



# MY LAPTOP TAKES FOREVER, NOW WHAT!

ERIC ADAMS  
PROGRAM MANAGER, RESEARCH COMPUTING SERVICES  
EWA@PURDUE.EDU

**PURDUE**  
UNIVERSITY.



# DATA WORKBENCH

# Outline

## Today's Topics

- RCAC Demographics
- Research Data
- Data Building Blocks
- Data Workbench
- The Ideal User
- Future Work



# Community Cluster Program

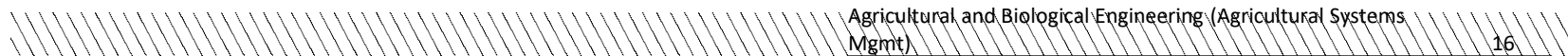


**303M hours delivered in 2017**  
**Almost 189 investors from 36 departments, from every College, and 3**  
**Purdue campuses**

The gold standard for condo-style computing  
 Today, the program is part of departments' faculty recruiting process.

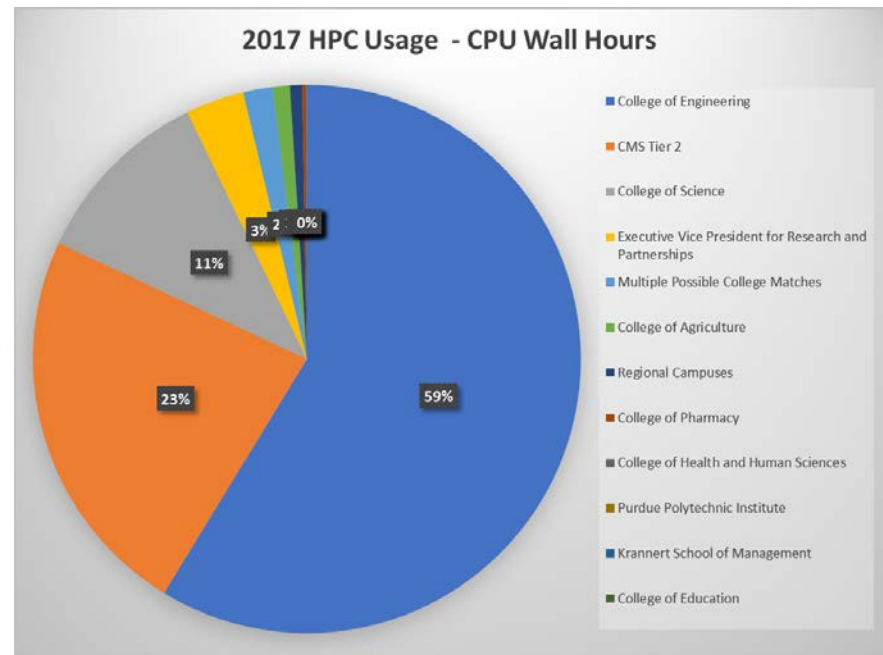
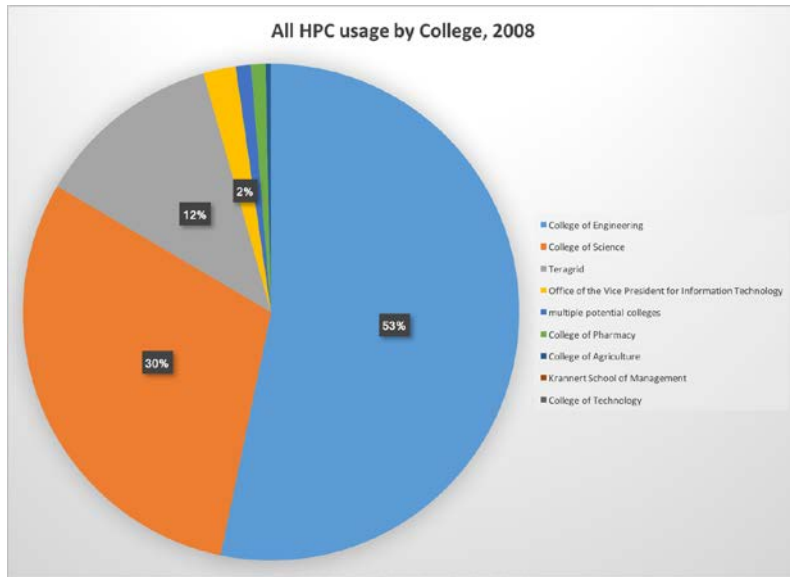
***A selling point to attract people to Purdue!***

