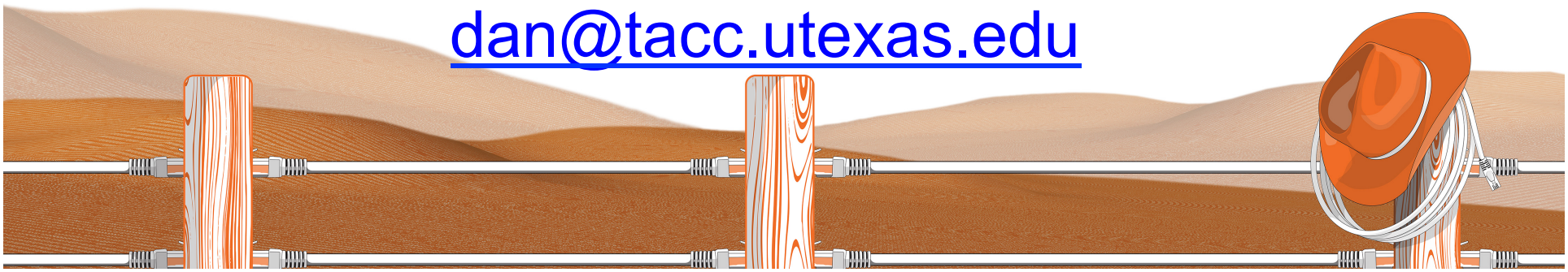
Wrangler : A New Generation of Data Intensive Cluster Computing

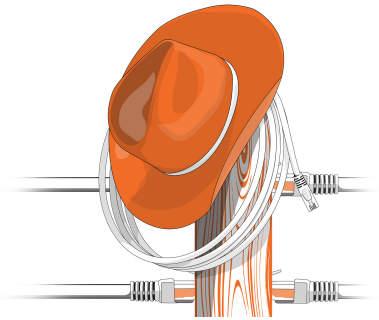
Dan Stanzione
Texas Advanced Computing Center
High Performance Big Data
June 29th, 2015
Columbus, Ohio

dan@tacc.utexas.edu



Acknowledgments

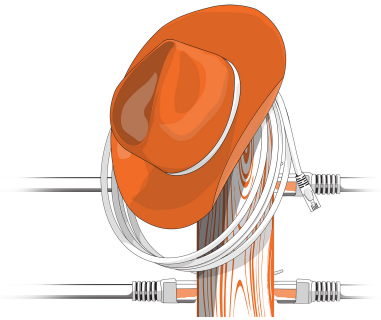
- The Wrangler project is supported by the Division of Advanced Cyberinfrastructure at the National Science Foundation.
  - Award #ACI-1447307 *"Wrangler: A Transformational Data Intensive Resource for the Open Science Community "*



What is Wrangler

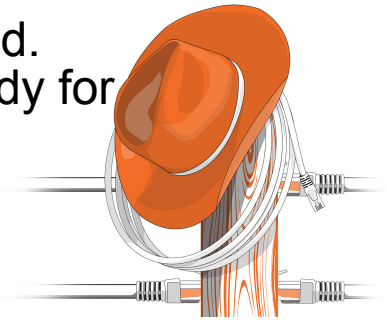
- Wrangler is a new data-intensive supercomputing system
- Built from the ground up for “data intensive” applications.
- HPC and “Big Data” have a lot in common
  - But the overlap isn’t 100% in all applications.
  - While Exascale computers will generate phenomenal amounts of data, not *every* data problem will map perfectly.
- New technologies can deal with the shortcomings in HPC Cluster architectures

TACC



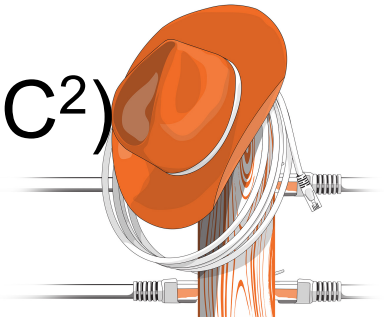
What Wrangler Enables

- Stampede is fantastic for tens of thousands of people
  - But has some limitations (metadata performance, I/O per node).
  - And makes assumptions about software (Distributed memory, MPI I/O for HPC, etc.).
- While *theoretically* we could fix all the software and workflows in the world to run well in this environment. . .
  - Practically, we will never have the time or resources.
  - Wrangler will simply lift up some of this “bad code”. Done your computation inside a SQL DB? No problem.
  - And the code that does get optimized will be able to do things we couldn’t imagine on Stampede. – Wrangler is *not* just the “bad code” system.
- On Wrangler, data isn’t just scratch; it’s a first class object to be protected, curated, kept on the system, and shared with the world. We think an enormous fraction of the scientific community is ready for this paradigm.

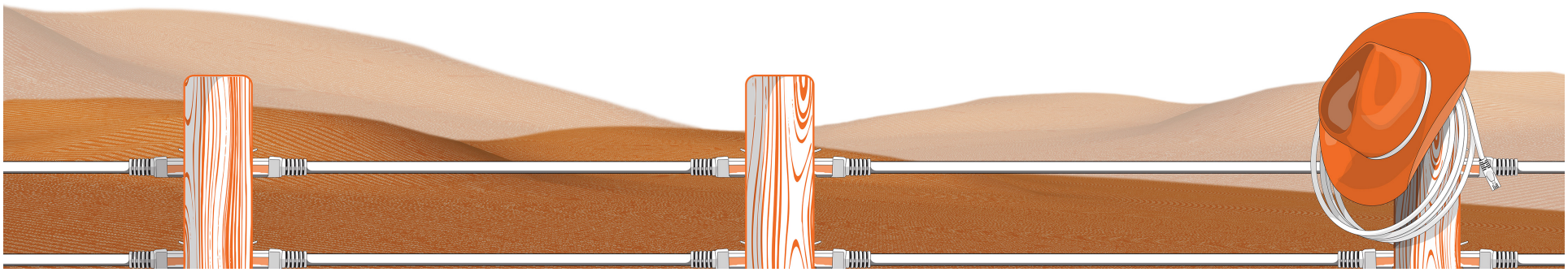


Project Partners

- Academic partners:
  - TACC – Primary system design, deployment, and operations
  - Indiana U. ; Hosting/Operating replicated system and end-to-end network tuning.
  - U. of Chicago: Globus Online integration, high speed data transfer from user and XSEDE sites.
- Vendors: Dell, DSSD (subsidiary of EMC^2)



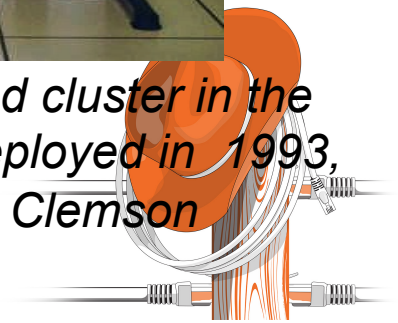
[How we got here]



Once upon a time, most of us built garage-style clusters

- We'd cobble together some machines and call it a cluster.
- The computer scientist or engineer would build the cluster (often including the furniture), add user accounts, and in theory the users should take it from there.
- We've gotten a little more sophisticated since then…

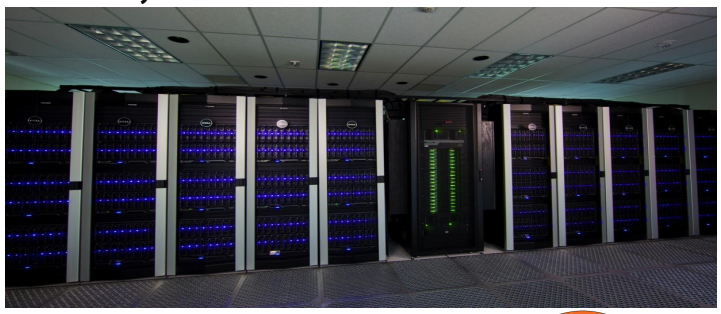
Grendel, the second cluster in the Beowulf project, deployed in 1993, at the PARL Lab at Clemson



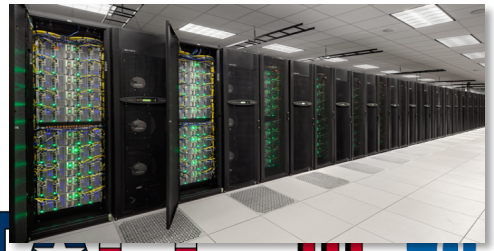
The Texas Advanced Computing Center: A World Leader in High Performance Computing

***1,000,000x** performance increase in UT computing capability in 10 years.*

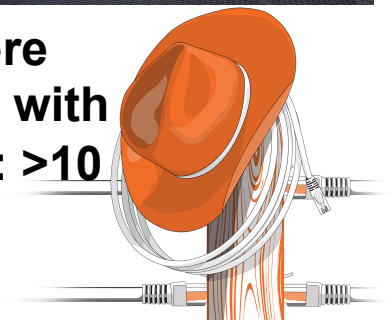
Ranger: 62,976 Processor Cores, 123TB RAM, 579 TeraFlops, Fastest Open Science Machine in the World, 2008



Lonestar: 23,000 processors, 44TB RAM, Shared Mem and GPU subsystems, #25 in the world 2011

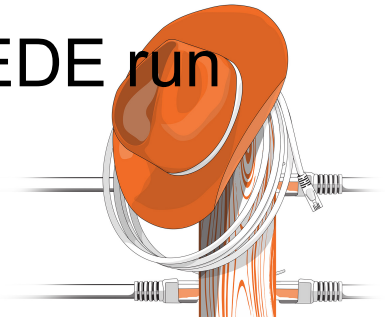


Stampede: #7 in the world, Somewhere around half a million processor cores with Intel Sandy Bridge and Intel MIC, Dell: >10 *Petaflops*.

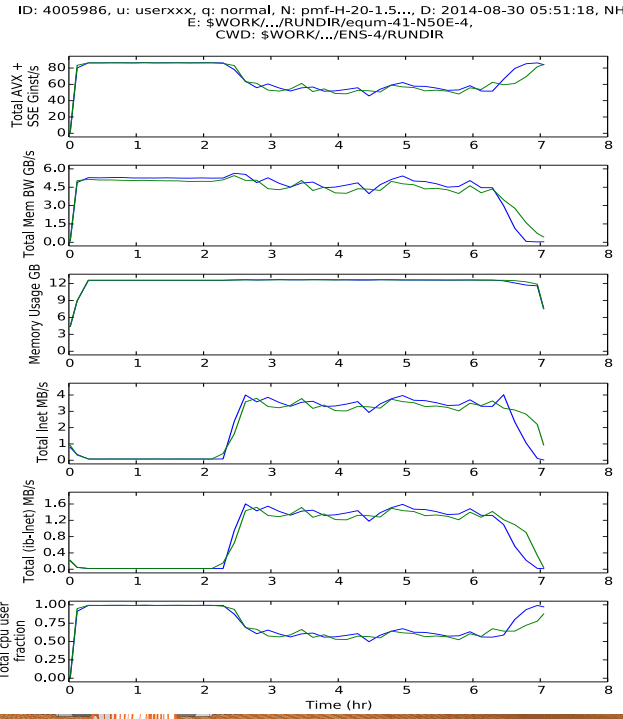


We've gotten pretty good at making big, useful clusters.

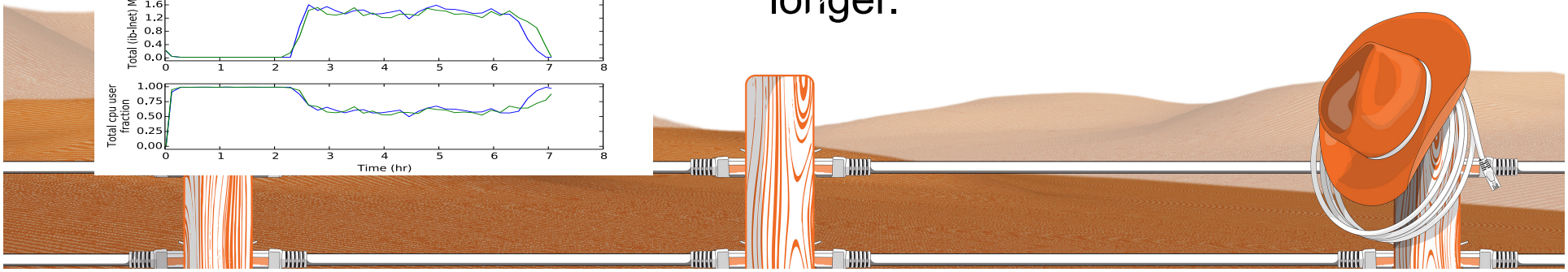
- Through 30 months of Production Operation, Stampede by all measures is remarkably successful.
  - Over *1.7 Billion* Service Units delivered.
  - Over 5 **million** successful jobs
  - 2,283 distinct projects received allocations
  - 6,926 Individuals have actually run a job (~10,000 accounts).
  - 98% cumulative uptime (target: 96).
  - 5,006 User Tickets Resolved last year
  - 2,000+ users attended training last year.
- Formal requests from the community from XSEDE run ~500% available hours



Even in the workload we *already* have. we'd see jobs like this:



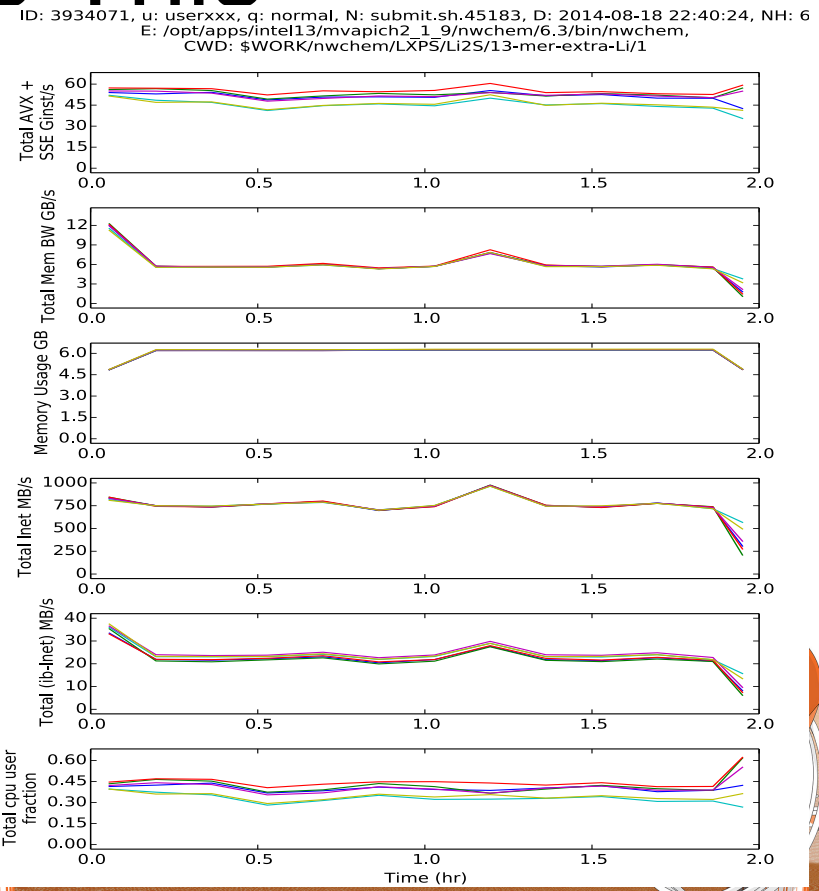
In this biophysics jobs, high metadata traffic starts to limit performance about 2 hours in, and in fact makes the job run 3 hours longer.



