From DevOps to DataOps: Cloud based Software Development and Deployment

Yuri Demchenko
Complex Cyber Infrastructure Research Group
University of Amsterdam
Amsterdam, The Netherlands
e-mail: y.demchenko@uva.nl

***Abstract*— This paper provides an analysis of trends in infrastructure tools to support modern data driven applications development, deployment, and operation to support the whole data lifecycle that are referred to as DataOps or MLOps. These new applications of the popular DevOps methods to the Data Science and Analytics, Machine Learning, and Artificial Intelligence technology domains that require the support of both application development and data lifecycle management to support an iterative process of the data analytics and machine learning development processes and ensure data lineage or provenance. Recent publications also stress the importance of consistent data management for industrial and technological data. The paper provides an overview of current trends in DevOps use by industry, business, and research. It also provides examples of the cloud based platforms and tools for DevOps and DataOps and their use for different project contexts. The presented research and experience is an outcome of the development of the curricula and course on the DevOps and Cloud based Software Development for Computer Science and Software Engineering masters. This includes the definition of the Body of Knowledge in the DevOps for Software Engineering (DevOpsSE BoK) that defines a set of knowledge areas required for the SE professionals to work efficiently as DevOps engineer or application developer. The DevOpsSE-BoK provides a basis for defining required professional competences and skills and allows consistent curricula structuring and profiling. The paper shares the experience of teaching the DevOps and Cloud based Software course at the University of Amsterdam.**

Keywords- DevOps and Cloud Based Software Engineering; DevOps Software Engieering Body of Knowledge; DataOps and Data Science Projects Managementn; Education; Instructional methodology; Project Based Learning.

I. INTRODUCTION

Demand for continuous development and improvement of the business services and applications motivates organisations to adopt agile services and applications development models to respond to market demand and technology change fast. Adoption of the DevOps methods and technologies allows organisations to incorporate the principles of Continuous Development – Continuous Integration – Continuous Delivery in their IT infrastructure and customer facing services which are foundation principles of DevOps [1, 2].

The DevOps and Agile software development practice is significantly accelerated with the development of cloud services, network services programmability and general convergence of infrastructure services based on cloud native services programmability. And provisioning automation. [3]. Clouds provide a native platform for the Infrastructure as Code (IaCode) model used in DevOps for integrating, deploying, managing cloud based services. Existing cloud platforms (both public such as Amazon Web Services (AWS) and Microsoft Azure, and Open Source such as OpenStack or CloudStack) provide rich functionality for applications deployment automation and cloud resources monitoring and cost management. Clouds provide a native environment for the development and operation of the Big Data applications providing a facility for both data storage and data processing also supporting instruments for data versioning and data lineage that is considered critical for Explainable Artificial Intelligence (XAI) applications [4].

This paper presents an overview of the DevOps technologies and their use in software/applications development that is extended with the analysis of the DevOps process for data driven projects, often referred to as DataOps or MLOps. This allowed to improve the definition of the DevOps Body of Knowledge for Software Engineering (DevOpsSE-BoK), that was initially defined in the authors' previous paper [5], and suggest extensions for the Data Science and Data Analytics applications. The proposed DevOpsSE-BoK can be used for DevOps related courses development, organisational roles profiling and competences assessment. The DevOpsSE-BoK defines a set of Knowledge Areas and Knowledge Units required for the Software Engineering professionals to work efficiently as DevOps engineer or application developer. The proposed DevOpsSE-BoK efficiently uses and extends the IEEE SWEBOK [6] and ACM/IEEE Computer Science BoK [7].

The paper also describes the experience of teaching DevOps and Cloud based Software Engineering course at the University of Amsterdam. The example of the curriculum design presents a complete course on DevOps for Computer Science and Software Engineering masters. The presented model curriculum defines key learning units that can be implemented in a form of lectures and practices or labs. Practical course configuration and content can be composed of the basic modules to target specific learners' groups or utilise existing resources.

The remainder of the paper is organized as follows. Section II explains the growing adoption of the DevOps model in developing and managing modern software and applications analysing also technology drivers behind this process. Section III provides an overview of existing DevOps definitions, which are given by DevOps practitioners and experts. Section IV analyses different aspects of the Data Science projects development, including Data Science process models. Section V

provides an overview and analyses existing education and training courses and programs. Section VI presents the definition of the DevOps and Software Engineering Body of Knowledge and proposed extensions for DataOps processes. Section VII briefly describes the experience in teaching the DevOps and Cloud based Software Engineering course at the University of Amsterdam. The paper concludes with the summary and future developments in section VIII.

II. DevOps Adoption by Industry and Business

A. DevOps as a Culture for Agile Organisations

DevOps became a major technology, approach and culture in applications and services development for modern agile data-driven companies, both information technology and business oriented. It is essential that DevOps implies the adoption of the DevOps culture by organisations that aligns people, processes, and tools toward a more unified customer focus [8].

