



z/OS® Operating System

SAM extended format and compression

@business on demand software

© 2008 IBM Corporation

Extended format

What are extended format data sets?

- Type of SMS-managed physical sequential data set.
- Same characteristics as a physical sequential data set.
- Internally, the data is stored in a different format.
- Sequential data striping is implemented through the use of extended format data sets.

Extended format data set can support data striping and compression.

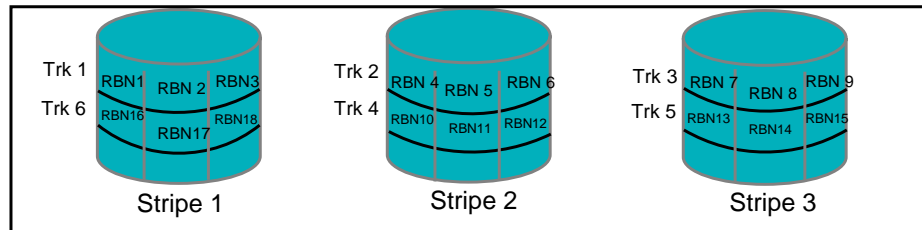
It is a type of SMS-managed physical sequential data set.

Externally, extended format data sets have the same characteristics as sequential data sets. However, records are not necessarily stored in the same format or order as they appear.

Sequential data striping is implemented through the use of extended format data sets. Data is interleaved across multiple volumes for striped data sets.

Extended format

Multi-Striped (Multi-volume) One logical volume



Single Stripe, Multi-Volume

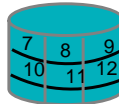
Data Class

DSNTYPE=Ext



Storage Class

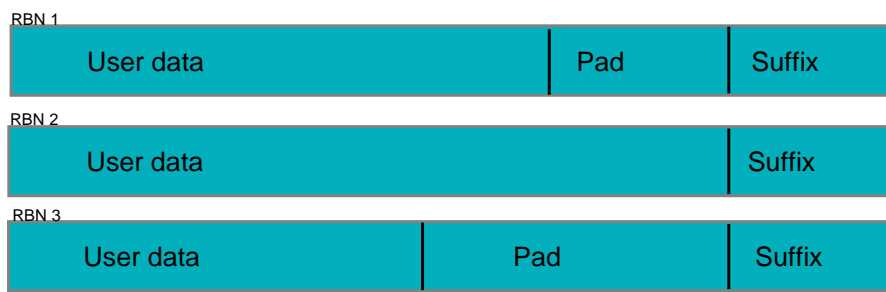
Sustained Data Rate=nnn



*/*where nnn is used to derive the number of stripes*/*

- Data striping distributes data for one data set across multiple volumes.
- A data set can have a maximum of 59 stripes.
- Each stripe must reside on one volume and cannot be extended to another volume.
- I/O can be done in parallel with striped data for better performance.
- To specify extended format, data set type (DSNTYPE) in data class needs to be set to 'Extended', and sustained data rate in storage class needs to be set to the number of stripes needed.
- Sustained data rate can be set to a maximum of 59 stripes. The default is 1, which will give a single striped data set.

Example of blocks in the extended data format



- User data
 - ▶ BSAM or QSAM Block
- Padding
 - ▶ Format 1 DSCB (DS1BLKL) contains user-perceived maximum block size
 - ▶ Above, RBN1 and RBN3 are short user blocks
- Suffix (length = 32 bytes) contains:
 - ▶ Length of user data
 - ▶ Relative block number
 - ▶ 3 bytes to detect control unit padding

Block in the extended data format:

-User data consists of a BSAM or QSAM block.

-Block can be short block.

-RBN1 and RBN3 are short user blocks and require padding up to the beginning of the suffix.

-Suffix (length of 32 bytes) contains length of user data, relative block number, and 3 bytes to detect control unit padding.

Characteristics of extended format data sets

- Maximum of 123 extents on each volume.
- Maximum of 59 volumes.
- Can occupy any number of tracks.
- Striped data set can contain up to 4GB blocks.
- All physical blocks are the same size.
- Each block has a 32-byte suffix

- Extended format sequential data sets have a maximum of 123 extents on each volume.
- Each extended format sequential data set can have a maximum of 59 volumes.
- An extended format data set can occupy any number of tracks.
- An extended format, striped sequential data set can contain up to 4GB blocks.
- All physical blocks in an extended format sequential data set are the same size, but when a program reads a block, the access method returns the length written by the writing program.
- Each block in an extended format data set has 32-byte suffix.

Compressed format

What are compressed format data sets?

- Type of extended format dataset which supports block level compression.
- Same characteristics as extended format sequential data sets.
- Internally, the data within each block is stored in a different format.
- Externally, a compressed format data set:
 - ▶ May consist of one or more stripes
 - ▶ May be multi-volume (single-stripe only)
 - ▶ May not be opened for update

-Compressed format data sets is a type of extended format data set which supports block level compression.

-An extended format data set can be allocated in the compressed format by specifying `COMPACTION=YES` in the data class.

-Internally, data within each block is stored in a different format.

-Externally, a compressed format data set may consist of one or more stripes, may be multi-volume, and may not be open for update.

