PROBLEM 1 – Fill in the Blanks

Java code which can be run directly

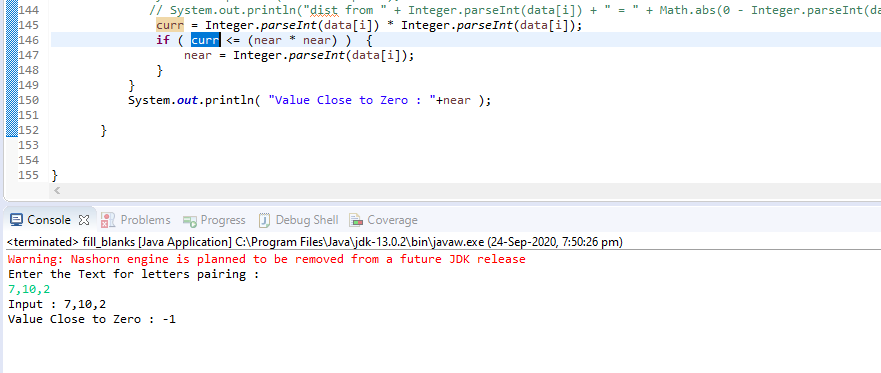
Input 1: -

Enter the Text for letters pairing :

7,10,2

Input : 7,10,2

Value Close to Zero : -1



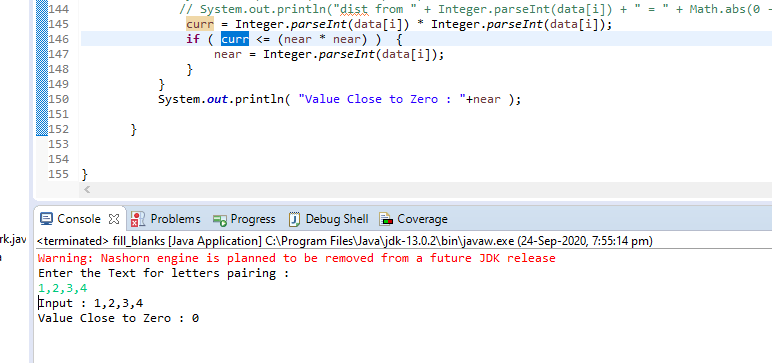
Input 2: -

Enter the Text for letters pairing :

1,2,3,4

Input : 1,2,3,4

Value Close to Zero : 0



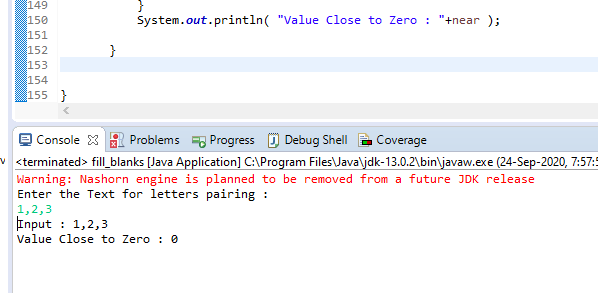
Input 3: -

Enter the Text for letters pairing :

1,2,3

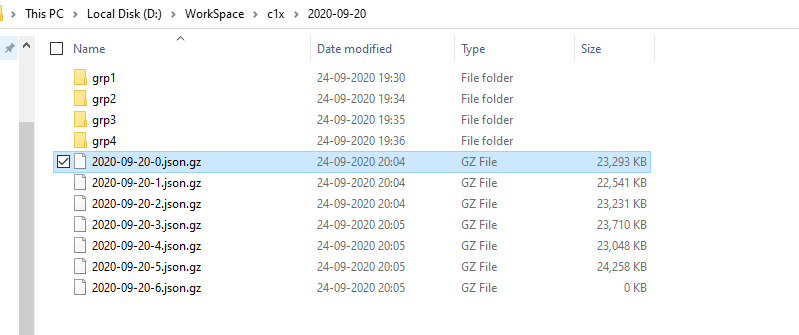
Input : 1,2,3

Value Close to Zero : 0



PROBLEM 2: - GitHub Archive

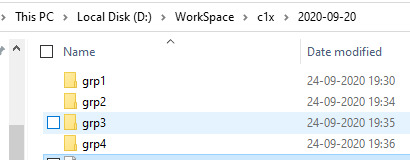
1. Github\_data.java – code which needs the date has input and will download all the data for the date in to the common folder “D:\WorkSpace\c1x\2020-09-20”



