The HPC Workforce Challenge

Steve Finn, Emagine IT

Steve Conway, IDC

Summary of talks from HPC User Forum meeting in Austin TX

Session Chair: Paul Muzio, City University of New York

Addressing the Critical Issue of HPC Workforce Development

- Data Center Management Minor at UT-Knoxville: Bridging the Gap between Computing and Facilities, Jim Serafin (ORNL)
- Computational Science Education at the Technical University of Munich, Hans-Joachim Bungartz
- HPC Curriculum at TACC: Melyssa Fratkin
- Preparing Students for HPC Careers in the Energy Sector: Jan Odegard, Rice University

See the full slides or watch the videos

- We'll be using the slides of the original presenters
- Some have been reduced to fit in our time slot
- Videos at:

http://insidehpc.com/video-gallery-hpc-user-forum-in-austin-sept-6-8-2016/

- The full presentations are available at:

http://hpcuserforum.com/presentations.html

Bridging the Gap Between Computer Science and Facility Engineering for Data Center Management

HPC User Forum

September 7, 2016

Jim Serafin, PE, PMP

Computational Complex Facilities Manager, UT-Battelle, LLC

ORNL is managed by UT-Battelle
for the US Department of Energy

OAK RIDGE National Laboratory

THE UNIVERSITY OF TENNESSEE KNOXVILLE

DATA Center Technology

Data Center Technology and Management

Minor Degree Program, UTK CoE

Dr. Mark E Dean, PhD

Fisher Distinguished Professor
UT Knoxville College of Engineering,
EECS Dept.

OAK RIDGE National Laboratory

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Data Center Technology and Management Minor Degree Program

The College of Engineering offers a minor in datacenter technology and management to establish broad awareness of datacenter requirements, design and management technologies and methodologies, including reliability, security, network systems, storage systems, industrial design, systems management, application environments/ management, operations, logistics and energy efficiencies.

The program is not expected to deliver deep knowledge in a specific area, but to give a student sufficient basic insight on datacenter operational characteristics and requirements.

Proposed Minor in "Datacenter Technology and Management".

Objectives: Establish broad awareness of datacenter requirements, design and management technologies and methodologies, including reliability, security, network systems, storage systems, industrial design, systems management, application environments/management, operations, logistics and energy efficiencies. The program is not expected to deliver deep knowledge in a specific area, but give a student sufficient basic insight on datacenter operational characteristics and requirements.

Disciplines appropriate to entry into this program include: CS, CpE, EE, IE, and ME.

The NSA will process clearances for anyone in the program who applies for the NSA's internship program. There will also be opportunities for students to complete the program and intern at commercial companies interested in this skill set.

Requirements:
Total number of courses required = 7
Total number of course hours required = 22
Note: internships should occur at the end of a student's sophomore year and at the end of a student's junior year...primarily during the summer semester. Internships will be hosted at ORNL, NSA datacenters and commercial datacenters (where appropriate).

Core Courses Required – 9 Hours:

IE 483 – Introduction to Reliability Engineering

3 Credit Hours: Probabilistic failure models and parameter estimation (maximum likelihood, Bayes techniques). Model identification and comparison, accelerated life tests, failure prediction, system reliability, preventive maintenance, and warranties. (Same as Chemical and Biomolecular Engineering 483; Nuclear Engineering 483; Mechanical Engineering 483.)
Registration Restriction(s): Minimum student level – junior.
Registration Permission: Consent of instructor.

IE 484 – Introduction to Maintainability Engineering

3 Credit Hours: Principles of maintenance and reliability engineering and maintenance management. Topics include information extraction from machinery measurements, rotating machinery diagnostics, nondestructive testing, life prediction, failure models, lubrication oil analysis, establishing a [illegible] maintenance program, and computerized maintenance [illegible]

OAK RIDGE National Laboratory

Data Center Technology and Management Minor Degree Program

Disciplines appropriate for entry into this program include:

Computer Science, Computer Engineering, Electrical Engineering, Industrial Engineering, and Mechanical Engineering.

Requirements:

Total number of courses required = 6
(+1 for non-EECS students, COSC 130 or equivalent)

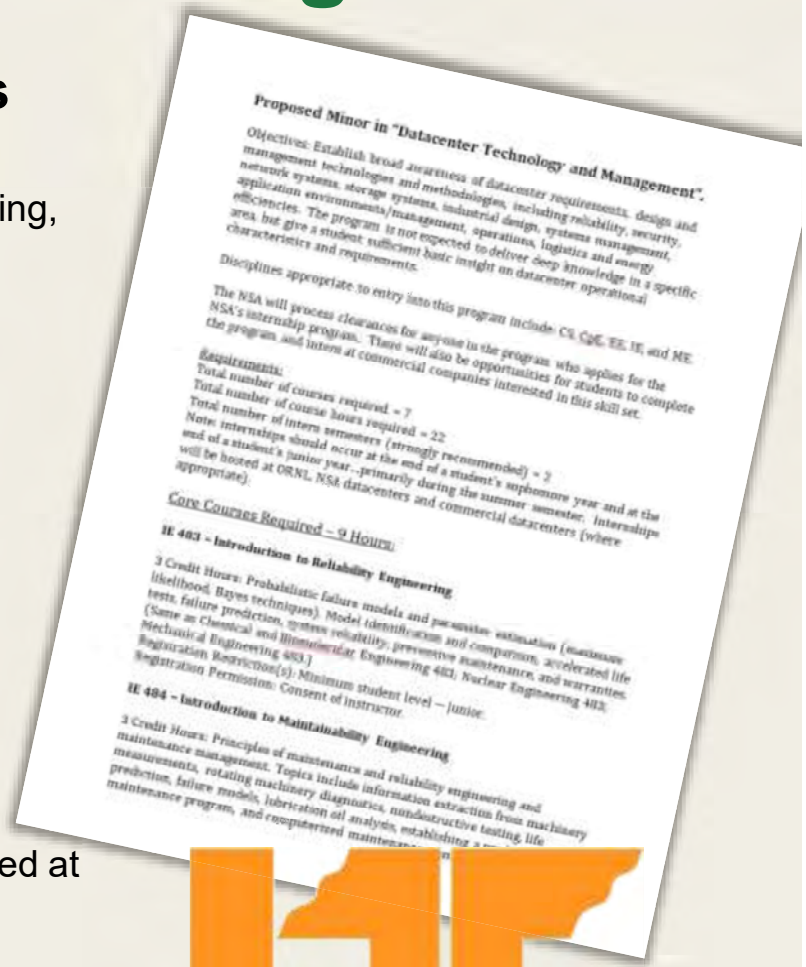
Total number of course hours required = 18
(+4 for non-EECS students, COSC 130 or equivalent)

Total number of intern semesters (strongly recommended) = 2

Note: internships should occur at the end of a student's sophomore year and at the end of a student's junior year… primarily during the summer semester. Internships will be hosted at ORNL and commercial datacenters (where appropriate).

Additional minor specific classes include:

- IE 483 - Introduction to Reliability Engineering
- IE 484 - Introduction to Maintainability Engineering



Proposed Minor in "Datacenter Technology and Management".

Objectives: Establish broad awareness of datacenter requirements, design and management technologies and methodologies, including reliability, security, network systems, storage systems, industrial design, systems management, application environments/management, operations, logistics and energy efficiencies. The program is not expected to deliver deep knowledge in a specific area, but give a student sufficient basic insight on datacenter operational characteristics and requirements.

Disciplines appropriate to entry into this program include: CS, CpE, EE, IE, and ME.

The NSA will process clearances for anyone in the program who applies for the NSA's internship program. There will also be opportunities for students to complete the program and intern at commercial companies interested in this skill set.

Requirements:
Total number of courses required = 7
Total number of course hours required = 22
Total number of intern semesters (strongly recommended) = 2
Note: internships should occur at the end of a student's sophomore year and at the end of a student's junior year…primarily during the summer semester. Internships will be hosted at ORNL NSA datacenters and commercial datacenters (where appropriate).

Core Courses Required – 9 Hours:

IE 483 – Introduction to Reliability Engineering

3 Credit Hours: Probabilistic failure models and parameter estimation (maximum likelihood, Bayes techniques). Model identification and comparison, accelerated life tests, failure prediction, system reliability, preventive maintenance, and warranties. (Same as Chemical and Biomedical Engineering 483; Nuclear Engineering 483; Mechanical Engineering 483.)
Registration Restriction(s): Minimum student level – junior.
Registration Permission: Consent of instructor.

IE 484 – Introduction to Maintainability Engineering

3 Credit Hours: Principles of maintenance and reliability engineering and maintenance management. Topics include information extraction from machinery measurements, rotating machinery diagnostics, nondestructive testing, life prediction, failure models, lubrication oil analysis, establishing a
maintenance program, and computerized maintenance

OAK RIDGE National Laboratory

ECE463 – Description and Goals

Class Description

This is a senior-level course focused on developing multi-disciplinary insights into datacenter technology. The primary objective of this coarse is to deliver broad awareness of datacenter requirements, design and management technologies and methodologies. This includes: reliability, security, network systems, storage systems, industrial design, systems management, operating environments, application environments/management, operations, logistics and energy efficiencies. The program is not expected to deliver deep knowledge in a specific area, but give a student sufficient basic insight on datacenter operational characteristics and requirements. Engineering disciplines which can take this course include EE, CpE, CS, ME and IE.

Rationale

The complexity and diversity of the computing services and computing environments has made datacenter design, operation and management challenging. This course fills a critical gap in knowledge around datacenter technology and best practices in the datacenter infrastructure to deliver information services.

Learning outcomes

Students will be presented with a broad overview of datacenter technology and the interactions and interdependencies of datacenter components (facilities, cooling, power, management, computers, networks, OSes, applications, security, safety, etc.). Students will gain knowledge in datacenter infrastructure, operations and management best practices.

Class Structure

Sections of class will be delivered by experts and practitioners in datacenter deployment and management from ORNL. Classes are meant to be interactive and foster lots of questions and discussion from students.

Text Book: Data Center Handbook, by Hwaiyu Geng, Wiley Publishing 2015

ECE463 – Introduction to Data Center Tech. (Course Syllabus/Material Covered)

- **Introduction to Course (1)**
  - ✓ What to expect from this course
  - ✓ What is a Data Center
- **Data Center Strategy (2)**
  - ✓ Datacenter Strategy overview
  - ✓ Models (Cost and Operations)
- **Cooling (2)**
  - ✓ Fundamentals
  - ✓ Technologies
  - ✓ Design
  - ✓ Operations
  - ✓ Future Direction
- **Efficiency metrics for data centers (1)**
- **Infrastructure/Space (2)**
  - ✓ Infrastructure/Space
  - ✓ Security and integration
  - ✓ Managing the space
- **Project Management (1)**
  - ✓ General Topics – What is a Project and what is project management
  - ✓ Project Life Cycle
  - ✓ Organizational Influences
  - ✓ Planning Process
- **Software Applications (1)**
- **Energy Efficiency for Mechanical and electrical systems (2)**
  - ✓ Measures of efficiency
  - ✓ Technologies (Inherent Efficiencies)
  - ✓ Heat Recovery/Reuse
  - ✓ Air Flow Management
  - ✓ Case Studies/ White Papers
- **Power (4)**
  - ✓ Power Systems
  - ✓ Metering, SCADA, EEM systems
  - ✓ Load Factors, capacity, …
  - ✓ Efficiency and sustainability
  - ✓ Reliability, Safety, Standards,
- **Data Center Network (2)**
- **Industrial Controls (2)**
  - ✓ SCADA Overview, Brands/Types of SCADA
  - ✓ System Hardware (Metering, Networking, Communications, Security Considerations)
- **Commissioning a Data Center (1)**
- **Management of a Data Center During Construction (1)**
- **Data Center Facility Operations (1)**
- **Computing Configuration Management DCIM (2)**

Computing and Facilities: Who knows both?

