Big Data Assignment

Name: Ayalew

ID: 985644

BigData

Week 2 Day 2

Build inverted index using the "Modified" Algorithm.

Input Split 1 (Doc ID 101) [ cat pat mat sat cat eat ]

Input Split 2 (Doc ID 201)  [ pat mat sat pat mat eat]

Input Split 3 (Doc ID 301) [ sat mat cat pat fat mat]

Assume three Mappers and two Reducers.

Let cat mat and sat be processed by Reducer 1 and rest by Reducer 2.

|  |  |  |  |
| --- | --- | --- | --- |
| Mapper 1 Input | Mapper 2 Input | | Mapper 3 Input |
| <(cat, 101), 2 >  <(pat, 101), 1 >  <(mat, 101), 1 >  <(sat, 101), 1 >  <(eat, 101), 1 > | <(pat ,201), 2 >  <(mat, 201), 2>  <(sat, 201), 1 >  <(eat, 201), 1 > | | <(sat, 301), 1 >  <(mat, 301), 2>  <(cat, 301), 1 >  <(fat, 301), 1 > |
| Shuffle and Sort | | | |
| Reducer 1 Input (cat, mat , sat) | | Reducer 2 input (pat, sat, fat, eat) | |
| <(cat, 101), [2] >  <(cat, 301), [1] >  <(mat, 101), [1] >  <(mat, 201), [2]>  <(mat, 301), [2]>  <(sat, 101), [1] >  <(sat, 201), [1] >  <(sat, 301), [1] > | | <(eat, 101), [1] >  <(eat, 201), [1] >  <(fat, 301), [1] >  <(pat, 101), [1]>  <(pat ,201), [2] > | |
| Reducer 1 Output | | Reducer 2 output | |
| <cat, (101,2)→ (301, 1) >  <mat, (101, 1)→ (201, 2)→(301, 2) >  <sat, (101,1)→ (201,1)→ (301,1)> | | <eat, (101,1)→ (201, 1) >  <fat, (301, 1)>  <pat, (101,1)→ (201,2)> | |
|  | |  | |