



Cloud Foundations



Chapter 2: Elastic Compute

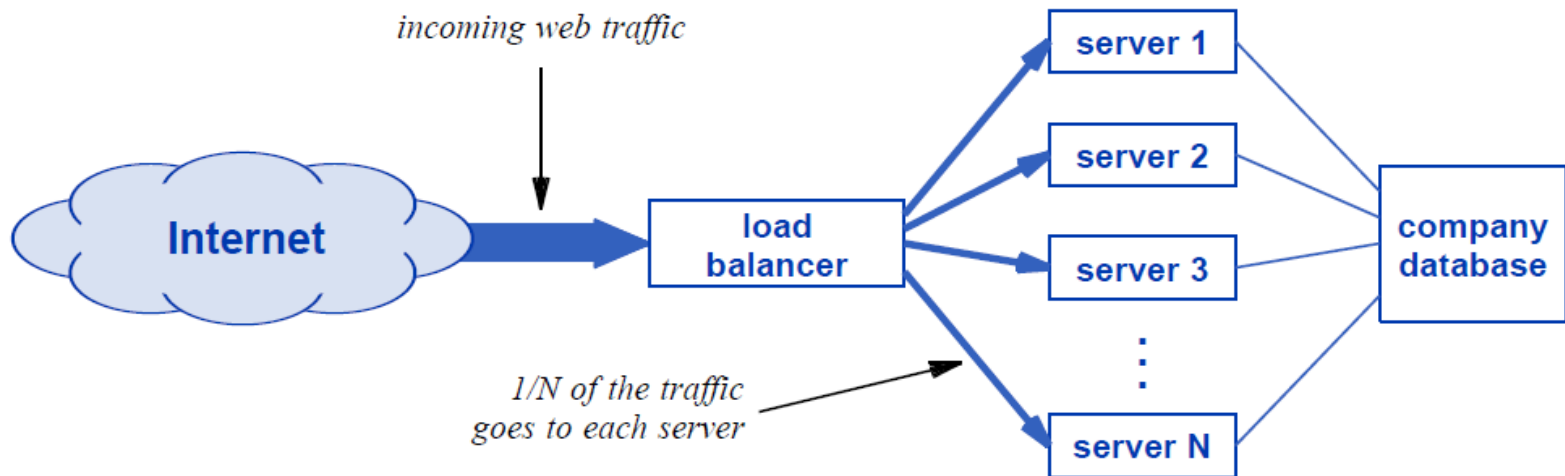
PR. OUASSIM KARRAKCHOU

UNIVERSITE INTERNATIONALE DE RABAT



The Origin of the Term “Cloud”

What do we refer to as “Cloud”?

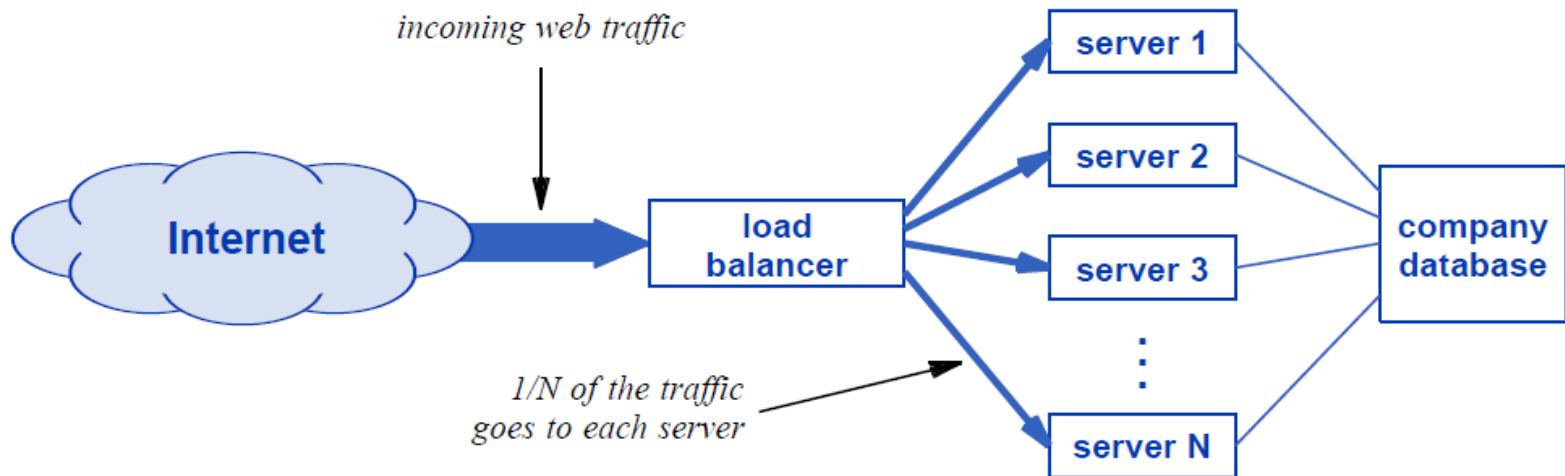


If you look at this figure, where is the Cloud?



