Parallel and Distributed Computing

Instructions:

1. You MUST use JAVA.

2. Your programs must compile without warnings and execute correctly for full credit. Use good programming style, including the use of an appropriate amount of comments and suitable naming conventions.

Problem Statement

Stack Exchange is a network of question-and-answer websites on topics in diverse fields, each site covering a specific topic, where questions, answers, and users are subject to a reputation award process. The reputation system allows the sites to be self-moderating.

The data files needed are attached.

You can find the metadata information here. You can also google if you have any

confusions. https://ia800107.us.archive.org/27/items/stackexchange/readme.txt

Create a **MapReduce** program for the questions below. You can run your code on standalone (local) Hadoop installation.

1. Find the average length of comments in Comments.txt file.