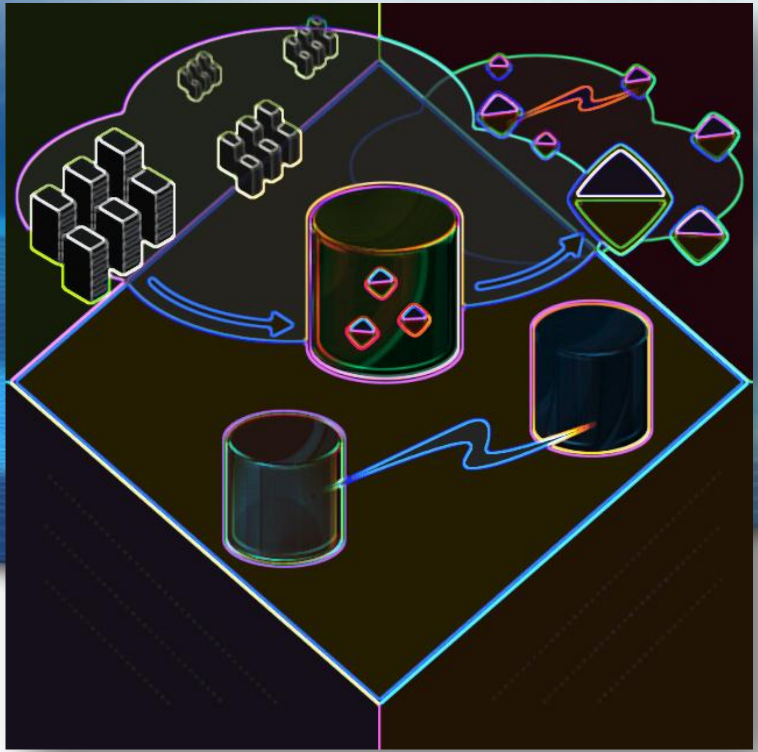
Microsoft

EXTREME COMPUTING GROUP



*Defining the future.*

# The Cloud Will Change Everything


James Larus

eXtreme Computing Group, Microsoft Research


MSR Cloud Futures Workshop

June 2, 2011

# Presentation Roadmap

What are the technology trends shaping the future – hardware, software, experiences?

What comes after the PC?

How can computing change today's life experiences – computers that become 'assistants'?

What is XCG doing to drive innovation in this new world?

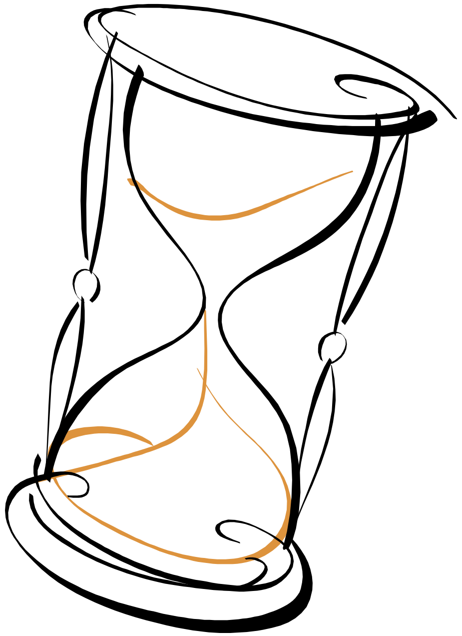


Microsoft

# It's Easy To Forget That Not Very Long Ago …

There were few or no experiences with…

- Web sites, email, spam, phishing, computer viruses
- e-commerce, digital photography, social networking, or video



Cell phones were rare and expensive

A portable cassette player was still cool

HiFi was more common than WiFi

A "friend" was someone you actually knew

# Understanding the Future

## Hard to predict, especially the future

- In the near term, we *overestimate* change
- In the long term, we *underestimate* change

## Outside their field of expertise

- Experts are often better at predictions

## Recognize exponentials

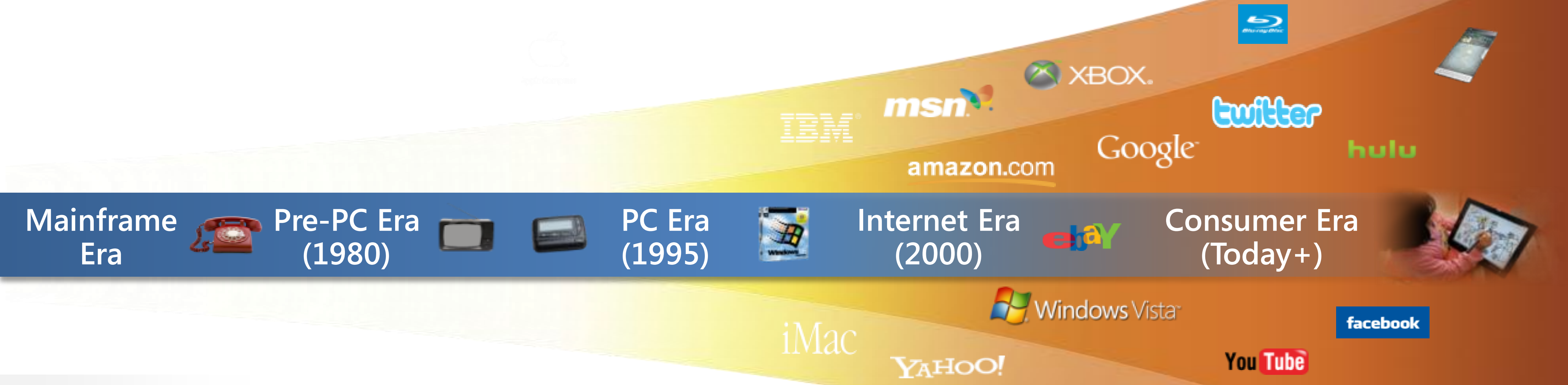
- Quantitative change brings qualitative change
- Multidisciplinary coupling shifts the balance

## Technological and social change

- Different rates with differing consequences

Everything that can be invented has been invented."
*-Charles Duell, head of the U.S. Patent Office, 1899*
"Who in their right mind would ever need more than 640k of ram!?"
*- Bill Gates, 1981*
"Computers in the future may weigh no more than 1.5 tons."
*- Popular Mechanics, 1949*
"I think there is a world market for maybe five computers."
*- Thomas Watson, chairman of IBM, 1943*
"There is no reason anyone would want a computer in their home."
*- Ken Olson, president, CEO of Digital Equipment Corp., 1977*
"This 'telephone' has too many shortcomings to be seriously considered as a means of communication. The device is inherently of no value to us."
*- Western Union internal memo, 1876*
"The wireless music box has no imaginable commercial value. Who would pay for a message sent to nobody in particular?"
*- David Sarnoff's associates about radio in the 1920s.*
"We don't like their sound, and guitar music is on the way out."
*- Decca Recording Co. rejecting the Beatles, 1962.*
"X rays are a hoax."
*-Lord Kelvin, physicist, c. 1900*
"The Bomb will never go off, and I speak as an expert in explosives."
*-Admiral William Leahy, advising Truman on atom bomb, 1945*
"Space travel is utter bilge."
*-Richard van der Riet Wooley, British Astronomer Royal, 1956*
"The cloning of mammals...is biologically impossible."
*-James McGrath and Davor Solter, Science, Dec. 14, 1984*

# Computing Eras: Exponential Change




## 21st century implicit and natural computing

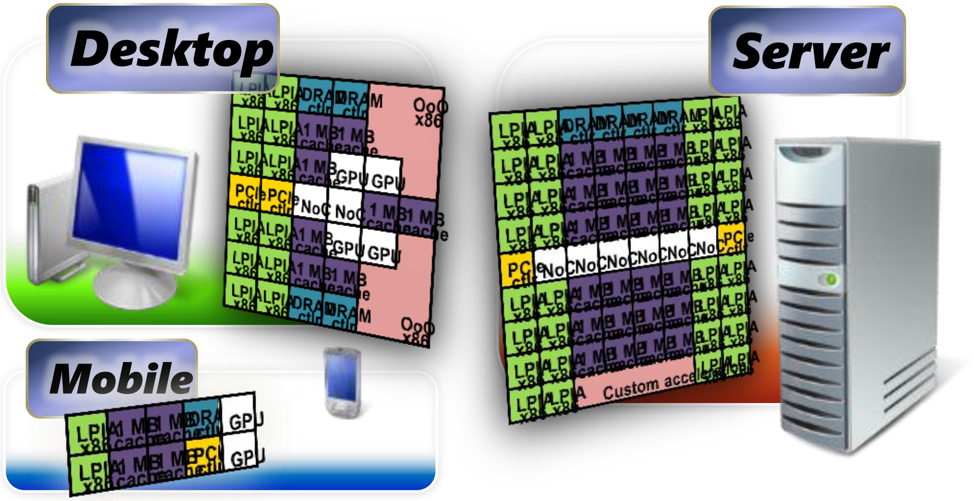
- Increasingly natural interfaces
- Embedded intelligence in everyday objects
- Ubiquitous network access and cloud services


Microsoft

# What Is Changing?

- System on a chip
  - Powerful mobile devices
- Graphics processing units
  - High quality graphics
- Explosive data growth
  - Ubiquitous sensors and media

- Inexpensive, embedded computing
  - Smart objects, CIP, …
- Wireless spectrum pressure
  - Mobile device growth
- New software models
  - Social networks, clients+clouds …




# Megatrends: The Many Device World

**System on a Chip Designs**
Powerful Mobile Devices

**Graphics Processing Units**
High Quality Graphics

**Explosive Data Growth**
Ubiquitous Sensors and Media

**Inexpensive Embedded Computing**
Everyday Smart Objects, CIP

**Mobile Device Growth**
Smart Phones and Feature Phones

**New Software Models**
Social Networks, Clients + Clouds…

# Moore's "Law" and Limiting Exponentials …

The experts look ahead

## Cramming more components onto integrated circuits

**With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip**

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

**The future of integrated electronics** is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wristwatch needs only a display to be feasible today.

But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Computers will be more powerful, and will be organized in completely different ways. For example, memories built of integrated electronics may be distributed throughout the machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

**Present and future**

By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional ones that result in electronics functions supplied to the user as irreducible units. These technologies were first investigated in the late 1950's. The object was to miniaturize electronics equipment to include increasingly complex electronic functions in limited space with minimum weight. Several approaches evolved, including microassembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and converged so that each borrowed techniques from another. Many researchers believe the way of the future to be a combination of the various approaches.

The advocates of semiconductor integrated circuitry are already using the improved characteristics of thin-film resistors by applying such films directly to an active semiconductor substrate. Those advocating a technology based upon films are developing sophisticated techniques for the attachment of active semiconductor devices to the passive film arrays.

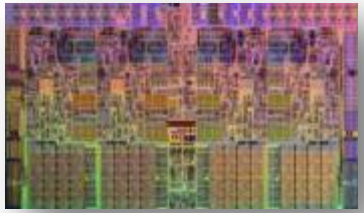
Both approaches have worked well and are being used in equipment today.