Modern technologies, particularly cloud and Big Data analytics, embrace DevOps methods. DevOps is often referred as the model for quick adoption of emerging technologies and subsequent integration of them into production.

DevOps proposes a new emerging software and applications development model that realises the continuous integration (CI), continuous delivery (CD) and continuous improvement of the services and applications, and it is deeply based on the cloud environment, on demand services deployment, and highly granular services monitoring currently available on the major cloud platforms. Organisations such as Netflix, Facebook, and Google are revolutionising their IT infrastructure by using DevOps and agile models.

DevOps culture is strongly supported by the professional community of developers. It is worth to mention the DevOps Manifesto that has been created in 2009 and is highly respected by the DevOps community [9].

DataOps and MLOps [10, 11] are other domains where the general DevOps model finds its application and development owing to the fact that Data Science and Data Analytics process involves iterative model development and continuous improvement after the application is moved to production.

Continuous search by agile organisations to increase the efficiency of their development and innovation process goes in the direction of improving tools and environment for the DevOps process and finding a new organisational model to increase team work and concentrate on achieving market goals. In this respect such organisational model as Holacracy worth mentioning [12]. It is experimented by many startups and fast growing companies that capitalise on the strong growth spirit in such organisations. The Holacracy is compliant to DevOps and Agile methods but focuses on individual and team motivational factors that are very important in DevOps and agile organisations. However, Holacracy is considered a controversial model as opposing individual leadership [13].

B. Technologies and Trends affecting DevOps

When assessing current and future development trends for DevOps and DataOps the following technology trends can be taken into account:

- Continuous development of the infrastructure automation technologies and tools (virtualization, microservices, composability, code libraries, API) which are increasingly offered as a part of the cloud platform
- The growing acceptance of the cloud native technologies for end-to-end infrastructure services provisioning what is one of the major components of the 5G technologies that features end-to-end network slicing provisioned on demand [14].
- Microservices and containerization, which adoption is growing from applications virtualisation to infrastructure services and network functions virtualization, which allows development and easy move from to production stage [15].
- Recent DevOps and CI/CD that trends to become integrated into the change management process to ensure the continuous evolution of the target system [16]
- Semantic Data Lakes as integrated data storage and data analytics platform that allow data storage at all stages of the data lifecycle, which examples are Azure Data Lake gen2 that offers storage for heterogeneous data and provide integrated data analytics [17, 18], DataBricks Delta Lake that is specifically adopted to streaming data analytics with the Spark and DataBricks platform [19].
- Infrastructure related security technologies that propose solutions for on-demand infrastructure services provisioning such as trust bootstrapping, data lineage, creating secure trusted virtual execution environment for data and for cooperating groups of developers.

Another important aspect that will affect future data driven applications development and consequently must be reflected in the DevOps and DataOps pipeline models is global data availability and access over the network that requires consistent security and privacy compliance to be addressed during the design process, in ideal case realizing security and privacy by design approach. For a cooperative group of researchers, including wide public access to scientific data, the data sharing and access policies must be compliant with a number of regulations and European GDPR (General Data Protection Regulation) in particular [20, 21].

III. DevOps Definition by Professional Community

A. DevOps according to DevOps experts and practitioners

DevOps term and concept have been actively used since 2015 [22] and 2016 [23] when it was introduced as an extension of the Agile software development for modern Internet companies that besides agile software development need also continuously improve and adapt their services to continuously changing technology and market environment.

Formal and "live" DevOps definition is given at the community maintained website [1]: "DevOps is the practice of operations and development engineers participating together through the entire service lifecycle; from the design and development process all the way to production support. DevOps is also characterized by operations staff making use of many of the same techniques as developers for their systems work."

The following DevOps features are admitted [1]:

- DevOps has strong affinities with Agile and Lean approaches.

- DevOps can be interpreted as an outgrowth of Agile software development and prescribes close collaboration of customers, product management, developers, and (sometimes) QA to fill in the gaps and rapidly iterate towards a better product
- DevOps is extending Agile principles beyond the boundaries of "the code" to the entire delivered service

B. Relationship to Agile and Continuous Delivery [2]

Agile and DevOps are closely related, however agile model is mostly applied for software development and represents a change in thinking and practice of the software development teams, while DevOps involves the whole lifecycle/process of the solution or service delivery and requires organizational and culture changes [2]:

- The need for DevOps was born from the increasing popularity of agile software development, as that tends to lead to an increased number of releases.
- One goal of DevOps is to establish an environment where releasing more reliable applications, faster and more frequently, can occur.

