

PROGRAMACIÓN AVANZADA EN SHELL



Contenido

Introducción.....	4
Características principales de BASH.....	4
Cuándo utilizar el intérprete de mandatos.....	5
Modos de operación.....	6
La línea de mandatos.....	6
Edición y ejecución de un guión.....	6
Recomendaciones de programación.....	7
Redirecciones.....	9
Redirección de entrada.....	9
Redirecciones de salida.....	9
Combinación de redirecciones.....	10
Redirección de entrada/salida.....	11
Tuberías.....	12
Variables.....	15
Matrices.....	19
Expresiones.....	21
Expresiones aritméticas.....	21
Expresiones condicionales.....	23
Expresiones de ficheros.....	24
Expresiones comparativas numéricas.....	26
Expresiones comparativas de cadenas.....	28
Entrecorillado y expansión.....	29
Entrecorillado.....	29
Expansión.....	29
Expansión de llaves.....	30
Expansión de tilde.....	30
Expansión de parámetro o de variable.....	31
Sustitución de mandato.....	32
Expansión aritmética.....	33
Sustitución de proceso.....	33
División en palabras.....	34
Expansión de fichero.....	34
Programación estructurada.....	37
Listas de mandatos.....	37
Listas condicionales.....	37
Estructuras condicionales y selectivas.....	38
Estructuras condicionales.....	38
Estructura selectiva.....	40
Bucles.....	42
Bucles condicionales “mientras” y “hasta”.....	43
Bucle de selección interactiva.....	44
Funciones.....	45
Características especiales.....	47
Control de trabajos.....	48
Intérprete de uso restringido.....	49
AWK.....	51
Introducción.....	51
¿Para qué sirve awk?.....	51
Prerrequisitos.....	52
Parámetros en tiempo de ejecución.....	52
Primeros pasos con awk.....	54
Campos de entrada.....	55
Variables.....	56
Operadores.....	57
Awk en acción.....	59
Plantillas de prólogo y epílogo.....	59

Instrucciones de control	60
Funciones	61
La solución definitiva	62
Conclusión	65
Uso básico de SED	66
Qué es sed	66
Formato de uso	66
Ejemplos de sustitución	67
Ejemplos de Inserción	67
Ejemplos de Selección/Visualización	68
Ejemplos de Borrado	69
Uso de salida selectiva	70

Introducción.

El **intérprete de mandatos** o "*shell*" es la interfaz principal entre el usuario y el sistema, permitiéndole a aquél interactuar con los recursos de éste. El usuario introduce sus órdenes, el intérprete las procesa y genera la salida correspondiente.

Por lo tanto, un intérprete de mandatos de Unix es tanto una interfaz de ejecución de órdenes y utilidades, como un lenguaje de programación, que admite crear nuevas órdenes – denominadas **guiones** o "*shellscripts*" –, utilizando combinaciones de mandatos y estructuras lógicas de control, que cuentan con características similares a las del sistema y que permiten que los usuarios y grupos de la máquina cuenten con un entorno personalizado ^[2].

En Unix existen 2 familias principales de intérpretes de mandatos: los basados en el intérprete de Bourne (BSH, KSH o BASH) y los basados en el intérprete C (CSH o TCSH).

Este curso de formación pretende ser un guía para el usuario de Linux, que le permitirá comprender, ejecutar y empezar a programar en la *Shell*, haciendo referencia especialmente a **BASH** (*Bourne Again Shell*) –evolución de BSH, con características de KSH y CSH–, ya que es el intérprete de mandatos más utilizado en Linux e incluye un completo lenguaje para programación estructurada y gran variedad de funciones internas.

Características principales de BASH.

Los principales características del intérprete BASH son:

- Ejecución síncrona de órdenes (una tras otra) o asíncrona (en paralelo).
- Distintos tipos de redirecciones de entradas y salidas para el control y filtrado de la información.
- Control del entorno de los procesos.
- Ejecución de mandatos interactiva y desatendida, aceptando entradas desde teclado o desde ficheros..
- Proporciona una serie de órdenes internas para la manipulación directa del intérprete y su entorno de operación.
- Un lenguaje de programación de alto nivel, que incluye distintos tipos de variables, operadores, matrices, estructuras de control de flujo, entrecomillado, sustitución de valores y funciones.
- Control de trabajos en primer y segundo plano.
- Edición del histórico de mandatos ejecutados.
- Posibilidad de usar una "*shell*" para el uso de un entorno controlado.

Cuándo utilizar el intérprete de mandatos.

Como se ha indicado anteriormente, una “*shell*” de Unix puede utilizarse como interfaz para ejecutar órdenes en la línea de comandos o como intérprete de un lenguaje de programación para la administración del sistema.

El lenguaje de BASH incluye una sintaxis algo engorrosa, pero relativamente fácil de aprender, con una serie de órdenes internas que funcionan de forma similar a la línea de comandos. Un programa o guión puede dividirse en secciones cortas, cómodas de depurar, permitiendo realizar prototipos de aplicaciones más complejas.

Sin embargo, hay ciertas tareas que deben ser resueltas con otros intérpretes más complejos o con lenguajes compilados de alto nivel, tales como ^[4]:

- Procesos a tiempo real, o donde la velocidad es un factor fundamental.
- Operaciones matemáticas de alta precisión, de lógica difusa o de números complejos.
- Portabilidad de código entre distintas plataformas.
- Aplicaciones complejas que necesiten programación estructurada o proceso multihilvanado.
- Aplicaciones críticas para el funcionamiento del sistema.
- Situaciones donde debe garantizarse la seguridad e integridad del sistema, para protegerlo contra intrusión o vandalismo.
- Proyectos formados por componentes con dependencias de bloqueos.
- Proceso intensivo de ficheros, que requieran accesos directos o indexados.
- Uso de matrices multidimensionales o estructuras de datos (listas, colas, pilas, etc.).
- Proceso de gráficos.
- Manipulación de dispositivos, puertos o “*sockets*”.
- Uso de bibliotecas de programación o de código propietario

Modos de operación.

La línea de mandatos.

La línea de mandatos es el interfaz del usuario con el sistema, que permite personalizar el entorno de operación y ejecutar programas y guiones.

El formato típico de una línea consta de una orden y unos modificadores y parámetros opcionales, aunque puede incluir algunos caracteres especiales, que modifican el comportamiento típico.

```
Mandato [Modificador ...] [Parámetro ...]
```

Un caso especial es el de las líneas que comienzan por la almohadilla (#), que se consideran comentarios. También puede insertarse un comentario en mitad de una línea, a la derecha de una orden.

```
# Esto es un comentario  
ls -al      # lista el contenido del directorio actual
```

Edición y ejecución de un guión.

Un guión interpretado por BASH es un fichero de texto normal que consta de una serie de bloques de código formados por líneas de mandatos que se ejecutan en secuencia. El usuario debe tener los permisos de modificación (escritura) en el directorio –para crear un nuevo programa– o sobre el propio fichero, para modificar uno existente.

Como cualquier otro programa binario, el usuario debe tener permiso de ejecución en el fichero del guión, y se lanza tecleando su camino completo junto con sus opciones y parámetros. Asimismo, como veremos en el Capítulo 4, si el programa se encuentra en un directorio incluido en la variable de entorno **PATH**, sólo se necesita teclear el nombre, sin necesidad de especificar el camino.

El proceso completo de edición y ejecución de un guión es el siguiente:

```
vi prueba.sh          # o cualquier otro editor de textos  
chmod u+x prueba.sh  # activa el permiso de ejecución  
./prueba.sh          # ejecuta el guión  
prueba.sh             # si está en un directorio de $PATH
```

Existe una manera especial para ejecutar un guión, precediéndolo por el signo punto, que se utiliza para exportar todas las variables del programa al entorno de ejecución del usuario (para más información sobre las variables de entorno revisar el Capítulo 4). El siguiente ejemplo ilustra el modo de ejecutar

apachectl start	# Ejecución normal de un guión.
. miprofile	# Ejecución exportando las variables.

Un “*script*” puede –y debe– comenzar con la marca **#!** para especificar el camino completo y los parámetros del intérprete de mandatos que ejecutará el programa. Esta marca puede usarse para ejecutar cualquier intérprete instalado en la máquina (BASH, BSH, PERL, AWK, etc.).

El siguiente cuadro muestra un pequeño ejemplo de guión BASH.

#!/bin/bash	
# ejemplo1: informe de la capacidad de la cuenta	
echo "Usuario: \$USER"	
echo "Capacidad de la cuenta en KB:"	
du -ks \$HOME	# suma total del directorio del usuario

Recomendaciones de programación.

Como cualquier otro programa, un guión BASH puede requerir un cierto mantenimiento, que incluya modificaciones, actualizaciones o mejoras del código. Por lo tanto, el programador debe ser precavido y desarrollarlo teniendo en cuenta las recomendaciones de desarrollo típicas para cualquier programa.

Una práctica ordenada permite una verificación y comprensión más cómoda y rápida, para realizar las modificaciones de forma más segura y ayudar al usuario a ejecutar el programa correctamente. Para ello, seguir las siguientes recomendaciones.

- El código debe ser fácilmente legible, incluyendo espacios y sangrías que separen claramente los bloques de código
- Deben añadirse comentarios claros sobre el funcionamiento general del programa principal y de las funciones, que contengan: autor, descripción, modo de uso del programa, versión y fechas de modificaciones.
- Incluir comentarios para los bloques o mandatos importantes, que requieran cierta aclaración.
- Agregar comentarios y ayudas sobre la ejecución del programa.
- Depurar el código para evitar errores, procesando correctamente los parámetros de ejecución.
- No desarrollar un código excesivamente enrevesado, ni complicado de leer, aunque esto haga ahorrar líneas de programa.
- Utilizar funciones y las estructuras de programación más adecuadas para evitar repetir código reiterativo.

- Los nombres de variables, funciones y programas deben ser descriptivos, pero no ha de confundirse con otras de ellas, ni con los mandatos internos o externos; no deben ser ni muy largos ni muy cortos.
- Todos los nombres de funciones y de programas suelen escribirse en letras minúsculas, mientras que las variables acostumbran a definirse en mayúsculas.

Redirecciones.

Unix hereda 3 ficheros especiales del lenguaje de programación C, que representan las funciones de entrada y salida de cada programa. Estos son:

- **Entrada estándar:** procede del teclado; abre el fichero descriptor **0** (`stdin`) para lectura.
- **Salida estándar:** se dirige a la pantalla; abre el fichero descriptor **1** (`stdout`) para escritura.
- **Salida de error:** se dirige a la pantalla; abre el fichero descriptor **2** (`stderr`) para escritura y control de mensajes de error.

El proceso de **redirección** permite hacer una copia de uno de estos ficheros especiales hacia o desde otro fichero normal.

El fichero especial `/dev/null` se utiliza para descartar alguna redirección e ignorar sus datos.

Redirección de entrada.

Para realizar estas pruebas copiaremos nuestro fichero `/etc/passwd` en otro llamado `misusuarios`

17:46:04 03/04/2017

```
daw2016@daw2016-VirtualBox:~/scripts$ grupos=$(wc -l < /etc/group)
daw2016@daw2016-VirtualBox:~/scripts$ usuarios=$(wc -l < /etc/passwd)
daw2016@daw2016-VirtualBox:~/scripts$ echo "tu sistema tiene $usuarios y $grupos"
tu sistema tiene 48 y 81
```

La **redirección de entrada** sirve para abrir para lectura el archivo especificado usando un determinado número descriptor de fichero. Se usa la entrada estándar cuando el valor del descriptor es 0 o éste no se especifica.

El siguiente cuadro muestra el formato genérico de la redirección de entrada.

<code>[N]<Fichero</code>

La redirección de entrada se usa para indicar un fichero que contiene los datos que serán procesados por el programa, en vez de teclearlos directamente por teclado. Por ejemplo:

<code>miproceso.sh < fichdatos</code>
--

```
daw2016@daw2016-VirtualBox:~/scripts$ wc -l < misusuarios
```

Sin embargo, conviene recordar que la mayoría de las utilidades y filtros de Unix soportan los ficheros de entrada como parámetro del programa y no es necesario redirigirlos.

Redirecciones de salida.

De igual forma a los descrito en el apartado anterior, la **redirección de salida** se utiliza para abrir un fichero –asociado a un determinado número de descriptor– para operaciones de escritura.

Se reservan 2 ficheros especiales para el control de salida de un programa: la salida normal (con número de descriptor 1) y la salida de error (con el descriptor 2).

En la siguiente tabla se muestran los formatos genéricos para las redirecciones de salida.

Redirección	Descripción
<code>[N]> Fichero</code>	Abre el fichero de descriptor <i>N</i> para escritura. Por defecto se usa la salida estándar (<i>N</i> =1). Si el fichero existe, se borra; en caso contrario, se crea.
<code>[N]> Fichero</code>	Como en el caso anterior, pero el fichero debe existir previamente.
<code>[N]>> Fichero</code>	Como en el primer caso, pero se abre el fichero para añadir datos al final, sin borrar su contenido.
<code>&> Fichero</code>	Escribe las salida normal y de error en el mismo fichero.

El siguiente ejemplo crea un fichero con las salidas generadas para configurar, compilar e instalar una aplicación GNU.

```
daw2016@daw2016-VirtualBox:~/scripts$ cat misusuarios >copiausuarios
daw2016@daw2016-VirtualBox:~/scripts$ cat copiausuarios
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
games:x:5:60:games:/usr/games:/usr/sbin/nologin
```

```
configure      > aplic.sal
make            >> aplic.sal
make install   >> aplic.sal
```

Combinación de redirecciones.

Pueden combinarse más de una redirección sobre el mismo mandato o grupo de mandatos, interpretándose siempre de izquierda a derecha.

Por ejemplo si ejecutamos el siguiente comando nos visualizará una serie de mensajes de error

```
daw2016@daw2016-VirtualBox:~/scripts$ find / -name apache2
```

Si lo ejecutamos de la siguiente forma, vemos que los mensajes de error lo envía

```
daw2016@daw2016-VirtualBox:~/scripts$ find / -name apache2 2>/dev/null
/var/cache/apache2
/var/lib/apache2
/var/log/apache2
/run/apache2
```

Y si lo ejecutamos de la siguiente forma

```
daw2016@daw2016-VirtualBox:~/scripts$ find / -name apache2 >donde_esta_apache 2>/dev/null
```

Nos crea un fichero llamado donde_esta_apache y los errores no se visualizan

Realizar un ejercicio similar al anterior pero que los errores los envíe a otro archivo

Como ejercicio, se propone interpretar las siguientes órdenes:

```
ls -al /usr /tmp /noexiste >ls.sal 2>ls.err  
find /tmp -print >find.sal 2>/dev/null
```

Conviene hacer notar, que –siguiendo las normas anteriores– las 2 líneas siguientes son equivalentes y ambas sirven para almacenar las salidas normal y de error en el fichero indicado:

```
ls -al /var/* &>ls.txt  
ls -al /var/* >ls.txt 2>&1
```

Sin embargo, el siguiente ejemplo muestra 2 mandatos que no tienen por qué dar el mismo resultado, ya que las redirecciones se procesan de izquierda a derecha, teniendo en cuenta los posibles duplicados de descriptores hechos en líneas anteriores.

```
ls -al * >ls.txt 2>&1 # Salida normal y de error a "ls.txt". ls  
-al * 2>&1 >ls.txt # Asigna la de error a la normal anterior  
# (puede haberse redirigido) y luego  
# manda la estándar a "ls.txt".
```

Redirección de entrada/salida.

