



getexample: Reducing Barriers to Entry on Shared HPC Resources

Dr. Dirk Colbry

Director of HPC Studies

Department of Computational Mathematics Science and
Engineering

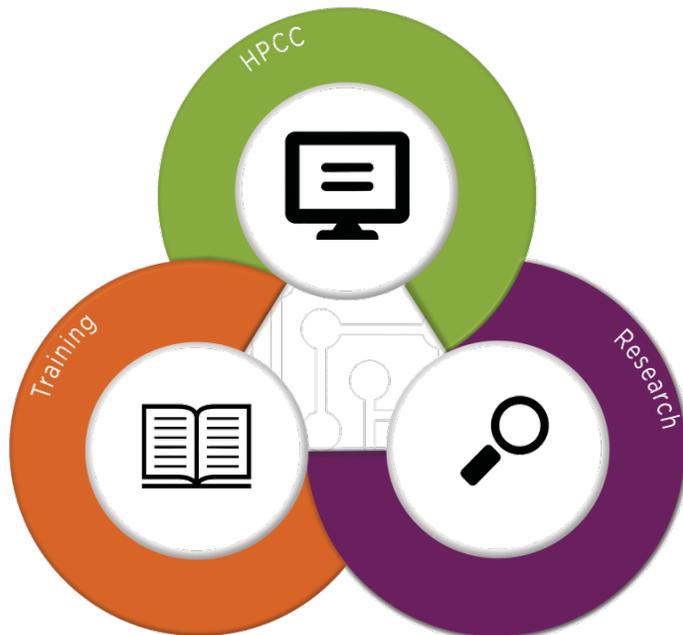
Michigan State University



Institute for Cyber-Enabled Research



- Computing Resources
- User Training
- Research Support



Vice President of Research and Graduate Studies



Department of Computational Mathematics Science and Engineering

Est. 2015

- Jointly operated by Colleges of Natural Science and Engineering
- Composed of 25-30 FTEs, including some current MSU faculty and a larger number of **new hires**.
- Most faculty will have joint appointments across campus.
- Faculty focus on data science and large-scale and high-performance computation
- Faculty are incentivized to engage in cross-discipline and cross-college research collaborations





Department of Computational Mathematics Science and Engineering

Computational science addresses the construction of mathematical models and quantitative analysis techniques and using computers to analyze and solve scientific problems.

Foundation:

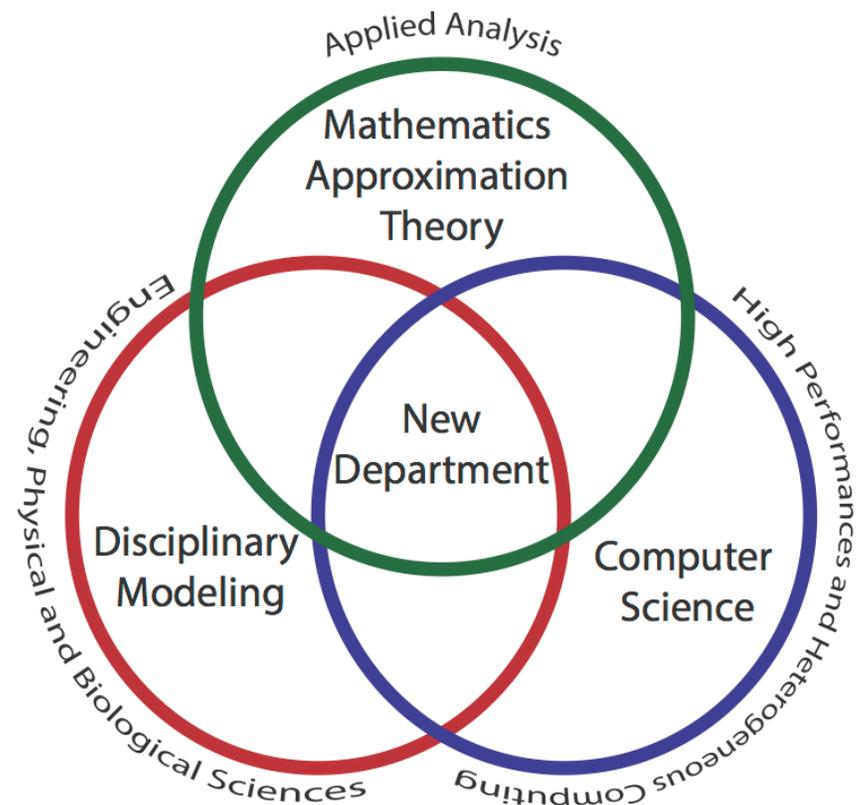
Discipline leverages:

