DOSSIER-Cloud

DEVOPS-BASED SOFTWARE ENGINEERING FOR THE CLOUD

http://www.dossier-cloud.eu

Deliverable D6.4

Short papers, posters, work-in-progress papers

Document details:

Editor :	Andreas Christoforou
Contributors :	Andreas Andreou, Panayiotis Christodoulou, Lambros Odysseos, Spyros Loizou, Michalis Pingos
Date:	22 Jan 2019
Version:	4.0

Document history:

Version	Date	Contributor	Comments
1.0	22/01/19	Lambros Odysseos	Initial document, structure and content
2.0	25/01/19	Andreas Christoforou	Additional content, first review
3.0	30/01/19	Panayiotis Christodoulou	Additional content, second review
4.0	04/02/19	Andreas Andreou	Final corrections
5.0	08/02/19	Partners	Final approval

Contents

1. Introduction

1.1 Purpose

This document introduces all types of research papers that have already been publsished so far by the project or describes studies are currently being conducted by CUT researchers in collaboration with the leading institutions. This work consists of research papers, master and undergraduate thesis and also parts of PhD work that is facilitated by DOSSIER-Cloud.

This deliverable is part of Workpackage-6 (WP6) that provides dissemination, communication and exploitation of successful transfer of knowledge from the leading institutions (POLIMI, UvT) to CUT targeting at enriching its knowledge base on research and technical issues.

1.2 Definitions, Acronyms, and Abbreviations

CUT: Cyprus University of Technology

UvT: University of Tilburg

POLIMI: Politecnico di Milano

JADS: Jheronimus Academy of Data Science

FCM: Fuzzy Cognitive Maps

FaaS: Function as a Service

MOGAS: Multi-Objective Genetic Algorithms

1.3 Synopsis of Publications

The following papers have already been published or have been accepted for publication in forthcoming conferences (see Appendix for full texts):

1. Christoforou, A., Garriga, M., Andreou, A. S., & Baresi, L. (2017, November). *Supporting the Decision of Migrating to Microservices Through Multi*-layer Fuzzy Cognitive Maps. In *International Conference on Service-Oriented Computing* (ICSOC 2017) (pp. 471-480). Springer, Cham.
2. Andreas Christoforou, Andreas Andreou, Luciano Baresi, Michael Papazoglou (September, 2018). *Devops-Based Software Engineering for the Cloud*, Proceedings of the 7th IFIP WG 2.14 *European Conference on Service-Oriented and Cloud Computing* (ESOCC 2018), Como, Italy.
3. Chatzis, S. P., Christodoulou, P., & Andreou, A. S. (2017, August). Recurrent Latent Variable Networks for Session-Based Recommendation. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems* (pp. 38-45). ACM-RECSYS.

4. Papazoglou, M., & Andreou, A. (Accepted/In press). Smart connected digital factories: Unleashing the power of industry 4.0. In D. Ferguson, V. Mendez Munoz, M. Helfert, & C. Pahl (Eds.), *Proceedings of the 8th International Conference on Cloud Computing and Services Science (CLOSER 2018)* (pp. 5-16). (Cloud Computing and Services Science Selected Papers 2018). Funchal,Madeira: Springer Verlag.
5. Christodoulou, P., Chatzis, S. P., & Andreou, A. S. (2018, July). A Recurrent Latent Variable Model for Supervised Modeling of High-Dimensional Sequential Data. In *2018 Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1-9). IEEE.
6. Andreas Christoforou and Andreas Andreou (December 2018), *An effective resource management approach in a FaaS environment,* European Symposium on Serverless Computing and Applications, Zurich, Switzerland.
7. Michalis Pingos, Amal Elgammal, Indika Priyantha, Panayiotis Christodoulou, Mike P.Papazoglou and Andreas S. Andreou. Smart Shop-Floor Monitoring Via Manufacturing Blueprints and Complex-Event Processing. *21st International Conference on Enterprise Information Systems (ICEIS)*, Heraclion, Crete, Greece, 2-5 May, 2019 (accepted, forthcoming).
8. Spyros Loizou, Amal Elgammal, Indika Priyantha, Panayiotis Christodoulou, Mike P.Papazoglou and Andreas S. Andreou. A Smart Product Co-Design and Monitoring Framework via Gamification and Complex Event Processing. *21st International Conference on Enterprise Information Systems (ICEIS)*, Heraklion, Crete, Greece, 2-5 May, 2019 (accepted, forthcoming).
9. Andreas Christoforou, Lambros Odysseos and Andreas S. Andreou. Microservice Synthesis: A Framework for Replacing Components with Microservices. *14th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)* Heraklion, Crete, Greece, 2-5 May, 2019 (accepted, forthcoming).

