2 This is a new talk, it maybe 30 minutes or 30 hours ...
What is a Big Ideas Class?
What is a Big Ideas Class?

Lectures and discussions of current research and technical developments in computer science for beginning graduate research students. Topics will emphasize open problems and recent scientific advances. Content may vary to reflect research advances in areas such as data analytics, scientific computing, graphics and visualization.
What is a Big Ideas Class?

Lectures and discussions of current research and technical developments in computer science for beginning graduate research students. Topics will emphasize open problems and recent scientific advances. Content may vary to reflect research advances in areas such as data analytics, scientific computing, graphics and visualization.
What is a Big Ideas Class?

Lectures and **discussions** of current **research** and technical developments in computer science for beginning graduate research students. Topics will emphasize **open problems** and recent scientific advances. Content may vary to reflect research advances in areas such as data analytics, scientific computing, graphics and visualization.

- Who are the NIU CS faculty with active research projects?
- What are NIU CS faculty interested?
- Where do I get more information?
Bit About Me

- Education
  - Northern Illinois University - Physics (BS)
  - University of Illinois @ Chicago - Computer Science (MS)
  - University of Chicago - Computer Science (MS, PhD)

- Experience
  - Fermi National Accelerator Laboratory
  - Argonne National Laboratory
  - Northern Illinois University
Bit About Me - Research *(Areas/Interests)*

- Advanced Display Environments
- Collaboration Technology
- High Performance Computing (Environments)
- Information Visualization
- Scientific Visualization and Analysis
- Virtual Reality (Augmented)
Scientific Visualization and Analysis

- vl3: volume rendering library
  - Parallel volume rendering library that exploits GPU hardware
  - Uses native data formats
- Integration with virtual and augmented reality
- Usability and collaboration
- Domain specific visualizations
vl3: volume rendering library
Parallel Distributed, GPU-Accelerated, Advanced Lighting Calculations For Large-Scale Volume Visualization.

Virtual Reality\textsuperscript{ab}


Virtual Reality

Usability and Collaboration

Usability and Collaboration

Domain Specific Visualizations

- Applied solutions to specific problems within domain
- Deep partnership with domain experts
- Current effort with NIU Chemistry - *Visualizing and Quantifying Structural Ordering Underlying Static Structure Factor Peaks from Molecular Dynamics Simulations* Travis Mackoy, Bharat Kale, Ralph Wheeler
Domain Specific Visualizations

Domain Specific Visualizations

Domain Specific Visualizations
Domain Specific Visualizations
High Performance Computing

- Applications
- Communication
- Operations
- Power


---

22 CS-600 - Big Ideas - Fall 2018
Evolving (scheduling constraints)

Evolving (complex workflows)

50% of the ATLAS papers based on 2015 data use the HPC-produced computing in a demonstrable manner.

These would still be written without the US HPC effort, but they probably would not exist today: the time-to-science has been dramatically shortened.
Evolving (increased engagement)

HPC Environments

*Usability and enabling*, that how do we enable scientists (users) to be the most *productive* from *start to finish*?
Workflow of Today

Simulation MIRA
Analysis
Analysis
Simulation MIRA
Simulation MIRA

Time

Storage
Workflow of Tomorrow (Today)
Facility of Tomorrow

Simulation

Analysis

User Community via Web and Experimental / Observational-Specific Clients

Jupyter Notebooks

Globus Online

PDACS

Machine Learning

Smaash

DESCQA

job submission/adaptation layer

Facility Storage

Experimental / Observation Datasets

Cosmology

APS

SNS

LHC

Climate

Brain Imaging

ALCF-hosted

Experimental / Observational-Specific Resources

(servers like JupyterHub, databases, web, indexed search, visualization)
Observations (Science Management)

- Data-intensive science (simulations and experiments) requires **capture, curation** and **analysis**
- Data comes from many sources, in many formats and multiple sizes
Observations *(Science Management)*

- Problem with science management:
  - Tracking simulations and output **[difficult]**
  - Finding and reproducing old simulations: **[difficult]**
  - Monitoring live simulations: **[inconvenient, idiosyncratic]**
  - Post-processing, analysis and archival of results: **[haphazard]**
  - Assessing simulation behavior/performance: **[difficult]**
Increased Access to Scientific Communities

