Machine Translated by Google

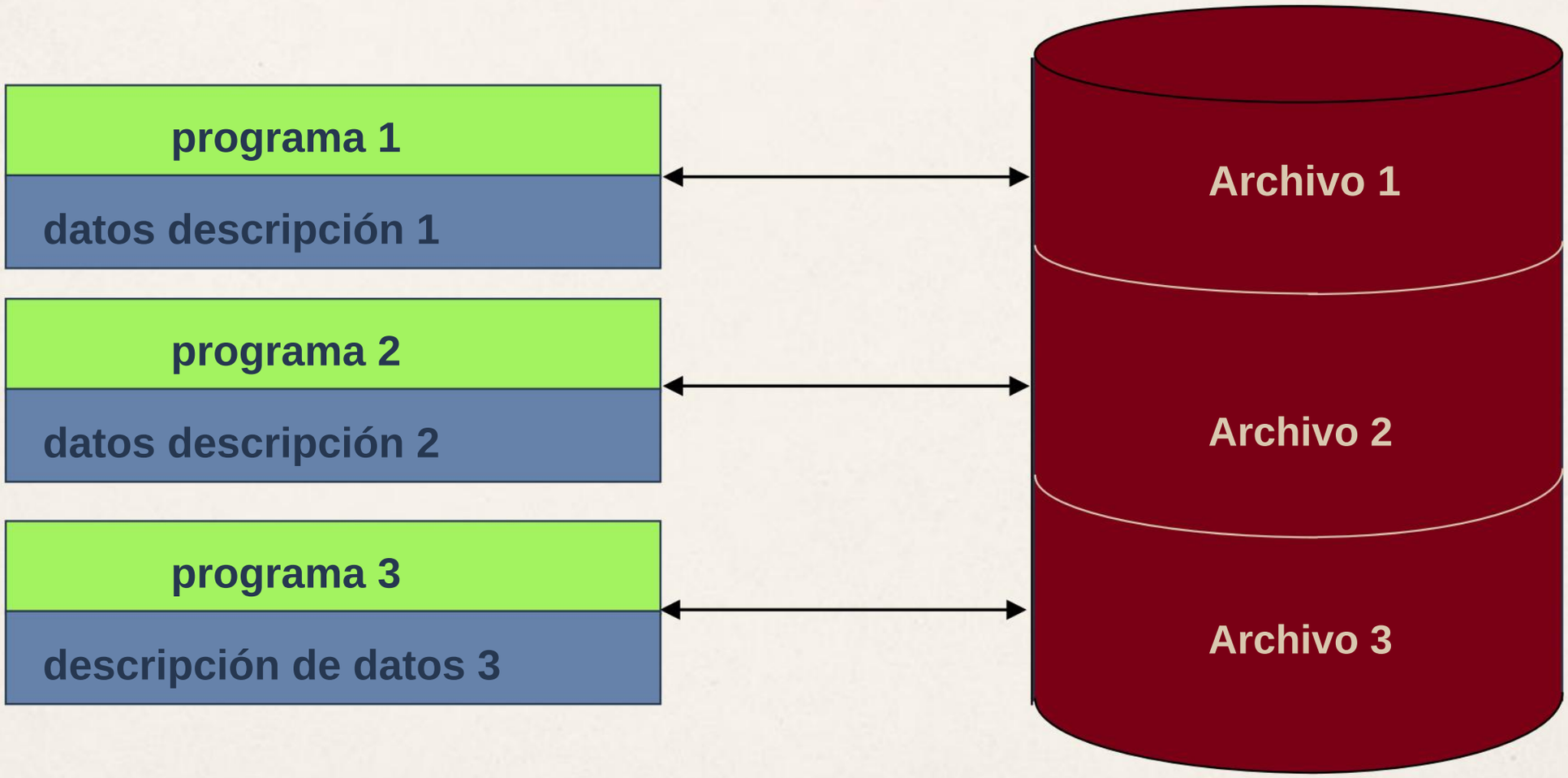
Esquema

- Introducción
  - ÿ ¿Qué es un DBMS distribuido?
  - ÿ Arquitectura DBMS distribuida
- Antecedentes
- Diseño de bases de datos distribuidas
- Integración de bases de datos •Control de datos semánticos
- Procesamiento de consultas distribuidas
- Procesamiento de consultas de bases de datos múltiples •Administración de transacciones distribuidas •Replicación de datos •Sistemas de bases de datos paralelos •SGBD de objetos distribuidos •Administración de datos punto a punto
- Administración de datos web •Problemas actuales