Continuous delivery and DevOps are similar in their meanings, but they are two different concepts. DevOps has a broader scope and centers around:

- Organizational change: specifically, to support greater collaboration between the various types of workers involved in software delivery:
- Automating the processes in software delivery what allows to make delivery of new application versions faster and more frequently, however preserving service continuity.

IV. DataOps: DevOps for Data Science Projects

A. Data Science and Data Analytics Development Process

The difference between SE and Data Science is in the outcome of the process. SE is developing a software product or service. Data Science is about developing a model that can be used in other applications, in particular in business analytics, Digital Twins, and Artificial Intelligence.

The Data Science and Data Analytics (DSDA) development process is dealing with the data pipelines. The typical DSDA process includes the following stages:

- Collecting data from multiple sources, also blending process or business data with external data such as environmental data or social media data that can be obtained via WebAPI or web scraping
- Working with data including data preparation, cleaning, filtering, and reformatting for modeling needs
- Combing datasets by joining on common attributes, consolidating attributes, build tabular data structure (such as used in popular analytics programming languages R, python, scala)
- Feature engineering, algorithm selection
- Testing before production and validating the model during production, in particular, detecting drift in predictive models
- Implement changes and deploy an updated model.

Automation of building and deployment can be done with CI tools such as Jenkins that manages the development and delivery pipeline in multiple steps, execution environment, and records test results.

In fact, DataOps is related to the Data Science Engineering (DSENG) process that is defined as one of the core Data Science domains, which besides technology, requires specific competences and skills such as defined in the EDISON Data Science Framework (EDSF) [24, 25]. DSENG includes the following tasks in the whole data pipeline:

- DSDA applications deployment as part of the CI/CD process
- Data operations tasks related to infrastructure, scalability, monitoring of availability and performance
- Data Preparation, including staging, cleaning, conforming, delivery
- Data Interfaces: APIs and data query tools.

Data Science and Data Analytics process often uses the standard model description framework/formats what is important to link development and operations stages:

- Predictive Models Markup Language (PMML) that have benefits of transferring a developed model to production, access to coefficients [27].
- Portable Format for Analytics (PFA), an emerging standard for statistical models and data transformation engines to ease portability across systems with algorithmic flexibility by defining composable models, pre-processing, and post-processing functions that can be built into complex workflows [28].

B. Applying DevOps to Data Science Projects

Using the DevOps process and environment allows speed up the Data Science applications move to productions, including addressing such aspects as security, access control, performance and log monitoring, backups, scalability, which are not directly addressed in the Data Science projects.

Applying Operations processes and tools will help to keep DSDA models accurate and up-to-date as production data evolve.

The Data Science project development will also benefit from adopting version control platforms such as Github [26], which is a backbone of the DevOps process.

When considering DevOps principles and research methods, we can notice their general similarity. The DevOps process is well aligned with the Research Methods and research cycle that includes such stages as hypothesis formulation, experiment design, data collection, data analysis, model building, evaluation and model improvement. It is essential that at all stages data are documented, stored and linked, ensuring their provenance and re-usability.

DataOps as DevOps continuous improvement model applied to the Data Science workflow/dataflow process naturally supports similar models in Data Science and Data Analytics such as widely accepted CRISP-DM model (Cross Industry Standard Data Mining) process [28] and ASUM, Analytics Solutions Unified Method [29] discussed in details below.

C. Data Analytics and Machine Learning Process Models

The following presents the reference and a short description of the popular data analysis process models.

1) CRISP-DM, CRoss-Industry Standard Process for Data Mining [28]

The CRISP-DM was the first model to formalise the data mining process and its relevance to the business processes and intended actionable outcome. The following 6 sequential phases are defined, which can also be organized into iterative/continuous development and improvement cycle:

Business understanding: Understanding business processes, problem definition, required outcome.

Data understanding: Obtaining available data, data inspection, data preparation, data observation and initial hypothesis formulation.

Data preparation: Prepare datasets for analysis by extracting necessary data from the raw data.

Modeling: Create one or several data models, select corresponding techniques, possibly test and experiment with several to assess performance.

Evaluation: Ensure that the solution fulfills the business objectives established in the first phase, and to obtain this certainty, a deep evaluation of the model must be performed during this phase. It is evaluated if the knowledge obtained with the model has the desired value for the customer.

Deployment: Solution deployment in a production environment for the customer, providing the results in an easy understandable form to the customer.

2) ASUM, Analytics Solutions Unified Method (IBM)[29]

ASUM is a hybrid of the agile and traditional process model. It has five (5) sequential phases and a set of processes to manage and monitor the progress and maintenance of the project:

Analyze. Requirements specified and agreed; contract or services agreement is signed.

Design. Define all components of the solution and their relationships and dependencies, identify necessary resources.

Configure and Build. The solution is developed, all components are integrated and configured.

Deploy. Create a plan to run and maintain the developed solution, including configuration management and migration plan if necessary.

