Confessions of a Costaholic:
An insider’s thoughts on software costs – and opportunities

John Baker, Senior Consultant
Agenda

• Sub-Capacity pricing
  • How we got here
  • How it works

• Lowering software costs
  • zIIP usage and crossover
  • Soft Capping
    • Static vs automated
  • IBM pricing options
  • Consumption
    • Lower it or move it – but first understand it!
In the beginning...
Sub-Capacity Pricing

- October 2000: IBM announces Workload License Charges
  - Sub-Capacity Pricing model at less than full machine capacity

- Two dimensions to determine costs:
  - Consumption (CPU/MSU) from SMF 70
  - Active Subsystems (IBM software products) from SMF 89

- IBM uses this report to generate MLC invoice
Dimension #1: Consumption
CPU time

• Accumulated dispatch time on one or more CPUs
• Typically reported in intervals (RMF type 7x)
  – SMF70PDT/EDT
• Also by address space (type 30)
MIPS: Millions of Instructions per Second, per Gary King

• Once upon a time, MIPS really meant Millions of Instructions Per Second

• As commonly used today, MIPS has become a RELATIVE indicator of AVERAGE processor CAPACITY

• MIPS are based on capacity RATIOS between processors

• MIPS are still in the ballpark of real MI/sec
  • Bear in mind that all instructions are not created equal
Service Units

• Accumulated measure of service from a z/OS (logical) perspective
• Per-processor value (see SRM constants)
• Used to normalize workloads across physical processors
• Different types of “service”
  – TCB, SRB, IOC, MSO
• RMF: R723CSRV
MSUs: Millions of Service Units
For Pricing –
NOT for Capacity Planning

• NOT related to WLM Service Units
• Software measurement of machine capacity and consumption
• Scope is CPC and LPAR – **NOT Workload**
• Published values (see LSPR) – along with PCI (MIPS)
• Commonly reported as MSUs per hour over four hour average (R4HA) for software pricing
R4HA or 4HRA

• On each LPAR...
  – Calculated and stored in RMF control block every 5 minutes
  – 48 intervals following IPL (12 per hour * 4 hours) an accurate value is derived (YES – your first four hours are FREE)
  – Continuously updated every five minutes (“rolling”)
  – Oldest interval dropped, newest added, and recalculated average

• RMF: SMF70LAC
Peak vs Peak

- CPU peak
- Billing peak
The SCRT Reporting Period

- From midnight (the very beginning) of the second day of the month up to midnight (the very beginning) of the second day of the next month

- Within each hour, the R4HA is calculated

- Independent of RMF interval (average)

- Peak of this value is the R4HA for an LPAR for billing purposes *

* IF there is no soft cap in place
Dimension #2: Active Subsystems
Sample: WLC Tiers

<table>
<thead>
<tr>
<th>Level</th>
<th>MSU Range</th>
<th>z/OS</th>
<th>DB2</th>
<th>CICS</th>
<th>IMS</th>
<th>MQ</th>
<th>Tier Rate</th>
<th>Accumulated MLC cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>3</td>
<td>4000</td>
<td>5000</td>
<td>6000</td>
<td>10000</td>
<td>2000</td>
<td>27000</td>
<td>$81,000</td>
</tr>
<tr>
<td>0</td>
<td>4 to 45</td>
<td>350</td>
<td>300</td>
<td>300</td>
<td>600</td>
<td>150</td>
<td>1700</td>
<td>$152,400</td>
</tr>
<tr>
<td>1</td>
<td>46 to 175</td>
<td>300</td>
<td>150</td>
<td>150</td>
<td>300</td>
<td>100</td>
<td>1000</td>
<td>$282,400</td>
</tr>
<tr>
<td>2</td>
<td>176 to 315</td>
<td>200</td>
<td>100</td>
<td>115</td>
<td>230</td>
<td>75</td>
<td>720</td>
<td>$383,200</td>
</tr>
<tr>
<td>3</td>
<td>316 to 575</td>
<td>100</td>
<td>80</td>
<td>85</td>
<td>170</td>
<td>50</td>
<td>485</td>
<td>$509,300</td>
</tr>
<tr>
<td>4</td>
<td>576 to 875</td>
<td>90</td>
<td>60</td>
<td>65</td>
<td>130</td>
<td>40</td>
<td>385</td>
<td>$624,800</td>
</tr>
<tr>
<td>5</td>
<td>876 to 1315</td>
<td>60</td>
<td>55</td>
<td>60</td>
<td>120</td>
<td>35</td>
<td>330</td>
<td>$770,000</td>
</tr>
<tr>
<td>6</td>
<td>1316 to 1975</td>
<td>50</td>
<td>50</td>
<td>55</td>
<td>110</td>
<td>25</td>
<td>290</td>
<td>$961,400</td>
</tr>
<tr>
<td>7</td>
<td>1976+</td>
<td>40</td>
<td>50</td>
<td>55</td>
<td>110</td>
<td>25</td>
<td>280</td>
<td></td>
</tr>
</tbody>
</table>

- Sample CPC @ 1000 MSUs: $666,050 per month
  - Accumulated costs of Base to Level 4 plus 125 MSUs @ level 5 rate
- Average Rate: $666.05
- Incremental Rate: $330
Which R4HA should I care about?

- Each hourly R4HA measurement from each LPAR where a subsystem is running is combined to create a peak R4HA by subsystem.

- **It does not matter if the workloads on that LPAR actually use the subsystem(s) or not!**

- Under CMP, all measurements within a geographic country are combined.

