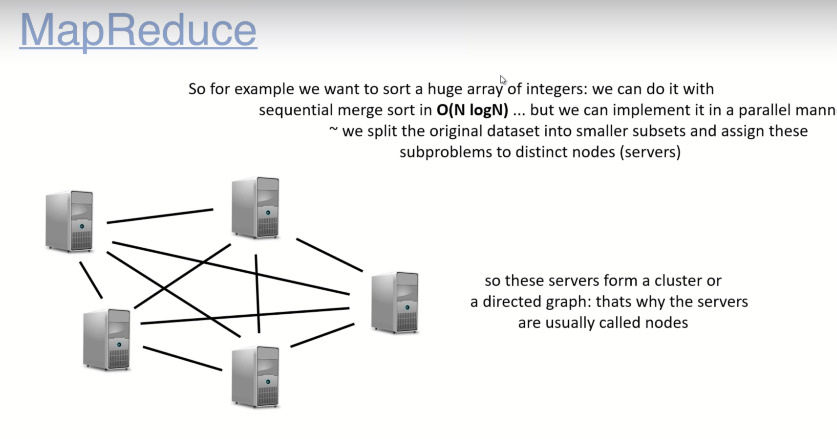
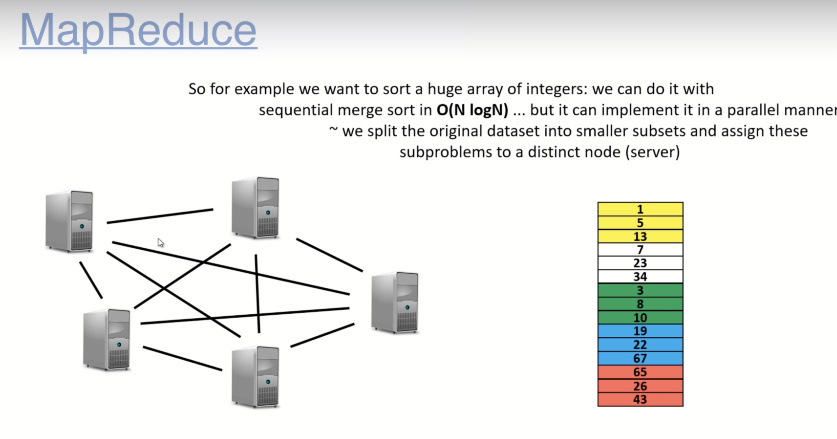
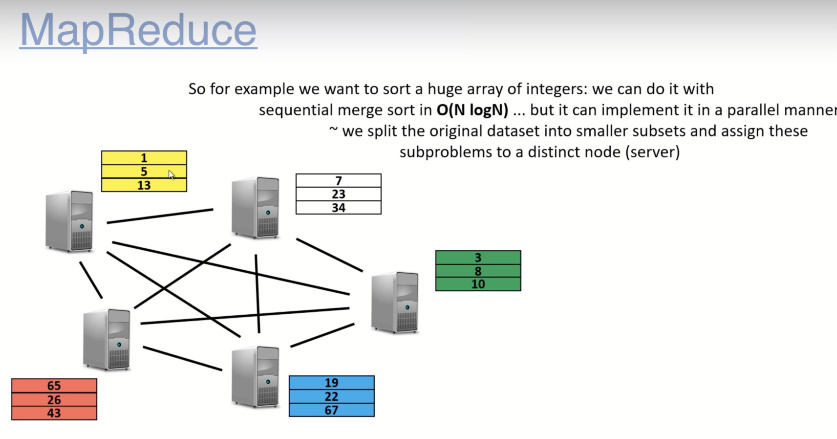
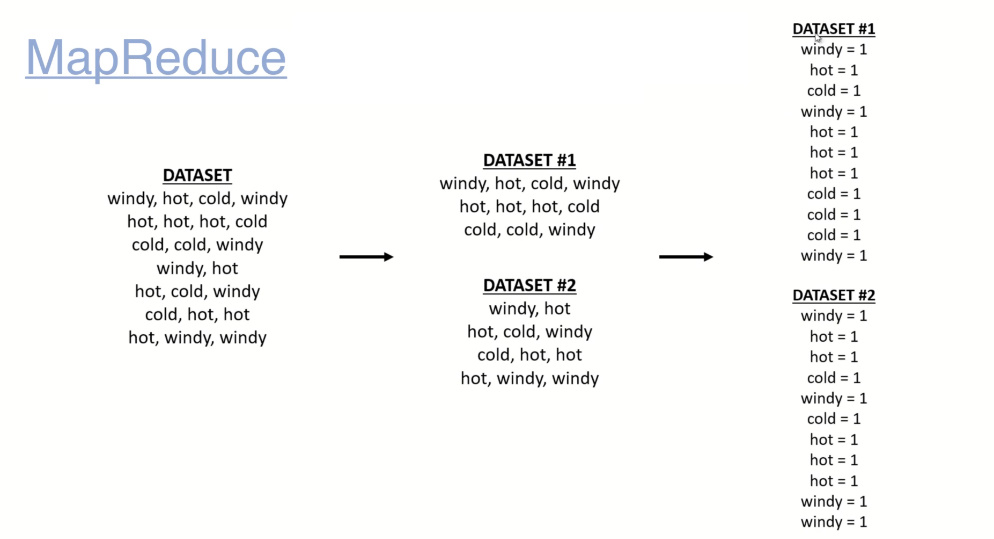
* MapReduce este blocul de baza pentru BigData
* **MapReduce** este un model de programare,un mod de a structura calculele care permite ca acesta sa fie rulat pe multe servere
* MapReduce e asemanator cu ForkJoin, doar ca nu e un Framework, si administreaza diferite servere cu diferite masini
* Cu ajutorul lui MapReduce putem face un algoritm paralel
* Alforitmul are 3 pasi:

1. **Map**: divizeaza multimea de date initiala in subprobleme
2. **shuffle&sort**: toate datele sunt rearanjate pentru a rula mai apoi in paralel. Se asigura ca itemele cu aceleasi keys sa fie luate de acelasi reducer
3. **reduce**: combina rezultatele finale

* 
* ForkJoin framework este pentru o singura masina, dar MapReduce este pentru a rula subprobleme pe diferite masini din retea
* fiecare server va rezolva o subproblema si apoi va returna raspunsul, dar toate o vor face in mod paralel, si anume de asta se ocupa MapReduce
* 
* 
* Deci, fiecare masina sorteaza o bucata de array

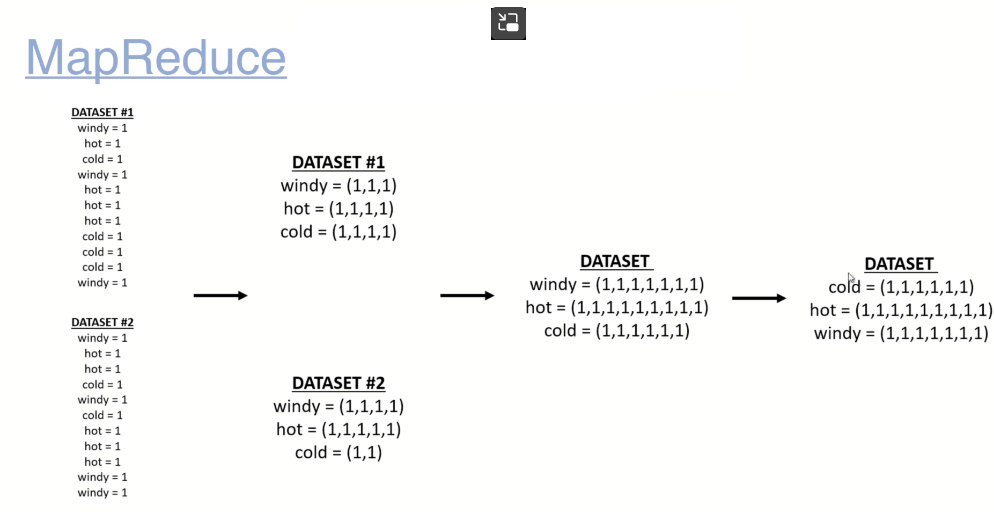
**Input and OutPut**

1. **MAP**: input este o multime de date. De ex, array de int. output: **List<key,value>** perechi