1.4 Overview

The rest of the document is structured based on the research work that CUT has conducted throughout the duration of the project in close collaboration with POLIM and UvT and was formulated by the research challenges identified in the project. Furthermore, new research challenges that emerged during the collaboration between the partners, are also being addressed. In the following sections the research studies and papers produced are presented per research thematic area and topics. The deliverable then provides a section with some conclusions. Finally, the Appendix closes the deliverable and gives the full text of the published papers.

2. Social Software Engineering

2.1 Master Thesis

Social Debt and Community Smells modeling through FCM

This research work involves investigating and creating a generic model that can predict through simulation whether a software organization is or will be in a state of Social Debt. In order to achieve this prediction, we employ Fuzzy Cognitive Maps (FCMs) that will specifically allow us to create a visual model of the interactions between the different concepts involved in community smells (i.e., the factors that dictate whether a state of Social Debt is present).

Work carried out up until now involved gathering and analysing the most common community smells that lead to this highly damaging and costly phenomenon from various published articles. From these articles we elicited all the concepts (variables, states, events or actions) that are known to take part either as causes or effects in a community smell. We then created individual FCM models for each community smell separately, depicting the associations between the concepts involved. Following this, we then constructed a consolidated FCM of all community smells in order to obtain a unified FCM model.

Next steps in this work will involve creating a survey targeting project managers that will aid in initializing the concepts of the FCM model, as well as the level of influence between them (i.e., the strength of the interaction), so that necessary scenarios can then be simulated.

Prediction of community smells with Bayesian models

This research is about creating an algorithm that, considering some specific variables as an input, can predict whether a software development team falls into a community smell. This algorithm will be created by converting and customizing existing Bayesian algorithms to facilitate data aggregation and prediction.

For the supervised learning of the algorithm, data will be used that relate to software development teams of specific projects that have already been developed. Projects like Firefox, Libre Office etc. In particular, they contain data relating to communication and cooperation between the individuals involved in the development teams and on the basis of some specific combinations of data, we observe when and where a community smell occurs.

The data is in the form of variables. We first understood and organized the data. Based on the information we provide, we choose which ones will be used as variables for learning. Then the algorithm will be implemented so that it has the ability, taking into account specific data, to predict when a community smell will occur. For the above work, several articles on community smells and a research work from POLIMI have been used that explain exactly what every variable in the data is showing us, since the data gathered for this work.

3. Cloud Resource Management

3.1 Research Papers

An Effective Resource Management Approach in a FaaS Environment (Conference Paper: Presented at ESSCA2018)

Serverless computing introduces a new Cloud service which consists of an increasingly popular architecture for building distributed applications. This paper investigates and proposes a new resource management approach in a FaaS platform, based on intelligent techniques. A number of experiments were conducted through an indicative framework consisting of a client application and a Lambda function. Three Genetic Algorithms were employed to deliver optimal solutions in a multi-objective environment.

Resource Management: Expansion of xSpark (Conference Paper: In Progress)

Big-data applications are batch applications that exploit dedicated frameworks to perform massively parallel computations across clusters of machines. The time needed to process the entirety of the inputs represents the application's response time, which can be subject to deadlines. Spark, probably the most famous incarnation of these frameworks today, allocates resources to applications statically at the beginning of the execution and deviations are not properly managed; to meet the application's deadlines, resources must be allocated carefully.

