



Shared IT Services for Higher Education & Research

# Conference 2017

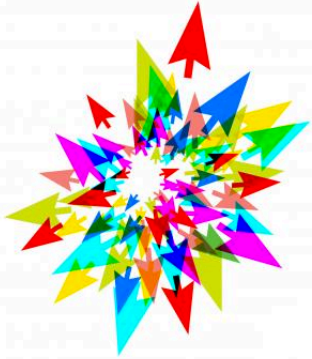


## Bringing the Thunder

Deploying Jupyter Notebooks for Research,  
Education and Innovation



Pacific Institute *for the*  
Mathematical Sciences



**compute** | **calcul**  
canada | canada

cybera

- Ian Allison
- Michael Lamoureux
- Dugan O'Neil
- Robin Winsor

# Outline

- Jupyter demo
- Technical overview
- The collaboration between Compute Canada, Cybera, and the Pacific Institute for the Mathematical Sciences
- Where are we trying to get to?
- Discussion

# Outline

- **Jupyter demo**
- Technical overview
- The collaboration between Compute Canada, Cybera, and the Pacific Institute for the Mathematical Sciences
- Where are we trying to get to?



# PIMS & SYZGY -- a conjunction of tools

- Formatted text
- Math formulas
- Code (Python and more)
- Data and data analysis
- Graphics
- Animations
- Slide show presentations

# SYZYG -- intro.syzygy.ca

Introduction to Syzygy



Introduction to Syzygy 0.0.8  
pimsmath/syzygy-intro

DOWNLOAD

★ STARS 1

Introduction

Getting Started

The Basic Elements

Python for Computing

Julia for Computing

R for Computing

Unix Tricks

Using Git and GitHub

Acknowledgements

email: [jupyter@pims.math.ca](mailto:jupyter@pims.math.ca)



# Syzygy -- Sample window

Dropbox/Code\_2016/Pyt... Embedded\_D3 ucalgary.syzygy.ca GridExperiments/ GridEigs

https://ucalgary.syzygy.ca/jupyter/user/04062091/notebooks/GridExperin 110% Search

Latest Headlines Apple Google Maps YouTube Wikipedia News Popular Sage Days 9 Work

Jupyter GridEigs Last Checkpoint: 02/16/2017 (autosaved) Control Panel Logout

File Edit View Insert Cell Kernel Widgets Help Julia 0.5.0

In [5]:

```
# We use a plotting package for Julia
using PyPlot
```

Slide Type Slide

**Dimension one.**

The second derivative operator is approximated by a central difference operator

$$D^2f(x) = \frac{-f(x - \Delta x) + 2f(x) - f(x + \Delta x)}{\Delta x^2}.$$

This operator is represented as a matrix with 2's on the diagonal, and -1 above and below.

In [6]:

```
# This code creates an n x x tridiagonal matrix
# It represents a finite difference operator for the second derivative
# We then compute the eigenvalues
```

Slide Type Slide

# Text -- Markdown format

- Easy to make:
  - Headings
  - Bold text
  - Italicized text
  - Lists and enumerations
- Simple commands like `__this__` or `## this` produces bold face, headers, etc.

# Math -- LaTeX format

- `$$ \int_0^1 \sin(2\pi x) dx $$` becomes

$$\int_0^1 \sin(2\pi x) dx = 0$$

- As Einstein once said: `$$ G_{\mu\nu} ... $$`

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

# Code -- Julia, Python, R, etc

To plot a surface, creat your x and y variables, expand to a grid, a

```
In [7]: n = 100
x = linspace(-3, 3, n)
y = linspace(-3, 3, n)

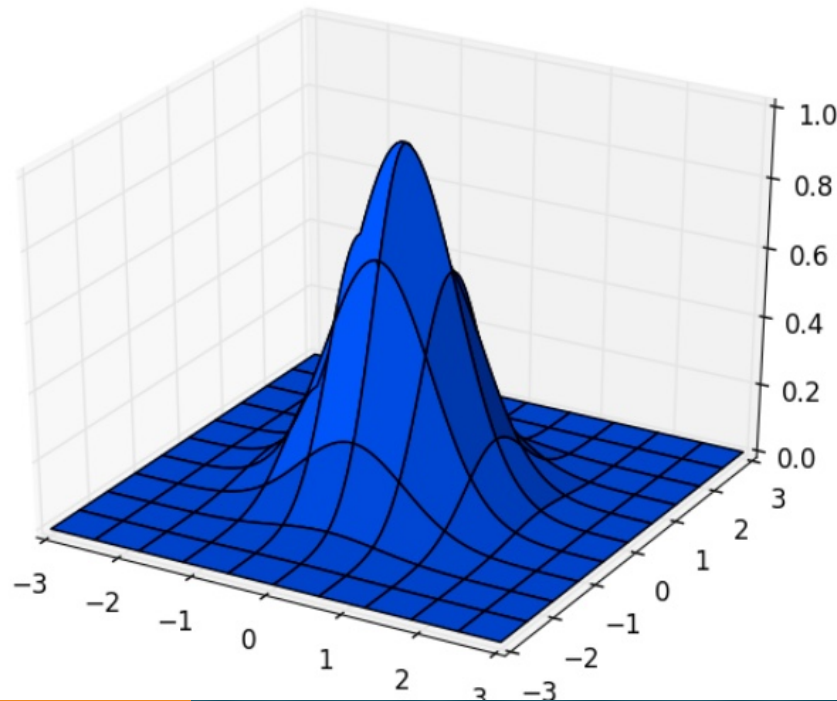
xgrid = repmat(x', n, 1)
ygrid = repmat(y, 1, n)

z = exp(-xgrid.^2 - ygrid.^2)

plot_surface(x, y, z)
```

