

CS 686: Special Topics in Big Data

RDDs

Lecture 28

Daily Cluster Status Report

- Just joking

Tweets

- "Fundamental data structure of Spark – RDD"
- "RDDs: Big Data Wrangling and Machine Learning Analysis, all in one system"
- "A resilient distributed dataset is your safest bet"
- "RDD: Keep all your data in the cache, wish your system has enough memory"
- "Exploit RDDs in batch analytics to produce a Spark!!"

Today's Discussion

- Groups of 4
- See:
 - <https://classroom.github.com/g/Wx38GHNW>