



## Caso de estudio

**MTI JESÚS E. ROMERO MORENO**

**DIVISIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN Y  
COMUNICACIÓN**

## Caso de estudio Nivel 1 administración de base de datos

### Perfil de la Empresa:

Nombre: DataSolutions Inc.

Industria: Tecnología y Servicios de Datos

Tamaño: 150 empleados

Sucursales: 3 (en diferentes ciudades)

Servidores: 4 servidores principales centralizados por VPN

Bases de Datos: 2 bases de datos principales de 1 TB y 1.8 TB

Consumo de Datos: Varía entre 5 TB durante horas pico y 2 TB en horas de menor actividad

### Características de los Servidores:

Procesadores: Quad-Core, 3.2 GHz

RAM: 32 GB por servidor

Tamaño de Disco Duro: 2 TB por servidor

Velocidad de la Red: 1 Gbps

Destino: Dos para bases de datos, uno para aplicaciones, uno para respaldo y recuperación

### Base de Datos:

Tipo: Relacional

Sistema de Gestión: PostgreSQL

Uso: Gestión de datos de clientes, transacciones y análisis de datos

### Esquema de Respaldo Actual:

Frecuencia: Respaldo completo semanal, respaldo incremental diario, durante el día de manera local.

### Fluctuaciones en el Consumo de Datos:

Mayor consumo durante horas laborales (9 am - 5 pm)

Menor consumo en horas no laborales y fines de semana

Incremento del 30% en el consumo al final de cada mes por reportes y análisis de datos, y crecimiento de la DB en 7% mensual.

### Problemas y Desafíos:

Gestión eficiente de la carga de trabajo durante las fluctuaciones

Mantenimiento de la integridad y seguridad de los datos

Optimización del rendimiento de la base de datos

Prever crecimiento anual

### Objetivo del Caso de Estudio:

Desarrollar un plan estratégico hipotético para mejorar la administración de la base de datos, enfocado en la optimización del rendimiento durante las fluctuaciones de carga de trabajo, manteniendo la seguridad de los datos y preparando la infraestructura para el crecimiento futuro.

# 1. Diagnóstico Inicial

DataSolutions Inc. es una empresa del sector de tecnología y servicios de datos, con una estructura operativa compuesta por 150 empleados distribuidos en tres sucursales ubicadas en diferentes ciudades. La infraestructura tecnológica actual se basa en un modelo centralizado mediante una red privada virtual (VPN), que conecta a todas las sucursales con 4 servidores principales ubicados en un nodo central. A manera de resumen la empresa actualmente opera con los siguientes recursos:

Recurso	Especificación
Servidores	4 (2 para BD, 1 para aplicaciones, 1 para respaldo)
Procesadores	Quad-Core 3.2 GHz
Memoria RAM	32 GB por servidor
Almacenamiento	2 TB por servidor
Red	1 Gbps
Base de Datos	2 (1 TB y 1.8 TB)
SGBD	PostgreSQL
Crecimiento	7% mensual
Incremento	30 % al final de cada mes
Tráfico	5 Tb en horas pico, 2 TB en horas bajas

La compañía opera con dos bases de datos principales, cuyas capacidades alcanzan 1 TB y 1.8 TB respectivamente. Estas bases de datos son críticas para el funcionamiento de los servicios y se encuentran alojadas en los servidores centrales. El consumo de datos varía significativamente, oscilando entre 5 TB durante las horas pico y 2 TB en periodos de baja actividad, lo que sugiere una alta demanda transaccional y de procesamiento en ciertos momentos del día.

Se detectó que la empresa atraviesa por los siguientes desafíos:

- Gestión eficiente de la carga de trabajo durante las fluctuaciones.
- Mantenimiento de la integridad y seguridad de los datos.
- Optimización del rendimiento de las bases de datos.
- Previsión y planificación del crecimiento anual.

## 2. Objetivos

1. Mejorar el rendimiento en horas pico.
2. Garantizar integridad y seguridad de los datos.
3. Planificar escalabilidad futura de almacenamiento y procesamiento.
4. Automatizar procesos críticos como respaldos y mantenimiento.

