The Elastic Processing of Data Streams in Cloud Environments: A Systematic Mapping Study

Floriment Klinaku, Michael Zigldrum, Markus Frank, and Steffen Becker
Reliable Software Systems, University of Stuttgart, Germany

Keywords: Elastic, Processing, Data Streams, Cloud.

Abstract: Ongoing efforts exist to exploit cloud elasticity for processing efficiently data streams generated by a variety of data sources. To contribute to these efforts an overview of existing research is required. To the best of our knowledge, a systematic overview of the field is missing. To fill this gap, we conduct a Systematic Literature Map (SLM). This way we offer a high-level overview of the literature on elastic data stream processing. We search four databases, evaluate 564 publications and identify 100 relevant publications. The identified publications show that the majority of work is validated research through proofs-of-concept and very few through case studies, surveys and field experiments. There are several frameworks, approaches and tools proposed, but, fewer metrics, models and processes.

1 INTRODUCTION

Motivation. Billions of data sources (i.e., sensors, devices, smartphones or things[1]) are connected to the Internet. They continuously produce data of different structure at variable rates. The variability in arrival rates and structure places a challenge on software systems that are responsible for processing such streams in an efficient and timely manner. The rise of the cloud computing paradigm (Mell et al., 2011) made it possible for businesses and customers to acquire computing resources according to their needs. Moreover, there exist ongoing efforts on designing processes and solutions (e.g., auto-scalers) which autonomously handle the provisioning of resources (Chen et al., 2018) in the presence of changing conditions and demands.
Problem. One class of software systems that tend to exploit cloud elasticity are *stream processing systems*. These systems perform computations on *events* that continuously arrive. They are required to efficiently compute low-latency queries on non-homogeneous data while it arrives and in presence of variable arrival rates. In literature these systems are also referred as *stream computing*, *complex event processing (CEP)* or *data stream management system (DSMS)* (Hummer et al., 2013).

The importance of these systems will continue to increase based on predicted increase of sensors and IoT devices in the next years. Thus, it is important for researchers and engineers to have an overview of existing and missing work in state-of-the-art research for elastic and efficient processing of data streams. To the best of our knowledge an overview is missing.
Solution. To provide an overview of existing research on elastic stream processing we conduct a Systematic Literature Map (SLM). Through this work we make the following contributions: 1. We systematically present work of researchers which tackles aspects in exploiting cloud elasticity for data stream processing applications. 2. We classify publications based on two facets: *research* and *contribution* and identify areas which need more attention in the future. 3. We discuss publications in areas where more work is required and provide researchers high-level aspects of their work.
Results. A set of one hundred publications constitute the effort of researchers tackling different aspects of elasticity in the context of stream processing. There are gaps in evaluation and opinion type of research. With respect to contribution type there are several frameworks and approaches found in the literature but less work is done in metrics, models and discussions.
Structure. The rest of the work is structured as follow: Section 1 presents details on the method. Section 3 presents the results; Section 4 discusses threats to validity; Section 5 highlights related work; finally Section 6 concludes the work and highlights plans for the future.

[1]Internet of Things vocabulary

Klinaku, F., Zigldrum, M., Frank, M. and Becker, S.
The Elastic Processing of Data Streams in Cloud Environments: A Systematic Mapping Study.
DOI: 10.5220/0007708503160323
In *Proceedings of the 9th International Conference on Cloud Computing and Services Science (CLOSER 2019)*, pages 316-323
ISBN: 978-989-758-365-0

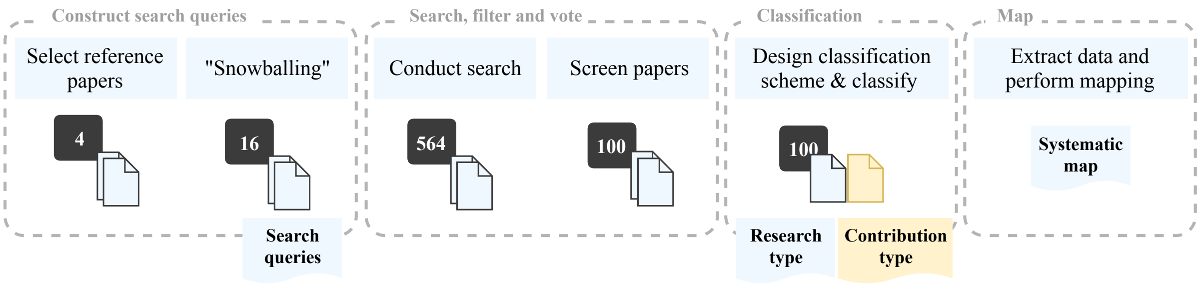

Figure 1: The overall study process.

