Cloud Broker: A Systematic Mapping Study Appendix A

Already, distinct works have been done for conducting SMS and designing a unique research methodology [1], [2]. One of the most prominent of these research methodologies is related to Peterson et al. [2]. The research methodology used in this paper is adapted from three SMS that were done by Peterson et al. [2], Ramaki et al. [3], and Javan et al. [4] and some updates and improvements in some phases have been done. Figure 1 illustrates the two primary phases of our review process: the planning phase and the conducting phase. In the following sections, we will delve into the specifics of each step in our methodology, providing a comprehensive understanding of the approach we followed.

I. Planning the Mapping Study

Figure 1 presents the breakdown of the planning phase in our methodology, which comprises several distinct stages. These stages encompass defining the scope and research questions, planning the search process, planning the study selection process, determining the evaluation strategy for the search and study selection, and planning the data extraction and classification process. In the subsequent sections, we will examine each stage individually, providing detailed explanations and insights.

A. Specifying the Scope and Research Questions (RQs)

Defining research questions is a crucial aspect of conducting a review research work. These research questions serve as a guide for researchers to address and answer throughout the review process. The primary objective of formulating research questions is to establish clear research goals. As a result, one of the fundamental steps in a SMS is to design pertinent and meaningful questions that, once answered, can shed light on the challenges, issues, and significant topics within the desired field of study. In this particular SMS, eight comprehensive and distinct research questions have been formulated. These research questions encompass a wide range of aspects, ensuring that they cover all the objectives of the study. Table I provides a detailed description of these research questions and explains the rationale behind each one, outlining their significance and relevance to the study at hand.

To comprehensively address the research questions mentioned earlier, it is crucial for the scope of this paper to encompass all published works in the field. Our review extends over a span of 23 years, starting from the emergence of cloud brokers in 2009 and continuing until the end of 2022.

B. Specifying the Search Strategy

To conduct the search process, specifying the exact search strategy is vital which should be carried out during the process of review. In this paper, as previously stated, two distinct search strategies (i.e. manual search and backward snowballing search) have been designed to find all related studies. In the manual search method, each venue within the search space (which refers to the current list of venues to be searched) is manually examined using the predefined keywords. This involves conducting individual searches in each venue to identify relevant papers. In backward snowballing, the objective is to identify additional related papers by exploring the reference lists of the papers that have already been added to the review.

TABLE I
Defined Research Questions

No.	Research Questions	Rationale
1	How active is the field of brokering and how is the distribution of selected studies by type over publication year (journal, conference, and workshop)?	To detect the current volume of researches and primary trends in order to better discernment the attraction of the filed. Investigating the publication volume of research studies in the field per year.
2	Which researchers and research venues are more active in this field and how are the active researchers distributed geographically?	The demographics of brokering techniques research provide a useful starting point for interested researchers by identifying active scholars, venues, and countries.
3	What are the core research topics in the field of brokering?	This will help the researchers to determine the significance of different areas within the domain of cloud brokering and unravel the research tree within the brokering field.
4	Which broker topics have the least/most corresponding attention and what is the publication trend and distribution for each topic?	Some objectives might be more prominent than others, but broker developers should take care to cover a varied spectrum of topics.
5	Which forms of empirical evaluation have been used? What are the tools available to support field approaches? Which techniques are more used in the field?	The empirical evaluation means whether the environment is real or simulation and supporting tools can describe frameworks, platforms, or simulation. Techniques can be game theory, optimization, and heuristic.
6	What is the relation between topics and broker roles in NIST category? Which NIST roles have the least/most corresponding attention?	General classification schemes might work to an extent, but a precise and comprehensive classification of broker roles should address broker-specific criteria.
7	In which environment and service layer is the broker mostly considered?	Environments are Multi-cloud, federated, etc. and the service layer is IaaS, SaaS, PaaS, and XaaS.
8	What is the broker Control orientation?	Generally, types of control orientation are centralized and distributed.