© MT Özsu & P. Valduriez

Machine Translated by Google

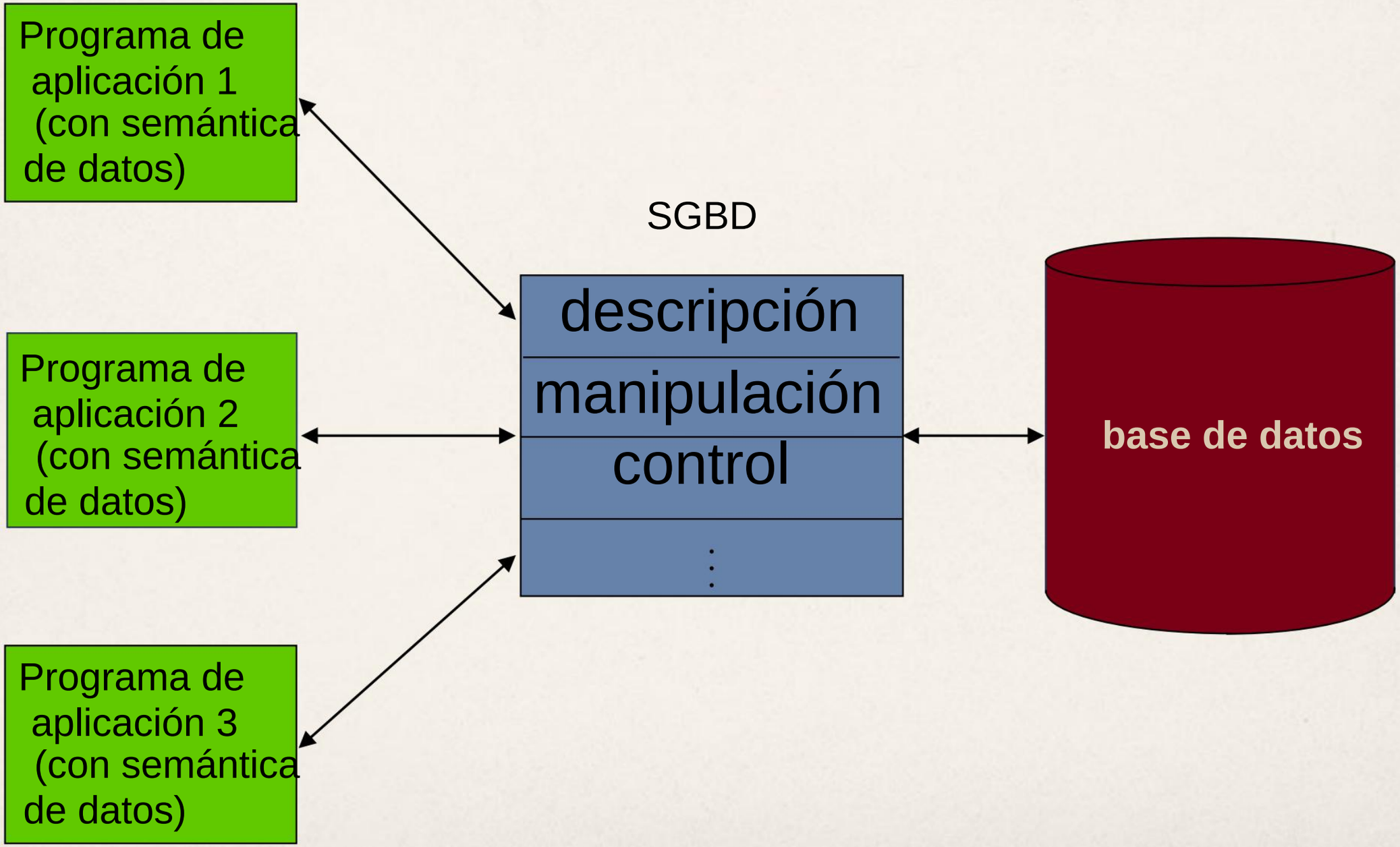
Sistemas de archivos



© MT Özsu & P. Valduriez

Machine Translated by Google

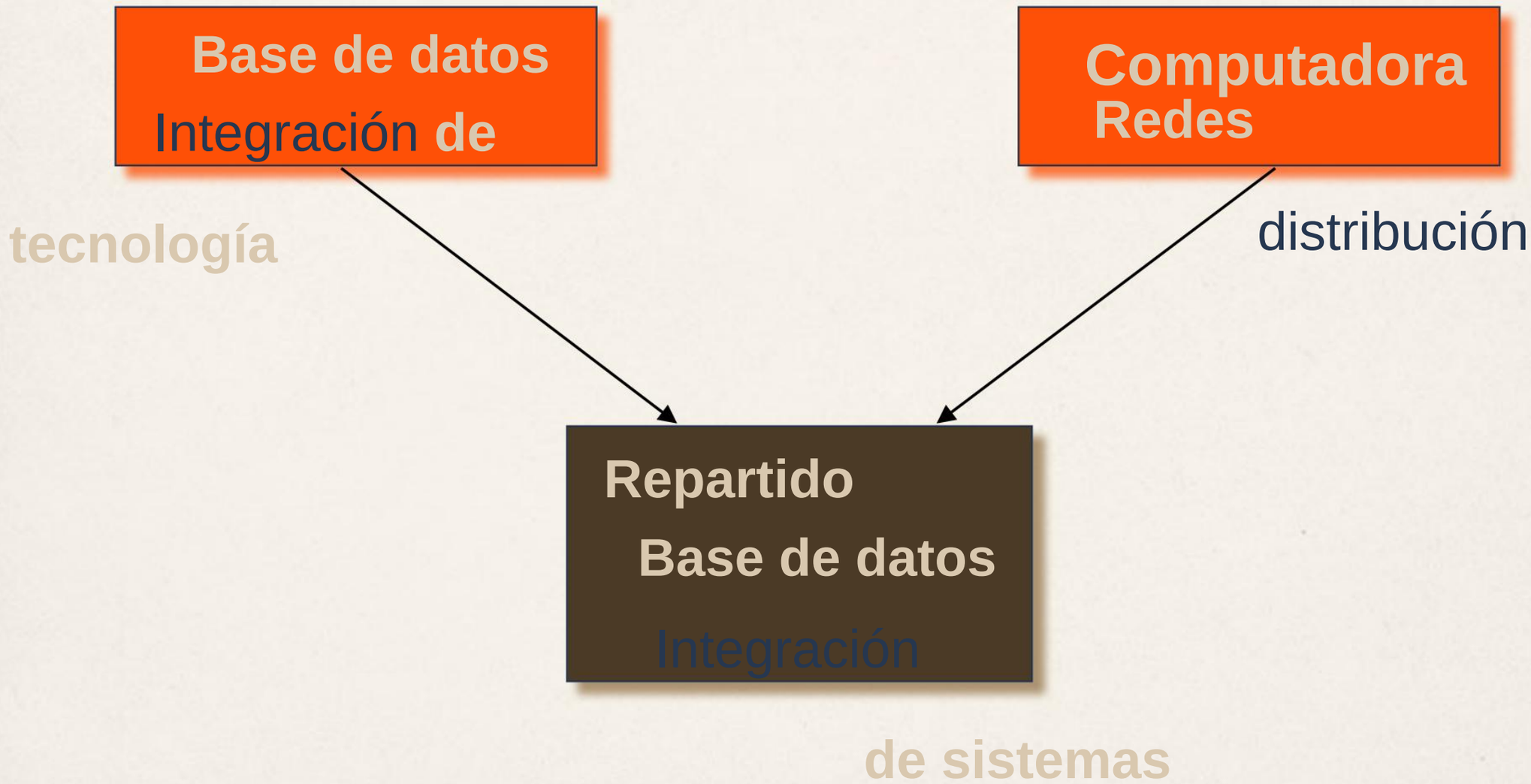
Gestión de base de datos



© MT Özsu & P. Valduriez

Machine Translated by Google

Motivación



integración ÿ centralización

© MT Özsu & P. Valduriez

Machine Translated by Google

Computación distribuída

- Una serie de elementos de procesamiento autónomos (no necesariamente homogéneos) que están interconectados por una red informática y que cooperan en la realización de las tareas asignadas.
- ¿Qué se está distribuyendo?
  - ÿ Lógica de procesamiento
  - ÿ Función
  - ÿ Datos
  - ÿ Controlar

© MT Özsu & P. Valduriez

Machine Translated by Google

¿Qué es un sistema de base de datos distribuida?

Una base de datos distribuida (DDB) es una colección de múltiples bases de datos interrelacionadas lógicamente distribuidas en una red informática.

Un sistema de gestión de bases de datos distribuidas (D-DBMS) es el software que gestiona la DDB y proporciona un mecanismo de acceso que hace que esta distribución sea transparente para los usuarios.

Sistema de base de datos distribuida (DDBS) = DDB + D–DBMS

© MT Özsu & P. Valduriez

Machine Translated by Google

¿Qué no es un DDBS?

- Un sistema informático de tiempo compartido
- Un sistema multiprocesador acoplado débil o estrechamente
- Un sistema de base de datos que reside en uno de los nodos de una red de computadoras: esta es una base de datos centralizada en un nodo de red

© MT Özsu & P. Valduriez

Machine Translated by Google

DBMS centralizado en un La red

© MT Özsu & P. Valduriez

Machine Translated by Google

SGBD distribuido Ambiente

© MT Özsu & P. Valduriez

Machine Translated by Google

suposiciones implícitas

- Datos almacenados en varios sitios ÿ cada sitio consta lógicamente de un solo procesador.
- Los procesadores en diferentes sitios están interconectados por una red informática ÿ no un sistema multiprocesador
  - ÿ Sistemas de bases de datos paralelas
- La base de datos distribuida es una base de datos, no una colección de archivos ÿ datos lógicamente relacionados como se muestran en los patrones de acceso de los usuarios
  - ÿ Modelo de datos relacionales
- D-DBMS es un DBMS completo
  - ÿ No es un sistema de archivos remoto, no es un sistema TP

© MT Özsu & P. Valduriez

Machine Translated by Google

Alternativas de entrega de datos

- Modos de entrega
  - Pull-only
  - Push-only
  - Híbrido
- Frecuencia
  - Periódico
  - Condicional
  - Ad-hoc o irregular
- Métodos de comunicación
  - Unidifusión
  - Uno a muchos
- Nota: no todas las combinaciones tienen sentido

© MT Özsu & P. Valduriez

Machine Translated by Google

Promesas DBMS distribuidas

ÿGestión transparente de datos distribuidos, fragmentados y datos replicados

ÿMejora de la confiabilidad/disponibilidad a través de transacciones distribuidas

ÿRendimiento mejorado

ÿExpansión del sistema más fácil y económica

© MT Özsu & P. Valduriez

Machine Translated by Google

Transparencia

- La transparencia es la separación de la semántica de nivel superior de un sistema de los problemas de implementación de nivel inferior.
- La cuestión fundamental es proporcionar

  independencia de datos

  en el entorno distribuido

  ÿ Transparencia de red (distribución)

  ÿ Transparencia de replicación

  ÿ Fragmentación transparencia ÿ

  fragmentación horizontal: selección ÿ

  fragmentación vertical: proyección ÿ

  híbrido

© MT Özsu & P. Valduriez

Machine Translated by Google

Ejemplo

EMP

ENO	ENAME	TITLE
E1	J. Doe	Elect. Eng
E2	M. Smith	Syst. Anal.
E3	A. Lee	Mech. Eng.
E4	J. Miller	Programmer
E5	B. Casey	Syst. Anal.
E6	L. Chu	Elect. Eng.
E7	R. Davis	Mech. Eng.
E8	J. Jones	Syst. Anal.

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36
E8	P3	Manager	40

PROJ

PNO	PNAME	BUDGET
P1	Instrumentation	150000
P2	Database Develop.	135000
P3	CAD/CAM	250000
P4	Maintenance	310000

