DOI:

A SYSTEMATIC REVIEW OF LITERATURE APPLIED TO THE RECOVERY OF DATA STORED IN CLOUD

UMA REVISÃO SISTEMÁTICA DA LITERATURA APLICADA À RECUPERAÇÃO DE DADOS ARMAZENADOS EM NUVEM

Luciano Ribeiro Duarte
UNIVERSIDADE FEDERAL DO PARÁ - ORCID: https://orcid.org/0000-0001-6519-3198
Erlon Fonseca Pinheiro
UNIVERSIDADE FEDERAL DO PARÁ - ORCID: https://orcid.org/0000-0001-7778-477X
Josivaldo De Souza Araújo
UNIVERSIDADE FEDERAL DO PARÁ - ORCID: https://orcid.org/0000-0001-6890-6923

Abstract
In order to identify and present the existing approaches, this paper aims to describe a Systematic Literature Review applied to recovery data on cloud servers.
However, data can be deleted on purpose or accidentally as in any storage device. Therefore, there is a need to recover them. The used data recovery techniques are almost always ineficiente on these devices when they are applied to stored data in the cloud.
Systematic Literature Review
The data show that 3,374 reviewed studies, 13 ones will be considered the inclusion criteria where the main used tools were presented and analyzed for data retrieval.
To identify and present the existing approaches to recovery data on cloud servers.
To identify and present the existing approaches to recovery data on cloud servers.

Key words: Cloud Computing, Data Recovery, Systematic Literature Review, Forensic Computing, Digital Data

Resumo
Com o objetivo de identificar e apresentar as abordagens existentes, esse trabalho descreve uma Revisão Sistemática da Literatura aplicada à recuperação de dados em servidores na nuvem.
Porém, como em todo dispositivo de armazenamento, os dados podem ser deletados de forma proposital, ou acidentalmente, havendo, com isso, a necessidade de recuperá-los. As técnicas utilizadas de recuperação de dados nestes dispositivos, mostram-se, quase sempre, ineficientes quando aplicadas em dados armazenados na nuvem.
Revisão Sistemática da Literatura
Dos 3.374 estudos retornados, 13 estudos passaram nos critérios de inclusão, onde foram apresentadas as principais ferramentas utilizadas para a recuperação de dados.
Apresentar as abordagens existentes à recuperação de dados em servidores na nuvem.
Apresentar as abordagens existentes à recuperação de dados em servidores na nuvem.

Palavras-chave: Computação na Nuvem, Recuperação de Dados, Revisão Sistemática da Literatura, Literatura, Computação Forense, Dados Digitais

UMA REVISÃO SISTEMÁTICA DA LITERATURA APLICADA À RECUPERAÇÃO DE DADOS ARMAZENADOS EM NUVEM

A SYSTEMATIC LITERATURE REVIEW APPLIED TO THE RECOVERY AND STORED DATA IN CLOUD

ABSTRACT: *Regardless of location, accessing to information has been something necessary and essential nowadays. Besides, acessing to personal files by mobile devices have become essential as well. For this purpose, cloud storage and processing services have evolved into a reliable technology and virtually unlimited in some ways. However, data can be deleted on purpose or accidentally as in any storage device. Therefore, there is a need to recover them. The used data recovery techniques are almost always ineficiente on these devices when they are applied to stored data in the cloud. In order to identify and present the existing approaches, this paper aims to describe a Systematic Literature Review applied to recovery data on cloud servers. The data show that 3,374 reviewed studies, 13 ones will be considered the inclusion criteria where the main used tools were presented and analyzed for data retrieval.*

Keywords: *Cloud Computing, Data Recovery, Systematic Literature Review, Forensic Computing.*

RESUMO: O acesso à informação, independentemente da localização, nos dias atuais, é algo necessário e essencial, e o acesso aos arquivos pessoais através dos dispositivos móveis, também se tornou algo imprescindível. Para isso, os serviços de armazenamento e processamento em nuvem evoluíram, tornando-se uma tecnologia confiável, e em muitos casos, virtualmente ilimitada. Porém, como em todo dispositivo de armazenamento, os dados podem ser deletados de forma proposital, ou acidentalmente, havendo, com isso, a necessidade de recuperá-los. As técnicas utilizadas de recuperação de dados nestes dispositivos, mostram-se, quase sempre, ineficientes quando aplicadas em dados armazenados na nuvem. Com o objetivo de identificar e apresentar as abordagens existentes, esse trabalho descreve uma Revisão Sistemática da Literatura aplicada à recuperação de dados em servidores na nuvem. Dos 3.374 estudos retornados, 13 estudos passaram nos critérios de inclusão, onde foram apresentadas e analisadas, as principais ferramentas utilizadas para a recuperação de dados.

Palavras-chave: Computação na Nuvem, Recuperação de Dados, Revisão Sistemática da Literatura, Computação Forense.

1. INTRODUÇÃO

O paradigma da computação em nuvem se consolidou nas últimas décadas, e trouxe impactos que excederam as estimativas globais, alcançando cada vez mais, novos clientes, mercados e serviços. Isso porque, as informações dos usuários são armazenadas em servidores de armazenamento compartilhados, em vez dos próprios discos rígidos dos usuários. Nessas condições, os usuários podem solicitar aumento do espaço de armazenamento, sempre que houver a necessidade, e podem também, acessar seus dados de uma forma remota e segura (Alsadhan, 2018).

O *National Institute of Standards and Technology* (NIST) é a agência reguladora de tecnologia dos Estados Unidos e define a Computação em Nuvem como sendo um modelo

para permitir acesso de rede onipresente, conveniente e sob demanda a um *pool* compartilhado de recursos de computação configuráveis, e que podem ser rapidamente provisionados e liberados, com esforço mínimo de gerenciamento, ou interação com o provedor de serviços (Ashraf, 2014).

O armazenamento de dados em nuvem é uma das demandas mais debatidas entre os profissionais da computação forense digital na contemporaneidade, isso porque, trouxe vários desafios, entre eles, a identificação do proprietário de um arquivo deletado, a alocação dinâmica e eficiente da distribuição dos dados, a integridade e a segurança dos dados armazenados, podem ser citados. A recuperação de dados nesses ambientes, além de necessitar de ferramentas especializadas para contribuir na análise, extração e recuperação de dados, pode necessitar, também, do trabalho conjunto de diversos profissionais, como especialistas em direito digital, peritos criminais e profissionais de Tecnologia da Informação (TI) (Mishra et al. 2012).

Nesse sentido, este trabalho apresenta uma Revisão Sistemática da Literatura com o objetivo de identificar abordagens e relatos de experiências envolvendo, especificamente, o a recuperação de dados em nuvem. Para isso, este trabalho é composto por cinco seções. A seção 2, relata os trabalhos relacionados ao tema proposto, bem como, a justificativa para o seu desenvolvimento. Na seção 3, é descrita a metodologia utilizada na Revisão Sistemática da Literatura. Já na seção 4, são apresentados os resultados, e as respostas às questões de pesquisa. E na seção 5, são discutidas as considerações finais e as propostas para o desenvolvimento de trabalhos futuros.

2. TRABALHOS RELACIONADOS E JUSTIFICATIVA.

No trabalho de (Gokulakrishnan and Gnanasekar, 2020) uma nova metodologia é proposta para recuperação e gerenciamento de dados, visando garantir escalabilidade de alto nível e alta confiabilidade da solicitação, fornecendo reconhecimento de falha e sistemas baseados em nuvem com tolerância a falhas. Utiliza-se a segmentação e geração de tokens para a divisão de dados, adicionando o endereço da nuvem ou locais de armazenamento. Dessa forma, o segmento que falta de qualquer nó com defeito é facilmente reconhecido dentro de um curto intervalo dos limites e obterá o backup de dados dos nós vizinhos.

O estudo de (Sorini and Scott 2020) apresenta o PYLOCKY, que é o nome de um *ransomware* desenvolvido recentemente, e que foi analisado com o objetivo de fornecer uma visão geral das ferramentas existentes que podem ajudar empresas a se recuperarem de um ataque desse tipo. Também, é explicado as limitações e princípios de funcionamento das ferramentas de recuperação, além de uma análise do código fonte do Pylocky, para apresentar falhas gerais na implementação de protocolos criptográficos que devem ser evitadas pelos desenvolvedores.