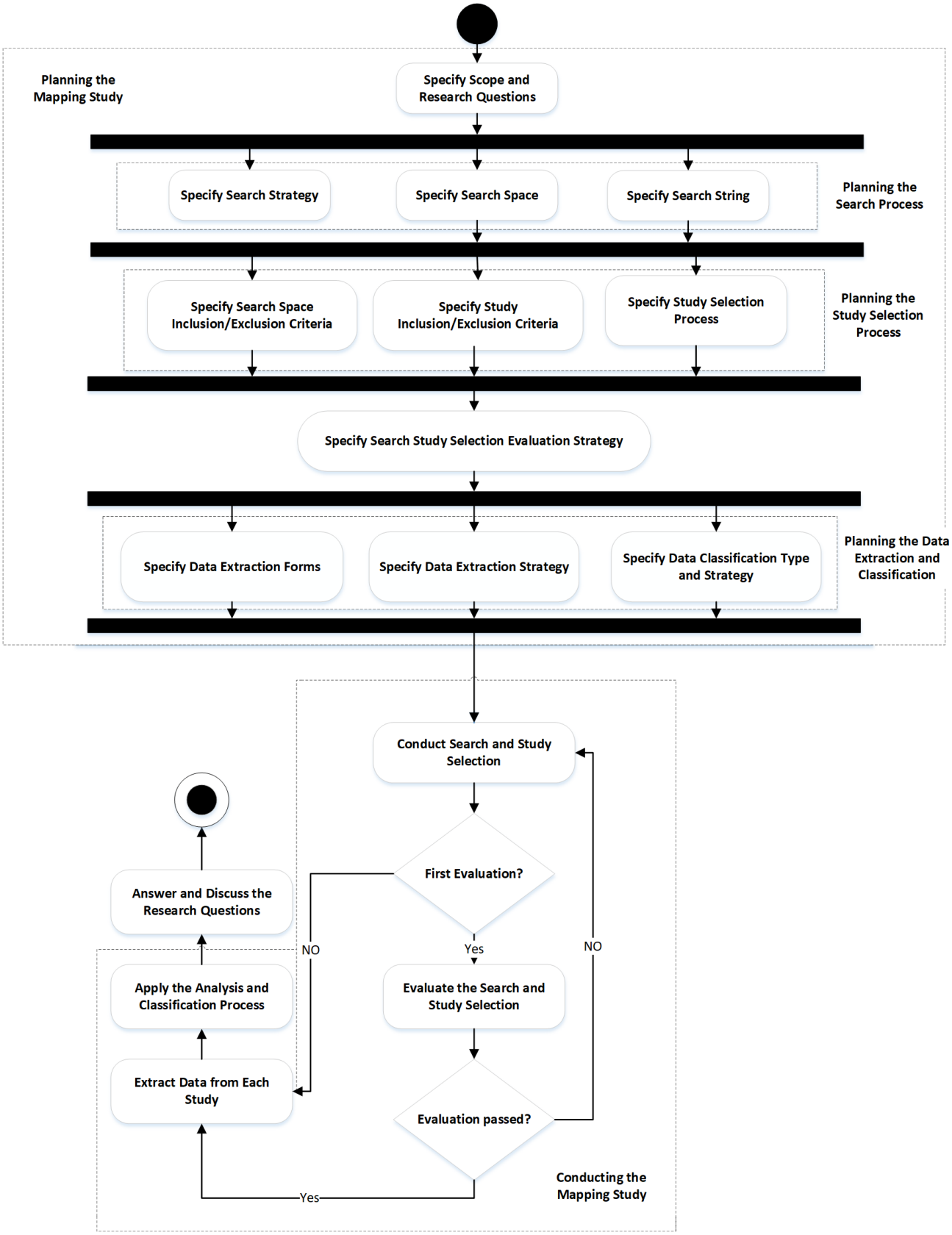

Fig. 1. Research methodology [4]

This process involves examining the reference section of each paper to identify any relevant sources that may have been cited. By doing so, we aim to expand the scope of the review and ensure a comprehensive coverage of the relevant literature.

Figure 2 illustrates the general strategy which comprises seven steps. It should be noted that the determining related search spaces (journals, conferences, and workshops) are not necessarily carried out after the manual search step. Hence, an S-Flag is defined and set with zero in step 2 and one in step 5. Current search strategy comprises six primary steps that are acquired by investigating the base reviews studies [2], [3], [4] and applying some improvements to them.

1) In this step, a set of initial secondary studies (i.e., surveys and reviews) shown in Table II are selected and investigated to determine a set of initial set of papers and keywords.
2) The search space of each included paper, which refers to the journal, conference, or workshop where the paper was published, is extracted and recorded.
3) The extracted set of search spaces is evaluated against a predefined set of quality criteria. Only those search spaces that meet these criteria are considered and included in the review. For these selected search spaces, manual searches are conducted using the specified keywords.
4) The study selection process is applied, wherein each paper is evaluated for inclusion in the review. If a paper meets the inclusion criteria, the process proceeds to step 5 for further analysis. However, if the paper is identified through backward snowballing, the process jumps directly to step 2 to explore additional related papers.
5) Backward snowballing is performed on papers that have not been previously snowballed. This involves examining the reference lists of the included papers to identify additional papers that may be relevant to the review. If a new paper is discovered through this process, the search process proceeds to step 4 for further investigation. However, if no new papers are found, the process continues to step 6.
6) In this step, backward snowballing is done on the all new found included papers and S-Flag is set to one. If the backward snowballing results new papers therefore, the execution jumps to step 4 to conduct the search space process, otherwise step 6 is done.
7) The study selection process is then carried out on the newly identified papers and the search process comes to an end.

A complete list of extracted studies can be seen in $SuppFile_{E9,T1}$ to $SuppFile_{E9,T5}$. Furthermore, $SuppFile_{W3,T4}$ to $SuppFile_{W3,T6}$ represents a complete list of all extracted search spaces.

