DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico Nº 3

Algoritmos en sistemas distribuidos

Primer Cuatrimestre de 2017 — Sistemas Operativos

Integrante	LU	Correo electrónico
Langberg, Andrés	249/14	andreslangberg@gmail.com
Sticco, Patricio	337/14	patosticco@gmail.com
Cadaval, Matías	345/14	matias.cadaval@gmail.com

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

http://www.fcen.uba.ar

Índice

1. Introduccion

En el siguiente trabajo práctico, se pide desarrollar un `DistributedHashMap`, una nueva versión del conocido `ConcurrentHashMap` del TP2, donde se busca poner en práctica el conocimiento sobre sistemas distribuidos a través del envío de mensajes para varias de sus funcionalidades.

Se utilizara la interfaz `MPI`, provista por `C++`. La idea es poder ejecutar el cómputo distribuido en varias máquinas, donde cada una de ellas podrá correr más de un proceso o *nodo*.

Entrando en detalle sobre la estructura del sistema, el mismo contará con un *nodo* distinguido (la **consola**), que se encargará de la entrada y salida de los comandos y sus resultados. A continuación, presentamos dichas funcionalidades:

- `load`(list<string> *params*): Toma como parámetro una lista de archivos de texto y carga las palabras de estos archivos en el `DistributedHashMap`. Deberá utilizar un nodo para cada archivo. Si hay más archivos que nodos, deberá enviarle un nuevo archivo a cada nodo que vaya liberándose.

- `addAndInc`(string *key*): Si *key* existe en el `DistributedHashMap`, incrementa su valor. Si no existe, crea el par (*key*, 1) en algún nodo. No tiene por qué ser procesada siempre por el mismo nodo, ni por uno que ya contenga la clave. Deberá enviarse el comando a todos los nodos; el primer nodo que lo tome deberá avisar que lo ha hecho, y el resto no deberá ejecutarlo.

- `member`(string key): Dado un string, determina si el mismo existe o no en el sistema.

- `maximum`(): Busca el string que más veces aparece y su valor. Luego, imprime el resultado por pantalla. La consola tendrá que pedir a cada nodo las palabras allí almacenadas, y deberá poder recibirlas de distintos nodos, sin que necesariamente todas las de un mismo nodo lleguen juntas. El `HashMap` dado provee un iterador (`HashMap::iterator`) que permite obtener una a una las palabras almacenadas. Una vez que la consola tenga el HashMap completo podrá calcular el resultado. Si un mismo nodo utiliza múltiples mensajes para enviar sus palabras a la consola, debe realizar una única llamada a `trabajarArduamente`(), en lugar de una por cada mensaje.

- `quit`(): Avisa a todos los nodos que deben finalizar y liberar sus recursos.

Más adelante en este informe será detallada la implementación de cada uno de estos comandos.

El resto de los *nodos* serán indistinguibles, y cada uno de ellos tendrá su propio `HashMap`, cuya implementación es provista por la cátedra.

2. Desarrollo

A continuación, haremos una breve descripción sobre la implementación de cada una de las funcionalidades previamente mencionadas, enfatizando sobre la manera en la que se realiza el cómputo distribuido, utilizando la interfaz `MPI`.

En cuanto al llamado de cada función, la implementación de cada una de ellas empieza mediante un envío bloqueante, a través del cual el *nodo* consola le avisa al conjunto de nodos la tarea a realizar. Por su parte, el resto de los nodos reciben de manera bloqueante el mensaje, para luego ejecutar el handler indicado.

La implementación que construimos siempre utiliza operaciones bloqueantes ya que los mensajes enviados y/o recibidos siempre son indispensables en el momento en que son enviados o recibidos, para el correcto funcionamiento del módulo. Por otra parte, el nodo consola siempre espera a que los nodos que están ejecutando tareas finalicen, antes de devolverle el control al usuario. Decidimos hacerlo de esta manera para evitar que se superpongan operaciones, ocasionando resultados indeseados (por ejemplo si un llamado a `load` y otro a `member` se superponen).

2.1. `load`

En esta función el nodo consola itera sobre la lista de archivos recibida por parámetro y, en cada iteración, le asigna un archivo al primer nodo de la cola de nodos libres. Si la cola está vacía, el nodo consola espera (recibiendo de manera bloqueante) a que un nodo le avise mediante un mensaje bloqueante, que finalizó de procesar un archivo, y que por ende ya está disponible. Una vez recibido el mensaje, el nuevo nodo libre es encolado en la cola previamente mencionada.

