

Continuous Architecting of Stream-Based Systems

Marcello Bersani, Francesco Marconi and Damian A. Tamburri
Politecnico di Milano
Milan, Italy

Email: [marcellomaria.bersani, francesco.marconi,
damianandrew.tamburri]@polimi.it

Pooyan Jamshidi, Andrea Nodari
Imperial College London
London, UK

Email: [p.jamshidi,
a.nodari15]@imperial.ac.uk

Abstract—Big data architectures have been gaining momentum in the last few years. For example, Twitter uses complex Stream topologies featuring frameworks like Storm to analyse and learn trending topics from billions of tweets per minute. However, maintaining the quality of said topologies often requires iterative experiments on expensive multi-node clusters. As an aid to designers and developers evaluating their Stream topologies in a DevOps fashion (i.e., through continuous feedback between development and operation), we developed OSTIA, that is, “On-the-fly Static Topology Inference Analysis”. OSTIA allows reverse-engineering of streaming topologies for the purpose of: (a) using previously existing verification&validation techniques on elicited models; (b) visualising and refactoring elicited models while maintaining constraints that would only be evaluated at deployment and run-time; (c) tackle the occurrence of common anti-patterns across considered topologies. We illustrate the uses and benefits of OSTIA on three real-life industrial case studies.

I. INTRODUCTION

Big data applications process vast amounts of data [1] for the purpose of gaining key business intelligence through complex analytics such as machine-learning [2]. Said applications are receiving increased attention in the last few years given their ability to yield competitive advantage by direct investigation of user needs and trends hidden in the enormous quantities of data produced daily by the average internet user. Gartner predicts¹ that said business intelligence and analytics applications will remain top focus for CIOs until at least 2017-2018. However, there are many costs and complexities behind harnessing said applications ranging from very high infrastructure costs to steep learning curves for the many frameworks involved in designing and developing applications for Big data, e.g., Apache Storm², Apache Spark³ or Hadoop⁴ to name a few.

In our own experience with designing and developing for Big Data, we observed that a key complexity lies in quickly and continuously evaluating the effectiveness of big-data architectures, or better, *topologies* - in the Big data jargon. Effectiveness in Big data terms means being able to design, deploy, operate, refactor and then (re-)deploy architectures/topologies continuously and consistently with runtime restrictions of necessary Big-data frameworks involved (e.g., Spark, Storm,

etc.). A very trivial example of said restrictions consists in the need for the topology to represent a Directed-Acyclic-Graph (DAG), a key constraint of the Storm and Spark frameworks. We argue that such effectiveness can be maintained starting from design time, by enacting a continuous architecting of big-data topologies consistently with a DevOps organisational structure [3], [4]. Supporting this effectiveness at design time eases the (re-)deployability of Big data topologies and save the time and effort to run trial-and-error experiments on expensive infrastructure.

To sustain this argument, we developed OSTIA, that stands for: “On-the-fly Static Topology Inference Analysis”. OSTIA allows designers and developers to infer the application topology through on-the-fly reverse-engineering and architecture recovery [5]. During this inference step, OSTIA analyses the operating topology to make sure it is consistent with restrictions and constraints of the Big data frameworks at hand. Also, in an effort to tackle said complexities and offer support in a DevOps fashion, OSTIA was engineered to act as a mechanism that closes the feedback loop between operating Streaming topologies (-Ops phase) and their development and improvement phase (Dev- phase).

Currently, OSTIA focuses on Storm, i.e., one of the most famous and established real-time stream processing Big data engine [6]. OSTIA hardcodes intimate knowledge on the streaming development framework (Storm, in our case) and its dependence structure in the form of a meta-model [7]. This knowledge is necessary to make sure that elicited topologies are correct, so that models may be used in at least four scenarios: (a) realising an exportable visual representation for said topologies; (b) checking said topologies against framework restrictions that would only become evident during infrastructure setup or runtime operation; (c) checking that said topologies do not show any anti-patterns [8] that may lower performance and limit deployability/executability; (d) finally, use said topologies for further analysis, e.g., through model verification [9].

