Q1 In-Map version of Pair approach

Method initialize(){

HM = new AssociateArray()

}

Method map(){

For all term u in record r do

For all term v in window(u) do

If(HM{u,v} is null)

HM{u,v} = new AssociateArray()

HM{u,v} = HM{u,v} + 1

}

Method close(){

For all key in HM

EMIT({u,v},HM{u,v})

}

Q2 In-Map version of Stripe approach

Method initialize(){

HM = new AssociateArray()

}

Method map(){

For all term u in record r do

For all term v in window(u) do

If(HM{u} is null)

HM{u} = new AssociateArray()

HM{u}{v} = HM{u}{v} + 1

}

Method close(){

For all key in HM

EMIT({u},HM{u})

}

Q3

Input Split 1 : [ {cat mat rat, cat}, {cat  bat cat pat},{cat bat rat bat}]    (Note : 3 records)

Input Split 2 : [{cat rat bat rat}, {bat mat pat bat}, {pat cat bat mat}]    (Note: 3 records)

**Let the neighborhood of X, N(X) be set of all term after X and before the next X.**

|  |  |
| --- | --- |
| Input Split 1 |  |
| Record 1 | N(cat) = {mat,rat}  N(mat)={rat,cat}  N(rat)={cat}  N(cat)={} |
| Record 2 | N(cat)={bat}  N(bat)={cat,pat}  N(cat)={pat}  N(pat)={} |
| Record 3 | N(cat)={bat,rat,bat}  N(bat)={rat}  N(rat)={bat}  N(bat)={} |
| Input Split 2 |  |
| Record 1 | N(cat)={rat,bat,rat}  N(rat)={bat}  N(bat)={rat}  N(rat)={} |
| Record 2 | N(bat)={mat,pat}  N(mat)={pat,bat}  N(pat)={bat}  N(bat)={} |
| Record 3 | N(pat)={cat,bat,mat}  N(cat)={bat,mat}  N(bat)={mat}  N(mat)={} |

1. Illustrate Pair approach

|  |  |
| --- | --- |
| Mapper 1 OutPut with input split 1 | Mapper 2 OutPut with input split 2 |
| ((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 Input 1 | Reducer Input 2 |
| ((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 1 Output | Reducer 2 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) |

1. Illustrate In-Mapper Combining Version of the Pair approach

|  |  |
| --- | --- |
| Mapper 1 OutPut with input split 1 | Mapper 2 OutPut with input split 2 |
| ((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 Input 1 | Reducer Input 2 |
| ((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 1 Output | Reducer 2 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) |

1. Illustrate Stripe approach.

|  |  |
| --- | --- |
| Mapper 1 OutPut with input split 1 | Mapper 2 OutPut with input split 2 |
| ((cat),[mat:1,rat:1])  ((mat),[cat:1,rat: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,[bat:1,rat:2))  ((rat,[bat:1])  ((bat,[rat:1])  ((bat,[mat:1,pat:1])  ((mat,[bat:1,pat:1])  ((pat,[bat:1])  ((pat,[bat:1,cat:1,mat:1])  ((cat,[bat:1,mat:1])  ((bat,[mat:1]) |

|  |  |
| --- | --- |
| Reducer Input 1 | Reducer Input 2 |
| ((bat,[[rat:1],[cat:1,pat:1],[rat:1),[mat:1,pat:1],[mat:1]])  ((cat),[[mat:1,rat:1],[bat:1],[pat:1],[bat:2,rat:1],[bat:1,rat:2][bat:1,mat:1]]) | ((mat),[[cat:1,rat:1],[bat:1,pat:1]])  ((pat,[[bat:1],[bat:1,cat:1,mat:1]])  ((rat),[[cat:1],[bat:1],[bat:1]]) |

|  |  |
| --- | --- |
| Reducer 1 Output | Reducer 2 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]) |

1. Illustrate In-Map combining version of Stripe approach:

|  |  |
| --- | --- |
| Mapper 1 OutPut with input split 1 | Mapper 2 OutPut with input split 2 |
| ((cat),[ bat:3,mat:1, pat:2,rat:1])  ((mat),[cat:1,rat:1])  ((rat),[ bat:1,cat:1])  ((bat),[cat:1,pat:1, rat:1]) | ((cat,[bat:2,mat:1,rat:2))  ((rat,[bat:1])  ((bat,[ mat:2,pat:1,rat:1])  ((mat,[bat:1,pat:1])  ((pat,[bat:2, cat:1,mat:1]) |

|  |  |
| --- | --- |
| Reducer Input 1 | Reducer Input 2 |
| ((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 1 Output | Reducer 2 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]) |