CLOUD COMPUTING:
technology and market drivers
implications for procurement and offering models
for users and suppliers

Corso Sistemi Informativi
Università Roma3
January 28th, 2010

Enzo Carboni
IBM - Global Techonology Services

Il cloud computing è un nuovo paradigma tecnologico destinato a sovvertire i modelli attuali

Socrate: Per l'ètra movo, e il sol dall'alto io guardo!

Strepsiade: E stando in terra, i Numi non li puoi guardar dall'alto? Ci vuole il corbello?

Socrate: I celesti fenomeni scrutare
giammai potrei dirittamente, senza
tener sospesa la mia mente, e mescere
il sottil pensier nell'omogeneo
ètra. Se dalla terra investigassi,
di giú le cose di lassú, non mai
le scoprirei; poiché la terra a forza
attira a sé l'umore dell'idea.

Le Nuvole – Aristofane
vv. 227-234
423 a.C.

Il cloud computing è soltanto una delle tante mode (hype) che i fornitori inventano per far crescere le vendite

Socrate: *Nuvole celesti, sono, Dee*
solenni degli scansafatiche. Esse le idee
ci dànno, la dialettica, la ciurmeria, l'ingegno,
la chiacchiera, il ghermire concetti, il dar nel segno!

Strepsiade: *Per questo, al solo udirle, sembra che metta piume*
il mio spirito, e cerca di parlar con acume,
di dir fumose ciance, di bucare concetti
con piú fini concetti, di opporre detti a detti.
Sicché, vorrei, se posso, veder come son fatte!

Le Nuvole – Aristofane
vv. 316 – 320

Agenda

1 **BASE CONCEPTS**

2 **ECONOMICS OF CLOUD COMPUTING**

3 **IT SOURCING OPTIONS IMPLICATIONS OF CLOUD COMPUTING**

4 **DYNAMICS OF THE TRANSFORMATION**

Varying opinions on the definition of cloud computing, but some common attributes are emerging

IT Customers:
- Flexible pricing/business models
- Outsourced, on demand provisioning
- Unlimited scaling
- SW developer platform
- Flexible

Globalization and Globally Available Resources

IT Analysts:
- Variable pricing
- No long term commitments
- Hosted, on demand provisioning
- Massive, elastic scaling
- Standard Internet technology
- Abstracted infrastructure
- Service-oriented Platforms

Billions of mobile devices accessing the World Wide Web

Common Attributes of Clouds

Rapid provisioning
Elastic scaling
Standardized Offerings
Advanced virtualization
Flexible pricing

Real-time data streams and information sharing

Press:
- Pay by consumption
- Lower costs
- On demand provisioning
- Grid and SaaS combination
- Massive scaling
- Efficient infrastructure
- Simple and easy

Rise of collaborative technologies

Financial Analysts:
- Utility pricing
- Hosted, a-a-s provisioning
- Parallel, on demand processing
- Scalable
- Virtualized, efficient infrastructure
- Flexible

Source: IBM Corporate Strategy analysis of MI, PR, AR and VCG compilations

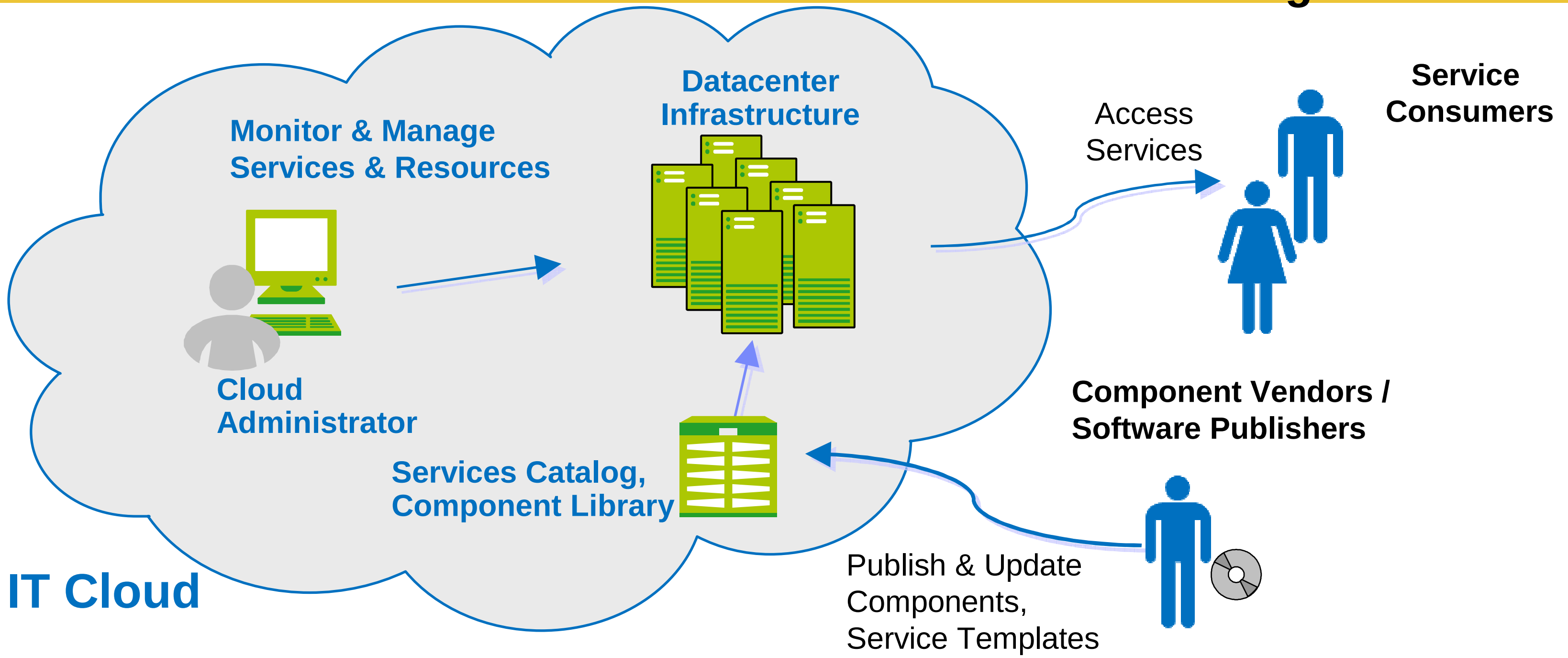
IBM's definition of Cloud Computing

A user experience and a business model

- Cloud computing is an emerging style of IT delivery in which applications, data, and IT resources are **rapidly provisioned** and provided as **standardized offerings** to users over the web in a **flexible acquisition model.**

An infrastructure management and services delivery methodology

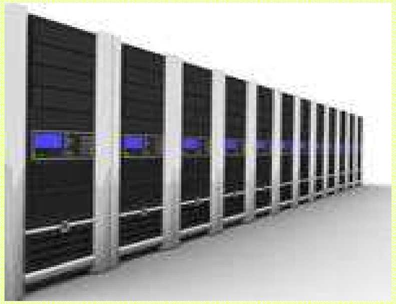
- Cloud computing is a way of managing large numbers of **highly virtualized resources** such that, from a management perspective, they resemble a single large resource. This can then be used to deliver services with **elastic scaling**.



