

## ***Monitor para determinar el crecimiento de un tablespace***

Universidad Nacional de Costa Rica

Administración de Bases de Datos

*Profesor: Steven R. Brenes Chavarría*

*II Ciclo - 2013*

### ***Modalidad:***

En grupos de cuatro personas o menos

### ***Sobre las fechas:***

Martes 6 de agosto .....	Entrega de enunciado
Jueves 11 de agosto.....	Continuación y evacuación de dudas
Semana del Martes 13 de agosto.....	Continuación y evacuación de dudas
Semana del Martes 20 de agosto.....	Continuación y evacuación de dudas
Martes 3 de septiembre .....	Entrega y exposición de los monitores

***Valor del proyecto:*** 10% en el rubro de proyectos.

### ***Sobre la aplicación***

Se pretende que el estudiante desarrolle una aplicación informática (desktop) en Java, con el IDE de preferencia. Esta aplicación deberá poder monitorear el uso de espacio físico del tablespace, pronosticar cuándo un tablespace se va llenar y notificarle al DBA sobre su pronta entrada a falla del disco duro.

Implementación de un modelo para la administración de espacio de disco de los objetos de la base de datos (tablas, índices) basado en cálculos de demanda, aproximaciones y restricciones sobre saturación del espacio asignado.

### ***Objetivos específicos***

1. Familiarizar al estudiante con el creación y administración de diferentes objetos de la base de datos.
2. Implementar herramientas de apoyo a las tareas de administración de bases de datos.
3. Conocer la importancia de los procesos de monitoreo como control preventivo para minimizar el impacto de la materialización de los riesgos de tecnologías de información asociados a los sistemas de bases de datos.
4. Aplicar los principios básicos de estadística en la administration de bases de datos

### ***Sobre el día de la entrega***

El día de la entrega final del monitor, los estudiantes deberán realizar una presentación oral ante la clase con el fin de compartir experiencias en el desarrollo del sistema. A la vez, este día los estudiantes podrán comentar sobre las posibles mejoras que le harían al sistema eventualmente. Esta presentación no tiene valor formal en la nota del monitor.

El mejor monitor será presentado en el evento de INFOCONTEXTO, evento que pretende visibilizar a nivel de Escuela de Informática y sociedad los aportes en el campo del desarrollo de software. El mejor monitor desarrollado por los estudiantes, obtendrá un beneficio adicional en la nota del curso.

### ***Requerimientos funcionales del sistema***

El sistema debe tener como mínimo las siguientes funciones:

- 1) El sistema deberá tener una ventana login, en la cual se ingrese el username, password y el puerto para la conexión a la base de datos. Puede ser conexión ODBC o propietaria del SGBD.
- 2) La ventana login, debe poder guardar los datos ingresados (recordar datos) con el fin de que la próxima vez que el programa arranque, no se le solicite. También es necesario poder modificar estos datos en alguna ventana de configuración.
- 3) El monitor debe contar con las siguientes funciones:
  - 4) Poder listar de forma gráfica y resumida los tablespaces del SGBD y sus respectivos tiempos de llenado (de preferencia mediante gráficos de barras, indicando espacio libre, utilizado y crítico). Las unidades deben ser dadas en días, meses, años según sea el valor estimado más conveniente.
  - 5) Cada vez que se le de click, a un tablespace el mismo deberá desplegar mediante un gráfico de pastel, todas aquellas tablas y objetos que son contenidos en dicho tablespace. Además deberá indicar el tamaño relativo de dichos objetos.
  - 6) Cada vez que se le de click, a un tablespace el mismo deberá desplegar mediante un gráfico continuo el tamaño histórico del tablespace a lo largo del tiempo (el histórico deberá ser configurable desde algún menú administrativo, por defecto se recomienda crear una cola circular de 365 días).
  - 7) En el gráfico de barras del punto 4, debe aparecer un indicador del nivel crítico del tablespace. Si dicho nivel es superado el sistema deberá enviar un correo electrónico a un email especificado en la ventana de configuración. Por ejemplo suponga un tablespace de 500Gb de tamaño total, 450Gb de tamaño utilizado y umbral de 480Gb. El sistema deberá notificar de forma automática al DBA que dicho tablespace está a punto de superar el tamaño permitido.
- 8) El sistema deberá poder exportar los datos históricos en alguno de los siguientes formatos (solo uno de los siguientes): archivo binario, archivo plano XML u archivo plano separado por comas CSV a escogencia del programador.
- 9) El sistema deberá poder calcular pronósticos mediante el método de los mínimos cuadrados también llamada regresión lineal de una serie de datos estadísticos los cuales se va a recoger en periodos de 24 horas. En ese sentido lo que el DBA puede escoger es la hora del día en la que se va correr la rutina de verificación. Por ejemplo todos los días a las 8pm.