This research work is based on a previous work conducted by POLIMI members and aims to modify and expand furthermore the xSpark framework. xSpark is an extension to Spark, that is able to allocate and redistribute resources to applications dynamically to meet deadlines and cope with the execution of unanticipated applications. More specifically, the purpose is to use specific machine learning algorithms to allocate resources dynamically depending on the size of the workloads that are executed. The main difference from the previous work is the substitution of the DAG scheduler (which calculates the execution times of local deadlines) with a machine learning algorithm to enhance accuracy.

3.2 Master Thesis

DevOps Oriented and Cloud Focused Metrics and Measurements

Monitoring of cloud computing is vital for the optimal management of the services offered like platforms, infrastructures, software, etc. Cloud monitoring requires standardized metrics that will enable efficient controlling of its resources, both human and technical, towards improved quality of the offered service. Similarly, in a DevOps software development environment, automation activities require efficient monitoring that will offer optimal support on build, deployment and operation processes.

The target of this research work is to deliver a set of metrics for the development and running of distributed software systems, along with the associated ways of measurement. The proposed set of metrics will be formatted in a straightforward process to measure distributed software systems aspects. Both the metrics and the process to acquire their values will constitute a framework that will be used as a dedicated guideline for efficient monitoring of DevOps and cloud services environments. The proposed framework will be evaluated and applied over various demonstration examples by utilizing synthetic and real-world cases.

Optimization of resource management in a serverless computing environment

Resource management support in a Function as a Service (FaaS) environment is essential for the software development process itself, which is directed towards satisfying the SLA and providing QoS assurance. The identification of the optimum scenario for resource allocation to serve adequately a specific workload is a computationally complex and time consuming process since multiple objectives need to be satisfied.

The main target of this research work is to investigate whether is feasible to apply intelligent algorithms in an efficient way, on a FaaS platform, aiming to achieve a balanced performance and cost through the management of the available resources. The assessment of the main FaaS providers will emerge all exposed resource management options and accordingly experimental environment will be defined. A series of available Multi-Objective approaches such as Multi Objective Genetic Algorithms (MOGAs) and Multi-Objective Particle Swarm Optimization (MOPSO) will be applied and compared in a multidimensional aspect, considering accuracy, complexity, scalability etc.

4. Microservice Architecture

4.1 Research Papers

Supporting the Decision of Migrating to Microservices through Multi-layer Fuzzy Cognitive Maps (Conference Paper: Presented at ICSOC2017)

Microservices architectures are gaining momentum for the development of applications as suites of small, autonomous, and conversational services, which are then easy to understand, deploy and scale. However, one of today's problems is that microservices introduce new complexities to the system and, despite the hype, many factors should be considered when deciding to adopt a microservices architecture. This paper proposes the first Decision Support System (DSS) to migrate to microservices, by identifying the key concepts and drivers regarding through a literature review and feedback from a group of experts from industry and academia. Then, these concepts were organized as a Multi-Layer Fuzzy Cognitive Map (ML-FCM), a graph-based computational intelligence model that captures the behavior of a given problem in nodes that represent knowledge in the domain, and offers the means to study their influence and interrelation. Static and dynamic analysis over the resulting ML-FCM helped us to identify the prevailing drivers towards the migration to a microservices architecture.

Migration of Software Components to Microservices: Matching and Synthesis (Conference Paper: Accepted at ENASE2019)

Nowadays, more and more software companies, as well individual software developers, adopt the microservice architecture for their software solutions. Although many software systems are being designed and developed from scratch, a significant number of existing monolithic solutions tend to be transformed to this new architectural style. What is less common, though, is how to migrate component-based software systems to systems composed of microservices and enjoy the benefits of ease of changes, rapid deployment and versatile architecture. This paper proposes a novel and integrated process for the decomposition of existing software components with the aim being to fully or partially replace their functional parts with a number of suitable and available microservices. The proposed process is built on semi-formal profiling and utilizes ontologies to match between properties of the decomposed functions of the component and those offered by microservices residing in a repository. Matching concludes with recommended solutions yielded by multi-objective optimization which considers also possible dependencies between the functional parts.

