**Question 1.**

Class Mapper

Method initialize

H= new AssociativeArray

Method Map(docid a, doc d)

For all term w in record r do

For all term u in Window(w) do

H{pair(w;u)}++;

Method Close

For all t in H do

Emit(t, H{t})

**Question 2.**

Class Mapper

Method initialize

G = new Associative Array

Method Map(docid a, doc d)

For all term w in record r do

H= new Associative Array

For all term u in Window(w) do

H{u}= H{u} + 1;

If (G{w} == null) G{w}= H;

Else G{w}= G{w} + H; // Elementwise addition

Method Close

For all key k in G do

Emit (k, G{k});

**Question3.a.**

Number of Input-Splits: 2

Number of Reducers: 2

**Mapper 0 Input**

{cat, mat, rat, cat}

{cat, bat, cat, pat}

{cat, bat, rat, bat}

**Mapper 1 Input**

{cat, rat, bat, rat}

{bat, mat, pat, bat}

{pat, cat, bat, mat}

|  |  |
| --- | --- |
| **Mapper 0 Output** | **Mapper 1 Output** |
| <(cat,mat ),1>  <(cat,rat),1>  <(mat,rat),1>  <(mat,cat),1>  <(rat,cat),1>  <(cat,bat),1>  <(bat,cat),1>  <(bat,pat),1>  <(cat,pat),1>  <(cat,bat),1>  <(cat,rat),1>  <(cat,bat),1>  <(bat,rat),1>  <(rat,bat),1> | <(cat,rat),1>  <(cat,bat),1>  <(cat,rat),1>  <(rat,bat),1>  <(bat,rat),1>  <(bat,mat),1>  <(bat,pat),1>  <(mat,pat),1>  <(mat,bat),1>  <(pat,bat),1>  <(pat,cat),1)>  <(pat,bat),1>  <(pat,mat),1>  <(cat,bat),1>  <(cat,mat),1>  <(bat,mat),1> |
| **Reducer 0 Input (<k)** | **Reducer 1 input (>k)** |
| <(bat,cat),[1]>  <(bat,mat),[1,1]>  <(bat,pat),[1,1]>  <(bat,rat),[1,1]>  <(cat,bat),[1,1,1,1,1]>  <(cat,mat ),[1,1]>  <(cat,pat),[1]>  <(cat,rat),[1,1,1,1]> | <(mat,bat),[1]>  <(mat,cat),[1]>  <(mat,pat),[1]>  <(mat,rat),[1]>  <(pat,bat),[1,1]>  <(pat,cat),[1])>  <(pat,mat),[1]>  <(rat,bat),[1,1]>  <(rat,cat),[1]> |
| **Reducer 0 output** | **Reducer 1 output** |
| <(bat,cat),1>  <(bat,mat),2>  <(bat,pat),2>  <(bat,rat),2>  <(cat,bat),5>  <(cat,mat),2>  <(cat,pat),1>  <(cat,rat),4> | <(mat,bat),1>  <(mat,cat),1>  <(mat,pat),1>  <(mat,rat),1>  <(pat,bat),2>  <(pat,cat),1)>  <(pat,mat),1>  <(rat,bat),2>  <(rat,cat),1> |

**Question 3.b**

Number of Input-Splits: 2

Number of Reducers: 2

**Mapper 0 Input**

{cat, mat, rat, cat}

{cat, bat, cat, pat}

{cat, bat, rat, bat}

**Mapper 1 Input**

{cat, rat, bat, rat}

{bat, mat, pat, bat}

{pat, cat, bat, mat}

|  |  |
| --- | --- |
| **Mapper 0 Output** | **Mapper 1 Output** |
| <{(cat,mat ),1}>  <{(cat,rat),2}>  <{(mat,rat),1}>  <{(mat,cat),1}>  <{(rat,cat),1}>  <{(cat,bat),3}>  <{(bat,cat),1}>  <{(bat,pat),1}>  <{(cat,pat),1}>  <{(bat,rat),1}>  <{(rat,bat),1}> | <{(cat,rat),2}>  <{(cat,bat),2}>  <{(rat,bat),1}>  <{(bat,rat),1}>  <{(bat,mat),2}>  <{(bat,pat),1}>  <{(mat,pat),1}>  <{(mat,bat),1}>  <{(pat,bat),2}>  <{(pat,cat),1)}>  <{(pat,mat),1}>  <{(cat,mat),1}> |
| **Reducer 0 Input (<k)** | **Reducer 1 input (>k)** |
| <{(bat,cat),[1]}>  <{(bat,mat),[2]}>  <{(bat,pat),[1,1]}>  <{(bat,rat),[1,1]}>  <{(cat,bat),[3,2]}>  <{(cat,mat ),[1,1]}>  <{(cat,pat),[1]}>  <{(cat,rat),[2,2]}> | <{(mat,bat),[1]}>  <{(mat,cat),[1]}>  <(mat,pat),[1]>  <{(mat,rat),[1]}>  <(pat,bat),[2]>  <(pat,cat),[1])>  <(pat,mat),[1]>  <{(rat,bat),[1,1]}>  <{(rat,cat),[1]}> |
| **Reducer 0 output** | **Reducer 1 output** |
| <(bat,cat),1>  <(bat,mat),2>  <(bat,pat),2>  <(bat,rat),2>  <(cat,bat),5>  <(cat,mat),2>  <(cat,pat),1>  <(cat,rat),4> | <(mat,bat),1>  <(mat,cat),1>  <(mat,pat),1>  <(mat,rat),1>  <(pat,bat),2>  <(pat,cat),1)>  <(pat,mat),1>  <(rat,bat),2>  <(rat,cat),1> |

