Introduction to Cloud Computing

Lakmal Warusawithana

Vise President, Apache Stratos

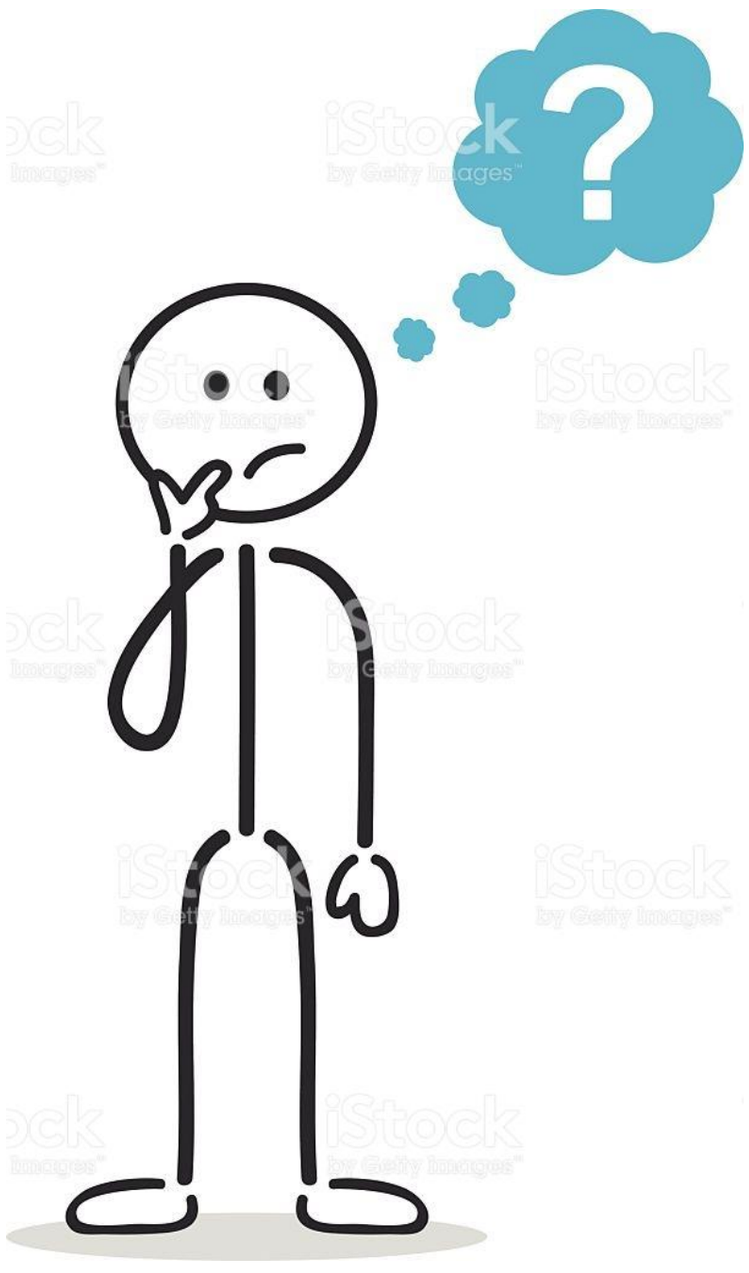
Director - Cloud Architecture, WSO2 Inc

lakmal@apache.org / lakmal@wso2.com

Twitter : lakwarus

2016 - University of Sabaragamuwa

What is Cloud Computing?



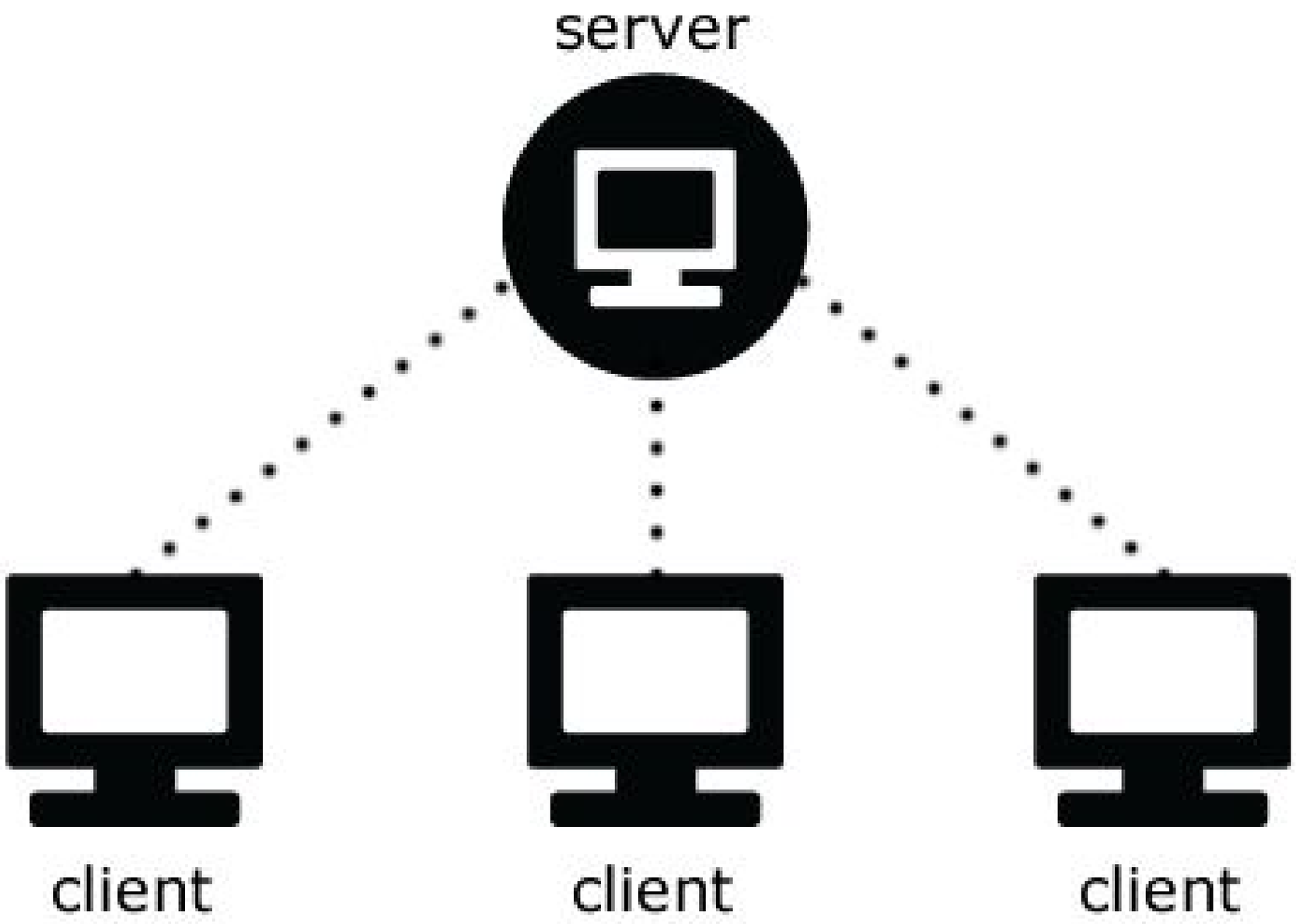
The Evolution of Software – Standalone

The Evolution of Software – Standalone

- Software installed on a local machine
  - e.g. most PC software that need installation
- Local access to software
- Problems - HR System on a PC
  - limited availability
  - limited reliability
  - limited capacity
  - high maintenance cost
    - e.g. software updates
  - high scaling cost

ex -
Imagine a small company using a payroll system installed only on one HR manager’s PC. If that PC crashes, payroll is stuck. That's the limitation of standalone software.

Software Evolution - Client-Server



Software Evolution - Client-Server

- Software installed on a (remote) machine (server)
- Client(s) access(es) server via network
  - communicate using a protocol
- Problems solved:
  - increase availability
    - if you have the proper client, you can use (hopefully)
- Problems remaining:
  - availability (still limited by server capacity)
  - reliability (depends on server)
    - need redundant server(s) & storage as back-up

Software Evolution - Web Applications

Software Evolution - Web Applications

- Still client-server applications
- Use web technologies (e.g. HTTP) to:
  - access application
  - connect components
- Problems solved:
  - thin client
    - limited processing on the client side
    - most processing are done on the server
  - use the “standardised” web interface and protocol
    - further increase availability & accessibility
- Problems remaining:
  - inherits client-server problems

Software Evolution - SuperComputing

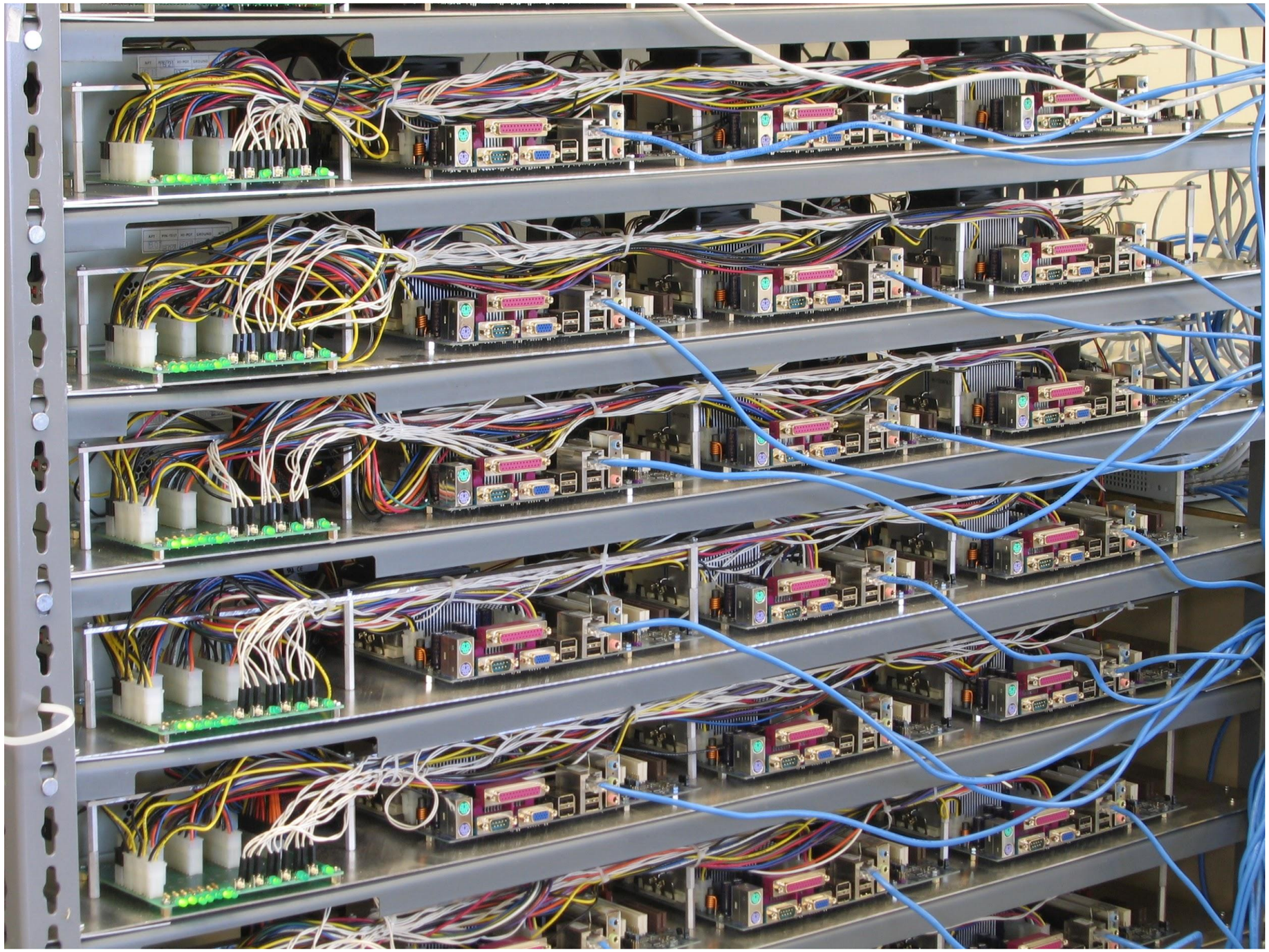
Software Evolution - SuperComputing

http://www.top500.org/

The Top 500 list

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	National Supercomputing Center in Wuxi China	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRCPC	10,649,600	93,014.6	125,435.9	15,371
2	National Super Computer Center in Guangzhou China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
3	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
4	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
5	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
6	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945

