



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# Uma proposta de migração de sistemas legados do governo para a nuvem

Breno Gustavo Soares da Costa

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Orientadora  
Prof.a Dr.a Priscila Solís

Brasília  
2018

Ficha catalográfica elaborada automaticamente,  
com os dados fornecidos pelo(a) autor(a)

Gp Gustavo Soares da Costa, Breno  
Uma proposta de migração de sistemas legados do governo para a nuvem / Breno Gustavo Soares da Costa; orientador Priscila América Solis. -- Brasília, 2018.  
110 p.

Dissertação (Mestrado - Mestrado Profissional em Computação Aplicada) -- Universidade de Brasília, 2018.

1. Migração de sistemas legados. 2. Computação em Nuvem. 3. Percepção de riscos. I. América Solis, Priscila, orient. II. Título.



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## Uma proposta de migração de sistemas legados do governo para a nuvem

Breno Gustavo Soares da Costa

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Prof.a Dr.a Priscila Solís (Orientadora)  
CIC/UnB

Prof. Dr. André Costa Drummond    Prof. Dr. Georges Daniel Amvame-Nze  
Universidade de Brasília                      Universidade de Brasília

Prof.a Dr.a Aletéia Favacho  
Coordenadora do Programa de Pós-graduação em Computação Aplicada

Brasília, 9 de agosto de 2018

# Dedicatória

Dedico este trabalho à minha esposa, Tatiana, e aos meus filhos, Pedro e Heitor. Foi com parte do tempo que eu poderia ter dedicado a eles que eu escrevi essa dissertação. Aos pequenos, espero ter dado o exemplo de que vale a pena sacrificar parte do que é precioso para nós, em busca de algo que nos complete e realize.

# Agradecimentos

Agradeço à Tatiana pelo apoio incondicional e irrestrito, pela cumplicidade e pelo incentivo que me dá, há 28 anos, desde que estamos juntos.

Agradeço aos meus colegas de trabalho, representados pelo José Renato, que foi quem viabilizou a minha participação no mestrado, por assumirem carga maior de trabalho nos períodos em que estive ausente priorizando o estudo.

Agradeço à minha orientadora, Priscila Solís, por me mostrar caminhos viáveis de forma construtiva, com seu afiado poder de síntese.

Agradeço aos demais professores e colegas do PPCA, pelos ensinamentos, pela paciência (nem todos!), pela disponibilidade e solicitude.

# Resumo

Com a utilização da Computação em Nuvem para provisionamento de serviços de Tecnologia da Informação, as organizações podem migrar sistemas legados para o novo ambiente e usufruir dos benefícios deste modelo. Há na literatura diversas propostas de modelagem dos elementos críticos a serem considerados na migração, validados por estudos de caso específicos. Há também alguns modelos de referência, definidos a partir das modelagens citadas e criados com o intuito de consolidar a literatura e ampliar suas aplicações. Este trabalho de pesquisa seleciona um modelo de referência de migração de sistemas legados para a nuvem e propõe um método para calcular um indicador de percepção de risco no processo de migração. Este indicador tem como objetivo definir a ordem de migração de um conjunto de sistemas de uma organização específica, no caso deste trabalho, uma organização da Administração Pública Federal brasileira. Prova de conceito no domínio do governo mostrou a aplicabilidade do método e um estudo de caso foi implementado com três sistemas legados do governo. Com o objetivo de validar o método proposto, foi realizada uma pesquisa de opinião, com uma amostra de 193 profissionais de tecnologia da informação do governo, para analisar o valor percebido da proposta. Os resultados obtidos mostraram que há percepção de valor para a proposta e permitiram ajustar os parâmetros do cálculo do indicador para que refletissem os achados da pesquisa. Ao aplicar os ajustes ao estudo de caso, entretanto, não houve alteração significativa no valor do indicador calculado, o que pode apontar um acerto nos parâmetros da proposta desde a concepção. Os achados deste trabalho apontam que um indicador de percepção de risco pode agregar valor real a um processo de migração de sistemas legados para Nuvem no âmbito da Administração Pública Federal brasileira.

**Palavras-chave:** computação em nuvem, migração de sistemas legados, percepção de risco

# Abstract

With the utilization of Cloud Computing, organizations can migrate legacy systems to the cloud in order to reach cloud related benefits and to preserve the investments made on the legacies. There are in the literature several proposals to modeling the critical elements of migration, validated by specific case studies. There are also some reference models, which were defined on top of these proposals with the intention of literature consolidation, trying to expand their applicability. This research selected a reference model for migrating legacy systems to the cloud and proposed a method to calculate a risk perception indicator to the migration process. This indicator can be used to define a rank of migration from a legacy systems set of a specific organization. In the case of this work, a Brazilian Federal Public Administration organization. A proof of concept in government domain has shown the method's applicability and a case study has been implemented with three government legacy systems. In order to validate the proposed method, a survey was conducted with a sample of 193 government information technology professionals to analyze the perceived value of the proposal. The results pointed out that this research's proposal is perceived as important and the parameters of indicator's calculation were adjusted to reflect them. When applying the adjustments to the case study, however, there was no significant change in the value of the calculated indicator, which may indicate a correctness in the parameters of the proposal from conception. The findings of this study indicate that a risk perception indicator can add real value to a legacy systems migration process to the Cloud, within the scope of the Brazilian Federal Public Administration.

**Keywords:** cloud computing, legacy systems migration, risk perception

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa do Tema . . . . .	1
1.2	Contribuição Esperada . . . . .	2
1.3	Estrutura do Trabalho . . . . .	3
<b>2</b>	<b>Fundamentação teórica</b>	<b>4</b>
2.1	Conceitos de computação em nuvem . . . . .	4
2.2	Benefícios no uso da computação em nuvem . . . . .	5
2.2.1	Benefícios específicos para governo . . . . .	5
2.3	Sistemas legados . . . . .	6
2.4	Migração para nuvem . . . . .	7
2.5	Percepção de risco . . . . .	8
2.6	Modelos de referência . . . . .	9
2.7	Conclusão deste capítulo . . . . .	10
<b>3</b>	<b>Migração de sistemas legados para a Nuvem</b>	<b>11</b>
3.1	Estratégias e métodos de migração . . . . .	11
3.2	Modelo de caracterização . . . . .	13
3.3	Modelo de avaliação . . . . .	15
3.4	Modelo de referência . . . . .	17
3.5	Conclusão deste capítulo . . . . .	19
<b>4</b>	<b>Proposta</b>	<b>22</b>
4.1	Migração de sistema legado como um projeto . . . . .	22
4.2	Avaliação do modelo de referência . . . . .	24
4.3	Método . . . . .	25
4.3.1	Passo 1 . . . . .	25
4.3.2	Passo 2 . . . . .	25
4.3.3	Passo 3 . . . . .	25
4.3.4	Passo 4 . . . . .	27



4.4	Prova de conceito . . . . .	27
4.4.1	Fase de planejamento . . . . .	28
4.4.2	Fase de projeto . . . . .	29
4.4.3	Representação visual do conjunto de avaliação de cada fase . . . . .	30
4.5	Estudo de caso . . . . .	31
4.5.1	Sistema A - Nuvem Cívica . . . . .	32
4.5.2	Sistema B - Pesquisa de jurisprudência . . . . .	32
4.5.3	Sistema C - Ambiente Virtual de Aprendizagem . . . . .	33
4.5.4	Indicador de percepção de risco e ranking de migração dos sistemas . . . . .	33
4.6	Conclusão deste capítulo . . . . .	34
<b>5</b>	<b>Análise experimental</b>	<b>35</b>
5.1	Método de coleta dos dados . . . . .	35
5.1.1	Pesquisa de opinião na Administração Pública . . . . .	35
5.1.2	Estrutura do questionário . . . . .	36
5.1.3	Descrição da amostra . . . . .	40
5.2	Análise dos resultados . . . . .	43
5.2.1	Verificação de confiabilidade . . . . .	43
5.2.2	Resultados . . . . .	45
5.3	Ajustes da proposta . . . . .	50
5.3.1	Fórmula 1 . . . . .	50
5.3.2	Fórmula 2 . . . . .	51
5.3.3	Fórmula 3 . . . . .	52
5.3.4	Discussão . . . . .	53
5.4	Conclusão deste capítulo . . . . .	54
<b>6</b>	<b>Conclusões e trabalhos futuros</b>	<b>56</b>
6.1	Trabalhos Futuros . . . . .	57
	<b>Referências</b>	<b>58</b>
	<b>Anexo</b>	<b>62</b>
<b>I</b>	<b>Convite enviado por e-mail</b>	<b>63</b>
<b>II</b>	<b>Questionário completo da Pesquisa de Opinião</b>	<b>65</b>
<b>III</b>	<b>Artigo publicado no congresso CLOSER 2018 - Ilha da Madeira - Portugal [1]</b>	<b>76</b>

IV Artigo publicado no congresso CLOSER 2018 - Ilha da Madeira -  
Portugal [2]

87

# Lista de Figuras

3.1	Cloud-RMM - modelo de caracterização proposto por Jamshidi et al. [3] . . .	14
3.2	Modelo de avaliação proposto por Gholami et al. [4]. . . . .	16
3.3	Elementos-chave do Modelo de Referência de migração de sistemas legados para a nuvem proposto por Gholami et al. [5]. . . . .	18
3.4	Fase de Planejamento do modelo de referência proposto por Gholami et al. [5] . . . . .	20
3.5	Fase de Projeto do modelo de referência proposto por Gholami et al. [5] . .	20
3.6	Fase de Habilitação do modelo de referência proposto por Gholami et al. [5]	21
4.1	Ciclo de vida de projetos, definido no PMBOK [6]. . . . .	22
4.2	Riscos e custos de projetos em relação ao tempo - PMBOK [6]. . . . .	23
4.3	Fase de Planejamento com indicativo visual do conjunto de avaliação. . . .	31
4.4	Fase de Projeto com indicativo visual do conjunto de avaliação. . . . .	31
5.1	Pergunta 1. . . . .	38
5.2	Pergunta 2. . . . .	39
5.3	Pergunta 3. . . . .	39
5.4	Pergunta 4. . . . .	40
5.5	Acompanhamento das respostas no LimeSurvey [7]. . . . .	42
5.6	Distribuição das respostas da Pergunta 1. . . . .	44
5.7	Distribuição das respostas por tarefa da Pergunta 4. . . . .	44
5.8	Importância de se usar um processo de apoio à migração. . . . .	46
5.9	Representatividade do indicador de percepção de risco, conforme proposto por esse estudo. . . . .	47
5.10	Se considera boa prática iniciar a migração por sistemas com baixa percep- ção de risco. . . . .	48
5.11	Importância atribuída a cada uma das seis tarefas selecionadas para compor o indicador de percepção de risco. . . . .	49

5.12 Importância atribuída à tarefa ‘Contratar o provedor de nuvem’ por faixa de experiência com TI. . . . . 49

# Lista de Tabelas

3.1	Alguns resultados selecionados da avaliação feita por Gholami et al. [4]. . .	17
4.1	Grupo de tarefas de avaliação da Fase de Planejamento. . . . .	28
4.2	Grupo de tarefas de avaliação da Fase de Projeto . . . . .	29
4.3	Avaliação da percepção de risco associada às tarefas no conjunto de avaliação das fases Planejamento e Projeto. . . . .	34
4.4	Cálculo do Indicador de Percepção de Risco para cada sistema avaliado. . .	34
5.1	Relacionamento dos objetivos específicos deste trabalho com as perguntas do questionário. . . . .	36
5.2	Resultados do teste Signed Rank de Wilcoxon para os itens propostos. . . .	46
5.3	Percentual dos respondentes que consideraram importante cada tarefa do cálculo do indicador. . . . .	48
5.4	Fórmula 1 - Pesos ajustados das tarefas que compõem o indicador. . . . .	51
5.5	Fórmula 1 - Mudança no valor do Indicador de Percepção de Risco do Sistema B (Seção 4.5), após ajustes nos pesos. . . . .	51
5.6	Fórmula 2 - Pesos ajustados das tarefas que compõem o indicador. . . . .	52
5.7	Fórmula 2 - Mudança no valor do Indicador de Percepção de Risco do Sistema B (Seção 4.5), após ajustes nos pesos. . . . .	52
5.8	Fórmula 3 - Pesos ajustados das tarefas que compõem o indicador. . . . .	52
5.9	Fórmula 3 - Mudança no valor do Indicador de Percepção de Risco do Sistema B (Seção 4.5), após ajustes nos pesos. . . . .	53
5.10	Pesos ajustados das tarefas que compõem o indicador de percepção de risco.	53
5.11	Cálculo do Indicador de Percepção de Risco para cada sistema avaliado. . .	54

# Lista de Abreviaturas e Siglas

**APF** Administração Pública Federal.

**CN** Computação em Nuvem.

**IaaS** Infrastructure as a Service.

**NIST** National Institute of Standards and Technology.

**PaaS** Platform as a Service.

**PMBOK** Project Management Book of Knowledge.

**SaaS** Software as a Service.

**SOA** Service Oriented Architecture.

**TCU** Tribunal de Contas da União.

**TI** Tecnologia da Informação.

# Capítulo 1

## Introdução

Neste capítulo será descrito o tema de pesquisa e os argumentos que o justificam, assim como os objetivos a serem alcançados com a proposta desenvolvida neste trabalho de dissertação de mestrado.

### 1.1 Justificativa do Tema

A Computação em Nuvem (CN) tem sido vista por especialistas do setor como algo capaz de revolucionar a Tecnologia da Informação (TI), uma vez que ela altera a forma como a TI é provida e consumida, mudando o panorama atual em que as organizações gerenciam o próprio parque computacional para outro em que a TI é consumida como serviço [8].

Segundo o Gartner [9], o uso de serviços em nuvem pública está crescendo a um ritmo acelerado e isso pode ser observado pelas taxas de crescimento significativas dos provedores de serviços, como a Amazon Web Services (AWS) [10] e Microsoft Azure [11]. Em uma pesquisa citada no estudo, realizada com quase 3000 participantes, 21% deles já utilizavam serviços de nuvem pública, enquanto que 56% planejavam implementar a nuvem até o final de 2017. Para muitas dessas organizações, a modernização da TI é quase sinônimo de uso crescente de computação em nuvem.

Alguns estudos procuram elucidar as principais vantagens ou benefícios na utilização ou na migração para este novo paradigma. O primeiro benefício é a potencial redução dos investimentos em infraestrutura de TI, apresentado por vários autores [12] [13] [9], colocando-se como vantagem especialmente no uso da nuvem pública. Mas há outros benefícios, como a transferência do risco de super provisionamento da infraestrutura para o provedor de nuvem e a rapidez e facilidade de se modificar a quantidade de recursos computacionais em uso, que é conhecida como elasticidade. A elasticidade, inclusive, pode evitar processos de compras adicionais, além de oferecer maior poder computacional simultâneo às cargas de trabalho intensivas em processamento, diminuindo o tempo de

execução sem causar acréscimo de custo. Um outro benefício citado é a redução do esforço operacional decorrente do uso de serviços de nuvem de mais alto nível, como uso da plataforma de execução de contêineres ou de bancos de dados na nuvem, cujas infraestruturas de apoio são gerenciadas pelo provedor de nuvem.

Por outro lado, de acordo com Armbrust et al. [13], a natureza multifacetada da Computação em Nuvem introduz desafios potenciais relacionados à segurança, à mudança organizacional, à incompatibilidade e interoperabilidade, à legislação, à regulação, ao custo de migração e ao aprisionamento pelo provedor.

O acórdão 1739/2015 [14] do Tribunal de Contas da União (TCU) faz uma análise detalhada das ofertas de Computação em Nuvem pública e apresenta os riscos inerentes ao modelo e os benefícios do seu uso em um domínio de governo. O estudo que deu origem ao acórdão foi solicitado pelo TCU para que servisse de base às fiscalizações futuras das contratações de CN da Administração Pública Federal (APF).

Em 2016, o Ministério do Planejamento Orçamento e Gestão (MPOG) publicou norma [15] que veta aos órgãos do Poder Executivo da APF a contratação e expansão de *data-centers*, orientando que esses órgãos contratem Computação em Nuvem como forma de expansão ou substituição de infraestrutura.

Assim, espera-se que haja diversas contratações de Computação em Nuvem nos próximos meses, iniciando um ciclo continuado de adoção desta plataforma pela APF e, como consequência, que haja a migração de parte dos sistemas legados dos órgãos para a nuvem pública, tanto para se adequar aos normativos vigentes, quanto para ter acesso aos benefícios de seu uso.

Segundo Pahl and Xiong [16], a migração para a nuvem é o processo de implantação, parcial ou total, de ativos digitais, serviços, recursos de TI ou sistemas de uma organização na Nuvem. O processo de migração para a Nuvem pode envolver a permanência de parte da infraestrutura dentro da organização. Por exemplo, um sistema legado pode ser fundido com uma solução em nuvem que o complementa, hospedada por um provedor de nuvem, com integração pela Internet.

## 1.2 Contribuição Esperada

Considerando essa tendência de aumento de uso da Computação em Nuvem no governo brasileiro, pode ser de benefício à APF a definição de diretrizes e métodos de migração de sistemas legados para a nuvem pública que considere as características do órgão de governo, as características do sistema a ser migrado e as características do provedor de nuvem pública que hospedará o sistema após a migração. Com o uso de Computação em Nuvem, as organizações de governo podem reduzir a quantidade de contratações e,



consequentemente, reduzir as oportunidades de irregularidades [14]. Podem, ainda, focar na prestação de serviços à população, a partir da redução do esforço operacional de manter a infraestrutura e a plataforma de serviços de TI.

O objetivo geral deste trabalho é propor um modelo de referência de migração de sistemas legados para a nuvem. O modelo incluirá também um Indicador de Percepção de Risco (IPR) da migração. A aplicabilidade do modelo será verificada com base em estudo de caso no domínio do governo. Os componentes fundamentais do modelo serão analisados por meio de pesquisa de opinião na APF.

Os objetivos específicos são:

1. Definir um modelo de referência de migração de sistemas legados para a nuvem cujo uso seja adequado no domínio do governo;
2. Propor um método de cálculo do Indicador de Percepção de Risco (IPR) da migração dos sistemas legados a partir das informações geradas pela aplicação do modelo de referência;
3. Realizar um estudo de caso no domínio do governo para mostrar a aplicabilidade do método proposto;
4. Analisar os componentes fundamentais do modelo por meio de uma pesquisa de opinião em órgãos da APF.

### **1.3 Estrutura do Trabalho**

Os próximos capítulos serão estruturados da seguinte forma: o Capítulo 2 contém a revisão bibliográfica e a base teórica da proposta apresentada. O Capítulo 3 apresenta a evolução dos modelos conceituais de migração para a nuvem e os critérios usados para a escolha de um modelo de referência. O Capítulo 4 contém o método de cálculo proposto para o indicador de percepção de risco, além do estudo de caso que mostra a aplicabilidade do método. O Capítulo 5 apresenta uma análise experimental, baseada em pesquisa de opinião, e o Capítulo 6 apresenta as conclusões e os trabalhos futuros.

# Capítulo 2

## Fundamentação teórica

Neste capítulo serão apresentados os conceitos e definições de Computação em Nuvem, sistemas legados, migração para a nuvem e modelos de referência. É o resultado da revisão da literatura associada ao tema.

### 2.1 Conceitos de computação em nuvem

Segundo o National Institute of Standards and Technology (NIST) [17], a Computação em Nuvem é um modelo que possibilita o acesso, conveniente e sob demanda, a um conjunto de recursos computacionais configuráveis, que podem ser adquiridos e liberados rapidamente com esforço gerencial mínimo e com pequena interação com o provedor de serviços.

Ainda segundo o NIST, o modelo de nuvem é composto por cinco características essenciais, três modelos de serviço e quatro modelos de implantação.

As características essenciais que compõem um ambiente de nuvem são:

**Autosserviço sob demanda** - Um consumidor pode provisionar recursos computacionais quando necessário sem a necessidade de interação humana.

**Ampla acesso a rede** - Os recursos são acessados por meio de mecanismos e protocolos padronizados.

**Compartilhamento de recursos** - Os recursos computacionais do provedor de nuvem são agrupados em um ambiente compartilhado e disponibilizados para os consumidores de acordo com a demanda.

**Rápida elasticidade** - Os recursos podem ser rapidamente provisionados e liberados para permitir aos consumidores aumentar ou diminuir o seu uso de acordo com a necessidade do momento.

**Serviço contabilizado** - Os serviços de nuvem automaticamente medem o uso dos recursos de um consumidor e permitem o monitoramento, o controle e a emissão de relatórios.

Os modelos de provisão de serviços são definidos de acordo com o nível de abstração computacional provido:

**Infrastructure as a Service (IaaS)** - Este modelo representa o provisionamento de recursos computacionais, incluindo processamento, armazenamento, rede e outros recursos.

**Platform as a Service (PaaS)** - Este modelo representa a entrega de uma plataforma de execução de *software* com ferramentas e recursos de desenvolvimento associados.

**Software as a Service (SaaS)** - Este modelo representa a entrega sob-demanda de um *software*, tipicamente uma aplicação *web*, por meio da Internet.

Os modelos de implantação são definidos de acordo com o público consumidor dos serviços:

**Nuvem privada** - A infraestrutura da nuvem é provisionada para uso exclusivo de uma única organização.

**Nuvem comunitária** - A infraestrutura da nuvem é provisionada para uso exclusivo de uma comunidade específica de consumidores de organizações que tem interesses compartilhados (por exemplo, órgãos de governo).

**Nuvem pública** - A infraestrutura da nuvem é provisionada para uso do público em geral.

**Nuvem híbrida** - A infraestrutura da nuvem é uma composição de duas ou mais infraestruturas de nuvem distintas (privada, comunitária ou pública) integradas.

## 2.2 Benefícios no uso da computação em nuvem

De acordo com o Gartner [9], uma das principais motivações para se mudar para a nuvem pública é a agilidade. A capacidade de se adaptar rapidamente a um ambiente em mudança é primordial no mundo digital e crítica para ganhar vantagem competitiva. A infraestrutura de TI local deve ser construída para a maior demanda prevista, e muito disso permanecerá ocioso ou subutilizado durante a operação normal. Os recursos de nuvem, por outro lado, podem ser dimensionados rapidamente para se adaptarem à demanda, eliminando grande parte do problema de superprovisionamento. Além disso, detalha o trabalho, porque grande parte da gestão da infraestrutura é agora gerenciada pelo provedor de serviços de nuvem, o departamento de TI pode colocar um foco maior em oferecer valor ao restante da organização em vez de dispendir esforços apenas para manter a infraestrutura em funcionamento.

### 2.2.1 Benefícios específicos para governo

Segundo o Tribunal de Contas da União [14], identificam-se as seguintes vantagens no uso de Computação em Nuvem, mais relacionadas a atividades estatais:

1. Maior agilidade da administração pública na entrega de serviços e em sua atualização tecnológica, pois os processos formais de contratação pública podem dificultar a manutenção de uma infraestrutura de TI própria atualizada e que responda rapidamente às demandas de seus usuários.
2. Suporte a iniciativas de Big Data e Dados Abertos, facilitando a abertura de informações governamentais que hoje encontram-se em sistemas que controlam as operações cotidianas do Estado e portanto são fechados, com acesso limitado aos seus operadores. O uso de nuvem pública permitiria ampliar o acesso a esses dados a um custo menor, sem comprometer a segurança, a disponibilidade e o desempenho operacional dos sistemas originais. Uma vez os dados governamentais estando facilmente acessíveis, torna-se possível maior participação da sociedade na criação de novos serviços baseados nesses dados.
3. Atendimento a picos de demanda sazonal de serviços públicos pela Internet sem necessidade de alocar grande quantidade de recursos fixos. Várias atividades estatais acarretam picos sazonais de demanda de serviços próximos a datas limite como: entregas de declarações de imposto de renda, inscrições e resultados do Enem, resultados eleitorais e listagem de gestores públicos inelegíveis, períodos de recadastramento do INSS, listagem dos percentuais do fundo de participação dos municípios, entre outros.
4. A contratação de serviços em nuvem de IaaS ou PaaS pode levar a uma redução de oportunidades de desvios e irregularidades, quando comparada às múltiplas contratações de máquinas, licenças de software, manutenção e suporte necessárias para a operação de *datacenter* próprio.
5. Agilidade e economia na entrega de serviços para instituições públicas com unidades descentralizadas, que podem ter serviços disponibilizados por meio de acesso à Internet, mais barato que as interconexões via redes privadas atualmente utilizadas.

Kundra [18] apresenta algumas vantagens na adoção da nuvem pelo governo norte-americano: melhoria da eficiência, melhor utilização dos ativos, redução da duplicação de recursos, redução do número de *datacenters*, aumento da produtividade, escalabilidade, agilidade, efetividade dos serviços e incentivo à cultura empreendedora.

## 2.3 Sistemas legados

Dedeke [19] define sistema legado como um pacote agregado de soluções de *hardware* e *software* cujas linguagens, padrões, códigos e tecnologias pertencem a uma geração

anterior de inovação.

Zhao and Zhou [20], de forma similar, definem um sistema legado como um sistema de computador desatualizado que permanece em uso mesmo depois que uma tecnologia mais moderna surge e que os principais motivos para essa permanência são: 1) a organização ter investido considerável volume de tempo e dinheiro e 2) o sistema legado conter dados valiosos.

Muitas organizações dependem fortemente dos sistemas legados, desenvolvidos ao longo de anos de operação. Como já foram extensamente testados em produção e preservam conhecimento acumulado do domínio do processo de negócio que suportam, reduzem o risco relativo ao atingimento de resultados organizacionais e, portanto, podem ser cruciais ao atingimento dos objetivos da organização. Mas por se basearem em tecnologia, método de desenvolvimento ou arquitetura ultrapassados, ou cuja adoção venha diminuindo consideravelmente ao longo do tempo, apresentam dificuldades de manutenção, como falta de pessoal qualificado, baixa interoperabilidade e necessidade de investimento adicional para adequá-los a novas tecnologias e plataformas vigentes [21].

Assim, apesar de haver problemas de manutenção, os sistemas legados ainda são necessários às organizações e a sua substituição é uma atividade complexa [3].

## 2.4 Migração para nuvem

A migração de *software* é o processo de mudança (ou adaptação) de um ambiente operacional para outro [22]. A migração de *software* pode ser vista como parte de um contexto mais amplo que é o da manutenção e reuso de *software* [21].

A migração para a Nuvem é o esforço de adaptar sistemas legados que executam localmente na infraestrutura da organização para transportá-los a um ambiente de nuvem [3].

Segundo Sun and Li [23], a migração para nuvem é um processo abrangente que envolve uma avaliação apropriada e precisa do suporte técnico para sua consecução, o que não é trivial e portanto precisa de um tempo adequado para ser executado.

Para aproveitar os benefícios da Computação em Nuvem e proteger o investimento existente em um sistema legado, as organizações tendem a migrar seus sistemas legados para a nuvem quando viável e de forma gradual [20] [9] [21].

Pahl and Xiong [16] alertam que o processo de migração requer análise, planejamento e execução com bastante cautela, para garantir segurança e integridade ao sistema depois de migrado e para que continue aderente aos requisitos da organização.

De acordo com Babar and Chauhan [24], uma migração para a nuvem mal sucedida pode causar falha nos processos de negócio, aumentar incidentes de segurança e aumentar custos de manutenção.

Além dessas considerações, há o questionamento acerca das estratégias e métodos de migração, complementado por Zhao and Zhou [20]. Para eles, diferentes estratégias de migração devem considerar diferentes processos de migração, com suas subdivisões em tarefas específicas.

## 2.5 Percepção de risco

Risco é definido pela norma ABNT NBR ISO 31000:2009 [25] como sendo o efeito da incerteza ou a medida probabilística de um evento impactar os objetivos de uma organização. Um risco pode ser negativo ou positivo, de acordo com o tipo de impacto que cause.

O PMBOK [6] descreve um grupo de processos de gerenciamento de riscos que tem por objetivo diminuir a probabilidade ou o impacto dos riscos negativos e aumentar a probabilidade ou o impacto dos riscos positivos, a fim de otimizar as chances de sucesso de um projeto.

Existem várias publicações que propuseram modelos de avaliação de risco com o objetivo de fornecer um método para quantificar os riscos com base na probabilidade e impacto e considerando também a gravidade dos ativos críticos. Eles têm sido frequentemente utilizados na adoção de Computação em Nuvem [26, 27]. Alguns deles estão relacionados à migração para a nuvem, mas apoiam a decisão sobre se um sistema específico deve ou não ir para a nuvem [28, 29].

Percepção de risco, diferentemente, refere-se aos julgamentos e avaliações das pessoas com relação aos perigos a que estas estão ou poderiam estar expostas, conforme descrito em Rohrman [30]. Tais percepções orientam decisões de aceitabilidade dos riscos.

Guimarães [31] destaca percepção de risco como sendo um “conjunto de fenômenos de natureza sociológica e psicológica que criam uma hierarquia de riscos subjetivos, particular a cada indivíduo e a cada grupo social”. A percepção de riscos coletiva (ou social) tende a avaliar os riscos subjetivos, superestimando a gravidade e subestimando as probabilidades. A percepção de risco individual, por outro lado, é influenciada pela avaliação em nível psicológico e nem sempre racional quanto aos benefícios e impactos.

Em um estudo baseado em uma revisão sistemática da literatura [32], os autores abordaram a percepção de risco com o objetivo de identificar e classificar os riscos relacionados à adoção de Computação em Nuvem. Eles distribuíram um questionário para especialistas em TI que estavam envolvidos na implementação de soluções de Computação em Nuvem

e coletaram suas percepções sobre os riscos que uma organização pode encontrar durante a adoção de Computação em Nuvem. Os respondentes indicaram o nível de importância de cada risco identificado de acordo com sua probabilidade, impacto e frequência de ocorrência. Com base nos dados coletados, os autores definiram uma fórmula para calcular a pontuação de importância para cada risco identificado e os classificaram com base na pontuação, criando uma lista chamada “10 principais riscos de Computação em Nuvem”. Eles descobriram que o potencial fracasso da adoção de Computação em Nuvem não pode ser simplesmente atribuído a riscos de privacidade e segurança, mas também é desencadeado por vários problemas operacionais, organizacionais e gerenciais relacionados a fornecedores e provedores de nuvem.

Em outra pesquisa realizada por Cicoria [33], o autor definiu um *Framework* dos Riscos de Migração para a Nuvem Pública, baseado em opiniões de 24 especialistas em Computação em Nuvem, que foram entrevistados e forneceram suas percepções sobre os riscos. Realizaram, ainda, uma pesquisa com 138 profissionais de TI e as percepções de risco coletadas na pesquisa foram confrontadas com as percepções dos especialistas. Um achado importante foi que, para alguns riscos, como os relacionados à migração de cargas de trabalho de baixa latência ou de alta taxa de transferência, uma maior experiência geral de TI, medida em anos, está diretamente relacionada com a maior conscientização sobre riscos.

Em um estudo metodológico que tratou da aplicação de questionários, foram investigadas as propriedades de vários formatos de resposta utilizados para estudar a percepção de risco. Verificou-se que as escalas de categorias com número limitado de alternativas, com 5 ou 7 por exemplo, fornecem dados mais eficientes para a discriminação entre os riscos [34].

Diferentemente dos estudos citados anteriormente e embora indiretamente relacionado à adoção do Computação em Nuvem, o Indicador de Percepção de Risco (IPR, definido na Seção 4.3) está focado na otimização do processo de migração, com objetivo de aumentar sua taxa de sucesso. O IPR atua dentro do processo de migração e não é um substituto dos processos de gerenciamento de risco que atuam em nível de projeto.

## 2.6 Modelos de referência

Segundo Fettke and Loos [35], um modelo de referência é um *framework* conceitual que pode ser usado como o modelo para desenvolvimento de sistemas de informação. Os modelos de referência também são chamados de modelos universais, modelos genéricos ou modelo padrão. Para usar modelos de referência, eles devem ser adaptados para os requisitos de um domínio específico.

Os autores citam que a efetividade e eficiência da aplicação de um modelo de referência são fortemente determinadas pela qualidade do modelo, mas que qualidade compreende diversos aspectos e pontos de vista. Assim, o objetivo de uma avaliação é determinar o valor e a utilidade de um modelo de referência. A avaliação de um modelo de referência valida e verifica sua adequação e as exigências reivindicadas pelo modelo.

O estudo apresenta diversas perspectivas de avaliação de modelos de referência. Entre elas, a perspectiva empírica, da qual citam-se duas abordagens que se relacionam com o objeto dessa pesquisa: o estudo de caso e a pesquisa de opinião.

O objetivo de um estudo de caso é investigar uma situação específica do uso do modelo de referência em uma organização em um ponto do tempo. O estudo de caso descreve o problema endereçado, a solução proposta e os resultados obtidos. As principais limitações do estudo de caso são uma menor objetividade e fraca generalização, principalmente se o caso investigado não for representativo. No entanto, esta abordagem de avaliação permite obter muitas informações úteis sobre a qualidade do modelo de referência, como por exemplo, informações sobre consistência e aplicabilidade de um novo modelo de referência.

O objetivo de uma pesquisa de opinião é examinar opiniões de pessoas em diferentes aspectos da construção e aplicação de modelos de referência por meio de questionários. As principais limitações do uso de pesquisas de opinião é que elas apresentam uma baixa taxa de resposta e analisam características subjetivas de qualidade e, em especial, novos modelos não são bem conhecidos na prática. Apesar disso, constituem uma boa ferramenta para avaliar a qualidade de um modelo de referência.