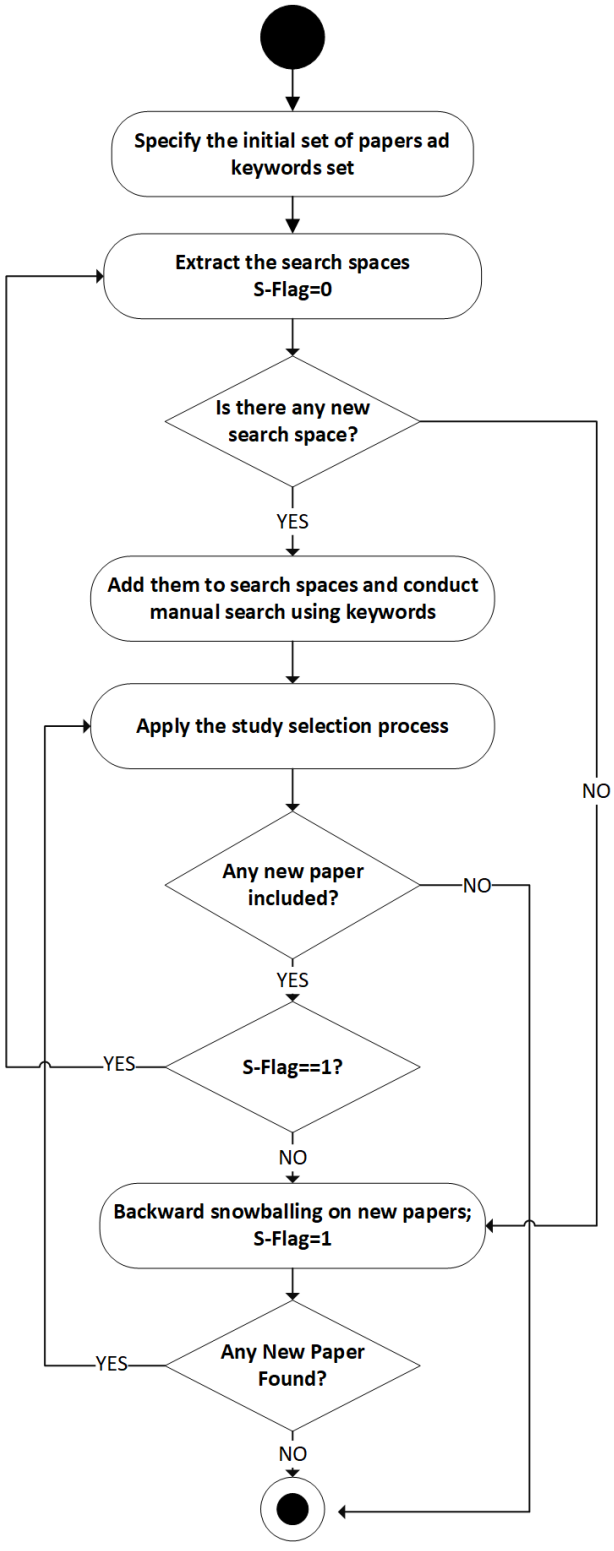

Fig. 2. The Search Strategy [4]

C. Specifying the Search Space

At the beginning of our process, the search space is empty. However, as we extract the initial set of papers from secondary

TABLE II
THE SECONDARY STUDIES USED FOR GENERATING INITIAL SET OF OUR SMS

No.	Secondary Study Title	Year	Ref.
1	Brokering in Interconnected Cloud Computing Environments A Survey	2018	[5]
2	A Review on Service Broker Algorithm in Cloud Computing	2017	[6]
3	A Comprehensive Study on Cloud Service Brokering Architecture	2017	[7]
4	Cloud Services Recommendation Reviewing the Recent Advances and Suggesting the Future Research Direction	2017	[8]
5	Service Provisioning in Cloud: A Systematic Survey	2017	[9]
6	A Survey on Various Cloud Aspects	2016	[10]
7	A Classification and Comparison Framework for Cloud Service Brokerage Architectures	2016	[2]
8	A Review on Broker Based Cloud Service Model	2016	[1]
9	Cloud Service Brokerage: A Systematic Literature Review using a Software Development Lifecycle	2016	[11]
10	Resource Provision Algorithms in Cloud Computing: A Survey	2016	[12]
11	Towards a Holistic Multi-Cloud Brokerage System: Taxonomy, Survey and Future Directions	2015	[13]
12	A survey on SLA-based Brokering for Inter-Cloud Computing	2015	[14]
13	Cloud Services Brokerage: A Survey and Research Roadmap	2015	[15]
14	Cloud Service Selection: State-of-the-art on Future Research Directions	2014	[10]
15	Cloud Computing Service Composition: A Systematic Literature Review	2014	[16]
16	A Review of Literature on Cloud Brokerage Services	2014	[17]
17	A Literature Review on Cloud Computing Adoption Issues in Enterprises	2014	[18]
18	A Survey on Needs and Issues of Cloud Broker for Cloud Environment	2014	[19]
19	Survey on important Cloud Service Provider attributes using SMI Framework	2013	[20]
20	A Comparison Framework and Review of Service Brokerage Solutions for Cloud Architectures	2013	[21]
21	A Survey on Interoperability in the Cloud Computing Environments	2013	[22]
22	Inter-Cloud Architectures and Application Brokering: Taxonomy and Survey	2012	[23]

studies, we identify specific journals, conferences, and workshops that are relevant to our research. These identified venues are then added to the search space. As our search progresses, the search space expands to include a comprehensive list of relevant journals, conferences, and workshops. For a detailed compilation of these venues, please refer to the Supplementary Material, specifically $SuppFile_{W3,T4}$ to $SuppFile_{W3,T7}$.

D. Specifying the Search String

The objective of this SMS is to review all papers in the field of cloud broker. To search related papers in the manual search phases, a set of keywords have been defined that can be seen in $SuppFile_{W3,T3}$. These keywords are gradually completed during the progress of the review process. Then, these keywords are merged and queries are created. These queries, which are the same as search strings, are used to search for related studies in the search spaces. Table III presents the queries employed in the proposed SMS.

TABLE III
CONSTRUCTED QUERIES FOR FINDING THE RELATED STUDIES

No.	Query
1	Cloud broker
2	Cloud AND service (arbitration OR intermediation OR aggregation OR integration OR customization OR Orchestrat)
3	(Multi cloud OR Federat cloud OR Cross cloud OR Inter cloud OR (third party AND Cloud)) AND (auction OR negotiation OR pricing OR interoperability OR management)
4	Cloud AND (”service composition” OR ”service selection”)

TABLE IV
JOURNAL EXCLUSION CRITERIA (JEC)

No.	Description
1	The journal is not indexed in the JCR
2	The scope of the journal is not related to our desired field

E. Specifying the search space inclusion/exclusion criteria

In order to identify search spaces that meet high quality standards and have relevant scope, specific criteria have been established to determine whether a search space should be included or excluded. The exclusion criteria for journals can be found in Table IV, while the exclusion criteria for conferences and workshops are presented in Table V. These criteria serve as guidelines to evaluate the suitability of each search space. By applying these exclusion criteria, any search spaces that are deemed irrelevant or of low quality are eliminated from consideration.

