



# DS-GA 1007

## Programming for Data Science

### Lecture 5

### CLI I - Working with Files and Folders

# Reminders

- ▶ Survey 2
  - ▶ Anonymous Survey due Monday October 7

# Reminders

- ▶ Survey 2
- ▶ Final Exam
  - ▶ Thursday December 19 4-6pm

# Reminders

- ▶ Survey 2
- ▶ Final Exam
- ▶ Forum
  - ▶ Contact Ziyu and Nimi through Graders Group on Messages
  - ▶ Post to Forum about
    - ▶ FAQ
    - ▶ Lecture, Section, Assignments
    - ▶ Project

# Reminders

- ▶ Labs

- ▶ Due on Wednesday 11:59pm

- ▶ Leave on JupyterHub

- ▶ Follow Links

- ▶ Sign off and sign on again

- ▶ Grading

- ▶ Lab01 not graded

- ▶ Lab02 collected with Lab03

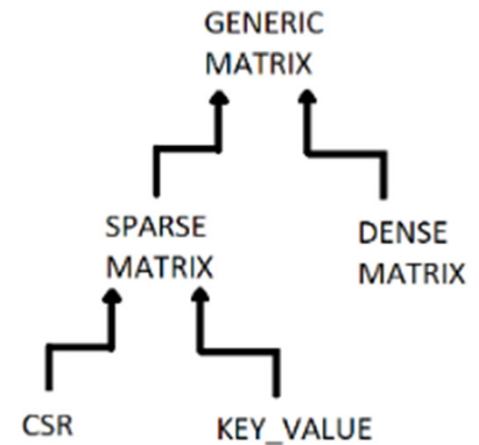
- ▶ Lab03 onward lowest grade dropped

# Reminders

- ▶ Homework
  - ▶ Due on Saturday 11:59pm
  - ▶ Submission
    - ▶ Leave notebook on JupyterHub
    - ▶ Export notebook to html
    - ▶ Print html to pdf
    - ▶ Upload pdf to Gradescope
    - ▶ [Mark responses](#)

# Homework 03

- Update
  - Check on JupyterHub
  - Clarified Signature of Constructor



```
# Implement this
class CSRMatrix(GenericSparseMatrix):
    def __init__(self, nrow, ncol, data):
        raise NotImplementedError

    def dense_to_sparse(self, arr):
        raise NotImplementedError

class KeyValueMatrix(GenericSparseMatrix):
    def __init__(self, nrow, ncol, data):
        raise NotImplementedError

    def dense_to_sparse(self, arr):
        raise NotImplementedError
```

# Reminders

- ▶ Project
  - ▶ Groups
    - ▶ Forum
    - ▶ Random
  - ▶ Problem
    - ▶ Software Development
    - ▶ Data Science
  - ▶ Dates
    - ▶ **October 31<sup>st</sup>: Project Proposal**
    - ▶ **November 28<sup>th</sup>: Project Milestone**
    - ▶ **December 14<sup>th</sup>: Project Report**



# Agenda

- Review
  - Patterns for classes and objects



# Agenda

- ▶ Review
- ▶ Lesson
  - ▶ Debugging
  - ▶ Logging



# Agenda

- ▶ Review
- ▶ Lesson
- ▶ Demo
  - ▶ Shell and Command Line Interface



# Debugging

# Shell and Command Line Interface

- ▶ Shell is application that passes directives from keyboard to the operating system
- ▶ Programmers use Command Line Interface to type directives for Shell
- ▶ Application running Command Line Interface called Terminal
- ▶ Terminal has prompt
  - ▶ Try PS1=DS-GA-1007...what happens to prompt?