2 RESEARCH PROCESS

To ensure reproducible and valid results we adhere to the process and guidelines described by Kuhrmann et al. (Kuhrmann et al., 2017) and by Petersen et al. (Petersen et al., 2008). In this section, we describe unique characteristics of the research process and avoid redundant guidance on how to conduct mapping studies. For details left out, readers should follow the process by Petersen et. al. To discover the state-of-the-art research and identify gaps in the literature we chose to quantify the body of work through a Systematic Literature Map (SLM). We define the research scope as follows:

Context. Processing data streams in cloud environments

Peculiarity. Elasticity

We design all the activities in the study process according to the *Context* and *Peculiarity* To determine how well the defined peculiarity is exploited in the defined context we pose two research questions:

RQ_1: How large is the body of work that tackles cloud elasticity related peculiarities in the field of data stream processing?

RQ_2: What type of research is required for the future?

Figure 1 depicts the overall step-by-step process which led to the results presented in this paper. From left to right: we start off with a set of **4 reference publications** (Cervino et al., 2012a; Gedik et al., 2014; Hochreiner et al., 2016; Abadi et al., 2005), which objectively belong to the publication space that is mapped. Through *snowballing* (Kuhrmann et al., 2017) we extended the initial set of relevant work to **16 publications**. We use this set of publications to obtain the search queries iteratively. The criteria for constructing successful search queries was to obtain the initial set of 16 publications when searching across the data sources. Since there were resources where the abstract of the work was not given we designed two queries: one for searching on titles and one for abstracts as shown in Table 1.

Table 1: Final search queries.

(a) Title search string
(elastic **or** auto?scale **or** borealis) **and** (stream **or** streaming) **and** (processing **or** system)
(b) Abstract search string
(elastic **or** auto?scale **or** borealis) **and** (data **or** cloud **or** event **or** query) **and** (stream) **and** (processing **or** computing **or** system)

We search in four prominent databases: ACM, IEEE, Springer, Science Direct. We decided to select these sources based on their focus on computer science and are also used in other mapping studies related to software engineering (e.g., (Ingibergsson et al., 2015)). Moreover, considering the search terms for our mapping study like ”elasticity” or ”streams”—which are very popular in other fields like physics or economics—we decide not to search in a meta-search database like Google Scholar. After searching the selected databases and filtering out duplicates and publications not relevant to our context (e.g., video-streaming), we obtain a set of **564 publications**. To further evaluate the found papers we follow the voting process proposed in (Kuhrmann et al., 2017), where reviewers vote independently and then in a workshop reviewers discuss publications which received different votes. Finally, after the voting process, there were **100 publications** left to be classified and mapped.

To get a better overview of the publication space, we classify the papers based on **research** and **contribution** type. For the first classification facet we follow the proposed scheme by Wieringa et al. (Wieringa et al., 2006). For the later—the contribution type—we created our own scheme. For the research type, there are seven classes, namely Evaluation, Solution Proposal, Validation, Philosophical, Opinion, Personal Experience, Review, and

Table 2: Research type classification.

Name	Abbr	Description - The work ...
		(a) Research type classification facet
Evaluation	EVAL	"results in new knowledge of causal relationships among phenomena. Causal properties are studied empirically, such as by case study, field study, field experiment, survey, etc."
Solution Proposal	SOL	"... proposes a solution ... and argues for its relevance, without a fullblown validation. A proof-of-concept may be offered by means of a small example, a sound argument, or by some other means." (Wieringa et al., 2006)
Validation	VAL	"... proposes a solution ... and argues for its relevance, without a fullblown validation. A proof-of-concept may be offered by means of a small example, a sound argument, or by some other means." (Wieringa et al., 2006)
Philosophical	PHIL	"... sketches a new way of looking at things, a new conceptual framework, etc." (Wieringa et al., 2006)
Opinion	OP	"... contain[s] the author's opinion about what is wrong or good about something, how we should do something, etc." (Wieringa et al., 2006)
Personal Experience	PERS	"... 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." (Wieringa et al., 2006)
Review and Summary	REV	"... where author/authors is/are reviewing or summarizing the evolution of an area of research in a historical fashion." (Wieringa et al., 2006)
		(b) Contribution type classification facet
Metric	MET	... contributes to forming comparable and meaningful metrics. The work presents a benchmark tool and provides metrics with which to compare.
Model	MOD	... provides models which aid to verify approaches, algorithms or other research.
Process	PROC	... describes a process or processes to reach a certain goal.
Discussion	DISC	... conducts a literature review or discusses and outlines challenges based on already existing research papers.
Approach/ Algorithm	APPR	... outlines a new approach to a known problem, or provides an algorithm which solves an existing problem.
Framework	FRMW	... provides a complete working framework which tackles several challenges. 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	TL	Contrary to a framework, work in this category contributes to a smaller part of a complete system or adds 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.