This paper outlines OSTIA, elaborating its major usage scenarios and its benefits while discussing and addressing its limitations. Also, we evaluate OSTIA using industrial case-study research featuring an open-source social-sensing application to show that OSTIA yields valuable design and development insights using an inference analysis of the three static topologies behind said industrial application. Finally, we

¹<http://www.gartner.com/newsroom/id/2637615>

²<http://storm.apache.org/>

³<http://spark.apache.org/>

⁴<https://hadoop.apache.org/>

elaborate on how OSTIA helps the continuous architecting of recovered topologies by means of valuable on-the-fly analyses as well as ancillary verification techniques. We conclude that OSTIA does in fact provide valuable insights for software architects to continuously architect and increase quality of their Big-data design and development featuring Storm so that expedite (re-)deployment can take place.

The rest of the paper is structured as follows. Section II-A outlines our research problem, research questions and our approach at tackling them. Sections III, IV and V outline OSTIA, discussing the usage scenarios and (anti-)patterns it was designed to support while outlining the main benefits we perceived in using it. Section VI elaborates further on the benefits by providing an actual evaluation of OSTIA using three cases from industrial practice. Section VII-B discusses the results and evaluation, also outlining OSTIA limitations and potential threats to its validity. Finally, Sections VIII and IX outline related work and conclude the paper.

II. RESEARCH DESIGN

The work we elaborated in this paper is stemming from the following research question:

“Can we assist the continuous architecting of Big-Data streaming designs?”

The results contained in this paper in response to this research question were initially elaborated within a free-form focus group [10] involving three experienced practitioners and researchers on Big data streaming technology such as Storm. Following the focus group, through self-ethnography [11] and brainstorming we identified the series of essential consistency checks, algorithmic evaluations as well as anti-patterns that can now be applied through OSTIA while recovering an architectural representation for Storm topologies. We designed OSTIA to support the incremental and iterative refinement of streaming topologies based on the incremental discovery and correction of the above checks and patterns.

In addition, the OSTIA prototype we developed⁵, was refined incrementally and evaluated through scenario-analysis and case-study research [12] involving an industrial partner.

A. Evaluating OSTIA

As previously stated, OSTIA was evaluated by means of an industrial case-study offered by one of the industrial partners in the DICE EU H2020 Project consortium⁶. The partner in question uses open-source social-sensing software to elaborate a subscription-based Big-Data application that: (a) aggregates news assets from various sources (e.g., Twitter, Facebook, etc.) based on user-desired specifications (e.g., topic, sentiment, etc.); (b) presents and allows the manipulation of data. Said application is based on the SocialSensor App⁷ which features the combined action of three complex streaming topologies

based on Apache Storm (see Fig. 1 for a sample of said topologies).

More in particular, the topology in Fig. 1 extracts data from sources and manipulates said data to divide and arrange contents based on type (e.g., article vs. media), later updating a series of databases (e.g., Redis) with these elaborations.

The models that OSTIA elicited from this application were showcased to our industrial partner in a focus group. Our qualitative analysis was based on questionnaires and open discussion.

Finally, we applied well-established verification approaches to integrate the value and benefits behind using OSTIA. More in particular, we engineered OSTIA to support exporting of elicited topologies for their further analysis using the Zot LTL model-checker [13] using the approach outlined in our previous work [9].

III. RESEARCH SOLUTION

@Pooyan,Andrea: here we should probably elaborate on OSTIA’s architecture and the design principles that led us to define it as such... also we might want to elaborate on its components, the structure I’m suggesting below is merely tentative but it will give us ahead start!!