**The author**

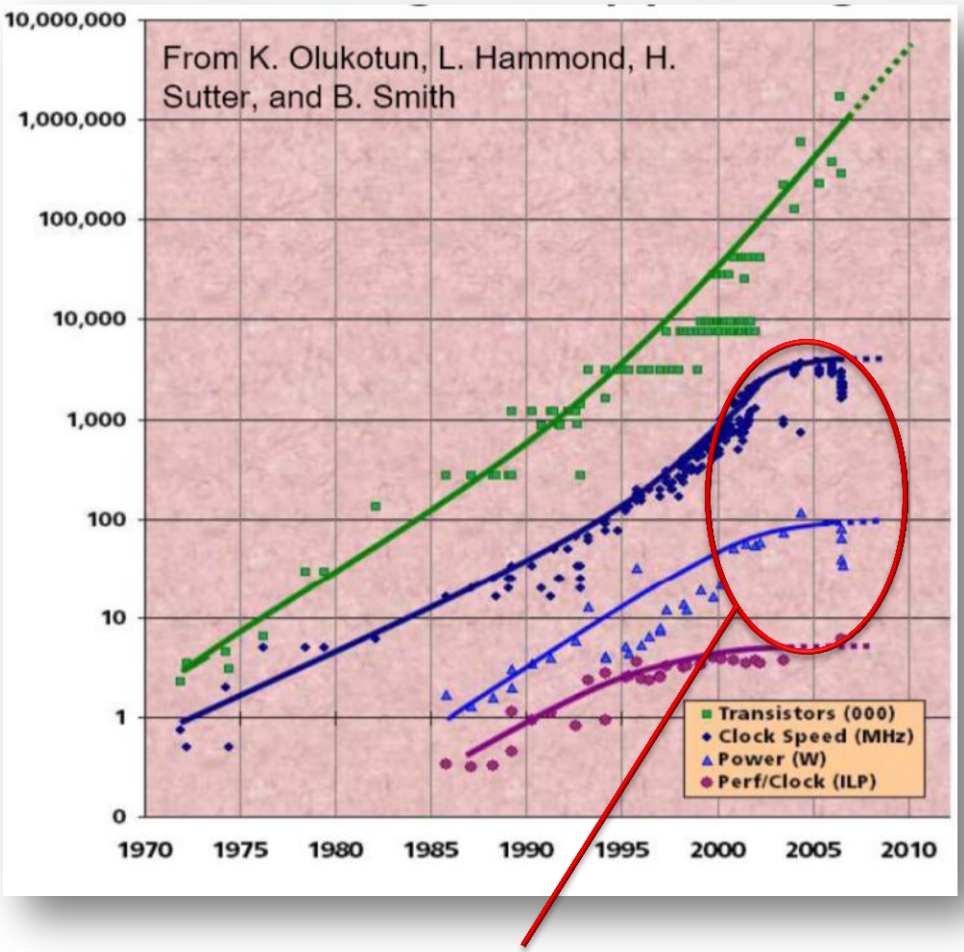
Dr. Gordon E. Moore is one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics. He earned a B.S. degree in chemistry from the University of California and a Ph.D. degree in physical chemistry from the California Institute of Technology. He was one of the founders of Fairchild Semiconductor and has been director of the research and development laboratories since 1959.

Electronics, Volume 38, Number 8, April 19, 1965

Intel 4004



Intel Core i7



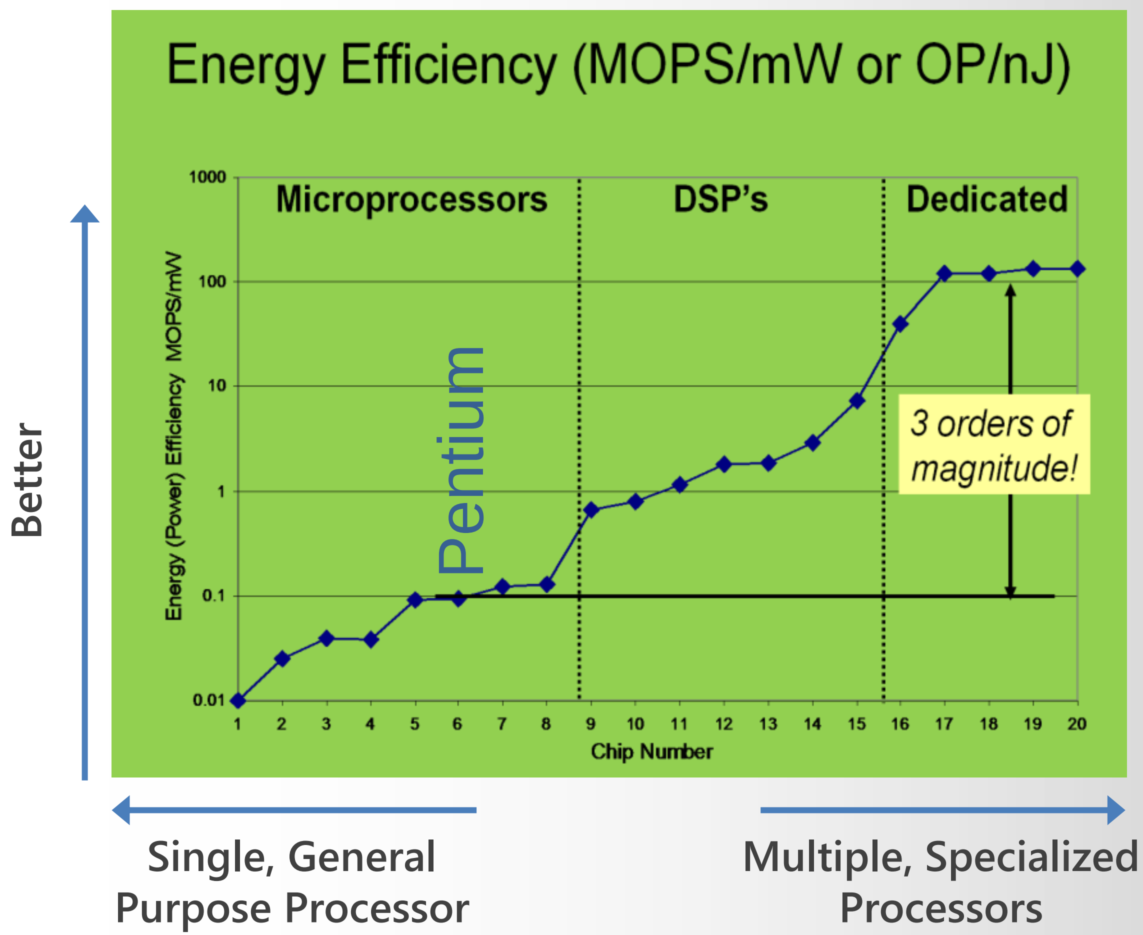

Intel's Inflection



Microsoft

# Silicon Scaling Challenges Extant Ecosystem

- Embrace heterogeneity
  - Functional and performance specialization
  - Optimize for function
- We're surrounded by "opportunities"
  - Devices and architectures
  - Algorithms and applications
  - Usage models and behaviors




Mihai Budiu, "On The Energy Efficiency of Computation," February 2004


Microsoft

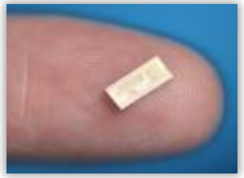
# Internet of Things and Systems on a Chip: Milliwatts Matter

## System on a Chip (SoC): The New Motherboard

- Core(s), memory controller, I/O
- Function-specific accelerators
  - Graphics, communications, sensors, security

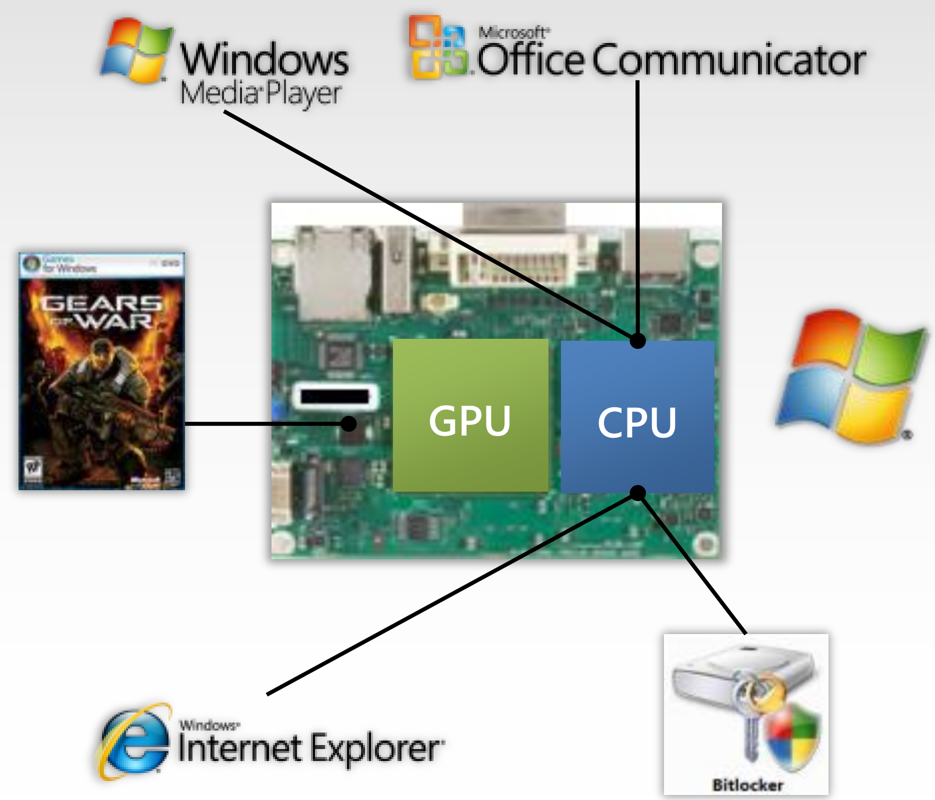
## Internet of Things (IoT)

- Embedded intelligence in everyday objects
- Experiences and natural user interfaces (NUIs)
- Resource discovery, security, services, programming

