NÁSKOK
DÍKY
ZNALOSTEM

PROFINIT

hadoop for Beginners

Cloud

Profinit Big Data team
Alisa Benešová

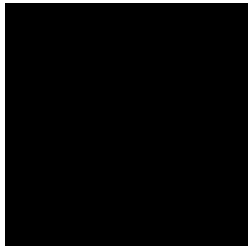
16.11.2022

Content

- Cloud
- Cloud vendors
- Cloud use cases
- Cloud architectures
- Terraform
- Azure, AWS cost calculator

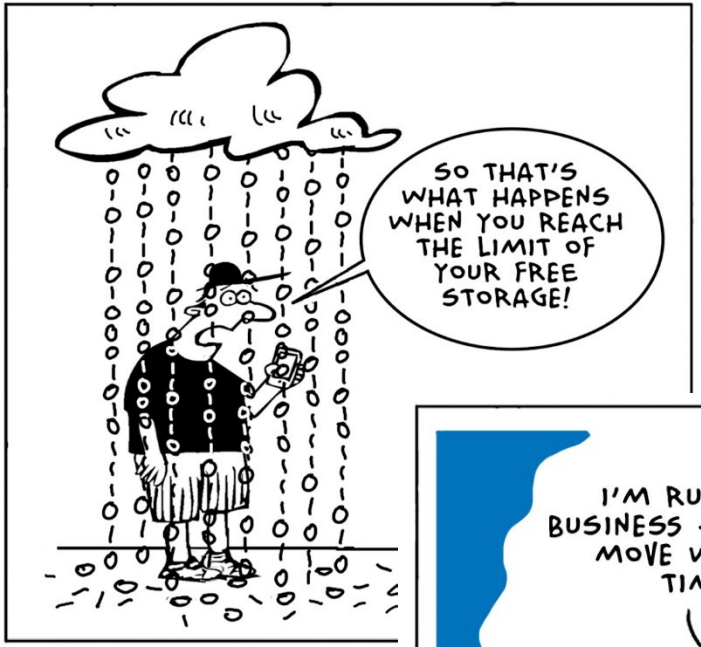
PROFINIT

Cloud

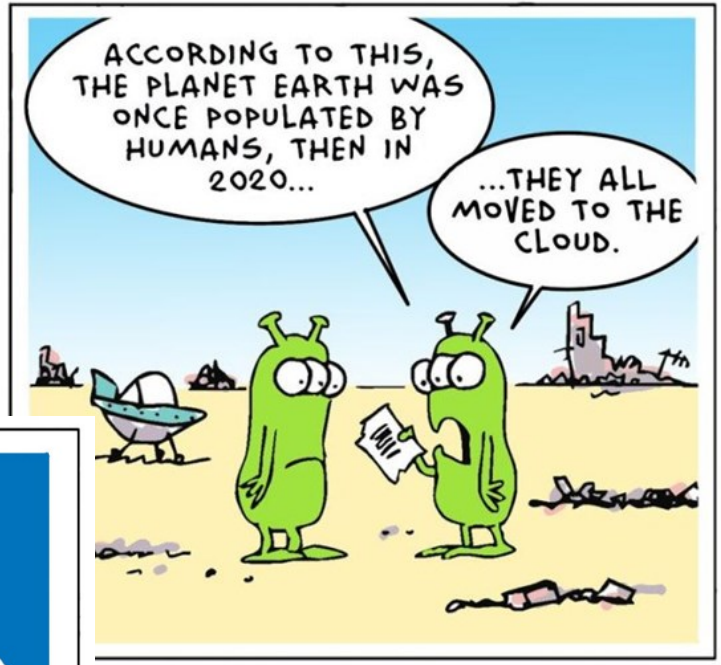


Do you know what cloud is?

PROFINIT


SO THAT'S WHAT HAPPENS WHEN YOU REACH THE LIMIT OF YOUR FREE STORAGE!


I'M RUNNING A BUSINESS – I HAVE TO MOVE WITH THE TIMES.

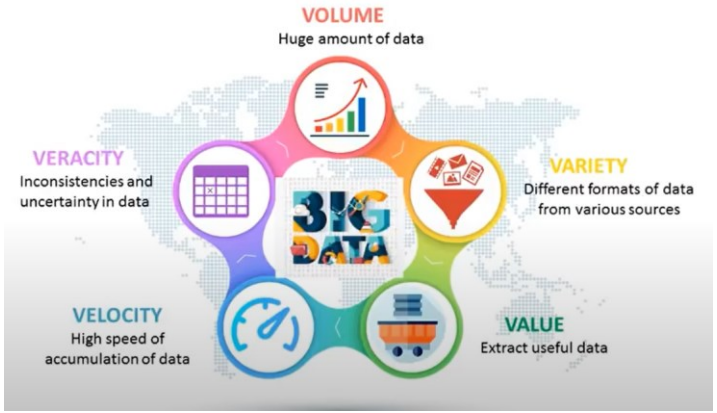

ACCORDING TO THIS, THE PLANET EARTH WAS ONCE POPULATED BY HUMANS, THEN IN 2020...
...THEY ALL MOVED TO THE CLOUD.

