




Executie distribuita folosind MapReduce



Bozgan Bianca

A decorative graphic on the left side of the slide consists of a large cyan hexagon in the center. Surrounding it are several smaller hexagons of varying shades of blue and cyan. Some of these hexagons contain white icons: a lightbulb, a thumbs-up, a smartphone, a magnifying glass, a gear, and a speech bubble. There is also a small network diagram icon with a central node and five connecting lines.

Ce este executia
distribuita?



Metoda traditionala

Date



Procesare centrata



Rezultate





Probleme

Redundanta

In cazul in care serverul central nu mai functioneaza atunci procesarea nu se mai realizeaza

Raportul pret/performanta

Pentru o putere de calcul performanta solutiile centralizate sunt mult mai scumpe in comparatie cu cele distribuite

Scalabilitate

Spre deosebire de un sistem distribuit expansiunea nu se poate face prin adaugarea de unitati de calcul

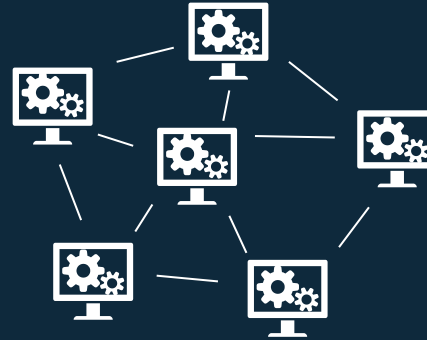


Solutie: procesarea distribuita (I)

- ◇ Sarcinile sunt distribuite pe intregul sistem
- ◇ Sistem cu redundanta si uptime crescut
- ◇ Sistemul este flexibil si scalabil in functie de nevoi



Un supercomputer cu
procesoare multiple



Mai multe calculatoare ce
comunica printr-o retea





Solutie: procesarea distribuita (II)

- ◇ Fiecare computer detine propriul CPU, RAM, HDD, NIC)
 - ◆ Grid-uri (diverse calculatoare conectate prin internet)
 - ◆ Cluestere (calculatoare asemanatoare in acelasi LAN)
- ◇ Util in a putea rezolva probleme mari/complexe, nu probleme mici mai rapid
 - ◆ Problemele necesita putere mare de procesare sau utilizarea volumelor mari de date

OBSTACOLE

Comunicare:

- distribuirea datelor;
- aparitia ineficientelor(overhead, timpi de asteptare).

Paralelizare: atat sarcinile cat si datele trebuie gandite din punctul de vedere al paralelizarii



MapReduce

Procesarea unor volume mari
de date in timp util





Indexarea cu MapReduce


Impartire **→** **Mapare** **→** **Grupare** **→** **Reducere**

Se imparte
volumul de date
in numar de M
fragmente.
Fiecare
fragment este
trimis unu
mapper

Se mapeaza
fiecare fragment
din care se
returneaza o
lista de perechi
cheie-valoare

Se part
perechile in
functie de cheie
in R grupuri si se
trimit reducerilor

Se citesc si se
proceseaza
perechile primite
(grupate dupa
valoarea cheii)



