IBM Systems Group

Introduction to Parallel Sysplex

Created By: Angelo Corridori
Presented By: Riaz Ahmad
IBM Washington Systems Center
Gaithersburg, Maryland

Trademarks

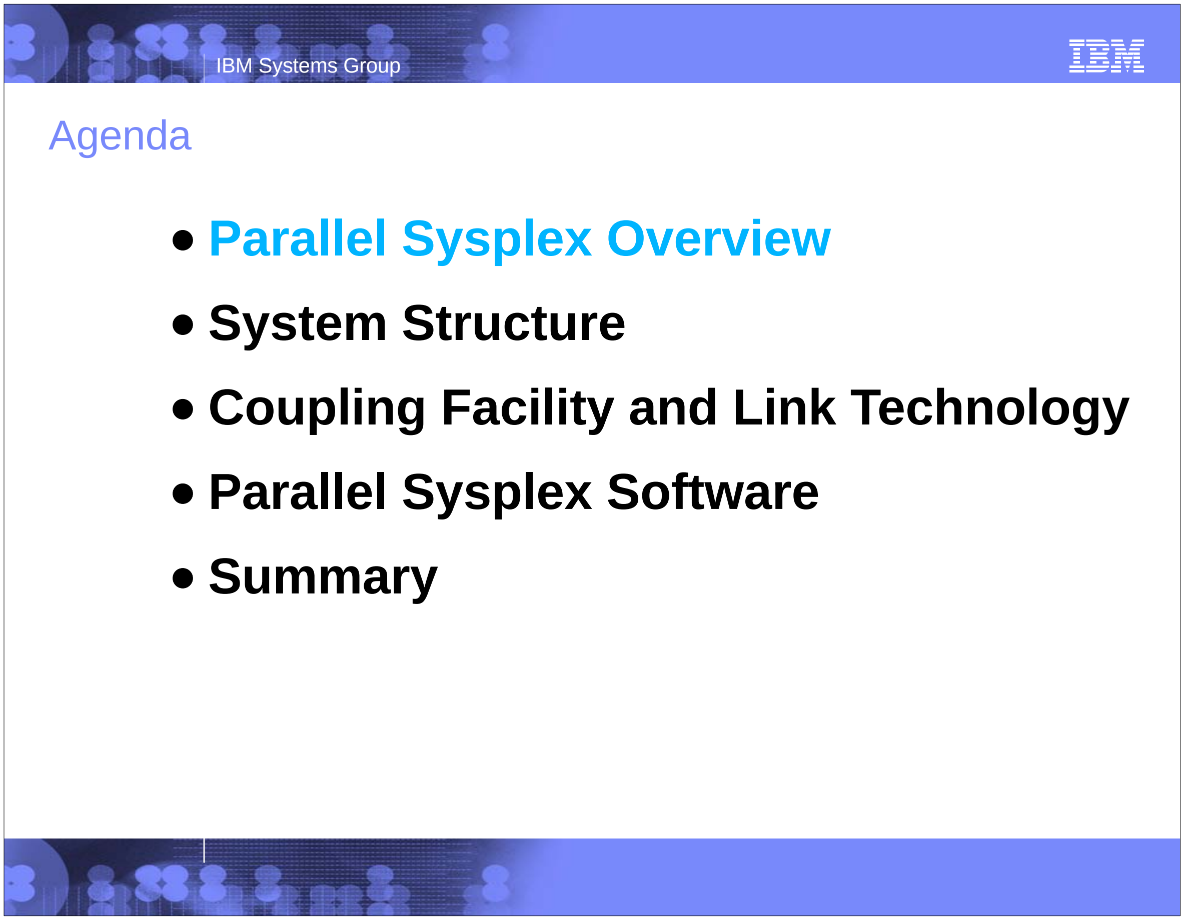
The following are trademarks of International Business Machines Corporation.

ACF/VTAM
AD/Cycle
ADSM
Advanced Function Printing
AFP
AIX*
AIX/ESA
AOEXPERT/MVS
Automated Operations Expert/MVS
CICS/ESA
DataHub
DATABASE 2
DataTrade
DB2*
DFDSM
DFSMS
DFSMS/MVS
DFSMdfp
DFSMSdss
DFSMShsm
DFSMSrmm
Distributed Relational Database Architecture
DRDA
Enterprise Systems Architecture/370
Enterprise Systems Architecture/390
Enterprise System/3090
Enterprise System/4381
Enterprise System/9000
Enterprise Systems Connection Architecture
ES/3090
ES/4381
ES/9000
ESA/370
ESA/390
ESCON
FASTService*
FlowMark
Hardware Configuration Definition
Hiperbatch
Hipersorting*
Hiperspace
IBM*
IBM S/390 Parallel Enterprise Server
IBM S/390 Parallel Enterprise Server - Generation 3
IMS/ESA
LANRES
Micro Channel*
MQ Series
MVS/DFP
MVS/ESA
NetView*
NQS/MVS
OPC
Open Blueprint
OpenEdition*
OSA
OSA 1
OSA 2
OS/2*
OS/390
OS/400*
Parallel Sysplex
Power Prestige
PR/SM
PS/2*
Processor Resource/Systems Manager
RISC System/6000
S/360
S/370
S/390
SAA
SAP R3
Sysplex Timer
System/370
System/390
Systems Application Architecture*
SystemView
VM/ESA
VSE/ESA
VTAM
3090

Note: **Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.**

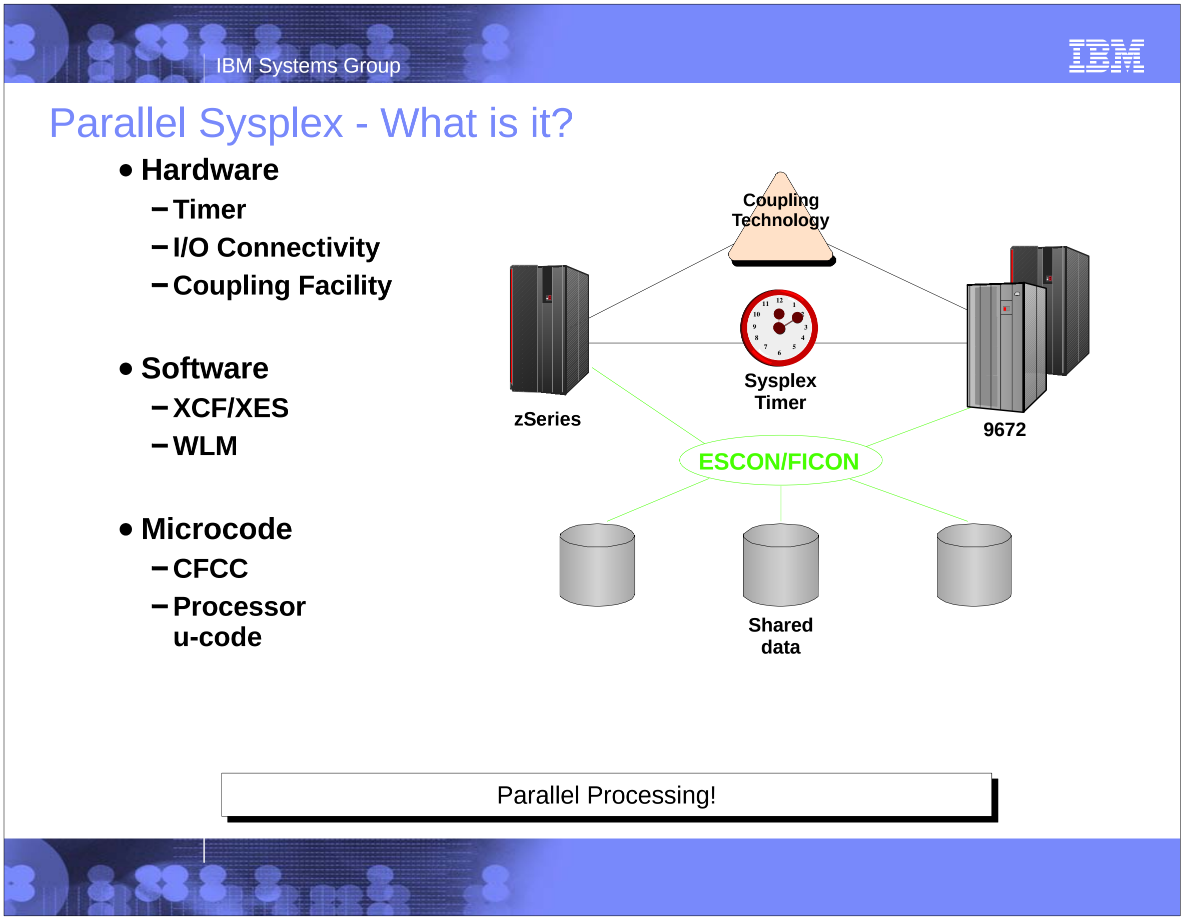
Actual performance and environmental costs will vary depending on individual customer configurations and conditions.