PAY

TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

© MT Özsu & P. Valduriez

Machine Translated by Google

Acceso transparente

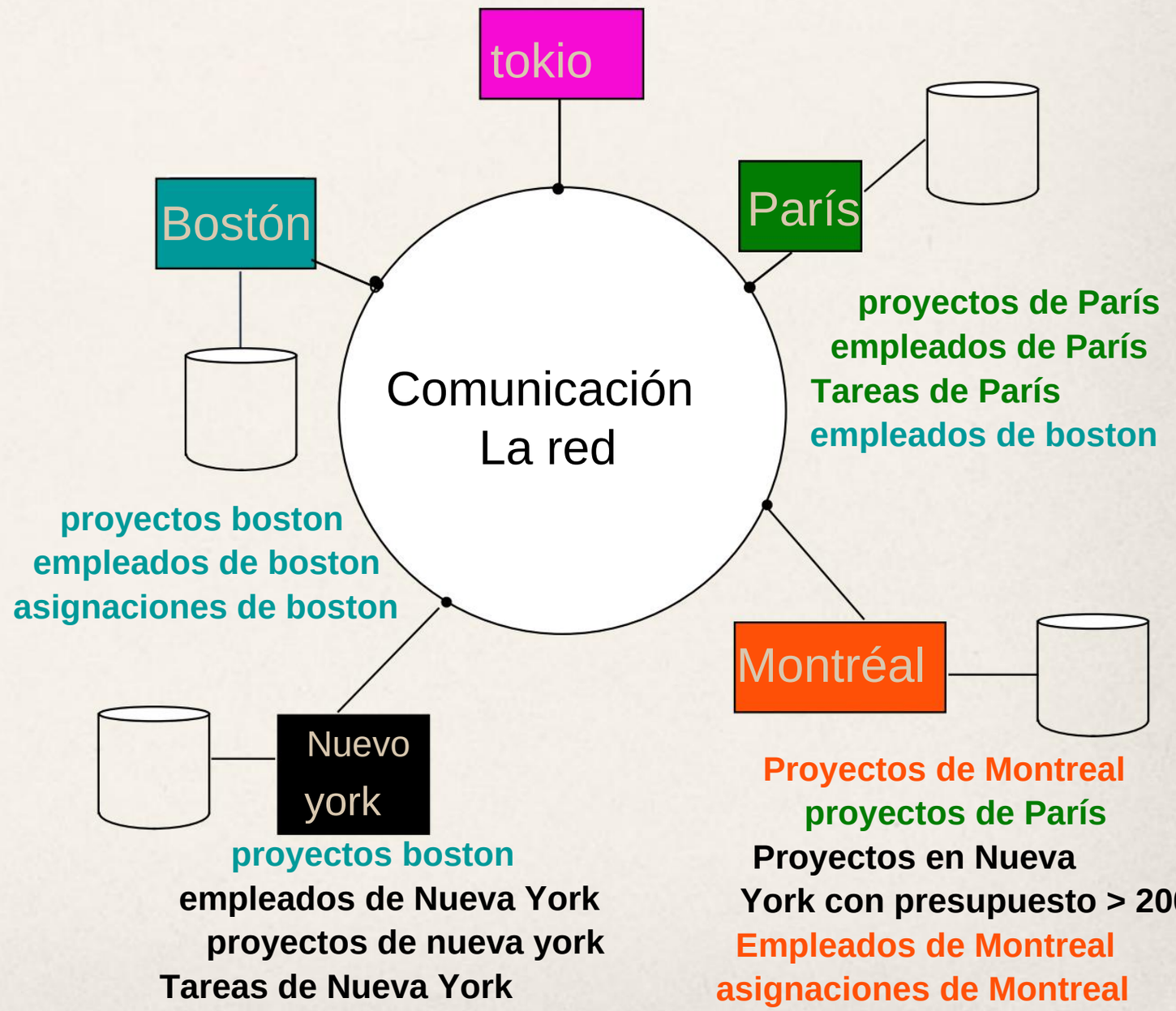
SELECCIONE NOMBRE, SAL

DESDE EMP,ASG,PAGO

DONDE DUR > 12

Y EMP.ENO = ASG.ENO

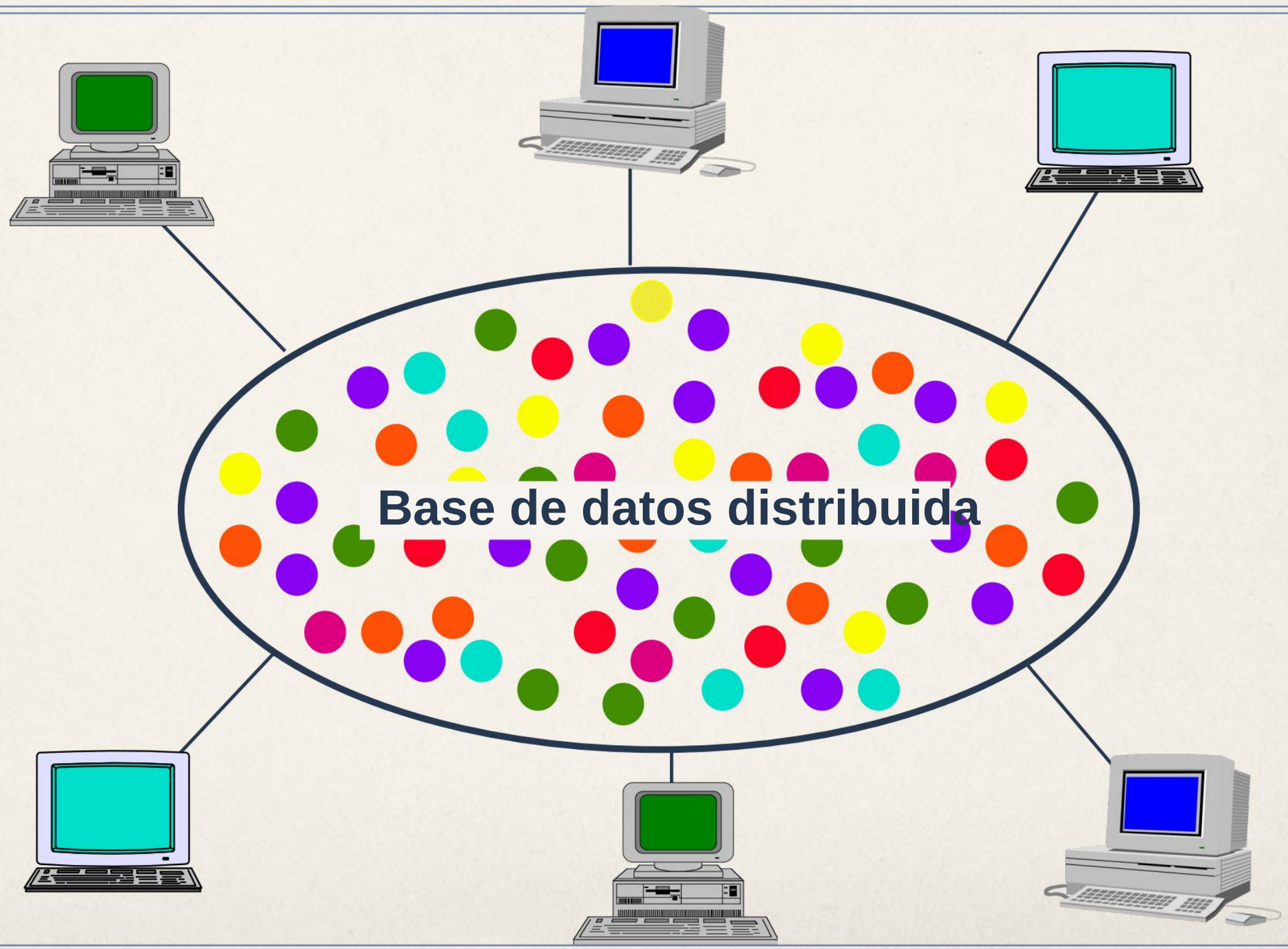
AND PAY.TITLE = EMP.TITLE



© MT Özsu & P. Valduriez

Machine Translated by Google

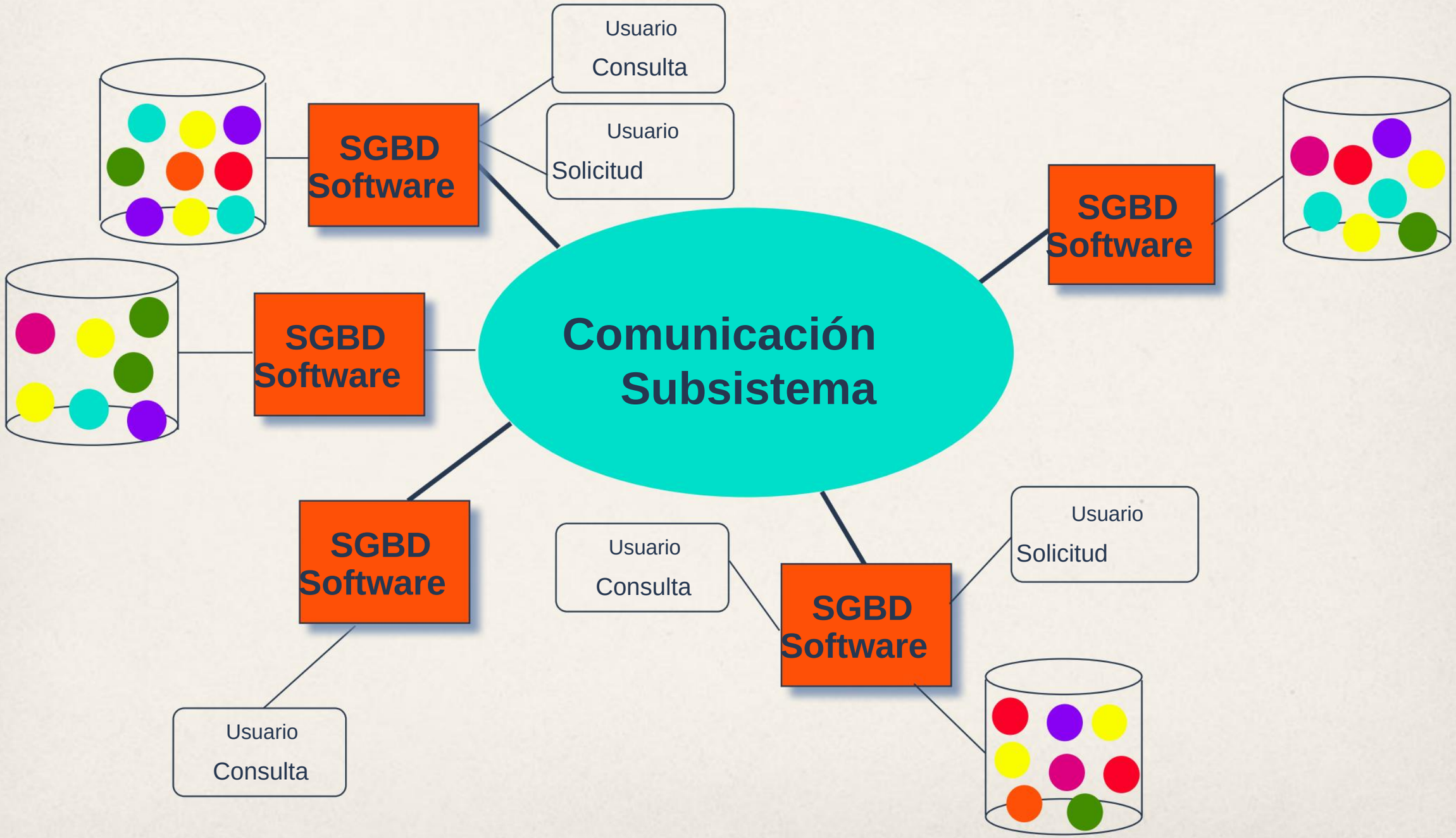
Base de datos distribuida - Usuario Vista



© MT Özsu & P. Valduriez

Machine Translated by Google

DBMS Distribuido - Realidad



© MT Özsu & P. Valduriez

Machine Translated by Google

Tipos de transparencia

- Independencia de datos
- Transparencia de red (o transparencia de distribución) ÿ

  Transparencia de ubicación ÿ Transparencia de fragmentación

- Transparencia de replicación
- Transparencia de fragmentación

© MT Özsu & P. Valduriez

Machine Translated by Google

Confiabilidad a través de Actas

- Los componentes y datos replicados deben hacer un DBMS distribuido más confiable.
- Las transacciones distribuidas proporcionan
  - ÿ Transparencia de concurrencia
  - ÿ Atomicidad de fallas
- El soporte de transacciones distribuidas requiere la implementación de ÿ Protocolos de control de concurrencia distribuida
  - ÿ Comprometer protocolos
- Replicación de datos
  - ÿ Genial para cargas de trabajo de lectura intensiva, problemático para actualizaciones
  - ÿ Protocolos de replicación

© MT Özsu & P. Valduriez

Machine Translated by Google

Potencialmente mejorado Actuación

- Proximidad de los datos a sus puntos de uso
  - ÿ Requiere algo de soporte para fragmentación y replicación
- Paralelismo en la ejecución
  - ÿ Paralelismo entre consultas
  - ÿ Paralelismo intra-consulta

© MT Özsu & P. Valduriez

Machine Translated by Google

Requisitos de paralelismo

- Tener la mayor cantidad de datos requeridos por cada aplicación en el sitio donde se ejecuta la aplicación
  - ÿ Replicación completa
- ¿Qué hay de las actualizaciones?
  - ÿ Coherencia mutua
  - ÿ Frescura de ejemplares

