

Implementación de script en bash para el manejo de archivos

Taller de Sistemas Operativos

Escuela de Ingeniería Informática

Ignacio Alvarado Torreblanca

Ignacio.alvarado@alumnos.uv.cl

Resumen. *El objetivo principal de este informe es el desarrollo de un script en bash, el cual sigue el estándar unix, que permite procesar datos de archivos, pertenecientes a simulaciones de un sistema de evacuación de personas, con extensión .txt para realizar estadísticas descriptivas de estos datos. Todo esto desarrollado bajo un servidor Ubuntu Server 18.04 contenido en un ambiente virtual.*

1 Introducción

GNU Bash o simplemente Bash (Bourne-again shell) es un lenguaje de comandos y shell de Unix escrito por Brian Fox para el Proyecto GNU como un reemplazo de software libre para el shell Bourne. Lanzado por primera vez en 1989, se ha utilizado ampliamente como el shell de inicio de sesión predeterminado para la mayoría de las distribuciones de Linux y MacOS Mojave de Apple y versiones anteriores. Este lenguaje es ampliamente utilizado para generar scripts, ya que no es compilado sino que necesita ser interpretado para ejecutar el código, cuya ventaja radica en la automatización de procesos y esto permite que el manejo de datos sea más eficiente.

El desarrollo del trabajo se realiza en el lenguaje de comandos Bash incluido en un servidor Ubuntu Server 18.04, para la programación de un script que automatiza comandos para realizar estadísticas descriptivas de los datos contenidos en archivos con extensión txt.

La estructura del informe, está compuesta por una sección que introduzca sobre los conceptos primordiales del objetivo del informe, para luego presentar el trabajo realizado, dando un contexto de los datos, para luego explicar la descripción de los datos del problema que se utilizaran para el script, para pasar al diseño del problema donde se muestra un diagrama de alto nivel de la solución, los resultados y por último la conclusión.

2 Descripción del problema

2.1 Contexto de los datos a utilizar

Los datos a procesar corresponden a estadísticas que provienen de cierta cantidad de experimentos de simulación de un sistema de evacuación de personas, con un total de 75000 personas. Existen tres modelos de personas: Residentes, Visitantes Tipo I y Visitantes Tipo II. Los residentes son personas las cuales viven en la ciudad y conocen su zona segura, los visitantes tipo 1 son los que durante la evacuación logran conocer su

zona segura, por último, los visitantes tipo II son los que durante la evacuación no logran determinar su zona segura.

También tenemos una clasificación por grupo etario, la que se resume en la tabla 1.

Grupo Etario	Intervalo de edad
G0	0 - 14
G1	15 - 29
G2	30 - 64
G3	65 o más

Tabla 1: Clasificación de grupo etario

2.2 Organización de los datos

Cada simulación entrega los resultados ordenados en una estructura de directorios, en total son 3 archivos de texto por simulación, con un total de 11 simulaciones, estos archivos están acompañados por un identificador de la simulación llamado NNN, representado por un código numérico de tres dígitos.

2.3 Descripción de los datos

2.3.1 Archivo executionSummary.txt

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 Tabla 2 se detallan los campos y su respectiva descripción. En la figura 1 se muestra la estructura del archivo para un mejor diseño de la solución.

Figura 1: Estructura del archivo executionSummary.txt

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

Tabla 2: Descripción de los campos del archivo executionSummary.txt

Campo	Descripción	Ejemplo
numExperiment	Número del experimento.	0 (Corresponde al código 000)
Tsim	Tiempo de simulación	3600: segundos de tiempo real que se simuló.
CalibrationTime	Tiempo de calibración del simulador	100: segundo de tiempo de real que se destina a la calibración inicial de las personas.
Residents	Cantidad de Residentes simulados	69000
Visitors	Cantidad de Visitantes simulados	6000

timeExecMakeAgents	Tiempo real que demora en crear en memoria las personas simuladas	28252: milisegundos
timeExecCal	Tiempo real que demora la calibración 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

2.3.2 Archivo summary.txt

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 Tabla 3 se detallan los campos y su respectiva descripción. En la figura 2 se muestra la estructura del archivo para un mejor diseño de la solución.

