

The 2009 Way of Reorganizing DB2 LUW

(Confidential Material at the time is removed)

Jens Seifert, IBM Deutschland R&D GmbH

Stefan Krämer, Consultant

DB2.
software

Disclaimer

© IBM Corporation 2009. All Rights Reserved.

The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

References in this publication to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth, savings or other results.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. IBM, the IBM logo, Lotus, Lotus Notes, Notes, Domino, Quickr, Sametime, WebSphere, UC2, PartnerWorld and Lotusphere are trademarks of International Business Machines Corporation in the United States, other countries, or both.

Unyte is a trademark of WebDialogs, Inc., in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

If you reference Linux® in your presentation, please mark the first use and include the following; otherwise delete:

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Agenda

- REORG
 - Known issues
 - Outlook
- Online table move
 - What is it ?
 - When and when not to use it
- User experience (Stefan Krämer)

REORG

DB2.
software



Reasons for reorganization

- Optimize disk efficiency
 - Space reclamation
 - Make space available for other tables
 - Release space to operating system (reduction of HWM)
 - Enable compression / Re-compress
- Optimize performance
 - Remove sparsely filled pages
 - Compact indexes (removed pseudo deletes, remove unused leaf pages)
 - Optimize clusterratio
- Fix effects of structural changes
 - Enablement of large RIDs
 - Fix REORG pending after column alteration

DB2 LUW REORG options

- OFFLINE REORG
- INPLACE REORG
- ONLINE TABLE MOVE stored procedure

Known issues of REORG

- Index REORG (cleanup only)
 - Does not release space back to tablespace
- Index REORG (rebuild)
 - All indexes are reorganized in one transaction → high active log space consumption → very critical on HADR systems
 - Lock to switch objects at end of index REORG could cause lock timeouts
- Inplace REORG
 - No support for MDC tables
 - Index reorg recommended after inplace REORG
 - Truncation phase could cause lock timeouts
 - LONG/LOB/XML columns are excluded from REORG
- Offline REORG
 - Offline (read-only access) not sufficient for 24x7 systems
- General REORG limitations
 - May or may not reduce HWM
 - Cannot be used to change table and tablespace characteristics

Planned REORG improvements in DB2 9.?

