IPython: modern tools for interactive & web-enabled scientific computing

Fernando Pérez

http://fperez.org, @fperez_org

Fernando.Perez@berkeley.edu

Helen Wills Neuroscience Institute, UC Berkeley

NERSC User Day, LBL, Berkeley,
Feb 13, 2013

Outline

Beyond (Floating Point) Number Crunching

Symbolic manipulation

Interval arithmetic

Rationals

Arbitrary precision integers

Extended precision floating point

Hardware floating point

FORTRAN

Text processing

Databases

Data formats: HDF5, XML, ...

Graphical user interfaces

Web interfaces

Hardware control

Multi-language integration

Scientific Python: a Rich Ecosystem

python™

NumPy

IPython

SciPy

matplotlib

SymPy

Mayavi

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

Cython

SM StatsModels
Statistics in Python

scikits learn
machine learning in Python

scikits-image
image processing in python

PyTables

NetworkX

Outline

The Lifecycle of a Scientific Idea (schematically)

1. Individual exploratory work
2. Collaborative development
3. Production work (HPC, cloud, parallel)
4. Publication (with reproducible results!)
5. Education
6. Goto 1.

The Problem with most tools

Barriers and discontinuities in workflow in between all the steps

The Lifecycle of a Scientific Idea (schematically)

1. Individual exploratory work
2. Collaborative development
3. Production work (HPC, cloud, parallel)
4. Publication (with reproducible results!)
5. Education
6. Goto 1.

The Problem with most tools

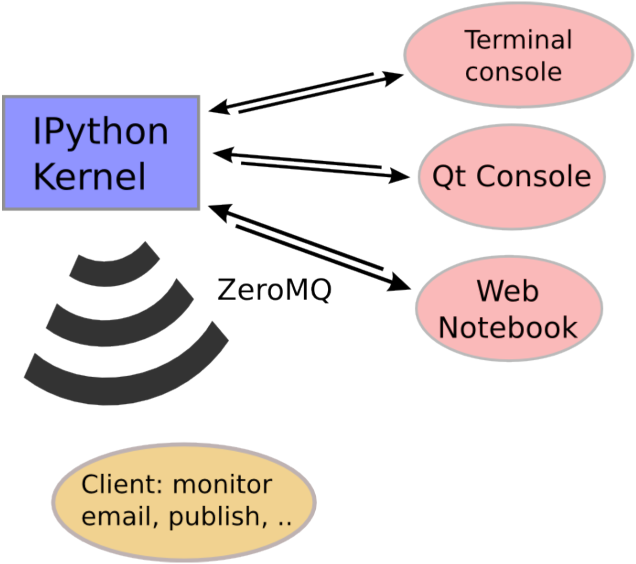
Barriers and discontinuities in workflow in between all the steps

IPython's goal:

Fluid transitions in all these steps

Demo

Pillar #1: An architecture for interactive computing



Pillar #2: the Notebook Format

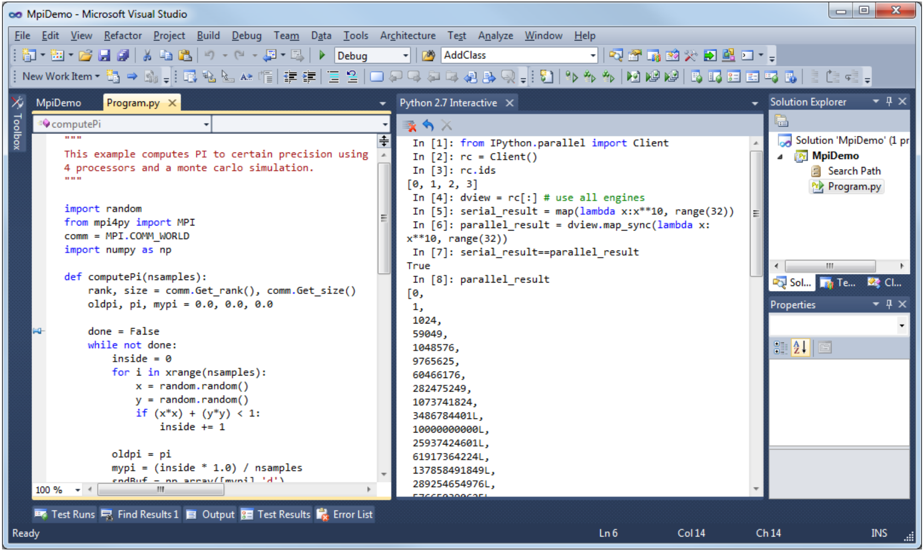
- **JSON** but version control-friendly
- Easy for machine processing, **fixable by hand if need be**.
- Lots of hooks for metadata
- **Not** Python-specific (R and Ruby notebooks exist, Julia planned)
- Produce Markdown, reST, LaTeX, HTML, etc...