- Pre-CMP, measurements are rolled up by CPC/CEC.
Things you can control
Hiperdispatch
First check if this is contributing to a R4HA peak! If so, then consider more zIIP capacity or SMT. If not, perhaps you shouldn’t care.
Soft Caps

- Defined Capacity or Group Capacity
- Single LPAR / multiple LPAR’s per CPC/CEC
- Static MSU value set via HMC
- Interval usage may exceed cap provided R4HA is below limit

- MLC invoice is based on R4HA peak OR cap limit – whichever is LOWER
Risks with static capping

- LPAR is capped
- Cap is impacting workloads
Automated Capping

- Software retrieves key values from WLM control blocks
  - IMSU, R4HA, capping

- Considers installation parameters such as:
  - Importance of Workloads and LPARs
  - Max combined MSU limit (LPAR groups)
  - Calculates lowest target DC’s for each LPAR protecting workloads

- Defined Capacity (DC) values are dynamically modified in real time by taking into account the behavior and needs of all the LPARs
Cap could be temporarily lowered to here with no application impact...
Why it works

- 12 * 5 minute R4HA intervals per hour = X
- 12 * 5 minute average DC per hour = Y
- SCRT uses: **LOWER** value of X OR Y
Don’t start stuff you don’t need

Software Stacks

MQ
IMS
CICS
DB2
z/os

LPAR1
LPAR2
LPAR3
LPAR4

0
5000
10000
15000
20000
25000

1/12/2018 8:00 AM
1/12/2018 9:00 AM
1/12/2018 10:00 AM
1/12/2018 11:00 AM
1/12/2018 12:00 PM
1/13/2018 8:00 AM
1/13/2018 9:00 AM
1/13/2018 10:00 AM
1/13/2018 11:00 AM
1/13/2018 12:00 AM
1/14/2018 8:00 AM
1/14/2018 9:00 AM
1/14/2018 10:00 AM
1/14/2018 11:00 AM
1/14/2018 12:00 AM
1/15/2018 8:00 AM
1/15/2018 9:00 AM
1/15/2018 10:00 AM
1/15/2018 11:00 AM
1/15/2018 12:00 AM

MQ
IMS
CICS
DB2
z/os

IBM ODM Standard
IMS/ESA
MQM MVS/ESA
z/OS
What’s IBM up to?
Recent announcements

• Mobile Workload Pricing (MWP)
• Countrywide Multiplex Pricing (CMP)
• Container pricing
• Solution Consumption License Charge (SCLC)
• ...
Understanding Mobile

Impact of MWP on 4HRA

- Post-MWP basis for your SW bill (‘Adjusted’ 4HRA)
- Pre-MWP basis for your SW bill (‘Real’ 4HRA)

• Chart courtesy of Frank Kyne, http://watsonwalker.com/
**Country Multiplex Pricing (CMP)**

*Country Multiplex License Charges (CMLC) and Multiplex zNALC (MzNALC) Structure*

(cumulative monthly pricing)

<table>
<thead>
<tr>
<th>Level</th>
<th>MSU Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base charge</td>
<td>Up to 3 MSUs</td>
</tr>
<tr>
<td>Level 0</td>
<td>4 - 45 MSUs</td>
</tr>
<tr>
<td>Level 1</td>
<td>46 - 175 MSUs</td>
</tr>
<tr>
<td>Level 2</td>
<td>176 - 315 MSUs</td>
</tr>
<tr>
<td>Level 3</td>
<td>316 - 575 MSUs</td>
</tr>
<tr>
<td>Level 4</td>
<td>576 - 875 MSUs</td>
</tr>
<tr>
<td>Level 5</td>
<td>876 - 1315 MSUs</td>
</tr>
<tr>
<td>Level 6</td>
<td>1316 - 1975 MSUs</td>
</tr>
<tr>
<td>Level 7</td>
<td>1976 - 2499 MSUs</td>
</tr>
<tr>
<td>Level 8</td>
<td>2500 - 3499 MSUs</td>
</tr>
<tr>
<td>Level 9</td>
<td>3500 - 4999 MSUs</td>
</tr>
<tr>
<td>Level 10</td>
<td>5000 - 6999 MSUs</td>
</tr>
<tr>
<td>Level 11</td>
<td>7000 - 9999 MSUs</td>
</tr>
<tr>
<td>Level 12</td>
<td>10000 - 13999 MSUs</td>
</tr>
<tr>
<td>Level 13</td>
<td>14000+ MSUs</td>
</tr>
</tbody>
</table>
CMP pros and cons

- Lowers the incremental cost per MSU on growth
- Allows one large tier for all machines in the same country
- Eliminates need for artificial Sysplexes ("shamplex")

- Price does not go down when you switch
- Primary benefit is from growth – are you growing?
- Initial price is based on existing MLC
  - “The Baseline”
  - Uses average of peaks of 3 months of SCRT data
- One time event

Lower Your Baseline First
Container Pricing

• Interesting solution for Dev/Test workloads
  – May consume 2/3x for the same price
• Similar entry method as CMP:
  – Measure current consumption
  – Baseline

• Bear in mind...
  – Capacity must be available
  – You must ensure other workloads don’t use it
New Application Solution

- Solution Consumption License Charges (SCLC)
- Total (accumulated) consumption vs peak (R4HA)

- New applications only*
- Two options:
  - Pay as you go
  - 20% discount with minimum commitment of 25,000 MSUs

Is IBM changing the game?

- If so, we may need to change the measurements we focus on
  - R4HA and capping may become irrelevant
  - Total consumption becomes key

- Efficiency becomes more important
  - Smart configurations
  - Well-tuned applications
  - We can’t just move/reschedule work
Thank you for listening

John.Baker@intellimagic.com