Department	Cores
Aeronautics and Astronautics	5740
Mechanical Engineering	5556
CMS Tier2	5440
Electrical and Computer Engineering	4344
Earth, Atmospheric, and Planetary Sciences	2540
Materials Engineering	2064
Nuclear Engineering	1564
Other College of Engineering	980
Chemistry	824
Physics and Astronomy	820
Biomedical Engineering	640
Other Executive Vice President for Research and Partnerships	600
Statistics	512
Chemical Engineering	424
Agricultural and Biological Engineering (Biological Engineering)	368
Biological Sciences	356
Industrial Engineering	296
Civil Engineering	276
Computer and Information Technology	248
Medicinal Chemistry and Molecular Pharmacology	248
Mathematics	232
Bioinformatics Core	200
Agronomy	180
ITaP	176
Computer Science	156
Horticulture and Landscape Architecture	156
Cancer Center	96
Computational Biology	96
Biochemistry	40
Botany and Plant Pathology	40
Industrial and Physical Pharmacy	40
Brian Lamb School of Communication	32
Agricultural Economics	20
Animal Sciences	20
Food Science	20
Health Sciences	20
Other College of Pharmacy	20
Agricultural and Biological Engineering (Agricultural Systems Mgmt)	16



# Batch Computing Community

2008 vs 2017



Engineering – 5.1x growth  
 Ag – 21x growth  
 Science – “only” 3.1x growth

# Growth of the Community

## From more than Engineering and Science

### PIs:

College	2014	2018	Growth 2014-2017	Growth Rate 2014-18
Agriculture	48	202	154	4.2x
Engineering	161	387	226	2.4x
Science	199	308	109	1.55x
Education	1	12	11	12x
Liberal Arts	1	15	14	15x
Management	20	49	29	2.45x
Pharmacy	5	18	13	3.6x
Polytechnic	13	34	21	2.62x
Health and Human Sciences	14	46	32	3.29x
Veterinary Medicine	0	15	15	

