CompSci 401: Cloud Computing

Elastic Computing

Prof. Ítalo Cunha

昆山杜克大学
DUKE KUNSHAN UNIVERSITY

Elastic Computing

- Cloud tenants can increase and decrease the amount of resources they lease on demand
- How does a cloud provider support this?

Elastic Computing

- Cloud tenants can increase and decrease the amount of resources they lease on demand
- How does a cloud provider support this?
  - Need fine-grained allocations
    - Amazon EC2 has hundreds of instance types (t2.nano has 1 vCPU and 0.5 GiB RAM)
  - Need to reallocate resources from one tenant to another seamlessly

Elastic Computing

- Cloud tenants can increase and decrease the amount of resources they lease on demand
- How does a cloud provider support this?
  - Need fine-grained allocations
    - Amazon EC2 has hundreds of instance types (t2.nano has 1 vCPU and 0.5 GiB RAM)
  - Need to reallocate resources from one tenant to another seamlessly
  - Allocating physical resources will fail on both accounts

Virtualization for rapid change

- Virtualization is the technology that supports *virtual machines (VM)* and *containers*
- Virtual machines have the following properties:
  - Rapid creation and removal
    - Controlled by software; physical machine and host operating system are not impacted

Virtualization for rapid change

- Virtualization is the technology that supports *virtual machines (VM)* and *containers*
- Virtual machines have the following properties:
  - Rapid creation and removal
    - Controlled by software; physical machine and host operating system are not impacted
  - Physical sharing
    - Multiple VMs can run on the same physical server

Virtualization for rapid change

- Virtualization is the technology that supports *virtual machines (VM)* and *containers*
- Virtual machines have the following properties:
  - Rapid creation and removal
    - Controlled by software; physical machine and host operating system are not impacted
  - Physical sharing
    - Multiple VMs can run on the same physical server
  - Logical isolation
    - VMs running on the same server cannot interact in any way
      - Programs on one VMs are not impacted by programs running in other VMs
      - Data on one VM cannot be accessed by other VMs

How virtualization aids providers

- Software control over creation, removal and modification
  - No human action involved
  - Physical servers
    - Rarely reboot
    - Only need to shut down to save energy
    - Can host multiple virtual machines
  - Virtual machines
    - Can be paused or migrated to a different server (e.g., due to failures)
    - Take seconds to create instead of hours to deploy a physical server

How virtualization aids providers

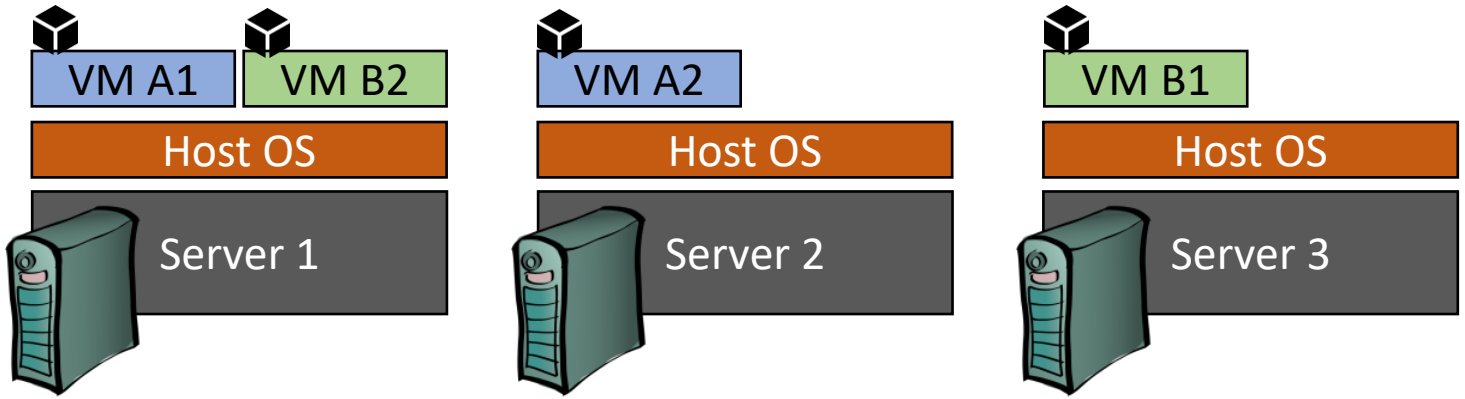
- Isolation
  - Cloud providers can place VMs on any server
  - VMs from multiple tenants can coexist on the same server
    - No interference, for example when a tenant's application misbehaves or crashes
    - No data "leaks" from one VM to another
  - Provider is oblivious to the owner, the applications they run, or the data they handle

