para cinco anos de uso para a alternativa on-premises, colocando os valores calculados nos lugares TCO on-premises 1 e TCO on-premises 5. As operações de cálculo estão descritas nos arcos de saída.

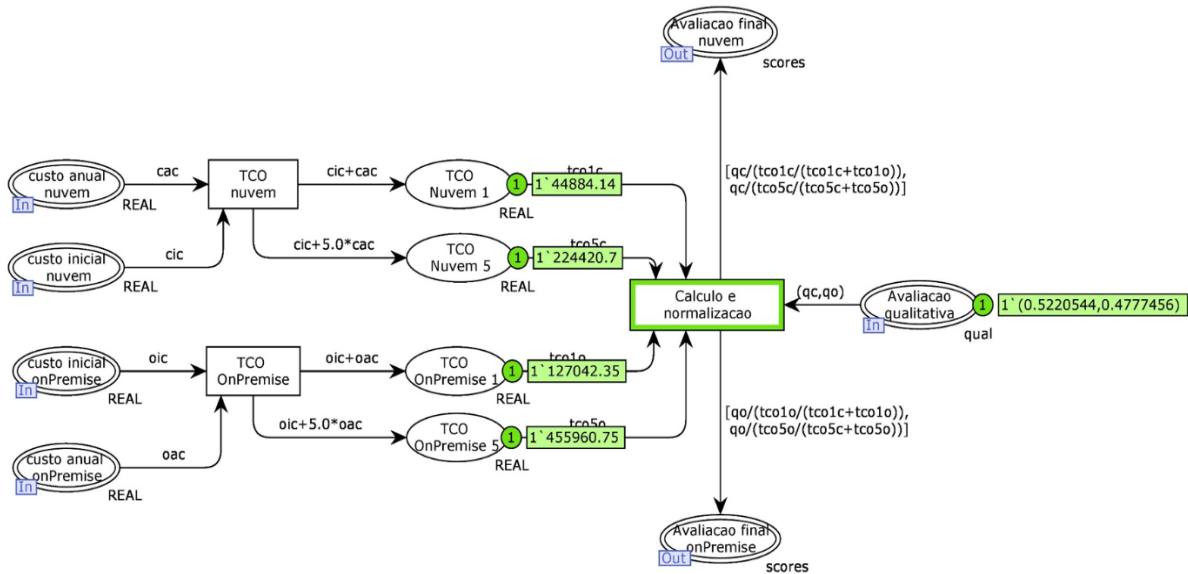
Figura 27 - Transição de substituição “Cálculo e Normalização” para o exemplo PaaS - estado inicial



Fonte: a autora

Finalmente, após execução dessas transições, os valores de TCO estão calculados e a transição “Cálculo e Normalização” estará habilitada para calcular a razão custo/benefício, conforme ilustrado na Figura 28. É importante ressaltar a necessidade de normalizar os custos (de modo a obter um valor entre 0 e 1) para que a razão custo/benefício seja compatível, já que a avaliação qualitativa é também um valor entre 0 e 1.

Figura 28 - Transição de substituição “Cálculo e Normalização” para o exemplo PaaS - estado intermediário



Fonte: a autora

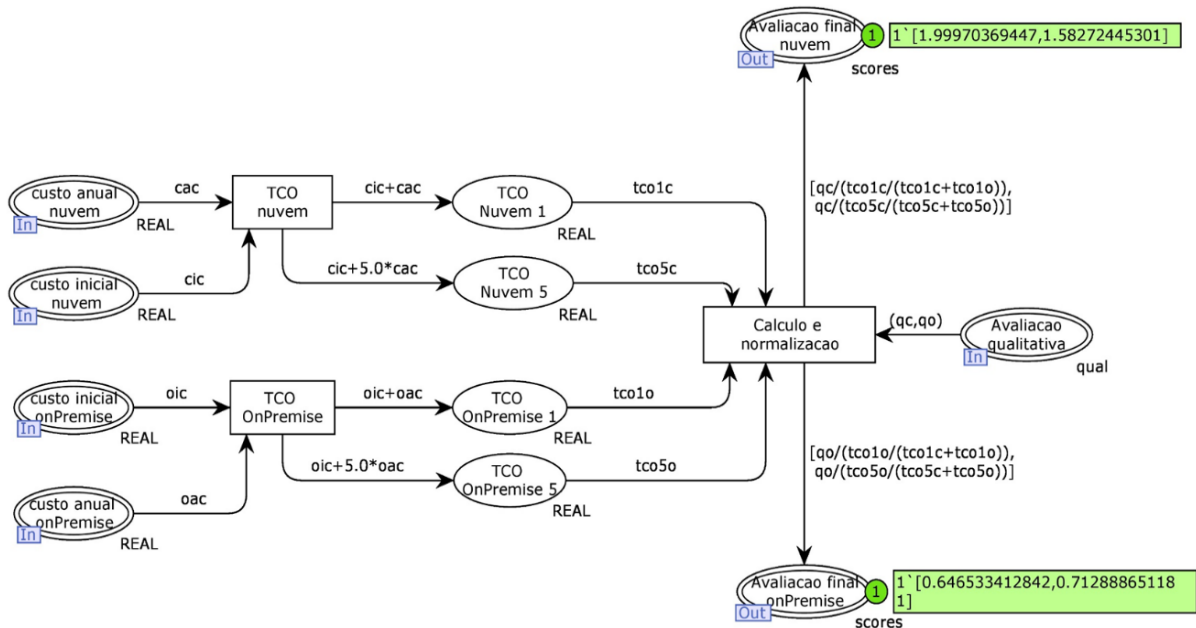
Nesse ponto, a transição “Cálculo e Normalização”, da rede principal, vai realizar o cálculo desejado, conforme descrito nos arcos de saída, e colocando o resultado final da simulação nos lugares ‘Avaliação Final Nuvem’ e ‘Avaliação Final *on-premises*’, que representam a saída final do modelo.

A Figura 29 ilustra o resultado final da simulação, onde foram calculados:

- Para a opção Nuvem: Score 1.999 para o primeiro ano e 1.588 após cinco anos de uso
- Para a opção On-premises: Score 0.646 para o primeiro ano e 0.713 após cinco anos de uso

Nesse exemplo, a avaliação final foi muito mais favorável para o serviço de nuvem, tanto para o primeiro ano, quanto para cinco anos de uso. Não houve uma diferença tão grande na avaliação qualitativa como no exemplo SaaS (o gestor entrevistado não foi o mesmo), por esse motivo a grande vantagem nos custos apurados no modelo nuvem (devida, em grande parte, ao custo com recursos humanos) se refletiu diretamente na avaliação final.

Figura 29 - Transição de substituição “Cálculo e Normalização” para o exemplo PaaS - final da simulação



Fonte: a autora

4.4.3 Exemplo IaaS

Para exemplificar o uso do modelo no cenário IaaS, serão utilizados os parâmetros de custo definidos no trabalho de (Principled Technologies, 2013). O trabalho compara custos de uma instalação local de máquinas virtuais para hospedagem de bancos de dados utilizando a solução integrada Dell PowerEdge VRTX, projetada especificamente para escritórios remotos e ambientes de escritórios pequenos. A solução integra servidores, armazenamento, rede e gerenciamento em um único chassi compacto com dimensão, acústica e segurança otimizadas para pequenos escritórios (Dell, 2015).

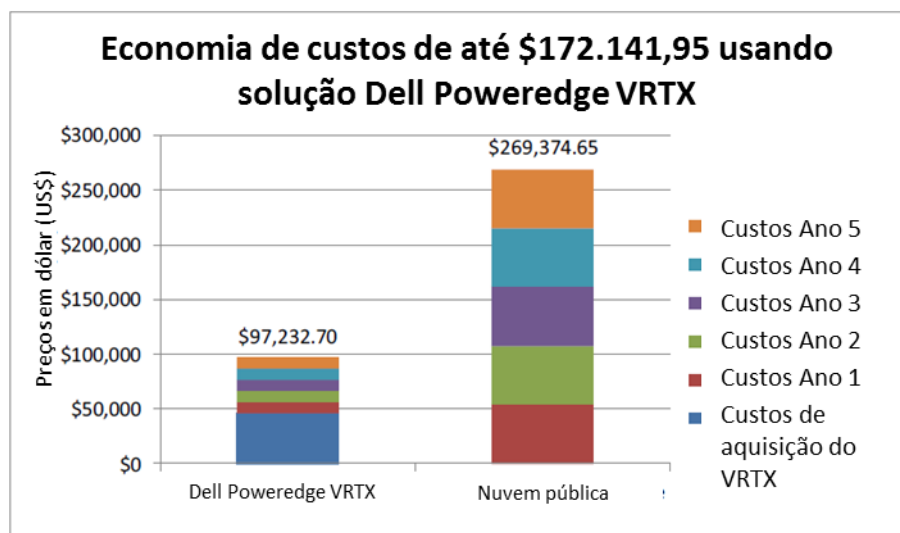
No teste realizado para o cenário local, foi utilizado o *software* de virtualização VMWare vSphere 5 para criar 10 máquinas virtuais na solução PowerEdge VRTX. Em cada uma das VMs foi instalada uma instância do gerenciador de banco de dados Microsoft SQL Server 2012. Para o cenário de nuvem, foi configurada uma instância semelhante de nuvem pública (EC2 da AWS) com 1 vCPU, 3.75GB RAM, e o menor volume de dados que oferecia 1,000 IOPS de desempenho no armazenamento dos dados, para provisionar a vazão necessária nos momentos de pico durante a carga de execuções de transações no banco de dados. Em ambos os cenários (local e nuvem), o sistema operacional utilizado foi Windows Server 2012,

atualizado até a correção de junho de 2013. O *software* gerenciador de bancos de dados SQL Server 2012 também foi atualizado com o Service Pack 1 nos dois cenários.

O teste comparativo realizado executava uma carga de transações da ordem de 1,000 solicitações por minuto. Essa carga representava o uso simultâneo de 250 usuários gerando aproximadamente quatro solicitações por minuto (SPM) em cada uma das VMs. Esse teste foi executado tanto na instância na nuvem como no ambiente local, e foi o parâmetro que norteou o uso de 10 VMs na solução local.

O custo total de propriedade calculado no trabalho foi bem inferior na solução local, conforme ilustrado na Figura 30.

Figura 30 - Comparação do Custo Total de Propriedade



Fonte: (Principled Technologies, 2013)

O custo total de propriedade das duas soluções foi baseado no custo de aquisição e suporte de *hardware* e *software*, além de custos de gerenciamento e de energia. Não foram considerados custos referentes a espaço físico e refrigeração, pois, segundo os autores, o equipamento tem um formato pequeno que permite alocação em pequenos espaços tais como debaixo de uma escrivaninha ou em um armário. Os elementos de custo, dados de entrada para o nosso *framework*, foram descritos na Tabela 14.

Para a parte qualitativa, foram utilizadas as respostas obtidas na pesquisa adicional de qualidade, que serão descritas em detalhe no capítulo 5 – Análise Qualitativa – Pesquisas Adicionais. Nessa pesquisa, os critérios foram resumidos para facilitar a coleta dos dados. Foi selecionado outro respondente (diferente dos utilizados nos exemplos anteriores) para exemplificar o uso do *framework* com base nas suas respostas.

Tabela 14 - Elementos de custo: parâmetros iniciais do modelo IaaS

Elementos de custo local		
Inicial	Aquisição de <i>hardware</i>	US\$ 22,710.00
Inicial	Aquisição de <i>software</i>	US\$ 23,380.00
Anual	Suporte ao S.O. (<i>software assurance</i> for Microsoft Windows Server 2012 Datacenter Edition)	US\$ 2,045.00
Anual	Suporte ao banco de dados (<i>software assurance</i> for Microsoft Windows SQL Server 2012 Standard)	US\$ 2,694.00
Anual	Suporte ao sistema de virtualização (VMWare vSphere 5 Standard Basic Suport)	US\$ 1,092.00
Anual	Suporte ao <i>hardware</i>	US\$ 1,300.00
Anual	Custos de energia	US\$ 937.54
Anual	Gerenciamento de <i>hardware</i> (Pessoal de TI)	US\$ 600.00
Anual	Gerenciamento da virtualização – VMWare (Pessoal de TI)	US\$ 1,200.00
Elementos de custo nuvem		
Anual	Custo da reserva da instância	US\$ 5,133.33
Anual	Custo das instâncias	US\$ 48,741.60

Fonte: (Principled Technologies, 2013)

Na comparação entre pares, o respondente indicou:

- Estratégia um pouco menos importante que Qualidade (5/7)
- Qualidade um pouco mais importante que Risco (8/6)
- Risco um pouco mais importante que Estratégia (6/5)

Na comparação entre alternativas, o respondente indicou:

- Em relação à estratégia, leve preferência pela nuvem (7/6)
- Em relação à qualidade, leve preferência pela nuvem (9/8)
- Em relação ao risco, forte preferência pelo ambiente local (4/7)

Tabela 15 - Pesos de alternativas (parâmetros de entrada)

Critério	Alternativa	Peso
Estratégia	<i>on-premises</i>	0.462
Estratégia	Nuvem	0.538
Qualidade	<i>on-premises</i>	0.471
Qualidade	Nuvem	0.529
Riscos	<i>on-premises</i>	0.65
Riscos	Nuvem	0.35

Fonte: a autora

Tabela 16 - Pesos de critérios (parâmetros de entrada)

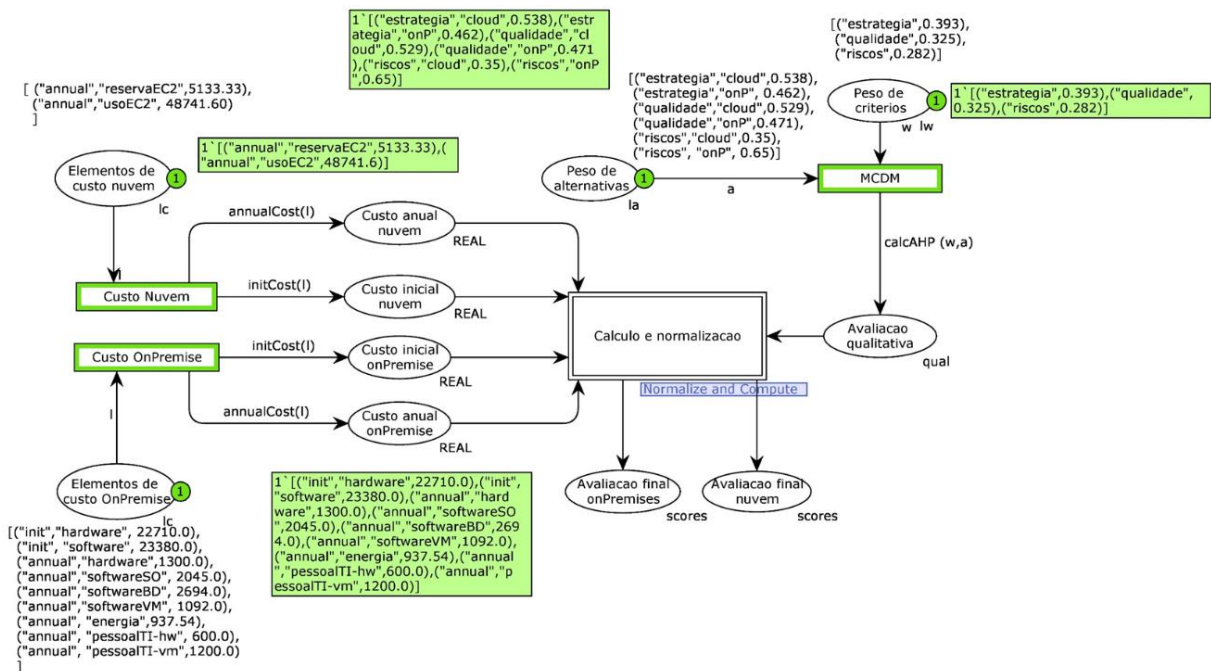
Critério	Peso
Estratégia	0.393
Qualidade	0.325
Riscos	0.282

Fonte: a autora

Após montar as matrizes e obter os autovetores, foi possível obter os pesos que foram utilizados como entrada para o método, conforme apresentados nas Tabelas 15 e 16.

Para esse exemplo, foi utilizada a mesma rede do exemplo PaaS, alterando apenas as marcações iniciais dos lugares que representam os parâmetros do modelo, ou seja, os elementos de custo e as matrizes de pesos de critérios e alternativas, conforme ilustrado na Figura 31.

Figura 31 - Início da simulação do exemplo IaaS



Fonte: a autora

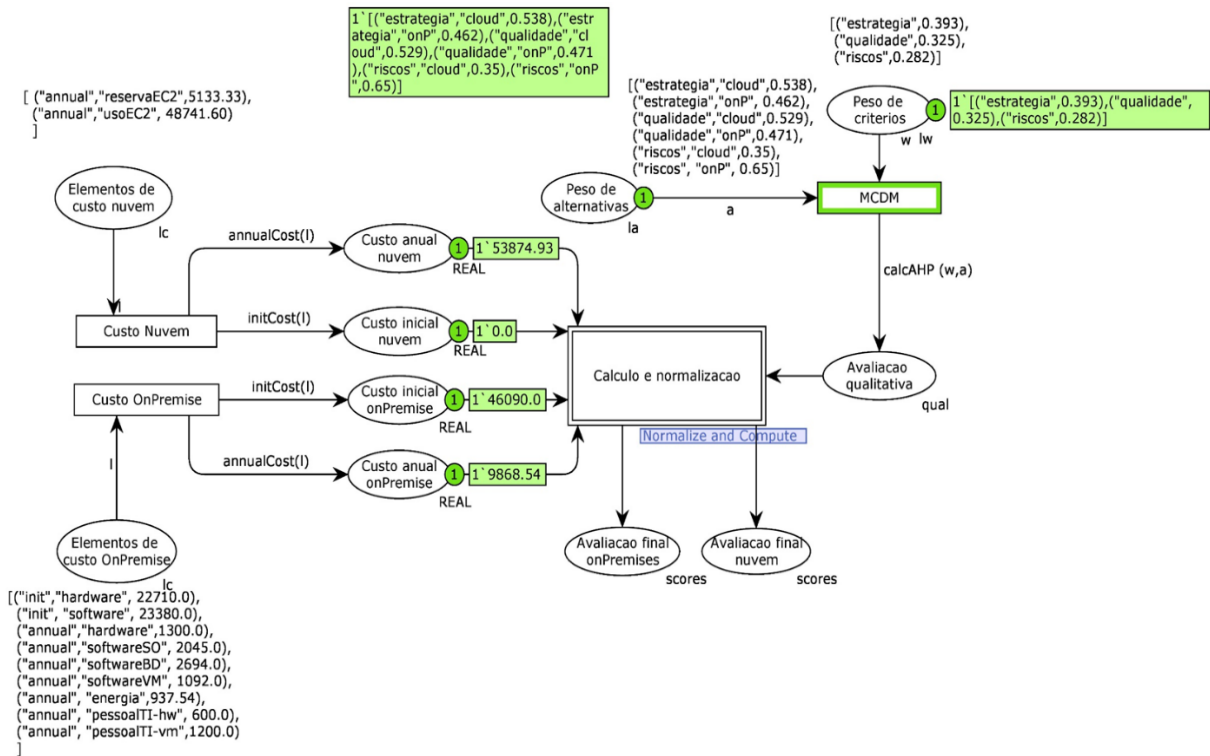
4.4.3.1 Passo 1: Estimativa de custos

Esse passo é realizado pelas transições ‘Custo nuvem’ e ‘Custo on-premises’, que estão habilitadas (destacadas com linhas verdes), conforme indicado na Figura 32. A marcação inicial dos lugares ‘Elementos de custo nuvem’ e ‘Elementos de custo on-premises’ refletem os dados de entrada do modelo descritos na Tabela 14.

