Creative Engineering

PDL Distinguished Alumni Talk

Erik Riedel, PhD

October 2010

Bedford Springs, PA

Active Disks - Remote Execution
for Network-Attached Storage

Erik Riedel

Parallel Data Laboratory,

Center for Automated Learning and Discovery

Carnegie Mellon University

www.pdl.cs.cmu.edu/Active

Carnegie Mellon

Parallel Data Laboratory
Center for Automated Learning and Discovery

http://www.pdl.cs.cmu.edu/Active

Active Disks
for Applications

Creativity

Exposure to diverse fields and diverse ideas is critical to be able to see the "big picture" when faced with a problem in your own domain.

Magazines on my nightstand

- Economist
- Business Week
- Wired
- Fast Company
- Technology Review
- Harvard Business Review
- Forbes
- Talk*
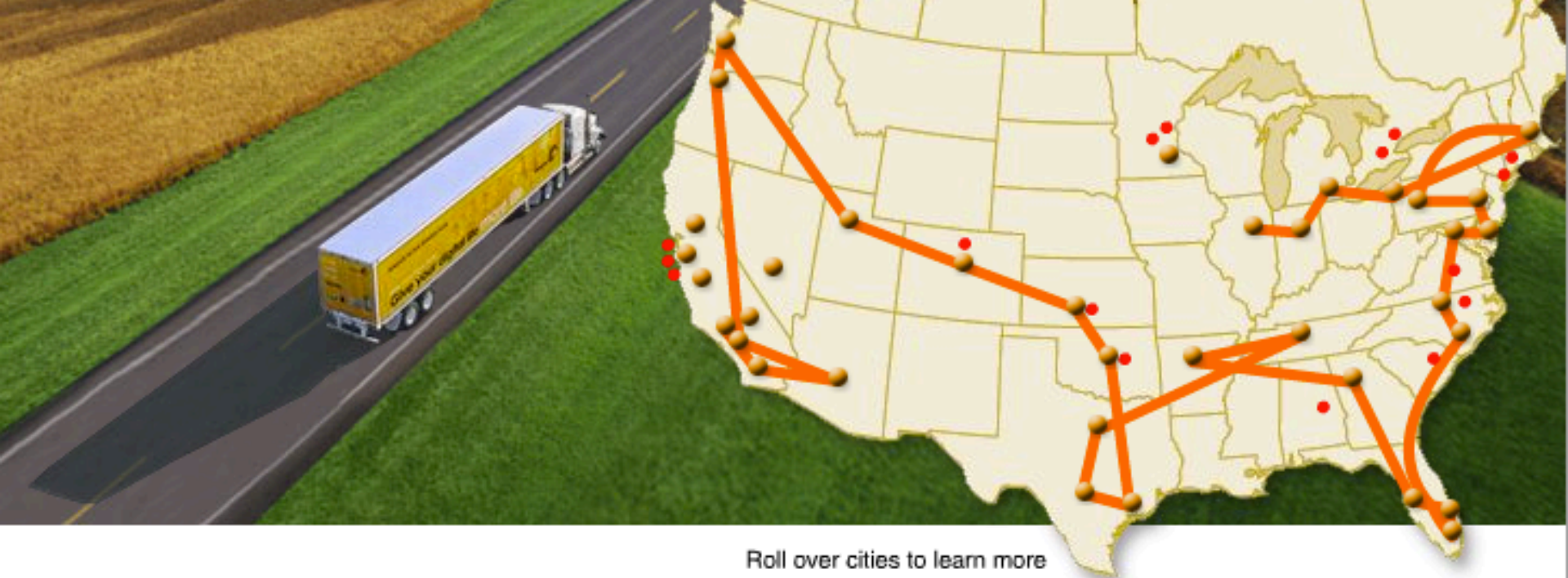
Conferences among my trip reports

- InfoVis
- CODES
- Int'l CES
- Emergence
- SOUPS

- ISCA
- ASPLOS

PLUS – any book with "innovation" in the title

*no longer in print, Tina Brown interview magazine

The same thing is true in discussion among your peers and colleagues.

Inclusion of diverse ideas and approaches leads to superior solutions.

Roll over cities to learn more

Experience Active Storage

The Seagate Active Storage Tour is now officially over. We hope you had the opportunity to view our truck at one of our many stops in North America. To learn more about our products that enhance your digital life, please go to our main site at www.seagate.com.

Most important advice – get out of the office regularly.

Whether to conferences, or to talks or to visit other divisions in your company, or (best) to visit your customers (the target users of your technology).

