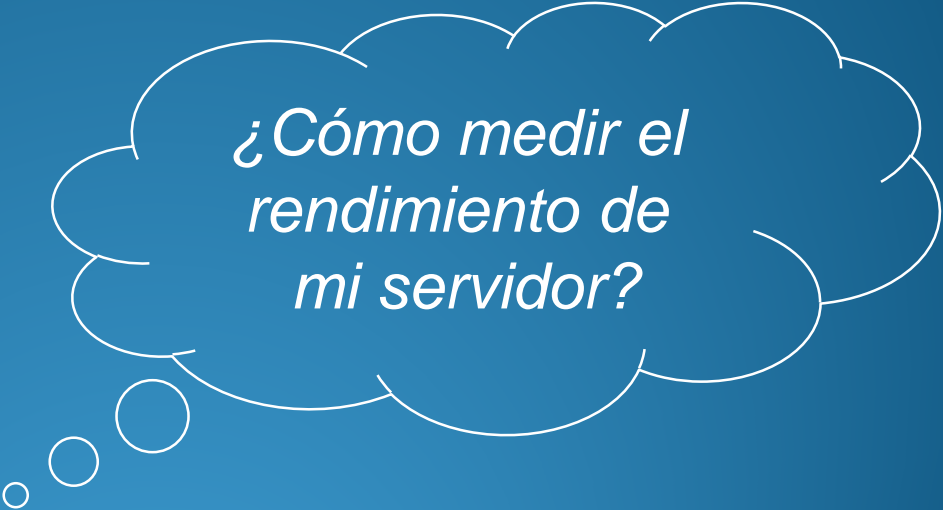


# Tema 3. Monitorización de servicios y programas



*¿Cómo medir el  
rendimiento de  
mi servidor?*

Analistas, administradores y diseñadores

# Objetivos del tema

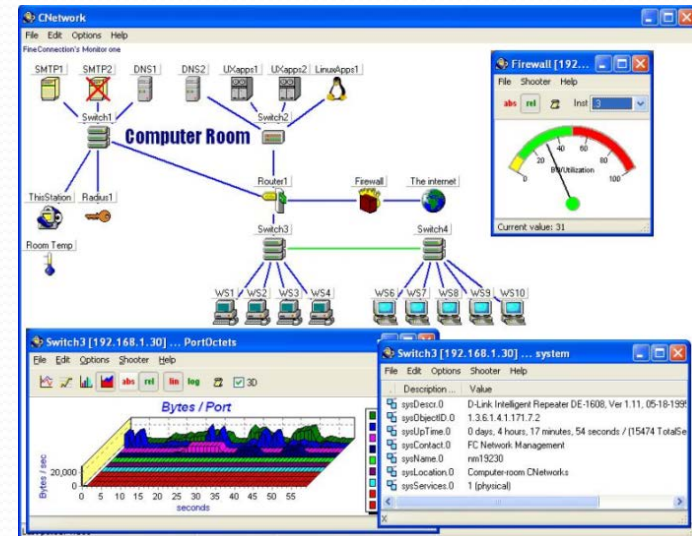
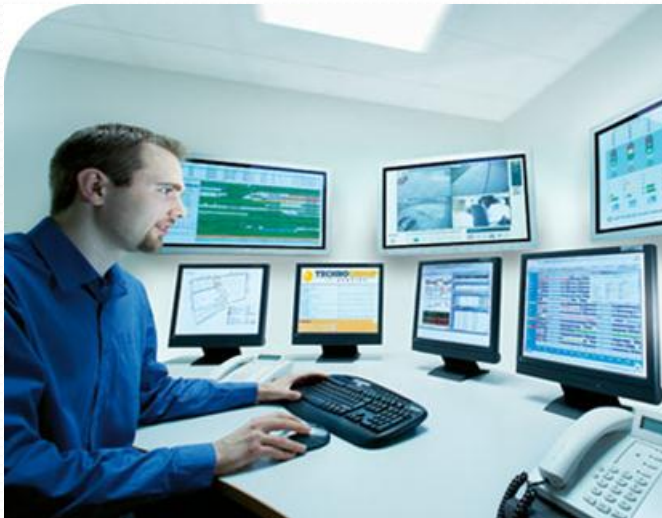
- Entender el concepto de monitor de actividad de un servidor y sus diferentes utilidades e implementaciones.
- Conocer las características fundamentales de un monitor a nivel de sistema operativo y a nivel de aplicación concreta (profilers).
- Comprender el papel que desempeñan los monitores para evaluar el rendimiento de un servidor ante una carga real.
- Saber interpretar adecuadamente la información que aporta un monitor.

# Bibliografía

- *Evaluación y modelado del rendimiento de los sistemas informáticos*. Xavier Molero, C. Juiz, M. Rodeño. Pearson Educación, 2004.
  - Capítulo 2
- *Measuring computer performance: a practitioner's guide*. D. J. Lilja, Cambridge University Press, 2000.
  - Capítulos 4 y 6
- *The art of computer system performance analysis*. R. Jain. John Wiley & Sons, 1991.
  - Capítulos 7 y 8
- *System performance tuning*. G.D. Musumeci, M. Loukides. O'Reilly, 2002.
  - Capítulo 2
- *Linux performance and tuning guidelines*. E.Ciliendo, T.Kunimasa. IBM Redpaper, 2007.
  - Capítulos 1 y 2
- *Linux man pages*. <http://www.linuxmanpages.com/>.

# Contenido

- Concepto de monitor de actividad.
- Monitorización a nivel de sistema.
- Monitorización a nivel de aplicación.



## 3.1. Concepto de Monitor de Actividad

# La carga y la actividad de un sistema

- **Carga** (*workload*): conjunto de tareas que ha de realizar un sistema. (= Todo aquello que demande recursos del sistema.)
- **Actividad** de un sistema: conjunto de operaciones que se realizan en el sistema como consecuencia de la carga que tiene.
- Algunas variables que reflejan la actividad de un sistema:
  - Procesadores: Utilización, temp.,  $f_{CLK}$ , nº de procesos, nº de interrupciones, cambios de contexto, etc.
  - Memoria: nº de accesos, memoria utilizada, fallos de caché, fallos de página, uso de memoria de intercambio, latencias, anchos de banda, voltajes, etc.
  - Discos: lecturas/escrituras por unidad de tiempo, longitud de las colas de espera, tiempo de espera medio por acceso, etc.
  - Red: paquetes recibidos/enviados, colisiones por segundo, sockets abiertos, paquetes perdidos, etc.
  - Sistema global: nº de usuarios, nº de peticiones, etc.
- Un sistema informático no es bueno ni malo *per se*, sino que se adapta mejor o peor a un tipo determinado de carga.





# Definición de monitor de actividad

- Herramienta diseñada para medir la actividad de un sistema informático y facilitar su análisis.



- Acciones típicas de un monitor:
  - Medir la actividad.
  - Procesar y almacenar la información recopilada.
  - Mostrar los resultados.



# Utilidad de los monitores de actividad

- Administrador/Ingeniero
  - Conocer la utilización de los recursos (detección de cuellos de botella) para saber:
    - Qué hardware hay que reconfigurar / sustituir/ añadir.
    - Cómo ajustar los parámetros del sistema (sintonización).
  - Predecir cargas futuras (capacity planning).
  - Tarificar a los clientes.
  - Obtener un modelo un componente o de todo el sistema para poder deducir qué pasaría si...
- Programador
  - Conocer las partes críticas (hot spots) de una aplicación.
- Sistema Operativo
  - Adaptarse dinámicamente a la carga.





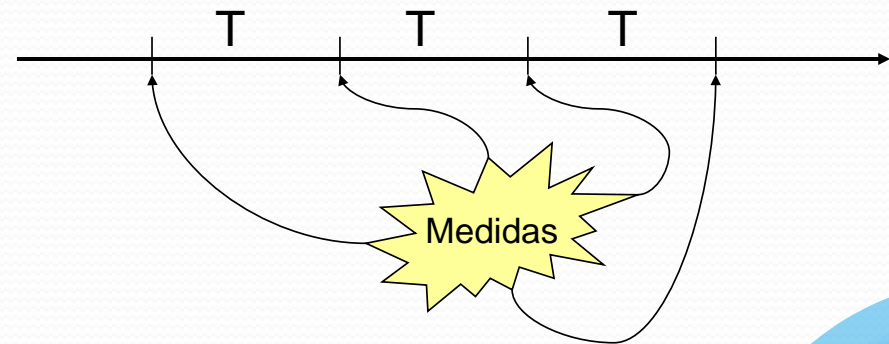
# Tipos de monitores: ¿cuándo se mide?

