

Attack of the DB2 9 for z/OS Clones Clone Tables That Is!

John Lyle
IBM Silicon Valley Lab. (Etats Unis)
jlyle@us.ibm.com



Réunion du Guide DB2A
Jeudi 29 mai 2008
Feel Europe, Vincennes (94)
France

John Lyle has worked at the Silicon Valley Laboratory in the DB2 organization since 1985.

Joining DB2 development in 1995, he has worked in the RDS area focusing on the DB2 system catalog, migration and fallback and the data definitional language (DDL) areas.

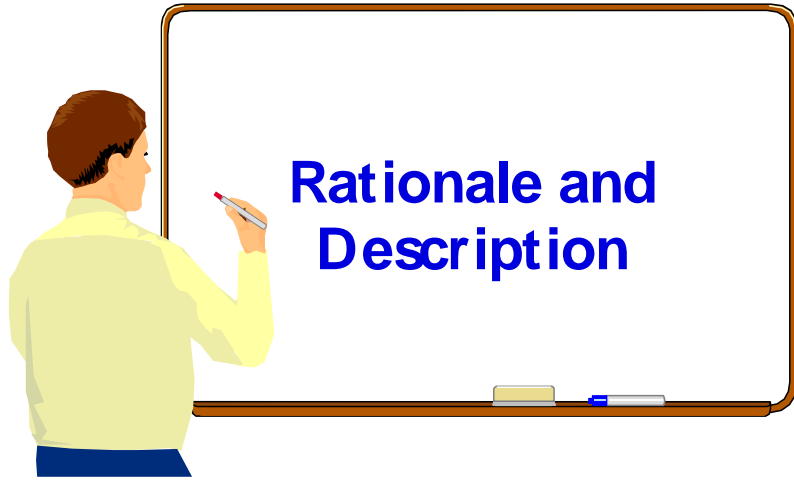
He was the lead designer and implementer of the catalog changes for V5, V6, V7 and V8, the fallback coordinator for V5, V6, and V7 and has worked on numerous other Line Items in these releases. He spent some time in DB2's index manager component and is currently working in the DB2 utilities department.

- Rationale and description
- DDL statements
 - ALTER TABLE . . . ADD CLONE
 - EXCHANGE DATA BETWEEN . . .
- Authorization
- Catalog and directory changes
- Commands and utilities



© 2008 IBM Corporation

In V9 DB2 for z/OS introduces Clone Tables. In this presentation I will introduce the concept of Clone Tables and detail how they work. Clone Tables have several dependencies on other new V9 functionality and I will briefly go over what these dependencies are. I will then indicate how you create Clone Tables tables, and how the data is magically 'exchanged' between the parent/base table and it's clone. I'll then finish up with a discussion of how Clone Tables can be used in a customer environment.



© 2008 IBM Corporation

- “. . . many of our processes involve completely emptying the table and then reloading it in its entirety. In order to accomplish this the table space must be offline which does impact the availability of the business application . . . “
- “. . . It is increasingly difficult to do LOAD REPLACES of data which is required for business as their business becomes more and more 24x7x365 with online transactions 24 hrs a day.”

Cloned table support provides the ability to generate a table with the exact same attributes as a table that already exists at the current server. This new table that's being created is referred to as the **clone** table of a **base** table

Not only is the clone table that's created identical to the base table in every way:

- same number of columns
- same column names
- same data types
- same table check constraints, etc.

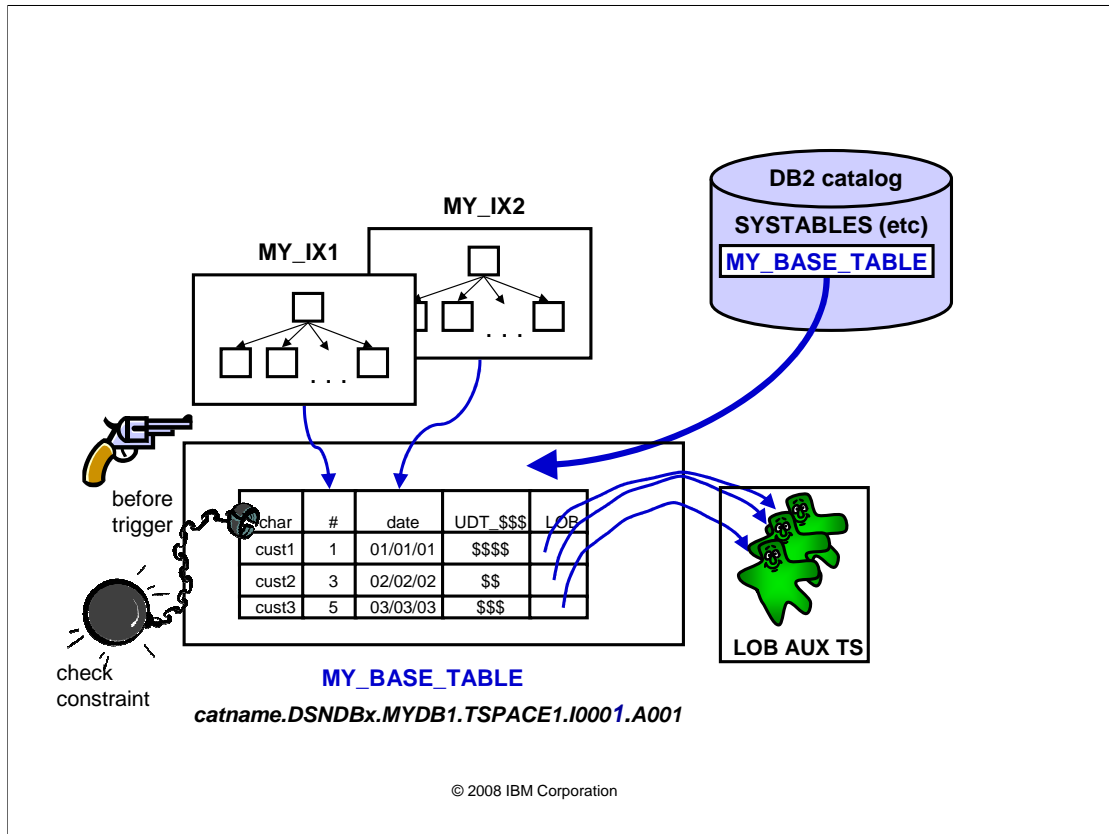
It's also created with the same

- indexes
- before triggers, etc.

