

Summer 2017 - W205 – Storing and Retrieving Data
Week 6 Live Class Session Agenda
Kevin R. Crook

- Survey – please fill out the survey today – we will spend the first 10 minutes of class allowing students time to fill out the survey
 - W205.5 <https://www.surveymonkey.com/r/56JDVZ3>
 - W205.6 <https://www.surveymonkey.com/r/569YSPT>

- This coming week:
 - Unit 6 – Data Processing and Aggregation
 - Break from labs – next lab due is Lab 6 on July 2nd.
 - Exercise 1 – continue working on it
 - Project - start working on forming groups and selecting a problem area
 - Team selection
 - 2 or 3 students
 - Mix domain knowledge
 - helps if 1 person has domain knowledge of the subject area
 - Mix technical skills
 - Examples: 1 person strong in Linux, 1 person strong in SQL, 1 person strong in Python, 1 person strong in Machine Learning, etc.
 - Take into consideration how to mix your domain knowledge person with technical skills people

- Exercise 1
 - To stay on track, you should have all of your data loaded using schema-on-read in Hive, be actively querying, be working on the analytics including ERD and query or Python.

- Spark web based user interface

- Review asynchronous material – Unit 5 – Ingestion: Storage and Maintenance

- Lab 4 – any questions?

- Project – briefly discuss

- Exercise 1 – any questions?

- Office hours after class