Deployment Coordination for Cross-Functional DevOps Teams

Daniel Sokolowski
sokolowski@cs.tu-darmstadt.de
Technical University of Darmstadt
Darmstadt, Germany

ABSTRACT

Software stability and reliability are the core concerns of DevOps. They are improved by tightening the collaboration between developers and operators in cross-functional teams on the one hand and by automating operations through continuous integration (CI) and infrastructure as code (IaC) on the other hand. Ideally, teams in DevOps are fully independent. Still, their applications often depend on each other in practice, requiring them to coordinate their deployment through centralization or manual coordination.

With this work, we propose and implement the novel IaC solution µIS ([mju:z] “muse”), which automates deployment coordination in a decentralized fashion. µIS is the first approach that is compatible with the DevOps goals as it enables truly independent operations of the DevOps teams. We define our research problem through a questionnaire survey with IT professionals and evaluate the solution by comparing it to other modern IaC approaches, assessing its performance, and applying it to existing IaC programs.

CCS CONCEPTS

• Computer systems organization → Cloud computing
• Software and its engineering → Orchestration languages

KEYWORDS

DevOps, Infrastructure as Code, Resource Orchestration, Cloud

ACM Reference Format:


1 INTRODUCTION

The goal of agile development—improving satisfaction through iterative feedback and higher velocity—has led to DevOps, aiming to reduce the friction between software development and operations. It unites developers, operators, and others in cross-functional teams, improving the collaboration between the professions to increase reliability and stability [20]. These goals are expressed in the software delivery and operational (SDO) performance of the organization, which is measured by Forsgren et al. [8] by the (1) deployment frequency, (2) lead time between development and production of changes, (3) required time to restore service on failure, and (4) the rate of failed changes.

Technically, DevOps drives operations automation through continuous integration (CI) and infrastructure as code (IaC) [19]. The latter leverages code-based techniques for operations, enabling the application of well-developed best practices from software development like versioning, testing, or reviewing. Earlier examples of IaC approaches are Chef [7] and Ansible [3], which allow server configuration through procedural scripts, and Puppet [26] that is declarative, i.e., its scripts describe a target state and the system automatically derives the operations to achieve it, providing better stability and less maintenance [29]. Meanwhile, these systems are often referred to as configuration as code (CaC) because they cannot provision infrastructure, which is required to use modern cloud infrastructure effectively, e.g., serverless functions, containers, databases, and storage. This demand led to declarative IaC solutions like Terraform [11], AWS CloudFormation [34], or Azure Resource Manager (ARM) [18], which use JSON, YAML, or other DSLs for their configuration. Lately, IaC solutions like Pulumi [24] and AWS CDK [33] became available, using general-purpose programming languages like TypeScript, Python, C#, or Go to describe the infrastructure. They allow developers to use already known languages with all their well-developed abstractions and tools.

Ideally, DevOps teams are independent. However, their applications often depend on each other in practice, requiring coordinating the times and order of deployments and changes. With today’s IaC solutions, this coordination requires either (1) centralization, which is against the independence of teams, or (2) manual coordination, e.g., via phone, email, or chat, which contradicts DevOps’ automation paradigm and requires synchronization between the teams, reducing the flexibility, reliability, and stability. In this work, we address deployment coordination for DevOps organizations with cross-functional teams to automate the deployment coordination in a decentralized fashion. Such decentralized automation enables DevOps organizations to be compatible with DevOps goals in the presence of inevitable application dependencies across teams, improving their SDO performance.

To identify the research problem in practice, we perform a questionnaire survey under IT professionals, assessing the existence of dependencies between their applications and whether these dependencies impact the order of deployment and undeployments. To automate the coordination of decentralized deployments, we propose µIS ([mju:z] “muse”), a novel IaC system that treats deployments as continuously running processes—in contrast to the common perception that deployments are one-off tasks. We connect such deployments and implement a protocol for the controlled
sharing and reactive updating of resources, achieving automated coordination across separate deployments and, thus, across independent teams. We plan to compare \( \mu S \) with other IaC solutions and apply it to existing IaC script.

