Scalability and Elasticity

Cloud Elasticity

Elasticity refers to the ability of a cloud to automatically expand or compress the infrastructural resources on a sudden up and down in the requirement so that the workload can be managed efficiently. This elasticity helps to minimize infrastructural costs. This is not applicable for all kinds of environments, it is helpful to address only those scenarios where the resource requirements fluctuate up and down suddenly for a specific time interval. It is not quite practical to use where persistent resource infrastructure is required to handle the heavy workload.

The Flexibility in cloud is a well-known highlight related with scale-out arrangements (level scaling), which takes into consideration assets to be powerfully added or eliminated when required. It is for the most part connected with public cloud assets which is generally highlighted in pay-per-use or pay-more only as costs arise administrations.

The Flexibility is the capacity to develop or contract framework assets (like process, capacity or organization) powerfully on a case by case basis to adjust to responsibility changes in the applications in an autonomic way.

Example: Consider an online shopping site whose transaction workload increases during festive season like Christmas. So for this specific period of time, the resources need a spike up. In order to handle this kind of situation, we can go for a Cloud-Elasticity service rather than Cloud Scalability. As soon as the season goes out, the deployed resources can then be requested for withdrawal.

Cloud Scalability

Cloud scalability is used to handle the growing workload where good performance is also needed to work efficiently with software or applications. Scalability is commonly used where the persistent deployment of resources is required to handle the workload statically.

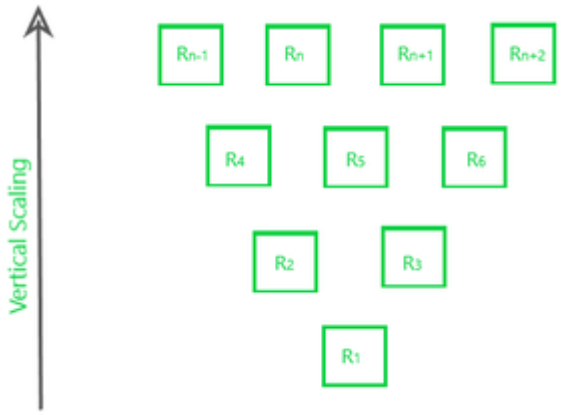
Example: Consider you are the owner of a company whose database size was small in earlier days but as time passed your business does grow and the size of your database also increases, so in this case you just need to request your cloud service vendor to scale up your database capacity to handle a heavy workload.

It is totally different from what you have read above in Cloud Elasticity. Scalability is used to fulfill the static needs while elasticity is used to fulfill the dynamic need of the organization. Scalability is a similar kind of service provided by the cloud where the customers have to pay-per-use. So, in conclusion, we can say that Scalability is useful where the workload remains high and increases statically.

Types of Scalability

1. Vertical Scalability (Scale-up)

In this type of scalability, increase the power of existing resources in the working environment in an upward direction.

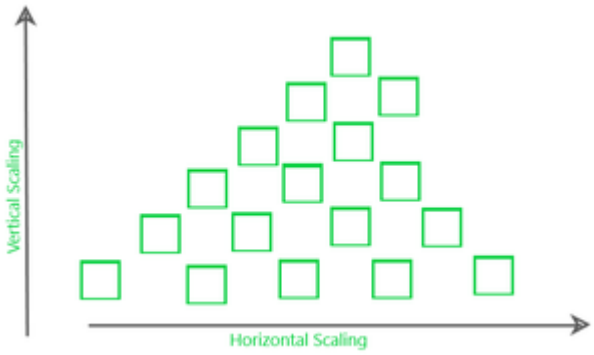


2. **Horizontal Scalability**

In this kind of scaling, the resources are added in a horizontal row.

3. **Diagonal Scalability**

It is a mixture of both Horizontal and Vertical scalability where the resources are added both vertically and horizontally.



Difference Between Cloud Elasticity and Scalability

	Cloud Elasticity	**Cloud Scalability**
1	Elasticity is used just to meet the sudden up and down in the workload for a small period of time.	Scalability is used to meet the static increase in the workload.
2	Elasticity is used to meet dynamic changes, where the resources need can increase or decrease.	Scalability is always used to address the increase in workload in an organization.
3	Elasticity is commonly used by small companies whose workload and demand increases only for a specific period of time.	Scalability is used by giant companies whose customer circle persistently grows in order to do the operations efficiently.
4	It is a short term planning and adopted just to deal with an unexpected increase in demand or seasonal demands.	Scalability is a long term planning and adopted just to deal with an expected increase in demand.

