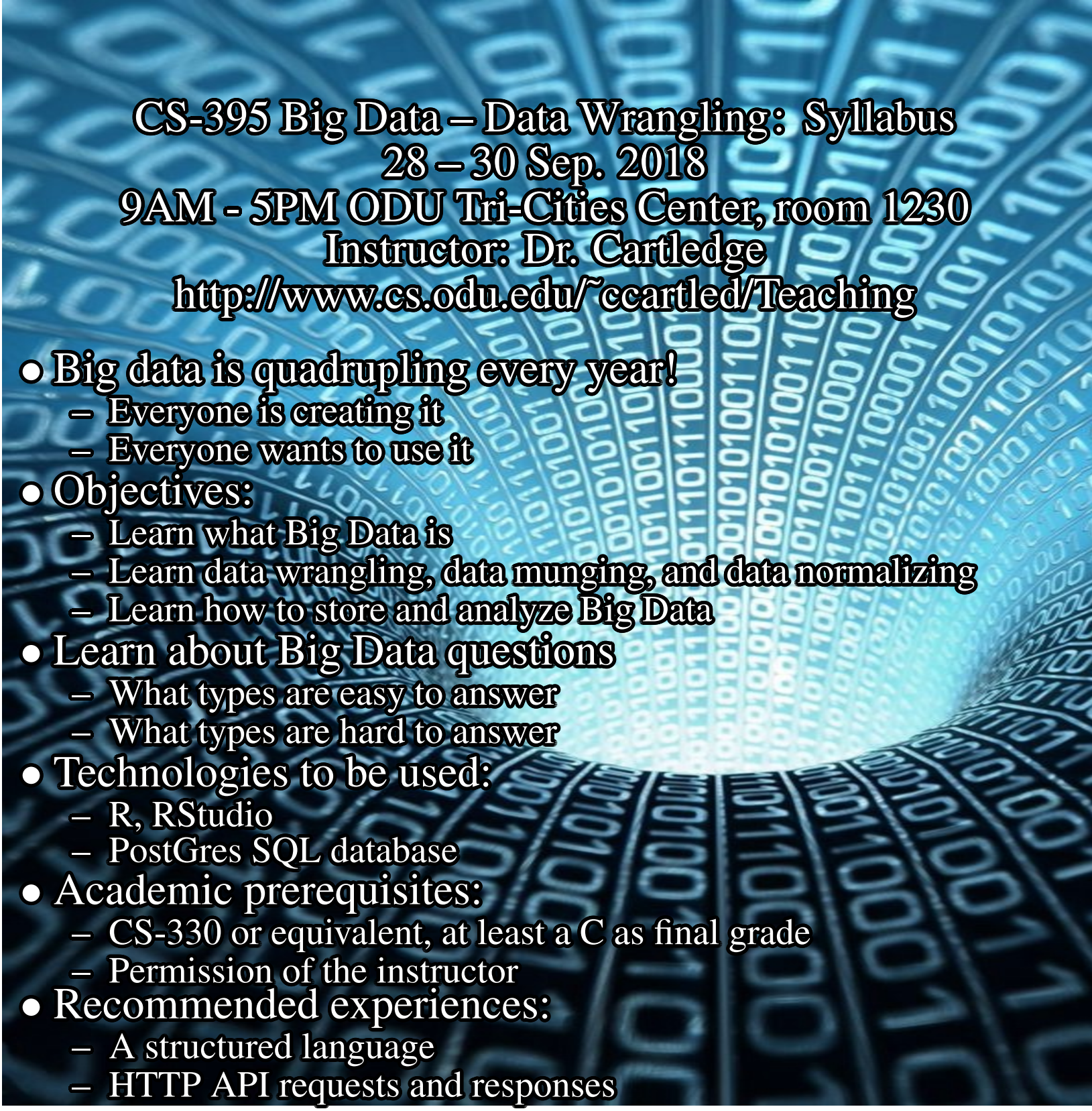

Other notes:

- This is a hands-on programming course. You will establish a Twitter developer's account.
- You will use Twitter's API to search for specific data.
- You will write real-time programs to conduct sentiment analysis of tweets.
- You will monitor sentiment over time.
- You will use different languages.

Contents

1 Course description

The field commonly known as "Big Data" has its origins in the world of business mergers and acquisions. Doug Laney in 2001 was credited with identifying the original 3Vs of Big Data: volume, velocity, and variety. Over the course of time, additional Vs expanded the original three, and the original definitions changed and evolved as well. In this course, we will cover aspects common to all Big Data investigations, including: defining Big Data, surveying tools and techniques for processing Big Data, and visualizing selected aspects of Big Data. We will focus primarily on velocity and variety.

The course will use tweets from Twitter as a standin for velocity, and use the text from the tweets to provide variety. Selected tweets will be downloaded in real-time, stored in a local database, and then analyzed in near real-time. Ideas, concepts, and experiences gained from intensive classroom training and programming projects will be generalized to enterprise level systems to solve problems at scale.

2 Course outline

Upon completion of this course students will be able define and characterize Big Data, analyze, design, and implement real-time programs that retrieve data using HTTP based application program interfaces (APIs), store data in an SQL database, analyze the data to detect sentiment and trends. More specifically a student will be able to:

1. Enumerate the characteristics that qualify a problem as a Big Data problem requiring Big Data tools and techniques,

2. Identify issues associated with handling and manipulating Big Data,

3. Collect real-time data via HTTP based APIs using different languages,
4. Store real-time data into different databases,
5. Analyze real-time data with different languages, and
6. Visualize real-time data

3 Assignments

There will be invidual programming assignments addressing different aspects of real-time Big Data. These include

- Collecting,
- Storing,
- Parsing, and
- Visualization.