© MT Özsu & P. Valduriez

Machine Translated by Google

Expansión del sistema

- El problema es el escalado de la base de datos
- Aparición de tecnologías de microprocesadores y estaciones de trabajo
  - ÿ Fallecimiento de la ley de Grosh
  - ÿ Modelo de computación cliente-servidor
- Costo de comunicación de datos vs costo de telecomunicaciones

© MT Özsu & P. Valduriez

Machine Translated by Google

Problemas de DBMS distribuidos

- **Diseño de base de datos distribuida**
  - ÿ Cómo distribuir la base de datos
  - ÿ Distribución de bases de datos replicadas y no replicadas ÿ Un problema relacionado con la gestión de directorios
- **Procesamiento de consultas**
  - ÿ Convertir transacciones de usuario en instrucciones de manipulación de datos ÿ Problema de optimización
    - ÿ min{costo = transmisión de datos + procesamiento local}
  - ÿ La formulación general es NP-dura

© MT Özsu & P. Valduriez

Machine Translated by Google

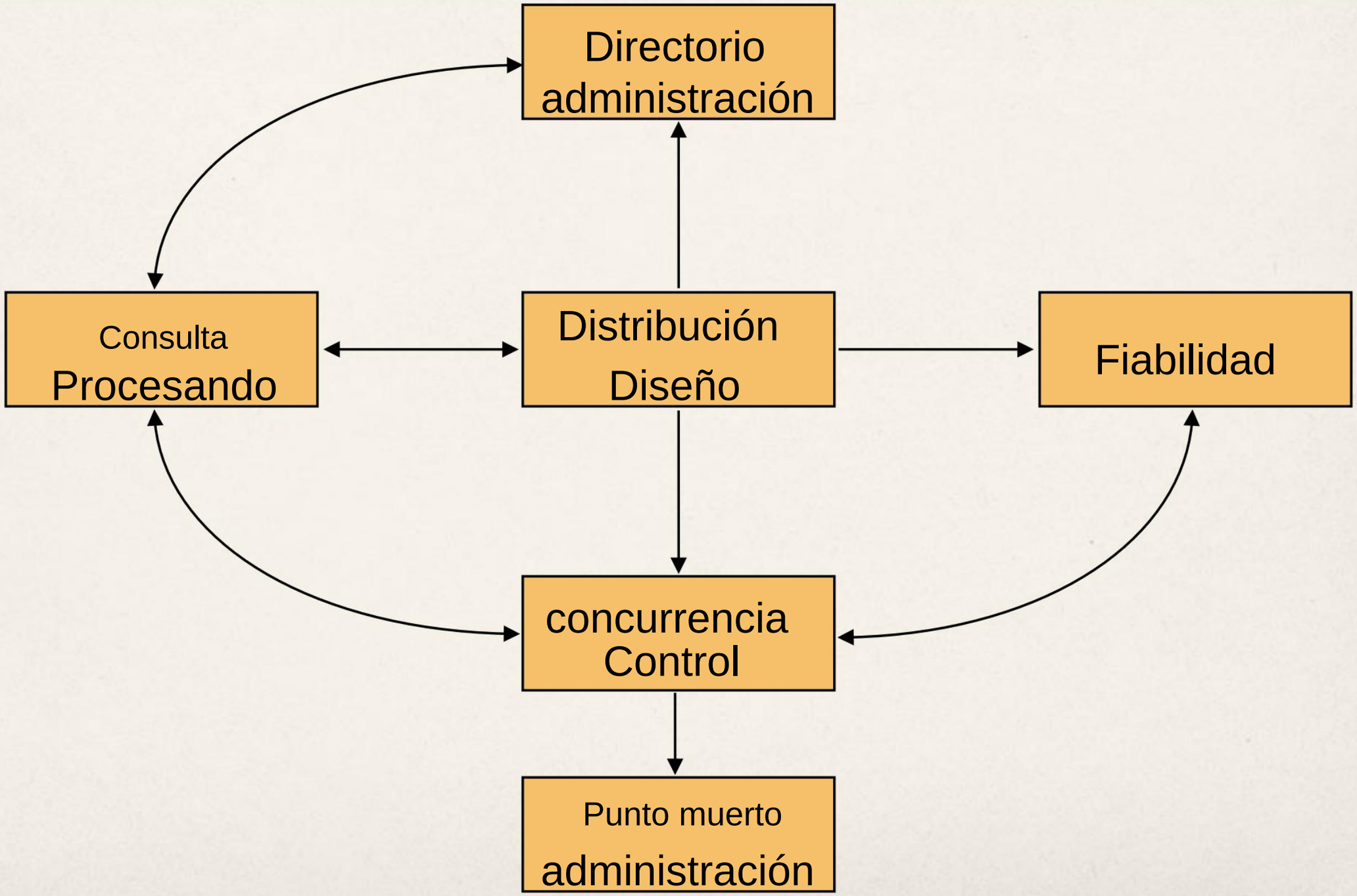
Problemas de DBMS distribuidos

- **Control de concurrencia**
  - ÿ Sincronización de accesos concurrentes
  - ÿ Consistencia y aislamiento de los efectos de las transacciones
  - ÿ Gestión de puntos muertos
- **Fiabilidad**
  - ÿ Cómo hacer que el sistema sea resistente a fallas
  - ÿ Atomicidad y durabilidad

© MT Özsu & P. Valduriez

Machine Translated by Google

Relación entre problemas



© MT Özsu & P. Valduriez

Machine Translated by Google

Asuntos relacionados

- **Soporte del sistema operativo**
  - ÿ Sistema operativo con soporte adecuado para operaciones de base de datos
  - ÿ Dicotomía entre requisitos de procesamiento de propósito general y requisitos de procesamiento de la base de datos
- **Sistemas Abiertos e Interoperabilidad**
  - ÿ Sistemas Distribuidos de Bases de Datos Múltiples
  - ÿ Escenario más probable
  - ÿ Cuestiones paralelas

© MT Özsu & P. Valduriez

Machine Translated by Google

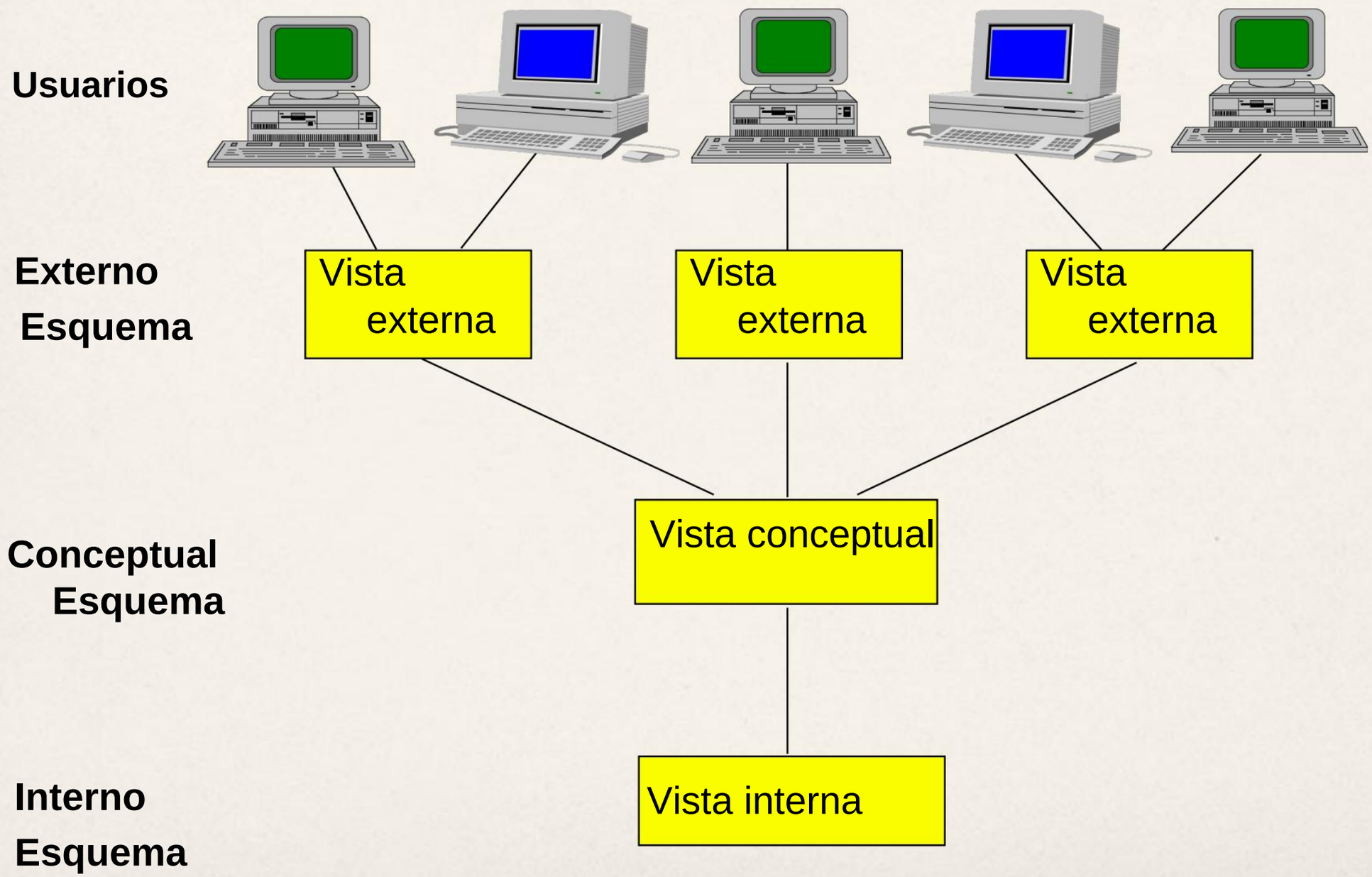
Arquitectura

- Define la estructura del sistema
  - ÿ componentes identificados
  - ÿ funciones de cada componente definidas
  - ÿ interrelaciones e interacciones entre componentes definidas