Após a execução do passo 1, ou seja, após o disparo de tais transições, a rede fica com a configuração ilustrada na Figura 32. Os custos foram calculados e colocados nos locais:

- Nuvem – custos anuais (‘Cloud annual cost’): US\$ 53,874.93
- Nuvem – custo inicial (‘Cloud initial cost’): US\$ 0.00
- On-premises – custo inicial (‘On-premises initial cost’): US\$ 46,090.00
- On-premises – custo anual (‘On-premises annual cost’): US\$ 98,68.54

Figura 32 – Rede de exemplo IaaS após a execução do Passo 1



Fonte: a autora

Pode-se também analisar a evolução dos custos, conforme realizado nos exemplos anteriores, para verificar a evolução dos custos usando a metodologia TCO e também NPV, utilizando como base as informações de custo inicial e anual de cada uma das alternativas. As Figuras 33 e 34 ilustram a evolução dos custos em ambas as metodologias.

O gráfico comparativo de percentuais da Figura 35 ilustra que o custo da opção nuvem é similar no primeiro ano de uso (aproximadamente 90%), e supera em muito o custo da opção *on-premises* a partir do segundo ano de uso, chegando a quase o triplo do valor (300%) após cinco anos, em qualquer uma das metodologias adotadas (TCO e NPV).

Figura 33 - Evolução dos custos no exemplo IaaS usando TCO

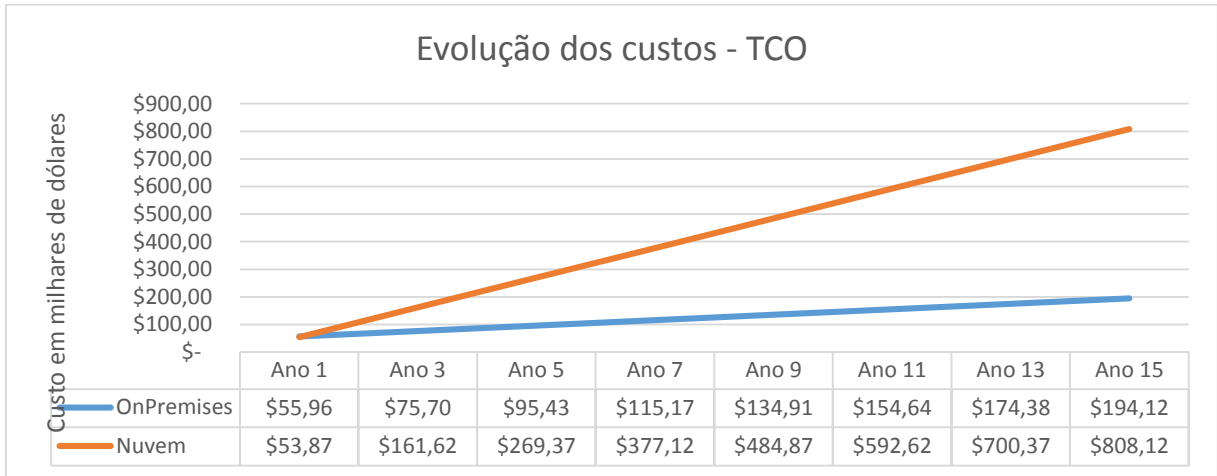


Figura 34 - Evolução dos custos no exemplo IaaS usando NPV

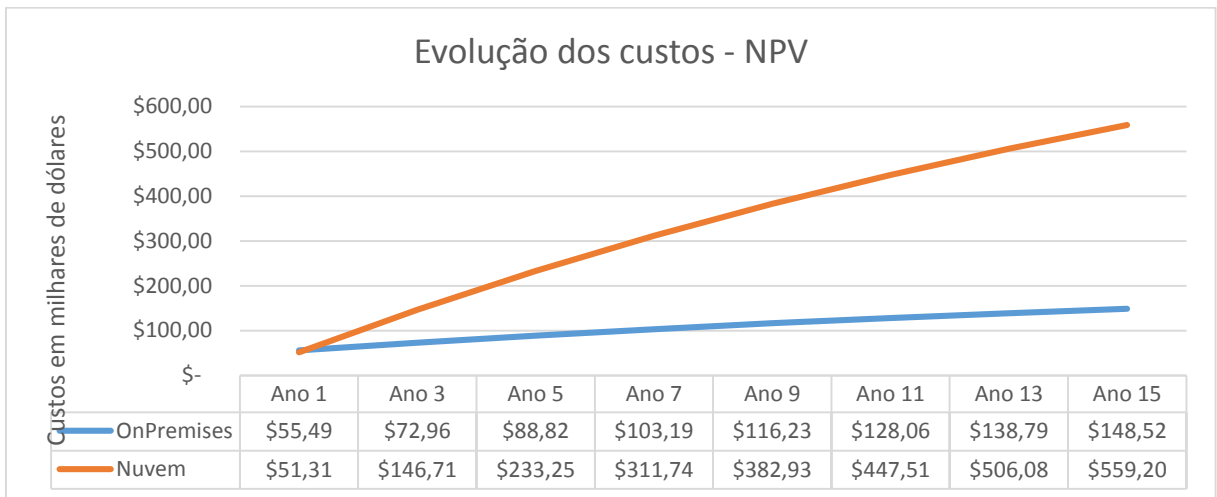
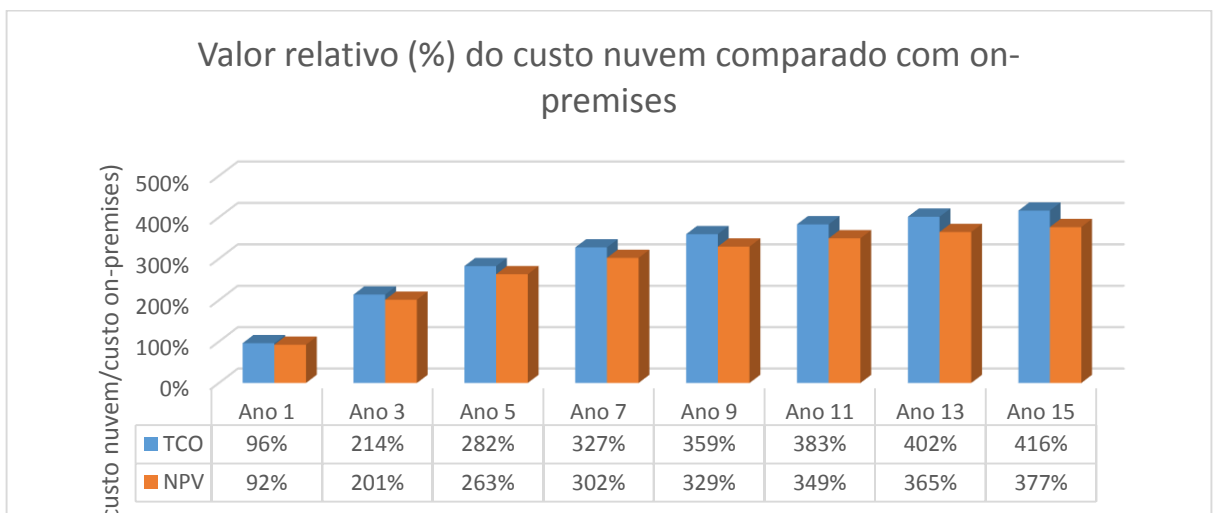
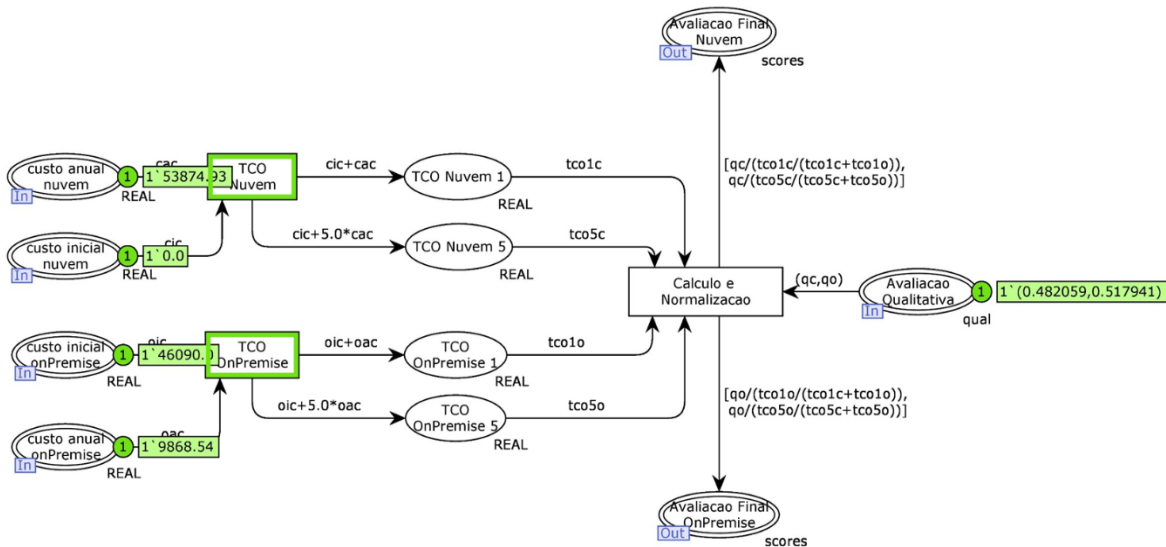


Figura 35 – Comparação de custos em percentual no modelo PaaS



Fonte das Figuras 33 a 35: a autora

Figura 38 - Transição de substituição “Cálculo e Normalização” para o exemplo IaaS - estado inicial



Fonte: a autora

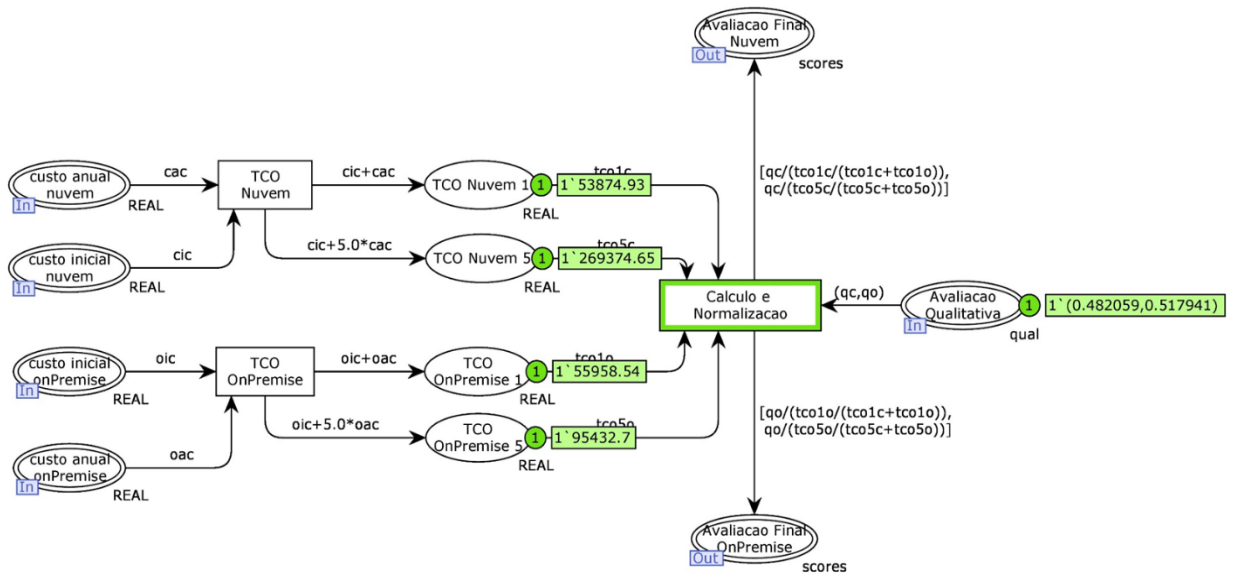
Nesse momento da simulação, as informações de custo já foram processadas e encontram-se nos lugares ‘Custo anual nuvem’, ‘Custo inicial nuvem’, ‘Custo inicial *on-premises*’, ‘Custo anual *on-premises*’, conforme ilustrado na Figura 38. Assim, as seguintes transições estão habilitadas:

- “TCO Nuvem”: calcula o TCO para o primeiro ano e para cinco anos de uso, para a alternativa nuvem, colocando os valores calculados nos lugares “TCO Nuvem 1” e “TCO Nuvem 5”. As operações de cálculo estão descritas nos arcos de saída.
- “TCO *on-premises*”: calcula o TCO para o primeiro ano e para cinco anos de uso, para a alternativa *on-premises*, colocando os valores calculados nos lugares “TCO *on-premises* 1” e “TCO *on-premises* 5”. As operações de cálculo estão descritas nos arcos de saída.

Após execução dessas transições, os valores de TCO estão calculados e a transição “Cálculo e Normalização” estará habilitada para calcular a razão custo/benefício, conforme ilustrado na Figura 39. É importante ressaltar a necessidade de normalizar os custos (de modo a obter um valor entre 0 e 1) para que a razão custo/benefício seja compatível, já que a avaliação qualitativa é também um valor entre 0 e 1.

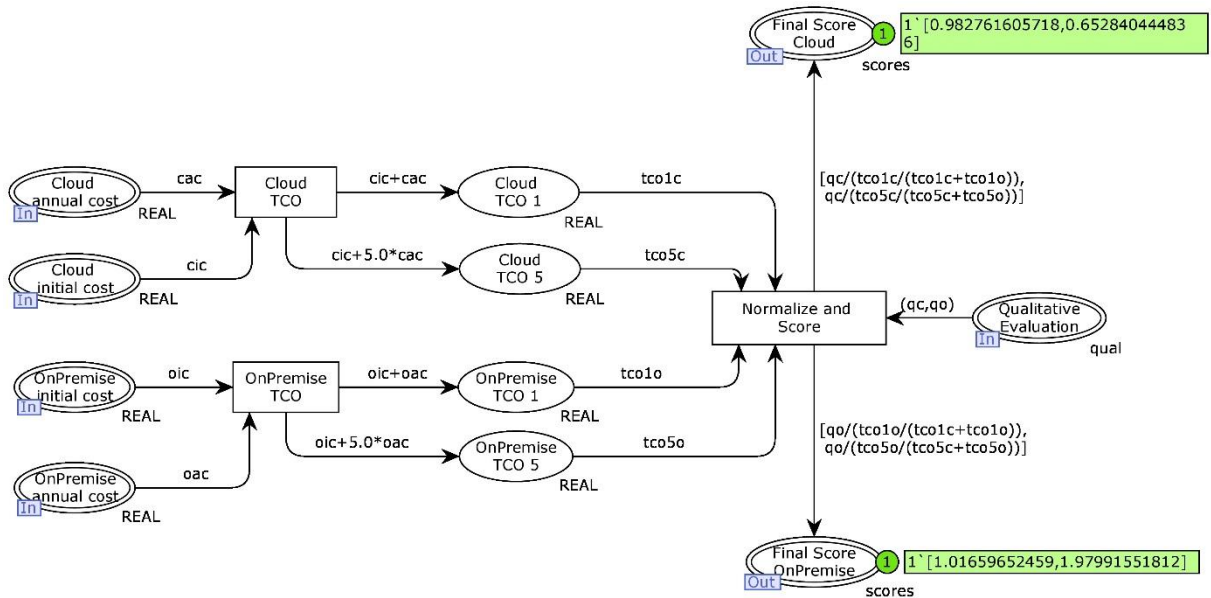
Nesse ponto, a transição “Cálculo e Normalização” realiza os cálculos de TCO e coloca o resultado nos lugares ‘Avaliação Final Nuvem’ e ‘Avaliação Final *on-premises*’, conforme ilustrado na Figura 40. Nesse exemplo, a avaliação foi mais favorável para a alternativa *on-premises*.

Figura 39 - Transição de substituição “Cálculo e Normalização” para o exemplo IaaS - estado intermediário



Fonte: a autora

Figura 40 - Transição de substituição “Cálculo e Normalização” para o exemplo IaaS - final da simulação



Fonte: a autora

4.4.4 Comparação entre os exemplos apresentados

Em cada exemplo apresentado, foi obtido um resultado diferente:

- SaaS: fator qualitativo levou à escolha do local
- PaaS: fator custo levou à escolha da nuvem
- IaaS: fator custo levou à escolha do local

Nota-se que os dados de entrada influenciam muito o resultado final, o que é bastante razoável, pois a decisão deve ser norteadada pelo cenário existente. No entanto, o *framework* construído permite uma análise visual e detalhada dos fatores que influenciam a decisão, atingindo assim o objetivo inicial do trabalho, que é de apoiar o processo decisório.

Capítulo 5

REDUÇÃO DE CUSTOS USANDO INSTÂNCIAS *SPOT*

O modelo de decisão proposto utiliza como entrada uma lista de elementos de custo e uma lista de critérios qualitativos e pesos, produzindo como saída avaliações numéricas para cada alternativa, na nuvem e local (*on-premises*). Os exemplos apresentados mostraram que, em algumas situações, os custos na nuvem precisam ser bem inferiores para convencer os tomadores de decisão, especialmente aqueles que indicam preferência maior pela opção local nos aspectos qualitativos.

A redução de custos na nuvem é, em si só, uma área de pesquisa abrangente. Para enriquecer este trabalho, foi realizado um estudo adicional, utilizando recursos avançados de redes de Petri, sobre o uso de instâncias *spot* para otimização de custos. Esse estudo, em sua versão preliminar, foi apresentado no congresso BDIM (Ribas, M. et al., 2015). Posteriormente, após coleta e análise dos resultados, foi publicado em periódico internacional (Ribas, M. et al, 2015).

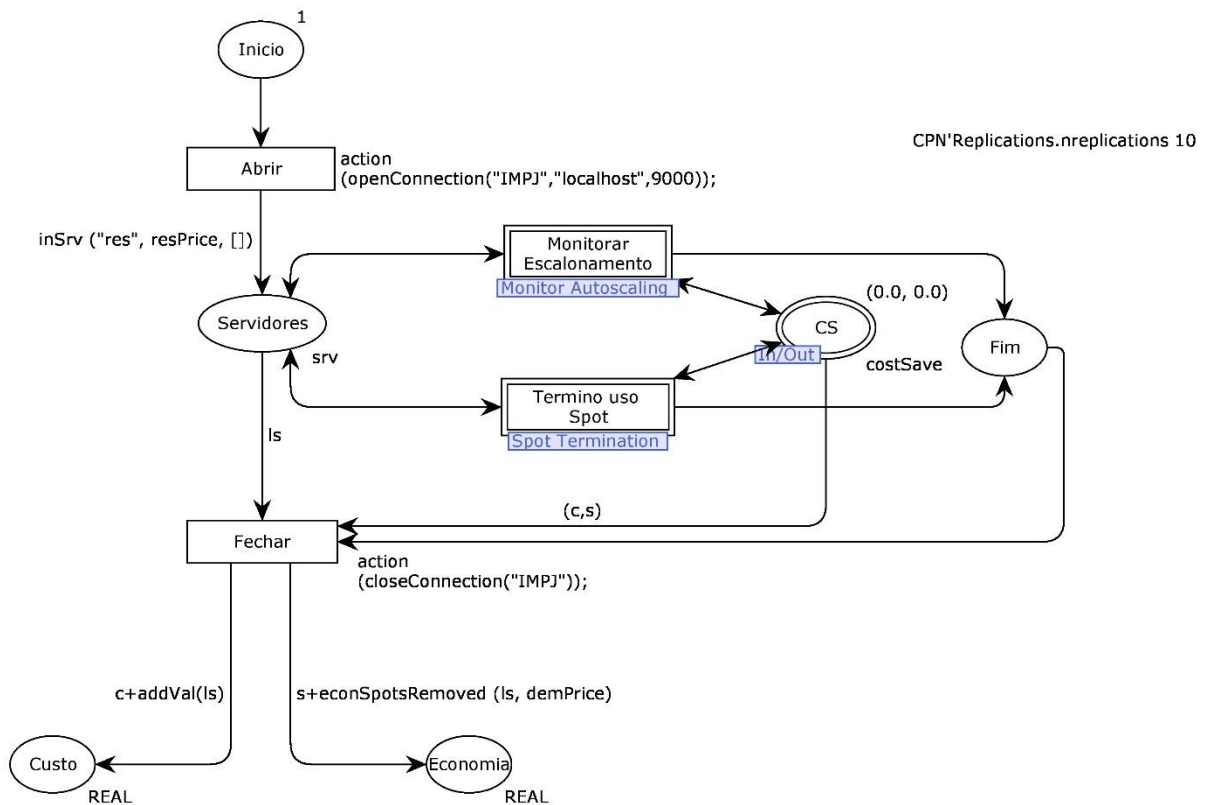
Um elemento de custo importante, quando se trata de IaaS, é o custo de utilização do servidor, ou seja, a conta mensal apresentada pelos provedores para utilização das máquinas virtuais na nuvem. Os provedores em geral utilizam uma cobrança baseada em uso, ou seja, o cliente paga por hora (ou minuto, ou mês) de uso da máquina ligada. Em geral, não há cobrança quando a máquina é desligada.

Cada provedor de IaaS tem um modelo próprio para cobrança pelo uso de máquinas virtuais. Neste capítulo, o trabalho investiga as opções de compra atuais da AWS – por demanda (*on-demand*), reservada e *spot*.

Para investigar como o uso de instâncias *spot* pode ajudar na redução de custos, este trabalho propõe a criação de uma rede de Petri colorida organizada em módulos hierárquicos que vai simular: 1) o custo mensal de uso de máquinas virtuais e 2) a economia utilizando instâncias *spot*. No modelo, haverá pelo menos uma instância reservada (poderiam ser mais, se necessário), que permanecerá constantemente ligada, para garantir disponibilidade do serviço o tempo todo. Nesse caso, a opção reservada é sem dúvida a de menor custo. Outras instâncias serão ligadas e desligadas de acordo com a necessidade, monitorando a demanda pelo serviço, simulando o processo de escalonamento automático (*auto scaling*). O objetivo deste experimento é simular o mecanismo de elasticidade e investigar a economia em potencial que o recurso de instâncias *spot* pode oferecer, se comparado com as instâncias *on-demand*. A

Figura 41 ilustra a rede principal do modelo, que será organizada hierarquicamente em outras quatro sub-redes: *Monitor Auto Scaling*, *Scale Up*, *Scale Down* e *Spot Termination*. Todas essas redes serão discutidas em detalhes a seguir.

Figura 41 - Rede principal do modelo - Simulação de uso de instâncias



Fonte: a autora

5.1 Rede Principal: Simulação do uso de instâncias

Essa rede representa o mecanismo proposto para uso de instâncias *spot* para redução de custo de elasticidade. O modelo possui quatro parâmetros de entrada, e dois de saída, conforme ilustrado na Tabela 17.

A Figura 42 apresenta um exemplo de customização da função *demand()*. Nesse exemplo, a função retorna o valor “1” (ou seja, é preciso de um servidor) durante o período noturno (23h às 6h) e retornará “1 ou mais” servidores no período diurno (6h às 22h). Para calcular quantos servidores adicionais serão necessários, utiliza-se uma distribuição Normal com média $\mu = 3$ e desvio padrão $\delta = 0.5$, como um padrão de uso hipotético.

Tabela 17 - Parâmetros de Entrada e Saída do Modelo de Redução de custos

Parâmetros de entrada	Descrição
Preço na opção <i>on-demand</i>	Preço por hora de uso de instâncias na modalidade de compra <i>on-demand</i> , representado pela constante <i>demPrice</i> no modelo.
Preço na opção Reserva	Preço por hora de uso de instâncias na modalidade Reserva, representado pela constante <i>resPrice</i> no modelo.
Preço na opção <i>spot</i>	Preço por hora de uso de instâncias na modalidade <i>spot</i> . Esses valores serão obtidos dinamicamente utilizando a interface de programação (API) da AWS, durante o período de simulação. Para tanto, foi utilizada uma interface de programação especial Java para a ferramenta CPN Tools desenvolvida no trabalho de (Gomes Furtado, Marques Soares, J., & Cordeiro Barroso, 2015).
Função demand()	Essa função representa a demanda por servidores durante o processo de escalada automática (<i>auto scaling</i>). A função devolve o número de servidores necessários em determinado ponto no tempo. Ela deve ser customizada de acordo com o cenário de uso do modelo.
Parâmetros de saída	Descrição
Custo	Custo mensal total de uso das instâncias EC2, incluindo a cobrança em todas as opções (Reservada, <i>spot</i> e <i>on-demand</i>)
Economia	Economia total obtida com o uso das instâncias <i>spot</i> , comparada com o valor que teria sido cobrado caso se utilizasse a modalidade <i>on-demand</i> .

Fonte: a autora

Figura 42 – Exemplo de função demand()

```

fun demand() =
  let
    val h = intTime() mod 24
    val extra = normal (3.0, 0.5)
  in
    if h < 7 or else h > 22 then
      1
    else
      1 + floor(extra)
  end;

```

Fonte: a autora

Assim, nossa necessidade adicional de servidores será de $\mu \pm 2\delta$ em 95% do tempo, ou seja, durante o dia, serão necessários entre dois e quatro servidores adicionais. Essa função deve ser adaptada para refletir a necessidade de servidores adicionais em cada cenário de escalada automática. Por exemplo, pode-se adaptar a função para monitorar o tamanho da fila de requisições, como no trabalho de (Han, R. et al., 2014). AWS oferece um serviço especial,

denominado *cloudwatch*, para monitorar métricas como uso de CPU, memória etc., que podem disparar a necessidade de servidores adicionais. A função `demand()` permite a fácil adaptação do modelo para qualquer necessidade específica.

A rede principal do modelo (Figura 41) possui os elementos a seguir.

5.1.1 Cores

A cor **server** representa uma instância em funcionamento. Constitui-se de uma tupla com três informações (tipo, preço, início), em que o **tipo** é uma sigla que identifica a opção de compra (*res*, *dem* ou *spot*) representando as opções Reservada, *on-demand* ou *spot*; o **preço** é o valor da hora de uso da instância; e **início** é um número inteiro entre 1 e 720 que representa a marcação do relógio global da rede quando a instância foi iniciada.

A cor **srv** é uma lista de **servers**, representando todos os servidores que estão em funcionamento no momento atual.

A cor **costSave** é uma tupla com dois números reais (r_1 , r_2) onde **r1** representa o custo total mensal e **r2** representa a economia total calculada.

5.1.2 Lugares

O lugar **Início** contém uma ficha temporizada para iniciar a simulação no tempo zero. O lugar **Fim** recebe uma ficha quando a simulação alcança o tempo predefinido para o seu final (720), que representa o número de horas em um mês ($720 \text{ horas} = 24 \text{ horas} * 30 \text{ dias}$), pois o modelo foi construído para analisar o custo de utilização em um mês. A marcação do lugar **Servidores**, de cor **srv**, representa os servidores em uso em um determinado momento da simulação. Ao final da simulação, a marcação do lugar **Custo**, representa o custo total de uso de servidores, e a marcação do lugar **Economia** representa a economia total obtida através do uso das instâncias *spot*. O lugar **CS** é utilizado como espaço temporário para acumular o custo de cada servidor, quando o mesmo é desligado. Ao desligar o servidor, pode-se calcular o número total de horas em que ele esteve ligado e multiplicar pelo preço por hora, obtendo o custo total de uso.

5.1.3 Transições, arcos e inscrições

A transição **Abrir** inicia a conexão com a interface de programação Java e cria a lista inicial de servidores em uso, incluindo uma instância adquirida na opção Reserva na lista. A transição **Fechar** fecha a conexão com a interface de programação Java e finaliza o processo de simulação, modelando a ação de desligar todos os servidores.

A transição **Monitorar Escalonamento** é uma transição de substituição que representa a sub-rede que modela o processo de escalonamento automático (ou seja, ligar e desligar servidores automaticamente de acordo com a demanda) e será discutida detalhadamente na seção 5.2.

A transição **Término uso Spot** é uma transição de substituição que representa a sub-rede que modela o processo de desligar uma instância *spot* devido às mudanças de preços no mercado *spot* e será discutida detalhadamente na seção 5.5.

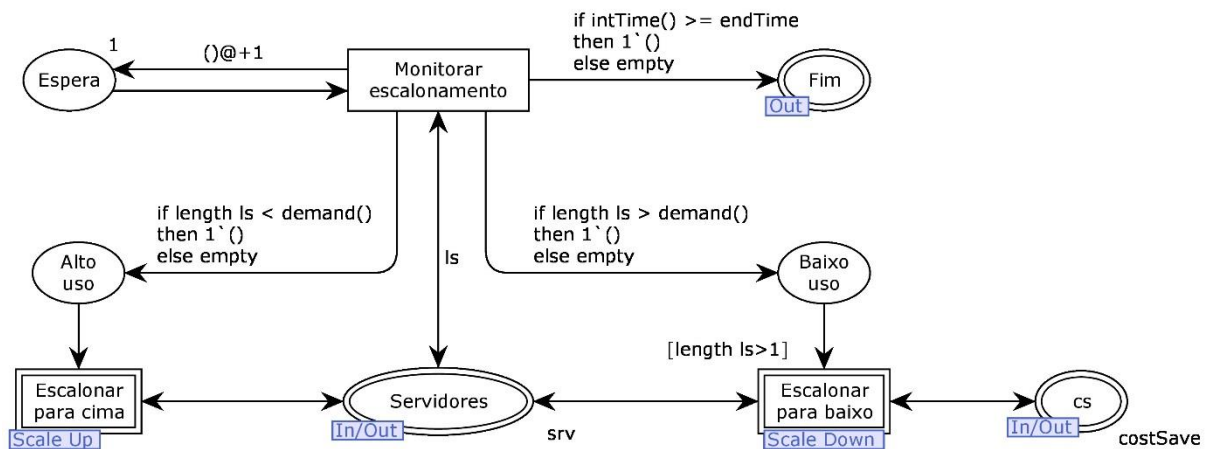
O modelo utiliza as funções auxiliares:

- **inSrv** (tipo, preço, lista): insere um servidor de um dado tipo (opção de compra) e preço em uma lista.
- **addVal** (lista): calcula o custo do uso de todos os servidores na lista, multiplicando as horas de uso pelo valor da hora de cada um, e somando todos os valores para obter o resultado final.
- **econSpotsRemoved** (lista): calcula a economia total ao usar as instâncias *spot*, multiplicando horas de uso de cada instância *spot* pela diferença entre o preço da hora na opção ‘*on-demand*’ e o preço *spot* utilizado, e somando todos os valores para obter o resultado final.

5.2 Sub-rede para monitorar escalonamento: *Monitor Auto Scaling*

Essa sub-rede modela o processo de escalonamento automático. A Figura 43 ilustra o modelo CPN que implementa a transição **Monitorar Escalonamento** na rede principal. Essa transição será executada a cada unidade de tempo e verifica se há necessidade de ligar ou desligar servidores. Para o objetivo de comparação de custos, a unidade de tempo será de uma hora. O modelo não teria alteração significativa se a unidade de tempo utilizada fosse menor, como minutos ou segundos. Em linhas gerais, a transição **Monitorar Escalonamento** compara a necessidade de servidores em um determinado tempo (obtida pela função `demand()`, parâmetro de entrada do modelo) com o número de servidores ativos (dado pelo tamanho da lista de servidores). Quando a demanda é maior, uma ficha é colocada no lugar **Alto uso**, o que vai disparar a transição de substituição **Escalonar para cima**, que será detalhado no item 5.3. Quando a demanda é menor, uma ficha é colocada no lugar **Baixo uso**, o que vai disparar a transição de substituição **Escalonar para baixo**, detalhado no item 5.4. Se a demanda não é maior nem menor, a transição vai apenas aguardar a próxima unidade de tempo, colocando uma ficha no lugar **Espera**. Essa ficha só estará disponível novamente para a transição **Monitorar Escalonamento** na próxima unidade de tempo. A transição **Monitorar Escalonamento** também verifica o final do período de simulação e, nesse caso, coloca uma ficha no lugar **Fim** (da rede principal, descrito em 5.1.2).

Figura 43 - Sub-rede para monitorar escalonamento: Monitor Auto Scaling



Fonte: a autora

5.2.1 Cores

As cores utilizadas são as mesmas descritas na rede principal (seção 5.1.1).

5.2.2 Lugares

Os lugares **Fim**, **CS** e **Servidores** são os mesmos descritos na rede principal (seção 5.1.2). Os lugares **Alto uso** e **Baixo uso** indicam, respectivamente, a necessidade de se aumentar ou reduzir a quantidade de servidores em uso, e possuem uma ficha quando essa necessidade é detectada. O lugar **Espera** é utilizado para implementar a temporização da checagem, para que ocorra a cada unidade de tempo, possuindo uma ficha quando for necessário aguardar o avanço do relógio.

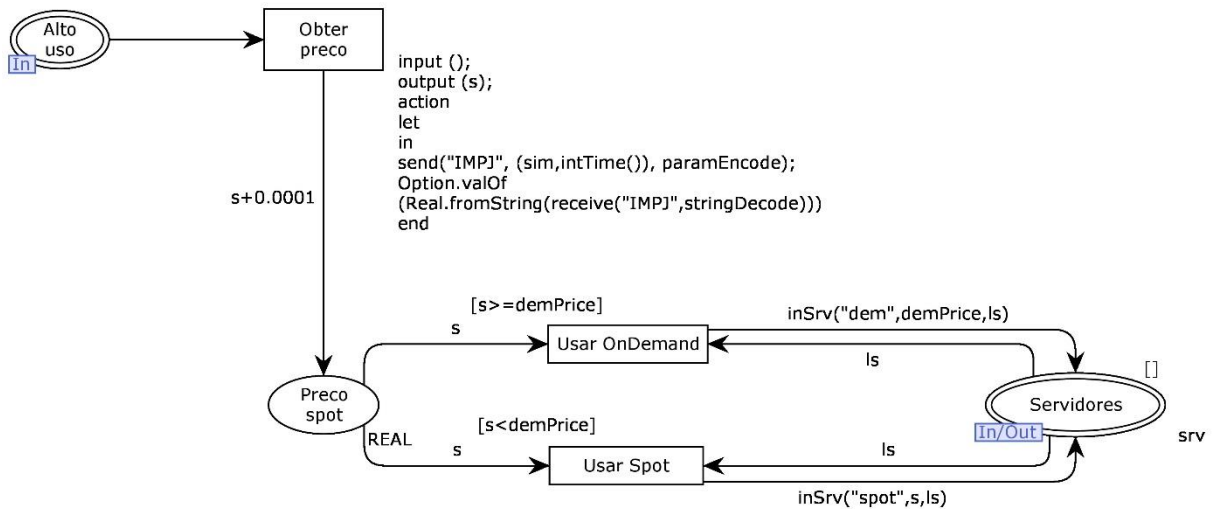
5.2.3 Transições, arcos e inscrições

A transição **Monitorar Escalonamento** realiza a verificação periódica da demanda e dispara os processos para reduzir ou aumentar o número de servidores. Ela também verifica o tempo para conferir se alcançou o final do período de simulação. As transições **Escalonar para cima** e **Escalonar para baixo** serão discutidas, respectivamente, nas seções 5.3 e 5.4.

5.3 Sub-rede para Escalonar para cima: *Scale Up*

Essa sub-rede modela o processo de escalonamento para cima, ou seja, as ações para ligar servidores adicionais, com o menor custo possível, para atender o aumento de demanda. A Figura 44 ilustra o modelo CPN que implementa a transição **Escalonar para cima** na sub rede **Monitor Auto Scaling** (5.2).

Figura 44 – Sub-rede Scale Up



Fonte: a autora

5.3.1 Cores

As cores utilizadas são as mesmas descritas na rede principal (seção 5.1.1).

5.3.2 Lugares

O lugar **Servidores** é o mesmo descrito na rede principal (seção 5.1.2), que representa a lista de servidores atual. O lugar **Alto uso** é o mesmo descrito na sub-rede **Monitor Auto Scaling** (seção 5.2), que tem uma ficha quando há necessidade de aumentar o número de servidores. A marcação do lugar **Preço Spot** representa o lance que será colocado para obter uma instância *spot*, cujo valor será o preço de mercado atual (um número real) retornado pela interface Java, acrescido de \$0.0001. Dessa maneira, garante-se que nosso lance é suficiente para obter uma instância *spot* ao menor custo possível.

5.3.3 Transições, arcos e inscrições

A transição **Obter preço** utiliza a interface Java para obter o preço *spot* atual para o modelo usando parâmetros (simulação, tempo), onde **simulação** é um número de simulação que representa um conjunto de parâmetros necessários para obter preços *spot*, segundo a API fornecida pela Amazon. Para poder executar a simulação repetidas vezes, os preços históricos das instâncias foram armazenados em arquivos e acessados através do número da simulação, conforme ilustrado na Tabela 18. O parâmetro **tempo** representa o dia e a hora do mês, conforme ilustrado na Tabela 19.

Tabela 18 - Parâmetro Simulação para obtenção de preço spot

Simulação	Região AWS	Sistema Operacional	Tamanho da instância
1	South America	Windows	m3.medium
2	South America	Windows	m3.2xlarge
3	South America	Linux/UNIX	m3.medium
4	South America	Linux/UNIX	m3.2xlarge
5	US-East	Windows	m3.medium
6	US-East	Windows	m3.2xlarge
7	US-East	Linux/UNIX	m3.medium
8	US-East	Linux/UNIX	m3.2xlarge
...			

Fonte: a autora.

Tabela 19 – Parâmetro Tempo e sua correspondência com hora e dia da simulação

Tempo	Dia do mês	Hora do dia
1	1	0
2	1	1
...
24	1	23
25	2	0
...
720	30	23

Fonte: a autora

A transição **Usar OnDemand** é disparada quando o valor da marcação no lugar **Preço Spot** é mais alto do que o preço da instância adquirido na opção ‘*On-demand*’ (esse preço é um dos parâmetros de entrada do modelo). Nesse caso, não é interessante adquirir a instância na modalidade *spot*, pois é mais caro. Essa transição utiliza a função *inSrv* (descrita em 5.1.3) para atualizar a lista, indicando a utilização de um servidor adicional do tipo *on-demand*.

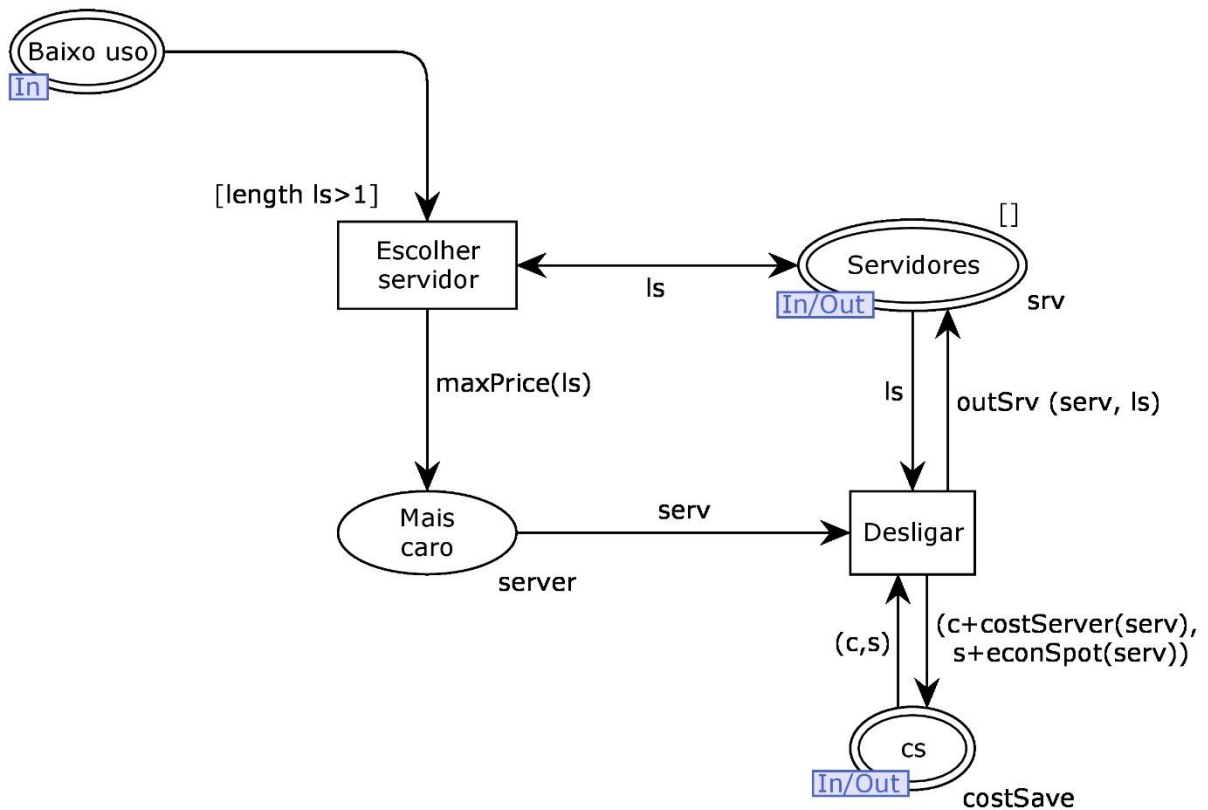
A transição **Usar Spot** é disparada quando o valor da marcação no lugar **Preço Spot** é mais baixo do que o preço da instância adquirida na opção ‘*On-demand*’. Nesse caso, é interessante adquirir a instância na modalidade *spot*, pois é mais barato. Essa transição utiliza a função *inSrv* (descrita em 5.1.3) para atualizar a lista, indicando a utilização de um servidor do tipo *spot*.

