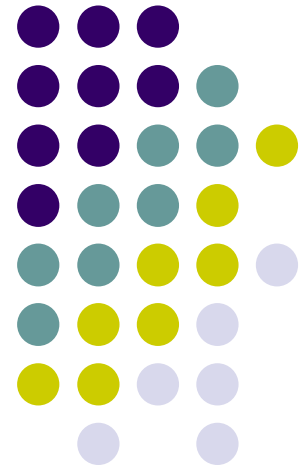


SIAD

Serviços e Infraestruturas de Alto Desempenho

An Introduction to Cloud Computing





Definition of Cloud Computing (I)

“What’s in a name? That which we call a rose by any other name would smell as sweet” (Shakespeare)

Gartner

*“Cloud computing is a style of computing where **massively scalable** IT-related capabilities are provided **as a service** across the Internet to multiple external customers”*



*“Cloud computing: **a pool of abstracted, highly scalable, and managed infrastructure** capable of hosting end-customer applications **and billed by consumption**”*



WIKIPEDIA
The Free Encyclopedia

*“Cloud computing is Web-based processing, whereby **shared resources, software, and information** are provided to computers and other devices (such as smartphones) **on demand** over the Internet.”*



Definition of Cloud Computing (II)

“What’s in a name? That which we call a rose by any other name would smell as sweet” (Shakespeare)



**National Institute of
Standards and Technology**
U.S. Department of Commerce

*“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 [delivery] models**, and four **deployment models**.”*



Service Delivery Models

Software as a Service (SaaS)

- **Application**
email, web-based applications...

Platform as a Service (PaaS)

- **Platform/Service**
Web hosting, blog sites, databases, java runtime

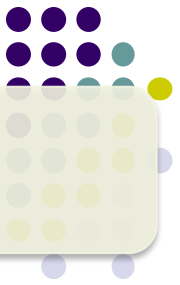
Infrastructure as a Service (IaaS)

- **Operating System**
(virtualized) servers, networking, raw storage...

physical hardware

Software as a Service (SaaS)

- **Application**
email, web-based applications...



“The capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email).

The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.” [NIST]

- Widely used by everyone (e.g., Gmail)
- Frequently “free”, easy to use and with good consumer adoption
- Proven business models (at least for some cases!)
- Uniform service for everyone (despite some customization), without significant flexibility
- Frequently without service level guarantees (especially when provided for free, but also for paid services)

Platform as a Service (PaaS)

- **Platform/Service**

Web hosting, blog sites, databases, java...

“The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider.

The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.” [NIST]

- Developers can upload a configured application and it “runs” within the platform’s framework
- Restricted to the platform’s ability only; sometimes dependent on the specific cloud provider
- Examples: application development, data, workflow middleware, security services (Single Sign-On, AAA), database management, directory services

Infrastructure as a Service (PaaS)

- **Operating System**

(virtualized) servers, networking, raw storage

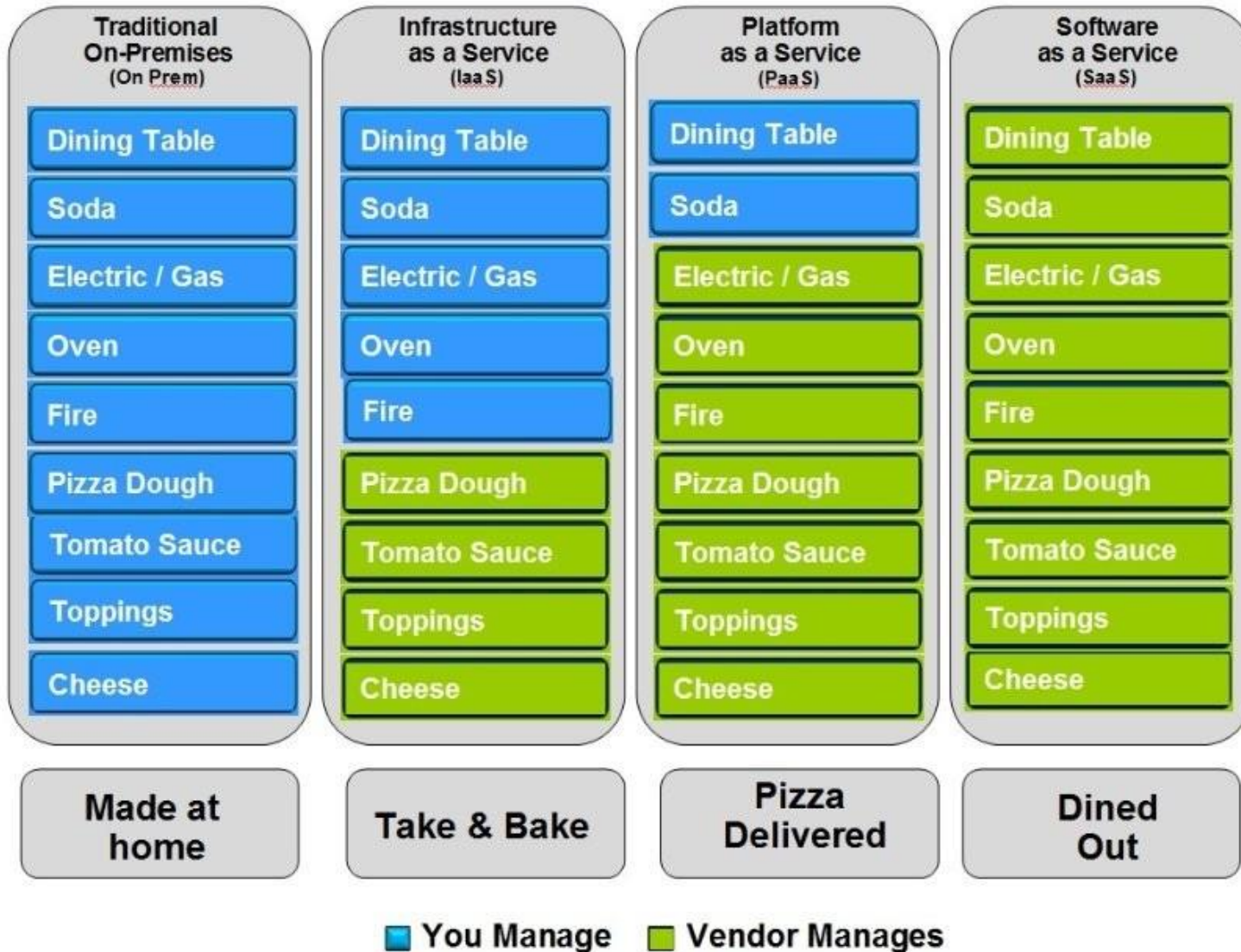
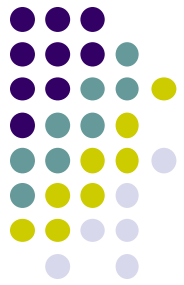
“The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls)” [NIST]