© MT Özsu & P. Valduriez

Machine Translated by Google

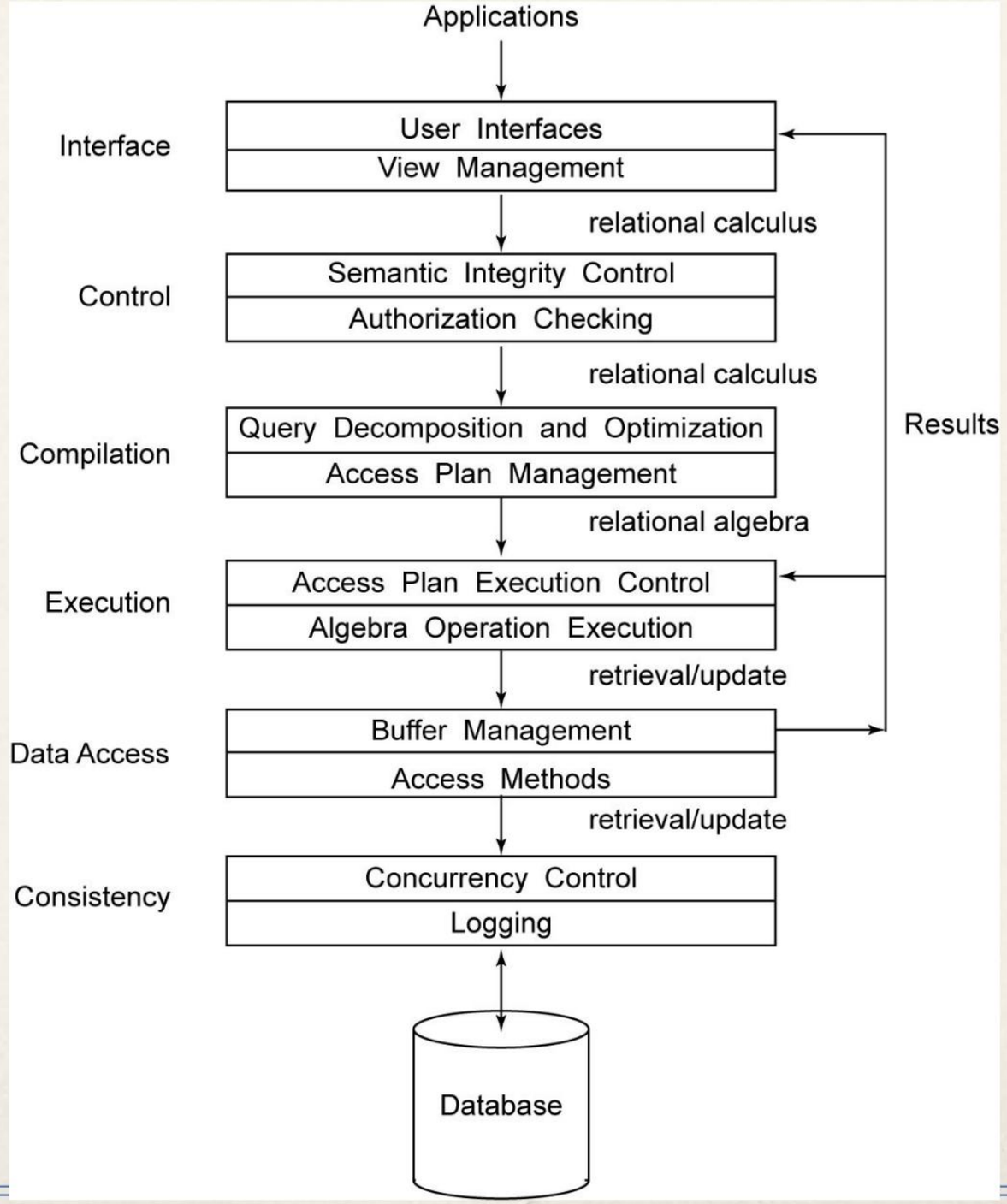
Arquitectura ANSI/SPARC



© MT Özsu & P. Valduriez

Machine Translated by Google

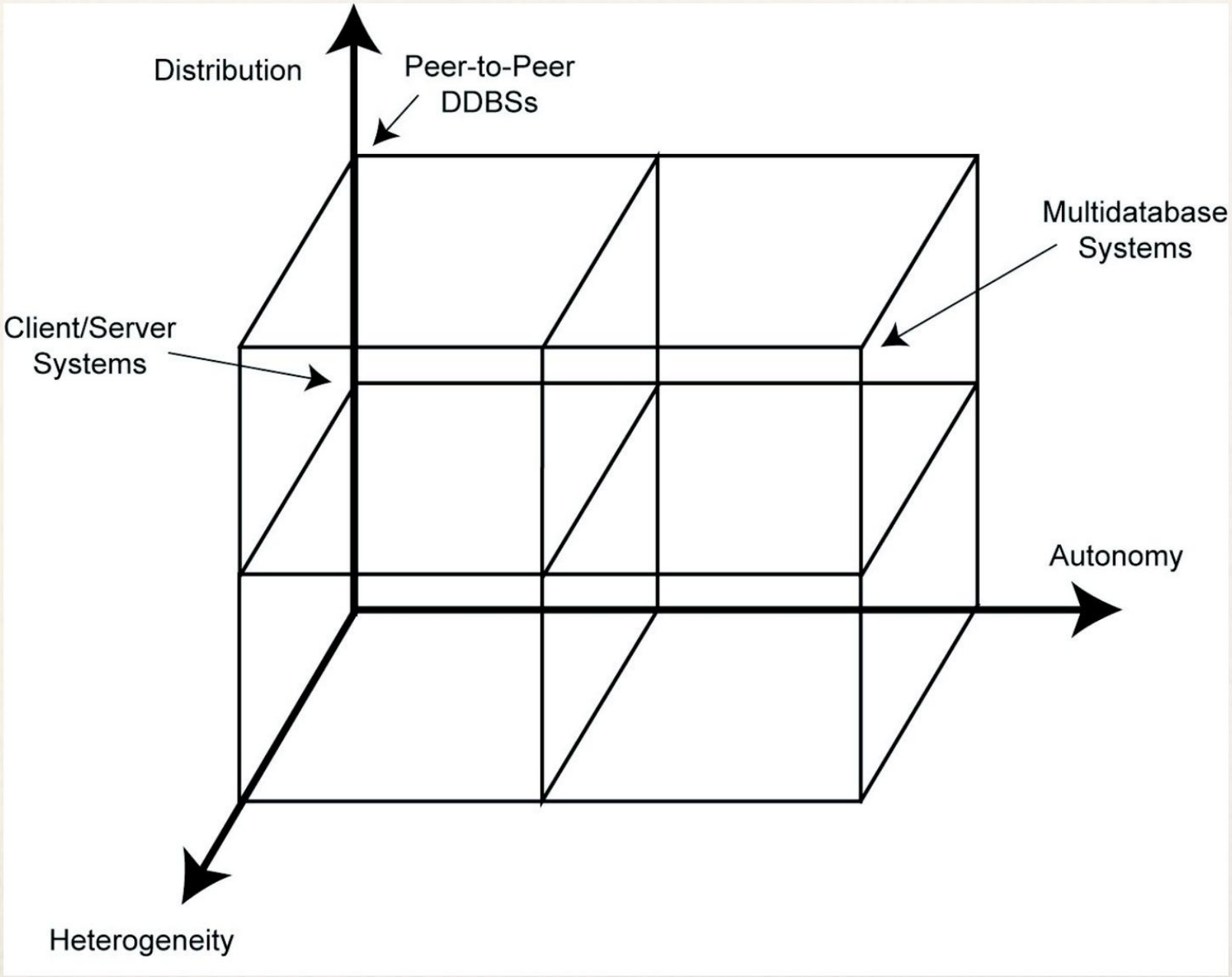
Arquitectura DBMS genérica



© MT Özsu & P. Valduriez

Machine Translated by Google

Implementación de SGBD Alternativas



© MT Özsu & P. Valduriez

Machine Translated by Google

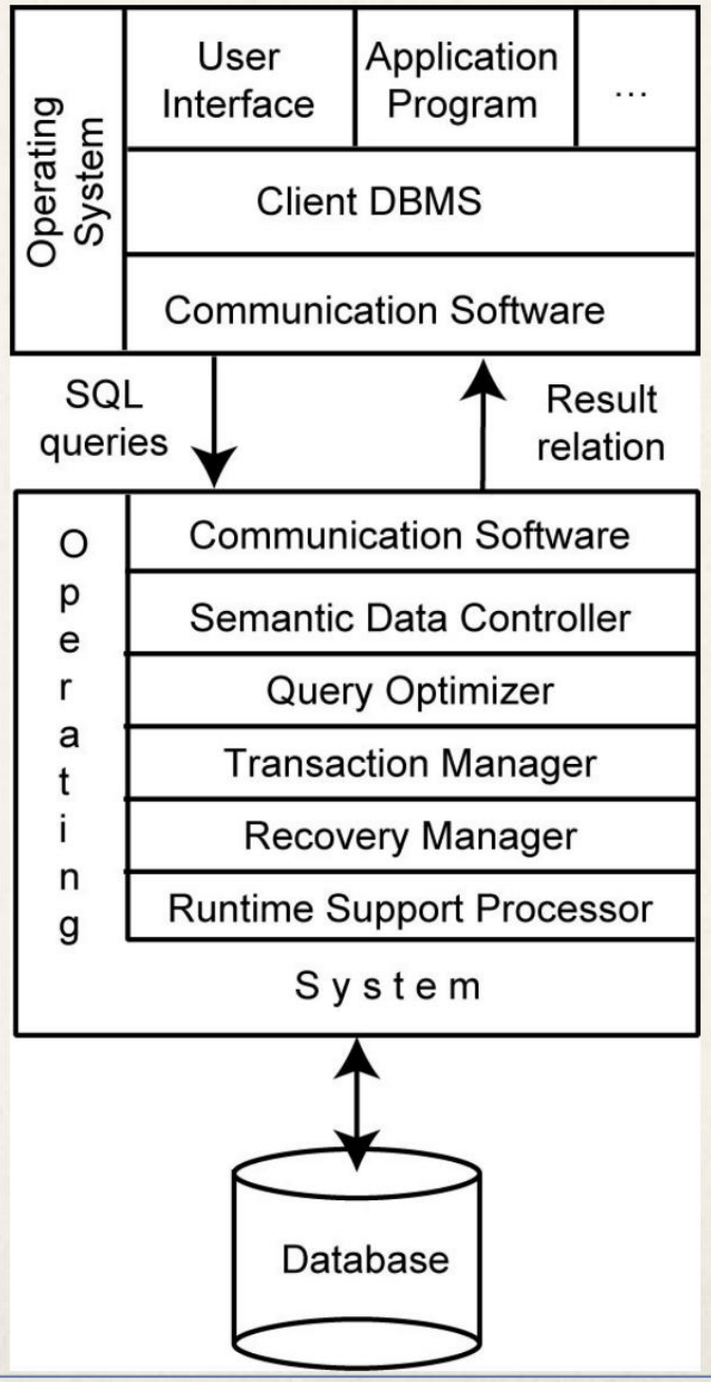
Dimensiones del Problema

- Distribución
  - ÿ Si los componentes del sistema están ubicados en la misma máquina O no
- Heterogeneidad
  - ÿ Varios niveles (hardware, comunicaciones, sistema operativo)
  - ÿ DBMS importante
    - ÿ modelo de datos, lenguaje de consulta, algoritmos de gestión de transacciones
- Autonomía
  - ÿ No bien entendido y muy problemático
  - ÿ Varias versiones
    - ÿ Autonomía de diseño: Capacidad de un componente DBMS para decidir sobre cuestiones relacionadas con su propio diseño.