Support for Application Teams

Simulation management and analysis system for Flash (Smaash)

- Tracking and coordination of data (simulation and meta)
- Run-time monitoring of simulations and automated analysis of simulation output
- Method for managing / executing common workflows

Prototype Partner - Flash

- Multi-physics
- Adaptive-mesh
Prototype Partner - Flash

- Meta-data output
  - .log: simulation progress, warnings, errors, resource use
  - .dat: integrated grid quantities
- Scientific data output (HDF5)
  - Checkpoint: complete information needed to restart simulation
  - Plotfile: data values of interest for analysis
  - Particle files: tracer particles of interest during analysis
Smaash Components

- Database (manages meta-data)
- Back end services (co-located with compute resources and scientific data)
- Front end interfaces (user facing)
Smaash Implementation

Python

MySQL

Matplotlib

Django

Dojo

Python-based web framework

2D, Python plotting library

JavaScript toolkit

Django app; bridge to Dojo

Dojango
Smaash Back End Services

- Collector - captures and stores meta-data in database about simulation
- Archiver - automates the archiving of data
- Verifier - cross checks output and database entries
- Associator - connects a current simulation with campaign
- Observer - responsible for updates to user (email)
- Visualizer - automatic running of user specified visualization scripts
Smaash Front End Interfaces (Views)

- **Tree** - collection of campaigns, simulations and runs
- **Graph** - quick graphs of results
- **Monitor** - automated visualizations
- **Summary** - details and notes
### Tree View

![Tree View Image]

**Filter by Date**
- Before: 2/24/2011
- After: 5/1/2010

**Filter by Tag**
- Flame Speed Study
- Flame Bubble
- RT Flame
- Resolution Study
- WD_def

**Filter by Site**
- ellipse.uchicago.edu
- franklin.nersc.gov
- antrepid.acl.fanl.gov

**Filter by Owner**
- Cal Jordan
- Carlo Graziani
- Chad Glendenin
- Chris Daley
- Dean Townsley
- Eva Wuyts

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Tags</th>
<th>Description</th>
<th>Dim</th>
<th>Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlameSpeed [35]</td>
<td>2010-11-16</td>
<td></td>
<td>/hudson/pe...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>flameBubble [34]</td>
<td>2010-08-12</td>
<td></td>
<td>/hudson/pe...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1km_s185_q3E9_r32</td>
<td>2010-08-15</td>
<td>Flame Bubble</td>
<td>Flame bubble resolution study...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>rundir_0001 [683]</td>
<td>2010-08-15</td>
<td>Flame Bubble</td>
<td>Antrepid-fs0/users/norris/pe...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>rundir_0002 [685]</td>
<td>2010-08-13</td>
<td>Flame Bubble</td>
<td>Antrepid-fs0/users/norris/pe...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>rundir_0003 [683]</td>
<td>2010-08-20</td>
<td>Flame Bubble</td>
<td>Antrepid-fs0/users/norris/pe...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>rundir_0004 [689]</td>
<td>2010-08-22</td>
<td>Flame Bubble</td>
<td>Antrepid-fs0/users/hudson/per...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>2km_s185_q3E9_r32</td>
<td>2010-08-12</td>
<td>Flame Bubble</td>
<td>Antrepid-fs0/users/hudson/per...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>4km_s185_q3E9_r32</td>
<td>2010-08-12</td>
<td>Flame Bubble</td>
<td>Antrepid-fs0/users/hudson/per...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>8km_s185_q3E9_r32</td>
<td>2010-08-12</td>
<td>Flame Bubble</td>
<td>Flame bubble resolution study...</td>
<td>16x16x16</td>
<td></td>
</tr>
<tr>
<td>16km_s185_q3E9_r32</td>
<td>2010-08-12</td>
<td>Flame Bubble</td>
<td>Flame bubble resolution study...</td>
<td>16x16x16</td>
<td></td>
</tr>
</tbody>
</table>
Graph View

http://flashdb.ci.uchicago.edu/graphBranches/410,425/using/v90/vs/v32/cstroke/png
Monitor View
FlameBubble problem on 2048 processors