### 3. Propuesta Hipotética

#### 3.1 Proyección anual de almacenamiento y tráfico de datos

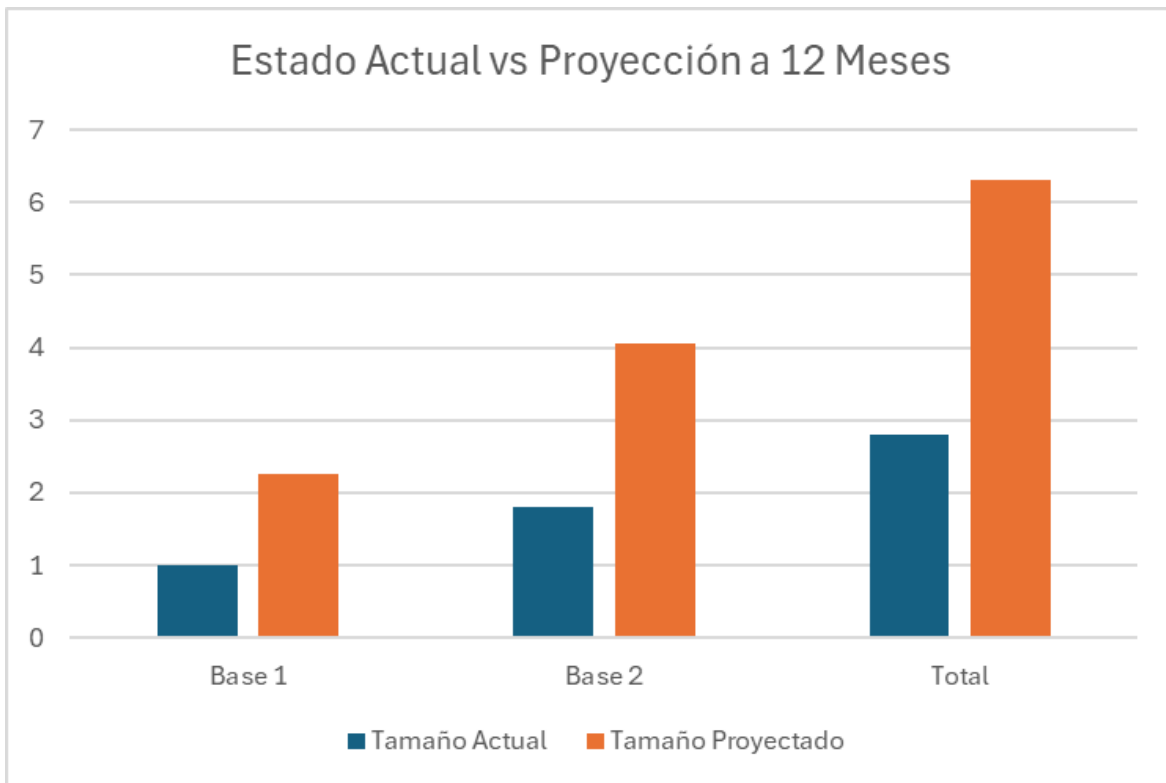
##### Proyección del Almacenamiento de Bases de Datos

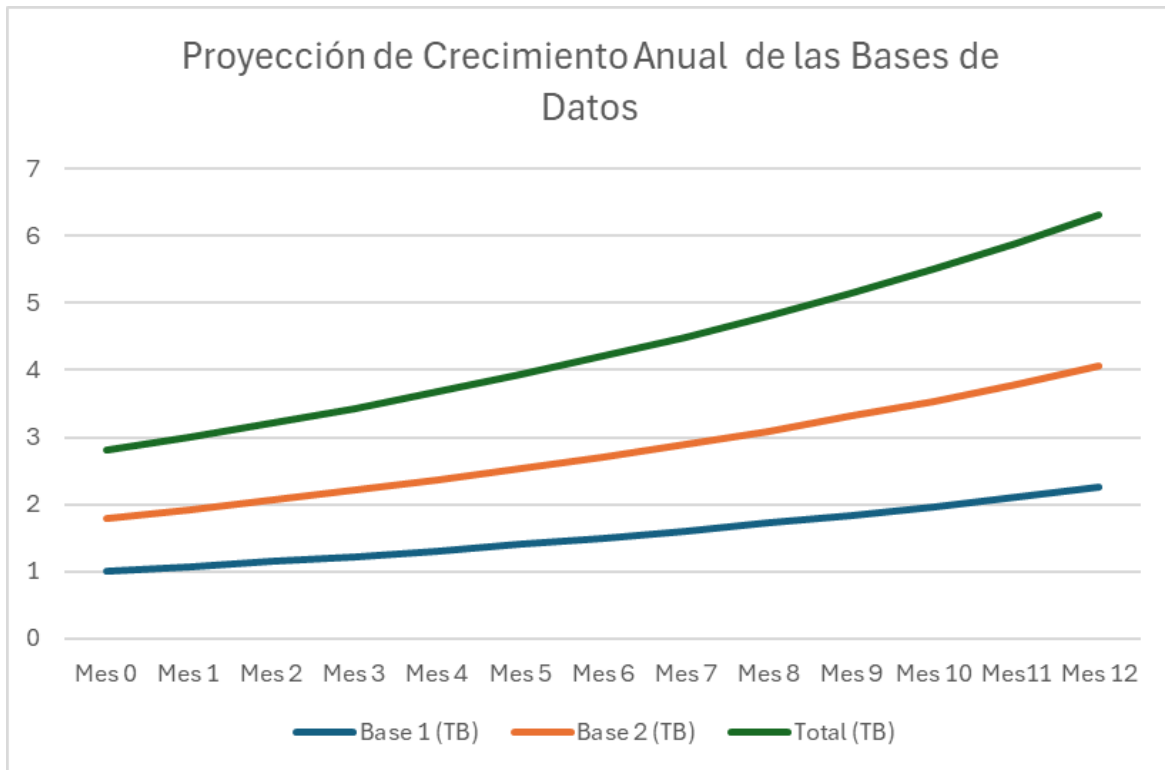
Actualmente, DataSolutions Inc., administra dos bases de datos relacionales en PostgreSQL con un tamaño combinado de 2.8 TB. Con base en el crecimiento mensual reportado del 7%, se proyecta un crecimiento compuesto durante 12 meses utilizando la fórmula de interés compuesto:

$$\text{Tamaño Futuro} = \text{Tamaño Actual} \cdot (1 + 0.07)^{12}$$

Base de Datos	Tamaño Actual	Tamaño Proyectado
Base 1	1.0 TB	2.25 TB
Base 2	1.8 TB	4.05 TB
Total	2.8 TB	6.30 TB

Se estima que el almacenamiento total requerido para bases de datos crecerá a más del doble en un año.





### Situación Actual del Tráfico de Datos

El tráfico de datos se distribuye entre horas pico (9 am a 5 pm), horas normales (resto del día). Los volúmenes de datos procesados actualmente son:

Periodo	Carga Total (Ct)	Duración	GB/hora	MB/segundo
Horas Pico	5 TB	8 horas	640 GB	182.04 MB/s
Horas Normales	2 TB	16 horas	128 GB	36.40 MB/s
Total	7 TB	24 horas		

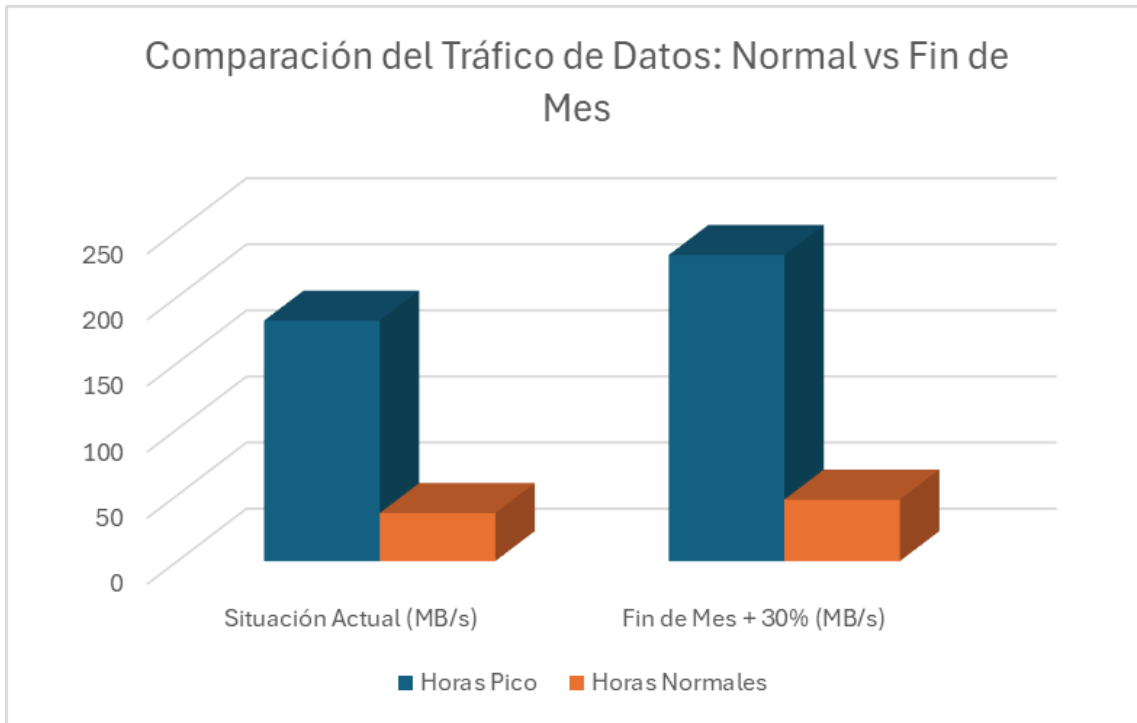
### Aumento del Tráfico al Final de Cada Mes