- add and comment the meta-model of storm and how OSTIA uses that as a reference to draw and check models which are consistent with the technology
- OSTIA Architecture
- we should probably elaborate the architecture part (or on a separate “implementation” part or paragraph) with a link to the downloadable technology - @Andrea: can we bundle it up as plugin for Eclipse? E.g., somehow using RCP?
- OSTIA Antipatterns Module
- OSTIA Visualisation Module
- OSTIA extensibility
- OSTIA explanation of use and simple usage scenario of continuous architecting

IV. TOPOLOGY DESIGN ANTI-PATTERNS

This section and section V elaborate on the Anti-patterns and algorithmic analysis supported in OSTIA. All figures in these sections use a simple graph-like notation where nodes may be any topological element (e.g., Spouts or Bolts in Apache Storm terms) while edges are to be interpreted as directed data-flow connections.

This elaborates on the anti-patterns we elicited through self-ethnography. These anti-patterns are elaborated further within OSTIA to allow for their detection during streaming topology inference analysis.

A. Multi-Anchoring

In order to guarantee fault-tolerant stream processing, tuples processed by bolts needs to be anchored with the unique id of the bolt and be passed to multiple acknowledgers (or “ackers” in short) in the topology. In this way, ackers can keep track of tuples in the topology.

⁵<https://github.com/maelstromdat/OSTIA>

⁶<http://www.dice-h2020.eu/>

⁷<https://github.com/socialsensor>

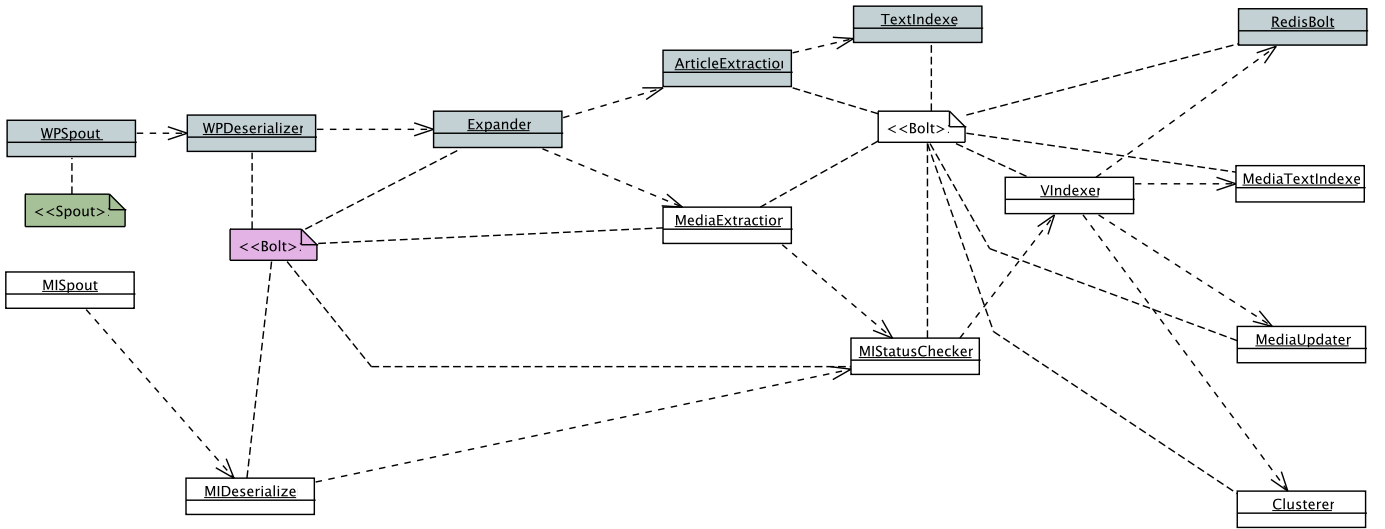


Fig. 1. A sample Storm topology in the SocialSensor App.

**TODO: what is the consequence of these anti-patterns?
How does OSTIA detect?**

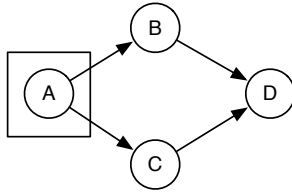


Fig. 2. Multi-anchoring.