Home Visits Project

Consumer

As a first step we went into the home of six families, talked to them and watched them interact with their data

Father and son look at photos together
James and Patrick spend a few moments looking at pictures on Shutterfly, with Patrick excitedly talking about who is in the picture and what's happening. Dad asks Patrick questions, Patrick answers. "Is that you?" "Yeah" "I on the wagon." This continues as Shutterfly automatically shows pictures in sequence. ...
27:10 Kiely 4

Seagate

What are the Techniques?

There are many methods to choose from, roughly falling into four categories:

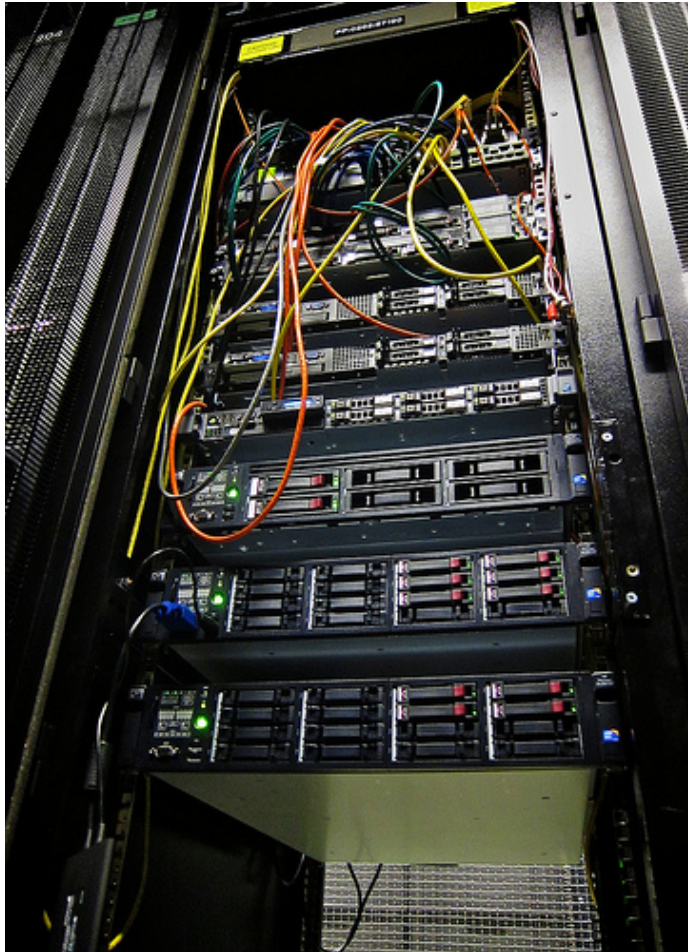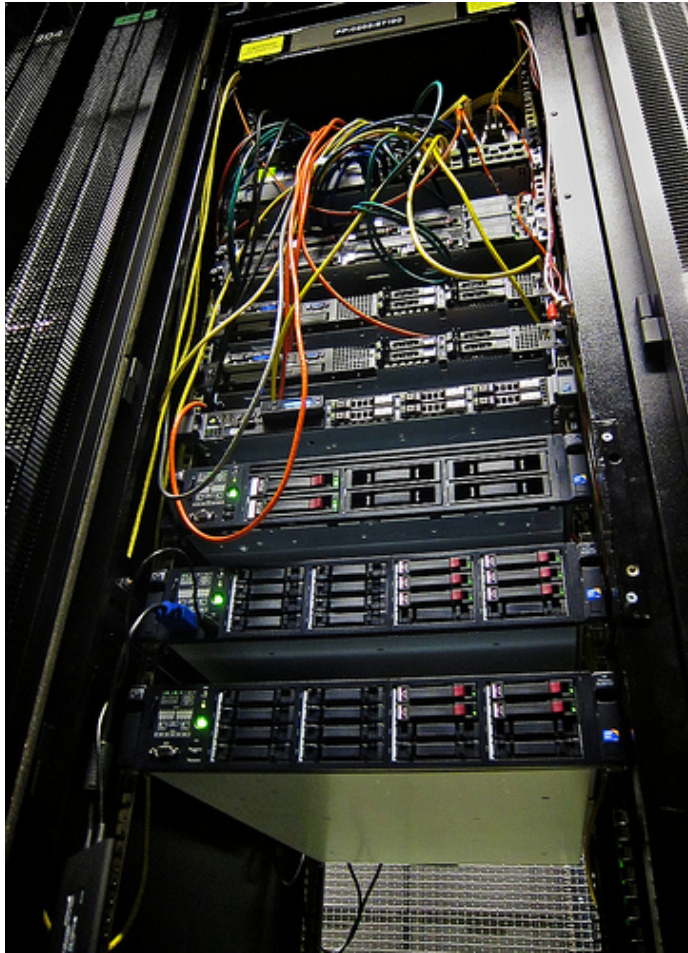**Observation**

Interviews

Self-reporting

Making

- Journals, cameras, audio recorder
- Triggered or open-ended

Enterprise

Go into someone's data center and look around

Look in the corners too

Passion

Enthusiasm.

