Sysplex: Key Coupling Facility Measurements

Peter Enrico
Peter.Enrico@EPStrategies.com
813-435-2297

Enterprise Performance Strategies, Inc
(z/OS Performance Education and Managed Service Providers)

http://www.epstrategies.com

Abstract and Offer

- **Abstract**
  - Contained in a coupling facility are structures that are used by z/OS Sysplex exploiters for intersystem communication and work coordination. These structures are either List, Lock, or Cache structures. Requests to these structures are either synchronous or asynchronous. During this presentation, Peter Enrico will provide an overview and usage of some of the key Coupling Facility measurements used to help understand Coupling Facility and z/OS Sysplex performance.

- **Report Generation Offer**
  - Please contact Peter directly if you are a customer installation that would like for Peter to generate a complete set of coupling facility reports (charts and table) with your data. Will process up to 24 hours of data.
  - Over 50 reports (Including will be host effect reports)
  - Send an email to peter.enrico@epstrategies.com for instructions for sending him raw SMF data.

If You Liked this Presentation...

- ...you will find great value by attending one of our other classes
  - Visit our web site and sign the guestbook to be notified of our class schedule
    - *WLM Performance and Re-evaluating of Goals*
      Instructor: Peter Enrico
      - **October 18 - 22, 2010, St. Paul, MN**
    - *Essential z/OS Performance Tuning*
      Instructor: Peter Enrico and Tom Beretvas
      - **No scheduled at this time**
    - *Parallel Sysplex and z/OS Performance Tuning*
      Instructor: Peter Enrico
      - **September 13 - 17, 2010, Philadelphia, PA**
- For more information, please visit: http://www.epstrategies.com

Contact, Copyright, and Trademark Notices

Questions?

Send email to Peter at Peter.Enrico@EPStrategies.com, or visit our website at http://www.epstrategies.com.

Copyright Notice:

Trademarks:

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

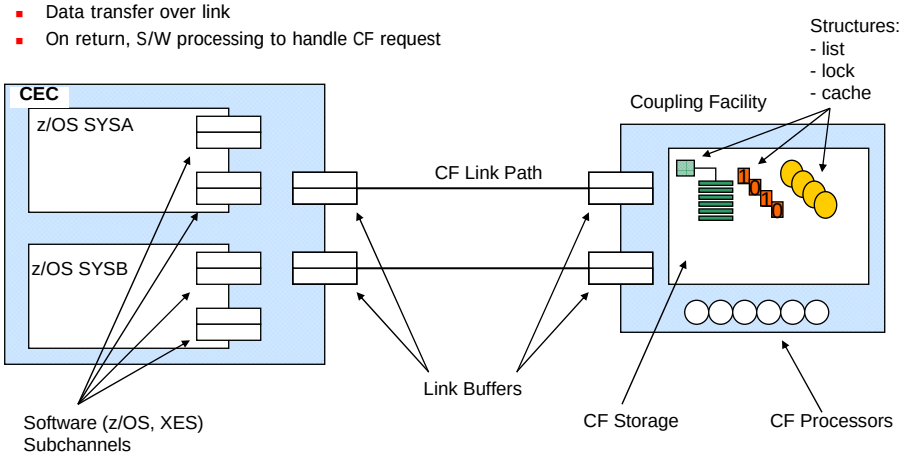
The following are trademarks of Enterprise Performance Strategies, Inc.: **Health Check®, Reductions®**

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

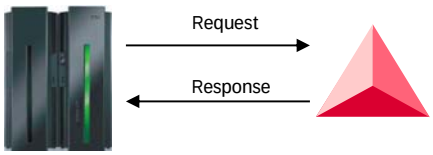
Performance Analyst View of CF Resource

- z/OS Processing
  - S/W processing to make CF request
  - Request a sub-channel
  - Request a path
  - Data transfer over link
  - On return, S/W processing to handle CF request
- Coupling Facility Processing
  - Link time (i.e. time on path)
  - CF busy processing request



Performance of Coupling Facility Requests

- Performance is heavily dependant on a number of variables:
  - Speed of requesting CPU
    - Larger processor will 'wait faster' for a response
  - Type of request – Synchronous versus Asynchronous
  - Busy conditions (Subchannel, path)
  - Time it takes to transmit data to the CF
    - CF link performance
    - Speed of data over link
  - Speed of CF processor
    - Shared LPAR or dedicated CF?
  - Storage of CF
  - Structures
  - Coupling facility duplexing



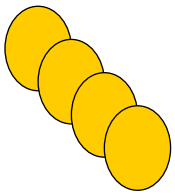
Many Questions Need to be Asked of Measurements

- Configuration / Setup questions
- Link and general load performance questions *(discussed SHARE 2010 Winter)*
- Host effect questions *(discussed SHARE 2010 Winter)*
- Processor related questions
- Storage Usage related questions
- List structure related questions *(discussed SHARE 2010 Winter)*
- Lock structure related questions *(discussed SHARE 2010 Winter)*
- Cache structure related questions (This presentation)

