

B13 Federated Database, "IBM's Killer Invention"*Scott Howard, Global Team Leader, BI and Information Integration, IBM*

"(IBM's) Killer Invention": Sam Palmisano, IBM CEO, eWeek, 2/25/2002 IBM's data management philosophy is one of inclusion. The DB2 approach is: keep your data where it makes sense, even if it's in a competitive product like Oracle, Informix or MS SQL Server or a non-relational structure like VSAM and IMS, and DB2 can manage it! This is in stark contrast to other vendors approaches which usually promote keeping your data anywhere, as long as anywhere is within their product set. Scott will introduce the DB2 UDB features that promote such inclusion including a detailed section on federated technology in both DB2 UDB and DB2 DataJoiner Products. Imagine this: an OPTIMIZED four-way join between DB2, IMS, ORACLE, and MS SQL Server. You'll learn the difference between IBM's true Global Optimization and less satisfactory approaches adopted by others. You'll see how DB2's knowledge of specific data sources makes it a powerful federated data mining and business intelligence enabler.

B13

DB2 Federated Technology: "IBM's Killer Invention"

Scott Howard, IBM Learning Services

scottho@us.ibm.com

A decorative graphic consisting of several green circles of varying sizes, some overlapping, arranged in a horizontal line. A central rounded rectangular box with a purple border is superimposed on this graphic, containing the text "IBM Data Management Technical Conference".

IBM Data Management Technical Conference

Anaheim, CA

Sept 9 - 13, 2002

DB2 Federated Technology: "IBM's Killer Invention"

Scott Howard, IBM Learning Services

scottho@us.ibm.com

Materials derived from ILS courses:

CF471, DB2 Federated Database: Integrating Diverse Data
and

DW470, DiscoveryLink: Integrating Diverse Life Sciences Data

For more information or to enroll:

<http://ibm.com/services/learning/us/>

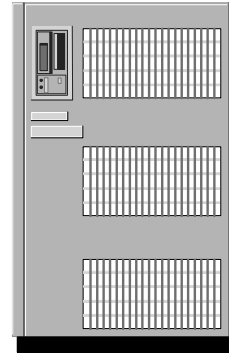
**IBM Employees should use Global Campus and search for course code CF471 or
DW470**

DB2 Relational Connect Home Page:

<http://www.software.ibm.com/data/db2/relconnect/>



Data, Data Everywhere



Flat Files
LOBS
BLOBS
CLOBS

Tabular

Audio

Image

Video



It is all related. What do I do?

How do I coordinate and manage it all?

Accessing and Managing Heterogeneous Data

Two major approaches:

- Move all data into the RDBMS
- Leave it be and access/manage it from within the RDBMS
 - ▶ all access via RDBMS
 - ▶ access via original applications and via RDBMS

Accessing and Managing Heterogeneous Data

Two major approaches:

- Move all data into the RDBMS
- Leave it be and access/manage it from within the RDBMS
 - ▶ all access via RDBMS
 - ▶ access via original applications and via RDBMS
 - ▶ IBM DB2's answer:
DB2 UDB Federated DB, DB2 DataJoiner
and DB2 DataLinks

Summary of Specifications

DB2 Federated Technology provides:

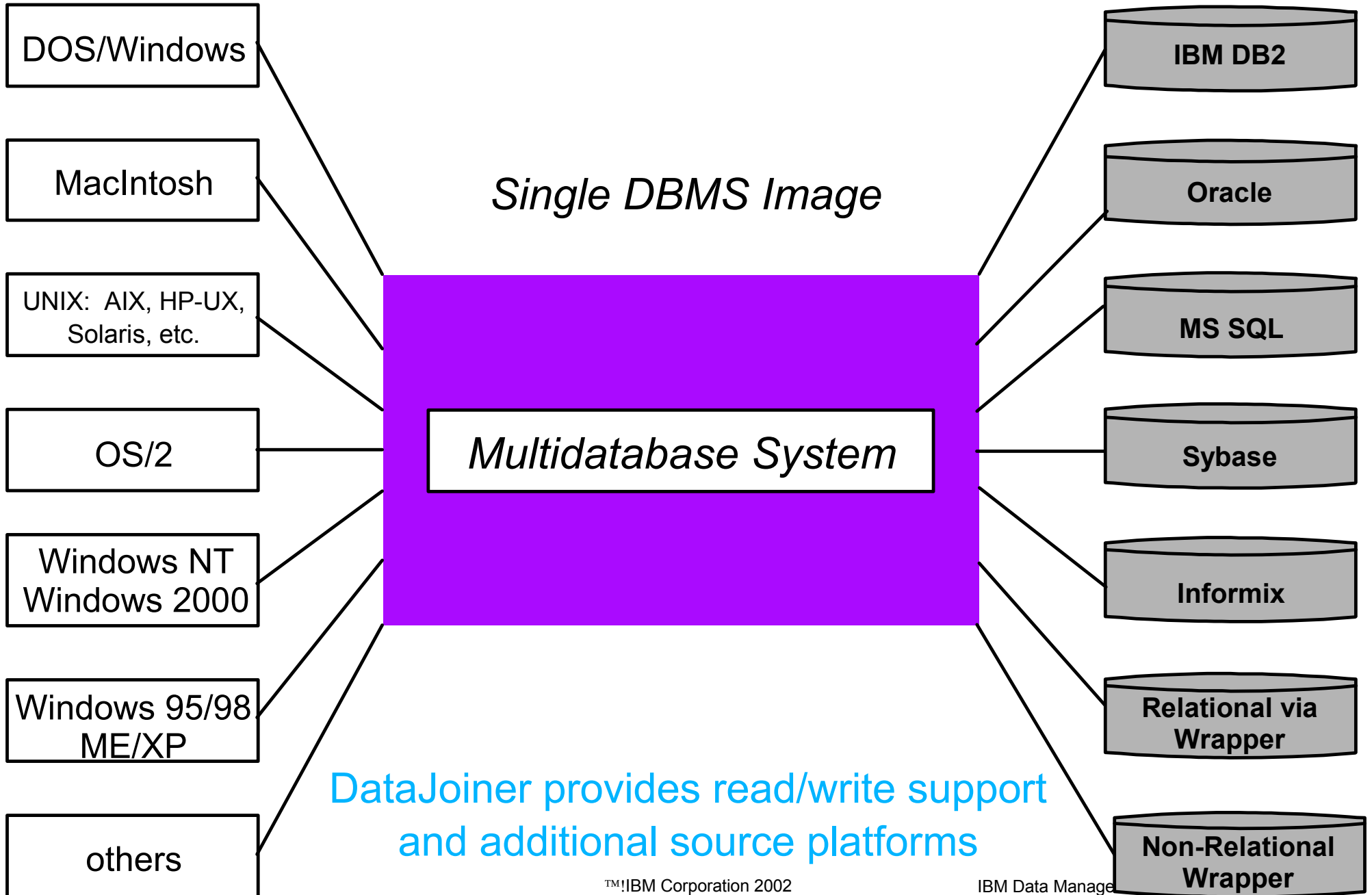
- Multidatabase server supporting multiple simultaneous applications
- Distributed queries including distributed join
 - Read/Write to DB2 Family and Informix
 - [DDL to Read/Write sites \(DDL Transparency\)](#)
- SQL dialect and error code translation
- SQL Compensation
 - Advanced DB2 function against non-DB2 data sources

Summary of Specifications (Cont)

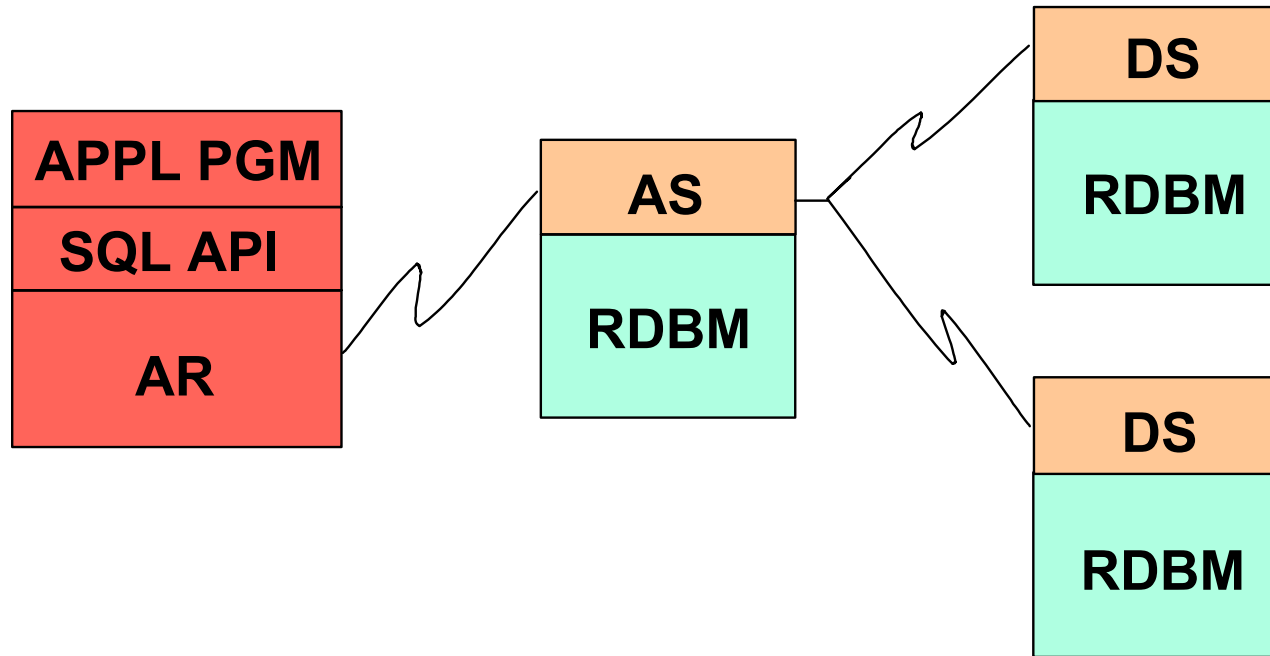
DB2 Federated Technology functions (Continued):

- Advanced optimization and Global Catalog
- Advanced SQL and data type support
- Pass-through option
- Complete local DBMS functions
- High runtime performance

One Global Database

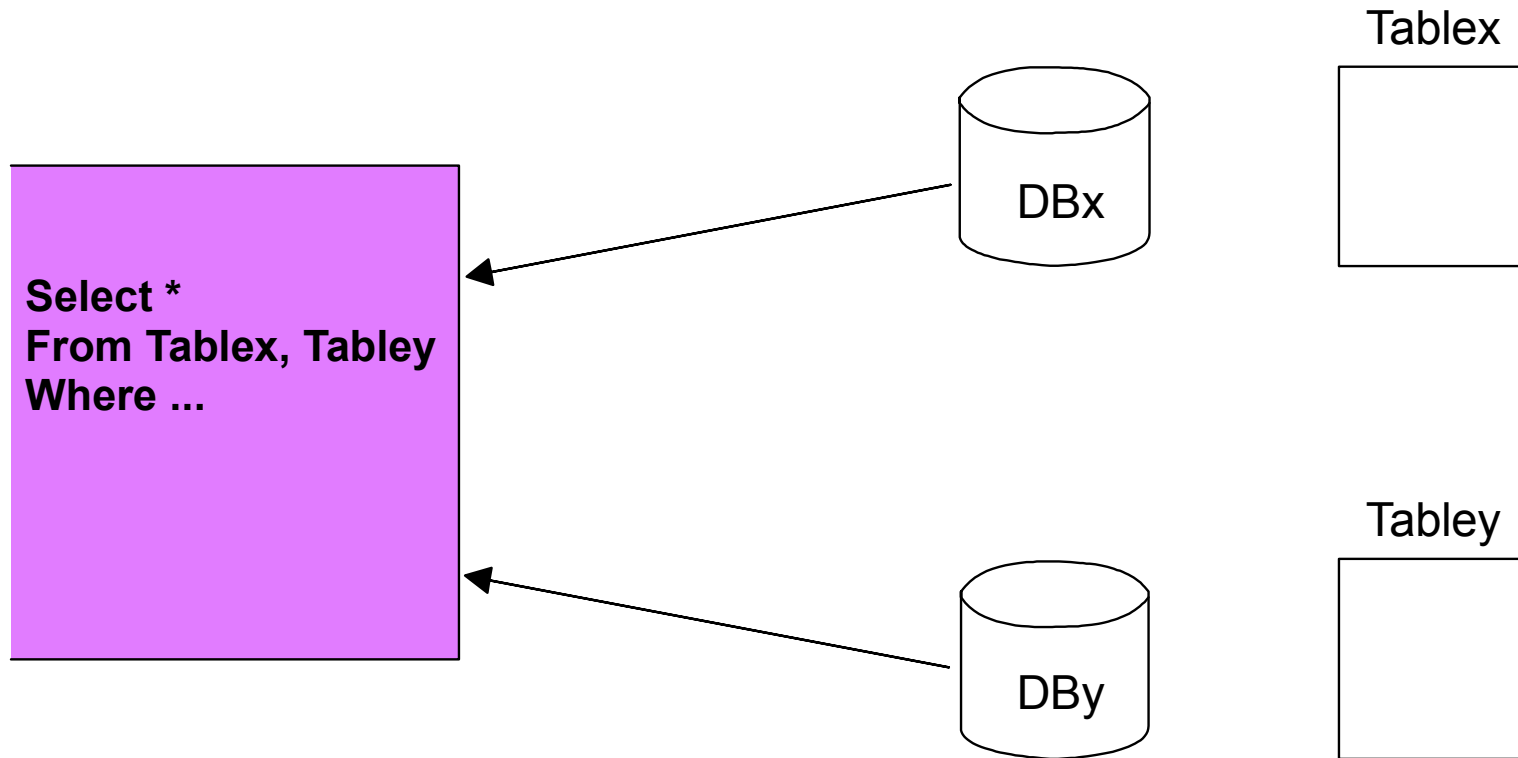


Distributed Request

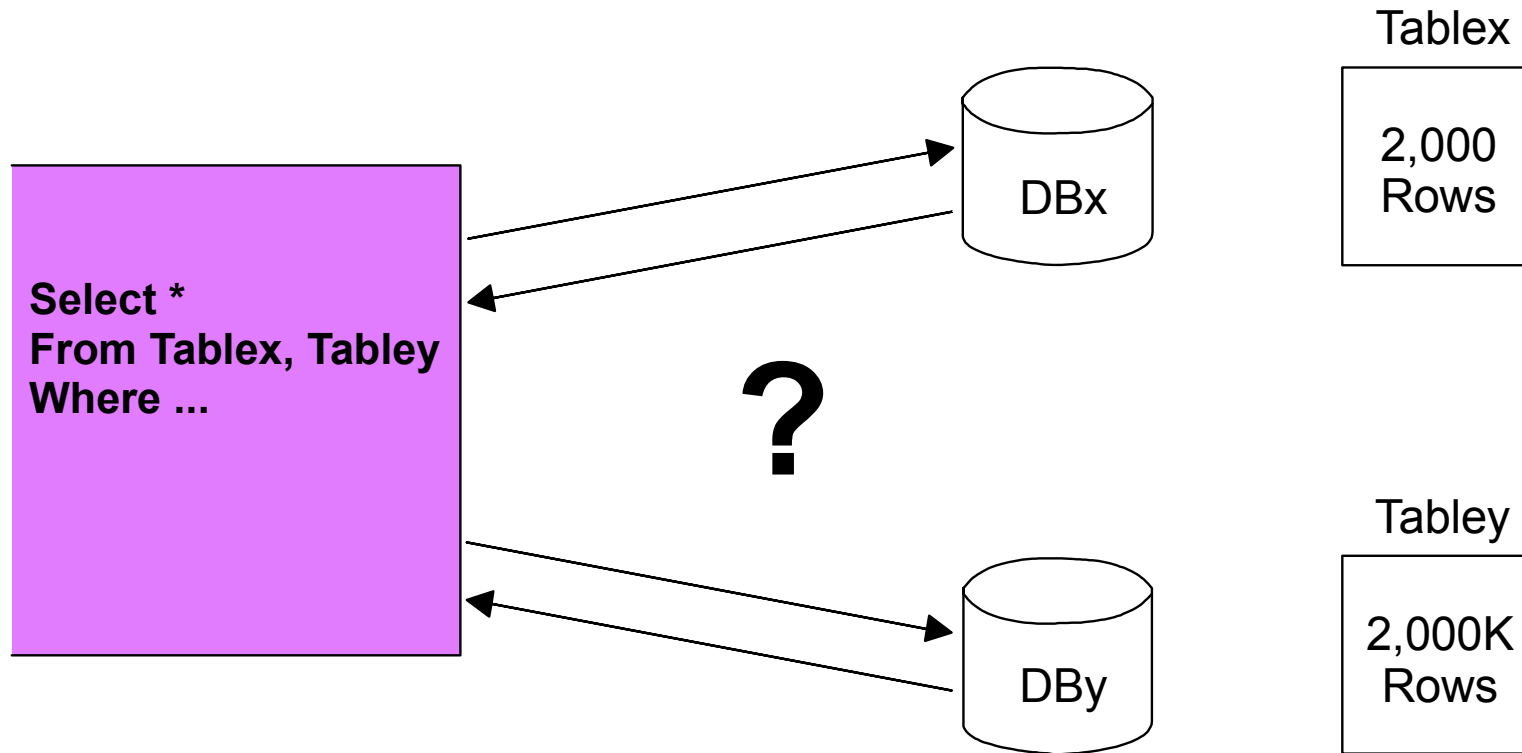


- Multiple RDBMs used in one unit of work
- **Multiple RDBMs used in one SQL statement**
- Requires two-phase commit between AS and DS
- Implies global optimization