O trabalho de (Alsadhan, 2018) debate e apresenta uma solução para o reconhecimento de propriedade de um arquivo depois que o mesmo é excluído. Esse problema é decorrente de que quando um arquivo é excluído, o sistema gerenciador da nuvem também exclui os seus metadados, e sem essas informações, ninguém pode mais saber quem é o proprietário do arquivo excluído. E torna-se ainda mais dificultoso identificar na nuvem, quem era o verdadeiro proprietário do arquivo. As dificuldades aumentam, para os profissionais forenses, quando um arquivo excluído é utilizado como prova contra um suspeito de crime.

O trabalho de (Mohite and Ardhapurkar, 2015), propõe uma ferramenta forense computacional para investigação em ambiente de nuvem, incluindo recuperação de dados.

O CBCFT (*Cloud Based Computer Forensic Tool*) permite que um profissional da área acesse a ferramenta, a partir do portal da nuvem, em sua máquina ou em qualquer máquina. A ferramenta ainda oferece vantagens, como: análise de uma grande quantidade de informações de forma rápida e eficiente; uso otimizado de mídia de armazenamento; mínimo de interrupção, oferecendo serviços de computação em nuvem ininterruptos para o usuário e a manutenção da infraestrutura é simplificada.

No estudo (Pichan et al. 2015), é realizado um levantamento sistemático dos desafios forenses na computação em nuvem e suas soluções e desenvolvimentos mais recentes. Diferentemente das pesquisas existentes sobre o tema, os problemas da computação em nuvem são abordados usando as fases de forense digital tradicional como base. Para cada fase do processo forense digital, foi incluída uma lista de desafios e análises de suas possíveis soluções. O trabalho procura evidenciar as diferenças entre os problemas e soluções para ambientes tradicionais e análise forense digital em nuvem.

2.1. Justificativa

A Revisão Sistemática tem como objetivo identificar as abordagens existentes para a recuperação de dados em servidores que oferecem o armazenamento de dados em nuvem, como um serviço. Com a identificação dessas abordagens, será possível avaliar os métodos e as práticas utilizadas. Desta forma, têm-se a seguinte estrutura, conforme proposto em (Pereira, 2019):

- ✓ **Analisar:** publicações científicas (trabalhos técnicos e relatos de experiências) através de um estudo baseado em revisão sistemática;
- ✓ **Com o propósito de:** identificar abordagens (métodos, técnicas, práticas e ferramentas) existentes que apoiem a Computação Forense na recuperação de dados na nuvem.
- ✓ **Com relação ao:** as práticas de ferramentas que auxiliem a recuperação de dados em nuvem;
- ✓ **Do ponto de vista:** de professores, pesquisadores e profissionais da área.
- ✓ **No contexto:** acadêmico e comercial.

3. REVISÃO SISTEMÁTICA DA LITERATURA (RSL)

A principal meta de uma RSL consiste em realizar pesquisa exaustiva na literatura, em busca de evidências que possam apoiar uma determinada hipótese, ou simplesmente a busca por conhecimento aprofundado acerca de certo fenômeno de interesse. Para tal, a revisão sistemática faz uso de estudos previamente publicados e validados pertinentes ao tópico de interesse: os estudos primários, estudos de natureza experimental que envolvem hipóteses e resultados obtidos com pesquisas e experimentação, a partir de diferentes métodos, como *surveys*, estudo de caso e experimentos (Mafra e Travassos, 2006).

A Revisão Sistemática consiste em um estudo secundário, por utilizar como base estudos primários previamente publicados. Assim, pode ser feita a integração de diversos estudos experimentais, de forma a comparar seus resultados, visto que nenhum estudo individualmente pode ser considerado definitivo (Mafra e Travassos, 2006), sendo necessária a confirmação de resultados obtidos a partir da análise de um número maior de estudos.

3.1. Fatores da Pesquisa

Para a RSL, neste artigo, foram definidas as seguintes perguntas bibliométricas e de pesquisa:

- Questões Bibliométricas (QB):
  - **QB.1**: Qual é a quantidade de trabalhos incluídos por indexador de estudos?
  - **QB.2**: Qual é a quantidade de estudos retornados por ano?
  - **QB.3**: Qual é a quantidade de estudos retornados por país?
  - **QB.4:** Quais foram os autores com maior quantidade de publicações?
  - **QB.5**: Quantidade de estudos por tipo de estudo? (Teóricos, experimentais, RSL e relatos de experiências comerciais).
  - **QB.6**: Qual a quantidade de estudos experimentais, por tipo de experimentação, estudo de caso, pesquisa ação, pesquisa de campo e etnográfico?
  - **QB.7:** Qual é a quantidade de estudos retornados por tipo de publicação? (Conferência, periódico ou workshop).
  - **QB.8:** Qual é a quantidade de estudos retornados por veículos de publicação?

- Questão de Pesquisa (QP):
  - **QP:** Quais são as abordagens existentes para a recuperação de dados, que apoiam as atividades da Computação Forense, no contexto do armazenamento em nuvem?

A questão levantada foi organizada conforme a estrutura *Population, Intervention, Context, Outcomes, Comparison* (PICOC), recomendada por (Kitchenham, 2007). Entretanto, apenas os itens População, Intervenção, Contexto e Resultados foram considerados relevantes para a pesquisa. Tal restrição, segundo (Pereira, 2019), caracteriza esta pesquisa como uma Revisão QUASI Sistemática da Literatura.

Na questão de pesquisa, objetiva-se em identificar abordagens utilizadas por professores e pesquisadores (População) existentes, na Computação Forense, para recuperar dados na nuvem (Intervenção), na qual, busca-se encontrar métodos, práticas e metodologias e ferramentas que possam ser empregadas para realizar essa recuperação (Resultados), (Santos, 2010). Logo, definiu-se a seguinte estrutura de pergunta de pesquisa, de acordo com o Quadro 1.

Quadro 1 - Estrutura da Questão de Pesquisa

População (*P*):	Professores, pesquisadores e profissionais da área.
Intervenção (*I*):	Sistemas de armazenamento na nuvem.
Contexto (C):	Recuperação de dados
Resultados (*O*)	Métodos, Práticas, Técnicas, Procedimentos, Abordagem e Ferramentas

Fonte: Elaborado pelo autor (2021).

Um conjunto de Questões Secundárias (QS) referentes à questão principal foram estabelecidas, questões estas para serem respondidas durante a fase de extração de informações. Tais questões têm o objetivo de esclarecer detalhes importantes que esta revisão procura identificar, para colaborar com o projeto onde este se insere:

- **QS.1:** Quais as atividades são apoiadas pelos sistemas de armazenamento de dados na nuvem?

- **QS.2:** Quais técnicas ou práticas são utilizadas para dar segurança aos dados em um ambiente de armazenamento na nuvem?
- **QS.3:** Quais técnicas são utilizadas pelos sistemas de armazenamento na nuvem para criptografar os dados?
- **QS.4:** Quais as técnicas e ferramentas utilizadas para a recuperação de dados em nuvem no contexto da Computação Forense?
- **QS.5:** Qual o contexto de aplicação da abordagem encontrada?
- **QS.6:** Quais os profissionais que utilizam ferramentas forenses para recuperar dados na nuvem?
- **QS.7:** Quais são as leis que tratam sobre a recuperação de dados em provedores de nuvem no Brasil?
- **QS.8:** Quais acordos ou cooperações internacionais tratam da recuperação de dados envolvendo o Brasil?
- **QS.9:** Quais os sistemas de arquivos usados na arquitetura de armazenamento de dados na nuvem?
- **QS.10:** Quais os sistemas operacionais utilizados nos provedores de armazenamento de dados na nuvem?
- **QS.11:** Quais as principais empresas que oferecem serviços de armazenamento de dados na nuvem?

3.2. Palavras-Chave

A partir das questões de pesquisa, palavras-chave foram identificadas em acordo com a estrutura População, Intervenção, Contexto e Resultados para a posterior formulação da *string* de busca. O Quadro 2, apresenta os componentes e os termos que compõem as palavras-chave utilizadas para responder a QP.1 (questão de pesquisa) e as QS (questões secundárias).

