Cloud Computing

Definitions (I)

We have redefined Cloud Computing to include everything that we already do. I do not understand what we would do differently other then change the working of some of our ads.

Larry Ellison
Oracle CEO

It's stupidity. It's worse than stupidity: it's a marketing hype campaign. Somebody is saying this is inevitable – and whenever you hear somebody saying that, it's very likely to be a set of businesses campaigning to make it true.

Richard Stallman
GNU & FSF father

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet).

Wikipedia

Definitions (III)

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

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

http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Definitions (III)

Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services.
The services themselves have long been referred to as Software as a Service (SaaS), so we use that term.
The datacenter hardware and software is what we will call a
Cloud.

http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf

Properties and Characteristics

Scalability
Elasticity

- Dynamic provision
- Multi-tenant design

Manageability
Interoperability

- Autonomic Control
- System monitoring
- Billing System

Availability
Reliability

- Fault tolerance
- System resilience
- Security

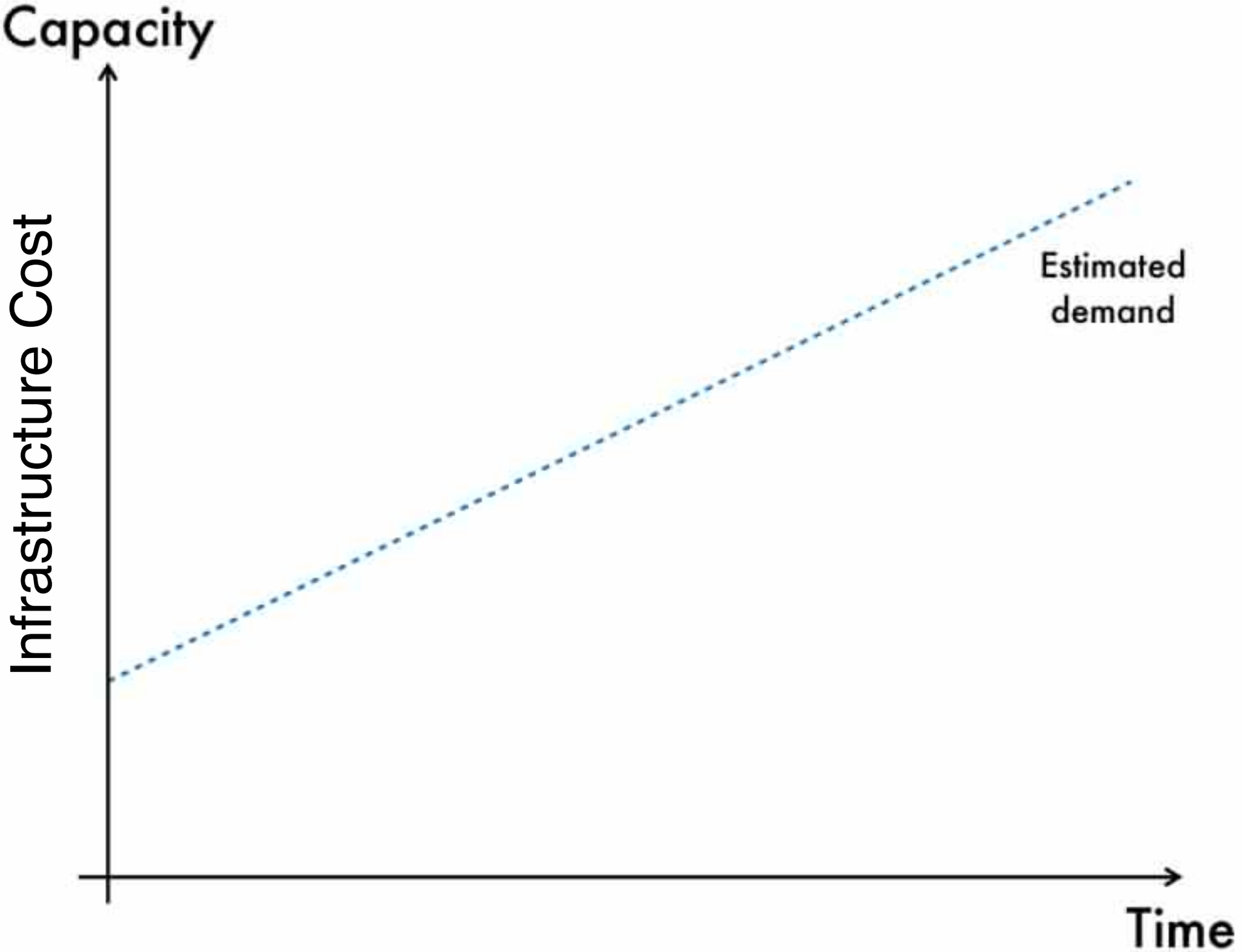
Accessibility
Portability

- Uniform access
- Thin client

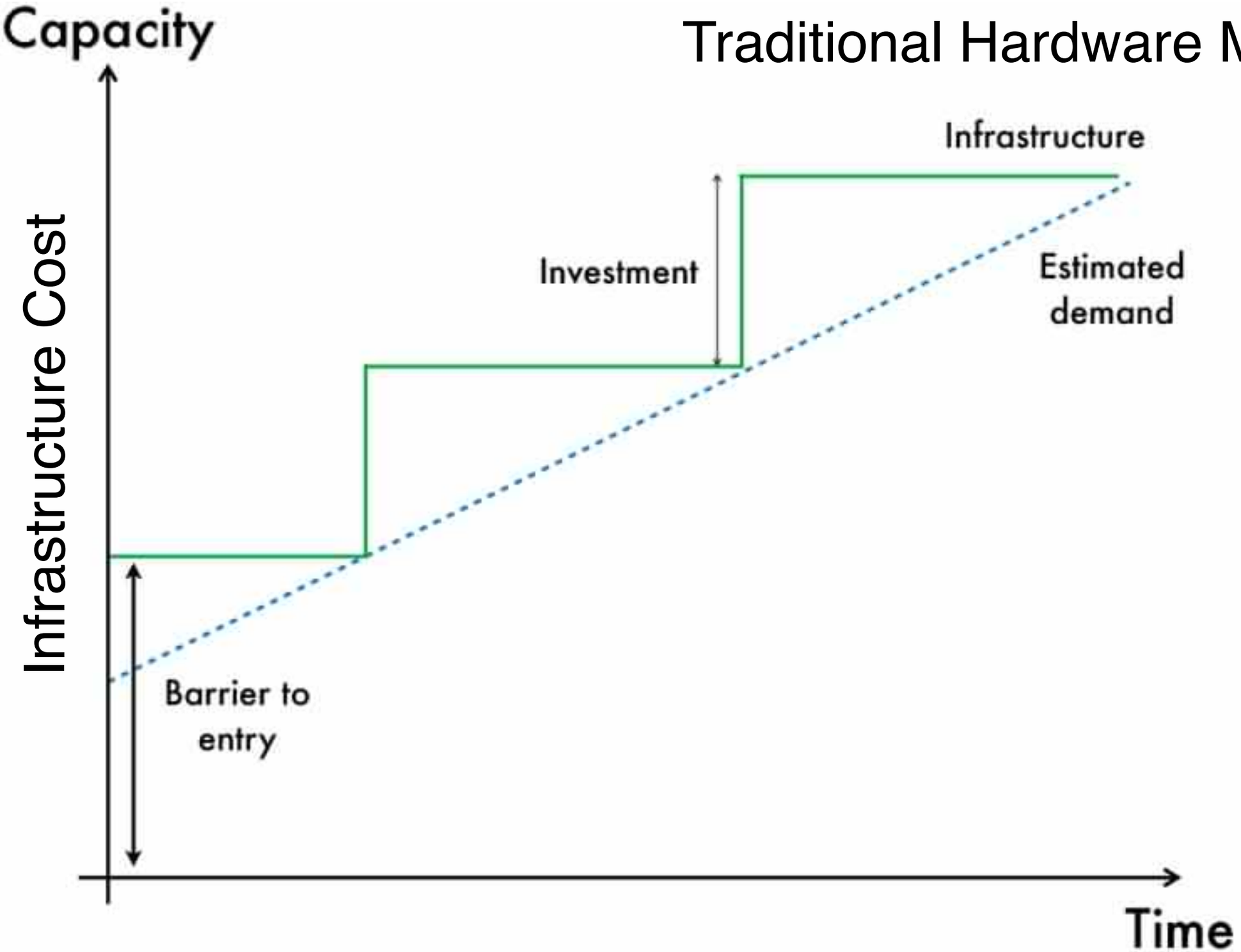
Performance
Optimization

- Parallel processing
- Load balancing
- Job scheduling

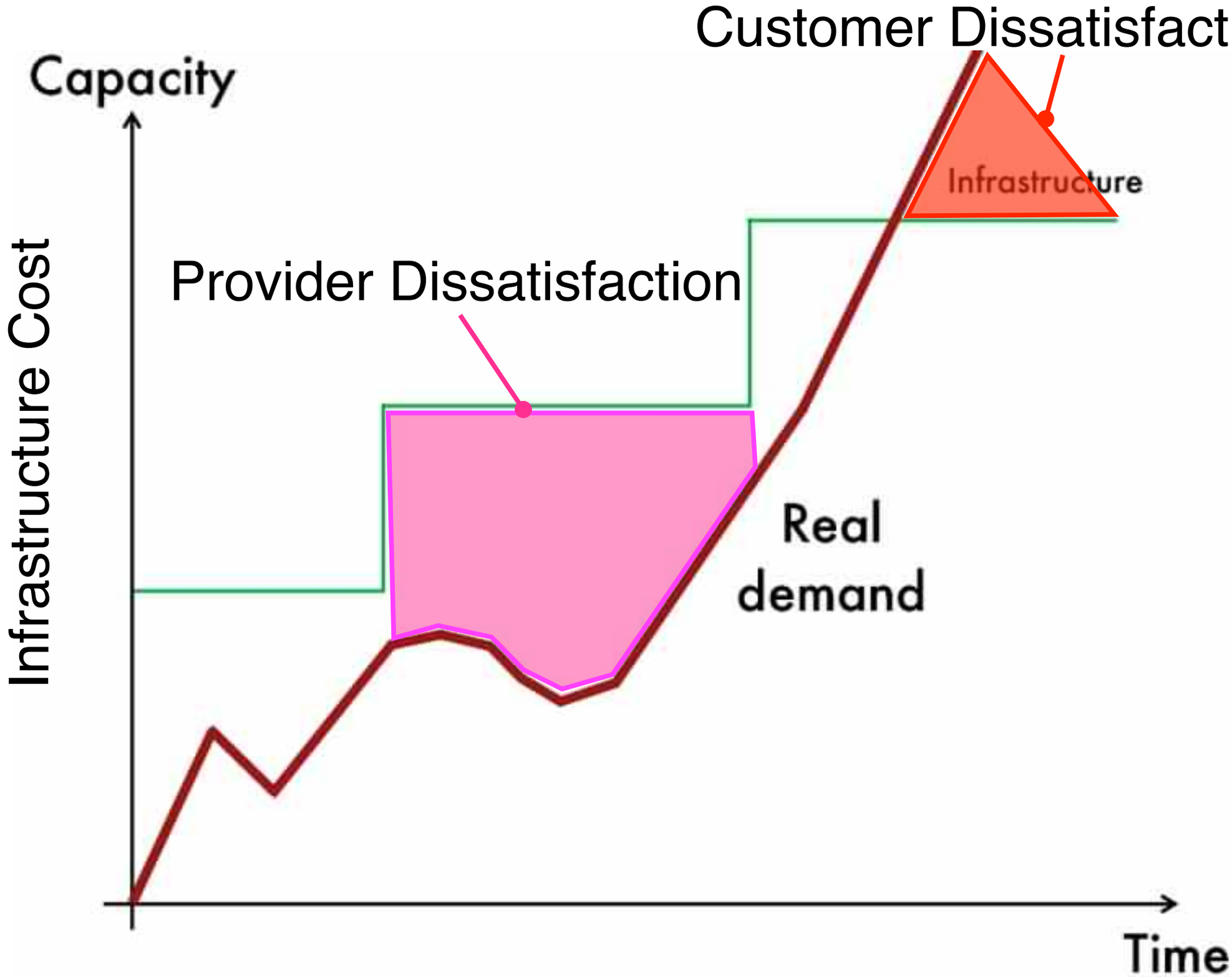
Elasticity (I)