Operate and Optimize. The solution is operational is monitoring data are collected and maintained.

3) TDSP, Team Data Science Process (Microsoft) [30]

TDSP is an agile and iterative process model. It has five sequential phases: Business Understanding, Data acquisition and understanding, Modeling, Deployment, Customer acceptance. TDSP framework document provides valuable information for Data Science applications developers for planning and managing the whole process from development to deployment and operation.

V. DevOps Competences, Education and Training

A. Developments by DevOps Professional Community

DevOps community is very committed and features a number of expert and community supported websites, blogs and online forums. We can admit two valuable initiatives and services that offer training and certification on DevOps and also contribute to the definitions of the DevOps competences and Body of Knowledge:

- DevOps Institute: The Association for DevOps Professionals [31]
- DevOps Agile Skills Association (DASA) [32]

The DASA DevOps Fundamentals course provides an example of well developed course that can be a starting point for anyone involved in an Agile and/or DevOps team. The course is also supported by the DASA competence model and DevOps certifications scheme. The DASA DevOps Competence model includes

- 8 knowledge areas: Architecture and Design; Business Value Optimisation; Business Analysis; Test Specification; Programming; Continuous Delivery; Security, Risk, Compliance; Infrastructure Engineering
- 4 skills: Courage, Teambuilding, DevOps Leadership, Continuous Improvement.

The DASA Qualification scheme includes 5 levels of proficiency: 1 - novice, 2 - competent, 3 - proficient, 4 - expert, 5 - master; where the DASA course fits into level 2 – competent. The course entry requirements include basic familiarity with Agile, Scrum, Lean, and ITSM principles.

B. DevOps Training for Practitioners

There is a strong demand for DevOps competences and skills, most of which are regarded as a necessary part of the modern developer profile. However there are many positions and job profiles that explicitly DevOps and Agile in their title. In this respect the Scrum Master is quite distinct profession that plays an important role in the organisational structure.

It is currently satisfied by professional training and certification provided by professional training organisations and big technology companies and software houses such as Microsoft Learn [33, 34] or AWS DevOps [35, 36]. There is a limited offer from Coursera [37] and Udacity [38] that are focused primarily on some specific topics or platforms and don't provide consistent materials for the academic degree. Courses on DevOps and related software engineering and cloud technologies and tools are well presented at Linkedin Learn [39], which can be recommended for students, practitioners, and teaching staff.

DevOps training and DevOps setup are popular topics for consulting services and organisational training/improvement/coaching. It is provided by many global and local consulting companies. It is also a popular business for small companies and individual consultants.

Online education company Edureka offers the extensive DevOps Engineer Masters Program with 10 Courses of a total duration of more than 200 hrs of interactive learning that includes Security Administration, Shell Scripting, Virtualization, User Administration, OOPs Concepts, Script Debugging, CRUD Operations, Python IDE, Git, Chef, and completed with the Capstone Project that practical business case to apply all learning competences [40]

Intellipaat DevOps Architect Master Course has been developed in collaboration with IBM [41]. The course is practically oriented and includes various tools like Docker, AWS SysOps, Git, Kubernetes, Puppet, Chef, Ansible, Python, and hands-on projects included in the program. The course is finished with certification.

C. DevOps Certification

Well defined and industry accepted certification programs play an important role in establishing a new profession. There

are a number of certification and training DevOps master programmes that require master in SE and provide additional training on specific DevOps competences, technologies, platforms and tools. Positively that some DevOps certification programs are supported by well-defined competences and skills specification.

An example of well organised certification program is the EXIN DevOps Master [42], an advanced-level certification program for IT professionals that are dealing with software development that tests candidates on their ability to develop and maintain sustainable work practices. It enables them to introduce and promote DevOps in their organization in order to better manage application and service life cycles whilst facilitating collaborative teamwork. Main subjects include: DevOps adoption, Planning, requirements, and design, Development and deployment, Operation and scaling, End-of-life.

D. University programmes on DevOps

There are not many academic programs that directly offer masters in DevOps as part of Software Engineering or Computer Science programs.

Few universities offer DevOps courses in their Computer Science or Software Engineering programmes adding DevOps qualification to their graduates. Here we give just two examples described below.

The John Hopkins Engineering part-time master program "DevOps Software Development" [43] is offered since 2016. It provides classical DevOps courses, however it lacks cloud technologies and tools orientation what is essential in modern DevOps. The universities still to find a proper form of teaching, such as practically oriented courses with software engineering focus, where the project based learning and group projects development can help mastering both the agile development process and team based development during the group project.

