ADMINISTRACIÓN DE BASE DE DATOS

Unidad III

BASE DE DATOS DISTRIBUIDAS

Ana E. Congacha

Necesidad de Datos Distribuidos

Un entorno de datos distribuido es aquél donde puede haber varias copias de la misma información en distintos servidores.

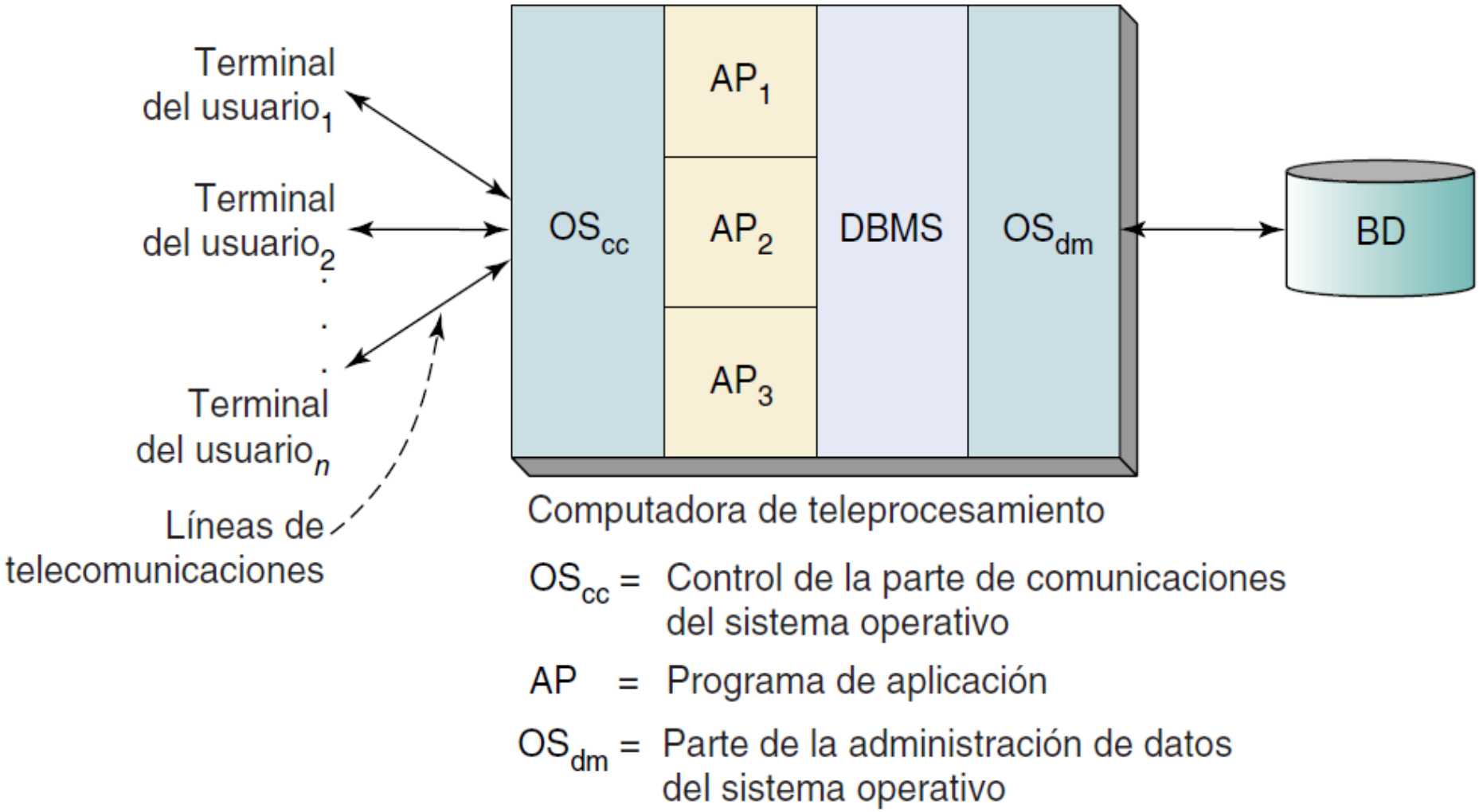
ARQUITECTURAS DEL PROCESAMIENTO DE LA BASE DE DATOS EMPRESARIAL

- Sistemas de Teleprocesamiento
- Sistemas de Archivos Compartidos
- Sistemas Cliente-Servidor
- Sistemas de Base de Datos Distribuidas

Sistemas de Teleprocesamiento

- Todo el procesamiento se realiza con una sola computadora.
- Los usuarios operan terminales tontas que transmiten los mensajes y los datos de una transacción a la computadora central.
- Históricamente fueron la alternativa más común para las Bases de datos de sistemas de multiusuario.

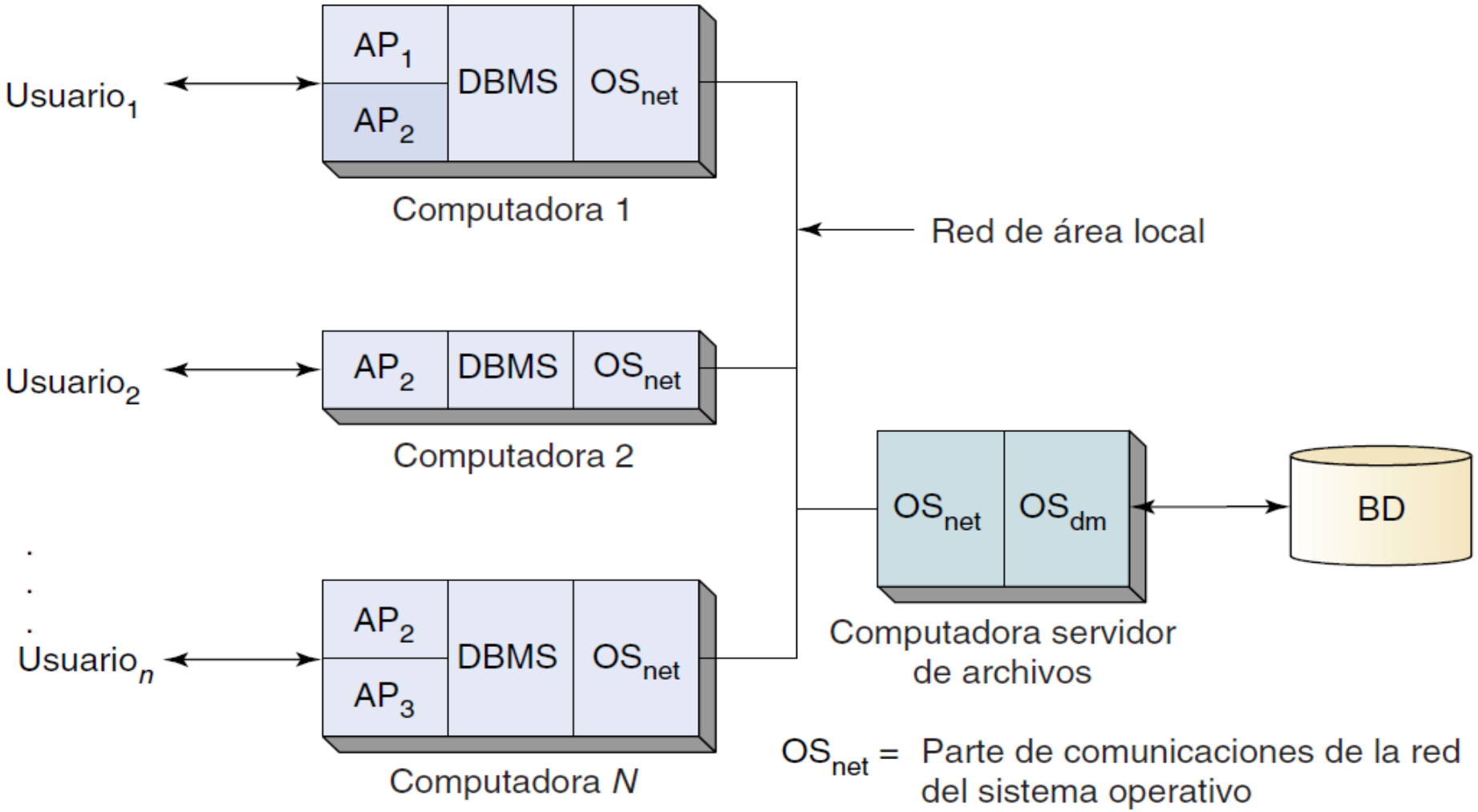
Sistema de teleproceso



Sistemas de archivos compartidos

- Distribuye los programas de aplicaciones y el DBMS a las computadoras de los usuarios.
- El servidor es un servidor de archivos y no uno de Base de Datos.
- El DBMS de cada computadora de un usuario envía peticiones al servidor de archivos para el procesamiento de nivel de archivos.
- Demasiado tráfico cruza por la LAN

