

ENTERPRISE INTEGRATION

Michael Stonebraker
Chief Technology Officer

COHERA CORPORATION

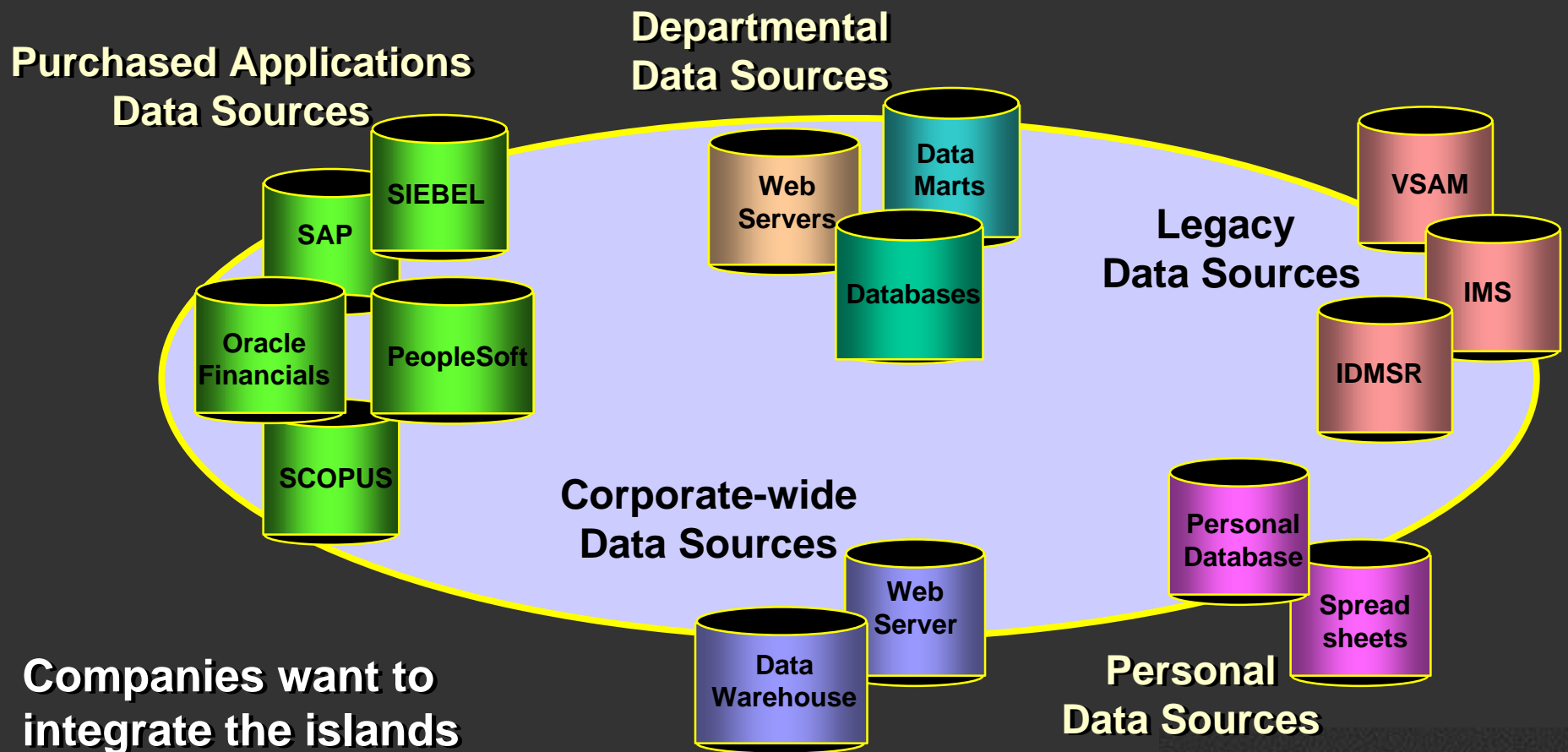


OUTLINE

- **The business problem - enterprise integration**
- **Candidate integration technologies**
- **Discussion**
 - **Do it at the data level!**
 - **The drive for coverage**

ENTERPRISE DATA ISOLATION

Most enterprises have distributed islands of dissimilar data.



AND EXISTING APPS

- **Running against these data sources**
- **Using a variety of programming systems and languages**
- **Some current and some legacy!**

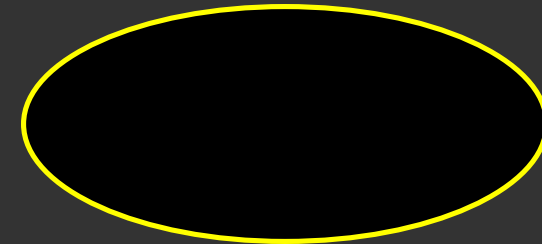
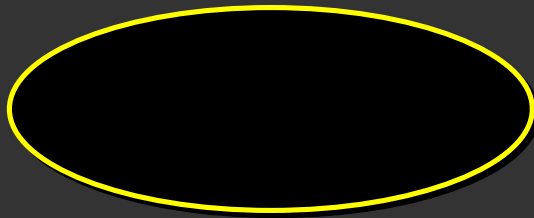
THE BIG PROBLEM TO A CIO

- **Integrate the enterprise to achieve**
 - **information integration (assemble data from multiple islands)**
 - **semantic data consistency (do the right thing to other islands when an island is updated)**

