B81

IMS Data Sharing Implementation

Rich Lewis

IMS

Technical Conference

Sept. 27-30, 2004

Orlando, FL

RETURN TO INDEX

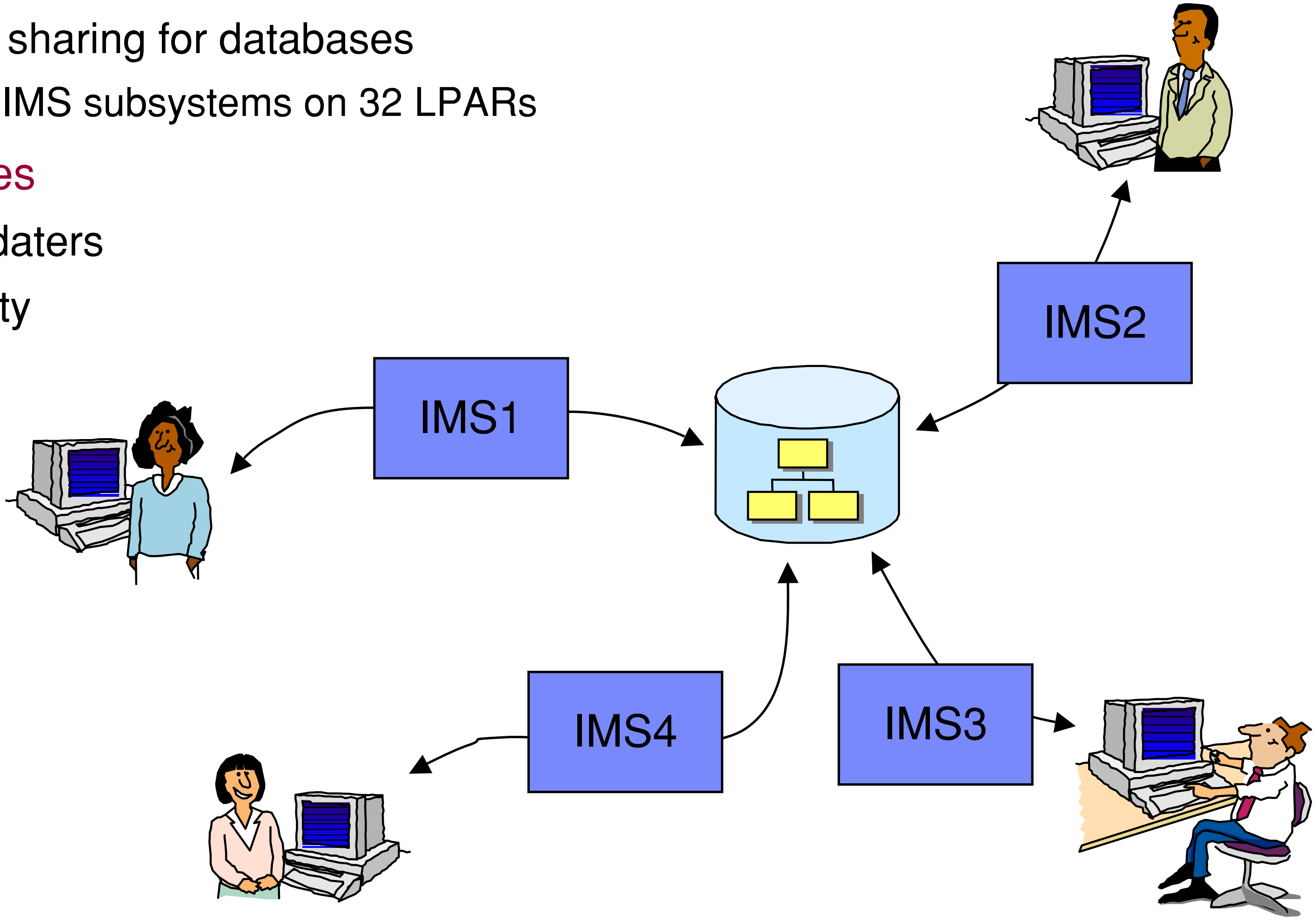
Topics

- DBRC
- System definition
- IMS execution parameters
- CF structures
- ILRM
- Lock contention
- Lock reporting
- Application considerations
- Availability and ease of operations
- Database recoveries
- System recoveries
- BMPs and batch jobs
- Implementation steps