Not cloned? Views and authorization

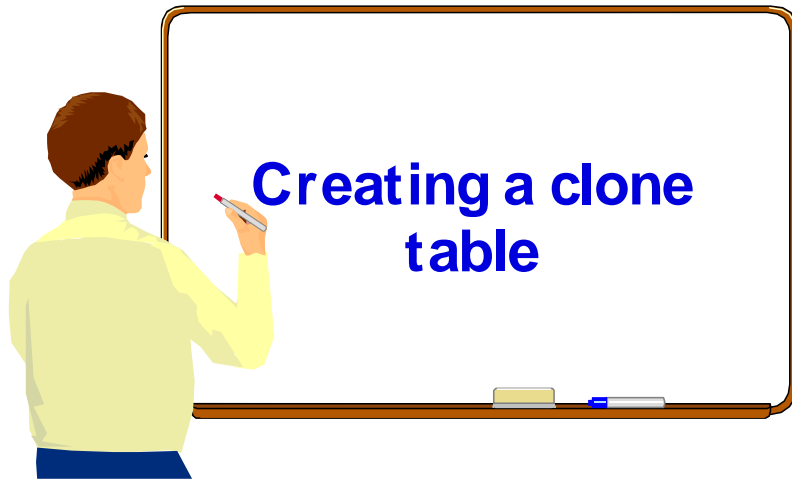
© 2008 IBM Corporation

Views and authorization are not created when the clone table is created. Base and clone tables can have different views and different authorization sets.

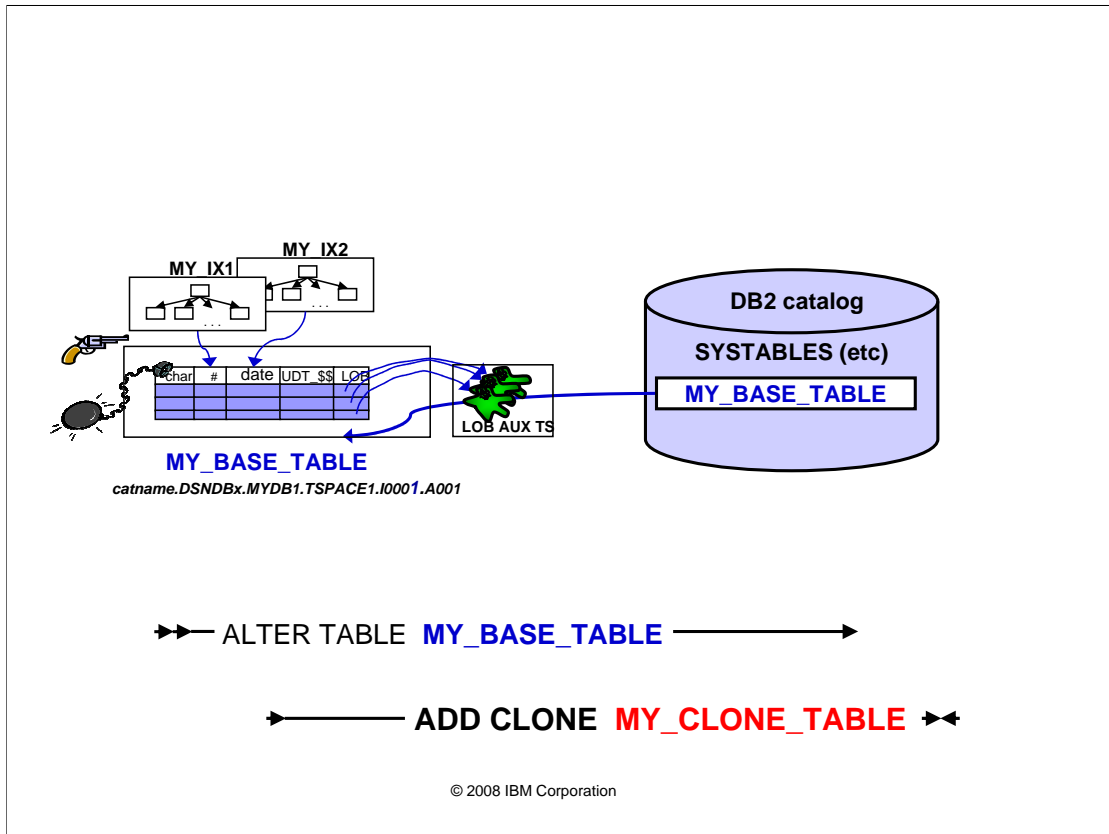


Here is an existing base table that has a couple indexes, a before trigger and a table check constraint. There is also a LOB column in the table so you can see there is an auxiliary table as well. This table already has some data in it. The picture would look the same for an XML column where there would also be an XML table space, but for simplicity that is not depicted here.

