Session z100832 What is Container Pricing Really All About?

Cheryl Watson cheryl@watsonwalker.com Watson & Walker





## 2018 IBM Systems Technical University

May 2, 2018 Orlando, Florida



#### Please note

IBM's latest Software Pricing initiative, Container Pricing, is still in its infancy. The information in this presentation is our best understanding of how Container Pricing works, based on the available documentation and discussions with IBM. However, as with any decision that has financial repercussions, you should verify your assumptions with IBM before making a final decision.

#### Introduction

Thank you for coming.

#### Who am I?

- President of Watson & Walker Inc. since 1986
- Working on IBM mainframes since 1965
- We publish Cheryl Watson's Tuning Letter (since 1991)
  - Now available to subscribers online at www.watsonwalkerpublications.com
- We teach classes, consult, and have three software products: BoxScore, BoxScore II, GoalTender
- Our latest SCRTPro Service Offering processes SCRT reports and helps to control your IBM Software costs
- z/OS evangelists, Subject Matter Experts in Software pricing, Parallel Sysplex, and Workload Manager.

#### What we are going to talk about today:

• IBM Container Pricing – the big picture, and how to prepare yourself for it.

Feel free to ask questions and make this session as interactive as possible.



For years, customers have grumbled about mainframe software costs.

And for years, IBM has delivered software pricing options to reduce the cost of a selected subset of your workloads:

- Parallel Sysplex Aggregation (PSLC and ULC)
- Sub-Capacity Pricing
- GSSP (Getting Started with Sub-Capacity Pricing)
- IWP (Integrated Workload Pricing)
- zNALC (z Systems New Application License Charge)
- MWP (Mobile Workload Pricing)
- zCAP (z Systems Collocated Application Pricing
- zWPC (z Systems Workload Pricing for Cloud)
- CMP (Country Multiplex Pricing)
- Many others, including zAAPs and zIIPs



Despite all the complaining about prices, many customers still don't use the available pricing options

- Don't understand, or even know about, all the pricing options.
- The available options are often complex, making it hard to select the *best* option for your environment.
- Don't like the hassle of having to create dedicated LPARs. (e.g. zNALC).
- Too hard to identify which transactions/workloads apply to which pricing option required information might not even be available in the SMF records. (MWP).
- Too much work on the system to churn through all the SMF records to extract the required information.
- Whole manual process of gathering and formatting the CPU consumption data seems like more hassle than it is worth.

Container Pricing was previewed as part of the z14 announcement. It is IBM's attempt to:

- Keep Dev/Test on the mainframe where it is eventually executed in production.
- Enable you to add new z/OS workloads without increasing the MLC cost of existing workloads.
- Address customer complaints that it is difficult, if not impossible, to accurately
  predict the software cost for proposed new applications "It depends" doesn't
  cut it.
- Provide a "cloud-like" software pricing model for customers that desire that.
- Break the connection between how much you pay for software and how you configure your systems.
  - Objective is to let you configure systems and applications in whatever way makes the most *technical* sense.
- Greatly simplify the whole process of identifying qualifying work, gathering and tracking resource consumption, getting that info into SCRT, and sending it back to IBM.
- Enable software bills to be based on metrics other than peak R4HA CPU consumption.

Some very important points before we get into details about Container Pricing:

- Container Pricing has NOTHING to do with Docker Containers!
- Most existing pricing options deliver savings by reducing the number of MSUs that your bill is based on.
- Two of the three currently-announced Container Pricing "Solutions" involve fixed costs for the Solution.
- In some scenarios, the Container Pricing Solution will be the most cost effective (or the most attractive for some other reason). In other scenarios, one of the existing pricing options might be the most cost-effective.
- IBM has not withdrawn the existing pricing options Container Pricing gives you a *additional* choices.
- In order to select the best option, you *must* understand the basics of software pricing and the Rolling 4 Hour Average.

THIS is the focus for this morning's session

There are **two** parts to Container Pricing:

- Set of infrastructure enhancements that will enable simpler and far more flexible software pricing on z:
  - The enhancements are intended to let you add new workloads to a z/OS environment with "no" impact on the MLC cost of the existing z/OS applications.
  - Over the longer term, the enhancements will enable software billing based on many different metrics, not just peak R4HA.
- New pricing options, called 'Solutions', that exploit the infrastructure.
  - Three Solutions are available since December 2017.
  - We will provide an overview here, but go into much more detail in Session z100833 this afternoon (3:15 pm in this room).