Global Optimization

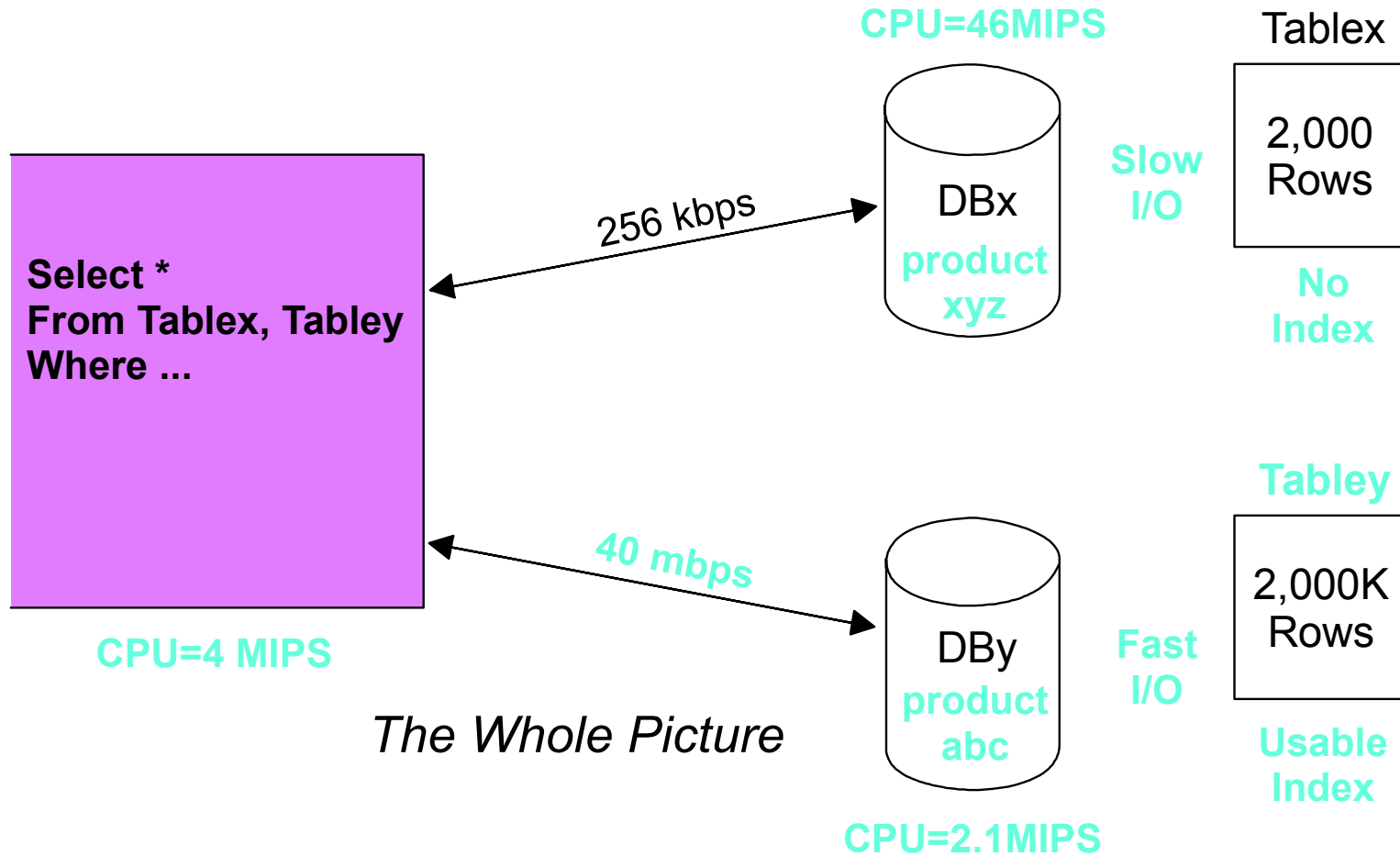


Global Optimization ...

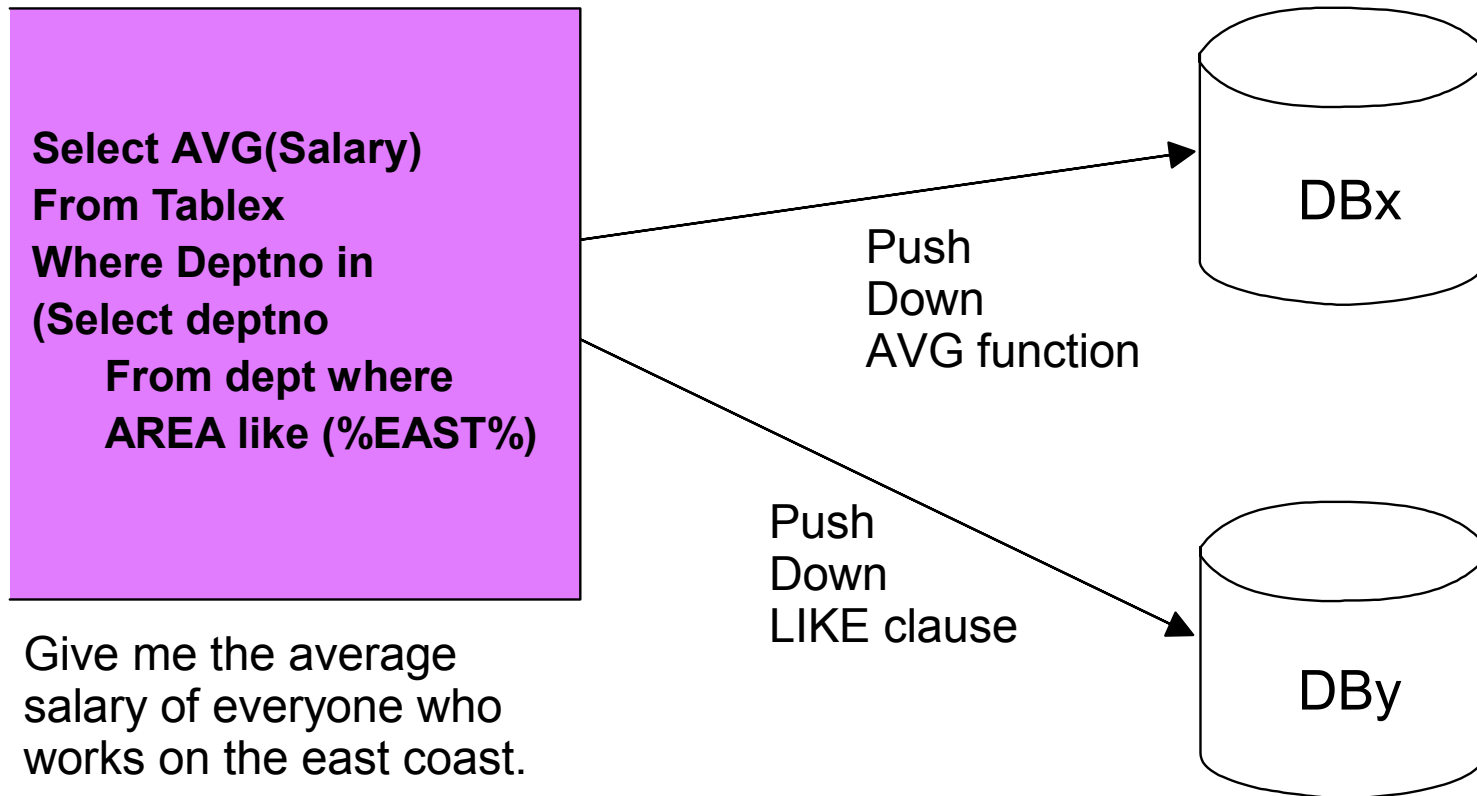


The Picture

Global Optimization ...



Push Down

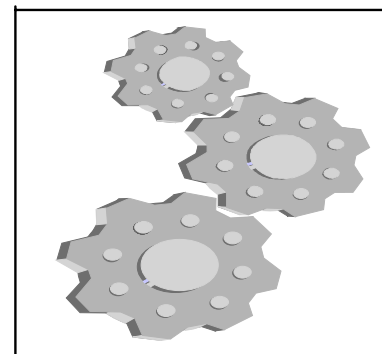


Query Rewrite

Subquery

```
SELECT EMPL.EMPNO  
FROM EMPL, DEPT  
WHERE EMPL.PROJNO =  
  (SELECT PROJ.PNO  
   FROM PDEP, PROJ  
   WHERE EMPL.PROJNO=PROJ.PNO  
   AND DEPT.DEPTNO=PDEP.DEPTNO  
   AND PDEP.PNO=PROJ.PNO )
```

DB2 Query Rewrite



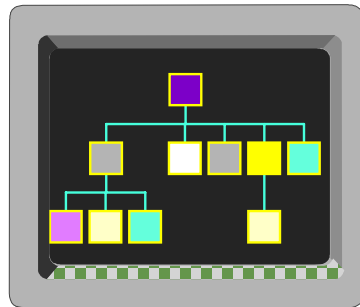
Join

```
SELECT E.EMPNO  
FROM EMPL E, DEPT D,PDEP P1,PROJ P2  
WHERE E.PROJNO=P2.PNO  
AND D.DEPTNO=P1.DEPTNO  
AND P1.PNO=P2.PNO  
AND E.DEPTNO=D.DEPTNO
```

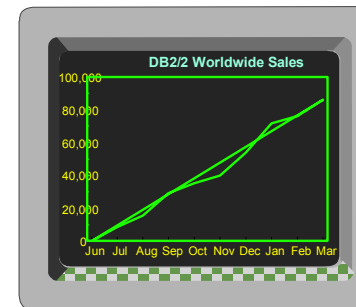
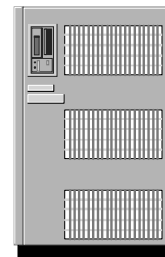
Visual Explain & PM

Visual Explain*

```
SELECT AVG(AVG_SAL)
FROM (
  SELECT DEPT,
         AVG(SALARY) AS AVG_SAL
  FROM DEPARTMENT
  GROUP BY DEPT
) AS AVG_DEPT_SAL
```

