Institut für Softwaretechnologie

Abteilung Zuverlässige Softwaresysteme

Universität Stuttgart
Universitätsstraße 38
70569 Stuttgart

Bachelorarbeit

Observation of current approaches to utilize the elastic cloud for Big Data Stream Processing

Michael Zigldrum

Studiengang: B.Sc. Informatik

Prüfer: Prof. Dr. Ing. Steffen Becker

Betreuer: Floriment Klinaku M.Sc.

begonnen am: 23.05.2018

beendet am: 23.11.2018

Abstract

This paper conducts a systematic literature map to collect information about current approaches to utilize the elastic cloud for data stream processing in the big data context. First is a description and setup of the used scientific methodology which adheres to generally accepted methods for systematic literature maps. After building a reference set and constructing search queries for the data collection came the data set cleaning where the publications were first automatically filtered and consecutively manually reviewed to determine the relevant papers. The collected data was evaluated and visualized to help answer the defined research questions and present information. Finally the results of the thesis are discussed and the limitations and implications addressed.

Kurzfassung

Diese Arbeit befasst sich mit der Durchführung einer Systematic Literature Map um einen Überblick über ein Feld zu gewähren. Das untersuchte Feld dieser Arbeit befasst sich mit der Verwendung der elastischen Eigenschaften der Cloud für Datenstrom Prozessierung im Big Data Umfeld. Bestandteil der Systematic Literature Map ist sowohl das Sammeln aller Publikationen, welche für das untersuchte Feld relevant sind, als auch die Auswertung und Präsentation der gesammelten Daten. Um die Informationen zielgerichtet zu evaluieren, wurden Forschungsfragen definiert, welche als Leitfaden dienen. Zu Beginn wurden die verwendeten wissenschaftlichen Methoden vorgestellt, welche sich an anerkannten Prozeduren orientieren. Nach dem zusammenstellen von einigen relevanten Publikationen, wurden auf deren Basis Suchanfragen für die Datensammlung erstellt. Danach wurden die Daten aus den Online Datenbanken bekannter Verleger exportiert und Duplikate entfernt. Um die endgültigen relevanten Publikationen festzustellen, wurden anhand von Schlagworten irrelevante Publikationen aussortiert und schließlich manuell einzeln bewertet. Die gesammelten Daten wurden teilweise automatisch ausgewertet und manuell klassifiziert um mit den Ergebnissen die vorher definierten Forschungsfragen zu beantworten. Abschließend werden die Ergebnisse diskutiert und die Einschränkungen und Implikationen dieser Arbeit behandelt.

Contents

List of Figures

List of Tables

List of Acronyms

CSV **C**omma **S**eparated **V**alue

DSP Data Stream Processing

DSPS Distributed Stream Processing System

IaaS Infrastructure as a Service

IoT Internet of Things

PaaS Platform as a Service

Listings

Chapter 1

Introduction

The advent of Internet of Things (IoT) saw a large increase in data that needed to be processed. The field of Big Data was created to develop techniques and approaches to handle those amounts of information. Soon Data Stream Processing (DSP) was developed to handle incoming streams of data. Those streams come from sensors that are connected via IoT and are continuously transmitting data which needs to be processed, hence the term DSP. With the help of cloud computing, companies soon started to offer solutions to customers as various services, such as Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). Cloud computing in this context follows the NIST Definition of Cloud Computing written by Mell & Grance [13].

In this definition, they define the essential characteristics of the cloud, which are:

On-demand self-service: The consumers can manage their resources by themselves, without the need of contacting a human from the provider. [13]

Broad network access: The computing resources are accessibly via networks by any client configuration. [13]

Resource pooling: The resources offered by the provider are used by different consumers. The consumers do not know which hardware they are using exactly. [13]

Rapid elasticity: Resources can be scaled up and down by the consumer (even automatically) in a rapid fashion. [13]

Measured service: The cloud system itself can optimize the used resources and observer and measure resource usage transparently. [13]

The “Rapid elasticity”[13] characteristic in combination with the other characteristics allow for more economical computing paradigms. Organizations only have to pay what they are using, and they can automatically and rapidly adapt the amount of resources they are buying to match their processing needs exactly. Most characteristics are being used already, however, utilizing elasticity is a recent idea for DSP and not a lot is known about the approaches to utilize this property of cloud computing for DSP.

This thesis aims to define the current state of utilizing elasticity and identify some areas where more work is required to further advance the field. To answer those question I will conduct a systematic literature map. Systematic mapping studies in software engineering are a new method, which has first been used in 2007 by Bailey et al. [3], according to Petersen et al. [14]. Contrary to a systematic literature review, a mapping study categorizes the publications of a field and presents a structured report, which gives a broad overview of the field and the work within it. During the course of this work, the field of utilizing elasticity for DSP will be referred to as “the field under study”, to simplify the writing.

Chapter 2

Research Design and Method

To ensure reproducible and valid results I adhered to the process described by Kuhrmann et al. [10] to follow when performing a literature map. I also relied on their paper for guidance regarding the search query construction, search result cleaning and voting procedure for the last step in determining whether papers are inside the field under study or not. For the execution of the actual mapping study I adhered to steps described in a paper by Petersen et al. [14].

Following the structure of Kuhrmann et al. [10] I grouped the steps into three major work packages (Figure: 2.1) for me to follow. The first work package contains all the steps necessary to build the search queries for each library used. Work package two handles the automated data collection by the online libraries and also contains the dataset cleaning procedure to exclude any duplicates and irrelevant publications. The last work package holds the rating and review process and finally the evaluation of the collected information.

Instead of doing a preliminary study or using the trial and error message I followed the snowballing process, namely the method of backwards snowballing recommended from the paper of Kuhrmann et al. [10]. This means, that I iteratively select relevant papers from the references of my initial reference papers and repeat the process with the newly added publications [4].

For the voting procedure I decided to follow the two person voting process as outlined by Kuhrmann et al. [10]. However the addition of at least one other person into the voting process ensures increased objectivity in the final result set and also helps to eliminate human errors.

After the final set of papers is determined I will evaluate the data and extract information which helps me answer my research questions as defined in Table 2.1. For this I will follow mapping procedures recommended by Petersen et al. [14]. An important part of the information extraction is the classification according to various facets, such as research and contribution type, which then enables me to draw conclusions about the research field. Finally I will present my findings visually in appropriate graphical representation.

RQ	Definition
1	How mature is the research field of cloud elasticity in the big data stream processing context?
1.1	How many publications are published in total?
1.2	How many publications are published per year?
1.3	How mature is the field judged by the composition of papers based on classification facets?
2	What type of research in the research field require more work?
2.1	How many papers are published per category in the research type classification?
2.2	How many papers are published per category in the contribution type classification?

Table 2.1: Research Questions.

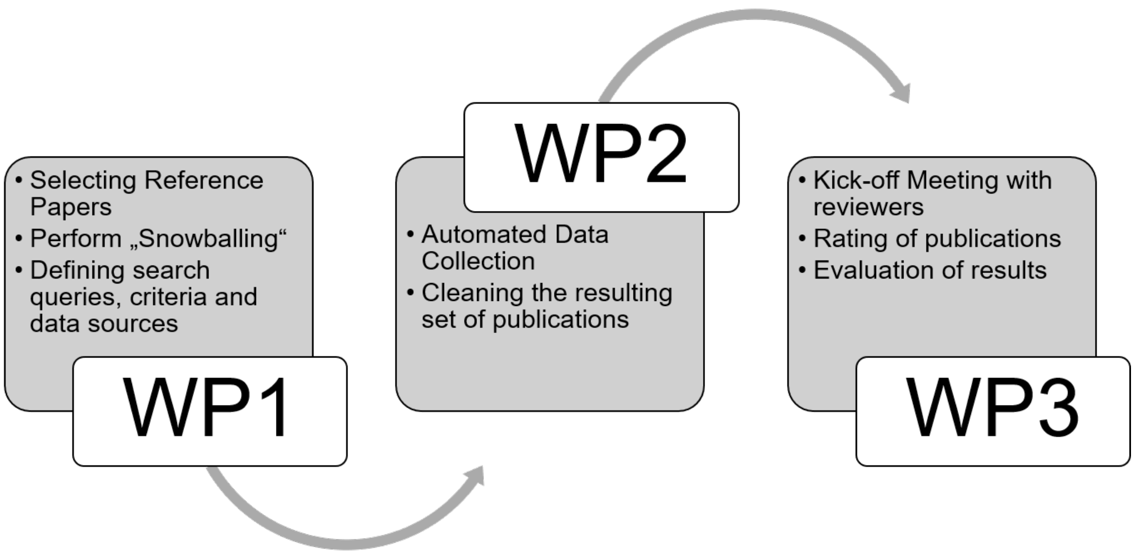

Figure 2.1: Work Packages.

As the first step I defined some inclusion and exclusion criteria to help identify whether a specific paper should be counted as inside my desired publication space, which can be seen in Tables 2.2 and 2.3. Those criteria are used as a rough guideline while doing the search query construction and later the in the reviewing process. I decided to use the ACM Digital Library, IEEE Xplore, Springer Link and ScienceDirect search engines, as those have a focus on computer science and are also used in other mapping studies with a focus on fields related to software engineering (see Thor et al. [15], or Kuhrmann & Konopka et al. [12]).

No.	Criterion
1	Title or Abstract make it clear, that the publication relates to Elastic Data Stream Processing in the cloud.
2	Title or Abstract contain 'elastic' or synonyms/words used in the same context.

Table 2.2: Inclusion Criteria.

No.	Criterion
1	The paper is a PHD Thesis or Workshop paper.
2	Title contains an exclusion keyword. (List of keywords appended to this thesis)
3	Abstract contains an exclusion keyword. (List of keywords appended to this thesis)

Table 2.3: Exclusion Criteria.

Chapter 3

Data Collection for the Literature Map

This chapter describes my work during the work packages as seen in Figure 2.1. First is a description of the search query building procedure including the techniques I used to get around the restrictions of some online libraries. The provided query examples and listings are written down in a SQL-like syntax. Next is the results of the data collection and a summary of the filtering process to narrow down the publications. Finally this chapter details the manual review process used to determine the final set of relevant publications.

