



IntelliMagic



## Performance Insights for the Newest areas of your z/OS Infrastructure



INTERNET • MOBILE • PERFORMANCE & CAPACITY • CLOUD • TECHNOLOGY  
NOVEMBER 6-9, 2017 | LOEWS HOTEL | NEW ORLEANS, LA

Dr. Gilbert Houtekamer, IntelliMagic BV

# Agenda

- Human Intelligence and Availability Intelligence
- **zHyperLink (not in proceedings)**
- Specialty Engines (aka PCI boards)
- WLM, Goals and Transactions
- Serialization and Locks
- TCP/IP
- **zIIP – SMT & Classic Metrics (see proceedings)**
- IntelliMagic Vision supported record types (11500+ reports)

# Human Intelligence and SMF Data





# Big Data Challenge

- Too much (SMF) data to look at
- SMF has clues to many problems in the making
- Reduced staff and old tooling makes it hard to keep track
- 'Predictive Analytics' to the rescue?

# Artificial Intelligence in 1991

## CMG Proceedings 1991

Topic was how to pick data that would well with IBM's small disk caches.

Resulted in clear  
Never/Must/May Cache  
Recommendation based  
on set of cache performance  
and usage metrics

*Table 1. Suitability for caching as a function of the measured cache cost and cache hit. Application means that the importance of good performance for the application will determine whether or not the high cost can be justified.*

	Low hit	High hit
Low cost	bad	good
High cost	bad	application

The resident set size and cost metrics as presented can be computed in the same way for subsystem and holding-time-based simulations. *This means that it is possible to determine which data sets are good caching candidates, without even performing a full subsystem simulation.* The cost metric can be refined further, for example to favor data sets with high hit fractions to data sets with lower hit fractions.



# Artificial Intelligence from my Perspective

2006:

- Predictive Analytics as basis for our Dashboards
  - Extremely 'dense' representation of issues

2011:

- Automated Human Intelligence with insight/rating engine

2017:

- Over 400 unique insights/rules are applied to your data



# Cognitive IT Operations

- Uses cognitive / AI techniques to improve human decision making in the delivery of continuous availability in a cost efficient way.
- Improve availability through superior and early insight in performance and configuration issues.
- Provide context on issues identified

# Why is predictability so important?



Real-time  
firefighting

Minimize Danger, Maximize Reward



E. Gordon, 2000; Rock, D. 2008



Identified  
future issue



## DELAYS EXPECTED AFTER TRAIN DERAILS NEAR PLUMSTEAD IN CT

The cause of the incident on the southern line is not yet clear.



A Metrorail train seen in Cape Town. Picture: EWN.

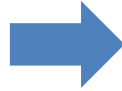
Metrorail Train derails Plumstead Station

Email Print Tweet Share 33

Shamiela Fisher | 2 days ago

CAPE TOWN - A train has derailed near Plumstead Station.

The cause of the incident on the southern line is not yet clear.



# This is when real-time analysis starts!

# Why is predictability so important?

## Unpredictable events:

- Application availability impacted
- Business impacted
- Staff ineffective and stressed
- Equipment used ineffectively



## Predictable events:

- Continuous application availability
- Customer confidence
- Staff productive and innovative
- Equipment used effectively





# Design Perspective

Look at SMF/RMF records looking for:

1. Risk of problems that are likely to affect production,
2. Root cause of problems (now and soon) and how to fix,
3. Inefficiency and waste in the application infrastructure.

Accelerate IT professionals have to get to and implement the solution

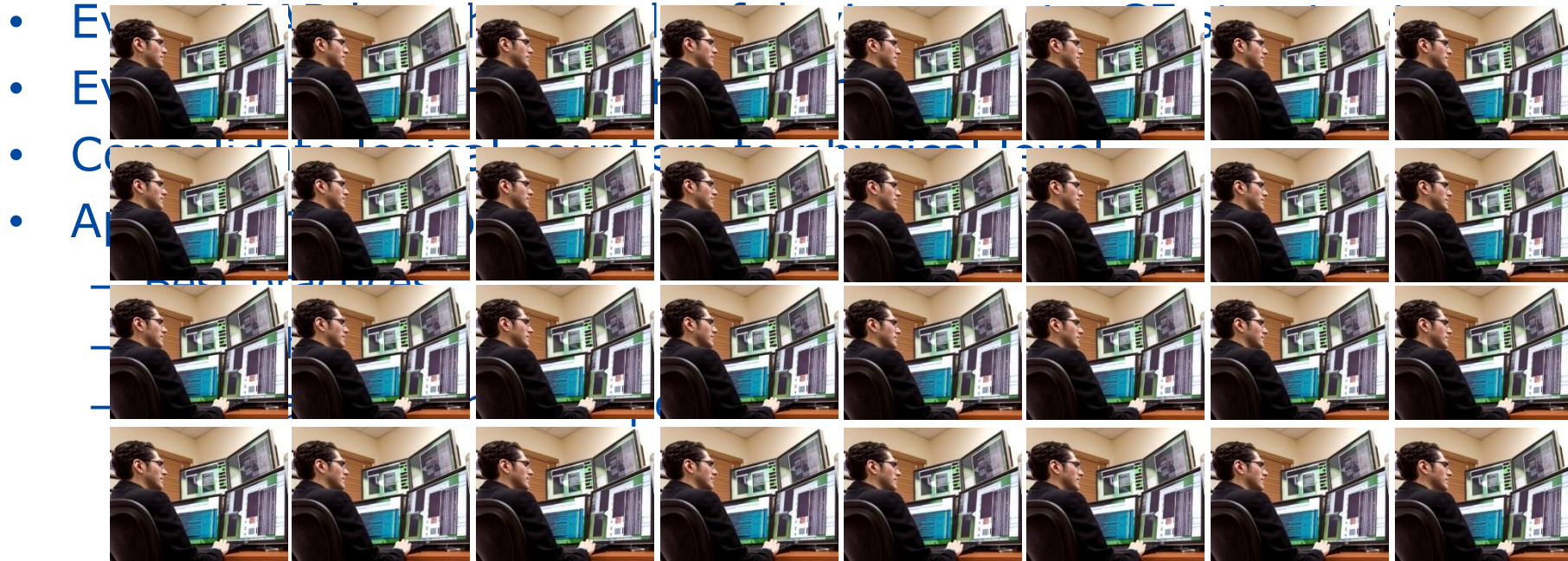
# The Human Intelligence Challenge

- Usually 96 sets of RMF records per day (one every 15 min)
- Every LPAR has thousands of devices, ports, CF structs etc
- Every item has 5-50 metrics to track
- Consolidate logical counters to physical level
- Apply hundreds of rules to 10-100 million data points
  - Best practices
  - Error checking
  - Hardware & workload specific
- Do you track all this with your SAS MICS/MXG reports?



# The Artificial (Machine) Intelligence Solution

- Usually 96 sets of RMF records per day (one every 15 min)



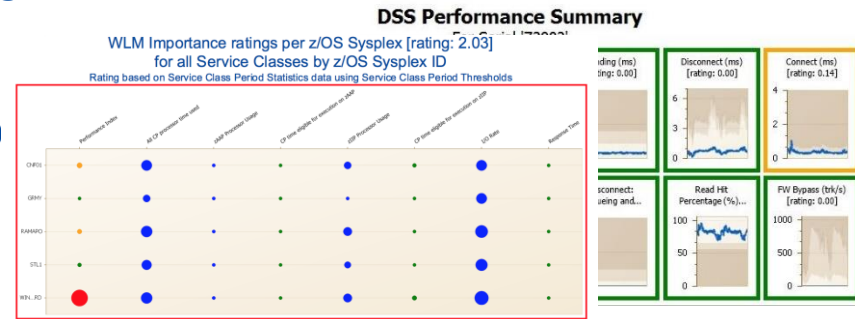
# Better Answers to Difficult IT Ops Questions

- What risky conditions exist right now across our entire environment? (exception charts, rated metrics)
- What related metrics are relevant to the context of this issue? (side-by-side mini-charts of related metrics that are clickable)
- Where do we go next to see root causes?

Coupling Facility Warnings and Exceptions: for all Coupling Facilities by CF Name

Report	Type	Element	Variable	Rating	Severity	Observation
CF_Cache	Bubble Chart	DSNPI10_GBP0	Dir Entries	0.48	Red	A very large portion of the allocated space is used, there may not be enough space left to handle workload peaks. Allocate more storage, and/or reassign storage to the structures that need it most.
CF_Struct	Bubble Chart	IN01	Async Service Time	0.48	Red	The response time is higher than acceptable. Very high response times are caused by one or more overloaded components.
CF_Inst	Bubble Chart	DSNPR09_GBP2	Rate for Queued Requests	0.44	Red	Many requests fail or are delayed.
CF_Cache	Bubble Chart	DSNPI10_GBP11	Dir Entries	0.37	Red	A very large portion of the allocated space is used, there may not be enough space left to handle workload peaks. Allocate more storage, and/or reassign storage to the structures that need it most.
CF_ListAid	Bubble Chart	H05_HEALTHH05LOG	List Entries	0.31	Red	A very large portion of the allocated space is used, there may not be enough space left to handle workload peaks. Allocate more storage, and/or reassign storage to the structures that need it most.
CF_ListAid	Bubble Chart	LOG_DPHLOG_013	List Entries	0.31	Red	A very large portion of the allocated space is used, there may not be enough space left to handle workload peaks. Allocate more storage, and/or reassign storage to the structures that need it most.
CF_ListAid	Bubble Chart	LOGGER_OPERLOG	List Entries	0.31	Red	A very large portion of the allocated space is used, there may not be enough space left to handle workload peaks. Allocate more storage, and/or reassign storage to the structures that need it most.
CF_Press	Bubble Chart	CPULP1	Delayed signal rate	0.29	Yellow	Some requests fail or are delayed.
CF_Struct	Bubble Chart	FOK7	Rate for Queued Requests	0.27	Yellow	Some requests fail or are delayed.
CF_Inst	Bubble Chart	DSNPR09_GBP1	Rate for Queued Requests	0.26	Yellow	Some requests fail or are delayed.
CF_Struct	Bubble Chart	AS15	Async Service Time	0.20	Yellow	The response time is higher than would be expected, the storage system is very busy or developed hotspots that begin to impact performance.
CF_Inst	Bubble Chart	DSNPR09_GBP2	Rate for Queued Requests	0.13	Yellow	Some requests fail or are delayed.
CF_Struct	Bubble Chart	G0F	Rate for Queued Requests	0.13	Yellow	Some requests fail or are delayed.

