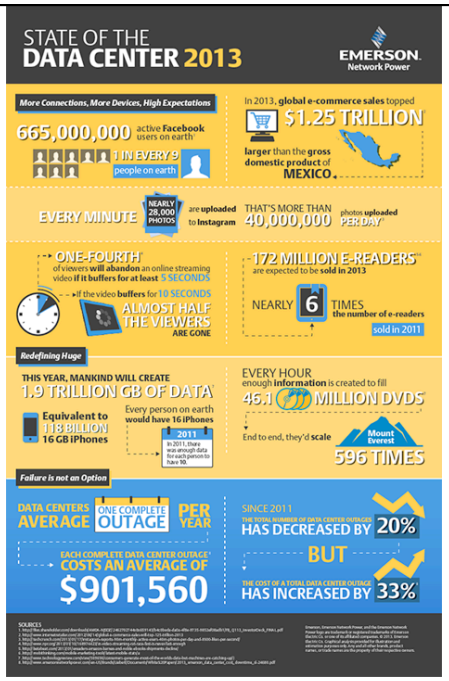
Today's Objectives

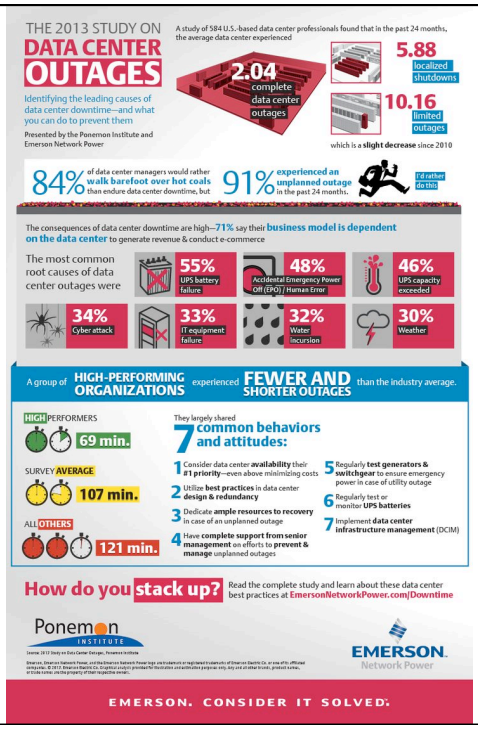
- Virtualization
- Cloud Computing
- Amazon Web Services

Data Center

- What did you think?


STATE OF THE
DATA CENTER 2013
EMERSON
Network Power
More Connections, More Devices, High Expectations
665,000,000 active Facebook users on earth
1 IN EVERY 9 people on earth
In 2013, global e-commerce sales topped
$1.25 TRILLION
larger than the gross domestic product of MEXICO.
EVERY MINUTE NEARLY 28,000 PHOTOS are uploaded to Instagram
THAT'S MORE THAN 40,000,000 photos uploaded PER DAY
ONE-FOURTH of viewers will abandon an online streaming video if it buffers for at least 5 SECONDS
If the video buffers for 10 SECONDS
ALMOST HALF THE VIEWERS ARE GONE
172 MILLION E-READERS are expected to be sold in 2013
NEARLY 6 TIMES the number of e-readers sold in 2011
Redefining Huge
THIS YEAR, MANKIND WILL CREATE
1.9 TRILLION GB OF DATA
Equivalent to 1.8 BILLION 16 GB IPhones
Every person on earth would have 16 IPhones
2011 In 2011, there was enough data for each person to have 16.
EVERY HOUR enough information is created to fill 45.1 MILLION DVDS
End to end, they'd scale Mount Everest 596 TIMES
Failure Is not an Option
DATA CENTERS AVERAGE ONE COMPLETE OUTAGE PER YEAR
EACH COMPLETE DATA CENTER OUTAGE COSTS AN AVERAGE OF
$901,560
SINCE 2011 THE TOTAL NUMBER OF DATA CENTER OUTAGES HAS DECREASED BY 20%
BUT
THE COST OF A TOTAL DATA CENTER OUTAGE HAS INCREASED BY 33%