Maryville University provides an online programme for DevOps engineers that is based on the Computer Science Bachelor and offered as a specialization of the master in software development [44]. The programme is focused on the following DevOps Engineer competences and functions besides writing code:

- Automate business processes to improve operations, implementing changes requested by customers or managerial staff, deploying updates and fixes, and in some cases, providing technical support.
- Write scripts and automation using various programming languages, such as Python, Java, and Ruby
- Build out and manage IT infrastructures and provide support for their users.
- Perform root-cause analysis of production errors
- Supervise and collaborate with the development, security, and operations (DevSecOps) team to improve a company's infrastructure security.

VI. DevOps and Software Engineering Body of Knowledge

Body of Knowledge (BoK) refers to several domains or operational categories into which the domain specific theory and practices break down. It is like a blueprint or framework in which students and practitioners can see how things fit together.

Definition of the Body of Knowledge is a cornerstone element in the proper definition of the new profession or qualification and a basis for establishing transferable competences, skills, and certification.

A typical Body of Knowledge defines and a set of Knowledge Areas (KA) that can grouped into the Knowledge Area Groups (KAG) and break down into Knowledge Units (KU). The proposed DevOpsSE-BoK is linked to existing industry and academia accepted IEEE/ACM Computer Science BoK (CS-BoK) [6], IEEE Software Engineering BoK (SWEBOK) [7], and Data Science Engineering BoK (DS-BoK) [45] developed by the authors.

A. DevOpsSE BoK

The following main BoK elements are defined for DevOpsSE BoK (updated and extended based on the initial version proposed in [5]):

1. DevOps fundamentals, Continuous Integration, Continuous Delivery, Continuous Testing. Relation to other agile development technologies Lean, CAMS, and ITSM.
2. Organisational impact of DevOps, Digital Transformation and DevOps; DevOps Team structure and operation, Leadership, collaboration, and structured problem solving; success factors and key performance indicators. DevOps and organisational change management. Project life cycle and management approaches.
3. Agile software development: Scrum, Kanban, Kaizen; Agile Scrum process and team management; Role of multi-disciplinary feature teams. Agile Manifesto.
4. DevOps Tools and Processes: CI/CD pipeline; DevOps Toolchain; Coding, versioning, collaboration, and testing. Versioning and team based development, Git, Automated testing.
5. DevOps Practices and Platforms: Software packaging, Containerization. Container technologies, Kubernetes, Tools
6. Cloud Computing Architectures, service and deployment models, Cloud IaaS and Infrastructure as Code, Cloud economics
7. Cloud powered software development: Cloud example and tools (AWS and Microsoft Azure)
8. Cloud monitoring tools and Operate for Development concept.
9. Cloud Automation Overview: Cloud based tools (e.g. AWS CloudFormation, Azure ARM); multicloud tools Chef, Puppet, Ansible, Terraform, others.
10. Popular cloud automation tools: Chef, Puppet, Ansible (depending on available platform)
11. Microservices, Containerisation, Docker, Kubernetes
12. Cloud Security Architecture and Models, Cloud compliance, CSA Consensus Assessment Initiative Questionnaire and Certification.
13. DevSecOps; Secure software development lifecycle; cloud based tools for secure Software development.

B. Extensions for DataOps and MLOps

The following BoK knowledge topics will cover specific methods, processes, and technologies to support Data Science and Data Analytics projects:

- Data Science and Data Analytics processes, CRISP-DM, ASUM, data lifecycle management
- Data modeling, Data lineage (data provenance), data management principles, and data quality assurance

- Platforms and tools for data pipeline automation (such as AWS Sage or Azure MLOps, supported with Data Lakes)

Data Scientists adopting DataOps methods should come with the necessary knowledge and competences in Data Science and Analytics from their Data Science education or training; Data Science professional and transversal skills are also important [46].

VII. Implementation: DevOps and Cloud based Software Engineering at the University of Amsterdam

A. General information and motivation

The presented in this paper analysis and approach to teaching DevOps competences is based on the experience of development and implementation of the DevOps and Cloud based Software Engineering course at the University of Amsterdam [47]. The course is taught as a part of the Software Engineering Master and counts for 6 credits. It runs 2 months and has 8 contact hours per week, of which 4 hours are lectures and 4 hours practice and labs.

The DevOps course is based on the author's previous experience in developing effective curricula for Cloud Computing [48] and Big Data [49] and implements the methodology and approach to teaching technology related courses in conditions of fast changing technologies.

This course teaches the basic concepts and technologies of DevOps, including its philosophy, workflow, monitoring methods and tools. The course gathers the recent development in cloud based services and applications development, deployment, and operations based on the best industry practices. Students have the opportunity to apply these concepts to understand how they can be best implemented to automate development, test, and release practices.

The course also includes the final projects to design cloud based application or service to support a selected business process or scientific workflow using one of the available cloud platforms: local cloud testbed, Amazon AWS, or Microsoft Azure. The project groups consist of 4-5 members and are required to function as small Agile teams using the Agile Scrum model (optionally Kanban or Kaizen, if motivated by the project team).