Summary. For the contribution type there are six claseses, namely Metric, Model, Process, Discussion, Approach/Algorithm, Framework, and Tool. Both are shown in Table 2 part (a) and (b).

3 RESULTS

In order to get an overview of research with respect to the utilization of cloud elasticity for efficient data stream processing we examine the number of papers that tackle this issue. We conduct a systematic mapping study to obtain the set of papers that contribute to the defined research context and peculiarity.

Generally speaking, the 100 publications (Table 3[2]) constitute a valuable information for researchers. Moreover, the data points highlight the lack of work in several types both for the contribution facet as well as the research facet. These gaps could constitute future research directions. The rest of this section reports on results and is divided in two subsections. The two subsections report on the data according to the designed research questions.

RQ1: How large is the body of work that tackles cloud elasticity related pecularities in the big data stream processing context? A set of one hundred papers constitutes the attempt of researchers to tackle different challenges in designing elastic and efficient big data stream processing applications. As seen in Figure 2 the first relevant publication came out in 2000. Then in 2012, the number of publications rose to a peak with 22 papers in 2015. After the peak, in the following two years, the number of publications fell. It is important to note that the data for the year 2018 is not complete, as we conduct the data search and export in August of 2018.

RQ2: What type of research is required for the future? As Figure 3 depicts, the biggest gaps are in evaluations and opinions. Evaluation work need to evaluate the research in practice whereas opinions are usually formed after investigation of multiple different approaches. Both research types have a low number of works. Also other research categories—

[2]Due to space reasons Table 3 contains selected publications, to see the full list of publications visit http://klinakuf.github.io/ms-elastic-datastreams

Table 3: Selected Publications.

RT	CT	Publications	Num
VAL	FRMW	(Gkolemis et al., 2017; Katsipoulakis et al., 2015; Wu and Tan, 2015; Gedik et al., 2018; Madsen and Zhou, 2013)	5/31
	APPR	(Cammert et al., 2008; Kombi et al., 2017; Isert and Schwan, 2000; Cervino et al., 2012b; Vasconcelos et al., 2016)	5/18
	TL	(Wu et al., 2017; Heinze et al., 2015; Zhang et al., 2013; Cardellini et al., 2018; Bellavista et al., 2013)	5/16
	MOD	(Mencagli, 2016; Heinze et al., 2014b; Nguyen et al., 2015; Lin et al., 2015; Qanbari et al., 2015)	5
SOL	FRMW	(Vakali et al., 2016; Zhang et al., 2015; Bhandari, 2012; Chen et al., 2014; Zacheilas et al., 2016)	5/8
	APPR	(Das et al., 2014; Vu et al., 2010; Heinze, 2011; Humayoo et al., 2014)	4
	DISC	(Ahmed et al., 2016; Reale et al., 2014; Martin et al., 2014)	3
	TL	(Martins et al., 2014; HoseinyFarahabady et al., 2017)	2
	MOD	(Imai et al., 2016; Vulpe and Frincu, 2017)	2
	PROC	(Heinze et al., 2014c)	1
	MET	(Vorona et al., 2014)	1
REV	DISC	(Heinze et al., 2014a; de Assuncao et al., 2018; Hummer et al., 2013)	3
PHIL	DISC	(Eyers et al., 2012; Sun et al., 2015)	2
	FRMW	(Hochreiner et al., 2015)	1
	APPR	(Bustamante et al., 2001)	1
OP	PROC	(Truong et al., 2016)	1
EVAL	FRMW	(Chun et al., 2013)	1