B. Cycle-in Topology

Technically it is possible to have cycle in Storm topologies. An infinite cycle of processing would create an infinite tuple tree and make it impossible for Storm to ever acknowledge spout emitted tuples. Therefore, cycles should be avoided or resulting tuple trees should be investigated additionally to make sure they terminate at some point and under a specified series of conditions.

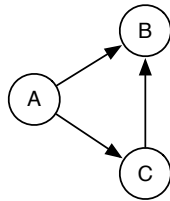


Fig. 3. Cycle-in Topology.

C. Persistent Data

This pattern covers the circumstance wherefore if two processing elements need to update a same entity in a storage, there should be a consistency mechanism in place.

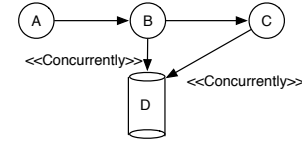


Fig. 4. Concurrency management.

V. ALGORITHMIC ANALYSIS ON STREAM TOPOLOGIES

A. fan-in/fan-out

For Bolts, fan-in is the number of processing elements (Bolts and Spouts) that are connected to this specific bolt in the topology and fan-out is the number of PEs that are dependent to this specific bolt. For Spouts, fan-in is the number of topics (from the messaging broker, i.e., Apache Kafka for Storm) that this specific spout is subscribed to and fan out is the number of bolts that is dependent to this spout.

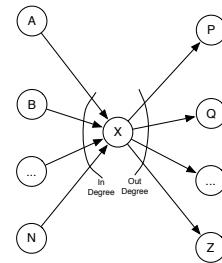


Fig. 5. Fan-in fan-out in Stream topologies.

B. topology cascading

By topology cascading, we mean connecting two different Storm topologies via a messaging framework (e.g., Apache Kafka). This circumstance, which is actually part of our evaluation and case-studies, may raise the complexity of the

overarching topology to unacceptable levels and may require additional attention.

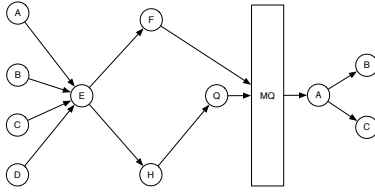


Fig. 6. cascading.

C. Topology clustering

Identifying the coupled processing elements and put the in a cluster in a way that elements in a cluster have high cohesion and less coupled with the elements in other clusters. Simple clustering or Social-Network Analysis mechanisms can be used to infer clusters. These clusters may require additional attention since they could turn out to become bottlenecks.

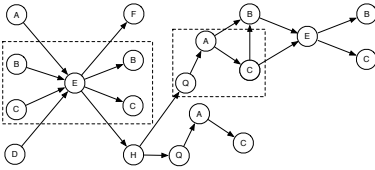


Fig. 7. clustering.

D. Computation funnel

A computational funnel emerges when there is not a path from data source (spout) to the bolts that sends out the tuples off the topology to another topology through a messaging framework or through a storage. This circumstance should be dealt with since it may compromise the availability of results under the desired performance restrictions.

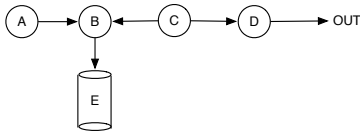


Fig. 8. computation funnel.

E. Linearising a topology

Sorting the processing elements in a topology in a way that topology looks more linear, visually. This step ensures that visual investigation and evaluation of the structural complexity of the topology is possible by direct observation. It is sometimes essential to provide such a visualisation to evaluate how to refactor the topology based on emerging needs.

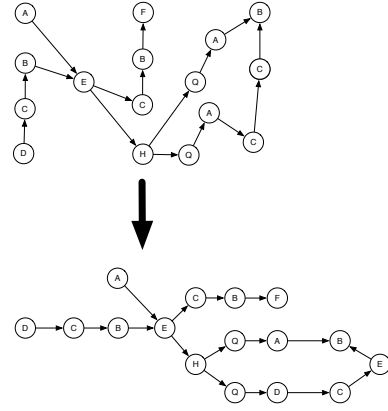


Fig. 9. linearizing.