PROFINIT

Cloud

- What Cloud is?
  - Almost unlimited space in cloud,
  - Collection of different servers, tools,
  - Infrastructure orchestration etc.
  - “pay what you use” (GB, cores, security…)

- BigData and Cloud?
  - **Scalability of computing power and storage,**
  - Cost prediction,
  - Click and Go solutions.



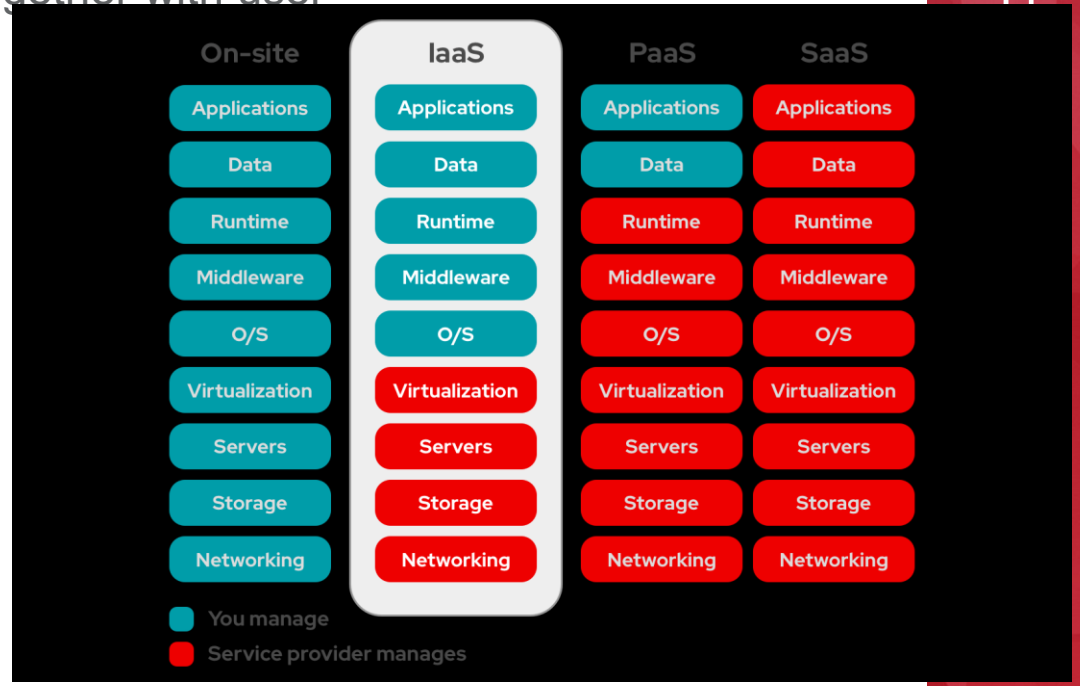
PROFINIT

Cloud services

- Cloud service = service provided by a cloud provider via Internet.

- Cloud vendor manages services together with user
  - 3 different levels

  - **IaaS – Infrastructure as a Service**
  - **PaaS – Platform as a Service**
  - **SaaS – Software as a Service**



Cloud service vendors

- Amazon, Microsoft, Google, Alibaba = world-wide leaders
- **Amazon AWS**
  - Leader in cloud computing (first in 2008),
  - AI, Serverless deployments, IaaS
- **Microsoft Azure**
  - Leader in SaaS (MS platform)
  - Enterprise customers
- **Google GCP**
  - Google platform services
- **Alibaba Cloud**
  - Primary cloud in China

PROFINIT

- Cloud has changed IT world (hardware, software, security, dataflow, infrastructure…)
- Almost every big player offers a cloud, cloud solution, cloud services to be hosted on cloud

PROFINIT

Gartner – magic quadrant

PROFINIT

Cloud vendors differs… (not fully true anymore)

Summary of Major Vendor Emphasis

	Build Private Services	Deliver Services	Services Delivered* IaaS	PaaS	SaaS	Private Offerings Enabling Tech.	Packaged Cloud
Amazon	○	●	●	◔	○	None	
salesforce.com	○	●	○	●	●	None	
Google	○	●	◕	◕	●	None	
Microsoft	●	●	◑	●	●		
IBM	●	◑	◔	◑	◑		
VMware	●	◕	○	◔	◕		
Oracle	●	◕	◕	◑	◑		
SAP	○	◑	○	◔	◑	None	
HP	●	◑	◔	◔	◑		

Note: This is not an evaluation of capabilities, but rather of emphasis.