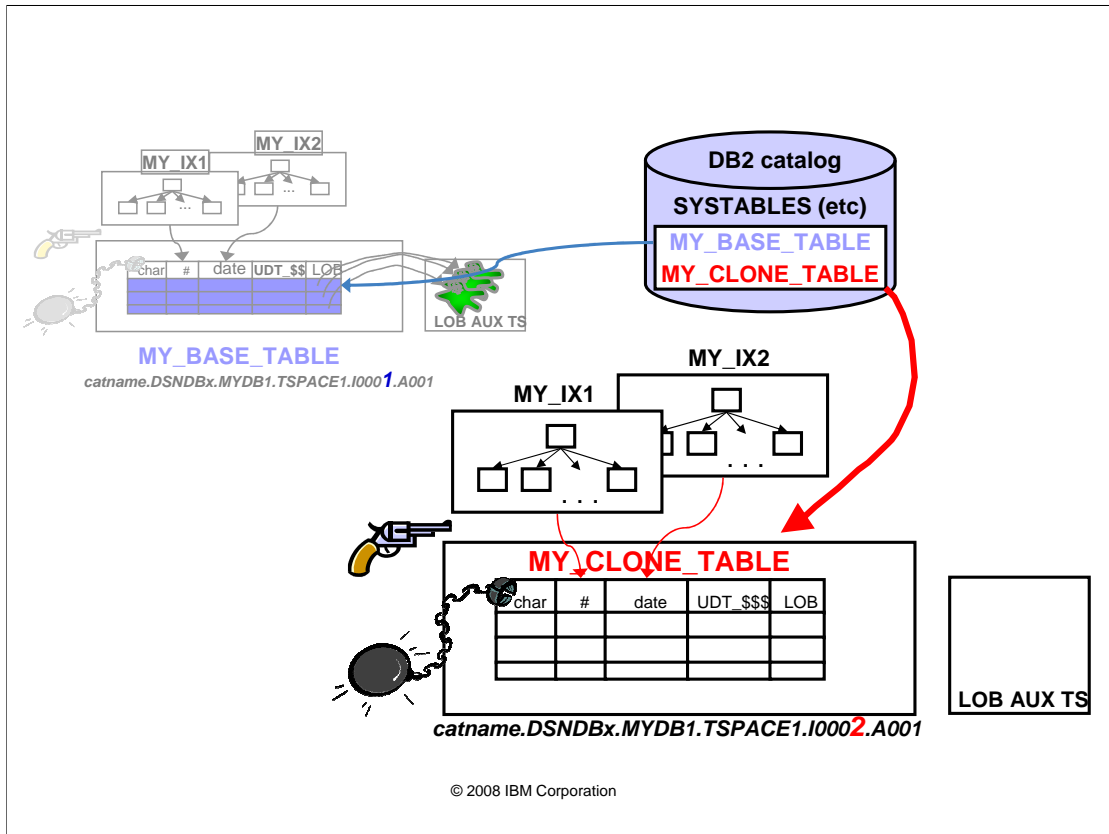
This table is referred to as the 'base' table. This should not be confused with the 'base' terminology which is also used with the description of LOB and XML objects.



© 2008 IBM Corporation



The base table MUST be in a Universal Table Space before a clone table can be created. Here we're going to create a clone table of base table MY_BASE_TABLE that already exists in a UTS. Although we'll be creating a clone table we have to use the ALTER TABLE syntax to create the clone table.



DB2 generates all objects:

- table space (a universal table space (UTS))
- table, with all column names and data types the same
- before triggers
- LOB objects (aux table space, indexes)
- XML objects (XML table space, indexes)
- Table check constraints
- indexes

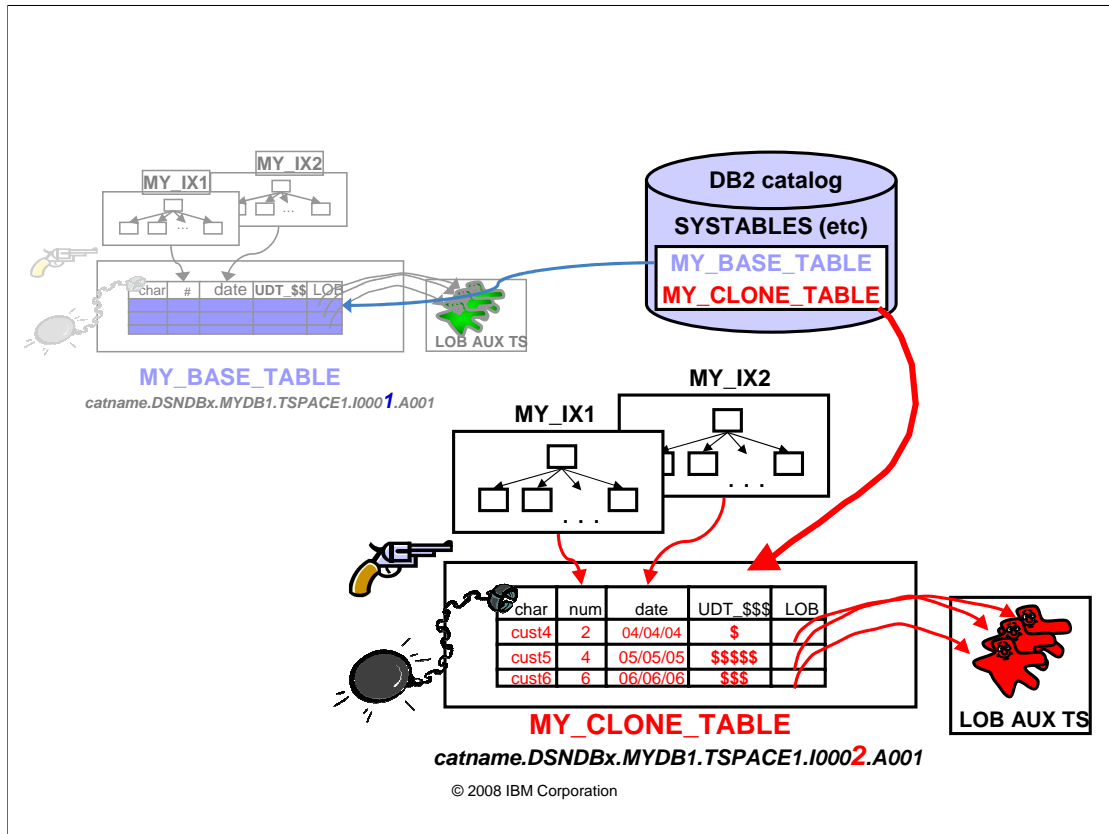
No data is cloned so after the create (via ALTER TABLE) completes successfully the new table is empty.

Note that the clone table is represented by instance number 2 data sets.

- Catalog tables cannot be cloned
- Could be created with a different schema (owner) than the base table
- Once created, access must be granted to users -- the clone functions like a “normal” DB2 table
- Like a “normal” DB2 table, after creation, the clone objects (table, indexes, aux objects, etc.) are empty. They can be populated by:
 - INSERT
 - LOAD
 - etc.