Ya asignados todos los archivos, el nodo consola espera a que todos los nodos le informen (otra vez mediante un mensaje bloqueante), que terminaron la ejecución de la tarea de carga que tenían asignada. Cuando todos terminan, se le indica al usuario que el load fue realizado con éxito.

2.2. `addAndInc`

Para realizar un `addAndInc` en nuestro `DistributedHashMap`, necesitamos de alguna manera "elegir" el nodo en el cual vamos a agregar la nueva clave, de manera no determinística. Por lo tanto, en primera instancia la consola recibe de cada nodo su *rank* mediante recibos bloqueantes. El primero en llegar, será el nodo en el que se agregará la clave.

A este nodo, se le enviará la clave mediante un envío bloqueante y el mismo la agregara en su `HashMap` local. Finalmente, se le avisa a la consola que ya se termino con el agregado.

Con respecto al resto de los nodos, se les comunica que no van a tener que incorporar la clave a su diccionario.

2.3. `member`

Como previamente mencionamos, esta función se encarga de indicarle al usuario si la clave dada como parámetro esta definida o no en el `DistributedHashMap`. Para ello, el nodo consola les envía mediante *broadcasting* bloqueante (usando `MPI_Bcast`) la clave a buscar, a todos los nodos.

Por su parte, los nodos reciben la clave y verifican si está definida en su `HashMap` mediante la función `member`. Luego, le envían el resultado al nodo consola mediante un mensaje bloqueante.

Mientras va recibiendo las respuestas, el nodo consola aplica sucesivamente la operación lógica `OR` entre el resultado guardado hasta el momento, y el nuevo resultado proveniente de

algún nodo. De esta manera, al recibir todos los mensajes, ya queda determinado el resultado final. Por último, le informa al usuario el resultado.

2.4. `maximum`

Inicialmente el nodo consola crea un `HashMap`, en el que irá agregando las claves (y sus repeticiones) pertenecientes a los `HashMaps` de cada uno de los nodos. Para lograrlo, itera indefinidamente en un ciclo, hasta que todos los nodos le informan que ya enviaron todas las palabras de sus respectivos `HashMaps`. En cada iteración, consola recibe mediante un mensaje bloqueante, un `string` proveniente de alguno de los nodos. Si el `string` es *"fin de hashmap"*, consola incrementa en uno una variable local, en la que almacena la cantidad de nodos que ya enviaron todas sus claves (esta sirve para determinar la condición de terminación del ciclo). Caso contrario, el nodo consola agrega mediante `addAndInc` el `string` recibido, a su propio `HashMap`.

Una vez que todos los nodos ya enviaron el mensaje de finalización, el ciclo finaliza, y el nodo consola busca el máximo en su `HashMap`, mediante la operación `maximum`. Finalmente le informa al usuario el resultado arrojado por dicha función.

2.5. `quit`

Esta operación es muy simple. Cuando los nodos reciben el mensaje de `quit` salen del ciclo en el que están iterando continuamente a la espera de nuevas tareas, mediante un `break`.

El nodo consola sólo le avisa a los nodos que finalicen, y termina.

3. Tests

Para probar el correcto funcionamiento de nuestra implementación para la clase `DistributedHashMap` realizamos una serie de tests que ejecutan las operaciones ofrecidas por la misma, en diferentes situaciones. Todos los tests disponibles están definidos en el archivo *Makefile*. Para entender más fácilmente su funcionamiento y finalidad, a continuación presentamos una breve explicación de cada uno de ellos. En algunos casos, adjuntaremos los comandos que ejecutan y los explicaremos uno a uno, para entender cada paso.

Además, para realizar estos tests agregamos el comando `print`, para fácilmente poder imprimir por stdout el contenido del `DistributedHashMap`.

3.1. Modo de uso

Antes de ejecutar cualquiera de los tests se debe compilar el código mediante `make`. Por su parte, los tests se puede ejecutar de varias formas:

- Todos: `make test-all` ejecuta todos los tests, uno tras otro.
- Por función: `make test-load` ejecuta todos los tests para la función `load`, uno tras otro. Es análogo para `addAndInc`, `member` y `maximum`.
- Test particular: `make <nombre-test>`. Ejemplo: `make test-maximum-3`

3.2. Tests para la función `load`

- `test-load-1`: invoca el comando `load` con una cantidad de archivos menor que la cantidad de nodos trabajadores (2 archivos, 3 nodos). Luego se llama a `print` y se compara el output con el resultado esperado.

