# Cloud for Data Analytics

## 1 - Cloud and Virtualisation

**Juan Ángel Lorenzo del Castillo**

École Internationale des Sciences du Traitement de l'Information

ING3 - Ingénierie Mathématique et Simulation Informatique

2019-2020


juanangel.lorenzodelcastillo@eisti.eu

Ecole Internationale des Sciences du Traitement de l'Information

# Table of Contents

# Table of Contents

# About me

## Juan Ángel Lorenzo del Castillo

- B.Sc. in Telecom Engineering majoring Electronics from the University of Valladolid (Spain).
- M.Sc. in Telecom Engineering from the University of Valladolid (Spain).
- M.Sc. in Distributed Systems. Inter-university Doctorate Program, USC-UDC (Spain).
- Ph.D. from the University of Santiago de Compostela (Spain).

- Visiting researcher at the EPCC (The University of Edinburgh).
- Visiting researcher at the Charles University in Prague (Czech republic).
- Collaborations with HP Labs USA.

- Assistant lecturer at the University of Santiago de Compostela.
- Lecturer and lab instructor in several computing courses.
- Collaborations in open source communities in Spain.
- Associate professor at EISTI. Co-head of the IMSI option.

- Software engineer at the Centre for Telecommunication Development in Spain.
- Research engineer at the Cloud and Security Lab (Hewlett Packard Labs, Bristol).
- Postdoctoral researcher at INRIA Bordeaux.
- Associate researcher at ESTIA Bidart.
- Researcher at EISTI.
- Research interests:

  Cloud computing infrastructures.
  Virtualisation.
  Operating systems scheduling.
  Performance monitoring of large-scale architectures.
  Data profiling in real and virtualised environments.

# About you

- Who are you?
- What is your background in Informatics?
  - What do you know about Distributed Systems?
  - What do you know about Programming?
  - Do you have any computing interests outside the school?
- What is Cloud Computing for you?
- What do you expect from this course?
  - Btw, make sure you have an *@eisti.eu* email account.

# Course outline

1. Introduction to Cloud Computing and virtualisation
   - Definitions
   - Types of Cloud infrastructures
   - Hypervisors
   - Cloud Platforms examples
2. Introduction to Docker
3. Microservices with Docker
4. The Google Cloud Platform
5. Data Analytics on the GCP

# Pedagogy and Assessment

**Pedagogy:**

- This course will be taught in English.
- The course slides and additional material will be either in French or English.
- *Constructivist approach*: I explain, I propose. You try by yourselves, you learn by experimenting.

**Assessment:**

- One single exam.
- Some exercises may be graded.

# Table of Contents

# What is the Cloud?

# What is the Cloud?

## You already are a cloud user

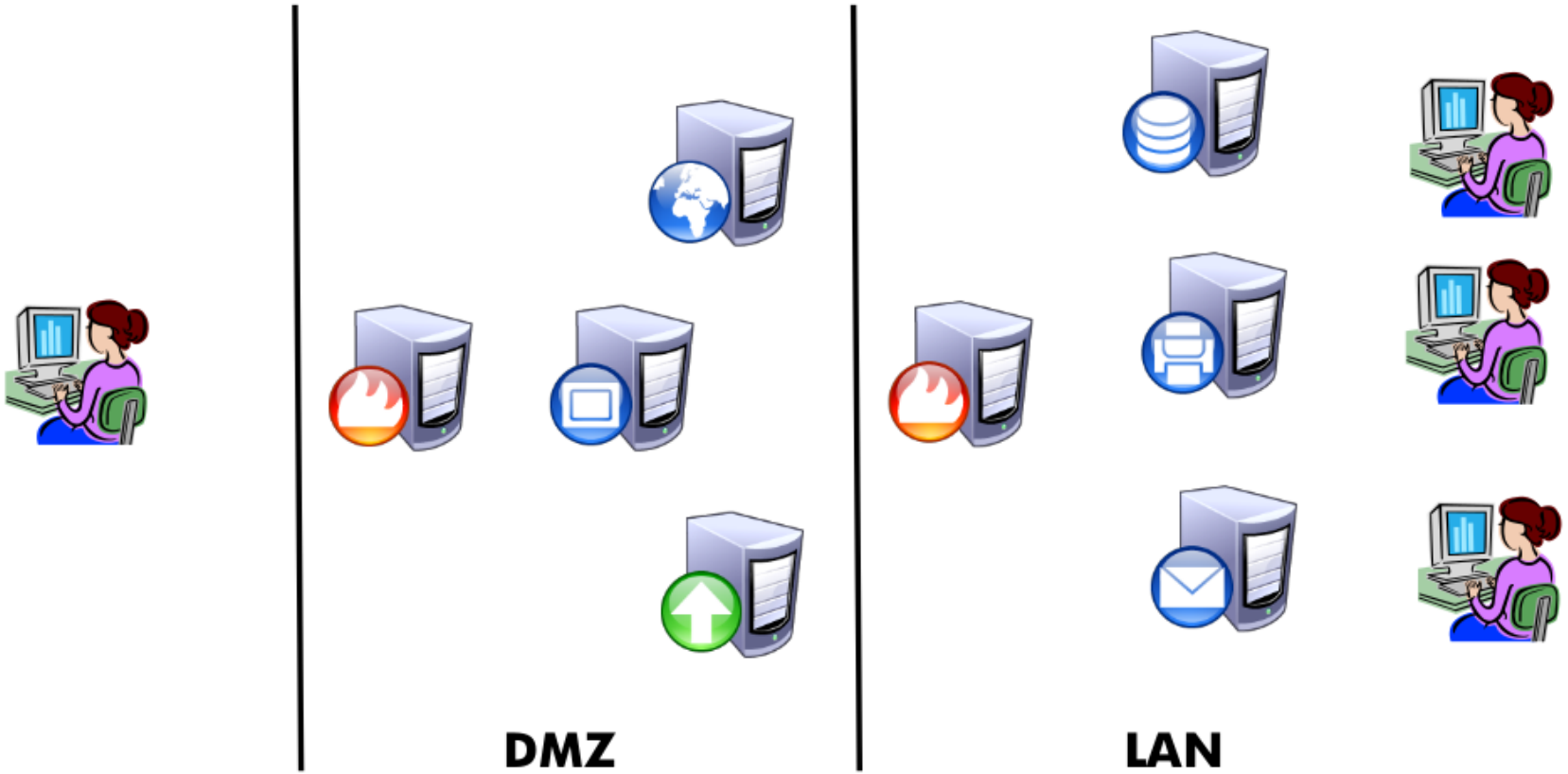
# What is the Cloud? Origins of the Cloud Computing

