Big Data – Hadoop

Hadoop Map-Reduce - Assignment 1

Question:

Find number of occurrences of each word in input file.

Objective:

Write a mapreduce program count number of occurrences of each word in input file.

Description:

The input file contains some random text. The text has multiple lines with each line contains multiple words. Obviously few words are repeated multiple times. Implement hadoop map-reduce program to count occurrence of each word in file.

Algorithm:

- **Input:** File containing some text.
- **Mapper:**
  - Read file line by line.
  - Separate each word from the line.
  - Emit word (as key) and its occurrence (as value). Each occurrence is recorded individually as single occurrence (i.e. occurrence count = 1).
- **Reducer:**
  - Sum up occurrence of each word to count total number occurrences.
  - Emit each word as key and its number of occurrences as value.
- **Output:** File containing word and its number of occurrences.

Example: If input file is:

```
my name is bond
james bond
adventure is my passion
my my
```

Then output should be:

```
adventure   1
bond        2
is          2
james       1
my          4
name        1
passion     1
```

Big Data – Hadoop

Hadoop Map-Reduce - Assignment 2

Question:

Find number of occurrences of each character in input file.

Objective:

Write a mapreduce program count number of occurrences of each character in input file.

Description:

The input file contains some random text. The text has multiple lines with each line contains multiple characters. Implement hadoop map-reduce program to count occurrence of each character in file.

Algorithm:

- **Input:** File containing some text.
- **Mapper:**
  - Read file line by line.
  - Separate each character from the line.
  - Emit character (as key) and its occurrence (as value). Each occurrence is recorded individually as single occurrence (i.e. occurrence count = 1).
- **Reducer:**
  - Sum up occurrence of each word to count total number occurrences.
  - Emit each word as key and its number of occurrences as value.
- **Output:** File containing word and its number of occurrences.

Example:

If input file is:

```
abcdabcd
aaabbbcc
aabbbbcc
```

Then output should be:

a	7
b	9
c	6
d	2

Big Data – Hadoop

Hadoop Map-Reduce - Assignment 3

Question:

Find maximum number from two consecutive columns in file.

Objective:

Implement a map-reduce program to find the maximum integer from two columns in file.

Description:

The input file contains integer values separated by comma. Each line contains 10 integers and there may be number of lines in the file. We should find maximum number from column 1 & 2, 3 & 4, 5 & 6, 7 & 8 and 9 & 10.

Algorithm:

- **Input:** File containing some 10 integers on each line separated by comma.
- **Mapper:**
  - Read file line by line.
  - Separate each integer from the line & store in an array.
  - Find maximum value from two consecutive integers (e.g. 1st & 2nd).
  - Emit column numbers (e.g. 1 & 2) as key and max integer as value.
- **Reducer:**
  - Go through maximum values from each row and find the maximum integer among all of them.
  - Emit column numbers (e.g. 1 & 2) as key and max integer as value.
- **Output:** File containing column numbers and corresponding max integer value.

Example: If input file is:

```
8,4,5,6,10,9,2,3,5,6
1,2,4,3,7,6,8,9,4,5
3,5,2,3,7,5,8,9,9,0
.......
5,7,7,4,8,6,5,7,9,0
```

Then output should be:

```
1st & 2nd Column 8
2nd & 3rd Column 6
....
9th & 10th Column 6
```