Cloud Computing & Storage

More than a File System in the Sky

Erik Riedel, PhD
Technology & Architecture
Cloud Infrastructure Group
EMC

EMC²

EMC²®

iomega
an EMC company

Search

HOME | PRODUCTS | SUPPORT | United States | English

About | Headlines | Where to Buy | Partners | Register | Subscribe

Products

- Desktop Hard Drives
- Portable Hard Drives
- Multimedia Drives
- Home Network
- Network Storage
- REV Removable
- Zip / Floppy
- Software
- Online Storage
- Data Recovery
- Factory Outlet

- Solutions
- Product Selector

Home Users | Office Users | Creative Professionals

Take your PC
Virtually Anywhere

A Star is Born!
ScreenPlay Director

New eGo USB
with Protection Suite

eGo Gold Edition
Spring Sweepstakes

mozy

Products | Sign Up | About | Support | Log In

It's more than a computer. It's your life.

View the Mozy intro video

Every photo.
Every document.
Every song.
Safe with Mozy.

When you think about it, all the important information in your life is now stored on a computer. Whether it's photos and music or business documents and financial records, everything is digital. With Mozy, you can be sure your digital life will always be there when you need it.

Backing up a home computer or an entire office?

Mozy has a plan just your size.

Learn More

Sign up for Mozy

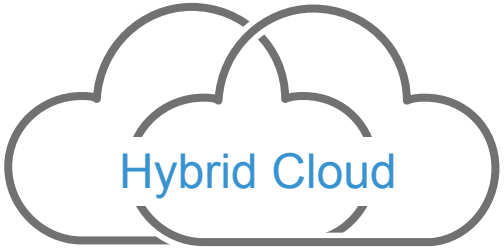
Cloud Computing

Courtesy – wordle.net

Supporting the Shift to Cloud Inside, Outside, and Across Organizations

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

Infrastructure deployed and operated exclusively for an organization or enterprise



Composition of two or more clouds, private and/or public

Infrastructure made available to general public or many industry groups/customers

Source: *National Institute of Standards and Technology, V15 October 2009

EMC²

Big Data

IN 2010 THE DIGITAL UNIVERSE WAS

1.2 ZETTABYTES

1,200,000,000,000,000,000,000

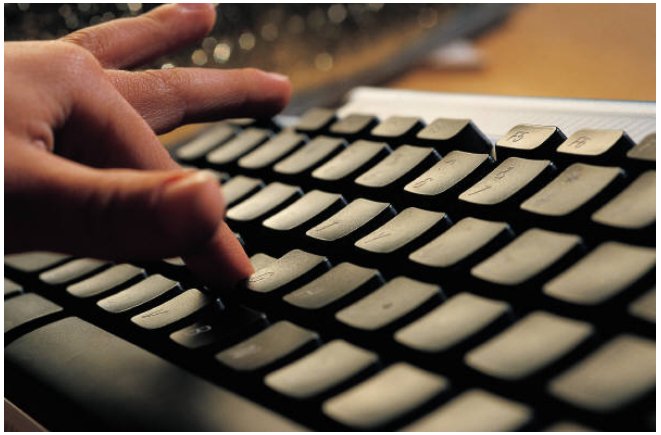
+ 600 million disk drives sold in 2011 (so another 1.2 ZB !)

Source: 2010 IDC Digital Universe Study

EMC²

Who Is It Really For

vs.



IT Managers

Programmers

Programmers buzz – Ruby/Rails, MapReduce/Hadoop

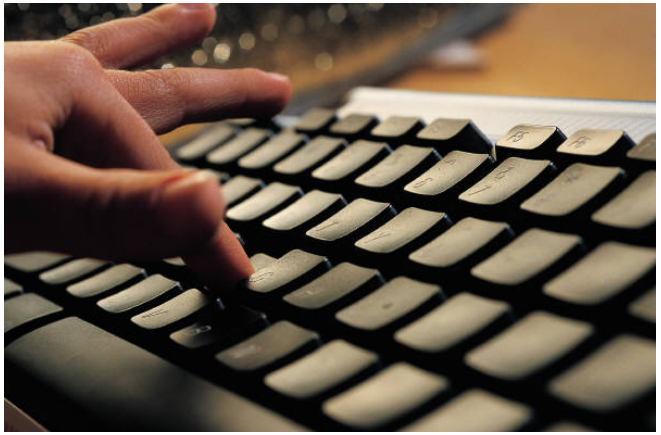
IT Managers buzz – VM images, vApps, VLANs

Marketing buzz – Virtualization, IaaS, PaaS, SaaS

The previously separate roles of software developer and operations have [become] increasingly intermeshed and intertwined. Things are materially different…

Ray Ozzie, Chief Software Architect, Microsoft

+



IT Managers

Programmers

Programmers buzz – Ruby/Rails, MapReduce/Hadoop

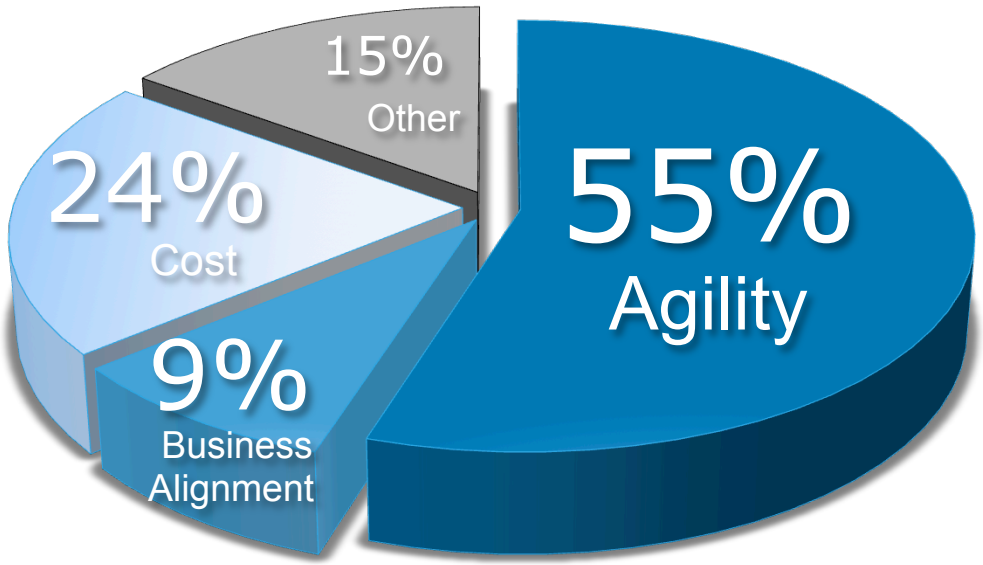
IT Managers buzz – VM images, vApps, VLANs