An open format for sharing, publishing and archiving executable computational work

Documented protocols and formats: a growing ecosystem around IPython

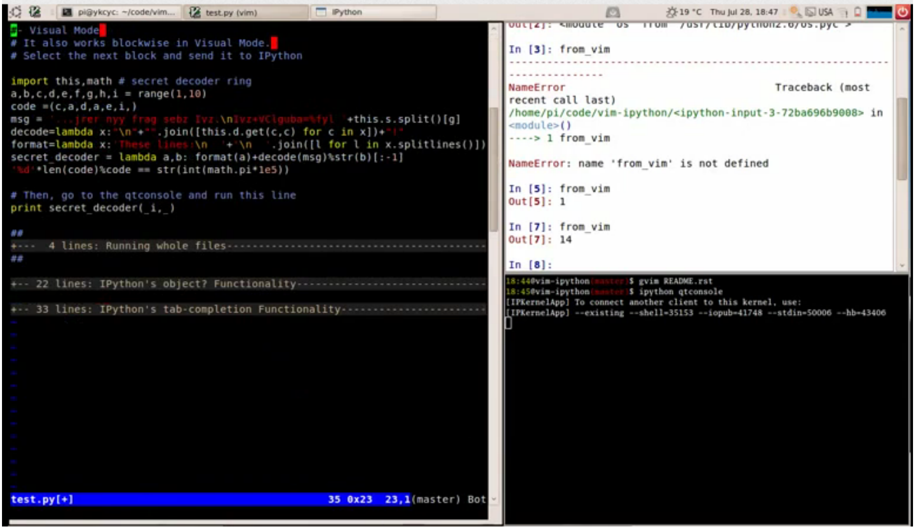
Microsoft Visual Studio 2010 integrated console

Dino Viehland and Shahrokh Mortazavi (Microsoft); http://pytools.codeplex.com



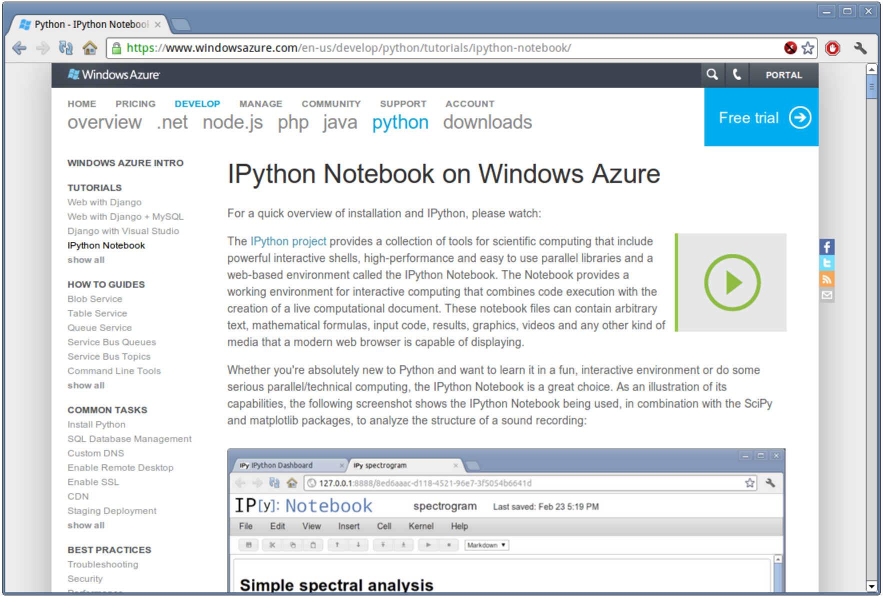
A vim client to control an IPython kernel/console

Paul Ivanov (Berkeley), `https://github.com/ivanov/vim-ipython`



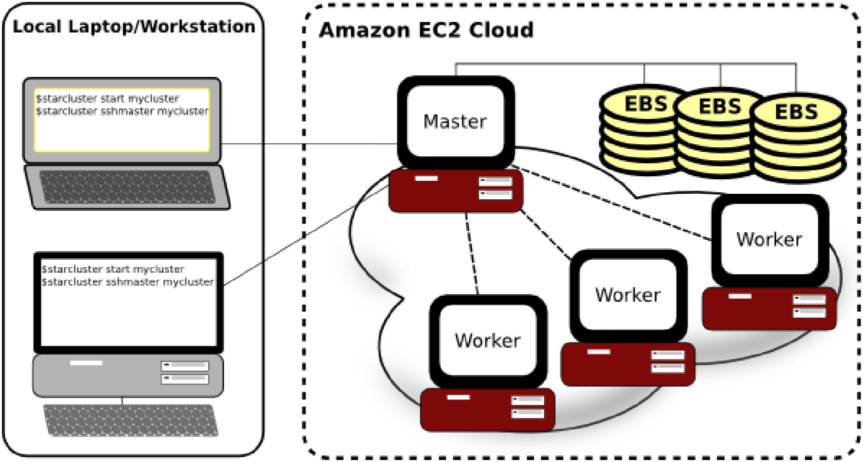
Notebooks on Windows Azure Cloud

Shahrokh Mortazavi (Microsoft), B.G., F.P.: `http://bit.ly/JQeojD`.