Results matter.

There are always more solutions than barriers, but there sure are a lot of barriers….

real and imagined

Budgets, skills, headcount, information, freedom-of-action, politics, processes

Don't mis-underestimate the power of getting your hands dirty.

Prototypes, systematic analysis, evaluation.

Make the time for rigor.

Engineering

There are a lot of ways for an engineer / scientist to make an impact.*

The difference between a scientist and an engineer is that the engineer stops to think how much it will cost.

Patience

Objects

Ideas

- U.S. 6,298,401 Anderson
- CMU NASD
  - Network-Attached Secure Disks
- NSIC NASD
  - Wide support & feedback
- T10 OSD (SNIA/ANSI)
  - Wide support & feedback
  - Worked out lots of details
- Storage Bricks (Gray)
  - Saw the win, spread the word

Results

- Lustre
- Panasas
- IETF pNFS
- EMC Atmos

Influences

- Google GFS
- Spinnaker
- S3
- ….

OSD Systems – 2006

A variety of Object-based Storage Devices being built today

- ➤ Disk array/server subsystem
- ➤ E.g. LLNL units with Lustre

- ➤ "Smart" disk for objects
- ➤ E.g. Panasas storage blade

- ➤ Highly integrated, single disk
- ➤ E.g. prototype Seagate OSD

➤ **File/ Security Manager**

- ➤ Orchestrates system activity
- ➤ Balances objects across OSDs
- ➤ Called clustered MDS in Lustre
- ➤ Called Mgmt Blade by Panasas
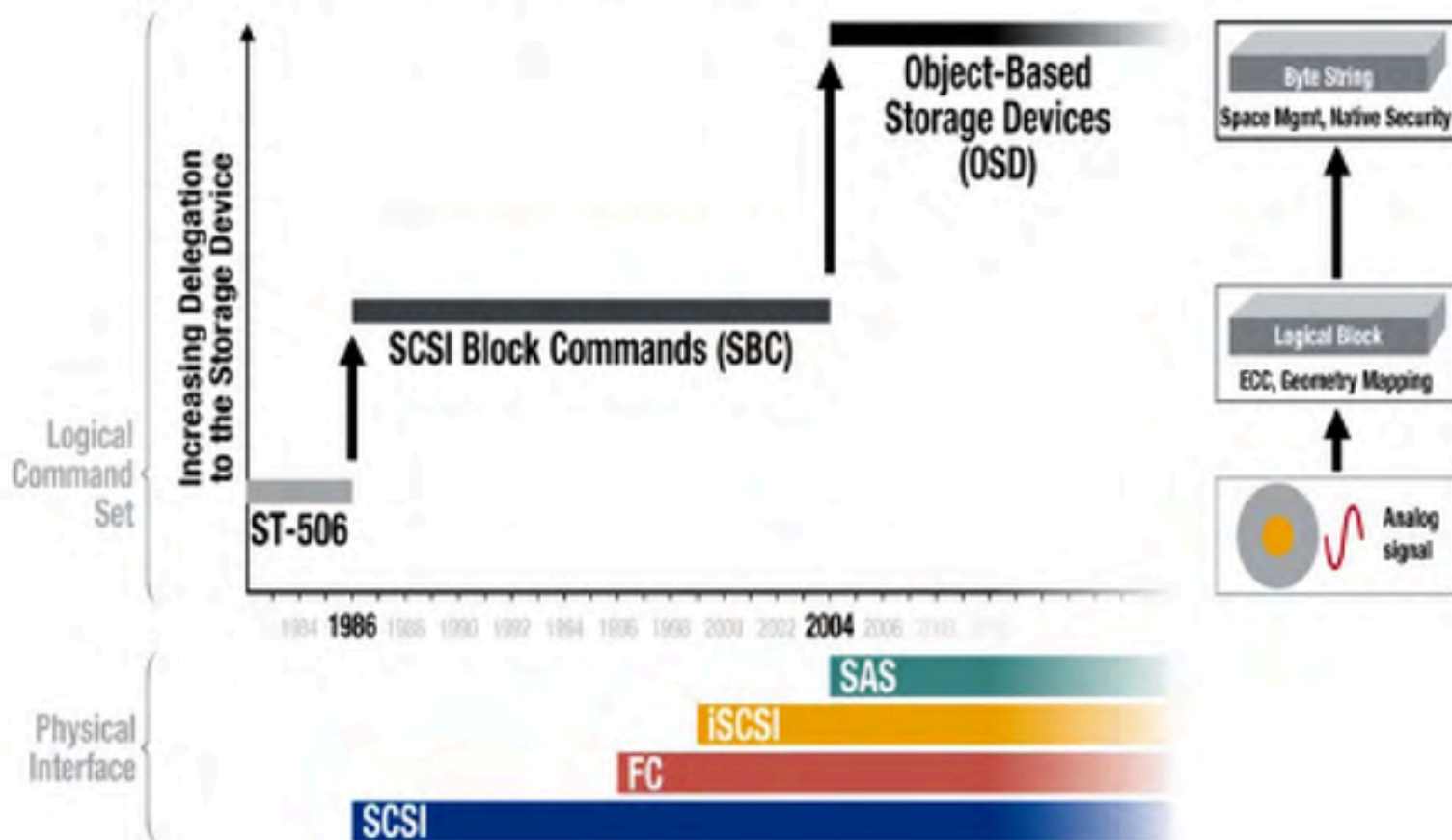- ➤ Called ST server cluster by IBM

➤ **Scalable Network**

- ➤ Connectivity among clients, managers, and devices
- ➤ Shelf-based GigE (Panasas)
- ➤ Specialized cluster-wide high-performance network (Lustre)
- ➤ Storage network (IBM)

Object Storage & Applications

4

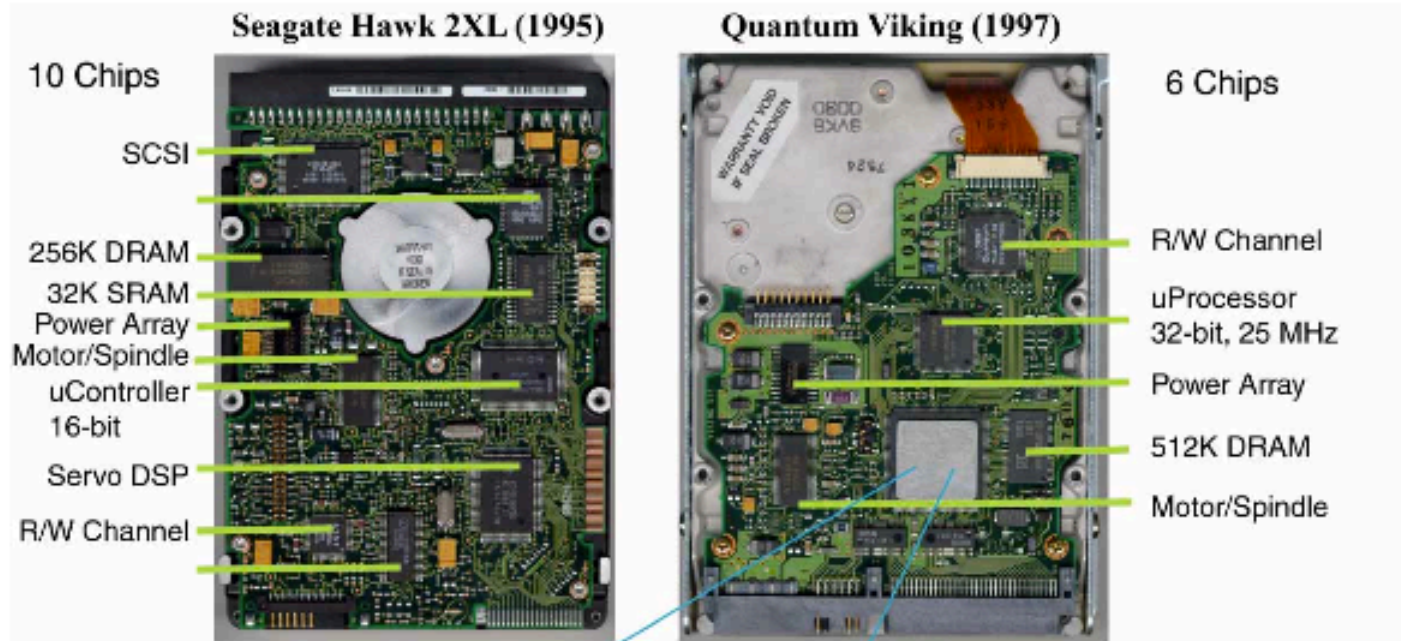# Object-Based Storage—Delegate Deeper

A Fundamental Change in Technology

Patience II

Evolution of Disk Drive Electronics

Seagate Hawk 2XL (1995)

10 Chips

- SCSI
- 256K DRAM
- 32K SRAM
- Power Array
- Motor/Spindle
- uController 16-bit
- Servo DSP
- R/W Channel

Quantum Viking (1997)

