



Rethinking IBM Software Cost Management Under Tailored Fit Pricing

Scott Chapman

Enterprise Performance Strategies, Inc.

Scott.chapman@EPStrategies.com



Contact, Copyright, and Trademarks



Questions?

Send email to performance.questions@EPStrategies.com, or al@Sherkow.com or visit our websites at <https://www.epstrategies.com> or <http://www.sherkow.com>.

Copyright Notice:

© Enterprise Performance Strategies, Inc. and I/S Management Strategies. All rights reserved. No part of this material may be reproduced, distributed, stored in a retrieval system, transmitted, displayed, published or broadcast in any form or by any means, electronic, mechanical, photocopy, recording, or otherwise, without the prior written permission of Enterprise Performance Strategies and I/S Management Strategies. To obtain written permission please contact Enterprise Performance Strategies, Inc. and I/S Management Strategies. Contact information can be obtained by visiting <https://www.epstrategies.com> and <http://www.sherkow.com>.

Trademarks:

Enterprise Performance Strategies, Inc. presentation materials contain trademarks and registered trademarks of several companies.

The following are trademarks of Enterprise Performance Strategies, Inc.: **Health Check[®]**, **Reductions[®]**, **Pivotor[®]**

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries: IBM[®], z/OS[®], zSeries[®], WebSphere[®], CICS[®], DB2[®], S390[®], WebSphere Application Server[®], and many others.

Other trademarks and registered trademarks may exist in this presentation

EPS: We do z/OS performance...



- Pivotor - Reporting and analysis software and services
 - Not just reporting, but analysis based reporting based on our expertise
- Education and instruction
 - We have taught our z/OS performance workshops all over the world
- Consulting
 - Performance war rooms: concentrated, highly productive group discussions and analysis
- Information
 - We present around the world and participate in online forums

Abstract



Are you considering moving from the rolling-4-hour average (R4HA) MLC pricing to the consumption model of Tailored Fit Pricing (TFP)? Or maybe you have moved from R4HA to TFP already and are now wondering about how to better manage your bill. Or maybe you'd just like an update on what TFP is and if you should consider it. In any of those cases, come to this webinar wherein ***Scott Chapman*** will review TFP and what we've been hearing from IBM and customers about it. He'll also explain how you must change your thinking about managing workload for optimal software cost management when moving from R4HA to TFP.

This presentation is about TFP not TPF, nor FTP, nor PTFs! 🤖

Agenda

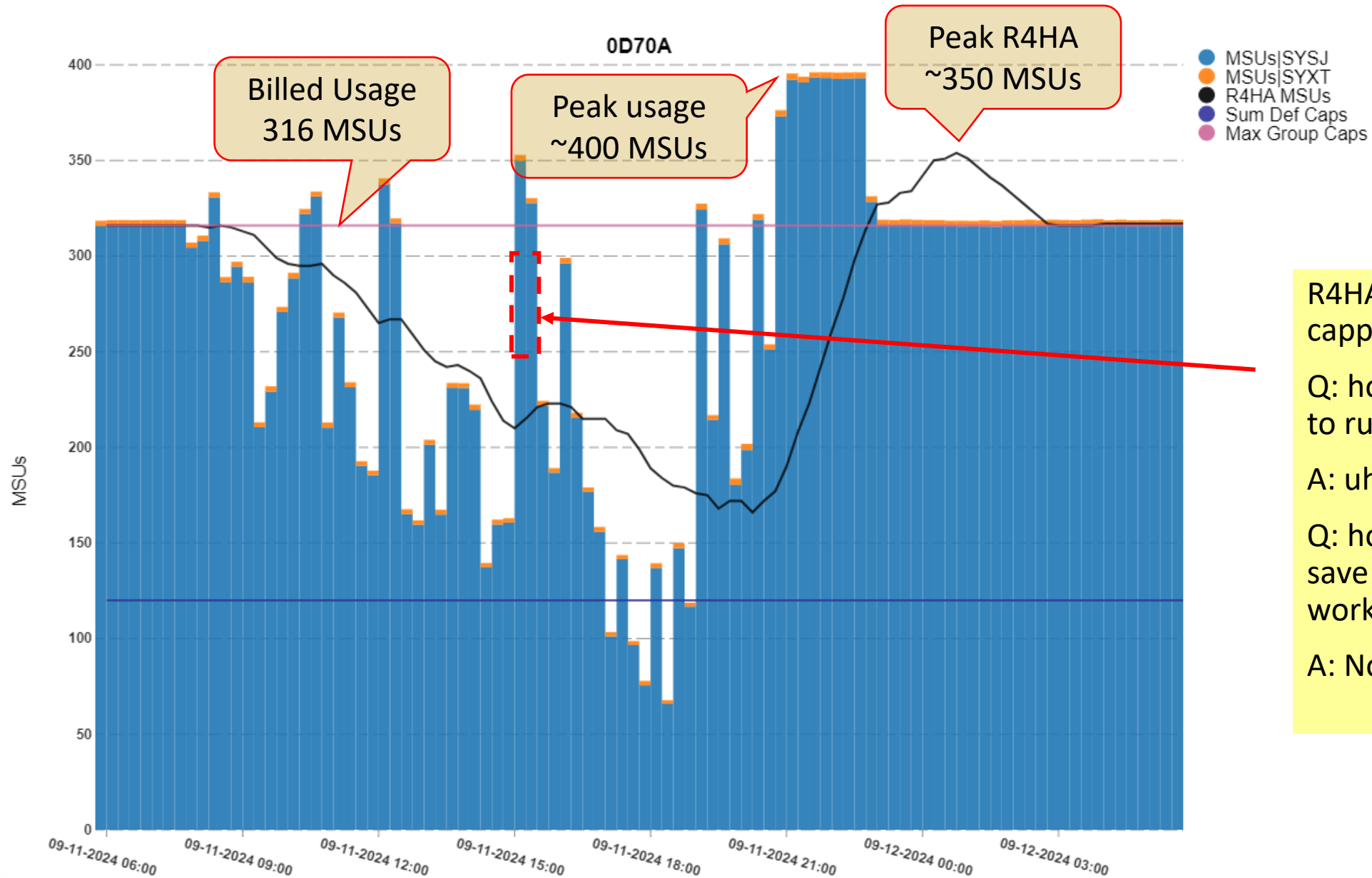


- Why are we revisiting TFP now
- Review of what Tailored Fit Pricing (TFP) is relative to R4HA
- How does TFP vs R4HA affect both performance and cost management
- What should you (maybe) do?
- If you're going to TFP what do you need to negotiate into your contract?

