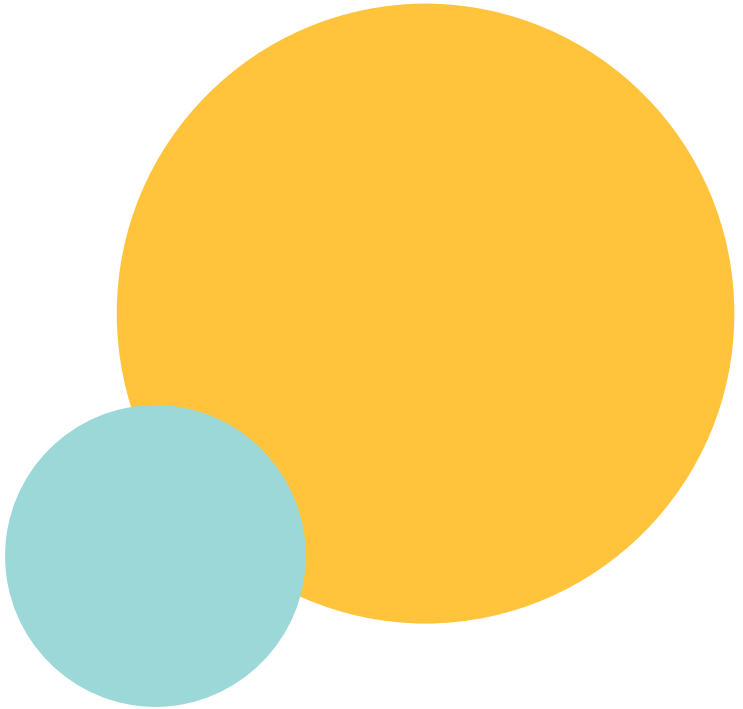
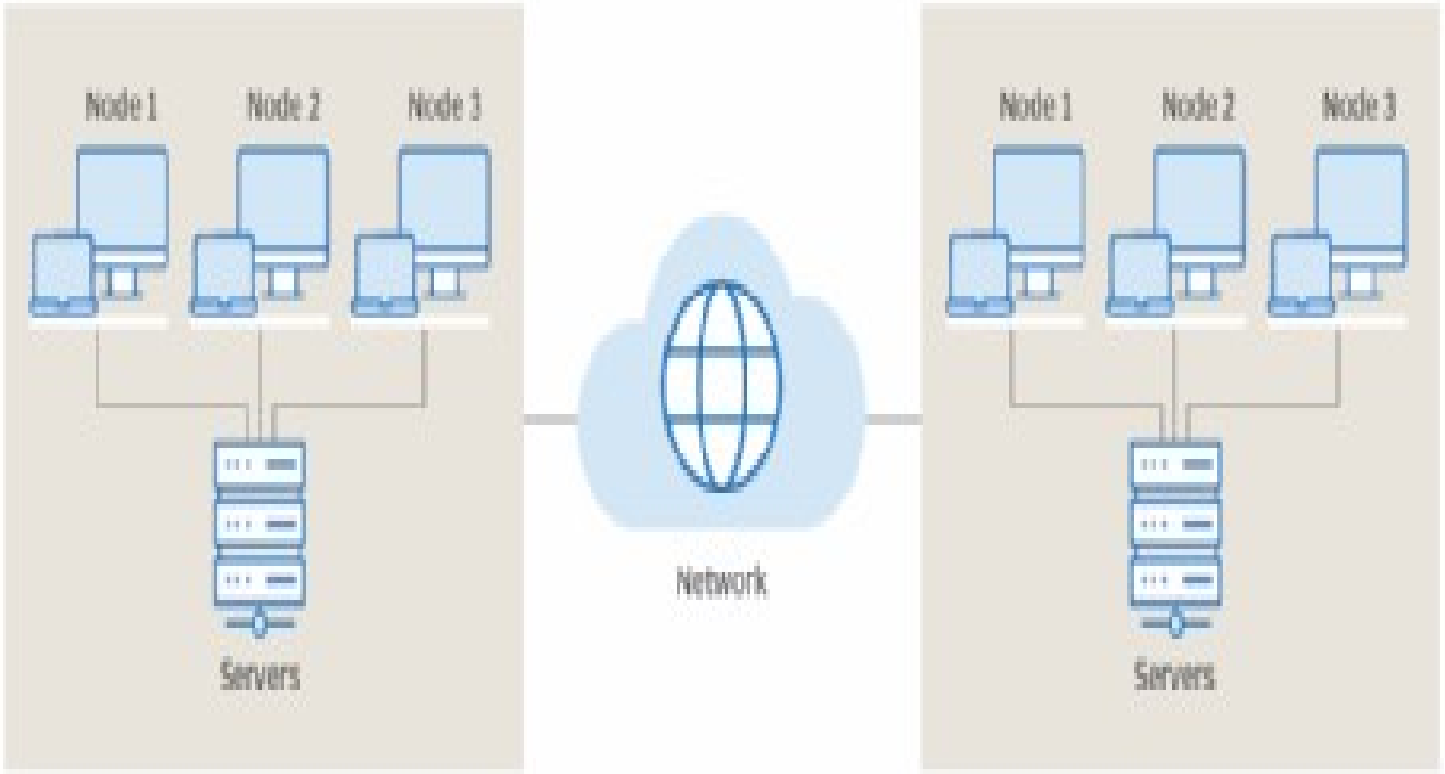
HISTORICAL DEVELOPMENTS



The distributed computing process



A distributed system is a collection of independent computers that appears to its users as a single coherent system.

Cloud computing and Distributed systems

- Clouds are large distributed computing facilities that make available their services to third parties on demand
- It is composed of multiple independent components and that these components are perceived as a single entity by users.
- Clouds hide the complex architecture and provide a single interface to users.
- The primary purpose of distributed systems is to share resources and utilize them better.
- This is true in the case of cloud computing, where this concept is taken to the extreme and resources (infrastructure, runtime environments, and services) are rented to users.

