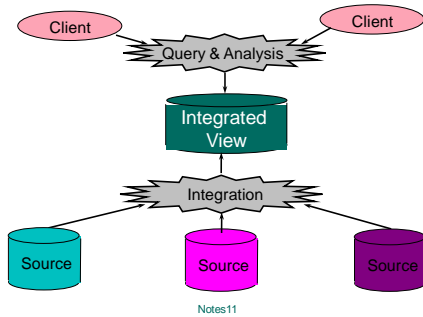
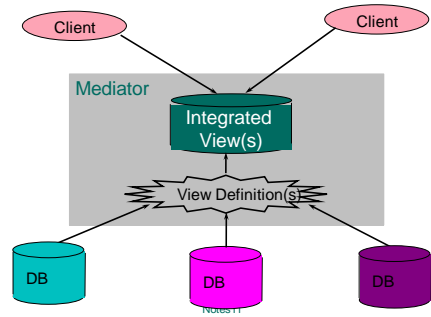


## Warehouses & Virtual Databases offer integrated views

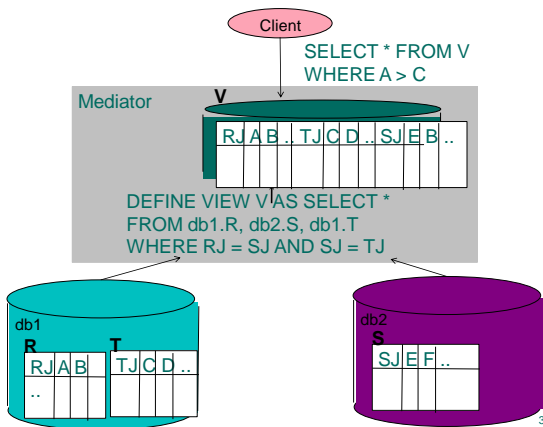


1

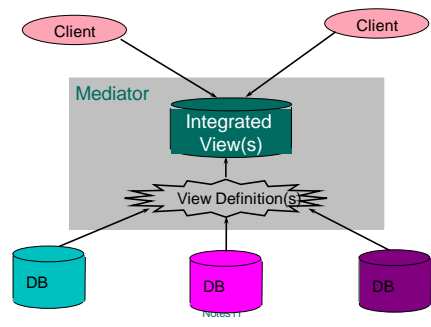
Focus: Sources are relational DBs. Integration specified by distributed view definition(s). Clients issue queries on views.



2

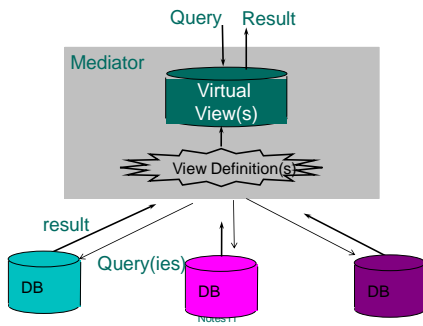


Virtual View -> Mediator needs Distributed Query Processor  
Materialized view (warehouse) -> Mediator also needs storage & Incremental View Maintenance



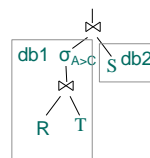
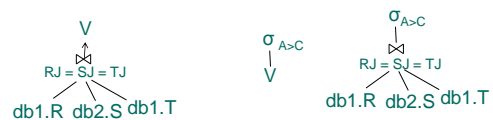
4

## Distributed Query Processing in Mediators



5

## Mediators as providers of view-based virtual views over distributed data



Translate subplans to SQL

+ How will the join happen?

- What if the source offers multiple data services instead of JDBC access? (besides CSE232 material)

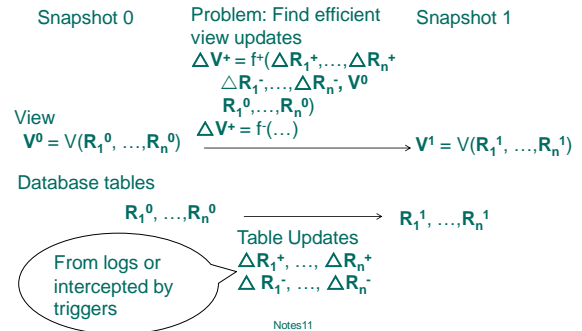
6

## Join types

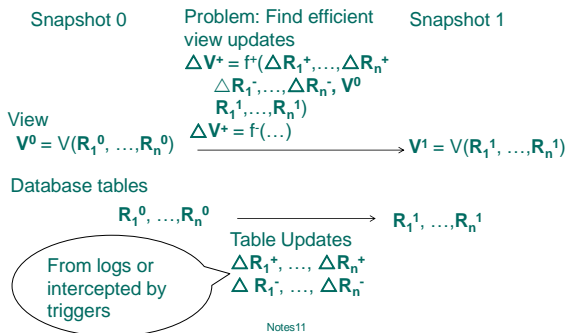
- Mediator-based Join
  - ◆ Ship results of queries at mediator
- Parameterized Join
  - ◆ Right subquery is enhanced with selection on join attribute
  - ◆ For each join value of left hand side, execute another right subquery
- Data Ship Join
  - ◆ Insert the result of left hand side (lhs) in the db of right hand side (rhs).
  - ◆ Execute join at db of right hand side
- Semijoin Reduction Join
  - ◆ Send rhs parameters to lhs
  - ◆ (Data ship alike variation) Lhs sends to rhs the semijoin of its subquery with the parameters set.
  - ◆ Execute join at db of rhs
  - ◆ Also, variation that looks like mediator-based join

7

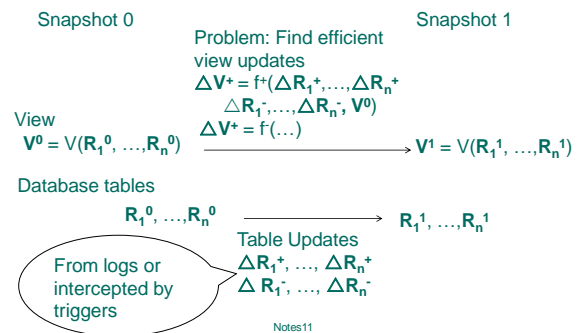
## IVM: Incremental (Materialized) View Maintenance. Eager version.



## IVM: Deferred version



## IVM: Self-maintaining version (not always possible)



## Generalizations

- Multiple views
  - ◆ self maintenance may involve a view utilizing the other views in its computation
- Genuine updates
  - ◆ Not simulated via insertions/deletions
- Insertions, deletions, updates on tables and views expressed as DML statements

11

## Comparisons

### Materialized View

- High query performance
- Queries not visible outside warehouse
- Local processing at sources unaffected
- Can operate when sources unavailable
- Extra information at warehouse
  - ◆ Modify, summarize (store aggregates)
  - ◆ Add historical information

### Virtual View

- No need for yet another database
  - More up-to-date data
    - ◆ Depending on specifics of IVM
  - Query needs can be unknown
  - Only query interface needed at sources
- => Lower Total Cost of Ownership

Notes11

12