- Not on the agenda: a whole lot of specific details because they're
 - Mostly immaterial to the larger questions
 - Deal with very specific situations

- Sub capacity pricing separated hardware and software growth
 - Because software upgrade costs were inhibiting hardware upgrades
 - Starting with Variable Workload License Charges (VWLC) in 2000
 - Based on peak rolling 4 hour average (R4HA) utilization for the “month”
- Customers started actively managing R4HA to try to reduce their bill
 - Using both IBM-provided capabilities and ISV software products
- To make sub-capacity work better IBM tweaked it over the years
 - GSSP, Solution Edition, IWP, SALC, MWLC, zWPC, zCAP, CMLC, SCLC
 - “work better” = lower costs of new workloads or workloads at risk of leaving
- But understanding the IBM software costs got increasingly complicated
 - Could quite easily have multiple acronyms in use at the same time
 - R4HA metric is divorced from overall capacity utilization and is hard to explain

LPAR Interval MSUs and CEC R4HA MSUs



R4HA example with capping in place.

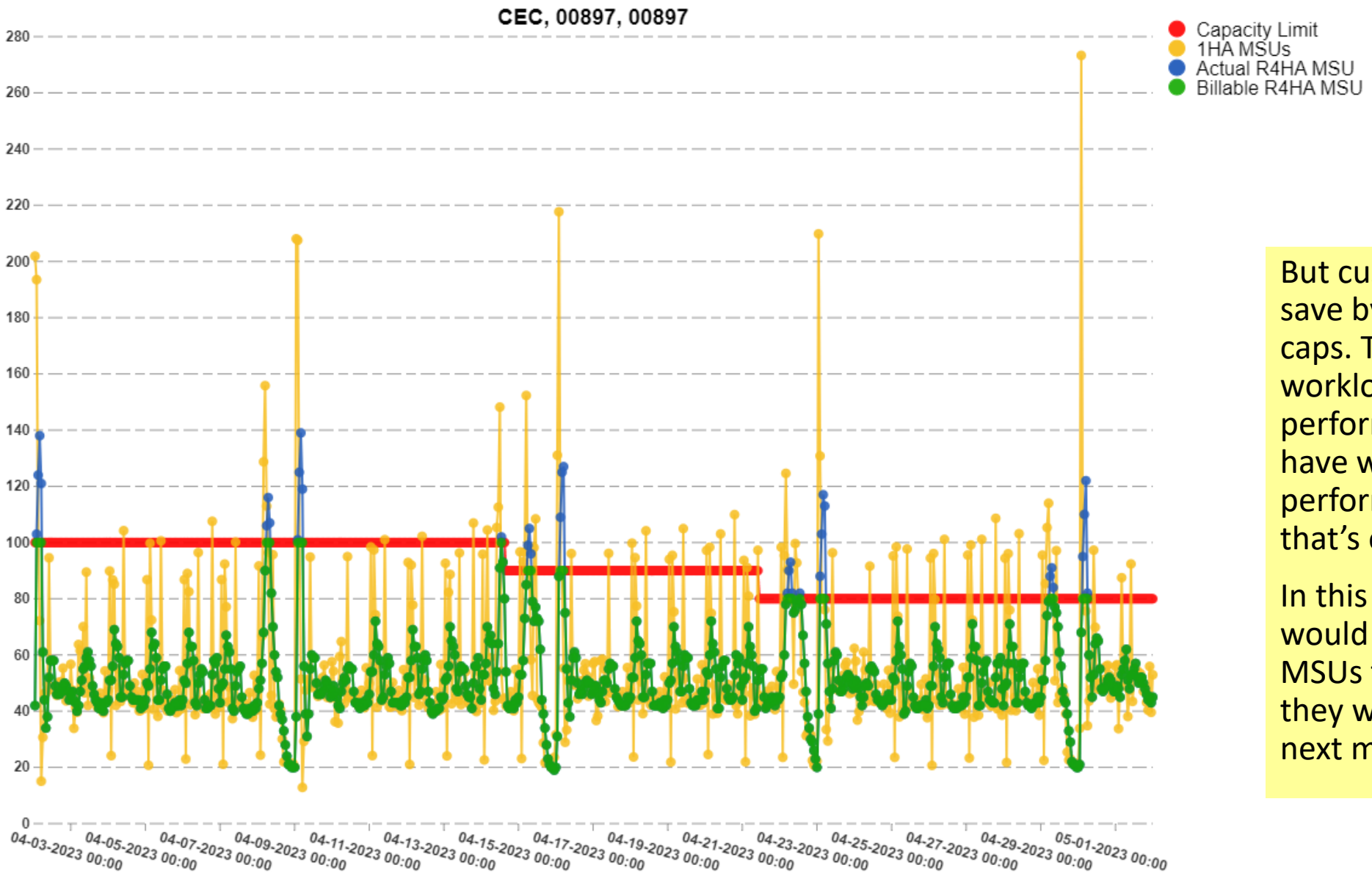
Q: how much did it cost to run this?

A: uhhh...

Q: how much will we save if we eliminate that work?