La **redirección de entrada y salida** abre el fichero especificada para operaciones de lectura y escritura y le asigna el descriptor indicado (0 por defecto). Se utiliza en operaciones para modificación y actualización de datos. El formato genérico es:

```
[N]<>Fichero
```

El siguiente ejemplo muestra una simple operación de actualización de datos en un determinado lugar del fichero ^[4].

```
echo 1234567890 > fich # Genera el contenido de "fich"  
exec 3<> fich # Abrir "fich" con descriptor 3 en E/S  
read -n 4 <&3 # Leer 4 caracteres  
echo -n , >&3 # Escribir coma decimal  
exec 3>&- # Cerrar descriptor 3  
cat fich # ➔ 1234,67890
```

Tuberías.

La **tubería** es una herramienta que permite utilizar la salida normal de un programa como entrada de otro, por lo que suele usarse en el filtrado y depuración de la información, siendo una de las herramientas más potentes de la programación con intérpretes Unix.

Pueden combinarse más de una tubería en la misma línea de órdenes, usando el siguiente formato:

```
Mandato1 | Mandato2 ...
```

Todos los dialectos Unix incluyen gran variedad de filtros de información. La siguiente tabla recuerda algunos de los más utilizados.

Mandato	Descripción
head	<p>Corta las primeras líneas de un fichero.</p> <p>11:27:03 03/04/2017</p> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ head -1 misusuarios root:x:0:0:root:/root:/bin/bash daw2016@daw2016-VirtualBox:~/scripts\$ head -2 misusuarios root:x:0:0:root:/root:/bin/bash daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin</pre>
tail	<p>Extrae las últimas líneas de un fichero.</p> <pre>daw2016@daw2016-VirtualBox:~\$ tail -f /var/log/apache2/access.log 127.0.0.1 - - [03/Apr/2017:05:32:10 -0400] "GET / HTTP/1.0" 200 275 "-" "Lynx/2. 8.8pre.4 libwww-FM/2.14 SSL-MM/1.4.1 GNUTLS/2.12.23" 127.0.0.1 - - [03/Apr/2017:05:34:07 -0400] "GET / HTTP/1.1" 200 312 "-" "Mozilla /5.0 (X11; Ubuntu; Linux x86_64; rv:47.0) Gecko/20100101 Firefox/47.0"</pre> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ tail -f /var/log/syslog Apr 3 05:25:01 daw2016-VirtualBox CRON[12287]: (daw2016) CMD (/home/daw2016/ 16/procesos.sh) Apr 3 05:25:01 daw2016-VirtualBox CRON[12284]: (CRON) info (No MTA insta led, discarding output) Apr 3 05:25:01 daw2016-VirtualBox CRON[12285]: (CRON) info (No MTA insta led, discarding output) Apr 3 05:26:01 daw2016-VirtualBox CRON[12399]: (root) CMD (/home/daw2016/ procesos.sh)</pre>
grep	<p>Muestra las líneas que contienen una determinada cadena de caracteres o cumplen un cierto patrón.</p> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ ps -ax grep apache2 1277 ? Ss 0:02 /usr/sbin/apache2 -k start 5062 ? SL 0:21 /usr/sbin/apache2 -k start 5063 ? SL 0:21 /usr/sbin/apache2 -k start</pre> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ grep root < misusuarios root:x:0:0:root:/root:/bin/bash daw2016@daw2016-VirtualBox:~/scripts\$ cat /etc/passwd grep /home syslog:x:101:104::/home/syslog:/bin/false usbmux:x:103:46:usbmux daemon,,,:/home/usbmux:/bin/false saned:x:108:115::/home/saned:/bin/false daw2016:x:1000:1000:daw2016,,,:/home/daw2016:/bin/bash mysql:x:998:998::/home/mysql: ana:x:1001:1001::/home/ana:</pre>

	<p>Corta columnas agrupadas por campos o caracteres.</p> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ cat /etc/passwd grep /home cut -d: -f1,6 syslog:/home/syslog usbmux:/home/usbmux saned:/home/saned daw2016:/home/daw2016 mysql:/home/mysql ana:/home/ana juan:/home/juan</pre>
cut	
	<p>Muestra o quita las líneas repetidas.</p> <p>12:18:53 03/04/2017</p> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ ps -axu cut -d " " -f1 sort uniq avahi colord daw2016 kernoops lp message+ nobody root rtkit syslog USER whoopsie www-data</pre>
uniq	
sort	<p>Lista el contenido del fichero ordenado alfabética o numéricamente.</p>
	<p>Cuenta líneas, palabras y caracteres de ficheros.</p> <p>12:12:26 03/04/2017</p> <pre>daw2016@daw2016-VirtualBox:~/scripts\$ wc -l < /etc/passwd 48 daw2016@daw2016-VirtualBox:~/scripts\$ n_usuarios=\$(wc -l </etc/passwd) daw2016@daw2016-VirtualBox:~/scripts\$ echo \$n_usuarios 48 daw2016@daw2016-VirtualBox:~/scripts\$</pre>
wc	<pre>daw2016@daw2016-VirtualBox:~/scripts\$ ps -axu cut -d " " -f1 sort uniq wc -l 13 daw2016@daw2016-VirtualBox:~/scripts\$ usuarios_con_procesos=\$(ps -axu cut -d " " -f1 sort uniq wc -l) daw2016@daw2016-VirtualBox:~/scripts\$ echo Hay \$usuarios_con_procesos en ejecucion Hay 13 en ejecucion</pre>
find	<p>Busca ficheros que cumplan ciertas condiciones y posibilita ejecutar operaciones con los archivos localizados</p>
sed	<p>Edita automáticamente un fichero. Ver anexo</p>
awk	<p>Procesa el fichero de entrada según las reglas de dicho lenguaje. Ver anexo</p>

El siguiente ejemplo muestra una orden compuesta que ordena todos los ficheros con extensión ".txt", elimina las líneas duplicadas y guarda los datos en el fichero "resultado.sal".

```
cat *.txt | sort | uniq >resultado.sal
```

La orden **tee** es un filtro especial que recoge datos de la entrada estándar y lo redirige a la salida normal y a un fichero especificado, tanto en operaciones de escritura como de añadidura. Esta es una orden muy útil que suele usarse en procesos largos para observar y registrar la evolución de los resultados.

El siguiente ejemplo muestra y registra el proceso de compilación e instalación de una aplicación GNU.

```
configure 2>&1 | tee aplic.sal  
make      2>&1 | tee -a aplic.sal  
make instal 2>&1 | tee -a aplic.sal
```

Se propone como ejercicio, interpretar la siguiente orden:

```
ls | tee salida | sort -r
```


Variables.

Al contrario que en otros lenguajes de programación, BASH no hace distinción en los tipos de datos de las variables; son esencialmente cadenas de caracteres, aunque –según el contexto– también pueden usarse con operadores de números enteros y condicionales. Esta filosofía de trabajo permite una mayor flexibilidad en la programación de guiones, pero también puede provocar errores difíciles de depurar^[4].

Una variable BASH se define o actualiza mediante operaciones de asignación, mientras que se hace referencia a su valor utilizando el símbolo del dólar delante de su nombre.

Suele usarse la convención de definir las variables en mayúsculas para distinguirlas fácilmente de los mandatos y funciones, ya que en Unix las mayúsculas y minúsculas se consideran caracteres distintos.

VAR1="Esto es una prueba	# asignación de una variable
VAR2=35	# asignar valor numérico
echo \$VAR1	# ➔ Esto es una prueba
echo "VAR2=\$VAR2"	# ➔ VAR2=35

Tipos de variables.

Las variables del intérprete BASH pueden considerarse desde los siguientes puntos de vista:

- Las **variables locales** son definidas por el usuario y se utilizan únicamente dentro de un bloque de código, de una función determinada o de un guión.
- Las **variables de entorno** son las que afectan al comportamiento del intérprete y al de la interfaz del usuario.
- Los **parámetros de posición** son los recibidos en la ejecución de cualquier programa o función, y hacen referencia a su orden ocupado en la línea de mandatos.
- Las **variables especiales** son aquellas que tienen una sintaxis especial y que hacen referencia a valores internos del proceso. Los parámetros de posición pueden incluirse en esta categoría.

Variables locales.

Las variables locales son definidas para operar en un ámbito reducido de trabajo, ya sea en un programa, en una función o en un bloque de código. Fuera de dicho ámbito de operación, la variable no tiene un valor preciso.

Una variable tiene un nombre único en su entorno de operación, sin embargo pueden –aunque no es nada recomendable– usarse variables con el mismo nombre en distintos bloques de código.

El siguiente ejemplo muestra los problemas de comprensión y depuración de código que pueden desatarse en caso de usar variables con el mismo nombre. En la primera fila se

presentan 2 programas que usan la misma variable y en la segunda, la ejecución de los programas (nótese que el signo > es el punto indicativo del interfaz de la “*shell*” y que lo tecleado por el usuario se representa en letra negrita).

#!/bin/bash # prog1 - variables prueba 1 VAR1=prueba echo \$VAR1	#!/bin/bash # prog2 - variables prueba 2 VAR1="otra prueba" echo \$VAR1
<pre>> echo \$VAR1 > prog1 prueba > prog2 otra prueba > prog1 prueba</pre>	

Por lo tanto, para asignar valores a una variable se utiliza simplemente su nombre, pero para hacer referencia a su valor hay que utilizar el símbolo dólar (\$). El siguiente ejemplo muestra los modos de referirse a una variable.

ERR=2	# Asigna 2 a la variable ERR.
echo ERR	# ➔ ERR (cadena "ERR").
echo \$ERR	# ➔ 2 (valor de ERR).
echo \${ERR}	# ➔ 2 (es equivalente).
echo "Error \${VAR}: salir"	# ➔ Error 2: salir

El formato `${Variable}` se utiliza en cadenas de caracteres donde se puede prestar a confusión o en procesos de sustitución de valores.

. Variables de entorno.

Al igual que cualquier otro proceso Unix, la “*shell*” mantiene un conjunto de variables que informan sobre su propio contexto de operación. El usuario –o un guión– puede actualizar y añadir variables exportando sus valores al entorno del intérprete (mandato **export**), lo que afectará también a todos los procesos hijos generados por ella. El administrador puede definir variables de entorno estáticas para los usuarios del sistema (como, por ejemplo, en el caso de la variable **IFS**).

La siguiente tabla presenta las principales variables de entorno.

Variable de entorno	Descripción	Valor por omisión
SHELL	Camino del programa intérprete de mandatos.	La propia <i>shell</i> .
PWD	Directorio de trabajo actual.	Lo modifica la <i>shell</i> .
OLDPWD	Directorio de trabajo anterior (equivale a <code>~-</code>).	Lo modifica la <i>shell</i> .
PPID	Identificador del proceso padre (PPID).	Lo modifica la <i>shell</i> .

IFS	Separador de campos de entrada (debe ser de sólo lectura).	ESP, TAB, NL.
HOME	Directorio personal de la cuenta.	Lo define <code>root</code> .
LOGNAME	Nombre de usuario que ejecuta la <i>shell</i> .	Activado por <code>login</code>
PATH	Camino de búsqueda de mandatos.	Según el sistema
LANG	Idioma para los mensajes.	
EDITOR	Editor usado por defecto.	
TERM	Tipo de terminal.	
PS1 ... PS4	Puntos indicativos primario, secundario, selectivo y de errores.	
FUNCNAME	Nombre de la función que se está ejecutando.	Lo modifica la <i>shell</i> .
LINENO	Nº de línea actual del guión (para depuración de código)	Lo modifica la <i>shell</i> .

Debe hacerse una mención especial a la variable **PATH**, que se encarga de guardar la lista de directorios con ficheros ejecutables. Si no se especifica el camino exacto de un programa, el sistema busca en los directorios especificados por **PATH**, siguiendo el orden de izquierda a derecha. El carácter separador de directorios es dos puntos.

El administrador del sistema debe establecer los caminos por defecto para todos los usuarios del sistema y cada uno de éstos puede personalizar su propio entorno, añadiendo sus propios caminos de búsqueda (si no usa un intérprete restringido).

Como ejercicio, interpretar la siguiente orden:

```
PATH=$PATH:/home/cdc/bin:/opt/oracle/bin
```

Recomendaciones de seguridad:

Siempre deben indicarse caminos absolutos en la definición de la variable **PATH** y, sobre todo, nunca incluir el directorio actual (.) ni el directorio padre (..).

Declarar la variable **IFS** de sólo lectura, para evitar intrusiones del tipo “caballos de Troya”.

Parámetros de posición.

Los parámetros de posición son variables especiales de BASH que contienen los valores de los parámetros que recibe un programa o una función. El número indica la posición de dicho parámetro en la llamada al código.

El 1^{er} parámetro se denota por la variable **\$1**, el 9º por **\$9** y a partir del 10º hay que usar la notación **\${Número}**. El mandato interno **shift** desplaza la lista de parámetros hacia la izquierda para procesar los parámetros más cómodamente. El nombre del programa se denota por la variable **\$0**.

Para observar el uso de parámetros posicionales y de variables locales, repasemos algunas partes del programa "usuario" (Apéndice C.4):

```
grep "^$1:" /etc/passwd
grep ":$GID:" /etc/group | cut -f1 -d:
```

1. Muestra la línea del usuario especificado en el 1^{er} parámetro recibido por el programa.
2. Presenta el nombre del grupo cuyo identificador se encuentra en la variable `GID`.

Variables especiales.

Las variables especiales informan sobre el estado del proceso, son tratadas y modificadas directamente por el intérprete, por lo tanto, son de sólo lectura. La siguiente tabla describe brevemente estas variables.

Variable especial	Descripción
<code>\$\$</code>	Identificador del proceso (PID).
<code>\$*</code>	Cadena con el contenido completo de los parámetros recibidos por el programa.
<code>\$@</code>	Como en el caso anterior, pero trata cada parámetro como un palabra diferente.
<code>\$#</code>	Número de parámetros.
<code>\$?</code>	Código de retorno del último mandato (0=normal, >0=error).
<code>\$!</code>	Último identificador de proceso ejecutado en segundo plano.
<code>\$_</code>	Valor del último argumento del comando ejecutado previamente.

La construcción `cat "$@"` puede usarse para procesar datos tanto de ficheros como de la entrada estándar ^[4].

La 1ª fila de la tabla del siguiente ejemplo muestra el código de un programa para convertir de minúsculas a mayúsculas. La 2ª línea indica cómo puede utilizarse el programa.