Durante los últimos días hábiles de cada mes, se genera un incremento aproximado del 30% en la carga de trabajo, debido a procesos de análisis y generación de reportes mensuales. Por lo tanto: se aplica el siguiente calculo:

$$C \cdot 1.30 = C \text{ con aumento del } 30\%$$

Periodo	Carga Ajustada	Duración	GB/hora	MB/segundo
Horas Pico	6.5 TB	8 horas	832 GB	231.65 MB/s
Horas Normales	2.6 TB	16 horas	162.5 GB	46.57 MB/s
Total	9.1 TB	24 horas		

Estos valores representan el escenario en donde se tiene una mayor exigencia en los servidores.



#### Proyección Anual del Consumo de Datos

Siguiendo el mismo criterio de crecimiento del almacenamiento, se estima que el volumen de datos transferido diariamente también aumentará en proporción. Aplicando el mismo factor de crecimiento compuesto ( $\times 2.25$ ):

Tráfico Base	Carga Diaria	Proyección Anual
Horas Pico	5 TB	11.26 TB
Horas Normales	2 TB	4.50 TB
Total	7 TB	15.76 TB

## 3.2 Gestión eficiente de la carga de trabajo durante fluctuaciones

### Datos del problema:

- Horas pico (9 a.m. - 5 p.m.): alto consumo (hasta 5 TB)
- Fin de mes: incremento del 30% adicional
- Infraestructura actual sin balanceo explícito de carga

### Posibles soluciones:

- Roles del Equipo de DBA:

Rol	Responsabilidades Clave
DBA Senior	Diseño de arquitectura, políticas de seguridad y acceso, auditorías, gestión de escalabilidad, decisiones estratégicas.
DBA de Operaciones	Monitoreo de rendimiento, ajustes de consultas (query tuning), mantenimiento de índices, gestión de tareas programadas, ejecución de respaldos.
DBA de Soporte Nocturno	Monitoreo preventivo, verificación de backups, tareas de mantenimiento profundo, pruebas de recuperación, generación de reportes nocturnos.

- Asignación de Turnos (Rotativos por semana):

Actualmente, hay 8 horas de trabajo (en horas pico) y 16 horas en que no se realizan operaciones en la empresa (Horas no laborales y fines de semana), por lo que se sugiere la asignación de turnos para aprovechar eficazmente las horas muertas.

Turno	Horario	Enfocado en
Turno A (Pico)	9:00 am - 5:00 pm	Supervisión activa, consultas críticas, apoyo a áreas de análisis.
Turno B (Tarde)	5:00 pm - 1:00 am	Ejecución de respaldos, rotación de logs, mantenimiento de rutina.
Turno C (Nocturno)	1:00 am - 9:00 am	Tareas pesadas (Re-indexación, limpieza), verificación de integridad, preprocesamiento de reportes.

Cada turno tendrá al menos 1 DBA de forma presencial o remota con acceso seguro.

Se establecen guardias de fin de semana rotativas quincenal con disponibilidad remota y acceso cifrado vía VPN. En días de fin de mes (pico de uso), se asigna un segundo DBA de respaldo.



Actualmente se tiene 1 Gbps (125 MB/s) en la velocidad de los servidores, sin embargo, el mayor consumo proyectado es de 231.65 MB/s en la base de datos de 5 TB durante las horas pico al final del mes, lo que generaría cuello de botella, por lo que se propone la contratación de un enlace dedicado de 10 Gbps (1250 MB /s) para los servidores principales. Esto permitiría descargar completamente el procesamiento intensivo hacia las 16 horas no laborales y los fines de semana, donde se pueden programar respaldos, sincronización y replicas, procesamiento de reportes y mantenimiento general, minimizando el impacto en el horario productivo. Esta medida pospone la necesidad inmediata de un clúster.

Uno de los servidores está llegando a su límite (1.8 TB). Realizando un diagnóstico, teniendo en cuenta que el crecimiento mensual es del 7%, con la fórmula de crecimiento de interés compuesto, el disco duro estaría llegando a su límite en menos de 30 días.

$$1.8 \text{ TB} \cdot (1 + 0.07) = 1.926 \text{ TB}$$

Por lo tanto, el cambió de disco duro a una capacidad de 4 TB (de acuerdo con las proyecciones) se debe realizar de manera urgente.