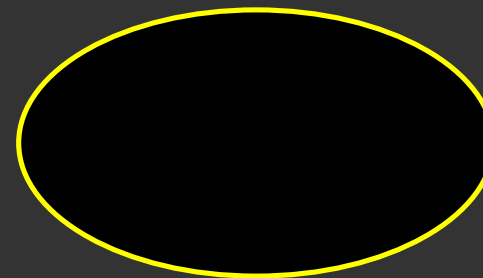
A SIMPLE EXAMPLE OF INFORMATION INTEGRATION

Emp (name, salary)
behind a web server

The user



Emp (name, salary, dept)
in DB2



Dept (dname, type, floor)
in Oracle

INTEGRATION PROBLEM

Find the average salary of employees who work in a service department

```
select avg (salary)
from Emp
where dept in
      (select dname
       from Dept
       where type = "service");
```

(wide area partitioned join with an aggregate)

BUT THERE IS (ALWAYS) SEMANTIC HETEROGENEITY!

- **US salary is gross salary in dollars**
- **French salary is net salary in francs with a lunch allowance**
- **And it might be possible for an employee to be part time in both organizations!**

SCALE OF THE PROBLEM

- **A typical enterprise has**
 - **10,000 data sources**
 - **on 1000 machines**
- **And 50,000 applications**

WHAT IS COMING?

- **Multi-enterprise data integration**
 - e.g. supply chain management
- **The web and the internet**
 - the web is one big data integration problem
 - and current tools are woefully inadequate

Lots more data sources!

WHAT IS COMING?

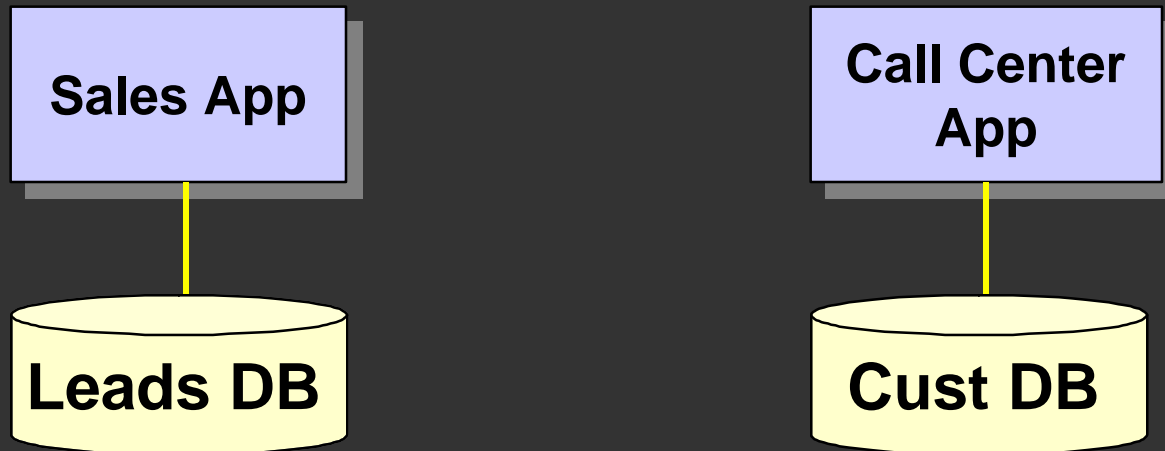
- **Mobile computing and the gizmo explosion**
 - we will all carry data on our gizmos!
- **Legacy system migration**
 - have to integrate “new” and “old” during migration

Lots more data sources!

CANDIDATE ENTERPRISE INTEGRATION TECHNOLOGIES

- **Application integration**
- **Publish/Subscribe systems**
- **TP monitors**
- **Warehousing and data marts**
- **Data federation systems**