Or Like This

In this chemistry job, high metadata traffic throughout (a proxy for many small I/O operations) keeps performance at 1/3rd of peak throughout the 2 hour run.

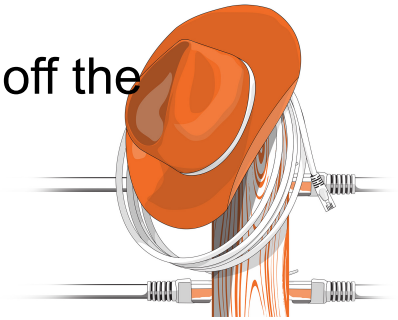
How fast is this computer" is a multi-dimensional quantity, and in a world of small, random access, it means different things.



Hadoop

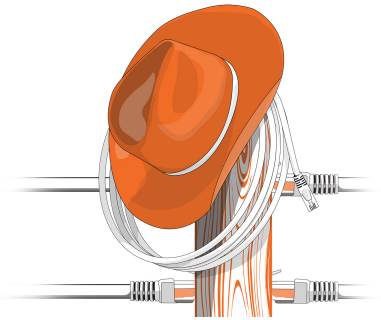
- In ~2011, we discovered an exciting new failure mode in large scale systems.
- We called this failure mode “Hadoop”.
- Recipe for success: take a big system
  - A huge central filesystem
  - Optimized for large, sequential, access
  - With a highly tuned, low level C interface
- And on that run software that:
  - Assumes a small, massively distributed filesystem.
  - Optimized for very small files.
  - With an untuned, well… Java.
- We deployed Rustler at our own expense to keep these users off the supercomputers.

TACC

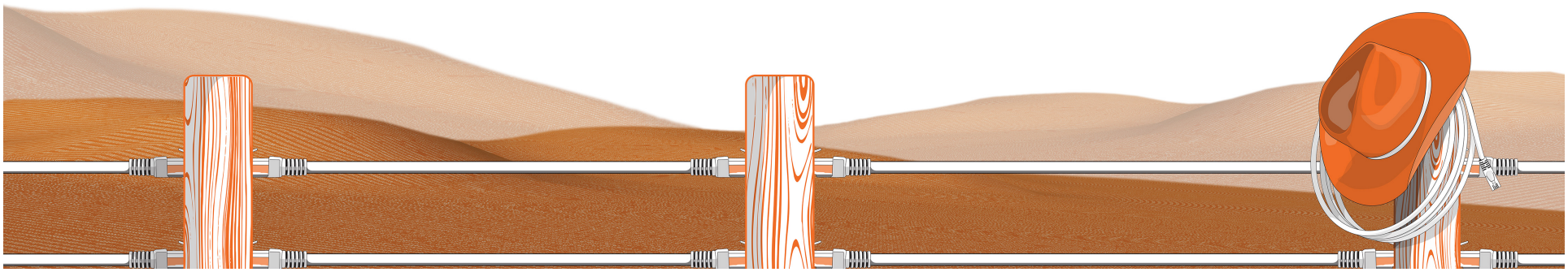


New User Communities

- A flood of new disciplines into computational sciences in last decade (led by life sciences).
  - Computational Economics requests as much time today as Computational Physics did in 2003!!!
- Most of these new areas have a few things in common
  - *Driven by data*, not equation based models
  - Mostly non-programmers
  - Less traditional languages, less performance tuning.



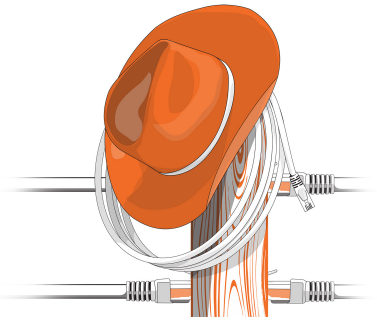
To address these problems, we proposed Wrangler



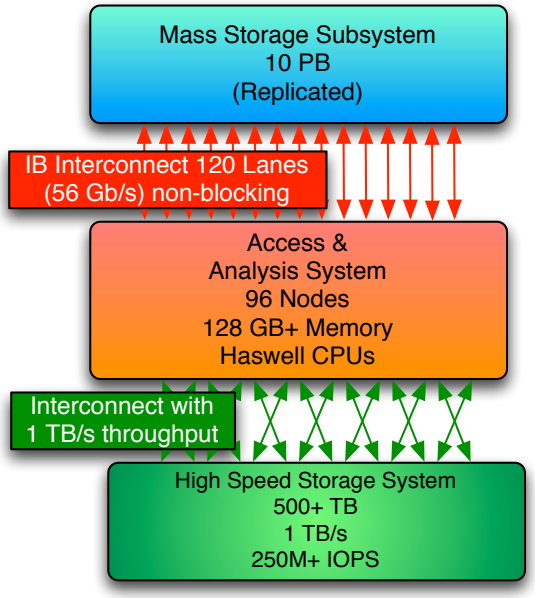
Goals of the Wrangler Project

- Our analysis of community needs indicated we needed:
  - To address the data problem in multiple dimensions
    - Big (and small), reliable, secure
    - Lots of data types: Structured and unstructured
    - Fast, but not just for large files and sequential access. Need high transaction rates and random access too.
  - To support a wide range of applications and interfaces
    - Hadoop, but not *just* Hadoop.
    - Traditional languages, but also R, GIS, DB, and other, perhaps less scalable things.
  - To support the full data lifecycle
    - More than scratch
    - Metadata and collection management support
- Wrangler is designed with these goals in mind.

TACC

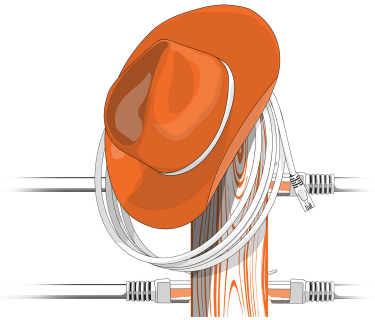


Wrangler Hardware

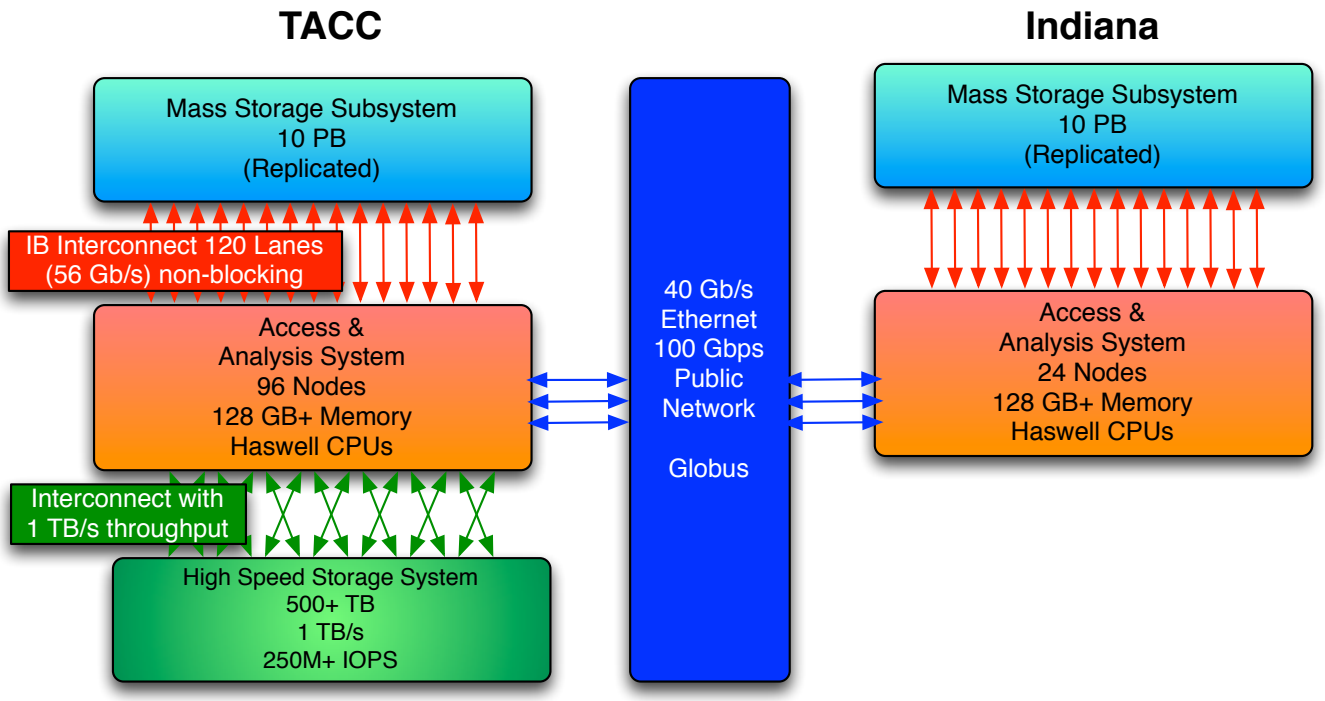


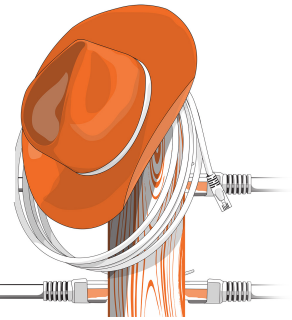
Three primary subsystems:

- A 10PB, replicated disk storage system.
- An embedded analytics capability of several thousand cores.
- A high speed global object store
  - 1TB/s
  - 250M+ IOPS



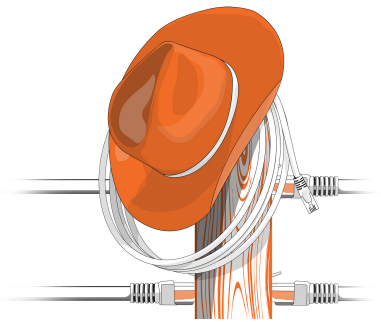
Wrangler At Large





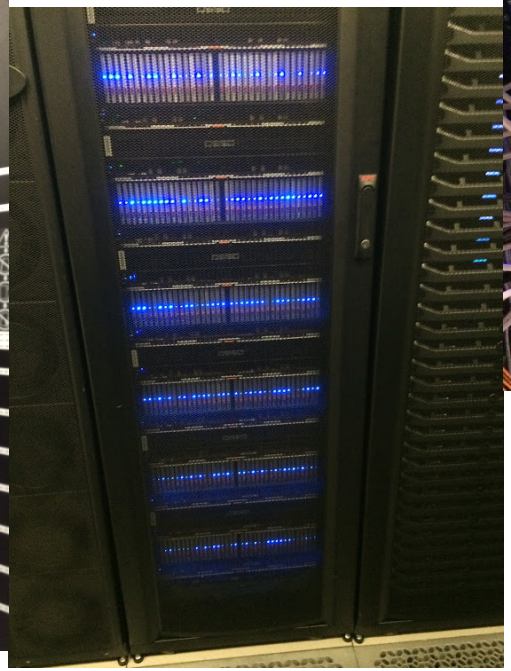
Storage

- The disk storage system will consist of more than 20PB of raw disk for “project-term” storage.
  - Geographically replicated between TACC and Indiana (more reliable than traditional scratch).
  - Ingest at either site.
  - Exposed to users on the system as a traditional filesystem



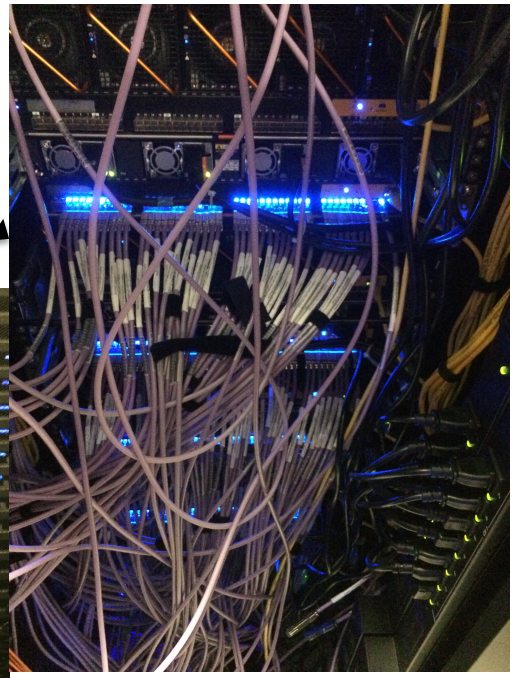
5 of the 35 storage JBODs and Lustre Metadata server

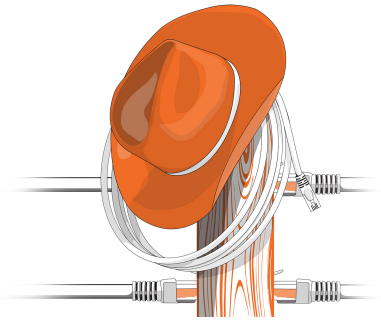
12 of 120 compute Haswell nodes



6 of the 10 D5s

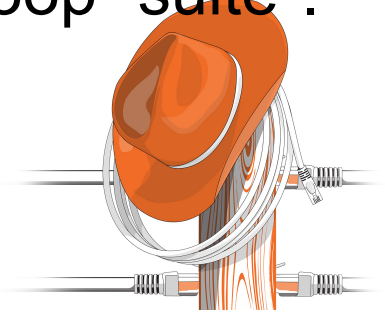
Half of the Final Cabling on 4 D5s





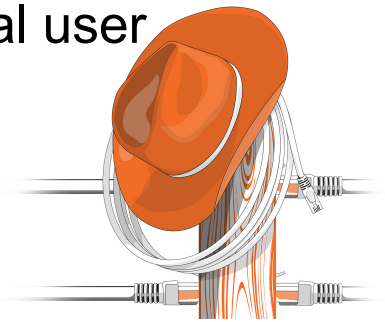
Wrangler Software and Use Cases

- The D5 storage is used in several ways:
  - As a GPFS filesystem, for POSIX based applications
  - As an HDFS filesystem, for Hadoop and other Map Reduce applications.
  - As a SQL database backend
  - Flood based object store, for novel data applications
- In addition to our “traditional” HPC stack, we support R, databases, NoSQL databases, and the full Hadoop “suite”.