Historical developments

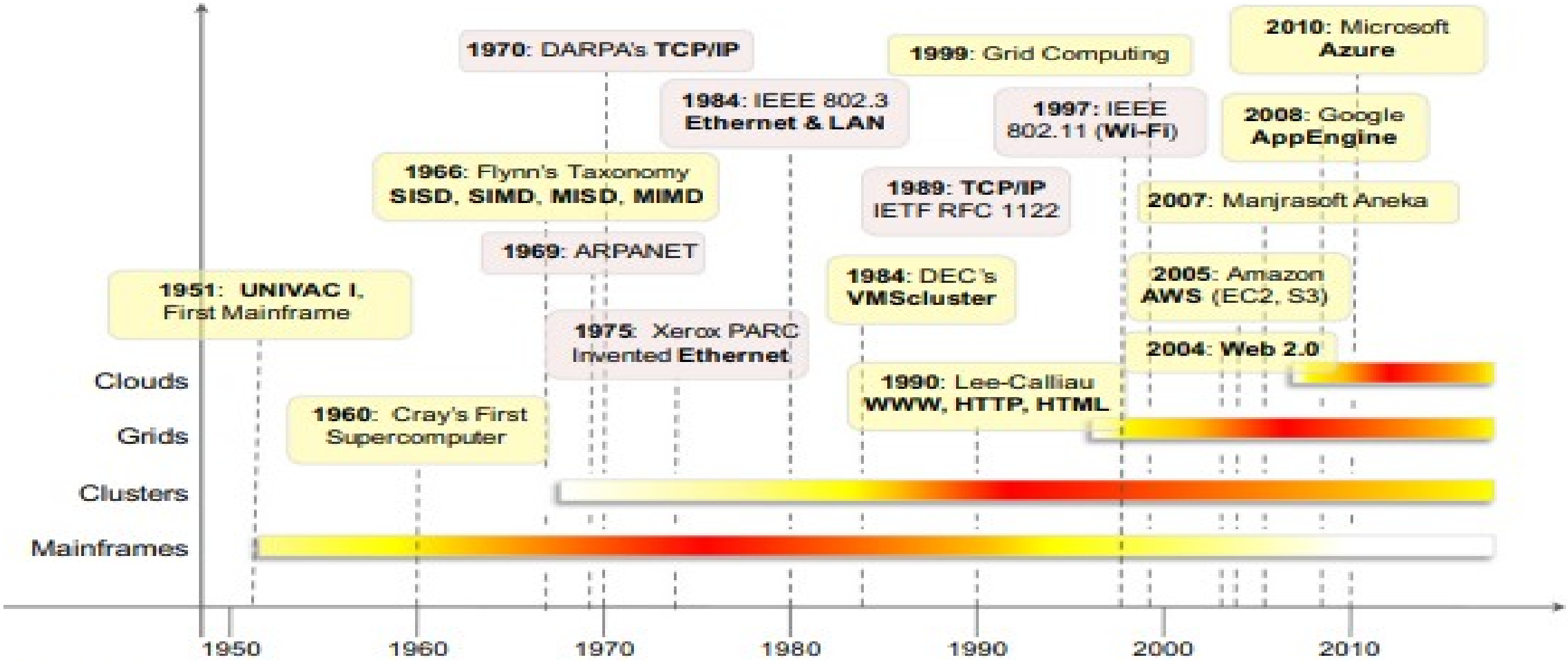


FIGURE 1.6

The evolution of distributed computing technologies, 1950s—2010s.

Characteristics of distributed systems

Share resources and utilize them better

Heterogeneity, openness, scalability, transparency, concurrency, continuous availability and independent failures.

Characterize clouds, especially in the context of scalability, concurrency, and continuous availability.

Major Milestones

- Main Frames
- Cluster Computing
- Grid Computing

Mainframes

- Mainframes were the **first example of large computing facilities** which leverage multiple processing units.
- They are **powerful, highly reliable computers** specialized for large data movement and large I/O operations.
- Mainframes are mostly used by large organizations for bulk data processing such as **online transactions, enterprise resource planning** and other big data operations.
- They are not considered as a distributed system; however they can perform big data processing and operations due to their **high computational power by multiple processors.**