Figura 2: Estructura del archivo summary.txt

```
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
```

Tabla 3: Descripción de los campos del archivo summary.txt

Campo	Descripción	Ejemplo
numExperiment	Número del experimento.	0 (Corresponde al código 000)
id	Identificador de la persona simulada	15
model	Identificar del modelo de la persona 0: Residente 1: Visitante tipo I 2: Visitante tipo II	0
gruoupAge	Identificador del grupo etario de la persona 0: G1, 1: G2, 2: G3, 3: G4	2

safeZone	Identificador de la zona segura de la persona Z1, Z2, Z3, Z4, Z5 NA: la persona no tiene zona segura asignada	Z1
distanceToTargetPos	Distancia a la que quedó la persona de su objetivo inicial	13.871407 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.	209.411742 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.	2300 segundos

2.3.3 Archivo usePhone.txt

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 Tabla 4 se detallan los campos y su respectiva descripción. En la figura 3 se muestra la estructura del archivo para un mejor diseño de la solución.

Figura 3: Estructura del archivo usePhone.txt

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

Tabla 4: Descripción de los campos del archivo usePhone.txt.

Campo	Descripción	Ejemplo
numExperiment	Número del experimento.	0 (Corresponde al código 000)

timeStamp	Tiempo de la medición	60
usePhone	Cantidad de persona que utilizaron el teléfono móvil en el tiempo especificado	4381

3 Análisis de los datos

El script que se desarrollará realizará estadísticas descriptivas de los datos, específicamente serán tres tareas:

- 1) Determinar Cantidad máxima, mínima y promedio para el tiempo de simulación total, este se define como: timeExecMakeAgents+ timeExecCal+ timeExecSim, y la Memoria utilizada por el simulador. Estos resultados de la tarea serán entregados en un archivo de texto metrics.txt En la tabla 5 se muestra la estructura de este archivo.

Tabla 5: Estructura del archivo metrics.txt

tsimTotal	promedio	min	max
memUsed	promedio	min	max

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

Los resultados de la tarea serán entregados en un archivo de texto evacuation.txt. En la tabla 6 se muestra la estructura del archivo.

Tabla 6: Estructura del archivo evacuation.txt

all	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

- 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 de la tarea serán entregados en archivo de texto usePhone-stats.txt. En la tabla 7 se muestra la estructura del archivo.

