

Script en bash para procesar datos y archivos

Taller de Sistemas Operativos
Escuela de Ingeniería Informática

Francisco Moretti Castro
francisco.moretti@alumnos.uv.cl

1. Introducción

El intérprete de comandos o shell es un programa que permite a los usuarios interactuar con el sistema, procesando las órdenes que se le indican. Además de comandos, los shells ofrecen otros elementos para mejorar su funcionalidad, tales como variables, funciones o estructuras de control.

Además de utilizar el shell desde la línea de comandos, puede emplearse para la interpretación de shell-scripts. Un shell-script o "guión de órdenes" es un fichero de texto que contiene un conjunto de comandos y órdenes interpretables por el shell.

Bash: Fue desarrollado para ser un superconjunto de la funcionalidad del Bourne Shell, siendo el intérprete de comandos asignado por defecto a los usuarios en las distribuciones de Linux, por lo que es el shell empleado en la mayoría de las consolas de comandos de Linux. [1]

La estructura del presente informe, está compuesta por una sección Introducción que provee un marco conceptual para comprender el objetivo del informe, después en la sección Descripción del problema se detalla la información las simulaciones, su estructura y ejemplos. Luego está la sección Tareas a realizar, donde se explica que se debe resolver y el formato de las soluciones para finalmente en la sección Diseño de la solución se muestra un diagrama de alto nivel de la solución, y aspectos importantes.

2. Descripción del problema

Los datos a procesar corresponden a estadísticas que provienen de cierta cantidad de experimentos de simulación de un sistema de evacuación de personas. Cada experimento es una simulación de una evacuación costera de la ciudad de Iquique ante un eventual tsunami, que considera 75000 personas. Existen tres modelos de personas las cuales se muestran en la Figura 1. [2]

Tipo de Persona	Descripción
Residente	Persona que vive en la ciudad y conoce su zona segura
Visitante Tipo I	Visitante que durante la evacuación logra determinar su zona segura
Visitante Tipo II	Visitante que durante la evacuación logra no determinar su zona segura

Figura 1

Además de clasificación mencionada, cada persona pertenece a un grupo etario, los que se resumen en la Tabla 1.

Tabla 1

Grupo Etario	Intervalo de edad
GO	0 – 14
G1	15 – 29
G2	30 - 64
G3	65 o más

2.1 Organización de los datos

Cada simulación entrega los resultados ordenados en una estructura de directorios, según el esquema que se visualiza en la Figura 1, donde NNN es el identificador de la simulación, representado por un código numérico de tres dígitos.

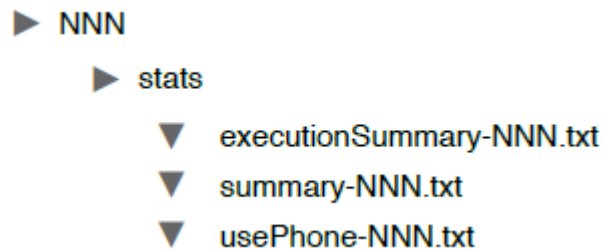


Figura 2

2.2 Descripción de los datos

2.2.1 El archivo **executionSummary-NNN.txt** contiene datos sobre el desempeño de la simulación NNN. Posee una cabecera y una segunda línea con los datos. Ésta tiene 10 campos, separados por el símbolo ‘:’. En la Figura 3 se muestra el contenido del archivo que corresponde a la simulación 000.

```
numExperiment:tsim:calibrationTime:Residents:Visitors:timeExecMakeAgents:timeExecCal:timeExecSim:maxMemory:agentsMem
0:3600:100:69000:6000:28252:33157:182800:288236:33648
```

Figura 3

En la Tabla 2 se detallan los campos y su respectiva descripción:

Tabla 2

Campo	Descripción	Ejemplo
numExperiment	Número del experimento.	0 (Corresponde al código 000)
Tsim	Tiempo de simulacion	3600: segundos de tiempo real que se simuló.
CalibrationTime	Tiempo de calibración del simulador	100: segundos de tiempo de real que se destina a la calibración inicial de las personas.
Residents	Cantidad de Residentes simulados	69000 personas
Visitors	Cantidad de Visitantes simulados	6000 personas
timeExecMakeAgents	Tiempo real que demora en crear en memoria las personas simuladas	28252: milisegundos
timeExecCal	Tiempo real que demora la calibracion de las personas	33157: milisegundos
timeExecSim	Tiempo real que toma la simulación	182800: milisegundos
maxMemory	Costo espacial del simulador	288236: Kbytes
agentsMem	Memoria utilizada por las estructuras de datos relacionados con las personas	33648: KBytes