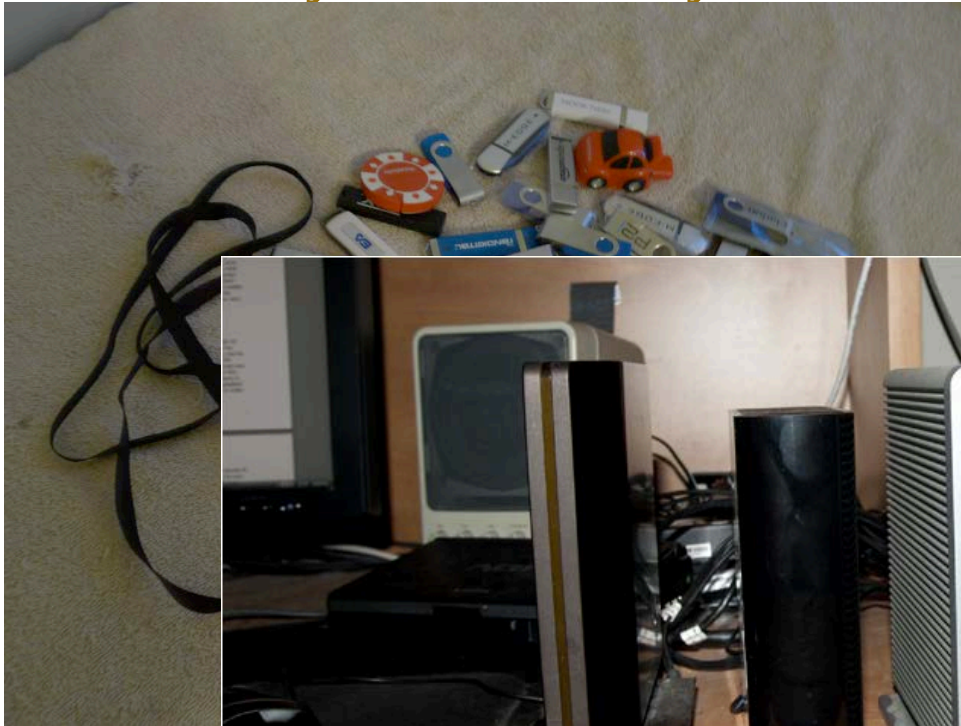


# RESEARCH DATA



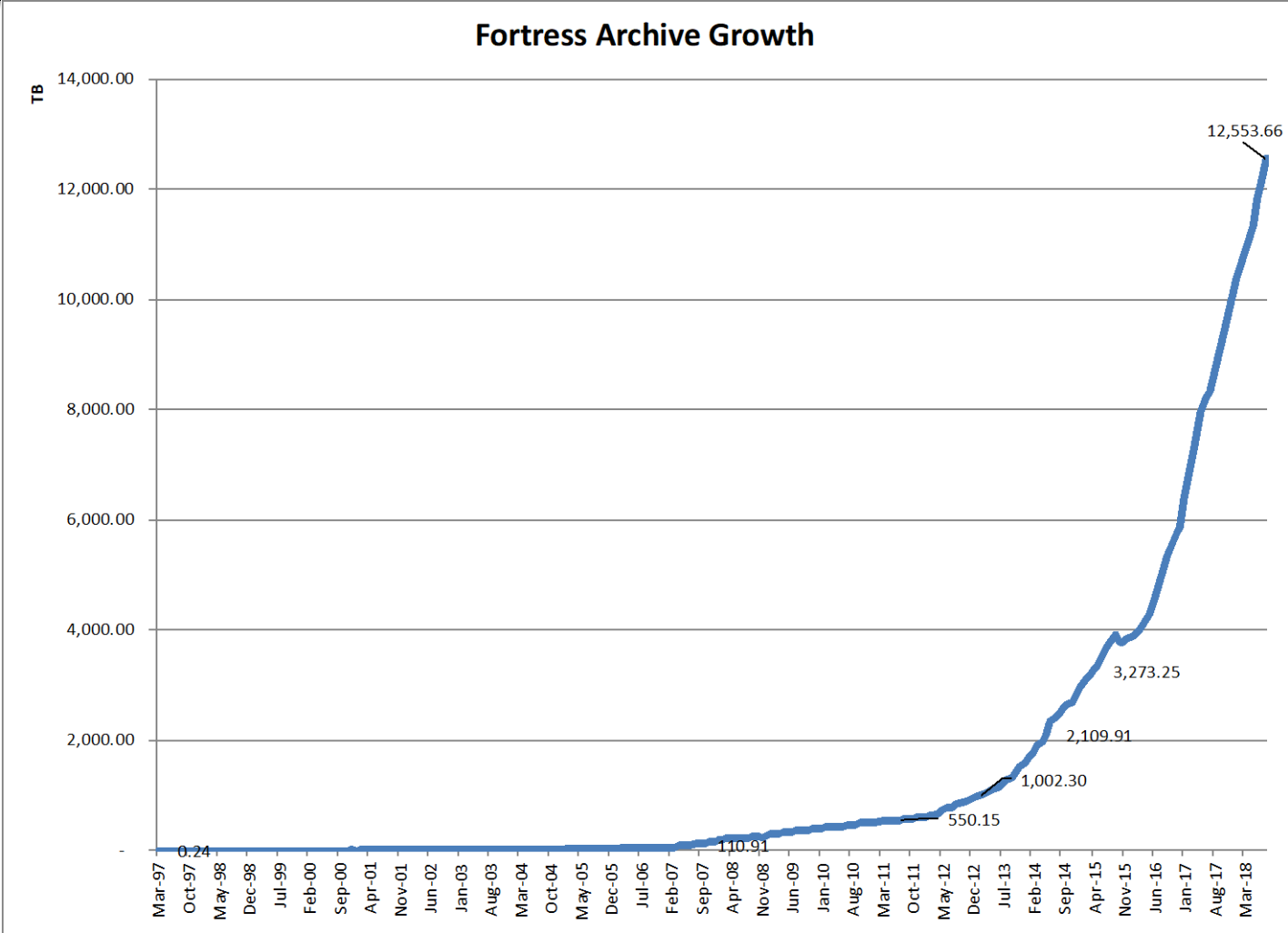
# On-Campus Research Storage Solutions

Left to fend for themselves,  
faculty will find a way





# Growth in Research Data Storage



# Research Data Depot



## Self-Service Lab Storage

- Built in 2014 – lab-oriented, high-performance, research data service for active data.
- \$75/TB per year
- Hardware refresh is in procurement now – adding approx. **8 PB of disk!**
- 490 labs are Depot investors
  - 60% are not HPC users!