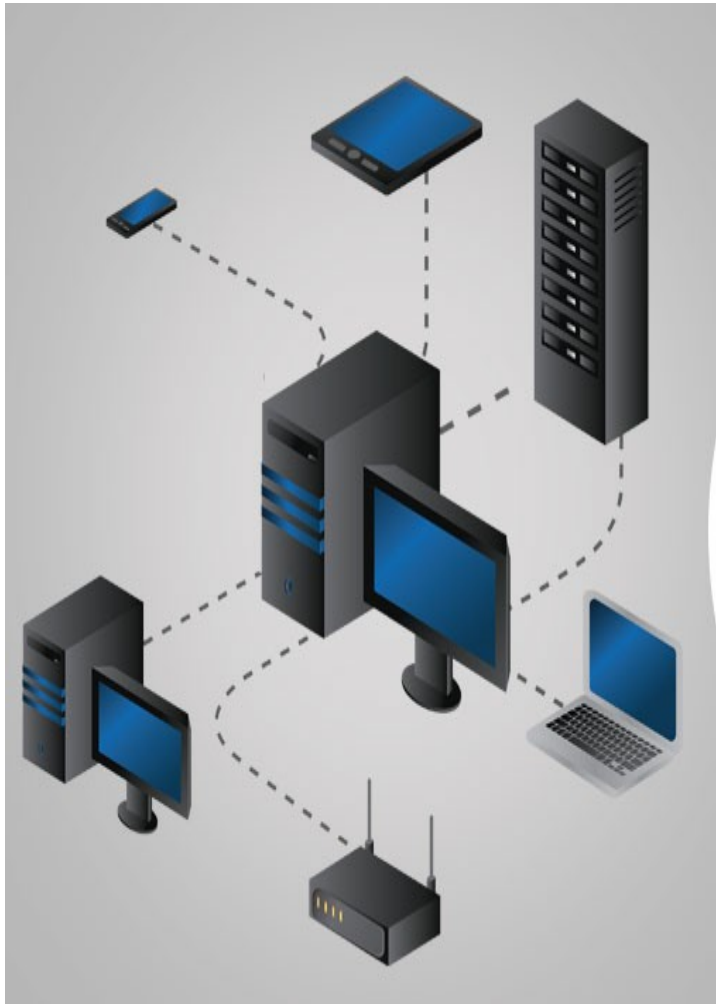
Mainframes

Mainframes

https://youtu.be/uTDqiPfEvyE

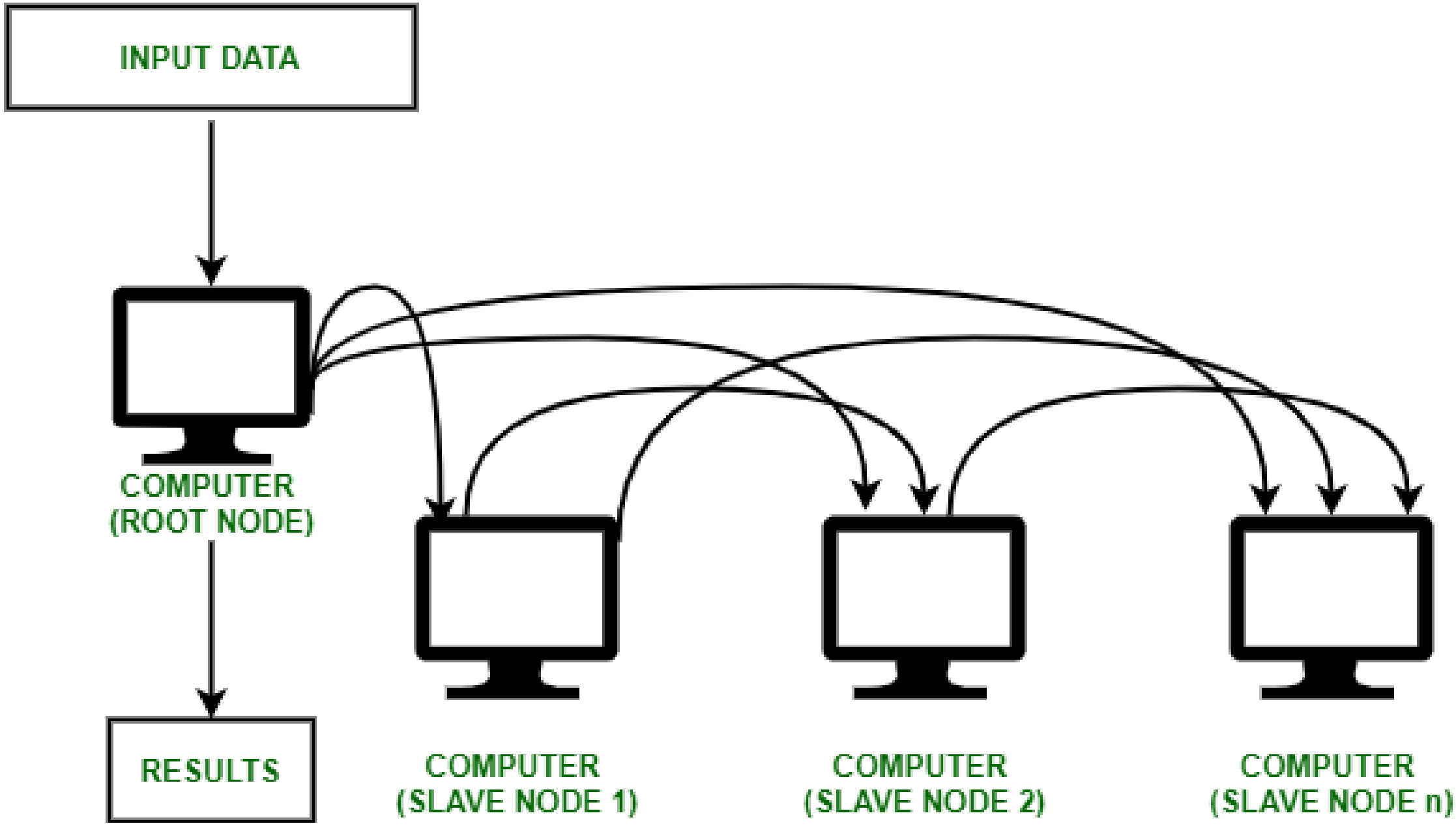
Cluster computing

- Cluster computing defines several computers linked on a network and implemented like an individual entity.
- Each computer that is linked to the network is known as a node.
- Cluster computing provides solutions to solve difficult problems by providing faster computational speed and enhanced data integrity.