3.1 Search Query Construction

To gain an understanding about the field of Elastic Data Stream Processing in the Cloud and use it for the search query construction, I followed the paper by Kuhrmann et al. [10] and started by snowballing a known set of publications. As the starting point for snowballing I chose four papers, which objectively belong into the publications space I will map. Those papers are: Cervino et al. [5], Gedik et al. [8], Hochreiner et al. [9] and Abadi et al. [1]

From the base papers I expanded the set of known publications to 16 entries including the reference papers I started out with. The full list of reference papers has been appended to this thesis as Appendix A.

After collecting the abstracts for every paper in my reference set I started working on the search query definition. It was clear I needed to design two separate queries, one to search the abstracts and another one for the title. The reason being, that not all papers have been entered into the databases with an abstract, or some papers might have a very small to non existant abstract. I counted how many times each word occurred in the title and abstract of every paper respectively, to determine the query buildup and

```
(elastic) and (data) and (stream) and (processing)
```

Listing 3.1: Starting Query.

```
(elastic or scalable or scaling or adaptive or borealis)
and
(stream or streaming)
and
(processing or system)
```

Listing 3.2: General Query for the Title.

keywords to search for. I then removed stopwords, such as “and”, “is” or “a”, from each list and ranked the words from most appearances to least appearances.

The next step was starting to do test searches in the online databases and check how many reference papers can be found in the search results. The goal was to have a 100% hit rate distributed accross the databases, which means that each reference papers has been returned from a query at least once. This method gives me guidance when construction and evaluating the search queries. I started out with a basic query related to the field of elastic data stream processing (Listing: 3.1) and started including synonyms based on the previously mentioned wordlists and the context of the papers. I also decided to include “borealis” as multiple papers did not reference “elastic” or a synonym in their abstract and title but instead referenced the Borealis Streaming Engine [2]. The resulting queries (Listing: 3.2 and Listing: 3.3) achieved the targeted 100% hit rate across the databases.

During the general query building I had to construct a special search query for each search engine used in the query construction process and later in the data collection step. To perform the data collection I had to convert my general queries into the search engine specific syntax, which presented some challenges. Some search engines for example would not allow more than eight binary operators (‘AND’ and ‘OR’) or only a certain number of keywords per search request. To overcome this problem I split up the general query. I ensured that the resulting set of queries were equivalent to the original one by dividing the first “OR” statement into separate sections and executing a search request for every section while appending the rest of the query. An example can be seen in Listing 3.4.

```
(elastic or auto?scale or borealis) and
(data or cloud or event or query) and
(stream) and
(processing or computing or system)
```

Listing 3.3: General Query for the Abstract.

```
Q:1  (elastic or scalable) and stream

 would become:

Q:1.1 elastic and stream
Q:1.2 scalable and stream
```

Listing 3.4: Splitting Queries Example.

This way I was able to merge the different result sets and still get the same results as if I had executed the whole query. Another challenge was the stemming feature of some search engines. For the engines which did not support stemming or placeholders for that matter, I had to include various different spellings and variations of the words in my query such as: ‘streaming’, ‘streams’ or ‘streamed’ , or ‘auto-scale’, ‘auto scale’ or ‘auto-scaling’ to make sure I would not miss any important publications with my search.

Lastly, some search engines such as Springer Link hold a wide range of publications not only related to computer science, so I also had to take care of selecting the relevant categories, since I was not interested in publications about elastic materials from the physics category.

3.2 Executing Data Search and Export

After determining search queries and adapting the queries for every search engine I would use, the next challenge was exporting the data from the results presented to me. The problem here was that most engines have an export limit which only allows downloading of the first 1000 entries. I used the same trick as before when search queries only supported a limited number of boolean operators or keywords, which was dividing the queries, so that there were fewer results per query and thus I was able to export the results. For the Springer Link search engine it was simpler to recursively go into subcategories within the search to stay within this limit, which in turn lead to results being displayed in two or more sub-categories.

I have exported all results as **C**omma **S**eparated **V**alue (CSV) files, since most of the formats are compatible or at least easily adaptable into the same form as other CSV files. After the data export the total number of publications was 7054, as can be seen in Table 3.1. The table also shows the composition of publications based on the online libraries that I used. The Springer Link library returned the most results, however this was expected as this library has the biggest number of publications in its database. As the first step in the dataset cleaning I removed the duplicates within each database, as the queries can have some overlap. Table 3.1 also shows the amounts of duplicates per database, as well as across all databases. Most notably, the results from the Springer Link library contained a lot of duplicates within the queries, which mainly came from the fact that the export size was limited to 1000 entries. The other libraries stayed within the expected range. There were no exceptional observations when removing duplicates between the databases. This step concludes the Work Package 2.

Step	Total	ACM	IEEE	Springer	Science Direct
Search	7054	874	308	5789	83
Duplicates per Database	4074	51	13	4006	5
Duplicates across Databases	204	29	106	66	1
After Duplicate Removal	2776	794	189	1717	77
After Automatic Filtering	564	178	26	349	11
After Review and Voting	100	67	16	13	4

Table 3.1: Total Publication Results.

3.3 Filtering and Review of Publications

3.3.1 Filtering and Exclusion

Since it is impractical to manually review and vote on more than 1000 papers, I had to filter out as many irrelevant papers as possible. In the paper of Kuhrmann et al. [10] they describe the filtering either as a separate step, or already incorporated into the query building step. In the latter, one would find keywords identifiying irrelevant papers during the query construction when comparing with the result set and then exclude papers in the query itself. This however leads to complications depending on the search engine used, as it increases the complexity of the queries. For this reason I opted to use the first approach, where I would export all the query results and then based on the results build filter lists with keywords of irrelevant papers.

I achieved this in two steps. In the first step I evaluated the occurrence of each word in the titles of my reference set of papers (see Appendix A about the reference set) and compared those with the occurrences in the titles of the exported papers from the search engines. I was then able to determine keywords which did not occurr in my reference set and rank them by number of occurrences. I had to ignore any words, which could be in any title, such as "a", "and", "also", "the" etc., and I also ignored any words which are obviously not related to the field under study, e.g.: “optical”, “tv”, or “music”.

I then filtered the publications based on those words. Afterwards I repeated the procedure stated above with the abstracts of the collected papers. This lead to a higher exclusion rate compared to the title exclusion because typically there are more words in the abstracts. For reproducability I have appended the full list of exclusion keywords for title and abstract to this thesis as Appendix B. Table 3.1 shows the amount of publications filtered out in total and on a per database basis. The keywording and exclusion process reduced the number of papers down to a total of 564 which seemed reasonable to work with (also compared to systematic maps such as a map by Fernández et al. [7]).

3.3.2 Manual Review

For the manual review I got help from another person to ensure a more objective result set. I decided to follow one of the alternative approaches described in the paper by Kuhrmann et al. [10] where the reviewers vote independently and then meetup for a workshop where they discuss and decide on any publications where they did not both agree with their votes. When deciding whether to count a publication as relevant or not, I have devised the following process based on the inclusion criteria in Table 2.2:

1. If the title makes it explicit, that the publication is not related to the field under study (namely: utilizing the elastic cloud for Big Data Stream Processing) then mark the publication as not relevant.
2. Otherwise read the Abstract. If the abstract makes it explicit, that the publication is not related to our field of interest then mark the publication as not relevant.
3. Otherwise mark the publication as relevant.

As can be seen in Table 3.1 the review and voting procedure resulted in 100 relevant publications. The table also displays the composition of those publications based on the online libraries I extracted them from.

Chapter 4

Data Evaluation and Mapping

After I completed the collection and filtering stages of the data collection process, and we finished the voting procedure, I had to start on evaluating and classifying the final resulting set of papers.

4.1 Classification

To give an insight into the maturity and areas of research in the field under study, I had to classify the papers based on research- and contribution type. The results of the classification were visualized according to an example from Petersen et al. [14] in a two dimensional scatter plot, as this type of visualization is quite good at conveying information about the field in one single glance.

4.1.1 Research Type

Petersen et al. [14] recommend using the classification schema from Wieringa et al. [17]. I used this schema to determine what research has been done in the publications. The following list states the original category definition taken from a paper by Wieringa et al. [17]:

Evaluation Papers: “This is the investigation of a problem in RE practice or an implementation of an RE technique in practice. If it reports on the use of an RE technique in practice, then the novelty of the technique is not a criterion by which the paper should be evaluated. Rather, novelty of the knowledge claim made by the paper is a relevant criterion, as is the soundness of the research method used. In general, research results in new knowledge of causal relationships among phenomena, or in new knowledge of logical relationships among propositions. Causal properties are studied empirically, such as by case study, field study, field experiment, survey, etc. Logical properties are studied by conceptual means, such as by mathematics or logic. Whatever the method of study, it should support the conclusions stated in the paper." [17]

"Evaluation criteria for this kind of paper are:
Is the problem clearly stated?
Are the causal or logical properties of the problem clearly stated?
Is the research method sound?
Is the knowledge claim validated?
In other words, is the conclusion supported by the paper?
Is this a significant increase of knowledge of these situations?
In other words, are the lessons learned interesting?
Is there sufficient discussion of related work?" [17]

Proposal of Solution: "This paper proposes a solution technique and argues for its relevance, without a fullblown validation. The technique must be novel, or at least a significant improvement of an existing technique. A proof-of-concept may be offered by means of a small example, a sound argument, or by some other means." [17]

"Evaluation criteria are:
Is the problem to be solved by the technique clearly explained?
Is the technique novel, or is the application of the techniques to this kind of problem novel?
Is the technique sufficiently well described so that the author or others can validate it in later research?
Is the technique sound?
Is the broader relevance of this novel technique argued?
Is there sufficient discussion of related work?
In other words, are competing techniques discussed and compared with this one?" [17]

Validation Research: "This paper investigates the properties of a solution proposal that has not yet been implemented in RE practice. The solution may have been proposed elsewhere, by the author or by someone else. The investigation uses a thorough, methodologically sound research setup. Possible research methods are experiments, simulation, prototyping, mathematical analysis, mathematical proof of properties, etc." [17]

"Evaluation criteria are similar to those for evaluation research:
Is the technique to be validated clearly described?
Are the causal or logical properties of the technique clearly stated?
Is the research method sound?
Is the knowledge claim validated (i.e. is the conclusion supported by the paper)?
Is it clear under which circumstances the technique has the stated properties?
Is this a significant increase in knowledge about this technique?
Is there sufficient discussion of related work?" [17]

Philosophical Papers: "These papers sketch a new way of looking at things, a new conceptual framework, etc." [17]

"Evaluation criteria are:
Is the conceptual framework original?
Is it sound?
Is the framework insightful?" [17]

Opinion Papers: "These papers contain the author's opinion about what is wrong or good about something, how we should do something, etc." [17]

"Evaluation criteria are:
Is the stated position sound?
Is the opinion surprising?
Is it likely to provoke discussion?" [17]

Personal Experience Papers: "In these papers, the emphasis is on what and not on why. The experience may concern one project or more, but it must be the author's personal experience. The paper should contain a list of lessons learned by the author from his or her experience. Papers in this category will often come from industry practitioners or from researchers who have used their tools in practice, and the experience will be reported without a discussion of research methods. The evidence presented in the paper can be anecdotal." [17]

"Evaluation criteria are:
Is the experience original?
Is the report about it sound?
Is the report revealing?

