
CDB/Auto-Online Unload CDB/Auto-Unload

CDB/Auto-Unload is a tool that extracts data from DB2 tables and puts it in sequential files in a fraction of the time it takes the IBM DB2 sample program to do the same function. CDB/Auto-Unload also offers more flexibility in the areas of data formatting and conversions. CDB/Auto-Online Unload and CDB/Auto-Unload are functionally equivalent products in all respects, except that CDB/Auto-Online Unload can produce consistent unload files while the data is being updated by DB2 applications.

Throughout this section, any reference to CDB/Auto-Unload refers to both products, unless otherwise noted.

CDB/Auto-Unload Features

The UNLD Statement

The UNLD statement, specified in the utility's SYSIN, is similar to a SELECT statement and permits the specification of most CDB/Auto-Unload options. The UNLD Statement has the following format:

```
UNLD DATA FORMAT format
                OUTDDN(ddname)
                FROM TABLE table
                columns
                WHERE where-clause
                ORDER BY sort-columns
```

The items in italics are variables, and are described in the next few sections.

Output Formats (*format*)

CDB/Auto-Unload supports the following pre-defined output formats.

DSNTIAUL - format compatible with IBM's sample unload program.

UNLOAD - format compatible with the format used by IBM Reorg when using the UNLOAD PAUSE option.

ARCHIVE - A CDB-specified format that is compatible with other CDB/Auto-Utilities such as CDB/Auto-Reorg and CDB/Auto-Load. This format is more efficient and produces more compact rows than the standard IBM formats, and is the preferred approach to exchange data among CDB/Auto-Utilities.

```
//SYSIN DD *
UNLD DATA OUTDDN(ITALY) FROM CUSTOMERS WHERE
COUNTRY="I"
UNLD DATA OUTDDN(FRANCE) FROM CUSTOMERS WHERE
COUNTRY="F"
UNLD DATA OUTDDN(FRANCE) FROM CUSTOMERS WHERE
COUNTRY="E"
UNLD DATA OUTDDN(OTHER) FROM CUSTOMERS WHERE
COUNTRY<"F" AND COUNTRY<"E"
/*
```

Multiple Output Dataset Example

ASCIIDEL - This format is equivalent to the “ASCII Delimited” standard used in personal computers and other ASCII devices. Data unloaded is always in character format, and can be used for subsequent load into any database or application that supports this format. You have full control over the delimiters to be used for decimal or character fields, and over the character to be used to separate fields.

REPORT - A format that generates tabular data. You have control over placement and width of each column.

Output Dataset Selection (*ddname*)

The *ddname* variable can be used to specify the output dataset for a given unload.

Multiple Output Datasets in a Single Pass

If multiple unload statements are found in the same SYSIN for the same input table, CDB/Auto-Unload will perform multiple unloads in a single pass of the input data. Each unload statement may have a different WHERE clause, and can be directed to separate datasets.

The figure entitled “*Multiple output dataset example*” shows how you might unload a table into multiple datasets, each of which contains data for a different country.

```
UNLOAD DATA FORMAT ASCIIDEL
  FROM TABLE PERSONNL.EDUCATION
    (EMPNO, CLASSNO, CLASSNAME)
  WHERE (EMPNO IN
    SELECT (EMPNO
      FROM TABLE PERSONNL.EMPLOYEES
      WHERE SEX = 'F'))
  ORDER BY CLASSNAME
```

Subselect Support Example

Table Selection (*table*)

A single table can be specified via the *table* variable in the UNLD statement. Joins are also supported by way of new keywords.

Column Selection (*columns*)

You have the option to unload all the columns of a row or only some columns. You have full control over column placement, column order and data format.

Data Selection (*where-clause*)

CDB/Auto-Unload allows you to select the rows to be unloaded via SQL.

You can specify a WHERE clause in the SYSIN unload statement. The WHERE clause is fully compatible with the IBM SQL as implemented by DB2/MVS, with the exception of column arithmetic with float or decimal columns.

Subselect Support

In addition to performing consistent copies from tables in Read-Write access mode, CDB/Auto-Online Unload provides support for subselects.