VI. EVALUATION

@Marcello,Francesco: we should also probably elaborate on the kind of verification technique we are using and how that can help in evaluating the topology.. remember here we do not have the DICE restriction so we can mention any kind of analysis that it would be possible to run, also analyses that are currently in the hands of other DICE partners!!

- we can use the ATC case study as much as we want - that yields already three topologies that we can infer
- ATC has agreed that we can mention their role in this exercise, I also showed them the topology that we elicited basically with OSTIA and they already made considerations on how to improve it
- in the evaluation we should also comment on how OSTIA can help you in visualizing the application topology that you may be considering to use by reusing a big-data application for something else... visualising the application topology and analysing it may allow you to improve it while you are using it as a starting point for your application
- another application that we can use is the one that NETF is considering for their own scenario, KILLRWEATHER - <https://github.com/killrweather/killrweather>
- any additional case that we can run?
- what do the results show? do we have a way to quickly quantify the time that is saved by using this approach? e.g., the time that is saved in setting up and running the infrastructure and how much would that time saved have costed these could be valuable evaluation insights

VII. DISCUSSION

This section discusses our findings, their value in improving the quality of big-data architectures as well as the insights we discovered to plan further improvements of our technology. Finally, this section discusses the limitations behind using OSTIA.

A. Findings and Quality-Improvement Insights

OSTIA represents one humble but significant step at supporting practically the necessities behind developing and maintaining high-quality big-data applications. In designing and developing OSTIA we encountered a number of insights that may lead to improving said quality by means of OSTIA.

First, we found that it is often more useful to develop a quick-and-dirty but “runnable” topology than improving the topology at design time for a tentatively perfect execution. This way of working is consistent with continuous improvement approaches typical in DevOps scenarios. Also, this is mostly the case with big-data applications that are developed stemming from previously existing topologies. OSTIA hardcodes this way of thinking by supporting reverse-engineering and recovery of deployed topologies for their incremental improvement. Although we did not carry out extensive qualitative or quantitative evaluation of OSTIA in this regard, we are planning additional industrial experiments for future work with the goal of increasing OSTIA usability and practical quality.

Second, big-data applications design is an extremely young and emerging field for which not many software design patterns have been discovered yet. The (anti-)patterns and approaches currently hardcoded into OSTIA are inherited from related fields, e.g., pattern- and cluster-based graph analysis. Nevertheless, OSTIA may also be used to investigate the existence of recurrent and effective design solutions (i.e., design patterns) for the benefit of big-data application design. We are planning to improve OSTIA in this regard by experimenting on two fronts: (a) re-design and extend the facilities with which OSTIA supports anti-pattern detection; (b) run OSTIA on multiple big-data applications stemming from multiple technologies beyond Storm (e.g., Spark, Hadoop Map Reduce, etc.) with the purpose of finding recurrent patterns.

Third, finally, a step which is currently unsupported during big-data applications design is devising an efficient algorithmic breakdown of a workflow into an efficient topology. However, OSTIA does support the linearisation and combination of multiple topologies, e.g., into a cascade. Cascading and similar super-structures may be an interesting investigation venue since they may reveal more efficient styles for big-data architectures beyond styles such as Lambda Architecture [14]. OSTIA may aid in this investigation by allowing the interactive and incremental improvement of multiple (combinations of) topologies together.

B. Approach Limitations and Threats to Validity

Although OSTIA shows promise both conceptually and as a practical tool, it shows several limitations by design in its current form.

First of all, OSTIA only supports streaming topologies enacted using Storm. Multiple other big-data frameworks exist, however, to support both streaming and batch processing.

Second, OSTIA only allows to recover and evaluate previously-existing topologies, its usage is limited to design improvement and refactoring phases rather than design. Although this limitation may inhibit practitioners from using

our technology, the (anti-)patterns and algorithmic approaches elaborated in this paper help designers and implementors to develop the reasonably good-quality and “quick” topologies upon which to use OSTIA for continuous improvement.