How virtualization aids providers

- Software control and isolation provide flexibility
  - Cloud provider can choose where to create VMs
    - Balance load across servers to minimize complaints
    - Maximize server utilization to reduce energy expenses

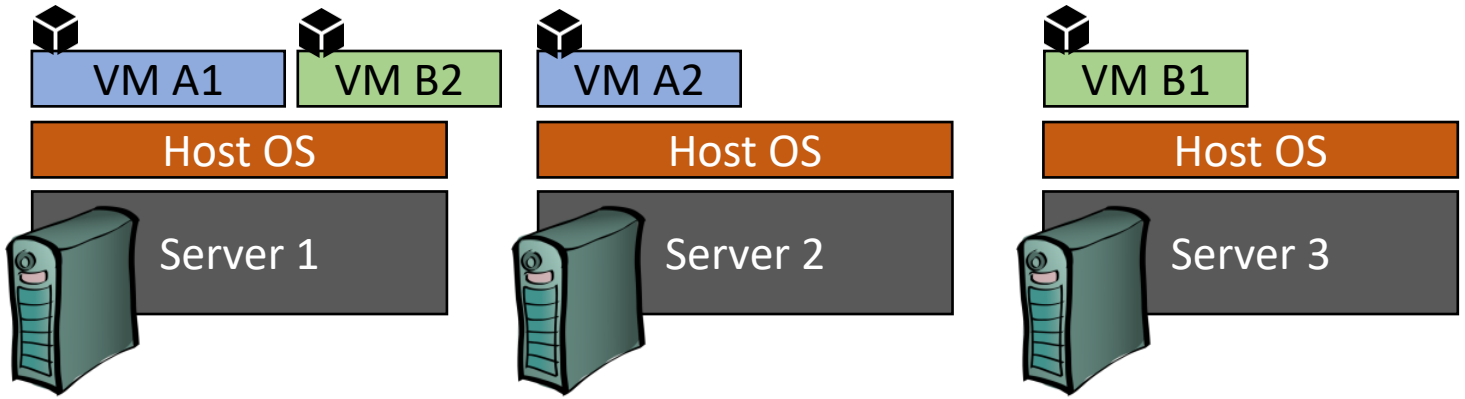
How virtualization aids providers

- Software control and isolation provide flexibility
  - Cloud provider can choose where to create VMs
    - Balance load across servers to minimize complaints
    - Maximize server utilization to reduce energy expenses
  - Cloud provider can migrate VMs
    - VMs can be migrated away from overloaded of failing servers