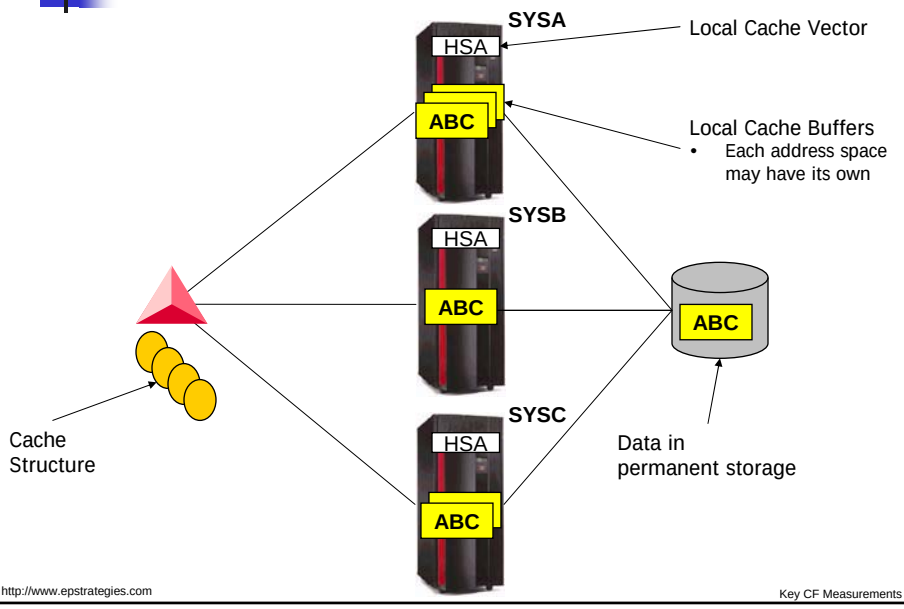
Review of Cache Structures

Cache Structures

- CF can be used as a high speed caching facility
- Cache structure made up of
  - directory to keep track of registered data elements
  - optionally, data elements
- Usage of cache structure
  - data consistency / buffer validation
    - ability to maintain a shared copy of data in cache structure in CF
    - ability to keep track of shared data that does not reside in CF
      - permanent storage (i.e. disk)
      - local storage (i.e. z/OS or subsystem buffers)
  - high speed data access
    - Shared data can be stored in cache structure and made available to every system in sysplex
    - Invalid local copy of data can be refreshed with CF cached copy
    - CF access faster than I/O subsystem cache



Cache Structure Components



Cache Structure Terminology

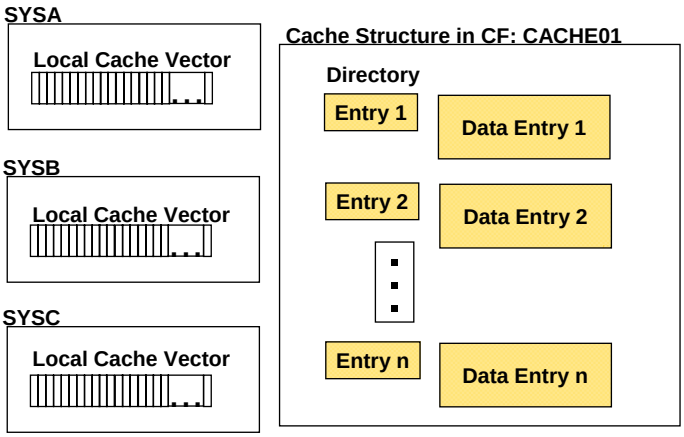
- Local Cache Buffers
  - Buffers in private area storage of cache structure exploiting subsystems
  - Required and allocated by every exploiter of cache structures
  - Contains copies of shared data
  - Populated by disk or CF cache structure
  - Used to refresh CF cache structure or disk copy
- Permanent Storage
  - Final and permanent repository for shared data - usually disk
  - Used to populate local cache buffers
- Local Cache Vectors
  - User defined vector in HSA
  - Allows connectors of a cache structure (i.e. those sharing data) to determine if their local cache buffers contain the latest copy of the data

Cache Structure Terminology

- The cache structure in the coupling facility has two primary components
- Directory Entries
  - Used to keep track of data entries that are shared among multiple systems
  - Every system that has a copy of a particular piece of shared data has a registration entry in this portion of the cache structure.
  - It is this directory whose entries are used to generate cross invalidation signals to indicate that a record in a local cache buffer may be invalid
- Data Entries
  - Used to contain a cached version of the data
  - Optional

Cache Structure Components cont…

- Directory - Used to keep track of share entries
- Data Entries - Used to optionally cache data



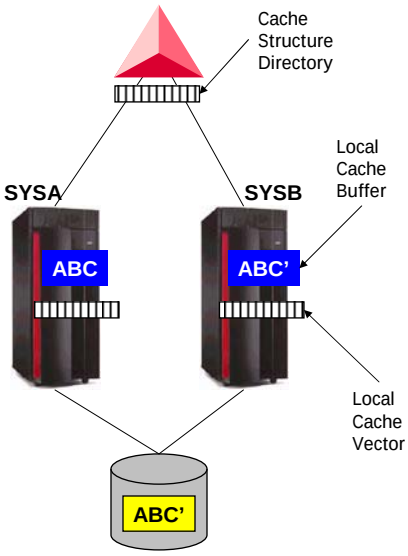
Cache Structure Usage

- There are three ways that CF cache structures are used
- Directory in Cache
  - CF structure is used to assist in maintaining consistency of data in local cache buffers
- Store Through Cache
  - Most recent copy of data is kept in both CF cache structure and DASD
- Store-In Cache
  - Most recent copy of data kept in CF cache structure and hardened to DASD asynchronously

NOTE: The way a cache structure is used and measured is based on the exploiter and how the exploiter is using making use of the cache structure.

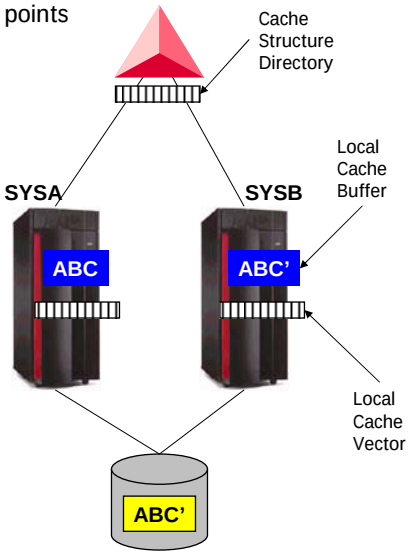
Directory in Cache Structure

- CF structure is used to assist in maintaining consistency of data in local cache buffers
- Writes
  - CF cache structure only contains directory entries; no data stored in CF
  - Data always written from local cache buffers to DASD
  - CF used to invalidate local other's buffers
- Reads
  - Local cache buffer version of data is used if local cache vector indicates that it is still valid
  - If local cache vector indicates that local copy is invalid, a fresh read from permanent storage is done



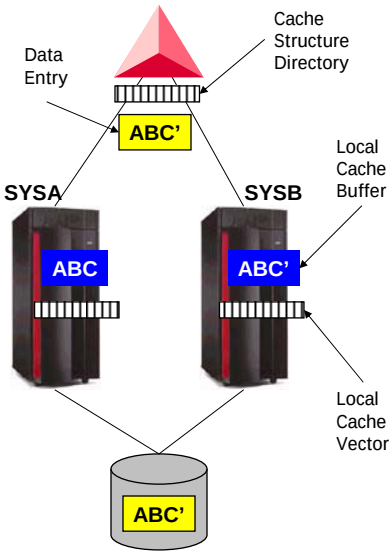
Directory In Cache Structure

- What do you think the performance sensitive points are for a Directory in Cache structure?
  - Size of directory