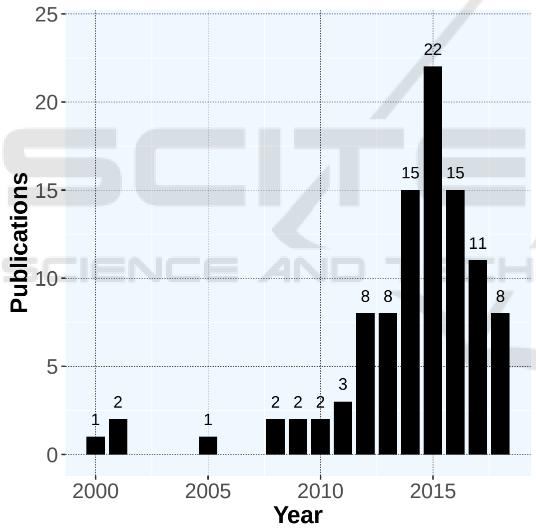

Figure 2: Number of publications per year.

excluding validation research—have a low number of publications. Especially summaries in which this study contributes to. With respect to contribution type it is evident that the majority of publications is about frameworks, approaches and tools and less about processes, models and metrics.

This study helps to capture state-of-the-art research in utilising elasticity for processing data streams in cloud environments. Even though the number of publications in the last years of the graph declined, the area can be considered an active area of research with 11 publications in 2017 and eight publications until August 2018. The curve showing the number of publications fits the Gartner curve that shows the adoption and maturity of an emerging technology (Linden and Fenn, 2003). The Gartner curve assumes that most technologies and concepts will progress through the pattern of overenthusiasm and disillusionment. One could judge the field as immature, passing the overenthusiastic phase in 2012 and entering the growth phase with a steady number of publications.

The study shows that the most significant gaps are in evaluations and opinions. The found gaps reinforce the judgment about the maturity of the field. Evaluation work needs to evaluate the research in practice whereas opinions are usually formed after investigating different approaches. Both research types have a low number of works. Also other research categories—besides validation research—have a small number of publications. Especially summaries in which this study contributes to. Regarding contribution type, it is evident that the majority of publications is about frameworks, approaches and tools, and less about processes, models, and metrics.

Processes. Concerning processes, (Truong et al., 2016) emphasize the difficulty in engineering IoT cloud platforms due to the lack of tools to test and evaluate complex designs. For elasticity, they split requirements in two parts: in analytical and control. For the first, it is essential for developers and other roles to have an end-to-end view on behavioral limits (*emergent behavior*) so to enable proper roles in the team to refine the software and improve control strategies. For the second—elasticity control—they define the granularity of control for different parts of the IoT system (e.g., the data reading frequency of sensors or