## **2.7 Conclusão deste capítulo**

Este capítulo apresentou os conceitos de CN definidos pelo NIST [17] e citou os benefícios do seu uso, com destaque para o aumento da agilidade no fornecimento de serviços aos clientes das organizações [9]. Em organizações de governo, um grande benefício citado pelo TCU [14] é o de melhorar o suporte a iniciativas de Big Data e de Dados Abertos, mas há referências também ao aumento da eficiência [18] e à diminuição das oportunidades de desvios e irregularidades [14]. Na sequência, conceituou-se sistema legado, migração para a nuvem e percepção de riscos. Por fim, o capítulo apresentou conceitos relacionados a modelos de referência, como o fato de eles necessitarem de adaptações para serem aplicados em um domínio específico e que podem ser validados empiricamente por meio de estudos de caso e por meio de pesquisas de opinião [35].



# Capítulo 3

## Migração de sistemas legados para a Nuvem

### 3.1 Estratégias e métodos de migração

Binz et al. [36] sugerem um *framework* de migração de aplicações para nuvem e entre provedores de nuvem. Definem também uma classificação das migrações em três classes: a Migração de Formato Padronizado (do Inglês *Standardized Format Migration*), Migração de Formato de Componente (do Inglês *Component Format Migration*) e Migração Holística (do Inglês *Holistic Migration*).

A Migração de Formato Padronizado é a migração de um componente autônomo, implementado em um formato padronizado. Os componentes são migrados entre instâncias de um mesmo *software* ou para outro *software* que suporte o formato. Há muitos formatos que são utilizados para facilitar as migrações dessa classe. Um exemplo é a migração de imagens de máquinas virtuais no formato OVF (do Inglês *Open Virtualization Format*).

Já na Migração de Formato de Componente, há transformação de formatos. Tem-se como exemplo a transformação de uma imagem de máquina virtual do formato nativo de um hipervisor – sistema que implementa a camada de virtualização de uma infraestrutura [37]– para outro.

A Migração Holística, que contempla os objetivos do *framework* definido em [36], é a migração completa da aplicação, através da migração de cada componente separadamente. Diferentemente da Migração de Formato de Componente, a Migração Holística decompõe a aplicação, removendo seus componentes e os remodelando para IaaS, PaaS e SaaS.

Binz et al. [36] identificaram três requerimentos para implementar a Migração Holística. O primeiro refere-se ao fato de que o *framework* deve ser independente de um modelo específico de aplicação. O segundo habilita a migração de componentes sem a necessi-

dade de acessar o código-fonte. O terceiro diz que a migração holística deve ser aberta e extensível para o reuso de transformações implementadas anteriormente no *framework*.

Outra visão relativa à migração para a nuvem é abordada por Andrikopoulos et al. [38], em que são identificados quatro tipos de migração que podem permitir a migração de aplicações através de adaptação. O primeiro tipo trata de substituir um ou mais componentes da aplicação pelos providos pela nuvem, sendo o menos invasivo. Como exemplo é citada a substituição de uma banco de dados MySQL [39] pelo Google Cloud SQL [40]. Outro tipo seria migrar parcialmente algumas das funcionalidades da aplicação ou um conjunto de componentes arquiteturais para a nuvem. O terceiro tipo seria realizar uma migração clássica em que todo *software* da aplicação é migrado para nuvem encapsulado em uma máquina virtual (VM, do Inglês *Virtual Machine*). O quarto tipo trata de realizar uma migração completa da aplicação para a Nuvem utilizando serviços componentizados do provedor.

Zhao and Zhou [20] analisaram e compararam as duas definições de migração já citadas [36] [38], além de outras três definições, e categorizaram a migração para nuvem em três grandes áreas: migração para IaaS, migração para PaaS e migração para SaaS. Esta última se subdivide em três: substituição por SaaS, revisão baseada em SaaS e reengenharia para SaaS.

Segundo os autores, a primeira categoria, migração para IaaS, é a mais indicada quando não há tempo para fazer reengenharia do sistema. A migração se dá com o transporte do sistema sem modificação para a infraestrutura de nuvem.

A segunda categoria, migração para PaaS, considera a reconstrução do sistema usando os componentes de PaaS do provedor de nuvem, podendo tornar o sistema em um sistema de nuvem nativo, com tratamento intrínseco de escalabilidade e elasticidade, por exemplo.

A primeira subdivisão da terceira categoria, a substituição por SaaS, trata apenas de exportar a base de dados local para a Nuvem e contratar o mesmo sistema no provedor de nuvem, no modelo SaaS. A revisão baseada em SaaS visa a substituir algumas funcionalidades do sistema por serviços oferecidos pelo provedor de nuvem. A reengenharia para SaaS significa remodelar o sistema para que seja decomposto em serviços e fique compatível com a arquitetura orientada a serviços (SOA, do inglês *Service Oriented Architecture*).

Antes do início do processo de migração é necessário classificar os sistemas quanto à viabilidade que têm de serem migrados para a nuvem. Os sistemas que não tiverem compatibilidade com o ambiente de Computação em Nuvem devem ser relacionados. Ou seja é necessário registrar quais são os sistemas em produção e em desenvolvimento para idealizar, preliminarmente, para qual modelo de serviço cada sistema pode ser migrado (IaaS, PaaS ou SaaS).

Um dos principais benefícios do uso de Computação em Nuvem é a redução de custos de investimentos em infraestrutura de TI [12] [13] [9]. No entanto, de acordo com Sun and Li [23], esse benefício normalmente considera apenas o custo dos serviços de nuvem após a migração. Os autores desenvolveram um método para estimar o custo do processo de migração dos sistemas legados para a Nuvem que é uma adaptação do método de Putnam [41] no contexto de Computação em Nuvem. O método considera que, embora descrito sequencialmente, o fluxo de execução das tarefas de migração tem um nível de iteratividade variável em que erros na execução de uma tarefa pode causar a repetição de tarefas anteriores. Assim, de acordo com o tamanho e complexidade do sistema legado e também com a maturidade da equipe responsável pela migração, o custo do processo de migração pode ser calculado com o uso do método que definiram.

Na revisão bibliográfica, foram encontradas muitas publicações de processos que estruturam a migração de sistemas legados para a Nuvem, mas que consideraram características específicas de um único domínio de aplicações, o que torna mais complexa a adaptação para o uso no domínio do governo [42][43]. Por outro lado, há alguns modelos conceituais que buscaram a generalização a partir dos casos específicos, de forma a se adequarem a múltiplos domínios. Serão analisados e comparados três destes modelos conceituais para identificar qual deles mais se adequa às migrações de sistemas legados do governo.

## 3.2 Modelo de caracterização

Passa-se a analisar inicialmente o modelo Cloud-RMM, exposto na figura 3.1.

Nesse modelo, há apenas uma consolidação de processos, tarefas e artefatos, encontrados nos vinte e três estudos analisados na revisão sistemática empreendida por Jamshidi et al. [3]. Não há no modelo um detalhamento ou mesmo um embasamento da necessidade e importância de cada processo e tarefa, mas apenas a citação de cada um deles pelo fato de estarem definidos nos modelos analisados.

O Cloud-RMM categoriza vinte e três estudos sobre a migração de sistemas legados para a nuvem e serve como guia da análise destes estudos. A Figura 3.1 mostra a estrutura do Cloud-RMM, composto por quatro processos, cada um representado por um grupo de tarefas. Cada processo referencia a quantidade de estudos que o contém e cada tarefa exibe os estudos que a citam. Vê-se, na figura 3.1, a notação  $Sn$ , em que o  $n$  é o número que determinado estudo (o S vem da inicial em inglês, *Study*) recebeu na lista disponível no artigo. No modelo, há ainda a indicação dos artefatos gerados ao fim de cada processo.

O Cloud-RMM é formado por três processos sequenciais que são: planejamento da migração, execução da migração e avaliação da migração, além de um processo guarda-chuva, que trata de preocupações transversais.

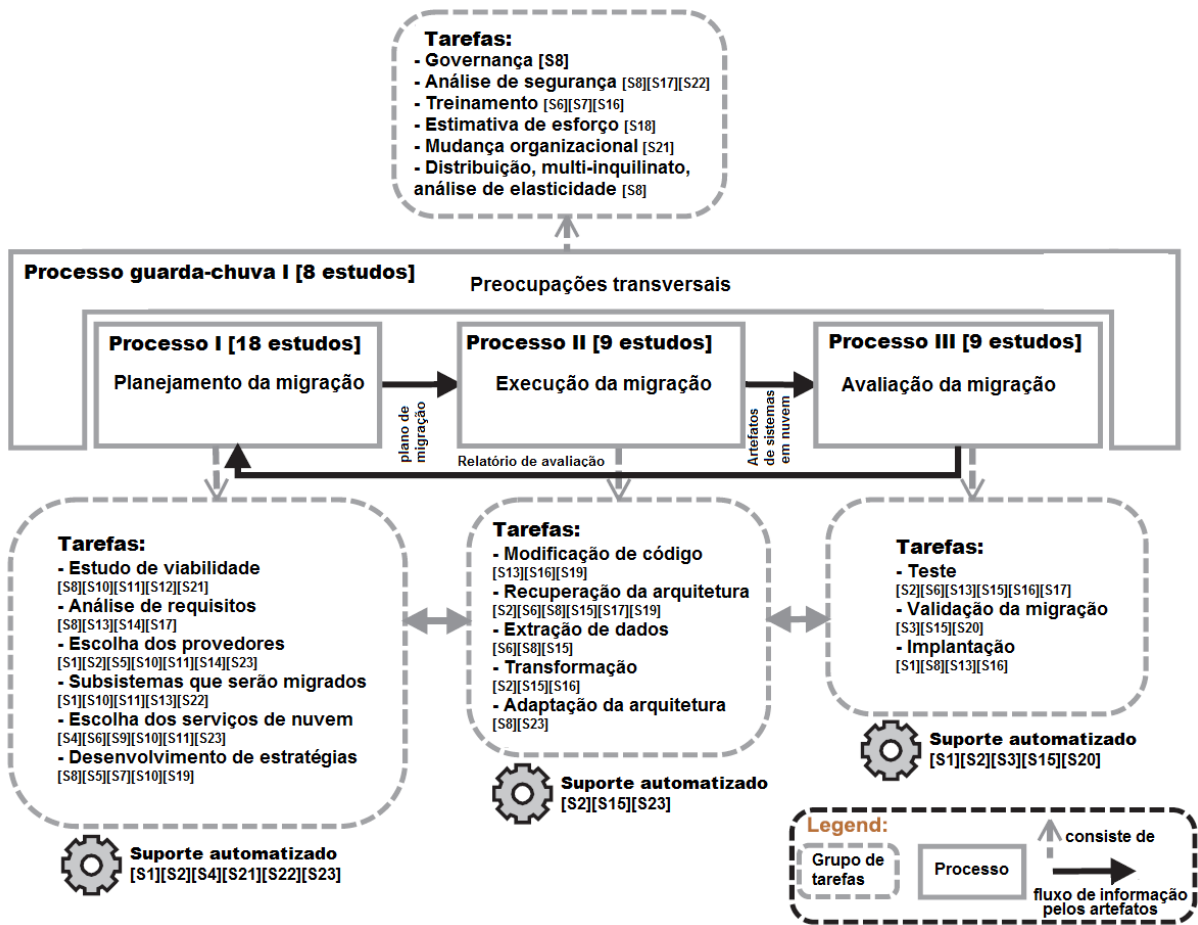


Figura 3.1: Cloud-RMM - modelo de caracterização proposto por Jamshidi et al. [3]

No processo de planejamento da migração foram encontradas, nos estudos analisados, as tarefas relacionadas a estudo de viabilidade, análise de requisitos de migração, suporte para a tomada de decisões a respeito da escolha do provedor, subsistemas a serem migrados e o desenvolvimento das estratégias da migração.

No segundo processo, estão as tarefas relativas à modificação dos sistemas e à migração propriamente dita. O modelo cita as seguintes tarefas: modificação de código, recuperação da arquitetura, extração dos dados, transformação e adaptação da arquitetura para a nuvem.

A avaliação da migração é o terceiro processo e registra as seguintes tarefas: teste, validação da migração e implantação das aplicações e sistemas migrados.

O processo que trata das questões transversais registra as seguintes tarefas: governança, análise de segurança, treinamento, estimativa de esforço, mudanças organizacionais e análises da distribuição dos componentes do sistema, da adaptação para múltiplos inquilinos (*multi-tenancy*) e da elasticidade.

Jamshidi et al. [3] apontaram que a pesquisa em migração para a nuvem estava ainda

em estágio inicial de maturidade e que um *framework* de migração aumentaria a maturidade e a confiança nos modelos de migração para a nuvem. Os autores identificaram que a pesquisa com foco em validação dos modelos existentes ainda precisava de atenção da comunidade científica e que eram necessários mais resultados de pesquisa baseados em estudos de caso e em relatórios de experiências reais.

### 3.3 Modelo de avaliação

Gholami et al. [4] fazem uma revisão sistemática da literatura com o intuito de fornecer uma revisão detalhada das abordagens de migração de nuvem existentes partindo da perspectiva da modelagem de processo e baseando-se fortemente nas metodologias de desenvolvimento de *software*. Para esse fim, propõem um modelo de avaliação para comparar as abordagens existentes, destacando seus recursos, semelhanças e diferenças fundamentais. A abordagem utilizada pelos autores difere dos outros trabalhos relacionados, pois foca no aspecto do processo da migração da nuvem para entender quais atividades e preocupações essenciais estão envolvidas durante essa transição.

Segundo esse artigo, nenhum dos estudos analisados fornece uma discussão aprofundada sobre as características e as atividades de migração propostas nas abordagens existentes, bem como a experiência útil de aplicar essas abordagens na prática. Além disso, o artigo fornece uma análise meticulosa das abordagens existentes através de um modelo de avaliação (ver Figura 3.2), que engloba vinte e oito critérios classificados em duas dimensões. Ou seja, onze critérios genéricos e dezessete critérios específicos de computação em nuvem. O quadro proposto foi derivado de uma extensa revisão da literatura e validado através de uma pesquisa de questionários na Web com 104 especialistas da academia e especialistas no campo da Computação em Nuvem (Gholami et al. [4]).

A Tabela 3.1 mostra resultados encontrados com o uso desse modelo de avaliação proposto quando confrontado com as informações presentes nos 43 artigos analisados. Foram selecionados alguns resultados que se relacionam mais fortemente com os objetivos de pesquisa desta dissertação.

O modelo de avaliação da Figura 3.2 foi utilizado pelos autores para classificar os 43 artigos em relação ao nível de atendimento de cada um dos 28 critérios. Para a maioria dos critérios, utilizou-se uma escala de 3 níveis de atendimento: atende completamente, atende parcialmente e não atende. Alguns dos critérios classificaram os artigos com respostas sim ou não e, em menor quantidade, há critérios que apenas coletam uma informação descritiva que caracteriza o artigo.

O resultado final da avaliação, defendem os autores, deve ser visto como um indicador de qual era o foco principal do artigo em vez de uma avaliação quanto à sua completude

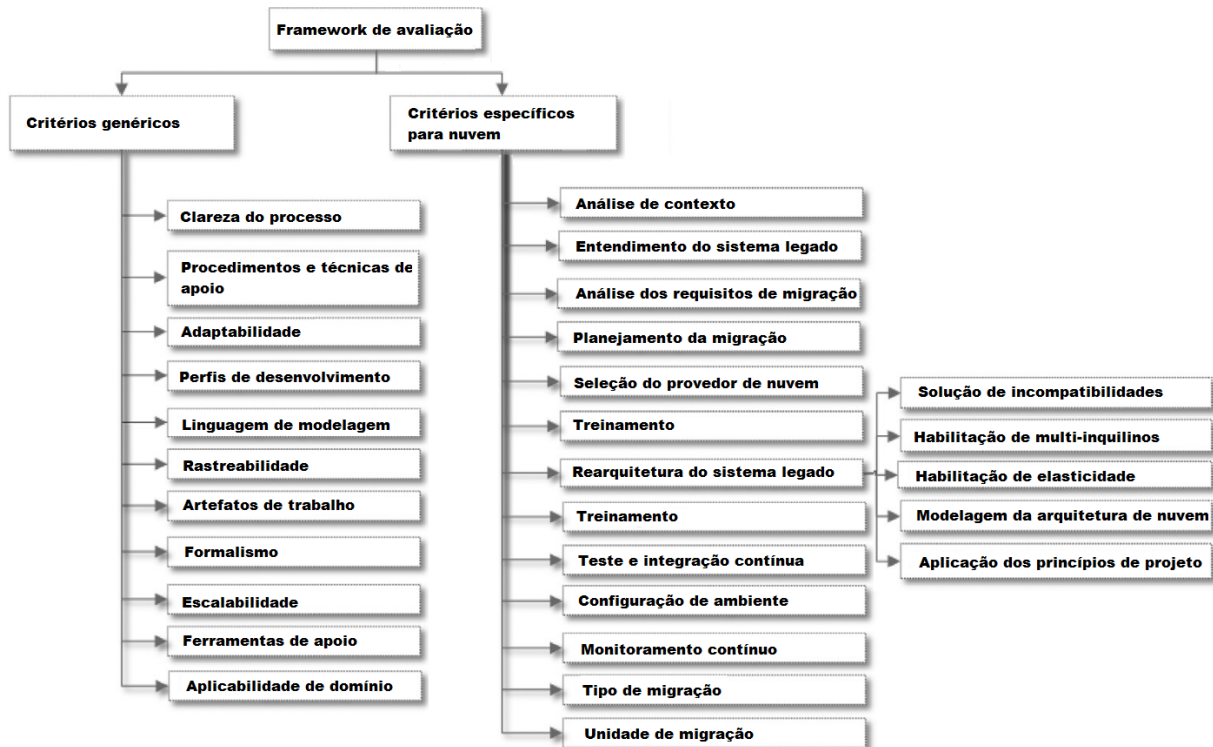


Figura 3.2: Modelo de avaliação proposto por Gholami et al. [4].

em comparação com os demais. Em outras palavras, os autores afirmam que se um artigo específico não atende a um determinado critério, isso não significa necessariamente que ele é inferior a outros artigos que atendam àquele critério. Em vez disso, significa que o critério não é o foco principal do artigo.

O modelo de caracterização de Jamshidi et al. [3] não inclui nenhum critério genérico oferecido pelo modelo de avaliação proposto em Gholami et al. [4]. Para a dimensão específica da nuvem, embora 11 dos 17 critérios tenham sido referenciados por Jamshidi et al. [3] em seu modelo, não houve investigação sobre eles como forma de avaliação de cada estudo revisado. Ainda, o estudo de Gholami et al. [4] cobre mais estudos (43 contra 23, com apenas 5 de interseção) e alcança estudos publicados entre os anos 2009 e 2015 (em comparação, o período analisado por Jamshidi et al. [3] foi de 2010 a 2013). Assim, o estudo de Gholami et al. [4] serve como ponto de referência mais recente na análise de estudos de migração de sistemas legados para a nuvem.

Como desafios para trabalhos futuros, Gholami et al. [4] reconhecem que há um grande volume de pesquisa em migração para a Nuvem, que atualmente está dispersa e fragmentada, e sugerem ser necessária a definição de um modelo de referência genérico (ver Seção 2.6) com o objetivo de consolidar a literatura existente. Afirmam que como a cada ano um número considerável de documentos de pesquisa é publicado no campo da Computação em Nuvem, e como cada um reporta diferentes soluções, usa relatórios experimentais

Tabela 3.1: Alguns resultados selecionados da avaliação feita por Gholami et al. [4].

Critério	Descrição	Resultado
<b>Clareza do processo</b>	As abordagens diferem em esclarecer e especificar detalhes das fases e atividades. Alguns oferecem um bom nível de detalhes enquanto outros, deliberadamente ou não, fornecem apenas breve descrição das atividades e deixam espaço para interpretação de como as atividades podem ser realizadas.	O único estudo que foi considerado não atendendo a esse critério foi o de Jamshidi et al. [3], pois não forneceu nenhum detalhe para as vinte atividades (rotuladas de tarefas) que compõem o modelo conceitual proposto
<b>Procedimentos e técnicas de suporte</b>	É desejável que a abordagem elabore suas atividades prescritas, definindo técnicas de suporte e exemplos, de forma que os desenvolvedores simplesmente possam entender e aplicar a abordagem.	Quatorze abordagens fornecem diretrizes para algumas atividades, mas não para todo o processo. Verificou-se que seis abordagens não oferecem nenhum detalhamento sobre a realização de atividades.
<b>Adaptação</b>	Como qualquer projeto de desenvolvimento de <i>software</i> , uma metodologia deve ser ajustada às características e contingências de um determinado cenário de migração da nuvem em mãos. Portanto, o pensamento de que uma metodologia universal de migração de nuvem pode atender a todos os cenários de migração é considerado falso.	Apenas três abordagens consideram a possibilidade de adaptação da metodologia proposta. A grande maioria considera que sua abordagem servirá a qualquer cenário de migração.

específicos e faz recomendações para mover ativos legados para ambientes em nuvem, isso em si é uma evidência de que o campo atingiu um ponto de maturidade em que o desenvolvimento de um modelo de referência genérico se faz obrigatório.

### 3.4 Modelo de referência

Em outro estudo mais recente, Gholami et al. [5] atenderam ao desafio proposto em [4] e trabalharam com o objetivo de descobrir as atividades críticas, os artefatos, as preocupações e as principais recomendações concernentes à migração de sistemas legados para a nuvem e os validaram empiricamente, com a coleta das percepções de especialistas da área para aumentar a confiabilidade dos resultados.

À luz da revisão de literatura que empreenderam com a análise de 78 estudos, os autores utilizaram uma abordagem de métodos mistos [44] na tentativa de fornecer uma compreensão exaustiva da migração de sistemas legados para a nuvem e identificar os principais elementos do processo envolvidos nesta transição.

Com base em uma análise qualitativa intensiva de construtos existentes na literatura, os autores criaram um modelo de referência do processo de migração para a nuvem. A relevância e solidez do modelo foi na sequência confirmada usando uma pesquisa quantitativa e *feedback* qualitativo de especialistas de domínio.

O trabalho dividiu-se em duas fases, cada uma empregando métodos qualitativos e quantitativos em diferentes formas. Na primeira fase, derivou-se um modelo de referência do processo de migração que mostra os elementos-chave relacionados. Na segunda fase, foram usados os resultados da pesquisa com especialistas em Computação em Nuvem para ilustrar como os vários itens identificados no modelo se comportam em cenários de

migração de nuvem do mundo real e, portanto, fornecem uma descrição mais aprofundada desses elementos.

Para a análise qualitativa na primeira fase, os autores utilizaram técnicas comuns para modelos de referência, incluindo a extração e seleção de um conjunto de elementos candidatos das fontes existentes, a reconciliação das suas definições, organização dos elementos em fases e classificação desses elementos com base nos tipos de migração, conforme descritos na Seção 3.1, e criaram um modelo de referência. Isso também está alinhado com a recomendação de Schwarz et al. [45], que afirma que um modelo de referência sintetiza pesquisas anteriores de uma forma que torna possível aos interessados usá-lo na prática.

A Figura 3.3 mostra um resumo do modelo de referência resultante para o processo de migração de sistemas legados para a Nuvem. Para simplificar a visualização, o resumo apresenta apenas os elementos-chave do modelo, sem as subdivisões das atividades em tarefas menores e sem o fluxo de informação entre elas.

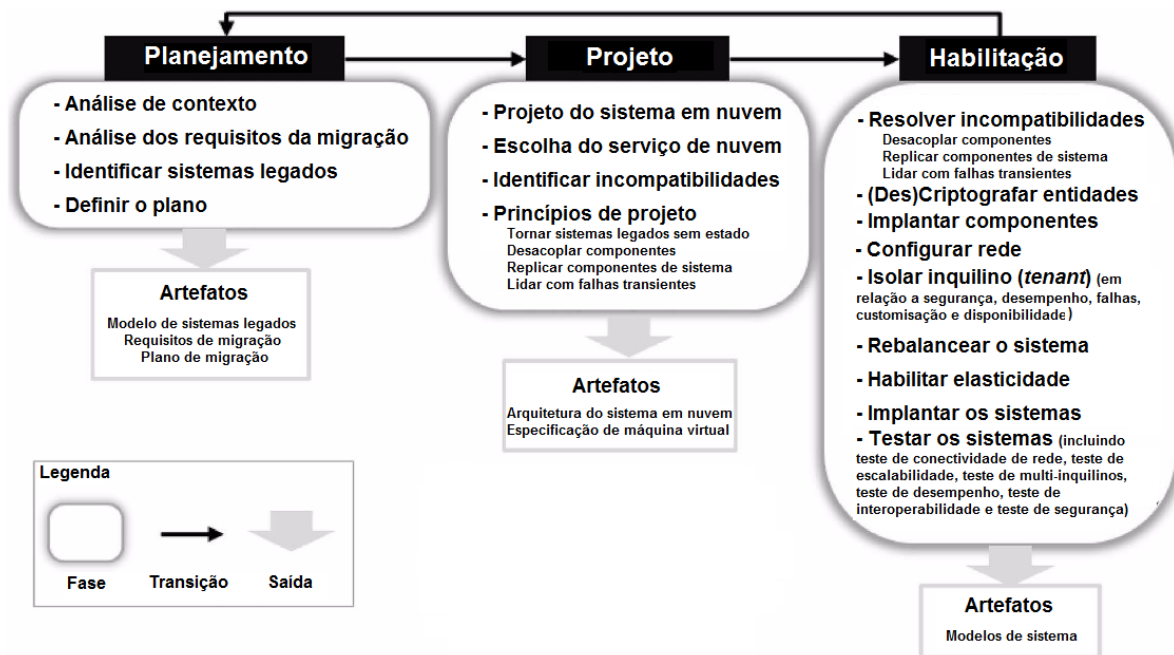


Figura 3.3: Elementos-chave do Modelo de Referência de migração de sistemas legados para a nuvem proposto por Gholami et al. [5].

O objetivo da segunda fase foi examinar se os elementos do modelo de referência, derivados na fase anterior, eram percebidos como relevantes na reengenharia de sistemas legados para plataformas em nuvem. Para a segunda fase, foi realizado um teste T de *Student* para confirmar a solidez dos elementos-chave no modelo. Além disso, os comentários qualitativos complementares fornecidos por especialistas em domínio foram utilizados para mostrar por que os elementos do modelo devem ser considerados em um cenário real de migração de sistemas legados para a nuvem.



O estudo apresenta uma explicação mais detalhada de todos os elementos-chave do modelo de referência. Apresenta também cada uma das três fases que o compõem, mostrando o relacionamento e o fluxo de informação entre as atividades e os artefatos que devem ser produzidos em cada fase, conforme exibem as Figuras 3.4, 3.5 e 3.6.

Esta pesquisa objetivava, inicialmente, definir um modelo de migração de sistemas legados para a Nuvem, específico para o domínio do governo, uma vez que não foi encontrado tal modelo na literatura. Por outro lado, com a publicação recente do Modelo de Referência de Gholami et al. [5], decidiu-se por adotá-lo, pois foi construído de forma robusta e com o propósito de ser utilizado em diversos domínios, embora ainda não tivesse sido validado no domínio do governo nas formas propostas por Fettke and Loos [35].

Assim, o modelo de referência foi selecionado para servir de base às migrações de sistemas legados do governo para a Nuvem. Ele foi escolhido por ter sido construído com o intuito de atender a diversos domínios de aplicação, ter sido validado por especialistas na área e por ter características que facilitam o seu uso, como as orientações acerca dos objetivos e detalhamento de cada atividade, sobressaindo-se aos outros modelos conceituais avaliados.

Pelo fato de haver, nessa pesquisa, objetivos específicos (Seção 1.2) direcionados a validar, no domínio do governo, o modelo de referência definido, entende-se que será minimizada a limitação da falta de utilização do modelo naquele domínio.

### **3.5 Conclusão deste capítulo**

Este capítulo fez uma revisão bibliográfica de temas relacionados a migração de sistemas legados para a Nuvem. Inicia com algumas definições e classificações de migração para a nuvem e, em seguida, apresenta, analisa e compara entre si, três modelos conceituais de migração: um modelo de caracterização, que agrupou processos, atividades e artefatos referenciados em 23 propostas de migração publicadas entre os anos de 2010 e 2013; um modelo de avaliação que, baseado em 28 critérios relevantes, avaliou 43 artigos, publicados entre 2009 e 2015; e um modelo de referência, definido a partir da consolidação de 78 propostas publicadas entre 2008 e 2015. O modelo de referência foi o escolhido por ter sido construído com o intuito de atender a diversos domínios de aplicação, ter sido validado por especialistas na área e por ter características que facilitam o seu uso, como as orientações acerca dos objetivos e detalhamento de cada atividade.

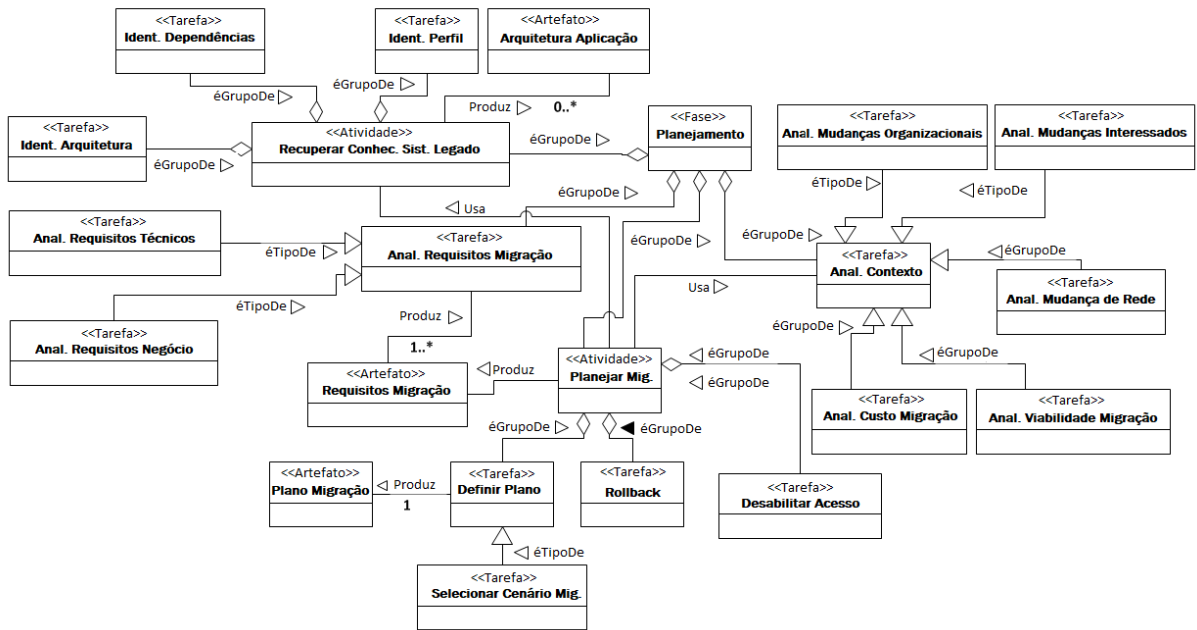


Figura 3.4: Fase de Planejamento do modelo de referência proposto por Gholami et al. [5]

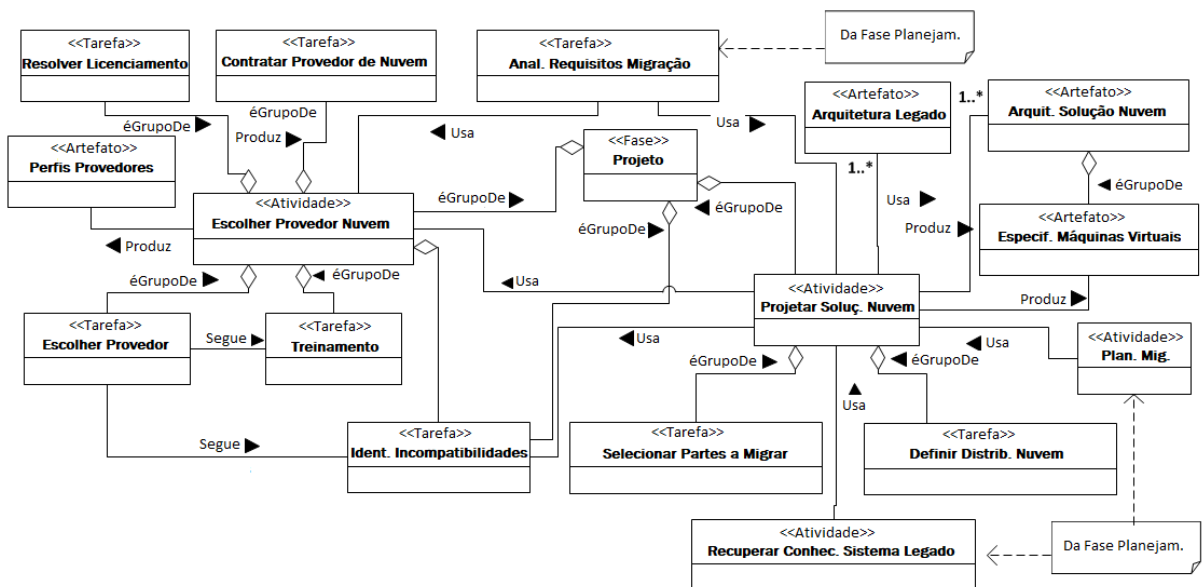


Figura 3.5: Fase de Projeto do modelo de referência proposto por Gholami et al. [5]

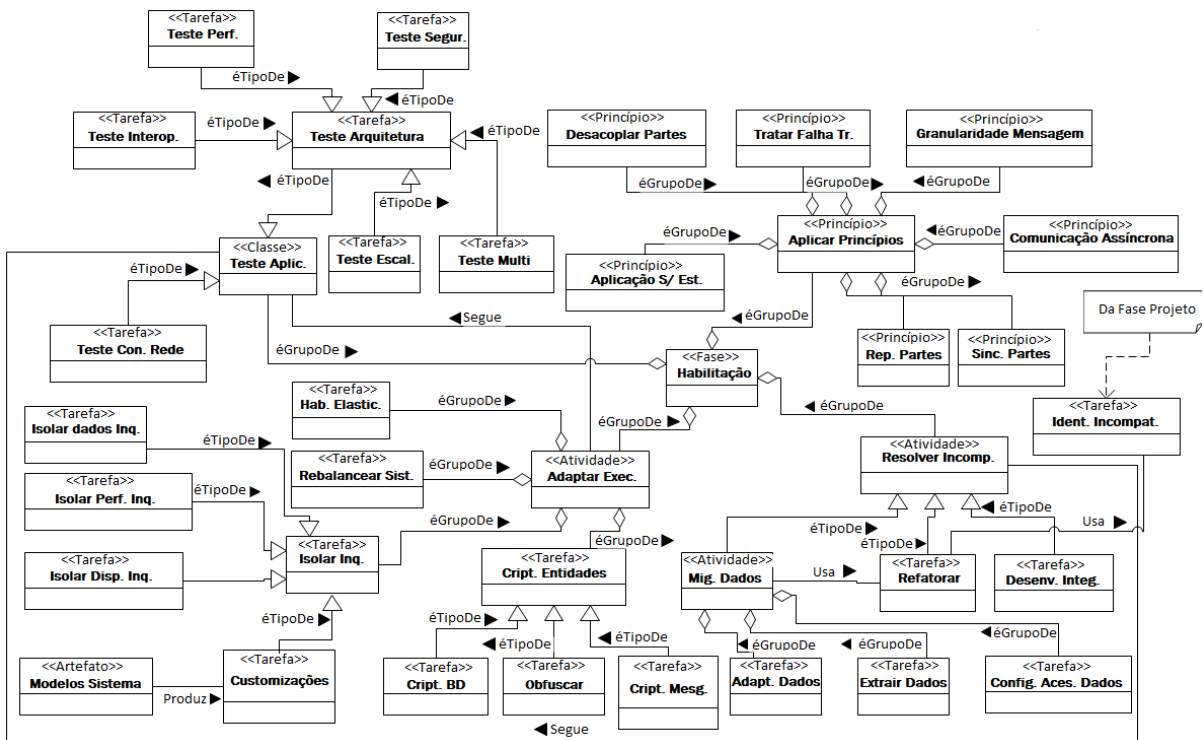


Figura 3.6: Fase de Habilitação do modelo de referência proposto por Gholami et al. [5]

# Capítulo 4

## Proposta

Neste capítulo será apresentado o método de cálculo do Indicador de Percepção de Risco do Modelo de Referência. Para tal, foi definida uma função de avaliação, que teve como base um conjunto de premissas que serão discutidas no capítulo. Uma prova de conceito será apresentada para mostrar a aplicabilidade do método e um estudo de caso, com três sistemas legados do governo, exemplificará o uso do indicador.