© 2008 IBM Corporation

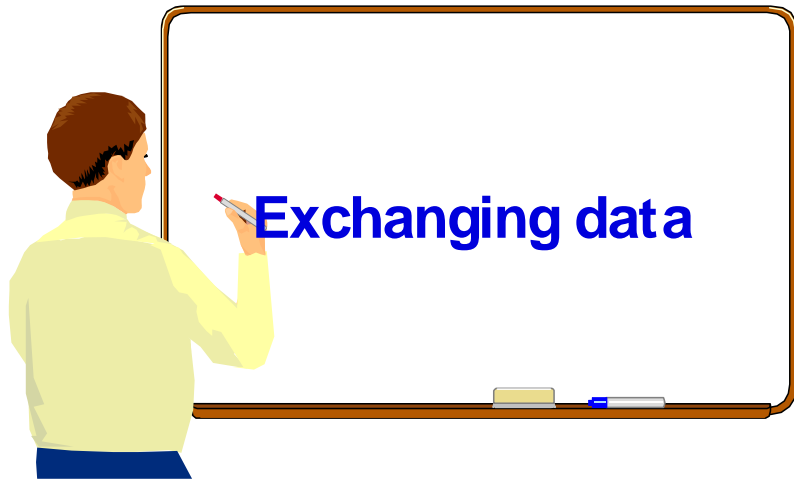
A clone table only exists if there was originally a base table. You cannot have a clone table without a base table. But after the clone table is created it is essentially it's own entity. You can grant users access to the clone, you can load it with completely separate data, etc. Conceptually you can think of it as being completely separate from the base table.



DB2 generates all objects:

- table space
- table, with all column names and data types the same
- before triggers
- LOB objects (aux table space, indexes)
- XML objects (XML table space, indexes)
- Table check constraints
- indexes

No data is cloned. But here we show the clone table after it's been LOADED (via LOAD, INSERT, etc) with data.



© 2008 IBM Corporation

↔ **EXCHANGE DATA BETWEEN** *table-name-1* →

← **AND** *table-name-2* ↔



© 2008 IBM Corporation

table-name-1 and table-name-2 -- base table and clone table names can be specified in any order.

- **EXCHANGE DATA BETWEEN . . .**

- Will switch the data set **instance numbers**. This **effectively** switches the data between the base and clone table objects. **No data is actually moved!**

- **Before EXCHANGE:**

```
catname.DSNDBx.MYDB1.TSPACE1.I0001.A001  BASE
catname.DSNDBx.MYDB1.TSPACE1.I0002.A001  CLONE
```

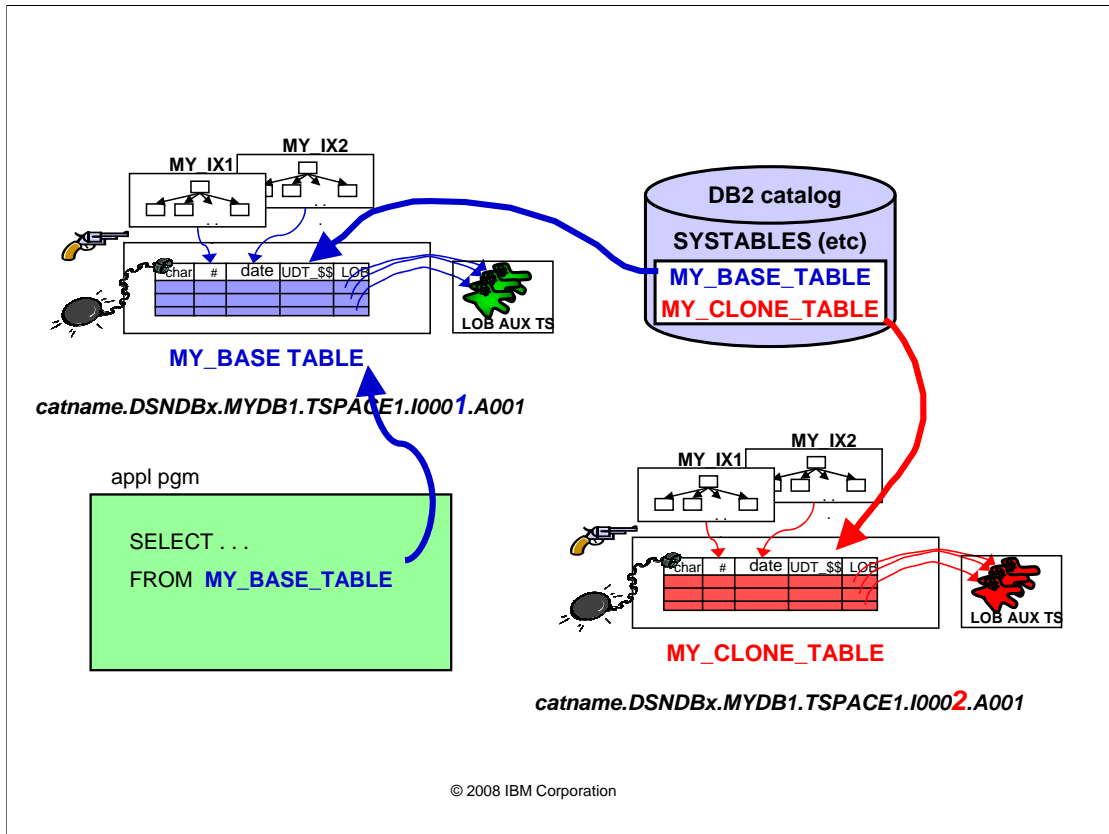
- **After EXCHANGE:**

```
catname.DSNDBx.MYDB1.TSPACE1.I0002.A001  BASE
catname.DSNDBx.MYDB1.TSPACE1.I0001.A001  CLONE
```

- Very quick operation.
- Only applicable if cloning in effect.

© 2008 IBM Corporation

No data is moved during the EXCHANGE process. We only 'switch' the data set instance numbers. This might make it look like we've moved data back and forth but we don't ever do that. The instance number changes happen for all objects that have underlying VSAM data sets (ie table and index data).