Quadro 2 - Construção da Palavra-chave para as questões principais

Palavras-chave	Termos
População	**Inglês:** *Forensic Experts,Cloud Solution Provider (CSP), Forensic Professionals,Digital Law Specialists, Forensic Software Company, Investigation Agencies,Cloud Criminal Investigations, Expert Reports..* **Português:** Peritos Forenses, Provedores de Soluções na Nuvem (CSP), Profissionais Forenses, Especialistas em Direito Digital, Empresa de Softwares Forenses, Investigações Criminais na Nuvem, Laudos Periciais.
Intervenção	**Inglês:** *Cloud Storage Systems, Data Carving, Data Protection Laws, International Cooperation, File Systems, Encryption, Cloud Forensic, Information Security, Digital Media, Data Mining, Operation Systems.* **Português:** Sistemas de Armazenamento, Esculpimento de Dados, Leis de Proteção dos Dados, Cooperação Internacional, Sistemas de Arquivos, Sistemas Operacionais, Criptografia, Segurança da Informação, Mídias Digitais, Mineração de Dados.
Contexto	**Inglês:** *Data Recovery.*

Palavras-chave	Termos
	Português: Recuperação de Dados
Resultados	**Inglês:** *Method, Practice, Technique, Methodology, Tool, Approach.* **Português:** Metodologia, Técnica, Prática, Método, Ferramenta, Abordagem;

Fonte: Elaborado pelo autor (2021).

3.3. Fontes de Pesquisa

Para a seleção das fontes de pesquisa, foram definidos os seguintes critérios:

- Disponibilidade para consultas web;
- Disponibilidade para busca de artigos através do domínio da Universidade;
- Disponibilidade de artigos na íntegra através do domínio da Universidade ou a partir da utilização da *engine* de busca Google ou Google Scholar ou Portal CAPES;
- Disponibilidade de artigos em inglês ou português;
- Que possuam máquinas de busca;

Sendo assim, as fontes definidas para a extração de dados dos estudos primários são as bases digitais, apresentadas no Quadro 3.

Quadro 3 – Fontes de Pesquisa.

Bases
IEEEXplore Digital Library
Ei Compendex
Scopus
ISI Web of Knowledge
Science Direct

Fonte: Elaborado pelo autor (2021).

3.4. *String* de Busca

As *strings* de busca foram adaptadas segundo as máquinas de buscas de cada fonte de pesquisa. No Quadro 4, são apresentadas as variações das *strings* utilizadas com base nas palavras-chave definidas.

Quadro 4 - Construção da *string* de busca.

Fonte	*String de busca*
IEEEXplore Digital Library	*((“Forensic*”AND (“Expert” OR “Professional” OR “Software Company”) OR (”Cloud*” AND (“Solution Provider” OR “Criminal Investigation”) OR “Investigation Agencies” OR “Expert Reports”))) AND ((“Cloud*” AND (“Storage System”) OR (“Data*” AND(“Carving” OR “Protection Laws ” OR “Mining”)) OR “International Cooperation” OR “File Systems” OR “Encryption” OR*

Fonte	***String de busca***
	“Information Security” OR “Digital Media” OR “Operation Systems”)) AND (“Procedure” OR “Methodology” OR “Method” OR “Technique” OR “Practice” OR “Tool” OR “Approach OR “Context” OR "Data Recovery" OR "Cloud Forensic Challenges" OR "Computer Forensics")
El Compendex	*Forensic AND (Experts OR Professionals OR Software Company OR Investigation Agencies OR Expert Reports OR Cloud Criminal Investigations OR Solution Provider OR Digital Law Specialists) AND (Cloud Storage Systems OR Data Carving OR Protection Laws OR Mining OR International Cooperation OR File OR Operating OR Systems OR Encryption OR Cloud Forensic OR Information Security OR Digital Media) AND (Procedure OR Methodology OR Method OR Technique OR Practice OR Tool OR Approach)*
Scopus	*Forensic AND (Experts OR Professionals OR Software Company OR Investigation Agencies OR Expert Reports OR Cloud Criminal Investigations OR Solution Provider OR Forensic OR Digital Law Specialists) AND (Cloud Storage Systems OR Data Carving OR Protection Laws OR Mining OR International Cooperation OR File OR Operating OR Systems OR Encryption OR Cloud Forensic OR Information Security OR Digital Media) AND (Procedure OR Methodology OR Method OR Technique OR Practice OR Tool OR Approach)*
ISI Web of Knowledge	*(("Forensic*" AND ("Experts" OR "Professionals" OR "Software Company" OR "Investigation Agencies" OR "Expert Reports" OR "Cloud" OR "Criminal Investigations" OR "Solution Provider" OR "Digital Law Specialists")) AND (("Cloud Storage Systems" OR "DATA" "Carving" OR "Protection Laws" OR "Mining" OR "International Cooperation" OR "File" OR "Operating" OR "Systems" OR "Encryption" OR "Cloud Forensic" OR "Information Security" OR "Digital Media")) AND (("Procedure" OR "Methodology" OR "Method" OR "Technique" OR "Practice" OR "Tool" OR "Approach")))*
Science Direct	*((“Forensic” AND (“Expert” OR “Professional” OR “Software Company”) AND “Cloud” AND “Solution Provider” OR “Criminal Investigation” OR “International Cooperation” OR “Cloud Forensic Challenges”)*

Fonte: Elaborado pelo autor (2021).

3.5. Escopo e Restrições de pesquisa

A pesquisa possui um escopo que obedece às restrições definidas no Quadro 5, que asseguram a viabilidade da pesquisa (Pereira, 2019).

Quadro 5 - Escopo e Restrições de pesquisa

Escopo	Restrições
▪ Artigos disponíveis na web e acessíveis por meio da rede de domínio da Universidade cuja a busca possa ser realizada via *engine* da *Google* e *Google Scholar;* ▪ Disponibilidade de artigos em Inglês e Português; ▪ Uso de mecanismos de busca utilizando palavras-chave;	▪ Os trabalhos devem mencionar pelo menos uma das palavras chaves propostas; ▪ A pesquisa não deve ocorrer em ônus financeiro aos pesquisadores, ou seja, deve-se selecionar apenas trabalhos que tenham acesso gratuito; ▪ A pesquisa deve estar restrita aos resultados publicados entre 01 de janeiro de 2006 a 31 de dezembro de 2020.

Fonte: Elaborado pelo autor (2021).

3.6. Critérios de Inclusão e Exclusão

Os critérios de inclusão e exclusão servem para avaliar a qualidade de um artigo científico, e assim criar uma lista de possíveis artigos primários selecionados, e outra, com os artigos excluídos (Pereira, 2019). Os critérios utilizados nesta pesquisa foram definidos pelos pesquisadores envolvidos nesta Revisão Sistemática. Assim, o Quadro 6, apresenta os critérios de Inclusão utilizados, e o Quadro 7, apresenta os critérios de Exclusão definidos para a RSL.

Quadro 6 – Critérios de Inclusão

Critérios de Inclusão
CI.01. Estudos que apresentem no seu contexto, de forma primária ou secundária, abordagens legais que apoiem o processo da recuperação de dados na nuvem;
CI.02. Estudos com abordagens de recuperação de dados e técnicas aplicada sem pesquisas de caráter experimental ou teórico na área de Computação Forense em ambiente de nuvem, contanto que apresentem exemplos de aplicação, descrição de experimentos ou casos reais de abordagens;
CI.3: Estudos que claramente sejam relevantes para a pesquisa deste trabalho;

Fonte: Elaborado pelo autor (2021).

Quadro 7 – Critérios de exclusão

Critérios de Exclusão
CE.01. Estudos que não estejam disponíveis livremente para consulta ou download (em versão completa) através das fontes de pesquisa ou através das ferramentas de busca Google (http://www.google.com.br/) e/ou Google Scholar (http://scholar.google.com.br/);
CE. 02. Estudos que claramente não atendam as questões de pesquisa;
CE. 03. Estudos repetidos (em mais de uma fonte de busca) terão apenas sua primeira ocorrência considerada;
CE. 04. Estudos enquadrados como resumos, *keynote speeches*, cursos, tutoriais, workshops e afins;
CE. 05. Estudos que não mencionem as palavras-chave da pesquisa no título, resumo ou nas palavras-chave do artigo, salvo trabalhos que abordem o processo de recuperação de dados, nos quais seja observada possibilidade da Computação Forense e técnicas computacionais a serem tratados ao longo do trabalho
CE.06. Estudos que não estejam inseridos no contexto da Computação Forense em nuvem;
CE. 07. Estudos que não estiverem apresentados nos idiomas aceitos (Português e Inglês);

Fonte: Elaborado pelo autor (2021).

