2017 Imperial College Computing Student Workshop

ICCSW'17, September 26–27, 2017, London, United Kingdom

Edited by

Fergus Leahy

Juliana Franco

OASICS

ICCSW17

Editors

Fergus Leahy
Department of Computing
180 Queen's Gate, London, SW7 2AZ
United Kingdom
`fergus.leahy@imperial.ac.uk`

Juliana Franco
Department of Computing
180 Queen's Gate, London SW7 2AZ
United Kingdom
`juliana.franco@imperial.ac.uk`

ACM Classification 1998
D.1.3 Concurrent Programming, D.2.2 Design Tools and Techniques, D.2.8 Performance measures, D.2.11 Software Architectures, D.3.3 - Language Constructs and Features, D.4.7 Distributed systems, D.4.8 Performance, E.1 Trees, F.1.2. Models of Computation - Probabilistic Computation, F.4.1. Mathematical Logic, F.4.2. Formal Languages, G.1.6 Optimization, G.3 Probability and Statistics, H.2.4 Concurrency, H.5.2 User Interfaces, I.2.6 Learning, K.0 General, K.2 History of Computing.

ISBN 424-2-424242-424-2

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/424-2-424242-424-2.

Publication date
(To be completed: Month, Year)

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Digital Object Identifier: 10.4230/OASIcs.CVIT.2016.0

OASIcs – OpenAccess Series in Informatics

OASIcs aims at a suitable publication venue to publish peer-reviewed collections of papers emerging from a scientific event. OASIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

ISSN 2190-6807

http://www.dagstuhl.de/oasics

Contents

Keynotes

Regular Papers

Abstracts

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco
OASICS Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
ICCSW17

Preface

Welcome to the 2017 Imperial College Computing Student Workshop (ICCSW'17), the sixth workshop in its series. ICCSW was initiated with a "by students, for students" spirit: a workshop organised solely by students to give student speakers the opportunity to present their work. The organising students gain the valuable experience of what is involved in conference organisation, including writing a call for sponsors, taking part in the reviewing process, and chairing a session. On the other hand, the participating students benefit from the interaction with international researchers who are at a similar stage in their career and developing skills in presenting their research to a non-specialist computer science audience.

This volume contains the papers accepted for presentation at the 2017 Imperial College Computing Student Workshop. ICCSW'17 received 12 submissions, including both papers and abstracts, from 6 different countries. After the thorough reviewing process and discussion by members of the Imperial College ACM Student Chapter 6 papers and 2 abstracts were accepted, representing a 75% acceptance rate.

After a year hiatus, ICCSW was back for more, more student talks, more keynotes, more socials and a new addition of a breakfast poster session. ICCSW'17 was a great success on all fronts, with over 30 students attending a variety of interesting and novel talks, covering systems, cloud, networking, programming languages and machine learning. This year we also hosted two exciting keynotes covering Google's V8 Javascript engine (Leszek Swirski) and How to write a great paper (Simon Peyton Jones), both of which saw upwards of 50 students and staff attend, and included some really insightful Q&A. To their merit, students arose first thing in the morning ready to be quizzed on their work at our breakfast poster session, as inquisitive wanderers chowed down on croissants, coffee and enlightening conversation. For our social event, we invited the students and our Googler to explore the Sky Garden atop the Walkie-Talkie and took them on a tour of the surrounding London sights, including the Tower of London and Tower Bridge, before settling down for 3 courses of pizza-and-pasta-riffic food at Pizza Express.

ICCSW'17 has been a great success - but absolutely could not have been done without the dedication and perserverence from the ICCSW committee, hard work and patience from the student authors, assistance from our network of ambassadors in disseminating the calls, financial support from our sponsors and the gratiutous support from the department here at Imperial; all of whom we would like to thank dearly.

We wish the best of luck to the new committee. Until next year!

Juliana Franco and Fergus Leahy,
ICCSW'17 Editors,
Chair & Vice-Chair,
ACM Student Chapter 2016 – 2017.