    - ÿ Autonomía de comunicación: Habilidad de un DBMS componente para decidir si y cómo comunicarse con otros DBMS.
    - ÿ Autonomía de ejecución: Habilidad de un componente DBMS para ejecutar operaciones locales de cualquier forma que quiera.

© MT Özsu & P. Valduriez

Machine Translated by Google

Arquitectura Cliente/Servidor



© MT Özsu & P. Valduriez

Machine Translated by Google

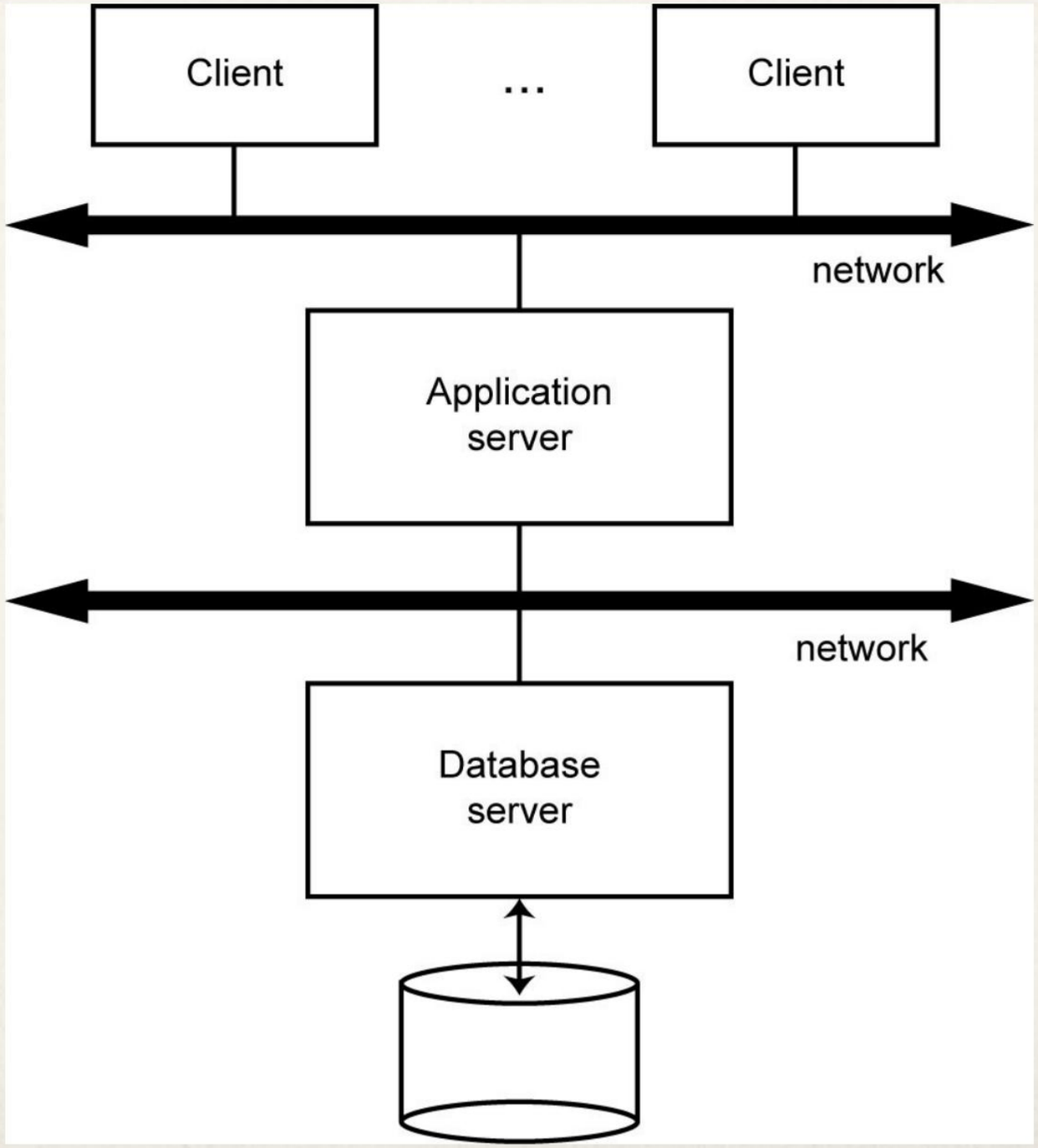
Ventajas de Cliente-Servidor arquitecturas

- División del trabajo más eficiente
- Escalado horizontal y vertical de recursos
- Mejor relación precio/rendimiento en las máquinas de los clientes
- Capacidad de usar herramientas familiares en las máquinas del cliente
- Acceso del cliente a datos remotos (a través de estándares)
- Funcionalidad DBMS completa proporcionada a las estaciones de trabajo del cliente
- Mejor precio/rendimiento general del sistema