### 4.1 Migração de sistema legado como um projeto

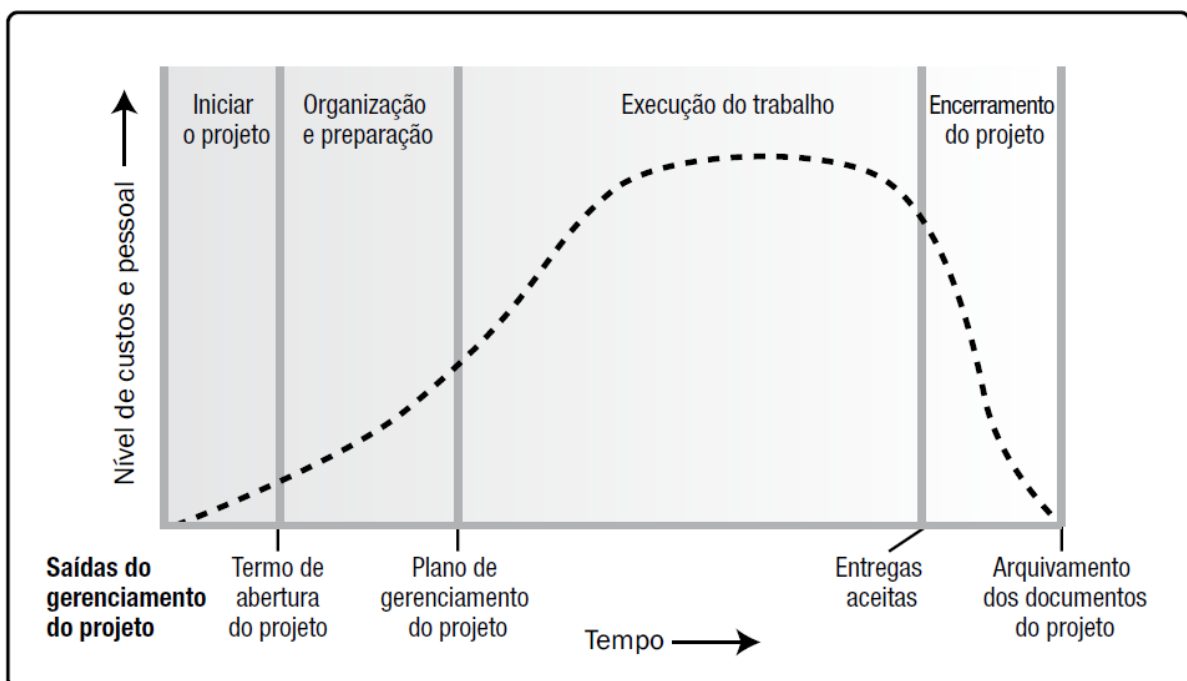


Figura 4.1: Ciclo de vida de projetos, definido no PMBOK [6].

Cada migração de sistema legado para a nuvem pode ser tratada como um projeto diferente, conforme conceitos descritos no Project Management Book of Knowledge (PMBOK), na medida em que tem início e fim definidos e entrega um produto único: o sistema legado migrado para a nuvem [6].

O PMBOK define ciclo de vida de um projeto como sendo a série de fases, geralmente sequenciais, por quais um projeto passa. A figura 4.1 mostra a estrutura genérica de um ciclo de vida de projeto e o nível de custos e pessoal necessário em cada fase do ciclo de vida.

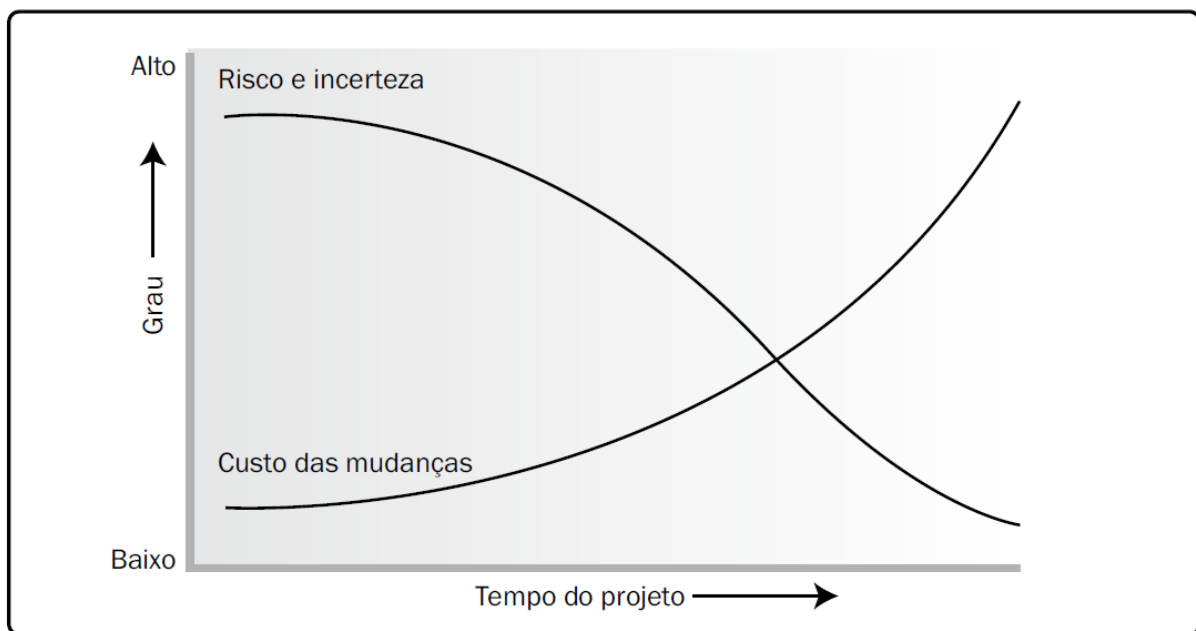


Figura 4.2: Riscos e custos de projetos em relação ao tempo - PMBOK [6].

Segundo o PMBOK, a estrutura genérica do ciclo de vida apresenta, entre outras, as seguintes características:

- Os níveis de custo e de pessoal são baixos no início, atingem um valor máximo enquanto o projeto é executado e caem rapidamente conforme o projeto é finalizado. A Figura 4.1 ilustra este padrão típico.
- Os riscos e incertezas são maiores no início do projeto. Esses fatores diminuem ao longo da vida do projeto à medida que as decisões são tomadas e as entregas são aceitas, conforme mostra a Figura 4.2.
- A capacidade de influenciar as características finais do produto do projeto, sem impacto significativo sobre os custos, é mais alta no início do projeto e diminui à medida que o projeto progride para o seu término. A Figura 4.2 ilustra a ideia de que

os custos das mudanças e correções de erros geralmente aumentam significativamente à medida que o projeto se aproxima do término.

Embora essas características continuem presentes até certo ponto nos ciclos de vidas de quase todos os projetos, elas não estão sempre presentes no mesmo grau.

## 4.2 Avaliação do modelo de referência

O Modelo de Referência divide-se em três fases: Planejamento, Projeto e Habilitação. A fase de Planejamento é responsável por coletar informações relevantes acerca do sistema a ser migrado (informações técnicas, contexto organizacional) e, em acordo com os requisitos definidos para a migração, criar um plano de migração apropriado. A fase de Projeto utiliza o conhecimento e artefatos gerados na fase de Planejamento e é responsável por escolher um ou mais provedores de nuvem e por definir a nova arquitetura que o sistema legado terá em cada provedor, quando houver mais de um, em termos de serviços na nuvem e da localização dos mesmos. A fase de Habilitação representa a migração propriamente dita. É a implementação das adaptações necessárias ao código do sistema legado, da construção de integradores quando for o caso, da configuração em cada provedor dos serviços de Computação em Nuvem definidos na fase de Projeto, dos testes e da entrada em produção do sistema migrado [5].

Conforme mostrado na Figura 4.1, o esforço necessário à execução da terceira fase é potencialmente maior que o que vai ser empregado nas duas fases iniciais. Isso se deve ao fato de as primeiras fases terem trabalho majoritariamente analítico, enquanto a terceira, de trabalho majoritariamente de implementação. Implementação, nesse contexto, significa tanto o desenvolvimento relacionado às adaptações e novas integrações, quando necessário, quanto a criação e configuração dos serviços de nuvem.

Assim, é importante ter-se uma avaliação precoce da execução da migração antes de ela chegar à fase de Habilitação. A avaliação pode indicar um risco aumentado de problemas (custos e prazos adicionais, por exemplo) para um determinado sistema, quando comparado com os demais, e assim fornecer uma medida objetiva de comparação. Pode-se ranquear os sistemas que serão migrados, por exemplo, e montar uma ordem de execução da migração dos sistemas legados que privilegie os sistemas com risco menor, aumentando assim a taxa de sucesso e a confiança da organização no projeto de migração, de forma a ganhar experiência com os provedores contratados, principalmente em relação aos primeiros sistemas a serem migrados. Além disso, pode-se rever os sistemas em que o resultado da avaliação foi mais negativo para, antes de se iniciar a fase de Habilitação, revisar as atividades e tarefas e entender se há como modificá-las de forma a melhorar a avaliação e conseqüentemente o risco de ocorrência de problemas na execução da migração.

## 4.3 Método

Como forma de prover uma avaliação objetiva e precoce acerca da migração de cada sistema legado para a nuvem, com o uso do Modelo de Referência, propõe-se adotar um Indicador de Percepção de Risco (IPR) com o método definido a seguir.

### 4.3.1 Passo 1

Em cada uma das duas fases iniciais do processo, Planejamento e Projeto, seleciona-se as tarefas que, em conjunto, melhor representam os pontos críticos de cada migração de sistema legado para a Nuvem. Algumas tarefas têm um trabalho meramente de coleta e agregação de informações, enquanto outras podem trazer restrições, decisões ou características intrínsecas ao sistema e à organização e que, com isso, têm o condão de impactar o restante da migração. Deve-se escolher entre as tarefas da segunda categoria. É importante perceber que diferentes domínios de aplicação podem constituir diferentes conjuntos de tarefas que melhor avaliam a migração de sistemas legados para aquele domínio.

### 4.3.2 Passo 2

Após a escolha do conjunto de tarefas de avaliação, é preciso definir os pesos que cada tarefa terá em relação às demais tarefas do conjunto. O peso da tarefa está relacionado com a importância relativa daquela tarefa, no domínio da aplicação, quando comparado com as demais tarefas. Por exemplo, no domínio do governo, a tarefa relacionada à aquisição é percebida como sendo mais arriscada do que a relacionada ao treinamento, devido a regras legais e de regulamentação.

### 4.3.3 Passo 3

Após escolhido o conjunto de tarefas de avaliação em cada fase e o peso de cada uma delas, é preciso avaliá-las. Note que o trabalho necessário em cada tarefa vai ser executado conforme definido no Modelo de Referência. Ao fim da execução de cada fase, e consequentemente da execução de todas as tarefas daquela fase, é que se executa o método de avaliação. Essa é uma consideração importante, pois é preciso conhecer em um nível de detalhes apropriado cada sistema para que a avaliação seja mais precisa e não se consegue alcançar o nível necessário executando isoladamente apenas as atividades que farão parte da função de avaliação.

Ao fim de cada fase, é necessário avaliar cada tarefa que faz parte do conjunto de avaliação, de acordo com a escala definida. A proposta é que seja utilizada uma escala *Likert*, um tipo de escala que se propõe a medir opiniões e atitudes das pessoas [46].

Propõe-se que se use uma escala de 5 valores, cujo significado pode ser: 1 - Risco muito alto, 2 - Risco alto, 3 - Neutro, 4 - Risco baixo e 5 - Risco muito baixo. Para cada tarefa, os 5 níveis da escala também podem ser definidos textualmente, de forma a facilitar a escolha e avaliação referente a cada sistema. Isso vai ser feito uma única vez e apenas para as tarefas que fazem parte do conjunto de avaliação. Uma vez a escala definida, assim que a Fase de Planejamento terminar, pode-se dar início à avaliação da mesma, escolhendo, para cada tarefa do conjunto de avaliação, o valor da escala que melhor caracterize aquele sistema e calcular o indicador de avaliação da Fase de Planejamento, o  $Pl$ , conforme Equação (4.2). O mesmo deve ser feito para a Fase de Projeto, calculando-se o Indicador de Avaliação para a Fase de Projeto, o  $Pr$ , conforme Equação (4.3). Com os dois indicadores calculados, pode-se calcular o Indicador de Percepção de Risco (IPR) da migração do sistema legado para a nuvem, conforme a Equação (4.1).

Portanto, o cálculo do IPR é realizado conforme a Equação (4.1) e os termos  $Pl$  e  $Pr$  são descritos na Equações (4.2) e (4.3).

$$IPR(Pl, Pr) = \frac{Pl + Pr}{2} \quad (4.1)$$

$$Pl = \frac{\sum_{i=1}^n (V_{Pl_i} \times P_{Pl_i})}{\sum_{i=1}^n (P_{Pl_i})} \quad (4.2)$$

e

$$Pr = \frac{\sum_{j=1}^n (V_{Pr_j} \times P_{Pr_j})}{\sum_{j=1}^n (P_{Pr_j})} \quad (4.3)$$

em que:

$Pl$  – Indicador de Avaliação da Fase de Planejamento.

$Pr$  – Indicador de Avaliação da Fase de Projeto.

$V_{Pl_i}$  – Valor de avaliação para a tarefa  $i$  do conjunto de avaliação da Fase de Planejamento.

$P_{Pl_i}$  – Peso da tarefa  $i$  no conjunto de avaliação da Fase de Planejamento.

$V_{Pr_j}$  – Valor de avaliação para a tarefa  $j$  do conjunto de avaliação da Fase de Projeto.

$P_{Pr_j}$  – Peso da tarefa  $j$  no conjunto de avaliação da Fase de Projeto.

A função definida em (4.1) considera que ambas as fases têm a mesma importância relativa no cálculo da percepção dos riscos de migração. Mas circunstâncias específicas



podem ser consideradas em outra função de avaliação que equilibre as fases de forma diferente.

#### 4.3.4 Passo 4

Com o IPR calculado para cada sistema que será migrado, pode-se estabelecer um ranking que indique, entre outras coisas, quais sistemas devem ser migrados primeiro, que serão os sistemas com os melhores valores de percepção de risco. Esta estratégia é reforçada por [47], que afirma que é melhor iniciar com um projeto piloto ao executar o processo de migração para a nuvem. Os sistemas com os piores valores de percepção de risco podem receber análise adicional para identificar se são possíveis e viáveis modificações no trabalho das tarefas de avaliação que causaram avaliação negativa. Ainda que não seja possível modificar algo nos sistemas com piores percepções de risco, saber precocemente do risco aumentado de problemas tem alto valor, na medida em que torna possível a gestão de expectativas do restante da organização e, inclusive, torna possível uma decisão de não ir para a nuvem, tudo isso antes que um maior investimento seja feito na migração do sistema.

Além de usar valores do IPR para ranquear os sistemas, é possível usá-los também para enquadrar os sistemas em uma escala de risco definida previamente. Para isso, podem ser definidos limiares de reavaliação. Por exemplo, pode-se definir que valores do IPR que estejam abaixo ou sejam iguais a um determinado limiar L1 significam que o sistema deve ter sua migração para a nuvem reavaliada, ou pelo menos postergada, e que os interessados devem ser comunicados dessa decisão. Um outro exemplo é o do sistema cujo valor do IPR ficou acima do limiar L1, mas cujo valor de um dos indicadores de fase ( $Pl$  ou  $Pr$ ) ficou abaixo de um limiar L2.

Os valores dos limiares podem ser definidos numericamente, a partir de um determinado valor da escala *Likert* utilizada, mas também podem ser definidos de acordo com o histórico dos valores do IPR calculados e considerando o resultado dessas migrações. Enquanto não há um histórico de valores de IPR calculados, sugere-se utilizar o valor central da escala para os limiares L1 e L2.

## 4.4 Prova de conceito

A prova de conceito que será descrita nesse trabalho seleciona as atividades que melhor constituem uma avaliação da percepção de risco para sistemas legados do domínio de governo, foco dessa pesquisa.

#### 4.4.1 Fase de planejamento

Para a fase de Planejamento foram selecionadas as tarefas e definidos os pesos que constam na Tabela 4.1. Os pesos foram definidos empiricamente, com base no conhecimento tácito e na experiência dos pesquisadores.

Tabela 4.1: Grupo de tarefas de avaliação da Fase de Planejamento.

Tarefa	Peso
Analisar custo da migração	3
Identificar dependências	3
Selecionar cenário de migração	1

Seguem os motivos pelos quais as atividades foram selecionadas:

**Analisar custo da migração** – Essa atividade tem como resultado uma estimativa do custo da migração. O custo está associado com esforço necessário a tornar o legado em serviços de Computação em Nuvem, assim como com o custo que a aplicação terá após migração.

Quanto maior o custo de migração de um sistema, maior o risco para todo o processo de migração, caso a migração tenha problemas (atrasos, diminuição de funcionalidades ou de desempenho). Por outro lado, um sistema cuja migração tenha baixo custo pode ser um bom candidato para validar o processo de migração e para se ganhar experiência tanto com o processo quanto com o provedor para o qual o sistema será migrado.

A medida de custo pode ser absoluta, baseada em faixas de valores na moeda corrente, ou pode ser um percentual em relação ao orçamento setorial da organização, ou ao orçamento específico de migração de sistemas legados para a Nuvem.

Caso o custo da migração somado ao custo da execução do sistema no provedor de nuvem não representem uma economia em relação ao custo atual, esse fato pode indicar um sistema que não se beneficiará da nuvem, caso não haja agregação de valor por outro meio, como o uso de características intrínsecas da nuvem (auto-escalabilidade, maior disponibilidade, por exemplo) [5].

De acordo com o Modelo de Referência, essa tarefa está inserida como parte da atividade Analisar Contexto. Um dos objetivos da atividade é a mitigação de riscos e, por isso, a escolha da tarefa para compor o Indicador [5].

**Identificar dependências** – Essa atividade tem como resultado identificar quais outros sistemas e componentes da infraestrutura de *hardware* e *software* local o sistema legado depende para funcionar após a migração. Com esse levantamento, é possível identificar quais outros sistemas teriam de ser migrados em conjunto ou previamente ao sistema que está sendo analisado. Ou ainda, em caso de nuvem híbrida e caso as dependências não venham a ser migradas, será preciso avaliar e considerar a latência de rede entre a es-

estrutura local e a estrutura na nuvem. De acordo com Reza Bazi et al. [47], este é um fator crítico a ser considerado, já que os sistemas legados foram desenvolvidos em plataformas mais antigas do que a versão atual suportada por provedores de nuvem.

De acordo com o Modelo de Referência, essa tarefa está inserida como parte da atividade Identificar Sistema Legado, cujo objetivo está descrito como ‘Entender as dependências do sistema legado’ [5].

**Selecionar cenário de migração** – De acordo com as características do sistema legado e também de acordo com a quantidade de esforço que se pretende seja dispendida na migração, escolhe-se o cenário de migração, dentre as cinco opções definidas em [38] e [4]. O cenário de migração do tipo V, também conhecido como *Lift & Shift* [48], é o que menos esforço demanda e que tem o menor risco de falha na migração, quando comparado com os demais cenários, para os sistemas que não tenham requisitos críticos de desempenho [33]. Assim, a escolha do cenário pode estar diretamente relacionada com a intensidade de risco que se aceita. Isso não quer dizer que um cenário de menor risco seja também a opção que trará o maior benefício (redução de custos, aumento da disponibilidade), pois o cenário V, por exemplo, traz consigo uma maior dificuldade na implementação da elasticidade e risco de perda de desempenho [38] [33].

De acordo com o Modelo de Referência, essa tarefa está inserida como parte da atividade Definir Plano, cujos objetivos – Gerenciamento de projetos e Mitigação de riscos - reforçam a sua escolha para compor o indicador [5].

#### 4.4.2 Fase de projeto

Para a fase de Projeto foram selecionadas as tarefas e definidos os pesos que constam na Tabela 4.2. Os pesos foram definidos empiricamente, com base no conhecimento tácito e na experiência dos pesquisadores.

Tabela 4.2: Grupo de tarefas de avaliação da Fase de Projeto

Tarefa	Peso
Negociar com o provedor de nuvem	3
Treinamento	1
Identificar incompatibilidades	1

**Negociar com o provedor de nuvem** – Essa tarefa é bastante crítica, na medida em que após sua execução tem-se um ou mais provedores de nuvens disponíveis para fornecer os serviços que suportarão os sistemas a serem migrados. Caso ainda não tenha havido a contratação do provedor no momento da migração de um determinado sistema legado, isso pode postergar a migração ou causar retrabalho. Veja que, mesmo após um provedor já ter sido contratado e alguns sistemas legados terem sido migrados para ele, é possível

que o contrato com esse provedor esteja acabando, ou mesmo que a prestação do serviço não esteja satisfatória e que a organização tenha decidido substituí-lo antes do fim do contrato. Ambas as situações demandarão nova contratação. Contratações públicas têm riscos intrínsecos, pois devem obedecer aos normativos e leis que as regem e que definem tempo máximo de contratação, maior transparência, dentre outros requisitos legais.

**Treinamento** – Essa tarefa tem o objetivo de introduzir e manter, na organização, o conhecimento necessário ao projeto, operação e acompanhamento de serviços de Computação em Nuvem. Uma necessidade maior de treinamento pode indicar um risco maior na gestão do projeto de migração de um determinado sistema legado. Nas primeiras migrações, o conhecimento acerca de Computação em Nuvem pode estar deficitário na organização. Com o início do uso de serviços em nuvem em um determinado provedor, quando ele é recém-contratado, há uma maior necessidade de treinamento também. Mesmo em momentos de estabilidade em relação à contratação de novos provedores, uma mudança na equipe que gerencia o processo de migração pode demandar mais treinamento. Além disso, há a atualização dos serviços já em uso, assim como oferecimento de novos serviços de Computação em Nuvem pelo provedor. De acordo com Reza Bazi et al. [47], as organizações devem estender o conhecimento da nuvem como forma de garantir um começo bem sucedido.

**Identificar incompatibilidades** – Essa atividade tem como objetivo listar as incompatibilidades entre o sistema legado, da forma como ele é executado atualmente, e o conjunto de serviços de computação em nuvem definidos na fase de projeto como sendo necessários à execução do sistema no provedor após a migração. Esse conjunto de incompatibilidades demandará esforço específico para serem resolvidas. Quanto maior o conjunto, ou quanto maior o esforço necessário para sanar as incompatibilidades, maior o risco da migração específica.

### 4.4.3 Representação visual do conjunto de avaliação de cada fase

Após detalhar as razões que motivaram a escolha das tarefas que serão a base da função de avaliação, é importante apresentar novamente as fases do Modelo de Referência com uma identificação visual das tarefas que farão parte do conjunto de avaliação.

A figura 4.3 mostra como cada elemento-chave compõe a fase de Planejamento do Modelo de Referência definido por Gholami et al. [5]. Cada um dos elementos-chave foi associado a uma cor, que cobre as tarefas que compõem aquele elemento-chave. Além disso, há uma indicação específica (a sigla IPR, de Indicador de Percepção de Risco) nas tarefas que fazem parte do conjunto de avaliação. O mesmo artifício visual foi utilizado para a Fase de Projeto, representada pela Figura 4.4.

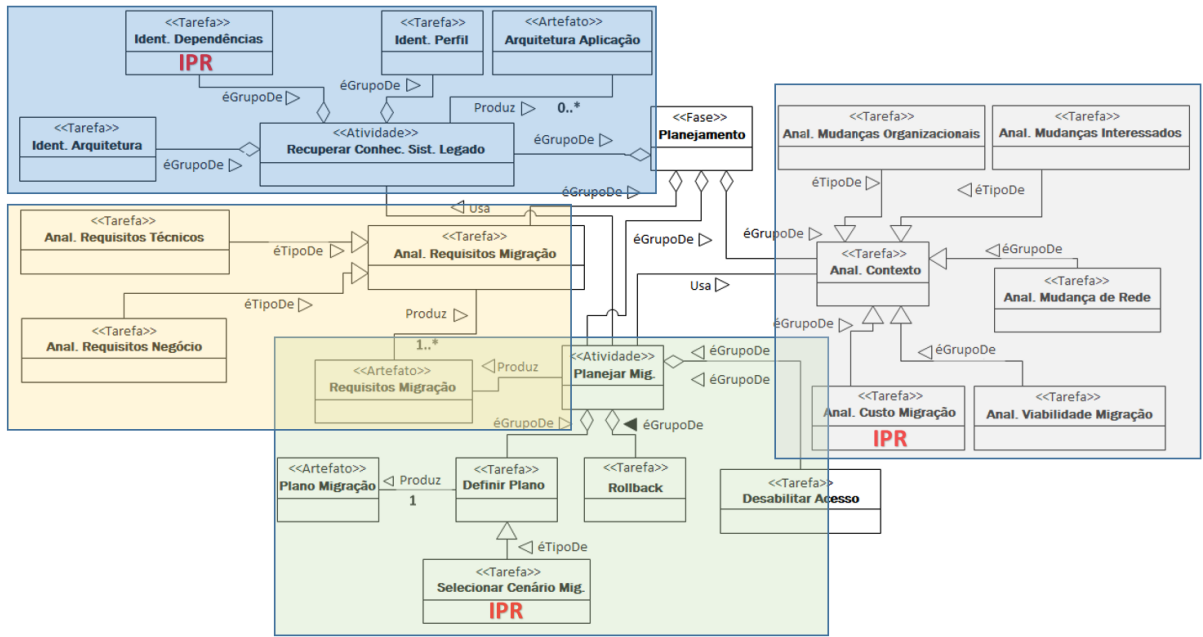


Figura 4.3: Fase de Planejamento com indicativo visual do conjunto de avaliação.

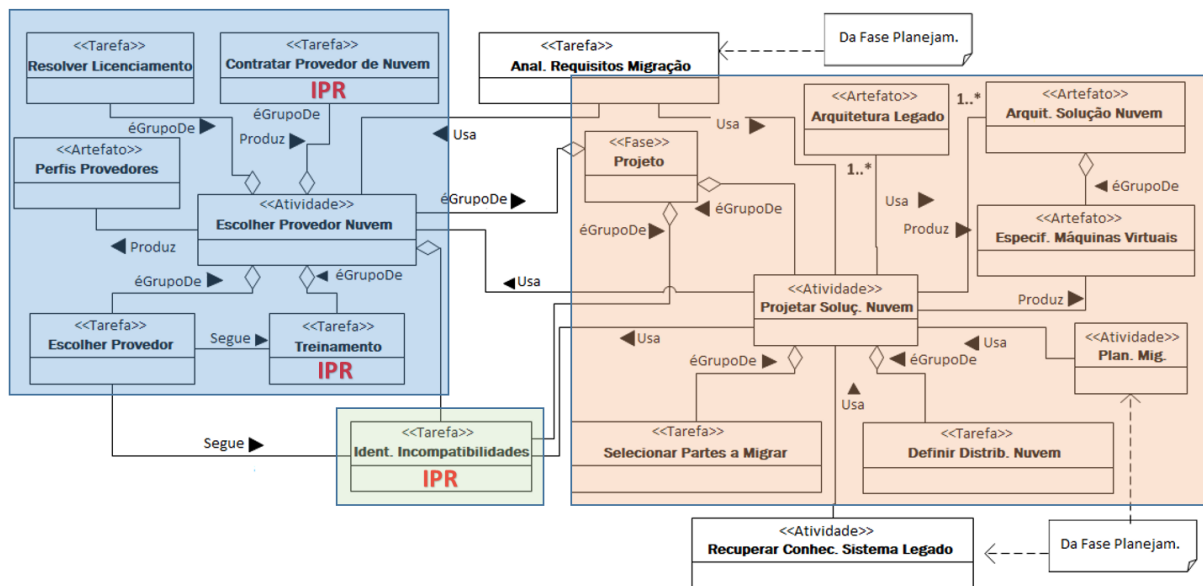


Figura 4.4: Fase de Projeto com indicativo visual do conjunto de avaliação.

## 4.5 Estudo de caso

Para mostrar a aplicabilidade do método, foram avaliados três sistemas legados de um órgão governamental que decidiu migrar seus sistemas legados para a Nuvem. Para cada um dos sistemas, vai ser calculado o Indicador de Percepção de Risco, conforme método exposto na seção 4.3, e sugerido uma ordem de migração a partir dos valores do indicador.

### 4.5.1 Sistema A - Nuvem Cívica

Este sistema legado expõe, na Internet, dados atualizados sobre instituições educacionais e de saúde em todo o país e dados sobre medicamentos que são autorizados pela agência governamental competente. Embora o sistema tenha Nuvem em seu nome, ele é executado em infraestrutura local.

O sistema é composto por um conjunto de serviços web, fracamente acoplados, que podem ser usados por desenvolvedores e empresas para agregar valor às suas aplicações. Por exemplo, pode-se desenvolver e publicar uma aplicação móvel que capture automaticamente a localização do usuário e sugira a instituição de saúde mais próxima do usuário. O sistema foi desenvolvido usando linguagem Java com o suporte do *framework* Spring MVC. Ele está sendo executado no servidor de aplicação JBOSS EAP, compartilhando recursos computacionais com outros sistemas legados. O sistema de gerenciamento de banco de dados é Oracle [49].

Como o sistema está disponível publicamente, qualquer desenvolvedor pode criar um aplicativo que se utilize dos serviços e dados fornecidos. Se uma dessas aplicações se tornar muito popular, com centenas ou milhares de transações por hora, há uma chance de que a infraestrutura de computação se torne inadequada devido à falta de elasticidade.

### 4.5.2 Sistema B - Pesquisa de jurisprudência

O órgão governamental é um tribunal e toma decisões que têm força normativa no âmbito administrativo do Estado brasileiro. De forma simplificada, dá-se o nome de jurisprudência ao conjunto destas decisões.

O Pesquisa de Jurisprudência é um sistema que permite que se façam buscas textuais nas quatro bases da jurisprudência do órgão: Acórdãos, Jurisprudência Seleccionada, Publicações e Súmulas. A busca pode ser em cada base individualmente ou, simultaneamente, em todas as bases. O resultado é uma lista com os documentos que contêm os termos buscados, que podem ser acessados diretamente pela mesma interface.

O sistema permite ainda o uso de operadores lógicos e de operadores de busca textual. Caso o usuário seja cadastrado na base de usuários do órgão, é possível ainda salvar pesquisas e marcar documentos como favoritos, de forma a agilizar chamadas posteriores a esses conteúdos.

O sistema foi desenvolvido em Java e faz uso do Solr [50], software de código aberto que implementa e gerencia a indexação e busca de diversos formatos de documentos.

### 4.5.3 Sistema C - Ambiente Virtual de Aprendizagem

Este sistema armazena e gerencia os diversos cursos – na modalidade ‘a distância’ – que o órgão governamental disponibiliza tanto para o consumo interno por seus servidores, quanto para consumo gratuito pela população. No momento da análise técnica realizada, havia uma oferta de 60 cursos, com variadas durações, com temas relacionados à Administração Pública em geral, assim como relacionados com a área de atuação do órgão.