● ◔ ◑ ◕ ○ Significant → None

* The provider may offer public, community or virtual private services

Gartner

PROFINIT

Private Cloud

Cloud deployment options

- Cloud deployment in an organization
  - Public
  - Private
  - Hybrid
- A public cloud is where an independent third-party provider, owns and maintains compute resources that customers can access over the internet.
- The private cloud removes this sharing aspect of cloud computing, instead dedicating infrastructure and services to a single “user”.
- A hybrid cloud is a model in which a private cloud connects with public cloud infrastructure, enabling an organization to orchestrate workloads across the two environments.

PROFINIT

Private clouds – how they are hosted and managed

- Virtual
  - Walled-off environment within a public cloud that enables an organization to run its workloads in logical isolation from every other user of the public cloud
  - Even though the server is shared by other organizations, the virtual logic ensures that a user's computing resources are private
- Hosted
  - The servers aren't shared with other organizations
  - The service provider configures the network, maintains the hardware and updates the software, but the server is occupied by a single organization
- Managed
  - The provider manages every aspect of the cloud for the organization, including deploying additional services such as identity management and storage

Private clouds – infrastructure differences

- Software-only
  - Vendor provides the necessary software which runs on an organization's preexisting hardware
  - OpenStack

- Software and hardware
  - All-in-one bundle
  - It's a simple platform that exists on the user's premises and may or may not be provider-managed environments.
  - Azure stack

PROFINIT

Cloud use cases

Use cases in real world

- **Prototyping (POC), Dev, Testing**
  - BD Architecture is summoned when you need it and comply your project needs
  - When you are not sure
    - What to use (sizing, platform)
    - Not ready to invest to hardware
    - If big data architecture is right for your project

PROFINIT

Use cases in real world

- **Prototyping (POC), Dev, Testing**
  - Usually cloud native tools, such as HDInsight, Databricks etc.
    - Quick launch
    - You do not care about underlay infrastructure
    - Minimal administration
  - Some ready-made images in cloud-shop
  - Managed software (Kafka, Airflow …)
  - This might not be most cost effective, if you’re summoning too often for a long time.

PROFINIT

Use cases in real world

- **Cloud Server instances with installed BD tools**
  - Long-term running, almost same as on-premise solution
  - Development, Testing, Production
  - Easy to boost server instances, if you need
  - Long-term running might be ~~~ cost effective (long term plans)
  - You have to care about infrastructure, administration etc.

PROFINIT

How to save money in cloud?

- **Question for million $$$**

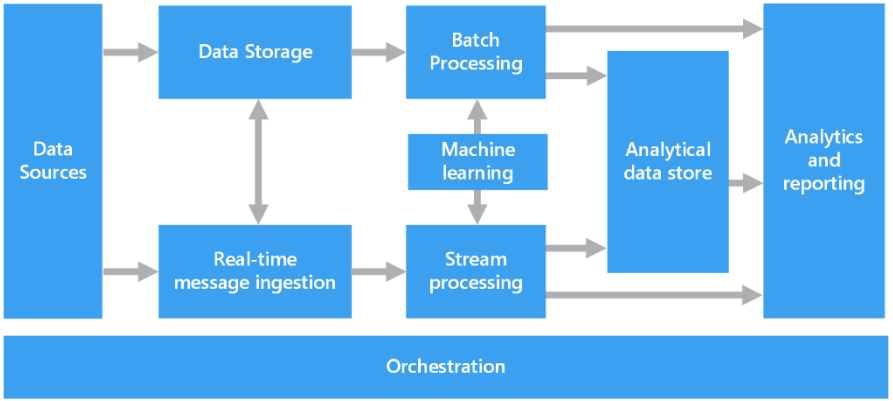
- **Turn off your instances, services**
  - Terraform is your friend
- Size you solution properly
  - Linear scaling?
- Think about arm
- Use reserved instances
- Shared responsibility is a budget killer