- Confidential Material - removed

Online table move

DB2.
software

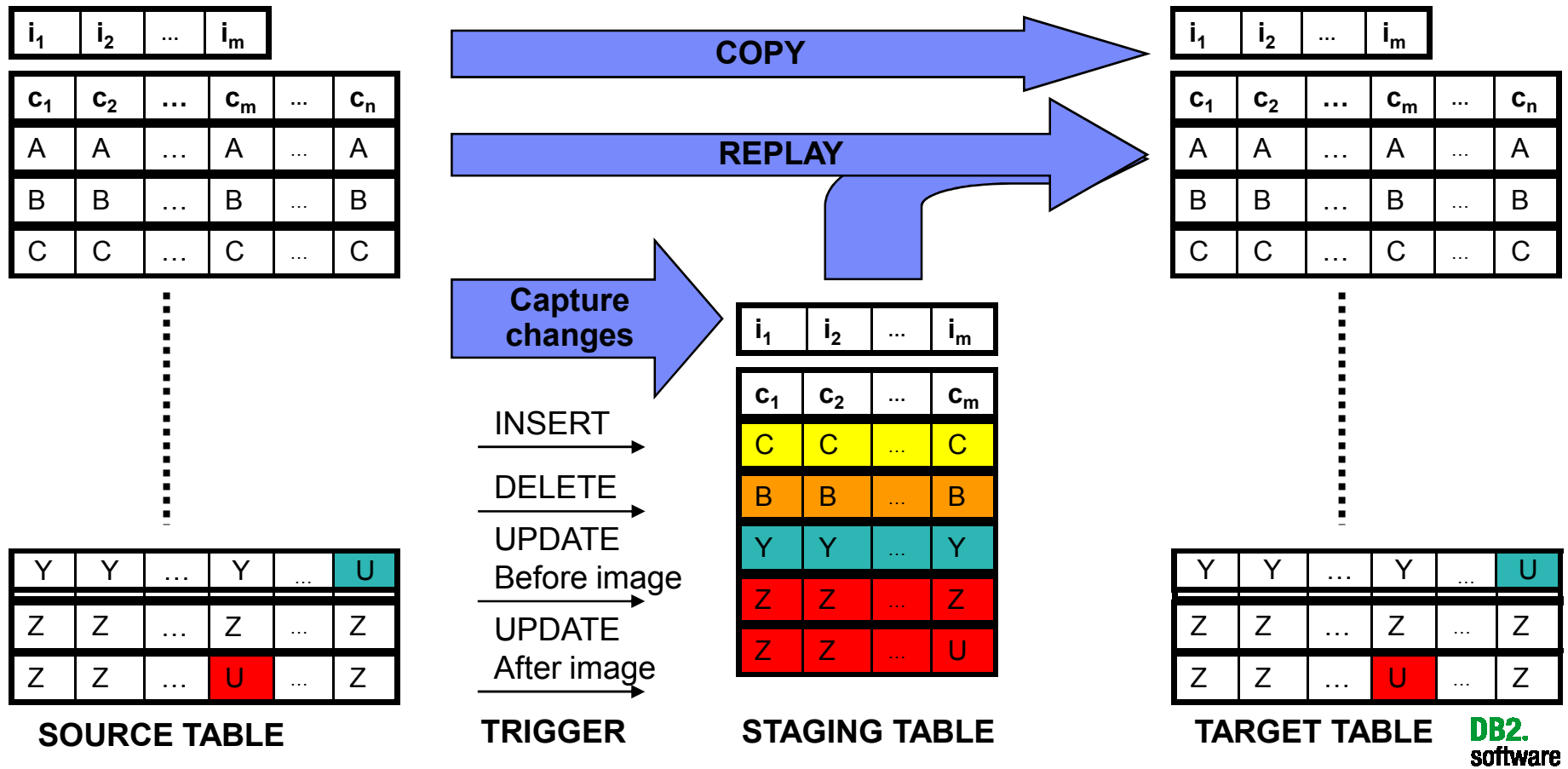
What is it ?

- Stored procedure to move a table online
- Online means
 - Fully logged operations:
 - ensure recoverability
 - ensure usability in HADR or log file shipping environment
 - Small window where a X-lock is required (seconds)
 - Small active log space usage
 - Minimal impact on UDI operations
 - No deadlocks***
 - No permanent locks on system catalogs
 - No problems with other utilities like REORG, online backup
 - Low resource consumption (CPU, Memory, DB resources like locks)