Advantages of Cluster Computing

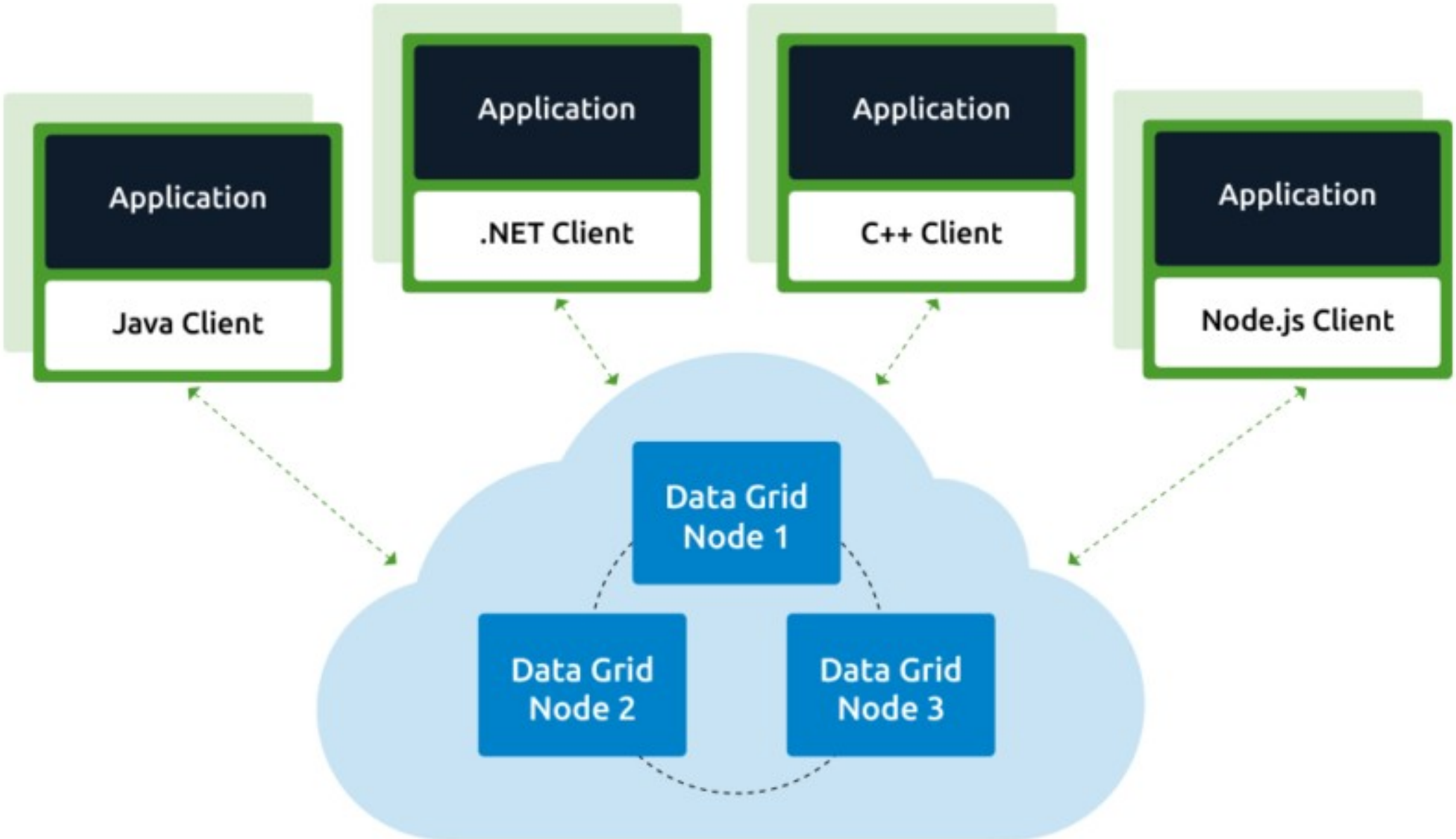
- Cost-Effectiveness
- Processing Speed
- Increased Resource Availability
- Improved Flexibility

