

*MASSACHUSETTS INSTITUTE OF TECHNOLOGY
SLOAN SCHOOL OF MANAGEMENT*

15.565 Integración de sistemas de información:

Factores tecnológicos, organizativos y estratégicos

15.578 Sistemas de información globales:

Comunicaciones y conectividad en sistemas de información

Primavera 2002

Clase 13

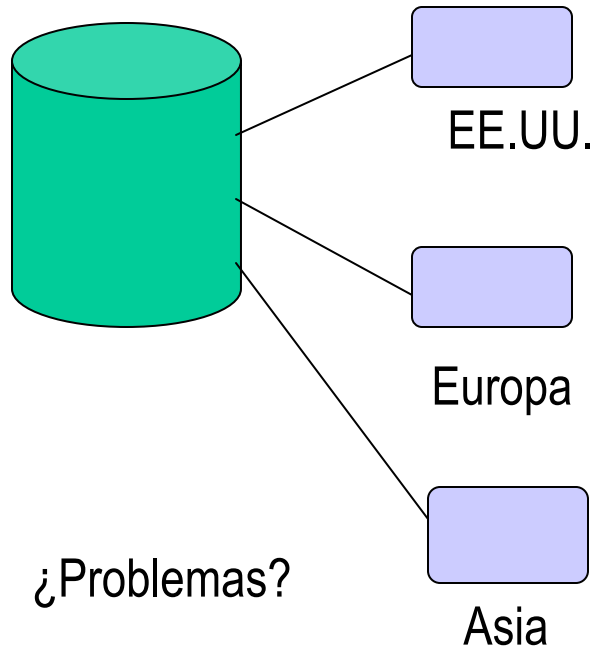
***BASES DE DATOS DISTRIBUIDAS
HOMOGÉNEAS***



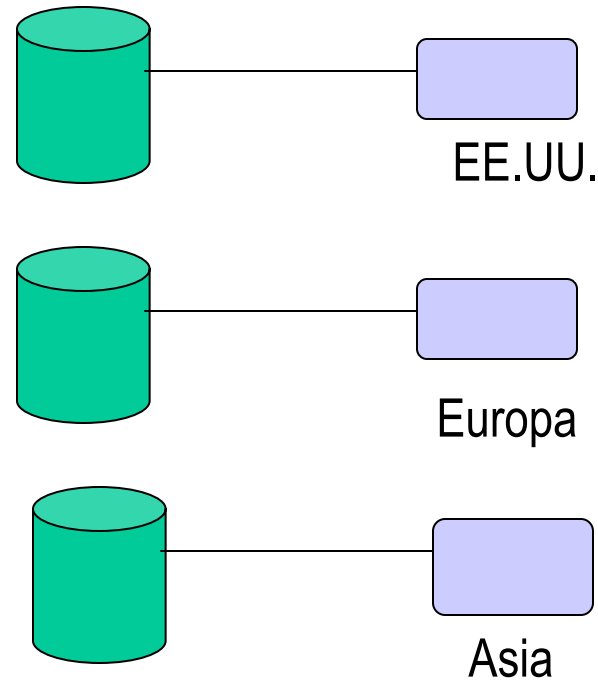
¿CENTRALIZAR O DESCENTRALIZAR?

Reto de diseño común -- Considere bases de datos de balance de cuentas para un banco internacional o una operación de cambio. ¿Qué haría?