**Need for Utility Computing**

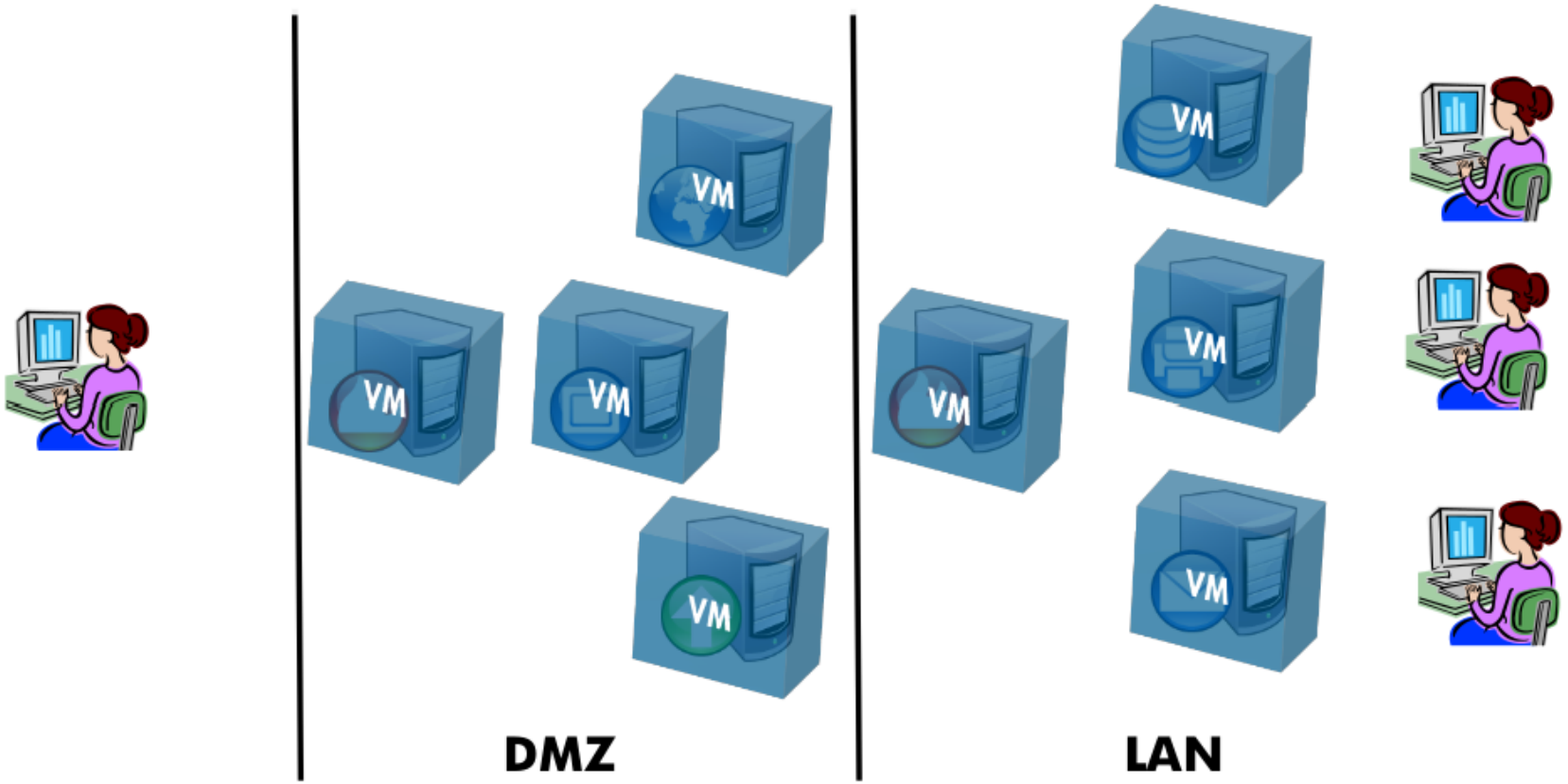
**Underutilised resources**

**Evolution of the virtualisation technology**

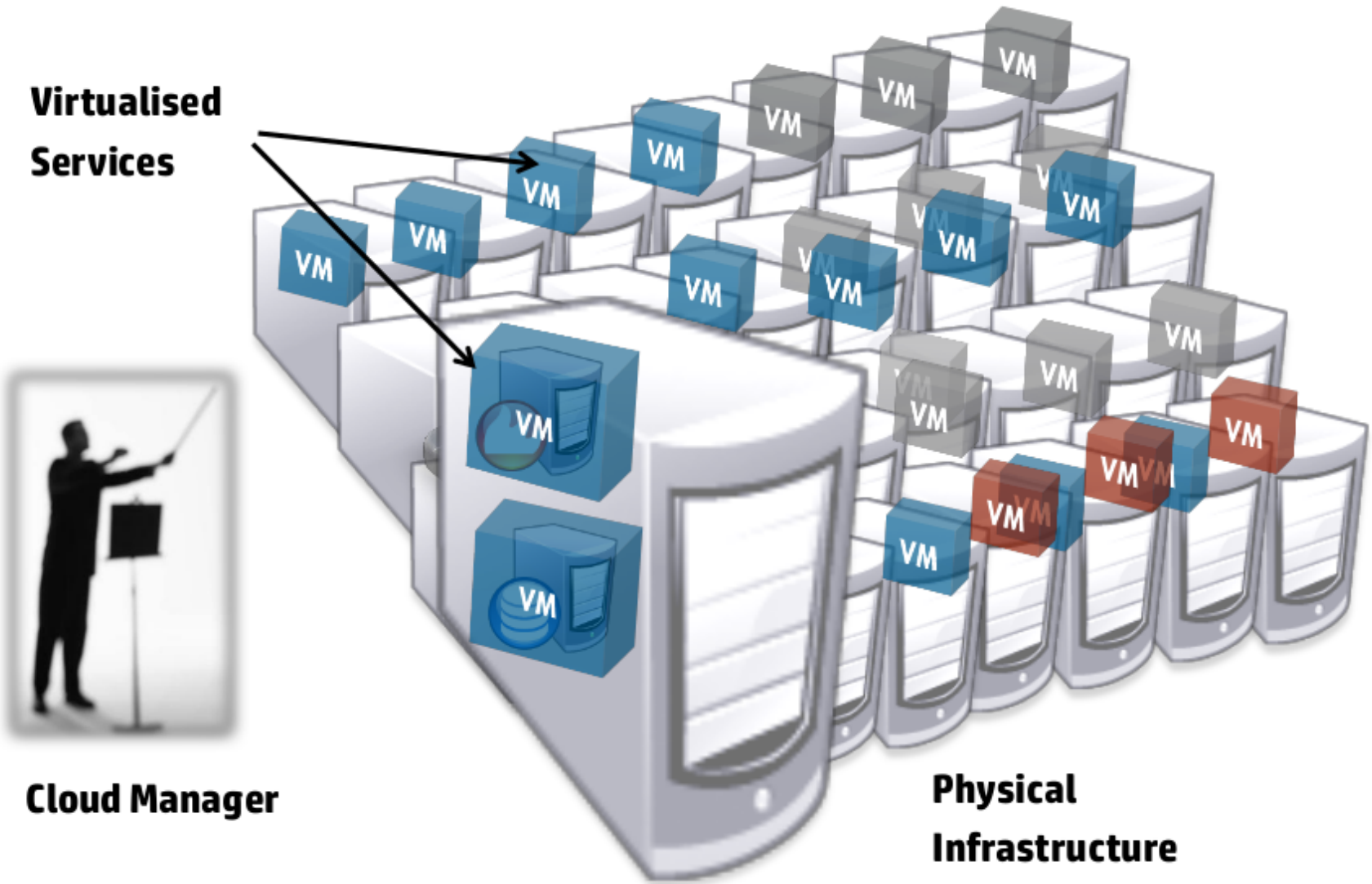
# What is the Cloud? Traditional approach