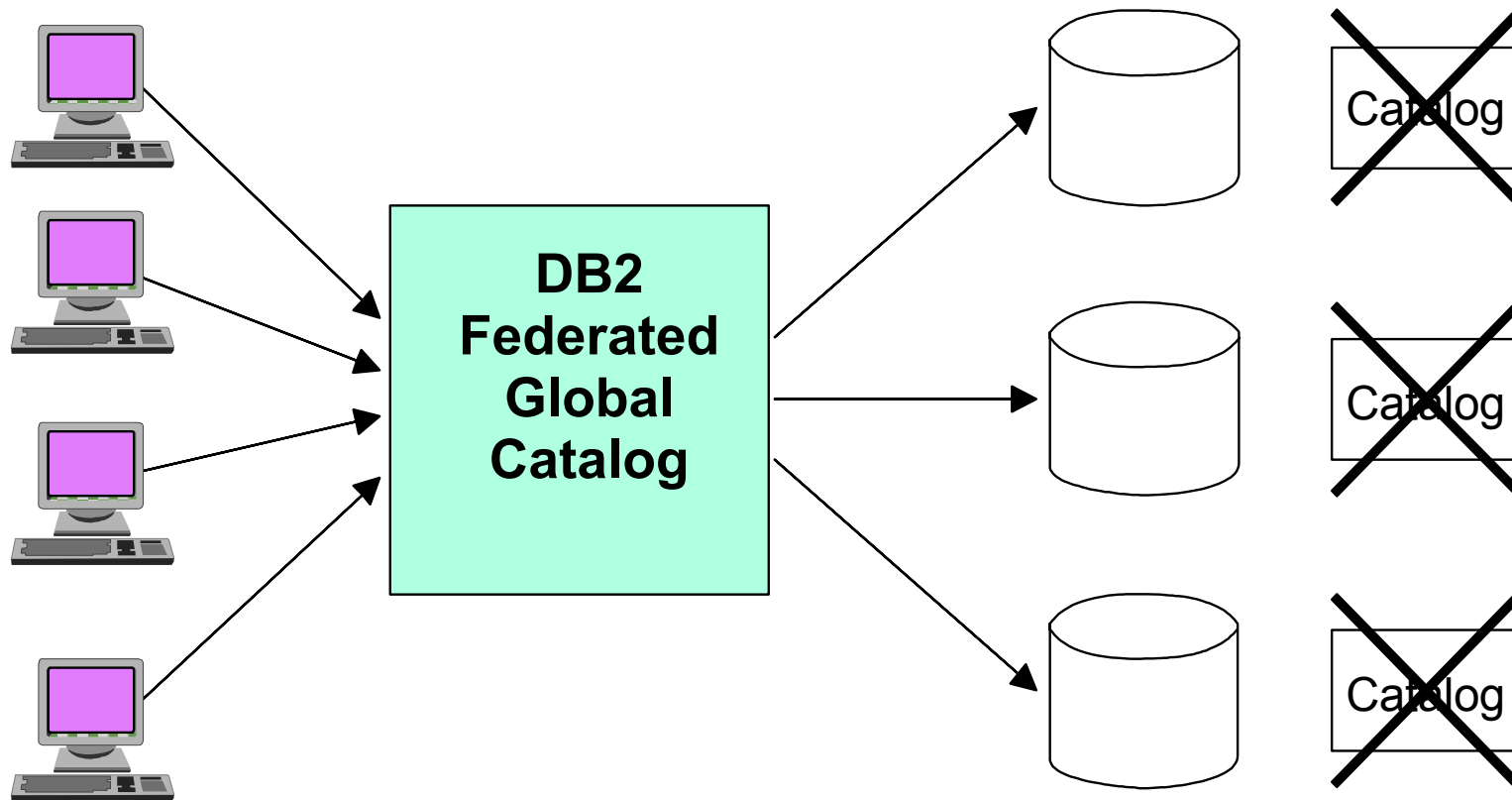


Performance Monitor*

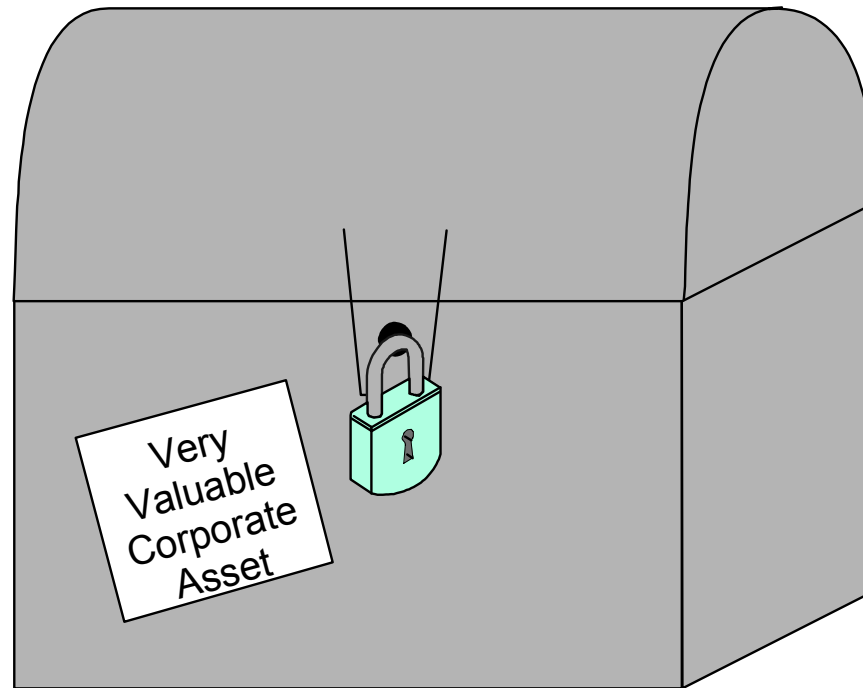


*** With extensions to
remote data sources**

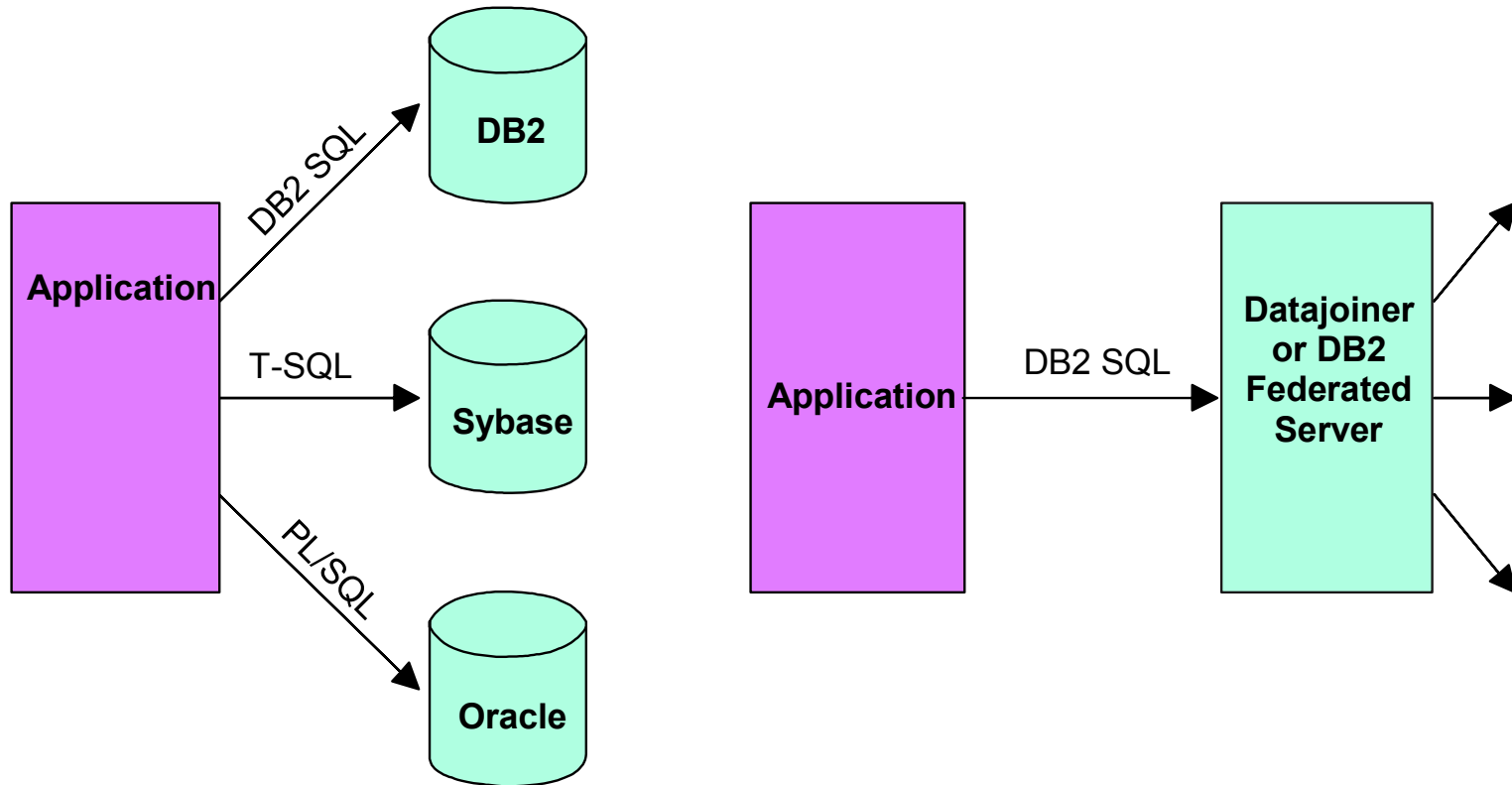
Global Catalog



Security, Integrity and Recovery

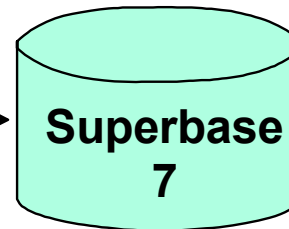


Single SQL



Compensation

```
DECLARE K9  
CURSOR WITH HOLD  
FOR SELECT NAME  
FROM USELESS.TABLE  
WHERE SALARY > :MINE
```



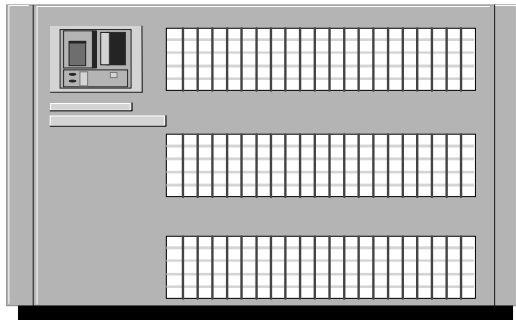
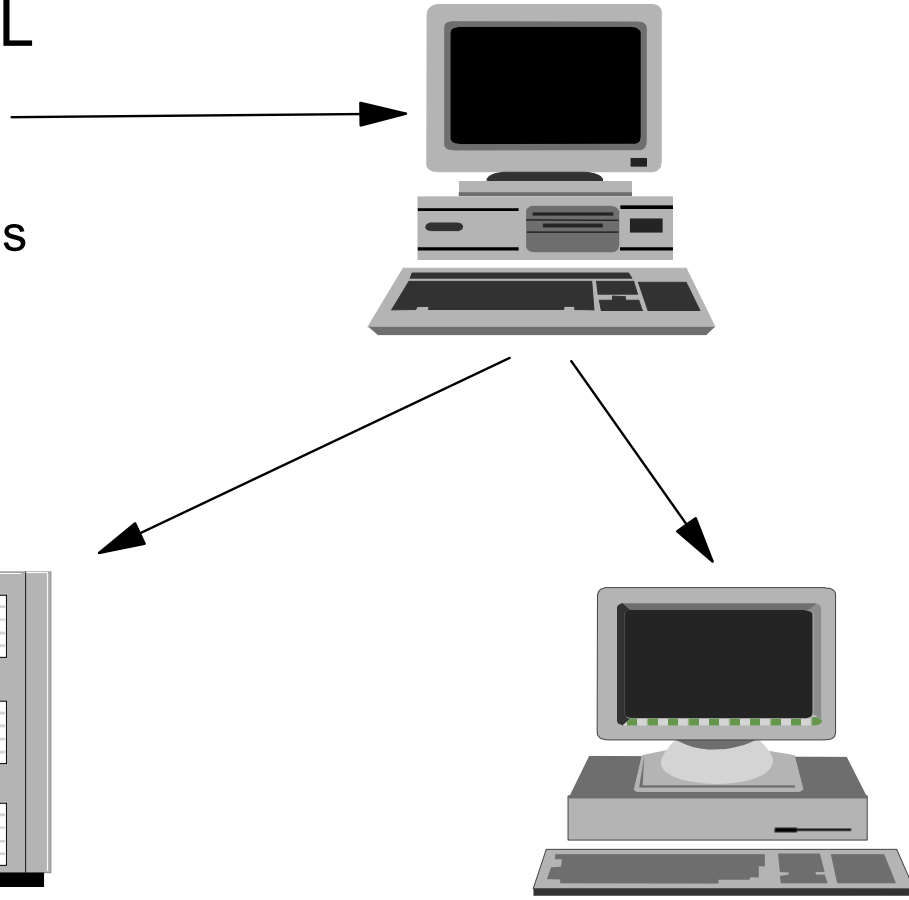
*I don't
do WITH
HOLD!*

Think of the possibilities:

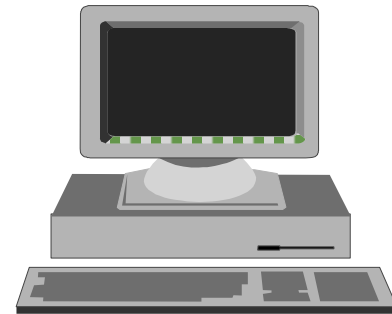
DB2 Federated Server

Advanced DB2 SQL

- Case expressions
- Recursive SQL
- User Defined Functions
- Nested Table Exp.
- Summary Tables*
- etc.



