Two LPARs Good, Four LPARs Better?

Anna Shugol & Martin Packer

IBM

November 2018

Session LG

Abstract

Many customers operate High Availability environments with two symmetric LPARs. But is two really the right number? Would four be better?

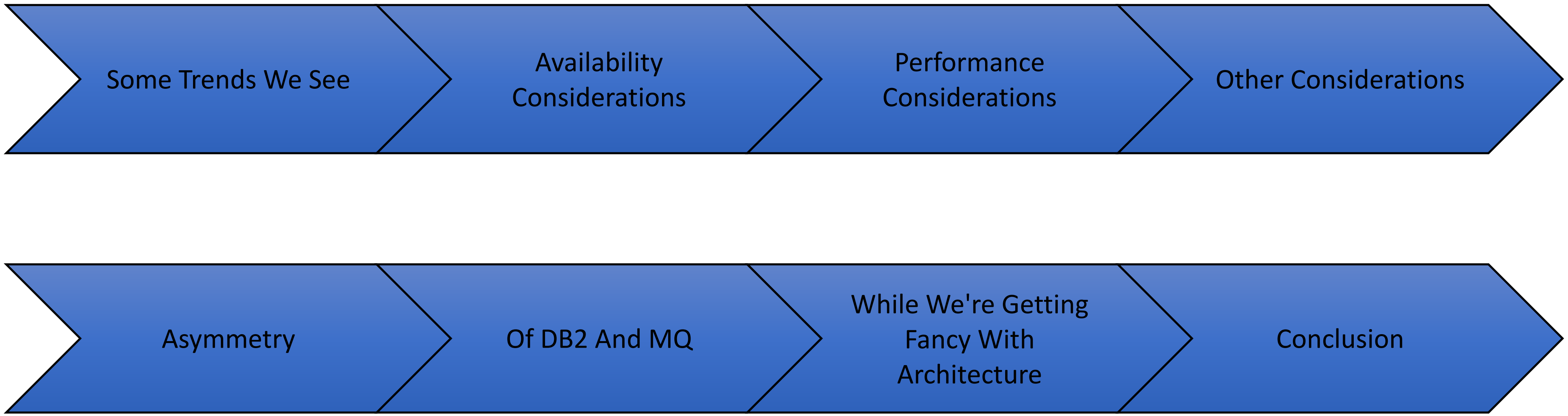
There are benefits of going to four, but there are also issues.

Quite a few customers are considering this question, so maybe you should.

This presentation explores the performance aspects of this topic, as well as some architectural ones.

Included is a look at some ways of looking at SMF data that can help you decide.

Topics



Some Trends We See

Availability Considerations

Performance Considerations

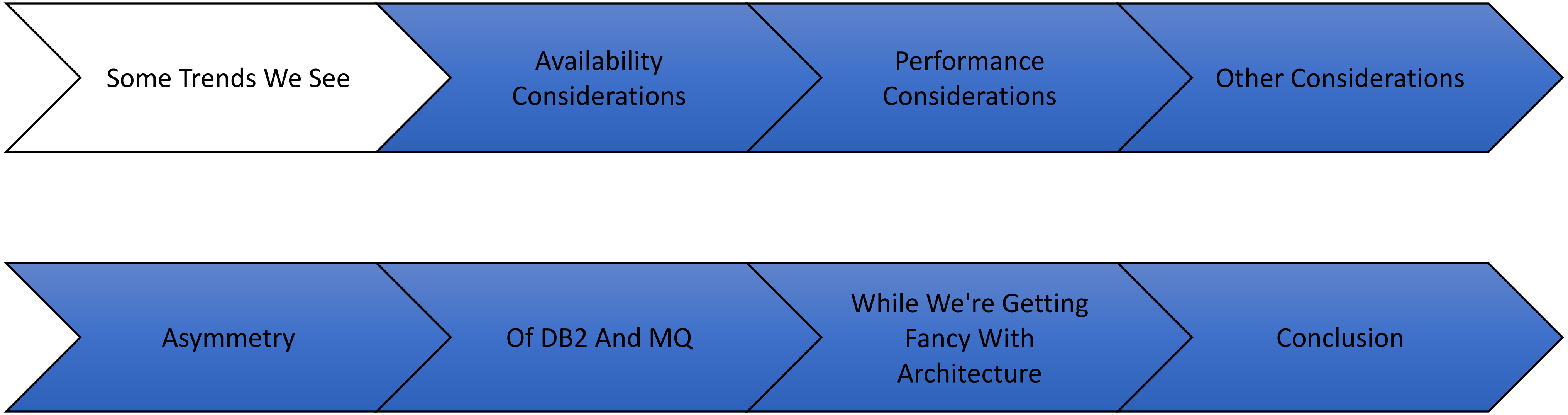
Other Considerations

Asymmetry

Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion

More LPARs Per Machine

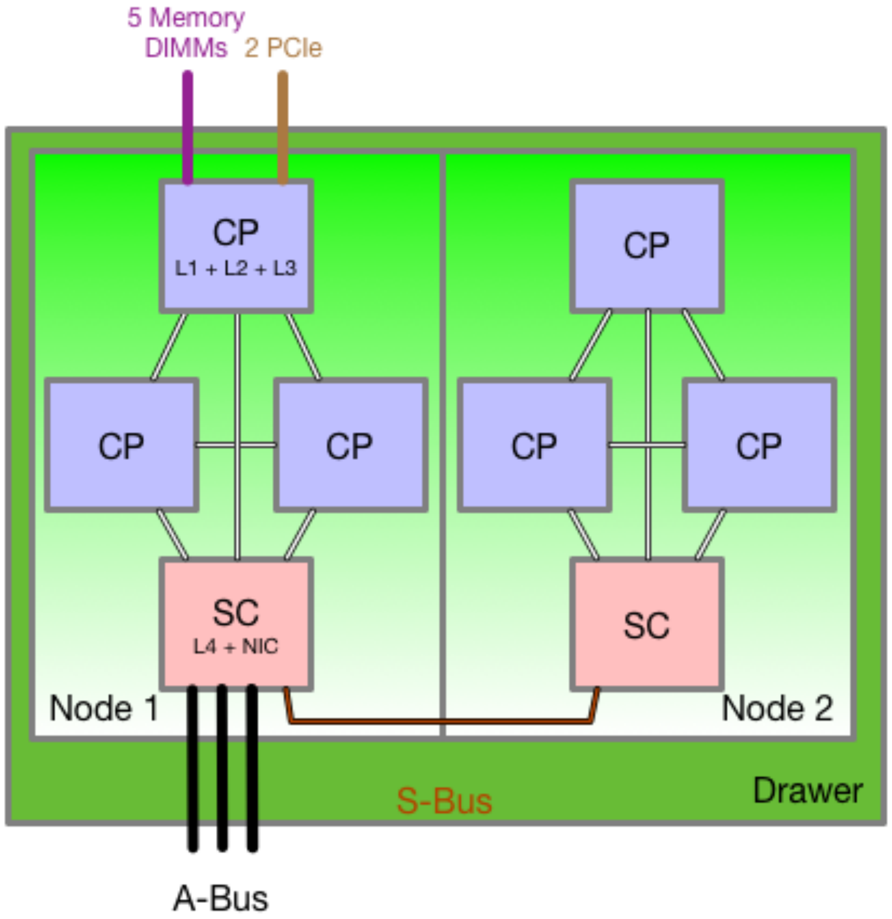
- When PR/SM was introduced 2 or 3 LPARs per machine was common
  - Now it's common to see dozens
  - Recent machines' limit is 85
- Wider diversity
  - Increasing use of Linux on Z & z/VM LPARs
  - Even z/OS LPARs look different from each other
- LPARs are bigger
  - More logical processors
  - More memory
- PR/SM's become more complex

Successive Machines Became More Complex

	z900	z990	z9	z10	z196	zEC12	z13	z14
Announcement Year	2000	2003	2005	2008	2010	2012	2015	2017
Configurable Processors	16	32	54	64	80	101	141	170
LPARs	15	30	60	60	60	60	85	85
Maximum Memory	64 GB	256 GB	512 GB	1.5 TB	3 TB	3 TB	10 TB	32 TB
LCSS	1	4	4	4	4	4	6	6