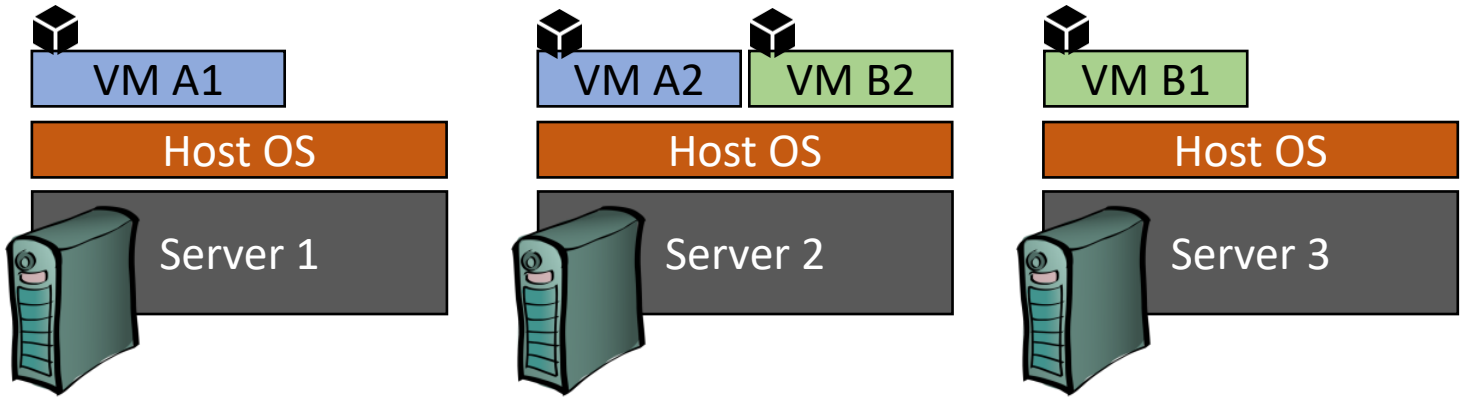
How virtualization aids providers

- Software control and isolation provide flexibility
  - Cloud provider can choose where to create VMs
    - Balance load across servers to minimize complaints
    - Maximize server utilization to reduce energy expenses
  - Cloud provider can migrate VMs
    - VMs can be migrated away from overloaded of failing servers
    - VMs can be migrated away from underloaded servers to power them off and save energy



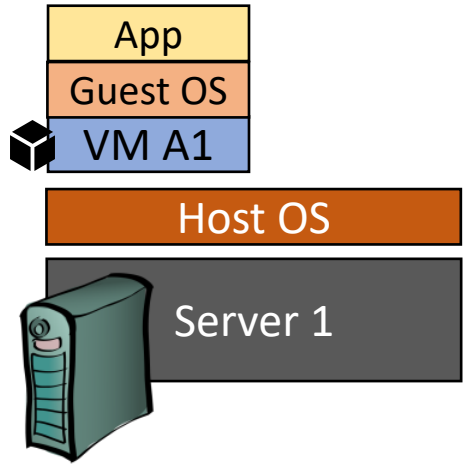
How virtualization aids providers

- Software control and isolation provide flexibility
  - Cloud provider can choose where to create VMs
    - Balance load across servers to minimize complaints
    - Maximize server utilization to reduce energy expenses
  - Cloud provider can migrate VMs
    - VMs can be migrated away from overloaded of failing servers
    - VMs can be migrated away from underloaded servers to power them off and save energy



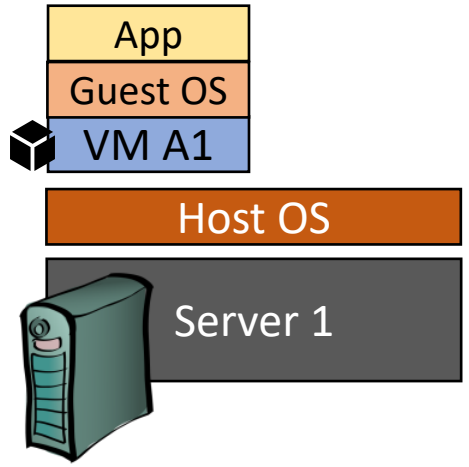
How virtualization aids customers

- A VM is mostly equivalent to a physical server
- Users have complete control over
  - Some hardware (e.g., direct access to network cards or GPUs)
  - Operating system (Linux distribution, Windows)
  - Applications



How virtualization aids customers

- A VM is mostly equivalent to a physical server
- Users have complete control over
  - Some hardware (e.g., direct access to network cards or GPUs)
  - Operating system (Linux distribution, Windows)
  - Applications
- VMs can communicate over the Internet
  - Serve users anywhere



How virtualization aids customers

- Ease of creating and deploying new services
  - Software can be deployed as soon as it is developed
  - Cloud providers have tooling to automate or ease software deployment, possibly to multiple VMs
- Rapid scaling of a service
  - Service capacity can be increased or decreased as a function of demand by creating or removing VMs
- Safe and rapid testing of new software
  - Organizations can test new software before deployment in a controlled but equivalent infrastructure using VMs

CompSci 401: Cloud Computing

Business Models

Prof. Ítalo Cunha

昆山杜克大学
DUKE KUNSHAN UNIVERSITY

Cloud provider structure

- How to generate revenue?
- What is the best structure for companies in the cloud industry?
- Should providers handle or delegate management of
  - Physical facilities (air conditioning, power, security)?
  - Computing facilities (network, servers, software)?