- Offers full control (except for physical hardware resources)
- Can be complex to build, manage and maintain
- Examples: virtualized servers and storage

It's all about food...

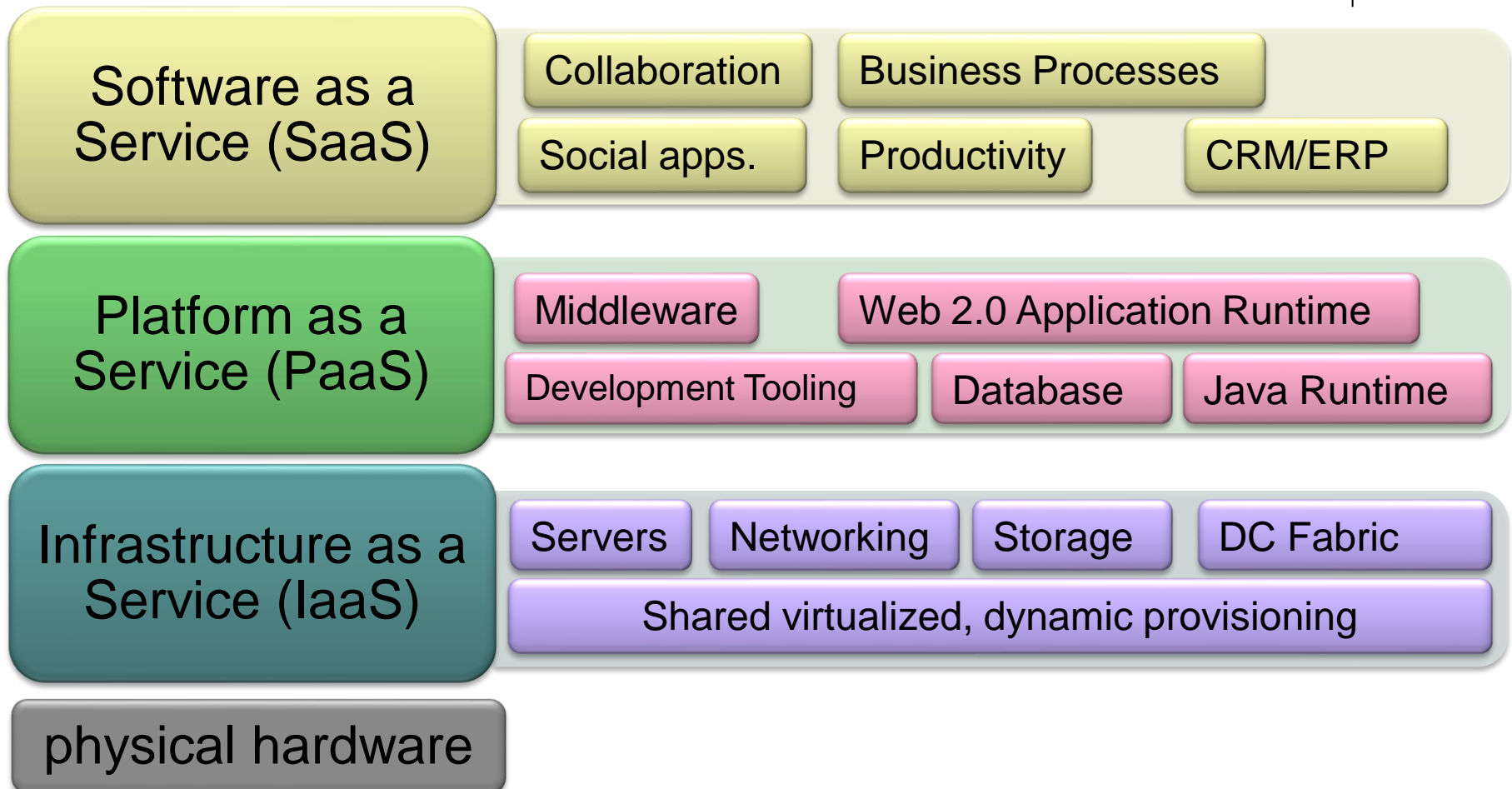
Pizza as a Service



Source: <http://www.episerver.com/blog/blog-start/fred-bals/pizza-as-a-service/>
(based on a LinkedIn post from Albert Barron)
<https://www.linkedin.com/pulse/20140730172610-9679881-pizza-as-a-service/>