2.2.2 El archivo **summary-NNN.txt** contiene datos sobre el comportamiento de las personas en la simulación NNN. Posee una cabecera, seguida de tantas líneas como la cantidad de personas que se simularon. Ésta líneas tienen ocho campos, separados por el símbolo ‘:’. En la Figura 3 se muestra el contenido del archivo que corresponde a la simulación 000.

```
numExperiment:id:model:groupAge:safeZone:distanceToTargetPos:responseTime:evacTime
0:0:0:2:Z1:8.271001:262.785961:1450.000000
0:1:0:1:Z2:7.345218:287.944671:1280.000000
0:2:0:0:Z5:17.635108:82.167863:1230.000000
0:3:0:1:Z1:23.365319:258.646227:2340.000000
0:4:0:2:Z3:5.798595:176.579702:1360.000000
0:5:0:2:Z5:28.269759:139.817949:1300.000000
.
.
.
0:74950:1:0:Z2:0.496621:136.855808:1940.000000
```

Figura 4

En la Tabla 3 se detallan los campos y su respectiva descripción:

Tabla 3

Campo	Descripción	Ejemplo
numExperiment	Número del experimento	1 (Corresponde al código 001)
id	Identificador de la persona simulada	14
model	Identificar del modelo de la persona 0: Residente 1: Visitante tipo I 2: Visitante tipo II	1
groupAge	Identificador del grupo etario de la persona 0: G1, 1: G2, 2: G3, 3: G4	3
safeZone	Identificador de la zona segura de la persona Z1, Z2, Z3, Z4, Z5 : Identificadores de zonas seguras	Z2

	NA: la persona no tiene zona segura asignada	
distanceToTargetPos	Distancia a la que quedó la persona de su objetivo inicial	8530 metros
responseTime	Número aleatorio que representa cuánto tiempo se demoró la persona en tomar la decisión de evacuar, desde que dieron la orden de evacuar. Si es -1, entonces la persona nunca pudo evacuar.	120 segundos
evacTime	Tiempo que la persona se demoró en llegar a la zona de evacuación. Si es 0, entonces la persona nunca llegó a la zona de evacuación.	1853 segundos

2.2.3 El archivo **usePhone-NNN.txt** contiene datos sobre el comportamiento de las personas en la simulación NNN, con respecto al uso de un teléfono móvil. Posee una cabecera y cada línea posterior, representa la cantidad de personas que utilizaron el teléfono en cierto instante de tiempo. Cada una tiene tres campos, separados por el símbolo ':'. En la Figura 4 se muestra el contenido del archivo que corresponde a la simulación 000.

```
numExperiment:timeStamp usePhone
0:0:0
0:10:4333
0:20:4383
0:30:4261
0:40:4410
```

Figura 5

En la Tabla 4 se detallan los campos y su respectiva descripción.

Campo	Descripción	Ejemplo
numExperiment	Número del experimento.	0 (Corresponde al código 000)
timeStamp	Tiempo de la medición	60 segundos
usePhone	Cantidad de persona que utilizaron el teléfono móvil en el tiempo especificado	4381 personas

3. Tareas a realizar

Diseñar un script en Bash usando parametros del estándar Unix , con el nombre stats.sh, que permita realizar estadísticas descriptivas de simulaciones entregadas.

El script debe realizar las siguientes tareas:

- 1) Determinar Cantidad máxima, mínima y promedio para las siguientes métricas de desempeño computacional del simulador.
 - Tiempo de simulación total: Este tiempo se define como: timeExecMakeAgents+timeExecCal+timeExecSim
 - Memoria utilizada por el simulador

Los resultados se deben entregar en un archivo metrics.txt , con la estructura que se muestra en la Figura 5.

```
tsimTotal:promedio:min:max  
memUsed:promedio:min:max
```

Figura 6

- 2) Determinar el tiempo promedio de evacuación, además del mínimo y el máximo para los siguientes grupos de personas:
 - Todas las personas simuladas
 - Sólo Residentes.
 - Sólo Visitantes Tipo I.
 - Sólo Residentes, separados por grupo etario.
 - Sólo Visitante Tipo I, separados por grupo etario.