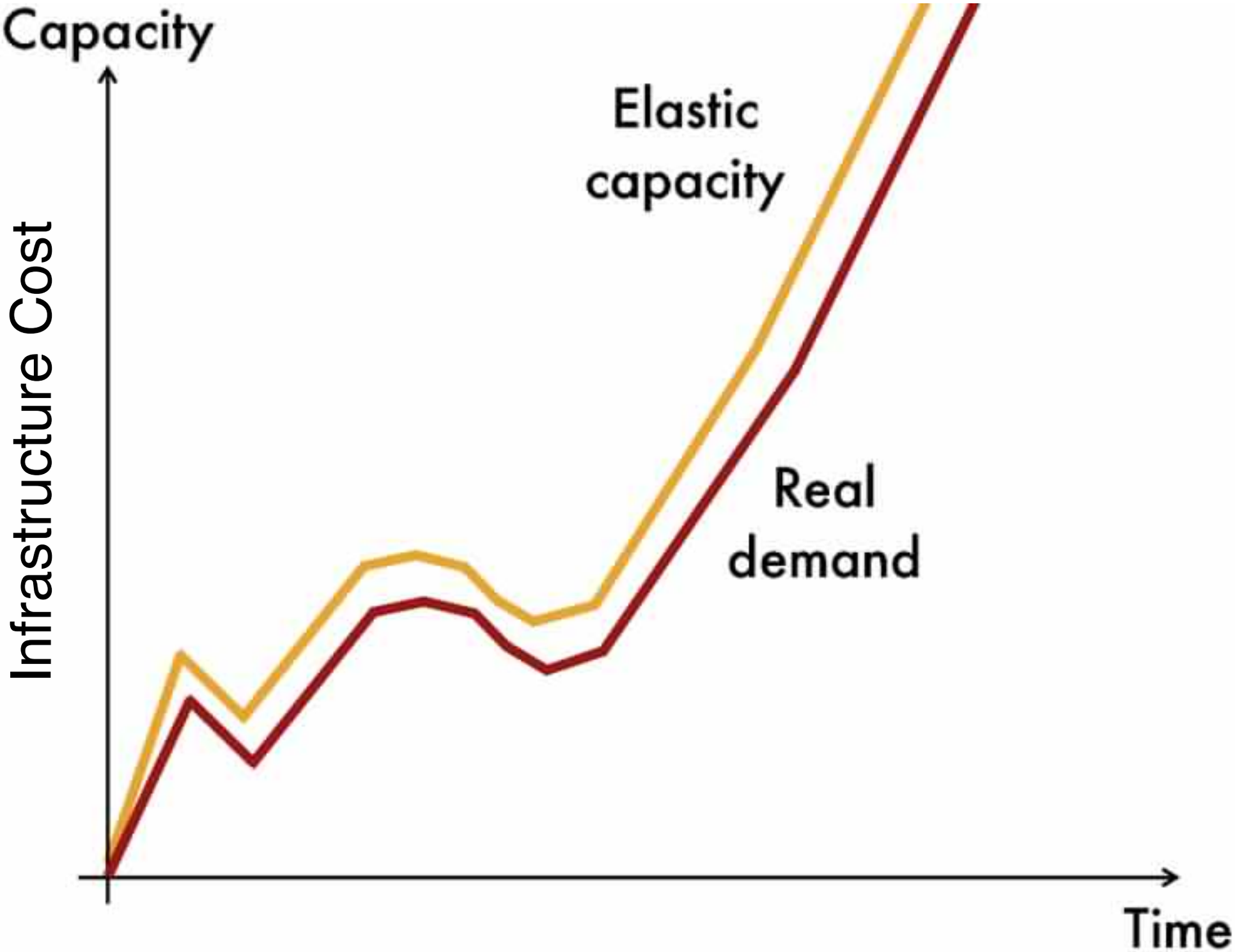
Elasticity (II)



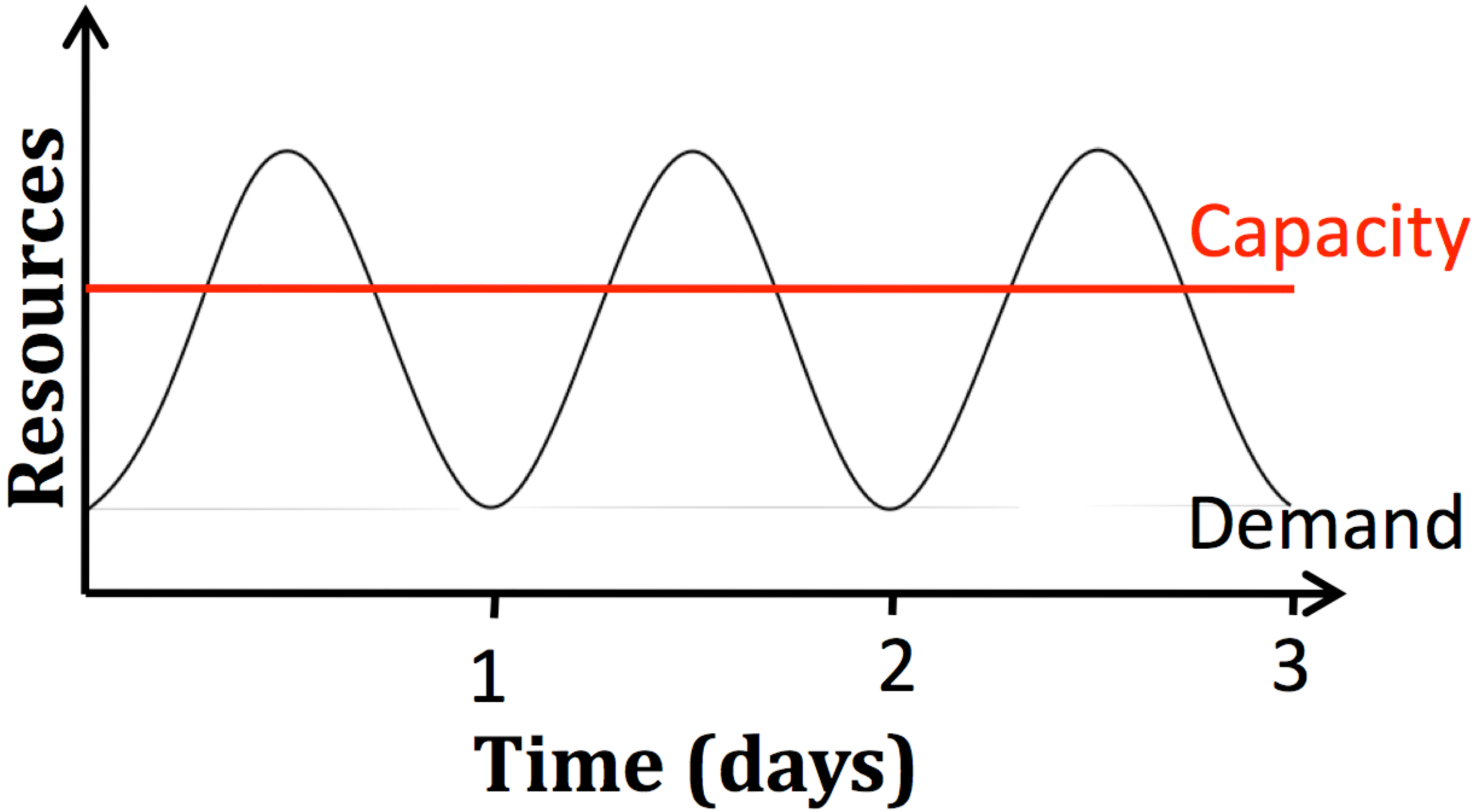
Elasticity (III)



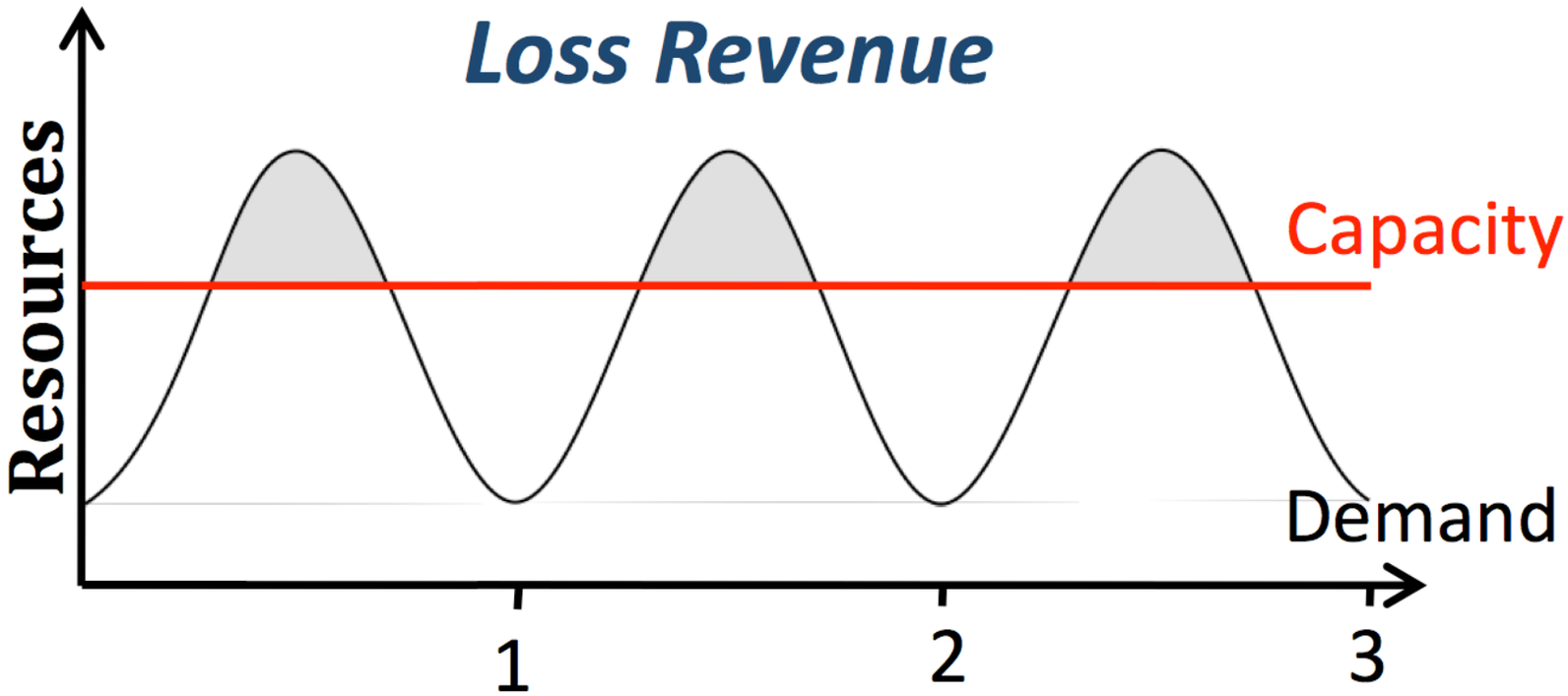
Elasticity (IV)