Third, OSTIA does offer essential insights to aid deployment as well (e.g., separating or *clustering* complex portions of a topology so that they may run on dedicated infrastructure), however, our tool was not meant to be used as a system to aid planning and infrastructure design. Rather, as specified previously in the introduction, OSTIA was meant to evaluate and increase the quality of topologies *before* they enter into operation since the continuous improvement cycles connected to operating the topology and learning from said operation are often costly and still greatly inefficient.

VIII. PREVIOUS AND RELATED WORK

The work behind OSTIA stems from the EU H2020 Project called DICE⁸ where we are investigating the use of model-driven facilities to support the design and quality enhancement of Big-Data applications. In the context of DICE, much previous work has been done to support the design, development and deployment of Big-Data applications. For example, we have been developing a series of ad-hoc technological specifications, i.e., meta-model packages that contain all concepts, constructs and constraints needed to develop and operate Big-Data applications for the frameworks coded into the technological specifications. These specifications can be used, for example, to instantiate Big-Data components following standard Model-Driven procedures, without the need to learn said frameworks at all. OSTIA uses insights gained in developing and using said frameworks to apply consistency checks in the context of recovering Big-Data architectures, specifically, for Storm. Much similarly to the DICE effort, the IBM Stream Processing Language (SPL) initiative [15] provides an implementation language specific to programming streams management (e.g., Storm jobs) and related reactive systems based on the Big-Data paradigm. Although SPL is specific to WebSphere and IBM technology, its attempt at providing a relatively abstract language to implement for streams management and processing is remarkably related to OSTIA, since one of its aims is to improve quality of streams management by direct codification of higher order concepts such as streams declarations.

In addition, there are several work close to OSTIA in terms of their foundations and type of support.

First, from a quantitative perspective, much literature discusses quality analyses of Storm topologies, e.g., from a performance [16] or reliability point of view [17]. Said works use complex math-based approaches to evaluating a number of Big data architectures, their structure and general configuration. However, although novel, these approaches do not suggest any significant design improvement method or pattern to make the improvements *deployable*. With OSTIA, we make available a tool that automatically elicits a Storm

⁸<http://www.dice-h2020.eu/>

topology and, while doing so, analyses said topology to evaluate it against a number of consistency checks that make the topology consistent with the framework it was developed for (Storm, in our case). As previously introduced, a very trivial example of said checks consists in evaluating whether the topology is indeed a Directed-Acyclic-Graph (DAG), as per constraints of the Storm framework. To the best of our knowledge, no such tool exists to date.

Second, in our previous work, we proposed BO4CO [18], an approach for locating optimal configurations using ideas of carefully choosing where to sample by sequentially reducing uncertainty in the response surface approximation in order to reduce the number of performance measurements. We have carried out extensive experiments with three different stream topologies running on Apache Storm. Experimental results demonstrate that BO4CO outperforms the baselines in terms of distance to the optimum performance with at least an order of magnitude.

Third, from a modelling perspective, approaches such as StormGen [19] offer means to develop Storm topologies in a model-driven fashion using a combination of generative techniques based on XText and heavyweight (meta-)modelling, based on EMF, the standard Eclipse Modelling Framework Format. Although the first of its kind, StormGen merely allows the specification of a Storm topology, without applying any consistency checks or without offering the possibility to *recover* said topology once it has been developed. By means of OSTIA, designers and developers can work hand in hand while refining their Storm topologies, e.g., as a consequence of verification or failed checks through OSTIA. As a consequence, tools such as StormGen can be used to assist the preliminary development of quick-and-dirty Storm topologies, e.g., to be further processed and improved with the aid of OSTIA.

Fourth, from a verification perspective, ...

