

Improving Quality in Storm-Based Big-Data Architectures Using OSTIA

Pooyan Jamshidi
Imperial College London
London, UK

Email: p.jamshidi@imperial.ac.uk

Andrea Nodari
Imperial College London
London, UK

Email: a.nodari15@imperial.ac.uk

Damian A. Tamburri
Politecnico di Milano
Milan, Italy

Email: damianandrew.tamburri@polimi.it

Abstract—[title suggestions: (i) **Topology Architecture Recovery for Stream Based Systems** (ii) **Automated Architecture Recovery and Consistency Checking for Stream Based Systems**]

Big data architectures have been gaining momentum in the last few years. For example, Twitter uses complex Stream topologies using fault tolerant frameworks like Storm to analyse and learn trending topics from billions of tweets per minute. However, verifying the consistency of stream topologies before actual deployment on multi-node clusters is expensive and time consuming. As an aid to designers and developers evaluating their Stream topologies at design-time, we developed OSTIA, that is, “On-the-fly Storm Topology Inference Analysis”. OSTIA offers round-trip engineering of Storm topologies so that designers and developers may: (a) use previously existing model-driven verification&validation techniques on elicited models; (b) visualise and evaluate these models against simple consistency checks that would only be available at deployment and run-time. We illustrate the uses and benefits of OSTIA on three real-life industrial case studies.

I. INTRODUCTION

- I would follow the path of the abstract, we should probably provide some numbers and info on storm
- mind you we should stress on the innovative aspects of the paper and tech. there is nothing strictly related to it
- we should comment on what could be done with OSTIA in combination with Eclipse Based tech.

II. RESEARCH APPROACH

- so we had a focus group to actually elaborate the approach
- then we used explorative prototyping to elicit the initial version of the prototype and then refined that by means of case study, we could mention that we used ATC as an action-research source (this is what we are doing now internally in WP2 actually)
- ...

III. RESEARCH SOLUTION

IV. EVALUATION

- we can use the ATC case study as much as we want - that yields already three topologies that we can infer
- 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

V. DISCUSSION

A. Findings and Quality-Improvement Insights

- we should probably enumerate the many issues we have found with the work in WP2 and WP3 to allow some good insights for the people using OSTIA
- also we could discuss the benefits one by one using some usage-scenarios

B. Approach Limitations and Threats to Validity

- here we could elaborate on the limitations of addressing only storm and why we are addressing storm (e.g., the main technology for streaming currently)
- we should probably discuss the fact that the tech is still eclipse-based and how this is not really a limitation after all
- are there any threats to validity? we should probably discuss this as well

VI. RELATED WORK

- mention DICE
- mention work by Len Bass on Big-Data
- other stuff on big data?
- feel free to extend this section with Previous work of course :)

VII. CONCLUSION

ACKNOWLEDGMENT

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.