Is the report relevant for practitioners?" [17]

Additional Category

I decided to add another category based on some papers, which did not fit in any of the categories proposed in the paper by Wieringa et al. [17], since they were not written from the personal accounts and viewpoints of the authors. The new Category can be described as follows:

Review and Summary Papers: Papers where the author/authors is/are reviewing or summarizing the evolution of an area of research in a historical fashion. Those papers do not have to be written from the authors personal opinion or viewpoint. Papers in this category often are Systematic Literature Reviews, Systematic Literature Maps or are simply recollections of the separate steps and advances in the particular area of research.

4.1.2 Contribution Type

In accordance with Petersen et al. [14] I also classified the papers based on contribution type. However, they did not recommend a specific classification scheme for this facet, so I followed their steps to build and create my own categories. As the first step I counted the occurrences of each word in the abstract and selected some categories based on the results. Approach, algorithm, framework and tool occurred frequently for example, and thus were the first categories in my classification scheme. As stated in Petersen et al. [14] the classification scheme evolved during the review of publications. I merged some categories and created new ones where a paper would not fit any other category. In the end I decided on seven categories which can be described as follows:

Metric: Papers in this category contribute a sort of metric for a Distributed Stream Processing System (DSPS), which is utilizing elastic properties of cloud environments. Papers which are presenting a benchmarking tool or system, which also provide the metrics for comparison within the benchmark should also be classified into this category. The reasoning is, that the main research work has gone into forming comparable and meaningful metrics.

Model: This category contains papers which are providing and researching a model on which approaches, algorithms or other research can be verified.

Process: Papers describing the steps of a process to reach a certain goal. The goal itself is irrelevant for the classification as the process itself is the main focus of the publication.

Discussion: These papers are usually conducting Literature Reviews and are discussing or outlining challenges based on already existing research papers. However, they may also handle personal opinions, experiences or original thoughts of the authors.

Approach/Algorithm: Papers which are outlining a new approach to a problem which does not necessarily have to be proposed by the authors themselves, or are detailing an algorithm which solves a problem. I have merged those categories into one during the review phase, because papers that describe a new algorithm also usually describe a new approach to a problem.

Framework: A complete working framework, may also be called "system", is the core of papers in this category. The frameworks do not have to be fully implemented or validated, as the contribution to the research field does not change based on its status.

Tool: Contrary to a framework, these papers contribute a smaller part of a complete system, or a new addition to an existing framework. Tools are distinct from approaches/algorithms, as they are either tailored to a specific framework or are self-contained solutions, which combine multiple algorithms into one component. An example would be a scheduler, because those are usually for specific frameworks and contain multiple algorithms and approaches to achieve their goal.

4.2 Additional Evaluation

Besides the two classification facets, I evaluated additional metadata from the collected publications. This information includes publication year, publication type and publisher. The data about the publication year will help me answer the first research question, as the information is directly linked to RQ 1.1 (Table 2.1). The other graphs are not directly related to any research question, however, they still contribute knowledge about the field under study.

Chapter 5

Results

5.1 RQ 1: Maturity of the Research Field

To determine the maturity of the field under study I applied the maturity phases described by Martha Vos [16]. Martha Vos [16] argued that the usage of general maturity measures is appropriate, due to the similarities between other approaches to measure maturity in other research fields. She determined the three phases for the field of IoT, however they are still applicable to the field under study. The applicability is justified by her statement that she based her ranking on a set of general maturity criteria. The maturity phases are defined as follows:

- "**Immature Phase:** Lower range of topics and methodologies, with a few researchers focusing on the area" [16]
- "**Growth Phase:** A range of methodologies with theory developing" [16]
- "**Mature Phase:** Quantitative hypothesis testing with a wide variety of research methods and approaches" [16]

I will use the contribution type classification to rank the field under study into the correct maturity phase, as the phases are described in relation to a range of topics and methodologies. To confirm the ranking I will evaluate the maturity based on the research type classification and the engineering cycle seen in Figure 5.1 which have been linked together in a paper by Wieringa et al. [17].

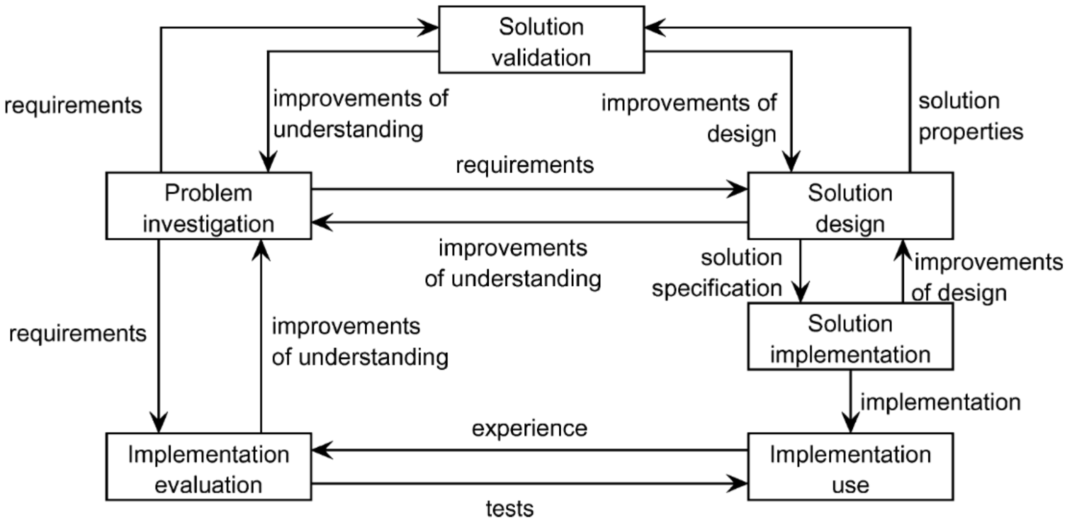

Figure 5.1: Engineering Cycle taken from Wieringa et al. [17].

They did not explicitly state a maturity ranking, however the relationship they established between the engineering cycle and their research type classification allows me to draw conclusions about the maturity of the field under study.
The following sections will answer the first research question as defined in Table 2.1 and rank the maturity of the field under study.

5.1.1 Number of Publications

As is visible from the publications per year histogram (Figure 5.2), the first relevant publication came out in 2000 as the only publication that year. The amount of publications released each year started to rise in 2012 and came to a peak with 22 publications released in 2015. After 2015 we can see the amount of publications released each year drop over the next two years. It is important to note, that the data for the year 2018 is not complete as the data search and export was executed in August 2018. If one assumes that the output of publications is steady over the course of one year, we can conclude, that the total number of publications released would be around the same level as in 2017.

Looking at the development of the total numbers of papers released (Figure 5.3) one can see, that the field under study started to grow more quickly in the years 2010 to 2012. Even after the peak in 2015, the field under study is growing at the same level as 2012. I was not able to find a valid explanation for the peak in the histogram, however I am examining a hypothesis, which might give an explanation, in the discussion part of the thesis.

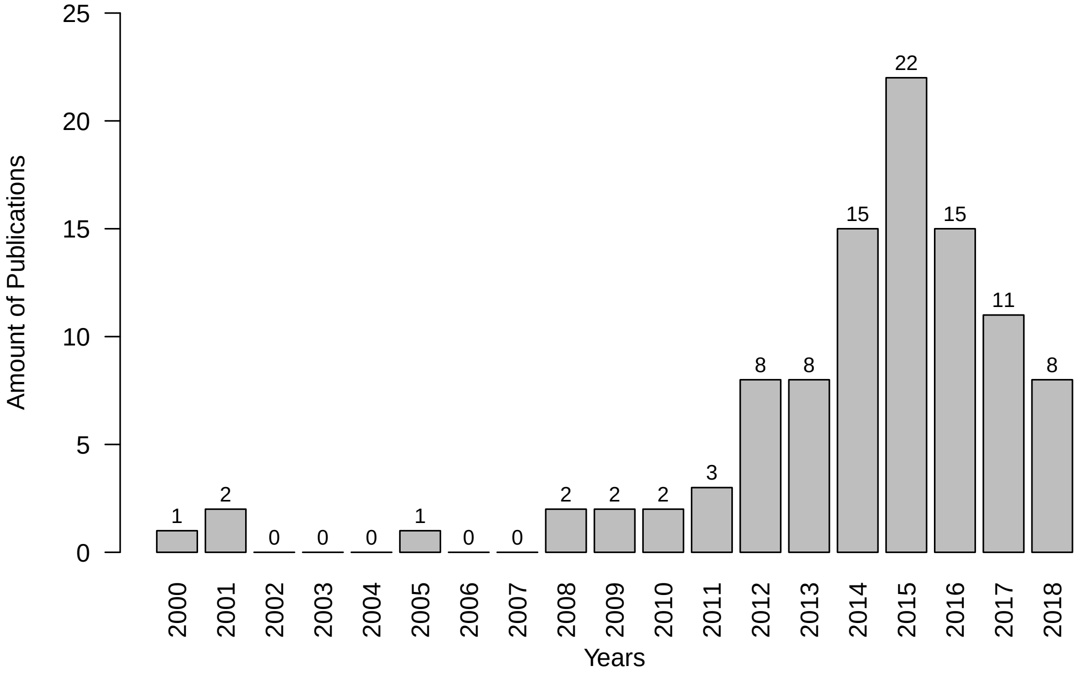

Figure 5.2: Histogram of the Publications released each year.