The key building blocks of Cloud Computing are familiar to IBM; a cloud implementation is a journey encompassing all these building blocks

Cloud

Simplification / Consolidation



- Consolidation, systems management,
- Reduce staffing needs, and costs
- Improve business resilience and utilization

Advanced Virtual Resource Pools

- Remove physical resource boundaries
- Improve scalability, increase utilization
- Reduce costs

Advanced Service Management

- Service catalog, metering, and automated deployment of virtualized resources
- Integrated virtualization management with IT processes

User Self Service

- Centralized, robust, self serve portal for 24X7 access to services
- Improve user satisfaction & productivity

Standardization

Elements that Drive Cloud Efficiency and Economics

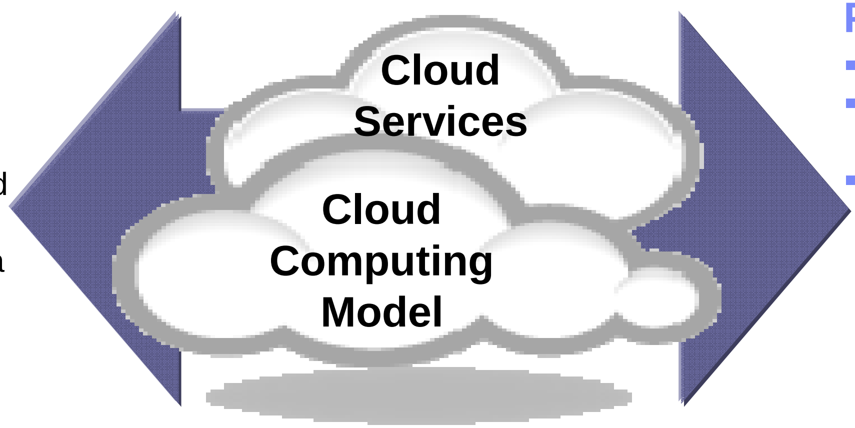
Infrastructure Leverage	Virtualization of Hardware	→	Drives lower capital requirements
Infrastructure Leverage	Utilization of Infrastructure	→	Virtualized environments only get benefits of scale if they are highly utilized
Labor Leverage	Self Service	→	Clients who can “serve themselves” require less support and get services
Labor Leverage	Automation of Management	→	Take repeatable tasks and automate
Labor Leverage	Standardization of Workloads	→	More complexity = less automation possible = people needed

Cloud Computing delivery models

Flexible Delivery Model

Public …

- Service provider owned and managed.
- Access by subscription.
- Delivers select set of standardized business process, application and/or infrastructure services on a flexible price per use basis



Private …

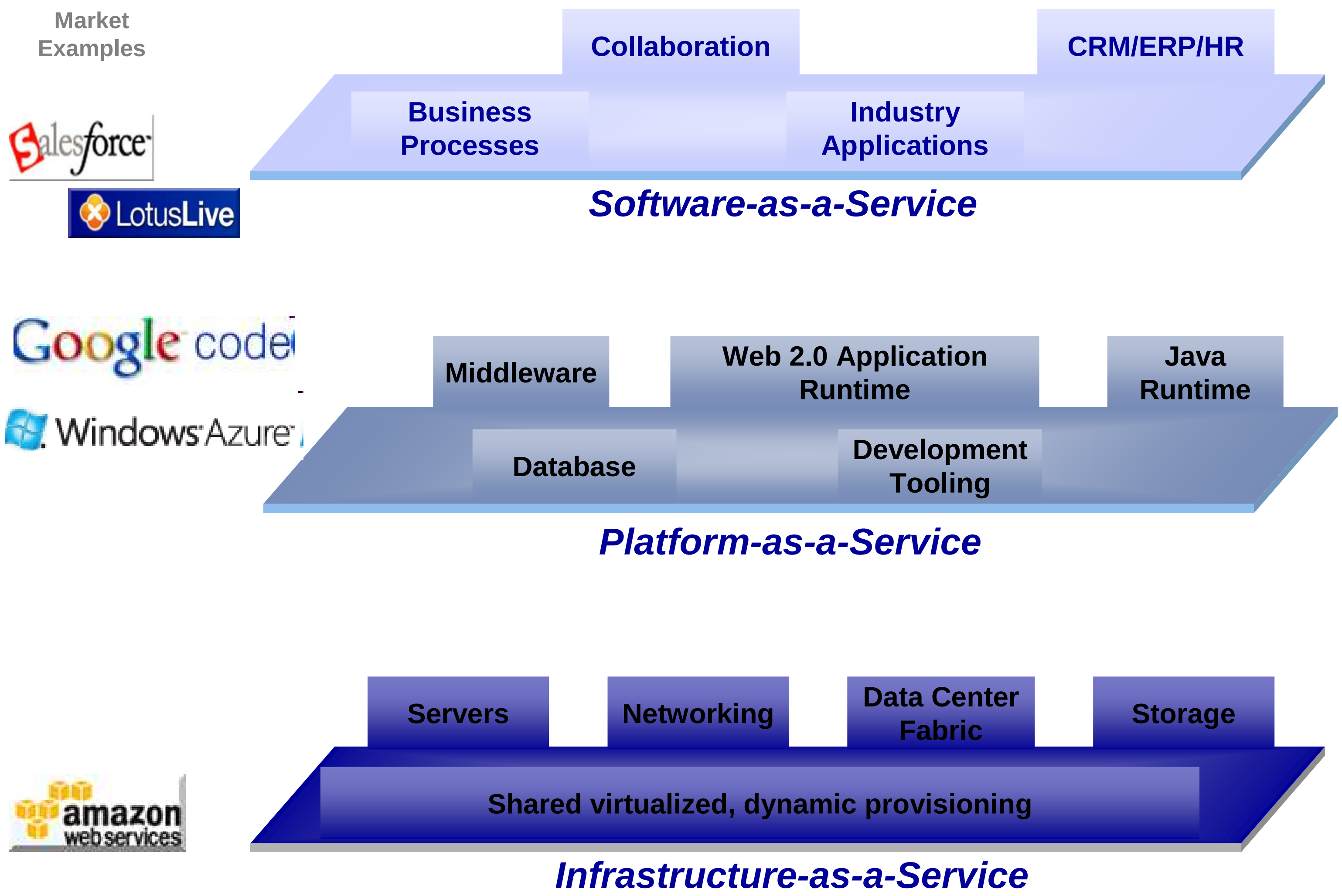
- Client owned and managed.
- Access limited to client and its partner network.
- Drives efficiency, standardization and best practices while retaining greater customization and control

….Standardization, capital preservation, flexibility and time to deploy

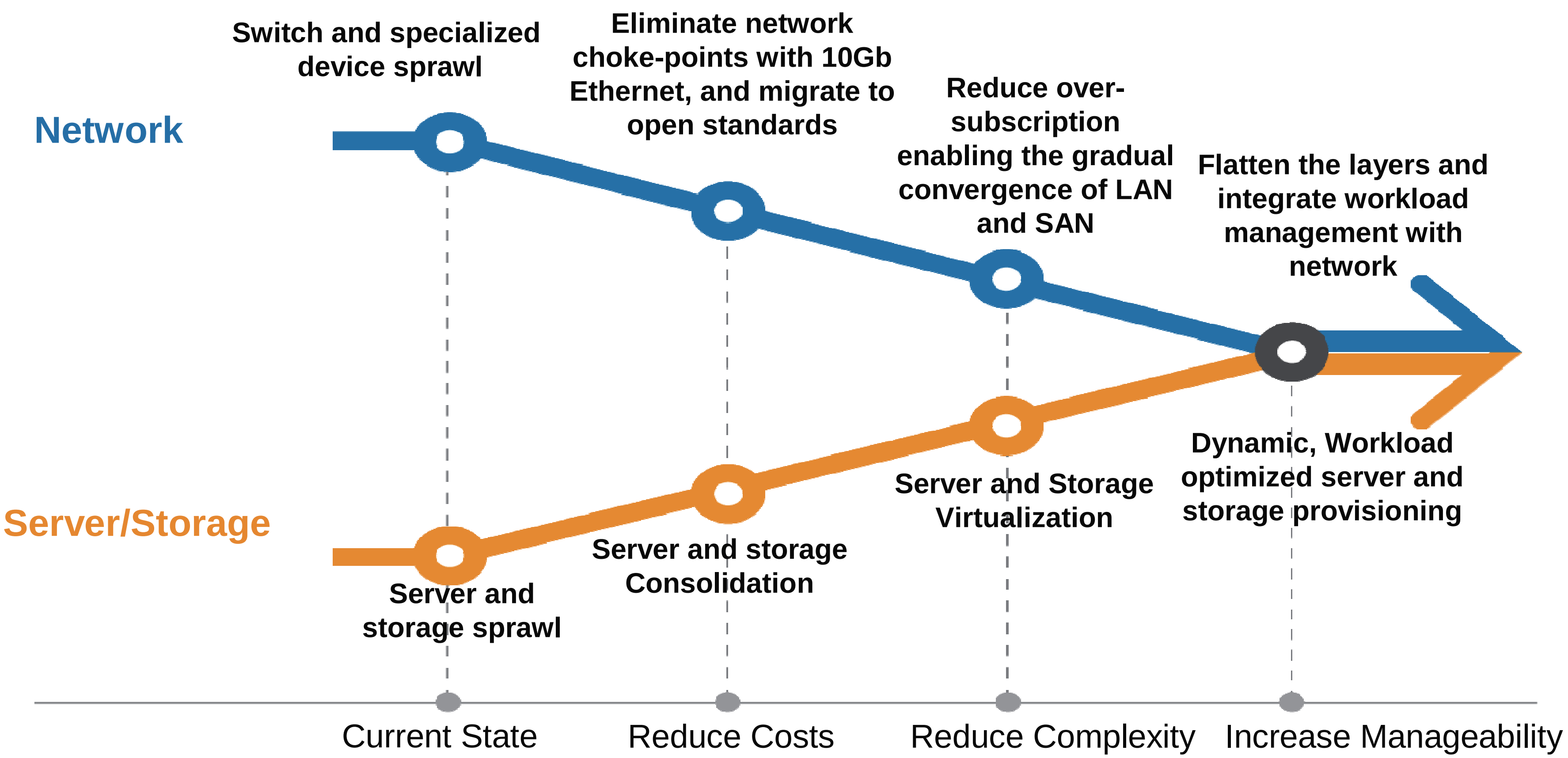
…. Customization, efficiency, availability, resiliency, security and privacy

