CERES Software Bulletin 97-09

September 15, 1997

Data Retrieval Time Comparison Between an HDF-EOS File and a Binary File

Lynn Jimenez (l.u.jimenez@larc.nasa.gov)

PURPOSE:

The purpose of this bulletin is to report the results of a test that compared data retrieval times using a binary file versus an HDF-EOS file. An ES-8 flat file and ES-8 HDF-EOS file were used. The two files were read, certain parameters that fell into a given region and had a particular scene id were extracted, and the parameters were written to a file. Each time listed in Table 1 is the average of five runs.

INPUT:

The specific files used and file sizes are shown below.

Binary file:	ES8_19861001_1_rap	487.373 MB
HDF-EOS file:	ES8_19861001_1_rap.hdfeos	491.261 MB

REQUIREMENTS:

Retrieve and write the required parameters (see specific tests below) that satisfy the following three constraints:

1. 30 degrees < colatitude < 37.5 degrees
2. 150 degrees < longitude < 175 degrees
3. scene id = 12, overcast

TEST 1:

If a measurement satisfied the above requirements, the unfiltered LW was written to a file. This resulted in an output file size of 0.52429 MB. Four Scientific Data Sets (SDSs) were read from the HDF-EOS file. The flat file was a direct access file that was read one record at a time.

TEST 2:

If a measurement satisfied the above requirements, the following parameters were written to a file:

1. colatitude
2. longitude
3. scene id
4. viewing zenith angle
5. solar zenith angle
6. relative azimuth
7. LW unfiltered radiance
8. SW unfiltered radiance
9. LW flux at TOA
10. SW flux at TOA

The output file had a size of 0.565865 MB. Ten SDSs were read from the HDF-EOS file.

TEST 3:

If a measurement met the above requirements, the following parameters were written to a file:

1. jultime
2. colatitude
3. longitude
4. sun position colatitude
5. sun position longitude
6. viewing zenith angle
7. solar zenith angle
8. relative azimuth
9. SW unfiltered radiance
10. LW unfiltered radiance
11. WN unfiltered radiance
12. SW flux at TOA
13. LW flux at TOA
14. radiometric, TOT
15. radiometric, SW
16. radiometric, WN
17. scene id

The output file had a size of 0.972338 MB. Seventeen SDSs were read from the HDF-EOS file.

Table 1. Summary of Results

Test	File Type	'/usr/bin/time' Command			csh Version of 'time' Command
		real	user	sys	time[a]
TEST 1	HDF-EOS	40.72 sec	34.94 sec	2.0 sec	36.99u 1.87s 0:39.23 99.0%
	Flat File	1 min 51.4 sec	55.76 sec	9.2 sec	56.33u 9.33s 1:49.43 59.5%
TEST 2	HDF-EOS	43.36 sec	40.94 sec	1.9 sec	40.87u 1.91s 0:43.16 99.1%
	Flat File	1 min 50.3 sec	56.66 sec	9.5 sec	56.60u 9.44s 1:48.24 60.9%
TEST 3	HDF-EOS	49.46 sec	45.54 sec	2.0 sec	44.81u 2.07s 0:47.26 99.1%
	Flat File	1 min 58.1 sec	57.08 sec	9.36 sec	56.78u 9.46s 2:00.94 54.9%

a. The csh Version of 'time' is defined as follows (using the first time as an example):
   36.99u seconds of CPU time devoted to the user's process
   1.87s seconds of CPU time consumed by the kernel
   0:39.230 minutes and 39.23 seconds of elapsed (wallclock) time
   99.0% total CPU time as a percentage of elapsed time

CONCLUSION:

These tests illustrate a typical investigation to retrieve a subset of data from an ES-8 file. In the case of the binary ES-8, every record must be read sequentially, tested to determine whether it meets the subset conditions, and selected parameters must then be written to a file. In the case of an HDF-EOS ES-8, only the SDSs which are required for either the subsetting test or for the output must be read. The following observations may be made from the timings listed in Table 1:

a) The timings for the binary flat file are relatively insensitive to the number of parameters which are output since the entire record must be read for each test.
b) The timings for the EOS-HDF file increase as the number of SDSs which must be read change from four in Test 1 to seventeen in Test 3.
c) Even as the number of SDSs increase, the EOS-HDF tests require less than half of the wall clock time required for the binary flat file tests.

HDF and HDF-EOS formats offer a more efficient way to retrieve subsets of the ES-8 data product.