A: Nothing

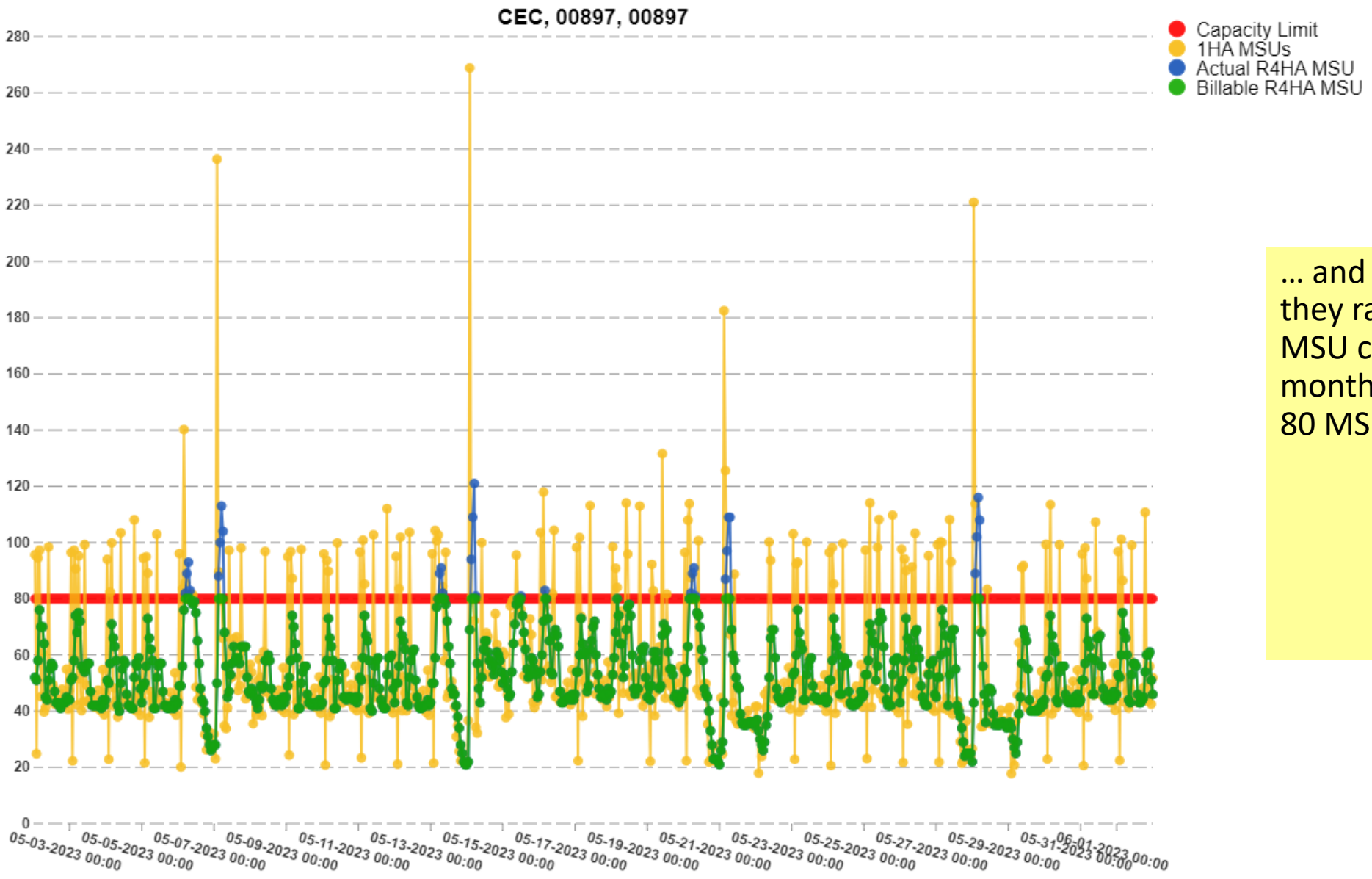
MSU Averages Comparisons



But customers could save by lowering their caps. This could restrict workload and impact performance. But if you have work that's over-performing, then maybe that's ok.

In this example they would have paid for 100 MSUs this month, but they were setting up for next month...

MSU Averages Comparisons



... and in the next month they ran with the 80 MSU cap the whole month and paid for just 80 MSUs.

After 19 years: Tailored Fit Pricing



- See announcement 5/14/2019:

http://www-01.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_ca/4/897/ENUS219-014/index.html

- New optional pricing method that does away with R4HA

- Enterprise Capacity Solution

- Charge is based on total installed capacity

Very, very few customers have a use case for this!

- Enterprise Consumption Solution

- Charge is based on total annual consumed MSU-hours
- “MSU-hours” abbreviated “MSUs” by IBM (to confuse us?)
 - I prefer to say “MSU-hours” to avoid confusion:
 - Under R4HA your incremental per MSU cost could be \$500
 - Under TFP your incremental per MSU cost could be \$0.85