Classes of cloud services

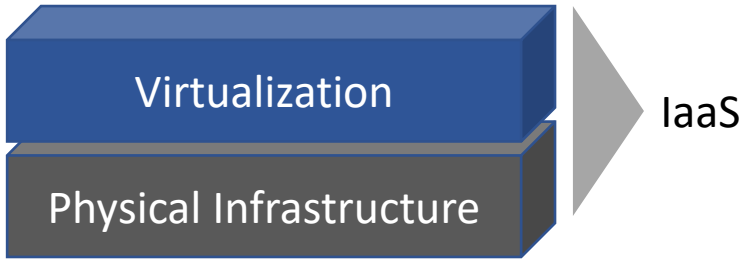
- Loose definitions and some overlap
- Help delineate what each company handles and what tenants handle

Infrastructure as a Service (IaaS)

- Physical resources
  - Building, power, air conditioning
  - Servers, network equipment, data storage, load balancers

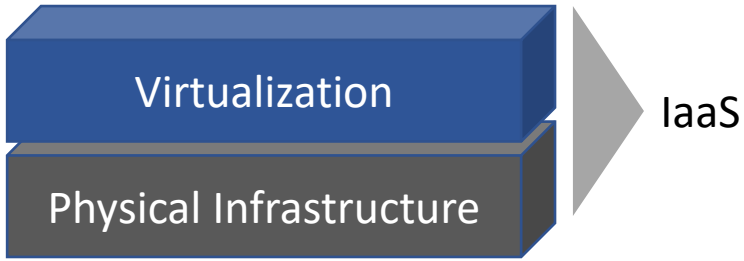
Infrastructure as a Service (IaaS)

- Physical resources
  - Building, power, air conditioning
  - Servers, network equipment, data storage, load balancers
- Virtualized resources equivalent to physical resources
  - Virtual machines, virtual networking, block storage, load balancers



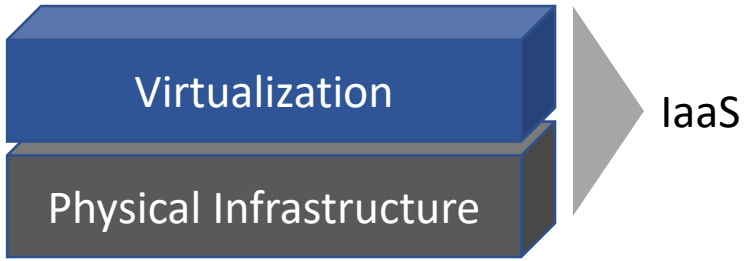
Infrastructure as a Service (IaaS)

- Physical resources
  - Building, power, air conditioning
  - Servers, network equipment, data storage, load balancers
- Virtualized resources equivalent to physical resources
  - Virtual machines, virtual networking, block storage, load balancers
- Complete control
  - Users have full control of operating systems and applications



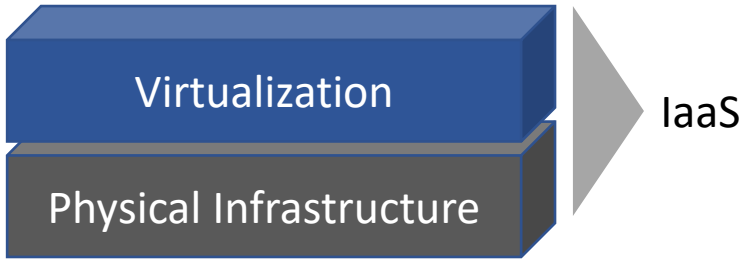
Infrastructure as a Service (IaaS)

- Physical resources
  - Building, power, air conditioning
  - Servers, network equipment, data storage, load balancers
- Virtualized resources equivalent to physical resources
  - Virtual machines, virtual networking, block storage, load balancers
- Complete control
  - Users have full control of operating systems and applications
- IaaS provider can provide complementary services
  - Autoscaling
  - Disk backups