The Origin of the Term “Cloud”

What do we refer to as “Cloud”?



Normally a Cloud represents the Internet in figures.

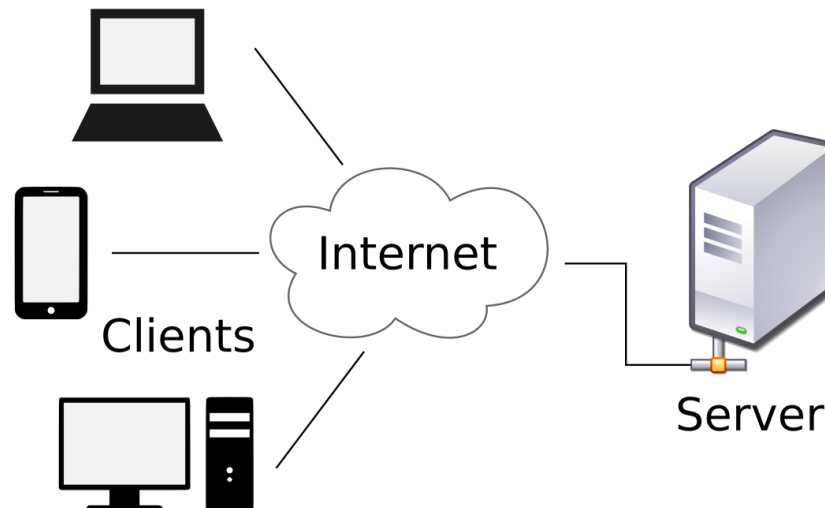
Why do we say that a data center is in the Cloud, or that an app has been deployed in the Cloud?



The Origin of the Term “Cloud”

What do we refer to as “Cloud”?

- Saying a data center is “in the cloud” is technically inaccurate because the servers in a data center are not part of the Internet itself. Instead, servers are merely **computers that attach to the Internet**, and should be depicted outside the cloud, with network connections leading to the cloud.



Why did the industry start using the terms *cloud computing* and say that the computing is in the cloud if it is not?



The Origin of the Term “Cloud”

What do we refer to as “Cloud”?

- Early **data centers** that supplied large-scale web services needed **high-capacity connections** to major Internet backbone networks. Placing a data center near an **Internet peering point** where major Internet backbones interconnect minimizes cost.
- Some providers actually placed their data centers **on the same floor in the same building** as an Internet peering point. Because such a data center was located with networking equipment rather than in a separate building, engineers said the **data center was in the cloud**, and the terminology caught on.





The Origin of the Term “Cloud”

What do we refer to as “Cloud”?

Although technically inaccurate, the phrase in the cloud arose because early data centers were located close to networking equipment at Internet peering points rather than in separate buildings.



How did we go from servers and data centers to the Cloud?



Elastic Compute and Virtualization

Multi-tenant Clouds

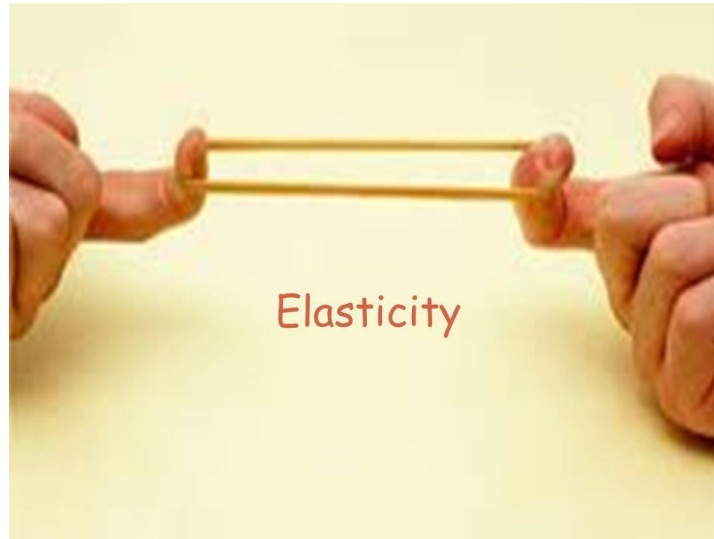
- Consolidating servers into a single data center has **economic advantages** for an organization.
- Cloud providers extend the advantages to a **larger scale**.
- In particular, instead of handling computing for one organization, a cloud provider builds a **data center** (or multiple data centers) that can handle computing for **many customers**.
- We use the term **multi-tenant** to refer to a data center that serves customers from multiple organizations.