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco
OASICS Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
ICCSW17

ICCSW'17 Social Photo

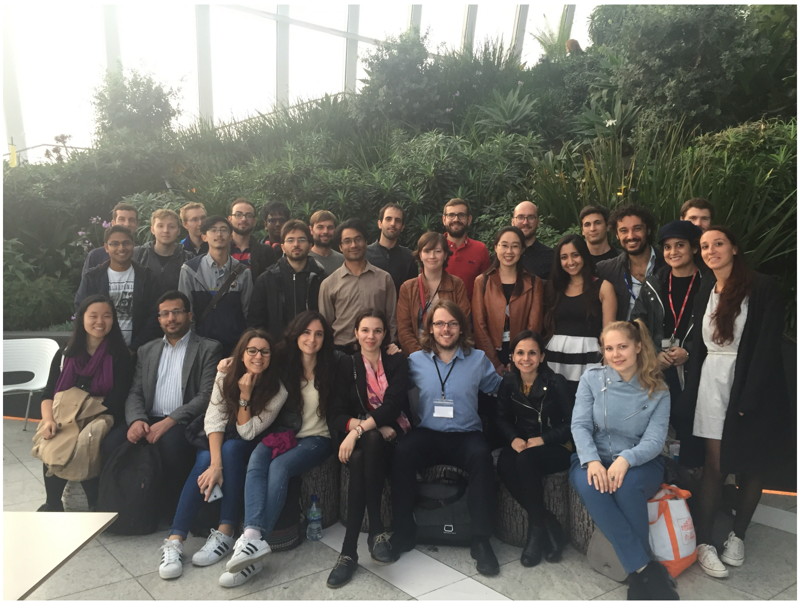

Figure 1 ICCSW visits the sky garden.

Conference Organisation

Organising Committee

Juliana Franco	Imperial College London
Fergus Leahy	Imperial College London
Kyriacos Nikiforou	Imperial College London
Simon Olofsson	Imperial College London
Mengjiao Wang	Imperial College London
Pamela Bezerra	Imperial College London
Shale Xiong	Imperial College London
Oana Cocarascu	Imperial College London
Assel Altayeva	Imperial College London
Nick Pawlowski	Imperial College London
Casper da Costa-Luis	Imperial College London

Imperial College London
ACM Student Chapter
http://acm.doc.ic.ac.uk/

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Ambassadors

Jasper Schulz	Kings College, UK
Fabio Niephaus	Hasso-Plattnner-Institut, Germany
Kiko Fernandez	Uppsala University, Sweden
Phuc Vo	Uppsala University, Sweden
Daniel Hillerstrom	University of Edinburgh, UK
Kim-Anh Tran	Uppsala University, Sweden
Stephan McQuistin	University of Glasgow, UK
Marija Jegorova	University of Edinburgh, UK
Dionysis Manousakas	University of Cambridge, UK
Ana-Maria Sutii	Eindhoven University of Technology, The Netherlands
Tatjana Davidovic	Serbian Academy of Science and Arts, Serbia
Darren Matthews	Royal Holloway, University of London, UJ

Supporters and Sponsors

Supporting Scientific Institutions

Imperial College London
http://www.imperial.ac.uk/

HiPEDS: Imperial College London
http://wp.doc.ic.ac.uk/hipeds/

Platinum Sponsor

Google Inc.
http://www.google.com/

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

How to Write a Great Research Paper

Simon Peyton Jones

Microsoft Research Cambridge

Abstract

Writing papers is a core research skill for any researcher, but they aren't easy. Writing is not just a way to report on great research; it's a way to do great research. Yet many papers are so badly written that, even if they describe excellent work, the work has much less impact than it should. In this talk I'll give you seven simple, actionable guidelines that will, I hope, help you to write better papers, and have more fun at the same time. I don't have all the answers—far from it—and I hope that the presentation will evolve into a discussion in which you share your own insights, rather than a lecture. The slides and video presentation are available online [1].

Simon Peyton Jones, FRS, graduated from Trinity College Cambridge in 1980. After two years in industry, he spent seven years as a lecturer at University College London, and nine years as a professor at Glasgow University, before moving to Microsoft Research (Cambridge) in 1998.

1998 ACM Subject Classification K.0 General

Keywords and phrases Academia, Research, Writing

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.01

[1] https://www.microsoft.com/en-us/research/academic-program/write-great-research-paper/

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 01; pp. 01:1–01:1
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

JavaScript: A Whirlwind tour

Leszek Swirski

Google Inc.

Abstract

A whirlwind tour through the history and state-of-the-art of JavaScript execution and optimization, with a focus on the V8 engine used by Chrome and Node.js, and how a 10-day prototype became one of the most important programming languages in the world.