At the time of writing, IBM has announced three Container Pricing based Solutions:

- Application Dev/Test Solution (217-490)
- New Application Solution (217-519)
- Payments Processing Solution (217-518)

There are also rumors that more Solutions are on the way.

9

## Container Pricing

- ALL of these solutions differ from previous software pricing options because they rely on the Container Pricing enhancements to deduct the R4HA of the container from the CPC's or Multiplex's R4HA.
- The other thing they have in common is that you must submit a request to IBM, describing the workload that you will place in the container. If IBM approves the request, you must retrieve the Solution ID from LMS and use it in the Tenant Resource Group (TRG) definition (for collocated containers) or in the SCRT control statements (for dedicated LPAR containers).
- Other than these two aspects, each solution offering is quite different from the others...

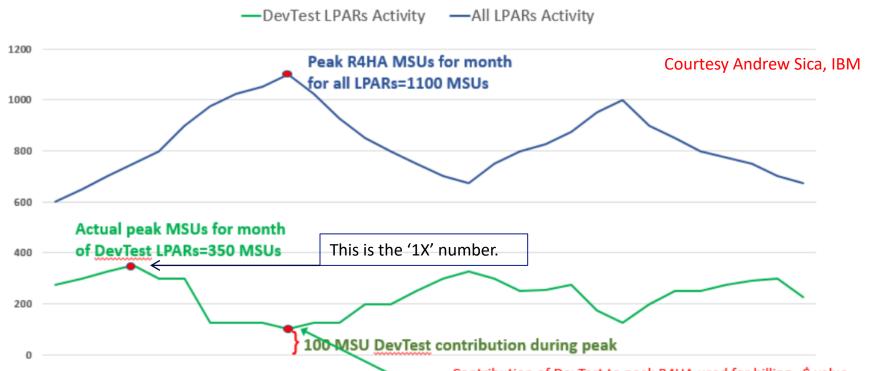
DevTest Solution  At the 50,000 foot level, the DevTest solution lets you grow your DevTest workload to up to 3x current size with no increase in your MLC costs (assuming nothing else changes).

- The <u>cost</u> of your DevTest workload:
  - Gives you a <u>fixed MLC price</u> for the container that you use for your Dev and Test workloads.
    - The workloads can be in dedicated LPARs or in shared LPARs.
  - The fixed price is based on the portion of your bill for each MLC product that is attributable to your DevTest workload, averaged over the last 3 months.

- The <u>cost</u> of your other workloads:
  - When determining the R4HA for your other workloads (Production), SCRT deducts the R4HA for the DevTest container from the total R4HA.
  - NOTE: The MSU reduction is based on the R4HA of the DevTest container for each corresponding hour. It is not based on the R4HA from the 3-month base period, nor on the peak R4HA for the container.
- The bill for the production part of your workload is based on the adjusted R4HA, after deducting the DevTest R4HA. Specifically, the fixed price you pay for your DevTest container has nothing to do with the bill for the production workload.

- The size of your DevTest container:
  - The average of the peak R4HA for the DevTest workload for the 3-month measurement period is deemed to be your base. This is likely to a higher number than the value that is used to set your fixed price.
  - You can grow the DevTest container up to 3X that size for that same fixed price.
  - You don't have to go for 3X you can select any maximum size between 1X and 3X.
  - If the DevTest container will use any IPLA software, you must purchase sufficient Value Units on day 1 to cover the maximum container size.

#### See later session for update!



In this example, actual peak for <u>DevTest</u> is seen as 350 MSUs so up to 1050 MSUs permitted at no extra MLC cost (subject to HW MIPs and <u>zIPLA</u> SW coverage being available)

Contribution of <u>DevTest</u> to peak R4HA used for billing. \$ value calculated by excluding from SCRT run and calculating delta versus BAU total \$ cost based on announcement letter terms – this becomes the MLC \$ cost for <u>DevTest</u> workload going forward.

Copyright 2018 Watson & Walker

- Based on customer feedback, the DevTest Solution is the one that is attracting the most attention at the moment.
- It is easy to implement, especially if you have dedicated Development and Test LPARs.
- Many installations cap their development LPARs during peak hours, to limit their impact on SW bills. This impacts programmer productivity and reflects poorly on z/OS. This Solution potentially lets you significantly increase the capacity available to Development and Test LPARs without increasing your MLC costs.
- The devil is always in the details, and like all SW pricing options, this one requires careful evaluation to see if it is appropriate for *your* shop.
- See this afternoon's session z100832 at 3:15 pm for more detail.

