The rise of CloudOps

Marek Wiewiórka, Tomasz Gambin, Agnieszka Szmurło

October 2024

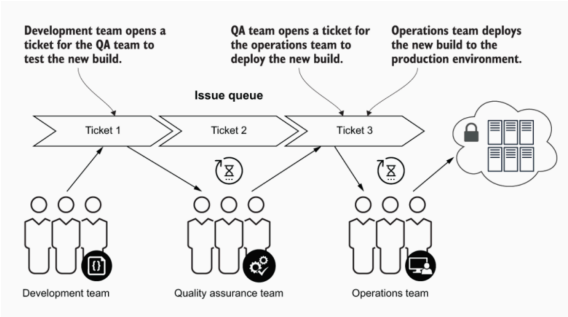
Agenda

1. What is CloudOps?
2. FinOps
3. GitOps and other XOps
4. Automation to the rescue

„Traditional“ Ops

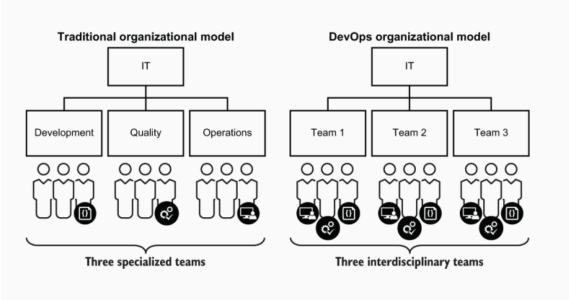
separation of:

- software development (Dev)
- quality assurance (QA)
- infrastructure configuration (IT Ops)
- software deployment (IT OPs)



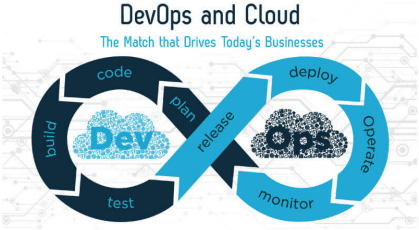
Dev(Sec)Ops

- *DevOps* is a set of software development practices that combine software development (Dev) and IT operations (Ops) to shorten the system development life cycle while delivering features, fixes, and updates frequently in close alignment with business objectives.
- *Automation* is a core principle for achieving DevOps success and CI/CD is a critical component.



CloudOps = DevOps for cloud computing

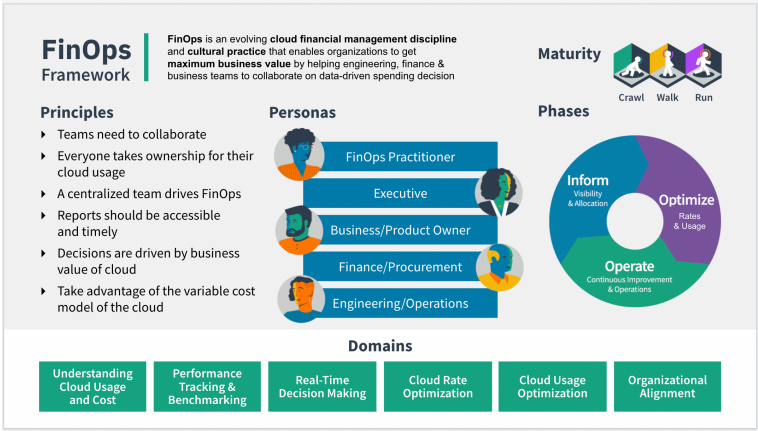
- lack of available cloud-related ops skills
- growing complexity - e.g. multicloud
- rising security threats - public clouds
- the shift to a utility consumption model - cloud cost management



- IaC and automation
- monitoring and logging
- security
- cost management

FinOps

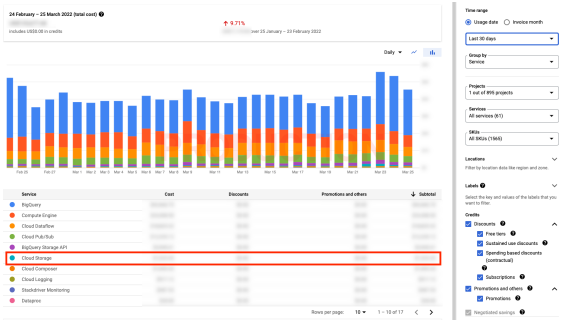
"If it seems that FinOps is about *saving* money, then think again. FinOps is about *making* money."