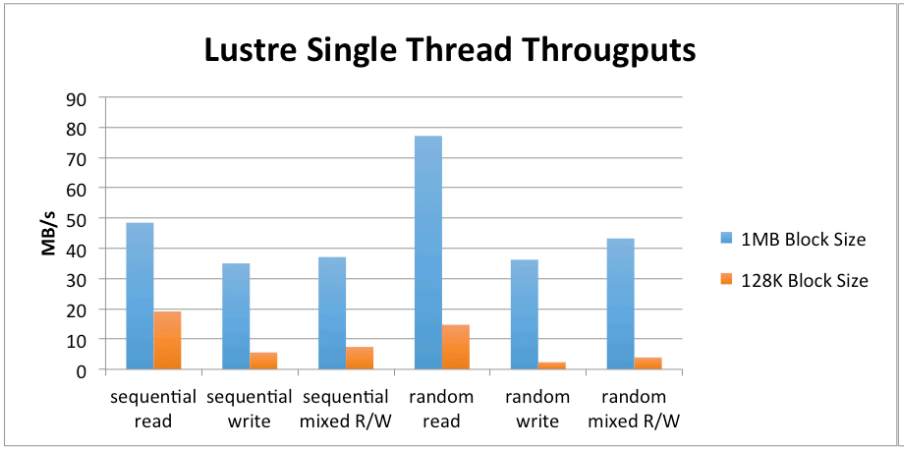


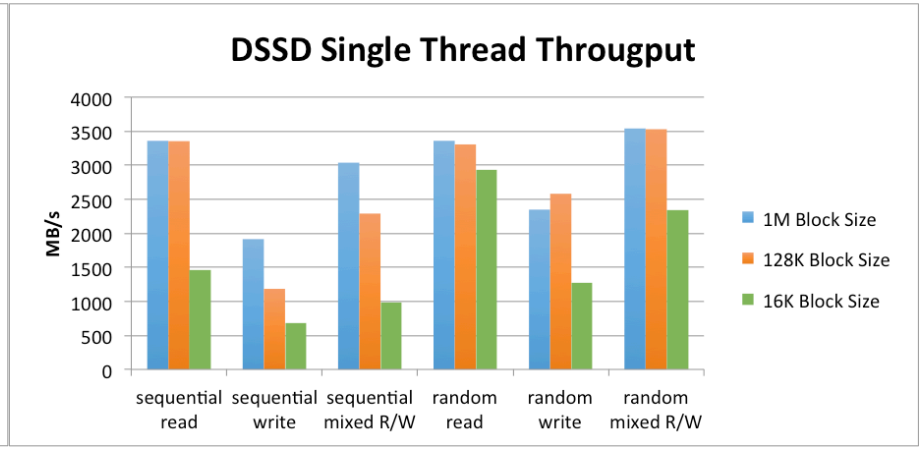
Status Today

- We are in early operations mode, with a number of projects on the system, and more being added each week (more on this in a moment).
- We have chosen to partition the Flash storage (to increase the number of compute nodes, and create some semi-persistent spaces).
  - Some nodes will still see all 10 DSSD devices.
  - Two pools of 48 nodes with >200TB available each
  - 16 additional nodes for persistent services.
  - Still experimenting with optimal configuration for actual user workloads

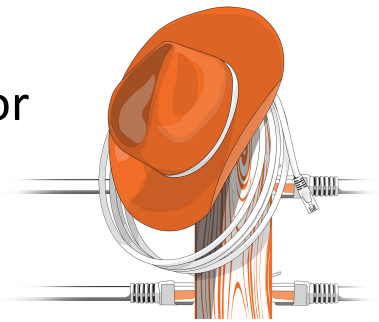


Where DSSD Really Shines



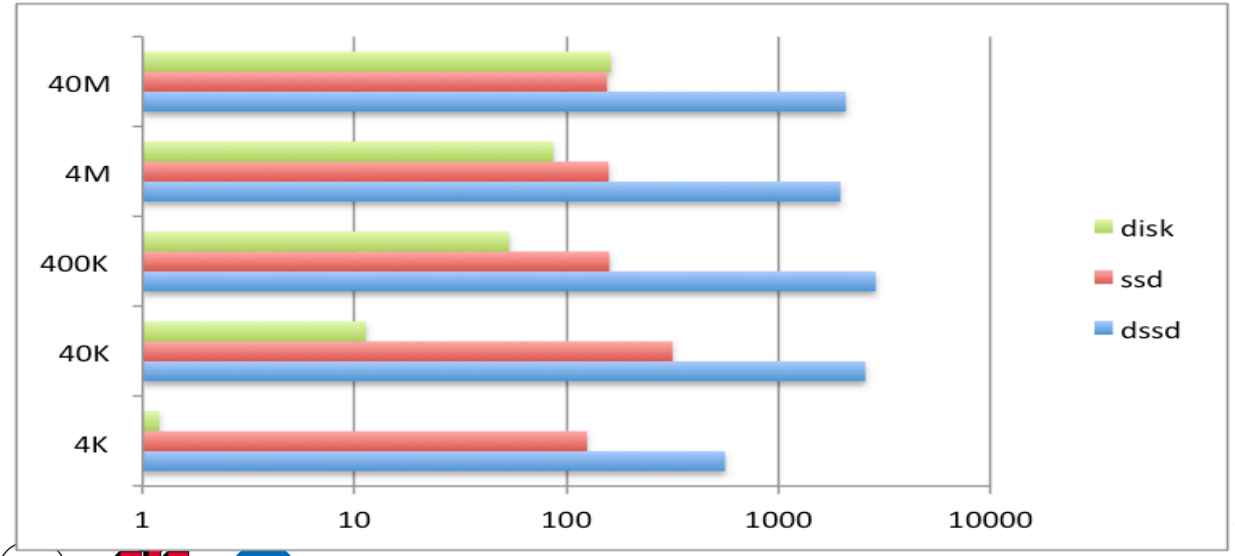


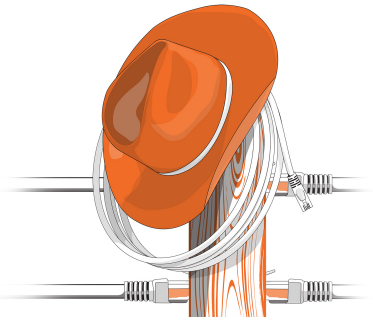
- Single thread IO for different block sizes
  - Flash is faster than single spinning disk (no surprise)
  - DSSD sustains most throughput for small block sizes and for sequential and random I/O paterns



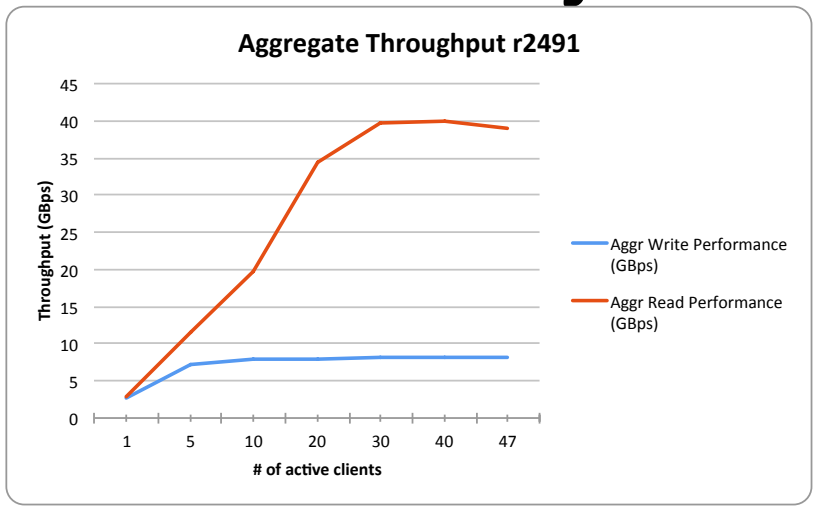
A little telling early data

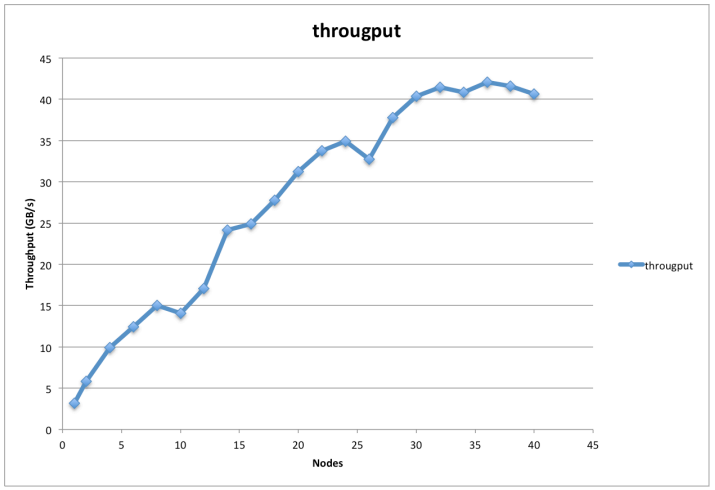
- At big write sizes, we're 10x better than "conventional" disk.
- At small write sizes, we're ~400x better than disk .



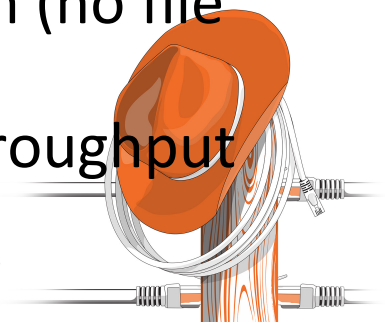


Object Store Rates



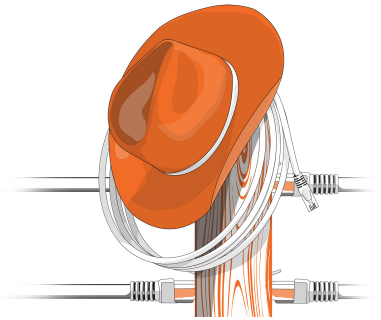


- Uses DSSD flood interface for direct access to Flash (no file systems)
- Current Object store can sustain 46+ GB/s read throughput



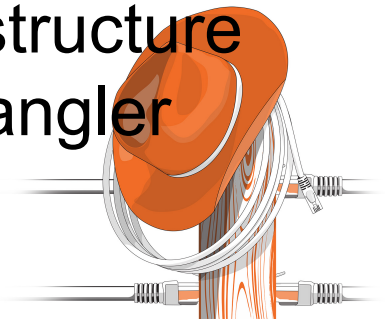
More Than the Parts

- We pride ourselves on our support of users…
  - but our HPC systems would fill if we weren't great at it.
  - A huge base of sophisticated, demanding users, some of whom have used HPC in their research for 30 years.
- With Wrangler, we are bringing in a myriad of new communities which will need much more support; and even the old ones will need to think differently.
- We've conceived the Wrangler *project* as more than just the system, but also a set of services.



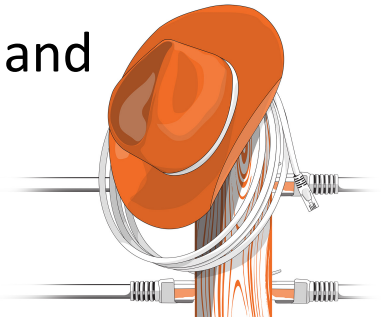
Collaborative Services

- Support for complex projects via XSEDE Extended Collaborative Data Support Services
  - Working directly with XSEDE staff
  - Focus on specific data management challenges
  - Work with teams to improve data workflows
  - Work to integrate tools into existing infrastructure
  - Transition of a projects processing to Wrangler



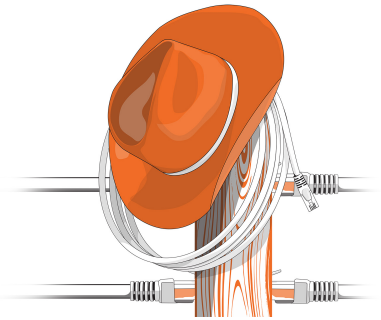
Managing Data

- Data Management & Autocuration Service
  - Leveraging Globus Online Dataset Services
  - Data Organization tying to research project to facilitate tracking of a projects data assets
  - Data Provenance leveraging emerging standards to ensure data accessibility and reusability
  - Data Fixity by automating extraction of technical metadata for files in the system
  - Data Usage by tracking overall size of a projects data and (internal and external) utilization information



A Flexible Comprehensive Software Environment

- My Hadoop style system
  - Working with Cloudera
  - More than just MapReduce (Spark? Mahout?)
- Apache Mesos
- RDB and noSQL databases with GIS integration
- System optimized R, Python, etc.

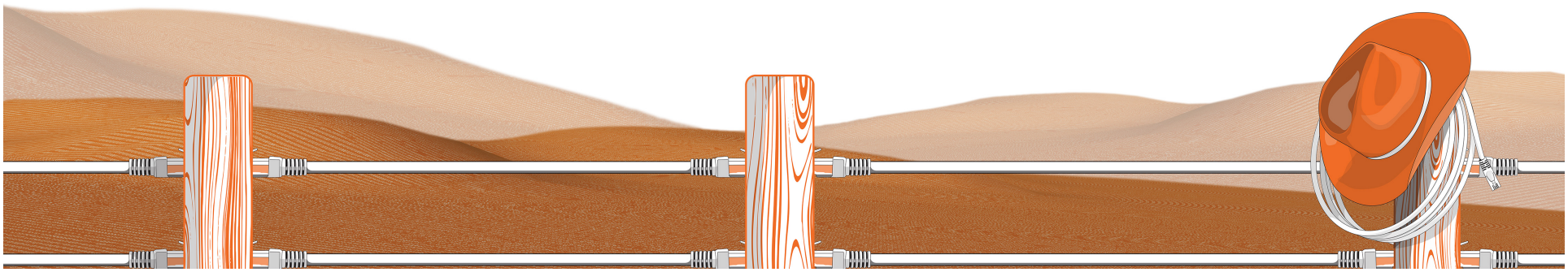


Moving Data Effectively

- End to End Network Performance Tuning
  - Leveraging I^2 and Globus based Perfsonar
    - XSP support to allocate bandwidth on demand
    - Software Defined Networking support via I2's AL2S
    - Network Performance Tools to monitor for bottlenecks and react to congestion
- Data Dock capability for the "last mile problem" where limited by network capabilities outside of I^2.