Replication

The simplest form of data replication in cloud computing environment is to store a copy of a file (copy), in expanded form, the copying and pasting in any modern operating systems. Replication is the reproduction of the original data in unchanged form. Changing data accesses are expensive in general through replication. In the frequently encountered master / slave replication, a distinction between the original data (primary data) and the dependent copies. In

peer copies (version control) there must be merging of data sets (synchronization). Sometimes it is important to know which data sets must have the replicas. Depending on the type of replication it is located between the processing and creation of the primary data and their replication in a certain period of time. This period is usually referred to as latency.

Array-Based Data Replication

An array-based data replication strategy uses built-in software to automatically replicate data. With this type of data replication, the software is used in compatible storage arrays to copy data between each. Using this method has several advantages and disadvantages.

Advantages:

- More robust
- Requires less coordination when deployed
- The work gets offloaded from the servers to the storage device

Disadvantages:

- Requires homogenous storage environments: the source and target array have to be similar
- It is costly to implement

Host-Based Data Replication

Host-based data replication uses the servers to copy data from one site to another site. Host-based replication software usually includes options like compression, encryption and, throttling, as well as failover. Using this method has several advantages and disadvantages.

Advantages:

- Flexible: It can leverage existing IP networks
- Can be customized to your business' needs: You can choose what data to replicate
- Can create a schedule for sending data: allows you to throttle bandwidth
- Can use any combination of storage devices on each end

Disadvantages:

- Difficult to manage with a large group of servers if there is no centralized management console
- Consumes host resources during replication
- Both storage devices on each end need to be active, which means you will need to purchase dedicated hardware and OS
- Not all applications can support this type of data replication
- Can be affected by viruses or application failure
- Host-based replication offers the safest option if a business is looking for close to zero impact on operations after a disaster.

Network-Based Data Replication

Network-based data replication uses a device or appliance that sits on the network in the path of the data to manage replication. The data is then copied to a second device. These devices usually have proprietary replication technology but can be used with any host server and storage hardware.

Advantages

- Effective in large, heterogeneous storage and server environments
- Supports any host platform and works with any array
- Works separately from the servers and the storage devices
- Allows replication between multi-vendor products

Disadvantages:

- Higher initial set-up cost because it requires proprietary hardware, as well as ongoing operational and management costs
- Requires implementation of a storage area network (SAN)

Monitoring

Cloud monitoring is a method of reviewing, observing, and managing the operational workflow in a cloud-based IT infrastructure. Manual or automated management techniques confirm the availability and performance of websites, servers, applications, and other cloud

infrastructure. This continuous evaluation of resource levels, server response times, and speed predicts possible vulnerability to future issues before they arise.

This technique tracks multiple analytics simultaneously, monitoring storage resources and processes that are provisioned to virtual machines, services, databases, and applications. This technique is often used to host infrastructure-as-a-service (IaaS) and software-as-a-service (SaaS) solutions. For these applications, you can configure monitoring to track performance metrics, processes, users, databases, and available storage. It provides data to help you focus on useful features or to fix bugs that disrupt functionality.

Monitoring is a skill, not a full-time job. In today's world of cloud-based architectures that are implemented through DevOps projects, developers, site reliability engineers (SREs), and operations staff must collectively define an effective cloud monitoring strategy. Such a strategy should focus on identifying when service-level objectives (SLOs) are not being met, likely negatively affecting the user experience. So, then what are the benefits of leveraging cloud monitoring tools? With cloud monitoring:

Benefits of cloud monitoring

- Scaling for increased activity is seamless and works in organizations of any size
- Dedicated tools (and hardware) are maintained by the host
- Tools are used across several types of devices, including desktop computers, tablets, and phones, so your organization can monitor apps from any location
- Installation is simple because infrastructure and configurations are already in place
- Your system doesn't suffer interruptions when local problems emerge, because resources aren't part of your organization's servers and workstations
- Subscription-based solutions can keep your costs low

Cloud monitoring is primarily part of cloud security and management processes. It is normally implemented through automated monitoring software that provides central access and control over cloud infrastructure.