Service Delivery Models



[Break Time: Stuff-as-a-Service]



- SaaS (Software)
- PaaS (Platform)
- IaaS (Infrastructure)
- DaaS (Desktop)
- MaaS (Monitoring)
- DNSaaS (DNS)
- MOBaaS (Mobility)
- DBaaS (Database)
- EPCaaS (Evolved Packet Core – 4G LTE networks)
- RANaaS (Radio Access Network)
- LBaaS (Load Balancing)
- AAAaaS (Authentication, Authorization & Accounting)
- CDNaaS (Content Distribution Network)
- SLAaaS (Service Level Agreement)
- DSSaaS (Digital Signage)
- IMSaaS (3GPP IP Multimedia Service – 3G networks)
- RCBaaS (Rating, Charging & Billing)

*All these acronyms were taken
from a single research project!*



...XaaS (anything as a service)

#!&*!%-as-a-Service?!



Cloud Deployment Models

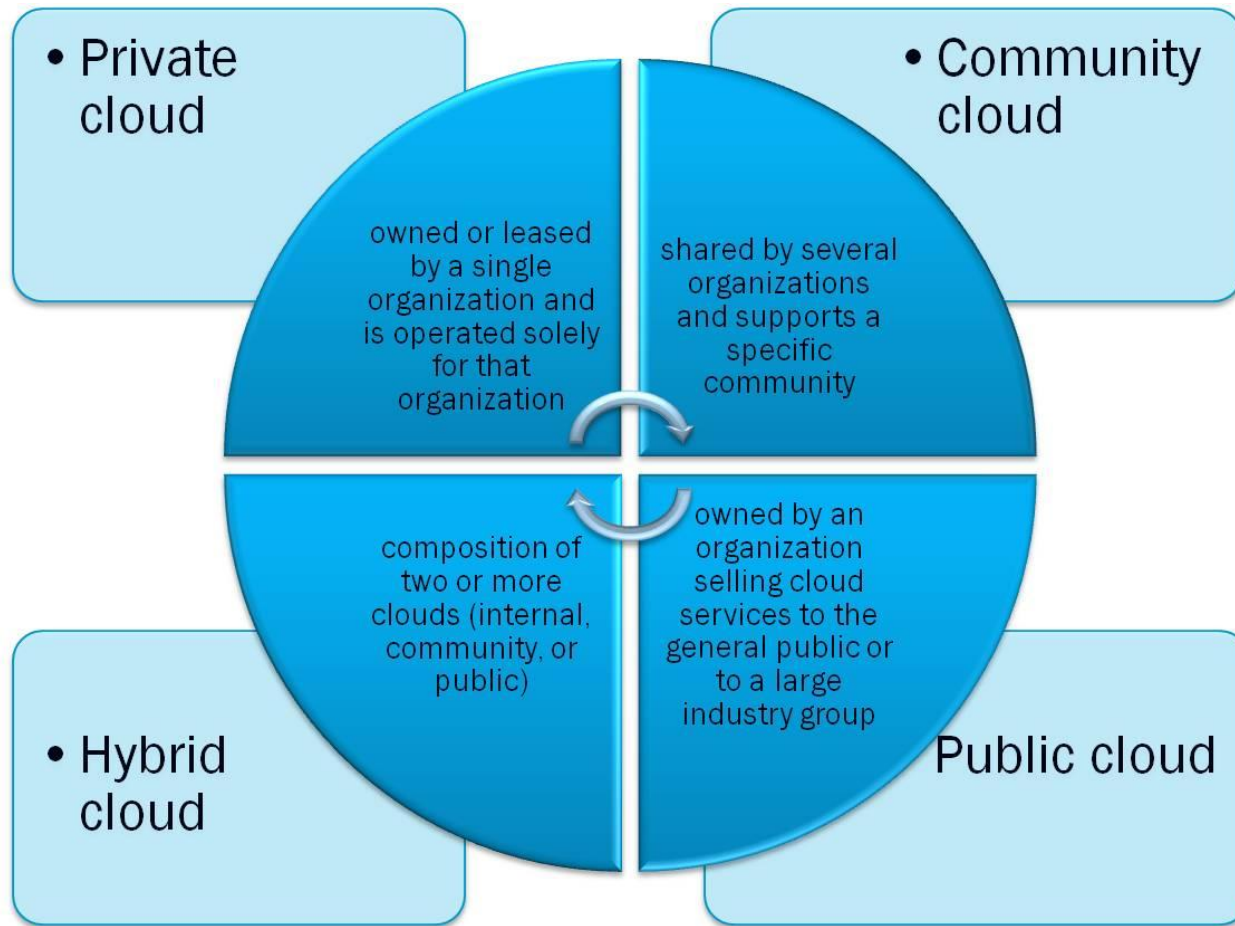


Image Credit: www.center4cloud.nl



Deployment Model – Private Cloud

“The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.”
[NIST]