@Marcello,Francesco: here we should probably elaborate on what kind of verification approach we are using and what other verifications may be done, e.g., using some related work at this point... e.g., is there any other verification attempt considering JSON as an interchange format? I would discuss these and compare them to OSTIA as a whole

Finally, several deployment modelling technologies may be related to OSTIA since their role is to model the deployment structure represented by Big data architectures so that it can actually be deployed using compliant orchestrators. One such example is Celar⁹, a deployment modelling technology based on the TOSCA OASIS Standard¹⁰. Celar and related technologies (e.g., Alien4Cloud¹¹) may be used in combination with OSTIA since their role is that of representing the infrastructure needed by modelled (Big data) applications so that they can

be deployed. This representation is realised by means of infrastructure blueprints to be run by compliant orchestrators. The role of OSTIA in this scenario, is that of helping the quality refinement of an application topology to represent the very infrastructure needed for its run-time environment.

IX. CONCLUSION

Big data applications are rapidly gaining interest and momentum by small to big players on the market, even beyond the IT corner. Applications that make heavy use of Big data application frameworks require intensive reasoning and continuous architecting of design aspects typically around the topology of the basic operations to be applied in manipulating the data.

We set out to answer the following research question: “Can we assist the continuous architecting of Big-Data streaming designs?” As an answer to this question this paper elaborates on OSTIA, a toolkit to assist designers and developers in this continuous architecting campaign. OSTIA helps designers and developers by recovering and analysing the architectural topology on-the-fly, assisting them in: (a) reasoning on the topological structure and how to refine it; (b) export the topological structure consistently with restrictions of their reference development framework so that further analysis (e.g., formal verification) may ensue. In addition, while performing on-the-fly architecture recovery, the analyses that OSTIA is able to apply focus on checking for the compliance to essential consistency rules specific to targeted big data frameworks. Finally, OSTIA allows to check whether the recovered topologies contain occurrences of key anti-patterns we elicited and studied in our own experience and previous work.

By running a case-study with a partner organization, we observed that OSTIA assists designers and developers in establishing and continuously improving the quality of topologies behind their big data applications in multiple ways.

In the future we plan to further our understanding of the anti-patterns that may emerge across big data topologies, e.g., as discussed, by learning said anti-patterns by using graphs analysis techniques inherited from social-networks analysis. Also, we plan to expand OSTIA to support further technologies beyond the most common application framework for streaming, i.e., Storm. Finally, we plan to further evaluate OSTIA using more ad-hoc empirical evaluation campaigns in industry.

ACKNOWLEDGMENT

The Authors’ work is partially supported by the European Commission grant no. 644869 (EU H2020), DICE. Also, Damian’s work is partially supported by the European Commission grant no. 610531 (FP7 ICT Call 10), SeaClouds.

REFERENCES

- [1] C. K. Emani, N. Cullot, and C. Nicolle, “Understandable big data: A survey.” *Computer Science Review*, vol. 17, pp. 70–81, 2015. [Online]. Available: <http://dblp.uni-trier.de/db/journals/csr/csr17.html#EmaniCN15>