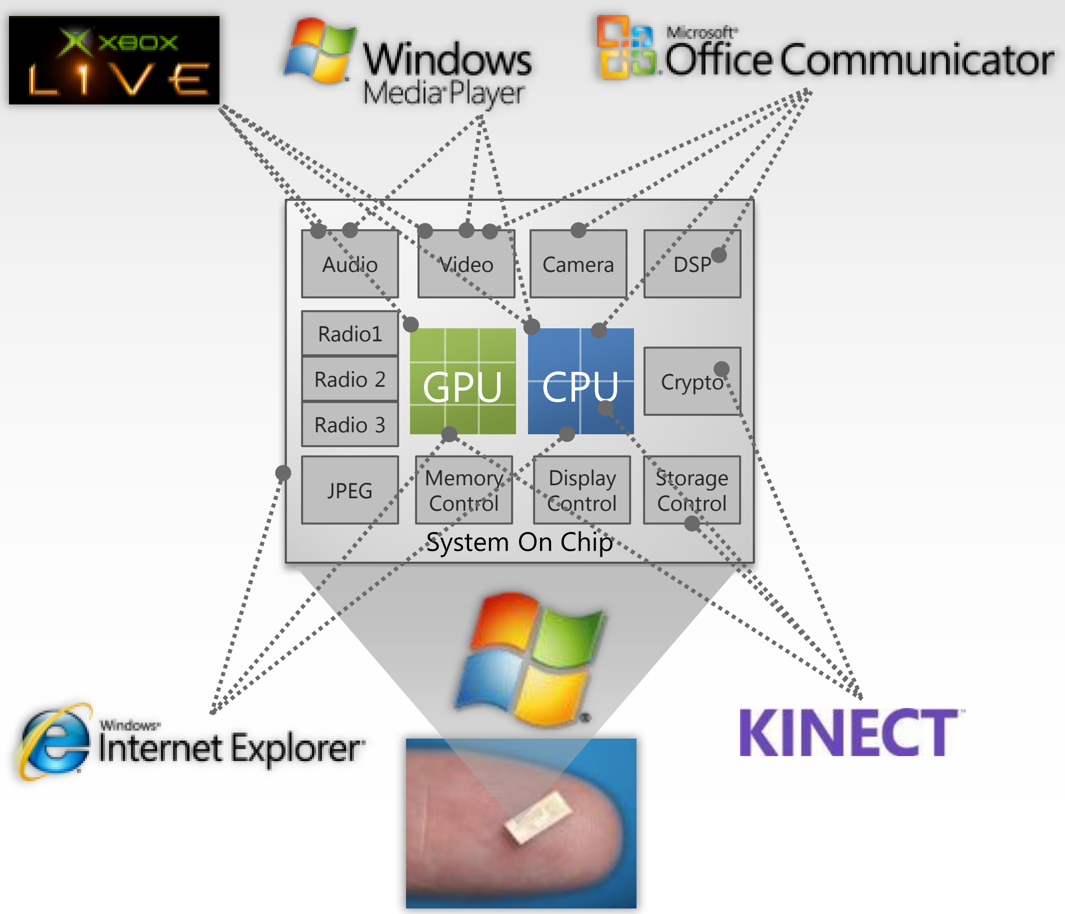


# System on a Chip (SoC) Implications

## Today




## Today and Tomorrow

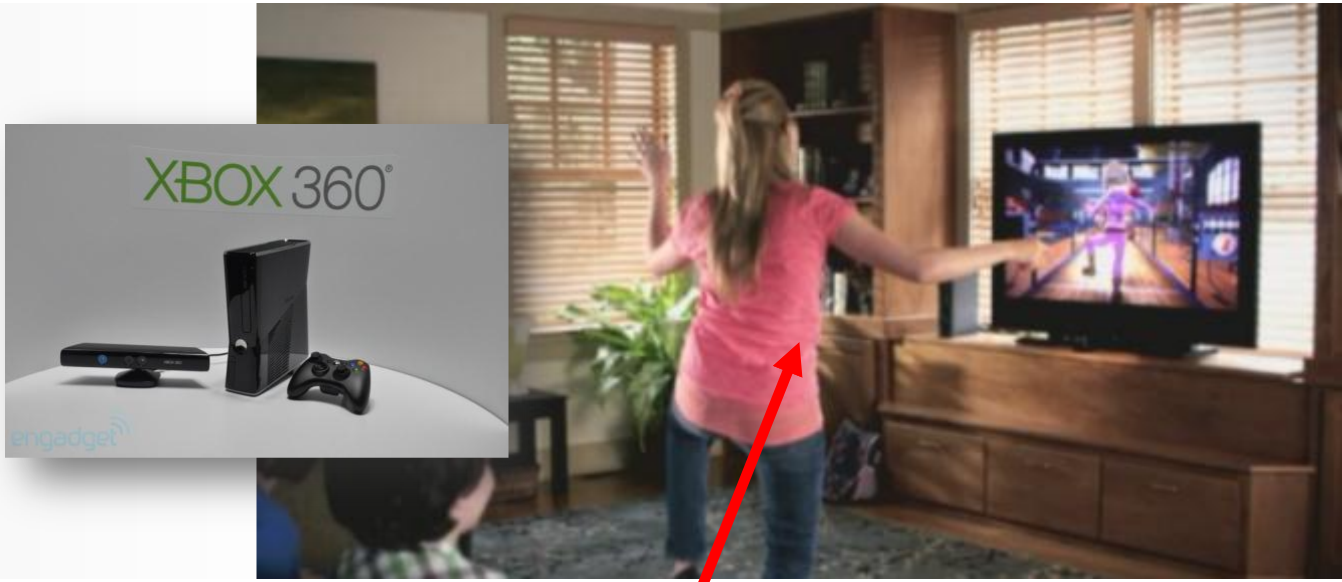




Microsoft

# Multicore and SoCs: What's An Application?

Microsoft Kinect




## An FFT?
- No, it's an algorithm

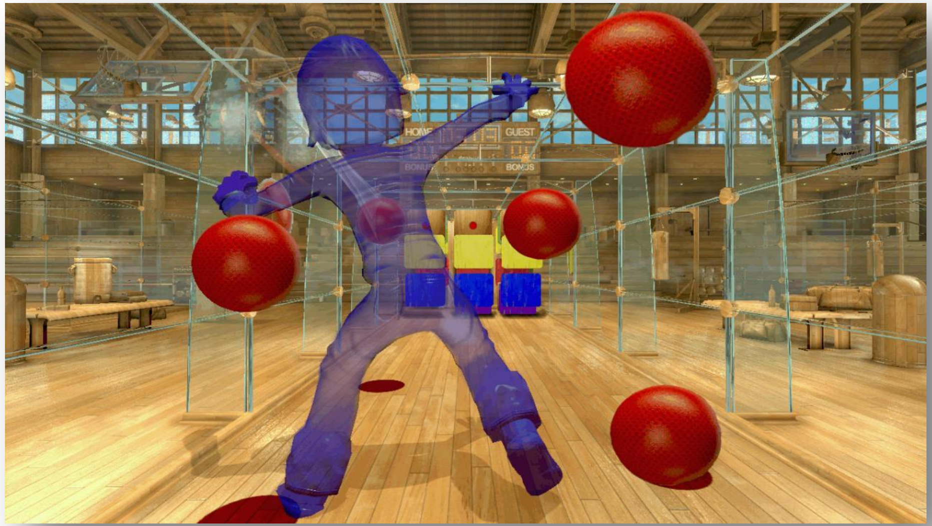
## A rendering pipeline?
- No, it's a software library

## A feature recognition system?
- No, it's a building block

## Our notion of "application" is increasingly complex
- Integrated and interoperating components

## Our tools must enable creativity
- Creation of integrated experiences

# Integrated Capabilities Are Increasingly Common




Hardware Companies ↑

Applications

Languages, Compilers, Operating Systems

Specialization (ASIC, GPUs, FPGAs) | Architecture (ISA) (x86, ARM, EDGE)

Microarchitecture

Chip Fabrication

Software Companies ↓




Microsoft

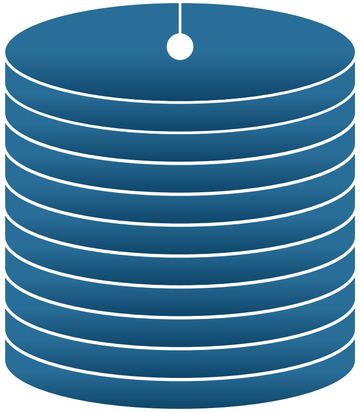

$1.2 \times 10^{21}$
New Bytes of Information in 2010
Source: IDC, as reported in The Economist, Feb 25, 2010
14

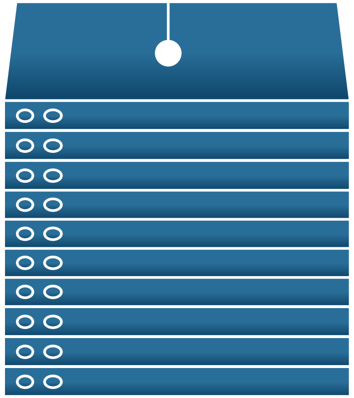
# Economics of Storage

$$40.576

$$10,2150

Disk Storage
(per gigabyte)





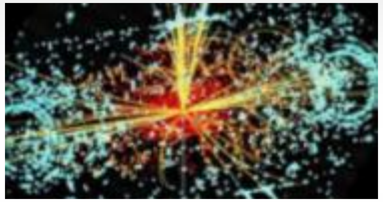
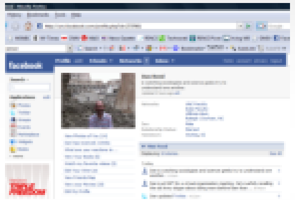
Web Storage
(per gigabyte)

2001

... free storage is like free puppies ...

Source: Wired Magazine April 2010; Figures represented in USD

# The Data Explosion

| Experiments | Simulations | Archives | Literature | Consumer |
| --- | --- | --- | --- | --- |
|  |  |  | Cornell University Library<br>arXiv.org |  |

Petabytes
Doubling & Doubling

## The Challenge
Enable Discovery

Deliver the capability to mine, search and analyze this data in near real time

## The Response

Discovery itself is evolving



Microsoft

In 2000 the Sloan Digital Sky Survey collected more data in its 1st week than was collected in the entire history of Astronomy

By 2016 the New Large Synoptic Survey Telescope in Chile will acquire 140 terabytes in 5 days - more than Sloan acquired in 10 years

The Large Hadron Collider at CERN generates 40 terabytes of data every second

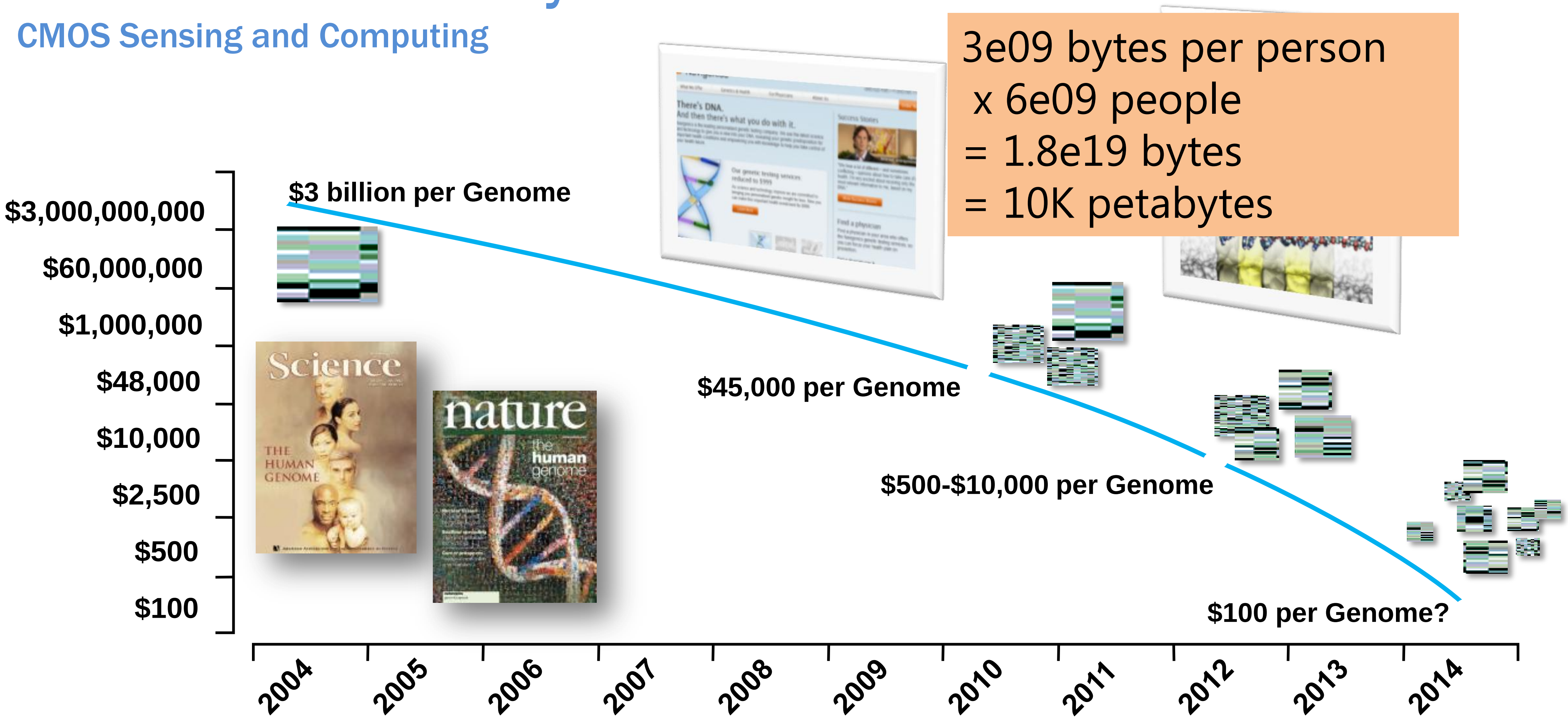


Sources:  The Economist, Feb ‘10; IDC

# Genetics Gets Really Personal

## CMOS Sensing and Computing

3e09 bytes per person
x 6e09 people
= 1.8e19 bytes
= 10K petabytes




Source: George Church, Harvard Medical School, as reported in *IEEE Spectrum*, Feb ’10. Figures represented in USD