3.7. Avaliação da Qualidade

A avaliação da qualidade de um artigo possibilita com que trabalhos estejam precisamente alinhados com os objetivos da RSL proposta, que tenham uma maior contribuição para as questões de pesquisa, e tenham maior notoriedade entre os estudos primários levantados. Ou seja, a avaliação da qualidade de um artigo científico é a mensuração de sua relevância e conteúdo. Esta avaliação é um dos critérios de inclusão ou exclusão aplicados aos estudos durante a seleção. Ao minimizar o viés da pesquisa, assegura-se a validação interna e externa (Kitchenham, 2007). Portanto, no Quadro 8, são apresentados os critérios de avaliação da qualidade dos estudos primários, adaptados de (Costa, 2010):

Como é possível notar, os critérios (1) a (4) são genéricos, ou seja, aplicam-se a todos os estudos primários avaliados, enquanto os critérios (5) a (7) são específicos, aplicam-se especificamente aos respectivos tipos de trabalhos mencionados.

3.8. Processo de Avaliação dos Estudos Primários

Os estudos primários selecionados são lidos em totalidade e então são avaliados quanto aos critérios de qualidade. Para avaliar o grau de adequação aos critérios de qualidade, foi adotada uma estratégia de avaliação semelhante à proposta por (Costa, 2010), onde se utiliza a escala de *Likert*-5, permitindo respostas gradativas de 0 (discordo totalmente) à 4 (concordo totalmente). Porém, foi utilizada a escala *Likert*-3, permitindo respostas gradativas de 0 (discordo totalmente) à 2 (concordo totalmente), pois com menos itens de *Likert* os critérios se tornam menos subjetivos. Para auxiliar a avaliação, seguindo a escala de *Likert*-3 para cada critério de qualidade, foram definidas escalas. Tais níveis são apresentados a seguir:

Quadro 8 – Critérios de Avalição da Qualidade dos Estudos Primários.

Critérios de Avaliação
1. Introdução/Planejamento a. Os objetivos ou questões do estudo são claramente definidos (incluindo justificativas para a realização do estudo)? b. O tipo de estudo está definido claramente?
2. Desenvolvimento a. Existe uma clara descrição do contexto no qual a pesquisa foi realizada? b. O trabalho é adequadamente referenciado (apresenta trabalhos relacionados ou semelhantes e, baseia-se em modelos e teorias da literatura)?
3. Conclusão a. O estudo relata de forma clara e não ambígua os resultados? b. Os objetivos ou questões do estudo são alcançados?
4. Critérios para a Questão de Investigação a. O estudo lista primária ou secundariamente as metodologias ou métodos ou técnicas ou práticas para apoiar as atividades de Recuperação de Dados? b. O estudo endereça explicitamente as abordagens para alguma atividade de Recuperação de Dados?
5. Critério Específico para estudos Experimentais a. Existe um método ou um conjunto de métodos descrito para a realização do estudo?
6. Critério Específico para estudos Teóricos a. Existe um processo não tendencioso na escolha dos estudos?
7. Critério Específico para Relato de Experiência Comerciais a. Existe uma descrição sobre a(s) instituição (ões) onde foi conduzido o estudo?

Fonte: Elaborado pelo autor (2021).

- Concordo totalmente (2): deve ser concedido no caso em que o trabalho apresente no texto os critérios que atendam totalmente a questão;
- Neutro (1): deve ser concedido no caso em que o trabalho não deixa claro se atende ou não a questão;
- Discordo totalmente (0): deve ser concedido no caso em que não existe nada no trabalho que atenda aos critérios da questão.

É definido uma escala de avaliação para cada critério de qualidade previamente estabelecido (Pereira, 2019). O Quadro 9, apresenta a escala utilizada para cada critério de qualidade.

Quadro 9 – Escala de *Likert -3* por critério

Critério	Escala
1a.	2 – Define e justifica o estudo claramente. 1 – Define claramente o estudo, mas não justifica. 0 - Não define os objetivos e nem justifica o estudo.
1b.	2 – Define o tipo de estudo, referenciado na literatura a metodologia. 1 – Define o tipo de estudo, porém sem referenciar a metodologia. 0 – Não é possível inferir o tipo de estudo.
2a.	2 – Define claramente uma seção com o contexto da pesquisa. 1 – O contexto da pesquisa está disperso ao longo do texto. 0 – O contexto da pesquisa não é abordado.
2b.	2 – O texto apresenta uma seção de trabalhos relacionados. 1 – O texto apresenta trabalhos relacionados dispersos ao longo do texto. 0 – O texto não apresenta trabalhos relacionados nem se apoia na literatura.
3a.	2 – Resultados são claramente apresentados na seção de conclusão. 1 – Resultados apresentados na conclusão não são claros. 0 – Não são apresentados resultados.
3b.	2 – Os resultados estão totalmente aderentes ao objetivo do estudo. 1 – Os resultados são parcialmente aderentes ao objetivo do estudo. 0 – Não é alcançado nenhum resultado.
4a.	2 – Alguns dos elementos é claramente descrito. 1 – Alguns dos elementos é avaliado, porém não descrito. 0 – Nenhum dos elementos é apresentado direta ou indiretamente.
4b.	2 – Endereça a abordagem para alguma atividade de forma detalhada (o nome, o que representa e como se usa a abordagem). 1 – Endereça a abordagem para alguma atividade de forma resumida (apenas o nome da abordagem). 0 – Não é apresentada nenhuma abordagem.
5a.	2 – O método de experimento é definido e referenciado claramente. 1 – O método de experimento é citado. 0 – Não é apresentada nenhuma abordagem.
6a.	2 – O texto descreve critérios para a escolha dos estudos. 1 – O texto descreve apenas estudos aderentes ao estudo apresentado. 0 – O texto não descreve estudos base.
7a.	2 – A área de atuação, tamanho e origem da organização são informados. 1 – Apenas uma das características do item 4 é informada. 0 – O estudo não foi conduzido em uma ou mais organizações.

Fonte: Adaptada de (Costa, 2010).

Para cada artigo avaliado foi atribuída uma pontuação, assim enquadrando-o em um dos cinco níveis de qualidade definidos por (Beecham, 2007), como mostra o Quadro 10.

Quadro 10 – Níveis de Qualidade

Faixa de Notas	Avaliação
Excelente	> 86%
Muito Boa	66% a 86%
Boa	46% a 65%
Média	26% a 45%
Baixa	< 26%

Fonte: (Beecham, 2007).

3.9. Procedimentos da RSL e da Seleção dos estudos Primários

Para a condução desta RSL foram alocados dois pesquisadores (um mestrando e um graduando), que realizaram os seguintes passos:

- A verificação e validação das *strings* de busca com o intuito de averiguar sua acurácia no retorno dos artigos primários e também, assim, poder criar múltiplas instâncias destas *strings* adaptadas para cada base de dados. No início da RSL foram selecionados 5 artigos base. Estes artigos foram selecionados pois estavam precisamente alinhados com o objetivo desta revisão. Tais artigos tinham seus retornos verificados após a aplicação da *string* de busca. Para assim, verificar a acurácia da *string* utilizada;
- Após os testes das *strings* de busca, os dois pesquisadores aplicaram a mesma nos indexadores de conteúdo científico por meio do domínio da Universidade para encontrar os possíveis artigos primários;
- No momento seguinte, os pesquisadores leram os títulos e os resumos dos artigos retornados pela *string* de busca. Para assim, criar uma lista com os possíveis artigos primários;
- Os artigos presentes na lista de possíveis artigos primários tiveram seus títulos, resumos, introduções e conclusões lidos. Neste momento, foram aplicados os critérios de inclusão e exclusão para descartar os falsos positivos, e assim criar uma lista dos artigos primários e uma lista dos artigos excluídos;
- As listas contendo os artigos primários foram comparadas e unificadas. Um artigo era incluído se ao menos um pesquisador o tivesse inserido em sua lista de possíveis artigos primários;
- Os estudos presentes na lista foram lidos em sua totalidade e os critérios de qualidade foram aplicados nestes. Assim, os artigos foram categorizados segundo os níveis de (Beecham, 2007);
- Após, os artigos presentes na lista gerada anteriormente, passaram pela etapa de extração de dados;
- Ademais, todos os documentos e procedimentos foram validados a partir de reuniões com o orientador desta pesquisa.

