IPython & StarCluster: Scalable Cloud Computing for the rest of us (a Qiime illustration)

Fernando Pérez, Min Ragan-Kelley, Brian Granger, Justin Riley
Greg Caporaso, Tony Walters, Rob Knight

http://ipython.org

Fernando.Perez@berkeley.edu

NIH Cloud Workshop, CU Boulder
April 3, 2012

A CU Boulder project

- Started when I was a graduate student in Physics (2001).
- Continued as a postdoc in Applied Mathematics.
- Brian Granger: CU Physics.

In brief

1. A better Python shell
2. Embeddable Kernel and powerful interactive clients
    1. Terminal
    2. Qt console
    3. Web notebook
3. Flexible parallel computing

A CU Boulder project

- Started when I was a graduate student in Physics (2001).
- Continued as a postdoc in Applied Mathematics.
- Brian Granger: CU Physics.

In brief

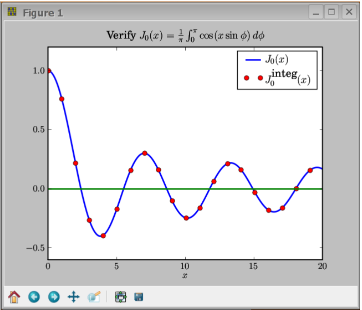
1. A better Python shell
2. Embeddable Kernel and powerful interactive clients
    1. Terminal
    2. Qt console
    3. Web notebook
3. Flexible parallel computing

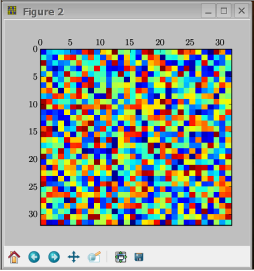
IPython: Matlab/IDL-like interactive use

fperez@longs:/home/fperez - Shell - Konsole

```
longs[~]> ipython -pylab
Python 2.4.3 (#2, Apr 27 2006, 14:43:58)
Type "copyright", "credits" or "license" for more information.

IPython 0.7.3.svn -- An enhanced Interactive Python.
?       -> Introduction to IPython's features.
%magic  -> Information about IPython's 'magic' % functions.
help    -> Python's own help system.
object? -> Details about 'object'. ?object also works, ?? prints more.

  Welcome to pylab, a matplotlib-based Python environment.
  For more information, type 'help(pylab)'.

In [1]: import math, numpy

In [2]: from scipy.integrate import quad

In [3]: from scipy.special import j0

In [4]: def j0i(x):
   ...:     """Integral form of J_0(x)"""
   ...:     def integrand(phi):
   ...:         return math.cos(x*math.sin(phi))
   ...:     return (1.0/math.pi)*quad(integrand,0,math.pi)[0]
   ...:

In [5]: x = numpy.linspace(0,20,200) # sample grid: 200 points between 0 and 20

In [6]: y = j0(x)  # sample J0 at all values of x

In [7]: x1 = x[::10]  # subsample the original grid every 10th point

In [8]: y1 = map(j0i,x1)  # evaluate the integral form at all points in x1

In [9]: # Make a plot with these values (the ; suppresses output)

In [10]: plot(x,y,label=r'$J_0(x)$');

In [11]: plot(x1,y1,'ro',label=r'$J_0^{\textrm{integ}}(x)$');

In [12]: axhline(0,color='green',label='_nolegend_');

In [13]: title(r'Verify $J_0(x)=\frac{1}{\pi}\int_0^{\pi}\cos(x \sin\phi)\,d\phi$');

In [14]: xlabel('$x$');

In [15]: legend();

In [16]: matshow(numpy.random.random((32,32)))
Out[16]: <matplotlib.figure.Figure instance at 0x4630042c>
```