TO ILLUSTRATE THE APPROACHES



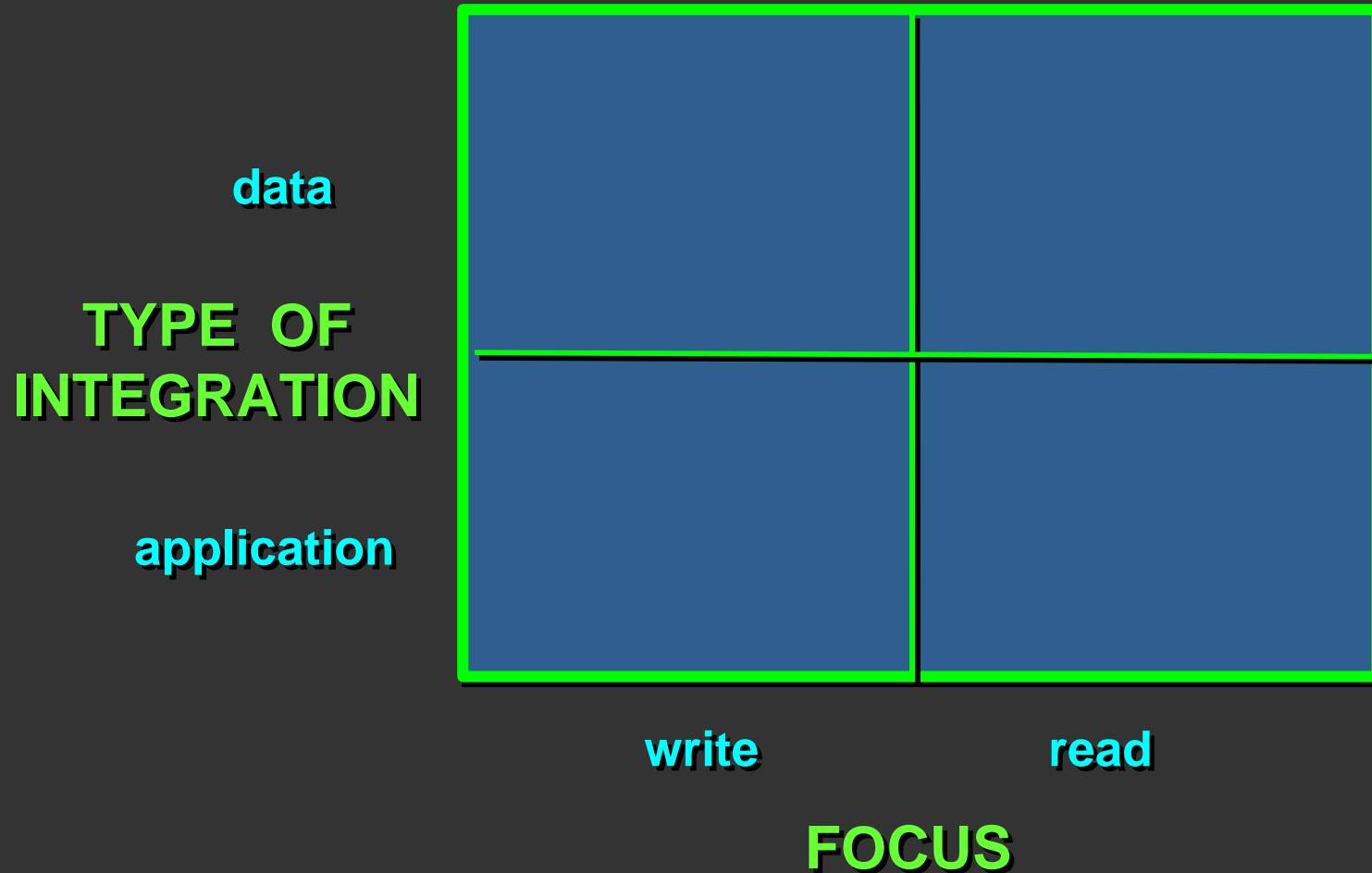
ONE WAY TO LOOK AT SOLUTIONS

- **Integrate at the data level**
 - **by doing integration through the data base**
- **Integrate at the application level**
 - **by having applications send/receive messages**

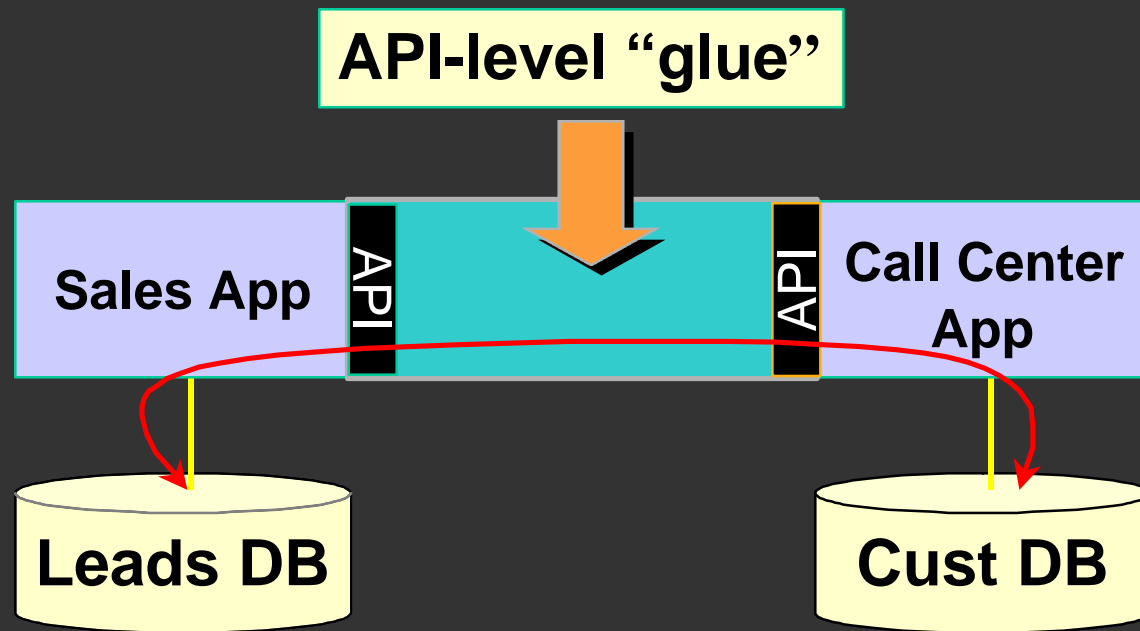
ANOTHER WAY

- **Focus on reads (information integration)**
 - **is anybody both a prospect and a customer**
- **Focus on writes (OLTP)**
 - **what to do when a prospect becomes a customer**

