Cloud Computing

The National Institute of Standards and Technology (NIST) Definition of Cloud Computing

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.

Essential Characteristics:

- **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.
- **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).

237

Cloud Computing

- **Resource pooling**. 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.

- **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.

- **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.

238

Cloud Computing

Service Models:

- **Software as a Service (SaaS).** The consumer uses the provider's applications running on a cloud infrastructure.
  - These applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface.
  - The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, apart from limited user specific application configuration settings.

- **Platform as a Service (PaaS).** The consumer can deploy consumer-created or acquired applications onto the cloud.
  - Again, 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 configuration settings for the application-hosting environment

239

Cloud Computing

- **Infrastructure as a Service (IaaS)** The consumer can deploy and run arbitrary software, including operating systems and applications.
  - The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

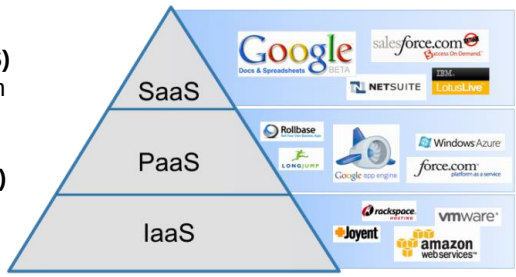
240

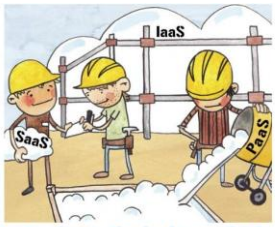
Cloud Deployment Models

Software as a Service (SaaS) Provides access to application services – usually via a browser.

Platform as a Service (PaaS) provides computing platforms which typically include operating systems, programming language execution environments, databases, web servers.

Infrastructure as a service (IaaS), provides the infrastructure such as virtual machines and other resources such as virtual-machine disk image library, block and file-based storage, firewalls, load balancers, IP addresses, and virtual local area networks.

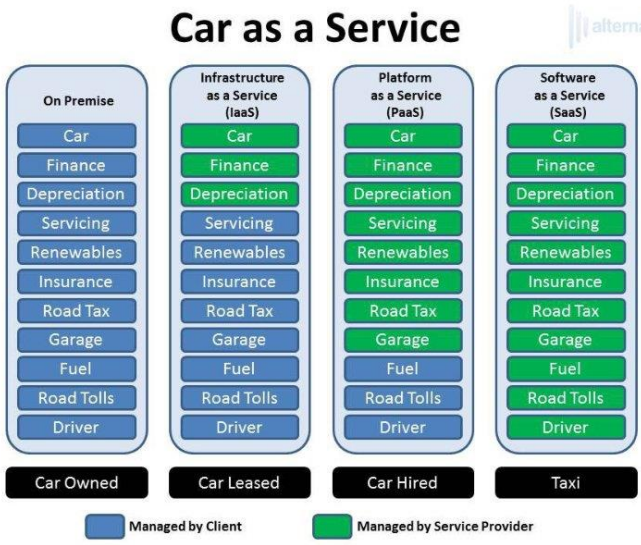




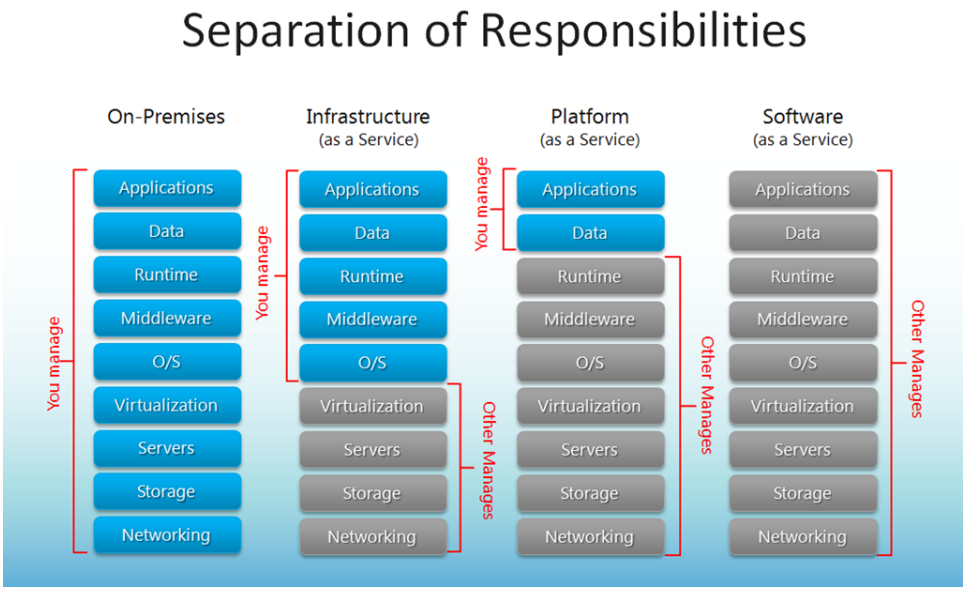
SaaS, PaaS and IaaS - The cloud Computing Services

241

242


Car as a Service
alternative
On Premise
Infrastructure as a Service (IaaS)
Platform as a Service (PaaS)
Software as a Service (SaaS)
Car
Finance
Depreciation
Servicing
Renewables
Insurance
Road Tax
Garage
Fuel
Road Tolls
Driver
Car Owned
Car Leased
Car Hired
Taxi
Managed by Client
Managed by Service Provider

243


Separation of Responsibilities
On-Premises
Infrastructure (as a Service)
Platform (as a Service)
Software (as a Service)
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
You manage
Other Manages

244

Cloud Computing

Types of Cloud:

- **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.

- **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 organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

- **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.

- **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).

245

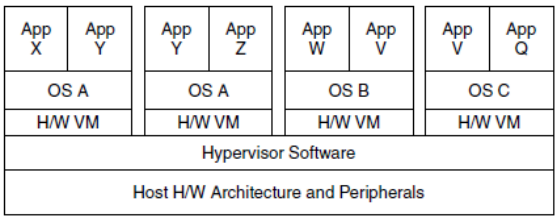
Hardware Virtualization

Hardware virtualization is a combination of hardware and software support that enables the simultaneous execution of multiple operating systems on a single physical computer.

To the user, each **virtual machine** running on the host computer appears to be a complete standalone computing system.

The **hypervisor** is a software component, much like an operating system kernel, that creates and manages instances of virtual machines.

The hardware provides necessary information to the hypervisor so that it can implement sharing policies for the CPU, storage, and I/O devices.



246

Hardware Virtualization

Benefits of Virtualization
System administrators can place multiple virtual machines on the same physical server and can move running virtual machines between servers to better distribute the total load.

Virtual machines give administrators fine-grained control over I/O device access. For example, the bandwidth of a virtualized network port could be partitioned based on user-service levels.

Challenges in Virtualization
No instruction should be able to access resources outside of the current virtual machine.

Memory access instructions (e.g., loads and stores) must only access physical memory allocated to the currently executing virtual machine. Typically, a processor supporting hardware virtualization will provide an additional page-mapping facility that maps *virtual machine physical memory pages to host machine physical memory pages*.

247

Hardware Virtualization

I/O instructions (including memory-mapped I/O) must not directly access physical I/O devices, since many virtualization policies partition access to I/O devices.

This fine-grained I/O control is typically implemented with interrupts to the hypervisor any time a virtual machine attempts to access an I/O device. The hypervisor can then implement the I/O resource access policy of its own choosing. Typically, some set of I/O devices is supported and the operating systems running in the virtual machines, called **guest operating systems** are expected to use these supported devices.

248