- Owned and managed by the enterprise
(keep the data center but change the way it is used)
- Access is limited to enterprise and partner networks
- High degree of control, privacy and security
- Enables business to more easily customize service
- Accessed from “inside the firewall”
- Still requires significant upfront capital expenditure (CAPEX) and operational expenditure (OPEX)
- Elasticity is bounded by existing hardware resources

For instance: <https://helpdesk.dei.uc.pt/configuration-instructions/cloud2-dei/>
(DEI account required)



Deployment Model – Public Cloud

“The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.” [NIST]

- Owned and managed by cloud service provider
- Delivers a select set of software, application or infrastructure services on a “pay per use” basis
- Highly standardized (despite the existence of proprietary frameworks from some cloud providers)
- More limited customization options
- Accessed from “outside the firewall”
- Access to/from the customer enterprise usually over VPNs
- Potentially less secure access from the Internet
- Still less upfront capital expenditure (CAPEX)
- Elasticity is typically higher



Deployment Model – Community Cloud

“The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.” [NIST]

- Owned and managed by the specific community involved
- Shares some of the advantages of private clouds (control, flexibility, customization) with some of the advantages of public clouds (e.g., more elasticity, less costs)
- CAPEX and OPEX costs are still relevant but shared among the community.
- Potentially less secure than private cloud, but still more secure than public cloud (*though your mileage may vary; it depends more on the quality of cloud management services than on the deployment model*)



Deployment Model – Hybrid Cloud

“The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).” [NIST]

- Takes advantage of both public and private clouds
- Provides a cost-effective solution for demand peaks: private clouds handle typical demand loads, and public clouds are used to provide additional resources for handling seasonal peak demand
 - Some other examples:
<https://www.netapp.com/us/info/what-is-hybrid-cloud.aspx>
- Useful also in disaster recovery scenarios: the private cloud can be replicated in a public cloud in the case of catastrophic failure
 - Se for instance <https://www.youtube.com/watch?v=1heZUPrGStU>
(apologies for some commercial content)



What defines Cloud Computing as... ...Cloud Computing?

“Dear Editor:

I have 3 servers in my room, running 24/7 and hosting 10 Virtual Machines (VMs) with several applications!

Do I own and operate a private cloud infrastructure?”

Cloud computing is not (just) virtualization of resources – it is a new paradigm that requires new tools and technologies.

Key characteristics of Cloud Computing:

- **On-demand Self Service**
- **Broad Network Access**
- **Resource Pooling**
- **Rapid Elasticity**
- **Measured Service**





#1 On-Demand Self Service

“A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.” [NIST]

- A consumer signs-in to create a Gmail or Dropbox account
- A web designer creates a new website and hosts it in a CSP
- An enterprise IT staff creates a new DNS server
- An enterprise IT staff creates a new “empty” virtual machine for testing a new application

Automated process (e.g., menu-driven)





#2 Broad Network Access

“Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).” [NIST]

Many services are directly available over the internet, anywhere, anytime.

Enterprise services may still be restricted to the corporate network, for security reasons, but they are still accessible “anywhere” (inside the corporate LAN/VPN) using protocols such as HTTP and TCP/IP.

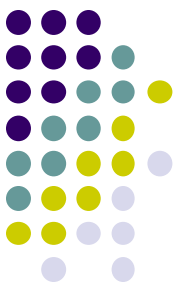




#3 Resource Pooling

“The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.” [NIST]

Providers can create a sense of infinite or immediately available resources by controlling resource adjustments, so that customers change their levels of service at will without being subject to limitations of physical or virtual resources.

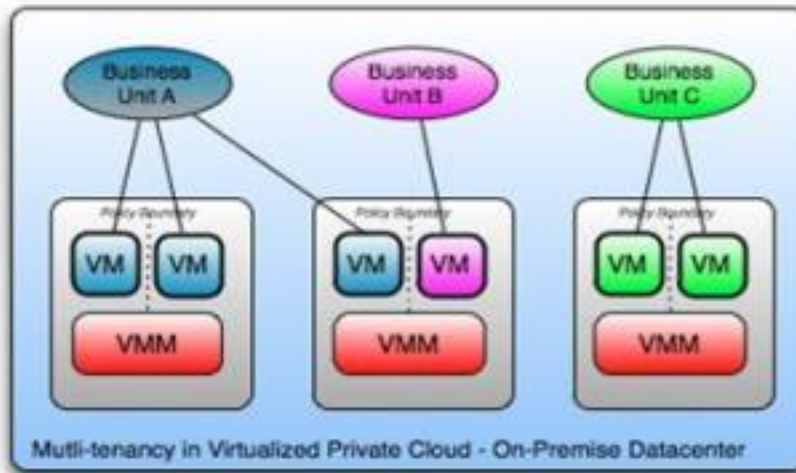


[Break Time: multi-tenancy]