Adopting Microservice Architectures: a Decision Support System based on Multi-Layer Fuzzy Cognitive Maps (In progress)

Microservice Architectures (MA) foster the development of applications offering a suite of small, atomic, and autonomous, which are then simple in their understanding, deployment and scaling. Nevertheless, a drawback with MA is the increase complexity introduced by microservices to the system under

development. Therefore, one must consider quite a few parameters and factors before deciding to use microservices. This paper proposes a Decision Support System (DSS) to adopt microservice architectures, conceived by identifying the key concepts and factors in such a decision, through a literature review and experts' feedback from the industry and academia. Then, these concepts are organized as a Multi-Layer Fuzzy Cognitive Map (ML-FCM), a graph-based computational intelligence model that captures the behavior of a given problem in nodes that represent knowledge in the domain, and edges that represent their influence and interrelation. The validation of our model was threefold: Static and dynamic analysis over the ML-FCM helped us to identify the prevailing factors towards the adoption of a microservice architecture. Then, a large scale survey involving both researchers and practitioners allowed us to fine-tune the importance of each concept. Finally, several industrial scenarios were elicited, modeled and executed. Results show that the outcome of our model is aligned with the actual decision/intention of adopting microservices and, moreover, it sheds light over the insights of such a decision in each particular scenario.

5. Smart Manufacturing

5.1 Research Papers

Smart Connected Digital Factories: Unleashing the Power of Industry 4.0. The paper is partially funded by the European union's H2020 project Dossier-Cloud.

Recent initiatives such as the Industrial IoT, or Industry 4.0, as it has been dubbed, are fundamentally reshaping the industrial landscape by promoting connected manufacturing solutions that realize a "digital thread" which connects all aspects of manufacturing including all data and operations involved in the production of goods and services. This paper focuses on Industry 4.0 technologies and how they support the emergence of highly-connected, knowledge-enabled factories, referred to as Smart Manufacturing Networks. Smart Manufacturing Networks comprise an ecosystem of connected factory sites, plants, and self-regulating machines able to customize output, and allocate resources over manufacturing clouds optimally to offer a seamless transition between the physical an digital worlds of product design and production.

Smart Shop-Floor Monitoring via Manufacturing Blueprints and Complex-Event Processing (Conference Paper: Accepted at ICEIS2019)

Nowadays, Product-Service-Systems (PSS) are being modernized into smart connected products that target to transform the industrial scenery and unlock unique prospects. This concept enforces a new technological heap and lifecycle models to support smart connected products. The intelligence that smart, connected products embed paves the way for more sophisticated data gathering and analytics capabilities ushering in tandem a new era of smarter supply and production chains, smarter production processes, and even end-to-end connected manufacturing ecosystems. The main contribution of this paper is a smart shop-floor monitoring framework and underpinning technological solutions, which enables the proactive identification and resolution of shop-floor distributions. The proposed monitoring framework is based on the synergy between the novel concept of Manufacturing Blueprints and Complex Event Processing (CEP) technologies, while it encompasses a middleware layer that enables loose coupling and adaptation in practice. The framework provides the basis for actionable PSS and production "intelligence" and facilitates a shift toward more fact-based manufacturing decisions. Implementation and validation of the proposed framework is performed through a real-world case study which demonstrates its applicability, and assesses the usability and efficiency of the proposed solutions.

A Smart Product Co-Design and Monitoring Framework via Gamification and Complex Event Processing (Conference Paper: Accepted at ICEIS2019)

In the traditional software development cycle, requirements gathering is considered the most critical phase. Getting the requirements right early has become a dogma in software engineering because the correction of erroneous or incomplete requirements in later software development phases becomes overly expensive. For product-service systems (PSS), this dogma and standard requirements engineering (RE) approaches are not appropriate because classical RE is considered concluded once a product service

is delivered. This paper proposes a novel framework that enables the customer and the product engineer to co-design smart products by integrating three novel and advanced technologies to support: view-based modelling, visualization and monitoring, i.e., Product-Oriented Configuration Language (PoCL), gamification and Complex Event Processing (CEP), respectively. These create a "*digital-twin*" model of the connected 'smart' factory of the future. The framework is formally founded on the novel concept of manufacturing blueprints, which are formalized knowledge-intensive structures that provide the basis for actionable PSS and production "intelligence" and a move toward more fact-based manufacturing decisions. Implementation and validation of the proposed framework through real-life case studies are ongoing to validate the applicability, utility and efficacy of the proposed solutions.

