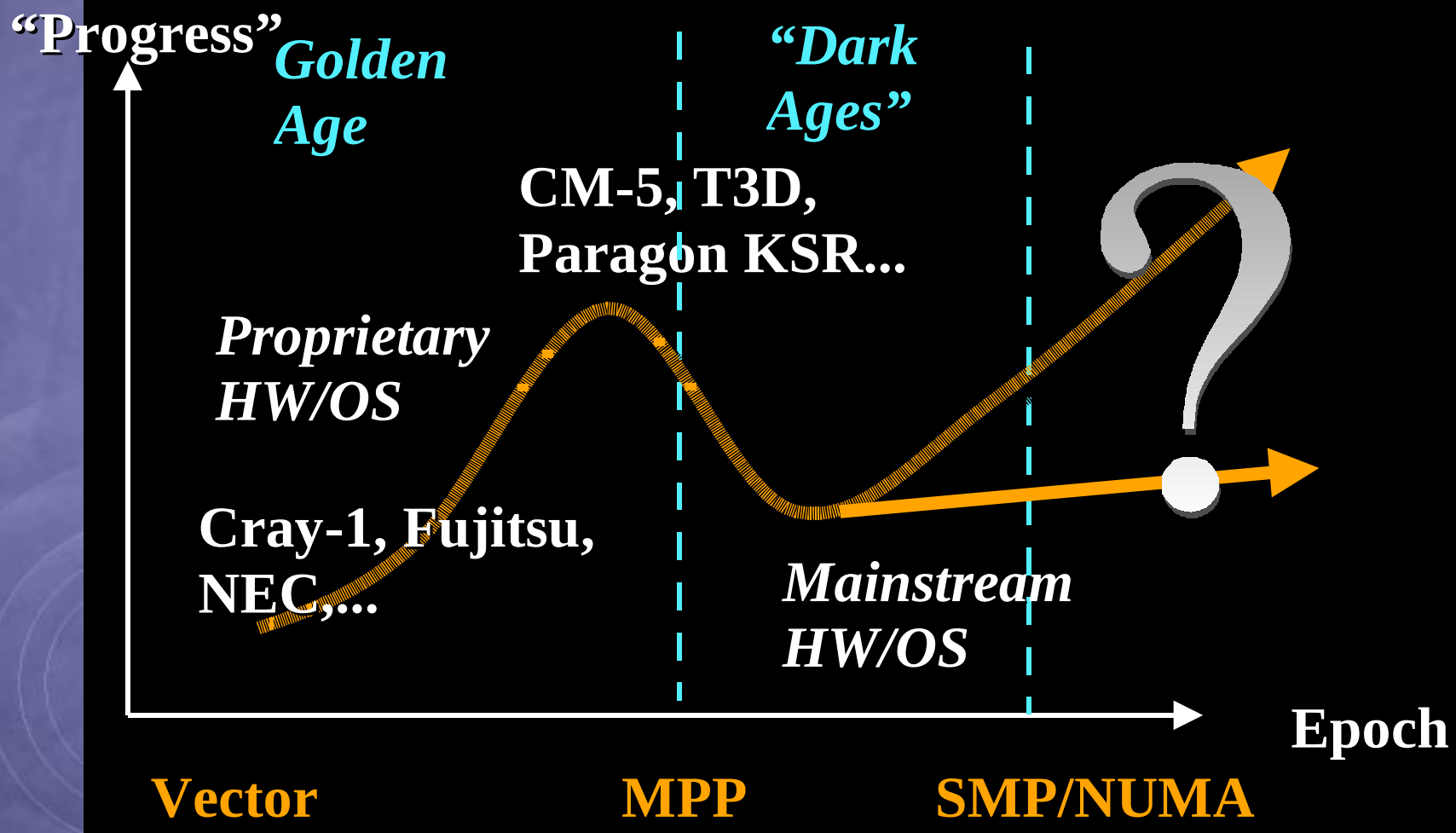
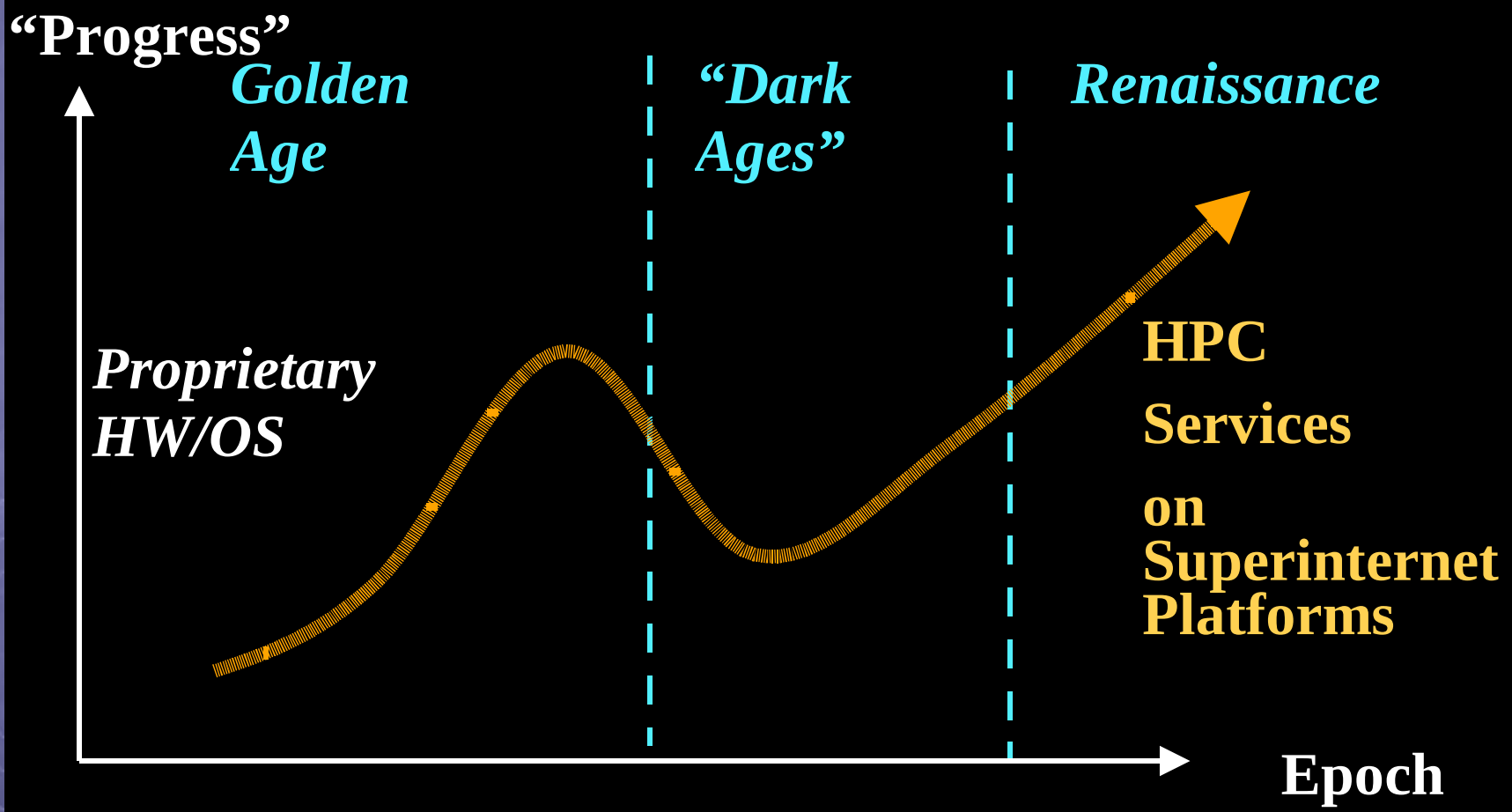
HPC meets .com
The convergence of Supercomputing and Superinternet architectures

Greg Papadopoulos
VP & CTO
Sun Microsystems, Inc

End of the Dark Ages?



