

# NetCDF数据格式及其IDL读写

激光雷达探测技术研究室

<http://lidar.aiofm.ac.cn>



## 概述

文件格式是指计算机存储和处理数据的方式，有的文件格式被设计用于储存特殊的数据。比如图像文件中的文件格式，多媒体文件格式等等。

在遥感与大气探测研究中我们经常使用到的文件格式有 **HDF**格式， **NetCDF**格式。

# NetCDF数据

## 简介

NetCDF全称为Network Common Data Format，中文译为“网络通用数据格式”，是由美国大气研究大学协会（University Corporation for Atmospheric Research, UCAR）的Unidata项目开发的。它是一种面向数据型数据，适于网络共享的数据描述和编码标准。其软件实现形式是一个免费的NetCDF软件包，内含可以访问NetCDF数据的工具程序和多种语言的接口库函数，其接口可以在C、FORTRAN、C++等编程环境下调用，使用起来非常便利。

# NetCDF数据



## 应用领域

NetCDF是一种网络通用数据格式，文件最初的目的是用于存储气象科学中的数据，由于其具有灵活性，能够传输海量的面向阵列（**array-oriented**）数据，现在已经成为许多数据采集软件生成文件的格式，被广泛用于陆地、海洋和大气科学。例如，NCEP（美国国家环境预报中心）发布的再分析资料，NOAA CDC9（气候数据中心）发布的海洋与大气综合数据集（**COADS**）均采用NetCDF作为标准。

# NetCDF数据



## 数据特点

NetCDF数据集(文件名后缀为.nc) 的格式不是固定的，一个NetCDF数据集包含维(dimensions)、变量(variables)和属性(attributes)三种描述类型，每种类型都会被分配一个名字和ID，这些类型共同描述了一个数据集，NetCDF库可以同时访问多个数据集，用ID来识别不同数据集。变量存储实际数据，维给出了变量维度信息，属性则给出了变量或数据集本身的辅助信息属性，又可以分为适用于整个文件的全局属性和适用于特定变量的局部属性，全局属性则描述了数据集的基本属性以及数据集的来源。

# NetCDF数据



## 数据结构

一个NetCDF文件的结构包括以下对象：

NetCDF name {

Dimensions: ... //定义维数

Variables: ... //定义变量

Attributes: ... //属性

Data: ...//数据

}

# NetCDF数据



## 数据结构

下面是一个假定的NetCDF文件结构体，包含了来自地球同步卫星的一副图像

Name:

image.nc

Dimensions:

xsize=1200

ysize=600

# NetCDF数据



## 数据结构

Variable:

```
byte image[xsize, ysize]
```

```
    long_name='Imager visible channel'
```

```
    units = 'Counts'
```

```
    valid_range = 0.255
```

```
double time[ysize]
```

```
    long_name='Seconds since 0000UTC, Jan 1 1970'
```

```
    units = 'seconds'
```

```
    valid_range = 0.0D+0, 10.0D+308
```



# NetCDF数据



## 数据结构

Global Attributes:

```
title = 'GOES Image'
```

```
history = 'Created Wed Jul 14 14:15:01 1993'
```

上面第一个被描述的是image.nc，它包含两个变量，第一个变量（image）是一个1200列、600行的浮点型数组。和image关联的变量属性（long\_name、units和valid\_range）指定了图像image的源、物理单位和有效范围。第二个变量time有600个元素，因为time和image共用一个维ysize，则可以假定time为image中每一行的时间值。Time变量也有属性long\_name、units和valid\_range。title和history为全局属性。

# NetCDF数据



## 语法原理

从数学上来说，NetCDF存储的数据就是一个多自变量的单值函数。

用公式表示就是 $f(x, y, z, \dots) = \text{value}$ 。函数的自变量 $x$ 、 $y$ 、 $z$ 等在NetCDF中叫做维（Dimension），函数值 $\text{value}$ 在NetCDF中叫做变量（Variables）。而自变量和函数值在物理学上的一些性质，比如计量单位（量纲）、物理学名称等在NetCDF中就叫做属性（Attributes）。

