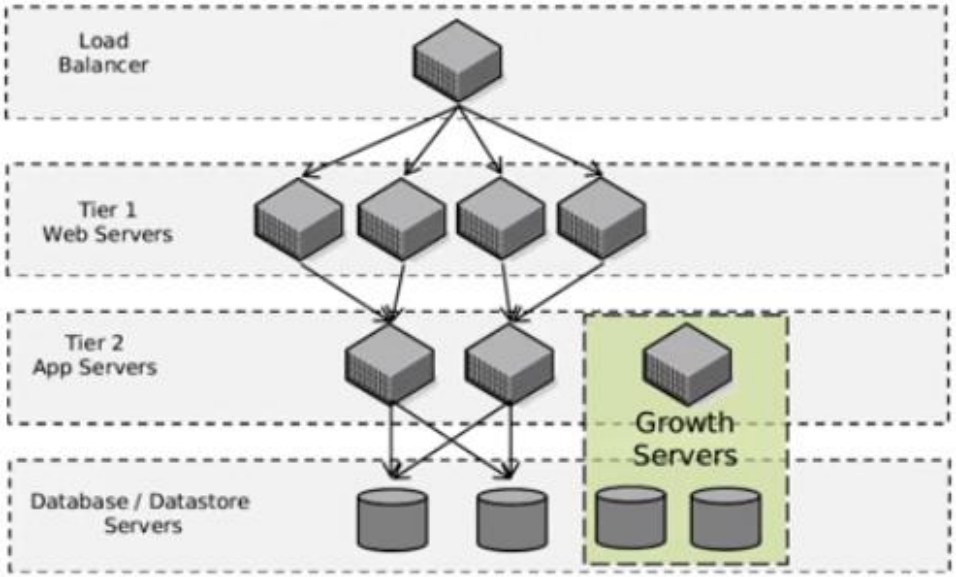
Module 5 - Pricing & Scaling Models

greatlearning
Learning for Life

Cloud Scaling model = Elasticity

Modern applications can leverage IaaS or PaaS for scaling only the layers that demand it!



30

Module 5 - Pricing & Scaling Models

27 ☐ Debate
28 ☐ Subscription model
29 ☐ Classical Scaling model
30 ☐ Cloud Scaling model = Elasticity
31 ☐ Cost economics - Classical model
32 ☐ Cost economics - Cloud model
33 ☑ Scaling in Google & AWS
34 ☐ Vertical/Specialized vs Horizontal/Commodity
35 ☐ Virtualization
36 ☐ A typical application stack
37 ☐ Virtualized stack
38 ☐ Cloud VM characteristics
39 ☐ Virtualization drawbacks

greatlearning
Learning for Life

Scaling in Google & AWS

Total number of instances	Average GPS*	Average Latency*	Average Memory
37 total	0.177	339.5 ms	72.0 MBytes

Instances

GPS*	Latency*	Requests	Errors	Age	Memory	Availability
0.300	347.7 ms	227	0	0:00:29	74.6 MBytes	Dynamic
0.300	329.5 ms	247	0	0:00:24	69.8 MBytes	Dynamic
0.217	343.4 ms	212	0	0:03:27	70.1 MBytes	Dynamic
0.317	304.1 ms	234	0	0:03:25	75.0 MBytes	Dynamic
0.100	370.6 ms	227	0	0:03:24	71.2 MBytes	Dynamic
0.383	318.3 ms	218	0	0:03:22	70.7 MBytes	Dynamic
0.400	341.3 ms	230	0	0:03:13	71.0 MBytes	Dynamic
0.167	346.6 ms	100	0	0:02:26	67.2 MBytes	Dynamic
0.350	343.9 ms	67	0	0:01:35	66.6 MBytes	Dynamic
0.250	386.2 ms	38	0	0:01:15	65.8 MBytes	Dynamic
0.300	388.9 ms	57	0	0:01:33	66.0 MBytes	Dynamic
0.100	376.8 ms	37	0	0:01:16	66.1 MBytes	Dynamic
0.017	269.0 ms	76	0	0:03:20	65.5 MBytes	Dynamic

A script simulated load and Google spun up instances to handle it automatically.

AWS configuration allowing 1 min 4 max instances of certain type with specific rules of scalability

Scaling

Environment type: Load balanced, auto scaling
Number instances: 1 - 4
Scale based on Average network out
Add instance when > 6000000
Remove instance when < 2000000

Instances

Instance type: t1.micro
Availability Zones: Any

33

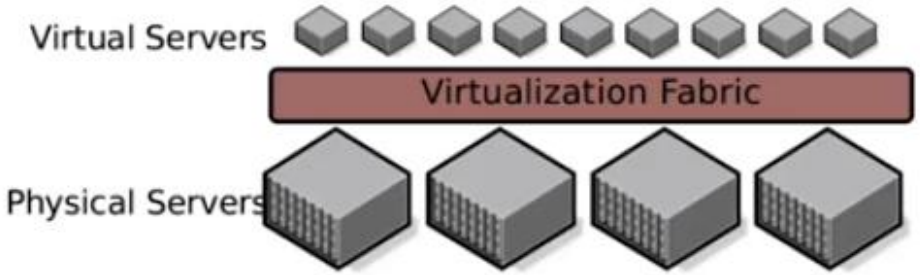
▶

Module 6 - Introduction to Virtualization

greatlearning
Learning for Life

Virtualization

- Virtualization of the computing resources, including servers, network, and storage, allows dynamic flexibility.
- Capacity can be more efficiently utilized.
- Quickly add new servers without delay due to procurement or installation.
- Easy to turn on or off virtual servers to handle scalability.
- Physical connectivity is done up front and configuration is done in software at provisioning time.
- Networking equipment and storage is virtualized as well.



35