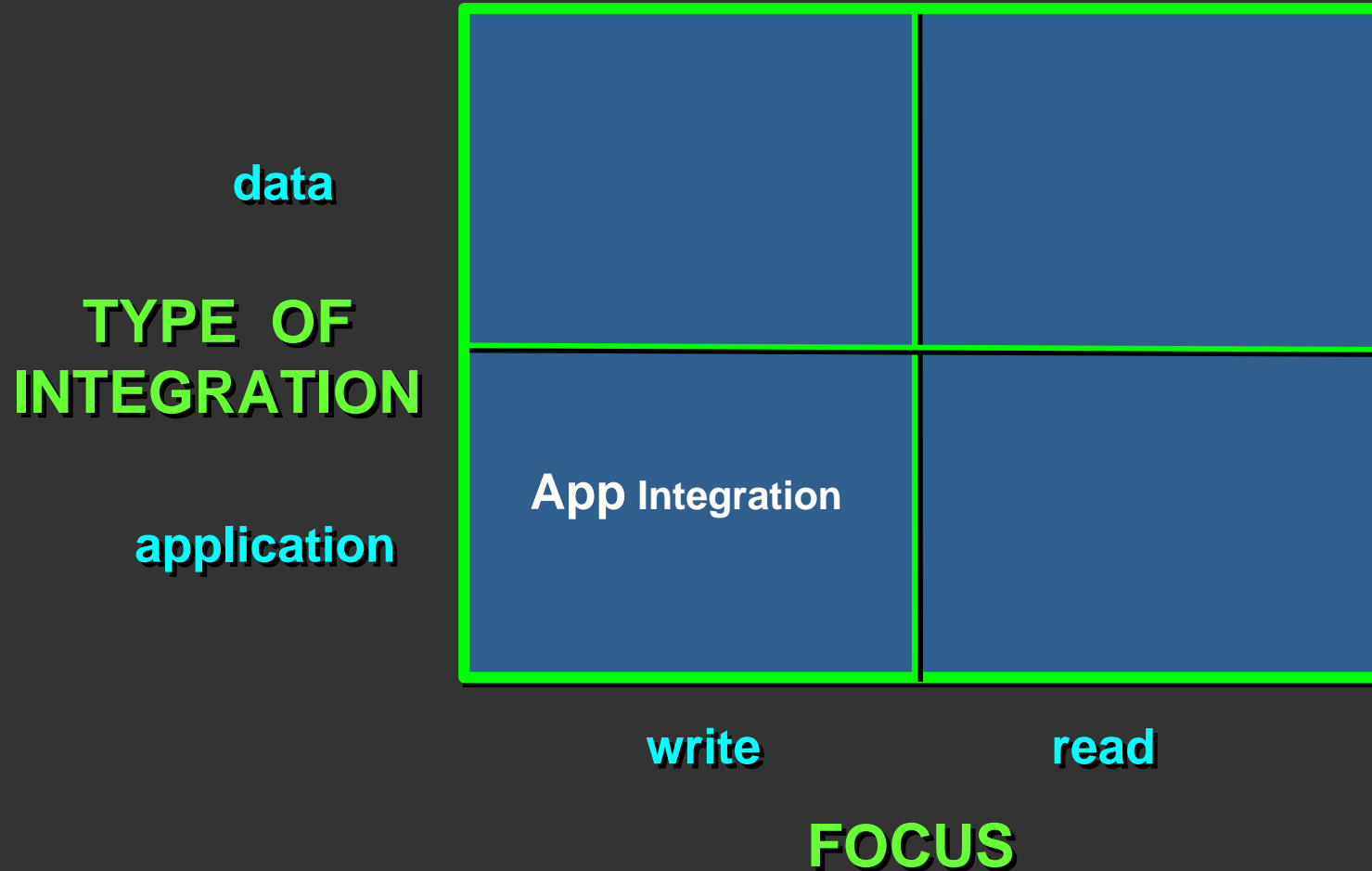
WHICH Solution To Use?



