Developing Data-Intensive Cloud Applications with Iterative Quality Enhancements

Software Engineering Book

Deliverable 7.4

Deliverable:	D7.4
Title:	Software Engineering Book
Editor(s):	Youssef RIDENE (NETF)
Contributor(s):	Giuliano Casale (IMP), Youssef RIDENE (NETF), Joas Yannick Kinouani(NETF), Laurie-Anne Parant (NETF), Chen Li (IMP), Lulai Zhu (IMP), Tatiana Ustinova (IMP), Pooyan Jamshidi (IMP), Danilo Ardagna (PMI), Marcello Bersani (PMI), Elisabetta Di Nitto (PMI), Eugenio Gianniti (PMI), Michele Guerriero (PMI), Matteo Rossi (PMI), Damian Andrew Tamburri (PMI), Safia Kalwar (PMI), Francesco Marconi (PMI), Gabriel Iuhasz (IEAT), Dana Petcu (IEAT), Ioan Dragan (IEAT), Matej Artač, (XLAB), Tadej Borovšak (XLAB), Craig Sheridan (FLEXI), David McGowran, Grant Olsson (FLEXI), Vasilis Papanikolaou (ATC), George Giotis (ATC), Christophe Joubert (PRO), Ismael Torres (PRO), Marc Gil (PRO), Simona Bernardi (ZAR), Abel Gómez (ZAR), José Merseguer (ZAR), Diego Pérez (ZAR), José-Ignacio Requeno (ZAR)
Reviewers:	Chen Li (IMP), Dana Petcu (IEAT)
Type (R/P/DEC):	DEC
Version:	1.0
Date:	31-January-2018
Status:	Final version
Dissemination level:	Public
Download page:	http://www.dice-h2020.eu/resources/
Copyright:	Copyright © 2018, DICE consortium – All rights reserved

DICE partners

ATC:	Athens Technology Centre
FLEXI:	Flexiant Limited
IEAT:	Institutul E Austria Timisoara
IMP:	Imperial College of Science, Technology & Medicine
NETF:	Netfective Technology SA
PMI:	Politecnico di Milano
PRO:	Prodevelop SL
XLAB:	XLAB razvoj programske opreme in svetovanje d.o.o.
ZAR:	Unversidad De Zaragoza

The DICE project (February 2015-January 2018) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644869

Table of contents

TABLE OF CONTENTS **3**
1. INTRODUCTION **4**
2. CONTENT **7**
3. CONCLUSION **8**

1. Introduction

Since the beginning of the project, the DICE consortium considered the communication and dissemination activities as a central element for the success of the project. An ambitious strategy has been established in order to reach a large audience including scientific/academic communities and industrial companies. As planned, the execution of the strategy has been successful thanks to the involvement of all partners. In fact, all the possible communication channels have been intensively used (Website, scientific publications, blog posts, industrial events, keynotes, meetups, social Medias, videos, Wiki, Webinars, training sessions, etc.). This viral multichannel communication generates a lot of interest and consequently potential business and scientific opportunities for mostly all the DICE partners. This estimation includes the website analytics, social media analytics, GitHub analytics, deliverable downloads, persons reached in events (online/offline), etc. And actually, **70% of the audience was reached online** which is not surprising regarding the profiles of the targeted audience. This audience analysis **motivated the idea to publish the DICE Book online**, making it available in the open to the community of end users.

From another perspective, the sustainability strategy of the DICE ecosystem were thought and built around open source software were anyone can download the DICE IDE and extend it within any new Big Data technology. And actually, this is already in action with people who started adapting/modifying DICE to their own needs. This situation was interesting but also challenging, i.e., does it make sense to publish a "static" book about a product which is already adopted by some external users in a highly changing Big Data ecosystem?

Taking into consideration the double dilemma of a up-to-date "living" book and reaching the largest possible audience, the consortium decided to publish the book in a collaborative community-oriented platform. After a test and assessment period, we decided to use WikiBooks (https://www.wikibooks.org/) which offers a bench of advantages such as a huge active community, collaborative editing, moderation, etc.