6 Chips

- R/W Channel
- uProcessor 32-bit, 25 MHz
- Power Array
- 512K DRAM
- Motor/Spindle
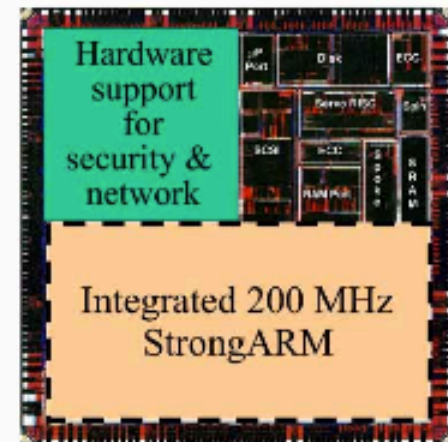
Integration

- **reduces chip count**
- **improves reliability**
- **reduces cost**
- **future integration to processor on-chip**
- **but there must be at least *one* chip**

μP Port | Disk | ECC
Servo RISC | Spin
SCSI | ECC | Spoke | SRAM
RAM Port

Trident ASIC

Hardware support for security & network

Integrated 200 MHz StrongARM
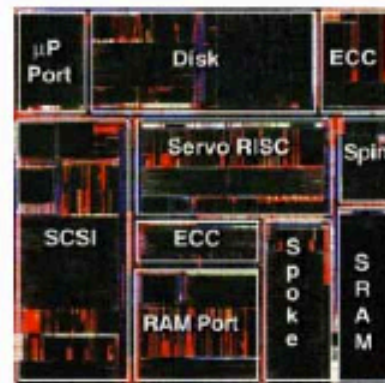
Future Generation ASIC

- **Active Disk architectures**
  - > Integrated processing power and memory into disk units
  - > Scaled processing power as the dataset grew

- **Decision support algorithms offloaded to Active Disks** to support key decision support tasks
  - > Active Disk architectures use stream-based model ideal for software architecture of relational databases

In Netezza's NPS® System: "Snippet Processing Units" take streams as inputs and generate streams as outputs

See U.S. Patent 7,464,106
(January 2006)

NETEZZA
The Power to Question Everything™

Active Disks as Intelligent Storage Nodes

Netezza Performance Server

Netezza added:

- **Highly optimized query planning**
- **Code generation**
- **Stream processing**

Snippet Processing Unit (SPU)

Result: 10X to 100X performance speedup over existing systems

Intelligent Query Streaming™

NIC
CPU
RAM
FPGA
Flash RAM

NETEZZA
The Power to Question Everything™
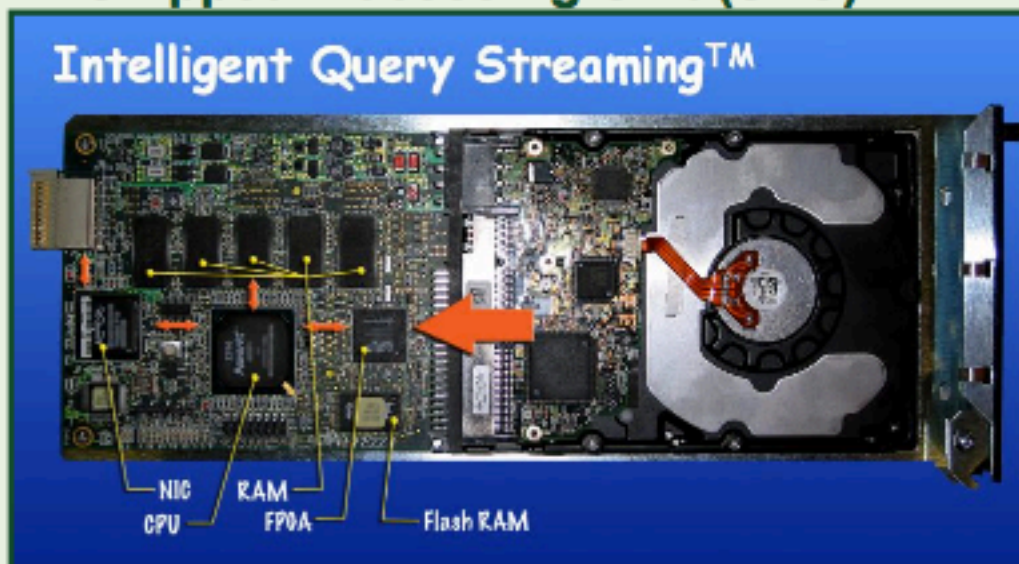
Big Data Means Big Sales: IBM to Buy Netezza for $1.7B

By Om Malik | Sep. 20, 2010, 5:57am PDT | 2 Comments

The era of Big Data arrived in full earnest this week when IBM decided to acquire Marlborough, Mass.,-based Netezza Corporation, a maker of data warehousing analytics for a whopping $1.7 billion in cash. IBM's decision came hours after Oracle announced a cloud-based Exadata Elastic Service at Oracle World in San Francisco.