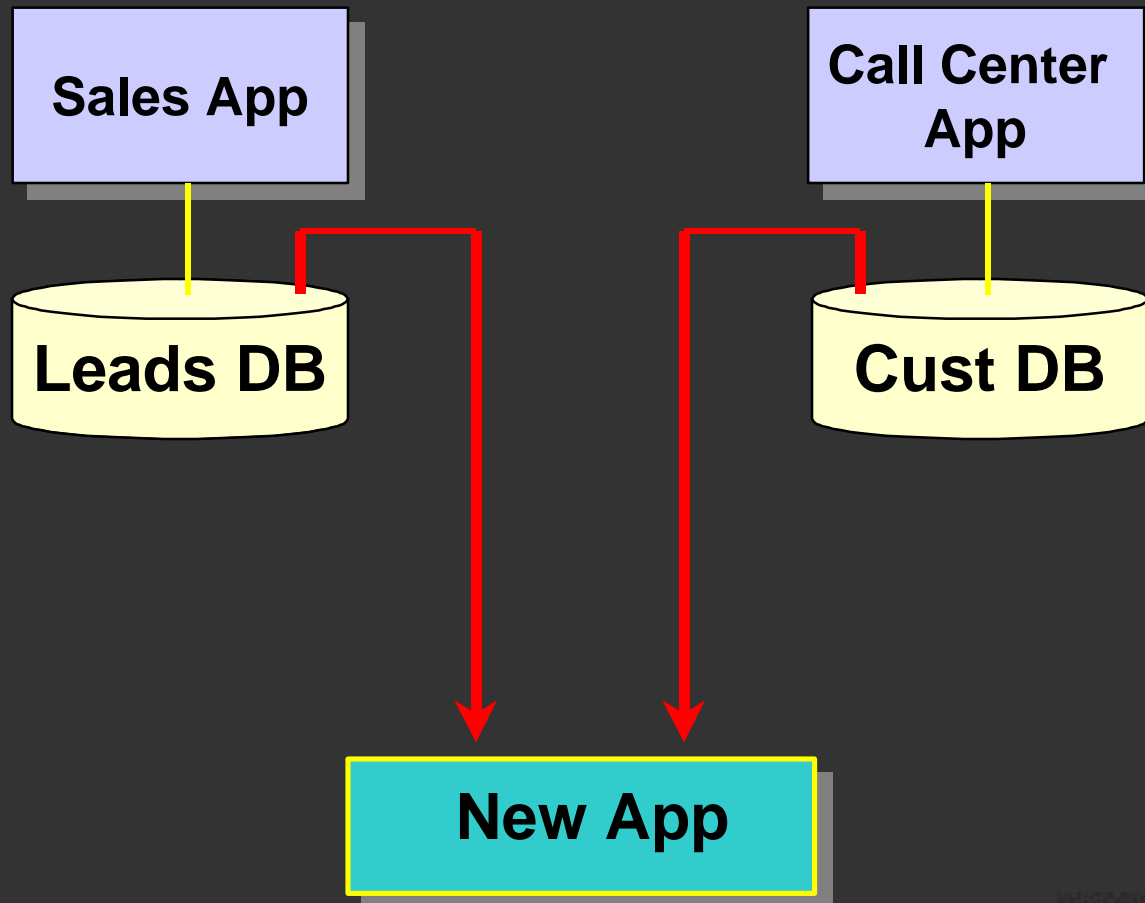
SOLUTION #1: APPLICATION Integration



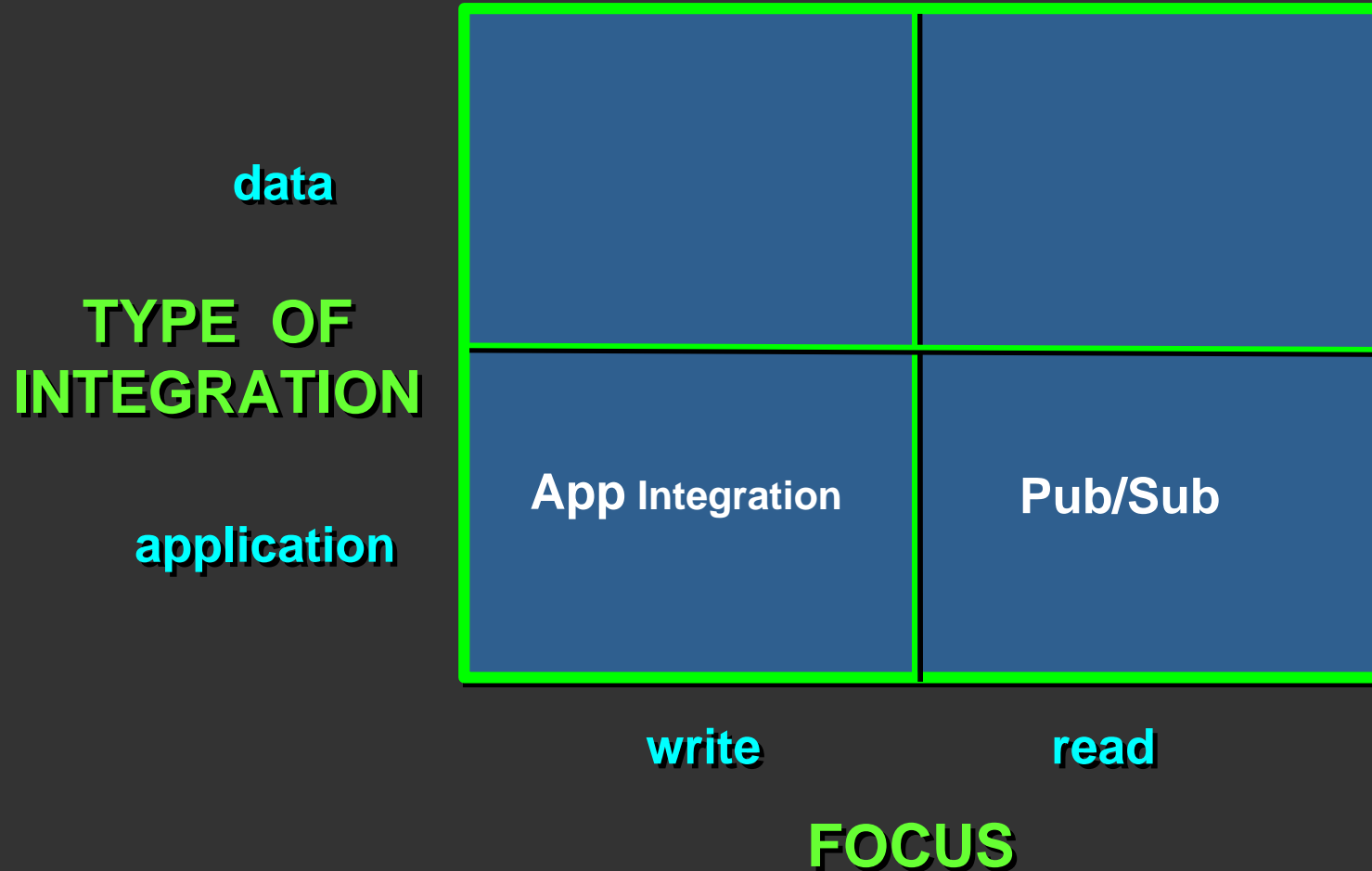
WHICH ONE TO USE?