Para participar de um dos cursos oferecidos, é preciso ser um usuário cadastrado e ter a participação aprovada, nos casos em que há limitação de vagas. O cadastro é gratuito, feito pelo portal público do órgão e um certificado é oferecido ao fim do curso, caso o aluno tenha atingido, na avaliação do curso, a pontuação mínima estabelecida.

Por mês, em média, 9 mil alunos realizam cursos a distância por meio deste sistema legado. As indisponibilidades eventuais, programadas ou não, do *datacenter* da organização causam problemas, principalmente quando ocorrem em períodos de avaliação dos cursos. Nestes casos, quando não é possível adiar uma indisponibilidade a ser causada por manutenção programada do *datacenter*, é necessário adiar as datas de avaliações e comunicar aos alunos, o que causa transtornos. Um aumento na disponibilidade do sistema, por estar na infraestrutura de um provedor de nuvem, é desejável e entendido como uma melhoria na prestação deste serviço educacional.

O sistema utiliza o Moodle [51] para a gestão educacional, cuja autenticação e autorização são fornecidas atualmente pelo Microsoft Active Directory [52].

### 4.5.4 Indicador de percepção de risco e ranking de migração dos sistemas

Foram realizadas reuniões técnicas com a equipe responsável por cada um dos três sistemas legados descritos e aplicado o modelo de referência de migração para a Nuvem em suas duas fases iniciais: Planejamento e Projeto. Após o trabalho prescrito em cada fase, o método definido na Seção 4.3 foi aplicado, usando as tarefas e pesos especificados na Seção 4.4.

As equipes avaliaram a percepção de risco para cada tarefa no conjunto de avaliação de ambas as fases para os três sistemas. Os valores podem ser vistos na tabela 4.3.

Usando os pesos definidos (Tabelas 4.1 e 4.2), as avaliações dadas pela equipe (Tabela 4.3) e aplicando a equação (4.1), calcularam-se os valores para o Indicador de Percepção de Risco da migração de cada sistema, o que permitiu ranqueá-los em ordem de menor risco para maior risco, conforme dados exibidos na Tabela 4.4. Após o cálculo do Indicador de Percepção de Risco, tem-se que o Sistema C deve ser o primeiro a ser migrado, seguido do Sistema A e do Sistema B.

Tabela 4.3: Avaliação da percepção de risco associada às tarefas no conjunto de avaliação das fases Planejamento e Projeto.

Fase	Tarefa	Sistema A	Sistema B	Sistema C
Planejamento	Analisar custo da migração	5	3	4
	Identificar dependências	4	3	4
	Selecionar cenário de migração	4	4	5
Projeto	Contratar o provedor de nuvem	4	5	5
	Treinamento	2	3	3
	Identificar incompatibilidades	5	5	5

Tabela 4.4: Cálculo do Indicador de Percepção de Risco para cada sistema avaliado.

Sistema	Pl	Pr	IPR	Ranking
Sistema C	4,14	4,60	<b>4,37</b>	<b>1</b>
Sistema A	4,43	3,80	<b>4,11</b>	<b>2</b>
Sistema B	3,14	4,60	<b>3,87</b>	<b>3</b>

Conforme sugerido na Seção 4.3, os limiares de reavaliação, L1 e L2, são definidos como sendo 3,0. Uma vez que  $IPR$ ,  $Pl$  e  $Pr$  são maiores que esses limiares, isso indica uma percepção de risco aceitável e não há, portanto, indicação de revisão para estas migrações.

## 4.6 Conclusão deste capítulo

Este capítulo apresentou uma proposta que consiste em adicionar o Indicador de Percepção de Risco (IPR) da migração ao Modelo de Referência adotado para apoiar as migrações de sistemas legados do governo. Para o cálculo do IPR, foi apresentado um método em quatro passos, que é independente do domínio dos sistemas. No passo 1, escolhe-se o conjunto de tarefas de avaliação, por fase. No passo 2, definem-se os pesos para cada tarefa. No passo 3, faz-se a avaliação de cada tarefa e aplica-se a fórmula definida. No passo 4, faz-se um ranking dos sistemas, ordenado pelo valor do IPR. Para mostrar a aplicabilidade do método, uma prova de conceito para o domínio do governo foi descrita, com a implementação dos passos 1 e 2. Em um estudo de caso com três sistemas legados do governo os passos 3 e 4 foram implementados, ou seja, foi calculado o IPR da migração destes sistemas para a Nuvem e fez-se um ranking dos três sistemas, ordenando-os do menor para o maior risco percebido da migração. O IPR fornece uma avaliação objetiva da execução do Modelo de Referência antes da fase de implementação da migração. Ele fornece as condições para que se inicie a migração por sistemas de baixo risco, o que pode aumentar a taxa de sucesso das migrações e, conseqüentemente, a confiança e experiência da equipe.



# Capítulo 5

## Análise experimental

O objetivo deste capítulo é validar metodologicamente o uso de um modelo de referência, apresentado no Capítulo 3, para apoiar a migração de sistemas legados do governo para a Nuvem e avaliar o indicador de percepção de risco da migração, conforme a proposta descrita no Capítulo 4. A análise está baseada na aplicação de um método para coleta de dados, validação e consolidação dos resultados de pesquisa de opinião no âmbito da Administração Pública Federal brasileira. Há uma descrição do planejamento da pesquisa, construção e aplicação do questionário, e da coleta das respostas. Espera-se que, após a análise dos resultados, sejam identificados os ajustes necessários aos parâmetros da proposta – tarefas que compõem o cálculo do Indicador de percepção de risco e respectivos pesos por fase – para que reflitam os achados da pesquisa.

### 5.1 Método de coleta dos dados

#### 5.1.1 Pesquisa de opinião na Administração Pública

O objetivo da pesquisa é coletar a opinião da comunidade de TI da APF acerca dos componentes fundamentais do que está sendo proposto por este estudo: o uso de um processo estruturado para apoiar a migração de sistemas legados do governo para a Nuvem, associado a um indicador de percepção de risco. A partir da consolidação dos resultados, os parâmetros da proposta podem ser reavaliados: tarefas usadas para cálculo do indicador e respectivos pesos sugeridos para elas.

Será calculada a margem de erro para o conjunto de respostas, que terá a confiabilidade avaliada pelo Alfa de Cronbach [53]. O teste Signed Rank de Wilcoxon [54] será utilizado para fazer inferências a partir dos níveis de importância fornecidos pelos respondentes na amostra. O teste de Wilcoxon pode ser visto como uma alternativa não-paramétrica ao método T de Student para os casos em que este não possa ser aplicado [55].

Foram seguidas as orientações para o planejamento da pesquisa de opinião, escolha da amostra e coleta de dados delineados por Pinsonneault and Kraemer [56]. O método de coleta escolhido foi o questionário eletrônico, com convite enviado por e-mail (ver Anexo I). A forma, organização e distribuição das perguntas seguiram as orientações contidas em Günther [57]. O questionário foi desenvolvido com o auxílio do LimeSurvey [7], que é um sistema web que auxilia na construção de questionários eletrônicos e também armazena as respostas registradas.

A pesquisa foi pública e o convite foi enviado para servidores públicos federais brasileiros com experiência em TI. Utilizou-se uma base de endereços de e-mails que continha pelo menos um e-mail relacionado com a gestão de Tecnologia da Informação de cada organização do Governo Federal dos três poderes e do Ministério Público.

Adicionalmente, foram contactadas duas empresas que organizam eventos de TI e que nos últimos anos fizeram eventos sobre Computação em Nuvem, sendo esses eventos de participação gratuita e exclusiva para servidores públicos. Ambas as empresas aceitaram encaminhar o convite aos participantes dos eventos que realizaram.

A participação na pesquisa foi voluntária. A cada dez dias, durante o período de coleta de dados, um lembrete amigável foi enviado aos convidados que não responderam à pesquisa.

### 5.1.2 Estrutura do questionário

Esta seção abordará a estrutura do questionário e apresentará as principais perguntas que o compõem. O Anexo II contém o questionário completo, incluindo os textos de apoio para cada pergunta.

O questionário foi dividido em dois grupos de perguntas:

Grupo 1 - as perguntas deste grupo dizem respeito aos componentes fundamentais da proposta (Capítulo 4). A tabela 5.1 mostra o relacionamento entre os objetivos específicos apresentado na Seção 1.2 e as perguntas.

Grupo 2 - as perguntas deste grupo se relacionam tanto com a organização em que o respondente trabalha, quanto com o próprio respondente.

Alguns dos passos e eventos mais significativos na construção do questionário:

Tabela 5.1: Relacionamento dos objetivos específicos deste trabalho com as perguntas do questionário.

<b>Objetivo específico</b>	<b>Perguntas</b>
Propor modelo de referência de migração de legados	1
Propor indicador de percepção de risco	2, 3 e 4

- Em 4/5/2018 - Primeira versão foi enviada a um grupo de controle com cinco participantes.

- Análise dos tempos de resposta, mostrou que o tempo para responder ao questionário era de dez minutos.

- Após os *feedbacks* do grupo de controle, foram retiradas duas perguntas que objetivavam ordenar as atividades do cálculo do indicador de percepção de risco, pois essa informação poderia ser derivada de uma das perguntas existentes (Pergunta 4).

- Mudança na apresentação das perguntas de ‘pergunta por pergunta’ para ‘por grupo’. Essa mudança reduziu o tempo de resposta.

- Foram feitas algumas pequenas correções. A principal, no texto das opções da escala *likert* para que ficasse simétrica e com um elemento neutro central. Esse ajuste ajuda a prevenir a ocorrência de viés causado pelo desbalanceamento da escala.

## **Introdução**

O questionário se inicia com uma breve explicação do seu objetivo, do contexto da pesquisa em andamento e com um breve direcionamento e chamamento aos servidores públicos de TI, público-alvo da pesquisa de opinião.

Este questionário foi criado com a intenção de coletar as opiniões dos colegas, servidores e empregados públicos, acerca da adoção de Computação em Nuvem na Administração Pública brasileira. Ele faz parte de pesquisa acadêmica que estou desenvolvendo sobre o tema na Universidade de Brasília - UnB.

O tempo para respondê-lo é inferior a 10 minutos. O objetivo é coletar as opiniões das pessoas e não de ter um posicionamento oficial das organizações, assim vários servidores de uma mesma organização podem respondê-lo.

Todas as suas declarações serão tratadas de maneira confidencial. Os resultados serão apresentados de maneira a não permitir a identificação de participantes individuais.

Se você tiver dúvidas ou quiser relatar um problema relacionado à pesquisa, entre em contato pelo e-mail [brenogc@tcu.gov.br](mailto:brenogc@tcu.gov.br), ou pelo celular/whatsApp 61 98401 1080. Tentarei responder da melhor maneira possível qualquer dúvida que haja.

Peço-lhe a gentileza de contribuir com esta pesquisa. Suas opiniões e experiências são muito importantes!

Seguindo as orientações de Günther [57], o questionário iniciou-se pelas perguntas mais importantes, relacionadas diretamente com os objetivos de pesquisa.

### **Pergunta 1 - Qual a importância do uso de processo como apoio à migração de sistemas para a nuvem?**

A pergunta 1 questiona o respondente sobre a importância de se ter um processo estruturado em apoio à migração de sistemas legados para a Nuvem. O respondente deve

escolher o nível de importância em uma escala de 1 a 5, conforme a Figura 5.1. O texto de apoio, com mais detalhes sobre o processo definido por esta pesquisa (Capítulo 3), estava disponível para o respondente e está incluído no Anexo II, após o questionário.

1 Qual o nível de importância que você atribui ao uso de um processo estruturado (com fases, atividades e documentos definidos) para apoiar a migração de sistemas legados para a nuvem? Use uma escala de 1 a 5, conforme opções a seguir:

Escolha uma das seguintes respostas:

- 5 - Totalmente importante
- 4 - Importante
- 3 - Neutro
- 2 - Sem importância
- 1 - Totalmente sem importância

Como parte da proposta que será feita na pesquisa acadêmica, estará a indicação de um processo com três fases, cada fase composta por atividades, cada atividade com um objetivo de coleta de informação, decisão técnica, ou mesmo documentação do sistema e dos serviços em Nuvem que o implementarão após a migração. Caso queira mais detalhes sobre as fases e atividades, clique [aqui](#).

Figura 5.1: Pergunta 1.

### **Pergunta 2 - Há representatividade no indicador de risco conforme cálculo proposto?**

A pergunta 2 explica resumidamente o cálculo do indicador de percepção de risco, conforme proposto por esta pesquisa (Capítulo 4) e questiona se o respondente o considera um indicador representativo. É uma pergunta do tipo Sim/Não, como mostra a Figura 5.2. Caso a resposta seja ‘Sim’, a Pergunta 3 aparecerá para o respondente e é obrigatório respondê-la. Caso a resposta seja ‘Não’, o respondente não verá a Pergunta 3.

A intenção é analisar, na amostra selecionada, qual a percepção da representatividade do indicador que está sendo proposto.

### **Pergunta 3 - Iniciar o processo de migração por um sistema com percepção de risco baixo é boa prática?**

A pergunta 3 coleta a percepção dos respondentes acerca do uso do indicador de percepção de risco para ranquear os sistemas, ordenando-os da menor percepção de risco para a maior percepção de risco. É uma pergunta do tipo Sim/Não, como mostra a Figura 5.3.

\* 2 Na migração de um sistema legado para a Nuvem, pode ser atribuído um score que indique a percepção do grau de risco nesse processo. A metodologia do cálculo do score está descrita a seguir. Na sua visão, a utilização desse score conforme essa metodologia é representativa como indicador?

Sim

Não

🔗 O cálculo de um valor (score) para o risco percebido pela equipe na migração de um sistema legado para a Nuvem utiliza as informações geradas por seis tarefas selecionadas nas duas primeiras fases do processo apresentado na pergunta anterior. Pontua-se cada tarefa em uma escala de 1 a 5 (1=Risco Muito Baixo e 5=Risco Muito Alto). Com esses valores calcula-se a média e obtém-se um score da percepção de risco da equipe para esta migração.

Figura 5.2: Pergunta 2.

\* 3 Você considera uma boa prática iniciar o processo de migração para a Nuvem por um sistema legado que apresente percepção de risco baixo (score), deixando para momentos posteriores os sistemas com percepção de risco alto?

Sim

Não

🔗 No contexto de vários sistemas a serem migrados, o score de percepção de risco seria utilizado para ordenar as migrações, iniciando pelas que têm menor percepção de risco.

Figura 5.3: Pergunta 3.

#### **Pergunta 4 - Qual o nível de importância de cada uma das tarefas selecionadas para compor o indicador de percepção de risco?**

A pergunta 4 (ver Figura 5.4) solicita que o respondente escolha o nível de importância de cada uma das tarefas que compõem o cálculo do indicador de percepção de risco. As tarefas foram selecionadas por coletarem ou produzirem informações que se relacionam fortemente com análise de risco, conforme explanado no Capítulo 4.

Quer-se analisar qual a aderência dos argumentos utilizados nesse estudo com a percepção dos respondentes no domínio do governo. A partir das diferentes frequências de valores para cada uma das faixas de importância de cada tarefa, pode-se retirar alguma das tarefas que não seja considerada importante e também ajustar os pesos que cada tarefa terá no cálculo do indicador.

Os pesos foram definidos empiricamente, com base no conhecimento tácito e na experiência dos pesquisadores (Seção 4.4). A intenção é reavaliá-los em acordo com o resultado das frequências de respostas em que cada tarefa seja considerada como importante e atribuir maior peso às maiores frequências.

As demais perguntas visam a coletar dados demográficos sobre os respondentes e sobre as organizações em que trabalham, de forma que seja possível caracterizar o perfil dos

\* 4 Qual o nível de importância que você atribui a cada uma das seguintes tarefas que subsidiarão o cálculo da percepção de risco (score) da migração de um sistema legado para a nuvem? Use a escala de valores de 1 a 5, conforme opções a seguir.

	5 - Totalmente importante	4 - Importante	3 - Neutro	2 - Sem importância	1 - Totalmente sem importância
Analisar custo da migração	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identificar dependências	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Selecionar cenário de migração	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contratar o provedor de nuvem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Treinar equipe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identificar incompatibilidades	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

As tarefas são subdivisões das atividades do processo de migração apresentado na Pergunta 1. Elas foram escolhidas como sendo as que melhor apontam os riscos que cada migração de sistema legado para a nuvem pode apresentar. Mais detalhes sobre cada atividade selecionada podem ser encontrados [aqui](#).

Figura 5.4: Pergunta 4.

respondentes e das organizações.

Em relação às organizações, há perguntas acerca da importância da Computação em Nuvem para a organização, assim como qual o uso atual da tecnologia e o prazo para intensificar esse uso.

Em relação aos respondentes, há perguntas sobre a experiência com TI, com Computação em Nuvem e com alguns provedores específicos. O questionário completo está disponível no Anexo II.

Os dados demográficos coletados serão utilizados nesta pesquisa apenas para apoiarem os objetivos definidos. Uma descrição detalhada do perfil das organizações e dos respondentes, assim como uma comparação dos dados demográficos coletados com a literatura, está fora do escopo deste trabalho, mas podem ser realizados posteriormente.

### 5.1.3 Descrição da amostra

O TCU realizou um diagnóstico na APF acerca dos servidores públicos federais que trabalham com TI [58]. O relatório aponta que há 448 organizações federais que empregam, exclusivamente na área de TI, 38114 servidores públicos, que foi a população do estudo.

A amostra foi composta de servidores públicos que tinham alguma experiência em TI, mesmo que não tivessem experiência em Computação em Nuvem.

Estima-se que a propensão para responder a questionários recebidos por e-mail para os profissionais de TI do governo seja a mesma da população em geral e, portanto, acredita-se não ter havido viés de seleção oriundo da ferramenta de coleta. Por outro lado, é comum

se reencaminhar um convite de resposta às pessoas mais qualificadas da organização para respondê-lo. Essa possibilidade de viés de seleção, enquanto diminui a aleatoriedade da amostra, pode aumentar a representatividade dela, uma vez que o conjunto de profissionais de TI do governo e **que trabalha ou estuda com Computação em Nuvem** é um subconjunto da população que está sendo estudada e, portanto, a amostra desta pesquisa nesse caso teria margem de erro menor para uma população também menor.

Os convites foram enviados em acordo com o seguinte cronograma:

- Em 24/5/2018, o questionário foi distribuído para diversos grupos de servidores públicos: responsáveis por departamentos de TI em órgãos públicos. 530 convites foram enviados por correio eletrônico.

- Em 12/6/2018, o HDI Brasil [59], empresa que organiza evento gratuito e exclusivo para servidores públicos, enviou o convite para a sua base de 650 alunos do governo.

- Em 15/6/2018, o IBGP [60], empresa que organiza eventos gratuitos e exclusivos para servidores públicos, enviou o convite para a sua base de 980 alunos.

- Em 25/6/2018, a pesquisa foi fechada com 242 respostas registradas.

Após o envio dos convites por e-mail, o total de respostas registradas aumentou diariamente conforme a seguir:

- Dia 1 - 45 respostas completas
- Dia 5 - 65 respostas completas
- Dia 10 - 109 respostas completas
- Dia 15 - 138 respostas completas
- Dia 20 - 201 respostas completas
- Dia 25 - 226 respostas completas
- Dia 30 - 242 respostas completas

Depois do trigésimo dia, a pesquisa foi bloqueada e não se aceitou mais o registro de novas respostas.

A título de ilustração, a Figura 5.5 mostra a página de acompanhamento do LimeSurvey [7], ferramenta utilizada para gerenciar o questionário e coletar as respostas.

Assim que a coleta de respostas terminou, fez-se necessário limpar e ajustar os dados. Primeiro, retirou-se do conjunto de dados de análise as respostas cujos respondentes não se enquadravam nos critérios de participação, que eram: ser servidor público e ter alguma experiência com TI. Em seguida, foram retiradas as colunas de preenchimento automático pela ferramenta de coleta utilizada, como registros de tempo de respostas, endereço IP do respondente, entre outros. Por fim, algumas mudanças de formato foram realizadas para que os dados se adequassem às análises, mas sem que houvesse perda ou modificação da informação coletada.

The screenshot shows the LimeSurvey interface for a survey titled 'Navegar respostas: (Pesquisa de opinião sobre migração de sistemas da Administração Pública para a Nuvem)'. The browser address bar shows 'https://contas.tcu.gov.br/limesurvey/index.php/admin/responses/sa/browse/surveyid/754...'. The interface includes a toolbar with various icons and a control panel with 'Mostrar registros: 50', 'Começando desde: 0', and 'Mostrar' button. Below the control panel is a table with the following columns: 'Qual o nível de importância', 'Na migração de um sistema legado para a', 'Você considera uma boa', 'Qual o nível de importância', 'Qual o nível de importância', 'Qual o nível de importância', 'Qual o nível de importância', 'Qual o nível de importância', 'Qual o nível de importância', 'Qual o nível de importância que você atribuiu a cada', and 'Considerando as três tarefas da'. The table contains 13 rows of data with various importance levels and migration status indicators.

Qual o nível de importância	Na migração de um sistema legado para a	Você considera uma boa	Qual o nível de importância	Qual o nível de importância	Qual o nível de importância	Qual o nível de importância	Qual o nível de importância	Qual o nível de importância	Qual o nível de importância que você atribuiu a cada	Considerando as três tarefas da
5 - Totalmente importante [A]	Não [N]		5 - Totalmente importante	5 - Totalmente importante [L001]	5 - Totalmente importante	4 - Importante [L002]	5 - Totalmente importante	5 - Totalmente importante	5 - Totalmente importante [L001]	
4 - Importante [B]	Sim [Y]	Sim [Y]	3 - Neutro [L003]	4 - Importante [L002]	3 - Neutro [L003]	3 - Neutro [L003]	3 - Neutro [L003]	3 - Neutro [L003]	5 - Totalmente importante [L001]	
4 - Importante [B]	Sim [Y]	Sim [Y]	5 - Totalmente importante	5 - Totalmente importante [L001]	3 - Neutro [L003]	3 - Neutro [L003]	4 - Importante [L002]	5 - Totalmente importante	5 - Totalmente importante [L001]	
3 - Neutro [C]	Sim [Y]	Sim [Y]	4 - Importante [L002]	5 - Totalmente importante [L001]	5 - Totalmente importante	3 - Neutro [L003]	5 - Totalmente importante	5 - Totalmente importante	5 - Totalmente importante [L001]	
4 - Importante [B]	Sim [Y]	Não [N]	4 - Importante [L002]	4 - Importante [L002]	4 - Importante [L002]	4 - Importante [L002]	3 - Neutro [L003]	3 - Neutro [L003]	3 - Neutro [L003]	
4 - Importante [B]	Sim [Y]	Sim [Y]	2 - Sem importância	5 - Totalmente importante [L001]	3 - Neutro [L003]	2 - Sem importância	4 - Importante [L002]	5 - Totalmente importante	5 - Totalmente importante [L001]	
5 - Totalmente importante [A]	Sim [Y]	Sim [Y]	5 - Totalmente importante	5 - Totalmente importante [L001]	5 - Totalmente importante	2 - Sem importância	5 - Totalmente importante	5 - Totalmente importante	5 - Totalmente importante [L001]	
4 - Importante [B]	Sim [Y]	Sim [Y]	4 - Importante [L002]	5 - Totalmente importante [L001]	3 - Neutro [L003]	5 - Totalmente importante	4 - Importante [L002]	5 - Totalmente importante	5 - Totalmente importante [L001]	
4 - Importante [B]	Sim [Y]	Sim [Y]	5 - Totalmente importante	4 - Importante [L002]	4 - Importante [L002]	4 - Importante [L002]	5 - Totalmente importante	4 - Importante [L002]	4 - Importante [L002]	
4 - Importante [B]	Sim [Y]	Sim [Y]	5 - Totalmente importante	4 - Importante [L002]	3 - Neutro [L003]	3 - Neutro [L003]	4 - Importante [L002]	5 - Totalmente importante	5 - Totalmente importante [L001]	
5 - Totalmente importante [A]	Sim [Y]	Sim [Y]	5 - Totalmente importante	5 - Totalmente importante [L001]	5 - Totalmente importante	5 - Totalmente importante	5 - Totalmente importante	5 - Totalmente importante	5 - Totalmente importante [L001]	

Figura 5.5: Acompanhamento das respostas no LimeSurvey [7].

No total, o convite com o acesso ao questionário eletrônico foi enviado para 2160 servidores públicos e 261 respondentes qualificados responderam à pesquisa. Das 261 respostas, 232 foram validadas para análise após a retirada de respostas incompletas e após a aplicação dos critérios de exclusão, resultando em uma taxa final de resposta aproximada de 11%. Importante registrar que aproximadamente 20% dos convites enviados retornaram pelo endereço não ter sido encontrado. Credita-se esse volume de erros no envio do convite à mobilidade de servidores entre organizações da APF, o que causa mudança no endereço de e-mail. Considerando apenas os convites que não tiveram erro no envio, a taxa de de resposta aumenta para aproximadamente 14%.

Não foram encontradas na literatura referências específicas acerca de taxas de resposta adequadas para questionários enviados por e-mail no cenário nacional. Duas pesquisas internacionais que compararam a taxa de resposta de questionários enviados por correio com envio por e-mail encontraram taxas de retorno de 17% [61] e 15%[62] para o envio por e-mail, similares a que esse trabalho teve.

Do total de respostas validadas foram retiradas as que diziam respeito a organizações das esferas estadual (37 respostas) e municipal (2 respostas), resultando em uma amostra com 193 registros, representando 135 diferentes organizações federais.



O tempo médio de experiência com TI do respondente é de 16 anos e de experiência com Computação em Nuvem é de 2 anos.

Os resultados dessa amostra apresentam uma margem de erro de 7% e um nível de confiança de 95%.

## 5.2 Análise dos resultados

### 5.2.1 Verificação de confiabilidade

Para verificar a consistência interna e a confiabilidade do questionário foi calculado o coeficiente Alfa de Cronbach. De acordo com da Hora et al. [53], o Alfa de Cronbach mede a correlação entre as respostas em um questionário por meio da análise do perfil das respostas dadas pelos respondentes [53]. É calculado a partir do somatório da variância dos itens individuais e da soma da variância de cada avaliador. Sua fórmula de cálculo está descrita na Equação 5.1:

$$\alpha = \left[ \frac{k}{k-1} \right] \times \left[ 1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right] \quad (5.1)$$

Na Equação 5.1, as variáveis são as seguintes:

$k$  - corresponde ao número de itens (perguntas) do questionário;

$S_i^2$  - corresponde à variância de cada item;

$S_t^2$  - corresponde à variância total do questionário (soma das variâncias dos avaliadores).

O Alfa de Cronbach retorna valores entre 0 e 1. Quanto maior o valor, mais confiável é o questionário. Apesar da ampla literatura científica a respeito das aplicações do Alfa de Cronbach nas diversas áreas do conhecimento, não existe um consenso entre os pesquisadores acerca do valor mínimo deste coeficiente que aponte a confiabilidade de um questionário [53]. Um limite inferior geralmente aceito para o Alfa de Cronbach é de 0,7, apesar de poder diminuir para 0,6 em pesquisas exploratórias [63].

Utilizou-se a ferramenta R [64] para o cálculo do Alfa de Cronbach. O resultado do Alfa foi 0,7826 para o conjunto de dados das respostas coletadas por essa pesquisa de opinião. Este resultado representa um forte indício da confiabilidade do questionário.

Como mostram os histogramas do conjunto das respostas às perguntas 1 e 4 (Figuras 5.6 e 5.7), os dados coletados têm distribuição assimétrica negativa, o que significa que a maioria das observações está concentrada nos valores mais altos, afastando-se assim da distribuição normal.

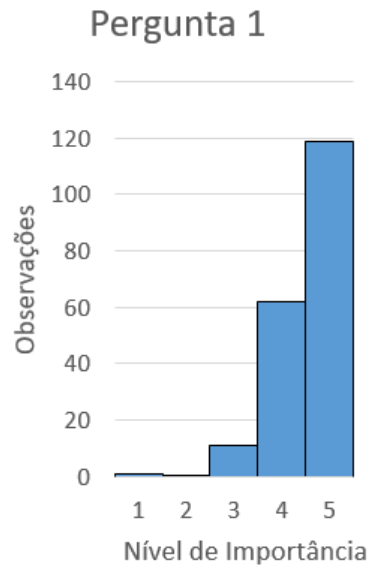


Figura 5.6: Distribuição das respostas da Pergunta 1.

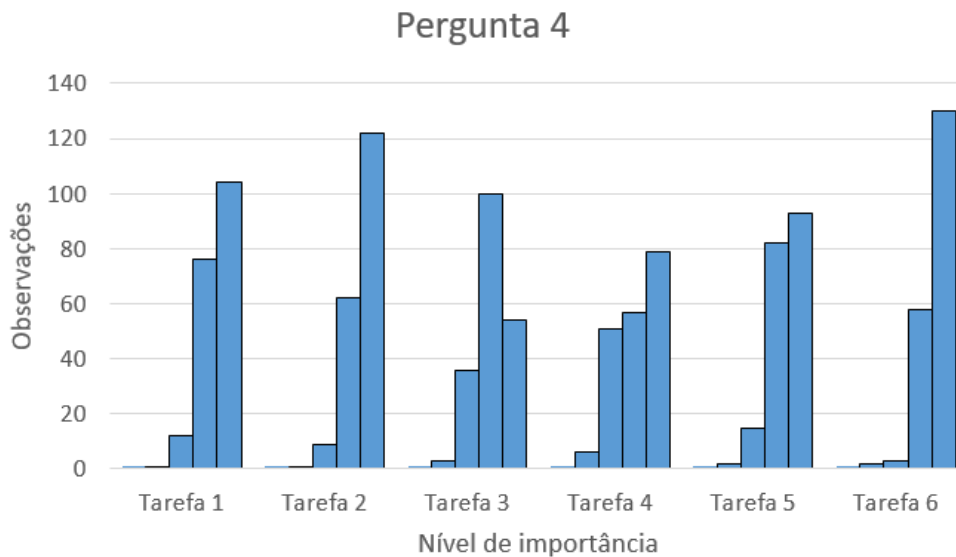


Figura 5.7: Distribuição das respostas por tarefa da Pergunta 4.

Pelo fato de a amostra apresentar uma distribuição que desvia da normalidade, assim como de não ser óbvio pela análise dos histogramas das Figuras 5.6 e 5.7 a qual distribuição os dados pertencem, testes estatísticos não-paramétricos devem ser utilizados. Os testes paramétricos assumem que a distribuição de probabilidade da população no qual retiramos

os dados seja conhecida e que somente os valores de certos parâmetros, tais como a média e o desvio padrão, sejam desconhecidos. Se os dados não satisfazem as suposições assumidas pelas técnicas tradicionais, métodos não paramétricos de inferência estatística devem ser usados. As técnicas não paramétricas assumem pouca ou nenhuma hipótese sobre a distribuição de probabilidade da população no qual se retiram os dados [65].

O teste Signed Rank de Wilcoxon [54] é um teste não paramétrico adequado para fazer análises estatísticas para a amostra coletada. O teste foi usado para examinar se as avaliações de importância, relativas ao uso do modelo de referência e a cada tarefa, estavam distribuídas a partir de uma mediana conhecida de 3, um valor de teste, representando ‘Neutro’. A hipótese alternativa foi a de que a mediana seria maior do que 3, o que significa que a maior parte das observações estariam acima de 3 e, portanto, nos valores que significam que os itens avaliados são considerados importantes.

Duas suposições foram verificadas antes de usar o teste estatístico de Wilcoxon. Em primeiro lugar, que as observações são independentes, o que se verificou ser correto, uma vez que nenhum respondente conhecia a identidade do conjunto total de respondentes. Houve representatividade de 135 órgãos diferentes do governo federal. Em segundo lugar, que as avaliações de importância têm distribuição simétrica na população, o que também se mostra correto, uma vez que há a mesma probabilidade na população de serem escolhidas tanto as opções que representam o processo ser importante quanto as que representam o contrário.

Uma vez que as suposições do teste de Wilcoxon [54] foram satisfeitas, o teste foi aplicado para análise estatística usando o software R [64] e os resultados estão consolidados na Tabela 5.2. Na terceira coluna estão os valores calculados para a estatística do teste (Estatística V) e na quarta coluna estão os valores do *p-value*, derivados a partir da Estatística V, para todos os itens. O  $p\text{-value} < 0,05$  significa que houve diferença estatisticamente significativa, validando a hipótese alternativa de que a mediana das avaliações de importância é maior do que 3. Esse resultado mostra que todos os itens avaliados – o uso do processo e as seis tarefas que compõem o indicador de percepção de risco – foram considerados importantes e reforça a possibilidade de adição de valor da proposta feita nesse estudo para migração de sistemas legados do governo para a Nuvem.

## 5.2.2 Resultados

As subseções seguintes apresentam os resultados consolidados da pesquisa de opinião.

Tabela 5.2: Resultados do teste Signed Rank de Wilcoxon para os itens propostos.

Item	Mediana	Estatística V	<i>p-value</i>
Uso de processo estruturado	5	16530	1,54e-31
Analisar custo da migração	5	16432	6,16e-35
Identificar dependências	5	17020	6,16e-35
Selecionar cenário de migração	4	12247	1,16e-28
Contratar o provedor de nuvem	4	9961	4,27e-25
Treinamento	4	15714	1,54e-31
Identificar incompatibilidades	5	18084	9,39e-36

### Uso de um processo estruturado para apoiar a migração de sistemas legados do governo para a nuvem

Conforme mostra a Figura 5.8, 93,78% (categorias 4 e 5 somadas) dos respondentes consideraram ser importante o uso de um processo estruturado de apoio à migração de sistemas legados do governo para a Nuvem.