The Renaissance: HPC Services on Superinternet Platforms



Thesis

HPC Benefits from Internet Computing Trends

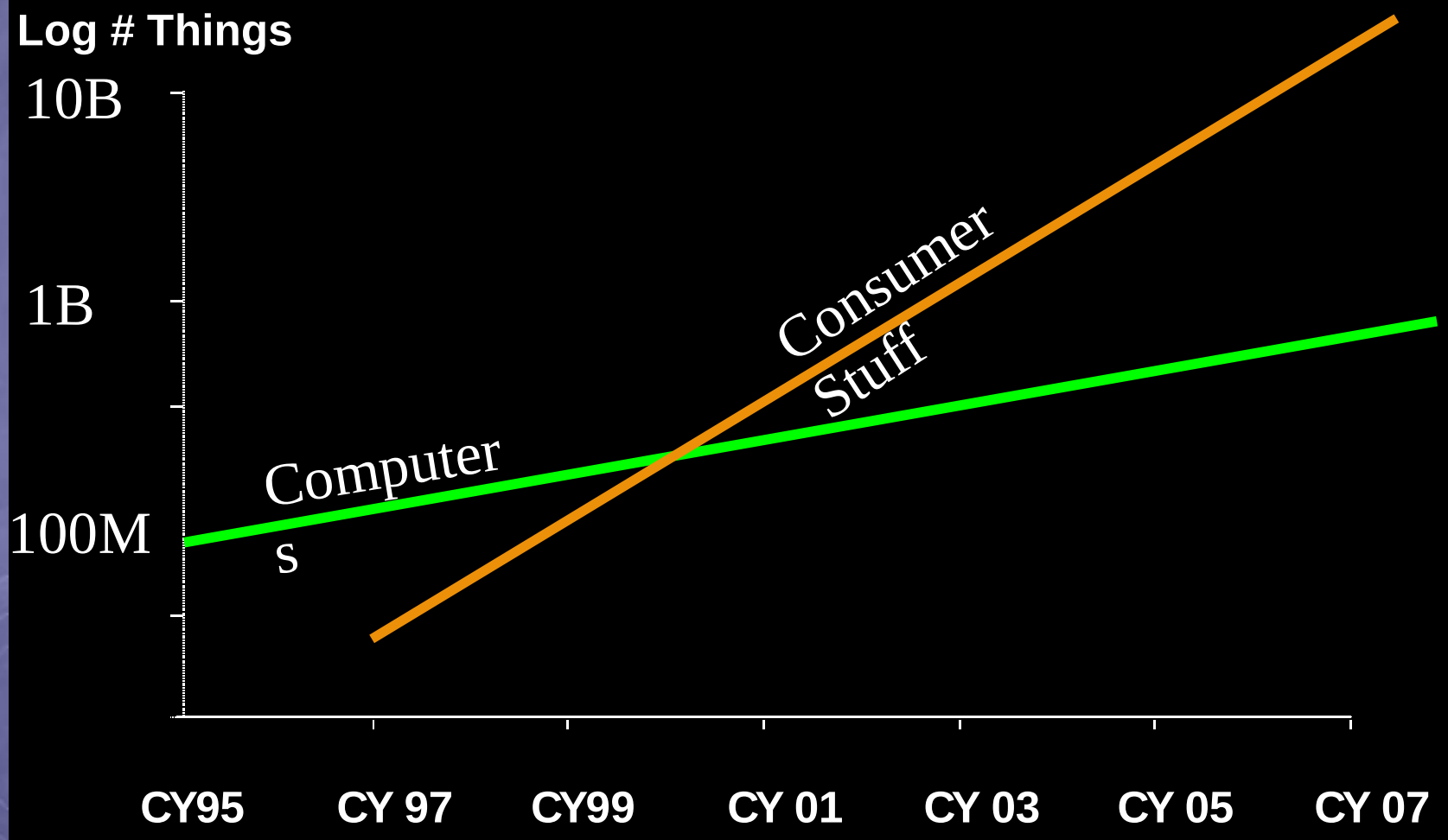
Prediction

HPC Converges with Internet Computing

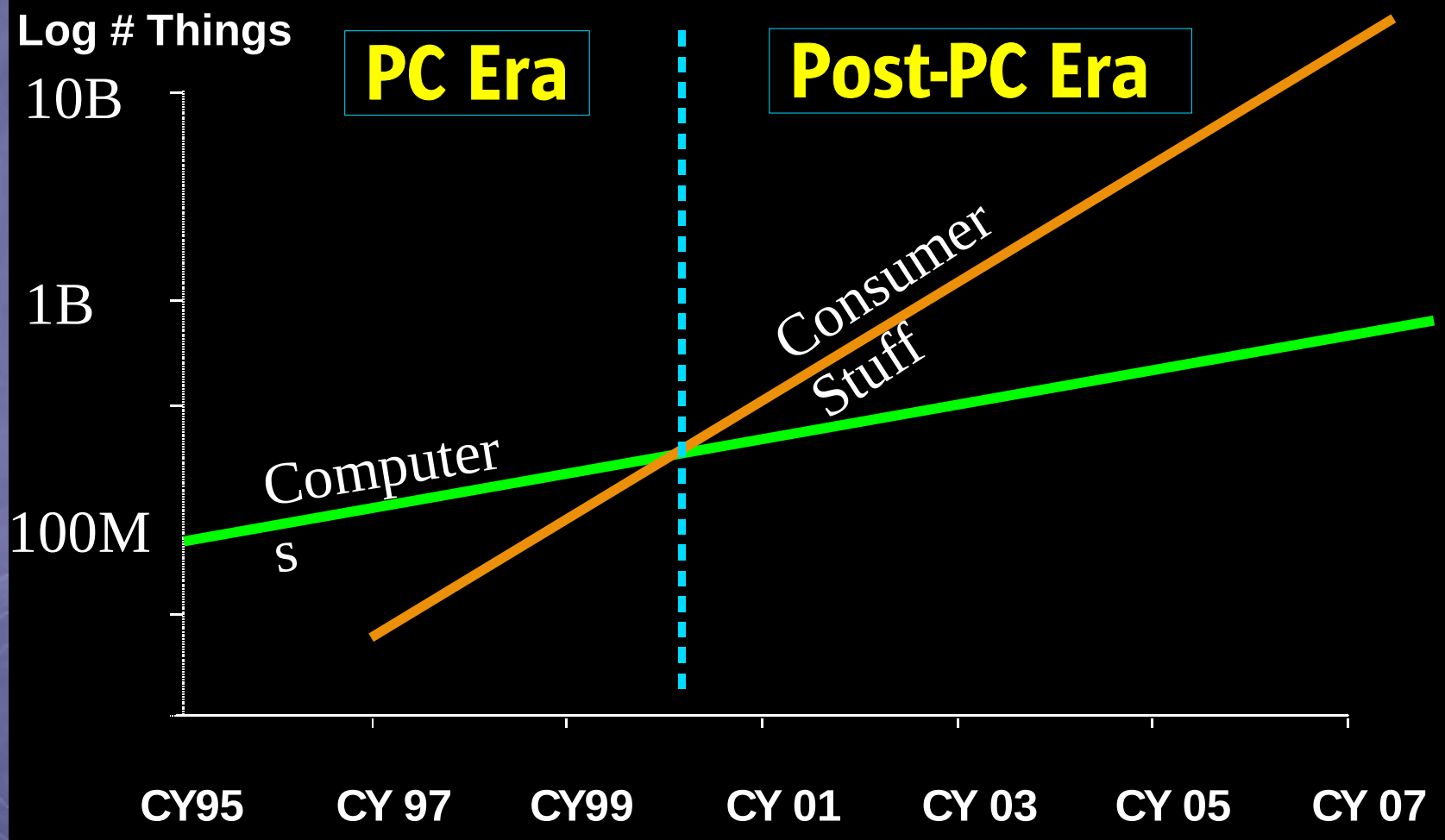
At the platform

Superinternet subsumes Supercomputing

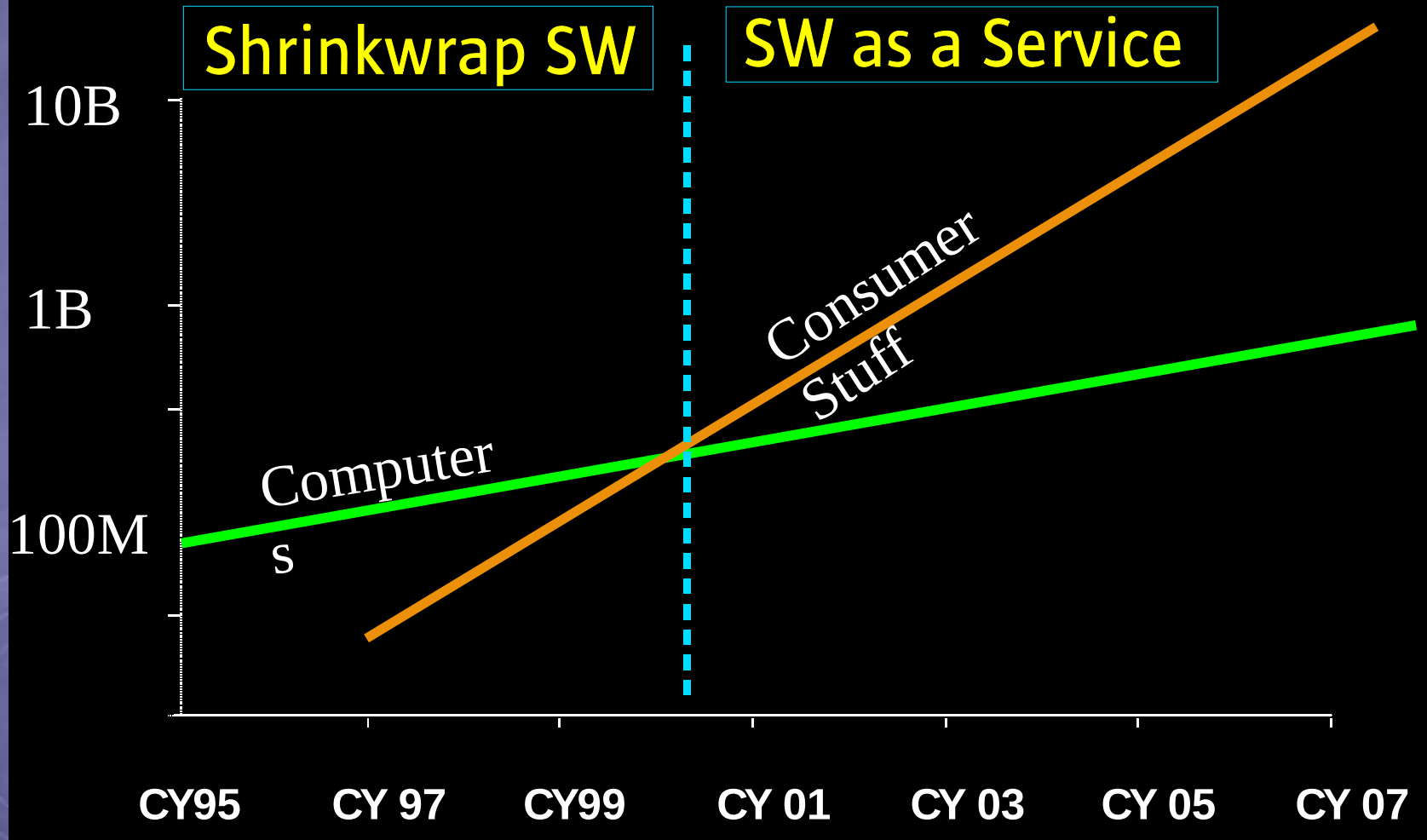
Things Networked



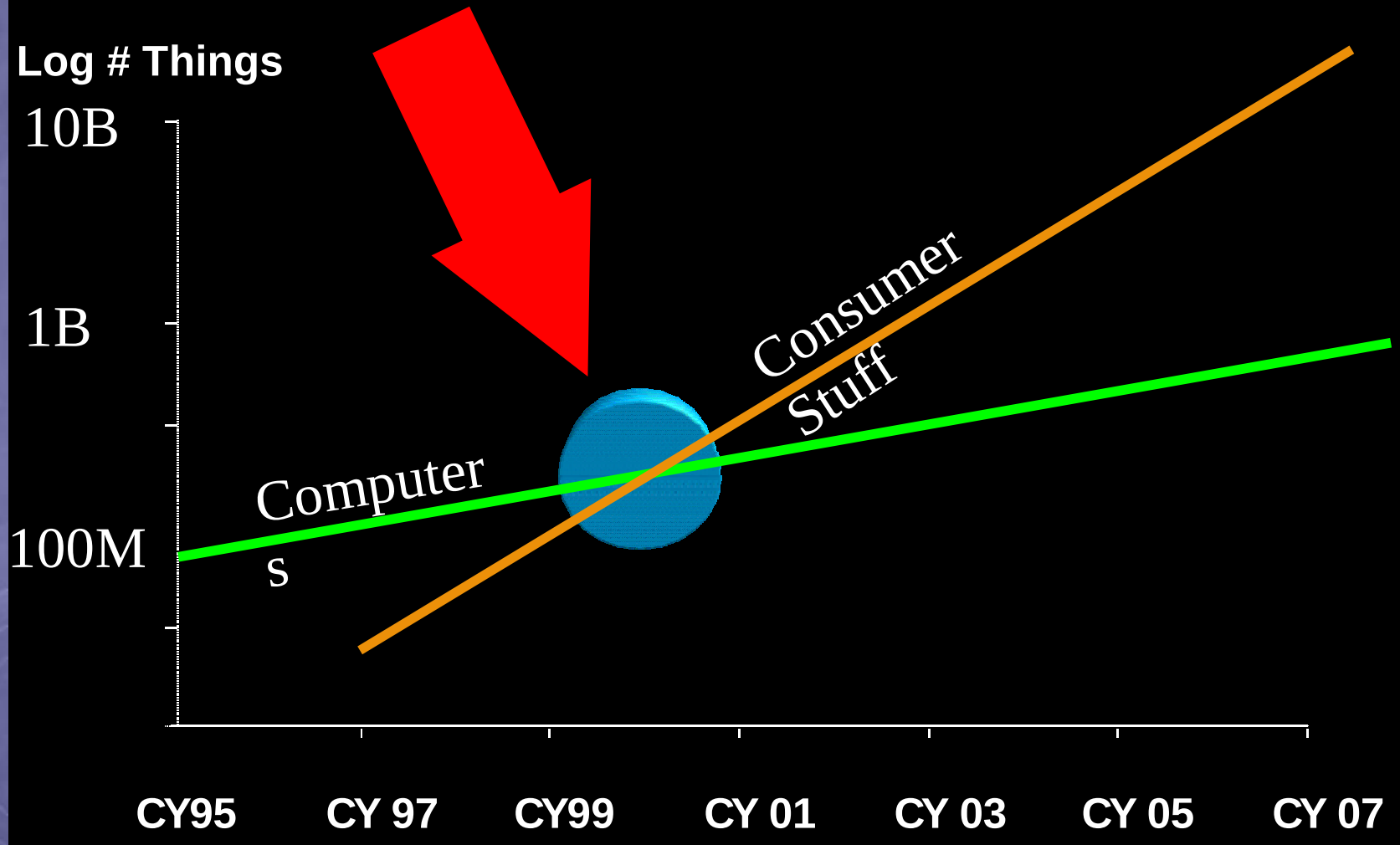
Things Networked



Things Networked



We are Here!

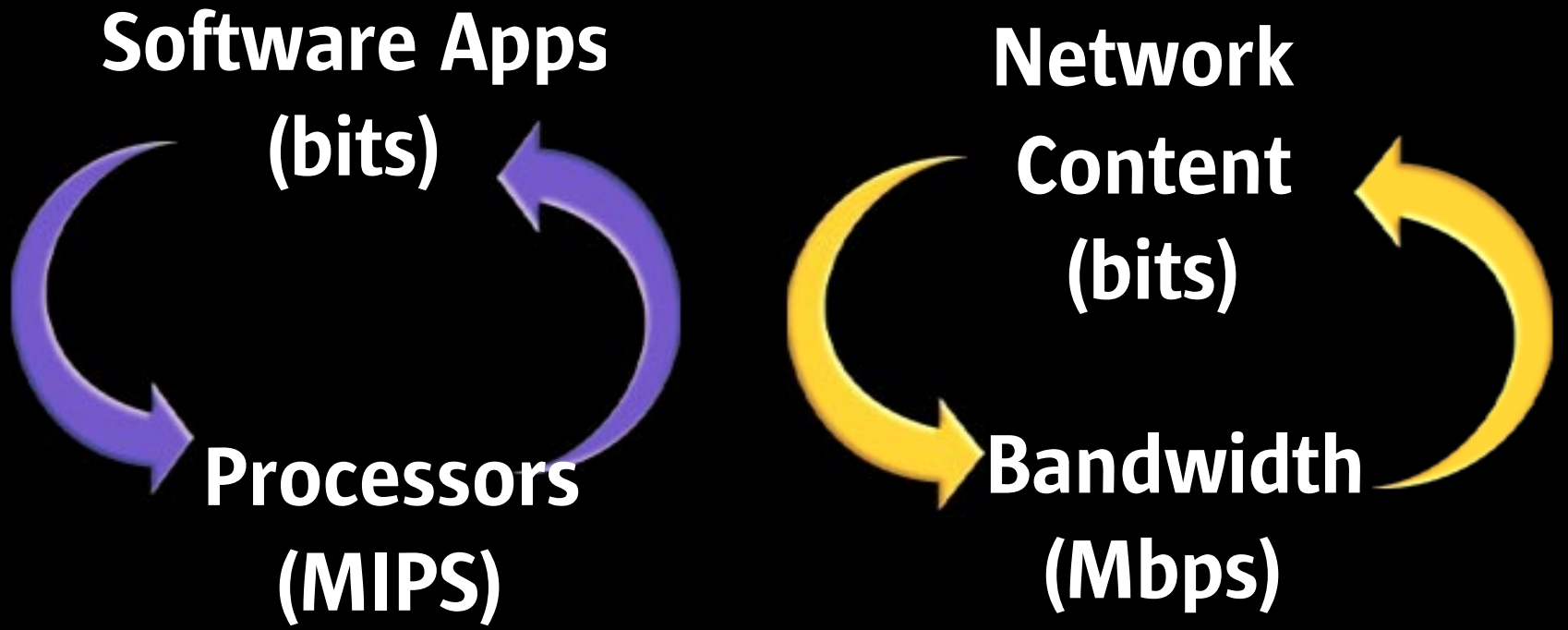

Log # Things
10B
1B
100M
Computer
s
Consumer
Stuff
CY95
CY 97
CY99
CY 01
CY 03
CY 05
CY 07

The Emerging Big Picture

The Challenge

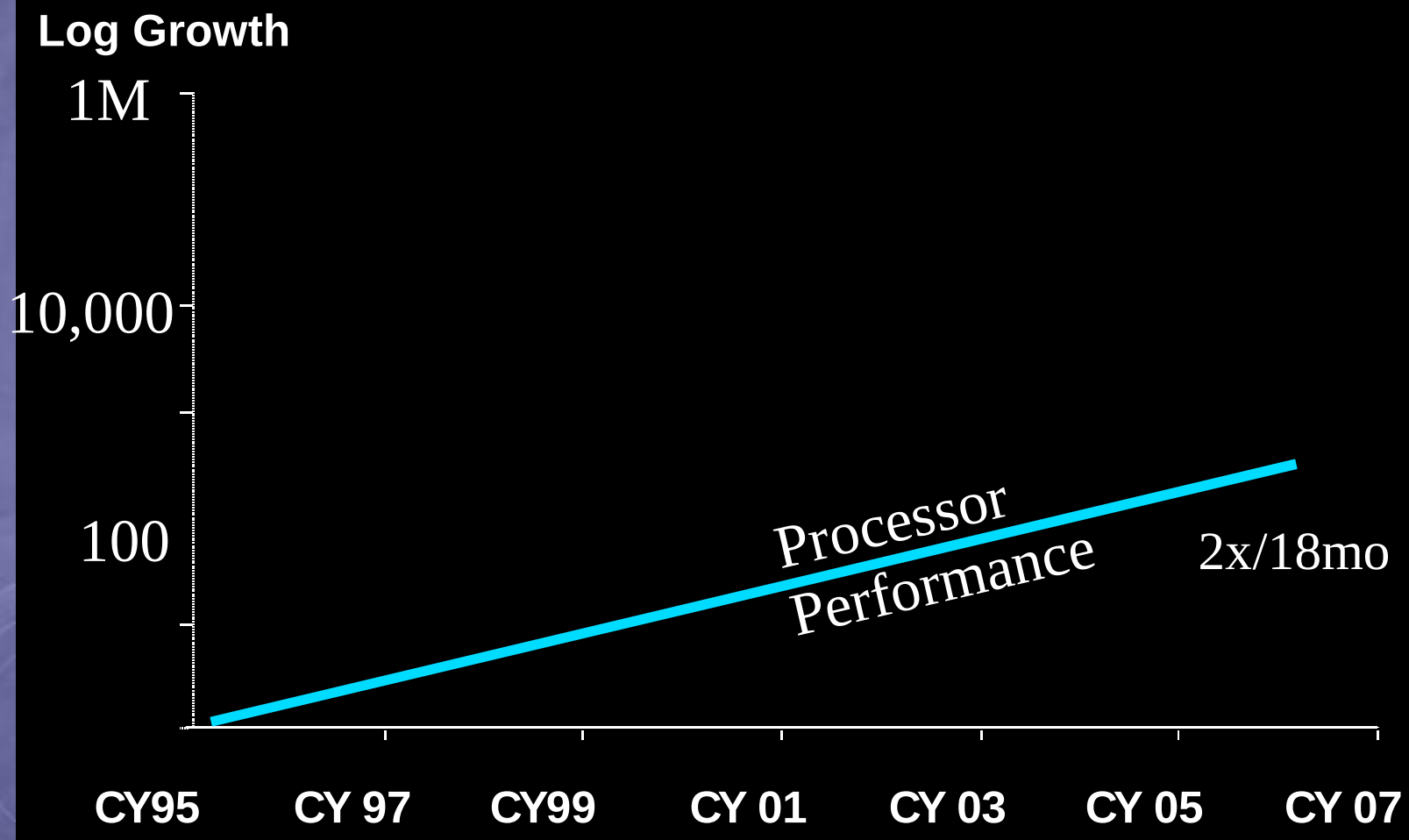
Billions of devices
interacting with
Millions of services,
Predictably,
Securely,
Globally.

