TACC's Stampede Project: Intel MIC for Simulation and Data-Intensive Computing

Jay Boisseau, Director

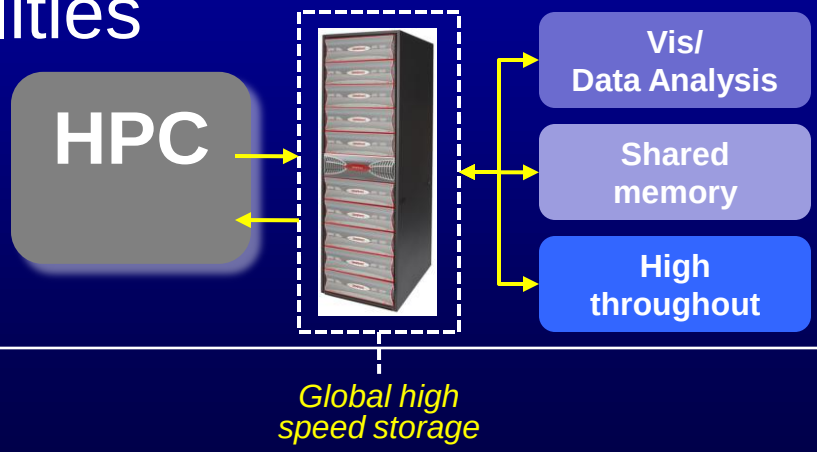
April 17, 2012

TACC Vision & Strategy

- *Provide the most powerful, capable computing technologies and techniques* that enable people—researchers, educators, developers, engineers, businessmen, etc.—to advance science and society.
- *Provide leadership in the advanced computing community* in technology R&D, support, education, and expertise to ensure maximum impact of current and future technologies in diverse applications.
- *Enable transformational science and societal impacts* that change, influence, and improve our understanding of the world, and the world itself.

Lonestar: Enabling World-Class Science

- HPC capabilities
  - 1888 Dell 1955 blades, dual socket
  - 6-core Intel Westmere (3.3GHz)
  - 302 Tflops, 43 TB memory
- Shared memory capabilities
  - 16 1TB shared memory nodes
- Remote visualization capabilities
  - 16 Nvidia Fermis -> 48 (2012)
- 750 US research projects!



TACC

Longhorn: Most Powerful Interactive Remote Visualization System in the World

Longhorn specs

- 256 Dell Quad-core Intel Nehalem Nodes
  - 8 cores/nodes, 2048 total cores
- 128 NVIDIA Quadroplexes
  - 4GPUs/node, 512 total GPUs)
- QDR InfiniBand Interconnect
- NSF award for $7M to enable remote visualization across the US

TACC

Stallion: Highest Resolution Display Environment in the World

TACC

Computational Science is Not Just Modeling and Simulation

- Modeling & simulation
  - Simulation of mathematical models
  - Must store, visualize/analyze simulation output
- Data-enabled science
  - Facilitated by digital data collection
  - Often, no formal mathematical laws -> statistics
    - Mine/analyze data (then work toward models)
  - “Physics vs. stamp collecting” (Rutherford): not true
  - “End of science (simulation)” (***Wired***): even less true

“Big Data”

- Computing power and storage growing rapidly, but…
- ***Digital data*** measurement & collection is growing even more rapidly
- Understanding the world requires analyzing data, not just solving questions
- “Big data” is the new big buzzword in computing, ***in business and government as well as science***

TACC

Massive Computing Requires Massive Data Storage: ***Corral…***

- 5 petabytes (replicated) of DataDirect Networks online disk storage
- Mulitple access mechanisms
  - MySQL & Postgres SQL databases
  - Lustre parallel filesystem
  - iRODS
  - Web-based access
- Can easily expand to petabytes
- Designed for hosting data collections!