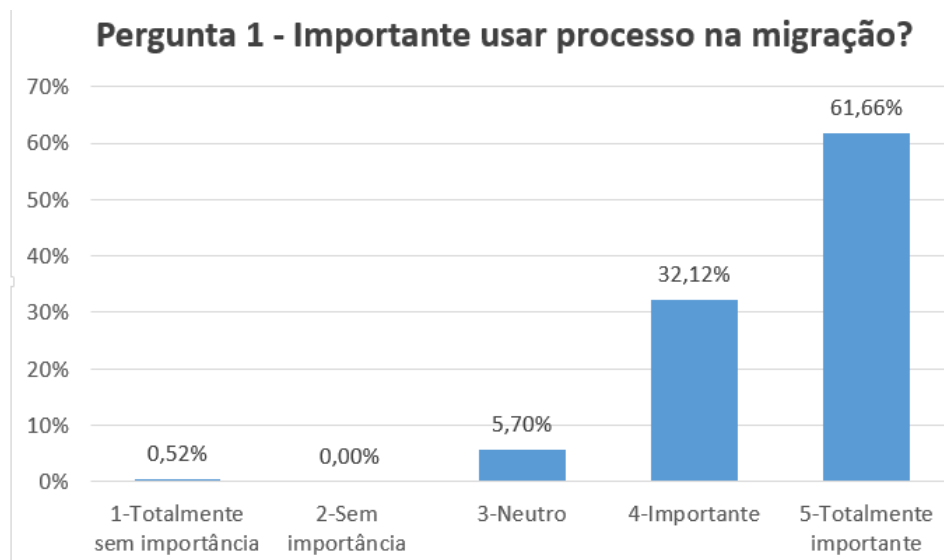


Figura 5.8: Importância de se usar um processo de apoio à migração.

Segundo os dados coletados, 82% das organizações representadas faz pouco ou nenhum uso atual de Computação em Nuvem, ao mesmo tempo que 43% dos respondentes têm experiência inferior a 1 ano com Computação em Nuvem.

O processo que esta pesquisa propõe que seja utilizado, e que cujo resumo foi apresentado aos respondentes, é um modelo de referência e, segundo Gholami et al. [5] facilita a compreensão da estrutura subjacente da migração para a Nuvem, condensando o conhecimento acumulado na literatura.

Uma vez que se decidiu preservar o investimento realizado nos sistemas legados [19] e se apropriar dos benefícios oferecidos pela Computação em Nuvem [14][18][13], as evidências

aqui coletadas apontam que há valor no fato de a migração ser realizada por um processo específico, principalmente nas organizações em que o uso da Computação em Nuvem ainda é pequeno e em que há pouca experiência da equipe de TI na área, como se caracteriza o cenário atual do governo brasileiro.

O resultado valida que, no escopo deste trabalho, o uso do processo de migração de sistemas legados do governo para a Nuvem, da forma que foi proposta no Capítulo 4, tem seu uso percebido como sendo importante no domínio do governo brasileiro.

### Indicador de percepção de risco

Em relação ao indicador de percepção de risco, conforme definição e método de cálculo apresentados na seção 4.3, 84,97% dos respondentes consideraram que ele é representativo (ver Figura 5.9). Esse é um resultado interessante, vez que o indicador pode ser calculado com pouco esforço e pode ser uma ferramenta adicional na gestão dos riscos específicos da migração de sistemas legados do governo.

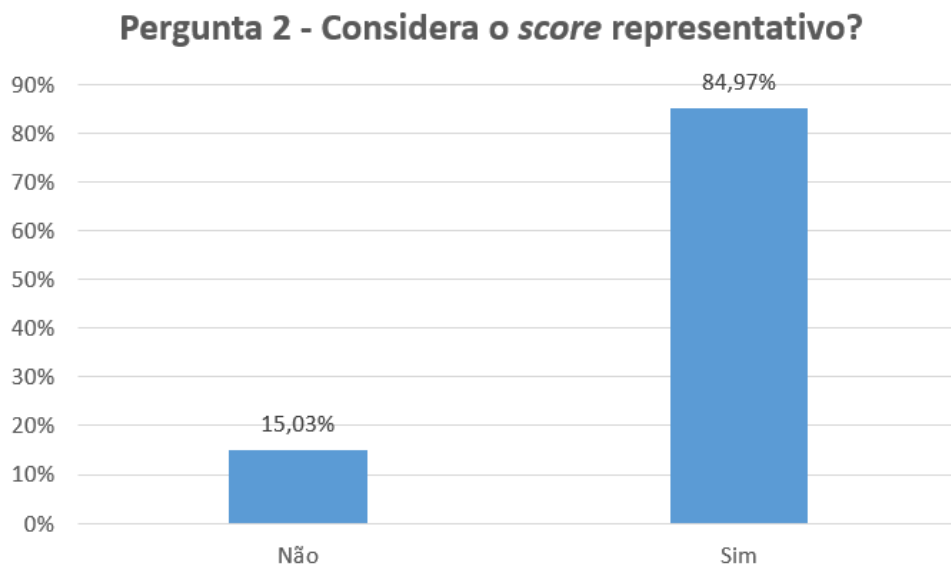


Figura 5.9: Representatividade do indicador de percepção de risco, conforme proposto por esse estudo.

Dos respondentes que consideraram o indicador representativo, 91,46% (Figura 5.10) percebem como boa prática iniciar a migração pelos sistemas que apresentem um indicador de percepção de risco baixo. Corroborando com a estratégia sugerida por Reza Bazi et al. [47], o sistema legado com a menor percepção de risco pode ser usado como piloto tanto para a migração inicial para a nuvem, quando nenhum sistema legado tiver sido migrado, quanto para a primeira migração para um novo provedor específico.

A pesquisa de opinião também solicitou que os respondentes avaliassem a importância de cada uma das tarefas pré-selecionadas para compor o cálculo do indicador. O resultado

### Pergunta 3 - É boa prática ranquear migrações?

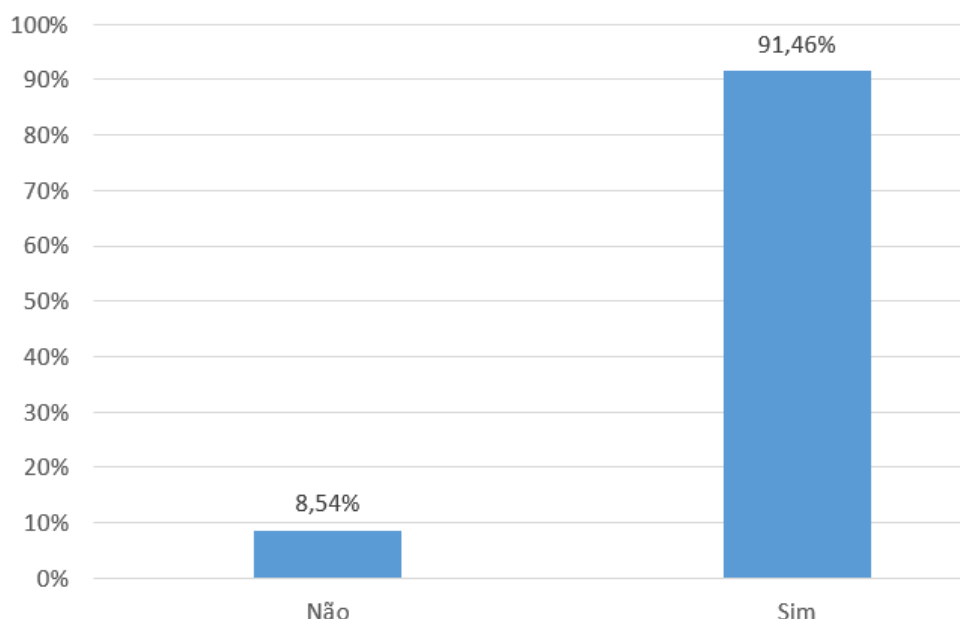


Figura 5.10: Se considera boa prática iniciar a migração por sistemas com baixa percepção de risco.

por nível de importância para cada tarefa pode ser visto na Figura 5.11. **Todas as seis tarefas foram consideradas importantes, com os percentuais (soma dos níveis ‘4-Importante’ e ‘5-Totalmente Importante’) resumidos na Tabela 5.3.**

Tabela 5.3: Percentual dos respondentes que consideraram importante cada tarefa do cálculo do indicador.

Fase	Tarefa	4-Import.	5-Tot. Imp.	Total
Planejamento	Analisar custo da migração	39,38%	53,89%	93,26%
	Identificar dependências	32,12%	63,21%	95,34%
	Selecionar cenário de migração	51,81%	27,98%	79,79%
Projeto	Contratar o provedor de nuvem	29,53%	40,93%	70,47%
	Treinamento	42,49%	48,19%	90,67%
	Identificar incompatibilidades	30,05%	67,36%	97,41%

Interessante perceber que ‘Contratar o provedor de nuvem’ foi a tarefa que obteve o menor percentual de importância, 70,47%. No entanto, há um senso comum nas organizações do Governo que considera a ação de contratar fornecedores como sendo de risco elevado, devido às questões legais inerentes e devido ao tempo elevado para sua consumação.

Ao analisar a avaliação de importância para esta tarefa de acordo com o nível de experiência com TI do respondente (ver Figura 5.12), verifica-se que a percepção de importância (soma dos níveis ‘4-Importante’ e ‘5-Totalmente Importante’) aumenta com

## Pergunta 4 - Importância das tarefas para o score

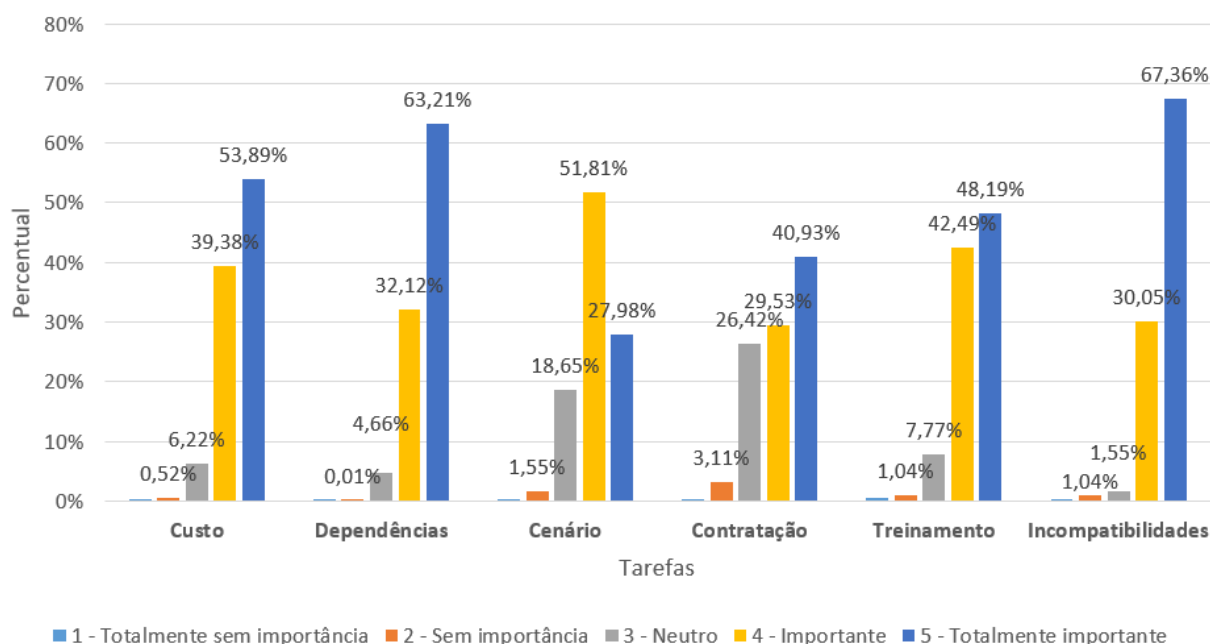


Figura 5.11: Importância atribuída a cada uma das seis tarefas selecionadas para compor o indicador de percepção de risco.

o aumento da experiência com TI. Passa de 53,85% no grupo com menos de 5 anos de experiência com TI para 82,86% no grupo com mais de 25 anos de experiência com TI.

## Tempo Exper. x 'Contratar o provedor de nuvem'

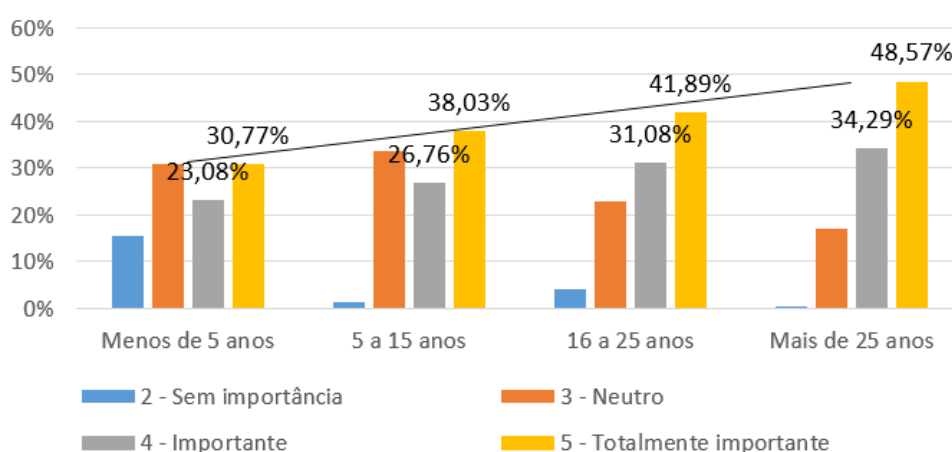


Figura 5.12: Importância atribuída à tarefa 'Contratar o provedor de nuvem' por faixa de experiência com TI.

Esse detalhamento indica que, embora a parcela dos respondentes com mais experiência perceba essa tarefa como sendo da mesma importância que as demais de sua fase, já que estão empatadas tecnicamente na margem de erro de 7%, no contexto de migração de

legados do governo para a Nuvem, a percepção majoritária é de ela ser menos importante que ‘Treinamento’ e ‘Identificar incompatibilidades’.

## 5.3 Ajustes da proposta

Como as equações do cálculo do indicador de percepção de risco por fase preveem pesos específicos por tarefa (Equações (4.2) e (4.3)), propõe-se ajustar os pesos para que reflitam a ordem de importância alcançada pelas tarefas na pesquisa de opinião.

Para que cada tarefa permanecesse no cálculo do indicador de percepção de risco, era necessário que a soma das opções ‘4-Importante’ e ‘5-Totalmente Importante’ superasse 57%. Considerando que a amostra tem uma margem de erro de 7%, um valor abaixo de 57% poderia indicar que não há como garantir que aquela tarefa seria considerada importante pelo conjunto de servidores públicos federais de TI.

Para a prova de conceito, os pesos das tarefas foram definidos empiricamente na Seção 4.3. O maior peso inicialmente definido foi de valor 3, e o menor, 1, sendo todos números inteiros. Como uma forma de simplificar a comparação entre a proposição inicial e os ajustes derivados das respostas à pesquisa de opinião, essa estrutura será mantida para dois cenários de ajustes.

Partindo destas premissas, há várias formas de derivar os pesos a partir das frequências de importância coletadas. Três delas serão abordadas e comparadas.

### 5.3.1 Fórmula 1

A primeira forma de ajustar os pesos das tarefas é a de dividir a faixa de valores totais, de 57% a 100%, em três intervalos iguais e associá-los aos pesos de 1 a 3. Às tarefas com percentuais de importância menores que 71,33% será atribuído o peso 1. Entre 71,34% e 85,67% será atribuído o peso 2. Às tarefas com percentuais maiores que 85,68% será atribuído o peso 3. Seguindo esse método, os pesos ajustados das tarefas ficariam como exibidos na Tabela 5.4. Vê-se que, por essa forma de ajuste, na fase de Planejamento, apenas a tarefa ‘Selecionar cenário de migração’ teria o peso alterado, de 1 para 2. Na fase de Projeto, todas as tarefas tiveram seus pesos alterados, significando uma inversão entre os menores e maiores pesos.

Para mostrar o impacto do ajuste dos pesos, será novamente calculado o Indicador de Percepção de Risco e comparado com o valor apresentado no estudo de caso (Seção 4.5). Esses valores podem ser observados na Tabela 5.5, que mostra também a diferença percentual. Houve uma variação negativa de 4,73%. Por questões de simplicidade, a tabela mostra apenas o valor do Indicador para o Sistema B, que foi o sistema que apresentou a maior diferença percentual absoluta na aplicação das fórmulas descritas nesse capítulo.



Tabela 5.4: Fórmula 1 - Pesos ajustados das tarefas que compõem o indicador.

Fase	Tarefa	Peso proposto	Peso ajustado
Planejamento	Analisar custo da migração	3	3
	Identificar dependências	3	3
	Selecionar cenário de migração	1	2
Projeto	Contratar o provedor de nuvem	3	1
	Treinamento	1	3
	Identificar incompatibilidades	1	3

Tabela 5.5: Fórmula 1 - Mudança no valor do Indicador de Percepção de Risco do Sistema B (Seção 4.5), após ajustes nos pesos.

Indicador	Valor original	Valor ajustado	Diferença %
Fase de Planejamento (Pl)	3,14	3,25	3,30%
Fase de Projeto (Pr)	4,60	4,14	-11,03%
<b>Indicador de percepção de risco (IPR)</b>	<b>3,87</b>	<b>3,70</b>	<b>-4,73%</b>

### 5.3.2 Fórmula 2

A segunda forma de ajustar os pesos considera avaliar não apenas a soma das opções de resposta, mas também os valores individuais de ‘4-Importante’ e ‘5-Totalmente Importante’. A fórmula utilizada foi a de ponderar as opções de resposta do seguinte modo: o percentual relativo à opção ‘5-Totalmente Importante’ é multiplicado por 3, o percentual relativo à opção ‘4-Importante’ é multiplicado por 2 e o percentual restante, a parcela de respostas que não considerou a tarefa importante, é multiplicado por -1. Esses valores são somados e o resultado é arredondado para o inteiro mais próximo.

O menor valor possível para essa soma ocorreria caso apenas 57% dos respondentes considerasse a tarefa importante e todos eles houvessem escolhido a opção ‘4-Importante’. O valor para a soma seria 0,67, que seria arredondada para 1. O maior valor possível para a soma seria o 3, com todos os respondentes escolhendo a opção ‘5-Totalmente Importante’. Mas como o resultado é arredondado para o inteiro mais próximo, valores da soma maiores que 2,5 já resultarão no maior valor do peso possível, seguindo as premissas estabelecidas.

Seguindo esse método, os pesos ajustados das tarefas ficariam como exibidos na Tabela 5.6. Por essa forma de ajuste, todas as tarefas teriam seus pesos alterados. Na fase de Planejamento, as duas primeiras tarefas teriam seus pesos diminuídos de 3 para 2 e a tarefa ‘Selecionar cenário de migração’ teria o peso aumentado de 1 para 2. Na fase de Projeto, ‘Contratar o provedor de nuvem’ teria diminuído de 3 para 2, enquanto que as demais tarefas teriam seus pesos aumentados: ‘Treinamento’, de 1 para 2; ‘Identificar incompatibilidades’ de 1 para 3.

Tabela 5.6: Fórmula 2 - Pesos ajustados das tarefas que compõem o indicador.

<b>Fase</b>	<b>Tarefa</b>	<b>Peso proposto</b>	<b>Peso ajustado</b>
Planejamento	Analisar custo da migração	3	2
	Identificar dependências	3	2
	Selecionar cenário de migração	1	2
Projeto	Contratar o provedor de nuvem	3	2
	Treinamento	1	2
	Identificar incompatibilidades	1	3

O impacto do ajuste dos pesos segundo esse método pode ser visualizado no cálculo do Indicador de Percepção de Risco, mais uma vez comparado com o valor apresentado no estudo de caso para o Sistema B. Esses valores podem ser observados na Tabela 5.7. Houve uma variação de 0,25%.

Tabela 5.7: Fórmula 2 - Mudança no valor do Indicador de Percepção de Risco do Sistema B (Seção 4.5), após ajustes nos pesos.

<b>Indicador</b>	<b>Valor original</b>	<b>Valor ajustado</b>	<b>Diferença %</b>
Fase de Planejamento (Pl)	3,14	3,33	5,71%
Fase de Projeto (Pr)	4,60	4,43	-3,87%
<b>Indicador de percepção de risco (IPR)</b>	<b>3,87</b>	<b>3,88</b>	<b>0,25%</b>

### 5.3.3 Fórmula 3

Uma terceira forma de ajuste, desconsiderando que o peso seja um número inteiro entre 1 e 3, seria usar o próprio percentual de importância alcançado por cada tarefa como o peso a ser utilizado. Assim, os pesos das tarefas ficariam como listados na Tabela 5.8 e o impacto negativo de 1,93% do ajuste no indicador do Sistema B pode ser visualizado na Tabela 5.9.

Tabela 5.8: Fórmula 3 - Pesos ajustados das tarefas que compõem o indicador.

<b>Fase</b>	<b>Tarefa</b>	<b>Peso proposto</b>	<b>Peso ajustado</b>
Planejamento	Analisar custo da migração	3	93,26
	Identificar dependências	3	95,34
	Selecionar cenário de migração	1	79,79
Projeto	Contratar o provedor de nuvem	3	70,47
	Treinamento	1	90,67
	Identificar incompatibilidades	1	97,41

Tabela 5.9: Fórmula 3 - Mudança no valor do Indicador de Percepção de Risco do Sistema B (Seção 4.5), após ajustes nos pesos.

<b>Indicador</b>	<b>Valor original</b>	<b>Valor ajustado</b>	<b>Diferença %</b>
Fase de Planejamento (Pl)	3, 14	3, 30	4,68%
Fase de Projeto (Pr)	4, 60	4, 30	-7,01%
<b>Indicador de percepção de risco (IPR)</b>	<b>3,87</b>	<b>3,80</b>	<b>-1,93%</b>

### 5.3.4 Discussão

As três fórmulas de ajustes dos parâmetros do Indicador de Percepção de Risco, considerando os resultados da pesquisa de opinião, obtiveram resultados similares, todas apresentando uma variação inferior a 5% quando comparadas com os parâmetros estabelecidos empiricamente.

O impacto detalhado de cada uma das fórmulas foi apresentado apenas para o cálculo do indicador do Sistema B pelo fato de este sistema ter sido o que apresentou a maior diferença absoluta entre os três, considerando os ajustes realizados pelas três fórmulas analisadas. Aplicando os pesos conforme definido em cada fórmula e calculando os indicadores para os três sistemas, em nenhum caso houve mudança no ranking apresentado na Tabela 4.4.

A validação do Indicador de Percepção de Risco foi realizada e mostrou que a metodologia proposta tem aplicabilidade no contexto em que foi estudada.

Uma vez que os resultados entre as fórmulas apresentaram valores próximos, optou-se por utilizar a Fórmula 3, pois ela é mais simples, já que não foi necessário discretizar os valores, utilizando diretamente os percentuais de importância alcançados como resultados da pesquisa de opinião.

Considerando o impacto pequeno que causará, propõe-se, ainda, arredondar os valores dos pesos para números inteiros, para facilitar o seu uso. Assim, os valores definitivos para os pesos das tarefas do cálculo do Indicador de Percepção de Risco da migração de sistemas legados do governo para a Nuvem podem ser vistos na Tabela 5.10.

Tabela 5.10: Pesos ajustados das tarefas que compõem o indicador de percepção de risco.

<b>Fase</b>	<b>Tarefa</b>	<b>Peso</b>
Planejamento	Analisar custo da migração	93
	Identificar dependências	95
	Selecionar cenário de migração	80
Projeto	Contratar o provedor de nuvem	70
	Treinamento	91
	Identificar incompatibilidades	97

Considerando os pesos definidos na Tabela 5.10, o ranking atualizado da migração dos sistemas legados para a Nuvem é apresentado na Tabela 5.11, assim como o cálculo do Indicador de Percepção de risco para cada sistema do estudo de caso descrito na Seção 4.5. Como já informado anteriormente, o ajuste nos pesos não causou mudança na ordem da migração.

Tabela 5.11: Cálculo do Indicador de Percepção de Risco para cada sistema avaliado.

<b>Sistema</b>	<b>PI</b>	<b>Pr</b>	<b>IPR</b>	<b>Ranking</b>
Sistema C	4,30	4,29	<b>4,30</b>	<b>1</b>
Sistema A	4,35	3,67	<b>4,01</b>	<b>2</b>
Sistema B	3,30	4,29	<b>3,80</b>	<b>3</b>

## 5.4 Conclusão deste capítulo

Este capítulo apresentou uma análise experimental, baseada em pesquisa de opinião na Administração Pública Federal (APF) brasileira sobre migração de sistemas legados para a Nuvem. O objetivo da pesquisa era o de verificar como os profissionais de TI da APF percebiam a proposta feita nesse estudo: o uso do Modelo de Referência para apoiar as migrações de sistemas legados para a Nuvem e o cálculo e utilização do Indicador de Percepção de Risco (IPR).

O questionário foi construído segundo as boas práticas disponíveis na literatura [57] e os convites enviados por e-mail, com lembretes periódicos, pelo período de um mês. A coleta resultou em uma amostra de 193 respostas válidas, o que para uma população de 38114 profissionais[58], representa uma margem de erro de 7%.

A amostra teve a confiabilidade aferida pelo Alfa de Cronbach, com resultado de 0,7826, o que aponta que o questionário pode ser considerado confiável[53]. Pelas características da distribuição do conjunto de respostas, foi utilizado um teste estatístico não-paramétrico, o teste Signed Rank de Wilcoxon[54], que validou a hipótese alternativa de que a mediana das avaliações de importância seria maior do que 3, valor que representa a neutralidade.

Esse resultado aduz que os profissionais de TI do Governo Federal consideram importante o uso proposto do Modelo de Referência, assim como consideram importante cada uma das seis tarefas escolhidas para compor o IPR (Seção 4.3), itens principais da proposta feita por esse estudo.

No cálculo do IPR, como os pesos das tarefas haviam sido definidos de forma empírica na prova de conceito descrita no Capítulo 4, foram definidas três fórmulas para derivá-los do conjunto das respostas à pesquisa. A análise do impacto de cada fórmula no

resultado do IPR mostrou que a variação no valor do indicador ficou abaixo de 5% para os três sistemas analisados no estudo de caso apresentado na Seção 4.5, o que reforça a adequação da escolha das tarefas e dos pesos para o domínio do governo.

A fórmula mais simples foi a selecionada: a de aplicação direta da frequência de importância de cada tarefa como peso. Os novos pesos foram definidos a partir dela e o IPR foi recalculado para cada um dos três sistemas do estudo de caso. Não houve variação do ranking dos sistemas após o ajuste dos pesos.

Os resultados obtidos com a pesquisa de opinião apontam que a adoção da proposta feita por esse estudo pode ser de grande valor à APF, na medida em que poderá auxiliar na obtenção dos benefícios do uso de Computação em Nuvem pelos órgãos de governo.

# Capítulo 6

## Conclusões e trabalhos futuros

As organizações estão iniciando ou aumentando o uso de Computação em Nuvem como uma forma de ter acesso aos benefícios da plataforma [9]. No caso dos governos, há benefícios específicos, como uma potencial diminuição de oportunidades de desvios e irregularidades [14] e um incentivo à cultura empreendedora [18]. Mas a Computação em Nuvem também traz riscos potenciais relacionados a segurança, aprisionamento pelo provedor, entre outros [13].

Este trabalho apresentou um processo estruturado, um modelo de referência, para apoiar as migrações de sistemas legados do governo para a Nuvem. O modelo é composto por três fases, Planejamento, Projeto e Habilitação. Cada fase divide-se em tarefas e gera alguns artefatos de documentação. O modelo inclui, ainda, um Indicador de Percepção de Risco (IPR) da migração. O IPR, calculado de forma rápida e com pouco esforço, baseia-se nas informações coletadas pelo uso do processo de migração e na experiência da equipe responsável pela migração.

Ao avaliar a literatura sobre o tema, alguns modelos conceituais de migração foram encontrados e avaliados. O modelo de referência definido por Gholami et al. [5] foi criado a partir da consolidação de 78 trabalhos na área e foi validado por 104 especialistas em Computação em Nuvem, de 32 países. Em comparação com os outros modelos disponíveis, este modelo sobressaiu-se pela robustez apresentada e foi o selecionado por essa pesquisa para apoiar a migração dos sistemas legados do governo.

Para adicionar valor ao modelo, foi proposto um método de cálculo de um Indicador de Percepção de Risco. O método consiste na escolha de um conjunto de tarefas de cada uma das duas fases iniciais. As tarefas serão avaliadas pela equipe responsável pela migração, por meio de uma escala de intensidade da percepção de risco quanto à migração. Cada tarefa tem um peso associado e uma função definida calcula o valor do indicador para a migração do sistema legado para a Nuvem.

Com o IPR calculado para cada sistema, é possível ranquear os sistemas em ordem do

risco percebido quanto à migração. Migrar primeiro os sistemas cujos riscos percebidos sejam os mais baixos, pode trazer confiança à equipe e aumento da experiência da organização em relação à migração para a Nuvem [47], itens que auxiliarão nas migrações com maiores percepção de risco.

Uma prova de conceito para o domínio do governo foi descrita para mostrar a aplicabilidade do método de cálculo proposto para o IPR. As tarefas foram selecionadas e os pesos definidos empiricamente. Um estudo de caso foi implementado no domínio do governo: para três sistemas de uma organização, as duas primeiras fases do modelo de referência foram implementadas e os indicadores de percepção de risco foram calculados, criando um ranking para a execução da migração.

Uma análise experimental foi implementada por meio dos resultados de uma pesquisa de opinião nas organizações do governo federal brasileiro. O objetivo da pesquisa era verificar se a proposta deste trabalho – uso de um modelo de referência para apoio às migrações e cálculo do IPR – eram percebidos como sendo importantes entre os profissionais de TI do governo, que farão as migrações para a Nuvem em suas organizações.

Uma amostra de 193 profissionais analisou a proposta e, com percentuais expressivos, considerou importantes o uso do processo para apoio às migrações, o indicador proposto, assim como cada uma das tarefas selecionadas para o cálculo do indicador. A partir dos resultados dessa pesquisa, inédita na APF brasileira, os pesos para as tarefas foram ajustados, refletindo a importância relativa de cada tarefa na avaliação dos respondentes.

Os resultados obtidos com a pesquisa de opinião apontam que a adoção da proposta feita por esse estudo pode ser de grande valor à APF, na medida em que poderá auxiliar na obtenção dos benefícios do uso de Computação em Nuvem pelos órgãos de governo. A proposta feita por esse estudo também agrega valor à base de conhecimento em Computação em Nuvem.

## 6.1 Trabalhos Futuros

Dado que o método proposto foi validado, os trabalhos futuros consistirão em aplicá-lo em diferentes sistemas da APF e propiciar a elaboração de um guia de migração de sistemas legados para a Nuvem nesse domínio.

Propõe-se uma validação continuada do método para avaliar a necessidade de incorporar as alterações dos cenários de migrações de sistemas legados do governo para a Nuvem que ocorram ao longo do tempo.

# Referências

- [1] B. G. S. Costa, M. A. S. Reis, A. P. F. Araújo, and P. Solis, “Performance and Cost Analysis Between On-Demand and Preemptive Virtual Machines,” in *Proceedings of the 8th International Conference on Cloud Computing and Services Science - Volume 1: CLOSER*,. SciTePress, 2018, pp. 169–178. ix, 76
- [2] B. G. S. Costa and P. Solis, “Risk Perception of Migrating Legacy Systems to the Cloud,” in *8th International Conference on Cloud Computing and Services Science*, May 2018, pp. 634–642. x, 87
- [3] P. Jamshidi, A. Ahmad, and C. Pahl, “Cloud migration research: a systematic review,” *IEEE Transactions on Cloud Computing*, vol. 1, no. 2, pp. 142–157, 2013. xi, 7, 13, 14, 16, 17
- [4] M. F. Gholami, F. Daneshgar, G. Low, and G. Beydoun, “Cloud migration process—A survey, evaluation framework, and open challenges,” *Journal of Systems and Software*, vol. 120, pp. 31–69, 2016. xi, xiii, 15, 16, 17, 29
- [5] M. F. Gholami, F. Daneshgar, G. Beydoun, and F. Rabhi, “Challenges in migrating legacy software systems to the cloud—an empirical study,” *Information Systems*, vol. 67, pp. 100–113, 2017. xi, 17, 18, 19, 20, 21, 24, 28, 29, 30, 46, 56
- [6] PMI - Project Management Institute, “PMBOK<sup>®</sup> Guide – Sixth Edition,” 2017. [Online]. Available: <https://www.pmi.org/pmbok-guide-standards/foundational/pmbok/sixth-edition> xi, 8, 22, 23
- [7] C. Schmitz and others, “LimeSurvey: An open source survey tool,” *LimeSurvey Project Hamburg, Germany*. URL <http://www.limesurvey.org>, 2012. xi, 36, 41, 42
- [8] ISACA, “Guiding principles for cloud computing adoption and use,” 2012. 1
- [9] Gartner, “Developing a Public Cloud IaaS Adoption and Migration Framework,” Mar. 2017. [Online]. Available: <https://www.gartner.com/doc/3645347/developing-public-cloud-iaas-adoption> 1, 5, 7, 10, 13, 56
- [10] Amazon, “AWS - Amazon Web Services,” 2017. [Online]. Available: <https://aws.amazon.com/> 1
- [11] Microsoft, “Azure,” 2017. [Online]. Available: <http://www.azure.com> 1



- [12] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and others, “Above the clouds: A berkeley view of cloud computing,” Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, Tech. Rep., 2009. 1, 13
- [13] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and others, “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010. 1, 2, 13, 46, 56
- [14] Tribunal de Contas da União, “Acórdão 1739/2015-Plenário,” Jul. 2015. [Online]. Available: [http://www.tcu.gov.br/Consultas/Juris/Docs/judoc/Acord/20150720/AC\\_1739\\_24\\_15\\_P.doc](http://www.tcu.gov.br/Consultas/Juris/Docs/judoc/Acord/20150720/AC_1739_24_15_P.doc) 2, 3, 5, 10, 46, 56
- [15] MPOG, “Boas práticas, orientações e vedações para contratação de Serviços de Computação em Nuvem,” May 2016. [Online]. Available: <https://www.governoeletronico.gov.br/documentos-e-arquivos/Orientacaoservicosemnuvem.pdf> 2
- [16] C. Pahl and H. Xiong, “Migration to PaaS clouds-migration process and architectural concerns,” in *Maintenance and Evolution of Service-Oriented and Cloud-Based Systems (MESOCA)*, 2013, pp. 86–91. 2, 7
- [17] P. Mell, T. Grance, and others, “The NIST definition of cloud computing,” 2011. 4, 10
- [18] V. Kundra, “Federal cloud computing strategy,” 2011. 6, 10, 46, 56
- [19] A. Dedek, “Improving legacy-system sustainability: A systematic approach,” *It Professional, IEEE*, vol. 14, no. 1, pp. 38–43, 2012. 6, 46
- [20] J.-F. Zhao and J.-T. Zhou, “Strategies and methods for cloud migration,” *international Journal of Automation and Computing*, vol. 11, no. 2, pp. 143–152, 2014. 7, 8, 12
- [21] A. D. Ionita, “Introduction to the migration from legacy applications to service provisioning,” *Migrating Legacy Applications: Challenges in Service Oriented Architecture and Cloud Computing Environments*, pp. 1–11, 2013. 7
- [22] ISO/IEC, *Systems and Software Engineering- Software Life Cycle Processes - Maintenance, ISO/IEC 14764:2006*, 2006. 7
- [23] K. Sun and Y. Li, “Effort estimation in cloud migration process,” in *Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on*. IEEE, 2013, pp. 84–91. 7, 13
- [24] M. A. Babar and M. A. Chauhan, “A tale of migration to cloud computing for sharing experiences and observations,” in *Proceedings of the 2nd international workshop on software engineering for cloud computing*. ACM, 2011, pp. 50–56. 8
- [25] ABNT - Associação Brasileira de Normas Técnicas, *Nbr iso 31000:2009 - gestão de riscos - princípios e diretrizes*, 2009. 8

- [26] C. Alberts, A. Dorofee, J. Stevens, and C. Woody, “Introduction to the OCTAVE Approach,” CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST, Tech. Rep., 2003. 8
- [27] A. Amini and N. Jamil, “A Comprehensive Review of Existing Risk Assessment Models in Cloud Computing,” *Journal of Physics: Conference Series*, vol. 1018, no. 1, p. 012004, 2018. 8
- [28] S. Islam, S. Fenz, E. Weippl, and H. Mouratidis, “A Risk Management Framework for Cloud Migration Decision Support,” *Journal of Risk and Financial Management*, vol. 10, no. 2, 2017. 8
- [29] G. Stone and P. Noel, “Cloud Risk Decision Framework,” Technical report. Microsoft. [http://download.microsoft.com/documents/australia/enterprise/SMIC1545\\_PDF\\_v7\\_pdf.pdf](http://download.microsoft.com/documents/australia/enterprise/SMIC1545_PDF_v7_pdf.pdf). Accessed 13 Aug, Tech. Rep., 2015. 8
- [30] B. Rohrmann, “Risk perception, risk attitude, risk communication, risk management: A conceptual appraisal,” in *Conferencia presentada en la Sociedad Internacional de Gerenciamiento de Emergencias*, 2008. 8
- [31] L. d. S. Guimarães, “Síntese de doutrina de segurança para projeto e operação de submarinos nucleares.” PhD Thesis, Universidade de São Paulo, 1999. 8
- [32] A. Dutta, G. C. A. Peng, and A. Choudhary, “Risks in enterprise cloud computing: the perspective of IT experts,” *Journal of Computer Information Systems*, vol. 53, no. 4, pp. 39–48, 2013. 8
- [33] S. Cicoria, *Qualitative Analysis of Application Migration to the Public Cloud and Framework for Application Architecture Migration Risks*. Pace University, 2017. 9, 29
- [34] L. Sjöberg, “Factors in risk perception,” *Risk analysis*, vol. 20, no. 1, pp. 1–12, 2000. 9
- [35] P. Fettke and P. Loos, “Multiperspective evaluation of reference models—towards a framework,” *Conceptual modeling for novel application domains*, pp. 80–91, 2003. 9, 10, 19
- [36] T. Binz, F. Leymann, and D. Schumm, “CMotion: A Framework for Migration of Applications into and between Clouds,” in *Service-Oriented Computing and Applications (SOCA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–4. 11, 12
- [37] A. Desai, R. Oza, P. Sharma, and B. Patel, “Hypervisor: A survey on concepts and taxonomy,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 2, no. 3, pp. 222–225, 2013. 11
- [38] V. Andrikopoulos, T. Binz, F. Leymann, and S. Strauch, “How to adapt applications for the cloud environment,” *Computing*, vol. 95, no. 6, pp. 493–535, 2013. 12, 29

- [39] MySQL, “MySQL - O mais popular banco de dados de código aberto,” 2018. [Online]. Available: <https://www.mysql.com> 12
- [40] Google, “Google Cloud SQL,” 2018. [Online]. Available: <https://cloud.google.com/sql> 12
- [41] L. H. Putnam, “A General Empirical Solution to the Macro software Sizing and Estimating Problem,” *IEEE Transactions on Software Engineering*, vol. SE-4, no. 4, pp. 345–361, 1978. 13
- [42] N. S. de Moraes, A. C. Drummond, and A. P. F. de Araújo, “Proposal for a virtualized data center migration model for private cloud model for the new Brazilian army information technology pole,” in *Information Systems and Technologies (CISTI), 2017 12th Iberian Conference on*. IEEE, 2017, pp. 1–7. 13
- [43] P.-J. Maenhaut, H. Moens, V. Ongenaes, and F. De Turck, “Migrating legacy software to the cloud: approach and verification by means of two medical software use cases,” *Software: Practice and Experience*, vol. 46, no. 1, pp. 31–54, 2016. 13
- [44] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2013. 17
- [45] A. Schwarz, M. Mehta, N. Johnson, and W. W. Chin, “Understanding frameworks and reviews: a commentary to assist us in moving our field forward by analyzing our past,” *ACM SIGMIS Database*, vol. 38, no. 3, pp. 29–50, 2007. 18
- [46] J. Hartley, “Some thoughts on Likert-type scales,” *International Journal of Clinical and Health Psychology*, vol. 14, no. 1, 2014. 25
- [47] H. Reza Bazi, A. Hassanzadeh, and A. Moeini, “A comprehensive framework for cloud computing migration using Meta-synthesis approach,” *Journal of Systems and Software*, vol. 128, pp. 87–105, 2017. 27, 29, 30, 47, 57
- [48] Briggs, Barry and Kassner, Eduardo, *Enterprise Cloud Strategy*. Redmond, WA: Microsoft Press, 2016. 29
- [49] Oracle, *Oracle Database.*, 2017. [Online]. Available: <https://www.oracle.com/database/index.html> 32
- [50] Apache, “Apache Solr -,” 2018. [Online]. Available: <http://lucene.apache.org/solr/> 32
- [51] Moodle, “Open-source learning platform,” 2018. [Online]. Available: <https://moodle.org/> 33
- [52] Microsoft, “Active Directory Domain Services,” 2018. [Online]. Available: <https://docs.microsoft.com/en-us/windows/desktop/ad/active-directory-domain-services> 33
- [53] H. R. M. da Hora, G. T. R. Monteiro, and J. Arica, “Confiabilidade em questionários para qualidade: um estudo com o Coeficiente Alfa de Cronbach,” *Produto & Produção*, vol. 11, no. 2, pp. 85–103, 2010. 35, 43, 54

- [54] F. Wilcoxon, “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. 35, 45, 54
- [55] E. A. Coelho Barros and J. Mazucheli, “Um estudo sobre o tamanho e poder dos testes t-Student e Wilcoxon,” *Acta Scientiarum. Technology*, vol. 27, no. 1, 2005. 35
- [56] A. Pinsonneault and K. Kraemer, “Survey research methodology in management information systems: an assessment,” *Journal of management information systems*, vol. 10, no. 2, pp. 75–105, 1993. 36
- [57] H. Günther, “Como elaborar um questionário (série: Planejamento de Pesquisa nas Ciências Sociais, nº 01),” *Brasília, DF: UnB, Laboratório de Psicologia Ambiental*, 2003. 36, 37, 54
- [58] Tribunal de Contas da União, “Levantamento de pessoal de TI,” 2015. [Online]. Available: <https://portal.tcu.gov.br/biblioteca-digital/levantamento-de-pessoal-de-ti-sumario.htm> 40, 54
- [59] H. BRASIL, “HDI Brasil: Capacitação de profissionais de TI em serviços,” 2018. [Online]. Available: <http://www.hdibrasil.com.br/> 41
- [60] IBGP, “Instituto Brasileiro de Governança Pública - IBGP,” 2018. [Online]. Available: <https://www.ibgp.net.br/> 41
- [61] M. Bech and M. B. Kristensen, “Differential response rates in postal and Web-based surveys in older respondents,” in *Survey Research Methods*, vol. 3, 2009, pp. 1–6. 42
- [62] E. Dannetun, A. Tegnell, and J. Giesecke, “Parents’ attitudes towards hepatitis B vaccination for their children. A survey comparing paper and web questionnaires, Sweden 2005,” *BMC public health*, vol. 7, no. 1, p. 86, 2007. 42
- [63] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham, *Análise multivariada de dados*. Bookman Editora, 2009. 43
- [64] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2016. [Online]. Available: <https://www.R-project.org/> 43, 45
- [65] S. Siegel, “Nonparametric statistics,” *The American Statistician*, vol. 11, no. 3, pp. 13–19, 1957. 45

# Anexo I

## Convite enviado por e-mail

# Pesquisa sobre Computação em Nuvem na Administração Pública

Breno Gustavo Soares da Costa

qua 11/07/2018 20:06

Para: Breno Gustavo Soares da Costa <brenogc@tcu.gov.br>;

## **Caro Colega Servidor Público,**

estou cursando o Mestrado na Universidade de Brasília - UnB - e desenvolvendo pesquisa na área de Computação em Nuvem. A pesquisa propõe o uso de um processo específico como apoio à migração de sistemas para a Nuvem na Administração Pública.

**A sua opinião sobre as linhas gerais dessa proposta é muito importante**, pois proporcionará um melhor entendimento do que pensam os Servidores e permitirá um ajuste na proposta que seja mais adequado à situação atual das organizações e das equipes de governo.

**Peço que contribua com essa pesquisa**, respondendo ao questionário a seguir:

<http://contas.tcu.gov.br/limesurvey/index.php/754816/lang-pt-BR>

Todas as suas declarações serão tratadas de maneira confidencial.

O objetivo é coletar as opiniões das pessoas e não de ter um posicionamento oficial das organizações, assim vários servidores de uma mesma organização podem responder ao questionário. **Peço-lhe encarecidamente que encaminhe esse e-mail a outros Colegas para que também possam contribuir com a pesquisa.**

Obrigado pelo apoio!

Atenciosamente,

Breno Costa

## Anexo II

# Questionário completo da Pesquisa de Opinião

## Pesquisa de opinião sobre migração de sistemas da Administração Pública para a Nuvem

Este questionário foi criado com a intenção de coletar as opiniões dos colegas, servidores e empregados públicos, acerca da adoção de Computação em Nuvem na Administração Pública brasileira. Ele faz parte de pesquisa acadêmica que estou desenvolvendo sobre o tema na Universidade de Brasília - UnB.

O tempo para respondê-lo é inferior a 10 minutos. **O objetivo é coletar as opiniões das pessoas e não de ter um posicionamento oficial das organizações, assim vários servidores de uma mesma organização podem respondê-lo.**

Todas as suas declarações serão tratadas de maneira confidencial. Os resultados serão apresentados de maneira a não permitir a identificação de participantes individuais.

Se você tiver dúvidas ou quiser relatar um problema relacionado à pesquisa, entre em contato pelo e-mail [brenogc@tcu.gov.br](mailto:brenogc@tcu.gov.br), ou pelo celular/whatsApp 61 98401 1080. Tentarei responder da melhor maneira possível qualquer dúvida que haja.

Peço-lhe a gentileza de contribuir com esta pesquisa. Suas opiniões e experiências são muito importantes!

Atenciosamente,

Breno Costa

Sair e apagar o questionário

Próximo



## Pesquisa de opinião sobre migração de sistemas da Administração Pública para a Nuvem

0%  100%

Este grupo de perguntas tentará captar suas percepções e opiniões acerca da importância de alguns recursos de apoio à migração de sistemas legados para Nuvem.

\* 1 Qual o nível de importância que você atribui ao uso de um processo estruturado (com fases, atividades e documentos definidos) para apoiar a migração de sistemas legados para a nuvem? Use uma escala de 1 a 5, conforme opções a seguir:

Escolha uma das seguintes respostas:

- 5 - Totalmente importante
- 4 - Importante
- 3 - Neutro
- 2 - Sem importância
- 1 - Totalmente sem importância

Como parte da proposta que será feita na pesquisa acadêmica, estará a indicação de um processo com três fases, cada fase composta por atividades, cada atividade com um objetivo de coleta de informação, decisão técnica, ou mesmo documentação do sistema e dos serviços em Nuvem que o implementarão após a migração. Caso queira mais detalhes sobre as fases e atividades, clique [aqui](#).

\* 2 Na migração de um sistema legado para a Nuvem, pode ser atribuído um score que indique a percepção do grau de risco nesse processo. A metodologia do cálculo do score está descrita a seguir. Na sua visão, a utilização desse score conforme essa metodologia é representativa como indicador?

- Sim
- Não

O cálculo de um valor (score) para o risco percebido pela equipe na migração de um sistema legado para a Nuvem utiliza as informações geradas por seis tarefas selecionadas nas duas primeiras fases do processo apresentado na pergunta anterior. Pontua-se cada tarefa em uma escala de 1 a 5 (1=Risco Muito Baixo e 5=Risco Muito Alto). Com esses valores calcula-se a média e obtém-se um score da percepção de risco da equipe para esta migração.

\* 3 Você considera uma boa prática iniciar o processo de migração para a Nuvem por um sistema legado que apresente percepção de risco baixo (score), deixando para momentos posteriores os sistemas com percepção de risco alto?

- Sim
- Não

No contexto de vários sistemas a serem migrados, o score de percepção de risco seria utilizado para ordenar as migrações, iniciando pelas que têm menor percepção de risco.

\* 4 Qual o nível de importância que você atribui a cada uma das seguintes tarefas que subsidiarão o cálculo da percepção de risco (score) da migração de um sistema legado para a nuvem? Use a escala de valores de 1 a 5, conforme opções a seguir:

	5 - Totalmente importante	4 - Importante	3 - Neutro	2 - Sem importância	1 - Totalmente sem importância
Analisar custo da migração	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identificar dependências	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Selecionar cenário de migração	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contratar o provedor de nuvem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Treinar equipe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identificar incompatibilidades	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

As tarefas são subdivisões das atividades do processo de migração apresentado na Pergunta 1. Elas foram escolhidas como sendo as que melhor apontam os riscos que cada migração de sistema legado para a nuvem pode apresentar.

Mais detalhes sobre cada atividade selecionada podem ser encontrados [aqui](#).

7 Caso queira comentar algo sobre as questões anteriores, ou sobre experiências específicas acerca da adoção de Computação em Nuvem, utilize o espaço abaixo! Caso deseje responder à próxima pergunta, basta clicar no botão Próximo no fim desta página.

O uso desse campo é opcional. Caso tenha comentários ou relatos a fazer, nós gostaríamos de ouvi-lo. Suas opiniões e experiências serão de grande valia na análise da adoção de Computação em Nuvem pela Administração Pública.

Sair e apagar o questionário

Anterior

Próximo

## Pesquisa de opinião sobre migração de sistemas da Administração Pública para a Nuvem

0%  100%

Concluindo, gostaríamos de fazer mais algumas perguntas para melhor caracterizar os respondentes desta pesquisa e as organizações em que trabalham.

8 Qual o nome da organização em que você trabalha?

\* 9 Quão importante é a Computação em Nuvem Pública para sua Organização? Use a escala de 1 a 5, conforme opções a seguir:

Escolha uma das seguintes respostas:

- 1 - Totalmente sem importância
- 2 - Sem importância
- 3 - Neutra
- 4 - Importante
- 5 - Totalmente importante

\* 10 Qual o uso atual de computação em nuvem na sua organização?

Escolha uma das seguintes respostas:

- Não há uso de computação em nuvem
- Há apenas uso experimental (testes efetuados ou em andamento, mas nada em produção)
- Há pouco uso (por ex.: serviço de e-mail na nuvem, pacotes de escritório ou outros softwares "de prateleira" contratados como serviço - SAAS)
- Há uso moderado (por ex.: provimento de infraestrutura (IAAS) ou de plataforma (PAAS) na nuvem, sistemas corporativos não-críticos em produção)
- Há uso intenso (por ex.: sistemas corporativos críticos em produção)

\* 11 Qual a sua expectativa de prazo para que a organização em que você trabalha faça uso moderado ou intenso de Computação em Nuvem, de acordo com a definição nos itens da questão anterior:

Escolha uma das seguintes respostas:

- Não ocorrerá nos próximos 5 anos
- Ocorrerá entre 3 e 5 anos
- Ocorrerá entre 2 e 3 anos
- Ocorrerá entre 1 e 2 anos
- Ocorrerá nos próximos 12 meses

\* 12 Em qual domínio/indústria a organização em que você trabalha atua?

Escolha uma das seguintes respostas:

- Governo
- Telecomunicações
- TI – Fabricante
- TI – Prestador de serviço
- TI – Provedor de nuvem
- Outro

\* 13 Quantos empregados trabalham na sua organização?

Escolha uma das seguintes respostas:

- 1-99
- 100-499
- 500-999
- 1000-4999
- 5000 ou mais

\* 14 Quantos anos de experiência profissional com TI você tem?

Escolha uma das seguintes respostas:

- Nenhuma experiência
- Menos de 5 anos
- 5 a 15 anos
- 16 a 25 anos
- Mais de 25 anos

\* 15 Quantos anos de experiência em qualquer tipo de nuvem (pública, privada ou híbrida) você tem?

Escolha uma das seguintes respostas:

- Nenhuma experiência
- Menos de 1 ano
- 1 a 2 anos
- 3 a 5 anos
- Mais de 5 anos


\* 16 Avalie seu nível de experiência com as seguintes áreas:

	Não-familiarizado	Familiarizado	Iniciante	Bem informado	Especialista / Experiente
Computação em Nuvem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nuvem Pública	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nuvem Privada	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nuvem Híbrida	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

\* 17 Avalie seu nível de experiência com os seguintes provedores de nuvem:

	Não-familiarizado	Familiarizado	Iniciante	Bem informado	Especialista / Experiente
AWS - Amazon Web Services	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Microsoft Azure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Cloud Platform	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Alibaba	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
IBM Softlayer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>


18 Caso deseje fazer algum comentário adicional ou detalhar algo que não foi possível por meio das opções disponíveis para resposta nas questões anteriores, por favor, faça-o no campo a seguir.

 Seus comentários, experiências e preocupações acerca da adoção de Computação em Nuvem acrescentarão muito valor a esse estudo!

19 Caso queira receber o resultado da pesquisa assim que estiver disponível, pedimos informar nome e e-mail nos campos a seguir.

Nome:

E-mail:

 Os resultados serão também publicados como parte da dissertação de mestrado que defenderei em 2018 acerca da migração de sistemas legados do governo para a Nuvem

Sair e apagar o questionário

◀ Anterior

Enviar

Obrigado por ter participado!

Qualquer dúvida, comentário ou sugestão, contate-me pelo e-mail [brenogo@tcu.gov.br](mailto:brenogo@tcu.gov.br) ou pelo celular/whatsApp 61 98401 1080.

### Modelo de referência para migração de sistemas legados para a Nuvem

Como parte do trabalho de pesquisa que empreendemos, avaliamos diversos modelos de processo para apoiar e estruturar a migração de sistemas legados para a Nuvem. Dentre eles, selecionamos o modelo de referência definido por Gholami *et al.*[1], resumido na Figura 1.

Ele possui três fases: Planejamento, Projeto e Habilitação. Cada fase é composta por algumas atividades, que dividem e especializam o trabalho e que coletam informações e registram decisões. As atividades subdividem-se ainda em tarefas, mas optamos por apresentar a seguir uma visão mais resumida do processo.

As fases de Planejamento e Projeto têm trabalho predominantemente analítico. O esforço empreendido nessas duas fases juntas é potencialmente menor que o que é empreendido na fase de Habilitação, momento em que se concentram a implementação das alterações no sistema legado, caso haja, assim como a configuração e testes dos serviços de computação em nuvem no(s) provedor(es).

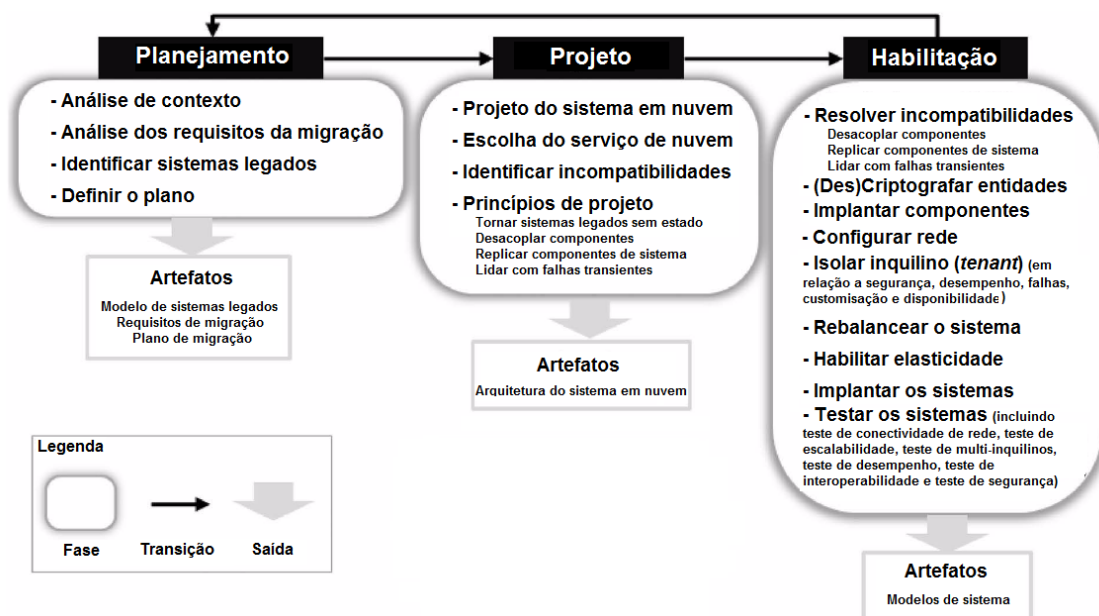


Figura 1 – Modelo de Referência de migração de sistemas legados para a nuvem proposto por Gholami *et al.*[1]

[1] M. F. Gholami, F. Daneshgar, G. Beydoun, and F. Rabhi, "Challenges in migrating legacy software systems to the cloud—an empirical study," *Information Systems*, vol. 67, pp. 100–113, 2017.

## Tarefas selecionadas para subsidiar o cálculo do score de risco percebido pela equipe

Para cada uma das fases de Planejamento e Projeto foram selecionadas três tarefas, cujas informações geradas subsidiarão o cálculo da percepção de risco da migração do sistema legado em análise.

### Fase de Planejamento

**Analisar custo da migração** – Essa tarefa tem como resultado uma estimativa do custo da migração. O custo está associado com esforço necessário a tornar o legado em serviços de computação em nuvem, assim como com o custo que a aplicação terá após migração.

Quanto maior o custo de migração de um sistema, maior o risco para todo o processo de migração, caso a migração tenha problemas (atrasos, diminuição de funcionalidades ou de desempenho). Por outro lado, um sistema cuja migração tenha baixo custo pode ser um bom candidato para validar o processo de migração e para se ganhar experiência tanto com o processo quanto com o provedor para o qual o sistema será migrado.

**Identificar dependências** – Essa tarefa tem como resultado identificar quais outros sistemas e componentes da infraestrutura de hardware e software local o sistema legado depende para funcionar após a migração. Com esse levantamento, é possível identificar quais outros sistemas teriam de ser migrados em conjunto ou previamente ao sistema que está sendo analisado. Ou ainda, em caso de nuvem híbrida e caso as dependências não venham a ser migradas, será preciso avaliar e considerar a latência de rede entre a estrutura local e a estrutura na nuvem.

**Selecionar cenário de migração** – De acordo com as características do sistema legado e também de acordo com a quantidade de esforço que se pretende seja dispendida na migração, escolhe-se o cenário de migração, dentre algumas opções presentes na literatura:

1. Substituição de componentes da arquitetura do sistema legado por serviço da nuvem;
2. Substituição do sistema legado por solução da Nuvem oferecida como SAAS – Software as a Service.
3. Transposição de toda a pilha de recursos do sistema legado encapsulado em máquinas virtuais que executarão na nuvem;
4. *Cloudificar* o sistema legado – reescrever o sistema para que se utilize apenas de serviços na nuvem e seguindo a filosofia de serviços com fraco acoplamento e outras características intrínsecas à Nuvem.

O cenário de migração do tipo 3. acima, também conhecido como Suspende e Desloca (*Lift and Shift*), é o que menos esforço demanda e que tem o menor risco de falha, ou de problemas da migração, quando comparado com os demais. Assim, a escolha do cenário pode estar diretamente relacionada com a intensidade de risco que se aceita. Isso não quer dizer que um cenário de menor risco seja também a opção que trará o maior benefício (redução de custos, aumento da disponibilidade), pois o cenário 3., por exemplo, traz consigo uma maior dificuldade na implementação da elasticidade.

### Fase de Projeto

**Contratar o provedor de nuvem** – Após a execução dessa tarefa tem-se um ou mais provedores de nuvens disponíveis para fornecer os serviços que suportarão os sistemas a serem migrados. Caso ainda não tenha havido a contratação do provedor no momento da migração de um



determinado sistema legado, isso pode postergar a migração ou causar retrabalho. Veja que, mesmo após um provedor já ter sido contratado e alguns sistemas legados terem sido migrados para ele, é possível que o contrato com esse provedor esteja acabando, ou mesmo que a prestação do serviço não esteja satisfatória e que a organização tenha decidido substituí-lo antes do fim do contrato.

**Treinar equipe** – Essa tarefa tem o objetivo de introduzir e manter, na organização, o conhecimento necessário ao projeto, operação e acompanhamento de serviços de computação em nuvem. Uma necessidade maior de treinamento pode indicar um risco maior na gestão do projeto de migração de um determinado sistema legado. Nas primeiras migrações, o conhecimento acerca de Computação em Nuvem pode estar deficitário na organização. Com o início do uso de serviços em nuvem em um determinado provedor, quando ele é recém-contratado, há uma maior necessidade de treinamento também.

Mesmo em momentos de estabilidade em relação à contratação de novos provedores, uma mudança na equipe que gerencia o processo de migração pode demandar mais treinamento. Além disso, há a atualização dos serviços já em uso, assim como oferecimento de novos serviços de computação em nuvem pelo provedor.

**Identificar incompatibilidades** – Essa atividade tem como objetivo listar as incompatibilidades entre o sistema legado, da forma como ele é executado atualmente, e o conjunto de serviços de computação em nuvem definidos como sendo necessários à execução do sistema no provedor após a migração. Esse conjunto de incompatibilidades demandará esforço específico para serem resolvidas. Quanto maior o conjunto, ou quanto maior o esforço necessário para sanar as incompatibilidades, pode ser maior o risco da migração específica.

## **Anexo III**

**Artigo publicado no congresso  
CLOSER 2018 - Ilha da Madeira -  
Portugal [1]**

# Performance and Cost Analysis Between On-Demand and Preemptive Virtual Machines

Breno G. S. Costa, Marco Antonio Sousa Reis, Aletéia P. F. Araújo and Priscila Solis  
*Department of Computer Science, University of Brasilia, Brasilia/DF, Brazil*

**Keywords:** Cloud Computing, Transient Server, Performance Analysis, Preemptive Machines.

**Abstract:** A few years ago, Amazon Web Services introduced spot instances, transient servers that can be contracted at a significant discount over regular price, but whose availability depends on cloud provider criteria and the instance can be revoked at any time. Google Cloud Platform offers preemptive instances, transient servers that have similar behavior and discount level to spot instances. Both providers advertise that their transient servers have the same performance level as servers contracted on-demand. Even with the possibility of revocation at the provider's discretion, some applications can benefit from the low prices charged by these servers. But the measured performance of both models, transient and on-demand, must be similar, and the applications must survive occasional or mass server revoking. This work compares the performance and costs of transient and on-demand servers from both providers. Results show there is no significant difference in performance measured, but there is real cost advantage using transient servers. On Amazon Web Services a MapReduce cluster composed of transient servers achieved a 68% discount when compared to the same cluster based on on-demand servers. On Google Cloud Platform, the discount achieved was 26% but it can be bigger when the clusters are larger.

## 1 INTRODUCTION

Cloud providers have introduced a new class of servers, called transient servers, which they can unilaterally revoke at any time (Singh et al., 2014). Transient servers increase the utilization of a cloud provider's infrastructure while enabling it to retrieve resources at any time to lease them to higher priority users.

Due to their preemptive nature, transient servers are not suitable for running interactive systems such as web services, or any system that does not tolerate downtime caused by server revocations. Cloud providers typically provide a brief early warning before revoking a transient server to allow the customer to shut it down properly. Batch-oriented interrupt-tolerant applications are particularly suitable for transient servers, as they can tolerate longer completion times caused by occasional inactivity. A common scenario is to use tens or hundreds of transient servers to run highly CPU-intensive or data-intensive systems at lower costs (compared to regular server prices contracted on-demand).

Different cloud providers have different pricing models for transient servers. The Google Cloud Plat-

form (GCP) transient servers, named preemptive instances (Google, 2017), have a fixed discount of about 80%, a maximum lifetime of 24 hours (with the possibility of preemption within lifetime) and with an alert of revocation of only 30 seconds.

Microsoft Azure recently announced that its transient server offering, called Low-priority virtual machine (VM), changed status from public preview to general availability. These VMs can only be used within a specific service called Batch (Microsoft, 2017). Their pricing structure is similar to GCP, with a fixed price and a discount of up to 80%. The Batch service re-queues a task when a low-priority VM that is executing the task is revoked. According to the announcement, preview pricing will be in effect for a few months, but it will move to regional pricing — the same as on-demand VMs — and this might cause a slight increase in pricing, depending on the region.

In contrast, Amazon Web Services (AWS) spot instances (SI) (AWS, 2017c) offer a variable discount. The price of SIs varies continuously based on market supply and demand for each type of server. The customer specifies a maximum price (a bid) that he is willing to pay when ordering SIs. AWS, based on the proposals submitted and following market criteria,

determines a market price for the server (Agmon Ben-Yehuda et al., 2013). If the market price increases and stays above the bid, the server will be revoked, but only after the two-minute notice period. While the SI price remains below the bid, the SI remains available and the customer will pay only the market price, even if his bid is higher.

Figure 1 shows the market price variation, in a one month period, of an M4.2xlarge type SI on the us-east-1e zone. Each availability zone is a different market and prices can be different. In the time frame, the SI price reached the on-demand price (US\$ 0.40) less than 10 times and for short periods.

Due to the high probability of changing prices and even the behavior of their transient servers in the short term, Microsoft Azure Low-priority VMs will be excluded from this analysis, but they will be included as future work as soon as the offer becomes stabilized.

### 1.1 Availability of Transient Servers

The availability of transient servers (in terms of average revocation time) can also vary significantly across server configurations and on the basis of changing market conditions. Unfortunately, cloud platforms do not directly expose the availability statistics of transient servers, requiring users to infer them indirectly, for example, through price history. Thus, it is challenging for a cloud system to select the most appropriate server configuration based on historical price or availability data to meet its needs. Recent research suggests that mitigating the risk of revocation requires a parallelized system to diversify its resource needs across various types of transient servers, further complicating decision making (Sharma et al., 2016).

The problem is exacerbated by the large number of transient server choices available from providers: there are over 2500 SI options in AWS Elastic Cloud Computing (EC2) and more than 300 GCP preemptive instances. This is because each availability zone has its own market value calculation for each available virtual machine configuration.

According to (Sharma et al., 2017), choosing a server configuration based only on price can produce sub-optimal results. The authors cite an example where server configurations with very low prices can also see greater market demand and consequently higher price volatility and more revocations. Frequent revocations generate additional verification, checkpointing, and system recovery efforts. Instead, they suggest that choosing a slightly more expensive server configuration and having a lower revocation rate can produce lower overall costs.

Due to the challenges listed, cloud providers such

as AWS have begun offering server selection tools. Amazon SpotFleet (AWS, 2015a) automatically replaces revoked servers. However, SpotFleet has a limited choice in terms of the combinations of server configurations that it offers and does not solve some of the challenges presented. Another tool, Amazon Spot Bid Advisor (AWS, 2015b), can help users select servers based on price, but exposes only superficial volatility information such as low, medium, or high categorization.

An important consideration is that if transient server performance was lower than the on-demand server, and adding up the additional complexity of dealing with revocations, the large discounts offered by the providers would not be worthwhile. Therefore, this study measures the performance of transient servers using benchmarking software and compares them to on-demand servers to verify whether the cost decrease advertised by cloud providers is also accompanied by a performance decrease. In addition, a scenario in which the use of a transient server is viable will be implemented and the costs of execution in the two server classes will be compared on both providers that offer them.

