

Dataset Download :

Dataset for the problem is a dataset on Airports which can be downloaded from moodle.

Problem 1:

Write Mapper and Reducer to get the number of Airports by :

1. Country
2. Type

Problem 2:

Write Mapper and Reducer to find the

1. Country
2. Region

having the highest number of airports

NOTE:

For both the problems and each part, write separate Mappers and Reducers and don't mix the problem.

Resources :

1. Unit 2 and Unit 3 from this online course (~ 12 hours). Unit 1 and 4 are not needed.
<https://in.udacity.com/course/intro-to-hadoop-and-mapreduce--ud617>
2. Chapter 6 should suffice which is also free to download. <http://go.cloudera.com/udacity-lesson-2>

Deliverables/Upload Format :

RollNo / ProblemNumber

Mapper.py

Reducer.py

You can upload the code from the Virtual Machine itself.