3,000 Rice genomes from the Philippines; yes, this is the actual box

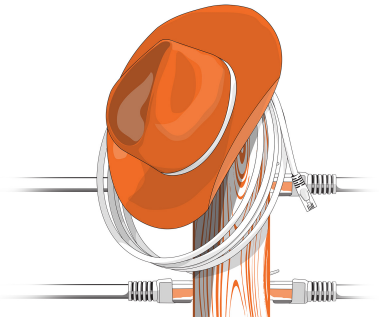
Early User Application Success Stories



OrthoMCL Science Case

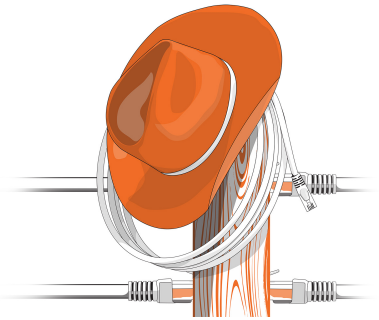
UT Austin Center for Computational Biology and Bioinformatics

- Dr Hans Hofmann, Dr. Rebecca Young
  - 6 and 8-species Gene expression comparison
  - Brain development/independent evolution of monogamous behavior
- Dr. Hans Hofmann, Dhivya Arasappan
  - *Rhazya stricta* gene family grouping
  - Medically important alkaloids



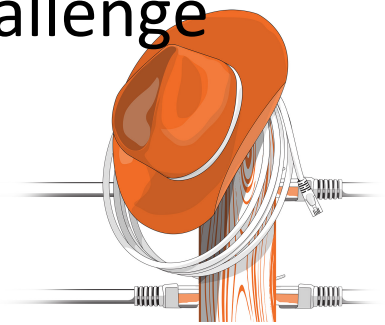
OrthoMCL Application

- "Orthologous Protein Grouping"
  - Multi-stage workflow
  - BLAST, protein grouping, results presentation
  - Protein grouping phase performed in-database
  - Both computational and I/O-Intensive
  - Order 10s of GBs databases typical
  - Computation done in SQL database



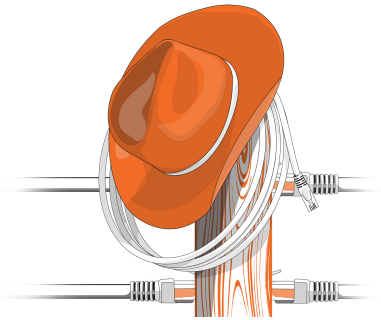
OrthoMCL Previous results

- Developers benchmark – 16-24 hours
- CCBB datasets on TACC Corral systems
  - Quad-socket servers, 64GB RAM, SAS-RAID6
  - Several steps took hours, some did not complete
  - Novel research data presented unique challenge



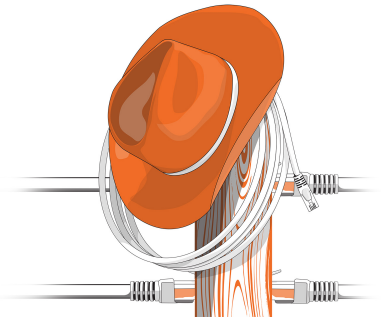
OrthoMCL Optimization Challenges

- Computational work performed in-database
- Ideal for ease-of-use but makes performance optimization difficult
- Data throughput/Random access are biggest factors in performance



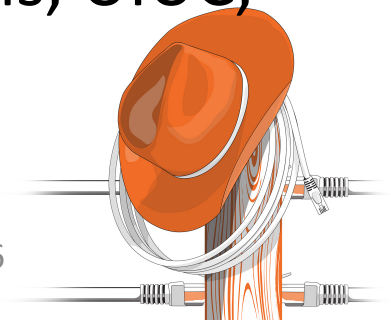
OrthoMCL/Wrangler Results

- Before Wrangler, TACC staff worked with researchers and OrthoMCL developers for > 1 month attempting to complete runs
- With Wrangler, multiple projects completed their research runs in less than a week
- All runs completed in 4-6 hours
- At least two publications in process
- Multiple new users coming for capability



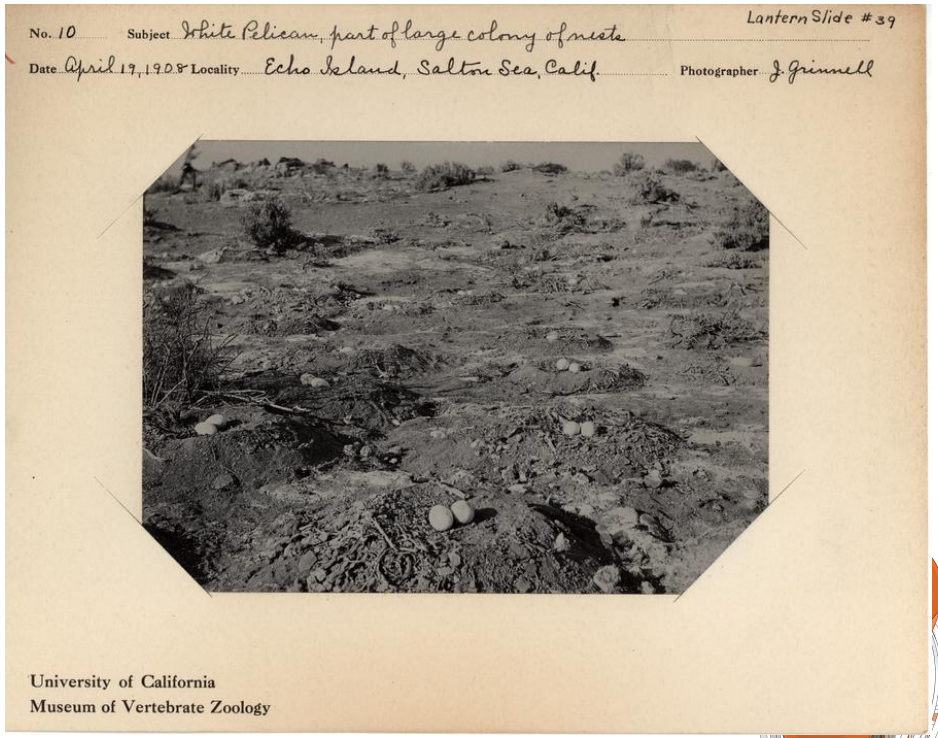
OrthoMCL Community

- Significant community of potential users
- At least one new project already allocated
- Sociogenomics RCN also planning to use OrthoMCL on Wrangler
  - Georgia Tech, Georgia State, Johns Hopkins, UIUC, Stanford, Harvard and UT Austin

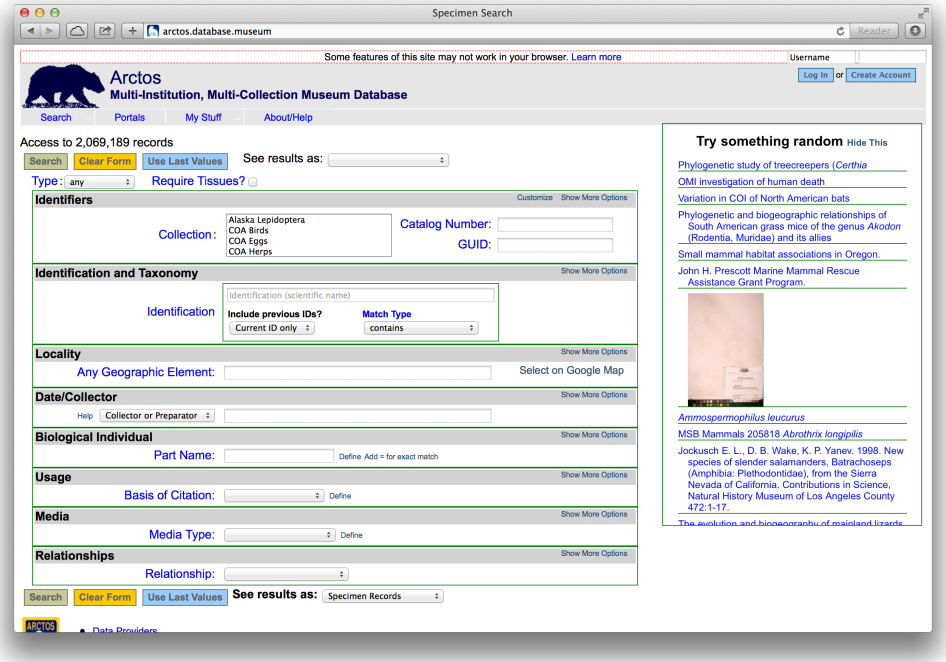


ARCTOS Specimen Catalog

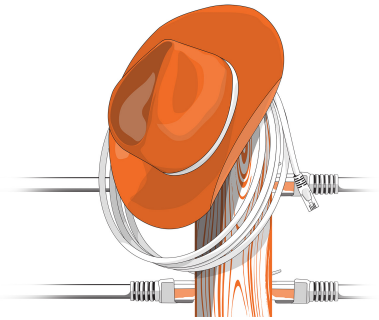
- Web-based natural history collection management, public/private data access
- 3 Million Taxon names
- 66,430 links to externally hosted datasets
- High data quality, internal consistency, automatically enforced within the database
- Key to research showing that flight causes genome shrinkage in birds.



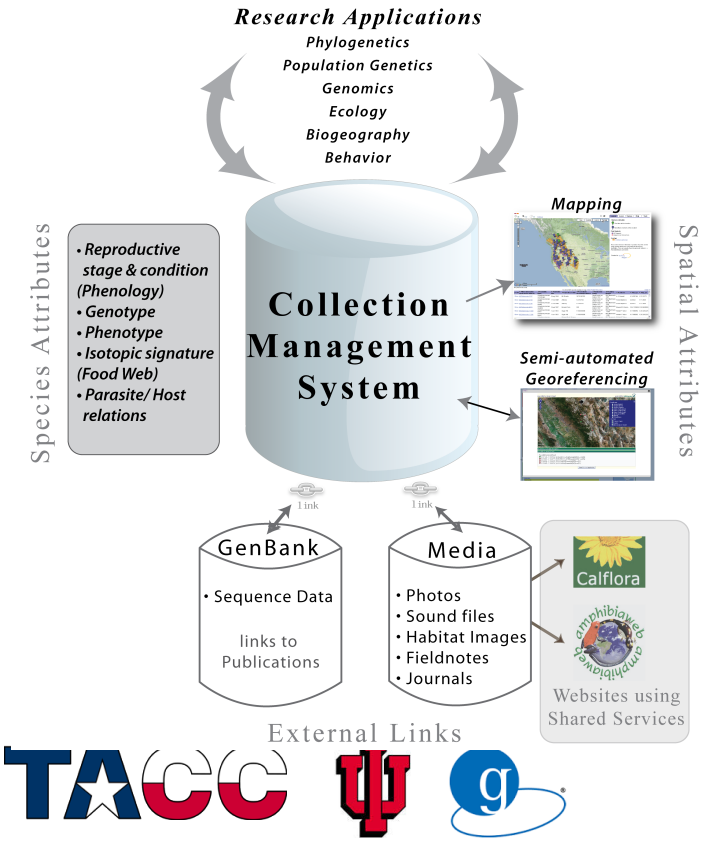
ARCTOS Specimen Catalog



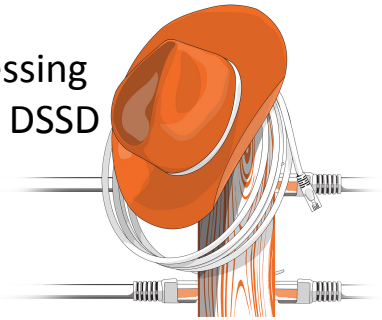
- MSB specimens with data in Arctos were key to research showing that flight causes genome shrinkage in birds.
- Arctos recently added over 15,000 specimens from the Denver Museum of Nature & Science Marine Invertebrate Collection



Arctos – Collections Catalog Workflow

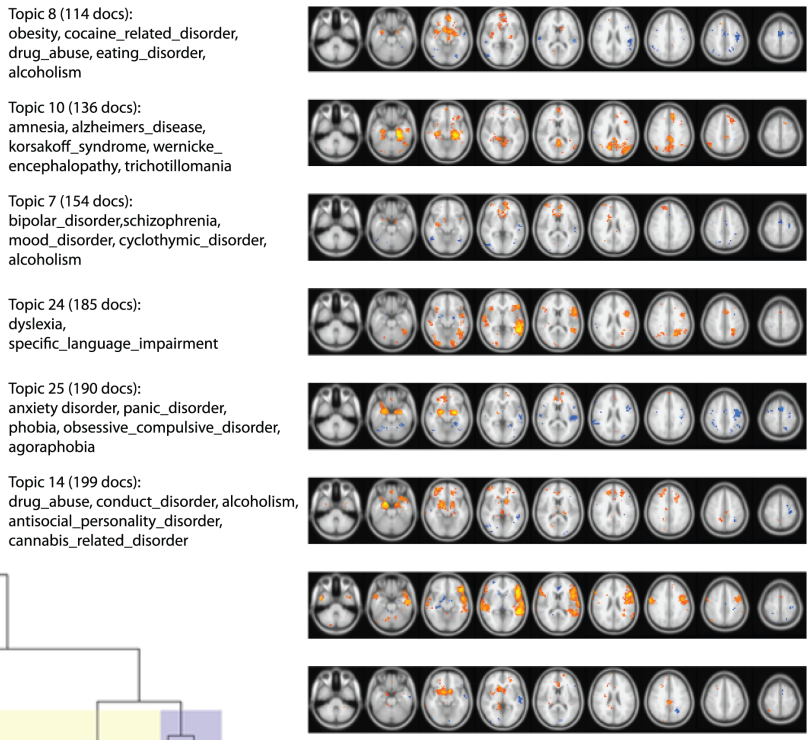


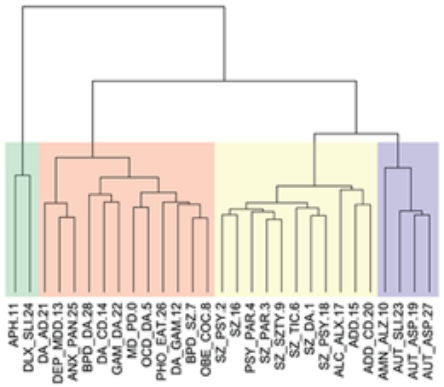
- A collection management information system for multiple biological collections with over 3M natural history museum records.
- Integrates collections for botany, entomology, herpetology, ornithology, paleontology, parasitology in a common structure
- Manages data, including specimen records, observations, tissues, endoparasites and ectoparasites, stomach contents, fieldnotes and other documents, and media such as images, audio recordings, and video.
- Pipeline for data ingest and processing nearly 3x faster on Wrangler with DSSD