Marketing buzz – Virtualization, IaaS, PaaS, SaaS

Cloud is often an “excuse” for enterprises to move to “New IT” – away from the old client/server model that has been used for the past ten years [toward Web 2.0 IT]

Werner Vogels, CTO, Amazon

Agility Is The #1 Private Cloud Driver



"The majority see agility and speed as the primary benefits of private cloud computing."

GARTNER

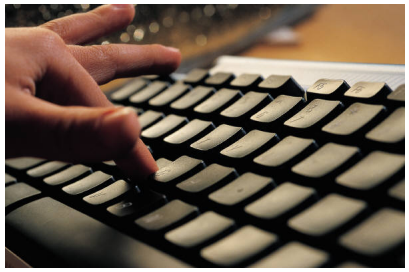
Source: "The Drivers And Challenges Of Private Cloud Computing", March 2011, Gartner

EMC²

A Few Details

It’s not possible to “start over” and re-write all applications using scale-out design patterns in the first few months of a cloud deployment, but it is possible to adapt many legacy applications with the help of virtualization, so cloud infrastructure can support and enable both development models, including mixing the two.

“Developers” Range Widely in Focus/Expertise



- **IT managers/admins** deploying applications encapsulated or pre-packaged into virtual machines
  - Language – configuration scripts, command lines
  - Input – catalog of vApp templates or pre-configured VMs
  - Output – VM images, VM configurations, system configurations
  - Runs on – vSphere/ESX, virtual networks, legacy storage + scale-out storage

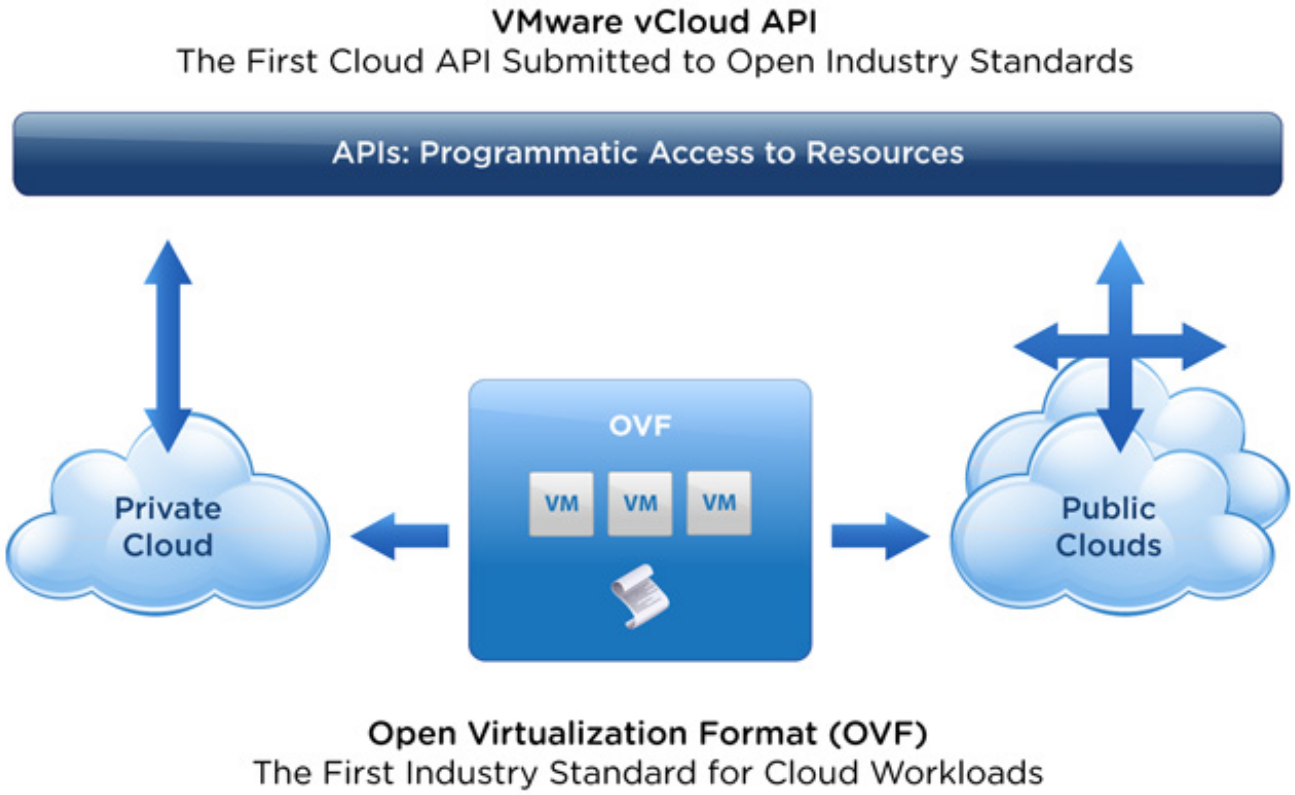
- **Programmers** using application frameworks such as Groovy/Grails or Hadoop
  - Language – Grails/Java, MapReduce/Hadoop
  - Input – code, with help of an IDE
  - Output – Rails + database configurations, job scripts
  - Runs on – Rails + MySQL, virtual networks, scale-out storage