### 3.3 Mantenimiento de la integridad y seguridad de los datos.

#### Posibles soluciones:

- Frecuencia y Herramientas de Respaldos:

Tipo de Respaldo	Frecuencia	Herramienta
Completo	Semanal (Domingo, 2:00 am)	pgBackRest, Barman
Incremental	Diario (2:00 am)	pgBackRest + WAL Archiving
En caliente (WAL)	Continuo	Streaming Replication + WAL archiving

- **Completo:** Es un respaldo que captura todos los datos de la base de datos.
  - Frecuencia: Se realiza semanalmente, específicamente los domingos a las 2:00 am.
  - Herramienta: Se utilizan herramientas como pgBackRest y Barman, ambas diseñadas para gestionar respaldos y restauraciones en bases de datos PostgreSQL.
- **Incremental:** Este tipo de respaldo solo captura los cambios realizados desde el último respaldo, ya sea completo o incremental.
  - Frecuencia: Se realiza diariamente a las 2:00 am.
  - Herramienta: Se usa pgBackRest en combinación con el archivo WAL (Write Ahead Log) para asegurar que todos los cambios se respalden eficientemente.
- **En caliente (WAL):** Se refiere a realizar respaldos de la base de datos mientras está en funcionamiento, sin necesidad de detenerla.
  - Frecuencia: Este es un proceso continuo, que se mantiene a lo largo del tiempo.
  - Herramienta: Utiliza replicación en streaming junto con el archivo WAL, lo que permite una copia de seguridad constante y mínima pérdida de datos.

- Rotación: Respaldos completos se almacenan en doble ubicación: local y remoto cifrado (S3, Azure Blob, etc.)

Los respaldos completos se almacenan en doble ubicación, lo que significa que hay copias de seguridad tanto en un entorno local (en la infraestructura propia de la empresa) como en un entorno remoto cifrado (servicios en la nube como S3 de Amazon o Azure Blob). La encriptación garantiza que los datos estén protegidos mientras se almacenan en la nube.

- Verificación automática: Comprobación con `pg_restore --list` y validación de checksums tras cada respaldo.

Se utiliza el comando `pg_restore --list` para listar el contenido del respaldo y después se realiza una validación de checksums. Donde los checksums son valores calculados a partir de los datos que permiten comprobar que no se han producido errores o modificaciones no deseadas después de realizar el respaldo. Esta verificación se realiza tras cada respaldo, lo que ayuda a garantizar que todas las copias de seguridad son precisas y confiables.

- Almacenamiento y Retención:

Destino	Duración de Retención
Servidor de Backups (local)	14 días
Nube cifrada (offsite)	90 días

- Cifrado: AES-256 con claves rotadas cada 30 días (criptoseguridad avanzada).

AES-256 es un algoritmo de cifrado de clave simétrica que utiliza una longitud de clave de 256 bits, considerado uno de los métodos más seguros para proteger datos.

- Todos los respaldos son versionados y tienen marcado de fecha para trazabilidad.

Respaldos versionados significa que cada copia de seguridad se guarda como una nueva versión, permitiendo a los usuarios acceder a diferentes estados de los datos en el tiempo.

Marcado de fecha para trazabilidad implica que cada respaldo incluye información sobre cuándo se realizó, lo cual es fundamental para identificar y recuperar versiones específicas de los datos si es necesario.

- Pruebas de Restauración y Recuperación:

Tipo de Prueba	Frecuencia
Restauración parcial (selectiva)	Mensual
Recuperación total (DRP)	Trimestral (simulado en staging)

Restauración parcial (selectiva): Esta prueba se realiza mensualmente y consiste en restaurar únicamente una parte específica de los datos. Es útil para verificar que los datos pueden ser recuperados de manera precisa sin necesidad de restaurar todo el sistema.

Recuperación total (DRP): Se lleva a cabo trimestralmente, y suele realizarse en un entorno de staging (pruebas). Este tipo de prueba implica simular una recuperación completa del sistema tras un fallo mayor, asegurando así que todos los datos y configuraciones pueden ser restaurados efectivamente.

- Simulación de fallos: Se incluyen escenarios como corrupción de archivos, pérdida de tabla, pérdida de red o servidor completo.
- Seguridad y Accesos:
  - Acceso restringido al rol backup\_user con permisos mínimos: Esto significa que solo las personas que tienen asignado el rol "backup\_user" pueden acceder a ciertas funciones o datos, y esos permisos son limitados para asegurar que no tengan acceso a más de lo necesario.
  - Conexión obligatoria por VPN cifrada para transferencias remotas: Se refiere a que, para realizar transferencias de datos de forma remota, es obligatorio utilizar una Red Privada Virtual (VPN) que cifra la conexión. Esto significa que los datos están protegidos durante su envío, lo que aumenta la seguridad.
  - Logs monitoreados en tiempo real: Esto sugiere que hay un sistema que registra (logs) las actividades de manera continua y que se está supervisando en tiempo real. Esto permite detectar actividades inusuales o problemas de inmediato.
  - Alertas ante fallas en tareas programadas (integrado con Zabbix/Nagios + email/Slack): Se establece que si alguna tarea programada falla, se envían alertas automáticamente. Zabbix y Nagios son herramientas de monitoreo que ayudan a gestionar el estado de los sistemas, y las alertas se envían a través de correo electrónico o Slack para que el personal pertinente pueda actuar rápidamente.