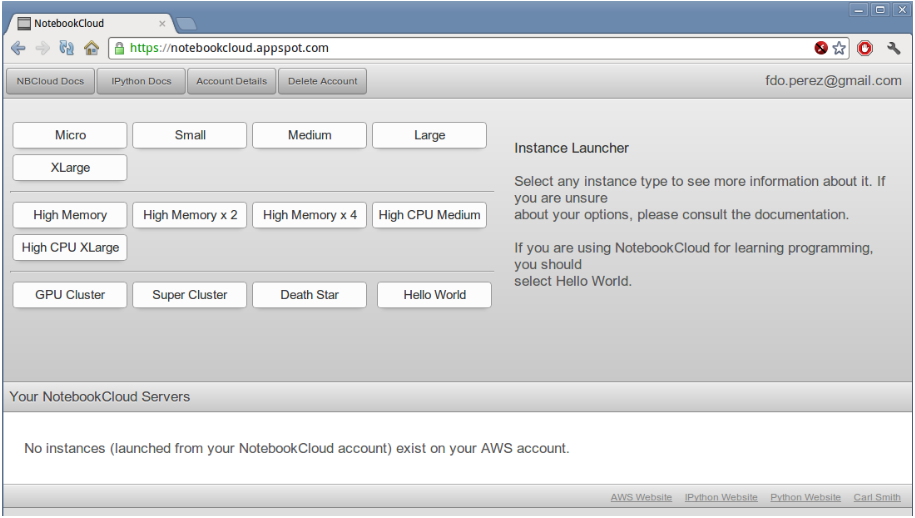
Star Cluster: IPython parallel+Notebook on Amazon EC2

Justin Riley (MIT): `http://web.mit.edu/star/cluster`



One-click single notebook on Amazon EC2

Carl Smith (UK): `https://notebookcloud.appspot.com`.



Other projects using IPython

Scientific

- **Software Carpentry**
- **EPD:** Enthought Python Distribution.
- **Continuum:** Anaconda, Wakari.
- **Sage**: open source mathematics.
- **PyRAF**: Space Telescope Science Institute
- **CASA**: Nat. Radio Astronomy Observatory
- **Ganga**: CERN
- **PyMAD**: neutron spectrom., Laue Langevin
- **Sardana**: European Synchrotron Radiation
- **ASCEND**: eng. modeling (Carnegie Mellon).
- **JModelica**: dynamical systems.
- **DASH**: Denver Aerosol Sources and Health.
- **Trilinos**: Sandia National Lab.
- **DoD**: baseline configuration.
- **Mayavi**: 3d visualization, Enthought.
- **NiPype**: computational pipelines, MIT.

Web/Other

- **Visual Studio 2010**: MS.
- **Django**.
- **Turbo Gears**.
- **Pylons** web framework
- **Zope** and **Plone** CMS.
- Axon Shell, BBC **Kamaelia**.
- **Schevo** database.
- **Pitz**: distributed task/bug tracking.
- **iVR** (interactive Virtual Reality).
- **Movable Python** (portable Python environment).
- ...

(Incomplete) Cast of Characters

- **Brian Granger** - Physics, Cal State San Luis Obispo
- **Min Ragan-Kelley** - Nuclear Engineering, UC Berkeley
- **Matthias Bussonnier** - Physics, Institut Curie, Paris
- **Brad Froehle** - Mathematics, UC Berkeley
- **Paul Ivanov** - Neuroscience, UC Berkeley.
- **Robert Kern** - Enthought
- **Thomas Kluyver** - Biology, U. Sheffield
- **Jonathan March**- Enthought
- **Evan Patterson** - Physics, Caltech/Enthought
- **Jörgen Stenarson** - Elect. Engineering, Sweden.
- John Hunter - TradeLink Securities, Chicago.
- Stefan van der Walt - Applied Math, U. Stellenbosch
- 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
- Many more! (~150 commit authors)

- **Enthought**, Austin, TX: **Lots!**
- **Microsoft**: WinHPC support, Visual Studio integration, Azure (thanks to Shahrokh Mortazavi).
- **DoD/DRC Inc**: funding through Sept. 2012 (thanks to Jose Unpingco and Chris Keees).
- NIH: via NiPy grant
- NSF: via Sage compmath grant
- Google: summer of code 2005, 2010.
- Tech-X Corp., Boulder, CO: Parallel/notebook (previous versions)

Recent Stable Funding

ALFRED P. SLOAN FOUNDATION

2-year funding, core team (7 people)

SIMONS FOUNDATION

Stanford: reproducibility & data sharing (neuroimaging, other data-intensive cases)

NumFOCUS: http://numfocus.org

Foundation to promote Open Source Scientific Computing

- Support the development of multiple projects.
- Community-created and driven.
- A neutral ground for industry, academia and government to support scientific open source.
- 501(c)3 - donations are tax-exempt in the USA