In this slide we're doing a SELECT from the base table which is currently represented by instance number 1 data sets. DB2 knows where to go to get the data (ie instance number 1 data sets) and which indexes to use.

- Table names in EXCHANGE statement can be specified in any order:

– EXCHANGE DATA BETWEEN MY_BASE_TABLE AND MY_CLONE_TABLE

or

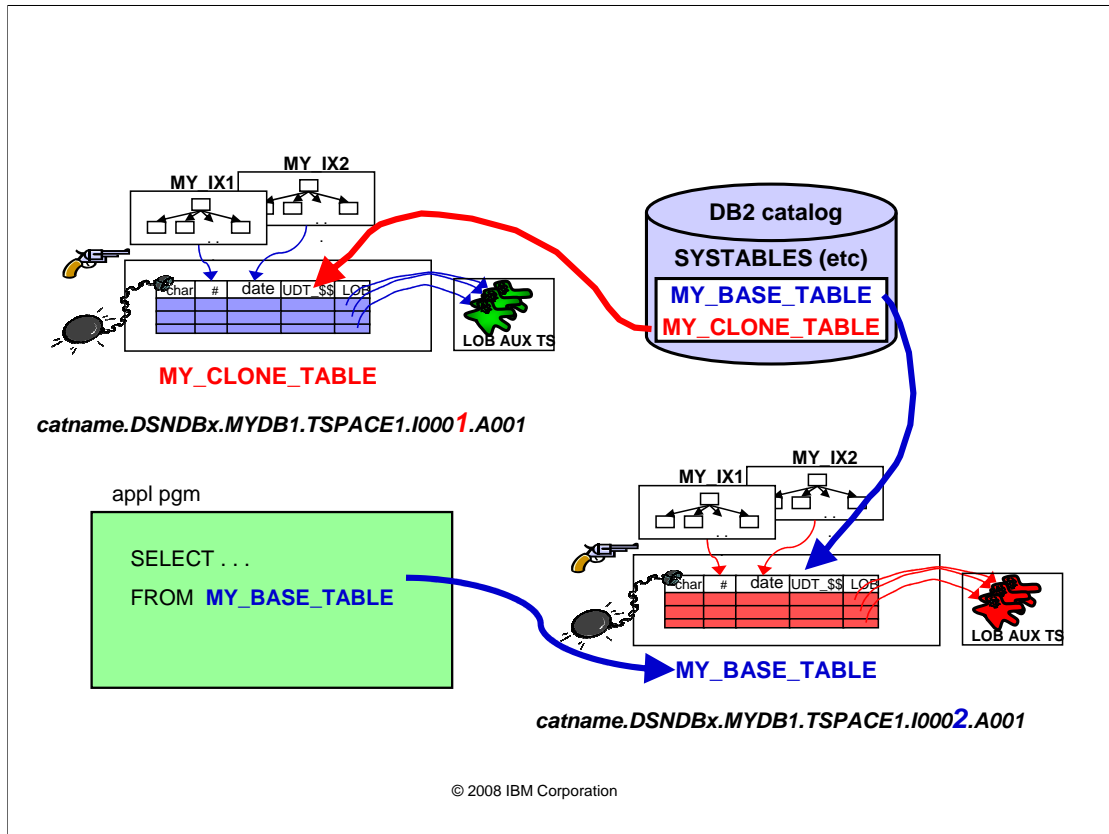
– EXCHANGE DATA BETWEEN MY_CLONE_TABLE AND MY_BASE_TABLE

© 2008 IBM Corporation

Applications are bound to the base table.

An authorized person submits the EXCHANGE statement.

Plans and packages are not invalidated.



After the EXCHANGE, the base table data is now found in the instance number 2 data sets and the clone table data in the instance number 1 data sets. When you SELECT data from the base table, DB2 knows that it now has to look in the instance number 2 data sets for its data. If you were to SELECT data from the clone table, DB2 would know that it now has to look in the instance number 1 data sets for its data. This is true for all data sets including those for the base table space, indexes, LOB table spaces, and XML table spaces.

- Can only create a clone if the table space is a Universal Table Space (**UTS**)
Can only create a clone table in a DB2 managed table space
Can only create a clone for a table if it's the only table in a table space
 - A UTS is partitioned by definition
No FASTSWITCH if there's a clone
 - Can create a clone if base table has a mix of I/J data sets.
However, cannot fix/change this as long as the clone exists.

© 2008 IBM Corporation

Clone tables can only be created if the base tables are in a universal table space. This means that clone tables (and their associated base table) will be partitioned.

FASTSWITCH is not allowed because the I/J data set information is kept in the DB2 catalog (for tables it's in SYSTABLEPART.IPREFIX and indexes it's kept in SYSINDEXPART.IPREFIX) and this information is 'shared' between the base and clone objects. This means that there will be a single row in SYSTABLEPART for a given partition and this single row is used for BOTH the base and clone table. If we were to allow FASTSWITCH to be used then the catalog would be updated for one object and then we'd be out of synch with the other object. To prevent this from happening we do not allow FASTSWITCH to be used.

- No ALTERs (online schema changes for base or clone table. For example, no ALTERing column data types, adding columns, etc.)
 - Because most of catalog is 'shared' by the base and clone objects
- Cannot clone a table that has RI or add RI to a table if it is involved in cloning
- After triggers NOT allowed on either table. Only before triggers allowed.
- No clones allowed on MQTs

© 2008 IBM Corporation

Most of these restrictions are there because catalog information is shared between the base and clone objects. Some of them (ie the restrictions with referential integrity, after triggers, etc) are there because of time and/or resource restrictions. We may see some of these restrictions eliminated in the future.

