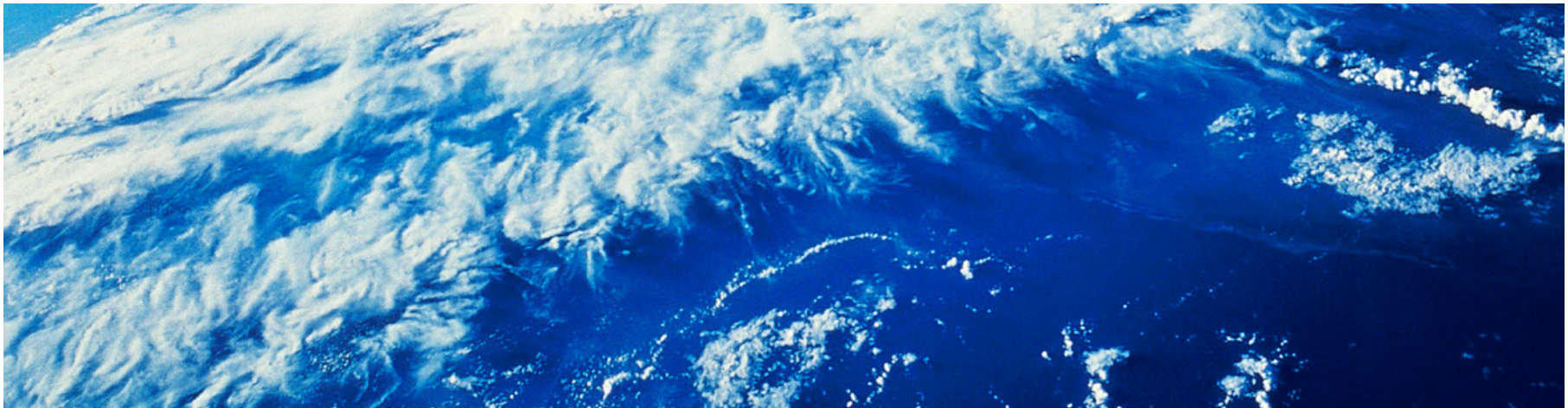


IBM WebSphere Commerce V7 FEP8

DBClean Offline Mode Large Data Cleanup



Melissa Singh

Agenda

- Introduction
- Traditional DBClean vs. Offline Mode
- Offline Mode Solution
- Assumptions/Limitations
- Troubleshooting

Traditional DBClean

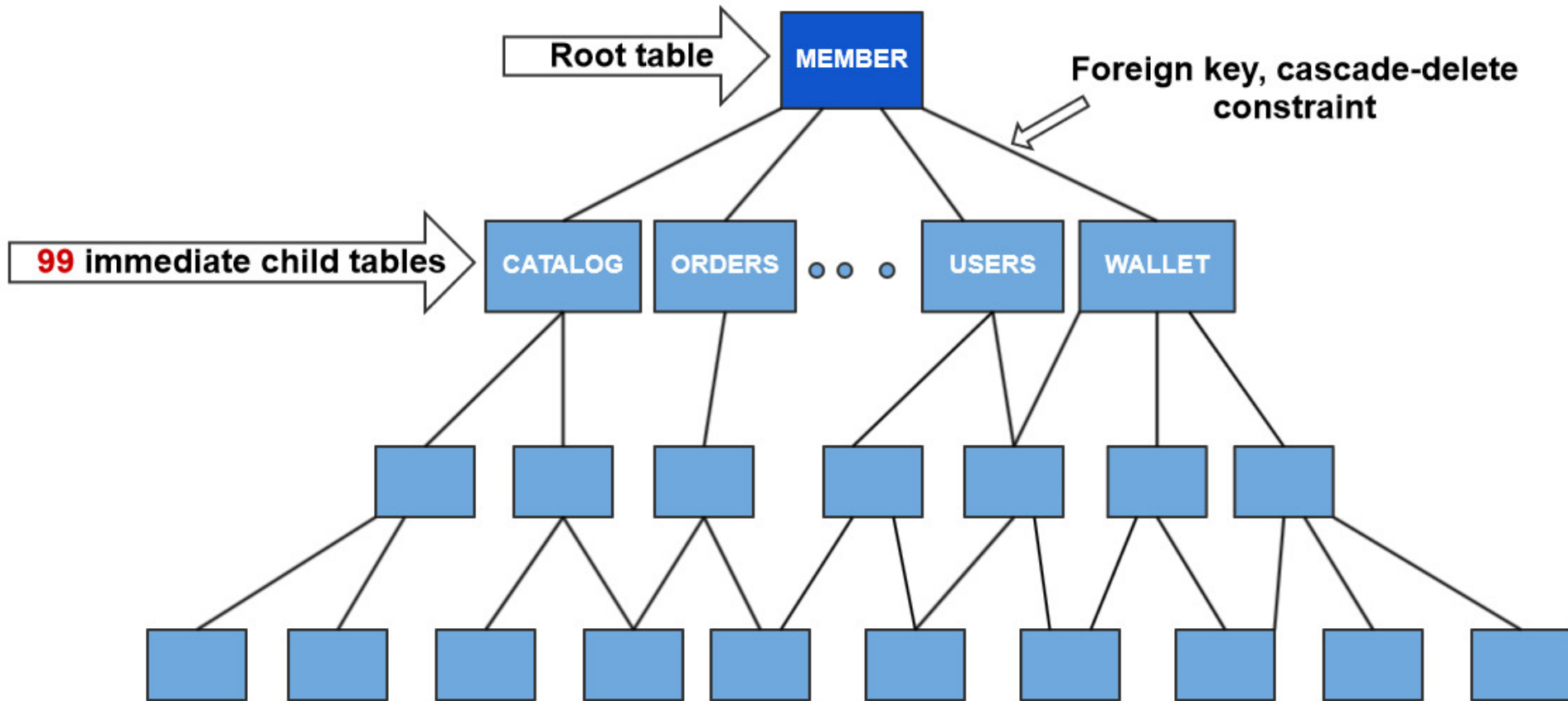
- Uses delete statements stored in the CLEANCONF table to perform deletes against a particular table
- The deletes then cascade delete from this **root** table to its descendant tables