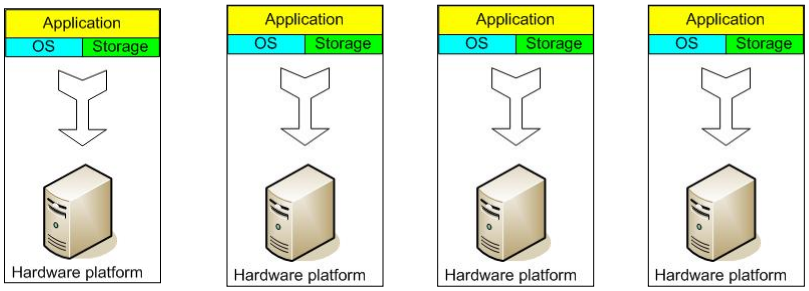

EMERSON
STATE OF THE
DATA CENTER 2011
The growth – and growing importance – of the data center
Fueling the Internet
$53 BILLION
$53 Billion in cyber weekend sales is larger than the entire economy of Bulgaria
BULGARIA
1 in every 12 people on earth
500,000,000
Active Facebook users on earth
EVERY SECOND 1,157 PEOPLE START WATCHING YOU TUBE
THATS 1,000,000 VIDEOS A DAY
TWITTER
IN FEBRUARY 2011, DAILY TWEETS AVERAGED 140 MILLION
ALMOST 3X 50 MILLION TWEETS
How big is it?
= 285,831,541 sq. ft.
THAT'S
509,147 DATA CENTERS WORLDWIDE
5,955 FOOTBALL FIELDS
THIS YEAR, MANKIND WILL CREATE 1.2 trillion GB OF DATA
Equivalent to 75 billion 16 GB iPods
EVERY PERSON ON EARTH WOULD HAVE 16 iPODS
EVERY HOUR enough information is consumed by Internet traffic to fill 7million DVDs. Side by side they'd scale Mount Everest 95 times.
95 TIMES
Too big to fail
509,147 data centers worldwide
X2.5 COMPLETE OUTAGES (average in US)
134 minutes
(average outage duration)
= 2,842,737 HOURS OF DOWNTIME AT AN AVERAGE COST OF $300,000 PER HOUR
For a total loss of $426 BILLION A YEAR, OR ENOUGH TO BUY EVERY PERSON IN MUNICH, GERMANY, A YACHT
40'
More with less
A SERVER PURCHASED IN 2011 HAS AN AVERAGE 45X MORE COMPUTE CAPACITY THAN A SIMILARLY CONFIGURED SERVER INSTALLED IN 2001.
2001
2011



Tiny Bookstore

- Due tomorrow night

The Traditional Server Concept

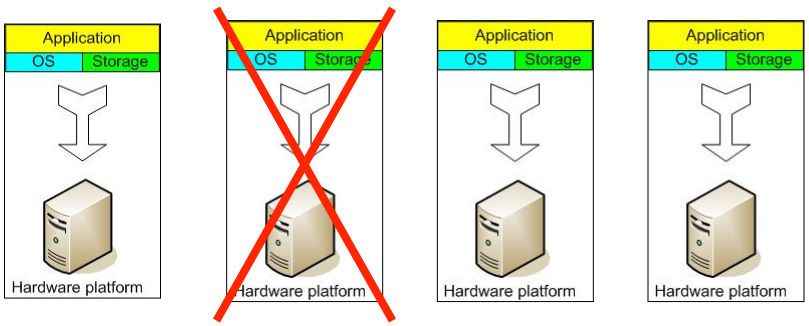


Web Server	App Server	DB Server	Email
Windows	Linux	Linux	Windows
IIS	Glassfish	MySQL	Exchange

The Traditional Server Concept

- System Administrators often talk about servers as a whole unit that includes the hardware, the OS, the storage, and the applications.
- Servers are often referred to by their function
  - Web server, SQL server, File server, etc.
- If the File server fills up, or the Web server becomes overtaxed, then sys admin must add a new server

And if something goes wrong ...



Web Server	App Server	DB Server	EMail
Windows	DOWN!	Linux	Windows
IIS		MySQL	Exchange

The Traditional Server Concept

- Unless there are multiple servers, if a service experiences a hardware failure, then the service is down.
- System Admins can implement clusters of servers to make them more fault tolerant.
- However, even clusters have limits on their scalability, and not all applications work in a clustered environment.