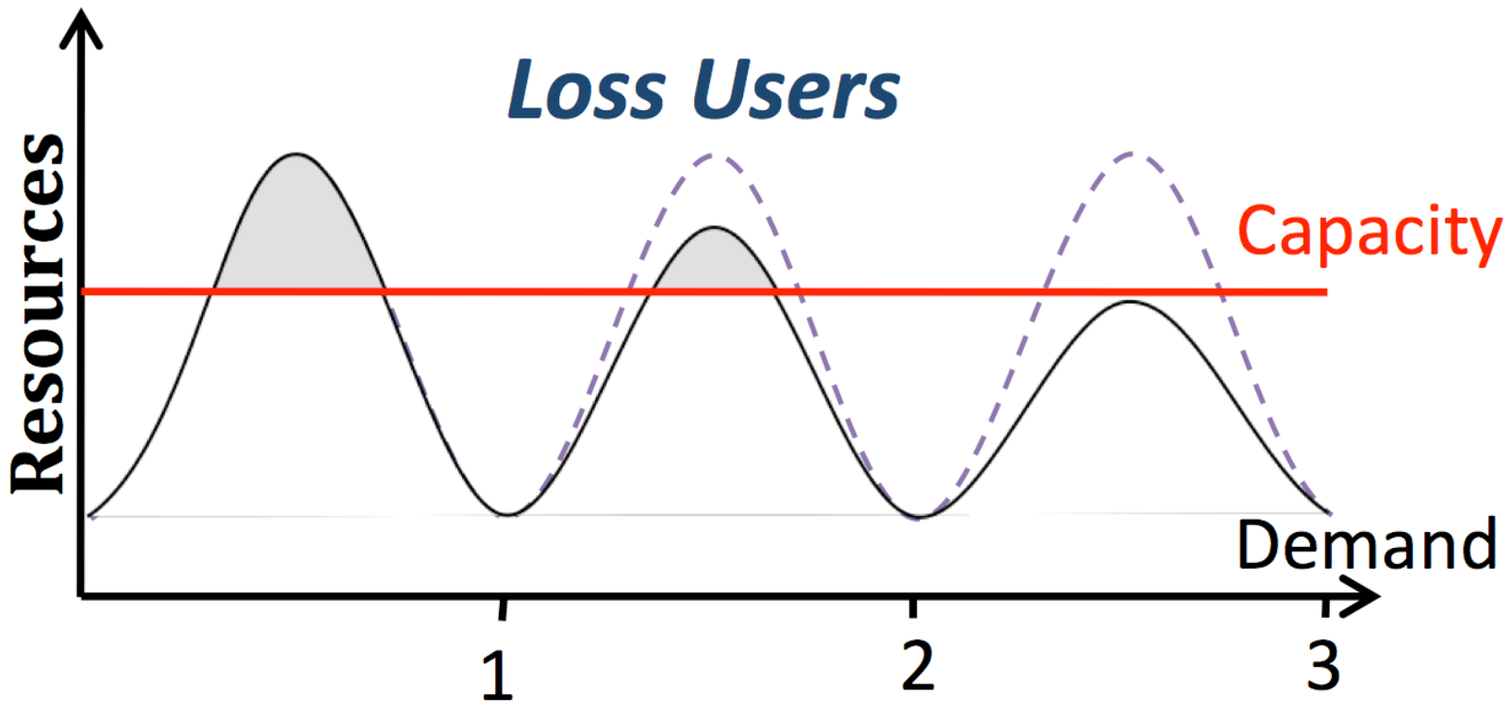


Traditional Provisioning

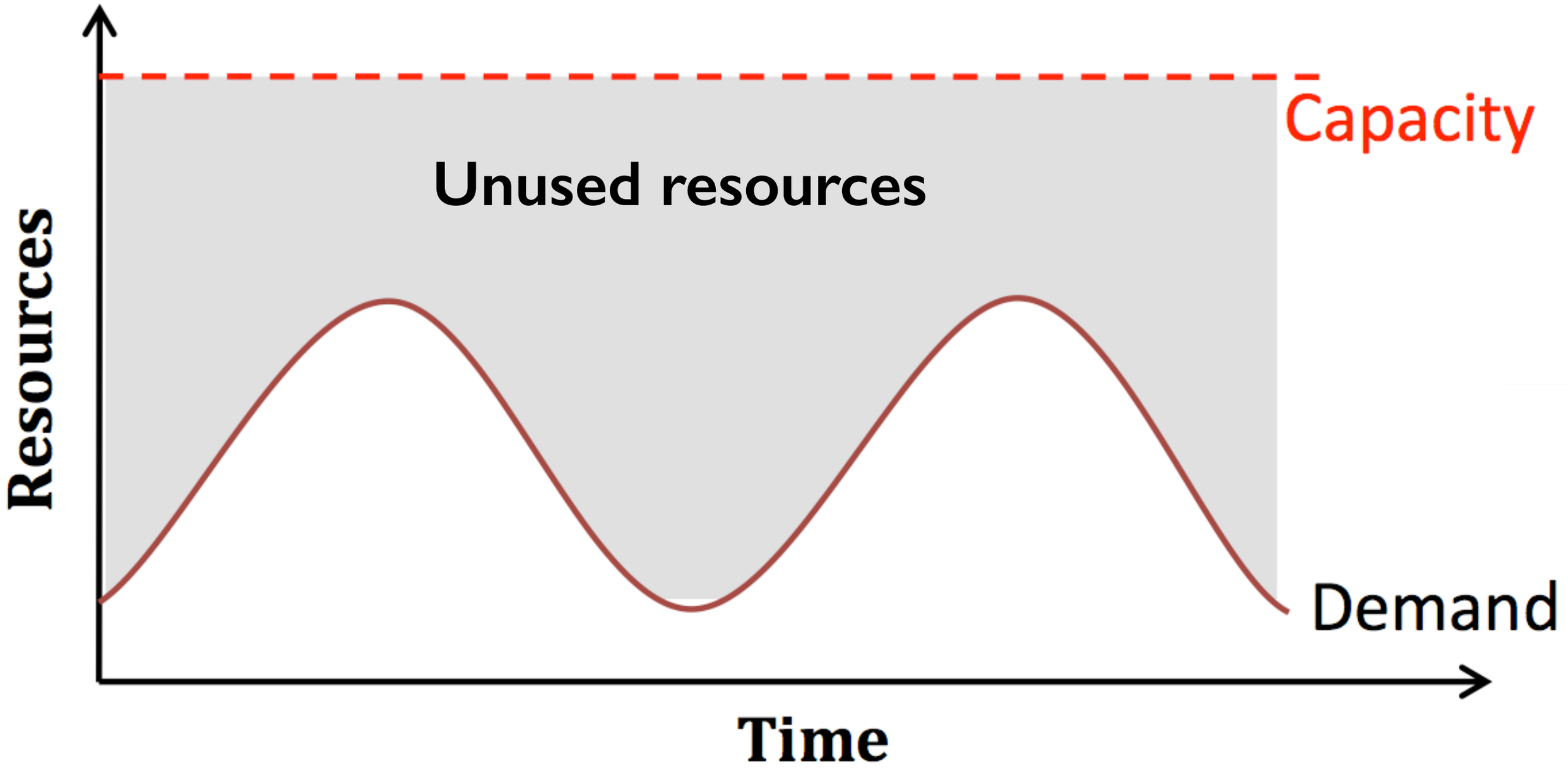


Under-provisioning



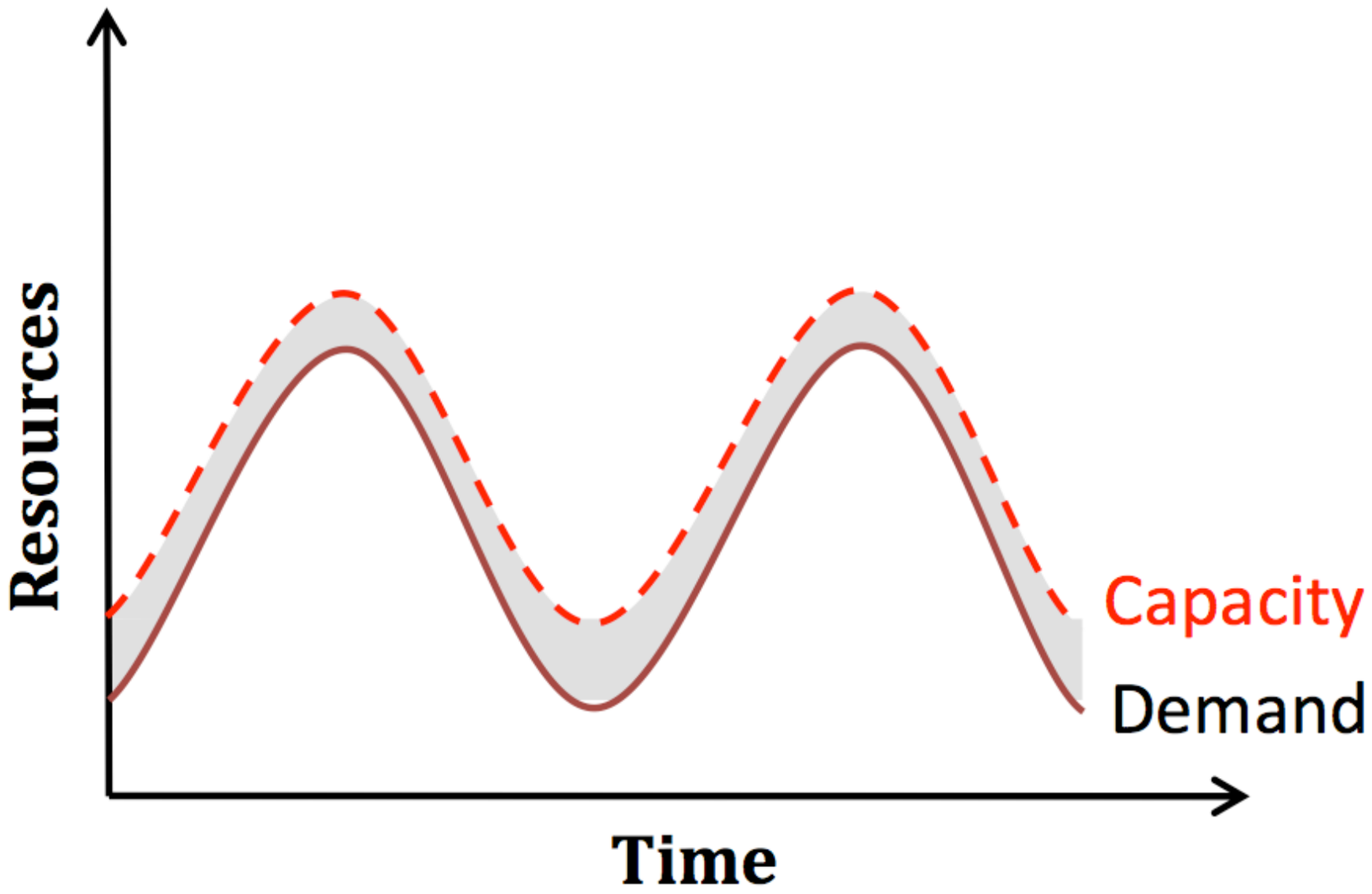


Over-provisioning



Dynamic Provisioning

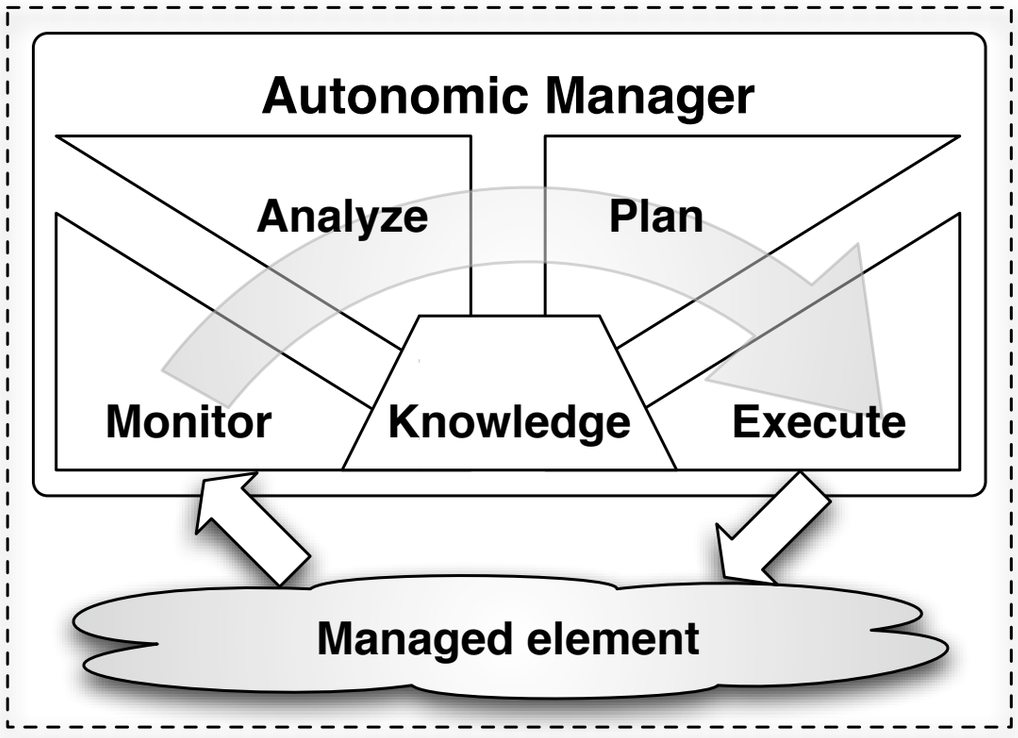
- Cloud resources should be provisioned dynamically
  - Meet seasonal demand variations
  - Meet demand variations between different industries
  - Meet burst demand for some extraordinary events

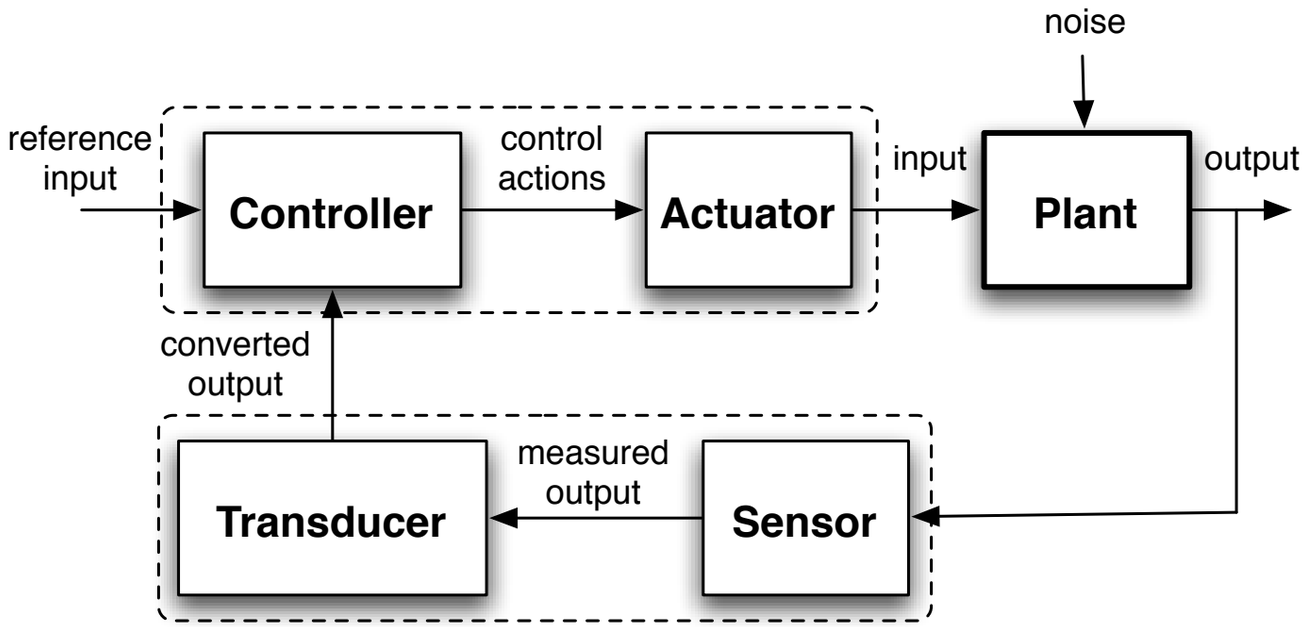


Manageability and Interoperability

- Manageability
  - Enterprise-wide administration of cloud computing systems.
  - Systems manageability is strongly influenced by network management initiatives in telecommunications.
- Interoperability
  - Interoperability is a property of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, without any restricted access or implementation.
- How to achieve?
  - Autonomic Control
  - System Monitoring
  - Billing System

Autonomic Control





- Autonomic Computing
  - Its ultimate aim is to develop computer systems capable of self-management, to overcome the rapidly growing complexity of computing systems management, and to reduce the barrier that complexity poses to further growth.

Self-* Properties

- Autonomic System Functional Areas (CHOP)
  - Self-Configuration
    - Automatic configuration of components.
  - Self-Healing
    - Automatic discovery, and correction of faults.
  - Self-Optimization
    - Automatic monitoring and control of resources to ensure the optimal functioning with respect to the defined requirements.
  - Self-Protection
    - Proactive identification and protection from arbitrary attacks.

System Monitoring

- A System Monitor in systems engineering is a process within a distributed system for collecting and storing state data.

- What should be monitored in the Cloud ?
  - Physical and virtual hardware state
  - Resource performance metrics
  - Network access patterns
  - System logs
  - etc...

Billing System

- Billing System in Cloud
  - Users pay as many as they used.
  - Cloud provider must first determine the list of service usage price.
  - Cloud provider have to record the resource or service usage of each user, and then charge users by these records.

- How can cloud provider know users' usage ?
  - Get those information by means of monitoring system.
  - Automatically calculate the total amount of money which user should pay.
  - Automatically request money from use's banking account.

Cloud Benefits

Markets & Enterprises

Users