****

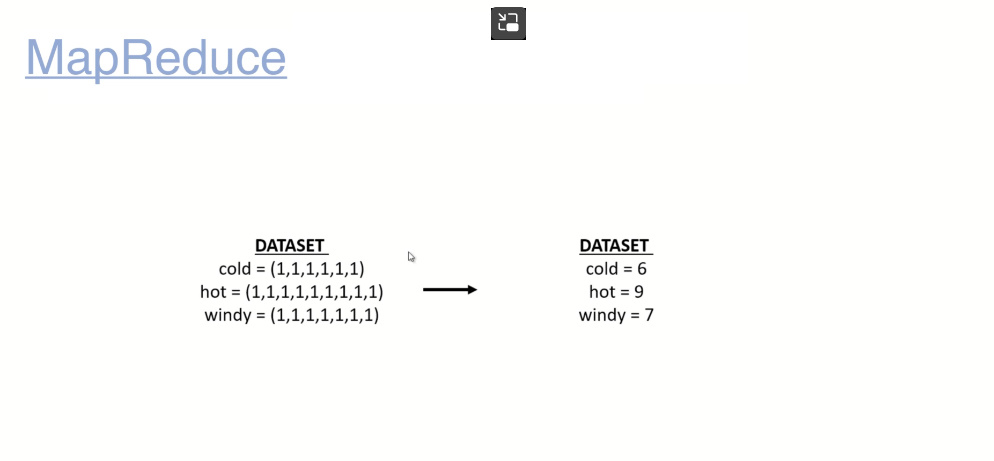


1. **Shuffle&sort**: este chemata functia de combinare. Combina toate key-value cu aceeasi key + le sorteaza. OutputL **List<key,List<Value>>** perechi

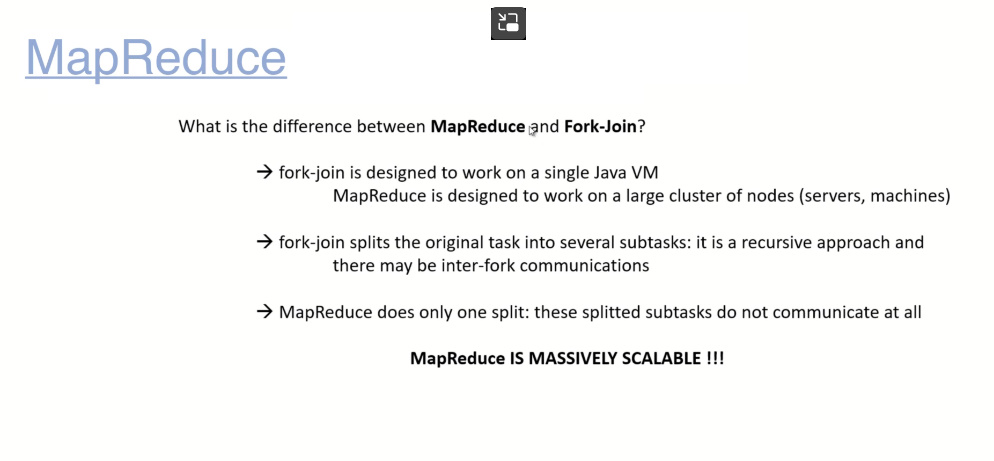
****



1. **Reduce**: combina subrezultatele. Output: **List<key,value>**



**MapReduce vs ForkJoin Framework**



Subtaskurile cu mapReduce nu pot comunica intre ele, asa cum sunt pe diferite masini