Microsoft

# The Changing Nature Of Research: Data Driven

## Experimental

Thousand
years ago

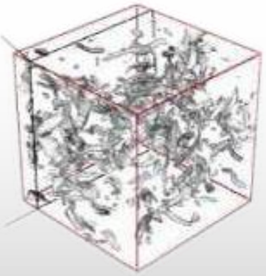
*Description of natural phenomena*

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$

## Theoretical

Last few
hundred years

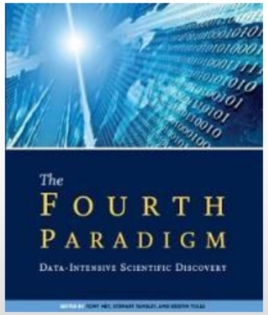
*Newton's laws, Maxwell's equations…*



## Computational

Last
few decades

*Simulation of complex phenomena*




## The Fourth Paradigm

Today and the
Future

*Unify theory, experiment and simulation with large multidisciplinary data*

*Using data exploration and data mining (from instruments, sensors, humans…)*

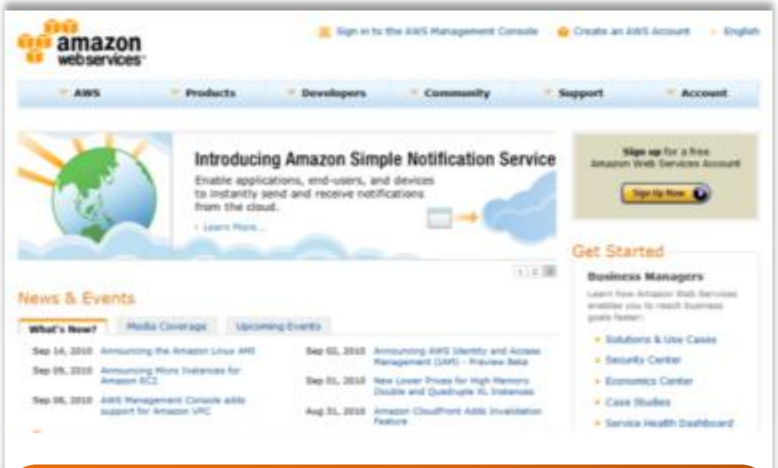
*Distributed communities*



Microsoft

# What’s A Cloud?

# Clouds: There Are Lots of Shapes




**Amazon Web Services**

User deploys and runs software; retains control over operating system and deployed applications


**Windows Azure & Office 365**

Cloud provider offers infrastructure and permits users to create or run applications


**Salesforce.com**

Applications run in the cloud

# What's Causing The Cloud Excitement?

**Cost Optimization**
- Reaping economies of scale
- Including virtualization

**Business Efficiencies**
- Focus on core competencies

**Transfer of Responsibilities**
- To cloud service provider

**Just-in-time Provisioning**
- Pay only when you need it


Microsoft

# What’s A Cloud? The Traditional View



Microsoft

# Microsoft's Data Center Evolution And Economics

2005 | 2006 | 2007 | 2008 | 2009 | 2010

## Data Center Co-Location Generation 1

## Quincy and San Antonio Generation 2

## Chicago and Dublin Generation 3

## Modular Data Center Generation 4


**Deployment Scale Unit**

### Server

Capacity

### Rack

Density & Deployment

### Containers

Scalability & Sustainability

### IT PAC

Time to Market
Lower TCO

# Generation 3 - Chicago Data Center

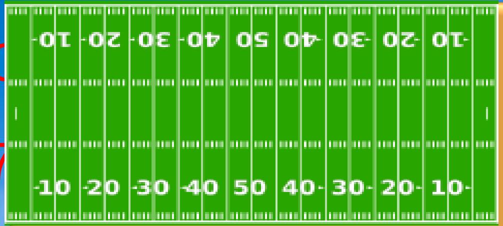
$500M+ investment

1.5 million person hours-of-labor

300

707

2400




Each data center is approximately

**17 times**

the size of a football field

and uses containers

7.5 miles of chilled water piping

26,000 cubic yards of concrete



Microsoft

# A Transition: Computing Power + Data

As individuals, we have more computing power than the fastest supercomputers once provided to a select few

We have enough computing and enough data that when combined with the power of the cloud, new kinds of experiences can emerge

**INTERNET**
1993-present
*Email*
*Web browsers*

**GUI**
1985-present
*Desktop publishing*
*Multimedia*

**DOS**
1981-1995
*Spreadsheets*
*Word processors*


TODAY


Microsoft

# Not Everyone Can Have An Assistant… Or Can They?

I'll research options, and let you know what will work.

There is only one connection that will work, but it requires you to stay an extra day.

I'll adjust your schedule, so that meetings don't conflict.

I'm running l… clients & staf… w.

I need to be … next week fo… Can you arra…

The team wil… in London by…

A Good Assistant:

- Leverages "Memory"
- Anticipates
- Holistically Completes Tasks
- Senses Emotion
- Recognizes Patterns

# VIDEO: The Future

http://www.officelabs.com/projects/futurevisionmontage/Pages/default.aspx



Microsoft

# Work/Office

Fully active workspace, leveraging sensors and projectors

Rich visualization, natural gesture 3-D space, touch and speech recognition

Ambient, translucent displays (OLED)




Integrated workflows, ability to access and share work via the cloud

Seamless interface with devices and cloud services
Rich graphics

Simple devices on surface to access and share data

# Mobile

Integrated life experience: smart device + cloud
= shared calendars, social spaces, next-gen interface

Digital boarding pass, touch from beneath,
dynamic information from the cloud


Smart device + cloud—preference-based communication routing (no decision-making)

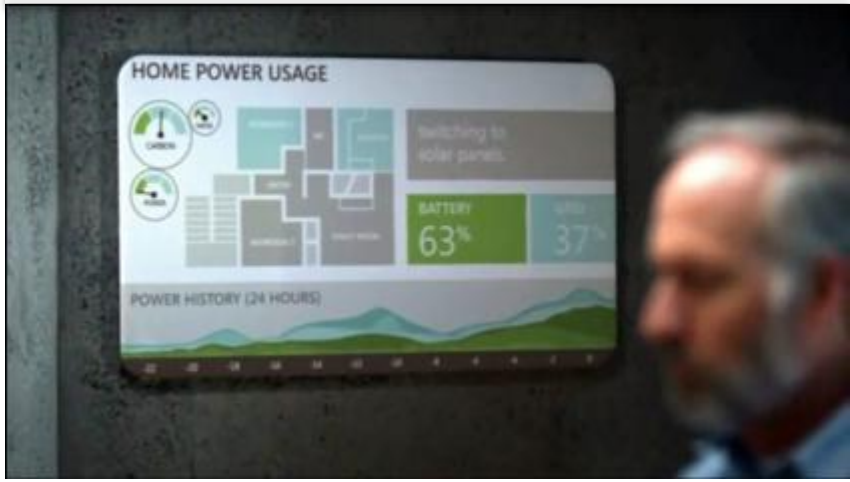
Location-based services, device projector within phone combined with sensor networks, context info retrieval, GPS, compass to “assist” in finding location

# Home

Different form factors, seamlessly integrated

Digital paper, OLED flexible displays, form factor – conductive battery set-up

"Smart" Home…




Switch to task pane for work; tasks pushed via personal profiles and data sharing in cloud

Switch surface, interconnected devices, Internet of Things

# The Connected Home and Lifestyle

Local Distributed Generation (Wind, Solar, etc…)

Cloud Energy Services

- Remote Control
- Reporting, Analytics, Alerts
- Remote Diagnostics
- CRM/Billing
- Competitive Retail Offerings
- Appliance Diagnostics and offers
- Smart Energy Wizards

Utility Z

Meter (bi-directional power flows)

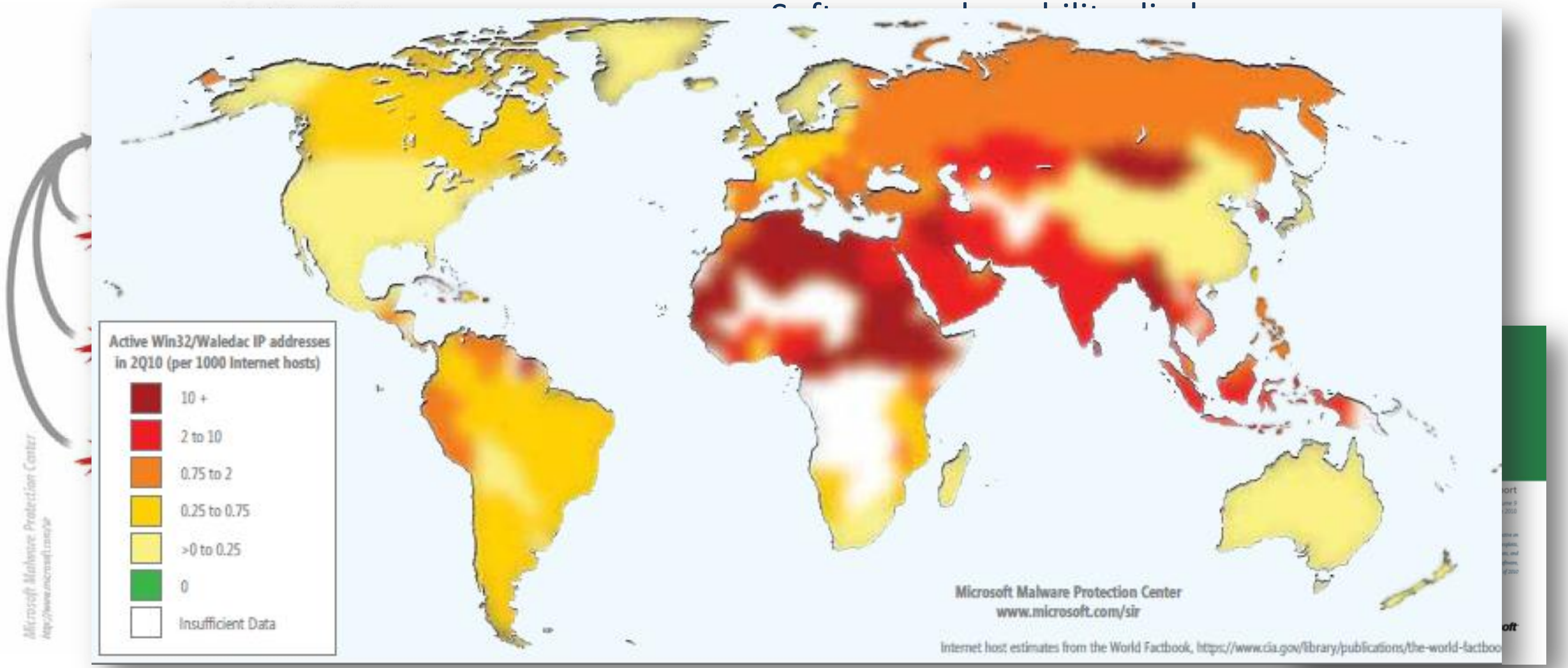
Home Energy Management System

Managed Circuits:

- HVAC Systems
- Hot Water
- Pool Pump
- Comfort lighting
- Accessories
- Other loads

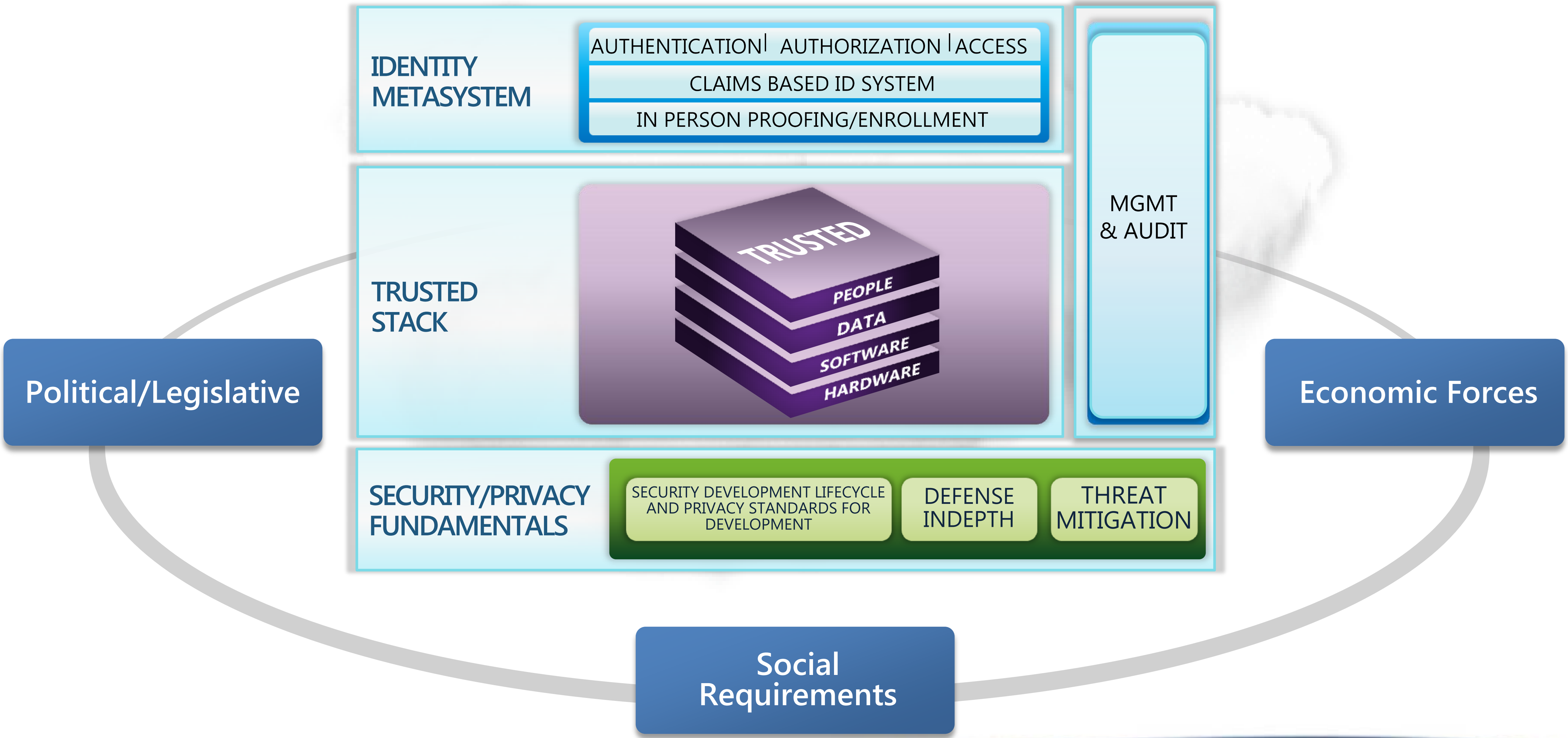
Hybrid/Electric Vehicle Interface

# Security and Privacy: The Threat Ecosystem




www.microsoft.com/sir

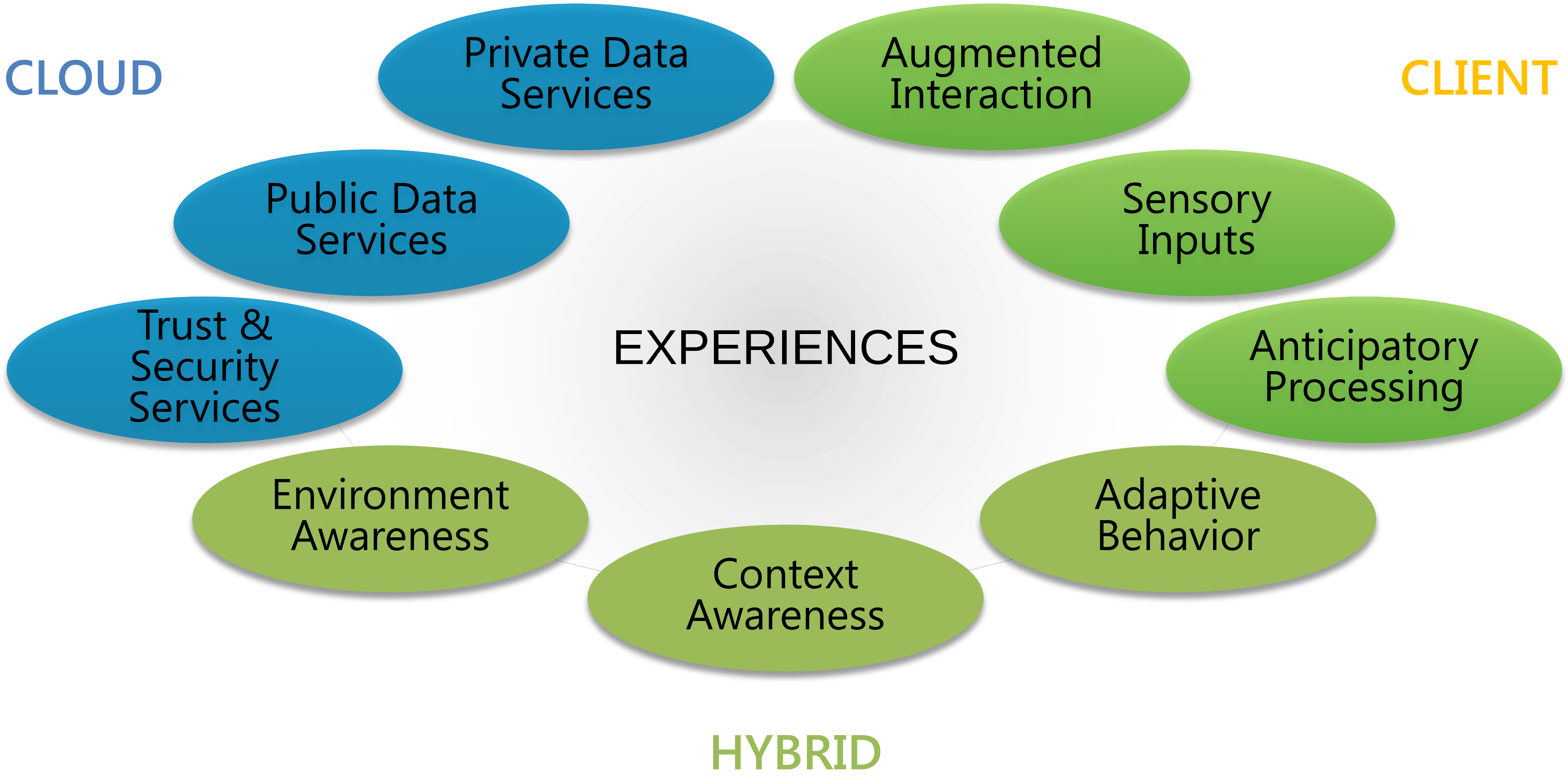
# Security, Privacy and End-to-End Trust