Transitioning to TFP

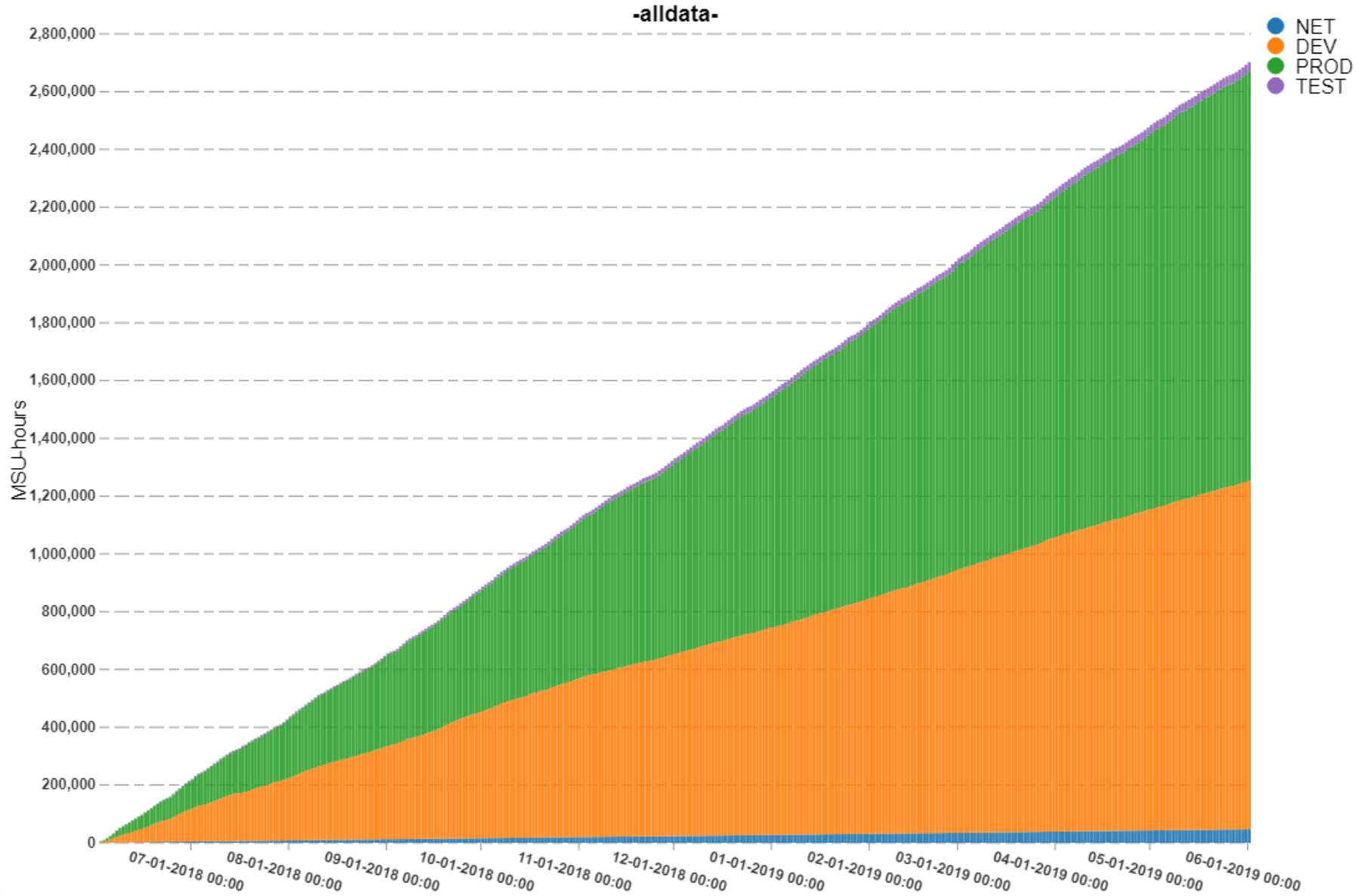


- TFP is not designed to reduce your existing spend for your existing work
- TFP is designed to make *growth* less expensive
- Charges based on past 12 months + agreed-upon reduced rate for growth
 - Baseline MSU-hours = total MSU-hours over the last 12 months before going to TFP
 - Also is a “committed” amount—you’re (maybe?) agreeing to pay IBM for at least that much
 - Baseline Price per MSU-hour = last 12 months charges / baseline MSU-hours
 - Growth Price per MSU-hour = baseline price less some discount (at *least* 50%)
 - That discount is one of the *many* negotiation points (ask for 70-75% to start?)
- If over the next year you use \leq baseline MSU-hours
 - You pay IBM what you paid them the prior 12 months
- If you use more MSU-hours
 - You pay IBM more, but at the Growth Price/MSU-hour for the usage above the baseline

Agreements individually negotiated!

LPAR MSU-hour Cumulative Totals

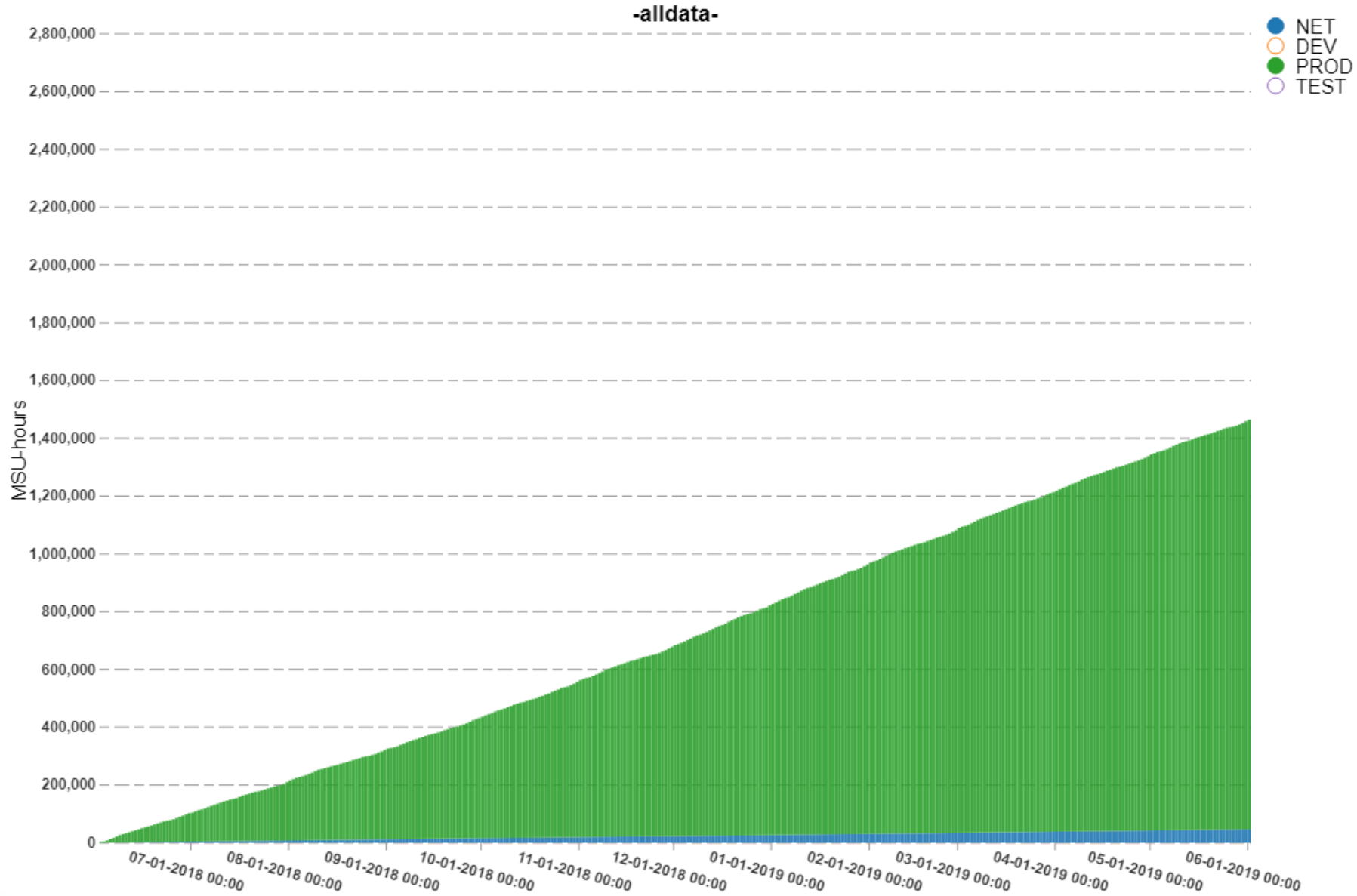
2018-06-02 - 2019-06-01



With TFP, all utilization contributes to the software cost.

LPAR MSU-hour Cumulative Totals

2018-06-02 - 2019-06-01



Dev/Test solution may be part of a TFP contract, so annual total will be driven by just production totals.

Dev/Test container



- Old opinion: Dev/test container should be part of every TFP deal
- New opinion: “probably should be”
 - Must pay attention to details: be wary of purchasing an overly large dev/test container (whether or not you’re going to TFP) as that can trigger zOTC charges on upgrades
 - If going to TFP, negotiate a single blended rate for production and dev/test with dev/test given a *very* high growth rate discount
- Dev/test containers are just one of the complications with TFP, but remember: **every deal is individually negotiated**
 - Fundamentally different than the “chose the correct license” WLC model

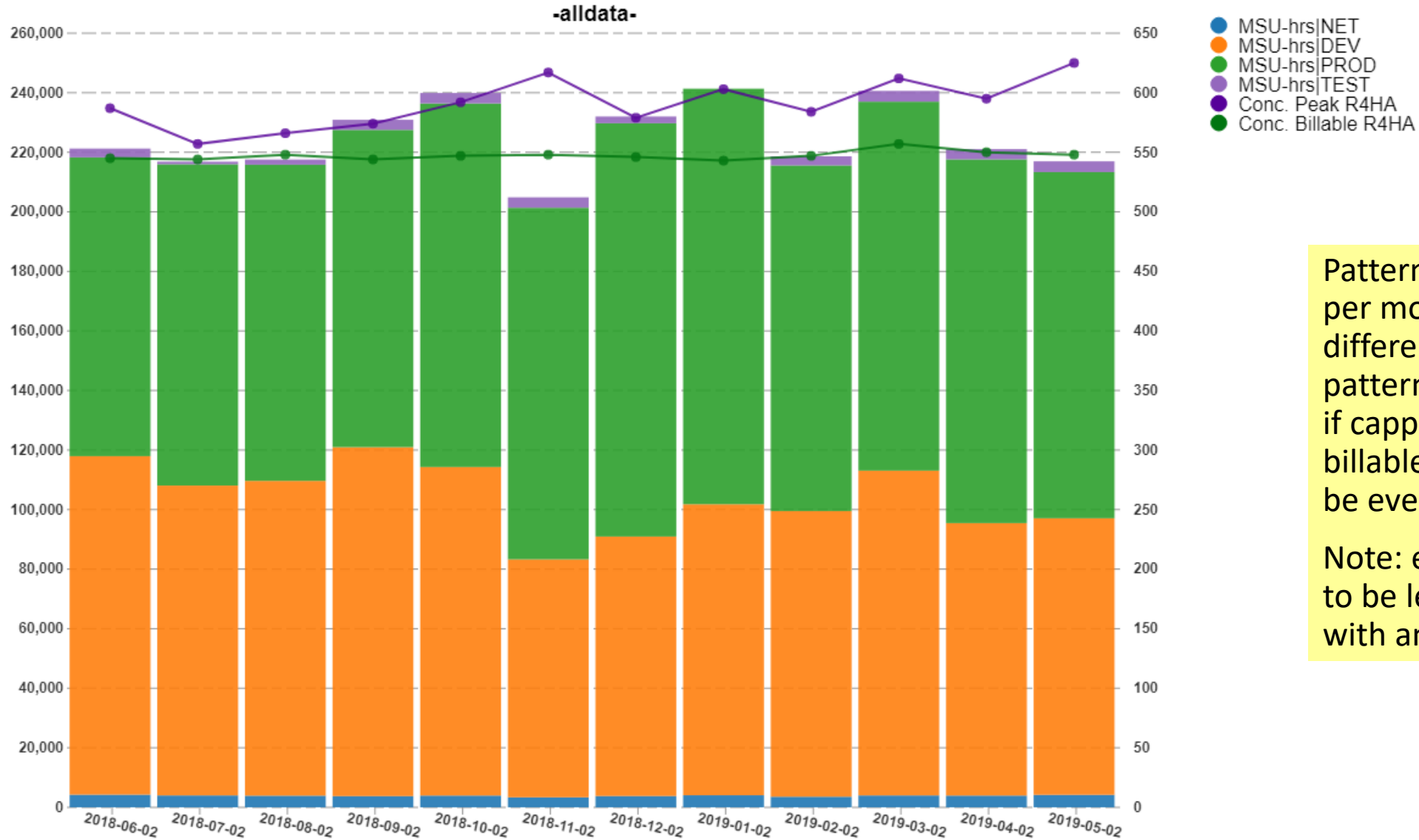
zOTC (aka IPLA)



- Converting zOTC to TFP involves an even more complicated formula
 - I've not tried to figure it out: make IBM explain it to you
- You do probably want to convert zOTC to TFP if you're doing TFP for MLC
 - Otherwise, zOTC will stay under R4HA, which means you'll be trying to manage to two completely different metrics !
 - Would need to run SCRT for both TFP and R4HA



MSU-hour Totals by MLC Month



Pattern of total usage per month is likely much different than the pattern of the R4HA and if capping is in place, billable R4HA usage may be even more different.

Note: expect TFP billing to be level each month with annual true-up.