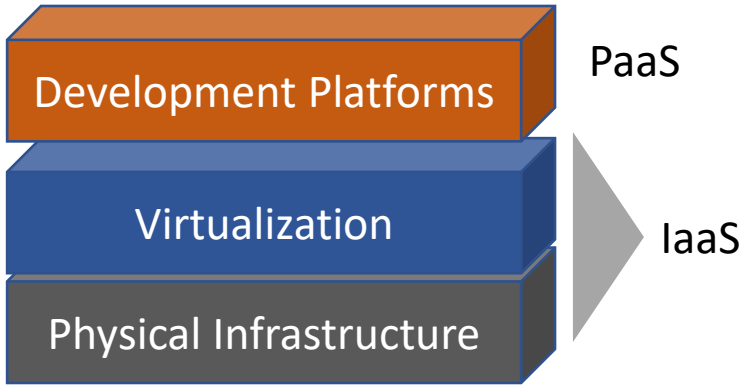
Infrastructure as a Service (IaaS)

- Physical resources
  - Building, power, air conditioning
  - Servers, network equipment, data storage, load balancers
- Virtualized resources equivalent to physical resources
  - Virtual machines, virtual networking, block storage, load balancers
- Complete control
  - Users have full control of operating systems and applications
- IaaS provider can provide complementary services
  - Autoscaling
  - Disk backups
- Examples: AWS, GCE, Azure, DigitalOcean, Linode



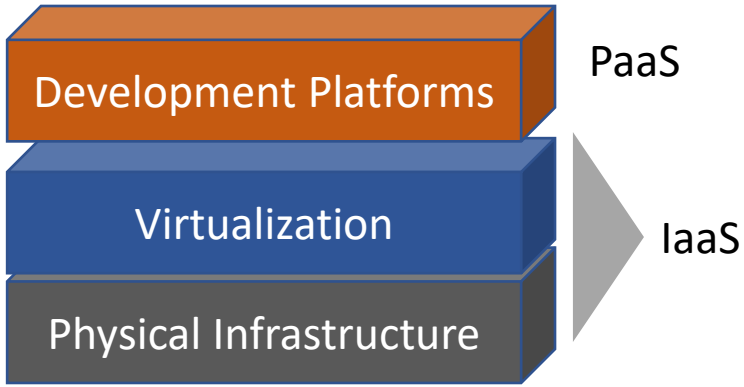
Platform as a Service (PaaS)

- Facility to allow customer to build and deploy software
  - Provider manages the facility and infrastructure for the customer
  - Provide compilers, libraries, middleware, automation of common tasks
  - Customers can focus entirely on their application



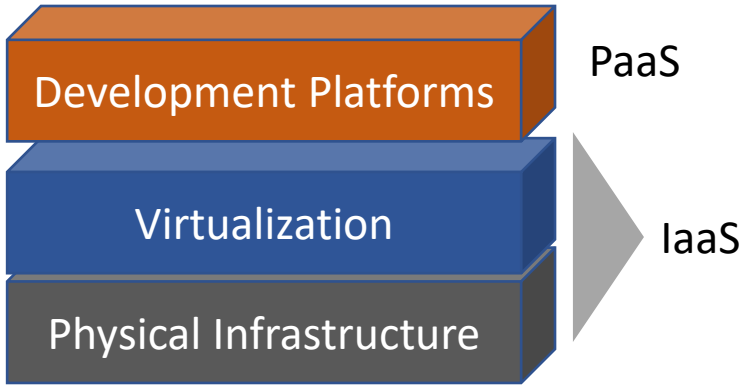
Platform as a Service (PaaS)

- Facility to allow customer to build and deploy software
  - Provider manages the facility and infrastructure for the customer
  - Provide compilers, libraries, middleware, automation of common tasks
  - Customers can focus entirely on their application
- Customer does not have control of lower layers
  - Operating system is fixed
  - Options at each layer of the application may be limited



Platform as a Service (PaaS)

- Facility to allow customer to build and deploy software
  - Manage the facility and infrastructure for the customer
  - Provide compilers, libraries, middleware, automation of common tasks
  - Customers can focus entirely on their application
- Customer does not have control of lower layers
  - Operating system is fixed
  - Options at each layer of the application may be limited
- Examples: Google App Engine, Heroku, Windows Azure, AWS Elastic Beanstalk