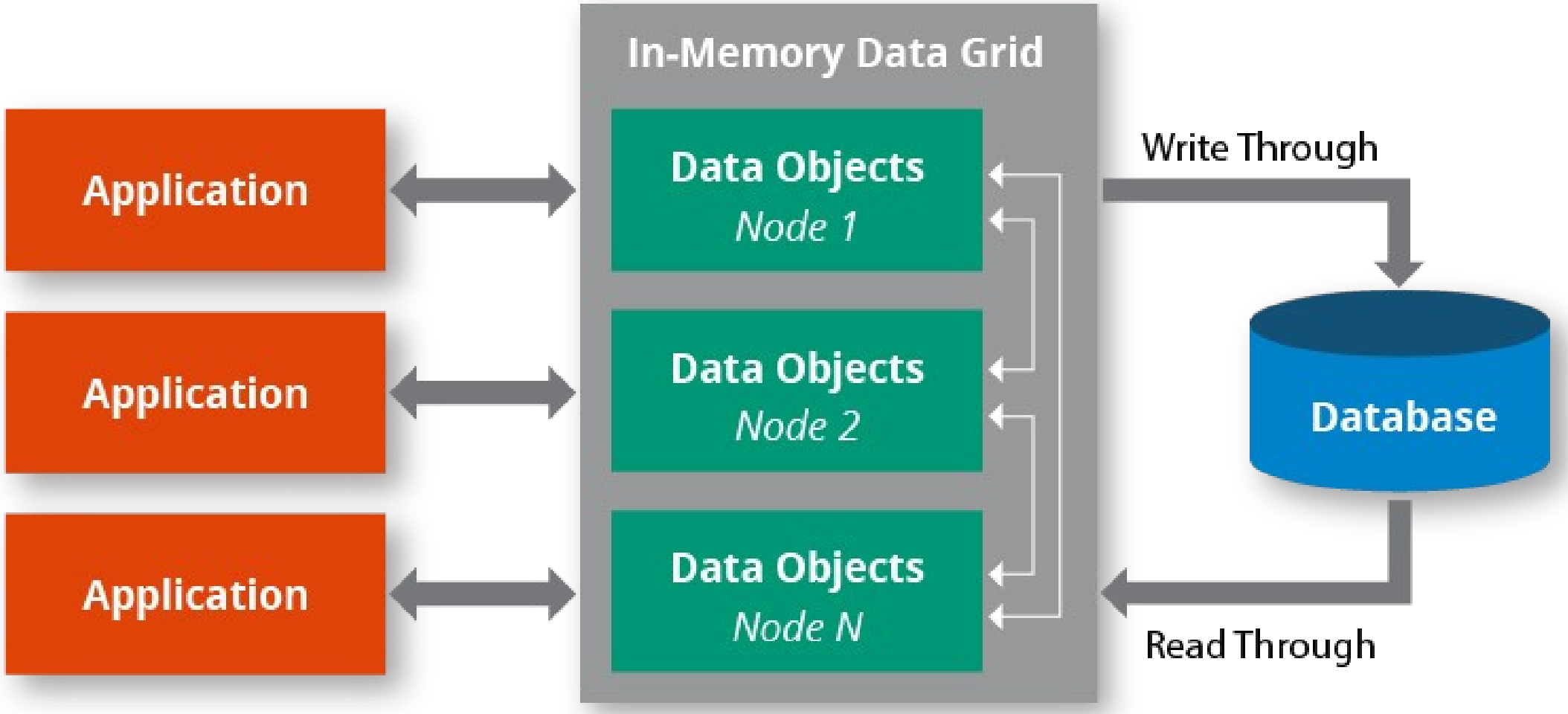

INPUT DATA
COMPUTER
(ROOT NODE)
RESULTS
COMPUTER
(SLAVE NODE 1)
COMPUTER
(SLAVE NODE 2)
COMPUTER
(SLAVE NODE n)

Grid computing

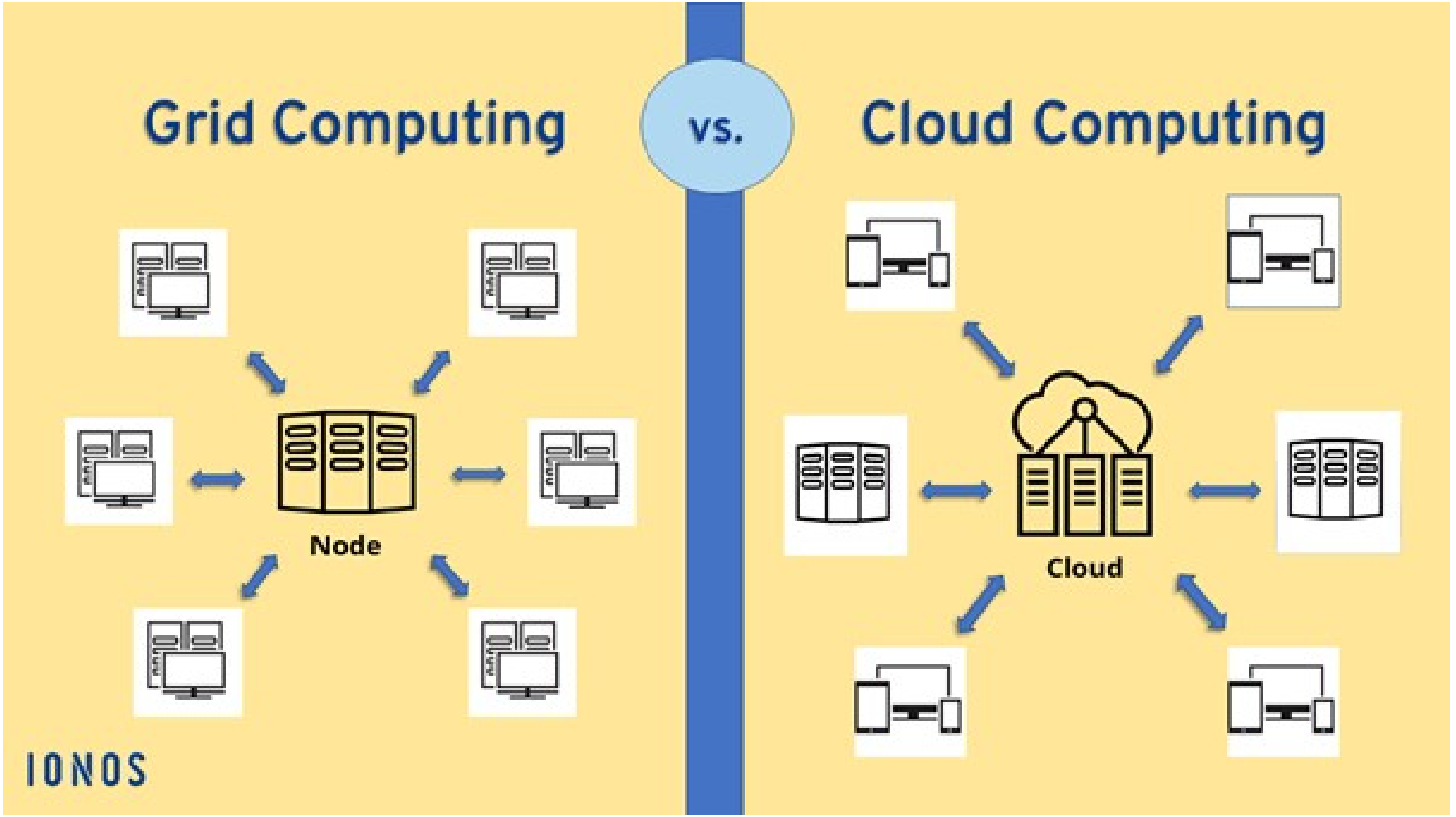
- Grid computing is the practice of leveraging multiple computers, often geographically distributed but connected by networks, to work together to accomplish joint tasks.

- It is typically run on a “data grid,” a set of computers that directly interact with each other to


Application
Java Client
Application
.NET Client
Application
C++ Client
Application
Node.js Client
Data Grid
Node 1
Data Grid
Node 2
Data Grid
Node 3


In-Memory Data Grid
Application
Application
Application
Data Objects
Node 1
Data Objects
Node 2
Data Objects
Node N
Write Through
Database
Read Through

The main difference between grid and cloud computing is that the grid computing refers to a collection of computer resources located at different locations to process a single task while the cloud computing refers to manipulating, configuring and accessing hardware and software resources remotely over the internet.


