HPC USER FORUM

Stuttgart Germany

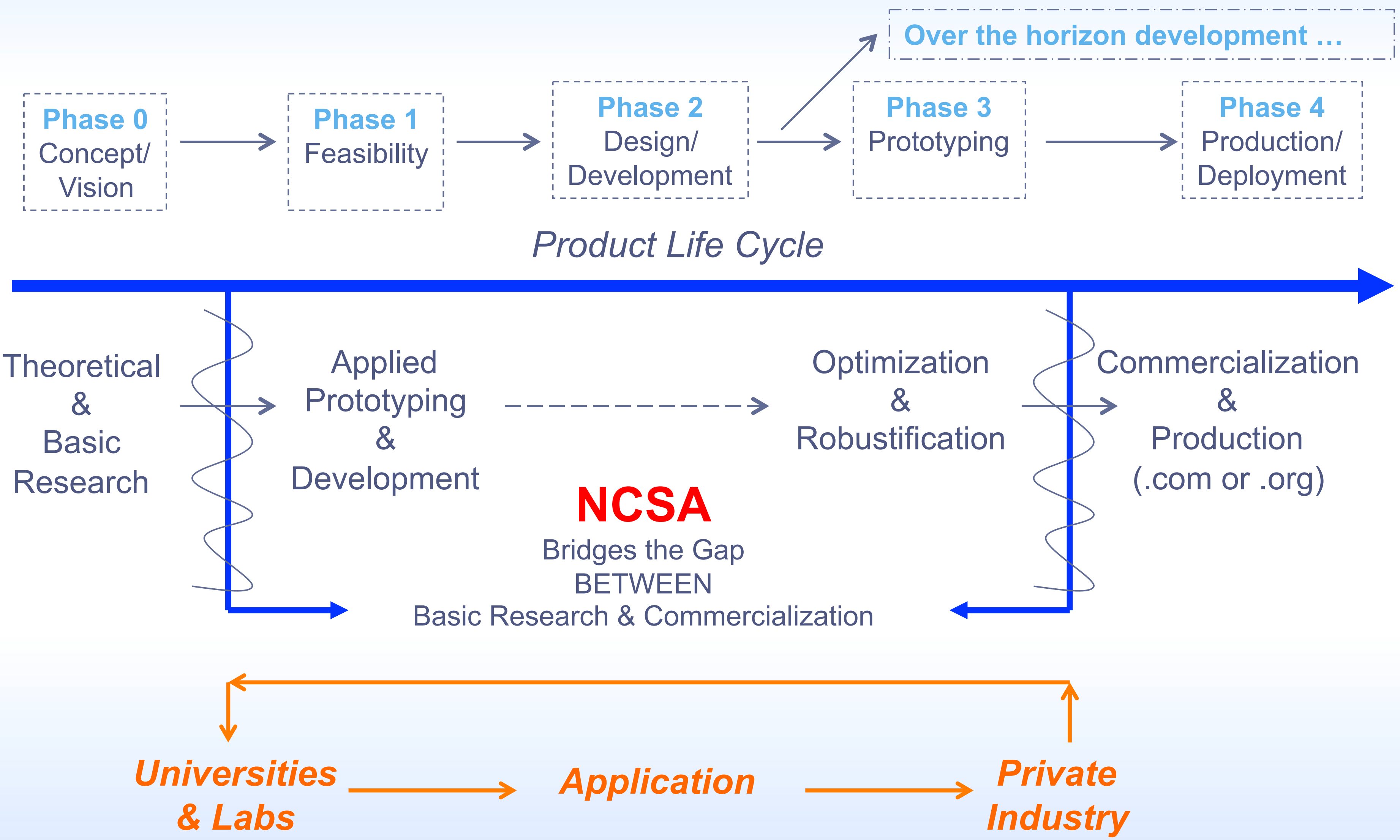
October 2010

Merle Giles
Private Sector Program & Economic Development
mgiles@ncsa.illinois.edu

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

Basic & Applied Research
CSL
Digital Computer Lab
College of Engineering
NCSA
ECE
Illinois Informatics Initiative
Invent. Imagine. Innovate.
Institute for Advanced Computing
Applications and Technologies
CSE
Information Trust
INSTITUTE
Computer Science
Beckman Institute
Institute for Genomic Biology
Biotechnology Center
NCSA

NCSA Bridges Basic Research and Commercialization with Application



Value Creation and Economic Development

- 3D virtual prototyping at NCSA => Caterpillar's Global Simulation Center in Champaign
- Full-scale simulation of cell tower activity
- Reduced reliance on wet labs thru computation
- Cluster design and architecture
- HPC-capable fast-network hard drives
- World's first Internet GUI interface
- Prototyping Windows® HPC Operating System

Value Creation and Economic Development

- The HDF Group (pdf-equivalent for data)
- Linux OS made standard for HPC
- Apache server software
- Telnet remote access
- R-systems hosts Wolfram Alpha
- Music data analytics at One Llama
- River Glass spinout – recent Boeing buyout