Apps + Data

- **Development**
  - new applications
  - explicitly scale-out (e.g. MapReduce, Hadoop)
  - built on higher-level frameworks (e.g. Ruby/Rails, Azure)
- **Deployment**
  - legacy applications
  - “packaged” into virtual machine containers
  - easy to replicate and migrate across virtual infrastructure
- **Data**
  - shared corporate data is the common ground (enterprise apps)
  - consumer value centered around their personal data (consumer apps)

EMC²

Example – Deployment



Marketing buzz – IaaS – Infrastructure as a Service

Example – Development

Marketing buzz – PaaS – Platform as a Service

Example – EMC Greenplum HD

Enterprise-Ready Hadoop Platform For Unstructured Data

- Addresses The Growth Of Unstructured Data
- More Reliable For The Enterprise
- Easier To Use With Existing Systems And Tools

COMMUNITY EDITION

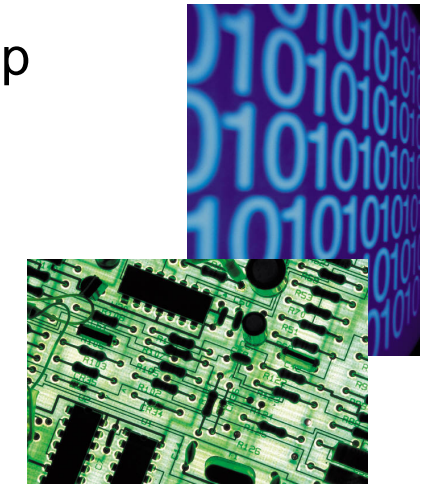
ENTERPRISE-EDITION

Marketing buzz – Big Data – MapReduce, Hadoop

EMC²

More About Apps + Data

- From the perspective of development & deployment, the key new technology component is a combined data + app (storage + compute) platform where apps are created, deployed, monitored & managed with a common set of tools.



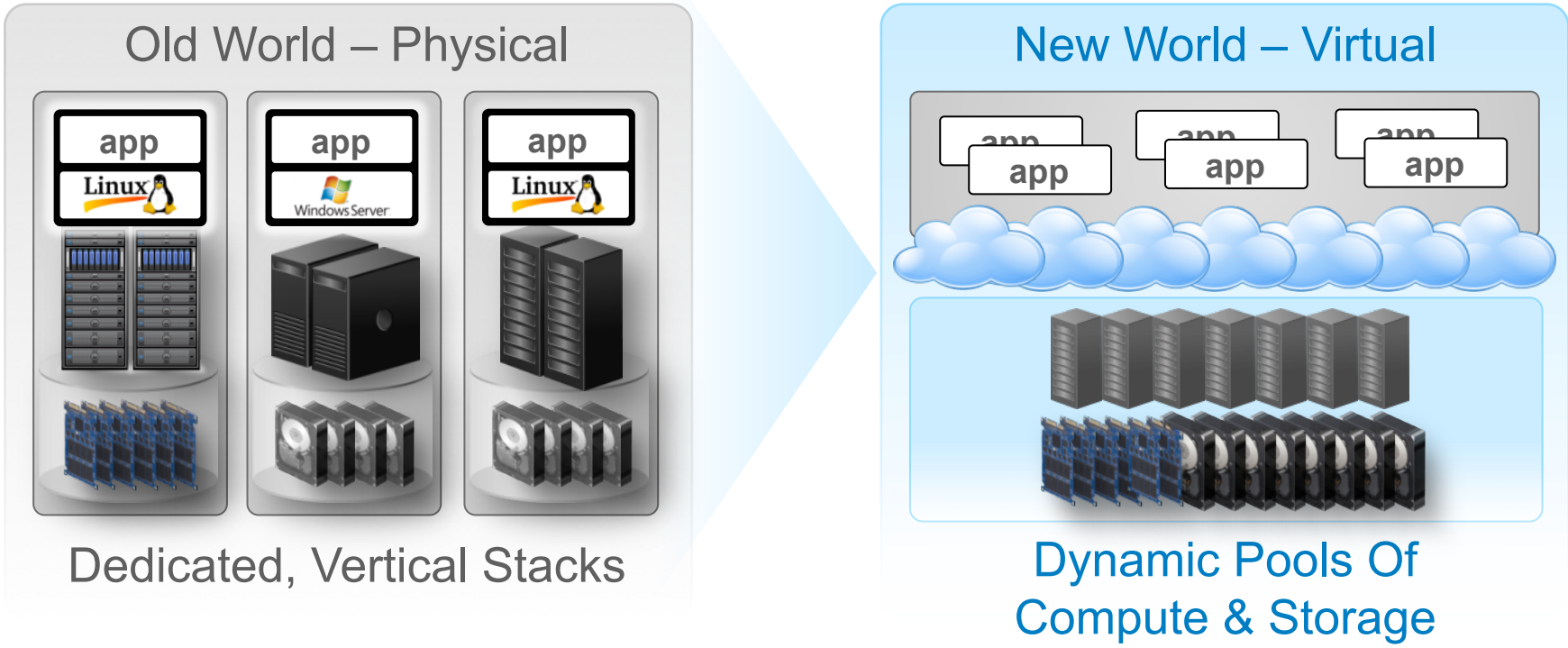
  - *Underlying enablers:*
- Common object space – apps, configs, user data
- Single identity store – public, private, enterprise, consumer
- Federation (public + private) – seamless across infrastructures

- Monitoring – continuous measurement to optimize (and generate bills)

Under The Covers

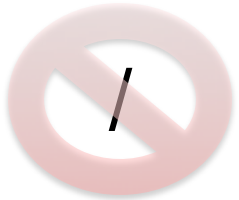
What About The Data?