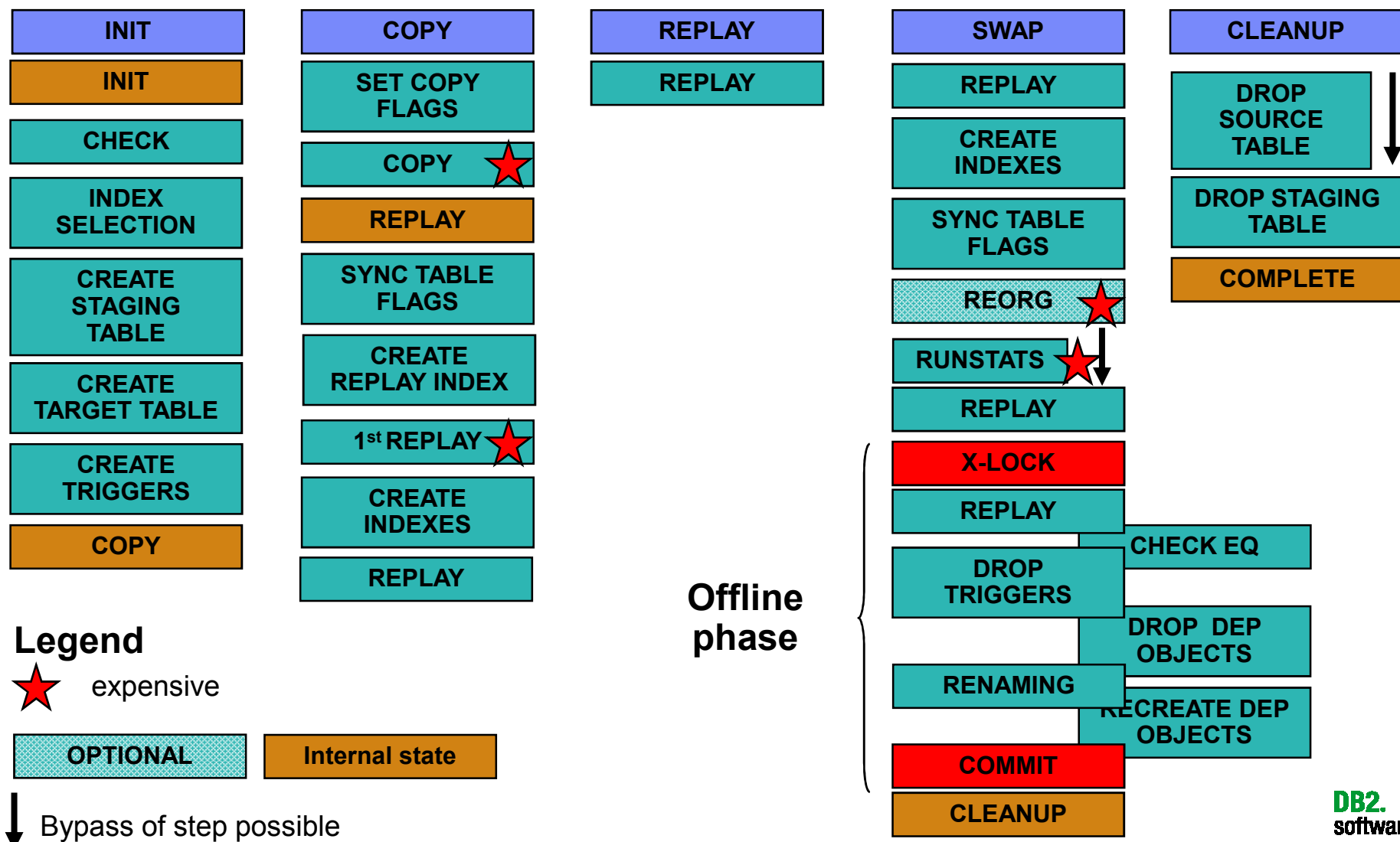
General concept

Protocol table SAPTOOLS.ONLINE_TABLE_MOVE

tabsche ma	tabna me	key	value	Long_value
SAPR3	UTAB	INDEX_SCHE MA	SAPR 3	-
...				



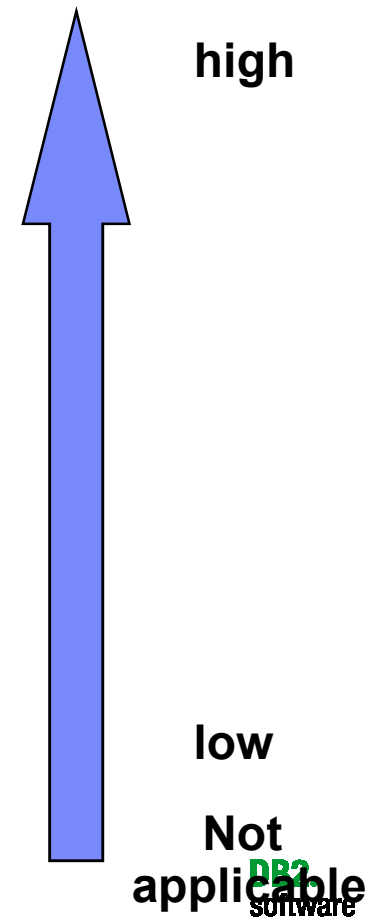
Basic Flow



Index selection

1. Primary index
2. Unique index with lowest COLCOUNT
3. Any index (non-unique or MDC block index)
 - a) With highest FULLKEYCARD
 - b) With highest COLCOUNT
4. Generated index. Select columns by following algorithm
 - a) With highest COLCARD
 - b) With lowest COLNO
 - c) At maximum 64 columns and less than 1024 bytes
 - d) Ignore LONG, LOB and XML columns
5. Fail (e.g. if table has only LONG/LOB columns)