1. **Application knowledge**
2. **Computer science**
3. **Mathematics**

To develop new methods for investigating complex problems through computation

Impact:

Analysis of complex experiments
virtual prototyping
virtual laboratory
etc...

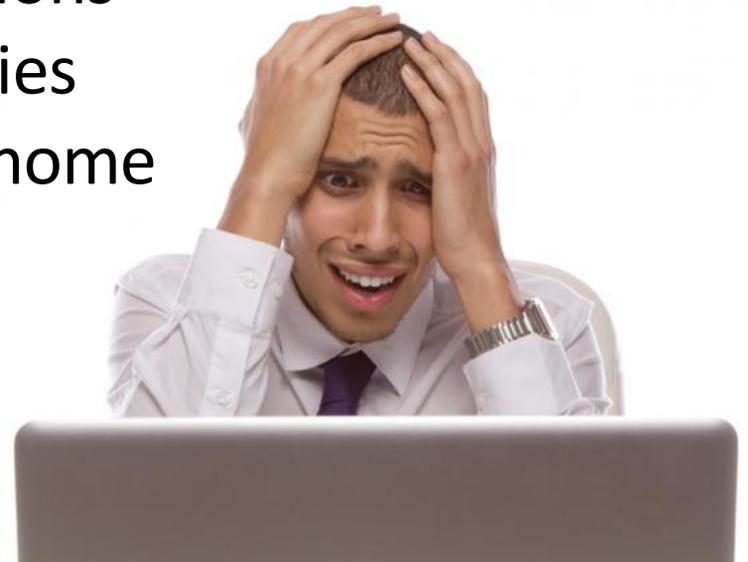


Examples Can...

- Enable independent learning
- Enable discovery of resources and capabilities
- Provide scaffolding to learning
- Reduce the “mean time to science”

Why are universal examples impossible?

- Different schedulers
- Different operating systems
- Different compiler commands
- Different recommended compiler flags
- Different module systems and paths
- Different module naming conventions
- Different versions of default libraries
- Different personal scratch, temp, home directories
- Different scheduling policies
- ...



The Trouble with Web Examples:

- Some common "novice" issues include:
 - Not knowing how to convert a bash script to an executable using `chmod`.
 - Fonts on websites may create errors when trying to copy-and-paste into a command script (such as incorrect conversion of special characters like as quotation marks or whitespace)
 - Users may not be familiar with text editors and or network copy programs such as `git`, `wget`, `scp`, etc.

getexample Design Goals

Tool to provide immediate access to a range of user examples

- It must be easy to use, even for novices
- Minimum number of steps to get working
- It must provide working examples, which can then be modified and updated by the user as appropriate for their own work
- Each example should include complete documentation in order to allow users to understand, use and modify the code for their own use

Exercise: getexample



- Run the “getexample” powertool
 `>getexample`
- Download the helloMPI example
 `>getexample helloworld`

Running *getexample*: no inputs

- Typing *getexample* by itself results provides a help message and a list of available examples

```
>>getexample
```

```
Download an HPC example:
```

```
usage:
```

```
  getexample <examplename>
```

Where <examplename> is the name of the example you want to download. This will create a directory named examplename which you can cd into and hopefully read the README file (if one is available) or just submit the *.qsub file.

