System course plan

        overview of system challenges (20min)

o   Why big data is different: (cost, efficiency), Scale->Geo, Mangement, Fault as constant, ->replication, redundancy, geo,

o   Stack: storage, computation, networking, scheduling, management, monitoring

        storage systems: GFS, (25 + 45)

o   consistency trade off

o   Replication,

o   Master (Paxos) chunk management unit

o   Better Design Colossus  (Flat Datacenter Storage, Ed Nightingale)

        BigTable, Dynamo, etc (45)

o   Different abstract

o   ACID (log, commit)

o   Bigtable design, tablet, reuse replication, master, minor compaction, major compaction,

o   CAP theorem

o   Dynamo weaken consistency

o   Advanced Ideas: Spanner, MegaStore, F1, AzureStore

        computation paradigms

o   Map Reduce  (45min)

o   Dryad (20)

o   High Level: Sawsal, Flumejava, DryadLinq, PigLatin, Scope.  (25)

        Management (45)

o   Resource Management (mesos, omega, yarn, Quincy, Berkeley fast scheduler)

o   Multi-tenency (Google)

        data center construction, networking and special hardware (45)

o   Networking (SDN, FatTree, ….)

o   Special Hardware (FPGA, ARM Storage, Customized SSD, GPU)

        Advanced (90 min)

o   Spark and related (RDD),

o   \*RAMCloud

o   Graph Computation (Pregel BSP, GraphChi, GraphLab/PowerGraph=>Async, GraphX, EdgePartion/Vertex Partition)

o   Stream Computation (Storm, xxx)