The BW Feedback Loop

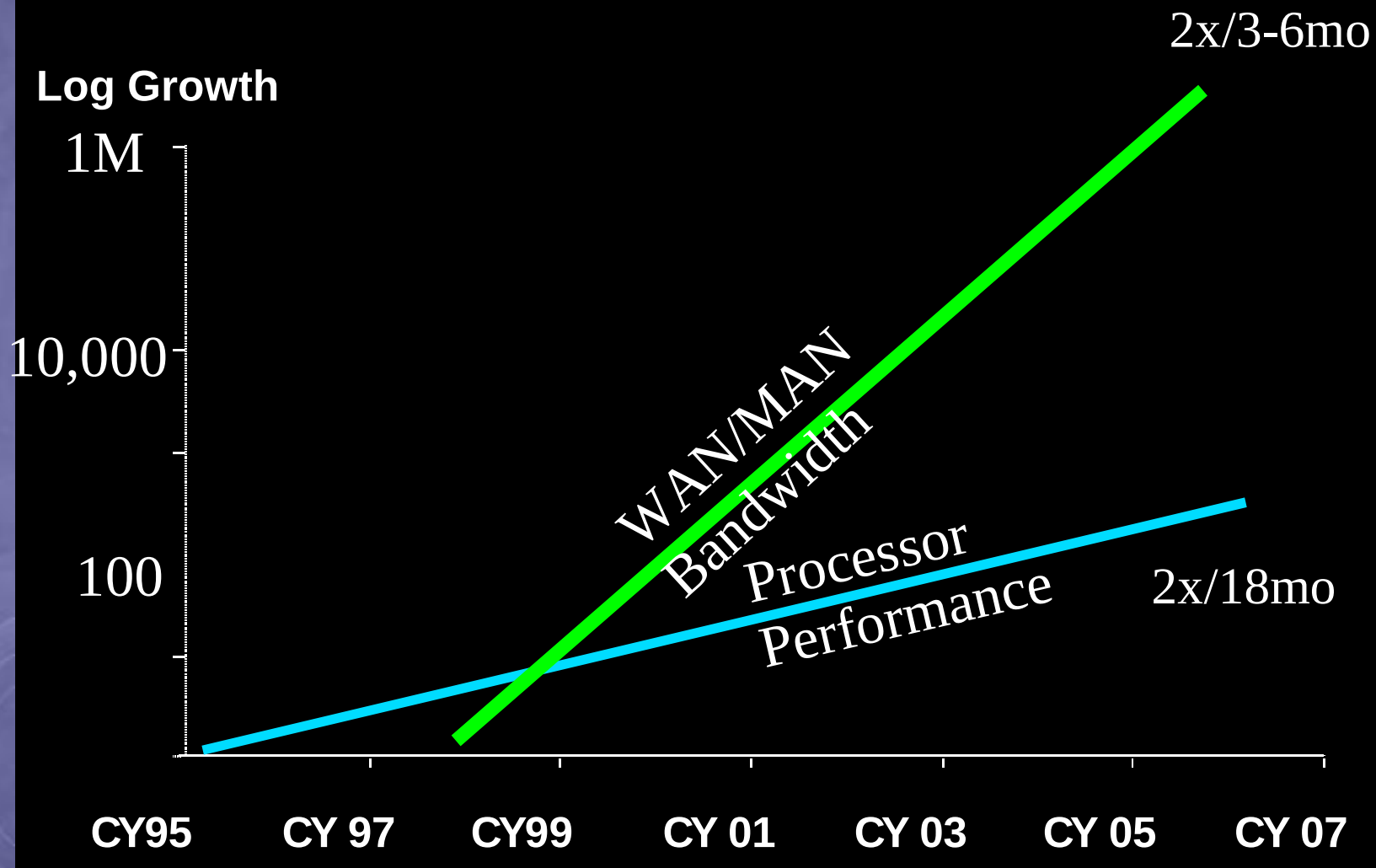


Footnote: The supercomputer technology cycle collapsed

Gilder's vs. Moore's Law



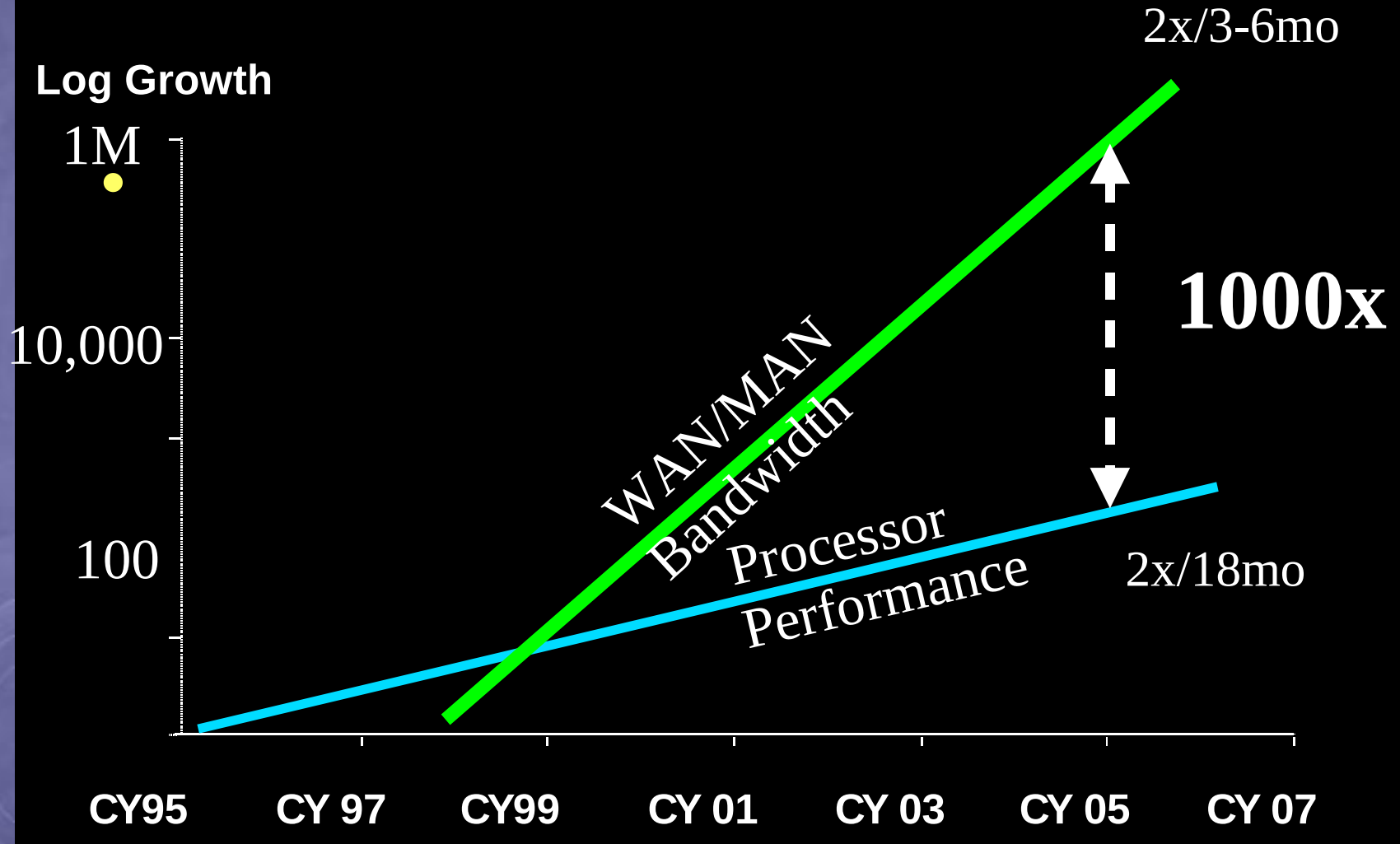
Gilder's vs. Moore's Law



Big Rule

NEVER
Bet Against
Bandwidth

Gilder's vs. Moore's Law



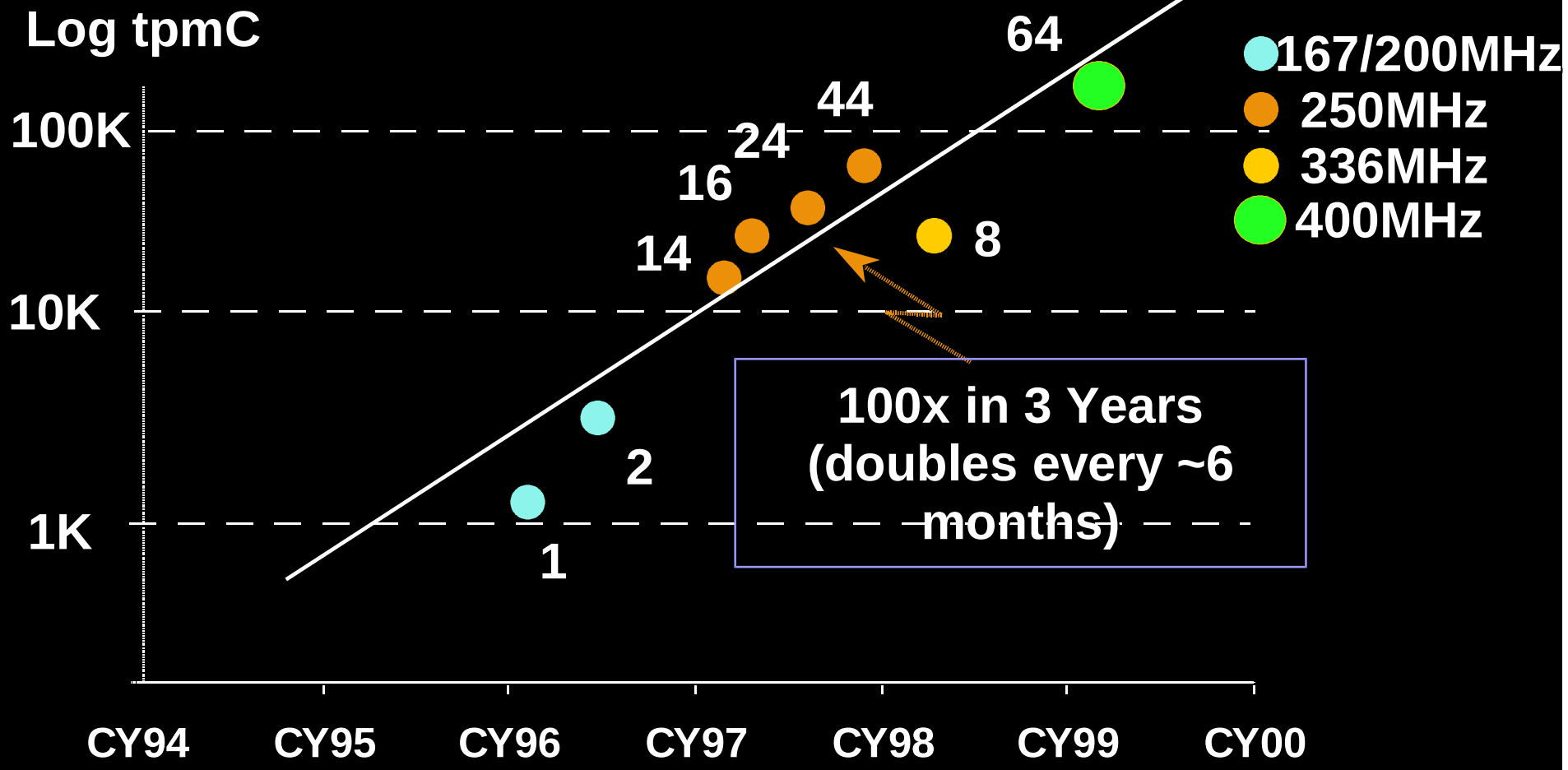
Without a lot of Imagination - 2005

- 10,000 CPU “plexes”
- 10's Millions of simultaneous users
- Multi-terabit network feeds
- 1 Terabyte RAM < $50K