Example for the **guest user** object:

```
DELETE from MEMBER where member_id in (select users_id  
from users where (days(CURRENT TIMESTAMP) -  
days(prevlastsession)) >= ? and (users_id > 0))
```

What happens when there are 100s of descendant tables to be cleaned?

Deeply/Widely nested schema



Performance Issues

- Cascade-delete to 100s of tables is too costly
- Without regular cleanup of guest user records, we've seen customers end up with 10M+ guest user records to be cleaned
- Traditional DBClean is inefficient

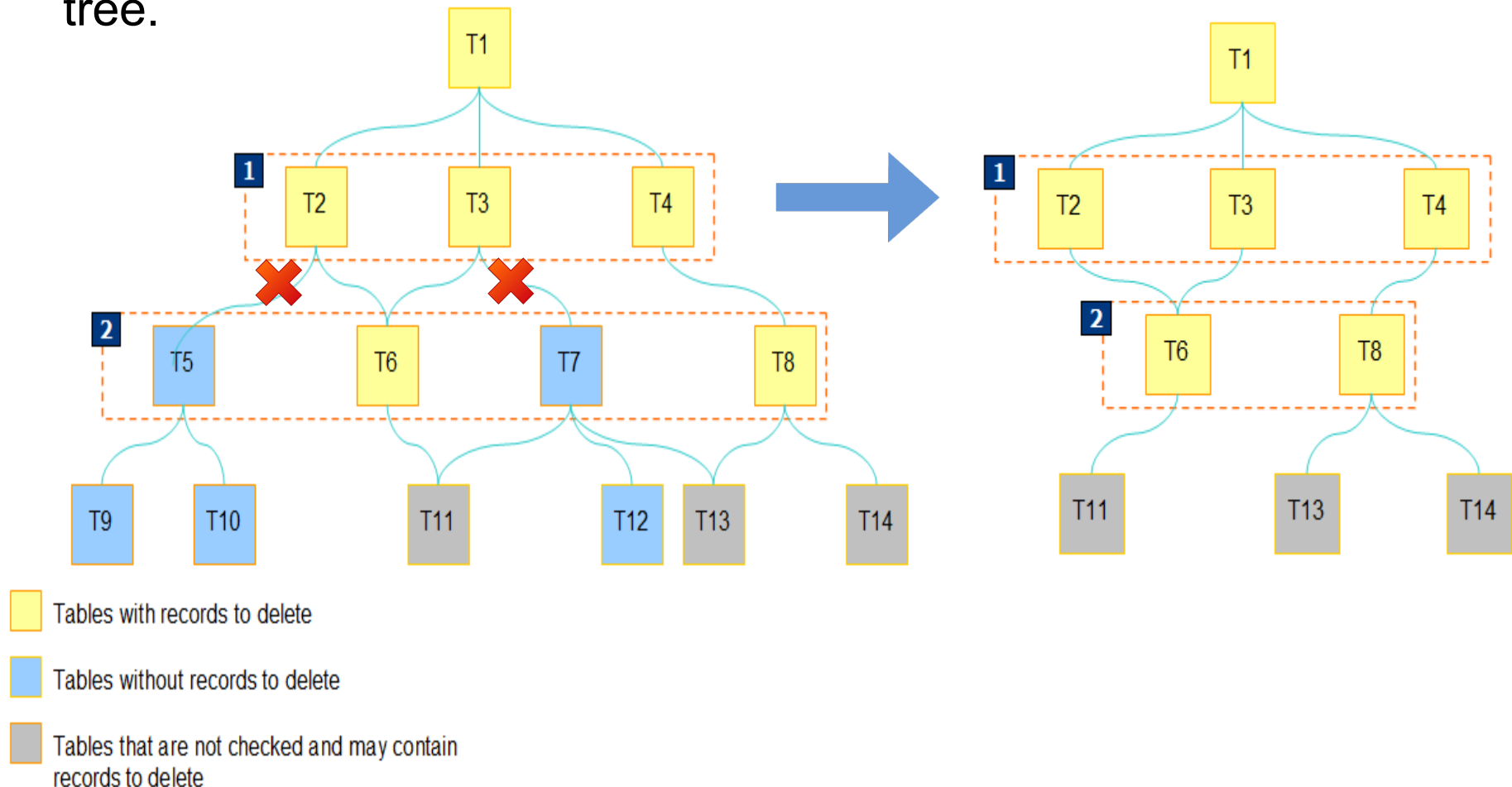
We need a solution to handle clean up of a **large amount of data** in schema which is **deeply/widely nested**!

