236620 Big Data Technologies Spring 2016 Syllabus

1. Chapter 1: Introduction (two weeks)
   • Week 1: course administration and logistics; intro lecture describing course outline, sampling a little bit of everything.
   • Week 2: principles of large scale systems: scalability, consistency, fault-tolerance, workload optimization.

2. Chapter 2: Batch data processing and analytics – Map/Reduce, Hadoop and beyond
   • Week 3: real-life problem motivating Map-Reduce (indexing), and the Map-Reduce programming paradigm at a high-level.
   • Week 4: Map-Reduce implementation in depth.
   • Week 5: high-level analytical languages (e.g. Pig, Hive). High-level abstractions (select, join, aggregate), and how they map to Map-Reduce.

3. Chapter 3: OLTP via NoSQL databases
   • Week 6: real-life use case that requires scalable key/value stores – maintaining a user profile store.
   • Week 7: OLTP basics, BigTable, HBASE.

4. Chapter 4: Stream processing
   • Week 8: canonical stream mining problems (approximate counting, frequent itemsets).
   • Week 9: Apache Storm.
   • Week 10: online optimization - multi-armed bandits.

5. Chapter 5: Misc
   • Week 11: motivation for incremental processing; transactions for NoSQL (in a nutshell), Percolator, Omid, Sieve.
   • Week 12: controlled experiments on the Web, A/B testing.
   • Week 13: real-life recommender system – architecture and challenges; short course summary.