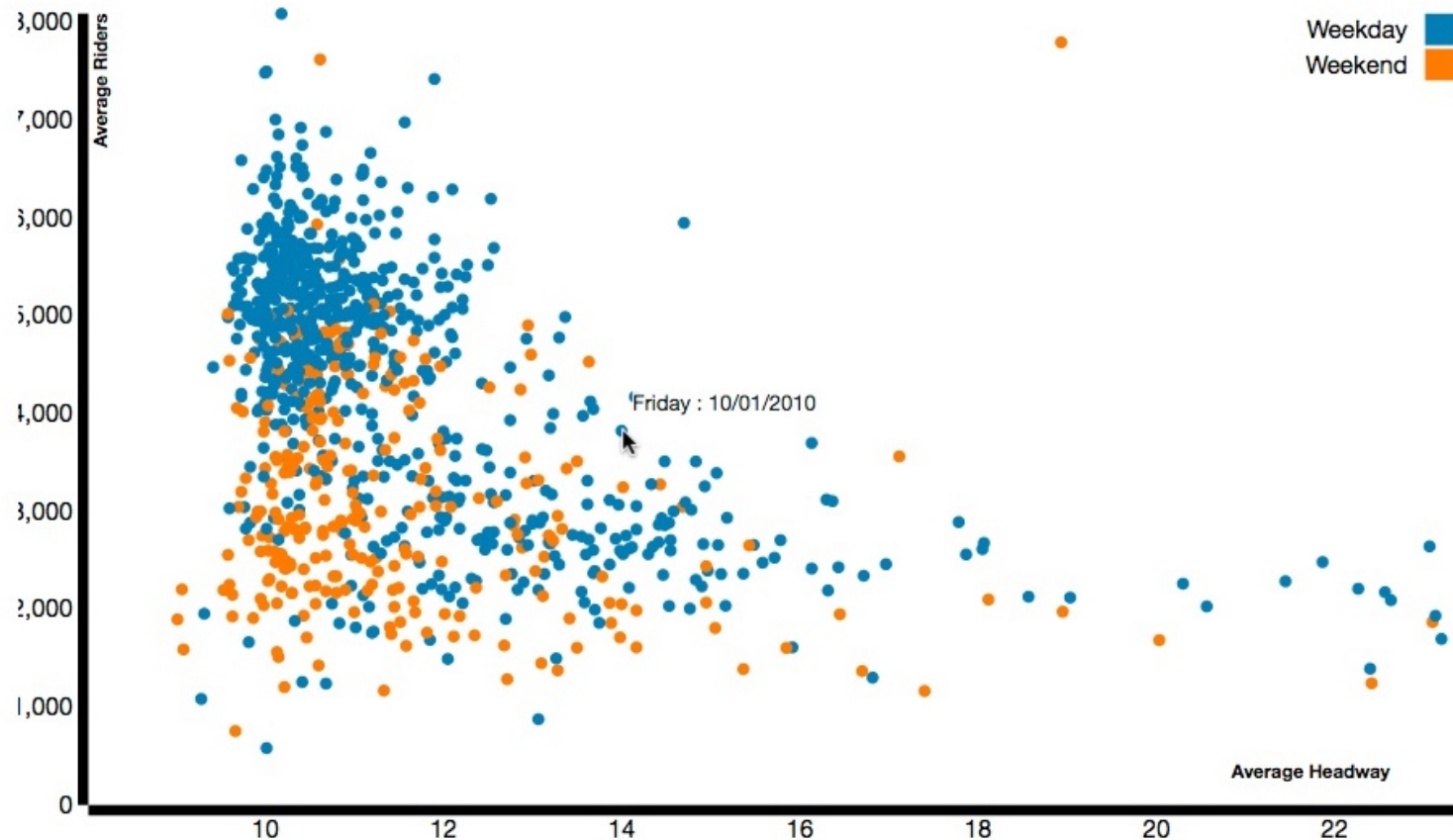
# Code -- Julia, Python, R, etc

```
In [7]: n = 100  
x = linspace(-3, 3, n)  
y = linspace(-3, 3, n)  
  
xgrid = repmat(x', n, 1)  
ygrid = repmat(y, 1, n)  
  
z = exp(-xgrid.^2 - ygrid.^2)  
  
plot_surface(x, y, z)
```