Benefits from Clouds

- For the market and enterprises
  - Reduce initial investment
  - Reduce capital expenditure
  - Improve industrial specialization
  - Improve resource utilization

- For the end user and individuals
  - Reduce local computing power
  - Reduce local storage power
  - Variety of thin client devices in daily l

Initial Investment

- Traditional process of enterprises to initiate business:
  - Survey and analysis the industry and market
  - Estimate the quantity of supply and demand
  - Purchase and deploy IT infrastructure
  - Install and test the software system
  - Design and develop enterprise specific business ser
  - Announce the business service to clients

- Some drawbacks :
  - The survey, analysis and estimation may not 100% correct
  - Infrastructure deployment is time consuming
  - Enterprises should take the risk of wrong investment

Initial Investment

- Initiate business with Cloud Computing services:
  - Survey and analysis the industry and market
  - Choose one cloud provider for enterprise deployme
  - Design and develop business service upon cloud environment
  - Announce the business service to clients

- Some benefits:
  - Enterprise do not need to own the infrastructure
  - Enterprise can develop and deploy business service in short time
  - Enterprise can reduce the business loss of wrong investment

Reduce Capital Expenditure

- Traditional capital expenditure of enterprises:
  - Each enterprise should establish its own IT departm
  - IT department should handle the listing jobs
    - Manage and administrate hardware and software
    - Apply regular data backup and check point process
    - Purchase new infrastructure and eliminate outdated one
    - Always standby for any unexpected IT problems

- Some drawbacks :
  - Enterprise pays for IT investment which is not its business focus
  - Enterprise should take the risk of hardware/software malfunction

Reduce Capital Expenditure

- Capital expenditure with Cloud Computing service:
  - Enterprise can almost dismiss its IT department
  - The jobs of IT department can be achieved by cloud provider
    - Dynamically update and upgrade hardware or software
    - Dynamically provision and deploy infrastructure for enterprise
    - Automatically backup data and check consistency
    - Self-recover from disaster or system malfunction

- Some benefits :
  - Enterprise can shift effort to its business focus
  - Enterprise can reconfigure its IT services in short time
  - Enterprise pays to cloud provider as many as the service

