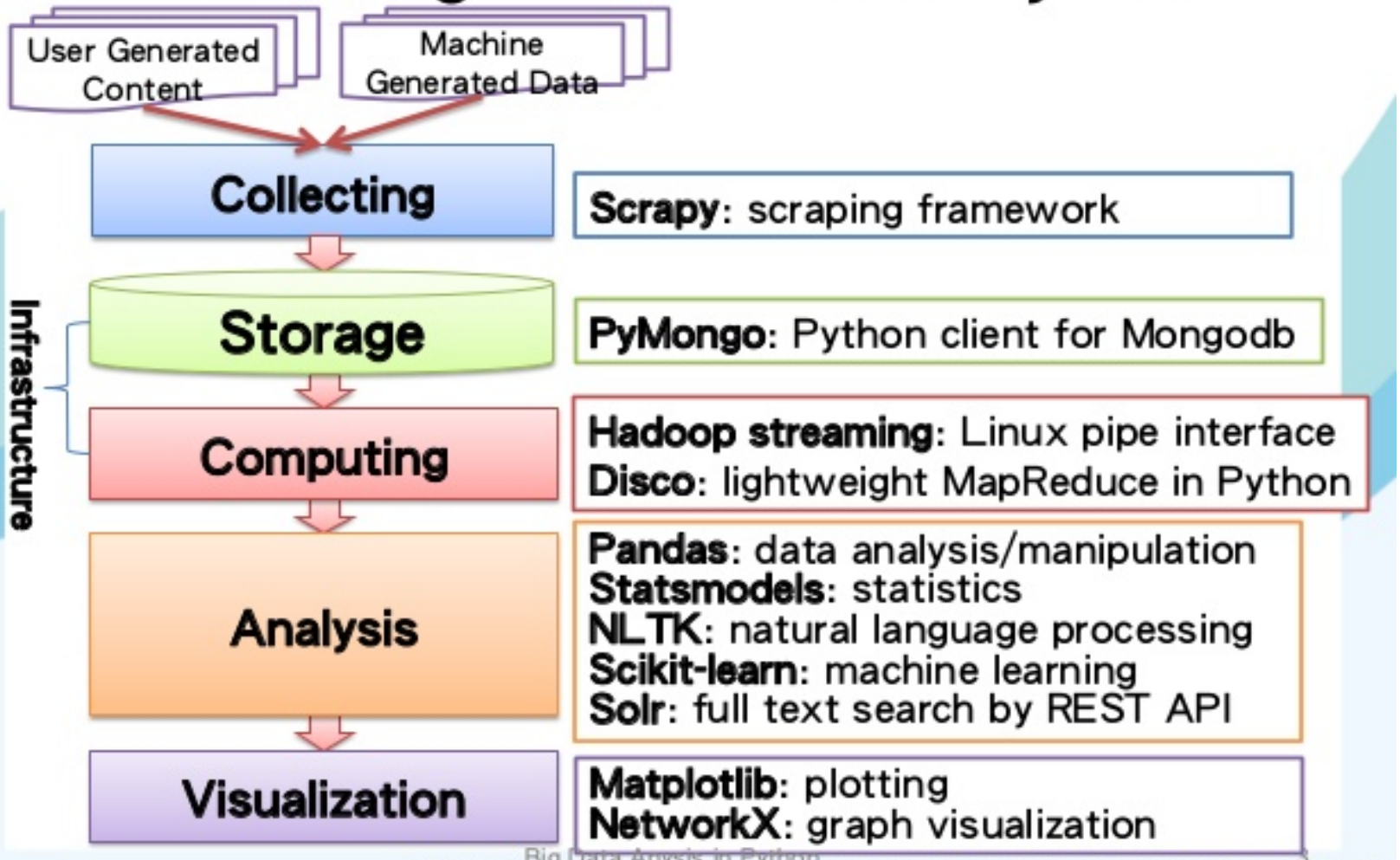
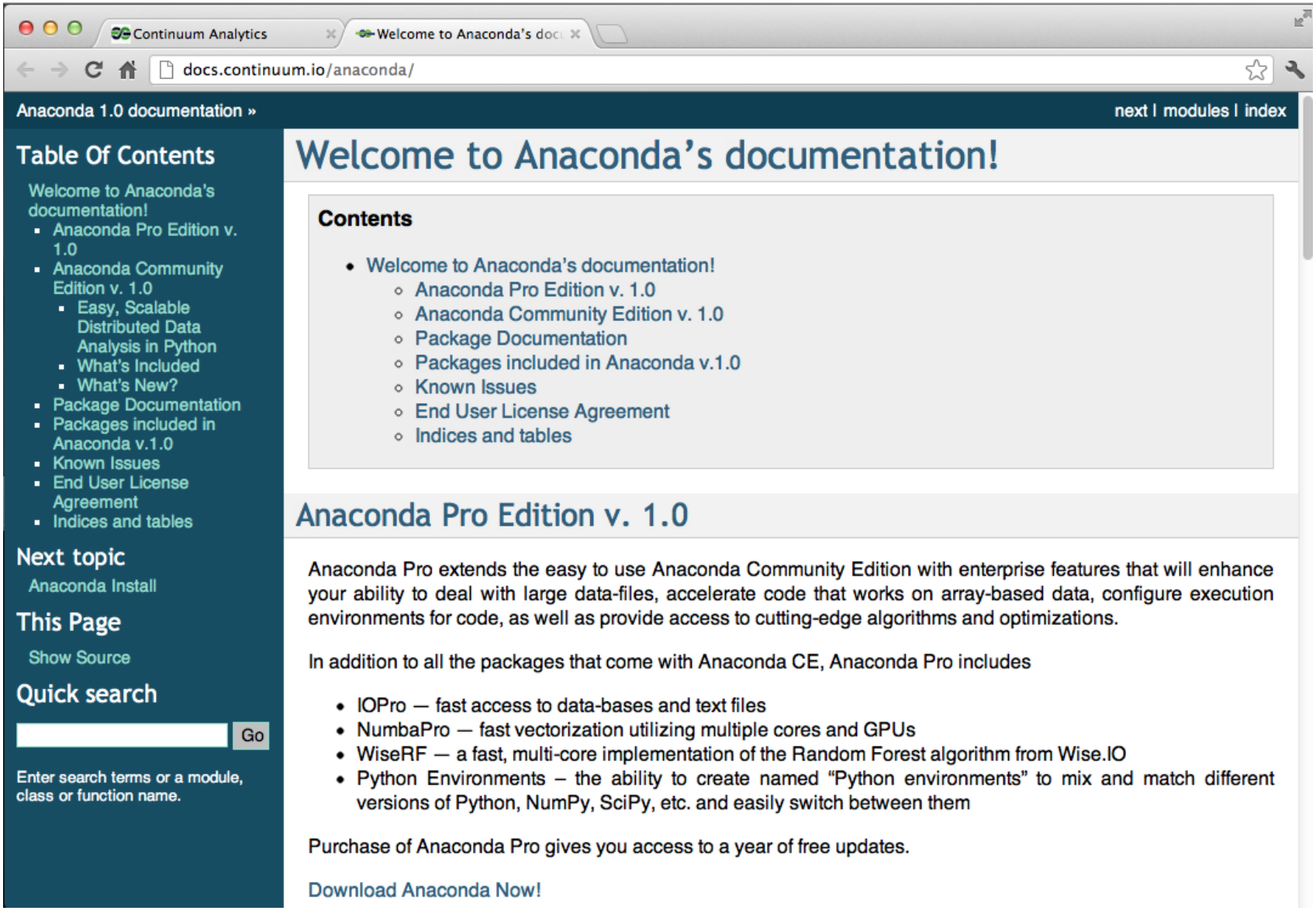
Important Components of the Python Scientific Stack