5.2 Master Thesis

Production Line Simulation and Anomaly Detection

The aim of this master thesis is to detect, prevent and repair faulty situations that may appear in a production line in order to eliminate any possibility of blackout during normal processes. Production Line Simulation will be based on data collection, aggregation and processing sourcing from various sensors. This simulation will be performed either in 'soft' mode (i.e. software based) or in 'hard' mode by utilizing a small robot able to execute different actions depending on the values of its sensors and the instructions received thus substituting machinery at the shop floor. At first, the production line simulation will extract data from various sensors and process them at a higher level, then advanced algorithms that will be developed will be utilized to automatically diagnose faulty states and execute actions for automatic self-healing or notify a control panel. This study targets to help in decision-making to minimize the effort of control panel administrator by self-healing faulty states of production line.

Optimization Techniques in Production Lines: The case of Poultry Meat Industry

The aim of this work is the use of optimization techniques, such as genetic algorithms and particle swarm optimization, to optimize the process of chicken slaughtering and packaging of poultry meat at a local industry in Cyprus and taking into consideration various characteristics and parameters of the different production lines.

6. Smart Data Processing and Data Analytics

6.1 Research Papers

Recurrent Latent Variable Networks for Session-Based Recommendation (Conference Paper: Presented at DLRS2017)

In this work, we attempt to ameliorate the impact of data sparsity in the context of session-based recommendation. Specifically, we seek to devise a machine learning mechanism capable of extracting subtle and complex underlying temporal dynamics in the observed session data, so as to inform the recommendation algorithm. To this end, we improve upon systems that utilize deep learning techniques with recurrently connected units; we do so by adopting concepts from the field of Bayesian statistics, namely variational inference. Our proposed approach consists in treating the network recurrent units as stochastic latent variables with a prior distribution imposed over them. On this basis, we proceed to infer corresponding posteriors; these can be used for prediction and recommendation generation, in a way that accounts for the uncertainty in the available sparse training data. To allow for our approach to easily scale to large real-world datasets, we perform inference under an approximate amortized variational inference (AVI) setup, whereby the learned posteriors are parameterized via (conventional) neural networks. We perform an extensive experimental evaluation of our approach using challenging benchmark datasets, and illustrate its superiority over existing state-of-the-art techniques.

6.2 Indergraduate Diploma Thesis

Intelligent Processing and Big Data Analytics: Visualization and Graphical Representation of Abnormalities in High Frequency and Range Data

The target of this thesis is the visualization of error conditions and messages for any abnormalities observed in the data obtained, which are analyzed by another work (see point 2 below). The purpose of this study is the graphical representation of existing data, but also the monitoring of the new flow of data entering the system in real-time to support decision making. This will assist decision-makers to take more informed decisions faster and more effective.

Intelligent Processing and Big Data Analytics: Collection and Processing of High Frequency and Range Data in Real Time for Anomaly Detection

The aim of this thesis was to collect and analyse Big Data to assist the process of taking decisions relying on the data using innovative technology to handle and analyse huge quantities of data and investigate how this would be able to contribute to business sectors, such as the automotive industry. The results of the case study used in this work sketched the role of the individual parts of vehicles and their normal

functioning in general. The results showed that the analytics techniques applied in huge automotive data are valuable for performing predictive maintenance.

6.3 Master Thesis

Big Data collection, homogenization and decision support

The evolution of technology and its wide application to industry, causes a constant increase in the need for management and beneficial use of the vast amount of information produced (Big Data). In a real industrial production line unit where digitization and automation processes have been adopted, it is possible to apply measurements and produce data at each stage of production. Despite the great availability of data, in the majority of the cases a very small percentage of these data is used and only for limited purposes.

The adoption and utilization of cost-effective Big Data Analytics processes in the industry is a major challenge for computer science and brings together engineers from different IT departments such as Data Scientists and Data Engineers. Implementing Big Data Analytics processes in organizations can deliver multiple benefits such as improving production processes, reducing costs and production time, improving the position against competition, assuring and improving quality, improving performance and so forth. These benefits can arise from the intelligent management of available data leading to more accurate predictions and subsequently to more accurate and efficient decisions.

