



Mid-Term Exam Review

Mid-Term Exam Details

- o Exam covers all the lectures so far except Advanced Pig.
- o Exam time – 2.5 hrs
- o Total points - 60
- o Total questions – 6
- o 10 MCQs
- o Closed book/Closed notes

Lesson-wise Points Distribution

- o HDFS & YARN – 10 points
- o MapReduce – 25 points
- o Apache Avro – 10 points
- o Apache Pig – 15 points

Topics Covered

o HDFS

- o Blocks, replications, replica placement policy
- o Pros and cons of HDFS
- o HDFS Daemons and their workings

o Hadoop 1 and Hadoop 2

- o Architecture
- o Data flow basics
- o MRv1 and MRv2 (YARN) Daemons and their job
- o HDFS Federation
- o NN HA

Topics Covered

o MapReduce

- o Signature of Map & Reduce methods
- o Combiner and Partitioner
- o In-mapper combiner
- o Shuffle/Sort

o Avro

- o How to write schema file
- o Use of AvroJob, AvroKey and AvroValue in code
- o Schema Evolution

Important Topics

- o Apache Pig
 - o Pig Latin Operators and Functions
 - o LOAD, DUMP, STORE, GROUP BY, GROUP ALL, FOREACH GENERATE, FILTER BY, ORDER BY, JOIN, DISTINCT, COUNT, SIZE, MAX, MIN...
 - o Diagnostic Operators
 - o Case Sensitivity

**Make sure to do TM before
coming for the Exam**

Good Luck!