TOTAL $$

- $1 Trillion per founder Larry Smarr

Industry Partners over time

CURRENT PARTNERS

BOEING

P&G

CAT

MOTOROLA

Microsoft

JOHN DEERE

ROLLS ROYCE

L&L Products

ADM

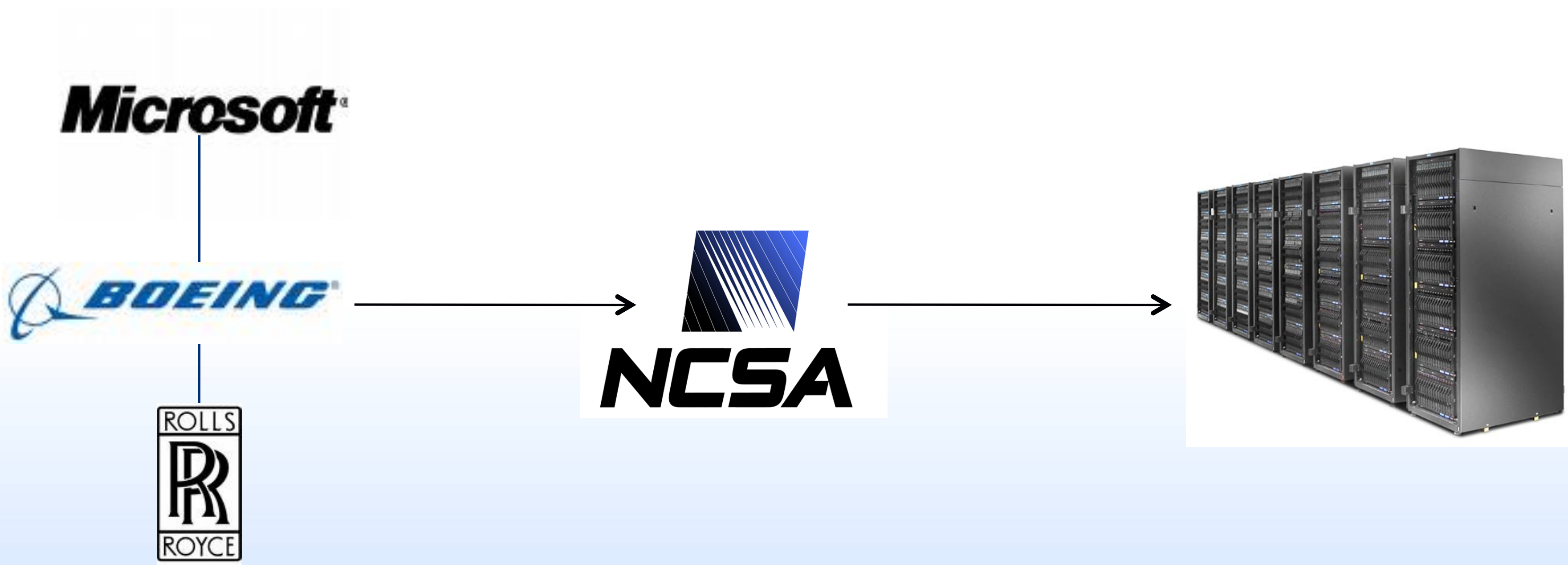
WATERBORNE ENVIRONMENTAL, INC.