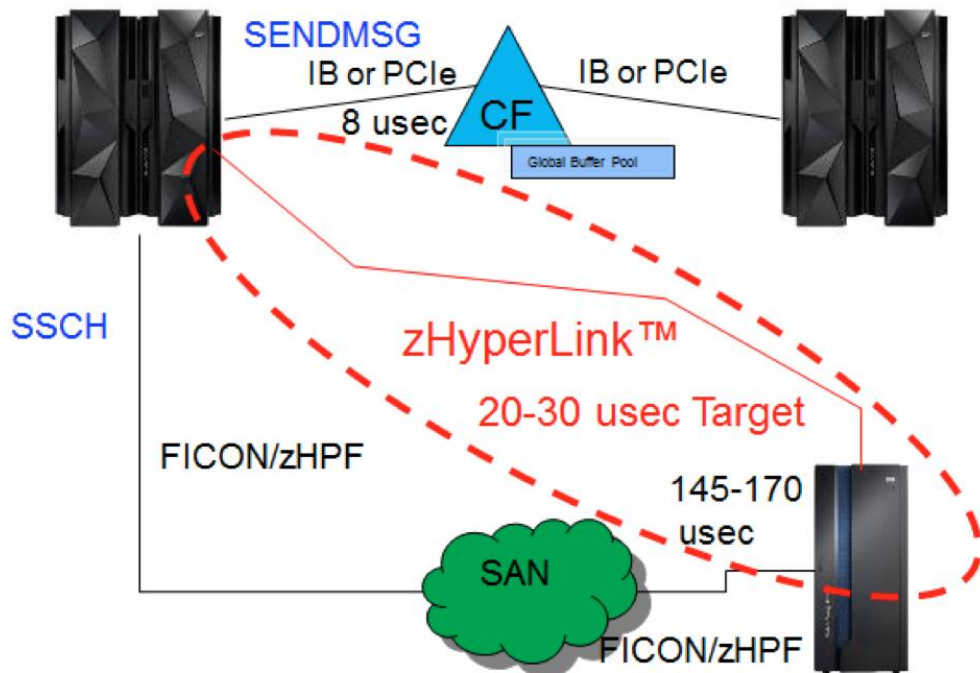
oluti



# zHyperlink – Some Thoughts



# Architecture: IBM's Design Specs



IBM Design spec as well as marketing slide

zHyperLink is about 10 times faster than FICON

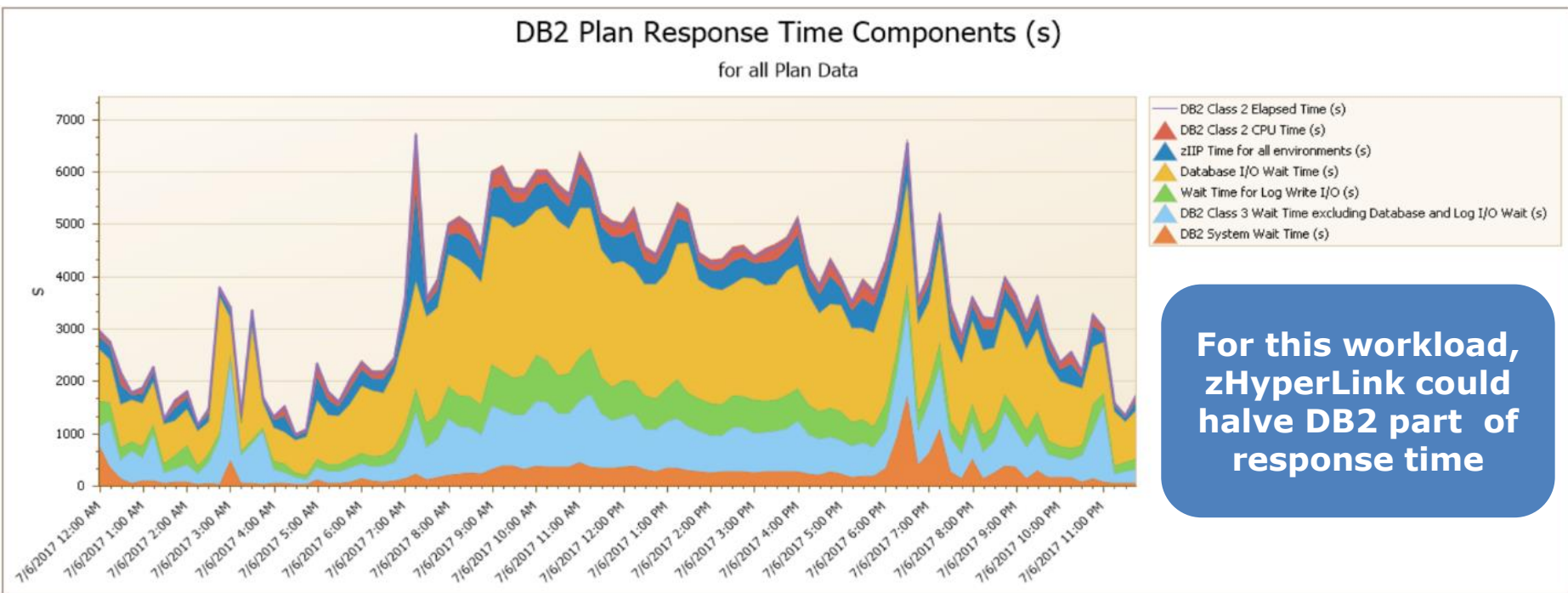
Like for CF, I/O is Synchronous, that is CP is busy waiting for completion

Related Intel paper "When Poll is Better than Interrupt"

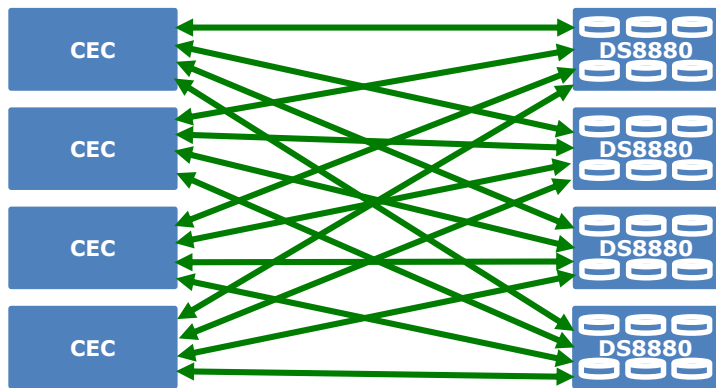
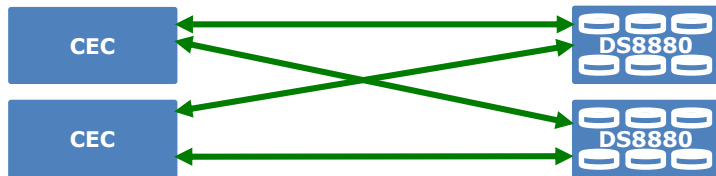
<http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=ca&infotype=an&supplier=897&letternum=ENUS117-005#sodx>



# For now DB2 Only – Potential to cut I/O and Log time

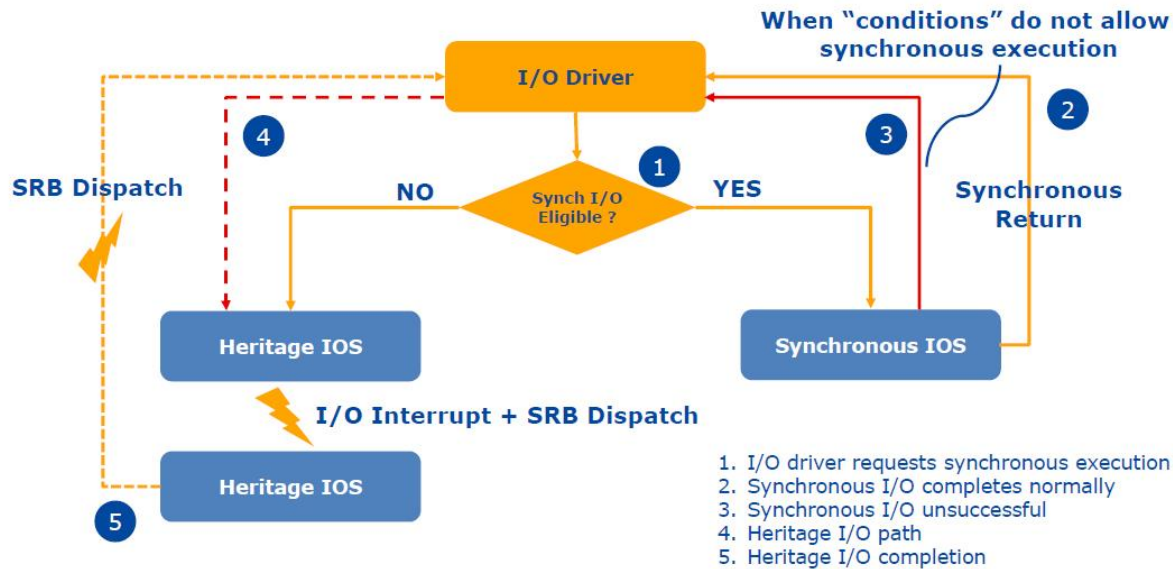


# Connectivity – Four CECs, for DS8880



- Point-to-point 8 Gbyte/s optical links
  - No directors
  - At least 2 for redundancy
  - Likely no more than 2 needed
- Architecturally, they are not channels, but links on a PCI board
- They are not part of Logical Control Units
  - so each LCUs can still have 8 FICON links

# Synchronous Operation – Decided by zOS

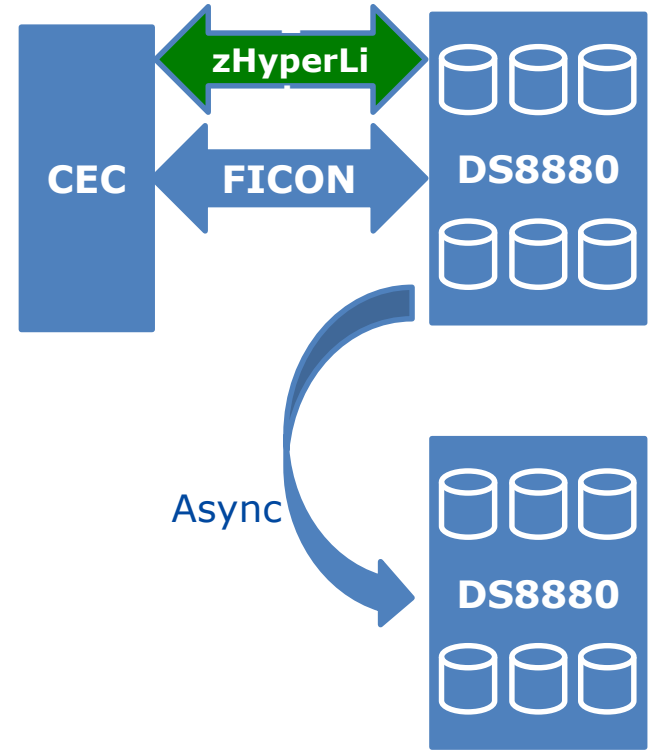


- Eligible I/Os are tried synchronously – cache hit only
- CP in spin-wait – no task switches needed
- Heritage (classic) I/O in case of cache miss
- MLC impact unlikely

# zHyperLink replication setup

- Synchronous copy must be local with zHyperWrite
  - Metro-distance synchronous copy not supported
- Hyperswap is supported (within 150m)
- Asynchronous copy is supported in combination with zHyperLink

⇒ **Replication approach needs to be redesigned**



# RMF & SMF Updates

## RMF 74.1 (Device Activity) & SMF records updated

- Synchronous I/O is new group of fields

## RMF 79.9 (PCI)

- Utilization, usage and error metrics
- Response time distribution in 10 microseconds buckets

## SMF 42.5 & 6 updated

- Using SMF 42 is key, since synchronous I/O decision is by data set, RMF will **not** help you understand application benefits
- Synchronous I/O included in existing microsecond counters
- New detailed counters too

# Specialty Engines (aka PCI boards)



## Specialty Engines (aka PCI boards)

- Processor speed does not increase much anymore
- Mainframe technology is more and more off-the-shelf hardware
- PCI boards are deployed for specific tasks, e.g.
  - Compression
  - Encryption
  - Flash
  - Channel interfaces
- Every new board means at least one new RMF subtype

# zEDC Compression Board

- Throughput information for card (RMF 74.9)
  - PCIe function section for counts/throughput
  - Hardware Accelerator section for response time
- Compression statistics for card (RMF 74.9)
  - Hardware Accelerator Compression section for effectiveness
- Compression usage for address spaces (SMF 30)
  - Find out what jobs are using it



# zEDC Compression Dashboard

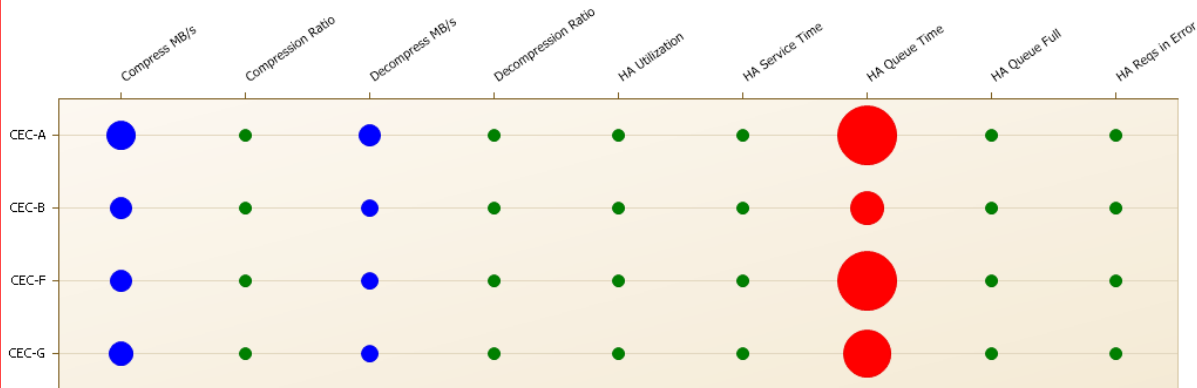
PCI Express and SCM Cards > Accelerators > CEC HA Health

Drill down to:  Systems  Time Charts  Details

## PCI Hardware Accelerator Function Dashboards [rating: 2.42]

for all PCIe Functions by Processor Complex serial

Rating based on PCIe Function Statistics data using System Thresholds



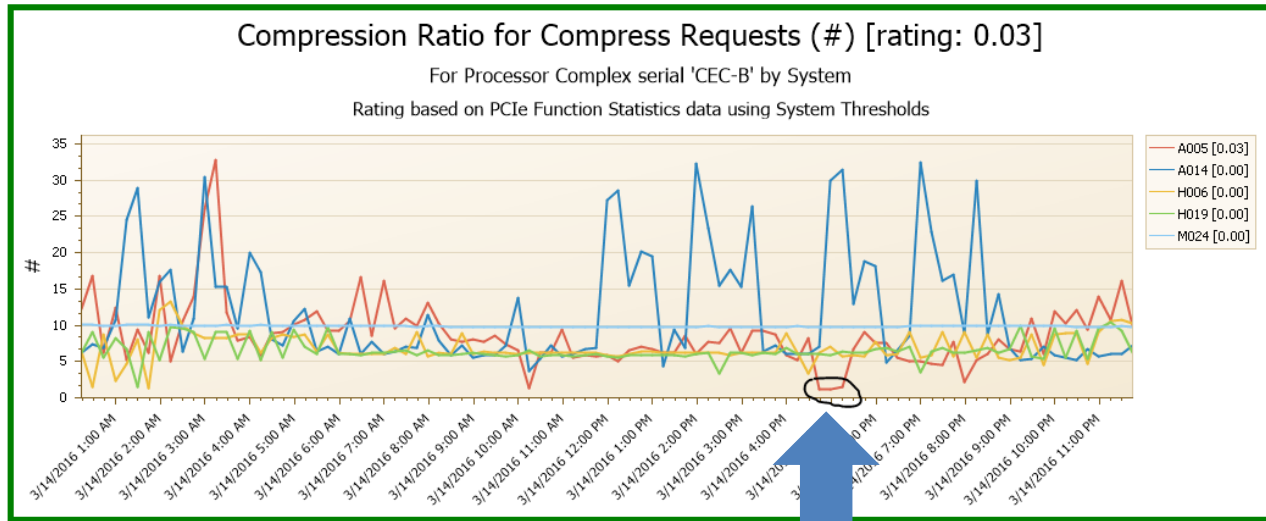
## Evaluate:

- Compression ratio
- Card Utilization
- Service Time
- Queue time
  - can be rather high, not always clear why
- Queue overflow
- Errors

# Not everything compresses well

PCI Express and SCM Cards ▶ zEDC Compression ▶ Compress Ratio ▶ Systems

Drill down to:  Physical Channels  Job  Jobs At Time  Identify  Month  By date  Average by day of week  Isolate



- Phased implementation recommended
- Review average compression
- Investigate when low

# Use Case: Job Driving Poor Compression Ratio

PCI Express and SCM Cards ▶ zEDC Compression ▶ Compress Ratio ▶ Systems ▶ Jobs At Time

Drill down to:  Compression Metrics  Usage Stats

Jobs that use compression (top 1000): For Processor Complex serial 'CEC-B', for System 'A005' by Address Space Name

Drag a column header here to group by that column

	System	Address Space N...	Compress and De...	Compress...	Compress...	Compress Ratio...	Uncompress...	Compressed O...	Decompress Ratio ...	Compressed L...	Decompress...
▶	<a href="#">A005</a>	<a href="#">TYG1PJTS</a>	727767.0	46.1	80.8	1.14	80105.3	70080.2	1.11	87161.0	96505.7
	<a href="#">A005</a>	<a href="#">TBJ1FDRS</a>	10.0	100.0	0.0	0.00	0.0	0.0	6.60	0.1	0.8
	<a href="#">A005</a>	<a href="#">TBJ1YDFA</a>		166.7	0.0		0.0	0.0	6.97	0.0	0.2

IGT 3/14/2016 4:14 PM - 5:14 PM (4:44 P Sysplex <all>, Shift <all>

very many requests .. that do not compress

# PCIe Flash cards: Storage Class Memory (SCM)

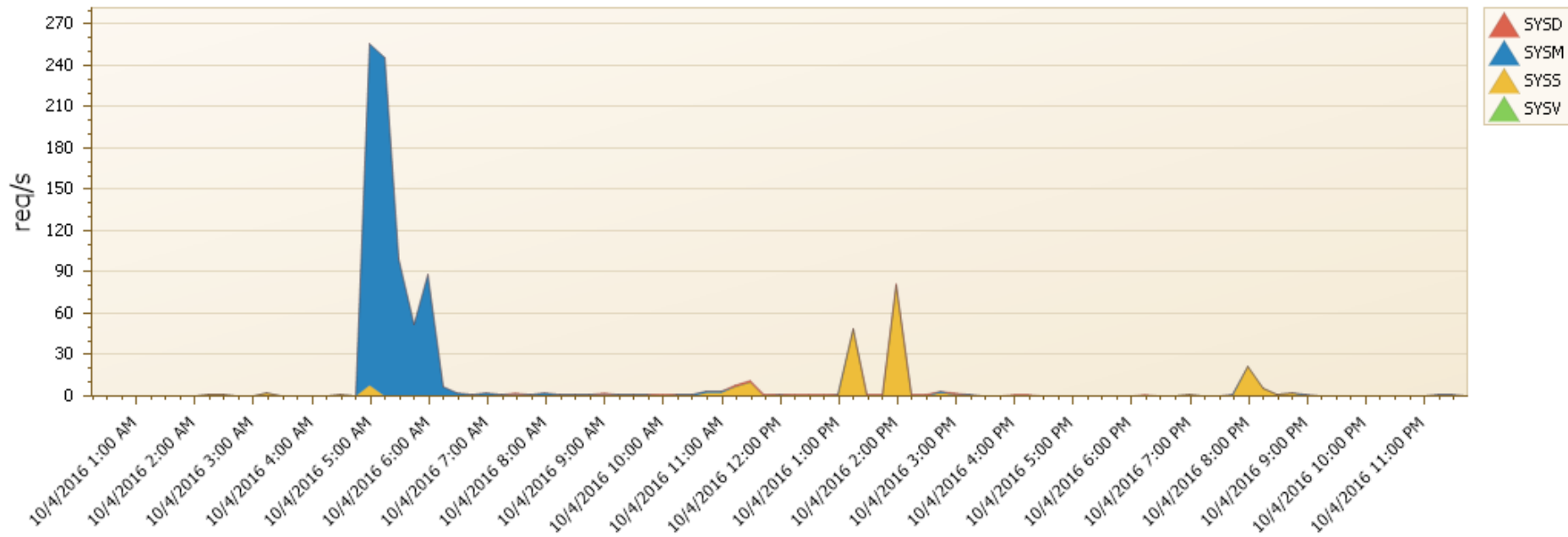
- Extra tier looking for z/OS exploiters
- Good use to back 1 Mbyte & 2 Gbyte pages
  - Recommended!
- Can also serve specialized storage needs (e.g. MQ CF structures)
- Cannot be replicated, so not used for traditional I/O

Already on the way out...

Drill down to:  Month  By date  Average by day of week  Isolate

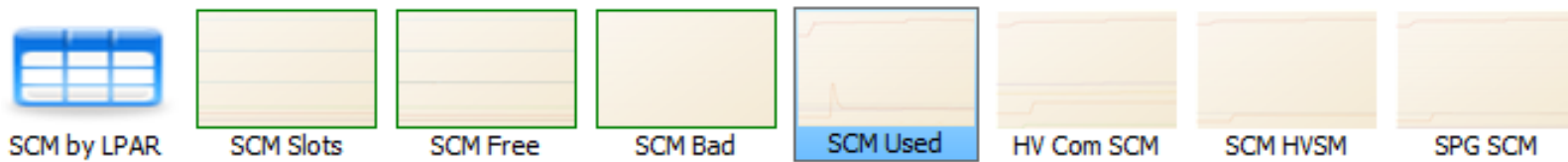
## SSCH Rate to all SCM devices (req/s)

For Processor Complex serial 'IBM-C9067' by System



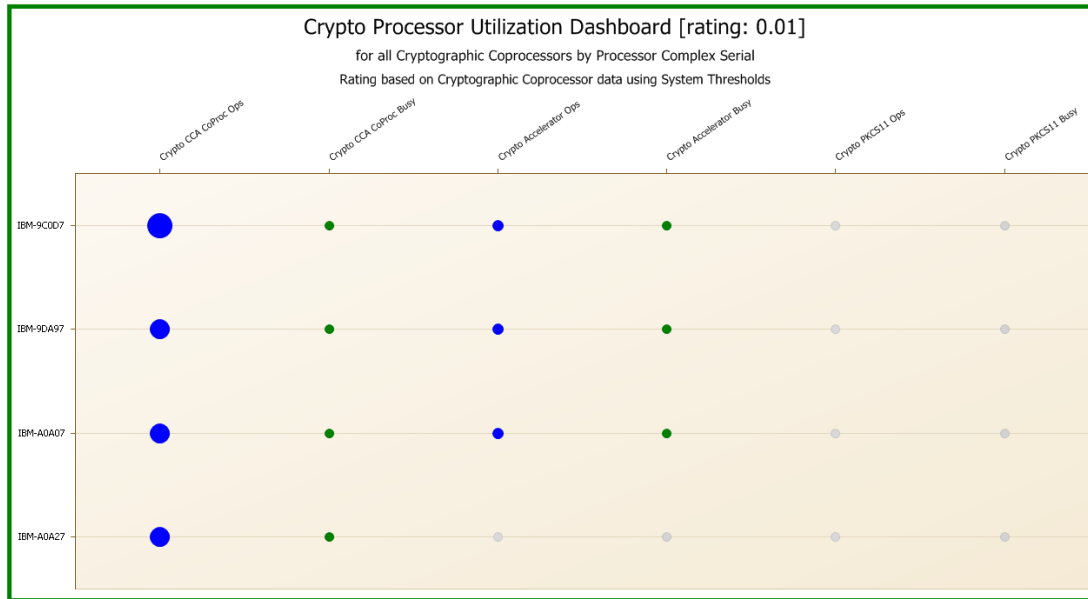
# SCM Page Blocks

## SCM Page Blocks Previews



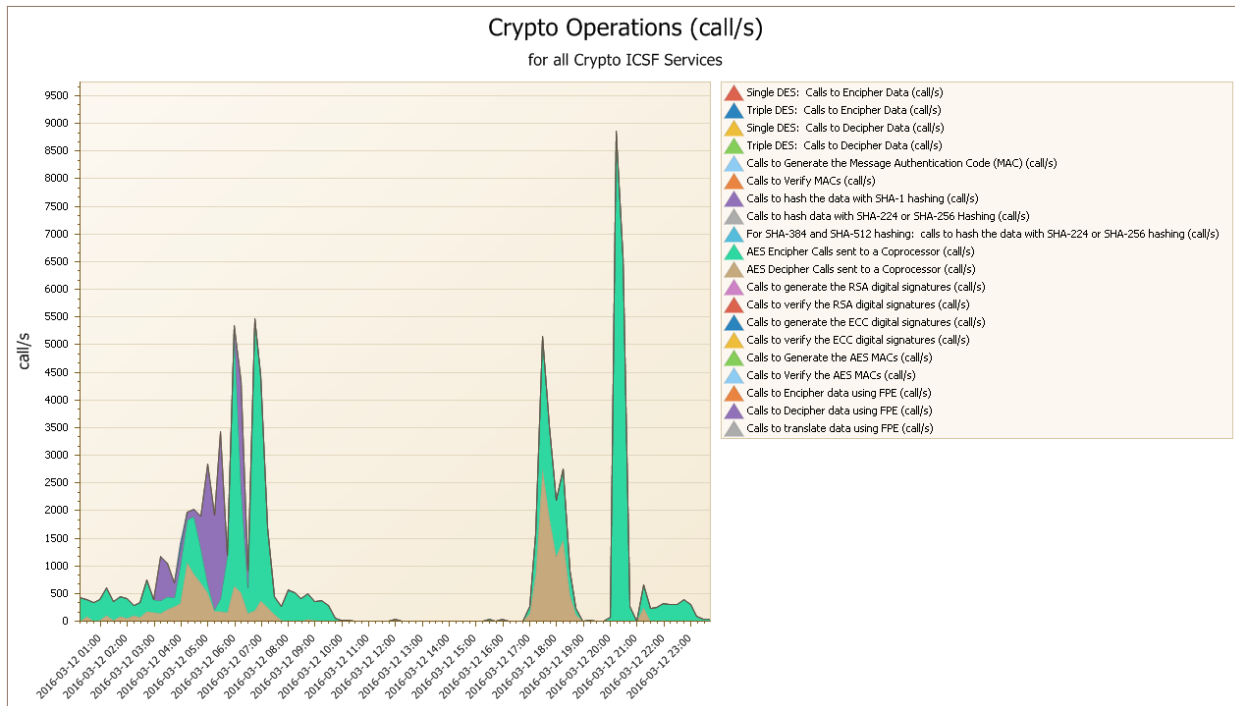
- Includes High Virtual Common and High Virtual Shared memory along with Shared Page Groups
- Report set located both here and in “Paging and Page Movement” focal point

# Crypto PCI Boards



- RMF 70.2 records provide overall utilization (health)
- No per LPAR view on cards
- Independent per LPAR counts by Crypto function

# Crypto Services Function Use



- All functions used through ICSF services

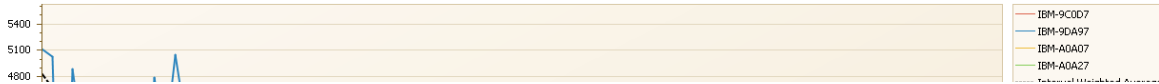


# Compress/Crypto Coprocessors – one per core

Crypto Activity Overview (Functions/call)

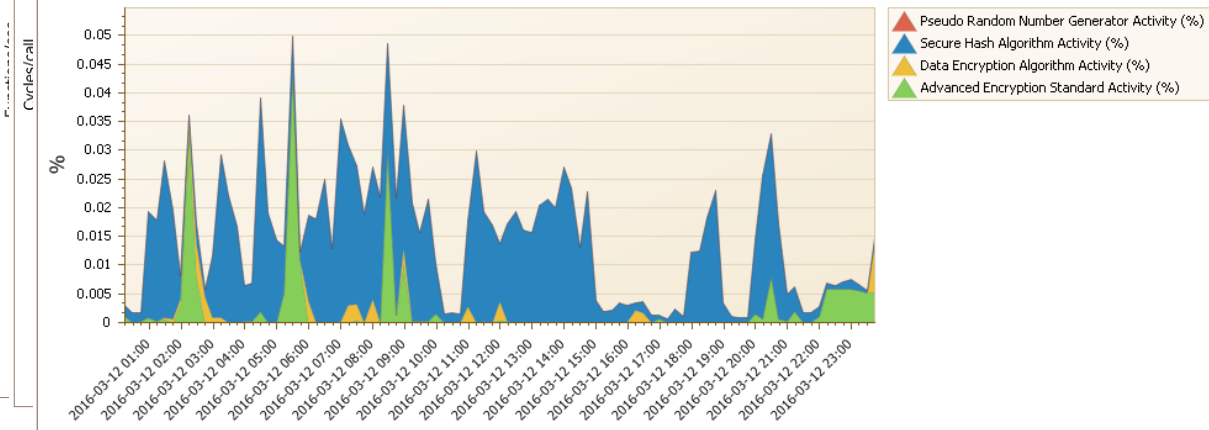
Secure Hash Algorithm Activity (Cycles/call)

for all Processor Hardware Counters by Processor Complex Serial



Average CP Core Usage for Crypto Functions (%)

For Processor Complex Serial 'IBM-A0A07'



- Processor is busy when coprocessor is working for it
- Calls and Cycles recorded in SMF 113 by type
- Core utilization typically very low
- z13 core has dedicated coprocessor, so blocking is unlikely