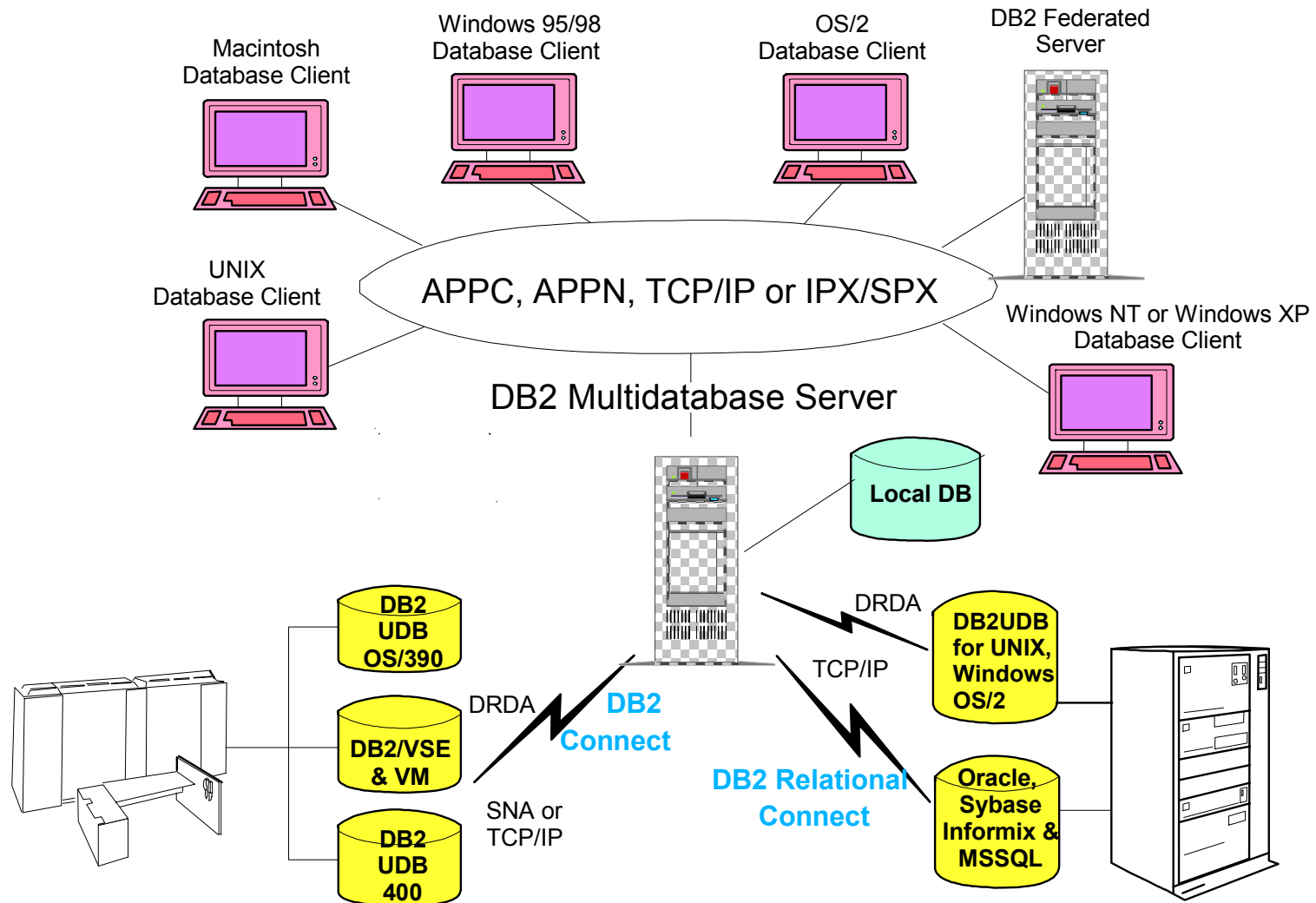
Informix



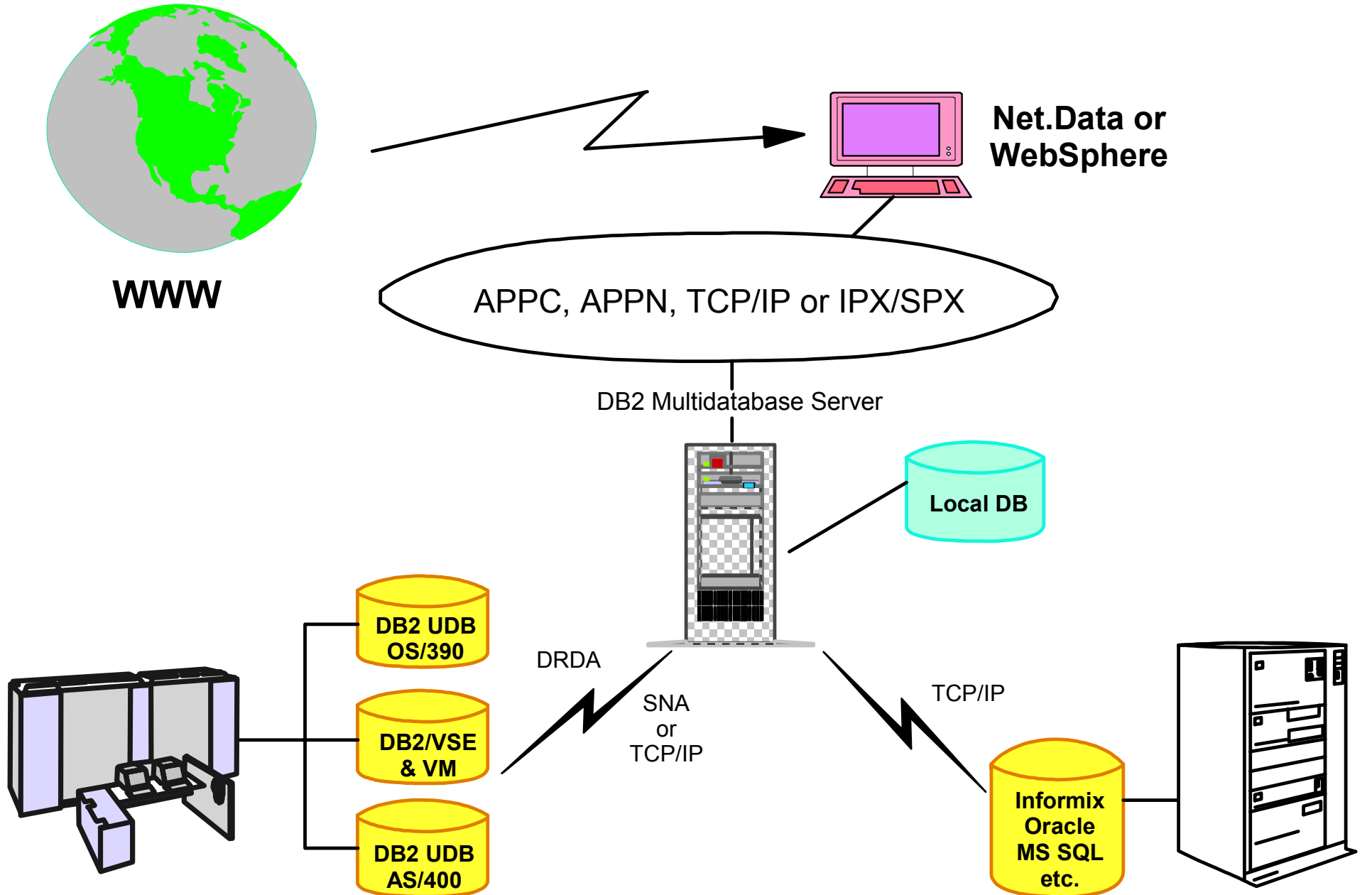
Oracle

*** Yes you can create Summary Tables (Materialized Query Tables) over nicknames!**

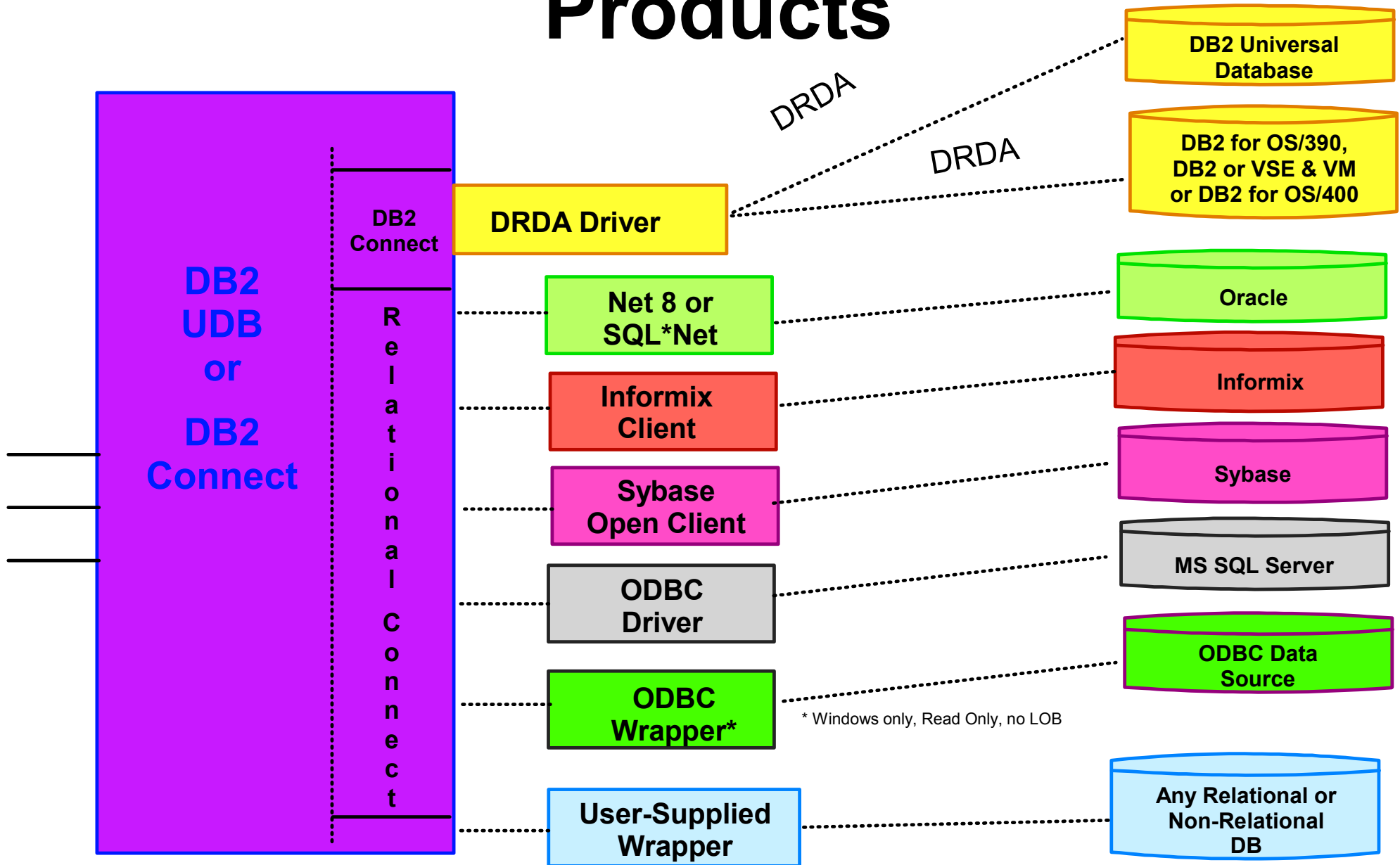
DB2 Federated Client/Server Environment



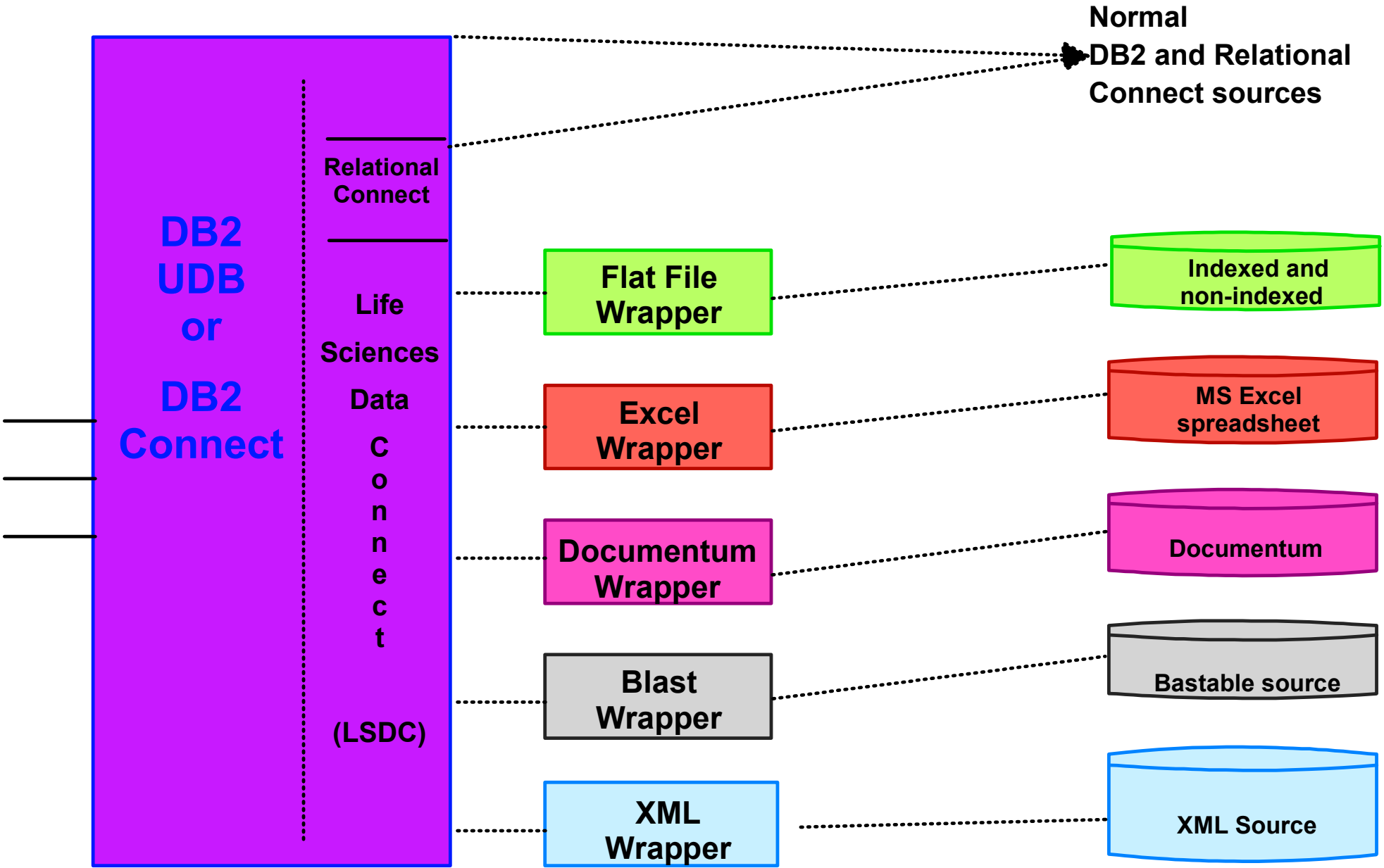
Don't Forget the WEB



Multidatabase Server Services and Products

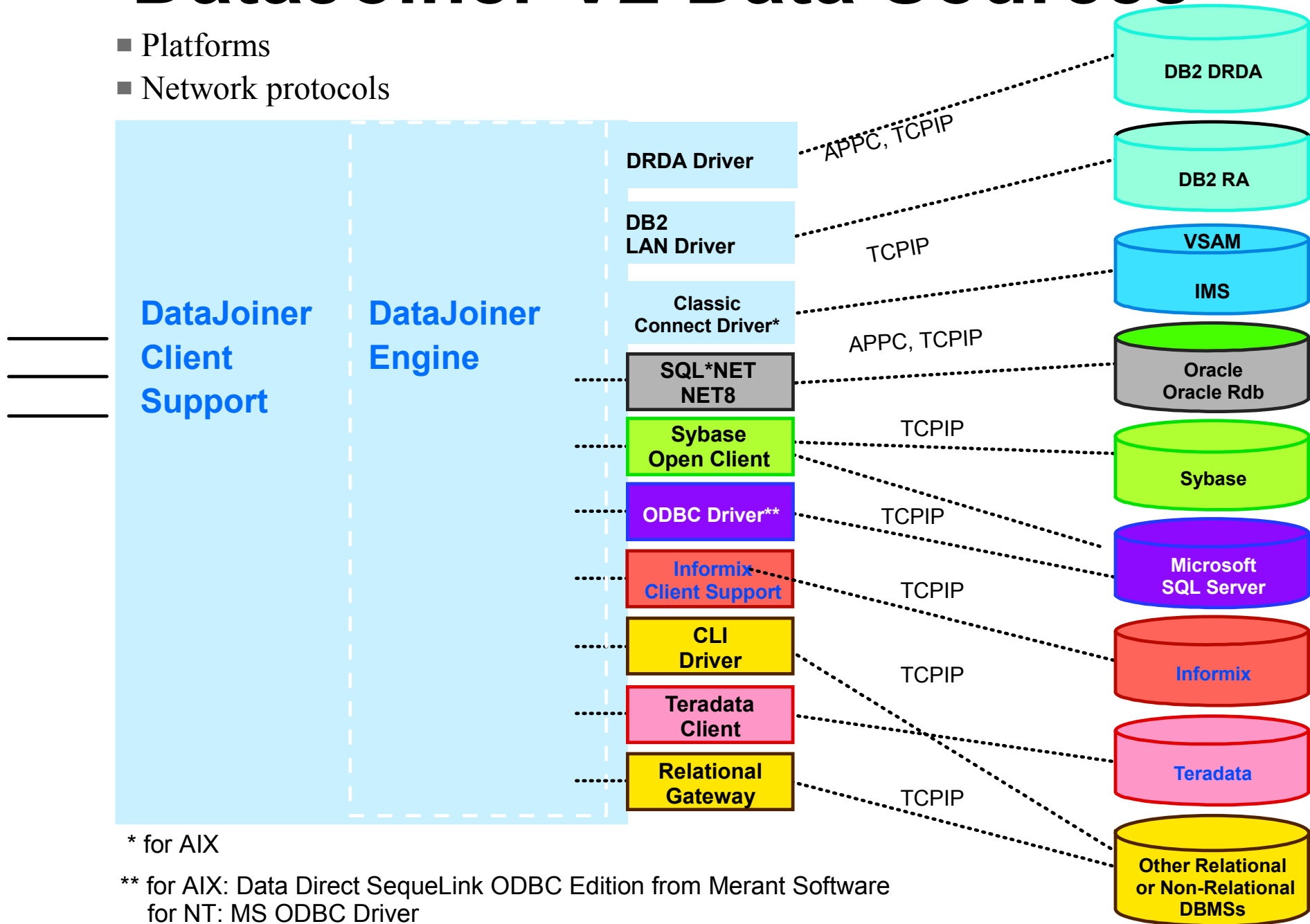


Life Sciences, DiscoveryLink



DataJoiner V2 Data Sources

- Platforms
- Network protocols



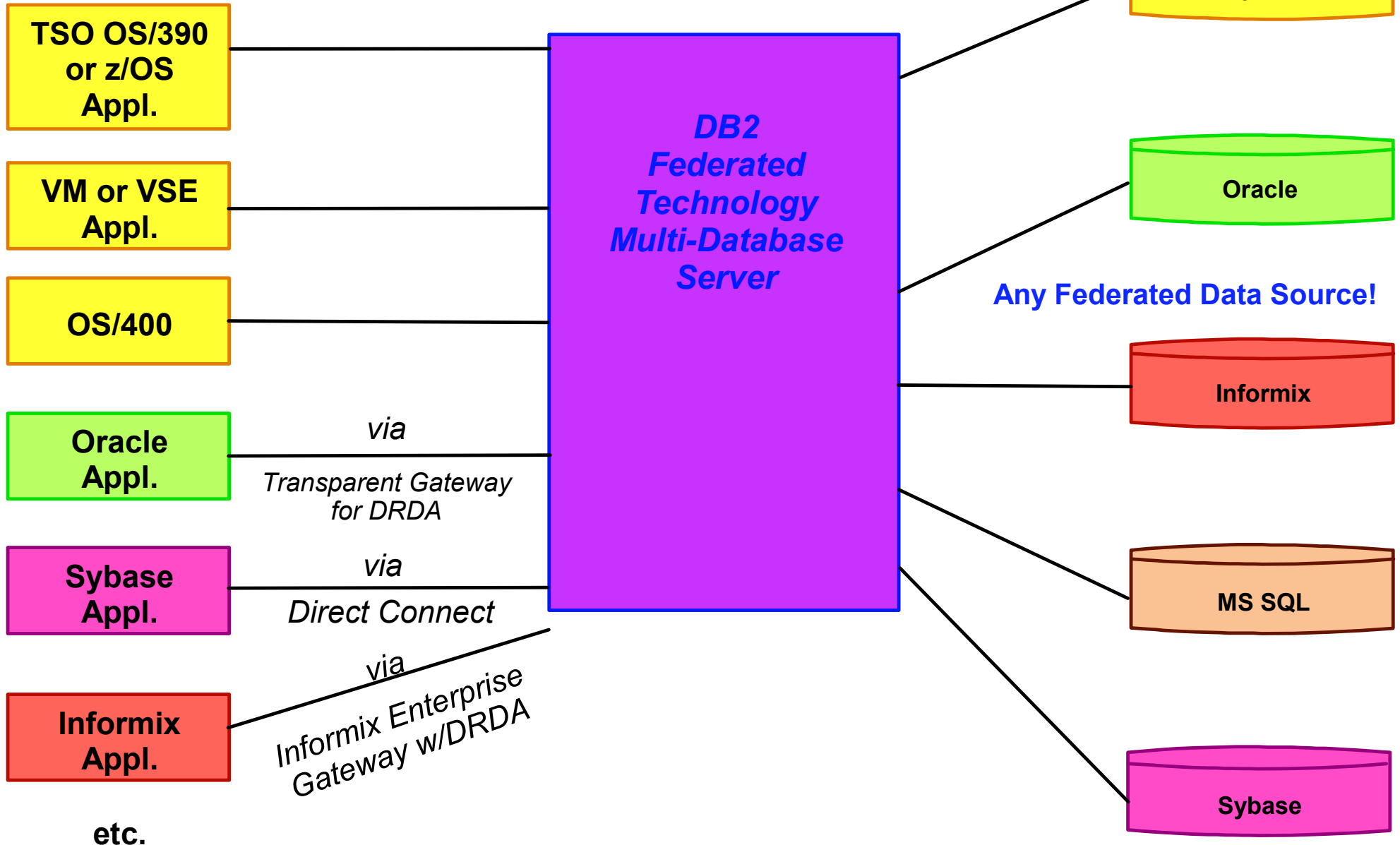
* for AIX

** for AIX: Data Direct SequeLink ODBC Edition from Merant Software
for NT: MS ODBC Driver



DRDA AS Support

Any DRDA AR Application:



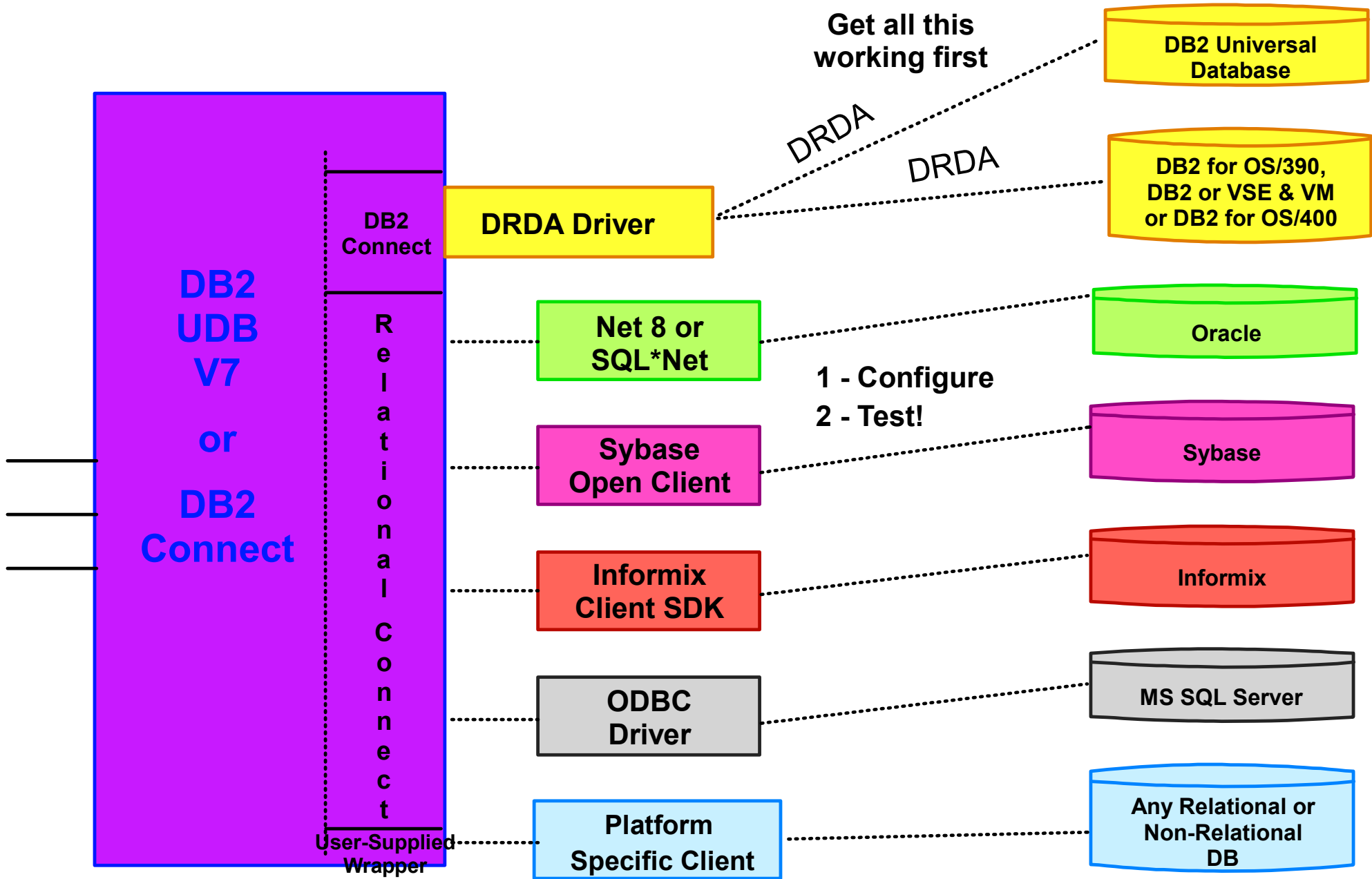


Configuring the Federated Environment

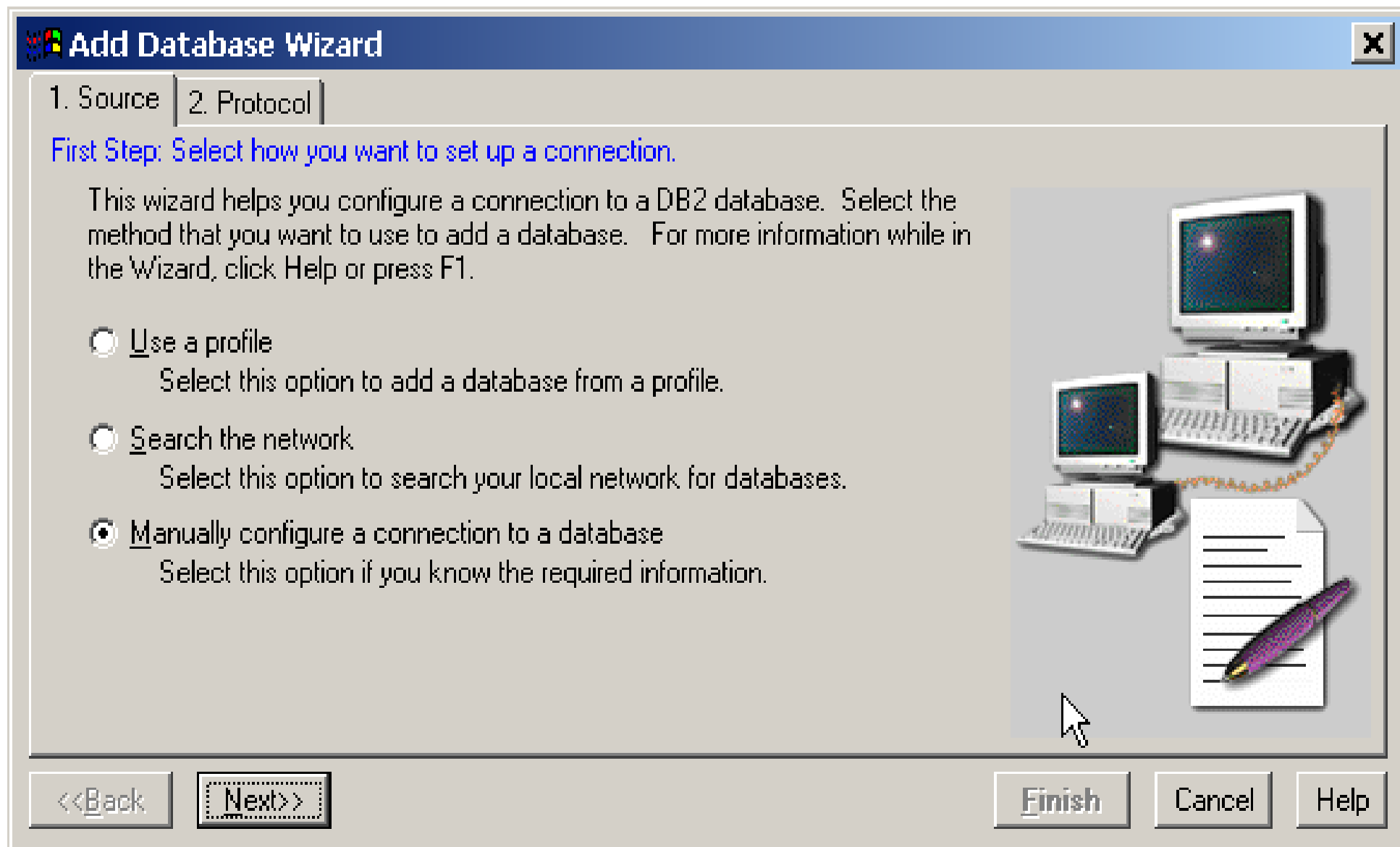
Installation Steps

1. **Install any non-DB2 client products**
 - For example, Oracle NET8
 - Ensure environmental variables are set (that is, ORACLE_HOME)
2. **Install DB2 UDB and/or DB2 Connect**
3. **Install DB2 Relational Connect**
 - Select the data sources that you intend to access
- 3a. **(UNIX only) Run djxlink**

