

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
• 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
**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 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

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abaqus example</td>
<td>fortran_openmpblas, MKL_Example</td>
</tr>
<tr>
<td>DDT examples</td>
<td>MATLAB_openmpblas, MKL_Example</td>
</tr>
<tr>
<td>MATLAB 2014_parfor2</td>
<td>build_velvet, helloMPI</td>
</tr>
<tr>
<td>OpenMP_profiling</td>
<td>helloMPI, MKL_FFTW</td>
</tr>
<tr>
<td>ADMB_example</td>
<td>shell_scripting, burn_heat, helloOpenMP</td>
</tr>
<tr>
<td>econ_examples</td>
<td>MKL_micro</td>
</tr>
<tr>
<td>MATLAB_basic</td>
<td>Circuitscape_examples, helloWorld, MKL_parallel</td>
</tr>
<tr>
<td>paraview_basic</td>
<td>simpleMatlab, Circuitscape_examples, helloWorld, MKL_parallel, simpleMatlab</td>
</tr>
<tr>
<td>ADMB_example2</td>
<td>R_example, blender_farm</td>
</tr>
<tr>
<td>espresso_benchmark</td>
<td>blast</td>
</tr>
<tr>
<td>MATLAB_bltcr</td>
<td>MATLAB_patternsearch</td>
</tr>
<tr>
<td>pcap_example</td>
<td>R_example, blender_farm</td>
</tr>
<tr>
<td>affinity</td>
<td>gmp_mpfr</td>
</tr>
<tr>
<td>fftw</td>
<td>MIC_examples, OpenCL_hello_world, BOOST_example</td>
</tr>
<tr>
<td>MATLAB Compiler</td>
<td>gmp_mpfr, MIC_programming</td>
</tr>
<tr>
<td>pcap_example</td>
<td>gromacs</td>
</tr>
<tr>
<td>allinea_map</td>
<td>gromacs</td>
</tr>
<tr>
<td>fluent3D</td>
<td>gromacs</td>
</tr>
<tr>
<td>MATLAB_compiler2</td>
<td>gromacs</td>
</tr>
<tr>
<td>PETSc_example</td>
<td>gromacs</td>
</tr>
<tr>
<td>Amber_CUDA_example</td>
<td>gromacs</td>
</tr>
<tr>
<td>fluentMPI</td>
<td>gromacs</td>
</tr>
<tr>
<td>MATLAB_many_jobs</td>
<td>gromacs</td>
</tr>
<tr>
<td>Python_MPI</td>
<td>gromacs</td>
</tr>
<tr>
<td>Amber_example</td>
<td>gromacs</td>
</tr>
<tr>
<td>fortran_openmp</td>
<td>gromacs</td>
</tr>
<tr>
<td>MATLAB_movie</td>
<td>gromacs</td>
</tr>
<tr>
<td>Python_numpy</td>
<td>gromacs</td>
</tr>
<tr>
<td>avida_bltcr</td>
<td>gromacs</td>
</tr>
</tbody>
</table>
Running *getexample*: with name

- Typing *getexample* `<examplername>` copies the example to the current directory.

```r
> 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

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

README

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

qsub_R_jog.qsub

```bash
#!/bin/bash

# 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 ab

# change to current directory
cd $PBS_O_WORKDIR

# run R commandline with the Rscommand command
Rscript -example.R
```
#!/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?