B. Learning outcomes

The DevOps course defines the following Intended Learning Outcomes (ILO)

- Understand how to build cloud based applications and use cloud automation tools in the DevOps process.
- Understand the theoretical background of DevOps and cloud automation techniques.
- Demonstrate the application of these techniques applied to a range of applications and software development scenarios/processes.
- Understand functionalities offered by the popular DevOps and cloud automation tools such as Chef, Puppet, Ansible, SlipStream
- Have hands on experience with one of such tools as a result of project completion.

The not formal course goal is to prepare the student to be ready to take one of DevOps related certifications exams.

The lab assignments include the basic hands-on tasks to learn the cloud platforms services and existing DevOps tools and services, including cloud monitoring, cloud automation. The practice can be done on one of two major cloud platforms by Amazon AWS and Microsoft Azure, what together provides a basis for the practical project realization. Selection of public clouds is essential for the students' future workplace alignment as majority of software development companies are using public cloud platforms.

C. Lectures

The course includes two general types of lectures. The first type is technology introduction lectures that provide an extended overview of the basic technologies used in the modern DevOps environment and practice, including DevOps concept and models, cloud technologies and services, CI/CD tools, cloud automation and monitoring tools, security, access control, compliance. The second type of lectures is more focused on use cases, practices, and case studies, where real life projects and development are used as examples.

Technology oriented lectures are provided at the beginning of the course and typically supported by practice and labs. Many technology specific topics are recommended for self-study as part of practical assignments. The students are required to submit reports on practice, which contribute to the final grade on the course. Lectures on research and industry practices and case studies are more interactive and require students' involvement in the discussion.

D. Practice

Practice in this course is organised by weekly topics to allow students sufficient time to master the technology and do reporting. The presented below topics are selected to gradually provide the necessary knowledge and experience for the students to start the project development, which is the main reporting material of the course.

Practice 01: Getting started with Cloud: AWS console and CLI, VM deployment, Accounts and security configuration.

Practice 02: Continuous Integration and Continuous Deployment process and tools on the example REST Web Services project

Practice 03: AWS/cloud Monitoring tools: Configure monitoring for deployed service.

Practice 04: AWS Cloud Formation to create a simple template for deployed service: Adding AWS monitoring to CloudFormation template

Practice 05: Ansible with AWS: Create Ansible playbook

Depending on available resources, platforms and local preference, the practice can include other cloud based DevOps platform such as Azure DevOps, or cloud automation tools such as Chef or Puppet.

E. Self-study and seminars

The course is built in such a way as to facilitate the students deeper study/research into the subject of DevOps and related technologies and tools, improve their skill in working with professional and scientific literature, facilitate collaborative learning and actively use their knowledge in the discussion.

To achieve this the course employs two types of self-study activities: (1) discussion of the selected papers in the classroom which is organised as a seminar at the second part of course when the student acquired sufficient knowledge to critically assess the papers content; and (2) literature based research on the topics related to the group project, the report on the literature study is included in the final project report, and contribute to the final course grade. The requirements to the literature study are described in the course guidelines.

F. Project development

The final project, including literature study, is the key components to achieve higher level of knowledge in DevOps and Software Engineering. The project topics are selected by the students themselves and course guidelines specify criteria to the project content and complexity criteria. In particular, the project should demonstrate the CI/CD process and provide facility for the service or application monitoring. Such topics as, for example, Content Management System (CMS) typically used for website content management is not sufficient.

The project development is also regarded as an activity for the students to learn Agile and Scrum/Kanban in practice. The students are also provided with the selection of topics from the Agile Practice guide published by the Project Management Institute [51]

VIII. Conclusion, Lessons Learned and Future Developments

This paper presents the authors experience on developing a new course on the DevOps and Software Engineering as a part of the Computer Science or Software Engineering masters [19] that is further extended with elements required for the Data Science projects that are also related to Machine Learning and Artificial Intelligence project that all have a similar goal in developing the actionable model that can be used in industrial processes, for business operations or for decision making in different domains. The paper is built on the authors' previous paper [5] and extended with the recent DevOps and DataOps trends, new technology development, and provides an example of the existing training and education programs that provide education and training for DevOps engineers. The paper also provides examples of certification programs and courses.

The described DevOps and Cloud based Software Engineering course has been taught in 2018/2019 (58 students) and 2019/2020 (39 students) academic years and provided an important practical experience of teaching essential technologies in the university environment to the students that in their majority have a high level of practical and technical knowledge and many of them have already a job. This category of students respects freedom in fulfilling course requirements and assignments. That is why the project based learning model was selected as the main educational model for this course. This also provided a working solution for combining research elements while keeping the important focus on the essential technologies mastering where the students have full freedom in selecting their projects, technology and tools, work in groups while complying with the complexity and reporting requirements for the project.