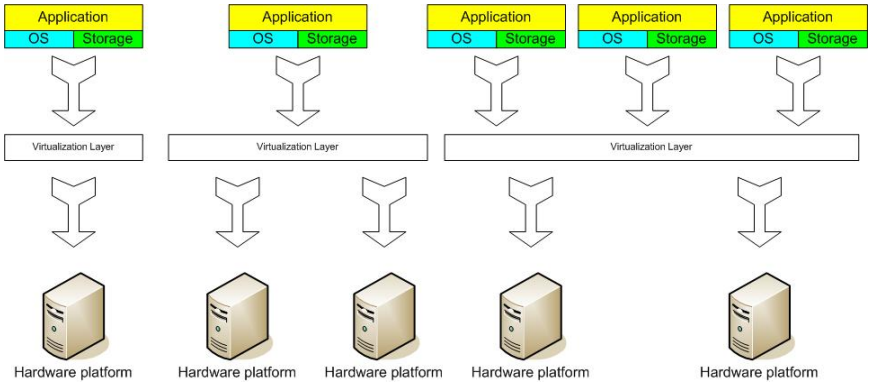
The Traditional Server Concept

- Pros
  - Easy to conceptualize
  - Fairly easy to deploy
  - Easy to backup
  - Virtually any application/ service can be run from this type of setup
- Cons
  - Expensive to acquire and maintain hardware
  - Not very scalable
  - Difficult to replicate
  - Redundancy is difficult to implement
  - Vulnerable to hardware outages
  - In many cases, processor is under-utilized

Historical Trends: Virtualization

- ***Virtualization***: The ability to run multiple operating systems on a single physical system and share the underlying hardware resources
- 1960s – OS-VM, VM-360 – Used to split mainframes into logical partitions.
- 1998 – VMWare – First practical implementation on X86 but at significant performance hit
- 2003 – Xen paravirtualization
  - Efficient, lightweight – any hardware
  - BUT requires kernel support
- Late 2000s – Intel and AMD add hardware support for virtualization.

The Virtual Server Concept



Virtualization layer between *Guest OS* and hardware

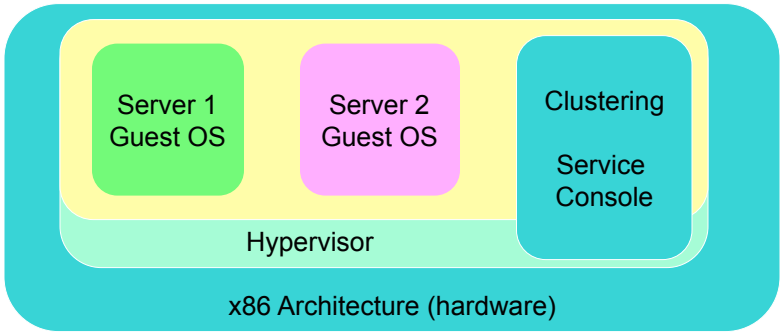
The Virtual Server Concept

- Virtual servers seek to encapsulate the server software (OS, applications, and storage) away from the hardware
- Servers end up as mere files stored on a physical box or in enterprise storage.
- One host typically houses many virtual servers (virtual machines or VMs)
- A virtual server can be serviced by one or more hosts, e.g. storage, services, etc.

Hypervisors or Virtual Machine Monitor

- A piece of computer software, firmware, or hardware that creates and runs virtual machines.
- ***Host*** machine: A computer on which a hypervisor is running one or more virtual machines
- Each virtual machine has a ***guest operating system***, which is managed by the hypervisor.
- Multiple instances of a variety of operating systems may share the virtualized hardware resources.

Hypervisors and Virtual Machines



Hypervisor: intercepts hardware requests

The Virtual Server Concept

- Virtual servers can still be referred to by their function (e.g., web, file, …)
- If the environment is built correctly, virtual servers will not be affected by the loss of a host
- Hosts may be removed and introduced almost at will to accommodate maintenance.

The Virtual Server Concept

- Virtual servers can be scaled out easily
  - If admins find that the resources supporting a virtual server are being taxed too much, they can adjust the amount of resources allocated to that virtual server
- Server templates can be created in a virtual environment to be used to create multiple, identical virtual servers
- Virtual servers themselves can be migrated from host to host almost at will.