Improve Industrial Specialization

- Traditional industry and market:
  - Each enterprise has to own its IT department
  - IT resources are managed directly by the enterprise
  - IT complexity should be addressed and managed with care by the enterprise itself

- Some drawbacks :
  - IT department is not the business focus of enterprises
  - Most enterprise do not maintain correctly their IT resources
  - Enterprises must optimize their IT resources usage

Improve Industrial Specialization

- Outsourcing to Cloud providers
  - Cloud providers centrally maintain IT infrastructure for enterprises
  - Cloud providers are business-focused on providing resources
  - Cloud providers employ experts for management and administration
  - Enterprises only rent and pay the services the

- Some benefits :
  - IT service performance is optimized by experts
  - Enterprises can focus on their business

Improve Resource Utilization

- Traditional resource utilization
  - Enterprises rarely take care about IT resource utilization
  - IT resources are not well managed by the enterpris
  - IT resources usually for peak demand

- Some drawbacks :
  - Power and spaces utilization wasted

Improve Resource Utilization

- Outsourcing to Cloud providers
  - Cloud providers centrally maintain IT infrastructure for enterprises
  - Cloud providers build performance optimized hardw
  - Cloud providers build consolidated cooling systems
  - Cloud providers take care of legal policy issues

- Some benefits :
  - IT resources can be shared among enterprise
  - IT infrastructure performance can be optimized
  - Large-scale integrated optimization can be applied

Reduce Local Computing Power

- Traditional local computing power requirements:
  - Need to buy your own personal computer
  - Buy powerful processor if you need intensive computing
  - Buy a lot of memory to meet application requirements
  - Install plenty of applications
  - Manage (security) upgrades

- Some drawbacks :
  - Hard to replicate the same system environment
  - Need to regularly update or upgrade software and hardware
  - Need to reinstall application if you reinstall the OS

Reduce Local Computing Power

- Using Cloud Computing services:
  - Can utilize remote computing power in the Cloud
  - Need a basic computer to connect to Internet
  - Applications automatically managed

- Some benefits:
  - Access personal computer anywhere through network
  - Dynamically request more resources on demand
  - Application must not be manually upgraded/managed/ reinstalled

Reduce Local Storage Power

- Traditional local storage power requirement:
  - User code and data files stored in local devices
  - Manual backup regularly preventing hardware failure
  - Physical / power / heating requirements

- Some drawbacks :
  - Storage space may not be enough
  - Storage space may be too much
  - Data consistency between computers is hard
  - Need to sacrifice storage for backups

Reduce Local Storage Power

- Using Cloud Computing services:
  - User code and data files stored in the Cloud
  - Cloud provider guarantees the data availability

- Some benefits:
  - Dynamic allocation of storage on demand
  - Seamless access to data through the network
  - No need to care about data consistency
  - No need to care about data losses and backups
  - No need to care about space/power

Variety of End Devices

- Traditional computing resources:
  - Connection to the Internet through personal computers
  - Only PCs can deliver reasonable computing power
  - Small devices have hardware and power limitations

- Some drawbacks :
  - Computing power is not portable
  - Small devices can only perform small works

Variety of End Devices

- Devices integrated with Cloud Computing:
  - Devices connect to the Internet through wireless networks
  - Devices access Cloud services through Web interfa
  - Devices outsource computing jobs to the cloud

- Some benefits:
  - Users can easily access Cloud services through small devices
  - User can access almost unlimited computing power
  - Small devices can be managed through Clouds (install/ upgrade apps)

Service Models for Housing

- You look for a place to live in a new city

- What do you do?

- Build a new house?
  - You can fully control everything you like to have in your house
  - It is a hard work...

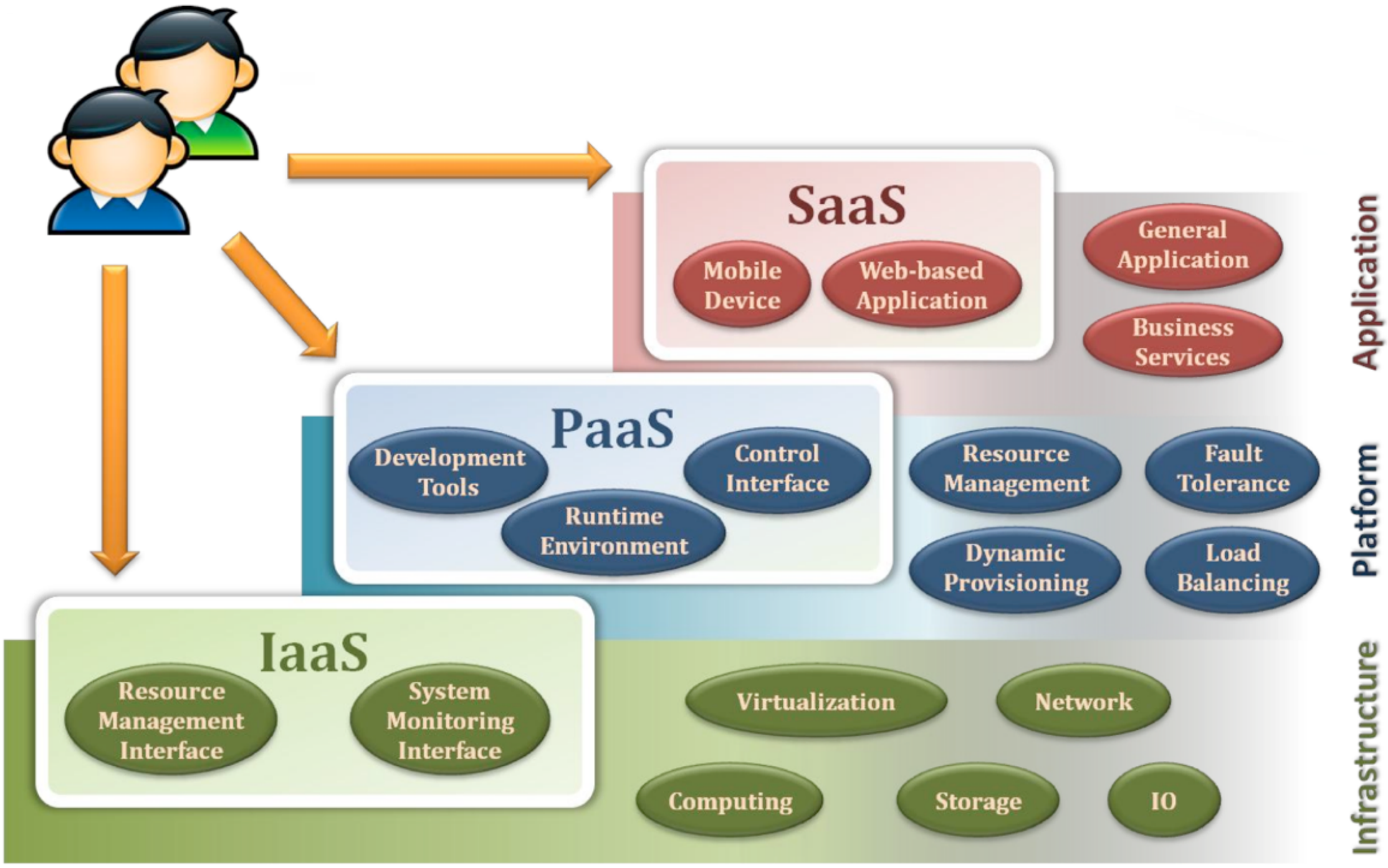
- Buy an empty house?
  - You can fully control only some parts of your house
  - Can not change original infrastructure

- Live in a hotel?
  - Great if you only want to enjoy your life
  - House just for living

Service Models for Clouds

- You want an IT department!
- You can rent some infrastructure and build up your IT system with these resources, under your full control
  - Build a new house
  - Technically, use **Infrastructure as a Service** (IaaS)
- You develop your IT applications through one Cloud platform, not caring about low level resource management
  - Buy an empty house
  - Technically, use **Platform as a Service** (PaaS)
- You directly use some existing IT solutions, provided by the Cloud and ignoring any further deteal
  - Live in a hotel
  - Technically, use **Software as a Service** (SaaS)

Service Models Summary



Infrastructure as a Service (I)

- The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

- The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components .

- Examples: Amazon EC2, Eucalyptus, OpenNebula

Infrastructure as a Service (II)

- Enabling technology: **virtualization**
  - An abstraction of logical resources away from underlying physical resources.
- Supported **properties**:
  - Manageability and interoperability
  - Availability and reliability
  - Scalability and elasticity
- Provided service:
  - **Virtual Machine** – As an IaaS provider, we should be able to provide the basic virtual machine operations, such as creation, suspension, resumption and termination, ... and be able to monitor some system states of each virtual machine, such as CPU loading, memory utilization, IO loading and internal network loading, ...
  - **Virtual Storage** – As an IaaS provider, we should be able to provide the basic virtual storage operations, such as space allocation, space release, data writing and data reading, ... and be able to monitor some storage states of each virtual storage, such as virtual space utilization, data duplication and storage device access bandwidth, ...
  - **Virtual Network** – As an IaaS provider, we should be able to provide the basic virtual network operations, such as IP address allocation, domain name register, connection establishment and bandwidth provision, ... and be able to monitor some network states of each virtual network, such as

Platform as a Service (I)

- The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider.

- The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

- Examples: Microsoft Windows Azure, Google App Engine,

Platform as a Service (II)

- Enabling technology: **runtime environment**
  - Refers to collection of software services available. Usually implemented by a collection of program libraries.
- Supported **properties**:
  - Manageability and interoperability
  - Performance and optimization
  - Availability and reliability
  - Scalability and elasticity
- Provided service:
  - **Programming IDE** – Users make use of programming IDE to develop their services among PaaS. This IDE should integrate the full functionalities which supported from the underling runtime environment. This IDE should also provide some development tools, such as profiler, debugger and testing environment.
  - **System Control Interface**
    - Police-Based Control
      - Typically described as a principle or rule to guide decisions and achieve rational outcome(s)
      - Make the decision according to some requirements
    - Workflow Control
      - Describe the flow of installation and configuration of resources

Software as a Service (I)