The remainder of this article is divided into six sections. Section 2 presents some related work. Section 3 describes the experiments environment. Section 4 describes the planning and the results of the experiments. Section 5 performs an analysis of the performance experiments. Section 6 presents a cost comparison of a MapReduce workload running on both classes of servers, and in Section 7 conclusion and future work are presented.

## 2 RELATED WORK

There are, in literature, several studies regarding spot instances. Many of them try to predict SI prices and find an optimal bid on the the spot market. The strategies undertaken by these researches are diverse. Time series forecasting is used by (Chhetri et al., 2017), whose results, using three specific metrics, show that successful estimation of bid prices in AWS spot markets is an implicit function of seasonal components and extreme spikes in the spot price history. Another study (Khandelwal et al., 2017) uses Regression Random Forests (RRFs) to predict spot prices. The authors use a one year trace of spot market prices and compare the results achieved by RRFs with existing non-parametric machine learning models. The paper reveal that RRF-based forecast accuracy outperforms other models.

In (Wolski and Brevik, 2016) a method is pro-

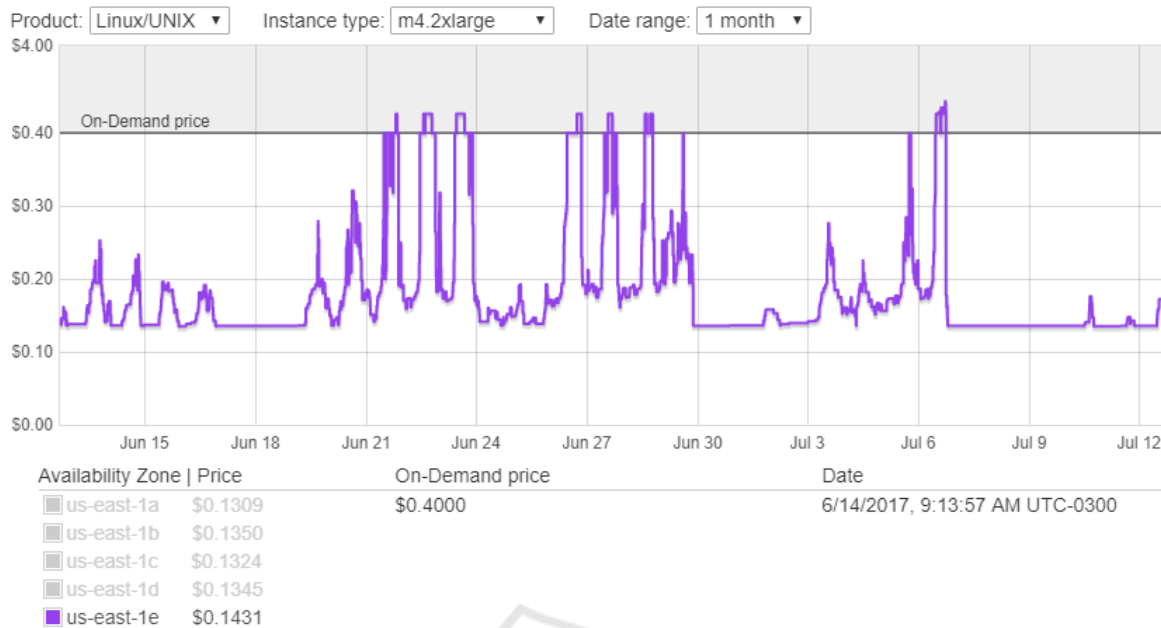


Figure 1: AWS SI price history for a one-month period in 2017.

posed to determine the probabilistic availability assurance for SIs: DrAFTS, acronym of Durability Agreements From Time Series. Thus, a prediction algorithm has been created for bid prices that will gain a certain SI in the AWS market. The prototype also determines the likelihood of how long these prices remain the same. The prediction, in this case, shows a combination of the maximum price of an SI and the time that this value guarantees the termination of the VM, considering that the termination is caused by the increase of bid prices.

To ensure predictability, DrAFTS (Wolski and Brevik, 2016) runs regression tests with the price history of each instance type and stores the predictions. Each time a new prediction is generated, the method selects a random sample of prices in the history and re-runs the DrAFTS algorithm. The fraction of correct predictions is reported with the probability of success.

Other studies consider also the workload that is running on the cloud to suggest a proper amount of resources in order to guarantee its completion. In (Huang et al., 2013), they proposed a tool that automatically profiles the application, builds a model to predict its performance, and infers a proper cluster size that can finish the job within its deadline while minimizing the total cost. Based on these parameters, the tool also chooses between on-demand or spot instances. In the work of (Sabyasachi et al., 2017), the proposal is about a framework that allows users to bid different prices depending on their perceived urgency

and nature of the running job. It allows them to negotiate the current bid price in a way that guarantees the timely completion of their jobs.

The work of (Chohan et al., 2010) uses SIs as the acceleration mechanism for MapReduce applications executed in benchmarks presented in the article. However, the unexpected termination of SIs can have adverse effects on application execution time, and could increase the final cost. The research then shows techniques that help mitigate these effects.

According to the authors of (Chohan et al., 2010), an SI is suitable for batch processing of MapReduce because of its fault tolerance characteristic. When a VM becomes unavailable, the internal mechanism of MapReduce automatically looks for another VM to complete the task. With this, the research concluded that the use of SIs accelerates processing and causes the consequent cost decrease.

In the experiments of (Chohan et al., 2010) four on-demand machines and one SI were used for acceleration. At some loads this acceleration reaches 200%, while the cost is only increased by 42%. However, the experiments conducted showed that in certain cases, failures negatively impacted processing by up to 27%, since SIs are less reliable than on-demand VMs.

This work differs from the others by comparing, in two cloud providers, the performance of on-demand VMs and transient ones. CPU, Input and Output (I/O), and network performance metrics will be used to identify whether performance is compati-

ble, or whether the price decrease of the transient VM also implies a decrease in performance. The chosen providers were AWS and GCP, since these providers offer both VM options and they are among the top three public cloud providers in the Gartner Magic Quadrant (Gartner, 2016). In addition to performance benchmarking, the cost of an execution scenario will be measured and compared for both VM classes from both providers.

### 3 TESTBEDS

The purpose of the study is to compare the performance of on-demand VMs with the transient versions of these same VMs from a public cloud provider. The on-demand version meets one of the cloud computing features defined by NIST (Mell et al., 2011), which is the perception that there is an infinite supply of resources. That is, it will always be possible to get additional on-demand resources at any time, from the point of view of a single customer.

The transient VMs, on the other hand, have diverse availability. If there are no idle resources in a given provider's availability zone, one may not get transient VMs while this scenario remains. And even when resources are available and they are in use by a customer, the provider can revoke them at their discretion. The counterpart to this decrease in availability is the cost, which is much lower in the model with possibility of preemption, and can reach up to 90% discount in relation to the on-demand price (AWS, 2017c).

We analyze the performance relationship between these two models of infrastructure as a service and verify if the lower price means, in addition to lower availability, lower performance according to three metrics: quantity of floating point operations per second (GFLOPS) that each vCPU supports, I/O throughput rate and network throughput. In addition, the costs of one workload that is transient-server-friendly will be compared, being executed in the two scenarios: one with the use of on-demand VMs and other with the use of transient VMs.

In public cloud providers, different availability zones represent data centers located in different geographic locations and possibly with a different infrastructure as well. For this study, before running the benchmarking software, both classes of VMs were initiated on the same availability zone and with the same configuration (cpu model, disk type).

Providers also offer different families of VMs and some of them are specialized on a computational function: processing, Input/Output (I/O), and net-

Table 1: AWS VM Configuration.

Name	Qty vCPUs	VM Type	Storage (GB)	Zone
On-Dem SML	1	on-demand	50	sa-east-1a
On-Dem MED	2	on-demand	50	sa-east-1a
On-Dem BIG	4	on-demand	50	sa-east-1a
SPOT SML	1	transient	50	sa-east-1a
SPOT MED	2	transient	50	sa-east-1a
SPOT BIG	4	transient	50	sa-east-1a

Table 2: GCP VM Configuration.

Name	Qty vCPUs	VM Type	Storage (GB)	Zone
On-Dem SML	1	on-demand	50	us-central1-a
On-Dem MED	2	on-demand	50	us-central1-a
On-Dem BIG	4	on-demand	50	us-central1-a
PREEM SML	1	transient	50	us-central1-a
PREEM MED	2	transient	50	us-central1-a
PREEM BIG	4	transient	50	us-central1-a

working, for instance. For the sake of performance comparison, the choice was, on each provider, the general-purpose family. This means all VMs used on performance experiments have a balance between the computational functions mentioned. On AWS, M-family was the choice and on GCP, n1-standard family. On cost experiments, the VM families that are appropriate for the specific workload being tested were used.

For comparison purposes, the experiment ran on three VMs with different amounts of vCPUs, in order to increase the evaluation sample and, therefore, to achieve more accurate results. The most powerful VM used in the experiment was the one with 4 vCPUs, referenced in the rest of this work as BIG. The other VMs have lower amounts of vCPUs and will be referenced as MED (for medium) and SML (for small). For each size, an on-demand VM (referenced as On-Dem on both providers) and a transient one (referenced as SPOT on AWS and PREEM on GCP) will be created, according to the tables 1 and 2. The AWS zone where the VMs were created was sa-east-1a and GCP availability zone used was us-central1-a.

For storage, Solid State Drive (SSD) options were used. The providers allows the user to choose between SSD and Hard Disk Drive (HDD) based storage. The SSD has higher throughput in terms of I/O operations per second (IOPS) and data transfers, and hence higher cost per GB of allocated space.

### 4 PLANNING

The study of (Coutinho et al., 2012) compared the results of CPU performance measurements among some available benchmark software and selected Lin-

Table 3: Experiment design.

SYSTEM	Virtual Machine
METRICS	CPU (GFLOPs), I/O (throughput, in MB/s), NETWORK (throughput, in Mbps)
PARAMETERS	For CPU, number of Linpack equations, array size, number of runs, and total size of data. For I/O, the size of files and records. For NETWORK, the load time and the amount of runs.
FACTORS	Size of VMs in amount of vCPUs (BIG, MED and SML)
TECHNIQUE	Measurement
WORKLOAD	Synthetic. Workload of benchmarks LINPACK (CPU), IOZONE (I/O) and IPERF (NETWORK)
DESIGN	For each VM, the benchmarks will be executed independently
DATA ANALYSIS	Interpretation of the results described in tables and graphs.
PRESENTATION OF RESULTS	Bar charts and tables

pack (Intel, 2017). They find compatible results between them when evaluating the same computing environment. Besides, Linpack uses less execution time and has more simplified configuration. Linpack was also the benchmark selected for the study presented here. Regarding I/O, the benchmark chosen was IOZone (Iozone, 2017), which is also widely used in such measurements in the literature. To evaluate network performance, iPerf (Iperf, 2017) was the benchmark software chosen.

The parameters used on CPU benchmark software were defined by (Coutinho et al., 2012). They defined as Linpack parameters 10000 equations (problem size), 10000 as matrix size, 30 repetitions and the data size as 10KB. For IOzone, which measures throughput in various I/O operations, the default parameters were used. IOZone varies the size of the file to be manipulated, as well as the data records that compose these files, in 8 different operations. For the purpose of this comparison, we will show the results of writing and reading operations only. For iPerf, which measures the network throughput, standard TCP protocol parameters were used.

The methodology for performing performance analysis is described in (Jain, 1991). It is possible to detail characteristics related to the design of the experiments, such as metrics and workloads, besides the factors and parameters. The details are described in Table 3.

#### 4.1 Experiment 1 - vCPU

The goal of this experiment is to compare VM performance results between both classes (on-demand

Table 4: AWS VMs performance measurements (GFLOPS).

GFLOPS	SPOT BIG	SPOT MED	SPOT SML	On-Dem BIG	On-Dem MED	On-Dem SML
Mean	145.43	73.53	35.709	145.87	72.51	35.99
Deviation	0.21	0.14	0.15	0.52	0.52	0.16
Minimum	144.45	72.99	35.10	143.69	69.97	35.42
Maximum	145.63	73.62	35.80	146.48	72.81	36.16

Table 5: GCP VMs performance measurements (GFLOPS).

GFLOPS	PREEM BIG	PREEM MED	PREEM SML	On-Dem BIG	On-Dem MED	On-Dem SML
Mean	125.70	60.13	31.82	125.95	65.14	32.39
Deviation	7.26	1.01	2.08	1.96	0.76	0.47
Minimum	111.68	57.63	22.16	117.84	62.39	30.45
Maximum	131.34	62.40	33.17	127.86	65.97	32.80

an transient) for each size. Only VMs of the same provider will be compared to each other.

Tables 4 and 5 show the values measured for each VM, as well as the standard deviation of the sample and the maximum and minimum values. It can be seen that in all of them the standard deviation was very small, which demonstrates stability in the delivery of the contracted vCPU resource in both contracting models.

Figure 2 compares each AWS on-demand VM size with the similar transient VM and it is seen that the measurement is practically the same in all cases. The results found for GCP VMs follow the same pattern, as seen in Table 5.

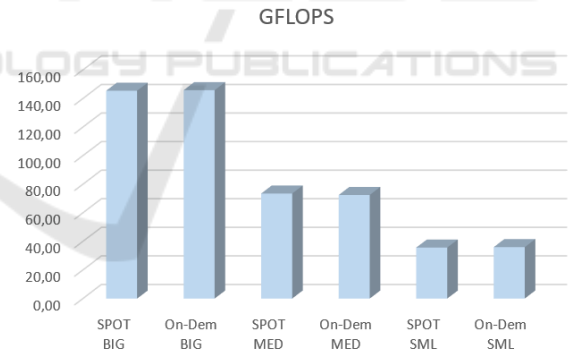


Figure 2: AWS vCPU comparative performance.

#### 4.2 Experiment 2 - I/O

The purpose of this experiment is to measure the I/O throughput of all VMs and compare them.

For this experiment IOZone was used. IOzone is a benchmark for file system. It generates and measures a variety of read and write operations on files. The tests used in this experiment were read, indicating the performance of reading a file that already exists in the file system, and write, indicating the performance of writing a new file in the file system. These tests create temporary test files of sizes ranging from 64KB to

Table 6: AWS I/O throughput, in MB/s, for write operations.

	SPOT BIG	SPOT MED	SPOT SML	On-Dem BIG	On-Dem MED	On-Dem SML
Mean	2504	2867	2211	2751	2456	2177
Minimum	497	420	506	552	458	475
Maximum	3407	3947	3070	3868	3425	2993

Table 7: AWS I/O throughput, in MB/s, for read operations.

	SPOT BIG	SPOT MED	SPOT SML	On-Dem BIG	On-Dem MED	On-Dem SML
Mean	5042	5515	4530	5265	4519	4248
Minimum	2031	2185	1339	2128	1882	1743
Maximum	11763	11956	11403	11185	10220	10119

512MB. The size of the records varies from 4KB to 16MB. All results are in MB/s.

Tables 6 and 7 show AWS throughput values measured by IOZone for each VM for write and read operations, respectively. By the analysis of Figures 3 and 4, it can be seen that the measurements are quite similar, with a slightly higher value of On-Dem BIG compared to SPOT BIG, but with slightly lower values for both On-Dem MED, as well as for On-Dem SML, when compared to their equivalent SPOT. The result seems to indicate that the variation was due to the I/O load of the infrastructure of the sa-east-1a zone at the time of measurement rather than to an actual difference between the analyzed VMs.

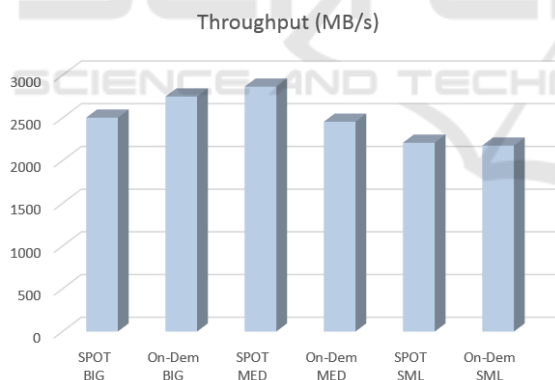


Figure 3: AWS I/O throughput, in MB/s, for write operations.

As can be seen in Tables 8 and 9, which show the results measured for the GCP provider, there was less than a 5% average I/O performance difference between the two types of VMs. The results also seem to indicate an expected variation on I/O load of the us-central1-a zone at the time of measurement.

### 4.3 Experiment 3 - Network

The purpose of this experiment is to measure the network throughput between on-demand VMs and tran-

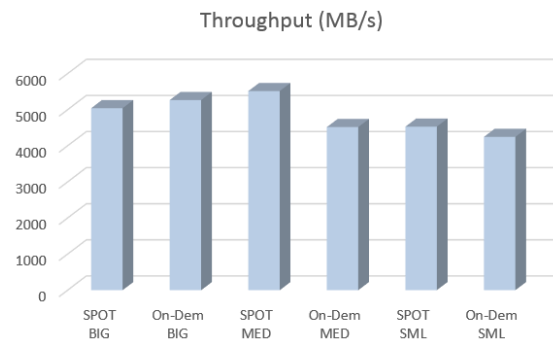


Figure 4: AWS I/O throughput, in MB/s, for read operations.

Table 8: GCP I/O throughput, in MB/s, for write operations.

	PREEM BIG	PREEM MED	PREEM SML	On-Dem BIG	On-Dem MED	On-Dem SML
Mean	1892	1886	2180	1858	1916	2071
Minimum	373	303	373	84	406	422
Maximum	2797	3145	2966	2917	3256	2947

sient ones on both providers. For this experiment the benchmark IPerf was used.

SML size VM was used as the server, while the other two VMs were used as client machines within the same zone.

The experiment consisted of generating a stream of TCP data on each client machine for one minute and measuring it second by second. This throughput was executed 40 times at different hours of the day, in order to record the throughput variation.

As there was a close similarity between the measured values of the two VMs that served as IPerf clients, the values were consolidated and are presented in Tables 10 and 11 with reference only to BIG VMs. As expected, there was a relevant variability in the measurements, which is reflected in the ratio of the deviations from the mean, as well as in the interval between the maximum and minimum values.

The difference between the measurements was quite small, as can be seen in Figure 5 for GCP. AWS has similar results as can be seen in Table 10. In the network measurement, the GCP VMs presented a difference of less than 1% in the throughput, which seems to show no actual difference between the VMs, but only a regular variability found in this type of measurement.

Table 9: GCP I/O throughput, in MB/s, for read operations.

	PREEM BIG	PREEM MED	PREEM SML	On-Dem BIG	On-Dem MED	On-Dem SML
Mean	3337	3006	3545	3319	3353	3802
Minimum	1443	1123	1539	1183	825	1592
Maximum	12075	12427	12075	10159	11852	10894



Table 10: AWS Network throughput (in Mbps).

	SPOT BIG	On-Dem BIG
Mean	303	313
Deviation	66	94
Minimum	276	296
Maximum	635	1,001

Table 11: GCP Network throughput (in Mbps).

	PREEM BIG	On-Dem BIG
Mean	3614	3592
Deviation	545	543
Minimum	1410	430
Maximum	4960	5100

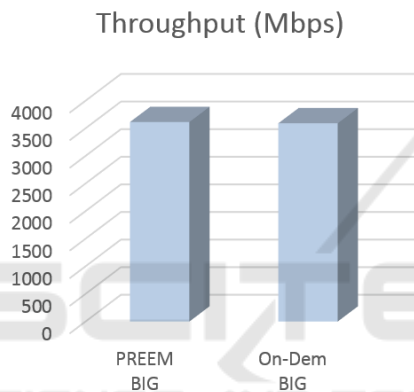


Figure 5: GCP Network throughput (in Mbps).

## 5 ANALYSIS

The experiments carried out showed that, in relation to vCPU performance, the greatest measurement difference found on AWS was small, only 1.4% in favor of VM SPOT MED. On the other hand, the On-Dem BIG and On-Dem SML VMs performed better than their similar ones by only 0.3% and 0.8%, respectively. On GCP, all on-demand VMs performed better than the preemptive ones. But the differences were tiny on average: 0.2%, 8.3% and 1.8% for BIG, MED and SML VMs, respectively.

In relation to the I/O throughput, the greatest measurement difference found on AWS was 18%, in the reading operations, in favor of VM SPOT MED. VM On-Dem BIG had a superior result of 4.4% and VM SPOT SML performed better than its equivalent at 6.22%. On GCP side, the greatest difference was also on reading operations. It was 11.5% in favor of VM On-Dem MED. VM PREEM BIG had a 0.55% better throughput and VM On-Dem SML did 7,24% better

than its preemptive equivalent.

Regarding the network throughput, the On-Dem BIG performed 3.3% higher when compared to its equivalent SPOT BIG on AWS. On GCP, PREEM BIG performed 0,61% better than On-Dem BIG.

The small differences found in measurements, as well as the fact that the VM size that presented the best performance varied according to the experiment and, in some cases, varied within the same experiment, allows us to state that there is no significant difference between on-demand and transient VMs from both providers.

## 6 COSTS

In order to proceed to cost analysis, additional experiments were performed to compare on-demand and transient VMs from each provider.

### 6.1 Costs in AWS

The VMs available in AWS for use in cost experiments are listed in Table 12. The AWS M family of VMs is used for general purpose applications (AWS, 2017b), while the R family is used for memory intensive applications.

Some VMs are available with Elastic Block Store (EBS), an AWS service that adds a virtual disk to the VMs for persistent data storage. Local SSD disks, however, are removed when the VM is turned off. EBS volumes preserve the data and can be copied as a backup. They can be attached later to other VMs (AWS, 2017a). VMs that use EBS are more robust and suitable to perform intensive processing, especially Hadoop applications, which is the focus of the additional experiments (AWS, 2017b).

Table 12: AWS instance types.

Type	CPU	Memory	Storage
m3.large	2	7.5	1 x 32
m3.xlarge	4	15	2 x 40
m3.2xlarge	8	30	2 x 80
m4.large	2	8	EBS-only
m4.xlarge	4	16	EBS-only
m4.2xlarge	8	32	EBS-only
r4.large	2	15.25	EBS-Only
r4.xlarge	4	30.5	EBS-Only
r4.2xlarge	8	61	EBS-Only

The AWS prices are shown on Table 13. For the additional experiments, EC2 instances running the

Elastic MapReduce (EMR) service were used. To calculate the total amount to be paid per hour of use, the columns for the hourly price of EC2 and EMR must be added together. The values are in US dollars. AWS us-east-1 region was used since it has the lowest price for on-demand VMs and then provides a fair comparison to SI prices.

Table 13: AWS prices for on-demand VMs (*us-east-1*).

Type	EC2 (US\$/hour)	EMR (US\$/hour)
m3.xlarge	0.266	0.070
m3.2xlarge	0.532	0.140
m4.large	0.100	0.030
m4.xlarge	0.200	0.060
m4.2xlarge	0.400	0.120
m4.4xlarge	0.800	0.240
r4.xlarge	0.266	0.067
r4.2xlarge	0.532	0.133

This experiment consists of indexing a file of 26 GB of data in BZ2 (compressed) format using MapReduce on a cluster composed of a master machine and two workers of VM type r4.xlarge, whose family of VMs is more appropriate to execute EMR clusters. This experiment was executed several times and an average execution time was calculated for both SI and on-demand EMR clusters.

In the SI mode, 3 clusters of the same configuration (3 VMs r4.xlarge) were created and named HDP3, HDP4 and HDP5, according to the values used as the SI bid: US\$ 0.03, US\$ 0.04 and US\$ 0.05, respectively. The SI market value for each zone of region us-east-1 for this VM type is shown in Table 14.

Table 14: Market value of r4.xlarge SPOT on region *us-east-1*

Zone	Value (US\$)
us-east-1a	0.047
us-east-1b	0.033
us-east-1c	0.039
us-east-1d	0.035
us-east-1e	0.032
us-east-1f	0.027

Clusters HDP4 and HDP5 were provisioned in 300 seconds, which is the average time verified in AWS for this type of VM. According to Table 14, the bid values were all above the market value for r4.xlarge in zone us-east-1f. However, HDP3 was not provisioned.

There is a large difference between the values of on-demand VMs and SIs. While an on-demand

r4.xlarge costs US\$ 0.266 per hour, an SI r4.xlarge costs between US\$ 0.027 and US\$ 0.047 at that time in the us-east-1 region. However, as verified in the experiment, there is no guarantee of provisioning, especially in cases where the bid value is close to the minimum value in the SI market.

Regarding performance, the on-demand and SI machines are similar, as seen in Section 5. The execution time of an indexing MapReduce application with the same dataset was, on average, between 4 and 5 hours in all EMR clusters created. The result can be verified in Table 15, with the confirmation that the performance does not suffer degradation during execution in SI, unless one or more instances are revoked during the execution.

HDP4 execution time was slightly lower than the on-demand EMR cluster execution time. HDP5 execution time was higher, in the opposite direction. This performance loss for HDP5 probably came from SI revocations and replenishment.

Table 15: AWS EMR cluster average execution time.

EMR cluster	Time
On-demand	4 hours e 16 minutes
HDP4	4 hours e 15 minutes
HDP5	4 hours e 52 minutes

## 6.2 Costs in GCP

The experiments in GCP were conducted using VMs of type 'n1-highmem-4', each of them with 200 GB disk and 26 GB of RAM. The datasets had 6 GB of data in BZ2 (compressed) format, and were processed using MapReduce.

GCP has a service called Dataproc, where one can create Hadoop clusters. On GCP, a Hadoop cluster should have at least the master node and 2 workers of on-demand VMs. Because of this, the cost experiment in this provider was a little bit different. That is, on AWS it is possible to create a cluster totally composed of transient servers, but on GCP it is possible only to create a partially composed one.

The GCP prices are showed on the table 16. On GCP there is no price variation for preemptive VMs. Similar to AWS, there is a fixed price for Dataproc service that must be added to VM execution price to achieve the total cost of cluster execution. The values are in US dollars and refer to region us-east1.

Table 16: GCP hourly prices (US\$) for n1-highmem-4 VM.

Type	On-dem	Preem	Dataproc
n1-highmem-4	0.2696	0.0535	0.04

Two clusters were created on GCP. The first one was composed of 1 master node and 4 workers, all of them on-demand. The second cluster was composed of 1 master node, 2 on-demand workers and 2 more preemptive workers. This cluster is therefore partially preemptive.

As seen in Table 17, the performance in hybrid clusters (preemptive and on-demand VMs) is similar to cluster with only on-demand VMs.

The partially preemptive cluster executed 7% slower on average, with a cost 26% smaller than the totally on-demand one. Considering that each preemptive VM has an 80% discount, that Dataproc service has a fixed cost and that the cluster is composed of at least 3 on-demand VMs, one can reach a bigger discount when executing larger clusters with only 3 mandatory on-demand VMs and the rest of them as preemptive VMs. For example, a cluster made of 3 on-demand VMs plus 7 preemptive VMs, executing the same workload described in this subsection (6.2) will reach a discount level of 52%.

Table 17: GCP cluster average execution time (seconds).

	Partially Preem	On-Demand
Mean	1330.00	1241.75
Deviation	32.11	32.01
Minimum	1289.00	1217.00
Maximum	1358.00	1285.00

## 7 CONCLUSION

This work compared the performance of on-demand VMs with equivalent preemptive VMs (same type, same configuration, running in the same zone). Metrics related to CPU, I/O and network were used, thus comparing execution costs in specific scenarios on each provider.

The results indicated that there were no significant performance differences in the metrics used for CPU, I/O and network. These results confirm that both VMs use the same computational resource set. This is additionally verified by a series of MapReduce application executions. The differences between the two modes are availability and cost. The availability of on-demand VMs seems unlimited, since in the vast majority of cases the VM can be started at any time and will be interrupted only at the request of the customer, or in case of technical failure.

The preemptive VMs, however, can be revoked anytime by criteria that are out of control by the customer, such as market price, lack of idle resources and others defined by the provider. The cost of preemp-

tive VMs can reach a 90% discount, on AWS, and an 80% discount on GCP. In the scenario executed by this study, a discount of 68% was obtained in relation to on-demand execution on AWS and a discount of 26% on GCP.

Although the AWS discount was greater than what was achieved on GCP, as all AWS cluster nodes were preemptive VMs, that increased the risk of not finishing the work or that have an increase on execution time caused by node revocations.

The performance and cost measurements performed in this study add value to applications that can benefit from the use of transient servers (AWS spot instances and GCP preemptive instances), thereby achieving a cost reduction, with little additional effort to automate the replenishment of instances that are claimed by the provider, when it is the case.

Future work is intended to analyze other transient server offers, as Microsoft Azure Low-priority VMs, and to evaluate other execution scenarios whose applications can deal with VM revocations and could benefit for cost reductions offered by the use of transient servers.

## REFERENCES

- Agmon Ben-Yehuda, O., Ben-Yehuda, M., Schuster, A., and Tsafir, D. (2013). Deconstructing amazon ec2 spot instance pricing. *ACM Transactions on Economics and Computation*, 1(3):16.
- AWS (2015a). Ec2 spot bid advisor. [online]. available from: <https://aws.amazon.com/ec2/spot/bid-advisor/>. 2017-05-17.
- AWS (2015b). Ec2 spot-fleet. [online]. available from: <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-fleet.html>. 2017-05-17.
- AWS (2017a). Amazon elastic block store. [online]. available from: <https://aws.amazon.com/efs/>. 2017-06-01.
- AWS (2017b). AWS EC2 Instance Types. [Online]. Available from: <https://aws.amazon.com/ec2/instance-types/>. 2017-10-12.
- AWS (2017c). AWS Free Tier. [online]. available from: <https://aws.amazon.com/free/>. 2017-05-17.
- Chhetri, M. B., Lumpe, M., Vo, Q. B., and Kowalczyk, R. (2017). On estimating bids for amazon ec2 spot instances using time series forecasting. In *Services Computing (SCC), 2017 IEEE International Conference on*, pages 44–51. IEEE.
- Chohan, N., Castillo, C., Spreitzer, M., Steinder, M., Tantawi, A. N., and Krintz, C. (2010). See spot run: Using spot instances for mapreduce workflows. *Hot-Cloud*, 10:7–7.
- Coutinho, E. F., Rego, P., Gomes, D. G., and de Souza, J. N. (2012). Análise de Desempenho com Benchmarks em um Ambiente Público de Computação em Nuvem. In

- Proc. of the X Workshop em Clouds e Aplicações. Belo Horizonte*, pages 96–109.
- Gartner (2016). Magic Quadrant for Cloud Infrastructure as a Service, Worldwide. [online]. available from: <https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519>. 2017-05-17.
- Google (2017). Google preemptible instances. [online]. available from: <https://cloud.google.com/compute/docs/instances/preemptible>. 2017-01-07.
- Huang, H., Wang, L., Tak, B. C., Wang, L., and Tang, C. (2013). Cap3: A cloud auto-provisioning framework for parallel processing using on-demand and spot instances. In *Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference on*, pages 228–235. IEEE.
- Intel (2017). Linpack - Intel math kernel. [online]. available from: <https://software.intel.com/en-us/mkl>. 2017-05-17.
- Iozone (2017). Iozone - filesystem benchmark. [online]. available from: <http://www.iozone.org/>. 2017-05-17.
- Iperf (2017). Iperf. Network performance benchmark. [online]. available from: <http://iperf.sourceforge.net/>. 2017-05-17.
- Jain, R. (1991). *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling*.
- Khandelwal, V., Chaturvedi, A., and Gupta, C. P. (2017). Amazon ec2 spot price prediction using regression random forests. *IEEE Transactions on Cloud Computing*.
- Mell, P., Grance, T., and others (2011). The NIST definition of cloud computing.
- Microsoft (2017). Azure batch low-priority vms. [online]. available from: <https://docs.microsoft.com/azure/batch/batch-low-pri-vms>. 2017-12-20.
- Sabyasachi, A. S., Kabir, H. M. D., Abdelmoniem, A. M., and Mondal, S. K. (2017). A resilient auction framework for deadline-aware jobs in cloud spot market. In *Reliable Distributed Systems (SRDS), 2017 IEEE 36th Symposium on*, pages 247–249. IEEE.
- Sharma, P., Guo, T., He, X., Irwin, D., and Shenoy, P. (2016). Flint: batch-interactive data-intensive processing on transient servers. In *Proceedings of the Eleventh European Conference on Computer Systems*, page 6. ACM.
- Sharma, P., Irwin, D., and Shenoy, P. (2017). Portfolio-driven Resource Management for Transient Cloud Servers. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(1):5.
- Singh, R., Sharma, P., Irwin, D., Shenoy, P., and Ramakrishnan, K. (2014). Here today, gone tomorrow: Exploiting transient servers in datacenters. *IEEE Internet Computing*, 18(4):22–29.
- Wolski, R. and Brevik, J. (2016). Providing statistical reliability guarantees in the aws spot tier. In *Proceedings of the 24th High Performance Computing Symposium*, page 13. Society for Computer Simulation International.

## Anexo IV

Artigo publicado no congresso  
CLOSER 2018 - Ilha da Madeira -  
Portugal [2]

# Risk Perception of Migrating Legacy Systems to the Cloud

Breno G. S. Costa and Priscila Solis

*Department of Computer Science, University of Brasília, Brasília-DF, Brazil*

**Keywords:** Cloud Computing, Legacy Systems, Migration, Reference Model, Risk Perception.

**Abstract:** With the utilization of Cloud Computing as the way to provide Information Technology services, the organizations can migrate legacy systems to the cloud in order to reach several benefits. There are in the literature several proposals to model the critical elements of migration and many have been validated by specific case studies. Also several reference models were defined on top of these proposals and were created with the intention to consolidate different researches, trying to expand their applicability. Based on the above, this paper selects a reference model for migrating legacy systems to the cloud and proposes a method for calculating a score of perceived risk associated with each legacy system migration. The paper presents a proof of concept on government domain to show the method's applicability.