- **Can** create an index on a base table

- Creates index on both base and clone tables
- Clone table indexes are immediately put into RBDP/PSRBD if clone table contains data

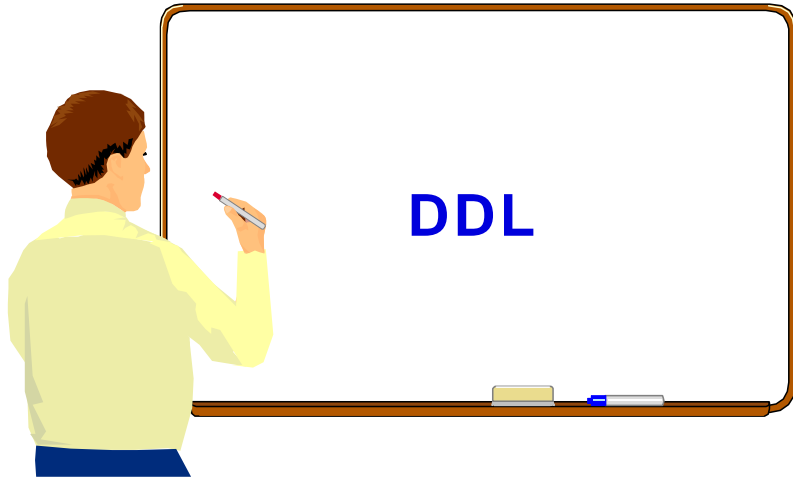
- **Can** create a before trigger on the base table

- Creates before trigger on both base and clone tables

- No RUNSTATS against clone table
- On EXCHANGE, the clone becomes the base and it assumes the base objects' statistics
 - Design premise: base and clone data will be similar enough that statistics can apply to both base and clone
 - This allows bound static SQL to function without a rebind

© 2008 IBM Corporation

Most of the catalog information is shared for base and clone objects. Because of this we will not allow/keep statistics for the clone objects. We believe that base and clone data will be similar enough that the same statistics could be used when accessing data from either set of objects. If the data is not that close then you may need to run RUNSTATS after an EXCHANGE has been done so that the statistics more closely match the actuals for the current base objects.



© 2008 IBM Corporation

- **ALTER TABLE *base-table-name* DROP CLONE**

- Drops only the clone table objects and their underlying data sets.

- **DROP TABLE** will drop base table and its clone.

- Cannot use DROP TABLE to drop only the clone table.

© 2008 IBM Corporation

If you only want to DROP the clone objects then you can do so by using the ALTER TABLE DROP CLONE syntax. This will delete only the clone objects. If you want to get rid of both base and clone objects then use the DROP TABLE syntax.



© 2008 IBM Corporation

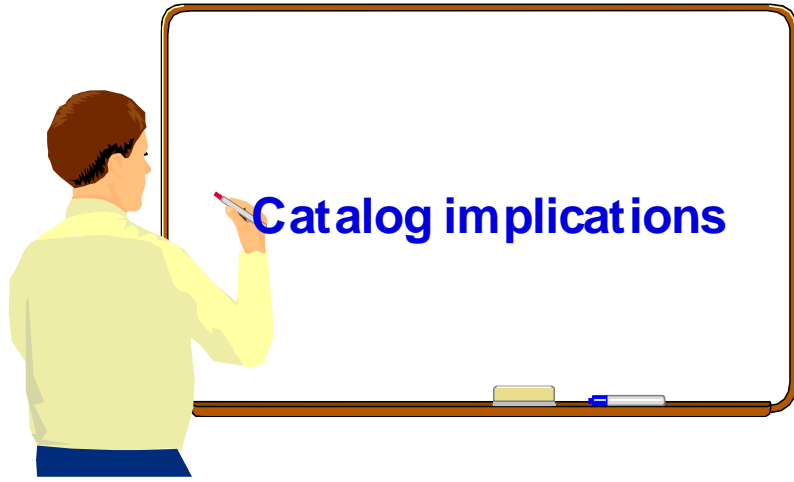
- DML authorization on base and clone can be different
- **EXCHANGE DATA BETWEEN . . .** operation requires one of these:
 - Ownership of both tables
 - INSERT and DELETE privileges for both tables
 - DBADM for the database
 - SYSADM auth
- **ALTER TABLE table-name ADD CLONE clone-table-name**
 - Same privilege set as for CREATE TABLE for table being cloned:
 - CREATETAB for the database
 - DBADM, DBCTRL, DBMAINT
 - SYSADM, or SYSCTRL

© 2008 IBM Corporation

For the EXCHANGE DATA statement, the privileges held by the authorization ID of the statement must include at least one of the following:

The INSERT privilege and DELETE privilege for both tables

- Ownership of both tables
- DBADM authority on the database that contains the tables
- SYSADM authority



© 2008 IBM Corporation

- SYSTABLESPACE
 - **INSTANCE** SMALLINT NOT NULL WITH DEFAULT 1
 - **CLONE** CHAR(1) NOT NULL WITH DEFAULT 'N'
- SYSCOPY
 - **INSTANCE** SMALLINT NOT NULL WITH DEFAULT 1
 - The timestamp of a data exchange will be tracked in SYSCOPY
 - ICTYPE = 'A' and STYPE = 'E' records
 - SYSCOPY EXCHANGE row for the base objects and also the clone objects

© 2008 IBM Corporation

The SYSTABLESPACE.INSTANCE column value will always be a '1' or a '2'. It will always reflect the instance number of the current base objects. The CLONE column indicates if there is currently cloning. If the clone objects are dropped when the base objects are represented by instance number 2 data sets then CLONE will be 'N' in this case and SYSTABLESPACE.INSTANCE will be a 2. This is OK. If/when a clone is created again in the future then the clone objects will be created using instance number 1 data sets.

The SYSCOPY catalog table keeps track of the EXCHANGES that are done. There will be one SYSCOPY row for each base and clone object (ie each index will have a row, the base and clone table space will have a separate row, etc).