When Big Data meet Python



http://www.slideshare.net/jimmy_lai/when-big-data-meet-python

Continuum Analytics Anaconda



Continuum Analytics Anaconda

- Anaconda, a free product of Continuum Analytics (www.continuum.io), is a virtually complete scientific stack for Python.
- It includes both the core Python interpreter and standard libraries as well as most modules required for data analysis.

Continuum Analytics Anaconda

- Anaconda is free to use and modules for accelerating the performance of linear algebra on Intel processors using the **Math Kernel Library** (MKL) are available (free to academic users and for a small cost to non-academic users).
- Continuum Analytics also provides other high-performance modules for reading large data files or using the GPU to further accelerate performance for an additional, modest charge.

Installing Anaconda

Most importantly, installation is extraordinarily easy onWindows, Linux and OS X. Anaconda is also simple to update to the latest version using

```
conda update conda
conda update anaconda
```

NumPy and SciPy

- **NumPy** provides a set of array and matrix data types which are essential for statistics and econometrics.
- **SciPy** contains a large number of routines needed for analysis of data.The most important include a wide range of random number generators, linear algebra routines and optimizers.
- Remark: SciPy depends on NumPy.
- More on them later.

IPython

IPython provides an interactive Python environment which enhances productivity when developing code or performing interactive data analysis.

IPy IPython Dashboard | IPy IPython Dashboard | IPy Untitled0

127.0.0.1:8889/a5989266-c452-4c65-b962-d921b364b3e0

IP[y]: Notebook Untitled0 Last saved: Jun 02 2:02 PM

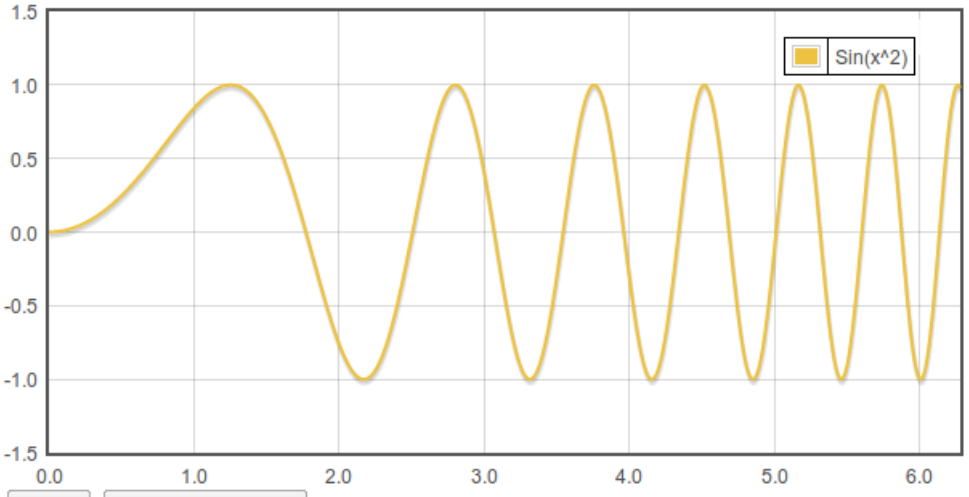
File Edit View Insert Cell Kernel Help

Code

In [8]:
```
import flot
```

In [9]:
```
fp = flot.Plot()
fplot = fp.plot_figure
```

In [10]:
```
x = linspace(0, 2*pi, 500)
y = sin(x**2)
fplot(x, y, label='Sin(x^2)')
```