*New: Cloud storage offering for research data being kicked off soon!*

# Research Data Depot



## Some Realizations

- Designed the service for HPC, but clearly everybody has data
- Adoption by non-HPC is large!
- Depot is now a key capability for research data.
- Data Science Initiative will require computing for all

## Some Questions

- Where does that non-HPC 60% analyze data?

# Data Science Initiative

- Purdue is launching an ambitious Integrative Data Science Initiative that will be driven by robust interdisciplinary collaboration from colleges across the campus
- Build on and advance Purdue's existing strengths to position the University as a leader at the forefront of advancing data science-enabled research and education, by tightly coupling theory, discovery and applications while providing students with an integrated, data science-fluent campus ecosystem.

# What Computing do all Researchers Do?

On whatever platform they use

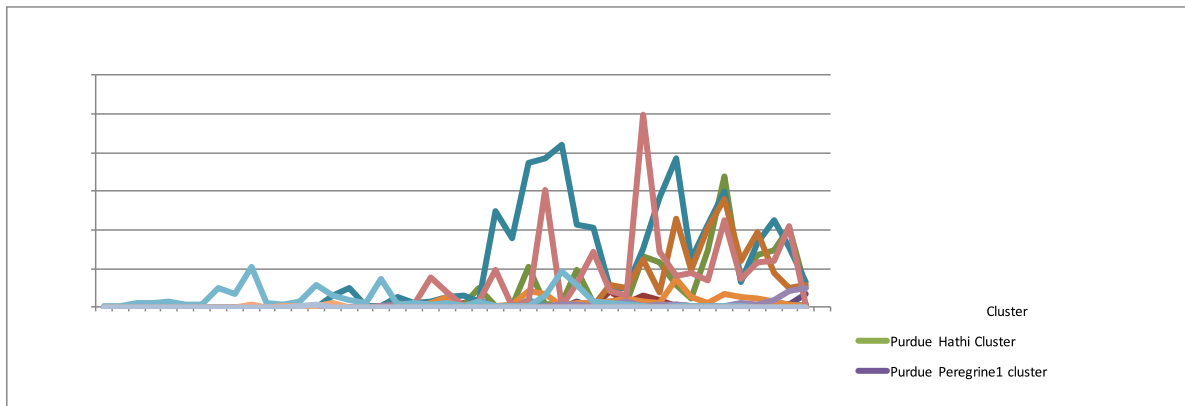
- Collect or simulate data
- Store data
- Analyze data
- Transfer and move data
- Share data
- Publish data