TACC

... and the ***Ranch*** Data Archival System

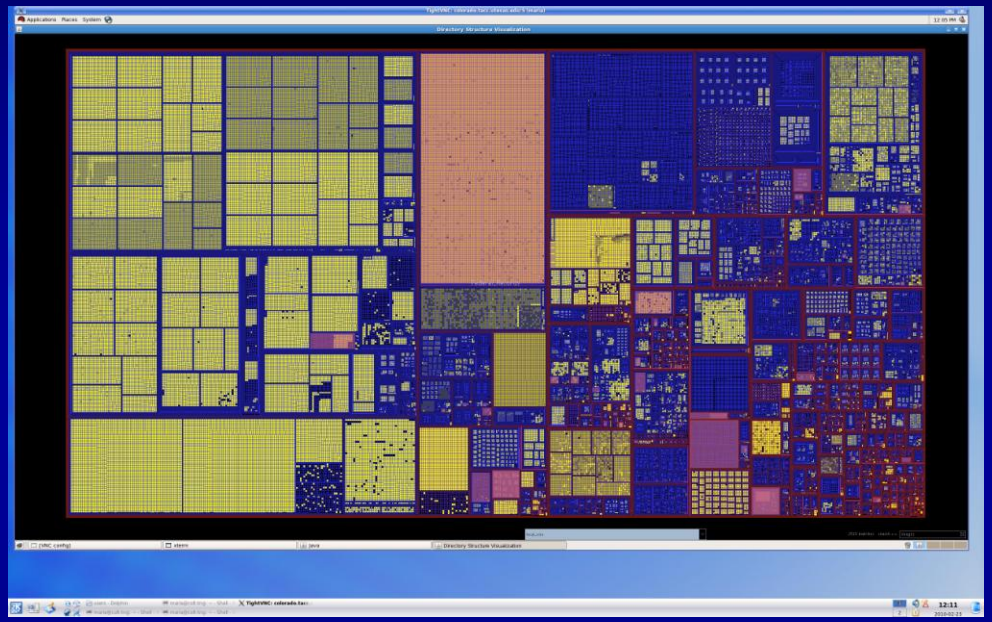
- 50 petabytes tape capacity in Sun StorageTek Silos
  - 10,000 1TB tapes, 6000 5TB
  - Used for long-term storage
  - Access provided to users of other TACC resources
  - Potential for up to 100PB

TACC

A Window on the Archive of the Future

PI: Maria Esteva, Weijia Xiu, Texas Advanced Computing Center

- Project with the National Archives and Records Administration (NARA) is developing experimental workflows and visualization tools to represent, analyze, and interact with massive datasets.



A tree map , information visualization representing the entire Federal Records Collection.

Impact: these tools will help future archivists organize government documents, while furthering public access to open records.

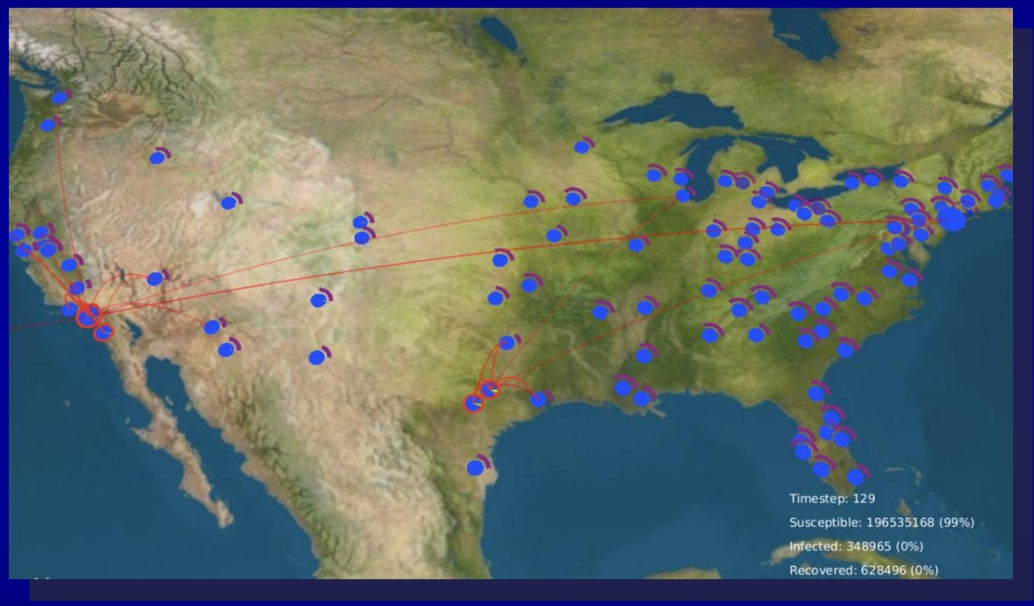
TACC

H1N1 Flu Outbreak Simulation

Ned Dimitrov, Lauren Meyers, UT Comp. Bio

- Project modeled potential spread of epidemic based on locations and transportation

- *Impact: produced insights toward helping understand how to minimize chance of pandemic by placement of antiviral drugs*



TACC

TACC & Sustainable Places Project

- TACC is new member of Capitol Area Texas Sustainability (CATS) “Sustainable Places Analytics Tool” project
- Modeling cities is a big data problem
  - Mining vast amounts of data (big fast storage)
    - People, buildings, autos, companies, power, water…
  - Executing huge statistical models (big fast compute)
    - Traffic, development, environmental quality….
  - And interpreting results at all scales (big viz)
    - Seeing citywide data with resolution of homes, autos
- Future : massive data mining, ensemble models, scenario/what-if decision simulations…