5.4 Sub-rede para escalonamento para baixo: *Scale Down*

Essa sub-rede modela o processo de escalonamento para baixo, ou seja, as ações para desligar servidores, de modo a obter o menor custo de uso possível quando a demanda diminui. Em

linhas gerais, é sempre desligado o servidor de maior custo. A Figura 45 ilustra o modelo CPN que implementa a transição **Escalonar para baixo** na sub-rede *Monitor Auto Scaling* (5.2).

Figura 45 – Sub-rede Scale Down



Fonte: a autora

A política implementada por esta sub-rede é a de selecionar o servidor que tem o preço de uso por hora mais alto, e desligá-lo. A sub-rede também implementa a política de deixar sempre ligado o servidor adquirido na modalidade Reserva (pela qual o cliente se compromete a pagar por período de reserva e não por horas de uso). Dessa forma, mesmo com baixa utilização, haverá um servidor disponível. Essa construção modela os cenários do mundo real em aplicações comerciais, onde a aplicação está disponível 24 horas por dia, sempre. Também pode modelar situações onde o servidor ‘reservado’ é, na verdade, um servidor hospedado internamente, em uma nuvem híbrida. Em termos de custo, em alguns cenários, um servidor ‘reservado’ é equivalente a um hospedado internamente (E-fiscal, 2013).

5.4.1 Cores

As cores utilizadas são as mesmas descritas na rede principal (seção 5.1.1).

5.4.2 Lugares

Os lugares **Servidores** e **CS** são os mesmos descritos na rede principal (seção 5.1.2), representando a lista de servidores atual, o custo total e a economia obtida. O lugar **Baixo uso** é o mesmo descrito na sub-rede **Monitor Auto Scaling** (seção 5.2), que tem uma ficha quando há necessidade de reduzir o número de servidores. A marcação do lugar **Mais caro** representa o servidor que tem o maior custo, candidato natural para ser desligado, dessa forma reduzindo ao máximo o custo total do uso de servidores.

5.4.3 Transições, arcos e inscrições

A transição **Escolher servidor** seleciona da lista o servidor com maior custo, para desligá-lo. Para isso, utiliza uma função denominada *maxPrice (list)*, que seleciona o servidor com o valor de uso por hora mais alto. Essa função desconsidera as instâncias reservadas, pois desligá-las não faz diferença no custo final, uma vez que serão sempre cobradas pelo período completo da reserva, independentemente de seu uso. A transição **Desligar** exclui da lista o servidor selecionado. Ao mesmo tempo, calcula o custo total do uso do servidor, por meio da apuração do número de horas que ele esteve ligado multiplicado pelo número de horas (função *CostServer*). Além disso, calcula a economia obtida, comparando com o custo de uma instância *on-demand* (função *EconSpot*). As funções utilizam as seguintes fórmulas:

- *CostServer*: (horas de uso) * (preço da hora)
- *EconSpot*: (horas de uso) * ((preço da hora na opção *on-demand*) – (preço da hora da instância *spot* sendo desligada)).

É importante notar que a economia obtida somente será calculada para as instâncias *spot*, pois serão sempre zero quando a instância a ser desligada foi adquirida no modelo *on-demand* (nesse caso, (preço da hora na opção *on-demand*) – (preço da hora da instância sendo desligada) = 0).

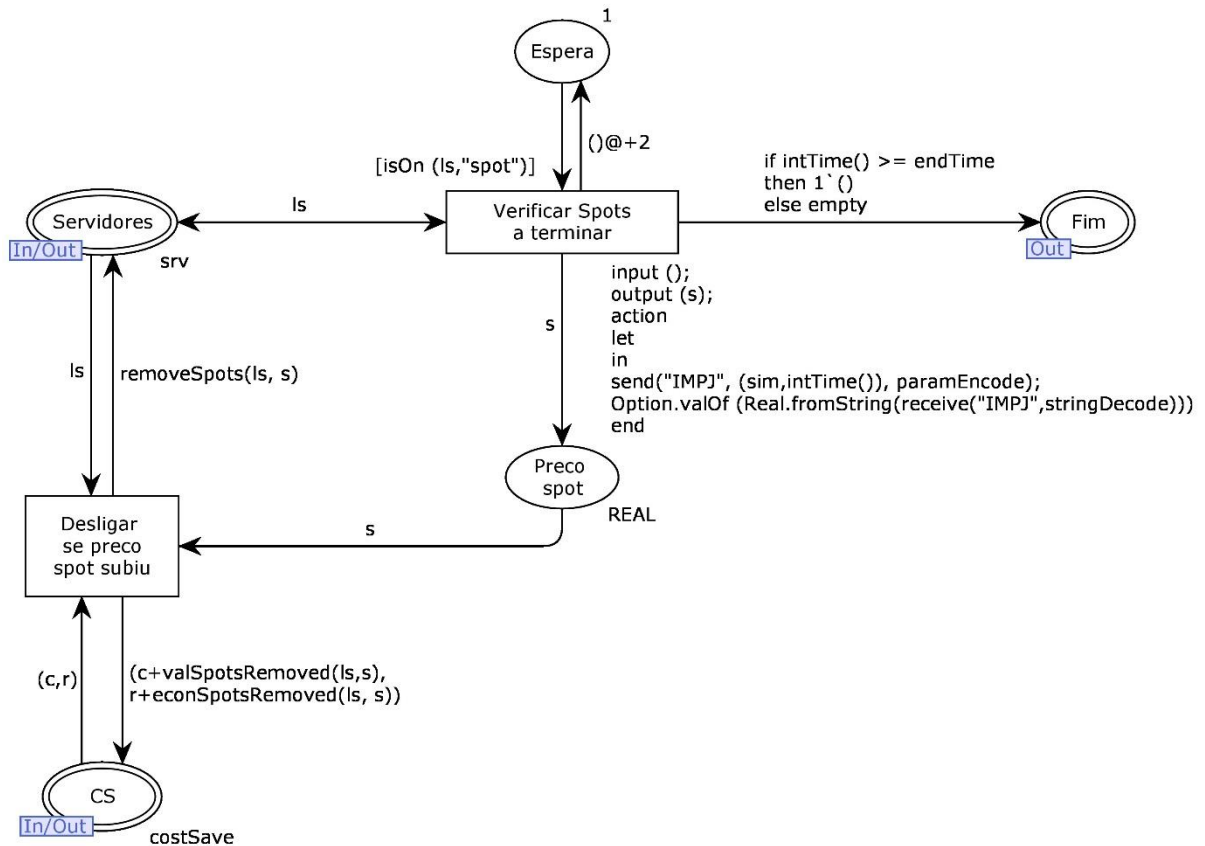
5.5 Sub-rede para simular término de instâncias *spot*: *Spot Termination*

Essa sub-rede simula a verificação regular dos preços no mercado *spot*. As instâncias *spot* são desligadas automaticamente pela AWS quando o valor da hora no mercado *spot* sobe e fica maior do que o valor que o cliente está atualmente pagando pelo uso da instância. A Figura 46 apresenta a sub-rede *Spot Termination*.

5.5.1 Cores

As cores utilizadas são as mesmas descritas na rede principal (seção 5.1.1).

Figura 46 - Sub-rede Spot Termination



Fonte: a autora

5.5.2 Lugares

Os lugares **Servidores**, **CS** e **Fim** são os mesmos descritos na rede principal (seção 5.1.2), representando a lista de servidores atual, o custo total e a economia obtida, e o indicador de final de simulação. O lugar **Espera** é utilizado para implementar a temporização da checagem, para que ocorra a cada duas unidades de tempo, possuindo uma ficha quando for necessário aguardar o avanço do relógio. A marcação do lugar **Preço Spot** representa o valor atual da hora no mercado *spot*, obtido através da interface de programação Java.

5.5.3 Transições, arcos e inscrições

A transição **Verificar Spots a terminar** é disparada a cada duas horas, para monitorar os preços do mercado *spot* e simular o desligamento das instâncias *spot*. Para tanto, utiliza a mesma interface de programação Java para obter o valor atual do preço *spot* que foi previamente utilizado na transição **Obter preço** na sub-rede *Scale Up*. Essa transição só é disparada quando existe pelo menos uma instância *spot* ligada. Após o disparo, a sub-rede aguarda duas horas, colocando uma ficha no lugar **Espera** que só estará disponível novamente depois de duas horas.

Essa transição também verifica se atingiu o final da simulação e, nesse caso, coloca uma ficha no lugar **Fim**.

A transição **Desligar se preço *spot* subiu** percorre a lista de servidores ativos buscando instâncias *spot* cujo valor da hora seja menor do que o preço *spot* atual. Essas instâncias serão desligadas. Ela utiliza funções *valSpotsRemoved()* and *econSpotsRemoved()* para calcular o custo de utilização das instâncias que foram desligadas, bem como a economia obtida, de modo idêntico ao cálculo descrito na sub-rede **Scale Down**. A única diferença é que a transição **Desligar se preço *spot* subiu** pode remover vários servidores da lista de servidores ativos de uma única vez. Nesse modelo, o desligamento dessas instâncias pode causar vários disparos da transição **Scale Up** caso o nível de utilização do sistema esteja alto e necessite repor as instâncias *spot* que foram desligadas.

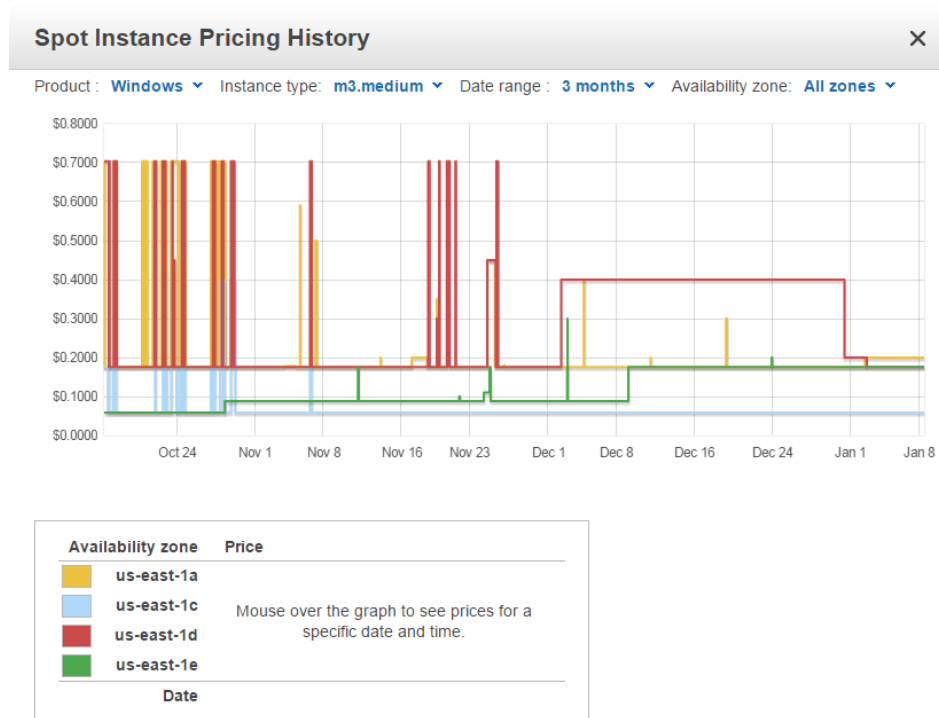
5.6 Resultados da simulação

O experimento realizado para similar a economia segue o formato 2^k fracionado para cada região geográfica onde a AWS possui *datacenters* e disponibiliza as instâncias *spot*. AWS atualmente oferece seus serviços de nuvem em nove regiões diferentes:

- US-East (N. Virginia),
- US-West (Oregon),
- US-West (N. California),
- EU (Ireland),
- EU (Frankfurt),
- Asia Pacific (Singapore),
- Asia Pacific (Tokyo),
- Asia Pacific (Sydney),
- South America (São Paulo).

Cada região tem pelo menos duas zonas de disponibilidade (AZs), que são *datacenters* em localizações distintas, conectados através de redes de baixa latência. Cada região tem sua tabela de preço para instâncias na modalidade *on-demand* e reservada. O preço das instâncias *spot* varia não apenas com a região, mas também com a zona de disponibilidade. A Figura 48 ilustra os preços *spot* para as diversas zonas de disponibilidade na região US-East. Pode-se visualizar que o preço é bastante aleatório, no entanto há longos períodos de tempo em que o preço permanece estável (e muito baixo). Note-se que, para a zona de disponibilidade **us-east-1c**, os preços ficaram constantemente baixos (\$0.0591) por dois meses – novembro e dezembro de 2014. O gráfico dinâmico da Figura 47 está disponível no console da AWS, mas é necessário criar uma conta na AWS para acessá-lo, na opção *EC2 – spot Requests*.

Figura 47 - Histórico de preços de instâncias spot



Fonte: AWS (AWS Console, 2014)

A transição **Obter preço** utilizada no modelo retorna o menor preço entre todas as zonas de disponibilidade de uma dada região, para garantir a maior economia possível. Outros fatores que influenciam os preços das instâncias são:

- **Sistema operacional (SO):** a AWS tem preços diferentes por horas de uso de instâncias EC2 dependendo do sistema operacional desejado: Linux/Unix, SUSE Linux e Windows.
- **Tipo (tamanho) da instância:** os preços das instâncias EC2 variam de acordo com seu tamanho. Para o experimento, foram considerados os tamanhos m3.medium (1 vCPU, com poder computacional de 3 ECU, 3.75 GB de memória e um HD do tipo SSD com capacidade de armazenamento de 4GB) e m3.2xlarge (8 vCPU, com poder computacional de 26 ECU, 30 GB de memória e dois HDs do tipo SSD com capacidade de armazenamento de 80 GB)

A Tabela 20 ilustra os fatores e seus níveis nos experimentos realizados para as simulações, e a Tabela 21 ilustra os preços correspondentes das instâncias. Cada uma das ($9 \times 2^2 = 36$) simulações foi executada 10 vezes para obter informação estatística.

Tabela 20 - Fatores e níveis do experimento

Fator	Níveis	Selecionadas para experimento
Região	9	Todas
SO	3	Windows, Linux/Unix
Tamanho da instância	Mais de 20	m3.medium, m3.2xlarge

Tabela 21 – Preços por hora de uso das instâncias utilizadas nas simulações

Simulação	Região	SO	Tamanho da instância	Valor da hora em dólares		
				On-demand	Reservada	Investimento inicial
1	South America	Windows	m3.medium	0,1580	0,1410	1.235,00
2	South America	Windows	m3.2xlarge	1,2650	1,1205	9.816,00
3	South America	Linux/UNIX	m3.medium	0,0950	0,0509	446,00
4	South America	Linux/UNIX	m3.2xlarge	0,7610	0,4063	3.559,00
5	US-East (N.Virginia)	Windows	m3.medium	0,1330	0,0855	749,00
6	US-East (N.Virginia)	Windows	m3.2xlarge	1,0640	0,6809	5.965,00
7	US-East (N.Virginia)	Linux/UNIX	m3.medium	0,0700	0,0425	372,00
8	US-East (N.Virginia)	Linux/UNIX	m3.2xlarge	0,5600	0,3412	2.989,00
9	US-West (Oregon)	Windows	m3.medium	0,1330	0,0855	749,00
10	US-West (Oregon)	Windows	m3.2xlarge	1,0640	0,6809	5.965,00
11	US-West (Oregon)	Linux/UNIX	m3.medium	0,0700	0,0425	372,00
12	US-West (Oregon)	Linux/UNIX	m3.2xlarge	0,5600	0,3412	2.989,00
13	US-West (California)	Windows	m3.medium	0,1400	0,0983	861,00
14	US-West (California)	Windows	m3.2xlarge	1,1200	0,7829	6.858,00
15	US-West (California)	Linux/UNIX	m3.medium	0,0770	0,0532	466,00
16	US-West (California)	Linux/UNIX	m3.2xlarge	0,6160	0,4216	3.693,00
17	EU (Ireland)	Windows	m3.medium	0,1330	0,0983	861,00
18	EU (Ireland)	Windows	m3.2xlarge	1,0640	0,7829	6.858,00
19	EU (Ireland)	Linux/UNIX	m3.medium	0,0770	0,0522	457,00
20	EU (Ireland)	Linux/UNIX	m3.2xlarge	0,6160	0,4187	3.668,00
21	EU (Frankfurt)	Windows	m3.medium	0,1460	0,0998	874,00
22	EU (Frankfurt)	Windows	m3.2xlarge	1,1690	0,7924	6.941,00
23	EU (Frankfurt)	Linux/UNIX	m3.medium	0,0830	0,0564	494,00
24	EU (Frankfurt)	Linux/UNIX	m3.2xlarge	0,6650	0,4522	3.961,00
25	Asia Pacific (Singapore)	Windows	m3.medium	0,1610	0,0973	852,00
26	Asia Pacific (Singapore)	Windows	m3.2xlarge	1,2880	0,7711	6.755,00
27	Asia Pacific (Singapore)	Linux/UNIX	m3.medium	0,0980	0,0578	506,00
28	Asia Pacific (Singapore)	Linux/UNIX	m3.2xlarge	0,7840	0,4623	4.050,00
29	Asia Pacific (Tokyo)	Windows	m3.medium	0,1510	0,0981	859,00
30	Asia Pacific (Tokyo)	Windows	m3.2xlarge	1,2060	0,7846	6.873,00
31	Asia Pacific (Tokyo)	Linux/UNIX	m3.medium	0,1010	0,0566	496,00
32	Asia Pacific (Tokyo)	Linux/UNIX	m3.2xlarge	0,8100	0,4589	4.020,00
33	Asia Pacific (Sydney)	Windows	m3.medium	0,1610	0,0973	852,00
34	Asia Pacific (Sydney)	Windows	m3.2xlarge	1,2880	0,7711	6.755,00
35	Asia Pacific (Sydney)	Linux/UNIX	m3.medium	0,0980	0,0578	506,00
36	Asia Pacific (Sydney)	Linux/UNIX	m3.2xlarge	0,7840	0,4623	4.050,00

Fonte: a autora

Em todos os experimentos, a variação entre as execuções foi bem pequena. A Tabela 22 apresenta o relatório de desempenho obtido pela ferramenta CPN Tools para a simulação 30 – região Asia Pacific (Tokyo), sistema operacional Windows e tamanho de instância m3.2xlarge, onde se pode observar o desvio padrão e os intervalos de confiança. Para o intervalo de confiança de 99%, a economia média ficou entre US\$ 1,337.73 e US\$ 1,403.95, muito próxima da média de US\$ 1,370.84. Por esse motivo, cada simulação foi executada apenas 10 vezes.