For Example:

```
  getexample helloworld
```

Possible example names:

getexample: some examples

```
abacus_example      fortran_openmp_blas  MKL_Example        stata_parallel
DDT_examples        MATLAB_parameter_sweepSAS_example      condor_basic
MATLAB_2014_parfor2 Python_scraping     build_velvet        intro_workshop
OpenMP_profiling    basic_array_job     helloMPI            MPI_pi
ADMB_example        FreeSurfer          MKL_FFTW           tbb_example
econ_examples       MATLAB_parfor        shell_scripting     condor_Python
MATLAB_basic        qsub_arraydepend    burn_heat          job_dependencies
paraview_basic      blast               helloOpenMP         multi_variable
ADMB_example2       GAMESS_example      MKL_mic            TotalView_MPI_example
espresso_benchmark MATLAB_patternsearch simpleMatlab        condor_R
MATLAB_blcr         R_example           Circuitscape_examples lammps_test
pbdR_examples       blender_farm        helloworld         myhadoop
affinity            gmp_mpfir          MKL_parallel       Trinity_Assembly
fftw                MIC_examples        SNPanalysis-model  condor_simple
MATLAB_compiler     OpenCL_hello_world Clang_example       LAPACK_example
pcap_example        BOOST_example       HFSS_example        NAMD_CUDA_example
allinea_map         gromacs             Molpro_example     Valgrind_example
fluent3D            MIC_programming     SNPPipeline         cuda
MATLAB_compiler2    RNAseq-model        cloudy_example      magma_example
PETSc_example       boostUnitTests      ImageJ              NAMD_example
Amber_CUDA_example  Hadoop_wordcount   mothur              VASP_example
fluentMPI           MKL_benchmark       splitBam           cuda_clock
MATLAB_many_jobs    RNAseq-semimodel   CMakePackageExamples makefile_example
Python_MPI          bowtie              intro2hpc           Octave_basic
Amber_example       HEEDS_test          mothur2            velvet_blcr
fortran_openmp      MKL_c_eigenvalues   STATA_example      condor_advanced
MATLAB_movie        R_plot              introToPython
Python_numpy        brother_test        mothur_example
avida_blcr          helloHPCC
```

Running *getexample*: with name

- Typing *getexample* <exemplename> copies the example to the current directory

```
> getexample R_example  
`/opt/software/powertools/share/examples//R_example' ->  
`./R_example'  
`/opt/software/powertools/share/examples//R_example/R_job.qsub' ->  
`./R_example/R_job.qsub'  
`/opt/software/powertools/share/examples//R_example/example.R' ->  
`./R_example/example.R'  
`/opt/software/powertools/share/examples//R_example/README' ->  
`./R_example/README'
```

Examples:

- README
- Example data
- Example Code:
- Submission Script
- Etc...

Example.R

```
z=rnorm(10000,mean=10,sd=2)
mean(z)
sd(z)
pdf(file="r_histogram.pdf")
hist(z,freq=FALSE,nclass=100)
```

README

```
#!/bin/bash
# to submit this job, type
qsub R_job.qsub
```

qsub_R_jog.qsub

```
#!/bin/bash -login
# how long?
#PBS -l walltime=00:10:00

# how much memory?
#PBS -l mem=400mb

# specify resources needed.
#PBS -l nodes=1:ppn=1

# you need this flag if you're generating graphics, which we are not in this case
#PBS -X

# email me
#PBS -m abe

# change to current directory
cd $PBS_O_WORKDIR

# run R commandline with the Rscript command
Rscript -example.R
```

getexample - code

```
#!/bin/bash
#DESCRIPTION Download user examples
#LABEL Files
location=/Path_to_example_parent_directory/

#Display help message if no example is given
if [ "$1" == "" ]
then
    echo "Download an HPC example:"
    echo "usage:"
    echo "  getexample <examplename>"
    echo ""
    echo "Where <examplename> is the name of the example you want to "
    echo "download. This will create a directory named examplename which"
    echo "you can cd into and hopefully read the README file (if one is"
    echo "available) or just submit the *.qsub file."
    echo ""
    echo "For Example:"
    echo "  getexample helloworld"
    echo ""
    echo "Possible example names:"
    ls $location
    exit 0
fi

cp -r -v -u ${location}/${1} .
```