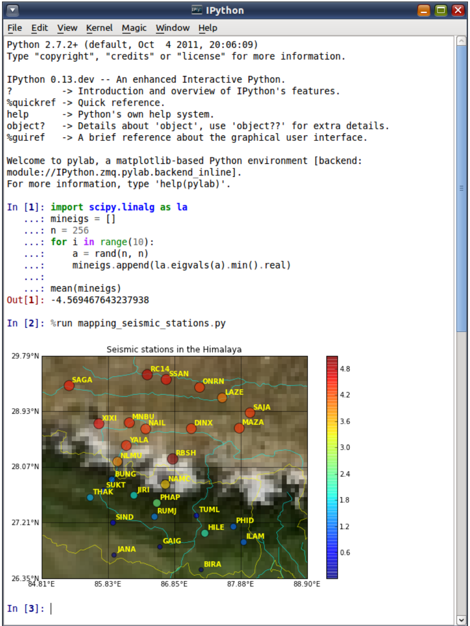
Shell





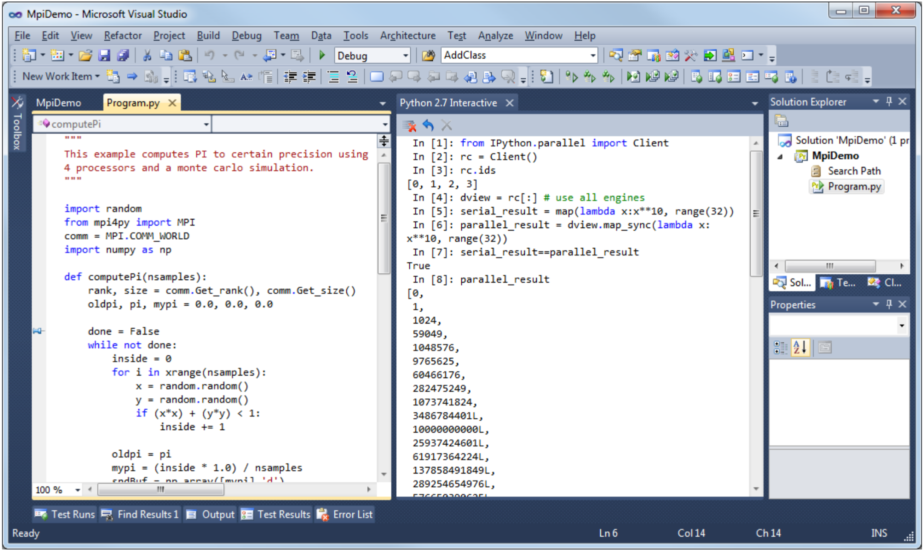
Qt console: inline plots, html, multiline editing, ...

Evan Patterson (Enthought)



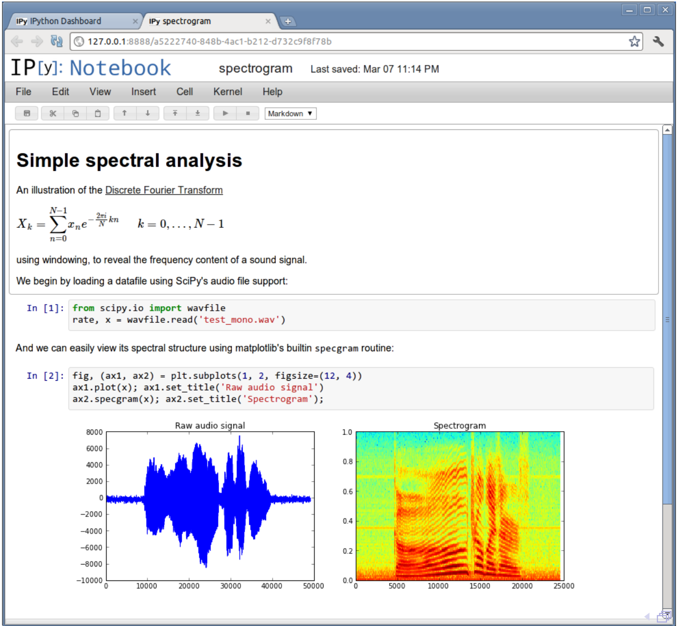
Microsoft Visual Studio 2010 integrated console