### ***Requerimientos no funcionales del instrumento***

1. El modelo debe ser capaz de calcular las necesidades de espacio en disco de almacenamiento secundario, tomando como referencia el modelo conceptual relacional así como las estimaciones de transacciones sobre las tablas que lo conforman.
2. El modelo utiliza como base de cálculo el modelo relacional directamente documentado en el diccionario de datos, sin intervención del DBA.

3. El modelo debe brindar al usuario una perspectiva en forma visual de la situación (relación de espacio usado y espacio estimado) de la base de datos
4. El modelo debe ser capaz de aplicar sus análisis a dos sistemas gestores de bases de datos diferentes (Oracle versiones 11+ ó PostgreSQL versión 9.1+)
5. El modelo permite la generación de proyecciones sobre uso de espacio y periodo de saturación para diferentes valores de consumo y tiempo.
6. El modelo debe ser programado utilizando, PlsSql para Oracle, PlpgSql para PostgreSQL, debe proveer un módulo de instalación y el mismo deberá construir los procedimientos, funciones o triggers en el sistema gestor.
7. El modelo incluye una interfaz de usuario con las facilidades requeridas para su utilización.
8. El modelo por su naturaleza y complejidad debe ser utilizado por usuarios administradores.
9. El modelo deberá ser construido en Java para aplicaciones de escritorio, el uso del IDE así como demás librerías queda a discreción del grupo de estudiantes.

### ***Otros datos de arquitectura***

El sistema debe realizar todos los días, una serie de corte de datos a la hora que el DBA configure, estos datos van a servir para realizar pronósticos sobre el comportamiento de la base de datos en el futuro y así recomendar al DBA.

Se debe crear en un archivo local de la maquina (con posibilidad de exportar) que contenga como mínimo la siguiente estructura:

Nombre de tabla	Nombre de tablespace	Fecha	Individuos totales	Tamaño total de la tabla	Nuevos individuos
T1	VENTAS	01/02/2012	0	0Mb	20
T1	VENTAS	02/02/2012	20	2Mb	20
T1	VENTAS	03/02/2012	40	4Mb	160
T1	VENTAS	04/02/2012	200	20Mb	-
T2	REPORTES	01/02/2012	100	2Mb	100

Con estos datos se pretende encontrar una función matemática que modele el comportamiento futuro del crecimiento del tablespace. Suponga que los siguientes datos son elementos del grafico de la función para T1:

Tamaño total de la tabla (Eje Y)	Individuos totales (Eje X)
0Mb	0
2Mb	20
4Mb	40
20Mb	200

De esta manera tenemos los siguientes pares ordenados para la tabla T1.

$$T1 = \{(0,0), (20,2), (40,4), (200,20)\}$$

Como lo pudieron notar existe una función fácilmente deducible, que permite calcular el tamaño en Mb de una tabla conociendo su crecimiento próximo:

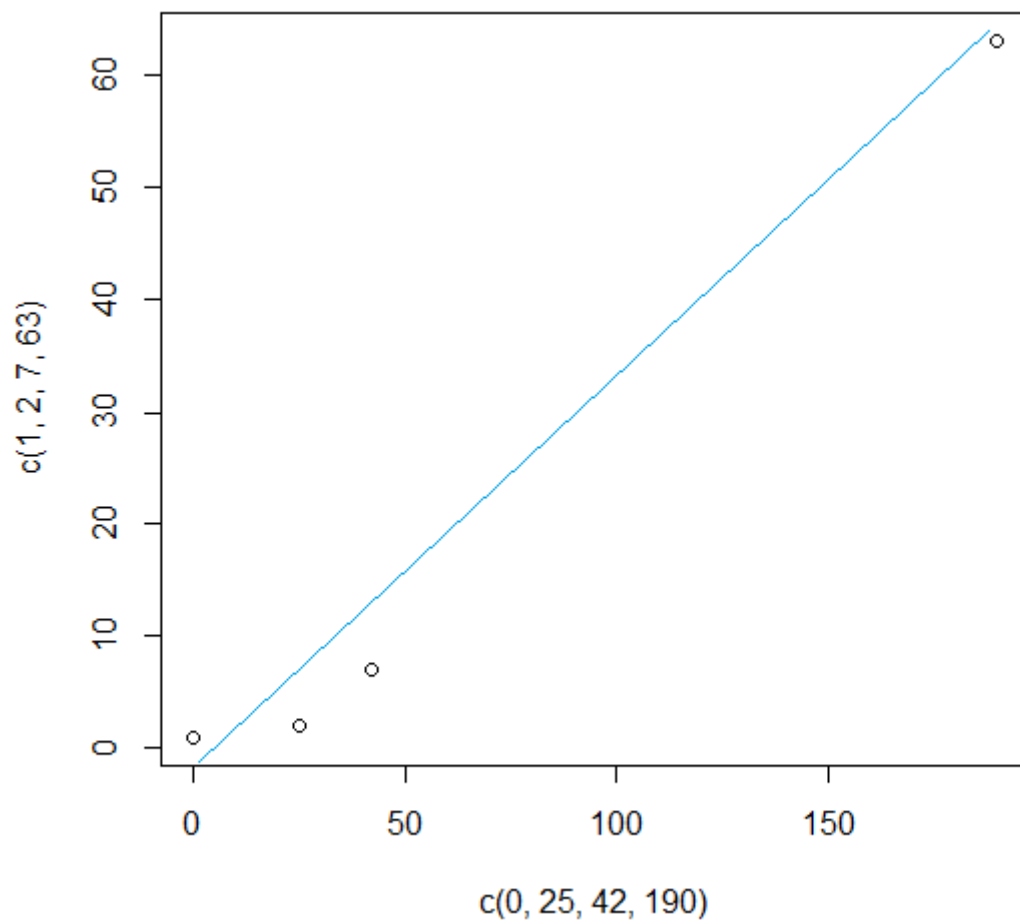
$$f(\text{individuos}) = \frac{\text{individuos}}{10}$$

Ahora para conocer el día siguiente sabemos que los individuos de la tabla T1 han crecido de la siguiente forma: 20,20,160 por lo que se esperaría que el día cuatro tenga un crecimiento de  $66.6 = \frac{20+20+160}{3}$ , a esta proyección se le llama serie de tiempo no móvil, que es realmente el promedio de los último N elementos.