Cada vez que ocurre un evento (*monitor por eventos*)

- Evento: Cambio en el estado del sistema.
- Volumen de información recogida: Depende de la frecuencia de los eventos.
- Información exacta.
- Ejemplos de eventos:
  - Inicio/fin de la ejecución de un programa.
  - Fallo en memoria cache.
  - Interrupción de un dispositivo periférico.
  - Abrir/cerrar un fichero, etc.

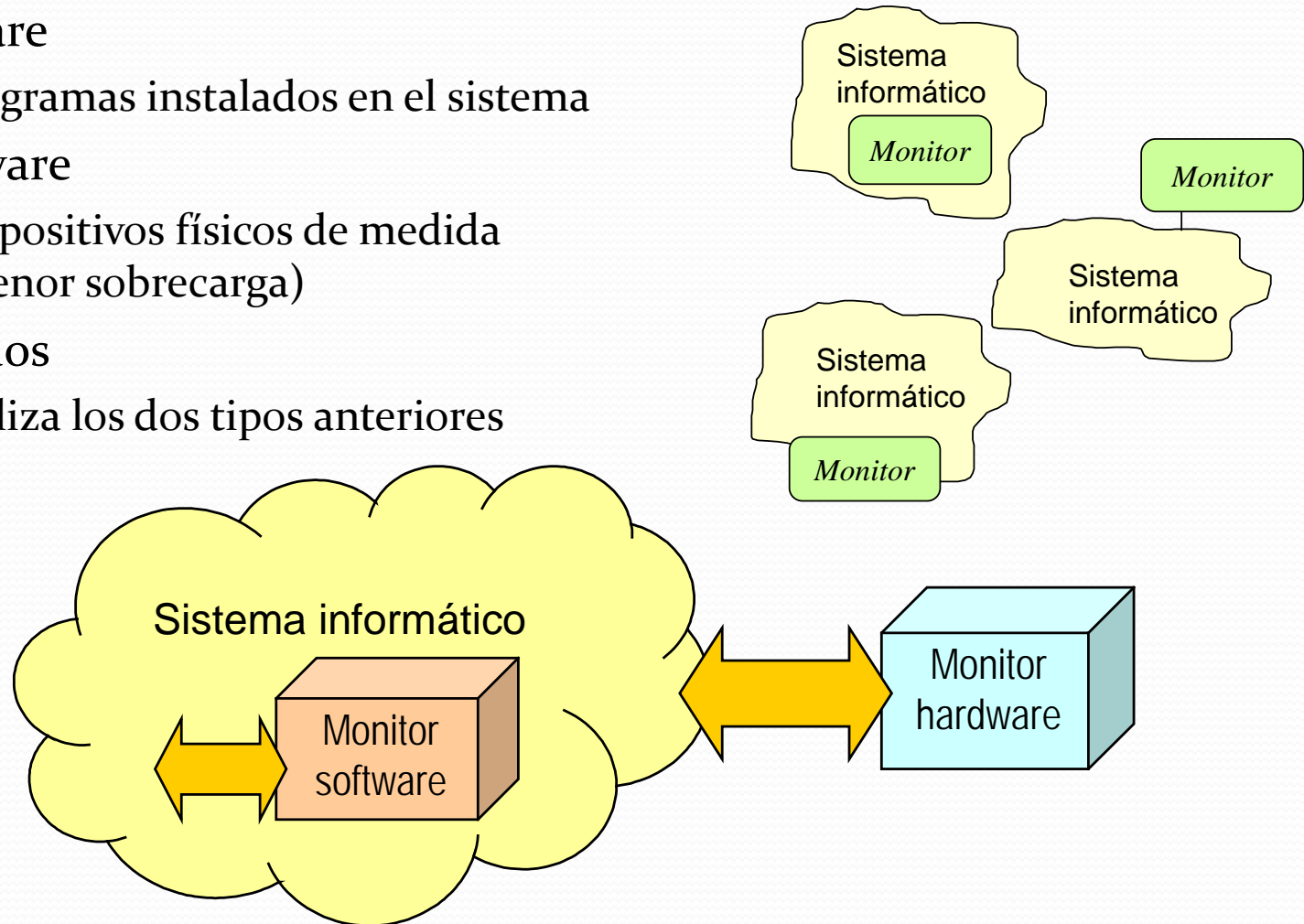
A intervalos regulares de tiempo (*monitor por muestreo*)

- El volumen de información global recogida depende del periodo de muestreo (T).
- Información estadística.



# Tipos de monitores: ¿cómo se mide?

- Software
  - Programas instalados en el sistema
- Hardware
  - Dispositivos físicos de medida (menor sobrecarga)
- Híbridos
  - Utiliza los dos tipos anteriores



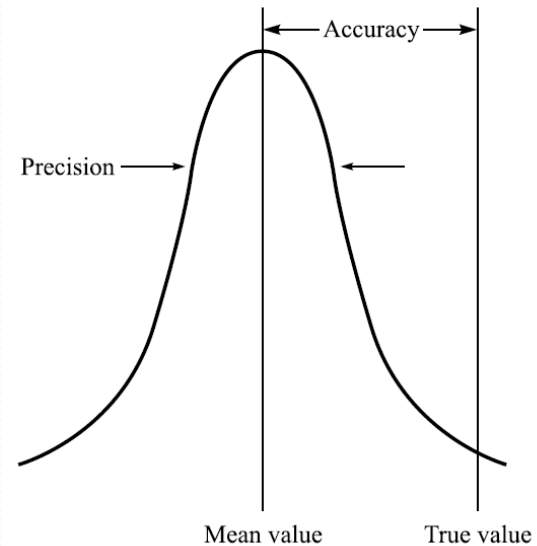
# Tipos de monitores: ¿existe interacción con el analista/administrador?

- No existe. La consulta sobre los resultados se realiza aparte mediante otra herramienta independiente al proceso de monitorización: monitores tipo batch, por lotes o en segundo plano (*batch monitors*).
- Sí existe. Durante el propio proceso de monitorización se pueden consultar los valores monitorizados y/o interactuar con ellos realizando representaciones gráficas diversas, modificando parámetros del propio monitor, etc.: monitores en primer plano o interactivos (*on-line monitors*).



# Atributos que caracterizan a una medida

- **Exactitud de la medida (Accuracy):** ¿Cómo se aleja el valor medido del valor real que se quiere medir?
- **Precisión (Precision):** ¿Cuál es la dispersión de las medidas?
- **Resolución del monitor/sensor:** ¿Cuánto tiene que cambiar el valor a medir para detectar un cambio?
- **Periodo de Muestreo (Sampling Time):** ¿Cada cuánto tiempo tomamos la medida? (monitores por muestreo)
- **Tasa Máxima de Entrada (Max Input Rate):** ¿Cuál es la frecuencia máxima de ocurrencia de los eventos que el monitor puede observar? (monitores por eventos)
- **Anchura de Entrada (Input Width):** ¿Cuánta información (nº de bytes de datos) se almacena por cada medida que toma el monitor?
- **Sobrecarga (Overhead):** ¿Qué recursos le “roba” el monitor al sistema?  
➡ *El instrumento de medida puede perturbar el funcionamiento del sistema.*



# Cálculo de la sobrecarga en un monitor

- La ejecución de las instrucciones del monitor se lleva a cabo utilizando recursos del sistema monitorizado.

$$\text{Sobrecarga}_{\text{RECURSO}} = \frac{\text{Uso del recurso por parte del monitor}}{\text{Capacidad total del recurso}}$$

- Ejemplo de cálculo:
  - El monitor se activa cada 5 s y cada activación del mismo usa el procesador durante 6 ms.