ORNL's HPC partners have expressed an interest in workplace development to address a lack of skilled managers understanding both computing and facilities to tackle unique challenges of large data center environments.

Added for HPC Forum

Rationale for this class

- The environment within which we design and operate data center facilities has become more complex.
  - Federal mandates to consolidate data centers causing many to go away in the next 10 years.
  - Federal sustainability goals.
  - 10MW is no longer unheard of for a single computer, current Exascale estimates are closer to 30MW.

- "Outside the box" thinking.
- Scalability of our data center infrastructure.

Added for HPC Forum

Develop Internship Pipeline

- How many of you landed in this industry by accident?
- This class will help ORNL and it's partners migrate to a structured plan to attract people into this field.
- Internships strongly encouraged.
- Example lesson: What is Total Cost of Ownership?
  - How do I contribute to the bottom line?
    programmers, system administrators, computer engineers
  - Tradeoffs between investing in computing performance versus redundancy?

DATA Center Technology

Class Project: Data Center Design

Oak Ridge National Laboratory

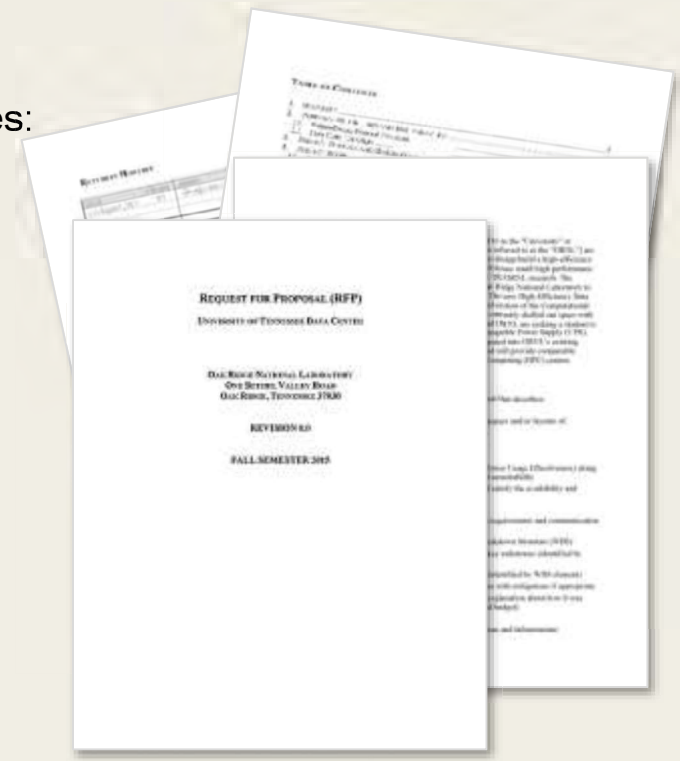
The University of Tennessee Knoxville

Project: Summary

The University of Tennessee Knoxville (hereinafter referred to as the "University" or “UTK”) and the Oak Ridge National Laboratory (hereinafter referred to as the "ORNL”) are seeking proposals from experienced and qualified students to design/build a high-efficiency state of the art data center for the UTK. This data center will house small high performance computers, business systems, and servers that will support UTK/ORNL research. The University is building a Remote Data Center (DC) at the Oak Ridge National Laboratory to meet the changing demands of the high density computing. The new High-Efficiency Data Center will be located on the second floor of a new expanded section of the Computational Science Building (CSB). This room and the floor below are currently shelled out space with the two floors above being finished office spaces. UTK and ORNL are seeking a student to design/build a data center with cooling, rack space, Un-interruptible Power Supply (UPS), Power distribution units (PDUs), generator, etc. that is integrated into ORNL’s existing electrical and Information Technology (IT) infrastructure and will provide comparable availability for the IT equipment as the High Performance Computing (HPC) centers.

Proposal Guidelines and Deliverables

Written Design Proposal Document



- At a minimum, the student is to provide a written proposal that describes:
  - Data center design strategy
  - Design narrative description and supporting schematics and/or layouts of:
    - Electrical distribution system
    - Cooling systems
    - Networking design
  - Measures proposed to obtain the required PUE (Power Usage Effectiveness) along with steps to be taken to maximize efficiency and sustainability
  - Electrical and mechanical configurations that will satisfy the availability and concurrent maintenance requirements.
  - Project Plan inputs:
    - Short description of customer requirements and communication strategy
    - Scope organized by Work Breakdown Structure (WBS)
    - Proposed schedule including key milestones (identified by WBS element)
    - Proposed budgetary estimate (identified by WBS element)
    - Identify risks to project success with mitigations if appropriate
    - Identify contingency with an explanation about how it was determined (both schedule and budget)
  - Future growth strategy
  - Integration into existing ORNL computer operations and infrastructure

Proposal Guidelines and Deliverables

Floor Plans / Drawings

- Also provide floor plans and other drawings that shows the following:
  - The layout of the data center showing IT equipment footprint, ramps to access raised floor areas if proposed, passageways in room, floor loading requirements, fire protection features, security barriers, floor trains, cabinets.
  - Location of electrical distribution equipment (e.g. generator, UPS, PDUs Remote Distribution Units (RDUs), etc.).
  - Routing of electrical circuits.
  - Routing of network circuits.
  - Location of the cooling equipment.
  - Routing of chilled water piping.
  - Network infrastructure to include network cable routing and type of cable to connect this new DC to existing network infrastructure.
  - Single line electrical drawings showing normal power source and generator backed up UPS source and the distribution from these systems for IT and cooling equipment (basic level single line, minimum of breaker/wire/device).
  - Piping and instrumentation drawings showing chilled water sources and piping required for cooling the new DC.

DATA Center Technology

Data Center Facilities Examples

Oak Ridge National Laboratory

The University of Tennessee Knoxville

Data Center Example: Yahoo Chicken Coop

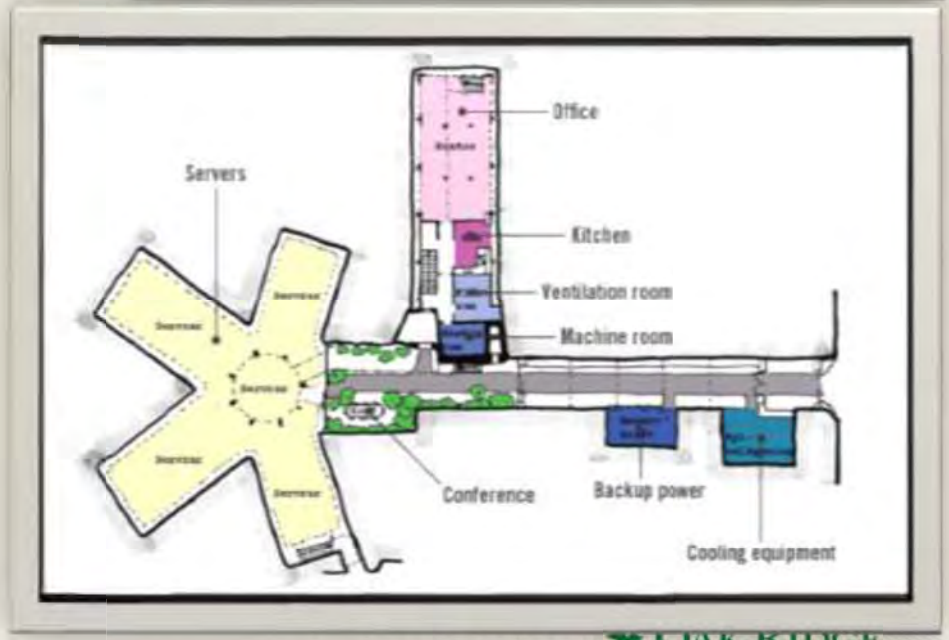
- The 155,000-square-foot data center in Lockport, NY
- Accommodates 50,000 servers
- Cooled almost 100 percent by outside air using 40 percent less energy than typical data centers.
- The building is important because it will effect the efficiency of the data center
- There are many ways to have an efficient building
  - Yahoo Chicken Coop
  - Designed to take advantage of “free cooling”
  - PUE 1.1

Data Center Example: Barcelona Supercomputing Center (BSC)

- The Barcelona Supercomputing Center
  - 19th century Torre Girona chapel in Barcelona, Spain.
- **MareNostrum Supercomputer**
  - "Divine Data Cruncher"
  - Once the most powerful supercomputer in Europe
  - 63.8 TFLOPS PowerPC-based supercomputer
  - ranked 465th in the world in June 2012
- Also hosts a newer 103.2 TFLOPS supercomputer
  - Xeon E5649 processors
- PUE 1.3

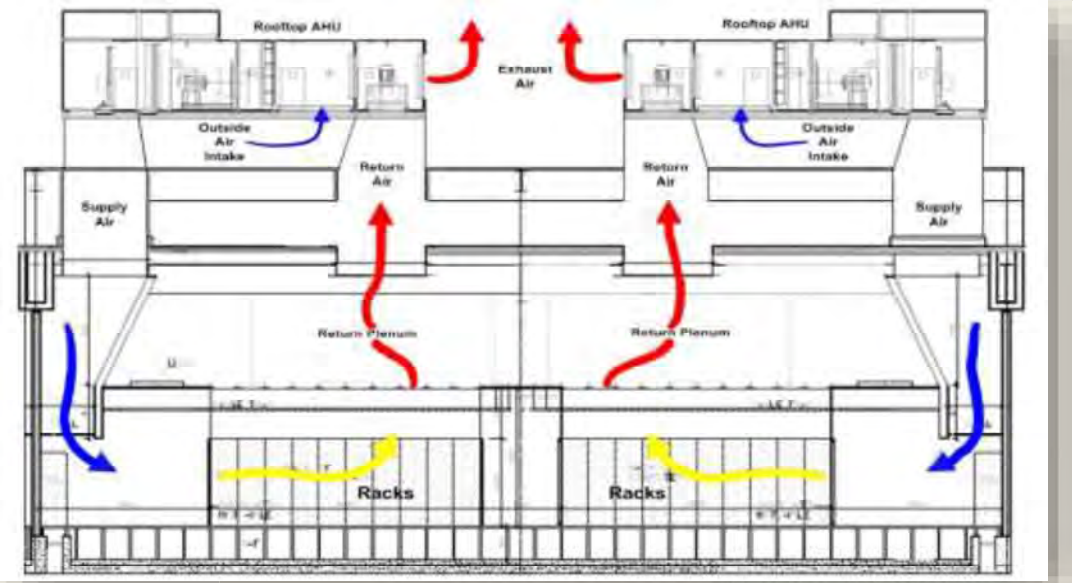
Data Center Example: Stockholm Data Center

- Headquarters of Bahnhof
  - a Swedish Internet Service Provider
- Named **"Pionen White Mountains"**
- Stockholm, Sweden converted nuclear bunker
- A hundred feet under the city of Stockholm
- Designed to take advantage of "free cooling"
  - PUE 1.1



Data Center Example: Microsoft

- **Microsoft Dublin data center**
  - Designed to take advantage of Ireland's low ambient temperatures for "free cooling" using airside economizers
  - Uses shipping container to create modularization to facilitate future expansi
  - PUE 1.25



Data Center Example: Sun

- A decommissioned coal mine in Chubu region on Japan's Honshu island.
- Contains 30 Blackbox self-contained data centers
  - containing a total of 10,000 servers (cores) - scalable to 30,000 cores
- **Blackbox concept** –
  - 250 servers mounted in seven racks inside a standard 20-foot shipping container.
- A subterranean data center easy to secure against unauthorized entry and terrorist attacks.
- The project was initially estimated to cost $405 million and was to start operations in April 2010.