SYSTABLES

- There is a separate SYSTABLES row for the base and clone tables.
- TYPE = 'C' for clone table
- TBCREATOR and TBNAME columns note related/partner table
 - For base table row, TBCREATOR and TBNAME are for clone table
 - For clone table row, TBCREATOR and TBNAME are for base table
 - OBID column value will be same for base and clone table
- SYSCOLUMNS
 - Contains rows for both base and clone tables
- SYSVIEWS
- SYSTABAUTH
- Rest of catalog
 - Most information is 'shared' (not duplicated)
 - SYSINDEXES, SYSINDEXPART, SYSKEYS, SYSTABLEPART

- Object identifiers

- Page set id

- High order bit (HOB) indicates instance number

- '0' indicates instance number 1

- '1' indicates instance number 2

Example: PSID = '0007'x for base object, = '8007'x for clone

- Catalog won't ever reflect a value where HOB = '1'B

- SYSTABLESPACE

- **INSTANCE** reflects the base table/object instance number (1 or 2) and will 'toggle' on an EXCHANGE (1 -> 2, 2 -> 1)
- **CLONE** remains 'Y'

- SYSCOPY

- **INSTANCE** There will be one SYSCOPY EXCHANGE row written (ICTYPE = 'A' and STYPE = 'E' record) for each instance (1 and 2). Does not indicate if it was base or clone object at the time of EXCHANGE.



© 2008 IBM Corporation

```

-DISPLAY DATABASE(MYDB*) SPACENAM(*) LIMIT(*)
DSNT360I ( *****
DSNT361I ( * DISPLAY DATABASE SUMMARY
. . .
NAME      TYPE      PART      STATUS
-----
CHKOUT2  TSB1      001      RW
CHKOUT2  TSB1      002      RW
. . .
XDEPT1   IXB1      001      RW
XDEPT1   IXB1      002      RW
. . .
CHKOUT2  TSC2      001      RO
CHKOUT2  TSC2      002      RW
. . .
XDEPT1   IXC2      001      ICOPY
XDEPT1   IXC2      002      RO
. . .
CHKOUT3  TS        001      RO
. . .
***** DISPLAY OF DATABASE MYDBX ENDED *****

```

© 2008 IBM Corporation

A 2-character type suffix is added to the end of these existing types as follows:

- The first suffix character will be either a 'B' to indicate a base object or a 'C' to indicate a clone object.
- The second suffix character will be the associated data set instance number. This will be either a '1' for data set instance number 1, or a '2' for data set instance number 2.

Here we have some base and clone objects. Note that the objects have different states for different partitions.

↔ EXCHANGE DATA BETWEEN *table-name-1* →

← AND *table-name-2* ↔



© 2008 IBM Corporation

The base or clone table name can be specified first.

```

-DISPLAY DATABASE(MYDB*) SPACENAM(*) LIMIT(*)
DSNT360I ( *****
DSNT361I ( * DISPLAY DATABASE SUMMARY
. . .
NAME      TYPE      PART      STATUS      . . .
-----
CHKOUT2  TSB2      001      RO          . . .
CHKOUT2  TSB2      002      RW          . . .
. . .
XDEPT1   IXB2      001      ICOPY       . . .
XDEPT1   IXB2      002      RO          . . .
. . .
CHKOUT2  TSC1      001      RW          . . .
CHKOUT2  TSC1      002      RW          . . .
. . .
XDEPT1   IXC1      001      RW          . . .
XDEPT1   IXC1      002      RW          . . .
. . .
CHKOUT3  TS        001      RO          . . .
. . .
***** DISPLAY OF DATABASE MYDBX ENDED *****
. . .

```

© 2008 IBM Corporation

After an EXCHANGE the instance numbers changed and the base objects are now represented by instance number 2 data sets. ***Note that the states followed the instance numbers!*** This means that if, for example, you have an ICOPY state on all the clone partitions (instance number 2 data sets) and RW on all the base object partitions.....and you do an EXCHANGE.... the base objects would now ‘inherit’ the ICOPY states and the clone objects would then ‘inherit’ the RW states.

```

-DISPLAY DATABASE(MYDB*) SPACENAM(*) LIMIT(*)
DSNT360I ( *****
DSNT361I ( * DISPLAY DATABASE SUMMARY
. . .
NAME      TYPE      PART      STATUS      . . .
-----
CHKOUT2   TSB2      001      RO
CHKOUT2   TSB2      002      RW
. . .
XDEPT1    IXB2      001      ICOPY
XDEPT1    IXB2      002      RO
. . .
CHKOUT3   TS        001      RO
. . .
***** DISPLAY OF DATABASE MYDBX ENDED *****
. . .

```

NOTE: If clone is dropped when the base objects are represented by instance number 2 (odd number of EXCHANGES done), DISPLAY will continue to display the 'B2' type suffix even though there is no clone. When new clone is created it will be given instance number 1 data sets.

If the clone table is dropped when the base objects are represented by instance number 2 data sets, a DISPLAY DATABASE will still have the 'B2' suffix. This indicates that the base objects are represented by instance number 2 data sets.

- **STOP/START DATABASE** gain a **CLONE** keyword
 - Stop/Start by default processes only base objects, unless **CLONE** specified then just **CLONEd** objects.

- **DISPLAY DATABASE**
 - No **CLONE** keyword. Displays both base and clone objects by default.

- **DSN1LOGP**

- DSN1LOGP can print the log records for both base and clone table objects. The base and clone objects are differentiated by differences in the OBID/PSID value.

- **DSN1COPY**

- DSN1COPY can operate on both base and clone table objects. The base and clone objects are differentiated by differences in the OBID/PSID value.

- **DSN1PRNT**

- DSN1PRNT operates on a named (SYSUT1) data set. Use data set name of base or clone object.

- ***catname.DSNDBx.MYDB1.TSPACE1.I0001A001***
- ***catname.DSNDBx.MYDB1.TSPACE1.I0002A001***

- CHECK DATA
- CHECK INDEX
- CHECK LOB
- COPY
- COPYTOCOPY
- DIAGNOSE
- LISTDEF
- MERGECOPY
- MODIFY RECOVERY
- QUIESCE
- REBUILD INDEX
- RECOVER
- REORG INDEX
- REORG TABLESPACE
- REPAIR
- UNLOAD

- Utilities will only ever process base or clone objects, **never** both.