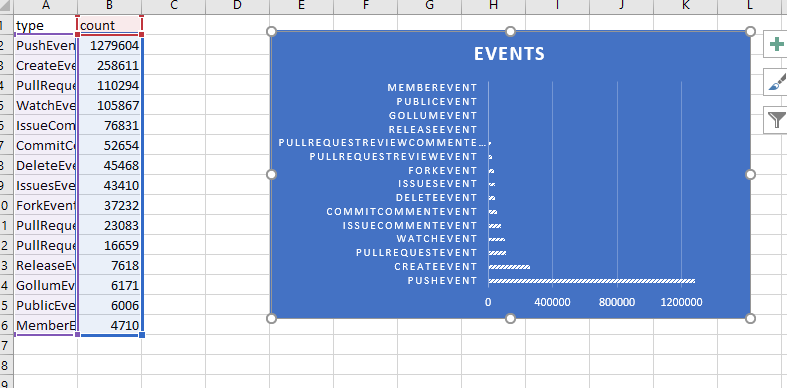
1. Spark\_read\_json.java code – it’s an spark code where we need to pass the date as an argument where from the common folder “D:\WorkSpace\c1x\2020-09-20”, it picks all the files and load in to spark dataset from which have got below 4 scenarios
   1. to get the event counts per day to see which events is performed more on daily basis
   2. to get the event counts per hour to see which events is performed more on hour basis
   3. Issues - Opened, Closed, Reopened
   4. Organization Name from URL - Top 20 Organization

Where each output will be stored on the common folder as csv files

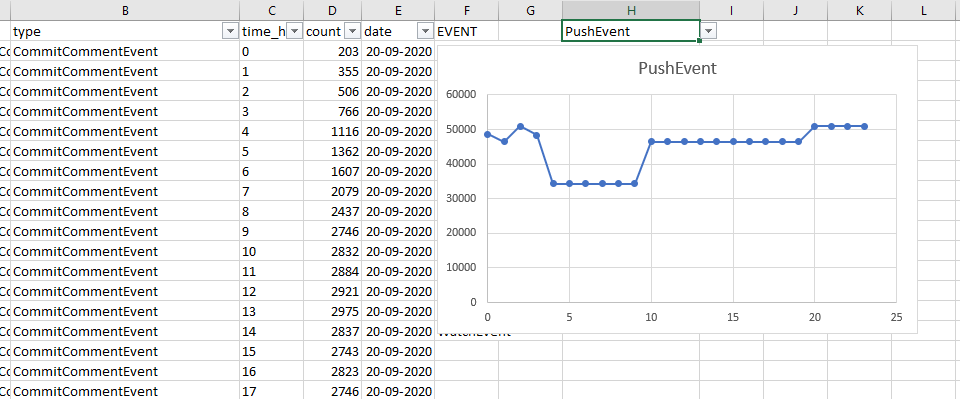
D:\WorkSpace\c1x\2020-09-20\grp1\output\ etc



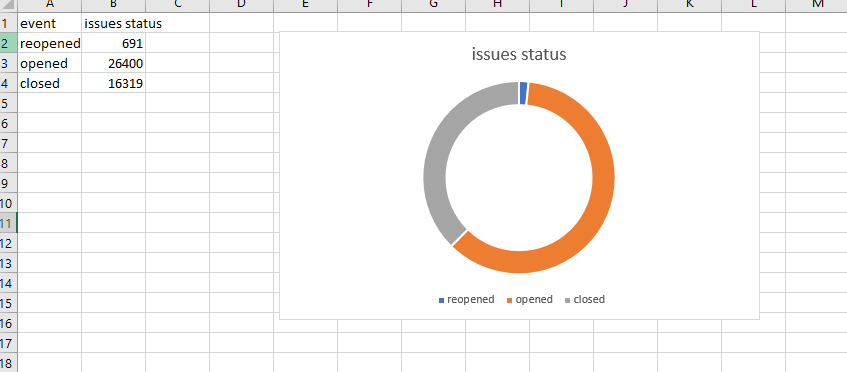
Use Case 1: - to get the event counts per day to see which events is performed more on daily basis



Use Case 2: - to get the event counts per hour to see which events is performed more on hour basis



Use Case 3: - Issues - Opened, Closed, Reopened



Use Case 4: - Organization Name from URL - Top 20 Organization