## New Application Solution:

- This is intended to protect you from the CPU consumption of a new application increasing the costs of your traditional applications.
- Also it provides a fixed price estimate earlier in the process.

# New Application Solution

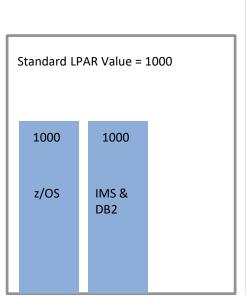
- The next most popular Solution is the New Application One.
- Like the other Container Pricing Based solutions, this one is intended to protect you from the CPU consumption of a new application increasing the costs of your traditional applications.
  - The R4HA of the container is deducted from the total R4HA, with the intent that the cost of the non-container workload would be roughly the same as it would be if the container workload didn't exist.
- It also addresses the requests of customers for 'line-of-sight pricing' in other words, they want predictable costs and the ability to know the cost in advance.

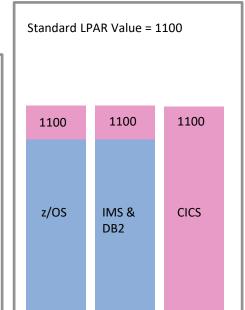
### New Application Solution

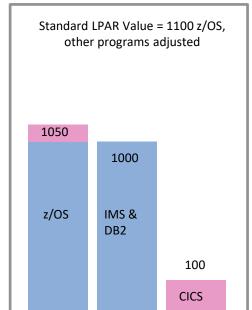
- This solution also supports both collocated applications and applications running in a dedicated LPAR.
- The fixed cost is based on information you provide about the expected size of the container and the set of products that the application will use.
  - z/OS is priced at a 50% discount on the MzNALC price.
  - All other products used by the application must be IPLA, and Value Units for the full container size must be purchased on Day 1.

#### What is Collocated?

Example of zCAP or container pricing application that is the first user of CICS in an IMS/DB2 shop (assumes all products have peak R4HA at same time).







© Watson & Walker 2017

Payments
Processing
Solution

- This solution is aimed at a very niche market, so we won't get into all the details.
- However, the model that it uses is very interesting as a possible indicator of what IBM has in mind for future offerings.

# Payments Processing Solution:

- Unlike the TestDev or New Application Solutions, this one does not have a fixed price for the container.
- Instead, the cost of the container is based on the number of payments that are processed by Financial Transaction Manager for z/OS, the heart of the offering.
- This provides a direct correlation between the price the customer pays for the container and the business value delivered by the product – if FTM processes more payments, that means more business value and therefore a higher bill. Fewer payments would mean a smaller bill.

# Payments Processing Solution

- This is interesting from a number of perspectives:
- It uses the container pricing infrastructure to measure the R4HA of the container so that the R4HA of all other work can be adjusted accordingly.
- It is supporting two metrics one metric (# of payments) determines the cost of the container, and a quite different metric (R4HA) is used to ensure that the cost of other workloads is not affected by the presence of the container workload.
- In the future, this concept could be extended to things like transaction counts, TBs read, total CPU time consumed, jobs run, just about anything that can be measured programmatically and audited could be used as the basis for the container bill.

Now back to the infrastructure. There are a number of objectives:

- Get the system to associate work with a pricing option, gather information about the resource consumption of the work and save it to SMF, pass that to SCRT, and have SCRT deduct that capacity from the total R4HA.
- Make it easier to define the qualifying workload to WLM.
- Make it easier to associate a given workload with a particular pricing option.
- Give customers the option to place the application in its own LPAR, *or* in shared LPARs, *or* in both.
- Allow the customer to limit the resource consumption of the new workload.

## How does it achieve these wondrous things?

- IT DEPENDS! (Remember that more flexibility usually results in more complexity).
- If the workload is running in a dedicated LPAR, you simply associate the Solution ID with that LPAR using SCRT control statements.
- IF the workload is collocated with other work (shared LPAR), then you use the new z/OS infrastructure enhancements that have been added for Container Pricing support.
- **NOTE**: Use of the new WLM constructs is ONLY required for collocated applications.

### How does it achieve these wondrous things?

- All the definition, tracking, and gathering for collocated applications is consolidated into WLM – this is becoming the focal point for software pricing controls as well as performance controls.
- Objective is that you define the work once in WLM, and everything after that (all the way through to sending the information to IBM) happens with minimal intervention.
- To provide this, WLM (<u>OA52312</u>), SMF (<u>OA53033</u>), RMF (<u>OA52694</u>), SDSF (<u>PI82528</u>), SCRT 25.2 (now a component of z/OS 2.3), and z/OSMF (<u>PI89361</u> & <u>PI89935</u>) have all been updated to support Container Pricing.