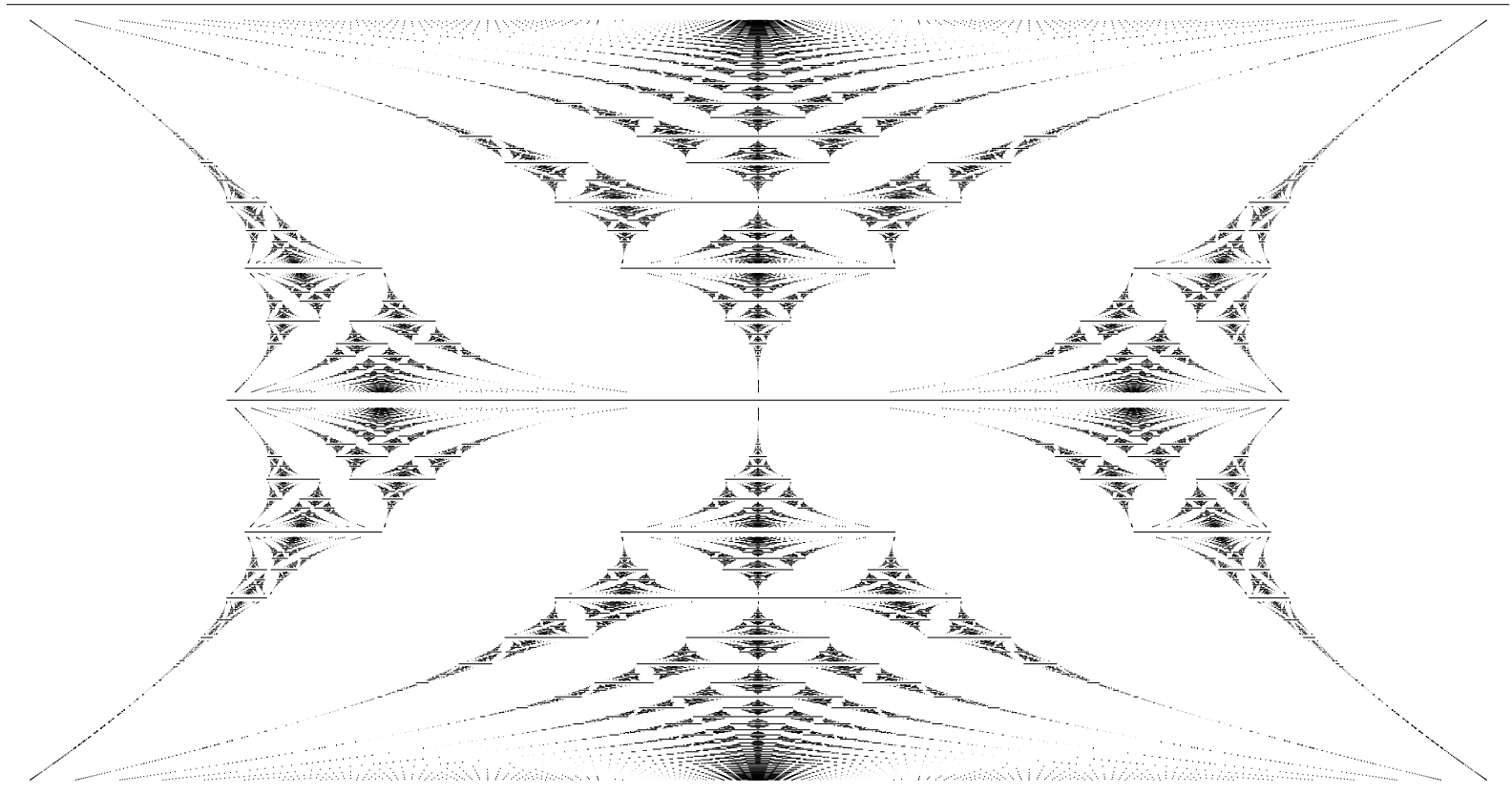


# Data -- Interactive D3 chart

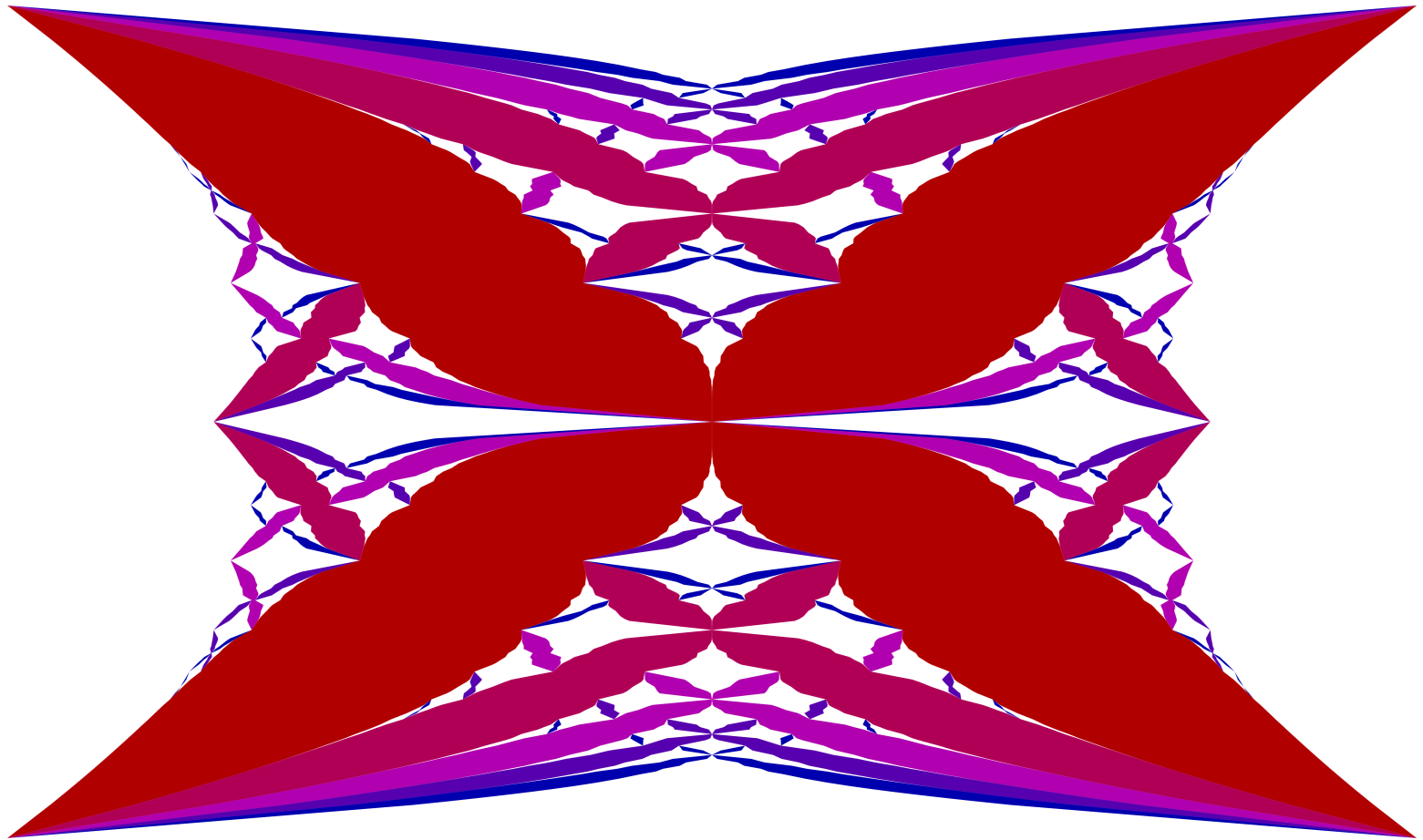




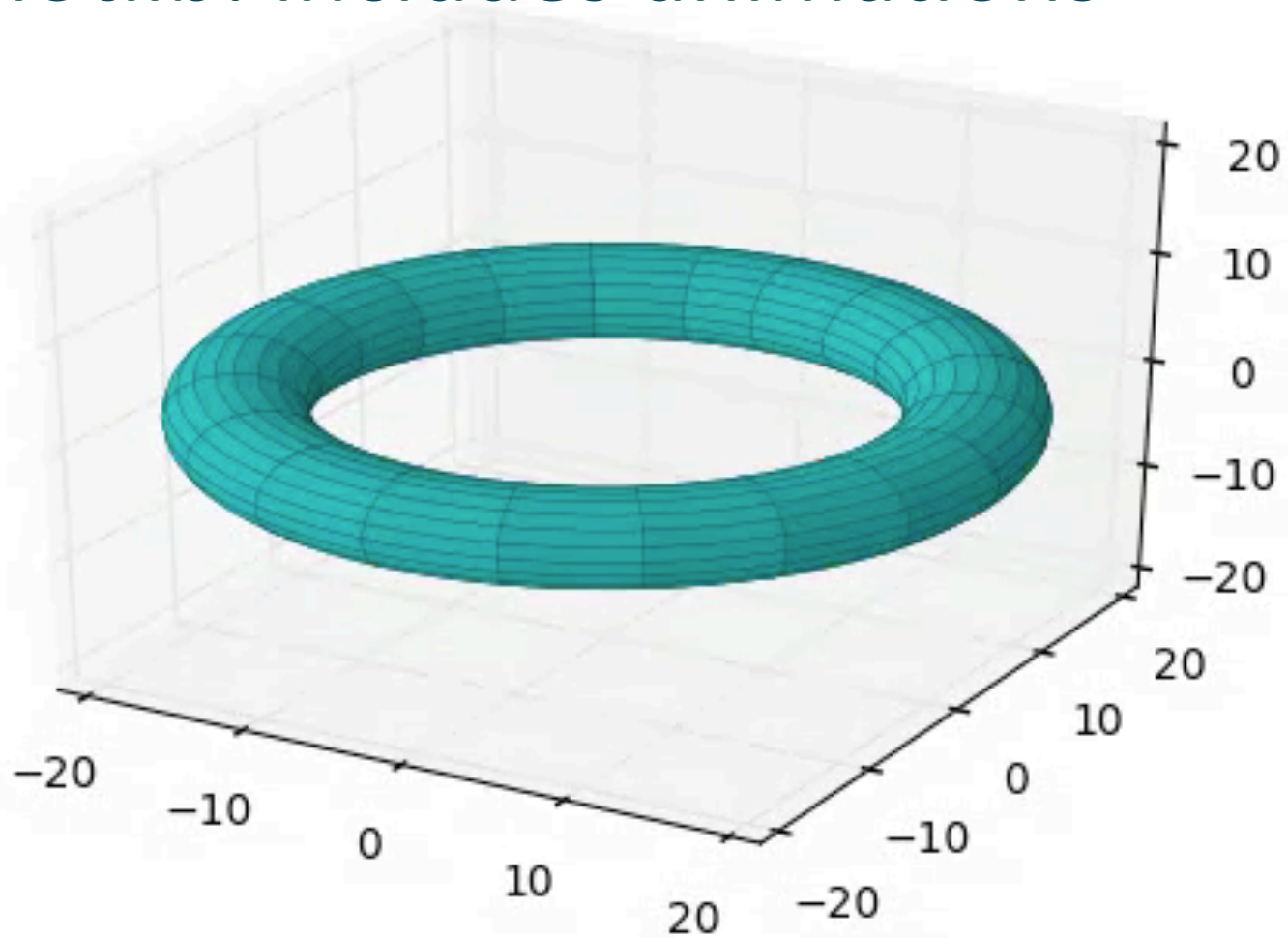
# Graphics -- Electron spectra



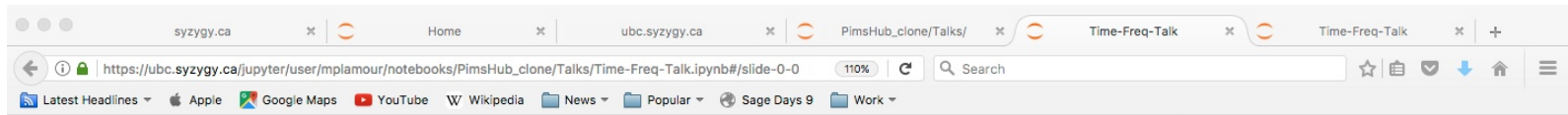
# Graphics -- Mathieu Butterfly



# Matplotlib: includes animations



# Slide Shows. Title page



## Time-Frequency Methods in Analysis

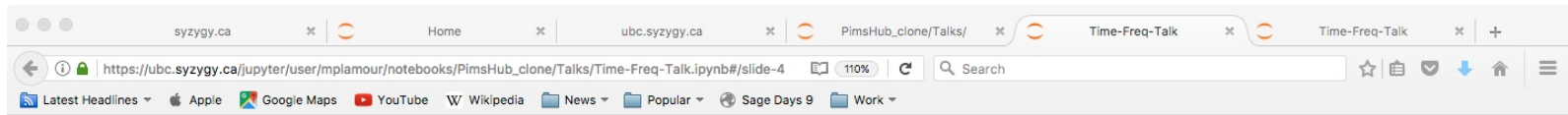
University of Calgary

March 23, 2016

Michael P. Lamoureux



# Slide Shows. Sample content



## ✕ Taylor series

We learn in Calculus that it is often useful to express certain functions as a sum of monomials. For instance,

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

or

$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots$$

Fourier suggested for periodic functions, it is useful to write them as sums of sines and cosines. For instance:

$$f(x) = \sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \dots$$



is a familiar function.



# Outline

- Jupyter demo
- **Technical overview**
- The collaboration between Compute Canada, Cybera, and the Pacific Institute for the Mathematical Sciences
- Where are we trying to get to?
- Discussion