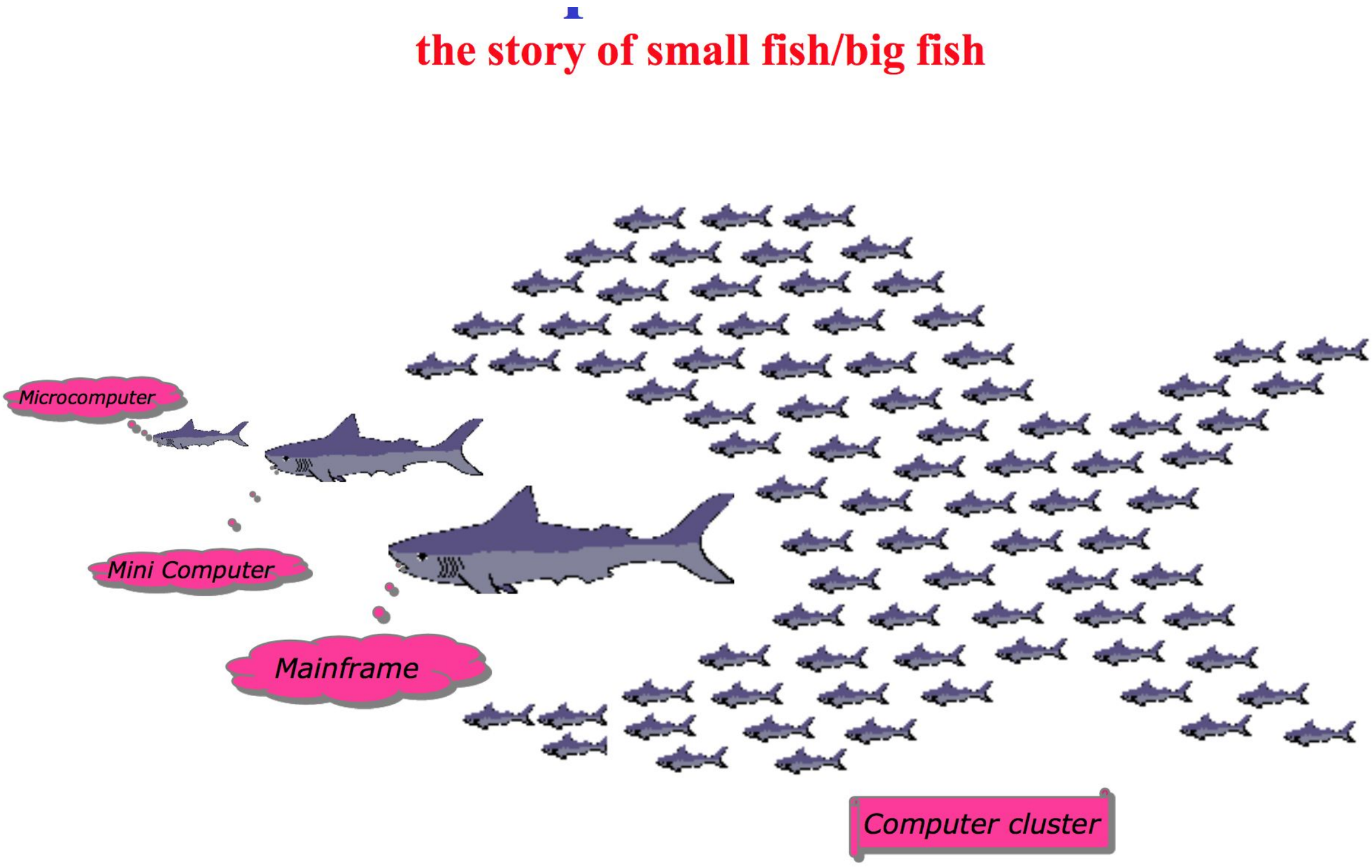
Cluster Computing



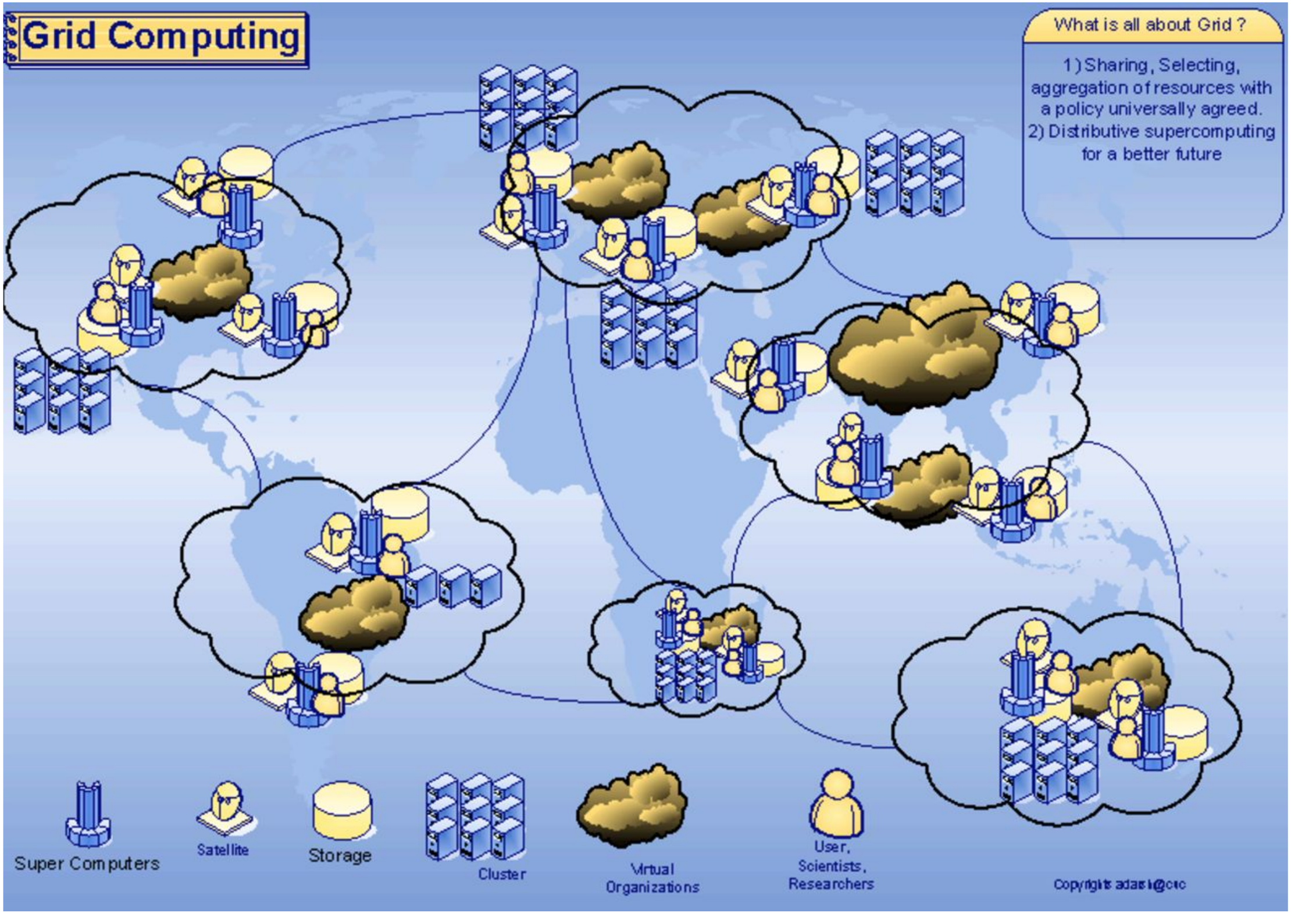
Cluster Computing

- Clusters are linked computer systems that can co-operate to perform computations and deliver services:
  - Often function / appear as single server.
  - Typically linked over LAN.
  - Offers scalability over single-server.
  - Used for high-availability (redundant), load balancing, shared compute

Cluster Computing



Grid Computing



Grid Computing

- Grids are autonomous and dynamic distributed resources contributed by multiple organisations.

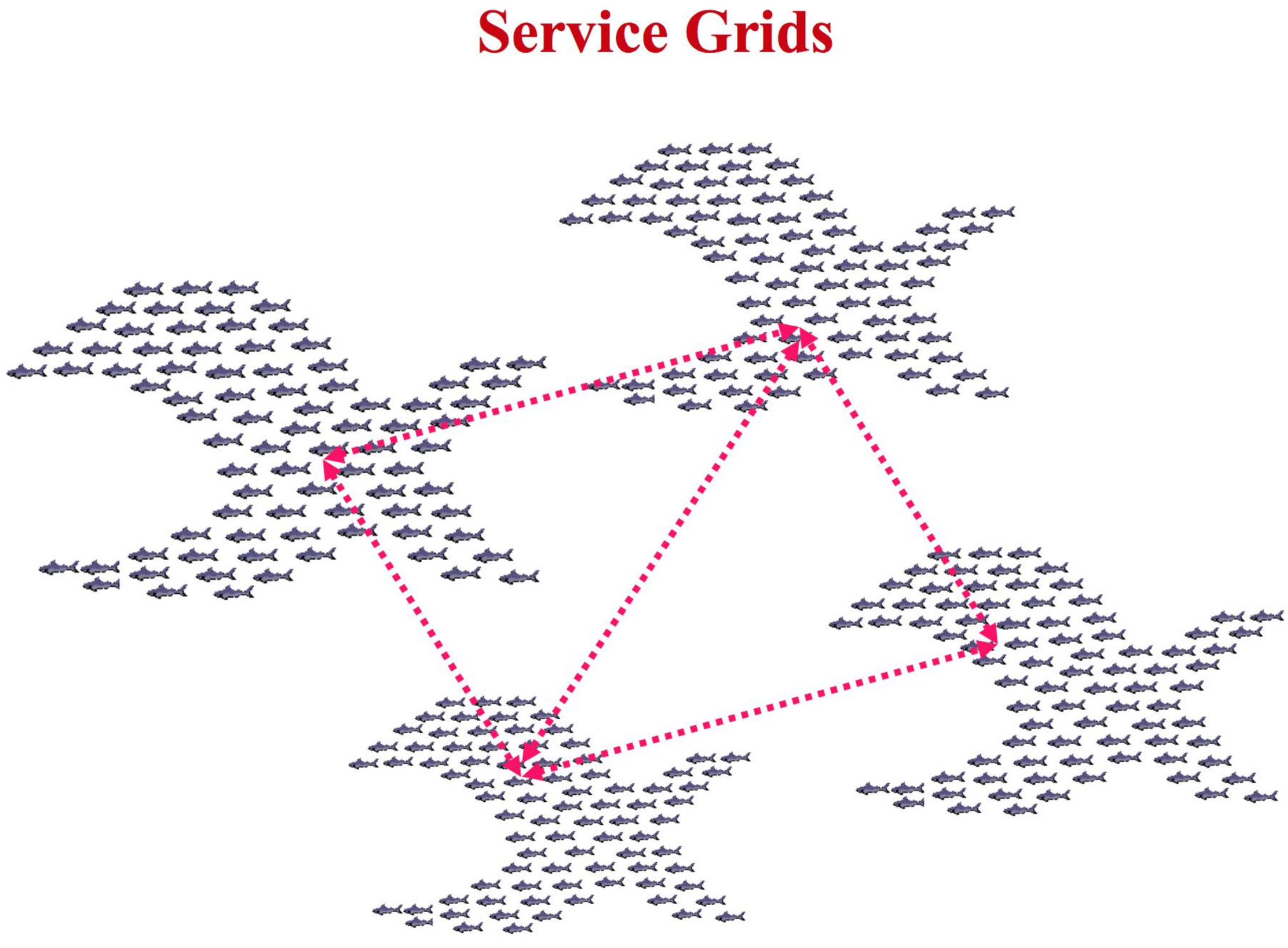
- Solve complex problems with large computational power

- Can offer computing, network, sensor and storage resources.

- Resources are loosely coupled, heterogeneous, geographically dispersed.

- Used in diverse fields: climate modelling, drug design and protein analysis to solve “Grand Challenges”

Grid Computing



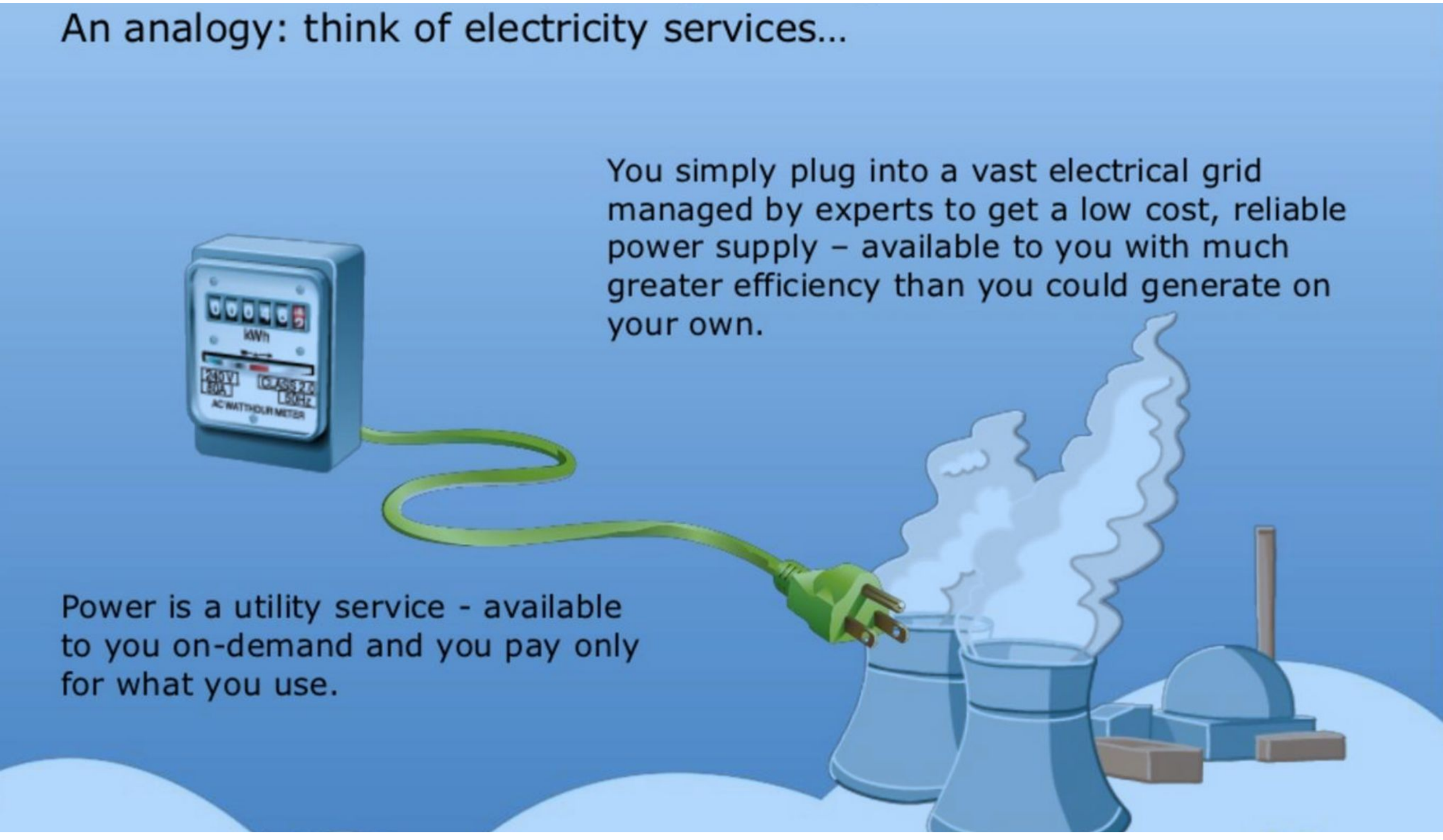
Grid Computing - Use case

- **BOINC** - Berkeley Open Infrastructure for Network Computing - Enables Grid Computing.

