IDUG

CCDUG
2023

Partitioning Advances: PBR and PBR RPN

Frank Rhodes

BMC Software

#IDUGDb2

Agenda

- History of Db2 Tablespace Structures
- Absolute Page numbering
- Partition By Range (PBR)
- PBR Relative Page Numbering (RPN)
- Impacts to page layouts and RIDs

History of Db2 Tablespace Structures

- Simple Tablespaces
- Partitioned Tablespace (Index controlled partitioning)
- Segmented Tablespace
- Large Partitioned Tablespaces
- Partitioned Tablespace (Table controlled partitioning)
- Universal Tablespaces
- Partition by Range using Relative page numbering

Simple Tablespaces

- Part of the first release of Db2
- Can hold 64GB of data
- Can hold multiple tables
- You can have multiple tables data stored on the same page
  - OBID (table identifier) is stored in each row header
- As rows are deleted and inserted, things can get messy
- Managing the Space maps becomes more complicated

- Deprecated with Db2 V9

Partitioned Tablespace

- Part of first release of Db2
- Hold a single table
- Partitioning is controlled by an Index (ICP)
- Maximum number of partitions
  - Originally of 64 parts
  - Increased to 254 with DB2 V5
  - Increased to 4096 with Db2 V8
- Today the max size is 128TB (This requires use of DSSIZE)
- PREVENT_NEW_IXCTRL_PART ZPARM could be set to prevent use of ICP with Db2 V11
- Deprecated with Db2 V12 M504

Segmented Tablespace

- Introduced with Db2 V2.1
- Can hold 64GB of data
- Can hold multiple tables
- Defined with a SEGSIZE referring to the number of pages in a segment
- Within one segment, only rows for one table are stored
- Deprecated with Db2 V12 M504

Large Partition Tablespace

- Introduced with Db2 V5.1
- Hold a single table
- Can hold 16TB of data
- Maximum number of partitions is 4096(Db2 V8)
- Maximum size of a partition is 4 GB

- Came out in V5, obsoleted with V6

Partitioned Tablespace (Table controlled partitioning)

- Introduced with Db2 V8.1
- Hold a single table
- Maximum of 4096
- Supports ADD PART and ROTATE PART
- Supports DPSIs

- OLS Conversion from ICP to TCP provided
- Deprecated with Db2 V12 M504

Universal Tablespaces (Partition by Growth)

- Introduced with Db2 V9.1
- No partitioning keys
- Holds a single table
- Define the maximum number of partitions the tablespace can grow to
- Use SEGSIZE and DSSIZE
- OLS supported to convert Segmented TS to PBG
- If multiple tables in Segmented TS can use MOVE command

Universal Tablespaces (Partition by Range)

- Introduced with Db2 V9.1
- Uses Absolute page numbering
- Holds a single table
- Use SEGSIZE and DSSIZE
- OLS supported to convert from TCP to PBG

PBR2 – Partition by Range using Relative page numbering

- Introduced with Db2 V12.1

Absolute page numbering for partitioned TSs

Page numbering has always been **absolute** within Db2.

- The internal page numbering is kept as a 4-byte value that includes a partition number and page number. Distinguishing which bits represent the partition and which represent the page number requires a shift value. The shift value is LOG base 2 (DSSIZE/(page -size)).

This is why DSSIZE, Page-size and Numparts have dependencies on each other

- The more parts a TS has, the fewer number of pages the TS can have. This in turn can result is a smaller DSSIZE for the TS partition