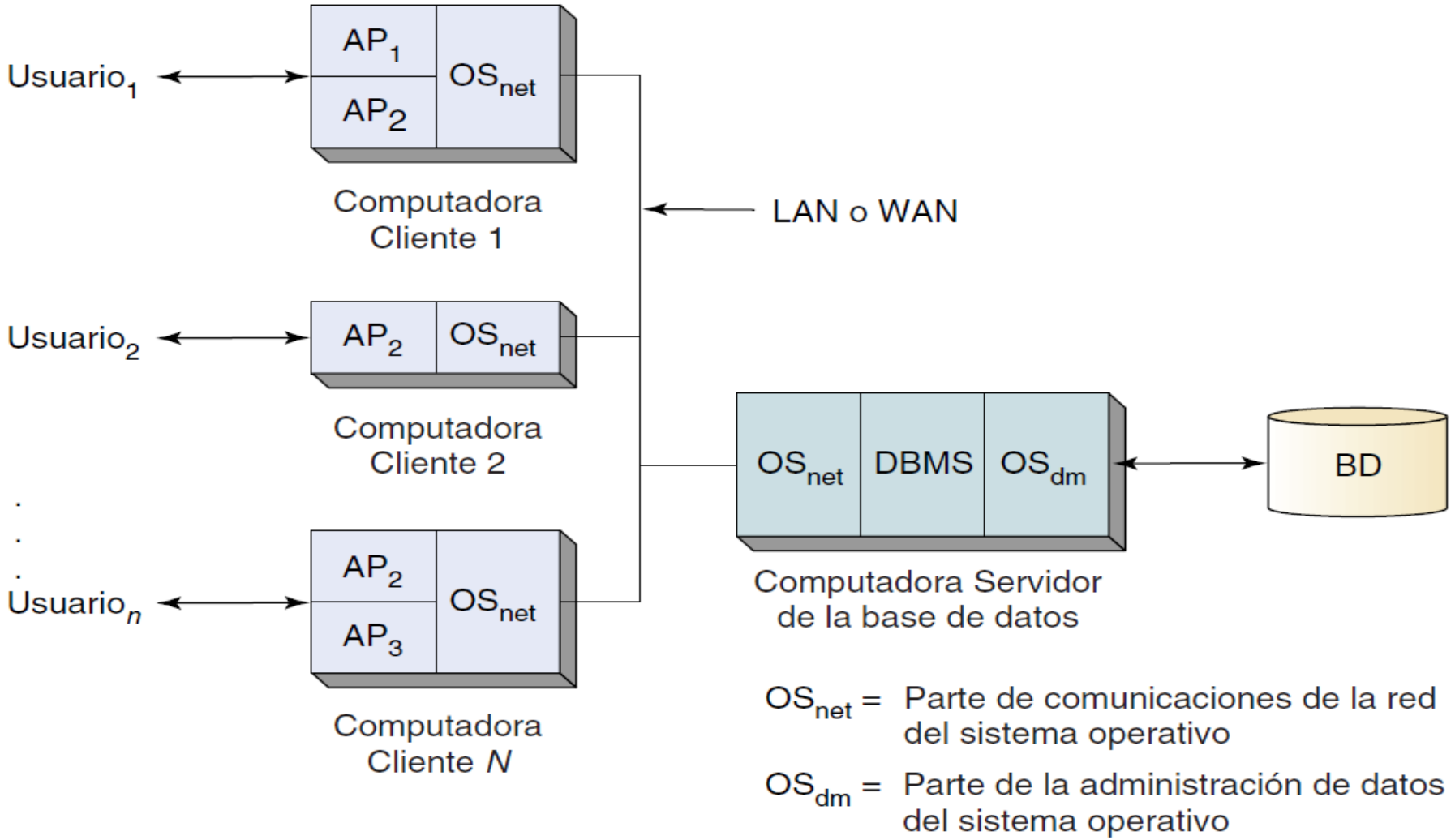
Arquitectura de Archivos Compartidos



Sistemas Cliente – Servidor

- Implica múltiples computadoras conectadas a una red
  - Clientes: procesa programas de aplicación.
  - Servidor: procesa la base de datos.
- Si hay múltiples servidores de procesamiento de Base de Datos, cada uno debe procesar una base de datos diferente.

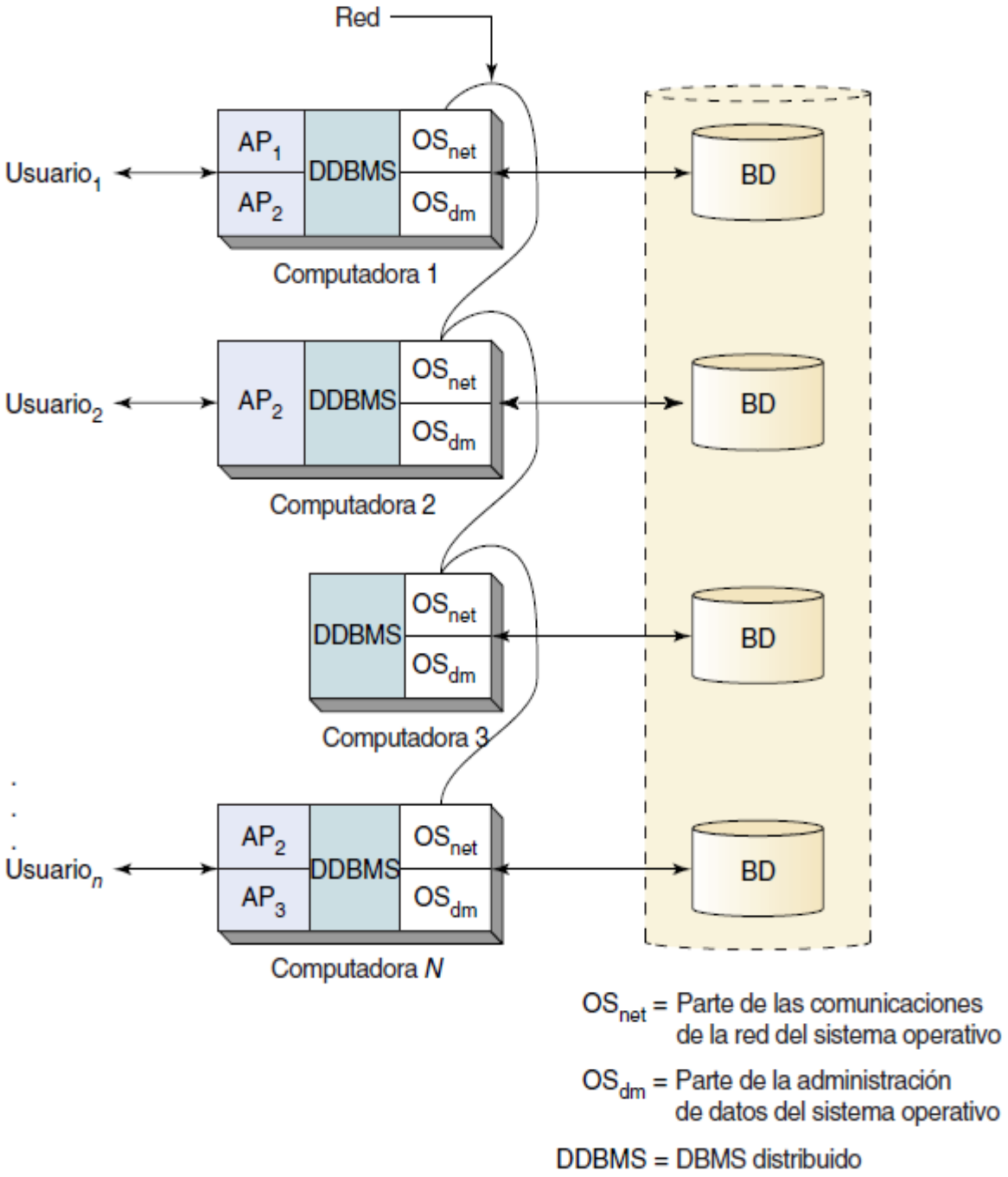
Arquitectura Cliente-Servidor



Sistemas de Base de Datos Distribuidas

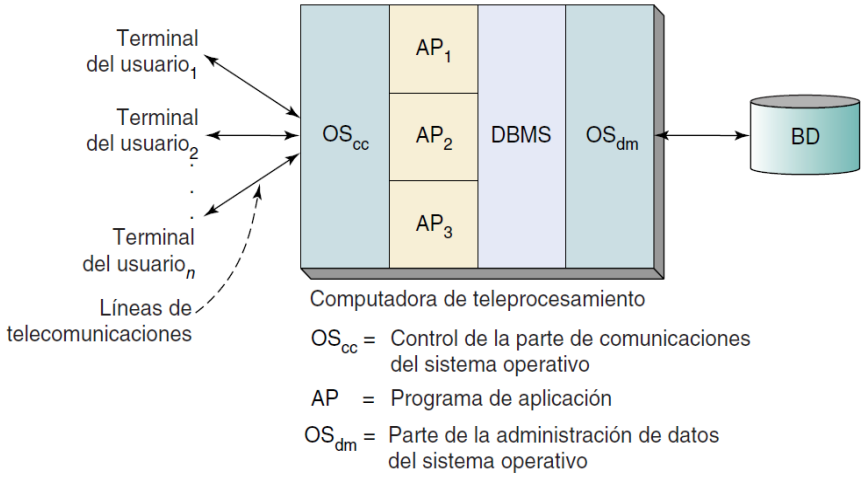
Los datos están repartidos entre más de una máquina, y un sistema de gestión de bases de datos distribuidas (SGBDD) como el software que gestiona una base de datos distribuida haciendo que la distribución de los datos sea transparente al usuario, es decir, los usuarios actúan como si todo estuviese junto en una sola base de datos.