4. ANÁLISE DOS RESULTADOS

Nesta seção será apresentada a condução das análises dos resultados dos estudos primários, visando responder as questões de pesquisa e as questões bibliométricas.

4.1.Busca Manual e Automática

Visando garantir uma maior confiabilidade para a RSL proposta, seguiu-se o procedimento de realizar um estudo piloto ao início desta RSL (Kitchenham, 2007), onde foi realizada uma busca manual em todas as bases utilizadas e em mais algumas outras bases que não compuseram a revisão sistemática devido terem alguma restrição de uso (vide Quadros 3). A busca manual visa analisar a qualidade dos trabalhos disponíveis sobre os temas referidos a cada uma das questões de pesquisa. Assim, foi identificada uma quantidade mínima de trabalhos relacionados com cada um dos temas. Em conseguinte, foram selecionados alguns dos trabalhos da busca manual como estudos chave, que, posteriormente, foram identificados nas bases digitais durante a busca automática.

Após isso, foi definido o protocolo da revisão sistemática, seus atributos, como questões bibliométricas, *strings* de busca, fontes de pesquisa, entre outras informações. Com as *strings* de busca definidas para cada base digital, cada pesquisador iniciou a busca automática em sua respectiva base. Os dados retornados pelas bases foram armazenados em um software especializado que fornece um banco de dados para o auxílio desta tarefa. Para esta fase, foram coletadas as seguintes informações de cada artigo: a base de onde foi extraído o artigo; o título; os autores; o ano de publicação; o nome do veículo no qual o artigo foi publicado; o resumo; e o endereço do artigo na sua base.

4.2.Seleção dos Estudos pelo Título e Resumo

Para realizar a seleção dos estudos primários foram utilizados: dois pesquisadores (um aluno de Mestrado e um aluno de Graduação); e o acesso às fontes de pesquisa por meio do domínio da Universidade.

Neste momento, cada pesquisador leu apenas o título e o resumo para a primeira etapa de seleção dos estudos primários, onde se buscou compreender o possível alinhamento do trabalho lido com as questões de pesquisa desta RSL a partir do entendimento do objetivo geral apresentado no trabalho. Caso o trabalho, em seu título e resumo, abordasse questões relevantes aos temas desta RSL, os mesmos eram considerados aceitos nesta fase.

Neste momento, foram identificados e excluídos os estudos duplicados entre as diferentes bases, e também, foram identificados e excluídos os trabalhos que estavam desalinhados aos temas desta RSL. Caso houvesse dúvida sobre o alinhamento do trabalho com as questões de pesquisa, os pesquisadores liam a introdução e a conclusão do artigo para decidir sobre a inclusão do artigo.

4.3. Seleção dos Estudos pela Introdução e Conclusão ou Leitura Completa

Os estudos primários foram coletados durante o período de 6 meses (de dezembro de 2020 a maio de 2021), utilizando as *strings* de busca nas fontes de pesquisa definidas no protocolo. Inicialmente, foram retornados 445 artigos e após uma filtragem inicial para a remoção dos artigos repetidos restaram 404 estudos. Tais artigos tiveram seus títulos, resumos e palavras-chave lidos, e após esta etapa, restaram um total de 400 artigos. Tais estudos foram declarados potencialmente relevantes, pois passariam pelos critérios de inclusão e exclusão. Enquanto que na terceira etapa, artigos não relevantes, duplicados, inacessíveis ou em outro idioma que não fossem em português ou inglês foram excluídos.

Posteriormente, restaram apenas 2 artigos incluídos, sendo que estes passaram pelos critérios de qualidade e pela etapa de extração de dados e consequentemente tiveram sua leitura feita em sua totalidade.

Na Tabela 1 é apresentado um resumo das etapas da RSL e o número de artigos retornados em cada etapa.

Tabela 1 – Seleção dos Estudos Primários

Fontes	**E.R.**	**F.R.**	**1° Sel.**	**2° seleção**									
				Excluídos							**Incluídos**		
				CE.1	CE.2	CE.3	CE.4	CE.5	CE.6	CE.7	CI.1	CI.2	CI.3
Ei Compendex	1.579	122	1.232	103	90	0	3	2	22	0	3	1	1
IEEE	1.043	117	390	320	151	0	1	0	58	0	4	0	2
Science Direct	602	18	502	13	62	0	0	0	5	0	1	0	1
ISI Web	137	1	128	2	5	0	1	0	0	0	0	0	0
Scopus	13	2	9	0	2	0	0	0	0	0	0	0	0
TOTAL	**3.374**	**260**	**2.261**	**438**	**310**	**0**	**5**	**2**	**85**	**0**	**8**	**1**	**4**
				840							**13**		

Fonte: Elaborado pelo autor (2021).

Legenda: E.R.: Estudos Retornados.
F.R.: Filtragem dos Repetidos.
1ª Sel.: 1ª Seleção (Títulos e Palavra-chave).

4.4. Avaliação da Qualidade

Foi utilizada uma planilha eletrônica para armazenar os dados dos artigos para responderem as questões bibliométricas e também calcular a nota (Excelente, Muito Boa, Boa, Média, Baixa) para artigo avaliado. A nota era calculada baseada nos atributos avaliados nos critérios de qualidade e na escala *Likert-3*, que representava a adesão destes atributos aos critérios de qualidade (Santos, 2010). A Tabela 2 apresenta os resultados da avaliação da qualidade.

Tabela 2 – Qualidade dos estudos primários

	Baixa	Média	Boa	Muito Boa	Excelente	TOTAL
Número de Estudos Primários	0	1	3	6	3	13
%	0	7,69	23,07	46,15	23,07	100

Fonte: Elaborado pelo autor (2021).

Como pode ser observado, nenhum dos trabalhos ficou na faixa Baixa, 1 estudo (7,69%) ficou na faixa Média, enquanto 3 estudos, (23,07%) ficaram na faixa Boa, 6 estudos (46,15%) estão na faixa Muito Boa e 3 estudos (23,07%) na faixa Excelente. A lista dos trabalhos incluídos e selecionados. A lista dos estudos primários (EP) incluídos e selecionados pode ser visualizado no Quadro 11.

Quadro 11 – Lista de Trabalhos Incluídos.

Trabalhos Incluídos	Fonte
EP.1: *Design and implementation of a cloud based computer forensic tool.*	*Ei Compedex.*
EP.2: *Deleted Data Attribution in Cloud Computing Platforms.*	*Ei Compedex*
EP.3: *Developing a cloud computing based approach for forensic analysis using OCR.*	*Ei Compedex*
EP.4: *Cloud forensics: Technical challenges, solutions and comparative analysis.*	*Ei Compedex*
EP.5: *PyLocky Ransomware Source Code Analysis.*	*Ei Compedex*
EP.6: *Augmenting Performance For Distributed Cloud Storage*	IEEE
EP.7: *Dynamic allocation and efficient distribution of data among multiple clouds using network coding.*	IEEE
EP.8: *A Data Distribution Service for Cloud and Containerized Storage Based on Information Dispersal*	IEEE
EP.9: *Disaster recovery in single-cloud and multi-cloud environments: Issues and Challenges.*	IEEE
EP.10: *An architecture for secure searchable cloud storage.*	IEEE
EP.11: *Data Integrity and Recovery Management in Cloud Systems.*	IEEE
EP.12: *Google Drive: Forensic analysis of data remnants.*	*Science Direct*
EP.13: *Forensic investigation of P2P cloud storage services and backbone for IoT networks: BitTorrent Sync as a case study.*	*Science Direct*

Fonte: Elaborado pelo autor (2021).

4.5. Extração dos Dados

Esta etapa consiste em organizar os dados extraídos para apresentação dos gráficos que serviram como panorama geral e base para futuras análises. A base para as respostas às questões de pesquisa faz parte da lista dos trabalhos incluídos e selecionados na RSL, listados no Quadro 11.

4.6. Respostas às Questões Bibliométricas

Esta seção descreve a extração dos dados referentes às questões bibliométricas da RSL conduzida.