CSE & HPC Education – Within & Beyond Classrooms

Hans-Joachim Bungartz

Technical University of Munich

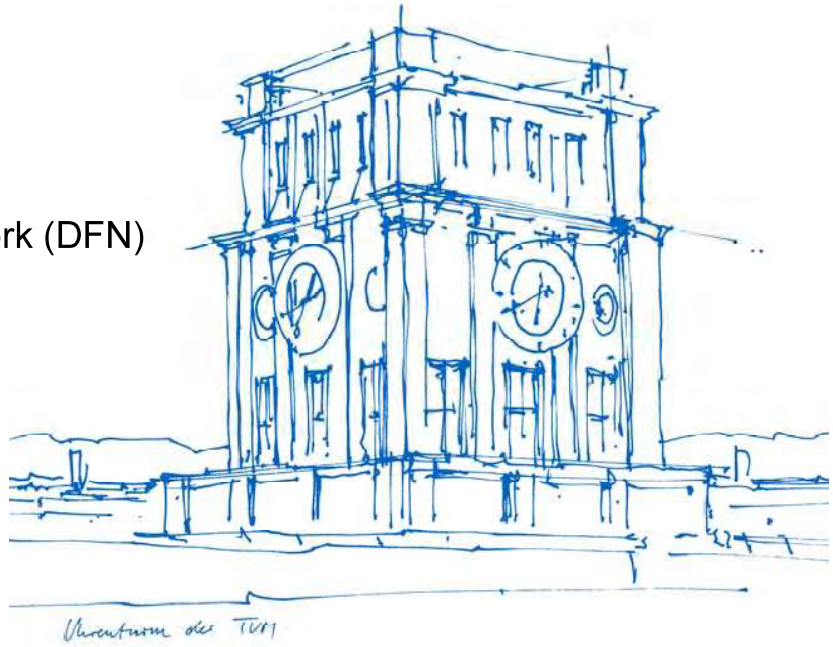
Dean of Informatics, Chair of Scientific Computing

TUM Graduate Dean

Board member, Leibniz Supercomputing Center (LRZ)

Chairman, German National Research & Education Network (DFN)

HPC User Forum, Austin, 7.9.2016



Preface: Why CSE & HPC together?

Computational Science & Engineering is the field with closest links to HPC

- **HPC drives CSE**: CSE – in particular that at extreme scales and for the famous grand challenges – essentially needs HPC
- **CSE drives HPC**: CSE – i.e. simulation, optimization, analytics, … – has always been a main driving force for HPC as well as a main justification for HPC investments
- CSE has its own programs – with frequent HPC modules
- Computational scientists and engineers are those who will most probably deal with HPC in their future careers, either in academia or industry, either as developers or practitioners/users – at a higher percentage than computer scientists, by the way

Hence an integral look at CSE & HPC education … and data belongs to that context, too!

Focus:

- University stuff
- HPC centers have a sophisticated course program (for HPC professionals, esp.)

Challenges

Neither CSE nor HPC "belong" to one specific discipline

- No field considers it as its own responsibility (look at the curricular recommendations of other fields' professional societies, etc., which are missing here)
- If done within classical programs: courses tailored to the respective students' needs
- If done via new/specific (master) programs: heterogeneous knowledge

No broad consensus on what's needed – beyond advanced programming skills and some hardware affinity

- Computational X / simulation technology programs come close, but can also be rather "implementation-free" (focus on modelling, e.g.)
- Endless debates on "is it really a discipline?", "does it really need new programs?", "shouldn't we rather just …?"

Raise the sex-appeal of HPC in times of Google & Co.

- "HPC is outdated, all is done by the cloud, …"

Hence: attracting new talents remains random-based and difficult

Recent Developments

Workshop "Future Directions in CSE Education & Research"

- August 2014, Breckenridge, CO
- Some 30+ participants from all over the world
- Co-organized by SIAM (its CSE branch) and EESI (European exascale)
- Goals:
  - In general: positioning – branding – marketing
  - More concrete: update of the report "Graduate Education in CSE" (2001)
- Understanding of "Computational", esp. w.r.t. HPC & Data Science

Close-to-final version just distributed

(authors: Rüde, Willcox, Curfman McInnes, De Sterck, et al.)

- Cover page
- Table of contents
- Central findings

Contents

Novel programs

Novel courses … in classrooms & labs

Ferienakademie: a summer school in the mountains

Mountains again: SPPEXA doctoral retreat

Student Cluster Competition @ SC ’15

Concluding remarks

CSE- and HPC-Related Programs @ TUM

Existing specializations in many classical undergrad & grad programs

Two established "Computational X" master's programs

- *Computational Mechanics* – since 2000 ➔ emerged from a domain's needs
- *Computational Science & Engineering* – since 2001 ➔ emerged from core technology
- Both are trans-disciplinary (4 or 7 depts., resp.) and international (English language)

One established "honours program"

- *Bavarian Graduate School of Computational Engineering* – since 2005, with Erlangen
- Additional modules, such as a "Team Software Project"
- Awards the "*BGCE Student Paper Prize*" at the SIAM CSE meetings – since 2007

Two new data-related master's programs

- *Mathematics in Data Science* – starts this fall
- *Data Engineering and Analytics* – starts this fall
- Both are pillars of TUM's "Integrative Study Program in Big Data"

Contents

Example #1: Interdisciplinary Lab
"Turbulent Flow Simulation on HPC Systems"

Occasion:

- "Ernst-Otto-Fischer Teaching Award"
- Annual TUM-internal competition for new concepts & innovative course formats

Idea: Teamwork across disciplines (Mechanical Engineering & Informatics)

Format: Labs (experiments & simulations) with accompanying lectures

Goals:

- Team-oriented learning
- Cross-disciplinary setting: 2 student sub-groups with different background

Contents:

- Extension of a 3D Navier-Stokes solver with a turbulence model & MPI
- Doing experiments and simulations, comparing resulting data

Example #1: Interdisciplinary Lab
"Turbulent Flow Simulation on HPC Systems"

Implementation:

- Lectures with core facts from both sides
- Small soft-skills sessions
- Supervised team meetings (fixed point-of-contact for student teams, feedback to lecturers)

Reward:

- Well, the idea got the prize …

Further reading:

- Neumann, Kowitz, Schranner, Azyrnykh: HPC Meets Engineering, Euro-Par 2015, pp. 125-134

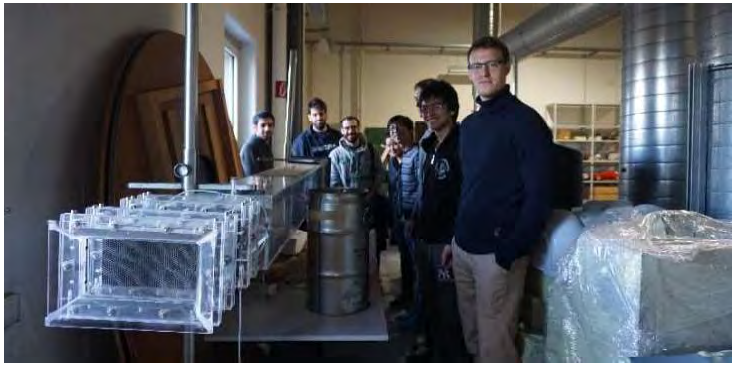
Example #1: Interdisciplinary Lab
"Turbulent Flow Simulation on HPC Systems"

Statistics:

- 14-20 participants, Informatics / CSE / Mechanical Engineering
- 2 weeks Intro (basic code), 6 weeks Extension (turbulence + MPI; mixed teams), 4 weeks Project (individual or team (preferred) – report up to 60 p.)

Reflection:

- Heterogeneity is a challenge – but it's feasible, and also enriching
- Very positive feedback & evaluation



Contents

Format

Ferienakademie: (cf. www.ferienakademie.de)

- Est. 1984, TUM + U's Erlangen & Stuttgart
- Summer school – but students work & present
- 2 weeks, in South Tyrol, during Oktoberfest …
- Funded by industry and by the 3 universities
- Application-based, for our best students
- ~ 12 courses, broad range of topics (even ethics)
- Leisure (hiking) plus intense work

Here (2015 – 2016 similar):

- "*Let's play! Simulated physics for interactive games*"
- 22 students (background: MA, CS, CSE, PH, ENG)
- 5 teams (models, simulation, I/O, interaction, parallel programming)
- Project mgt. by students (profs. are advisers)

Format

Technical details:

- Students were provided with template-code and a server for version management
- Server set up beforehand, i. e. installation of libraries, Git-server, documentation, …
- On-site installation of a LAN and WiFi for development
- Limited internet access: preparation in advance necessary
- Assignment of topics for the talks (incl. a supervisor) beforehand

Contents

SPPEXA Doctoral Retreat & Coding Weeks

SPPEXA:

- Priority Program "Software for Exascale Computing" of the German Research Foundation (DFG), see www.sppexa.de (nation-wide funding scheme)
- Installed in 2012, two 3-y funding phases (2013-2015, 2016-2018)
- 17 project consortia funded – each inter-institutional and interdisciplinary, some international ones due to a joint call with agencies from France and Japan

Education:

- Has been one core argument for SPPEXA's installation

Organized training for PhD candidates:

- Annual 1-to-2-week "doctoral retreat", 12-20 participants
- Practical components (hands-on sessions) plus tutorials
- Networking – foster exchange across the SPPEXA consortia
- Each doctoral candidate within SPPEXA must attend at least one retreat

So far: 2013 Darmstadt, 2014 Alps, 2015 Passau, 2016 Strasbourg

Contents

TUMuch Phun – Hardware Setup

Just one Host Node:

- 2x Intel Xeon E5-2697v3 - 2x14 Cores @ 2.3 – 3.6 GHz
- 128GB RAM
- Intel SSD DC S3710 – 1.2 TB Storage

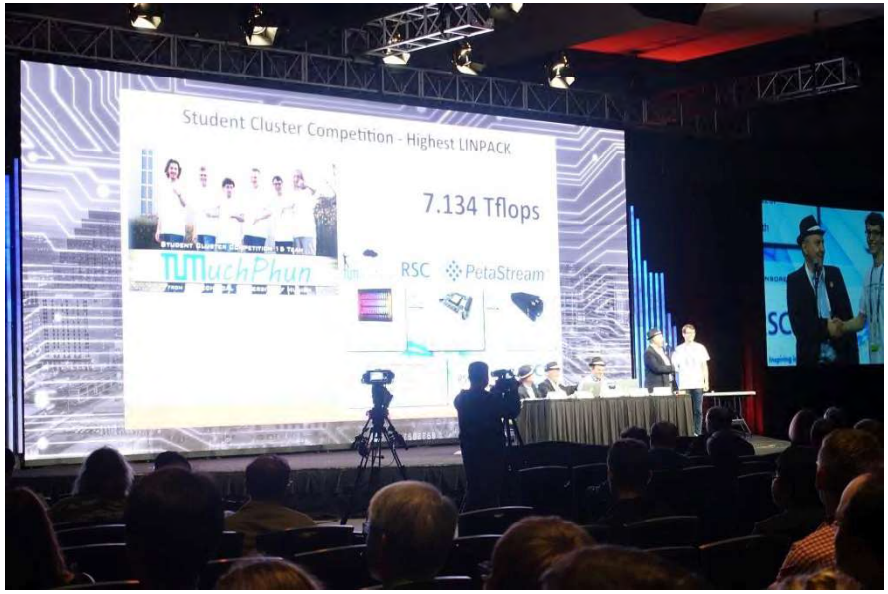
8x Intel Xeon Phi Coprocessor 7120A/P

- each with 61 Cores @ 1.238 GHz
- each with 16GB GDDR5 RAM

Proprietary Interconnect from RSC (basis: external PCIe 3.0 switch)

- Xeon Phis directly communicate with each other, no detour via host
- Mellanox SB7790 Infiniband EDR Switch
- 2x Mellanox Connect X4 Dual Port EDR Adapter (Ein IB Adapter for 4 Xeon Phis)
- Host communicates via two PCIe 3.0 16x ports with all Xeon Phis via switch

Finally: the Ceremony!



#1 in LINPACK performance

#3 in overall performance

☺

TEXAS ADVANCED COMPUTING CENTER

EDUCATION & WORKFORCE DEVELOPMENT

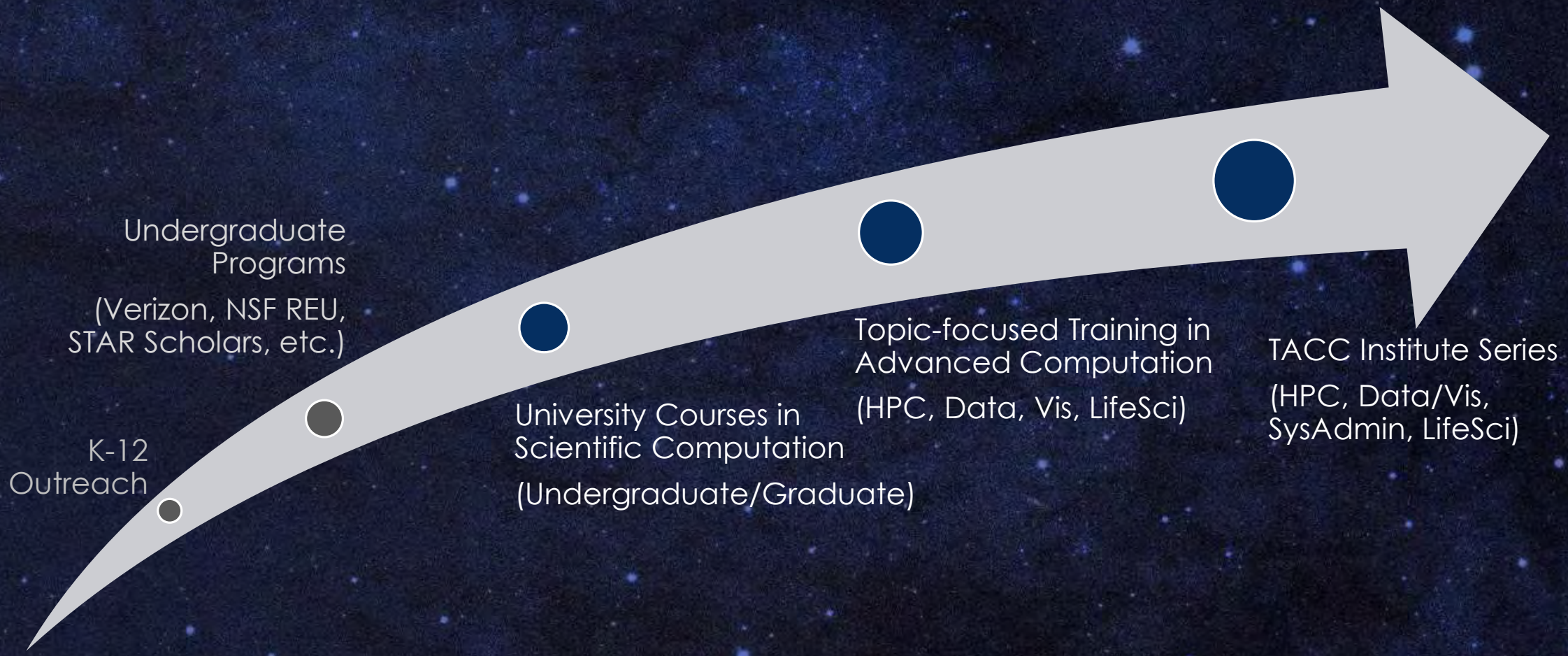
Melyssa Fratkin

Industrial Programs Director

IDC HPC User Forum

Sept 7, 2016

LIFELONG LEARNING IN ADVANCED COMPUTATION



UNDERGRADUATE PROGRAMS

- NSF Research Experiences for Undergraduates (REU) grants
  - 10 undergraduates majoring in science and engineering disciplines
  - Participants explore grand challenges such as climate modeling, weather forecasting, drug delivery, brain mapping, energy exploration and understanding the human genome, among others.
- STAR Scholars (Funded by TACC industry partners)

STAR SCHOLARS

- Graduate or undergraduate interns, depending on project
- Students work under a TACC mentor, on projects of interest to Partners
- Students submit papers or posters at conferences
- Partner funded ($50K per student for 2 semesters)
- Funding from BP, Shell, NASA

STAR SCHOLARS - OUTCOMES

- *Interactive Parallelization Tool* - paper by Ritu Aurora and student Madhav Gupta https://www.youtube.com/watch?v=L4a19kF6q48
  - IPT can be used by domain-experts and students to semi-automatically generate parallel programs based on multiple parallel programming paradigms (MPI, OpenMP, and CUDA) and learn about these paradigms through observation and comparison.
- Student paper accepted to SC16: "*A Tool for Semi-Automatic Application-Level Checkpointing*"
- Work on TACCStats – Todd Evans and student Vivek Pradhan
  - Student benchmarked a NoSQL database for analysis of all of the job-level data from our systems that is collected by TACCStats.
  - Working on visualization to facilitate detection of poorly written programs that are not using resources effectively

UNIVERSITY COURSES IN ADVANCED COMPUTATION

- UT Austin dept. of Statistics and Data Sciences
- Dual-listed undergraduate/graduate
- Five (5) courses covering many aspects of advanced computation
  - Introduction to Scientific Programming – programming concepts using C/C++ and Fortran
  - Scientific and Technical Computing – basic numerical methods, fundamentals of floating-point arithmetic, common tools for scientific software development (git, make), profiling and optimization, debugging
  - Parallel Computing – developing parallel applications using OpenMP and MPI
  - Visualization and Data Analysis – tools for data analysis (Hadoop, R) and visualization of scientific datasets (VisIt, Paraview)
  - Distributed and Grid Computing – executing data parallel ensembles which do not require MPI using grid-enabled tools (Globus, Condor, etc.)
- Content and instructor notes made publicly available through a grant from Chevron
  - https://www.tacc.utexas.edu/education/academic-courses

TOPIC-FOCUSED TRAINING IN ADVANCED COMPUTATION

- Offered in both spring and fall academic semesters
- 1 and 2-day courses in HPC, Data, and Visualization
  - Mix of lecture and hands-on exercises
- Open to TACC users, academic and industry partners
- In many cases courses are live-streamed and recorded on YouTube
  - http://bit.ly/TACCtraining
- Upcoming and past training at: https://portal.tacc.utexas.edu/training

Courses have included:

- ✓ MPI/OpenMP
- ✓ Profiling and Optimization
- ✓ Debugging
- ✓ Programming the Intel Xeon Phi (KNC/KNL)
- ✓ Python
- ✓ R
- ✓ Using Hadoop/Spark
- ✓ Data Management Practices and Techniques
- ✓ Introduction to Scientific Visualization

…and more

SUMMER SUPERCOMPUTING INSTITUTE (SSI)

- Held each summer since 2007
- Originally an aggregation of TACC's topic-focused training courses
  - MPI, OpenMP, profiling, debugging, visualization
- Offers one-on-one meetings with computational staff to discuss research-specific problems
- From 2007-2015, enrollment steady around 40
- Registration fee for attendees

SUMMER SUPERCOMPUTING INSTITUTE (SSI)

- In 2016, SSI was expanded
- Two tracks
  - Parallel Applications
    - MPI, OpenMP, debugging, profiling, many-core programming (Xeon Phi)
  - Data Analysis and Visualization
    - Python, R, Data Management, Hadoop/Spark, VisIt/Paraview, High Throughput Computing
- Plenary talks showcasing newest research
- 90 participants
  - 5 countries represented
  - 18% female participation

COMING SOON – TACC INSTITUTE SERIES

- Series of week long training events modeled on SSI
  - High Performance Computing
  - Visualization and Data Analysis
  - Cluster design and administration
  - Life Sciences Computing
- Registration fee TBD
- To be offered during the summer months
- More details coming at SC16 in November!

MEASURING IMPACT

- More than 1,000 registered for training between Aug. 2015 and July 2016
- We also track number of independent views of our recorded training events
  - More than 11,000 views between Aug. 2015 and July 2016
    - Up 25% over previous year
- In-person participants are asked to answer surveys after training events
  - Satisfaction with content: > 70% "very satisfied"
  - Satisfaction with instructors: > 90% "very satisfied"
  - Would recommend to others: > 70% "definitely"

CONTACT US

- Melyssa Fratkin, Industrial Programs Director, mfratkin@tacc.utexas.edu
- Lucas Wilson, Director of Training & Professional Development, lwilson@tacc.utexas.edu
- https://portal.tacc.utexas.edu/training

Preparing Students for HPC Careers in the Energy Sector

Jan E. Odegard

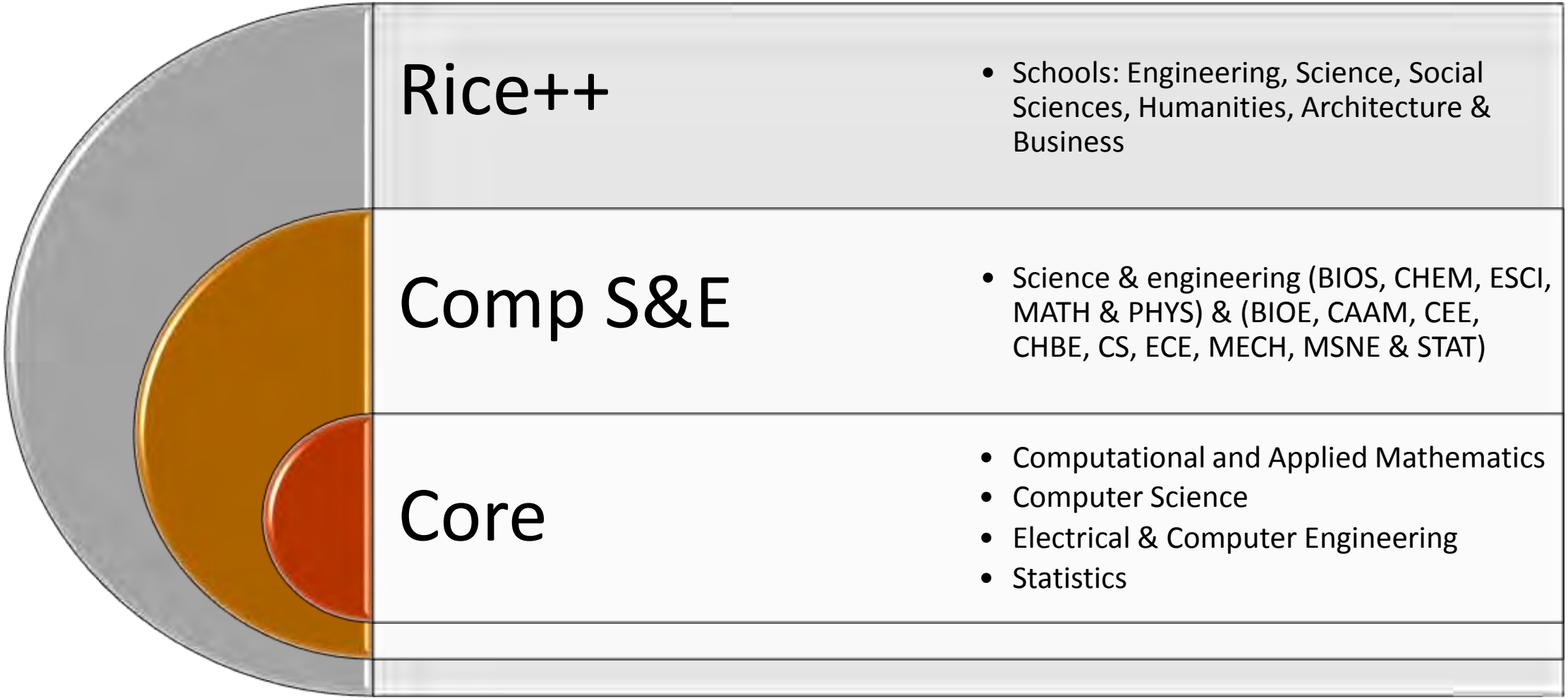
Executive Director, Ken Kennedy Institute for Information Technology

Associate Vice President, Research Computing & Cyberinfrastructure

odegard@rice.edu / 713.348.3128

Ken Kennedy Institute Community

IT research & education enabler



Rice: Computing as a Universal Enabler

Data analytics, data mining, and machine learning …

Numerical methods, solvers, modeling, and simulation …

Programming models, languages, compilers, and tool

Tool makers

Fundamental R&D in: computer science, electrical & computer engineering, applied mathematics, statistics, signal processing, imaging, modeling, machine learning, data-science, cyber-security, energy, materials, biomedicine, …

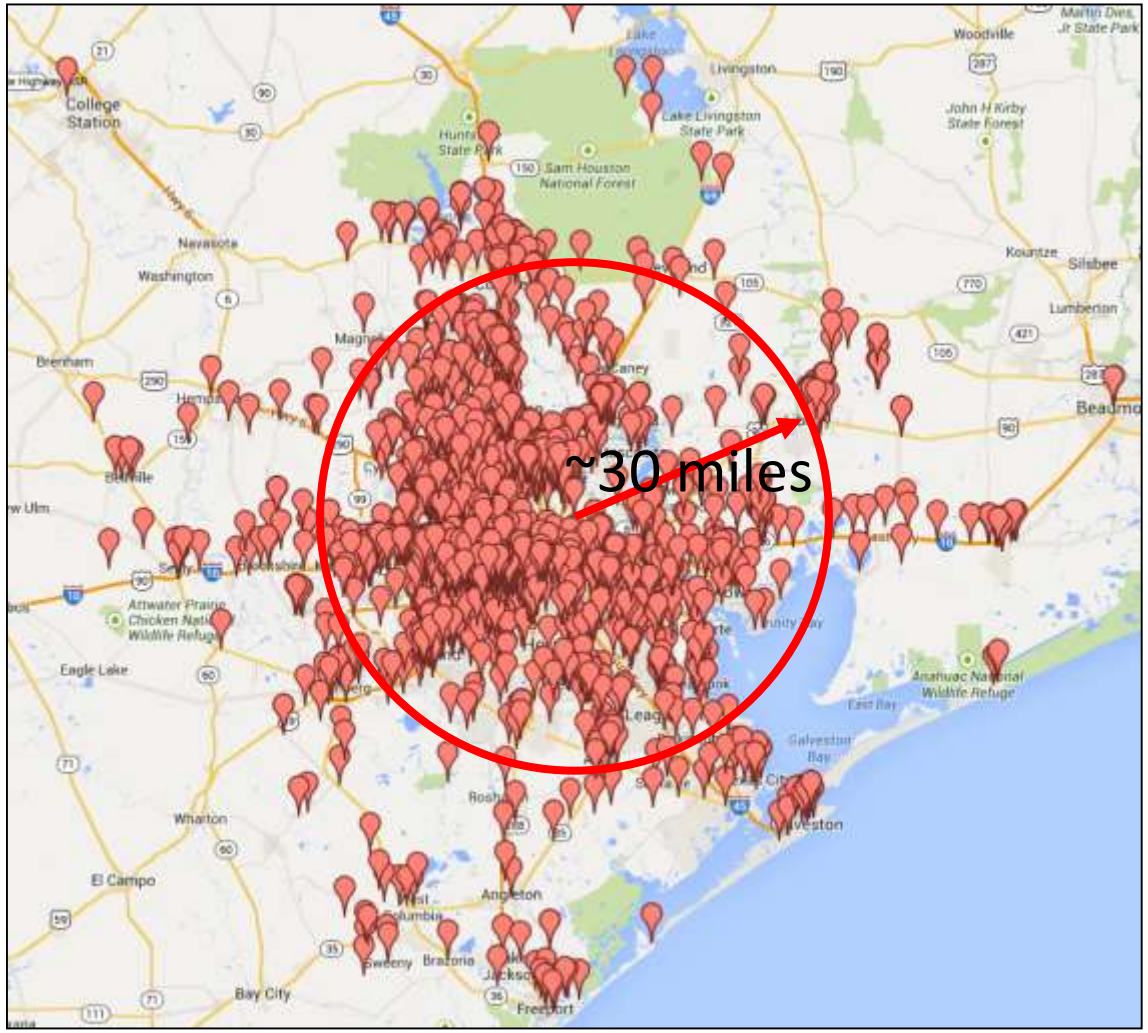
Cyber-Infrastructure to enable and accelerate research

Research Computing: Rice's profile in HPC

"The [global] computational science community views Rice [University] as a tier-1 HPC site"

2014 IDC Study: Best Practices in Securing Funding for University-based HPC Centers

Energy Capital of the World (data from 2014)



Houston

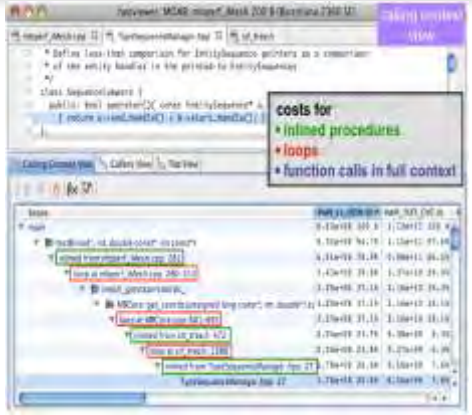
- 3,600+ oil & gas companies around Houston
- 50% of Houston's employment in energy sector
  - 10% in oil & gas
- 1 in 5 jobs created since 2010 in oil & gas
- 29% of total U.S. oil & gas employment

Texas

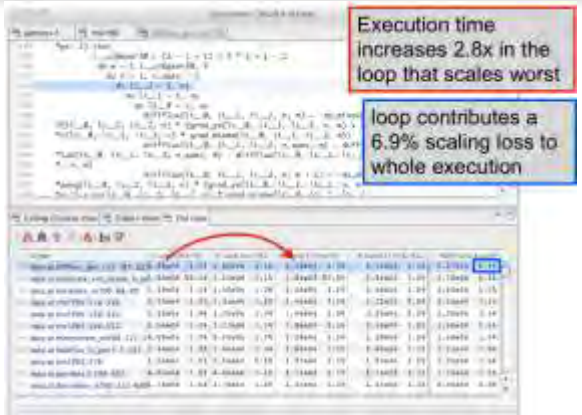
- 36% of the oil produced in U.S. comes from Texas
- 23% of crude oil produced in the U.S. comes from GOM offshore production
- 389,000 Texans are directly employed by oil & gas
- 1.8 million additional jobs are supported by economic gains from oil & gas industry spending
- **Economic Impact on Texas: $308B**

Performance Tools: HPCToolkit

John Mellor-Crummey, CS

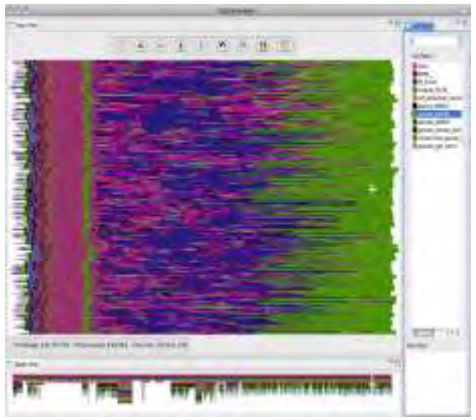


Attribute Costs to Code

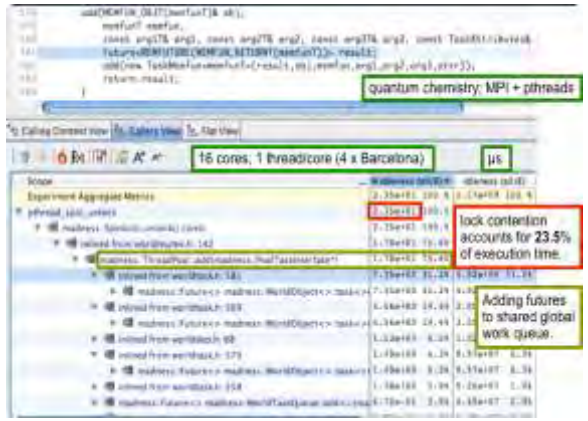


Pinpoint & Quantify Scaling Bottlenecks

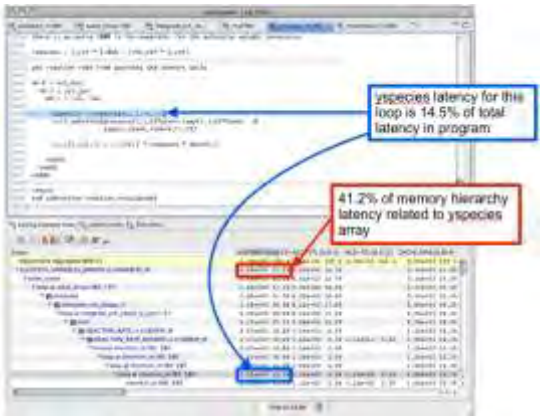
Assess Imbalance and Variability



Analyze Behavior over Time



Shift Blame from Symptoms to Causes



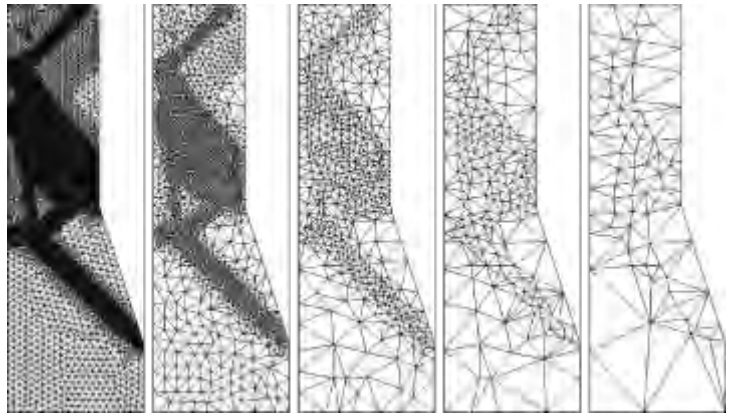
Associate Costs with Data

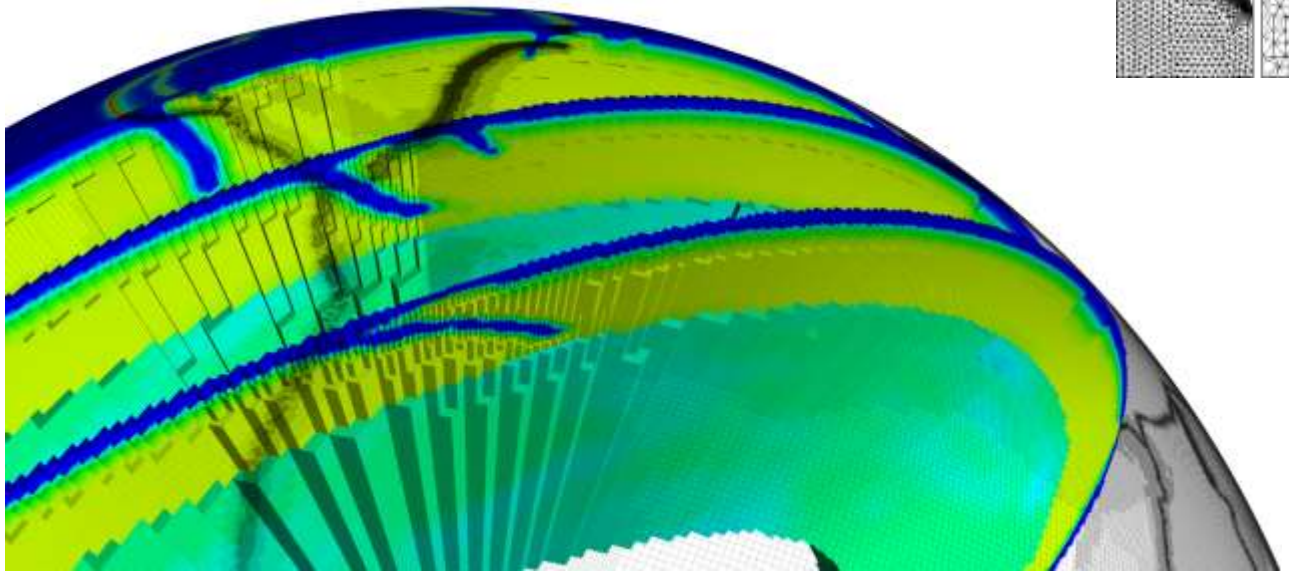
Performance tools can play an important role in guiding development to maximally leverage infrastructure

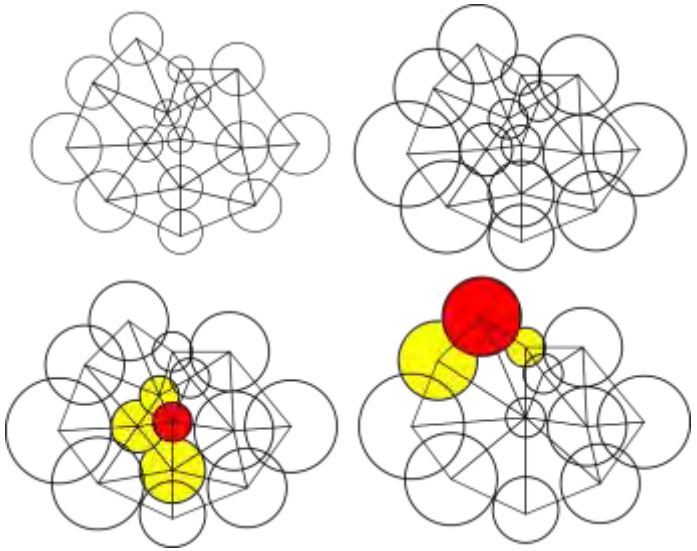
Scalable Solvers & Numerical Libraries

Matthew Knepley, CAAM

- Unstructured grids
- Node-aware optimization
- Intel Parallel Computing Center
  - ✓Accelerating and expanding PETSc

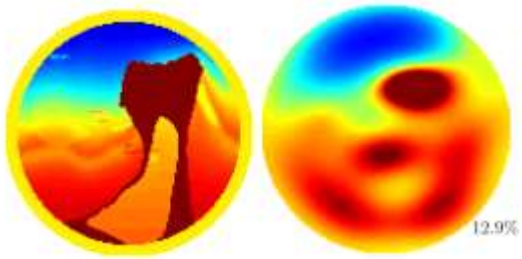




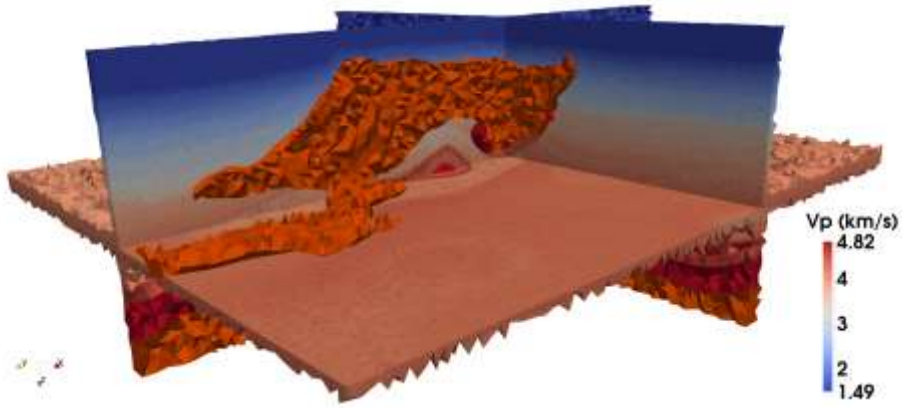


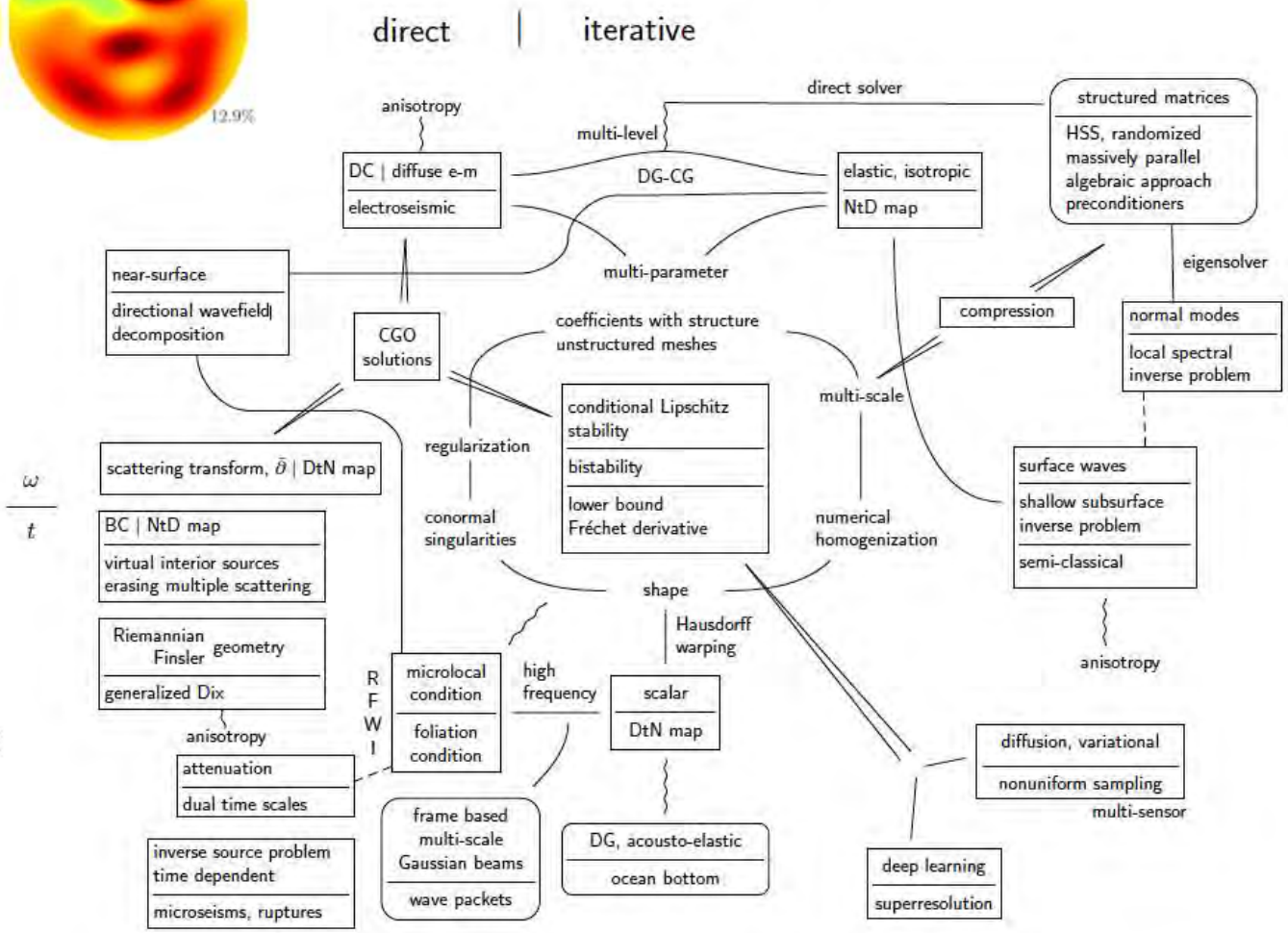
Geo-Mathematical Imaging

- Direct nonlinear methods
- Iterative methods (FWI)
- Attenuation
- Spectral methods
- Micro-seismicity
- Deep learning



Maarten de Hoop, CAAM

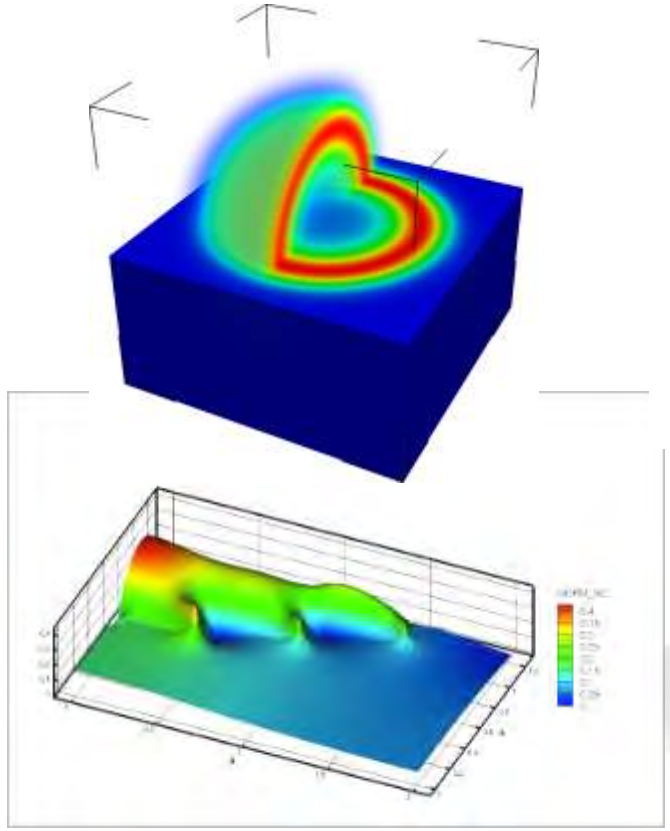


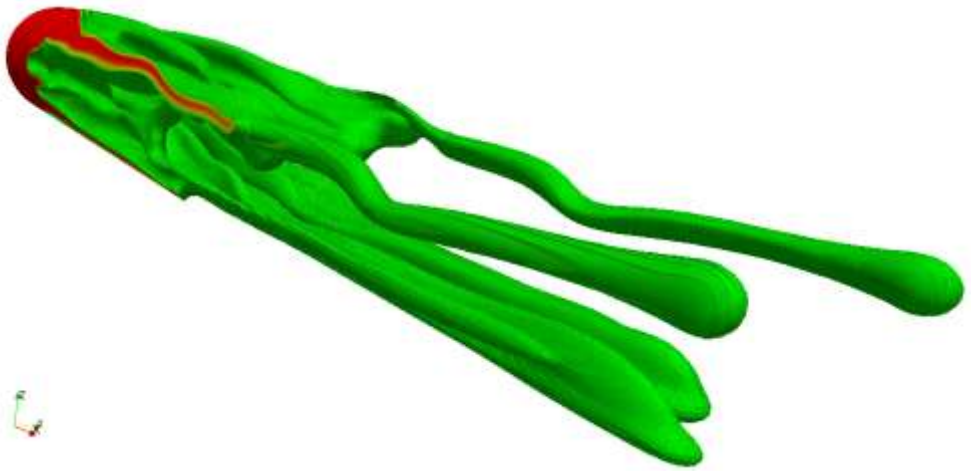


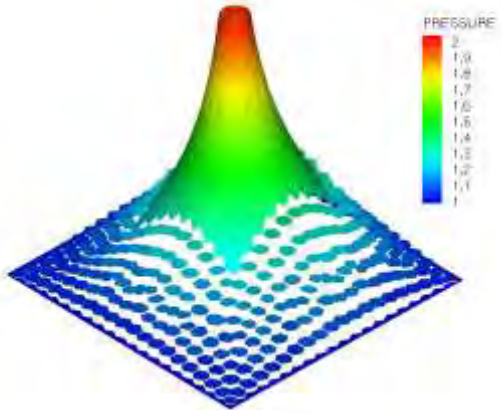
Coupled Flow & Reservoir Simulation

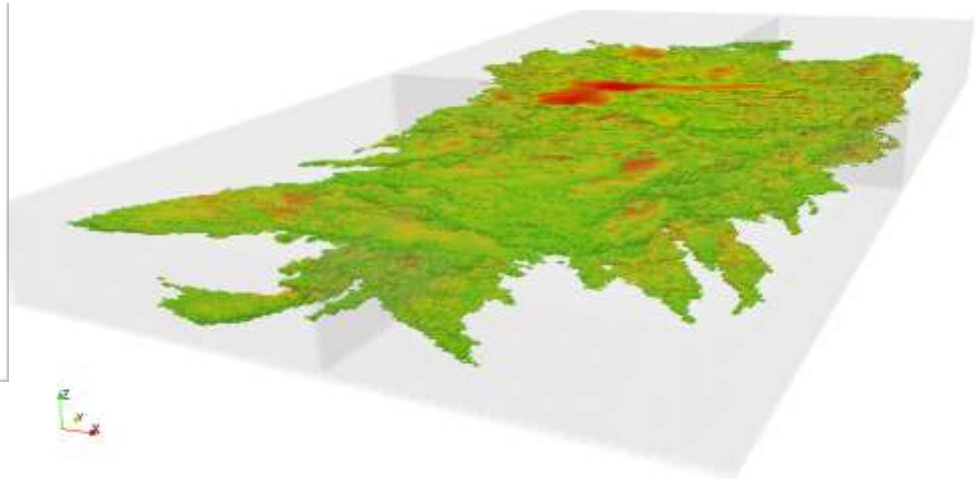
Beatrice Riviere, CAAM

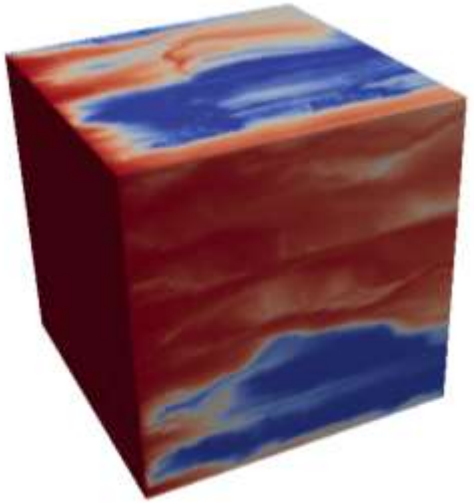
Modeling and Simulation of Processes in Porous Med and Pore Scales