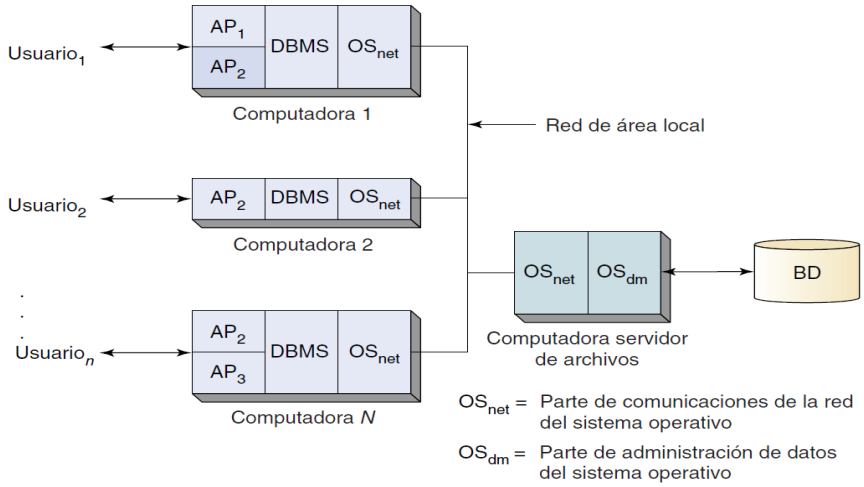
Sistemas de Base de Datos Distribuidas

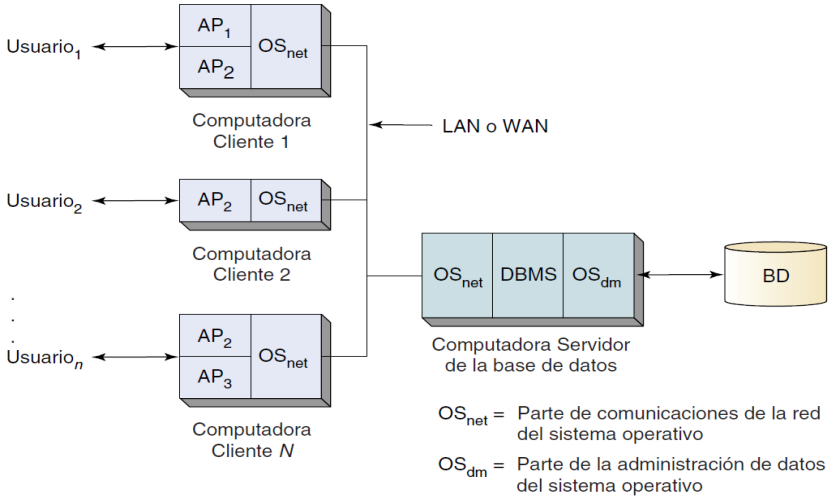


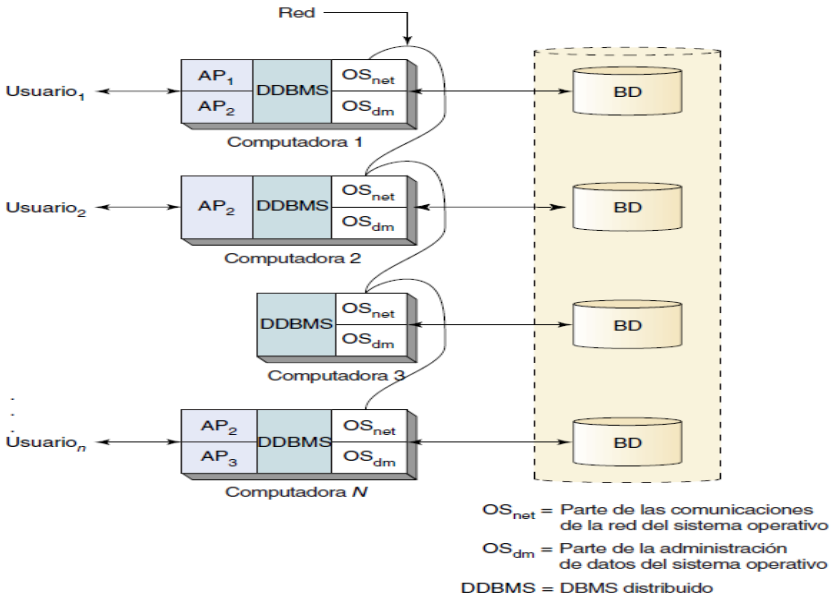
- La BD o parte de ésta se almacena en todas las N computadoras.
- La computadora 1, 2 y n procesan las aplicaciones y BD
- Computadora 3 procesa solamente la BD
- La línea punteada indica que la BD está compuesta de todos los segmentos de la BD en todas las n computadoras.
- Se pueden ubicar físicamente en el mismo lugar, en diferentes partes del mundo o en algún lugar intermedio.

PROCESAMIENTO DISTRIBUIDO CONTRA PROCESAMIENTO DE BASE DE DATOS DISTRIBUIDAS.



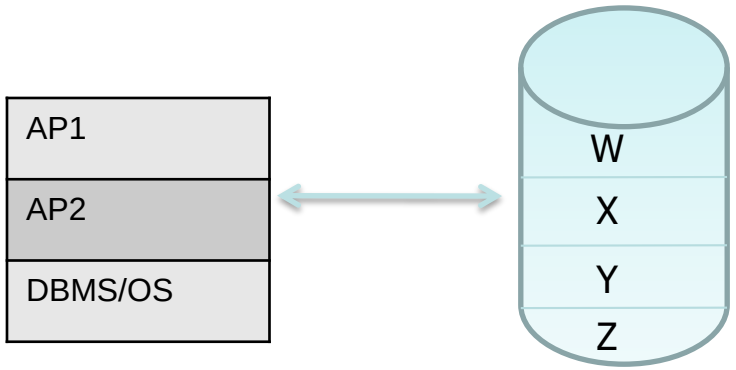






Tipos de BD Distribuidas

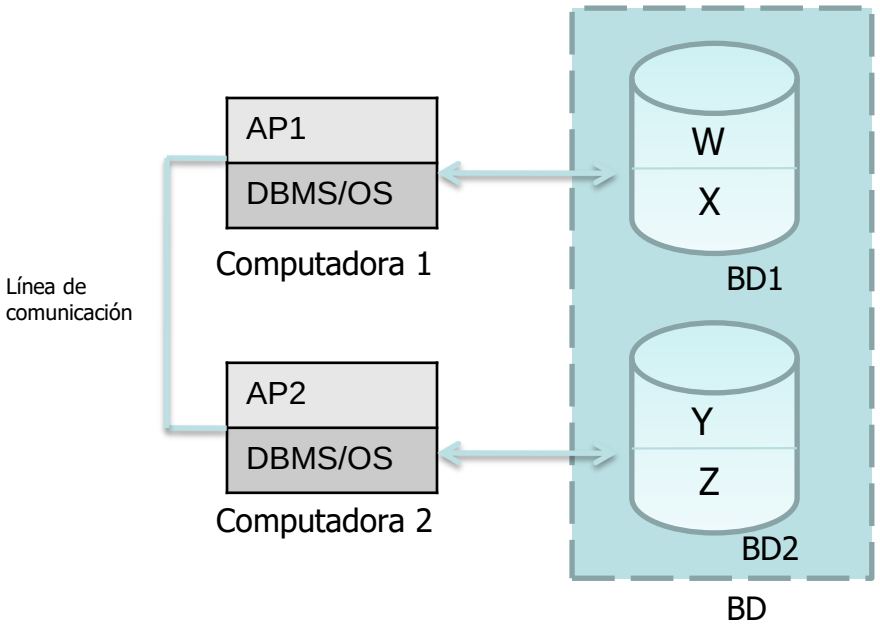
Partamos de:



Procesamiento en una sola computadora

BD no distribuida con cuatro partes, las mismas que se localizan en una sola base de datos y no hay duplicación de datos.

