

Data lake migration to cloud

a modern infrastructure approach

Ramon Gonzalez

Cloud and Analytics Architect

REAL use case



High number of files to be stored
non-structured and semi-structured data



Automation
Solution should be reusable without "any kind"
of human effort



Collaborative
It should implement a collaborative way of work
usable by a distributed team

About me

Ramón González Castro

INDITEX Cloud Architect (2019 -)



BI & Analytics PFE (2014 – 2019)

@ramonbgc 

<https://www.linkedin.com/in/ramongonzalezcastro> 

Automation

- Has been used for a while (for several tasks, processes...)
- Really(?) well defined for some of them

- Configurations:  **ANSIBLE**

- CI/CD:  **Jenkins**  **GitLab**

- Planned executions:  **bmc** Control-M

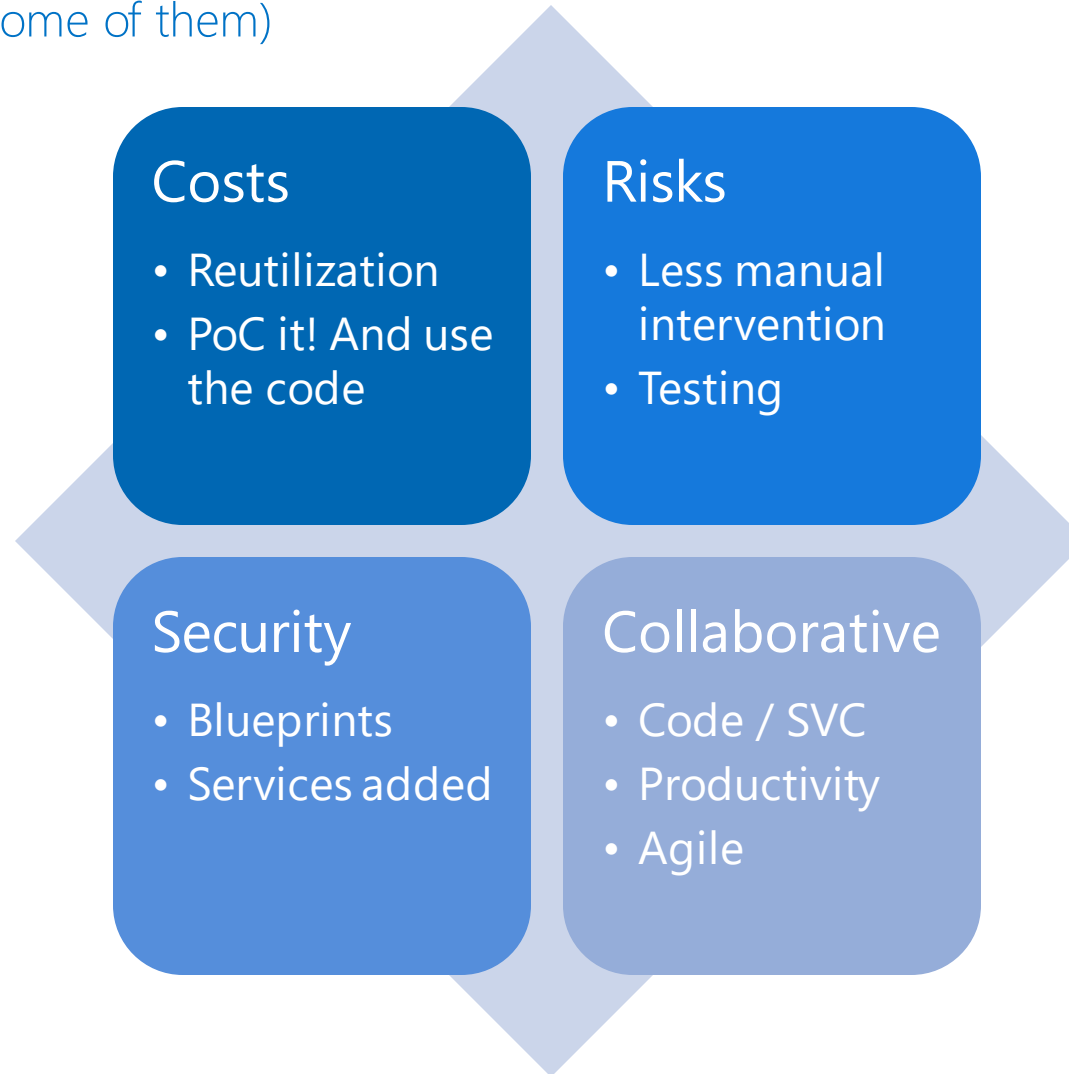
Automation

What is (was?) infrastructure automation really?