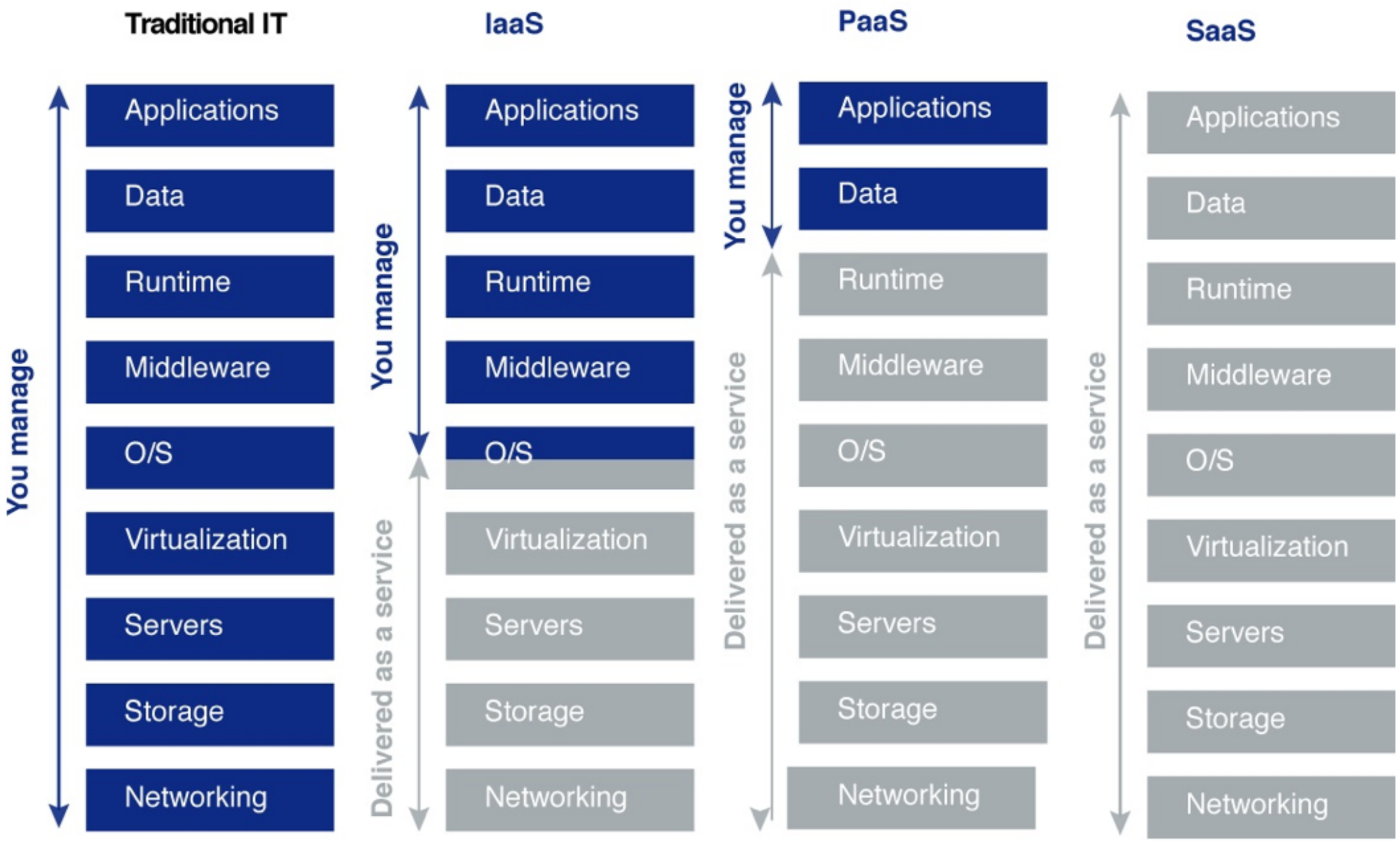
- The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email).

- The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

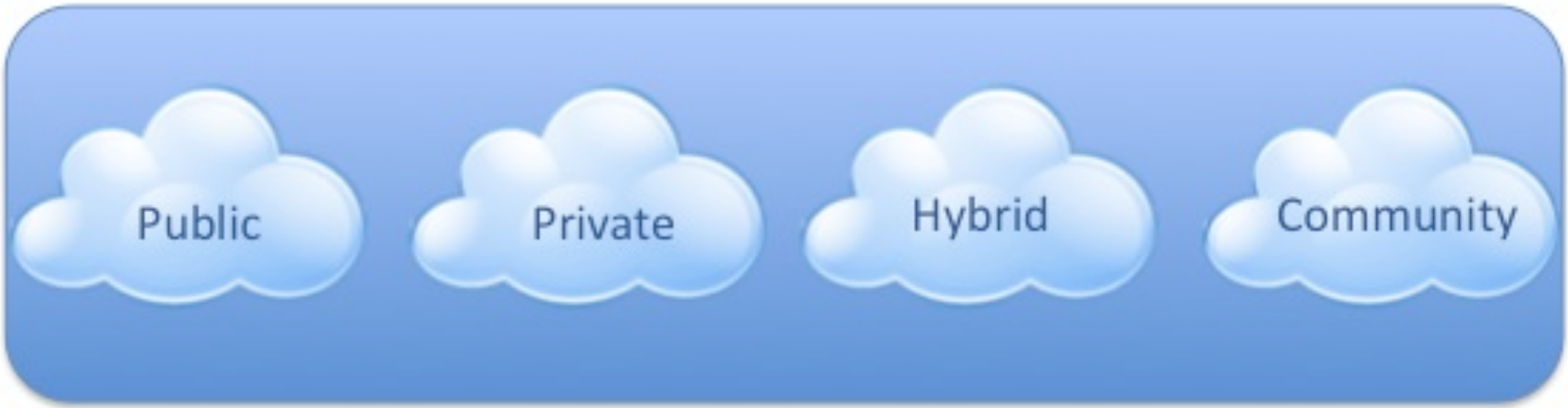
Software as a Service (II)

- Enabling technology: **web service**
  - Refers to a method of communication between two electronic devices over the World Wide Web.
- Supported **properties**:
  - Accessibility and portability
- Provided service:
  - **Web-based Applications**
    - General Applications – Applications which are designed for general propose, such as office suit, multimedia and instant message, ...
    - Business Applications – Application which are designed for business propose, such as ERP, CRM and market trading system, ...
    - Scientific Applications – Application which are designed for scientific propose, such as aerospace simulation and biochemistry simulation, ...
    - Government Applications – Applications which are designed for government propose, such as national medical system and public transportation system service, ...
  - **Web Portal**
    - Apart from the standard search engine feature, web portals offer other services such as e-mail, news, stock prices, information, databases and entertainment.
    - Portals provide a way for enterprises to provide a consistent look and feel with access control and

Service Models Summary

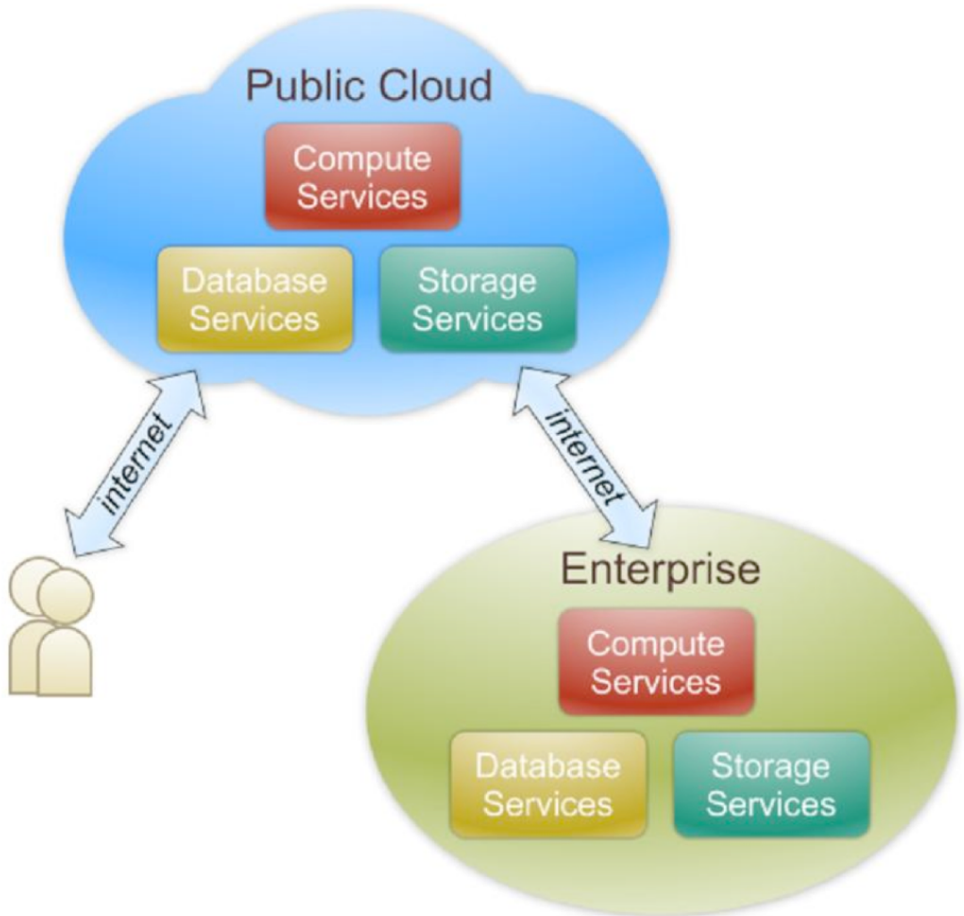


Deployment Models



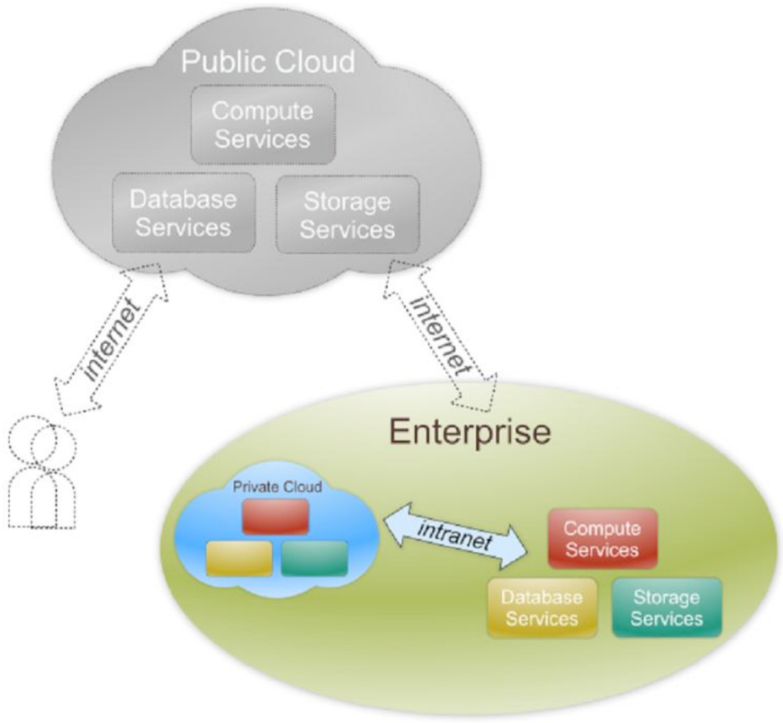
Public Cloud

- The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- Also known as external cloud or multi-tenant cloud, this model