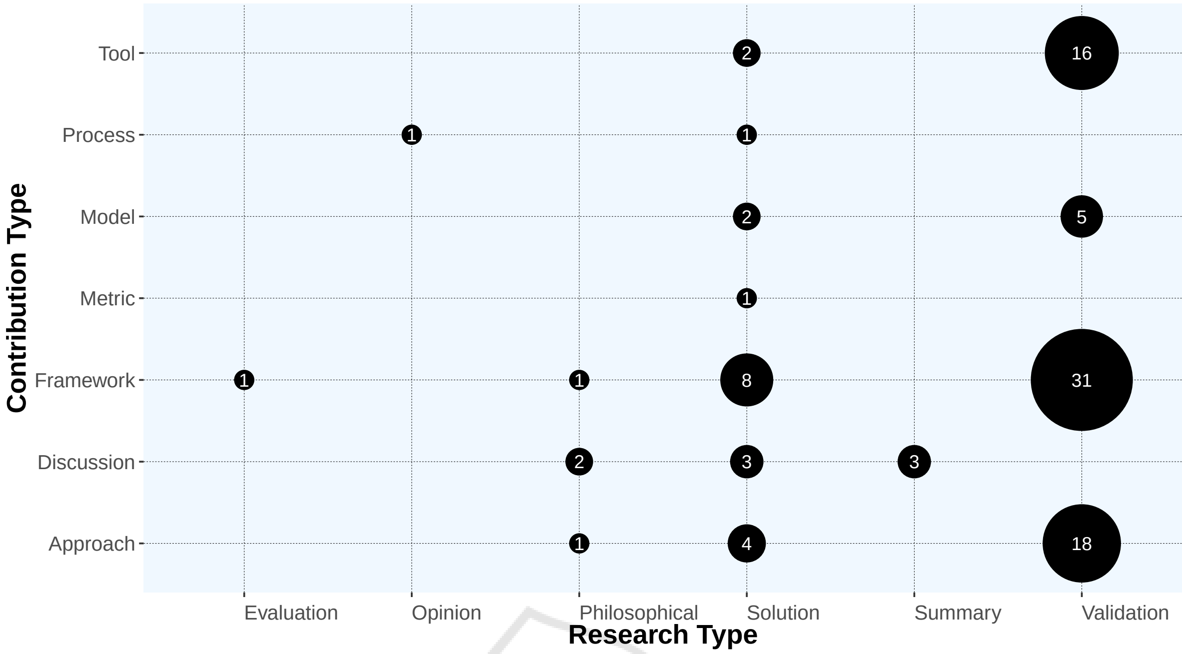

Figure 3: Publications mapped to research and contribution type.

adding/removing virtual machines).

Metrics and Benchmarks. (Vorona et al., 2014) contribute to the identified gap by benchmarking analytical platforms for answering complex business queries while leveraging the elastic infrastructure of the cloud. Moreover, they propose two metrics *scaling overhead* and *elasticity overhead*. The first represents the time wasted while the system is stabilizing after a provisioning action whereas the second—elasticity overhead—represents the time lost because of sub-optimal scaling decisions. It is of importance to notice that these elasticity metrics differ when compared to application agnostic elasticity metrics like mean time to quality repair (MTTQR) proposed in (Lehrig et al., 2015).

Models. Our study reveals various kind of modelling approaches to enable elasticity for data stream processing systems. These kinds include application-agnostic performance modelling solutions (Imai et al., 2016), game-theoretic approaches (Mencagli, 2016), models which estimates the latency spike created by a set of operator movements (Heinze et al., 2014b) used then to built latency aware elastic placement operator algorithm, models for data assets in the context of Data-as-a-Service (DaaS) (Nguyen et al., 2015) where providers could handle different quality requirements for results.

4 THREATS TO VALIDITY

In this section we describe threats to the following classes of validity as presented by (Claes et al., 2000): (1) conclusion validity, (2) internal validity and (3) construct validity.

Conclusion Validity. The choice of data sources constitutes one threat to the validity of the conclusion for the identified gaps. The study might potentially miss relevant publications that exist on other sources.

Internal Validity. The classification of publications is another threat to the internal validity of the study. The classification might be impacted by human error both from authors of this study when classifying known as judgmental error (Petersen et al., 2008), and authors of the relevant publications. Potential errors on both sides could lead to incorrect classification.

Construct Validity. The conclusion of the mapping study about uncovered contribution types, e.g., metrics or processes, is subject to the threat that classes of the contribution classification facet may subsume each other, e.g., a publication classified as an approach may also contribute with metrics for evaluating the approach. However, we argue that this does not impact the identified gaps since there is a low number of publications which have these classes as their primary contribution.

5 RELATED WORK

Our mapping study revealed other works which contribute to summarizing the problem area. In a recent survey, Assunção et al. (de Assuncao et al., 2018) discuss challenges, solutions and techniques for elastic and efficient data stream processing. Their study shows problems at a finer granularity whereas our research gives an overview at a higher level and identifies gaps for future research directions.

Heinze et al. (Heinze et al., 2014a) present open challenges for next generations of data stream processing systems. One aspect they foresee as important is *advanced elasticity* where they argue that current elasticity strategy optimize utilization of the system only and other metrics like latency or bandwidth are less considered.

In a similar direction goes also the work of Hummer et al. (Hummer et al., 2013) where they pinpoint one challenge which also our study reflects: how to define service-level objectives in the best way and how are metrics that indicate the elasticity of a data stream processing system related to parameters that reflect the quality guarantees for a tenant in a multi-tenant data stream as a service scenario.

6 CONCLUSION

In this paper, we present the results of a systematic literature map on elastic processing of data streams. We search 4 databases, evaluate 564 publications and identify 100 relevant publications which tackle different elasticity aspects of data stream processing. The study aids researchers in the field to get a high-level overview of contributions and forms a basis for a detailed study. Moreover, the visual map based on contribution and research type helps practitioners to identify work which is relevant for their specific problems in the area of elastic data stream processing.