The successful solution of the above mentioned problem for the DevOps course was essentially based on the authors' previous experience in developing effective curricula for Cloud Computing [48] and Big Data [49]. At the same time, setting up the DevOps course provided an experience for the author and the whole course team that was used in other course development on the Big Data Technologies for Data Analytics that included such topics/lectures related to DataOps as DSDA process management, CI/CD, cloud services monitoring and agile projects management.

The important outcome of the presented research and development is the validation of the proposed DevOpsSE Body of Knowledge and suggested extensions for DataOps related knowledge topics that should help future course developers related to Software Engineering and Data Science programs.

Acknowledgment

This work is supported by EU funded projects SLICES-DS (Grant Agreement No. 951850), FAIRsFAIR (Grant Agreement No. 831558), GN4-3 (Grant Agreement No. 856726). The DevOps and DataOps methodology is based on the methodology developed in the EU funded project EDISON (2015-2017, grant 675419)

The successful teaching of both DevOps and BDIT4DA courses became possible thanks to the AWS Educate support program for educators and students that provided sufficient resources for the practical and lab assignments in the courses.

References

[1] What is DevOps? The Agile Admin [online] https://theagileadmin.com/what-is-devops/

[2] DevOps, Wikipedia [online] https://en.wikipedia.org/wiki/DevOps

[3] Louis Columbus, Jinesh Varia, Building Powerful Web Applications in the AWS Cloud [online] https://d1.awsstatic.com/whitepapers/AWS_Cloud_Best_Practices.pdf

[4] Alejandro Barredo Arrietaa et al, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Information Fusion Volume 58, June 2020, Pages 82-115 https://doi.org/10.1016/j.inffus.2019.12.012

[5] Yuri Demchenko, Zhiming Zhao, Spiros Koulouzis, Jayachander Surbiryala, Zeshun Shi, Xiaofeng Liao and Jelena Gordiyenko, Teaching DevOps and Cloud based Software Engineering in University Curricula, Proc. 5th IEEE STC CC Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science (DTW2019), part of the eScience 2019 Conference, September 24 – 27, 2019, San Diego, California, USA

[6] Software Engineering Body of Knowledge (SWEBOK) [online] https://www.computer.org/web/swebok/v3

[7] ACM and IEEE Computer Science Curricula 2013 (CS2013) [online] http://dx.doi.org/10.1145/2534860

[8] What is DevOps Culture? How DevOps culture helps align people, processes, and tools toward a more unified customer focus. Blog by Tom Hall [online] https://www.atlassian.com/devops/what-is-devops/devops-culture

[9] The modern DevOps Manifesto [online] https://medium.com/ibm-garage/the-modern-devops-manifesto-f06c82964722

[10] MLOps: Methods and Tools of DevOps for Machine Learning, 23 Jul, 2020 [online] https://www.altexsoft.com/blog/mlops-methods-tools/

[11] Design a machine learning operations (MLOps) framework to upscale an Azure Machine Learning lifecycle, Microsoft Azure [online] https://docs.microsoft.com/bs-latn-

ba/azure/architecture/example-scenario/mlops/mlops-technical-paper

[12] Holacracy Constitution https://www.holacracy.org/constitution

[13] Beyond the Holacracy Hype, By Ethan Bernstein, John Bunch, Niko Canner, and Michael Lee, From the Magazine (July–August 2016) [Online] https://hbr.org/2016/07/beyond-the-holacracy-hype)

[14] 5G and The Cloud, 5G Americas White Paper, December 2019 [online] https://www.5gamericas.org/5g-and-the-cloud/

[15] Shihabur Rahman Chowdhury; Mohammad A. Salahuddin; Noura Limam; Raouf Boutaba, et al, Re-Architecting NFV Ecosystem with Microservices: State of the Art and Research Challenges, Published in: IEEE Network (Volume: 33, Issue: 3, May/June 2019), DOI: 10.1109/MNET.2019.1800082

[16] DevOps & Change Management In The Enterprise World [online] https://clearbridgemobile.com/devops-change-management-in-the-enterprise-world/

[17] Ten Reasons to Dive into the Smart (Semantic) Data LakeCambridge Semantics, 2017 [online] https://blog.cambridgesemantics.com/ten-reasons-to-dive-into-the-smart-semantic-data-lake

[18] Building your Data Lake on Azure Data Lake Storage gen2, Blog by Nicholas Hurt, March 2020 [onine] https://medium.com/@Nicholas_Hurt/building-your-data-lake-on-adls-gen2-3f196fc6b430

[19] Delta Lake and Dlta Engine guide, DataBricks [online] https://docs.databricks.com/delta/index.html