Note: **IBM hardware products are manufactured from new parts, or new and used parts. Regardless, our warranty terms apply.**



Agenda

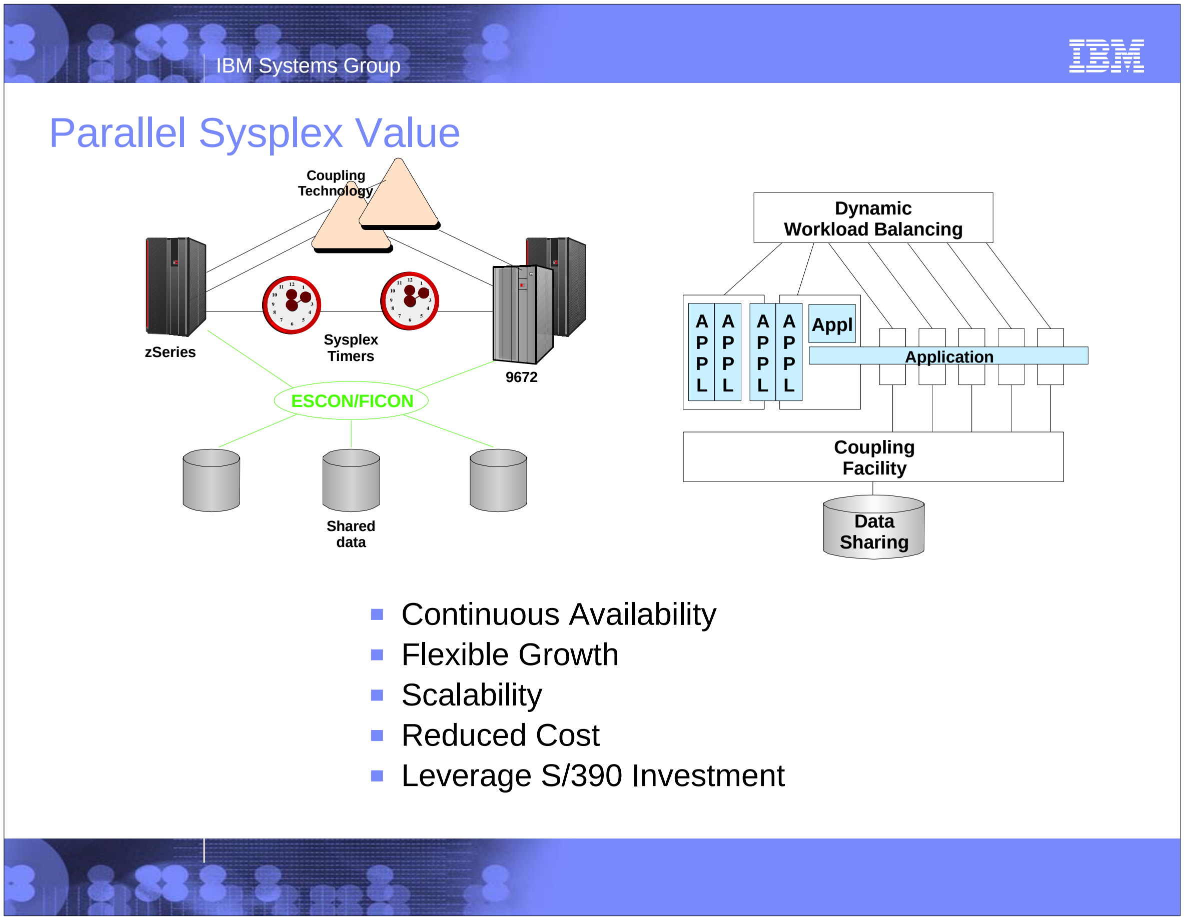
- **Parallel Sysplex Overview**
- **System Structure**
- **Coupling Facility and Link Technology**
- **Parallel Sysplex Software**
- **Summary**



Parallel Sysplex - What is it?

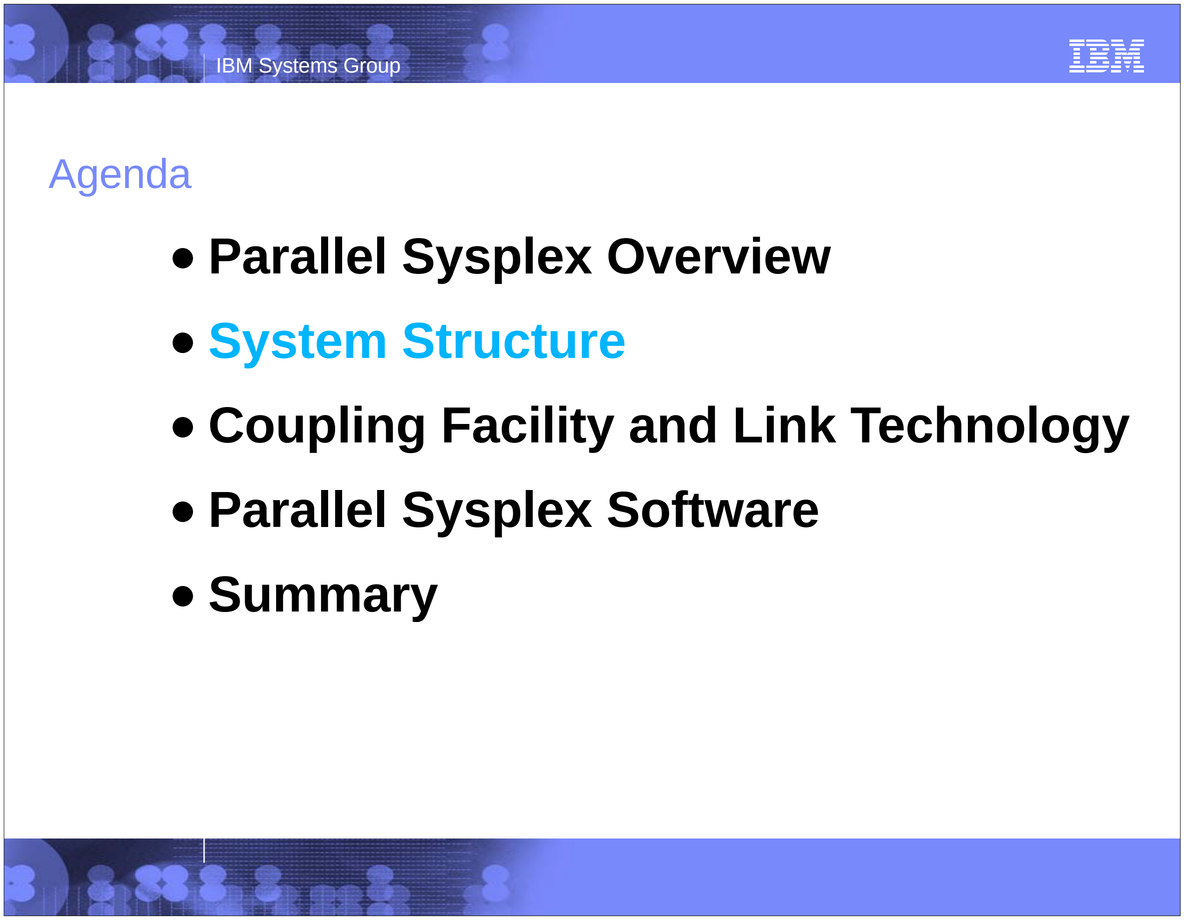
- **Hardware**
  - **Timer**
  - **I/O Connectivity**
  - **Coupling Facility**
- **Software**
  - **XCF/XES**
  - **WLM**
- **Microcode**
  - **CFCC**
  - **Processor u-code**

Parallel Processing!