TACC Is Going Big Into Big Data

- ✓ Just upgraded Corral, Ranch
- ✓ Already supporting many data-driven projects
- ✓ Hiring new data mining, data mgmt experts
- ✓ Stampede project will seek out new data – driven projects
- ✓ Just received $10M commitment for designing, deploying new data services

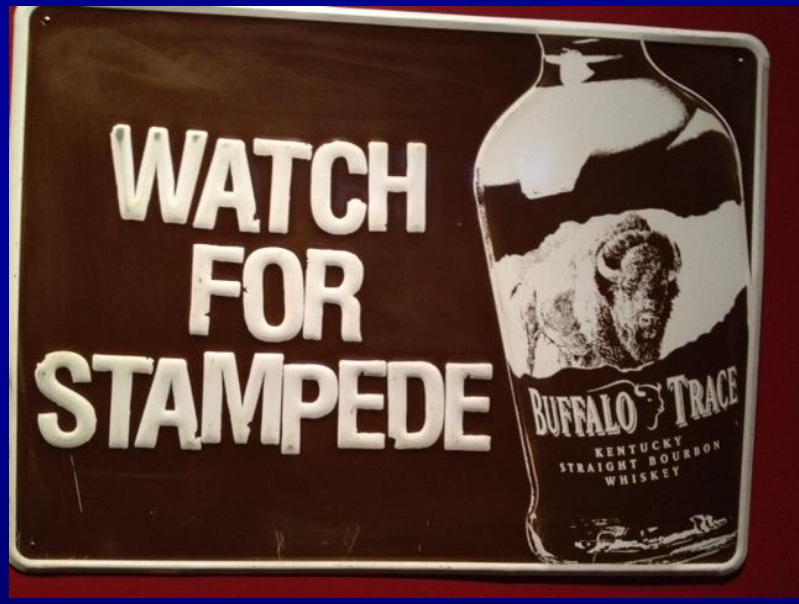
TACC

Forthcoming Data Infrastructure

- Received $10M commitment from O'Donnell Foundation for new data intensive computing infrastructure
  - Start massive parallel file system (~25PB)
  - Start large MapReduce-style cluster (~18PB)
  - Expand Corral further for data collections?
  - Science portal/gateway hosting infrastructure

Stampede – Coming January 2013

- 10 petaflops (PF) peak performance
  - 2 PF Linux cluster: Dell nodes w/Intel Xeon E5 procs
  - 8 PF Intel MIC co-processors
- 56Gb FDR InfiniBand
- 250+ TB memory
- 14+ PB disk, 300 GB/s
- 16 1TB shared memory nodes
- 128 Nvidia Kepler 2 GPUs



TACC

What is MIC?

- MIC is Intel's answer to NVIDIA's and AMD's GPUs
  - For the compute market! (not graphics)
- Stage 1: Knights Ferry (KNF)
  - SDP: Software Development Platform
  - Intel has granted early access to KNF to several dozen institutions
  - Training has started
  - TACC had access to KNF hardware since early March 2011 and is hosting a system since mid March 2011
  - TACC continues to evaluate the hardware and programming models
  - Talk based on our experiences with KNF
- Stage 2: Knights Corner (KNC)
  - First product expected early 2013

MIC Ancestors

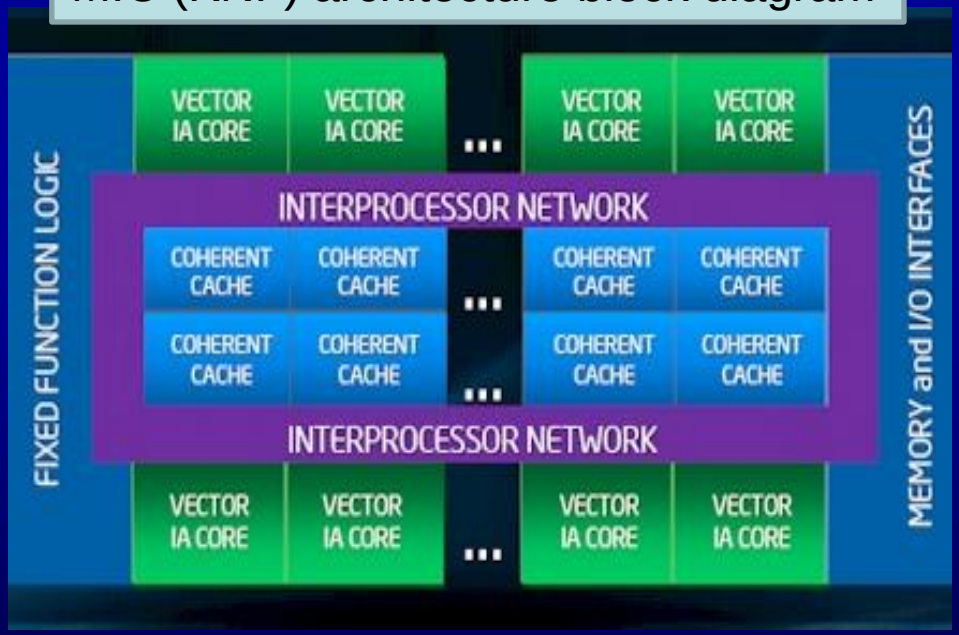
- 80-core Terascale research program
- Larrabee many-core visual computing
- Single-chip Cloud Computer (SCC)

