A tentative paper list with the presentation and discussion date (*tentative, subject to change*), the name of designated student will be added to each paper later.

1. (Jan 30) MapReduce-Simplified Data Processing on Large Clusters *(Vahid Behzadan)*

**Big Data Storage**

2. (Feb 1) Cassandra - A Decentralized Structured Storage System (short) + Hive A Warehousing Solution Over a MapReduce Framework (short) *(Vladislav Savranschi)*
3. (Feb 1) The Hadoop Distributed File System *(Joe Ravenna)*
4. (Feb 6) Bigtable: A Distributed Storage System for Structured Data *(Stephen Williams)*
5. (Feb 6) MD-HBase: A Scalable Multi-dimensional Data Infrastructure for Location Aware Services *(Ahsan Ali)*
6. (Feb 8) Analysis of HDFS Under HBase: A Facebook Messages Case Study *(Amirhesam Yazdi)*
7. (Feb 8) Pyro: A Spatial-Temporal Big-Data Storage System *(Syed Zawad)*

**Big Data Processing Frameworks**

8. (Feb 13) Improving MapReduce Performance in Heterogeneous Environments *(Biplav Timalsina)*
9. (Feb 13) Delay Scheduling: A Simple Technique for Achieving Locality and Fairness in Cluster Scheduling *(Maxwell Wiegant)*
10. (Feb 15) Starfish: A Selftuning System for Big Data Analytics *(Jalal Kiswani)*
11. (Feb 15) DyScale: a MapReduce Job Scheduler for Heterogeneous Multicore Processors *(instructor)*
12. (Feb 22) Apache Hadoop YARN: Yet Another Resource Negotiator *(Di Li)*
13. (Feb 22) Spark: Cluster Computing with Working Sets *(Megan Nilsen)*
14. (Feb 27) Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing *(Sui Cheung)*
15. (Feb 27) Mesos-A Platform for Fine-Grained Resource Sharing in the Data Center *(Naresh Kumar Giri)*
16. (March 1) Omega flexible, scalable schedulers for large compute clusters *(Tyler Defoor)*
17. (March 1) Multi-Resource Fair Sharing for Datacenter Jobs with Placement Constraints *(Gautham Yerroju)*

**Big Data Analytics Tools**

18. (March 6) Dynamo: Amazon’s Highly Available Key-value Store *(Eric Garza)*
19. (March 6) Building a High-Level Dataflow System on top of Map-Reduce: The Pig Experience *(Tianyi Jiang)*
20. (March 8) Dremel: Interactive Analysis of Web-Scale Datasets (Avishek Bose)
21. (March 8) Megastore: Providing Scalable, Highly Available Storage for Interactive Services (Sonu Jose)
22. (March 13) Spanner: Google’s Globally-Distributed Database (Tehman Tariq)
23. (March 13) S4: Distributed Stream Computing Platform (Emerson Havener)
24. (March 15) Storm @Twitter (Phillip Vong)
25. (March 15) Pregel: A System for Large-Scale Graph Processing (Brian Smith)
26. (March 27) GraphX: Unifying Data-Parallel and Graph-Parallel Analytics (Monique Dingle)
27. (March 27) One Trillion Edges: Graph Processing at Facebook-Scale (Tim Kwist)

**Big Data Machine Learning**

28. (March 29) A Few Useful Things to Know about Machine Learning (Ryan Lieu)
29. (March 29) Large Scale Distributed Deep Networks (Athira Pillai)
31. (April 3) Scaling Distributed Machine Learning with the Parameter Server (Kevin Muran)
33. (April 5) SERF: Efficient Scheduling for Fast Deep Neural Network Serving via Judicious Parallelism (instructor)
34. (April 10) DjiNN and Tonic: DNN as a Service and Its Implications for Future Warehouse Scale Computers (Liam Gomez)
35. (April 10) TensorFlow: A System for Large-Scale Machine Learning (Alex Buettner)
36. (April 12) Automating Model Search for Large Scale Machine Learning (Sankeerth Ankam)