Leszek Swirski has been a software engineer at Google for two years, first in California, now in London, working on the performance of the Android camera and the V8 Javascript engine. Before joining Google, Leszek did a PhD in the University of Cambridge, researching gaze estimation (a.k.a. eye tracking) on stereoscopic (a.k.a. "3D") displays.

1998 ACM Subject Classification D.3.3 - Language Constructs and Features, K.2 History of Computing

Keywords and phrases Javascript, NodeJS

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.02

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 02; pp. 02:1–02:1
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

ICCSW17

Improving the Latency and Throughput of ZooKeeper Atomic Broadcast

Ibrahim EL-Sanosi[1,2] and Paul Ezhilchelvan[2]

1 **Faculty of Information Technology, Sebha University, Sebha, Libya**
`i.elsanosi@sebhau.edu.ly`

2 **School of Computing, Newcastle University, Newcastle Upon Tyne, UK**
`{i.s.el-sanosi,paul.ezhilchelvan}@ncl.ac.uk`

Abstract

ZooKeeper is a crash-tolerant system that offers fundamental services to Internet-scale applications, thereby reducing the development and hosting of the latter. It consists of $N \geq 3$ servers that form a replicated state machine. Maintaining these replicas in a mutually consistent state requires executing an Atomic Broadcast Protocol, *Zab*, so that concurrent requests for state changes are serialised identically at all replicas before being acted upon. Thus, ZooKeeper performance for update operations is determined by Zab performance. We contribute by presenting two easy-to-implement Zab variants, called *ZabAC* and *ZabAA*. They are designed to offer small atomic-broadcast latencies and to reduce the processing load on the primary node that plays a leading role in Zab. The former improves ZooKeeper performance and the latter enables ZooKeeper to face more challenging load conditions.

1998 ACM Subject Classification D.2.8 Performance measures, D.4.7 Distributed systems

Keywords and phrases Atomic Broadcast, Server Replication, Protocol Latency, Throughput

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.03

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 03; pp. 03:1–03:1
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

ICCSW17

Demand for Medical Care by the Elderly: A Nonparametric Variational Bayesian Mixture Approach

Christoph F. Kurz[1] **and Rolf Holle**[1]

1 **Helmholtz Zentrum München,**
Institute of Health Economics and Health Care Management, Neuherberg, Germany
christoph.kurz@helmholtz-muenchen.de

Abstract

Outpatient care is a large share of total health care spending, making analysis of data on outpatient utilization an important part of understanding patterns and drivers of health care spending growth. Common features of outpatient utilization measures include zero-inflation, over-dispersion, and skewness, all of which complicate statistical modeling. Mixture modeling is a popular approach because it can accommodate these features of health care utilization data. In this work, we add a nonparametric clustering component to such models. Our fully Bayesian model framework allows for an unknown number of mixing components, so that the data, rather than the researcher, determine the number of mixture components. We apply the modeling framework to data on visits to physicians by elderly individuals and show that each subgroup has different characteristics that allow easy interpretation and new insights.

1998 ACM Subject Classification G.3 Probability and Statistics

Keywords and phrases machine learning, health care utilization, Bayesian statistics

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.04

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 04; pp. 04:1–04:1
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

ICCSW17

Discriminative and Generative models for clinical risk estimation: An empirical comparison

John Stamford[1] and Chandra Kambhampati[2]

1 **Department of Computer Science, The University of Hull, Kingston upon Hull, HU6 7RX**
`j.stamford@2014.hull.ac.uk`
2 **Department of Computer Science, The University of Hull, Kingston upon Hull, HU6 7RX**
`c.kambhampati@hull.ac.uk`

Abstract

Linear discriminative models, in the form of Logistic Regression, are a popular choice within the clinical domain in the development of risk models. Logistic regression is commonly used as it offers explanatory information in addition to its predictive capabilities. In some examples the coefficients from these models have been used to determine overly simplified clinical risk scores. Such models are constrained to modeling linear relationships between the variables and the class despite it known that this relationship is not always linear. This paper compares the conditions under which linear discriminative and linear generative models perform best. This is done through comparing logistic regression and naïve Bayes on real clinical data. The work shows that generative models perform best when the internal representation of the data is closer to the true distribution of the data and when there is a very small difference between the means of the classes. When looking at variables such as sodium it is shown that logistic regression can not model the observed risk as it is non-linear in its nature, whereas naïve Bayes gives a better estimation of risk. The work concludes that the risk estimations derived from discriminative models such as logistic regression need to be considered in the wider context of the true risk observed within the dataset.