QB.1: Qual é a quantidade de trabalhos incluídos por indexador de estudos?

Após a aplicação do protocolo, descrito na seção 3, foram incluídos: 6 estudos da base *IEEEXplore Digital Library*, 5 estudos da base *Ei Compendex* e 2 estudos da base *Science Direct*. Os estudos estão listados no Quadro 11.

BQ.2: Qual é a quantidade de estudos retornados por ano?

A quantidade de estudos pode ser observada na Figura 1, onde observa-se que no ano de 2015, retornou um número um pouco maior de artigos.

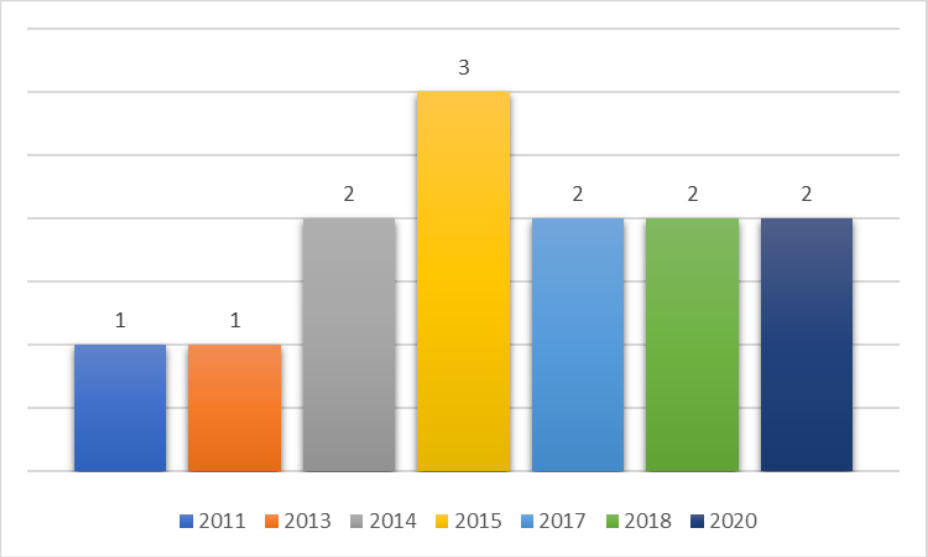

Figura 1 – Quantidade de trabalhos retornados por ano.
Fonte: Elaborado pelo autor (2021).

Para responder as questões (principal e secundárias) levantadas neste trabalho, foram utilizados trabalhos publicados entre 01 de janeiro de 2006 e 31 de dezembro de 2020, não contemplando, dessa forma, os trabalhos publicados em 2021. Este período foi estabelecido a partir de um acontecimento relevante, uma vez que em 2006, a computação em nuvem começa, de fato, a ser oferecida comercialmente, e quando empresas de diversos portes (pequenas, médias e grandes) passam a adotar essa tecnologia como parte do universo corporativo. Pode-se dizer que a computação em nuvem se popularizou, no mercado corporativo, com a *Amazon*, quando esta lançou seu produto "*Amazon Web Services*" (AWS) em 2006.

BQ.3: Qual é a quantidade de estudos retornados por país?

Dos estudos retornados, Alemanha, Austrália e Índia, foram os que mais retornaram, 2 estudos cada país. Outros países foram citados, porém, com apenas 1 estudo, cada. O resultado pode ser observado na Figura 2.

BQ.4: Quais foram os autores com maior quantidade de publicações?

O autor que obteve o maior número de estudos retornados, foi o *Mohammad Matar Al-Shammari*, com duas publicações. Todos os outros autores citados, obtiveram apenas uma publicação cada.

BQ.5: Quantidade de estudos por tipo de estudo? (Teórico, ou Relato de experiências).

Dos estudos retornados, 5 foram estudos teóricos: [EP.2], [EP.3], [EP.4], [EP.9] e [EP.11], representando 38% dos estudos incluídos. Outros 8 estudos, foram experimentais: [EP.1], [EP.5], [EP.6], [EP.7], [EP.8], [EP.10], [EP12] e [EP.13], o que equivale a 62% dos estudos retornados.

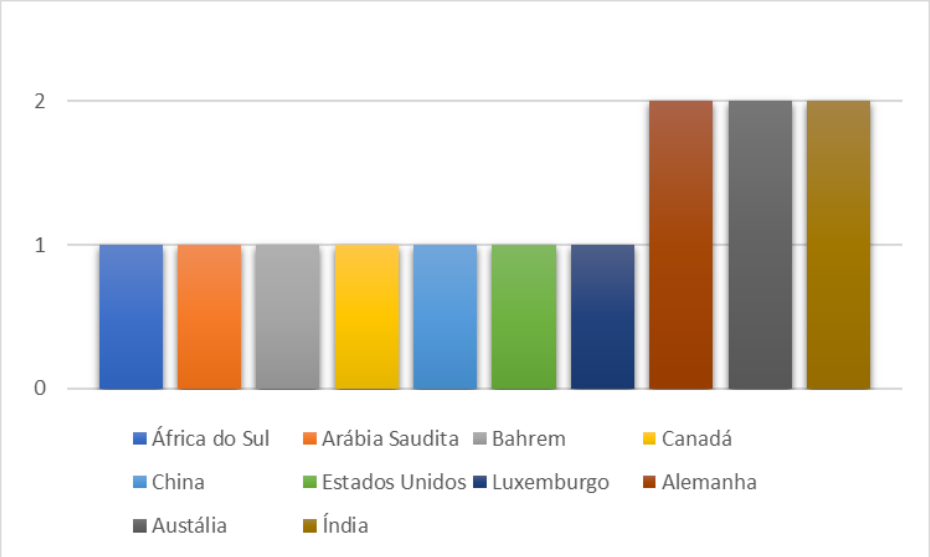

Figura 2 – Quantidade de trabalhos retornados por país.
Fonte: Elaborado pelo autor (2021).

BQ.6: Qual a quantidade de estudos experimentais, por tipo de experimentação, estudo de caso, pesquisa ação, pesquisa de campo e etnográfico?

Dos estudos retornados, 5 foram estudos de caso: [EP.2], [EP.4], [EP.9], [EP.11] e [EP.12], representando 38% dos estudos incluídos. Outros 8 estudos, foram pesquisa ação: [EP.1], [EP.3], [EP.5], [EP.6], [EP.7], [EP.8], [EP10] e [EP.13], o que equivale a 62% dos estudos retornados.

BQ.7: Qual é a quantidade de estudos retornados por tipo de publicação? (Conferência, periódico ou workshop).

A maior parte dos estudos retornados foram publicados em conferências: [EP.1], [EP.2], [EP.3], [EP.5], [EP.6], [EP.7], [EP8], [EP.9], [EP.10] e [EP.11], e 3 em Journal: [EP.4], [EP.12] e [EP.13]. Os percentuais podem ser visualizados na Figura 3.

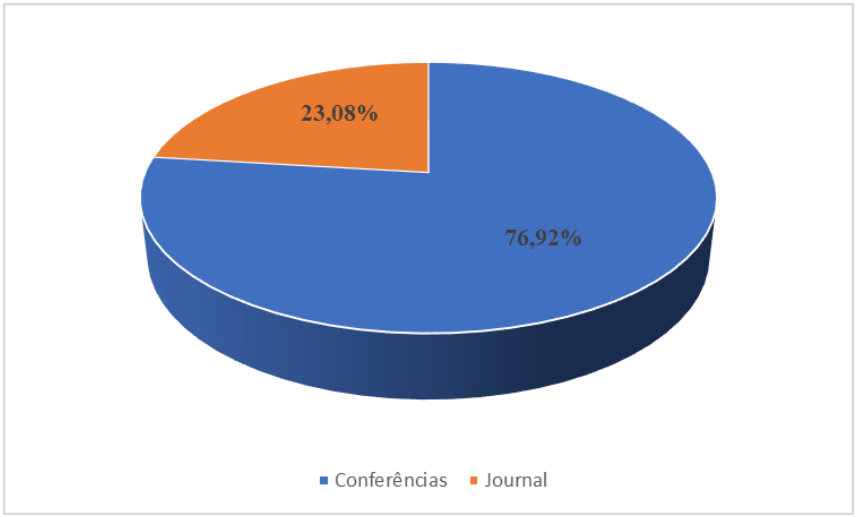

Figura 3 – Porcentual de estudos retornados por tipo de publicação.
Fonte: Elaborado pelo autor (2021).