In the following, we present related work (§2) and our IaC solution for cross-functional DevOps teams in §3. Then we describe our evaluation plan (§4) before reporting the already achieved results in §5 and providing evidence for our contribution to knowledge (§6).

2 RELATED WORK

Resource Orchestrators. Weerasiri et al. [42] provide an overview on resource orchestration for the cloud. Following their reference architecture, \( \mu S \) is a rule engine (it performs rule-based operations for decisions) and a policy enforcement engine (it takes decisions based on policies and external signals). Ranjan et al. [31] summarize the programming of resource orchestration operations. Various centralized orchestration solutions for virtualized containers exist, e.g., Kubernetes, Kubernetes Federation, Mesos, and Docker Swarm [6]. DOCMA [12] is an orchestrator for IoT applications that is distributed and decentralized. However, the applications globally define all resources in their scope requiring a centralized view of the system. COPE [16] is a distributed policy enforcement engine for cloud orchestrators, enforcing orchestration policies expressing constraints and service-level agreements (SLAs).

Resource orchestrators create, update and delete resources, which is also required in our IaC approach. \( \mu S \) uses resource orchestrators internally for these tasks. However, the focus of this work is on coordinating dependency fulfillment in a decentralized way. In contrast, resource orchestrators rely on centralized mechanisms to fulfill dependencies if they handle them at all.

Infrastructure as Code. Infrastructure as code [19] uses machine-readable code to configure and provision, i.e., deploy, systems. It enables applying best practices from software engineering to software operations, e.g., version control, static analysis, or code reviews. Various industrial IaC solutions are available (cf. §1) and Rahman et al. [27] performed a mapping study on their discussion in academia. A Terraform based approach for repeatable auto-scaling infrastructure in scientific computing is proposed by Balis et al. [4]. Guerriero et al. [10] performed semi-structured interviews with 44 developers, noticing that maintenance and evolution of IaC require better support and tooling.

Code quality has been studied for CaC systems. Sharma et al. [35] analyzed ~ 5K Puppet projects, identified common code smells, and showed that configuration and design smells often go along. Schwarz et al. [32] generalized these findings because such smells are independent of the used technology. Rahman and Williams [30] conducted an empirical study on defective CaC scripts, identifying source code properties that correlate with defects. They also identify code smells through qualitative studies of Ansible, Chef, and Puppet scripts and propose corresponding static linters [28, 29].

\( \mu S \) is a representative of the youngest generation of IaC solutions: It is declarative and leverages a general-purpose programming language. \( \mu S \) supports various cloud providers and is based on Pulumi. In contrast to all other IaC solutions, \( \mu S \) enables decentralized deployments with automated coordination.

Modeling Languages. Modeling languages specify architectures and behavior, e.g., the OASIS standard TOSCA [21] models cloud applications and their management. It describes topologies as a graph of components with their relationships. Operational behavior is described in management plans using existing workflow modeling languages, e.g., BPEL or BPMN. Bellendorf and Mann [5] give an overview of TOSCA cloud orchestration techniques, extensions, and tools. TOSCA was also applied to DevOps to integrate heterogeneous automation artifacts [43, 44].

Wurster et al. [46] describe the Essential Deployment Metamodel (EDMM), a least denominator metamodel of IaC solutions. TOSCA Light [47] is the EDMM-compatible TOSCA subset that was shown to be deployable with 13 IaC solutions using TOSCA Lightning [45].

Declarative IaC solutions and modeling languages describe system architectures. However, modeling languages also describe operational behavior, which is automatically derived in declarative IaC. The resource graph model of \( \mu S \) is inspired and compatible with the EDMM. Therefore, system descriptions in modeling languages like TOSCA could be automatically converted to \( \mu S \), which then provides the operationalization and runtime to execute these specifications in DevOps settings out-of-the-box. This also enables to use (graphical) modeling tools, e.g., from the TOSCA ecosystem, as description languages for \( \mu S \).