IBM

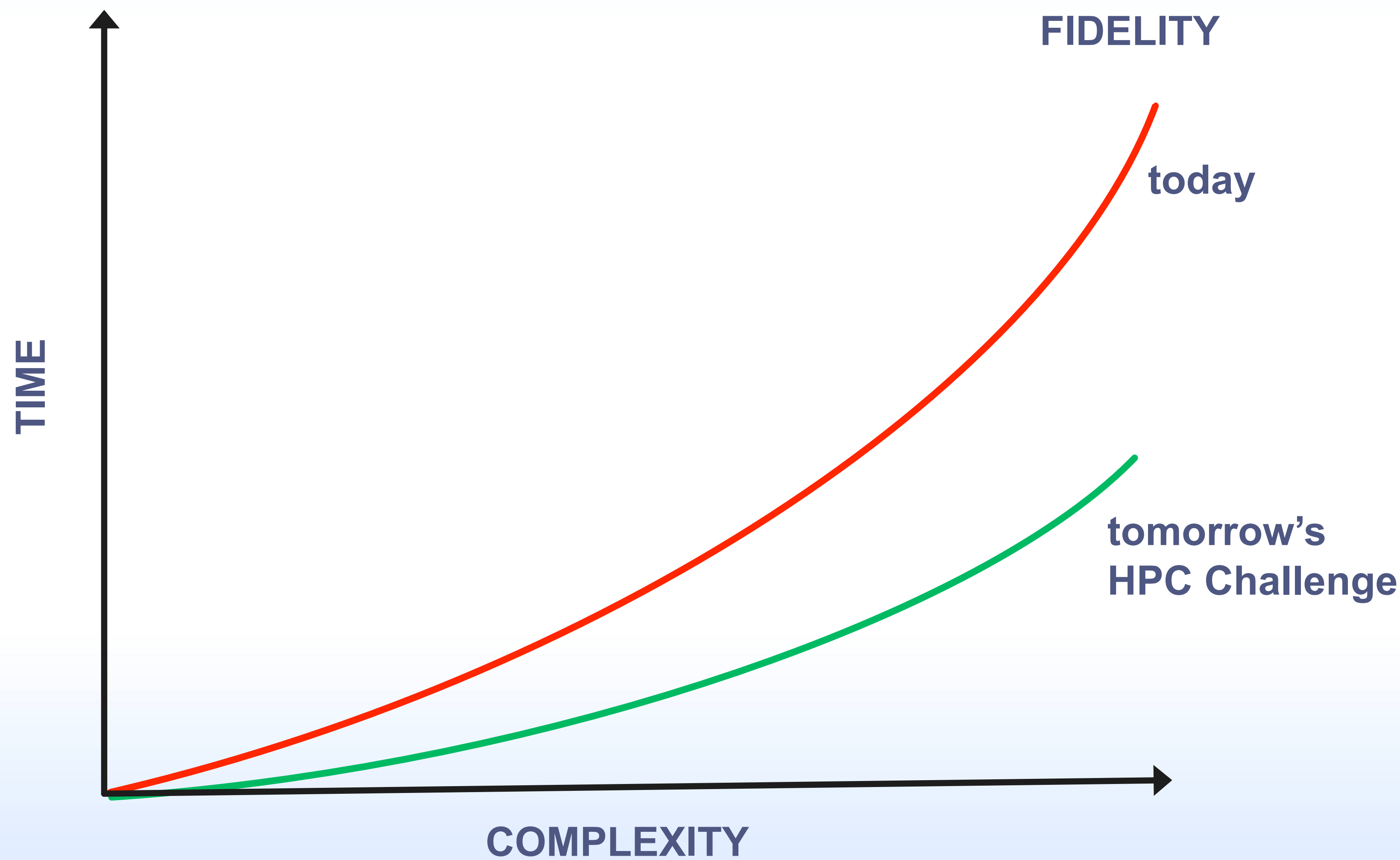
Illinois Rocstar

CLASSICAL HPC MISSION

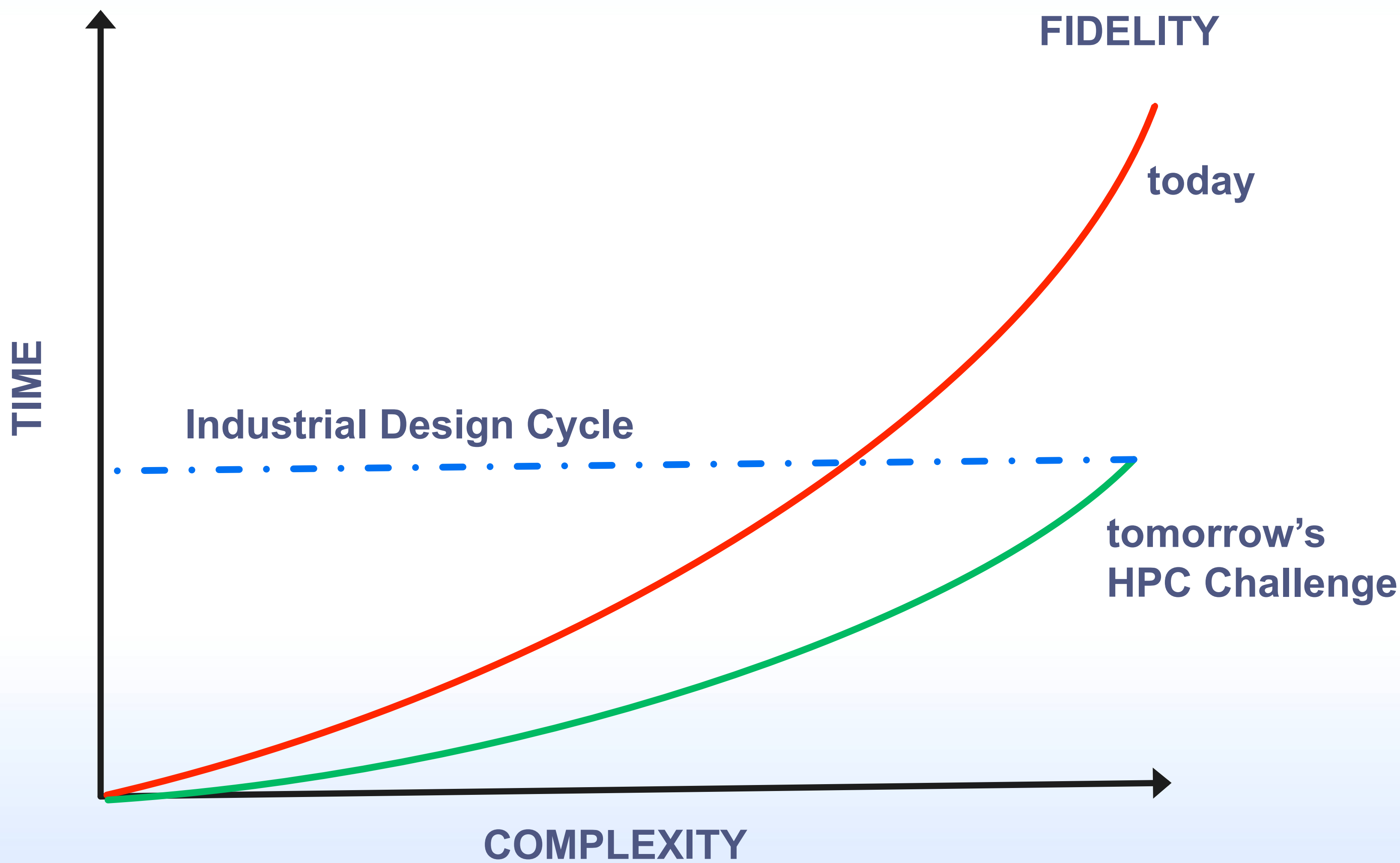
- Historically, HPC has focused on science *discovery*
- Economic value has also been achieved in HPC derivatives
- Industrial value keys on discovery and *optimization*
- Increasingly, industry brings world-class problems