Application of BOINC

- **Folding@home** - Simulates protein folding, drug design and molecular dynamics
- **Einstein@home** - Searches for signals from rotating neutron stars in the data generated from telescopes
- **SETI@home** - Analyze radio signals, searching for signs of extraterrestrial intelligence

What is Cloud Computing?



*

What is Cloud Computing?

Cloud Computing is also a utility service - giving you access to technology resources managed by experts and available on-demand.

You simply access these services over the internet, with no up-front costs and you pay only for the resources you use.

Cloud Computing….

- Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand.

- It is a model for enabling, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services), which can be rapidly provisioned and released with minimal management effort.

Wikipedia

Cloud Computing….

- “Cloud computing has the following characteristics:
  (1) The illusion of infinite computing resources…
  (2) The elimination of an up-front commitment by Cloud users…
  (3). The ability to pay for use…as needed…”
  – **UCBerkeley RADLabs**

Cloud Computing….

- “... a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

- National Institute of Standards and Technology (NIST)

Cloud Computing….

- Cloud computing, often referred to as simply “the cloud,” is the delivery of on-demand computing resources—everything from applications to data centers—over the Internet on a pay-for-use basis.

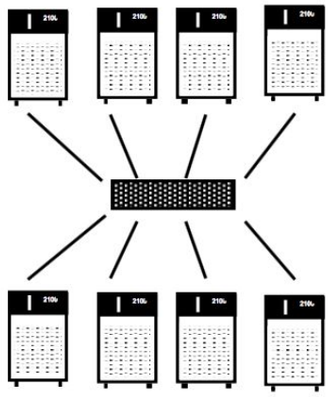
- **IBM**

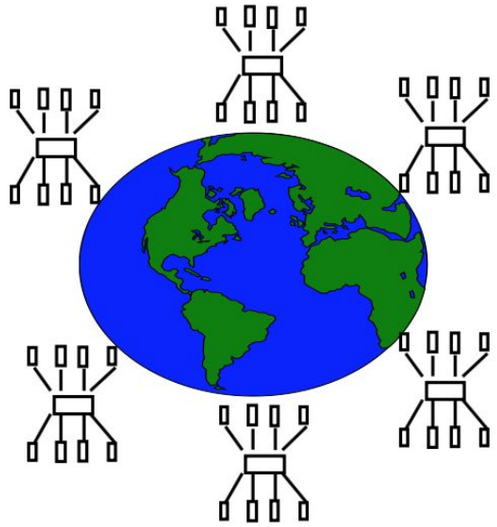
Common ground - Cloud Computing..

- Pay as you go. Pay for what you use
- Self-service interface
- Elastic capacity - scale up/down on demand.
- Resources are abstracted / virtualized and endless.
- Faster time to market. Focus on your business

Progress to Cloud Computing

Ability to solve massive tasks efficiently





Computers

PCs

Super-computer

Cluster

Grid

Cloud

Enabling Technologies

- Virtual Machines / Containers
- Virtualized Storage
- Load Balancing
- AutoScaling
- Virtualized Network
- Monitoring
- Metering
- Multi Tenancy
- Self-healing

Virtualization

Virtual Machines - Run multiple instances on the same machine. Shared machines

Virtual Network - Provides a dedicated network segments in one LAN. Shared network

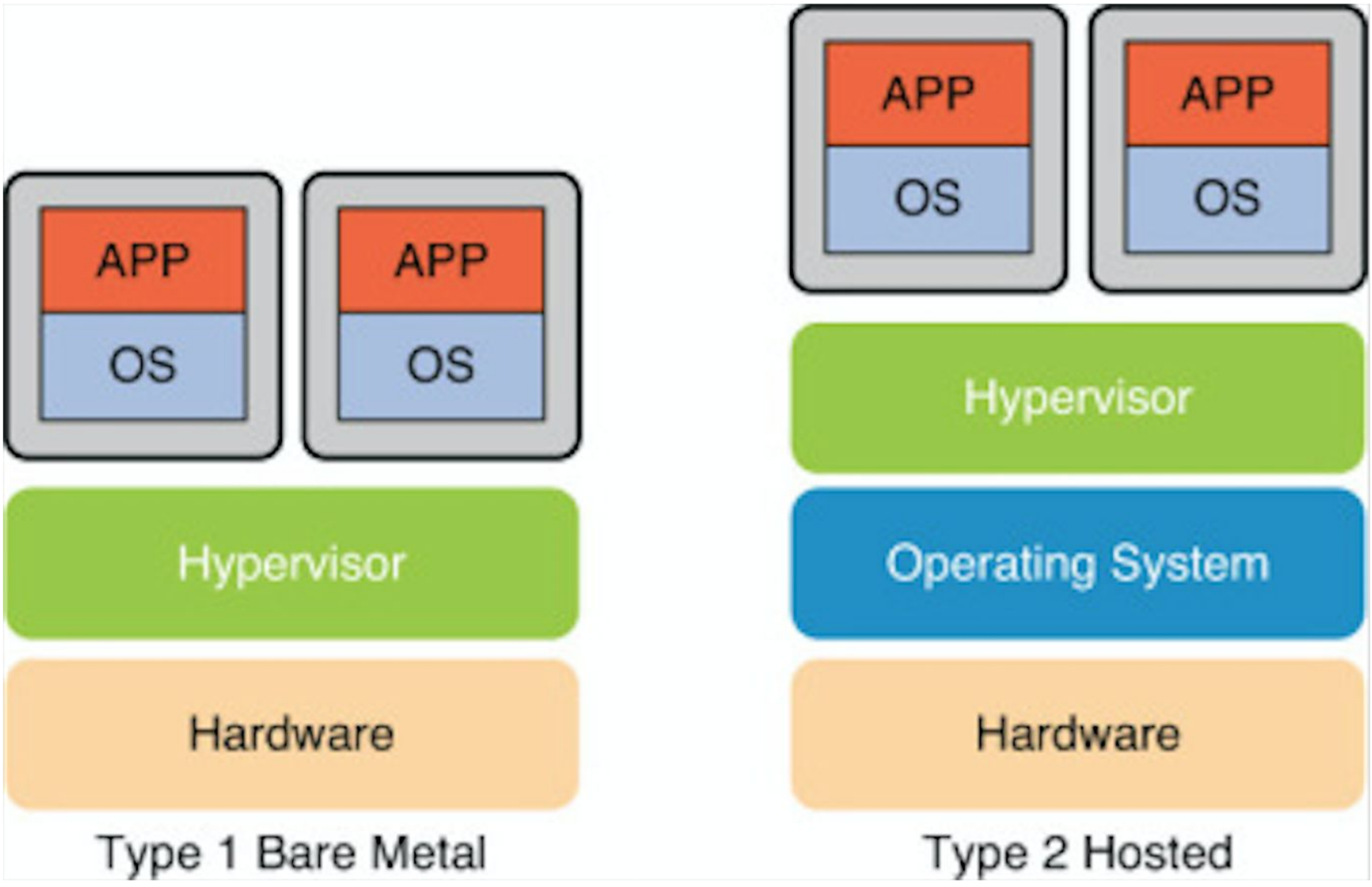
Virtual Storage - Provides a logical space for you to store data. Actual physical location (probably shared) of the data is handled by the udnerlying system.

Shared, shut-down, restored, re-allocated on-demand!

Enabling Technologies

- Virtual Machines / Containers - Cloud resources
- Virtualized Storage - Cloud resources
- Load Balancing - Cloud resources
- AutoScaling - Cloud resources / Elasticity
- Virtualized Network - Cloud resources
- Monitoring - Pay-as-you go model
- Metering - Pay-as-you go model
- Multi Tenancy - Cloud resources
- Self-healing - Cloud resources

Virtualization



Hypervisors

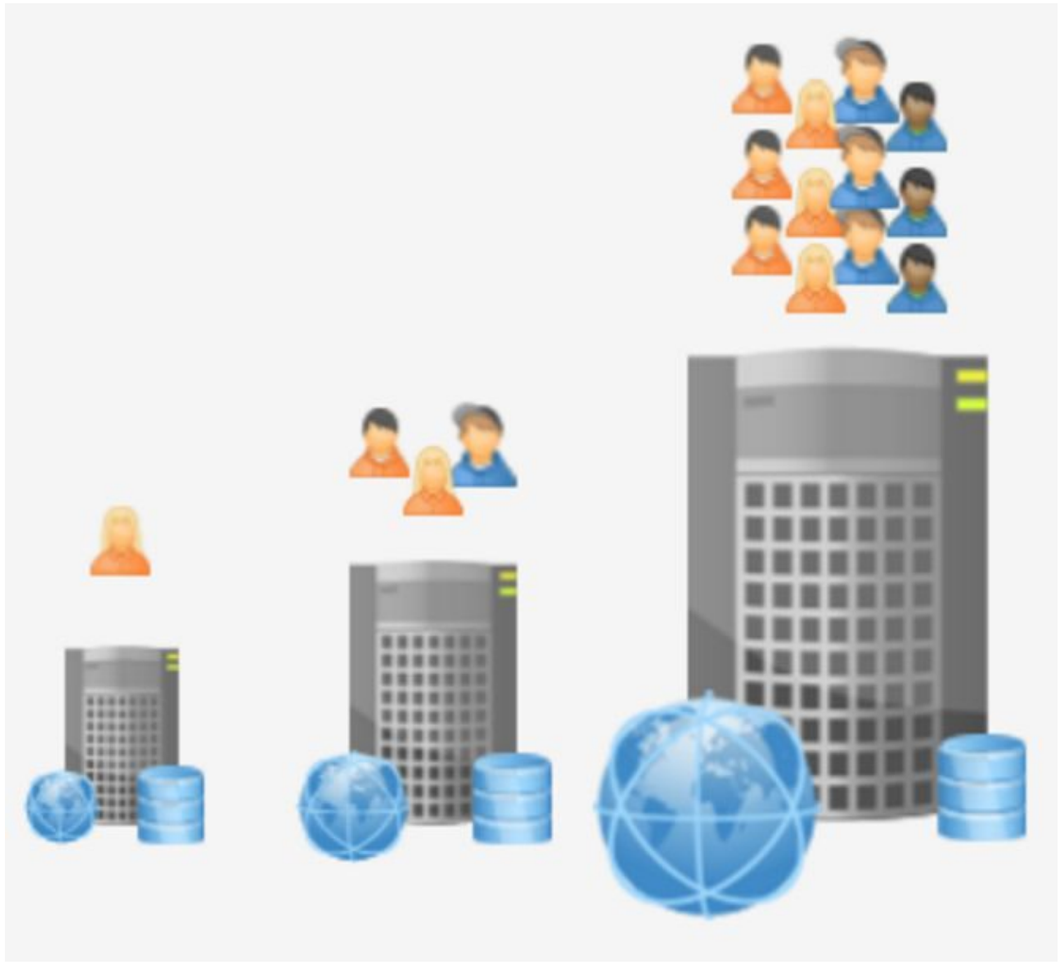
- Type 1, which is considered a bare-metal hypervisor and runs directly on top of hardware. The Type 1 hypervisor is often referred to as a hardware virtualization engine.
- Type 2, which operates as an application on top of an existing operating system.