Introducing ... IaC (Infra as Code)

The benefits (just some of them)



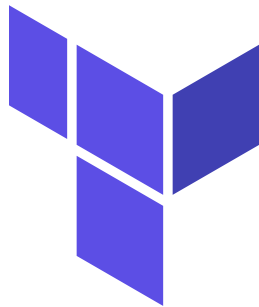
Introducing ... IaC (Infra as Code)

The REAL reason



Introducing ... IaC (Infra as Code)

- How to start?
- What “tool” to choose?*



Cloud
Deployment
Manager

HashiCorp

Terraform



*But it's just a suggestion to start playing with IaC! I don't work for HashiCorp!!

Introducing ... IaC (Infra as Code)

- Terraform benefits (also in other tools*):

- Works for the (multi) cloud

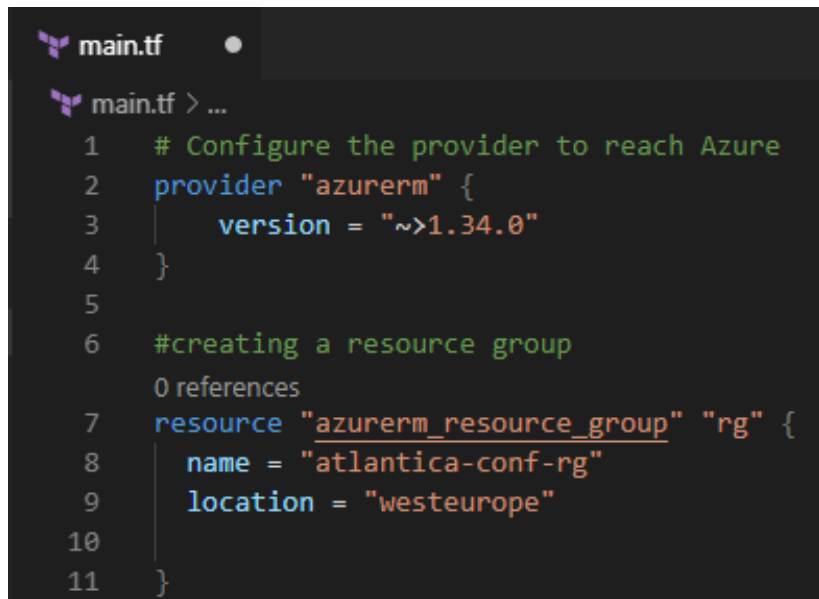


- As well as for “on-prem” (virtualization)



Introducing ... IaC (Infra as Code)

- Terraform benefits (also in other tools):
 - Declarative



```
main.tf
main.tf > ...
1  # Configure the provider to reach Azure
2  provider "azurerm" {
3      version = "~>1.34.0"
4  }
5
6  #creating a resource group
7  0 references
8  resource "azurerm_resource_group" "rg" {
9      name = "atlantica-conf-rg"
10     location = "westeurope"
11 }
```

- Idempotent

Idempotence (UK: /[ɪdɛmˈpʊtəns](#)/,^[1] US: /[aɪdəm-/](#))^[2] is the property of certain [operations](#) in [mathematics](#) and [computer science](#) whereby they can be applied multiple times without changing the result beyond the initial application. The concept of idempotence arises in a number of places in [abstract algebra](#) (in particular, in the theory of [projectors](#) and [closure operators](#)) and [functional programming](#) (in which it is connected to the property of [referential transparency](#)).

Introducing ... IaC (Infra as Code)

- Terraform benefits (also in other tools):

- Interpolation
- Execution plans
- Resource Graph

```
resource "azurerm_storage_account" "sa" {
  name = "${var.project_name}${var.resource_group_name}${var.location}"
}
```

An execution plan has been generated and is shown below.

```
Terraform v0.12.9
Initializing plugins and modules...
2019/10/03 16:09:17 [DEBUG] Using modified User-Agent: Terraform/0.12.9 TFC/51146a6042
azurerm_resource_group.rg: Creating...
azurerm_resource_group.rg: Creation complete after 2s [id=/subscriptions/ce358b9a-f63d-43c0-bb0e-0800202c3602/resourceGroups/rg]
azurerm_data_factory.df: Creating...
azurerm_storage_account.sac: Creating...
azurerm_data_factory.df: Creation complete after 8s [id=/subscriptions/ce358b9a-f63d-43c0-bb0e-0800202c3602/dataFactories/df]
azurerm_storage_account.sac: Still creating... [10s elapsed]
azurerm_storage_account.sac: Still creating... [20s elapsed]
azurerm_storage_account.sac: Creation complete after 24s [id=/subscriptions/ce358b9a-f63d-43c0-bb0e-0800202c3602/storageAccounts/sa]
azurerm_template_deployment.pipe: Creating...
azurerm_storage_container.sc: Creating...
azurerm_storage_container.sc: Creation complete after 1s [id=https://atlanticaconfsa01.blob.core.windows.net/atlanticaconfsa01]
azurerm_template_deployment.pipe: Still creating... [10s elapsed]
```

```
}
```

Introducing ... IaC (Infra as Code)

Pleaaaaaaase stop talking and ...

show me the ~~money~~ DEMO

THANK YOU!!!