

Hierarchical Data Format Release 4 (HDF4)

Существует две реализации формата HDF (4.x и предыдущие версии) и HDF5. Причем они очень сильно отличаются друг от друга, являясь несовместимыми. Драйвер GDAL HDF4Image предназначен для работы с данными формата HDF4. Данный драйвер собран на основе библиотеки NCSA HDF library, поэтому если GDAL собирается вручную, то необходимо её наличие для поддержки HDF4. NASA's Earth Observing System (EOS) разрабатывает собственную реализацию HDF, называемую HDF4-EOS. Данная модификация предназначена для хранения данных дистанционного зондирования (телеметрия со спутников 'Terra' и 'Aqua') и полностью совместима с базовым форматом HDF4. Драйвер GDAL HDF4Image имеет также возможность импортировать данные этого формата. Стоит отметить, что EOS планирует в будущем перейти на хранение данных в формате HDF5-EOS для предоставления данных со спутника 'Aqua'.

Subdatasets

HDF представляет собой контейнер для различных наборов данных. Чаще всего данные хранятся в специальных многомерных массивах (SDS, Scientific Datasets).

Один HDF файл может содержать несколько различных SDS массивов. Они могут различаться размером, размерностью и могут содержать совершенно разнообразную информацию. Если HDF файл содержит один SDS массив с изображением, то выгрузить его можно непосредственным образом. Если же в HDF хранится несколько изображений, то может потребоваться дополнительный шаг по выгрузке. Первый шаг - это получение информации о содержащихся изображениях (массивах SDS), используя *gdalinfo* и второй шаг - непосредственно выгрузка нужного изображения по имени, которое было определено на первом шаге. *gdalinfo* позволяет просматривать список всех SDS массивов входного HDF файла. Имя каждого изображения содержит поле метаданных **SUBDATASET_n_NAME**. Описание каждого изображения хранится в поле метаданных **SUBDATASET_n_DESC**. Для HDF4 имена изображений имеют формат:

```
HDF4_SDS:subdataset_type:file_name:subdataset_index
```

где *subdataset_type* - тип SDS массива, *file_name* - имя входного HDF файла, *subdataset_index* - индекс изображения (используется для внутренних целей GDAL).

Выгрузка данных из HDF в GeoTIFF

1. Выгрузка для случая, если HDF содержит один массив SDS:

```
gdal_translate input.hdf output.tif
```

2. Выгрузка для случая, если HDF содержит несколько массивов HDF:

Если попробовать аналогичным способом выгрузить данные HDF файла, содержащего несколько массивов SDS, то мы получим сообщение:

```
Input file contains subdatasets. Please, select one of them for
reading.
```

В этом случае, как уже было сказано выше, нужно сначала определить имя выгружаемого изображения. Допустим, у нас есть файл с данными MODIS Level 1B с именем MOD02QKM.A2008096.0705.005.hdf, содержащий несколько SDS массивов. Чтобы узнать их имена, выполним команду:

```
gdalinfo MOD02QKM.A2008096.0705.005.hdf > hdf_info.dat
```

После чего открываем файл *hdf_info.dat*:

```
Driver: HDF4/Hierarchical Data Format Release 4
Files: MOD02QKM.A2008096.0705.005.hdf
Size is 512, 512
Coordinate System is ``
Metadata:
  HDFEOSVersion=HDFEOS_V2.9
  HDFEOS_FractionalOffset_40*nscans_MODIS_SWATH_Type_L1B=0.5
  HDFEOS_FractionalOffset_4*Max_EV_frames_MODIS_SWATH_Type_L1B=0
  LOCALGRANULEID=MOD02QKM.A2008096.0705.005.hdf
  PRODUCTIONDATETIME=2009-08-15T07:04:16.000Z
  .....
Subdatasets:

SUBDATASET_1_NAME=HDF4_EOS:EOS_SWATH:"MOD02QKM.A2008096.0705.005.hdf":MODIS_SWATH_Type_L1B
SUBDATASET_1_DESC=[2x8120x5416] EV_250_RefSB MODIS_SWATH_Type_L1B
(16-bit unsigned integer)

SUBDATASET_2_NAME=HDF4_EOS:EOS_SWATH:"MOD02QKM.A2008096.0705.005.hdf":MODIS_SWATH_Type_L1B
SUBDATASET_2_DESC=[2x8120x5416] EV_250_RefSB_Uncert_Indexes
MODIS_SWATH_Type_L1B (8-bit unsigned integer)
Corner Coordinates:
Upper Left ( 0.0, 0.0)
Lower Left ( 0.0, 512.0)
Upper Right ( 512.0, 0.0)
Lower Right ( 512.0, 512.0)
Center ( 256.0, 256.0)
```