Several important points should be noted regarding the evaluation of journal and conference/workshop quality in this study:

- The field of cloud broker encompasses a large volume of research studies, and therefore, empirical thresholds for the exclusion criteria have been carefully selected. The aim is to ensure that small changes in the exclusion thresholds do not have a significant impact on the number of included or excluded papers. Additionally, the thresholds have been set to avoid excluding highly cited papers.
- While the focus of this systematic mapping study (SMS) is on the field of cloud broker, search spaces related to web services, distributed computing, and parallel computing have also been considered. This is because these areas often publish papers in the field of cloud computing. For example, the IEEE Transactions on Smart Grid is included because it has published relevant papers in the field of cloud computing.

TABLE V
OTHER EXCLUSION CRITERIA (OEC)

No.	Description
1	((Qualis<A5) OR (ERA<A) OR [(H5_Index<15) AND ((Qualis<A5) OR (ERA<A))] OR (Metrics Not Available))
2	Aims and scopes are not related

TABLE VI
EXCLUSION CRITERIA FOR EXTRACTED STUDIES

No.	Description
1	The study is not a primary study (e.g. survey)
2	Study cannot be accessed (e.g. not indexed)
3	The study is out of our primary scope (e.g. security)
4	The study belongs to an excluded search space
5	The contribution of the study is not related to the cloud broker

- Certain search spaces, such as the International Conference on Software Engineering (ICSE), have been excluded because they have limited publications in the field of cloud computing. ICSE primarily focuses on software engineering, which is not directly aligned with the scope of this SMS.
- In some cases, certain search spaces may have multiple names or undergo name changes over time. For example, the ACM International Symposium on High-Performance Distributed Computing and the International Symposium on High-Performance Parallel and Distributed Computing refer to the same search space, which is HPDC. Similarly, the IEEE Global Telecommunications Conference changed its name to the Global Communications Conference after 1972. Another example is the merger of two search spaces: the European Conference on Machine Learning (ECML) and the European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) merged to form the European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) after 2007.

F. Specifying the study inclusion/exclusion criteria

After finding the related and included search spaces, they are searched using constructed queries in Table III to find the related papers. The inclusion criterion for an extracted study is its relevance to our designated scope. Table VI illustrates the exclusion criteria for extracted studies. To determine the inclusion/exclusion state of each extracted paper, title, abstract, and keywords have been generally investigated. In some cases, the entire paper has been reviewed to ensure its inclusion or exclusion status. A complete list of all included papers can be found $SuppFile_{E9,T1}$. It is worth mentioning that all extracted review studies during conducting the search process in this SMS have been excluded.

G. Specifying the Study Selection Strategy

The study selection strategy comprises two main parts. In the first part, the relevance of a paper is assessed by reviewing its title, abstract, and keywords. If these elements do not provide sufficient information to make a conclusive decision about its relevance, the full text of the paper is examined. If the paper is determined to be irrelevant, it is discarded. In cases where the relevance is uncertain, a third-party is involved to assist in judging its relevance. In the second part of the process, for the papers deemed relevant, they undergo evaluation based on the criteria specified in Table VI.

H. Specifying the Search and Study Selection Evaluation Strategy

In this phase, the goal is to examine the completeness of the search strategies utilized for finding the related studies. To achieve a more objective evaluation, the sensitivity metric is selected to evaluate the applied search and study selection, which is calculated as follows:

$$Sensitivity = \frac{\text{Number of studies in the SMS}}{\text{Number of studies overall}} \times 100 \quad (1)$$

Since the exact number of the denominator is not available, we have employed a methodology known as the Quasi-Gold Standard (QGS). This involves selecting a set of studies from reputable sources within the research community. To accomplish this, a separate team, referred to as the *shadow team*, visited the home pages of active researchers in the field of cloud computing. They extracted any papers relevant to the cloud brokering field. Subsequently, by applying inclusion and exclusion criteria, we obtained our Quasi-Gold Standard (QGS) dataset.

We utilized the QGS to calculate the quasi-sensitivity, which involved dividing the number of included papers in our SMS that were also present in the QGS by the total number of papers in the QGS. The purpose of employing the QGS was to determine the quasi-sensitivity and compare the obtained result with a predefined threshold. If the result falls below the threshold, it indicates that the search and study selection process should be repeated using the QGS. According to [8], an acceptable threshold range lies between 70% and 80%. During the evaluation phase, a total of 32 articles were discovered, out of which 25 articles had already been identified in the main phases of the systematic review process. Thus, a sensitivity around 80% was achieved, which is above the current research's predefined threshold. In other words, the probability of not finding a paper related to the cloud broker was less than 20%. It can then be concluded that the results acquired from the proposed review have satisfactory accuracy and validity. Section 1.6 of Appendix A presents the details of the steps taken to evaluate the search strategies and the metrics for calculating the accuracy and completeness of these strategies so as to extract the related studies.