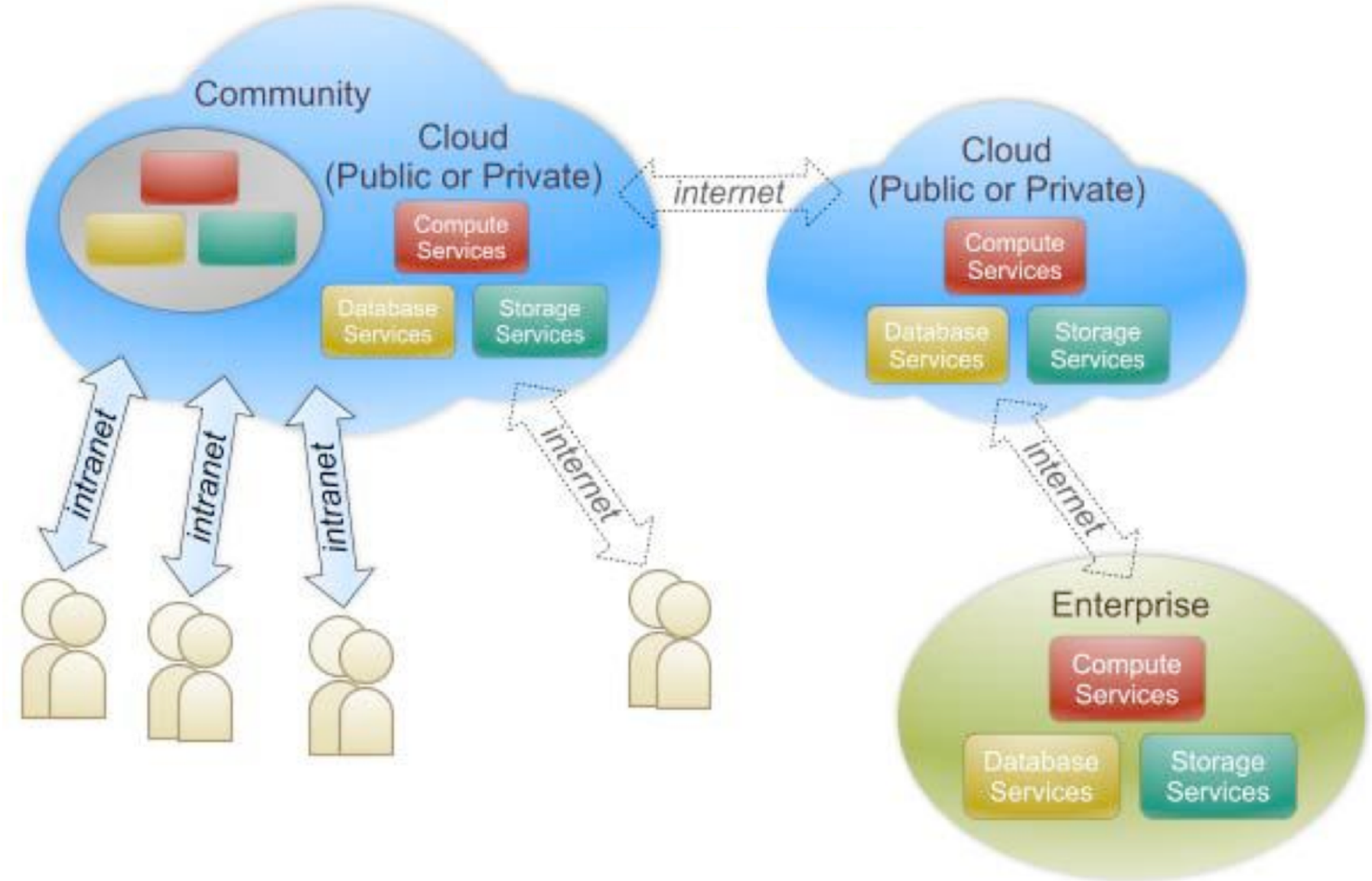
Private Cloud

- The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- Also referred to as internal cloud or on-premise cloud, a private cloud intentionally limits access to its resources to service consumers that belong to the same organization that owns the cloud.