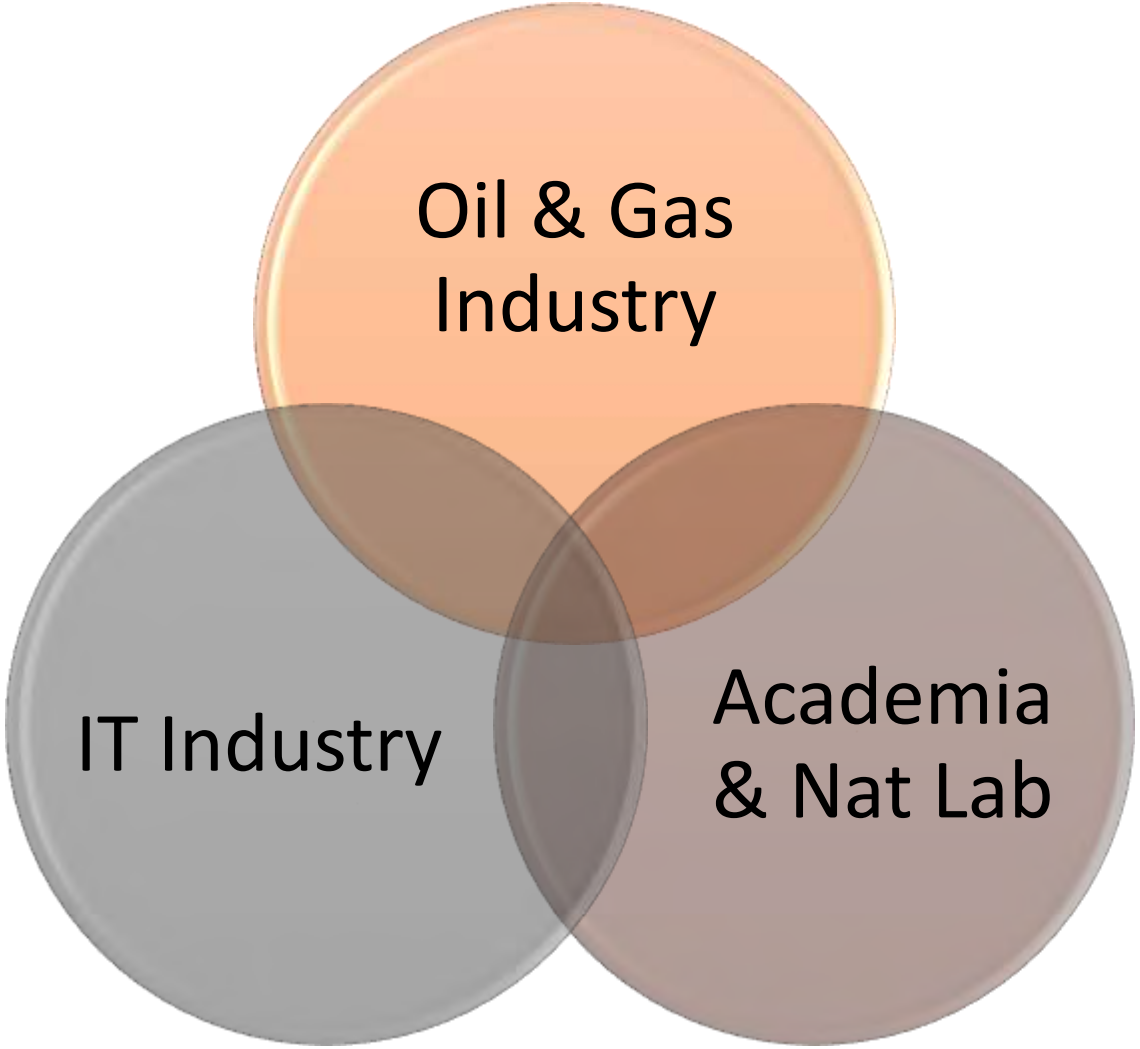




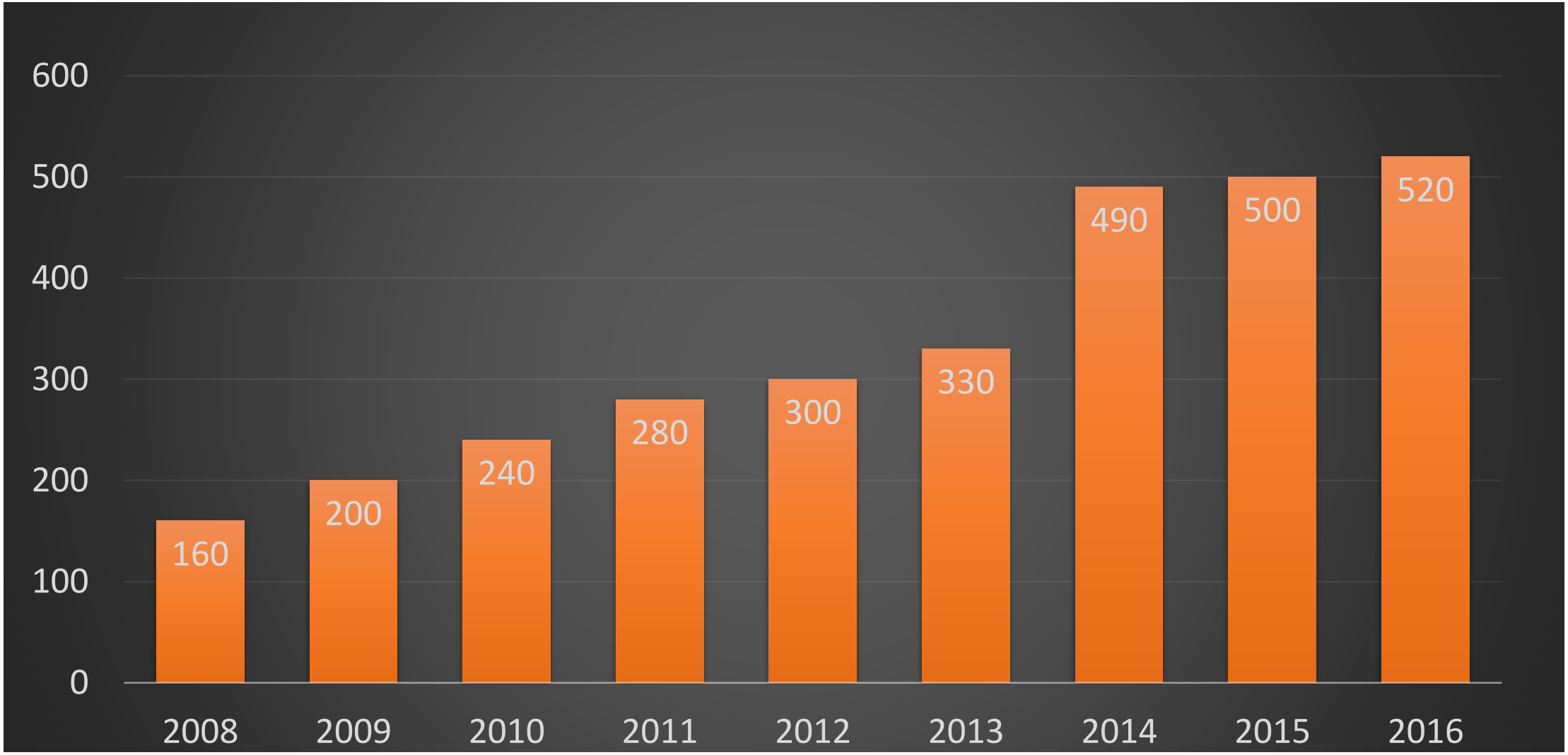
Private-Public Partnership

Annual Rice Oil & Gas HPC Conference

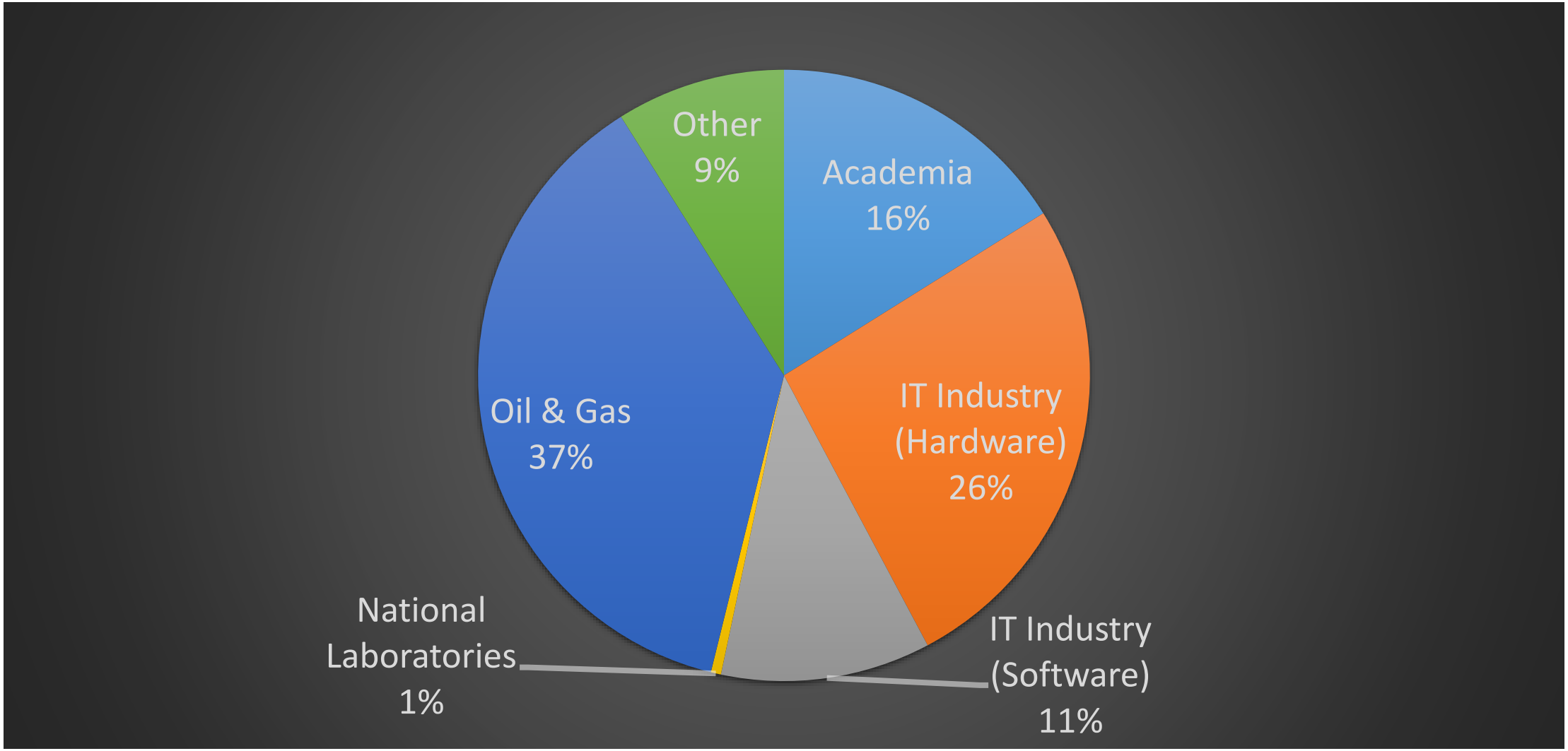
A private-public partnership to support cyberinfrastructure operation and development challenges, develop a sustainable and diverse high performance computing, computational science & engineering, and data-science workforce directed at industry needs



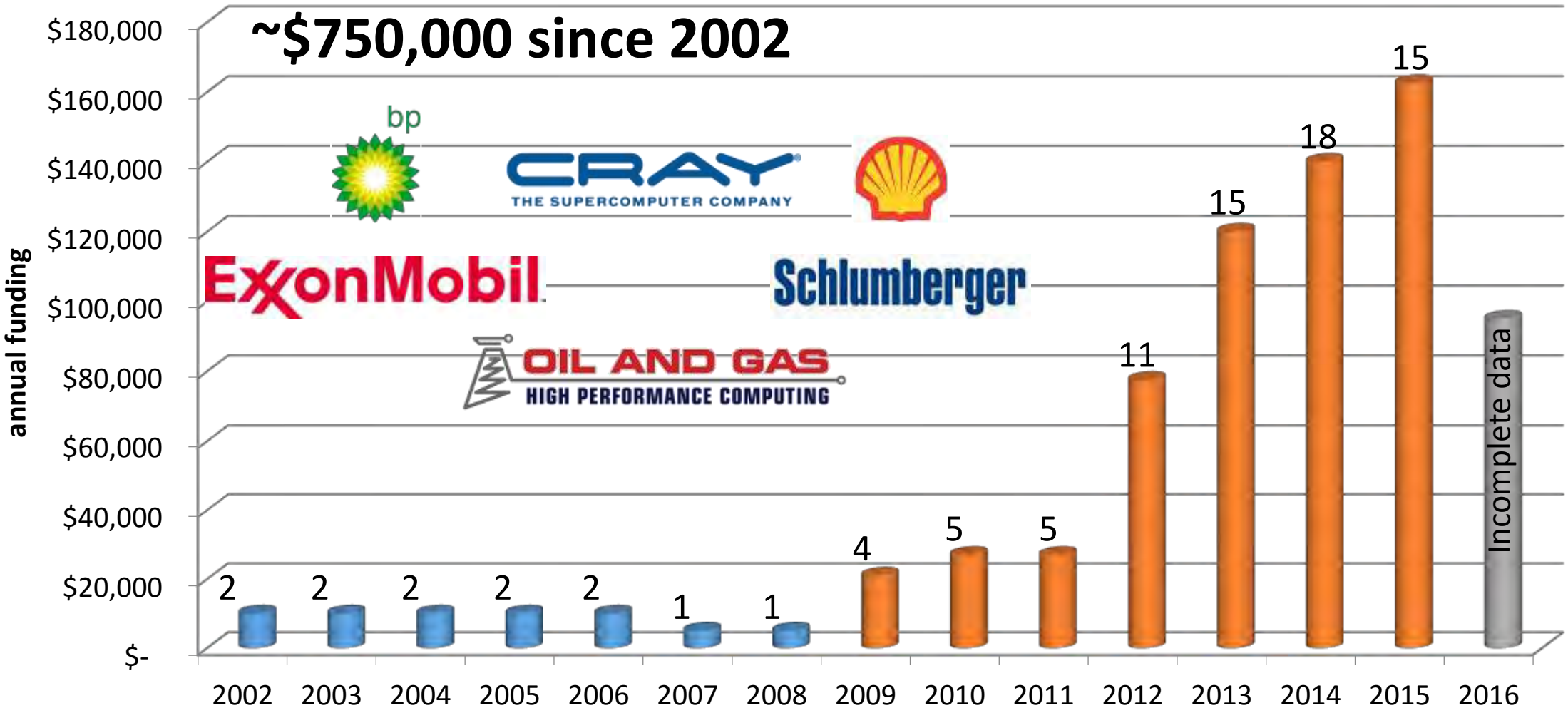
Conference Attendee History



Attendance by Industry Segment



Graduate Student Support to Create Awareness



Where/how are fellowships allocated?

- In computational departments
  - Computational and Applied Mathematics
  - Computer Science
  - Earth Science
  - Electrical and Computer Engineering
  - Statistics
  - Applied Physics
- Used to support students already in these programs at Rice
- Used to recruit students to computational programs at Rice
  - Support MSc & PhD students
- Industry partnership also led to the development of the professional masters degree in Computational Science and Engineering
  - 30 credit hours of course work, can be part time

Boot Camps: Industry Summer Training

High-Performance Computing **Data-Science**

Boot camps offered by Ken Kennedy Institute at Rice with leading faculty and researchers as instructors.
Intensive lecture/lab training to quickly acquire basic skills for beginners or intermediate users (can be mid career boost vehicle).

- Intro to thread-based parallelism
- Intro to MPI
- Performance analysis tools
- GPU Accelerated computing
- Intro to parallel I/O

- Intro to R & Python
- Intro to AWS, Hadoop and Spark
- Intro Unsupervised Learning
- Intro Supervised Learning

SAVE THE DATE:

March 15-16, 2017
10th Anniversary

2017 Rice Oil & Gas HPC Conference

http://rice2017.og-hpc.org

Abstracts due by November 23, 2016.

Subscribe to mailing list to stay informed!