5.1.2 Maturity of the research field

An important part of the maturity ranking is the classification of result sets into various facets. For the ranking from Vos [16] the classification facet was the contribution type which grouped publications into one of seven categories (namely: Metric, Process, Model, Discussion, Tool, Approach/Algorithm and Framework). Figure 5.4 displays the composition as a bar chart showing the total amount of papers in each category. One can see that a minority of the categories (Approach/Algorithm and Framework) are containing the majority of the publications (64 out of 100). This in turn means, that the range of topics in the field under study is concentrated on mainly two categories. According to the maturity phases the field fits in between the immature and the growth phase, as the range of topics is focused mainly on a few categories, however there are some publications in other categories as well which suggest some growth is happening within the field.

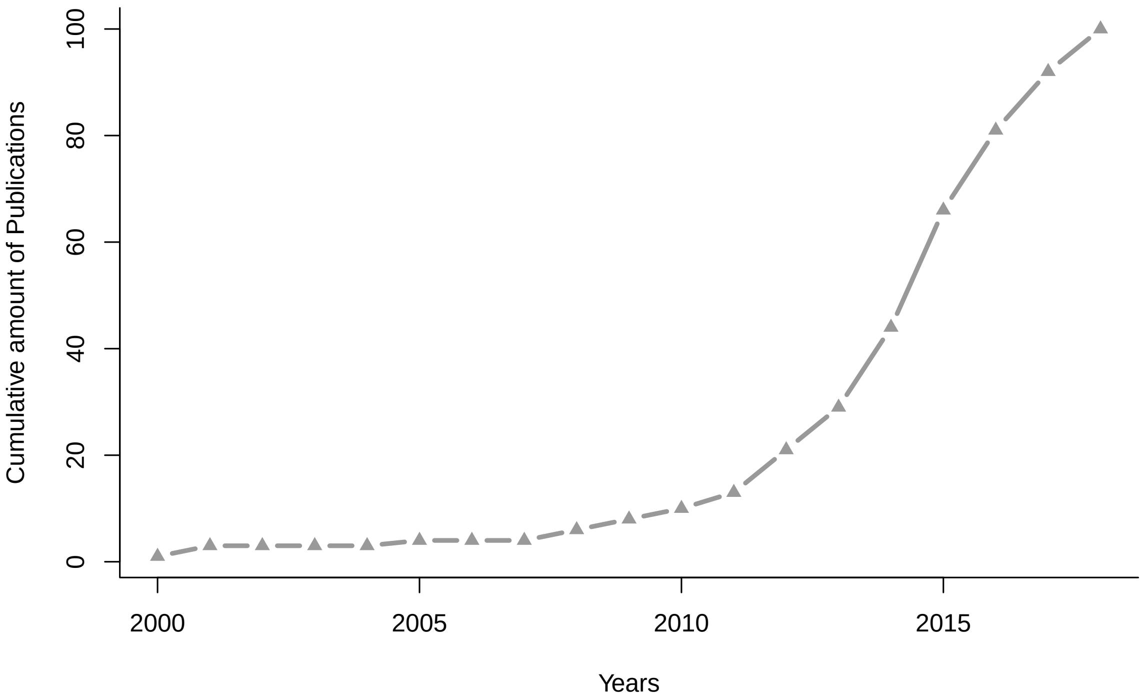

Figure 5.3: Total Sum of Publications over the years.

Additionally the theoretical categories (such as Metric, Process, Model and Discussion) contain the least amount of papers which also reinforces the claim that the field under study is in the immature stage of research based on the description of the growth phase, as this phase should show a developing theory within the field.
To confirm my assessment of the maturity of the field under study I determined the maturity based on the paper from Wieringa et al. [17]. In their paper Wieringa et al. [17] presented their adaptation of the engineering cycle shown in Figure 5.1 which they also linked to their publication classification scheme. Even though Wieringa et al. [17] did not specify an order in which the engineering cycle has to be traversed I can draw conclusions by their mapping of the engineering cycle to the classification categories.
From Figure 5.5 it is evident that the most work has gone into validating and proposing solutions to problems. Since those two steps are only a minor part of the complete cycle, and the field could have only cycled between those two steps, I conclude that the field is quite immature in the research progress. Especially because there is little to no evaluation research compared to validation research, so the solutions have not been proven to work in actual practice.

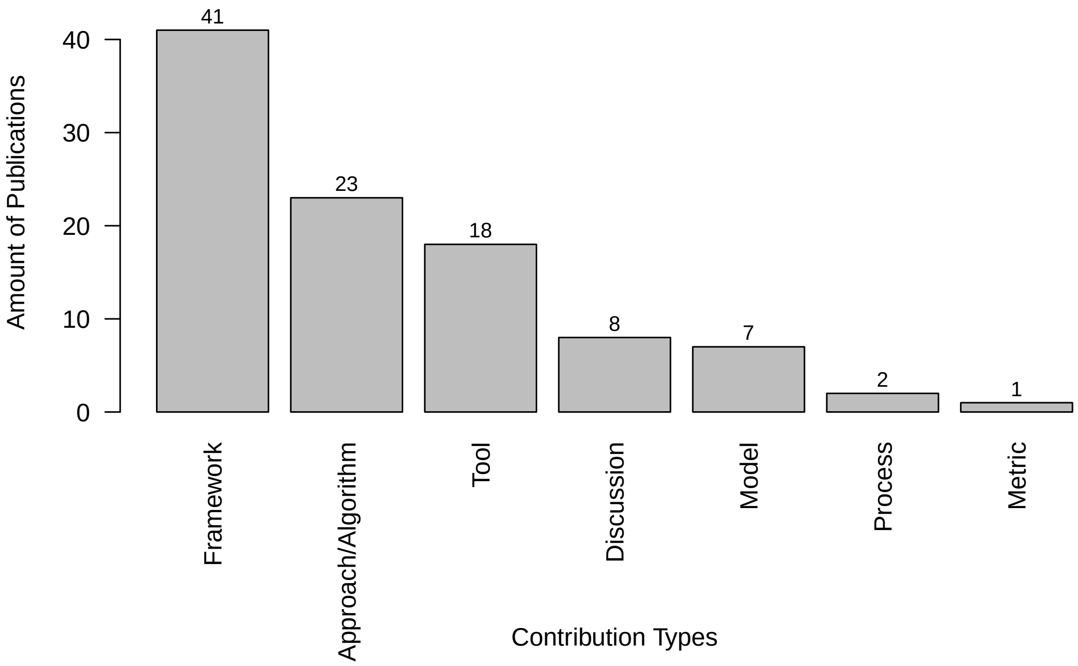

Figure 5.4: Contribution Type Classification.

5.2 RQ 2: Areas of Research Requiring more Work

This question can be answered by evaluating the research and contribution type classification and analyzing where there is a lack of publications and therefore a lack of work in the field.

5.2.1 Contribution Type Gaps

Looking at Figure 5.4 one can see that I was able to identify seven areas of research within the current state of the research field. A definition of those categories can be found at Section 4.1.2. It is evident that the focus of the research community is on frameworks and approaches/algorithms. The largest amount of work has gone into developing and designing frameworks to enable the utilization of elasticity. The next biggest area of contribution are Approaches/Algorithms which also fits in with the observation from research question 1. As the research field is very immature the main focus has been laid on getting the basics of the field right, to have a baseline from which to improve processes and techniques.

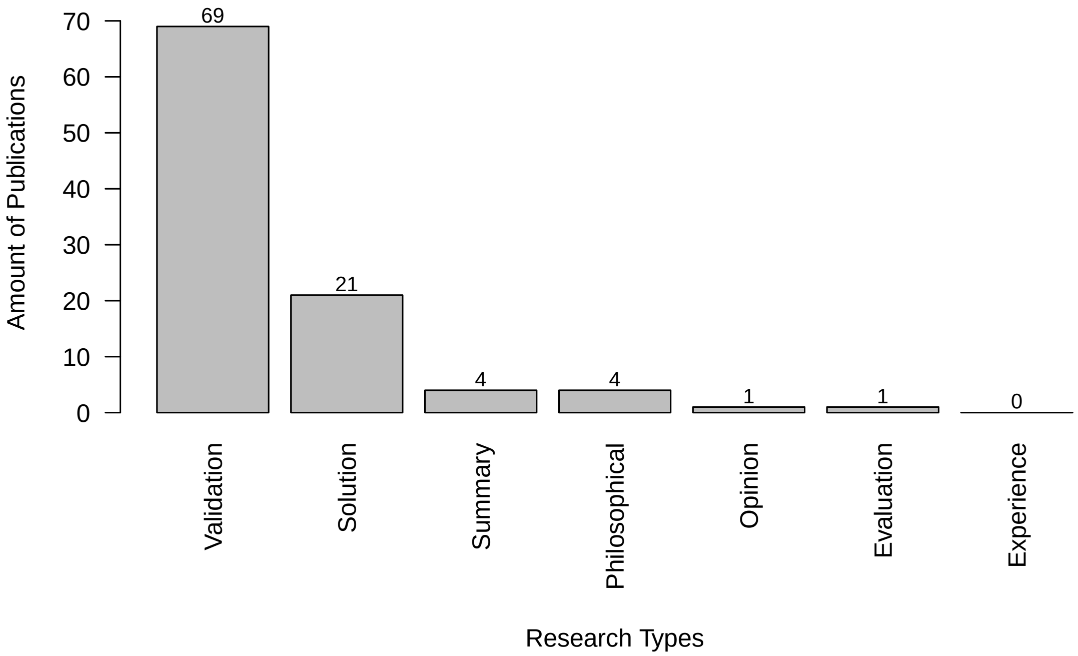

Figure 5.5: Research Type Classification.