The need to Scale

- Demand to services fluctuates
  - fluctuation occurs in different time scales
    - year-long cycle/pattern: Christmas shopping, summer travelling
    - fluctuation within 90 min: happy hour credit card deal
- Fixed resource provisioning results in under/over-provisioning
  - **under-provisioning**: slow system response, unable to serve clients
  - **over-provisioning**: waste of resources

Vertical Scaling

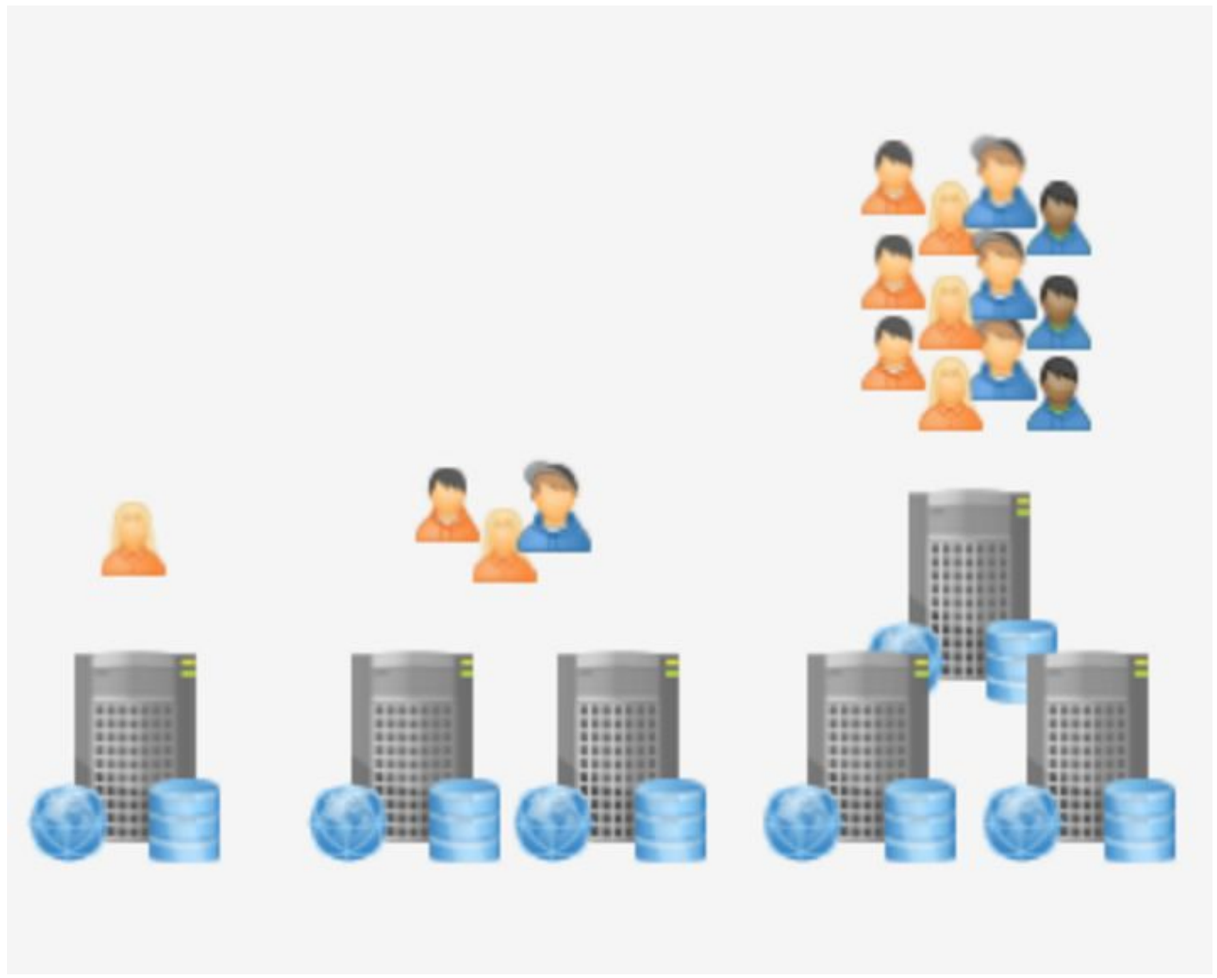
- **Vertical scaling** means that you **scale** by adding more power (CPU, RAM) to an existing machine
- **Scale up** : more powerful server
- **Scale down** : less powerful server



Auto Scaling

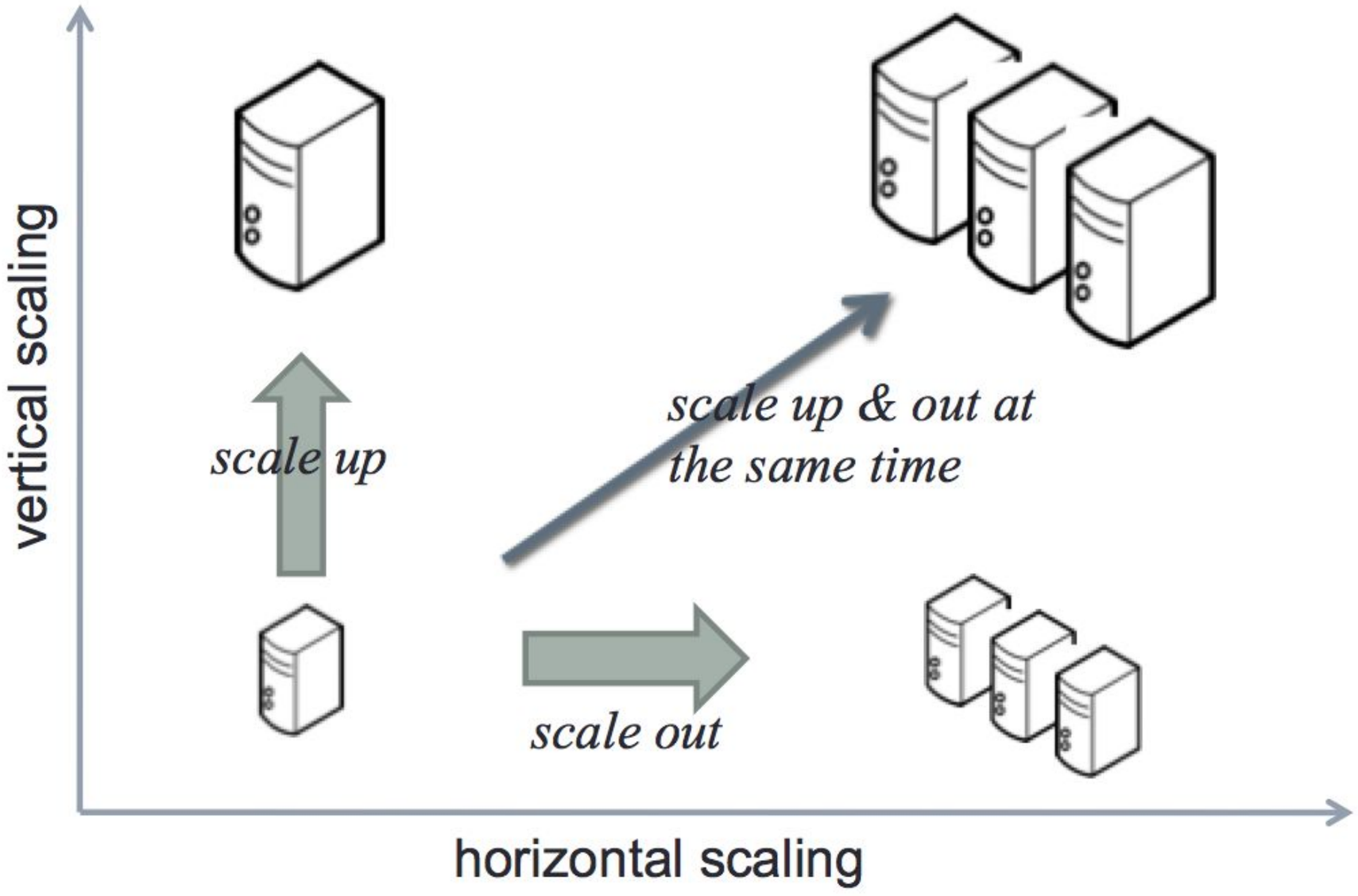
Horizontal Scaling

- **Horizontal scaling** means that you **scale** by adding more machines into your pool of resources
- **Scale out** : more instances
- **Scale in** : few instances