I. Planning the Data Extraction and Classification Process

After conducting the study search process and determining the included papers, to answer the RQs, some information is needed that should be extracted from these papers. Here, the data extraction forms are specified and also the data extraction strategy is determined. To prepare useful information to answer the RQs, the type of data extraction and strategy should be specified. After the determination of which studies are included, based on the defined quality criteria, different parts of the articles (i.e., the title, abstract, keywords, and body) are examined to extract the information needed for later analysis. The extracted data are organized into tables and utilized as the information for responding to the RQs of the SMS ($SuppFile_{E1}$ to $SuppFile_{E9}$ includes all the information needed to answer the RQs).

II. Conducting the Mapping Study

Once the search and study selection process has been defined during the planning phase of the mapping study, we can proceed to conduct the mapping study. This phase consists of several stages, each involving specific activities. These stages include conducting the search and study selection, evaluating the search and study selection process, data extraction, and analysis and classification. In the following section, we will provide a detailed explanation of the activities carried out in each stage.

A. Conducting the process of Search space and study Selection

The search process begins by providing an initial set of papers obtained from secondary studies, which were identified through an informal search process using keywords such as *cloud broker*, *survey*, and *review*. The search methodology employs a variable called *depth* to describe the type of search conducted. Once a set of secondary studies is obtained (as shown in Table II), the initial value of the depth variable is set to zero, and relevant cited papers from those studies are extracted, forming an initial set of included studies (primary studies). Subsequently, manual search and backward snowballing are performed on the papers in the initial set. Sample records of journals, conferences, and workshops are presented in Table VII and Table VIII, which indicate their inclusion or exclusion status based on the criteria outlined in Table IV and Table V. The *Reason for Exclusion* column in these tables specifies the rationale for excluding each search space.

For instance, referring to Table VII, the journal with JID=11 is excluded due to its scope being unrelated to the defined scope of the cloud broker. Similarly, the journal with JID=12 is excluded because it has not been indexed in JCR. Table VIII demonstrates that a conference with OID=52 is excluded as its metrics, such as Qualis and ERA, fall below the defined threshold in Table V. Additionally, OID=62 is excluded because its scope does not align with the field of cloud brokers (refer to $SuppFile_{W3,T1}$). Detailed information regarding all extracted search spaces, including the reasons for their inclusion or exclusion in journals and other categories (conferences and workshops), can be found in $SuppFile_{W3,T4}$ and $SuppFile_{W3,T6}$ respectively.

Table IX and Table X provide samples of extracted papers along with their exclusion and inclusion states, determined based on the defined exclusion criteria. In Table IX, for instance, the first study's presented contribution is not aligned with the aim and scope of a cloud broker, leading to its exclusion from the analysis.

$SuppFile_{W3,T4}$ and $SuppFile_{W3,T6}$ provide a list of extracted search spaces along with detailed information about the reason for the exclusion of each venue. Additionally, the aims and scope of each search space are introduced in $SuppFile_{W3,T5}$ and $SuppFile_{W3,T7}$, which also provide the reasons for the exclusion of each search space. For comprehensive information about all conferences and workshops, you can refer to $SuppFile_{E9,T3}$, while $SuppFile_{E9,T2}$ contains detailed information about all journal papers. Furthermore, $SuppFile_{E9,T1}$ presents a complete list of all included papers, including journals, conferences, and workshops.

B. Evaluating the Search and Study Selection Process

The obtained results from the previous phase are utilized to extract a list of active researchers, which can be found in $SuppFile_{E1,T3}$. Additionally, information about the geographical distribution publications can be found in $SuppFile_{E1,T1}$. To validate this SMS, an evaluation phase was conducted according to the planned strategy in Section I-H. In this phase, the homepage of authors was manually examined by the *shadow team* to identify any unseen papers that were not discovered during the search phases. A list of these unseen papers can be found in $SuppFile_{E9,T4}$. Based on the provided information, there are 6 included studies that were not initially found during the search process. Considering the conducted processes to evaluate the completeness of the search strategy, it can be concluded that the set of studies satisfactorily covers the field of cloud brokers.

C. Data Extraction

After the completion of the set of studies, the next step is the data extraction and classification process. During this process, the relevant information from each study is extracted and categorized according to predefined criteria. It is important to note that disagreements may arise during the data extraction process. To address these disagreements, discussions were held among the team members involved in the study. Through these discussions, a consensus was reached on the appropriate extraction and classification of the data.

D. Analysis and Classification of data

Once the data has been organized into tables, the next step is to investigate and analyze the data in order to address the research questions of the SMS. One of the main objective of this SMS is to determine the primary topics and sub-topics in the field of cloud broker. In the following section, the process of determining the topics and sub-topics is explained in detail.

In order to determine the topics and sub-topics, a clustering technique proposed by researchers [24], [25], [26] was employed. This technique utilizes a co-occurrence matrix of keywords to group them into clusters based on their similarities. By analyzing the co-occurrence patterns, related topics and sub-topics can be identified, taking into consideration their similar cognitive orientation [26]. The extraction process of level-one topics in the research tree is depicted in Figure 3. A research tree is a multilevel tree structure that represents the main topics and sub-topics in the field of cloud broker.