5.2.2 Research Type Gaps

Looking at Figure 5.5 the differences between categories are more apparent when compared to the contribution classification in Figure 5.4. By far the most papers have been categorized as validation and solution papers. As with the contribution type, this observation also reinforces the maturity ranking from research question 1. In the research type context the biggest lack of effort can be seen in evaluation and experience research, which shows that there have not been any notable practical uses of techniques from the field under study. There is also a lack of papers addressing the state of the field itself in the form of summary papers, or sketching new views on the field as philosophical papers.

5.2.3 Combining Classifications into a visual Map

To enhance the visual presentation of the classification I have combined both classification facets into one graph, which I designed according to an example from Petersen et al. [14] in their 2008 paper. Figure 5.6 shows this map of the field under study.

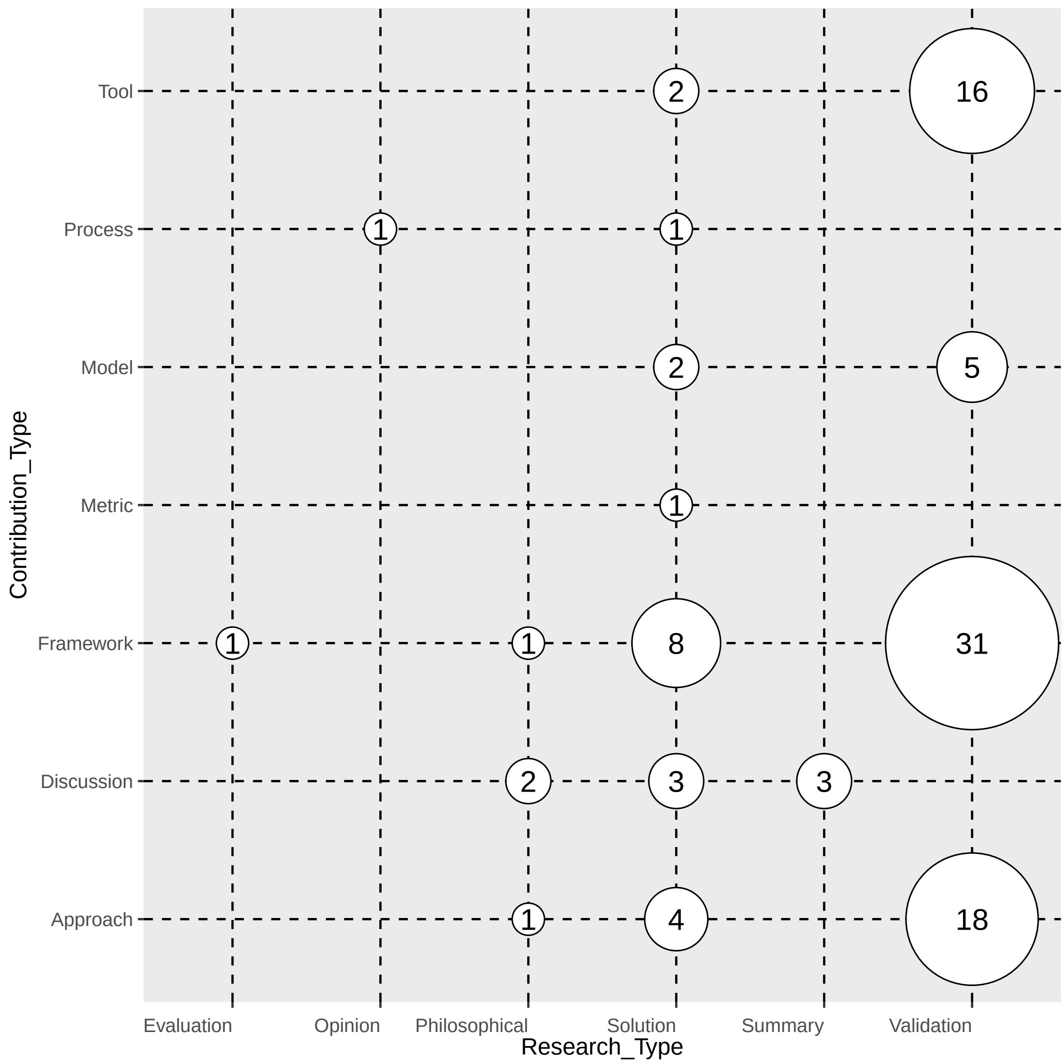

Figure 5.6: Visual Map of Contribution- and Research Type.

From the map it is visible, that there are multiple areas of research which have not been covered at all yet, especially in the left half of the map. Another observation is, that there has been a lot of work put into validation papers for most contribution types, however no papers have been submitted for the discussion, metric and process categories. The solution category shows no significant gaps of research, as there is at least one paper for each contribution type category. Even though the total amount of papers is less than in the validation category, the work into the field has been distributed more evenly across contribution types. In general there are a number of gaps identifiable by observing the visual map.

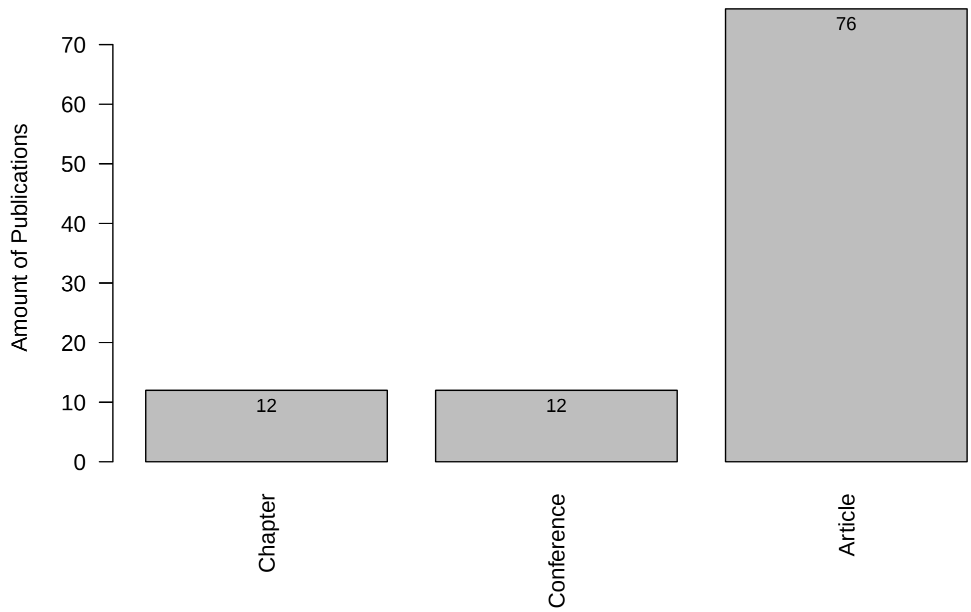

Figure 5.7: Publication Type.

5.3 Additional Evaluation

In Addition to answering the research questions above, I have analyzed the publication types and publishers of the field. This information shows which publication channels are used by researchers of the field under study. An overview of the different publication types can be seen in Figure 5.7, which shows three different types. The biggest part of papers have been published as articles with equal parts of conference papers and chapters. In the publishers chart (Figure 5.8) we can observe that the most relevant publications have come from the "IEEE" and "ACM" publishers. Other notable publishers are "Springer" and "Elsevier", even though they are contributing less than 25 % of the total relevant papers. Of course this chart is influenced by my choice of Search engines, as each Search engine mostly has their own publications in the search index. This result was also expected because "IEEE" and "ACM" are the typical software engineering and computer science publishers.

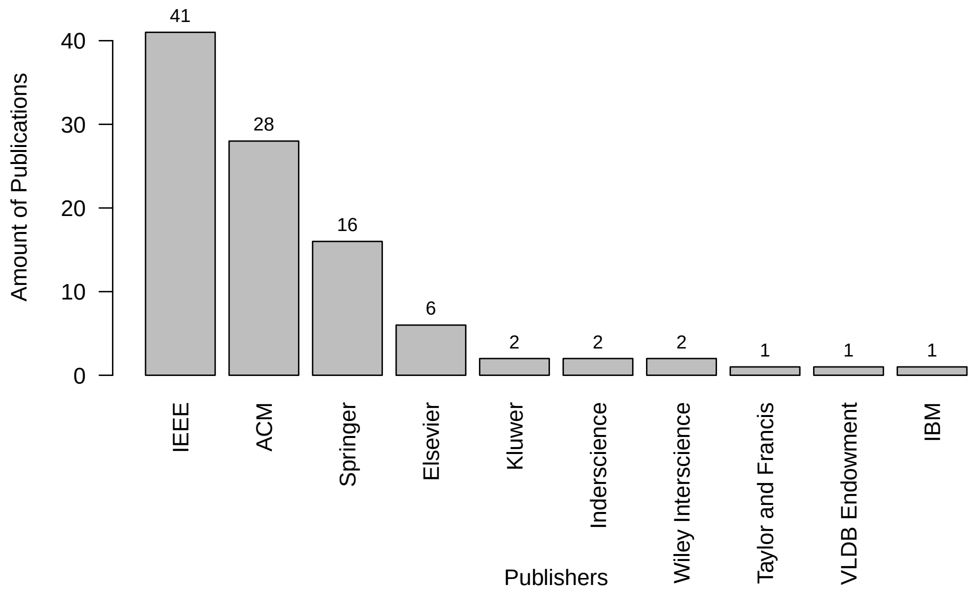

Figure 5.8: Composition of the publishers.

