Joining Tables

CompSci 590.04

Instructor: Ashwin Machanavajjhala

A SQL Query walks into a bar. In the corner of the bar are two tables. The Query walks up to the tables and asks,

"Mind if I join you?"

Duke UNIVERSITY

Image credit: http://www.swcasinoproductions.com/images/party-rentals/lighted-tables-buffets-furniture/party-rentals-lighted-bar-acrylic.jpg

Why Joins?

- Cornerstone of the relation data model!
- Normalization of schema helps reduce redundancy and inconsistencies.
  - Social network: Table of edges and Table of vertices
- To reason about relationships between multiple types of entities
  - How many triangles are there in a graph
- To answer queries over data split over multiple tables
  - Need to join a table of income and table of education to get correlation between the two attributes.

A short detour through datalog

- Suppose we have a relational schema R(A, B) and S(B, C).
  - R is a table with two attributes A and B
  - S is a table with two attributes B and C

- Natural join: Join every row in R with every row in S whenever they agree on the B attribute.

  J(a, b, c) :- R(a,b) ∧ S(b, c)

- Triangle counting:

  T(a, b, c) :- E(a,b) E(b,c) E(c,d)

- We can express most SQL queries in datalog

  Z(a,c) :- R(a, b) ∧ S(b, c) ∧ b='b'

Question: What is the best method for computing a join

J(a,b,c) :- R(a,b), S(b,c)

Question: What is the best method for computing a join

J(a,b,c) :- R(a,b), S(b,c)

Many techniques are known (this is a 40 year old problem)!

- X - Nested Loop Join (X = "", Block, Index, …)
- Hash Join
- Sort Merge Join
- …

Joining 2 tables in parallel systems

Hash Join:

- Hash records to reducers
- Join in the reducer.

Fragment Replicated Join:

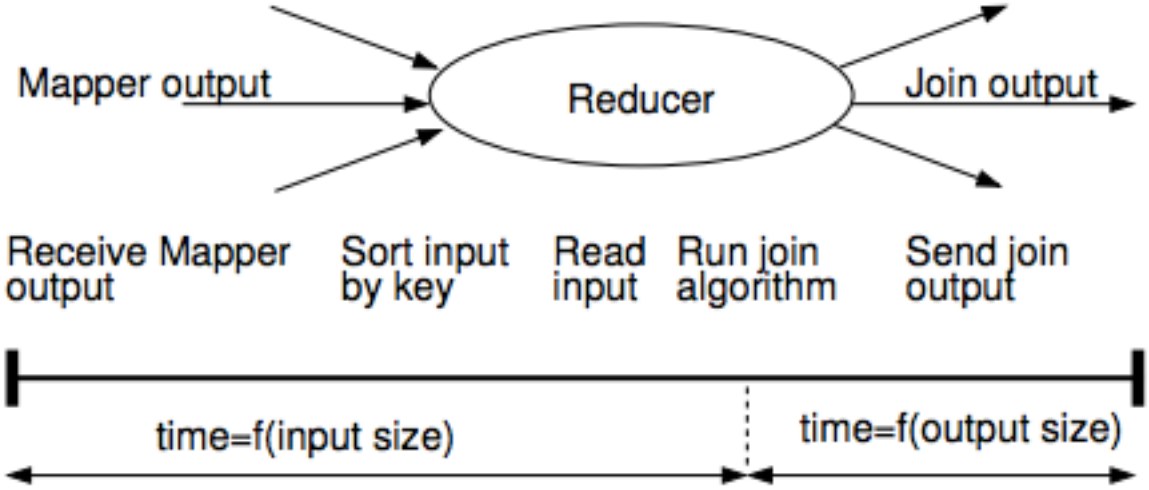
- When one of the tables is small enough to fit in memory.
- Replicate the "small" table to all mappers containing the other "large" table.

Merge Join:

- When two datasets are already sorted on the join key
- Use sort merge join.

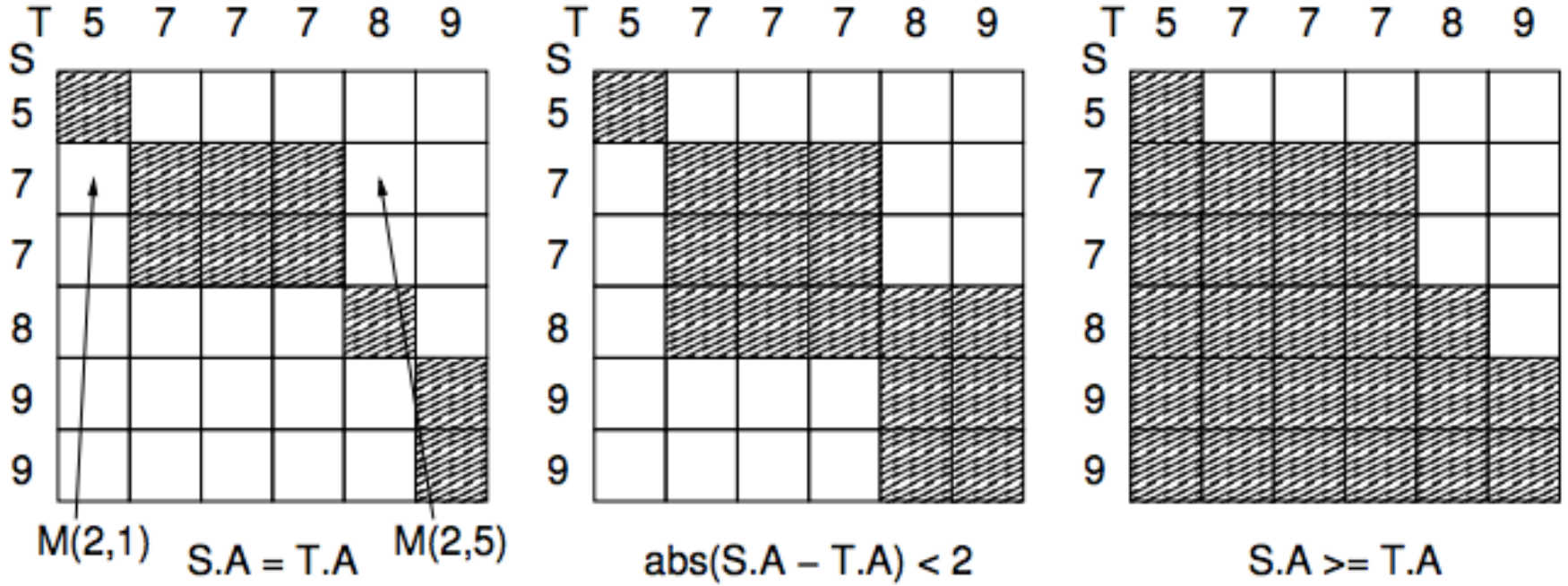
Join as an optimization problem

- Objective: minimize job completion time
- Cost at a reducer:



- Input-size dominated: Reducer input processing time is large
- Output-size dominated: Reducer output processing time is large

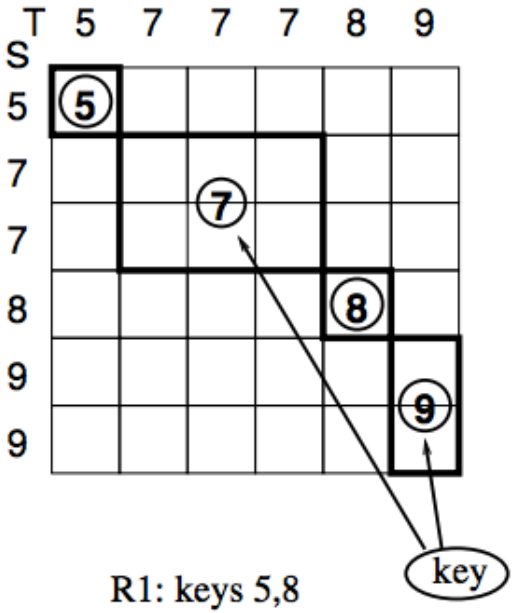
Join Matrix



Goal: find a mapping between join matrix cells to reducers that minimizes completion time.

Duke UNIVERSITY

Join Alternatives



R1: keys 5,8
Input: S1,S4
T1,T5
Output: 2 tuples

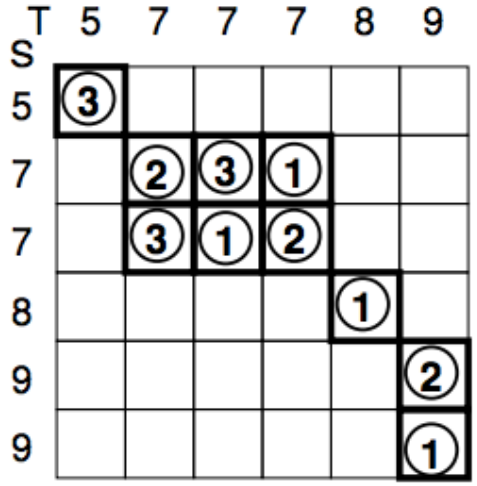
R2: key 7
Input: S2,S3
T2,T3,T4
Output: 6 tuples

R3: key 9
Input: S5,S6
T6
Output: 2 tuples

max–reducer–input = 5
max–reducer–output = 6

- Standard join algorithm
- Group both tables by the key, send all tuples with same key to a single reducer
- Skew in 7 leads to long completion time.

Join Alternatives



- Fine grained load balancing.
  - Divide the cells in the join matrix equally amongst reducers
- Leads to replication of tuples to multiple reducers
  - Higher communication cost

R1: key 1
Input: S2,S3,S4,S6
T3,T4,T5,T6
Output: 4 tuples

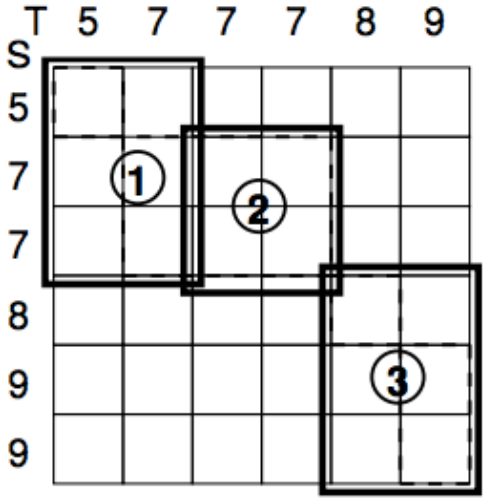
R2: key 2
Input: S2,S3,S5
T2,T4,T6
Output: 3 tuples

R3: key 3
Input: S1,S2,S3
T1,T2,T3
Output: 3 tuples

max–reducer–input = 8
max–reducer–output = 4

Duke UNIVERSITY

Join Alternatives



R1: key 1
Input: S1,S2,S3
T1,T2
Output: 3 tuples

R2: key 2
Input: S2,S3
T3,T4
Output: 4 tuples

R3: key 3
Input: S4,S5,S6
T5,T6
Output: 3 tuples

max–reducer–input = 5
max–reducer–output = 4

- Best of both worlds
- Key 7 is broken into two different reducers
- Limits replication of input as well as reduces skew

General Strategy

- Identify the regions in the join matrix that appear in the join
  - Sufficient to identify a superset of the shaded cells in the join matrix

- Map regions of the join matrix to reducers such that each shaded cell is covered by a reducer.

Multi-way Joins

J(a,b,c) :- R(a,b) S(b,c) T(a,c)

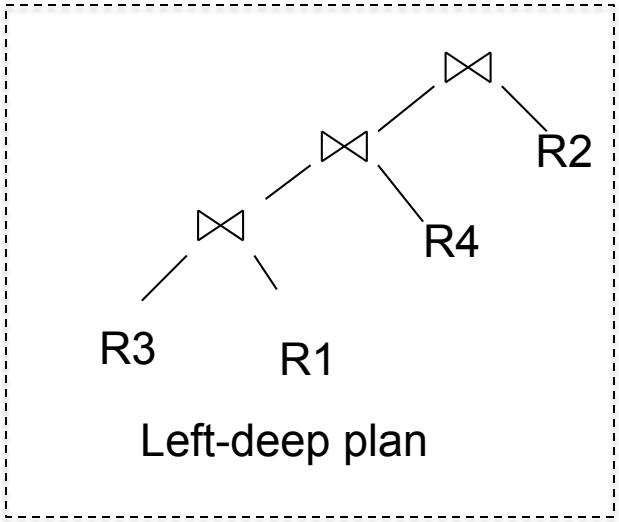
//This is triangle counting

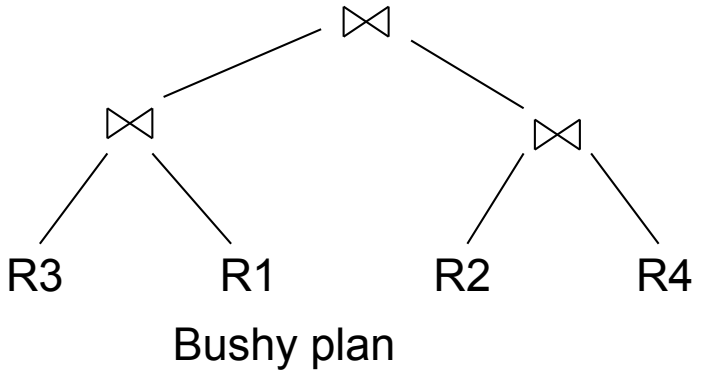
// Suppose each table has the same size N

Multi-way Joins

J(a,b,c) :- R(a,b) S(b,c) T(a,c)

- Historically databases designers decided that the best way to handle multi-way joins is to do them one pair at a time.
  - For efficiency reasons.





Duke UNIVERSITY