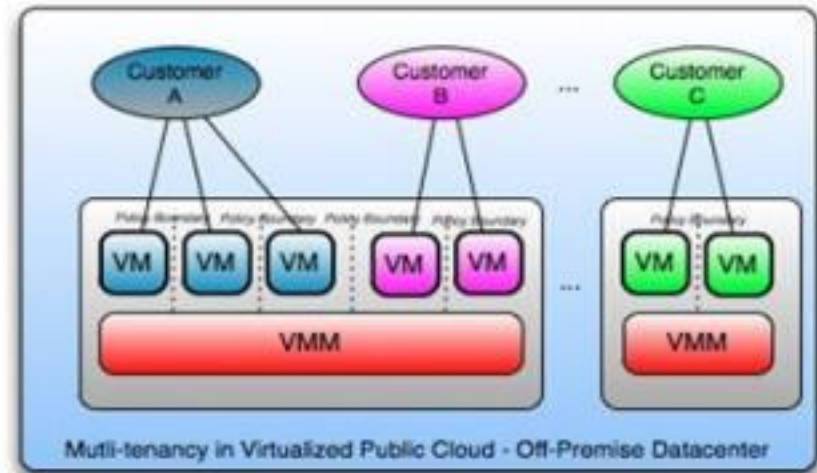
Resource pooling is the keystone for **multi-tenancy**: the ability to enable instances of the program to serve different consumers (tenants) whereby each is isolated from the other.

Resources are dynamically assigned and reassigned according to demand, through statistical multiplexing.

Whilst multi-tenancy is not mandatory, it is applied in most scenarios, in both private and public clouds.



Private Cloud of Company XYZ with 3 business units, each with different security, SLA, governance and chargeback policies on shared infrastructure



Public Cloud Provider with 3 business customers, each with different security, SLA, governance and billing policies on shared infrastructure



#4 Rapid Elasticity

“Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.” [NIST]

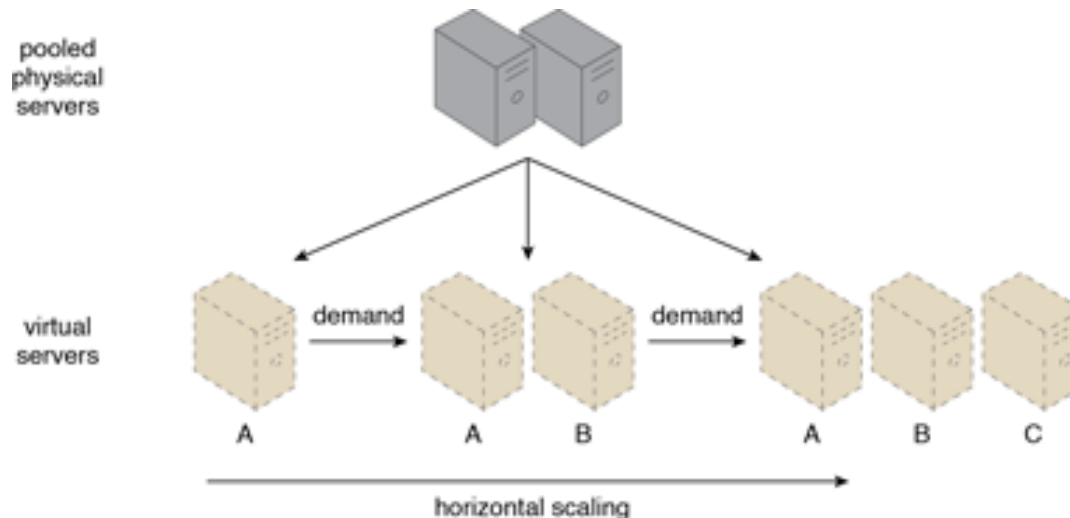




[Break Time: scale-up, scale-out] 1/3

Horizontal scaling (scale out / scale in):

- Add more nodes to the system
(e.g., add one additional Web or DNS server to handle peak loads)
- Requires the application to be specifically designed to support multiple servers – can be complex in some cases
- Complexity of distributed application may impose growth limits
(e.g., synchronization issues between different nodes)

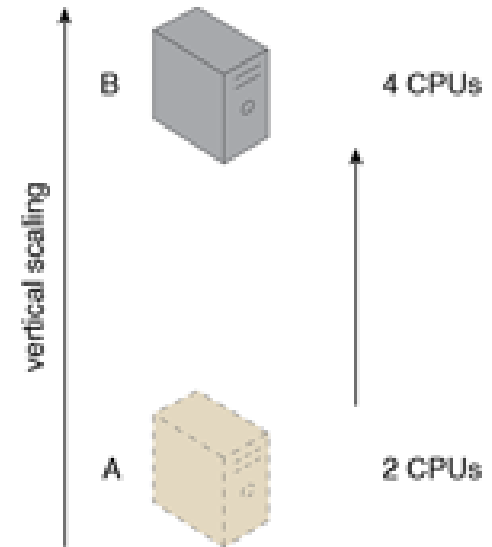




[Break Time: scale-up, scale-out] 2/3

Vertical scaling (scale up / scale down):

- Add more resources to the existing node(s)
(e.g., add more memory and CPU to the existing Web server)
- Requires the virtualization framework and the operating system to handle dynamic resizing of computing resources – sometimes requires rebooting the servers or the applications
- Operating systems may impose scalability boundaries (e.g., maximum supported number of CPUs or maximum amount of RAM)