  => 10’s of Terabytes commonplace
- 1 Terabyte DISK < $5K

  => 10’s of Petabytes commonplace

Processor x Scale: *TPC-C*



Key Driver Is Relationship Between

System Scaling and Application Management Complexity

Scalability, Redefined

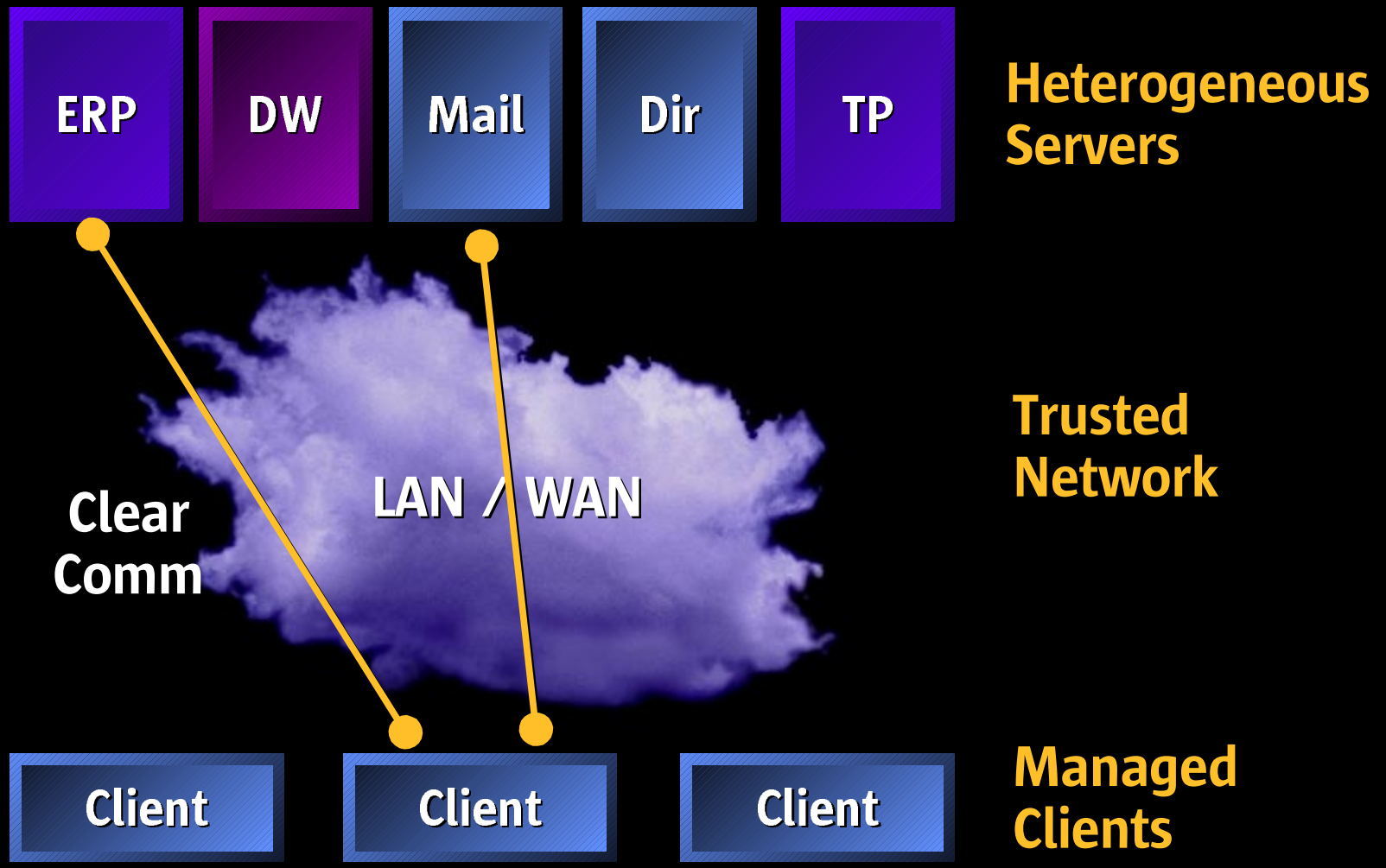
A system Is Scalable Only If You Can Add Resources Without Adding Complexity

Two Questions

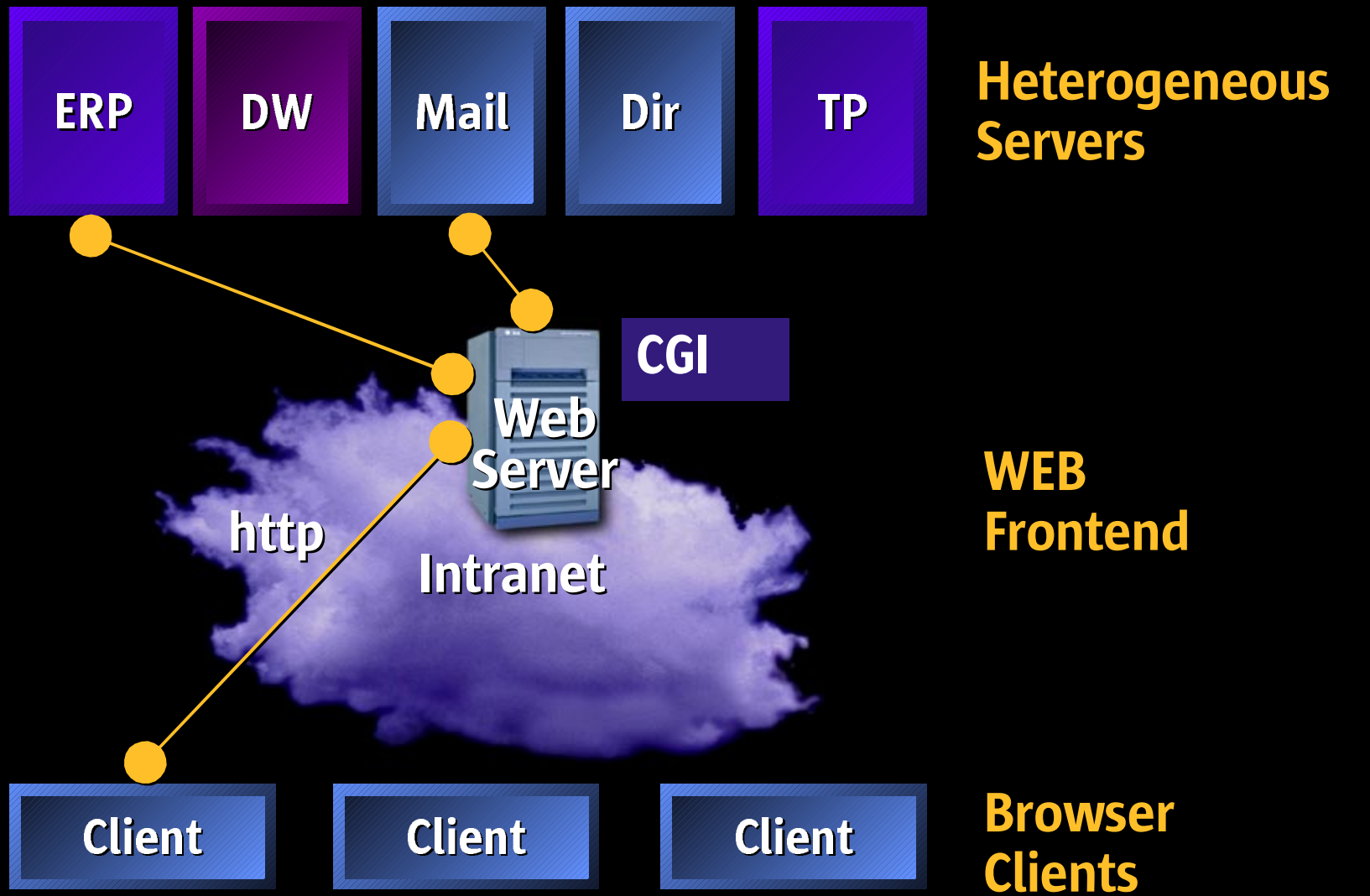
Which Application Technologies?

Which Platform Technologies?