The Virtual Server Concept

- Pros
  - Resource pooling
  - Highly redundant
  - Highly available
  - Rapidly deploy new servers
  - Easy to deploy
  - Reconfigurable while services are running
  - Optimizes physical resources by doing more with less
- Cons
  - Slightly harder to conceptualize
  - Slightly more costly (must buy hardware, OS, Apps, **and** abstraction layer)

Virtual-* Allows for the Scale of Abstraction to Increase Over Time

- Run one process within certain resource limits.
  - Op Sys has virtual memory, virtual CPU, and virtual storage (file system).
- Run multiple processes within certain resource limits.
  - Resource containers (Solaris), virtual servers (Linux), virtual images (Docker)
- Run an entire operating system within certain limits.
  - Virtual machine technology: VMWare, Xen, KVM, etc.
- Run a set of virtual machines connected via a private network.
  - Virtual networks (SDNs) provision bandwidth between virtual machines.
- Run a private virtual architecture for every customer.
  - Automated tools replicate virtual infrastructure as needed.

CLOUD COMPUTING

Historical Trends

- Shared Utility Computing
  - 1960s – MULTICS – Concept of a Shared Computing Utility
  - 1970s – IBM Mainframes – rent by the CPU-hour. (Fast/ slow switch.)
- Data Center Co-location
  - 1990s-2000s – Rent machines for months/years, keep them close to the network access point and pay a flat rate. Avoid running your own building with utilities!
- Pay as You Go: Utility computing
  - Early 2000s - Submit jobs to a remote service provider where they run on the raw hardware. Sun Cloud ($1/ CPU-hour, Solaris +SGE), IBM Deep Capacity Computing on Demand (50 cents/hour)

Cloud Computing

- Cloud Computing: general term used to describe the class of network-based computing that takes place over the Internet,
  - basically a step on from Utility Computing
  - a collection/group of shared, integrated, and networked hardware, software and Internet infrastructure (called a platform).
  - Using the Internet for communication and transport provides hardware, software and networking services to clients
- Platforms hide complexity and details of underlying infrastructure from users and applications
  - Provide graphical interface or API (Applications Programming Interface)
- Resources are provided to computers and other devices on demand – pay per use.

Basic Cloud Characteristics

- Cost-effective means of virtualizing and making use of resources more effectively
  - Low start-up costs – pay for use helps to kick-start companies
- "no-need-to-know" in terms of the underlying details of infrastructure
  - applications interface with the infrastructure via the APIs.
- "flexibility and elasticity" allows these systems to scale up and down at will
  - utilizing the resources of all kinds
    - CPU, storage, server capacity, load balancing, and databases
  - Scaling is proportional to demand (revenue) so it's a good business model
- "pay as much as used and needed" type of utility computing and the "always on, anywhere and any place" type of network-based computing.

Basic Cloud Characteristics

- **Remotely hosted**: Services or data are hosted on remote infrastructure.
- **Ubiquitous**: Services or data are available from anywhere.
- **Transparent** to users and applications → can be built in multiple ways
  - branded products, proprietary open source, hardware or software, or just off-the-shelf PCs.
  - In general, built on clusters of PC servers and off-the-shelf components plus Open Source software combined with in-house applications and/or system software
- **Commodified**: utility computing model similar to traditional utilities, like gas and electricity
  - you pay for what you would want!

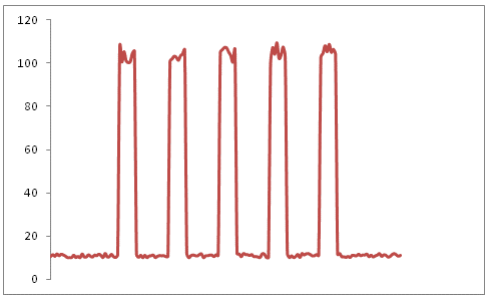
Cloud Computing Use

- Vast range of Cloud Computing applications
  - Virtual private servers, Web hosting, data servers, fail-over services, etc.

Motivation Example: Forbes.com

- You offer on-line real time stock market data
- Why pay for capacity weekends, overnight?

Rate of Server Accesses



9 AM - 5 PM, M-F

ALL OTHER TIMES

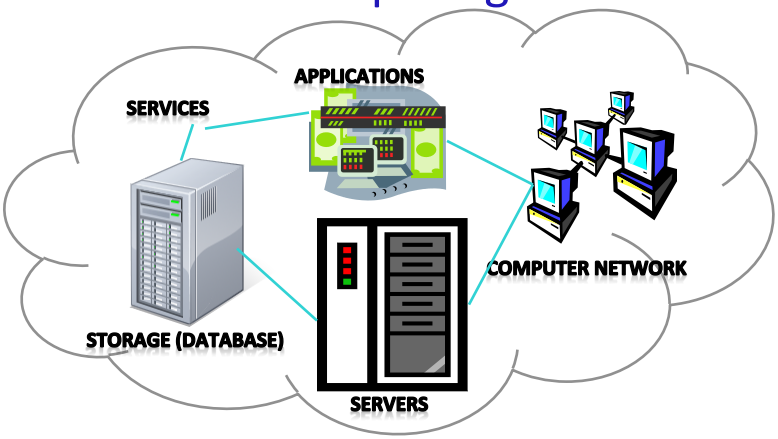
Forbes' Solution

- Host the web site in Amazon's EC2 Elastic Compute Cloud
- Provision new servers every day, and deprovision them every night
- Pay just $0.10* per server per hour
  - * more for higher capacity servers
- Let Amazon worry about the hardware!