Chapter 6

Discussion

The findings in this thesis help to define the state of the art of the field under study and are supporting my expectations of the field. In Section 5.1 there are some interesting observations about the number of publications in total, as there is a peak of submissions between 2013 and 2016. This would mean that the field under study is currently in regression but this is not necessarily what is happening. Since the field is quite small with 100 publications in total, there is a strong collaboration between authors. Together with the fact that most publications are Validation Research with a typically longer research time than experience papers for example, could lead to peaks in the submission history, as authors release their papers at the same time, and then spend the next year or years researching the next topic. This peak could be especially noticeable if the authors all started work on the field at the same time, in this case around 2012 probably.

Besides that, the results show conclusively, that the field under study is an immature area of research. There are some aspects which show that the field is developing and moving to the growth stage. This result correlates with the number of papers, which have only significantly increased in 2012 (six years before this thesis).

The immaturity of the field fits in with the observation of research areas within the field under study as seen in Figures 5.4, 5.5 and 5.6, which show that there are significant gaps in research. When looking at the visual map (Figure 5.6), it is easy to see the main gaps especially in the left half of the chart. However, there are many areas where there have been some publications, but those can still be mentioned as gaps compared to areas with more contribution. My research shows that the biggest gaps are in the evaluation and opinion research type. This reinforces the maturity ranking, as those categories require some experience in practice. Evaluation papers need to evaluate the research in practice, and opinions are usually formed after investigation of multiple different approaches, which could not happen yet.

6.1 Limitations

Even though I followed a methodology that was accepted by the research community as valid, and is used in other maps such as: Condori-Fernandez et al. [6] and Kuhrmann, Fernández & Tiessler [11], there are a few limitations to my research. The biggest one is my choice of online libraries, as those influence all other steps in the research process. It is possible, that there exists an online library which offers a lot more papers in the field under study which I did not know about.

Additionally the research- and contribution type classification relied on the abstracts of the publications. Depending on the authors, the abstracts might be quite short or not contain relevant information. This could lead to false classifications. However for most papers the abstracts were of sufficient length and detail to be classified correctly.

6.2 Future Work

Due to the classification and identification of research gaps, it is possible to show areas in the field under study which require more work and research. Even though there are clear gaps, the maturity rank shows that the field in general requires some more work to increase the number of approaches. This is necessary to reach the growth phase, after which the next steps are closing the research gaps. In general, future work could go into evaluation papers, as this explores the utilization of solutions in practice. Opinion papers are another paper type in need of more work. Those publications can give some guidance to new researchers, which have not had a lot of contact with the field.

Chapter 6

Bibliography

[1] D. J. Abadi, Y. Ahmad, M. Balazinska, U. Çetintemel, M. Cherniack, J.-H. Hwang, W. Lindner, A. Maskey, A. Rasin, E. Ryvkina, N. Tatbul, Y. Xing, S. B. Zdonik, S. Schneider, H. Andrade, B. Gedik, A. Biem, K. L. Wu, S. Schneider, M. Hirzel, K. L. Wu, C. Hochreiner, M. Vogler, S. Schulte, S. Dustdar. "Elastic scaling for data stream processing." In: *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)* 25.6 (2014), pp. 100–107. ISSN: 1530-2075 (cit. on p. 7).

[2] D. J. Abadi, Y. Ahmad, M. Balazinska, U. Çetintemel, M. Cherniack, J.-H. Hwang, W. Lindner, A. Maskey, A. Rasin, E. Ryvkina, N. Tatbul, Y. Xing, S. B. Zdonik. "The Design of the Borealis Stream Processing Engine." In: *Cidr* (2005), pp. 277–289 (cit. on p. 8).

[3] J. Bailey, M. Turner. "Evidence relating to Object-Oriented software design : A survey." In: (2007), pp. 482–484. DOI: 10.1109/ESEM.2007.58 (cit. on p. 2).

[4] F. Bezerra, C. H. Favacho, R. Souza, C. D. Souza. "Towards Supporting Systematic Mappings Studies: An Automatic Snowballing Approach." In: *SBBD Proceedings* October (2014), pp. 167–176 (cit. on p. 3).

[5] J. Cervino, E. Kalyvianaki, J. Salvachua, P. Pietzuch. "Adaptive Provisioning of Stream Processing Systems in the Cloud." In: *2012 IEEE 28th International Conference on Data Engineering Workshops* (2012), pp. 295–301. DOI: 10.1109/ICDEW.2012.40. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6313696 (cit. on p. 7).

[6] N. Condori-Fernandez, M. Daneva, K. Sikkel, R. Wieringa, O. Dieste, O. Pastor. "A systematic mapping study on empirical evaluation of software requirements specifications techniques." In: *2009 3rd International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2009, pp. 502–505. ISBN: 978-1-4244-4842-5. DOI: 10.1109/ESEM.2009.5314232. URL: http://ieeexplore.ieee.org/document/5314232/ (cit. on p. 30).

[7] D. M. Fernández, S. Ognawala, S. Wagner, M. Daneva. "Where do we stand in requirements engineering improvement today? First results from a mapping study." In: (2017). ISSN: 19493789. DOI: 10.1145/2652524.2652555. arXiv: 1701.05497 (cit. on p. 11).

[8] B. Gedik, S. Schneider, M. Hirzel, K. L. Wu. "Elastic scaling for data stream processing." In: *IEEE Transactions on Parallel and Distributed Systems* 25.6 (2014), pp. 1447–1463. ISSN: 10459219. DOI: 10.1109/TPDS.2013.295 (cit. on p. 7).

[9] C. Hochreiner, M. Vogler, S. Schulte, S. Dustdar. "Elastic Stream Processing for the Internet of Things." In: *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*. IEEE, 2016, pp. 100–107. ISBN: 978-1-5090-2619-7. DOI: 10.1109/CLOUD.2016.0023. URL: http://ieeexplore.ieee.org/document/7820260/ (cit. on p. 7).

[10] M. Kuhrmann, D. M. Fernández, M. Daneva. "On the pragmatic design of literature studies in software engineering: an experience-based guideline." In: *Empirical Software Engineering* 22.6 (2017), pp. 2852–2891. ISSN: 15737616. DOI: 10.1007/s10664-016-9492-y. arXiv: 1612.03583 (cit. on pp. 3, 7, 10, 11).

[11] M. Kuhrmann, D. Méndez Fernández, M. Tiessler. "A mapping study on the feasibility of method engineering." In: *Journal of Software: Evolution and Process* 26.12 (2014), pp. 1053–1073. ISSN: 20477473. DOI: 10.1002/smr.1642. URL: http://doi.wiley.com/10.1002/smr.1642 (cit. on p. 30).

[12] M. Kuhrmann, C. Konopka, P. Nellemann, P. Diebold, J. Münch. "Software process improvement: where is the evidence?: initial findings from a systematic mapping study." In: *Proceedings of the 2015 International Conference on Software and System Process - ICSSP 2015*. New York, New York, USA: ACM Press, 2015, pp. 107–116. ISBN: 9781450333467. DOI: 10.1145/2785592.2785600. URL: http://dl.acm.org/citation.cfm?doid=2785592.2785600 (cit. on p. 5).

[13] P. M. Mell, T Grance. *The NIST definition of cloud computing*. Tech. rep. Gaithersburg, MD: National Institute of Standards and Technology, 2011. DOI: 10.6028/NIST.SP.800-145 (cit. on p. 1).

[14] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson. "Systematic Mapping Studies in Software Engineering." In: *12Th International Conference on Evaluation and Assessment in Software Engineering* 17 (2008), p. 10. ISSN: 02181940. DOI: 10.1142/S0218194007003112. arXiv: 339:b2535. URL: http://www.robertfeldt.net/publications/petersen_ease08_sysmap_studies_in_se.pdf (cit. on pp. 2, 3, 13, 16, 24).

[15] J. Thor, M. Ingibergsson, U. P. Schultz. "On the Use of Safety Certification Practices in Autonomous Field Robot Software Development : A Systematic Mapping Study." In: (2015), pp. 335–352 (cit. on p. 5).

[16] M. Vos. "Maturity of the Internet of Things Research Field: Or Why Choose Rigorous Keywords." In: (2015). URL: https://arxiv.org/ftp/arxiv/papers/1606/1606.01452.pdf (cit. on pp. 19, 21).

[17] R. Wieringa, N. Maiden, N. Mead, C. Rolland. "Requirements engineering paper classification and evaluation criteria: A proposal and a discussion." In: *Requirements Engineering* 11.1 (2006), pp. 102–107. ISSN: 09473602. DOI: 10.1007/s00766-005-0021-6 (cit. on pp. 13–16, 19, 20, 22).

All links were last followed on November 20th, 2018.