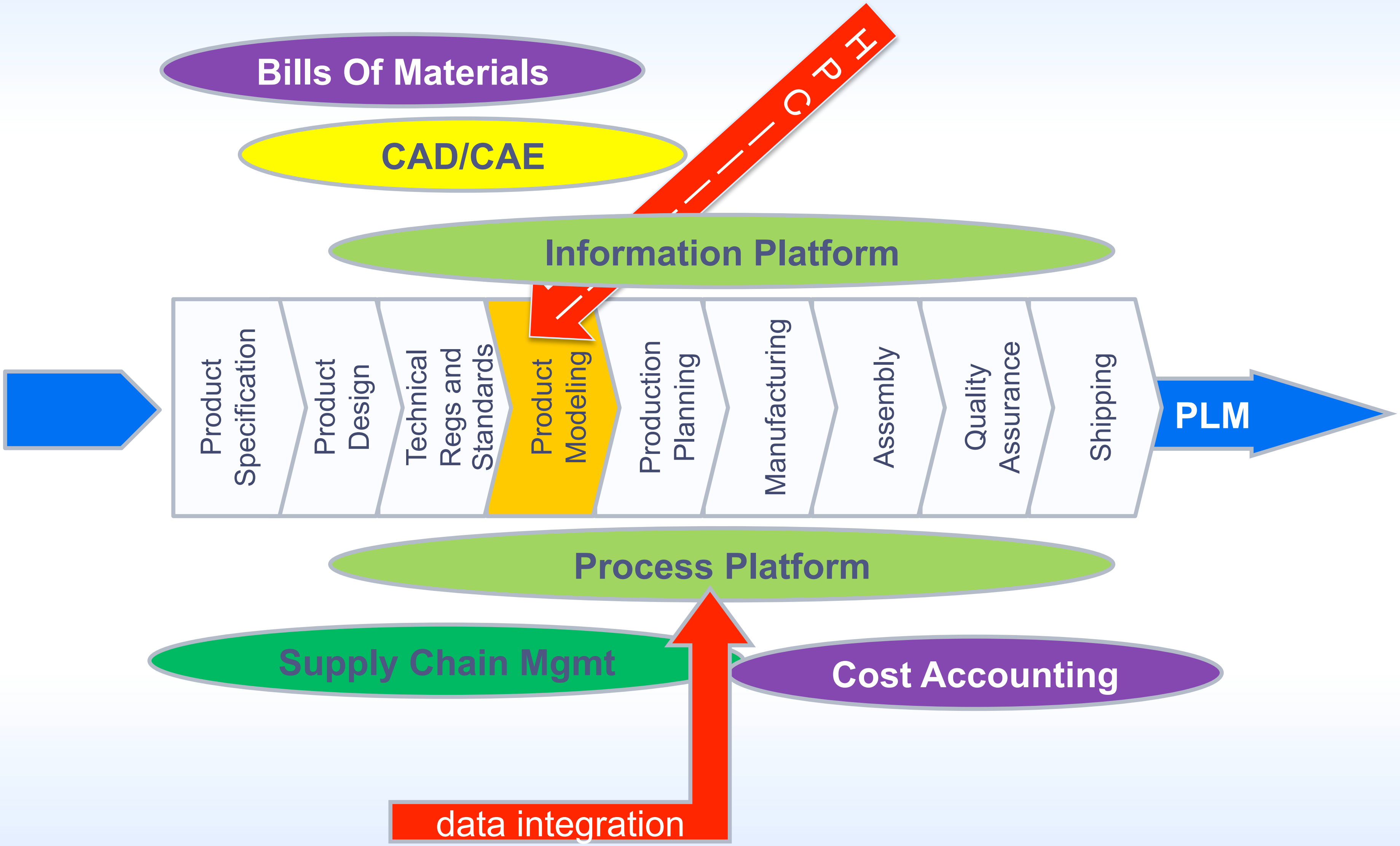
Tomorrow’s Industrial Challenge



Discovery is No Longer Sufficient



Product Lifecycle Management Proves Why



Forbes Magazine: Publisher Rich Karlgaard
Digital Rules, September 13, 2010

Smart-aggregation rules: Some forms of content will always need human curating.

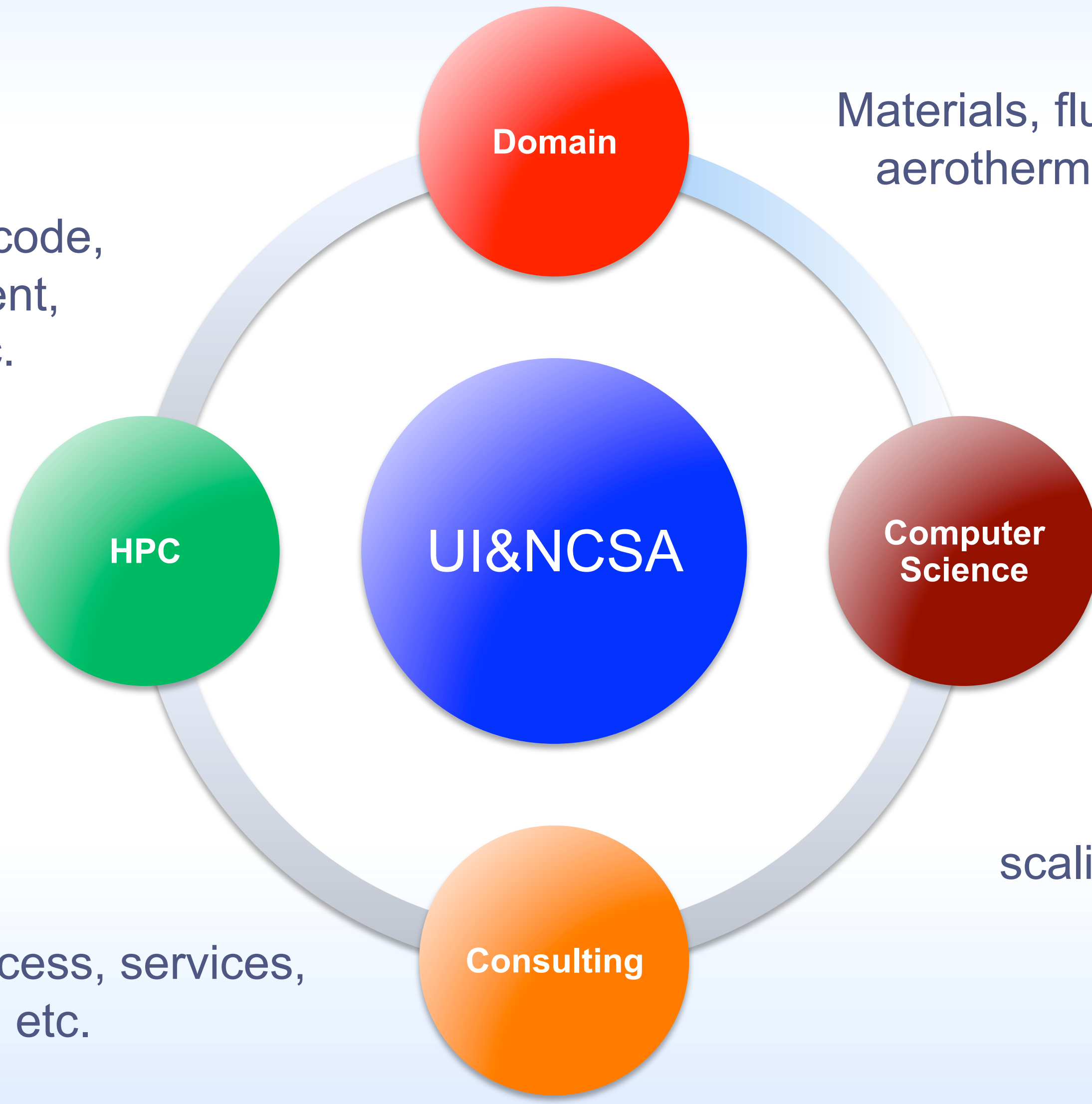
Dumb-aggregation rules: And some forms of content won't.