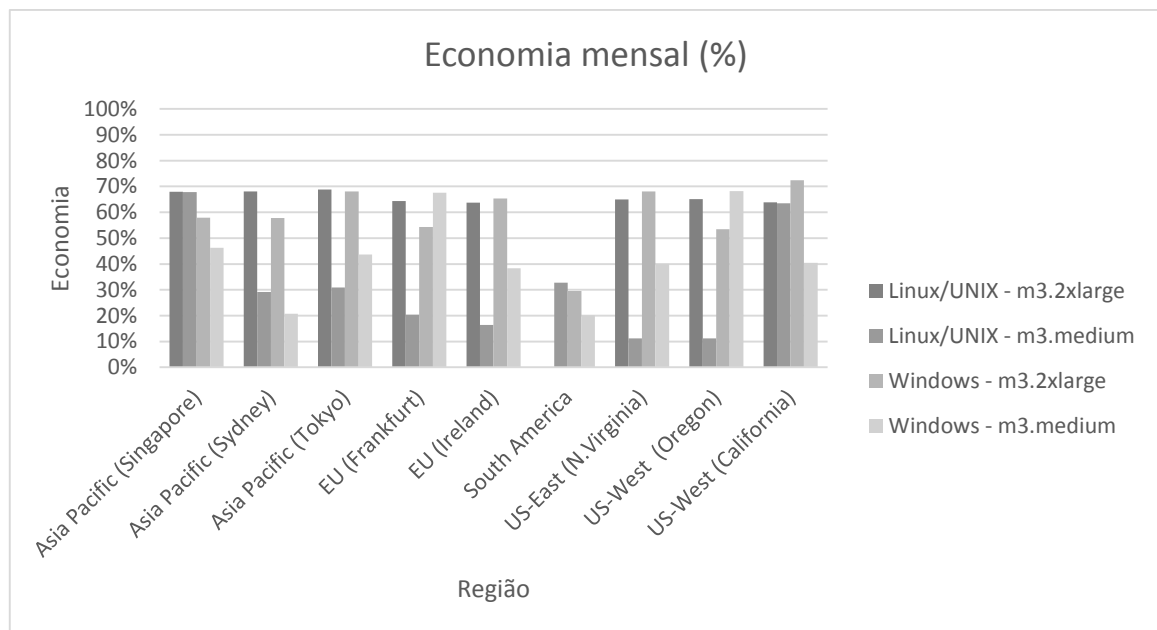
Tabela 22 - Relatório de desempenho obtido para a simulação 30

Estatísticas						
Avrg	90% Half Length	95% Half Length	99% Half Length	StD	Min	Max
Custo						
641.992	1.049	1.295	1.861	1.811	640.220	646.190
Economia						
1370.843	18.676	23.047	33.113	32.219	1339.329	1445.517

Fonte: a autora, obtido através da ferramenta CPN Tools

Para esses experimentos, foi utilizado o mês mais recente disponível na época (com o histórico de preços *spot* mais atual), de modo que as simulações foram realizadas com os dados de dezembro de 2014. As Figuras 48 e 49 apresentam um resumo dos resultados das simulações.

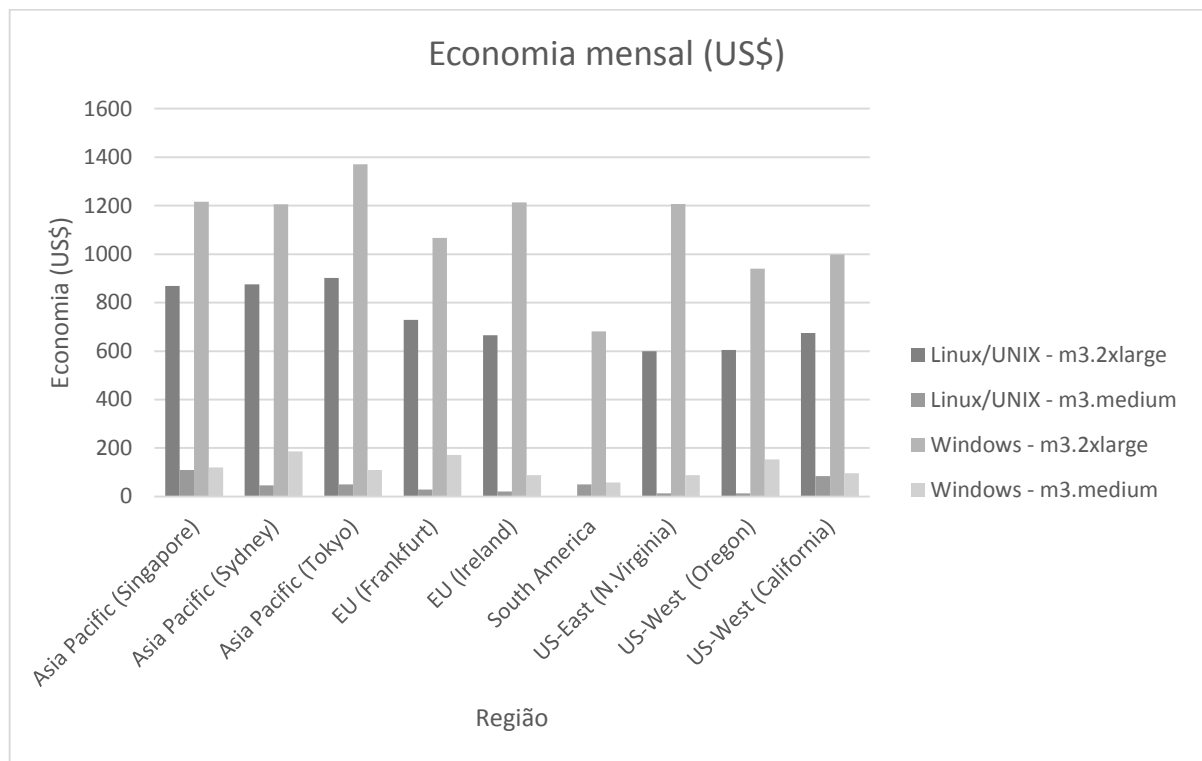
Figura 48 – Resultados das simulações: economia obtida em todas as regiões para todos os fatores (em %)



Fonte: a autora

A economia obtida varia entre 0 e 70%, dependendo da região, sistema operacional e tamanho da instância. O único caso em que não houve economia alguma foi na região *South America*, para Linux/UNIX e instância m3.2xlarge. Isso ocorreu devido ao fato de que, no período da simulação, os preços *spot* permaneceram acima dos preços *on-demand* (US\$ 1.2240 para *spot* e US\$ 0.7610 para *on-demand*). Nesse caso, o modelo não utilizou as instâncias *spot* e, conseqüentemente, não houve economia.

Figura 49 - Economia mensal em dólares



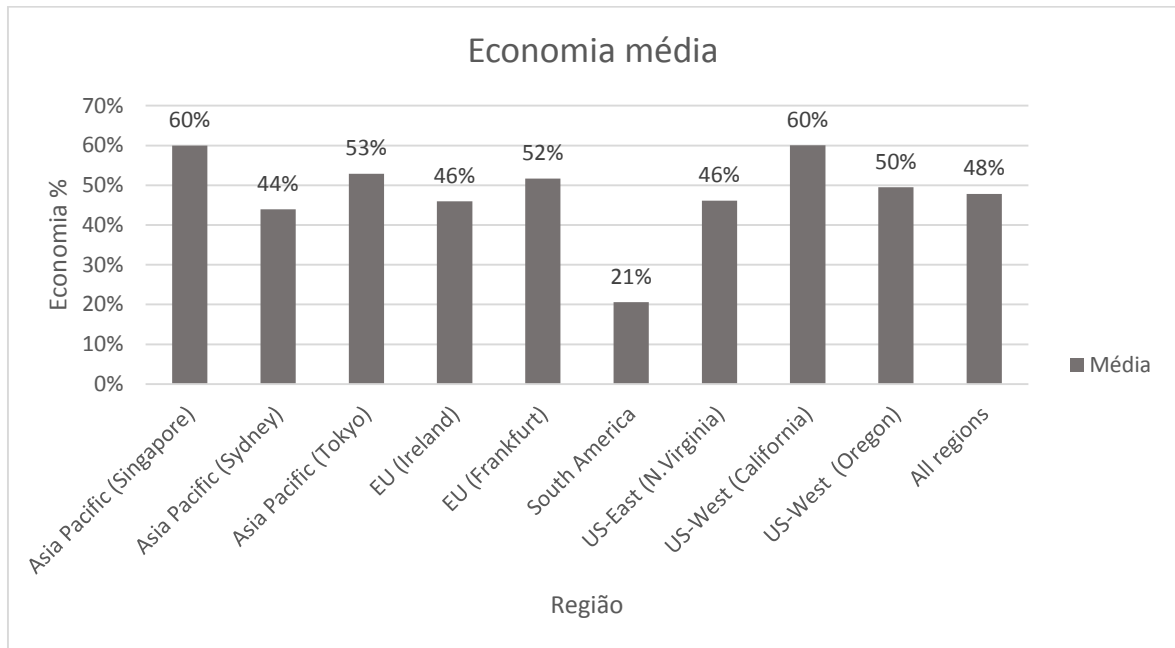
Fonte: a autora

Note-se que as instâncias maiores (m3.2xlarge) apresentaram maior economia, principalmente em valores absolutos. Em praticamente todas as regiões, a economia foi superior a US\$ 600 dólares, tanto no sistema operacional Windows como Linux. No sistema operacional Windows, a economia foi ainda maior, superando os US\$ 800 na maioria das regiões. Essa observação sugere que o fator **tamanho** influencia positivamente a média, ou seja, quanto maior a instância, maior será a economia com o uso de *spot*. Em termos absolutos isso seria esperado, pois as instâncias maiores são também as mais caras, mas a observação também se repete em termos percentuais. Observa-se uma economia acima de 60% na maioria das regiões no gráfico da Figura 49.

A região *South America* foi onde se verificou o menor percentual de economia, indicando que, nessa região e período, o preço das instâncias *spot* foi próximo ao das instâncias *on-demand*, conforme pode-se observar na Figura 49.

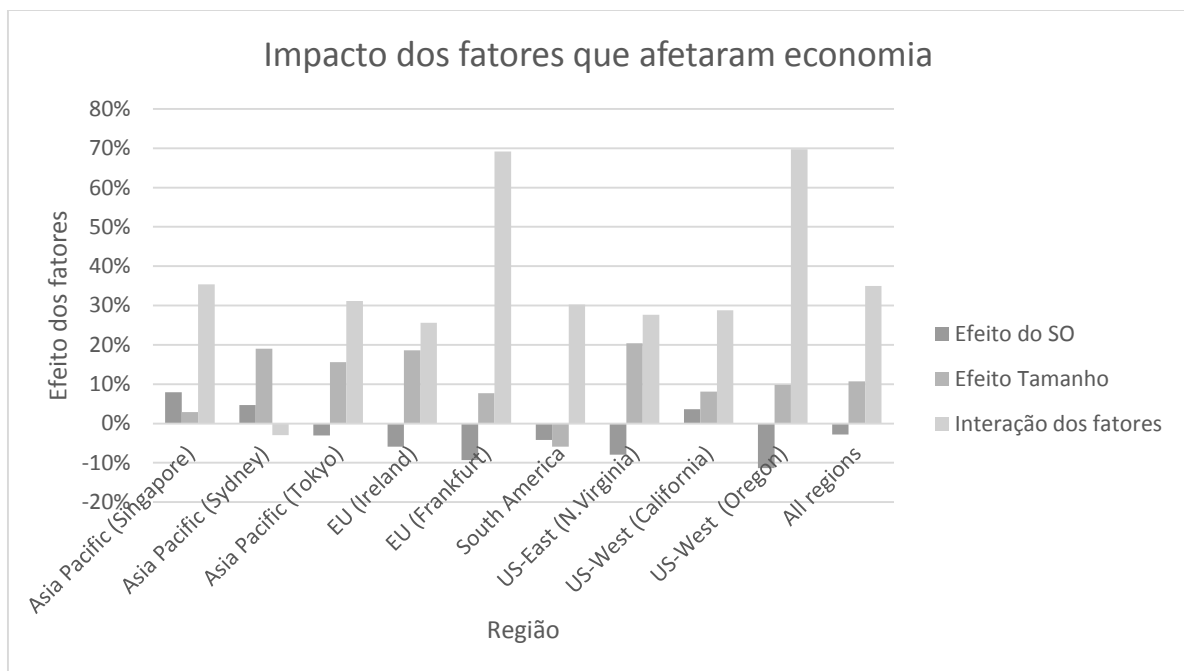
A Figura 50 apresenta a média obtida no experimento do tipo 2^k -fatorial (Jain, R., 1991). A média global foi de 48%, sendo maior do que 20% em todas as regiões.

Figura 50 - Economia média em todas as regiões



Fonte: a autora

Figura 51 - Impacto dos fatores



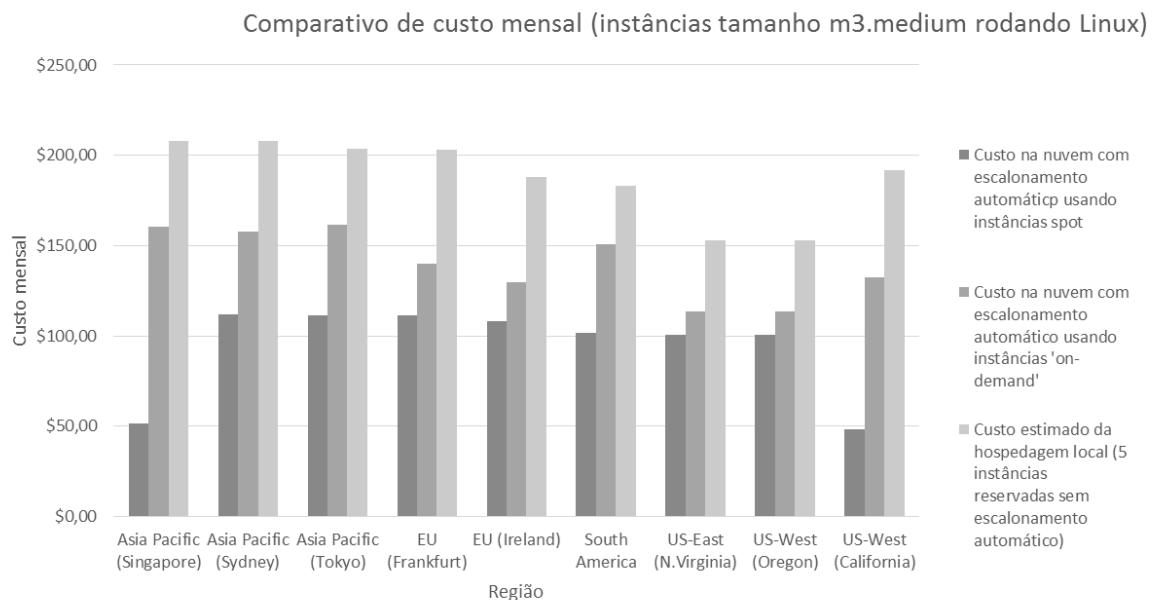
Fonte: a autora

A Figura 51 apresenta o impacto dos dois fatores utilizados no experimento. Conforme esperado, pode-se observar um impacto positivo do fator **tamanho** da instância na maioria das regiões, pois o valor médio de economia obtido era maior nas instâncias maiores.

Também se pode observar um impacto negativo do fator **sistema operacional**, na maioria das regiões, pois se verifica uma economia ligeiramente maior, tanto em termos

absolutos como percentuais, nas instâncias com sistema operacional Linux (convencionando-se que Linux é um sistema “menor”, no sentido de ser mais leve). A interação dos dois fatores (sistema operacional e tamanho da instância) produz um impacto levemente positivo na maioria das regiões. No entanto, nas regiões *EU (Frankfurt)*, *US-West (Oregon)* e *South America*, observa-se um alto impacto da interação entre os fatores. Nessas regiões, o padrão observado de economia foi um pouco diferente – nenhum fator isolado apresentou grande impacto, somente a interação entre os fatores.

Figura 52 - Comparativo de custos de hospedagem local e na nuvem



Fonte: a autora

Finalmente, é possível comparar o custo mensal de utilização de servidores implementando a política proposta para escalonamento automático usando instâncias *spot* com o uso de servidores hospedados localmente (*on-premises*). Para conduzir essa comparação, foi adotado o custo das instâncias reservadas como uma aproximação para o custo de servidores locais, pois em alguns cenários eles são bastante similares (E-fiscal, 2013). No cenário de hospedagem local, é preciso provisionar servidores adicionais para atender a demanda nos momentos de pico de utilização. Na função hipotética utilizada, $demand()$, a carga dos servidores varia seguindo uma distribuição Normal com média $\mu = 3$ e desvio padrão $\delta = 0.5$, o que significa que 95% dos valores ficam na região entre $\mu \pm 2\delta$, ou seja, durante o horário diurno, a necessidade de servidores adicionais varia entre 2 e 4. Assim, é necessário provisionar

cinco servidores, e alguns ficarão inativos parte do tempo. A Figura 52 apresenta a comparação de custos para instâncias de tamanho m3.medium utilizando sistema operacional Linux.

Pode-se observar uma grande variação entre os resultados obtidos nas diversas regiões. A variação de custo, considerando-se o (custo da hospedagem local – custo usando instâncias *spot*), varia muito de acordo com a região, de US\$ 52 (*US-East* e *US-West Oregon*) até US\$ 156 (*Asia Pacific – Singapore*). Apesar da alta variação, causada basicamente pela aleatoriedade dos preços *spot*, pode-se notar que a estratégia é sempre economicamente vantajosa. Além disso, o custo de implementação da estratégia é extremamente baixo, devido à simplicidade do modelo. A automação da estratégia não requer altos investimentos. Dessa forma, pode-se concluir que o modelo proposto é muito interessante para redução de custos, o que pode resultar numa avaliação bem mais favorável para a alternativa da nuvem utilizando o *framework* proposto no capítulo 4.

A economia utilizando instâncias *spot* era esperada, pois o preço *spot* é muitas vezes inferior em grande parte do tempo, na maioria das regiões. A simulação assume que a aplicação esteja preparada para ser desligada e reiniciada sem perda de informações ou grande impacto para o usuário final, o que restringe o uso desse modelo a aplicações concebidas para esse tipo de situação.

Capítulo 6

VALIDAÇÃO DO MODELO

Para validação do modelo apresentado, foi utilizado o método da validade de aparência (Kitchenham & Pfiieger, 2002), ou seja, verificar se o modelo aparenta ser útil, prático, pertinente e adequado para o seu propósito. Utilizando essa metodologia, o modelo foi apresentado a 13 gerentes da área de sistemas de informação, dos quais 10 responderam um questionário (detalhado no Apêndice C) indicando, em sua visão, se o modelo parece ser adequado. O questionário foi construído de modo similar ao apresentado em (Lima A. , 2010), para avaliar as hipóteses de preferência, utilidade, acurácia, efetividade e aprimoramento.

A seguir, são detalhadas as hipóteses avaliadas e os resultados obtidos.

6.1 Hipótese 1 – Preferência

Hipótese nula: não há diferença na preferência dos gestores quanto ao uso do modelo e do método atual de decisão.

Hipótese alternativa: preferências diferentes em relação ao uso do modelo e do método atual de decisão.

Medição necessária: preferência dos gestores em relação ao método de decisão.

Em relação à preferência, todos os gestores (10) responderam que preferem o modelo apresentado ao método atual de decisão.

6.2 Hipótese 2 – Utilidade

Hipótese nula: os gestores não consideram o modelo útil.

Hipótese alternativa: o modelo é útil.

Medição necessária: utilidade do modelo.

Em relação à utilidade, todos os gestores (10) consideraram o modelo útil para subsidiar o processo decisório de adoção de serviços de nuvem.