Microsoft

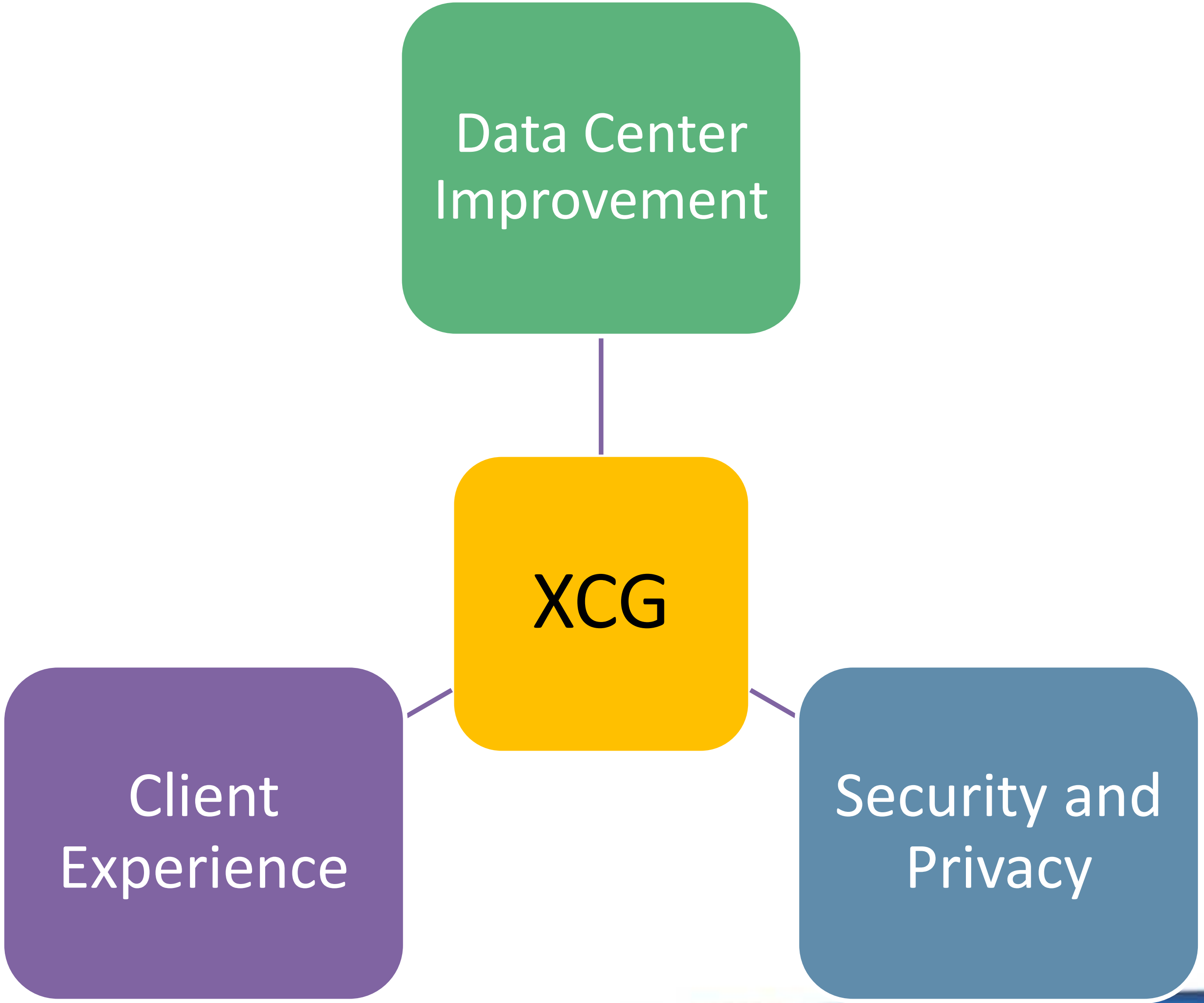
# The Future of Experiences





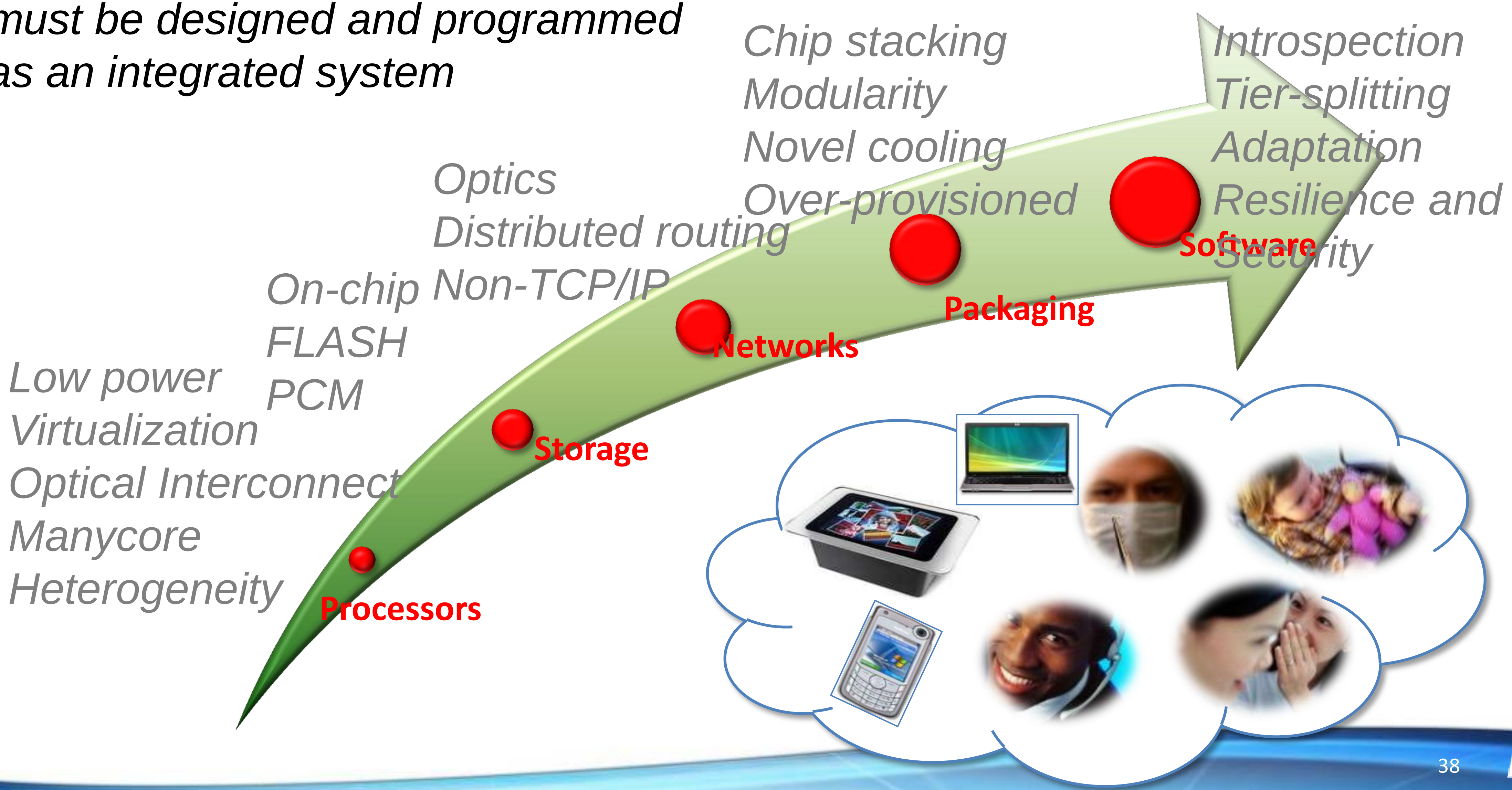
Microsoft

# eXtreme Computing Group (XCG)



# XCG Vision: Rethinking Everything

**Philosophy:** *Cloud infrastructure must be designed and programmed as an integrated system*

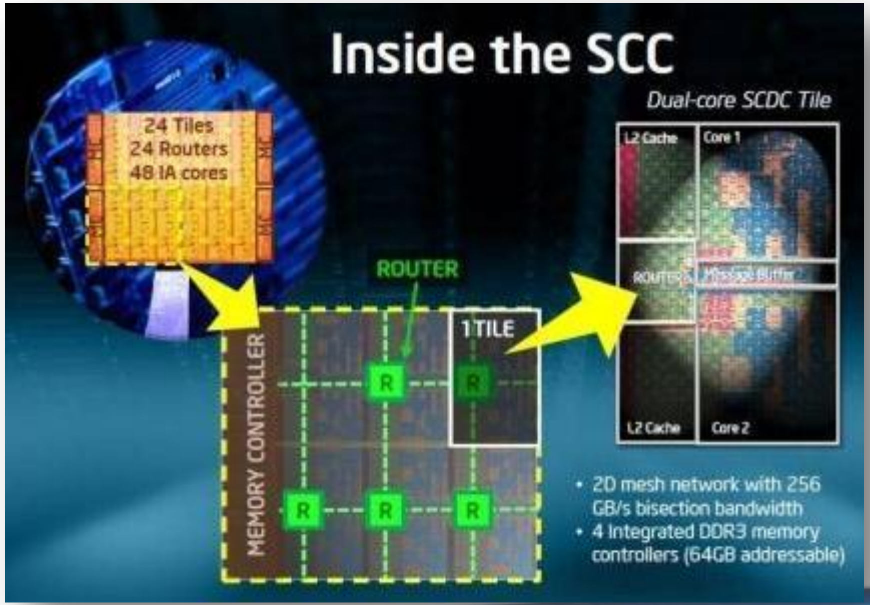




Microsoft

# Rethinking Node Architecture

- Low power systems for web search
  - Workload adaptation and SLAs
- Intel Cloud Chip
  - 48 x86 cores
  - Software power control
    - Energy management
  - On-chip mesh network
    - Low latency
    - High bandwidth

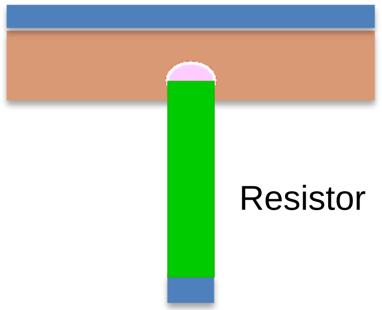





Microsoft

# Rethinking Storage: Beyond Disks/FLASH

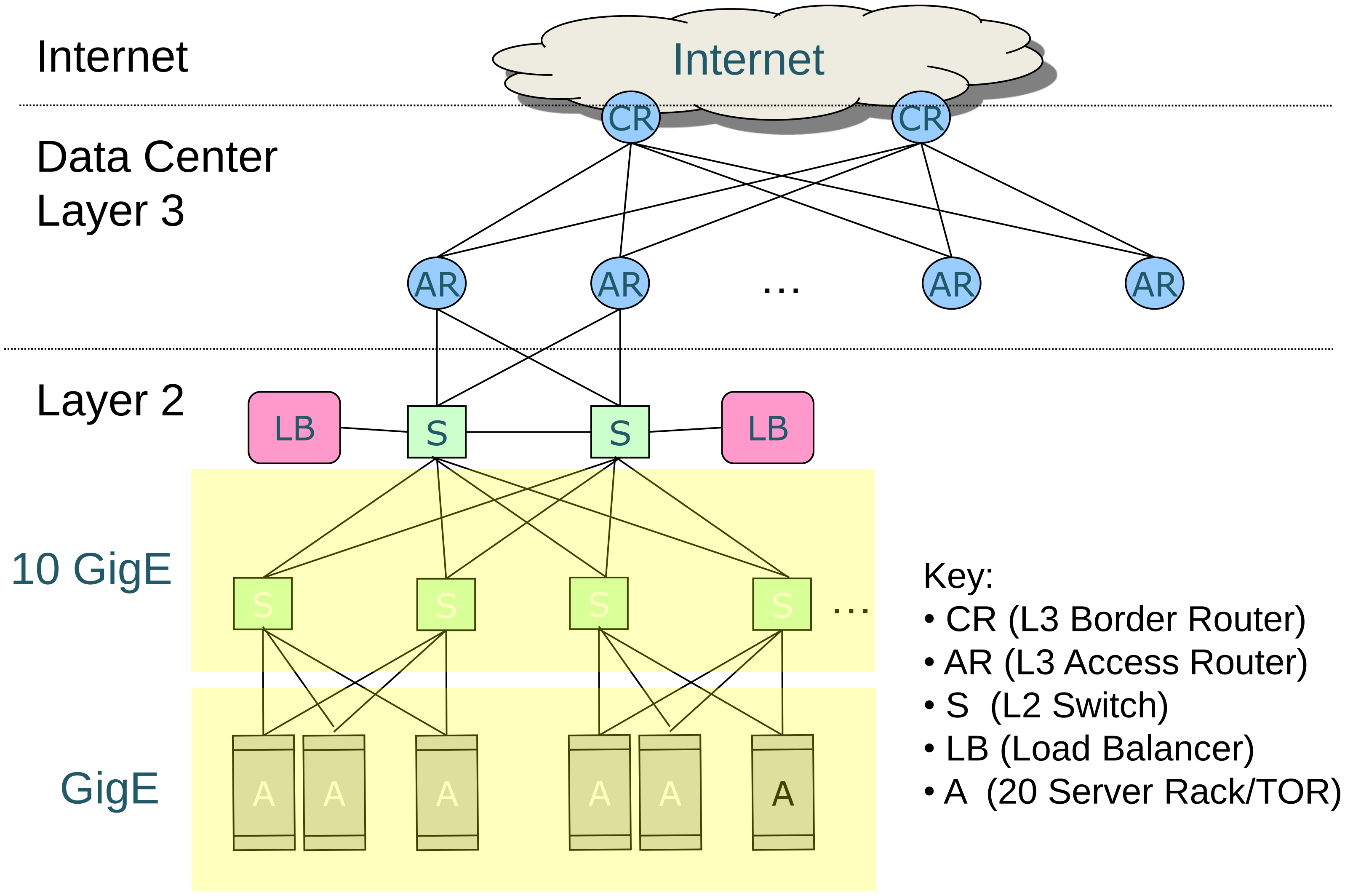
- The tyranny of disks
  - Last mechanical component
  - Most common failure mode
  - Capacity/bandwidth mismatch
- Rethink the storage hierarchy
  - Mixed processes and DRAM
  - Chip stacking and PIM
  - NVRAM futures
    - FLASH (transition)
    - Phase change memory (PCM)
      - Crystalline (1) and amorphous (0) states
      - Word/byte addressable with lower latency




Museum)