Ipython Notebook / Jupyter

Evolved from the IPython Project

The language-agnostic parts of IPython are getting a new home in Project Jupyter

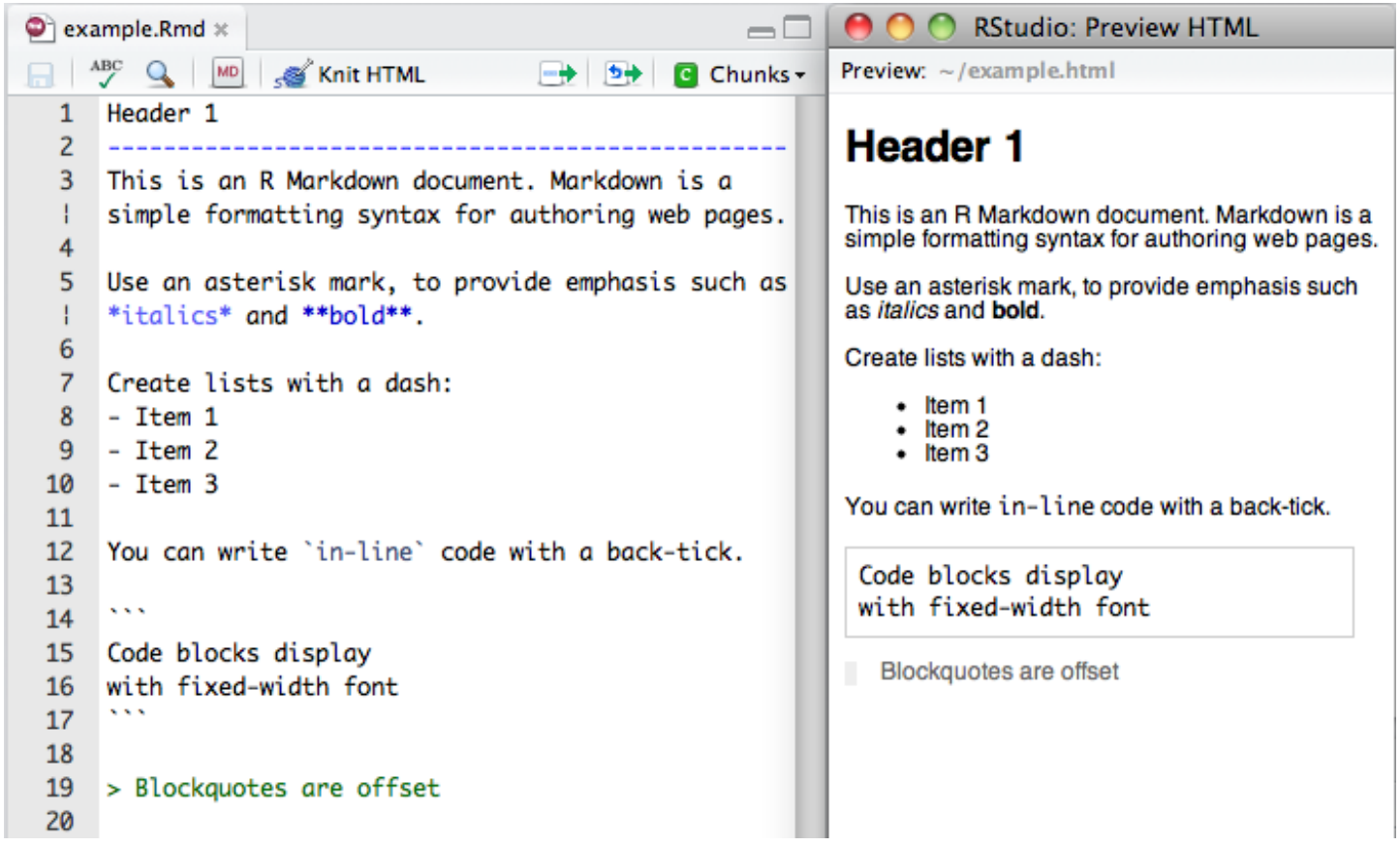
IPython

- Interactive Python shell
- Python kernel for Jupyter
- Interactive Parallel Python

Jupyter

- Rich REPL Protocol
- Notebook (format, environment, conversion)
- JupyterHub (multi-user notebook server)
- More...