# Data Building Blocks

## On Community Clusters and Scholar

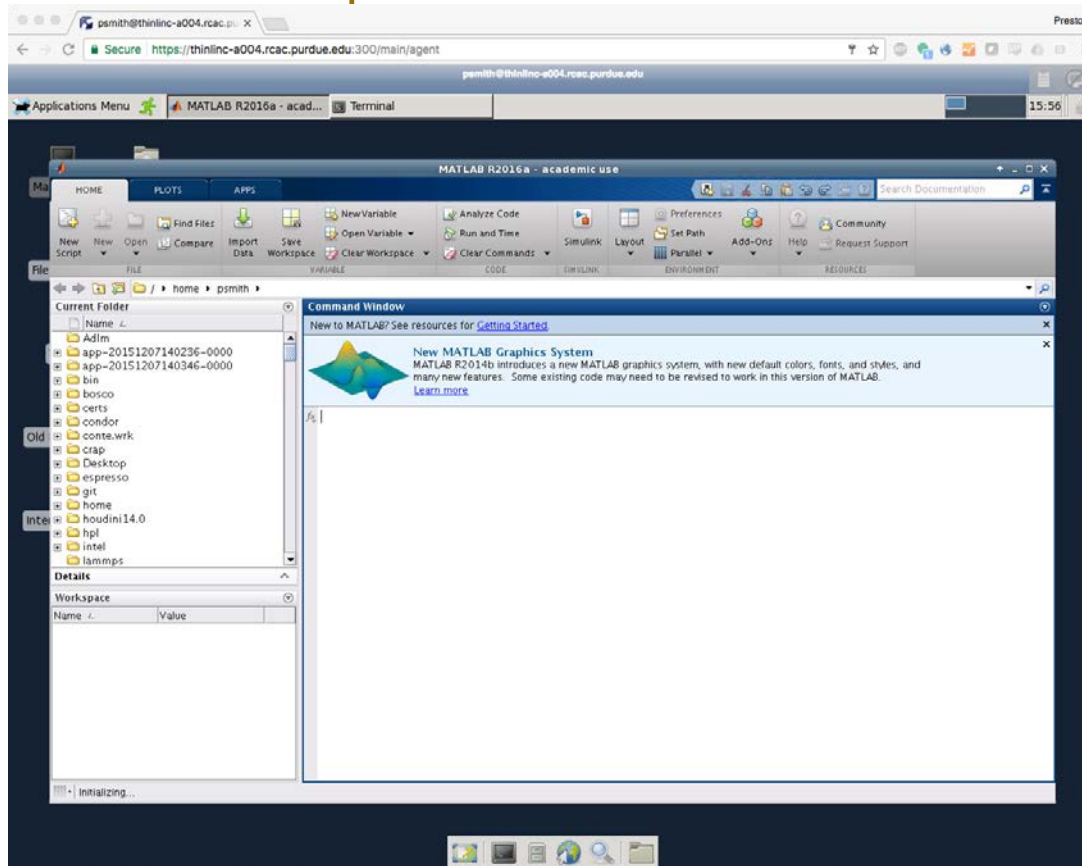


- Calendar Year 2017:
  - 2.3PB moved
  - Avg 194 TB, and 100 users/mo



# Data Building Blocks

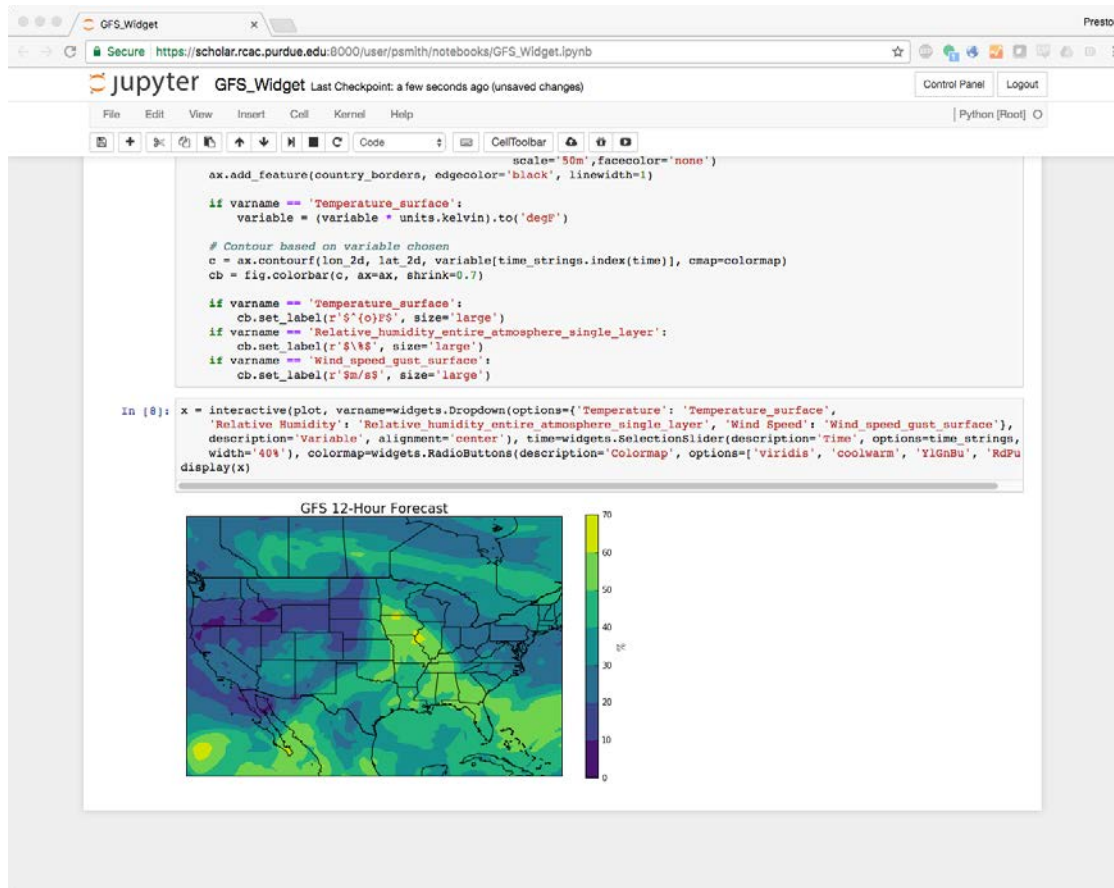
## Remote Desktops - Thinlinc





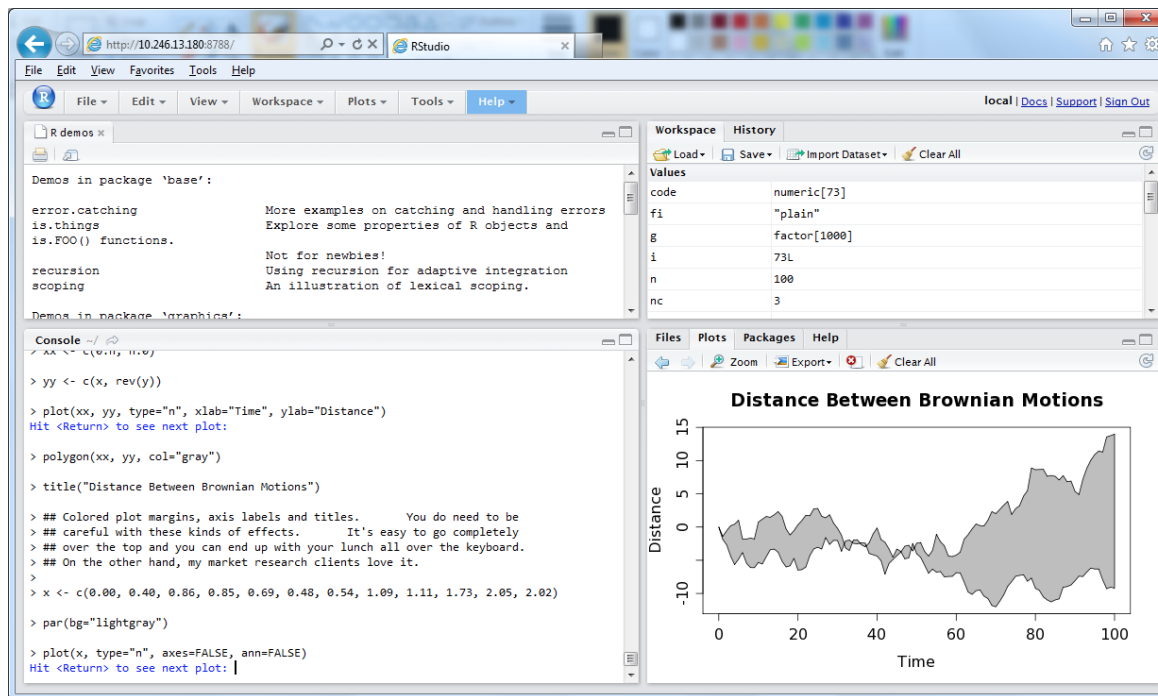
# Data Building Blocks

## Jupyter Notebooks



# Data Building Blocks

## Rstudio in the Browser



The screenshot shows the RStudio interface in a browser window. The browser address bar displays `http://10.246.13.180:8788/`. The RStudio interface includes a menu bar (File, Edit, View, Favorites, Tools, Help), a toolbar, and a workspace area. The console window shows the following R code:

```
Demos in package 'base':  
error.catching      More examples on catching and handling errors  
is.things            Explore some properties of R objects and  
is.FOO() functions.  
  
recursion            Not for newbies!  
scoping              Using recursion for adaptive integration  
                     An illustration of lexical scoping.  
  
Demos in package 'graphics':  
  
Console  
/ xx <- c(0:n, n:0)  
> yy <- c(x, rev(yy))  
> plot(xx, yy, type="n", xlab="Time", ylab="Distance")  
Hit <Return> to see next plot:  
> polygon(xx, yy, col="gray")  
> title("Distance Between Brownian Motions")  
> ## Colored plot margins, axis labels and titles.      You do need to be  
> ## careful with these kinds of effects.              It's easy to go completely  