Los resultados se deben entregar en un archivo evacuation.txt., utilizando la estructura de archivo que se muestra en la Figura 6.

```
alls:promedio:min:max  
residents:promedio:min:max  
visitorsI: promedio:min:max  
residents-G0:promedio:min:max  
residents-G1:promedio:min:max  
residents-G2:promedio:min:max  
residents-G3:promedio:min:max  
visitorsI-G0: promedio:min:max  
visitorsI-G1: promedio:min:max  
visitorsI-G2: promedio:min:max  
visitorsI-G3: promedio:min:max
```

Figura 7

- 3) Determinar el promedio de uso de teléfonos móviles, además del mínimo y el máximo para los instantes de tiempo especificados en los archivos usePhone-NNN.txt. Los resultados se deben guardar en un archivo usePhone-stats.txt. con la estructura de la Figura 7:

```
timestamp:promedio:min:max
```

Figura 8

4. Diseño de la solución

El trabajo se realizará en el servidor Ubuntu Server 18.04 instalado en una máquina virtual, en el cual como se ve en el Diagrama de Secuencia [3] de la Figura 8, provee los procedimientos para ejecutar un script en bash el cual nos indica que el fuerte está en la creación de un script que analice y permita realizar estadísticas descriptivas de las simulaciones entregadas en archivos de texto.

DIAGRAMA DE SECUENCIA

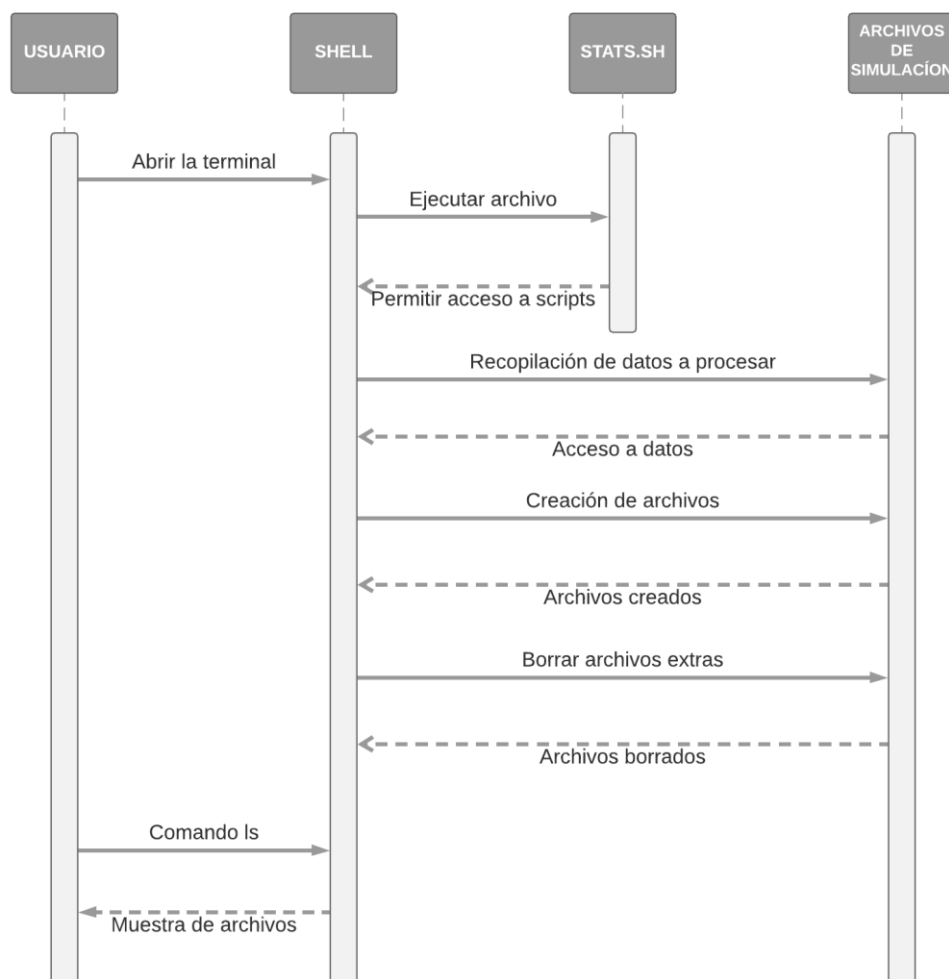


Figura 9

*Además esta solución puede resolverse creando archivos intermedios, los cuales deben ser eliminados antes de la finalización del script.

Referencias

- [1] Departamento de Ingeniería Telemática, ultimo acceso 11 de junio de 2020.
<http://trajano.us.es/~fjfj/shell/shellscript.htm>
- [2] Taller 01, Taller de Sistemas Operativos.
<https://github.com/taller-sistemas-operativos-uv/U1/blob/master/taller01/taller01.pdf>
- [3] Lucidchart, herramienta de diagramación basada en la web
<https://www.lucidchart.com/pages/es>