Monitoring the behavior of a simple data series program

by Indra Bayu

Producer

Produces a random duration, and delays based on it

Delivers the integer to the consumer by using NamedPipe (blocking)

Continues with step #1

Main

Program

Fires up task 1 and task 2

Waits for the input from the Producer (blocking)

Pairs a timestamp (to determine its age) with it

Delivers it to task 1 and task 2 (non-blocking)

Calls V on semaphore 1 and 2 (non-blocking)

Continues with step #2

Consumer

Task 1

Task 2

Reads the stream from the Main Program after calling P on semaphore 1 (blocking)

Flushes the input to Console.Out, as a line of string

Continues with step #1

Fires up thread which is terminated 1 second later

Waits for 1 second while worker thread is crunching numbers and prints the results to Worker Console.Error Thread (see the details in the

next slide).

Stops that worker thread

Continues with step #1

What constitutes one second in the worker thread which is fired up by Task 2 in the previous slide?

Removal of inputs (from the head of the linked list) which will be outdated by the next 1 second. Iterating the linked list to get the first (oldest) and last (youngest) elements, and compute the average and total. Running the blocking GC to sweep the released items from step 1 (important to minimize the chance that GC runs randomly in other steps).

While not yet aborted by the creator thread (Task 2), moving the incoming data from the queue's head into the back of the linked list.

How to view the analysis from the Consumer process?

- In the SETTINGS region, change the boolean constant variable DEBUG to true. Even though the computation will run as usual, this setting will disable the supposed output from Output & Error streams, and replaces them with a line to the Default Output every 1 second (the interval you normally would expect from stream 2).
- Besides the 4 mandatory things (average {0}, QTI {1}, oldest element {2}, and youngest element {3}), this analysis outputs also: jitter {5}, memory in MB {6}, effective cycle duration in milliseconds {7}, items served in the linked list {8}, GC duration in milliseconds {9}, and how long has the consumer process run in minutes {10} & seconds {11}.
- In the terminal's title you can see the time taken by Task 2 outside of the 1 second wait. This number estimates the overall excess so far.

How to estimate the memory usage?

- At the beginning of the program, Consumer asks Producer for its Process ID.
- At the end of each 1-second cycle, the Consumer sums its memory usage with the Producer's.
- In .NET, this is done through the property of a Process class instance, called WorkingSet64. This property returns the amount of physical memory allocated for the associated process. The program divides it with 1048576, in order to convert it to megabyte.
- It's basically a shortcut for reading this performance counter:

new PerformanceCounter("Process", "Working Set", "ProcessName").RawValue;

How to estimate the stream-to-stream latency?

- A timestamp (reflecting the current time) is attach to the input at its arrival in stream 0. So after parsing the string input into an integer, we pair the timestamp with the integer (making a kind of tuple out of them).
- The semaphore in Task 1 waits for a flag from the main program (recall that the main program receives input from the Producer and copies the value to the corresponding queues available: each for Task 1 and 2). Semaphore is chosen to avoid busy wait.
- After being flushed to stream 1 (or the *out* stream), Task 1 gets the time difference between the current time and the timestamp attached to the input. That's the latency.
- Now, each time the 1-second-cycle wants to print, it will take the current latency value available. NOTE: at the beginning of every cycle, the latency value is reset to zero, in order to read an *incorrect* value (one already used in previous cycle) whenever there's no new latency value available!

How to estimate the jitter?

- The jitter is total time taken by the whole duration (across cycles) of program for the purpose of computation only, and divide it by how many successful iterations has happened so far.
- A cycle is considered successful if it can manage to finish crunching numbers for inputs within the 10 seconds timeframe (regardless of how many they are, or how quick they come).
- You can modify the program to show the total amount of time spent inside the cycle, which means counting in 1 the time involved in removing inputs older than 10 seconds, and 2 the time used by semaphore of Task 2 waiting for new inputs to come (and inserting that new value at the back of the linked list).
- Such hack is done by changing these values in the SETTINGS region:
 - StopwatchForRawInputs to Timing.ComputationAndWait
 - ConsiderOutdated to true

What's the output look like?

- It depends on how often the input comes! To make it simple, I will split it into three scenarios. The last two will be assumed as edge cases.
- In scenario 1, the inputs come between 0 to 2000 milliseconds.
- In scenario 2, they come a little late, between 5 to 15 seconds. Hence it is possible that there is nothing to process within the 10 seconds time frame (idle). You can increase these numbers as high as you want.
- In scenario 3, they come as often as your computer can produce input to stream 0. To put it simple: no delay between inputs.
- We do this by monitoring the first 44 seconds of the Consumer (if you also want to do this, remember to set the DEBUG to true).

Are those 1 second spans? Somewhat yes.

- The last column is a good indicator.
- Currently, the computation time considers <u>only</u> the time needed to process inputs within the 10 seconds time frame.
- But within the 1 second cycle, there are other things happening too, as mentioned in slide #3, but minus the GC time: (1) removing items ≤ 10 seconds and (2) waiting for input to come from stream 0 then put it to the back of the linked list. So *ideally* the sum of these two plus the GC time = 1 second (except when there's nothing in the list).
- Therefore, I will include those times into the computation time, to see whether the sum from the durations really constitutes the one second. You can see this in the next slide (slide #13).

Scenario 3: no delay between inputs (with the wait time considered)

In scenario 3, the computation time doesn't take the majority of the time. Why so?

- The time needed to transfer between streams (and between processes too) seems to be the major bottleneck. This makes the thread to use most of it's time waiting for input instead of doing calculation on massive amount of elements in the list.
- Can this be improved, if inter-process communication is not the case? YES. This is logical, since we can use shared collection instead.
- Currently the maximum time for processing items ≤ 10 seconds (as seen in slide #11) is 135 milliseconds, and for the number of items in the list itself is 1.1 millions. We will compare this to when there's only 1 process, so everything is done using shared variable/collection.
- I'll do this twice, for blocking and unblocking GC.

Scenario 3 (no delay) with blocking GC in a single process Producer-Consumer.



Scenario 3 (no delay) with **non-**blocking GC in a single process Producer-Consumer.

What just happened when the program is ported to run in a single process?(and redoing scenario 3)

- The time taken to do the computation increased from the maximum of 135 milliseconds into the maximum of 527 milliseconds (for blocking GC) and 658 milliseconds (for non-blocking GC).
- The number of items in the list being processed within that time is up to 5.0 millions (compare to the 1.1 millions previously). This means increase in throughput to 4.54 times.
- Did you notice the (blocking) GC time raised from ±75-83 to 254-347 milliseconds? That corresponds to the increase of memory consumed from ±69-74 (60-62 if Producer's isn't considered) to 189-216 MB. Obviously the size of the data stored in memory corresponds to how long GC takes to run.
- Jitter raised from \pm 0.11 to 0.38 seconds, and latency is reduced by \pm 50%.
- Conclusion: by porting to single process, the overall computer utilization can be raised to ±4 times whenever there's no delay between inputs.