Legac-E Education

Count records without processing the file

A question was raised in March 2003 asking whether with very large files, was it possible to count the number of records in the file, using REXX. As REXX I/O isn't particularly efficient the logical conclusion would be to use some other means, however for Fixed Length data sets (RECFM=F, FB, FBS, FBA) it is possible to estimate the number of records without reading the actual data.

The original code was intended to run on z/OS 1.3 in TSO foreground and used a SAY instruction to prompt for the data set name. This updated version was tested in April 2018 on a z/OS 2.2 system and has the SAY instruction removed so that the required data set name can be presented via SYSTSIN.

The job stream which follows was run four times to test the concept. The second step produces 50,000 80-byte records as test data but allocates space in four different ways. The results of the test runs were:

SPACE=	No. of Tracks	No. of Extents	Estimated Capacity	REXX Count
(CYL,(10,10),RLSE)	75	1	52350	52350
(TRK,(10,10),RLSE)	72	8	50256	50256
SPACE=(27920,(10,10),RLSE)	72	15	50256	50256
SPACE=(80,(10,10),RLSE) AVGREC=K	72	5	50256	50256

The estimated count uses the formula:

(BLKSIZE/LRECL) * blks per track * number of tracks

Where BLKSIZE will be 27920 having been left to the system to determine, knowing that two blocks of that size will fit on a track.

The accuracy varied between 0.5 and 4.7%. The reason the figures in columns 4 and 5 do not match the 50,000 records generated are twofold:

- Although the last three rows represent data sets being 100% full, the last block is a short block, i.e. has fewer than 349 records.

- This is exacerbated on the first row by the fact that the RLSE parameter only releases unused cylinders, not unused tracks within the last cylinder containing data. Hence the greater inaccuracy of 4.7%.

Count records without processing the file

```
//         EXPORT SYMLIST=HLQ
//         SET  HLQ=&SYSUID
//S0010    EXEC PGM=IEFBR14
//SYSUT2   DD   DISP=(MOD,DELETE),DSN=&HLQ..TEST.DATA,
//              SPACE=(TRK,0)
//S0020    EXEC PGM=IEBDG
//SYSPRINT DD   SYSOUT=*
//SYSUT2   DD   DISP=(,CATLG),DSN=&HLQ..TEST.DATA,
//              RECFM=FB,LRECL=80,BLKSIZE=0,
//              SPACE=(TRK,(10,10),RLSE)
//SYSIN    DD   *
  DSD OUTPUT=(SYSUT2)
  FD NAME=A,LENGTH=6,STARTLOC=1,PICTURE=6,P'000000',INDEX=2
  FD NAME=B,LENGTH=74,FORMAT=AL,ACTION=RP
  CREATE QUANTITY=50000,NAME=(A,B)
  END
//         IF   RC = 0 THEN
//S0030    EXEC PGM=IKJEFT01
//SYSTSPRT DD   SYSOUT=*
//SYSEXEC  DD   UNIT=SYSDA,SPACE=(80,(100,50)),AVGREC=U,
//              LRECL=80,BLKSIZE=0,RECFM=FB,
//              DSN=&HLQ..SEQ.REXX.EXEC,DISP=(,DELETE)
//SYSTSIN  DD   *,SYMBOLS=JCLONLY
 REPRO INFILE(REXXCODE),OUTFILE(SYSEXEC)
PROFILE NOPREFIX
EXEC &HLQ..SEQ.REXX EXEC
&HLQ..TEST.DATA
//REXXCODE DD *,DLM='?@'
/* REXX Code to estimate the number of records in a file */
Parse Upper Pull dsn .
x = LISTDSI(dsn)                          /* obtain d/s info   */
/*---------------------------------------------------------*/
/*  Ensure the data set has fixed length records.          */
/*---------------------------------------------------------*/
if substr(SYSRECFM,1,1) ª= 'F' Then Do
   Say "Data Set -" dsn "not fixed format - run aborted"
   Exit 16
   End
```

Count records without processing the file

```
/*--------------------------------------------------------*/
/*  Perform the calculation                               */
/*--------------------------------------------------------*/
bfactor = SYSBLKSIZE / SYSLRECL      /* Records per block */
rtrack  = bfactor * SYSBLKSTRK         /* Records per track */
Select
  When SYSUNITS = "CYLINDER" Then Do
    dssize = SYSALLOC * SYSTRKSCYL
    count  = dssize * rtrack
    End
  When SYSUNITS = "TRACK"    Then Do
    dssize = SYSALLOC
    count  = dssize * rtrack
    End
  When SYSUNITS = "BLOCK"    Then Do
    dssize = SYSALLOC
    count  = dssize * bfactor
    End
  Otherwise
    Say dsn " Unknown size - Estimation aborted"
    Exit 16
End
Say dsn "Estimated to hold" count "records."
Exit 0
?@
//        ENDIF
```