# Motivation: *Make Jupyter widely available*

**Why?** Because it could benefit a *lot* of people

- ⇒ A lot of people are nibbling round the edges, let's dive in!
- ⇒ Benefits and work can be shared
- ⇒ Collaborate and expand

**How?** Do what we can, then collaborate to go further

- ⇒ Start with modest & resources (make a case for more ;-)
- ⇒ Work within existing framework (e.g. FIPPA)
- ⇒ Keep things simple, we're a really small group!

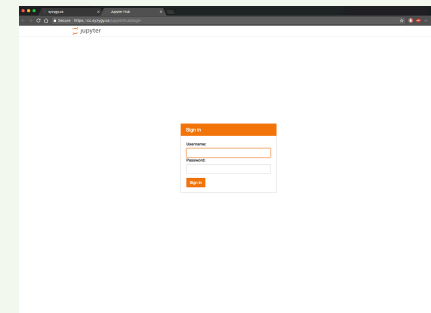
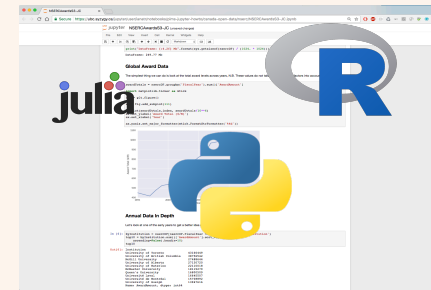
***Automate, Share & Collaborate***

**Caveat!** This *isn't* a finished product, but it is a good first step!

# IPython/Jupyter/JupyterHub

IP[y]:  
IPython

```
ana — IPython: Users/ana — ipython — 80x24
In [1]: import scipy.integrate as integrate
In [2]: import scipy.special as special
In [3]: result = integrate.quad(lambda x: special.jv(2.5,x), 0, 4.5)
In [4]: result
Out[4]: (1.117817938783253, 7.86631721638718e-09)
In [5]: import numpy as np
In [6]: A = np.mat('1 2;3 4')
In [7]: A
Out[7]:
matrix([[1, 2],
        [3, 4]])
In [8]: A.dot(A.I)
Out[8]:
matrix([[ 1.00000000e+00,  0.00000000e+00],
        [ 0.00179420e+16,  1.00000000e+00]])
In [9]:
```