TACC

MIC Architecture

- Many cores on the die
- L1 and L2 cache
- Bidirectional ring network
- Memory and PCIe connection

MIC (KNF) architecture block diagram



Knights Ferry SDP

- Up to 32 cores
- 1-2 GB of GDDR5 RAM
- 512-bit wide SIMD registers
- L1/L2 caches
- Multiple threads (up to 4) per core
- Slow operation in double precision

Knights Corner (first product)

- 50+ cores
- Increased amount of RAM
- Details are under NDA
- Double precision half the speed of single precision (canonical ratio)
- 22 nm technology

TACC

What we at TACC like about MIC

(and we think that **you** will like this, too)

- Intel's® MIC is based on x86 technology
  - x86 cores w/ caches and cache coherency
  - SIMD instruction set
- Programming for MIC is similar to programming for CPUs
  - Familiar languages: C/C++ and Fortran
  - Familiar parallel programming models: OpenMP & MPI
  - MPI on host and on the coprocessor
  - Any code can run on MIC, not just kernels
- Optimizing for MIC is similar to optimizing for CPUs
  - Make use of existing knowledge!

Key elements of this talk highlighted!

TACC

Coprocessor vs. Accelerator

- Differences
  - Architecture: x86 vs. streaming processors
    coherent caches vs. shared memory and caches
  - HPC Programming model:
    extension to C++/C/Fortran vs. CUDA/OpenCL
    OpenCL support

    Threading/MPI:
    OpenMP and Multithreading vs. threads in hardware
    MPI on host and/or MIC vs. MPI on host only
  - Programming details
    offloaded regions vs. kernels
  - Support for any code: serial, scripting, etc.
    Yes No
- Native mode: Any code may be “offloaded” as a whole to the coprocessor

Programming Models

Ready to use on day one!

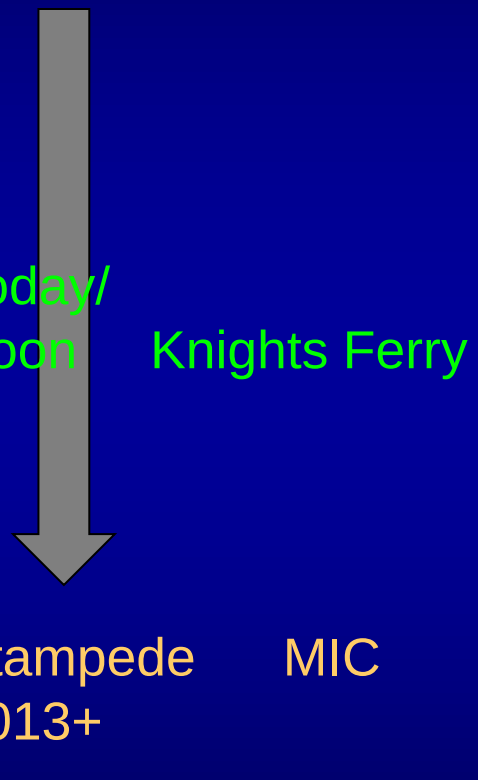
- TBB's will be available to C++ programmers
- MKL will be available
  - Automatic offloading by compiler for some MKL features
- Cilk Plus
  - Useful for task-parallel programing (add-on to OpenMP)
  - May become available for Fortran users as well
- OpenMP
  - TACC expects that OpenMP will be the most interesting programming model for our HPC users

Adapting and Optimizing Code for MIC

- Hybrid Programming
  - MPI + Threads (OpenMP, etc.)

- Optimize for higher thread performance
  - Minimize serial sections and synchronization
- Test different execution models
  - Symmetric vs. Offloaded
- Optimize for L1/L2 caches

- Test performance on MIC
- Optimize for specific architecture
- Start production



TACC

Roadmap

What comes next?

How to get prepared?

- General expectation:
  Many of the upcoming large systems will be accelerated
- Intel's MIC coprocessor is on track for 2013
  - TACC's assessment!
  - Viable alternative to GPUs for compute
  - Expect that (large) systems with MIC coprocessors will become available in 2013 at/after product launch
- OpenMP is the main HPC programming model for MIC
  - If you are not using TBBs
  - If you are not spending all your time in libraries (MKL, etc.)
- Training has begun; Stampede-specific training in Dec12