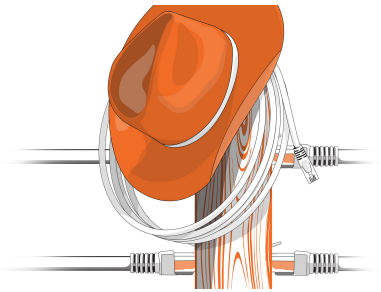


Early Use Cases

- Russ Poldrack, Stanford
  - fMRI Processing Pipelines
  - Freesurfer, R, SciPy, Hadoop/Spark
- Guatemalan National Police Historical Archive
  - 10 Million scanned document images
  - Large-scale image processing challenge



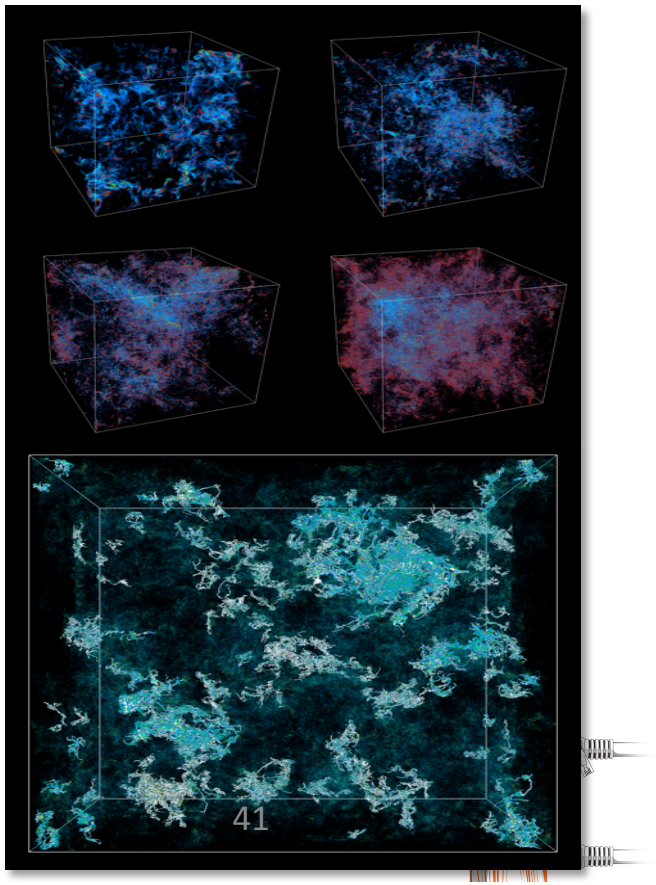




Early Use case: Analyzing Large Scale Turbulent Flow Data

P.K. Yeung, Georgia Tech; *Diego Donzis, Texas A&M; Kelly Gaither et al, TACC*

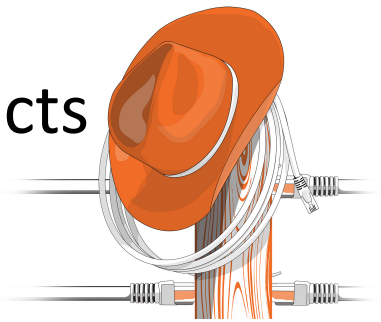
- Remote interactive visualization and data analysis of 17 time-steps (34 TB) of a turbulent flow simulation (4096^3)
- Equal parts data mining and remote interactive visualization – goal is to characterize flow behavior over time
- Data must be pre-processed and written to a Silo format to be useable for interactive manipulation
- With Longhorn, took ~3 hours to read the data set in for all 17 timesteps
- Theoretically possible to read data set in under 60 seconds on Wrangler



Gaither, K., Childs, H., Schulz, K., Harrison, C., Barth, W., Donzis, D., and Yeung, P.K., "Using Visualization and Data Analysis to Understand Critical Structures in Massive Time Varying Turbulent Flow Simulations," IEEE Computer Graphics and Applications, 32(4), Jul/Aug 2012.

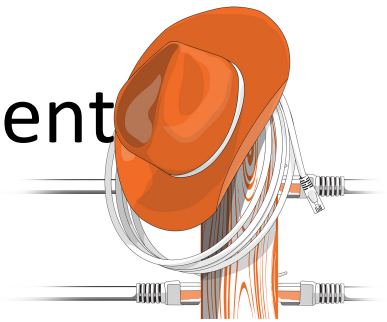
Genomics Projects and Applications

- iPlant and NCGAS projects
  - Key applications: BLAST and Trinity
  - DNA and RNA sequence processing
  - I/O Intensive due to large input datasets, need for global comparisons, and minimal reduction
  - Often the first step in longer life sciences workflows, used by many additional projects



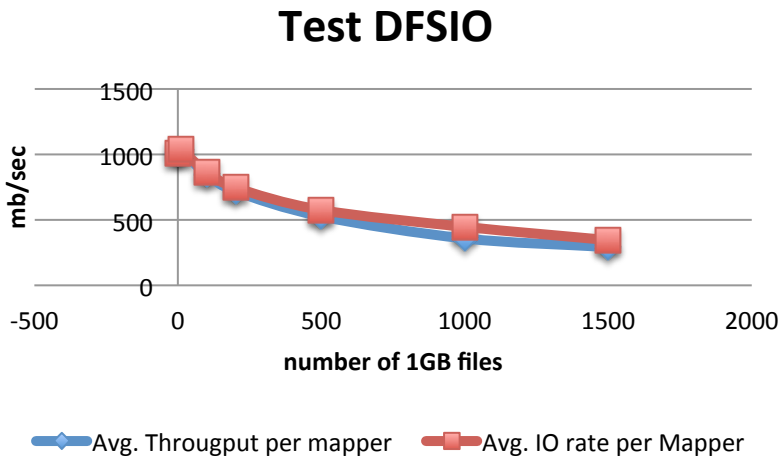
Improved Genomics Workflows

- BLAST and Trinity often run off of RAMdisk on current cluster systems
  - Wrangler eliminates this requirement
- Preliminary Trinity runs exhibit performance better than any existing TACC system
- Significant scope for further improvement

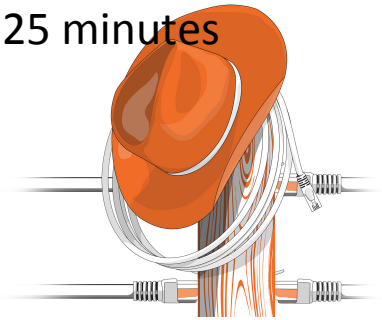


Performance at an glance

- TestDFSIO – HDFS IO Throughput Benchmark
  - Measures average mapper's throughput on 1 GB file for different MR job sizes

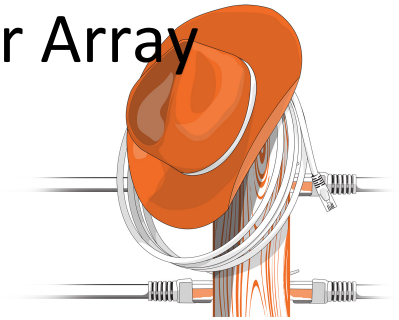


- Crossbow Genome Mo17
  - Full genome sequence workflow paralelized
  - 46 million reads vs 2 Billion BP, < 5 mins (> 1 hour on Longhorn, 25 minutes on 10 node Ivy Bridge disk system over Lustre)



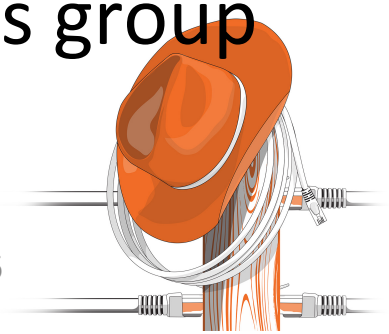
Other Projects

- Exploring daily stock market historical records since 1926
- Exploring large astrophysical data outputs from Stampede and Blue Waters
- Suporting future data processing for Astronomical missions such as the Hobby Eberly Telescope Dark Energy eXperiments and the Square Kilometer Array
- Doing out of core quantum chemistry



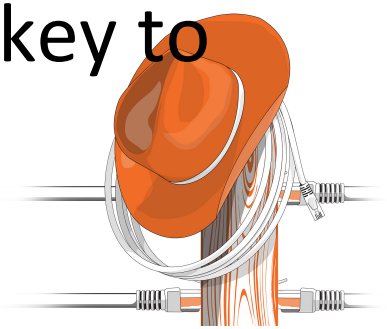
Early User Lessons Learned

- Process of allocating and configuring Flash storage is most complex task
- Reservation/Scheduling system crucial
- Many use cases will be generalizable
- *The complexity may be more than this group of users is prepared for (I'll come back to that).



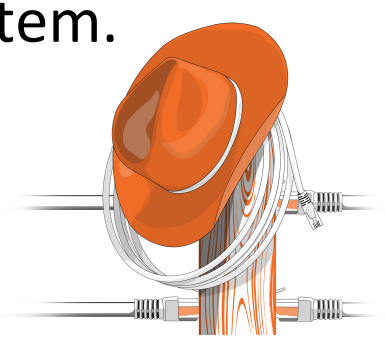
Early User Lessons Learned 2

- Wrangler is a unique resource
- Users will not always know if it is appropriate
- If it is appropriate, users will not necessarily know how best to make use of it
- Consulting and technical expertise are key to making effective use of Wrangler



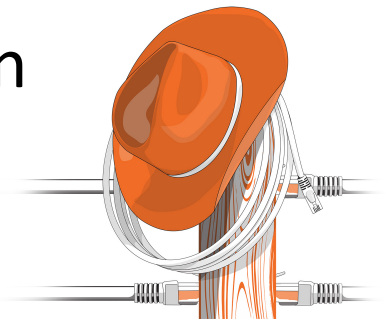
Different Allocation Model

- Wrangler is not a normal HPC cluster
- The definition of a Wrangler Computing Service Unit:
  - One *NODE* hour = One SU
    - 24 CPU Cores
    - 128 GB RAM
    - 4 TB allocation on flash filesystem during that time.
    - Bandwidth from one node to DSSD storage
- Standard storage allocation for long term disk system.
- Time on system manage differently
  - Data Campaigns



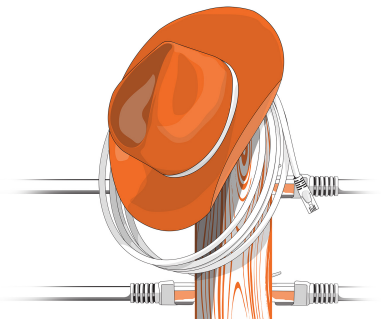
Reservation Model

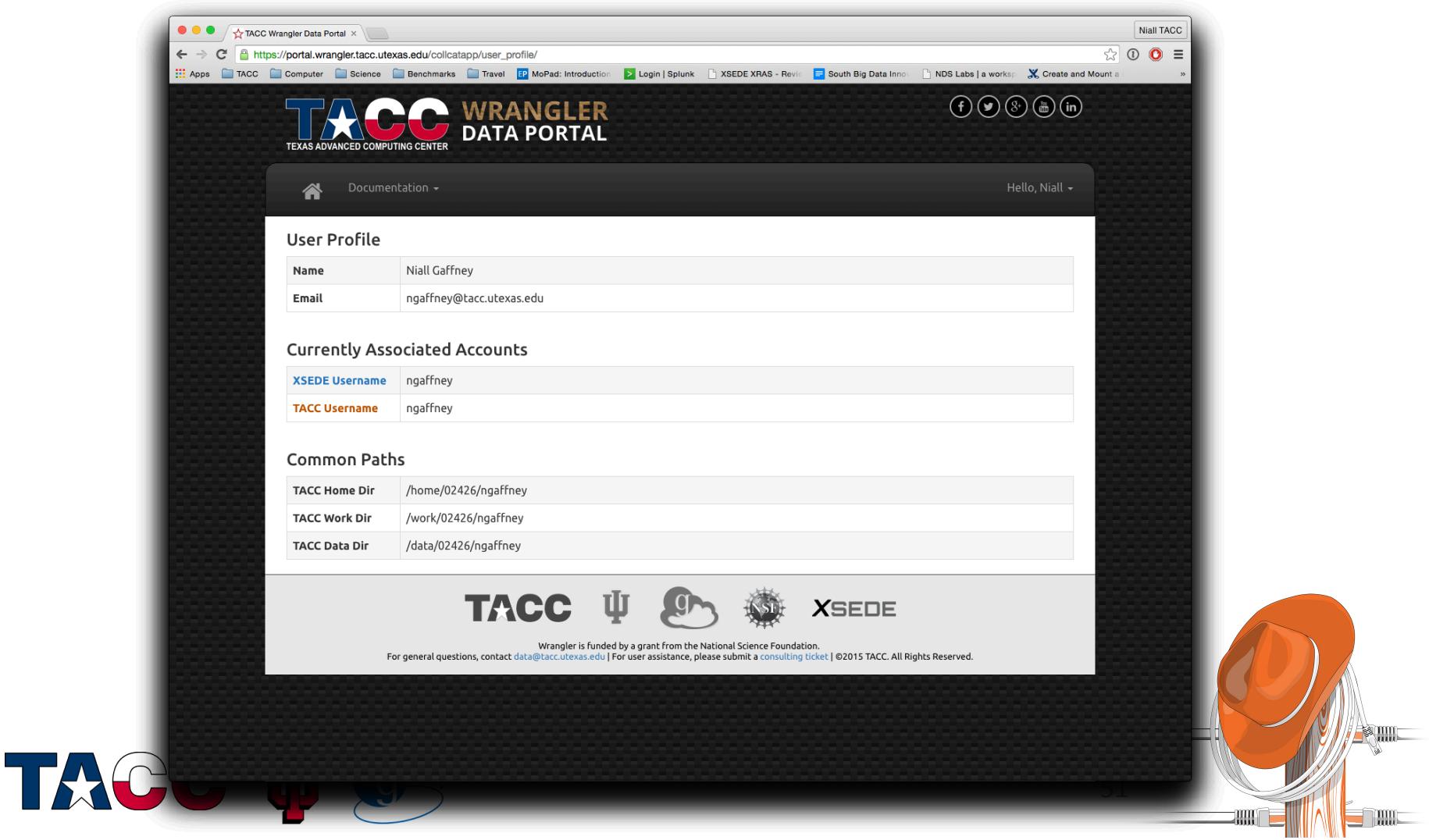
- To reduce data motion, promoting use of reservations on Wrangler
  - Can reserve a set of nodes for a specific purpose
    - Database, Workflow, Hadoop Cluster…
  - Each node allocation associated with 4 TB of Flash storage for the duration of the reservation