Architecture Description Languages. Architecture Description Languages (ADLs) define the component-level structure of an application. Medvidovic and Taylor [17] define that ADLs specify components, connections between them and their configuration, and they provide tools for development and evolution. ADLs exist on various levels, e.g., ArchJava [2] defines components in Java and ORS [15] treats entire services as components. Terra and Valente [38, 39] propose a DSL to enforce constraints on structural dependencies in object-oriented software.

ADLs can be used to verify that an application’s architecture complies with its specification. Descriptive IaC solutions are similar to ADLs, because they define the system’s architecture as resources (components) and their dependencies (connections). However, ADLs do not provide an executable specification, constructing the system from the specification, which is required for deployments. Moreover, they do not cover mechanisms to coordinate decentralized deployments.

3 AUTOMATING DECENTRALIZED DEPLOYMENT COORDINATION

In descriptive IaC systems, users define a directed acyclic graph (DAG) where each node is a resource, e.g., a database, container, or network ACL entry, and arcs are dependencies between them, typically due to a contained-in or requires relationship [46] between the two resources. These dependencies are transitive and order the deployment, i.e., if resource \( R \) depends on \( S \), \( S \) must be deployed before \( R \), and \( R \) must not be deployed when \( S \) is undeployed.

We use a simple static website with a single index.html page hosted in an AWS S3 bucket as a running example. The deployment description is in Listing 1 and defines Figure 1a; the index must be deployed after and undeployed before the bucket. While this is unrealistically simple, it suffices to showcase our approach.
Listing 1: Centralized deployment description of the website.
11 const bucket = new aws.s3.Bucket('website', { (12 website: { indexDocument: 'index.html' } (13 }) ); // Creates the S3 bucket for the static website (14 const index = new aws.s3.BucketObject('index', { (15 bucket, content, key: 'index.html' (16 }) ); // Saves the index.html page in the website's bucket

(b) Centralized.

(a) Centralized deployment description of the website.
(1) The provider generates the deployment description and a snapshot of the current values of all external signals, e.g., changes in other deployments. The runtime of a \(\mu S\) deployment comprises three components: (1) The interpreter generates the target state from the deployment description and a snapshot of the current values of all external signals, e.g., remote offers. The target state is a snapshot of the resource graph described by the deployment description where all resources depending on unsatisfied wishes are pruned, e.g., if the offer to provide resources to the connected deployment is not available, the index page is absent in the target state of the editor's deployment. (2) The driver reads the current deployment state from persistent storage, deploys, updates, and undeploys resources to achieve the target state, and persists the new deployment state for its next invocation. (3) The reactive engine is the only continuously executing component and records changes of external signals, e.g., when a remote offer changes. It triggers the interpreter on changes, which subsequently runs the driver.

To ensure correct behavior, the connections between deployments and inter-deployment resource dependencies must be explicit. For this we propose three new resource types: RemoteConnection for connections to other deployments, Offer to provide resources or information to a connected deployment, and Wish to access the offer of a connected deployment. Using them, the provider and editor specify their connection (Lines 2.1 and 3.1). The provider offers its bucket (Lines 2.2 to 2.4) to the editor’s deployment in Line 2.5. The editor specifies their expectation of the offer by defining a wish in Line 3.2, allowing to use the offered bucket via wish.offer (Line 3.4). Together Listings 2 and 3 define the resource graph Figure 1b.

3.2 Enabled Use Cases

*Asynchronous deployment across teams.* \(\mu S\) enables teams to start their deployments independently and deploys resources asynchronously once their dependencies are fulfilled. E.g., the editor starts their deployment before the provider. First, the index is not deployed because its dependency is unsatisfied. Once the provider deploys the bucket, \(\mu S\) automatically deploys the index.

Figure 1: Resource graph of the website deployment.