Ready the Client Connectivity

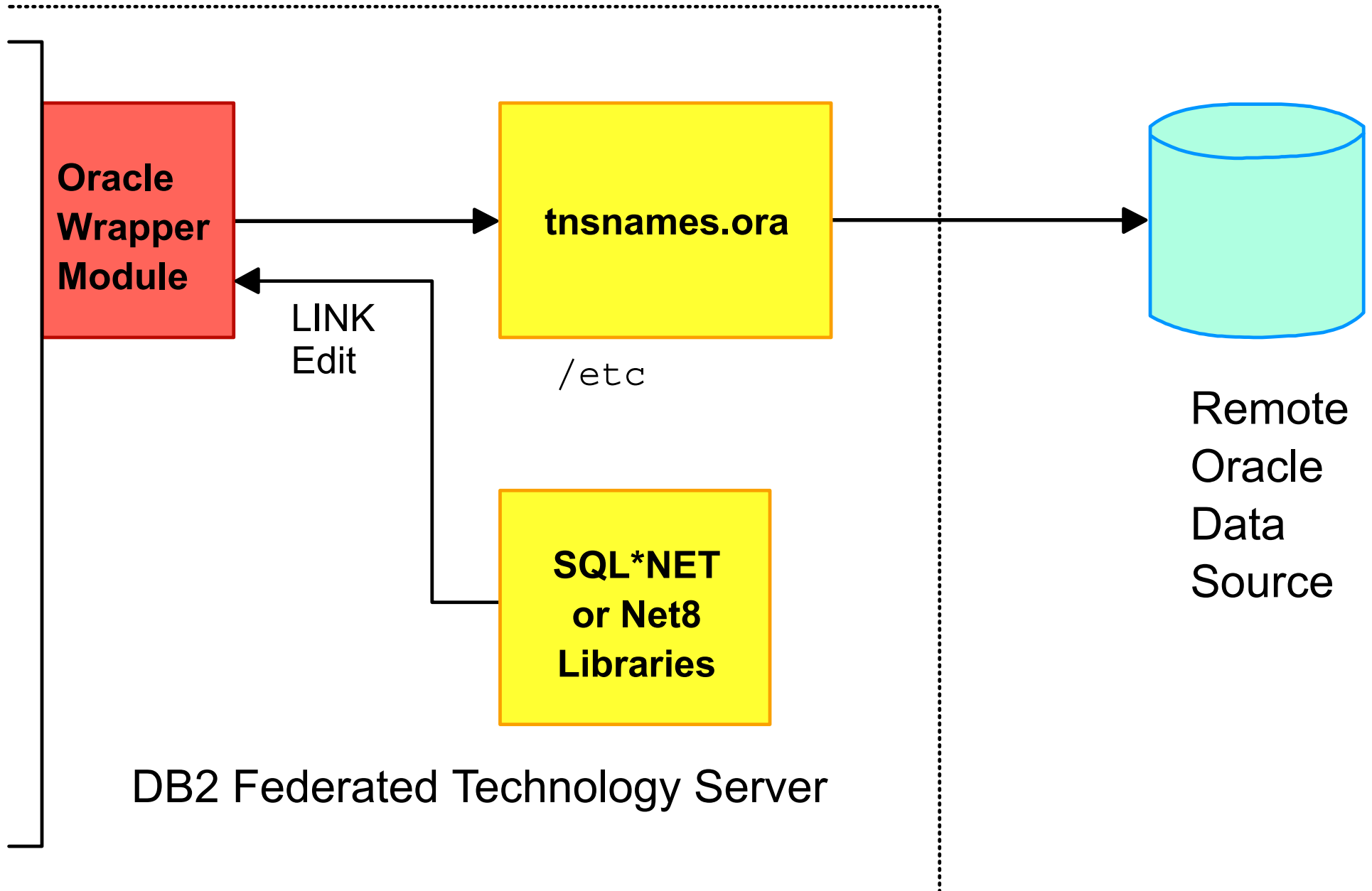


DB2 Data Sources



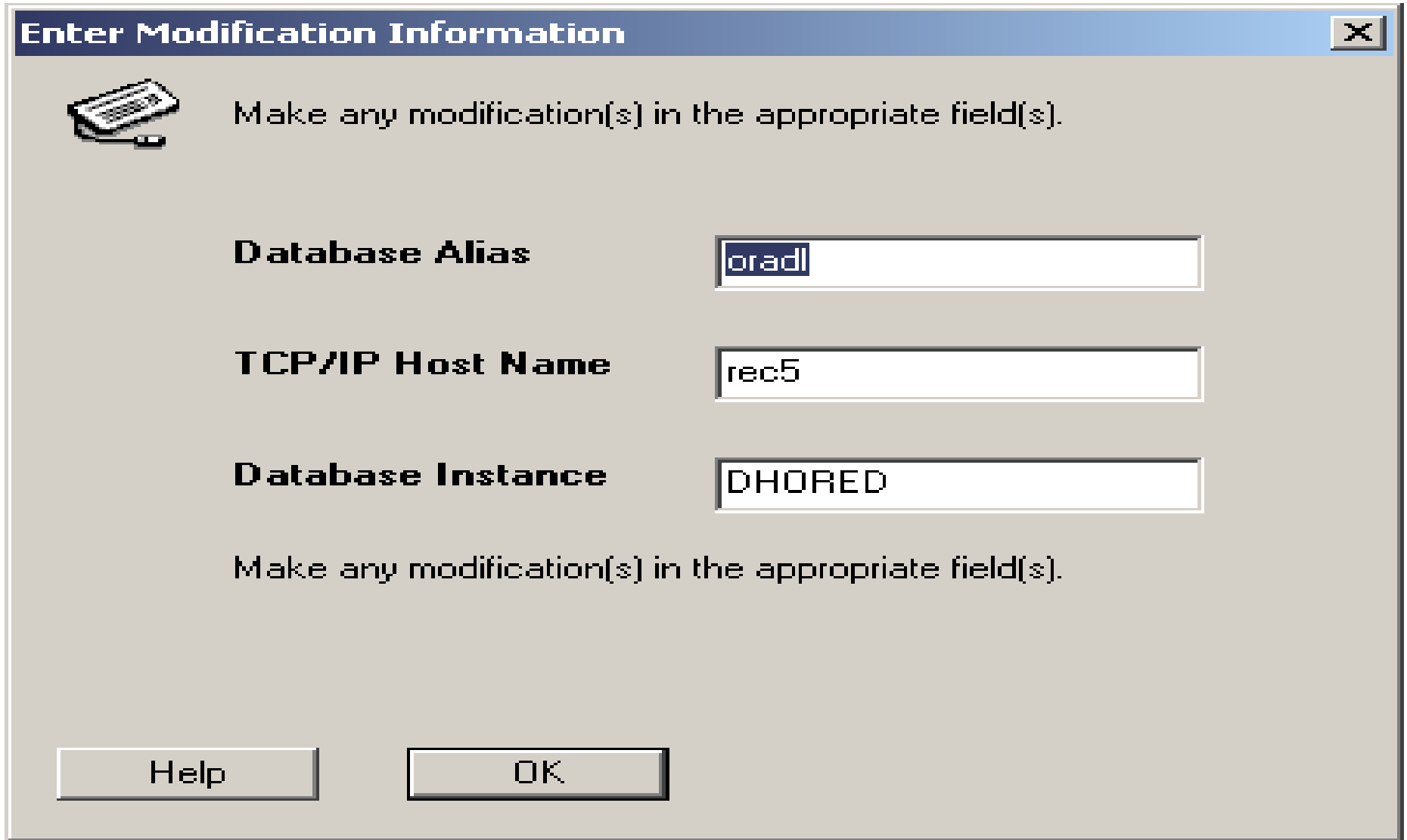
! Test via Command Center, then create the DRDA Wrapper !

ORACLE Wrapper Module Setup




Oracle Setup Tools

Windows: Use the SQL*Net Easy setup or Easy config tool to configure connectivity.



Enter Modification Information [X]

 Make any modification(s) in the appropriate field(s).

Database Alias

TCP/IP Host Name

Database Instance

Make any modification(s) in the appropriate field(s).

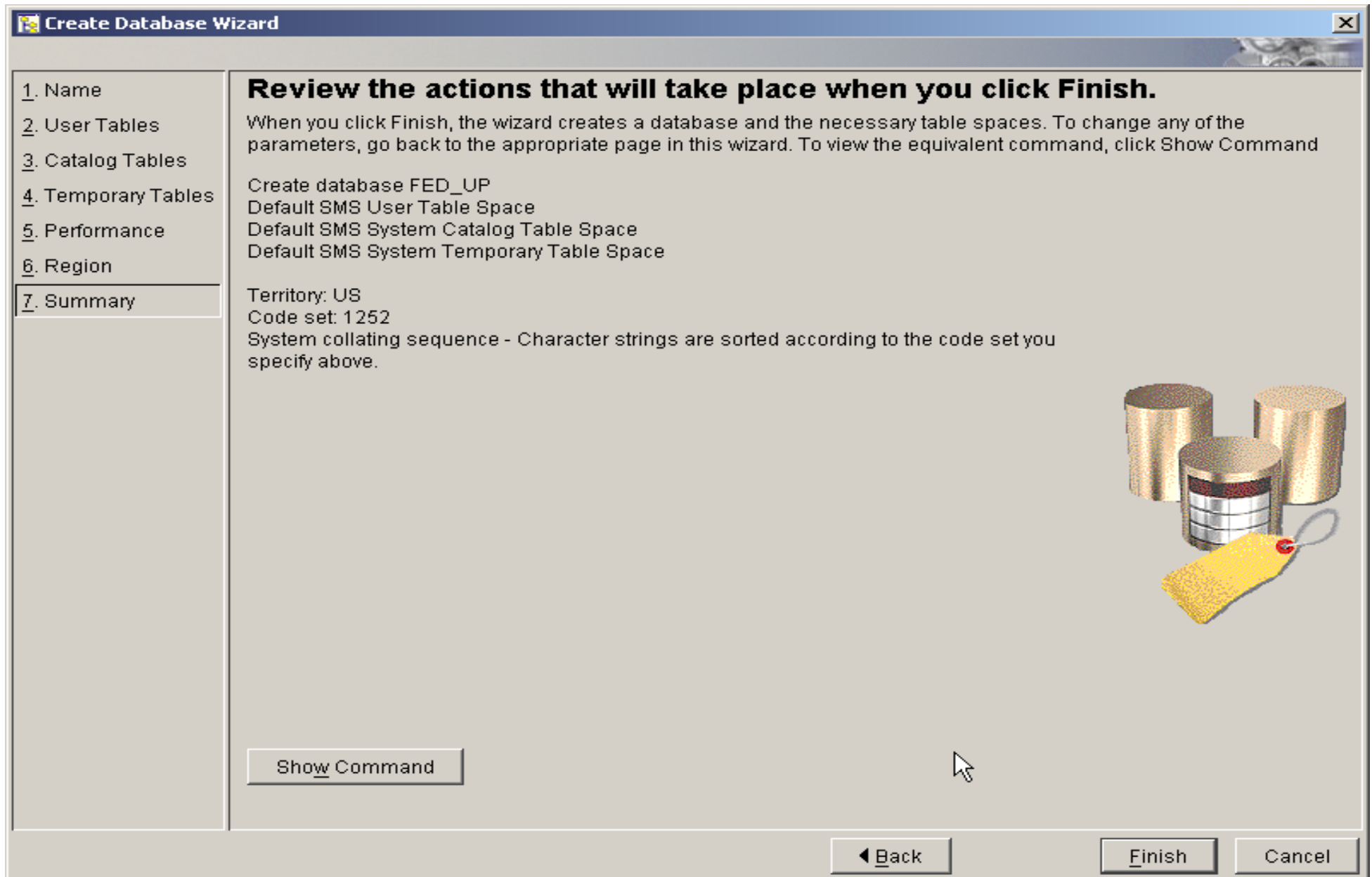
Verify Your Oracle Connection

Use:

1. ping
2. tnsping
3. sqlplus

Continue by installing and configuring any other required client enablement on the Federated Server (i.e. Informix Client, MS SQL Server ODBC drivers...)

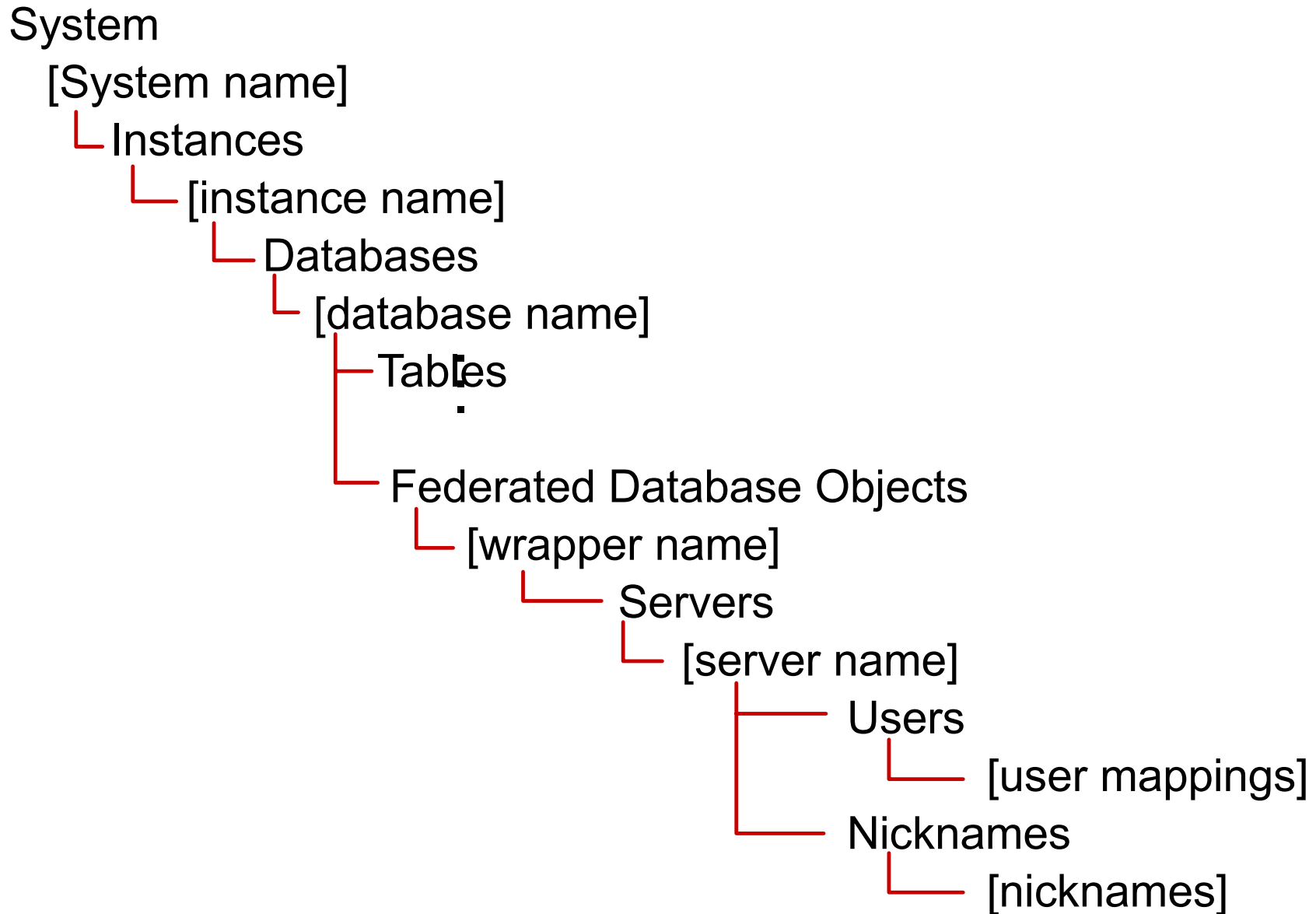
Create the DB2 Federated Server Database



Wrappers - Servers - Users - Nicknames

- In DB2 UDB, there is a RI relationship between Wrappers, Servers, Nicknames, and User Mappings
 - ▶ CONNECT to a DB2 UDB Database
 - ▶ 1st: CREATE WRAPPER
 - ▶ 2nd: CREATE SERVER
 - ▶ 3rd-a: CREATE USER MAPPING's
 - ▶ 3rd-b: CREATE NICKNAME's
- DROPPing a SERVER will drop all dependent USER MAPPINGS and NICKNAMES
- DROPPing a WRAPPER will drop all dependent SERVERS, USER MAPPINGS, and NICKNAMES.
- Also, this is evident in the DB2 Control Center 'tree'

DB2 UDB Control Center



CREATE WRAPPER

Wrapper-names:

DRDA	For all DB2 data sources
SQLNET	For Oracle via SQL*NET
NET8	For Oracle via net8
ctlib	For Sybase via ctlib
dblib	For Sybase and Microsoft SQL Server via Sybase dblib
DJXMSSQL3	Microsoft SQL Server via ODBC 3.0 (or higher) driver (Windows NT/2000)
MSSQLODBC3	Microsoft SQL Server via Merant Data Direct Connect ODBC (AIX)
INFORMIX	For Informix via Informix client
OLEDB	All OLE DB providers supported by Microsoft OLE DB