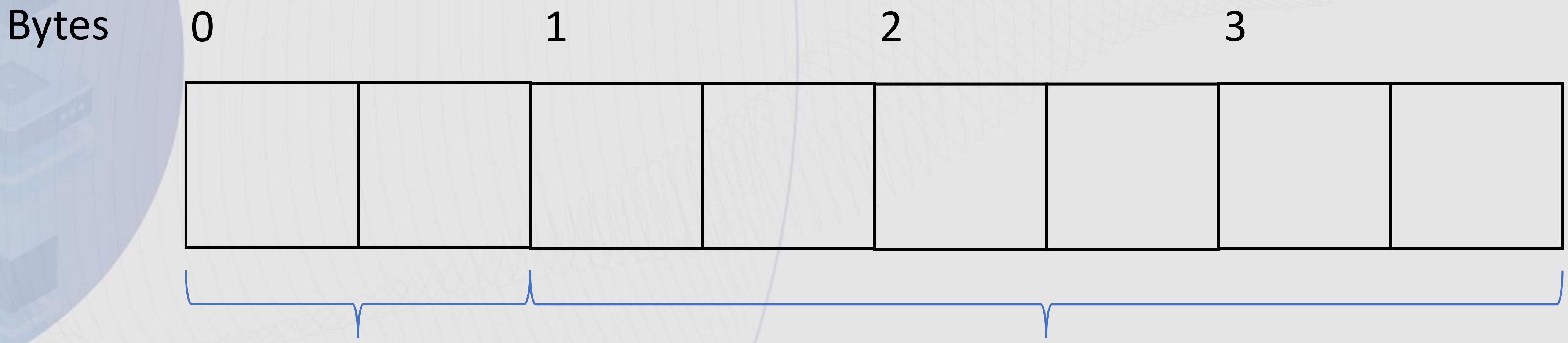
Maximum MAXPARTITIONS value for a given page size and DSSIZE value

Taken from Db2 V12 SQL Reference

DSSIZE value	**4K page size**	**8K page size**	**16K page size**	**32K page size**
1 G - 4 G	4096	4096	4096	4096
8 G	2048	4096	4096	4096
16 G	1024	2048	4096	4096
32 G	512	1024	2048	4096
64 G	254	512	1024	2048
128 G	128	256	512	1024
256 G	64	128	256	512

4 Byte Page Number in More Detail

```
CREATE TABLESPACE MYDB.MYTS… NUMPARTS 256 PGSIZE 4K
SEGSIZE 0 DSSIZE 64G
```



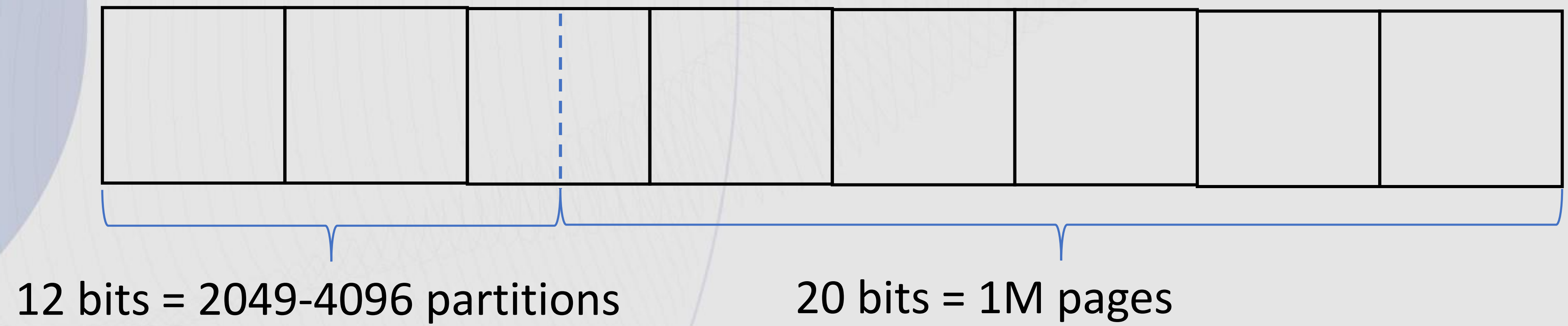
Bytes 0 1 2 3

8 bits = 1-256 partitions

24 bits = 16M pages = 64GB
(128GB for 8K page size)

Pop Quiz

- What should the maximum DSSIZE be, for a 3000 part 16K tablespace?



12 bits = 2049-4096 partitions

20 bits = 1M pages

1M pages * 16K gives a max DSSIZE of 16GB

Introduction of Partition by Range Universal Tablespace

Based upon the use of Table Controlled Partitioning

- Tablespace is created with NUMPARTS and SEGSIZE
  - This tells you that the TS is Universal and will be used for PBR
- Regular PBR TSs use absolute page numbering
- Index Controlled and Table controlled partitioned Tablespaces have SEGSIZE set to zero.