In Figure 3, the process of constructing the research tree begins with the extraction of keywords from the included papers. These keywords can be obtained directly from the papers or, if unavailable, they are manually extracted by an expert. Subsequently, the most frequent keywords, surpassing a specified threshold value, are selected. Based on these selected keywords, a co-occurrence matrix is constructed. An example of the co-occurrence matrix can be seen in Table XI. Each

TABLE VII
SAMPLE EXTRACTION TABLE FOR JOURNAL SEARCH SPACE

#JID	Journal Name	ISSN	$IF_{JCR-2017}$	$Q_{JCR-2017}$	Topic	Publisher	Depth	Reaseon of Exclusion
6	Future Generation Computer Systems	1872-7115	4.639	Q1	Computer science a Theory and methods	Elsevier	0	-
7	IEEE Internet Computing	1941-0131	1.929	Q1	Computer science a Software engineering	IEEE	0	-
11	Annals of Telecommunications	1958-9395	1.168	Q4	Telecommunication	Springer	0	JEC2
12	Journal of Cloud Computing: Advances, Systems and Applications	N/A	-	-	N/A	Springer	0	JEC1

TABLE VIII
SAMPLE EXTRACTION TABLE FOR CONFERENCE AND WORKSHOP SEARCH SPACE

#OID	Name	H5_Index	Qualis	ERA	Indexing	Depth	Reaseon of Exclusion
46	ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)	77	A1	A	ACM	0	-
52	International Conference on High Performance Computing and Communications (HPCC)	20	-	B	IEEE	0	OEC1
62	International Conference on Information and Knowledge Management (CIKM)	49	-	A	ACM	0	OEC2
70	International Conference on Data Engineering (ICDE)	53	A1	A	IEEE	0	-

TABLE IX
SAMPLE EXTRACTION TABLE FOR JOURNAL PAPERS

Paper Title	Journal Name	Year	Exclusion Criteria
Privacy-preserving and sparsity-aware location-based prediction method for collaborative recommender systems	Future Generation Computer Systems	2019	PEC1
A trust centric optimal service ranking approach for cloud service selection	Future Generation Computer Systems	2018	-
A privacy-preserving cryptosystem for IoT E-healthcare	Information Sciences	2019	PEC1
Implementation of a real-time network traffic monitoring service with network functions virtualization	Future Generation Computer Systems	2019	PEC1
A hybrid multi criteria decision method for cloud service selection from Smart data	Future Generation Computer Systems	2019	-

cell x_{ij} in the matrix represents the number of times the keyword from the ith row appears together with the keyword from the jth column in the same paper. For instance, the keyword *service composition* appears in 88 studies, and it co-occurs with the keyword *QoS* in 31 studies. The complete co-occurrence matrix, which consists of 40 keywords, can be found in $SuppFile_{E3,T5}$. This matrix provides comprehensive information about the relationships and associations between different keywords, forming the basis for further analysis and clustering.

After constructing the co-occurrence matrix, the next step is to normalize it and compute the similarity between each pair of keywords based on their co-occurrence patterns. This is achieved by creating a vector for each keyword, where each element of the vector represents the co-occurrence count with another keyword. The similarity between keyword pairs is computed using the cosine similarity index, as shown in Equation 2 [26].

$$Sim(i,j) = \frac{\sum_{k=1}^{n}(x_{ik}y_{kj})}{\sqrt{((\sum_{k=1}^{n}x_{ik}^2)(\sum_{k=1}^{n}y_{kj}^2))}} \quad (2)$$

where x_{ij} is the number of co-occurances of keyword i with keyword j. Once the similarity matrix is obtained, the keywords need to be classified into clusters. In this study, the K-means algorithm [27] is employed as a common clustering method. The parameter K, representing the number of clusters to be formed, is initialized by an expert.

The process of constructing the research tree and detecting top-level topics based on keyword frequency is described as follows:

- Collection of Keywords: The keywords from all included papers (581 papers) are gathered. In cases where a paper does not have author-provided keywords, an expert generates keywords. In the $SuppFile_{E1,T5}$, keywords manually extracted by an expert are marked with yellow cell color. The extraction process involves analyzing the paper's title, abstract, context, and the defined queries in Table III. The collected keywords are stored in a keyword pool, and the frequency of each keyword is computed. For example, the keyword *broker* appears in 100 included papers. After the first phase, the keyword pool size is reduced to 1713.

TABLE X
SAMPLE EXTRACTION TABLE FOR CONFERENCE PAPERS

Paper Title	Conference Name	Year	Exclusion Criteria
A model for evaluating the economics of cloud federation	International Conference on Cloud Networking (CloudNet)	2015	-
Cloudlet Scheduling with Particle Swarm Optimization	International Conference on Communication Systems and Network Technologies (CSNT)	2015	PEC1
Incentivizing Microservices for Online Resource Sharing in Edge Clouds	International Conference on Distributed Computing Systems (ICDCS)	2019	-
The Elasticity and Plasticity in Semi-Containerized Co-locating Cloud Workload: A View from Alibaba Trace	ACM Symposium on Cloud Computing	2019	PEC1

TABLE XI
A SAMPLE OF THE CO-OCCURRENCE MATRIX

Keywords	Service Composition	QoS	Service Selection	Resource Management	Scheduling
Service Composition	88	31	12	3	0
QoS	31	68	11	7	4
Service Selection	12	11	56	1	1
Resource Management	3	7	1	51	10
Scheduling	0	4	1	10	32

