Big Data Analytics

----BI2 NI He

## Question 1: Count the number of words of each length present in the text (in order to establish a histogram of word lengths).

IntWritable one = new IntWritable(1);

Text wlength = new Text();

map (LongWritable key, Text value, Context context)

{

StringTokenizer itr = new StringTokenizer(value.toString());

while (itr.hasMoreTokens())

{

String wordToCheck = itr.nextToken();

wlength.set(String.valueOf(wordToCheck.length()));

context.write(wlength, one);

}

}

IntWritable result = new IntWritable();

reduce (Text key, Iterable<IntWritable> values, Context context)

{

int sum = 0;

for (IntWritable val : values)

{

sum += val.get();

}

result.set(sum);

context.write(key, result);

}

## Question 2: Count the number of words of 1 to 5 characters (inclusive), 6 to 10 characters (inclusive), 11 to 15 characters (inclusive) and more than 15 characters present in the text.

IntWritable one\_to\_five = new IntWritable(1);

IntWritable six\_to\_ten = new IntWritable(2);

IntWritable eleven\_to\_fifteen = new IntWritable(3);

IntWritable more\_than\_fifteen= new IntWritable(4);

Text wlength = new Text();

map (LongWritable key, Text value, Context context)

{

        StringTokenizer itr = new StringTokenizer (value.toString() );

while (itr.hasMoreTokens())

{

String wordToCheck = itr.nextToken();

                     if (wordToCheck.length() <= 5)

{

                                     wlength.set(String.valueOf(wordToCheck.length()));

                     context.write(wlength, one\_to\_five);

}

                            else if (wordToCheck.length() > 5 && wordToCheck.length() =<10)

{

                                     wlength.set(String.valueOf(wordToCheck.length()));

                     context.write(wlength, six\_to\_ten);

}

                            else if (wordToCheck.length() > 10 && wordToCheck.length() =<15) {

                                      wlength.set(String.valueOf(wordToCheck.length()));

                     context.write(wlength, eleven\_to\_fifteen); }

                            else {

                                      wlength.set(String.valueOf(wordToCheck.length()));

                     context.write(wlength, more\_than\_fifteen); }

        }

    }

}

IntWritable result = new IntWritable();

reduce (Text key, Iterable<IntWritable> values, Context context)

{

int sum = 0;

for (IntWritable val : values)

{

sum += val.get();

}

result.set(sum);

context.write(key, result);

}

## Question 3: Obtain word lists of 1 to 5 characters (inclusive), 6 to 10 characters (inclusive), 11 to 15 characters (inclusive) and more than 15 characters present in the text. It is not required to sort the words within a list, nor to eliminate duplicates.

IntWritable one\_to\_five = new IntWritable(1);

IntWritable six\_to\_ten = new IntWritable(2);

IntWritable eleven\_to\_fifteen = new IntWritable(3);

IntWritable more\_than\_fifteen= new IntWritable(4);

Text word = new Text();

map (LongWritable key, Text value, Context context)

{

        StringTokenizer itr = new StringTokenizer (value.toString() );

while (itr.hasMoreTokens())

{

String wordToCheck = itr.nextToken();

                     if (wordToCheck.length() <= 5)

{

word.set(wordToCheck);

context.write(word, one\_to\_five);}

                            else if (wordToCheck.length() > 5 && wordToCheck.length() =<10)

{

                                     word.set(wordToCheck);

context.write(word, one\_to\_five);}

}

                            else if (wordToCheck.length() > 10 && wordToCheck.length() =<15)

{

                                      word.set(wordToCheck);

context.write(word, one\_to\_five);}

}

  else {

                                     word.set(wordToCheck);

context.write(word, one\_to\_five);}

}

        }

    }

}

String result = new String();

reduce (Text key, Iterable<IntWritable> values, Context context)

{

        StringBuffer buf=new StringBuffer();

        for (String val : values)

{

Buf += val.get() + ”,” ; // ” /r ”

        }

String result = buf.toString();

         context.write(key, result);

    }

}