Partition by Range and Online Schema support

To convert Table Controlled partitioned TS to PBR

  - ALTER TS <MyDB>.<MyTS> SEGSIZE ##
  - It is a pending DDL ALTER, a REORG needed to materialize the change

- DSSIZE
  - Must be between 1G -256G based on a power of 2. (1,2,4,8,16 …)
  - Size must increase
  - Applies to all partitions in the tablespace
  - To specify a value greater than 4G, the data sets for the table space must be associated with a DFSMS data class that has been specified with extended format and extended addressability.

- Member Cluster

Relative page numbering for PBR

Introduced with Db2 V12

- Indicates that internal page numbering is kept as a 4-byte value without a partition number. The page number is a relative page from the start of the partition, and the partition number is kept only in the header page. When PAGENUM RELATIVE is specified, the data sets for the table space must be associated with a DFSMS data class that is specified with extended format and extended addressability.
- I refer to Partition by Range tablespaces using relative page numbering as PBR2

DB2 Catalog changes to support Relative page numbering

- SYSIBM.SYSTABLESPACE
  - PAGENUM
- SYSIBM.SYSTABLEPART
  - PAGENUM
  - DSSIZE
- SYSIBM.SYSINDEXES
  - DSSIZE
  - PAGENUM
- SYSIBM.SYSINDEXPART
  - DSSIZE
  - PAGENUM

DSNZPARM control of Relative page numbering

ZPARM: PAGESET_PAGENUM specifies whether partition-by-range table spaces and associated partitioned indexes are created to use absolute page numbers across partitions or relative page numbers.

DSNZPxxx: DSN6SPRM.PAGESET_PAGENUM
Acceptable values: ABSOLUTE or RELATIVE
Default: ABSOLUTE for Db2 V12 and RELATIVE for Db2 V13

Relative page numbering for PBR Attribute and size Maximums

- Maximum NUMPARTS is 4096.

- Maximum DSSIZE is 1024G.
  - DSSIZE can be specified at the partition level vs. just the TS level
  - DSSIZE is any value between 1G-1024G, default is 4G. Does **not** have to be a multiple of 2.
  - DSSIZE is **not** limited by pagesize or number of partitions
  - NOTE: If DSSIZE specified at Part level or if DSSIZE over 4G, you must have use DFSMS data class that has been specified with extended format and extended addressability

- Maximum Table size is a theoretical 4 Peta bytes

Relative page numbering for PBR DSSIZE Alters

- ALTER DSSIZE to a larger value are immediate alters
  - Alter at the PART level or the Tablespace level
- ALTER DSSIZE to a value smaller value will cause a pending DDL if any partitions DSSIZE is reduced

Relative page numbering for Indexes

- Only applies if the base Tablespace is using RPN

- DSSIZE is supported
  - Not valid on nonpartitioned secondary indexes
  - Any integer between 1G-1024G, default is 4G. Does not have to be a multiple of 2.
  - For DSSIZE greater that 4G, the data sets for the table space must be associated with a DFSMS data class that has been specified with extended format and extended addressability

- If the index is a partitioned index using relative page numbering, the value of DSSIZE for a particular partition is given by the first of these choices that applies:
  - The value of DSSIZE given in the PARTITION clause for that partition.
  - The value given by a DSSIZE keyword that is not in any PARTITION clause.
  - The default value is inherited from the base table space.

Relative page numbering for Indexes

- ALTER DSSIZE to a larger value are immediate alters
  - Alter at the PART level or the Index level

- ALTER DSSIZE to a smaller value is not allowed if the datasets have already been created.

Relative page numbering in Table Create

Only valid if creating table with implicit tablespace

- Can specify PAGENUM and DSSIZE along with most anything you can specify in the USING block of a CREATE.
  - STOGROUP
  - BUFFERPOOL
  - . . .
- If you do not explicitly specify DSSIZE or PAGENUM you will get whatever is in ZPARM.
  - This may not be what you are expecting
- Minimum record length for table data is 3 bytes if using RPN
  - I am not sure why, but it would not make sense to adopt RPN to handle large amounts of data, if you will have row lengths less than 3 bytes.