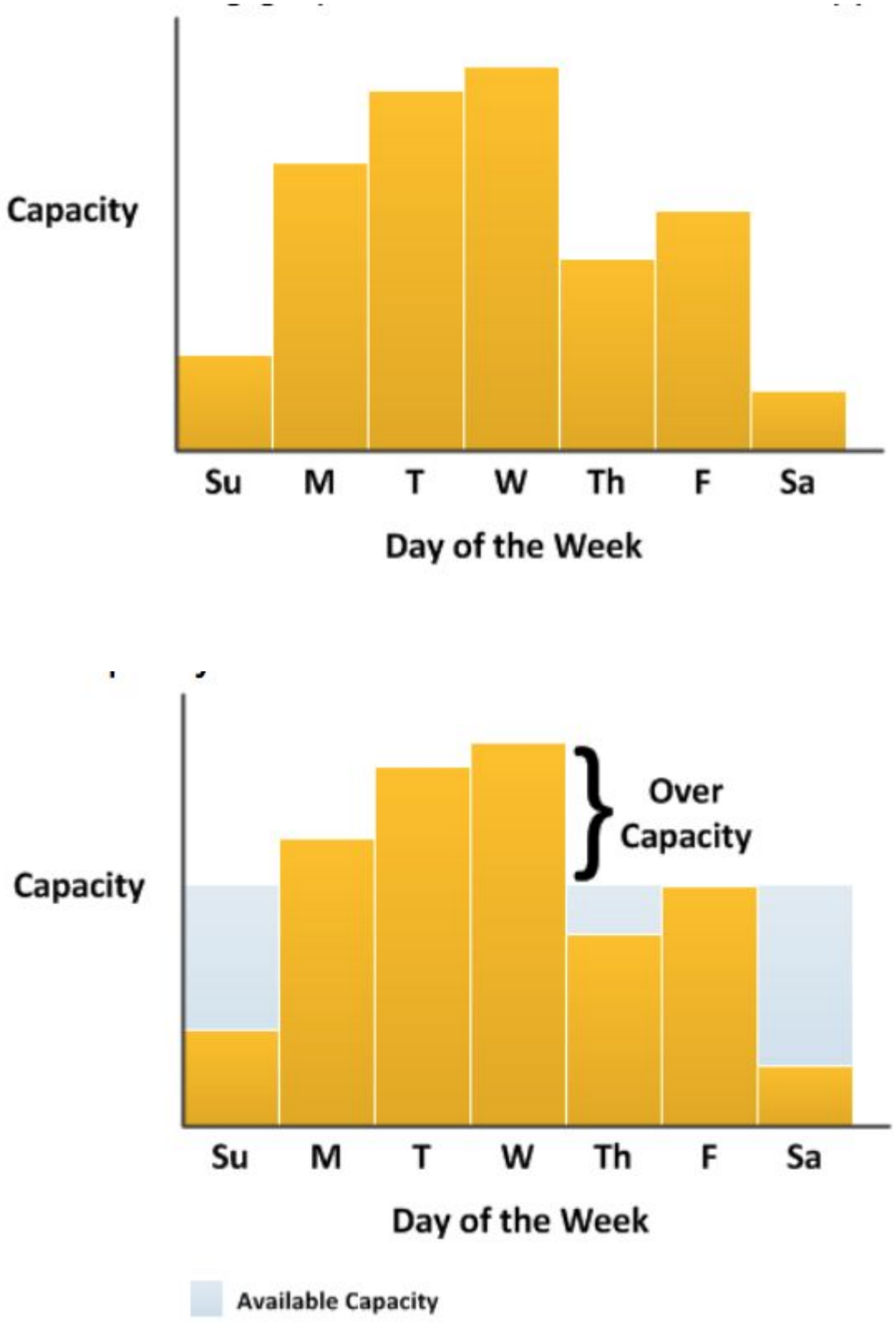


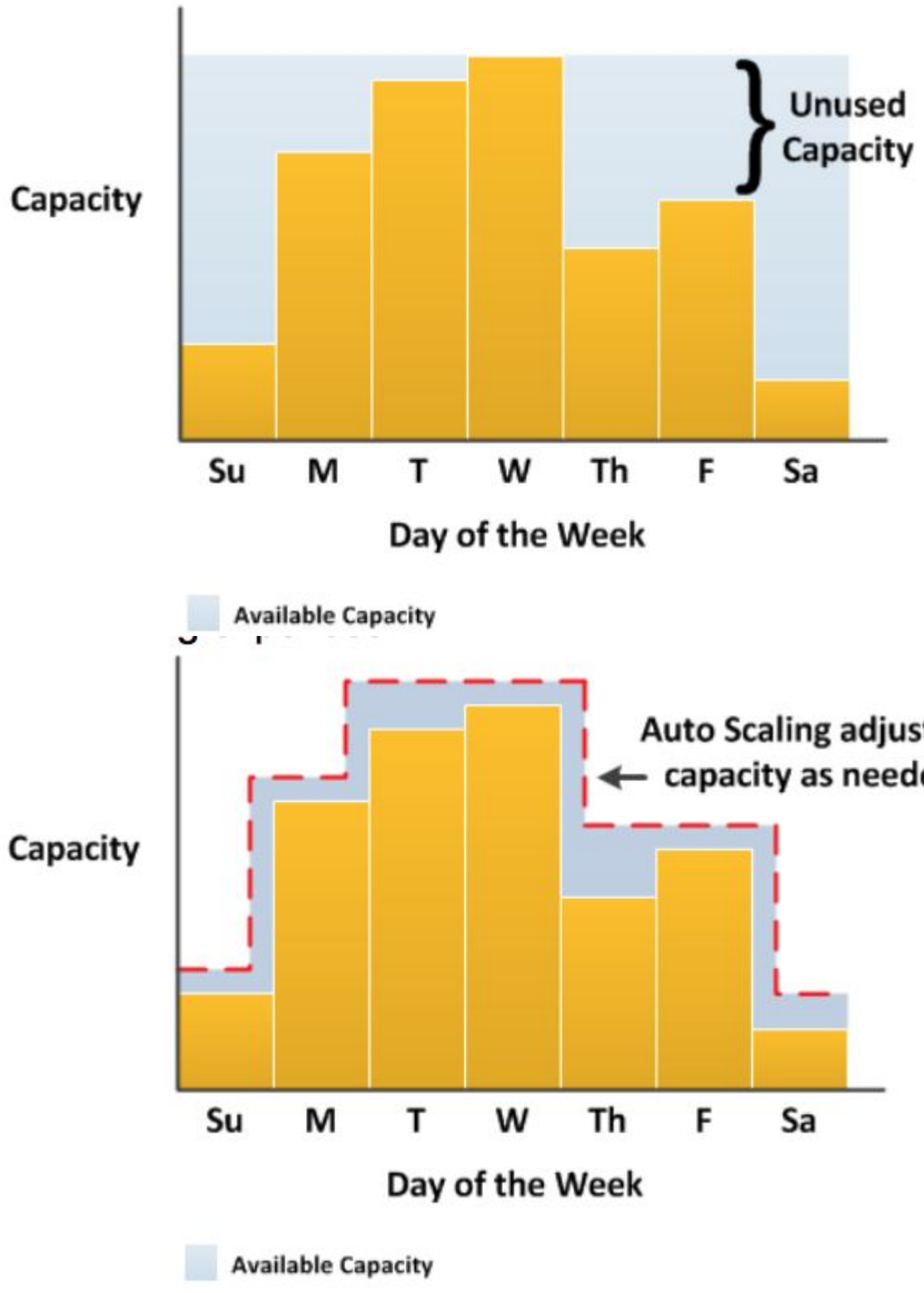
Auto Scaling

Combine scaling

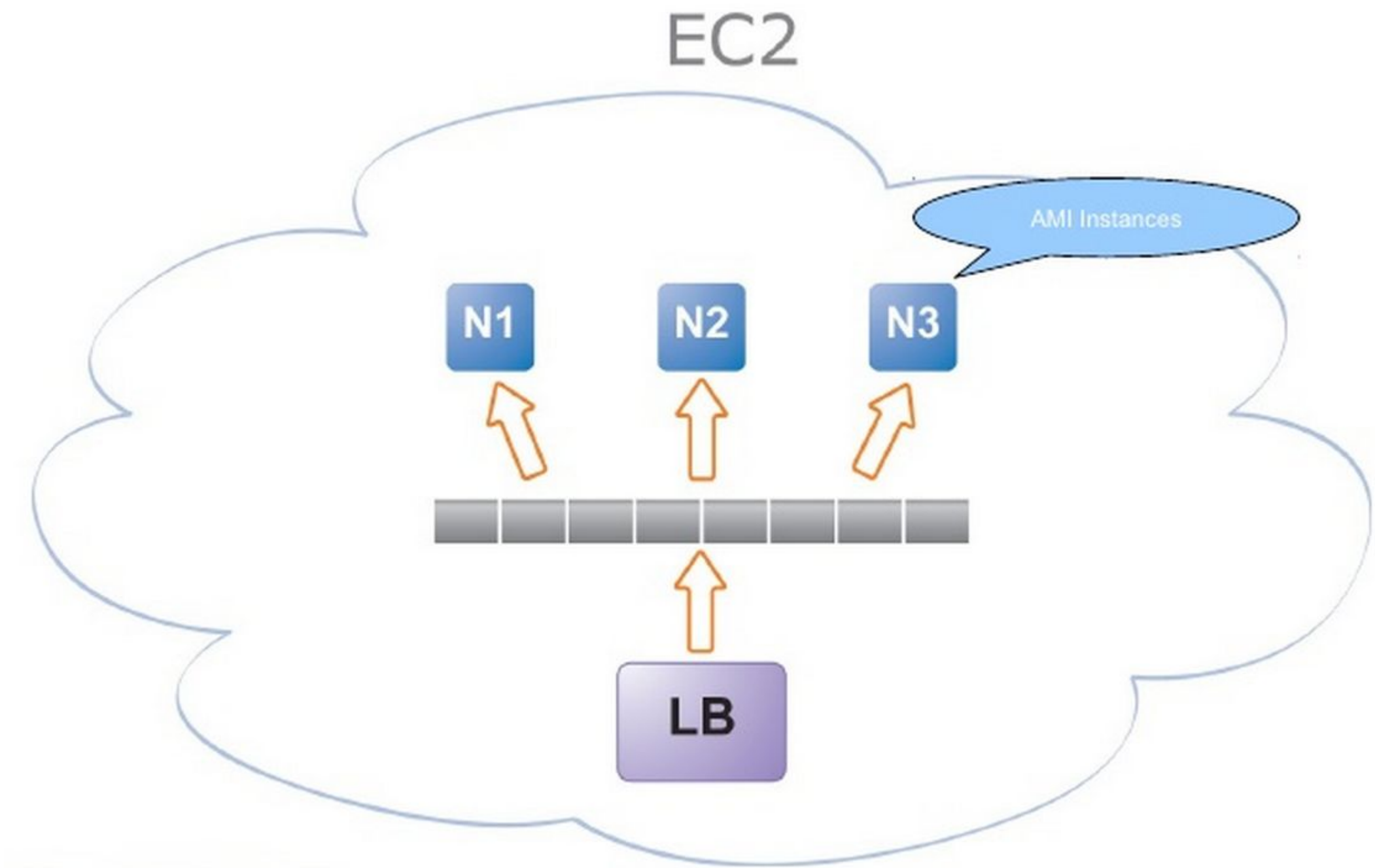


Handling variable Demands

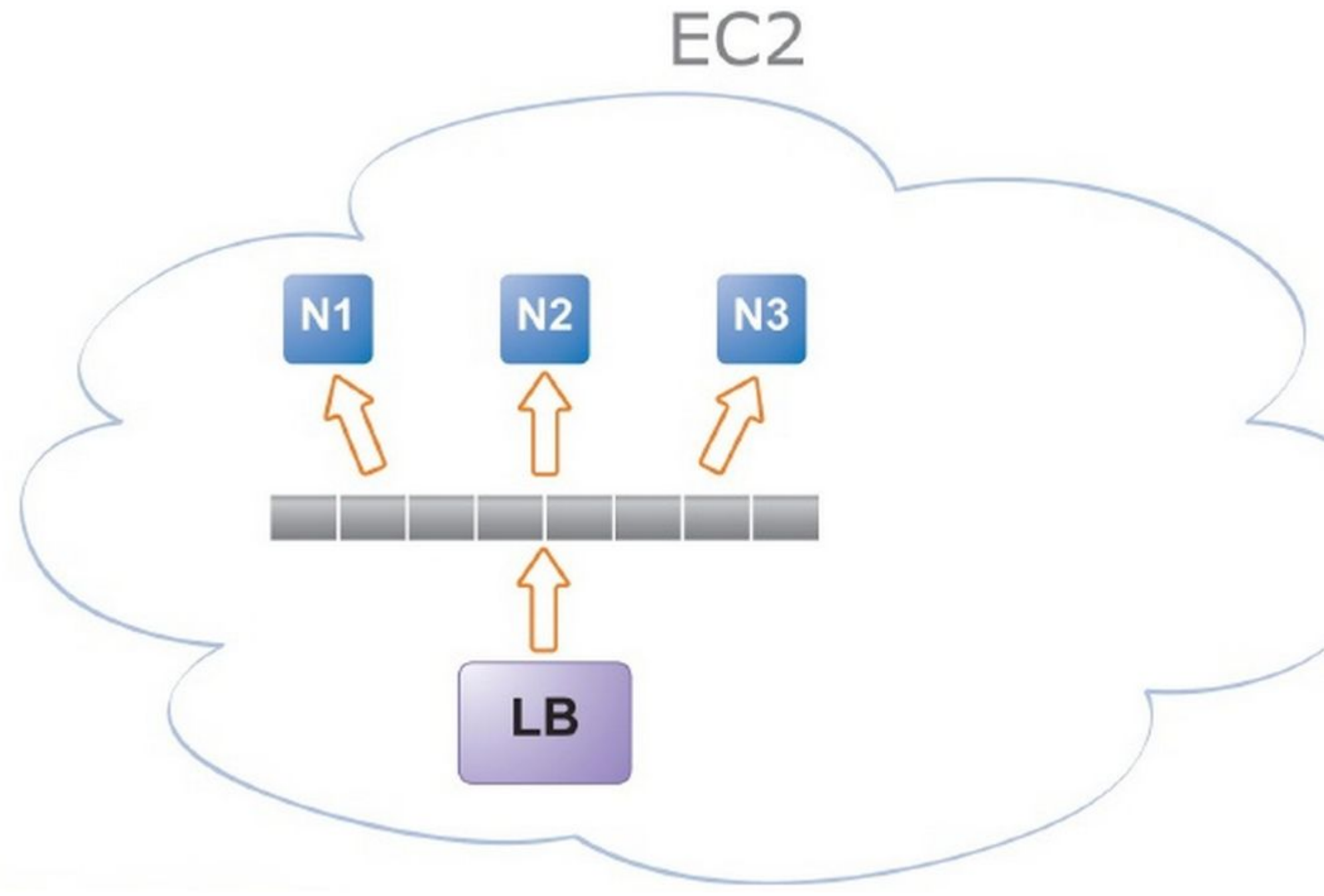




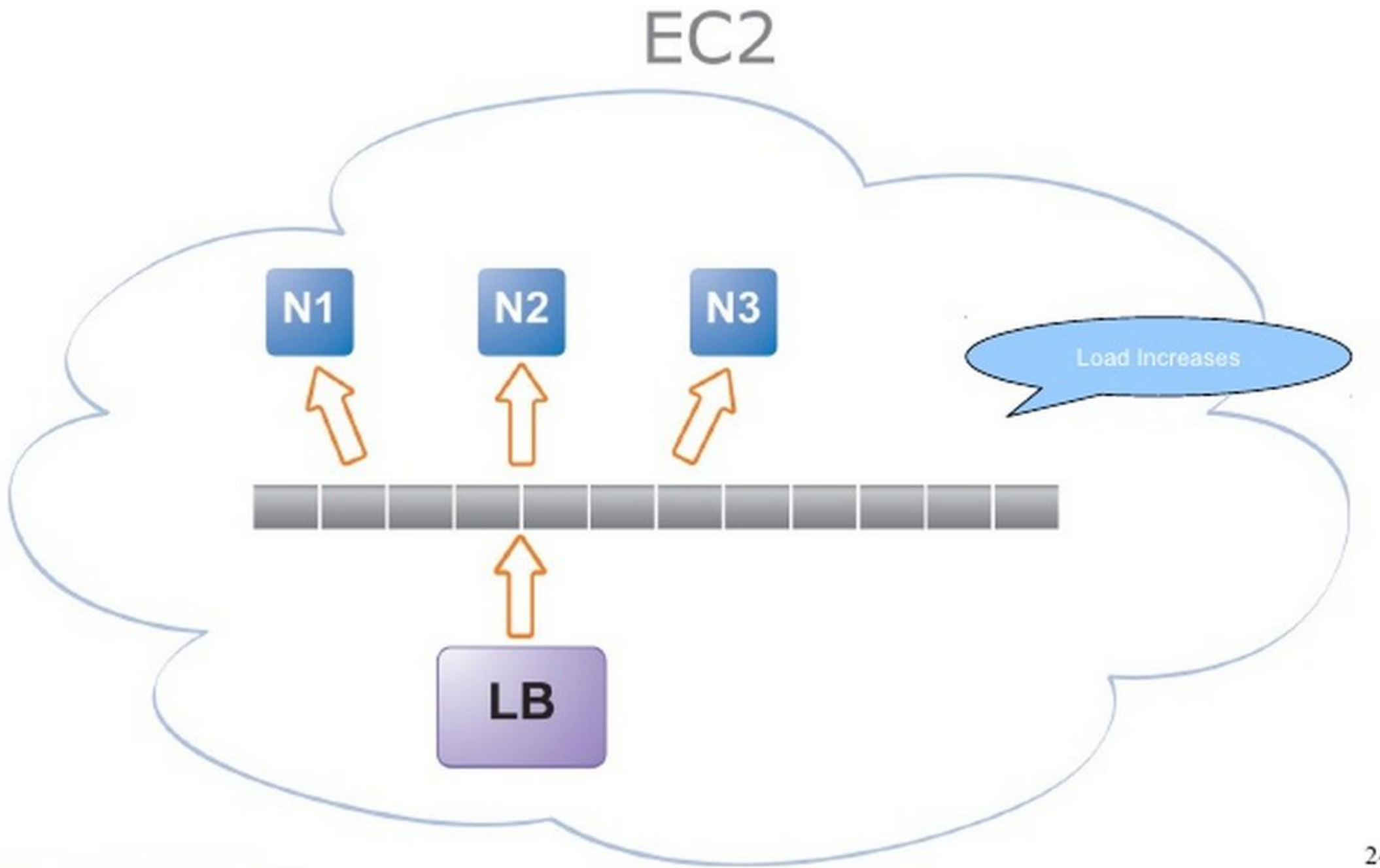
Load Balancing / Autoscaling



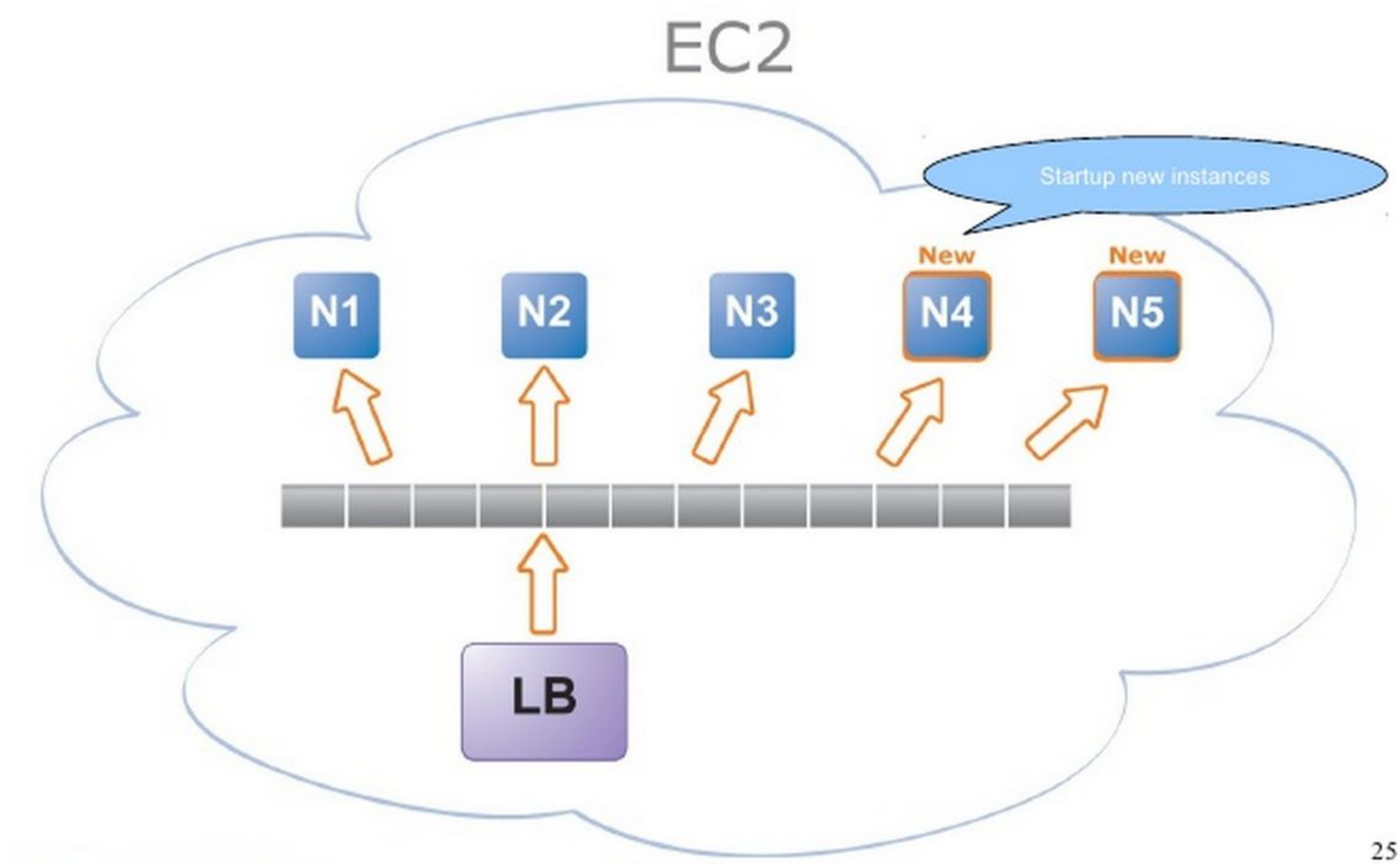
Load Balancing / Autoscaling



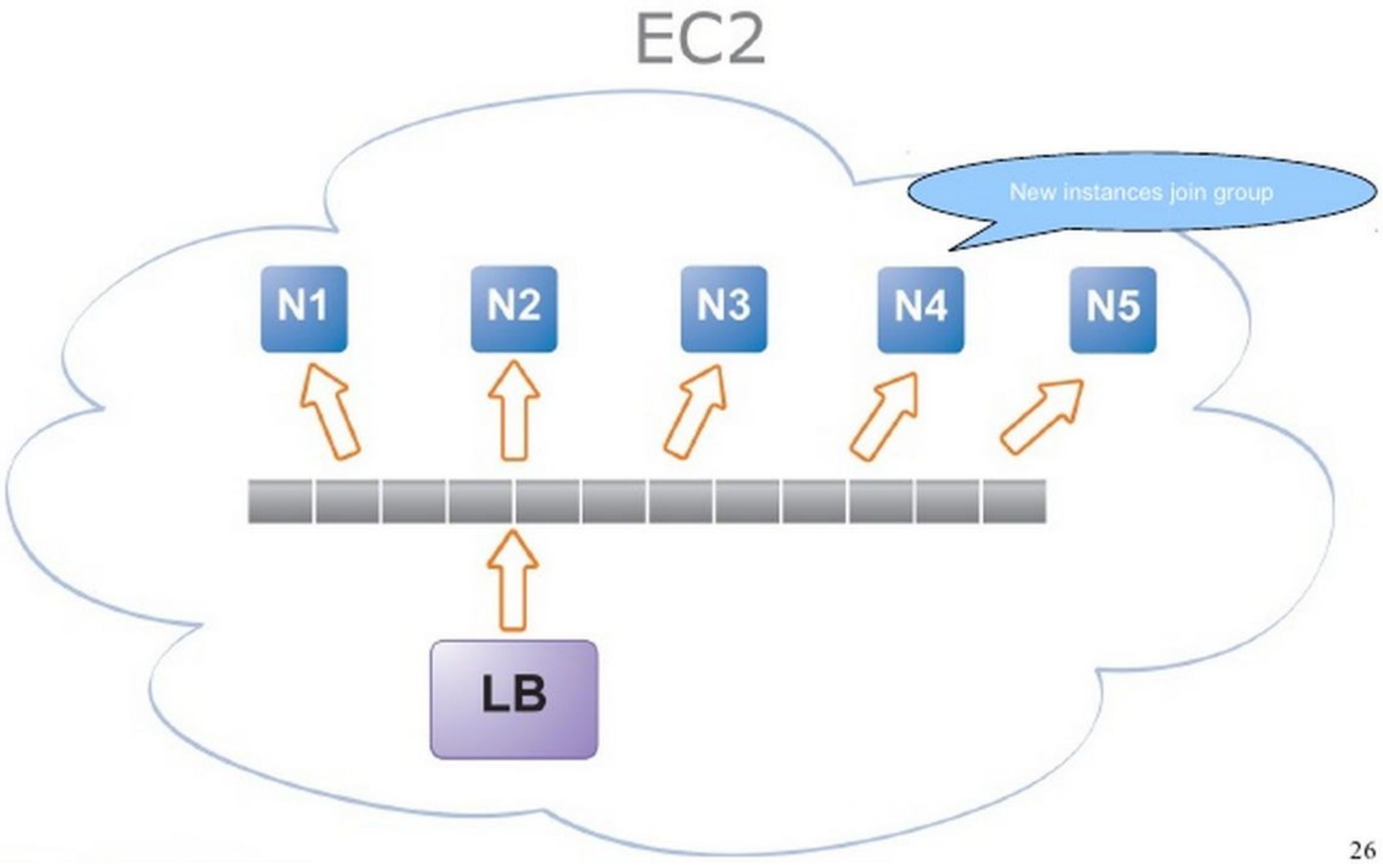
Load Balancing / Autoscaling



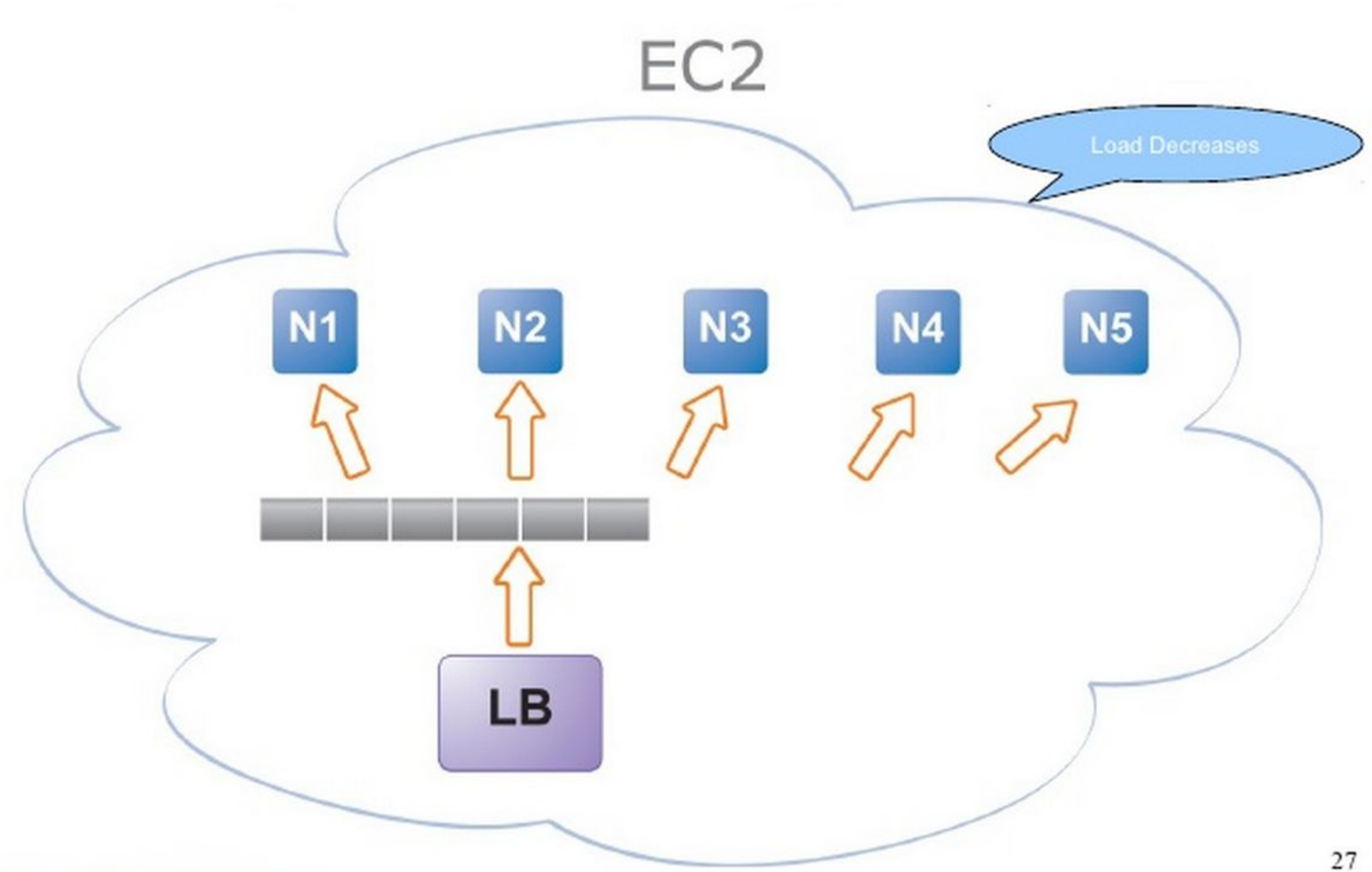
Load Balancing / Autoscaling



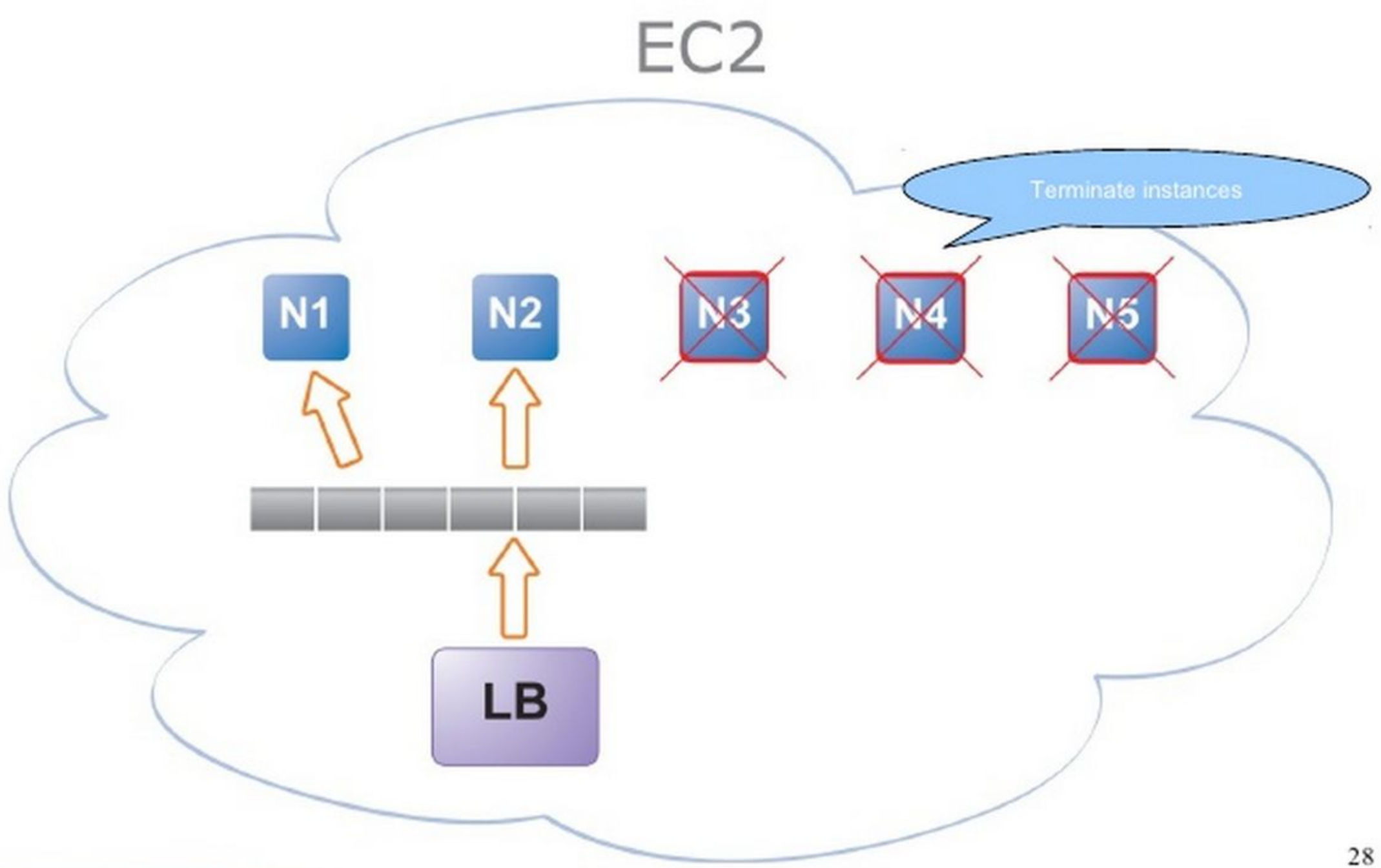
Load Balancing / Autoscaling