Appendices

Appendix A

Base Reference Papers before Snowballing

The reference set after the snowballing procedure:

Title:	Authors
The Design of the Borealis Stream Processing Engine	Daniel J. Abadi,Yanif Ahmad,Magdalena Balazinska,Ugur Cetintemel,Mitch Cherniack,Jeong-Hyon Hwang,Wolfgang Lindner,Anurag S. Maskey,Alexander Rasin,Esther Ryvkina,Nesime Tatbul,Ying Xing,Stan Zdonik
Elastic scaling of data parallel operators in stream processing	Scott Schneider,Henrique Andrade,Bugra Gedik,Alain Biem,Kun-Lung Wu
Elastic Stream Processing for Distributed Environments	Christoph Hochreiner,Stefan Schulte,Schahram Dustdar,Freddy Lecue
Network-aware operator placement for stream-processing systems	Peter Pietzuch, Jonathan Ledlie, Jeffrey Shneidman, Mema Roussopoulos, Matt Welsh, Margo Seltzer

Title:	**Authors**
Streamcloud: An elastic and scalable data streaming system	Vincenzo Gulisano,Ricardo Jimenez-Peris,Marta Patino-Martinez,Claudio Soriente,Patrick Valduriez
Load management and high availability in the borealis distributed stream processing engine	Nesime Tatbul,Yanif Ahmad,Ugur Cetintemel,Jeong-Hyon Hwang,Ying Xing,Stan Zdonik
Elastic stream processing in the cloud	Waldemar Hummer, Benjamin Satzger, Schahram Dustdar
Esc: Towards an elastic stream computing platform for the cloud	Benjamin Satzger, Waldemar Hummer ,Philipp Leitner, Schahram Dustdar
Elastic complex event processing	Thomas Heinze
Elastic stream computing with clouds	Atsushi Ishii, Toyotaro Suzumura
Elastic scaling for data stream processing	Bugra Gedik, Scott Schneider, Martin Hirzel, Kun-Lung Wu
Auto-scaling techniques for elastic data stream processing	Thomas Heinze, Valerio Pappalardo, Zbigniew Jerzak, Christof Fetzer
Flood: Elastic Streaming MapReduce	David Alves,Pedro Bizarro,Paulo Marques
Adaptive Provisioning of Stream Processing Systems in the Cloud	Javier Cervino, Evangelia Kalyvianaki, Joaquin Salvachua, Peter Pietzuch
Elastic Stream Processing for the Internet of Things	Christoph Hochreiner ,Michael Voegler , Stefan Schulte , Schahram Dustdar
Adaptive Control of Extreme-scale Stream Processing Systems	Lisa Amini, Navendu Jain , Anshul Sehgal, Jeremy Silber , Olivier Verscheure

Appendix B

List of Exclusion Keywords

B.1 Title

video,multimedia,mobile,wireless,live,virtual,3d,optical,http,p2p,method,peer-to-peer,rate, mining,bandwidth,multicast,delivery,support,content,ip,flecible,reality,object,joint,grid, tcp,cellular,end-to-end,graphics,channel,physical,rendering,video-on-demand,vod, mpeg-4,augmented,facial,videos,multiservice,communications,proxy,mapping,networks, radio,cross-layer,playout,presentations,tv,ad,images,mobility,markov,streammine3g,cell, http,bit,music,multi-stream,concurrent,delivery

B.2 Abstract

recognition,classification,mobile,mining,search,images,digital,wireless,networks,speed, packet,interaction,neural,survey,transfer,3d,link,training,protocol,robot,rdf,reliability, remote,market,buffer,distance,traffic,artificial,embedded,audio,object,medical,chapter, function,functions,provenance,cipher,secure,index,phase,queries,news,metadata,surface, sampling,ice,message,intrusion,sea,client,session,cyber,dimension,urban,fluid,main,linear, dataflow,numerical,vector,criminal,transaction,graphical,gpu,spectral,in-network,games, cognitive,viewers,file,road,blockchain,trajectory,wind,buffering,animation,visualization, fourier,multicast,compression,facial,motor,crime,sql,methane,bit,ip,call,geospatial, ontology,drug,channels.,cities,region,ontologies,momentum,gas,bytes,particle,liquid, biometric,vehicular,oscillations,streaming,biofilter,velocity,wave,filtered,verification, multi-core,floating-point,languages,plasma,millwheel,emotion,fpga-based,cyber-physical, friendship,algebra,hive,manufacturing

Appendix C

Final Set of Papers

The final set of publications relevant to the field under study:

title	year	authors
A Container-Based Elastic Cloud Architecture for Pseudo Real-Time Exploitation of Wide Area Motion Imagery (WAMI) Stream	2017	Ryan Wu and Bingwei Liu and Yu Chen and Erik Blasch and Haibin Ling and Genshe Chen
A Cost-Based Approach to Adaptive Resource Management in Data Stream Systems	2008	Michael Cammert and Jurgen Kramer and Bernhard Seeger and Sonny Vaupel
A Game-Theoretic Approach for Elastic Distributed Data Stream Processing	2016	Gabriele Mencagli
A multi-layer software architecture framework for adaptive real-time analytics	2016	A. Vakali; P. Korosoglou; P. Daoglou
A Preventive Auto-Parallelization Approach for Elastic Stream Processing	2017	R. K. Kombi; N. Lumineau; P. Lamarre
A QoS-Aware Resource Allocation Controller for Function as a Service (FaaS) Platform	2017	Mohammad Reza, Hoseiny Farahabady, Young Choon Lee, Albert Y. Zomaya, Zahir Tari
A Task-level Adaptive MapReduce Framework for Real-time Streaming Data in Healthcare Applications	2015	Fan Zhang and Junwei Cao and Samee U. Khan and Keqin Li and Kai Hwang
ACDS: Adapting Computational Data Streams for High Performance	2000	C. Isert, K.Schwan

title	year	authors
Active Streams-An Approach to Adaptive Distributed Systems	2001	F. E. Bustamante and P. Widener
Adaptive Provisioning of Stream Processing Systems in the Cloud	2012	Javier Cervino and Evangelia Kalyvianaki and Joaquin Salvachua and Peter Pietzuch
Adaptive Stream Processing Using Dynamic Batch Sizing	2014	Tathagata Das and Yuan Zhong and Ion Stoica and Scott Shenker
An Adaptive Event Stream Processing Environment	2012	Samujjwal Bhandari
An Adaptive Replication Scheme for Elastic Data Stream Processing Systems	2015	Thomas Heinze and Mariam Zia and Robert Krahn and Zbigniew Jerzak and Christof Fetzer
AuDy: Automatic Dynamic Least-Weight Balancing for Stream Workloads Scalability	2014	Pedro Martins and Maryam Abbasi and Pedro Furtado
Automatic Scaling of Resources in a Storm Topology	2018	Evangelos Gkolemis, Katerina Doka, Nectarios Koziris
Auto-scaling Techniques for Elastic Data Stream Processing	2014	Thomas Heinze and Valerio Pappalardo and Zbigniew Jerzak and Christof Fetzer
Benchmarking Elastic Query Processing on Big Data	2015	Dimitri Vorona, Florian Funke, Alfons Kemper, Thomas Neumann
C-Cube: Elastic Continuous Clustering in the Cloud	2013	Zhenjie Zhang and Hua Lu and Hu Shu and Yin Yang and Zhihong Chong
CE-Storm: Confidential Elastic Processing of Data Streams	2015	Nick R. Katsipoulakis and Cory Thoma and Eric A. Gratta and Alexandros Labrinidis and Adam J. Lee and Panos K. Chrysanthis
ChronoStream: Elastic stateful stream computation in the cloud	2015	Y. Wu; K. L. Tan
Cloud-based Data Stream Processing	2014	Thomas Heinze and Leonardo Aniello and Leonardo Querzoni and Zbigniew Jerzak
Cost-Efficient Elastic Stream Processing Using Application-Agnostic Performance Prediction	2016	S. Imai; S. Patterson; C. A. Varela
C-Stream: A Co-routine-Based Elastic Stream Processing Engine	2018	Semih Sahin and Busra Gedik

title	year	authors
Decentralized self-adaptation for elastic Data Stream Processing	2018	Valeria Cardellini, Francesco Lo Presti, Matteo Nardelli, Gabriele Russo Russo,
Demo: Elastic Mapreduce-style Processing of Fast Data	2013	Kasper Grud Skat Madsen and Yongluan Zhou
Distributed data stream processing and edge computing: A survey on resource elasticity and future directions	2018	Marcos Dias de Assunção, Alexandre da Silva Veith, Rajkumar Buyya,
Distributed Operation in the Borealis Stream Processing Engine	2005	Yanif Ahmad and Bradley Berg and Ugur Cetintemel and Mark Humphrey and Jeong-Hyon Hwang and Anjali Jhingran and Anurag Maskey and Olga Papaemmanouil and Alexander Rasin and Nesime Tatbul and Wenjuan Xing and Ying Xing and Stan Zdonik
DRS: Auto-Scaling for Real-Time Stream Analytics	2017	Tom Z. J. Fu and Jianbing Ding and Richard T. B. Ma and Marianne Winslett and Yin Yang and Zhenjie Zhang
Dynamic Aggregate: An Elastic Framework for QoS-Aware Distributed Processing of RFID Data on Enterprise Hierarchy	2014	Wen-sheng Chen and Ying-Jun Chen and Shiow-yang Wu
Dynamic and coordinated software reconfiguration in distributed data stream systems	2016	Rafael Oliveira Vasconcelos, Igor Vasconcelos, Markus Endler
Dynamic Datacenter Resource Provisioning for High-performance Distributed Stream Processing with Adaptive Fault-tolerance	2013	Paolo Bellavista and Antonio Corradi and Spyros Kotoulas and Andrea Reale
Dynamic Load Balancing Techniques for Distributed Complex Event Processing Systems	2016	Nikos Zacheilas, Nikolas Zygouras, Nikolaos Panagiotou, Vana Kalogeraki, Dimitrios Gunopulos
Dynamic Resource Management In a Massively Parallel Stream Processing Engine	2015	Kasper Grud Skat Madsen and Yongluan Zhou
Dynamically Scaling Apache Storm for the Analysis of Streaming Data	2015	Jan Sipke van der Veen and Bram van der Waaij and Elena Lazovik and Wilco Wijbrandi and Robert J. Meijer