There will be three assignments. They are:

1. A paper describing the general aspects of Big Data (- 10 pages, due at the start of the first contact period),
2. Satisfactory attendance and participation in the Boot Camp,
3. A final assignment based on discussions with the instructor. Either
   (a) A system design and demonstration document showing how to scale the system(s) developed during the contact period to enterprise size (due 2 weeks after the contact period), or
   (b) A data wrangling programming assignment of comparable scale and difficulty as those taught during the boot camp.

4 Grading

4.1 Overall grading scale

Overall grade for the course will be based on the student's performance in: class attendance and participation (20%), assignments (40%), reports (40%).

The grading scale follows:

Table 1: Grading scale

Range	Grade	Grade points
94 - 100	A	4.00
90 - 93	A-	3.70
87 - 89	B+	3.30
82 - 86	B	3.00
80 - 81	B-	2.70
77 - 79	C+	2.30
73 - 76	C	2.00
70 - 72	C-	1.70
67 - 69	D+	1.30
63 - 66	D	1.00
60 - 62	D-	0.70
0 - 59	F	0.00
N/A	WF	0.00

4.2 Late assignments

Assignments are due by midnight of the due date. The time of submission is the timestamp of the e-mail saying that the submission is ready. Assignments that are late will be penalized at the rate of one half of a letter grade per 24 hour period. Late submissions will be accepted up to 4 days late (see Table 2).

Table 2: Late submission maximum grade.

Hours late	Max. grade
0	A
24	A-
48	B+
72	B
96	B-
>96	F

5 Course Policies

5.1 Attendance Policy

You are responsible for the contents of all lectures. If you know that you are going to miss a lecture, have a reliable friend take notes for you although slides will be available. Of course, there is no excuse for missing due dates or exam days. During lectures, we will be covering material from the textbook. Lectures will also consist of the exploration of real world problems not covered in the book. You will be given a reading assignment at the end of each lecture for the next class.

I expect you to attend class and to arrive on time. Your grade may be affected if you are consistently tardy. If you have to miss a class, you are responsible checking the course website to find any assignments or notes you may have missed.

5.2 Classroom Conduct

Be respectful of your classmates and instructor by minimizing distractions during class. Cell phones must be turned off during class.

5.3 Seeking Help

The course website should be your first reference for questions about the class. Announcements will be posted to the course website. The best way to get help is to set up an appointment for a Skype or Google+ conference

I will be establishing virtual office hours using Skype, Google+ as DrChuckCartledge, and will use Google calendar to coordinate. I am available via email, but do not expect or rely on an immediate response.

5.4 Disability Services

In compliance with PL94-142 and more recent federal legislation affirming the rights of disabled individuals, provisions will be made for students with special needs on an individual basis. The student must have been identified as special needs by the university and an appropriate letter must be provided to the course instructor. Provision will be made based upon written guidelines from the University's Office of Educational Accessibility. All students are expected to fulfill all course requirements.

6 Academic Integrity / Honor Code

By attending Old Dominion University you have accepted the responsibility to abide by the honor code. If you are uncertain about how the honor code applies to any course activity, you should request clarification from the instructor. The honor pledge is as follows:

> *"I pledge to support the honor system of Old Dominion University. I will refrain from any form of academic dishonesty or deception, such as cheating or plagiarism. I am aware that as a member of the academic community, it is my responsibility to turn in all suspected violators of the honor system. I will report to Honor Council hearings if I am summoned."*

In particular, submitting anything that is not your own work without proper attribution (giving credit to the original author) is plagiarism and is considered to be an honor code violation. It is not acceptable to copy source code or written work from any other source (including other students), unless explicitly allowed in the assignment statement. In cases where using resources such as the Internet is allowed, proper attribution must be given.

Any evidence of an honor code violation (cheating) will result in a 0 grade for the assignment/exam, and the incident will be submitted to the Department of Computer Science for further review. Note that honor code violations can result in a permanent notation being placed on the student's transcript. Evidence of cheating may include a student being unable to satisfactorily answer questions asked by the instructor about a submitted solution. Cheating includes not only receiving unauthorized assistance, but also giving unauthorized assistance. For class files kept in Unix space, students are expected to use Unix file permission protections (chmod) to keep other students from accessing the files. Failure to adequately protect files may result in a student being held responsible for giving unauthorized assistance, even if not directly aware of it.

Students may still provide legitimate assistance to one another. Students should avoid discussions of solutions to ongoing assignments and should not, under any circumstances, show or share code solutions for an ongoing assignment. All students are responsible for knowing the rules. If you are unclear about whether a certain activity is allowed or not, please contact the instructor.

7 Boot Camp Schedule

A detailed class schedule is provided in Table 3.

Table 3: The 3 day boot camp schedule. Attendees who are taking the boot camp as part of a 2 credit course will have pre and post boot camp assignments. The texts will be used extensively during days 2 and 3[**?**, **?**].

Day 1	**Day 2**	**Day 3**
What is Data Wrangling? What is BD? What are BD Vs?	Textual sentiment analysis on Tweets using	Processing of static web pages using CSS selectors and R
Lunch	Lunch	Lunch
BD sources. BD tools. Create Twitter development accounts.		 Conclusion. Presentation of certificates.