Start to Finish Data Science with Anaconda

Anaconda 4.3, early 2017

Data science is about a systematic way to apply numerical and visual techniques to analyze and interpret data. To do that you need easy access to the best numerical computing, data processing, and visualization tools today.

This material is part of a 60 minute workshop introducing you to Anaconda: your gateway to a rich ecosystem of open source data science tools and libraries.

Navigate to the Anaconda Cloud page for this asset which can be found at:

[http://anaconda.org/ijstokes/start-to-finish-data-science] (http://anaconda.org/ijstokes/start-to-finish-data-science)

[https://github.com/ijstokes/start-to-finish-ds] (https://github.com/ijstokes/start-to-finish-ds) *(advanced users)*

Anaconda Cloud provides a public instance of Anaconda Repository. That particular asset has been published in the `ijstokes` channel as a publicly accessible Jupyter Notebook. From there you can download a template version of the Notebook that will allow you to follow along.

Part A (10 minutes)

- Introduction
- Setup
- Conda Environments

Part B (20 minutes)

- Pandas data injest and manipulation
- Bokeh data visualization
- Scikit-Learn machine learning
- Q&A

Part C (10 minutes)

- Anaconda Project specifications

- Project deployment

Part D (10 minutes)

- Anaconda Enterprise v4 workflow
- Anaconda Enterprise v5 workflow

Part E (10 minutes)

- Announcements
- Survey
- Q&A

A - Setup & Smoke Test

1 - Save this Notebook into your Anaconda working space

- Anaconda Distribution for people using their laptop directly: save these inside the Anaconda Project directory
- Anaconda Enterprise for people using a web-interface to access server-deployed Anaconda: create a new project then upload the notebook file into the project

2 - Terminal Window

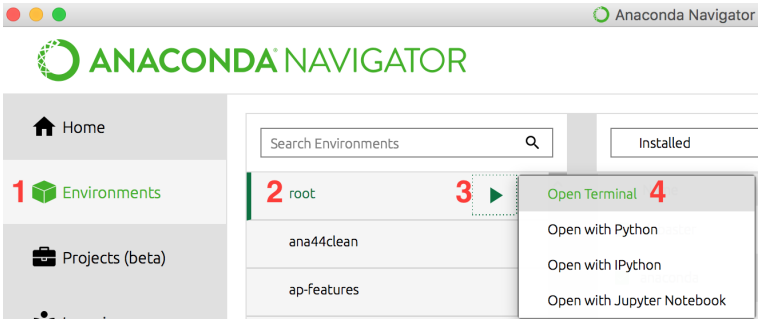
Get to a terminal window either through Navigator (if you are running the Anaconda Distribution on your local system) or through the Project view (if you are using Anaconda Enterprise), as shown in the screenshots below

2a - Anaconda Navigator (local users)

In [1]:
```
from IPython.display import Image
```

In [2]:
```
Image("imgs/AnacondaNavigatorOpenTerminal.png")
```

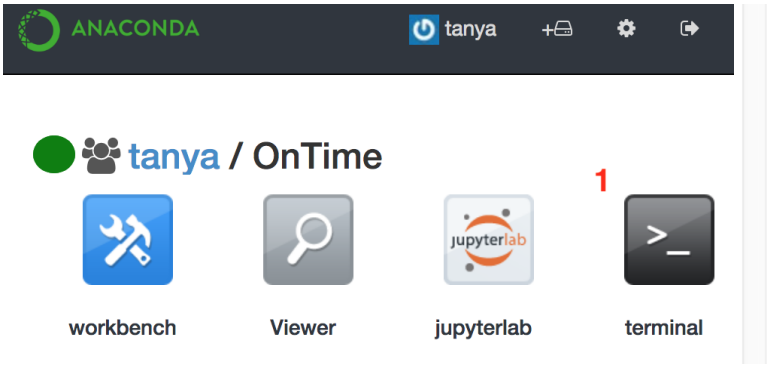
Out[2]:



2b - Anaconda Enterprise 4 (server based)

In [3]:
```
Image("imgs/AnacondaEnterprise4OpenTerminal.png")
```

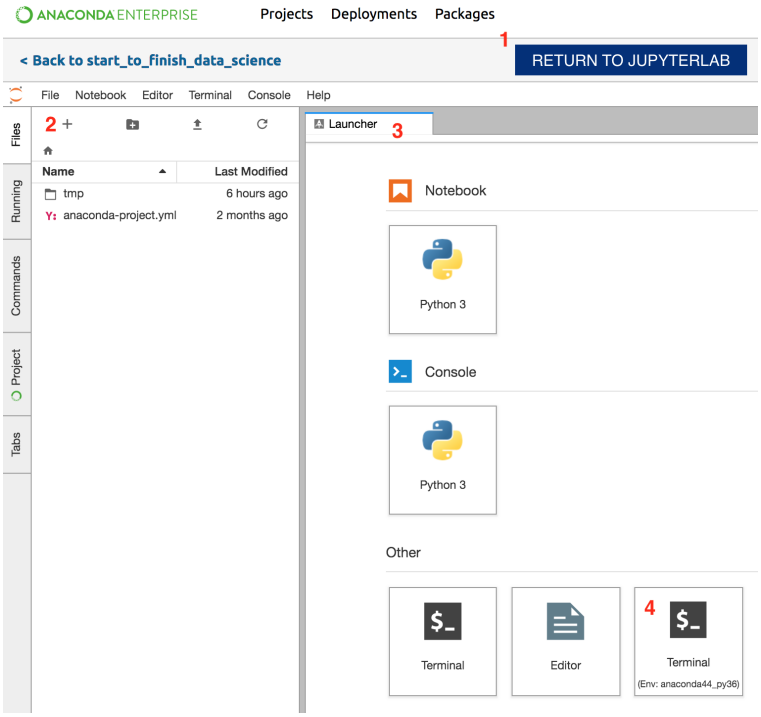
Out[3]:



2c - Anaconda Enterprise 5 (server based)

In [4]:
```
Image("imgs/AnacondaEnterprise5OpenTerminal.png")
```

Out[4]:



3 - Create a conda environment

We want everyone on the same page which means creating a reference conda environment and using that for this workshop. Since enterprise users of Anaconda are the target audience, and sometimes in an enterprise setting it can take many months to get the latest version of a piece of software, we'll use the early 2017 release of Anaconda, version 4.3, as our reference.

From the terminal window you opened, execute the following command:

```
conda create -n anaconda43 anaconda=4.3
```

If that doesn't work properly for you then an alternate strategy is to execute:

```
conda create -n approx43 jupyter bokeh=0.12.4 pand
as=0.19.2 scikit-learn=0.18.1
```

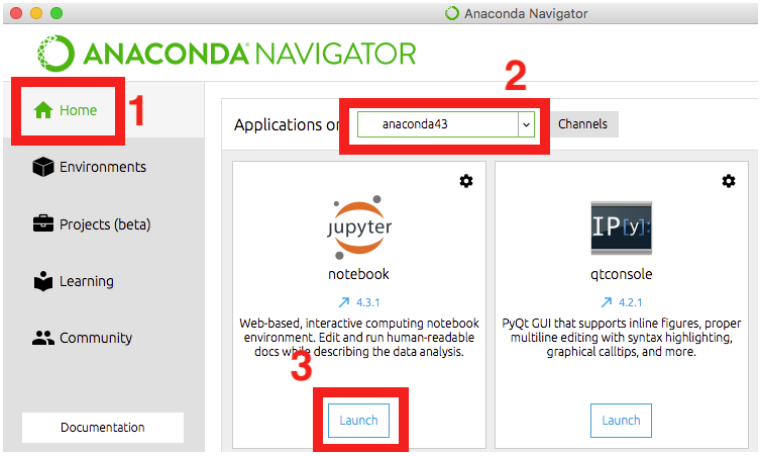
4 - Start Jupyter Notebook with this conda environment

You now need to get Jupyter started in the right environment:

4a - Jupyter from Navigator

In [5]:
```
Image("imgs/AnacondaSelectHomeEnvJupyter.png")
```

Out[5]:

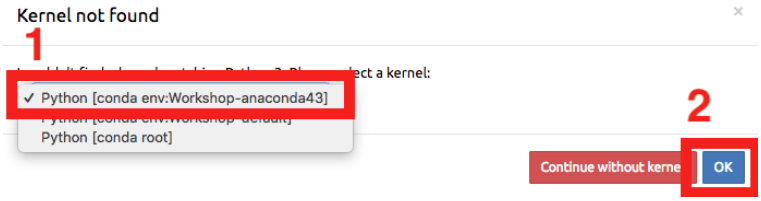


4b - Anaconda Enterprise 4 Change Kernel

From within your project, select **"Jupyter"** and then pick the `anaconda43`kernel. You may get a "*Kernel not found*" error.

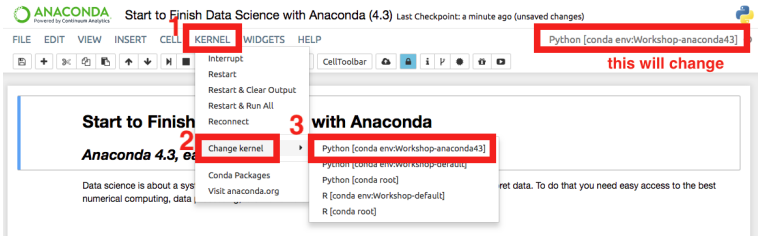
In [6]:
```
Image("imgs/AnacondaEnterprise4KernelNotFound.png")
```

Out[6]:



In [7]:
```
Image("imgs/AnacondaEnterprise4ChangeKernel.png")
```

Out[7]:



5 - Smoke Test

Execute each of the following by pressing `SHIFT-ENTER`. They all should work.

In [ ]:
```
import sys
from __future__ import print_function, with_statement
```

In [ ]:
```
print(sys.executable)
```

In [ ]:
```
print(sys.version)
```

In [ ]:
```
import pandas as pd
```

In [ ]:
```
import bokeh as bk
```