Load Balancing / Autoscaling

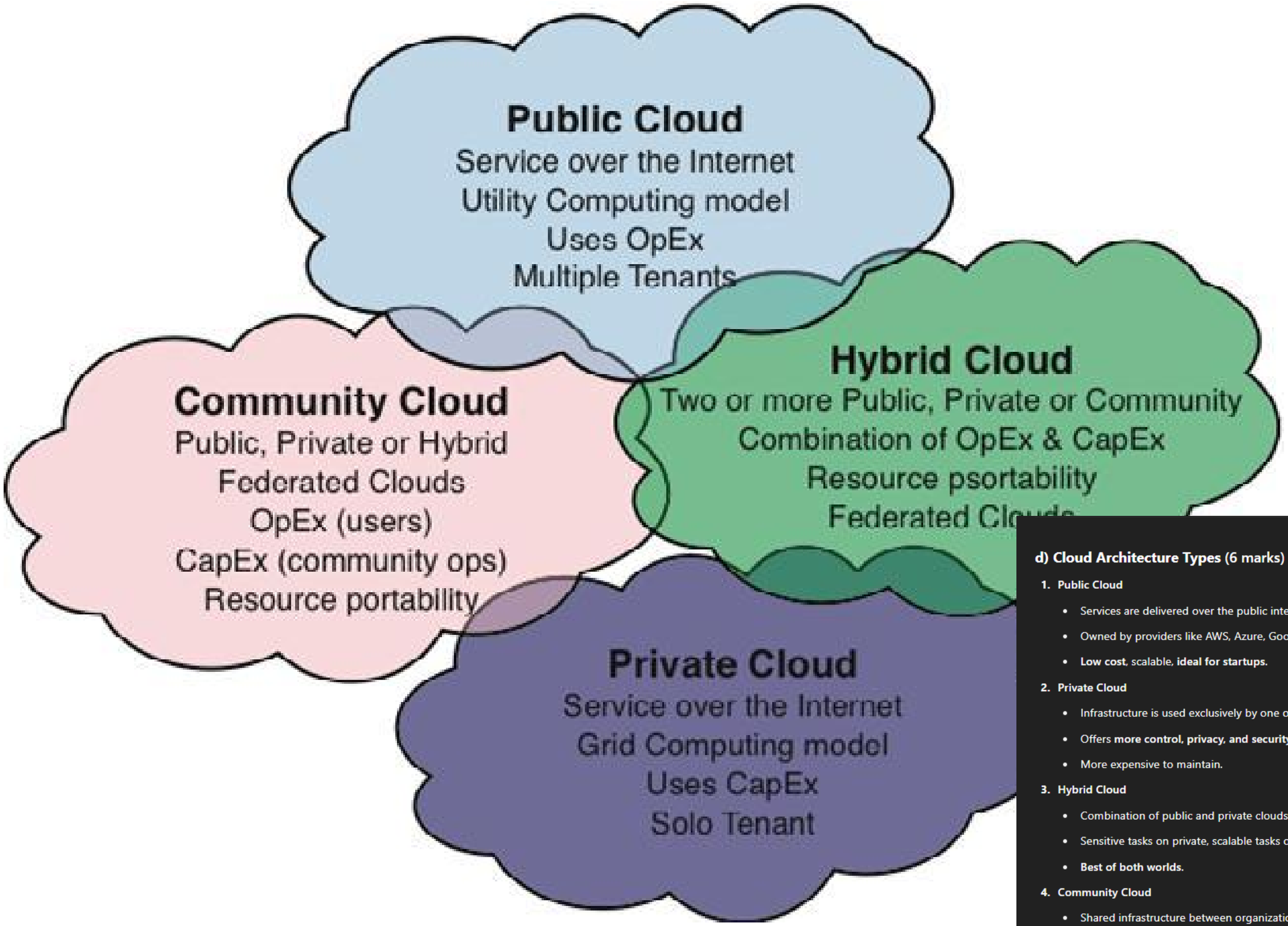


Load Balancing / Autoscaling



Demo

Type of Clouds



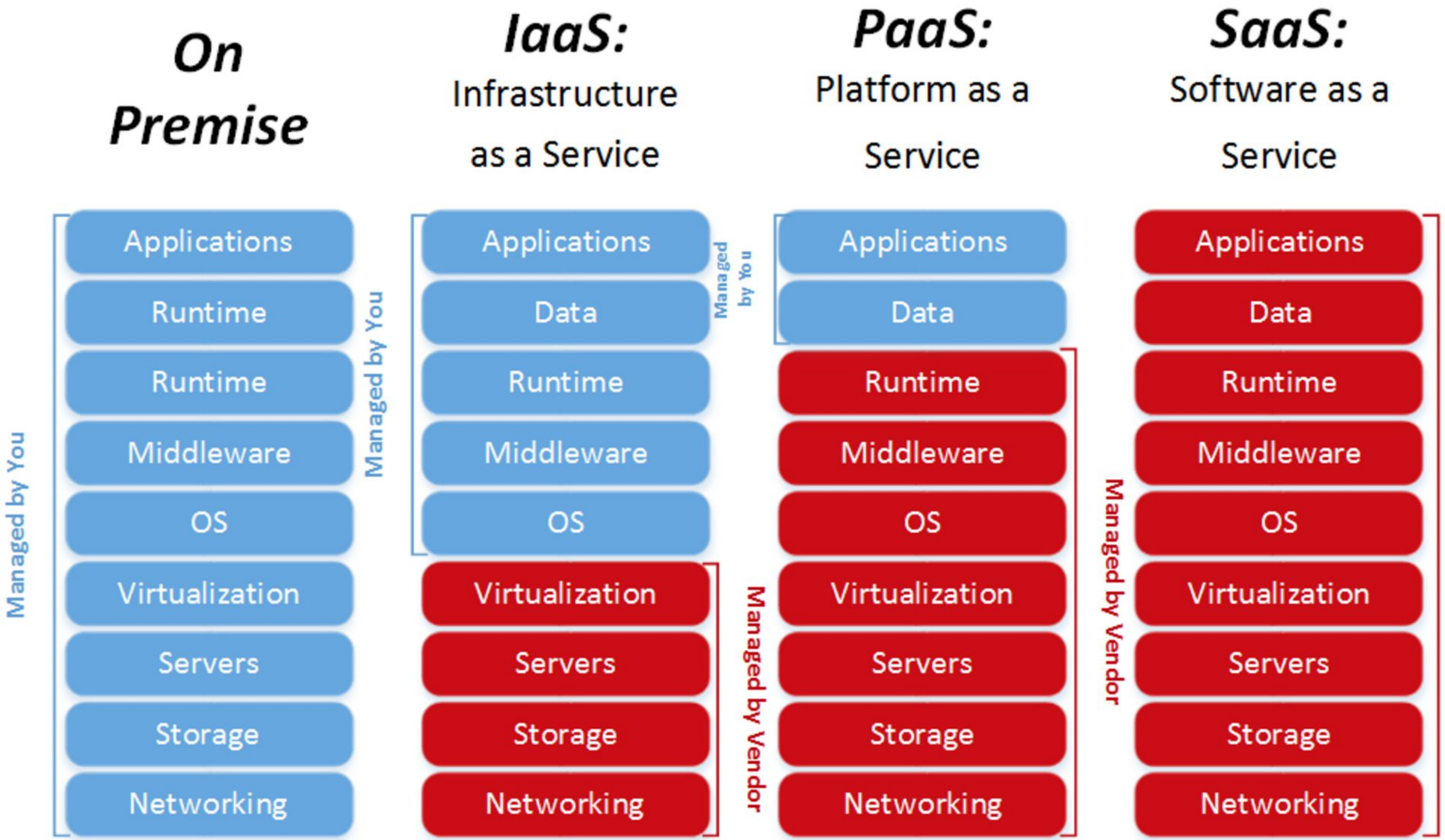
d) Cloud Architecture Types (6 marks)

1. **Public Cloud**
   - Services are delivered over the public internet.
   - Owned by providers like AWS, Azure, Google Cloud.
   - **Low cost**, scalable, **ideal for startups.**
2. **Private Cloud**
   - Infrastructure is used exclusively by one organization.
   - Offers **more control, privacy, and security.**
   - More expensive to maintain.
3. **Hybrid Cloud**