Started today: <https://github.com/colbrydi/getexample>

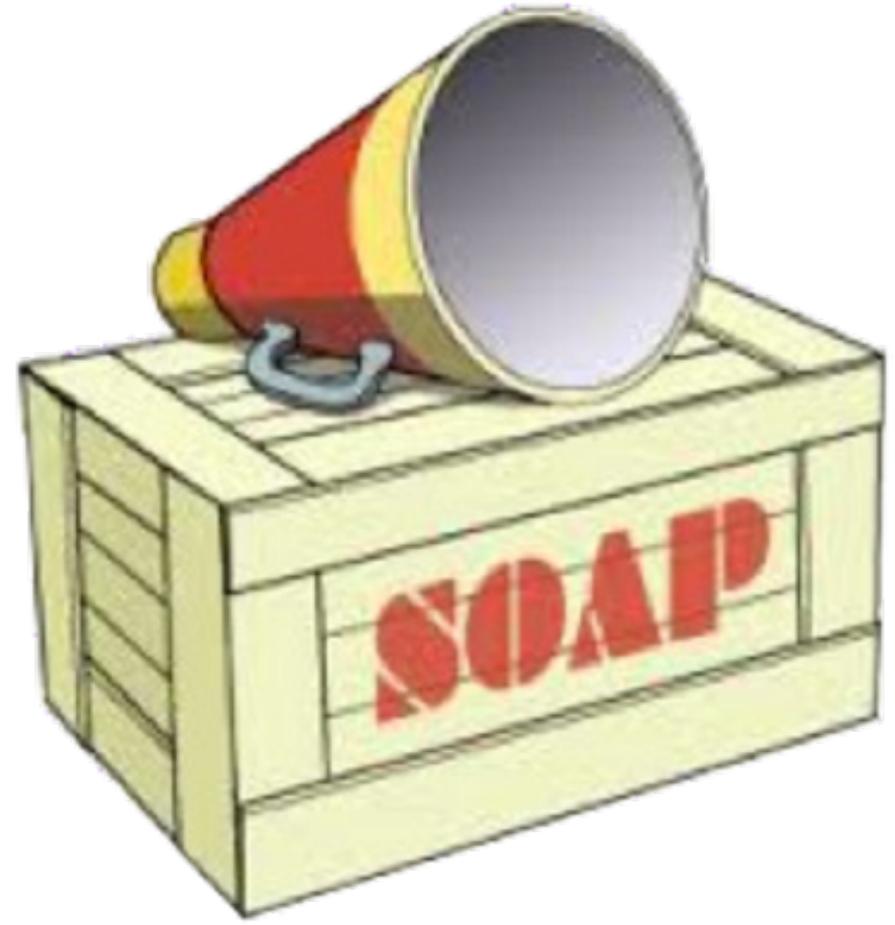
Curation is key!!

- Does the example include a readme file?
- Is it clear (without reading the readme file) what needs to be done to run the example?
- Is it clear (with the readme file) what needs to be done to run the example?
- Does the example run without any changes to the files?
- If changes are needed, is it clear how to make the changes (assuming little or no familiarity with Linux CLI)?
- Is it clear what the example is doing?
- Is it clear whether the example ran successfully?
- Is it clear to someone familiar with the software how to modify the example to use their own data/inputs?
- Is it clear to someone unfamiliar with the software how to modify the example to use their own data/inputs?
- Are there any confusing or missing steps or instructions in the example?

Community Challenge

- Lets put well curated examples on every shared system
- Make the *getexample* command ubiquitous





Questions?