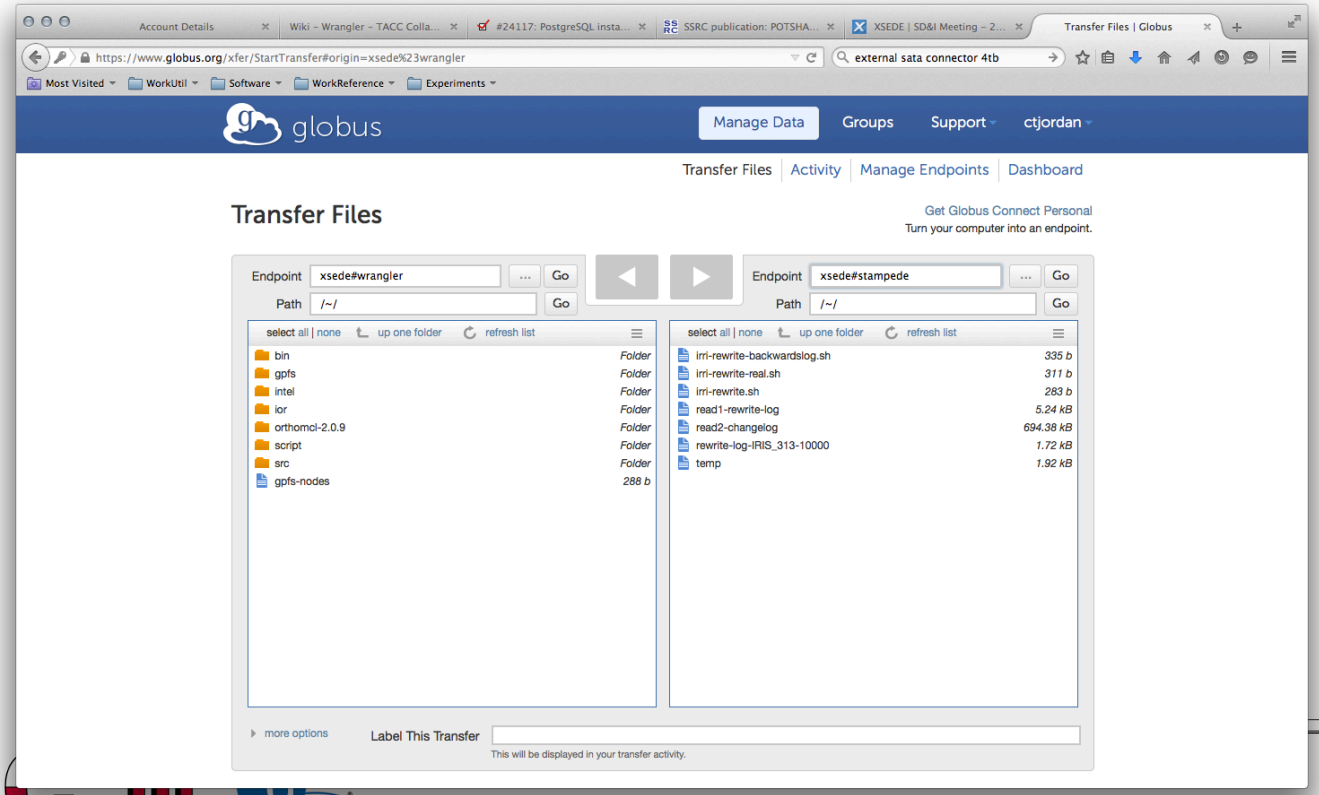
Wrangler Portal

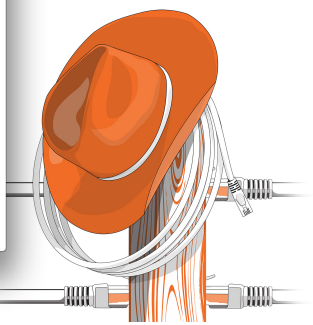
- Wrangler Portal is online...



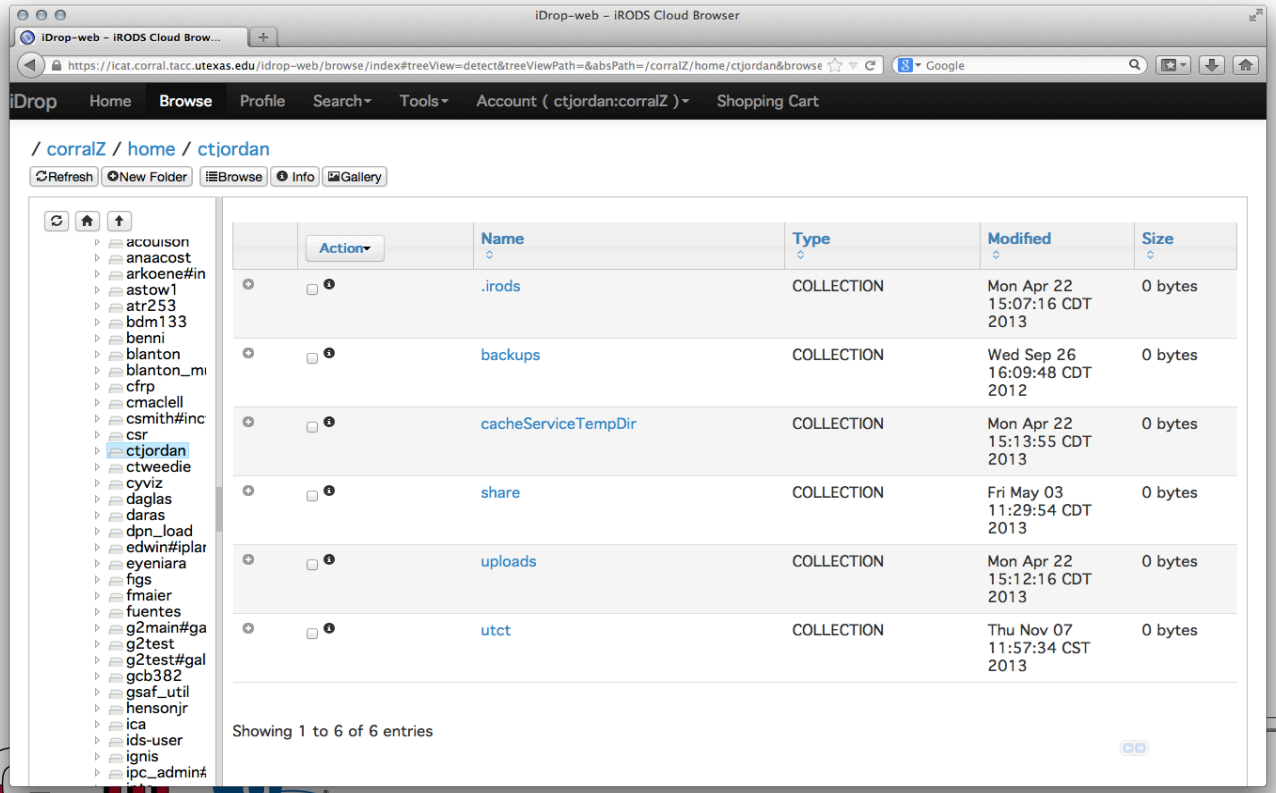

TACC Wrangler Data Portal
Niall TACC
https://portal.wrangler.tacc.utexas.edu/collcatapp/user_profile/
TACC
TEXAS ADVANCED COMPUTING CENTER
WRANGLER
DATA PORTAL
Documentation
Hello, Niall
User Profile
Name
Niall Gaffney
Email
ngaffney@tacc.utexas.edu
Currently Associated Accounts
XSEDE Username
ngaffney
TACC Username
ngaffney
Common Paths
TACC Home Dir
/home/02426/ngaffney
TACC Work Dir
/work/02426/ngaffney
TACC Data Dir
/data/02426/ngaffney
TACC
XSEDE
Wrangler is funded by a grant from the National Science Foundation.
For general questions, contact data@tacc.utexas.edu | For user assistance, please submit a consulting ticket | ©2015 TACC. All Rights Reserved.

Globus Transfer



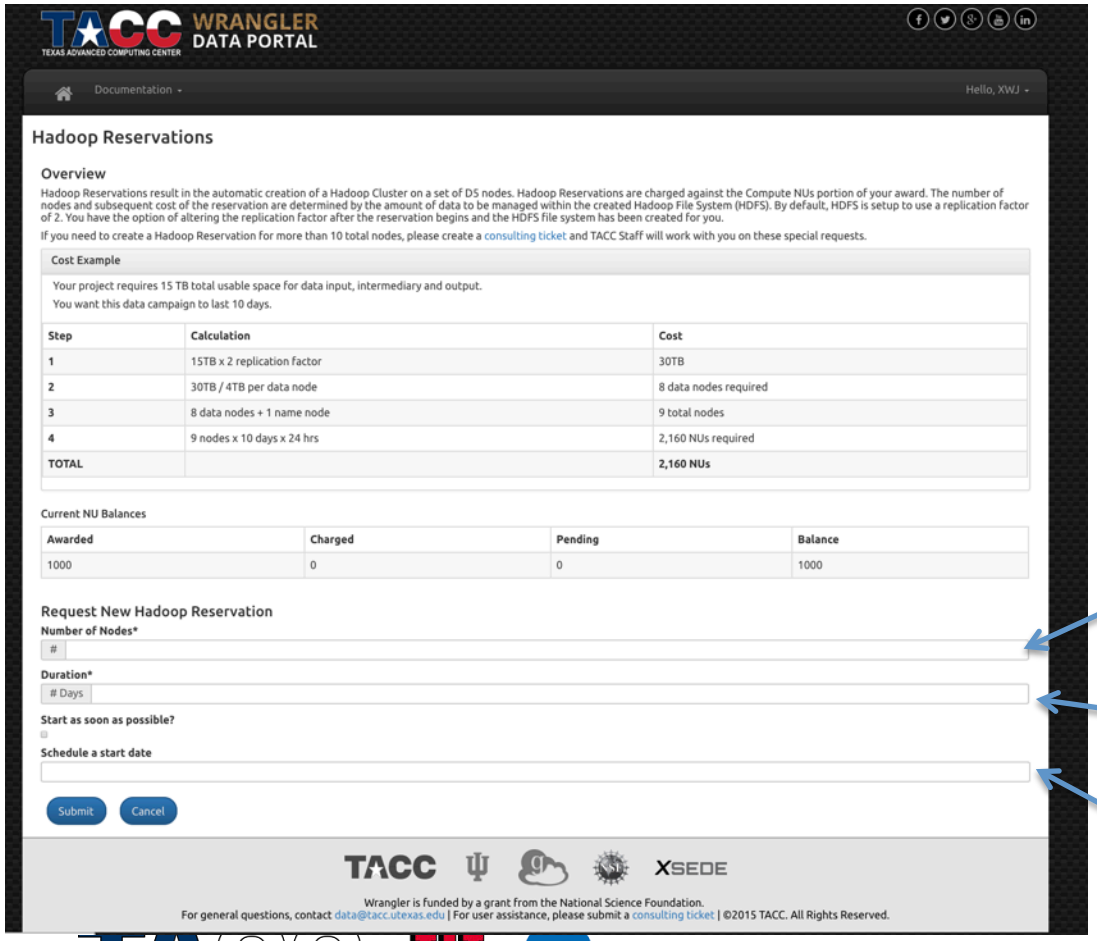


iRODS Web Interface



Easy Hadoop Setup

(still way too hard)

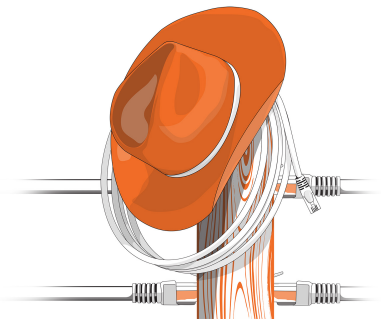


User input

the number of nodes to be used for the Hadoop cluster.

Duration

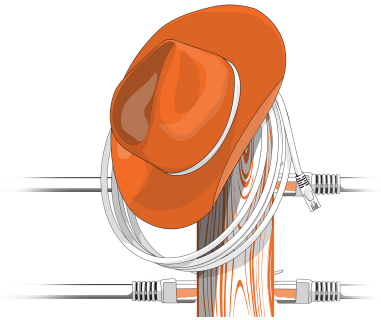
Start time



Check Hadoop Reservation

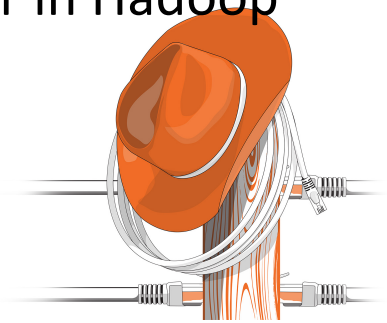
```
login1.wrangler(1)$ scontrol show reservation
ReservationName=hadoop-test8 StartTime=2015-04-22T12:11:17 EndTime=2015-04-25T12:11:17 Duration=3-00:00:00
   Nodes=c252-[101-115] NodeCnt=15 CoreCnt=720 Features=(null) PartitionName=D5_2 Flags=
   Users=xwj,root,rhuang,lprakash,gruan,tg828112 Accounts=(null) Licenses=(null) State=ACTIVE

login1.wrangler(2)$
```

- User can check the reservation status with `scontrol` command
  >scontrol show reservation
- The reservation will include all users from the user allocation
- One node in the reservation will be used for name node and other nodes will be used as data node
- The hadoop cluster will start with a set of default settings
  - User may override most settings such as duplication factor, block size at run time per application.
  - Hadoop cluster with specific settings upon request



Running Analysis with Hadoop Cluster

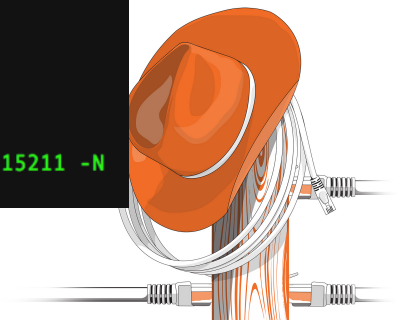
- User can access through slurm job:
  - VNC job: starts a vnc session on one of the node in Hadoop cluster
  - idev job: grant user console access to one of the node in Hadoop cluster
  - Batch job: submitting jobs to YARN resource manager in Hadoop cluster.