6.3 Hipótese 3 – Acurácia

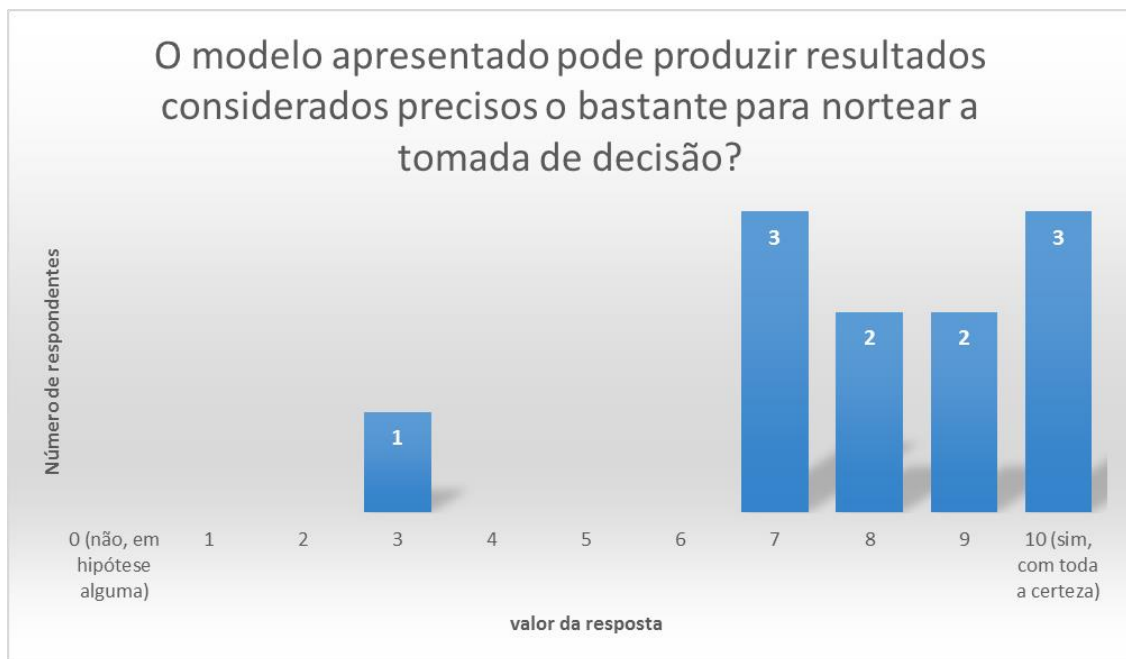
Hipótese nula: os gestores não consideram que o modelo possui acurácia em sua estimativa.

Hipótese alternativa: o modelo possui acurácia.

Medição necessária: acurácia do modelo.

Em relação à acurácia, 9 dos 10 gestores consideraram que o modelo apresentado pode produzir resultados considerados precisos o bastante para nortear a tomada de decisão de adoção de serviços de nuvem, atribuindo uma nota igual ou superior a 7, conforme ilustrado na Figura 53.

Figura 53 - respostas dos gestores em relação à acurácia do modelo



Fonte: a autora

6.4 Hipótese 4 – Efetividade

Hipótese nula: o modelo formulado não é efetivo para representar o processo decisório. Ao usar o modelo para representar a tomada de decisão, menos da maioria absoluta dos gerentes será capaz de identificar os elementos importantes no processo decisório.

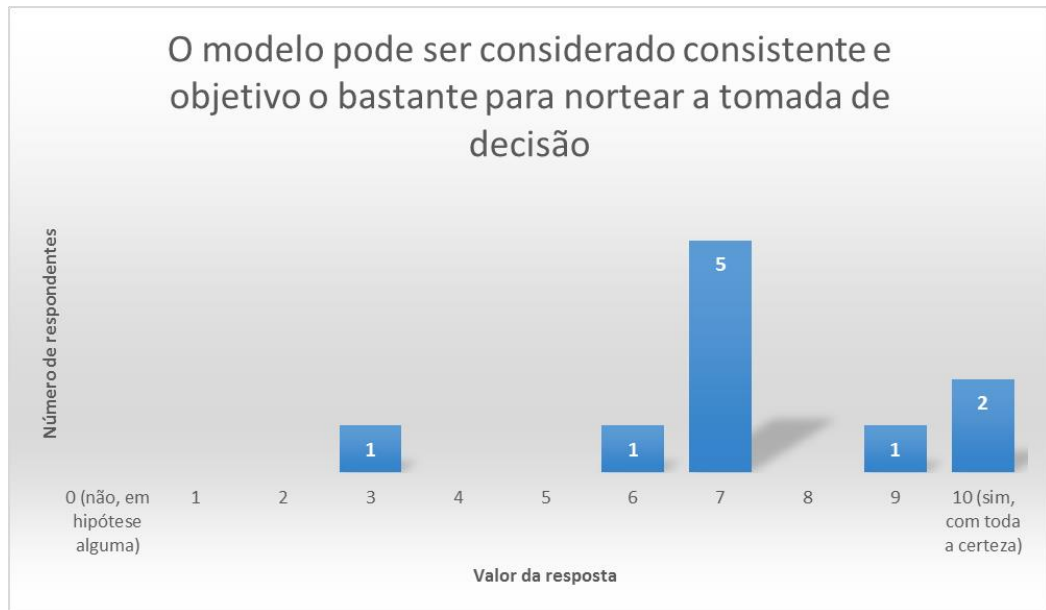
Hipótese alternativa: o modelo é efetivo para representar o processo decisório. Ao usar o modelo para representar a tomada de decisão, pelo menos a maioria absoluta dos gerentes será capaz de identificar os elementos importantes no processo decisório.

Por efetivo compreende-se:

- os resultados apresentados pelo modelo são consistentes e objetivos o bastante para basear a tomada de decisão;
- os resultados obtidos da aplicação do modelo de valor são compatíveis com os resultados obtidos por outros métodos que apoiam o processo de tomada de decisão nas organizações.

Em relação à efetividade, 8 dos 10 gestores consideraram o modelo consistente e objetivo o bastante para nortear a tomada de decisão, conforme ilustrado na Figura 54.

Figura 54 - Respostas dos gestores em relação à efetividade do modelo



Fonte: a autora

6.5 Hipótese 5 – Aprimoramento

Hipótese nula: o modelo construído não aprimora o processo decisório.

Hipótese alternativa: o modelo construído aprimora o processo decisório.

Em relação ao aprimoramento, todos os gestores (10) avaliaram que o modelo apresentado melhora o processo de tomada de decisão.

6.6 Dificuldades enfrentadas e necessidades adicionais

Além das respostas às hipóteses avaliadas, os gestores avaliaram as dificuldades na utilização do modelo. A grande maioria (7 dos 9 que responderam a essa questão) destacou a complexidade das redes de Petri. Alguns (4 dos 9) reforçaram também a necessidade de muitos dados de entrada. Todos os gestores consideraram que seria necessário a implementação de uma ferramenta de *software* automatizada para facilitar o uso do modelo, sendo que alguns (4 dos 9) também consideraram necessário, além da ferramenta, um programa de treinamento.

6.7 Sugestões e comentários

Alguns gestores forneceram comentários e sugestões de melhoria do modelo, que serão descritos neste tópico.

“Com relação à precisão do modelo, entendo que ele é bastante suscetível às premissas, que podem ser subjetivas. Portanto, ele precisaria ser reaplicado de tempos em tempos, com refinamento dos dados de acordo com a experiência dos responsáveis pelos dados qualitativos. Com essa consideração, o modelo é bastante útil para gestores tomarem decisões ou para vendedores demonstrarem os benefícios de seus produtos e serviços; para isso, esses gestores e vendedores precisam ser capacitados na utilização do modelo e do *software* programado para o fim desejado” (Wânia Gomes Pedrosa Ferrari).

“O modelo permite uma tomada de decisão consciente, e o fato de ter uma representação ‘visual’ é o que mais ajuda nesse sentido. Talvez seria interessante ter algum mecanismo/procedimento para garantir que o conjunto de dados de entrada está completo e que os pesos dos critérios qualitativos estão adequados. Quem sabe esses itens poderiam ser fornecidos por mais de uma pessoa e confrontados depois para fechar um consenso” (Henrico Scaranello).

“Quanto à precisão e à consistência do resultado, ficou bem claro que não se trata de uma decisão exclusiva da área de TI – quanto maior o conhecimento acerca das diversas áreas que correspondem aos fatores de peso inseridos, maior a eficácia do modelo. Portanto o modelo propõe indicativos bastante concretos, desde que os envolvidos conheçam bem cada fator analisado pelo modelo, desde custo e riscos a projeções futuras do mercado” (João Paulo A. de Aguiar).

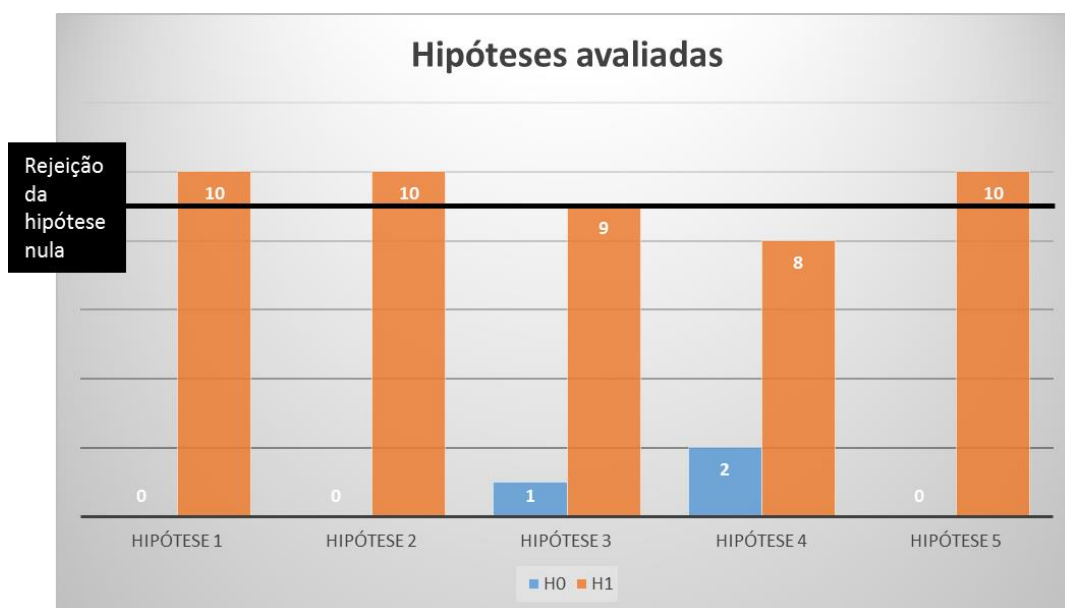
“O modelo permite uma avaliação e simulações do negócio, não somente da TI. Através dele é possível fazer simulações das diversas áreas da empresa, por exemplo, suas diretorias, e se conseguir através disso uma visão do negócio como um todo, ajudando a TI a cada vez mais se alinhar aos objetivos estratégicos do negócio” (Gustavo Costa de Carvalho).

6.8 Consolidação dos resultados

Para testar as hipóteses nulas, foi utilizado um teste binomial com nível de significância 5% e com 10 tentativas, correspondente ao número de respondentes. Resultado: a hipótese nula (H_0) pode ser rejeitada nos testes de hipóteses, quando pelo menos 9 pessoas entre 10 respondem com sucesso, o que corresponde ao percentual de 90%.

Considerando que uma resposta de valor 7 ou superior nas questões onde o respondente escolhia em uma escala de 0 a 10 indica que o respondente concorda com a afirmação, os resultados indicam que a hipótese nula foi rejeitada em todas as hipóteses, exceto na hipótese 4, que avalia a efetividade e objetividade do modelo, conforme ilustra a Figura 55.

Figura 55 - Comparativo entre as hipóteses avaliadas



Fonte: a autora

Capítulo 7

CONCLUSÃO E TRABALHOS FUTUROS

Apesar das dificuldades inerentes ao processo de tomada de decisão na adoção de serviços em nuvem, o modelo proposto nesta pesquisa se mostrou efetivo para o suporte a esse processo. O trabalho explorou aspectos quantitativos e qualitativos, de modo a obter uma visão abrangente da problemática que se deseja resolver.

Por utilizar uma ferramenta matemática (redes de Petri coloridas), os estudos puderam ser realizados através de extensivas simulações, levando a resultados interessantes. As facilidades do simulador CPN permitiram a obtenção desses resultados sem os custos financeiros que teriam sido necessários para obter as mesmas informações. As principais contribuições do trabalho são:

- A organização do problema decisório em um diagrama visual para facilitar o correto entendimento da influência dos principais fatores;
- O estudo detalhado de fatores importantes na adoção do modelo nuvem, especialmente na área de custos, com um exemplo do uso de instâncias *spot*.

O problema da tomada de decisão na adoção de serviços de nuvem é extremamente complexo e possui inúmeras especificidades a serem analisadas em cada cenário de uso. Assim, a criação de um modelo que apoie a decisão com base em fundamentos sólidos é uma tarefa muito extensa e requer um esforço considerável.

A avaliação dos custos envolvidos é o primeiro desafio. É uma tarefa que exige conhecimentos detalhados das diversas opções existentes, e de aspectos às vezes desconhecidos para muitos gestores sem familiaridade com as possibilidades oferecidas pelos serviços de nuvem. Além disso, as ofertas de serviços estão em constante atualização, dificultando ainda mais um conhecimento detalhado das opções. Por esses motivos, o levantamento inicial dos elementos de custo que serão utilizados como entrada de dados para o modelo é uma tarefa não trivial.

O capítulo 5 apresenta um exemplo onde um único elemento de custo (horas de uso de servidores) pode exigir uma simulação sofisticada para obter um valor de comparação minimamente confiável. Nesse exemplo, os custos são praticamente impossíveis de se estimarem sem uma ferramenta adequada, devido à inexistência de um valor fixo por hora. Soma-se a essa dificuldade, a imprevisibilidade na demanda pelos recursos computacionais.

Com esse exemplo, busca-se ilustrar as dificuldades enfrentadas pelos gestores ao estimar seus custos com os serviços de nuvem, bem como apresentar uma ferramenta de apoio a essa tarefa.

A avaliação qualitativa, por sua vez, também apresenta alto nível de complexidade. São diversos fatores envolvidos, que variam de acordo com as preferências e necessidades de cada cenário de uso. Para trabalhar um número elevado de fatores, o método AHP fornece uma ferramenta prática e bastante utilizada em cenários de serviços em nuvem. A vantagem do uso do AHP é permitir que o modelo se adapte facilmente às análises que o gestor precisar realizar, incluindo ou excluindo critérios de acordo com o cenário. Os critérios mais comuns, conforme a literatura, servem como um bom ponto de partida para elencar e trabalhar a avaliação qualitativa necessária.

Durante a validação de aparência, nota-se que o modelo concebido parece ter atingido os objetivos alcançados. A participação dos gestores durante a palestra de apresentação do modelo e os exemplos de utilização foi crescente, formulando questões cada vez mais consistentes sobre o problema da decisão. Os gestores que participaram do processo de validação expressaram uma forte aprovação do mesmo. Dessa forma, conclui-se que o modelo, apesar das inúmeras dificuldades inerentes ao próprio processo de decisão, é capaz de apoiar e auxiliar nessa difícil e complicada tomada de decisão.

7.1 Trabalhos futuros

Como trabalho futuro, a construção de uma ferramenta automatizada parece ser o ponto prioritário para utilização efetiva do modelo. As opiniões coletadas durante a validação de aparência indicam a forte necessidade dessa ferramenta.

Essa ferramenta facilitaria a coleta dos elementos de custo envolvidos e também da avaliação qualitativa, produzindo diretamente o resultado final, mas apresentando um relatório analítico indicando os fatores que mais influenciaram a decisão. Tais fatores poderiam, então, ser revisados pelos gestores, para garantir que não houve erro de avaliação que pudesse ter afetado significativamente a decisão.

Outro trabalho futuro inspirado nos comentários dos gestores que participaram da validação de aparência seria a incorporação de elementos de custo referentes à área fiscal. Especialmente no Brasil, onde a carga tributária é alta, e pode haver diferença significativa entre os valores de impostos e taxas cobrados nos dois modelos, é justificável o tratamento especial para esses elementos de custo. O modelo atual permite que essa análise seja realizada à parte e depois incorporada como um elemento de custo qualquer. No entanto, uma extensão

do modelo que tivesse um tratamento diferenciado poderia ajudar ainda mais o gestor na tomada de decisão.

A extensão do modelo para tratar os parâmetros de qualidade de serviço definidos no SLA também é um trabalho futuro vislumbrado. Os custos para se obter um padrão de qualidade mais apurado certamente são mais altos, e a avaliação qualitativa também muda de acordo com o padrão acordado. As simulações de diversos cenários de SLA distintos poderiam trazer informações interessantes para a tomada de decisão. A análise dos riscos, na área qualitativa, também é um trabalho futuro bastante promissor.

REFERÊNCIAS

ALFONSO, C. et al. An economic and energy-aware analysis of the viability of outsourcing cluster computing to a cloud. **Future Generation Computer Systems**, v. 29, p.704-712, mar.2013.

ALTMANN, J.; KASHEF, M. M. Cost model based service placement in federated hybrid clouds. **Future Generation Computer Systems**, v.41, p.79–90, dez. 2014.

AMAZON WEB SERVICES. Simple Monthly Calculator.
<http://calculator.s3.amazonaws.com/index.html>. Acesso em: set.2015.

AMAZON WEB SERVICES. Viewing Spot Instance Pricing History.
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances-history.html>. Acesso em: set.2015.

AMAZON WEB SERVICES. Purchase reserved instances. AWS Console.
<https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#ReservedInstances:sort=reservedInstancesId>. Acesso em: set.2015.

AMAZON WEB SERVICES. Instâncias reservadas do Amazon EC2.
<http://aws.amazon.com/pt/ec2/pricing/#reserved-instances>. Acesso em: set.2015.

AMAZON WEB SERVICES . AWS Management Console.
<http://aws.amazon.com/pt/console/>. Acesso em: dez.2014.

AVRAM, M.G. Advantages and challenges of adopting cloud computing from an enterprise perspective. **Procedia Technology**, v.12, p.529 – 534, 2014.

BAARS, T. et al. Chargeback for cloud services. **Future Generation Computer Systems**. v. 41, p. 91–103, dez.2014.

BARROSO, G.C. Introdução às Redes de Petri, Redes de Petri Coloridas e CPN Tools. [S.I:s.n].

BAYRAK, T. A decision framework for SME Information Technology (IT) managers: Factors for evaluating whether to outsource internal applications to Application Service Providers. **Technology in Society**, v.35, p.14–21, fev.2013.

BAZERMAN, M.; MOORE, D. **Processo Decisório**. Elsevier. 2004.

BENLIAN, A.; HESS, T. Opportunities and risks of software-as-a-service: Findings from a survey of IT executives. **Decision Support Systems**. v.52, p.232-246, dez.2011.

BESERRA, P. et al. Cloudstep: A Step-by-Step Decision Process to Support Legacy Application Migration to the Cloud. In: IEEE 6TH INTERNATIONAL WORKSHOP ON THE MAINTENANCE AND EVOLUTION OF SERVICE-ORIENTED AND CLOUD-BASED SYSTEMS (MESOCA), 6, 2012, Trnto, p.7-16.

BIBI, S.; KATSAROS, D.; BOZANIS, P. Business Application Acquisition: On-Premise or SaaS-Based Solutions? **IEEE Software**, v.29, p.86-93, jun.2012.

BOAMPONG, P.; WAHSHEH, L. Different Facets of Security in the Cloud. 15TH COMMUNICATIONS AND NETWORKING SIMULATION SYMPOSIUM, 15, 2012, Orlando.

BUYYA, R. et al. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. **Future Generation Computer Systems**, v.25, p.599-616, jun.2009.

CARDOSO, J.; VALETTE, R. **Redes de Petri**. Florianópolis: Editora da UFSC, 1997.