- `test-load-2`: invoca el comando `load` con una cantidad de archivos mayor que la cantidad de nodos trabajadores (5 archivos, 2 nodos). Luego se llama a `print` y se compara el output con el resultado esperado.
- `test-load-3`: invoca el comando `load` con una cantidad de archivos igual que la cantidad de nodos trabajadores (4 archivos, 4 nodos). Luego se llama a `print` y se compara el output con el resultado esperado.

Explicación paso a paso de `test-load-2`:

```
test-load-2:
  ########## test-load-2 ##########

  # Genera el archivo de salida que es esperado como resultado, para luego compararlo
  awk -f corpus.awk corpus | sort > corpus-post

  # Genera archivos de entrada usados para el load a partir de corpus (divide corpus en 5)
  for i in 0 1 2 3 4; do sed -n "$$((i * 500 + 1)),$$(((i + 1) * 500))p"
  corpus > corpus-"$$i"; done

  # Genera lista de comandos que recibe el nodo consola y los guarda en input
  printf " load corpus-0 corpus-1 corpus-2 corpus-3 corpus-4 \n print \n q \n" > input

  # Ejecuta el codigo con la lista de comandos guardada en input y luego procesa el output mediante
  # sed y sort, para compararlo contra corpus-post usando diff. El awk es para mostrar un mensaje
  # de Test OK si es exitoso, o Test FAILED en caso contrario
  cat input | mpiexec -np 3 ./dist_hashmap | sed -e '1,/finalizado/d' | sed '/>/d' | sort
  | diff -u - corpus-post | awk '{ print } END { if (NR) print "Test FAILED"; else print "Test OK" }'

  # Elimina todos los archivos generados
  rm -f input corpus-post corpus-[0-4]
```

3.3. Tests para la función `addAndInc`

- `test-addAndInc-1`: invoca el comando `addAndInc primerstring` sobre un `DistributedHashMap` vacío (np = 3). Luego se llama a `print` y se compara el output con el resultado esperado.
- `test-addAndInc-2`: invoca los comandos `addAndInc nuevapalabrauno` y `addAndInc nuevapalabrados` sobre un `DistributedHashMap` con los archivos *corpus-0* y *corpus-1* cargados (np = 3). Luego se llama a `print` y se compara el output con el resultado esperado.
- `test-addAndInc-3`: invoca los comandos `addAndInc archivo`, `addAndInc estabilidad` y `addAndInc funcional` sobre un `DistributedHashMap` con el archivo *corpus* cargado (np = 3). Luego se llama a `print` y se compara el output con el resultado esperado.

Explicación paso a paso de `test-addAndInc-2`:

```
test-addAndInc-2:
  ########## test-addAndInc-2 ##########

  # Genera lista de comandos que recibe el nodo consola y los guarda en input
  printf " load corpus-0 corpus-1 \n addAndInc nuevapalabrauno \n addAndInc nuevapalabrados
  \n print \n q \n" > input

  # Le agrega a corpus las dos nuevas claves y lo guarda en corpus-aux
  cat corpus > corpus-aux && printf "\nnuevapalabrauno\nnuevapalabrados" >> corpus-aux

  # Genera el archivo de salida que es esperado como resultado, para luego compararlo
  awk -f corpus.awk corpus-aux | sort > corpus-post

  # Genera archivos de entrada usados para el load a partir de corpus (divide corpus en 2)
  for i in 0 1; do sed -n "$$((i * 1250 + 1)),$$(((i + 1) * 1250))p" corpus > corpus-"$$i"; done

  # Ejecuta el codigo con la lista de comandos guardada en input y luego procesa el output mediante
  # sed y sort, para compararlo contra corpus-post usando diff. El awk es para mostrar un mensaje
  # de Test OK si es exitoso, o Test FAILED en caso contrario
  cat input | mpiexec -np 3 ./dist_hashmap | sed -e '1,/nuevapalabrados/d' | sed '/>/d' | sort
  | diff -u - corpus-post | awk '{ print } END { if (NR) print "Test FAILED"; else print "Test OK" }'

  # Elimina todos los archivos generados
  rm -f input corpus-aux corpus-post corpus-[0-1]
```