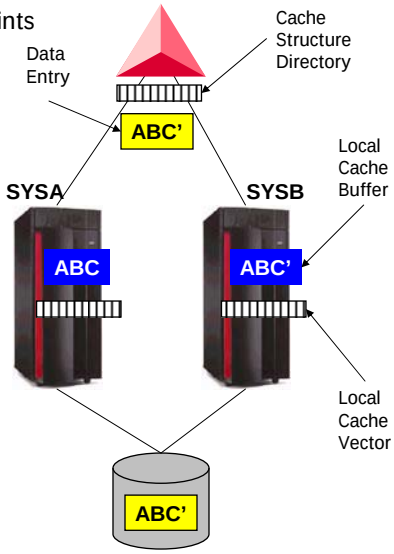
Store Through Cache Structure

- Most recent copy of data is kept in both CF cache structure and permanent storage
- Writes
  - CF cache structure contains directory entries
  - CF cache structure contain both changed and unchanged data
  - Data written to CF and DASD at the same time (serialized)
  - No cast out processing
- Reads
  - Local cache buffer version of data is used if local cache vector indicates that it is still valid
  - If local cache vector indicates that local copy is invalid, a fresh read from CF



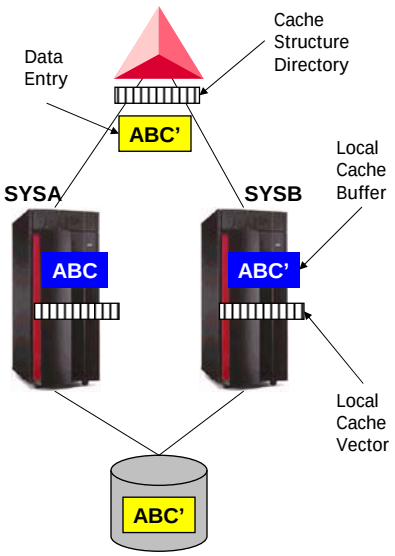
Store Through Cache Structure

- What do you think the performance sensitive points are for a Store Through Cache structure?
  - Size of directory in structure
  - Size of data entry in structure
  - Size of data entry relative to directory entry
  - Reads from Cache Structure



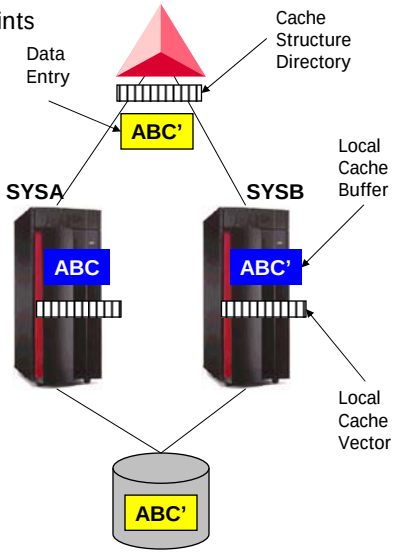
Store-In Cache Structure

- Most recent copy of data kept in CF cache structure and hardened to DASD later
  - (Asynchronously)
- Writes
  - CF cache structure contains directory entries
  - Changed data stored in CF
  - Updated data written to DASD later by subsystem. Known as cast out processing.
  - Data in CF may be different than data on DASD
- Reads
  - Local cache buffer version of data is used if local cache vector indicates that it is still valid
  - If local cache vector indicates that local copy is invalid, a fresh read from CF
  - DASD is check as last resort

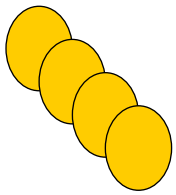


Store-In Cache Structure

- What do you think the performance sensitive points are for a Store-In Cache structure?
  - Size of directory in structure
  - Size of data entry in structure
    - Forcing of writes
  - Size of data entry relative to directory entry
  - Reads from Cache Structure



More Popular Cache Structure Exploiters



- Enhanced Catalog Sharing (ECS)
  - store in cache structure
- RACF
  - store through cache structure
  - For frequently accessed data
- DB2
  - Buffer Pools – store in cache structure and store-through
- VSAM RLS
  - System buffer Pools – store through cache structure
- IMS
  - IMS VSO DEDB – store in cache structure
  - OSAM – directory only cache structure
  - OSAM – store through cache structure
  - VSAM – directory only cache structure

Cache Structure Measurements

Tuning Cache Structures

- Cache structures consist of two main components
  - Directory entries
  - Data elements
- When shortage of space occurs
  - Directory entries for unchanged data are reclaimed via LRU algorithm
  - Buffer invalidation on host systems must occur
    - CF notifies all systems with a registered interest in the structure
  - Access times will suffer if the data needs to be re-accessed
    - I/O must occur
- Balance
  - Too large a structure wastes storage, may cause spikes in CPU during invalidation processing
  - Too small may cause invalidated data to be re-accessed from DASD

Some questions of interest

- Directory Entries
  - Is the size too small
  - Forcing cross invalidates?
  - Forcing castout processing?
- Data Entries
  - Is the size too small
  - Forcing cross invalidates?
  - Forcing castout processing?
- Reads and Writes
  - For data written to the structure, is system benefiting from reads
    - Example: High writes and low reads?
High writes and high reads?
- Castout processing
  - Natural or being forced due to too small size of data entry