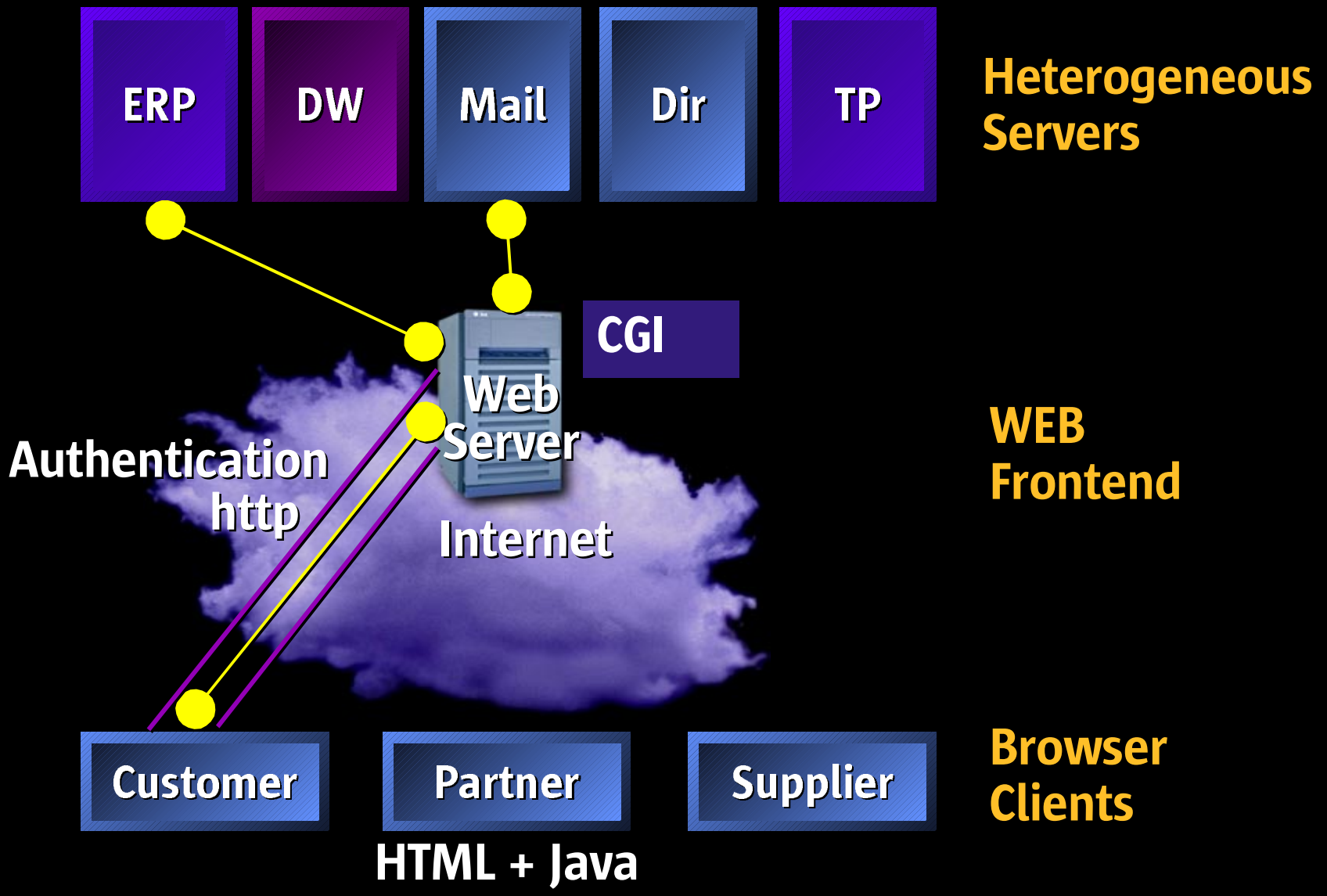
“Old” Client/Server



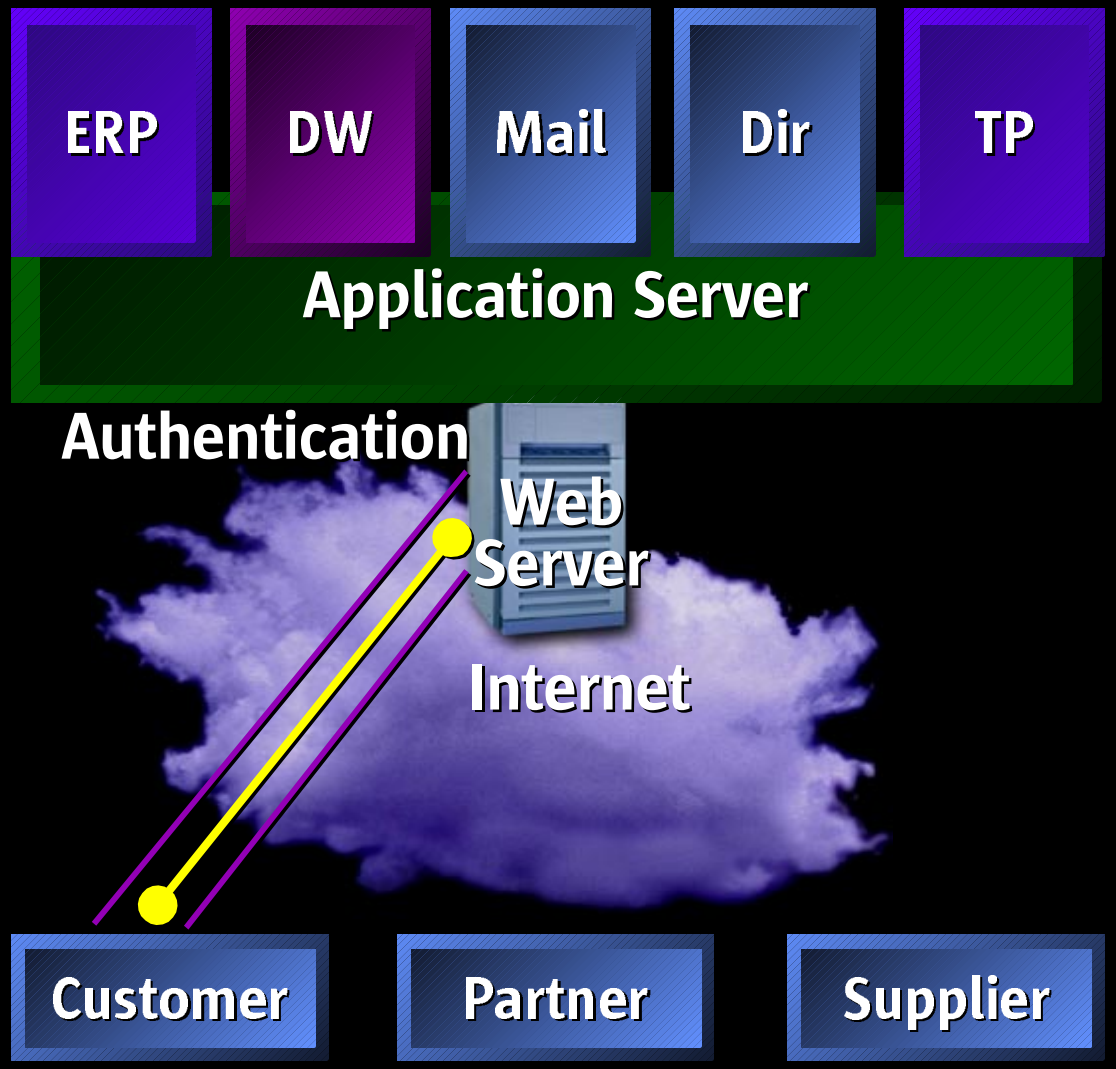
Transitional



Transitional



Application Servers



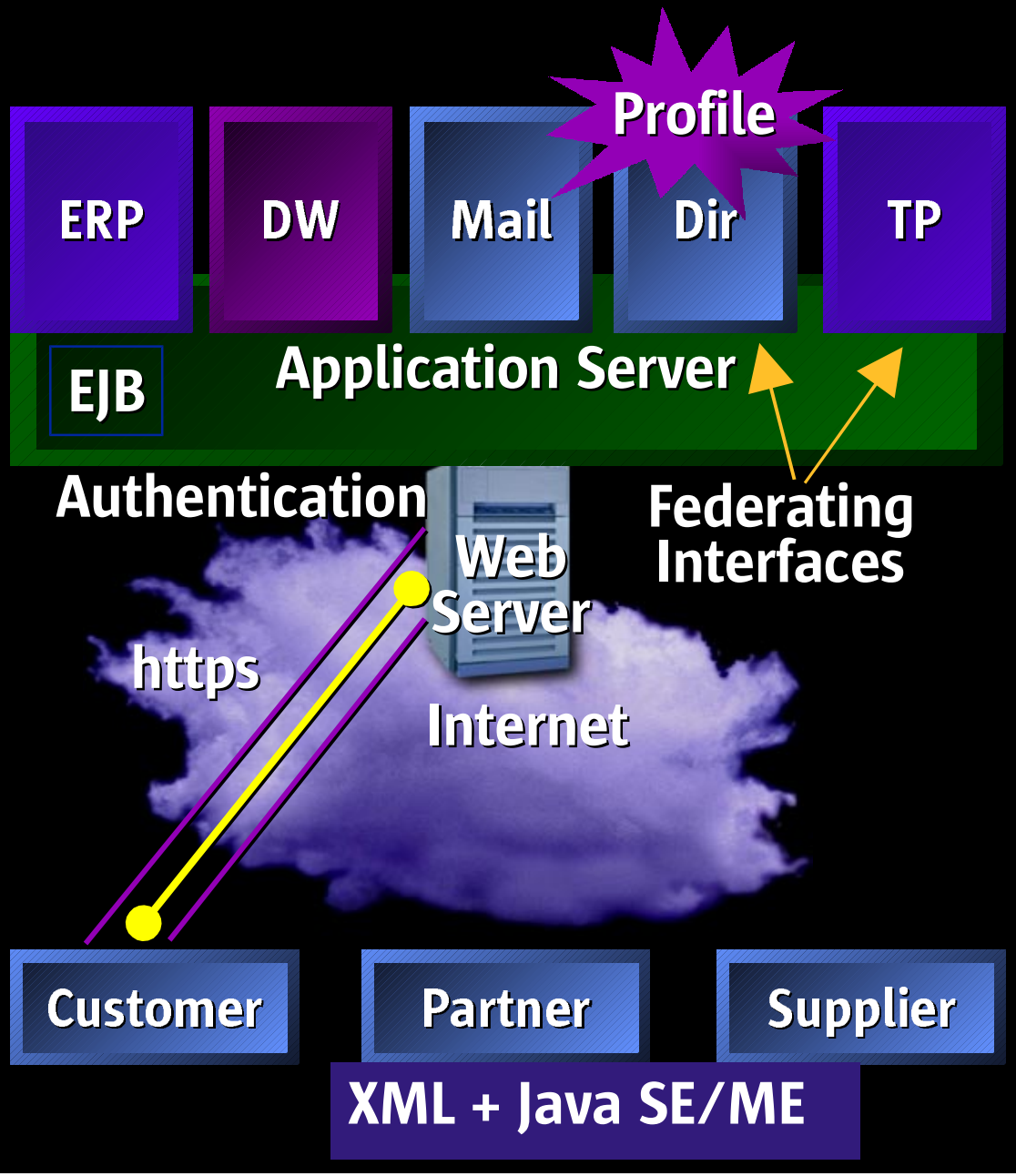
Heterogeneous Servers

Distributed Apps

WEB Frontend

Uncontrolled Clients

Java 2 Enterprise Edition



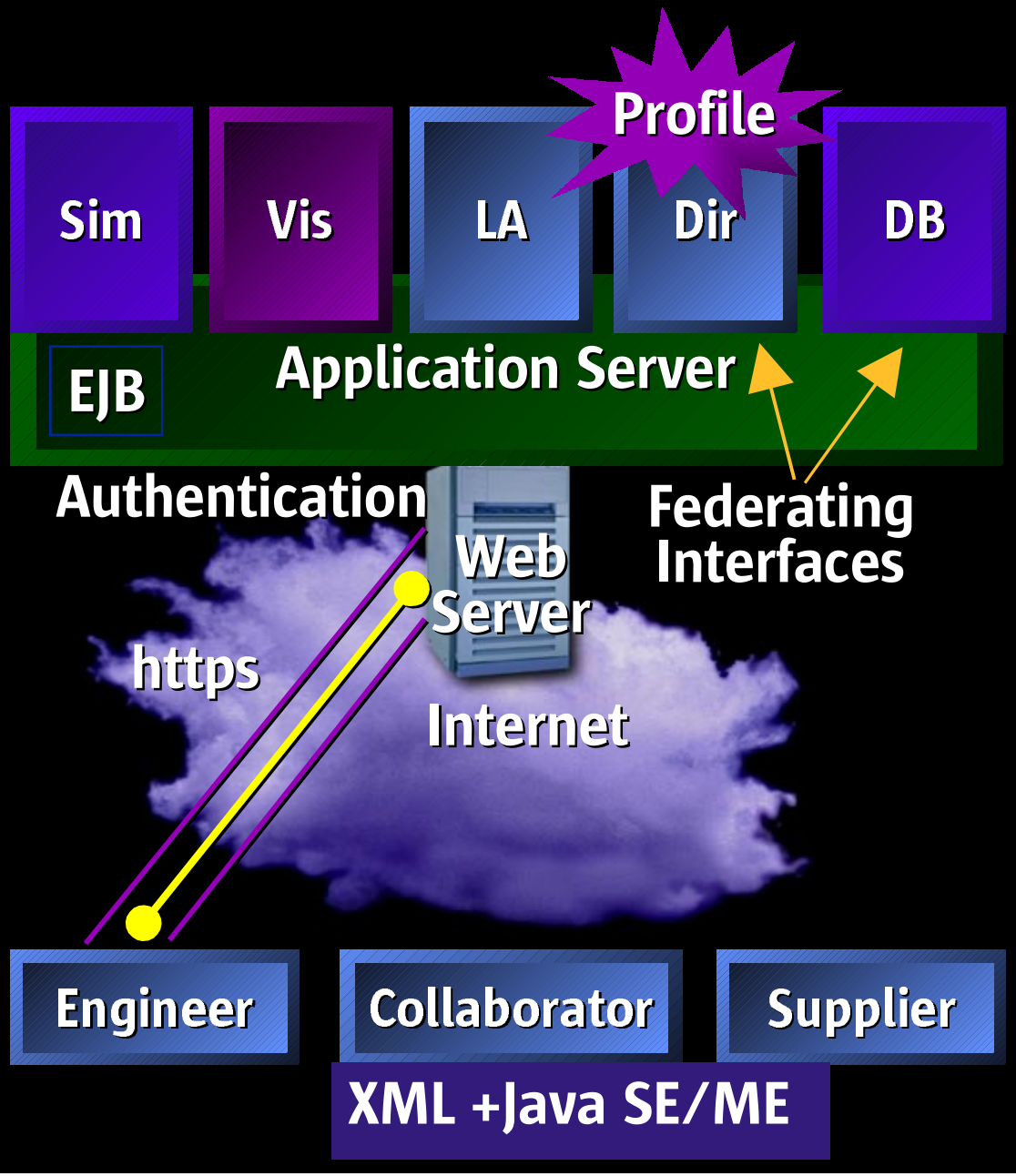
Heterogeneous Servers

Distributed Apps

WEB Frontend

Uncontrolled Clients

HPC Bindings



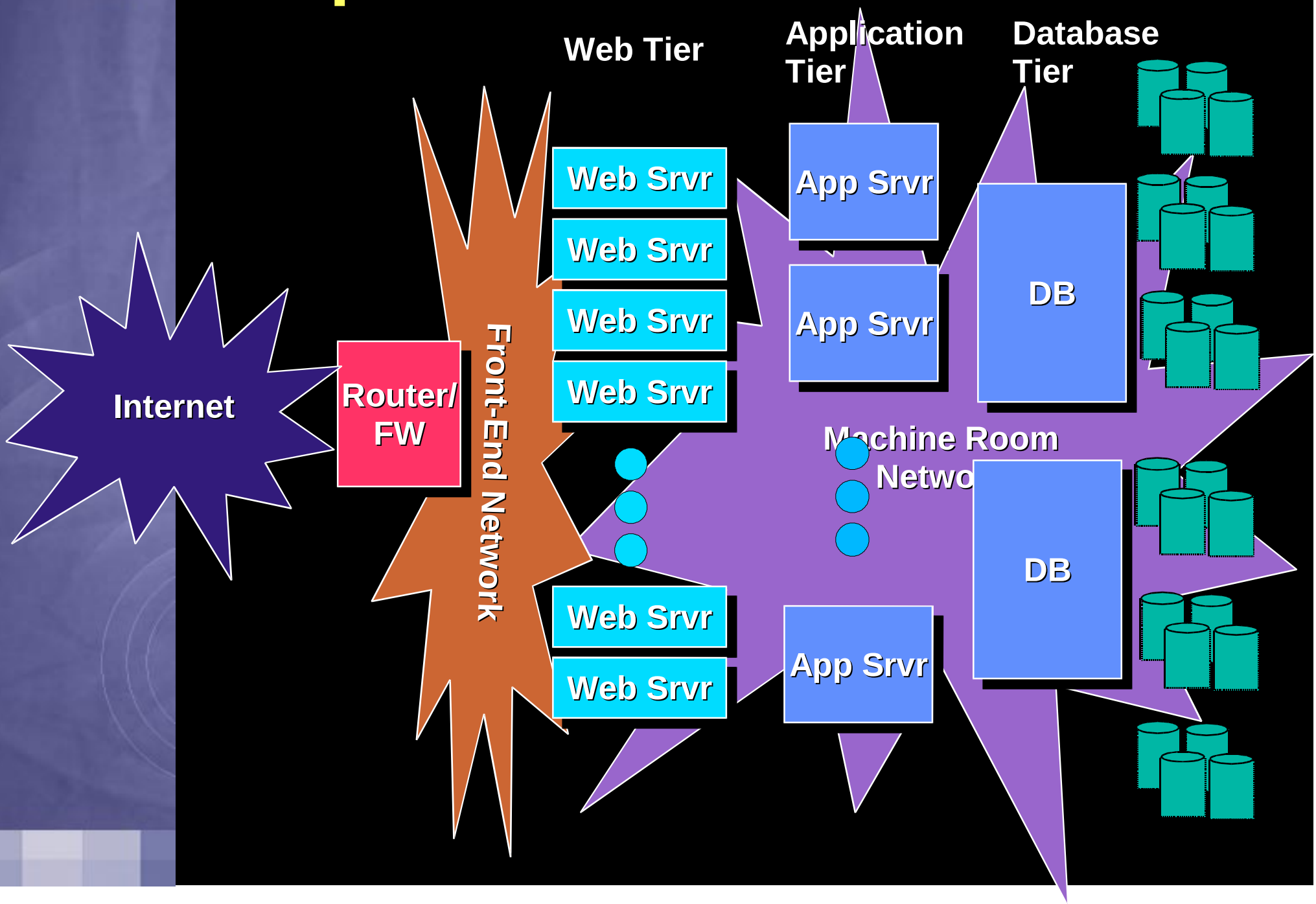
Heterogeneous Servers

Distributed Apps

WEB Frontend

Uncontrolled Clients

Super-Internet Architecture



Qualities Demanded

- Reactive Scalability
- 5 & 6 9's Availability
- Wirespeed Performance
- Extreme Trust
- Rational Scaling of Management

Here is an interesting system...

- >40,000 Processors
- >4 Terabytes of RAM
- >1 million simultaneous users
- 7x24 global operation
- > $1B/yr operating expense

What is it?

Answer

AOL

Here is an (even more?) interesting system…

- >500,000 Processors
  - 1000 Processor-Years every two days
- >10 Terabytes of RAM
- Highly heterogeneous
  - IA32, Sparc, PowerPC,…
  - Windows, Solaris, Linux, MacOS,…
- Wide area distributed

What is it?