Parallel Sysplex Value

- Continuous Availability
- Flexible Growth
- Scalability
- Reduced Cost
- Leverage S/390 Investment

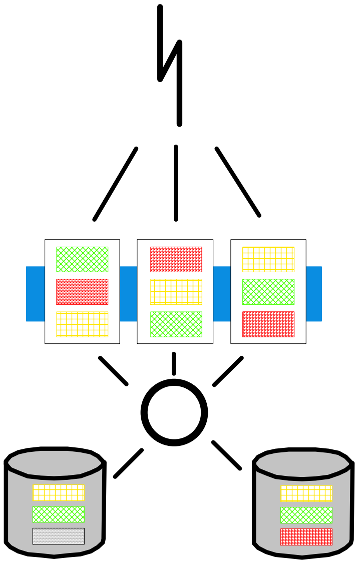


Agenda

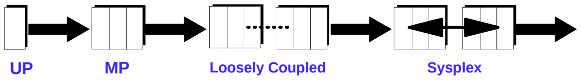
- **Parallel Sysplex Overview**
- **System Structure**
- **Coupling Facility and Link Technology**
- **Parallel Sysplex Software**
- **Summary**

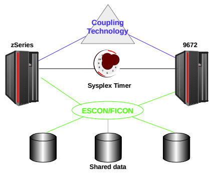
The zSeries Parallel Sysplex Solution

- Shared data
- Dynamic workload balancing
- Continuous application availability
- Incremental Non-disruptive growth

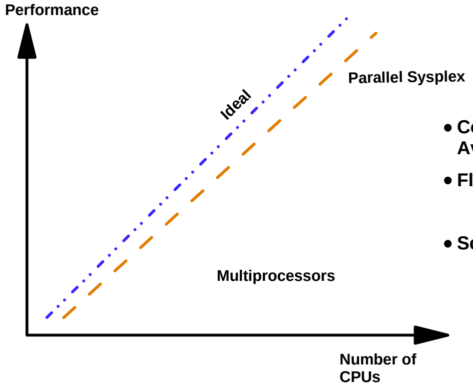


Evolution in Information Processing





zSeries Parallel Sysplex

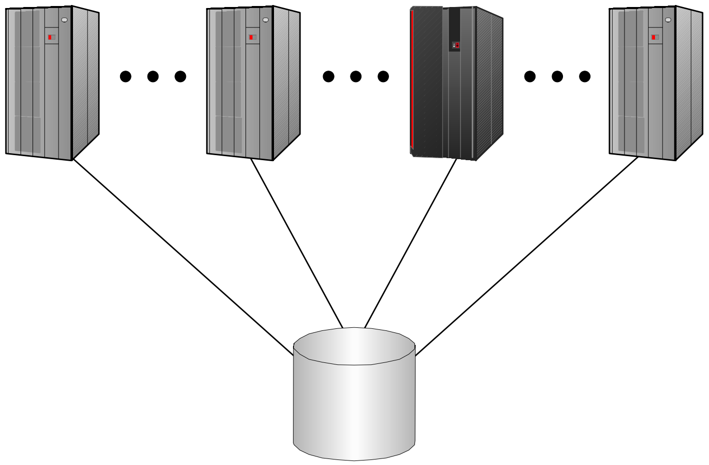


- Continuous Availability
- Flexible Growth
  - Granular
- Scalable
  - Nearly Unlimited

- Data Sharing
- Dynamic Workload Balancing
- Single System Image

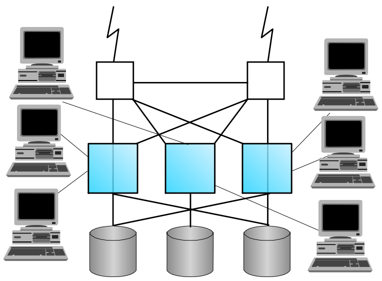
- Unique Design of IBM Hardware and Software
- Base for Future Enhancements

The Challenge



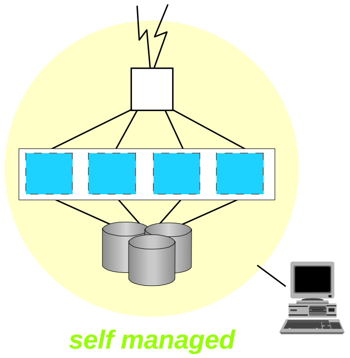
Single System Image/Systems Management
No Application Changes

Simplified Systems Management



people managed

Complex!



self managed

Simple!

Design Objective

A Parallel Sysplex must

- **Look like a single image**
- **Be managed more easily than today's single image**

Through

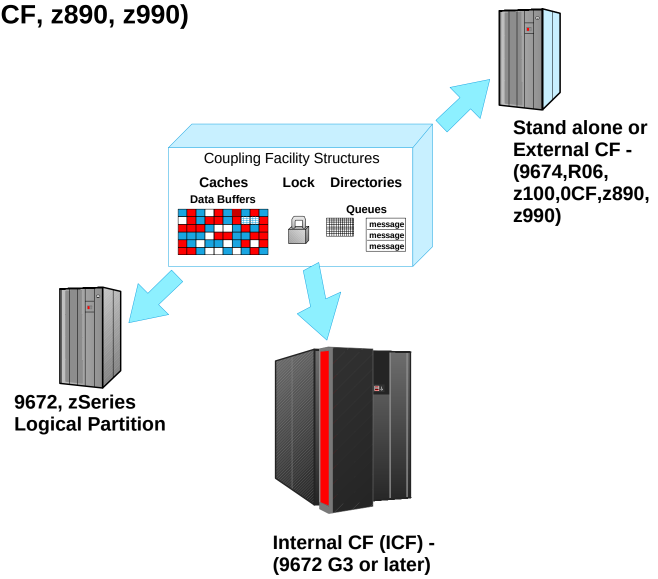
- **Elimination of tasks**
- **Reduction of complexity**
- **Automation**
- **Cloning Systems**

Agenda

- Parallel Sysplex Overview
- System Structure
- Coupling Facility and Link Technology
- Parallel Sysplex Software
- Summary

Coupling Facility Options

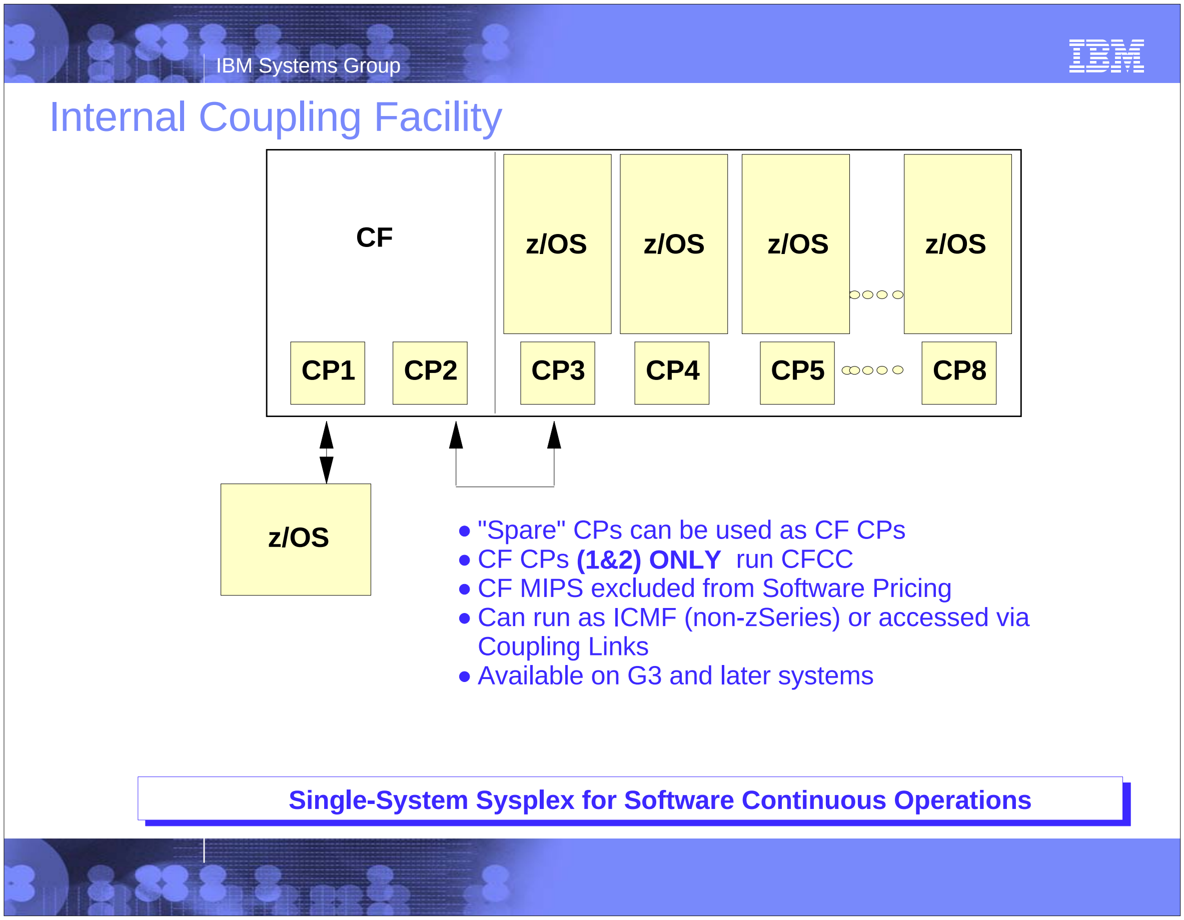
- Coupling Facility Control Code (CFCC) microcode creates a CF
  - Stand Alone (9674, R06, z100, 0CF, z890, z990)
  - Integrated (G3 +)
  - In an LPAR
- Integrated Coupling Migration Facility (ICMF)
  - For Test & Migration
  - 9672 only



Stand alone or External CF - (9674,R06, z100,0CF,z890, z990)