For future work, we plan a more in-depth study of the publication set. We plan to extract information concerning contribution areas that show a low number of publications, e.g., metrics, models and processes. One possible way is to analyse publications that are contributing to frameworks or tools. From these publications, one could extract and compare information about metrics, models and processes they have used. This in-depth study would benefit the body of work with a better overview of the identified gaps.

REFERENCES

Abadi, D. J., Ahmad, Y., Balazinska, M., Çetintemel, U., Cherniack, M., Hwang, J.-H., Lindner, W., Maskey, A., Rasin, A., Ryvkina, E., Tatbul, N., Xing, Y., and Zdonik, S. B. (2005). The Design of the Borealis Stream Processing Engine. *Cidr*.

Ahmed, T. M., Zulkernine, F. H., and Cordy, J. R. (2016). Proactive auto-scaling of resources for stream processing engines in the cloud. In *Proc. of the 26th Annual Intl. Conf. on Comp. Science and Software Engineering*.

Bellavista, P., Corradi, A., Kotoulas, S., and Reale, A. (2013). Dynamic datacenter resource provisioning for high-performance distributed stream processing with adaptive fault-tolerance. In *Proc. Demo & Poster Track of ACM/IFIP/USENIX Intl. Middleware Conf.*

Bhandari, S. (2012). An adaptive event stream processing environment. In *Proc. of the on SIGMOD/PODS PhD Symp.*

Bustamante, F. E., Eisenhauer, G., and Schwan, K. (2001). The active streams approach to adaptive distributed systems. In *High Performance Distributed Computing. Proc. 10th IEEE Intl. Symp. on*.

Cammert, M., Kramer, J., Seeger, B., and Vaupel, S. (2008). A cost-based approach to adaptive resource management in data stream systems. *IEEE Transactions on Knowledge and Data Eng.*, 20(2).

Cardellini, V., Presti, F. L., Nardelli, M., and Russo, G. R. (2018). Decentralized self-adaptation for elastic data stream processing. *Future Generation Comp. Systems*.

Cervino, J., Kalyvianaki, E., Salvachua, J., and Pietzuch, P. (2012a). Adaptive provisioning of stream processing systems in the cloud. In *IEEE 28th Intl. Conf. on Data Engineering Ws.*

Cervino, J., Kalyvianaki, E., Salvachua, J., and Pietzuch, P. (2012b). Adaptive provisioning of stream processing systems in the cloud. In *Data Engineering Ws.s (ICDEW), IEEE 28th Intl. Conf. on*.

Chen, T., Bahsoon, R., and Yao, X. (2018). A survey and taxonomy of self-aware and self-adaptive cloud autoscaling systems. *ACM Comput. Surv.*

Chen, W.-s., Chen, Y.-J., and Wu, S.-y. (2014). Dynamic aggregate: An elastic framework for qos-aware distributed processing of rfid data on enterprise hierarchy. *IEEE Transactions on Parallel and Distributed Systems*.

Chun, B.-G., Condie, T., Curino, C., Douglas, C., Matusevych, S., Myers, B., Narayanamurthy, S., Ramakrishnan, R., Rao, S., Rosen, J., et al. (2013). Reef: Retainable evaluator execution framework. *Proc. of the VLDB Endowment*.

Claes, W., Per, R., Martin, H., Magnus, C., Björn, R., and Wesslén, A. (2000). Experimentation in software engineering: an introduction. *Online Available: http://books. google. com/books*.

Das, T., Zhong, Y., Stoica, I., and Shenker, S. (2014). Adaptive stream processing using dynamic batch sizing. In *Proc. of the ACM Symp. on Cloud Computing*.

de Assuncao, M. D., da Silva Veith, A., and Buyya, R. (2018). Distributed data stream processing and edge computing: A survey on resource elasticity and future directions. *Journal of Network and Comp. Applications*.

Eyers, D., Freudenreich, T., Margara, A., Frischbier, S., Pietzuch, P., and Eugster, P. (2012). Living in the present: On-the-fly information processing in scalable web architectures. In *Proc. of the 2Nd Intl. Ws. on Cloud Computing Platforms*.