Relative page numbering for AUX Tablespaces

- XML Tablespace will inherit from the base tablespace.
  - If Base TS is using PBR RPN, the XML TS will also use RPN
  - The XML TS will use the DSSIZE from the Base TS, but can be altered after it is created
- LOB Tablespaces remain unchanged. Still need one LOB TS per partition per column
  - They will remain using PAGENUM ABSOLUTE

Converting a PBR to a PBR2

ALTER PAGENUM for a PBR tablespace to RELATIVE

- This will be a pending DDL alter.
- REORG required to materialize the ALTER

You cannot ALTER PAGENUM to ABSOLUTE.

- If you want to change PAGENUM RELATIVE to ABSOLUTE, you will need to drop recreate your object

Internals for Regular Partitioned Tablespace

```
PARTITION: # 0002
PAGE: # 00100000
HEADER PAGE:PGCOMB='10'X  PGBIGRBA='00000000000000000000'X  PGNUM='00100000'X  PGFLAGS='18'X
  HPGOBID='94220008'X  HPGHPREF='00100004'X  HPGCATRL='00'X  HPGREL='Q'  HPGZLD='.’
  HPGCATV='00'X  HPGTORBA='000000000000'X  HPGTSTMP='20220812143032024960'X
  HPGSSNM='DEJM'  HPGFOID='0007'X  HPGPGSZ='1000'X  HPGSGSZ='0000'X  HPGPARTN='0003’
  HPGZ3PNO='000000'X  HPGZNUMP='00'X  HPGTBLC='0001'X  HPGROID='0009'X
  HPGZ4PNO='00000000'X  HPGMAXL='00CD'X  HPGNUMCO='0005'X  HPGFLAGS='0008'X
  HPGFLAGS2='00'X  HPGFLAGS3='80'X  HPGCONTM='20220812143047462751'X
  HPGSGNAM='DEMSGSMS'  HPGVCATN='DEJMCAT '  HPGRBRBA='000000000000'X
  HPGLEVEL='000000000000'X  HPGPLEVL='000000000000'X  HPGCLRSN='000000000000'X
  HPGSCCSI='0025'X  HPGDCCSI='0000'X  HPGMCCSI='0000'X  HPGPARTNUM='0002'X
  HPGDSSZ='00400000'X  HPGFLAG2='00'X  HPGEPOCH='0001'X  HPGRBLP='000000000000'X
  HPGDNUMB='01'X  HPGDNUMC='0100'X  HPGDFSG='00000000'X  HPGDLSG='00000000'X
  HPGSISP='00000000'X  HPGBIGTORBA='00000000000000000000'X
  HPGBIGRBRBA='000000003101B2CCC9C6'X  HPGBIGLEVEL='000000003101B2CCC9C6'X
  HPGBIGPLEVL='000000003101B2CB6EE1'X  HPGBIGCLRSN='000000003101B2CBD395'X
  HPGBIGRBLP='00000000000000000000'X  FOEND='52'X
  DVI HASH BUCKET:  HPGDBKT#='01'X  HPG1BEYE='4E'X
    HPGDCOLL#='0100'X  HPG1OBID='0009'X  HPG1V='00'X  HPG1RID='0010000201'X
```