SOLUTION #2: Publish/Subscribe SYSTEMS



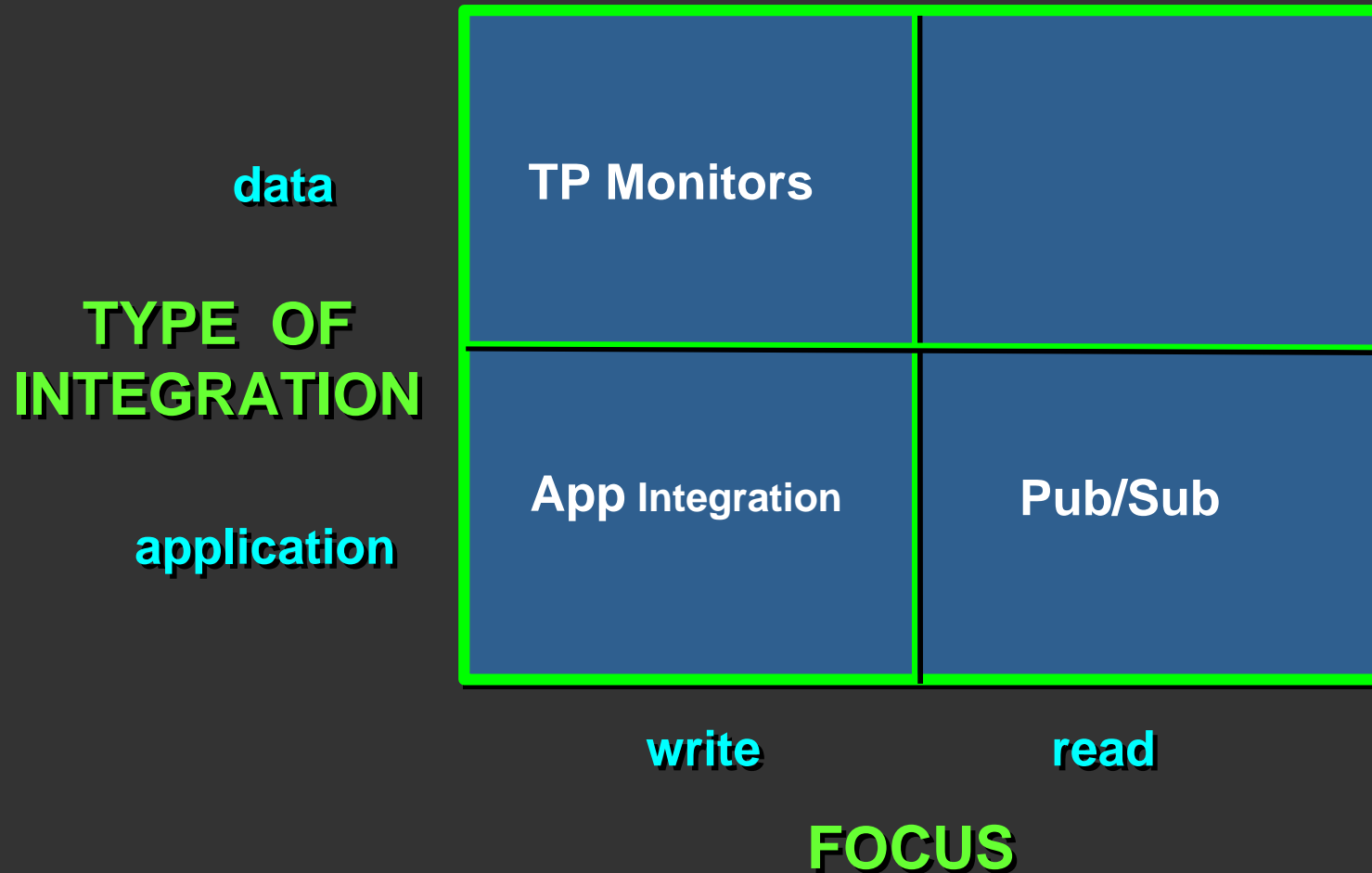
WHICH ONE TO USE?