Offline Mode Solution

- Remove the use of the cascade deletes
- Generate explicit deletes for each table in the hierarchy
- Execute the deletes concurrently in a bottom-up fashion

Solution: Offline Mode Cleanup

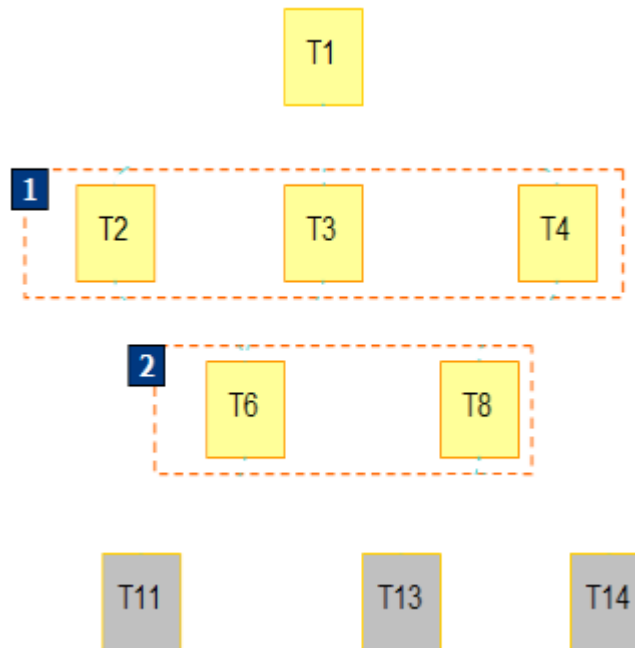
Step 1: Shrink the hierarchal tree by checking tables to see if they have actual data to be deleted. If they don't, the table and subsequently it's descendant child tables are pruned from the tree.



Solution: Offline Mode Cleanup

Step 2: Generate explicit delete statements for each table in the reduced hierarchy.