Subselect support is implemented by first obtaining the result set of the subselect clause and then loading it in memory. The main table is then unloaded by matching each row against the subselect result set. Since this matching does not require I/O, this implementation is extremely fast, and is ideal for situations in which the subselect result size is relatively small.

The figure “*Subselect Support Example*” shows an example of an UNLD statement with a subselect that creates a sequential file in Delimited ASCII format listing the education courses attended by all female employees.

Output Sorting (*sort-columns*)

CDB/Auto-Unload will automatically sort the unloaded data via a standard ORDER BY clause.

Parallel Processing of Partitions

When unloading a partitioned tablespace, you may unload partitions in parallel if the output is sent to different output datasets. The figure entitled “*Parallel Unload of Partitions*” illustrates this. If you use a single UNLD statement, then only one output dataset is used, and output processing is serialized. On the other hand, you can use multiple UNLD statements with multiple WHERE clauses, then the output may be sent to multiple output datasets, and both reading and writing are performed in parallel.

Unloading Subsets of Data

You can reduce the size of the output file by skipping “n” qualifying rows and then allowing the next row to be qualified for unloading. A qualifying row is a row that would normally have been selected for unloading because it matches the criteria imposed by a WHERE clause.

Limiting the Rows to Unload

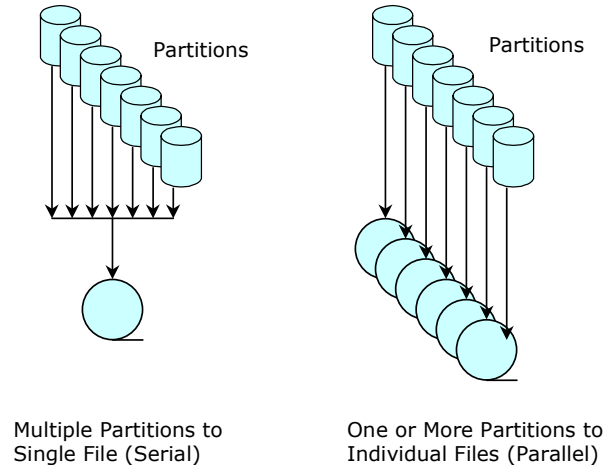
With the +MAXROWS(n) parameter, you can set a limit to the number of rows unloaded. After “n” rows have been unloaded, CDB/Auto-Unload will stop processing.

Unloading from Image Copy

Normally, the input for CDB/Auto-Unload is the DB2 tablespace containing the table referenced in the unload statement. However, you can also unload data from an image copy or DSN1COPY of a tablespace.

Specifying a user-supplied OBID

CDB/Auto-Unload allows you to override the DB2 defined OBID value when qualifying rows for unloading. By default, CDB/Auto-Unload uses the DB2 assigned value. This capability is useful when you need to unload data from an image copy after the corresponding DB2 table has been dropped.



Parallel Unload of Partitions

Default Values

You may supply default values for null columns through the specification of an IFNULL. If the column value for a row being unloaded is null, CDB/Auto-Unload substitutes the value supplied in the IFNULL clause for the null value. A separate IFNULL clause may be specified for each nullable column in the input table.

CDB/Auto-Online Unload

CDB/Auto-Online Unload offers the possibility of performing consistent unloads from a table that is available to applications in RW mode. A quiesce is performed prior to starting the unload, but the table remains in RW mode for the duration of the unload. CDB/Auto-Online Unload relies, like CDB/Auto-Online Copy and CDB/Auto-Online Reorg, on the CDB Subsystem for operation. The CDB Subsystem is a started task that monitors DB2 activity against the dataset being processed by the CDB utility. When a page is changed by DB2, the CDB Subsystem saves the previous copy of the page, and later feeds it to the utility as needed. This guarantees that the unload will be consistent as of the time of the quiesce.

CDB/Auto-Unload Benchmarks

Auto-UnLoad (V7)
125 Million Rows, 3 Partitioned Indexes, 1 NPI and
18 Partitions