$$\text{Sobrecarga}_{\text{CPU}} = \frac{6 \times 10^{-3} \text{ s}}{5 \text{ s}} = 0.0012 = 0.12\%$$



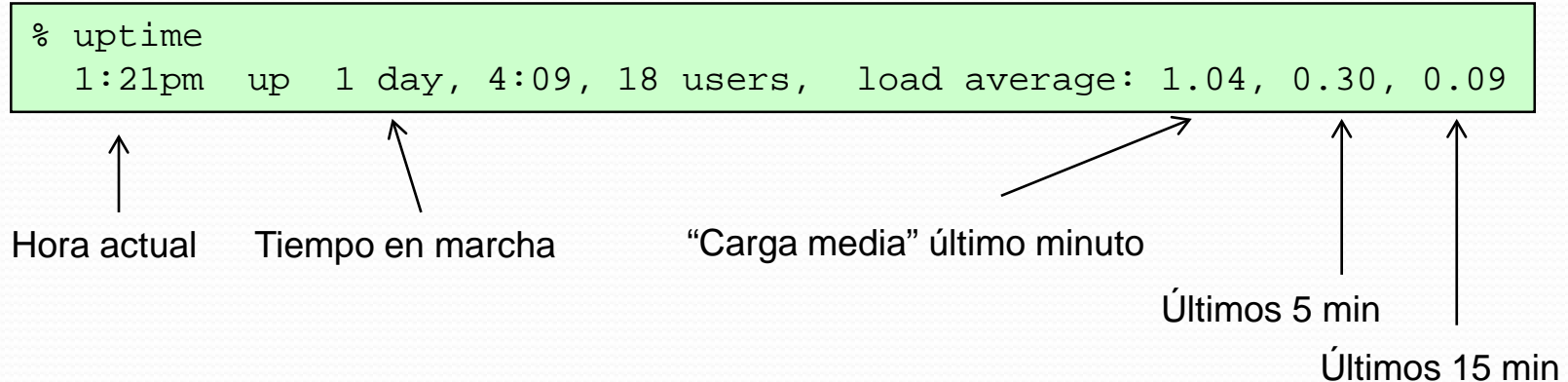
## 3.2. Monitorización a nivel de sistema

# El directorio `/proc` (Unix)

- Es una carpeta en RAM utilizada por el núcleo de Unix como interfaz con las estructuras de datos del kernel.
- A través de `/proc` podemos:
  - Acceder a información global sobre el S.O. : `loadavg`, `uptime`, `cpuinfo`, `meminfo`, `mounts`, `net`, `kmsg`, `cmdline`, `slabinfo`, `filesystems`, `diskstats`, `devices`, `interrupts`, `stat`, `swap`, `version`, `vmstat` ...
  - Acceder a la información de cada uno de los procesos del sistema.  
(`/proc/[pid]`): `stat`, `status`, `statm`, `mem`, `smaps`, `cmdline`, `cwd`, `environ`, `exe`, `fd`, `task`...
  - Acceder y, a veces, modificar algunos parámetros del kernel del S.O.  
(`/proc/sys`): `dentry_state`, `dir-notify-enable`, `dquot-max`, `dquot-nr`, `file-max`, `file-nr`, `inode-max`, `inode-nr`, `lease-break-time`, `mqueue`, `super-max`, `super-nr`, `acct`, `domainname`, `hostname`, `panic`, `pid_max`, `version`, `net`, `vm`...
- La mayoría de los monitores de Unix usan como fuente de información este directorio.

# uptime

- Tiempo que lleva el sistema en marcha y la “carga media” que soporta



[man uptime \(http://man7.org/linux/man-pages/man1/uptime.1.html\)](http://man7.org/linux/man-pages/man1/uptime.1.html)

## NAME

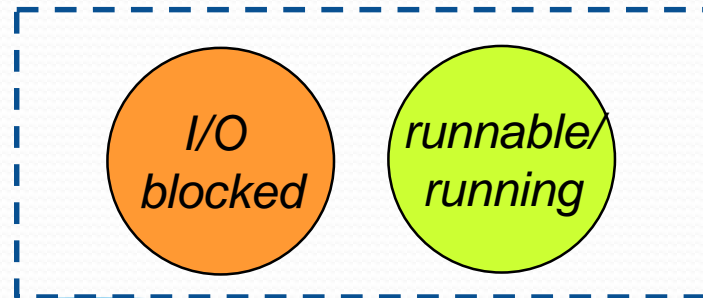
uptime - Tell how long the system has been running.

## DESCRIPTION

- **uptime** gives a one line display of the following information. The current time, how long the system has been running, how many users are currently logged on, and the system load averages for the past 1, 5, and 15 minutes. This is the same information contained in the header line displayed by `w`.

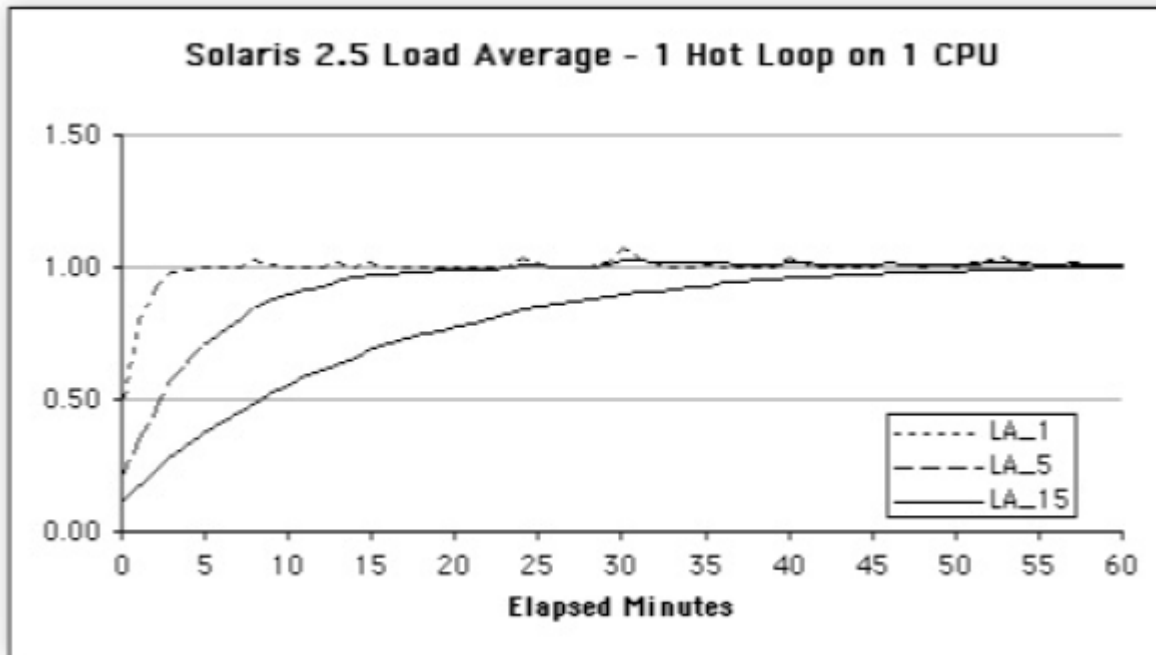
# Carga del sistema Unix

- Estados básicos de un proceso:
  - En ejecución (*running*) o en la cola de ejecución (*runnable*).
  - Bloqueado esperando a que se complete una operación de E/S para continuar (*uninterruptible sleep* = *I/O blocked*).
  - Durmiendo esperando a un evento del usuario o similar (p.ej. una pulsación de tecla) (*interruptible sleep*).
- La cola de procesos del núcleo (*run queue*) está formada por aquellos que pueden ejecutarse (*runnable* + *running*).
- Carga del sistema (*system load*): número de procesos en modo *running*, *runnable* o *I/O blocked*.



# ¿Cómo mide la carga media el S.O.?

- Experimento: Ejecutamos 1 único proceso (bucle infinito). Llamamos a uptime cada cierto tiempo y representamos los resultados.