Gedik, B. et al. (2018). C-stream: A co-routine-based elastic stream processing engine. *ACM Transactions on Parallel Computing*.

Gedik, B., Schneider, S., Hirzel, M., and Wu, K. L. (2014). Elastic scaling for data stream processing. *IEEE Transactions on Parallel and Distributed Systems*.

Gkolemis, E., Doka, K., and Koziris, N. (2017). Automatic scaling of resources in a storm topology. In *Intl. Ws. on Algorithmic Aspects of Cloud Computing*.

Heinze, T. (2011). Elastic complex event processing. In *Proc. of the 8th Middleware Doctoral Symp.*

Heinze, T., Aniello, L., Querzoni, L., and Jerzak, Z. (2014a). Cloud-based data stream processing. In *Proc. of the 8th ACM Intl. Conf. on Distributed Event-Based Systems*.

Heinze, T., Jerzak, Z., Hackenbroich, G., and Fetzer, C. (2014b). Latency-aware elastic scaling for distributed data stream processing systems. In *Proc. of the 8th ACM Intl. Conf. on Distributed Event-Based Systems*.

Heinze, T., Pappalardo, V., Jerzak, Z., and Fetzer, C. (2014c). Auto-scaling techniques for elastic data stream processing. In *Data Engineering Ws.s, IEEE 30th Intl. Conf. on*.

Heinze, T., Zia, M., Krahn, R., Jerzak, Z., and Fetzer, C. (2015). An adaptive replication scheme for elastic data stream processing systems. In *Proc. of the 9th ACM Intl. Conf. on Distributed Event-Based Systems*.

Hochreiner, C., Schulte, S., Dustdar, S., and Lecue, F. (2015). Elastic stream processing for distributed environments. *IEEE Internet Computing*.

Hochreiner, C., Vogler, M., Schulte, S., and Dustdar, S. (2016). Elastic Stream Processing for the Internet of Things. In *IEEE 9th Intl. Conf. on Cloud Computing (CLOUD)*.

HoseinyFarahabady, M., Lee, Y. C., Zomaya, A. Y., and Tari, Z. (2017). A qos-aware resource allocation controller for function as a service (faas) platform. In *Intl. Conf. on Service-Oriented Computing*.

Humayoo, M., Zhai, Y., He, Y., Xu, B., and Wang, C. (2014). Operator scale out using time utility function in big data stream processing. In *Intl. Conf. on Wireless Algorithms, Systems, and Applications*.

Hummer, W., Satzger, B., and Dustdar, S. (2013). Elastic stream processing in the cloud. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.

Imai, S., Patterson, S., and Varela, C. A. (2016). Cost-efficient elastic stream processing using application-agnostic performance prediction. In *16th IEEE/ACM Intl. Symp. on Cluster, Cloud and Grid Computing*.

Ingibergsson, J. T. M., Schultz, U. P., and Kuhrmann, M. (2015). On the use of safety certification practices in autonomous field robot software development: A systematic mapping study. In *Intl. Conf. on Product-Focused Software Process Improvement*.

Isert, C. and Schwan, K. (2000). Acds: Adapting computational data streams for high performance. In *Parallel and Distributed Processing Symp. IPDPS. Proc. 14th Intl.*

Katsipoulakis, N. R., Thoma, C., Gratta, E. A., Labrinidis, A., Lee, A. J., and Chrysanthis, P. K. (2015). Ce-storm: Confidential elastic processing of data streams. In *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*.

Kombi, R. K., Lumineau, N., and Lamarre, P. (2017). A preventive auto-parallelization approach for elastic stream processing. In *Distributed Computing Systems, IEEE 37th Intl. Conf. on*.

Kuhrmann, M., Fernández, D. M., and Daneva, M. (2017). On the pragmatic design of literature studies in software engineering: an experience-based guideline. *Empirical software engineering*, 22(6).

Lehrig, S., Eikerling, H., and Becker, S. (2015). Scalability, elasticity, and efficiency in cloud computing: A systematic literature review of definitions and metrics. In *Proc. of the 11th Intl. ACM SIGSOFT Conf. on Quality of Software Architectures*.

Lin, Q., Ooi, B. C., Wang, Z., and Yu, C. (2015). Scalable distributed stream join processing. In *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, SIGMOD '15.

Linden, A. and Fenn, J. (2003). Understanding gartner's hype cycles. *Strategic Analysis Report Nº R-20-1971. Gartner, Inc*.