Tabla 7: Estructura del archivo usePhone-stats.txt

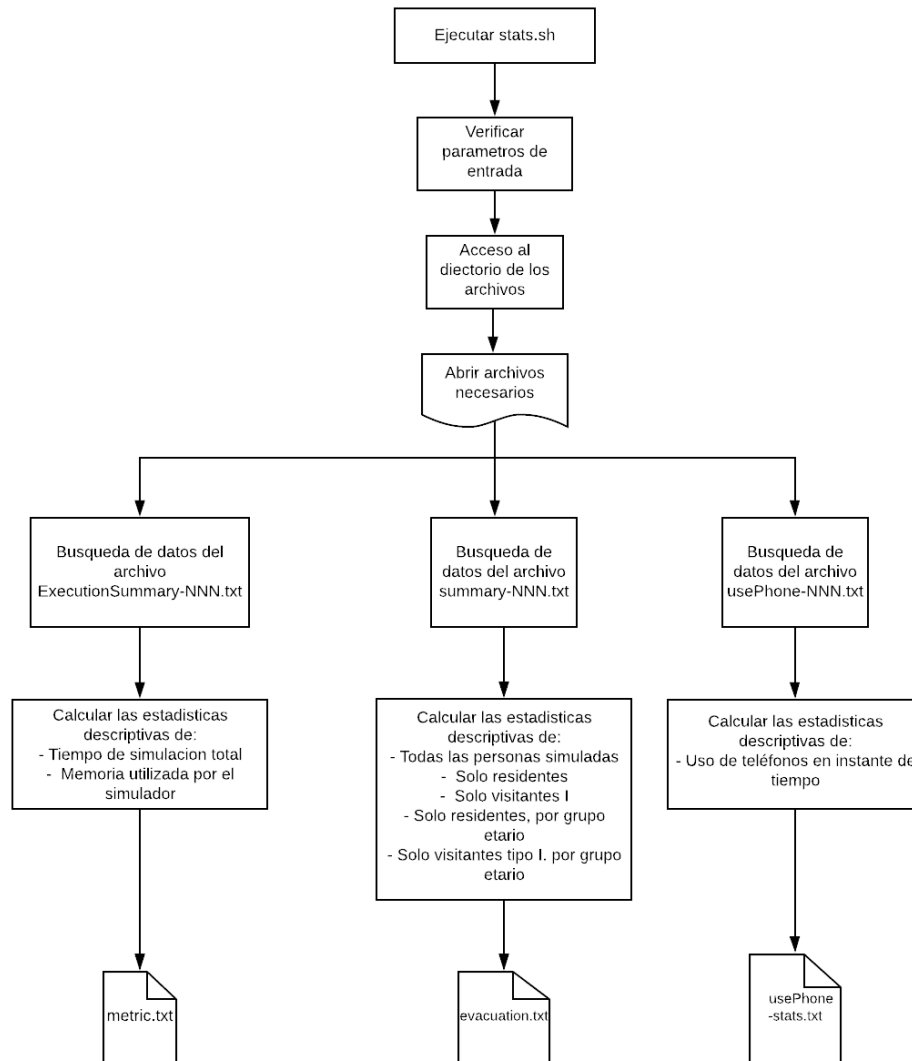
timestamp	promedio	min	max
-----------	----------	-----	-----

4 Diseño del problema

En este problema se dividen los procesos del script en tareas, cada tarea hace una lectura de cada archivo descrito anteriormente: executionSummary.txt, summary.txt y usePhone.txt. Estas tareas cumplen con distintos requerimientos que permiten la lectura de archivo y el procesamiento de los datos necesarios para realizar las estadísticas descriptivas. El código de cada tarea en este script es similar, por lo que varían en el archivo que leen y los datos de estos.

Para poder tener un desarrollo correcto del problema, se tienen que tener en cuenta los requerimientos que se necesitan, para este problema tenemos: Pasar por parámetro el directorio donde se encuentran los archivos a analizar, verificar que exista el directorio, acceder a los archivos de la simulación, toma de los datos requeridos en el problema, estadísticas descriptivas de estos datos (Mínimo, máximo y promedio), almacenar las estadísticas con la estructura de cada archivo correspondientemente. En la figura 4, se puede visualizar el diseño de alto nivel que explica a gran escala una solución del problema.

Figura 4: Diagrama de flujo del problema



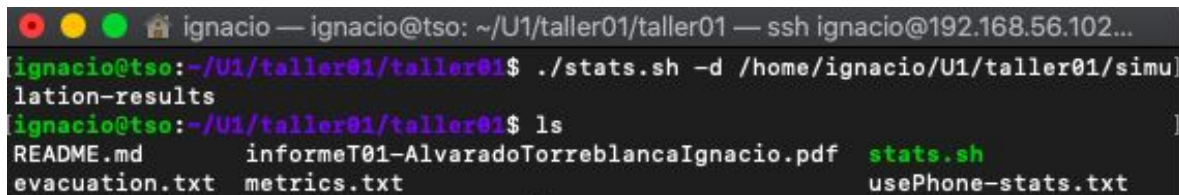
Vista la figura 4, se debe notar que al ejecutar el script tiene que existir parámetros de entrada, los cuales serán: “-d ” para ejecutar el script con el directorio correspondiente y “-h” para mostrar la forma de uso del script. Al tener el problema dividido en tres tareas, cada una de estas funciona de una forma similar, toma los archivos correspondientes, los leer, procesa los datos, calcula las estadísticas descriptivas y almacena los resultados dentro de los archivos correspondientes por tarea : metric.txt, evacuation.txt y usePhone-stats.txt. También se considerará utilizar archivos temporales para una mayor organización de los datos de cada tarea,

ya que necesitamos tener un orden y claridad de estos para el cálculo de las estadísticas descriptivas, una vez utilizados estos archivos temporales para almacenar los datos en el archivo final de cada tarea, serán removidos ya que no contendrán datos relevantes.