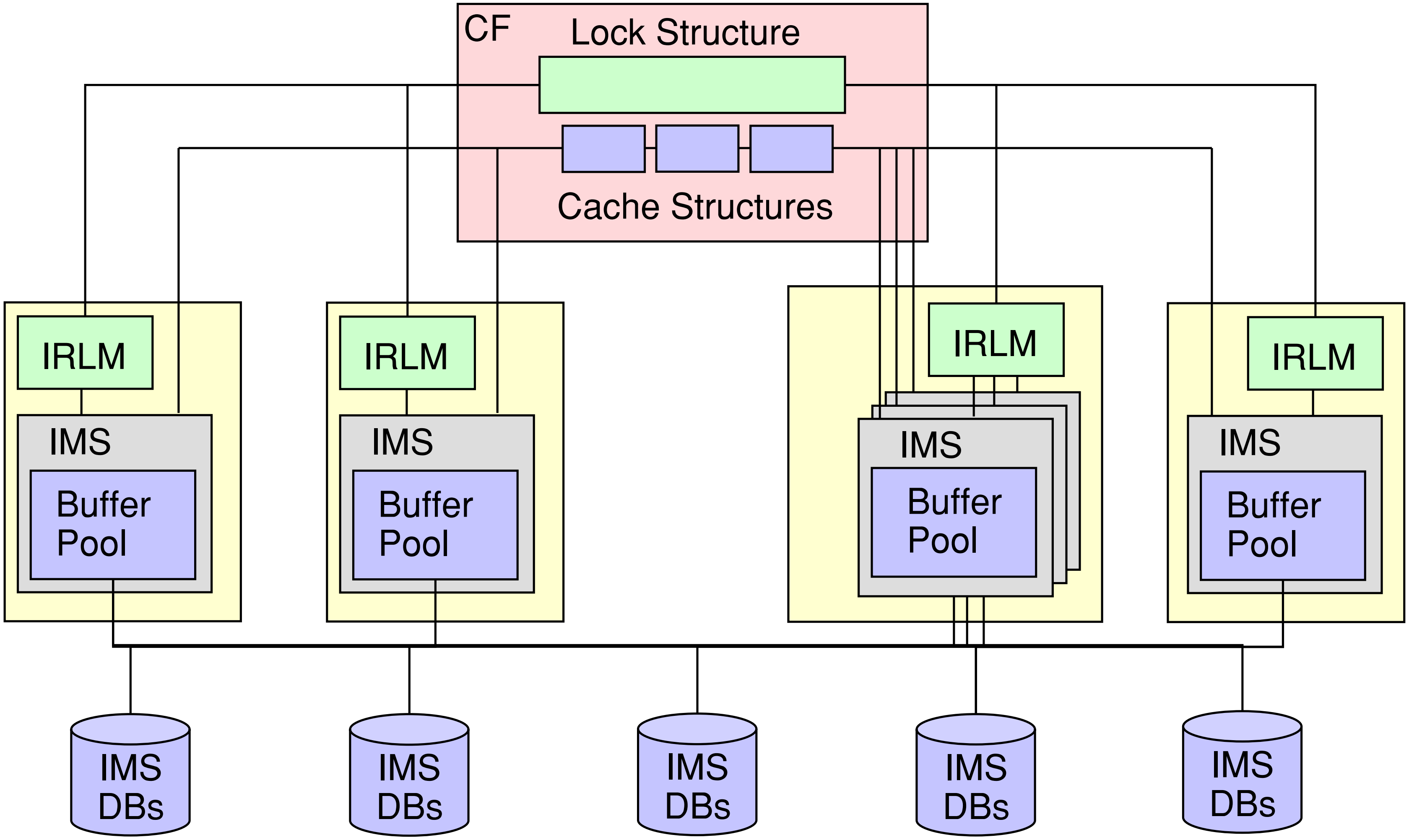
Advanced Technical Support
TECHNICAL SALES SUPPORT
AMERICAS

Block Level Data Sharing

- Block level data sharing (BLDS)
  - N-way data sharing for databases
    - Up to 255 IMS subsystems on 32 LPARs
- Full capabilities
  - Multiple updaters
  - Data integrity