As of today, the DICE Book "**Practical DevOps for Big Data**" is available at

https://en.wikibooks.org/wiki/Practical_DevOps_for_Big_Data

The book even started receiving inputs of people external to DICE, as we could establish from the edit history, and has enjoyed the constant review of WikiBooks volunteer reviewers, which in a number of occasions corrected small problems to the book, as illustrated in the screenshot in Figure 1.

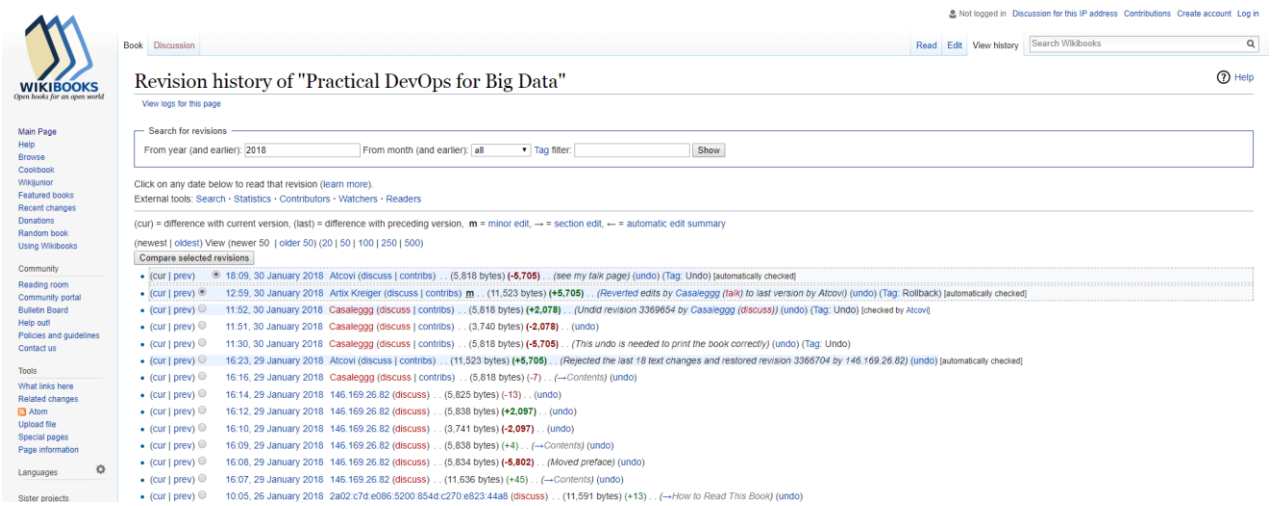

Figure 1 Book editing history - the two topmost contributions are from WikiBooks reviewers

Moreover, the WikiBooks platform offers an easy PDF export feature. This feature may be used, for instance, to publish a free version on other platforms such as Amazon Kindle and Google Play Books. Moreover, through its PressPedia service, WikiBooks allows to obtain a hardcopy of the book in return of a small price (about 20 euros).

Figure 2 PressPedia service to print the WikiBook

Figure 3 further illustrates the advertisement we have done on the DICE website of our book, so that end users can rapidly access to the relevant WikiBooks page.