The trick is to figure out where algorithms beat humans, and vice versa.

Competitive Advantage needs Human Expertise

Materials, fluids, mechanical, aerothermal, electrical, etc.

Petascale, ISV code, data management, architecture, etc.



Parallelization, scaling, multiphysics, algorithms, etc.

Strategy, process, services, partnerships, etc.

Headlines

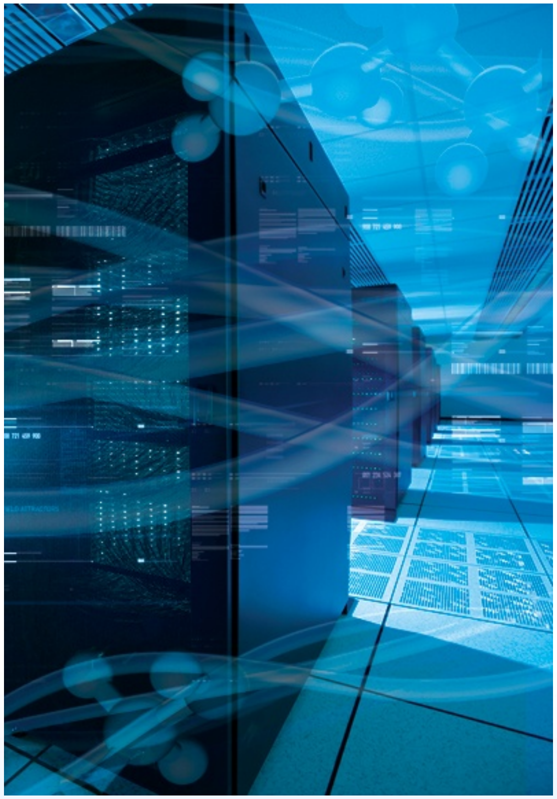
- **GERMANY Trade & Invest** – Partnership is the key to country's thriving R&D landscape.
- **IBM Smarter Planet** – Thanks to pervasive instrumentation and global interconnection, we are now capturing data in unprecedented volume and variety.
- **IBM Smarter Planet** – World's network traffic will soon total more than half a zettabyte.
- **WSJ Steve Conway** – "There is growing recognition of the close link between supercomputing and scientific advancement as well as industrial competitiveness."
- **Forbes** – Consumer technology is now ahead of most industrial technology.

NCSA

EXTREME COMPUTING

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

New Performance Driver



NCSA’s Blue Waters is the first open-access system tasked to achieve ≥ 1 petaflop/s on *real* applications.

Guess What This Is ?

From 1956 . . .

Guess What This Is ?

From 1956 . . .

A hard disk drive

with 5 MB storage

Leading-Edge Collaboration

BLUE WATERS
BREAKING THROUGH THE LIMITS

Mega- 10^6
Giga- 10^9
Tera- 10^{12}
Peta- 10^{15}
Exa- 10^{18}
Zetta- 10^{21}
Yotta- 10^{24}

(Bits, Bytes, Flop/s)

IBM

GREAT LAKES CONSORTIUM
FOR PETASCALE COMPUTATION

Blue Waters Expected to Beat 2008's TOP500® COMBINED!

U.S. Leadership Computing Programs

- **U.S. Department of Energy**
  - Oak Ridge National Laboratory: Jaguar + Follow-on
  - Argonne National Laboratory: Intrepid + Follow-on
  - Lawrence Livermore National Lab: Dawn + Sequoia
- **National Science Foundation**
  - University of Illinois/NCSA: Blue Waters
- **NASA**
  - Ames Research Center: Pleiades

Petaflop/s Comparison

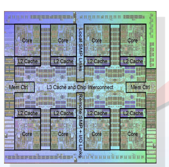
SYSTEM ATTRIBUTE	NCSA Abe	DOE Jaguar	NCSA Blue Waters
Vendor	Dell	Cray	IBM
Processor	Intel Xeon 5300	AMD 2435	IBM Power7
Peak Performance (Pf/s)	0.088	2.33	~10.0
Sustained Performance (Pf/s)	~0.005	??	≥1.03
# Cores/Chip	4	6	8
# Cores (total)	9,600	224,256	>300,000
Memory (Terabytes)	14.4	360	>1,200
Online Disk Storage (Terabytes)	100	10,000	>18,000
Archival Storage (Petabytes)	5	20	*up to 500*
Sustained Disk Transfer (TB/s)	na	.240	> 1.5