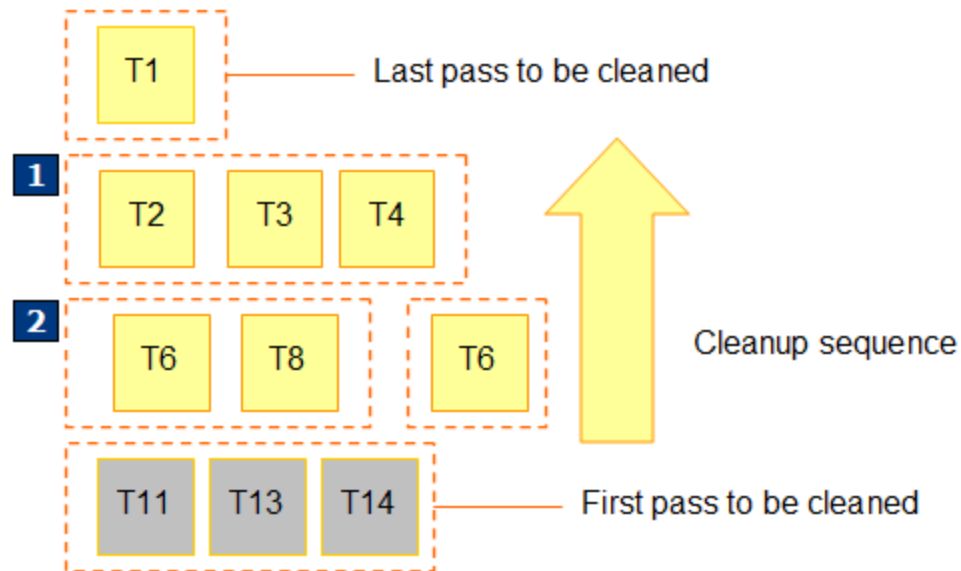
Step 3: Temporarily disable the foreign-key relationships amongst the tables of the hierarchy



Solution: Offline Mode Cleanup

Step 4: Execute the explicit deletes bottom-up. The deletes can be executed concurrently for each hierarchy level.

- The utility creates **passes** to group the delete statements which can be executed concurrently.
- Each pass cannot contain multiple delete statements for the same table (this is done to avoid waiting scenarios due to table level locking)



Solution: Offline Mode Cleanup

Step 6: Temporarily disable referential integrity checking for each disabled constraint.

- These checks are done to ensure there aren't any foreign key constraint violations (no holes in the data).
- When the foreign key constraints are re-enabled, this check is kicked off and it can be incredibly time consuming (especially for tables with lots of data)

Step 7: Re-enable the disabled foreign key constraints

Step 8: Re-enable referential integrity checks

How to run DBClean in Offline Mode

Optional parameters:

- offlineMode: 'yes' indicates the utility is to run in offline mode. (Default 'no')
- pruneLevel: The number of hierarchy levels to check for tables to be pruned from the table hierarchy. (Default 2)
- threadCount: The number of threads (delete statements) that can be executed concurrently. (Default 8)
- validateDel: 'yes' indicates the utility will check child tables for any foreign-key constraint violations.

From command line:

```
dbclean -dbtype db2 -db mall -dbuser wcs -dbpasswd **** -object  
guest -type user -offlineMode yes -pruneLevel 1 -threadCount 4  
-validateDel no
```

Assumptions/Limitations

- Support for DB2 and Oracle only
- Utility cannot detect if the environment is offline
- Replication services should be turned off during the clean
- Should only be used in scenarios where the traditional DBClean cannot be used (i.e. 10M+ records to be cleaned in a deeply/widely nested schema)

Assumptions/Limitations

- If the utility fails before or during the clean, the disablement and re-enablement of the foreign keys and referential integrity checking will still be executed
- Best used for the cleanup of guest user records but this is a generic solution
- Given that the referential integrity checking is disabled, the data for the tables is left in an 'unchecked' state! Thus, we need to ensure the environment is in fact offline to avoid any deletes in the environment.

References

- Database Cleanup Utility

[http://www-](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cduccl)

[01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cduccl](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cduccl)
[ean.htm?lang=en](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cduccl)

- Database Cleanup utility command script

[http://www-](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/refs/rducommand.htm?cp=SSZLC2_7.0.0%2F6-8-1-1-5)

[01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/refs/rducommand.htm?](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/refs/rducommand.htm?cp=SSZLC2_7.0.0%2F6-8-1-1-5)
[cp=SSZLC2_7.0.0%2F6-8-1-1-5](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/refs/rducommand.htm?cp=SSZLC2_7.0.0%2F6-8-1-1-5)

- Database maintenance guidelines

[http://www-](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cdcdbmaintenance.htm)

[01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cdcdbmaintena](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cdcdbmaintenance.htm)
[nce.htm](http://www-01.ibm.com/support/knowledgecenter/SSZLC2_7.0.0/com.ibm.commerce.admin.doc/concepts/cdcdbmaintenance.htm)

Thank You!

Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, Coremetrics, DB2, PowerVM, Rational, WebSphere, and z/VM are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "

Copyright and trademark information" at <http://www.ibm.com/legal/copytrade.shtml>

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

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, 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 PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION 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 ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2014. All rights reserved.