   - Combination of public and private clouds.
   - Sensitive tasks on private, scalable tasks on public.
   - **Best of both worlds.**
4. **Community Cloud**
   - Shared infrastructure between organizations with similar needs.
   - Managed internally or by a third party.
   - Used in **research or government sectors.**

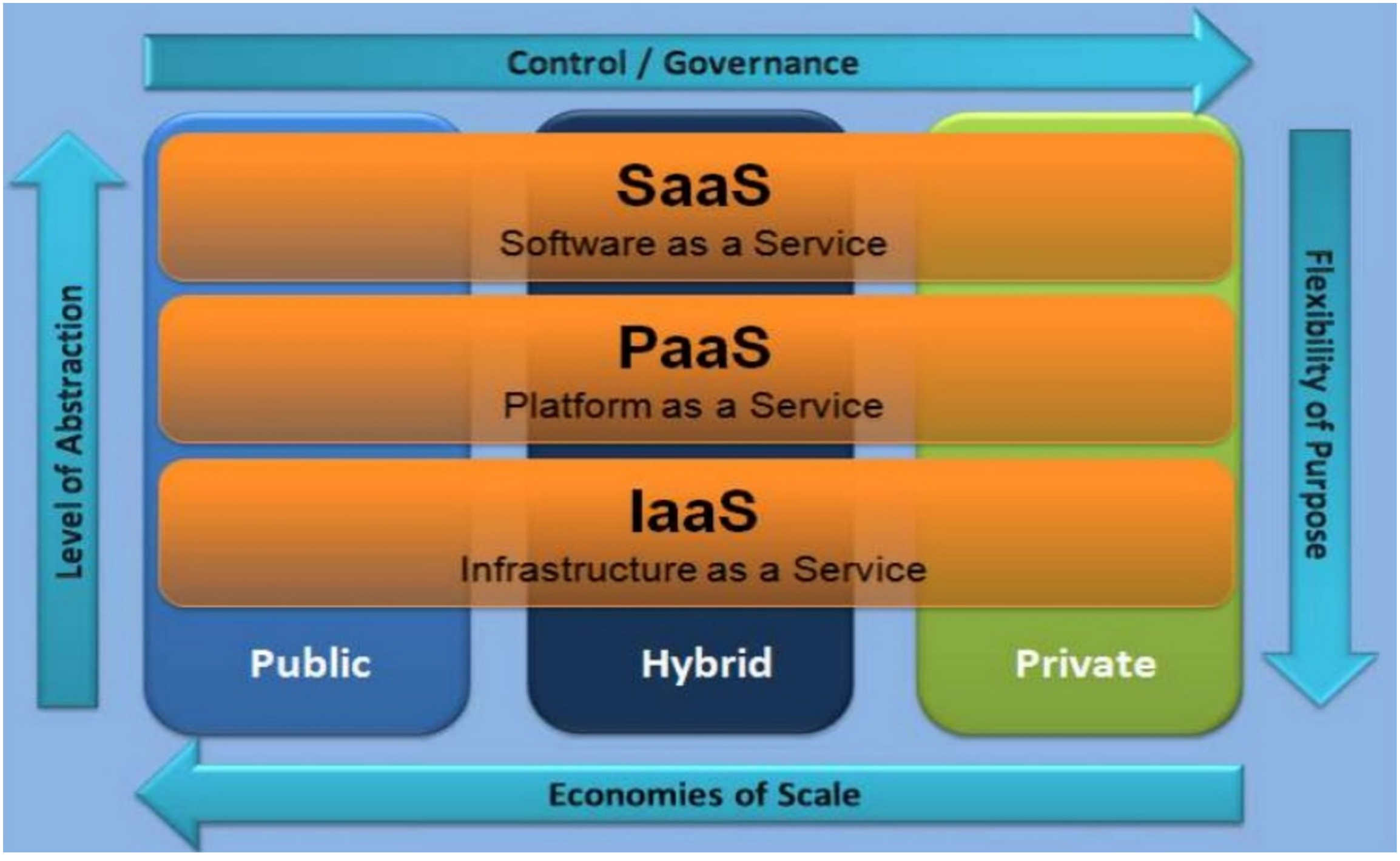
IaaS, PaaS and SaaS

- IaaS - Infrastructure as a Service. Such as diskspace, machine power, networking from the cloud
- PaaS - Platform as a Service. Platform such as tomcat from the Cloud managed
- SaaS - Software as a Service. Software from the Cloud