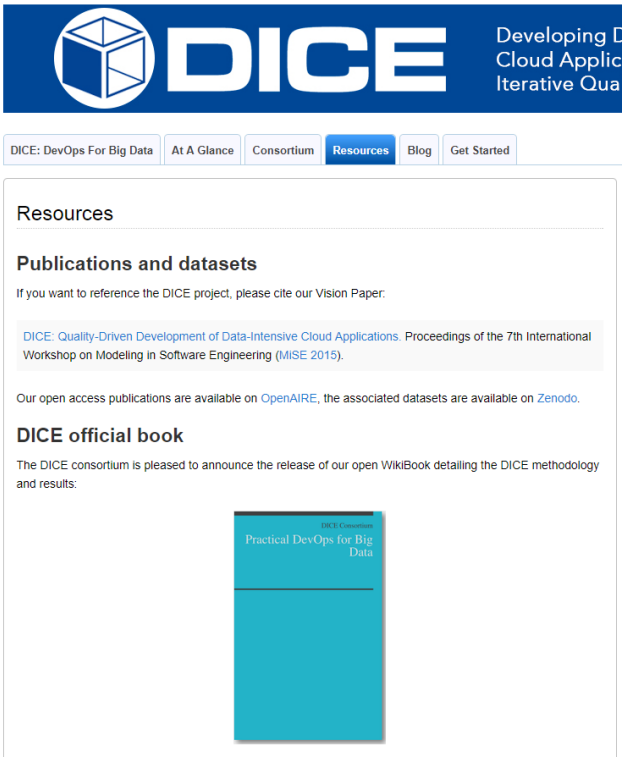

Figure 3 DICE Book advertisment on project website

Professor Franck Barbier (University of Pau[1]), has kindly accepted to preface the DICE Book. Actually this is a supplementary sign arguing for the quality of the contribution, as Prof. Franck Barbier is an international expert with over 20 years experience both in Academia and Industry. He has published around 150 books/papers mainly related to Model-Driven Engineering (MDE) and Unified Modeling Language (UML). He also was deputy head of the ICTS department at the French National Research Agency. Currently he is deputy member of the Council of the French Universities.

[1] http://www.univ-pau.fr/en/home.html

2. Content

Practical DevOps for Big Data is about a methodology for constructing big data applications. A methodology exists for the purpose of solving software development problems. It is made of development processes—workflows, ways of doing things—and tools to help concretise them. The ideal and guiding principle of a methodology is to facilitate the job and guarantee the satisfaction of stakeholders involved in a software project—end-users and maintainers included. Our methodology addresses the problem of reusing complex and not easily learned big data technologies to effectively and efficiently build big data systems of good quality. To do so, it takes inspiration from two other successful methodologies: DevOps and model-driven engineering. Regarding prerequisites, we assume the reader has a general understanding of software engineering, and, from a tool point of view, a familiarity with the Unified Modeling Language (UML) and the Eclipse IDE. The book is composed of eight parts, an overview of their content is directly available on the WikiBook version and in the print attached to this deliverable.

1. Introduction
    - Introduction
    - Related Work
2. DevOps and Big Data Modelling
    - Methodology
    - Review of UML Diagrams Useful for Big Data
3. Modelling Abstractions
    - Introduction to Modelling
    - Platform-Independent Modelling
    - Technology-Specific Modelling
    - Deployment-Specific Modelling
4. Formal Quality Analysis
    - Quality Verification
    - Quality Simulation
    - Quality Optimisation
5. From Models to Production
    - Delivery
    - Configuration Optimisation
    - Quality Testing
    - Fault Injection
6. From Production Back to Models
    - Monitoring
    - Anomaly Detection
    - Trace Checking
    - Iterative Enhancement
7. Case Studies
    - Fraud Detection
    - Maritime Operations
    - News and Media
8. Conclusion
    - Future Challenges
    - Closing Remarks

3. Conclusion

The consortium spent a significant effort in order to meet a high quality document which can actually used by any external parties with limited resources to succeed its Big Data project. We foresee that the DICE community will continue working on the WikiBook in order to enrich it within any new feature added to their tools, i.e. a future-proof strategy for a volatile IT ecosystems.

The PDF version of the book is available in attachment to this deliverable. Due to a limitation of the WikiBooks platform itself, **the PDF print from browser does not always reflect the correct alignment of figure and the font sizes that is visible in the online version**, we therefore invite the readers to directly check the online WikiBooks version:

https://en.wikibooks.org/wiki/Practical_DevOps_for_Big_Data