9672, zSeries Logical Partition

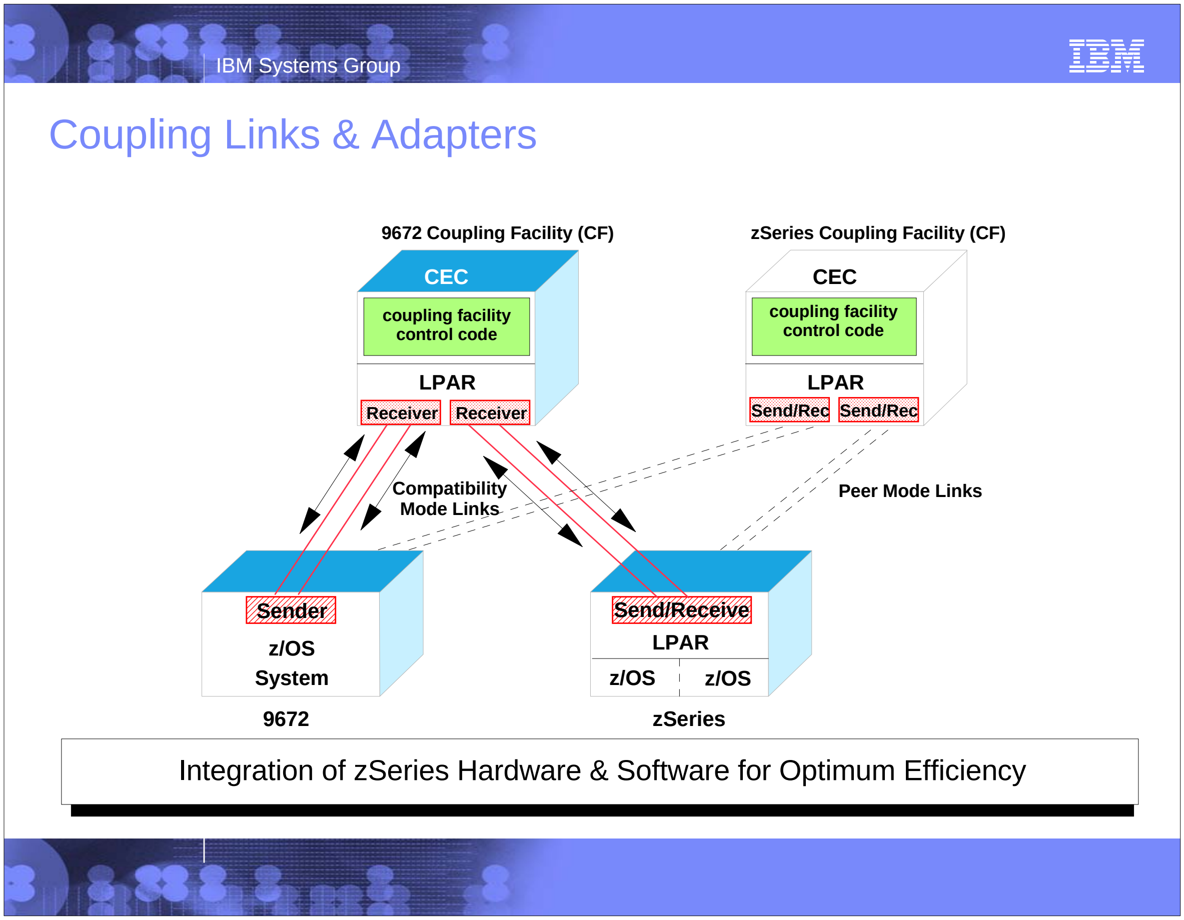
Internal CF (ICF) - (9672 G3 or later)



Internal Coupling Facility

- "Spare" CPs can be used as CF CPs
- CF CPs **(1&2) ONLY** run CFCC
- CF MIPS excluded from Software Pricing
- Can run as ICMF (non-zSeries) or accessed via Coupling Links
- Available on G3 and later systems

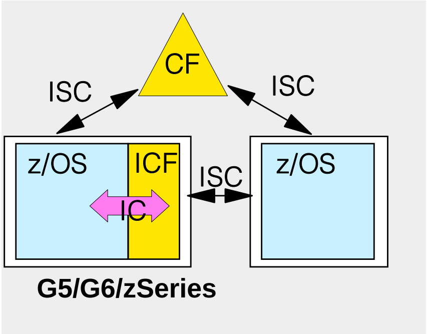
Single-System Sysplex for Software Continuous Operations

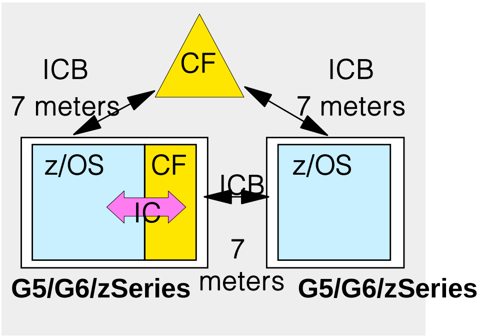


Coupling Links & Adapters

Integration of zSeries Hardware & Software for Optimum Efficiency

Link Technology





Inter System Channel (ISC)

- Original Coupling Channel
- Provides long distance connections

Internal Coupling (IC) Channel

- Fastest coupling performance
- Reduced complexity
- Increased reliability
- Standard on 9672 G5 and up

Integrated Cluster Bus (ICB)

- High bandwidth link for short distance
- Fastest interconnection link
- Improved processor utilization

- **Cluster Technology Scales with Processor Speed**
- **Peer mode available between zSeries servers/CFs**

zSeries Coupling Technology

- 64 Bit Architecture
- Dedicated or Shared CPs
- Up to 15 LPs
- Up to 32 GB of storage
- zSeries Peer Channels
  - InterSystem Channels-3 (ISC3)
  - Integrated Cluster Bus-3 (ICB3)
  - Integrated Cluster Bus-4 (ICB4)
- zSeries Compatibility Channels
  zSeries to 9672/9674
  - InterSystem Channels (ISC)
  - Integrated Cluster Bus (ICB)
- Dynamic CF Dispatch
- Dynamic ICF Expansion

z990 Model 300

- Up to 32 ICFs
- Up to 16 ICB-3
- Up to 32 ISC-3
- Up to 16 ICB-4
- Up to 64 GB/book
- Upgrade to z990
- Cannot upgrade directly from z100

z900 Model 100

- Up to 9 ICFs
- Up to 16 ICB-3
- Up to 32 ISC-3
- Upgrade to z900
- Upgrade from R06

z800 Model 0CF

- Up to 4 ICFs
- Up to 6 ICB-3
- Up to 24 ISC-3
- Upgrade to z800

z890...

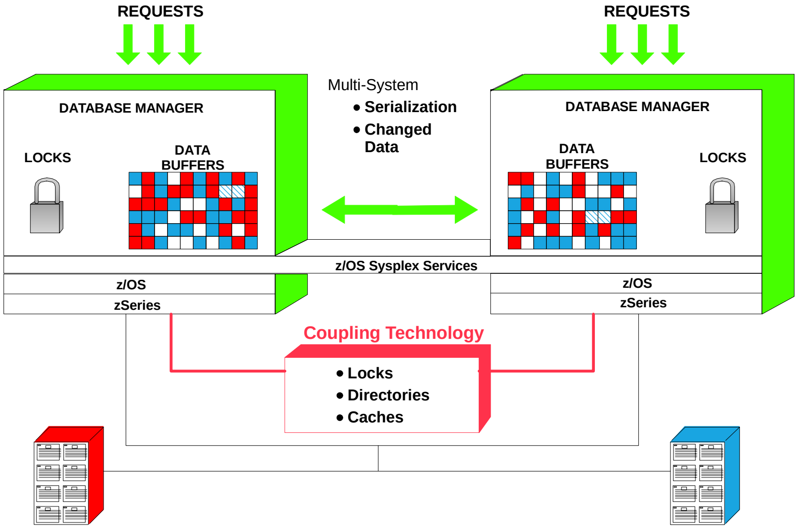
- Up to 4 ICFs
- Up to 16/8 ICB-3/4
- Up to 48 ISC-3
- Upgrade to z990

Coupling Facility CFCC Levels

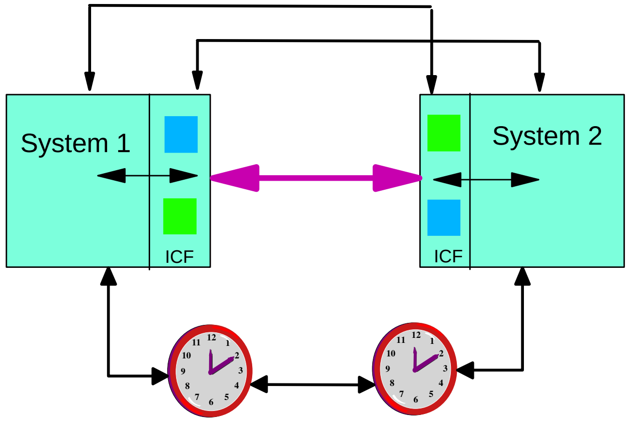
CF Level	Function	G3	G4	G5	G6	z800	z900	z890/990
14	**CFCC Dispatcher Restructure**							**x**
13	**DB2 castout processing performance enhancements**					**x**	**x**	**x**
12	64-bit CFCC addressability Message Time Ordering DB2 Performance SM Duplexing support for zSeries CFs Toleration for LPAR id >15 on z9xx					X X X X X	X X X X X X	X X X X X X
11	SM Duplexing support for 9672 G5/G6/R06 Toleration for LPAR id >15 on z9xx			X X	X X			
10	z900 GA2 Level						X	X
9	Intelligent Resource Director IC3 / ISC3 / ICB3 peer mode MQSeries Shared Queues WLM Multi-System Enclaves			 X X	 X X	X X X X	X X X X	X X X X
8	Dynamic ICF Expansion into shared ICF pool Systems-Managed Rebuild	 X	 X	X X	X X	X X	X X	X X
7	Shared ICF partitions on server models DB2 Delete Name optimization	 X	 X	X X	X X	X X	X X	X X
6	ICB & IC TPF support	 X	 X	X X	X X	X X	X X	X X
5	DB2 cache structure duplexing DB2 castout performance improvement Dynamic ICF expansion into shared CP pool	X X X	X X X	X X X	X X X	X X X	X X X	X X X
4	Performance optimization for IMS & VSAM RLS Dynamic CF Dispatching Internal Coupling Facility IMS shared message queue extensions	X X X X	X X X X	X X X X	X X X X	X X X X	X X X X	X X X X
3	IMS shared message queue base	X	X	X	X	X	X	X
2	DB2 performance VSAM RLS 255 Connectors / 1023 structures for IMS Batch DL1	X X X	X X X	X X X	X X X	X X X	X X X	X X X
1	Dynamic Alter support CICS temporary storage queues System logger	X X X	X X X	X X X	X X X	X X X	X X X	X X X

Details available in "Coupling Facility Level (CFLevel) Consideration" at URL **ibm.com**/servers/eserver/zseries/pso/cftable.html

Role of the Coupling Facility



System Managed CF Structure Duplexing



OS: z/OS v1.2 or later
ICFs: zSeries G5, G6
CFs: R06 or zSeries
CFCC: Level 11 (G5/G6) or Level 12 or higher (zSeries)

- - Automatic Rebuild for planned reconfiguration
- - Automatic switchover for unplanned outages
- - Automatic duplexing re-establishment
- - Overlapped requests for high performance

- - Consistent Recovery Mechanism
  - Reduced complexity
- - Faster than structure rebuild
- - Enables a robust "all-ICF" configuration

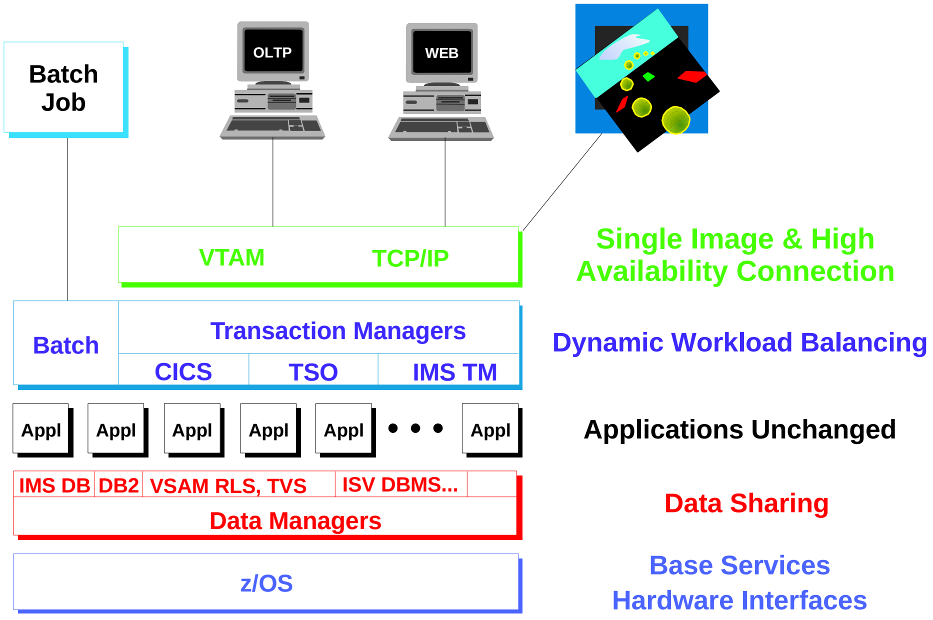
Parallel Sysplex Hardware Cluster Technology

Hardware Component	Function
Sysplex Timer (9037)	Consistent Multi-system Time Reference
Coupling Links	High Performance sysplex communications
- multi-mode ISC	50 MB/sec
- Single mode ISC, ISC-3 (peer)	100 MB/sec, 200 MB/sec
- HiPer Links	100MB/sec (w/improved adapters)
- ICB, ICB-3, ICB-4 (peer)	333 MB/sec, 1000 MB/sec, 2000 MB/sec
- IC, IC-3 (peer)	700 MB/sec, 1250 MB/sec
Coupling Facility (7th generation)	High performance processor
CFCC (14th level)	CF structures (list, lock, cache) and operations (high performance contention detection, etc.)
ESCON/FICON I/O Architecture and Directors	Flexible, high availability I/O connectivity
I/O Fencing	Failure Isolation
PPRC Freeze	Data Consistency for Disaster Recovery
IRD	CP, I/O balancing across workloads
CF Structure Duplexing	High availability CF data and faster failover

Agenda

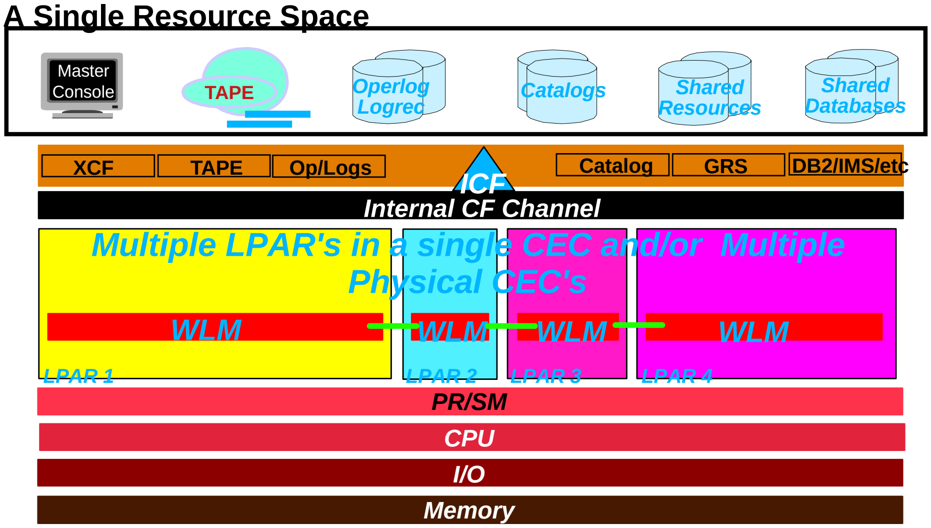
- Parallel Sysplex Overview
- System Structure
- Coupling Facility and Link Technology
- Parallel Sysplex Software
- Summary

Parallel Sysplex OLTP Software Structure



Parallel Sysplex Resouce Sharing

A Single Resource Space



Master Console

TAPE

Operlog Logrec

Catalogs

Shared Resources

Shared Databases

XCF | TAPE | Op/Logs | ICF | Catalog | GRS | DB2/IMS/etc

Internal CF Channel

Multiple LPAR's in a single CEC and/or Multiple Physical CEC's

WLM | WLM | WLM | WLM

LPAR 1 | LPAR 2 | LPAR 3 | LPAR 4

PR/SM

CPU

I/O

Memory

Single-CEC Value

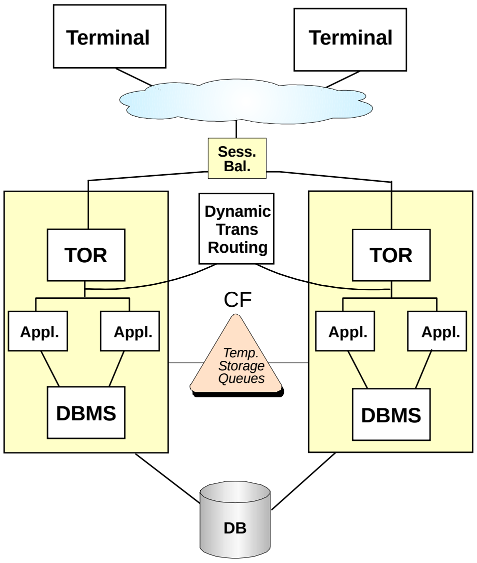
Multiple-CEC Value

Improved Performance

Simplified Systems Mgmt

Shared Resources

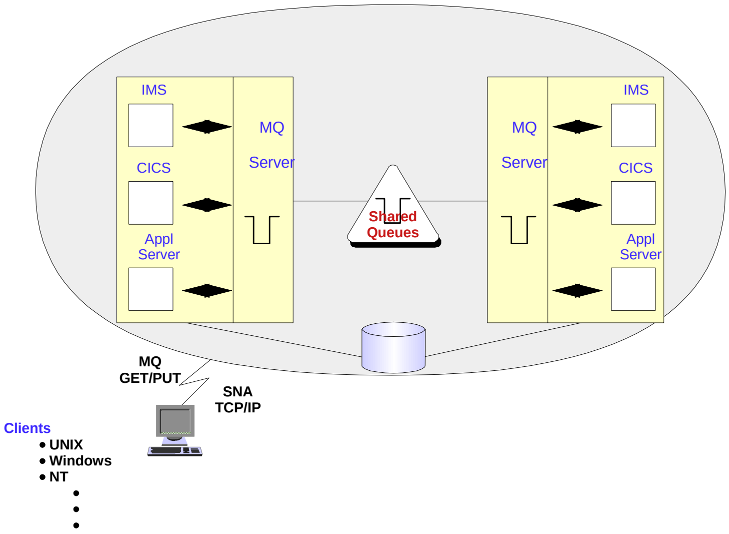
Transaction Management



- Single System Image
- Dynamic Session Balancing
- Dynamic Transaction Routing

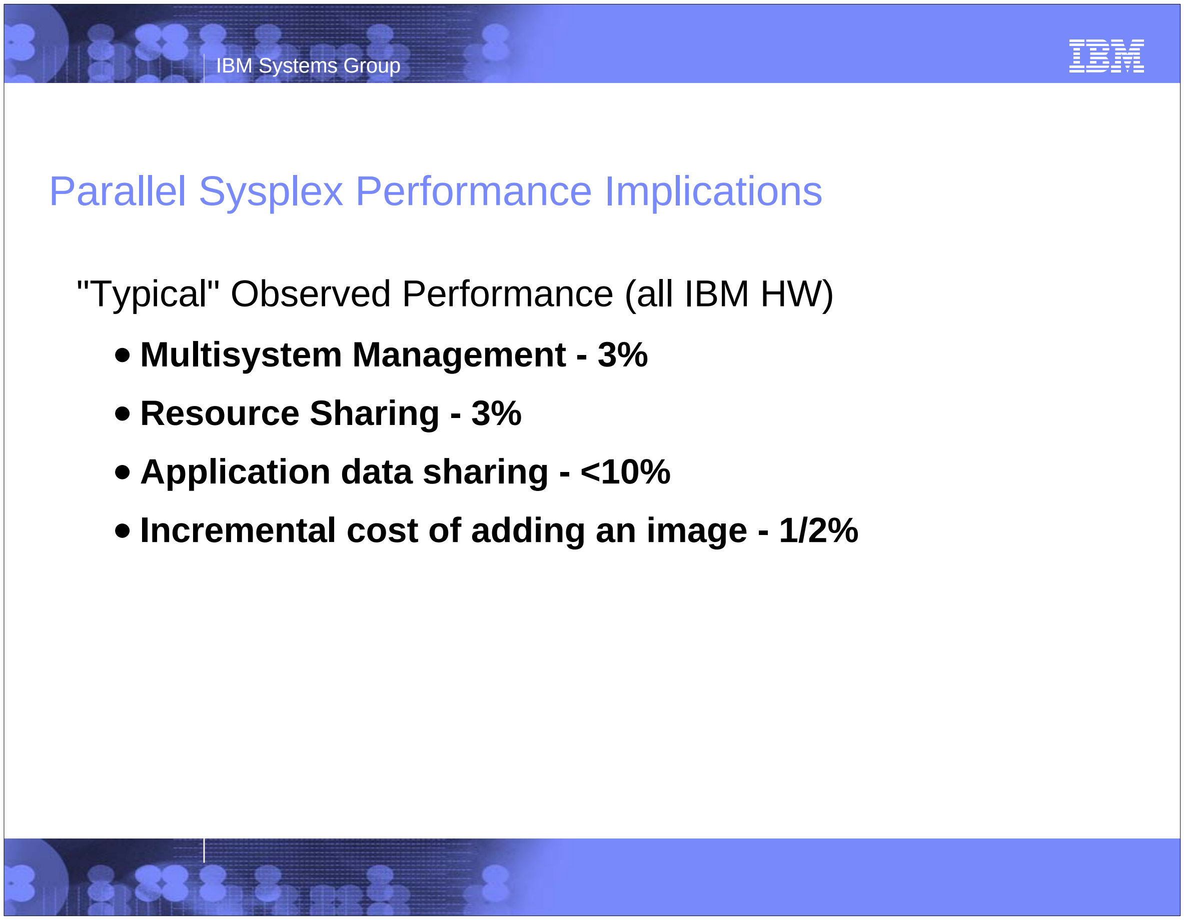
- Applications
  - *Any transaction can run anywhere*
  - *All data used by the application can be shared*

Websphere MQ Series



Parallel Sysplex Software Cluster Technology

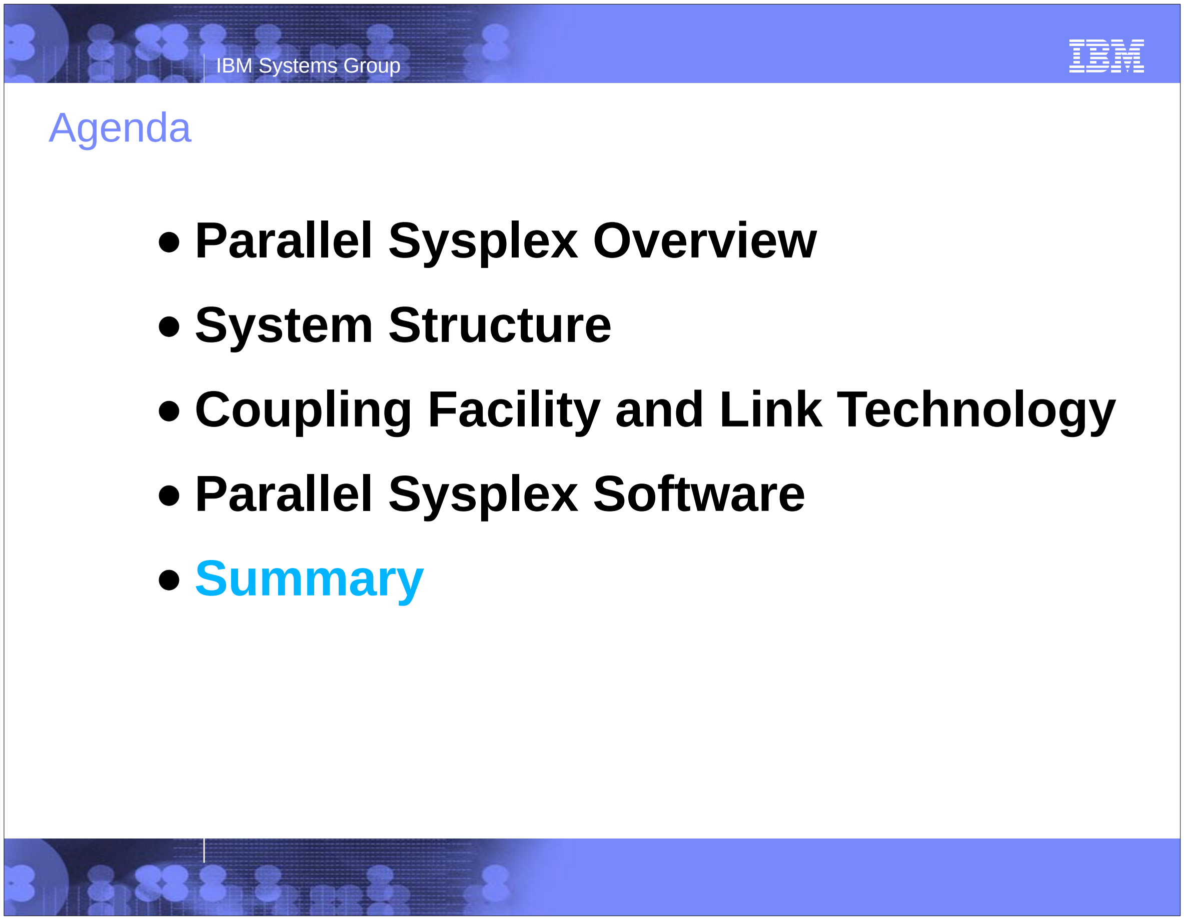
Software Component	Function
XCF	Sysplex Communication/Status Monitoring/Group Services
ARM	Subsystem restart (within CEC or cluster)
CFRM	CF Resource Management Policy
System Logger	High performance logging, Merged logs
WLM	Goal oriented unit of work management
WLM Enclaves	Mult-system unit of work
VTAM Generic Resource	Network Single System Image
VTAM MNPS	High Availability Network Connection
TCP/IP VIPA	Network Single System Image
TCP/IP VIPA take over/take back	High Availability Network Connection
CICSPlex/SM, IMS and MQ SMQ	Transaction routing/balancing
DB2 Sysplex Query Parallelism	SQL Query de/re-composition
Batch PipePlex	Cluster I/O Piping
ESCON Manager	ESCON I/O Systems Mangement
DB2, VSAM TVS, IMS/DB	Full read/write data sharing
IRLM	Sysplex database locking
Base Operating System Exploitation	Resource Sharing
Additional Subsystem Exploitation	Resource/Data Sharing



Parallel Sysplex Performance Implications

"Typical" Observed Performance (all IBM HW)

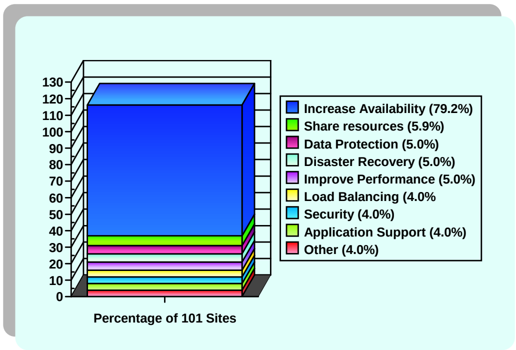
- **Multisystem Management - 3%**
- **Resource Sharing - 3%**
- **Application data sharing - <10%**
- **Incremental cost of adding an image - 1/2%**



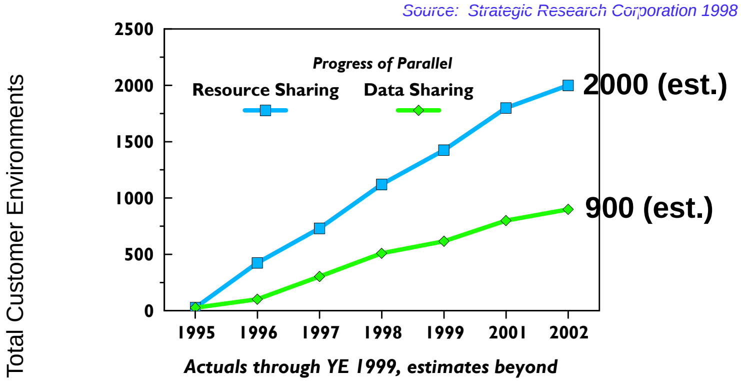
Agenda

- **Parallel Sysplex Overview**
- **System Structure**
- **Coupling Facility and Link Technology**
- **Parallel Sysplex Software**
- **Summary**

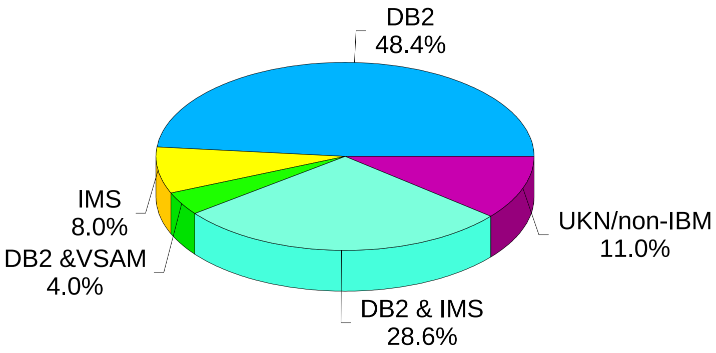
Parallel Sysplex Status



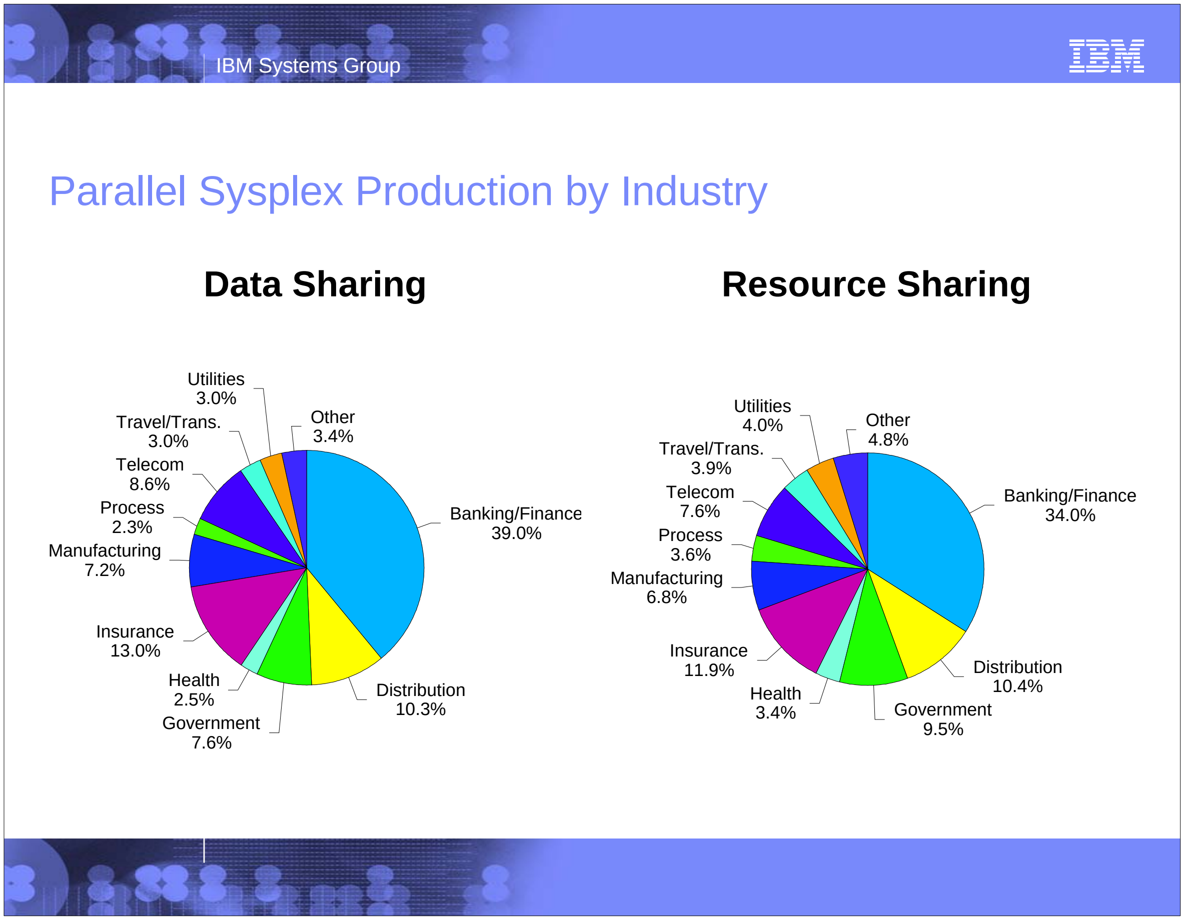
Source: Strategic Research Corporation 1998 Clustering Practices Profile



Data Sharing Database Summary



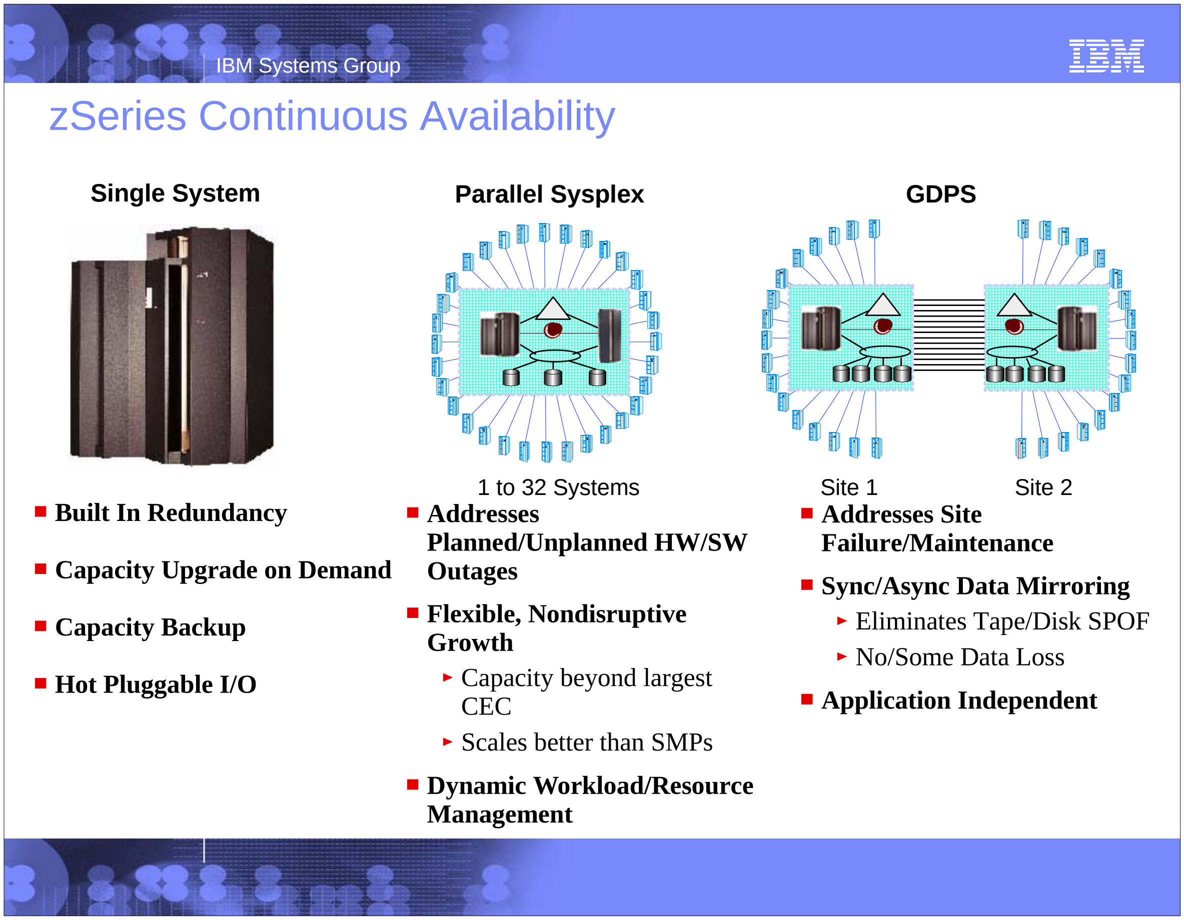
		Number	%
TOTAL SITES		525	100
Sites with DB2		425	81.0
Sites with IMS		192	36.6



Parallel Sysplex Production by Industry

Data Sharing

Resource Sharing



zSeries Continuous Availability

Single System

- **Built In Redundancy**
- **Capacity Upgrade on Demand**
- **Capacity Backup**
- **Hot Pluggable I/O**

Parallel Sysplex

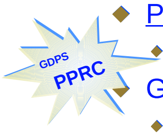
1 to 32 Systems

- **Addresses Planned/Unplanned HW/SW Outages**
- **Flexible, Nondisruptive Growth**
  - Capacity beyond largest CEC
  - Scales better than SMPs
- **Dynamic Workload/Resource Management**

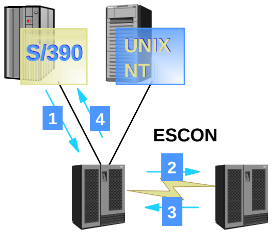
GDPS

- **Addresses Site Failure/Maintenance**
- **Sync/Async Data Mirroring**
  - Eliminates Tape/Disk SPOF
  - No/Some Data Loss
- **Application Independent**

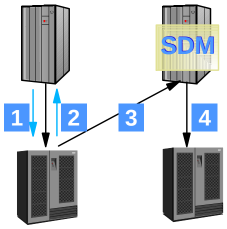
GDPS/PPRC and GDPS/XRC



- Peer to Peer Remote Copy (PPRC)
  - Synchronous data mirroring
- GDPS manages secondary data consistency
  - No or limited data loss in failover - user policy
- Production site exception condition monitoring
  - GDPS initiates and executes failover
- Distance between sites up to 40KM (fiber)
- **Continuous Availability and Disaster Recovery solution**

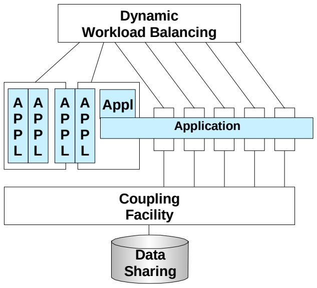


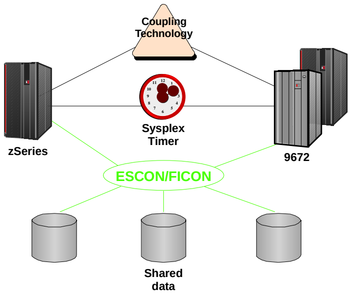
- eXtended Remote Copy (XRC)
  - Asynchronous data mirroring
  - Limited data loss to be expected in unplanned failover
- XRC manages secondary data consistency
- GDPS executes parallel sysplex restart - limited user involvement
- Supports any distance
- **Disaster Recovery solution**



Parallel Sysplex Value

Parallel Sysplex





The best server for diverse workloads . . .

- **Traditional**
- **New network computing**

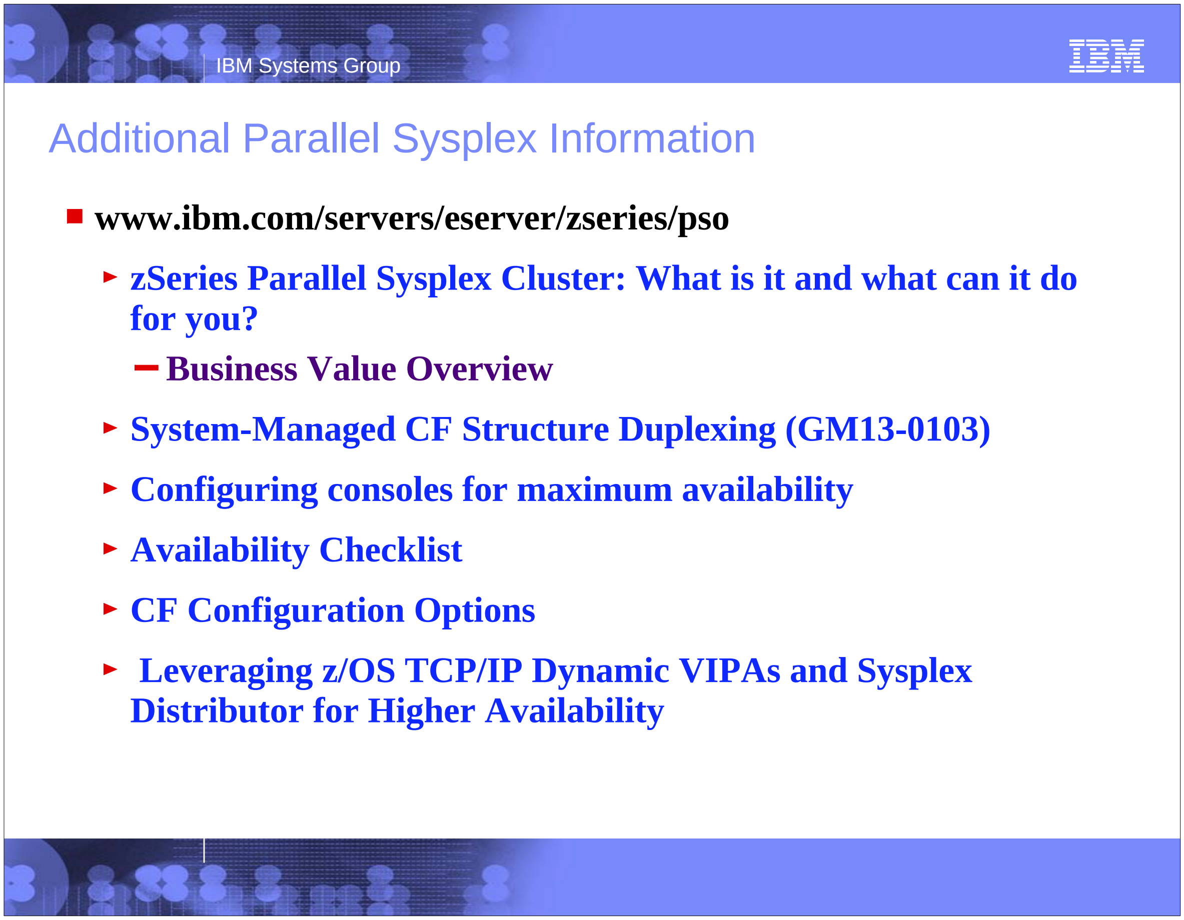
continuous application availability

virtually unlimited capacity

leverage existing investments

lowest incremental cost

classic strengths!



IBM Systems Group

Additional Parallel Sysplex Information

- **www.ibm.com/servers/eserver/zseries/pso**
  - **zSeries Parallel Sysplex Cluster: What is it and what can it do for you?**
    - **Business Value Overview**
  - **System-Managed CF Structure Duplexing (GM13-0103)**
  - **Configuring consoles for maximum availability**
  - **Availability Checklist**
  - **CF Configuration Options**
  - **Leveraging z/OS TCP/IP Dynamic VIPAs and Sysplex Distributor for Higher Availability**