Según timer.c  
(kernel de Linux):  
$$LA(t) = c \cdot \text{load}(t) + (1-c) \cdot LA(t-5)$$

LA = Load Average  
Se actualiza cada 5 segundos su valor.

- $c$  es una constante. A mayor valor, más influencia tiene la carga actual en el valor medio ( $c_1 > c_5 > c_{15}$ ). Si  $c = c_1$  calculamos  $LA_1(t)$ , etc.
- $\text{load}(t)$  es la carga del sistema en el instante  $t$ .



# ps (*process status*)

- Información sobre el estado actual de los procesos del sistema.
  - Es una de las herramientas más importantes empleadas en tareas de monitorización.

```
$ ps aur
USER      PID %CPU %MEM  VSZ   RSS TTY      STAT START  TIME COMMAND
miguel    29951 55.9   0.1 1448   384 pts/0    R      09:16  0:11 tetris
carlos    29968 50.6   0.1 1448   384 pts/0    R      09:32  0:05 tetris
javier    30023  0.0   0.5 2464  1492 pts/0    R      09:27  0:00 ps aur
```

- Tiene una gran cantidad de parámetros (aquí se presentan algunos).

```
***** simple selection ***** selection by list *****
-A, -e all processes                                -C by command name
-N negate selection                                -G by real group ID
-a all w/ tty except session leaders                -U by real user ID
-d all except session leaders                       -g by session
-p by process ID                                    t  by tty
T  all processes on this terminal                   -s proc. in sessions given
a  all w/ tty, including other users                 -t by tty
-u by effective user ID                             r  only running processes
U  processes for specified users                     x  processes w/o controlling ttys
```

# Información aportada por ps

- USER
  - Usuario que lanzó el proceso
- %CPU, %MEM
  - Porcentaje de procesador y memoria física usada
- SIZE (o VSIZE)
  - Memoria (KiB) virtual total asignada al proceso
- RSS (*resident set size*)
  - Memoria (KiB) física ocupada por el proceso
- STAT
  - R (*running or runnable*), D (*I/O blocked*), S (*interruptible sleep*), T (*stopped*), Z (*zombie: terminated but not died*)
  - N (*lower priority = running niced*), < (*higher priority = not nice*)
  - s (*session leader*), + (*in the foreground process group*), W (*swapped*)

# top

- Muestra cada T segundos: carga media, procesos, consumo de memoria...
- Normalmente se ejecuta en modo interactivo (se puede cambiar T, las columnas seleccionadas, la forma de ordenar las filas, etc.)

```
8:48am up 70 days, 21:36, 1 user, load average: 0.28, 0.06, 0.02
47 processes: 45 sleeping, 2 running, 0 zombie, 0 stopped
CPU states: 99.6% user, 0.3% system, 0.0% nice, 0.0% idle
Mem: 256464K av, 234008K used, 22456K free, 0K shrd, 13784K buff
Swap: 136512K av, 4356K used, 132156K free 5240K cached
```

PID	USER	PRI	NI	SIZE	RSS	SHARE	STAT	LC	%CPU	%MEM	TIME	COMMAND
9826	carlos	0	0	388	388	308	R	0	99.6	0.1	0:22	simulador
9831	miguel	19	0	976	976	776	R	0	0.3	0.3	0:00	top
1	root	20	0	76	64	44	S	0	0.0	0.0	0:03	init
2	root	20	0	0	0	0	SW	0	0.0	0.0	0:00	keventd
4	root	20	19	0	0	0	SWN	0	0.0	0.0	0:00	ksoftiq
5	root	20	0	0	0	0	SW	0	0.0	0.0	0:13	kswapd
6	root	2	0	0	0	0	SW	0	0.0	0.0	0:00	bdf flush
7	root	20	0	0	0	0	SW	0	0.0	0.0	0:10	kdated
8	root	20	0	0	0	0	SW	0	0.0	0.0	0:01	kinoded
11	root	0	-20	0	0	0	SW<	0	0.0	0.0	0:00	recoved

# vmstat (*virtual memory statistics*)

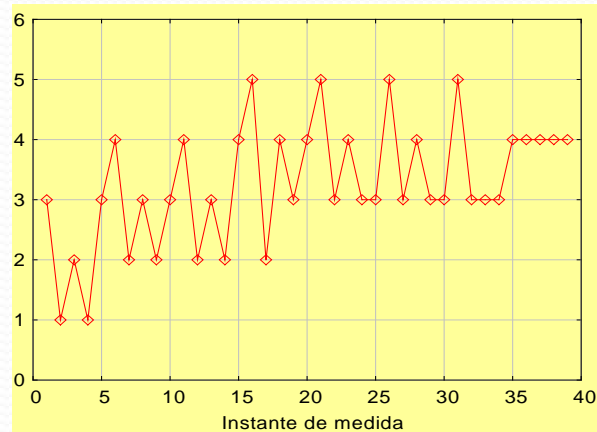
- *Paging* (paginación), *swapping*, interrupciones, cpu
  - La primera línea no sirve para nada (info desde el inicio del sistema)

% vmstat 1 6																	
procs			memory				swap		io		system		cpu				
r	b	swpd	free	buff	cache	si	so	bi	bo	in	cs	us	sy	id	wa	st	
0	0	868	8964	60140	342748	0	0	23	7	222	199	1	4	80	15	0	
0	0	868	8964	60140	342748	0	0	0	14	283	278	0	7	80	23	0	
0	0	868	8964	60140	342748	0	0	0	0	218	212	6	2	93	0	0	
0	0	868	8964	60140	342748	0	0	0	0	175	166	3	3	94	0	0	
0	0	868	8964	60140	342752	0	0	0	2	182	196	0	7	88	5	0	
0	0	868	8968	60140	342748	0	0	0	18	168	175	3	8	69	20	0	

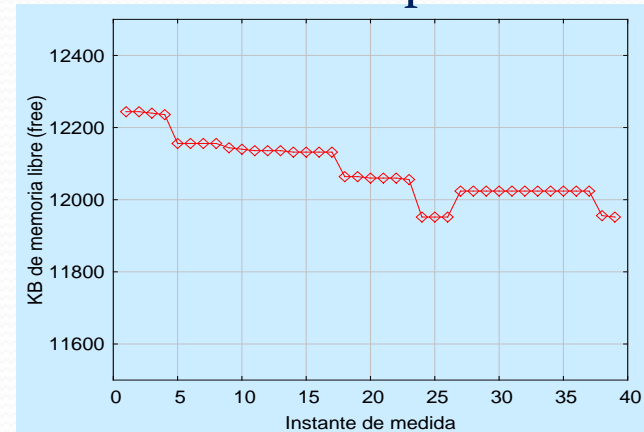
- Procesos: *r* (*runnable*), *b* (*I/O blocked*)
- Bloques por segundo transmitidos: *bi* (*blocks in*), (*blocks out*)
- KB/s entre memoria y disco: *si* (*swapped in*), *so* (*swapped out*)
- *in* (*interrupts per second*), *cs* (*context switches per second*)
- Con otros argumentos, puede dar información sobre acceso a discos (en concreto la partición de swap) y otras estadísticas de memoria.