Cloud – A New Architecture



Operating Systems & Frameworks "disappear" into the cloud fabric

Builds on 20 Years of Storage Research



- APIs vs. mount points – “no slashes required”
  - blocks vs. files vs. objects vs. “APIs”

- App-driven and policy-automated
  - self-configuring, self-organizing, self-tuning, self-*

- Built in data services
  - self-healing
- Unlimited namespace, dynamic
  - billions and billions of objects, large and small
- Native multi-tenancy
  - security/auth, monitoring, resource isolation

EMC²

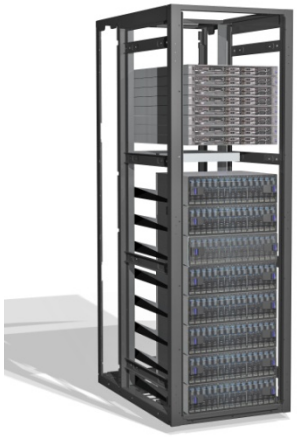
EMC Atmos

BIG. SMART. ELASTIC.

EMC²

Atmos Gen 2 Hardware Configurations

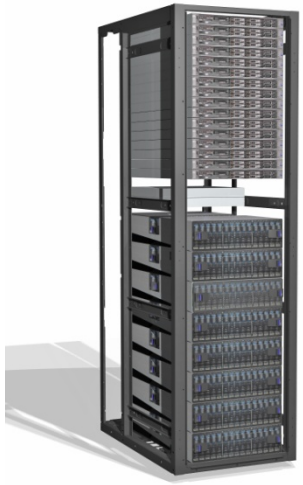
WS2-120



Small/Medium Scale Compute

- 1:15 servers-disk ratio
- 60 – 120 disks
- 60 TB entry point / 240 TB max
- 2 Node / 30 Disk expansions
- (7.2K) PWR: 5KW / 16.7 BTU/Hr

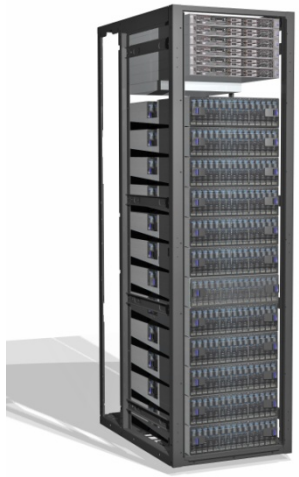
WS2-240



Dense Compute

- 1:15 servers-disk ratio
- 60 – 240 disks
- Up to 480 TB total capacity
- 2 Node / 30 Disk expansions
- 7.2K PWR: 10kW HEAT: 34.2k BTU/hr

WS2-360



Capacity

- 1:60 server-disk ratio
- 240 or 360 disks
- Up to 720 TB total capacity
- 2 Node / 120 Disk expansions
- 7.2k PWR: 10.3 kW HEAT: 35.1k BTU/hr

Intel® Xeon® 5500 “Nehalem” Micro architecture

1 / 2 / 3 TB SATA Drives

Expansion Nodes (2 Nodes per expansion)

- commodity SATA drives (as many as possible)
- x86 servers/controllers (as few as possible)
- SAS backplanes/cables (just the right number)