Internals for Partitioned by Range

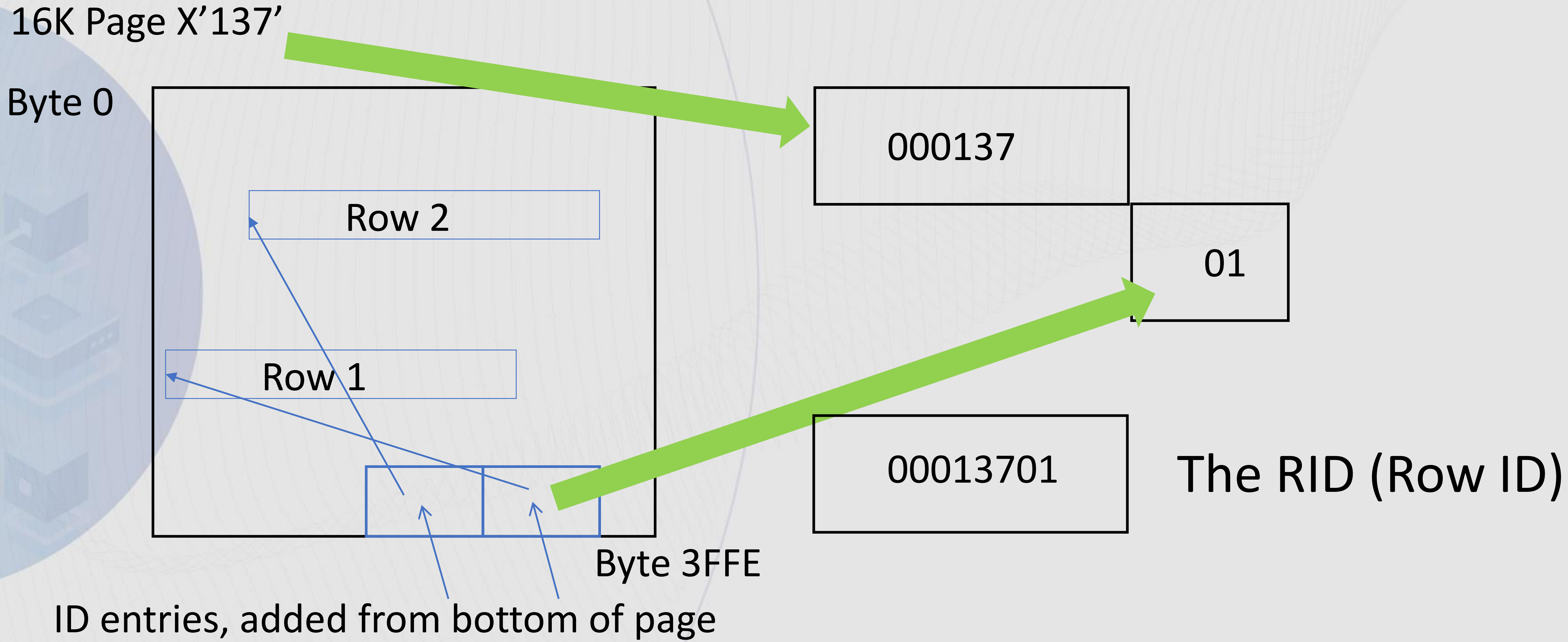
```
PARTITION: # 0002
PAGE: # 00100000
HEADER PAGE: PGCOMB='10'X  PGBIGRBA='00000000000000000000'X  PGNUM='00100000'X  PGFLAGS='38'X
  HPGOBID='9E420002'X  HPGHPREF='00100042'X  HPGCATRL='00'X  HPGREL='Q'  HPGZLD='.’
  HPGCATV='00'X  HPGTORBA='000000000000'X  HPGTSTMP='20220812135704578999'X
  HPGSSNM='DEJM'  HPGFOID='0001'X  HPGPGSZ='1000'X  HPGSGSZ='0020'X  HPGPARTN='0003’X
  HPGZ3PNO='000000'X  HPGZNUMP='00'X  HPGTBLC='0001'X  HPGROID='0003'X
  HPGZ4PNO='00000000'X  HPGMAXL='00CD'X  HPGNUMCO='0005'X  HPGFLAGS='010C'X
  HPGFLAGS2='00'X  HPGFLAGS3='80'X  HPGCONTM='20220812135706273079'X
  HPGSGNAM='SYSDEFLT'  HPGVCATN='DEJMCAT '  HPGRBRBA='000000000000'X
  HPGLEVEL='000000000000'X  HPGPLEVL='000000000000'X  HPGCLRSN='000000000000'X
  HPGSCCSI='0025'X  HPGDCCSI='0000'X  HPGMCCSI='0000'X  HPGPARTNUM='0002'X
  HPGDSSZ='00400000'X  HPGFLAG2='00'X  HPGEPOCH='0001'X  HPGRBLP='000000000000'X
  HPGMASSDELETETIMESTAMP='000000000000'X  HPGDNUMB='01'X  HPGDNUMC='0100'X
  HPGDFSG='00000001'X  HPGDLSG='00000001'X  HPGSISP='00000000'X
  HPGBIGTORBA='00000000000000000000'X  HPGBIGRBRBA='000000003101AB0FC571'X
  HPGBIGLEVEL='000000003101AB0FC571'X  HPGBIGPLEVL='000000003101AB0F10B8'X
  HPGBIGCLRSN='000000003101AB0F185E'X  HPGBIGRBLP='00000000000000000000'X
  HPGBIGMASSDELETETIMESTAMP='00000000000000000000'X  FOEND='52'X
  DVI HASH BUCKET:  HPGDBKT#='01'X  HPG1BEYE='4E'X
    HPGDCOLL#='0100'X  HPG1OBID='0003'X  HPG1V='00'X  HPG1RID='0010000201'X
  SI HASH BUCKET:  HPGDBKT#='01'X  HPGS1BEYE='E2'X  HPGS1OBI='0003'X  HPGS1FSG='00000002'X

                HPGS1CSG='00000002'X  HPGS1LSG='00000002'X
```