Compressed format (continued)

Types of compression

- **Generic (DBB-based)**

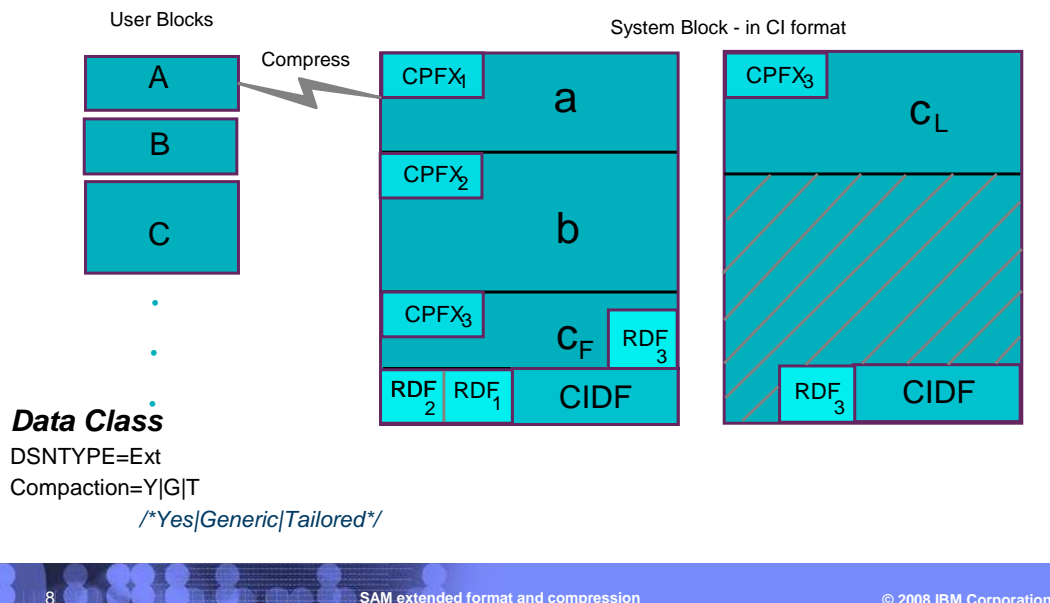
The system selects a set of Dictionary Building Blocks (DBB) which best reflects the initial data written to the data set. The system later can reconstruct the dictionary by using the information in the dictionary token stored in the catalog.

- **Tailored**

The system attempts to derive a compression dictionary tailored specifically to the initial data written to the data set. Once derived, the compression dictionary is stored in system blocks which are imbedded within the data set itself.

- Two compression techniques are available for compressed format data sets.
- They are DBB-based compression and tailored compression.
- **Generic (DBB-based) compression:** The system selects a set of dictionary building blocks (DBBs), which best reflects the initial data written to the data set. The system can later reconstruct the dictionary by using the information in the dictionary token stored in the catalog.
- **Tailored compression:** The system attempts to derive a compression dictionary tailored specifically to the initial data written to the data set. Once derived, the compression dictionary is stored in system blocks which are imbedded within the data set itself.
- The form of compression the system is to use for newly created compressed format data sets can be specified at either or both the data set level and at the installation level.

Compressed format (continued)



-Each block in the CI has an associated compression record prefix (CPFX) and a record definition field (RDF).

-A compression record prefix (CPFX) contains information about the compressed block, such as whether the record is compressed or not, uncompressed length of record, and some flags.

-A record definition field (RDF) contains information about the compressed record, such as whether it is a spanned record, which segment the record is spanned, and record length.

-A CI has an associated control interval definition field (CIDF).

-A control interval definition field (CIDF) contains information about the CI such as the offset to the free space and length of the free space.

Characteristics of compressed format data sets

- Might or might not contain compressed records.
- Data format consists of physical blocks whose length has no correlation to the logical block size of the data set.
- Compressed format data set cannot be opened for update.
- System attempts to derive a compression token when enough data is written to the data set (between 8K and 64K for generic compression and much more for tailored compression).

•A compressed format data set might or might not contain compressed records.

-The data format for a compressed format data set consists of physical blocks whose length has no correlation to the logical block size of the data set in the DCB, DCBE, and the data set label. The actual physical block size is calculated by the system and is never returned to the user. However, the system maintains the user's block boundaries when the data set is created so that the system can return the original user blocks to the user when the data set is read.

-A compressed format data set cannot be opened for update.

-When the data set is created, the system attempts to derive a compression token when enough data is written to the data set (between 8K and 64K for generic compression and much more for tailored compression). If the system is successful in deriving a compression token, the access method attempts to compress any additional records written to the data set. However, if an efficient compression token could not be derived, the data set is marked as non-compressible and there is no attempt to compress any records written to the data set.

References

- Using Data Sets, SC26-7410
- Macro Instructions for Data Sets, SC26-7408
- SAM Logic, ZW84-4030-03 (IBM confidential)

Feedback

Your feedback is valuable

You can help improve the quality of IBM Education Assistant content to better meet your needs by providing feedback.

- Did you find this module useful?
- Did it help you solve a problem or answer a question?
- Do you have suggestions for improvements?

Click to send e-mail feedback:

mailto:iea@us.ibm.com?subject=Feedback_about_V1R0-SAM-Extended-Format-100808.ppt

This module is also available in PDF format at: [../V1R0-SAM-Extended-Format-100808.pdf](..//V1R0-SAM-Extended-Format-100808.pdf)

You can help improve the quality of IBM Education Assistant content by providing feedback.

Trademarks, copyrights, and disclaimers

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM z/OS

A current list of other IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements or changes in the products or programs described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

Information is provided "AS IS" without warranty of any kind. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (for example, IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products.

IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance 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 or performance improvements equivalent to the ratios stated here.

© Copyright International Business Machines Corporation 2008. All rights reserved.

Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.