8.6 drives/U

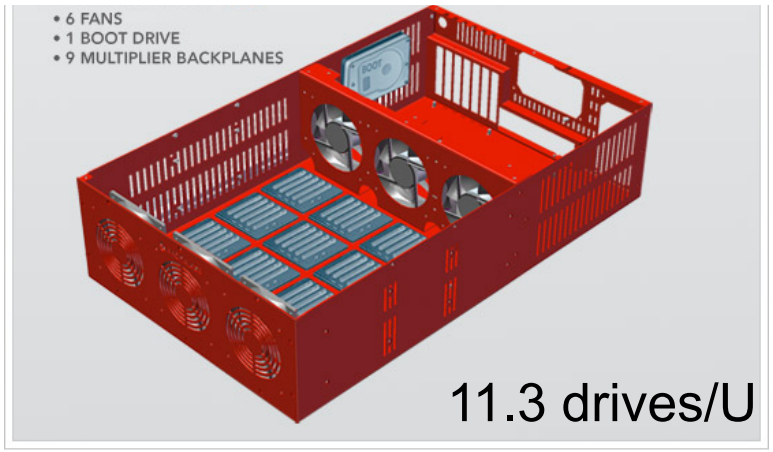
SGI® CloudRack™ C2

12 drives/U

Dell

6 drives/U

Backblaze

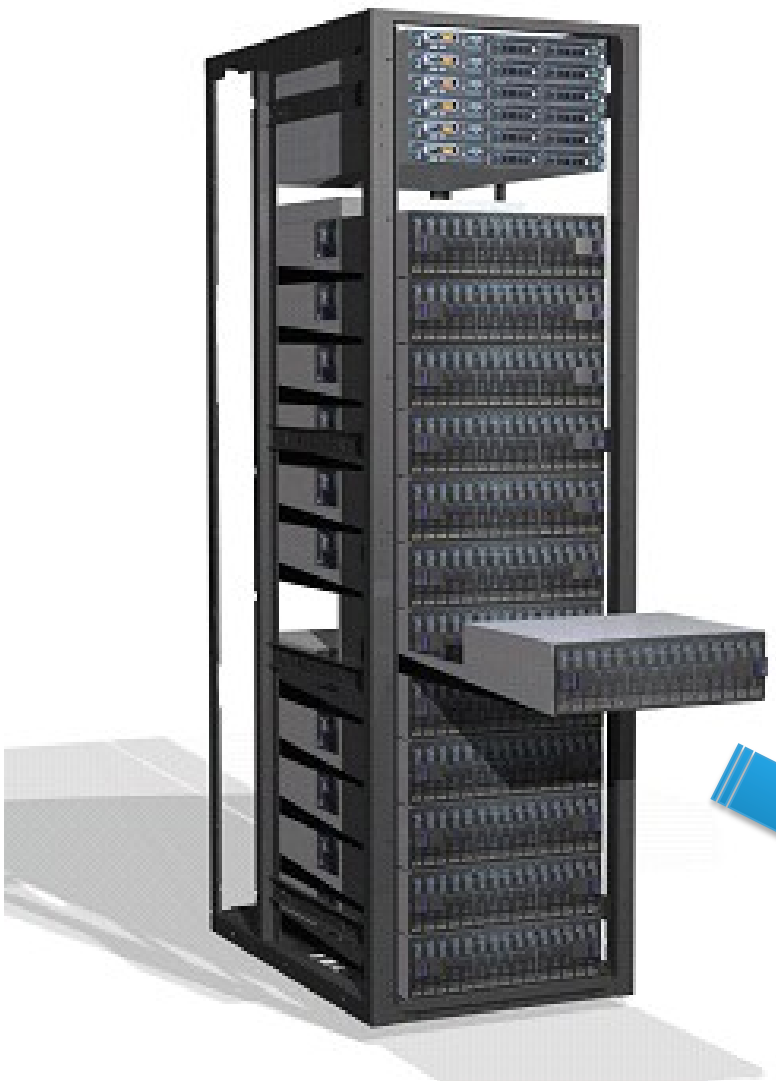


11.3 drives/U

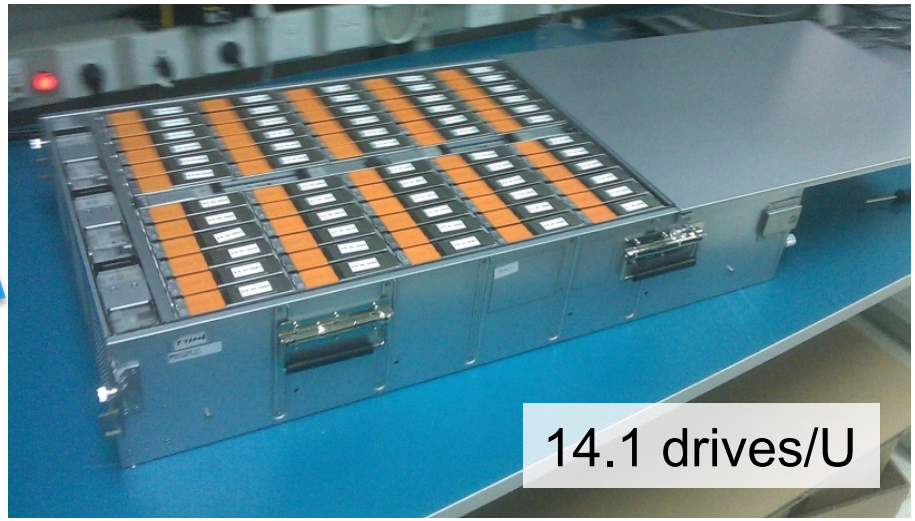
Supermicro

11.3 drives/U

EMC²

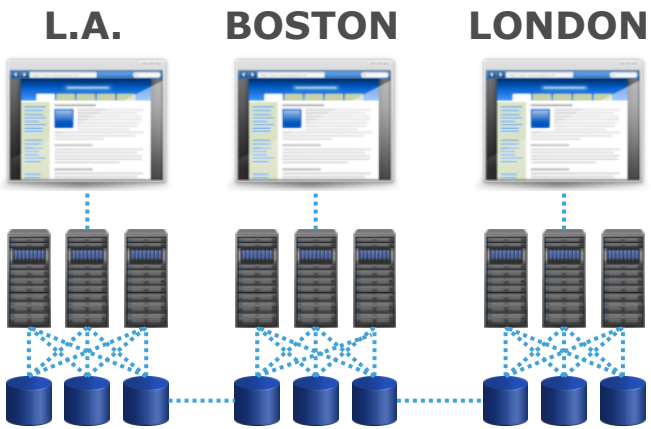


- commodity SATA drives (as many as possible)
- x86 servers/controllers (as few as possible)
- SAS backplanes/cables (just the right number)



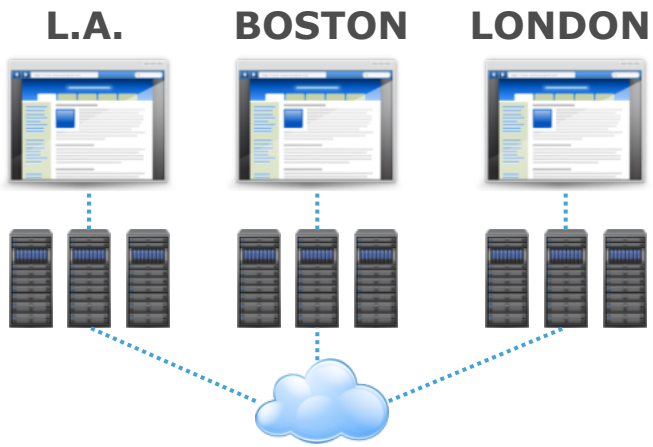
EMC²

A New Approach For Distributed Big Data



Storage Islands

- Disparate Systems
- Manual Administration
- One Tenant, Many Systems
- IT Provisioned Storage

