**Lecture 02 – notes and reminders**

File Edit Options Buffers Tools Help

Lecture topic(s): Introduction to GitHub and Jetstream environments; Introduction to free-of-charge data commons (Dataverse); Introduction to the MapReduce paradigm

Assignment topic(s): (2) Parse text files and use a dictionary to convert Unicode strings to ASCII.

Jetstream: Not yet discussed

Project: Not yet discussed

Paper: Jeffrey Dean and Sanjay Ghemawat “MapReduce: Simplified Data Processing on Large Clusters”

Video: What we learned from 5 million books -- https://www.ted.com/talks/jean\_baptiste\_michel\_erez\_lieberman\_aiden\_what\_we\_learned\_from\_5\_million\_books

Students’ to-do list:

* Make sure your environment is properly working
  + Git and GitHub
  + Python
  + Anaconda
* Learn about free-of-charge data storage (i.e., Dataverse)
* Learn how to process text
* Complete Assignment 2 and upload the solution in GitHub --- DUE next Friday (lecture 3) at 8 AM ET --- You have a private repository for your assignments
  + This assignment counts toward your final score. You will get an F in the course if you miss the submission.
* Read the paper on MapReduce
* LAST CALL FOR YOUR GITHUB ACCOUNT: Submit your GitHub account to the instructor by completing this form (DUE TODAY): <https://forms.gle/my61LrtNj1zXHkmEA>
  + The repository will become private on Monday, and you will only have access if we add you to the group.

Before the following lecture:

* Pull Lecture 3
* Pull Assignment 3
* Read Assignment 3 and complete the google form –-- DUE next Friday (lecture 3) before noon ET-- https://forms.gle/7uS9FdEZj4fL3PA77

“Every week, Professor Michela Taufer will share an assignment three days before the class. Each **student must read the assignment and submit up to 3 questions or comments before the lecture so that these questions can be discussed during the breakout sessions.**

**Submit on Friday before noon ET.**”