Managing Very Large Databases

Dick Weiland
CA Technologies

CA IDMS™ Technical Conference

Framingham MA
December 2-5, 2014

Abstract

As CA IDMS database sizes grow, organizations face a number of challenges generated by very large volumes of data. This session examines these challenges and covers strategies for managing these environments, focusing on the use of database segmentation and mixed page group support, as well as new techniques which can be used with release 18.5.

CA IDMS™ Technical Conference

Agenda

- Introduction
- Space
- Time
- Summary

What are the major considerations of a very large database (VLDB)?

- Space
  – The number of pages available to the database

- Time
  – Application processing
  – Maintenance processing

Space

- Page Groups
  - Dividing the areas of the database across multiple pages groups increases the number of pages available to the database

CA IDMS™ Technical Conference

© 2014 CA. ALL RIGHTS RESERVED.

Space (cont.)

- Altering the number of records that can be stored on a database page (radix point)
  - The standard maximum number of user records per page is 255
  - If the database contains a large number of very small records
    - The maximum records per page could be increased up to 2727 to store more occurrences per page
    - This reduces the number of available pages
  - If the database contains extremely large records
    - The number of records per page can be reduced to a maximum of 3
    - This increases the total number of pages that can be defined within the page group

CA IDMS™ Technical Conference

© 2014 CA. ALL RIGHTS RESERVED.

Space
Page Groups

- Page groups and radix points are defined on a segment basis

- All segments within the same page group should have the same radix point

- Mixed Page Group Support
  - Allows a single run-unit to access areas from multiple page groups
  - Enabled when defining the DBNAME to be used for the run-unit

```
CREATE DBNAME R170DBTB.RJW
   MIXED PAGE GROUP BINDS ALLOWED
   INCLUDE SEGMENT RJWSEG
      INCLUDE SEGMENT SYSDICT
   INCLUDE SEGMENT SYSMSG;
```

CA IDMS™ Technical Conference

7   © 2014 CA. ALL RIGHTS RESERVED.

Space
Page Groups

- When using Mixed Page Group Support there are restrictions that must be observed
  - Chained sets and referential constraints can not cross page group boundaries
  - Record names must be unique across all areas accessed under a single run-unit
  - Unqualified FIND/OBTAIN DML commands should be avoided
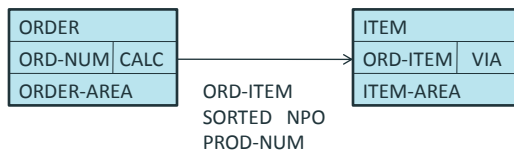
CA IDMS™ Technical Conference

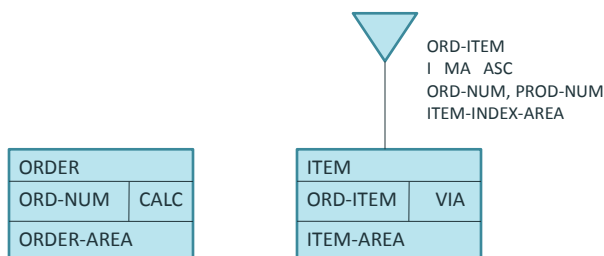8   © 2014 CA. ALL RIGHTS RESERVED.

Space
Page Groups

Scenario 1:

- The following database structure is to be separated into two page groups, each containing a single area

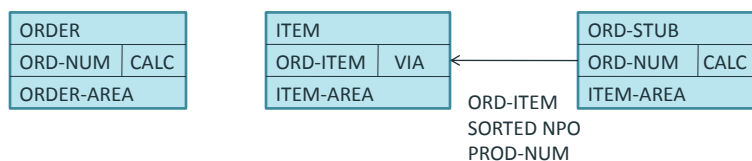- To allow for this change, set ORD-ITEM must be replaced by some other construct

ORDER	
ORD-NUM	CALC
ORDER-AREA	

ORD-ITEM
SORTED NPO
PROD-NUM

ITEM	
ORD-ITEM	VIA
ITEM-AREA	

Space
Page Groups

- Option 1: ensure that the CALC key of the owner exists within the member and replace set ORD-ITEM set with a single system-owned index

ORD-ITEM
I MA ASC
ORD-NUM, PROD-NUM
ITEM-INDEX-AREA

ORDER	
ORD-NUM	CALC
ORDER-AREA	

ITEM	
ORD-ITEM	VIA
ITEM-AREA	

Space
Page Groups

- Option 2: Define a 'stub' record in an area of the ITEM-AREA page group that contains only the CALC key of the owner and use it as the owner of the ORD-ITEM set

ORDER	
ORD-NUM	CALC
ORDER-AREA	

ITEM	
ORD-ITEM	VIA
ITEM-AREA	

ORD-ITEM
SORTED NPO
PROD-NUM

ORD-STUB	
ORD-NUM	CALC
ITEM-AREA	

Space
Page Groups

- Starting with CA IDMS 18.5 it is possible to convert the chain set to a user-owned index set which may cross the boundaries of areas defined in different page groups

- See "How to Expand the Employee Demo Database With Mixed Page Group Index Sets" in CA IDMS Scenario Guide