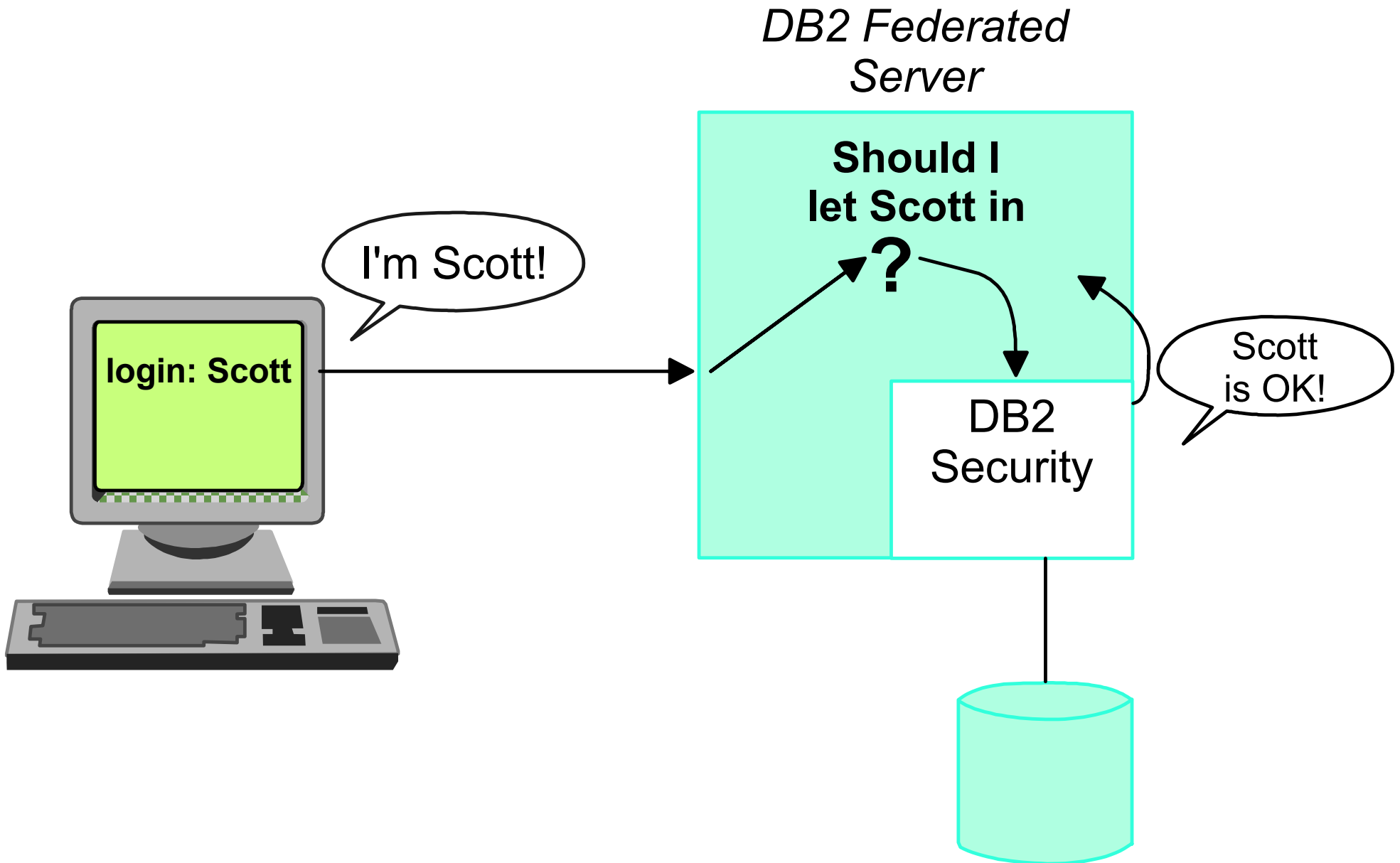
CREATE SERVER

```
CREATE SERVER server_name TYPE server_type  
VERSION server_version WRAPPER wrapper_name  
AUTHORIZATION remote_id PASSWORD remote_pw  
OPTIONS (...)
```

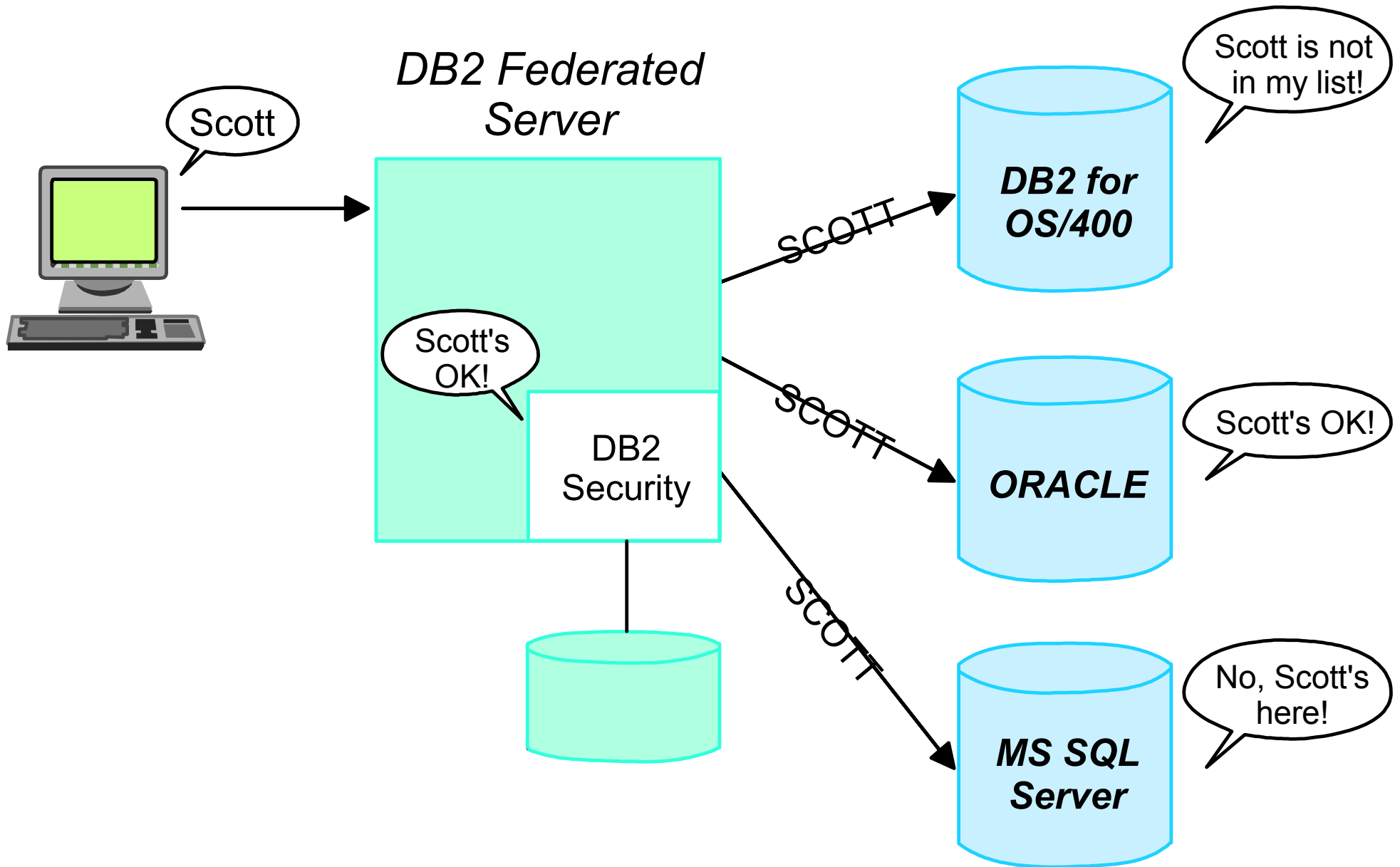
- NODE option is always required
- DBNAME option required except for Oracle

```
create server nydb type db2/390 version 6.1  
wrapper drda authorization myid password mypw  
options (node 'TOMVS', dbname 'DSNT')
```

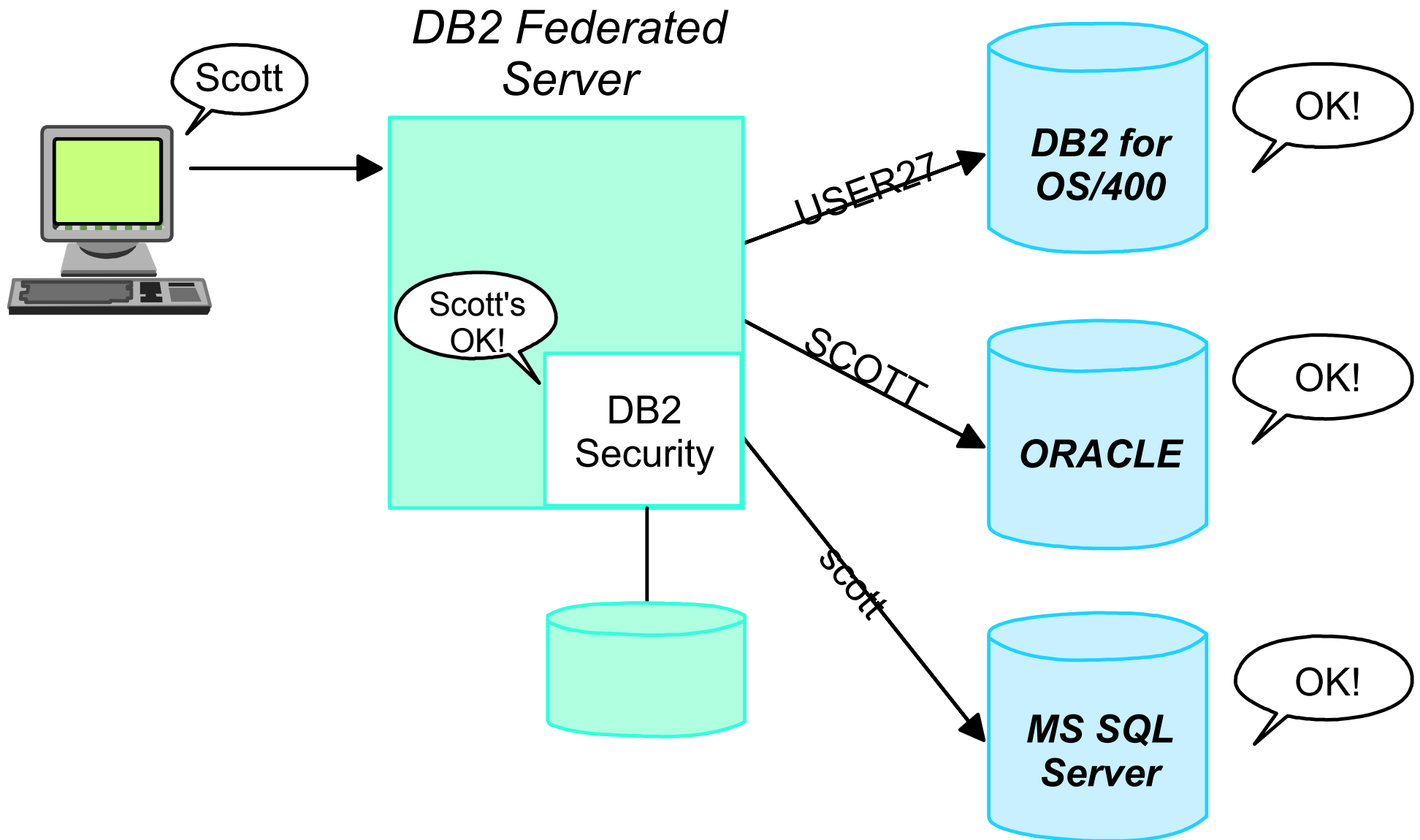
Single Server Security



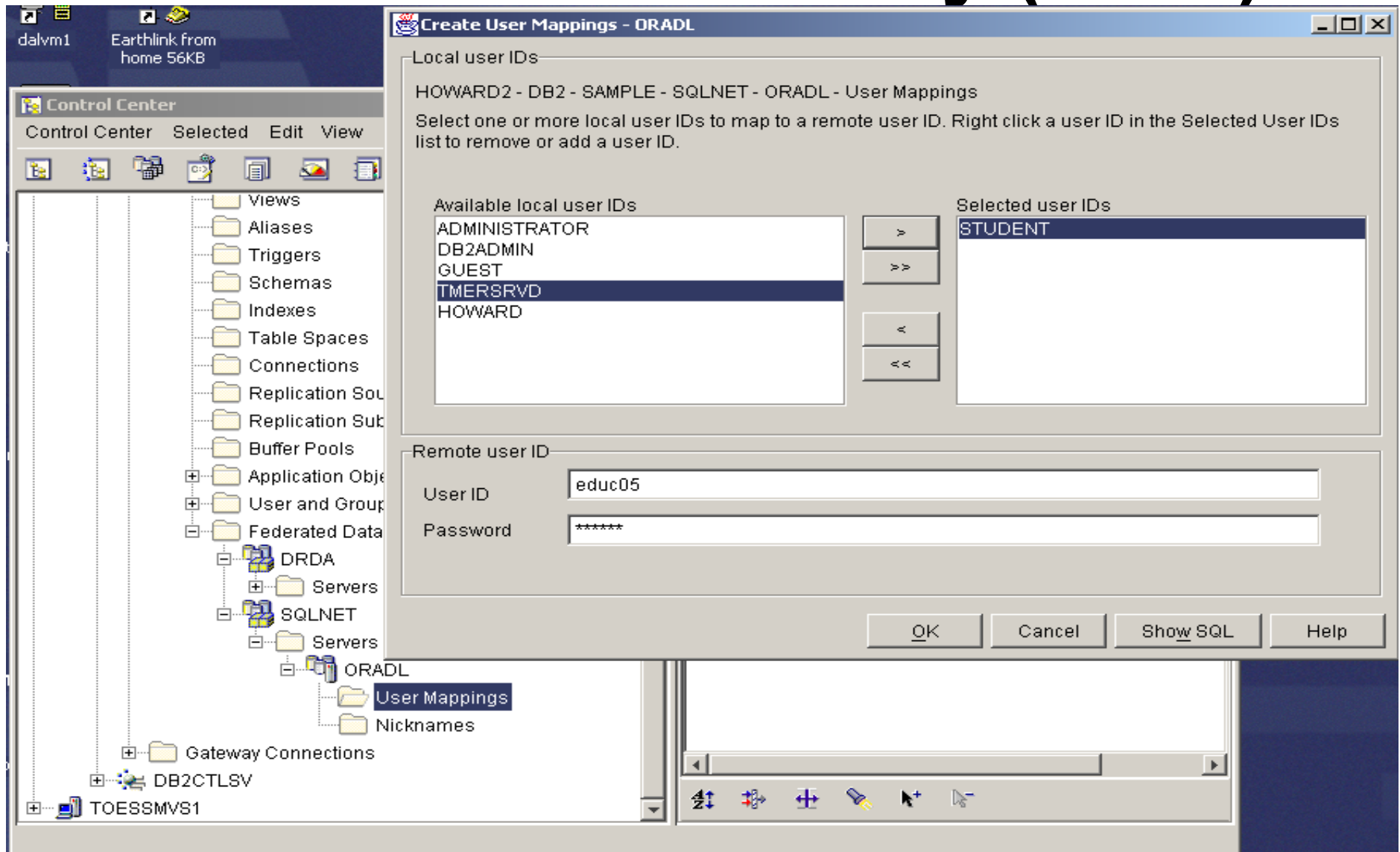
Multi-Server Security



Multi-Server Security (Cont)

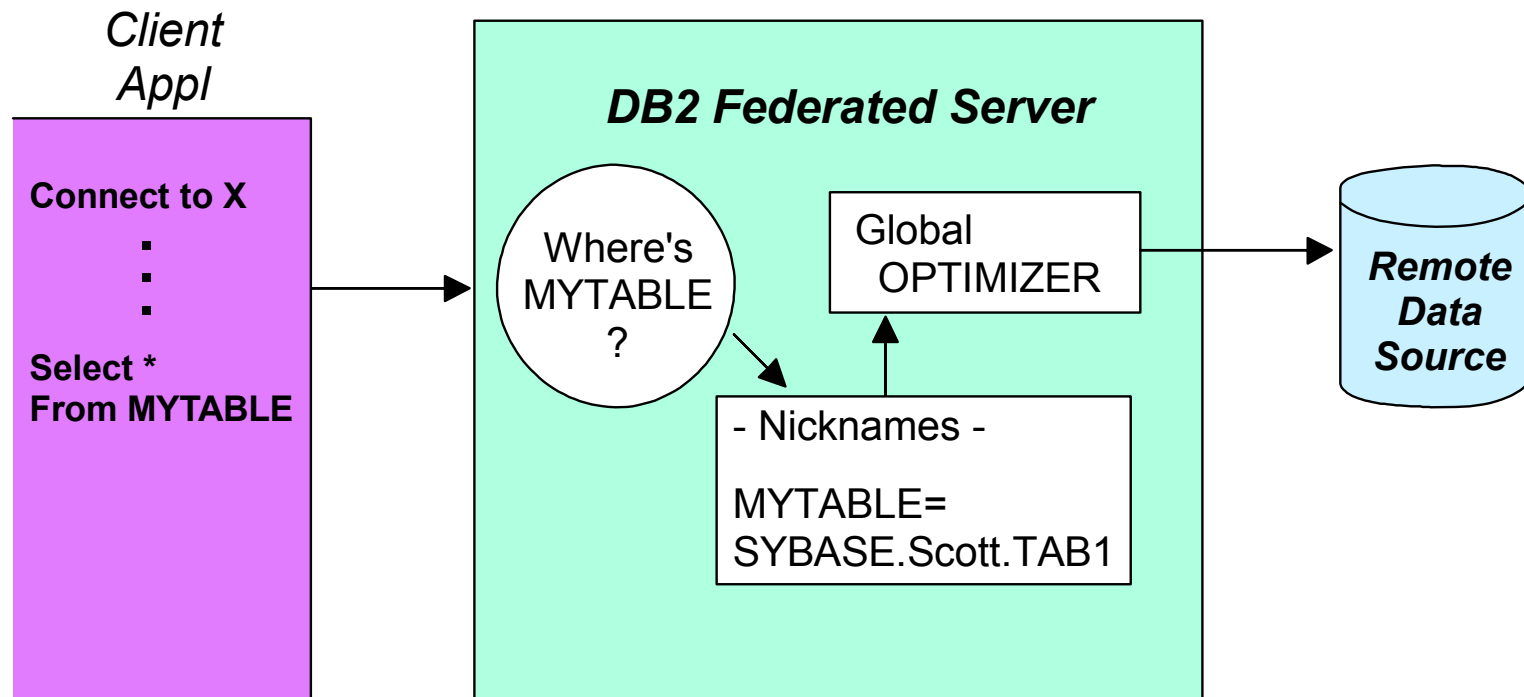


Multi-Server Security (Cont)



```
c:\ > db2 "CREATE USER MAPPING FOR scott
SERVER MSSQL OPTIONS (REMOTE_AUTHID 'scott' REMOTE_PASSWORD 'scottspw')"
```

Nicknames



CREATE NICKNAME

▶▶ CREATE NICKNAME — *nickname* — FOR — *remote-tablename* —▶

▶▶ DROP NICKNAME — *nickname* —▶

▶▶ ALTER NICKNAME — *nickname* — SET COLUMN — *old-local-column-name* —▶

▶ LOCAL NAME — *new-local name* —▶
▶ LOCAL TYPE — *data-type* —▶
▶ REMOTE TYPE — *remote data-type* —▶

Examples:

```
CREATE NICKNAME MYTABLE FOR "SYBASE.scott.TAB1"
```

```
DROP NICKNAME MYTABLE
```

```
ALTER NICKNAME DEPT SET COLUMN address LOCAL NAME daddress
```

CREATE NICKNAME Process

User Scott enters:

```
CREATE NICKNAME dept FOR oracle1.scotto.dept
```

SYSCAT.TABLES

TABNAME	DEFINER	TYPE	
DEPT	SCOTT	N	

	CARD	NPAGES	FPAGES	
	?	?	?	