The challenge is how to do this correctly. What challenges can you think of?



Elastic Compute and Virtualization

The challenge of Elasticity



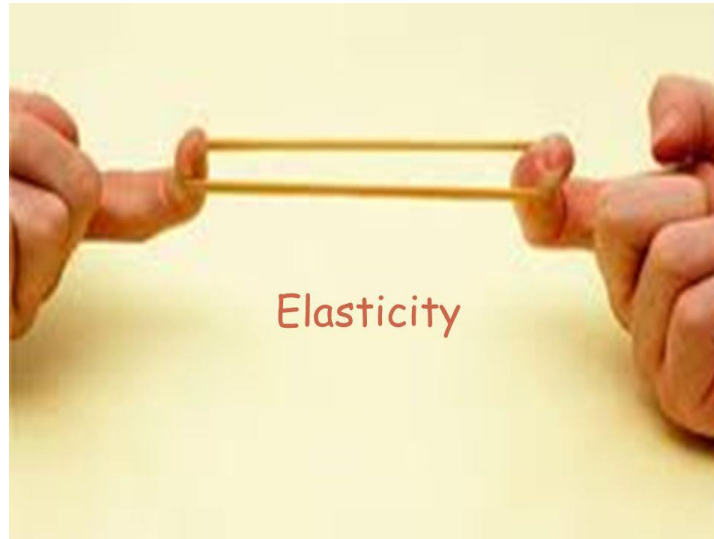
- One of the main motivations for Cloud was to reserve just enough computation resources to adapt to the demand, nothing more nothing less.
- We call this property Elasticity

How can we make the infrastructure Elastic?



Elastic Compute and Virtualization

A not so good solution to Elasticity

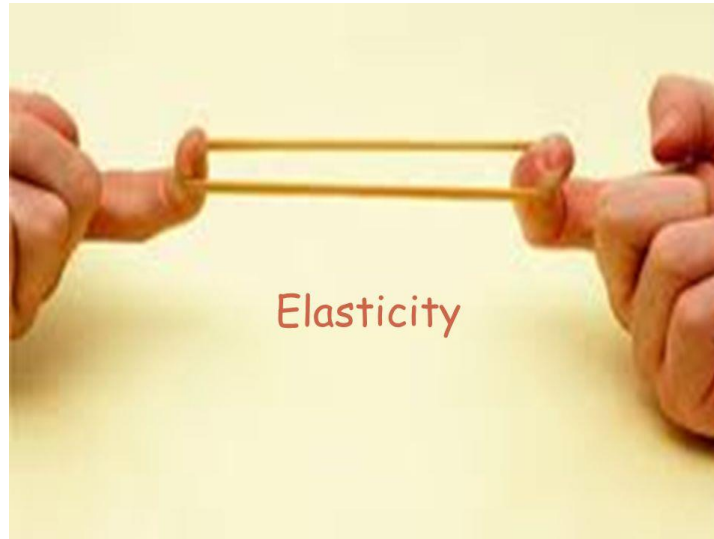


- A simple solution is to give the ability to a customer to **lease servers** and only pay for the number of servers they need. A customer can choose to lease a few servers or many. More important, a customer can **change the allocation dynamically**, adding servers during peak times and decreasing the number of servers during times they are not needed.