# syzygy

- **What is syzygy?**
- Targeted JupyterHubs for Canadian researchers
  - Institutional credentials
  - Customized environments
- Documentation and materials
  - [intro.syzygy.ca](http://intro.syzygy.ca)
- Supporting services:
  - [discourse.syzygy.ca](http://discourse.syzygy.ca)
  - Slack
- We're doing this in partnership (RPP + MOUs)

# Existing Installations



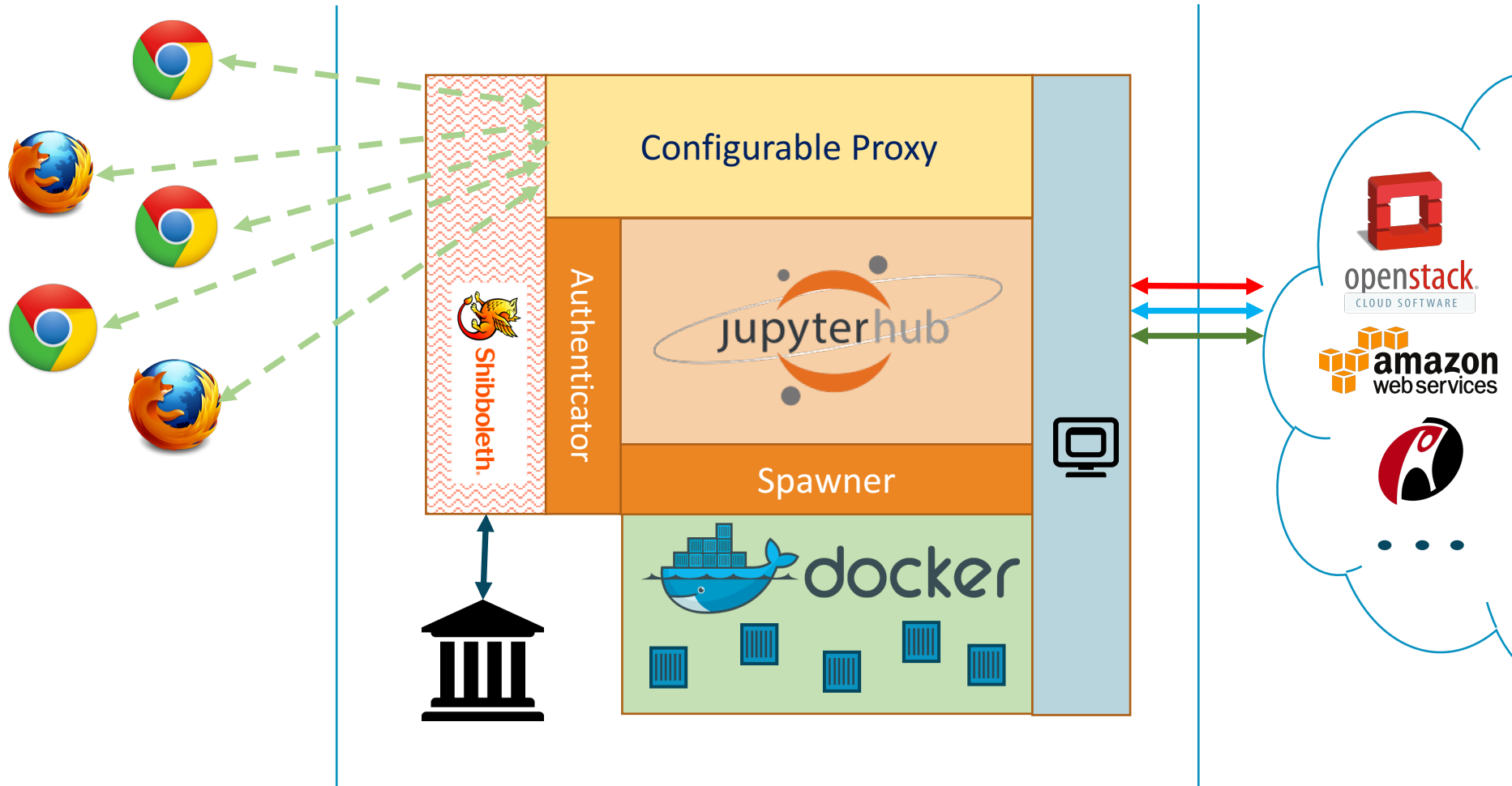
University of  
Lethbridge














cybera



# Simplified Syzygy Hub Overview



# Automation

Infrastructure	Provisioning	Software Stacks	Services
  	 ANSIBLE 	  	  

# Authentication: Shibboleth

Users have home institutions, use those credentials: *e.g. shibboleth*

- *User tries to access a protected resource*
- *shibboleth session created*
- *Pass secure headers to JupyterHub*
- *JupyterHub can make auth (and other) decisions*

The real point is that authenticators are modular

- LDAP/sssd, Google/Github/.. OAuth, ...
- Talking to Canarie about CAF



- Containers let us manage resources
- Spawn new single-user-servers that are
  - Reliably the same (*automation!*)
  - Resource appropriate and constrained
  - (relatively) isolated
- It also lets us keep an eye on expansion
  - Swarm
  - Mesos + Kubernetes

# Next Steps 1: More users

Our templated VMs OK for around a few hundred active users, but there will be/are limitations...

- Distributed filesystems
- Load Balancing
- More complicated routing

Most components scale simply (e.g. database), the main limitation (empirically) seems to be the proxy.

⇒ Experimenting with Nginx + LUA + Hub API

# Next Steps 2: Other flavours

- Félix-Antoine Fortin
  - Batchspawner - access deeper resources
  - Exercise existing allocations on clusters
  - CVMFS and external software repositories
- Interest in GPU + TensorFlow
  - Large storage integrations (google drive, globus...)
- Different Docker deployments
  - OpenStack Docker driver
  - **Kubernetes + Mesos**
  - Swarm



# Outline

- Jupyter demo
- Technical overview
- **The collaboration between Compute Canada, Cybera, and the Pacific Institute for the Mathematical Sciences**
- Where are we trying to get to?
- Discussion