3.1 Decentralized Deployment

In DevOps, ideally, each team operates its resources independently, including deployment. In the website example, the provider could be responsible for the bucket and the editor for the page in it. To ensure that the index page is only deployed when the bucket is, both teams need to manually coordinate whenever the bucket is deployed, updated or undeployed. This decreases the flexibility of the teams and wastes time due to synchronization. To decouple the deployment times, we propose to treat deployments not as one-off tasks—as all common IaC systems do—but as continuously running processes, updating the deployed resources reactively based on the deployment description and external signals, e.g., changes in other deployments. The editor can start their deployment independently and run it continuously. Whenever the provider starts or updates their deployment, the editor’s deployment automatically deploys, updates, or undeploys the index page, without manual intervention.

To enable such behavior, the connections between deployments and inter-deployment resource dependencies must be explicit. For this we propose three new resource types: RemoteConnection for connections to other deployments, Offer to provide resources or information to a connected deployment, and Wish to access the offer of a connected deployment. Using them, the provider and editor specify their connection (Lines 2.1 and 3.1). The provider offers its bucket (Lines 2.2 to 2.4) to the editor’s deployment in Line 2.5. The editor specifies their expectation of the offer by defining a wish in Line 3.2, allowing to use the offered bucket via wish.offer (Line 3.4). Together Listings 2 and 3 define the resource graph Figure 1b.

3.3 System Architecture and Design

Safe undeployment across teams. \(\mu S\) ensures that all resources depending on a resource \(R\) are not deployed anymore when \(R\) is undeployed. If \(R\) shall be undeployed, the undeployment of all resources depending on it is triggered and \(R\) is only undeployed after their undeployment is completed. E.g., when the provider undeploys the bucket, \(\mu S\) automatically undeploys the index before.

Reactive updates across teams. \(\mu S\) automatically transports configuration changes across the teams’ deployments and triggers reactive updates. E.g., the editor might show the bucket’s name (wish.offer.name) in the page’s content. If the provider updates the bucket name, this change is transported to the editor’s deployment, automatically updating the index page’s content.

Listing 2: \(\mu S\) deployment description of the provider.
21 const editor = new RemoteConnection('editor'); (22 const bucket = new aws.s3.Bucket('website', { 23 website: { indexDocument: 'index.html' } 24 }) ); (25 new Offer(editor, 'bucket', bucket);

Listing 3: \(\mu S\) deployment description of the editor.
3.1 const provider = new RemoteConnection('provider'); (3.2 const wish = new Wish<aws.s3.Bucket>(provider, 'bucket'); (3.3 const index = new aws.s3.BucketObject('index', { (3.4 bucket: wish.offer, content, key: 'index.html' (3.5 }) );
of another deployment delays the deployment of its resources. Also, unavailability delays the undeployment of a resource on which resources of the unavailable deployment (transitively) depend.

4 EVALUATION

We evaluate the existence of the research problem and $\mu$S’ effectiveness, performance, and applicability to existing IaC scripts. To identify the demand for decentralized automated deployment coordination, we perform an empirical study with IT professionals that follows the ACM SIGSOFT guidelines for questionnaire surveys [1] and advice from Kasunic [13] and Kitchenham and Pfleeger [14]. We want to find out: (1) How many dependencies do applications have with other applications? (2) Do such dependencies constrain their order of deployment and undeployment? (3) How are deployments coordinated in practice? (4) Do practitioners believe that automated coordination provides better SDO performance than manual coordination? (5) Does the organization’s SDO performance influence the answers to the previous questions? For the SDO performance-related questions, validated instruments from Forsgren et al. [8] are reused.

