

Schedule (updated 2016-11-15)

All required readings should be completed by the following week.

Date	Topic / Readings	Deadlines
2016-08-30	<p>Introductions; computing setup: Jupyter notebook and command line shell basics; Git and GitHub basics.</p> <p><u>Readings for next week:</u> Required: Software Carpentry Lesson: The Unix Shell, http://swcarpentry.github.io/shell-novice/</p> <p>Required: Roger Peng on Reproducible Research (three videos): http://tinyurl.com/jhu-reproducible-research</p> <p>Optional: Software Carpentry Lesson: Version Control with Git, http://swcarpentry.github.io/git-novice/</p>	Exercise #1, Friday, 9/2, 12pm
2016-09-06	<p>The command line shell: input, output, and pipelines; csvkit; data types.</p> <p><u>Readings</u> Required: Wickham, "Tidy Data." http://vita.had.co.nz/papers/tidy-data.pdf</p> <p>Optional: Data Science at the Command Line, chapters 1-5</p>	Exercise #2, Friday, 9/9, 12pm
2016-09-13	<p>Command line filters in the shell and Python; parallel processing in the shell.</p> <p><u>Readings</u> Required: Software Carpentry Lesson: Using Databases and SQL, Topics 1-5, http://swcarpentry.github.io/sql-novice-survey/</p> <p>Optional: Data Science at the Command Line, chapters 6-8</p>	Project #1, Friday, 9/23, 12pm
2016-09-20	<p>RDBMS: schema, keys, basic SQL operations, aggregate functions.</p> <p><u>Readings</u> Required: Software Carpentry Lesson: Using Databases and SQL, Topics 6-10, http://swcarpentry.github.io/sql-novice-survey/</p> <p>Optional: Learning SQL, chapters 1-4; Database System Concepts, chapters 1-3</p>	Review #1, Tuesday, 9/27, 7pm
2016-09-27	<p>RDBMS: subqueries, joins, integrity, transactions, functions, triggers, schema design and E-R models, normal forms.</p>	Exercise #3, Friday 9/30, 12pm

	<u>Readings</u> Optional: Learning SQL, chapters 5, 6, 7, 9, 10 Optional: A Gentle Introduction to Algorithm Complexity Analysis (online at http://discrete.gr/complexity/) Optional: Visualizing Algorithms (online at http://bost.ocks.org/mike/algorithms/)	
2016-10-04	RDBMS: advanced SQL, ETL, indexes, query processing, analysis, and optimization, SQL from Python. Note: no office hours on Tuesday, October 4. <u>Readings</u> Required: Star Schema, chapters 1-5 Optional: Learning SQL, chapters 12, 13, 14	
2016-10-11	No class Note: no office hours on Tuesday, October 11.	
2016-10-18	Warehouses: facts and dimensions, architectures, schemas <u>Readings</u> Required: Star Schema, chapters 4-7	Exercise #4, Monday, 10/24, 12pm
2016-10-25	No class (fall break)	
2016-11-01	Warehouses: dimension design <u>Readings</u> Required: Star Schema, chapter 11 Required: AWS Redshift. https://aws.amazon.com/redshift/	Project #2, Friday, 11/11, 12pm
2016-11-08	Warehouses: fact table design <u>Readings</u> Required: Dean and Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters." http://research.google.com/archive/mapreduce.html Required: Drake, "Command-line tools can be 235x faster than your Hadoop cluster." http://aadrake.com/command-line-tools-can-be-235x-faster-than-your-hadoop-cluster.html	Review #2, Tuesday, 11/15, 7pm

	<p>Optional: Chang et al. "Bigtable: A Distributed Storage System for Structured Data." http://research.google.com/archive/bigtable.html</p> <p>Optional: DeCandia et al. "Dynamo: Amazon's Highly Available Key-value Store", http://www.read.seas.harvard.edu/~kohler/class/cs239-w08/de-candia07dynamo.pdf</p>	
2016-11-15	<p>Contemporary data management tools: Map/Reduce, Hadoop, Trifacta</p> <p><u>Readings</u> Required: Apache Spark. https://spark.apache.org/ Required: Lambda Architecture. http://lambda-architecture.net/</p>	<p>Exercise #5, Friday, 11/18, 12pm</p> <p>Final Project, Friday 12/9, 12pm</p>
2016-11-22	<p>Contemporary data management tools: Spark introduction</p> <p><u>Readings</u> Required: CAP theorem. https://en.wikipedia.org/wiki/CAP_theorem Required: Kudu. http://getkudu.io/ Required: AWS Kinesis. https://aws.amazon.com/kinesis/</p>	<p>Exercise #6, Tuesday 11/29, 12pm</p>
2016-11-29	<p>Contemporary data management tools: Spark SQL, DataFrames, MLib, Streaming</p>	<p>Final Project presentations Tuesday, 12/6</p>
2016-12-06	<p>Final Project presentations, course wrap-up</p>	<p>Review #3, Tuesday, 12/12 12pm</p>