

1.

Dibuja el esquema de MapReduce que permita contar la frecuencia de palabras de los siguientes fragmentos

F1: to be or
F2: not to be
F3: that is the question
F4: be not afraid
F5: he was not afraid

El programador indica usar 4 nodos para el mapeo y 2 para reducción .En el esquema se debe mostrar claramente cada uno de los pasos y resultados para cada uno de los nodos.

Solución

	mapeo	Agrupar por llave	reducir
F1: to be or	N1 (to:1)	(to:1)	
F2: not to be	(be:1)	(to:1)	(to:2)
F3: that is the question	(or:1)	(be:1)	(be:3)
F4: be not afraid	(not:1)	(be:1)	(or:1)
F5: he was not afraid	(to:1)	(or:1)	(not:3)
	(be:1)	(not:1)	(that:1)
	(That:1)	(not:1)	(is:1)
	(is:1)	(not:1)	(the:1)
	(the:1)	(that:1)	(question:1)
	(question:1)	(is:1)	(afraid:2)
	(be:1)	(the:1)	(he:1)
	(not:1)	(question:1)	(was:1)
	(afraid:1)	(afriad:1)	
	(he:1)	(afriad:1)	
	(was:1)	(he:1)	
	(not:1)	(was:1)	
	(afraid:1)		

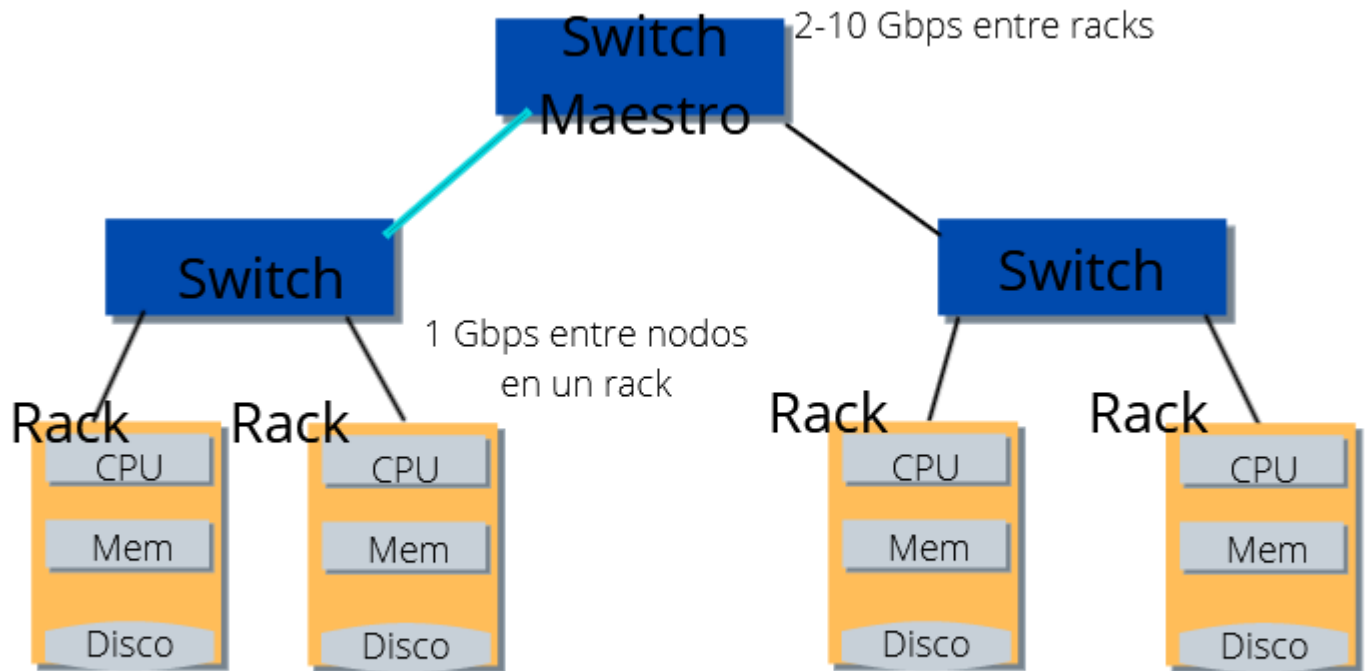
Figura 1: MapReduce con cuatro nodos de mapeo y dos de reducción

2.

Diseña una arquitectura basada en clúster que contenga 4 racks en total, 2 racks deben estar conectados a un switch y los otros 2 racks deben estar conectados a otro switch.

Ambos switches deben estar conectados a un switch principal. En la arquitectura diseñada etiqueta: la velocidad de comunicación entre nodos y entre racks, así mismo indica el límite inferior y superior de nodos que debe contener cada rack, y menciona cuáles son los elementos que componen a un nodo.

Solución



Cada rack tiene entre 16 y 64 nodos

Figura 2: Arquitectura basada en clúster

La velocidad entre dos nodos en un rack es de 1 Gbps, en cambio entre dos racks conectados por un switch es de 2 a 10 Gbps.

Cada rack tiene entre 16 y 64 nodos en cada uno. Y un nodo está compuesto por el CPU, la memoria y el Disco.