> ## over the top and you can end up with your lunch all over the keyboard.  
> ## On the other hand, my market research clients love it.  
>  
> x <- c(0.00, 0.40, 0.86, 0.85, 0.69, 0.48, 0.54, 1.09, 1.11, 1.73, 2.05, 2.02)  
> par(bg="lightgray")  
> plot(x, type="n", axes=FALSE, ann=FALSE)  
Hit <Return> to see next plot: |
```

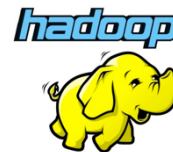
The workspace area shows the following variables:

Variable	Value
code	numeric[73]
fi	"plain"
g	factor[1000]
i	73L
n	100
nc	3

The plot window displays a line graph titled "Distance Between Brownian Motions". The x-axis is labeled "Time" and ranges from 0 to 100. The y-axis is labeled "Distance" and ranges from -10 to 15. The plot shows a jagged, gray-filled area representing the distance between two Brownian motions over time.

# Data Building Blocks

## Miscellaneous Capabilities



# Data Building Blocks

What's Missing – All being in one Package!



# Data Workbench

## Service

- Easy access to data analysis tools
- Run virtual private Windows desktops
  - e.g. for GIS software*
- Run virtual private Linux images in containers
- Integrate with RCAC services
  - Depot, Fortress, Globus, Github, Self Service, etc.
- Grow to batch HPC as your needs grow!



# DEMOS

Workbench, Data Depot and Globus

# The Ideal Data Workbench User

## Who should consider it

- Doing Data Analytics, statistical software
- Wish to have private, custom Windows environments
- Want top-end access to Depot Storage or Fortress