13:48:10 03/04/2017

```
daw2016@daw2016-VirtualBox:~/scripts$ echo hola | tr 'a-z' 'A-Z'
HOLA
```

```
#!/bin/bash
# mayusculas - convierte a mayúsculas usando ficheros o stdin
# Uso: mayusculas [ [<]fichero ]

cat "$@" | tr 'a-zñáéíóü' 'A-ZÑÁÉÍÓÜ'

> mayusculas datos.txt >datos.sal
> mayusculas <datos.txt >datos.sal
> mayusculas
Esto es una prueba de ejecución del
programa. ^D
ESTO ES UNA PRUEBA DE EJECUCIÓN DEL PROGRAMA.
```

En uso común de la variable `$$` es el de asignar nombres para ficheros temporales que permiten el uso concurrente del programa, ya que al estar asociada al PID del proceso, éste valor no se repetirá nunca al ejecutar simultáneamente varias instancias del mismo programa.

Matrices.

Una **matriz** (o “*array*”) es un conjunto de valores identificados por el mismo nombre de variable, donde cada una de sus celda cuenta con un índice que la identifica. Las matrices deben declararse mediante la cláusula interna **declare**, antes de ser utilizadas.

BASH soporta matrices de una única dimensión –conocidas también como **vectores**–, con un único índice numérico, pero sin restricciones de tamaño ni de orden numérico o continuidad.

Los valores de las celdas pueden asignarse de manera individual o compuesta. Esta segunda fórmula permite asignar un conjunto de valores a varias de las celdas del vector. Si no se indica el índice en asignaciones compuestas, el valor para éste por defecto es 0 o sumando 1 al valor previamente usado.

Utilizar los caracteres especiales `[@]` o `[*]` como índice de la matriz, supone referirse a todos los valores en su conjunto, con un significado similar al expresado en el apartado anterior.

El siguiente ejemplo describe la utilización de matrices.

```
declare -a NUMEROS                # Declarar la matriz.
NUMEROS=( cero uno dos tres )    # Asignación compuesta.
echo ${NUMEROS[2]}               # → dos
NUMEROS[4]="cuatro"              # Asignación simple.
echo ${NUMEROS[4]}               # → cuatro
NUMEROS=( [6]=seis siete [9]=nueve ) # asigna celdas 6, 7 y 9.
echo ${NUMEROS[7]}               # → siete
echo ${NUMEROS[*]}               # ==> uno dos tres cuatro seis siete nueve
```

Configuración del entorno.

El intérprete de mandados de cada cuenta de usuario tiene un entorno de operación propio, en el que se incluyen una serie de variables de configuración.

El administrador del sistema asignará unas variables para el entorno de ejecución comunes a cada grupo de usuarios –o a todos ellos–; mientras que cada usuario puede personalizar algunas de estas características en su perfil de entrada, añadiendo o modificando las variables.

Para crear el entorno global, el administrador crea un perfil de entrada común para todos los usuarios (archivo `/etc/bashrc` en el caso de BASH), donde –entre otros cometidos– se definen las variables del sistema y se ejecutan los ficheros de configuración propios para cada aplicación.

Estos pequeños programas se sitúan en el subdirectorio `/etc/profile.d`; debiendo existir ficheros propios de los intérpretes de mandatos basados en el de Bourne (BSH, BASH,

PKSH, etc.), con extensión **.sh**, y otros para los basados en el intérprete C (CSH, TCSH, etc.), con extensión **.csh**.

El proceso de conexión del usuario se completa con la ejecución del perfil de entrada personal del usuario (archivo **~/.bash_profile** para BASH). Aunque el administrador debe suministrar un perfil válido, el usuario puede retocarlo a su conveniencia. En el siguiente capítulo se presentan las variables de entorno más importantes usadas por BASH.

Expresiones.

El intérprete BASH permite utilizar una gran variedad de expresiones en el desarrollo de programas y en la línea de mandatos. Las distintas expresiones soportadas por el intérprete pueden englobarse en las siguientes categorías:

- **Expresiones aritméticas:** las que dan como resultado un número entero o binario.
- **Expresiones condicionales:** utilizadas por mandatos internos de BASH para su evaluar indicando si ésta es cierta o falsa.
- **Expresiones de cadenas:** aquellas que tratan cadenas de caracteres (se tratarán a fondo en el Capítulo 6).

Las expresiones complejas cuentan con varios parámetros y operadores, se evalúan de izquierda a derecha. Sin embargo, si una operación está encerrada entre paréntesis se considera de mayor prioridad y se ejecuta antes.

A modo de resumen, la siguiente tabla presenta los operadores utilizados en los distintos tipos de expresiones BASH.

Operadores aritméticos:	+ - * / % ++ --
Operadores de comparación:	== != < <= > >= -eq -nt -lt -le -gt -ge
Operadores lógicos:	! &&
Operadores binarios:	& ^ << >>
Operadores de asignación:	= *= /= %= += -= <<= >>= &= ^= =
Operadores de tipos de ficheros:	-e -b -c -d -f -h -L -p -s -t
Operadores de permisos:	-r -w -x -g -u -k -O -G -N
Operadores de fechas:	-nt -ot -et
Operadores de cadenas:	-z -n

Expresiones aritméticas.

Las expresiones aritméticas representan operaciones números enteros o binarios (booleanos) evaluadas mediante el mandato interno **let**. BASH no efectúa instrucciones algebraicas con números reales ni con complejos.

La valoración de expresiones aritméticas enteras sigue las reglas:

- Se realiza con números enteros de longitud fija sin comprobación de desbordamiento, esto es, ignorando los valores que sobrepasen el máximo permitido.

- La división por 0 genera un error que puede ser procesado.
- La prioridad y asociatividad de los operadores sigue las reglas del lenguaje C.

La siguiente tabla describe las operaciones aritméticas enteras y binarias agrupadas en orden de prioridad.

Operación	Descripción	Comentarios
<i>Var++</i> <i>Var--</i>	Post-incremento de variable. Post-decremento de variable.	La variable se incrementa o decrementa en 1 después de evaluarse su expresión.
<i>++Var</i> <i>--Var</i>	Pre-incremento de variable. Pre-decremento de variable.	La variable se incrementa o decrementa en 1 antes de evaluarse su expresión.
<i>+Expr</i> <i>-Expr</i>	Más unario. Menos unario.	Signo positivo o negativo de la expresión (por defecto se considera positivo).
<i>! Expr</i> <i>~ Expr</i>	Negación lógica. Negación binaria.	Negación de la expresión lógica o negación bit a bit.
<i>E1 ** E2</i>	Exponenciación.	E1 elevado a E2 ($E1^{E2}$).
<i>E1 * E2</i> <i>E1 / E2</i> <i>E1 % E2</i>	Multiplicación. División. Resto.	Operaciones de multiplicación y división entre números enteros.
<i>E1 + E2</i> <i>E1 - E2</i>	Suma.. Resta.	Suma y resta de enteros.
<i>Expr << N</i> <i>Expr >> N</i>	Desplazamiento binario a la izquierda. Desplazamiento binario a la derecha.	Desplazamiento de los bits un número indicado de veces.
<i>E1 < E2</i> <i>E1 <= E2</i> <i>E1 > E2</i> <i>E1 >= E2</i>	Comparaciones (menor, menor o igual, mayor, mayor o igual).	
<i>E1 = E2</i> <i>E1 != E2</i>	Igualdad. Desigualdad.	
<i>E1 & E2</i>	Operación binaria Y.	
<i>E1 ^ E2</i>	Operación binaria O Exclusivo.	
<i>E1 E2</i>	Operación binaria O.	
<i>E1 && E2</i>	Operación lógica Y.	
<i>E1 E2</i>	Operación lógica O.	

$E1 \ ? \ E2 \ : \ E3$	Evaluación lógica.	Si E1=cierto, se devuelve E2; si no, E3.
$E1 = E2$ $E1 \ Op = E2$	Asignación normal y con pre-operación (operadores válidos: \ast , $/$, $\%$, $+$, $-$, $<=$, $>=$, $\&$, \wedge , $ $).	Asigna el valor de E2 a E1. Si es específica un operador, primero se realiza la operación entre las 2 expresiones y se asigna el resultado ($E1 = E1 \ Op \ E2$).
$E1, E2$	Operaciones independientes.	Se ejecutan en orden.

El cuadro que se muestra a continuación ilustra el uso de las expresiones aritméticas.

let a=5	# Asignación a=5.
let b=\$a+3*9	# b=a+(3*9)=32.
echo "a=\$a, b=\$b"	# ➔ a=5, b=32
let c=\$b/(\$a+3)	# c=b/(a+3)=4.
let a+=c--	# a=a+c=9, c=c+1=5.
echo "a=\$a, c=\$c"	# ➔ a=9, c=5
let CTE=\$b/\$c, RESTO=\$b%\$c	# CTE=b/c, RESTO=resto(b/c).
echo "Cociente=\$CTE, Resto=\$RESTO"	# ➔ Cociente=6, Resto=2

Los números enteros pueden expresarse en bases numéricas distintas a la decimal (base por defecto). El siguiente ejemplo muestra los formatos de cambio de base.

let N=59	# Base decimal (0-9).
let N=034	# Base octal (0-7), empieza por 0.
let N=0x34AF	# Base hexadecimal (0-9A-F), empieza por 0x.
let N=[20#]G4H2	# Base 20 (especificada entre 2 y 64).

Expresiones condicionales.

Las **expresiones condicionales** son evaluadas por los mandatos internos del tipo **test**, dando como resultado un valor de cierto o de falso. Suelen emplearse en operaciones condicionales y bucles, aunque también pueden ser empleadas en órdenes compuestas.

Existen varios tipos de expresiones condicionales según el tipo de parámetros utilizado o su modo de operación:

- **Expresiones con ficheros**, que comparan la existencia, el tipo, los permisos o la fecha de ficheros o directorios.
- **Expresiones comparativas numéricas**, que evalúan la relación de orden numérico entre los parámetros.
- **Expresiones comparativas de cadenas**, que establecen la relación de orden alfabético entre los parámetros.

Todas las expresiones condicionales pueden usar el modificador de negación (`! Expr`) para indicar la operación inversa. Asimismo, pueden combinarse varias de ellas en una expresión compleja usando los operadores lógicos Y (`Expr1 && Expr2`) y O (`Expr1 || Expr2`).

Expresiones de ficheros.

Son expresiones condicionales que devuelven el valor de cierto si se cumple la condición especificada; en caso contrario da un valor de falso. Hay una gran variedad de expresiones relacionadas con ficheros y pueden agruparse en operaciones de tipos, de permisos y de comparación de fechas.

Conviene recordar que “todo en Unix es un fichero”, por eso hay bastantes operadores de tipos de ficheros. La siguiente tabla lista los formatos de estas expresiones.

Formato	Condición (cierto si...)
<code>-e Fich</code>	El fichero (de cualquier tipo) existe
<code>-s Fich</code>	El fichero no está vacío.
<code>-f Fich</code>	Es un fichero normal.
<code>-d Fich</code>	Es un directorio.
<code>-b Fich</code>	Es un dispositivo de bloques.
<code>-c Fich</code>	Es un dispositivo de caracteres.
<code>-p Fich</code>	Es una tubería.
<code>-h Fich</code> <code>-L Fich</code>	Es un enlace simbólico.
<code>-S Fich</code>	Es una “socket” de comunicaciones.
<code>-t Desc</code>	El descriptor está asociado a una terminal.
<code>F1 -ef F2</code>	Los 2 ficheros son enlaces hacia el mismo archivo.

Las condiciones sobre permisos establecen si el usuario que realiza la comprobación puede ejecutar o no la operación deseada sobre un determinado fichero. La tabla describe estas condiciones.

Formato	Condición (cierto si...)
<code>-r Fich</code>	Tiene permiso de lectura.
<code>-w Fich</code>	Tiene permiso de escritura (modificación).
<code>-x Fich</code>	Tiene permiso de ejecución/acceso.
<code>-u Fich</code>	Tiene el permiso SUID.

-g <i>Fich</i>	Tiene el permiso SGID.
-t <i>Fich</i>	Tiene permiso de directorio compartido o fichero en caché.

-o <i>Fich</i>	Es el propietario del archivo.
-G <i>Fich</i>	El usuario pertenece al grupo con el GID del fichero.

Las operaciones sobre fechas –descritas en la siguiente tabla– establecen comparaciones entre las correspondientes a 2 ficheros.

14:02:09 03/04/2017

```
daw2016@daw2016-VirtualBox:~/scripts$ if [ -r misusuarios ];then echo es de lectura;fi
es de lectura
```

Tened mucho cuidado con los espacios,, debe haber despues y antes de los corchetes.

Formato	Condición (cierto si...)
-N <i>Fich</i>	El fichero ha sido modificado desde al última lectura.
<i>F1</i> -nt <i>F2</i>	El fichero F1 es más nuevo que el F2.
<i>F1</i> -ot <i>F2</i>	El fichero F1 es más antiguo que el F2.

Veamos algunos ejemplos extraídos del fichero de configuración /etc/rc.d/rc.sysinit.

```
# Si /proa/mdstat y /etc/raidtab son ficheros; entonces ...
if [ -f /proc/mdstat -a -f /etc/raidtab ]; then
...
# Si el caminto representado por el contenido de la variable
# $afile es un directorio; entonces ...
if [ -d "$afile" ]; then
...
```

Se propone como ejercicio, interpretar la siguiente expresión.

```
if [ -e /proc/lvm -a -x /sbin/vgchange -a -f /etc/lvmtab ]; then
```

Expresiones comparativas numéricas.

Aunque los operadores de comparación para números ya se han comentado en el apartado anterior, la siguiente tabla describe los formatos para este tipo de expresiones.

Formato	Condición (cierto si...)
<i>N1</i> -eq <i>N2</i> <i>N1</i> -ne <i>N2</i> <i>N1</i> -lt <i>N2</i> <i>N1</i> -gt <i>N2</i>	Se cumple la condición de comparación numérica (respectivamente igual, distinto, menor y mayor).

Continuando con el ejemplo, se comentan algunas líneas del fichero de configuración /etc/rc.d/rc.sysinit.

14:07:20 03/04/2017

```
daw2016@daw2016-VirtualBox:~/scripts$ result=-1
daw2016@daw2016-VirtualBox:~/scripts$ if [ $result -gt 0 ];then echo mayor;fi
daw2016@daw2016-VirtualBox:~/scripts$ if [ $result -gt 0 ];then echo mayor;else echo menor;fi
menor
```

```
# Si la variable RESULT es > 0 y
#    /sbin/raidstart es ejecutable; entonces ...
if [ $RESULT -gt 0 -a -x /sbin/raidstart ]; then
...
```

```
# Si el código de la ejecución del 1er mandato es 0 y
#   el del 2º es distinto de 0; entonces ...
if grep -q /initrd /proc/mounts && \
    ! grep -q /initrd/loopfs /proc/mounts ; then
...
# Si la expresión de que existe el fichero /fastboot es cierta o
#   el código de salida del mandato es correcto; entonces ...
if [ -f /fastboot ] || \
    grep -iq "fastboot" /proc/cmdline 2>/dev/null ; then
...