5 Resultados

En base al diseño propuesto se desarrolló el script, y se obtuvieron los resultados, donde como se puede ver en la figura 5, se tiene como salida a la ejecución de stats.sh a los archivos generados por cada tarea.

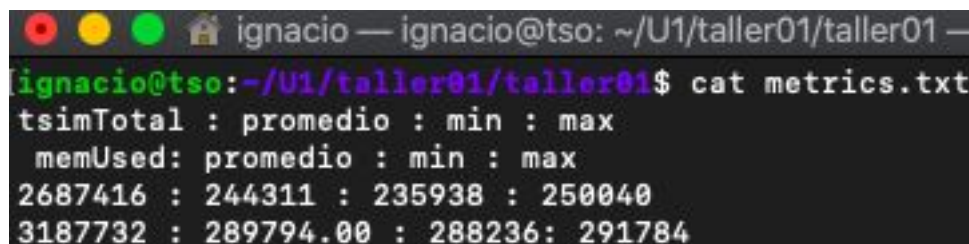
Figura 6: Archivos generados por el stats.sh



```
ignacio@tso: ~/U1/taller01/taller01 — ssh ignacio@192.168.56.102...
ignacio@tso:~/U1/taller01/taller01$ ./stats.sh -d /home/ignacio/U1/taller01/simu
lation-results
ignacio@tso:~/U1/taller01/taller01$ ls
README.md      informeT01-AlvaradoTorreblancaIgnacio.pdf  stats.sh
evacuation.txt metrics.txt                                usePhone-stats.txt
```

Para cada tarea, se generaban distintos archivos los cuales contenían las estadísticas descriptivas calculadas por el script. Para la tarea 1 se genera el archivo metrics.txt, el cual posee el tiempo de simulación total de todas las simulaciones, junto al promedio, mínimo y máximo; considerando también la memoria utilizada por toda la simulación, promedio, mínimo y máximo de esta. En la figura 7, se puede ver el resultado final de la tarea 1, la que está dentro del archivo metrics.txt.

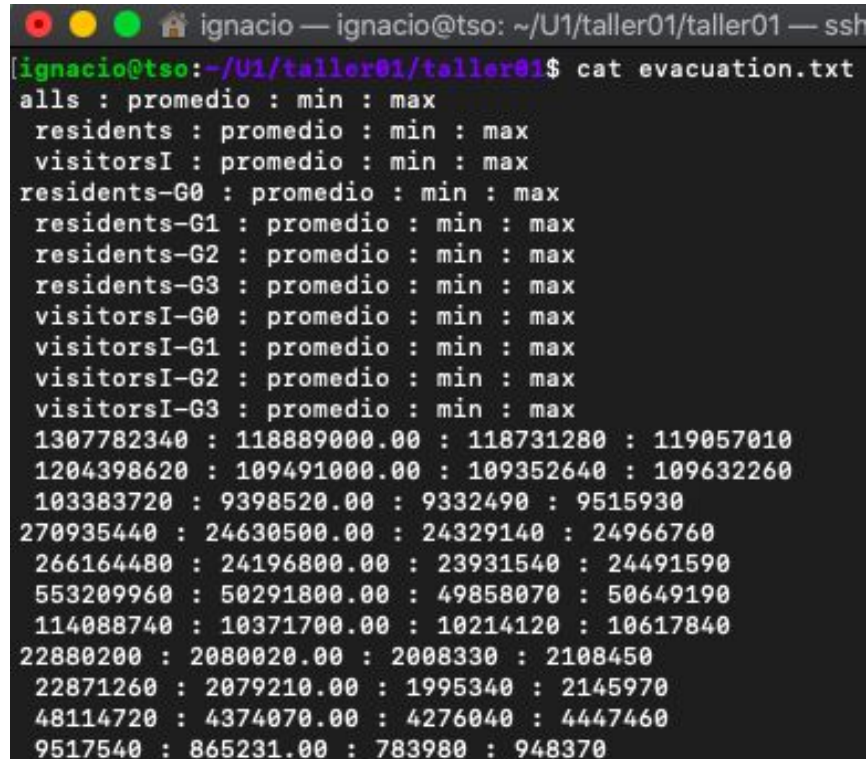
Figura 7: Datos del archivo metrics.txt



```
ignacio@tso: ~/U1/taller01/taller01 —
ignacio@tso:~/U1/taller01/taller01$ cat metrics.txt
tsimTotal : promedio : min : max
memUsed: promedio : min : max
2687416 : 244311 : 235938 : 250040
3187732 : 289794.00 : 288236: 291784
```