GCP - Cloud billing report

- *SKU* is the resource used by your service, e.g. for Cloud Storage:
  - class A operations (object adds, bucket/object list)
  - class B operations (object gets, retrieve bucket/object metadata)
  - Monthly amount of data retrieved
  - Monthly data transfer from Cloud

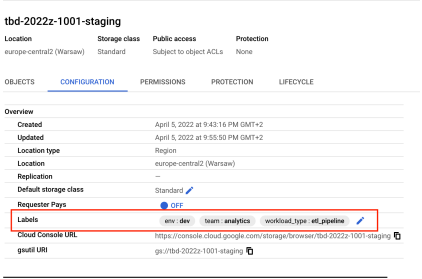
  other cost factors include:
  - Standard, Nearline, Coldline or Archive storage class
  - total size of bucket



GCP resource labeling for costs breakdown

```
resource "google_storage_bucket"
 ↪  "tbd-staging-bucket" {
  name =
   ↪  "${var.project_name}-staging"
  location = var.region
  uniform_bucket_level_access =
   ↪  false
  force_destroy = true
  labels = {
    env = var.environment
    workload_type = "etl_pipeline"
    team = "analytics"
  }
}
```

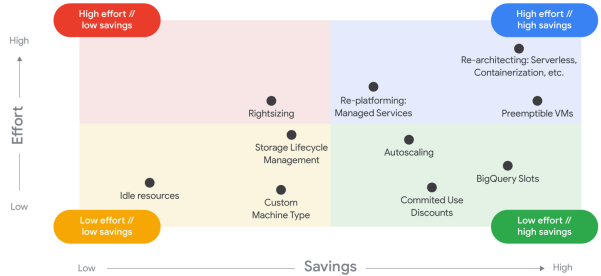
- develop cloud resources labeling convetions[a]
- export cloud billing to Big Query
- develop custom reporting using labels for grouping



[a] FinOps for data pipelines on Google Cloud Platform

FinOps - cost optimizations

- preemptible(spot) instances
- VM instance schedules
- autoscaling group of instances
- cloud storage lifecycle management
- serverless - but not because it's cheaper
- budgets and alerting
- organization/platform policies

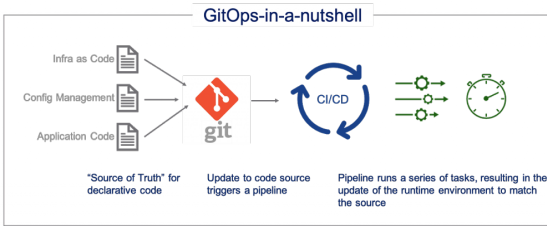


- understand pricing models
  - what are the main cost drivers?
  - Pay-as-you-go or reserved

GitOps

GitOps is a DevOps process characterized by:

- best practices for deployment, management, and monitoring of (containerized) applications
- a developer-centric experience for managing applications, with fully automated pipelines/workflows using Git for development and operations
- manage infrastructure and application configurations using Git (not only application code)



DataOps

- application of the best practices of Agile methodology, DevOps to data processing
- automated deployments
- *data quality* and unit testing
- automated *metadata* generation e.g. lineage
- automated *security* via tagging (tag-based policies)

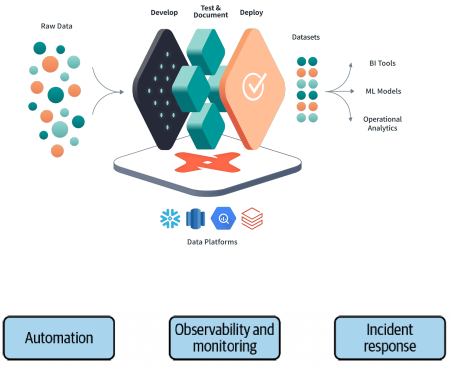


Figure: The three pillars of DataOps

MLOps 1/2

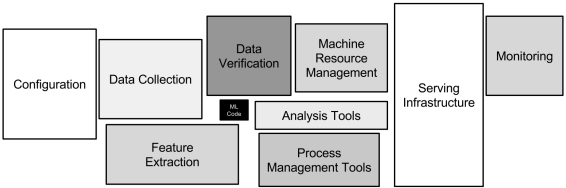


Figure: Elements of ML systems

- adaptation of DevOps principles to ML domain
- standardized ML model lifecycle aka models factory
- pioneered in 2015
- various kinds of technical debts, e.g.:
  - pipeline jungles
  - configuration debt
  - data testing debt