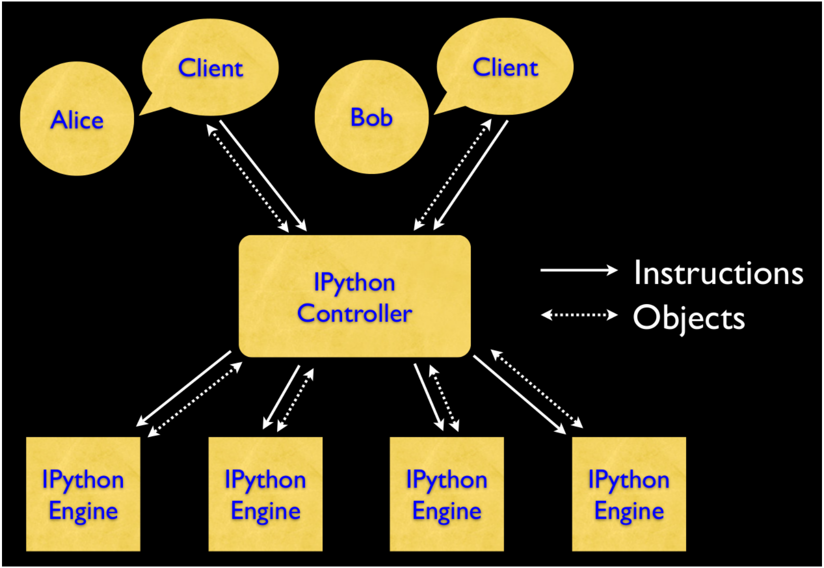
Dino Viehland and Shahrokh Mortazavi; `http://pytools.codeplex.com`



Browser-based notebook: rich text, code, plots, ...

Brian Granger, James Gao (Berkeley), rest of the team




Client
Alice
Client
Bob
IPython
Controller
Instructions
Objects
IPython
Engine
IPython
Engine
IPython
Engine
IPython
Engine

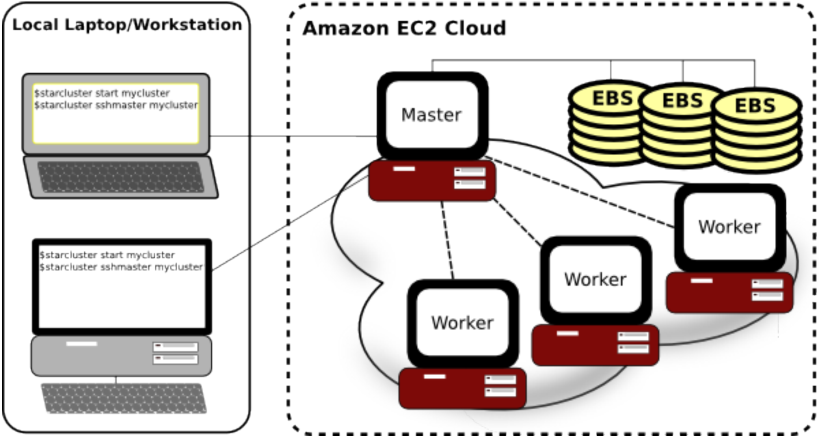
Support

- **Enthought**, Austin, TX: **Lots!**
- Tech-X Corporation, Boulder, CO: Parallel/notebook (previous versions)
- Microsoft: WinHPC support, Visual Studio integration
- NIH: via NiPy grant
- NSF: via Sage compmath grant
- Google: summer of code 2005, 2010.
- DoD/HPTi.

(Incomplete) Cast of Characters

- **Brian Granger** - Cal State San Luis Obispo Physics
- **Min Ragan-Kelley** - UC Berkeley Nuclear engineering.
- **Thomas Kluyver** - U. Sheffield Plant biology
- **Jörgen Stenarson** - SP Technical Research Institute of Sweden
- **Paul Ivanov** - UC Berkeley neuroscience
- **Robert Kern** - Enthought
- **Evan Patterson** - Caltech Physics/Enthought
- Stefan van der Walt - UC Berkeley
- John Hunter - TradeLink Securities, Chicago.
- Prabhu Ramachandran - Aerospace Engineering, IIT Bombay
- Satra Ghosh- MIT Neuroscience
- Gaël Varoquaux - Neurospin (Orsay, France)
- Ville Vainio - CS, Tampere University of Technology, Finland
- Barry Wark - Neuroscience, U. Washington.
- Ondrej Certik - Physics, U Nevada Reno
- Darren Dale - Cornell
- Justin Riley - MIT
- Mark Voorhies - UC San Francisco
- Nicholas Rougier - INRIA Nancy Grand Est
- Thomas Spura - Fedora project
- Julian Taylor - Debian/Ubuntu
- Many more! (~140 commit authors)

Simplify the process of building, configuring, and managing clusters of virtual machines on Amazon's EC2 cloud.



Resource Links

IPython

http://ipython.org/videos.html

http://ipython.org/documentation.html

StarCluster

http://web.mit.edu/star/cluster

StarCluster/IPython/Qiime public AMI:

ami-2faa7346