3.4. Tests para la función `member`

- `test-member-1`: invoca los comandos `member noestoy` y `member yotampocoestoy` (no definidas) sobre un `DistributedHashMap` vacío (np = 3). Luego se filtra el output y se comparan los resultados arrojados por la función contra los resultados esperados.
- `test-member-2`: invoca los comandos `member noestoy` y `member yotampocoestoy` (no definidas) sobre un `DistributedHashMap` con los archivos *corpus-0*, *corpus-1*, *corpus-2* y *corpus-3* cargados (np = 5). Luego se filtra el output y se comparan los resultados arrojados por la función contra los resultados esperados.
- `test-member-3`: invoca los comandos `member archivo` y `member estabilidad` (sí definidas) sobre un `DistributedHashMap` con los archivos *corpus-0*, *corpus-1*, *corpus-2* y *corpus-3* cargados (np = 5). Luego se filtra el output y se comparan los resultados arrojados por la función contra los resultados esperados.
- `test-member-4`: invoca el comando `member nuevaclave`, luego `addAndInc nuevaclave` y `member nuevaclave` (primero no definida, luego sí) nuevamente, sobre un `DistributedHashMap` con el archivo *corpus* cargado (np = 2). Luego se filtra el output y se comparan los resultados arrojados por la función contra los resultados esperados.

Explicación paso a paso de `test-member-4`:

```
test-member-4:
  ########## test-member-4 ##########

  # Genera lista de comandos que recibe el nodo consola y los guarda en input
  printf " load corpus \n member nuevaclave \n addAndInc nuevaclave \n member nuevaclave \n
  q \n" > input

  # Genera el archivo de salida que es esperado como resultado, para luego compararlo
  printf "La clave nuevaclave NO ESTA DEFINIDA\nLa clave nuevaclave ESTA DEFINIDA\n" > output

  # Ejecuta el codigo con la lista de comandos guardada en input y luego procesa el output mediante
  # sed y grep, para compararlo contra ouput usando diff. El awk es para mostrar un mensaje
  # de Test OK si es exitoso, o Test FAILED en caso contrario
  cat input | mpiexec -np 2 ./dist_hashmap | grep DEFINIDA | sed 's/> //' | diff -u - output
  | awk '{ print } END { if (NR) print "Test FAILED"; else print "Test OK" }'

  # Elimina todos los archivos generados
  rm -f input output
```

3.5. Tests para la función `maximum`

- `test-maximum-1`: invoca el comando `maximum` sobre un `DistributedHashMap` con el archivo *corpus-max* cargado (np = 3). Luego se filtra el output y se compara el resultado arrojado por la función contra el resultado esperado.
  Resultado esperado: <clavemaxima, 9>

- `test-maximum-2`: invoca el comando `maximum` sobre un `DistributedHashMap` con los archivos *corpus-max-0*, *corpus-max-1*, *corpus-max-2*, *corpus-max-3* y *corpus-max-4* cargados (np = 6). Luego se filtra el output y se compara el resultado arrojado por la función contra el resultado esperado. En este caso *clavemaxima* no es maximo en ninguno de los archivos localmente, pero si globalmente.
  Resultado esperado: <clavemaxima, 9>

- `test-maximum-3`: invoca el comando `maximum` sobre un `DistributedHashMap` con el archivo *corpus-max* cargado (np = 3). Hecho esto, ejecuta diez veces seguidas `addAndInc nuevomax`, y nuevamente ejecuta `maximum`. Luego se filtra el output y se compara el resultado arrojado por la función contra el resultado esperado.
  Resultado esperado: <clavemaxima, 9> y luego <nuevomax, 10>

Explicación paso a paso de `test-maximum-3`:

```
test-maximum-3:
  ########## test-maximum-3 ##########

  # Genera lista de comandos que recibe el nodo consola y los guarda en input
  printf " load corpus-max \n maximum \n" > input && for i in 0 1 2 3 4 5 6 7 8 9;
  do printf " addAndInc nuevomax \n" >> input; done && printf " maximum \n q \n" >> input

  # Genera el archivo de salida que es esperado como resultado, para luego compararlo
  printf "El maximo es <clavemaxima, 9>\nEl maximo es <nuevomax, 10>\n" > output

  # Ejecuta el codigo con la lista de comandos guardada en input y luego procesa el output mediante
  # sed y grep, para compararlo contra ouput usando diff. El awk es para mostrar un mensaje
  # de Test OK si es exitoso, o Test FAILED en caso contrario
  cat input | mpiexec -np 3 ./dist_hashmap | grep maximo | sed 's/> //' | diff -u - output
  | awk '{ print } END { if (NR) print "Test FAILED"; else print "Test OK" }'

  # Elimina todos los archivos generados
  rm -f input output
```