Reasons to go to TFP



- Potential to reduce the cost of growth
- Simplification, especially if you have multiple WLC licenses in place
- Run any product anywhere: no more separating products by LPARs
 - Still do so where there's a good technical reason, but avoid being forced to to save \$
- Eliminate need for capping (at least for IBM software)
- Easier to correlate application consumption with a cost
- Eliminates 12-month version upgrade windows
- **Individually negotiated contracts**: maybe you negotiate well!

Reasons to avoid TFP



- You're *actively* shrinking your consumption
 - Easier to reduce costs under WLC, assuming you can reduce the peak R4HA
 - Not actively shrinking still has a potential path to TFP with a reduced baseline
- Cost management now involves all consumption instead of just the R4HA
 - Total CPU consumption should be more closely monitored
- Have potential for a significant bill at annual true-up that has to be planned for
- You're outsourcing your machine
 - Owner of the machine generally has to pay IBM for MLC, zOTC may stay with customer
 - Outsourcers can't use TFP
- **Individually negotiated contracts:** maybe you negotiate poorly!
 - Because of the way the conversion from R4HA to TFP works, two customers with the same R4HA and software stack may pay different amounts under TFP
 - Hard to say whether a growth rate of \$1/MSU-hour is good or bad



Negotiating a better TFP agreement

Even small customers have leverage



- “These outsourcers/MSPs say they can save me money”
 - Possibly true, especially for smaller customers
 - But if so: IBM is likely going to lose software revenue too, so they’d probably prefer that you stay on your own contract
 - Outsourcing has its own set of issues and you should consider more than just the cost
 - But good point to make to IBM if you’re trying to negotiate a favorable contract
- “It costs too much to move or grow workload on the mainframe”
 - That’s obviously what TFP is meant to help, so easy argument to make
 - **Push for largest possible growth rate discount: don’t accept 50%**
 - Maybe start by asking for 75%
 - If procurement crows about getting you 55%... that’s nothing to crow about

Protect yourself



- You can always go back to R4HA with 30 day notice
 - But that may be hard, if you've taken advantage of some of TFP's benefits
 - So maybe go slow on rolling some of those back, or only roll back what actually helps
- **Push for a "pay lesser of" addendum**
 - I.E. at the end of the year when doing the true-up calculate what the cost would have been under R4HA and if the R4HA would have been less: give you the right to only pay that lesser amount and re-open that the TFP contract for negotiation
 - R4HA *probably* wouldn't be cheaper if:
 - You've negotiated well
 - Your workload hasn't declined

What if consumption isn't growing?



- Remember the baseline is a committed spend
 - If you use less than the baseline MSU-hours/year, you've paid for MSU-hours you haven't used
 - In general, contract should allow you to roll-over unused MSU-hours to next year
 - You can potentially negotiate to be able to get some "blue dollar" rebates for unused MSU-hours
 - This might actually not be as good of a deal as rolling them over to the following year though
- **Push for an option to be able to reset the baseline amount at least once during the contract period**
 - I wouldn't burn that on a small decrease in consumption, but if it's large enough, definitely worth considering

Consider starting the baseline low



- **Push to start the baseline at something less than the measured amount**
 - E.G. ask for the baseline to only be 75%-80% of the measured amount
 - You can probably justify this by talking about tuning efforts, retiring applications, etc.
- Note that the cost will only be reduced by the growth rate
- If usage next year is same as last year, you'll pay IBM the same amount
 - But if you use less, you'll pay less!
 - Gives you some incentive for some performance tuning! 😊
- This can protect you from year to year workload fluctuations
 - Don't have to worry (as much?) about deciding when to switch to TFP
 - Timing may still matter though

What if you expect to grow a lot?



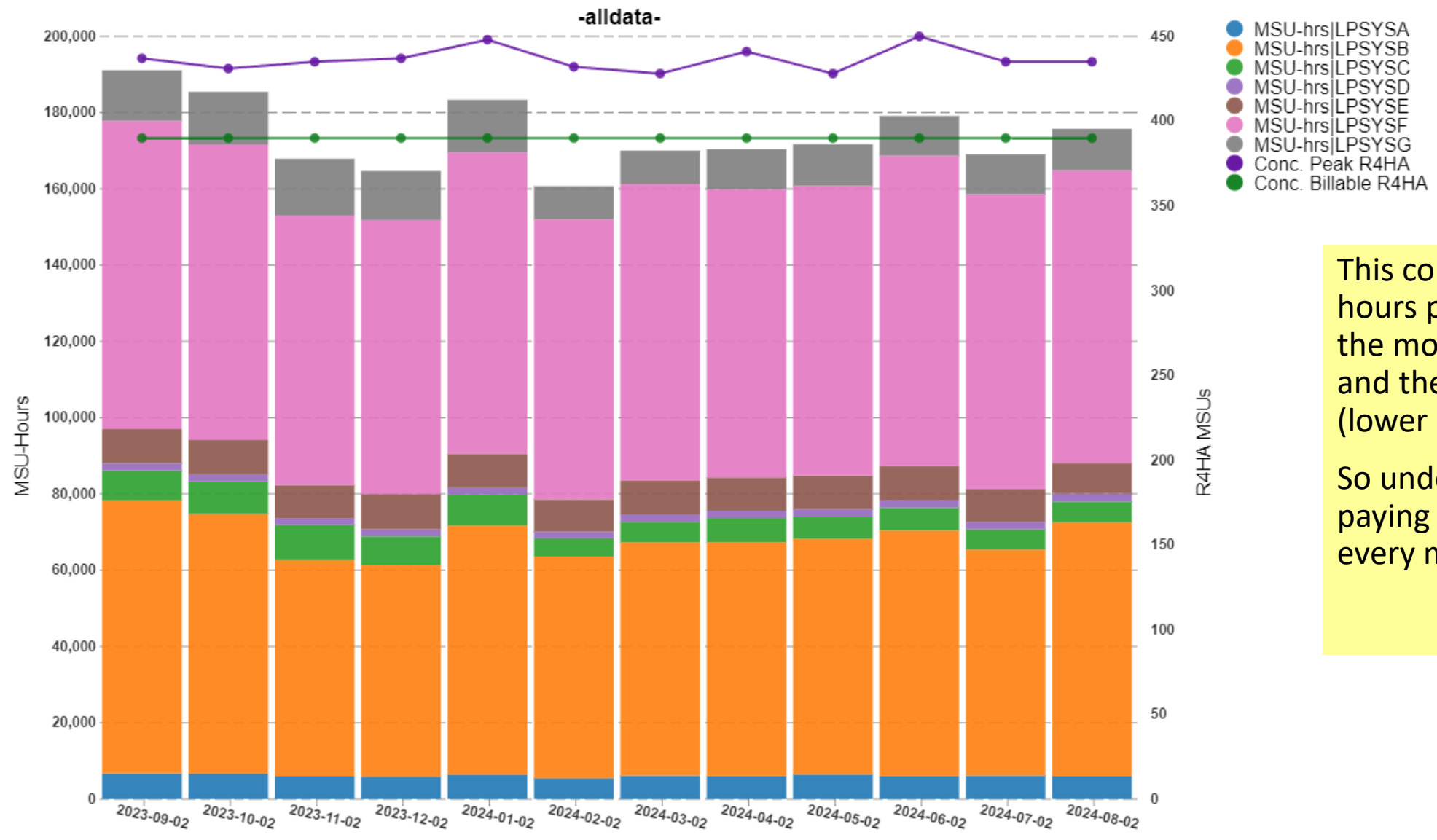
- Try to get a breaker inserted where growth over a certain amount will let you renegotiate
 - Obviously if you go above a certain level, you'd like to pay even less
 - Probably has to be a significant increase to trigger that
 - Most TFP agreements are 3 years, so those will be opportunities as well

TFP Example

Purely Hypothetical!



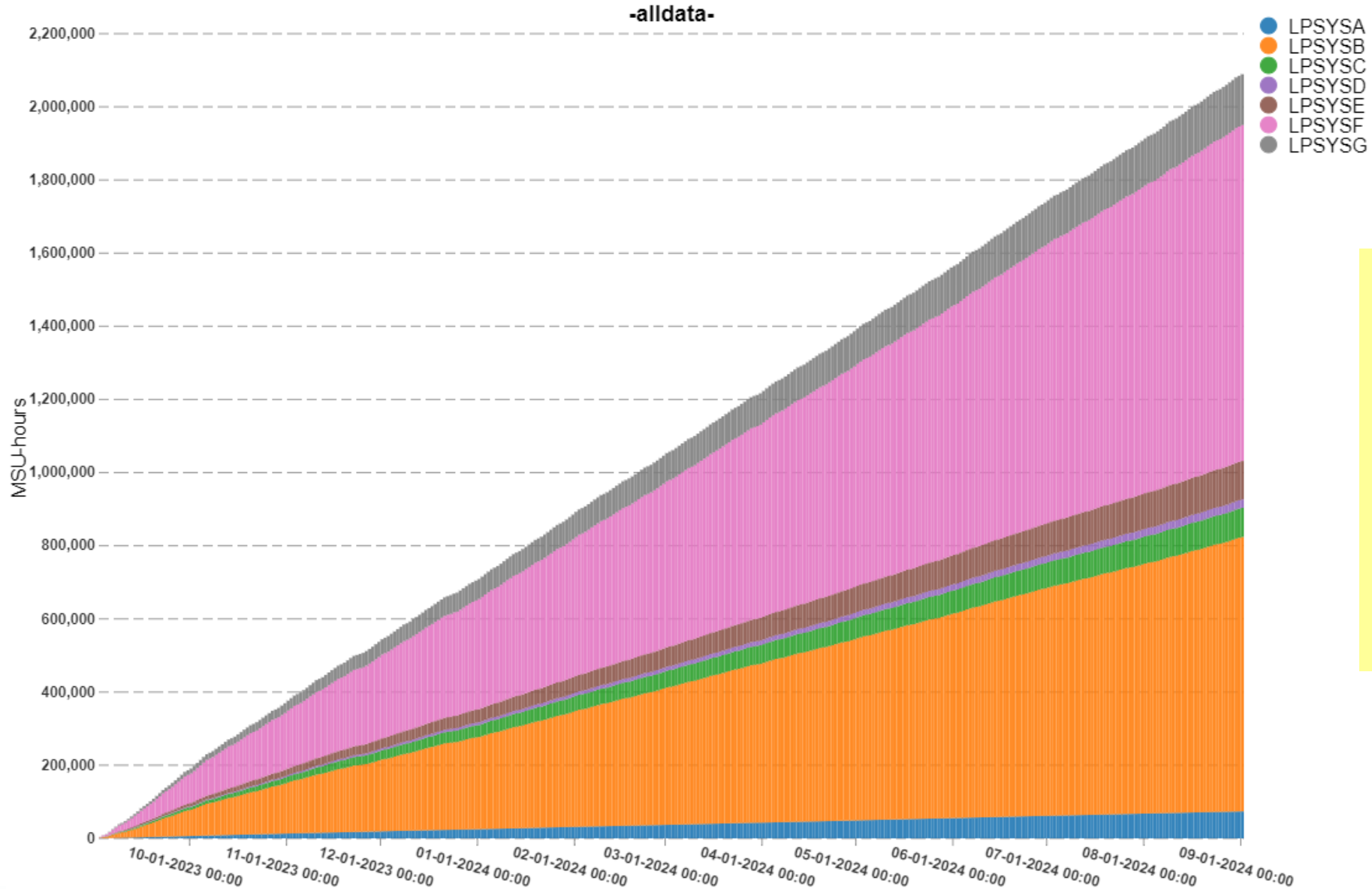
MSU-hour Totals by MLC Month



This compares MSU-hours per month with the monthly peak R4HA and the billable R4HA (lower due to capping). So under R4HA they're paying for 390 MSUs every month.

LPAR MSU-hour Cumulative Totals

2023-09-02 - 2024-09-01



Here's how the MSU-hours grew over the course of the year.

In total, they consumed 2.1 million MSU-hours.

Hypothetically...



- Let's pretend their monthly MLC bill is \$500K. \$6M/year.
 - Ignoring zOTC here for simplicity
- IBM is going to say:
 - Your baseline amount is 2.1 million MSU-hours and your baseline \$ is \$6M/year
 - $\$6M / 2.1M \text{ MSU-hours} = \$2.86/\text{MSU-hour}$
 - So growth above 2.1M MSU-hours will be at only \$1.43/MSU-hour
 - Send us \$500K/month and at the end of the year if you used more than 2.1M MSU-hours we'll bill you at \$1.43/MSU-hour
 - And if you use less than 2.1M, we'll roll those over to next year
- Don't take this deal!
 - You haven't done the **individually negotiated** part!

So you'll say...



- *We only want to pay \$0.70/growth MSU-hour*
 - They'll say no and you'll settle on something like \$.95/MSU-hour (66%, not 50%)
- *We want the baseline MSU-hours set at 1.7M because we're getting ready to do some tuning and blah, blah...*
 - IBM is going to (maybe) say ok, that means instead of a baseline amount of \$6M, it's really $\$6M - 400,000 * 0.95 = \$5.62M$, that will now be \$468K/month 👍
 - Note 400,000 is the difference between the 1.7M and 2.1M MSU-hours
 - Now at the end of the year if you used 2.1M, they're going to come with a bill for the \$380K that you should have been paying them all year
 - So make sure you have that budgeted!! 😬
 - But if you reduce usage, you can end up paying them less than the full \$6M 😊

You'll also say...



- *Thanks, resetting the baseline is good, but we do worry about application leaving that might leave our utilization much lower so... **We want to renegotiate the baseline at some point in the contract to protect ourselves from something like that***
 - You'll probably get this without much fuss
- *Great, BTW, at the end of the year we want you to **calculate what we would have paid you under R4HA and if that would have been less, that's all we want to pay you that lower amount***
 - Also, should not be a difficult, although it does require somebody to do some work
 - Note: if you remove any capping (because that doesn't matter under TFP) the R4HA charges very well may increase, making this ask less important to you

Why negotiation is important



	"Standard"	"Negotiated"	
Baseline MSU-hours	2,100,000	1,700,000	Lowered baseline by ~20%
Baseline \$/MSU-hours	\$2.86	\$2.86	
Growth \$/MSU-hours	\$1.43	\$0.95	66% growth discount
Cost of Baseline	\$6,000,000	\$5,620,000	

What-if usage becomes:	"Standard"	"Negotiated"
2,100,000	\$6,000,000	\$6,000,000
2,300,000	\$6,286,000	\$6,190,000
1,800,000	\$6,000,000	\$5,715,000
1,700,000	\$6,000,000	\$5,620,000
1,600,000	\$6,000,000	\$5,620,000

We have heard from customers who have had great difficulty getting IBM to reset the baseline in the middle of the contract period. Hence, be sure to have that reset clause added!



TFP affect on performance management

Performance tuning matters!



- Under WLC you're managing just the peak R4HA time period per month
 - So tuning outside the peak doesn't much matter (at least directly)
- Under TFP, you're concerned about all consumption at all times
 - This means there's *more* opportunities to save money!
 - Assuming you're trending to use more than the baseline (set baseline wisely!)
 - E.G. reducing the CPU consumption of your performance reporting could reduce your TFP costs
- Capping doesn't help lower total consumption
 - If caps are restricting performance, removing them can improve performance
 - Unless you need caps for your ISV contracts
 - Consider renegotiating those to consumption similar to TFP at renewal

Use less CPU = Fewer MSU-hours



- Reducing CPU means reducing MSU-hours
 - If looking for something to reduce, it doesn't really matter how the CPU is measured: find the big things
- Don't run unnecessary things
 - Obvious, but maybe after going to TFP is a good time to chase down those old jobs/reports that nobody has looked at in 10 years!
- Don't run your machine at 100% busy!
 - Work will generally consume less CPU when run in a less busy environment
 - Additional hardware capacity to run more efficiently may be less expensive than the software costs
- Consider your upgrades carefully
 - More/slower CPs generally will be more efficient than fewer/faster CPs
 - "efficient" meaning fewer MSU-hours consumed per unit of useful work
 - More/faster would be even better, but probably cost-prohibitive for most

How many MSU-hours is that work?



- Sometimes people want to know how many MSU-hours some work is
- Easy conversion:
 - MSUs Per CP = MSU rating of machine / number of GP CPUs
 - z15 603 = 396 MSUs / 3 CPs = 132 MSU per CP
 - MSU-hours for work = CPU time in hours * MSUs Per CP
 - Batch job consumed 1200 CPU seconds
 - $(1200 / 3600) * 132 = 44$ MSU-hours
 - Note that this is a subtly different than the question of “how much of my capacity did this consume” because that involves an element of execution time
 - 1200 CPU seconds over 15 minutes is different than over the course of the day!
- Not if growth cost is \$0.95/MSU-hour and we have 132 MSUs per CP:
 - 1 CPU hour = $\$0.95 * 132 = \125.40
 - 1 CPU second $\$125.40 / 3600 = 3.5¢$

Note this is growth cost, not the baseline committed cost!

Chargeback



- If you want to chargeback (or even just showback) costs, remember that growth rate is just a part of the picture
 - Baseline costs are at a different rate
 - Can't save below baseline cost without resetting/renegotiating
 - Doesn't include hardware costs
 - Doesn't include ISV costs
- But can be a very useful number for accessing incremental changes!
 - Either in a savings or growth situation

Final thoughts



- Tailored Fit Pricing has evolved to be potentially useful to a wider variety of customers
- Customers who are growing capacity should definitely consider it
- Customers who are flat to slightly decreasing might want to consider it
 - Assuming can handle the budgeting uncertainty around the true-ups
- Customers who are cutting significantly may want to stay on R4HA
- **Negotiate well and include:**
 - Greater than 50% discount for the growth rate
 - A baseline less than the measured baseline (in most cases)
 - Allows a possibility of reducing costs
 - Agreement to pay lesser of R4HA or TFP at end of the year
 - Ability to reset the baseline at least once over the contract period



Questions?!?