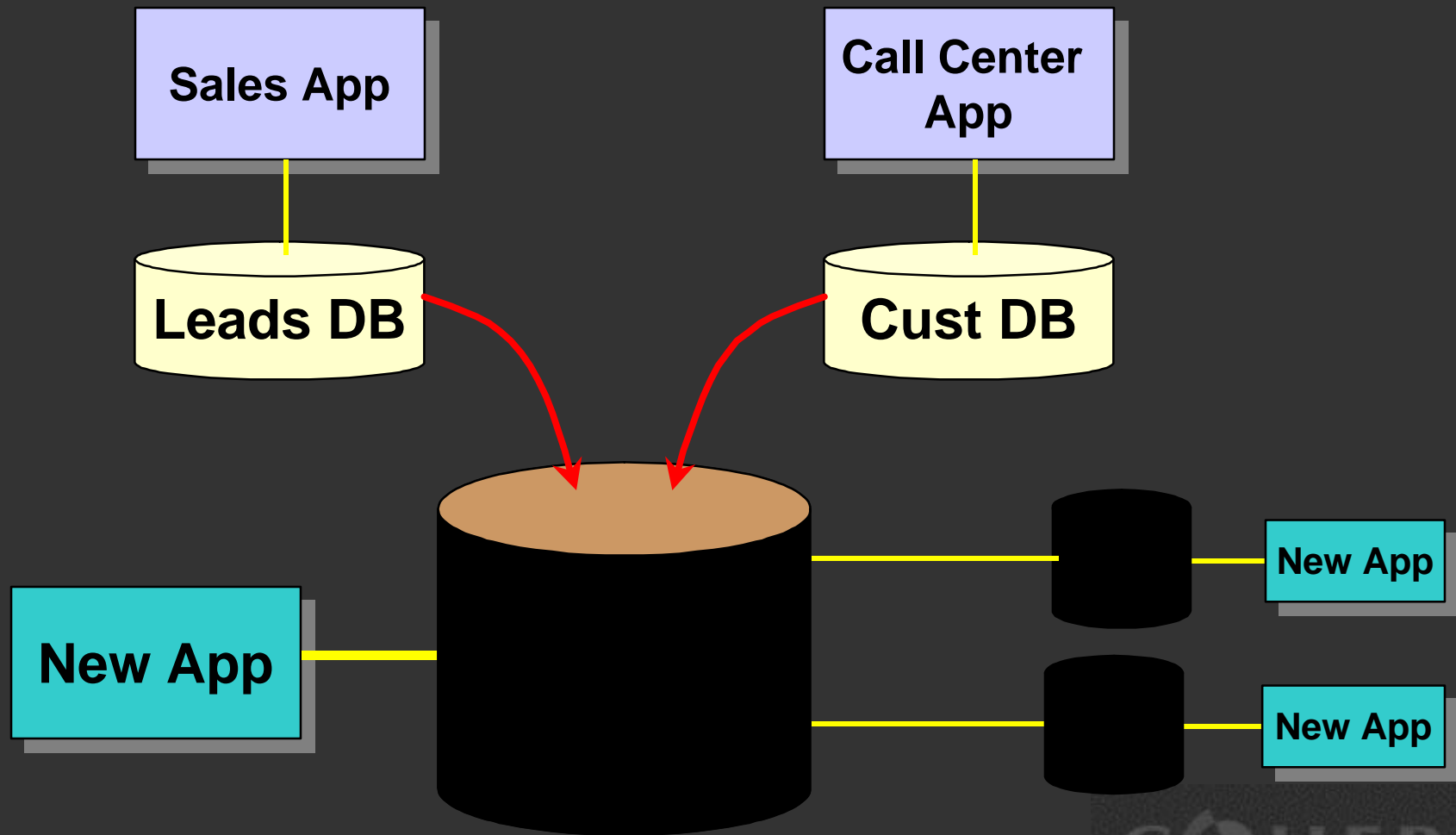
SOLUTION #3: TP Monitors



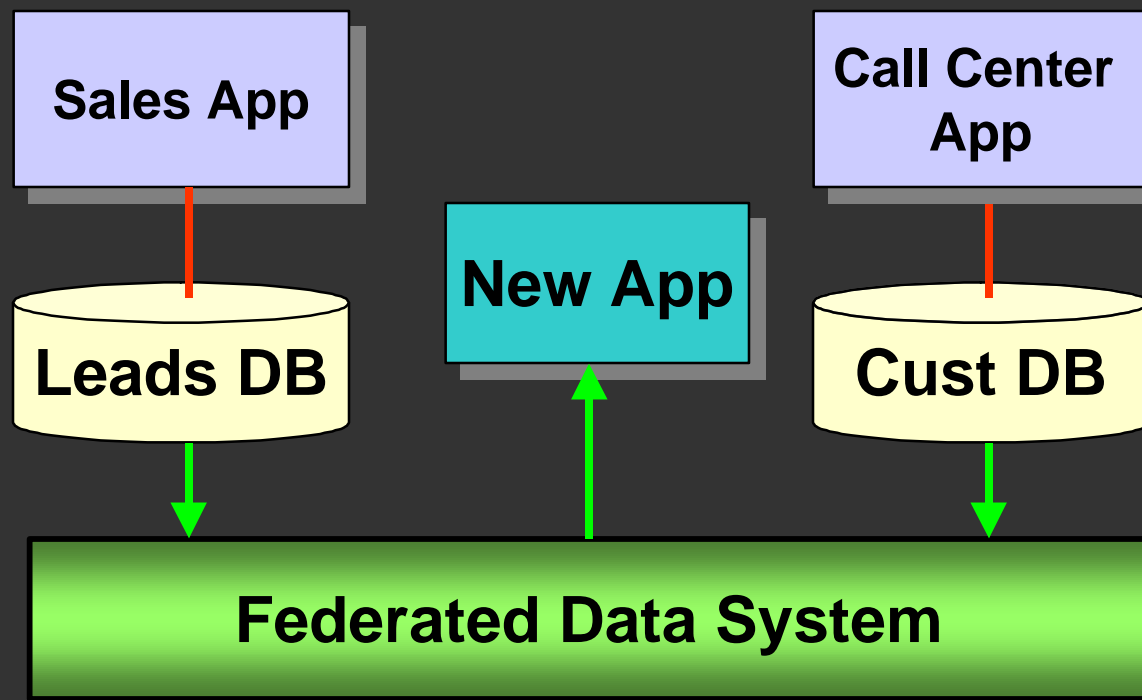
WHICH ONE TO USE?