CULTURE **ORGANIZATION** **CHANGE**

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services



Networks for cloud computing must become flexible, responsive and managed together with the rest of the IT infrastructure.



Agenda

1 BASE CONCEPTS

2 **ECONOMICS OF CLOUD COMPUTING**

3 IT SOURCING OPTIONS IMPLICATIONS OF CLOUD COMPUTING

4 DYNAMICS OF THE TRANSFORMATION

Cloud Computing is a new term for a long-held dream of computing as a utility that has the potential to have the same impact on software that foundries have had on the hardware industry.

- A semiconductor fabrication line costs over $3B today, so only a handful of major “merchant” companies with very high chip volumes, such as Intel and Samsung, can still justify owning and operating their own fabrication lines
- Foundries enable “fab-less” semiconductor chip companies whose value is in innovative chip design: A company such as nVidia can now be successful in the chip business without the capital, operational expenses, and risks associated with owning a state-of-the-art fabrication line.
- The construction and operation of extremely large-scale, commodity-computer datacenters at lowcost locations was the key necessary enabler of Cloud Computing, for they uncovered the factors of 5 to 7 decrease in cost of electricity, network bandwidth, operations, software, and hardware available at these very large economies of scale.
- These factors, combined with statistical multiplexing to increase utilization compared a private cloud, meant that cloud computing could offer services below the costs of a medium-sized datacenter and yet still make a good profit.
- The advantages of the economy of scale and statistical multiplexing may ultimately lead to a handful of Cloud Computing providers who can amortize the cost of their large datacenters over the products of many “datacenter-less” companies.

- *By the year 2000, the most successful computer companies will be those that buy computers rather than build them … computers have become too powerful for the uses to which they are being put … Value derives from scarcity. In the computer industry, scarcity now resides in the gap between power—what computers and their underlying semiconductor technologies are capable of doing—and utility—what human imagination and software engineering are capable of enabling computers to do.*
  - “The Computerless Computer Company”
  - by Andrew S. Rappaport and Shmuel Halevi – Harvard Business Review – July-August 1991

A necessary condition for a company to become a Cloud Computing provider is that it must have investments in very large datacenters and in large-scale software infrastructure and operational expertise required to run them

- Elastic Compute Cloud (EC2) from Amazon Web Services (AWS) sells 1.0-GHz x86 ISA “slices” for 10 cents per hour, and a new “slice”, or instance, can be added in 2 to 5 minutes
- Amazon’s Scalable Storage Service (S3) charges $0.12 to $0.15 per gigabyte-month, with additional bandwidth charges of $0.10 to $0.15 per gigabyte
- A necessary but not sufficient condition for a company to become a Cloud Computing provider is that it must have existing investments not only in very large datacenters, but also in large-scale software infrastructure and operational expertise required to run them.

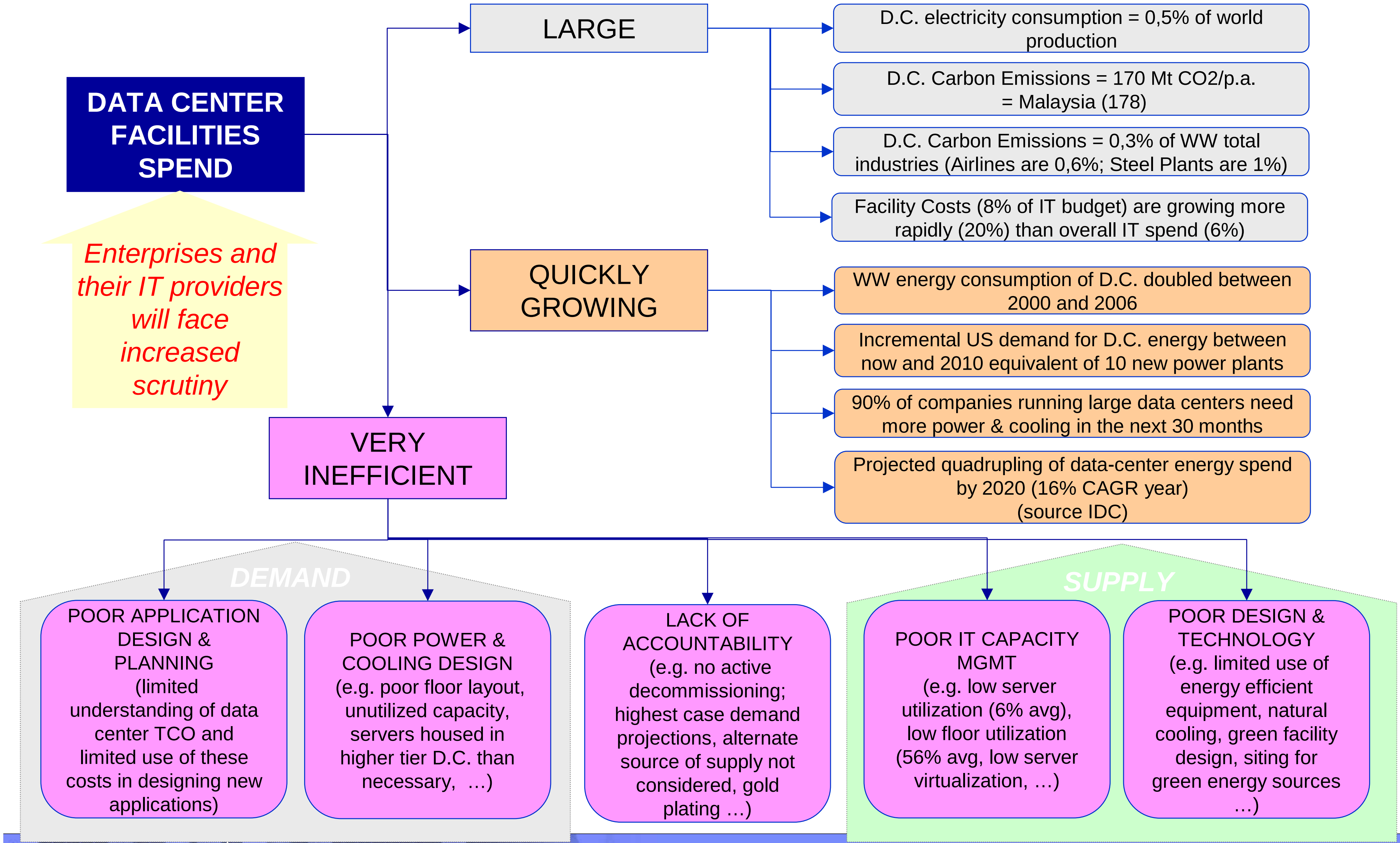
Technology	Cost in Medium-sized DC (≈ 1.000 servers)	Cost in Very Large DC (≈ 50.000 servers)	Ratio
Network	$95 per Mbit/sec/month	$13 per Mbit/sec/month	7.1
Storage	$2.20 per GByte / month	$0.40 per GByte / month	5.7
Administration	≈ 140 Servers / Administrator	> 1000 Servers / Administrator	7.1

