DS-GA 1007 Programming for Data Science

Lecture 1

Reminders

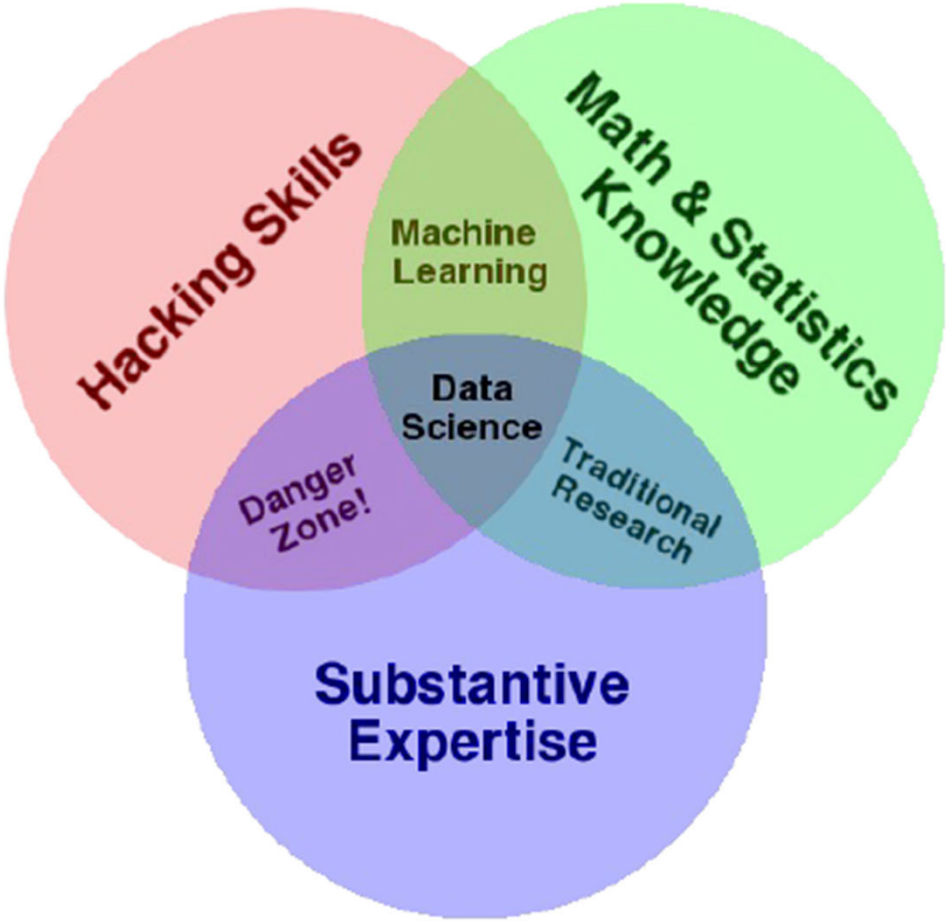
- Section
  - Mondays from 7:00pm-7:50pm at 60th 5th Avenue, Room 150
- Syllabus
  - Please review the course policies about assignments and grading.
- Surveys
  - Please complete Survey 1

Agenda

- Overview
  - Data Science
  - DS-GA 1007
- Lesson
- Demo
- Logistics

What is Data Science?

- Drawing useful conclusions from data using computation
- **Exploration**
  - Identifying patterns in information
  - Uses charts and tables
- **Prediction**
  - Making informed guesses
  - Uses procedures to determine appropriate hypotheses
- **Inference**
  - Quantifying whether those patterns are reliable
  - Uses randomization