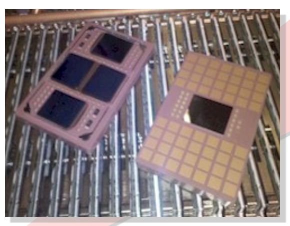
Machine Comparison

SYSTEM ATTRIBUTE	ASC Purple IBM	DOE Jaguar CRAY	Blue Waters IBM *****1 Rack*****	Blue Waters IBM ***Complete***
Year Deployed	2005	2009	2011	2011
Processor	IBM P5	AMD	IBM P7	IBM P7
# Cores	12,000	224,256	3,072	300,000+
Peak Performance (Tf/s)	100	2,330	100	10,000+
Disk Storage (Terabytes)	2,000	7,000	153.5	18,000+
# Disk drives	10,000	??	384	40,000
Memory (Terabytes)	50	300	24.6	1,200
Cost	$290M	??	??	$208M

Integrated/Scalable System

Blue Waters will be the most powerful computer in the world for scientific research when it comes on line in 2011.





Power7 Chip

8 cores, 32 threads
L1, L2, L3 cache (32 MB)
Up to 256 GF (peak)
128 Gb/s memory bw

Quad-chip Module

4 Power7 chips
128 GB memory
512 GB/s memory bw
1 TF (peak)

Hub Chip

1.128 TB/s bw

IH Server Node

8 QCM's (256 cores)
  8 TF (*peak*)
1 TB memory
  4 TB/s memory bw
8 Hub chips
Power supplies
PCIe slots

Fully water cooled

Blue Waters Building Block

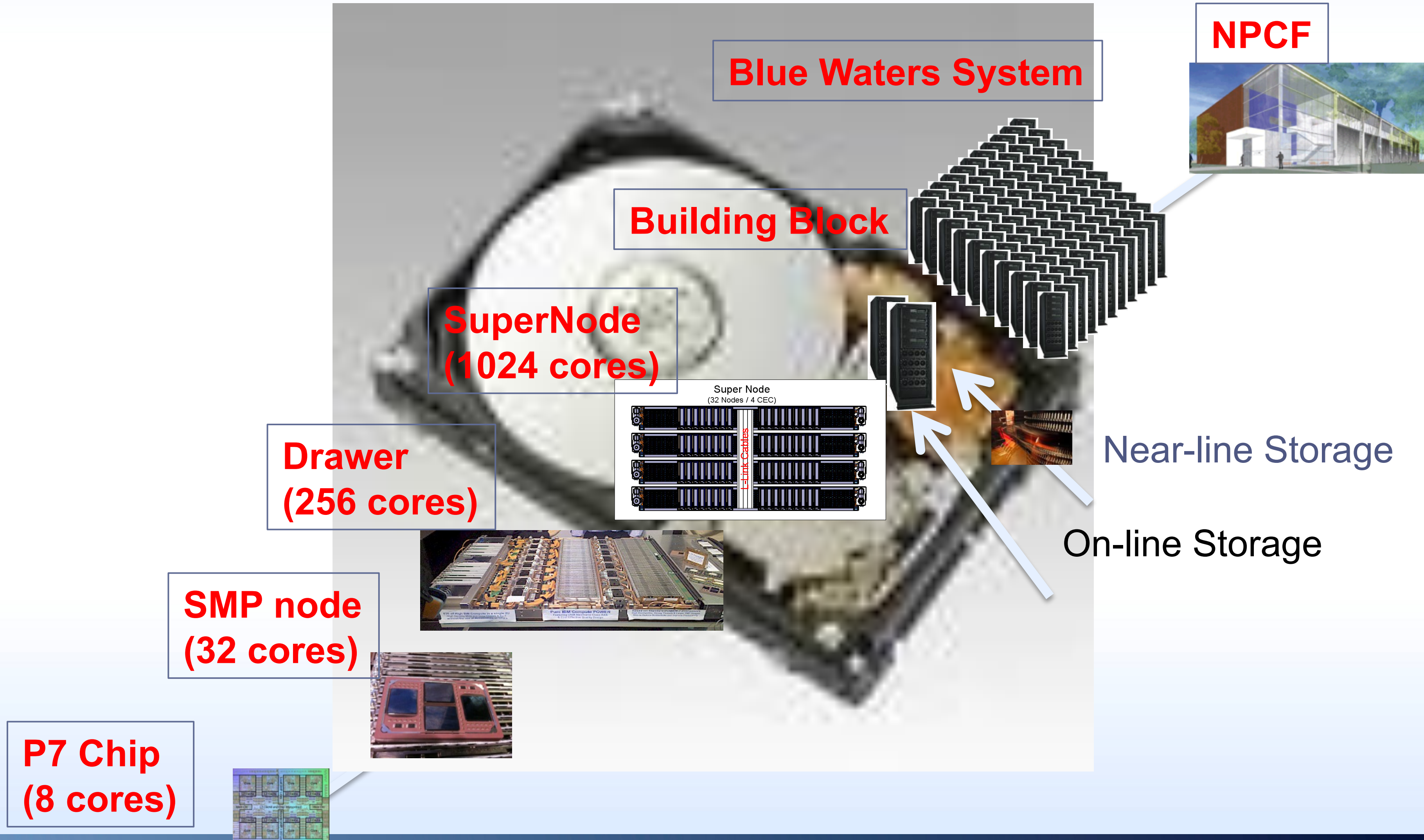
32 IH server nodes
  256 TF (peak)
  32 TB memory
  128 TB/s memory bw
4 Storage systems (>500 TB)
10 Tape drive connections