## Tenant Resource Groups (TRGs)

- Tenant Resource Groups (TRG) (think traditional WLM Resource Groups with a few new frills). These are used to track resource usage by qualifying work.
- Even though we are used to using Resource Groups as a way to limit the capacity used by some workload, it is NOT necessary to cap a TRG. It is just a mechanism for tracking the resource consumption of all the work in that TRG.
- When you define a TRG to WLM, you MUST specify a 'Solution ID' that you get from IBM if you want to test this capability or if you have signed a Solution contract. The Solution ID is uniquely associated with a specific Solution in your company. The Solution ID is like a tag that can be used to identify which Solution used how much CPU time.

## Tenant Resource Groups (TRGs)

- While you can have multiple TRGs associated with a single Solution, this is not necessary unless you want to use capping and don't want to cap all parts of the Solution.
- The SMF type 70 record has new Tenant Resource Group data sections, 1 per TRG – this information comes from a new IWM4QTNT WLM interface. The new section contains most of the info from the TRG definition, plus SUs for zAAP, zIIP, and GCPs, plus the GCP R4HA MSUs for that TRG.
- RMF PP Reporting on TRGs is based on the existing WLM Resource Groups reporting. The RMF Overview reports have been enhanced to support the new Type 70 fields.

```
<u>Ienant-Resource-Group Notes Options Help</u>
                   Create a Tenant Resource Group
Command ===>
Enter or change the following information:
Tenant Resource Group Name <u>IESTIRG</u> (required)
Description . . . . . . . <u>Test TRG</u>
                                                      Optional
Tenant ID . . . . . . . .
   D9F4555-114E99A-2BE9DC541E-4D5664C9-409A-4AA8-9D47-7B017E-10A2DF
                                                               Optional
Define Capacity: 4 1. In Service Units (Sysplex Scope)
                   2. As Percentage of the LPAR share (System Scope)
                   3. As a Number of CPs times 100 (System Scope)
                   4. In accounted workload MSU (Sysplex Scope)
Include Specialty Processor Consumption NO (YES or NO)
     Selection List empty. Define a Tenant Resource Group. (IWMAM530)
```

## What is a Container?

- This is purely a pricing construct. It is the set of TRGs and dedicated LPARs that have the same Solution ID.
- It is used mainly in SCRT:
  - For dedicated LPARs, to assign the Solution ID associated with that LPAR.
  - For all Solutions, to assign a name that is a little more user-friendly than the 64-byte Solution ID.
  - In SCRT Report sections that are related to Solutions (SCRT uses the term 'Containers' rather than 'Solutions').

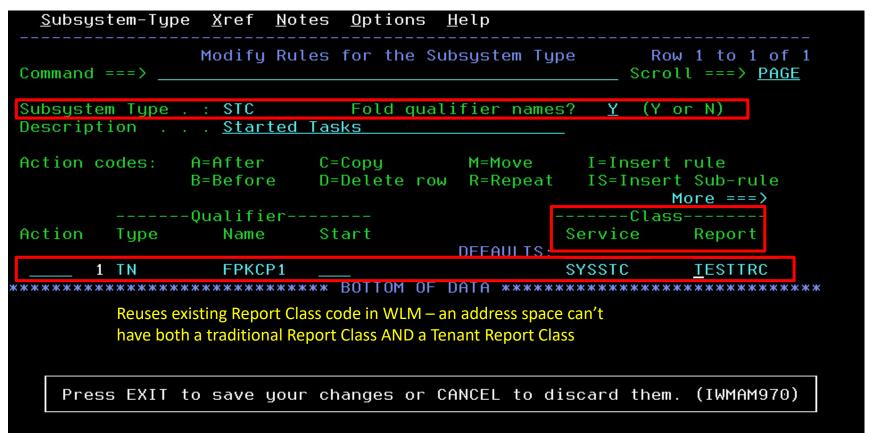
## What is a Container?

- The container is the z/OS representation of the Solution. The combined R4HA of the TRGs AND the dedicated LPAR(s) (where appropriate) in the Container represents the CPU consumption of the workloads in the Solution.
- The Solution will have an agreed size (specified in the contract with IBM). If the Container peak R4HA exceeds that size, additional charges will be applied.
- You can define a cap for each TRG, but there is no way to specify a cap at the container level (assuming that there are multiple TRGs in the container).