Elastic Compute and Virtualization

A not so good solution to Elasticity

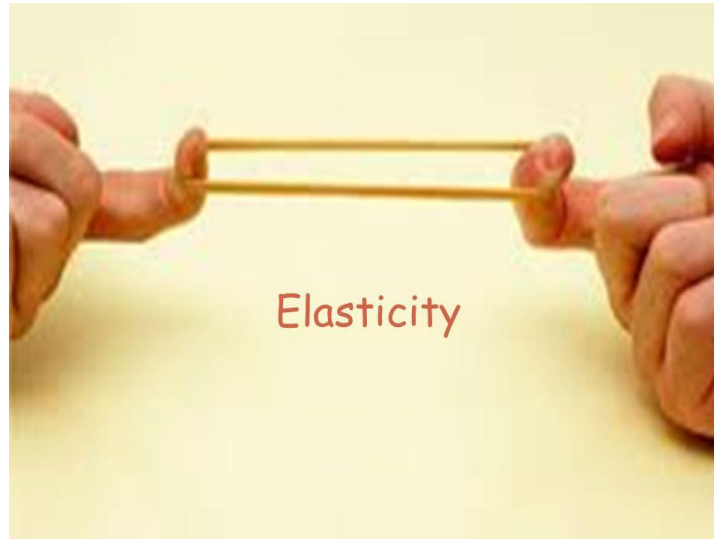


- This early approach **allocated a set of physical servers**. A provider restricted a customer to a few sizes. For example, a customer could lease a full rack of servers, a half rack, or a quarter rack, or multiples of the sizes.
- However, such an approach is relatively **inflexible** because it requires a provider to dedicate a set of physical resources to each customer.



Elastic Compute and Virtualization

Software comes to the rescue for Elasticity

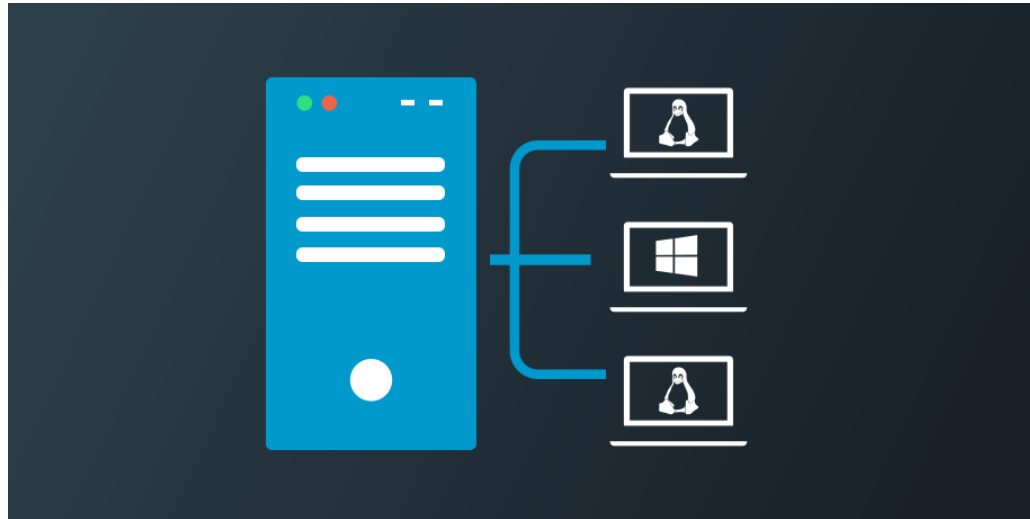


- A technology emerged that allows a cloud provider to offer fast, flexible allocation at low cost.
- Surprisingly, the technology does not involve allocating sets of physical servers. Instead, a cloud owner runs **software** on each physical server that allows the cloud owner to create a set of **virtualized servers**.