1 ERA . ALTERNATIVA DE DISTRIBUCIÓN

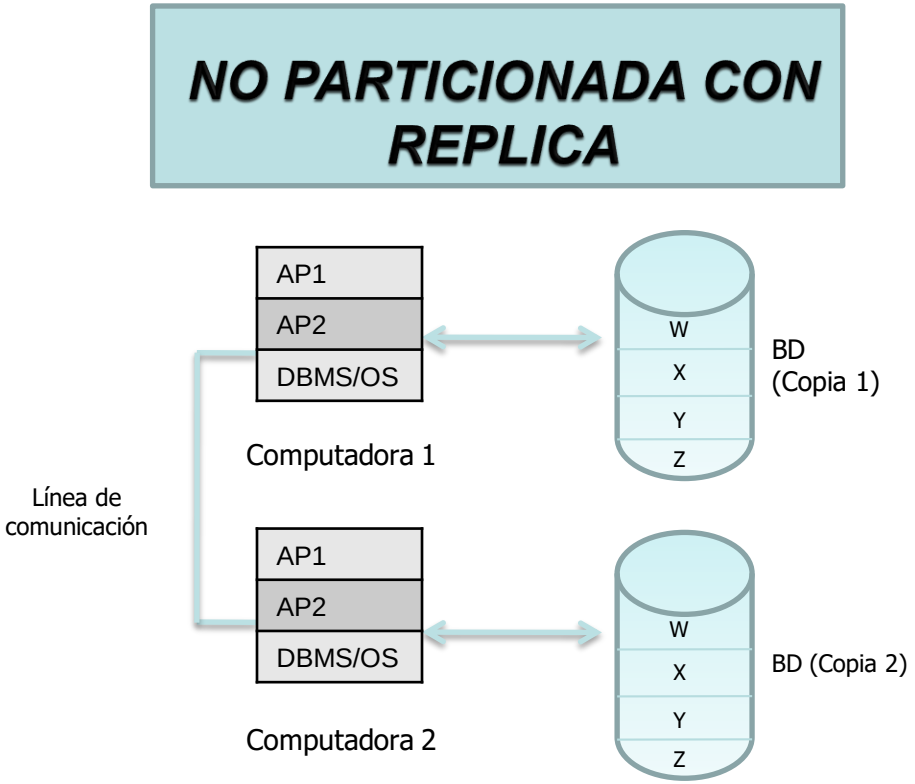


PARTICIONADA SIN REPLICA

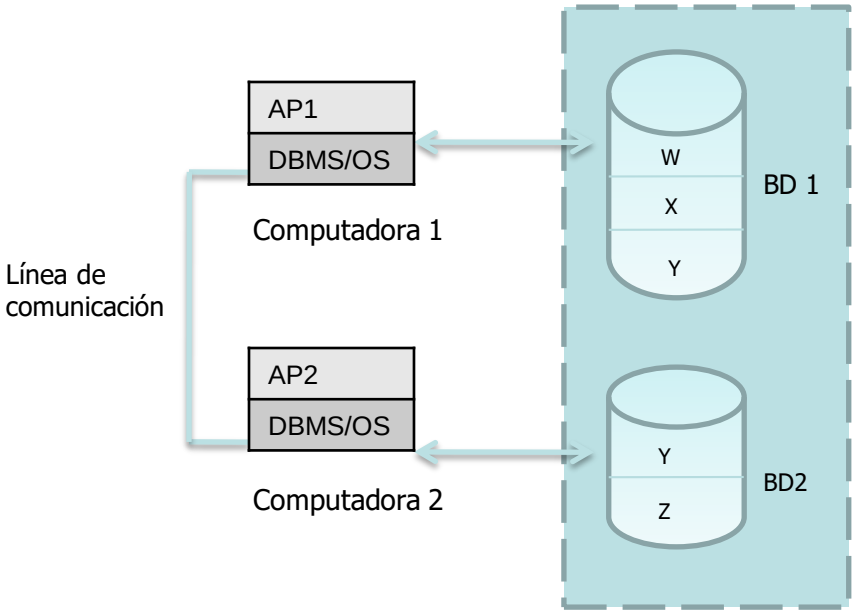
- La BD se ha distribuido a 2 computadoras
- Las piezas W y X se almacenan en la computadora 1 y las piezas Y y Z en la computadora 2

2 da. ALTERNATIVA DE DISTRIBUCIÓN

- La BD completa ha sido duplicada en dos computadoras



3 ERA . ALTERNATIVA DE DISTRIBUCIÓN



PARTICIONADA CON REPLICA

- La BD ha sido particionada y una parte (Y) tiene réplica

ESQUEMAS DE FRAGMENTACIÓN

División de los datos en varios emplazamientos

Partición Vertical: Fragmento Vertical

- Se refiere a una tabla que se divide en dos o más conjuntos de columnas.
  - Ejemplo: R(C1, C2, C3, C4)

    P1(C1, C2)
    P2(C3, C4)

    } 2 partes verticales

  La llave de R comúnmente sería colocada también en P2 para formar:

    P1(C1, C2)
    P2(C1,C3, C4)

Partición Horizontal: Fragmento horizontal

- Se refiere a los renglones de una tabla cuando están divididas en piezas.
  - Ejemplo: R(C1, C2, C3, C4)

En la relación R, si los primeros 1000 renglones se colocan en R1 (C1, C2, C3, C4) y los renglones restantes en R2(C1, C2, C3, C4). Dara como resultado 2 particiones horizontales.

Necesidad de datos distribuidos

Al crear un entorno de datos distribuido, debe diseñar una solución que consiga lo siguiente:

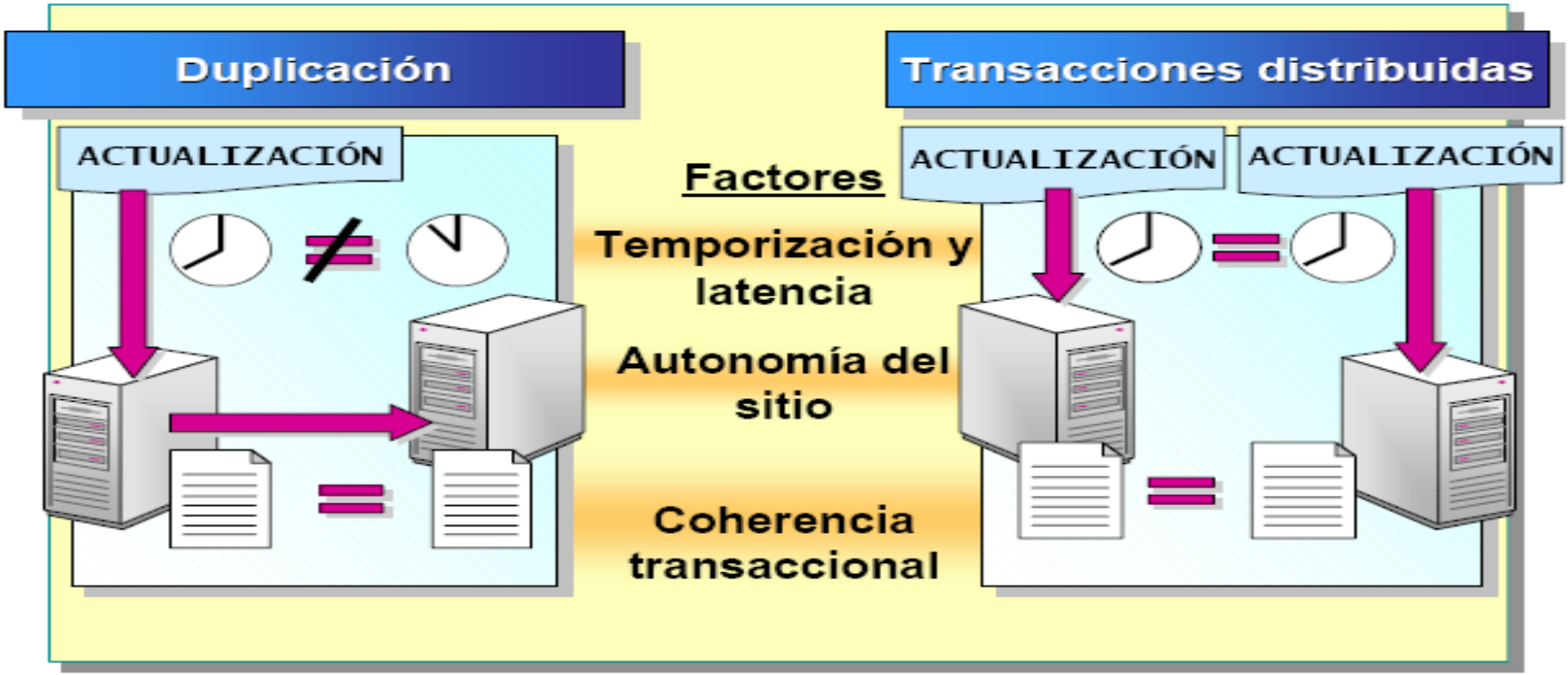
- Acercar los datos al usuario.
- Permitir que los sitios operen independientemente (de forma autónoma).
- Reducir los conflictos

Consideraciones para la distribución de datos.

Estrategias principales para la distribución de datos:

- *Duplicación*
- *Transacciones distribuidas*

- En ambas, es posible mantener copias actualizadas de los datos
- También es posible diseñar un entorno distribuido que incluya aspectos de las dos estrategias.



DUPLICACIÓN

Duplica y distribuye copias recientes de los datos desde una base de datos de origen a una de destino.

Sitios autónomos – conectados intermitentemente

TRANSACCIONES DISTRIBUIDAS

- Garantizan que las copias de los datos tienen los mismos valores en cada momento.
- Servidores conectados, capaces de confirmar su parte de la transacción.
- Solo debe usarse si los datos deben estar sincronizados en todo momento.

Factores para decidir el modo de distribuir los datos

Temporización y latencia ¿Es necesario que todas las copias de los datos sean exactamente iguales en cada instante o es aceptable un cierto grado de latencia, o retardo? Este intervalo puede variar entre unos segundos o varios días, según la estrategia de duplicación elegida.

Autonomía del sitio ¿Qué grado de autonomía (independencia) se puede permitir? Por ejemplo, una vez que haya una copia de los datos disponible localmente, ¿cuánto tiempo puede transcurrir antes de que sea necesario volver a conectarse con el servidor central para realizar una sincronización?

Coherencia transaccional Una transacción es una serie de modificaciones de datos que debe completarse o no producirse. En la duplicación hay un retardo desde la actualización inicial, pero la coherencia transaccional se mantiene.

Introducción a la replicación en SQL Server

Metáfora

La replicación utiliza una metáfora del sector editorial para representar los componentes de una topología de replicación, que incluyen el **publicador**, el **distribuidor**, los **suscriptores**, las **publicaciones**, los **artículos** y las **suscripciones**.

Resulta útil pensar en la replicación de Microsoft SQL Server como si fuera una revista:

- El publicador (editor) de una revista produce una o más publicaciones.
- Una publicación contiene artículos.
- El publicador distribuye la revista directamente o a través de un distribuidor.
- Los suscriptores reciben las publicaciones a las que se han suscrito.

publicador

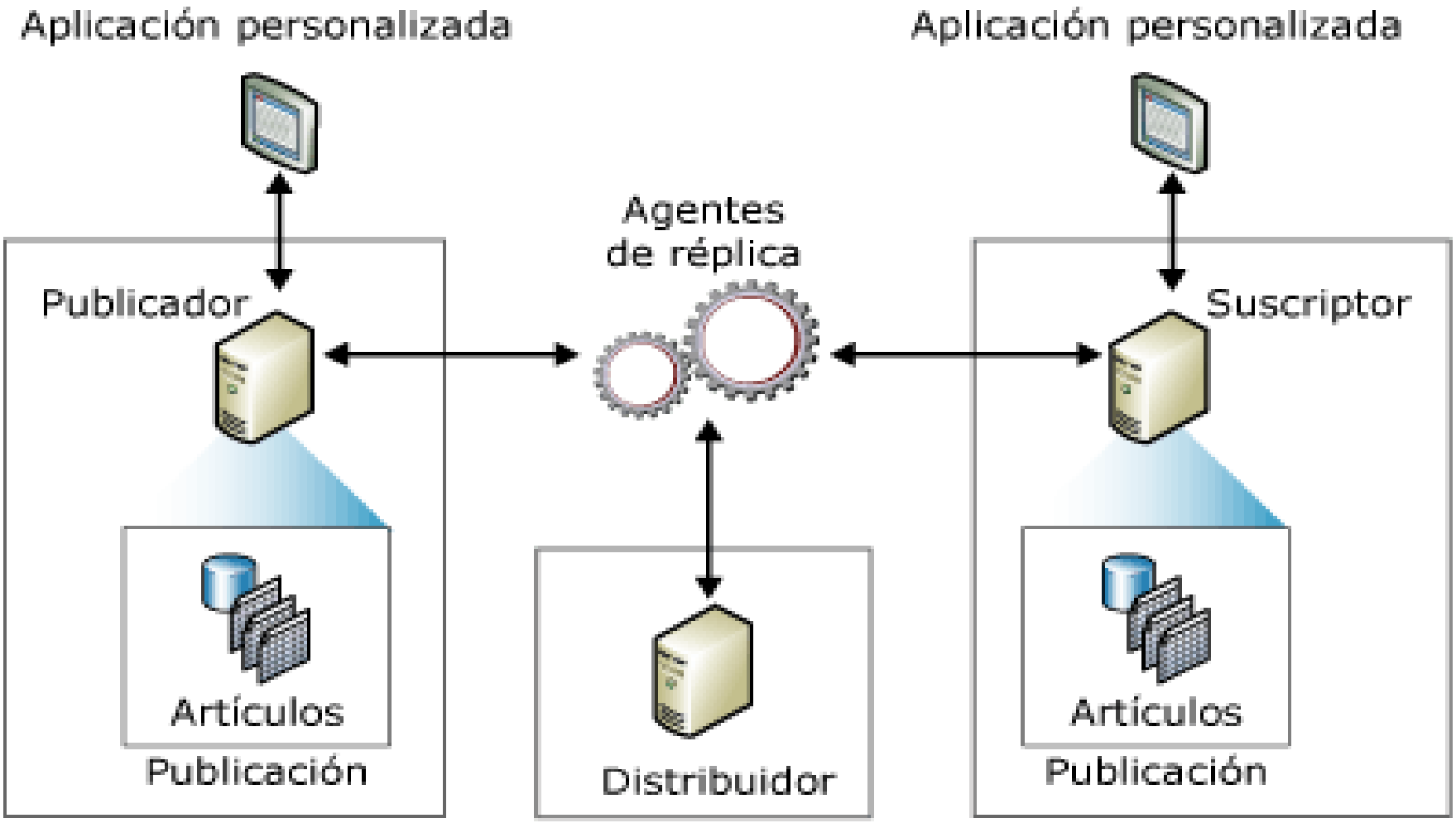
distribuidor

suscriptor

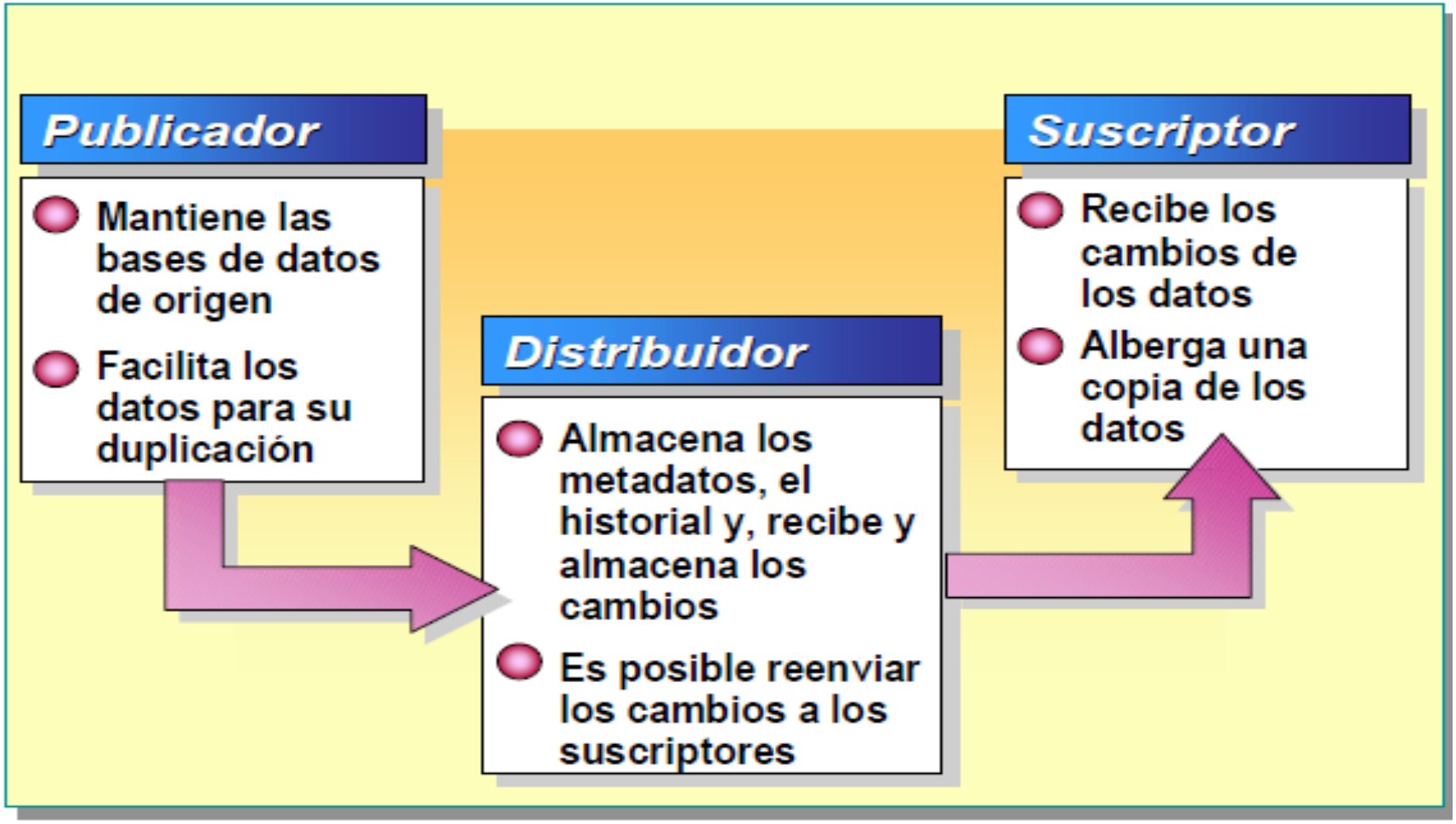
En la replicación de SQL Server incluye funciones que no están representadas en esta metáfora, en particular, ***la posibilidad de que un suscriptor realice actualizaciones y de que un publicador envíe cambios incrementales a los artículos de una publicación.***

Hay varios procesos de replicación (denominados *agentes*) que son responsables de copiar y mover los datos entre el publicador y los suscriptores.

En la siguiente ilustración se muestra información general acerca de los componentes y procesos que participan en la replicación.



El módelo del publicador y el suscriptor



- **Publicador**
  - Mantiene las bases de datos de origen, facilita los datos para duplicarlos y detecta y envía los cambios de los datos publicados al distribuidor, que puede ser el mismo servidor u otro distinto.

- **Distribuidor**
  - Es un servidor que contiene la base de datos **distribution y** almacena los metadatos, los datos de historial y las transacciones.

- **Suscriptor**
  - Alberga una copia de los datos y recibe los cambios de las publicaciones modificadas. Puede permitirse que un suscriptor actualice los datos y los vuelva a duplicar en el publicador para que se dupliquen a su vez en otros suscriptores.

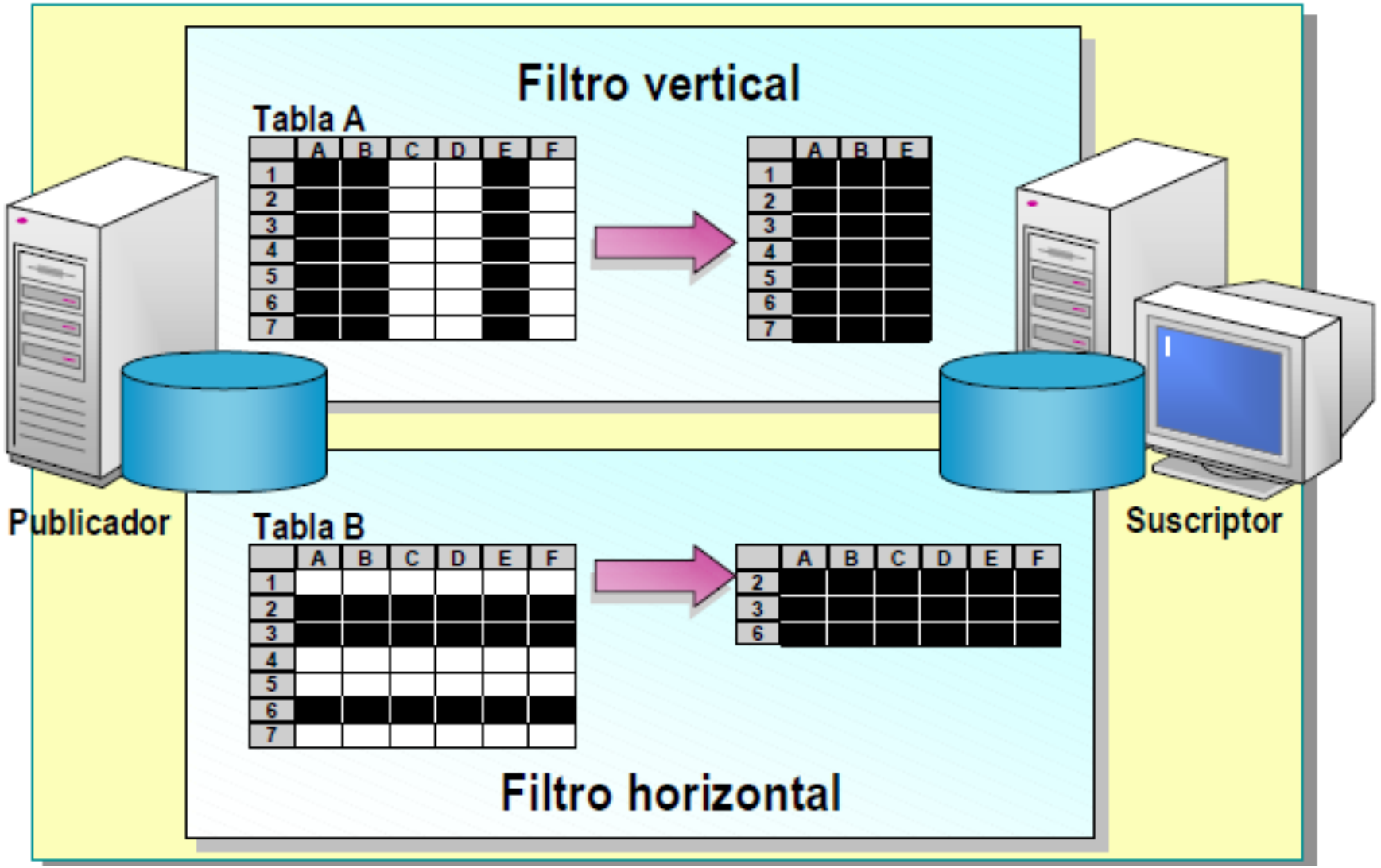
Publicación

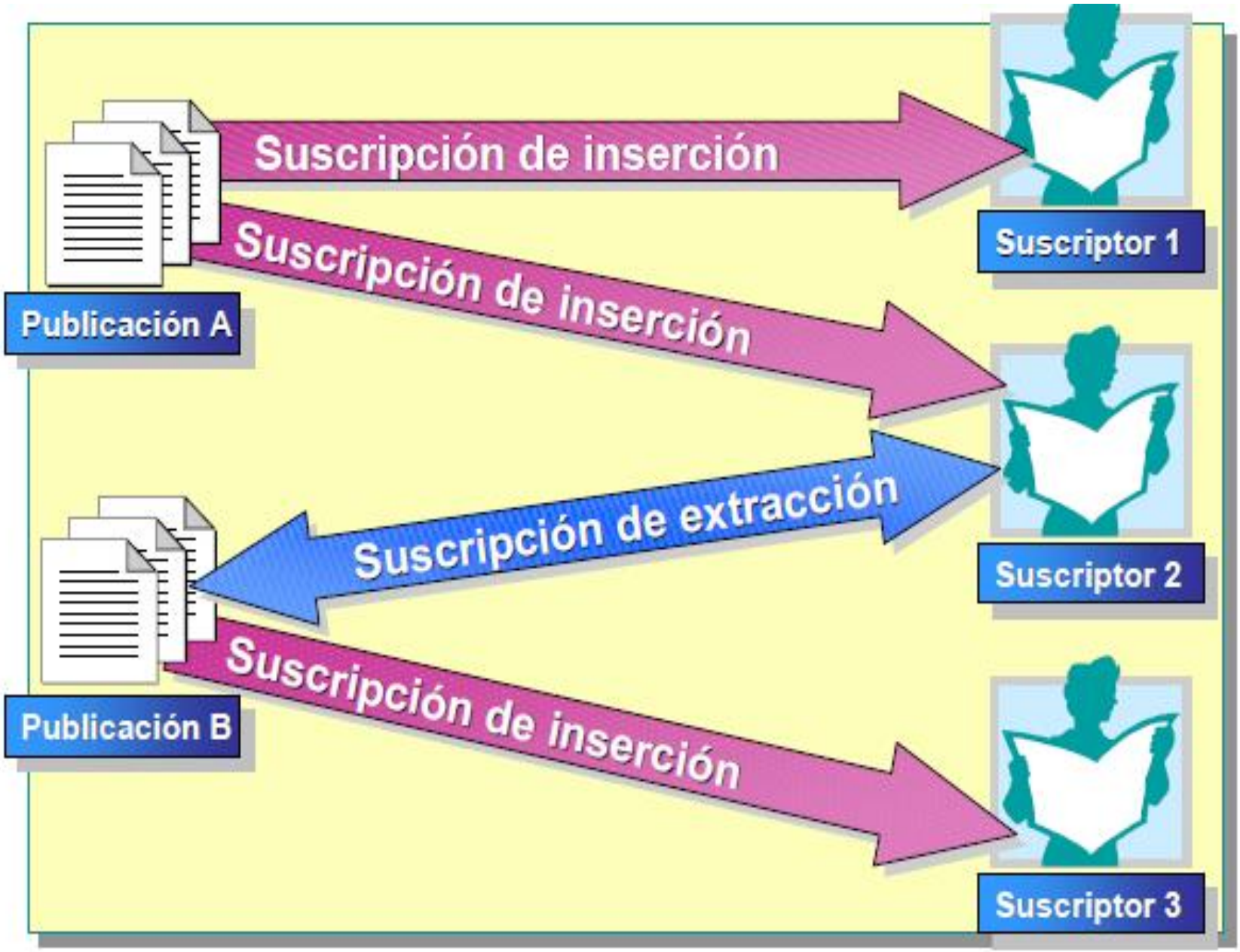
- Una colección de uno o más artículos
- La base de una suscripción
- Puede haber una o más publicaciones por base de datos

Artículo

- Parte de una publicación
- Datos globales o partición filtrada de una tabla u objeto de base de datos

Filtros de datos




Suscripción de inserción
Suscriptor 1
Publicación A
Suscripción de inserción
Suscripción de extracción
Suscriptor 2
Publicación B
Suscripción de inserción
Suscriptor 3

Suscripción	Características	Se utiliza si
Suscripción de inserción	Con una suscripción de inserción, el publicador propaga los cambios al suscriptor sin que éste lo pida. Los cambios pueden insertarse en los suscriptores a petición, de manera continua o según una programación. De forma predeterminada, el Agente de distribución o el Agente de mezcla se ejecutan en el distribuidor.	• Normalmente, los datos se sincronizarán de forma continua o con una frecuencia determinada. • Las publicaciones requieren que el movimiento de datos sea casi en tiempo real. • La sobrecarga del procesador en el distribuidor no afecta al rendimiento. • Se utiliza frecuentemente en la replicación de instantáneas y transaccional.
Suscripción de extracción	En una suscripción de extracción, el suscriptor solicita los cambios efectuados en el publicador. Las suscripciones de extracción permiten al usuario del suscriptor determinar cuándo se sincronizan los cambios en los datos. El Agente de distribución o el Agente de mezcla se ejecutan en el suscriptor.	• Los datos se sincronizarán, generalmente, a petición o en función de una programación, en lugar de hacerlo de forma continuada. • La publicación dispone de un gran número de suscriptores y/o la ejecución de todos los agentes en el distribuidor supone un uso demasiado intensivo de recursos. • Los suscriptores son autónomos, están desconectados o se desplazan. Los suscriptores determinan cuándo se conectarán y sincronizarán los cambios. • Se utiliza frecuentemente en la replicación de mezcla.

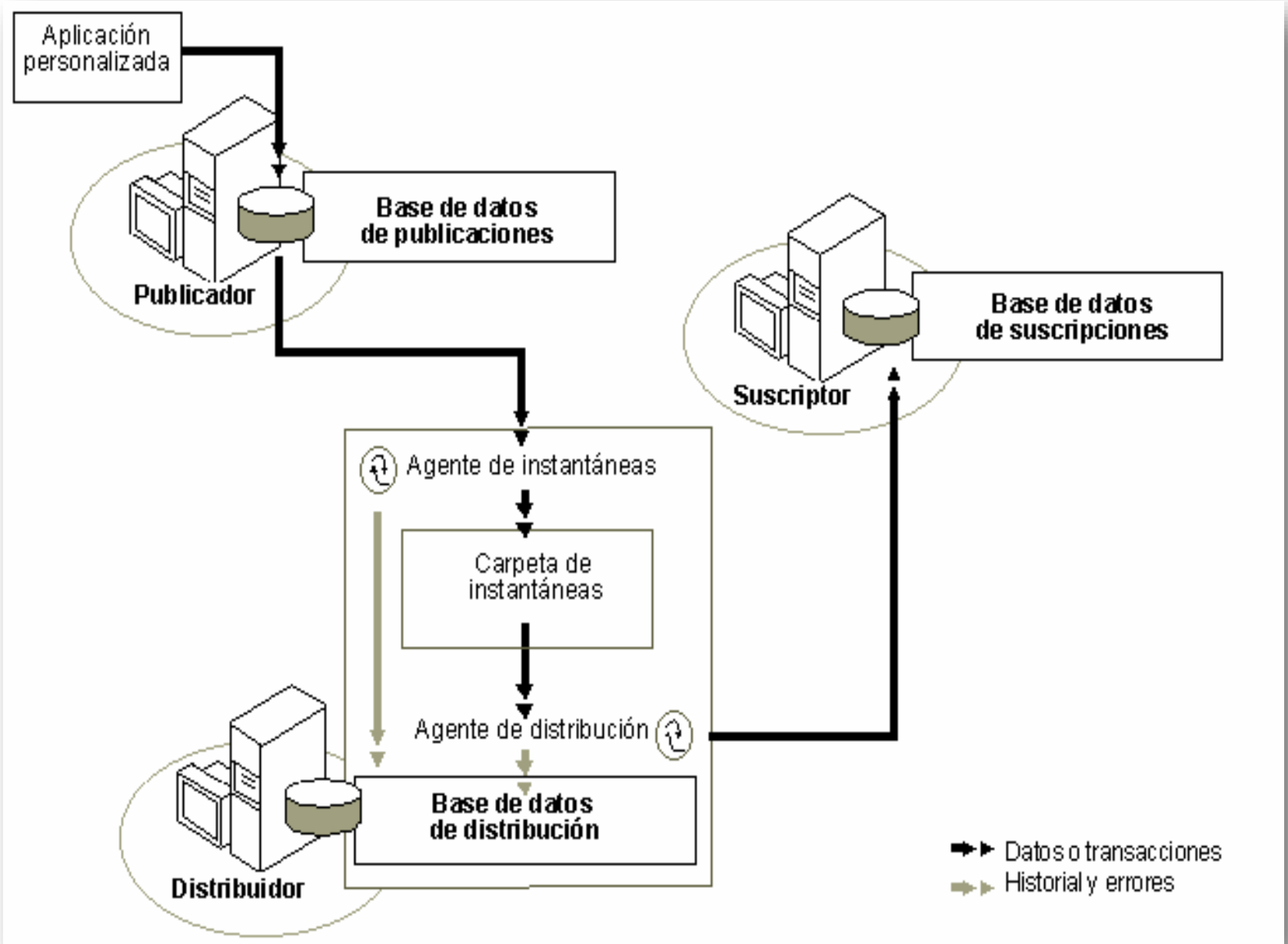
TIPOS DE REPLICACIÓN SQL SERVER

- Replicación de instantáneas (snapshot)
- Replicación transaccional
- Replicación de mezcla

Replicación de instantáneas

La replicación de instantáneas es de utilidad en los entornos:

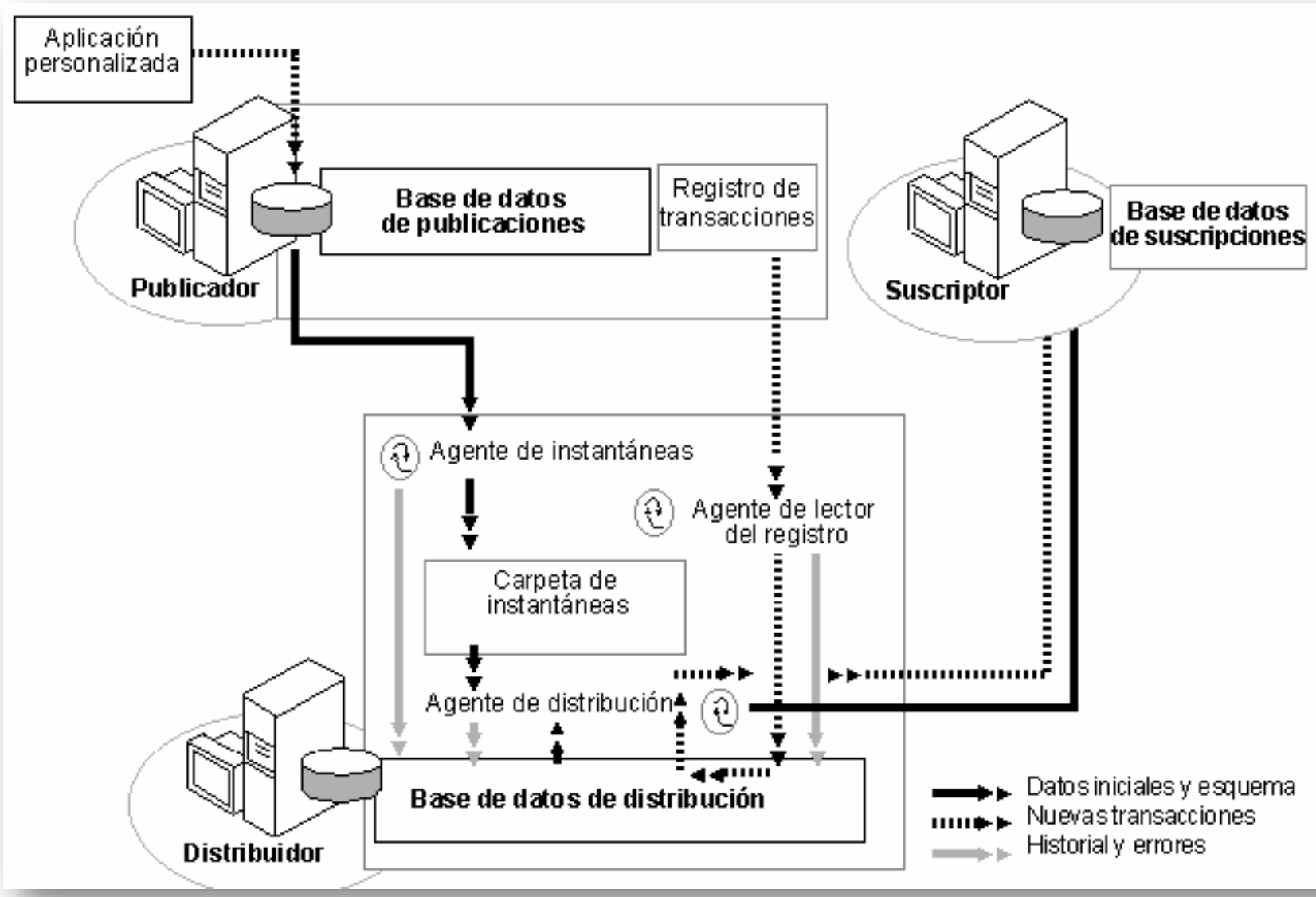
- Si los cambios son sustanciales pero poco frecuentes.
- Que requieren datos de sólo lectura en los suscriptores.
- Que admiten una latencia alta, ya que, normalmente, los datos sólo se actualizan periódicamente.
- Donde los suscriptores requieren autonomía de sitio.