MLOps 2/2

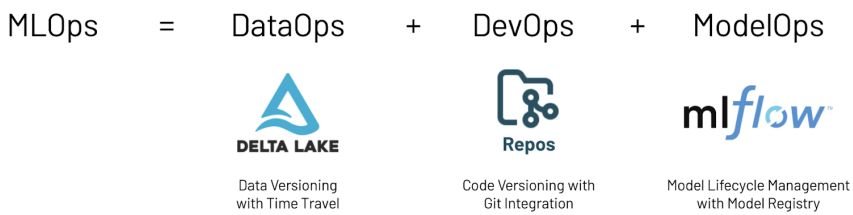


Figure: DataOps vs DevOps vs ModelOps

AIOps

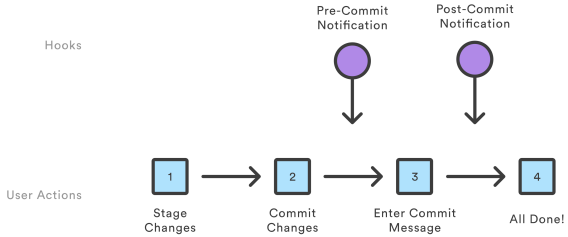
- **AIOps** stands for **Artificial Intelligence for IT Operations**.
- Combines AI and machine learning to automate and improve IT operations.
- Aims to enhance monitoring, troubleshooting, and incident response in complex environments.
- a few examples:
  - k8sgpt - Kubernetes clusters triaging and troubleshooting with LLMs
  - LLM-powered automatic labeling, commit summaries, even code reviews
  - automatic test cases/ data quality tests generation

Automation to the rescue - IaC

- Terraform is great…but have some limitations
- Terragrunt and Atlantis for more robust CI/CD processes - e.g. DRY principle - versions, variables
- modules vs stacks
  - radius blast
  - speed
  - resources lifecycle - e.g. compute vs storage

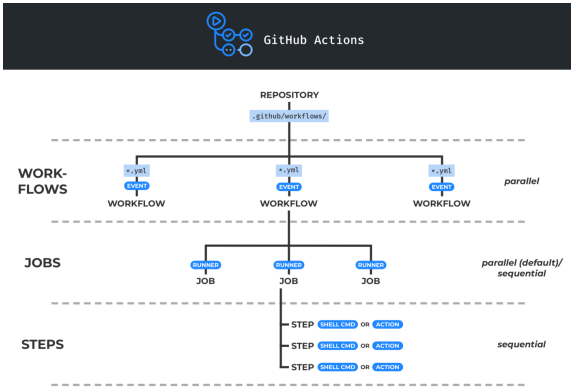
Automation to the rescue - git hooks

pre-commit – a framework for managing and maintaining multi-language pre-commit hooks.




```
➜  thd-20222-infra-internal git:(feature/ci-cd) ✗ git commit -m "Check"
Terraform fmt....................................................................Passed
Terraform docs...................................................................Passed
Terraform validate with tflint...................................................Failed
- hook id: terraform_tflint
- exit code: 2


TFLint in ./:
4 issue(s) found:

Warning: variable "group_id" is declared but not used (terraform_unused_declarations)

  on variables.tf line 1:
   1: variable "group_id" {

Reference: https://github.com/terraform-linters/tflint/blob/v0.35.0/docs/rules/terraform_unused_declarations.md
```

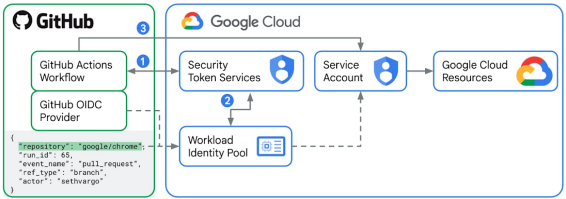
Automation to the rescue - Github Actions




```yaml
name: Pull request workflow
on:
  pull_request:
    branches: [dev]
jobs:
  pull-request:
    runs-on: ubuntu-latest
    steps:
    - uses: actions/checkout@v2
    - uses: hashicorp/setup-terraform@v1
      with:
        terraform_version: 1.1.7
    - name: Terraform fmt
      id: fmt
      run: terraform fmt -check
```