Design Drives The Performance

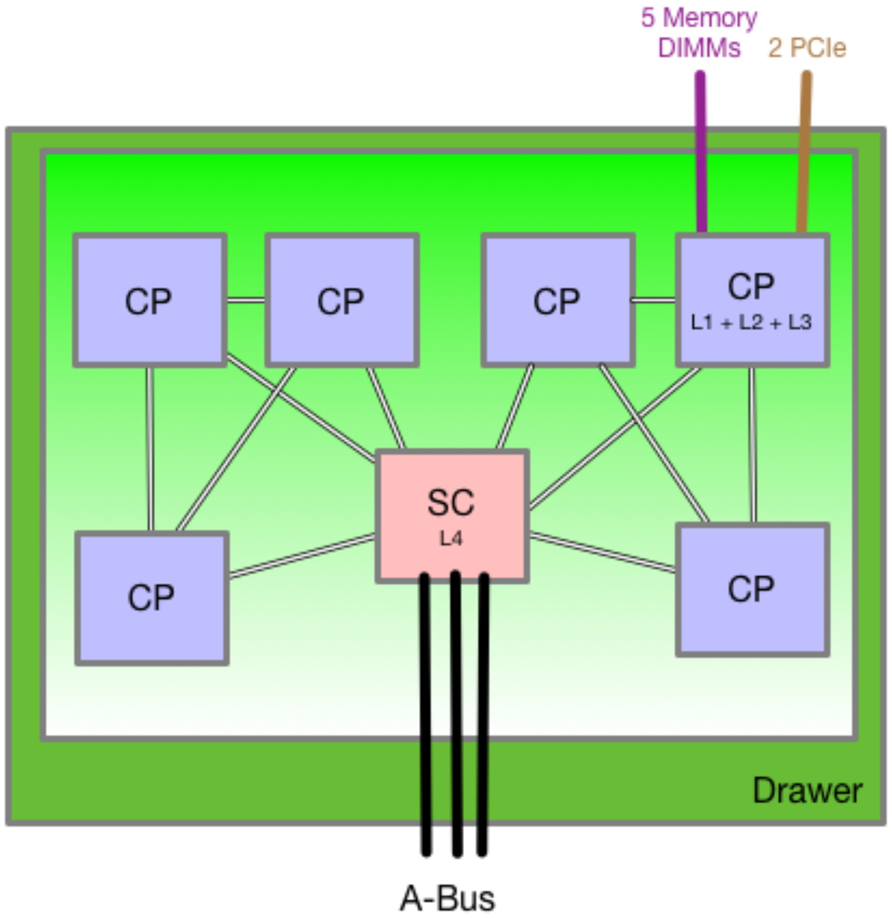
z13



8 cores per CP chip

5.0 GHz

z14



10 cores per CP chip

5.2 GHz

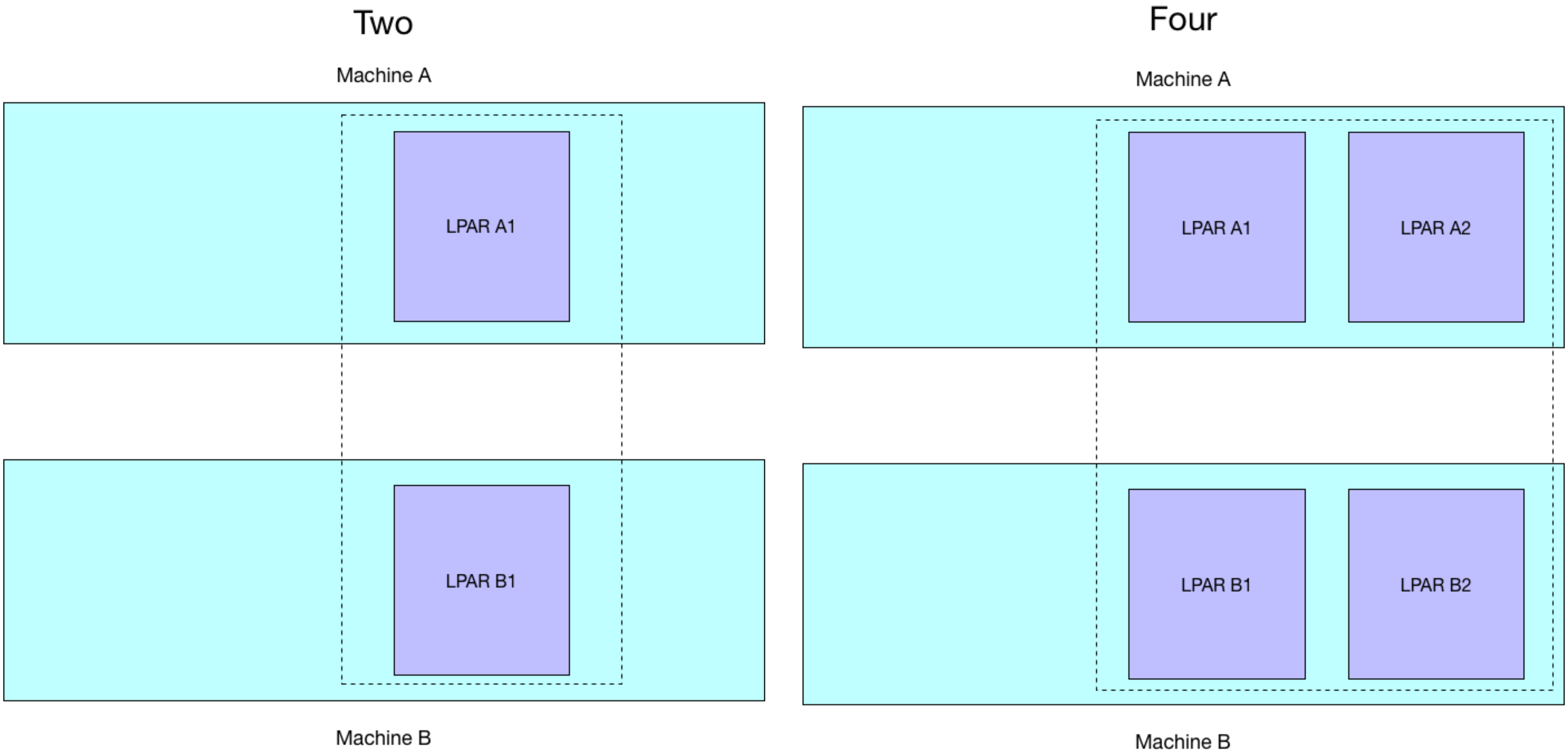
Larger L1, L2, L3 caches

Smaller L4 cache but unified

MORE LPARs Per Sysplex

- Initially diverse LPARs were brought into a sysplex
  - Driven by Parallel Sysplex Licence Charge
    - Country Multiplex Pricing might change this
    - 32-LPAR limit rarely a problem
      - But large sysplexes quite common
  - Availability benefits limited by their heterogeneity
- Some homogeneous sysplexes early on
  - Driven by the need to increase availability
- Recently more "cloning"
  - On different machines
  - On the same machine