Sun microsystems

SETI@home

The Search for Extraterrestrial Intelligence at HOME

http://setiathome.ssl.berkeley.edu

Data Analysis

Chirping data 100%

Doppler drift rate: -0.4385 Hz/sec

Frequency resolution: 0.074506 Hz

Strongest Peak: power 168.15

(2937.9 Hz at 0.00 seconds, drift rate 0.117 Hz/sec)

Strongest Gaussian: power 0.71, fit 3.003

(6696.0 Hz at 26.00 seconds, drift rate 8.527 Hz/sec)

Overall: 68.9988% done

CPU time: 1 hr 59 min 46.3 sec

Data Info

From: 22 hr 52 min 44 sec RA, + 22 deg 16 min 12 sec Dec

Recorded on: Fri Mar 8 16:56:20 1929 GMT

Source: Arecibo Radio Observatory

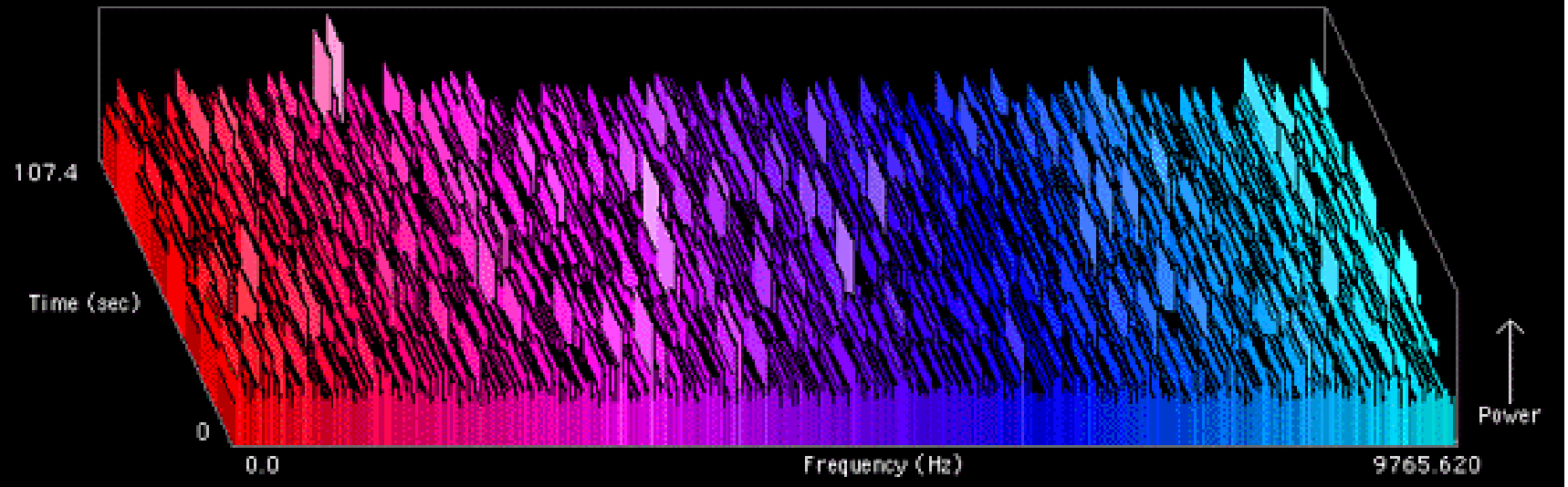
Base Frequency: 1.419648436 GHz

User Info

Name: Ron Hipschman

Data units completed: 46

Total computer time: 359 hr 09 min 47.6 sec



Okay. No more games

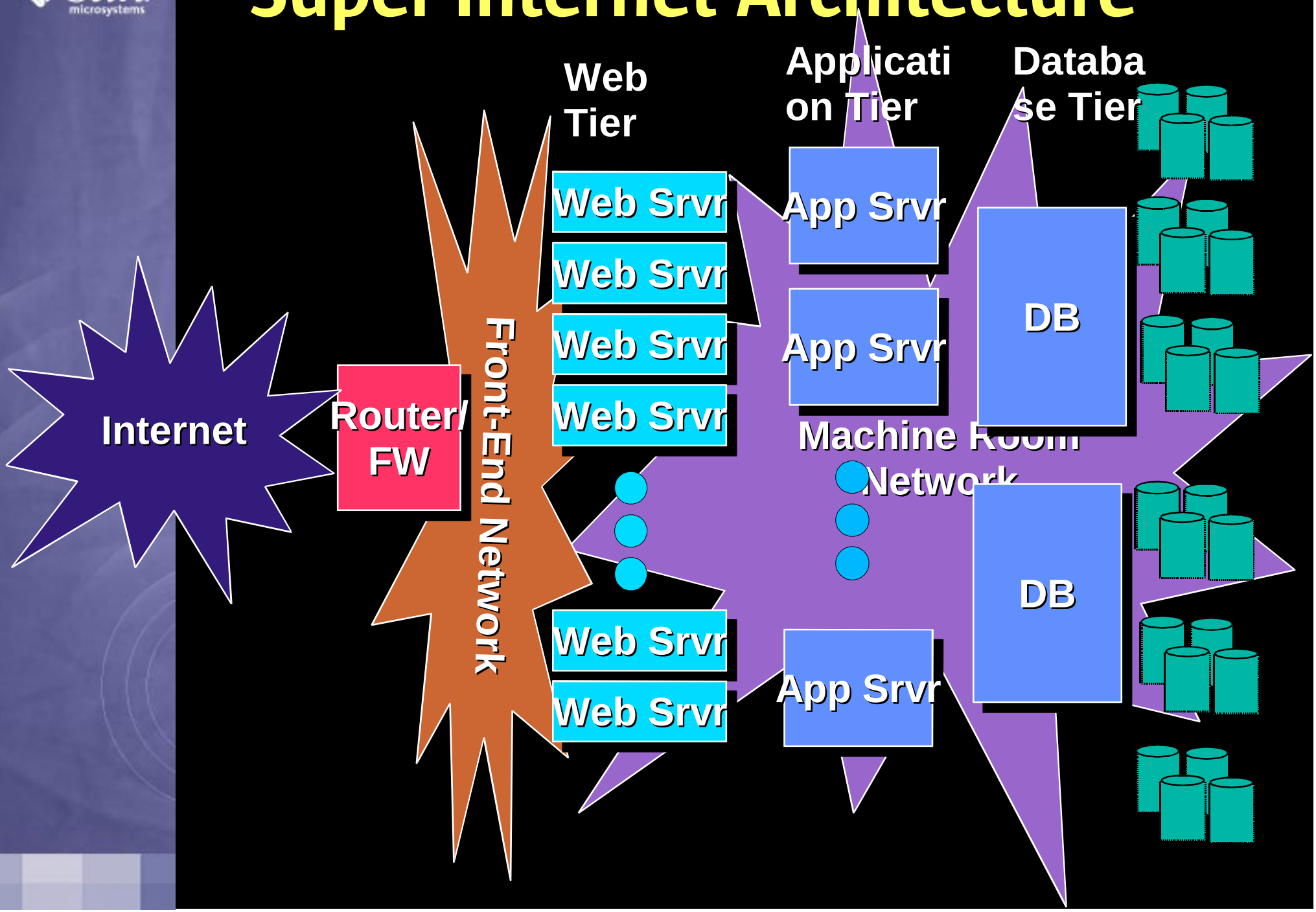
- 10,000 Processors
- 40 Terabytes of RAM
- 500 power users (designers)
- Huge design data files and simulation output
- Deploy 2001-ish

What is it?

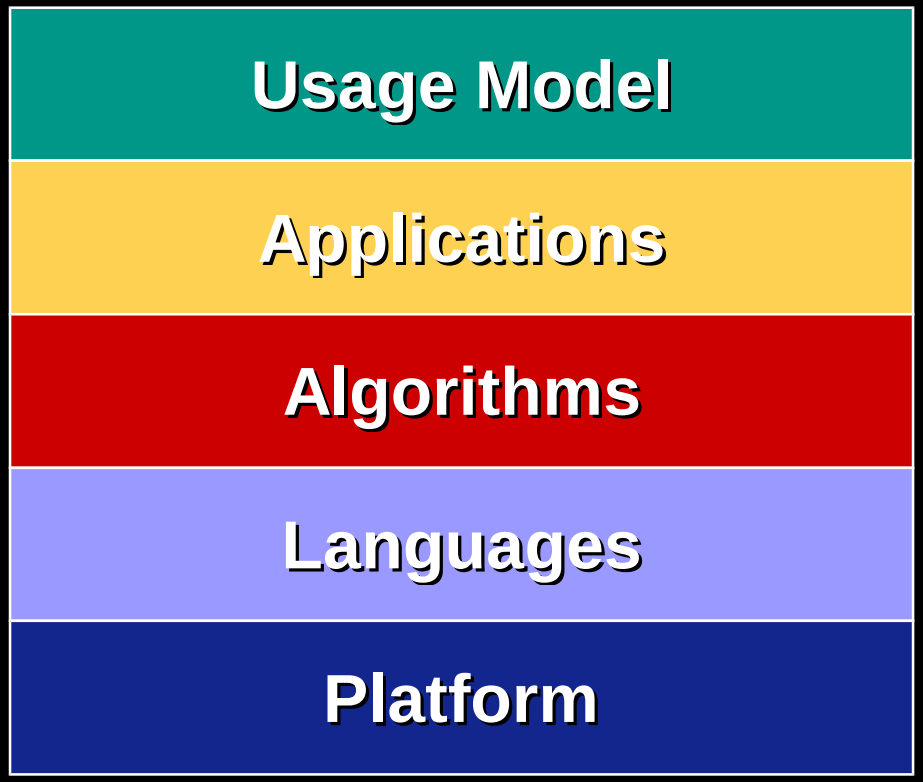
Answer

Sun's Chip Design Server "Ranch"