Example: AWS Elastic Beanstalk

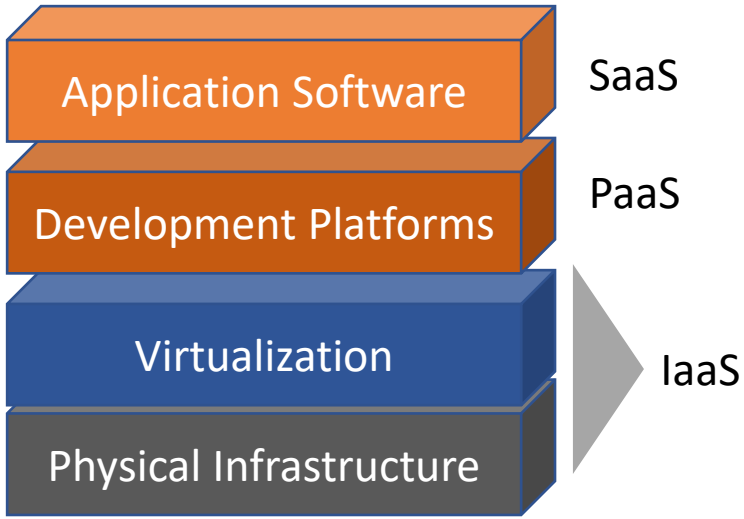
- From the introductory video:
  - “easy to use service to deploy and scale Web servers”
    - Specific type of apps
  - “using Java, .Net, Node.js, Python”
    - Selection of languages and frameworks
  - “familiar services like Nginx and IIS”
    - Limitations on infrastructure
  - “without you spending time managing and configuring infrastructure”
    - Given the constraints, PaaS provider manages everything but the application

PaaS limitations and concerns

- Lack of flexibility
  - The set of available functionality may be inadequate
    - Language or middleware of choice may not be supported
    - Legacy applications may be left out entirely
- Vendor lock-in
  - There are some common standards, but solutions may differ across PaaS providers

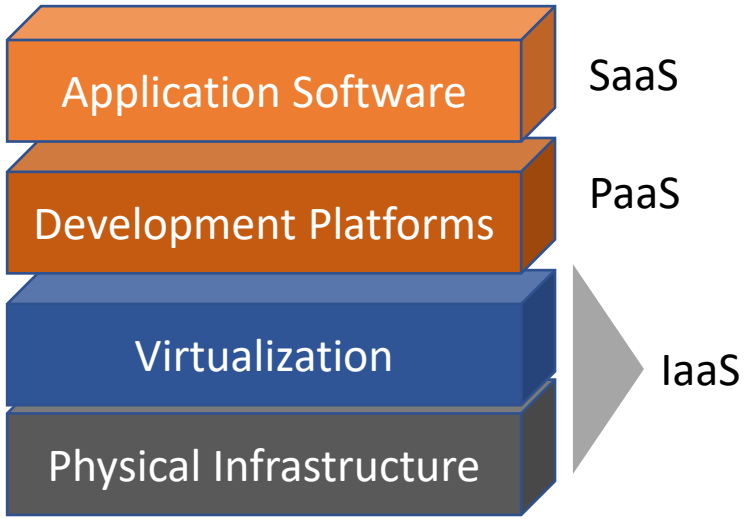
Service as a Service (SaaS)

- Subscription model where customers pay a fee
- SaaS offerings rely on IaaS to support arbitrarily many customers
- Universal access
  - Accessed from desktops, laptops, tablets, phones, or even servers
  - Web apps
- Guaranteed synchronization
  - Even when data are modified by multiple devices
  - All state kept in the cloud
  - Identical data across all devices, no “conflicts”
- High availability
  - Reliance on IaaS ensures redundancy and resilience against failures