Price per KWH	Where	Possible Reasons Why
3.6¢	Idaho	Hydroelectric power; not sent long distance
10.0¢	California	Electricity transmitted long distance over the grid; limited transmission lines in Bay Area; no coal fired electricity allowed in California.
18.0¢	Hawaii	Must ship fuel to generate electricity

But additional technology trends and new business models are also playing a key role in making cloud computing a reality this time around

- The emergence of Web 2.0 was a shift from “high-touch, high-margin, high-commitment” provisioning of service “low-touch, low-margin, low-commitment” self-service
  - Web 1.0, accepting credit card payments from strangers required a contractual arrangement with a payment processing service such as VeriSign
  - With the emergence of PayPal, however, any individual can accept credit card payments with no contract, no long-term commitment, and only modest pay-as-you-go transaction fees
- Several important classes of existing applications will become even more compelling with Cloud Computing and contribute further to its momentum
  - Mobile interactive applications
    - services that respond in real time to information provided either by their users or by nonhuman sensors
  - Parallel batch processing
    - batch-processing and analytics jobs that analyze terabytes of data and can take hours to finish can take advantage of the cloud’s new “cost associativity”: using hundreds of computers for a short time costs the same as using a few computers for a long time
  - The rise of analytics
    - while online transaction volumes will continue to grow slowly, decision support is growing rapidly, shifting the resource balance in database processing from transactions to business analytics

Data Centers are a large source of GreenHouse Gas (GHG) emissions



Based on McKinsey analysis

Cloud computing could become an option to reduce GHG emissions

- Data Center spending is so big, destined to grow so fast starting from an average efficiency so poor that the problem will take a relevant priority in CIOs' agendas
- Because of the relevance of carbon footprint related to the Data Centers, it can be expected that in the next years the regulation will become very heavy and initiatives to reduce GHG emissions will put pressure on IT, outside its evolutionary mission
- Improving the efficiency of the data centers is a very complex multi-faceted problem; moreover the role of IT in the enterprises is supposed to shift more and more toward supporting the business; it can be expected that propensity to buy infrastructure as-a-service will grow significantly

Example of actual existing Data Center site and in-flight improvement to double CADE (Corporate Average Data Efficiency) rating

Within an actual data center currently on CADE tier 1

CADE: 1.8%
Tier 1

=

Facility Efficiency: 36%

x

IT Asset Efficiency: 5%

- CPU utilization across 20,000 servers averages 5%

Facility Energy Efficiency: 48%

x

Facility Utilization: 75%

- Actual energy delivered to IT of 10,550 kW divided by actual energy draw from grid of 22,200 kW

- Energy reaching IT servers, storage and networking equipment of 10,550 kW divided by Facility capacity of 14,000 kW

Year 1 improvements underway to enable doubling of CADE by 2012

- Remove 4,000 dead servers. Average CPU utilization increases by 10%*
- Virtualize 8,000 servers on 4 to 1 ratio with 50% utilization. Further increase avg CPU utilization to 15% to 20%*
- Implementing full suite of industry best practices. increases facility energy efficiency to ~53%
- Defer new data center construction allowing 15% annual organic IT growth to increase facility utilization

* Removing servers will decrease facility utilization and facility efficiency; newly freed capacity assumed to be consumed by organic growth

Source: Uptime Institute

Several obstacles still exist and cannot be ignored; one of the key lessons we all hopefully learned from the dot-com bubble is that complex, global initiatives take time.

	Obstacle	Possible work arounds / solutions
1	Availability of Service	Use Multiple Cloud Providers
2	Data Lock-In	Standardize APIs
3	Security, Data Confidentiality and Auditability	Deploy Encryption, VLANs, Firewalls; Geographical Data Storage
4	Data Transfer Bottlenecks	FedExing Disks; Data Backup/Archival; Higher BW Switches
5	Performance Unpredictability	Improved VM Support; Flash Memory;
6	Scalable Storage	Invent Scalable Store
7	Bugs in Large Distributed Systems	Invent Debugger that relies on Distributed VMs
8	Scaling Quickly	Invent Auto-Scaler that relies on ML; Snapshots for Conservation
9	Reputation Fate Sharing	
10	Software Licensing	Pay-for-use licenses; Bulk use sales

The first three are technical obstacles to the adoption of Cloud Computing, the next five are technical obstacles to the growth of Cloud Computing once it has been adopted, and the last two are policy and business obstacles to the adoption of Cloud Computing.

Agenda

Today's enterprises can increasingly "program" the business by selecting from a variety of established modules or disparate components

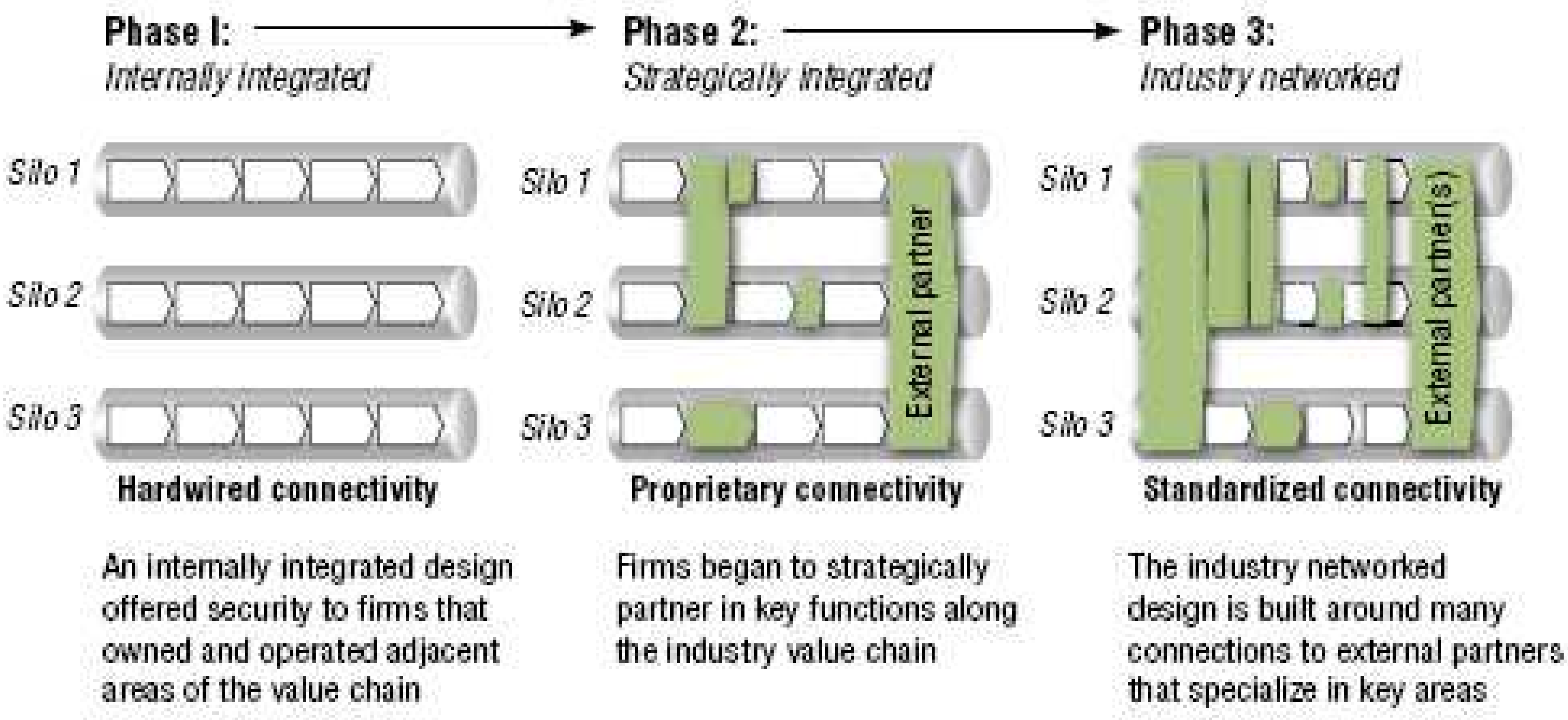
- Over the last 5-7 years, a number of diverse business and technology architectures have matured and converged to form a **global connectivity platform** that supports **widespread collaboration**.
- By slashing the cost of coordination both within the firm and externally, with partners, this new platform represents a **de facto weakening of traditional business structures and boundaries**.
- The rapid **decline of transaction costs** is having an especially profound and lasting effect on ownership decisions

1. **Communication networks**, specifically broadband and wireless technologies, have made digital connectivity faster and more affordable.
2. With the consolidation of the enterprise software market and the proliferation of business integration software, companies evolve toward a **common IT platform** upon which broader and better functionality can be built.
3. **Open standards** – both technology and business – are optimizing interoperability and creating the potential for truly modularized infrastructures.