Super-Internet Architecture



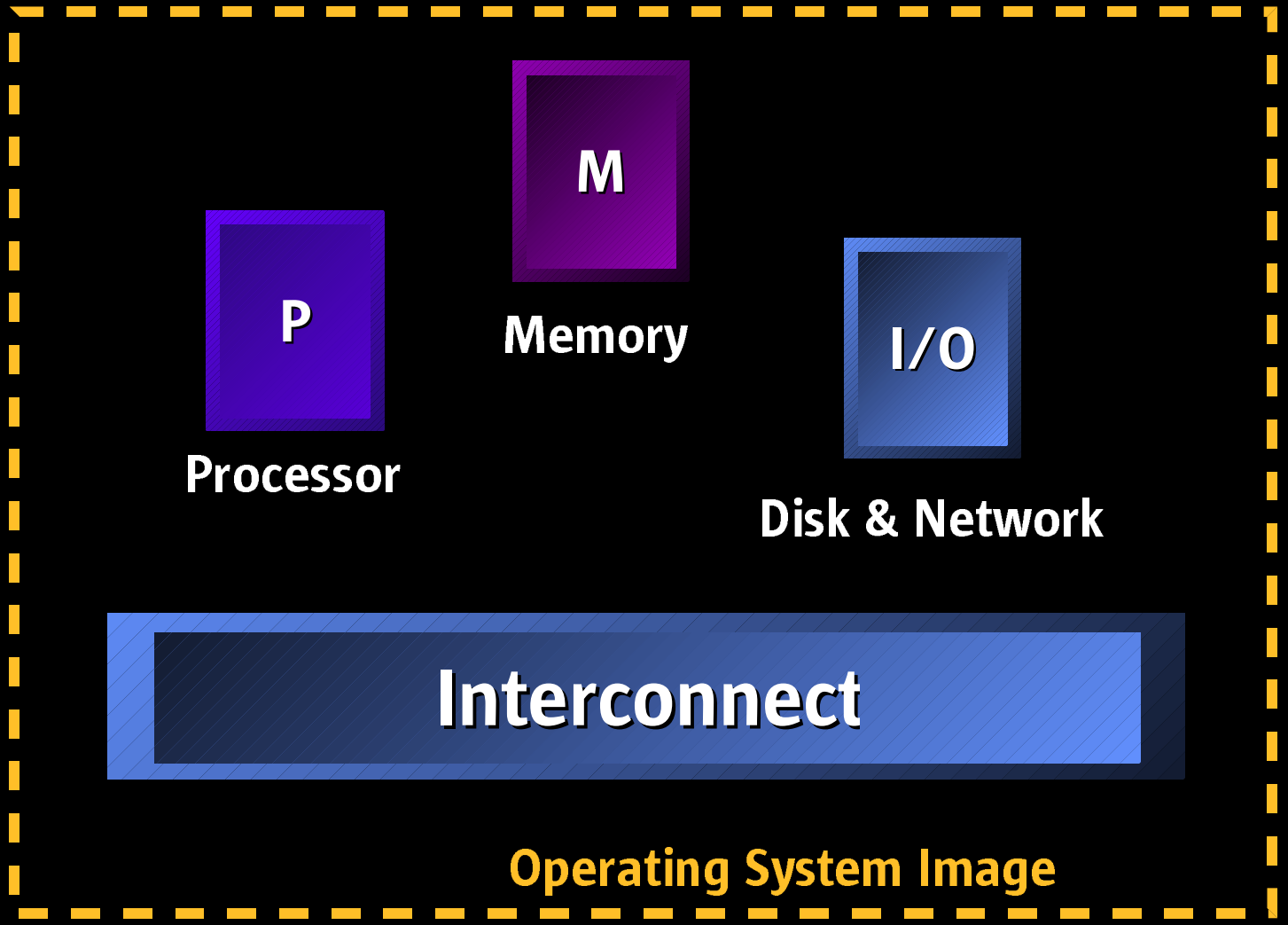
HPC Stack



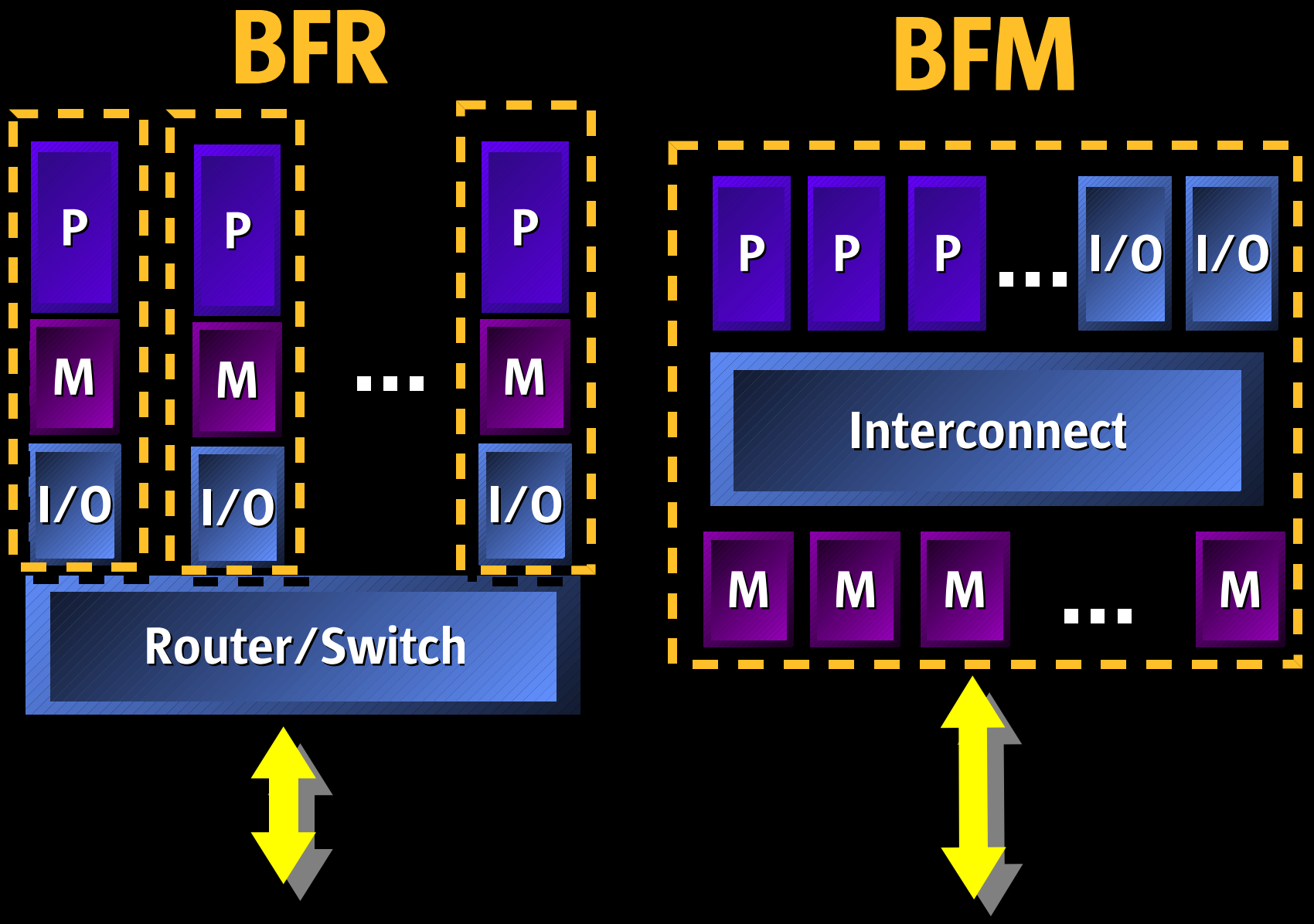
Platform

- Machine Organization & O/S
- IP I/O
- Local Interconnection Network
- Storage

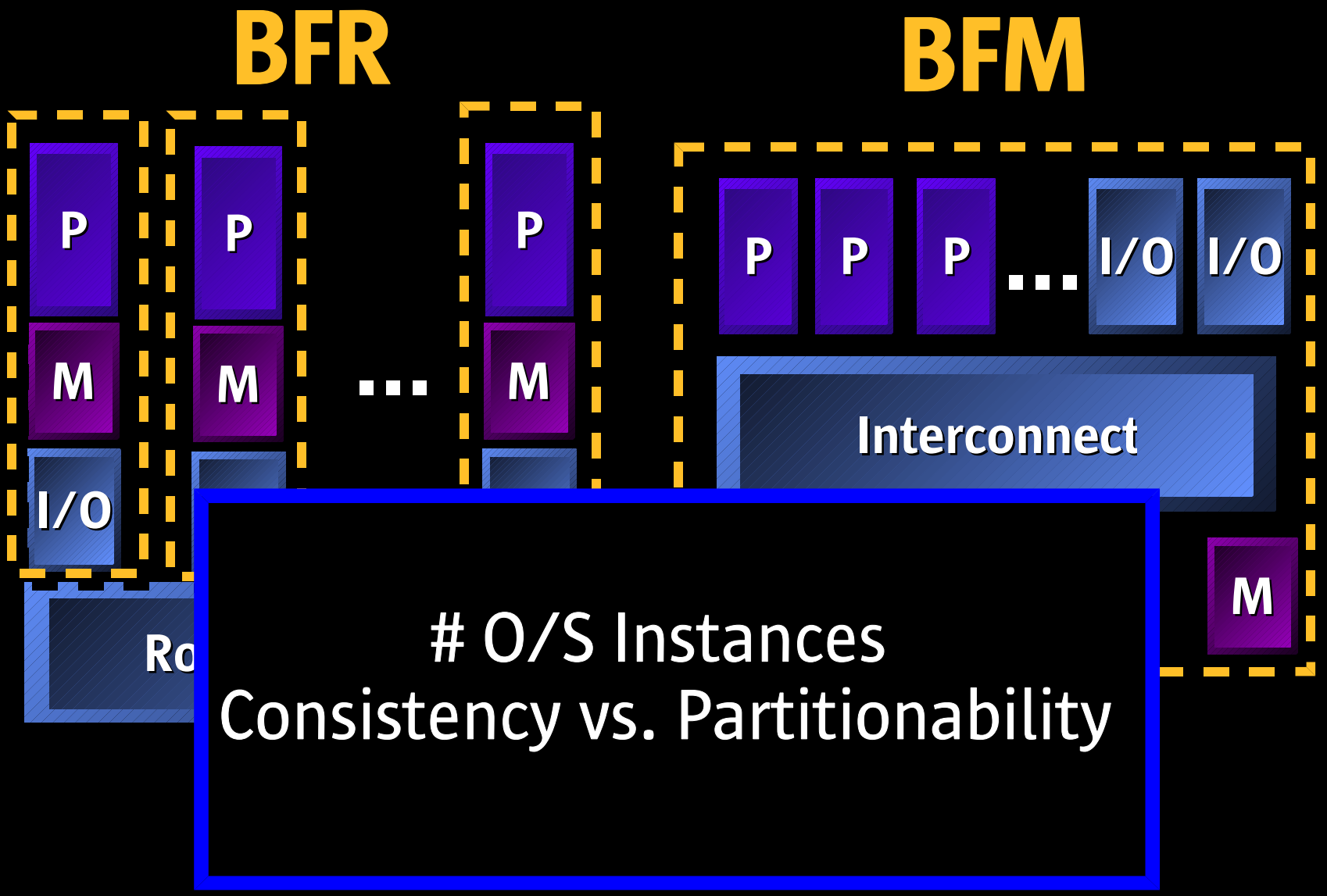
The Five Architecture Building Blocks



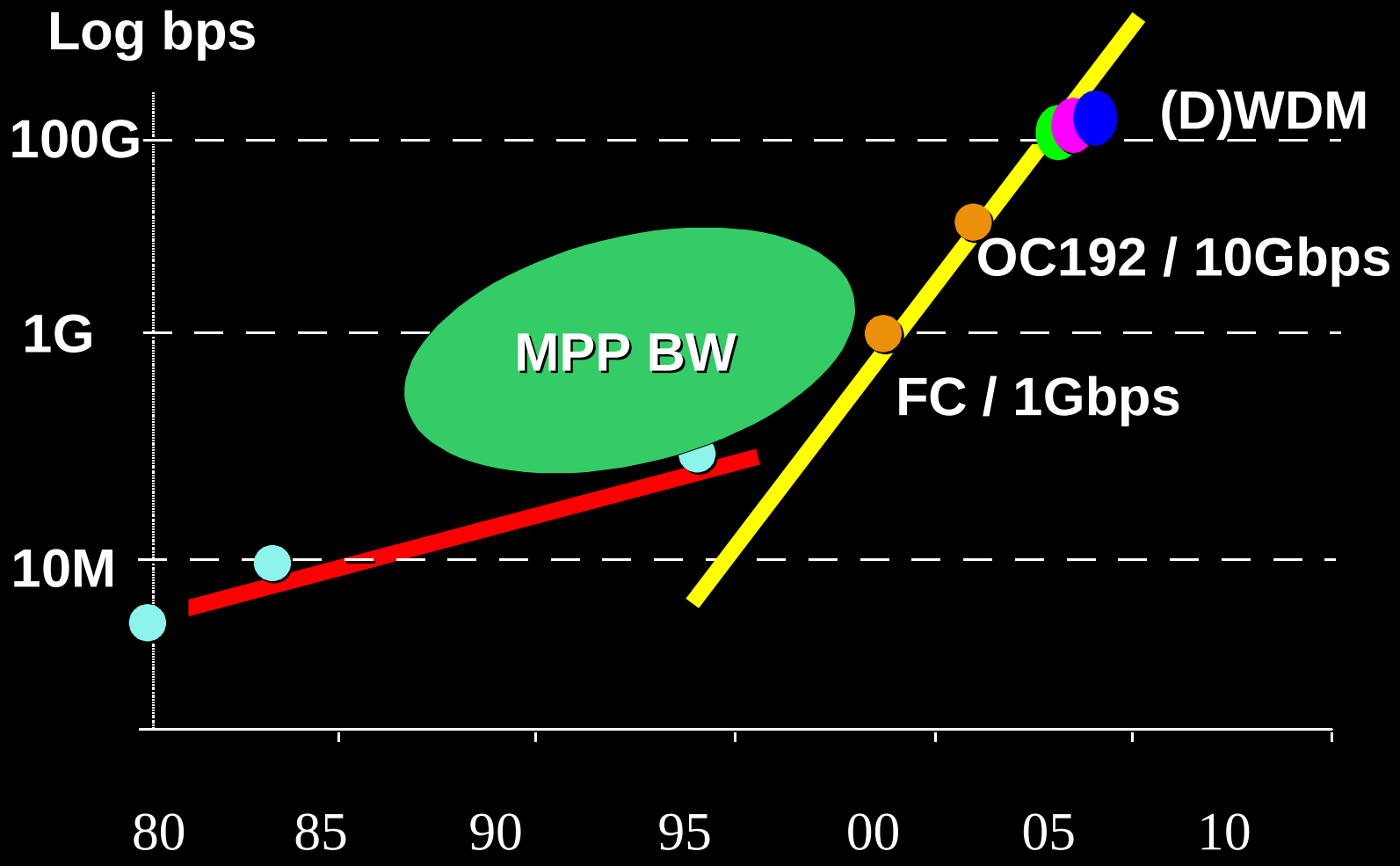
Two Design Camps



Select When...


BFR
BFM
P
M
I/O
Interconnect
O/S Instances
Consistency vs. Partitionability

Interconnect BW Growth



The 10Gbps Watershed

- 10G Ethernet = OC192 rate
- Time-of-flight = Serialization delay
  e.g., @ distance 30m = 150ns
  256B/10Gbps = 200ns
- W x 10Gbps WDM @ 2x/Year
- All-Optical circuit switching networks
- Overhead gets pounded down

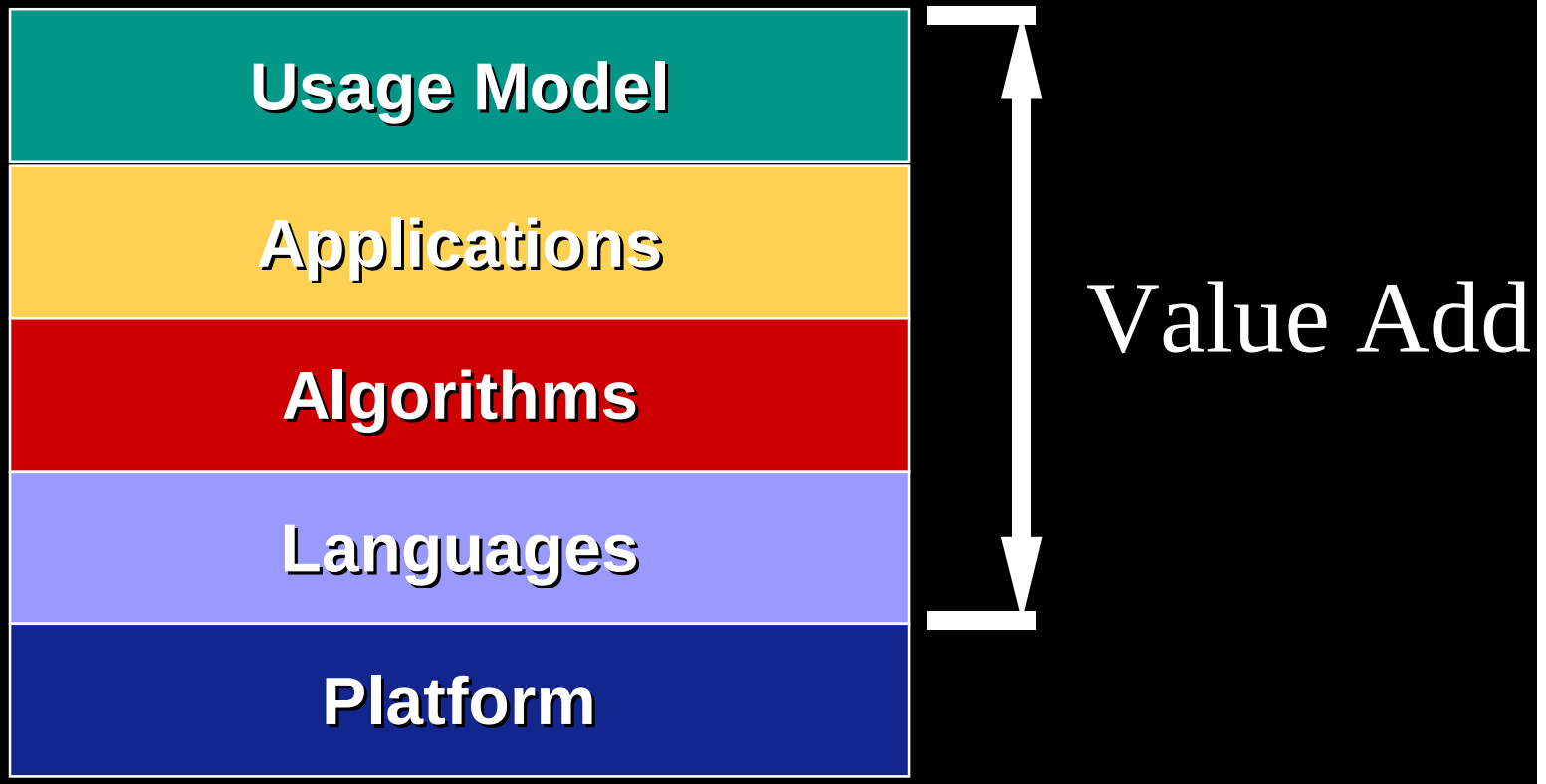
I/O and Storage

- SIO (Infiniband) becomes standard
  I/O pipe = 6GBytes/sec
- Storage attaches to SAN
  - FC today
  - IP or Infiniband tomorrow
- HUGE industry