**Question 3.c**

Number of Input-Splits: 2

Number of Reducers: 2

**Mapper 0 Input**

{cat, mat, rat, cat}

{cat, bat, cat, pat}

{cat, bat, rat, bat}

**Mapper 1 Input**

{cat, rat, bat, rat}

{bat, mat, pat, bat}

{pat, cat, bat, mat}

|  |  |
| --- | --- |
| **Mapper 0 Output** | **Mapper 1 Output** |
| <(cat,{mat=1,rat=1})>  <(mat,{rat=1,cat=1})>  <(rat,{cat=1})>  <(cat,{bat=1})>  <(bat,{cat=1,pat=1})>  <(cat,{pat=1})>  <(cat,{bat=2,rat=1})>  <(bat,{rat=1})>  <(rat,{bat=1})> | <(cat,{rat=2,bat=1})>  <(rat,{bat=1})>  <(bat,{rat=1})>  <(bat,{mat=1,pat=1})>  <(mat,{pat=1,bat=1})>  <(pat,{bat=1})>  <(pat,{cat=1,bat=1,mat=1})>  <(cat,{bat=1,mat=1})>  <(bat,{mat=1})> |
| **Reducer 0 Input (<k)** | **Reducer 1 input (>k)** |
| <bat,[{cat=1,pat=1},{mat=1},  {mat=1,pat=1},{rat=1},{rat=1}]>  <cat,[{bat=1},{bat=1,mat=1},{bat=2,rat=1},  {mat=1,rat=1},{pat=1},{bat=1,rat=2}]> | <mat,[{bat=1,pat=1},{cat=1,rat=1}]>  <pat,[{bat=1},{cat=1,bat=1,mat=1}]>  <rat,[{bat=1},{bat=1},{cat=1}]> |
| **Reducer 0 output** | **Reducer 1 output** |
| <bat,{cat=1, mat=2, pat=2, rat=2}>  <cat,{bat=5, mat=2, pat=1, rat=4}> | <mat,{bat=1, cat=1, pat=1, rat=1}>  <pat,{bat=2, cat=1, mat=1}>  <rat,{bat=2, cat=1}> |

**Question 3.d**

Number of Input-Splits: 2

Number of Reducers: 2

**Mapper 0 Input**

{cat, mat, rat, cat}

{cat, bat, cat, pat}

{cat, bat, rat, bat}

**Mapper 1 Input**

{cat, rat, bat, rat}

{bat, mat, pat, bat}

{pat, cat, bat, mat}

|  |  |
| --- | --- |
| **Mapper 0 Output** | **Mapper 1 Output** |
| <(cat,{mat=1, rat=2, bat=3, pat=1}>  <(mat,{rat=1, cat=1})>  <(rat,{cat=1, bat=1})>  <(bat,{cat=1, pat=1, rat=1})> | <(cat,{rat=2, bat=2, mat=1})>  <(rat,{bat=1})>  <(bat,{rat=1, mat=2,pat=1})>  <(mat,{pat=1,bat=1})>  <(pat,{bat=2, cat=1, mat=1})> |
| **Reducer 0 Input (<k)** | **Reducer 1 input (>k)** |
| <(bat,[{cat=1, pat=1, rat=1},{mat=2, pat=1, rat=1}])>  <(cat,{bat=3, mat=1, pat=1, rat=2},  {bat=2, mat=1, rat=2})> | <(mat,[{cat=1, rat=1},{ bat=1, pat=1}])>  <(pat,[{bat=2, cat=1, mat=1}]  <(rat,[{bat=1, cat=1}, {bat=1}])> |
| **Reducer 0 output** | **Reducer 1 output** |
| <bat,{cat=1, mat=2, pat=2, rat=2}>  <cat,{bat=5, mat=2, pat=1, rat=4}> | <mat,{bat=1, cat=1, pat=1, rat=1}>  <pat,{bat=2, cat=1, mat=1}>  <rat,{bat=2, cat=1}> |