Microsoft

# Current Cloud Data Center Networks

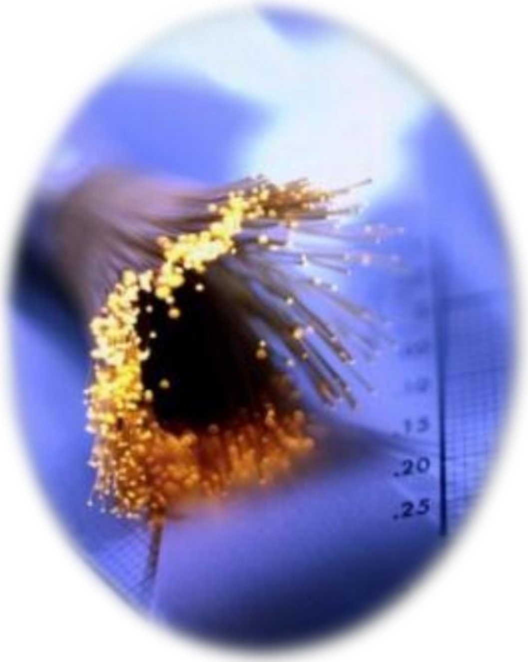




Microsoft

# Rethinking LAN/WAN Networking

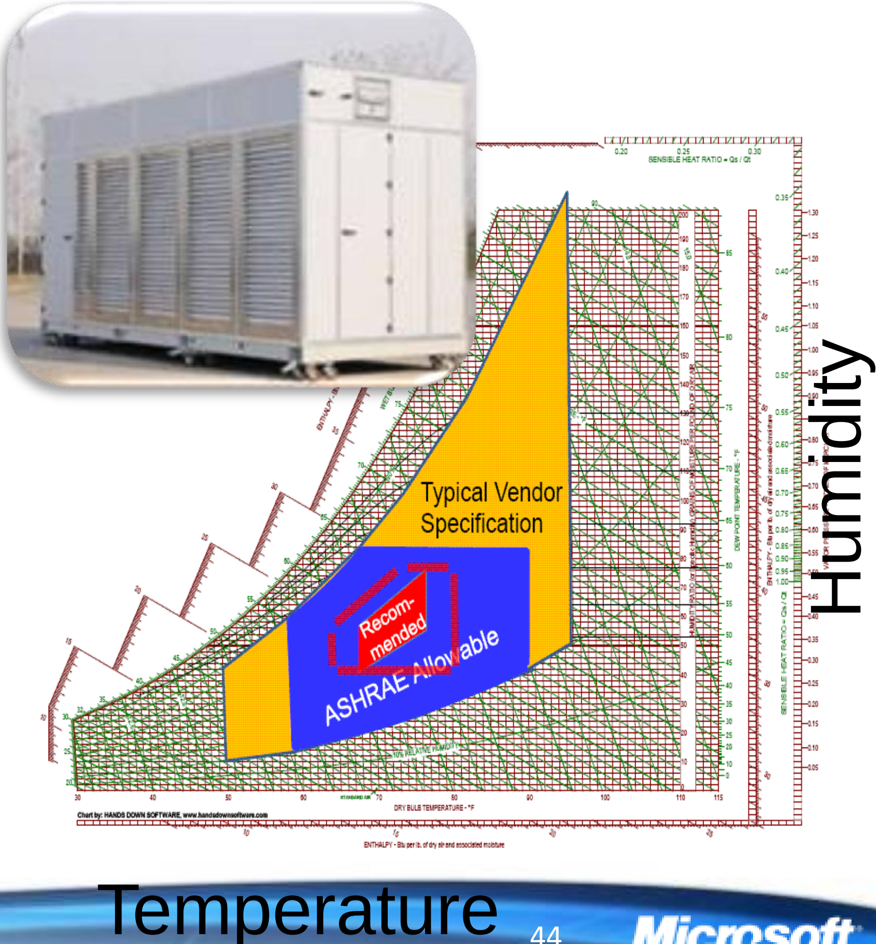
- Break the LAN hierarchy
  - Multiple paths, commodity components
  - High bisection bandwidth
- We build WAN islands, not continents
  - Isolated facilities with limited connectivity
- Change the landscape
  - Serious, multiple terabit WANs
  - Many lambdas entering a facility
  - Fused node/LAN/WAN infrastructure



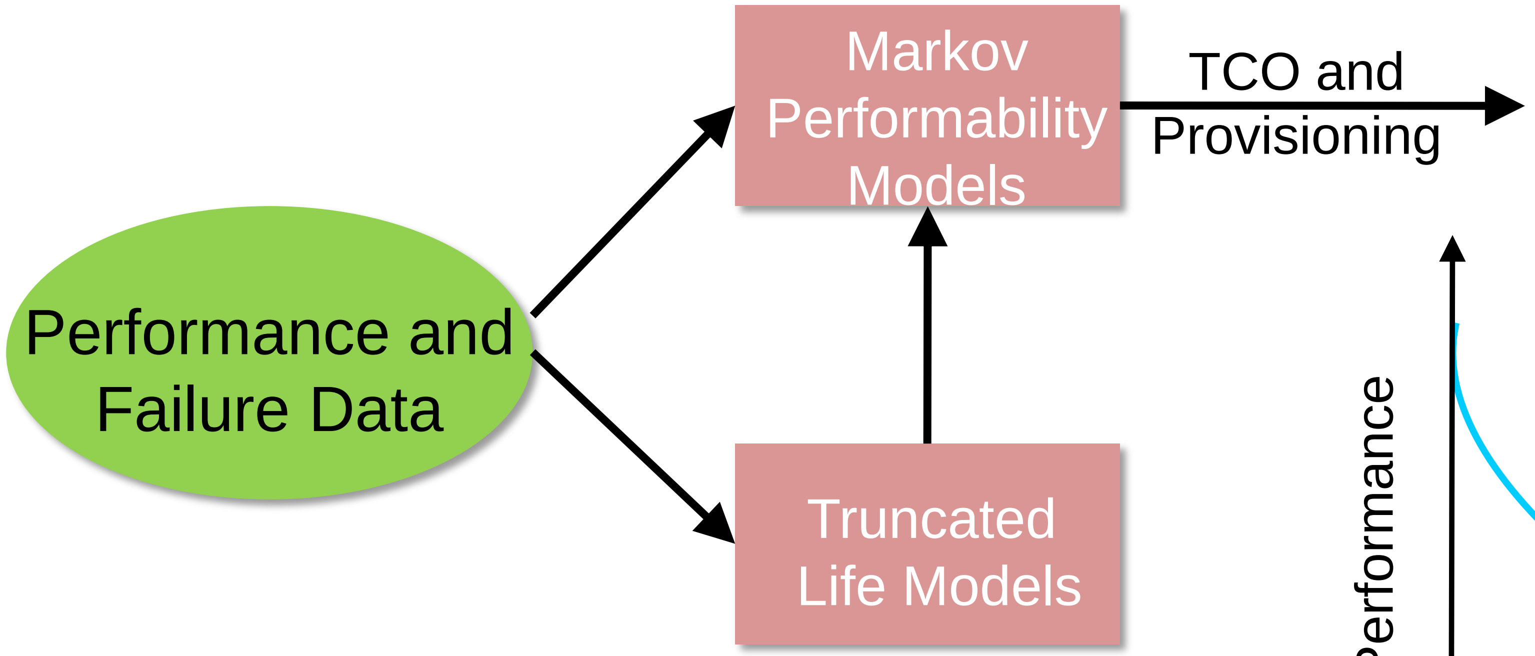

Microsoft

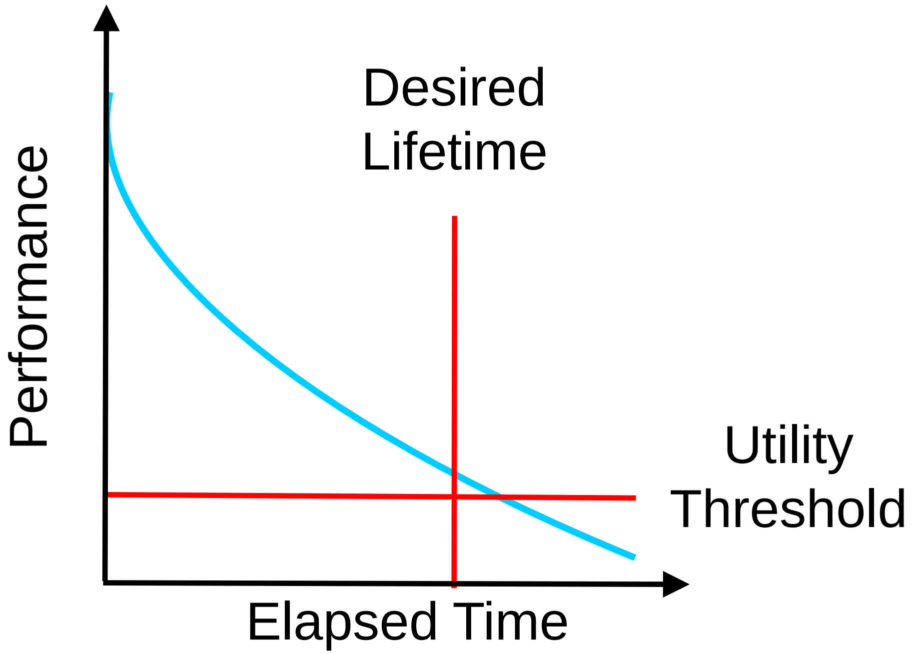
# Rethinking Packaging and Cooling

- People and hardware need not mix
- Hardware cooling standards are conservative
  - Reliable at high temperature/humidity
- Optimize for efficiency
  - Cooling is (often) unnecessary
    - Design for ambient environments
  - Energy reliability is (often) unnecessary
    - Design for power outages
  - Use larger building blocks
    - Accept component failures



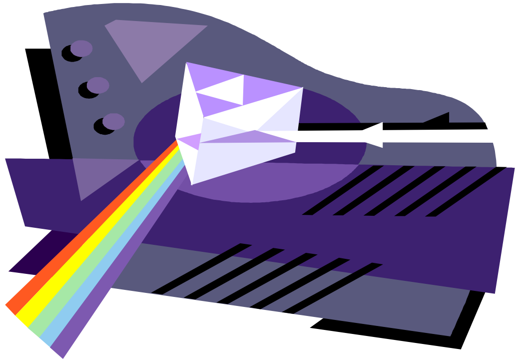
# Rethinking Reliability: Fail In Place




- Factory sealed units (FRUs)
  - Over-provisioned for failure
  - Dynamic reconfiguration
  - Real-time, adaptive control