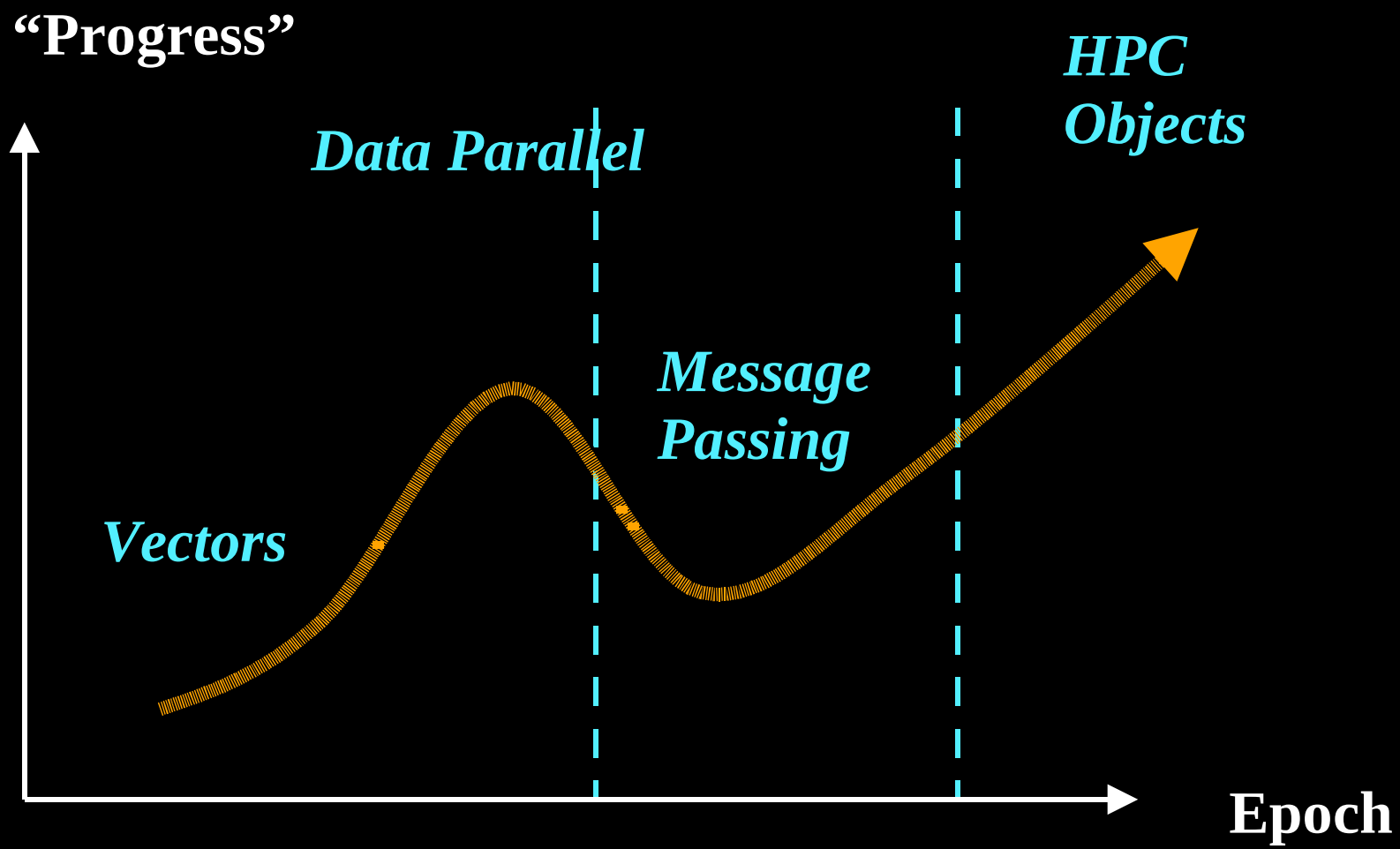
Platform negatives

- Random access memory latency & BW
- Collective/Structured communications
- Some numerical efficiency

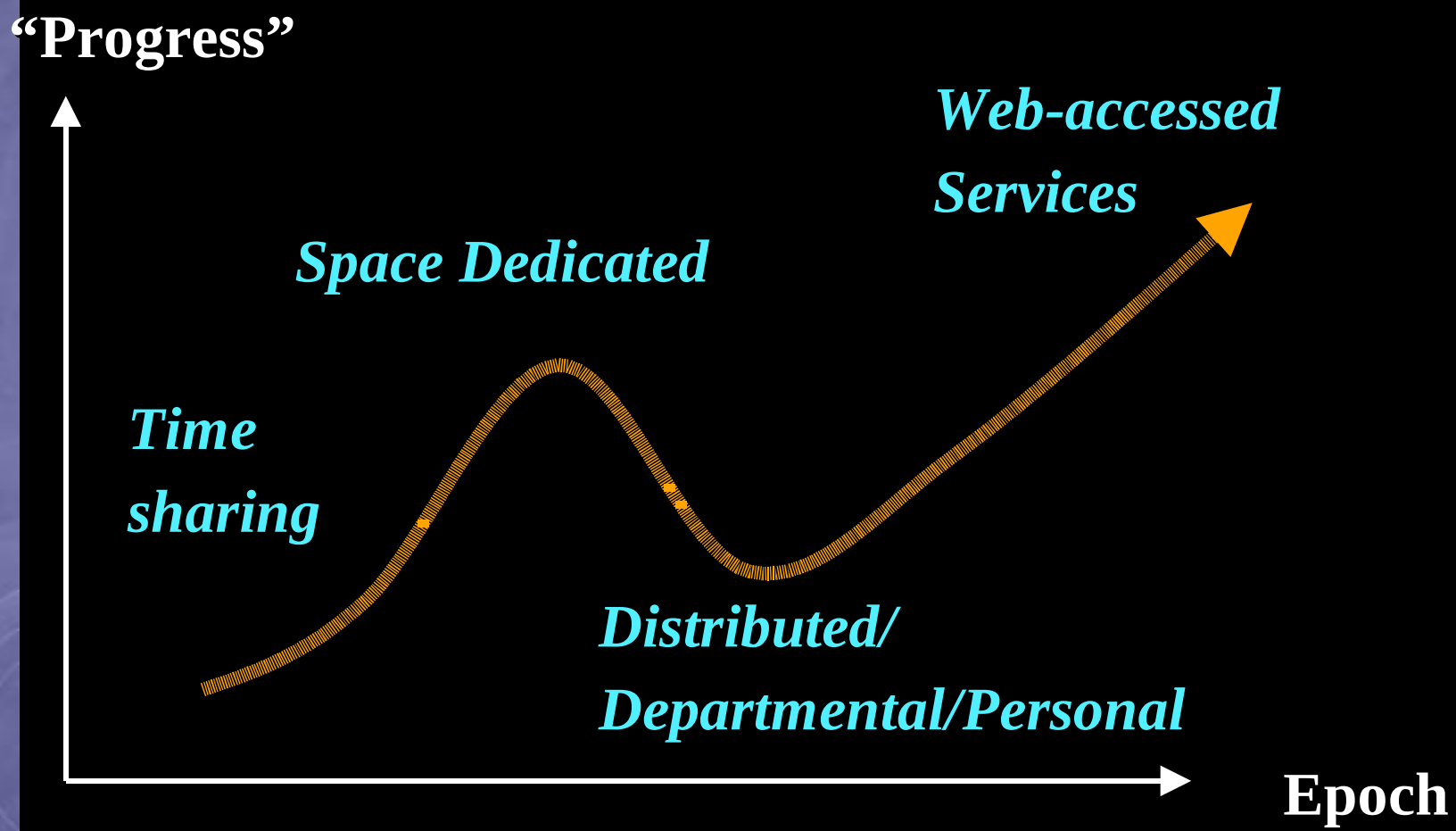
HPC Stack


Usage Model
Applications
Algorithms
Languages
Platform
Value Add

Languages & Parallelism



Usage Model

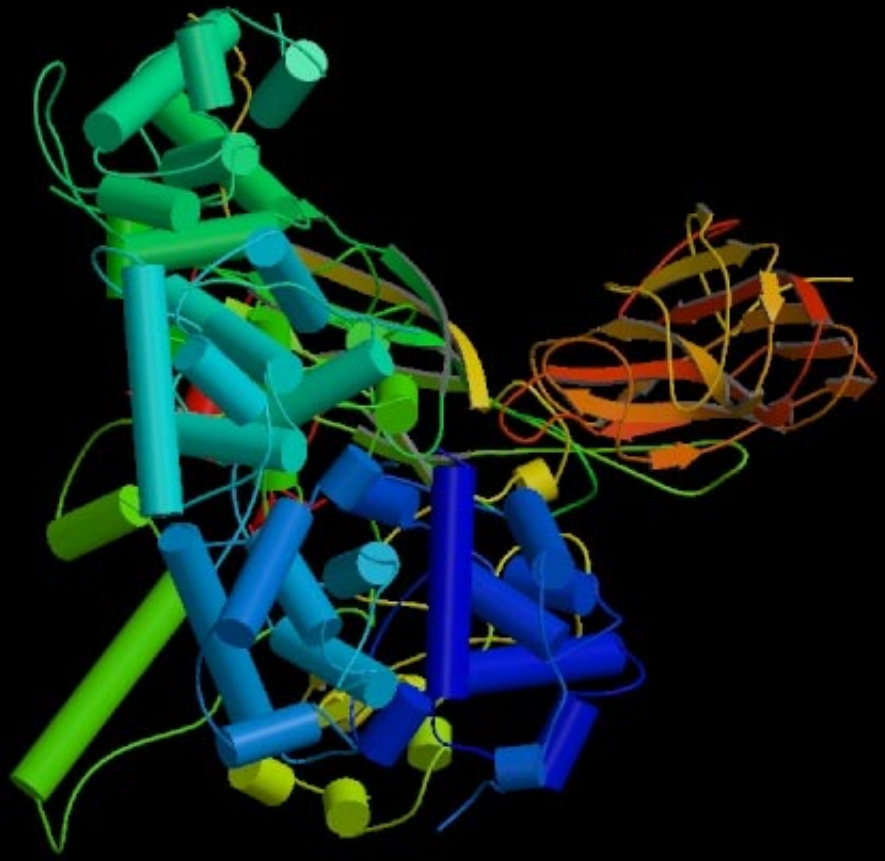

"Progress"
Web-accessed Services
Space Dedicated
Time sharing
Distributed/ Departmental/Personal
Epoch

Services

- Storage & Archiving
- Intelligent Search
- Simulation
- Visualization
- ...

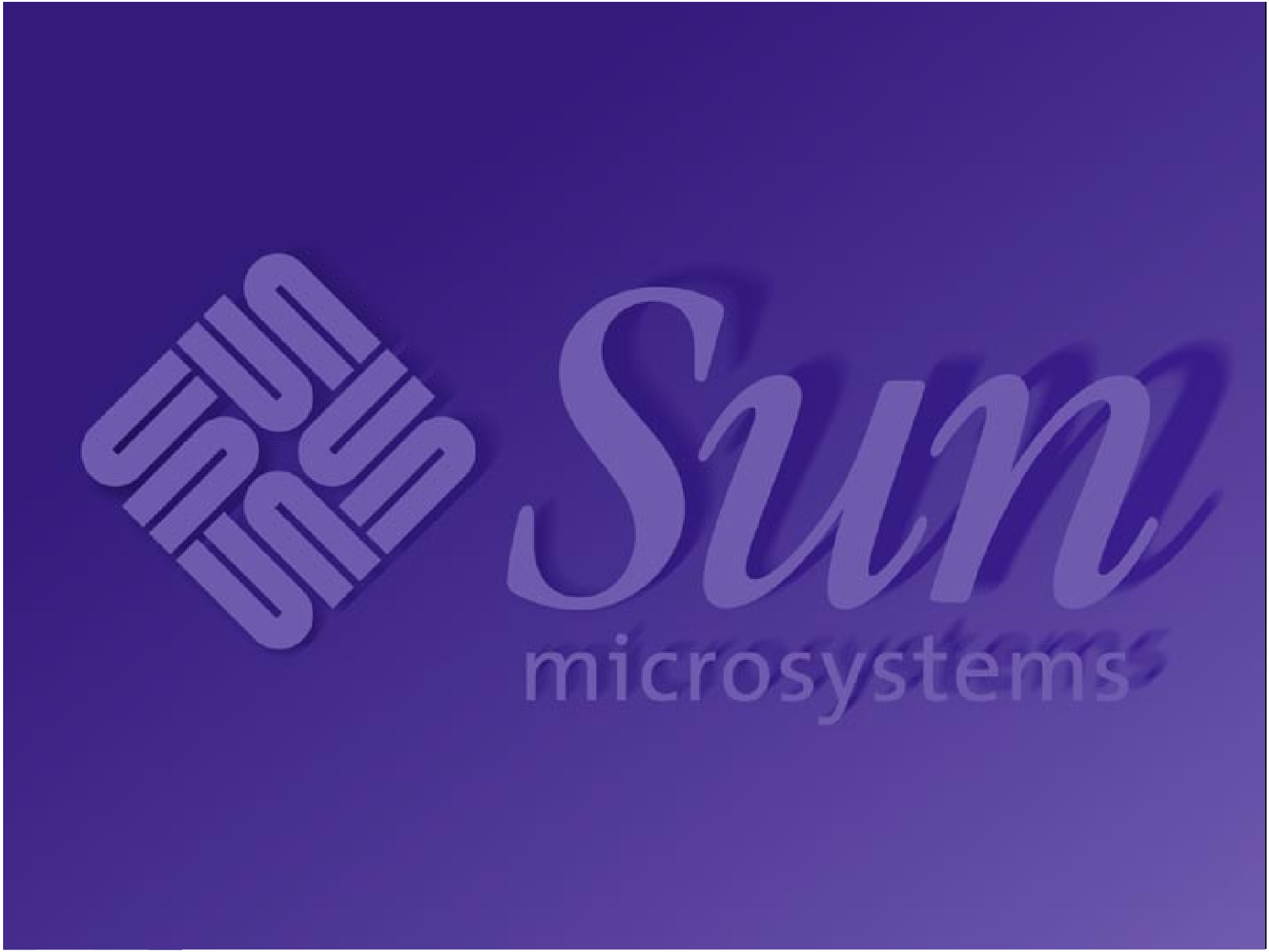
Example: Protein Data Base www.rcsdb.org

PDB™ PROTEIN DATA BANK



Take Aways

- The superinternet platform is the supercomputer.
- Supercomputing moves up the food chain
- Everything becomes a network service


Sun
microsystems