Grid Computing
vs.
Cloud Computing
Node
Cloud
IONOS

GRID COMPUTING
VERSUS
CLOUD COMPUTING

GRID COMPUTING	CLOUD COMPUTING
Use of widely distributed computer resources to reach a common goal	Technology that enables access to shared pools of configurable system resources and higher-level services over the internet
Computing resources are distributed among different devices located in different locations	Computing resources are managed centrally in data centers belonging to the cloud service providers
Task is divided into several independent subtasks, and each machine on the grid is assigned with a subtask	Provides resources according to the requirements
Users can access the data in the grid computing devices via cooperate networks such as internet or a low-speed network	Users can access the resources through the internet
Management is decentralized	Management is centralized
Uses distributed architecture	Uses client-server architecture

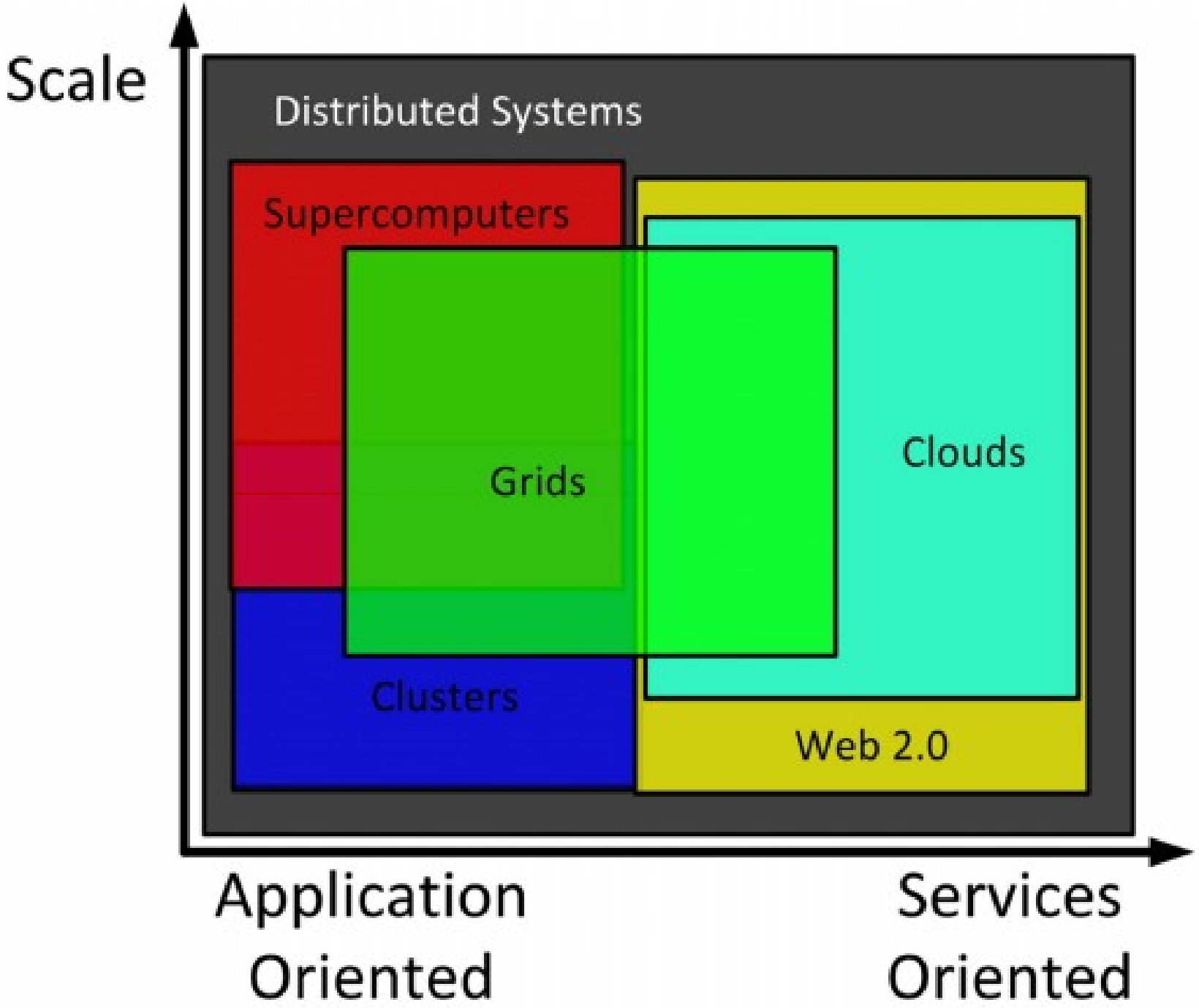

Scale
Distributed Systems
Supercomputers
Grids
Clouds
Clusters
Web 2.0
Application Oriented
Services Oriented

Table 1 Summary of cluster, grid, and cloud computing paradigms

	Cluster Computing	Grid Computing	Cloud Computing
Basic Idea	Aggregation of resources.	Segregation of Resources.	Consolidation of Resources.
Running Processes	Same processes run on all computers over the cluster at the same time.	Job is divided into sub-jobs each is assigned to an idle CPU so they all run concurrently.	Depends on service provisioning. Which computer offers a service and provisions it to the requesting clients.
Operating System	All nodes must run the same operating system.	No restriction is made on the operating system.	No restriction is made on the operating system.
Job Execution	Execution depends on job scheduling. So, jobs wait unit it's assigned a runtime.	Execution is scalable in a way that moves the execution of a job to an idle processor (node).	Self-Managed.
Suitable for Apps	Cascading tasks. If one tasks depends on another one.	Not suitable for cascading tasks.	On-demand service provisioning.
Location of nodes	Physically in the same location	Distributed geographically all over the globe.	Location doesn't matter
Homo/Heterogeneity	Homogenous	Heterogeneous	Heterogeneous
Virtualization	None	None	Virtualization is a key
Transparency	Yes	Yes	Yes
Security	High	High, but doesn't reach the level of cluster computing.	Lower than both types.
Interoperability	Yes	Yes	No

https://www.youtube.com/watch?v=fBURiN1omLo

https://www.youtube.com/watch?v=RWgW-CgdIk0

Historical developments

Five core technologies that played an important role in realization of Cloud Computing

Distributed Systems, Virtualization, Web 2.0, Service oriented computing and utility oriented computing

1. **Distribute systems**

A distributed system is a collection of independent computers that appears to its users as a single coherent system.