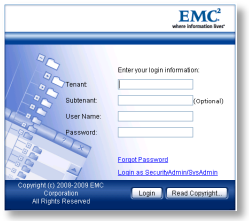


Single Storage Pool

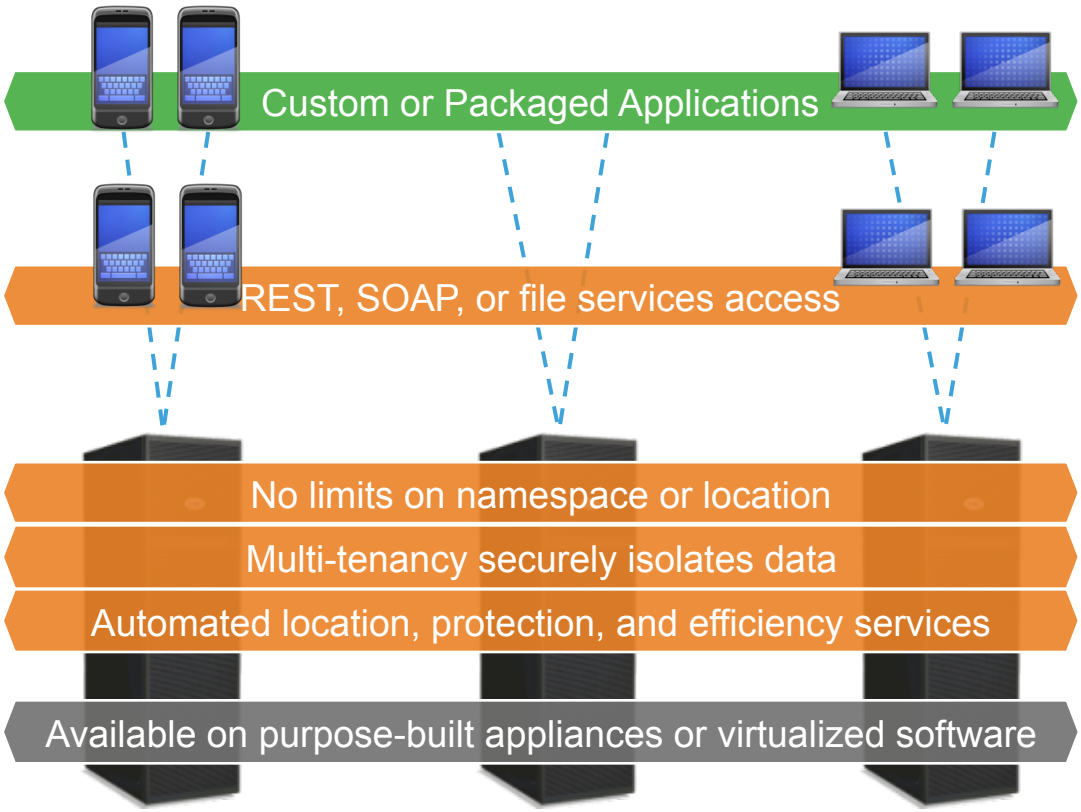
- Single System Across Locations
- Automated Policies
- Many Tenants One System
- Self-Service Access

What is EMC Atmos?

EMC Atmos



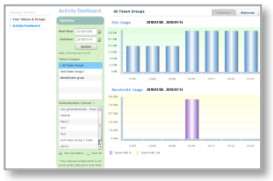
Single GUI



SITE #1
Los Angeles

SITE #2
New York

SITE #3
London

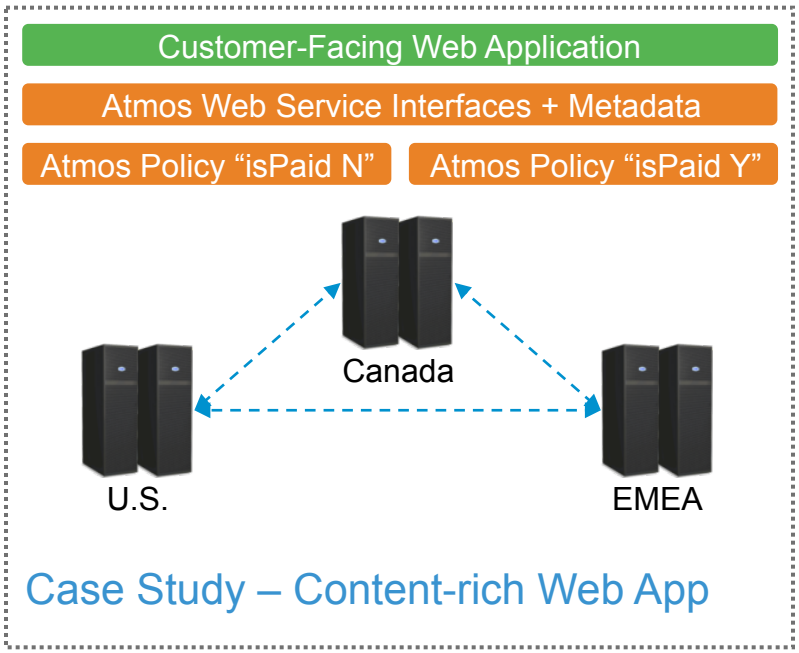


Self-Service Experience

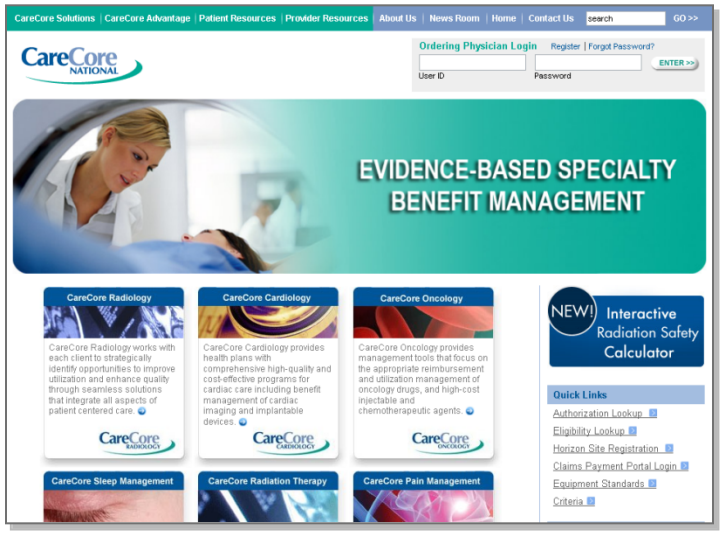
EMC²

Case Studies

- Petabyte-scale
- Geographic distribution
- Policy-driven storage



Case Study – Content-rich Web App



Content-Rich Web App on Atmos

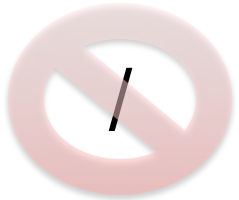
- Global distribution, content mix
- Multi-tenancy, scale to multiple sites
- Policy supports business models