Как видно, во входном HDF файле содержится 2 массива SDS. Допустим мы хотим выгрузить первый SDS, можно предварительно узнать информацию о нём:

```
gdalinfo
HDF4_EOS:EOS_SWATH:"MOD02QKM.A2008096.0705.005.hdf":MODIS_SWATH_Type_L1B:EV_250_RefSB
> sds_info.dat
```

Открываем файл *sds_info.dat*:

```
Driver: HDF4Image/HDF4 Dataset
Files: MOD02QKM.A2008096.0705.005.hdf
Size is 5416, 8120
Coordinate System is ``
GCP Projection = GEOGCS["WGS 84",DATUM["WGS_1984",SPHEROID["WGS
84",6378137,298.257223563,AUTHORITY["EPSG","7030"]],TOWGS84[0,0,0,0,0,0,0],AUTHORITY["
GCP[ 0]: Id=, Info=
(0.5,1.5) -> (43.5514907836914,65.9675598144531,0)
GCP[ 1]: Id=, Info=
(492.5,1.5) -> (52.816333770752,65.9040145874023,0)
...
Geolocation:
SRS=GEOGCS["WGS 84",DATUM["WGS_1984",SPHEROID["WGS
84",6378137,298.257223563,AUTHORITY["EPSG","7030"]],TOWGS84[0,0,0,0,0,0,0],AUTHORITY["
X_DATASET=HDF4_EOS:EOS_SWATH_GEOL:"MOD02QKM.A2008096.0705.005.2009227070416.hdf":MODIS
X_BAND=1
Y_DATASET=HDF4_EOS:EOS_SWATH_GEOL:"MOD02QKM.A2008096.0705.005.2009227070416.hdf":MODIS
Y_BAND=1
PIXEL_OFFSET=0
PIXEL_STEP=4
LINE_OFFSET=1
LINE_STEP=4
Corner Coordinates:
Upper Left ( 0.0, 0.0)
Lower Left ( 0.0, 8120.0)
Upper Right ( 5416.0, 0.0)
Lower Right ( 5416.0, 8120.0)
Center ( 2708.0, 4060.0)
Band 1 Block=5416x1 Type=UInt16, ColorInterp=Gray
NoData Value=65535
Band 2 Block=5416x1 Type=UInt16, ColorInterp=Gray
NoData Value=65535
```

Помимо прочей полезной информации из данного файла видно, что интересующее нас изображение содержит 2 канала (Band 1 и Band 1). Допустим, мы хотим выгрузить только первый канал. Это можно сделать следующей командой:

```
gdal_translate -b 1
HDF4_EOS:EOS_SWATH:"MOD02QKM.A2008096.0705.005.hdf":MODIS_SWATH_Type_L1B:EV_250_RefSB
output_band01.tif
```