[Break Time: scale-up, scale-out] 3/3

Keep in mind the need to also scale-down/scale-in:

- Free the no longer necessary allocated resources
- Can be triggered automatically by the application or the cloud management framework

Automatic scale-up/down and scale-out/scale-in may be tricky:

- Conflicting interests (CSP wants to charge more – may force unnecessary scale up/out operations)
- Need to combine monitoring data from the application (often owned by the customer) and from the cloud platform (owned by the CSP)
- May require explicit support from the CSP (for instance to achieve geographic distribution of application servers)



#5 Measured Service

“Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.” [NIST]

Measuring is necessary for monitoring, efficient management of existing resources, SLA management and, when applicable, charging





What defines Cloud Computing as... ...Cloud Computing?

“Dear Editor:

I have 2 servers in my room, running 24/7 and hosting 10 Virtual Machines (VMs) with several applications! Do I own and operate a private cloud infrastructure?”

- Can you create new services/virtual resources on demand?
- Are the resources and services widely available on the Internet?
- Can you provide resources such as storage and VMs in such a way that they are independent of your hardware?
- Can you quickly provide/take back more resources for demand peaks?
- Can you measure the resources consumed by each service/VM?





[Status Update]

What is Cloud Computing?

Service Models (SaaS, PaaS, IaaS)

Deployment Models (private, public, community, hybrid clouds)

Key characteristics of Cloud Computing (on-demand self-service, broad network access, resource pooling, rapid elasticity, measured)

And also...

- **scaling concepts** (scale up/out/down/in)
- **multi-tenancy vs. single-tenancy**

[Next]

Roles in Cloud Computing

Some additional fundamental concepts...



Roles in Cloud Computing (1/2)

Cloud Service Provider

- Organization that provides cloud-based services (SaaS, PaaS, IaaS). May own their own cloud infrastructure or resell from 3rd parties.

Cloud Consumer

- User or organization that consumes cloud-based services (not always a clear role – e.g., rent a VM to provide cloud services).

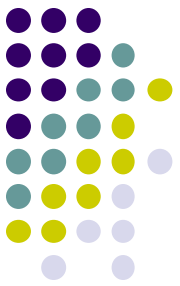
Cloud Service Owner

- Owns the cloud service. It can be the CSP (e.g., Gmail), the cloud consumer (e.g., a company that used IaaS to install new applications for its own staff or the public in general), or a third party

Cloud Administrator

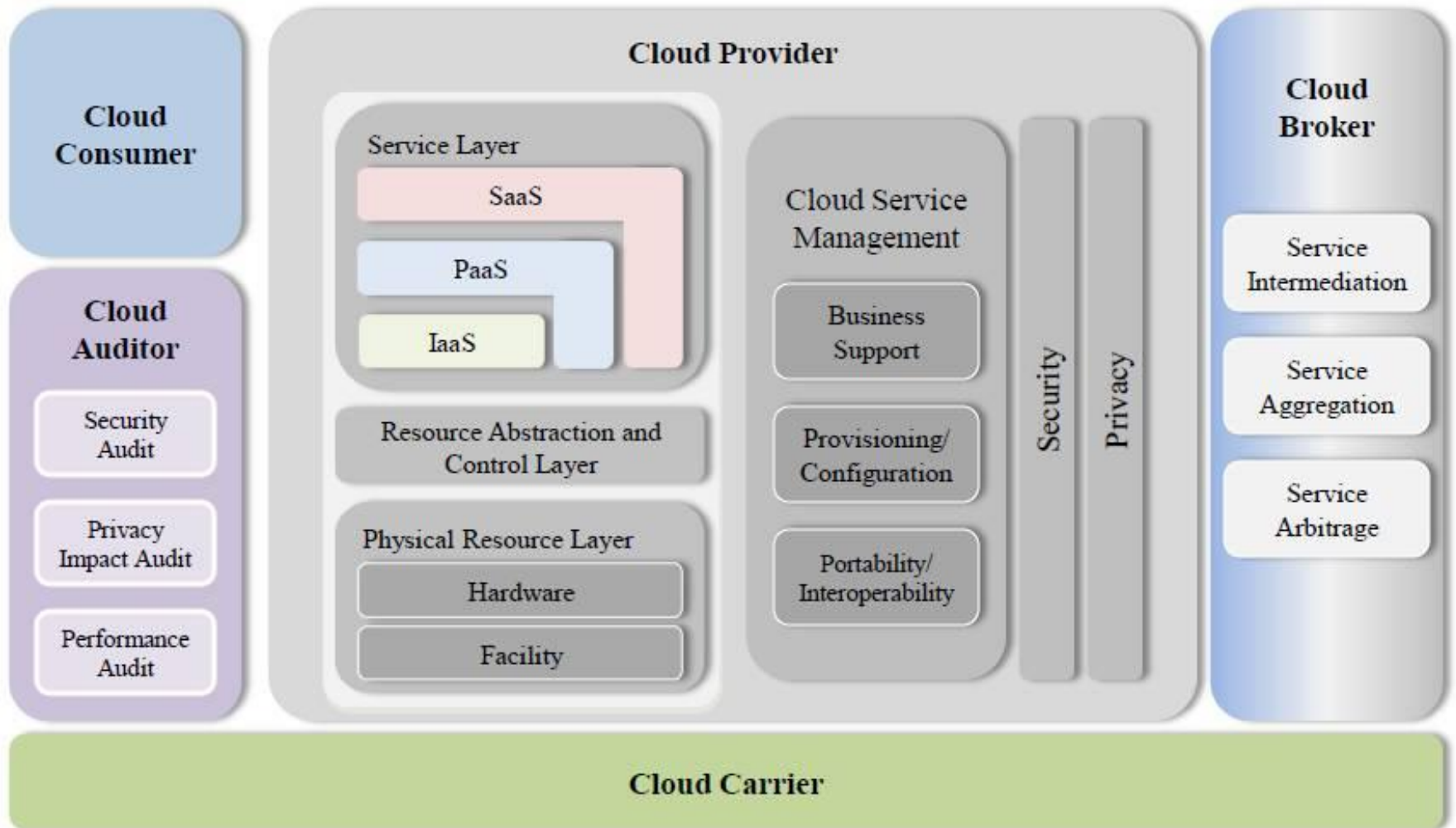
- Person or organization that administers cloud-based services and resources. It can be/belong to the CSP, the consumer, the service owner or a 4th party