```

Expresiones comparativas de cadenas.

También pueden realizarse comparaciones entre cadenas de caracteres. La tabla indica el formato de las expresiones.

Formato	Condición (cierto si...)
<i>Cad1</i> = <i>Cad2</i> <i>Cad1</i> != <i>Cad2</i>	Se cumple la condición de comparación de cadenas (respectivamente igual y distinto).
[-n] <i>Cad</i>	La cadena no está vacío (su longitud no es 0).
-z <i>Cad</i>	La longitud de la cadena es 0.

Como en los párrafos previos, se revisa parte del código del fichero `/etc/rc.d/rc.sysinit`.

```
# Si LOGNAME es una variable vacío o
#   tiene el valor "(none)"; entonces ...
if [ -z "$HOSTNAME" -o "$HOSTNAME" = "(none)" ]; then
...
# Si la variable $fastboot no tiene contenido y
#   la variable $ROOTFSTYPE no es "nfs"; entonces ...
if [ -z "$fastboot" -a "$ROOTFSTYPE" != "nfs" ]; then
...

```

Se propone como ejercicio interpretar la siguiente orden.

```
if [ -n "$PNP" -a -f /proc/isapnp -a -x /sbin/sndconfig ]; then

```

Entrecomillado y expansión.

Entrecomillado.

Cada uno de los caracteres especiales —**metacaracteres**— usados en BASH tienen un comportamiento especial, según la sintaxis del lenguaje. El **entrecomillado** es el procedimiento utilizado para modificar o eliminar el uso normal de dicho metacaracteres. Obsérvese el siguiente ejemplo.

```
# El ";" se usa normalmente para separar comandos.
echo Hola; echo que tal           # ➔ Hola
                                   #   que tal
# Usando entrecomillado pierde su función normal.
echo "Hola; echo que tal"         # ➔ Hola; echo que tal
```

Los 3 tipos básicos de entrecomillado definidos en BASH son ^[2]:

- **Carácter de escape** (`\Carácter`): mantiene el valor literal del carácter que lo precede; como último carácter de la línea, sirve para continuar la ejecución de una orden en la línea siguiente. La secuen
- **Comillas simples** (`"Cadena"`): siempre conserva el valor literal de cada uno de los caracteres de la cadena.
- **Comillas dobles** (`'Cadena'`): conserva el valor de literal de la cadena, excepto para los caracteres dólar (\$), comilla simple (') y de escape (`\$, \\", \', \"`, ante el fin de línea y secuencia de escape del tipo ANSI-C).

El entrecomillado con formato `$"Cadena"` se utiliza para procesos de traducción según el idioma expresado por la variable **LANG**. Si se utiliza el valor de idioma por defecto (C o POSIX), la cadena se trata normalmente con comillas dobles.

Veamos unos ejemplos:

```
echo "Sólo con permiso \"root\"" # ➔ Sólo con permiso "root"
echo 'Sólo con permiso \"root\"' # ➔ Sólo con permiso \"root\"
```

Expansión.

Como se lleva observando en todo el curso, la línea de comandos se divide en una serie de elementos que representan cierto significado en la semántica del intérprete. La **expansión** es un procedimiento especial que se realiza sobre dichos elementos individuales.

BASH dispone de 8 tipos de expansiones, que según su orden de procesamiento son ^[2]:

- **Expansión de llaves**: modifica la expresión para crear cadenas arbitrarias.
- **Expansión de tilde**: realiza sustituciones de directorios.

- **Expansión de parámetro y variable:** tratamiento general de variables y parámetros, incluyendo la sustitución de prefijos, sufijos, valores por defecto y otras operaciones con cadenas.
- **Sustitución de mandato:** procesa el mandato y devuelve su salida normal.
- **Expansión aritmética:** sustituye la expresión por su valor numérico.
- **Sustitución de proceso:** comunicación de procesos mediante tuberías con nombre de tipo cola (FIFO).
- **División en palabras:** separa la línea de mandatos resultante en palabras usando los caracteres de división incluidos en la variable `IFS`.
- **Expansión de fichero:** permite buscar patrones con comodines en los nombres de ficheros.

Expansión de llaves.

La expansión de llaves es el preprocesado de la línea de comandos que se ejecuta en primer lugar y se procesan de izquierda a derecha. Se utiliza para generar cadenas arbitrarias de nombre de ficheros, los cuales pueden o no existir, por lo tanto puede modificarse el número de palabras que se obtienen tras ejecutar la expansión. El formato general es el siguiente:

Formato	Descripción
<code>[Pre]{C1,C2[,...]}[Suf]</code>	El resultado es una lista de palabras donde se le añade a cada una de las cadenas de las llaves –y separadas por comas– un prefijo y un sufijo opcionales.

Para ilustrarlo, véanse los siguientes ejemplos.

```
echo a{b,c,d}e      # ➔ abe ace ade
mkdir $HOME/{bin,lib,doc} # Se crean los directorios:
                        # $HOME/bin, $HOME/lib y $HOME/doc.
```

Expansión de tilde.

Este tipo de expansión obtiene el valor de un directorio, tanto de las cuentas de usuarios, como de la pila de directorios accedidos. Los formatos válidos de la expansión de tilde son:

Formato	Descripción
<code>~[Usuario]</code>	Directorio personal del usuario indicado. Si no se expresa nada <code>\$HOME</code> .
<code>~+</code>	Directorio actual (<code>\$PWD</code>).
<code>~-</code>	Directorio anterior (<code>\$OLDPWD</code>).

Véase este pequeño programa:

```
#!/bin/bash
# capacidad - muestra la capacidad en KB de la cuenta del
#             usuario indicado
# Uso:  capacidad usuario

ls -ld ~$1
du -ks ~$1
```

Es recomendable definir un alias en el perfil de entrada del usuario para cambiar al directorio anterior, ya que la sintaxis del comando es algo engorrosa. Para ello, añadir la siguiente línea al fichero de configuración ~/.bashrc.

```
alias cda='cd ~-'
```

Expansión de parámetro o de variable.

Permite la sustitución del contenido de la variable siguiendo una amplia variedad de reglas. Los distintos formatos para la expansión de parámetros son ^[1]:

Formato	Descripción
<code>\${!Var}</code>	Se hace referencia a otra variable y se obtiene su valor (expansión indirecta).
<code>\${Parám:-Val}</code>	Se devuelve el parámetro; si éste es nulo, se obtiene el valor por defecto .
<code>\${Parám:=Val}</code>	Si el parámetro es nulo se le asigna el valor por defecto y se expande.
<code>\${Parám:?Cad}</code>	Se obtiene el parámetro; si es nulo se manda un mensaje de error .
<code>\${Parám:+Val}</code>	Se devuelve el valor alternativo si el parámetro no es nulo.
<code>\${Parám:Inic}</code> <code>\${Parám:Inic:Long}</code>	Valor de subcadena del parámetro, desde el punto inicial hasta el final o hasta la longitud indicada.
<code>\${!Pref*}</code>	Devuelve los nombres de variables que empiezan por el prefijo.
<code>\${#Parám}</code> <code>\${#Matriz[*]}</code>	El tamaño en caracteres del parámetro o en elementos de una matriz.
<code>\${Parám#Patrón}</code> <code>\${Parám##Patrón}</code>	Se elimina del valor del parámetro la mínima (o la máxima) comparación del patrón, comenzando por el principio del parámetro.
<code>\${Parám%Patrón}</code> <code>\${Parám%%Patrón}</code>	Se elimina del valor del parámetro la mínima (o la máxima) comparación del patrón, buscando por el final del parámetro.
<code>\${Parám/Patrón/Cad}</code> <code>\${Parám//Patrón/Cad}</code>	En el valor del parámetro se reemplaza por la cadena indicada la primera comparación del patrón (o todas las comparaciones).

BASH proporciona unas potentes herramientas para el tratamiento de cadenas, sin embargo la sintaxis puede resultar engorrosa y requiere de experiencia para depurar el código. Por lo tanto, se recomienda crear guiones que resulten fáciles de comprender, documentando claramente las órdenes más complejas.

Unos ejemplos para estudiar:

```
# Si el 1er parámetro es nulo, asigna el usuario que lo
ejecuta. USUARIO=${1:-`whoami`}

# Si no está definida la variable COLUMNS, el ancho es de
89. ANCHO=${COLUMNS:-80}

# Si no existe el 1er parámetro, pone mensaje de error y
sale. : ${1:? "Error: $0 fichero"}

# Obtiene la extensión de un fichero (quita hasta el
punto). EXT=${FICHERO##*.}

# Quita la extensión "rpm" al camino del fichero.
RPM=${FICHRPM%.rpm}

# Cuenta el nº de caracteres de la variable
CLAVE. CARACTERES=${#CLAVE}

# Renombra el fichero de enero a
Febrero. NUEVO=${ANTIGUO/enero/febrero}

# Añade nuevo elemento a la matriz
(matriz[tamaño]=elemento). matriz[${#matriz[*]}]="nuevo"
```

Por último, interpretar los siguientes ejercicios.

```
DATOS=ls -ld $1
TIPO=${DATOS:1}
PERM=${DATOS:2:9}

if [ ${#OPCIONES} -gt 1 ]; then ...

f="$f${1%$e}"
```

Sustitución de mandato.

Esta expansión sustituye el mandato ejecutado –incluyendo sus parámetros– por su salida normal, ofreciendo una gran potencia y flexibilidad de ejecución a un “*shellscript*”. Los formatos válidos son:

Formato	Descripción
<code>\$ (Mandato)</code>	Sustitución literal del mandato y sus parámetros.
<code>`Mandato`</code>	Sustitución de mandatos permitiendo caracteres de escape.

Cuando la sustitución de mandatos va en una cadena entre comillas dobles se evita que posteriormente se ejecute una expansión de ficheros.

El siguiente programa lista información sobre un usuario.

```
#!/bin/bash
# infous - lista información de un usuario.
# Uso:  infous usuario
TEMPORAL=`grep "^$1:" /etc/passwd 2>/dev/null`
USUARIO=`echo $TEMPORAL | cut -f1 -d:`
echo "Nombre de usuario: $USUARIO"
echo -n "Identificador (UID): " echo
$TEMPORAL | cut -f3 -d:
echo -n "Nombre del grupo primario: "
GID=`echo $TEMPORAL | cut -f4 -d:`
grep ":$GID:" /etc/group | cut -f1 -
d: echo "Directorio personal: "
ls -ld `echo $TEMPORAL | cut -f6 -d:`
```

Se propone como ejercicio interpretar la salida de la siguiente orden:

```
echo "`basename $0`: \"$USER\" sin permiso de ejecución." >&2
```

Expansión aritmética.

La expansión aritmética calcula el valor de la expresión indicada y la sustituye por el resultado de la operación. El formato de esta expansión es:

Formato	Descripción
$\$((Expresión))$ $\$[Expresión]$	Sustituye la expresión por su resultado.

Véase el siguiente ejemplo:

```
# Cuenta el n° de espacios para centrar una cadena
#      espacios = ( ancho_pantalla - longitud (cadena) ) /
2. ESPACIOS=$(( (ANCHO-${#CADENA})/2 ))
```

Y ahora un ejercicio algo complejo de leer; interpretar la siguiente línea:

```
if [ ${#cad} -lt ${${#1}-1} ]; then ...
```

Sustitución de proceso.

La sustitución de proceso permite utilizar un fichero especial de tipo cola para intercambiar información entre 2 procesos, uno que escribe en la cola y el otro que lee de ella en orden (el primero en llegar es el primero en salir). Los formatos válidos para esta expansión son:

formato	Descripción
<i>Fich</i> <(Lista) <i>Descr</i> <(Lista)	La lista de órdenes escribe en el fichero para que éste pueda ser leído por otro proceso.
<i>Fich</i> >(Lista) <i>Descr</i> >(Lista)	Cuando otro proceso escribe en el fichero, el contenido de éste se pasa como parámetro de entrada a la lista de órdenes.

División en palabras.

Una vez que se hayan realizado las expansiones previas, el intérprete divide la línea de entrada en palabras, utilizando como separadores los caracteres especificados en la variable de entorno **IFS**. Para evitar problemas de seguridad generados por un posible “Caballo de Troya”, el administrador debe declarar esta variable como de sólo lectura y establecer unos valores fijos para los separadores de palabra; que por defecto éstos son espacio, tabulador y salto de línea. Una secuencia de varios separadores se considera como un único delimitador.

Por ejemplo, si se ejecuta el siguiente mandato:

```
du -ks $HOME
```

el intérprete realiza las sustituciones y –antes de ejecutar la orden– divide la línea en las siguientes palabras.

```
"du" "-ks" "/home/ramon" "du" "-ks" "/home/ramon"
```

Expansión de fichero.

Si algunas de las palabras obtenidas tras la división anterior contiene algún caracteres especial conocido como **comodín** (*, ? o []), ésta se trata como un patrón que se sustituye por la lista de nombres de ficheros que cumplen dicho patrón, ordena da alfabéticamente^[2]. El resto de caracteres del patrón se tratan normalmente.

Los patrones válidos son:

Formato	Descripción
*	Equivale a cualquier cadena de caracteres, incluida una cadena nula.
?	Equivale a cualquier carácter único.
[Lista]	Equivale a cualquier carácter que aparezca en la lista. Pueden incluirse rangos de caracteres separados por guión (-). Si el primer carácter de la lista es ^, se comparan los caracteres que no formen parte de ella.

La siguiente tabla describe algunos ejemplos.

<pre># Listar los ficheros terminados en .rpm ls *.rpm # Listar los ficheros que empiecen por letra minúscula y tengan</pre>
--

```
#      extensión
.rpm ls [a-z]*.rpm
# Listar los ficheros que empiezan por ".b", ".x" y
".X" ls .[bxX]*
# Listar los ficheros cuya extensión tenga 2
caracteres ls *.*??
```

Por último, se propone como ejercicio evaluar la siguiente orden:

```
cd /var/log; tar cvf copialog.tar syslog.[0-9] messages.[0-9]
```

Programación estructurada.

Las estructuras de programación se utilizan para generar un código más legible y fiable. Son válidas para englobar bloques de código que cumplen un cometido común, para realizar comparaciones, selecciones, bucles repetitivos y llamadas a subprogramas.

Listas de mandatos.

Los mandatos BASH pueden agruparse en bloques de código que mantienen un mismo ámbito de ejecución. La siguiente tabla describe brevemente los aspectos fundamentales de cada lista de órdenes.