Base de datos centralizada



Bases de datos descentralizadas



SISTEMA DE BOLSA GLOBAL DE NUEVA GENERACIÓN (FXX) -- REVISIÓN

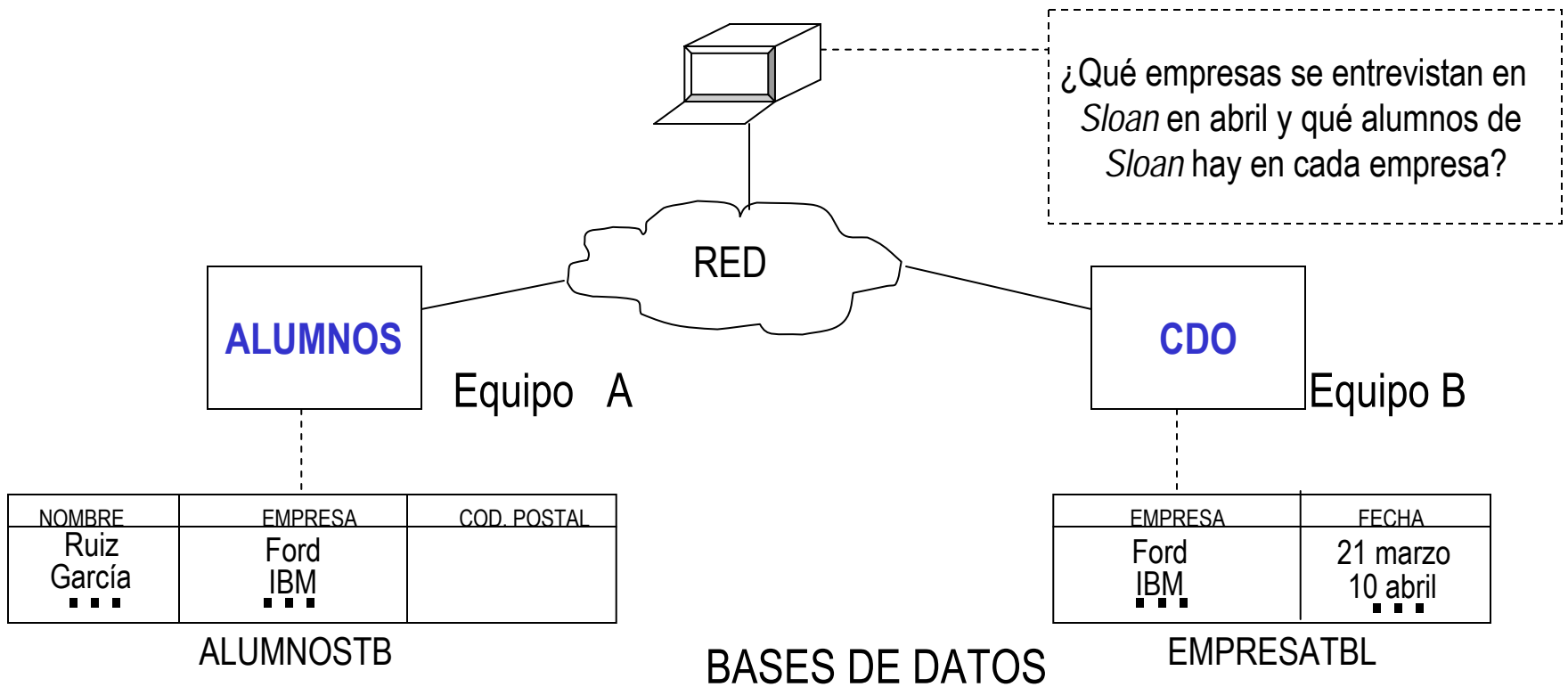
Base de datos global

La *base de datos global* contiene una *copia* de la información de cada uno de los servidores regionales y locales, basados en la zona de horario local y en GMT. La base de datos global se actualiza constantemente por filiales y regiones en todo el mundo, de forma asíncrona y en tiempo real. El retraso nominal atribuido al tiempo de transmisión no suele superar el periodo de un minuto. Todas las ciudades de nivel 1 se enlazan mediante RDSI para administrar el exceso de tráfico de la red y los posibles errores de red.

Los datos de la base de datos global se registran en horario GMT y en el horario local de la filial desde la que se reciben los datos. El registro de los datos en horario GMT favorece la coherencia de consultas de varias zonas de horarios. El registro de la hora de los datos en horario local favorece las consultas locales de cada sitio.

La base de datos global incluirá todos los tratos de regiones y filiales, operaciones al final de la jornada, cifras diarias de riesgo, incluidas las de tipo *Money-at-Risk*, información de socios y de crédito. Los gestores de negocio globales y los gestores senior pueden utilizar la base de datos global para administrar, supervisar y analizar los valores, el riesgo, la actividad de los socios, los límites de crédito, los márgenes de corretaje y el gestor de pedidos FX global. También puede utilizarse cualquier herramienta de *sistemas de información ejecutiva* (“EIS”) de terceros para comparar datos con la base de datos global.