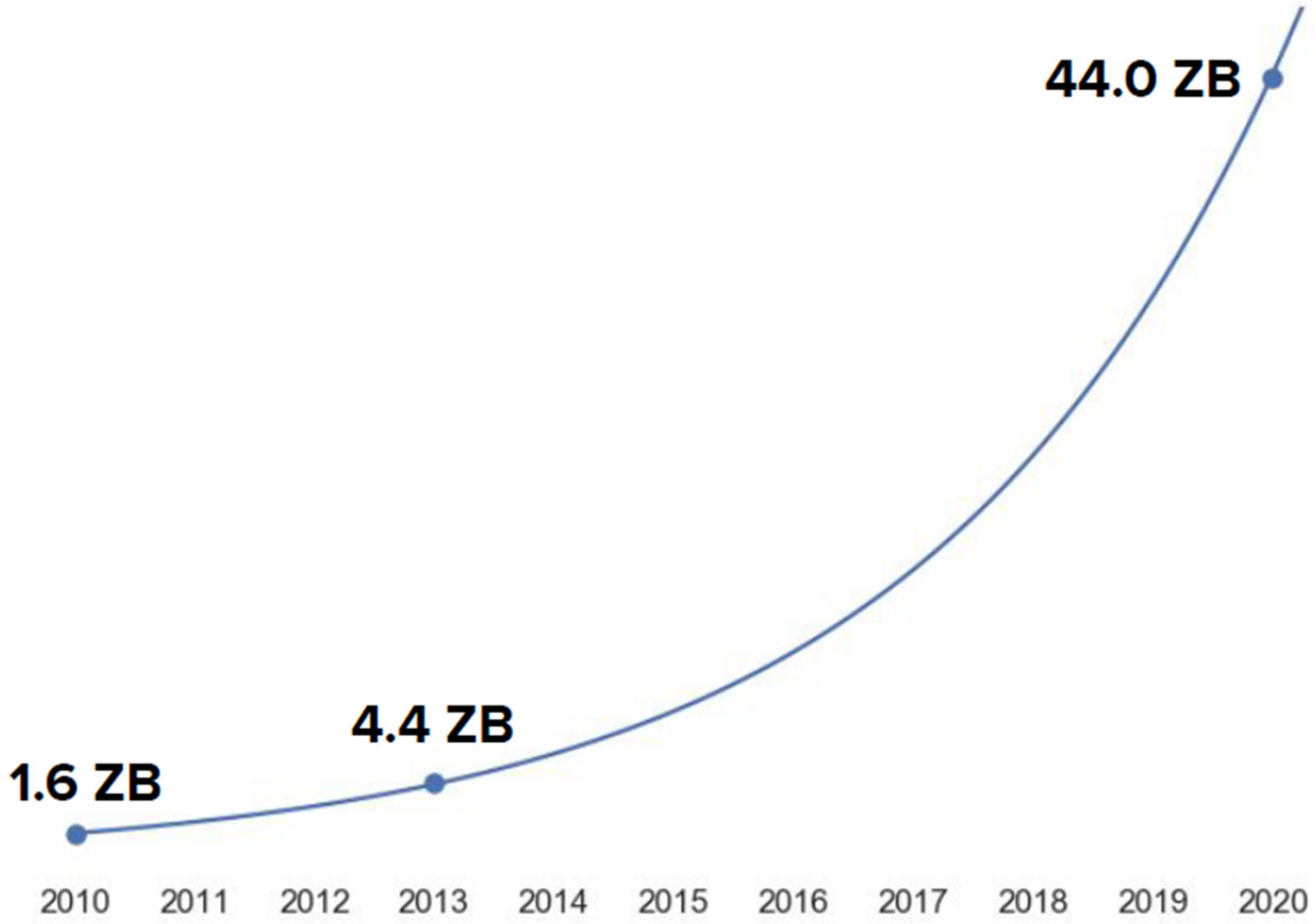

Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

Who uses Data Science?