## Tenant Report Classes (TRC)

- Traditionally, work is assigned to a WLM Resource Group indirectly – the service class is assigned to the RG, and work is assigned to the Service Class.
- This would not work for Solutions because single address spaces (DB2DBM1, for example) can be shared between your traditional workloads AND by work in the Solution. If you assigned the Service Class to the TRG, you would pick up more work than is qualified.
- To get around this, IBM created Tenant Report Classes (TRCs).
   A TRC is associated with one, and only one, TRG.
- Work is assigned to the TRC using the WLM classification rules.

```
<u>Ienant-Report-Class</u> <u>N</u>otes <u>O</u>ptions <u>H</u>elp
                         Create a Tenant Report Class
Command ===>
Enter or change the following information:
Tenant Report Class Name . . . <u>IESTIRC</u> (Required)
Description . . . . . . . <u>Tenant Report Class for testing</u>
Tenant Resource Group Name . . <u>IESTIRG</u> (Required; name or ?)
        Tenant Report Class MUST be assigned to a TRG
    Press EXIT to save your changes or CANCEL to discard them. (IWMAM970)
```



#### **Container Pricing Considerations**

There is generally no need to create new WLM service classes for the Solution workload

- Remember the RoT to have not > 30 active service class periods in a system.
- We were concerned that potentially having a subset of the work in a service class capped (if you cap the associated TRG) could cause a problem. To avoid potential performance issues, we recommend splitting off a new service class IF you have work that you will assign to a heavily capped TRG.

#### **Container Pricing Considerations**

There is generally no need to create new WLM service classes for the Solution workload

- Service classes that are assigned to a traditional WLM Resource Group must not contain work that is assigned to a TRC.
- You can specify an upper limit (cap) for a TRG, but not a minimum limit.
- IBM recommends that you avoid capping a TRG unless it is really necessary.
- At the time of writing, specifying a memory limit for TRGs is not supported. We expect this to be addressed soon.

# Considerations for TRCs

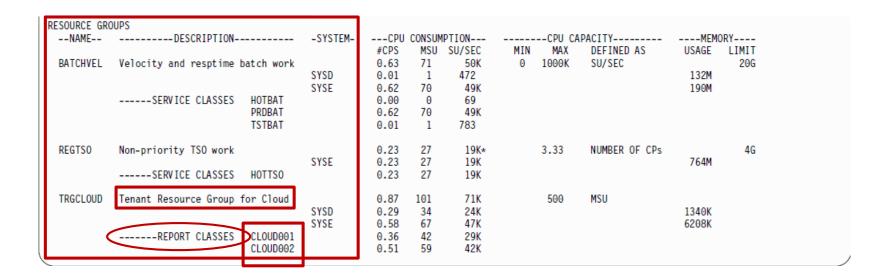
- Work can only be assigned to a TRC at the address space or independent enclave level.
  - This means that you cannot use shared CICS regions with Solutions.
  - On the other hand, if it is worth going to the trouble to get the application qualified, then it is probably large enough to justify its own region(s).
  - This has the added benefit that all of the CICS processing, including the 'overhead' associated with those transactions, is included in the container.

# Considerations for TRCs

- The following WLM subsystem types can be assigned to a TRC: JES, STC, CB, NETV, LDAP, TCP.
- IBM recommends having a different TRC for each service class that has work in the container.
- Never ever classify heterogeneous work to the same TRC. If you try to do this, WLM will issue a warning message, but you can ignore that and proceed – DON'T!

#### RMF considerations

 TRGs and RGs are reported in the same RMF report, so use Descriptions that make it easier to identify the TRGs...



#### **SCRT Considerations:**

- Now that you have everything set up, the information in saved in the new TRG sections of the SMF 70.1 and 89 records, and then processed by SCRT.
- SCRT contains a number of new reports specifically in support of Container Pricing

Q7					
Product Name	Product ID	SYS1 TGCICS21	SYS2 TGCICS21	SYS4 TGCICS21	SYS5 TGCICS21
z/OS V2 CICS TS for z/OS V5	5650-ZOS 5655-Y04	0.60% 0.60%	0.60% 0.60%	0.60% 0.60%	0.60% 0.60%

• New Q7 section reports on which products were used in which TRGs. This can be used to help you verify that your understanding of which products are being used in each TRG is accurate.

#### **SCRT Considerations:**

- The T4 report shows the Container Max Contributors for the container shown on the last line CPS1 in this example.
- This report shows that TRG TGCICS21 was active in 4 LPARs, however it only consumed enough capacity in the SYS1 LPAR to make a contribution to the LPARs R4HA.