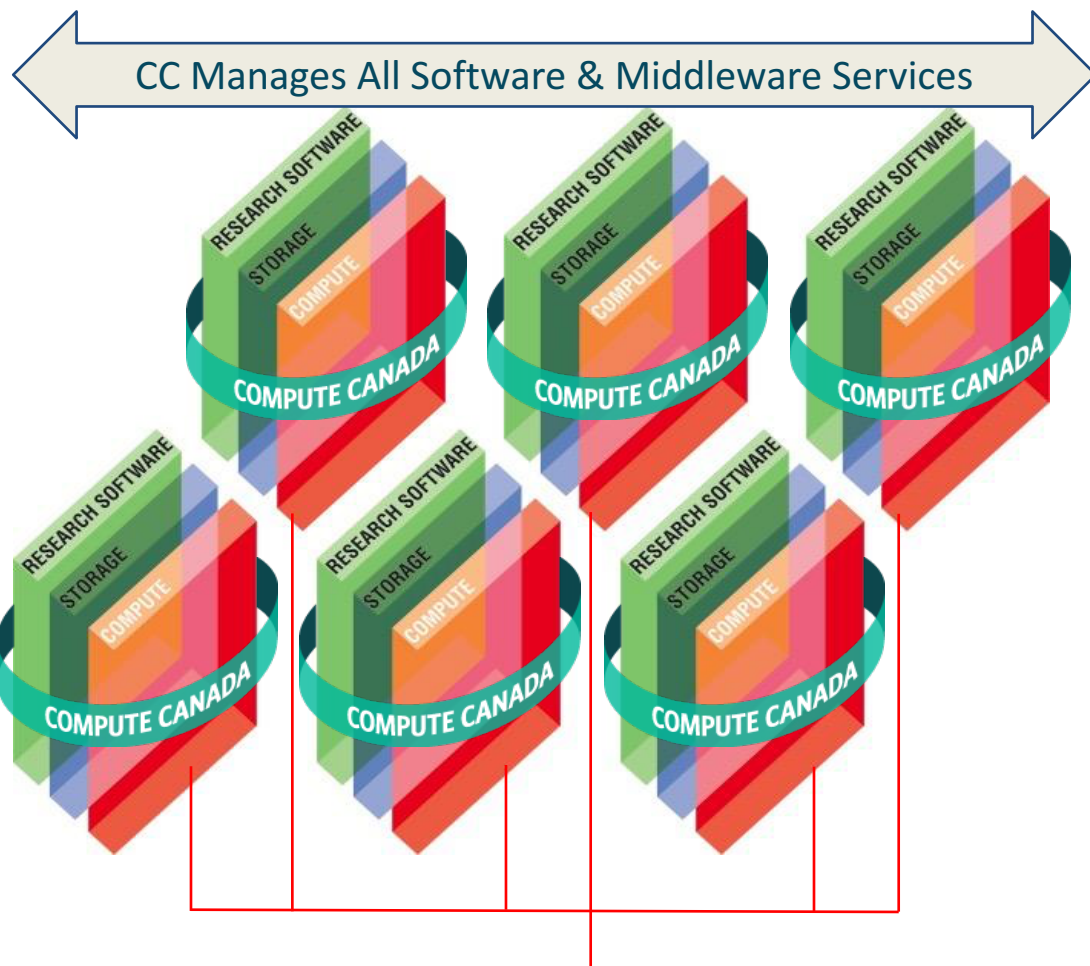
# Compute Canada

Tremendous computing capacity:

Cedar commissioned last week  
~3.6PF

Tremendous data capacity:  
Deploy more than 40PB this  
year (more tape), more in 2018.  
National Data  
Cyberinfrastructure.

How do users exploit the  
resources?



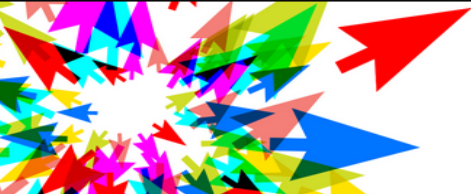
# Traditional Compute Canada

```
[root@localhost ~]# ping -q fa.wikipedia.org
PING text.pmtpa.wikimedia.org (208.80.152.2) 56(84) bytes of data.
^C
--- text.pmtpa.wikimedia.org ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 540.528/540.528/540.528/0.000 ms
[root@localhost ~]# pwd
/root
[root@localhost ~]# cd /var
[root@localhost var]# ls -la
total 72
drwxr-xr-x. 18 root root 4096 Jul 30 22:43 .
drwxr-xr-x. 23 root root 4096 Sep 14 20:42 ..
drwxr-xr-x.  2 root root 4096 May 14 00:15 account
drwxr-xr-x. 11 root root 4096 Jul 31 22:26 cache
drwxr-xr-x.  3 root root 4096 May 18 16:03 db
drwxr-xr-x.  3 root root 4096 May 18 16:03 empty
drwxr-xr-x.  2 root root 4096 May 18 16:03 games
drwxrwx--T.  2 root gdm  4096 Jun  2 18:39 gdm
drwxr-xr-x. 38 root root 4096 May 18 16:03 lib
drwxr-xr-x.  2 root root 4096 May 18 16:03 local
lrwxrwxrwx.  1 root root   11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x. 14 root root 4096 Sep 14 20:42 log
lrwxrwxrwx.  1 root root   10 Jul 30 22:43 mail -> spool/mail
drwxr-xr-x.  2 root root 4096 May 18 16:03 nis
drwxr-xr-x.  2 root root 4096 May 18 16:03 opt
drwxr-xr-x.  2 root root 4096 May 18 16:03 preserve
drwxr-xr-x.  2 root root 4096 Jul  1 22:11 report
lrwxrwxrwx.  1 root root   6 May 14 00:12 run -> ../run
drwxr-xr-x. 14 root root 4096 May 18 16:03 spool
drwxrwxrwt.  4 root root 4096 Sep 12 23:50 tmp
drwxr-xr-x.  2 root root 4096 May 18 16:03 yp
[root@localhost var]# yum search wiki
Loaded plugins: langpacks, presto, refresh-packagekit, remove-with-leaves
rpmfusion-free-updates                               | 2.7 kB      00:00
rpmfusion-free-updates/primary_db                     | 206 kB      00:04
rpmfusion-nonfree-updates                             | 2.7 kB      00:00
updates/metalink                                       | 5.9 kB      00:00
updates                                                 | 4.7 kB      00:00
updates/primary_db                                     73% [=====] ] 62 kB/s | 2.6 MB      00:15 ETA
```

Very useful and very powerful, but not for everyone.

# Compute Canada (2014+ - data)

[Log In](#) [Sign Up](#)



**compute**  
canada

**calcul**  
canada

## Sign In

[Sign Up with Globus](#)

Compute Canada has partnered with Globus to offer this high performance file transfer service.