# Representación gráfica de los resultados

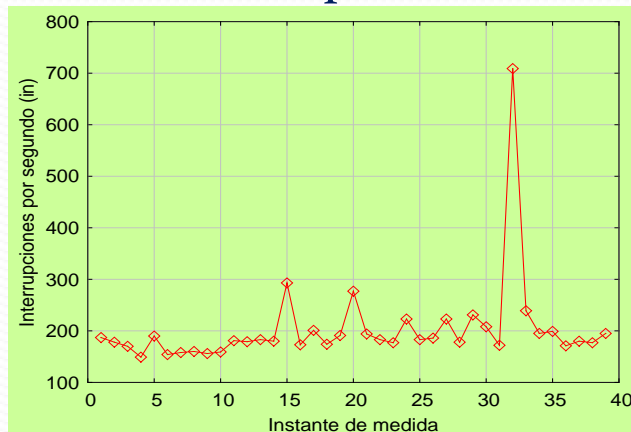
## Procesos activos



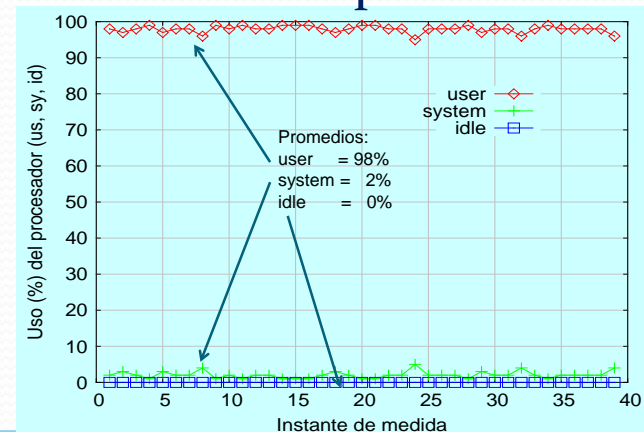
## Memoria disponible



## Interrupciones



## Utilización procesador





# El paquete de monitorización: SysStat



☆ 📁 🖨️ 🌱 ↩️ sebastien.godard.pagesperso-orange.fr 🔍 systat sar ☰

Welcome to the  
**SYSSTAT**  
Utilities Home Page!

[News](#) | [Features](#) | [Download](#) | [GitHub](#) | [Documentation](#) | [Tutorial](#) | [FAQ](#) | [Contact](#)

Useful links:  
[Freecode](#)  
[IceWalkers](#)  
  
[The Linux Foundation](#)  
[Linux Kernel Archives](#)  
[Linux Kernel Mailing List](#)  
[The Linux Documentation Project](#)

[Donate](#)

W3C HTML 4.01

You are visitor #  
**564920**

## News

**Sunday October 19th, 2014**

Sysstat 11.1.2 released (development version).  
Sysstat 11.0.2 released (stable version).  
Version 11.1.2 includes two new commands in the contrib directory: irqtop (contributed by Robert Elliott from HP), which monitors differences in /proc/interrupts and /proc/softirqs per CPU along with CPU statistics, and irqstat (contributed by Lance Shelton from SanDisk/Fusio-io) designed for NUMA systems with many processors. These two commands should be merged in a future sysstat version.  
Apart from that, this version includes mainly bug fixes (which you can find in the stable 11.0.2 version).

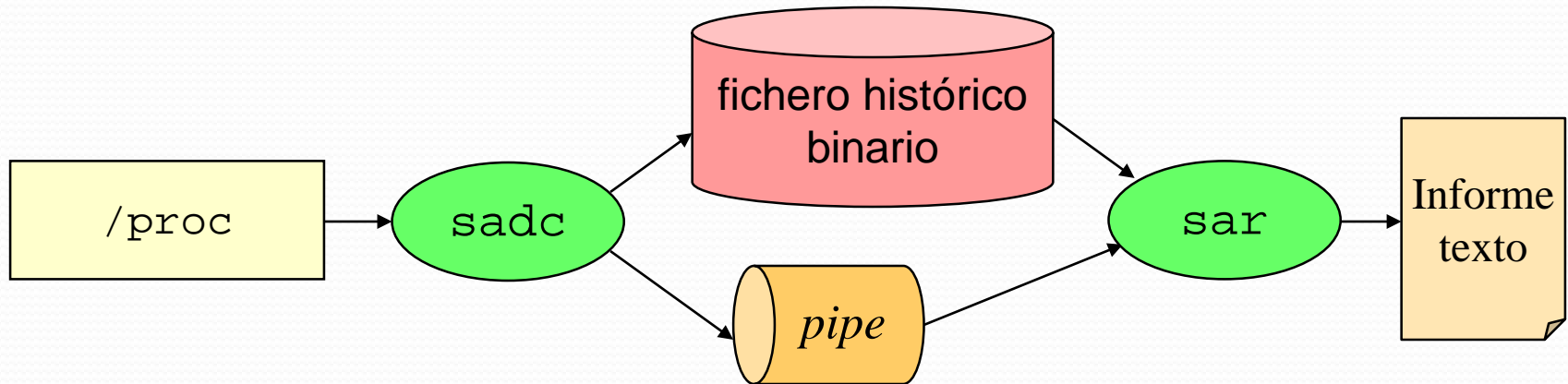
**Saturday August 30th, 2014**

# El monitor `sar`

- `sar` (*system activity reporter*)
  - Muy utilizado por los administradores de sistemas Unix en la detección de cuellos de botella (*bottlenecks*).
  - Información sobre todo el sistema.
    - Actual: qué está pasando el día de hoy, o ahora mismo, en el sistema.
    - Histórica: qué ha pasado en el sistema en otros días pasados.
      - Ficheros históricos en `/var/log/sysstat/saDD`, donde los dígitos DD indican el día del mes.
  - Hace uso de contadores estadísticos del núcleo del sistema operativo ubicados en el directorio `/proc`.
- Disponibilidad en Internet y ya incluido en muchas distribuciones de Linux.
  - <http://sebastien.godard.pagesperso-orange.fr/>

# Funcionamiento del monitor sar

- Se basa en dos órdenes complementarias
  - `sadc` (*system-accounting data collector*)
    - Recoge los datos estadísticos (lectura de contadores) y construye un registro en formato binario (*back-end*)
  - `sar`
    - Lee los datos binarios que recoge `sadc` y los traduce a un formato legible por nosotros en formato texto (*front-end*)



# Parámetros de sar

- Gran cantidad de parámetros (puede funcionar en modo batch o en modo interactivo)

Modo interactivo: [tiempo\_muestreo, [nº muestras]]

Modo no interactivo:

-f Fichero de donde extraer la información, por defecto: hoy  
-e Hora de fin de la monitorización  
-s Hora de comienzo de la monitorización

-u Utilización del procesador (opción por defecto)  
-P Mostrar estadísticas por cada procesador (-P ALL: todos)  
-I Estadísticas sobre interrupciones  
-w Cambios de contexto  
-q Tamaño de la cola y carga media del sistema  
-b Estadísticas de transferencias de E/S  
-d Transferencias para cada disco  
-n Conexión de red  
-r Utilización de memoria  
-R Estadísticas sobre la memoria  
-A Toda la información disponible  
...

# Ejemplo de salidas del monitor sar

- Utilización de los procesadores (sistema biprocesador)

```
$ sar (=sar -u)
00:00:00      CPU      %user      %nice      %system      %idle
00:05:00      all       0.09       0.00        0.08       99.83
00:10:00      all       0.01       0.00        0.01       99.98
...
11:15:00      all       0.02       0.00        0.02       99.96
11:20:00      all       0.44       0.00        0.20       99.36
11:25:00      all       0.05       0.00        0.02       99.92
```

- Actividad del sistema de entrada/salida

```
$ sar -b
00:00:00      tps      rtps      wtps      bread/s      bwrtn/s
00:05:00      0.74      0.39      0.35        7.96        3.27
00:10:01      0.09      0.00      0.09        0.00        0.91
00:15:00      0.15      0.00      0.14        0.03        1.36
00:20:00     65.12     59.96      5.16     631.62     162.64
```



# Otros ejemplos de ejecución de sar

- Ejecución interactiva (tiempo de muestro, nº muestras)

- `sar -u 2 30`

- Información recogida sobre el día de hoy

- `sar (= sar -u)`

- `sar -d -s 10:00:00 -e 12:00:00`

- `sar -A`

- `sar -b`

- Información recogida en otro día anterior

- `sar -f /var/log/sysstat/sa02`

- `sar -P ALL -f /var/log/sysstat/sa06`

- `sar -d -s 10:00:00 -e 12:00:00 -f /var/log/sa/sa04`

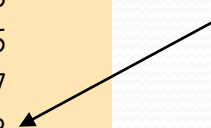
# Los datos sobre la actividad

- Se utiliza un fichero histórico de datos por cada día.
- Se programa la ejecución de `sadc` un número de veces al día con la utilidad “cron” del sistema Unix.
  - Por ejemplo, una vez cada 5 minutos comenzando a las 0:00 de cada día.
- Cada ejecución de `sadc` añade un registro binario con los datos recogidos al fichero histórico del día.

```
%ls /var/log/sysstat
```

```
-rw-r--r-- 1 root root 3049952 Sep 30 23:55 sa30
-rw-r--r-- 1 root root 3049952 Oct 1 23:55 sa01
-rw-r--r-- 1 root root 3049952 Oct 2 23:55 sa02
-rw-r--r-- 1 root root 3049952 Oct 3 23:55 sa03
-rw-r--r-- 1 root root 3049952 Oct 4 23:55 sa04
-rw-r--r-- 1 root root 3049952 Oct 5 23:55 sa05
-rw-r--r-- 1 root root 3049952 Oct 6 23:55 sa06
-rw-r--r-- 1 root root 3049952 Oct 7 23:55 sa07
-rw-r--r-- 1 root root 2372320 Oct 8 18:45 sa08
```