### 3.4 Prever crecimiento anual

#### Datos del problema:

- La base de datos crece un 7% mensual.
- Actualmente hay 2 bases de datos: 1 TB y 1.8 TB
- Se tiene 2 servidores dedicados a bases de datos, cada uno con 2 TB de disco duro. En total: 4 TB de almacenamiento disponible.

#### Posibles soluciones:

De acuerdo con los cálculos realizados en la primera parte; actualmente se tiene 1 Gbps (125 MB/s) en la velocidad de los servidores, sin embargo, el mayor consumo proyectado es de 231.65 MB/s en la base de datos de 5 TB durante las horas pico al final del mes lo que genera cuello de botella por lo que se propone la contratación de un enlace dedicado de 10 Gbps (1250 MB /s) para los servidores principales. Esto permitiría descargar completamente el procesamiento intensivo hacia las 16 horas no laborales y los fines de semana, donde se pueden programar respaldos, sincronización y replicas, procesamiento de reportes y mantenimiento general, minimizando el impacto en el horario productivo. Esta medida pospone la necesidad inmediata de un clúster.

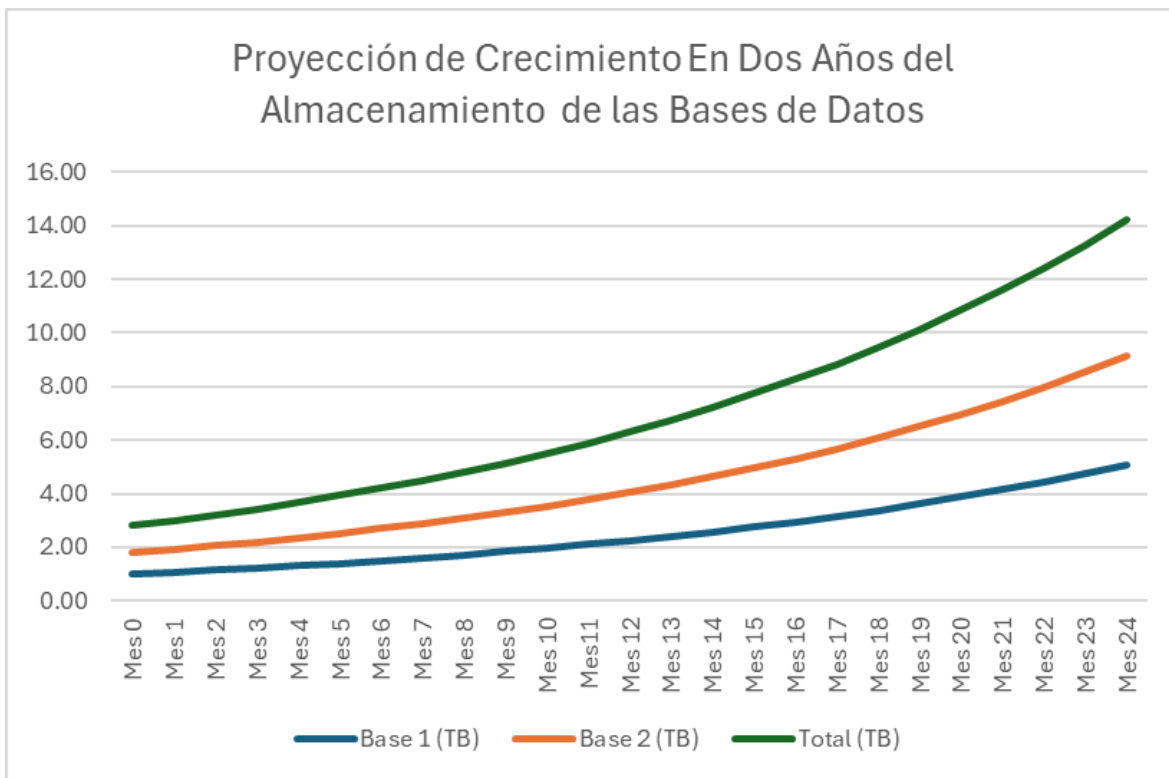
Por otra parte, uno de los servidores está llegando a su límite (1.8 TB). Realizando un diagnóstico, teniendo en cuenta que el crecimiento mensual es del 7%, con la fórmula de crecimiento de interés compuesto, el disco duro estaría llegando a su límite en menos de 30 días.

