More Examples: Relational Join

Assigned Departments

Employee

Name	SSN
Sue	999999999
Tony	777777777

EmpSSN	DepName
999999999	Accounts
777777777	Sales
777777777	Marketing

$SSN = EmpSSN$

Emplyee ⋈ Assigned Departments

Name	SSN	EmpSSN	DepName
Sue	999999999	999999999	Accounts
Tony	777777777	777777777	Sales
Tony	777777777	777777777	Marketing

Relational Join in MapReduce: Before Map Phase

Key idea: Lump all the tuples together into one dataset

Employee

Name	SSN
Sue	999999999
Tony	777777777

Assigned Departments

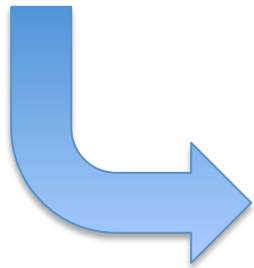| EmpSSN	DepName
999999999	Accounts
777777777	Sales
777777777	Marketing

Employee, Sue, 999999999
Employee, Tony, 777777777
Department, 999999999, Accounts
Department, 777777777, Sales
Department, 777777777, Marketing

What is this for?

Relational Join in MapReduce: Map Phase

Employee, Sue, 999999999
Employee, Tony, 777777777
Department, 999999999, Accounts
Department, 777777777, Sales
Department, 777777777, Marketing

key=999999999, value=(Employee, Sue, 999999999)
key=777777777, value=(Employee, Tony, 777777777)
key=999999999, value=(Department, 999999999, Accounts)
key=777777777, value=(Department, 777777777, Sales)
key=777777777, value=(Department, 777777777, Marketing)
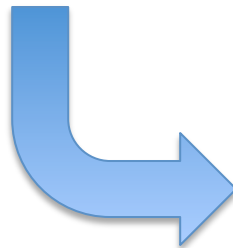
why do we use this as the key?

Relational Join in MapReduce: Reduce Phase

key=999999999, values=[(Employee, Sue, 999999999),
                       (Department, 999999999, Accounts)]

Employee   Dept

| Sue, 999999999, 999999999, Accounts |

key=777777777, values=[(Employee, Tony, 777777777),
                       (Department, 777777777, Sales),
                       (Department, 777777777, Marketing)]

Emp   Dept

| Tony, 777777777, 777777777, Sales |
| Tony, 777777777, 777777777, Marketing |

Relational Join in MapReduce, again

Order(orderid, account, date)   LineItem(orderid, itemid, qty)

1, aaa, d1	1, 10, 1
2, aaa, d2	1, 20, 3
3, bbb, d3	2, 10, 5
	2, 50, 100
	3, 20, 1

Map   **tagged with relation name**

Order

1, aaa, d1	→	1 : "Order", (1,aaa,d1)
2, aaa, d2	→	2 : "Order", (2,aaa,d2)
3, bbb, d3	→	3 : "Order", (3,bbb,d3)

Line

1, 10, 1	→	1 : "Line", (1, 10, 1)
1, 20, 3	→	1 : "Line", (1, 20, 3)
2, 10, 5	→	2 : "Line", (2, 10, 5)
2, 50, 100	→	2 : "Line", (2, 50, 100)
3, 20, 1	→	3 : "Line", (3, 20, 1)

Reducer for key 1

"Order", (1,aaa,d1)
"Line", (1, 10, 1)
"Line", (1, 20, 3)

(1, aaa, d1, 1, 10, 1)
(1, aaa, d1, 1, 20, 3)

Simple Social Network Analysis: Count Friends

Input

Jim, Sue
Sue, Jim
Lin, Joe
Joe, Lin
Jim, Kai
Kai, Jim
Jim, Lin
Lin, Jim

MAP

Jim, 1
Sue, 1
Lin, 1
Joe, 1
Jim, 1
Kai, 1
Jim, 1
Lin, 1

SHUFFLE

Jim, (1, 1, 1)
Lin, (1, 1)
Sue, (1)
Kai, (1)
Joe, (1)

REDUCE

Jim, 3
Lin, 2
Sue, 1
Kai, 1
Joe, 1