Block Level Data Sharing Configuration



IMS systems include TM/DB, DBCTL, and IMS batch jobs

DBRC Considerations

- RECON data sets
  - Placement and reserve considerations
    - Important, but no additional considerations for data sharing
- All IMSs sharing a database must use the same RECONs
  - Dynamic allocation of RECONs is recommended
- Shared databases must be registered
  - Registration of databases increases accesses to RECONs
  - Sharing databases does not increase accesses to RECONs
- Share level for databases
  - SHARELVL(3)
    - Multiple LPAR block level data sharing
      - Multiple updaters on multiple LPARs with multiple IRLMs

DBRC Considerations

- DEDB Shared VSO
  - CHANGE.DBDS
    - CFSTR1(structure-name-1)
      - Defines cache structure for area
    - CFSTR2(structure-name-2)
      - Defines duplicate cache structure for area
    - LKASID | NOLKASID
      - Specifies if look aside buffering will be used for area

- IMS V9 DEDB Shared VSO
  - Allows multiple areas to share a structure
  - Duplexing of structures is done with system-managed duplexing

IMS System Definition

- DATABASE macro
  - ACCESS=UP
    - Full sharing allowed, updates allowed
- IMSCTRL macro
  - IRLM=Y or N
    - Establishes default execution setting
      - May be overridden
  - IRLMNM=
    - Establishes default IRLM name
      - May be overridden
  - DBRC=YES, NO, or FORCE
    - Establishes batch default
      - May be overridden

IMS Execution Parameters

- DFSPBxxx member
  - IRLM=Y
  - IRLM=irlm name
- DFSVSMxx member
  - CFNAMES statement
    - CFIRLM= lock structure name
    - CFOSAM=(OSAM structure name, dirratio, elemratio)
      - Used for OSAM database data sets
      - If you are not caching OSAM, make the elemratio value 0
    - CFVSAM=VSAM structure name
      - Used for full function VSAM database data sets
    - All keywords must be specified
      - CFOSAM= value may be omitted if you have no OSAM data sets
    - Values must be same in all IMSs in the data sharing group

Database Data Sets

- VSAM
  - On DEFINE CLUSTER
    - SHAREOPTION(3 3) must be specified
  - DISP=SHR must be specified
  - If either of these is not specified,
    - The data set will not be opened when the ILRM is used with DBRC SHARELVL(1, 2, or 3)

- OSAM
  - DISP=SHR must be specified

Coupling Facility Structures

- Must be defined in CFRM policy
  - Definition includes name, location, and size parameters
- Structures
  - Lock structure
  - Full function VSAM structure
  - Full function OSAM structure
  - DEDB VSO structures
    - One or two per shared area in V7 and V8
    - Shared by multiple areas in V9 (optional)
- CFSIZER tool
  - May be used to estimate structure size
    - http://www-1.ibm.com/servers/eserver/zseries/cfsizer/

Coupling Facility Structures

- IRLM lock structure
  - Size lock table to avoid false contention
    - Size is power of 2
    - Recommendation: 1000 entries per lock held
      - Size of entries determined by IRLM MAXUSRS parameter
  - Size record list to hold all locks protecting updates
    - Locks acquired with PROCOPT allowing updates
      - Full function database record locks and block locks
      - Fast Path CI locks
    - 2M + 175 bytes per lock
  - Maximum requirement usually depends on batch and BMP jobs
    - They can hold many, many locks

Coupling Facility Structures

- IRLM lock structure
  - Most installations can use 64M structure
    - 32M lock table + 32M record list
  - What happens if the structure is too small?
    - If the lock table is too small, more false contentions occur
      - Overhead of communications with other systems
    - If the record list is too small, lock requests fail
      - Applications abend

Coupling Facility Structures

- IRLM lock structure placement
  - Never place a lock structure on the same machine with an IRLM using it
    - Unless you are using system managed duplexing for the structure
    - Concurrent failure of IRLM and non-duplexed lock structure causes IMSplex-wide data sharing failure
      - No new locks may be granted until the failed IMS is emergency restarted