B. DE DATOS DISTRIBUIDA -- APARECE COMO B. DE DATOS CENTRALIZADA

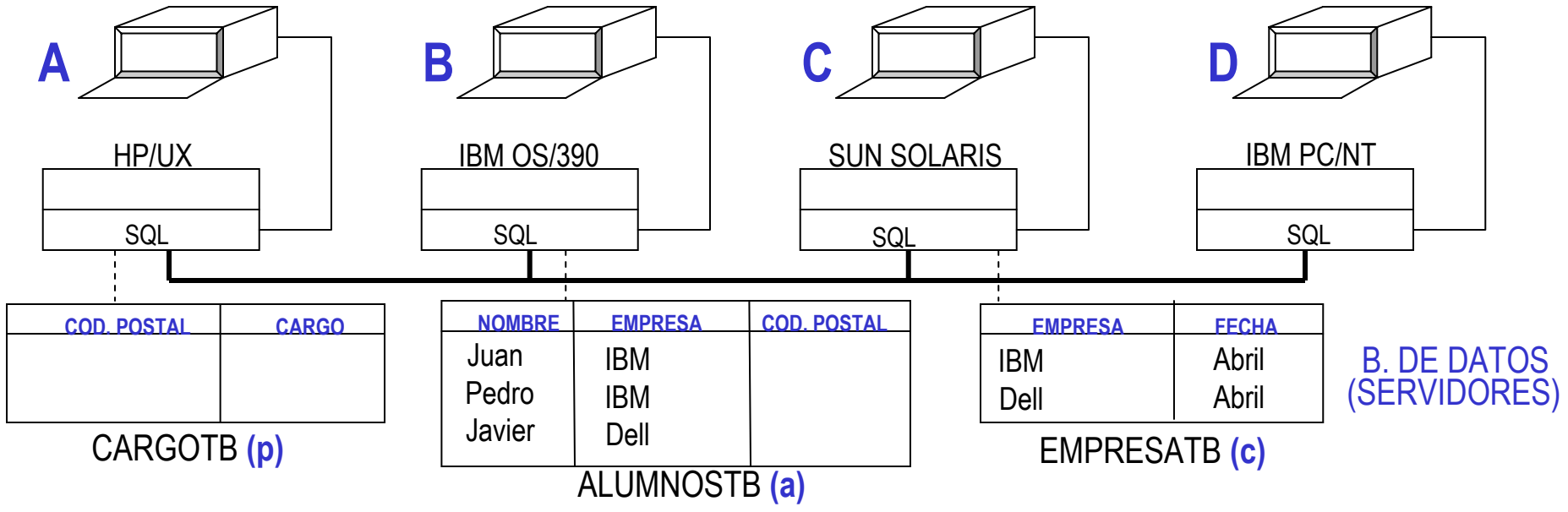


PROBLEMAS:

1. ¿QUÉ CONOCIMIENTOS SE NECESITAN?
2. ¿QUÉ DATOS SE OBTIENEN? ¿DE DÓNDE?
3. ¿CÓMO SE RELACIONAN LOS DATOS?
4. ¿DÓNDE SE COPIAN?
5. ¿CÓMO SE HACE ESTO DE FORMA **TRANSPARENTE** PARA EL USUARIO Y EL DESARROLLADOR?

SISTEMAS DE BASES DE DATOS DISTRIBUIDAS HOMOGÉNEAS

(P.EJ., ORACLE/STAR y INGRES DISTRIBUIDAS -- MISMAS DBMS EN TODOS LOS SISTEMAS)

