



# CONTAINERS AT TACC

October 11, 2017



# NOT A BACKGROUND SLIDE

- ▶ (Since I'm the 4<sup>th</sup> or 5<sup>th</sup> speaker, not going to try and explain what containers are or why we use them. TACC local details only!)

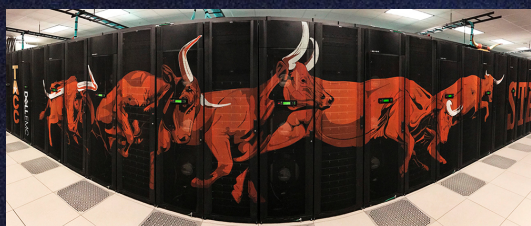


# OUR GENERAL CONTAINERS PHILOSOPHY

- ▶ Develop with Docker
- ▶ Run locally with Docker
- ▶ *Run your image at scale with us in Singularity!!!*



# CONTAINERS ARE SUPPORTED ON EVERY TACC PRODUCTION PLATFORM



## **Stampede Stampede-2**

•~ #12 HPC system in the world for computation 350k CPU core 18 PF  
**Lonestar 5** Cray XC40 30,000 Intel Haswell cores 1.25 PF

**Wrangler** Data Analytics 0.6 PB usable DSSD flash storage w 1 TB/s read rate + 10 PB Lustre

**Maverick** Interactive and GPU computing NVIDIA Kepler K40 GPGPU

**Jetstream** Production Cloud

**Chameleon** Cloud Testbed

**Hikari** Secure Computing





# HOW DO I DO THIS?

- ▶ *module load tacc-singularity/2.3.1*
- ▶ In job script (usually Slurm):
  - ▶ *singularity exec <Image name> <command> <parameters>*
  - ▶ Or, pull from a repo:
    - ▶ *singularity exec docker://<registryname>/<imagename> <command> <params>*
- ▶ That's it, you are running!
- ▶ *If you use one of our supported repository, we have hacked in a few extra bells and whistles that will automatically be supported in your image.*



# A BIT OF HISTORY, AND SOME UPGRADES WE'VE MADE

- ▶ We started public support in October of 2016 on Stampede
  - ▶ But only in some one off cases for particular users (originally FMRI processing).
- ▶ We now support more than 3,500 containerized apps!
- ▶ *We are committed to continuing to support containers, from Day 1, on every production platform, now and future.*
- ▶ Images pulled from BioContainers (3,500+) are automatically modded to have \$HOME, \$SCRATCH, and \$WORK defined and mounted, with the same filesystems you see native on TACC Systems.
- ▶ Making future changes to our Modules systems, so cloud/container friendly things like *"Module load Python latest"* will work.



# CROWDSOURCING RESEARCH SOFTWARE

- ▶ Conda
  - ▶ Package, dependency and environment management for any language
- ▶ BioConda
  - ▶ Conda “channel” specializing in Bioinformatics. 2000+ Github recipes + build system that turns them into Conda packages. 135 contributors!
- ▶ BioContainers
  - ▶ Community framework that maintains BioConda packages as Docker containers



BioContainers Registry UI

Secure <https://biocontainers.pro/registry/#/>

BioContainers Registry UI BioContainers GitHub

All Containers and Tools in BioContainers (2471 !!!)

Search...

Container	Description	Real Name	Last Modified	Starred/Starts	Popularity	Registry Link
<a href="#">blast</a>	basic local alignment search tool	biocontainers/blast	23/06/2017			
<a href="#">comet</a>	an open source tandem mass spectrometry sequence database search tool	biocontainers/comet	23/06/2017			
<a href="#">biocontainers</a>	Biocontainers base Image	biocontainers/biocontainers	26/06/2017			
<a href="#">tpp</a>	a collection of integrated tools for MS/MS proteomics	biocontainers/tpp	23/06/2017			
<a href="#">vcftools</a>	"tools written in Perl and C++ for working with VCF files	biocontainers/vcftools	23/06/2017			
<a href="#">bedtools</a>	a powerful toolset for genome arithmetic	biocontainers/bedtools	23/06/2017			
<a href="#">bwa</a>	Burrow-Wheeler Aligner for pairwise alignment between DNA sequences	biocontainers/bwa	23/06/2017			
<a href="#">samtools</a>	Tools for manipulating next-generation sequencing data	biocontainers/samtools	23/06/2017			

*All of these available on Stampede 2!*



# OUTCOMES

- ▶ Use nearly any public Docker image
  - ▶ Bioinformatics & computational biology
  - ▶ Machine & deep learning
  - ▶ Data analytics applications
  - ▶ New user-developed codes
- ▶ Develop locally. Package with Docker. **Share & use nearly anywhere (Stampede, Comet, AWS, Azure, Jetstream, localhost, etc.)**
  - ▶ Increase collaborativity
  - ▶ Increase velocity of exploration



**THANKS!**  
**DISCUSSION...**