Cloud auditor, Cloud Broker, Cloud Carrier...



Roles in Cloud Computing (2/2)

NIST Reference Architecture



Cloud Computing fundamentals (some of this stuff not addressed here!)

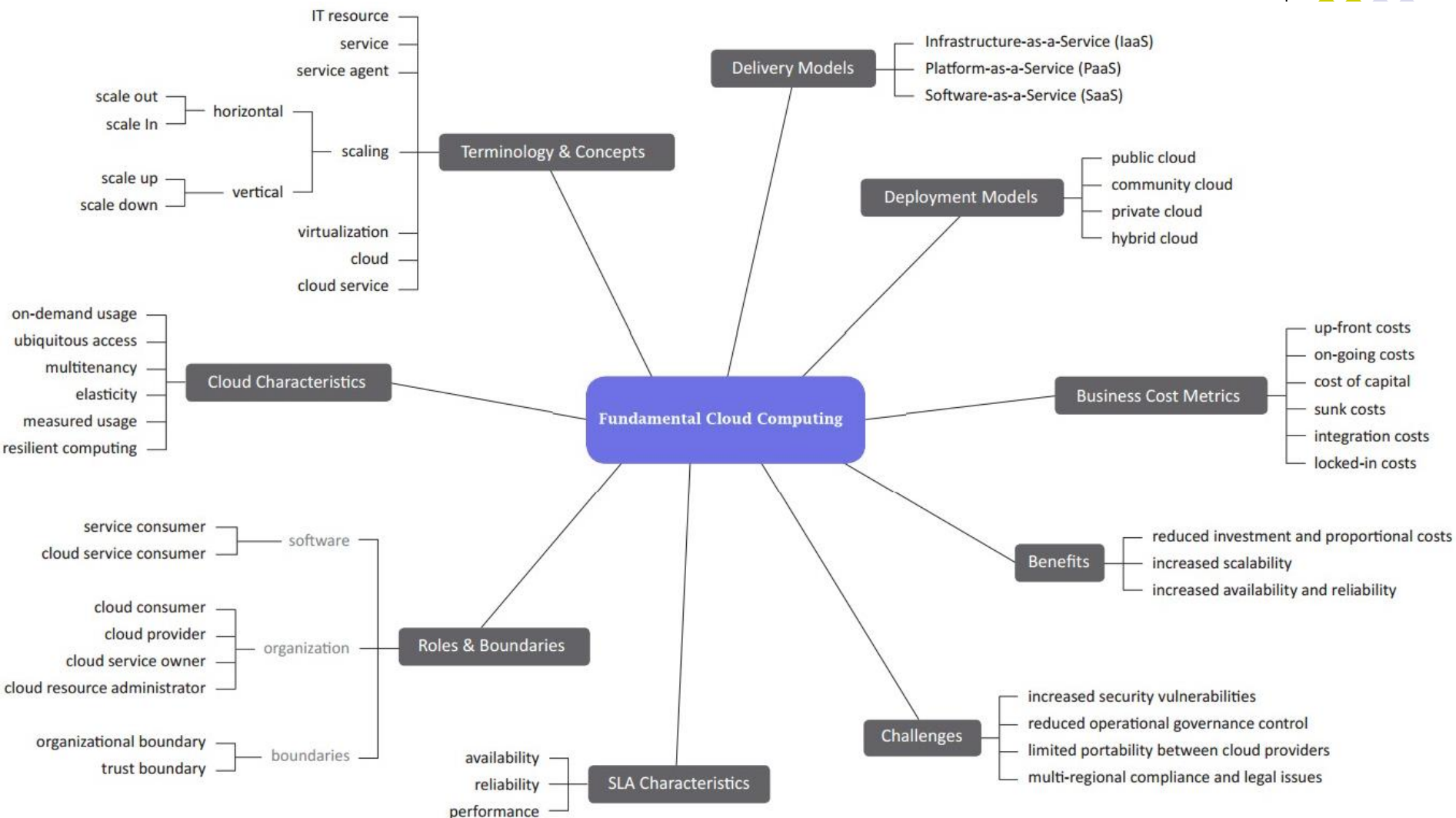


Image credit: www.whatiscloud.com



How does it work altogether? (1/2)