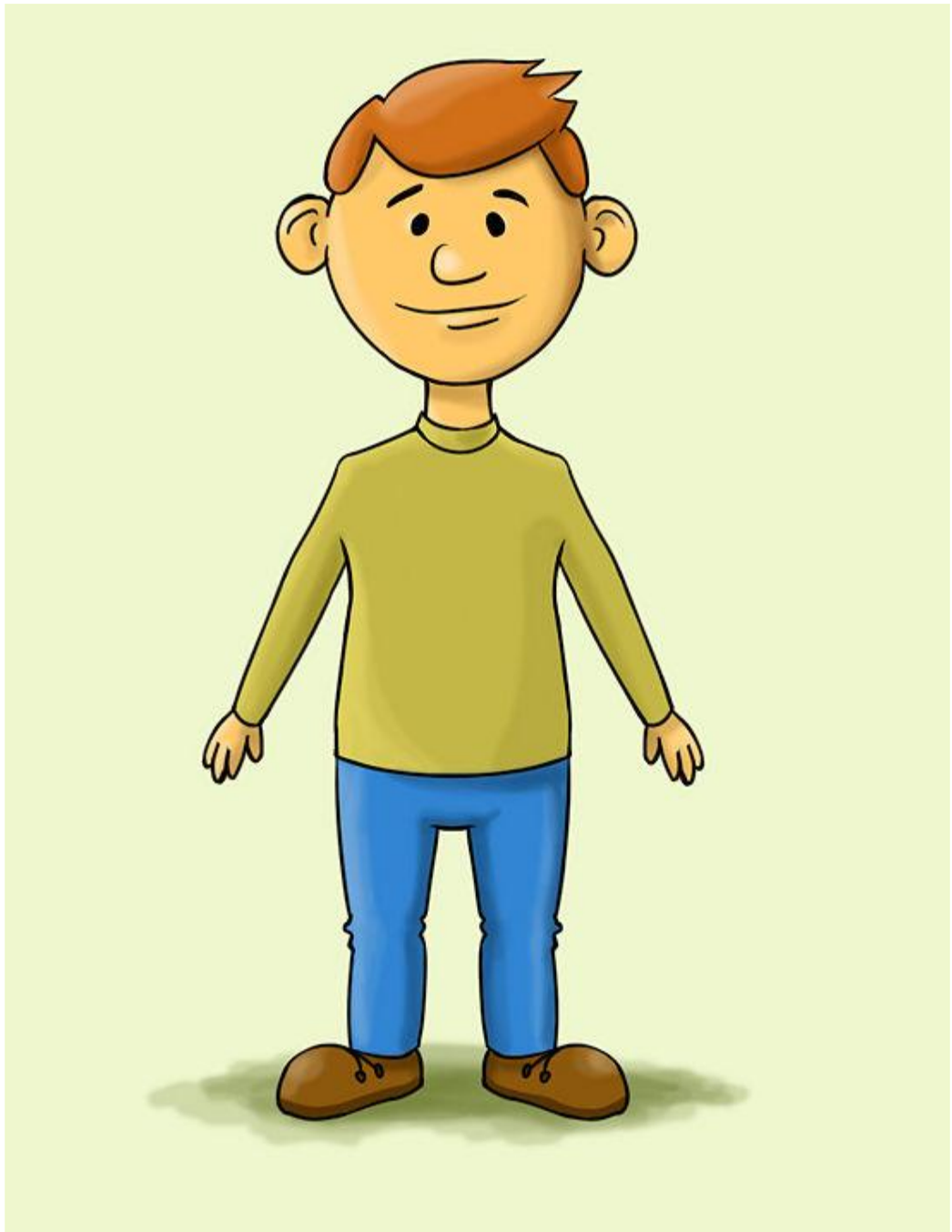
Service Types



Service Model



Please meet Joe...



- My Name is Joe
- I am a software Architect
- I love cloud computing
- I love to take challengers
- I want to find a job

Vacancy - Cloud Architect

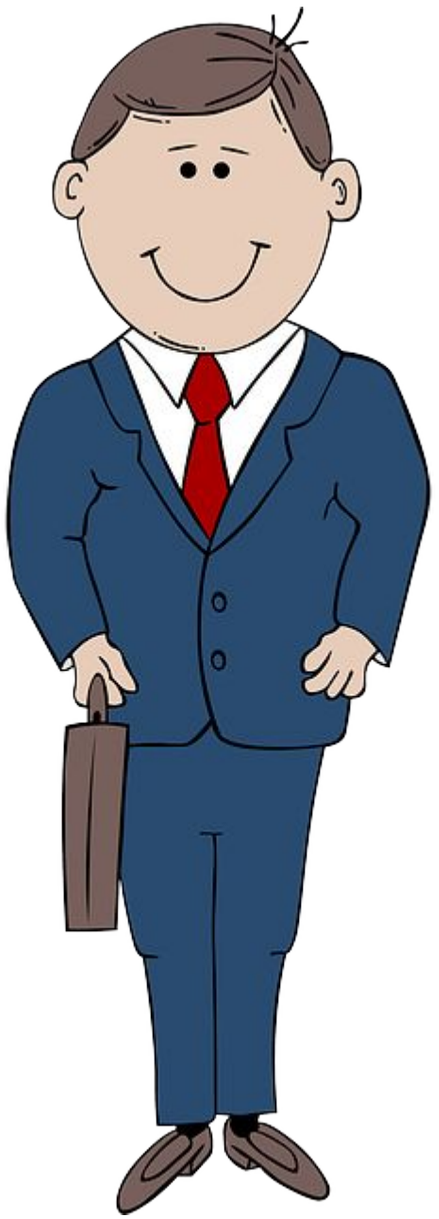
A cloud architect is an IT professional who is responsible for overseeing a company's **cloud computing strategy**. This includes cloud adoption plans, cloud application design, and cloud management and monitoring. Cloud architects oversee application architecture and deployment in cloud environments -- including **public cloud**, **private cloud** and **hybrid cloud**. Additionally, cloud architects act as consultants to their organization and need to stay up-to-date on the latest trends and issues.

Qualifications for the position should include a strong understanding of cloud computing technology and infrastructure as well as experience designing and migrating applications to the cloud. Cloud architects should have experience in a consultant role, as they need to build relationships with customers and team members.

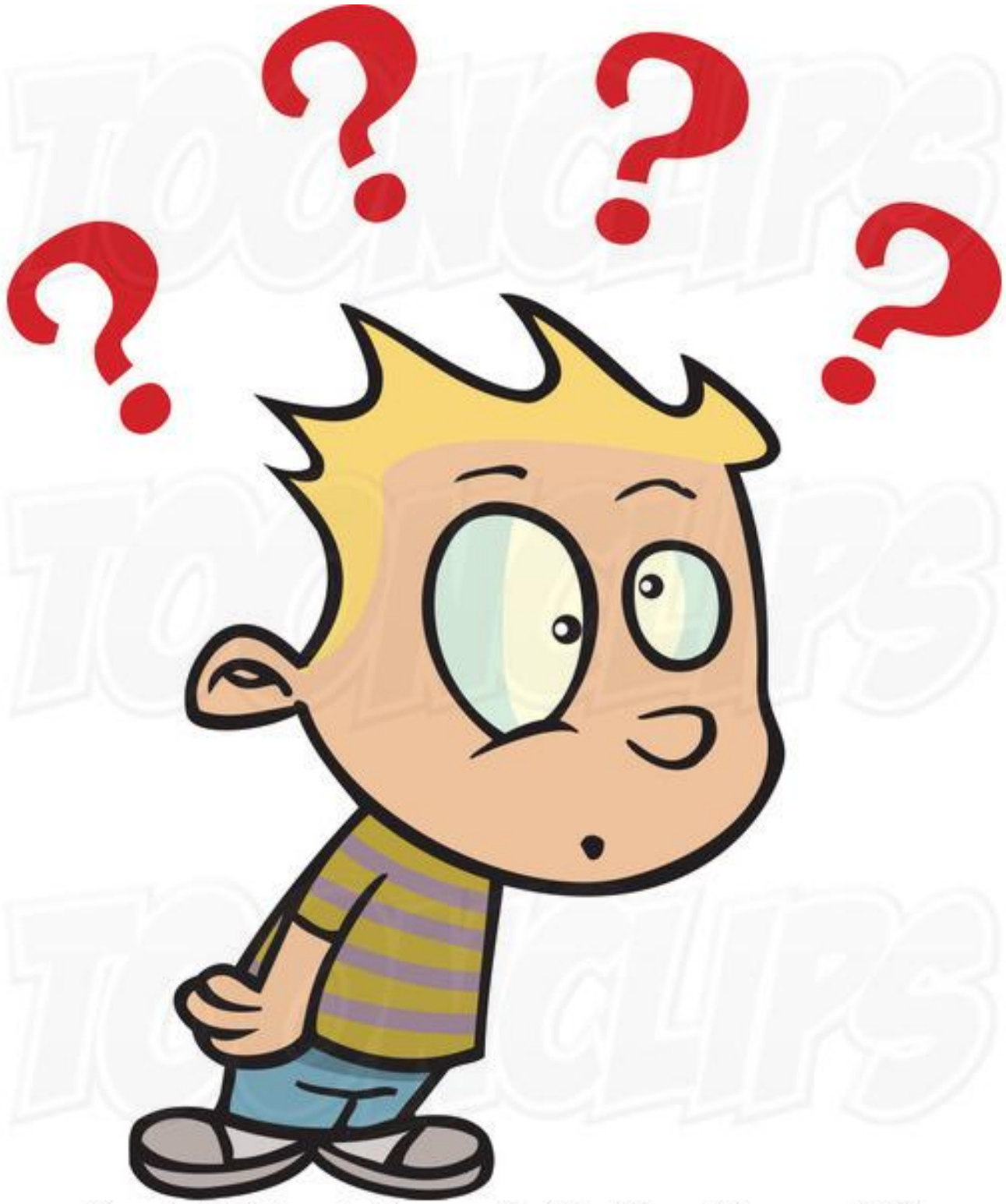
Meeting @ First day

Meeting notes

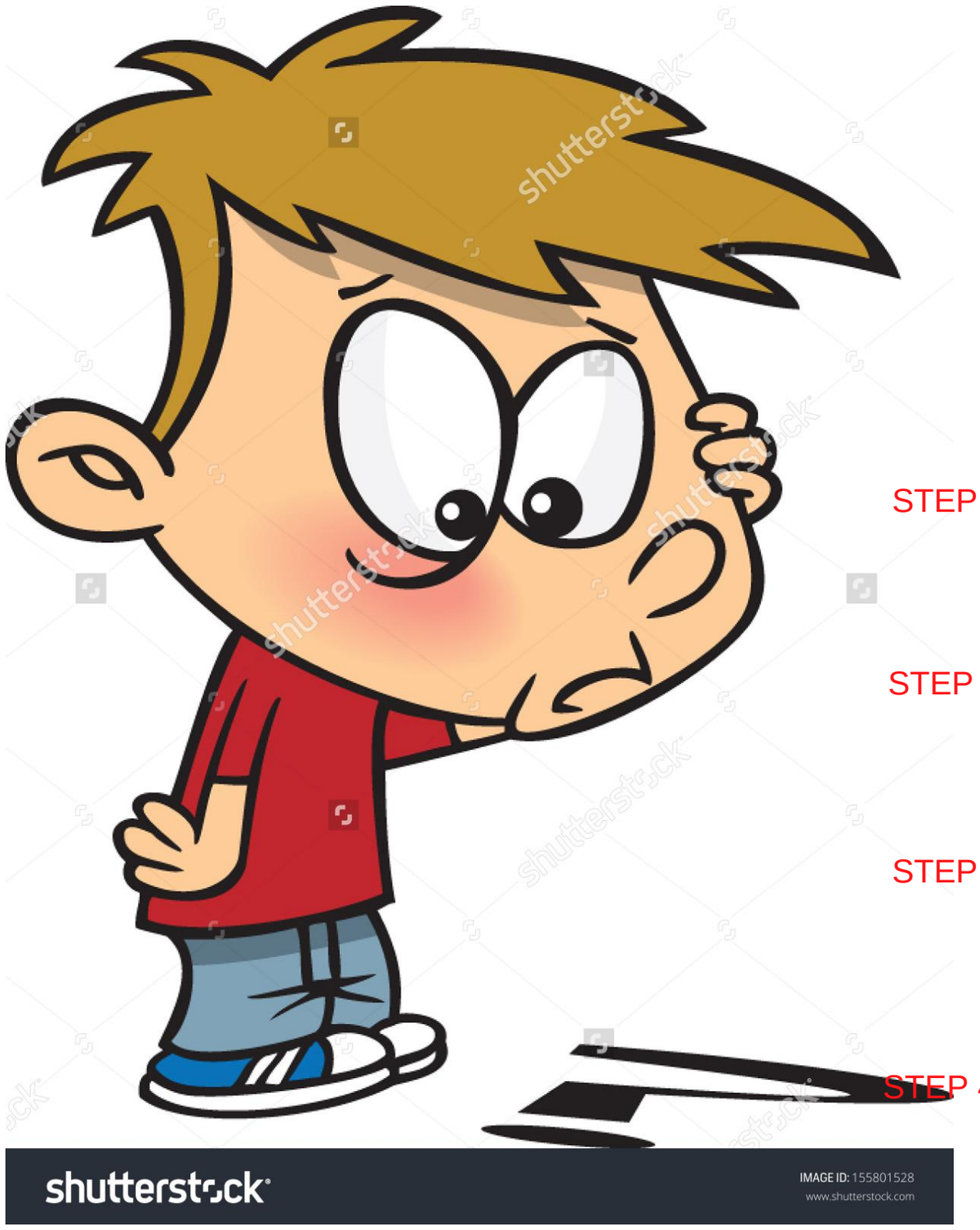
- They are selling Michael Jackson’s music albums
- They have a web App
- Written in PHP
- They have bunch of developers
- They are experts in their business domain
- They want to go to market quickly. (ASAP)
- After launching within couple of weeks they are expecting million of users come to website
- They have an Operation team
- Hired me for Operation team
- My team members
  - Only me :O



My Feeling :)



I start thinking….



- I have learn Cloud Computing
- I trust myself

STEP 1: Choose the Right Cloud Model

Joe starts with IaaS to rent virtual machines and storage from a provider like AWS or Azure.Why? He needs control over PHP setup, but doesn't want to manage physical servers.

STEP 2: Use PaaS to Speed Up Deployment

joe deploys the PHP app using Google App Engine or Heroku, which takes care of:Runtime management, Server provisioning, Automatic scaling

STEP 3: Add Auto Scaling and Load Balancing

oe configures cloud auto-scaling groups.If traffic increases, more instances launch automatically.If traffic drops, unnecessary instances shut down (saving cost).

STEP 4: Monitor, Meter, and Optimize

Uses CloudWatch or Datadog for real-time monitoring.Uses cloud metering to track billing and optimize resources.Sets up alerts (e.g., if CPU > 80%, add more servers).