© MT Özsu & P. Valduriez

Machine Translated by Google

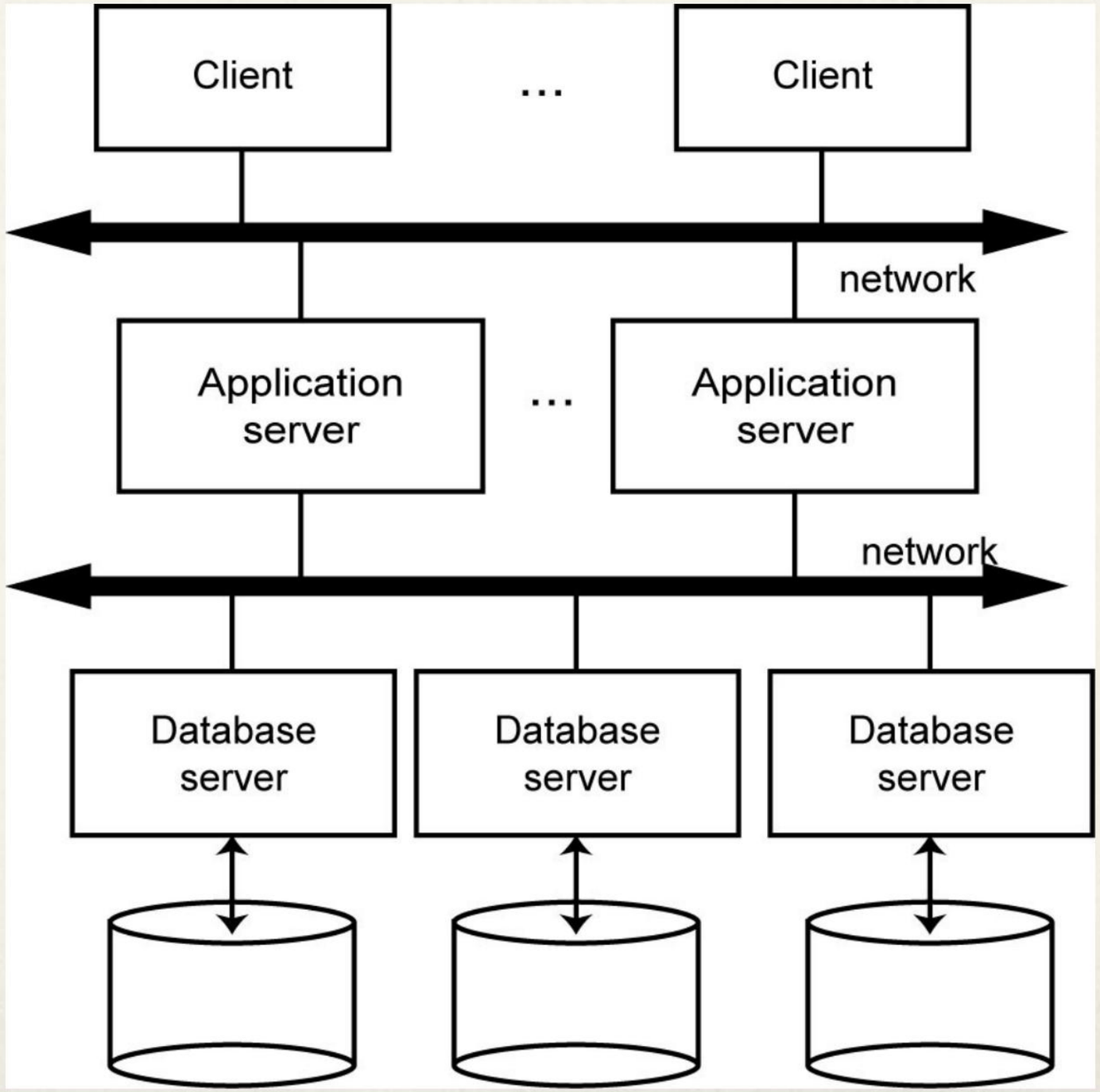
Servidor de base de datos



© MT Özsu & P. Valduriez

Machine Translated by Google

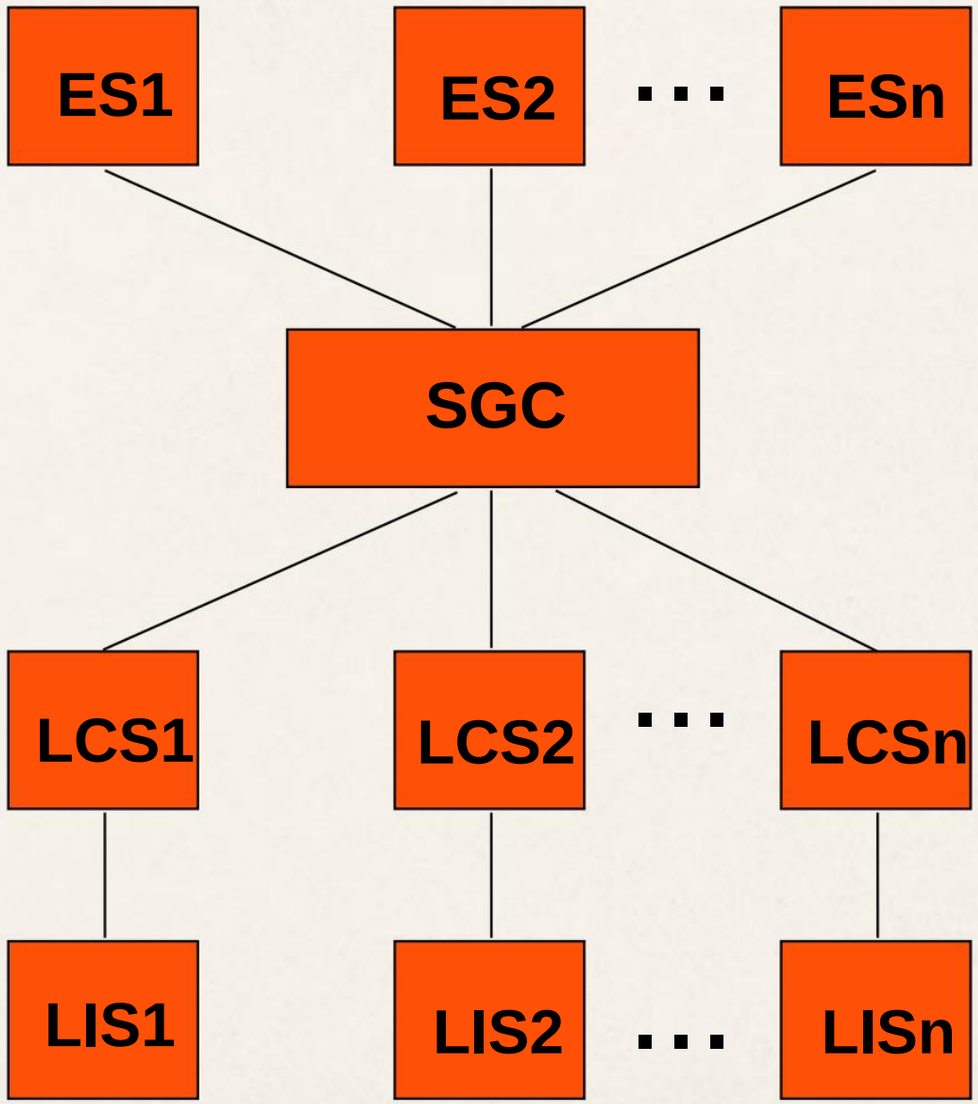
Servidores de bases de datos distribuidas



© MT Özsu & P. Valduriez

Machine Translated by Google

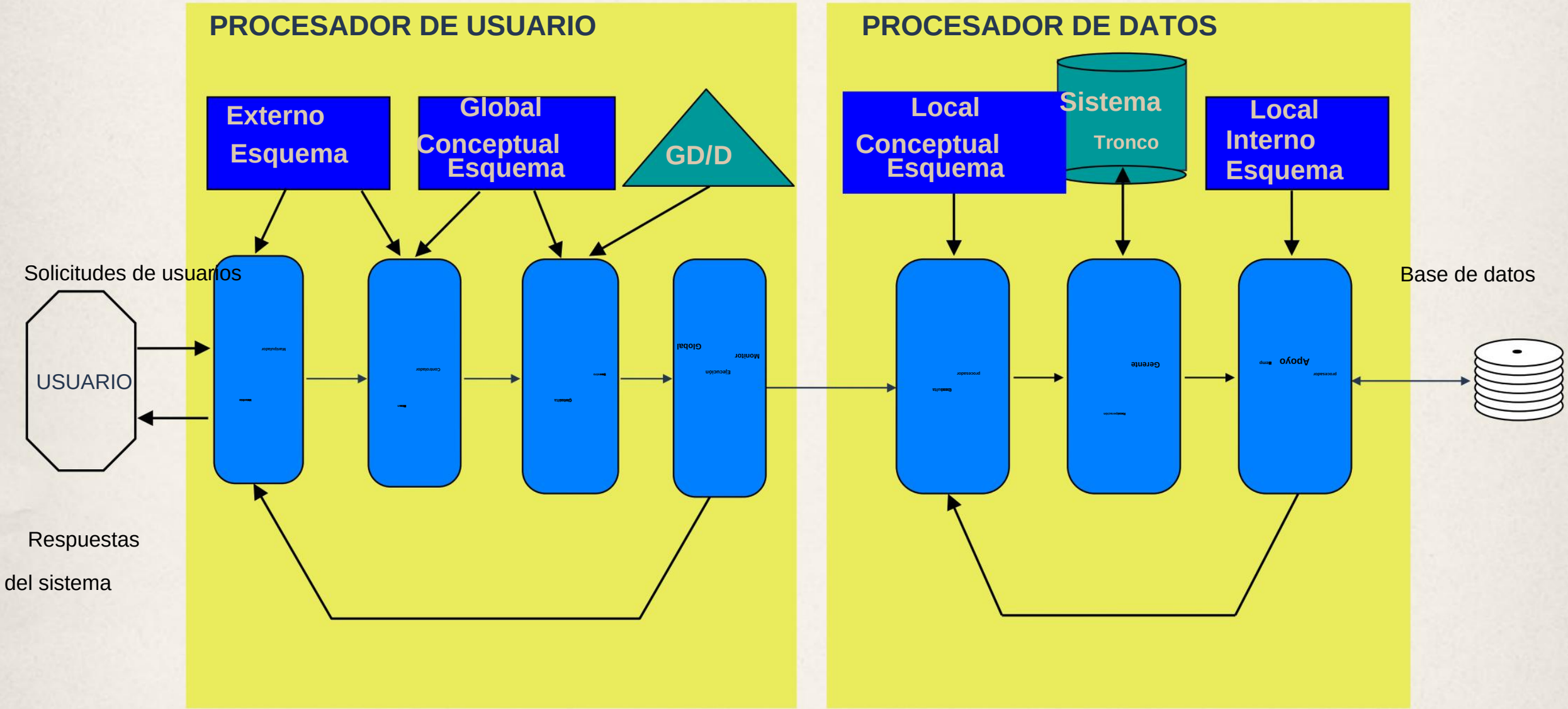
DBMS distribuido de datos lógicos Arquitectura