Changes in interaction costs are causing entire industries to reorganize rapidly and dramatically.

- As business interactions move on to electronic networks such as the Internet, basic assumptions about corporate organization will be overturned
- Activities that companies have always believed to be central to their businesses will suddenly be offered by new, specialized competitors that can do those activities better, faster, and more efficiently. Executives will be forced to ask the most basic and discom-fiting question about their companies: what business are we really in?
- Developing or keeping a capability in-house confers no differentiation if an outside specialist can provide the same more effectively or efficiently

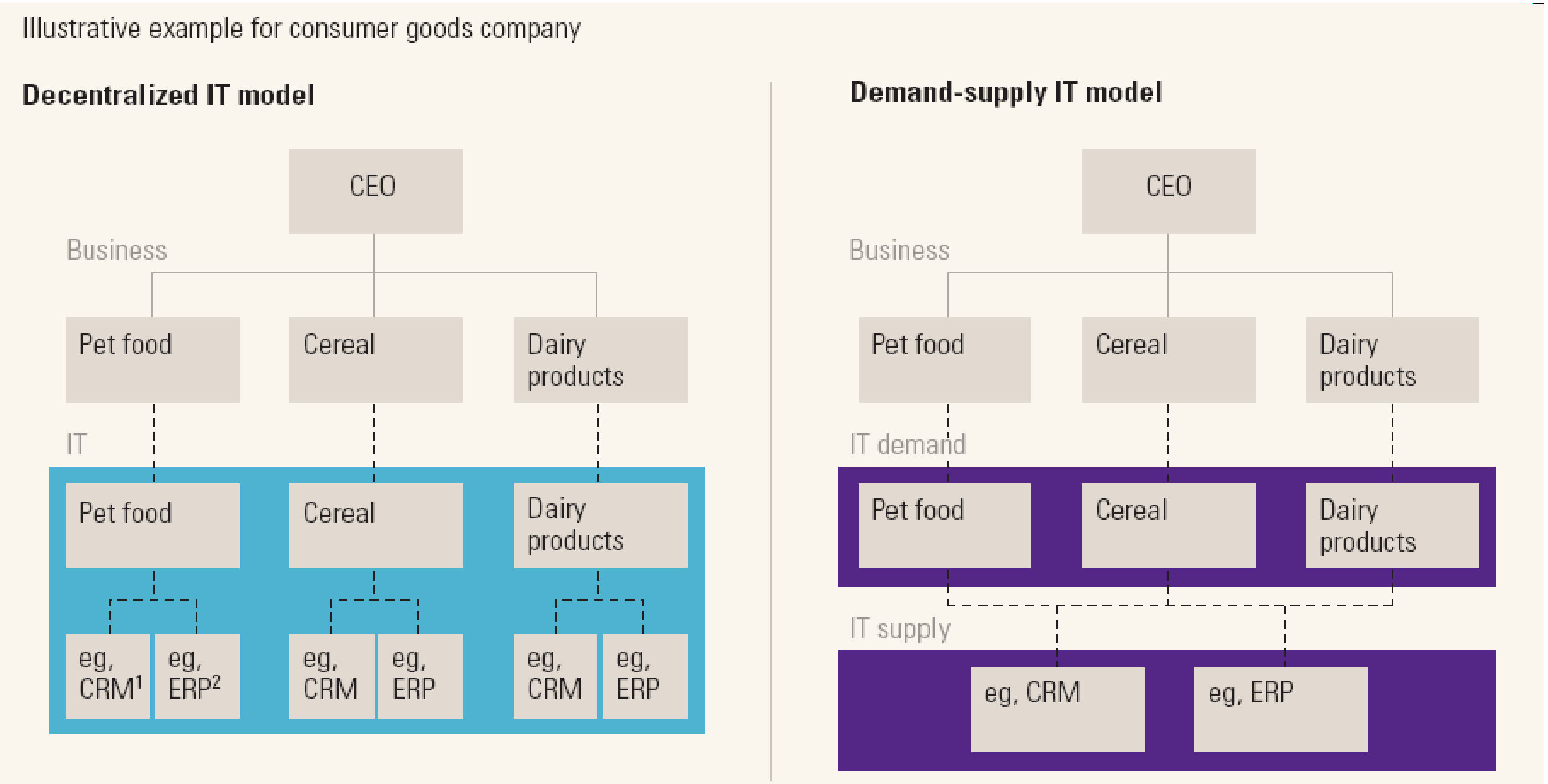
As standards-driven internal specialization matures, firms gain the ability to leverage the benefits of lower transaction costs by engaging with partners through collaborative industry networks.



Source: IBM Institute for Business Value.

- The reduction of the transaction costs changes the rule of the business.
- The entire Value Net must be analysed and re-designed to exploit the emerging opportunities

Companies struggle to achieve both agility and efficiency by splitting IT supply units, as service centers aligned along broad functional pillars, from IT demand management units aligned with BUs, but



- Opex are expenses accounted on the basis of a fixed amount per year, and are proportional to the assets being managed → difficult to variabilize
- Capex are investments related to the projects needed to deliver new services to the BUs → assets are purchased from start → difficult to variabilize

- **Cloud Computing appears as an oppotunity to reduce and variabilize both**

Considering the growing pressure on IT costs both demand managers and IT operations will look at the opportunities offered by the cloud market. Adoption will be shaped by workload affinity and characteristics

Risk and migration cost may be too high

- Database
- Transaction processing
- ERP workloads
- Highly regulated workloads
- Security

Workloads that can be standardized for cloud

- Web infrastructure applications
- Collaboration infrastructure
- Development and test
- High performance computing

Workloads that are made possible by cloud

- High volume, low cost analytics
- Collaborative Business Networks
- Industry scale “smart” applications

Different utility computing offerings will be distinguished based on the level of abstraction presented to the programmer and the level of management of the resources.

- Any application needs a model of computation, a model of storage, and a model of communication.
  - The statistical multiplexing necessary to achieve elasticity and the illusion of infinite capacity requires each of these resources to be virtualized to hide the implementation of how they are multiplexed and shared
- Different utility computing offerings will be distinguished based on the level of abstraction presented to the programmer and the level of management of the resources.
- Amazon EC2 is at one end of the spectrum. An EC2 instance looks much like physical hardware, and users can control nearly the entire software stack, from the kernel upwards.
  - This low level makes it inherently difficult for Amazon to offer automatic scalability and failover, because the semantics associated with replication and other state management issues are highly application-dependent.
- At the other extreme of the spectrum are application domains pecific platforms such as Google AppEngine.
  - AppEngine is targeted exclusively at traditional web applications, enforcing an application structure of clean separation between a stateless computation tier and a stateful storage tier.