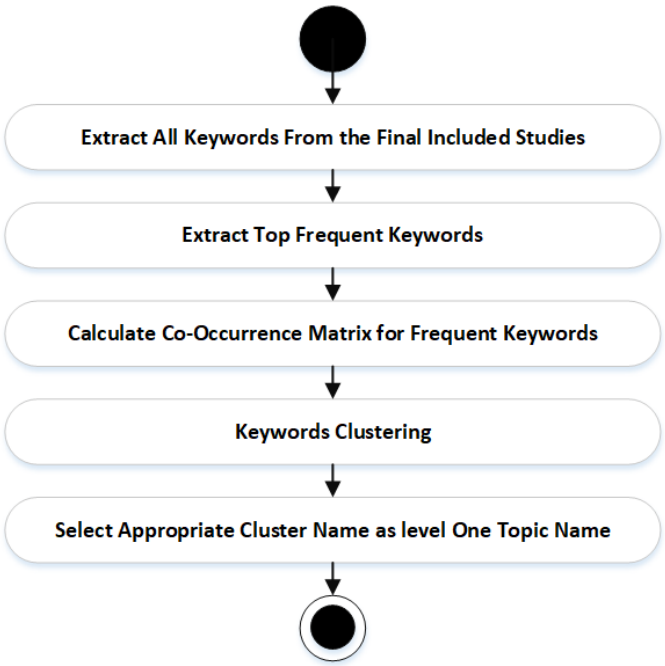

Fig. 3. The extraction process of level-one topics in research tree [3]

- Categorization of Lexically Similar Keywords: Keywords that are lexically similar are grouped together. For example, keywords like *SaaS*,*SaaS applications,* and *software as a service* are categorized under the *SaaS* group. This step reduces the number of keywords to 395 items.
- Removal of Repetitive Keywords: Certain highly repetitive keywords that do not contribute to new concepts in the research tree are removed. Examples of such keywords include *broker, cloud computing, edge computing, distributed system, IoT, mobile cloud computing,* and *autonomic computing.*
- Construction of Co-occurrence Matrix: A co-occurrence matrix is generated for words with a repetition threshold equal to or greater than 9. The threshold value is determined through trial and error, and it coincidentally yields the best semantic output for the matrix. A total of 70 keywords with a repetition count of 9 or more form the primary matrix. In each iteration, the two words with the highest co-occurrence count compared to other word pairs are selected, forming a new subject. This process continues until no new co-occurrences are found among the matrix members. After 27 iterations, a final set of 18 topics is obtained. Table XIII displays the resulting compounds from the co-occurrence matrix, along with the corresponding level-one topics of the research tree. Through expert examination and aggregation of related topics, a total of 11 main topics are identified as the level-one topics of the research tree (shown in the second column of Table XIII).
- Division of Topics: The topics are further divided into two categories, namely client-centric and provider-centric, based on the evaluation of a team of experts.

The client-centric and provider-centric topics in the cloud broker field are distinguished based on the activities performed by the broker in response to user and provider requests, respectively. Client-centric topics encompass activities such as service discovery, resource management (client-side), and service composition, which are initiated as a result of user requests. Provider-centric topics, on the other hand, include activities like pricing, resource management (provider-side), and energy management, which are triggered by provider requests.

Figure 4 presents the research tree obtained in this SMS. The percentage of included papers in each topic is indicated below the respective topic. It's important to note that the determination of paper subjects considers thematic similarities, as outlined in Table XII. Each row of Table XII associates the topics listed in the *Topic* column with the words specified in the *Similar Concept* column, denoting their equivalence.

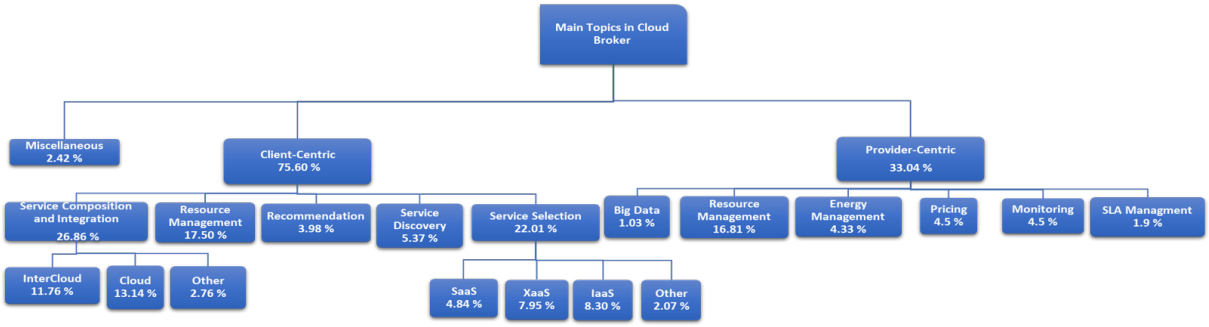

Fig. 4. Multilevel research tree containing the main topics and sub-topics in the field

TABLE XII
THEMATIC SIMILARITIES BETWEEN THE KEYWORDS OF INCLUDED PAPERS

Topic	**Similar Concept**
Service composition and integration	Composition + Orchestration + Integration
Resource Management (client)	Scheduling + Provisioning
Energy management	Energy management + Green computing
SLA management	SLA violation + SLA management and etc.
Resource Management (provider)	Resource allocation + Resource management + VM scheduling
Pricing	Pricing + Cloud market
Recommendation	Recommendation + Data management

TABLE XIII
THE TOPICS OBTAINED AS THE LEVEL-ONE TOPICS OF THE RESEARCH TREE

No.	**Topic**	**Integrated Keywords**
1	Monitoring	Performance/Monitoring
2	Resource Management	Resource Management/Scheduling IaaS / Virtualization / Provisioning
3	Pricing	Pricing/Cloud market Game/Auction Data management/Cost
4	Energy Management	/Green Computing
5	Service Discovery	XaaS/Discovery Semantic
6	Recommendation	Recommendation
7	Service Composition and Integration	Web service/Service Composition/QoS Aggregation Integration/SOA
8	Service Selection	Service Selection/Fuzzy/AHP/Decision making
9	Big Data	Big data/Healthcare Data-intensive/Evolutionary algorithm/optimization
10	SLA management	SLA/Negotiation/Dependability Cloud provider/Cloud user
11	Other	Agent/Multi-agent systems Graph-based methods/Datacenter Linear programming/Cloud service/Learning Failure/Workflow

REFERENCES

[1] N. Rajganesh and T. Ramkumar, "A review on broker based cloud service model," *Journal of computing and information technology*, vol. 24, no. 3, pp. 283–292, 2016.