PROFINIT

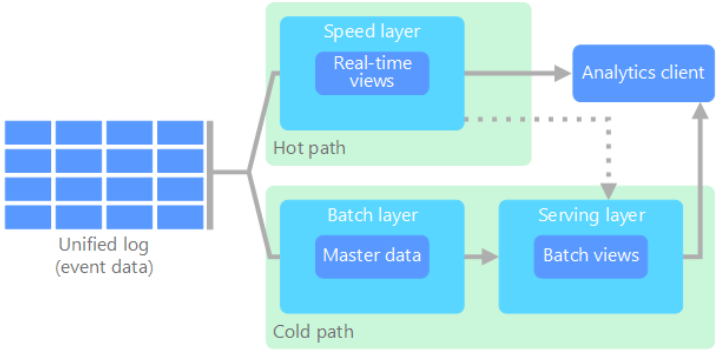
Cloud Architectures

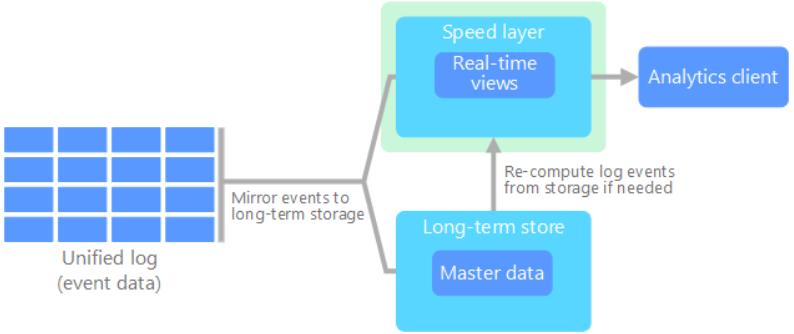
Big Data Architectures

- Cloud fits to BD architectures
  - Native components
  - Services
  - Hybrid solutions



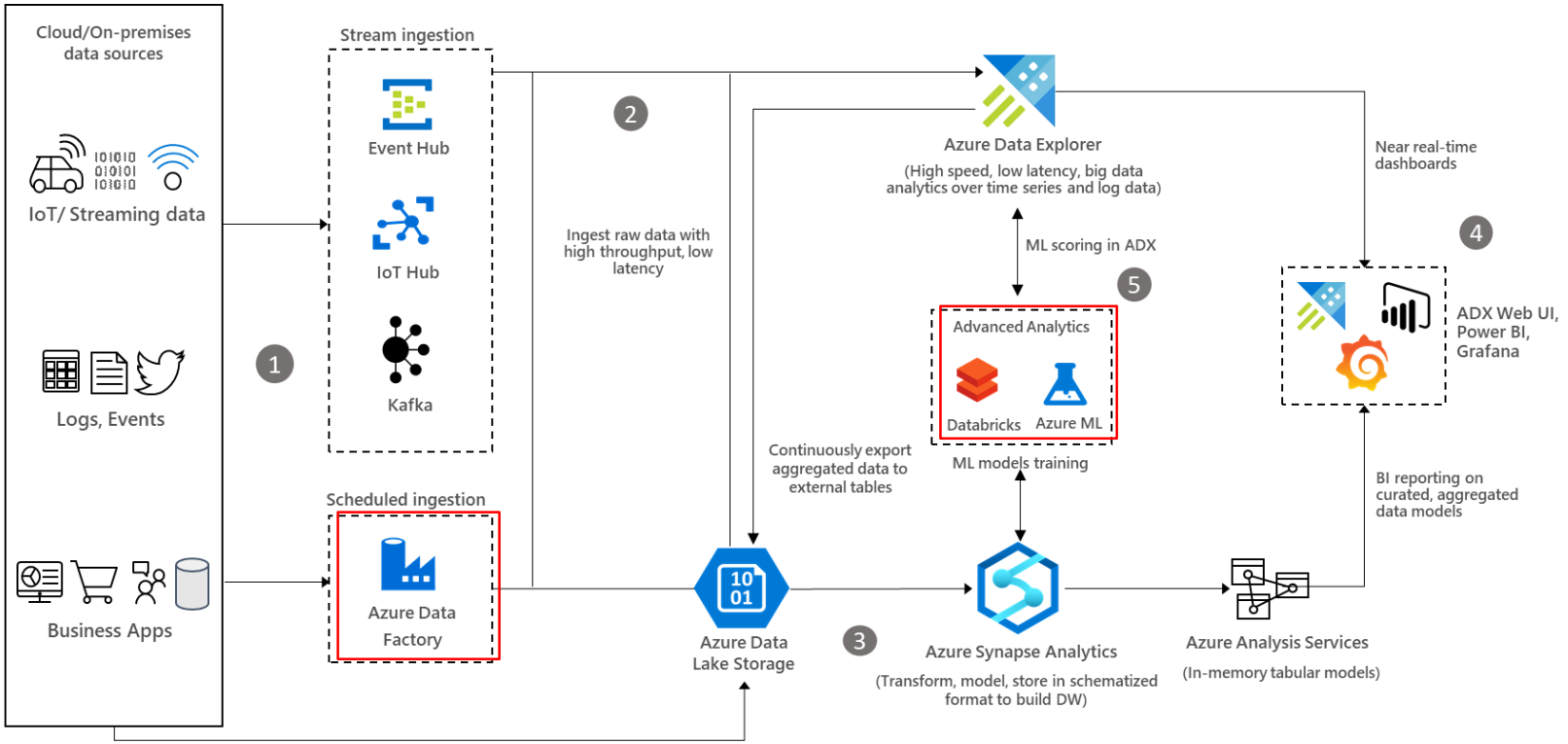
- Lambda and Kappa, IoT





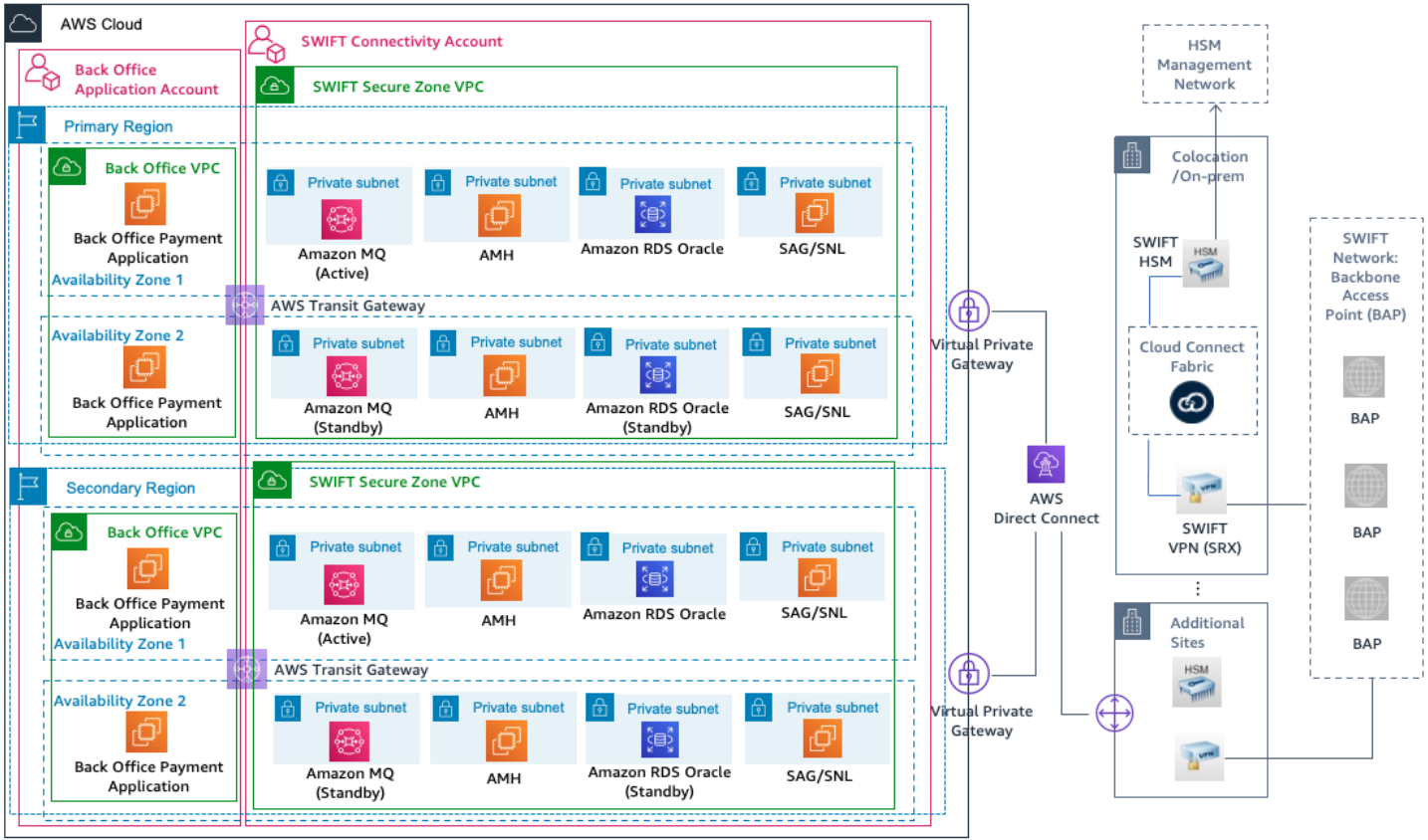
PROFINIT

Example of complex Azure BD analytic architecture



PROFINIT

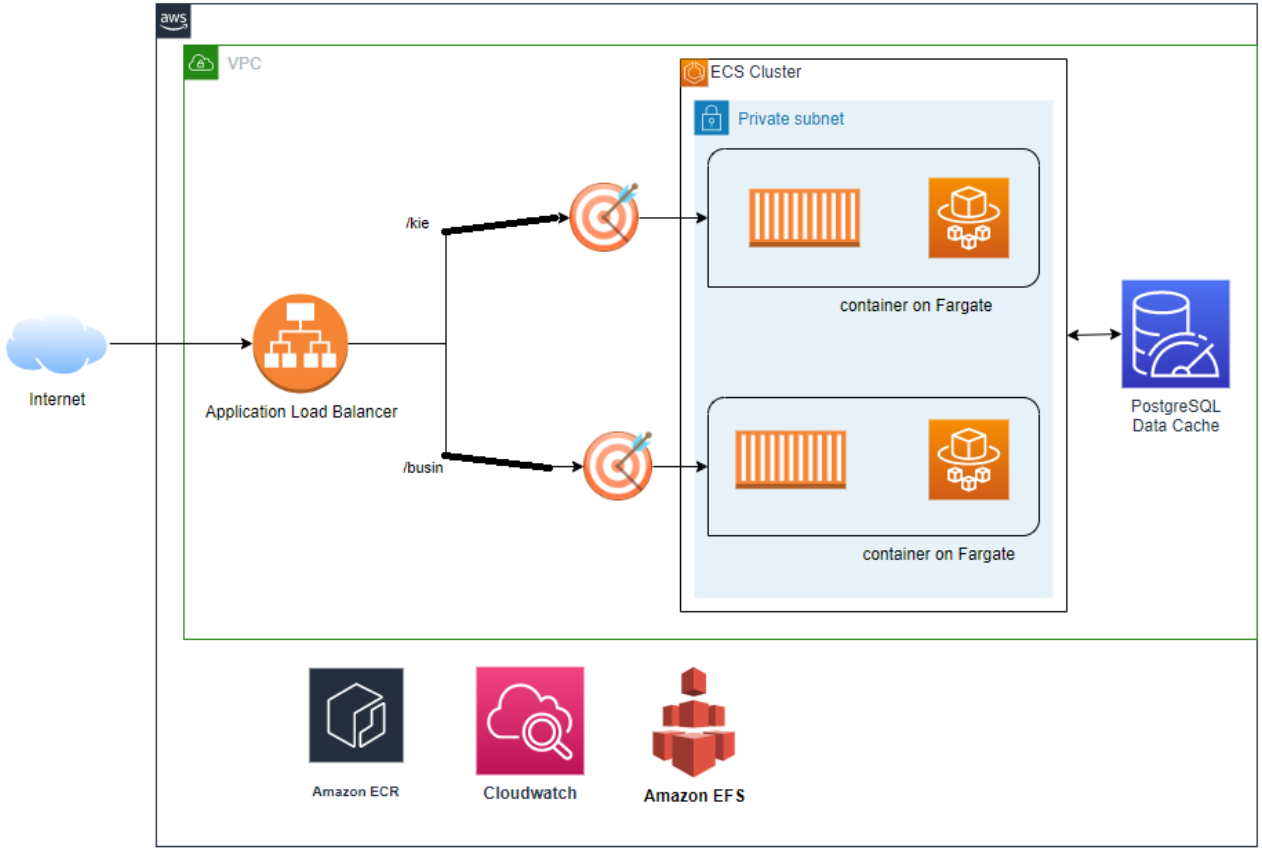
Examples of AWS architecture (reference)



https://aws.amazon.com/blogs/architecture/architecting-swift-connectivity-on-amazon-web-services-aws/

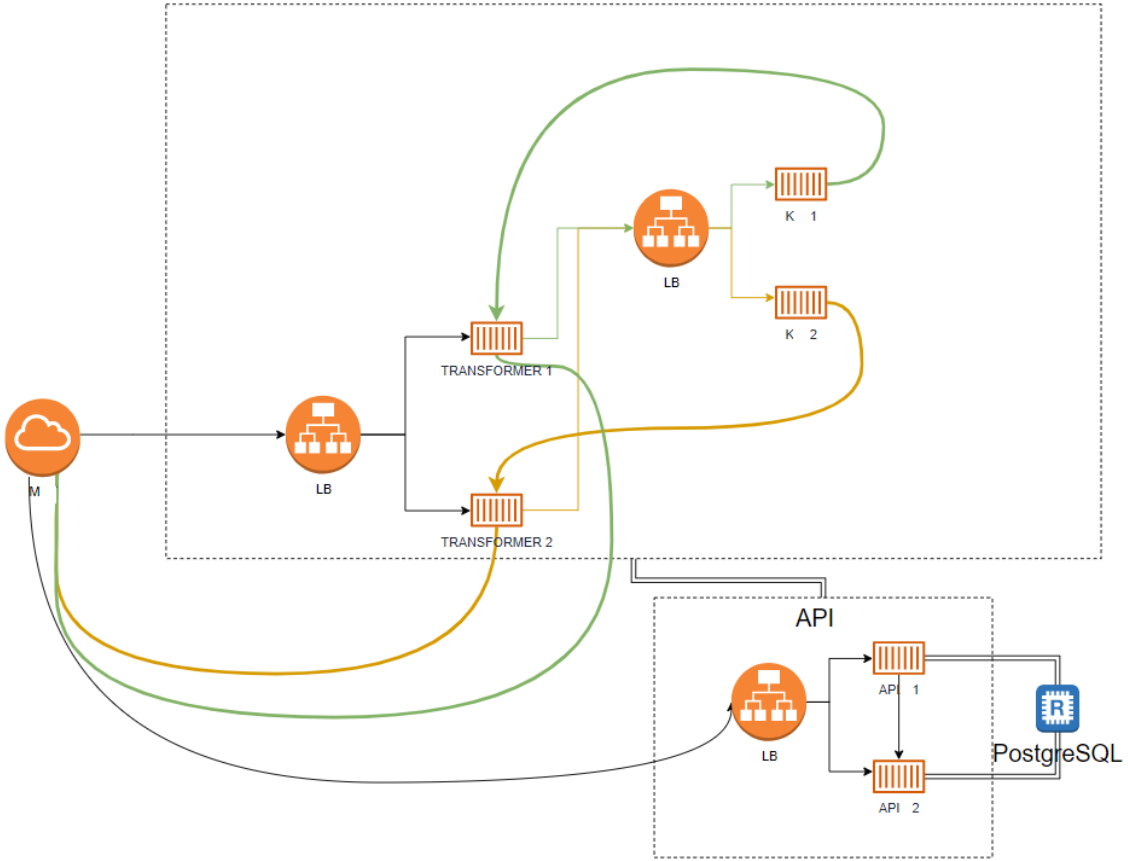
PROFINIT

Examples of AWS architecture (draw.io)



PROFINIT

Examples of AWS architecture



PROFINIT

Reference architectures

https://aws.amazon.com/blogs/architecture/
https://learn.microsoft.com/en-us/azure/architecture/browse/

PROFINIT

Terraform – infrastructure as a code

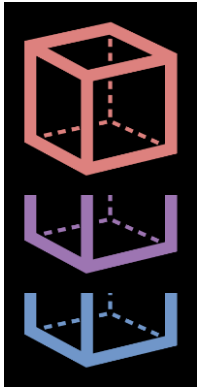
Terraform

HashiCorp Terraform

- Open-Source automatization and management of your
  - (cloud) infrastructure
  - Your platform
  - Services
- Declarative language used – define WHAT result you want

Declarative

- Define what you have
  - Red, violet, blue cube
- Define what you want
  - A tower od red, violet, blue cubes



Procedural

- Define what you have
  - Red, violet, blue cube
- Define how to make what you want
  - Put blue cube
  - Put violet cube on blue cube
  - Put red cube on violet cube

PROFINIT

Ansible vs. Terraform

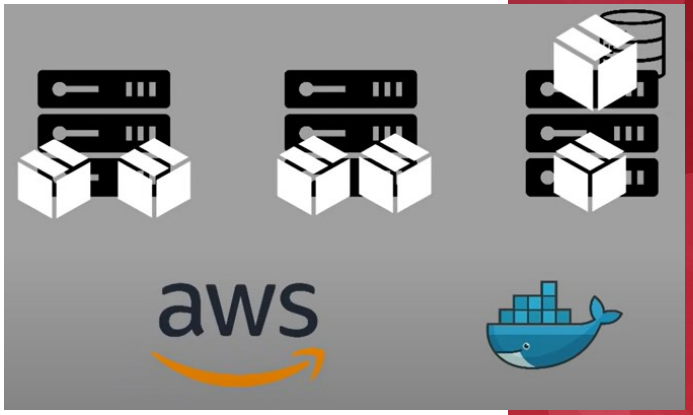
- Infrastructure as Code

Ansible	Terraform
Mainly a config tool (once infra is done)	Mainly infra provisioning
Deploy apps	Can deploy apps
Install/update software	
More mature	Relatively new tool
	Advanced in orchestration
Better for configuring infrastructure	**Better for provisioning infrastructure**

- DevOps usually use both tools

PROFINIT

Terraform, case study



- 3 servers,
- Several microservices and database in Docker
- In usual, you must do steps:
  - Prepare private network
  - EC2 server instances
  - Install Docker, tools
  - Security, firewalls etc.
  - Deploy Docker containers.

1. PROVISIONING INFRASTRUCTURE

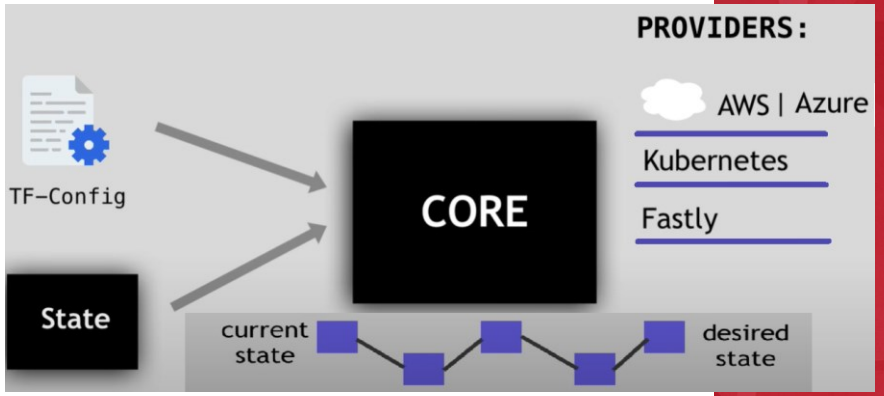
2. DEPLOY APPS

- 2 separate teams, usually
- Adding new servers, security setup, replicating from dev to prod…

⇨ AUTOMATION VIA TERRAFORM

Terraform, how it works

- **Core** takes **input** and plans what needs to be created, updated, destroyed… from current **state** = execution plan
- Steps are executed with platform specific tools
- **Providers** (100 providers)
  - AWS, Azure (IaaS)
  - Kubernetes (Paas)
  - Fastly (SaaS)
- Each provider offers resources you can work with



PROFINIT

Terraform - steps

- **Have the latest version of terraform**

- **terraform init**
  - Initialize the environment
- **terraform plan**
  - To see what will happen
  - Save the plan and apply it otherwise you are not sure what will be executed
- **terraform apply**
  - Do the job!

PROFINIT

Terraform – steps - example

```
-/+ resource "aws_ecs_task_definition" "transformer" {
      ~ arn                      = "arn:aws:ecs:eu-west-1:70332xxxxxx2:task-definition/test-transformer-uat:12" -> (known after apply)
      ~ container_definitions    = (sensitive) # forces replacement
      ~ id                       = "test-transformer-uat" -> (known after apply)
      - ipc_mode                 = "" -> null
      - pid_mode                 = "" -> null
      ~ revision                 = 12 -> (known after apply)
      - tags                     = {} -> null
        # (9 unchanged attributes hidden)
    }
Plan: 3 to add, 12 to change, 3 to destroy.
Saved the plan to: plan.tfplan
To perform exactly these actions, run the following command to apply:
    terraform apply "plan.tfplan"
```

PROFINIT

Terraform – steps - example

- Apply complete! Resources: 3 added, 12 changed, 3 destroyed.

- Outputs:

- common_container_sg = "sg-0ae9ceXXXXXXXXXXda"
- db_endpoint = “test-uat.ccxxxxxxwugh.eu-west-1.rds.amazonaws.com"

PROFINIT

Terraform – state file

- Terraform stores information about your infrastructure in a state file
- This state file keeps track of resources created by your configuration and maps them to real-world resources
- Extremely important!

- **Can be stored on**
  - Localhost
  - S3 / another shared cloud storage
  - HashiCorp paid solution - consul

PROFINIT

Terraform – sections

- Resource
  - Managed by terraform
- Data
  - Managed by others
- Module
  - Larger independent part
- Output
  - Output of the terraform script
- Provider
  - “Connector”
- Variable
  - Variable

PROFINIT

Terraform – modules

- Structure your code
- Have a logical structure
- Create reusable parts
- Work with variables

PROFINIT

Terraform – architecture

- Typically executed from GitHub actions or Jenkins
- First of all you compile your artifacts and store them somewhere
- Then run terraform to create infrastructure and use artifacts from the storage
- You can destroy your dev environment every day and have it off during the weekends ☺ (to be sure that you can build it from the scratch)

PROFINIT

Terraform – recommended online course

- Pluralsight
  - Ned Bellavance
    - Terraform and terraform deep dive
    - Approximately 8 hours

https://www.pluralsight.com/courses/terraform-getting-started

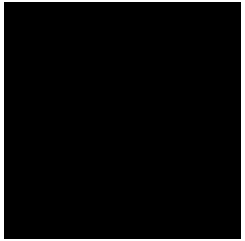
https://www.pluralsight.com/courses/terraform-deep-dive

PROFINIT

Cloud Costs

Cost Calculator

- Check how much you will pay what for
  - Total Cost of Ownership (TCO) Calculator | Microsoft Azure
    - Even electricity cost etc.
  - AWS Pricing Calculator
    - Just UseCase of your setup.



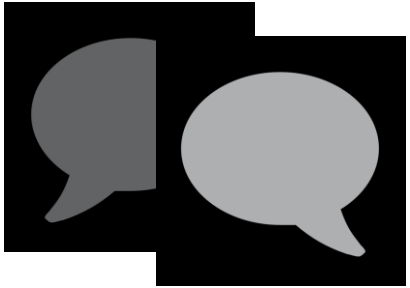
PROFINIT

For next hands-on Azure-cloud course

- **Create a NEW account as a student for free**
  - Free space, free computing etc.
  - **https://azure.microsoft.com/en-us/free**

- Alternative:
- Create a NEW FREE account with **200 USD to spend for 30 days**
  - Be careful, you have to add your credit card !
  - https://azure.microsoft.com/en-us/free/

PROFINIT



Dotazy

PROFINIT