* Source: Above the Clouds: A Berkeley View of Cloud Computing; February 10, 2009

The cost of wide-area networking has fallen more slowly than all other IT hardware costs; this has to be taken into account when considering to split data from computation

	WAN bandwidth/mo.	CPU hours (all cores)	disk storage
Item in 2003	1 Mbps WAN link	2 GHz CPU, 2 GB DRAM	200 GB disk, 50 Mb/s transfer rate
Cost in 2003	$100/mo.	$2000	$200
$1 buys in 2003	1 GB	8 CPU hours	1 GB
Item in 2008	100 Mbps WAN link	2 GHz, 2 sockets, 4 cores/socket, 4 GB DRAM	1 TB disk, 115 MB/s sustained transfer
Cost in 2008	$3600/mo.	$1000	$100
$1 buys in 2008	2.7 GB	128 CPU hours	10 GB
cost/performance improvement	2.7x	16x	10x
Cost to rent $1 worth on AWS in 2008	$0.27–$0.40 ($0.10–$0.15/GB x 3 GB)	$2.56 (128x2 VM's@$0.10 each)	$1.20–$1.50 ($0.12–$0.15/GB-month x 10 GB)

Examples of available cloud services from IBM

Business & IT Workloads

Deployment models	**Analytics**	**Collaboration**	**Development and Test**	**Desktop and Devices**	**Infrastructure Storage**	**Infrastructure Compute**	**Infrastructure Back-up and recovery**	**Business Services**
Public Cloud - Standardized services provided by a supplier		IBM Lotus Live IBM Lotus iNotes®	IBM Smart Business Development & Test on the IBM Cloud (open beta)	IBM Smart Business Desktop on the IBM Cloud Smart Business End User Support – Service Assist	IBM Smart Business Storage on the IBM Cloud	Smart Business Compute on the IBM Cloud	IBM Information Protection Services	Smart Business Expense Reporting on the IBM Cloud
Private Cloud - services, behind your firewall, built and/or managed by a supplier	IBM Smart Analytics Cloud		IBM Smart Business Test Cloud	IBM Smart Business Desktop Cloud	IBM Smart Business Storage Cloud			
Hybrid Cloud – Customized services provided by a supplier	IBM Smart Analytics System		IBM CloudBurst™ family		IBM Information Archive			Smart Business for SMB

Agenda

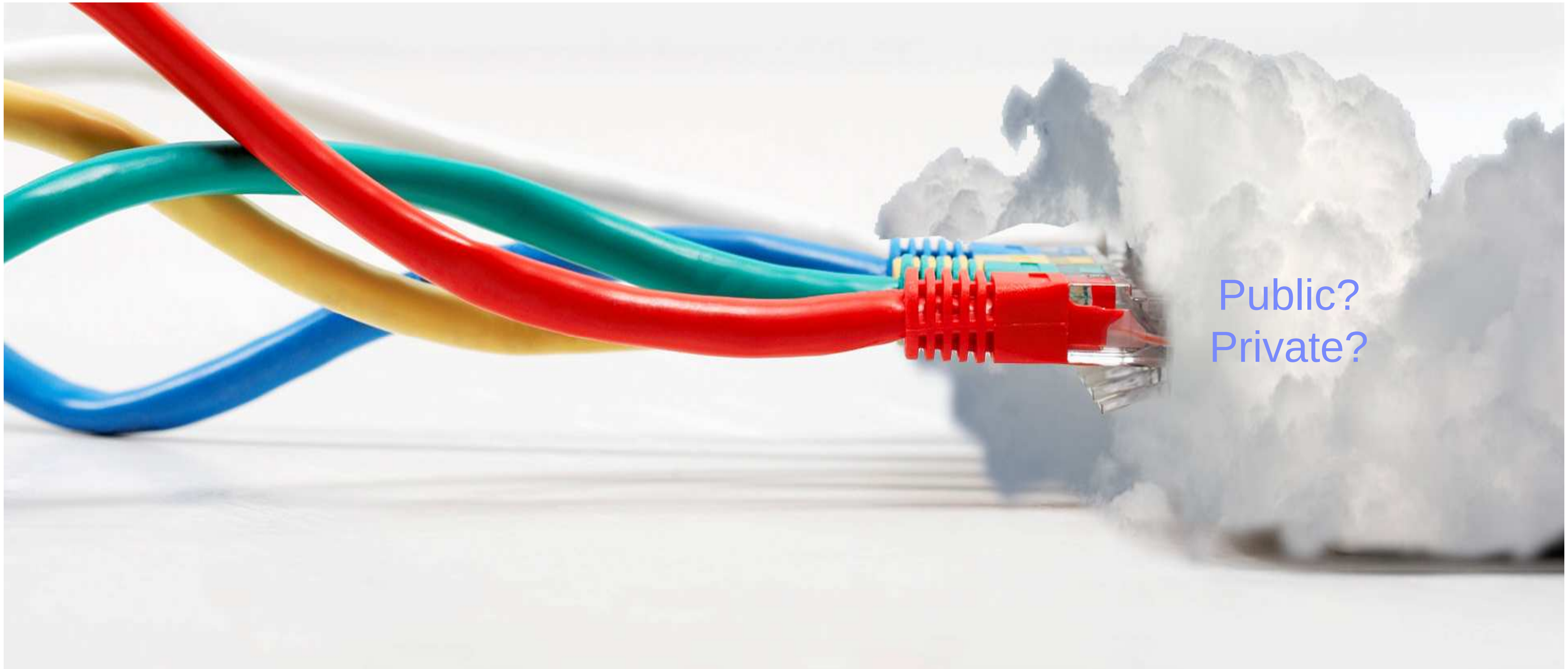
1 BASE CONCEPTS

2 IT SOURCING OPTIONS IMPLICATIONS OF CLOUD COMPUTING

3 DYNAMICS OF THE TRANSFORMATION

4 DYNAMICS OF THE TRANSFORMATION

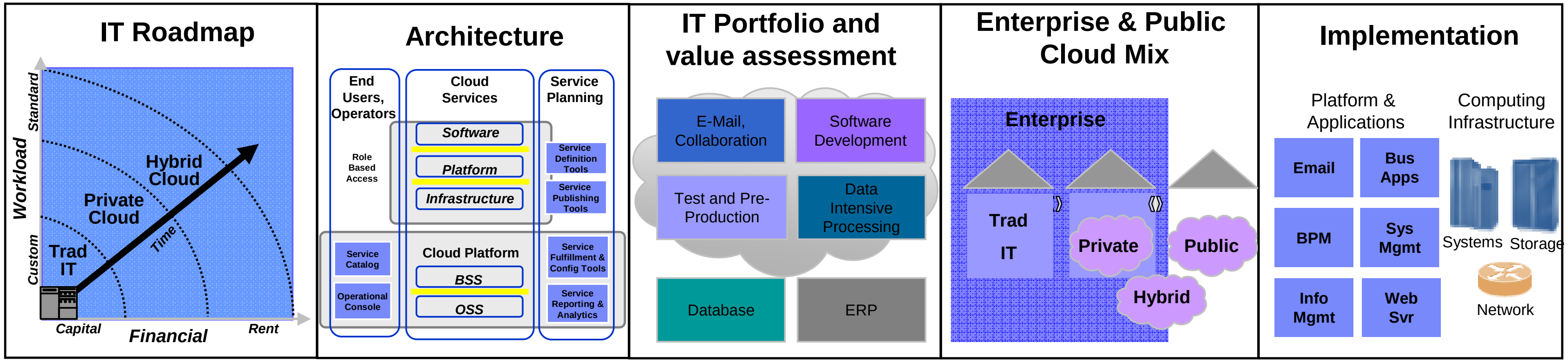
So, can a company just plug into the cloud?