Особый случай

Файл LISOTD_HRAC_V2.2.hdf

gdalinfo LISOTD_HRAC_V2.2.hdf

Обращаем внимание что нормальная структура SUBDATASETS отсутствует, но каналы с данными есть

```
Driver: HDF4Image/HDF4 Dataset
Files: LISOTD_HRAC_V2.2.hdf
Size is 365, 720
Coordinate System is ``
Metadata:
  valid_range=0, 0.3163778
  long_name=Combined Flash Rate Annual Climatology
  units=Flashes/km^2/day
  _FillValue=0
Geolocation:
  SRS=GEOGCS["WGS 84",DATUM["WGS_1984",SPHEROID["WGS
84",6378137,298.257223563,AUTHORITY["EPSG","7030"]],TOWGS84[0,0,0,0,0,0,0],AUTHORITY["E
X_DATASET=HDF4_SDS:UNKNOWN:"LISOTD_HRAC_V2.2.hdf":2
X_BAND=1
Y_DATASET=HDF4_SDS:UNKNOWN:"LISOTD_HRAC_V2.2.hdf":3
Y_BAND=1
PIXEL_OFFSET=0
PIXEL_STEP=1
LINE_OFFSET=0
LINE_STEP=1
Corner Coordinates:
Upper Left ( 0.0, 0.0)
Lower Left ( 0.0, 720.0)
Upper Right ( 365.0, 0.0)
Lower Right ( 365.0, 720.0)
Center ( 182.5, 360.0)
Band 1 Block=365x273 Type=Float32, ColorInterp=Gray
Band 2 Block=365x273 Type=Float32, ColorInterp=Gray
Band 3 Block=365x273 Type=Float32, ColorInterp=Gray
Band 4 Block=365x273 Type=Float32, ColorInterp=Gray
```

Для обращения к конкретному каналу, необходимо выполнить команду **без** префикса HDF4_SDS

```
gdal_translate -of GTiff -b 201 LISOTD_HRAC_V2.2.hdf output.tif
```

Геопривязка

На данный момент нет общепринятого способа хранения информации о геопривязке внутри HDF файлов. Однако некоторые типы продуктов имеют собственные механизмы хранения информации о геопривязке и некоторые из них поддерживаются библиотекой GDAL (*subdataset_type* показан в скобках):

- HDF4 files created by GDAL (GDAL_HDF4)
- ASTER Level 1A (ASTER_L1A)
- ASTER Level 1B (ASTER_L1B)
- ASTER Level 2 (ASTER_L2)
- ASTER DEM (AST14DEM)
- MODIS Level 1B Earth View products (MODIS_L1B)
- MODIS Level 3 products (MODIS_L3)
- SeaWiFS Level 3 Standard Mapped Image Products (SEAWIFS_L3)

Замечание: Как видно из примеров в нашем случае *subdataset_type=MODIS_SWATH_Type_L1B*. Приведенного типа массива SDS среди поддерживаемых GDAL-ом нет (возможно, это MODIS_L1B), возможно из-за этого при попытке выгрузки данных из HDF GDAL выдал сообщение:

```
Warning 1: Lost metadata writing to GeoTIFF ... too large to fit in tag
```

И если сравнивать результат выгрузки изображений с помощью GDAL и с помощью MRT Tools, то выходные файлы будут отличаться в плане геопривязки. К сожалению возможности проверить совпадение геопривязок для других типов данных не было. Поэтому, если у кого есть желание может проделать это и опистать результаты.

BTW, GeoTIFFs have limits on the number of GCPs that can be recorded (FrankW) Try saving in another format that supports .aux.xml file if you want a lot of GCPs. (ie. PNG)

Выгрузка данных из GeoTIFF в HDF

GDAL поддерживает создание HDF4 SDS массивов. Вы можете создавать как двумерные массивы (одно измерение на каждый канал) или трехмерные, где в третьем измерении хранится номер канала. Все метаданные и описания каналов исходного GeoTIFF изображения сохраняются в атрибуты HDF4. Информация о проекции (если есть) и коэффициенты аффинного преобразования также сохраняются в HDF атрибуты. HDF файлы, созданные при помощи GDAL, имеют специальный атрибут:

```
"Signature=Created with GDAL (http://www.remotesensing.org/gdal/)"
```

который автоматически определяется при чтении для восстановления информации о проекции и матрице трансформаций. При создании HDF средствами GDAL может использоваться всего один ключ:

RANK=n, где **n** - размерность создаваемого SDS. По умолчанию создается трехмерный SDS массив.

Пример создания HDF файла:

```
gdal_translate -of HDF4Image input.tif output.hdf
```

<http://lists.osgeo.org/pipermail/gdal-dev/2009-August/021747.html>