Data Science
Course Logistics

Risa Myers

Rice University

Class syllabus

- Syllabus
- Logistics
- Communication
- Attendance
- Teaching mechanisms
- Grading
- Honor code
- Letters of recommendation

- Time for Students / Student / Office hours
    - TBD
    - Or by appointment
    - TA times TBD
    - Watch Piazza for updates

Communications

- Canvas `canvas.rice.edu`
    - Graded work
    - Grades
    - Lecture notes
- Piazza [1]
- Email
    - COMP 553 in subject line

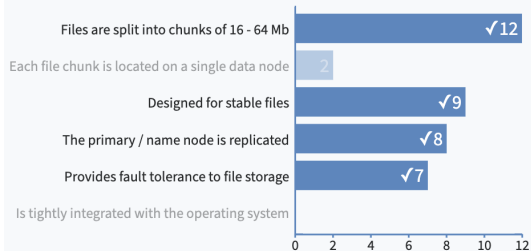[1] contribution version this semester

Attendance

- Strongly encouraged
- Required for lab days (don't be late)
- Required for quiz days
- Excused absences (e.g. athletic events, conferences, etc.) require 48 hours notice
- Expected to have access to Rice resources through the date & time of the course final

- Polls
- Practice problems / Think-Pair-Share
- In class exercises
- Muddiest point
- Flipped classroom

■ Pollev.com/risamyers905

Which are true of HDFS

Files are split into chunks of 16 - 64 Mb	√12
Each file chunk is located on a single data node	2
Designed for stable files	√9
The primary / name node is replicated	√8
Provides fault tolerance to file storage	√7
Is tightly integrated with the operating system	

0 2 4 6 8 10 12

1. Quietly **Think** about the question / problem posed on your own. Formulate a response
2. **Pair** up with your neighbor to discuss
3. **Share** what you learned with the class

- Opportunity to work out a problem and ask questions
- Often combined with Think–Pair–Share

Muddiest point

- Opportunity to anonymously ask for clarity on a topic covered in class
- I will review the questions and provide supplemental resources online or review in a future class

Flipped classroom

- "Flipped" means there is prep to do before coming to class
- Labs in the course are flipped

- Labs
    - 10 short activities to get initial hands-on experience with a practical concept
- Assignments
    - 8-9 theoretical or programming assignments of varying complexity
    - Number of points is proportional to expected level of effort
- Quizzes
    - Short, closed book, in-class quizzes demonstrating your understanding of each assignment and material covered for that assignment

- Labs - 10%
- Quizzes - 20%
- Assignments - 70%
- Grades
    - 93+: A
    - 90: A-
    - 87: B+
    - 83: B
    - 80: B-
    - ...
    - 70: C-
    - ...
    - 60: D-
    - May be scaled

- Methods matter!
- Building blocks for later work
- Approaches that deviate from those covered in class may be given reduced credit

- **Assignment** due dates
  - Typically due BEFORE 5:00 PM
  - Final assignment is due at the **end of the course final exam slot**
  - 1 second – 24 hours late = 10% penalty
  - 24 hours + 1 second – 48 hours late = 20% penalty
  - > 48 hours late: NOT ACCEPTED
  - There are exceptions (e.g. typically the assignment(s) due on the last day of class and during finals may not be turned in late)
- Canvas is the time keeper – if Canvas says it's late, it's late
- Exceptions will only be made for EXTENDED Canvas outages
- Submit early (and often)!
- CHECK THAT YOU SUBMITTED THE RIGHT FILE

- In class
- On paper
- Closed book
- First school day after last day to turn in assignment
- Assignment grade may go up or down, based on how you do
- Covers the assignment and material for the assignment

- 3 per person
- Used in full day increments
- Automatically credited until used up
- May be used only on Assignments
- There are exceptions (e.g. typically the assignment(s) due on the last day of class and during finals may not be turned in late)

- Must be requested within 3 days of work being returned
- Intended for errors in grading or MINOR errors
- Not an extension to the assignment
- Process
    - Meet with me to discuss
    - Write up request with complete details
    - Email to instructor with "COMP 553 regrade" subject

Assignments

1. Python
2. Relational Calculus / Relational Algebra
3. Declarative SQL
4. Imperative SQL programming
5. pandas / command line
6. Spark counting
7. Spark kNN
8. Outliers
9. Gradient descent

- Can discuss any course material (except quizzes, during the quiz) with other classmates

- Rice Honor Code
- No code sharing of ANY kind
    - Email
    - Whiteboard
    - Sharing / showing your screen
    - Piazza posts
    - Verbally

- No outside help
    - No StackOverflow or other posting
    - No Googling solutions
    - No asking someone who took the class last year/semester
- "2 line rule" WITH CITATION
- OK (encouraged) to look up syntax
- CITE your source

Letters of recommendation

- Participate!
- In class
- During Student / Office hours
- On Piazza
  - Answer each other's questions
  - Start discussions

PollEv.com/risamyers905

1. What's the best way to get an answer to a question?
   - A Google
   - B Post on Piazza
   - C Email the instructor
   - D Go to student/office hours

2. How late can I turn in an assignment and still get credit?
   - A Late assignments are not accepted
   - B Typically 1 day
   - C Typically 2 days
   - D The last day of finals

Questions?

?