Three major milestones have led to cloud computing: mainframe computing, cluster computing, and grid computing.

2. **Virtualization**

Virtualization is essentially a technology that allows creation of different computing environments. These environments are called virtual because they simulate the interface that is expected by a guest

EX: Hardware virtualization, Storage and Network virtualization

Historical developments

3. Web 2.0

- Web is the primary interface through which cloud computing delivers its services
- A set of technologies and services that facilitate interactive information sharing, collaboration, user-centered design, and application composition.
- Web 2.0 brings interactivity and flexibility into Web pages, providing enhanced user experience by gaining Web-based access to all the functions that are normally found in desktop applications.
- These capabilities are obtained by integrating a collection of standards and technologies such as XML, Asynchronous JavaScript and XML (AJAX), Web Services, and others

  Ex: Google Documents, Google Maps, Flickr, Facebook, Twitter, YouTube Blogger, and Wikipedia.

Historical developments

4. Service-oriented computing

- Service orientation is the core reference model for cloud computing systems.
- This approach adopts the concept of services as the main building blocks of application and system development.
- Service-oriented computing (SOC) supports the development of rapid, low-cost, flexible, interoperable, and evolvable applications and systems.
- A service is an abstraction representing a self-describing and platform-agnostic component that can perform any function—anything from a simple function to a complex business process
- A service is supposed to be loosely coupled, reusable, programming language independent, and location transparent.

Historical developments

5. Utility-oriented computing

- Utility computing is a vision of computing that defines a service-provisioning model for compute services in which resources such as storage, compute power, applications, and infrastructure are packaged and offered on a pay-per-use basis

- With the advent of cloud computing, the idea of providing computing as a utility like natural gas, water, power, and telephone connection has become a reality today

Building cloud computing environments

The creation of cloud computing environments encompasses

1. **Application development**

   Web applications, enterprise applications resource-intensive applications.

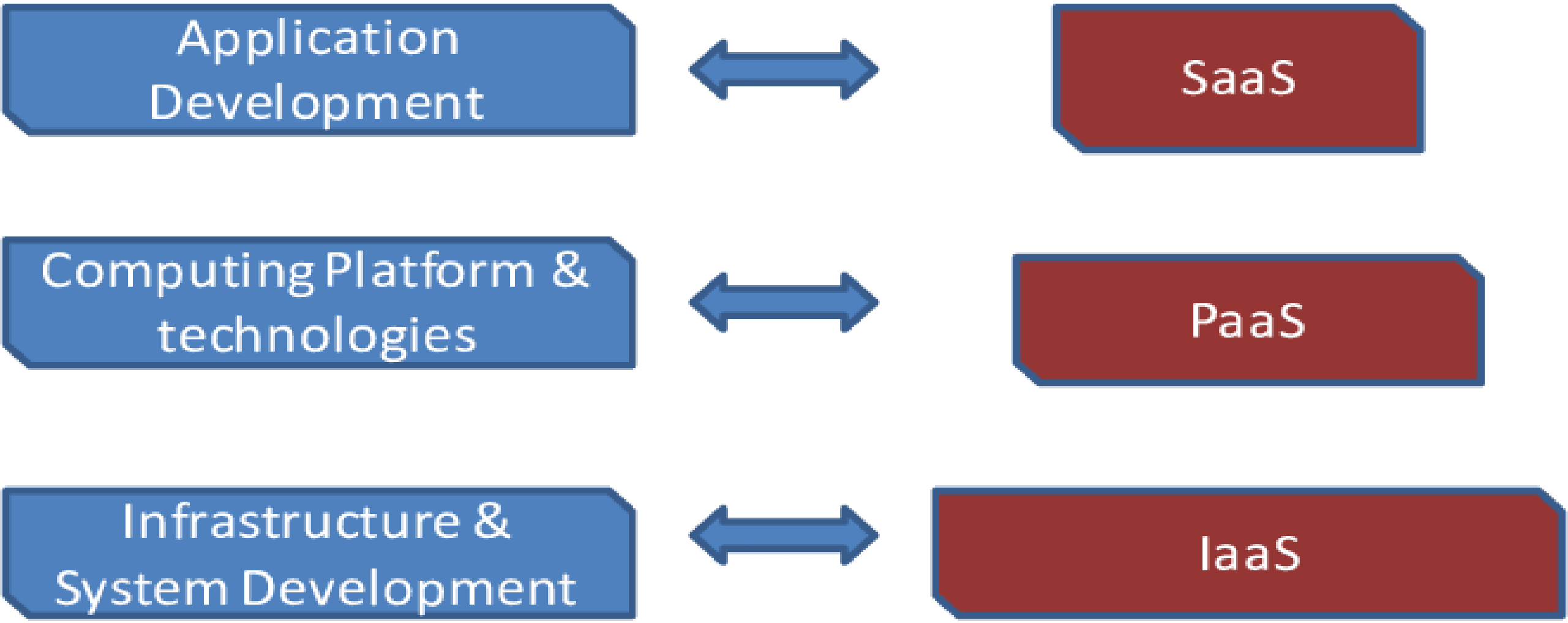
   Data-intensive

   compute-intensive applications.