Indexarea cu MapReduce



M1

M2

M3

M4


M5

M6

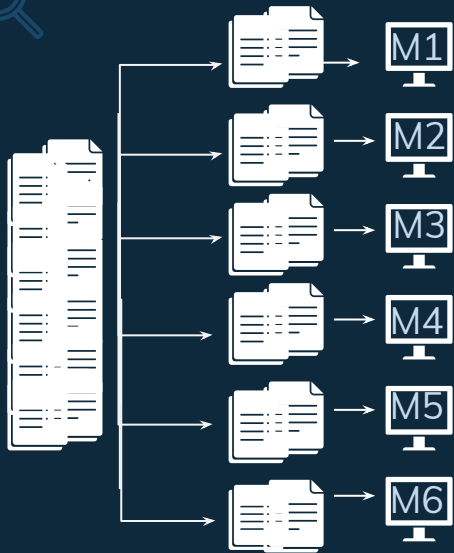
R1

R2





Indexarea cu MapReduce



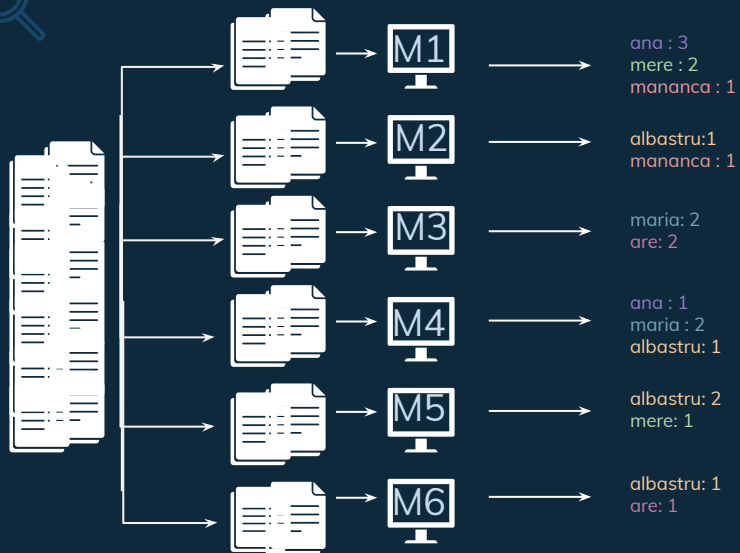
impartire

R1

R2



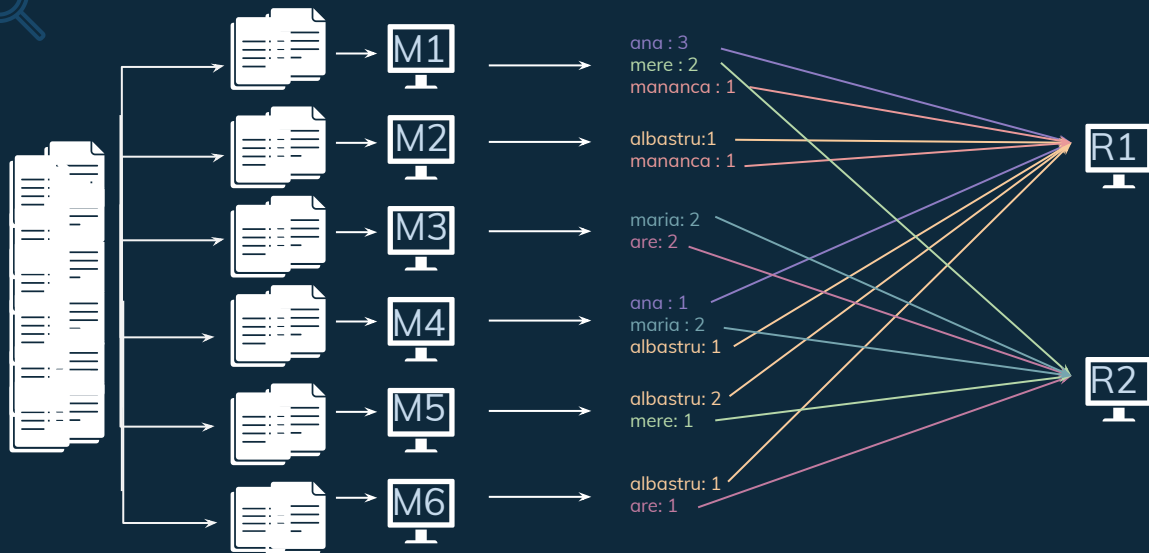
Indexarea cu MapReduce



impartire

mapare

Indexarea cu MapReduce

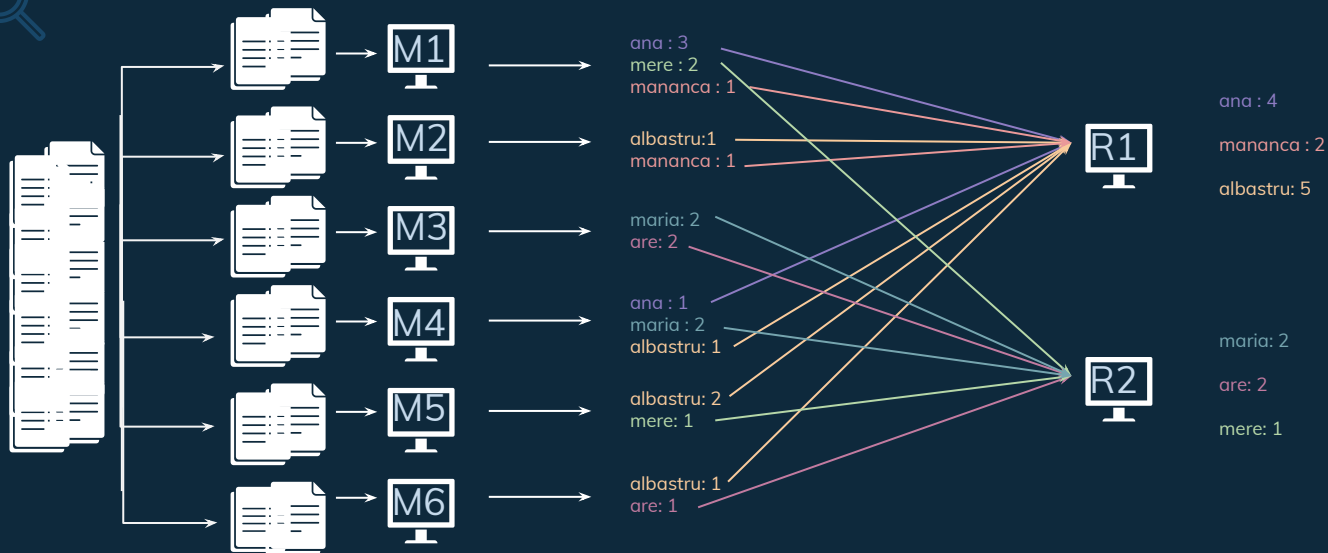


impartire

mapare

grupare

Indexarea cu MapReduce




impartire

mapare

grupare

reducere

The background of the image is a perspective view of a server room. On both sides, there are long rows of server racks. Each rack is filled with numerous horizontal server units, each of which has a glowing green and blue light strip. The floor is a light-colored, reflective surface. The ceiling is visible, showing a grid of lights and structural beams. In the center of the image, there is a large, semi-transparent grey rectangular box containing the text "Timp de executie" in a blue, serif font. Surrounding this central box and scattered throughout the image are several hexagonal icons in various shades of blue and cyan. These icons include a lightbulb (top center), a smartphone (top left), a magnifying glass (top left), a thumbs up (top right), a speech bubble (bottom left), a gear (bottom center), and a network node (bottom right).

Timp de
executie



Va

multumesc!