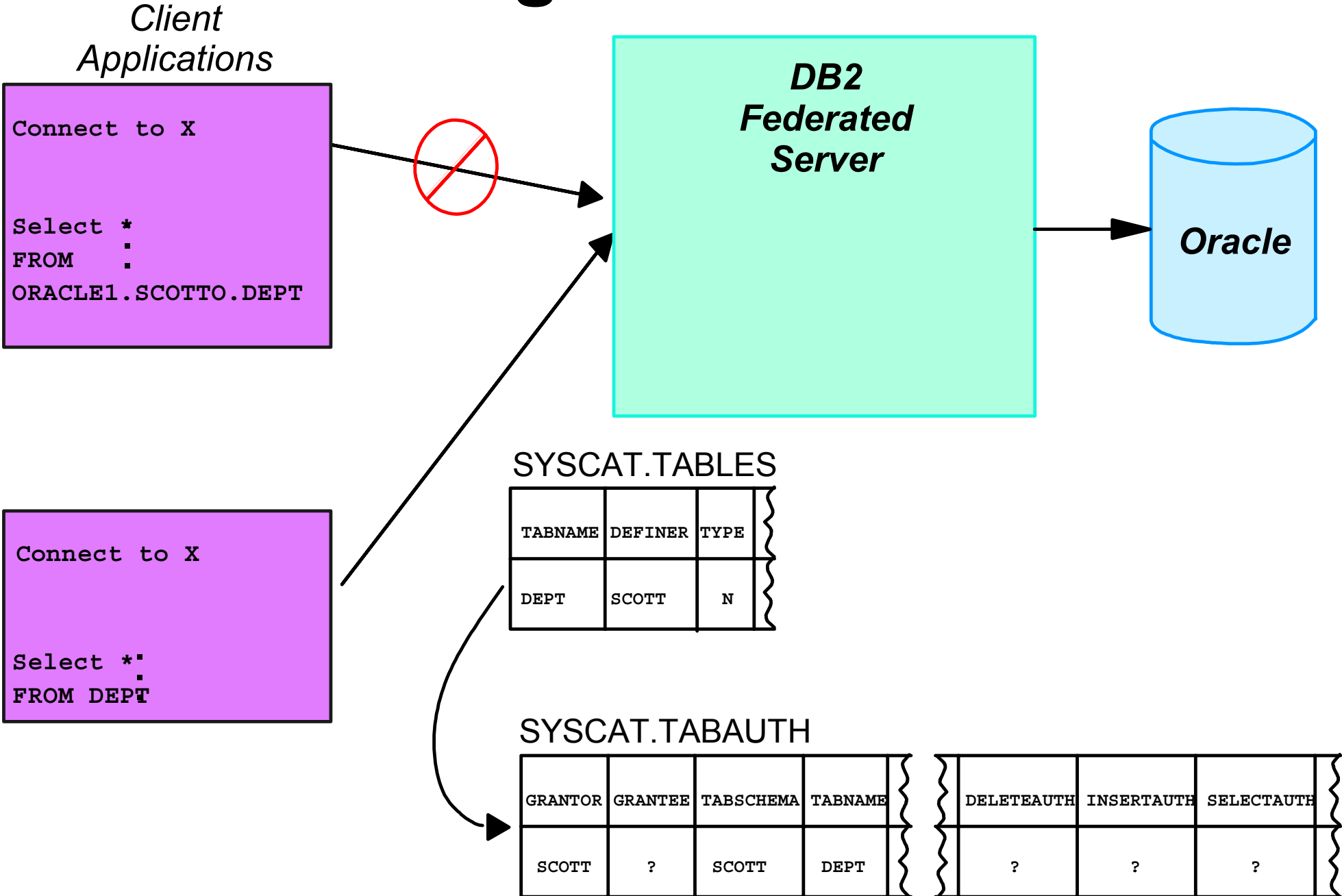
*We need to go to oracle1 and
gather statistics!*

DB2 UDB V7 Create Nickname Process ...

SYSCAT.COLOPTIONS

TABSCHEMA	TABNAME	COLNAME	OPTION	SETTING
SCOTT	DEPT	NAME	REMOTE_COLUMN_TYPE	VARCHAR
SCOTT	DEPT	NAME	REMOTE_COLUMN_LENGTH	25
SCOTT	DEPT	ADDRESS	REMOTE_COLUMN_TYPE	VARCHAR
SCOTT	DEPT	ADDRESS	REMOTE_COLUMN_LENGTH	50
SCOTT	DEPT	BONUS	REMOTE_COLUMN_TYPE	NUMBER
SCOTT	DEPT	BONUS	REMOTE_COLUMN_LENGTH	38
SCOTT	DEPT	BONUS	REMOTE_COLUMN_SCALE	0

Using Nicknames



"Unified Schema"

Multi-Location Views

*DB2 for
OS/390*

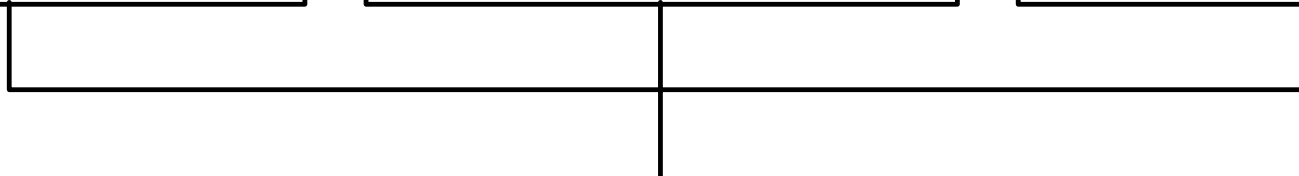
Sybase

Oracle

CUST SFBR		
ACCT NO	NAME	...
101	Ajax	...
303	Super	...
099	Acme	...
.

SJBR		
ACCT NO	Balance	...
101	250,000	...
303	100,000	...
099
.

ACCT NO	Crllimit	...
101	319,000	...
303	300,000	...
099

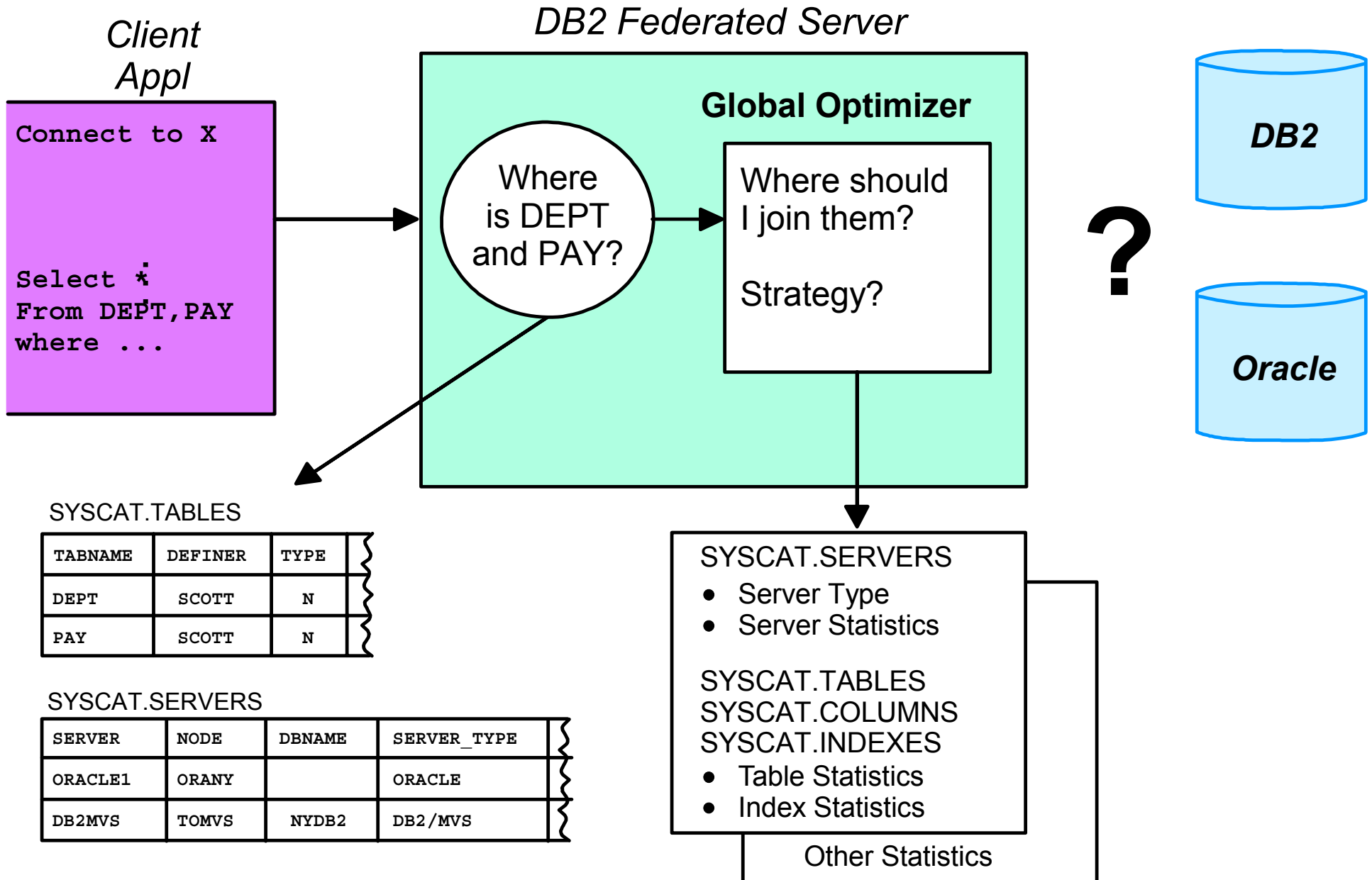


*DB2 Federated
Technology*

```
Select *  
From CUSTVIEW
```



Global Optimization



Updating Global Statistics

DB2 Federated Server

DB2 Data Source

Global Catalogs

SYSCAT.INDEXES

INDNAME	INDSCHEMA	TABNAME	TABSCHEMA	COLNAME
DEPTIX	RUSS	EMPLOYEE	RUSS	NAME

SYSCAT.INDEXES

INDNAME	INDSCHEMA	TABNAME	TABSCHEMA	COLNAME
DEPTIX	RUSS	EMPLOYEE	RUSS	NAME
EMPNOIX	RUSS	EMPLOYEE	RUSS	EMPNO

```
CREATE INDEX RUSS.EMPNOIX  
ON EMPLOYEE (EMPNO ASC);
```

How can DB2 Federated Technology discover this new remote index?

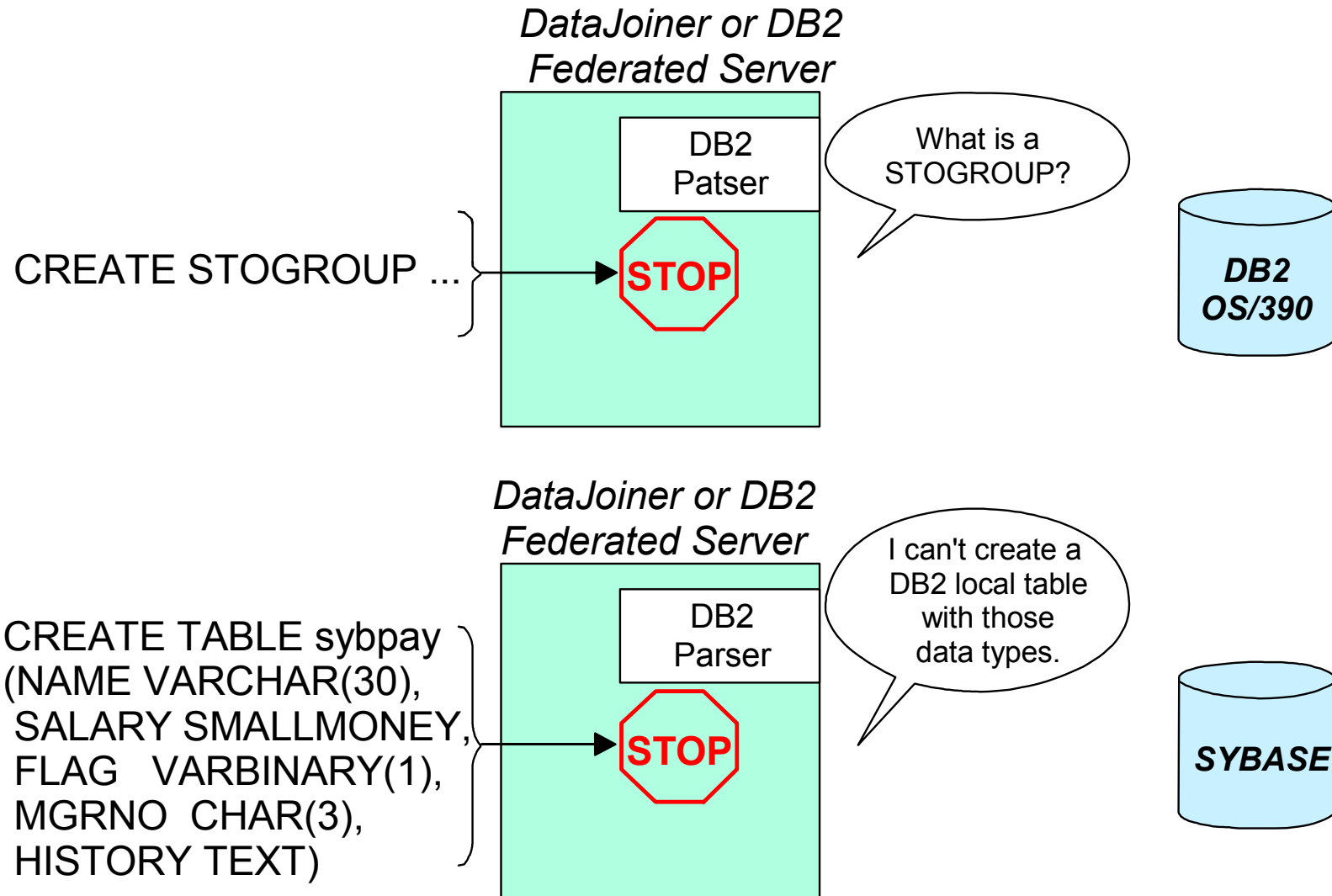
CREATE INDEX Specification

```
CREATE INDEX index-name ON nickname  
(column_name specification) SPECIFICATION ONLY
```

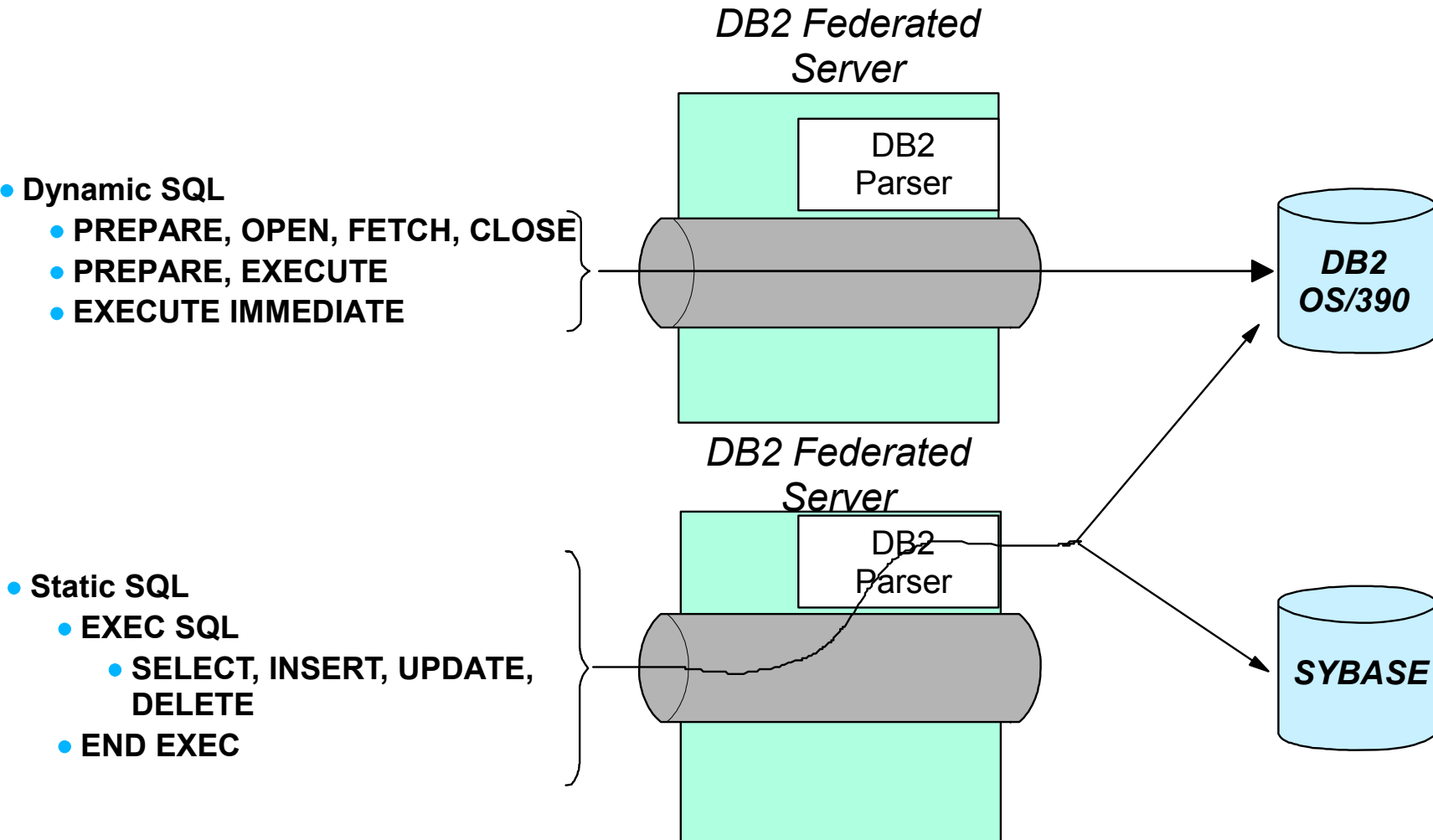
RUNSTATS Considerations

- **RUNSTATS will not run against nicknames**
 - Extrapolate statistics and modify the DB2 Federated Technology catalogs
 - Use GETSTATS tool