Analyzing Cache Structure Activity Example

```
                                    COUPLING  FACILITY  STRUCTURE  ACTIVITY
STRUCTURE NAME = DSNDB3G_GBP10     TYPE = CACHE  STATUS = ACTIVE

            # REQ     ------------- REQUESTS -------------    ------------- DELAYED REQUESTS -------------
 SYSTEM     TOTAL             #     % OF  -SERV TIME(MIC)-     REASON     #   % OF  ---- AVG TIME(MIC) -----
 NAME       AVG/SEC           REQ    ALL     AVG   STD_DEV             REQ    REQ   /DEL    STD_DEV   /ALL
 SYSA        3448K   SYNC    3326K   22.0    17.0     13.0     NO SCH    6    0.0   96.3       130.4    0.0
             1915    ASYNC    122K    0.8    98.4     61.6     PR WT     0    0.0    0.0         0.0    0.0
                     CHNGD       6    0.0  INCLUDED IN ASYNC   PR CMP    0    0.0    0.0         0.0    0.0
                                                              DUMP      0    0.0    0.0         0.0    0.0
 SYSB        2041K   SYNC    1940K   12.8    16.0     14.9     NO SCH  149    0.0   2163        4787    0.2
             1134    ASYNC    101K    0.7   201.9   1388.9     PR WT     0    0.0    0.0         0.0    0.0
                     CHNGD     149    0.0  INCLUDED IN ASYNC   PR CMP    0    0.0    0.0         0.0    0.0
                                                              DUMP      0    0.0    0.0         0.0    0.0
 SYSC        2504K   SYNC    2415K   16.0    13.8     11.3     NO SCH  419    0.0  125.8       129.3    0.0
             1391    ASYNC     89K    0.6   141.6    297.1     PR WT     0    0.0    0.0         0.0    0.0
                     CHNGD     206    0.0  INCLUDED IN ASYNC   PR CMP    0    0.0    0.0         0.0    0.0
                                                              DUMP      0    0.0    0.0         0.0    0.0
 SYSD        3312K   SYNC    3103K   20.5    16.4     13.2     NO SCH 5704    0.2  472.2        5263    0.8
             1840    ASYNC    203K    1.3   137.4    548.5     PR WT     0    0.0    0.0         0.0    0.0
                     CHNGD    5700    0.0  INCLUDED IN ASYNC   PR CMP    0    0.0    0.0         0.0    0.0
                                                              DUMP      0    0.0    0.0         0.0    0.0
 SYSE        3799K   SYNC    3681K   24.4    16.1     11.4     NO SCH  422    0.0   2818       10743    0.3
             2111    ASYNC    118K    0.8   193.5   2760.8     PR WT     0    0.0    0.0         0.0    0.0
                     CHNGD     421    0.0  INCLUDED IN ASYNC   PR CMP    0    0.0    0.0         0.0    0.0
                                                              DUMP      0    0.0    0.0         0.0    0.0
 TOTAL      15104K   SYNC      14M   95.8    16.0     12.7     NO SCH 6700    0.0  635.5        5633    0.3    -- DATA ACCESS ---
             8391    ASYNC    633K    4.2   151.2   1356.3     PR WT     0    0.0    0.0         0.0    0.0    READS      567928
                     CHNGD    6482    0.0                      PR CMP    0    0.0    0.0         0.0    0.0    WRITES    1684727
                                                              DUMP      0    0.0    0.0         0.0    0.0    CASTOUTS   813051
                                                                                                             XI'S        45216
```


Analyzing Cache Structure Activity Example

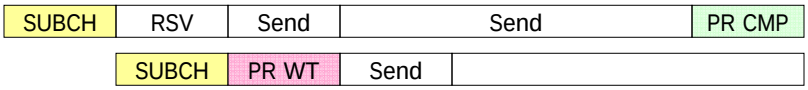
- Majority of report contains standard structure measurement
  - # Requests – Average and rate
  - Breakdown of sync versus async requests
  - Service times
    - Sync requests should be less than 20 microseconds
    - Async request should be less than 300 microseconds
    - Times depend on duplexing, H/W technology, and distance between CF an z/OS system

```
                                    COUPLING  FACILITY  STRUCTURE  ACTIVITY
STRUCTURE NAME = DSNDB3G_GBP10     TYPE = CACHE  STATUS = ACTIVE

            # REQ     ------------- REQUESTS -------------    ------------- DELAYED REQUESTS -----------
 SYSTEM     TOTAL             #     % OF  -SERV TIME(MIC)-     REASON     #   % OF  ---- AVG TIME(MIC) ---
 NAME       AVG/SEC           REQ    ALL     AVG   STD_DEV             REQ    REQ   /DEL    STD_DEV   /AL
 SYSA        3448K   SYNC    3326K   22.0    17.0     13.0     NO SCH    6    0.0   96.3       130.4    0.(
             1915    ASYNC    122K    0.8    98.4     61.6     PR WT     0    0.0    0.0         0.0    0.(
                     CHNGD       6    0.0  INCLUDED IN ASYNC   PR CMP    0    0.0    0.0         0.0    0.(
                                                              DUMP      0    0.0    0.0         0.0    0.(
```

Analyzing Cache Structure Activity Example

- Since many cache structures are duplex
  - SUBCH – delay due to Subchannel busy
  - PR WT – delay due to waiting on a peer to send
  - PR CMP – delay due to waiting on a peer to complete



COUPLING FACILITY STRUCTURE ACTIVITY

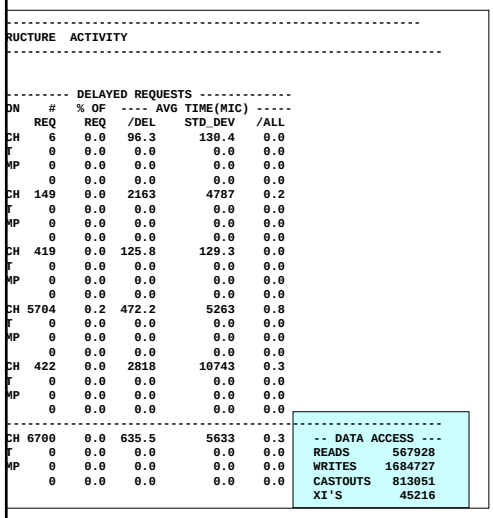
STRUCTURE NAME = DSNDB3G_GBP10 TYPE = CACHE STATUS = ACTIVE

SYSTEM NAME	# REQ TOTAL AVG/SEC		REQUESTS # REQ	% OF ALL	-SERV TIME(MIC)- AVG	STD_DEV	REASON	DELAYED REQUESTS # REQ	% OF REQ	---- AVG TIME(MIC) --- /DEL	STD_DEV	/AL
SYSA	3448K	SYNC	3326K	22.0	17.0	13.0	NO SCH	6	0.0	96.3	130.4	0.
	1915	ASYNC	122K	0.8	98.4	61.6	PR WT	0	0.0	0.0	0.0	0.
		CHNGD	6	0.0	INCLUDED IN ASYNC		PR CMP	0	0.0	0.0	0.0	0.
							DUMP	0	0.0	0.0	0.0	0.

Tuning Structure Delays

- If PR WT > 10% then improve links for secondary structure
  - Upgrade link technology
  - Add additional links
  - Additional Share senders CPs
- If PR CMP is high then improve CF speed of secondary structure
  - Perhaps a configuration change
    - Turn dynamic dispatching off
    - Dedicate CPs
  - Upgrade technology
  - Add additional CF CPs
- Monitor CF to CF service times
- Make sure duplexing is necessary
  - Very expensive from a performance point-of-view
  - Cost of duplexing is 2X the cost of not duplexing

Analyzing Cache Structure Activity Example



```
-------------------------------------------------------
RUCTURE  ACTIVITY
-------------------------------------------------------


-------- DELAYED REQUESTS ------------
ON    #   % OF  ---- AVG TIME(MIC) -----
    REQ    REQ   /DEL    STD_DEV   /ALL
CH    6    0.0   96.3      130.4    0.0
T     0    0.0    0.0        0.0    0.0
MP    0    0.0    0.0        0.0    0.0
      0    0.0    0.0        0.0    0.0
CH  149    0.0   2163       4787    0.2
T     0    0.0    0.0        0.0    0.0
MP    0    0.0    0.0        0.0    0.0
      0    0.0    0.0        0.0    0.0
CH  419    0.0  125.8      129.3    0.0
T     0    0.0    0.0        0.0    0.0
MP    0    0.0    0.0        0.0    0.0
      0    0.0    0.0        0.0    0.0
CH 5704    0.2  472.2       5263    0.8
T     0    0.0    0.0        0.0    0.0
MP    0    0.0    0.0        0.0    0.0
      0    0.0    0.0        0.0    0.0
CH  422    0.0   2818      10743    0.3
T     0    0.0    0.0        0.0    0.0
MP    0    0.0    0.0        0.0    0.0
      0    0.0    0.0        0.0    0.0
-------------------------------------------------------
CH 6700    0.0  635.5       5633    0.3
T     0    0.0    0.0        0.0    0.0
MP    0    0.0    0.0        0.0    0.0
      0    0.0    0.0        0.0    0.0

   -- DATA ACCESS ---
   READS      567928
   WRITES    1684727
   CASTOUTS   813051
   XI'S        45216
```

- Lower right hand side of report contains cache structure specific measurements
- Hardware measurements from the coupling facility
  - Since hardware measurement they are represented just once since they should be exactly the same from each system
  - If your own reports make sure you only count once and do not add up for each system

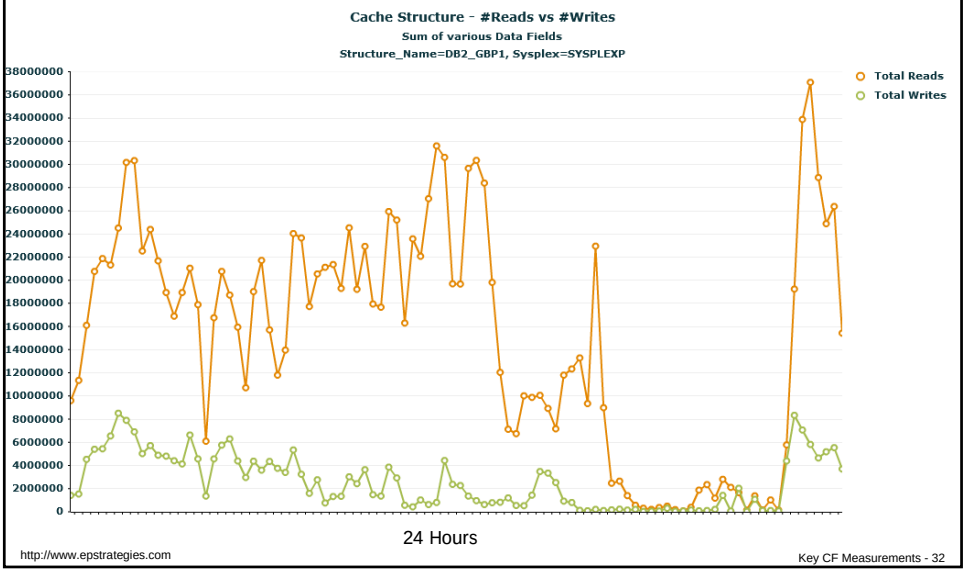
Analyzing Cache Structure Activity

- READS - Number of read hits
  - Count of the number of times the CF returned data on a read request by any connector
  - Note: Directory only caches will have a 0 value reported since the structure contains no "data"
- WRITES - Number of writes to the CF structure
  - Count of times a connector placed changed or unchanged data into the CF structure
  - Note: Directory only caches will have a 0 value reported since the structure contains no "data"
  - Note: Changed/unchanged is an attribute assigned to the data when written by the connector.
    - From a performance/capacity view point, the importance of the attribute is: changed data cannot be reclaimed from the structure should directory or data elements become scarce

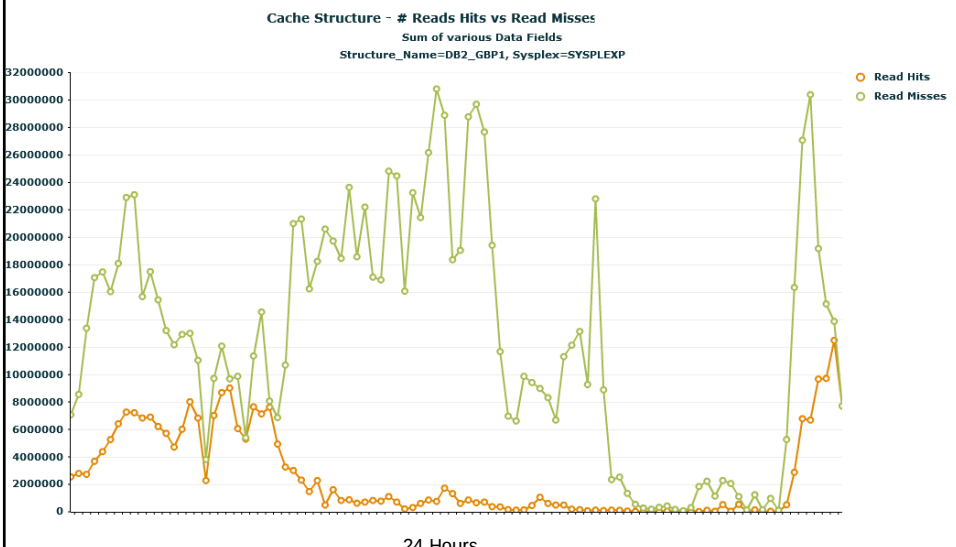
Analyzing Cache Structure Activity

- Conditions of Interest – Reads versus Writes
  - One key usage of a cache structure is to take advantage of caching the data in the CF for data sharing
  - Prefer to avoid file I/O
- High Writes versus Low Reads
  - Never getting the benefit of caching the data
  - Condition may indicate:
    - Insufficient structure space allocated, and data entries (and perhaps directory entries) are being discarded by the coupling facility space management routines
    - Inappropriate allocation of the ratio of directory entry to data elements is causing the data entries to be discarded by the coupling facility space management routines
- Note: For duplexed structures, expect secondary structure to have no/few reads

Reads versus Writes



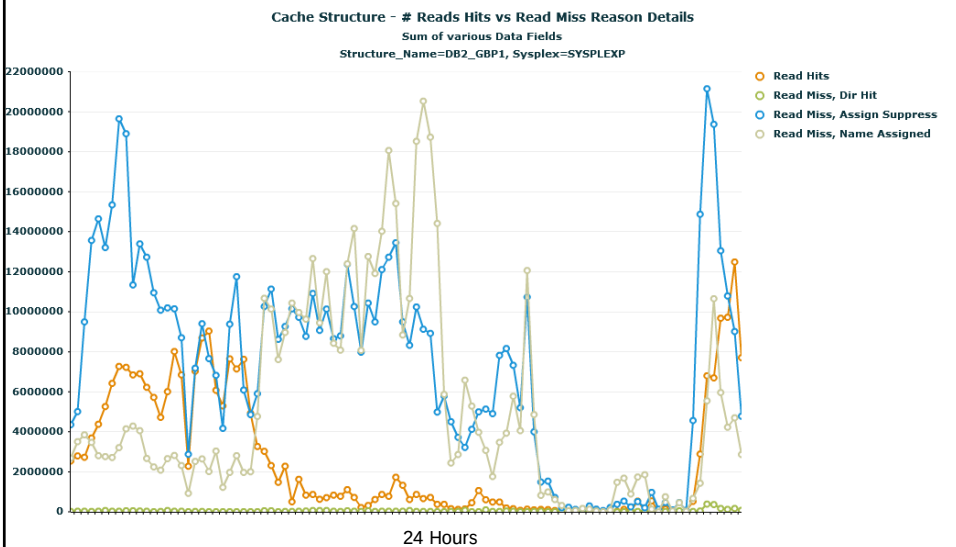
Read Hits versus Read Misses



http://www.epstrategies.com

Key CF Measurements - 33

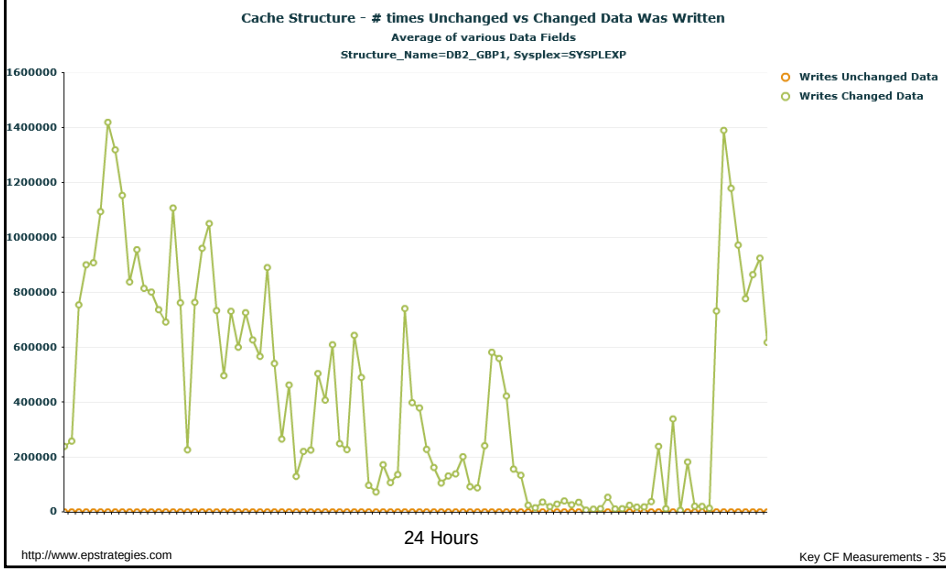
Read Hits versus Read Misses (by Reason)



http://www.epstrategies.com

Key CF Measurements - 34

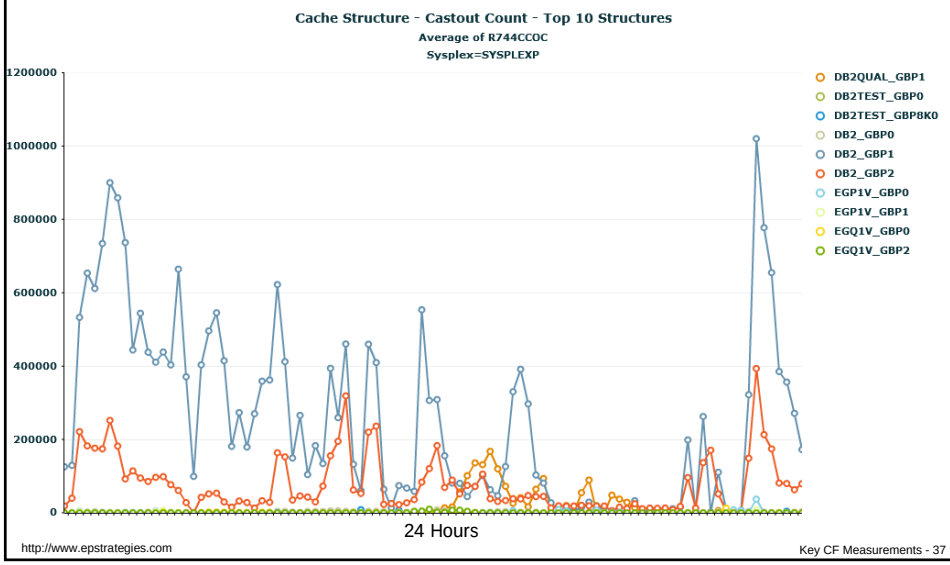
Writes Unchanged versus Changed



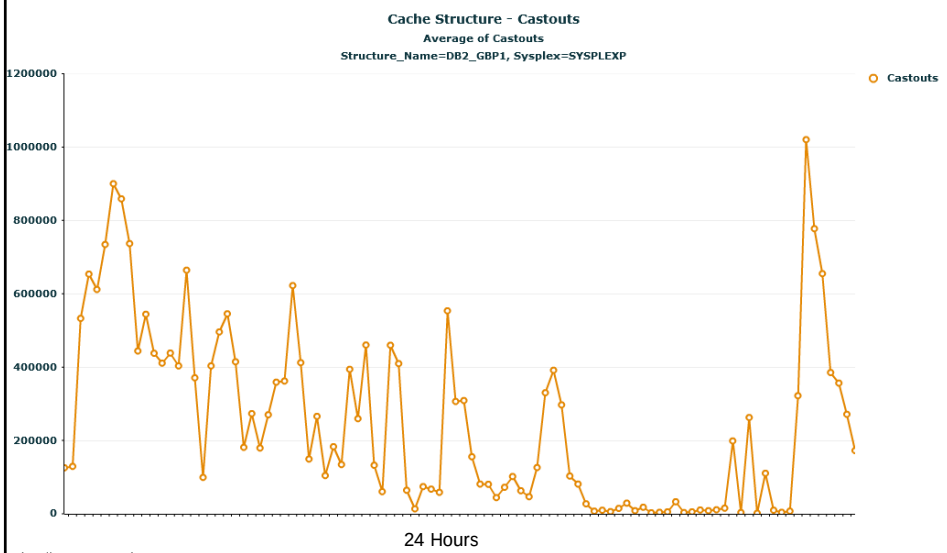
Analyzing Cache Structure Activity

- CASTOUTS - Number of times cast-out processing occurred (changed data)
  - This is a count of the number of times a connector retrieved a changed data entry, wrote the data to DASD and caused the changed attribute to be reset to unchanged.
  - Castouts due to reclaims is not desirable and will adversely effect the data base manager and/or the user of the data base manager
  - This counter is of interest for store-in cache structures (i.e. DB2 group buffer pool structures) in determining the volume of changed data being removed from the structure
    - Note: This counter is not an indicator of the number times cast out processing was performed during the RMF interval.
  - A large amount of cast out activity on a single structure may warrant additional cache structures and redirecting locally buffered data to different cache structure.
  - Cast out processing by the connectors must keep pace with the rate at which changed data is placed in the structure
  - When all directory or data elements are associated with changed data, no new data items can be registered or written to the structure.

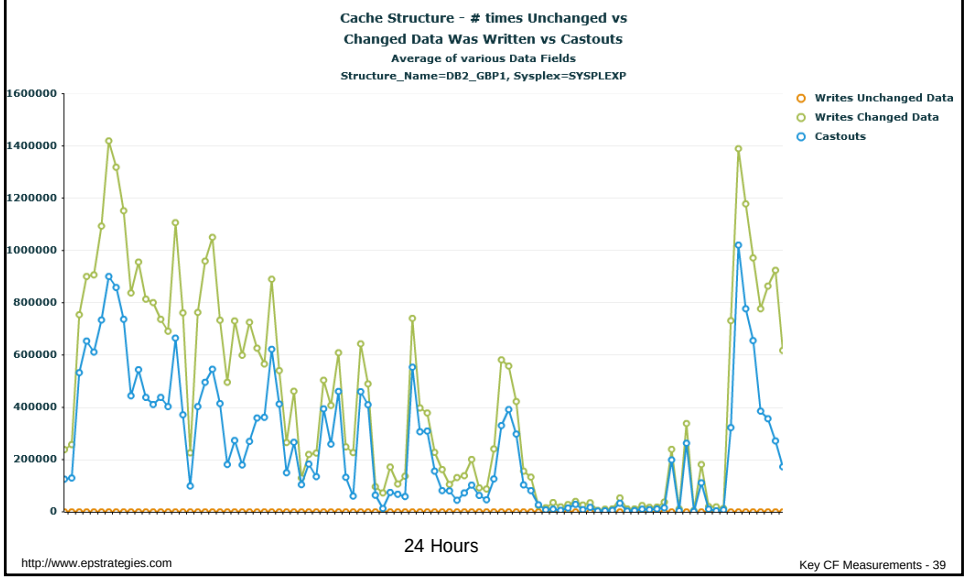
Castout Processing



Castout Processing



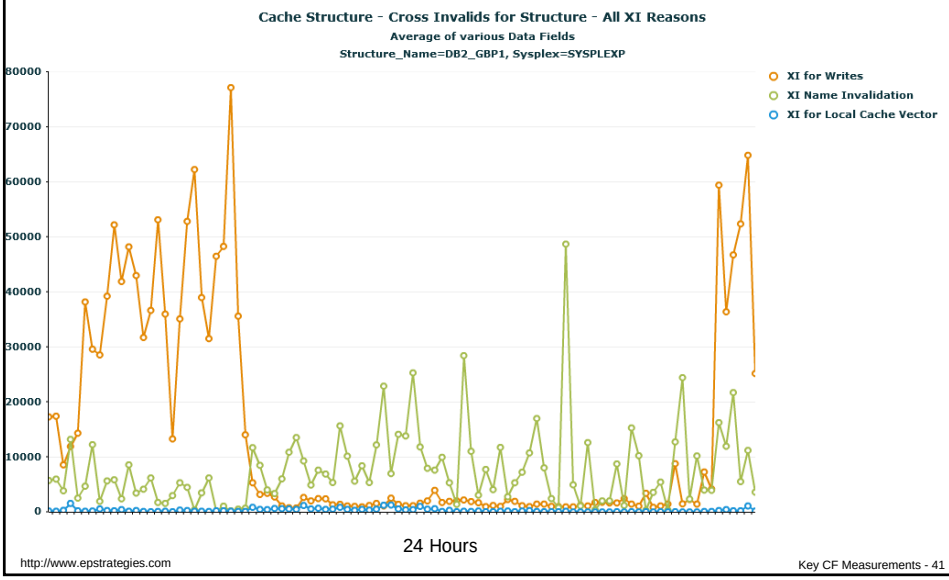
Castout Processing vs Writes



Analyzing Cache Structure Activity

- Data particular to Cache Structures (DATA ACCESS)
  - XI's - This is the number of times a data item residing in a local buffer pool was marked invalid by the coupling facility during the interval
    - XI'S count values are seen for directory, store-in and store-thru caches. This count reflects both the amount of data sharing among the users of the cache and the amount of write/update activity against the data bases.
  - To the cache structure user, this means the data item must be re-acquired from DASD or perhaps the coupling facility structure, and interest in the item must be re-registered in the coupling facility structure.
  - There are several "XI counts" obtained from the coupling facility which are consolidated into this value. They are:
    - XI for Directory Reclaim
    - XI for Write
    - XI for Name Invalidation
    - XI for Complement Invalidation
    - XI for Local Cache Vector Entry Replacement

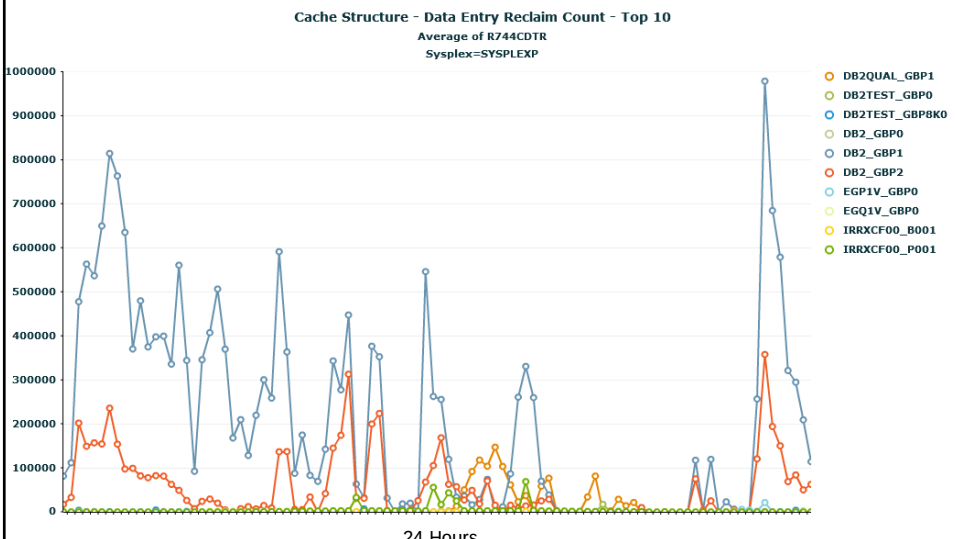
Cross Invalidates



Other Measurements of Interest

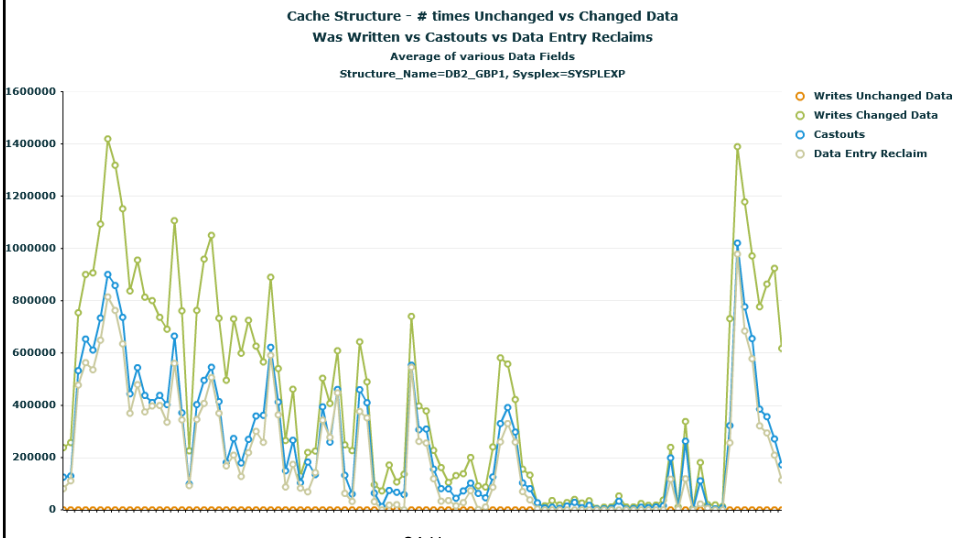
- Directory Entry Reclaims
- Data Entry Reclaims

Data Entry Reclaims



24 Hours

Data Entry Reclaims



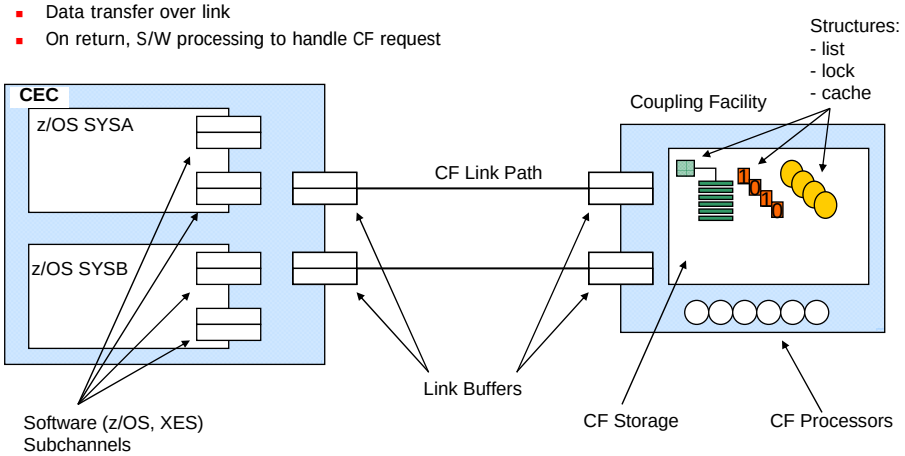
24 Hours

Many Questions Need to be Asked of Measurements

- Configuration / Setup questions
- Link and general load performance questions *(discussed SHARE 2010 Winter)*
- Host effect questions *(discussed SHARE 2010 Winter)*
- Processor related questions
- Storage Usage related questions
- List structure related questions *(discussed SHARE 2010 Winter)*
- Lock structure related questions *(discussed SHARE 2010 Winter)*
- Cache structure related questions (This presentation)

Performance Analyst View of CF Resource

- z/OS Processing
  - S/W processing to make CF request
  - Request a sub-channel
  - Request a path
  - Data transfer over link
  - On return, S/W processing to handle CF request
- Coupling Facility Processing
  - Link time (i.e. time on path)
  - CF busy processing request



If You Liked this Presentation...

- ...you will find great value by attending one of our other classes
  - Visit our web site and sign the guestbook to be notified of our class schedule
    - *WLM Performance and Re-evaluating of Goals*
      Instructor: Peter Enrico
      - **October 18 - 22, 2010, St. Paul, MN**
    - *Essential z/OS Performance Tuning*
      Instructor: Peter Enrico and Tom Beretvas
      - **No scheduled at this time**
    - *Parallel Sysplex and z/OS Performance Tuning*
      Instructor: Peter Enrico
      - **September 13 - 17, 2010, Philadelphia, PA**
- For more information, please visit: http://www.epstrategies.com