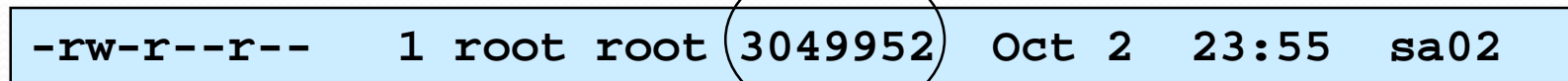
Día actual



# Análisis de un fichero histórico

- Ejemplo

- El fichero histórico de un día ocupa 3.049.952 bytes (unos 3 MB)
- La orden `sadc` se ejecuta cada 5 minutos
  - Cada hora se recogen 12 muestras
  - Al día se recogen  $24 \times 12 = 288$  muestras
- Anchura de entrada del monitor:  
Cada registro ocupa de media 10.3 KB



```
-rw-r--r--  1 root root 3049952 Oct 2  23:55 sa02
```

Fichero `sa02`  
(día 2 de octubre)

# Otras herramientas de Sysstat

- `mpstat` (*processors related statistics*)

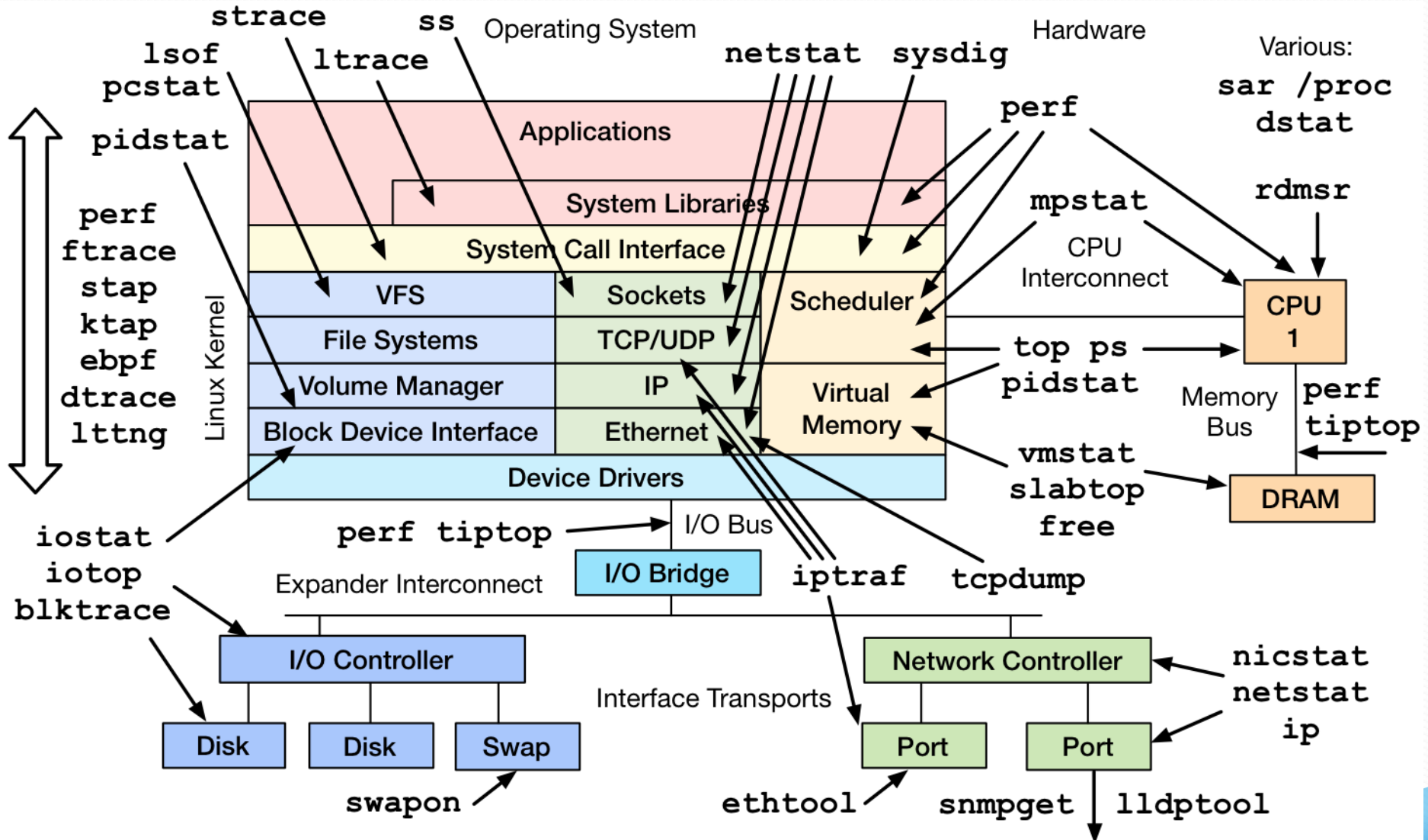
```
$ mpstat -P 1 3 5
12:07:03      CPU      %user    %nice  %system    %idle    intr/s
12:07:06        1  100.00     0.00     0.00     0.00     63.00
12:07:09        1  100.00     0.00     0.00     0.00     66.00
12:07:12        1  100.00     0.00     0.00     0.00     44.00
12:07:15        1  100.00     0.00     0.00     0.00     74.00
12:07:18        1  100.00     0.00     0.00     0.00     50.00
Average:        1  100.00     0.00     0.00     0.00     59.40
```

- `iostat` (*input/output statistics*)

```
$ iostat
cpu-avg:  %user    %nice    %sys    %idle
          3.70     0.02     0.48    95.81

Device:      tps    Blq_read/s    Blq_wrtn/s    Blq_read    Blq_wrtn
dev2-0        0.00         0.00         0.00        133         0
dev3-0        0.55         4.53         6.62    11726226    17108122
dev3-1        0.01         0.00         0.61       2698    1590072
```

# Otras herramientas para monitorización



# Otras herramientas para monitorización (II)

- CollectL
  - <http://collectl.sourceforge.net/>. Parecido a sar. Es capaz de ejecutarse de forma interactiva o como un demonio para recopilar datos históricos.
- Nagios
  - <https://www.nagios.org/>. Permite la monitorización y la generación de alarmas de equipos distribuidos en red (tanto recursos HW como servicios de red). Se puede personalizar mediante la programación de plugins propios.
- Otras herramientas que permiten la monitorización de equipos distribuidos en red: Ganglia, Munin, Zabbix, Pandora FMS...



# Programa SarCheck

- Herramienta para
  - Análisis de prestaciones
  - Sintonización, planificación de la capacidad
- Sistemas Sun Solaris, HP-UX, AIX y Linux x86
- Basado en el monitor sar
- Utiliza gnuplot para generar gráficos
- Genera informes en formato HTML
  - Sección de recomendaciones
  - Sección de análisis de recursos
  - Sumario de estadísticas, etc.
- [www.sarcheck.com](http://www.sarcheck.com)

**SarCheck™**

*UNIX Performance Tuning Simplified...  
and Linux Performance Tuning too!*



# SarCheck®

UNIX Performance Tuning Simplified...  
and Linux Performance Tuning too!

**APTITUNE**  
CORPORATION

Kernel activity reported in English

Easily understand resource utilization

Text or HTML output with graphs

Kernel tuning recommendations

Capacity planning



**SarCheck** is a Linux & UNIX performance analysis and performance tuning tool. It is designed to help you with performance management on most Oracle Solaris, AIX, Linux, and HP-UX systems by making recommendations and explaining them with plain text, supporting graphs, and tables.