**Depends on the actual scenarios and perspective
(consumer, provider, administrator) – mileage varies a lot!!!**

Some examples:

- **CloudSigma – IaaS provider**
www.cloudsigma.com →
use a trial account to create your own servers

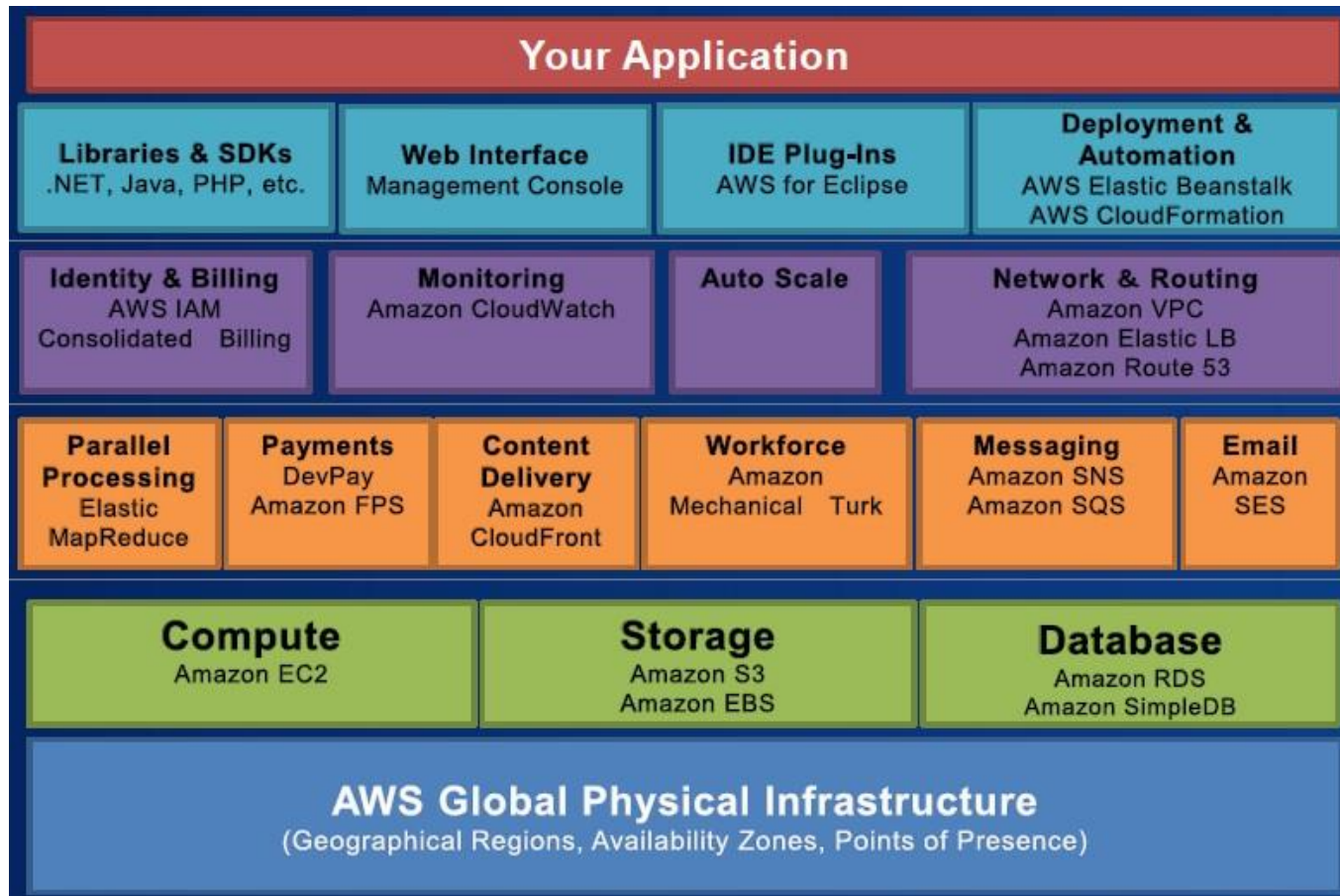
How to manage everything from the administrator's perspective?

- Computing, networking, and storage hardware (next classes...)
- Virtualization of servers and storage (next classes...)
- Software-defined networking to link everything together
see <https://www.youtube.com/watch?v=nunM-GGVJYU>
- Overall orchestration and resource management



How does it work altogether? (2/2)

Amazon Web Services (AWS)



*Don't worry with all the acronyms.
Focus on the big picture
(we will not use AWS anyway!)*



Readings & Links

NIST Special Publication 800-145 (Definition of Cloud Computing):

<https://csrc.nist.gov/publications/detail/sp/800-145/final>

VMware Cloud Fundamentals Course

<https://mylearn.vmware.com/>

(free, requires registration, interesting entry point for more advanced courses)

Cloud Sigma tutorial materials (how to create a server)

www.cloudsigma.com, <https://www.youtube.com/watch?v=yplkf0N71vk>