BQ.8: Qual é a quantidade de estudos retornados por veículo de publicação?
Dos veículos utilizados para a publicação, todos tiveram apenas um único estudo publicado. As conferências, bem como os *Journals* utilizados para a publicação, podem ser visualizados no Quadro 12.

Quadro 12 – Lista de Trabalhos por veículo de publicação

Veículo de Publicação	Tipo
2011 Information Security for South Africa	Conferência
2013 Seventh International Conference on IT Security Incident Management and IT Forensics	Conferência
2014 IEEE 3rd International Conference on Cloud Networking (CloudNet)	Conferência
Journal of Network and Computer Applications	*Journal*
2015 Fifth International Conference on Communication Systems and Network Technologies	Conferência
Digital Investigation	*Journal*
2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing	Conferência
2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)	Conferência
Computers & Electrical Engineering	*Journal*
2018 1st International Conference on Computer Applications & Information Security (ICCAIS)	Conferência
2018 IEEE Symposium on Service-Oriented System Engineering (SOSE)	Conferência
2020 IEEE Symposium on Product Compliance Engineering - (SPCE Portland)	Conferência
2020 Fourth International Conference on Inventive Systems and Control	Conferência

4.7. Respostas às Questões de Investigação ou de Pesquisa.
Este trabalho responde as questões principais levantadas no protocolo desta Revisão quasi Sistemática da Literatura como meio de investigar o estado da arte no âmbito de analisar as

técnicas ou métodos que tratam da recuperação de dados em nuvem. A seguir, são apresentadas as respostas das questões de pesquisa, e as considerações dos autores a respeito delas.

QP.1: Quais são as abordagens existentes para a recuperação de dados, que apoiam as atividades da Computação Forense, no contexto do armazenamento em nuvem?

A recuperação de dados em nuvem possui diversas e complexas variáveis inerentes as suas características, que se tornam grandes desafios aos profissionais desta área, pois não se trata de recuperar arquivos deletados de um determinado dispositivo de armazenamento que esteja ao alcance, mas sim, de servidores de arquivos que podem estar localizados em qualquer país do mundo. Nos estudos avaliados, entre as principais abordagens encontradas que necessitam recuperar dados na nuvem, estão: os desastres naturais [EP.9], *ransomware* [EP.5], utilização de kits de ferramentas computacionais para investigação em ambiente de nuvem [EP.1], metodologia de gerenciamento de dados [EP.11] e na estrutura de um *middleware* [EP.6].

QS.1. Quais as atividades são apoiadas pelos sistemas de armazenamento de dados na nuvem?

O armazenamento em nuvem já está bastante inserido no cotidiano de atividades de ensino, comerciais e de pesquisa. O estudo [EP.6] desenvolveu, visando melhorar todos os aspectos que um usuário espera de um provedor de nuvem, um armazenamento indexado distribuído na nuvem para interface com os CSPs (*Cloud Services Providers*), que são empresas especializadas que oferecem serviços remotos de processamento, rede, armazenamento, infraestrutura completa e/ou de softwares, e que são utilizados por meio de conexão dedicada ou internet.

QS.2. Quais técnicas ou práticas são utilizadas para dar segurança aos dados em um ambiente de armazenamento na nuvem?

Entre os estudos analisados, não foram identificados relatos de técnicas ou práticas relacionadas especificamente à segurança dos dados no armazenamento em nuvem. No entanto, no estudo [EP.11], é apresentada uma metodologia para recuperação e gerenciamento de dados que utiliza um método de embaralhamento dos dados armazenados, com o objetivo de proporcionar garantia dos backups dos dados dos usuários, minimizando o risco de perda.

QS.3. Quais técnicas são utilizadas pelos sistemas de armazenamento na nuvem para criptografar os dados?

Nos estudos avaliados, foram identificados o preenchimento OAEP com criptografia RSA, em [EP.5]; além dos algoritmos de função *Hash* (MD5, SHA1 e SHA265) em [EP.4].

QS.4. Quais as técnicas e ferramentas utilizadas para a recuperação de dados em nuvem no contexto da Computação Forense?

Entre as técnicas e ferramentas utilizadas para recuperar os dados na nuvem, foram identificadas no [EP.1]: o AIR (*Automated Image Restore*), o TSK (*The Sleuth Kit*), o AFB (*Autopsy Forensic Browser*), o CBCFT (*Cloud Based Computer Forensic Tool*), e o PYLOCKY, no estudo [EP.5].

QS.5. Qual o contexto de aplicação da abordagem encontrada?

Nos estudos avaliados, foram identificadas, em sua grande maioria, aplicações comerciais, como nos estudos: [EP.3, EP.4, EP.5, EP.6, EP.7, EP.8, EP.9, EP.10, EP.11] com soluções voltadas para empresas que atuam na área de computação forense. Como no [EP.5] que trata de uma análise de um *ransomware* desenvolvido, chamado PYLOCKY, fornecendo uma visão geral das ferramentas existentes que podem ajudar empresas a se recuperarem de um ataque *ransomware*, bem como, da análise do código fonte. No entanto, foram identificados, também, estudos na área acadêmica [EP.1], [EP.2], [EP.12] e [EP.13], no qual são avaliados trabalhos em universidades com o objetivo de aprimorar as pesquisas já desenvolvidas. No [EP.12] é realizado um estudo de caso que utilizou o Google Drive para um experimento, identificando artefatos em um disco rígido de computador e em um iphone.

QS.6. Quais os profissionais que utilizam ferramentas para recuperar dados na nuvem?

Entre os estudos avaliados, os profissionais mais citados, são os peritos forenses [EP.13], [EP.11], [EP.10], [EP.9], [EP.8], [EP.7], [EP.6], [EP.5] e o [EP.4], que utilizam ferramentas para analisar e recuperar dados, principalmente em atividades voltadas para o mercado e para perícias criminais. Em seguida, estão os pesquisadores acadêmicos e praticantes [EP.1], [EP.2], [EP.3] e [EP.12], que testam e avaliam técnicas de recuperação com a finalidade de aprimorar os resultados das ferramentas que estão sendo desenvolvidas.

QS.7. Quais são as leis que tratam sobre a recuperação de dados em provedores de nuvem?

Não foram identificados estudos que tratassem de leis sobre recuperação de dados ligados à provedores em nuvem.

QS.8. Quais acordos e cooperações internacionais tratam da recuperação de dados?

Foi identificado no estudo [EP.4] o acordo de cooperação internacional MLAT (*Mutual Legal Assistence Treaty*) que é um tratado de assistência jurídica mútua assinado entre dois ou mais países para a execução das tarefas de investigação, ação penal e prevenção do crime. Um exemplo desse tipo de tratado foi o celebrado entre o Brasil e os Estados Unidos em 1997 e promulgado pelo Decreto 3.810, de 02 de maio de 2001 [Direito&TI 2015]. O MLAT é o meio bilateral mais usado por autoridades brasileiras para solicitar cooperação jurídica internacional e solicitar diligências ao governo dos Estados Unidos.

QS.9. Quais os sistemas de arquivos usados na arquitetura de armazenamento de dados na nuvem?

Entre os estudos selecionados, grande parte não aborda qual, ou quais sistemas de arquivos são utilizados no processo de armazenamento dos dados na nuvem, porém o [EP.4] e o [EP.13] citam que utilizam o XtremeFS, e o [EP.5] e o [EP.12], mencionam o NTFS.

QS.10. Quais os sistemas operacionais utilizados nos provedores de armazenamento de dados na nuvem?

Nenhum dos estudos selecionados abordaram a questão dos tipos de sistemas operacionais utilizados nos provedores de armazenamento de dados na nuvem.

QS.11. Quais as principais empresas que oferecem serviços de armazenamento de dados na nuvem?

As principais empresas citadas nos estudos são: Microsoft (*One Drive*) [EP.6], Google (*Google Drive*) [EP.12], Amazon (*Elastic Compute Cloud* - EC2) [EP.10], Dropbox Inc (*Dropbox*) [EP.6], BitTorrent (*BitTorrent Sync*) [EP.13].