[Evaluation Copy](#) [Sample Reports](#) [SarCheck Video](#) [Platforms](#) [Customer Reviews](#) [About](#) [Contact](#)

• Solaris

• AIX

• Linux

• HP-UX

**Learn more about SarCheck...**

1

**Play the  
SarCheck Video**

2

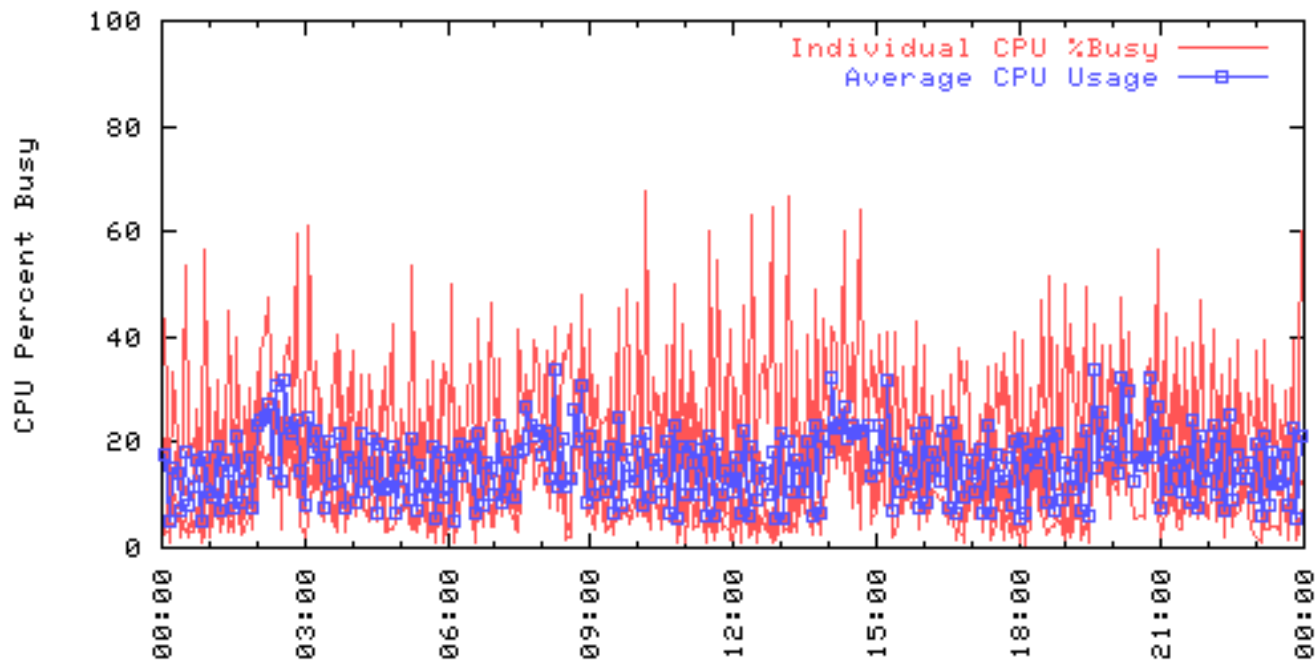
**View a Sample  
SarCheck Report**

3

**Try a FREE  
Evaluation Copy**

# Ejemplo informe generado por SarCheck

Average CPU utilization was only 15.7 percent. This indicates that spare capacity exists within the CPU. If any performance problems were seen during the monitoring period, they were not caused by a lack of CPU power. CPU utilization peaked at 34.00 percent from 08:10:01 to 08:15:01. A CPU upgrade is not recommended because the current CPU had significant unused capacity.



# Informe generado por SarCheck (cont.)

Change the virtual memory parameter 'swappiness' from 60 to 70. This controls the systems likelihood of swapping out memory pages which have not been recently used. Larger values will swap out unused pages of memory, speeding the time required to load new pages into memory.

Try to balance the disk I/O load across time or among other disk devices. The load on at least one disk was clearly excessive at peak times. By distributing the load over a greater time period or across other disk devices, the load will not be as likely to cause performance degradation.

Change the bdflush parameter 'nfract' from 50 to 55. This is the percentage of dirty buffers allowed in the buffer cache before the kernel flushes some of them.

Change the bdflush parameter 'nfract\_sync' from 80 to 85. This is the percentage of dirty buffers in the buffer cache before the kernel aggressively flushes them synchronously.

To change the value of the bdflush parameters immediately as described in the above recommendations, use the following command:

```
echo "55 500 8 0 500 3000 85 45 0" > /proc/sys/vm/bdflush
```

### **3.3. Monitorización a nivel de aplicación (profilers)**

# Monitorización a nivel de aplicación (*profiling*)

- Objetivo
  - Observar el comportamiento de una aplicación concreta con el fin de obtener **información para poder optimizar su código**.
- Información que pueden proporcionar las herramientas de análisis de aplicaciones (*profilers*):
  - ¿Cuánto tiempo tarda en ejecutarse el programa? ¿Qué parte de ese tiempo es de usuario y cuál de sistema? ¿Cuánto tiempo se pierde por las operaciones de E/S?
  - ¿En qué parte del código pasa la mayor parte de su tiempo de ejecución?
  - ¿Cuántas veces se ejecuta cada línea de programa?
  - ¿Cuántas veces se llama a un procedimiento y desde dónde?
  - ¿Cuánto tiempo tarda en ejecutarse (el código propio) de un procedimiento?
  - ¿Cuántos fallos de caché/página genera cada línea del programa?
  - ...



# time (/usr/bin/time)

Mide el tiempo de ejecución de un programa y muestra algunas estadísticas sobre su ejecución.

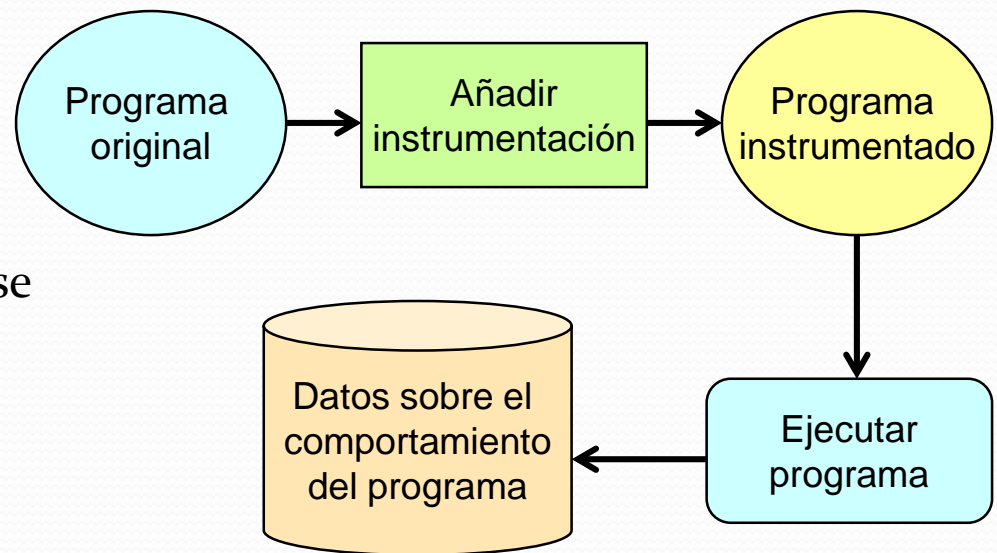
- **real**: tiempo total usado por el sistema (wall-clock CPU time = tiempo de respuesta).
- **user**: tiempo de CPU ejecutando en modo usuario.
- **sys**: tiempo de CPU ejecutando código del núcleo.
- Cambios de contexto voluntarios: al tener que esperar a una operación de E/S cede la CPU a otro proceso.
- Cambios involuntarios: Expira su “time slice”.
- Major page faults: requieren acceder al disco.

```
% time ./matr_mult2
real    0m4.862s
user    0m4.841s
sys     0m0.010s

% /usr/bin/time -v ./matr_mult2
User time (seconds): 4.86
System time (seconds): 0.01
Percent of CPU this job got: 99%
Elapsed (wall clock) time 0:04.90
Maximum RSS (kbytes): 48784
Major page faults: 0
Minor page faults: 3076
Voluntary context switches: 1
Involuntary context switches: 195
Swaps: 0
etc.
```

# Etapas típicas para usar un *profiler*

- Compilar el programa habilitando la recogida de información.
- Ejecutar el programa instrumentado.
  - Ejecución más lenta porque se ha de recoger y dejar la información en un fichero (*profile data*).
- Analizar la información contenida en el fichero de comportamiento.