Elastic Compute and Virtualization

Main properties of virtualization

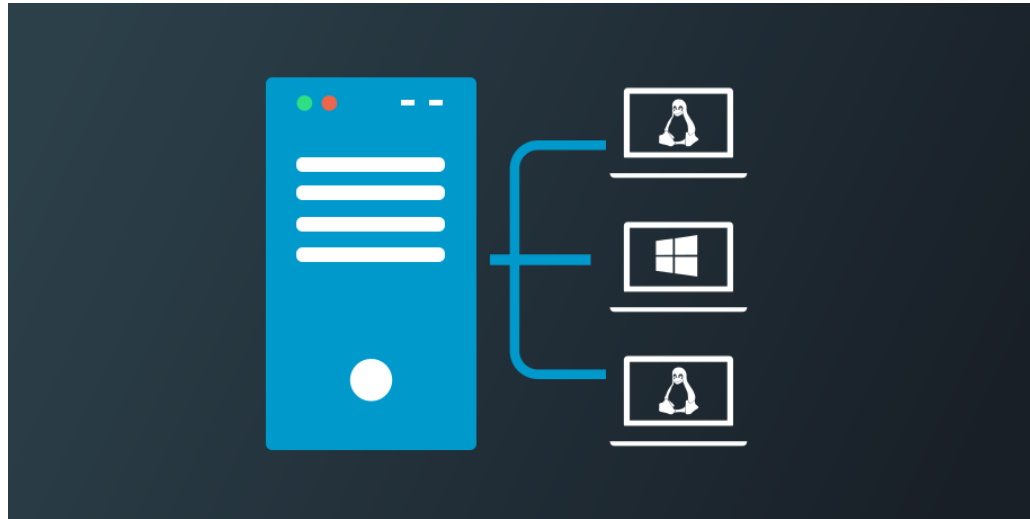


What are the main properties of virtualization compared to physical servers?



Elastic Compute and Virtualization

Main properties of virtualization



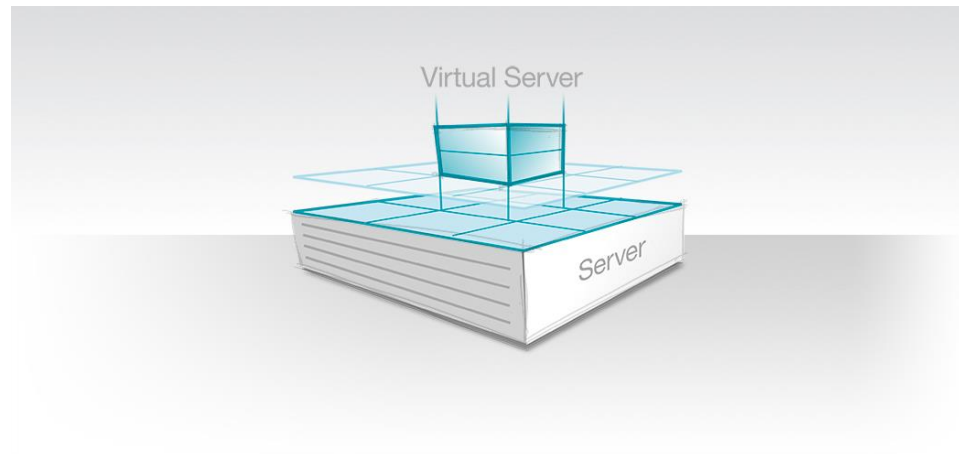
Three important properties of virtualized servers:

- Rapid Creation and Removal
- Physical Sharing
- Logical Isolation



Elastic Compute and Virtualization

Virtualization: Rapid creation and Removal

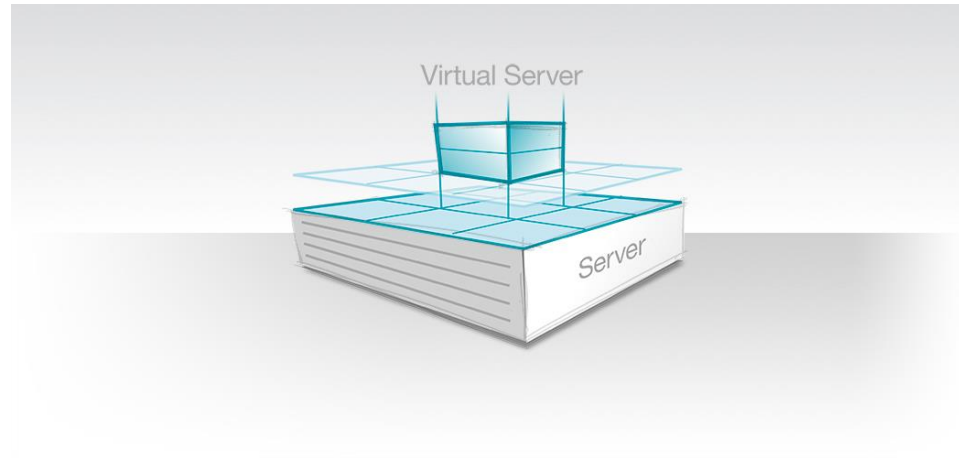


- Virtualized servers are managed entirely by **software**.
- Management software can **create or remove** a new virtualized server at any time, without changing or rebooting physical servers.