用来描述NetCDF文件结构的一种语法格式称为CDL（Network Common Data form Description Language），它包括NetCDF对象（变量、维、属性）的具体定义。

# NetCDF数据



## IDL中常用的NetCDF程序

名称	功能	名称	功能
<code>ncdf_open()</code>	打开一个NetCDF文件	<code>ncdf_attname()</code>	返回一个属性名称
<code>ncdf_close</code>	关闭一个NetCDF文件	<code>ncdf_create()</code>	创建一个NetCDF文件
<code>ncdf_varid()</code>	返回一个变量标识符	<code>ncdf_dimdef()</code>	创建一个维度
<code>ncdf_varget</code>	读取一个变量	<code>ncdf_vardef()</code>	创建一个变量
<code>ncdf_attget</code>	读取一个属性	<code>ncdf_attput</code>	写入属性数据
<code>ncdf_inquire()</code>	返回文件信息	<code>ncdf_control</code>	开始或介绍定义模式
<code>ncdf_varing()</code>	返回变量信息	<code>ncdf_varput</code>	写入变量数据

# NetCDF数据



使用IDL从NetCDF中读取一个变量

例如，从NetCDF文件image.nc中读取并显示变量

“image”

```
IDL>cdfid=ncdf_open('image.nc')
```

```
IDL>varid=ncdf_varid(cdfid,'image')
```

```
IDL>ncdf_varget, cdfid, varid, data
```

```
IDL>ncdf_close, cdfid
```

```
IDL>tvsc1, data
```

# NetCDF数据



## 使用IDL从NetCDF中读取一个变量

在该例中，使用函数`ncdf_open`打开文件，并返回文件标识符`cdfid`。然后使用函数`ncdf_varid`从名为“image”的变量中获取变量标识符`varid`，并调用过程`ncdf_varget`读取变量中的所有内容，文件由`ncdf_close`关闭。最后，结果数组`data`由`tvsci`显示出来。

# NetCDF数据



使用IDL从NetCDF中读取一个属性

例如，从sao.nc文件中读取属性long\_name

```
IDL>cdfid=ncdf_open('sao.nc')
```

```
IDL>varid=ncdf_varid(cdfid,'PRECIP')
```

```
IDL>ncdf_attget, cdfid, varid, 'long_name',attvalue
```

```
IDL>print, attvalue
```

```
Precipitation amount
```

# NetCDF数据



## 使用IDL从NetCDF中读取一个变量

在该例中，调用函数`ncdf_open`打开文件，并返回文件标识符`cdfid`。而函数`ncdf_varid`返回变量`PRECIP`的标识符，最后，调用过程`ncdf_attget`读取属性`long_name`及变量`varid`。

# NetCDF数据



## 解开NetCDF文件中的内容

前面例子介绍了在知道特定变量或属性存在的情况下如何读取NetCDF文件中的数据。但是，用户有时可能需要从一个预先不知道内容的NetCDF中读取数据。在这种情况下，首先要做的是探测文件内容包含哪些内容，然后才能决定那一部分是需要读取的。

为了解开NetCDF文件中的内容，可以调用调查程序被包括在一个函数中。当给定一个文件的标识符`cdfid`后，下面的函数将返回一个字符串数组，该数组包含一个打开NetCDF文件所有变量的名称。如果文件中没有变量，则返回空字符串。此外，一些软件例如HDF explorer等也可以直接查看NetCDF数据。



# NetCDF数据



## 写入NetCDF文件

为了将一个变量写入NetCDF文件中，必须先定义变量的名称和类型，然后再写入变量的数据。例如，下面的语句将例子数据文件ctscan.dat写入到一个NetCDF文件中。

首先，从ctscan.dat中载入数据：

```
IDL>file=filepath('ctscan.dat', subdir='examples/data)
```

```
IDL>openr, lun, file, /get_lun
```

```
IDL>readu, lun, data
```

```
IDL>free_lun, lun
```