⁹<https://github.com/CELAR/c-Eclipse>

¹⁰<https://www.oasis-open.org/apps/org/workgroup/tosca/>

¹¹<http://alien4cloud.github.io/>

- [2] B. Ratner, *Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data*. CRC PressINC, 2012. [Online]. Available: <http://books.google.de/books?id=0KuJTAznLUC>
- [3] D. A. Tamburri, P. Lago, and H. van Vliet, "Organizational social structures for software engineering," *ACM Comput. Surv.*, vol. 46, no. 1, p. 3, 2013.
- [4] L. J. Bass, I. M. Weber, and L. Zhu, *DevOps - A Software Architect's Perspective.*, ser. SEI series in software engineering. Addison-Wesley, 2015.
- [5] I. Ivkovic and M. W. Godfrey, "Enhancing domain-specific software architecture recovery," in *IWPC*. IEEE Computer Society, 2003, pp. 266–273. [Online]. Available: <http://dblp.uni-trier.de/db/conf/iwpc/iwpc2003.html#IvkovicG03>
- [6] R. Evans, "Apache storm, a hands on tutorial." in *IC2E*. IEEE, 2015, p. 2. [Online]. Available: <http://dblp.uni-trier.de/db/conf/ic2e/ic2e2015.html#Evans15>
- [7] D. Frankel, *Model Driven Architecture: Applying MDA to Enterprise Computing*. New York, NY, USA: John Wiley & Sons, Inc., 2002.
- [8] D. C. Schmidt, M. Stal, H. Rohnert, and F. Buschmann, *Pattern-oriented software architecture. Vol. 2, Patterns for concurrent and networked objects*. John Wiley, Chichester, England, 2000.
- [9] M. M. Bersani, S. Distefano, L. Ferrucci, and M. Mazzara, "A timed semantics of workflows." in *ICSOFT (Selected Papers)*, ser. Communications in Computer and Information Science, A. Holzinger, J. Cardoso, J. Cordeiro, T. Libourel, L. A. Maciaszek, and M. van Sinderen, Eds., vol. 555. Springer, 2014, pp. 365–383. [Online]. Available: <http://dblp.uni-trier.de/db/conf/icsoft/icsoft2014s.html#BersaniDFM14>
- [10] D. L. Morgan, *Focus groups as qualitative research*. Sage Publications, Thousand Oaks, 1997.
- [11] M. Hammersley and P. Atkinson, *Ethnography*. London: Routledge, 2003.
- [12] S. B. Merriam, *Case study research in education: A qualitative approach*. Jossey-Bass Publishers, 1991, vol. 1, no. 1980. [Online]. Available: <http://psycnet.apa.org/psycinfo/1989-97006-000>
- [13] C. A. Furia, D. Mandrioli, A. Morzenti, and M. Rossi, "Modeling time in computing: A taxonomy and a comparative survey," *ACM Comput. Surv.*, vol. 42, no. 2, pp. 6:1–6:59, Mar. 2010. [Online]. Available: <http://doi.acm.org/10.1145/1667062.1667063>
- [14] M. Quartulli, J. Lozano, and I. G. Olaizola, "Beyond the lambda architecture: Effective scheduling for large scale eo information mining and interactive thematic mapping." in *IGARSS*. IEEE, 2015, pp. 1492–1495. [Online]. Available: <http://dblp.uni-trier.de/db/conf/igarss/igarss2015.html#QuartulliLO15>
- [15] M. Hirzel, H. Andrade, B. Gedik, G. Jacques-Silva, R. Khandekar, V. Kumar, M. P. Mendell, H. Nasgaard, S. Schneider, R. Soul, and K.-L. Wu, "Ibm streams processing language: Analyzing big data in motion." *IBM Journal of Research and Development*, vol. 57, no. 3/4, p. 7, 2013. [Online]. Available: <http://dblp.uni-trier.de/db/journals/ibmrd/ibmrd57.html#HirzelAGJKKMNSSW13>
- [16] D. Wang and J. Liu, "Optimizing big data processing performance in the public cloud: opportunities and approaches." *IEEE Network*, vol. 29, no. 5, pp. 31–35, 2015. [Online]. Available: <http://dblp.uni-trier.de/db/journals/network/network29.html#WangL15>
- [17] Y. Tamura and S. Yamada, "Reliability analysis based on a jump diffusion model with two wiener processes for cloud computing with big data." *Entropy*, vol. 17, no. 7, pp. 4533–4546, 2015. [Online]. Available: <http://dblp.uni-trier.de/db/journals/entropy/entropy17.html#TamuraY15>
- [18] P. Jamshidi and G. Casale, "An uncertainty-aware approach to optimal configuration of stream processing systems." in *VLDB (under evaluation)*, 2016.
- [19] K. Chandrasekaran, S. Santurkar, and A. Arora, "Stormgen - a domain specific language to create ad-hoc storm topologies." in *FedCSIS*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2014, pp. 1621–1628. [Online]. Available: <http://dblp.uni-trier.de/db/conf/fedcsis/fedcsis2014.html#ChandrasekaranSA14>