CSC 9010-003: Seminar in Big Data and Data Science

Fall, 2014: Pool of Topics

|  |
| --- |
| **Group 1** |
| Big Data | Data Science | NoSQL; types of systems | Brewer’s CAP Theorm; consistencies |
| Map Reduce |  |  |  |
| **Group 2** |
| Data Visualization | Cloud Computing | Hadoop & Mahout |  |
| **Group 3** |
| Big Table & GFS | HDFS | The Hadoop “Ecosystem” | HBase |
| Hadoop Pig | AWS |  |  |
| **Group 4** |
| JSON & XML | data mining | parallelism | statistics with R |
| social networks | benchmarks | large datasets for free |  |
| Cassandra | MegaStore | Riak | CouchDB |
| MongoDB | Project Voldemort | Lucene | Sphinx |
| Sensage | Splunk | Redis | MySQL Cluster |
| DynamoDB | MemcacheDB | Zookeeper | (AWS) SimpleDB |
| (AWS) Elastic MR |  |  |  |