# Shell and Command Line Interface

- ▶ Directives include commands for operating system
- ▶ Commands refer to programs
  - ▶ ls will list the content of folder
- ▶ Commands take 0 or more arguments separated by spaces
- ▶ Commands take special characters that modify behavior

# Shell and Command Line Interface

- ▶ Remember difference between Python and IPython...
- ▶ Python
  - ▶ Interpreter taking scripts with file extension .py
  - ▶ Integrated Development Environment (Visual Studio Code) combines interpreter with text editor

# Shell and Command Line Interface

- ▶ Remember difference between Python and IPython...
- ▶ IPython
  - ▶ Command line interface for interactively passing directives to interpreter
    - ▶ Read
    - ▶ Evaluate
    - ▶ Print
    - ▶ Loop
  - ▶ Instead of Terminal use Web based Integrated Development Environment
  - ▶ Jupyter Notebook file extension .ipynb



## Connection between IPython and Bash?

- ▶ IPython allows access to Bash without switching between multiple windows IPython
- ▶ Execute bash commands directly from within Jupyter Notebook using exclamation point
- ▶ Anything appearing after ! on a line will be executed not by the Python interpreter but by the shell.

# Connection between IPython and Bash?

```
!echo "Hello World"
```

Hello World

```
!pwd
```

/scratch/work/jupyterhub/2019-FA-DS-GA-1007/cp126/DS-GA-1007-Public/demos/demo02

# Connection between IPython and Bash?

```
: %%bash
```

```
pwd;
```

```
echo "Hello World";
```

```
/scratch/work/jupyterhub/2019-FA-DS-GA-1007/cp126/DS-GA-1007-Public/demos/demo0  
2  
Hello World
```

# Connection between IPython and Bash?

```
current_directory = !pwd  
print(current_directory)  
print(type(current_directory))
```

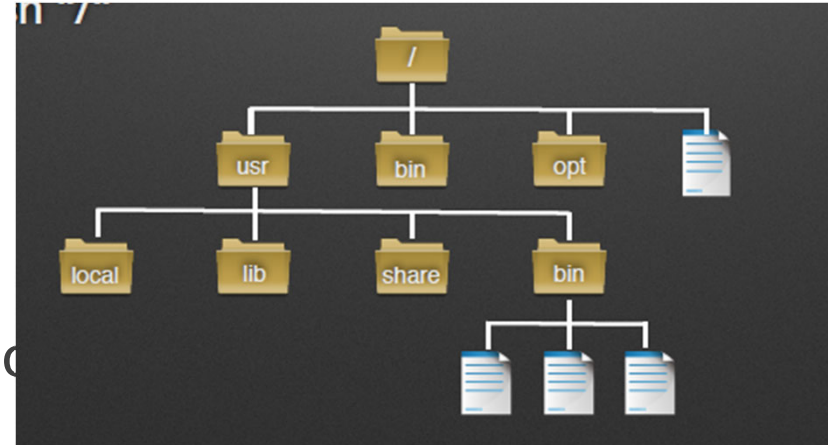
```
['/scratch/work/jupyterhub/2019-FA-DS-GA-1007/cp126/DS-GA-1007-Public/demos/dem  
o02']  
<class 'IPython.utils.text.SList'>
```

```
message = "Hello World"  
!echo {message}
```

```
Hello World
```

# Files and Directories

- ▶ A filesystem is a hierarchical arrangement of
  - ▶ Files containing data
  - ▶ Directories containing files and other directories
  - ▶ Folder used by analogy in graphical user interfaces (GUI)
- ▶ Top of filesystem called root. Denoted by /



# Files and Directories

## ▶ Paths

- ▶ List of directories leading to location of file or directory
- ▶ Absolute (starts with /) or relative to current directory
- ▶ Directory and file names separated by /

## ▶ Special characters

- ▶ . current directory
- ▶ .. parent directory
- ▶ Files beginning with . are hidden

# Commands, Options, Arguments

- ▶ Shell commands use format *command -option ... argument ...*
  - ▶ *command* name of shell command
  - ▶ *-option* used to change behavior
  - ▶ *argument* data passes to commands
- ▶ Options
  - ▶ Sometimes have two hypens (e.g. --help)
  - ▶ Sometimes no hyphen (e.g. chmod +r)