CareCore “gets in the cloud” with Atmos

- Wrote to Atmos REST API in one week
- Bought Atmos and deployed in three weeks
- Adding over 2 million objects a day to Atmos
- Started with one app, spreading to many more

EMC²

Builds on 20 Years of Storage Research



- APIs vs. mount points – “no slashes required”
  - blocks vs. files vs. objects vs. “APIs”

- App-driven and policy-automated
  - self-configuring, self-organizing, self-tuning, self-*

- Built in data services
  - self-healing
- Unlimited namespace, dynamic
  - billions and billions of objects, large and small
- Native multi-tenancy
  - security/auth, monitoring, resource isolation

EMC²

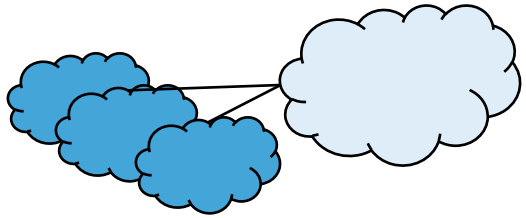
Summary

What Changes

Summary – Structural Changes

Enterprise IT challenges/pain points

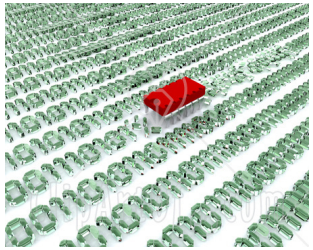
- Adapting to the business model changes of cloud
- Answer: private + public clouds with federation



- Adapting to development model changes of cloud
- Answer: leverage new tools, frameworks to develop Web 2.0 and scale-out apps

- Migrating legacy applications to cloud
- Answer: virtualization to encapsulate legacy OS + apps

- Managing data across apps & users – governance
- Answer: a combined + app platform to manage the data flow among apps and virtual machines



Questions?

References

- Geoff Moore *"Partly Cloudy: Business and Innovation in the Internet Era"* September 2010
  - www.snia.org/cloud/Cloudburst/Moore_SNIA_Keynote.pdf
- Peter Mell & Tim Grance *"The NIST Definition of Cloud Computing"* October 2009
  - csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc
- Any business or computing magazine published at any point in 2009 or 2010

EMC²


EMC²

Big Data Challenges

Unstructured Content

Prepare for digital universe explosion — 34 zettabytes of growth to 2020

Distributed Big Data

Aggregate data as a business advantage; manage as one system

Accessibility

Make available around the globe—from any device—any location

IN A DECADE THE DIGITAL UNIVERSE WILL BE

35 ZETTABYTES

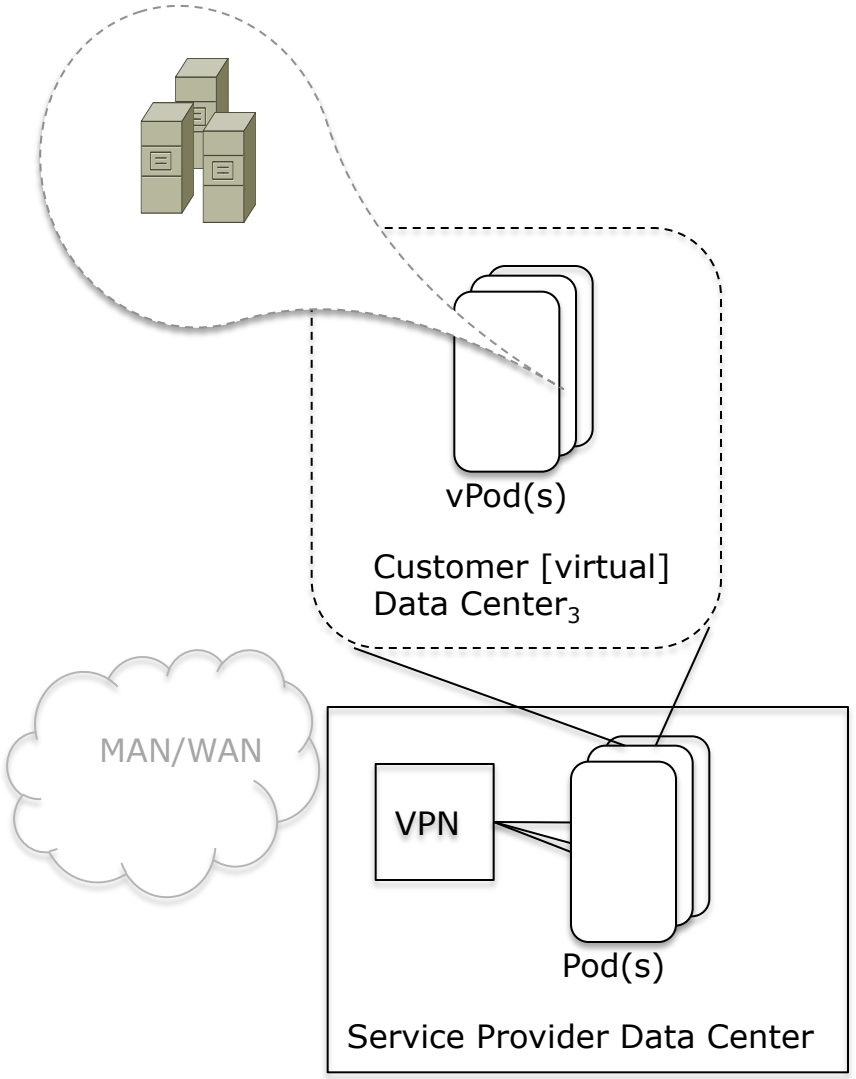
35,000,000,000,000,000,000,000

Source: 2010 IDC Digital Universe Study

EMC²

Key Technology Components

- Policy-driven Orchestration
  - Application mgmt – via virtualization
  - Data mgmt – ILM is finally required
  - Continuous measurement & monitoring – to meter/bill; to maintain high efficiency
- vPods (virtualization)
  - Enables easy migration and replication of containerized applications
  - Drives highly efficient resource utilization
  - Eases rapid deployment of new applications & new services
- Pods (packaged racks)
  - Rack-level deployment of infrastructure (compute + network + storage)
  - Drives highly efficient acquisition and deployment vs. traditional full custom or semi-custom design-per-app