Applicability of Online table move



Deployment

- For DB2 8.2.2 and DB2 9
 - Get binary from SAP (SAP OSS Note 1039544)
 - Copy the shared library to sqllib/function
 - Use db2 invoke online_table_move_sp to get stored procedure catalogued
 - Only available for selected platforms
- For DB2 9.5
 - Online table move is part of DB2 9.5 delivery as part of db2sap library
 - Use db2 invoke db2sap to get stored procedure catalogued
 - Available for all DB2 LUW platforms
- DB2 9.?
 - Confidential Material - removed

Usage

- DB6CONV version ≥ 4.00 (SAP OSS Note 362325)
use new “online” option
- On the command line (single step)
`call saptools.online_table_move('SAPR3','SVERS','TGT_D','TGT_I','','','MOVE')`
- On the command line (multi step)
`call saptools.online_table_move('SAPR3','SVERS','TGT_D','TGT_I','','','INIT')`
`call saptools.online_table_move('SAPR3','SVERS','','','','COPY')`
`call saptools.online_table_move('SAPR3','SVERS','','','','REPLAY')`
`call saptools.online_table_move('SAPR3','SVERS','','','','SWAP')`

Usage scenario sample: enable row compression

- Scenarios
 - Create dictionary online
 - Compress whole table without need for offline REORG
 - (Re-)create optimal dictionary
 - Dictionary degrades over time
 - Dictionary not optimal, if it was created with bad sample of data (e.g. during R3load)
- How does online table move achieve this ?
 - Load sample of about 20MB to target table using Bernoulli sampling
 - 10MB is the maximum sample size the dictionary algorithm is using to build dictionary
 - Why 20MB ? Size estimation and calculation of Bernoulli sampling rate is based on logical data size. If pages are sparsely filled, the sample could be less than 10MB. Assumption: pages are at least 50% filled
 - Create dictionary using `INSPECT TABLE ROWCOMPESTIMATE`
 - Clean target table without dropping the dictionary using `IMPORT from /dev/null (TRUNCATE)`
 - Restart COPY → all rows are compressed with optimal dictionary

Special Options

- KEEP: keep original table
- FORCE: skip some tests
- COPY_USE_LOAD: use non-recoverable LOAD (fast, but requires intermediate backup)
- COPY_WITH_INDEXES: create indexes before copying table
- STATS_NO: no RUNSTATS on target
- STATS_COPY: (experimental) copy statistics
- REORG: extra REORG with RESETDICTIONARY on target (no longer required)

When online table move helps ? (1)

- Online enablement of features
 - Large RID
 - Row Compression
- Online structural changes of table
 - Add/remove columns
 - Change column type for example CLOB to VARCHAR
 - Splitting of LONG/LOB data into separate tablespaces
 - Enable/disable MDC
 - Enable/disable Range-partitioning
- Change of tablespace physical layout required
 - Change of page size, extent size
 - Change to tablespace with large object table (DB2_OBJECT_TABLE_ENTRIES=65535)

When online table move helps ? (2)

- Misc
 - Reduce HWM
 - Solves active log space consumption problem
 - Online recompression of tables
 - Online REORG of LONG/LOB data
 - Online redistribute
- Swap phase which requires X-Lock can be performed in times with low system activity
- DB2 9.7 feature enablement
 - Confidential Material - removed

Online table move – The dark side (1)

- Types of tables that cannot be moved
 - Tables with only LONG, LOB or XML columns
 - Tables of active event monitor
 - The online table move protocol table
 - System tables
 - Tables with
 - Foreign-key relationship
 - RCT, MQT, typed tables
 - Generated columns
- Other
 - XML indexes are not copied
 - Not compatible with DB2_SKIP_DELETED=ON
 - Not compatible with LOAD, IMPORT...REPLACE, ALTER TABLE ... NOT LOGGED

Online table move – The dark side (2)

- Resource consumption
 - High space consumption (more than twice the size of table+indexes)
 - Additional pressure on buffer pool
 - High logging overhead (required to be fully recoverable)
- Usability dependant on index “quality”
 - Primary/unique index exists → no problems
 - Non-unique index
 - no deadlocks only if only 1 writer present
 - REPLAY efficiency dependant on index selectivity
 - No index
 - Generated index may not be optimal
 - Additional space required for generated index
 - No parallelism (REORG runs in parallel on DPF)

Online table move on SAP systems ?