Blue Waters

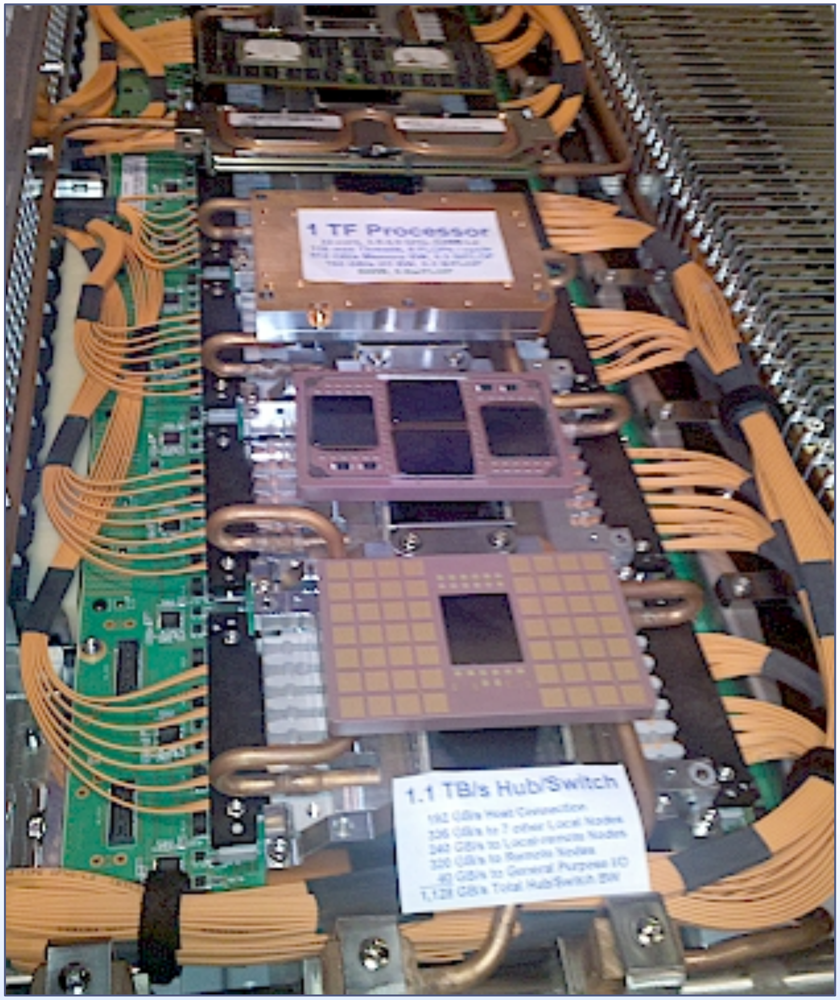
~10 PF Peak
~1 PF sustained
>300,000 cores
~1.2 PB of memory
>18 PB of disk storage
500 PB of archival storage
≥100 Gbps connectivity

Blue Waters is built from components that can be used to build other systems with a wide range of capabilities—from servers to beyond Blue Waters.

From Chip to Entire Integrated System



IBM P7IH Supernode = 128 CPUs/1024 cores



Data Center in a Rack

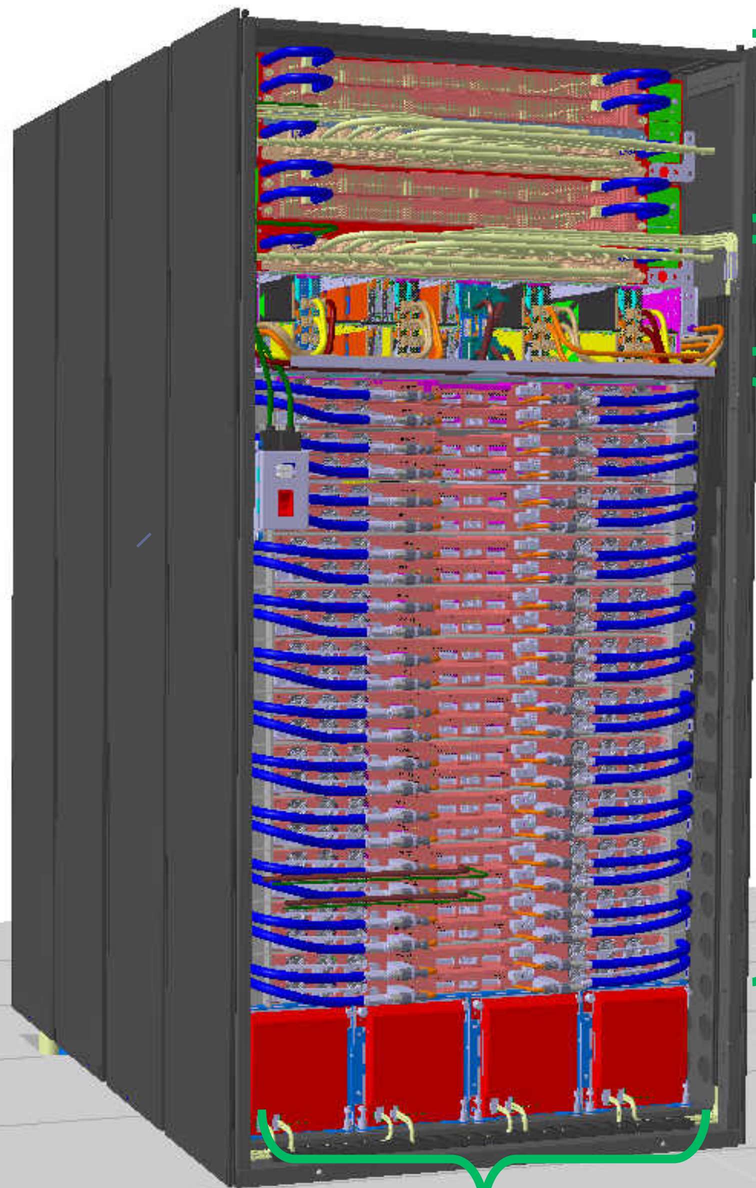
Rack
- 990.6w x 1828.8d x 2108.2
- 39”w x 72”d x 83”h
- ~2948kg (~6500lbs)