The Problem



Directed SQL



Pass-thru SQL

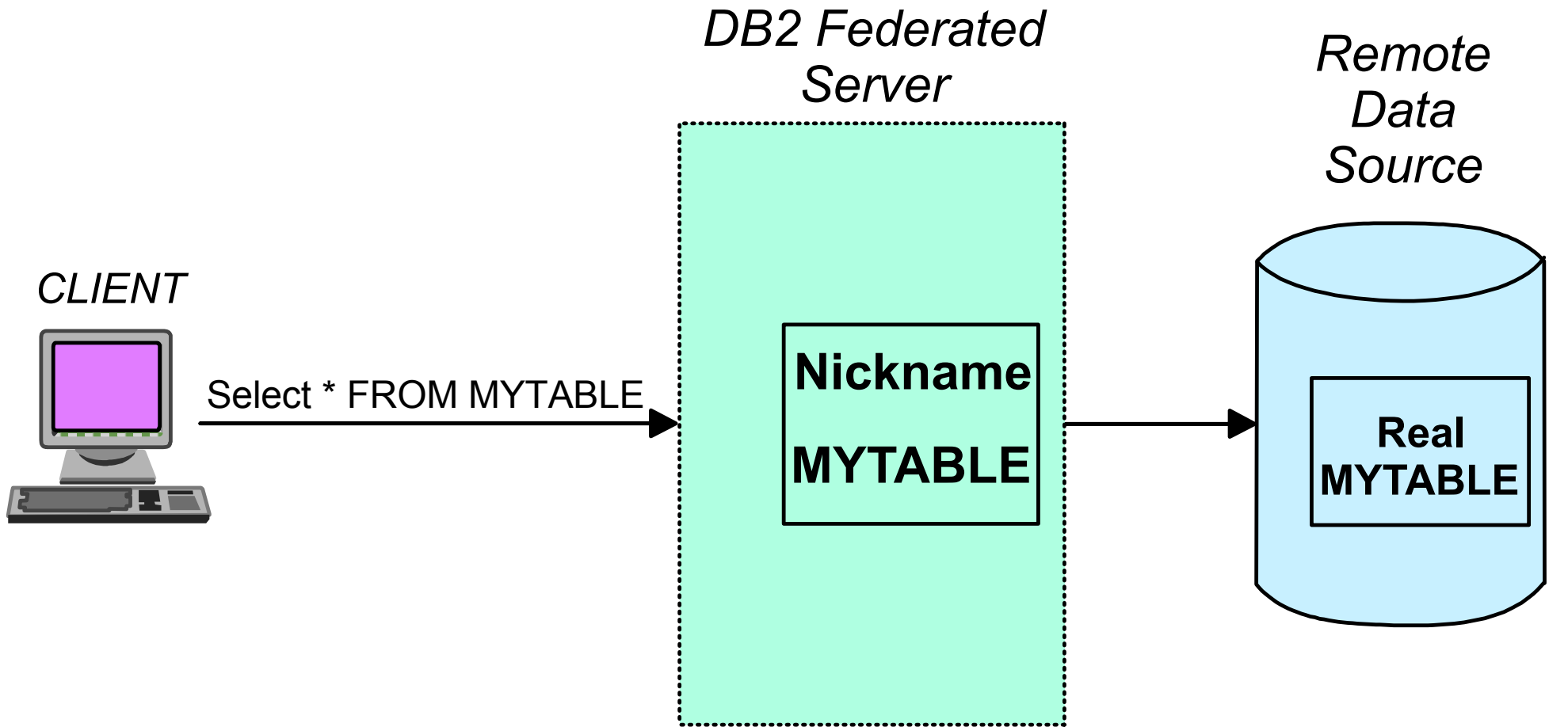
```
SET PASSTHRU sybase1
```

```
CREATE TABLE sybpay  
(NAME VARCHAR(30),  
SALARY SMALLMONEY,  
FLAG VARBINARY(1),  
MGRNO CHAR(3),  
HISTORY TEXT)
```

```
SET PASSTHRU RESET  
PASSTHRU authority can be granted or revoked
```

Note that remote tables can also be created directly via DDL

Remote Table Access



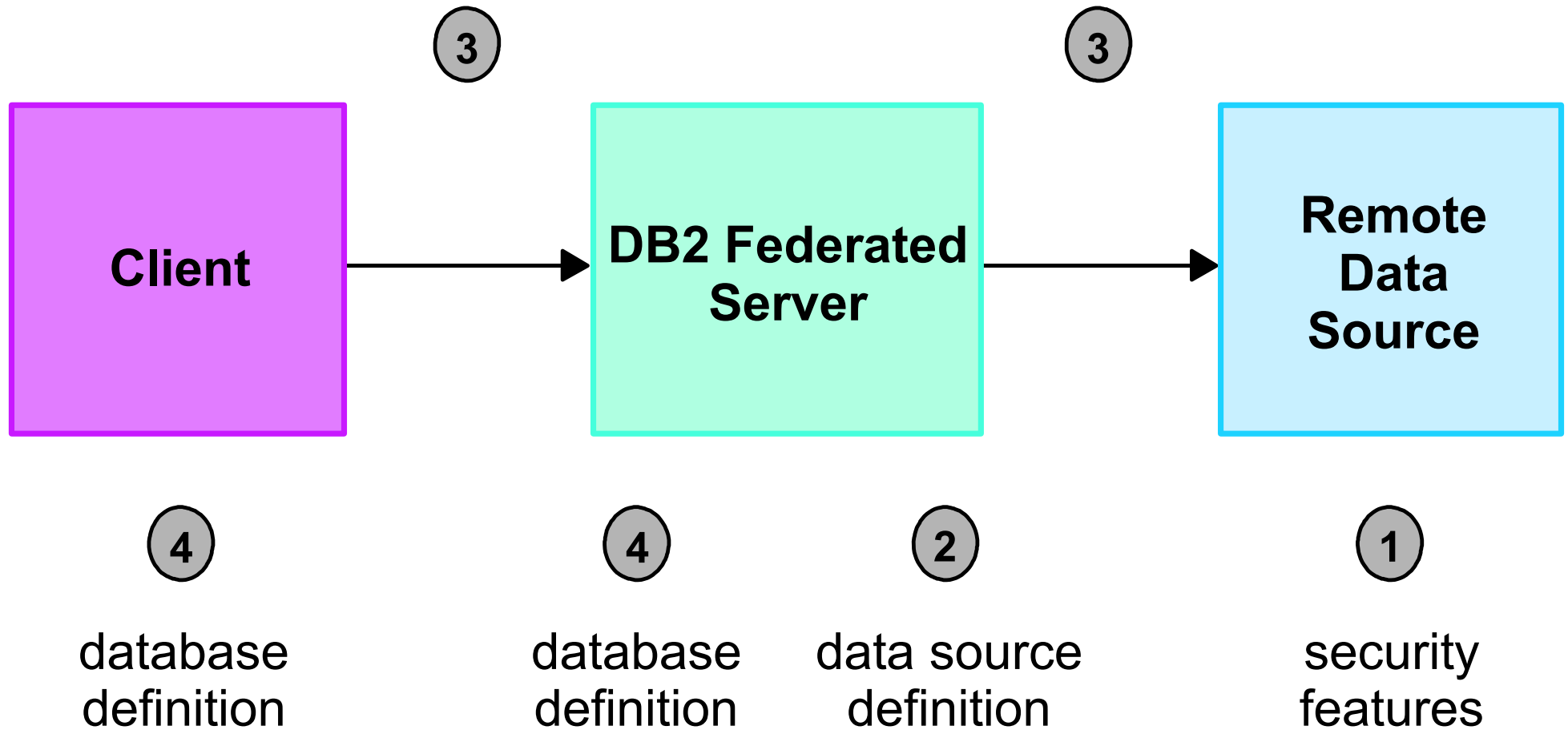
Requires:

Select privilege
on the nickname
MYTABLE

+

Select
privilege on
the real table

The Four Security Knobs





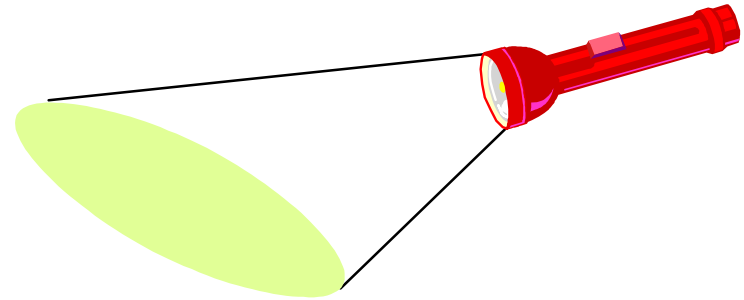
Tuning the Federated Environment



Tuning Diagnostics

- **Isolate the problem:**

- DB2 Federated Server?
- Remote database server?
- Network?
- Clients?



- **If performance related, read the "Federated Database Query Compiler Phases" in *Administration Guide: Performance***

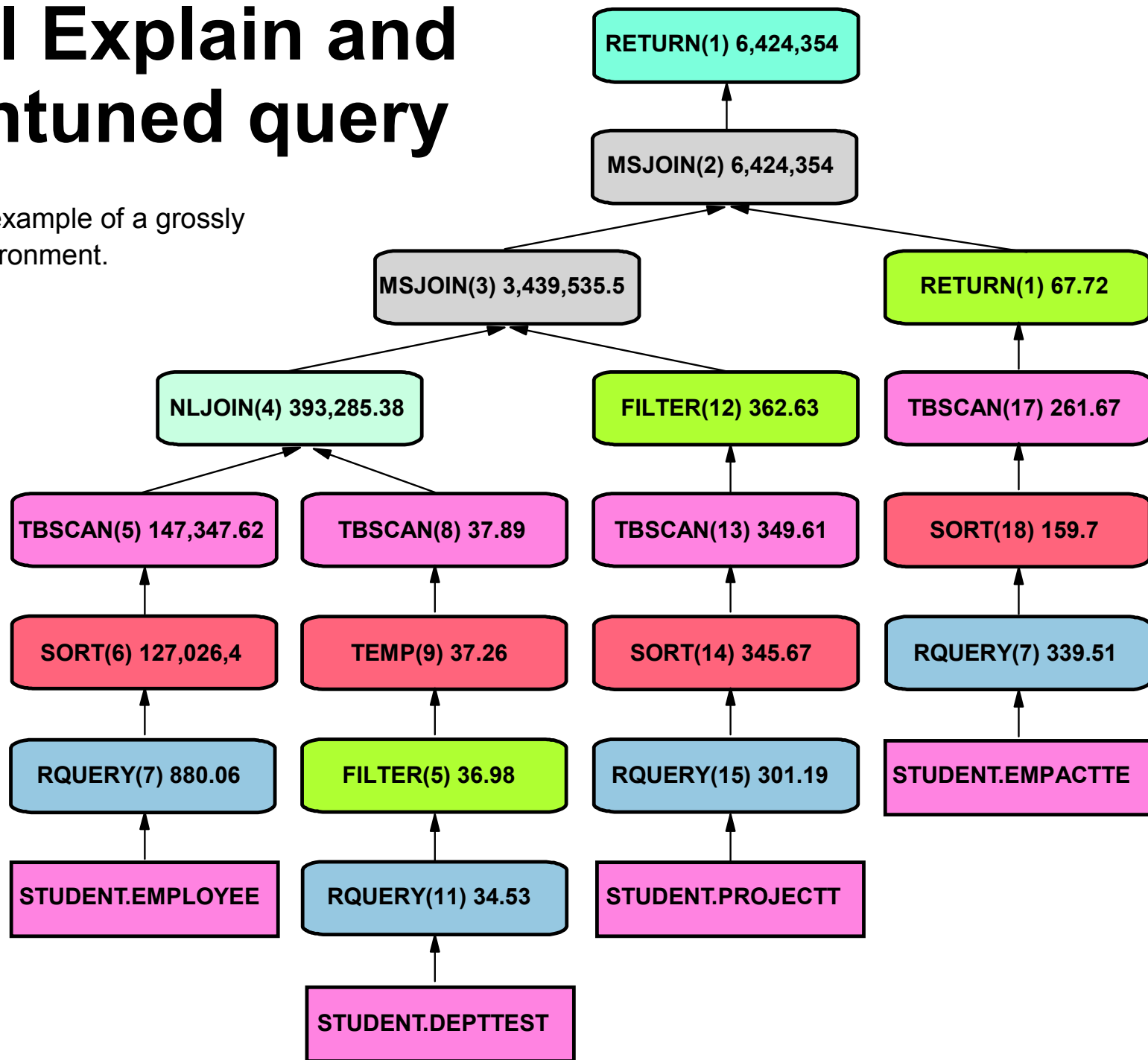
- Use Visual Explain
- Check sorts
- Check pushdown of SQL and functions
- Reconsider query if compensation forces poor path

- **Issue SQL remotely against data source:**

- From the DB2 Federated Server
- At data server

Visual Explain and the untuned query

Hypothetical example of a grossly mistuned environment.

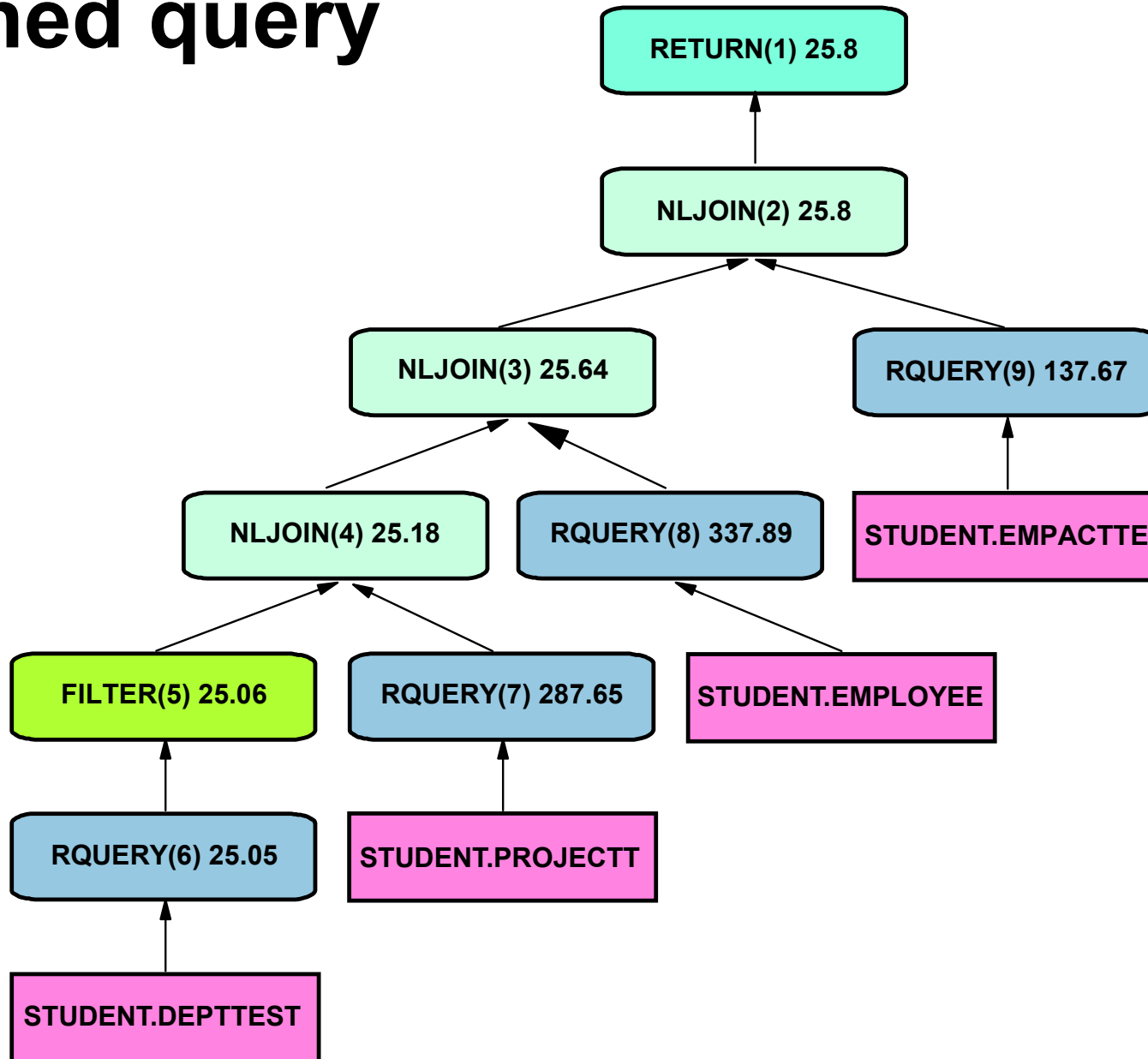


Pushdown Analysis

DB2 should Pushdown operations close to data

- Analyze query to determine query blocks that can be executed at the remote data source
 - Predicate pushdown (In, Exists, RelOp, Between)
 - Operation pushdown (Group, Order, Union, and so on)
 - Head expression pushdown
 - Function pushdown
- Significantly affects performance
- Several factors affect the ability to pushdown
 - Server characteristics
 - Nickname characteristics
 - Query characteristics

Visual Explain and the tuned query



Where can I get more training?

Consider IBM Learning Services Course:

CF471, [DB2 Federated Database: Integrating Diverse Data](#)

DW470, [DiscoveryLink: Integrating Diverse Life Sciences Data](#)

For more information or to enroll:

<http://ibm.com/services/learning/us/>

IBM Employees should use Global Campus and search for course code CF471