IBM is offering $27 a share for the Neteeza and hopes that the smaller company would help with its growing business analytics. Over the past four years, IBM has invested more than $12 billion in 23 analytics related acquisitions.
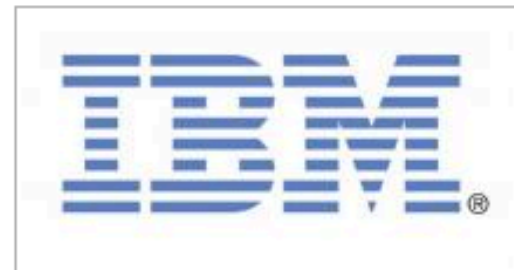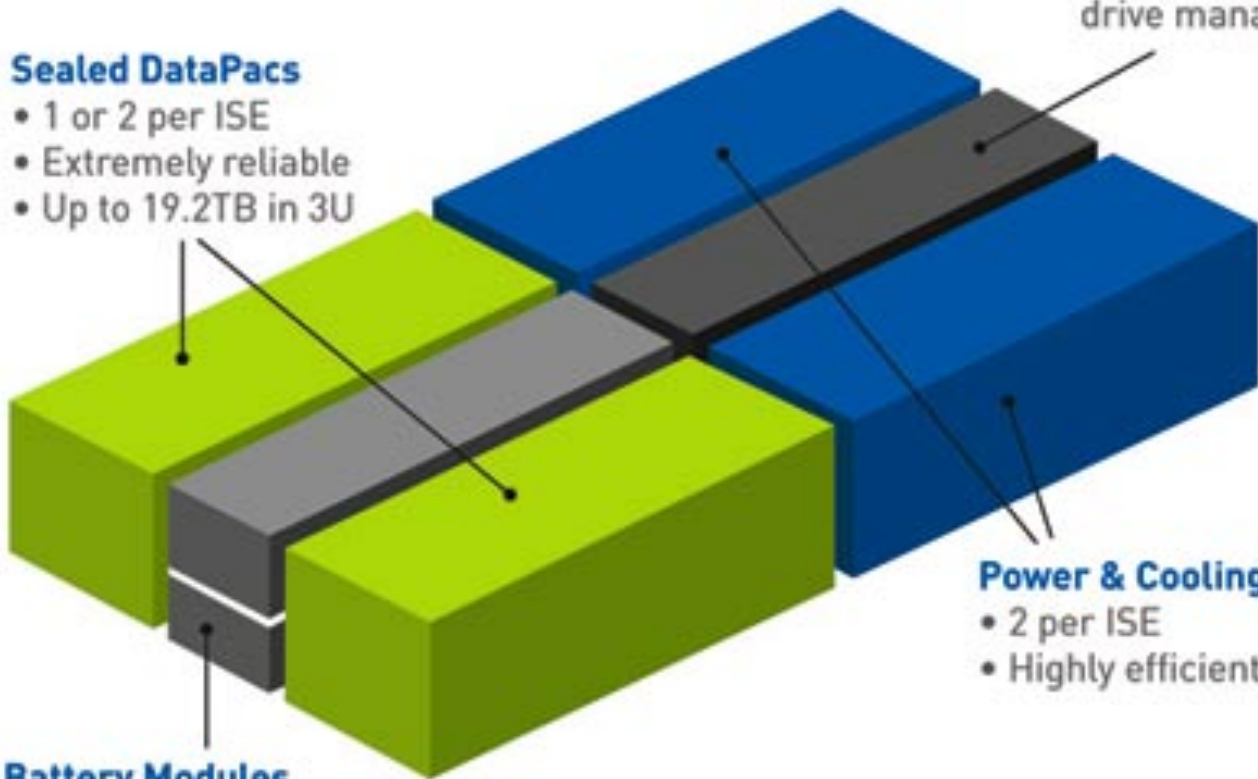
Patents

- 6,732,241   Technique for migrating data between storage devices for reduced power consumption

- 7,200,747   System for ensuring data privacy and user differentiation in a distributed file system

- 7,405,941   Storage array with enhanced RVI

Inside ISE

Sealed DataPacs
- 1 or 2 per ISE
- Extremely reliable
- Up to 19.2TB in 3U

Managed Reliability Controllers
- 2 per ISE
- High Performance
- Local RAID, cache, drive management