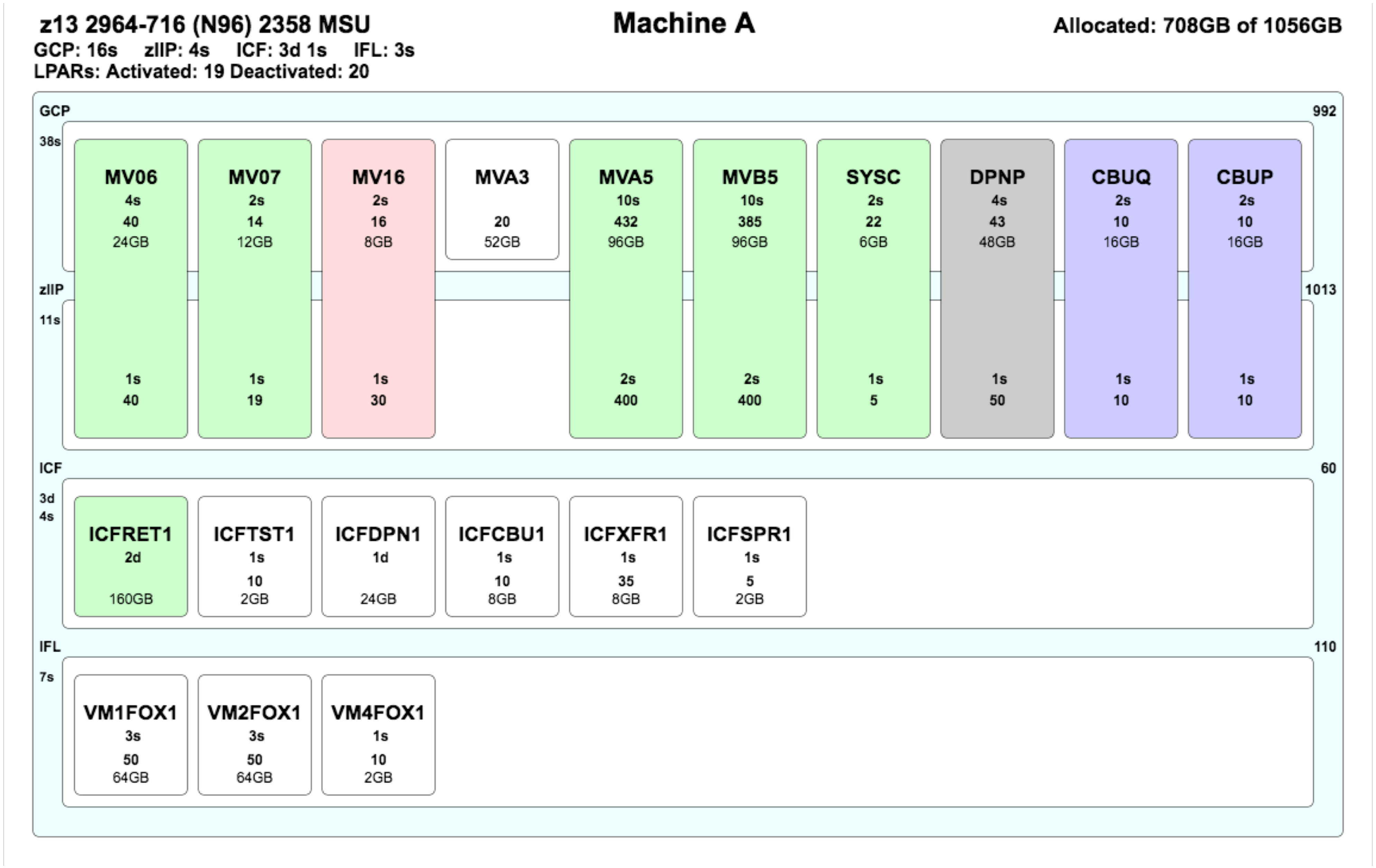
What's 2 And What's 4?



We see these things with SMF

- RMF
  - SMF Type 70 Partition Data Report describes LPARs
    - Weights and Logical Processor Definitions
    - Hiperdispatch Parking and Unparking
  - SMF 72 Workload Activity Report describes workloads
    - Including proportions of CPU used by each Service Class
  - SMF 74-2 XCF Traffic Between DB2 IRLMs, MQ queue managers, CICS Regions
    - Gives Datasharing and Queue Sharing groups
- SMF 30 Address Space describes resource usage one layer down
  - CICS regions, DB2 subsystems, MQ queue managers
  - Also topology with Usage Data Section
    - For example which MQ or DB2 a CICS region connects to
- SMF 89 yields lots of information
  - Software levels
  - DB2 subsystems and queue managers

Life Is More Complex Than That



z13 2964-716 (N96) 2358 MSU

Machine A

Allocated: 708GB of 1056GB

GCP: 16s zIIP: 4s ICF: 3d 1s IFL: 3s

LPARs: Activated: 19 Deactivated: 20

GCP 38s — 992

zIIP 11s — 1013

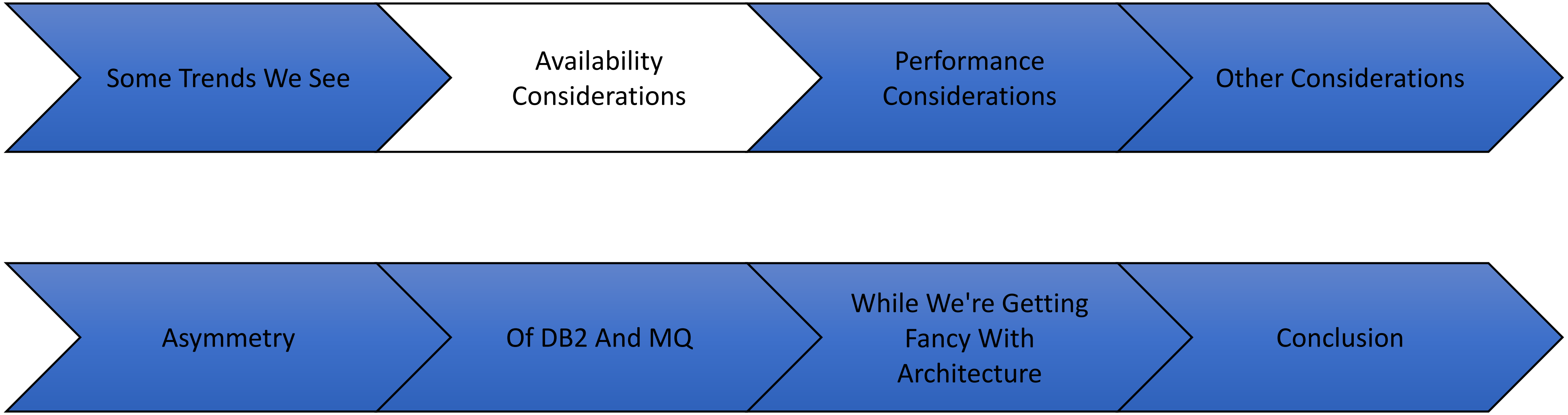
MV06	4s	40	24GB	1s	40
MV07	2s	14	12GB	1s	19
MV16	2s	16	8GB	1s	30
MVA3		20	52GB		
MVA5	10s	432	96GB	2s	400
MVB5	10s	385	96GB	2s	400
SYSC	2s	22	6GB	1s	5
DPNP	4s	43	48GB	1s	50
CBUQ	2s	10	16GB	1s	10
CBUP	2s	10	16GB	1s	10

ICF 3d 4s — 60

ICFRET1	2d		160GB
ICFTST1	1s	10	2GB
ICFDPN1	1d		24GB
ICFCBU1	1s	10	8GB
ICFXFR1	1s	35	8GB
ICFSPR1	1s	5	2GB

IFL 7s — 110

VM1FOX1	3s	50	64GB
VM2FOX1	3s	50	64GB
VM4FOX1	1s	10	2GB



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

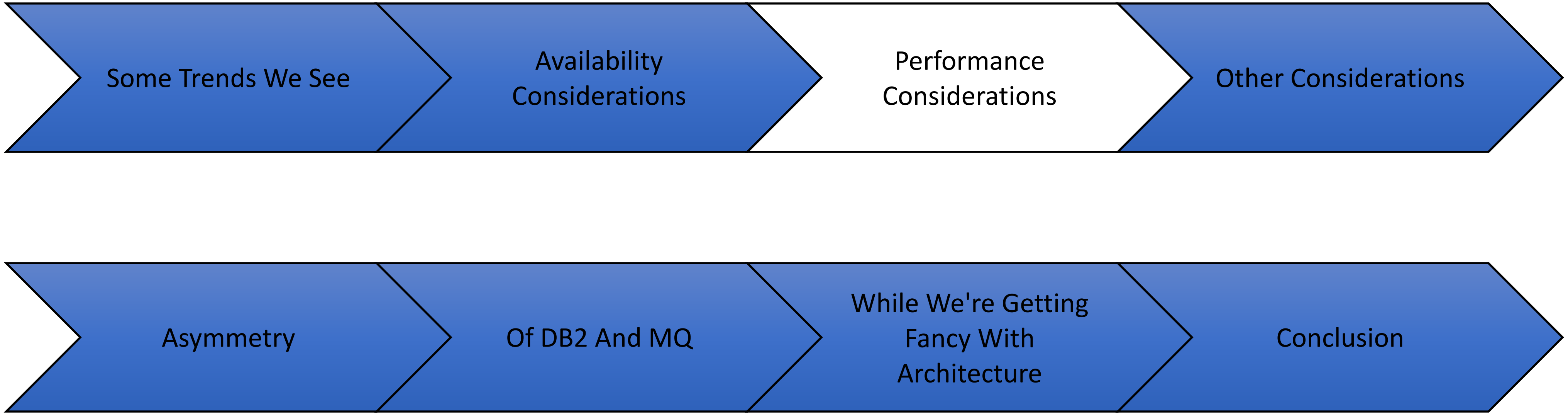
Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion

More LPARs More Resilient

- Four survive single member outage better
  - With two LPARs what brought one down might bring the other down
  - With four the workload can be spread
    - Example: DB2 Virtual Storage
- Heterogeneous LPARs isolate individual services
  - Some might be more troublesome than others
  - Some services more important than others
- Resilience not just about up vs down
  - Connectivity
  - Performance