<table>
<thead>
<tr>
<th>Parent: rundir_0001_06831</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details</td>
</tr>
</tbody>
</table>

```
/intrepid-fs0/users/jnorris/persistent/2010/ResolutionStudy/lkm_sl
```

**System Info**
- Linux login5 2.6.16.60-0.42.8-ppc64 #1 SMP Tue Dec 15 17:28:00 UTC 2009 pp

**Setup Syntax**
- /intrepid-fs0/users/gjordan/persistent/2010/flameBubble/src/20100610/trunk/bin/s
  - maxblocks=40

**FORTRAN Compiler Flags**
- mpi90,lbm -g -O4 -cintsize=4 -qrealsize=8 -qfixed -qosave -c -qsuffix=cpp=F -qsuffix=f=F90 -cpp=F90 -qfree=f90 -WF -D MAXBLOCKS=40 -WF -DNXB=16 -W

**C Compiler Flags**
- mpiocc.lbm -I/include -Lsoft/apps/hdf5-1.6.6/include -DNOUNDERSCORE -Vbgsy -arch=450 -qnum=auto -ocache=auto -qmaxmem=16384 -D_FILE_OFFSET_BITS
  - DN_DIM=3 -D HAVE_MALLOCINFO

**Max Number of Blocks/Proc**
- 40

**Max Number of Particles/Proc**
- 1000
Smaash Outcomes (Simulation State)
Smaash Outcomes (Analysis)
Smaash Outcome (Notebook)
Last Topic

Information Visualization

- Connection to X science
Thank You

Most of my funding is provided by the Argonne Leadership Computing Facility, a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357 with additional support from the National Science Foundation.

Thanks to all the staff of ALCF, colleagues at NIU and ANL, and the students of the ddiLab.
If I have seen further it is by standing on the shoulders of giants.

— Sir Isaac Newton
Extra Slides
HPC Landscape (Yesterday)

Simulation Applications

- 64bit floating point
- memory bandwidth
- random access to memory
- sparse matrices
- distributed memory jobs
- synchronous input/output multinode
- scalability limited communication
- low latency high bandwidth
- large coherency domains (sometimes)
- output typically greater than input
- output rarely read
- output is data
## HPC Landscape (Today)

<table>
<thead>
<tr>
<th>Simulation Applications</th>
<th>Big Data Applications</th>
<th>Deep Learning Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>64bit floating point</td>
<td>64bit and integer important</td>
<td>lower precision &lt;= 32bit</td>
</tr>
<tr>
<td>memory bandwidth</td>
<td>data analysis pipelines</td>
<td>inferencing can be 8bit (TPU)</td>
</tr>
<tr>
<td>random access to memory</td>
<td>databases including NoSQL</td>
<td>scaled integer possible</td>
</tr>
<tr>
<td>sparse matrices</td>
<td>MapReduce/SPARK</td>
<td>training dominates development</td>
</tr>
<tr>
<td>distributed memory jobs</td>
<td>millions of jobs</td>
<td>inference dominates pro</td>
</tr>
<tr>
<td>synchronous input/output multinode</td>
<td>input/output bandwidth limited</td>
<td>reuse of training data</td>
</tr>
<tr>
<td>scalability limited communication</td>
<td>data management limited</td>
<td>data pipelines needed</td>
</tr>
<tr>
<td>low latency high bandwidth</td>
<td>many task parallelism</td>
<td>dense float point typical SGEMM small DFT, CNN</td>
</tr>
<tr>
<td>large coherency domains (sometimes)</td>
<td>large-data in and large-data out</td>
<td>ensembles and search</td>
</tr>
<tr>
<td>output typically greater than input</td>
<td>input and output both important</td>
<td>single models small</td>
</tr>
<tr>
<td>output rarely read</td>
<td>output is read and used</td>
<td>input more important than output</td>
</tr>
<tr>
<td>output is data</td>
<td>output is data</td>
<td>output is models</td>
</tr>
</tbody>
</table>