Power & Cooling Units
- 2 per ISE
- Highly efficient & reliable

Battery Modules
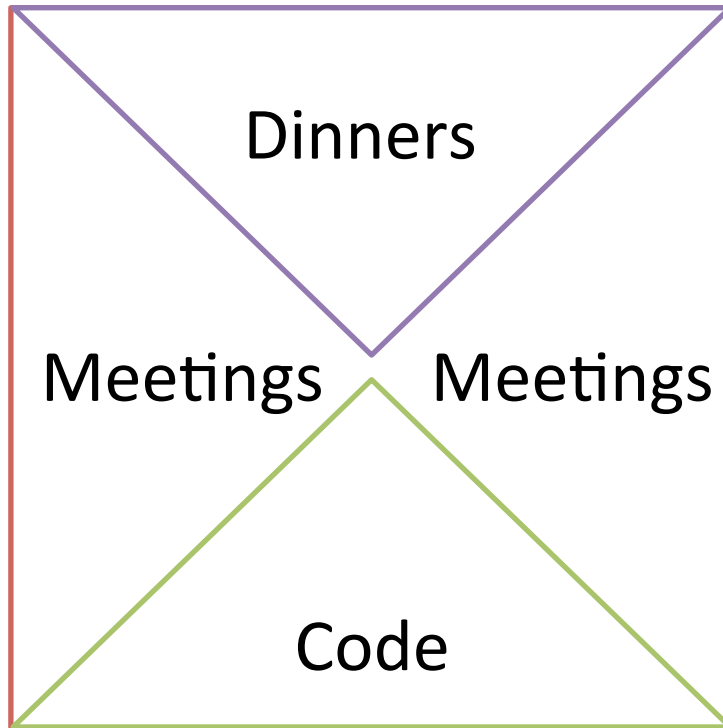- 2 per ISE
- Up to 96 hours protection

xiotech.

Influence

- Research vs. Product Groups

- CTOs have an impact almost exclusively by influencing the "real" parts of the business
  - Product Management
  - Sales & Marketing
  - Business GMs – caveat: don't confuse power with influence

Communication

Dinners

Meetings   Meetings

Code

Amount of time spent

Career Progression

*Advice courtesy Jim Gray, as retold in May 2007 at UC Berkeley on a day both sad and uplifting

Never underestimate the power of talking and **listening**.

*PowerPoint is **way** over-rated.*

There are numerous studies that show that your personal network is the single biggest determinant of career success.

These studies are correct.

Summary

Creativity

Passion

Results & Rigor

Patience

Influence

Communication

References

- *Listen: Discover the Secret to Getting Through to Absolutely Anyone* by Mark Goulston
- *The Tipping Point: How Little Things Can Make a Big Difference* by Malcolm Gladwell
- *Made to Stick: Why Some Ideas Survive and Others Die* by Chip & Dan Heath
- Design Thinking (IDEO and others)
- *The Future of Enterprise IT* by Geoff Moore at 2010 SNIA CloudBurst event

- Gobioff/Riedel Emigration talks – startups vs. labs

What wouldn't fit into the talk
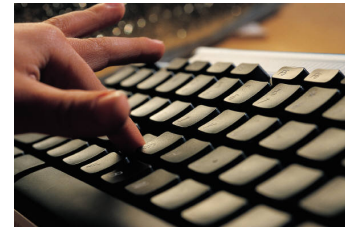
OTHER THOUGHTS