Coupling Facility Structures

- VSAM cache structure
  - Entry required for each VSAM CI in a buffer pool
    - Count the number of VSAM buffers in all IMSs
      - Include batch jobs
      - Include Hiperspace buffers (or delete them)
    - Size:
      - 2M + 300 bytes per entry

Coupling Facility Structures

- OSAM cache structure
  - Entry required for each OSAM block in a buffer pool
    - Count the number of OSAM buffers in all IMSs
      - Include batch jobs
      - Include OSAM sequential buffering
    - Space required for any cached blocks
    - Size:
      - 2M + 300 bytes per entry + space for caching
        - Percent of structure used for caching is determined by CFOSAM parameters

Coupling Facility Structures

- OSAM and VSAM cache structures
  - 32M is large enough for 100,000 database buffers
    - 2M + (100,000 x 300)
  - What happens if the structure is too small?
    - If the structure is too small, IMS database buffers are invalidated
      - Blocks or CIs must be reread

Coupling Facility Structures

- DEDB VSO cache structures
  - Entry required for each CI in direct portion of PRELOAD area
    - CI0 and REORG UOW are not stored in the structure
  - User determines size for non-PRELAD areas
    - Depends on the amount of data wanted in the cache structure
  - 2M + 300 bytes per entry + space to hold CIs

- Duplexing of VSO structures is recommended
  - Improves availability
  - Loss of structure without duplexing causes area outage
    - Area must be recovered
  - IMS V9 shared structures use system-managed duplexing

IRLM Execution Parameters

- IRLMID=
  - 1 to 256
  - Each IRLM in data sharing group must have a unique value
- IRLMNM=
  - 4 byte subsystem name
  - Must be unique on an LPAR
    - Typically, there is only one IRLM on an LPAR
  - All IRLMs can (probably should) have the same name
    - Allows IMS restarts to be done on any LPAR without changing execution parameters
    - Allows IMS batch jobs to be run on any LPAR without changing execution parameters

IRLM Execution Parameters

- SCOPE= execution parameter
  - LOCAL
    - One IRLM, lock structure is not used
  - GLOBAL
    - Multiple IRLMs allowed, lock structure is used
  - NODISCON
    - Same as GLOBAL but IRLM does not disconnect from the lock structure when it has no IMSs connected to it
    - Recommended
      - Especially valuable with sharing of IMS batch jobs

IRLM Execution Parameters

- DEADLOK='lll,ggg' execution parameter
  - lll
    - Number of milliseconds or seconds between deadlock detection cycles
      - Values from 1 to 5 are seconds
      - Values from 100 to 5000 are milliseconds
      - Values from 6 to 99 are converted to 5 seconds
      - 1 is a reasonable value for most installations
  - ggg
    - Number of local deadlock detection cycles in a global cycle
      - Value is ignored
      - Every local cycle is also a global cycle

IRLM Execution Parameters

- IRLMGRP=
  - XCF group name for the IRLMs
    - All IRLMs must have the same value
    - Does not have to be defined to MVS
- LOCKTABL=
  - Ignored if CFNAMES statement is in DFSVSMxx for IMS
  - Specifies the IRLM lock structure name

IRLM Execution Parameters

- MAXUSRS= number of IRLMs
  - Determines size of lock table entries in lock structure
    - 1-6 have the same meaning - 2 byte entries
    - 7-22 have the same meaning - 4 bytes entries
    - 23-32 have the same meaning - 8 bytes entries
  - If 7th IRLM is started with MAXUSRS < 7, structure is rebuilt with larger entry
  - If 23rd IRLM is started with MAXUSRS < 23, structure is rebuilt with larger entry
  - Always specify the maximum number of IRLMs to be used
  - Do not specify > 6 if no more than 6 IRLMs will be used
  - Do not specify > 22 if no more than 22 IRLMs will be used

IRLM Execution Parameters

- LTE=
  - Number of lock table entries in units of 1 meg
    - Must be specified as a power of 2
    - Defaults to half of the space in the lock structure
    - Size of lock table entries is determined by MAXUSRS
- PGPROT=YES or NO
  - IRLM common storage load modules placed in MVS page protected storage
- TRACE=YES or NO
  - Default is NO
  - May be turned on by command
    - Only use trace when it is necessary

IRLM Execution Parameters

- PC=NO or YES
  - NO (ignored by IRLM 2.2)
    - Uses slightly less CPU
    - Lock information in ECSA
  - YES (always used by IRLM 2.2)
    - Uses slightly more CPU
    - Lock information in IRLM extended private
      - Less likely to fail due to out-of-space reasons
- MAXCSA= (ignored by IRLM 2.2)
  - 1M to 999M
  - Limits CSA and ECSA usage with PC=NO
  - IRLM uses approximately 250 bytes per lock

Commands to Modify IRLM Parameters

- Change deadlock detection cycle time
  - MODIFY irlmproc,SET,DEADLOCK=nnnn
- Change number of lock table entries on next connect to the lock structure
  - MODIFY irlmporc,SET,LTE=nnnn
- Change maximum (E)CSA usage
  - MODIFY irlmproc,SET,CSA=nnnn
  - Not used with IRLM 2.2

Lock Contention

- Applications which run well without data sharing usually run well with data sharing
  - The exceptions are discussed later
- Applications which have lock contention without data sharing almost always run worse with data sharing
  - Locks are held a bit longer and there are more locks

Lock Contention

- Fast Path locks
  - Same locks with or without data sharing
    - CI
    - UOW
  - Typically, no new contention
- Full function locks
  - Database record locks
    - With and without data sharing
  - Block locks
    - Only for updates to blocks
    - Only for data sharing
  - Busy locks
    - Open, close, data set extension, KSDS updates

IMS Monitor Reporting of Lock Waits

- Reports lock waits in Program I/O report
  - "PI" and database name in "DDN/FUNC" column
    - "PI" is used even though the IRLM is the lock manager

```
IMS MONITOR ****PROGRAM I/O**** TRACE START 2003 092 14:00:18 TRACE STOP 2003 022 14:02:20
                                          .........IWAIT TIME..........
PSBNAME  PCB NAME      IWAITS        TOTAL        MEAN     MAXIMUM   DDN/FUNC  MODULE


PROGDE1A TRMNALDA          20        62468        3123        6825   TRMNALDA   VBH
                            1       275811      275811      275811 PI TRMNALDA....

        PCB TOTAL          21       338279       16108
```

  - REGION IWAIT report also contains lock wait information
    - Reported as "PI" in FUNCTION column

Deadlock Report

- DEADLOCK report
  - When a deadlock occurs, IMS and IRLM gather information
    - Information is written on the log of the "victim"
- DFSERA10 utility with DFSERA30 exit creates reports of all deadlocks on a log
  - SYSIN control statements:

```
//SYSIN *
OPTION   PRINT  OFFSET=5,FLDLEN=2,FLDTYP=X,VALUE=67FF,COND=M
OPTION   PRINT  OFFSET=33,FLDLEN=8,FLDTYP=C,VALUE=DEADLOCK,COND=E,     X
                EXITR=DFSERA30
END
/*
```

Deadlock Report

- Sample report:

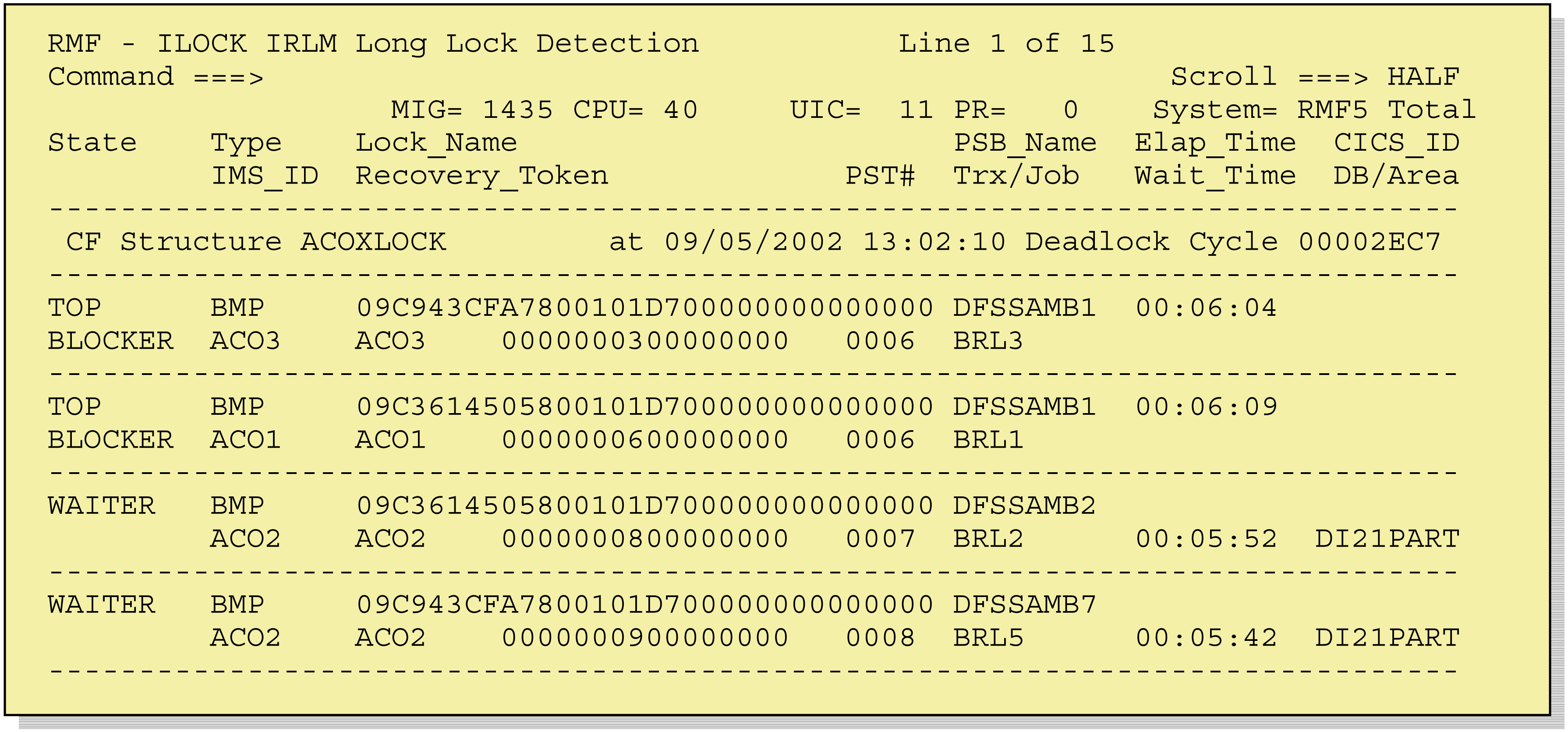
```
DEADLOCK ANALYSIS REPORT - LOCK MANAGER IS IRLM


   RESOURCE DMB-NAME LOCK-LEN LOCK-NAME     - WAITER FOR THIS RESOURCE IS VICTIM
   01 OF 02 DHVNTZ02    08    00000BC4800501D7

   KEY IS ROOT KEY OF DATA BASE RECORD ASSOCIATED WITH LOCK
   KEY=(KK360)

          IMS-NAME TRAN/JOB PSB-NAME PCB--DBD PST# RGN  CALL LOCK  LOCKFUNC STATE
   WAITER IMS4     NQF1     PMVAPZ12 DLVNTZ02 0002 MPP  GET  GRIDX 30400358  06-P
   HOLDER IMS3     DDLKBMP1 PLVAPZ22 -------- 0003 BMP  ---- ----- --------  06-P



   RESOURCE DMB-NAME LOCK-LEN LOCK-NAME
   02 OF 02 DHVNTZ02    08    000009248005O1D7

   KEY IS ROOT KEY OF DATA BASE RECORD ASSOCIATED WITH LOCK
   KEY=(KK130)

          IMS-NAME TRAN/JOB PSB-NAME PCB--DBD PST# RGN  CALL LOCK  LOCKFUNC STATE
   WAITER IMS3     DDLKBMP1 PLVAPZ22 DLVNTZ02 0003 BMP  GET  GRIDX 30400358  06-P
   HOLDER IMS4     NQF1     PMVAPZ12 -------- 0002 MPP  ---- ----- --------  06-P

   DEADLOCK ANALYSIS REPORT - END OF REPORT
```

RMF IRLM Long Lock Detection Report

- RMF reports lock waits greater than specified time
  - MODIFY irlmproc,SET,TIMEOUT=nnnn,ssname
    - nnnn is 1 to 3600 seconds
    - ssname is IMS subsystem name
    - "Timeout" does not cause lock wait to end
      - It only reports that a long wait has occurred
  - Monitor II ILOCK report
  - Uses SMF record type 79 subtype 15),
    - Specify:
      - S RMF,,,(SMFBUF(RECTYPE(70:78,79(15))))

RMF IRLM Long Lock Detection Report

- Sample report:



```
RMF - ILOCK IRLM Long Lock Detection           Line 1 of 15
Command ===>                                                  Scroll ===> HALF
                   MIG= 1435 CPU= 40     UIC=  11 PR=   0    System= RMF5 Total
State    Type    Lock_Name                        PSB_Name  Elap_Time  CICS_ID
         IMS_ID  Recovery_Token             PST#  Trx/Job   Wait_Time  DB/Area
------------------------------------------------------------------------------
 CF Structure ACOXLOCK         at 09/05/2002 13:02:10 Deadlock Cycle 00002EC7
------------------------------------------------------------------------------
TOP      BMP     09C943CFA7800101D700000000000000 DFSSAMB1  00:06:04
BLOCKER  ACO3    ACO3    0000000300000000   0006  BRL3
------------------------------------------------------------------------------
TOP      BMP     09C3614505800101D700000000000000 DFSSAMB1  00:06:09
BLOCKER  ACO1    ACO1    0000000600000000   0006  BRL1
------------------------------------------------------------------------------
WAITER   BMP     09C3614505800101D700000000000000 DFSSAMB2
         ACO2    ACO2    0000000800000000   0007  BRL2      00:05:52  DI21PART
------------------------------------------------------------------------------
WAITER   BMP     09C943CFA7800101D700000000000000 DFSSAMB7
         ACO2    ACO2    0000000900000000   0008  BRL5      00:05:42  DI21PART
------------------------------------------------------------------------------
```

PSB PROCOPTs

- E (exclusive)
  - Exclusive scheduling within an IMS online subsystem
  - Locking for data sharing is done
- A (update and read with integrity)
  - Database record lock for updates held until sync point
- G (read with integrity)
  - Use when possible
- GO (read without integrity)
  - No locking
  - Increased exposure to wrong data
  - Increased exposure to abends

Application Considerations

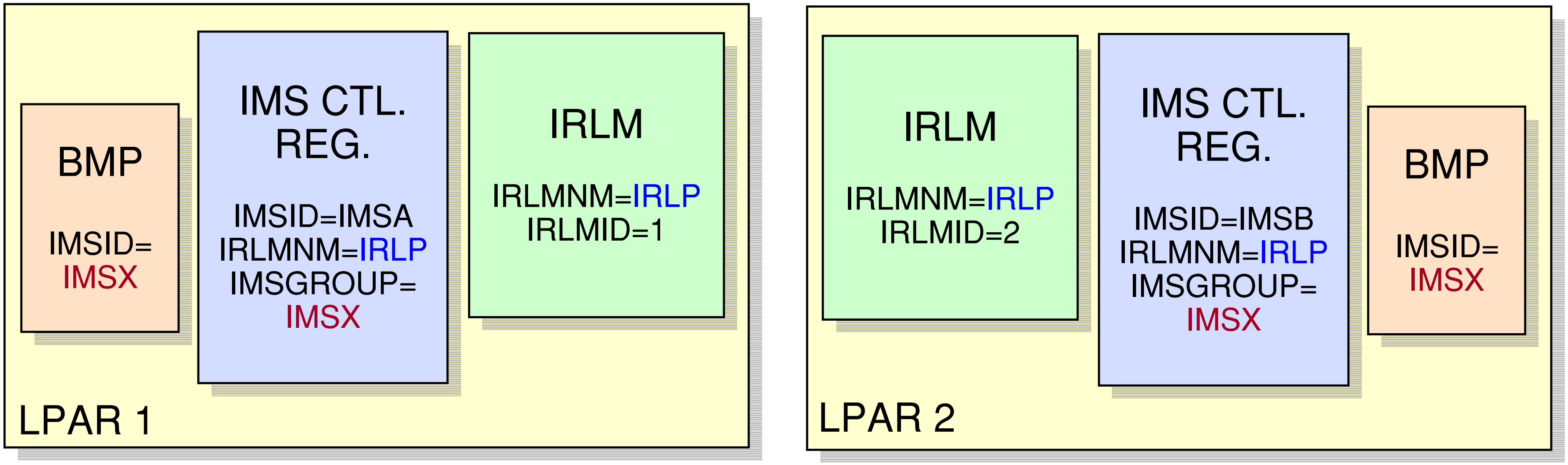
- Typical locking and invalidation problems
  - Application control records
  - Next invoice number, next order number, etc.
  - These can be a serialization problem without data sharing
    - Data sharing makes this worse
- Hot spots
  - Frequently updated blocks
    - Very small database with many updates
    - Frequent inserts to databases without free space
      - All inserts go to the end of the database
    - Keys based on current time
      - Often a problem with secondary indexes
    - Empty (P)HIDAM databases
      - New records are always added at the end of the database