- Academia
- Industry
- Government

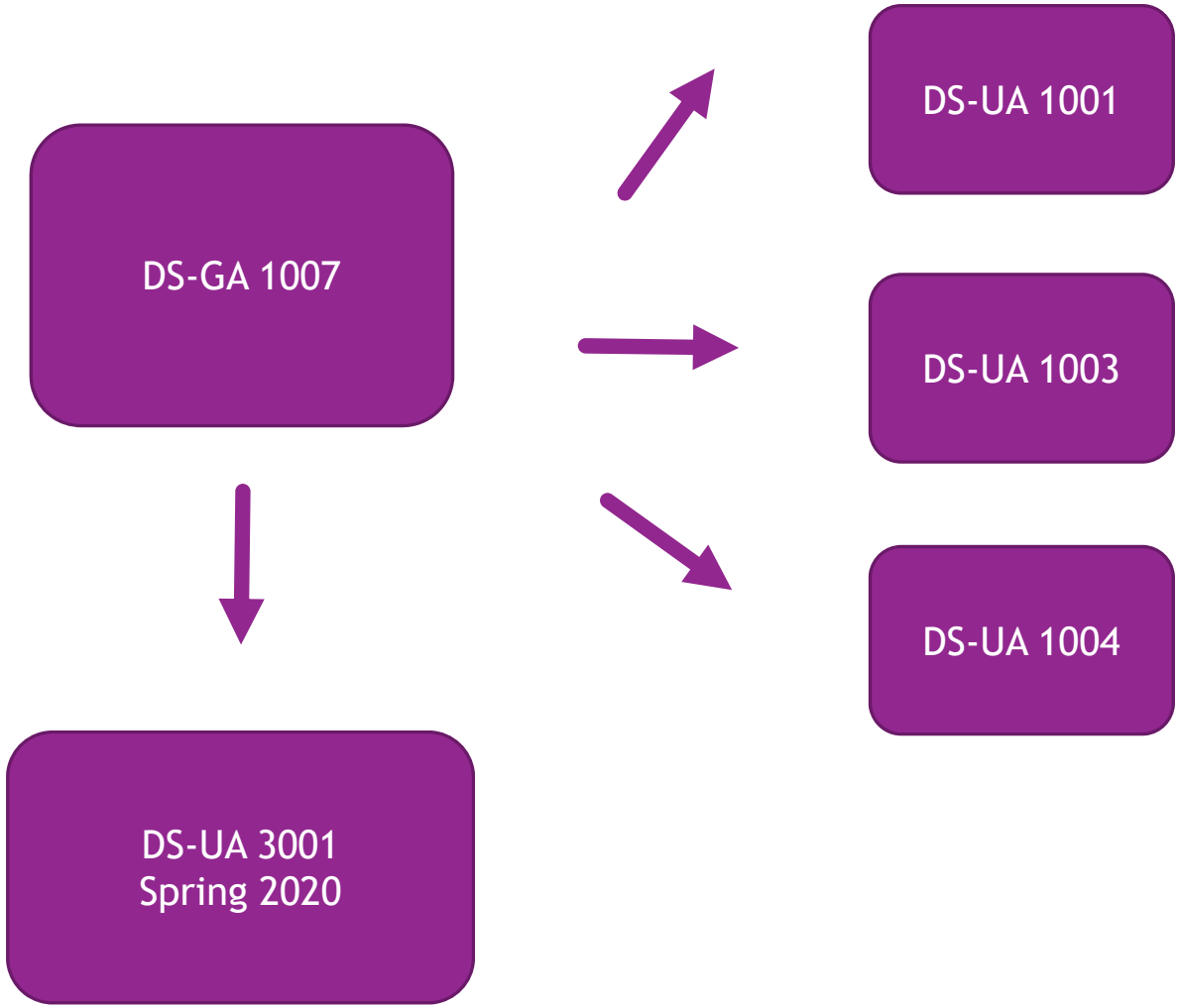
Why Study Data Science?