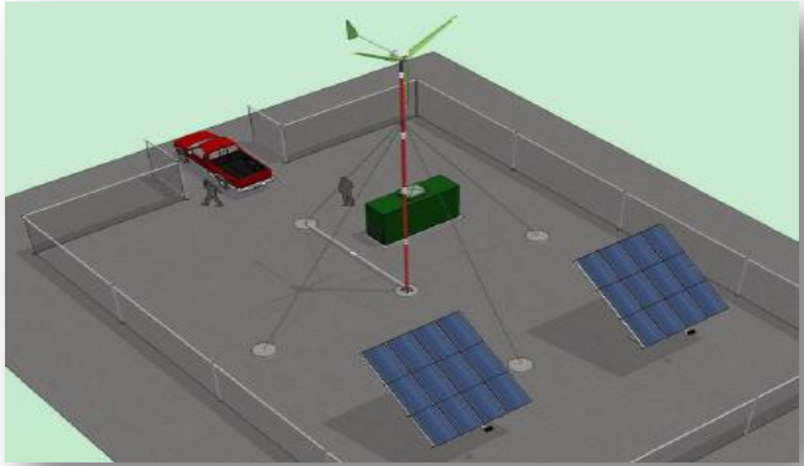
# Rethinking Energy Provisioning

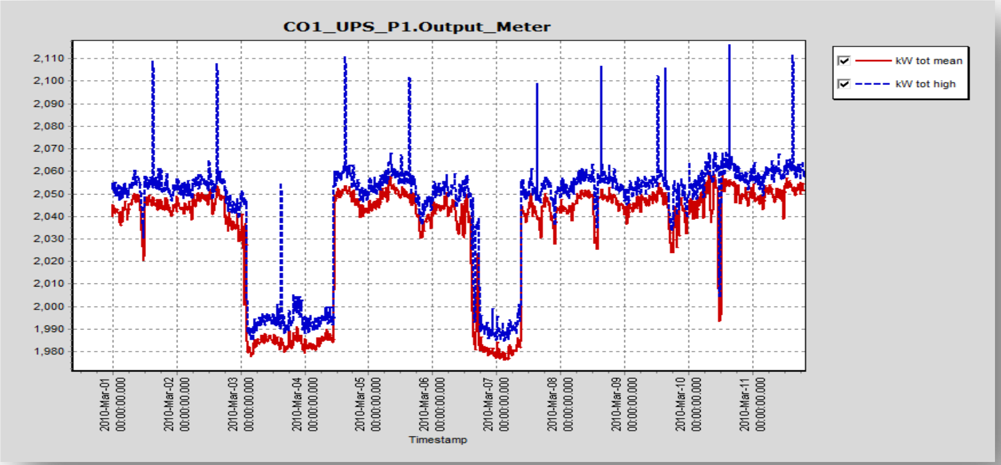


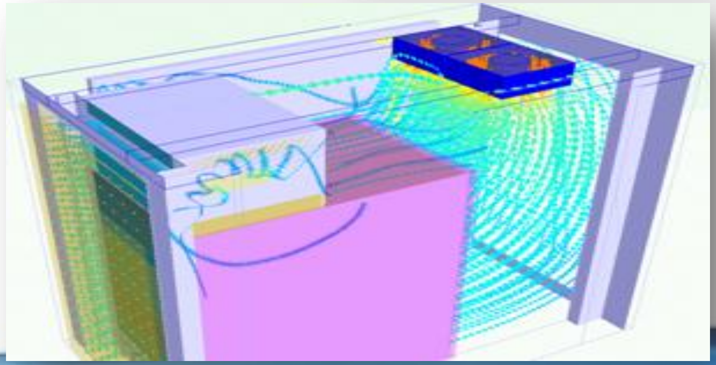
- Power redundancy is a major cost
  - Batteries to supply up to 15 minutes
- Use multiple sites, based on energy cost and carbon footprint
  - Electrical grid, solar, wind, fuel cell, …
  - Workload dispatching based on models
- Real-time optimization and prediction
  - Workload demand
  - Weather and seasonal models
  - Auction-based energy pricing
  - Infrastructure
    - UPS, optical fiber and computing

# Micro Datacenter Prototype

- Early test vehicle
  - 1500 W target
    - Solar and wind renewables
    - Grid as backup
- Power smoothing

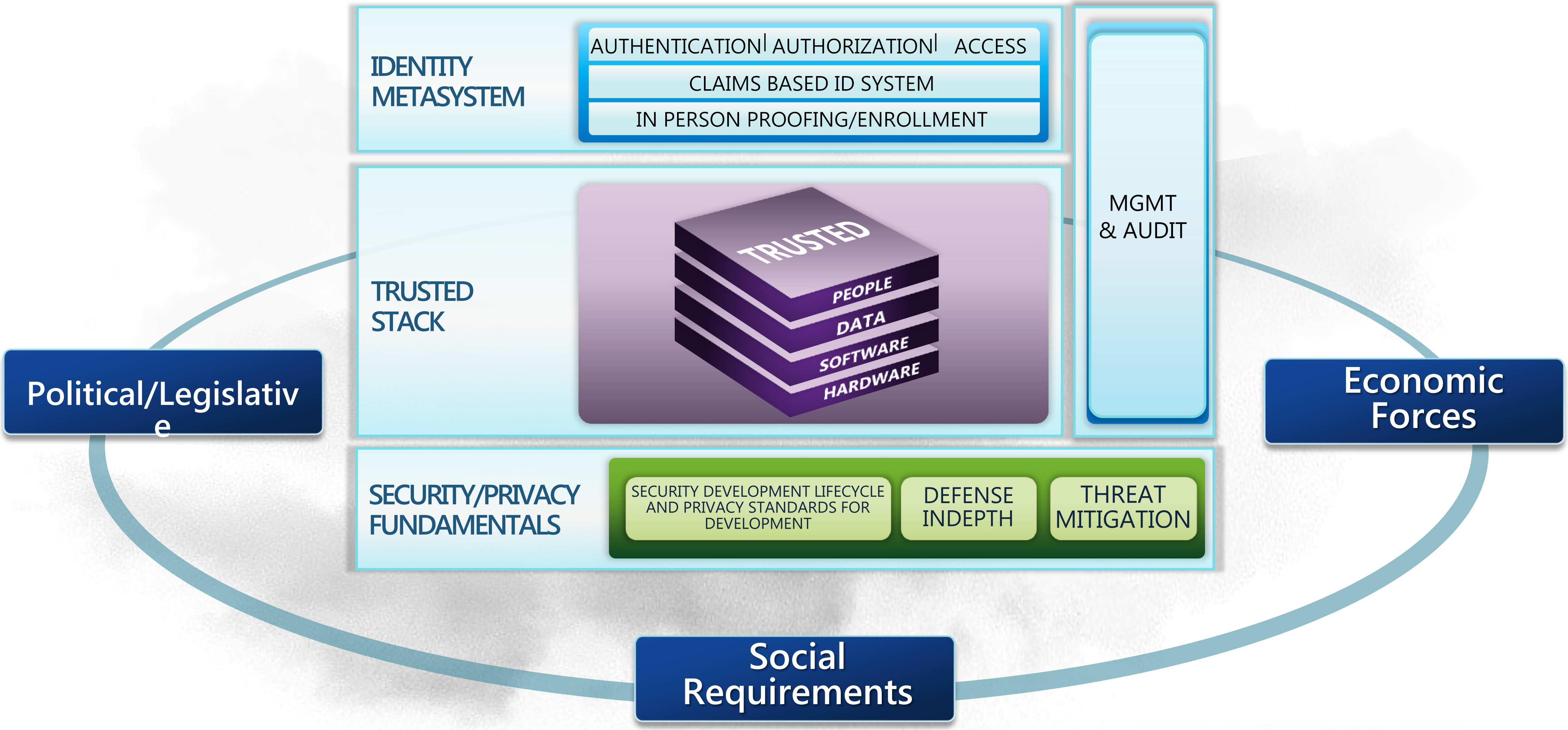




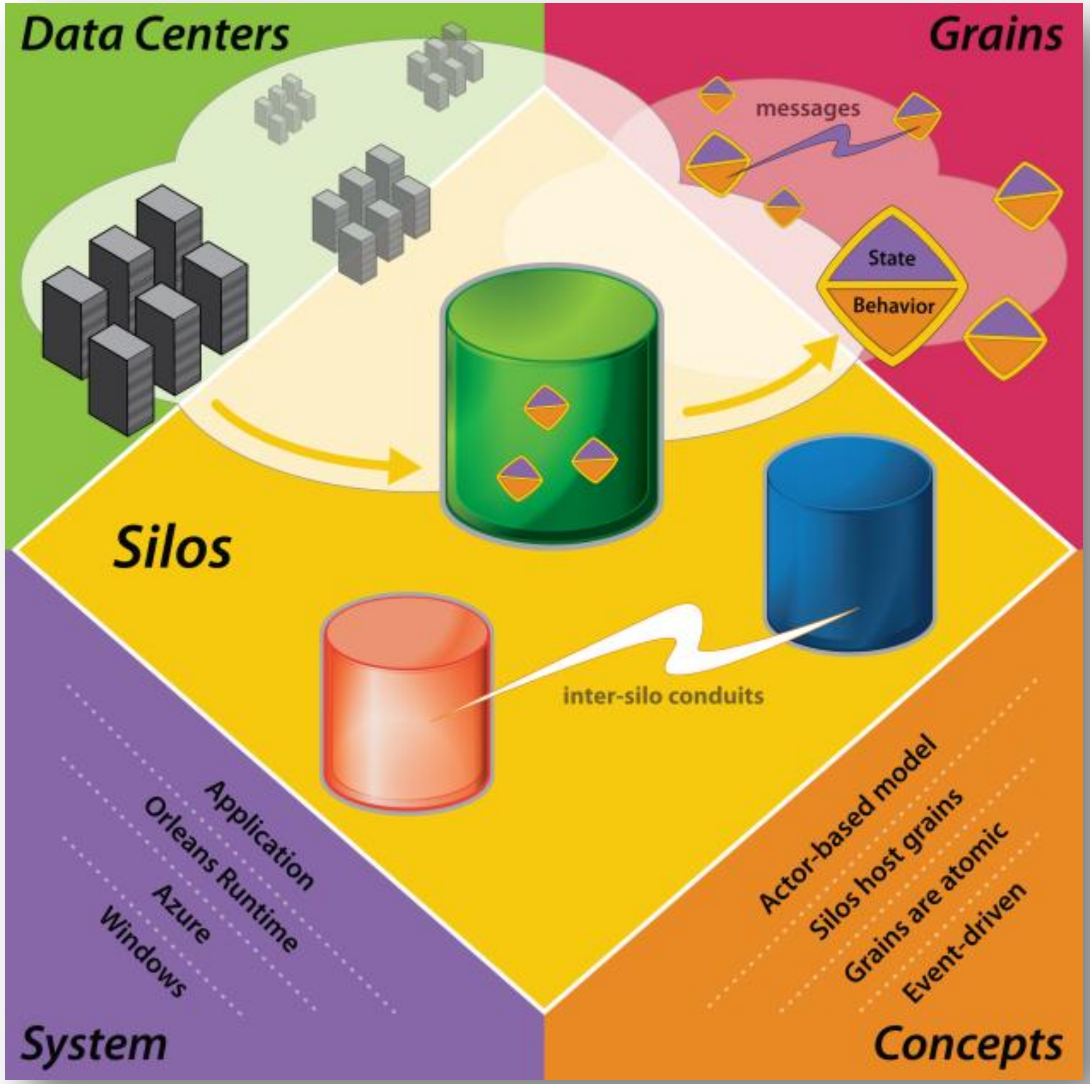

Microsoft

# Rethinking Trust: End-to-End Balance



# Rethinking Data Center Software: Orleans

- Key concepts
  - Grains
  - Activations
  - Message passing
  - Promises
  - Transactions



# A Vision of the Future

- A web of modular cloud infrastructure
  - Intelligent energy management
  - Adaptive failure resilience
  - Rugged environment tolerant
  - Configurable components
  - Designer hardware
- Flexible enabling software
  - Device agnostic adaptation
  - Fine grained mobility
  - SLA driven resource management
  - Secure, attested environment
- Rich and diverse client experiences
  - Mediated by wired and wireless devices
  - Contextually aware and responsive
  - Supported by rich infrastructure

Microsoft®

Your potential. Our passion.™

© 2010 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.
The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.
MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.