# WLM, Goals and Transactions



# Change Drivers

- Mobile pricing schemas
  - Need for aggregated processor data in RMF records
  - ⇒ RMF Report Class info now includes CICS, IMS, DB2 transactions
- WLM to Manage Short Transactions
  - Sites run with mostly **Importance Level 1** during MLC peak
  - Very fast I/O creates need to manage millisecond response time
  - ⇒ High resolution response time counters for Service Class / Report Class so WLM has meaningful data

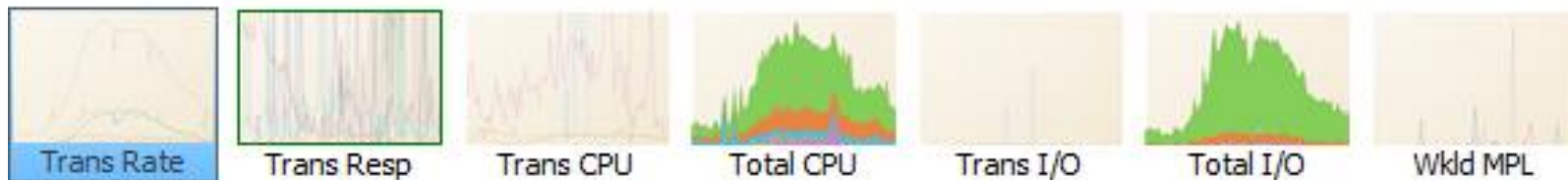
# Transaction Reporting

- Newly-introduced transaction level reporting in RMF 72 records displays key metrics by service class and report class
  - Do not define more than 25 service classes
  - Not a problem to have 100's of report classes200 report classes
- Previously required processing transaction level records
  - CICS 110s, DB2 101s, IMS log
  - Often 10s or 100s of millions500,000 transactions

CICS 5.3, IMS 14

# Service Class / Report Class Transactions

Service Class Transactions Previews

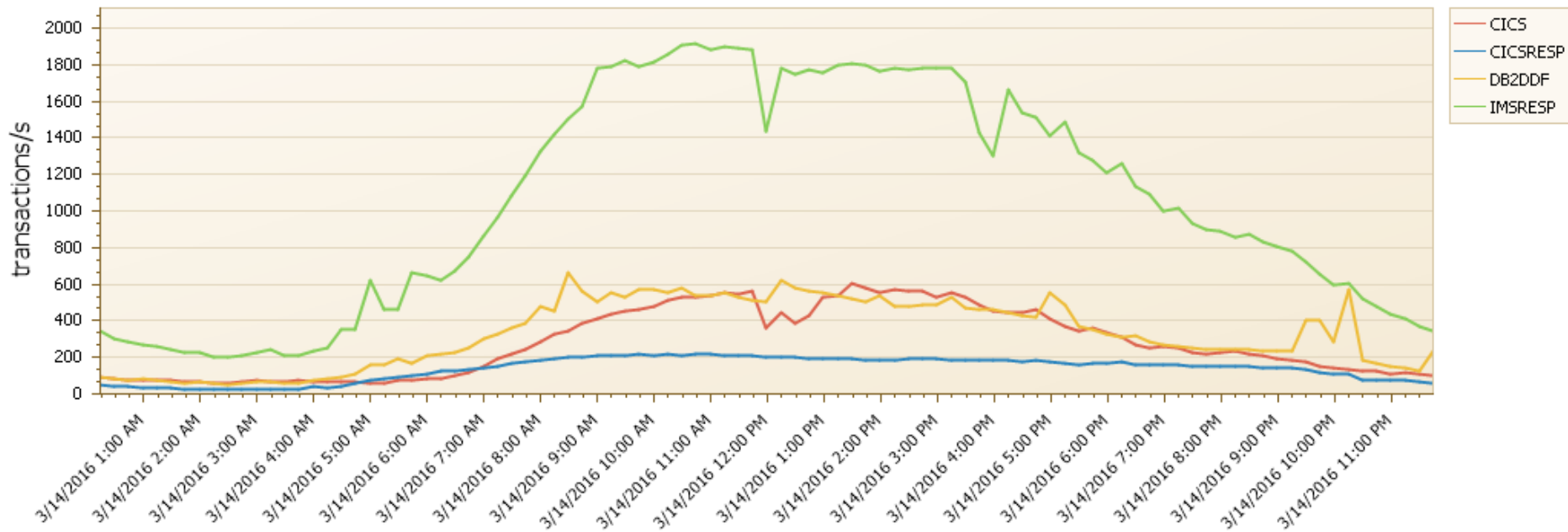


- Key transaction metrics provided (by service or report class)
  - Transaction rate
  - Transaction response times
  - CPU per transaction (and total CPU)
  - I/O per transaction (and total I/Os)
  - Average number of concurrent transactions

**No interval recording!  
Only for ended  
transactions**

Drill down to:  Periods  Systems  Report Class  Identify  Month  By date  Average by day of week  Isolate

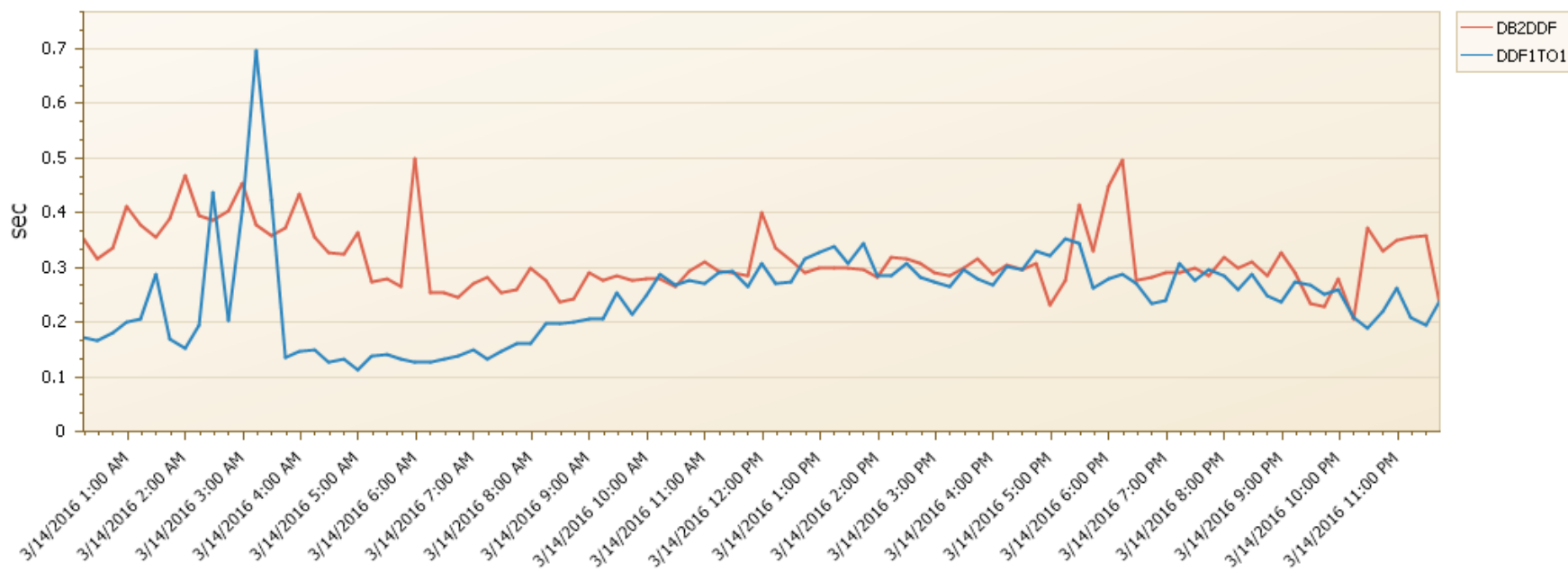
## Transaction Rate for Service Classes with a Response Time Goal (transactions/s) for all Service Classes by Service Class with filter on Service Class



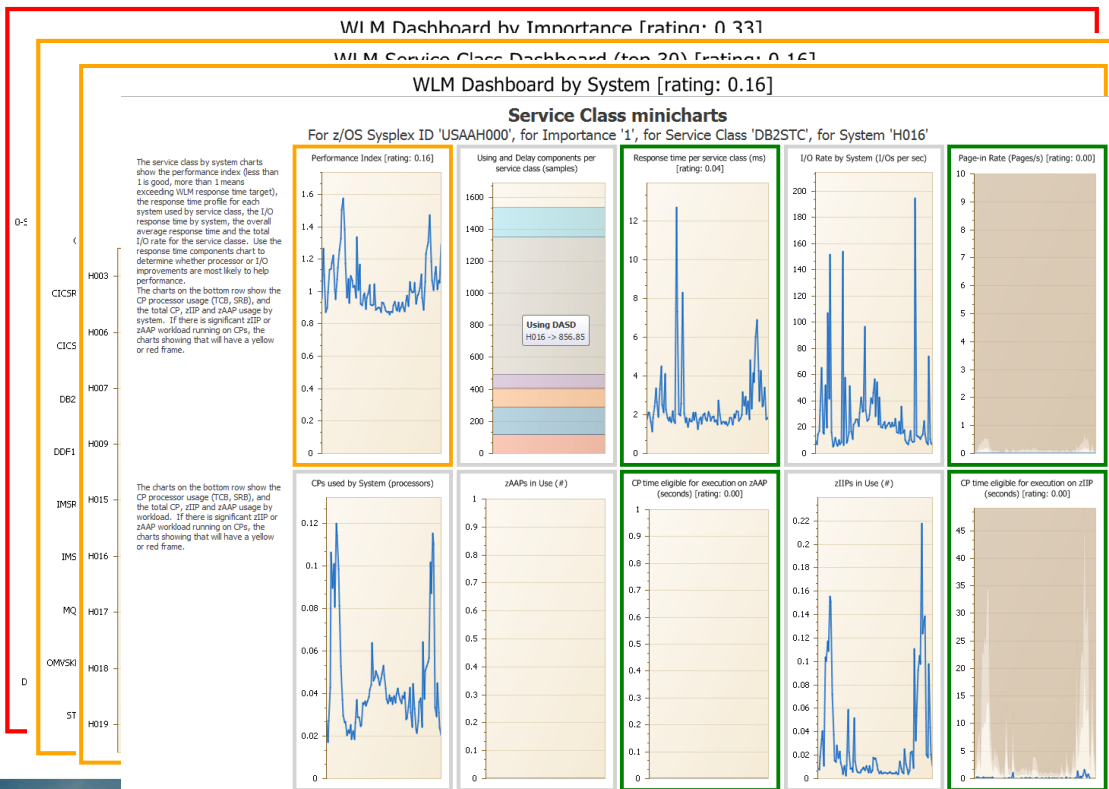
# CPU/Tran & IO/Tran Requires CICS 5.3, IMS 14

CP Processor Time used per Transaction for Service Classes with Reponse Time Goal (sec)

for all Service Classes by Service Class



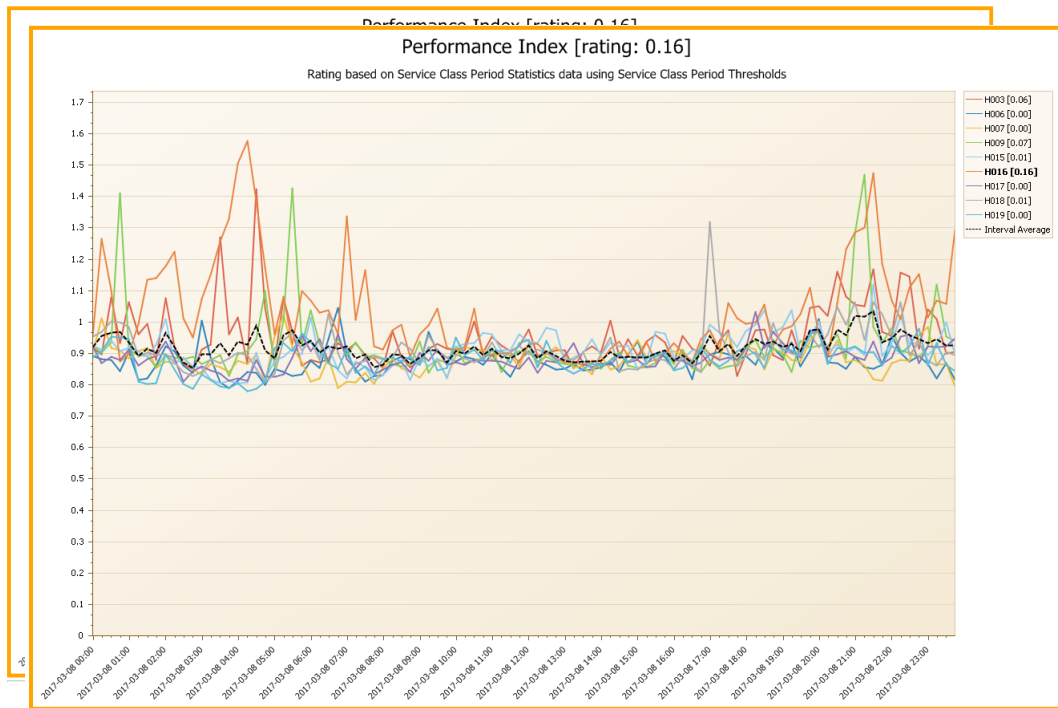
# Performance Index



- Performance Index is what drives WLM decisions
- Achievable goals are very important
- Investigate exceptions
  - Lower expectations?
  - Tune environment?
  - Upgrade?



# Realistic Goal?



- Goal never met for this system ..
- Not met other systems either ..

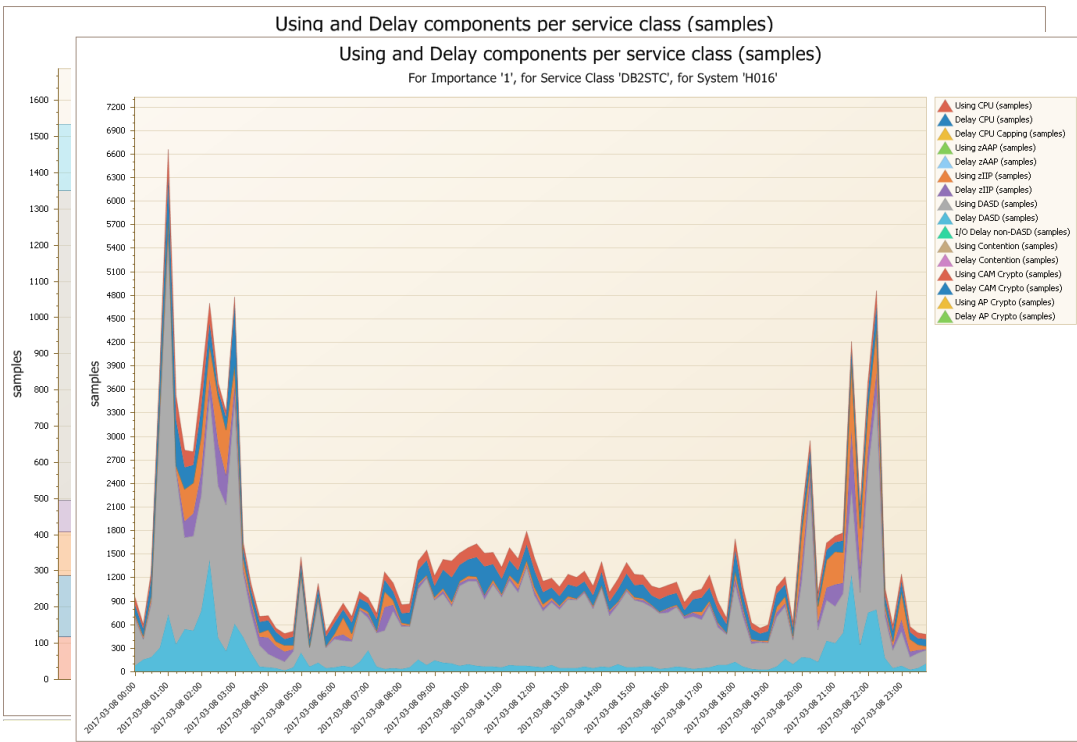
So not realistic, why?

# WLM Workload Response Time Profile

Using and Delay components per service class (samples)

Using and Delay components per service class (samples)

For Importance '1', for Service Class 'DB2STC', for System 'H016'

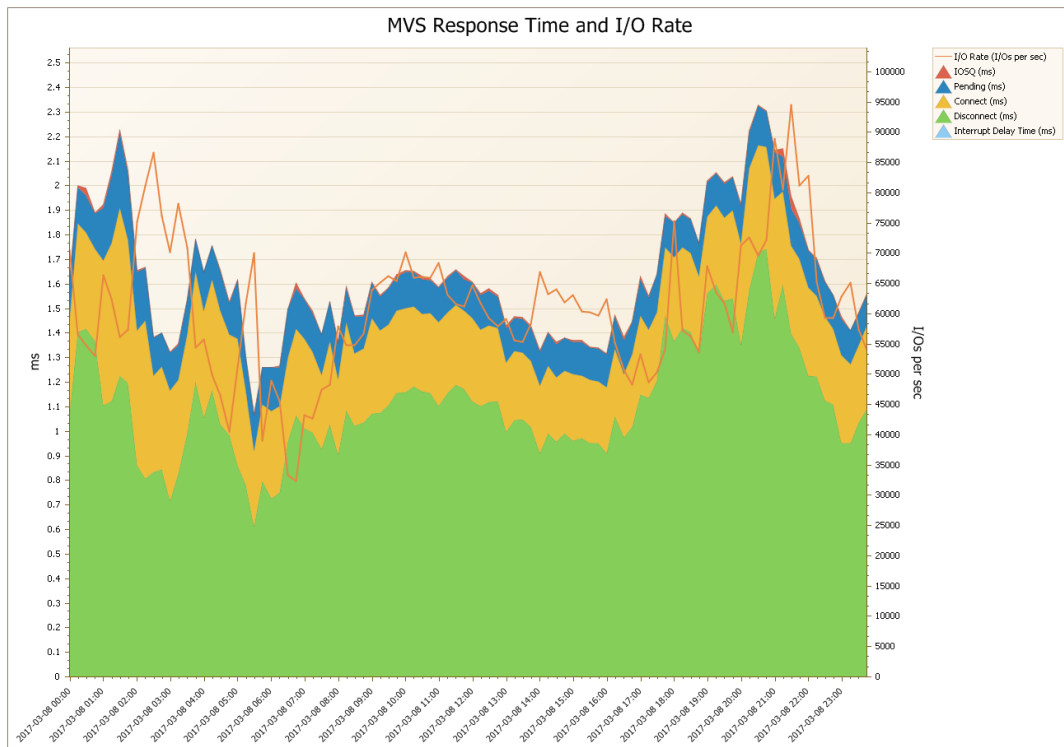


- Mostly storage delay

- All day

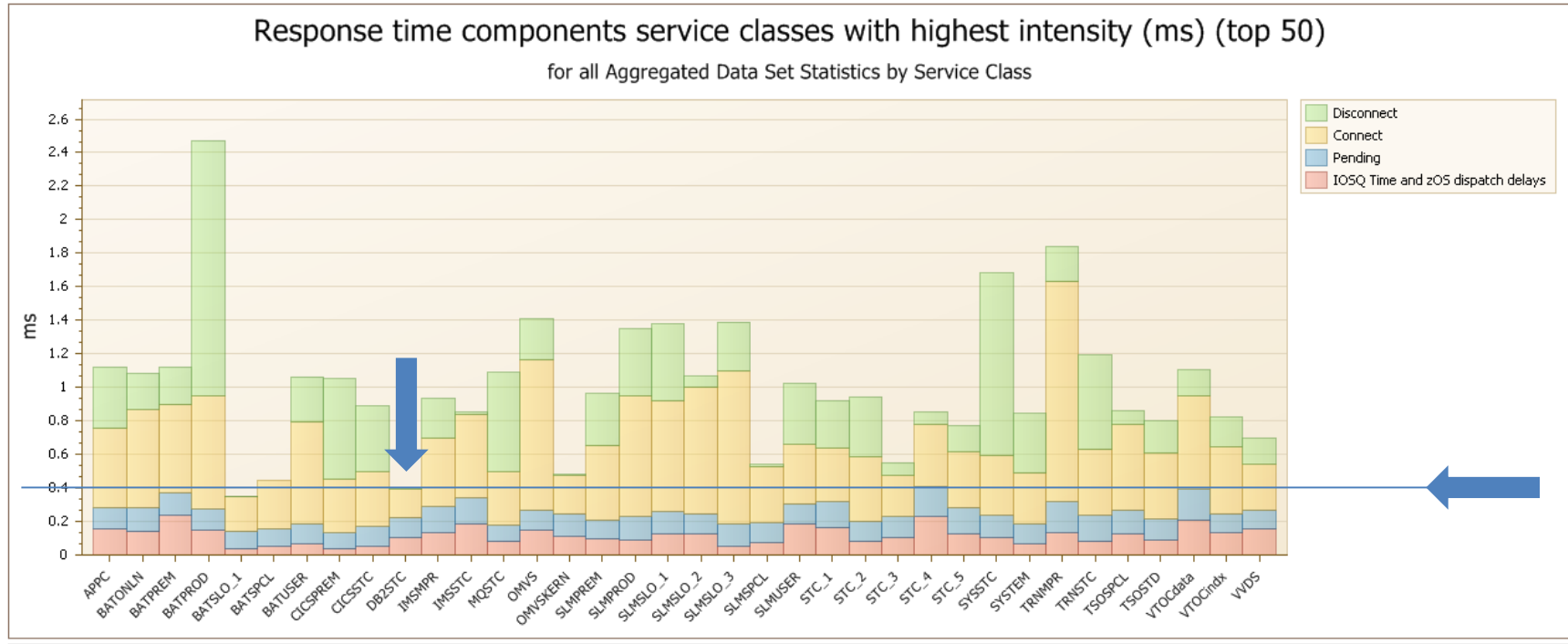
Faster DASD response time would help

# Overall Storage Response Time



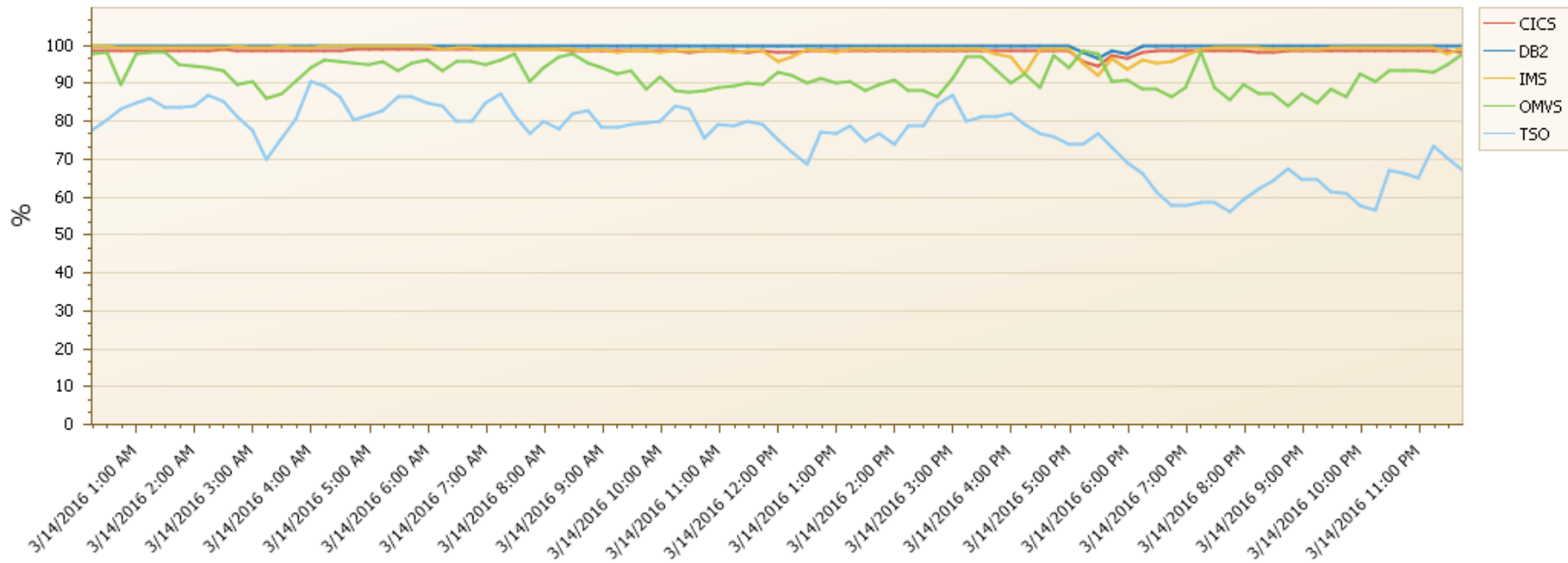
- Storage Response time around 2ms overall
- New hardware would help make this lower, but ...

# Averages can be deceiving



Drill down to:  Systems  WLM Importance  Service Class  Percentiles  Identify  Month  By date  Average by day of week  Isolate

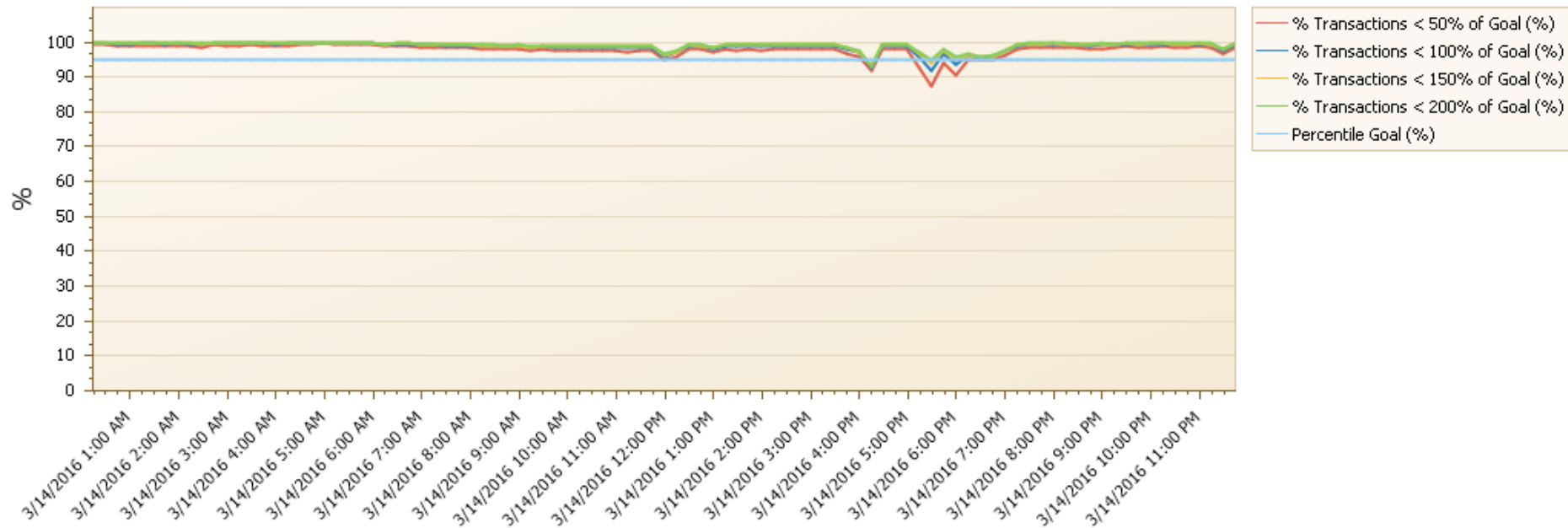
### % Transactions < 100% of Goal (%) for all Service Classes by Workload



Drill down to:  Identify

## % Transaction Completion (%)

For Workload 'IMS'



# Serialization and Locks



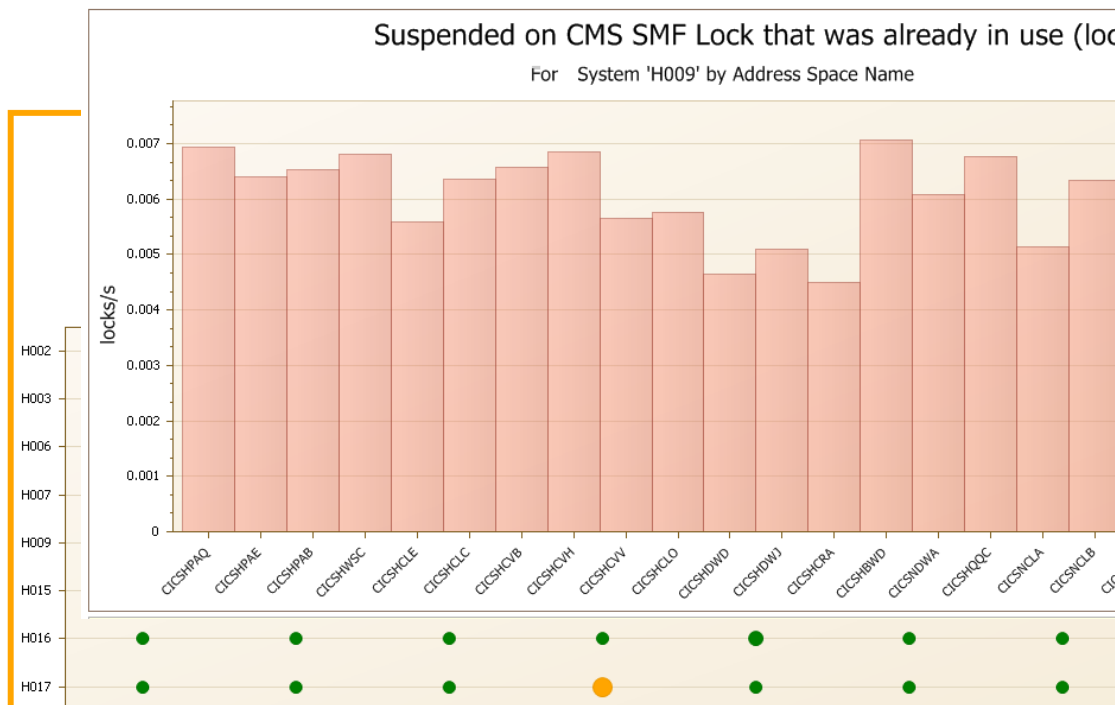
# Serialization and Locks

- RMF 72.5 record introduced with z/OS 2.1
- Information on Locks that cause suspend
  - CMS Locks (CMS, Enq, Latch, SMF)
  - CML Locks (Local Locks)
- Summary information for GRS Locks
  - Step, System, Systems
  - RMF 77 has more detailed data (but most sites use job log)
- Top-20 address spaces included for each lock type



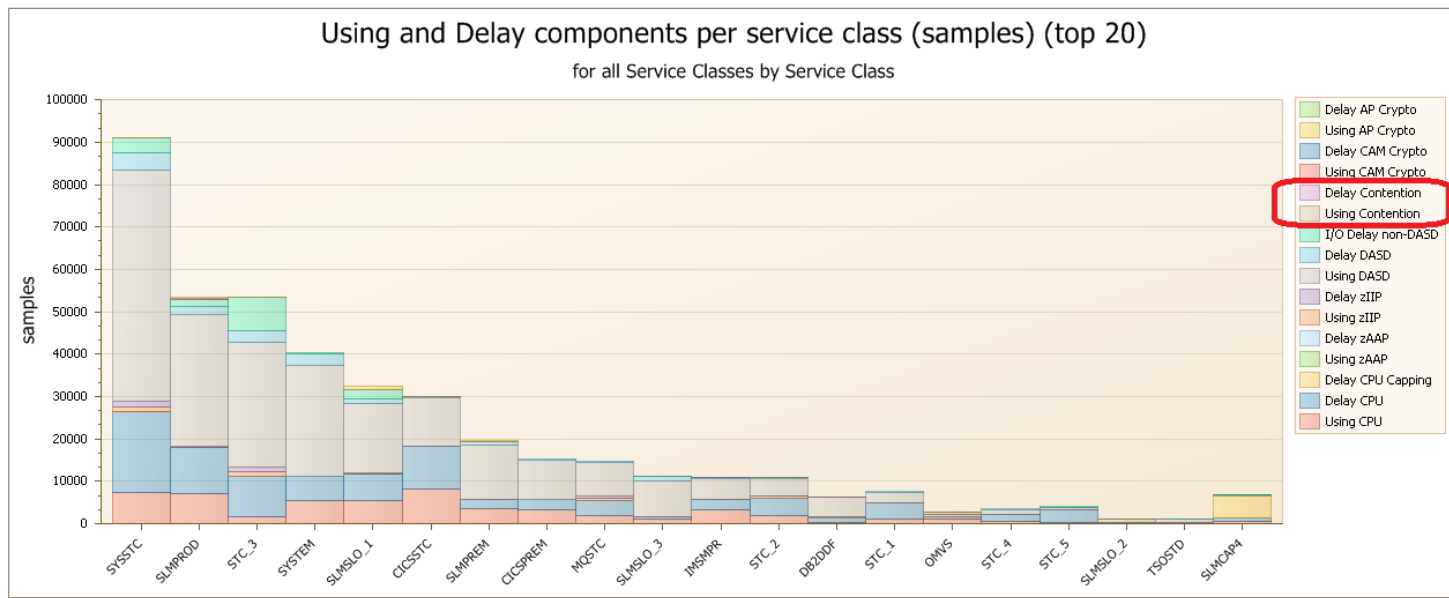
# How to interpret the 72.5 Lock Data?

- Records provide probability that lock is already in use
  - This yields surprisingly high results (10-50%)
  - Suggests that locks come in bursts, given activity
- Impact from 72.3



# Does it matter?

- Workload manager using/delay samples



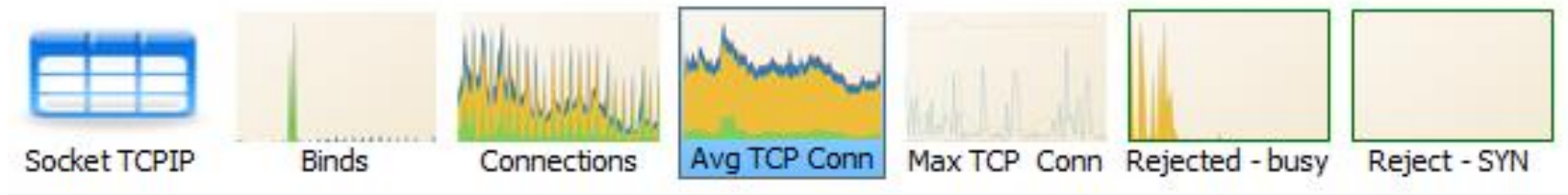
# TCP/IP



# TCP/IP Traffic and Ports

- Network interfaces
- TCP/IP binds and connections by socket and port
- UDP activity and throughput by socket and port

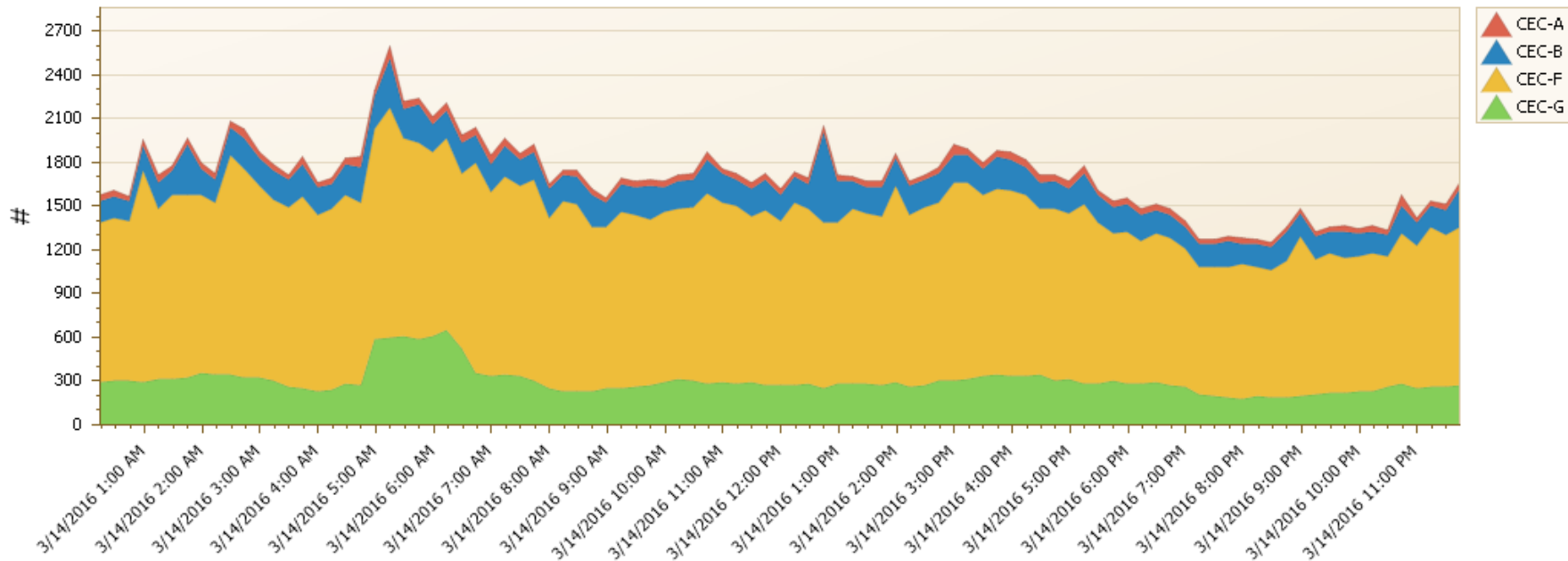
## TCP/IP Previews



Drill down to:  Systems  Sockets  IP Port No  Identify  Month  By date  Average by day of week  Isolate

## Avg No of Active TCP Connections (#)

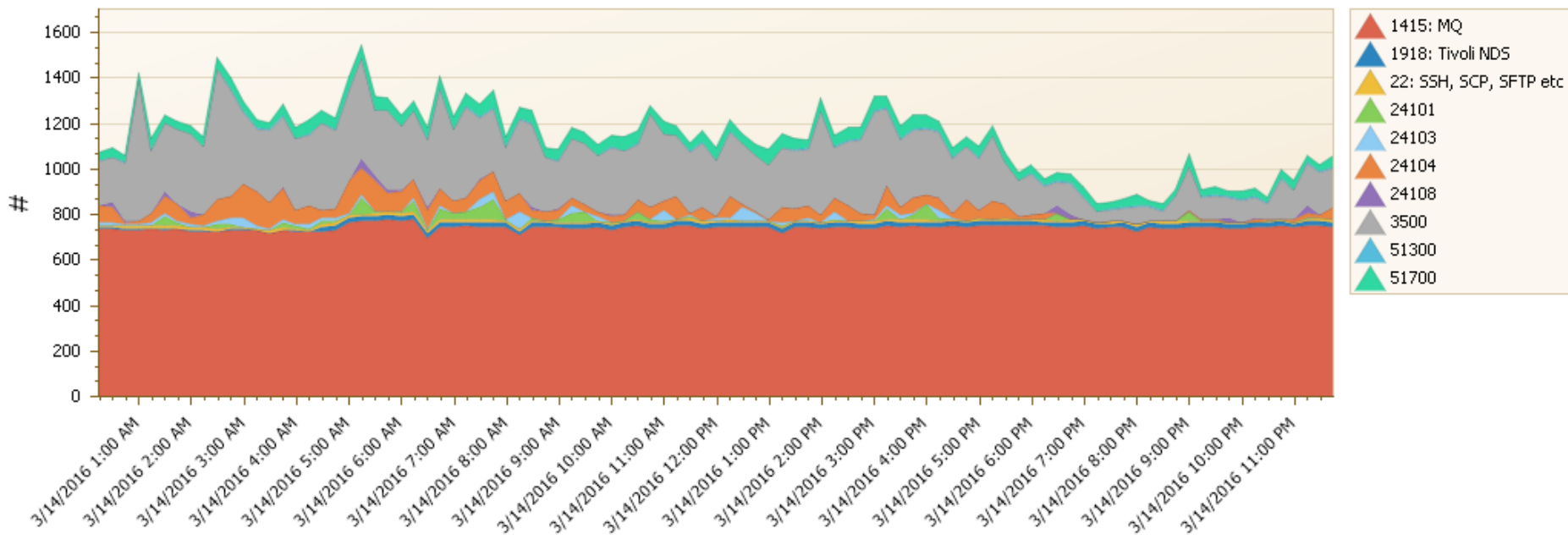
for all TCP/IP Ports by Processor Complex serial



Drill down to:  Systems  Sockets  Identify  Month  By date  Average by day of week  Isolate

## Avg No of Active TCP Connections (#) (top 10)

For Processor Complex serial 'CEC-F' by Port Number



# TCP/IP Statistics

- Aggregated TCP/IP metrics for all interfaces and users
- Includes set of dashboards

## ▲ TCP/IP Statistics

Dashboards

IP Statistics

TCP/IP Statistics

SMC-R Statistics

UDP Statistics

ICMP Statistics

IPv6 Statistics

ICMP v6 Statistics

Virtual Storage

### – Dashboards Previews



Drill down to:  TCP/IP Addr Space  Time Charts  Details

## TCP/IP Error Conditions [rating: 1.00]

For Processor Complex serial 'CEC-B' by System

Rating based on TCP/IP Statistics data using System Thresholds

Open Conn Failure

Input Error

Dupl ACK

Receive Discarded

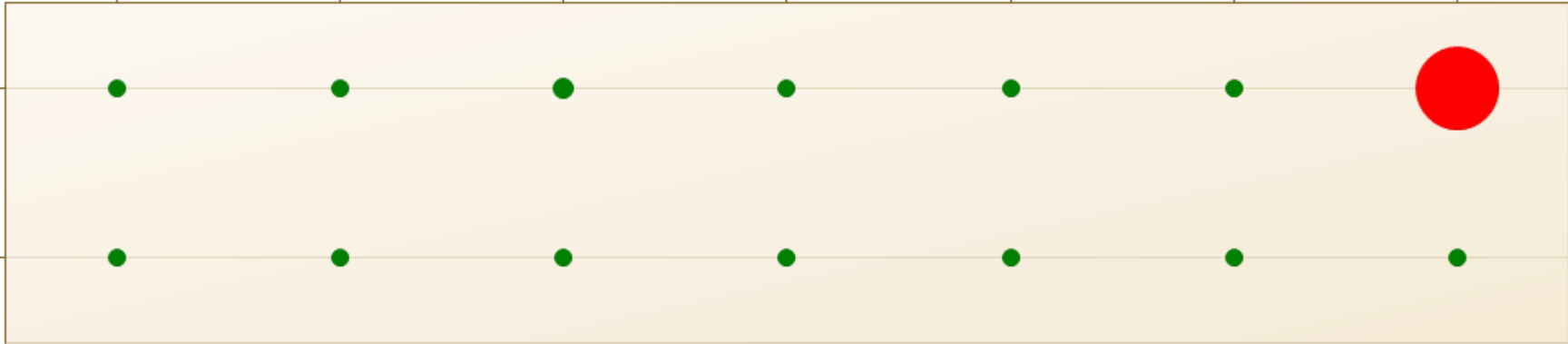
Dropped Conn

Ephemeral Ports Exhausted

Ephemeral Port Usage

A005

A014





# zIIP SMT





## zIIP SMT: Free Extra Capacity

- Multithreading: one or two threads (processors) per core
- Threads share some core resources, so the maximum throughput improvement is around 60%
  - Reflected in the Maximum Capacity Factor (MCF)
- When active, individual threads run slower (@ 160% / 2)
- zIIP throughput will increase, response time may increase too

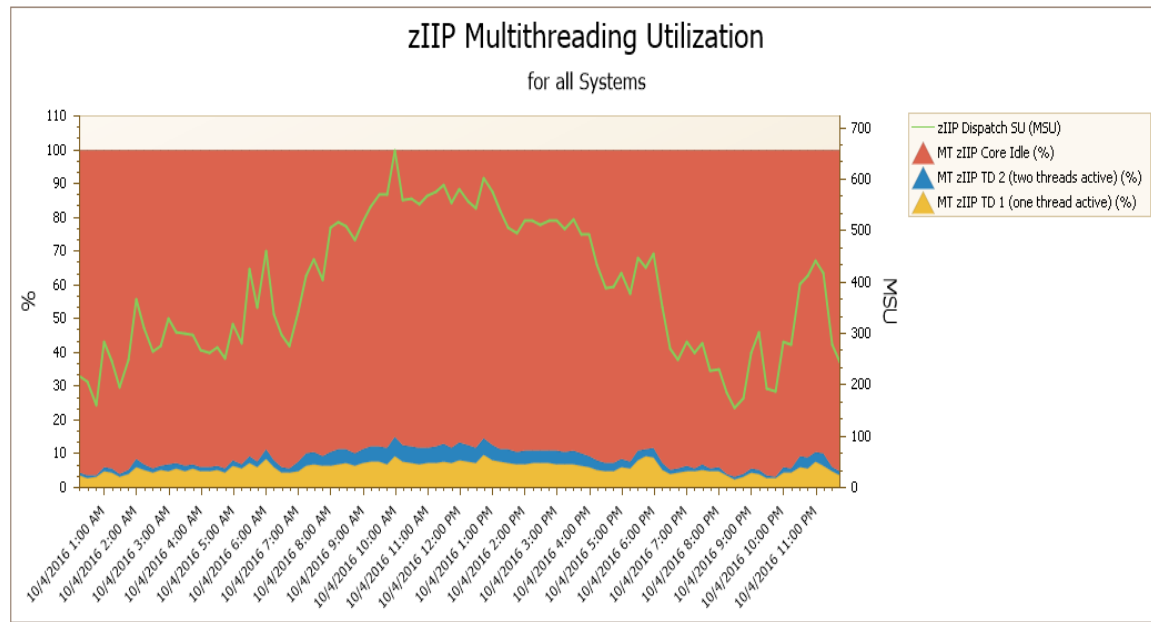
## Why only for zIIPs (not for CPs)

- It kills most existing processor reporting
- Charge back is more complicated
  - Actual time per transactions varies more
  - Depends on average concurrency / competition on core
  - IBM solutions: MT1 Equivalent Time & Counting Instructions
- Harder to determine how much capacity remains available
  - RMF 70 utilization is 'any thread active'
  - MCF is workload dependent MIPS multiplier

# SMT: Idle, One or Two Threads Active

zIIP and zAAP Reporting > zIIP MT > zIIP Usage

Drill down to:  Identify



## CP Two States:

- Idle
- Busy

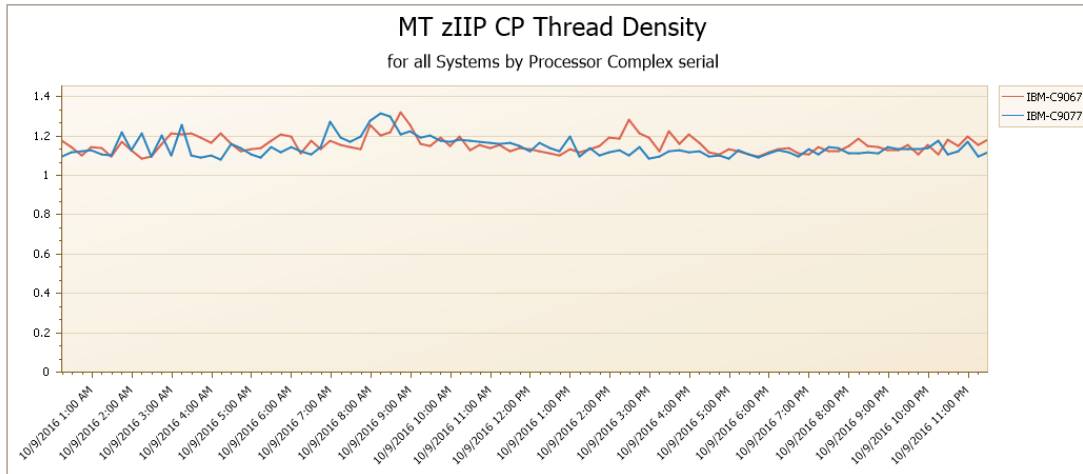
## zIIP SMT Three States:

- Idle
- 1 Thread Active - TD1
- 2 Threads Active - TD2

# Thread Density

zIIP and zAAP Reporting > zIIP MT > MT zIIP TD

Drill down to:  Systems  Identify  Month  By date  Average by day of week  Isolate



- Average **number of active threads** when work is being processed (i.e., core is not idle)
- Between 1 and 2
- For your enjoyment

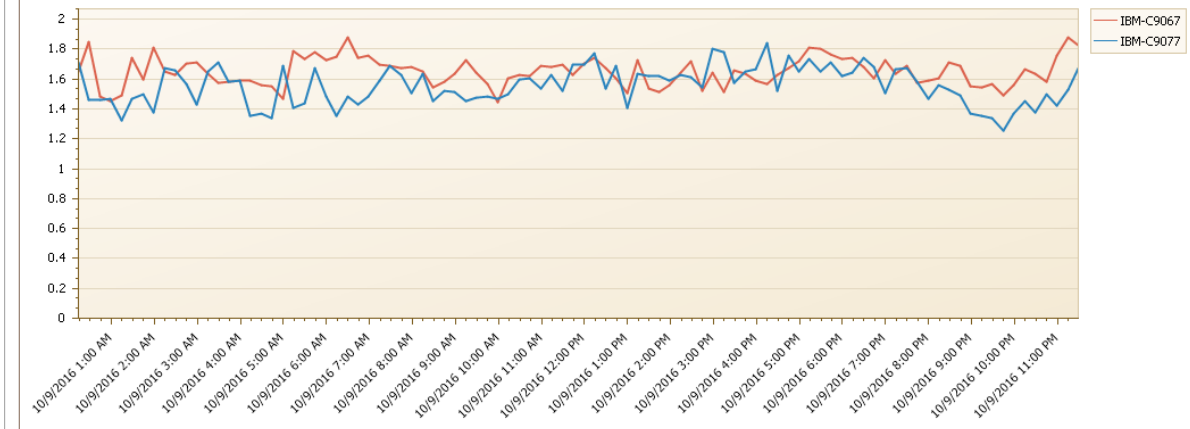
# Maximum Capacity Factor

zIIP and zAAP Reporting > zIIP MT > MT zIIP MCF

Drill down to:  Systems  Identify  Month  By date  Average by day of week  Isolate

## MT zIIP Maximum Capacity Factor

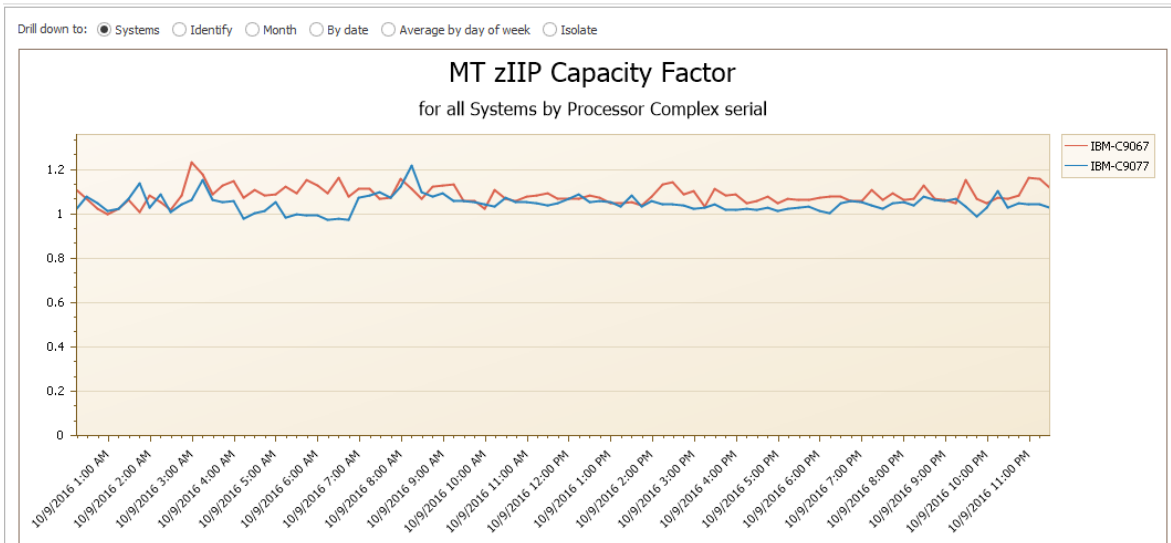
for all Systems by Processor Complex serial



- RMF's estimate how much work the zIIP **could** process in SMT-mode relative to MT=1 core
- Designed as a planning metric
- z14 MCF should be higher

# Capacity Factor

zIIP and zAAP Reporting ▶ zIIP MT ▶ MT zIIP CF

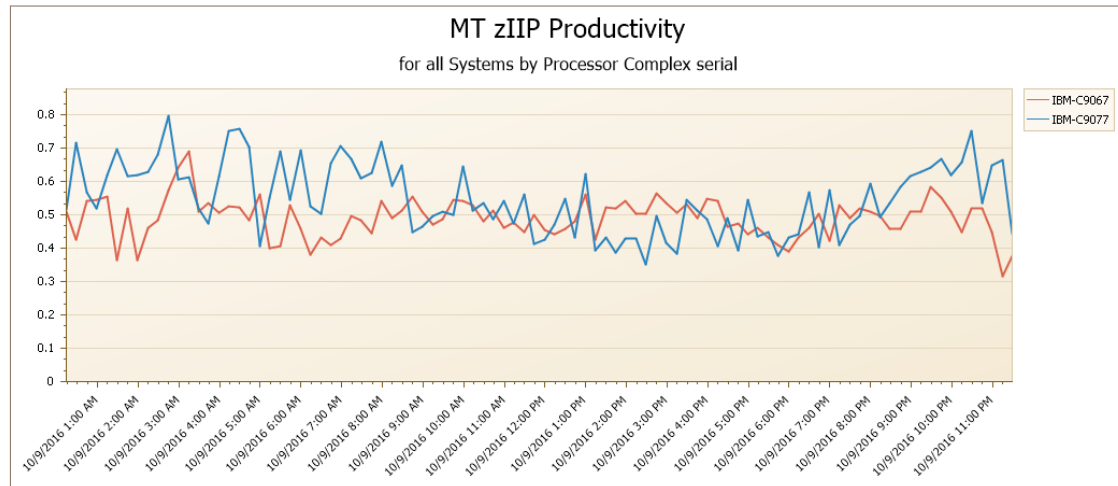


- How much work is actually completing relative to MT=1 core.
- For your enjoyment

# zIIP Productivity

zIIP and zAAP Reporting > zIIP MT > MT zIIP Prod

Drill down to:  Systems  Identify  Month  By date  Average by day of week  Isolate



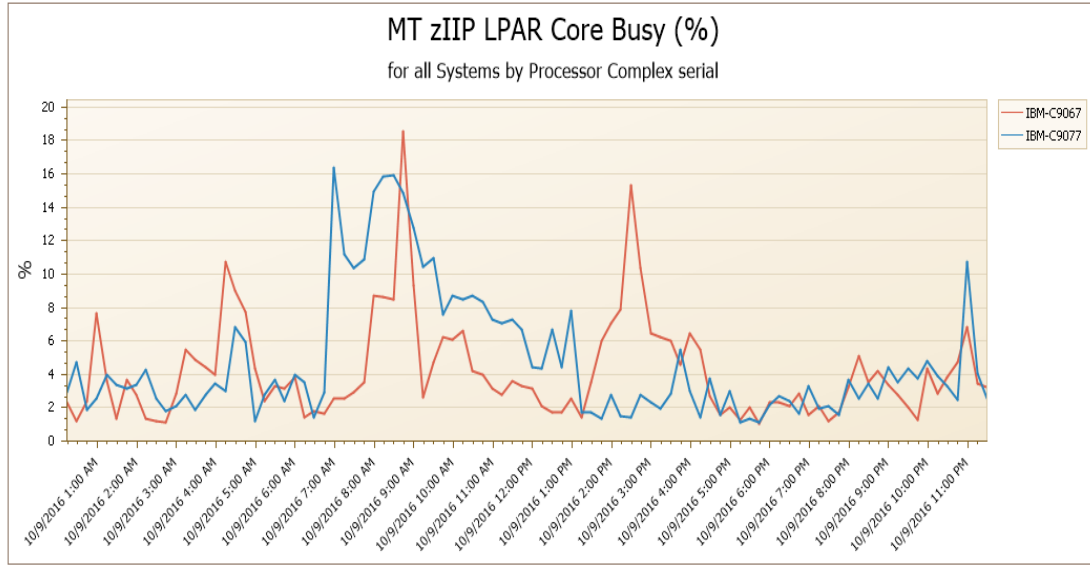
- zIIP Capacity used relative to (estimated) Maximum Capacity Factor
- Shows what fraction of MCF is used
- Expect 100% when 2 threads are always active



# LPAR Core Busy

zIIP and zAAP Reporting > zIIP MT > MT zIIP Util%

Drill down to:  Systems  Identify  Month  By date  Average by day of week  Isolate



- Core is busy when one or threads are active
- Consider this along with Productivity from previous slide

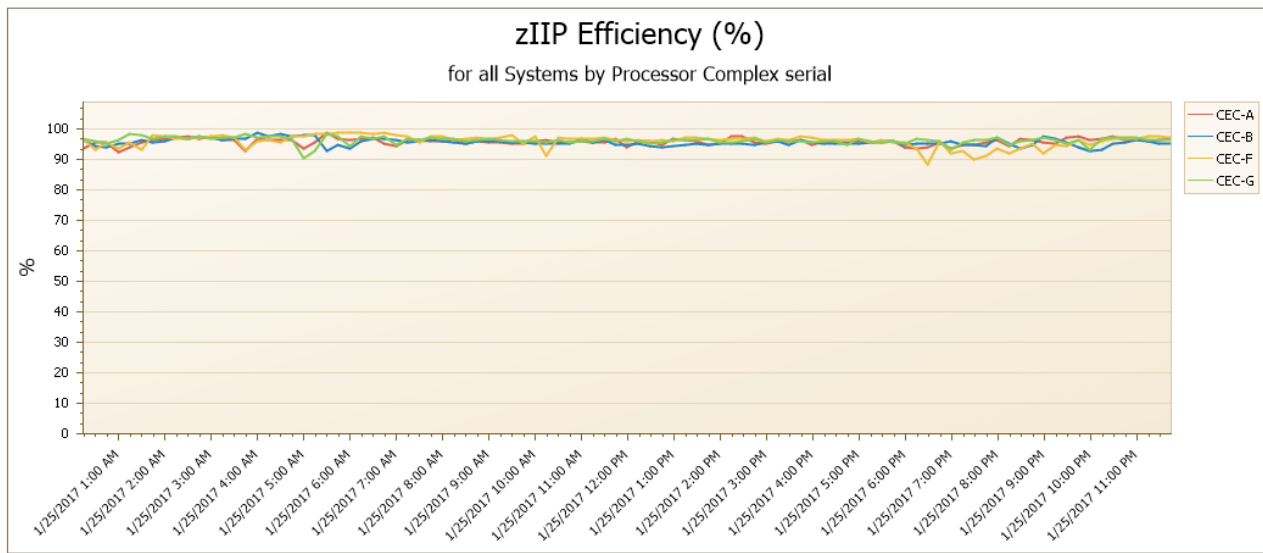
## More zIIP



# How much of zIIP eligible work runs on zIIPs?

zIIP and zAAP Reporting > zIIPs and CPs > zIIP Efficiency

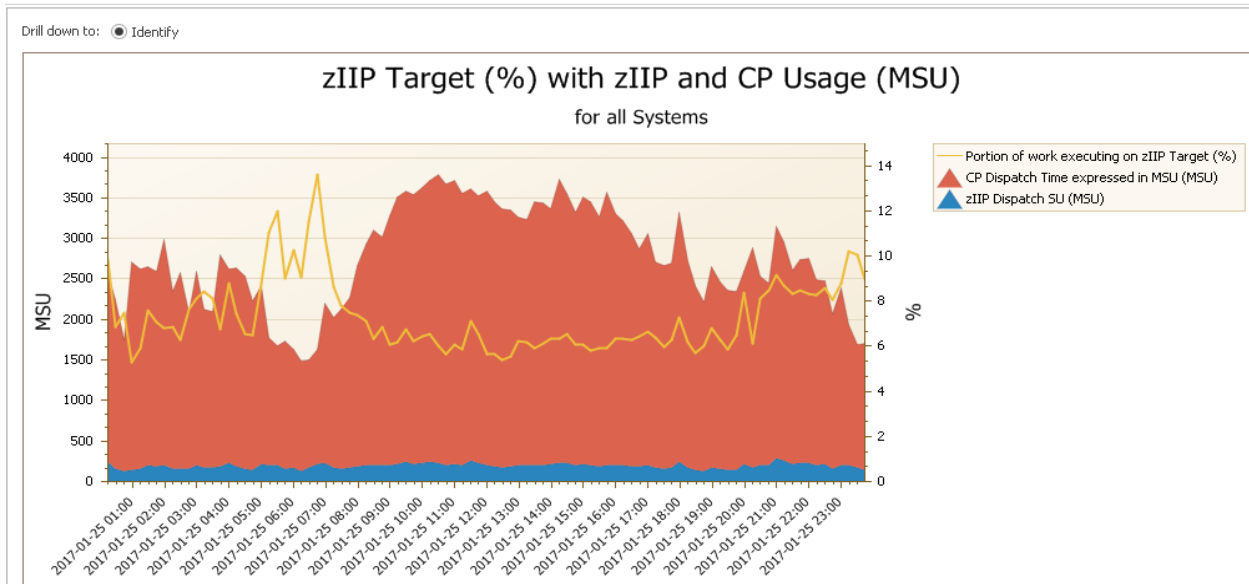
Drill down to:  Systems  Identify  Month  By date  Average by day of week  Isolate



- 100% if no 'zIIP on CP'
- Should be close to 100% during MLC peak

# How much of the work runs on zIIPs?

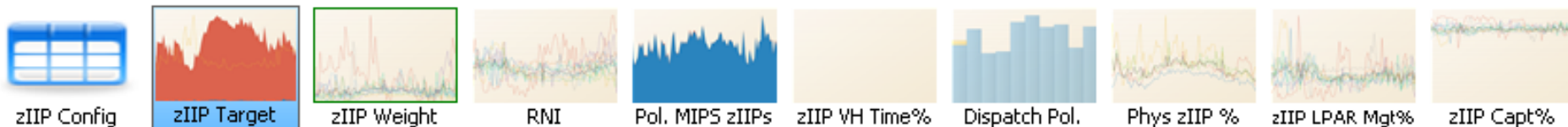
zIIP and zAAP Reporting > zIIP Management > zIIP Target



- More is better
- No particular 'good' number, workload dependent
- IBM small print has maximum value

# zIIP Management

## zIIP Management Previews



- Reports to enhance management of zIIP configuration and optimize processor cache efficiency
- Extends many concepts from GCPs, such as weights, RNI, polarity, LPAR overhead, and capture ratios
- New metric “zIIP Target %” reflects % of total processor workload executing on zIIPs

Questions?



# IntelliMagic Vision Product Scope



More than 11500 unique  
report definitions

# IntelliMagic Vision **Disk & Tape** Modules

## Disk

- Devices (74)
- Channels (73, 78)
- Storage Systems (74, 78)
- Links & Replication (78)
- FICON Dir. (74)
- Page Datasets (75)
- Data sets (42)
- Address Space (30)
- Logger (88) - June

## Replication

- XRC (42) & GDPS (105)
- EMC SRDF/A

## Tape

- SMF (14/15, 21, 30)
- BVIR for IBM TS7700
- SMF for Oracle STK
- Tape Catalogs



# IntelliMagic Vision **Systems** Module – Overview

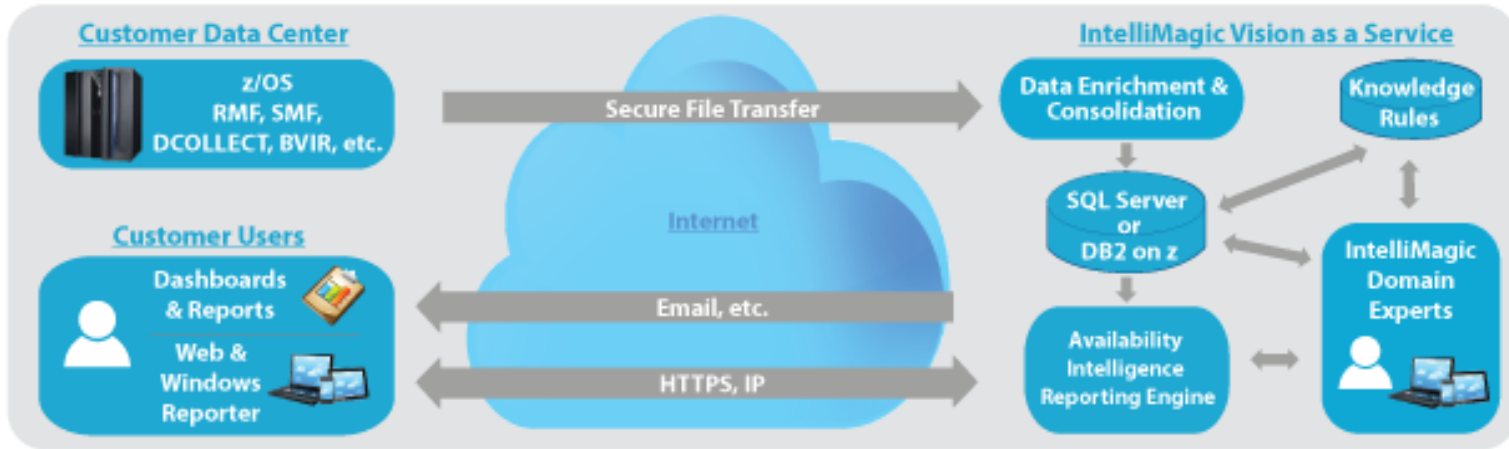
## Classic

- CEC, LPAR, 4HRA (70)
- Real Storage, Paging (71, 75)
- WLM (72)
- Channels (73)
- CF, XCF, FICON Dir. (74)
- Virtual Storage (78)
- Address Space (30)
- TCP/IP (119)
- MQ (115), (116, coming 2018)
- Logger (88)
- DB2 (100, 101) – coming soon

## Emerging

- PCIe/zEDC (74)
- LPAR Topology (99)
- Processor Cache (113)
- zIIP SMT (70)
- Transaction (72)
- SCM (Flash) (74)
- SCRT/Usage (89, 30)
- Crypto (70)
- Serialization (72.5), Enq (77)

# IntelliMagic Vision as a Service (or on premise)



- Good problem to solve with Software as a Service
- Different roles in your organization get easy access to custom intelligence
- Access to IntelliMagic experts for knowledge transfer, analysis
- Solution infrastructure is managed for you, creating more focus