# Seguridad y Privacidad en Bases de Datos – Material complementario

## 1. Identificación de Amenazas a la Seguridad de los Datos

Antes de proteger los datos, es vital comprender contra qué los estamos protegiendo. Las amenazas no solo provienen de atacantes externos, sino también de errores internos y configuraciones deficientes.

- **Ataques de Inyección SQL:** Esta es una de las amenazas más comunes y peligrosas. Un atacante inserta código malicioso en una consulta SQL a través de un formulario de entrada de una aplicación web. Si la base de datos no está configurada para sanitizar o validar estas entradas, el código malicioso se ejecuta, permitiendo al atacante:
    - **Robar información confidencial:** Obtener nombres de usuario, contraseñas, datos personales, etc.
    - **Modificar o eliminar datos:** Realizar cambios no autorizados en la base de datos.
    - **Tomar el control de la base de datos:** En casos severos, obtener acceso de administrador al sistema.
- **Abuso de Privilegios y Accesos No Autorizados:** Esta amenaza se da cuando un usuario legítimo (o un atacante que ha obtenido sus credenciales) utiliza privilegios de acceso de forma indebida. Por ejemplo, un empleado con acceso a la base de datos de salarios podría intentar ver o modificar el salario de un compañero. Esto subraya la importancia de otorgar privilegios solo cuando son estrictamente necesarios para el rol del usuario.
- **Malware (Software Malicioso):** Programas como virus, troyanos o ransomware pueden infectar el servidor de la base de datos. El ransomware, en particular, puede cifrar los datos, dejándolos inaccesibles a menos que se pague un rescate. El malware puede robar datos, corromper archivos o usar el servidor para lanzar otros ataques.
- **Pérdida de Datos:** Las amenazas no siempre son maliciosas. La pérdida de datos puede ocurrir por:
    - **Errores humanos:** Un administrador que ejecuta accidentalmente un comando DELETE sin una cláusula WHERE.
    - **Fallos de hardware:** Un disco duro que deja de funcionar o un servidor que colapsa.
    - **Desastres naturales:** Inundaciones o incendios que destruyen la infraestructura física.

## 2. Medidas de Protección y Seguridad

Las bases de datos deben ser protegidas con un enfoque de defensa en profundidad, combinando múltiples capas de seguridad.

- **Autenticación y Autorización:**
    - **Autenticación:** Es el proceso de verificar la identidad de un usuario. Esto se logra mediante credenciales (usuario y contraseña). Para fortalecerla, se recomienda el uso de contraseñas fuertes y complejas, y la autenticación de múltiples factores (MFA).
    - **Autorización:** Una vez autenticado, este mecanismo define qué puede hacer el usuario. Se basa en el principio del **"menor privilegio"**: cada usuario o rol debe tener solo los permisos mínimos indispensables para realizar sus funciones. Por ejemplo, un empleado de ventas no necesita permiso para ver la tabla de recursos humanos. Los comandos GRANT y REVOKE en SQL son esenciales para gestionar estos privilegios.
- **Cifrado de Datos:**
    - **Datos en reposo (at rest):** Cifrar los datos almacenados en el disco duro del servidor de la base de datos. Esto protege la información incluso si un atacante logra robar los medios de almacenamiento físicos.
    - **Datos en tránsito (in transit):** Cifrar los datos mientras viajan por la red, por ejemplo, entre la aplicación del usuario y el servidor de la base de datos. Esto previene que un atacante intercepte y lea la información.
- **Auditoría y Monitoreo:**
    - **Auditoría:** Consiste en registrar y revisar todas las actividades importantes en la base de datos, como los inicios de sesión, los cambios de privilegios y las operaciones sobre datos sensibles. Un registro de auditoría detallado es fundamental para detectar actividades sospechosas y, en caso de una brecha, determinar qué ocurrió y cómo.
    - **Monitoreo:** El monitoreo en tiempo real de la actividad de la base de datos permite detectar patrones anómalos o comportamientos que sugieran un ataque en curso.
- **Políticas de Respaldo y Recuperación (Backups):** Las copias de seguridad regulares son la última línea de defensa contra la pérdida de datos. Una política de respaldo bien definida debe incluir:
    - La frecuencia de los respaldos.
    - El lugar donde se almacenan los respaldos (idealmente fuera del sitio).

- Pruebas periódicas de recuperación para asegurar que los datos pueden ser restaurados correctamente.

## 3. Normativas Legales sobre Privacidad

El manejo de datos personales está regulado por leyes que buscan proteger los derechos de las personas. Ignorar estas normativas puede llevar a multas cuantiosas y a la pérdida de confianza de los usuarios.

- **Ley 25.326 de Protección de Datos Personales (Argentina):** Esta es la principal normativa que regula el tratamiento de datos personales en Argentina. Establece principios fundamentales como:
  - **Consentimiento:** Los datos personales solo pueden ser recolectados si el titular da su consentimiento expreso.
  - **Derecho de Acceso, Rectificación, Cancelación y Oposición (ARCO):** Las personas tienen el derecho de acceder a sus datos, corregirlos si son incorrectos, solicitar que se cancelen (eliminen) y oponerse a su uso.
  - **Finalidad:** Los datos deben ser recolectados para un fin específico y no pueden ser usados para otros propósitos sin un nuevo consentimiento.
  - **Seguridad:** Quienes manejan datos personales tienen la obligación de implementar medidas de seguridad adecuadas para proteger la información.
- **Reglamento General de Protección de Datos (GDPR) de la Unión Europea:** Aunque es una normativa europea, ha tenido un impacto global y muchas empresas la utilizan como estándar. Se basa en principios similares a la Ley 25.326, pero es más estricta en cuanto a las multas y la responsabilidad de las empresas. El GDPR es un excelente ejemplo de cómo las regulaciones modernas exigen una gestión proactiva de la privacidad.

Comprender estas leyes no es solo una cuestión de cumplimiento legal, sino un pilar fundamental para construir sistemas de bases de datos éticos y confiables.

## 4. Cifrado Homomórfico

El **cifrado homomórfico** es una forma avanzada de criptografía que permite realizar operaciones matemáticas directamente sobre datos cifrados sin necesidad de descifrarlos previamente. Una vez completadas las operaciones, el resultado, que también está cifrado, al ser descifrado, es idéntico al resultado que se hubiera obtenido al operar sobre los datos originales en texto plano.

- **Principio Fundamental:** La clave está en la "homomorfía", que en matemáticas se refiere a una propiedad en la que la estructura de una operación se preserva después de una transformación. En este caso, la operación (por ejemplo, una suma) sobre el texto cifrado (cifrado(A) + cifrado(B)) corresponde a la misma operación sobre el texto plano subyacente (A + B).

- **Diferencia con el Cifrado Tradicional:** Mientras que el cifrado tradicional protege los datos **en reposo** (almacenados) y **en tránsito** (moviéndose por la red), el cifrado homomórfico extiende esta protección a los datos **en uso** (durante el procesamiento). En un sistema tradicional, los datos deben ser descifrados en la memoria del servidor para poder ser procesados, creando un momento de vulnerabilidad. El cifrado homomórfico elimina esta ventana de riesgo.

- **Ventajas y Desventajas:**

    - **Ventajas:** Permite a servicios externos (como la nube) realizar cálculos con datos sensibles sin poder ver el contenido real. Esto es crucial para la privacidad en el análisis de grandes datos, la investigación médica colaborativa y la minería de datos. Reduce significativamente el riesgo de filtraciones.

    - **Desventajas:** La principal limitación es su **eficiencia computacional**. Las operaciones sobre datos cifrados son significativamente más lentas y consumen muchos más recursos que las mismas operaciones sobre datos en texto plano.

- **Aplicaciones Prácticas:**

    - **Análisis de datos en la nube:** Un hospital podría enviar datos genéticos cifrados a un proveedor de nube para que se realicen análisis sin que el proveedor pueda ver la información del paciente.

    - **Aprendizaje automático (Machine Learning) con privacidad:** Permite entrenar modelos de IA utilizando datos cifrados de múltiples fuentes, garantizando que ninguna parte pueda ver los datos de los demás.

    - **Sistemas de votación electrónica:** Los votos podrían ser cifrados homomórficamente para permitir un recuento seguro sin que nadie, ni siquiera la autoridad electoral, pueda ver los votos individuales.

## 5. Auditoría y Monitoreo de Seguridad

La **auditoría** y el **monitoreo** son herramientas esenciales para la seguridad reactiva y proactiva de una base de datos. Mientras que los controles de acceso son preventivos, la auditoría y el monitoreo permiten detectar incidentes de seguridad y actividades sospechosas.

- **Propósito y Eventos a Monitorear:**

    - **Detección de Amenazas:** Sirven para identificar comportamientos anómalos o maliciosos en tiempo real.

    - **Análisis Forense:** En caso de una brecha de seguridad, los registros de auditoría son fundamentales para determinar qué sucedió, cuándo, y cómo fue explotada una vulnerabilidad.

- **Cumplimiento Legal:** Proveen la evidencia necesaria para demostrar el cumplimiento de normativas de privacidad.
- **Eventos clave a monitorear:**
  - **Intentos de inicio de sesión fallidos:** Podrían indicar ataques de fuerza bruta.
  - **Cambios en los permisos de usuario:** Señalan un posible escalamiento de privilegios.
  - **Accesos a tablas con datos sensibles:** Actividad inusual en tablas que contienen información personal o financiera.
  - **Ejecución de sentencias DDL:** Creación o modificación de la estructura de la base de datos.

- **Herramientas SIEM:** Un **Sistema de Gestión de Información y Eventos de Seguridad (SIEM)** es una plataforma que centraliza y analiza los registros de auditoría de múltiples fuentes (sistemas operativos, bases de datos, firewalls, etc.). Los SIEM son cruciales para el monitoreo porque:
  - **Correlacionan eventos:** Pueden identificar patrones complejos que un humano podría pasar por alto, como un intento de inicio de sesión fallido seguido de un acceso a una tabla crítica desde una IP inusual.
  - **Generan alertas:** Notifican a los administradores de seguridad en tiempo real sobre posibles amenazas.
  - **Simplifican el cumplimiento:** Permiten generar informes automatizados para demostrar el cumplimiento con normativas como la Ley 25.326 o el GDPR.

## 6. Estrategias de Respaldo y Recuperación

Las copias de seguridad (backups) no son solo una medida de seguridad, sino una parte esencial de un plan de continuidad de negocio y recuperación ante desastres.

- **Tipos de Respaldo:**
  - **Respaldo Completo (Full Backup):** Copia la totalidad de los datos en un momento dado. Es el más simple de restaurar, pero el que más espacio consume y el más lento de realizar.
  - **Respaldo Diferencial (Differential Backup):** Solo copia los datos que han cambiado desde el **último respaldo completo**. La restauración requiere el último respaldo completo y el último respaldo diferencial.
  - **Respaldo Incremental (Incremental Backup):** Solo copia los datos que han cambiado desde el **último respaldo de cualquier tipo** (completo o incremental). La restauración es la más compleja, ya que requiere el respaldo

completo y todos los incrementales subsiguientes. Es el más rápido y consume menos espacio.

- **Respaldo del Registro de Transacciones (Transaction Log Backup):** Utilizado en bases de datos que registran cada operación. Permite una recuperación muy precisa hasta un punto exacto en el tiempo, minimizando la pérdida de datos.

- **Regla 3-2-1:** Esta es una estrategia de respaldo fundamental que garantiza la redundancia y resiliencia de los datos.
  - **3 copias:** Mantén al menos **tres** copias de tus datos: el original y dos respaldos.
  - **2 soportes:** Guarda los respaldos en al menos **dos** tipos de medios de almacenamiento diferentes (por ejemplo, un disco duro local y una cinta, o un disco duro y la nube).
  - **1 copia externa:** Almacena al menos **una** copia de seguridad fuera de las instalaciones físicas (off-site), para protegerte contra desastres locales como incendios, robos o inundaciones.
- **RPO y RTO (Objetivos de Recuperación):** Estos dos conceptos definen los objetivos de un plan de recuperación ante desastres.
  - **RPO (Recovery Point Objective - Objetivo de Punto de Recuperación):** Es la cantidad máxima de datos que una organización está dispuesta a perder en caso de un incidente. Se mide en tiempo (ej. 1 hora, 15 minutos). Un RPO de 1 hora significa que las copias de seguridad deben realizarse al menos cada hora. Está directamente relacionado con la **frecuencia de los respaldos**.
  - **RTO (Recovery Time Objective - Objetivo de Tiempo de Recuperación):** Es el tiempo máximo que un sistema puede estar inactivo después de un fallo. Se mide en tiempo (ej. 2 horas). Un RTO de 2 horas significa que el sistema debe estar completamente restaurado y operativo en menos de dos horas. Está directamente relacionado con la **velocidad de la restauración** y la eficiencia del plan de recuperación.