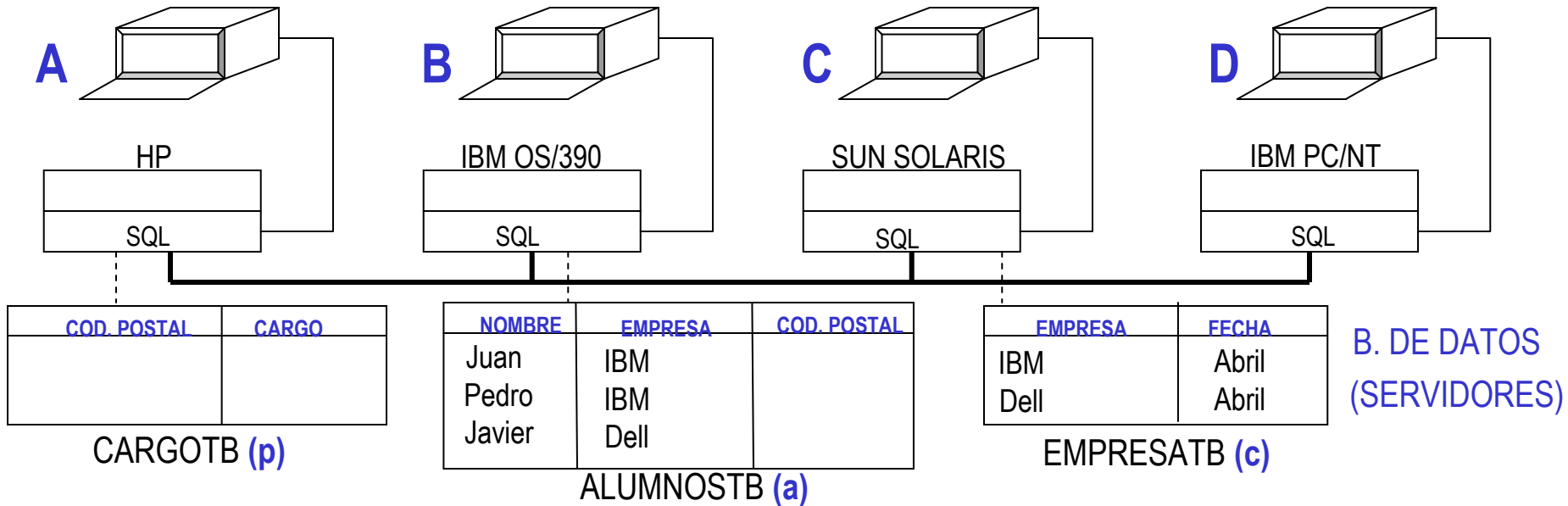


OBJETIVOS DE TRANSPARENCIA:

- 1. RECUPERACIÓN** -- MISMOS RESULTADOS EN CUALQUIER SITIO
- 2. ACTUALIZACIÓN** -- ACTUALIZAR DESDE CUALQUIER SITIO A UNA ÚNICA TABLA

Seleccione un nombre, fecha y empresa, de la "a" la "c" donde fecha = 'abril' y a.empresa = c.empresa