To evaluate the effectiveness of $\mu$S, we re-implement the deployment of existing applications and compare them with deployments in systems like Pulumi or AWS CDK. The amount of coordination overhead reduction shall be evaluated, and code-based metrics indicating the complexity shall be compared, e.g., lines of code, resource objects, etc. To evaluate $\mu$S’ performance, we implement a set of microbenchmarks in $\mu$S and other systems like Pulumi and AWS CDK, covering simple, typical deployments. Pulumi’s examples repository [23] is the starting point. With this benchmark suite, we compare the deployment duration, obtain insight into $\mu$S’ behavior, measure the delay to adapt to changes reactively, and assess the resource consumption. To assess the adaptability, we migrate existing IaC programs to $\mu$S. We start with Pulumi TypeScript programs using stack references [25]. Pulumi’s feature to explicitly model dependencies across deployments. These scripts are compatible with $\mu$S out-of-the-box; however, to leverage automated coordination, the stack references need to be converted to offers and wishes. To evaluate the migration from other IaC solutions, focusing on their features explicitly modeling dependencies across deployments, we use import and conversion tools from the Pulumi community [22] to obtain a $\mu$S compatible version. Then we enable $\mu$S’ automated deployment coordination by applying transformations based on information from the original programs.

5 ACHIEVED RESULTS

We organized the Dependencies in DevOps Survey 2021 from January to April 2021. It was filled by 134 IT professionals working in industry, who were advertised through snowball sampling [9] on a DevOps mailing list, social media, and in the authors’ personal networks. The central insights are that (1) the majority of applications depend on others, (2) such dependencies usually impact the deployment and undeployment order, and (3) deployments across teams typically rely on manual coordination, even though (4) automated coordination promises better SDO performance. The results show the demand for automated deployment coordination.

A first version of $\mu$S is implemented as a TypeScript library using Pulumi as driver and Hareactive [40] for the reactive engine [37] and presented in [36]. With it, a set of initial evaluations was performed. First, the deployment of the TeaStore [41] was implemented in $\mu$S, Pulumi, and AWS CDK. $\mu$S required 14% and 35% more lines of code. $\mu$S enables decentralized automated deployment coordination with small definition overhead. Second, in a microbenchmark $\mu$S’ deployment duration for a single microservice was compared, showing similar performance to Pulumi and better performance than AWS CDK. Third, the microservice was deployed multiple times with a serial and a parallel chain of dependencies between them. $\mu$S deploys the services—as expected—in sequence and parallel, each requiring roughly the same time measured in the single-service experiment. Fourth, we transformed stack references in 64 Pulumi projects from Github to $\mu$S offers and wishes. It is easy to adopt $\mu$S’ automated deployment coordination in decentralized Pulumi deployments using stack references.

6 CONTRIBUTION TO KNOWLEDGE

Research on IaC so far either focuses on modeling approaches like TOSCA or CaC solutions like Ansible, Chef, or Puppet (cf. §2). Recent IaC systems that support infrastructure provisioning and leverage general-purpose programming languages, e.g., Pulumi and AWS CDK, are not discussed in scientific research yet. $\mu$S is an entrance into this direction, and we presume the discussion of various problems in this field. For instance, such IaC solutions start dissolving the separation of infrastructure and application code, which could be further blurred for, e.g., safe updating and dynamic adaption. Also, we expect work on specializing debugging and testing techniques of general-purpose languages for IaC, improving the currently underdeveloped field of debugging and testing IaC [10].

The proposed decentralized mechanism to automate the deployment coordination is novel and enables further decoupling of cross-functional teams, improving DevOps. It can be further contributed back to modeling languages and implemented at the level of resource orchestrators, enabling decentralized federations suitable to span across organizations, which is not the case for today’s centralized orchestrator federation solutions.

7 CONCLUSION

DevOps aims for decoupled cross-functional teams, each independently developing and operating their applications. However, applications depend on other teams’ applications, requiring decentralized, asynchronous deployment coordination. We evaluate the practical relevance of this problem and solve it by proposing the $\mu$S IaC system, which automates decentralized deployment coordination. We implement $\mu$S and describe its evaluation for effectiveness, performance, and applicability to existing IaC scripts.

ACKNOWLEDGMENTS

The author is advised by Prof. Dr. Guido Salvaneschi, who is associate professor at the University of St. Gallen. This work has been co-funded by the German Research Foundation (DFG, No. 383964710, SFB 1119) and by the Hessian LOEWE initiative (emergenCITY and Software-Factory 4.0).