Lista de órdenes	Descripción
<i>Mandato &</i>	Ejecuta el mandato en 2º plano. El proceso tendrá menor prioridad y no debe ser interactivo..
<i>Man1 Man2</i>	Tubería. Redirige la salida de la primera orden a la entrada de la segundo. Cada mandato es un proceso separado.
<i>Man1; Man2</i>	Ejecuta varios mandatos en la misma línea de código.
<i>Man1 && Man2</i>	Ejecuta la 2ª orden si la 1ª lo hace con éxito (su estado de salida es 0).
<i>Man1 Man2</i>	Ejecuta la 2ª orden si falla la ejecución de la anterior (su código de salida no es 0).
<i>(Lista)</i>	Ejecuta la lista de órdenes en un subproceso con un entorno común.
<i>{ Lista; }</i>	Bloque de código ejecutado en el propio intérprete.
<i>Línea1 \</i> <i>Línea2</i>	Posibilita escribir listas de órdenes en más de una línea de pantalla. Se utiliza para ejecutar mandatos largos.

Ya se ha comentado previamente la sintaxis de algunas de estas combinaciones de mandatos. Sin embargo, el siguiente epígrafe presta atención especial a las listas de órdenes condicionales.

Listas condicionales.

Los **mandatos condicionales** son aquellos que se ejecutan si una determinada orden realiza correctamente –lista Y (**&&**)– o si se produce algún error –lista O (**||**)–.

A continuación se comentan algunos ejemplos, sacados del fichero `/etc/rc.d/rc.sysinit`.

```
# Si se ejecuta correctamente "converquota", se ejecuta
"rm". /sbin/convertquota -u / && rm -f /quota.user

# Si da error la ejecución de "grep", se ejecuta "modprobe".
grep 'hid' /proc/bus/usb/drivers || modprobe hid 2> /dev/null
```

Se propone como ejercicio, interpreta la siguiente lista de órdenes.

```
loadkeys $KEYMAP < /dev/tty0 > /dev/tty0 2>/dev/null && \ success
"Loading def keymap" || failure "Loading def keymap"
```

Estructuras condicionales y selectivas.

Ambas estructuras de programación se utilizan para escoger un bloque de código que debe ser ejecutado tras evaluar una determinada condición. En la estructura condicional se opta entre 2 posibilidades, mientras que en la selectiva pueden existir un número variable de opciones.

Estructuras condicionales.

La **estructura condicional** sirve para comprobar si se ejecuta un bloque de código cuando se cumple una cierta condición. Pueden anidarse varias estructuras dentro del mismo bloques de código, pero siempre existe una única palabra **fi** para cada bloque condicional.

La tabla muestra cómo se representan ambas estructuras en BASH.

Estructura de programación	Descripción
<pre>if Expresión; then Bloque; fi</pre>	Estructura condicional simple: se ejecuta la lista de órdenes si se cumple la expresión. En caso contrario, este código se ignora.
<pre>if Expresión1; then Bloque1; [elif Expresión2; then Bloque2; ...] [else BloqueN;] fi</pre>	Estructura condicional compleja: el formato completo condicional permite anidar varias órdenes, además de poder ejecutar distintos bloques de código, tanto si la condición de la expresión es cierta, como si es falsa.

Aunque el formato de codificación permite incluir toda la estructura en una línea, cuando ésta es compleja se debe escribir en varias, para mejorar la comprensión del programa. En caso de

teclear la orden compleja en una sola línea debe tenerse en cuenta que el carácter separador (;) debe colocarse antes de las palabras reservadas: **then**, **else**, **elif** y **fi**.

Hay que resaltar la versatilidad para teclear el código de la estructura condicional, ya que la palabra reservada **then** puede ir en la misma línea que la palabra **if**, en la línea siguiente sola o conjuntamente con la primera orden del bloque de código, lo que puede aplicarse también a la palabra **else**).

Puede utilizarse cualquier expresión condicional para evaluar la situación, incluyendo el código de salida de un mandato o una condición evaluada por la orden interna **test**. Este último caso se expresa colocando la condición entre corchetes (formato: [Condición]).

Véanse algunos sencillos ejemplos de la estructura condicional simple extraídos del “*script*” `/etc/rc.d/rc.sysinit`. Nótese la diferencia en las condiciones sobre la salida normal de una orden –expresada mediante una sustitución de mandatos– y aquellas referidas al estado de ejecución de un comando (si la orden se ha ejecutado correctamente o si se ha producido un error).

```
# La condición más simple escrita en una línea:
# si RESULT>0; entonces rc=1
if [ $RESULT -gt 0 ]; then rc=1; fi
#
# La condición doble:
# si la variable HOSTNAME es nula o vale "(none)"; entonces ...
if [ -z "$HOSTNAME" -o "$HOSTNAME" = "(none)" ];
    then HOSTNAME=localhost
fi
#
# Combinación de los 2 tipos de condiciones:
# si existe el fichero /fastboot o la cadena "fastboot" está
#   en el fichero /proc/cmdline; entonces ...
if [ -f /fastboot ] || grep -iq "fastboot" /proc/cmdline
    \ 2>/dev/null; then
    fastboot=yes
fi
```

Obsérvense los ejemplos para las estructuras condicionales complejas, basados también en el programa de configuración `/etc/rc.d/rc.sysinit`.

```
# Estructura condicional compleja con 2 bloques:
# si existe el fichero especificado, se ejecuta; si no, se da
#   el valor "no" a la variable NETWORKING
if [ -f /etc/sysconfig/network
]; then
.
/etc/sysconfig/network else
NETWORKING=no
fi
# Estructura anidada:
# si rc=0; entonces ...; si no, si rc=1; entonces ...; en caso
#   contrario; no se hace nada.
if [ "$rc" = "0" ];
    then success
    "$STRING" echo
elif [ "$rc" = "1" ];
    then passed
    "$STRING" echo
fi
```

Por último, evaluar el siguiente ejercicio:

```
if [ -f /etc/sysconfig/console/default.kmap ];
then KEYMAP=/etc/sysconfig/console/default.kmap
else
  if [ -f /etc/sysconfig/keyboard ]; then
    .
    /etc/sysconfig/keyboard fi
  if [ -n "$KEYTABLE" -a -d "/usr/lib/kbd/keymaps" -o -
d "/lib/kbd/keymaps" ]; then
    KEYMAP=$KEYTABLE
  fi
fi
fi
```

Estructura selectiva.

La **estructura selectiva** evalúa la condición de control y, dependiendo del resultado, ejecuta un bloque de código determinado. La siguiente tabla muestra el formato genérico de esta estructura.

Estructura de programación	Descripción
<pre>case Variable in ([Patrón1) Bloque1 ;; ... esac</pre>	<p>Estructura selectiva múltiple: si la variable cumple un determinado patrón, se ejecuta el bloque de código correspondiente. Cada bloque de código acaba con ";;".</p> <p>La comprobación de patrones se realiza en secuencia.</p>

Las posibles opciones soportadas por la estructura selectiva múltiple se expresan mediante patrones, donde puede aparecer caracteres comodines, evaluándose como una expansión de ficheros, por lo tanto el patrón para representar la opción por defecto es el asterisco (*). Dentro de una misma opción pueden aparecer varios patrones separados por la barra vertical (|), como en una expresión lógica O.

Si la expresión que se comprueba cumple varios patrones de la lista, sólo se ejecuta el bloque de código correspondiente al primero de ellos, ya que la evaluación de la estructura se realiza en secuencia.

Obsérvense el siguientes ejemplos:

```
# Según el valor de la variable UTC:
# - si es "yes" o
  "true", ... CLOCKFLAGS="$CLOCKFLAGS --utc";
            CLOCKDEF="$CLOCKDEF (utc)";
            # - si es "no"
  o "false", ...
case "$UTC" in
  yes|true) CLOCKFLAGS="$CLOCKFLAGS --localtime";
            CLOCKDEF="$CLOCKDEF (localtime)";
;;
```



```
no|false)
```

```
;;
```

```
esac
```

Y, como en los casos anteriores, describir el modo de ejecución de la siguiente estructura selectiva:

```
case "$SO" in
    AIX)    echo -n "$US: "
            lsuser -ca expires $US|fgrep -v "#"|cut -f2 -d:`"
            ;;
    SunOS)  echo "$US: `logins -aol $US|cut -f7 -d:`"
            ;;
    Linux)  echo "$US: `chage -l $US|grep Account|cut -f2 -d:`"
            ;;
    *)      echo "Sistema operativo desconocido" ;;
esac
```

Bucles.

Los bucles son estructuras reiterativas que se ejecutan repetitivamente, para no tener que teclear varias veces un mismo bloque de código. Un bucle debe tener siempre una condición de salida para evitar errores provocados por bucles infinitos.

La siguiente tabla describe las 2 órdenes especiales que pueden utilizarse para romper el modo de operación típico de un bucle.

Orden	Descripción
break	Ruptura inmediata de un bucle (debe evitarse en programación estructurada para impedir errores de lectura del código).
continue	Salto a la condición del bucle (debe evitarse en programación estructurada para impedir errores de lectura del código).

Los siguientes puntos describen los distintos bucles que pueden usarse tanto en un guión como en la línea de mandatos de BASH.

Bucles genéricos.

Los bucles genéricos de tipos “para cada” ejecutan el bloque de código para cada valor asignado a la variable usada como índice del bucle o a su expresión de control. Cada iteración debe realizar una serie de operaciones donde dicho índice varíe, hasta la llegar a la condición de salida.

El tipo de bucle **for** más utilizado es aquél que realiza una iteración por cada palabra (o cadena) de una lista. Si se omite dicha lista de valores, se realiza una iteración por cada parámetro posicional.

Por otra parte, BASH soporta otro tipo de bucle iterativo genérico similar al usado en el lenguaje de programación C, usando expresiones aritméticas. El modo de operación es el siguiente:

- Se evalúa la primera expresión aritmética antes de ejecutar el bucle para dar un valor inicial al índice.
- Se realiza una comprobación de la segunda expresión aritmética, si ésta es falsa se ejecutan las iteraciones del bucle. Siempre debe existir una condición de salida para evitar que el bucle sea infinito.
- como última instrucción del bloque se ejecuta la tercera expresión aritmética –que debe modificar el valor del índice– y se vuelve al paso anterior.

La siguiente tabla describe los formatos de los bucles iterativos genéricos (de tipo “para cada”) interpretados por BASH.

Bucle	Descripción
for <i>Var</i> [in <i>Lista</i>]; do <i>Bloque</i> done	Bucle iterativo: se ejecuta el bloque de mandatos del bucle sustituyendo la variable de evaluación por cada una de las palabras incluidas en la lista.
For ((<i>Exp1</i> ; <i>Exp2</i> ; <i>Exp3</i>)) do <i>Bloque</i> done	Bucle iterativo de estilo C: se evalúa <i>Exp1</i> , mientras <i>Exp2</i> sea cierta se ejecutan en cada iteración del bucle el bloque de mandatos y <i>Exp3</i> (las 3 expresiones deben ser aritméticas).

Véanse algunos ejemplos:

```
# Se asigna a la variable "library" el camino de cada uno de
# los archivos indicados en la expansión de ficheros y se
# realizan las operaciones indicadas en el bloque de código. for
library in /lib/kernel/$(uname -r)/libredhat-kernel.so* ; do
    ln -s -f $library /lib/
    ldconfig -n /lib/
done
...
# Se establece un contador de hasta 20 iteraciones para
# ejecutar el bucle.
for (( times = 1; times < 20; times++ )); do /usr/sbin/rpcinfo
    -p | grep ypbind > /dev/null 2>&1 && \
        ypwhich > /dev/null 2>&1
done
```

Y el ejercicio a evaluar:

```
for US in `cut -sf2 -d: /home/cdc/*.lista`; do
    grep "^$US:" /etc/shadow | cut -sf1-2 -d:
>>$FICHTEMP done
```

Bucles condicionales “mientras” y “hasta”.

Los bucles condicionales evalúan la expresión en cada iteración del bucle y dependiendo del resultado se vuelve a realizar otra iteración o se sale a la instrucción siguiente.

La siguiente tabla describe los formatos para los 2 tipos de bucles condicionales soportados por el intérprete BASH.

Bucle	Descripción
while <i>Expresión</i> ; do <i>Bloque</i> done	Bucle iterativo “mientras” : se ejecuta el bloque de órdenes mientras que la condición sea cierta. La expresión de evaluación debe ser modificada en algún momento del bucle para poder salir.
until <i>Expresión</i> ; do <i>Bloque</i> done	Bucle iterativo “hasta” : se ejecuta el bloque de código hasta que la condición sea cierta. La expresión de evaluación debe ser modificada en algún momento del bucle para poder salir.

Ambos bucles realizan comparaciones inversas y pueden usarse indistintamente, aunque se recomienda usar aquél que necesite una condición más sencilla o legible, intentando no crear expresiones negativas. Véase el siguiente ejemplo:

```
# Mientras haya parámetros que procesar, ...
while [ $# != 0 ] ; do
    processdir
    "$1" shift
done
```

Se proponen como ejercicios convertir este bucle **while** en un bucle **until** y en uno **for** de estilo C.

```
i=5
while [ $i -ge 0 ] ; do
    if [ -f /var/log/ksyms.$i ] ; then
        mv /var/log/ksyms.$i /var/log/ksyms.$((i+1))
    fi i=$((i-1))
done
```

Bucle de selección interactiva.

La estructura **select** no es propiamente dicho un bucle de programación estructurada, ya que se utiliza para mostrar un menú de selección de opciones y ejecutar el bloque de código correspondiente a la selección escogida. En caso de omitir la lista de palabras, el sistema presenta los parámetros posicionales del programa o función. Este tipo de bucles no suele utilizarse.

La siguiente tabla describe el formato del bucle interactivo.

Bucle	Descripción
select <i>Var</i> [in <i>Lista</i>]; do <i>Bloque1</i> ... done	Bucle de selección interactiva : se presenta un menú de selección y se ejecuta el bloque de código correspondiente a la opción elegida. El bucle se termina cuando se ejecuta una orden break . La variable <code>PS3</code> se usa como punto indicativo.

Funciones.

Una función en BASH es una porción de código declarada al principio del programa, que puede recoger parámetro de entrada y que puede ser llamada desde cualquier punto del programa principal o desde otra función, tantas veces como sea necesario.

El uso de funciones permite crear un código más comprensible y que puede ser depurado más fácilmente, ya que evita posibles errores tipográficos y repeticiones innecesarias.

Los parámetros recibidos por la función se tratan dentro de ella del mismo modo que los del programa principal, o sea los parámetros posicionales de la función se corresponden con las variables internas \$0, \$1, etc.

El siguiente cuadro muestra los formatos de declaración y de llamada de una función..

Declaración	LLamada
<pre>[function] NombreFunción () { Bloque ... [return [Valor]] ... }</pre>	<pre>NombreFunción [Parámetro1 ...]</pre>

La función ejecuta el bloque de código encerrado entre sus llaves y –al igual que un programa– devuelve un valor numérico. En cualquier punto del código de la función, y normalmente al final, puede usarse la cláusula **return** para terminar la ejecución y opcionalmente indicar un código de salida.