Elastic Compute and Virtualization

Virtualization: Physical Sharing

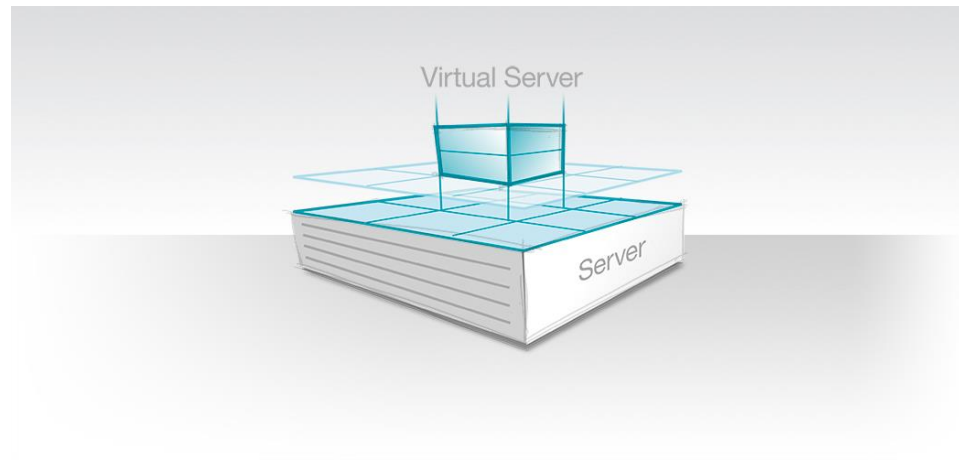


- Because it consists of a software artifact, a virtualized server is similar to a computer program.
- Each virtualized server must run on a physical server, and multiple virtualized servers can run on a given physical server concurrently.



Elastic Compute and Virtualization

Virtualization: Logical Isolation



- Although multiple virtualized servers can run on a single physical server at the same time, **each virtualized server is completely isolated** from the others.
- The data and computations performed by one virtualized server **cannot be observed or affected by another**.



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Providers



- From a cloud provider's point of view, the ability to virtualize servers provides the **basis for elastic computing** and makes cloud computing **economically viable**.
- A cloud provider only needs to use computer **software** to **increase or decrease the number of servers** a customer is leasing.
- Because the provider does not need to reconfigure or reboot physical servers, a provider can **accommodate the creation of thousands of virtualized servers** as needed.



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Providers



- Because virtualization technologies guarantee **isolation**, a cloud provider can place virtualized servers on physical servers **without regard to** the owner, the apps they will run, or the data they will handle.
- Isolation allows a provider to mix **virtualized servers from multiple customers on the same physical server** without any interference and without any chance of data from one customer's virtualized server "leaking" to another customer.



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Providers



- A provider needs to **choose where to run a virtualized server**, the choice can be made in a way that provides the best benefit to the provider.
- In particular, a provider tries to **avoid placing too many virtualized servers on the same physical server** because each virtualized server will receive less processing power, causing customer complaints.
- A provider can use virtualized server placement to **balance the load** across all physical servers in the data center.



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Providers



Using virtualized servers provides invaluable advantages for cloud providers, including the ability to scale the service and the ability to balance the load and avoid overloading a physical server while other physical servers remain idle.



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Customers



- To a customer, a virtualized server **appears to act like a physical server**.
- A virtualized server allows apps to **communicate over the Internet**. That is, like a physical server, each virtualized server is **assigned an Internet address**.

But if a virtualized server merely acts like a physical server, how does a customer benefit?



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Customers



- Ease of creating and deploying new services:
 - Cloud providers and third-party vendors offer software that makes it easy to create new apps for a cloud environment.
 - In addition, they offer software that can deploy an app in the cloud, including software that replicates an app on multiple virtualized servers.



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Customers



- Rapid scaling of a service:
 - **Scaling** an app to handle more users means **adding more copies**.
 - If an app runs on virtualized servers, **new copies can be created quickly** (e.g., as requests arrive over the Internet).



Elastic Compute and Virtualization

Advantages of Virtualization for Cloud Customers



- Safe and rapid testing of new software or new versions:
 - Before deploying new apps or new versions of software, most large organizations deploy the software on an **isolated test system** before installing it in production.
 - Virtualized servers allow an organization to **create isolated virtualized servers** for a test system **without interfering with the production systems**.



Elastic Compute and Virtualization

Summary

- *Elasticity forms the key concept behind cloud computing: a customer can lease as many servers as needed, the number can vary over time, and a customer only pays for the number of servers used.*
- *Instead of dedicating physical servers to each customer, cloud providers use software technology to create virtualized servers as needed.*
- *Multiple virtualized servers can run on a given physical server.*
- *Virtualization has many benefits for both the Cloud Providers and Customers.*