

Part 9. pureXML Utility Features and Impact on DBAs

Hello DB2 friends! My name is Guogen Zhang. I'm a Senior Technical Staff Member at IBM Silicon Valley Lab, responsible for the delivery of pureXML in DB2 for z/OS server.

This is part 9 of the pureXML podcast series. The title is “pureXML Utility Features and Impact on DBAs.”

The design principle of pureXML in DB2 for z/OS is to leverage existing mature optimized infrastructure to support XML features. Especially, the Universal Table Spaces are used for XML data, i.e. Partition-By-Growth (PBG) or Partition-By-Range (PBR) table spaces. And B-tree index structures are used for XML related indexes, including NodeID indexes and XML value indexes. Therefore, all the utilities and tools work as usual over XML objects as you would expect.

I will quickly go over the enhancements made to individual utilities, then talk about backup and recovery operation and performance monitoring briefly.

All DB2 utilities are enhanced to handle XML data type and XML related database objects properly. Here is a list of utility features and restrictions related to XML:

- CHECK DATA: checking consistency between a base table space which contains XML columns and its XML table spaces.
- CHECK INDEX: checking of the DocID, NodeID and XML value indexes.
- CHECK LOB: adds error checking to disallow processing of XML table spaces.
- COPY INDEX: support taking full image copies and concurrent copies of DocID, NodeID and XML value indexes.
- COPY TABLESPACE: support taking full, incremental image copies and concurrent copies of XML table spaces.
- COPYTOCOPY: support the replication of image copies of XML table spaces, DocID, NodeID and XML value indexes.
- EXEC SQL: adds error checking to disallow cross loading of tables with XML columns. This is a restriction.
- LISTDEF: implements a new XML keyword for constructing lists with or without XML objects.
- LOAD: supports loading of a table with XML columns - file reference is supported for XML.
- MERGECOPY: supports merging of image copies of XML table spaces with existing function.
- QUIESCE TABLESPACESET: includes XML table spaces and index spaces in the set of quiesced objects.
- REAL TIME STATISTICS: gathers existing statistics on the new XML objects.
- REBUILD INDEX: supports rebuilding of DocID, NodeID and XML value indexes. Restriction: XML value index shrlevel reference only.

- RECOVER INDEX: supports recovery of DocID, NodeID and XML value indexes and includes XML objects during consistency checking of point-in-time recoveries.
- RECOVER TABLESPACE: supports recovery of XML table spaces and includes XML objects during consistency checking of point-in-time recoveries.
- REORG INDEX: supports reorganization of DocID, NodeID and XML value indexes.
- REORG TABLESPACE: supports reorganization of XML table spaces and of base table spaces with XML columns with some restrictions (REBALANCE not supported, UNLOAD EXTERNAL not supported).
- REPORT TABLESPACESET: includes XML table spaces, DocID, NodeID and XML value indexes in the set of reported objects.
- RUNSTATS INDEX: processes base table DocID indexes normally, collects basic statistics for NodeID indexes and XML value indexes.
- RUNSTATS TABLESPACE: processes the base table space DocID column normally and collects basic statistics for all XML table space columns.
- UNLOAD: supports unloading of a table containing XML columns. File reference is supported. UNLOAD of XML data FROMCOPY is not supported.

In addition, all the standalone utilities also support XML objects. Since an XML table space contains string IDs in XML data that are specific to each catalog, (stringIDs are in SYSIBM.SYSXMLSTRINGS catalog table) an XML table space copied by DSN1COPY cannot be moved to another DB2 subsystem.

The database operation and recovery are similar to that of a database with LOB data. The following provides guidelines.

- To recover base table space, take image copies of all related objects.
 - Use LISTDEF ... XML to obtain a list of related objects including XML objects.
 - QUIESCE is no longer needed in DB2 9 for consistent point-in-time recovery before taking image copies.
- COPYTOCOPY may be used to replicate image copies of XML objects.
- MERGECOPY may be used to merge incremental copies of XML table spaces.
- Point in Time recovery (RECOVER TORBA, TOLOGPOINT)
 - All related objects, including XML objects must be recovered to a consistent point in time.
- Optional: CHECK utilities to validate base table spaces with XML columns, XML indexes and related XML table spaces.
- If there is an availability issue with one object in the related set, availability of the others may be impacted.

Check consistency when an XML column is involved:

1. Check DOCID index on the base table, check NODEID index and XML indexes on the XML table space.
2. Check data against the base table space, which will check the consistency of base table with XML table using the NodeID index.

When there is inconsistency between DOCID index and base table data, because DOCID index is just a regular index so whatever you do normally will apply, for example, if you

confirm base table data is correct, then you rebuild DOCID index. When there is inconsistency between NODEID index and XML table data, rebuild or recover NODEID index if data is correct, or REPAIR LOCATE RID DELETE to delete orphan rows if index is correct.

Based on CHECK DATA output, use the REPAIR LOCATE KEY or RID to locate base table rows or XML table rows to delete.

For performance monitoring, since native XML support in DB2 9 is built on top of regular table spaces, index structures, and buffer pools, there are no special changes in DB2 Performance Expert other than the new XML document lock (type ID x'35'). XML performance problems can be analyzed through accounting traces and performance traces as usual.

Some configuration information may help you. DB2 9 introduces a new load module called DSNXML in the DBM1 address space for most of native XML processing. Implicit or explicit XMLPARSE invokes z/OS XML System Services within the same address space. Stored procedures are used for XML schema registration, and a user-defined function is used for XML schema validation. All other query processing is within the DBM1 address space.

Again, if you have any questions or feedback, send us emails at db2zxml@us.ibm.com
Bye-bye.