# Monitor gprof

- Da información sobre el tiempo de ejecución y número de veces que se ejecuta cada función del programa.
- Utilización de gprof
  - Instrumentación en la compilación
    - `gcc prog.c -o prog -pg -g`
  - Ejecución del programa y recogida de información
    - `./prog`
    - La información recogida se deja en el fichero `gmon.out`
  - Visualización de la información referida a la ejecución del programa
    - `gprof prog > prog.gprof`

# Utilización del monitor gprof

```
float a=0.3; float b=0.8; float c=0.1;
void main(void) {
    unsigned long i;
    for (i=0;i<80000000;i++) a=a*b/(1+c);
    bucle1(); bucle1(); bucle2();
}
void bucle1(void) {
    unsigned long i;
    for (i=0;i<20000000;i++) {
        c=(c+c*c)/(1+a*c);c=a*b*c; }
    for (i=1;i<=5;i++) bucle3();
}
void bucle2(void) {
    unsigned long i;
    for (i=0;i<5000000;i++) {
        c=(c+c*c+c*c*c)/(1+a*c*c);
        c=a*b*c; }
    for (i=1;i<=8;i++) bucle3();
}
void bucle3(void) {
    unsigned long i;
    for (i=0;i<1000000;i++) {
        c=a*b*c; c=1/(a+b*c);
    }
}
```

## • Pasos

- Instrumentación (-pg) en la compilación
- Ejecución del programa y recogida de información
- Obtención de la información referida a la ejecución del programa

```
% gcc bucles.c -pg -o bucles
% ./bucles
% gprof bucles > bucles.prof
```

# Salida del monitor gprof

flat profile

% time	cumulative seconds	self seconds	calls	self s/call	total s/call	name
44.47	18.14	18.14				main
34.20	32.08	13.95	2	6.97	8.98	bucle1
17.71	39.30	7.22	18	0.40	0.40	bucle3
4.32	41.07	1.76	1	1.76	4.97	bucle2

- **% time:** Tanto por ciento del tiempo total de ejecución que usa el código propio de la subrutina (código propio es el que pertenece a la subrutina y no a las subrutinas a las que llama).
- **cumulative seconds:** La suma acumulada de los segundos consumidos por la subrutina y por las subrutinas que aparecen encima de ella en la tabla (código propio).
- **self seconds:** tiempo total de ejecución del código propio de la subrutina. Es el criterio por el que se ordena la tabla.
- **self s/call:** tiempo medio de ejecución del código propio por cada llamada a la subrutina.
- **total s/call:** tiempo medio total de ejecución por cada llamada a la subrutina (contando las subrutinas a las que llama)

# Salida del monitor gprof

call profile

index	% time	self	children	called	name
[1]	100.0	18.14	22.93		main [1]
		13.95	4.01	2/2	bucle1 [2]
		1.76	3.21	1/1	bucle2 [4]
-----					
[2]	43.7	13.95	4.01	2/2	main [1]
		13.95	4.01	2	bucle1 [2]
		4.01	0.00	10/18	bucle3 [3]
-----					
		3.21	0.00	8/18	bucle2 [4]
		4.01	0.00	10/18	bucle1 [2]
[3]	17.6	7.22	0.00	18	bucle3 [3]
-----					
		1.76	3.21	1/1	main [1]
[4]	12.1	1.76	3.21	1	bucle2 [4]
		3.21	0.00	8/18	bucle3 [3]
-----					

*bucle1* ha llamado a *bucle3* 10 de las 18 veces que es llamado *bucle3* en total.

# Salida del monitor gprof

call profile

index	% time	self	children	called	name
[1]	100.0	18.14	22.93		main [1]
		13.95	4.01	2/2	bucle1 [2]
		1.76	3.21	1/1	bucle2 [4]
-----					
[2]	43.7	13.95	4.01	2/2	main [1]
		13.95	4.01	2	bucle1 [2]
		4.01	0.00	10/18	bucle3 [3]
-----					
		3.21	0.00	8/18	bucle2 [4]
		4.01	0.00	10/18	bucle1 [2]
[3]	17.6	7.22	0.00	18	bucle3 [3]
-----					
		1.76	3.21	1/1	main [1]
[4]	12.1	1.76	3.21	1	bucle2 [4]
		3.21	0.00	8/18	bucle3 [3]
-----					

De las 18 veces que *bucle3* es llamado, 8 proceden de *bucle2* y 10 de *bucle1*

# Monitor gcov

- Aporta información sobre el número de veces que se ejecuta cada línea de código del programa.
- Utilización de gcov
  - Instrumentación en la compilación
    - `gcc prog.c -o prog -fprofile-arcs -ftest-coverage`
  - Ejecución del programa y recogida de información
    - `./prog`
    - La información recogida se deja en varios ficheros
  - Visualización de la información referida a la ejecución del programa
    - `gcov prog.c` (genera el fichero `prog.c.gcov`)

# Salida del monitor gcov

```
%gcc bucles2.c -o bucles2 -fprofile-arcs -ftest-coverage  
%./prog  
%gcov bucles2.c  
100.00% of 16 source lines executed in file bucles2.c  
Creating bucles2.c.gcov
```



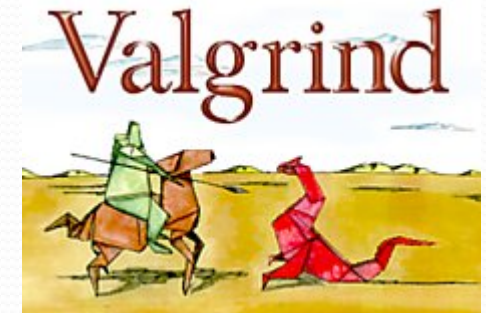
bucles.c.gcov

```
void main()  
{  
1   producto();  
1   producto();  
1   producto();  
1   division();  
1   division();  
1   atangente();  
1 }  
  
150000003 producto() {  
150000000   for (i=0;i<500000000;i++)  
3         c=a*b;  
        }  
  
600000002 division() {  
600000000   for (i=0;i<300000000;i++)  
2         c=a/b;  
        }  
  
300000001 atangente() {  
300000000   for (i=0;i<300000000;i++)  
1         c=atan(a);  
        }
```



# Otros Profilers: Valgrind

- Valgrind es un conjunto de herramientas para el análisis y mejora del código. Entre éstas, encontramos:
  - Memcheck, un detector de errores de memoria.
  - Cachegrind, un profiler de la memoria caché.
  - Callgrind, una versión más refinada de gprof.
- Valgrind puede analizar cualquier binario ya compilado (no necesita instrumentar el programa a partir de su código fuente). Valgrind actúa, esencialmente, como una máquina virtual que emula la ejecución de un programa ejecutable en un entorno aislado.
- Como desventaja, el sobrecoste computacional es muy alto. La emulación del programa ejecutable puede tardar decenas de veces más que la ejecución directa del programa de forma nativa.



# Otros profilers: V-Tune y CodeXL

- Son monitores de aplicaciones (profilers) que combinan tanto monitorización software mediante muestreo como monitorización hardware por eventos (hacen uso de **contadores hardware** disponibles en los últimos microprocesadores de Intel y AMD). Ambos programas funcionan tanto para Windows como para Linux, y permiten (adicionalmente a la información que nos proporciona gprof y gcov) obtener información sobre los fallos de caché, fallos de TLB, bloqueos/rupturas del cauce, fallos en la predicción de saltos, cerrojos y esperas entre hebras, etc. asociados a cada línea del programa.