ORDER	
ORD-NUM	CALC
ORDER-AREA	

ORD-ITEM
SORTED IO
PROD-NUM

ITEM	
ORD-ITEM	VIA
ITEM-AREA	
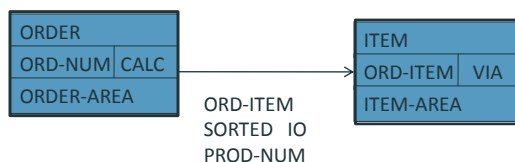
Space
Page Groups

- All record names accessed by a run-unit must be unique
  (as with other non-SQL defined databases)
  - May become an issue when Mixed Page Group Support is used with a
    segmentation strategy

- Use of unqualified FIND/OBTAIN DB-KEY commands can lead
  to erroneous results

Change	**OBTAIN DB-KEY IS WS-DBKEY**
To	OBTAIN ITEM DB-KEY IS WS-DBKEY
Or	OBTAIN DB-KEY IS WS-DBKEY PAGE-INFO WS-PAGE-INFO

Time

- The amount of time required to process a VLDB can have a
  major impact on an installation

- Time constraints typically fall into two categories
  - Application processing
  - Maintenance operations

- Horizontal segmentation of the database frequently relieves
  time pressure in both of these areas

Time
Application Processing

- Horizontal segmentation
  - Divides occurrences of database entities along a high-level business specification
  - Allows portions of the database to be independent of each other

- Potential segmentation criteria
  - Geographic/company/division
  - Time/date relationships
  - A manufactured key value

- Typically a portion of the database cannot be segmented across the selected criteria and is referred to as the 'common' segment

Time
Application Processing

- Horizontal segmentation like Mixed Page Group Support has the same type of restrictions
  - Chain sets and referential constraints cannot cross segment boundaries
  - Record names within non-SQL defined database must be unique

- If Mixed Page Group Support is used with segmentation
  - Unqualified DML commands should be avoided

Time
Application Processing

Scenario 2

- A site does business in Europe and North America and has a single database for each of these locales

- Batch processing cannot occur until all online processing is done for the day

- Waiting for the North America online window to close causes the batch work to run into the opening of the next day's European window

Customer Database
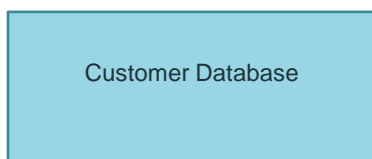
Time
Application Processing

- First, separate those areas that contain customer specific data that can be divided by location from data that is common to both European and North American customers

Common Areas

Customer Specific Areas

Time
Application Processing

- Define each group of areas to their own segment
- Create a third segment for the customer specific areas
- Assign each segment to their own page group if needed

Common Areas	Customer Specific Areas	Customer Specific Areas
CUSTCOM	CUSTEUR	CUSTNA
Page Group 0	Page Group 1	Page Group 2

Time
Application Processing

CUSTCOM

CUSTEUR

CUSTNA
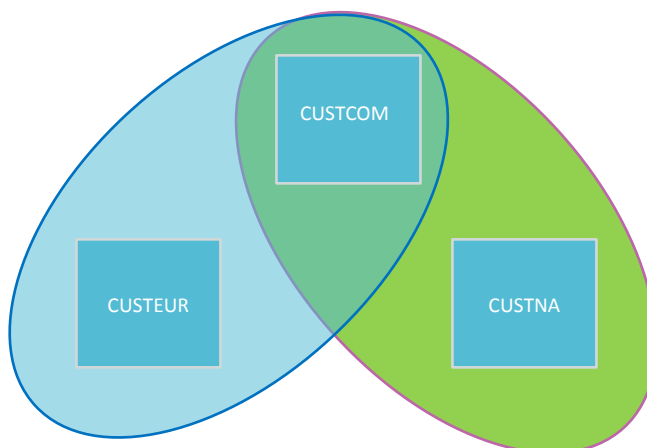
Care should be taken to avoid updating the common segment and another segment in the same run-unit due to recovery considerations

Time
Maintenance Operations

Segmentation

- Can reduce the volume of data to be processed and therefore the time to perform the following processes
  - Unload/reload of the database
  - Index tuning

- May minimize the portion of the database affected and the related outage times for the following operations
  - Restructures
  - Recovery

Time
Maintenance Operations

- The following utilities will reduce the processing time or the length of an outage when processing a VLDB
  - REORG
    - Reduces the time to perform an unload/reload operation by processing the database in multiple concurrent slices
  - EZ-REORG (third-party)
    - Reduces the outage time for a database during an unload/reload operation by allowing the operation to occur against a copy of the database while concurrent access is allowed against the original database
  - TUNE INDEX
    - Can eliminate outages by concurrently tuning indexes while normal processing is occurring against the database

Summary

- To implement a very large database using CA IDMS requires the ability to provide
  - Adequate space for the data
  - A method to perform application and maintenance processing within the necessary timeframe

- CA IDMS provides the architecture and the support code to fulfill both of these needs

- However, to successfully implement this type of environment requires planning on both the database and application design levels

FOR INFORMATION PURPOSES ONLY
Terms of this Presentation

Online Session Evaluation

Please provide your feedback about this session: D2

On the CA Communities web site:
http://communities.ca.com

More details in your conference bag

Questions and Answers