SOLUTION #4: DATA AGGREGATION



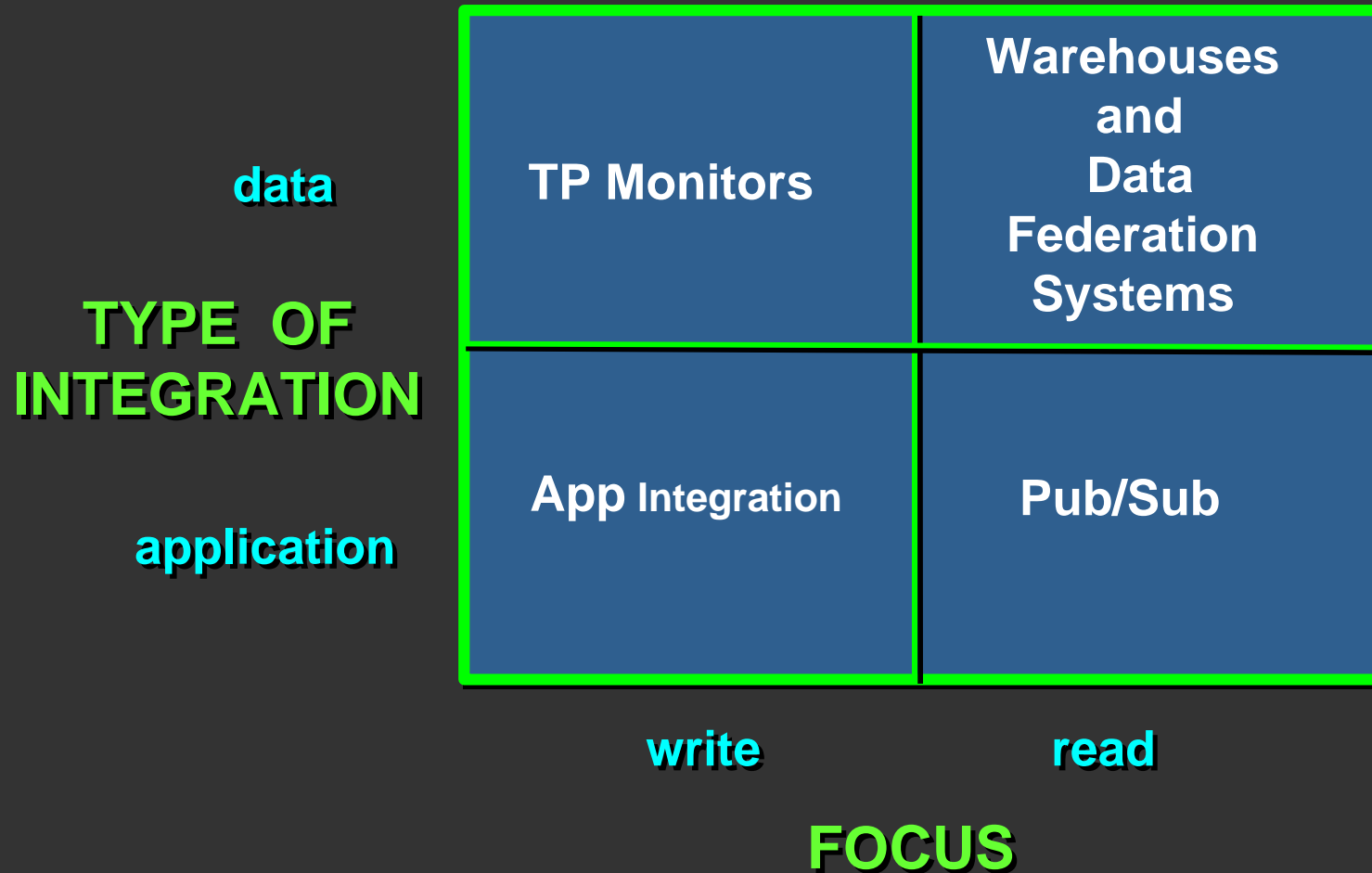
SOLUTION #5: DATA FEDERATION SYSTEM



WHAT IS A Data Federation System?

- **Middleware SQL system**
- **Communicates with local storage manager to run actual queries**
- **Utilize gateway technology to support heterogeneous access**
- **Usually contains a replication engine**

WHICH ONE TO USE?



TECHNOLOGY COMMENT

- **App integration is very hard!**
 - **Integrating two employee records easier than two methods which put an employee on probation**

Forrester reports a factor of 10-20!

TECHNOLOGY COMMENT

- **App integration doesn't scale**
 - all pairs of apps must be “glued”
 - will never get 50,000 pairs done!

ADAGE #1: DO IT AT THE DATA LEVEL!

- **Unless....**
 - **no data base level is present (the sins of our forefathers)**
 - **the packaged app won't let you (e.g. SAP)**

ADAGE #2: THE DRIVE FOR COVERAGE

- **Want to cover the whole matrix because customers have problems in all quadrants**
 - **will drive partnerships**
 - **will drive “reach extension”**

The Drive For Coverage Will Drive Partnerships

- TP monitor plus a data federation system
 - covers top row
- App integration plus Pub/Sub
 - covers bottom row
- Pub/Sub plus data federation system
 - covers right column
- Others may make sense...

THE DRIVE FOR COVERAGE WILL DRIVE REACH EXTENSION

- TP monitors are adding Corba
 - extending down to lower left quadrant
- Data federation systems will add OLTP focus
 - extending to the upper left quadrant

SUMMARY

- **Four kinds of technologies**
- **With different architectures**
- **Good for different things**
- **With different issues**

Use the appropriate one for your problem!