OSAM Caching

- OSAM blocks may be cached in structure
  - Caching by OSAM pool
    - All data sets in the pool are cached
  - Options:
    - Cache all blocks read
    - Cache only blocks which are updated
  - Overhead
    - Writes to cache structure
  - Recommendation:
    - Use only when invalidations are a problem
      - Such as "hot spots"

Availability and Ease of Operations

- Use same IRLM name for all IRLMs
  - Allows IMS to run on any MVS system with an IRLM
    - No JCL or execution parameter changes required
- Use IMSGROUP= for control regions
  - Give all control regions the same IMSGROUP name
  - Allows any dependent region (BMP) to run on any IMS
    - No JCL or execution parameter changes required



Database Recovery

- Database recovery must merge logs
  - Change Accumulation before Database Recovery utility

  or

  - Database Recovery Facility (DRF) tool
    - Merges logs automatically

- Disaster recovery
  - Test your procedures
  - RSR and DRF have good capabilities with data sharing

Failure Recovery

- Many new recovery scenarios
  - CF failures, CF link failures, IRLM failures, IMS subsystem failures, etc.
  - Plan recovery procedures
  - Test recovery procedures

- Recover as quickly as possible
  - Requests for locks held by a failed subsystem are rejected
    - Requester abends
  - FDBR backs out in-flight updates and releases locks very quickly

BMPs and Batch Jobs

- BMPs are usually preferred
  - BMP abends are backed out automatically
    - Do not cause lock rejects
  - Batch (DLI and DBB) abends are not backed out automatically
    - Cause lock rejects until back out completes
    - May cause multiple abends
  - BMPs use online log
    - Makes log management simpler

- You may keep a batch window without data sharing
  - Batch (DLI and DBB) jobs without data sharing
    - No IRLM
    - Will get exclusive authorization from DBRC

Data Sharing Implementation Steps

- Potential Steps (steps may be combined):

1. Register Databases SHARELVL(1)
2. Define ACCESS=UP for databases
3. VSAM SHAREOPTION(3 3) and DISP=SHR
4. IRLM with SCOPE=LOCAL
5. Define structures in CFRM policy
6. CFNAMES statement
   - Each FF DB I/O requires CF access for OSAM/VSAM structures
7. IRLM SCOPE=NODISCON
8. Register Databases SHARELVL(3)
   - All locks are placed in the lock structure
9. Establish second IMS subsystem
   - Buffer invalidations may occur and lock conflicts may increase

This list does not include testing and procedure changes!

These steps add overhead.

Things to Remember

- Performance
  - Most (99% ?) applications run very well with data sharing
    - Applications which run poorly without data sharing will usually run worse with data sharing
  - Many DB performance problems may be addressed by DBAs
    - More free space, spreading of data over more blocks, ...

- Size CF structures carefully
  - Use CFSIZER tool on the Web

More Information

- Redbooks
  - IMS in the Parallel Sysplex
    - *Volume I: Reviewing the IMSplex Technology*
      - SG24-6908
    - *Volume II: Planning the IMSplex*
      - SG24-6928
    - *Volume III: IMSplex Implementation and Operations*
      - SG24-6929