SISTEMAS DE BASES DE DATOS DISTRIBUIDAS HOMOGÉNEAS



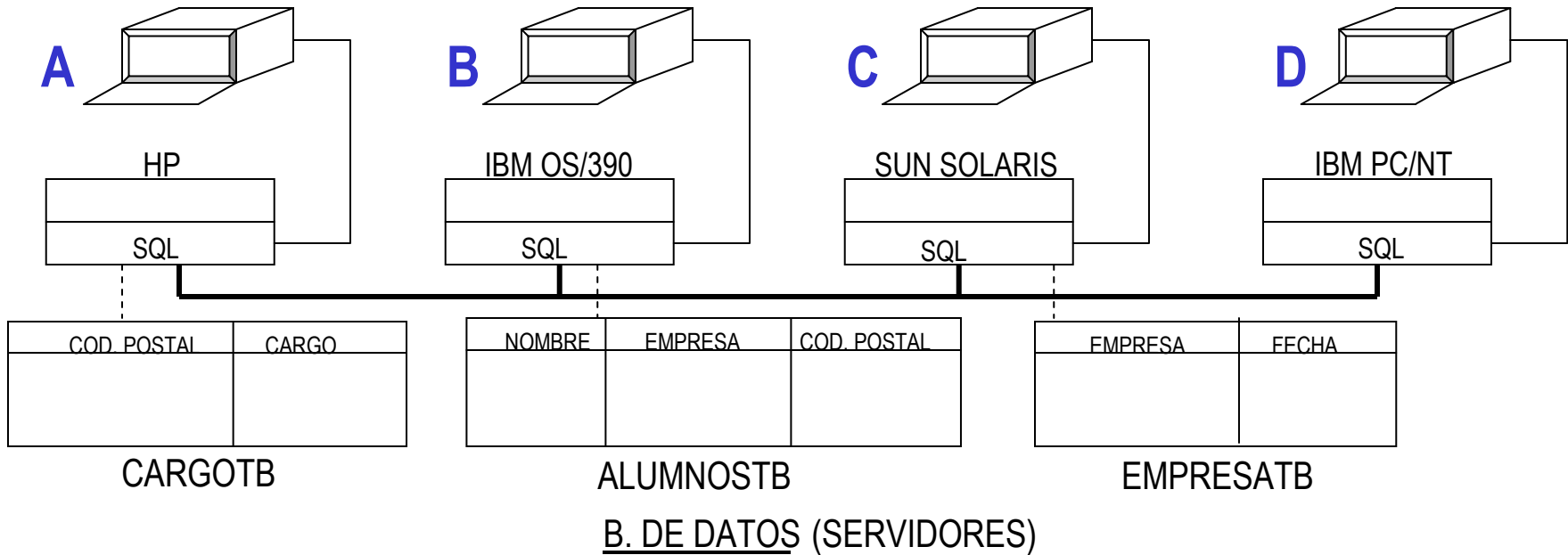
OBJETIVOS DE TRANSPARENCIA:

3. **ESQUEMA** – CAMBIO DE ESQUEMA VISIBLE EN CUALQUIER SITIO (AUTOMÁTICAMENTE)

4. **OPTIMIZACIÓN DEL RENDIMIENTO** – (MEJOR) VELOCIDAD IGUAL POSIBLE DESDE CUALQUIER SITIO

Seleccione un nombre, fecha y empresa, de la "a" a la "c" donde fecha = 'abril' y a.empresa = c.empresa

SISTEMAS DE BASES DE DATOS DISTRIBUIDAS HOMOGÉNEAS



OBJETIVOS DE TRANSPARENCIA:

5. TRANSACCIÓN -- ACTUALIZACIÓN MÚLTIPLE CORRECTA (CONSULTAR SIG. DIAP.)

6. COPIA -- MANTENIMIENTO Y UTILIZACIÓN EFICAZ DE COPIAS REPETIDAS

DIFICULTAD DE ACTUALIZACIÓN Y TRANSPARENCIA DE COPIAS

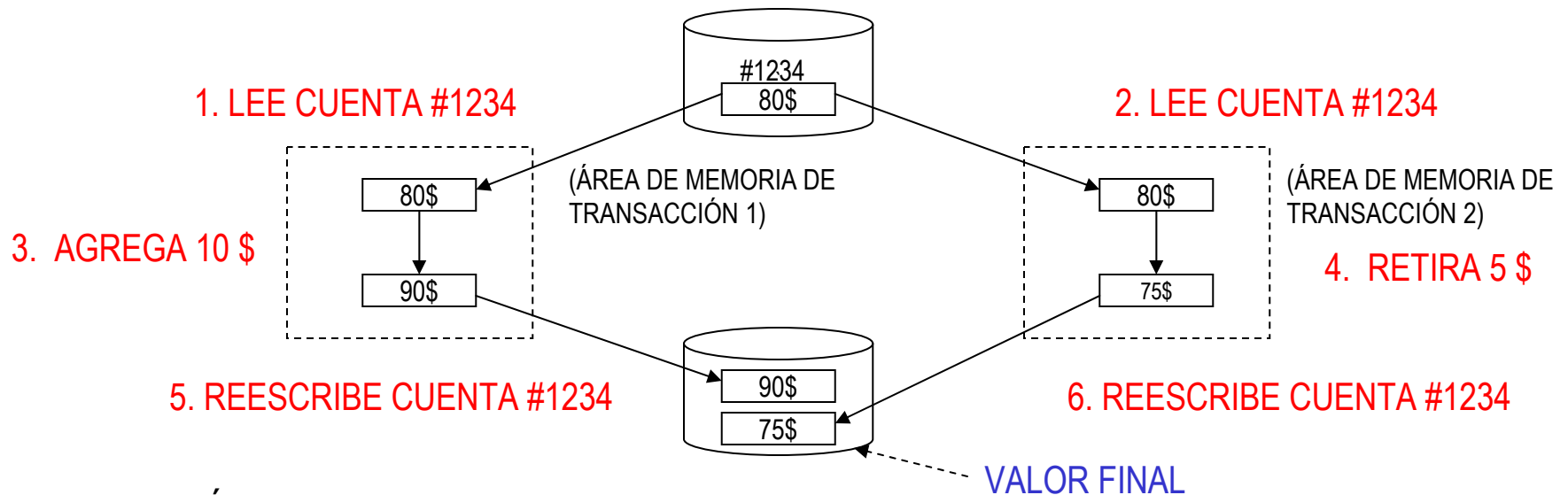
- “CONTROL DE CONCURRENCIA” EN LA B. DE DATOS CENTRAL -- EJEMPLO:

TRANSACCIÓN 1

AGREGA 10 \$ A LA CUENTA #1234

TRANSACCIÓN 2

RETIRA 5 \$ DE LA CUENTA #1234



-- SOLUCIÓN AL PROBLEMA (CENTRALIZADO)

- COORDINAR LA TRANSACCIÓN MEDIANTE BLOQUEOS
- SE NECESITAN VARIOS BLOQUEOS DURANTE LA TRANSACCIÓN PARA COORDINAR TODOS LOS ELEMENTOS DE DATOS NECESARIOS (LECTURA Y ESCRITURA)
- PROBABILIDAD DE EXISTENCIA DE INTERBLOQUEO

PROBLEMAS ADICIONALES DE CONTROL DE CONCURRENCIA

- EN B. DE DATOS DISTRIBUIDAS CON ACTUALIZACIÓN GENERAL Y TRANSPARENCIA DE COPIA
 - COMPLEJIDAD PARA COORDINAR **BLOQUEOS DISTRIBUIDOS**
 - **SOBRECARGA EN COMUNICACIÓN** DE AJUSTE DE BLOQUEO (NO SE SABE SI REALMENTE EL BLOQUEO ES NECESARIO)
 - DIFICULTAD PARA LA **DETECCIÓN DE INTERBLOQUEOS**
 - AGREGACIÓN DE COPIAS REPLICADAS:
 - ACTUALIZACIÓN DE SOBRECARGA (SE DEBEN ACTUALIZAR TODAS LAS COPIAS)
 - BLOQUEO DE SOBRECARGA (SE DEBEN BLOQUEAR TODAS LAS COPIAS)
 - COMPLEJIDAD DE RENDIMIENTO (MUCHAS MÁS SECUENCIAS)
 - PELIGRO DE ERROR EN EL SITIO
 - ¿QUIÉN RESTABLECE EL BLOQUEO SI OCURRE UN ERROR?
 - PARTICIÓN Y REINTEGRACIÓN DE LA RED
- CIENTOS DE ARTÍCULOS DE INVESTIGACIÓN SOBRE ESTE TEMA

PROPIEDADES DESEABLES DE INDEPENDENCIA

1. **ERROR**

- SOLO AFECTA AL SITIO "CON ERRORES"

2. **RECUPERACIÓN**

- SE RECUPERA AUTOMÁTICAMENTE

3. **RED**

- FUNCIONA EN TODAS LAS REDES

4. **HARDWARE/SO**

- SE EJECUTA EN VARIOS EQUIPOS

5. **PROVEEDOR SQL**

- PERMITE VARIOS PROVEEDORES SQL (P.EJ., DB2, ORACLE)