The purpose of this work is to develop and deliver a suite of tools for implementing Big Data Analytics processes in a production unit of an industrial factory. This suite will incorporate tools and applications that will deal with data collection and recording, intelligent processing and visual representation of data and results. Additionally, within this framework the performance of various computational and artificial intelligence models and techniques for intelligent data processing will be studied and compared, indicating the most suitable for the tasks in hand.

7. Conclusions

The aim of DOSSIER-Cloud was to facilitate transfer of scientific knowledge and expertise, as well as of best research practices from UvT and POLIMI to CUT. This aim led to the ultimate goal, that is, the increase of CUT's research capacity and prowess through the investigation of a number of significant and hot topics in the field of Distributed Systems development.

In the context of Workpackage-6 (WP6), we have organized all the research works that have been or are being carried out from CUT research group. In this deliverable, research papers, bachelor and master thesis are presented. The main research topics that our efforts focus on are coming from those that are descripted by the initial project objectives and in addition from those that emerged through the collaboration with the project partners. The amount of research work along with the corresponded results can be summarized in 7 master thesis, 2 bachelor thesis and 8 research papers. Research results produced so far are quite successful and encouraging and that leads and defines the future research steps.

APPENDIX: Published Papers

1. Christoforou, A., Garriga, M., Andreou, A. S., & Baresi, L. (2017, November). *Supporting the Decision of Migrating to Microservices Through Multi*-layer Fuzzy Cognitive Maps. In *International Conference on Service-Oriented Computing* (ICSOC 2017) (pp. 471-480). Springer, Cham.
2. Andreas Christoforou, Andreas Andreou, Luciano Baresi, Michael Papazoglou (September, 2018). *Devops-Based Software Engineering for the Cloud*, Proceedings of the 7th IFIP WG 2.14 *European Conference on Service-Oriented and Cloud Computing* (ESOCC 2018), Como, Italy.
3. Chatzis, S. P., Christodoulou, P., & Andreou, A. S. (2017, August). Recurrent Latent Variable Networks for Session-Based Recommendation. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems* (pp. 38-45). ACM-RECSYS.
4. Papazoglou, M., & Andreou, A. (Accepted/In press). Smart connected digital factories: Unleashing the power of industry 4.0. In D. Ferguson, V. Mendez Munoz, M. Helfert, & C. Pahl (Eds.), *Proceedings of the 8th International Conference on Cloud Computing and Services Science (CLOSER 2018)* (pp. 5-16). (Cloud Computing and Services Science Selected Papers 2018). Funchal,Madeira: Springer Verlag.
5. Christodoulou, P., Chatzis, S. P., & Andreou, A. S. (2018, July). A Recurrent Latent Variable Model for Supervised Modeling of High-Dimensional Sequential Data. In *2018 Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1-9). IEEE.
6. Andreas Christoforou and Andreas Andreou (December 2018), *An effective resource management approach in a FaaS environment,* European Symposium on Serverless Computing and Applications, Zurich, Switzerland
7. Michalis Pingos, Amal Elgammal, Indika Priyantha, Panayiotis Christodoulou, Mike P.Papazoglou and Andreas S. Andreou. Smart Shop-Floor Monitoring Via Manufacturing Blueprints and Complex-Event Processing. *21st International Conference on Enterprise Information Systems (ICEIS)*, Heraclion, Crete, Greece, 2-5 May, 2019 (accepted, forthcoming).
8. Spyros Loizou, Amal Elgammal, Indika Priyantha, Panayiotis Christodoulou, Mike P.Papazoglou and Andreas S. Andreou. A Smart Product Co-Design and Monitoring Framework via Gamification and Complex Event Processing. *21st International Conference on Enterprise Information Systems (ICEIS)*, Heraklion, Crete, Greece, 2-5 May, 2019 (accepted, forthcoming).
9. Andreas Christoforou, Lambros Odysseos and Andreas S. Andreou. Microservice Synthesis: A Framework for Replacing Components with Microservices. *14th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)* Heraklion, Crete, Greece, 2-5 May, 2019 (accepted, forthcoming).