Aplicación personalizada
Base de datos de publicaciones
Publicador
Base de datos de suscripciones
Suscriptor
Agente de instantáneas
Carpeta de instantáneas
Agente de distribución
Base de datos de distribución
Distribuidor
Datos o transacciones
Historial y errores

Replicación transaccional

La replicación transaccional es de utilidad en los entornos:

- Los cambios se propaguen a los suscriptores a medida que ocurren.
- Requiere una latencia baja entre el momento en que se realizan los cambios en el publicador y los cambios llegan al suscriptor.
- El publicador o suscriptor es una base de datos que no es de SQL Server, como Oracle

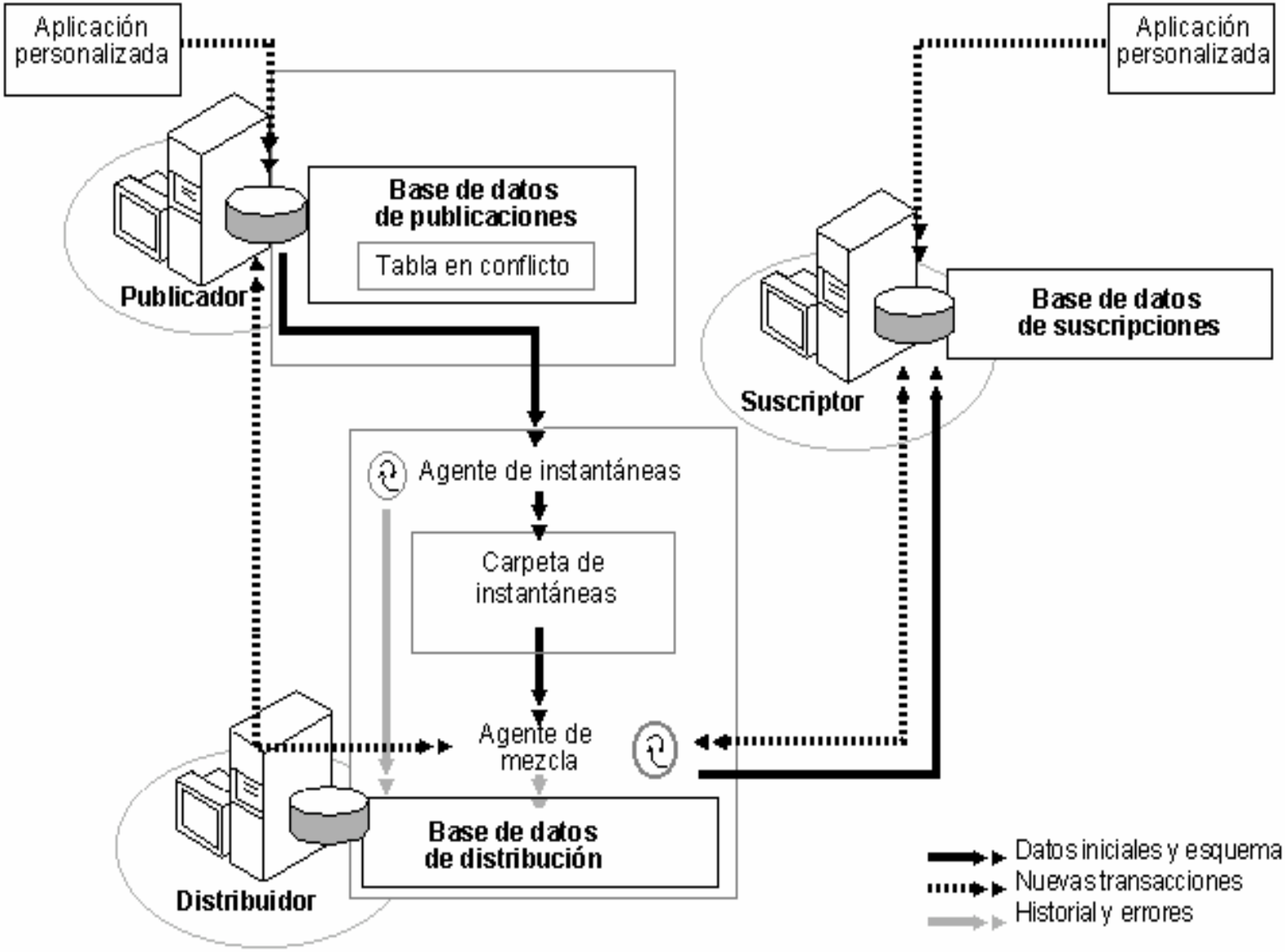
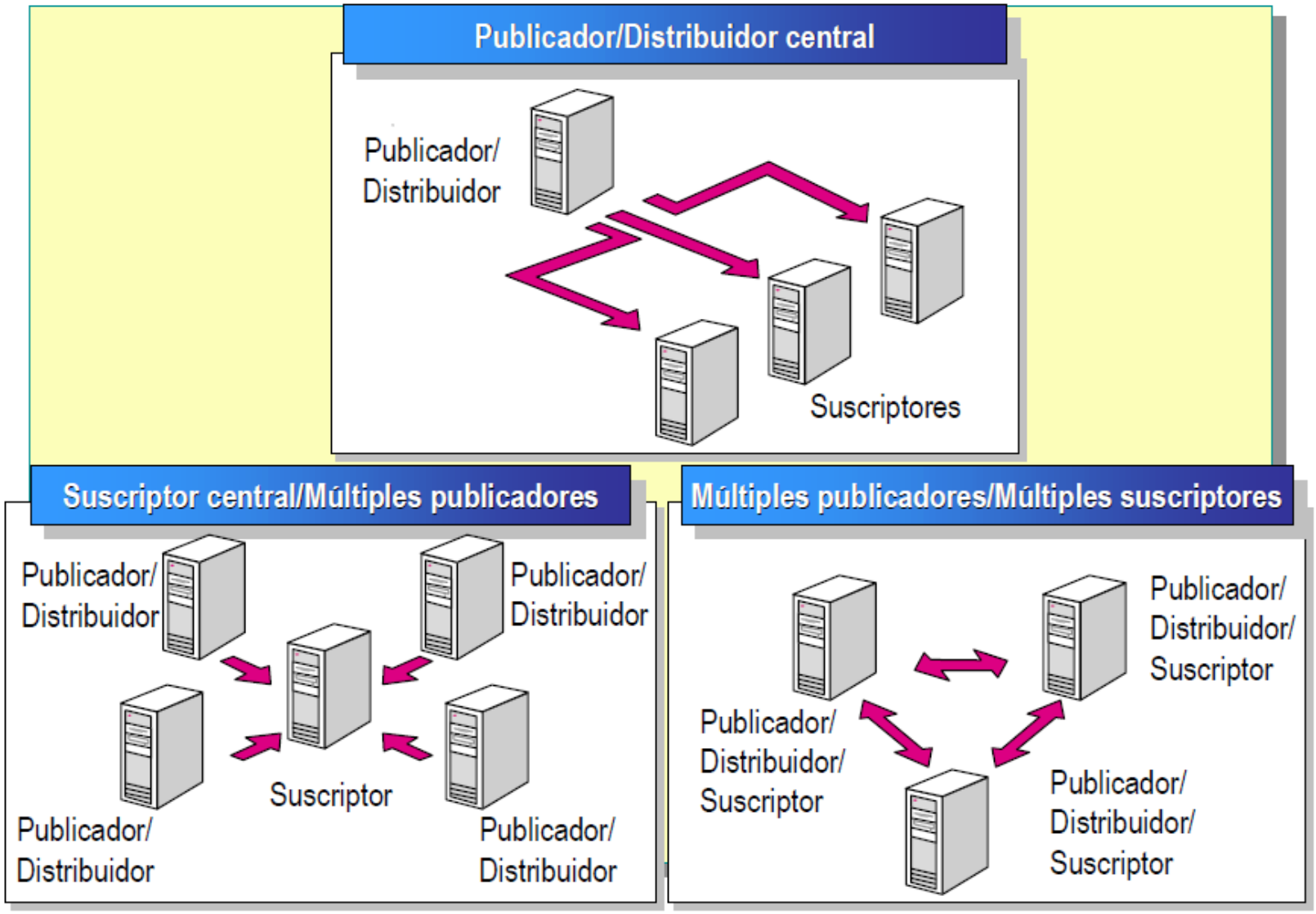

Aplicación personalizada
Base de datos de publicaciones
Registro de transacciones
Publicador
Base de datos de suscripciones
Suscriptor
Agente de instantáneas
Agente de lector del registro
Carpeta de instantáneas
Agente de distribución
Base de datos de distribución
Distribuidor
Datos iniciales y esquema
Nuevas transacciones
Historial y errores

Replicación de mezcla

La replicación de mezcla es apropiada en las siguientes situaciones:

- Varios suscriptores pueden actualizar los datos en distintos momentos y propagar dichos cambios al publicador y a otros suscriptores.

- Los suscriptores deben recibir los datos, realizar cambios sin conexión y sincronizar posteriormente los cambios con el publicador y otros suscriptores.


Aplicación personalizada
Aplicación personalizada
Base de datos de publicaciones
Tabla en conflicto
Publicador
Base de datos de suscripciones
Suscriptor
Agente de instantáneas
Carpeta de instantáneas
Agente de mezcla
Base de datos de distribución
Distribuidor
Datos iniciales y esquema
Nuevas transacciones
Historial y errores

Introducción a los modelos de duplicación físicos