6. **DBMS**

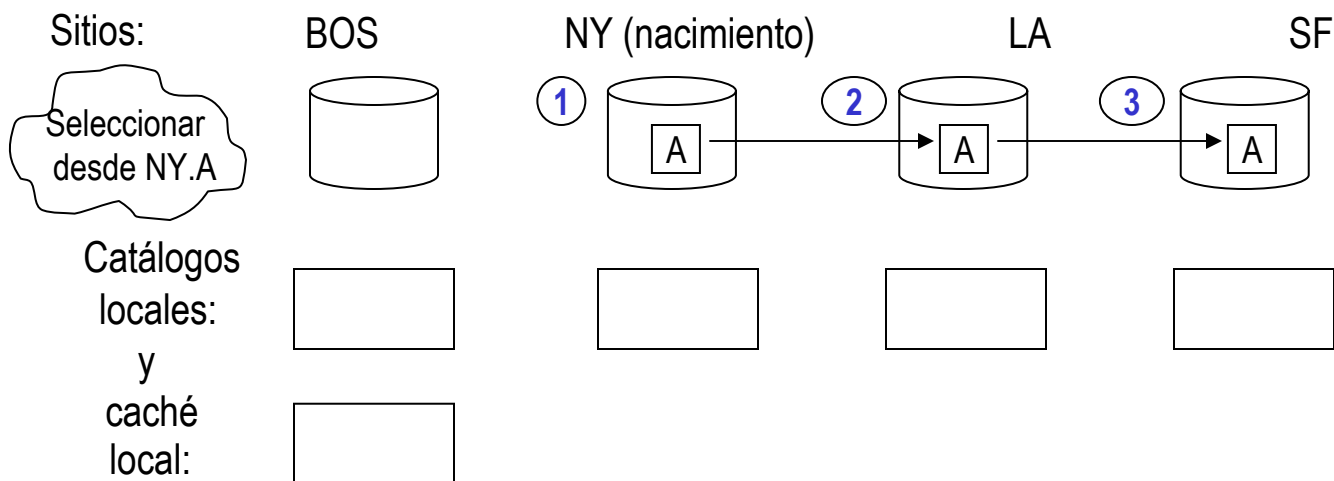
- ADMITE OTROS MODELOS DE DBMS (JERÁRQUICAS, REDES), ASÍ COMO OTROS LENGUAJES

TÉCNICAS Y PROBLEMAS DE IBM R*

- GESTIÓN DISTRIBUIDA DEL CATÁLOGO
- ACCESO REMOTO AL CATÁLOGO
- DEFINICIÓN DISTRIBUIDA DE DATOS
- **MIGRACIÓN DE TABLAS**
- PLANIFICACIÓN DISTRIBUIDA DE CONSULTAS
- DISTRIBUCIÓN DE CONSULTAS
- PROTOCOLOS DISTRIBUIDOS DE COMPROMISOS
- DETECCIÓN DISTRIBUIDA DE INTERBLOQUEOS

GESTIÓN DEL CATÁLOGO R* Y ALTERNATIVAS

- CATÁLOGO CENTRAL (INGRES)
- CATÁLOGOS REPLICADOS (ORACLE)
- CATÁLOGOS LOCALES (IMB R*)
 - IDENTIFICAR SITIO "DE NACIMIENTO" A PETICIÓN
 - "ALMACENAR EN LA CACHÉ" LA INFORMACIÓN PARA REDUCIR EL ACCESO AL CATÁLOGO LOCAL
 - NÚMERO DE VERSIÓN DEL CATÁLOGO UTILIZADA PARA VALIDAR CORRECCIÓN



ALGUNOS PROBLEMAS DE DIFICULTAD AÑADIDA

- **BASES DE DATOS CON PARTICIÓN VERTICAL**

BD1:

Nombre	Dirección

BD2:

Nombre	Teléfono

- **BASES DE DATOS CON PARTICIÓN HORIZONTAL**

BD
ESTE:

Nombre	Dirección

BD
OESTE:

Nombre	Dirección

RESUMEN

- LOS SISTEMAS GESTIÓN DE B. DE DATOS DISTRIBUIDAS HOMOGÉNEAS SON UNA REALIDAD
- AÚN ESTÁN LIMITADOS POR LA EXPERIENCIA PRÁCTICA
- TAMBIÉN NECESITAN SEGUIR AFRONTANDO DISTINTOS ENTORNOS Y VARIACIONES DURANTE UN TIEMPO (DBMS HETEROGÉNEAS)