

IBM Terabyte Club Size Queries



This document contains three queries to help you calculate the size of your database. Each query is explained below followed by the actual queries.

Important: After running the queries against your DB2 Catalog for the table spaces (and tables) in your data warehouse, please enter the size in the appropriate field on the [Terabytes Club online application](#) and also forward the detailed output results via email to wfaver0@us.ibm.com.

Please report any problems with these queries via email to Willie Favero at wfaver0@us.ibm.com.

=====

- **Query 1** has a single query that multiplies the average row length (AVGROWLEN) and the number of rows (CARDF) from SYSIBM.SYSTABLEPART for an approximate number of bytes in each table space partition. It then converts that number to terabytes. The result is given in floating point and decimal. If the decimal value is not greater than one (1), you have less than a terabyte of data in that partition, simple table space, segmented table space or universal table space. If compression is on, the row length is expanded to its approximate full row length and bytes calculated. Each row in the query's output is labeled COMPRESSION ON or OFF.
- **Query 2** contains a single query that calculates the size of your table spaces based on the number of pages with rows (NPAGESF), the number of active pages (NACTIVEF), and average row length times cardinality. This is a simple query that is kind of a sanity check for verifying that all queries have similar results and there were no problems that occurred when running any of the queries.
- **Query 3** is a GROUP BY operation summarizing database.

RUNSTATS must be run prior to using any of the following queries.

These queries were tested using DB2 Version 8 for z/OS NFM and DB2 9 for z/OS.

Changes that need to be made to queries prior to running:

- In all of the following queries, change the “**database_name**” and the “**table_space_name**” variables in the WHERE predicates to the database and table space names that contain your warehouse before running the query.
- If all the table spaces in a database are all part of the data warehouse, then the phrase **AND TSNAME = 'table_space_name'** can be eliminated from the query.
- If the warehouse (or warehouses) are contained in multiple databases and table spaces in those databases participate in the data warehouse, you can substitute **DBNAME IN ('database_name 1','database_name 2', etc...)** for the phrase **DBNAME = 'database_name' AND TSNAME = 'table_space_name'**

Query 1

```
--
SELECT SUBSTR(DBNAME,1,8), SUBSTR(TSNAME,1,8),
       PARTITION,
       CASE COMPRESS
         WHEN 'Y' THEN 'ON'
         ELSE 'OFF' END AS COMPRESSION,
       PAGESAVE, AVGWLEN, CARDF,
       AVGWLEN*CARDF AS "AVGWLEN*CARDF",
       ((FLOAT
        (AVGWLEN)/((100-FLOAT(PAGESAVE))/100))*CARDF)
        /1099511627776 AS TERABYTES,
       DEC( ((FLOAT
        (AVGWLEN)/((100-FLOAT(PAGESAVE))/100))*CARDF)
        /1099511627776,15,6) AS "TERABYTES IN DECIMAL"
       FROM SYSIBM.SYSTABLEPART
       WHERE DBNAME IN ('DSNDB06','DBDWH10')
              AND TSNAME = ('TSLINEIT')
       WITH UR
;
--
```

Query 2

NPAGESF & NACTIVEF

```
--
SELECT  SUBSTR(A.NAME,1,8) AS "TABLE NME",
        SUBSTR(A.DBNAME,1,8) AS DBNAME,
        SUBSTR(A.TSNAME,1,8) AS TSNAME, B.PGSIZE, A.NPAGESF,
        B.NACTIVE, B.NACTIVEF, A.AVGROWLEN, CARDF,
        B.PGSIZE*1024 AS "BYTES/PAGE",
        A.NPAGESF*(B.PGSIZE*1024) AS "TSP IN BYTES",
        (A.NPAGESF*(B.PGSIZE*1024))/1099511627776 AS TERABYTES,
        B.NACTIVEF*(PGSIZE*1024) AS "BYTES/ACTIVE PGS",
        (B.NACTIVEF*(PGSIZE*1024))/1099511627776 AS TERABYTES,
        A.AVGROWLEN*A.CARDF AS "AVGROWLEN*CARDF"
FROM SYSIBM.SYSTABLES A, SYSIBM.SYSTABLESPACE B
    WHERE A.TSNAME = B.NAME AND A.TYPE = 'T'
        AND A.DBNAME = B.DBNAME
        AND A.DBNAME = 'database_name'
        AND A.TSNAME = 'table_space_name'

WITH UR
;
--
```

Query 3

```
SELECT TB.DB, TB.TS, SUM(TB.TERABYTES_DEC) AS UNCOMPRESS_SUM
FROM (
    SELECT SUBSTR(DBNAME,1,8) AS DB,
           SUBSTR(TSNAME,1,8) AS TS,
           PARTITION, COMPRESS, PAGESAVE, AVGROWLEN, CARDF,
           ((FLOAT (AVGROWLEN)/
           ((100-FLOAT(PAGESAVE))/100))*CARDF)/1099511627776
           AS TERABYTES, DEC(
           ((FLOAT(AVGROWLEN)/
           ((100-FLOAT(PAGESAVE))/100))*CARDF)/1099511627776,
           15,5) AS TERABYTES_DEC
    FROM SYSIBM.SYSTABLEPART
    WHERE DBNAME = 'database_name' AND TSNAME = 'table_space_name'
        AND PAGESAVE > 0 AND COMPRESS = 'Y'
    ) AS TB
GROUP BY TB.DB, TB.TS
WITH UR
;
```