Data Center In a Rack
Compute
Storage
Switch
100% Cooling
PDU Eliminated

Input: 8 Water Lines, 4 Power Cords

*Out: **~100TFLOPs / 24.6TB / 153.5TB***
192 PCI-e 16x / 12 PCI-e 8x

WCU
- Facility Water Input
- 100% Heat to Water
- Redundant Cooling
- CRAH Eliminated



BPA
- 200 to **480Vac**
- 370 to 575Vdc
- Redundant Power
- Direct Site Power Feed
- PDU Elimination

Storage Unit
- 4U
- 0-6 / Rack
- **Up To 384 SFF DASD/Unit**
- File System

CECs
- 2U
- 1-12 CECs/Rack
- 256 Cores
- 128 SN DIMM Slots / CEC
- 8,16, (32) GB DIMMs
- 17 PCI-e Slots
- Imbedded Switch
- Redundant DCA
- NW Fabric
- **Up to: 3072 cores, 24.6TB (49.2TB)**

Diverse Large Scale Computational Science

Science areas	Multi-physics, Multi-scale	Dense linear algebra (DLA)	Sparse linear algebra (SLA)	Spectral Methods (FFT)s (SM-FFT)	N-Body Methods (N-Body)	Structured Grids (S-Grids)	Unstructured Grids (U-Grids)	Data Intensive
Nanoscience	X	X	X	X	X	X		
Chemistry	X	X	X	X	X			
Fusion	X	X	X			X	X	X
Climate	X		X	X		X	X	X
Combustion	X		X			X	X	X
Astrophysics	X	X	X	X	X	X	X	X
Biology	X	X					X	X
Nuclear		X	X		X			X
System Balance Implications	General Purpose balanced System	High Speed CPU, High Flop/s rate	High Performance Memory	High Interconnect Bisection bandwidth	High Performance Memory	High Speed CPU, High Flop/s rate	Irregular Data and Control Flow	High Storage and Network bandwidth

Programming Environment

Resource manager: Batch and interactive access

Performance tuning: HPC and HPCS toolkits, open source tools

Parallel debugging at full scale

Environment: Traditional (command line), Eclipse IDE (application development, debugging, performance tuning, job and workflow management)

Languages: C/C++, Fortran (77-2008 including CAF), UPC

IO Model: Global, Parallel shared file system (>10 PB) and archival storage (GPFS/HPSS) MPI I/O

Libraries: MASS, ESSL, PESSL, PETSc, visualization…

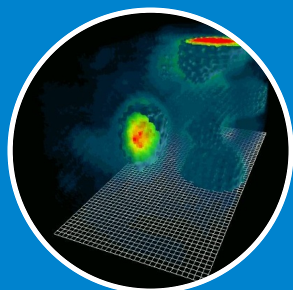
Programming Models: MPI/MP2, OpenMP, PGAS, Charm++, Cactus

Low-level communications API supporting active messages (LAPI)

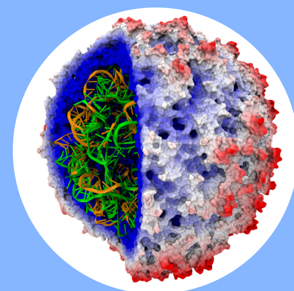
Full – featured OS (AIX or **Linux**), Sockets, threads, shared memory, checkpoint/restart

Hardware
Multicore POWER7 processor with Simultaneous MultiThreading (SMT) and Vector MultiMedia Extensions (VSX)
Private L1, L2 cache per core, shared L3 cache per chip
High-Performance, low-latency interconnect supporting RDMA

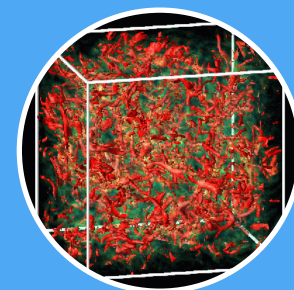
Blue Waters Benchmark Codes



MILC
(lattice
QCD)



NAMD
(molecular
dynamics)



Pseudospectral
Method
(turbulence)

NSF Challenge: ≥1 Sustained petaflop/s

Photos courtesy of NERSC, UIUC, IBM

Path to Petascale

USERS

- Aerospace
- Automotive
- Bio/Chemical
- Oil & Gas
- Pharma
- Energy
- Finance
- DOE/DoD

DEVELOPERS

- Proprietary 50%
- Commercial 30%
- Open Source 20%

WORKFORCE

- Corporate
- Technical
- University
- HPC experts
- Domain experts
- Federal labs

2 Paths to Blue Waters

NSF Allocation

- Allocation 80%
- Peer Review
- Faculty, Labs
- Industry
- Tech Support
- FREE cycles

PSP

- Allocation 7%
- Proprietary work
- Supply Chain
- Com'l licensing
- Tech Support
- Cost-Recovery

National Petascale Computing Facility

$72.5M, 25MW, LEED Gold+
Military-grade security
Non-classified
88,000 ft^2

THANK YOU!

http://industry.ncsa.illinois.edu

www.ncsa.illinois.edu/BlueWaters

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign