Welcome!..

The Second International Workshop on
DATA-AWARE DISTRIBUTED COMPUTING
DADC 2009

JUNE 9, 2009
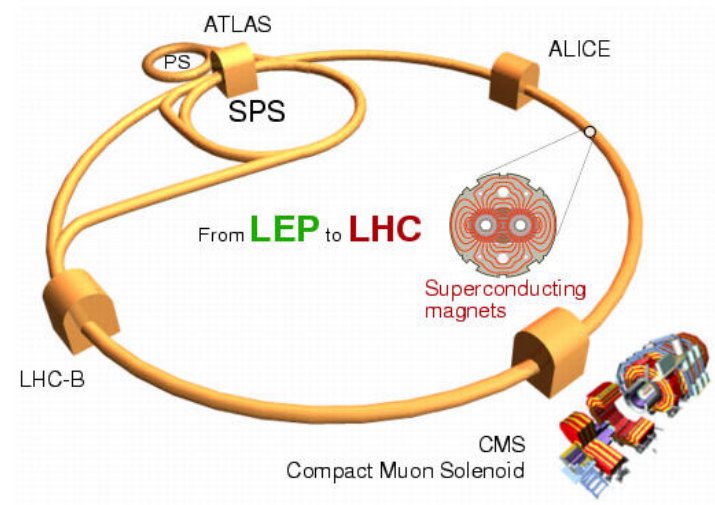MUNICH, GERMANY

Tevfik Kosar
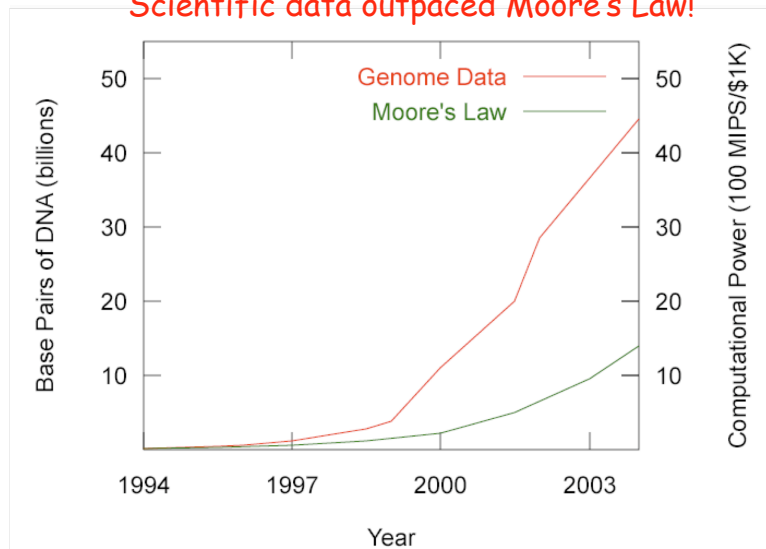DADC'09 Workshop Chair

The Data Deluge

Demand for data in all areas of science!

Application	Area	Data Volume
VISTA	Astronomy	100 TB/year
LIGO	Astrophysics	250 TB/year
WCER EVP	Educational Technology	500 TB/year
LSST	Astronomy	1000 TB/year
BLAST	Bioinformatics	1000 TB/year
ATLAS/CMS	High Energy Physics	5000 TB/year

The Large Hadron Collider (LHC)

ATLAS
PS
SPS
ALICE
From LEP to LHC
Superconducting magnets
LHC-B
CMS
Compact Muon Solenoid

Scientific data outpaced Moore's Law!

Genome Data
Moore's Law

Base Pairs of DNA (billions)
Computational Power (100 MIPS/$1K)
Year

Need to Change in Computing Paradigm?

New Paradigms

☑ Data-aware Distributed Computing

⭐ Data-aware scheduling

⭐ Data-aware storage

⭐ Data-aware workflow management

⭐ Data-aware resource allocation

⭐ .. and other solutions tailored for data-intensive applications

Last Year's DADC @HPDC'08

⭐ **Keynote Address - Malcolm Atkinson**: "Data Streaming Patterns for Distributed Data Exploration"

⭐ **Invited Talk - Jacek Becla**: "Extremely Large Databases for Data Intensive Computing"

⭐ 7 Paper presentations

⭐ **Panel Discussion - Ian Foster, Geoffrey Fox, Reagan Moore, Dan Reed:** "Challenges and New Trends in Data Intensive Science"

⭐ *Slides available at http://www.cct.lsu.edu/~kosar/dadc08*

Today @DADC'09

- ⭐ **Invited Talk - Florian Schintke**: "Scalaris - Methods for a Globally Distributed Key-Value Store with Strong Consistency"

- ⭐ **Paper 1**: "Data-Driven Batch Scheduling"

- ⭐ **Paper 2**: "A Distributed Architecture for Data Mining and Integration"

- ⭐ **Invited Talk - Daniel S. Katz**: "Data Challenges Faced by the Teragrid"

- ⭐ **Paper 3**: "Balancing TCP Buffer vs Parallel Streams in Application Level Throughput Optimization"

- ⭐ **Paper 4**: "Abstract Storage: Moving File Format-specific Abstractions into Petabyte-scale Storage Systems "

Program Committee:

Micah Beck, *University of Tennessee*
John Bent, *Los Alamos National Laboratory*
Ann Chervenak, *USC Information Sciences Institute*
Alok Choudhary, *Northwestern University*
Ewa Deelman, *USC Information Sciences Institute*
Renato Figueiredo, *University of Florida*
Geoffrey Fox, *Indiana University*
Peter Kacsuk, *Hungarian Academy of Sciences*
Dan Katz, *University of Chicago*
Peter Kunszt, *Swiss National Computing Center*
Erwin Laure, *CERN*
Reagan Moore, *San Diego Supercomputing Center*
Don Petravick, *Fermi National Accelerator Laboratory*
Ioan Raicu, *University of Chicago*
Sanjay Ranka, *University of Florida*
Doron Rotem, *Lawrence Berkeley National Laboratory*
Florian Schintke, *Zuse Institute Berlin*
Jennifer Schopf, *Argonne National Laboratory*
Alex Sim, *Lawrence Berkeley National Laboratory*
Ian Taylor, *Cardiff University*
Douglas Thain, *University of Notre Dame*
Brian Tierney *Lawrence Berkeley National Laboratory*
Bernard Traversat, *Sun Microsystems*
Sudharshan Vazhkudai, *Oak Ridge National Laboratory*
Andrew Wendelborn, *University of Adelaide*
Mike Wilde, *Argonne National Laboratory*