# What is the Cloud? Virtualised approach



# What is the Cloud? Virtualised approach



# What is the Cloud? Cloud definition

## Definition by NIST[1]

*Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.*

[1] https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

# What is the Cloud? Cloud definition

## Definition by NIST (cont.)

### Essential characteristics:

- **On-demand self-service:** consumer asks for resources (storage/processing/memory/network bandwidth) without human intervention
- **Broad network access:** resource available over internet
- **Resource pooling:** resources location not specified or specified at a higher level (country, region, ...)
- **Rapid elasticity:** Capabilities can be elastically provisioned and released
- **Measured services:** Control and optimization of the resource usage by leveraging a metering capability

# What is the Cloud? Cloud definition

## Definition by Wikipedia[2]

*Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. The term is generally used to describe data centers available to many users over the Internet. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is relatively close, it may be designated an edge server.*

[2] https://en.wikipedia.org/wiki/Cloud_computing

# What is the Cloud? Cloud definition

**Virtualisation**

**Scalability and Elasticity**

**Resources optimisation**

**Pay-per-use**

**SLAs (Service Level Agreement)**

# Table of Contents

# Virtualisation

Hardware Infrastructure

# Virtualisation

Hardware Infrastructure

Guest OS

# Virtualisation

Hardware Infrastructure

Guest OS

Virtual Environment

# Virtualisation

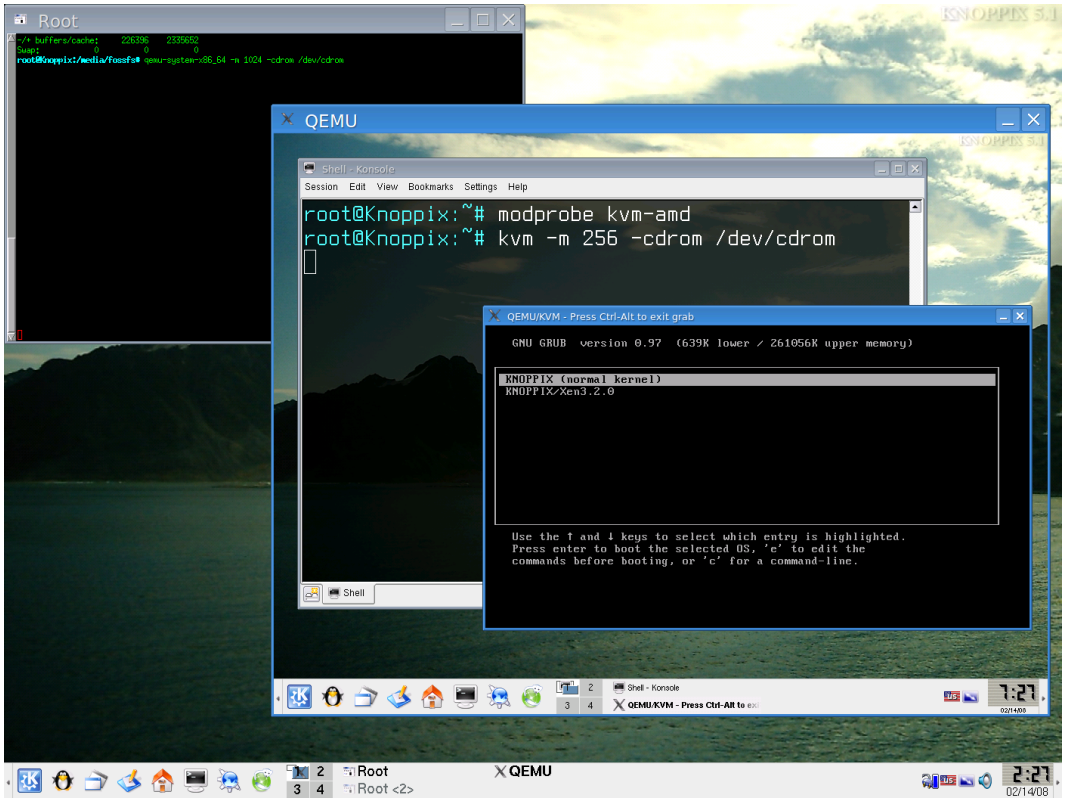
Hardware Infrastructure

Guest OS

Virtual Environment

*Virtualization refers to the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, storage devices, and computer network resources (Wikipedia).*

# Hypervisors

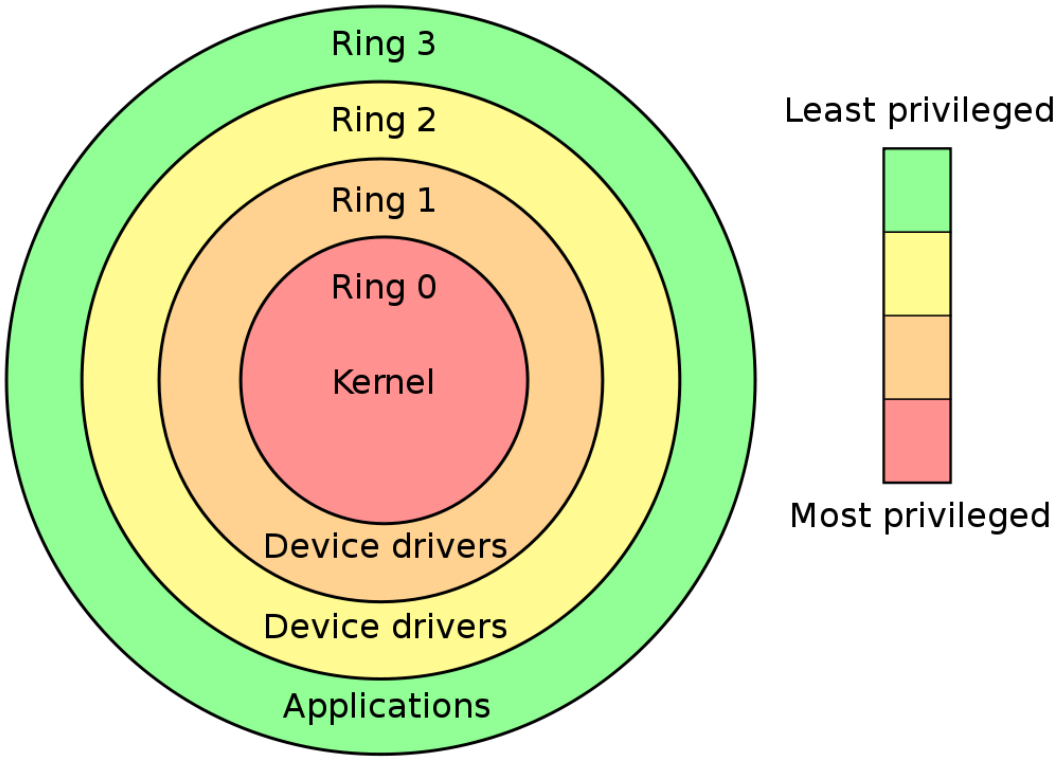






# Virtualisation types

## The Privilege (or Protection) Rings




Privilege rings for the x86 available in protected mode (source: Wikipedia).

# Virtualisation types

**Binary Translation**

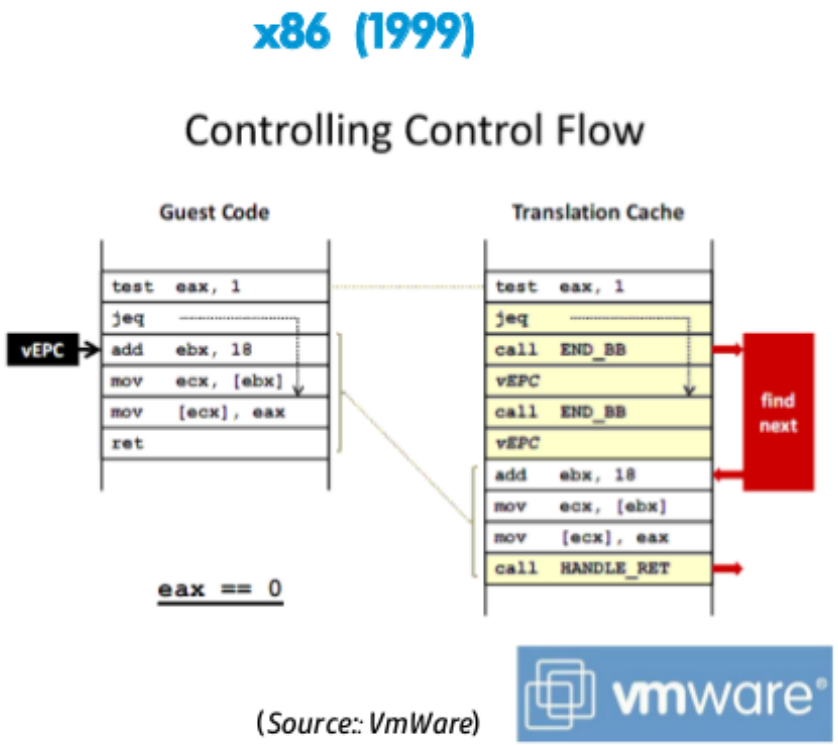
**Paravirtualisation**

**Hardware Virtualisation**

# Virtualisation types

## Binary Translation

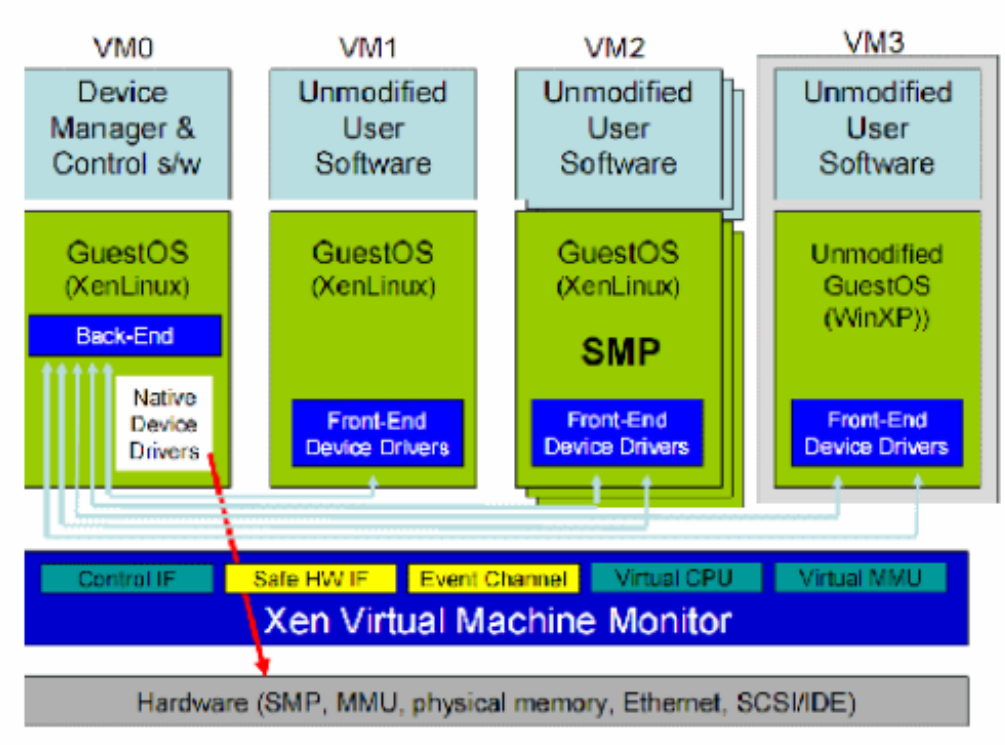
Replacement of “critical” or “dangerous” code to safe code **in runtime**.




- Translation of the binary code that the kernel of a guest OS wants to execute on the fly
- User mode applications are assumed to be safe so they are executed directly as if they were running natively.

# Virtualisation types Paravirtualisation

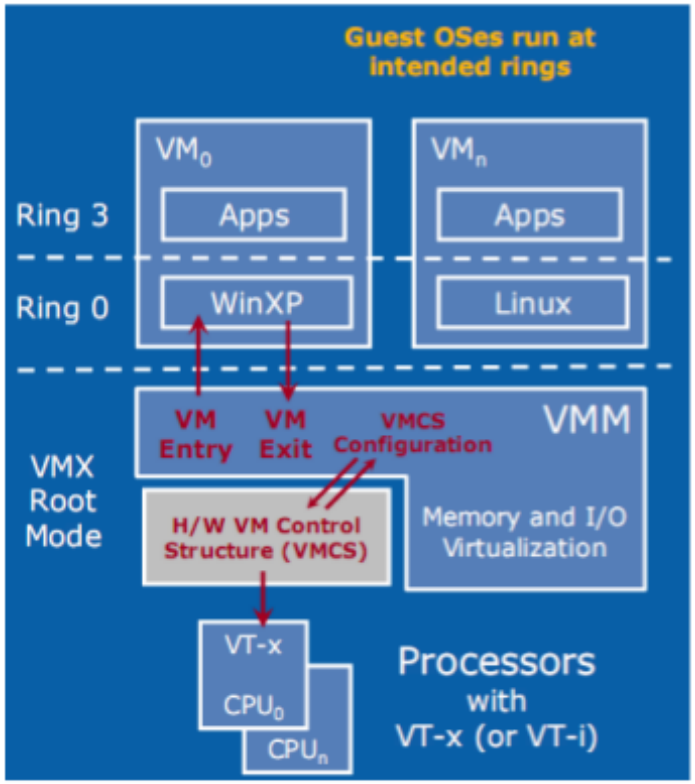
Replacement of “critical” or “dangerous” code to safe code **in the source code**.




(*Source: Xen*)

# Virtualisation types

## Hardware Virtualisation




(Source:: VmWare)

KVM

- Fixes the problem that the x86 instructions architecture cannot be virtualised.
- Traps all exceptions and privileged instructions by forcing a transition from the guest OS to the VMM (*Virtual Machine Manager*): "VMexit".
- Can be considered an improved version of the IBM S/370 virtualization methods.
- Advantage: the guest OS runs at its intended privilege level (ring 0) and the VMM is running at a new ring with an even higher privilege level (Ring -1, or Root mode").
- Efficient: as long as system calls do not involve critical instructions, the guest OS can provide kernel services to the user applications.

# Virtualisation types

## Hypervisor classification

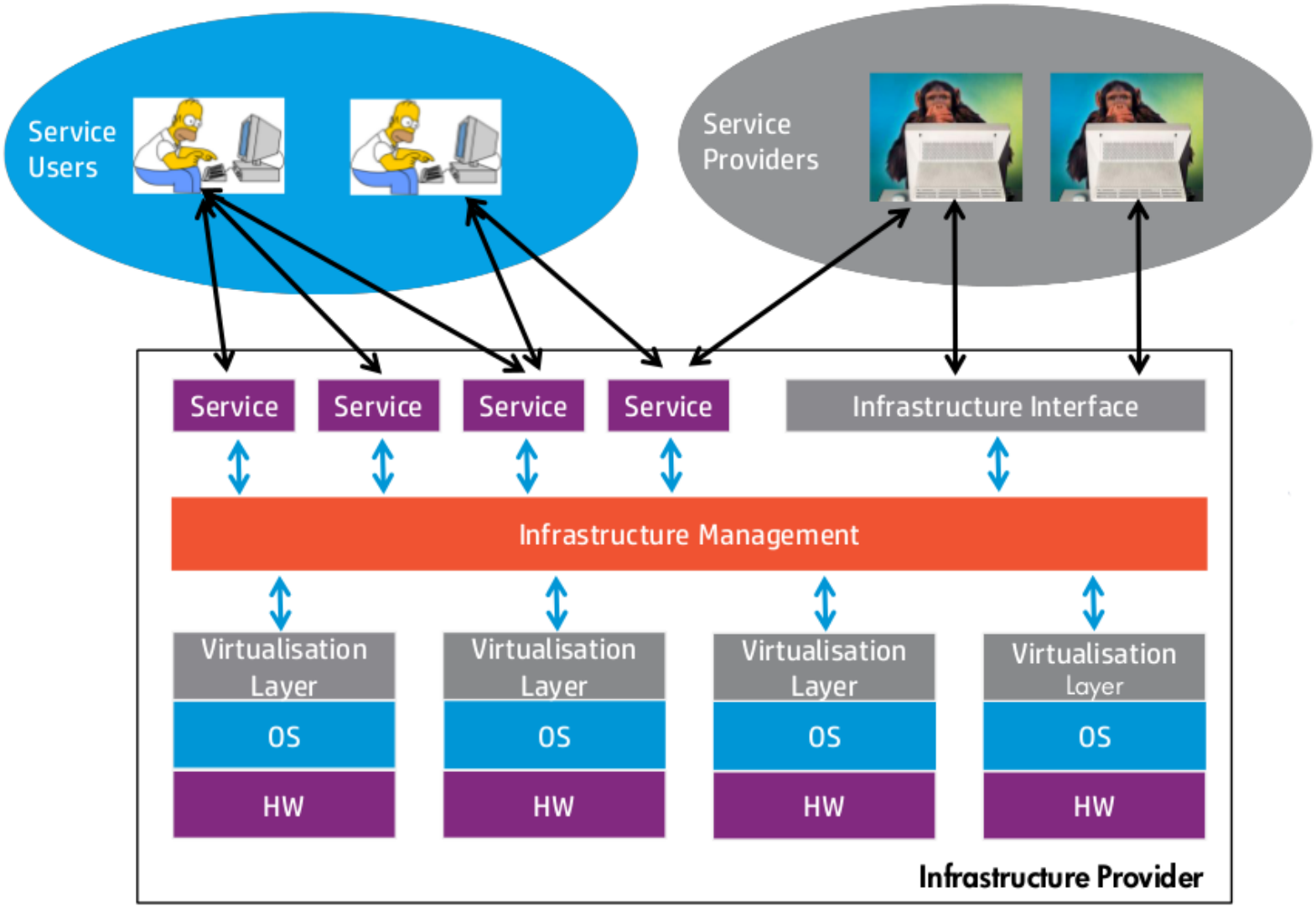
**Type I (Bare Metal)**



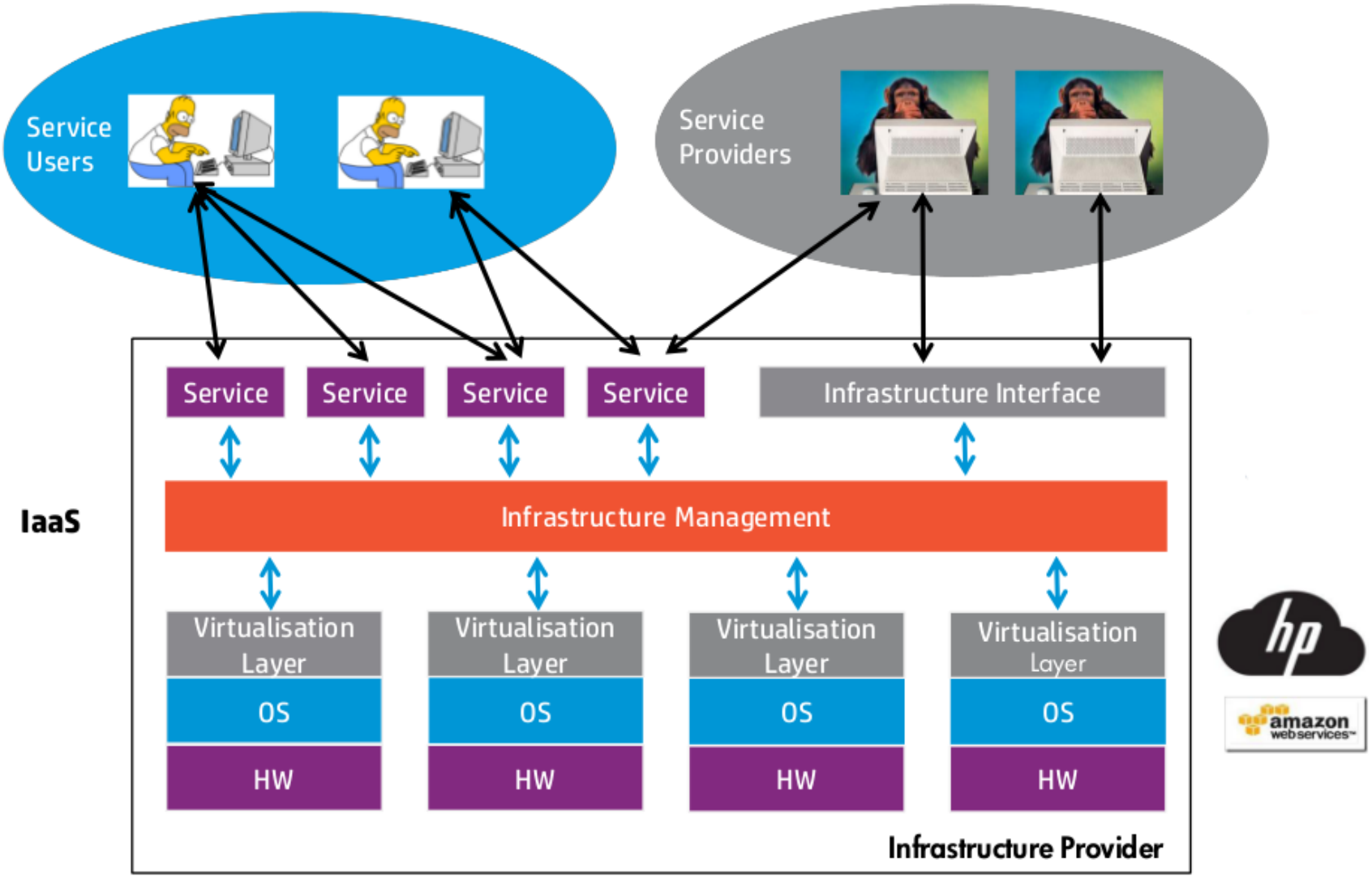

**Type II**


# Table of Contents

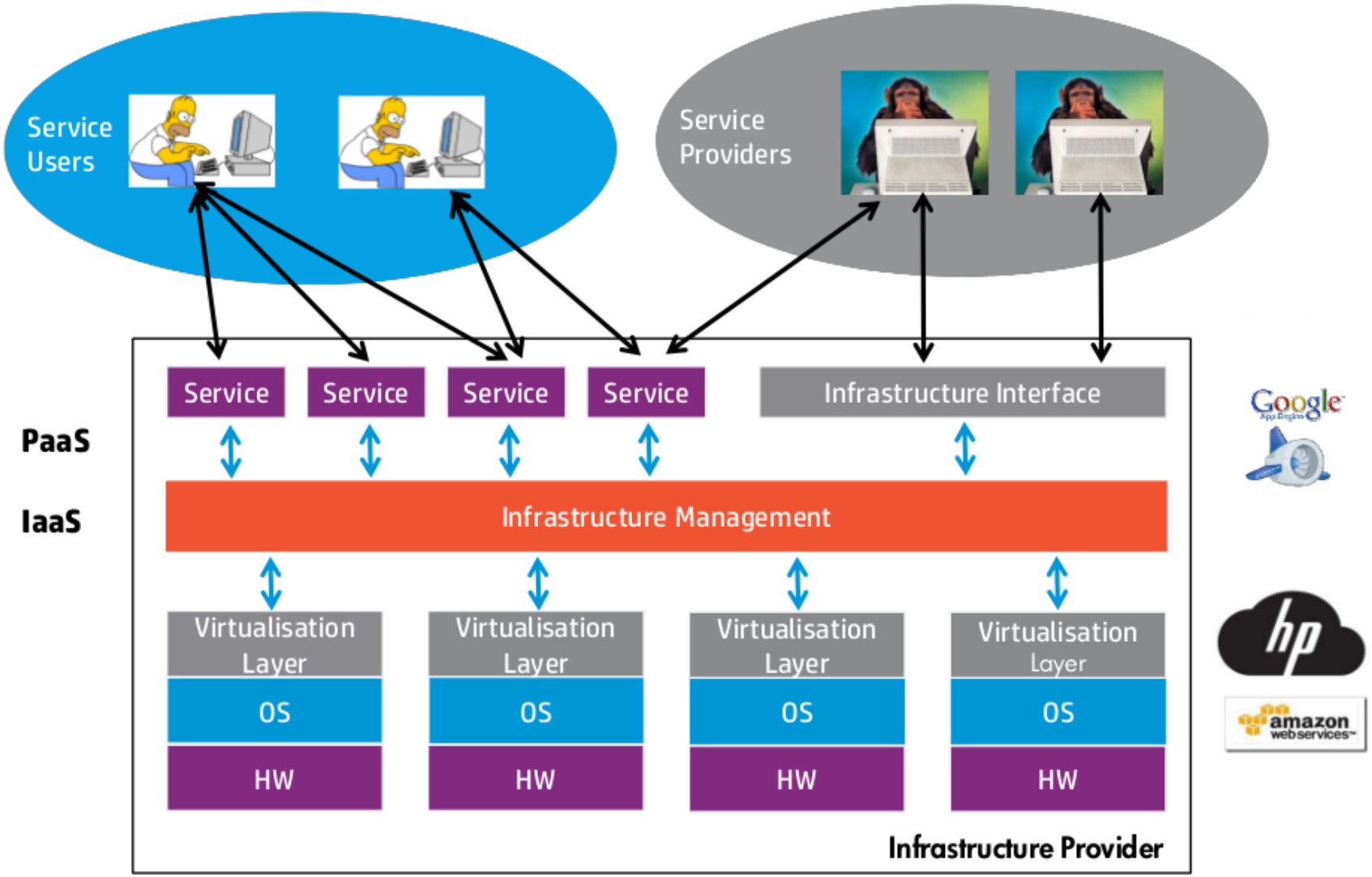
# Cloud Actors - Service models



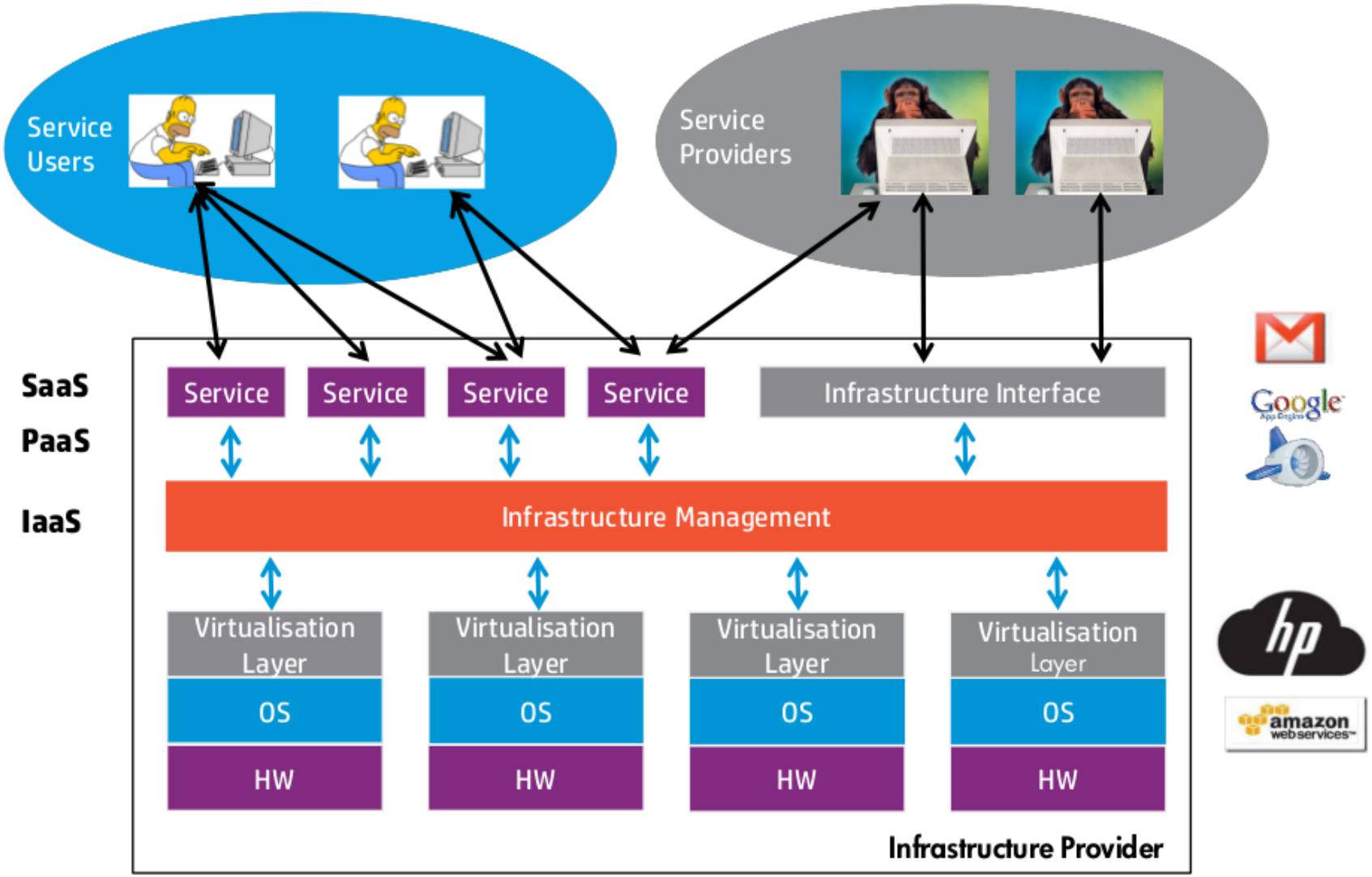
# Cloud Actors - Service models



# Cloud Actors - Service models

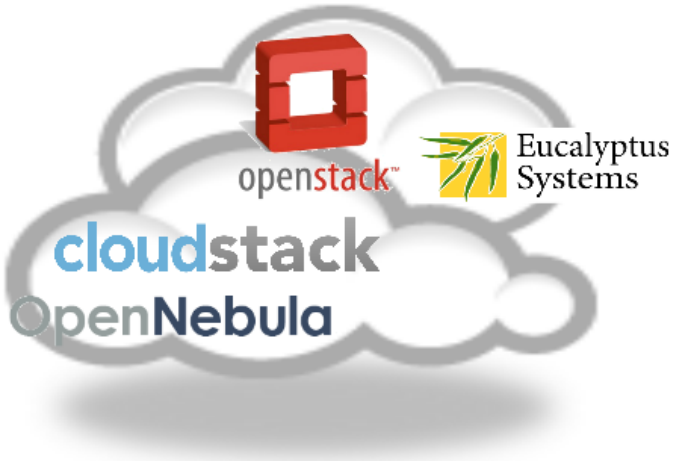


# Cloud Actors - Service models



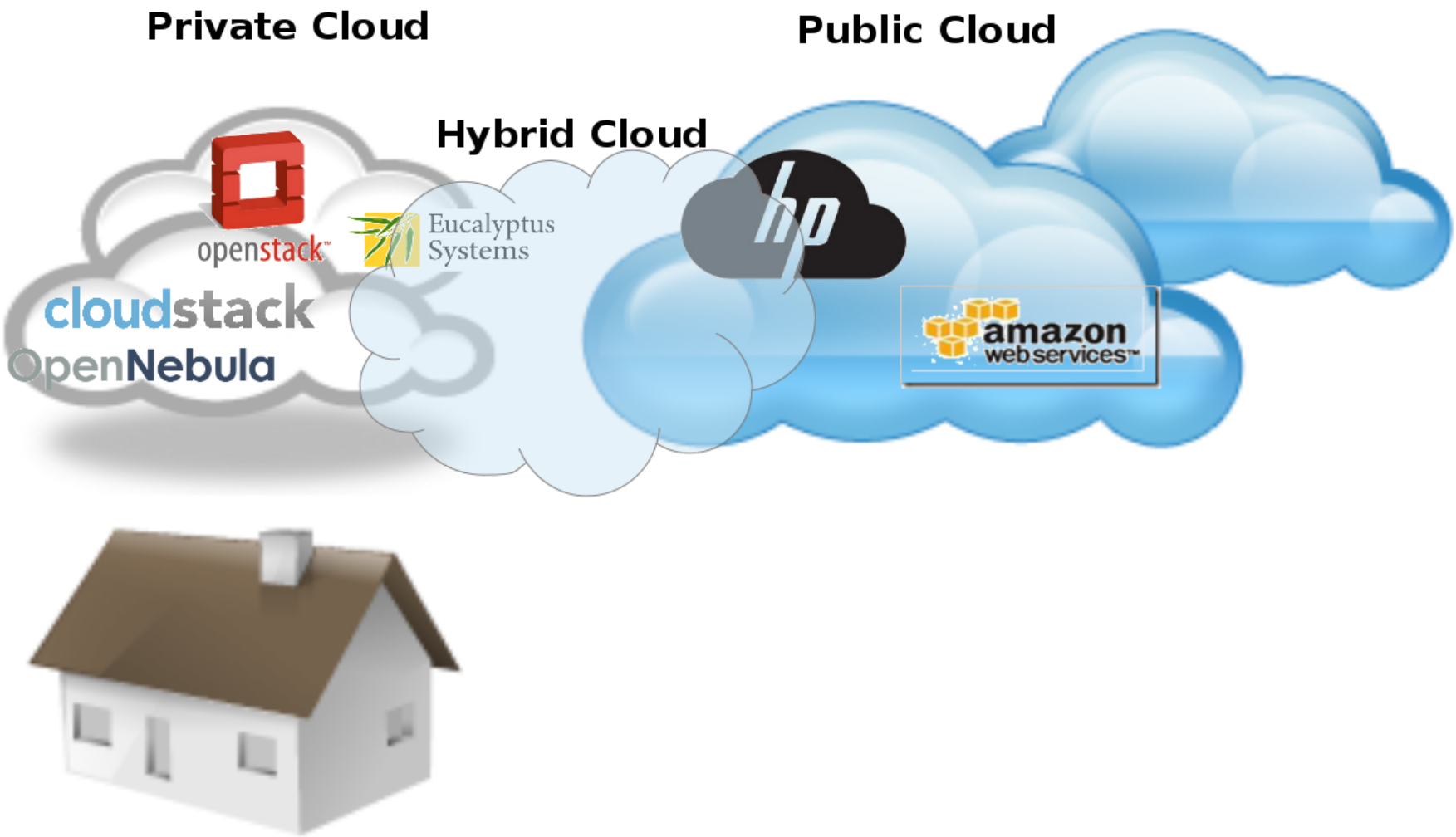
# Deployment models

**Private Cloud**




**Public Cloud**

# Deployment models



# IaaS Challenges

- How can I provision a new VM?

  **Context & Images Management**

- Where do I store disks?

  **Storage**

- How can I configure the network to provide multi-layer services?

  **Networking & VLANs**

- Where should I deploy a VM containing my web server?

  **Monitoring & Scheduling**

- How do I manage my hypervisors?

  **Virtualisation**

- Who has access to the Cloud resources?

  **Users & Role Management**

- How can I manage my distributed infrastructure?

  **Interfaces & APIs**

# Cloud for Data Analytics

## 1 - Cloud and Virtualisation

**Juan Ángel Lorenzo del Castillo**

École Internationale des Sciences du Traitement de l'Information

ING3 - Ingénierie Mathématique et Simulation Informatique

2019-2020


juanangel.lorenzodelcastillo@eisti.eu

Ecole Internationale des Sciences du Traitement de l'Information