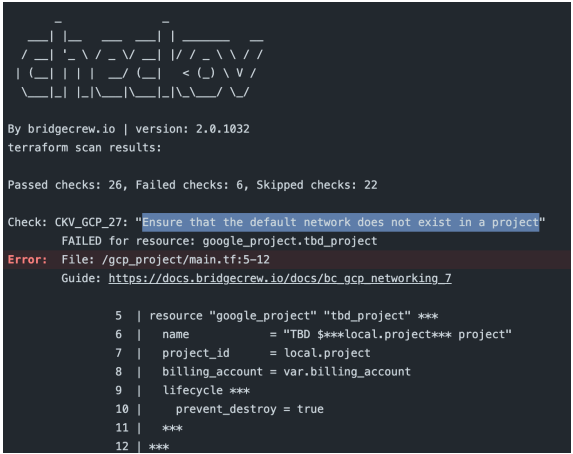
Github Actions – keyless authentication

- fine-grained scoping – e.g. GitHub repository, Organization
- short-lived credentials – by default 1 hour, greater security even if compromised
- minimal management overhead – no need for JSON service keys, short-lived tokens rotated every deployment



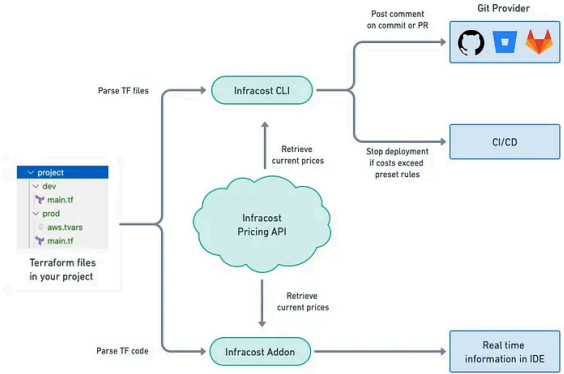
Automation to the rescue - SecOps

- static application security testing (SAST) tools like checkov, tfsec or terrascan
- policy-as-code
- support for many IaC technologies (e.g. Terraform, Docker files, Helm charts)



Automation to the rescue - FinOps

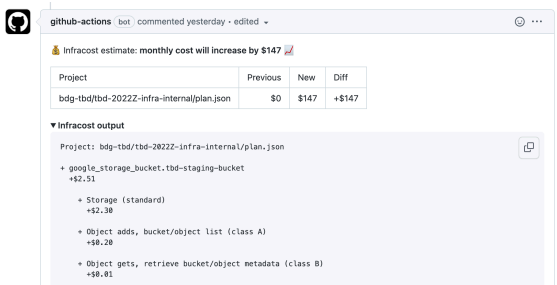
- analyses Terraform code
- support for CI/CD and code editor
- unfortunately some data services - such as GCP Composer or DataProc still not supported :-(



Automation to the rescue - FinOps

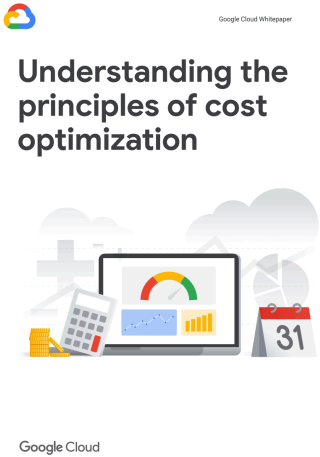
Infracost Cloud cost estimates for Terraform in pull requests

```
#usage file
version: 0.1
resource_usage:
google_storage_bucket.tbd-staging-bucket:
storage_gb: 100
monthly_class_a_operations: 40000
monthly_class_b_operations: 20000
monthly_data_retrieval_gb: 1000
```



Books


Google Cloud
Getting started with FinOps on GCP
Authors:
Sam Moss, Kinjal Tanna, Tan-Minh Truong


Google Cloud Whitepaper
Understanding the principles of cost optimization
31
Google Cloud

Thank you !

Q&A

marek@getindata.com
marek.wiewiorka@gmail.com