$$1.8 \text{ TB} \cdot (1 + 0.07) = 1.926 \text{ TB}$$

Por lo tanto, el cambio de disco duro a una capacidad de 4 TB (de acuerdo con las proyecciones) se debe realizar de manera urgente. Dando como resultado un total nuevo de 8 TB de espacio lo que sería suficiente para 1 año y medio o más de acuerdo con la siguiente tabla.

Mes	Base 1 (TB)	Base 2 (TB)	Total (TB)
Mes 0	1.00	1.80	2.80
Mes 1	1.07	1.93	3.00
Mes 2	1.14	2.06	3.21
Mes 3	1.23	2.21	3.43
Mes 4	1.31	2.36	3.67
Mes 5	1.40	2.52	3.93
Mes 6	1.50	2.70	4.20
Mes 7	1.61	2.89	4.50
Mes 8	1.72	3.09	4.81
Mes 9	1.84	3.31	5.15
Mes 10	1.97	3.54	5.51
Mes 11	2.10	3.79	5.89
Mes 12	2.25	4.05	6.31
Mes 13	2.41	4.34	6.75
Mes 14	2.58	4.64	7.22

Mes 15	2.76	4.97	7.73
Mes 16	2.95	5.31	8.27
Mes 17	3.16	5.69	8.84
Mes 18	3.38	6.08	9.46
Mes 19	3.62	6.51	10.13
Mes 20	3.87	6.97	10.84
Mes 21	4.14	7.45	11.59
Mes 22	4.43	7.97	12.41
Mes 23	4.74	8.53	13.27
Mes 24	5.07	9.13	14.20



Algunas otras soluciones podrían ser:

- Mover datos viejos
- Archivar datos de más de 1 año para liberar espacio.
- Guardarlos en un servidor de respaldo o en la nube.
- Tener un plan de escalamiento futuro

Solución a posibles cuestiones del profesor con “subir cosas a la nube es caro” o con la parte de los servidores

En lugar de contratar nube, se pueden usar los discos del servidor de respaldo para guardar datos antiguos, y hacer limpieza cada 6 meses para evitar llenar el espacio principal.

Antes de gastar en más servidores, podemos reducir el tamaño de la base eliminando o archivando datos viejos.

**CIERRE DE TEMA****CASO DE ESTUDIO**

Por Martinez de la Cruz Jose Julian

Fecha: 28/05/2025

Palabras clave: Administración de base de datos, proyección de crecimiento, solución ante fluctuaciones, seguridad, crecimiento del tráfico, interés compuesto.

**1. Objetivo del Tema**

El propósito de este caso de estudio es lograr una comprensión y resolución de los desafíos asociados con la administración de bases de datos enfocados en entornos más empresariales, abordando aspectos clave como la gestión eficiente de cargas fluctuantes, la seguridad e integridad de los datos, la proyección del crecimiento de la información y la optimización del rendimiento del sistema.

**2. Resumen**

DataSolutions Inc. es una empresa del sector de tecnología y servicios de datos, con una estructura operativa compuesta por 150 empleados distribuidos en tres sucursales ubicadas en diferentes ciudades. La infraestructura tecnológica actual se basa en un modelo centralizado mediante una red privada virtual (VPN), que conecta a todas las sucursales con 4 servidores principales ubicados en un nodo central de los cuales 2 con 2 TB cada uno; son dedicados a bases de datos una con 1 TB y otro de 1.8 TB y con un crecimiento mensual del 7% en las mismas.

La empresa, presenta las siguientes dificultades y desafíos en su operación con respecto a la administración de base de datos:

1. **Gestión eficiente de la carga de trabajo en fluctuaciones**
2. **Mantenimiento de la integridad y seguridad de los datos**
3. **Prever el crecimiento anual**
4. **Optimización del rendimiento**

Con base a estas problemáticas se desarrolló una propuesta hipotética, que incluyo:

**La proyección de crecimiento**

Con un 7% mensual de crecimiento, la proyección de los tamaños bases fue calculada con la fórmula de interés compuesto tal que:

Base 1: de 1.0 TB a 2.25 TB

Base 2: de 1.8 TB a 4.05 TB

Dando como resultado un total aproximado de: 6.30 TB en un año.