title	year	authors
DynLW: Balancing and Scalability for Heavy Dynamic stream-DB Workloads	2014	Pedro Martins and Pedro Furtado
Efficient Stream Processing in the Cloud	2012	Dung Vu, Vana Kalogeraki, Yannis Drougas
Elastic Allocator: An Adaptive Task Scheduler for Streaming Query in the Cloud	2014	Zheng Han and Rui Chu and Haibo Mi and Huaimin Wang
Elastic and Scalable Processing of Linked Stream Data in the Cloud	2013	Danh Le-Phuoc and Hoan Nguyen Mau Quoc and Chan Le Van and Manfred Hauswirth
Elastic Complex Event Processing	2011	Thomas Heinze
Elastic Complex Event Processing Exploiting Prediction	2015	Nikos Zacheilas and Vana Kalogeraki and Nikolas Zygouras and Nikolaos Panagiotou and Dimitrios Gunopulos
Elastic Resource Provisioning for Batched Stream Processing System in Container Cloud	2017	Song Wu, Xingjun Wang, Hai Jin, Haibao Chen
Elastic Scaling for Data Stream Processing	2014	Bugra Gedik and Scott Schneider and Martin Hirzel and Kun-Lung Wu
Elastic Scaling for Distributed Latency-Sensitive Data Stream Operators	2017	T. d. Matteis; G. Mencagli
Elastic Scaling of Data Parallel Operators in Stream Processing	2009	Scott Schneider and Henrique Andrade and Bugra Gedik and Alain Biem and Kun-Lung Wu
Elastic stateful stream processing in storm	2016	V. Cardellini; M. Nardelli; D. Luzi
Elastic Stream Computing with Clouds	2011	Atsushi Ishii and Toyotaro Suzumura
Elastic Stream Processing for Distributed Environments	2015	C. Hochreiner; S. Schulte; S. Dustdar; F. Lecue
Elastic Stream Processing for the Internet of Things	2016	C. Hochreiner; M. Vögler; S. Schulte; S. Dustdar
Elastic Stream Processing in the Cloud	2013	Waldemar Hummer and Benjamin Satzger and Schahram Dustdar
Elastic Stream Processing with Latency Guarantees	2015	B. Lohrmann; P. Janacik; O. Kao
Elastic Streaming Semantic Engine for Web of Things	2015	Xi Chen and Huajun Chen and Ningyu Zhang and Jue Huang

title	year	authors
Elastic Symbiotic Scaling of Operators and Resources in Stream Processing Systems	2018	F. Lombardi; L. Aniello; S. Bonomi; L. Querzoni
Elastic-PPQ: A two-level autonomic system for spatial preference query processing over dynamic data streams	2018	Gabriele Mencagli, Massimo Torquati, Marco Danelutto,
Enabling A Load Adaptive Distributed Stream Processing Platform on Synchronized Clusters	2014	Xing Wu and Yan Liu
Enabling Elastic Stream Processing in Shared Clusters	2016	J. Li; C. Pu; Y. Chen; D. Gmach; D. Milojicic
Enabling Self-adaptivity in Component-based Streaming Applications	2009	Onur Derin and Alberto Ferrante
Esc: Towards an Elastic Stream Computing Platform for the Cloud	2011	Benjamin Satzger and Waldemar Hummer and Philipp Leitner and Schahram Dustdar
Evaluating CP Techniques to Plan Dynamic Resource Provisioning in Distributed Stream Processing	2014	Andrea Reale, Paolo Bellavista, Antonio Corradi, Michela Milano
FAS: A Flow Aware Scaling Mechanism for Stream Processing Platform Service Based on LMS	2017	Yongfeng Wu and Ruonan Rao and Pei Hong and Jin Ma
Flood: Elastic Streaming MapReduce	2010	David Alves and Pedro Bizarro and Paulo Marques
Flower: A Data Analytics Flow Elasticity Manager	2017	Alireza Khoshkbarforoushha and Rajiv Ranjan and Qing Wang and Carsten Friedrich
Integrating Fault-tolerance and Elasticity in a Distributed Data Stream Processing System	2014	Kasper Grud Skat Madsen and Philip Thyssen and Yongluan Zhou
Keep Calm and React with Foresight: Strategies for Low-latency and Energy-efficient Elastic Data Stream Processing	2016	Tiziano De Matteis and Gabriele Mencagli
Large-scale Real-time Semantic Processing Framework for Internet of Things	2015	Xi Chen and Huajun Chen and Ningyu Zhang and Jue Huang and Wen Zhang

title	year	authors
Latency-aware Elastic Scaling for Distributed Data Stream Processing Systems	2014	Thomas Heinze and Zbigniew Jerzak and Gregor Hackenbroich and Christof Fetzer
Living in the Present: On-the-fly Information Processing in Scalable Web Architectures	2012	David Eyers and Tobias Freudenreich and Alessandro Margara and Sebastian Frischbier and Peter Pietzuch and Patrick Eugster
Load adaptive and fault tolerant distributed stream processing system for explosive stream data	2016	M. Lee; M. Lee; S. J. Hur; I. Kim
Load Management and High Availability in the Borealis Distributed Stream Processing Engine	2008	Nesime Tatbul and Yanif Ahmad and Ugur Cetintemel and Jeong-Hyon Hwang and Ying Xing and Stan Zdonik
Low Cost Energy Forecasting for Smart Grids Using Stream Mine 3G and Amazon EC2	2014	Andre Martin and Rodolfo Silva and Andrey Brito and Christof Fetzer
Monitoring and Autoscaling IaaS Clouds: A Case for Complex Event Processing on Data Streams	2013	Omran Saleh and Francis Gropengießer and Heiko Betz and Waseem Mandarawi and Kai-Uwe Sattler
On Developing and Operating of Data Elasticity Management Process	2015	Tien-Dung Nguyen, Hong-Linh Truong, Georgiana Copil, Duc-Hung Le, Daniel Moldovan, Schahram Dustdar
On Engineering Analytics for Elastic IoT Cloud Platforms	2016	Hong-Linh TruongGeorgiana CopilSchahram DustdarDuc-Hung LeDaniel MoldovanStefan Nastic
Online Parameter Optimization for Elastic Data Stream Processing	2015	Thomas Heinze and Lars Roediger and Andreas Meister and Yuanzhen Ji and Zbigniew Jerzak and Christof Fetzer
Operator Scale Out Using Time Utility Function in Big Data Stream Processing	2014	Mahammad Humayoo and Yanlong Zhai and Yan He and Bingqing Xu and Chen Wang
Optimization of Load Adaptive Distributed Stream Processing Services	2014	Xing Wu and Yan Liu
Optimized Elastic Query Mesh for Cloud Data Streams	2015	Fatma Mohamed, Rasha M. Ismail, Nagwa L. Badr, M. F. Tolba

title	year	authors
Performance Characterization of the Serviotic API: An IoT-as-a-Service Data Management Platform	2015	Juan Luis Perez and David Carrera
Proactive Auto-scaling of Resources for Stream Processing Engines in the Cloud	2016	Tarek M. Ahmed and Farhana H. Zulkernine and James R. Cordy
Proactive Elasticity and Energy Awareness in Data Stream Processing	2017	Tiziano De Matteis and Gabriele Mencagli
Proactive Scaling of Distributed Stream Processing Work Flows Using Workload Modelling: Doctoral Symposium	2016	Thomas Cooper
Project Hoover: Auto-scaling Streaming Map-reduce Applications	2012	Rajalakshmi Ramesh and Liting Hu and Karsten Schwan
Prospects, Challenges and Latest Developments in Designing a Scalable Big Data Stream Computing System	2015	Dawei Sun and Chunxiao Liu and Dongfeng Ren
REEF: Retainable Evaluator Execution Framework	2015	Markus Weimer and Yingda Chen and Byung-Gon Chun and Tyson Condie and Carlo Curino and Chris Douglas and Yunseong Lee and Tony Majestro and Dahlia Malkhi and Sergiy Matusevych and Brandon Myers and Shravan Narayanamurthy and Raghu Ramakrishnan and Sriram Rao and Russel Sears and Beysim Sezgin and Julia Wang
Resa: Realtime Elastic Streaming Analytics in the Cloud	2013	Tian Tan and Richard T.B. Ma and Marianne Winslett and Yin Yang and Yong Yu and Zhenjie Zhang
Rethinking Elastic Online Scheduling of Big Data Streaming Applications over High-velocity Continuous Data Streams	2018	Dawei Sun and Hongbin Yan and Shang Gao and Xunyun Liu and Rajkumar Buyya
Revenue-Based Resource Management on Shared Clouds for Heterogenous Bursty Data Streams	2012	Rafael Tolosana-Calasanz, José Ángel Bañares, Congduc Pham, Omer F. Rana
Runtime-aware Adaptive Scheduling in Stream Processing	2016	Yuan Liu and Xuanhua Shi and Hai Jin

title	year	authors
Scalable Distributed Stream Join Processing	2015	Qian Lin and Beng Chin Ooi and Zhengkui Wang and Cui Yu
Scheduling Data Stream Jobs on Distributed Systems with Background Load	2017	Anca Vulpe and Marc Frincu
SCIMITAR: Scalable Stream-Processing for Sensor Information Brokering	2013	K. Rohloff; J. Cleveland; J. Loyall; T. Blocher
SEEP: Scalable and Elastic Event Processing	2010	Matteo Migliavacca and David Eyers and Jean Bacon and Yiannis Papagiannis and Brian Shand and Peter Pietzuch
Self-adaptive Processing Graph with Operator Fission for Elastic Stream Processing	2017	Nicolas Hidalgo and Daniel Wladdimiro and Erika Rosas
Stormy: An Elastic and Highly Available Streaming Service in the Cloud	2012	Simon Loesing and Martin Hentschel and Tim Kraska and Donald Kossmann
StreamCloud: An Elastic and Scalable Data Streaming System	2012	Vincenzo Gulisano and Ricardo Jimenez-Peris and Marta Patino-Martinez and Claudio Soriente and Patrick Valduriez
Study and Implementation of Elastic Stream Computing In The Cloud	2015	Hanfeng Zhu and Gang Wu
Telemetry for Elastic Data (TED): Middleware for MapReduce Job Metering and Rating	2015	Soheil Qanbari and Ashkan Farivarmoheb and Parham Fazlali and Samira Mahdizadeh and Schahram Dustdar
The Active Streams Approach to Adaptive Distrubuted Systems	2001	F. Bustamante
Towards Hierarchical Autonomous Control for Elastic Data Stream Processing in the Fog	2018	Valeria Cardellini, Francesco Lo Presti, Matteo Nardelli, Gabriele Russo Russo
Using Application Data for SLA-Aware Auto-scaling in Cloud Environments	2015	Andre Abrantes D. P. Souza and Marco A. S. Netto
VISP: An Ecosystem for Elastic Data Stream Processing for the Internet of Things	2016	C. Hochreiner; M. Vogler; P. Waibel; S. Dustdar

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Datum und Unterschrift:

Declaration

I hereby declare that the work presented in this thesis is entirely my own. I did not use any other sources and references that the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

Date and Signature: