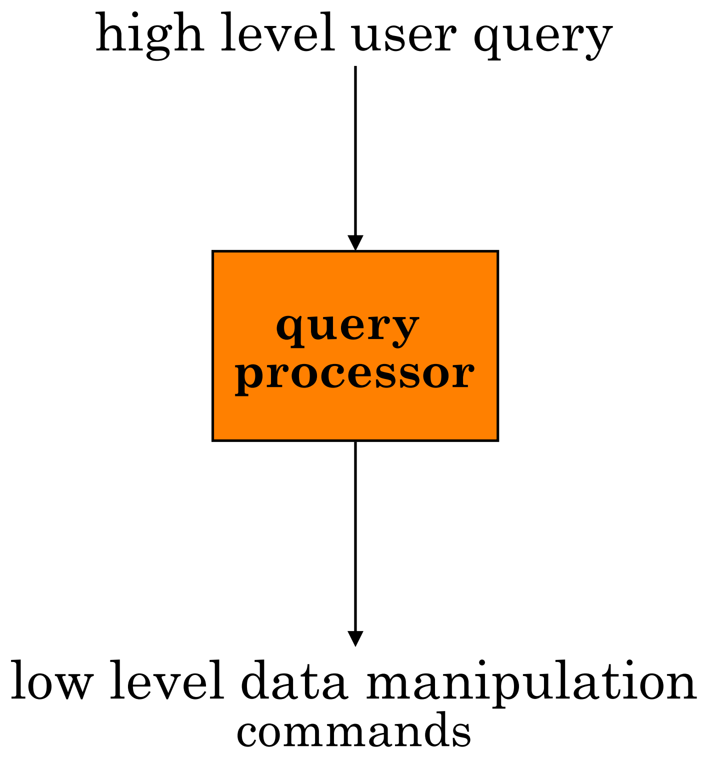
Outline

- Introduction
- Background
- Distributed DBMS Architecture
- Distributed Database Design
- Distributed Query Processing
  - Query Processing Methodology
  - Distributed Query Optimization
- Transaction Management
- Building Distributed Database Systems (RAID)
- Mobile Database Systems
- Privacy, Trust, and Authentication
- Peer to Peer Systems

Useful References

- Textbook *Principles of Distributed Database Systems*, Chapter 6

Query Processing



Query Processing Components

- Query language that is used
  - SQL: “intergalactic dataspeak”
- Query execution methodology
  - The steps that one goes through in executing high-level (declarative) user queries.
- Query optimization
  - How do we determine the “best” execution plan?

Selecting Alternatives

```
SELECT    ENAME
FROM      EMP,ASG
WHERE     EMP.ENO = ASG.ENO
AND       DUR > 37
```

Π Project

σ Select

× Join

Strategy 1

$$\Pi_{ENAME}(\sigma_{DUR>37 \wedge EMP.ENO=ASG.ENO}(EMP \times ASG))$$

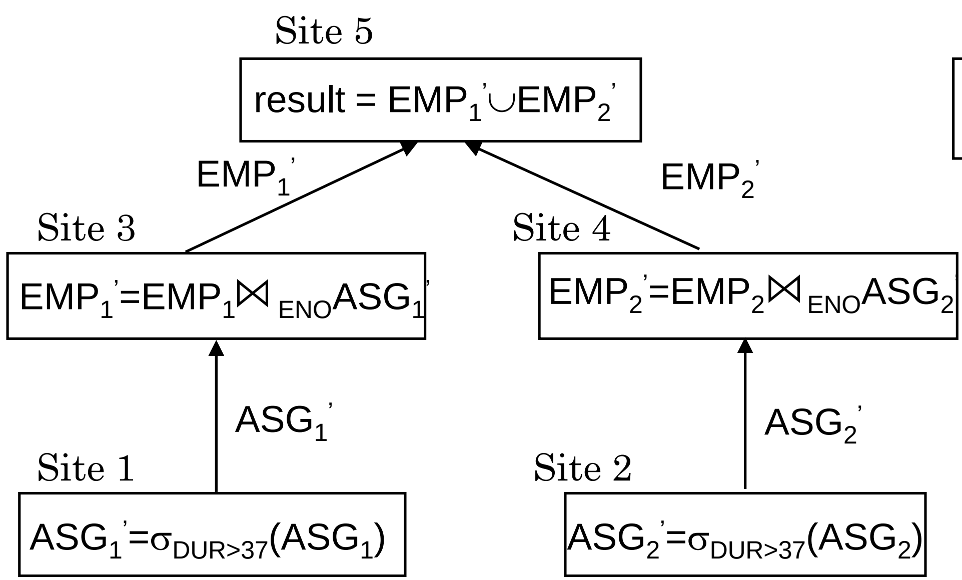
Strategy 2

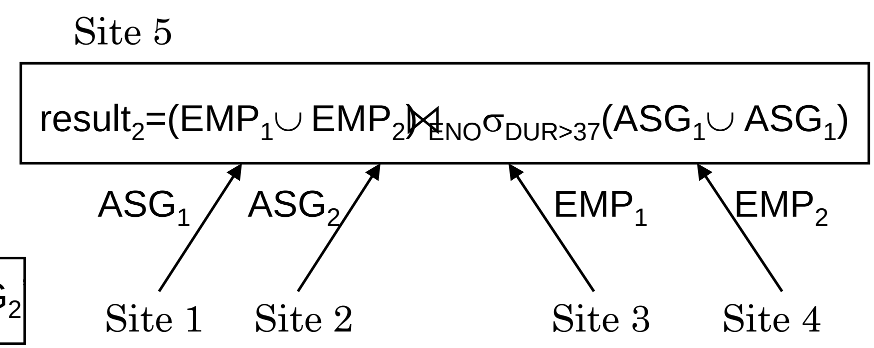
$$\Pi_{ENAME}(EMP \bowtie_{ENO} (\sigma_{DUR>37}(ASG)))$$

Strategy 2 avoids Cartesian product, so is "better"

What is the Problem?

Site 1	Site 2	Site 3	Site 4	Site 5
$ASG_1=\sigma_{ENO\leq\text{“E3”}}(ASG)$	$ASG_2=\sigma_{ENO>\text{“E3”}}(ASG)$	$EMP_1=\sigma_{ENO\leq\text{“E3”}}(EMP)$	$EMP_2=\sigma_{ENO>\text{“E3”}}(EMP)$	Result





Cost of Alternatives

- Assume:
  - *size*(EMP) = 400, *size*(ASG) = 1000
  - tuple access cost = 1 unit; tuple transfer cost = 10 units
- Strategy 1

produce ASG': (10+10)∗tuple access cost	20
transfer ASG' to the sites of EMP: (10+10)∗tuple transfer cost	200
produce EMP': (10+10) ∗tuple access cost∗2	40
transfer EMP' to result site: (10+10) ∗tuple transfer cost	200
Total cost	460

- Strategy 2

transfer EMP to site 5:400∗tuple transfer cost	4,000
transfer ASG to site 5 :1000∗tuple transfer cost	10,000
produce ASG':1000∗tuple access cost	1,000
join EMP and ASG':400∗20∗tuple access cost	8,000
Total cost	23,000

Query Optimization Objectives

Minimize a cost function

I/O cost + CPU cost + communication cost

These might have different weights in different distributed environments

Wide area networks

- communication cost will dominate (80 – 200 ms)
  - low bandwidth
  - low speed
  - high protocol overhead
- most algorithms ignore all other cost components

Local area networks

- communication cost not that dominant (1 – 5 ms)
- total cost function should be considered

Can also maximize throughput

Complexity of Relational Operations

- Assume
  - relations of cardinality n
  - sequential scan

Operation	Complexity
Select Project (without duplicate elimination)	$O(n)$
Project (with duplicate elimination) Group	$O(n\log n)$
Join Semi-join Division Set Operators	$O(n\log n)$
Cartesian Product	$O(n^2)$

Query Optimization Issues – Types of Optimizers

- Exhaustive search
  - cost-based
  - optimal
  - combinatorial complexity in the number of relations
- Heuristics
  - not optimal
  - regroup common sub-expressions
  - perform selection, projection first
  - replace a join by a series of semijoins
  - reorder operations to reduce intermediate relation size
  - optimize individual operations

Query Optimization Issues – Optimization Granularity

- Single query at a time
  - cannot use common intermediate results
- Multiple queries at a time
  - efficient if many similar queries
  - decision space is much larger

Query Optimization Issues – Optimization Timing

- Static
  - compilation ⇒ optimize prior to the execution
  - difficult to estimate the size of the intermediate results ⇒ error propagation
  - can amortize over many executions
  - R*
- Dynamic
  - run time optimization
  - exact information on the intermediate relation sizes
  - have to reoptimize for multiple executions
  - Distributed INGRES
- Hybrid
  - compile using a static algorithm
  - if the error in estimate sizes > threshold, reoptimize at run time
  - MERMAID

Query Optimization Issues – Statistics

- Relation
  - cardinality
  - size of a tuple
  - fraction of tuples participating in a join with another relation
- Attribute
  - cardinality of domain
  - actual number of distinct values
- Common assumptions
  - independence between different attribute values
  - uniform distribution of attribute values within their domain

Query Optimization Issues – Decision Sites

- Centralized
  - single site determines the "best" schedule
  - simple
  - need knowledge about the entire distributed database
- Distributed
  - cooperation among sites to determine the schedule
  - need only local information
  - cost of cooperation
- Hybrid
  - one site determines the global schedule
  - each site optimizes the local subqueries

Query Optimization Issues – Network Topology

- Wide area networks (WAN) – point-to-point
  - characteristics
    - low bandwidth
    - low speed
    - high protocol overhead
  - communication cost will dominate; ignore all other cost factors
  - global schedule to minimize communication cost
  - local schedules according to centralized query optimization
- Local area networks (LAN)
  - communication cost not that dominant
  - total cost function should be considered
  - broadcasting can be exploited (joins)
  - special algorithms exist for star networks