- Data is not just numbers, it includes
  - Text
  - Images
  - Videos
- Technology has made data ubiquitous
  - Storage
  - Communication
  - Digitization


44.0 ZB
4.4 ZB
1.6 ZB
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

DS-GA 1007 Programming for Data Science

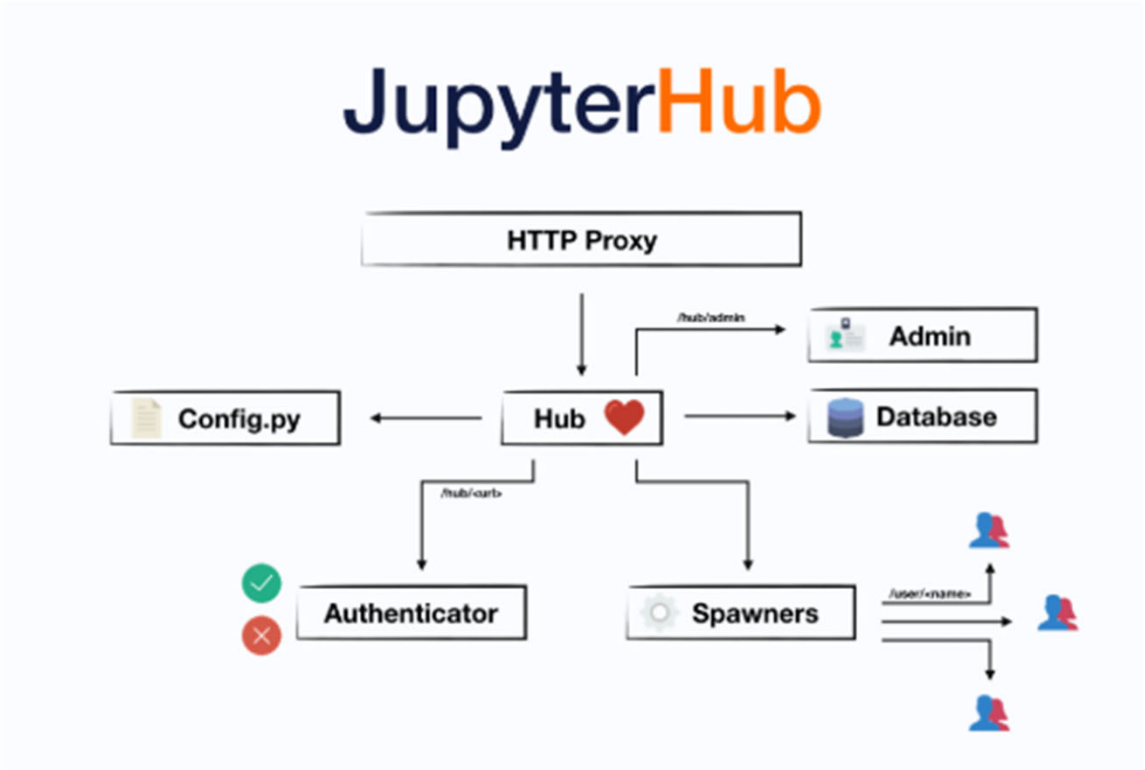
- Interdisciplinary course for students in the sciences, engineering and humanities
- Goals
  - Prepare
  - Empower
  - Enable


DS-GA 1007
DS-UA 1001
DS-UA 1003
DS-UA 1004
DS-UA 3001
Spring 2020

Lesson

Lesson

Lesson



https://pds-f19.jupyter.hpc.nyu.edu

Demo

- Manipulating data in a table
  - How to summarize information?
  - How can text be data?

Demo

- Manipulating data in a table
  - How to summarize information?
  - How can text be data?

- Goals
  - Packages

Demo

- Manipulating data in a table
  - How to summarize information?
  - How can text be data?

- Goals
  - Package
  - Variable

Demo

- Manipulating data in a table
  - How to summarize information?
  - How can text be data?

- Goals
  - Package
  - Variable
  - Output

Demo

- Set-Up
  - Launch in Browser
  - Files and Running Files
  - New Folder and New Notebook with Specific Kernel
- Apps
  - Notebooks
  - Text Editor
  - Command Line Interface
- Python Versions 2/3

Demo

- Cell Types
  - Markdown
  - Code
  - Magic (line or Cell)
- Cell Mode
  - Edit Command
  - Command

Demo

- Getting Help
  - %quickref, help(), ?
  - %lsmagic
  - Tab completion, Introspection (keyboard shortcuts)
- Exporting
  - HTML
  - Python Script
  - PDF