5. Descrição e Discussão das Ferramentas e Técnicas Encontradas

Durante as pesquisas realizadas para a elaboração deste estudo, as principais ferramentas e técnicas encontradas para recuperação de dados em ambiente de nuvem, foram: *Autopsy Forensic Browser* (AFB), no [EP.1], que é uma ferramenta de código aberto e gratuita, que reuniu as ferramentas do TSK (*The Sleuth Kit*) em uma interface gráfica, permitindo uma análise dos arquivos, diretórios, blocos de dados e *inodes* (alocados ou apagados) presentes na imagem de um sistema de arquivos. Através da interface gráfica do AFB, as imagens dos sistemas de arquivos da máquina invadida podem ser examinadas no nível de abstração de arquivos, *inodes* ou blocos de dados. Também, é possível buscar por palavras-chave e expressões regulares nas imagens, bem como, criar uma linha de tempo contendo os *MAC times* dos arquivos e diretórios. A interface utilizada pelo AFB é baseada em HTML, segundo um modelo cliente-servidor. O AFB corresponde ao servidor, e qualquer navegador HTML pode ser usado como cliente. Com isto, é permitido executá-lo diretamente no sistema comprometido, por meio de uma mídia removível, fornecendo acesso remoto e protegido contra escritas ao investigador, que utiliza um navegador HTML no sistema de análise.

O *Automated Image and Restore* (AIR), também citado no [EP.1], é um *front-end* de interface gráfica do usuário para os utilitários dd/dc3dd (comandos do sistema Linux, que permitem converter, copiar e formatar arquivos), usados para limpar definitivamente dados de um disco. Essa ferramenta AIR foi projetada para tornar a tarefa de criação de imagens, em mídia digital, mais fácil para investigadores e equipes de resposta a incidentes.

O CBCFT (*Cloud Based Computer Forensic Tool*), propõe uma ferramenta forense computacional para investigação, em ambiente de nuvem, incluindo recuperação de dados, que permite que um profissional da área acesse a ferramenta a partir do portal da nuvem, em sua máquina ou em qualquer outra. A ferramenta ainda oferece vantagens, como: análise de uma grande quantidade de informações de forma rápida e eficiente; uso otimizado de mídia de armazenamento; mínimo de interrupção, oferecendo serviços de computação em nuvem ininterruptos para o usuário e a manutenção da infraestrutura é simplificada. Essa ferramenta está descrita em [EP1].

The *Sleuth Kit* (TSK) é uma coleção de ferramentas de linha de comando para Unix que permite investigar um computador, isto é, trata-se de uma coleção de ferramentas forenses. O objetivo destas ferramentas é o acesso aos arquivos e aos sistemas de arquivos. O TSK suporta os sistemas de arquivos FAT, Ext2/3, NTFS, UFS, e ISO 9660. Ao todo são 27 ferramentas que fornecem informações ou conteúdo de arquivos, unidades de dados,

volumes, partições, sistemas de arquivos e metadados como *Mac times* e *inode*. Mais detalhes desta ferramenta podem ser encontrados no [EP1].

PYLOCKY, é o nome de um *ransomware*, e foi analisado, para fornecer uma visão geral das ferramentas existentes que podem ajudar empresas a se recuperarem de um ataque desse tipo, além de explicar as limitações e princípios de funcionamento das ferramentas de recuperação. [EP5].

6. CONCLUSÃO E TRABALHOS FUTUROS

Com o objetivo de fornecer uma análise dos principais métodos existentes de recuperação de dados em ambiente de nuvem, foi desenvolvida esta Revisão Sistemática da Literatura para contribuir com os profissionais e pesquisadores. Porém, no desenvolvimento do estudo, percebeu-se que a quantidade de trabalhos que especificamente abordam esse tema, ainda é bem limitada. Isso, em razão das dificuldades encontradas em recuperar dados nesse tipo de ambiente, que possui particularidades que dificultam imensamente o trabalho dos investigadores, seja no âmbito computacional ou jurídico. Infere-se, que o processo de recuperação de dados em nuvem, ainda tem muito a evoluir com a criação de novas ferramentas, técnicas e na criação de uma legislação que facilite o trabalho de investigação. O investimento em pesquisas para desenvolver propostas que tenham como meta soluções viáveis para esse tipo de problema, é um caminho a ser considerado para que acelerem de forma eficiente a recuperação de dados em nuvem.

6.1. Trabalhos Futuros

Pretende-se pesquisar mais as soluções disponíveis para a recuperação de dados em ambiente de nuvem, focando na compreensão do funcionamento de softwares específicos, que possam auxiliar e acelerar, de forma eficiente, as investigações, facilitando o trabalho dos profissionais da área.

REFERÊNCIAS BIBLIOGRÁFICAS

ASHRAF, I. (2014) *An Overview of Service Models of Cloud Computing*. International Journal of Multidisciplinary and Current Research, vol. 2, pp. 779-783.

ALSADHAN, A., ALHUSSEIN, M. (2018), *Deleted Data Attibution in Cloud Computing Platforms*. International Conference on Computer Applications & Information Security (ICCAIS). 4-6 April de 2018, Riyadh – Saudi Arabia. DOI: 10.1109/CAIS.2018.8441961.

BEECHAM, S.; BADDOO, N.; HALL, T.; ROBINSON, H.; SHARP, H. (2007). *Motivation in Software Engineering: A systematic literature review. Information and Software Technology*: Elsevier, v. 50, n. 860 -878.

COSTA, C. S. (2010). Uma abordagem baseada em evidências para o gerenciamento de projetos no desenvolvimento distribuído de software. Dissertação de Mestrado – Programa de Pós-Graduação em Ciência da Computação - Universidade Federal de Pernambuco, Recife.

EASTERBROOK, S., SINGER, J., STOREY, M. A., & DAMIAN, D. (2008). *Selecting empirical methods for software engineering research. In Guide to advanced empirical software engineering* (pp. 285-311). Springer, London.

GOKULAKRISHNAN, S.; GNANASEKAR, J.M. (2020) "*Data Integrity and Recovery Management in Cloud Systems*". Conference: 2020 Fourth International Conference on Inventive Systems and Control (ICISC). DOI:10.1109/ICISC47916.2020.9171066.

KITCHENHAM, B. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Technical Report EBSE-2007-01, Department of Computer Science Keele University, Keele.

MAFRA, S.; TRAVASSOS, G. (2006). Estudos Primários e Secundários apoiando a busca por Evidencia em Engenharia de Software - Relatório Técnico: RT-ES-687/06 – Programa de Engenharia de Sistemas e Computação - COPPE/UFRJ – Rio de Janeiro.

MISHRA, A. K.; MATTA, P.; PILLI, E. S.; JOSHI, R. C. (2012). "*Cloud Forensics: State-of-the-Art and Reasearch Challenges". International Symposium on Cloud and Services Computing*. International Symposium on Cloud and Services Computing (ISCOS) – Índia. DOI: 10.1109/ISCOS.2012.32.

MOHITE, M. P.; ARDHAPURKAR, S. B. (2015). "*Design and Implementation of a Cloud Based Computer Forensic Tool*". Department of Computer Technology Y.C.C.E, Índia. DOI: 10.1109/CSNT.2015.180.

PICHAN, A.; LAZARESCU, M.; SOH, S. T. (2015) "*Cloud forensics: Technical Challenges, Solutions and Comparative Analysis*". Department of Computing, Curtin University, Kent Street, Bentley, Perth, WA 6102, Australia. DOI: 10.1016/j.diin.2015.03.002.

PEREIRA, E. F. DE O.; ARAÚJO, J. S.; SILVA, SAWAKI, W. DE M.; OLIVEIRA, S. R. B. (2019). Revisão Sistemática da Literatura na Computação Forense: Um Estudo de Caso Aplicado na Recuperação de Dados em Mídias Digitais. 16th *International Conference on Information Systems & Technology Management* (CONTECSI) 2019. DOI: 10.5748/16 Contecsi/sec - 6138.

SANTOS, G. (2010). Revisão Sistemática, Minicurso. Simpósio Brasileiro de Qualidade de Software – SBQS 2010, Belém – PA.

SORINI, A.; SCOTT, G. D. (2020). "*PyLocky Ransomware Source Code Analysis*". IEEE Symposium on Product Compliance Engineering - (SPCE Portland). DOI: 10.1109/SPCE50045.2020.9296183.