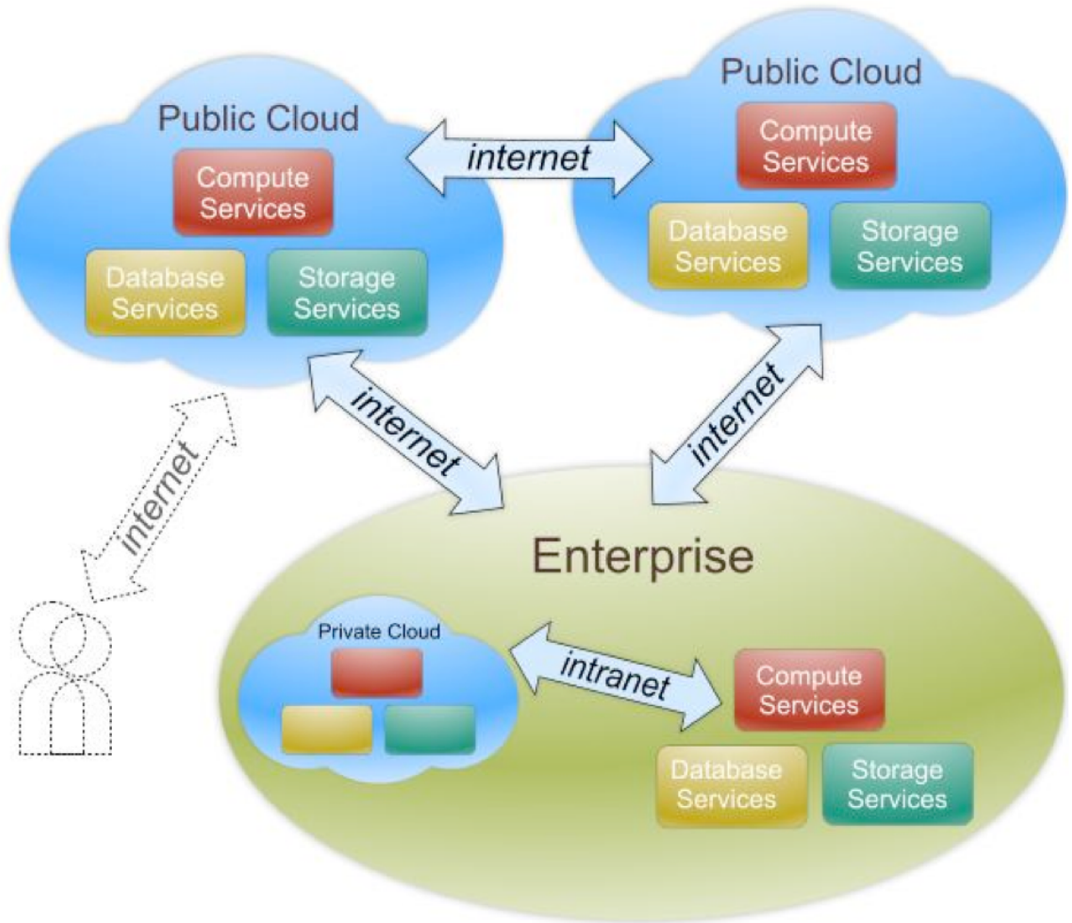
Community Cloud

- The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance conside



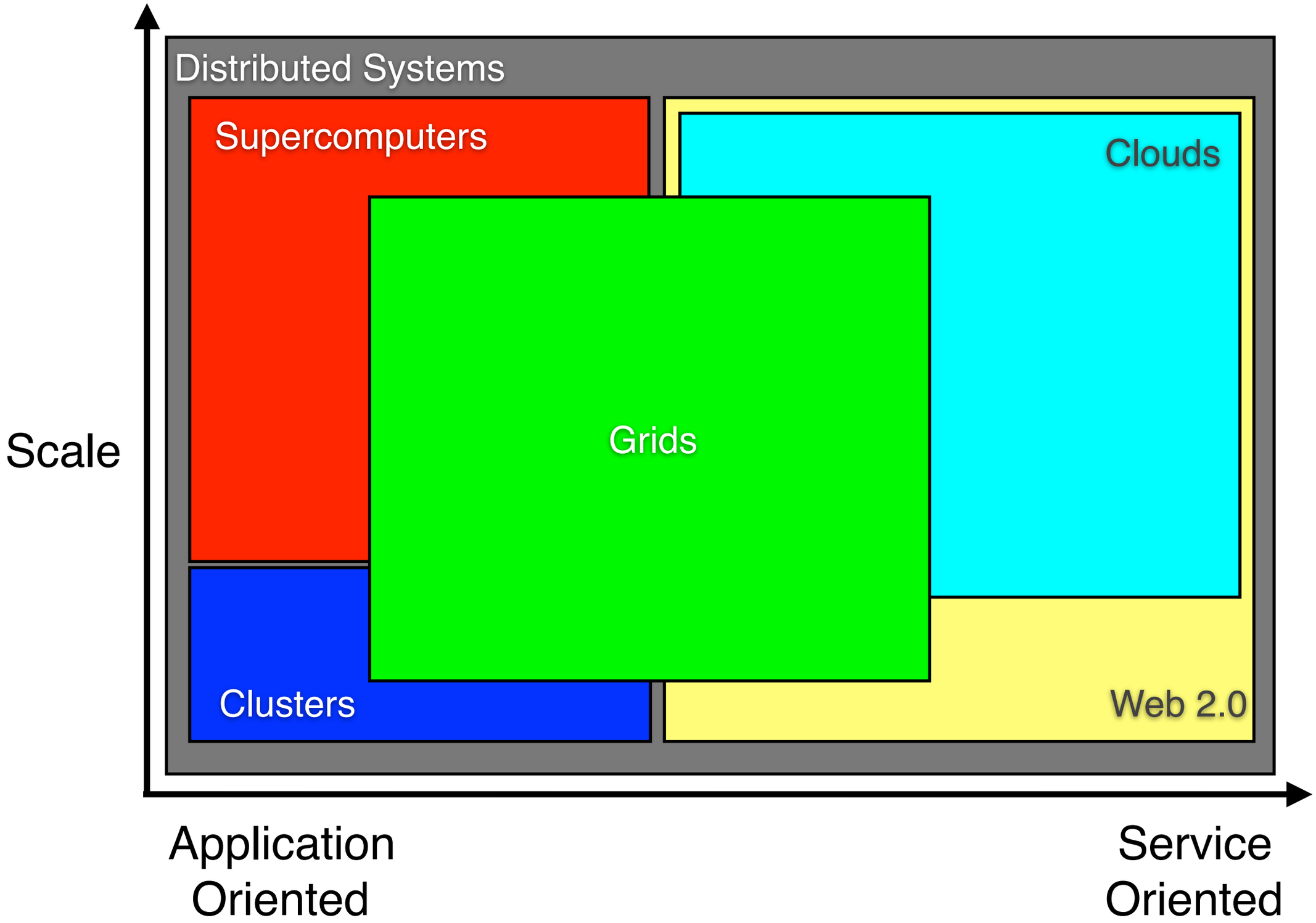
Hybrid Cloud

- The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-



Grids versus Clouds

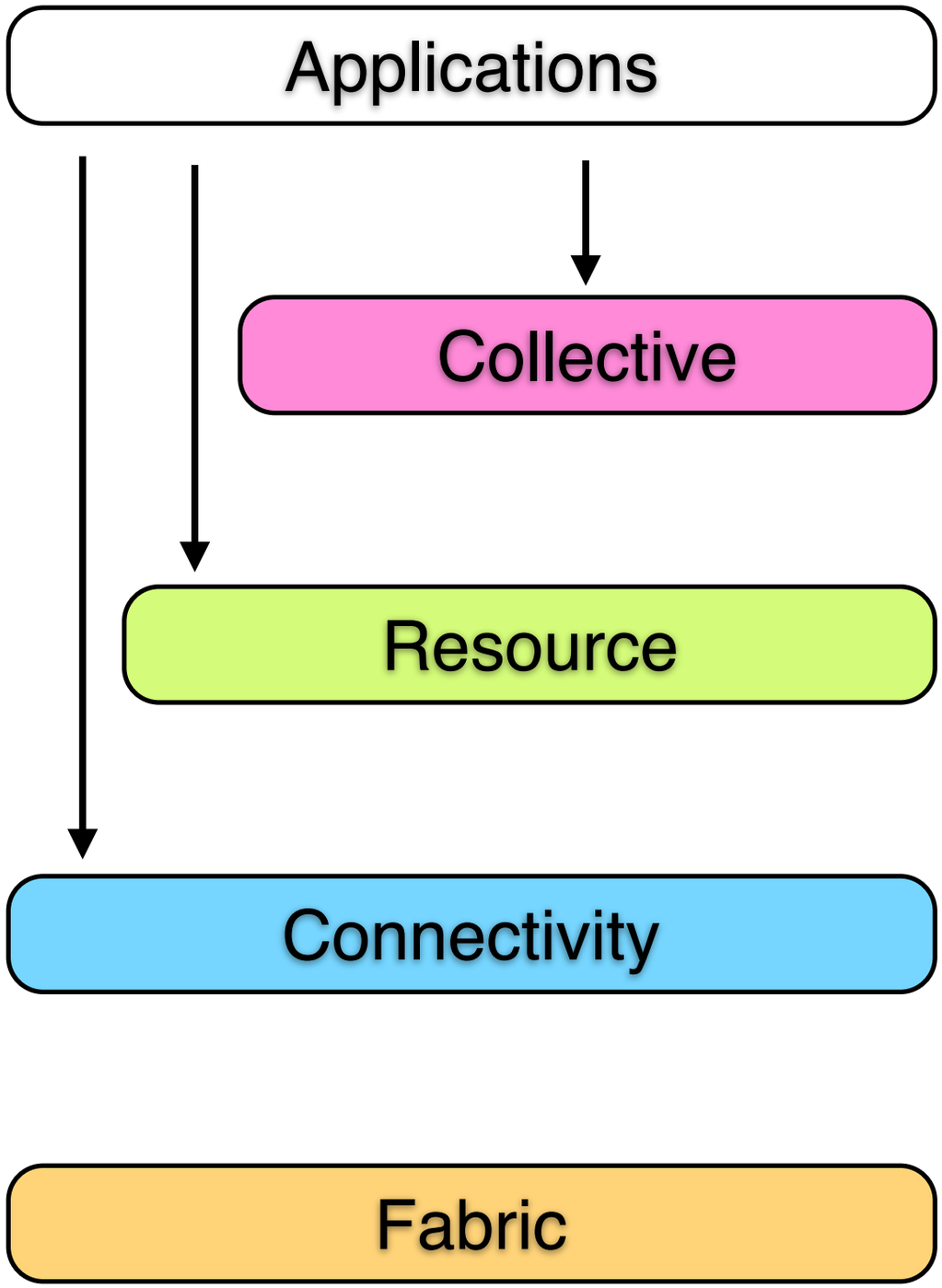
Grids and Clouds Overview

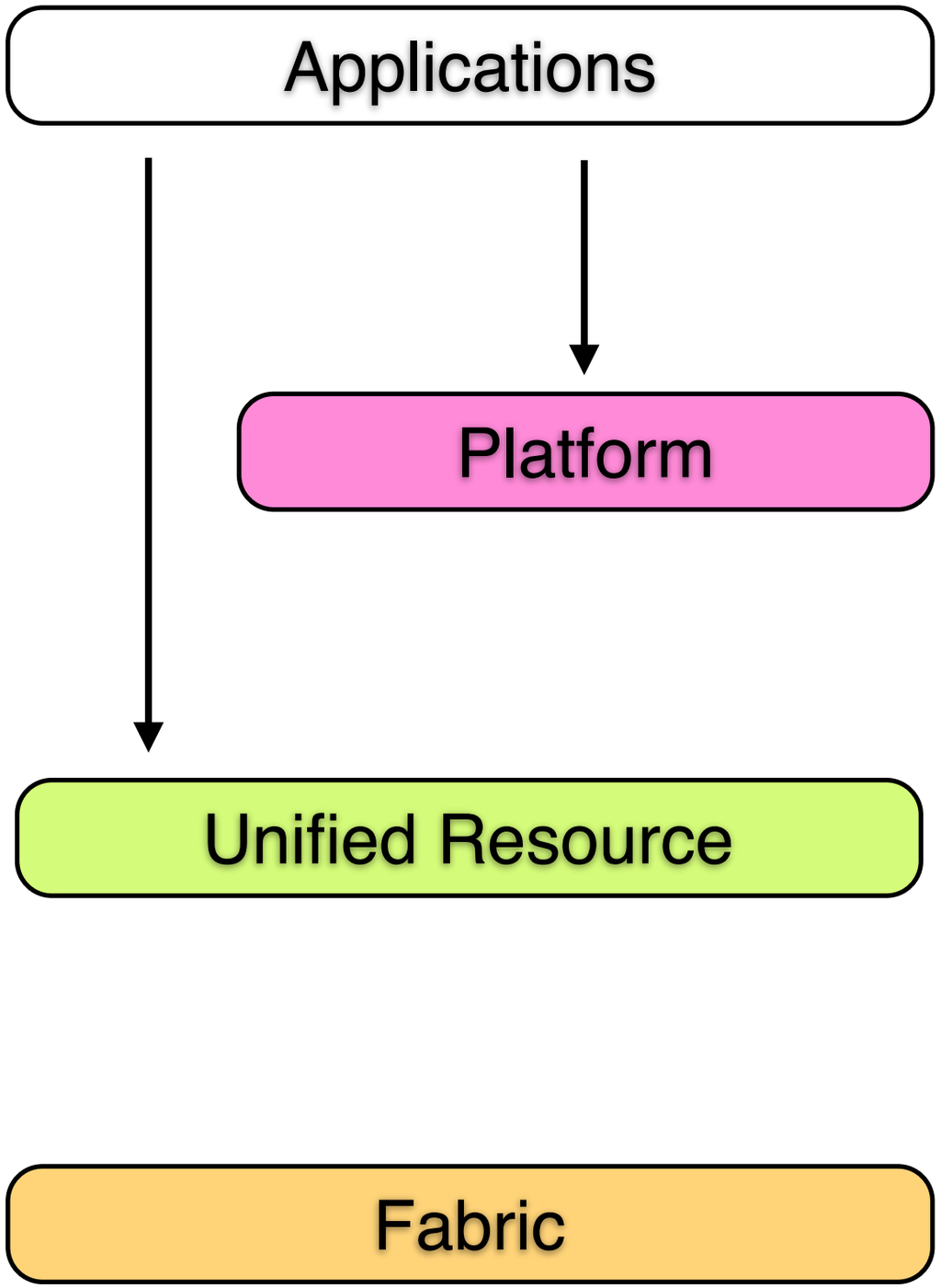


Business Model

- **Classical Computing**
  - One-time payment for unlimited use
  - Capital expenditure
  - Subject to de-pricing
- **Grid Computing**:
  - Project-oriented service units (i.e. CPU hours)
  - Prevent capital expenditure
  - Shared operational expenditure
- **Cloud Computing**
  - Pay-per-use
  - Economy of scale
  - Amazon charges instance-hours, GB/month (disk), TB/month

Architecture





Compute and Data Models

- **Grid Computing**:

- Batch-scheduled compute model (LRM + GRAM)
- Dedicated resources governed by queue scheduling systems
- No native support for interactive applications
- Space-sharing allocations
- Data-to-code

- **Cloud Computing**

- Time sharing
- All users share everything
- Elasticity
- Pay-per-use
- Code-to-data (but multicores can limit)

Programming Model

- **Grid Computing**:
  - Large-scale scientific computations
  - Heterogenous resources
  - Different administrative domains
  - Fast and efficient codes
  - Complex coding (focus on non-functional logic)
  - Example: MPI
- **Cloud Computing**
  - Large-scale data processing
  - Homogeneous resources
  - Same administrative domain
  - Data crunching code
  - Simple coding (focus on business logic)

Application Model

- **Grid Computing**:
  - High Performance Computing
    - Tightly coupled parallel jobs
  - High Throughput Computing
    - Loosely coupled parallel jobs
    - Small Number of Large Batch Jobs
- **Cloud Computing**
  - Still at infancy
  - Independent Jobs
  - Large Number of Small Interactive Jobs

Security Model

- **Grid Computing**:

- Heterogenous Resources
- Multiple Administrative Domains
- Large Interoperability
- Based on GSI, PKI, SSL/TLS, SSO, Delegation

- **Cloud Computing**

- Homogeneous Resources
- Single Administrative Domain
- Limited Interoperability
- Based on Web forms over SSL, emails, credit card

Typical Security Concerns

1. **Privileged user access**: sensitive data processed outside the enterprise needs the assurance that they are only accessible and propagated to privileged users
2. **Regulatory compliance**: a customer needs to verify if a Cloud provider has external audits and security certifications and if their infrastructure complies with some regulatory security requirements
3. **Data location**: since a customer will not know where her data will be stored, it is important that the Cloud provider commit to storing and processing data in specific jurisdictions and to obey local privacy requirements on behalf of the customer
4. **Data segregation**: one needs to ensure that one customer's data is fully segregated from another customer's data;
5. **Recovery**: it is important that the Cloud provider has an efficient replication and recovery mechanism to restore data if a disaster occurs;
6. **Investigative support**: Cloud services are especially difficult to investigate, if this is important for a customer, then such support needs to be ensured with a contractual commitment;
7. **Long-term viability**: your data should be viable even the Cloud provider is acquired by another company.

Reading Assignments

- NIST (National Institute of Standards and Technology), *NIST Cloud Computing Reference Architecture*, http://csrc.nist.gov/groups/SNS/cloud-computing/

- M. Armbrust et al., *Above the Clouds: A Berkeley View of Cloud Computing*, Technical Report No. UCB/EECS-2009-28, University of California at Berkeley, 2009.

- I. Foster et al., *Cloud Computing and Grid Computing 360-Degree Compared*, Grid Computing Environments Workshop, 2008. GCE '08 , pp.1-10, 12-16 Nov. 2008

- R. Buyya et al., *Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility*, FGCS 25(6), pp. 599-616, Jun 2009