Developing a Cloud Strategy and Plan

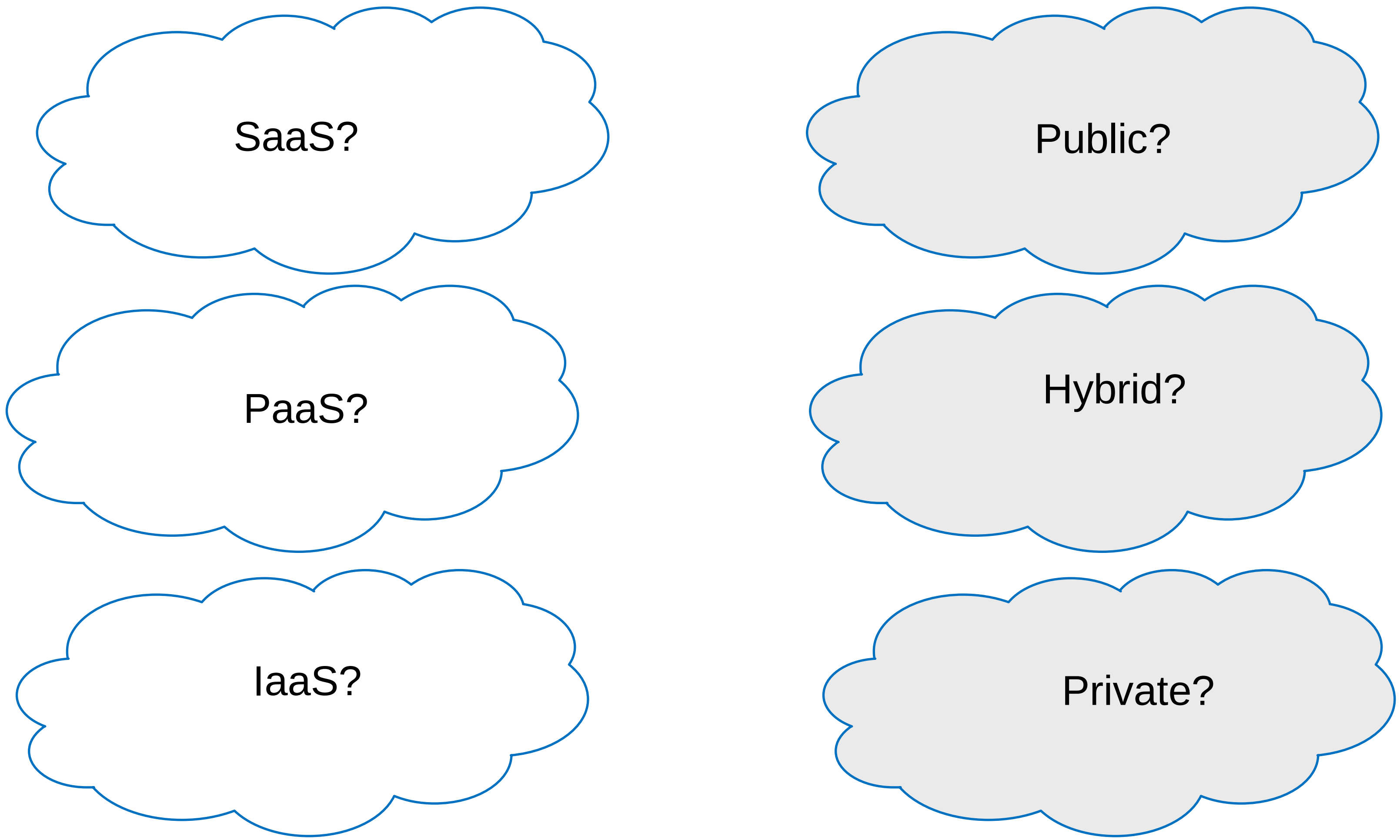
- **Without a strategy, Cloud computing can be a threat to the CIO and IT team**
  - Reduced control of IT services delivered over the Internet
  - Perceived cost gap between a cloud service and traditional IT
- **With a strategy, Cloud computing is a huge opportunity for the CIO**
  - Lower costs, more responsive IT, optimized delivery
  - Greater range of services and capabilities
  - Greater visibility in billing / chargeback to LOBs
  - Better control of the users' systems, desktops, and services access

5 Steps to Cloud



Basically, a cloud computing solution consists of

identifying the right cloud service(s) and the right deployment model(s) to help an enterprise achieve strategic business and IT objectives.



Delivery Models

Deployment Models

Step 1: Understand the IT Transformation Roadmap

Centralize

- Consolidate from many to fewer data centers
- Reduce infrastructure complexity
- Improve facilities management
- Improve operational cost

Consolidate

- Consolidate many servers into fewer physical resource boundaries
- Reduce system management complexity
- Reduce physical footprints

Virtualize

- Remove physical resource boundaries
- Increase hardware utilization
- Allocate less than physical boundary
- Simplify deployments
- Reduce software licensing costs
- Standardize virtual images

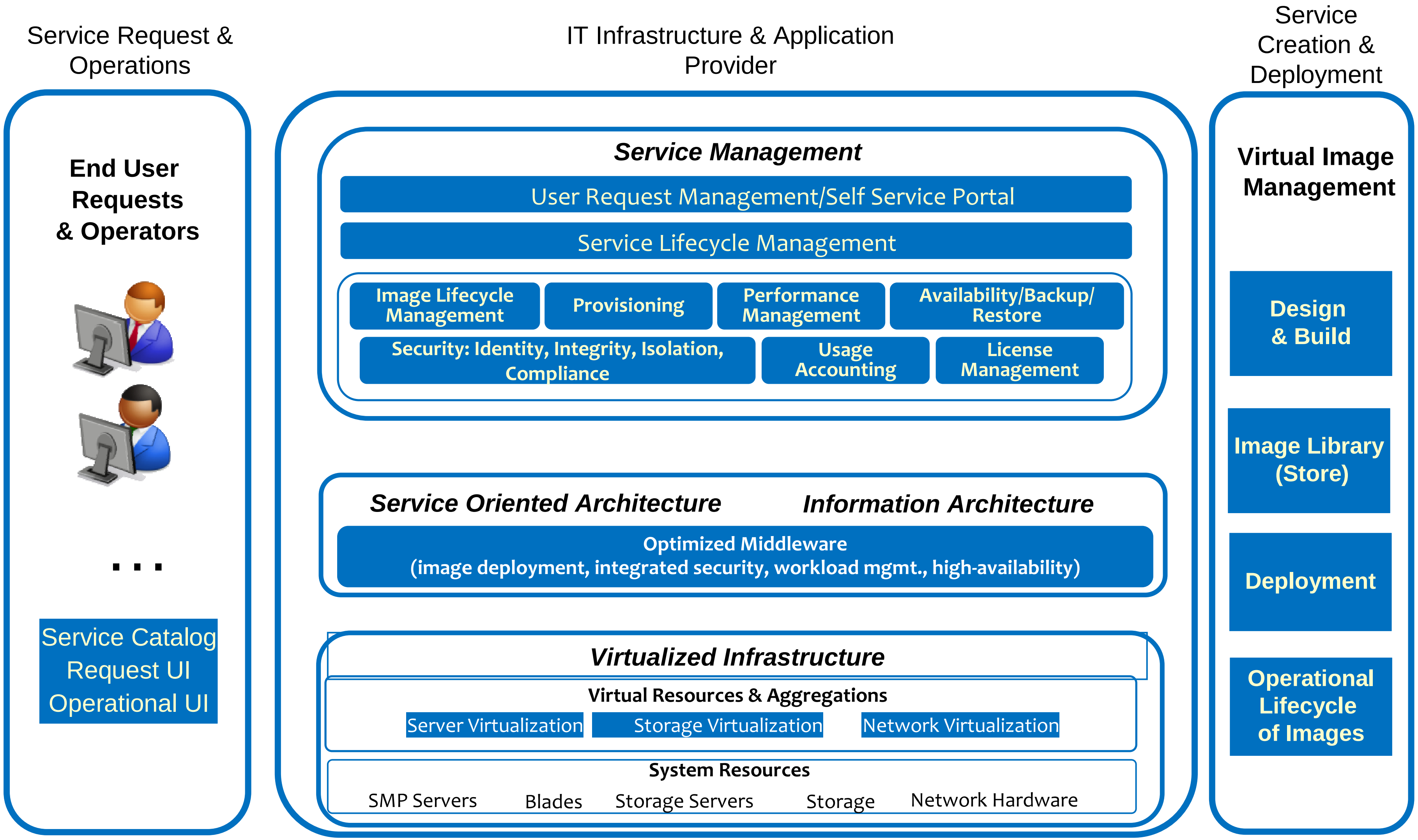
Automate

- Migrate many applications into fewer images
- Simplify IT environment
- Reduce operations resources
- Improve application specific monitoring and tuning
- Automate provisioning and de-provisioning of services and resources