Why The Cloud Is Here To Stay

Enterprise vs. Consumer Technology

EMC²

Another angle – cloud computing is really about bringing enterprise computing technology and applications up to the norms and expectations of consumer computing technology.

The way we run our lives has forever changed. The employees we are hiring right out of school are appalled by the technology we use to run our companies. They are more productive at home than they are in the office.

Marc Benioff, CEO, Salesforce.com

The barrier is becoming less and less between enterprises and consumers in application terms [expectations and functionality often are very much the same]”

Eric Schmidt, CEO, Google

The Big Disconnect

How can it be
I am so **powerful** as a consumer
And so **LAME** as an employee!!??

How disruptive do you think Consumer IT will be to Enterprise IT?

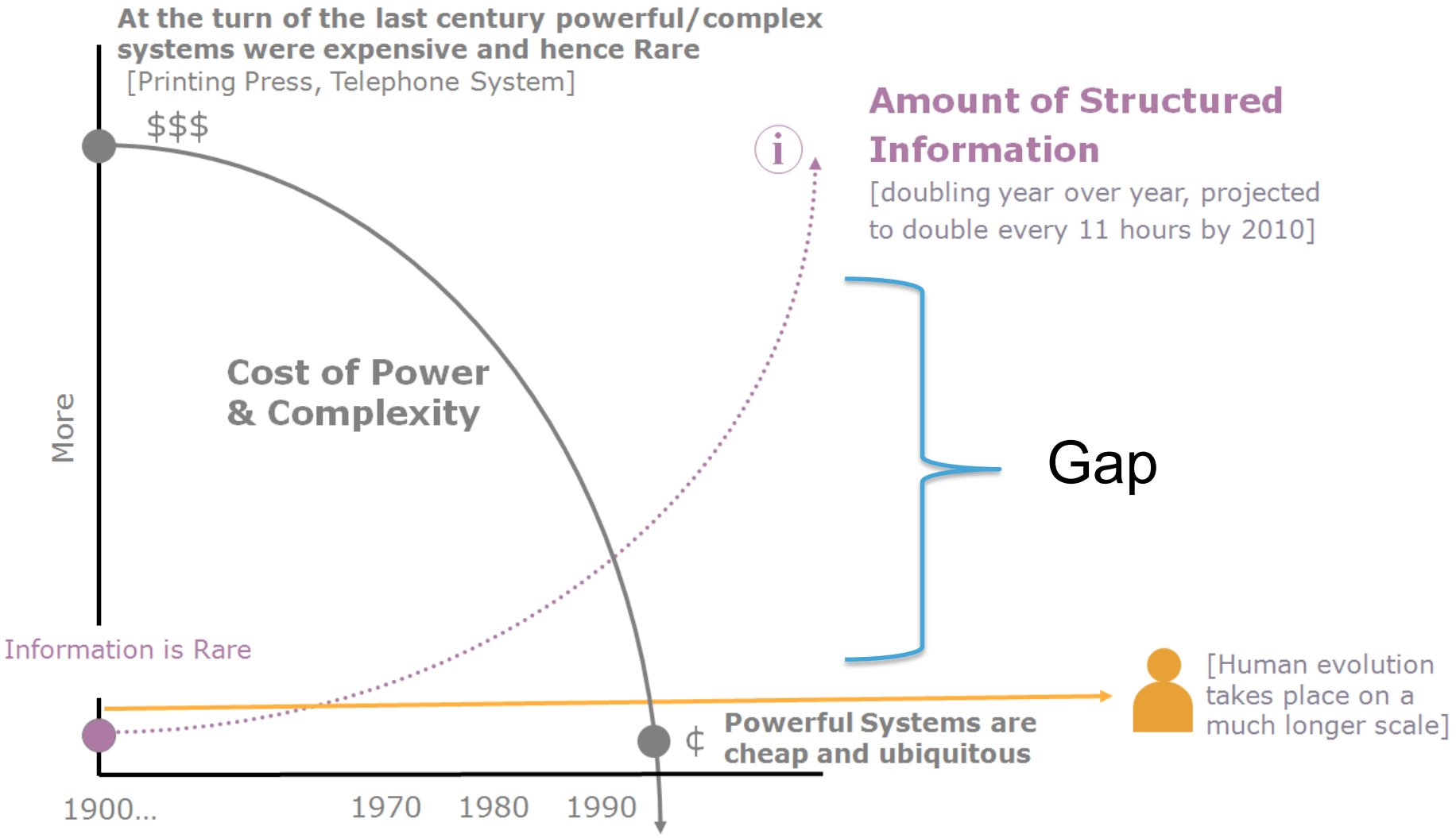
tcg Advisors

From September 2010, SNIA CloudBurst keynote by Geoffrey Moore

Why should employees accept a 50% reduction in their productivity when they come to the office on Monday morning? On the weekend, Google can answer any question I have, on Monday, I can't get the answer to "who are my five biggest customers?" On the weekend, someone from my high school can find me and try to be my friend, on Monday, I can't find my VP of Finance.

Geoff Moore, Author, Crossing the Chasm

Consumer Attention



Courtesy Mick McManus, MAYA Design