Las variables declaradas con la cláusula **local** tienen un ámbito de operación interno a la función. El resto de variables pueden utilizarse en cualquier punto de todo el programa. Esta característica permite crear funciones recursivas sin que los valores de las variables de una llamada interfieran en los de las demás.

En el ejemplo del siguiente cuadro se define una función de nombre *salida*, que recibe 3 parámetros. El principio del código es la definición de la función (la palabra **function** es opcional) y ésta no se ejecuta hasta que no se llama desde el programa principal. Asimismo, la variable `TMPPGREP` se declara en el programa principal y se utiliza en la función manteniendo su valor correcto.

```
#!/bin/bash  
# comprus - comprueba la existencia de usuarios en listas y en  
#           el archivo de claves (normal y NIS).  
#           Uso: comprus ? | cadena  
#           ?: ayuda.  
# Ramón Gómez - Septiembre 1.998.  
  
# Rutina de impresión.  
# Parámetros:  
#     1 - texto de cabecera.  
#     2 - cadena a buscar.
```

```

# 3 - archivo de búsqueda. salida
()
{
    if egrep "$2" $3 >$TMPGREP 2>/dev/null; then
        echo " $1:"
        cat $TMPGREP
    fi
}

## PROGRAMA PRINCIPAL ##

TMPGREP=/tmp/grep$$
DIRLISTAS=/home/cdc/listas

if [ "x$" = "x?" ]
then
    echo "
Uso:      `basename $0` ? | cadena
Propósito: `basename $0`: Búsqueda de usuarios. cadena:
           expresión regular a buscar.
"
    exit 0
fi
if [ $# -ne 1 ]; then
    echo "`basename $0`: Parámetro incorrecto.
Uso: `basename $0` ? | cadena
     ?: ayuda" >&2
    exit 1
fi
echo

for i in $DIRLISTAS/*.lista; do
    salida "$1" "`basename $i | sed 's/\.lista//'\`" "$i"
done
salida "$1" "passwd" "/etc/passwd" [ -
e "$TMPGREP" ] && rm -f $TMPGREP

```

Para terminar el capítulo se propone desarrollar en BASH:

- Una función que reciba un único parámetro, una cadena de caracteres, y muestre el último carácter de dicha cadena. Debe tenerse en cuenta que el primer carácter de una cadena tiene como índice el valor 0.
- Una función que reciba una cadena de caracteres y que la imprima centrada en la pantalla. Tener en cuenta que el ancho del terminal puede estar definido en la variable de entorno COLUMNS, en caso contrario usar un valor por defecto de 80 columnas.

Características especiales

En última instancia se van a describir algunas de las características especiales que el intérprete BASH añade a su lenguaje para mejorar la funcionalidad de su propia interfaz y de los programas que interpreta.

Seguidamente se enumeran algunas de las características especiales de BASH y en el resto de puntos de este capítulo se tratan en mayor profundidad las que resultan más interesantes:

- Posibilidad de llamar al intérprete BASH con una serie de opciones que modifican su comportamiento normal.
- Mandatos para creación de programas interactivos.
- Control de trabajos y gestión de señales.
- Manipulación y personalización del punto indicativo.
- Soporte de alias de comandos.
- Gestión del histórico de órdenes ejecutadas.
- Edición de la línea de mandatos.
- Manipulación de la pila de últimos directorios visitados.
- Intérprete de uso restringido, con características limitadas.
- Posibilidad de trabajar en modo compatible con la norma POSIX 1003.2 ^[2].

Programas interactivos.

BASH proporciona un nivel básico para programar “*shellscripts*” interactivos, soportando instrucciones para solicitar y mostrar información al usuario.

Las órdenes internas para dialogar con el usuario se describen en la siguiente tabla.

Mandato	Descripción
read [-p <i>Cadena</i>] [<i>Var1</i> ...]	Asigna la entrada a variables: lee de la entrada estándar y asigna los valores a las variables indicadas en la orden. Puede mostrarse un mensaje antes de solicitar los datos. Si no se especifica ninguna variable, REPLY contiene la línea de entrada.
echo [-n] <i>Cadena</i>	Muestra un mensaje: manda el valor de la cadena a la salida estándar; con la opción -n no se hace un salto de línea.

<code>printf</code> <i>Formato Parám1 ...</i>	Muestra un mensaje formateado: equivalente a la función <code>printf</code> del lenguaje C, manda un mensaje formateado a la salida normal, permitiendo mostrar cadena y números con una longitud determinada.
---	---

Véanse los siguientes ejemplos.

```
# Las siguientes instrucciones son equivalentes y muestran
# un mensaje y piden un valor
echo -n "Dime tu nombre:
" read NOMBRE
#
read -p "Dime tu nombre: " NOMBRE
...
# Muestra los usuarios y nombres completos
# (la modificación de la variable IFS sólo afecta al
bucle) while IFS=: read usuario clave uid gid nombre ignorar
do
    printf "%8s (%s)\n" $usuario
    $nombre done </etc/passwd
```

Como ejercicio se propone explicar paso a paso el funcionamiento del bucle anterior y los valores que se van asignando a las distintas variables en cada iteración.

Control de trabajos.

Las órdenes para el control de trabajos permiten manipular los procesos ejecutados por el usuario o por un guión BASH. El usuario puede manejar varios procesos y subprocesos simultáneamente, ya que el sistema asocia un número identificador único a cada uno de ellos.

Los procesos asociados a una tubería tienen identificadores (PID) propios, pero forman parte del mismo trabajo independiente.

La siguiente tabla describe las instrucciones asociadas con el tratamiento de trabajos.

Mandato	Descripción
<i>Mandato &</i>	Ejecución en 2º plano: lanza el proceso de forma desatendida, con menor prioridad y con la posibilidad de continuar su ejecución tras la desconexión del usuario.
bg %Nº Trabajo	Retorna a ejecutar en 2º plano: continúa la ejecución desatendida de un procesos suspendido.
fg %Nº Trabajo	Retorna a ejecutar en 1º plano: vuelve a ejecutar el proceso asociado al trabajo indicado de forma interactiva.
jobs	Muestro los trabajos en ejecución, indicando el nº de trabajo y los PID de sus procesos.

kill <i>Señal PID1 %Trab1 ...</i>	Manda una señal a procesos o trabajos , para indicar una excepción o un error. Puede especificarse tanto el nº de señal como su nombre.
suspend	Para la ejecución del proceso , hasta que se recibe una señal de continuación.
trap [<i>Comando</i>] [<i>Señal1 ...</i>]	Captura de señal : cuando se produce una determinada señal (interrupción) en el proceso, se ejecuta el mandato asociado. Las señales especiales <code>EXIT</code> y <code>ERR</code> se capturan respectivamente al finalizar el “ <i>script</i> ” y cuando se produce un error en una orden simple.
wait [<i>PID1 %Trab1</i>]	Espera hasta que termine un proceso o trabajo : detiene la ejecución del proceso hasta que el proceso hijo indicado hay finalizado su ejecución.

Revisar los siguientes ejemplos.

```
# Se elimina el fichero temporal si en el programa aparecen
# las 3 señales 0 (fin), 1, 2, 3 y 15.
trap 'rm -f ${FICHTEMP} ; exit' 0 1 2 3 15
...
# Ordena de forma independiente las listas de ficheros de
# usuarios, espera a finalizar ambos procesos y compara
# los resultados.
(cat alum03.sal prof03.sal pas03.sal | sort | uniq > usu03)
& (cat alum04.sal prof04.sal pas04.sal | sort | uniq >
usu04) & wait
diff usu03 usu04
```

Intérprete de uso restringido.

Cuando el administrador del sistema asigna al usuario un intérprete de uso restringido, éste último utiliza un entorno de operación más controlado y con algunas características limitadas o eliminadas.

Existen 3 formas de ejecutar BASH en modo restringido:

```
rbash
bash -r
bash --restricted
```

BASH restringido modifica las siguientes características ^[2]:

- No se puede usar la orden `cd`.
- No se pueden modificar las variables `SHELL`, `PATH`, `ENV` ni `BASH_ENV`.

- No se pueden ejecutar mandatos indicando su camino, sólo se ejecutarán aquellos que se encuentren en los directorios especificados por el administrador..
- No se pueden especificar caminos de ficheros como argumento del mandato “.”.
- No se pueden añadir funciones nuevas en los ficheros de inicio.
- No se admiten redirecciones de salida.
- No se puede reemplazar el intérprete RBASH (no se puede ejecutar la orden **exec**).
- No se pueden añadir o quitar mandatos internos.
- No se puede modificar el modo de operación restringido.

AWK

Introducción

Me pide Bernat que le ayude a hacer una faena rápida. Es tan fácil como procesar automáticamente el contenido de una hoja de cálculo. En ella, cada fila contiene datos de un alumno de la Universidad. Se pretende extraer de la hoja el nombre de usuario y la contraseña de cada estudiante, y autorizar su acceso al contenido proporcionado por un servidor web.

Estoy en franca desventaja debido a mis problemas cutáneos (padezco una urticaria que se manifiesta cada vez que me acerco a un ordenador con *Windows*) Un tercer investigador de la Compañía para la que trabajamos se abalanza sobre el problema, debidamente pertrechado con sus cientos de *gigabytes*, gigaciclos y gigaeuros dudosamente gastados en *visualestudios* variados.

Cuando nuestro jovial amigo ha pasado un buen rato buceando por las profundidades del *API* de Excel, extraños problemas relacionados con el flamenco y las Fuerzas Armadas --aparece en la pantalla algo relacionado con *OLE* y el *General Failure*-- le apartan de la cuestión principal. Yo, por mi parte, decido volver sobre el tema realizando una aproximación más humilde.

¿Por qué no *awk*? Al fin y al cabo, esta herramienta fue diseñada en los albores del *UNIX* precisamente para resolver con poco esfuerzo situaciones en las que se desea leer un archivo línea a línea, y tomar decisiones en función del contenido particular de cada registro.

Los adictos a las mil y una variedades del clic (*single click*, *double click*, *right button click*, *close your eyes and clap your hands click*, etc.) encontrarán seguramente alguna ocupación alternativa a la lectura de este artículo. Solamente los locos de la pantalla negra podrán seguir leyendo sin experimentar unas ganas irresistibles de vomitar. El que avisa no es traidor.

Supongamos que el lector selecto --es decir, el que sigue leyendo a pesar de la enorme oferta televisiva del momento-- dispone de un ordenador con su correspondiente sistema operativo y de lo necesario para ejecutar *awk*. (A decir verdad, *Windows* también puede servir en este caso, ya que existe una versión de *awk* para *msdos*, *win16*, *win32*, etc. Con un pequeño esfuerzo adicional los usuarios de este tipo de *me-reservo-el-calificativo* pueden seguir este artículo a pesar de no tener un sistema decente que llevarse al procesador)

¿Para qué sirve *awk*?

Como ya se ha indicado, esta utilidad permite procesar secuencialmente uno o varios archivos, los cuales constituyen la información "entrante" para *awk*. El contenido de cada una de las líneas de la entrada se compara con una o varias plantillas, en las que debe

especificarse un patrón de búsqueda. A cada patrón se le asocia una acción, que, normalmente, procesa la información entrante y la envía a la salida estándar.

Por tanto, los usos típicos de *awk* se centran en la selección de registros de un archivo en función de uno o varios criterios de búsqueda. Si solamente resultara relevante una parte del registro seleccionado, también sería posible separarla e ignorar el resto.

La información "saliente" se puede transformar según las necesidades del usuario, por lo que es posible crear informes con muy poco esfuerzo de programación.

La sintaxis es muy semejante a la del lenguaje C, con la particularidad de que, al ser *awk* un intérprete, no es necesario pasar por procesos de compilación y *linkedición*. El rendimiento de *awk* no es comparable al de un ejecutable, pero, dada la orientación al proceso secuencial de archivos, esta circunstancia solamente es relevante en condiciones extremas.

Awk puede, además, leer y grabar simultáneamente otros archivos, además de la entrada y salida estándar. Otras capacidades avanzadas quedan, por su mayor dificultad, fuera del alcance de este artículo.

Conviene, finalmente, recalcar la circunstancia de que *awk* solamente procesa archivos de texto. Para manejar los datos de una hoja Excel o de cualquier otro documento en formato *propietario* es necesario exportar la información a algún formato de texto.

Prerrequisitos

Se presupondrá en adelante que el lector dispone de alguna variedad de *UNIX* (permítaseme un inciso para reconocer el mérito del pedante anónimo que acuñó la expresión *sabores de UNIX...*)

No se requiere el uso de una *shell* concreta. En los ejemplos se utiliza *bash* por ser capaz de interpretar caracteres de control escritos con la sintaxis de escape con "barra invertida".

Y, evidentemente, se precisa disponer de *awk*. Si se abre una sesión en la consola o una ventana de terminal se averiguará fácilmente la presencia de *awk* en la máquina tecleando la información que aparece en tipografía negrita de paso fijo:

```
$ awk  
Usage: awk [-f programfile | 'program'] [-F fieldsep] [-v var=value] [files]
```

Parámetros en tiempo de ejecución

Awk sigue, como hemos indicado, un guión previamente escrito por el programador. Está compuesto por plantillas, que indican --mediante una expresión-- qué registros se desea

procesar, y --mediante una acción-- qué proceso se pretende realizar con los registros previamente seleccionados.

Se puede escribir el guión directamente en la línea de mandatos, encerrado entre comillas simples, o en un archivo de texto separado. Si se opta por la segunda opción, debe escribirse el nombre del archivo tras el parámetro `-f`. La `"f"` debe ser minúscula obligatoriamente. Por tanto, cualquiera de las dos formas mostradas a continuación produce resultados equivalentes:

```
$ awk '{ print $1 }' entrada.txt
$ awk -f primer.awk entrada.txt
```

Más adelante se mostrará cómo *awk* procesa cada registro como una sucesión de "campos" separados por un carácter determinado que indica dónde termina un campo y empieza el siguiente. La situación más habitual es aquella en la cual cada registro contiene una frase y los campos son las palabras de la frase. El carácter delimitador es el espacio que separa cada palabra de la siguiente. Mediante el parámetro `-F` (ahora mayúscula) se indica a *awk* qué carácter debe considerar como separador de campos.

```
$ awk -F "#" -f primer.awk entrada.txt
```

También se mostrará la capacidad para usar "variables" dentro del guión, igual que en otros lenguajes de programación. Mediante el parámetro `-v` se puede asignar un valor a una variable desde la línea de mandatos:

```
$ awk -v fecha="16/Dic/2001" -f primer.awk entrada.txt
```

El resto de los parámetros son los nombres de los archivos de datos "entrantes". El proceso se realiza accediendo a los archivos en el orden especificado. Toma el conjunto de los archivos como si fueran uno solo. *Awk* espera recibir por la entrada estándar (*stdin*) los datos que debe procesar. Por ejemplo:

```
$ cat ejemplo.txt | awk -f primer.awk
```

Alternativamente, los parámetros indicados al final de la línea de mandatos se interpretan como nombres de archivo:

```
$ awk -f primer.awk ejemplo.txt
```

Suponga el lector que se desea actuar sobre un archivo de texto como el que se muestra a continuación. Contiene cuatro líneas de texto terminadas por el carácter de salto de línea “\n”, que no se muestra en la pantalla, salvo por el efecto de escribir en la línea siguiente:

```
$ cat ejemplo.txt
La bella y graciosa moza
marchóse a lavar la ropa
la frotó sobre una piedra
y la colgó de un abedul.
$ _
```

Primeros pasos con awk

Todo el proceso gira en torno a la comparación de todas las líneas del archivo, leídas una por una en secuencia, con una o más plantillas. Cada plantilla contiene una expresión, y lleva asociada una acción. La sintaxis es, en general, de la forma:

```
expresión { acción }
```

Si la evaluación de la expresión resulta positiva, se realiza la acción asociada. El resultado de una expresión vacía se considera siempre positivo. Si la acción se omite, se asume que lo que se desea es mostrar el registro en la salida estándar. Este proceso se reitera con cada línea, hasta llegar al final del archivo.