[2] F. Fowley, C. Pahl, P. Jamshidi, D. Fang, and X. Liu, "A classification and comparison framework for cloud service brokerage architectures," *IEEE Transactions on Cloud Computing*, vol. 6, no. 2, pp. 358–371, 2016.

[3] A. A. Ramaki, A. Rasoolzadegan, and A. G. Bafghi, "A systematic mapping study on intrusion alert analysis in intrusion detection systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–41, 2018.

[4] A. J. Jafari and A. Rasoolzadegan, "Security patterns: A systematic mapping study," *Journal of Computer Languages*, vol. 56, p. 100938, 2020.

[5] S. S. Chauhan, E. S. Pilli, R. C. Joshi, G. Singh, and M. C. Govil, "Brokering in interconnected cloud computing environments: A survey," *Journal of Parallel and Distributed Computing*, vol. 133, pp. 193–209, 2019.

[6] R. Jain, T. Sharma, and N. Sharma, "A review on service broker algorithm in cloud computing," *International Journal of Computer Applications*, vol. 975, p. 8887, 2017.

[7] A. Koneru and M. Sreelatha, "A comprehensive study on cloud service brokering architecture," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2017, pp. 47–53.

[8] F. Aznoli and N. J. Navimipour, "Cloud services recommendation:

Reviewing the recent advances and suggesting the future research directions," *Journal of Network and Computer Applications*, vol. 77, pp. 73–86, 2017.

[9] A. Bhattacharya and S. Choudhury, "Service provisioning in cloud: A systematic survey," in *Advanced Computing and Systems for Security*. Springer, 2017, pp. 37–63.

[10] R. Rakhi, L. M. Biju, S. S. K. Rinta, M. Biju, and M. L. Madhavu, "A survey on various cloud aspects," *International Journal of Advanced Research in Computer Science*, vol. 7, no. 5, 2016.

[11] V. Paulsson, V. C. Emeakaroha, J. P. Morrison, and T. Lynn, "Cloud service brokerage: A systematic literature review using a software development lifecycle," in *Americas Conference on Information Systems*, 2016.

[12] J. Zhang, H. Huang, and X. Wang, "Resource provision algorithms in cloud computing: A survey," *Journal of Network and Computer Applications*, vol. 64, pp. 23–42, 2016.

[13] B. Aldawsari, T. Baker, and D. England, "Towards a holistic multi-cloud brokerage system: Taxonomy, survey, and future directions," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. IEEE, 2015, pp. 1467–1472.

[14] E. Mostajeran, B. I. Ismail, M. F. Khalid, and H. Ong, "A survey on sla-based brokering for inter-cloud computing," in *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*. IEEE, 2015, pp. 25–31.

[15] A. Barker, B. Varghese, and L. Thai, "Cloud services brokerage: A survey and research roadmap," in *2015 IEEE 8th International Conference on Cloud Computing*. IEEE, 2015, pp. 1029–1032.

[16] A. Jula, E. Sundararajan, and Z. Othman, "Cloud computing service composition: A systematic literature review," *Expert systems with applications*, vol. 41, no. 8, pp. 3809–3824, 2014.

[17] J. Akilandeswari and C. Sushanth, "A review of literature on cloud brokerage services." *International Journal of Computer Science and Business Informatics*, vol. 10, no. 1, pp. 25–40, 2014.

[18] R. F. El-Gazzar, "A literature review on cloud computing adoption issues in enterprises," in *International Working Conference on Transfer and Diffusion of IT*. Springer, 2014, pp. 214–242.

[19] D. Geetha, R. Hayat, and M. Thamizharasan, "A survey on needs and issues of cloud broker for cloud environment," *Int. J. Dev. Res*, vol. 4, no. 5, pp. 1035–1040, 2014.

[20] L. Monteiro and A. Vasconcelos, "Survey on important cloud service provider attributes using the smi framework," *Procedia Technology*, vol. 9, pp. 253–259, 2013.

[21] F. Fowley, C. Pahl, and L. Zhang, "A comparison framework and review of service brokerage solutions for cloud architectures," in *International Conference on Service-Oriented Computing*. Springer, 2013, pp. 137–149.

[22] B. Rashidi, M. Sharifi, and T. Jafari, "A survey on interoperability in the cloud computing environments." *International Journal of Modern Education & Computer Science*, vol. 5, no. 6, 2013.

[23] N. Grozev and R. Buyya, "Inter-cloud architectures and application brokering: taxonomy and survey," *Software: Practice and Experience*, vol. 44, no. 3, pp. 369–390, 2014.

[24] N. M. Calcavecchia, A. Celesti, and E. Di Nitto, "Understanding decentralized and dynamic brokerage in federated cloud environments," in *Achieving federated and self-manageable cloud infrastructures: theory and practice*. IGI Global, 2012, pp. 36–56.

[25] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity." *BMJ: British Medical Journal*, vol. 308, no. 6943, p. 1552, 1994.

[26] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Information and Software Technology*, vol. 53, no. 6, pp. 625–637, 2011.

[27] H. Li and K. Yamanishi, "Text classification using esc-based stochastic decision lists," *Information processing & management*, vol. 38, no. 3, pp. 343–361, 2002.