Optimize

- Dynamically compose, develop, and deploy application services
- Manage virtualized infrastructure in response to demand
- Extend services delivery platform dynamically with external service providers
- Deliver services based on requested Qualities of Service and established SLAs

Step 2: Define an architectural model for cloud computing



Step 3: Analyze IT services workloads

EXAMINE FOR RISK

Database
Transaction processing
ERP workloads

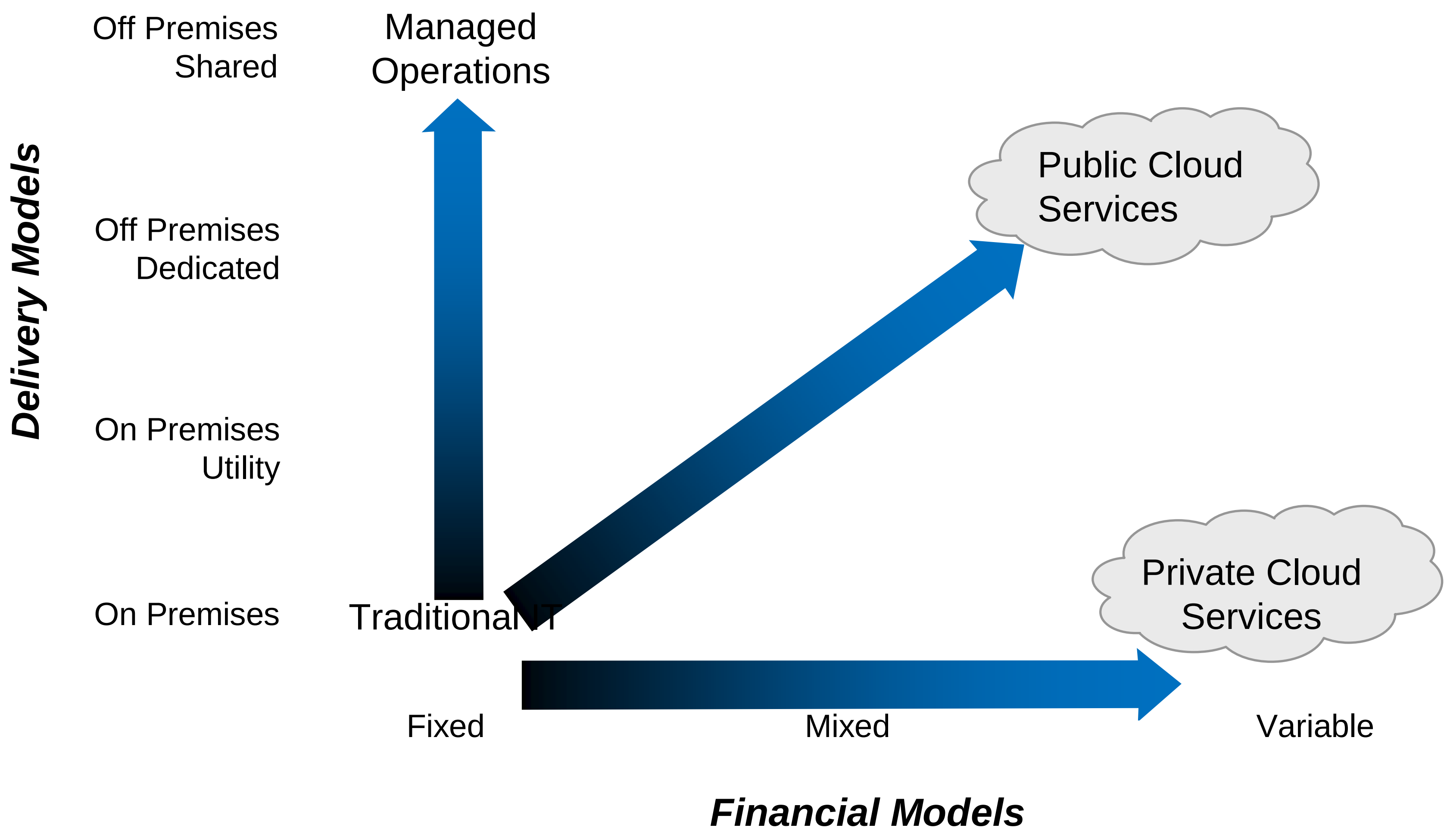
TEST FOR STANDARDIZATON

Web infrastructure applications
Collaborative infrastructure
Development and test
High Performance Computing

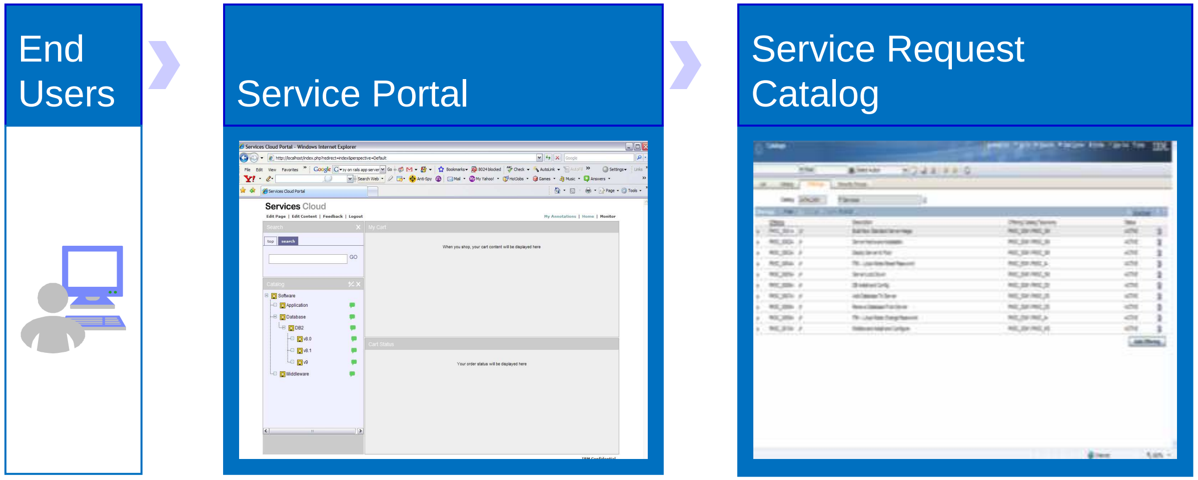
EXPLORE NEW WORKLOADS

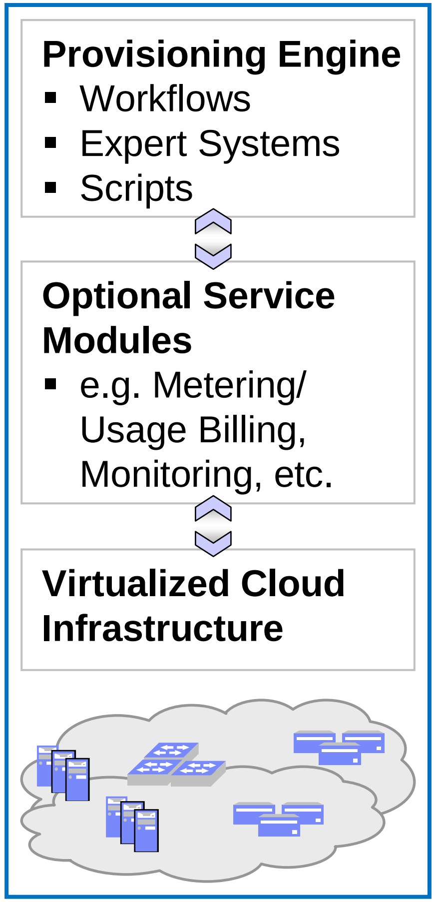
High volume, low cost analytics
Collaborative Business Networks
Industry scale “smart” applications

Step 4: Decide the right mix for your enterprise



Step 5: Implement the strategy and plan





- Easy to access, easy to use Service Request Catalog
- Hides underlying complex infrastructure from user and shifts focus to services provided
- Enables the ability to provide standardized and lower cost services
- Facilitates a granular level of services metering and billing
- Workload standardization eases complexity

Questions ?