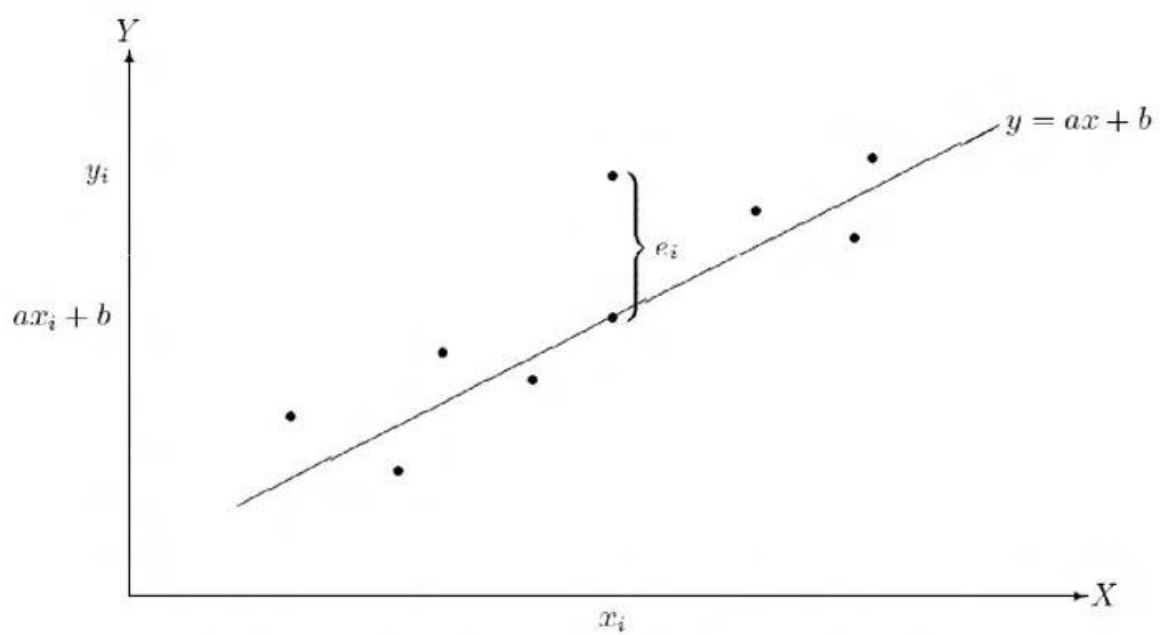
En el ejemplo anterior, los datos eran fácilmente deducibles; pero que pasaría si por ejemplo tenemos los siguientes datos:

Tamaño total de la tabla (Eje Y)	Individuos totales (Eje X)
1Mb	0
2Mb	25
7Mb	42
63Mb	190

Como se ve en el siguiente grafico, conociendo 4 puntos del grafico es posible trazar una línea recta de la forma  $y = mx + b$  que pronostique el comportamiento futuro del crecimiento de la tabla en cuestión. Van a haber tantas rectas como tablas tenga la base de datos.



Un método genérico para determinar los componentes de la función, es aplicando el método de mínimos cuadrados o de regresión lineal simple, el cual lo que hace es minimizar los errores de la recta.



Para ello es necesario encontrar el a y b de la siguiente recta  $f(x) = ax + b$ .

Donde:

$X = \text{Individuos de las tablas}$

$Y = \text{Tamaño en Mb}$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

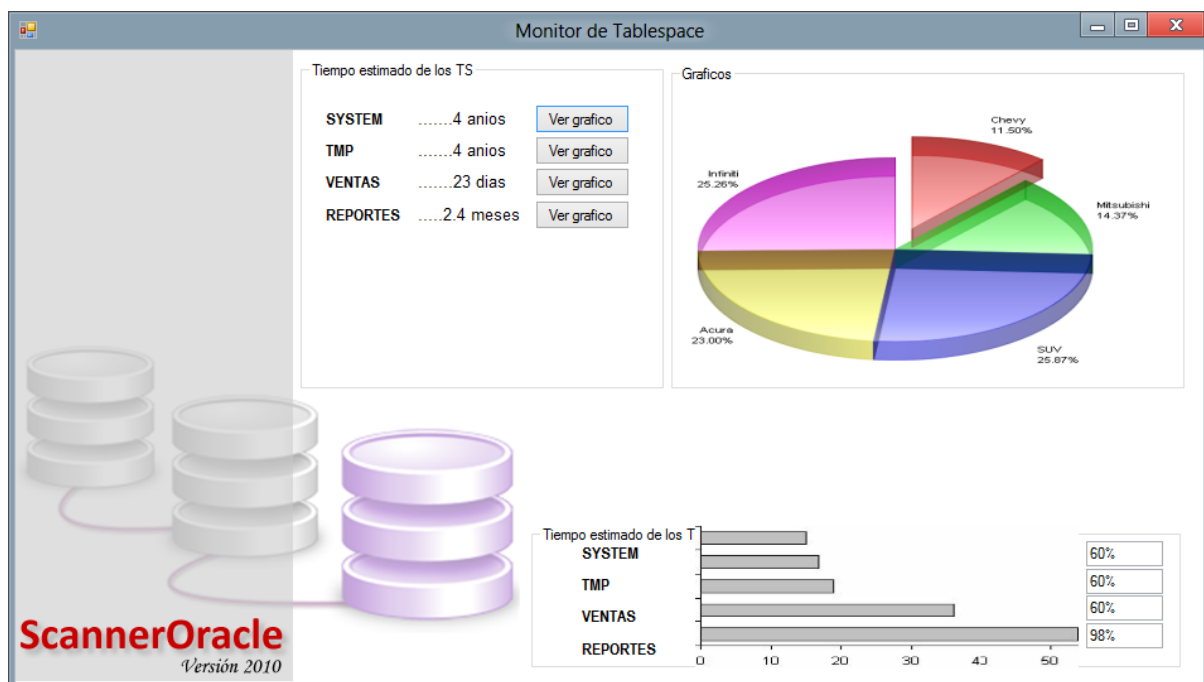
Con esto podemos calcular el valor de a y b.

$$a = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n * \text{var}(X)}$$

$$b = \bar{Y} - a\bar{X}$$

### Resultado visual de la aplicación

La aplicación principal debe verse de la siguiente forma:



Junto con esta ventana el sistema debe tener la opción de exportar datos, la ventana de login y la ventana de configuración del sistema.

## ***Sugerencias para empezar***

La siguiente consulta muestra la cantidad de Mb usados y libres de cada tablespace.

```
SELECT t1.tablespace_name, usada, libre, 100*(usada/libre) PorcentajeUsado
FROM (SELECT tablespace_name, ROUND(sum(bytes)/1024/1024,0) usada FROM
dba_free_space WHERE tablespace_name NOT LIKE 'TEMP%' GROUP BY
tablespace_name) t1, (SELECT tablespace_name, ROUND(sum(BYTES/1024/1024),0)
libre FROM dba_data_files bWHERE tablespace_name NOT LIKE 'TEMP%' GROUP BY
b.tablespace_name) t2 WHERE t1.tablespace_name = t2.tablespace_name;
```

La siguiente consulta muestra la cantidad de Mb usados de cada tabla dentro de un tablespace

```
SELECT a.table_name, ROUND(SUM (b.bytes)) / 1024 / 1024 Mb
FROM (SELECT tablespace_name, table_name FROM user_tables) a,
(SELECT segment_name, tablespace_name, bytes FROM user_extents) b
WHERE a.table_name = b.segment_name
AND a.tablespace_name = b.tablespace_name
AND upper(a.TABLESPACE_NAME) like '%SYSTEM%'
GROUP BY a.tablespace_name, a.table_name;
```

La siguiente consulta muestra la cantidad de Mb usados de cada tabla dentro de **TODOS LOS TABLESPACES**.

```
SELECT a.table_name TABLA, a.tablespace_name TABLESPACE, ROUND(SUM
(b.bytes))/1024/1024 Mb FROM (SELECT tablespace_name, table_name FROM
user_tables) a, (SELECT segment_name, tablespace_name, bytes FROM
user_extents) b WHERE a.table_name = b.segment_name AND a.tablespace_name =
b.tablespace_name GROUP BY a.tablespace_name, a.table_name
ORDER BY a.tablespace_name, a.table_name;
```