***Anybody who says: “My laptop is no longer powerful enough to process my data, but community clusters/high performance computing is more than I need”.***



# The Ideal Data Workbench User

## What is not a good fit?

- Applications with very intensive I/O requirements
- Applications with very large memory requirements
- Applications requiring specialized hardware (GPUs)
- Bioinformatics
- Parallel computing (MPI)
- Batch computing

# Data Workbench Growth Track

## Laptop to Workbench to HPC

- Easily scale and evolve from workbench to community clusters
- Same user experience as community clusters
- ***Are you Community Cluster partner already?***
  - Workbench's key features are available on community clusters as well!

# Data Workbench Service

Pricing?

- **\$300 annual charge per lab**

# Geospatial Tools

## Packages available from RCAC

- Esri ArcGIS
- ERDAS IMAGINE
- Harris ENVI
- Trimble eCognition
- Purdue MultiSpec

# Windows on HPC Top 5

Darrell Schulze Lab - <https://soilexplorer.net/>

Calls	Alias	Name	Dept	Queue
192	dschulze	Darrell Schulze	Agronomy	dschulze
96	dogas	Daelen Ogas	Agronomy	dschulze
76	screger	Stephen Creger	Chemical Engineering	jlaskin
74	erwyss	Erica Wyss	Agronomy	dschulze
52	sesquive	Silenze Silvae Esquivel Benjamin	Botany	Thook



# THANK YOU

[www.rcac.purdue.edu/compute/workbench](http://www.rcac.purdue.edu/compute/workbench)



**WE ARE PURDUE.** WHAT WE MAKE MOVES THE WORLD FORWARD.

**PURDUE**  
UNIVERSITY.