Markdown is a text-to-HTML conversion tool for web writers. Markdown allows you to write using an easy-to-read, easy-to-write plain text format, then convert it to structurally valid XHTML (or HTML).



matplotlib and seaborn

Graphics Packages

- **matplotlib** provides a plotting environment for 2D plots, with limited support for 3D plotting.
- **seaborn** is a Python package that improves the default appearance of matplotlib plots without any additional code.

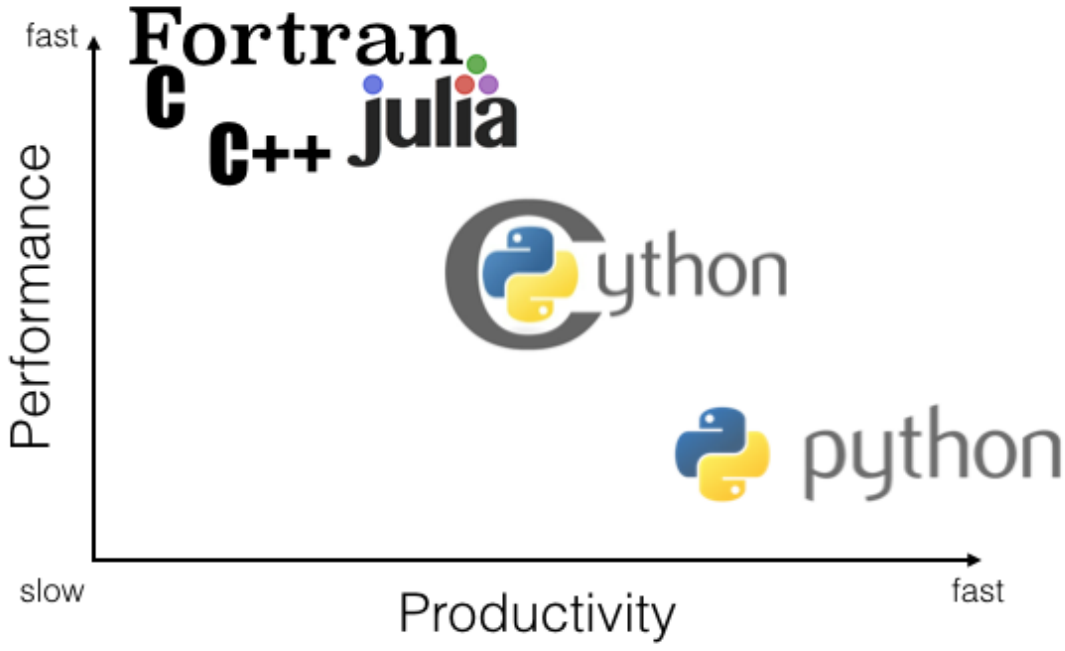
pandas

- *pandas* is a high-performance module that provides a comprehensive set of structures for working with data.
- *pandas* excels at handling structured data, such as data sets containing many variables, working with missing values and merging across multiple data sets.

pandas

- While extremely useful, *pandas* is not an essential component of the Python scientific stack unlike NumPy, SciPy or matplotlib, and so while *pandas* doesnt make data analysis possible in Python, it makes it much easier.
- *pandas* also provides high-performance, robust methods for importing from and exporting to a wide range of formats.
- - example `read.csv()`


Cython


fast
Performance
slow
Fortran
C
C++
julia
Cython
python
Productivity
fast

Performance Modules : Cython and Numba

A number of modules are available to help with performance. These include Cython and Numba.

Cython Cython is a Python module which facilitates using a simple Python-derived creole to write functions that can be compiled to native (C code) Python extensions.

Numba Numba uses a method of just-in-time compilation to translate a subset of Python to native code using *Low-Level Virtual Machine* (LLVM).