# NetCDF数据



## 写入NetCDF文件

如何创建一个新的NetCDF文件并定义变量:

```
IDL>cdfid = ncdf_create('ctscan.nc', /lobber)
```

```
IDL>xdimid = ncdf_dimdef(cdfid, 'xdim', 256)
```

```
IDL>ydimid = ncdf_dimdef(cdfid, 'ydim', 256)
```

```
IDL>varid = ncdf_vardef(cdfid, 'image', $
```

```
IDL> [xdimid, ydimid], /byte)
```

# NetCDF数据

## 写入NetCDF文件

最后，写入属性和数据：

```
IDL>ncdf_attput, cdfid, varid, 'long_name', $
```

```
IDL> 'CT scan from IDL examples'
```

```
IDL>ncdf_attput, cdfid, 'creation_date', systime(), /global
```

```
IDL>ncdf_control, cdfid, /endef
```

```
IDL> ncdf_varput, cdfid,varid, data
```

```
IDL>ncdf_close, cdfid
```

# NetCDF数据

在该例中，先调用函数`ncdf_create`创建文件。如果使用关键字`clobber`，则存在的文件被覆盖。当使用`ncdf_create`创建一个新的NetCDF文件时，则文件自动转入定义模式，这意味着新的变量、属性和维度可以加入文件中。该变量的维度是通过调用函数`ncdf_dimdef`定义的（注意，几个变量可以共享一个维度）。然后创建了一个名为`image`的变量，其维度由`xdimid`和`ydimid`给定。如果当调用`ncdf_vardef`时没有设置数据类型关键字（如`/byte`），则默认的类型为浮点型。同时加入变量属性`long_name`来描述`image`的内容。并加入全局属性`creation_date`。当所有的定义完成之后，则调用过程`ncdf_control`结束定义模式，并进入数据模式。

# NetCDF数据



## 创建坐标变量

坐标变量包含了一个数值数组，其中数值表示的量对应于一个维度。例如，创建一个NetCDF文件来描述不同气压下大气温度的垂直分布图，这就需要创建一个名为**Pressure**的维度。

```
IDL>cdfid = ncdf_create('profile.nc', /clobber)
IDL>dimid = ncdf_dimdef(cdfid, 'Pressure', 10)
IDL>varid = ncdf_vardef(cdfid, 'Temperature', [dimid], /float)
IDL>ncdf_attput, cdfid, varid, 'units', 'K'
IDL>ncdf_control, cdfid, /endef
IDL>data = 300.0 - 5.0*findgen(10)
IDL>ncdf_varput, cdfid, varid, data
```

# NetCDF数据



## 创建坐标变量

为了存储在每个大气高度实际的气压值，则可以创建一个名 **Pressure** 的坐标变量，也可以使用 **Pressure** 维度：

```
IDL>ncdf_control, cdfid, /redef
IDL>dimid = ncdf_dimdef(cdfid, 'Pressure')
IDL>varid = ncdf_vardef(cdfid, 'Temperature', [dimid], /float)
IDL>ncdf_attput, cdfid, varid, 'units', 'hPa'
IDL>ncdf_control, cdfid, /endef
IDL>data = 1000.0 - 15.0*findgen(10)
IDL>ncdf_varput, cdfid, varid, data
IDL>ncdf_close, cdfid
```

# NetCDF数据



## 创建坐标变量

坐标变量含有一个与变量同名的单独维度，它们可以像其他变量一样使用（例如，属性可以像上面例子那样被加入）。坐标变量使用的例子包括为数据的存储定义时间间隔或为数据存储于n维网格而定义网格坐标等。

# NetCDF数据

## 数据主要特点

自描述性

它是一种自描述的二进制数据格式，包含自身的描述信息

易用性

它是网络透明的，可以使用多种方式和操作这些数据

高可用性

在读取大数据集中的子数据集时可以不按顺序直接读取需要访问的数据

可追加性

对于新数据，可沿某一维进行追加，不用复制数据集和重新定义数据结构

平台无关性

NetCDF数据集支持在异构的网络平台间进行数据传输和数据共享。



# 谢谢！

如有问题可与孙路联系，电子邮件：[louissun2009@gmail.com](mailto:louissun2009@gmail.com)