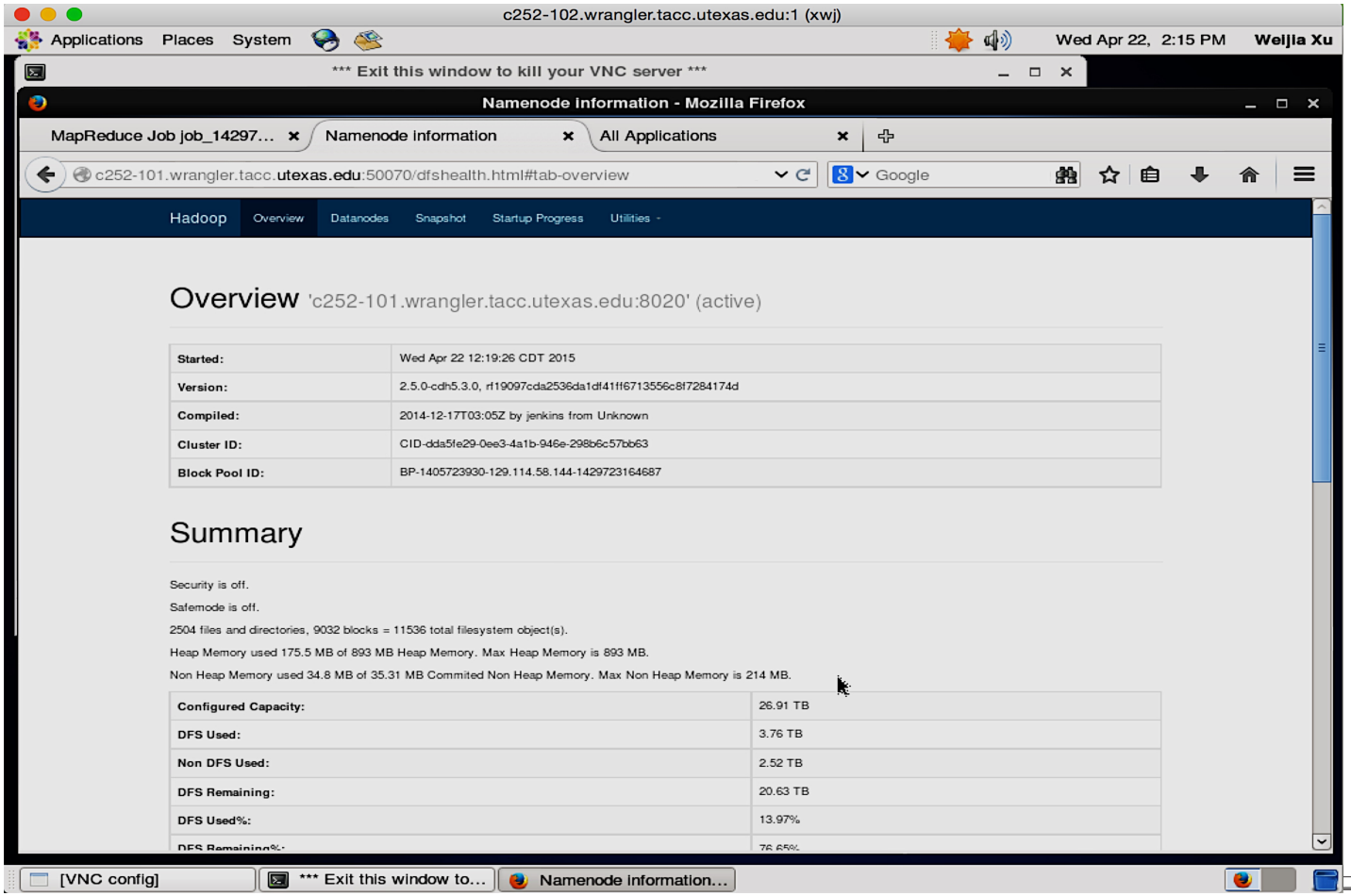


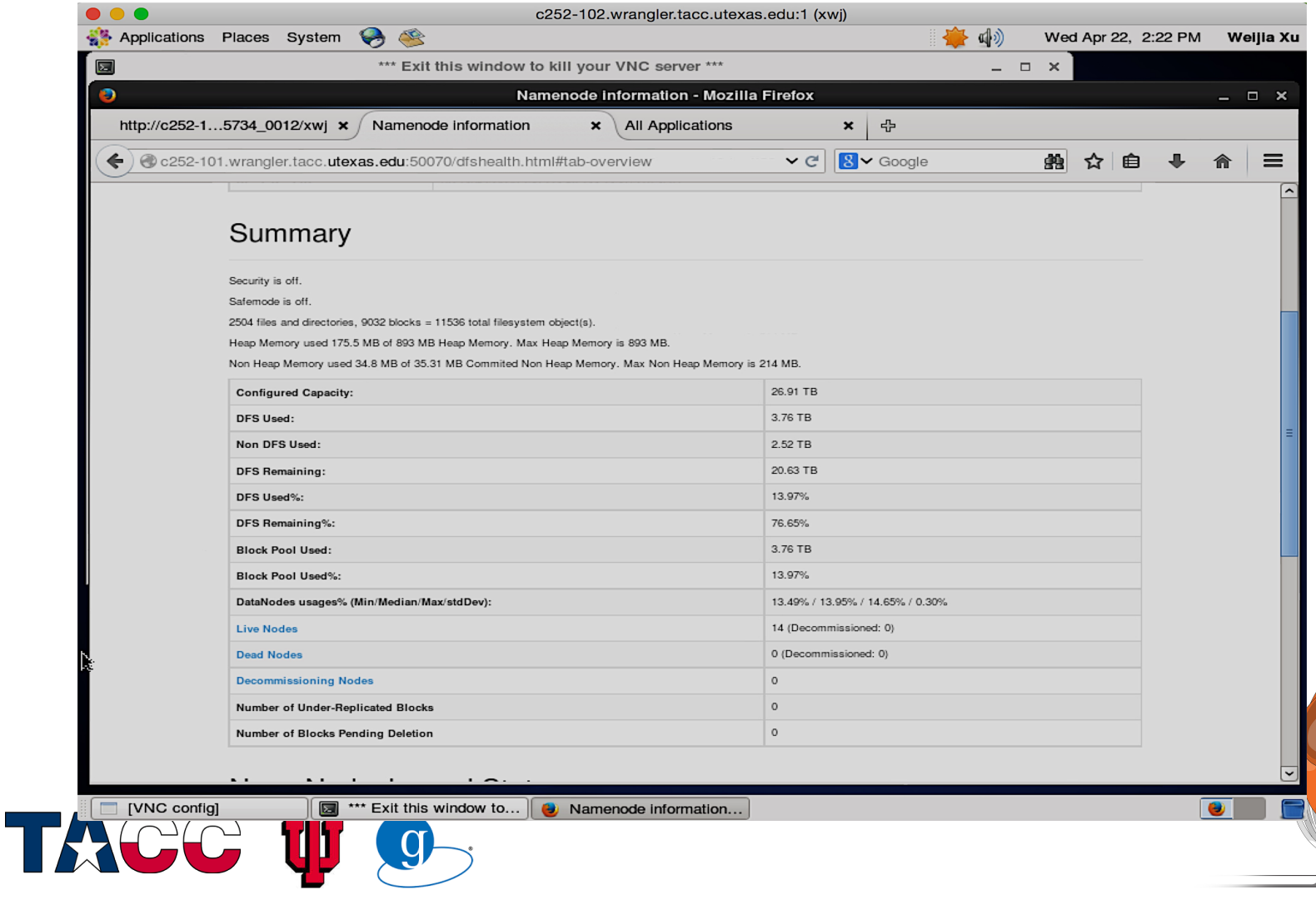
Checking Hadoop cluster status via web UI through VNC

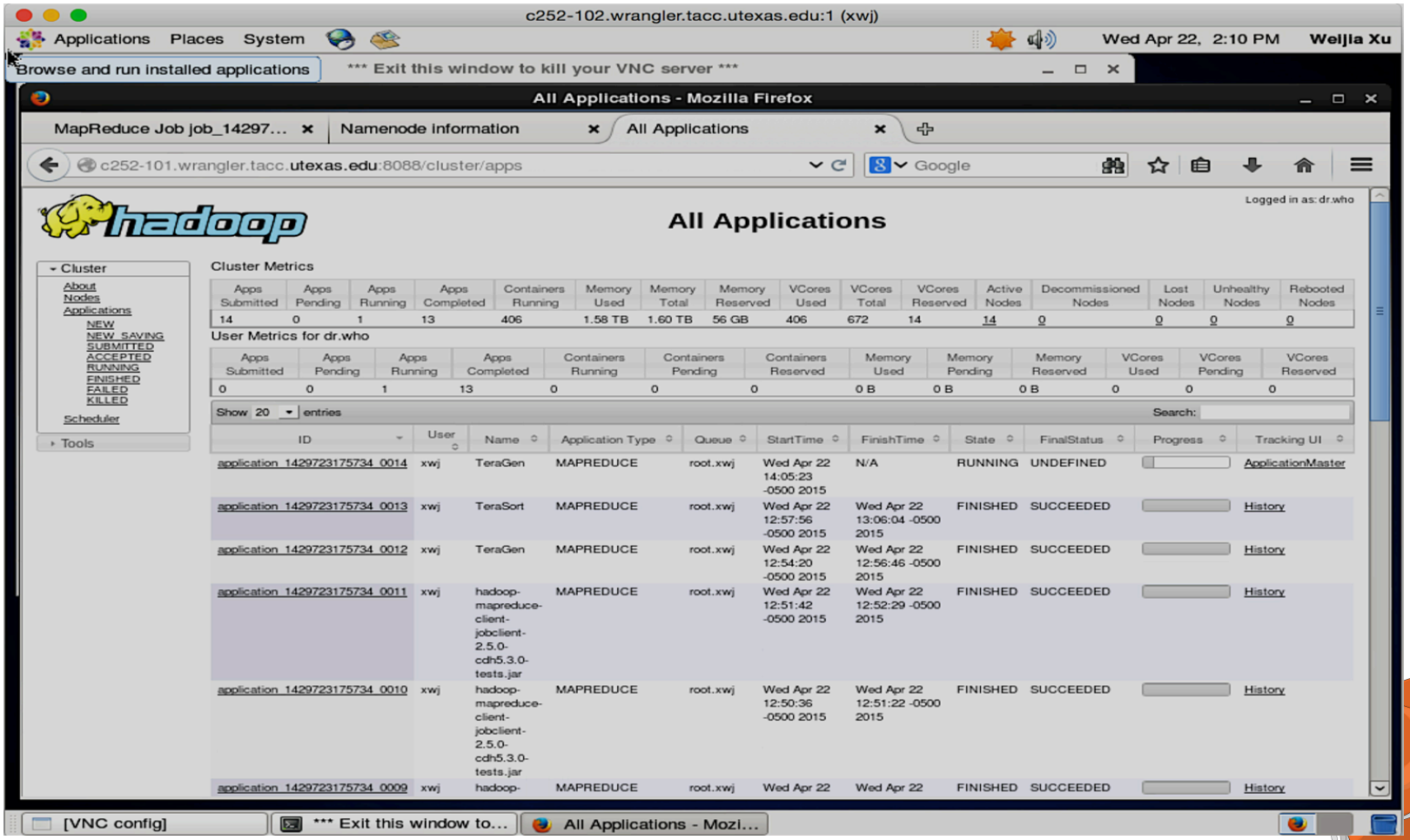
- Exemplar slurm job script for VNC session is available at
  - /share/doc/slurm/job.vnc
- Submitting vnc job to the existing Hadoop cluster e.g.
  - ***>sbatch –reservation=hadoop-test job.vnc***
- Following instruction in output file to connecting to VNC

```
login1.wrangler(8)$ tail vncserver.out
memory limit set to 125318847 kilobytes
set wayness to
got VNC display :1
local (compute node) VNC port is 5901
got login node VNC port 15211
Created reverse ports on wrangler logins
Your VNC server is now running!
To connect via VNC client:  SSH tunnel port 15211 to wrangler.tacc.utexas.edu:15211
                            i.e. ssh -f xwj@wrangler.tacc.utexas.edu -L 15211:wrangler.tacc.utexas.edu:15211 -N
                            Then connect to localhost:15211
```