CHUN, S.; CHOI, B. Service models and pricing schemes for cloud computing. **Cluster Computing**, v.17, p.529–535, jun.2014.

COUTINHO, E. et al. Elasticidade em Computação na Nuvem: Uma Abordagem Sistemática. In: XXXI SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS. SBRC 2013 - Minicursos. [S.l.: s.n.], 2013.

CPN GROUP. CPN Tools. <http://cpntools.org/>. Acesso em: set.2015

DELL. Plataforma de infraestrutura compartilhada PowerEdge VRTX. <http://www.dell.com/br/empresa/p/poweredge-vrtx/pd>. Acesso em: set.2015.

E-FISCAL. Computing e-Infrastructure cost estimation and analysis - Pricing and Business Models. http://efiscal.eu/files/deliverables/D2%203%20Computing%20e-Infrastructure%20cost%20calculations%20and%20business%20_models_vam1-final.pdf. Acesso em: set.2013.

GAGNON, S. et al. The next web apps architecture - challenges for SaaS vendors. **IT Professional**, v.13, p.44-50, set. 2011.

GARG, S.K.; VERSTEEG, S.; BUYYA, R. A framework for ranking of cloud computing services. **Future Generation Computer Systems**, v.29, p.1012-1023, jun.2013.

GARG, R. STILLER, B. TrAdeCIS: Trade-off-based Adoption Methodology for Cloud-based Infrastructures and Services. In: 8th International Conference on Autonomous Infrastructure, Management and Security (AIMS 2014), Brno, Czech Republic, jul.2104.

GARTNER GROUP. Top 10 Strategic Technology Trends for 2014. In: GARTNER SYMPOSIUM ITXPO, 2013, Orlando.

GOMES FURTADO, C.; MARQUES SOARES, J.; CORDEIRO BARROSO, G. CacheSIM: A Web Cache Simulator Tool Based on Coloured Petri Nets and Java Programming. **Revista IEEE America Latina**, v.13, n.5, p.1511-1519, mai.2015.

HAAS, R.; MEIXNER, O. An illustrated guide to the analytic hierarchy process. <http://alaskafisheries.noaa.gov/sustainablefisheries/sslmc/july-06/ahptutorial.pdf>. Acesso em: nov.2013.

HAN, R. ET AL. Enabling cost-aware and adaptive elasticity of multi-tier cloud applications. **Future Generation Computer Systems**, v.32, p.82-98, mar.2014.

HO, W.; XU, X.; DEY, P. Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. **European Journal of Operational Research**, vol. 202, n.1, p.16-24, abr.2010.

INFOTECH RESEARCH GROUP. Database as a Service TCO Calculator. <http://www.infotech.com/research/ss/it-use-dbaas-to-drag-drop-and-deploy-enterprise-databases/it-database-as-a-service-tco-calculator>. Acesso em: set.2015.

JAIN, R. **Art of Computer Systems Performance Analysis Techniques For Experimental Design Measurements Simulation And Modeling**. [S.I.], Wiley Computer Publishing, 1991.

JAMSHIDI, P., AHMAD, A., & PAHL, C. Cloud Migration Research: A Systematic Review. **IEEE Transactions On Cloud Computing**, v.1, no.2, p.142-157, out.2013.

JAVADI, B.; THULASIRAM, R.; BUYYA, R. Characterizing spot price dynamics in public cloud environments. **Future Generation Computer Systems**, v.29, n.4, p.988-999, jun.2013.

JENSEN, K.; KRISTENSEN, L. **Coloured petri net - modelling and validation of concurrent systems**. New York, USA, Springer Verlag, 2009.

KAPLAN, R.; NORTON, D. **The Balanced ScoreCard**. Harvard Business Review Press, [S.I.], 1996.

KITCHENHAM, B.; PFIEEGER, S. Principles of Survey Research. Part 4: Questionnaire Evaluation, **ACM SIGSOFT Software Engineering Notes**, New York, NY, v.27, n.3, p.20-23, mai.2002.

LEE, S.; PARK, S.; LIM, G. Using balanced scorecards for the evaluation of “Software-as-a-service”. **Information & Management**, v.50, p. 553-561, 2013.

LENK, A. ET AL. What’s Inside the Cloud? An Architectural Map of the Cloud Landscape. In: ICSE WORKSHOP ON SOFTWARE ENGINEERING CHALLENGES OF CLOUD COMPUTING, 2009, Vancouver. p 23-31.

LIMA, A.S. **Modelo para melhoria contínua de serviços de Tecnologia da Informação sob o ponto de vista do negócio**. 2011. 371p. Tese (Engenharia de Teleinformática). Universidade Federal do Ceará, Fortaleza, 2011.

LIMA, A.; SAUVÉ, J.; SOUZA, J. Capturing the Quality and Business Value of IT Services Using a Business-Driven Model. **IEEE Transactions on Network and Service Management**, v.9, n.4, p.421-432, ago.2012.

MALAWSKI, M. FIGIELA, K.; NABRZYSKI, J. Cost minimization for computational applications on hybrid cloud infrastructures. **Future Generation Computer Systems**, v.29, n.7, p.1786-1794, set.2013.

MARSTON, S. et al. Cloud computing — The business perspective. **Decision Support Systems**, v.51, n.1, p.176-189, abr.2011.

MCGOUGH, A.S. et al. Comparison of a cost-effective virtual Cloud cluster with an existing campus cluster. **Future Generation Computer Systems**, v.41, p.65-78, dez.2014.

MERIGÓ, J.; GIL-LAFUENTE, A.M. New decision-making techniques and their application in the selection of financial products. **Information Sciences**, v.180, n.11, p.2085–2094, jun.2010.

MOURATIDIS, H. et al. A framework to support selection of cloud providers based on security and privacy requirements. **The Journal of Systems and Software**, v.86, n.9, p.2276-2293, set.2013.

MURATA, T. Petri Nets: properties, analysis and applications. **Proceedings of the IEEE**, v.77, n.4, p. 541-580, abr.1989.

NIST. The NIST Definition of cloud computing. Gaithersburg, 2011.
<http://csrc.nist.gov/publications/nistpubs/.../SP800-145.pdf>. Acesso em: set.2015.

PETERSON, J. **Petri Net Theory and the modeling of Systems**. The University of Texas at Austin, [s.n.], 1981.

PETRI, C.A. **Kommunikation mit Automaten**. Bonn:Institut für Instrumentelle Mathematik, Schriften des IIM Nr.2, 1962.

PRINCIPLED TECHNOLOGIES. Meet Database performance needs while reducing TCO with Dell Poweredge VRTX vs. a leading public cloud service, 2013.
http://www.principledtechnologies.com/Dell/Microsoft_Private_Cloud_TCO_0914.pdf. Acesso em set.2015

RIBAS, M. et al. Modeling the use of spot instances for cost reduction in cloud computing adoption using a Petri net framework. In: 2015 IFIP/IEEE INTERNATIONAL SYMPOSIUM ON INTEGRATED NETWORK MANAGEMENT (IM), 10th workshop on business-driven IT management (BDIM 2015), Ottawa, Canada, mai.2015,p.1428 – 1433.

RIBAS, M. et al. A Petri net-based decision-making framework for assessing cloud services adoption: The use of spot instances for cost reduction. **Journal of Network and Computer Applications**, v.57, p.102-118, nov.2015.

RIBAS, M. et al. Assessing cloud computing SaaS adoption for enterprise applications using a Petri net MCDM framework. In: Network Operations and Management Symposium (NOMS), 9th workshop on business-driven IT management (BDIM 2014), Krakow, mai.2014, p.1-6.

SARIPALLI, P.; PINGALI, G. MADMAC: Multiple Attribute Decision Methodology for Adoption of Clouds. In: 2011 IEEE INTERNATIONAL CONFERENCE ON CLOUD COMPUTING (CLOUD). Washington, DC, jul.2011, p.316-323.

SAUVÉ, J. et al. An Introductory Overview and Survey of Business-Driven IT Management. In: The First IEEE/IFIP International Workshop on Business-Driven IT Management, 2006. BDIM '06. p.1-10.

SOUSA, E. et al. A Modeling Approach for Cloud Infrastructure Planning Considering Dependability and Cost Requirements. **IEEE Transactions On Systems, Man, And Cybernetics: Systems**, v.45, n.4, p.549-448, abr.2015

SOUSA, F. et al. Gerenciamento de dados em nuvem: Conceitos, sistemas e desafios. In: SIMPÓSIO BRASILEIRO DE BANCOS DE DADOS, 2010, Belo Horizonte, p.101-130.

SRIPANIDKULCHAI, K.; SUJICHANTARARAT, S. A Business-Driven Framework for Evaluating Cloud Computing. IN: NETWORK OPERATIONS AND MANAGEMENT SYMPOSIUM (NOMS), 2012, 7th Workshop on Business Driven IT Management (BDIM), p.1335-1342.

TAMANINI, I.; PINHEIRO, P.R.; SANTOS, C.N. An hybrid approach of verbal decision analysis and machine learning. **Lecture Notes in Computer Science**, v.7413, p.126–131, 2012.

TANG, S. et al. A Framework for Amazon EC2 Bidding Strategy under SLA Constraints. **IEEE Transactions On Parallel And Distributed Systems**, v.25, n.1, p.2-11, jan.2014.

TECHNE ENGENHARIA E SISTEMAS LTDA. Plataforma Cronos - relatório de atividades de Marketing, 2014.

THE YANKEE GROUP. Understanding Total Cost of Ownership of a Hosted vs. Premises-Based CRM Solution.

http://www.avecon.gr/photos/research/yankee%20group_undestanding%20tco%20of%20a%20hosted%20vs.%20premises-based%20csm%20solution.pdf. Acesso em: out.2013.

TRIANANTAPHYLLOU, E. **Multi-Criteria Decision Making Methods: A Comparative Study**. [S.I.], Springer US, 2000.

ULLMAN, J. **Elements of ML Programming**. [S.I.], Prentice Hall. 1998

VAN DEN BOSSCHE, R. et al. Online cost-efficient scheduling of deadline-constrained workloads on hybrid clouds. **Future Generation Computer Systems**, v.29, n.4, p.973-985, jun.2013.

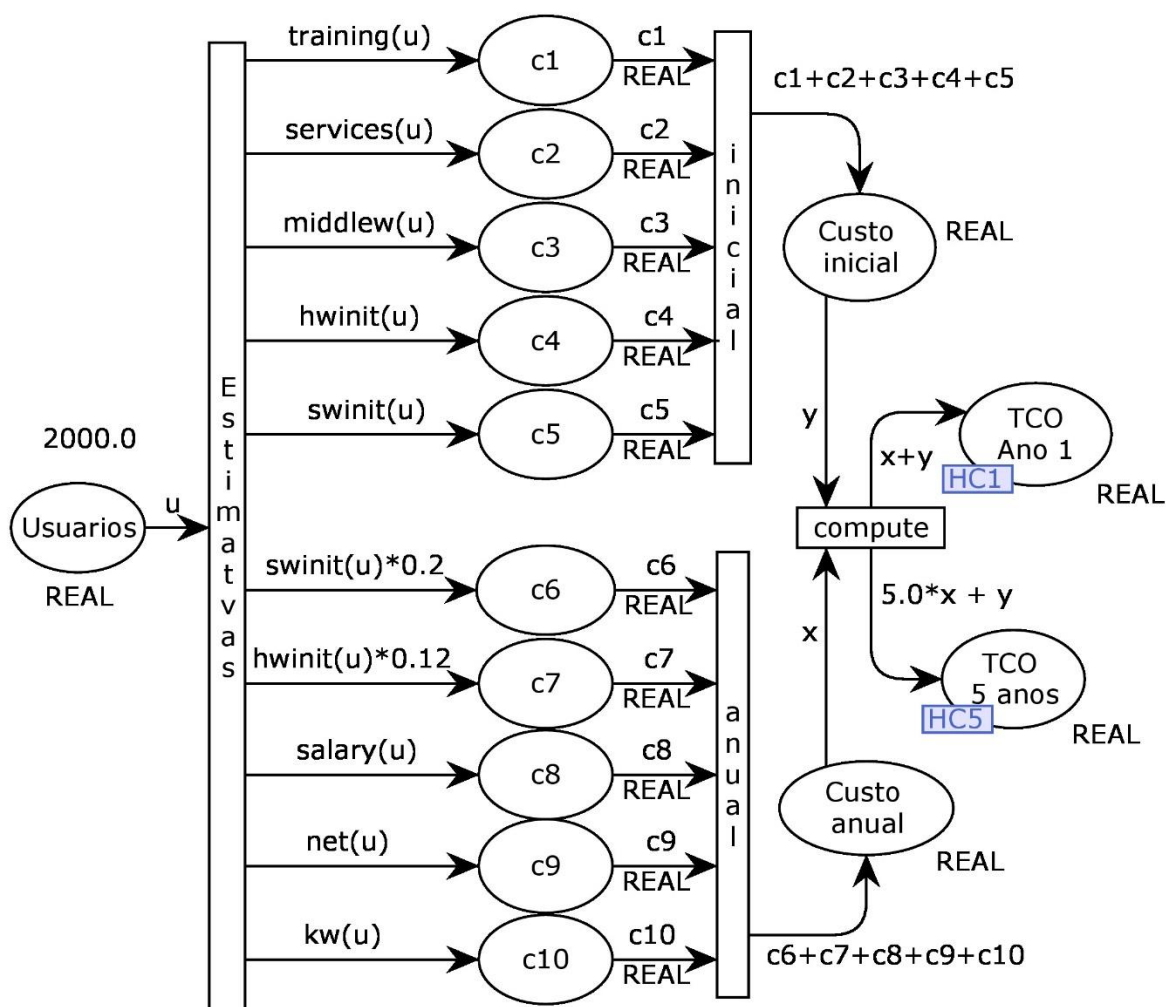
WU, W.W., LAN, L.; LEE, Y. Exploring decisive factors affecting an organization's SaaS adoption: A case study. **International Journal of Information Management**, v.31, n.6, p.556-563, dez.2011.

YUEN, K. (2012). Software-as-a-Service Evaluation in Cloud Paradigm: Primitive Cognitive Network Process Approach. In: *2012 IEEE INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING, COMMUNICATION AND COMPUTING*. Hong Kong, ago.2012, p.119-124.

APÊNDICE A – Rede preliminar apresentada no congresso BDIM 2014

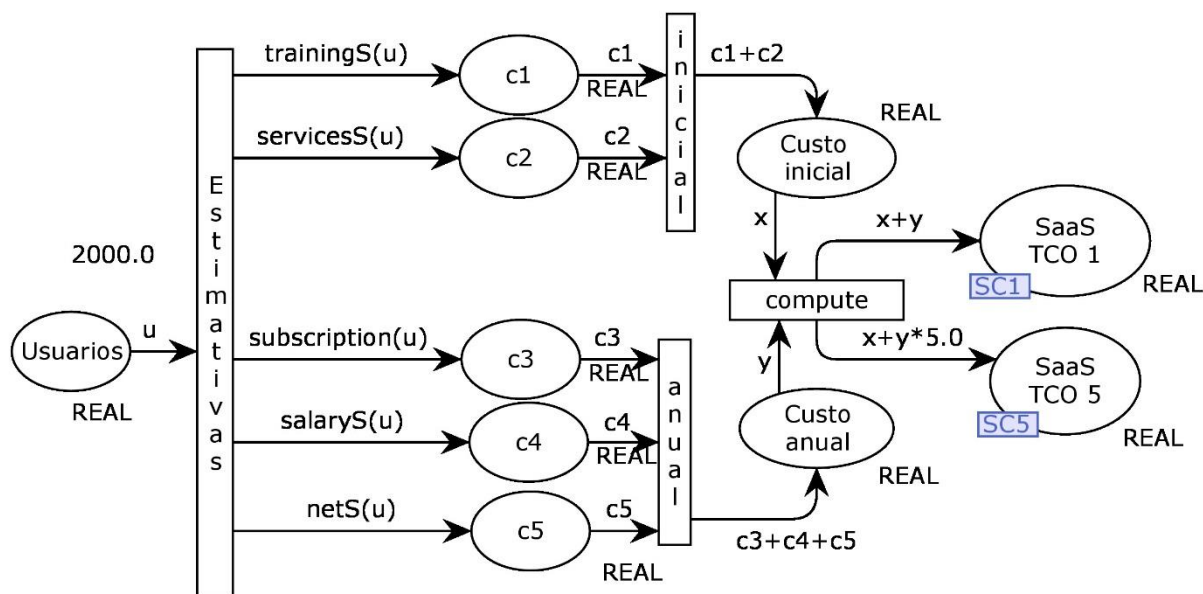
Este trabalho teve início com um artigo apresentado no congresso internacional BDIM 2014 (Ribas, et al., 2014). Na época, foi construído um modelo, também usando redes de Petri, mas específico para o cenário SaaS, onde os elementos de custo não eram livremente parametrizados como entrada do modelo. Inspirado no trabalho de (Bibi, Katsaros, & Bozanis, 2012), os principais elementos de custo no panorama SaaS foram explicitamente colocados na rede, e a única parametrização permitida era a quantidade de usuários, pois em geral é o principal parâmetro para avaliação de custo de *softwares* no modelo SaaS. As Figuras 56 e 57 ilustram o cálculo dos custos no modelo SaaS.

Figura 56 - Rede para cálculo de custos On-premises no modelo SaaS



Fonte: (Ribas, et al., 2014).

Figura 57 - Rede para cálculo de custos nuvem no modelo SaaS

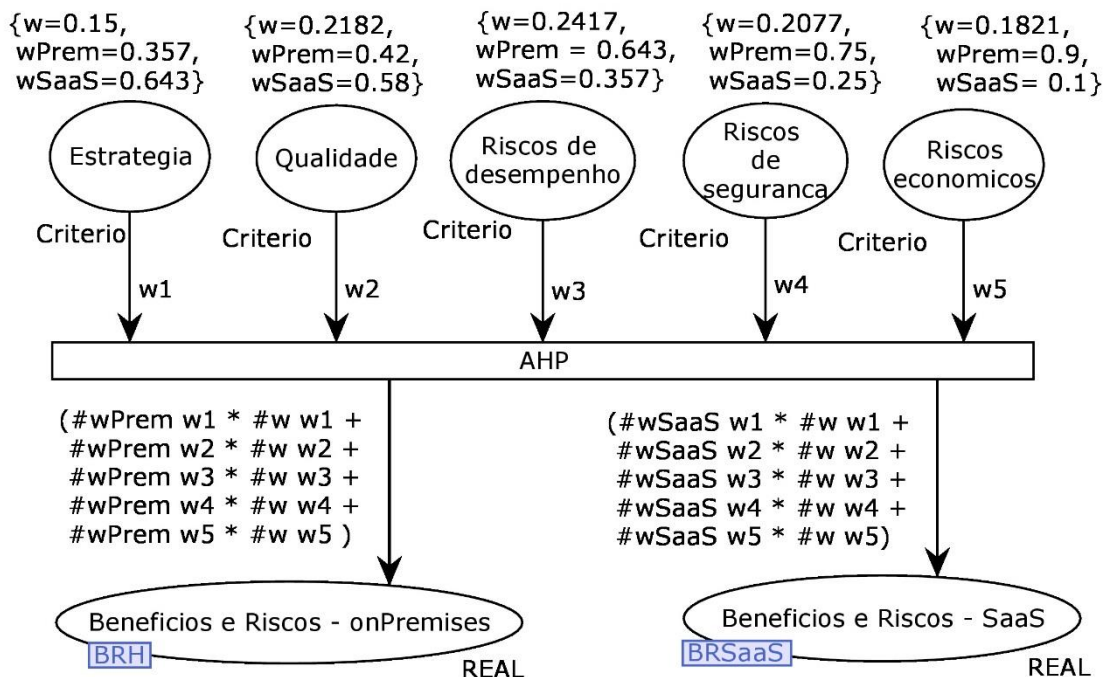


Fonte: apresentado em (Ribas, et al., 2014).