En el siguiente ejemplo se pasa como único parámetro la plantilla entre comillas simples:

```
$ cat ejemplo.txt | awk '1'
La bella y graciosa moza
marchóse a lavar la ropa
la frotó sobre una piedra
y la colgó de un abedul.
$ _
```

Esta plantilla contiene una expresión trivial (1), cuya evaluación resulta siempre positiva. En ausencia de una acción asociada explícitamente, *awk* asume la acción por omisión, que consiste en mostrar la línea completa en *stdout*. Se puede añadir una acción algo más sofisticada entre llaves:

```
$ cat ejemplo.txt | awk '1 { print "línea" }'  
línea  
línea  
línea  
línea  
$ _
```

Como puede observarse, la acción `print` se realiza una vez por cada línea que se obtiene de `stdin`. Para emplear más de una plantilla se emplea la construcción “expresión { acción }” repetidas veces, como en el siguiente ejemplo:

```
$ cat ejemplo.txt | awk '1 { print "línea" } 1 { print "recta" }'  
línea  
recta  
línea  
recta  
línea  
recta  
línea  
recta  
$ _
```

También es posible incorporar varias instrucciones en la misma acción, separadas por el signo “punto y coma”:

```
$ cat ejemplo.txt | awk '1 { print "Yo uso"; print "Linux" }'  
Yo uso  
Linux  
Yo uso  
Linux  
Yo uso  
Linux  
Yo uso  
Linux  
$ _
```

Para no confundir al lector, se omite deliberadamente la mención de la posibilidad de usar expresiones regulares semejantes a las que usa “*grep*”. Si el distinguido público lector manifiesta un ferviente deseo de profundizar en el tema, el autor considerará seriamente la posibilidad de ampliar la información al respecto.

Campos de entrada

Para utilizar una expresión que no sea trivial se debe introducir el concepto de “campo”. Se considerará cada registro del archivo de entrada como una sucesión de campos delimitados por un carácter dado. Este carácter es, por omisión, el espacio. Como se ha visto anteriormente, se puede indicar a *awk* que considere otro carácter como separador de campos mediante la opción `-F` (mayúscula) en la línea de mandatos.

Se hace referencia al contenido de cada campo usando el carácter “\$” (dólar) seguido del ordinal del campo: `$1`, `$2`, `$3`, etc. Con la notación `$0` se hace referencia a la línea entera.

Se puede forzar el proceso del registro carácter a carácter dejando la variable “separador de campos” `FS` sin contenido. De este modo, en `$1` se tendrá el primer carácter de la

línea, en \$2 el segundo, etc. Incluso en este caso, el campo \$0 contendrá la línea completa.

Como quiera que se está anticipando en estos ejemplos el uso de la instrucción *"print"*, se muestra a continuación el guión anterior con las modificaciones necesarias para mostrar en stdout el contenido de la primera y la segunda palabra de cada línea:

```
$ cat ejemplo.txt | awk '1 { print "Primera y segunda palabras ==>", $1, $2, "<=== final del registro." }'
```

Primera y segunda palabras ==> La bella <=== final del registro.
Primera y segunda palabras ==> marchóse a <=== final del registro.
Primera y segunda palabras ==> la frotó <=== final del registro.
Primera y segunda palabras ==> y la <=== final del registro.
\$ _

Se estudiará más adelante el efecto de la coma en la instrucción *"print"*. Si no se indica otra cosa, *awk* inserta un espacio por cada coma que aparezca en esta instrucción.

Variables

También es necesario en este punto citar la existencia de variables predefinidas en el lenguaje. El concepto tradicional de "variable" (almacén de datos en memoria, caracterizado por un nombre no repetido en el mismo contexto y capaz de contener un dato que puede ser modificado por el programa) se aplica perfectamente en este caso.

El programador puede crear sus propias variables simplemente haciendo referencia a ellas en expresiones. Las variables pueden ser escalares (con un solo nombre y valor) o vectoriales (con nombre, subíndice y valor) El subíndice puede ser una cadena arbitraria, lo cual permite crear tablas asociativas. Se emplea la notación "nombre[subíndice]" para referirse a un elemento de la tabla. También se admite la definición de variables de naturaleza matricial, con nombre, varios subíndices y varios valores, si se usa la notación "nombre[subíndice1, subíndice2, subíndice3]".

Las siguientes variables predefinidas están relacionadas con la información entrante:

FS (Field separator): contiene el carácter que indica a *awk* en qué punto del registro acaba un campo y empieza el siguiente. Por omisión es un espacio. Se puede indicar un carácter alternativo mediante una instrucción de asignación como `FS = "/"`. Si se deja vacío, cada lectura se realiza dejando un carácter en cada campo.

NF (Number of fields): contiene el número total de campos que contiene el registro que se está leyendo en cada momento.

RS (Record separator): contiene el carácter que indica a awk en qué punto del archivo acaba un registro y empieza el siguiente. Es “\n” por omisión.

NR (Number of record): contiene el número de orden del registro que se está procesando en cada momento.

Del mismo modo, se dispone de variables relativas a la información de salida, como las siguientes:

OFS (Output FS): contiene el separador de campos para la salida generada por awk. La instrucción print inserta en la salida un carácter de separación cada vez que aparece una coma en el código. Por ejemplo:

```
$ echo "manejo un apple" | awk -v OFS="\t" '1 {print $1, $2, $3}'
manejo      un        apple
$ _
```

ORS (Output RS): Contiene el carácter que awk escribirá al final de cada registro. Es “\n” por omisión. Obsérvese que en esta modificación del ejemplo anterior la entrada contiene DOS registros, dada la aparición del salto de línea en el centro de la cadena de caracteres entre comillas. Al modificar el carácter separador de la salida se genera un solo registro:

```
$ echo -e "manejo un apple\ncaro y elegante" | awk -v OFS = "\t" -v ORS = "\n" '1 {print $1, $2, $3}'
manejo      un        apple      caro      y        elegante
$ _
```

Operadores

Son aplicables los operadores comunes en C para realizar operaciones aritméticas sobre variables con valores numéricos, y para concatenar los valores contenidos en variables con valores alfanuméricos. Así, se tiene:

+	Suma
-	Resta
*	Multiplicación
/	División

%	Resto
^	Exponenciación
Esp	Concatenación

También se dispone de los operadores comunes

!	Negación
A ++	Incrementar positivamente en una unidad
A --	Incrementar negativamente en una unidad
A += N	Incrementar positivamente en N unidades
A -= N	Incrementar negativamente en N unidades
A *= N	Multiplicar por N
A /= N	Dividir por N
A %= N	Reemplazar por el resto de la división por N
A ^= N	Sinceramente, no sé qué demonios hace este operador
>	Mayor que
>=	Mayor o igual que
<	Menor que
<=	Menor o igual que

==	Igual a
!=	Distinto de
?:	Formato alternativo de decisión (if/else)

Awk en acción

Para recapitular lo visto hasta ahora, se realizará a continuación algún proceso sencillo con el archivo de ejemplo mostrado en párrafos anteriores. El archivo entrante contiene un número teóricamente no determinado de registros, terminados por el carácter contenido en RS. El proceso comienza leyendo el primer registro, aplicando a éste todas las plantillas existentes con el valor de la variable NR igual a 1. Se reitera el proceso para cada registro, incrementando en una unidad el valor de NR cada vez.

Cada registro está compuesto por NF campos, delimitados por las apariciones del carácter FS en el mismo. El ordinal de cada campo se designa, como se indicó arriba, con \$1, \$2, \$3... y así hasta \$NF.

Con campos, variables y operadores se puede escribir una plantilla algo más útil que las anteriores:

```
$ awk '$5 == "moza" { print "La línea", NR, "contiene MOZA en el quinto campo"; }' ejemplo.txt
La línea 1 contiene MOZA en el quinto campo
$ _
```

A medida que la lógica de las plantillas se complique, será de utilidad el parámetro -f (minúscula) seguido del nombre del guión o script que contendrá las plantillas. Cuando las plantillas se codifican dentro del script, no se usan las comillas simples para delimitarlas. Un guión típico como el que se muestra ("proceso.awk" en el directorio actual) podría tener el siguiente contenido:

```
$ cat proceso.awk
{ print "Procesando la línea", NR }
NF > 1 { print "La línea tiene más de una palabra" }
$1 == "Linux" {print "La primera palabra es Linux"}
$ _
```

La primera plantilla siempre se ejecuta por no tener expresión delante de las llaves. La segunda solamente se activa para las líneas que tienen más de una palabra. La tercera es aplicable si la línea comienza con la palabra "Linux".

Plantillas de prólogo y epílogo

Dos expresiones especiales, BEGIN y END, permiten realizar acciones antes de procesar el primer registro, y después de procesar el último, respectivamente.

El sencillo script "suma.awk" pone a cero la variable "total" antes de iniciar el proceso del archivo de entrada, añade a esa variable el valor correspondiente al contenido del primer campo de cada registro, y muestra el valor de la variable al final:

```
$ cat suma.awk

BEGIN { total = 0 }
{ total += $1 }
END { print "El total es", total }
```

Para ejecutarlo, se puede usar lo siguiente (obsérvese que la entrada tiene cuatro registros):

```
$ echo -e "10\n10\n10\n70" | awk -f ./suma.awk

El total es 100
```

Dependiendo de la implementación de UNIX que se esté usando, el script puede fallar si el campo que interviene en la operación (\$1 en este caso) contiene un valor no numérico. En este caso se puede forzar la conversión sumando "cero" a la variable.

Instrucciones de control

Algunas instrucciones del lenguaje C están presentes en *awk* de manera simplificada. Se dispone de la construcción condicional "if/else", que se codifica de la manera habitual:

```
if( expression ) statement [ else statement ]
```

En algunas implementaciones se puede complicar hasta añadir varias instrucciones mediante anidamiento de acciones:

```
{
    print "Procesando registro", NR;
    if( $1 > 0 )
    { print "positivo";
      print "El primer campo tiene:", $1;
    }
    else
    {
      print "negativo";
    }
}
```

La primera forma de la construcción "for" se maneja igual que en el lenguaje C, es decir:

```
for( expression; expression; expression ) statement
```

donde la primera expresión suele poner valor inicial al índice. En la segunda se controla hasta qué momento la ejecución debe continuar. La tercera marca el nivel de incremento de la variable índice.

```
{
  for( i = 0; i < 10; i ++ )
  {
    j = 2 * i;
    print i, j;
  }
}
```

En su segunda forma, la construcción "*for*" recorre los elementos de una matriz (*array*). Dado el carácter avanzado de este concepto, quedará para un segundo artículo su manejo.

La construcción "*do/while*" se emplea en la forma habitual. En el caso actual se realiza un sencillo contador de cero a diez:

```
BEGIN { i = 0; do {print i; i ++} while ( i < 10 ) }
```

La instrucción "*break*" permite abandonar un bucle anticipadamente. Para descartar el resto del código del bucle y empezar una nueva iteración se usa "*continue*". La instrucción "*next*" descarta todas las plantillas que siguen, y pasa a realizar el proceso del siguiente registro. Finalmente, la instrucción "*nextfile*" termina anticipadamente el proceso de un archivo de entrada para pasar al siguiente.

Funciones

Las funciones predefinidas del lenguaje responden al concepto habitual en otros lenguajes de programación. La llamada a una función se resuelve en tiempo de ejecución. Cuando se evalúa una expresión, se realiza la sustitución de la llamada a la función por el valor que ésta devuelve. Por ejemplo:

```
$ awk 'BEGIN { print "El seno de 30 grados es", sin( 3.141592653589 / 6 ); }' /dev/null
El seno de 30 grados es 0.5
$ _
```

Se dispone de las funciones matemáticas y trigonométricas `exp`, `log`, `sqrt`, `sin`, `cos` y `atan2`. Además, son de mucha utilidad las siguientes:

<code>length()</code>	Longitud del parámetro en bytes
<code>rand()</code>	Número al azar entre cero y uno
<code>srand()</code>	Inicia la semilla de generación al azar
<code>int()</code>	Devuelve el parámetro convertido en un entero
<code>substr(s,m,n)</code>	Devuelve la subcadena de <code>s</code> en <code>m</code> con longitud <code>N</code>
<code>index(s,t)</code>	Posición de <code>s</code> donde aparece <code>t</code> , o cero si no está.
<code>match(s,r)</code>	Posición de <code>s</code> en que se cumple la expresión <code>r</code> .
<code>split(s,a,fs)</code>	Devuelve <code>s</code> en elementos separados por <code>fs</code>
<code>sub(r,t,s)</code>	Cambia en <code>s</code> la cadena <code>t</code> por <code>r</code> . Es <code>\$0</code> por omisión
<code>gsub(r,t,s)</code>	Igual, pero cambia todas las <code>t</code> en la cadena
<code>sprintf(f,e,e...)</code>	Devuelve cadena de formato <code>f</code> con las “e”
<code>system(cmd)</code>	Ejecuta <code>cmd</code> y devuelve el código de retorno
<code>tolower(s)</code>	Devuelve <code>s</code> convertida en minúsculas
<code>toupper(s)</code>	Devuelve <code>s</code> convertida en mayúsculas
<code>getline</code>	Fuerza una lectura de fichero

La solución definitiva

El lector avezado debería estar, a estas alturas, preparado para afrontar el desafío del principio del artículo. Con los mimbres mostrados en las páginas anteriores se tratará de

hacer el cesto deseado. Será suficiente con que el gui3n *awk* resultante sea capaz de construir un *shell script* que contenga los mandatos necesarios para autorizar el acceso de los alumnos al servidor web.

La informaci3n de partida est1 contenida en el archivo "alumnos.txt", que contiene una l3nea por cada alumno. En esa l3nea encontraremos el nombre, los apellidos, la cuenta de usuario que le corresponde y la contrase1a que va a emplear para acceder.

```
$ cat alumnos.txt
Sebasti1n Palomo Linares      spl18901      jalisco
Pedro P3rez Pino             ppp0          hatalue
John Jensen                  jotajota      whatafew
Mar3a del Carmen Garc3a Feo   mc98819       uglymom
$ _
```

En la salida se desea generar un *shell script* con el comentario que indica el nombre de la shell a emplear ("#!/bin/bash", por ejemplo) en la primera l3nea, y con una l3nea como "/usr/bin/htpasswd -b usuario contrase1a" por cada alumno.