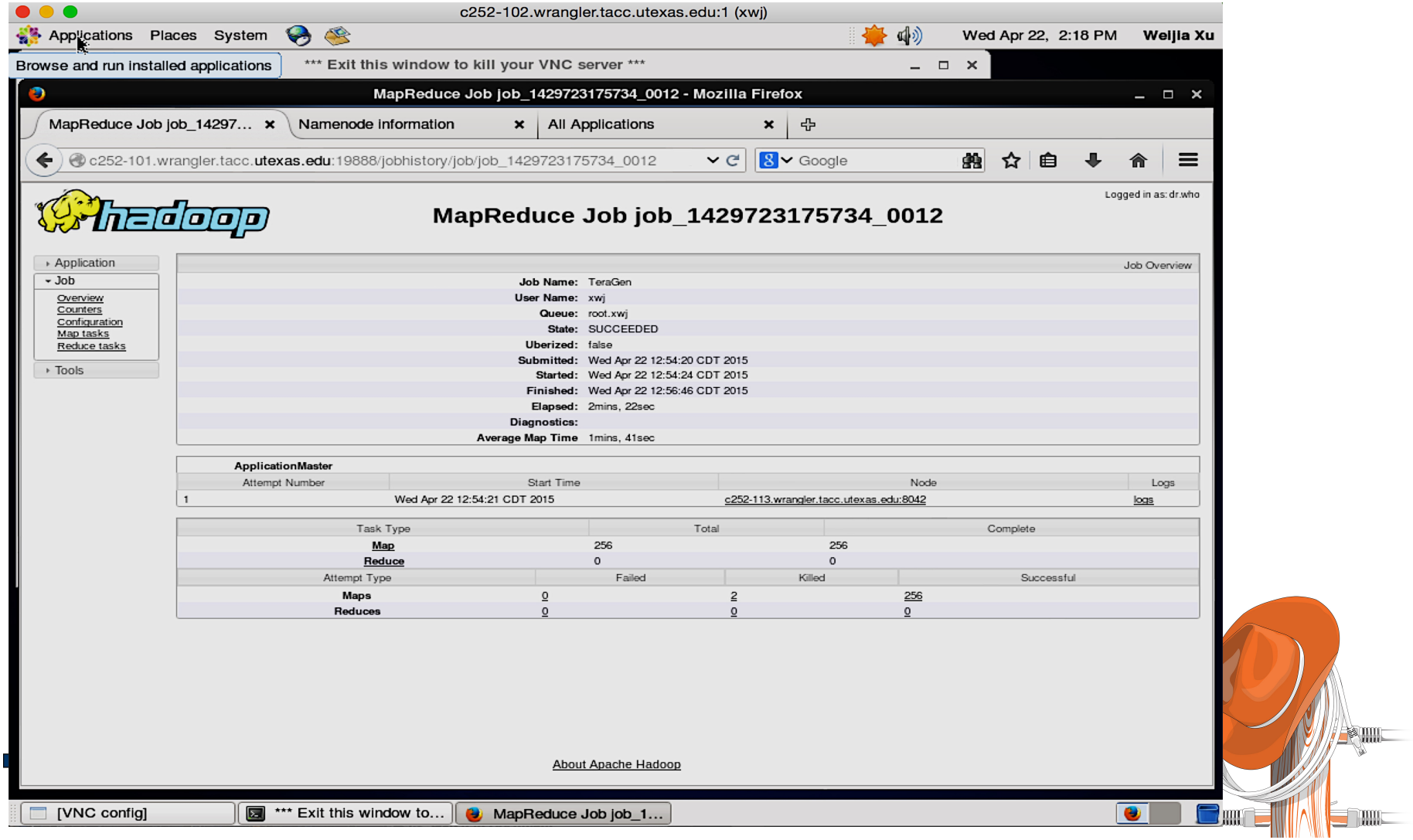


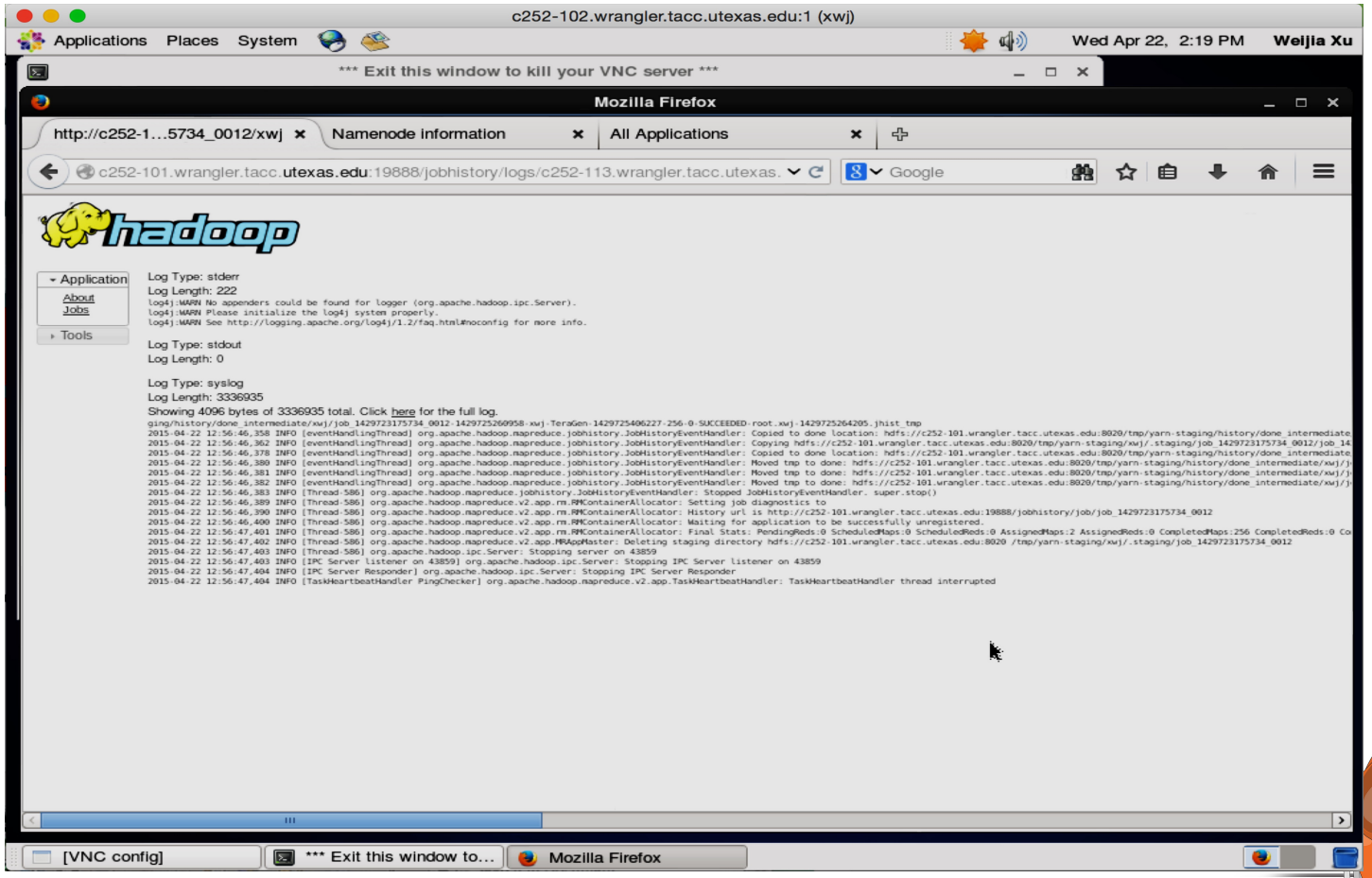

c252-102.wrangler.tacc.utexas.edu:1 (xwj)
Applications Places System
Wed Apr 22, 2:15 PM Weijia Xu
*** Exit this window to kill your VNC server ***
Namenode information - Mozilla Firefox
MapReduce Job job_14297...
Namenode information
All Applications
c252-101.wrangler.tacc.utexas.edu:50070/dfshealth.html#tab-overview
Google
Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities
Overview 'c252-101.wrangler.tacc.utexas.edu:8020' (active)
Started: Wed Apr 22 12:19:26 CDT 2015
Version: 2.5.0-cdh5.3.0, r19097cda2536da1df41ff6713556c8f7284174d
Compiled: 2014-12-17T03:05Z by jenkins from Unknown
Cluster ID: CID-dda5fe29-0ee3-4a1b-946e-298b6c57bb63
Block Pool ID: BP-1405723930-129.114.58.144-1429723164687
Summary
Security is off.
Safemode is off.
2504 files and directories, 9032 blocks = 11536 total filesystem object(s).
Heap Memory used 175.5 MB of 893 MB Heap Memory. Max Heap Memory is 893 MB.
Non Heap Memory used 34.8 MB of 35.31 MB Commited Non Heap Memory. Max Non Heap Memory is 214 MB.
Configured Capacity: 26.91 TB
DFS Used: 3.76 TB
Non DFS Used: 2.52 TB
DFS Remaining: 20.63 TB
DFS Used%: 13.97%
[VNC config]
*** Exit this window to...
Namenode information...


c252-102.wrangler.tacc.utexas.edu:1 (xwj)
Applications Places System
Wed Apr 22, 2:22 PM Weijia Xu
*** Exit this window to kill your VNC server ***
Namenode information - Mozilla Firefox
http://c252-1...5734_0012/xwj
Namenode information
All Applications
c252-101.wrangler.tacc.utexas.edu:50070/dfshealth.html#tab-overview
Google
Summary
Security is off.
Safemode is off.
2504 files and directories, 9032 blocks = 11536 total filesystem object(s).
Heap Memory used 175.5 MB of 893 MB Heap Memory. Max Heap Memory is 893 MB.
Non Heap Memory used 34.8 MB of 35.31 MB Commited Non Heap Memory. Max Non Heap Memory is 214 MB.
Configured Capacity: 26.91 TB
DFS Used: 3.76 TB
Non DFS Used: 2.52 TB
DFS Remaining: 20.63 TB
DFS Used%: 13.97%
DFS Remaining%: 76.65%
Block Pool Used: 3.76 TB
Block Pool Used%: 13.97%
DataNodes usages% (Min/Median/Max/stdDev): 13.49% / 13.95% / 14.65% / 0.30%
Live Nodes 14 (Decommissioned: 0)
Dead Nodes 0 (Decommissioned: 0)
Decommissioning Nodes 0
Number of Under-Replicated Blocks 0
Number of Blocks Pending Deletion 0
[VNC config]
*** Exit this window to...
Namenode information...


c252-102.wrangler.tacc.utexas.edu:1 (xwj)
Applications Places System
Wed Apr 22, 2:10 PM Weijia Xu
Browse and run installed applications
*** Exit this window to kill your VNC server ***
All Applications - Mozilla Firefox
MapReduce Job job_14297...
Namenode information
All Applications
c252-101.wrangler.tacc.utexas.edu:8088/cluster/apps
Google
Logged in as: dr.who
All Applications
Cluster
About
Nodes
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted | Apps Pending | Apps Running | Apps Completed | Containers Running | Memory Used | Memory Total | Memory Reserved | VCores Used | VCores Total | VCores Reserved | Active Nodes | Decommissioned Nodes | Lost Nodes | Unhealthy Nodes | Rebooted Nodes
14 | 0 | 1 | 13 | 406 | 1.58 TB | 1.60 TB | 56 GB | 406 | 672 | 14 | 14 | 0 | 0 | 0 | 0
User Metrics for dr.who
Apps Submitted | Apps Pending | Apps Running | Apps Completed | Containers Running | Containers Pending | Containers Reserved | Memory Used | Memory Pending | Memory Reserved | VCores Used | VCores Pending | VCores Reserved
0 | 0 | 1 | 13 | 0 | 0 | 0 | 0 B | 0 B | 0 B | 0 | 0 | 0
Show 20 entries
Search:
ID | User | Name | Application Type | Queue | StartTime | FinishTime | State | FinalStatus | Progress | Tracking UI
application_1429723175734_0014 | xwj | TeraGen | MAPREDUCE | root.xwj | Wed Apr 22 14:05:23 -0500 2015 | N/A | RUNNING | UNDEFINED | | ApplicationMaster
application_1429723175734_0013 | xwj | TeraSort | MAPREDUCE | root.xwj | Wed Apr 22 12:57:56 -0500 2015 | Wed Apr 22 13:06:04 -0500 2015 | FINISHED | SUCCEEDED | | History
application_1429723175734_0012 | xwj | TeraGen | MAPREDUCE | root.xwj | Wed Apr 22 12:54:20 -0500 2015 | Wed Apr 22 12:56:46 -0500 2015 | FINISHED | SUCCEEDED | | History
application_1429723175734_0011 | xwj | hadoop-mapreduce-client-jobclient-2.5.0-cdh5.3.0-tests.jar | MAPREDUCE | root.xwj | Wed Apr 22 12:51:42 -0500 2015 | Wed Apr 22 12:52:29 -0500 2015 | FINISHED | SUCCEEDED | | History
application_1429723175734_0010 | xwj | hadoop-mapreduce-client-jobclient-2.5.0-cdh5.3.0-tests.jar | MAPREDUCE | root.xwj | Wed Apr 22 12:50:36 -0500 2015 | Wed Apr 22 12:51:22 -0500 2015 | FINISHED | SUCCEEDED | | History
application_1429723175734_0009 | xwj | hadoop- | MAPREDUCE | root.xwj | Wed Apr 22 | Wed Apr 22 | FINISHED | SUCCEEDED | | History
[VNC config]
*** Exit this window to...
All Applications - Mozi...

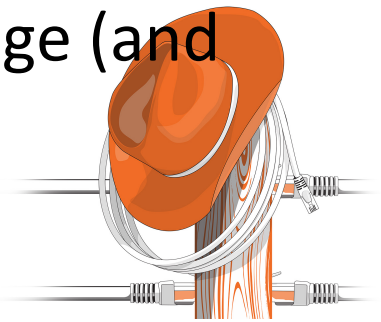
TACC


c252-102.wrangler.tacc.utexas.edu:1 (xwj)
Applications Places System
Wed Apr 22, 2:18 PM Weijia Xu
Browse and run installed applications
*** Exit this window to kill your VNC server ***
MapReduce Job job_1429723175734_0012 - Mozilla Firefox
MapReduce Job job_14297...
Namenode information
All Applications
c252-101.wrangler.tacc.utexas.edu:19888/jobhistory/job/job_1429723175734_0012
Google
Logged in as: dr.who
hadoop
MapReduce Job job_1429723175734_0012
Application
Job
Overview
Counters
Configuration
Map tasks
Reduce tasks
Tools
Job Overview
Job Name: TeraGen
User Name: xwj
Queue: root.xwj
State: SUCCEEDED
Uberized: false
Submitted: Wed Apr 22 12:54:20 CDT 2015
Started: Wed Apr 22 12:54:24 CDT 2015
Finished: Wed Apr 22 12:56:46 CDT 2015
Elapsed: 2mins, 22sec
Diagnostics:
Average Map Time 1mins, 41sec
ApplicationMaster
Attempt Number
Start Time
Node
Logs
1
Wed Apr 22 12:54:21 CDT 2015
c252-113.wrangler.tacc.utexas.edu:8042
logs
Task Type
Total
Complete
Map
256
256
Reduce
0
0
Attempt Type
Failed
Killed
Successful
Maps
0
2
256
Reduces
0
0
0
About Apache Hadoop
[VNC config]
*** Exit this window to...
MapReduce Job job_1...


c252-102.wrangler.tacc.utexas.edu:1 (xwj)
Applications Places System
Wed Apr 22, 2:19 PM Weijia Xu
*** Exit this window to kill your VNC server ***
Mozilla Firefox
http://c252-1...5734_0012/xwj
Namenode information
All Applications
c252-101.wrangler.tacc.utexas.edu:19888/jobhistory/logs/c252-113.wrangler.tacc.utexas.
Google
hadoop
Application
About
Jobs
Tools
Log Type: stderr
Log Length: 222
log4j:WARN No appenders could be found for logger (org.apache.hadoop.ipc.Server).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Log Type: stdout
Log Length: 0
Log Type: syslog
Log Length: 3336935
Showing 4096 bytes of 3336935 total. Click here for the full log.
[VNC config]
*** Exit this window to...
Mozilla Firefox

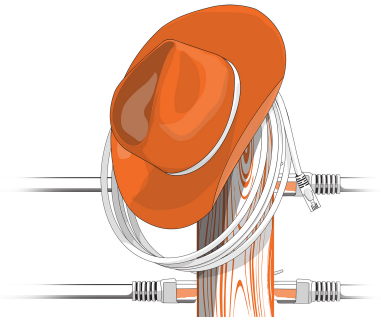
Still Need to Make This Easier

- In most ways, Wrangler is much friendlier than Stampede
- …but users aren't running
  - mpicc myjob.c
  - ./a.out
- …either.
- We have a tougher job in this community, as there are a lot of frameworks with a lot of tribal knowledge (and not our usual HPC Tribe!).



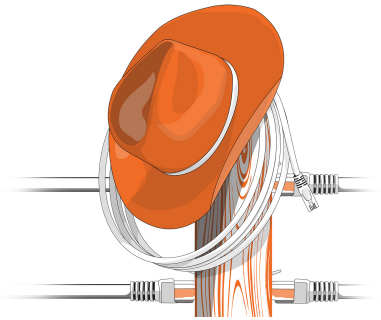
Supported Use Cases

- (In addition to traditional batch jobs)
- Hadoop
  - Pig
  - Mahout
  - Spark
- Databases -- Persistent and Transient
  - PostgreSQL
  - MariaDB/MySQL
  - MongoDB / NoSQL
  - PostGIS
- iRODs for data collections



Wrangler in the TACC Ecosystem

- TACC is traditionally a provider of HPC, Visualization, and storage systems.
  - And we still are.
- But these new communities provide kinds of data-intensive problems our HPC systems just aren't built for
  - Run Hadoop on your favorite supercomputer to see what we need.
  - Or do a bunch of random access to a bunch of really small files.
- Wrangler is not to replace our supercomputer, vis, or cloud offerings; it supplements this environment.



Thank You!

Dan Stanzione

dan@tacc.utexas.edu