Madsen, K. G. S. and Zhou, Y. (2013). Elastic mapreduce-style processing of fast data. In *Proc. of the 7th ACM Intl. Conf. on Distributed event-based systems*.

Martin, A., Silva, R., Brito, A., and Fetzer, C. (2014). Low cost energy forecasting for smart grids using stream mine 3g and amazon ec2. In *Proc. of the IEEE/ACM 7th Intl. Conf. on Utility and Cloud Computing*.

Martins, P., Abbasi, M., and Furtado, P. (2014). Audy: Automatic dynamic least-weight balancing for stream workloads scalability. In *IEEE Intl. Congress on Big Data*.

Mell, P., Grance, T., et al. (2011). The nist definition of cloud computing.

Mencagli, G. (2016). A game-theoretic approach for elastic distributed data stream processing. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*.

Nguyen, T.-D., Truong, H.-L., Copil, G., Le, D.-H., Moldovan, D., and Dustdar, S. (2015). On developing and operating of data elasticity management process. In *Intl. Conf. on Service-Oriented Computing*.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *EASE*, volume 8.

Qanbari, S., Farivarmoheb, A., Fazlali, P., Mahdizadeh, S., and Dustdar, S. (2015). Telemetry for elastic data (ted): Middleware for mapreduce job metering and rating. In *IEEE Trustcom/BigDataSE/ISPA*.

Reale, A., Bellavista, P., Corradi, A., and Milano, M. (2014). Evaluating cp techniques to plan dynamic resource provisioning in distributed stream processing.

In *Intl. Conf. on AI and OR Techniques in Constriant Programming for Combinatorial Optimization Problems*.

Sun, D., Liu, C., and Ren, D. (2015). Prospects, challenges and latest developments in designing a scalable big data stream computing system. *Intl. Journal of Wireless and Mobile Computing*.

Truong, H.-L., Copil, G., Dustdar, S., Le, D.-H., Moldovan, D., and Nastic, S. (2016). On engineering analytics for elastic iot cloud platforms. In *Intl. Conf. on Service-Oriented Computing*.

Vakali, A., Korosoglou, P., and Daoglou, P. (2016). A multi-layer software architecture framework for adaptive real-time analytics. In *IEEE Intl. Conf. on Big Data*.

Vasconcelos, R. O., Vasconcelos, I., and Endler, M. (2016). Dynamic and coordinated software reconfiguration in distributed data stream systems. *Journal of Internet Services and Applications*.

Vorona, D., Funke, F., Kemper, A., and Neumann, T. (2014). Benchmarking elastic query processing on big data. In *Ws. on Big Data Bench.*

Vu, D., Kalogeraki, V., and Drougas, Y. (2010). Efficient stream processing in the cloud. In *Intl. Conf. on Heterogeneous Networking for Quality, Reliability, Security and Robustness*.

Vulpe, A. and Frincu, M. (2017). Scheduling data stream jobs on distributed systems with background load. In *17th IEEE/ACM Intl. Symp. on Cluster, Cloud and Grid Computing (CCGRID)*.

Wieringa, R., Maiden, N., Mead, N., and Rolland, C. (2006). Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. *Requirements Engineering*, 11(1).

Wu, R., Liu, B., Chen, Y., Blasch, E., Ling, H., and Chen, G. (2017). A container-based elastic cloud architecture for pseudo real-time exploitation of wide area motion imagery (wami) stream. *Journal of Signal Processing Systems*.

Wu, Y. and Tan, K.-L. (2015). Chronostream: Elastic stateful stream computation in the cloud. In *IEEE 31st Intl. Conf. on Data Eng.*

Zacheilas, N., Zygouras, N., Panagiotou, N., Kalogeraki, V., and Gunopulos, D. (2016). Dynamic load balancing techniques for distributed complex event processing systems. In *Distributed Applications and Interoperable Systems*.

Zhang, F., Cao, J., Khan, S. U., Li, K., and Hwang, K. (2015). A task-level adaptive mapreduce framework for real-time streaming data in healthcare applications. *Future Generation Comp. Systems*.

Zhang, Z., Shu, H., Chong, Z., Lu, H., and Yang, Y. (2013). C-cube: Elastic continuous clustering in the cloud. In *IEEE 29th Intl. Conf. on Data Engineering*.