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

Of DB2 And MQ

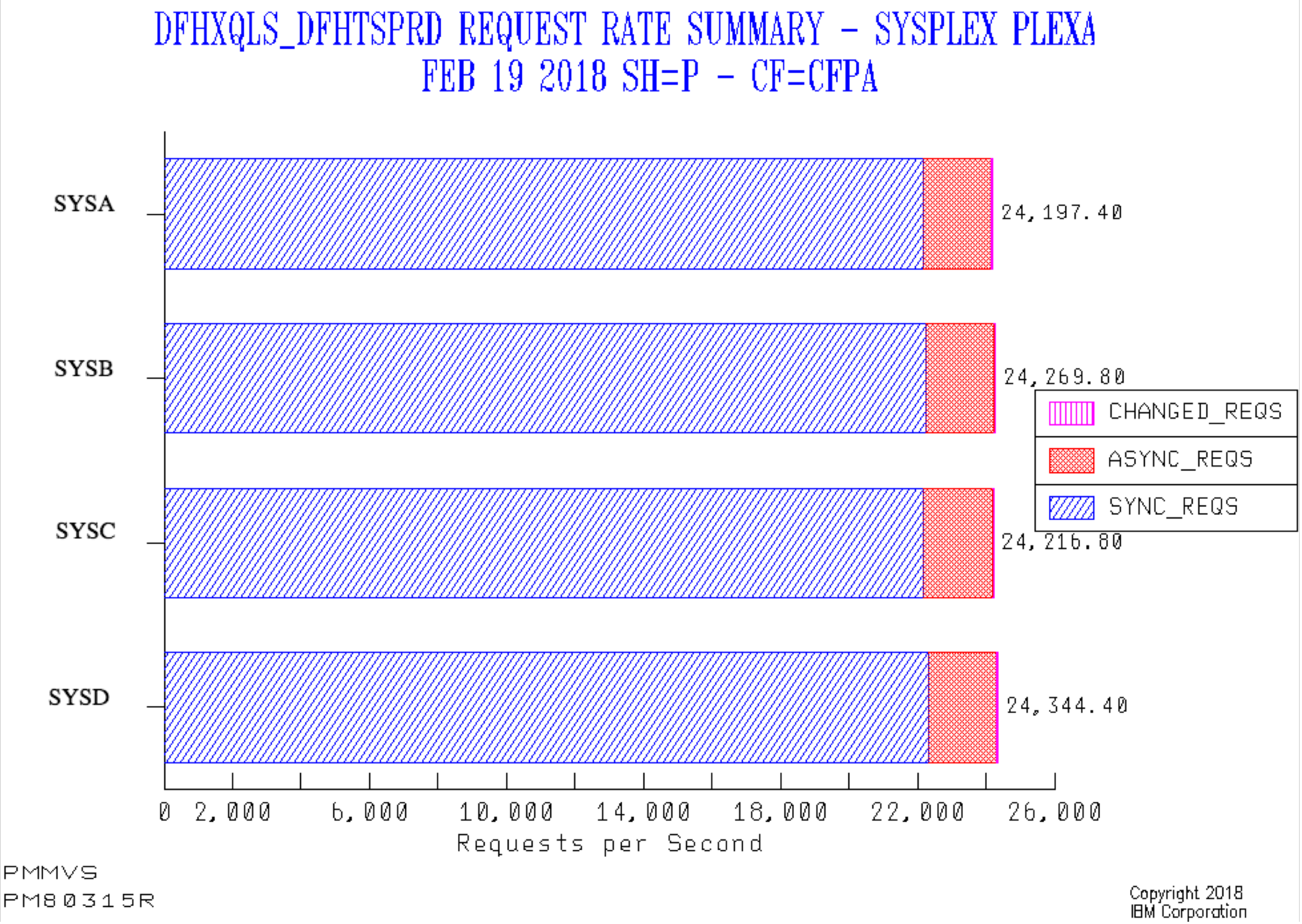
While We're Getting Fancy With Architecture

Conclusion

Greater overhead

- Sysplex
    - Much of the cost going 1-way to 2-way
- Datasharing
    - More XCF communication
    - Cloned DB2 jobs spread across more individual members
        - More Inter-DB2 read/write interest
- More sets of buffer pools means fewer hits per GB
    - Probably more database I/Os overall
- PR/SM overhead increases
    - 70-1 PHYSICAL
        - Only part of it
        - Cache effects not recorded
- Careful design mitigates these effects somewhat

Transactions Calling Others Create More Cross-System Traffic: 75% v 50%



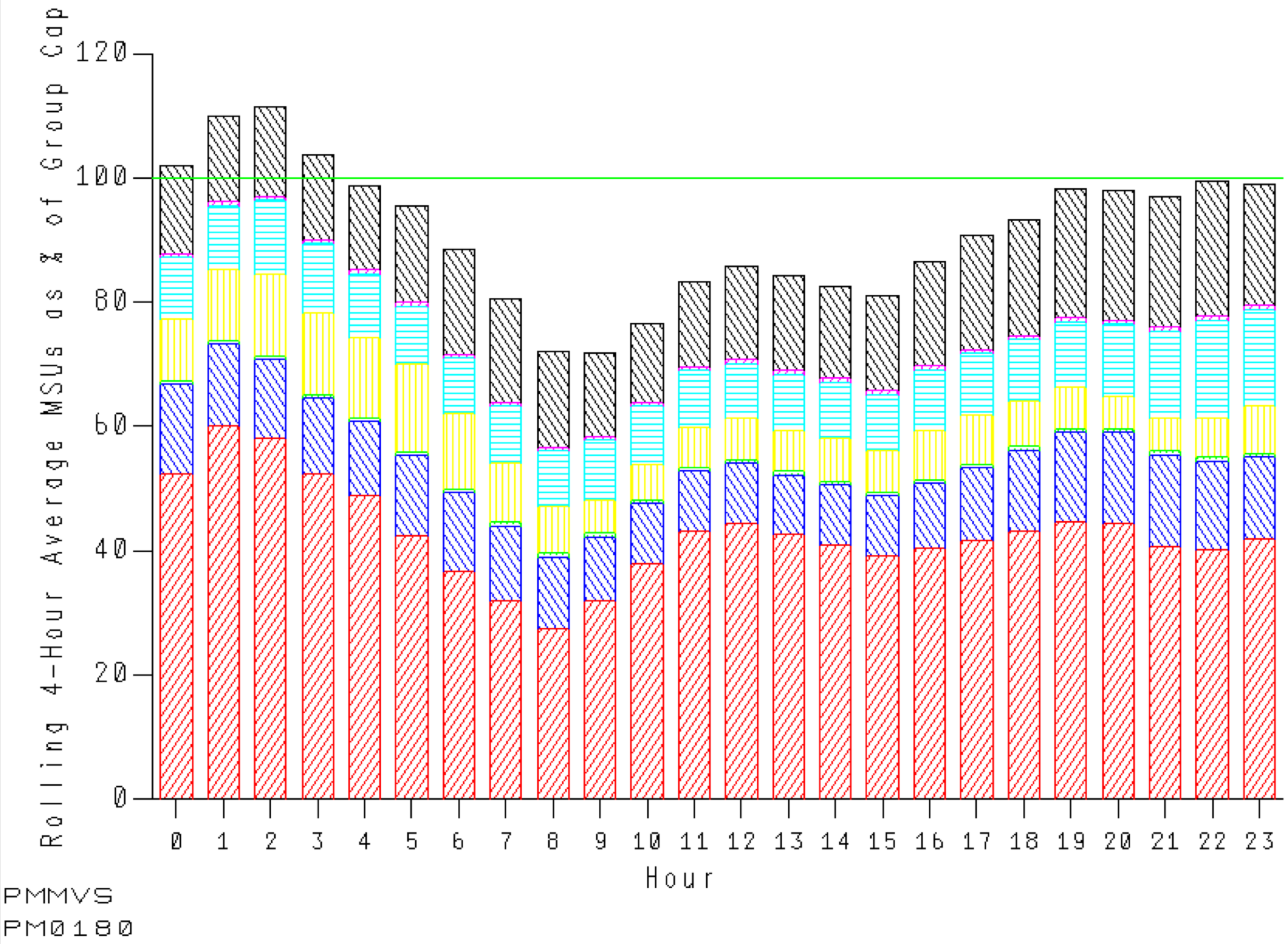
LPAR Design Pragmatics

- Processor topology impacts on LPAR size and quantity
- Hiperdispatch significantly affects LPAR design
- PR/SM tries to allocate PUs and memory in a single CPC drawer
    - If not possible then the same set of CPC drawers
- z14:
    - Up to 85 LPARs
    - Up to 16TB (OS dependent) of memory per LPAR
- Single-Drawer LPARs a reasonable design
- Other Considerations Limit LPAR size
    - Common Area virtual storage
        - e.g CSA for IMS
        - Can limit 24- and 31-bit Private too much
    - Below 16MB real storage
    - Historical example: "No LPAR shall be more than 1000 MIPs"

Group Capacity More Complex Than Defined Capacity

- For many customers softcapping is a fact of life
- Cloning an LPAR leads to managing two LPARs' Rolling 4 Hour Average
    - Group Capacity indicated
    - Defined Capacity still usable

Heterogeneous Group - Who To Protect?



Transaction Routing - General

- Transaction routing considerations become more important and complex
- Monitor transaction outcomes
- WLM needs to be properly set up
    - IRD interacts with routing decisions
    - Other routing mechanisms rely on goal attainment

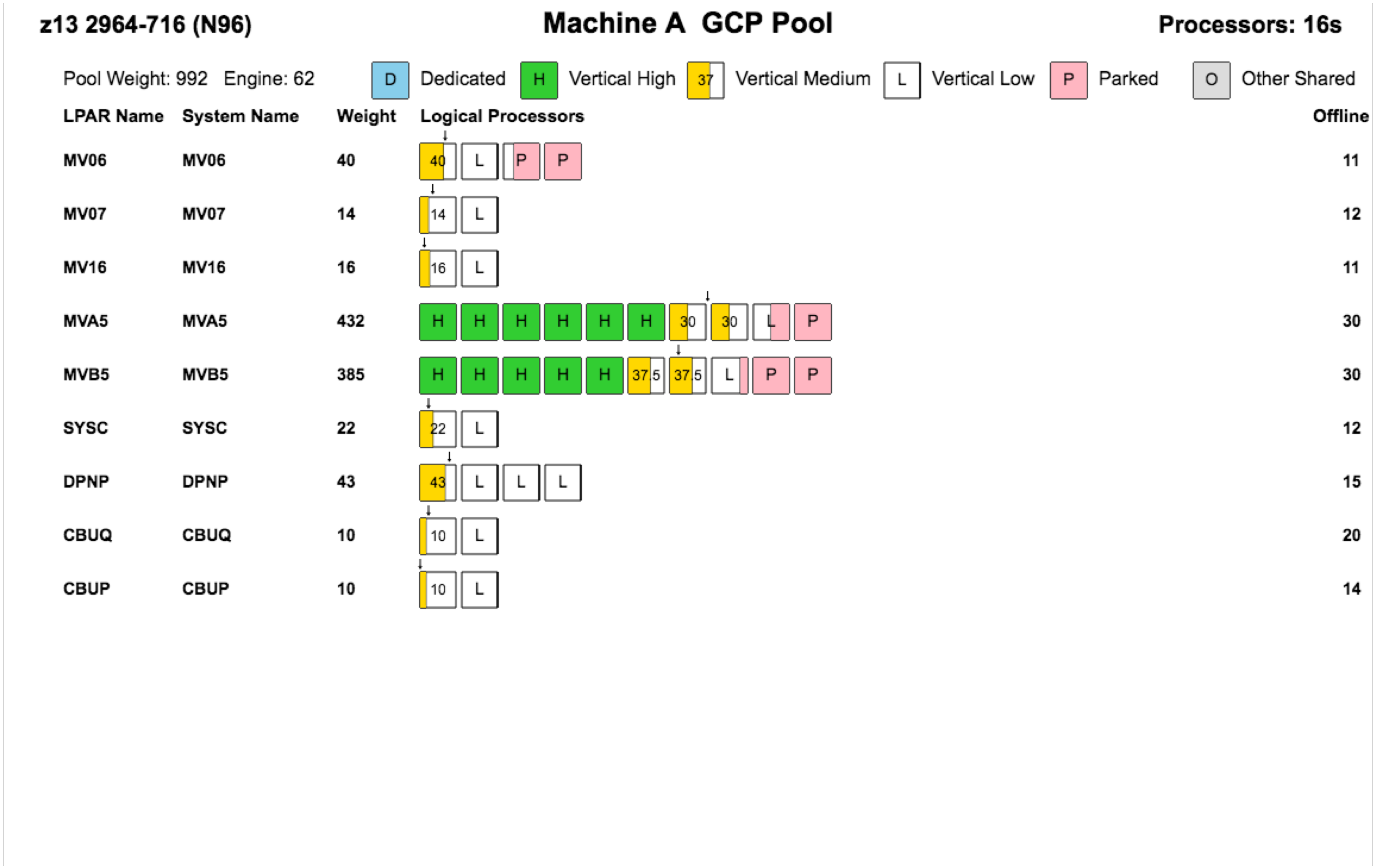
Transaction Routing - Middleware

- CPSM routes CICS transactions
  - **Queue Algorithm** - Queue length relative to e.g. MAXTASKS
  - **Goal-oriented** - Likelihood of meeting z/OS WLM Average Response Time goal
  - Cogniscent of transaction affinities
  - CICS transaction-level instrumentation is SMF 110 Monitor Trace
- DDF
  - SMF 30 Enclave CPU, Transaction Rate, Response Times
  - SMF 101 more detailed transaction statistics
    - “Sloshing” is a condition where routing is uneven
  - Server Health is important
    - Reported by DB2 to WLM

HiperDispatch

- HiperDispatch (HD) was introduced for z/OS and subsequently for z/VM
- HD - exploitation of PR/SM Vertical CPU management
- HiperDispatch creates affinity nodes and tries to redispatch the workload on this nodes
- IBM LSPR tables assume HD=YES
- HiperDispatch corrals engines effectively
    - Careful with unparked Vertical Lows doing work
- Weight shifting with IRD affects HiperDispatch

HiperDispatch Parks Engines



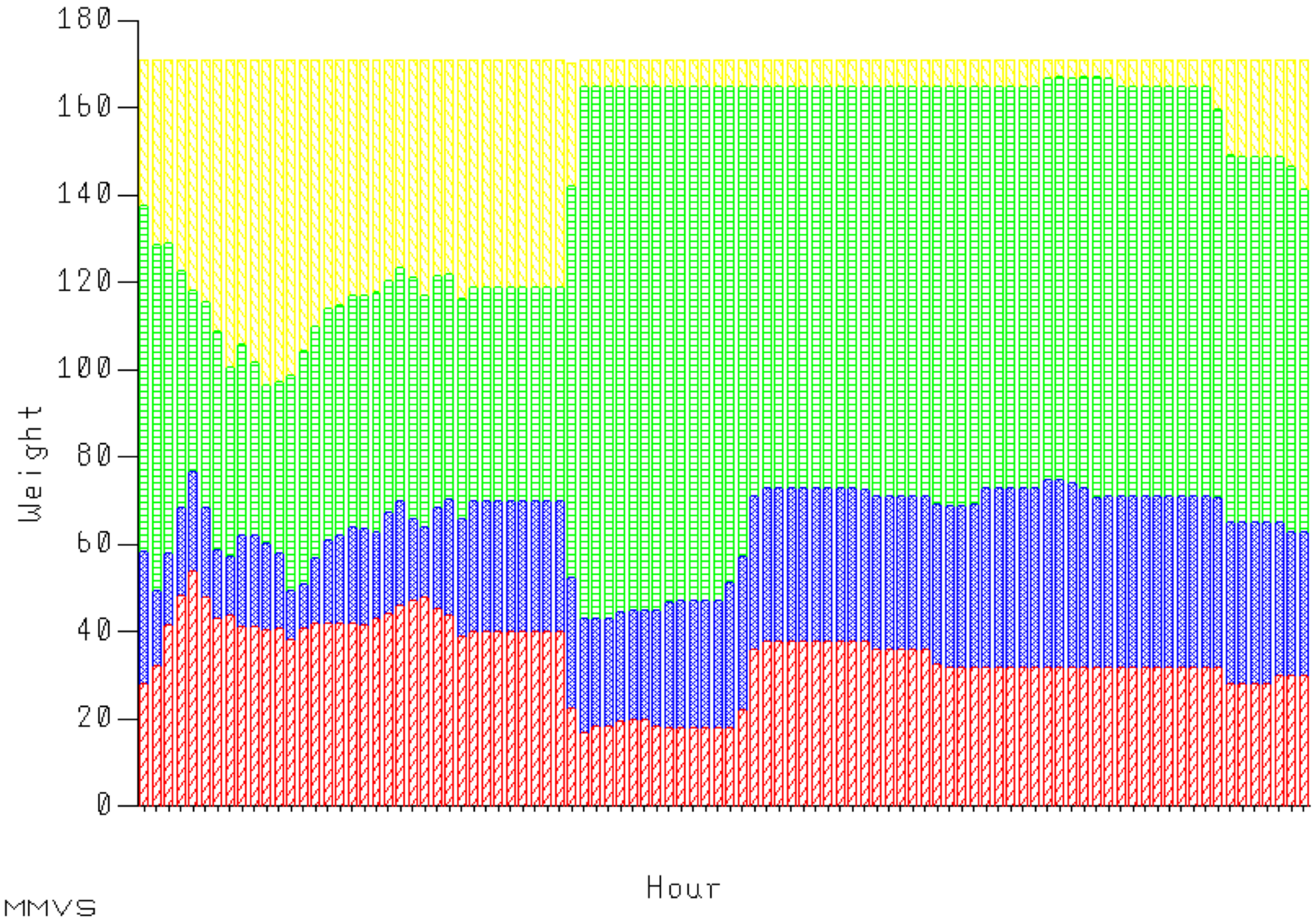
z13 2964-716 (N96) **Machine A GCP Pool** **Processors: 16s**

Pool Weight: 992 Engine: 62

D Dedicated H Vertical High 37 Vertical Medium L Vertical Low P Parked O Other Shared

LPAR Name	System Name	Weight	Logical Processors	Offline
MV06	MV06	40	40 L P P	11
MV07	MV07	14	14 L	12
MV16	MV16	16	16 L	11
MVA5	MVA5	432	H H H H H H 30 30 L P	30
MVB5	MVB5	385	H H H H H 37.5 37.5 L P P	30
SYSC	SYSC	22	22 L	12
DPNP	DPNP	43	43 L L L	15
CBUQ	CBUQ	10	10 L	20
CBUP	CBUP	10	10 L	14

IRD Shifts Weights - Affecting HiperDispatch



PMMVS
PM0980

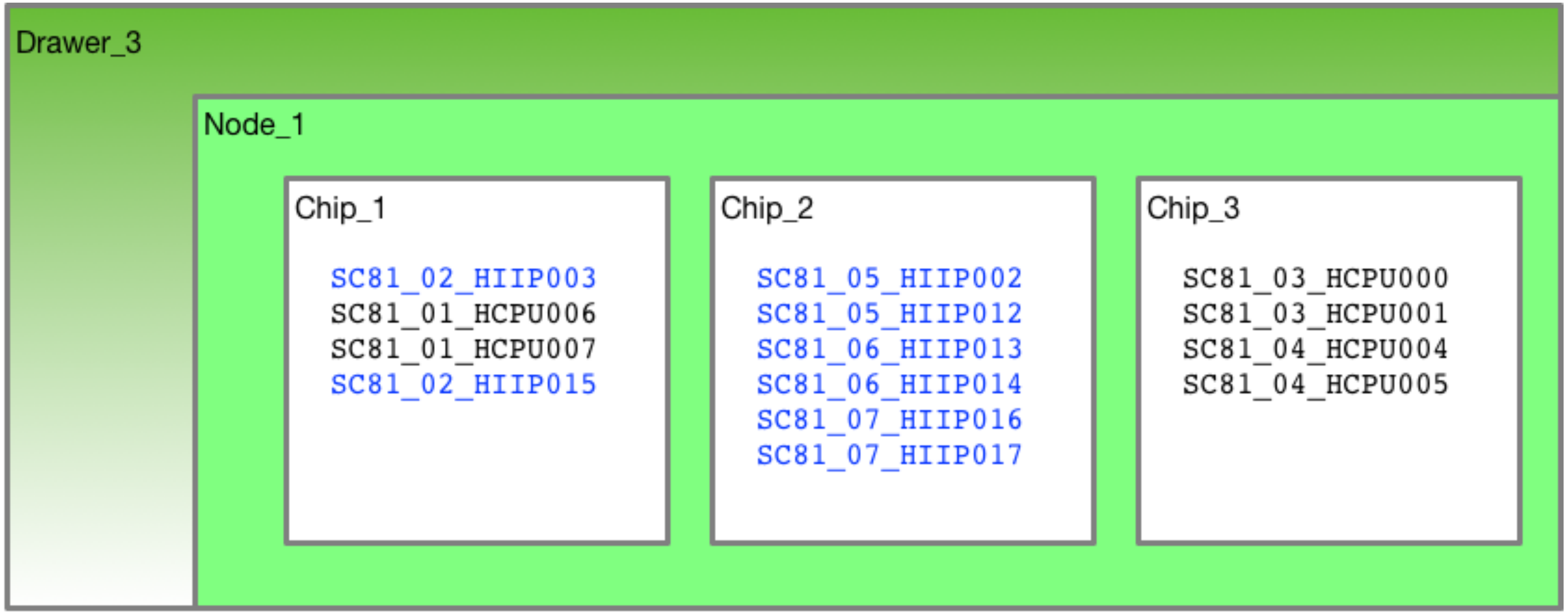
HiperDispatch Recent Enhancements

- z14 and z/OS 2.3 Enhancements:
  - Hiperdispatch workload balancing algorithm considers processor topology and takes it into account for:
    - memory consumers - creates memory affinity
    - large TCBs that are split across address spaces

HiperDispatch Instrumentation

- RMF SMF 70-1 CPU Activity
    - Individual engine utilisations
    - Vertical Polarisation / Weights & Parking information for Logical Processors
    - Behaviour of I/O interrupt handling at Logical Processor level
- WLM SMF 99-14
    - WLM Topology Report
    - Location of logical processors in the hardware
    - Interval-based to reflect dynamic nature
- Hardware Instrumentation Services (HIS) SMF 113
    - Cache Effectiveness and Cycles Per Instruction (CPI)
- Above helps analysis of the LPAR design & its implications
- Include **individual processor** level in your analysis

WLM Topology Report Fragment - From SMF 99 Subtype 14



Drawer_3

Node_1

Chip_1

SC81_02_HIIP003
SC81_01_HCPU006
SC81_01_HCPU007
SC81_02_HIIP015

Chip_2

SC81_05_HIIP002
SC81_05_HIIP012
SC81_06_HIIP013
SC81_06_HIIP014
SC81_07_HIIP016
SC81_07_HIIP017

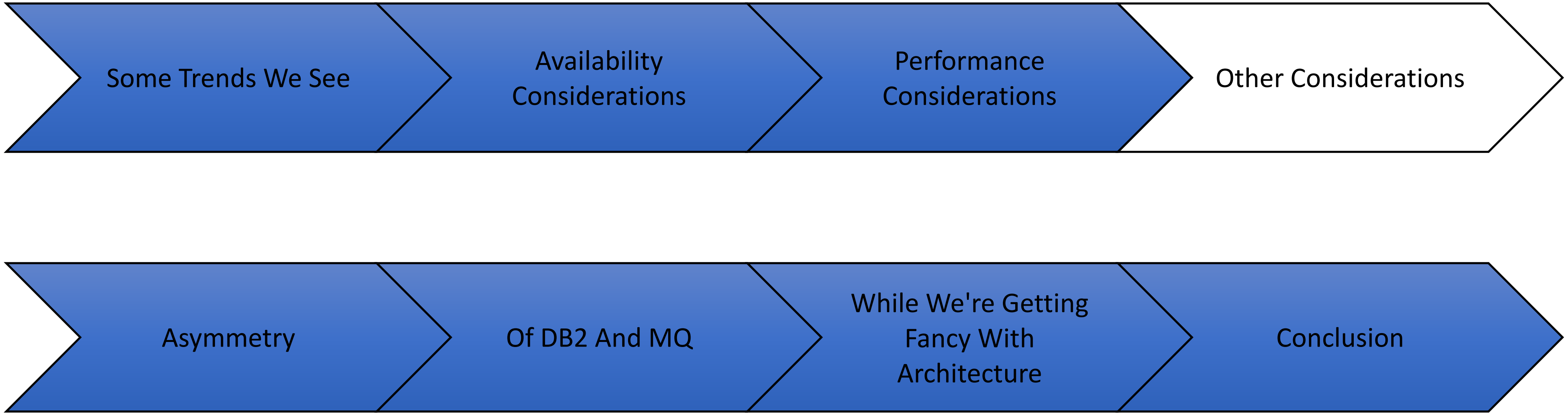
Chip_3

SC81_03_HCPU000
SC81_03_HCPU001
SC81_04_HCPU004
SC81_04_HCPU005

Logical Processor Decoding: ssss_NN_vtttnnn where:

ssss	SMF ID	eg	SC81
NN	Affinity Node Number	eg	02
v	Polarization	eg	H
ttt	Processor Type	eg	IIP
nnn	Processor Number	eg	003

Example: SC81_02_HIIP003



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

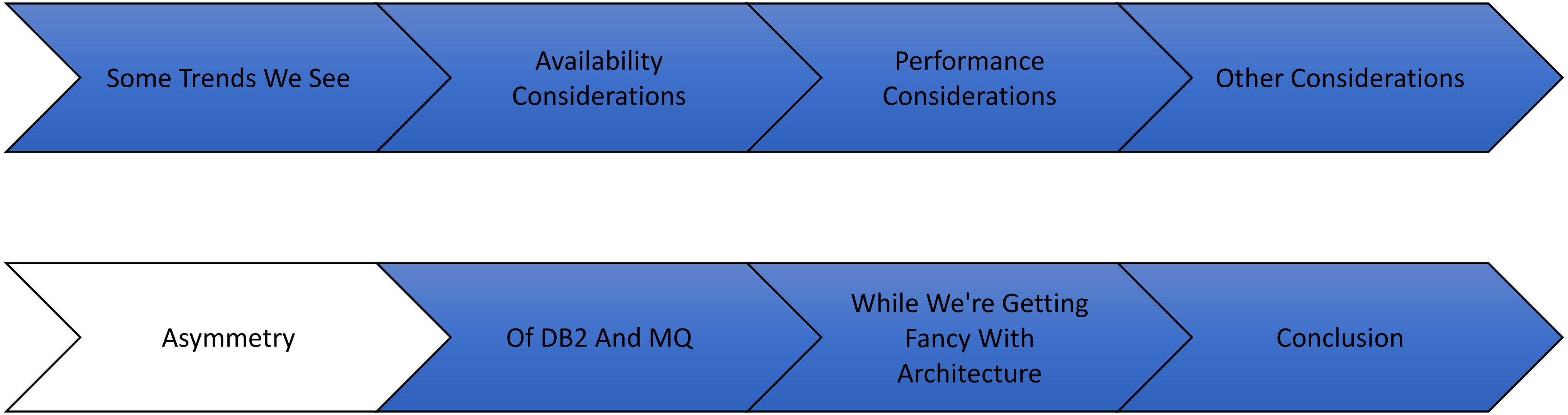
Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion

Operational

- LPAR naming convention is important to avoid mistakes (eg reIPLing wrong LPAR)
    - Same applies to the SW subsystems (Db2 for z/OS, CICS etc)
- The devops complexity grows with the number of LPARs
- All operational processes must be tested and automated



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion

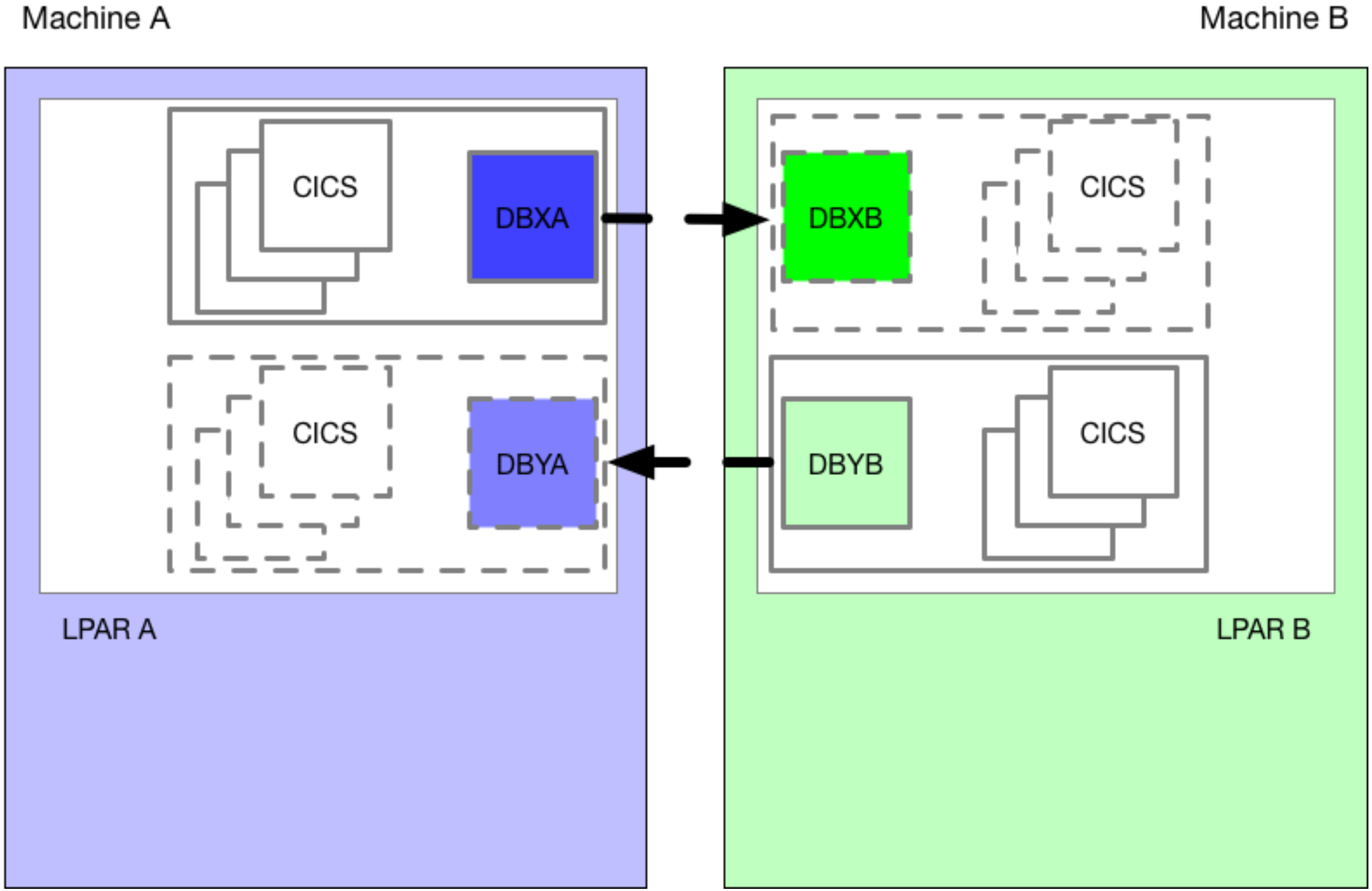
Why Not Three?

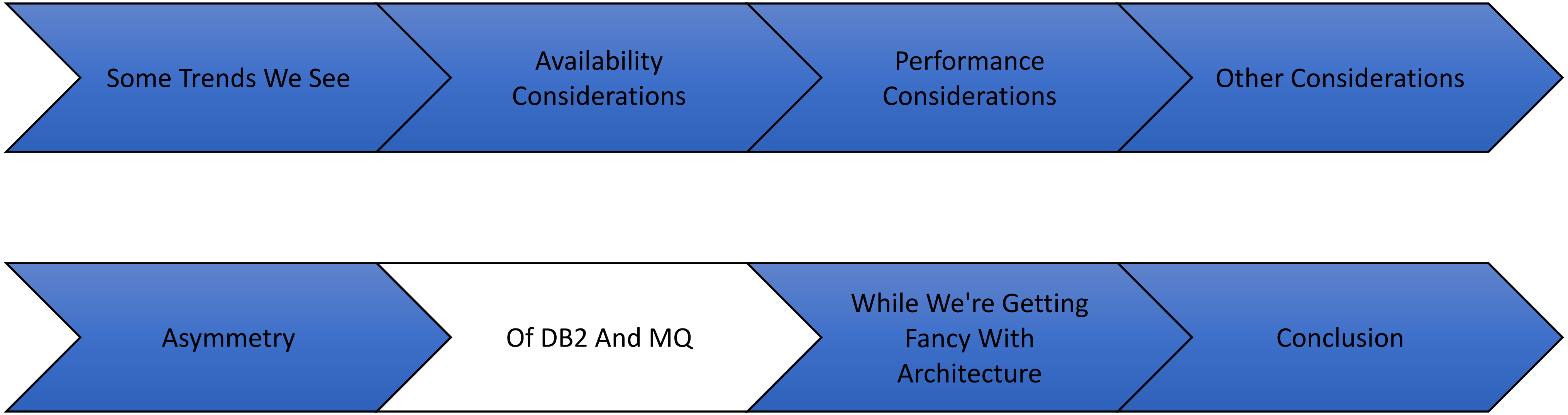
- Availability benefits somewhere between two and four
    - Likewise performance
- Imbalance
    - Unless you have a third machine

Possibly Already Asymmetric

- Asymmetry inhibits availability
  - Takeover requires a peer to actually exist
- If dissimilar LPARs are brought into a sysplex complete symmetry is unlikely
  - Workload affinities are hard to eradicate
    - For example, DB2 subsystems have specific roles
- Cloning often used to create symmetry
  - CICS regions, DB2 subsystems, MQ queue managers
  - Sometimes names give the game away
    - e.g. ATMCICS on SYSA, ATMCICS2 on SYSB
  - Cloning in both directions
    - Can lead to plethoration

To Get The Availability Benefits Need To Replicate





Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion

Proliferation Of Members And Queue Managers

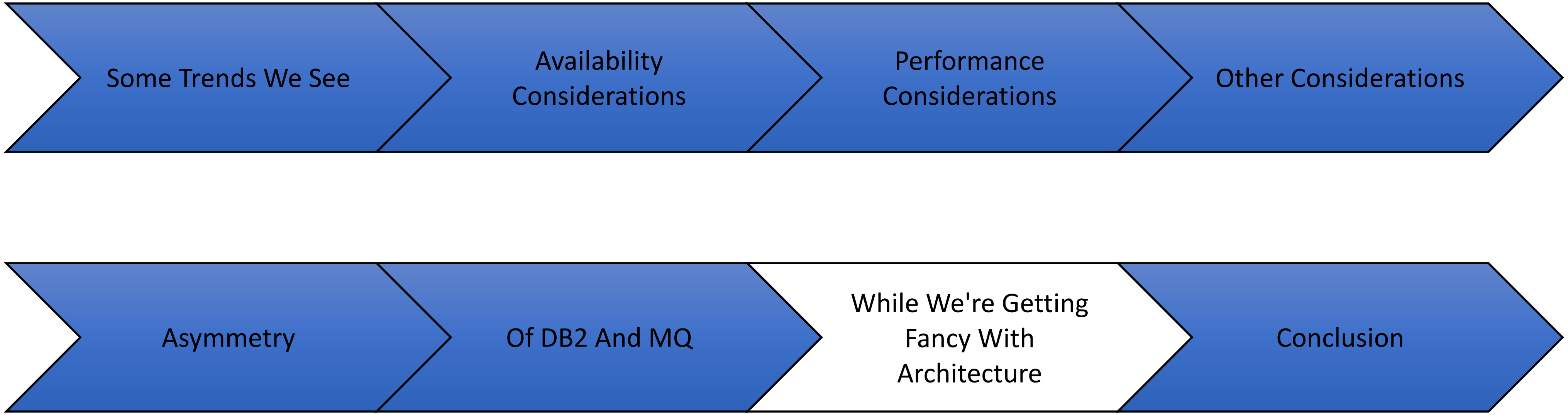
- SMF 30 detecting DB2 / MQ
  - And their roles
- One customer starts with 6 DB2s on 2 LPARs
  - Going to 4 LPARs don't want to go 12 DB2s
    - Some merging required
- How Did We Get To So Many DB2s and MQs?
  - Going from 2 to 4 LPARs a good chance to consolidate
- MQ rarely has scalability limitations
- CICS proliferation generally beneficial
  - Particularly if cloning alleviates QR TCB constraint
    - Less of an issue now with Threadsafe

DB2 Scalability

- Some inhibitors to merging DB2 subsystems have been
  - Virtual Storage
    - **Massively** relieved each release up to Version 10
    - IFCID 225 in DB2 Statistics Trace documents usage well
  - Logging Bandwidth
    - zHyperWrite should help duplexed Active Log case
  - Buffer pool scalability
  - Prefetch and Deferred Write Engines
    - Limits are 600 each for Prefetch and Deferred Write
  - Data and application ownership
    - "This organisation owns this DB2"
    - Each SAP application has its own data sharing group

Roles of DB2 Subsystems

- Definitive view from DB2 Accounting Trace
    - Lots of SMF 101 data
    - Detailed view of clients and the CPU used in DB2
- First Pass / "Lighter Touch" view from SMF 30
    - Usage Data Section shows which DB2 each CICS region connects to
        - Likewise batch jobs, etc
    - No view of in-DB2 CPU available from SMF 30
    - DDF detectable from Enclave fields
        - Transaction rates and CPU
        - No information on client machines
- Even lighter "Zeroth Touch" view from SMF 89
    - NO89 in DSNZPARM prevents DB2 CPU showing
- Most of the above points apply equally to MQ
    - NO89 not a problem for MQ



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

Of DB2 And MQ

While We're Getting Fancy With Architecture

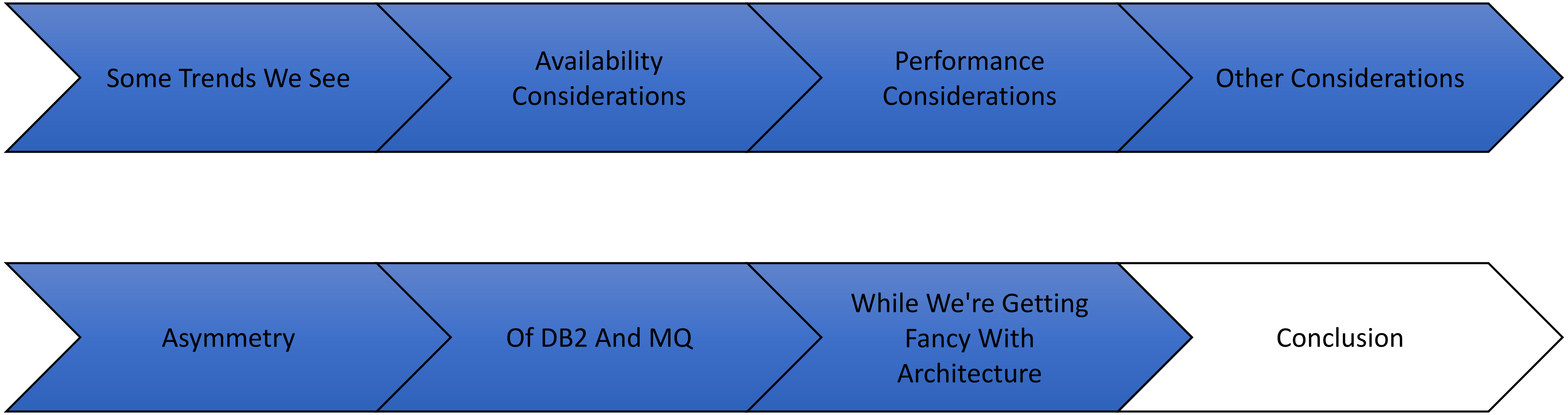
Conclusion

What About One LPAR?

- You’d probably never go back there
    - Either two or four should have much better resilience
- One is simplest to operate
    - But not much simpler than if already two- or four-up
- Country Multiplex Pricing might enable single-LPAR
    - But only with growth where Sysplex wasn't **technically** wanted

How About Four Machines?

- Not much availability benefit over two machines
    - Two LPARs per machine in a sysplex is usually fine
    - True for **unplanned** outages
        - **Planned** outages are a different story
- Environmentals probably worse
- Less opportunity to share physical links
    - Disk and tape channels
    - Coupling Facility links
- SMC-D & HiperSockets might have to be replaced by physical links



Some Trends We See

Availability Considerations

Performance Considerations

Other Considerations

Asymmetry

Of DB2 And MQ

While We're Getting Fancy With Architecture

Conclusion

Moving To More LPARs In A Sysplex Is Common

- Generally for resiliency
    - Can be for efficiency
- Moving to more LPARs needs care
    - Especially to get the performance right
- Architecture is important
    - Performance people need to engage with it
    - Performance people can use instrumentation to provide insight
- Review your LPAR design with each new processor generation

We want your feedback!

- Please submit your feedback online at ….
    - ➢http://conferences.gse.org.uk/2018/feedback/LG

- Paper feedback forms are also available from the Chair person

- This session is LG