Para poner una pizca de dificultad al problema, pondremos de manifiesto un hecho que seguramente no extra1ar1 al lector; los nombres no tienen un n3mero fijo de palabras, y no se garantiza que todos los alumnos est3n expresados con los dos apellidos. As3 que el planteamiento simplista que har3a un anglosaj3n no muy despierto, consistente en asociar \$1 con el nombre, \$2 con el apellido, \$3 con la cuenta y \$4 con la contrase1a, no conduce al resultado correcto:

```
$ awk '{print "/usr/bin/htpasswd -b", $3, $4}' alumnos.txt
/usr/bin/htpasswd -b Linares spl18901
/usr/bin/htpasswd -b Pino ppp0
/usr/bin/htpasswd -b jotajota whatafew
/usr/bin/htpasswd -b Carmen Garc3a
```

Otra dificultad que puede presentarse viene dada por la eventual inexistencia de la base de datos de autorizaciones del servidor *web* en el momento de la ejecuci3n del *script*. A los efectos de la programaci3n, implica a1adir la opci3n "-c" a la primera llamada al mandato "*htpasswd*" dentro del *script* que se genera.

La aproximaci3n correcta al problema consiste en obtener el n3mero de campos (palabras) del registro (l3nea), y seleccionar la pen3ltima y la 3ltima. Afortunadamente, la variable NF contiene en cada momento el n3mero de campos de la l3nea que se est1 procesando. En este ejemplo puede verse c3mo se hace uso de esta capacidad:

```
$ awk '{ print NF }' alumnos.txt
5
5
4
7
```

Pero, lo que es realmente maravilloso en este caso es que podemos acceder al valor del campo que se encuentra en el último lugar con tan solo `$NF` en lugar de `NF`:

```
$ awk '{ print NF, ":", $NF }' alumnos.txt
5 : jalisco
5 : hatalue
4 : whatafew
7 : uglymom
```

Rizando más el rizo, se puede acceder también al penúltimo campo empleando una expresión aritmética:

```
$ awk '{ print $(NF - 1) }' alumnos.txt
spl18901
ppp0
jotajota
mc98819
```

Combinando ambos, es fácil obtener cuenta y contraseña con la instrucción:

```
$ awk '{ print $(NF - 1), $NF }' alumnos.txt
spl18901 jalisco
ppp0 hatalue
jotajota whatafew
mc98819 uglymom$ _
```

Para dar los toques finales con más comodidad se escribiera el guión en un archivo llamado `"altaweb.awk"` :

```
$ cat altaweb.awk
BEGIN { print "#!/bin/bash" }
{
    print "/usr/bin/htpasswd",
        ( NR == 1 ? "-c -b" : "-b" ),
        $(NF - 1), $NF
}
END { print "# Shell script generado OK." }
$ _
```

Puede observarse que se ha usado la sintaxis abreviada de la estructura de control `"if/else"`, típica del lenguaje C, en la que una expresión antecede al signo de interrogación. Al evaluarse positivamente se toma en cuenta el código que aparece antes del signo de dos puntos. Si la evaluación es negativa, el código que se ejecuta es el que sigue a los dos puntos.

La ejecución del proceso se realiza, como antes, con el mandato que se muestra a continuación:


```
$ awk -f altaweb.awk alumnos.txt > altaweb.sh
```

Solamente resta verificar el contenido del *shell script* y ejecutarlo. Cuando se tenga confianza en el resultado se puede obviar el paso intermedio y ejecutar directamente:

```
$ awk -f altaweb.awk alumnos.txt | sh
```

Conclusión

Awk es un poderoso intérprete, que permite realizar tareas con archivos de tipo secuencial con una elevada efectividad. La eficacia de su uso en procesos automatizados es innegable, a la vista de que mi jovial colaborador ha desistido, y se dispone a reinstalar *Windows*. Sin embargo, disfrutará de lo lindo con su salvapantallas de tiburones simultaneado con apariciones esporádicas de una misteriosa pantalla azul en la que se glosan las maravillas de la "*Excepción OE*"... Hasta más ver.

Uso básico de SED

Qué es sed

SED (Stream EDitor) es un editor de flujos y ficheros de forma no interactiva. Permite modificar el contenido de las diferentes líneas de un fichero en base a una serie de comandos o un fichero de comandos (-f fichero_comandos).

Sed recibe por stdin (o vía fichero) una serie de líneas para manipular, y aplica a cada una de ellas los comandos que le especifiquemos a todas ellas, a un rango de las mismas, o a las que cumplan alguna condición.

Por ejemplo:

```
Sustituir apariciones de cadena1 por cadena2 en todo el fichero:
```

```
# sed 's/cadena1/cadena2/g' fichero > fichero2
```

12:44:20 03/04/2017



```
daw2016@daw2016-VirtualBox:~/scripts$ sed 's/;/;/g' copiausuarios
```

```
Sustituir apariciones de cadena1 por cadena2 en las líneas 1 a 10:
```

```
# comando | sed '1,10 s/cadena1/cadena2/g'
```

```
Eliminar las líneas 2 a 7 del fichero
```

```
# sed '2,7 d' fichero > fichero2
```

```
Buscar un determinado patrón en un fichero:
```

```
# sed -e '/cadena/ !d' fichero
```

```
Buscar AAA o BBB o CCC en la misma línea:
```

```
# sed '/AAA/!d; /BBB/!d; /CCC/!d' fichero
```

```
Buscar AAA y BBB y CCC:
```

```
# sed '/AAA.*BBB.*CCC/!d' fichero
```

```
Buscar AAA o BBB o CCC (en diferentes líneas, o grep -E):
```

```
# sed -e '/AAA/b' -e '/BBB/b' -e '/CCC/b' -e d
```

```
# gsed '/AAA\|BBB\|CCC/!d'
```

Formato de uso

El formato básico de uso de sed es:

```
# sed [-ns] '[direccion] instruccion argumentos'
```

Donde:

- **[direccion]** es opcional, siendo un número de línea (N), rango de números de línea (N,M) o búsqueda de regexp (/cadena/) indicando el ámbito de actuación de las instrucciones. Si no se especifica [direccion], se actúa sobre todas las líneas del flujo.
- **Instruccion** puede ser:
 - **i** = Insertar línea antes de la línea actual.
 - **a** = Insertar línea después de la línea actual.
 - **c** = Cambiar línea actual.
 - **d** = Borrar línea actual.
 - **p** = Imprimir línea actual en stdout.

- **s** = Sustituir cadena en línea actual.
 - **r fichero** = Añadir contenido de "fichero" a la línea actual.
 - **w fichero** = Escribir salida a un fichero.
 - **!** = Aplicar instrucción a las líneas no seleccionadas por la condición.
 - **q** = Finalizar procesamiento del fichero.
- **-n**: No mostrar por stdout las líneas que están siendo procesadas.
 - **-s**: Tratar todos los ficheros entrantes como flujos separados.

Ejemplos de sustitución

Reemplazar cadenas:

```
# sed 's/^Host solaris8/Host solaris9/g' fichero > fichero2
```

Reemplazar cadenas sólo en las líneas que contentan una cadena:

```
# sed '/cadena_a_buscar/ s/vieja/nueva/g' fichero > fichero2
```

12:52:05 03/04/2017

```
daw2016@daw2016-VirtualBox:~/scripts$ sed '/x/ s/x/!/g' misusuarios
root;!;0;0;root;/root;/bin/bash
daemon;!;1;1;daemon;/usr/sbin;/usr/sbin/nologin
bin;!;2;2;bin;/bin;/usr/sbin/nologin
```

Reemplazar cadenas sólo en en determinadas líneas:

```
# sed '5,6 s/vieja/nueva/g' fichero > fichero2
```

Reemplazar multiples cadenas (A o B):

```
# sed 's/cadenasrc1\|cadenasrc2/cadena_nueva/g'
```

Sustituir líneas completas (c) que cumplan o no un patrón:

```
# echo -e "línea 1\nlínea 2" | sed '/1/ cPrueba'
```

```
Prueba
línea 2
```

```
# echo -e "línea X 1\nlínea 2" | sed '/1/ !cPrueba'
```

```
línea 1
Prueba
```

Ejemplos de Inserción

Insertar 3 espacios en blanco al principio de cada línea:

```
# sed 's/^/   /' fichero
```

Añadir una línea antes o después del final de un fichero (\$=última línea):

```
# sed -e '$i Prueba' fichero > fichero2
```

```
# sed -e '$a Prueba' fichero > fichero2
```

Insertar una línea en blanco antes de cada línea que cumpla una regex:

```
# sed '/cadena/{x;p;x;}' fichero
```

Insertar una línea en blanco detrás de cada línea que cumpla una regex:

```
# sed '/cadena/G' fichero
```

Insertar una línea en blanco antes y después de cada línea que cumpla una regex:

```
# sed '/cadena/{x;p;x;G;}' fichero

Insertar una línea en blanco cada 5 líneas:
# sed 'n;n;n;n;G;' fichero

Insertar número de línea antes de cada línea:
# sed = filename | sed 'N;s/\n/\t/' fichero

Insertar número de línea, pero sólo si no está en blanco:
# sed '/./=' fichero | sed '/./N; s/\n/ /'

Si una línea acaba en \ (backslash) unirla con la siguiente:
# sed -e :a -e '/\\$/N; s/\\n//; ta' fichero
```

Ejemplos de Selección/Visualización

```
Ver las primeras 10 líneas de un fichero:
# sed 10q

Ver las últimas 10 líneas de un fichero:
# sed -e :a -e '$q;N;11,$D;ba'

Ver un rango concreto de líneas de un fichero:
# cat -n fich2 | sed -n '2,3 p'
  2    línea 2
  3    línea 3
(Con cat -n, el comando cat agrega el número de línea).
(Con sed -n, no se imprime nada por pantalla, salvo 2,3p).

Ver un rango concreto de líneas de varios ficheros:
# sed '2,3 p' *
  línea 2 fichero 1
  línea 3 fichero 1
  línea 2 fichero 2
  línea 3 fichero 2
(-s = no tratar como flujo sino como ficheros separados)

Sólo mostrar la primera línea de un fichero:
# sed -n '1p' fichero > fichero2.txt

No mostrar la primera línea de un fichero:
# sed '1d' fichero > fichero2.txt

Mostrar la primera/ultima línea de un fichero:
# sed -n '1p' fichero
# sed -n '$p' fichero

Imprimir las líneas que no hagan match con una regexp (grep -v):
# sed '/regexp/!d' fichero
# sed -n '/regexp/p' fichero

Mostrar la línea que sigue inmediatamente a una regexp:
# sed -n '/regexp/{n;p;}' fichero

Mostrar desde una expresión regular hasta el final de fichero:
# sed -n '/regexp/, $p' fichero

Imprimir líneas de 60 caracteres o más:
# sed -n '/^.{60\}/p' fichero
```

```
Imprimir líneas de 60 caracteres o menos:  
# sed -n '/^\.{65}\!/p' fichero  
# sed '/^\.{65}\!/d' fichero
```

Ejemplos de Borrado

```
Eliminar un rango concreto de líneas de un fichero:  
# sed '2,4 d' fichero > fichero2.txt
```

```
Eliminar todas las líneas de un fichero excepto un rango:  
# sed '2,4 !d' fichero > fichero2.txt
```

```
Eliminar la última línea de un fichero  
# sed '$d' fichero
```

```
Eliminar desde una línea concreta hasta el final del fichero:  
# sed '2,$d' fichero > fichero2.txt
```

```
Eliminar las líneas que contentan una cadena:  
# sed '/cadena/ d' fichero > fichero2.txt  
# sed '/^cadena/ d' fichero > fichero2.txt  
# sed '/^cadena$/ d' fichero > fichero2.txt
```

```
Eliminar líneas en blanco (variación del anterior):  
# comando | sed '/^$/ d'  
# sed '/^$/d' fichero > fichero2.txt
```

```
Eliminar múltiples líneas en blanco consecutivas dejando sólo 1:  
# sed '/./,/^$/!d' fichero
```

```
Añadir una línea después de cada línea:  
# echo -e "línea 1\nlínea 2" | sed 'aPrueba'  
línea 1  
Prueba  
línea 2  
Prueba
```

```
Eliminar espacios al principio de línea:  
# sed 's/^ *//g' fichero
```

```
Eliminar todos los espacios que haya al final de cada línea:  
# sed 's/ *$//' fichero
```

```
Eliminar espacios sobrantes a principio y final de línea, o ambos:  
# sed 's/^[ \t]*//' fichero  
# sed 's/[ \t]*$//' fichero  
# sed 's/^[ \t]*//;s/[ \t]*$//' fichero
```

```
Eliminar tags HTML:  
# sed -e :a -e 's/<[^>]*>//g;/</N;//ba' fichero
```

```
Borrar líneas duplicadas no consecutivas de un fichero:  
# sed -n 'G; s/\n/&&/; /\n\([ -~]*\n\).*\n\1/d; s/\n//; h; P' fichero
```

```
Eliminar líneas en blanco y comentarios bash:  
# comando | sed '/^$/ d'  
# sed '/^$/d; / *#/d' fichero > fichero2.txt
```

Uso de salida selectiva

Salir a nuestra voluntad antes de acabar el fichero:

```
# sed -e '/uno/ s/uno/1/' -e '/salir/ q' fichero > fichero2.txt
# sed 10q fichero
# sed q fichero
```

Equivalencia de -e con ";":

```
# sed -e '/AAA/b' -e '/BBB/b' -e 'd' == sed '/AAA/b;/BBB/b;d'
```

Usar 'q' apropiadamente reduce tiempo de procesamiento:

```
# sed -n '10,20p' fichero
# sed -n '21q;10,20p' fichero -> Más rápido que el anterior.
```

Conversión de CRLF de DOS a formato UNIX (LF):

```
# sed 's/.$//' fichero
```

Conversión de LF de UNIX a formato DOS (CRLF):

```
# sed 's/$''\r'\`echo \\r`/" fichero
```

Obtener el Subject de un correo, pero sin cadena "Subject: ":

```
# sed '/^Subject: */!d; s///;q' fichero
```

Imprimir párrafo (cadenas entre 2 líneas en blanco) si contiene XXX:

```
# sed -e '/./{H;$!d;}' -e 'x;/XXX/!d;' fichero
```

Imprimir párrafo si contiene (1) XXX y ZZZ o bien (2) XXX o ZZZ :

```
# sed -e '/./{H;$!d;}' -e 'x;/XXX/!d;/ZZZ/!d'
# sed -e '/./{H;$!d;}' -e 'x;/XXX/b' -e '/ZZZ/b' -e d
```