# Connection between IPython and Bash?

```
!ls
```

```
demo02.ipynb  images  large-file-01.csv  large-file-03.csv
```

```
!ls images/
```

```
for.PNG  if.PNG
```

```
!ls ..
```

```
demo01  demo02  demo03  demo04
```



# Connection between IPython and Bash?

```
!ls -l
```

```
total 92
```

```
-rwxrwx---+ 1 cp126 cp126 72356 Oct  3 14:10 demo02.ipynb  
drwxrwx---+ 2 cp126 cp126  4096 Sep 17 12:18 images  
-rwxrwx---+ 1 cp126 cp126  5365 Sep 17 12:18 large-file-01.csv  
-rwxrwx---+ 1 cp126 cp126  5127 Sep 17 12:18 large-file-03.csv
```

# Connection between IPython and Bash?

```
!ls -l -a .
```

```
total 112
drwxrwx---+ 4 cp126 cp126 4096 Oct  3 14:22 .
drwxrwx---+ 7 cp126 cp126 4096 Sep 30 10:39 ..
-rwxrwx---+ 1 cp126 cp126 73877 Oct  3 14:22 demo02.ipynb
drwxrwx---+ 2 cp126 cp126 4096 Sep 17 12:18 images
drwxrwx---+ 2 cp126 cp126 4096 Oct  3 14:02 .ipynb_checkpoints
-rwxrwx---+ 1 cp126 cp126 5365 Sep 17 12:18 large-file-01.csv
-rwxrwx---+ 1 cp126 cp126 5127 Sep 17 12:18 large-file-03.csv
-rwxrwx---+ 1 cp126 cp126  651 Sep 17 12:18 .script.py
```

# Connection between IPython and Bash?

```
!ls -la
```

```
total 108
drwxrwx---+ 4 cp126 cp126 4096 Oct  3 14:21 .
drwxrwx---+ 7 cp126 cp126 4096 Sep 30 10:39 ..
-rwxrwx---+ 1 cp126 cp126 72356 Oct  3 14:10 demo02.ipynb
drwxrwx---+ 2 cp126 cp126 4096 Sep 17 12:18 images
drwxrwx---+ 2 cp126 cp126 4096 Oct  3 14:02 .ipynb_checkpoints
-rwxrwx---+ 1 cp126 cp126 5365 Sep 17 12:18 large-file-01.csv
-rwxrwx---+ 1 cp126 cp126 5127 Sep 17 12:18 large-file-03.csv
-rwxrwx---+ 1 cp126 cp126  651 Sep 17 12:18 .script.py
```

# Files and Directories

- ▶ Searching for files and directories require matching patterns
- ▶ Wildcards allow for multiple patterns

*	Matches any characters
?	Matches any single character
[ <u>characters</u> ]	Matches any character that is in the set of <u>characters</u>
[! <u>characters</u> ]	Matches any character that is not in the set of <u>characters</u>

# Files and Directories

<code>*</code>	All filenames
<code>g*</code>	All filenames beginning with “g”
<code>b*.txt</code>	All filenames that begin with “b” and end with “.txt”
<code>Data???</code>	Any filename that begins with “Data” followed by exactly three more characters
<code>[abc]*</code>	Any filename that begins with “a”, “b”, or “c” followed by any other characters
<code>Data[!ef]*</code>	Any filename that begins with “Data” followed by any characters that are not “e” or “f” followed by any other characters

# Question

▶ Law\*

▶ Law, La, Lawyer, aw

▶ \*Law\*

▶ Law, La, Lawyer, aw

▶ ?at

▶ Bat, at, bat

▶ [CB]at

▶ Cat, cat, bat, Bat

▶ Letter[0-9]

▶ Letter0, Letters, Letter10, Letter0

▶ [!C]at

▶ Cat, cat, bat, Bat

# Take-Aways

- ▶ Divide and Conquer
  - ▶ Split large problem into small problems. Call function recursively on smaller problems.
- ▶ Encapsulation and Inheritance
  - ▶ Classes includes attributes and methods. Compare to Abstraction and Decomposition for functions
- ▶ Debugging
  - ▶ Halt the execution of code to locate bugs