Cloud computing: Virtualization

- You don't have to own the hardware
- You "rent" it as needed from a cloud
- There are public clouds
  - e.g. Amazon EC2, and now many others (Google, Microsoft, IBM, Sun, and others ...)
- A company can create a private one
  - With more control over security, etc.

What is Cloud Computing



- Shared pool of configurable computing resources
- On-demand network access
- Provisioned by the Service Provider

Adopted from: Effectively and Securely Using the Cloud Computing Paradigm by Peter Mell, Tim Grance

Cloud Computing Characteristics

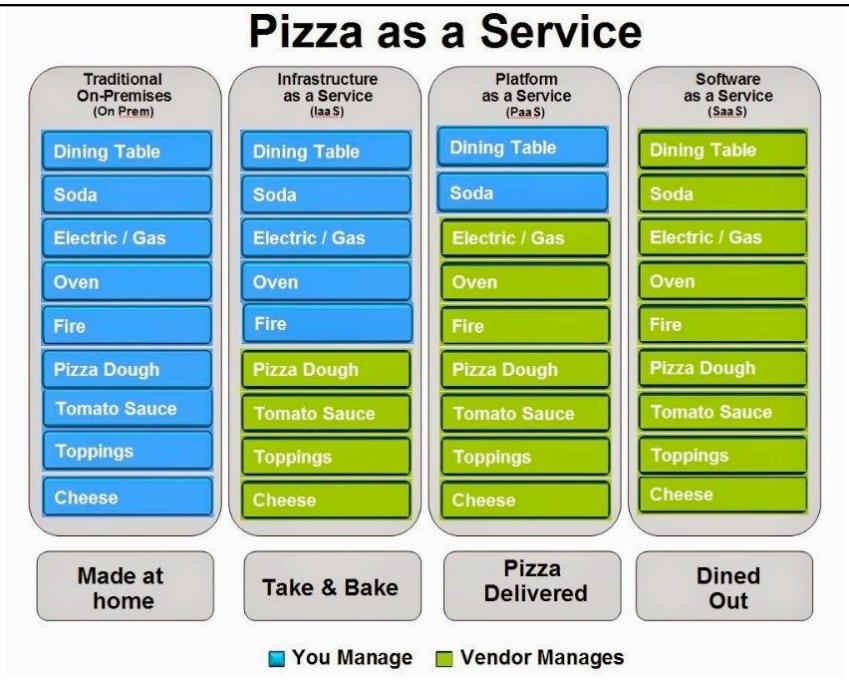
Common Characteristics:

- Massive Scale
- Resilient Computing
- Homogeneity
- Geographic Distribution
- Virtualization
- Service Orientation
- Low-Cost Software
- Advanced Security

Essential Characteristics:

- On Demand Self-Service
- Broad Network Access
- Rapid Elasticity
- Resource Pooling
- Measured Service

Adopted from: Effectively and Securely Using the Cloud Computing Paradigm by Peter Mell, Tim Grance



SaaS and PaaS

- **SaaS: Software as a Service**
  - an application is hosted as a service provided to customers across the Internet
  - Saas alleviates the burden of software maintenance/ support
  - but users relinquish control over software versions and requirements
- **PaaS: Platform as a Service**
  - provides a computing platform and a solution stack as a service
  - Consumer creates the software using tools and/or libraries from the provider
  - Consumer controls software deployment and configuration settings.
  - Provider provides the networks, servers, storage and other services

IaaS: Infrastructure as a Service

- IaaS providers offer virtual machines, virtual-machine image libraris, raw (block) and file-based storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles.
- Pools of hypervisors can scale services up and down according to customers' varying requirements
- All infrastructure is provided on-demand

Looking Ahead

- Map Reduce – tonight
- Tiny Bookstore – tomorrow