- Most SAP tables have a primary or unique index
- SAP only uses some table features
 - foreign key relationships only used in very rare cases on SAP systems
 - XML indexes, sequences, security labels, row change timestamp,... not used by SAP
- Controlled environment
 - LOAD: SAP does not use LOAD except for R3load
 - IMPORT REPLACE: Not used except for TRUNCATE TABLE, but this is now has been made aware of existence of DELETE trigger
 - ALTER TABLE ... NOT LOGGED INITIALLY:
 - Was used in SAP BI, but all operations has been replaced by DBSL truncate
 - DB2_SKIP_DELETED=OFF is SAP default
- Ease of use by DB6CONV tooling

Best practices (1)

- Only use online table move if DB2 methods are not applicable
 - Don't use it as general replacement for INPLACE REORG
 - Check if offline methods are the better choice
- Be careful using the stored procedure on tables without unique index
 - Each entry in a non-unique index may address some or thousands of rows. In case the index has a bad selectivity (1 index entry addresses thousands of rows) a lot of resource for replaying table changes is required
 - Deadlocks are possible
 - Long lock-wait times possible, because large replay operations can take some time
- Avoid moves on tables that have no index

Best practices (2)

- Performance:
 - Rule of thumb: 50GB per day
- Don't issue multiple moves into same tablespace at a time to avoid fragmentation on target tablespace.
- Run move during time of low activity on the table
 - avoid mass data loads/deletes
 - Parallel read access is no problem
- Separate online table move steps
 - INIT+COPY at any time
 - REPLAY several times to keep staging table small
 - Issue SWAP during a time of no/low activity on the table

Summary

- Online table move is expensive and slow.
- Online table move has its own limitations.
- Online table move can help where REORG sucks.

User Experience

Real Life Experience - Stefan Krämer

DB2.
software

One Page - Successful reorg

- Traditional table reorgs, reclaim space
 - Smaller tables with lots of freespace
 - Tables which are not permanently needed
 - Weekend – time for large BW tables
- Most important are indexes
 - Before table reorgs !
 - Always with "ALLOW WRITE ACCESS", READ is the default
 - Use "CLEANUP ONLY ALL" first
 - Use full index reorg only if required – will need exclusive locks
- Inplace table reorg
 - Only with nottruncate table
 - Recluster by specific indexes if you understand the application requirements
 - Specify index name for tables with > 50% freespace to reduce the runtime
- Other
 - Outside main business hours if possible
 - Try to prevent locking conflicts (perform lock checks before reorg)
 - Runstats after reorg
 - HADR may add limitations (LOGINDEXBUILD=YES)
 - Limit reorgs in relationship to the available temp and log space

Online table move - Overview

- Online table move ?
 - "ONLINE" is true for most tables
 - Downtime minimized for exceptions
 - Use the phased approach if you want to gain control
 - Create indexes before copy – safe implementation
- Planning
 - Runtime may be long - plan to start early before planned outages
 - Make sure there are no major changes to the table content, mass deletions, etc.
 - Outside SAP TMS Import All Windows
 - SWAP as early as possible
 - Find HOT tables

Online table move – Phases

- INIT Phase
 - CREATE TRIGGER requires exclusive locks
 - For HOT tables – choose the best time for INIT outside main business hours
- COPY Phase
 - Some tables have very long runtime, plan INIT early before planned outages
 - Runtime: up to 3 weeks with INSERT for extremely large and slow tables
- REPLAY Phase
 - Replay daily to reduce time to SWAP
 - Runtime: up to 2 weeks with INSERT for extremely large and slow tables
- SWAP Phase
 - Choose best possible time window, it may impact current applications on the table

Online table move – Versions

- SAP Version 2.7 is the current version for DB2 V9
 - Reclustering by secondary index
 - Unrecoverable LOAD
 - Improvements for tables without unique primary key