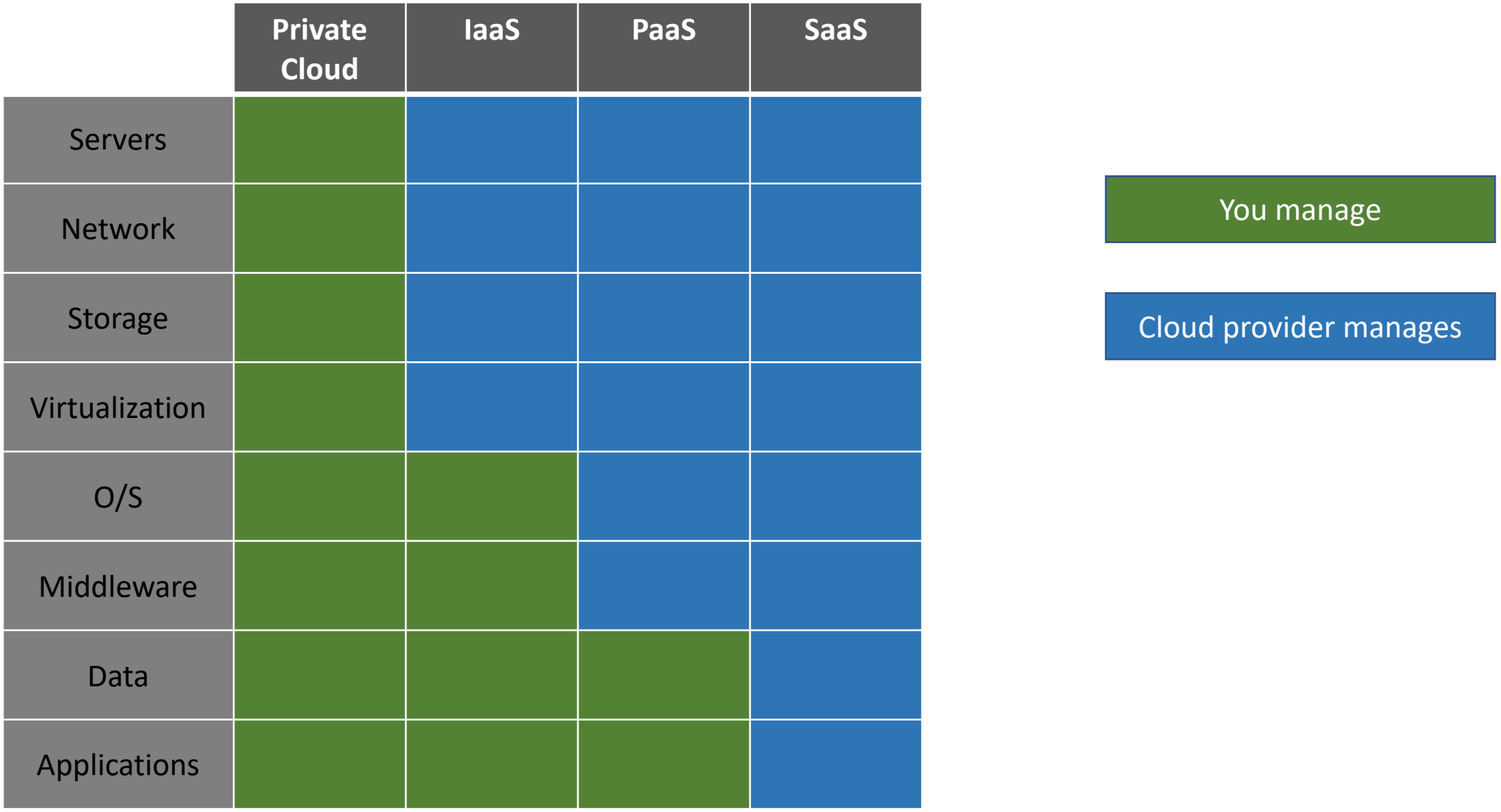
Service as a Service (SaaS)

- Examples:
  - Office 365
  - Evernote
  - DotA Plus
  - Todoist
  - Google Workspace
  - Dropbox
  - Salesforce
  - Zoom
- Applications run on the browser or cell phone
  - Also called “thin clients” because data and heavy computation remain on the cloud



SaaS limitations and concerns

- Set features
  - The set of features and development effort is set by the provider
  - Limited customization
  - Lack of control
- Interoperability
  - SaaS applications may not be designed to adhere to open standards
  - May be hard or impossible to integrate with in-house applications
- Vendor lock-in
  - It may be hard to get your data out of the application


Private Cloud
IaaS
PaaS
SaaS
Servers
Network
Storage
Virtualization
O/S
Middleware
Data
Applications
You manage
Cloud provider manages