© MT Özsu & P. Valduriez

Machine Translated by Google

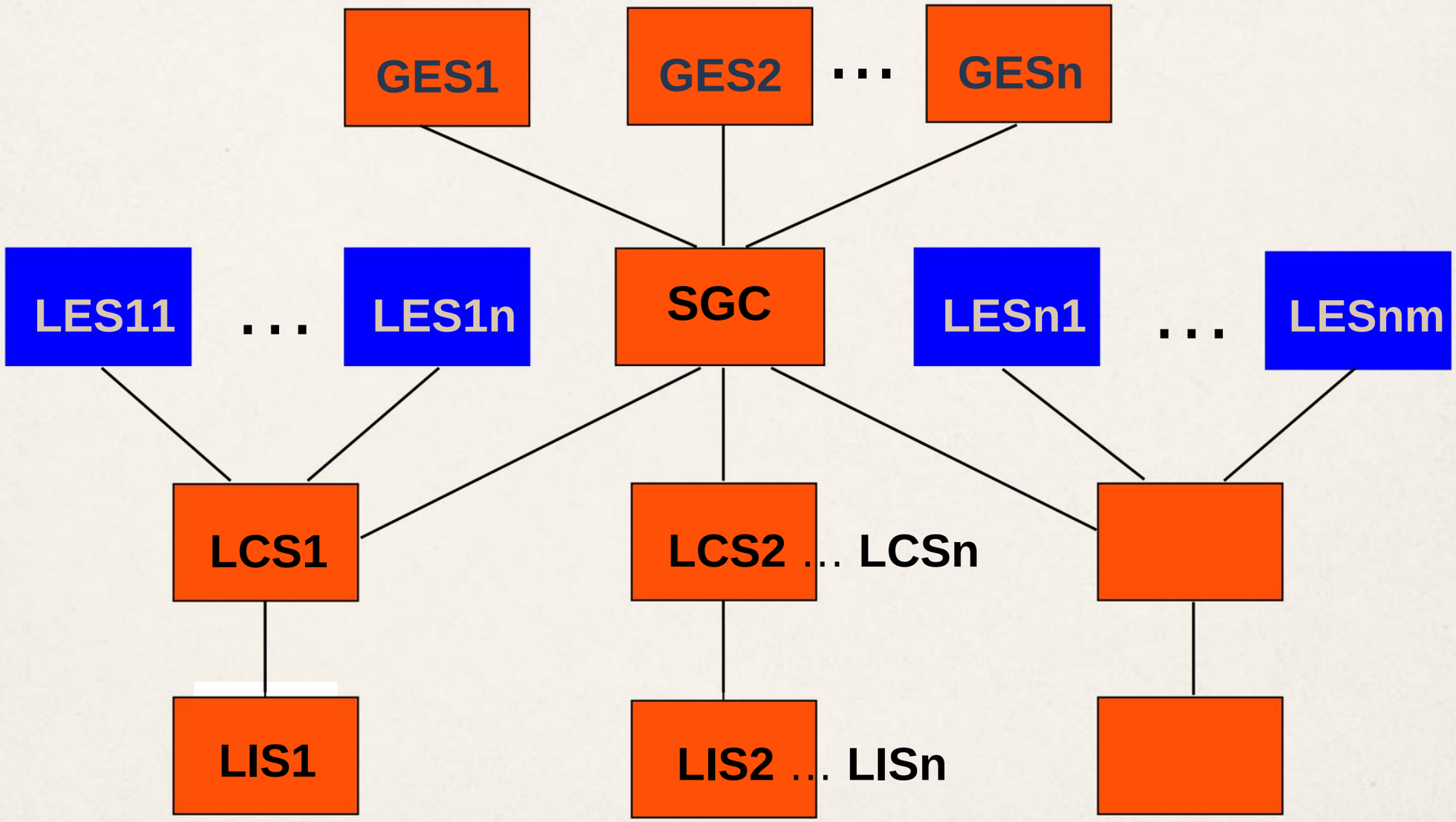
Componente de igual a igual Arquitectura



© MT Özsu & P. Valduriez

Machine Translated by Google

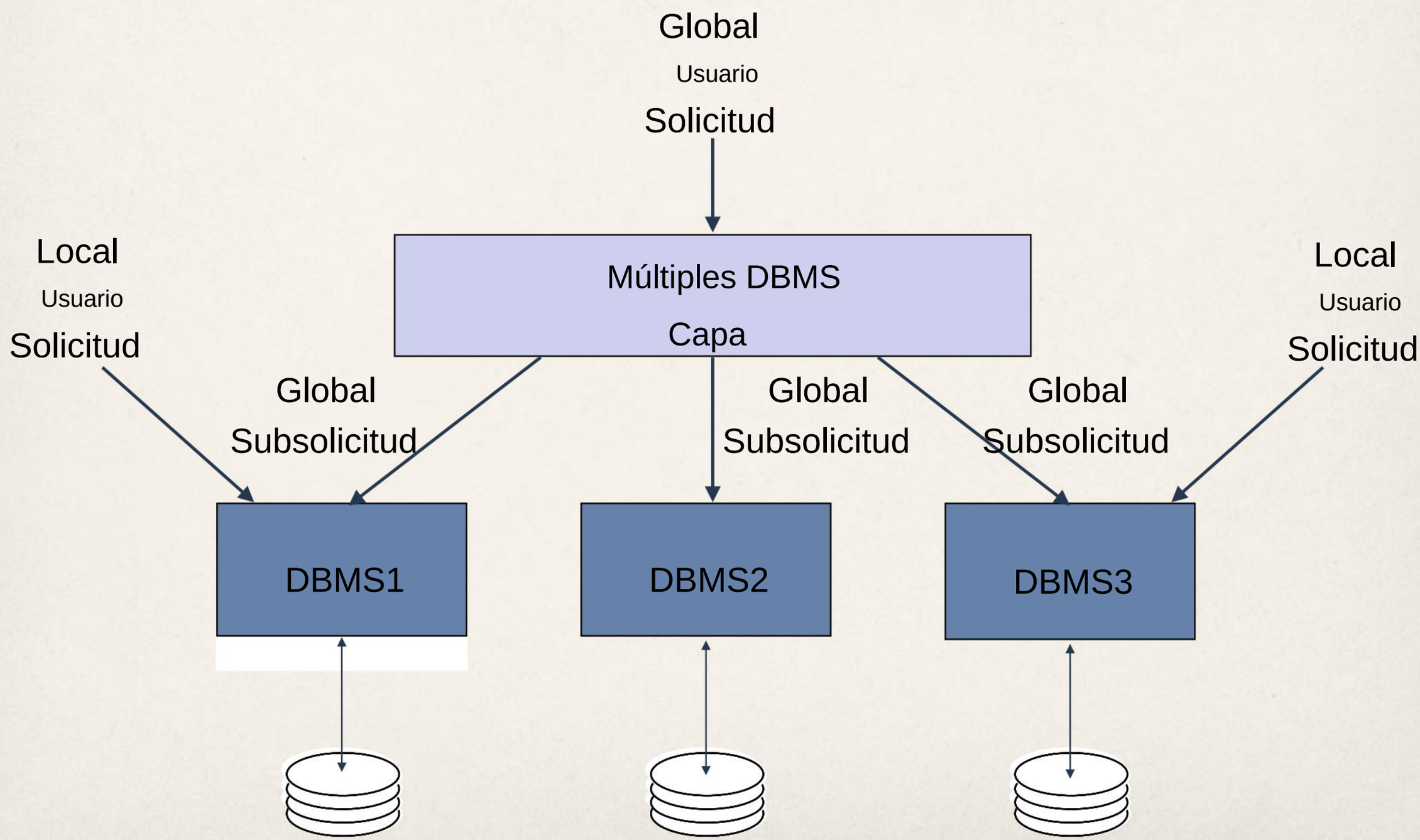
Multi-DBMS de datos lógicos Arquitectura



© MT Özsu & P. Valduriez

Machine Translated by Google

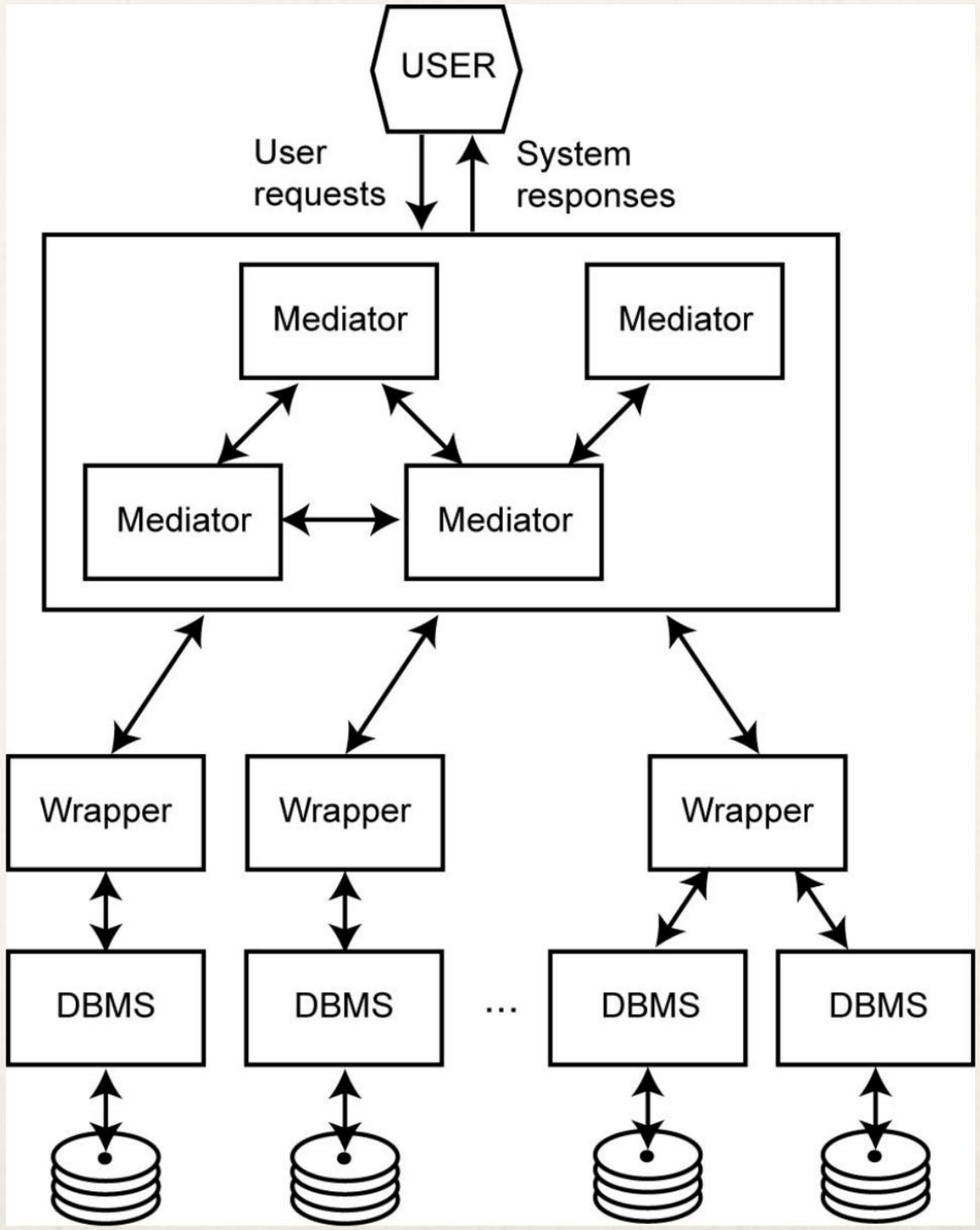
Componentes MDBS y Ejecución



© MT Özsu & P. Valduriez

Machine Translated by Google

Mediador/Envoltorio
Arquitectura



© MT Özsu & P. Valduriez