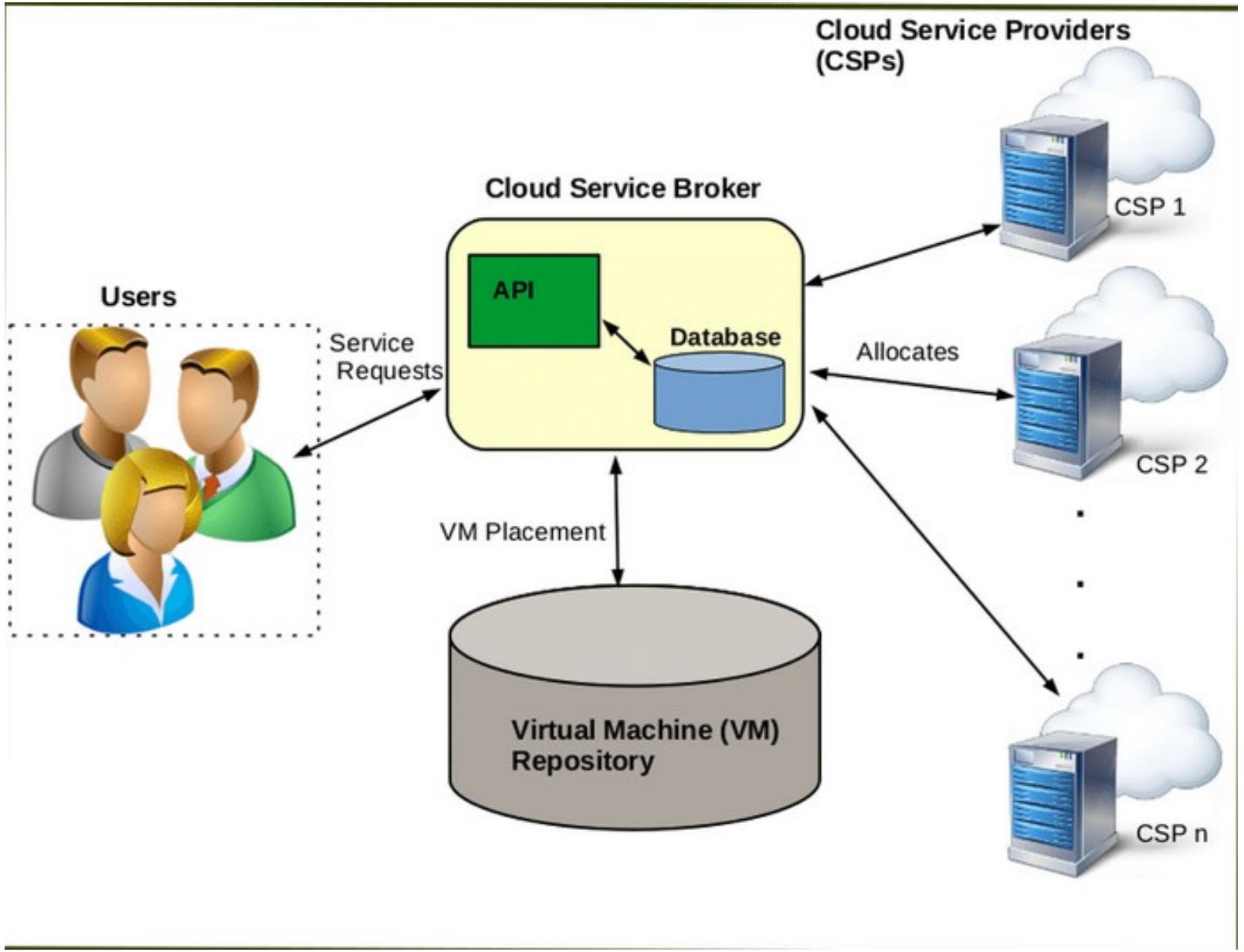
2. **Infrastructure and system development**

   Distributed computing, virtualization, service orientation, and Web 2.0 form the core technologies enabling the provisioning of cloud services from anywhere on the globe.

Building cloud computing environment



Building cloud computing environments



Computing platforms and technologies

1. **Amazon web services (AWS)**

   AWS offers

   comprehensive cloud IaaS services ranging from virtual compute, storage, and networking to complete computing stacks.

   EX: Elastic Compute Cloud (EC2)

   Simple Storage Service (S3)

2. **Google AppEngine**
   - Google AppEngine is a scalable runtime environment mostly devoted to executing Web applications.
   - The languages currently supported are Python, Java, and Go.

Computing platforms and technologies

3. Microsoft Azure

- It provides a scalable runtime environment for Web applications and distributed applications
- Applications in Azure are organized around the concept of roles, which identify a distribution unit for applications and embody the application's logic. Currently, there are three types of role: Web role, worker role, and virtual machine role.
- The Web role is designed to host a Web application,
- the worker role used to perform workload processing,
- the virtual machine role provides a virtual environment in which the computing stack can be fully customized, including the operating systems.

Computing platforms and technologies

4. Hadoop

- Apache Hadoop is an open-source framework that is suited for processing large data sets on commodity hardware
- Hadoop is an implementation of MapReduce,
- MapReduce is programming model which provides two fundamental operations for data processing
  - map
  - reduce.
- Map transforms and synthesizes the input data provided by the user;
- Reduce aggregates the output obtained by the map operations.

Computing platforms and technologies

5. Force.com and Salesforce.com

- Force.com is a cloud computing platform for developing social enterprise applications.
- SalesForce.com is a Software-as-a-Service solution for customer relationship management.

6. Manjrasoft Aneka

- Manjrasoft Aneka is a cloud application platform for rapid creation of scalable applications and their deployment on various types of clouds in a seamless and elastic manner.
- It supports a collection of programming abstractions
- Developers can choose different abstractions to design their application