Using your [Globus login](#). [alternate login](#)

Username

Password

[Forgot password?](#)

Manage Data ▾ Groups ▾ Account



**compute**  
canada

**calcul**  
canada

[Transfer Files](#) | [Activity](#) | [Endpoints](#) | [Bookmarks](#) | [Publish](#) | [Console](#)

Transfer Files

RECENT ACTIVITY 0 0 ▾ 0 0

Endpoint

Path

select none

up one folder

refresh list

gridstore

scratch

ComputeCanadaSP\_Draft\_Mar20\_EN.docx

PhDThesis.pdf

hhntup.root

Folder

Folder

23.45 KB

1.80 MB

2.49 GB

Endpoint

Path


select all

up one folder

refresh list

Hs.data.gz

164.09 MB



doneil

Password

☐ remember

Log in

# Compute Canada (platforms+portals)



The Canadian Writing Research Collaboratory  
Le Collaboratoire scientifique des écrits du Canada

Search...

ABOUT CWRC / CSEC ▾ CWRC NEWS ▾ EVENTS ▾ PROJECTS ▾ INFRASTRUCTURE PROJECTS ▾ RESEARCH PROJECTS ▾ CONTACT ▾ PARTICIPA



**The Digital Page**

THE COLLECTED WORKS OF

The Digital Page is a digital edition of the writings of P.K. Page and of the visual art she created under her married name P.K. Irwin, employing an open source text-image tool, the Digital Page Reader, created especially for this project. P.K. Page's career as a writer [...]

What is CWRC?


The Canadian Writing Research Collaboratory is an online project designed to provide unprecedented avenues for people to most move people in and out of the digital space.

## CBRAIN - Login

Login

Password


[Forgot your password?](#)



# IRIDA

## INTEGRATED RAPID INFECTIOUS DIS

Whole-genome sequencing has proven itself to be an effective epidemiological tool (2008 List of Infectious Diseases). The use of whole-genome sequencing in epidemiology to researchers and public health workers is inhibited by the ad-hoc nature and cost of the technology. The FWS/GC Bioinformatics Platform is to design and implement a software platform to make genomic data accessible to microbiologists and researchers.



## Genetics and Genomics Analysis Platform


GenAP The Computing Gateway for Life Sciences

[About](#) [Contact Us](#)

GenAP is a computing platform for life sciences researchers that leverages both the CANARIE high-speed network and Compute Canada's High Performance Computing (HPC) resources to give researchers access to modern and specialized services closely integrated to HPC resources. Being fully connected to the Compute Canada's users database, you can use GenAP as soon as you have a [Compute Canada account](#).

**GenAP offers:**

- ✓ Private Instances of the Galaxy Web application
- ✓ Solutions to share and publish your research data
- ✓ A collection of bioinformatics data analysis pipelines



# General Platforms, Partnerships

- It is important to Compute Canada that we make the resources accessible to ALL researchers
- We should do this by
  - Making the “bare” platform easier to use
  - Making it easier for research platforms to be built
- Big effort ongoing on building a common service/middleware layer across new CC sites (authentication, scheduling, monitoring, etc.)

# General Platforms, Partnerships

- Partnerships - CC has neither the capacity nor the expertise to “do it all” for Canadian researchers.
- Partner with organizations that have complementary strengths:
  - CARL - Federated Research Data Repository
  - PIMS+Cybera - Jupyter
- CC provides some underlying infrastructure, scale, software expertise but would miss some key ingredients for each.