1998 ACM Subject Classification D.4.8 Performance

Keywords and phrases Discriminative, Generative, Naïve Bayes, Logistic Regression, Clinical Risk

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.5

Discriminative and Generative models for clinical risk estimation: An empirical comparison.
Editors: Fergus Leahy; Article No. 5; pp. 5:1–5:1
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Hey there's DALILA: a DictionAry LearnIng LibrAry

Veronica Tozzo[1], Vanessa D'Amario[1], and Annalisa Barla[1]

1 **Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa**
Genoa, I-16146, Italy
{veronica.tozzo, vanessa.damario}@dibris.unige.it
annalisa.barla@unige.it

Abstract

Dictionary Learning and Representation Learning are machine learning methods for decomposition, denoising and reconstruction of data with a wide range of applications such as text recognition, image processing and biological processes understanding. In this work we present DALILA, a scientific Python library for regularised dictionary learning and regularised representation learning that allows to impose prior knowledge, if available. DALILA, differently from the others available libraries for this purpose, is flexible and modular. DALILA is designed to be easily extended for custom needs. Moreover, it is compliant with the most widespread ML Python library and this allows for a straightforward usage and integration. We here present and discuss the theoretical aspects and discuss its strength points and implementation.

1998 ACM Subject Classification G.1.6 Optimization, D.1.3 Concurrent Programming, D.2.2 Design Tools and Techniques

Keywords and phrases Machine learning, dictionary learning, representation learning, alternating proximal gradient descent, parallel computing

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.06

Faster Concurrent Range Queries with Contention Adapting Search Trees Using Immutable Data

Kjell Winblad[1]

1 **Department of Information Technology, Uppsala University, Sweden**
kjell.winblad@it.uu.se

Abstract

The need for scalable concurrent ordered set data structures with linearizable range query support is increasing due to the rise of multicore computers, data processing platforms and in-memory databases. This paper presents a new concurrent ordered set with linearizable range query support. The new data structure is based on the contention adapting search tree and an immutable data structure. Experimental results show that the new data structure is as much as three times faster compared to related data structures. The data structure scales well due to its ability to adapt the sizes of its immutable parts to the contention level and the sizes of the range queries.

1998 ACM Subject Classification D.2.8 Performance measures, E.1 Trees, H.2.4 Concurrency

Keywords and phrases linearizability, concurrent data structures, treap

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.07

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 07; pp. 07:1–07:1
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
ICCSW17

Gesture Recognition and Classification using Intelligent Systems

Norah Alnaim[1] and Maysam Abbod[2]

1,2 Department of Electronic and Computer Engineering
Brunel University London
Uxbridge UB8 3PH
UK
1 `Norah.alnaim@brunel.ac.uk`
2 `Maysam.abbod@brunel.ac.uk`

Abstract

Gesture Recognition is defined as non-verbal human motions used as a method of communication in HCI interfaces. In a virtual reality system, gestures can be used to navigate, control, or interact with a computer. Having a person make gestures formed in specific ways to be detected by a device, like a camera, is the foundation of gesture recognition. Finger tracking is an interesting principle which deals with three primary parts of computer vision: segmentation of the finger, detection of finger parts, and tracking of the finger. Fingers are most commonly used in varying gesture recognition systems.

Finger gestures can be detected using any type of camera; keeping in mind that different cameras will yield different resolution qualities. 2-dimensional cameras exhibit the ability to detect most finger motions in a constant surface called 2-D. While the image processes, the system prepares to receive the whole image so that it may be tracked using image processing tools. Artificial intelligence releases many classifiers, each one with the ability to classify data, that rely on its configuration and capabilities. In this work, the aim is to develop a system for finger motion acquisition in 2-D using feature extraction algorithms such as Wavelets transform (WL) and Empirical Mode Decomposition (EMD) plus Artificial Neural Network (ANN) classifier.

WL is an image processing algorithm that performs signal analysis with one signal frequency differing at the end of time. EMD is an innovative technology used in both non-stationary and non-linear data. The primary function of this method is decomposing a signal into Intrinsic Mode Functions consistently through the domain. For classification, ANN is used which is defined as a system that processes information and has structure much like that of the biological nervous system. What is most unique is that this system inhibits an abstract but familiar structure as an information processing system.