Note-se que, nessa versão preliminar, o modelo só calculava o TCO, não permitindo o uso de outros métodos como o NPV. Além disso, a lista de elementos de custo era fixa. Para a opção *on-premises*, os custos avaliados eram: treinamento (função *training*), serviços de customização (função *services*), licenças de *software* básico como sistema operacional e servidor de banco de dados (função *middle*), custos iniciais com *hardware* (função *hwinit*), custos iniciais com *software* (função *softwinit*), custos com pessoal (função *salary*), custos com conectividade (função *net*) e custos com eletricidade (função *kw*). Para a opção nuvem, os custos avaliados eram: treinamento (função *trainingS*), serviços de customização (função *servicesS*), assinatura do sistema no modelo SaaS (função *subscription*), custos com pessoal (função *salaryS*) e custos com conectividade (função *netS*).

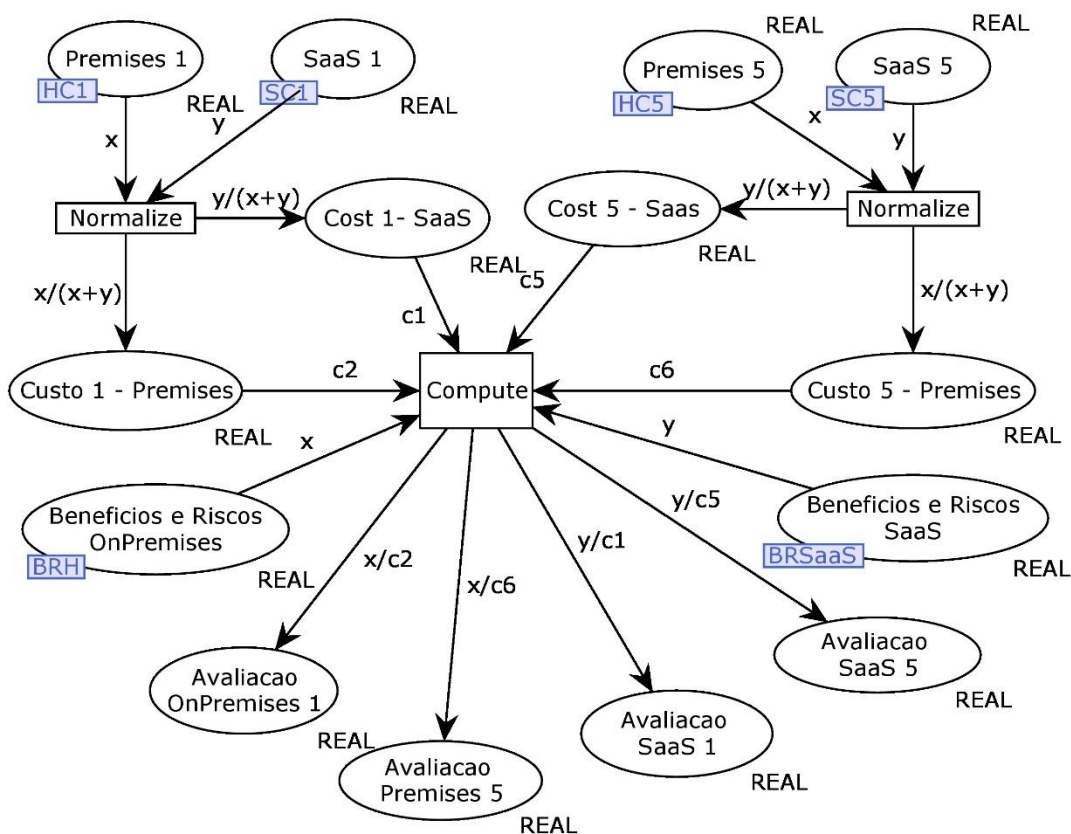
A análise qualitativa considerou os critérios pesquisados na época, que foram os mesmos utilizados ao longo deste trabalho, com algumas variações. A rede para avaliação explicitava os cálculos finais do método AHP, conforme ilustrado na Figura 58, não permitindo nenhum tipo de adaptação ou customização do método de decisão multicritério adotado. Nessa rede inicial, a avaliação qualitativa era armazenada em locais distintos, o que posteriormente levaria a um aspecto visual da parte final da rede (Figura 59) menos organizado, se comparado à versão final da rede apresentada neste trabalho.

Figura 58 - Rede para cálculo de avaliação qualitativa no modelo SaaS



Fonte: apresentado em (Ribas, et al., 2014).

Figura 59 - Rede para avaliação final das soluções



Fonte: apresentado em (Ribas, et al., 2014).

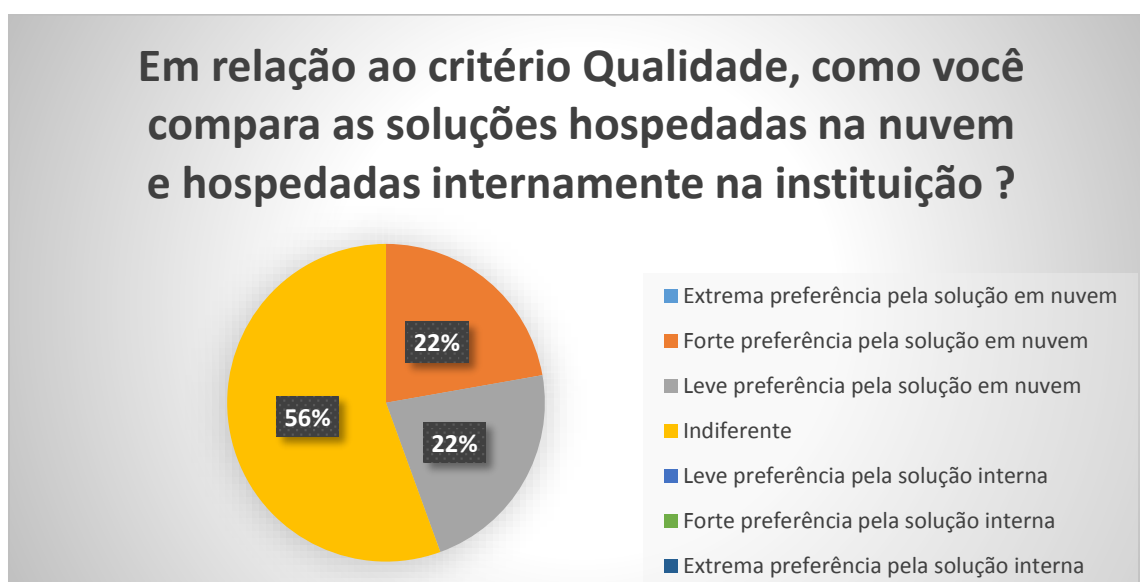
APÊNDICE B – Pesquisa qualitativa utilizada nos exemplos de uso do modelo

Para investigar com mais detalhe a opinião dos gerentes de TI em relação à nuvem, foi realizada uma pesquisa junto aos participantes do evento 6º Seminário de TI do Sindicato das Mantenedoras do Ensino Superior (Semesp), realizado em outubro de 2014, cujos resultados serão apresentados a seguir. A pesquisa foi realizada pela empresa Techne Engenharia e Sistemas, que atualmente conta com apoio financeiro da Financiadora de Estudos e Projetos (Finep) para desenvolvimento de um projeto que visa criar uma plataforma de desenvolvimento de aplicativos de gestão na nuvem (Techne Engenharia e Sistemas LTDA, 2014), e buscou seguir os padrões recomendados para fins científicos (Kitchenham & Pfieeger, 2002). O formulário utilizado para coleta de informações está no anexo A deste documento. A pesquisa busca analisar a preferência dos participantes comparativamente ao modelo tradicional *on-premises* e o modelo na nuvem.

B.1 Comparação entre alternativas: Critério Qualidade

O critério qualidade refere-se à eficiência e eficácia dos serviços que a aplicação fornecerá para melhorar os processos de negócio suportados. Em relação a esse critério, os resultados obtidos são apresentados na Figura 60.

Figura 60 - Avaliação qualitativa do critério Qualidade



Fonte: adaptado de (Techne Engenharia e Sistemas LTDA, 2014).

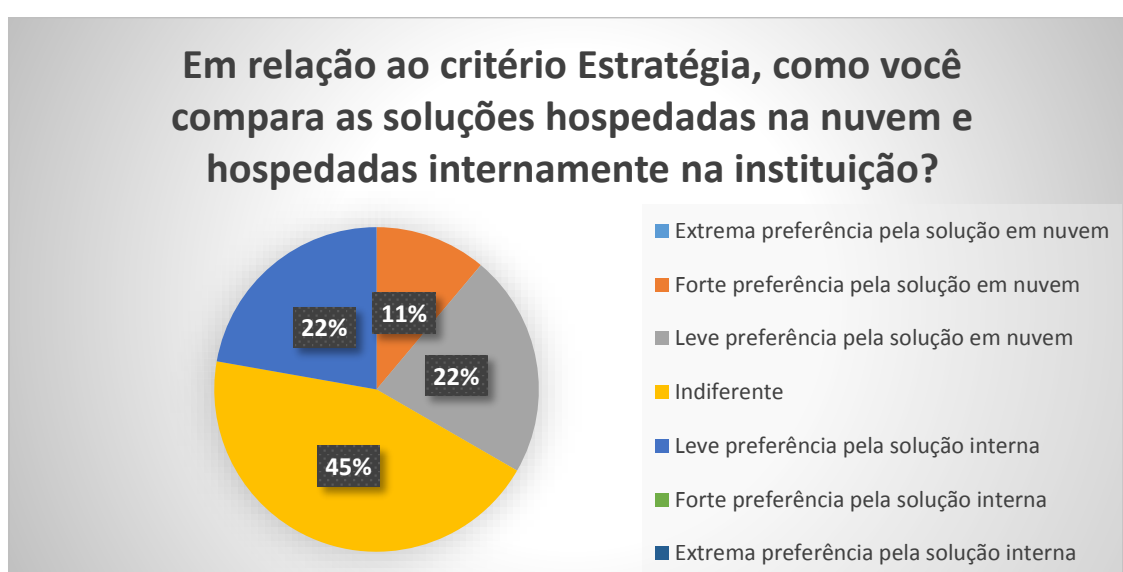
A maioria dos entrevistados (56%) respondeu que é indiferente em relação ao critério qualidade, considerando eficazes ambos os modelos. Uma parcela menor (44%)

manifestou preferência, leve ou forte pelo modelo nuvem, o que demonstra uma variação positiva para o modelo nuvem.

B.2 Comparação entre alternativas: Critério Estratégia

O critério estratégia refere-se a questões tais como: 1) flexibilidade para mudar de provedor de TI, 2) escalabilidade, 3) possibilidade de implantação mais rápida do *software*, 4) redução da dependência do fornecedor (*lock-in*) devido a menores custos iniciais e 5) possibilidade de concentrar esforços nas competências específicas da empresa. Em relação a esse critério, os resultados obtidos são apresentados na Figura 61.

Figura 61 - Avaliação qualitativa do critério Estratégia



Fonte: adaptado de (Techne Engenharia e Sistemas LTDA, 2014).

Novamente a maioria (45%) se diz indiferente ao modelo adotado, e tem-se uma parcela significativa (33%) que prefere o modelo nuvem. Nesse critério, diferentemente do critério qualidade, aparece uma parcela (22%) que possui leve preferência pelo modelo de solução interna.

B.3 Comparação entre alternativas: Critério Risco

O critério risco refere-se a questões tais como: possibilidade de corrupção ou vazamento de dados, erros de autenticação, ou possibilidade de ter que pagar mais para atingir o nível esperado de serviço do que o inicialmente previsto - os chamados "custos escondidos". Em relação a esse critério, os resultados são apresentados na Figura 62.

No critério risco fica evidente que uma parcela significativa ainda possui ressalvas ao modelo nuvem, pois embora a maioria dos entrevistados (34%) declare ser indiferente,

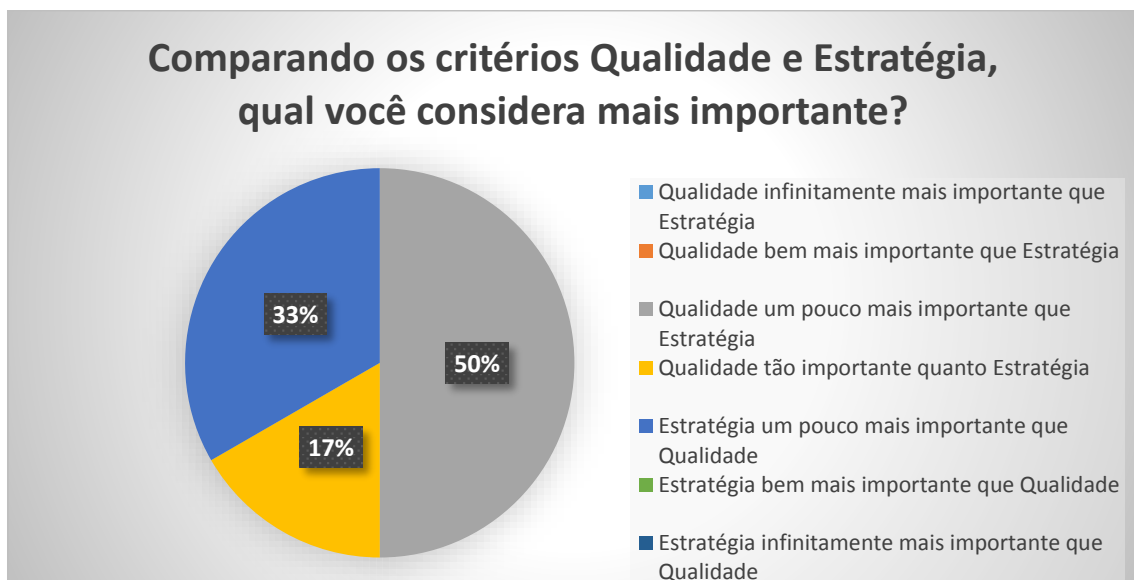
somente 22% tem preferência pelo modelo nuvem contra um total de 44% dos que disseram ter preferência leve ou forte pelo modelo de solução interna.

Figura 62 - Avaliação qualitativa do critério Risco



Fonte: adaptado de (Techne Engenharia e Sistemas LTDA, 2014).

Figura 63 - Comparação entre critérios estratégia e qualidade



Fonte: adaptado de (Techne Engenharia e Sistemas LTDA, 2014).

B.4 Comparação entre critérios: Qualidade x Estratégia

Na comparação entre os critérios, pode-se notar, na Figura 63, que qualidade prevalece sobre a estratégia, com a maioria 50% contra os 33% que opinaram que a estratégia é um pouco mais importante que a qualidade. O que demonstra claramente que o critério qualidade é extremamente importante na escolha de uma solução.

B.5 Comparação entre critérios: *Qualidade x Riscos*

Na comparação entre os critérios nota-se na Figura 64 que aspectos relacionados à eficiência e eficácia dos serviços são mais importantes que questões como a possibilidade de corrupção ou vazamento de dados. Os entrevistados consideraram em maior ou menor escala que o critério qualidade ainda é mais importante que o risco potencial.

Figura 64 - Comparação entre critérios qualidade e riscos



Fonte: adaptado de (Techne Engenharia e Sistemas LTDA, 2014).

B.6 Comparação entre critérios: *Estratégia x Riscos*

A maioria dos entrevistados respondeu que as questões relativas ao risco são mais importantes que as questões relativas a estratégia, conforme ilustrado na Figura 65. Somados aos que responderam riscos bem mais importante que estratégia (16%) aos que responderam riscos um pouco mais importante que estratégia serão 49% contra 34% da soma dos que responderam estratégia um pouco mais importante que riscos (17%) e estratégia bem mais importante que riscos (17%).

B.7. Considerações adicionais

Esta pesquisa não tem a responsabilidade de diagnosticar todos os aspectos qualitativos em relação à adoção do modelo nuvem em instituições de ensino superior, porém serve de indicador de tendências e dá algum norte de direção sobre o que pensam os gestores da infraestrutura dessas instituições. Os resultados demonstram que o mercado ainda está

receoso com esse novo paradigma, porém os gestores de TI possuem a visão de que a infraestrutura na nuvem pode trazer benefícios, apesar de ainda ser vista com reservas, pelo menos no conservador segmento educacional.

Figura 65 - Comparação entre critérios estratégia e riscos



Fonte: adaptado de (Techne Engenharia e Sistemas LTDA, 2014).

B.8 Instrumento de Coleta De Dados – Pesquisa Qualitativa

QUESTIONÁRIO

Por favor, atribua uma nota de 1 a 9 para cada critério abaixo:

- 1) **Estratégia:** Aspectos relacionados a:
 - a. flexibilidade para mudança de provedor de infraestrutura,
 - b. escalabilidade (capacidade de atender aumento de demanda),
 - c. foco na atividade fim da empresa
- 2) **Qualidade:** como as soluções se comparam em termos da eficiência e da efetividade dos processos de negócio que o *software* suporta
- 3) **Riscos:**
 - a. desempenho,
 - b. segurança,
 - c. econômicos

Por favor, indique a importância relativa de cada critério. Por exemplo, se acha que estratégia é muito mais importante que qualidade, coloque 9 para qualidade e 1 para estratégia:

	Nota									Em relação a	Nota								
	1	2	3	4	5	6	7	8	9		1	2	3	4	5	6	7	8	9
Estratégia										Qualidade									
Qualidade										Riscos									
Riscos										Estratégia									

Como você avalia as soluções de aplicativos em nuvem, em relação às soluções de aplicativos instalados internamente no setor de TI da sua empresa – *on-premises*?

Notas comparativas:

	Software em nuvem (nota)									Em relação a	Software instalado internamente (nota)								
	1	2	3	4	5	6	7	8	9		1	2	3	4	5	6	7	8	9
Estratégia										Estratégia									
Qualidade										Qualidade									
Riscos										Riscos									

Qual a característica que você considera mais positiva no modelo nuvem?

E a negativa?

APÊNDICE C – Questionário para validação de aparência

O questionário foi criado usando uma ferramenta (*surveymonkey*) que permite a coleta de dados através de um navegador web. A seguir são apresentadas as imagens das telas que foram apresentadas aos gestores.

Questionário de Entrevista – Avaliação do modelo para Tomada de Decisão na Adoção de Serviços de Nuvem

Avaliação do Modelo

1. Preferência, Utilidade, Melhoria

	Sim	Não
Você prefere o modelo avaliado em relação ao processo atual executado na organização?	<input type="radio"/>	<input type="radio"/>
O resultado geral do modelo em relação ao tratamento dos itens avaliados é considerado útil para subsidiar o processo decisório de adoção de serviços de nuvem	<input type="radio"/>	<input type="radio"/>
Você acha que o modelo apresentado melhora o processo de tomada de decisão?	<input type="radio"/>	<input type="radio"/>

2. Precisão e consistência

	0 (não, em hipótese alguma)	1	2	3	4	5	6	7	8	9	10 (sim, com toda certeza)
Você acha que o modelo apresentado pode produzir resultados considerados precisos o bastante para nortear a tomada de decisão de adoção de serviços de nuvem ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O modelo pode ser considerado consistente e objetivo o bastante para nortear a tomada de decisão ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. Entre as dificuldades que você enfrentou ao conhecer o modelo, marque as que estiverem na lista abaixo:

- Alta curva de aprendizado do modelo matemático
- Muitos dados de entrada
- Redes de Petri são complexas, mas eu não preciso conhecer esses conceitos para utilizar o modelo, em uma situação normal na organização

4. Marque entre as opções abaixo as que você considera como necessárias para uma implantação efetiva do modelo avaliado em uma organização:

- Automação em uma ferramenta de software
- Programa de treinamento
- Consultoria externa

5. Comente nesta área, sobre alguns aspectos dessa pesquisa que você considera que devem ser enfatizados sobre o modelo avaliado, além das respostas objetivas que foram respondidas (possíveis utilizações e aplicabilidades, etc.):

6. Nome (opcional)

7. Cargo/função (opcional)

8. Data e Hora

9. Experiência e participação

	Alta	Média	Baixa
Experiência em atividades relacionadas a serviços de nuvem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participação no processo de decisão	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>