## 1 INTRODUCTION

In the last years Cloud Computing (CC) changed the way Information Technology (IT) is provided and consumed and created a new landscape in which organizations manage their own computing infrastructure to another, where IT is consumed as a service (ISACA, 2012).

According to (Gartner, 2017), the massive use of public cloud services appears in the significant growth rates of service providers such as Amazon Web Services (AWS) (AWS, 2017) and Microsoft Azure (Microsoft, 2017), among others. The study surveyed almost 3000 participants and found that 21% of them have already used public cloud services, while 56% planned to implement the cloud by the end of 2017. For many of these organizations, IT modernization is almost synonymous of increasing the use of CC.

The migration to this new paradigm has a potential cost reduction of IT infrastructure, noticed by several authors (Armbrust et al., 2009; Armbrust et al., 2010; Gartner, 2017), mainly when related to the use of public cloud. According to (Pahl and Xiong, 2013), migration to the cloud is the process of deploying, in whole or in part, the digital assets, services, IT resources, or systems of an organization on the cloud. With the use of CC, government organizations can reduce the number of contracts and consequently reduce the opportunities for irregularities while improving efficiency. They can also focus on providing services to the population by reducing the operational

effort to maintain the infrastructure and IT services platform (Kundra, 2011).

Migration to the cloud may involve retention of part of the infrastructure within the organization. For example, a legacy system can be merged with a complementary cloud solution, with integration over the Internet. The trend of increased use of CC, including on government organizations, may benefit of using a process (guidelines and methods) to migrate legacy systems to the public cloud.

The goal of this paper is to present an extension to a reference model of legacy systems migration to the cloud. The extension will be validated with a proof of concept in government domain. This paper makes the following contributions:

- based on a review of related work, proposes a method for calculating a score of perceived risks associated with the legacy migration to the cloud;
- describes a proof of concept showing the applicability of the method for the government domain.

The remainder of this article is divided into four sections. Section 2 presents the theoretical concepts and related work and also describes the motivation of evaluating the risk perception of applying the reference model. Section 3 presents the proposed evaluation method. Section 4 describes the proof of concept. Finally, Section 5 presents the conclusions and future work of this research.

## 2 THEORETICAL CONCEPTS AND RELATED WORK

### 2.1 Migration Methods and Strategies

Software migration is the process of moving (or adapting) a system from one operating environment to another (ISO/IEC, 2006). Software migration can be seen as part of a broader context that is software maintenance and reuse (Ionita, 2013). Migration to the cloud is the effort to adapt legacy systems running locally on the organization's infrastructure to transport them to the cloud (Jamshidi et al., 2013).

To take advantage of CC benefits and protect existing investment in a legacy system, organizations tend to migrate their legacy systems to the cloud gradually when feasible (Zhao and Zhou, 2014; Gartner, 2017; Ionita, 2013). As discussed in (Pahl and Xiong, 2013), the migration process requires careful consideration, planning, and execution to ensure system security and integrity after migration, and to remain compliant with the organization's requirements.

According to (Babar and Chauhan, 2011), unsuccessful cloud migration can cause business processes to fail, increase security incidents, and increase maintenance costs. In addition to these considerations, the work of (Zhao and Zhou, 2014) discusses migration strategies and methods. For them, different migration strategies should consider different migration processes, with their subdivisions into specific tasks. In the study of (Binz et al., 2011), the authors propose an application migration framework for the cloud and between cloud providers.

One alternative point of view regarding migration to the cloud is the use of adaptation, addressed by (Andrikopoulos et al., 2013). The authors identify four types of legacy systems migration. In (Zhao and Zhou, 2014), the authors analyze and compare the two migration definitions cited in (Binz et al., 2011) and (Andrikopoulos et al., 2013), among other definitions and categorized migration to the cloud into three major areas: migration to Infrastructure as a Service, migration to Platform as a Service and migration to Software as a Service.

One of the main benefits of using CC is the potential cost reduction of IT infrastructure investments (Armbrust et al., 2009; Armbrust et al., 2010; Gartner, 2017). However, according to (Sun and Li, 2013), this benefit typically only considers the cost of cloud services after migration. The authors developed a method to estimate the cost of the migration process from legacy systems to cloud. The method considers that, although described sequentially, the execution flow of the migration tasks iterates in a way in

which errors in the execution of a task can cause the repetition of previous tasks. Thus, according to the size and complexity of the legacy system and also to the maturity of the team responsible for the migration, the cost of the migration process can be calculated using the method they have defined.

### 2.2 Characterization Model

Many publications address how to structure the migration of legacy systems to the cloud. The Cloud-RMM model (Jamshidi et al., 2013) categorizes twenty-three studies on the migration of legacy systems to the cloud and serves as a guide to the analysis of these studies. The model is composed of four processes and each process is a group of tasks. In the model, there is also an indication of the artifacts generated at the end of each process.

In this model, there is only a consolidation of processes, tasks and artifacts, found in the twenty-three studies analyzed in the systematic literature review undertaken by (Jamshidi et al., 2013). There is no discussion about the need and relevance of each process and task.

### 2.3 Evaluation Model

The work of (Gholami et al., 2016) provides a detailed assessment of existing CC migration approaches from the perspective of process modeling and software development methodologies. The authors propose an evaluation model to compare existing approaches, highlighting their resources, similarities and differences. The approach used by the authors differs from other related works because it focuses on the aspect of the cloud migration process to understand which core activities and concerns are involved during this transition. According to their analysis, none of the reviewed studies provides an in-depth discussion of proposed migration features and activities and also did not bring useful experience of applying those approaches in practice. In addition, the article provides a meticulous analysis of existing approaches through an evaluation model which encompasses twenty-eight criteria classified into two dimensions. That is, eleven generic criteria and seventeen specific criteria for cloud computing. The proposed framework was derived from a literature review and validated through a web-based questionnaire survey of 104 academics and experts in the field of cloud computing (Gholami et al., 2016).

As challenges for future work, (Gholami et al., 2016) acknowledge that there is a large amount of research on cloud migration, which is currently dis-

persed and fragmented, and suggest that a generic reference model must be defined with the goal of consolidating the existing literature.

### 2.4 Reference Model

According to (Fettke and Loos, 2003), a reference model is a conceptual framework that can be used as the model for information systems development. Reference models are also called universal models, generic models, or standard models. To use reference models, they must be tailored to the requirements of a specific domain. The study presents several perspectives for the evaluation of reference models. Among them, the empirical perspective, from which we mention two approaches that relate to the object of this research: case study and survey.

In a more recent study, (Gholami et al., 2017) responded to the challenge proposed in (Gholami et al., 2016) and worked to find out the critical activities, artifacts, concerns and key recommendations regarding the migration of legacy systems to the cloud. Results were validated empirically, through the collection of experts perceptions to increase the reliability.

Based on an intensive qualitative analysis of existing constructs in the literature, the authors created a migration to the cloud reference model. The relevance and robustness of the model was further confirmed using quantitative research and qualitative feedback from domain experts. Figure 3 shows a summary of the resulting reference model. To simplify the view, the summary shows only the model's key elements, without the subdivisions of activities into smaller tasks and without the flow of information between them. The detailed view of phases Plan and Design will be better described in Section 4.

Although this reference model has been created from the literature and evaluated using survey with CC experts, the authors do not claim its general applicability. In the opposite sense, they assert that there is no universally superior or applicable method for

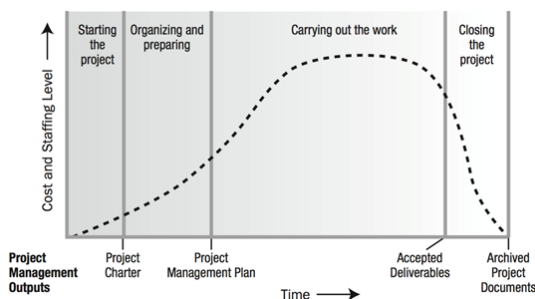


Figure 1: Project life cycle as defined on PMBOK (PMI, 2013).

all cloud migration scenarios and therefore, methods must be tailored to the specific characteristics of the application domain.

### 2.5 Legacy Systems Migration Viewed as an IT Project

Each legacy system migration to the cloud can be viewed as a different project, following what is outlined in the Project Management Body Of Knowledge (PMBOK)(PMI, 2013): it is a temporary endeavor and creates a unique product: the legacy system migrated to the cloud. PMBOK defines the life cycle of a project as being the series of phases, usually sequentially linked, that a project goes through. Figure 1 shows the generic structure of a project life cycle and the level of costs and personnel required at each stage of the life cycle.

According to the PMBOK, the generic structure of the life cycle presents, among others, the following characteristics:

- The personnel and cost levels are low at the beginning, they reach a maximum value while the project is running, and fall quickly as the project is finalized. Figure 1 illustrates this pattern.
- The risks and uncertainties are greater at the beginning of the project. These factors decrease over the life of the project as decisions are made and deliveries are accepted, as shown in Figure 2.
- The ability to influence the final characteristics of the project's product, without significant impact on costs, is higher at the beginning of the project and decreases as the project progresses to its end. Figure 2 illustrates the idea that change costs and bug fixes generally increase significantly as the project nears completion.

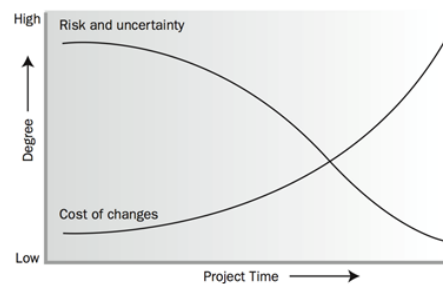


Figure 2: Project risks and cost of changes through the time (PMI, 2013).



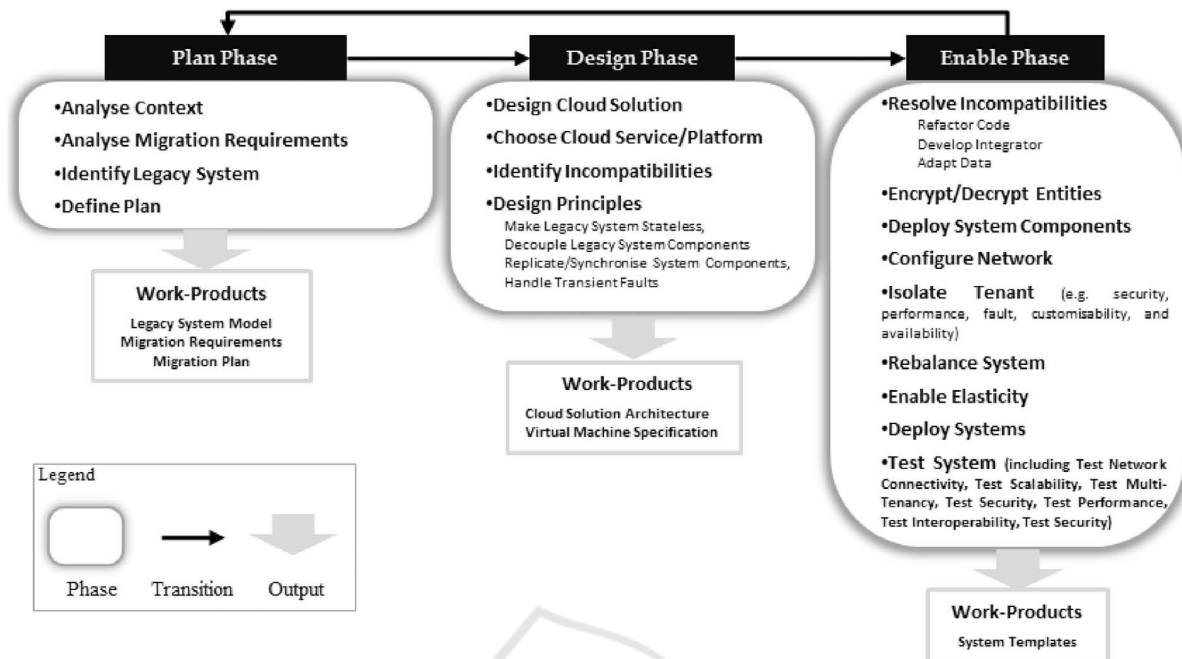


Figure 3: Key elements of Legacy to Cloud migration Reference Model proposed by (Gholami et al., 2017).

### 3 REFERENCE MODEL EVALUATION

The previous section presented some definitions and classifications of migration models and analyzed three of them: a characterization model that grouped processes, activities and artifacts referenced in 23 migration proposals published between 2010 and 2013; an evaluation model that, based on 28 relevant criteria, evaluated 43 articles published between 2009 and 2015; and a reference model (Gholami et al., 2017), selected as the basis for this study, defined from the consolidation of 78 proposals published between 2008 and 2015. The selected reference model is divided into three phases: Plan, Design and Enable as shown on Figure 3. The Plan phase is responsible for collecting relevant information about the system to be migrated (technical information, organizational context) and, in accordance to migration requirements, for creating an appropriate migration plan. The Design phase uses the knowledge and artifacts generated in Plan phase and is responsible for choosing one or more CC providers and also for defining the new architecture that the legacy system will have on the cloud. The Enable phase represents the actual migration. It comprehends the implementation of legacy system code adaptations, the build of integrators when appropriate, the configuration of the CC services defined in Design phase, testing and go live of migrated system.

As shown in Figure 1, the effort required to execute the third phase is potentially bigger than what will be spent in the two initial phases. This is due to the fact that first phases have mostly analytical work, while the third, mostly implementation work. Implementation in this context means both the development related to the adaptations and new integrations, when necessary, and cloud services configuration.

Thus, it is important to have an early assessment of migration risks before Enable phase. The assessment may indicate an increased risk perception (migration additional cost and time, for instance) for a given system when compared to the others, and thus provide an objective measure of comparison. It is possible to rank the systems that will be migrated and set up an order of execution that prioritizes the systems with less perceived risk, thus increasing the success rate and organization's confidence on the migration project.

As a way to provide an objective and early assessment of perceived risks of migration of each legacy system to the cloud using the Reference Model, we defined the following procedure: in each of the two initial phases (Plan and Design), one must select the tasks that together better represent the critical points of each legacy system migration to the cloud. Some tasks have a job of merely collecting and aggregating information, while others implement constraints, decisions or characteristics that are intrinsic to the system and the organization and thus have the potential

to impact the rest of the migration. One should choose between tasks in the second category. It is important to realize that different application domains may have different sets of tasks that better evaluate the risk perception of migrating legacy systems to the cloud on that domain.

After choosing the set of evaluation tasks, it is necessary to define the weights that each task will have in relation to the other tasks on the set. The weight of tasks is related to the importance of that task in the domain of application. For instance, on government domain the task related to procurement is perceived as being more risky than the one related to training, due to legal and regulation rules.

The evaluation function defined here has the following format:

$$EvF(Pl, De) = \frac{Pl + De}{2} \tag{1}$$

where:

$$Pl = \frac{\sum_{i=1}^n V_{Pl_i} W_{Pl_i}}{\sum_{i=1}^n W_{Pl_i}} \tag{2}$$

$$De = \frac{\sum_{j=1}^n V_{De_j} W_{De_j}}{\sum_{j=1}^n W_{De_j}} \tag{3}$$

and:

*Pl* – Plan Phase Evaluation Indicator.

*De* – Design Phase Evaluation Indicator.

*V<sub>Pl<sub>i</sub></sub>* – Risk perception rate for task *i* of the Plan Phase evaluation set.

*W<sub>Pl<sub>i</sub></sub>* – Weight of task *i* in Plan Phase evaluation set.

*V<sub>De<sub>j</sub></sub>* – Risk perception rate for task *j* of Design Phase evaluation set.

*W<sub>De<sub>j</sub></sub>* – Weight of task *j* in Design Phase evaluation set

The function defined in (1) considers that both phases have the same relative importance on calculating the the score of perceived migration risks. But specific circumstances could be considered on another evaluation function that balance the phases differently.

After selecting the set of evaluation tasks in each phase and the weight of each of them, they must be evaluated. Note that the work required on each task must be performed as defined in the Reference Model. At the end of the execution of each phase, and consequently at the end of the execution of all tasks of that phase, is when the evaluation method is performed. This is an important consideration because it is necessary to know each system in an appropriate level of detail to do a more accurate evaluation and this can

Table 1: Plan Phase evaluation task set.

Task	Weight
Analyze migration cost	3
Identify dependencies	3
Select migration scenario	1

not be done by executing only the tasks on the evaluation set. So, after the end of each phase's work, it is necessary to evaluate each task that is part of the evaluation set. The proposal is to use a scale of 5 values for perceived risk, whose meaning could be 1 – High risk, 2 – Moderate risk, 3 – Average risk, 4 – Somewhat risk and 5 - Low risk. For each task, the 5 levels of the scale can also be defined textually, in order to facilitate the choice and evaluation of each system. This must be done only once and only for the tasks that are part of the evaluation set.

Once the scale is defined and the Plan Phase is finished, the evaluation can start by defining, for each task in the evaluation set, the rating of perceived risk associated with that task's subject in relation to legacy migration and calculating the evaluation indicator of the Plan Phase, *Pl*, as described in (2). The same should be done to Design Phase by calculating the evaluation indicator for the Design Phase, *De*, described in (3). After calculating the two indicators, one can calculate the perceived risk score of migrating the legacy system to the cloud, *EvF(Pl, De)*, as described in (1).

After calculating the evaluation function for each system that will be migrated, a ranking can be established that indicates, among other things, which systems should be migrated first, which will be the systems with higher evaluation values (higher perceived risk scores). This strategy is reinforced by (Reza Bazi et al., 2017) who states that it is better to select a pilot project at first, when performing migration process. Systems with lower evaluation values may receive additional analysis to identify whether work modifications to the evaluation tasks that are causing low evaluation are feasible. Even when could not be possible to change something in systems with low evaluation values, knowing that there is an increased risk perception is of great value. In addition, when using evaluation function values to rank systems, one can also use them to categorize systems on a previously defined risk scale. For this purpose, evaluation thresholds can be defined. For example, one can define that evaluation function results that are below or equal to a certain threshold *T1* mean that the system should have its migration decision reanalyzed. Another example is when the value calculated by evaluation function is above the threshold *T1*, but the value of one of the

indicators ( $Pl$  or  $De$ ) is below a given threshold  $T2$ .

The threshold values can be defined by using a certain value of the Likert scale used, but can also be defined by the history of evaluations already done and considering the result of these migrations. While there is no history of assessments, it is suggested to use the central value of the scale as the threshold  $T1$  and  $T2$  and adjust them as soon as new values have been calculated from actual migrations.

## 4 PROOF OF CONCEPT

The proof of concept in this section presents an analysis of the selected activities and defines weights that best support a risk perception evaluation for the migration of legacy systems in the government domain.

### 4.1 Plan Phase

For the Plan phase, we selected the tasks and define the weights shown in the Table 1. The reasons why the tasks were selected are the following:

**Analyze migration cost** - This activity results in an estimate of migration cost. Cost is associated with the effort required to migrate the legacy to cloud as well as the cost of running the application after migration. As cost of migrating a system increases, also increases the risk perception for the entire migration process. On the other hand, a low-cost migration system can be a good candidate to validate the migration process and to gain experience with both the process and the provider to which the system will be migrated.

If the cost of the migration added to the cost of running the system in the cloud provider does not represent an economy relative to the current cost, this may indicate a system that will not benefit from the cloud if there is no value aggregation by other means, such as the use of intrinsic cloud characteristics (self-scalability, higher availability, for example) (Gholami et al., 2017). According to (Gholami et al., 2017), the objectives of Analyze Context activity (which is an aggregate of Analyze migration cost and other tasks) are cost estimate and risk mitigation.

**Identify dependencies** - This activity has the effect of identifying which other local systems and components the system being analyzed depends on to function properly on the cloud after migration. With this task, it is possible to identify which other systems would have to be previously migrated to the cloud. Or, in the case of a hybrid cloud and if the dependencies are not migrated yet, one will need to evaluate and consider the network latency between the local infrastructure and the cloud. According to (Reza Bazi

Table 2: Design phase evaluation task set.

Task	Weight
Negotiate with cloud provider	3
Train	1
Identify incompatibilities	1

et al., 2017), this is a critical factor to be considered, since legacy systems were developed on older platforms than the current version supported by cloud providers. The objective of Recover Legacy application knowledge (activity which aggregates Identifying dependencies task) is Understand legacy system dependencies (Gholami et al., 2017).

**Select migration scenario** - According to the characteristics of the legacy system and also according to the amount of effort that is intended to be spent in the migration, one may choose the migration type among five options (Andrikopoulos et al., 2013; Gholami et al., 2016). Migration type V, for instance, is a type associated to a low risk of failure when compared with other types, since it is based on moving the whole system stack to the cloud. Thus, the choice of the scenario may be directly related to the risk intensity that is accepted. This is not the same to say that a scenario of low risk is also the option that will bring the greatest benefit (cost reduction, increased availability), because scenario V, for instance, could make costly the implementation of elasticity (Andrikopoulos et al., 2013). The objectives of Define Plan (activity that aggregates Select migration scenario task) are Project management and Risk mitigation (Gholami et al., 2017).

### 4.2 Design Phase

For the Design phase, we selected the tasks and define the weights shown in the Table 2.

**Negotiate with cloud provider** - This task is critical because only after its execution one or more cloud providers are available to provide the cloud services that will support migrated systems. If one have not still hired the provider at migration time of a particular legacy system, this may delay the migration or cause rework. Note that even after a provider has already been hired and some legacy systems have been migrated to the cloud, it is possible that the contract could be reaching its end, or even that the provision of the service is not satisfactory and that the organization has decided to replace the provider, even before the end of the contract.

**Training** - This task is responsible for acquiring and maintaining the necessary knowledge to properly design, operate and monitor the CC services. A

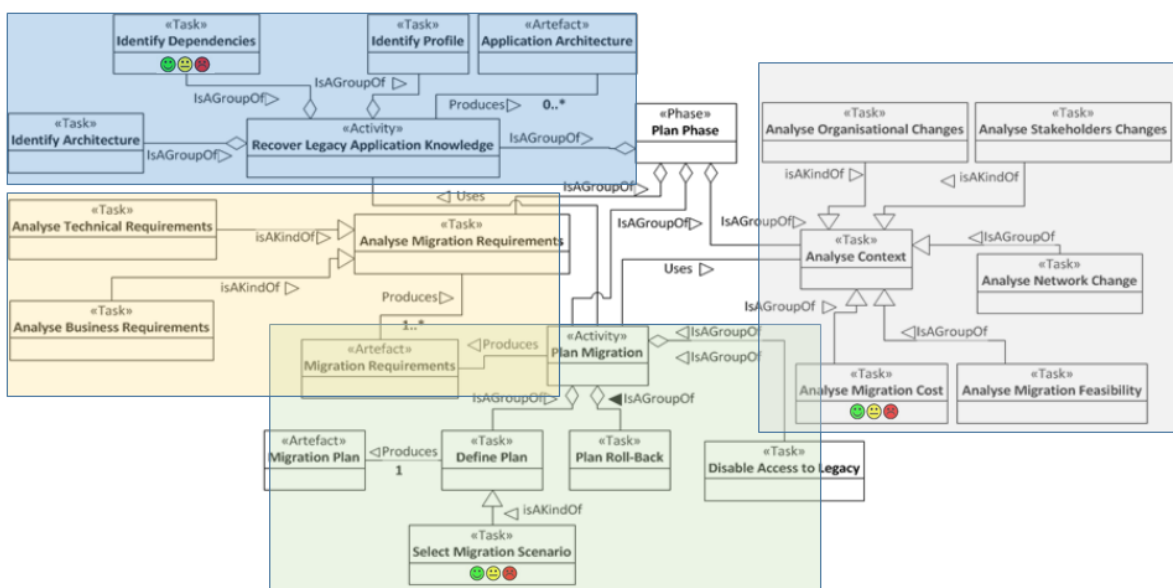


Figure 4: Reference Model's Plan phase with visual indicative of task evaluation set.

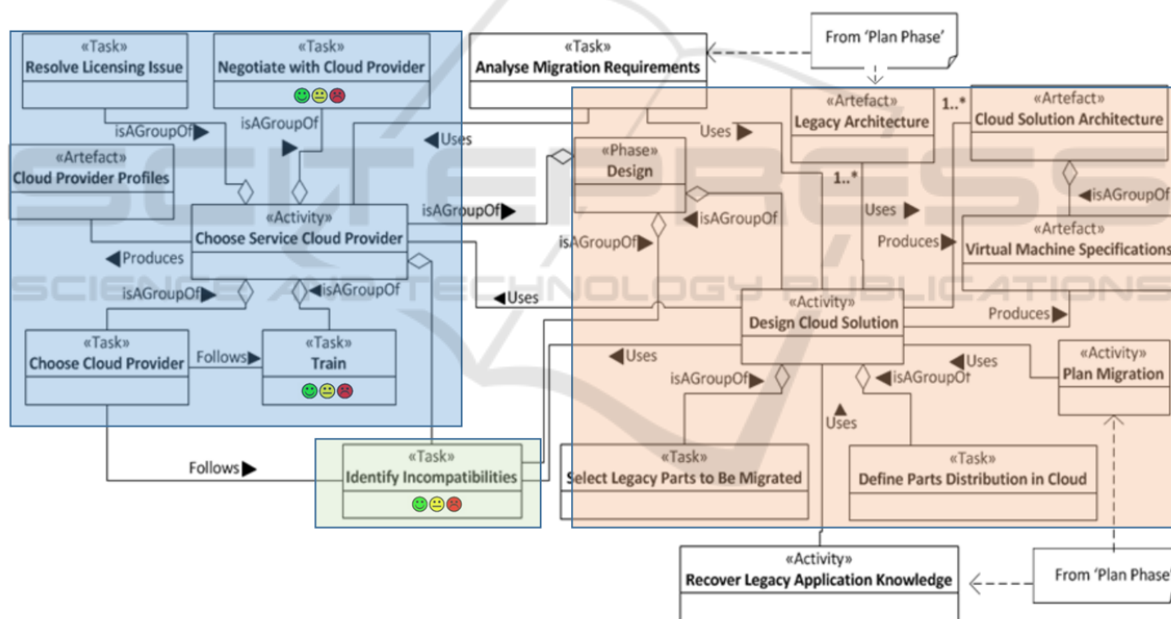


Figure 5: Reference Model's Design phase with visual indicative of task evaluation set.

higher need for training may indicate a higher risk in managing the migration project of a given legacy system. According to (Reza Bazi et al., 2017) organizations must extend their cloud knowledge as a way to guarantee a successful start. The objective of Choose Cloud Provider (activity that aggregates both Negotiate with the cloud provider and Train tasks) is to find the best providers that meet migration requirements (Gholami et al., 2017).

**Identifying incompatibilities** - This task is re-

sponsible to find the incompatibilities between the legacy system and the set of CC services defined in the design phase as being required to run the system after migration to the cloud. These incompatibilities will require specific effort to be solved. The objective of the activity Identifying incompatibilities is to estimate effort and cost to resolve incompatibilities (Gholami et al., 2017).

The above explanations of what motivated task selection for the government domain are illustrated in a

Table 3: Rating of risk perception associated with tasks on evaluation set (shaded lines:Plan phase; white lines:Design phase).

Task	Weight	Rate	Reason
Analyze migration cost	3	5	Low complexity system and low estimated migration cost.
Identify dependencies	3	4	System that has almost no dependency.
Select migration scenario	1	4	Scenario type V, lift and shift.
Negotiate with cloud provider	3	4	Providers selected.
Train	1	2	High necessity of training. No experience on CC.
Identify incompatibilities	1	5	No incompatibilities found.

detailed view of Reference Model phases with a visual identification of the tasks in Figures 4 and 5. The figures show how each key element is associated to the others. Each key element has been associated with a color, which covers the tasks that compound that key element. In addition, there is a specific indication (three colored smileys) in the tasks that are part of the evaluation set.

### 4.3 Use Case

To show the method's applicability, we chose a legacy system in use by a government body that decided to migrate its legacy systems to the cloud. The legacy system is called Civic Cloud and exposes, on the Internet, updated data about educational and health institutions around the country and data about medicines that are authorized by the proper government agency. Although the system has cloud on its name, it runs on local infrastructure.

The system is composed by a set of loosely coupled web services that can be used by developers and companies to add value to their applications. For instance, one can develop and publish a mobile application that automatically captures user location and suggests the health institution that is closer to the user. The system was developed using Java language with the support of Spring MVC framework. It is running on application server JBOSS EAP, sharing computational resources with other legacy systems. The database management system is Oracle (Oracle, 2017).

As the system is publicly available, any developer can build an application that makes calls to its services and uses the provided data. If one of these applications becomes a killer application, with hundreds or thousands of transactions by hour, there is a chance that the computing infrastructure becomes inadequate due to the lack of elasticity.

We worked with the team responsible by the system in order to apply the reference model for migration to the cloud. After doing the work prescribed on initial phases, Plan and Design, we applied the method described on the Section 3, using the tasks

and weights defined on the Section 4.

The team shared their perception that the system was a good choice to be the first legacy to be migrated by this government body. This is explained due to the low number of integrations with other legacy systems and to the low complexity and size of system code.

The team then rated the risk perception for each task on the evaluation set of both phases. The values can be viewed on table 3, along with the main reason that justified the rating.

Using the weights defined on tables 1 and 2, the rates given by the team (table 3) and applying equations (2), (3) and (1), we have:  $Pl = 4.43$ ,  $De = 3.80$  and  $EvF = 4.11$ . The calculated value of  $EvF$  is the score of risk perception to migrate legacy system Civic Cloud to the cloud with the use of Reference Model. As suggested on Section 3,  $T1$  and  $T2$  are defined to be 3.0. Since  $EvF$ ,  $Pl$  and  $De$  are higher than these thresholds, this indicates a low risk perception for this system migration.

## 5 CONCLUSIONS AND FUTURE WORK

This paper presented a review of software migration strategies, compared three conceptual migration models and selected a reference model as a basis for legacy systems migration to the cloud. Based on the concepts, we proposed a method for calculating a score of perceived risk to the systems that will be migrated to the cloud. Method's steps are described along with a proof of concept in the government domain. The proposed method can be used to rank the systems to be migrated to the cloud, offering an opportunity to assess migration risk perception before migration execution.

The future work of this research aims to apply the method on other government systems and collect the results in different scenarios. Also, these results could be validated with a survey, applied over different government organizations, to improve and verify the proposed method.

## REFERENCES

- Andrikopoulos, V., Binz, T., Leymann, F., and Strauch, S. (2013). How to adapt applications for the cloud environment. *Computing*, 95(6):493–535.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and others (2010). A view of cloud computing. *Communications of the ACM*, 53(4):50–58.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., Lee, G., Patterson, D. A., Rabkin, A., Stoica, I., and others (2009). Above the clouds: A Berkeley view of cloud computing. Technical report, EECS Department, University of California, Berkeley.
- AWS (2017). Amazon Web Services. [online]. available from: <https://aws.amazon.com/>.
- Babar, M. A. and Chauhan, M. A. (2011). A tale of migration to cloud computing for sharing experiences and observations. In *Proceedings of the 2nd international workshop on software engineering for cloud computing*, pages 50–56. ACM.
- Binz, T., Leymann, F., and Schumm, D. (2011). CMotion: A Framework for Migration of Applications into and between Clouds. In *Service-Oriented Computing and Applications (SOCA), 2011 IEEE International Conference on*, pages 1–4. IEEE.
- Fettke, P. and Loos, P. (2003). Multiperspective evaluation of reference models—towards a framework. *Conceptual modeling for novel application domains*, pages 80–91.
- Gartner (2017). Developing a Public Cloud IaaS Adoption and Migration Framework. [online]. available from: <https://www.gartner.com/doc/3645347/developing-public-cloud-iaas-adoption>.
- Gholami, M. F., Daneshgar, F., Beydoun, G., and Rabhi, F. (2017). Challenges in migrating legacy software systems to the cloud—an empirical study. *Information Systems*, 67:100–113.
- Gholami, M. F., Daneshgar, F., Low, G., and Beydoun, G. (2016). Cloud migration process—A survey, evaluation framework, and open challenges. *Journal of Systems and Software*, 120:31–69.
- Ionita, A. D. (2013). Introduction to the migration from legacy applications to service provisioning. *Migrating Legacy Applications: Challenges in Service Oriented Architecture and Cloud Computing Environments*, pages 1–11.
- ISACA (2012). Guiding principles for cloud computing adoption and use. [online]. available from: <http://www.isaca.org/Knowledge-Center/Research/Pages/Cloud.aspx>.
- ISO/IEC (2006). Systems and Software Engineering- Software Life Cycle Processes - Maintenance, ISO/IEC 14764:2006.
- Jamshidi, P., Ahmad, A., and Pahl, C. (2013). Cloud migration research: a systematic review. *IEEE Transactions on Cloud Computing*, 1(2):142–157.
- Kundra, V. (2011). Federal Cloud Computing Strategy. Washington: The White House.
- Microsoft (2017). Microsoft Azure. [online]. available from: <https://www.azure.com/>.
- Oracle (2017). Oracle Database. <https://www.oracle.com/database/index.html>.
- Pahl, C. and Xiong, H. (2013). Migration to PaaS clouds-migration process and architectural concerns. In *Maintenance and Evolution of Service-Oriented and Cloud-Based Systems (MESOCA), 2013 IEEE 7th International Symposium on the*, pages 86–91. IEEE.
- PMI (2013). A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Fifth Edition.
- Reza Bazi, H., Hassanzadeh, A., and Moeini, A. (2017). A comprehensive framework for cloud computing migration using meta-synthesis approach. *Journal of Systems and Software*, 128:87–105.
- Sun, K. and Li, Y. (2013). Effort estimation in cloud migration process. In *Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on*, pages 84–91. IEEE.
- Zhao, J.-F. and Zhou, J.-T. (2014). Strategies and methods for cloud migration. *international Journal of Automation and Computing*, 11(2):143–152.