

FOCUS: server consolidation

Virtualization allows to have:

- Hardware independence of the software,
- High flexibility,
- The operative system and the applications can be handed as a "single entity"

Consolidation means migrating from physical to virtual machines, servers are connected one another so it is possible to:

- Perform live migration, which is moving one server to another physical architecture without switching it off, thus making the system more scalable;
- Automatically balance the workloads, without the machines feeling the difference;
- Protect the machines from hardware or system failures, as if a physical machine fails the other continue running, so the virtual machines do not need to be stopped even if they could be slowed down.

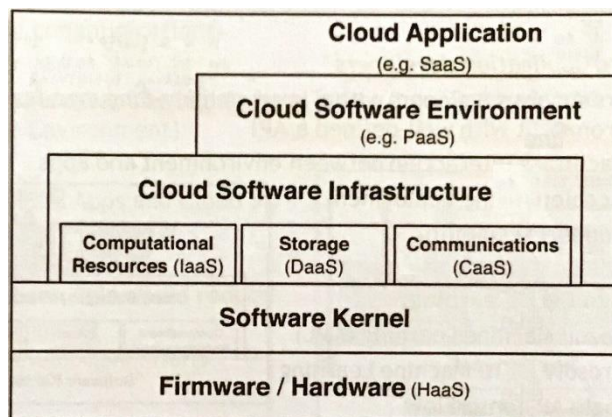
It leads to a higher hardware utilization, it allows different operative systems to run on the same hardware and to continue to use legacy software (Linux applications on Linux, Windows on Windows, ...) as the applications are independent from the hardware.

Cloud computing is a model in which resources as CPU and storage are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion.

Cloud computing (DEF)

It is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources, as network, servers, storage, applications and services that can be rapidly provisioned and released with minimal management effort or service provider interaction.

1. Cloud application (software as a service): the user accesses resources which are provided through the internet without needing to install or know anything.
2. Cloud software environment (platform as a service): the users are application developers, they are supplied with a platform where they can code and with the needed API, to accelerate the deployment and support scalability of the so-built applications.
3. Cloud software infrastructure: it provides resources to the higher level, it might be bypassed by the lower ones.
 - a. Computational resources (infrastructure as a service): virtual machines allow flexibility and root access to the machines, but they could lead to performance inefficiencies and to the inability to provide strong guarantees about the service levels.
 - b. Storage (data as a service): they allow users to store and retrieve data (usually through specific applications), the key points are a high dependability, replication and data consistency.
 - c. Communications (communication as a service): it becomes a vital component, it must provide security (encryption) and monitoring and be service-oriented, configurable, schedulable, reliable and predictable, so it is necessary to make previsions on the traffic with the aim of guaranteeing a maximum delay of message.
4. Software kernel: it provides the basic software management for the servers.



5. Firmware/hardware (hardware as a service): the users are big IT enterprises with high software requirements, they do not want to worry about the hardware, so they rent it, the provider must guarantee also the needed management and upgrade.

Usually, the role of service provider is divided into two parts: the service provider, who serves the end users, rents its infrastructure by another provider who manages the cloud platform and is called infrastructure provider; the chain can become even longer, as the latter can become the user for a lower-level product.

Cloud infrastructures can belong to four different categories.

1. Public

A company rents its infrastructure to private users or to other companies, it is characterised by the virtualization of the operative systems and by a service level agreement. The main feature is the fact that it is entirely web-based, the "pay-as-you-go" policy allows users to select the service they need, to pay for it and to use it almost immediately.

2. Private

It is an internally managed data-centre, a huge company buys the hardware and takes care of everything, the total control has the side effect of loose flexibility.

3. Community

Some organizations share a single cloud, so resources can be used by one of the partners when the others are not using them; it is similar to private cloud, but it has a more complex accounting system. Usually it is hosted locally, the participants share their infrastructures, but sometimes it is submitted to a specific separate organization or to a small subset of the partners.

4. Hybrid

It is a mixture of the others, it can be used by an IT company that has its own resources but is subject to unpredictable and rare peaks in computation demand so, when they happen, it rents the resources to deal with them only for the needed time

Cloud computing has many positive aspects:

- Lower IT costs,
- Improved performances,
- Instant software updates,
- Unlimited storage capacity,
- Data reliability,
- Documents can be accessed everywhere and on every device.

However, it can also have some disadvantages:

- It requires a constant and fast internet connection, as it does not work well with low-speed connections,
- The features might be limited,
- Can be slow due to virtualization,
- Stored data might not be secure and might be lost.

A further development: fog computing

It stays in the middle between the objects and the cloud: the computation is split. The fog pre-processes the data and, if it is able, it takes a decision, otherwise it sends them to the cloud to exploit its major computation power.