Online table move – Use Cases

- DB2 V8 and later - Tablespace Capacity or HWM issues
- DB2 V9 Large RID conversions
- DB2 V9 Deep Compression
- DB2 V9.x (future release) – confidential material removed

DB2 Deep Compression – Customer Expectation

- For large databases is the total possible space saving estimated:
 - Total saving of used space = ca. 40-50%
 - Total saving of allocated space = ca. 50-60%
- The total space saving, is estimated to be in average:
 - 55% of all savings by compressing table content
 - 5% of all savings by reclaiming space within tables
 - 20% of all savings by reducing freespace in tablespaces under the HWM
 - 20% of all savings by changes in the tablespace freespace pre-allocation

DB2 Deep Compression – Overview

- INSPECT always before
 - Store the result for later comparison
- Candidates – define your criteria's
 - Estimated compression rate > 50%
 - > 10000 rows
 - No DB2 volatile tables
 - No other volatile tables (may uncompress if the compression rate drops)
- Possible fallback plans
 - Alter table compress NO
 - Uncompress tables – reverse compression implementation

DB2 Deep Compression – Valid Criteria's ?

- One example R/3 system – 14,5 TB used space
 - 1800 of 35000 are well compressable and satisfy own conditions
 - 72 tables represent 80% of all compressable space
 - 206 tables represent 95% of all compressable space

- Is it necessary to compress more tables?
 - One key driver is the space saving
 - Compressing more tables does not provide major space savings, but may also help to reduce additional growth
 - Work per tablespace to get most space saving !

DB2 Deep Compression & Reclaim Space from Tablespaces

- How to kill two birds with one stone ?
 - Rebuild tablespaces completely, use the following steps
 - Rename old tablespaces to e.g. "_old"
 - Create new tablespaces with the original tablespace names
 - SAP data class -> correct new tablespace, no issues with SAP Transports
 - Move all tables from the "old" tablespace into the new tablespace
 - Drop "old" tablespaces
- Other aspects
 - Move large tables into own tablespaces to allow later easier space reclaim
 - Tablespace autoresize - keep the increase size large to prevent fragmentation for file containers, e.g. 16-128 MB * containers

Space Saving Cockpit - Example

Tables compressable from v9_tb_stat:	1449	
Tables compressed from syscat.tables:	1938	= 133.74 %
Tables not compressable from v9_tb_stat:	122	
Tables in syscat.tables:	33673	
Tables compressable with too small rows:	0	
Space MB currently in all tables:	535084	= 41.04 % of all used space
Space MB currently in all indexes:	572573	= 43.92 % of all used space
Space MB unused in table np < fp:	41631	= 3.19 % of all used space
Space MB compressable currently unc. tables:	1035	= .07 % of all used space
Space MB to be saved by compression:	622	= .04 % of all used space
Space MB currently in compressed tables:	523099	= 40.12 % of all used space
Space MB is saved in compressed tables:	2039004	= 99.95 % of compression work completed
Space MB other content:	154362	= 11.84 % of all used space
NB: other/missing space may be in extents of empty tables, LONG and LOB objects, addition space for indexes		
Current compressed tables compression factor:	79.57 %	- target is: 79.65 %
Current total table compression factor:	79.20 %	- target is: 79.22 %
Current total used space compression factor:	60.97 %	- target is: 60.99 %
TBS total space MB usable (allocated):	1278628	
TBS total space MB used:	1253843	= 98.06 % used

References

- IBM DB2 Best Practices:
<http://www.ibm.com/developerworks/db2/bestpractices>
- IBM DB2 V9 Information Center:
<http://publib.boulder.ibm.com/infocenter/db2luw/v9>
- SAP Note 1039544 DB6: Stored procedure for online table move
<https://service.sap.com/sap/support/notes/1039544>
- DSAG 2007 – Detailed Reorg Presentation
<http://www.skc-group.com>

Thank YOU !

Stefan Krämer

D-88682 Salem

<http://www.skc-group.com>

Other points mentioned ...

- Use AIX fileplace to see the fragmentation at the filesystem level
- Resize target tablespaces to full size, to minimize fragmentation
- Minimal performance impact of online table move