# **Core vs Thread CPU Scaling**

Author: Craig Shallahamer (craig@orapub.com), Version 1a, 9-June-2011

## **Background and Purpose**

As CPU subsystems become more complex and employ various methods to utilize multiple CPU cores and multiple threads per core, determining the CPU requirements, capacity, and the resulting utilization is not entirely clear. But it is important we understand what the utilization tells us about our system. As an Oracle DBA, I'm OK with not knowing the specific OS utilization calculations, but I'm not OK with blindly stating a utilization figure without understanding what that means in relation to performance. Hence my quest...

The purpose of this notepad is to:

- 1. Observe how a CPU subsystem handles multiple processes launched simultaneously.
- 2. Understand if a CPU subsystem is more core or thread processing focused.

Where C is the total number of CPU cores, a core power focused system will complete C number of processes launched simultaneously at pretty much the same time. Once the number of processes is greater than C, the all process completion time will increase.

Where T is the total number of CPU threads, a thread powered focused system will complete T number of processes launched simultaneously at pretty much the same time. Once the number of processes is greater than T, the all process completion time will increase.

My observations has shown that CPU subsystems without threads tend to be completely core powered. (As you would expect.) However, CPU subsystems with threads can be either more core or more thread focused.

This focus (core or thread) impact CPU requirement, capacity, and utilization calculations.

The notepad was used in support of my blog entry, Cores vs Threads... Part 3 posted on June 10, 2011.

### Data Loading, Processing, and Basic Statistics

This data was gathered on a single 4 four core CPU with no threads. The number of very CPU intensive processes where simultaneously launched and when they all completed the elapsed time was gathered.

I created the histogram based on the elapsed time when 2 processes where launched simultanteously. The histogram visually does not look normal and the distribution fit test produced a P-value less than 0.05 (my chosen alpha) and therefore is stratistically NOT normally distributed.

```
In[199]:=
        sampleSet = Import["/Users/cshallah/Desktop/cpuThreadTestResults.txt", "Data"];
        procsMax = 10;
        theGrid = {{"Simul\nProcs", "Avg\nElapsed (s)",
             "Std Dev\nElapsed", "Median\nElapsed (s)", "Samples"}};
        theResults = {};
        Table[
          elapsedSet = Select[sampleSet, #[[2]] == procs &][[All, 3]];
          elapsedMean = N[Mean[elapsedSet]];
          elapsedSD = N[StandardDeviation[elapsedSet]];
          elapsedMed = N[Median[elapsedSet]];
          sampleCount = Length[elapsedSet];
          AppendTo[theResults, {procs, elapsedMean, elapsedSD, elapsedMed, sampleCount}];
          AppendTo[theGrid, {procs, elapsedMean, elapsedSD, elapsedMed, sampleCount}];
           {procs, 1, procsMax}
         ];
        theResults
        Grid[theGrid, Frame \rightarrow All]
        ListPlot[Table[theResults[[procs]][[2]], {procs, 1, 10}],
         Filling \rightarrow Axis, AxesLabel \rightarrow {"Simul Procs", "Elapsed (s)"}]
        Export["/Users/cshallah/Desktop/coreVsThreadGrid.jpg", Grid[theGrid, Frame → All]]
Out[204]=
        \{\{1, 78.129, 0.805892, 78., 31\}, \{2, 85.3226, 0.944708, 85., 31\},\
         \{3, 85.7742, 0.920495, 86., 31\}, \{4, 85.8387, 0.582911, 86., 31\},\
```

```
{5, 128.323, 0.701764, 128., 31}, {6, 129.387, 1.72583, 129., 31},
{7, 150.935, 1.76891, 150., 31}, {8, 171.613, 0.667204, 172., 31},
{9, 200.839, 1.86363, 200., 31}, {10, 215.258, 1.43684, 215., 31}}
```

Cimul	7	C+d Dee	Madian	Complea
SIMUL	Avg	sta Dev	меатап	Sampres
Procs	Elapsed (s)	Elapsed	Elapsed (s)	
1	78.129	0.805892	78.	31
2	85.3226	0.944708	85.	31
3	85.7742	0.920495	86.	31
4	85.8387	0.582911	86.	31
5	128.323	0.701764	128.	31
6	129.387	1.72583	129.	31
7	150.935	1.76891	150.	31
8	171.613	0.667204	172.	31
9	200.839	1.86363	200.	31
10	215.258	1.43684	215.	31



2

Out[205]=

Out[207]=

/Users/cshallah/Desktop/coreVsThreadGrid.jpg

## **Data Analysis**

Just to be a little more thorough, I wanted to ensure that statistically there is a real difference between the elapsed times when 4 and 5 processes where launched simultaneously. This plot above makes it look like there is obviously a difference, but I wanted to perform a statistical significance test just be sure.

First I need to ensure my two sample sets (of elapsed times) are normally distributed. If so I can perform a simple t-test but if not, I will need to perform, what is called, a location test. A good introductory reference is: http://www.statsoft.com/textbook/nonparametric-statistics

I created the histogram based on the elapsed time when 2 processes where launched simultanteously. The histogram visually does not look normal and the distribution fit test produced a P-value less than 0.05 (my chosen alpha) and therefore is stratistically NOT normally distributed.

The cell directly below is just a check to ensure I'm doing the significant testing correctly. So I create a normally distributed data set (1000 values) based on the mean and standard deviation from the elapsed times from, in this case, when the processes were 4. And then, I do a normality check using *Mathematica's* DistributionFitTest function. The result should be clearly greater than 0.05. This will reinforce to me that to demonstrate normality, I need a P-value greater than 0.05.

```
In[208]:=
```

```
procs4ss = Select[sampleSet, #[[2]] == 4 &][[All, 3]];
procs5ss = Select[sampleSet, #[[2]] == 5 &][[All, 3]];
```

DistributionFitTest[ RandomVariate[NormalDistribution[Mean[procs4ss], StandardDeviation[procs4ss]], 1000]]

Out[210]=

#### 0.442154

Below are two histrograms directly followed by their normality test P-value results. If the P-value is greater than 0.05 we can say the are normal. My data shows, because the P-value is less than 0.05, the data sets are NOT normal.



In both elapsed time data sets above, they failied the normality test. This means to test if there is a real difference in their elapsed times, I can not perform a simple t-test, but a location test.

Comparing the two historgram, it appears the two data sets are clearly different. But we want to be statistically sure. I picked the standard alpha value of 0.05. The null hypothesis is the two data sets are the same. If they are the same, the resulting P-value will be greater than the alpha value, which I choose to be 0.05. Remember above, to show normality (comparing our data with a normal distribution) we wanted the the P-value to be large to indicate the normal distribution and our data are statisticially similar. In this case, we want to show they are statistically different, so we want a small P-value.

And indeed, using my data, there is a statistically significant difference between the two data sets.

