Hotel Rating analysis on customer review Using Hadoop.

**Tushar Jalindar Gatkal**
MSc in Big Data

Abstract:

The main aim of the project is to perform analysis on the hotel ratings based on the customer reviews. Initially data is in unstructured format, the data is being distributed among the hotels, further the ratings are being reviewed, based on that the barchart is being generated. Map-reduce is used for scaterring the data,PIG is used for counting of the reviews within one dataset,finally R programming is used for the last stage of analysis.

Input   **Process**   **Result (After Analysis)**

Block Diagram

Methodology:

❏ Map-Reduce is the primary step which is being carried out to get the data seperated for every single hotel in different dataset.

❏ Regular expressions is being used to seperated out the rating number from the given dataset.

❏ Pig is used to collect the rating from the every reviews(eg: review1,review2 etc) and store it within one dataset so it becomes easy to analyse.

❏ The analysed data is being supplied to the Rstudio to generate the final stacked barchart which completes the final analyses part of the project.
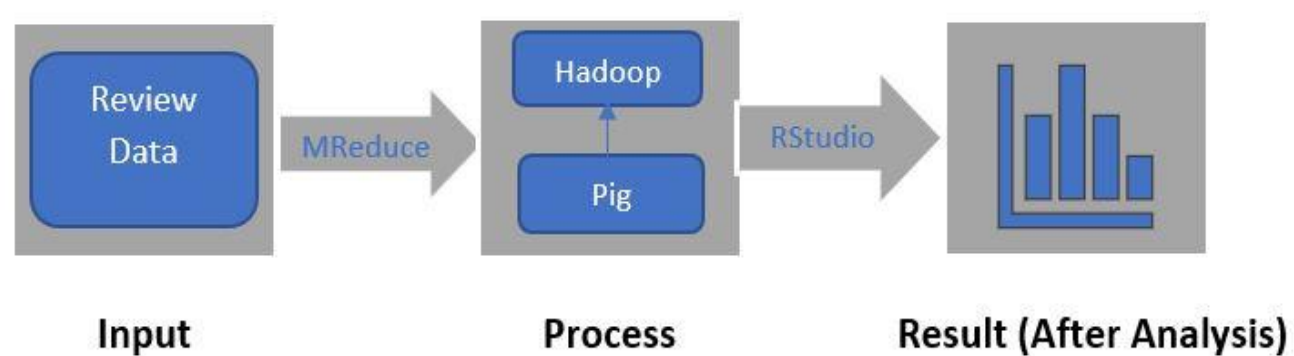
Conclusion:

The unstructured dataset is being provided on which map-reduce,pig,rstudio methodology is being applied to get the analysis of the customer rating for every hotel in the form of stacked barchart.

stir.ac.uk

For further information contact: *Tushar Jalindar Gatkal*

✉ *tug00001@students.stir.ac.uk*

BE THE DIFFERENCE