[20] Yuri Demchenko, Fatih Turkmen, Mathias Slawik, Defining Intercloud Security Framework and Architecture Components for Multi-Cloud Data Intensive Applications. Sixth IEEE International Workshop on Cloud Computing Interclouds, Multiclouds, Federations, and Interoperability (Intercloud 2017), In Proc. 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. Madrid, Spain, May 14-17, 2017

[21] GDPR Compliance, GDPR.EU [online] https://gdpr.eu/

[22] Ernest Mueller, DevOps State of the Union 2015, Slideshare [online] https://www.slideshare.net/mxyzplk/devops-state-ofthe-union-2015

[23] Jason Bloomberg, The Agile Architecture Revolution: How Cloud Computing, REST-Based SOA, and Mobile Computing Are Changing Enterprise IT [online] https://intellyx.com/author/jbloomberg/

[24] The Data Science Framework, A View from the EDISON Project, Editors Juan J. Cuadrado-Gallego, Yuri Demchenko, Springer Nature Switzerland AG 2020, ISBN 978-3-030-51022-0, ISBN 978-3-030-51023-7

[25] EDISON Data Science Framework (EDSF). [online] Available at https://github.com/EDISONcommunity/EDSF

[26] Github version control system [online] https://guides.github.com/introduction/git-handbook/

[27] Predictive Models Markup Language, PMML 4.4 - General Structure PMML [online] http://dmg.org/pmml/v4-4/GeneralStructure.html

[28] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0, Cross Industry Standard Process for Data Mining Reference Model, http://edison... (2000).

[29] IBM, ASUM, Analytics solutions unified method [online] https://www01.ibm.com/marketing/iwm/iwm/web/pick.do?source=swerpb a- basimext&lang=en US

[30] TSDP, Team Data Science Process, https://docs.microsoft.com/en-us/azure/machinelearning/team-data-science-process/overview

[31] DevOps Institute: The Association for DevOps Professionals [online] https://devopsinstitute.com/

[32] DevOps Agile Skills Association (DASA) [online] https://www.devopsagileskills.org/certifications/dasa-devops-fundamentals/

[33] Microsoft Learn [online] https://docs.microsoft.com/en-gb/learn/

[34] Azure DevOps [online] https://azure.microsoft.com/en-gb/services/devops/

[35] AWS Training [online] https://www.aws.training/

[36] AWS DevOps [online] https://aws.amazon.com/devops/

[37] DevOps at Coursera [online] https://www.coursera.org/courses?query=devops

[38] DevOps at Udacity [online] https://eu.udacity.com/courses/all?keyword=devops

[39] Linkedin Learning [online] https://www.linkedin.com/learning/

[40] Edureka DevOps Master [online] https://www.edureka.co/masters-program/devops-engineer-training

[41] DevOps Architect Master's Training Program [online] https://intellipaat.com/devops-architect-masters-training-program/

[42] EXIN DevOps Master training and certification program https://www.exin.com/certifications/exin-devops-master-exam

[43] John Hopkins Engineering part-time master program 605.609 - DevOps Software Development [online] https://ep.jhu.edu/programs-and-courses/605.609-devops-software-development

[44] How to Become a DevOps Engineer, Maryville University [online] https://online.maryville.edu/online-masters-degrees/software-development/careers/how-to-become-a-devops-engineer/

[45] Data Science Body of Knowledge (DS-BoK). EDSF Community Initiative [online] https://github.com/EDISONcommunity/EDSF/tree/master/data-science-body-of-knowledge

[46] Yuri Demchenko, Tomasz Wiktorski, Steve Brewer, Juan José Cuadrado Gallego, EDISON Data Science Framework (EDSF) Extension to Address Transversal Skills required by Emerging Industry 4.0 Transformation, Proc. 5th IEEE STC CC Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science (DTW2019), part of the eScience 2019 Conference, September 24 – 27, 2019, San Diego, California, USA

[47] DevOps and Cloud Based Software, Software Engineering at UvA [online] http://studiegids.uva.nl/xmlpages/page/2018-2019/zoek-vak/vak/63471

[48] Demchenko, Yuri, David Bernstein, Adam Belloum, Ana Oprescu, Tomasz W. Wlodarczyk, Cees de Laat, New Instructional Models for Building Effective Curricula on Cloud Computing Technologies and Engineering. Proc. The 5th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2013), 2-5 December 2013, Bristol, UK.

[49] Demchenko, Yuri, Emanuel Gruengard, Sander Klous, Instructional Model for Building effective Big Data Curricula for Online and Campus Education. 1st IEEE STC CC and RDA Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science, in Proc. The 6th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2014), 15-18 December 2014, Singapore.

[50] Agile Practice guide, Project Management Institute, 2017 ISBN 978-1-62825-199-9