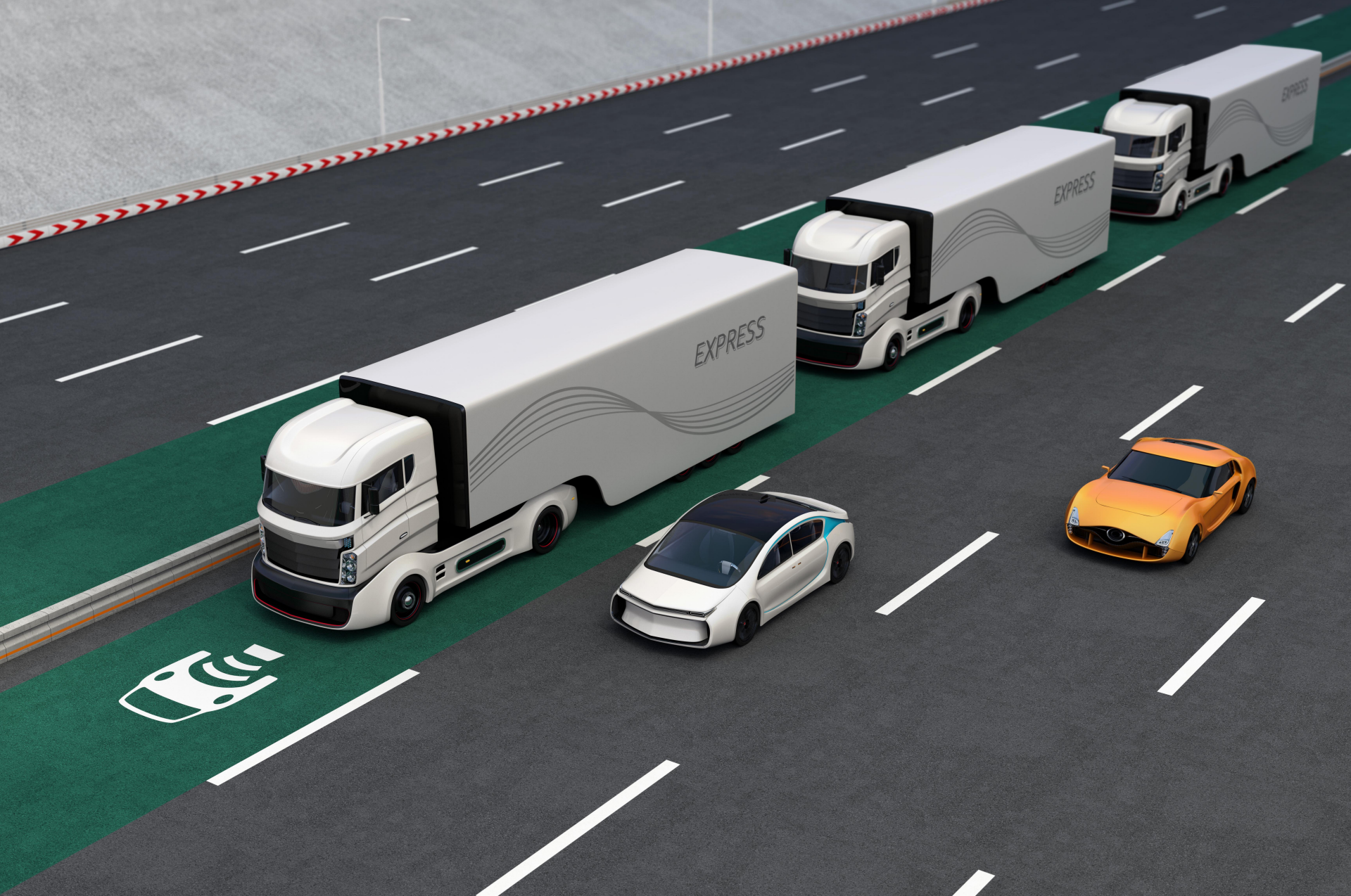
# PIMS + Cybera + CC

- PIMS is dedicated to making a standard platform (Honda Civic) available to all Canadian researchers.
- Cybera has a broader mandate than CC (e.g., K-12) and different set of use-cases.
- Together we want to give everyone a Civic and the tools to transform it into a Ferrari or a Hummer if that is more suitable for the task.
- Integrate and interoperate with CC services, CARL services, research portals.



# Outline

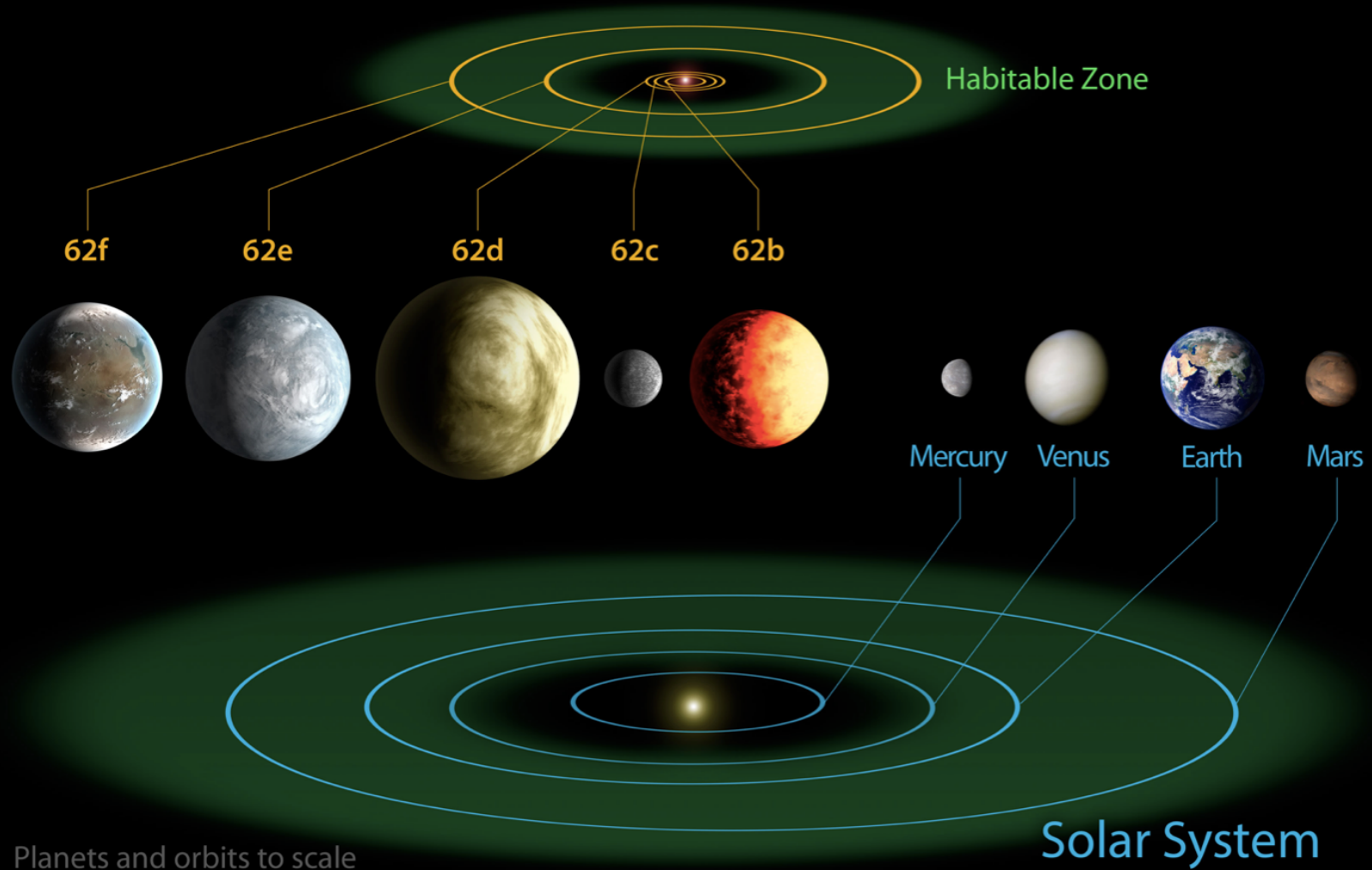
- Jupyter demo
- Technical overview
- The collaboration between Compute Canada, Cybera, and the Pacific Institute for the Mathematical Sciences
- **Where are we trying to get to?**
- Discussion







# Kepler-62 System





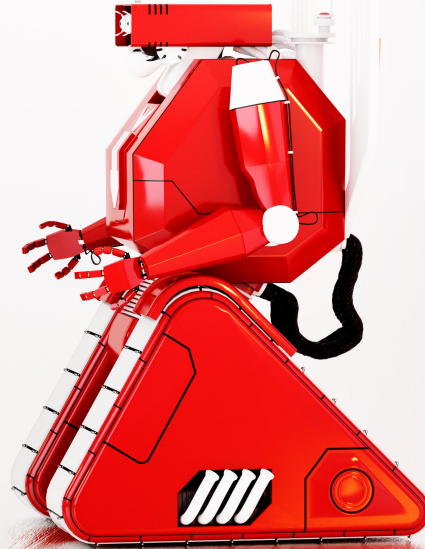








# JOB INTERVIEW THIS WAY



"In the same way that electricity revolutionized manufacturing and the microprocessor reinvented how we gather, analyze and communicate information, artificial intelligence will cut across nearly every industry....



It will shape the world that our kids and our grandkids grow up in.



# Outline

- Jupyter demo
- Technical overview
- The collaboration between Compute Canada, Cybera, and the Pacific Institute for the Mathematical Sciences
- Where are we trying to get to?
- **Discussion**

More information:  
[intro.syzygy.ca/](http://intro.syzygy.ca/)