In this work, three different finger motions are recorded using an iPhone 6s Plus camera. The gesture classification system is developed for three types of finger gesture recognition. WL and EMD algorithms are used to extract features which are fed to ANN for gesture classification. The classification results of training, validation, and testing mean square error using WL are 5.1312E-4/0.01245/0.0079 respectively, while the classification mean square error using EMD are 1.1035E-11/9.676E-09/2.5616E-9 respectively. Feature extraction execution time, in seconds, for Wavelet Transform is 131 and EMD is 7200. The classification accuracy for training, validation, and testing using WT are 0.9984/0.9909/0.9953 and using EMD are 1.0/1.0/1.0. The results of this experiment clearly identify EMD being a suitable method to extract features from the image but it is time consuming.

1998 ACM Subject Classification H.5.2 User Interfaces, I.2.6 Learning

Keywords and phrases Wavelets, Empirical Model Decomposition, Artificial Neural Network, Gesture Recognition, HCI.

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 08; pp. 08:1–08:2
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
ICCSW17

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.08

KubeNow: a cloud agnostic platform for microservice-oriented applications*

Marco Capuccini[1,2], Anders Larsson[3], Salman Toor[1], and Ola Spjuth[2]

1 **Department of Information Technology, Uppsala University**
`marco.capuccini@it.uu.se`
`salman.toor@it.uu.se`

2 **Department of Pharmaceutical Biosciences, Uppsala University**
`marco.capuccini@farmbio.uu.se`
`ola.spjuth@farmbio.uu.se`

3 **Department of Cell and Molecular Biology, Uppsala University**
`anders.larsson@icm.uu.se`

Abstract

KubeNow is a platform for rapid and continuous deployment of microservice-based applications over cloud infrastructure. Within the field of software engineering, the microservice-based architecture is a methodology in which complex applications are divided into smaller, more narrow services. These services are independently deployable and compatible with each other like building blocks. These blocks can be combined in multiple ways, according to specific use cases. Microservices are designed around a few concepts: they offer a minimal and complete set of features, they are portable and platform independent, they are accessible through language agnostic APIs and they are encouraged to use standard data formats. These characteristics promote separation of concerns, isolation and interoperability, while coupling nicely with test-driven development. Among many others, some well-known companies that build their software around microservices are: Google, Amazon, PayPal Holdings Inc. and Netflix [?].

Cloud computing is a new technology trend that enables the allocation of virtual infrastructure on demand, giving place to a new business model where organizations can purchase resources with a pay-per-use pricing arrangement [?]. Microservices in cloud environments can help to build scalable and resilient applications, with the goal of maximing resource usage and reducing costs. At the time of writing, Docker and Kubernetes are the most broadly adopted container engine and container orchestration framework [?, ?]. Even though these software tools ease microcroservices operations considerably, their setup and configuration is still complex, tedious and time consuming. When allocating cloud resources on demand this becomes a critical issue, since applications need to be continuously deployed and scaled, possibly over different cloud providers, to minimize infrastructure costs. This new challenging way of provisioning infrastructure was the main motivation for the development of KubeNow.

KubeNow provides the means to rapidly deploy fully configured clusters, automating Docker and Kubernetes configuration, while providing a mechanism for the application layer setup. We designed KubeNow using the Infrastructure as Code (IaC) paradigm, meaning that the virtual resources and the provisioning process are defined as machine-readable language. A natural consequence of this choice is that KubeNow is immutable and repeatable over different cloud providers, being cloud agnostic in this sense. In addition, IaC enables infrastructure version control and collaborative development.

KubeNow has been adopted by the PhenoMeNal H2020 consortium as the platform used to launch on demand Cloud Research Environments (CRE) [?]. The PhenoMeNal CRE allows

* This work was supported by the PhenoMeNal H2020 consortium.

2017 Imperial College Computing Student Workshop (ICCSW 2017).
Editors: Fergus Leahy and Juliana Franco; Article No. 9; pp. 9:1–9:2
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
ICCSW17

for running reproducible large-scale medical metabolomics analysis. In addition, we are currently developing additional software layers for large-scale analysis on top of KubeNow including: Apache Spark [?], Pachyderm [?] and Slurm [?]. KubeNow supports Amazon Web Services [?], Google Compute Engine [?] and OpenStack [?]. The software is generally applicable and publicly available as open source on GitHub [?].

1998 ACM Subject Classification D.2.11 Software Architectures

Keywords and phrases Microservices; Cloud computing; Infrastructure as Code; Docker; Kubernetes

Digital Object Identifier 10.4230/OASIcs.ICCSW.2017.9