In the absence of a CLONE option or keyword the utilities will operate against base objects only by default

LISTDEF can have INCLUDEs and EXCLUDEs for both base and clone objects but when the final list is used with a utility the utility will only process base or clone objects.

- Example: If your LISTDEF result has 100 objects in it and only 5 of them have clones then when you use this list with the COPY utility with no CLONE keyword then 100 base objects will be copied. If you use the same list with COPY and specify the CLONE keyword then only the 5 clone objects will be copied.

- **CHECK DATA *table-space-spec* . . . CLONE . . .**
 - Indicates that CHECK DATA is to check the clone table in the specified table space. Because clone tables cannot have referential constraints, the utility checks only table check constraints and for consistencies between the clone table data and the corresponding LOB data. If you do not specify CLONE, CHECK DATA operates on the base table.
- **CHECK INDEX . . . CLONE**
 - Indicates that CHECK INDEX is to check only the specified indexes that are on table clones.
- **CHECK LOB TABLESPACE *lob-table-space-spec* . . . CLONE**
 - Indicates that CHECK LOB is to check the LOB table space data for only the table clone, note the LOB data for the base table.

- **COPY . . . CLONE . . .**
 - Indicates that COPY is to copy only table clone data in the specified table spaces or indexes on table clones in the specified index spaces. If you use the LIST keyword to specify a list of objects, COPY processes only those objects in the list that contain table clones or indexes on table clones. The other objects in the list are ignored.
- **COPYTOCOPY . . . CLONE . . .**
 - Indicates that COPYTOCOPY is to process only image copy data sets that were taken against table clones or indexes on table clones.
- **DIAGNOSE . . . CLONE . . .**
 - Indicates that DIAGNOSE is to display information for only the specified objects that are table clones, table spaces that contain table clones, indexes on table clones, or index spaces that contain indexes on table clones.

- LISTDEF

- CLONE

- Indicates that the INCLUDE or EXCLUDE expression is to return only clone tables, table spaces that contain clone tables, indexes on table clones, or index spaces that contain indexes on clone tables. If you specify ALL, LOB objects are also included. The CLONE keyword is ignored if you also specify a table name.

- ALL

- Specifies that both BASE and LOB objects are to be included in the list. Auxiliary relationships are to be followed from all objects that result from the initial object lookup, and both BASE and LOB objects are to remain in the final enumerated list. Clone objects are not included.

- TABLE *creator-id.table-name*

- Specifies the table that is to be used for the initial search for the object. . . . If you specify a table name with CLONE, the CLONE keyword is ignored.

- **MERGECOPY . . . CLONE . . .**
 - Indicates that MERGECOPY is to process only image copy data sets that were taken against clone objects.

- **MODIFY RECOVERY . . . CLONE . . .**
 - Indicates that MODIFY RECOVERY is to delete SYSCOPY records and any related SYSLGRNX records for only clone objects. DBD entries for clone tables are not deleted.

- **QUIESCE . . . CLONE . . .**
 - Indicates that QUIESCE is to create a quiesce point for only the specified table spaces that contain clone tables.

- REBUILD INDEX . . . CLONE . . .

- ALL

- Specifies that all indexes in the table space that is referred to by the TABLESPACE keyword are to be rebuilt. If you specify ALL, indexes on table clones are not included; only indexes on the base table are included.

- CLONE

- Indicates that both REBUILD INDEX is to reconstruct only the specified indexes that are on table clones. If you specify CLONE, you cannot specify STATISTICS. Statistics are not collected for table clones.

- RECOVER . . . CLONE . . .

There are no special recovery considerations for a cloned table space or cloned index. **Cannot do PIT recovery to a time that precedes that last exchange.**

- CLONE

- Indicates that RECOVER is to recover only table clone data in the specified table spaces or the specified index spaces that contain index on table clones.

- REORG INDEX . . . CLONE . . .

- Indicates that REORG INDEX is to reorganize only the specified index spaces that contain indexes on table clones.

- REORG TABLESPACE . . . CLONE . . .

- Indicates that REORG TABLESPACE is to reorganize only table clones from the specified table spaces. Base tables in those table spaces are not reorganized. If you specify CLONE, you cannot specify STATISTICS. Statistics are not collected for table clones.

Note: FASTSWITCH cannot be specified for REORG INDEX and REORG TABLESPACE utilities on any object involved with cloning (base table, clone table, indexes)!

- REPAIR . . . CLONE

- Indicates that REPAIR is to process only the specified objects that are table spaces that contain table clones, indexes on table clones, or index spaces that contain indexes on table clones. If you specify CLONE, you cannot specify VERSIONS, because table clones do not have versions. **Clones cannot be created for tables with active versions.**

If you specify SET with CLONE, the status is changed for only the specified clone tables and their indexes. The CLONE keyword applies to all SET statements and LOCATE statements within the same REPAIR utility control statement.

- UNLOAD . . . CLONE . . .

- Indicates that UNLOAD is to unload data from only clone tables in the specified table spaces. If you specify the name of the clone table in the FROM TABLE clause, you do not need to specify the CLONE keyword.

- MODIFY STATISTICS
- REPORT
- RUNSTATS

- **MODIFY STATISTICS . . .**

The online MODIFY STATISTICS utility deletes unwanted statistics history records from the corresponding catalog tables. You can remove statistics history records that were written before a specific date, or you can remove records of a specific age. You can delete records for an entire table space, index space, or index.

Run MODIFY STATISTICS regularly to clear outdated information from the statistics history catalog tables. By deleting outdated information from those tables, you can improve performance for processes that access data from those tables.

Restriction: MODIFY STATISTICS does not process table clones, because statistics are not collected for these tables.

- **REPORT**

- The report provides information for both base and clone objects. If you specify TABLESPACESET, REPORT also processes related LOBs.

- **RUNSTATS**

- No statistics are kept for a clone table so CLONE cannot be specified.

- Clone tables gives us an Online Load Replace type of capability.
- Lots of capability
- Online schema restrictions
- Future?

Attack of the DB2 9 for z/OS Clones Clone Tables That Is!

John Lyle
IBM Silicon Valley Lab. (Etats Unis)

jlyle@us.ibm.com