Cloud Computing

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

Programmers

Ruby/Rails, MapReduce

IT Managers

VM images, vApps, VLANs

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

Consumer vs. Enterprise

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.
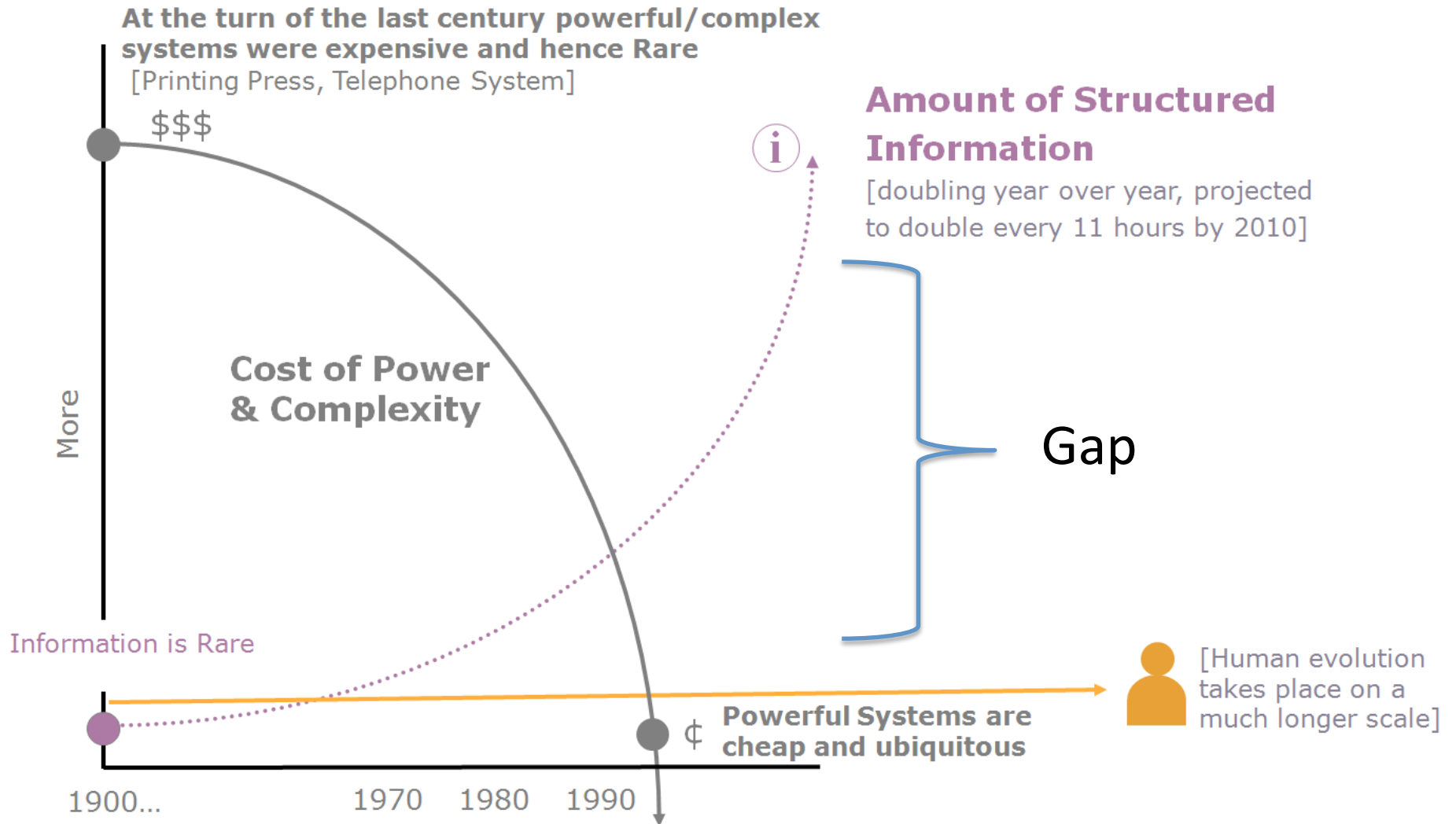
Marc Benioff, CEO, Salesforce.com

The Big Disconnect

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

How disruptive do <u>you</u> think Consumer IT will be to Enterprise IT?

From September 2010, SNIA CloudBurst keynote by Geoffrey Moore

Consumer Attention

At the turn of the last century powerful/complex systems were expensive and hence Rare
[Printing Press, Telephone System]

$$$

Amount of Structured Information
[doubling year over year, projected to double every 11 hours by 2010]

Cost of Power & Complexity

More

Gap

Information is Rare

¢

Powerful Systems are cheap and ubiquitous

[Human evolution takes place on a much longer scale]

1900...   1970   1980   1990
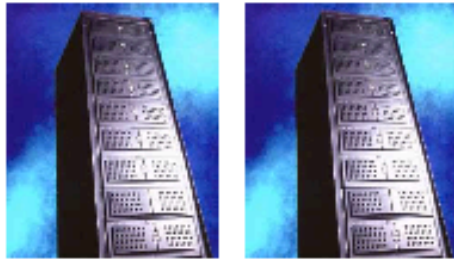
Courtesy Mick McManus, MAYA Design

Small Nodes vs. Big Nodes

Network "Appliances" Can Win Today

Dell PowerEdge & PowerVault System

Dell PowerVault 650F $40,354 x 12 =484,248
  512 MB cache, dual link controllers, additional 630F cabinet,
  20 x 9 GB FC disks, software support, installation
Dell PowerEdge 6350 $11,512 x 12 = 138,144
  500 MHz PIII, 512 MB RAM, 27 GB disk
3Com SuperStack II 3800 Switch 7,041
  10/100 Ethernet, Layer 3, 24-port
Rack Space for all that 20,710

NASRaQ System

Cobalt NASRaQ $1,500 x 240 =360,000
  250 MHz RISC, 32 MB RAM, 2 x 10 GB disks
Extra Memory (to 128 MB each) $183 x 360= 65,880
3Com SuperStack II 3800 Switch $7,041 x 11= 77,451
  240/24 = 10 + 1 to connect those 10
Rack Space (estimate 4x as much as the Dells) 82,840
Installation & Misc 50,000

Comparison

	Dell	Cobalt
Storage	2.1 TB	4.7 TB
Spindles	240	480
Compute	6 GHz	60 GHz
Memory	12.3 GB	30.7 GB
Power	23,122 W	12,098 W
Cost	$650,143	$636,171

From April 1999, Active Disks talk