Internals for Partitioned by Range Relative Page numbering

```
PARTITION: # 0002
PAGE: # 00000000 -
HEADER PAGE: PGCOMB='10'X  PGBIGRBA='00000000000000000000'X  PGNUM='00000000'X  PGFLAGS='38'X
 HPGOBID='9E420005'X  HPGHPREF='00000042'X  HPGCATRL='00'X  HPGREL='Q'  HPGZLD='.’
 HPGCATV='00'X  HPGTORBA='000000000000'X  HPGTSTMP='20220812135704578999'X
 HPGSSNM='DEJM'  HPGFOID='0004'X  HPGPGSZ='1000'X  HPGSGSZ='0020'X  HPGPARTN='0003’
 HPGZ3PNO='000000'X  HPGZNUMP='00'X  HPGTBLC='0001'X  HPGROID='0006'X
 HPGZ4PNO='00000000'X  HPGMAXL='00CD'X  HPGNUMCO='0005'X  HPGFLAGS='010C'X
 HPGFLAGS2='08'X  HPGFLAGS3='80'X  HPGCONTM='20220812135706796546'X
 HPGSGNAM='DEMSGSMS'  HPGVCATN='DEJMCAT '  HPGRBRBA='000000000000'X
 HPGLEVEL='000000000000'X  HPGPLEVL='000000000000'X  HPGCLRSN='000000000000'X
 HPGSCCSI='0025'X  HPGDCCSI='0000'X  HPGMCCSI='0000'X  HPGPARTNUM='0002'X
 HPGDSSZ='00400000'X  HPGFLAG2='00'X  HPGEPOCH='0001'X  HPGRBLP='000000000000'X
 HPGMASSDELETETIMESTAMP='000000000000'X  HPGDNUMB='01'X  HPGDNUMC='0100'X
 HPGDFSG='00000001'X  HPGDLSG='00000001'X  HPGSISP='00000000'X
 HPGBIGTORBA='00000000000000000000'X  HPGBIGRBRBA='000000003101AB126F72'X
 HPGBIGLEVEL='000000003101AB126F72'X  HPGBIGPLEVL='000000003101AB11BB14'X
 HPGBIGCLRSN='000000003101AB11C2EE'X  HPGBIGRBLP='00000000000000000000'X
 HPGBIGMASSDELETETIMESTAMP='00000000000000000000'X  FOEND='52'X
 DVI HASH BUCKET:  HPGDBKT#='01'X  HPG1BEYE='4E'X
  HPGDCOLL#='0100'X  HPG1OBID='0006'X  HPG1V='00'X  HPG1RID='0000000201'X
 SI HASH BUCKET:  HPGDBKT#='01'X  HPGS1BEYE='E2'X  HPGS1OBI='0006'X  HPGS1FSG='00000002’X
        HPGS1CSG='00000002'X  HPGS1LSG='00000002'X
```

RIDs and how they have changed



RIDs increase in size to allow for more data

RIDs increased to 5 bytes to support 4096 parts

- 4 bytes for left justified partition, rest for page number
- 1 byte for ID

RID have increased to 7 bytes

- 2 bytes for part number
- 4 bytes for page number
- 1 byte for ID

CCDUG

Partitioning Advances:
PBR and PBR RPN

Frank Rhodes
BMC Software
frank_rhodes@bmc.com

Please fill out your session evaluation!