CONTAIN	ER MAX CONTRIBUT	TORS		
	TRG	Highest	Date/Time	Contribution to Highest
SYS2	TGCICS21			0
SYS4	TGCICS21			0
SYS1	TGCICS21			14
SYS5	TGCICS21			0
CPS1		14	25 Sep 2017 - 15:00	

#### **SCRT Considerations:**

 The final report shows the total MSUs consumed by each TRG in each LPAR over the reporting period. Note that this is the total, not the peak. None of the currentlyannounced Solutions charge on total MSU consumption.

	MSU CONSUMPTION	N CONTRIBUTORS
CPS1	Z194E15-nnnnn	nn-nnnnnnnnn-nnnnnnn-nnnn-nnnn-nnnn-nnnn
	TRG	Contribution to Total Consumed
LPAR1 LPAR1 LPAR1 LPAR2 LPAR2 LPAR3	TRG1 TRG2 TRG3 TRG4 TRG5	26712 48081 16027 29383 48081 16027
CPS1		184311

## Resource Group capping

- If you want to cap the Solution, remember that the Container size is based on the Rolling 4-Hour Average, but the Resource Group caps (both traditional Resource Groups and Tenant Resource Groups) are based on rolling 60-second intervals.
  - Even the new type 4 MSU caps are based on these 60-second intervals.

## Resource Group capping

 Effectively, Resource Group caps behave in a similar way to absolute caps. If you set the Resource Group cap to the same value as the Container size, it is unlikely that the Container R4HA will ever reach the agreed Container size, unless it is CPU bound for a long time.

# Effects of Collocation

- Even though Container Pricing allows both dedicated and collocated options, there are some things to remember about the collocated option:
  - Adding work to an LPAR will generate more work for system components – for example, Master, SMF, XCF, GRS, Catalog, batch scheduler, JES, etc. It is not possible to apportion out the system work caused by the Container workload back to the TRGs.
  - Similarly, while DB2 charges back most of its CPU time to requesters of its services, about 20% remains in DB2. If the DB2 is shared between a new Solution and traditional applications, the subsystem cannot be assigned to a TRG..

# Effects of Collocation

- Therefore, adding collocated work to an LPAR will result in some increase in MSUs for the traditional part of the system.
- Additionally, IBM's guidance is that every 10% increase in physical CPC utilization results in an increase of between 3% and 5% in the CPU consumption of work running in that system. This applies regardless of whether the new workload is collocated or in its own LPAR it is a result of running more work on the same CPC.

## Summary

- You CAN get started with Container Pricing without making any changes in WLM if the workload will run in dedicated LPARs.
- In parallel with that, it would be a good idea to set up some 'dummy' TRGs, just to get experience with this new capability.
- IBM has provided a set of sample Solution IDs that can be used for testing. For more information about the sample Solution IDs, see <a href="https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ZSL03543USEN&">https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ZSL03543USEN&</a>
- Make sure that you use the IGNORE CONTAINER statement to exclude the impact of test Solution IDs before sending the SCRT file to IBM

## Summary

- IBM is putting a lot of time and money into this initiative this is not just yet another pricing option.
- This is a little like z/OSMF it will grow in function and pervasiveness over time, so don't ignore it.
- The infrastructure changes are long overdue, but will enable many different pricing models in the future and also make it easier to exploit the available options.
- IBM has not yet removed any of the previously-available pricing options, so you currently have fixed price and variable price options one of them can very likely save you money.

## Summary

- It is not possible to make a blanket statement about which is the best option – it really does depend on many variables specific to your environment and software stack, so you need to evaluate the options using *your* numbers and projections for the future.
- Attend our z10833 session at 3:15 pm today, here in this room for more information about these options and about Country MultiPlex Pricing.

For more information about Container Pricing, refer to:

- Announcement letters:
- Container Pricing preview <u>117-044</u>.
- Application Development and Test Solution <u>217-490</u>.
- New Application Solution <u>217-519</u>.
- Payments Processing Solution <u>217-518</u>.
- Container Pricing White Paper <u>WP102719</u>.
- List of sample Solution IDs
- SCRT User's Guide
- SHARE in Sacramento Session <u>22548</u>, Container Pricing Overview and Sub-Capacity Reporting, by **Andrew Sica**.

## Summary



We think that about 90% of our clients will be using containers at some point in their future. You can learn it now or learn it later.

# Please complete the session survey!

Z100832