Además, se descubrió que el disco duro de 1.8 se agotará en menos de 1 mes, por lo que, con base a esto, se pudo anticipar la compra inmediata de un disco duro de 4TB que pueda servir por lo menos un año y medio.

### Crecimiento del tráfico

Se estimó el tráfico diario considerando las cargas en horas pico y en horas normales aplicando un crecimiento del 30% al final de cada mes para obtener el máximo consumo estimado, lo que resultó en: 231.65 MB/s en las horas pico. Esta información pudo solucionar el cuello de botella que existe en las horas pico, debido al ancho de banda de 1Gbps (125MB/s) con el que contaba la empresa, por lo que se sugirió adquirir un servicio de 10Gbps.

### Proyección Anual del Tráfico

Se estimó un aumento proporcional usando el mismo factor de crecimiento compuesto, alcanzando 15.76 TB diarios, que es casi el triple del tráfico actual.

### Solución Ante Fluctuaciones

Se descubrió que hay 16 horas muertas dentro de la empresa en la que no se absolutamente nada y se están desaprovechando, por lo que se propuso crear una estrategia de gestión por turnos del equipo de Administración de Base de Datos para aprovechar esas horas no laborales.

### Seguridad

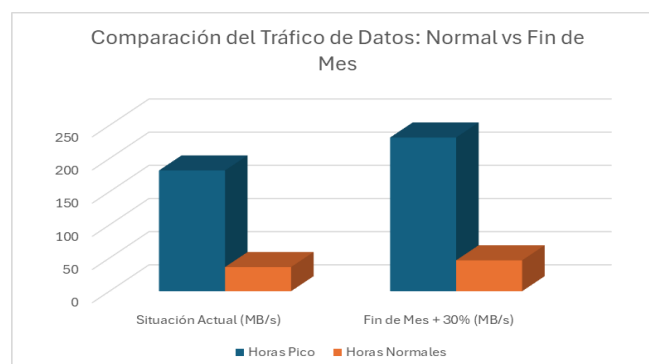
Finalmente se estableció una política de respaldo y recuperación robusta incluyendo cifrado, validación y replicación.

## 3. Conclusión

En resumen, la administración de bases de datos en contextos empresariales es un elemento de vital importancia para garantizar la continuidad de la operación, el funcionamiento del sistema y la integridad de la información. Al revisar los problemas que enfrentaba DataSolutions Inc., he podido precisar la relevancia de proyecciones adecuadas del crecimiento de almacenamiento y tráfico a través de la fórmula de interés compuesto, así como la necesidad de identificar y resolver cuellos de botella y la escasez de recursos a tiempo.

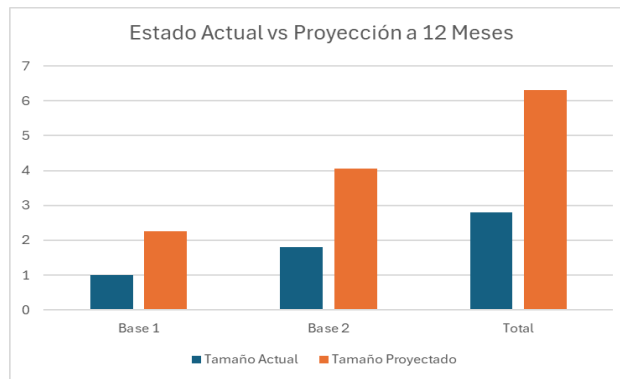
## 4. Gráficos o Ejemplos

### Gráfica de Comparación del Tráfico de Datos: Normal vs Fin de Mes (+ 30%)





### Gráfica de Comparación Entre el Estado Actual del Almacenamiento de la Base de Datos vs La Proyección a 12 meses



### 5. Referencias

- Presentación: *La Administración de Bases de Datos es un Aspecto Crítico*, MTI Jesús E. Romero, UTTEC.
- Presentación: *Caso de Estudio*, MTI Jesús E. Romero, UTTEC.
- Communications. (21 de Marzo de 2025). *Interés compuesto: ¿qué es y cómo calcularlo?*, BBVA

### 6. Reflexiones Finales

1. **Relevancia de las habilidades analíticas y técnicas del administrador de bases de datos:** Definitivamente un administrador de base de datos debe tener las competencias para el análisis de métricas, la proyección del crecimiento, la anticipación a las fallas, y evaluar soluciones técnicas efectivas. Lo anterior, permite que su cargo lo convierta en una figura estratégica para cualquier empresa tecnológica.
2. **Importancia de la previsión en la gestión de datos en las empresas:** Proporcionar proyecciones de crecimiento de volumen de datos precisas puede evitar crisis de operación como la saturación del almacenamiento, cuellos de botella o la caída del rendimiento.