Para la tarea 2 se genera el archivo evacuation.txt, el cual posee el promedio del tiempo de evacuación, el mínimo y el máximo de los distintos grupos de personas requeridos, como lo son: todas las personas simuladas, todos los residentes y visitantes tipo I, separando a cada uno de estos por grupo etario. En la figura 8, se puede ver el resultado final de la tarea 2, la que está dentro del archivo evacuation.txt.

Figura 8: Datos del archivo evacuation.txt

A terminal window with a dark background and light-colored text. The window title bar shows three colored circles (red, yellow, green) and the text 'ignacio — ignacio@tso: ~/U1/taller01/taller01 — ssh'. The prompt is '[ignacio@tso:~/U1/taller01/taller01\$'. The command 'cat evacuation.txt' has been executed, displaying the following data:

```
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
1307782340 : 118889000.00 : 118731280 : 119057010
1204398620 : 109491000.00 : 109352640 : 109632260
103383720 : 9398520.00 : 9332490 : 9515930
270935440 : 24630500.00 : 24329140 : 24966760
266164480 : 24196800.00 : 23931540 : 24491590
553209960 : 50291800.00 : 49858070 : 50649190
114088740 : 10371700.00 : 10214120 : 10617840
22880200 : 2080020.00 : 2008330 : 2108450
22871260 : 2079210.00 : 1995340 : 2145970
48114720 : 4374070.00 : 4276040 : 4447460
9517540 : 865231.00 : 783980 : 948370
```

Por último, para la tarea 3 se genera el archivo usePhone-stats.txt, el cual posee el promedio de uso de teléfonos móviles, el mínimo y el máximo para los instantes de tiempo. En la figura 9, se puede ver el resultado final de la tarea 3, la que está dentro del archivo usePhone-stats.txt.

Figura 9: Datos del archivo usePhone.txt

```
ignacio — ignacio@tso: ~/U1/taller01/taller01 — ssh ignacio
ignacio@tso:~/U1/taller01/taller01$ cat usePhone-stats.txt
#timestamp:promedio:min:max
1:4453.91:4530:4333
2:4428.55:4481:4369
3:4451.82:4585:4261
4:4439.91:4558:4372
5:4454.18:4529:4327
6:4415.64:4531:4297
7:4424.45:4480:4316
8:4437.36:4532:4367
9:4452.73:4506:4414
10:4434:4505:4350
11:4452.18:4517:4364
12:4435.73:4542:4298
13:4433.27:4558:4342
14:4400.55:4479:4291
15:4437.73:4550:4308
16:4407:4491:4306
17:4444.45:4553:4323
18:4400.55:4511:4315
19:4450.73:4571:4359
20:4446.36:4549:4329
```

6 Conclusión

En conclusión, los archivos de salida del script entregan los datos requeridos de las estadísticas descriptivas de cada simulación, donde en base a estos se pueden tomar medidas para una evacuación real en la ciudad de Iquique.

También, como recomendación a una futura reestructuración del script, se puede visualizar que se repiten ciertos pedazos de código en cada tarea, la cuales podrían ser optimizadas para una mayor eficiencia en los tiempos de ejecución del script. Como también podría reducirse el uso de archivos temporales, ya que se hizo un uso muy elevado de estos.