

附录 2

HDF 格式说明

目 录

2.1 HDF 简介	3
2.2 HDF 库介绍	3
2.3 HDF 的基本数据结构	3
2.4 HDF 文件的 3 层交互	4
2.5 FY-3A 卫星数据文件中使用的 HDF 对象	5
2.5.1 文件（全局）属性	5
2.5.2 科学数据集（SDS）	6
2.5.3 虚拟数据（Vdata）.....	6

2.1 HDF 简介

HDF (Hierarchical Data Format) 是用于存储和分发科学数据的一种自我描述、多对象文件格式。HDF 是由美国国家超级计算应用中心 (NCSA) 创建的, 以满足不同群体的科学家在不同工程项目领域之需要。HDF 主要特征是:

- ◇ 自述性: HDF 文件里的每一个数据对象, 有关于该数据的综合信息 (元数据)。在没有任何外部信息的情况下, 应用程序能够解读 HDF 文件的结构和内容。
- ◇ 通用性: 许多数据类型都可以被嵌入在一个 HDF 文件里。例如, 通过使用合适的 HDF 数据结构, 符号、数字和图形数据可以同时存储在一个 HDF 文件里。
- ◇ 灵活性: HDF 允许用户把相关的数据对象组合在一起, 放到一个分层结构中, 可在数据对象中添加描述和标签。它还允许用户把科学数据放到多个 HDF 文件里。
- ◇ 扩展性: HDF 极易容纳新增加的数据模式, 容易与其他标准格式兼容。
- ◇ 跨平台性: HDF 是一个与平台无关的文件格式。HDF 文件无需任何转换就可以在不同平台上使用。

HDF 是存储科学数据的容器, 能够存储图像、多维数组、表格等, 强调存储和 I/O 效率。HDF 在国际科学界和产业界得到广泛支持。HDF 是一个开放式的数据格式, 针对不断出现的新应用, HDF 也在不断的升级。HDF 新的版本支持网络化环境下的资料服务应用需求, 目前已经有大量基于 Internet 的 HDF 数据文件发布和应用程序 (如 JVH 等)。更重要的是, 这种升级与应用软件的编制是无关的, 保证了程序和数据的向上兼容性。

2.2 HDF 库介绍

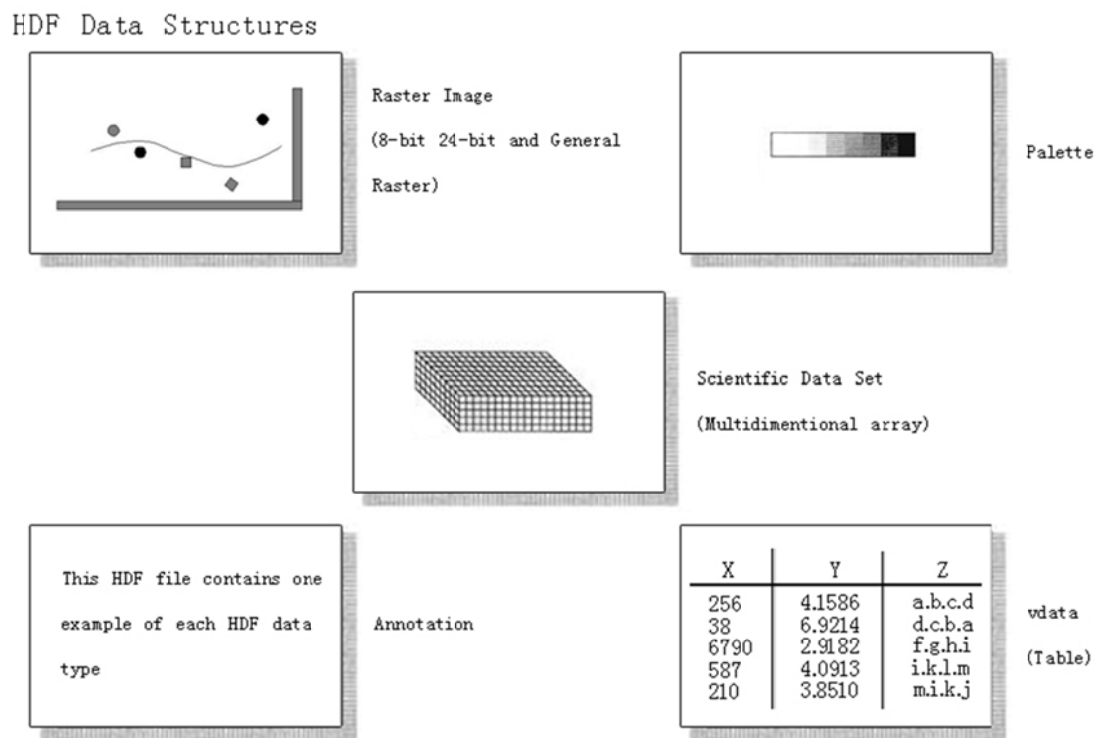
NCSA 提供了各类 UNIX 平台、Linux 平台、Windows 平台的 HDF 库。并且提供的语言接口有 FORTRAN 和 C 语言, 也有用于 Java 程序员访问 HDF 文件的 Java HDF 接口程序。

2.3 HDF 的基本数据结构

HDF 采用分层数据结构存储文件, 包括两种基本数据结构: HDF 组与 HDF 数据集。

- ◇ HDF 组: 用于存储多个 HDF 数据对象的集合, 可以包含多个 HDF 组、HDF 数据集及其元数据。
- ◇ HDF 数据集: 用于记录各种类型的多维数据及其元数据。

HDF 数据集使用的数据类型有：光栅图像 (Raster Image)，调色板 (Palette)，科学数据集 (Scientific Data Set)，注解 (Annotation)，虚拟数据 (Vdata)。如附图 2.3-1 所示。

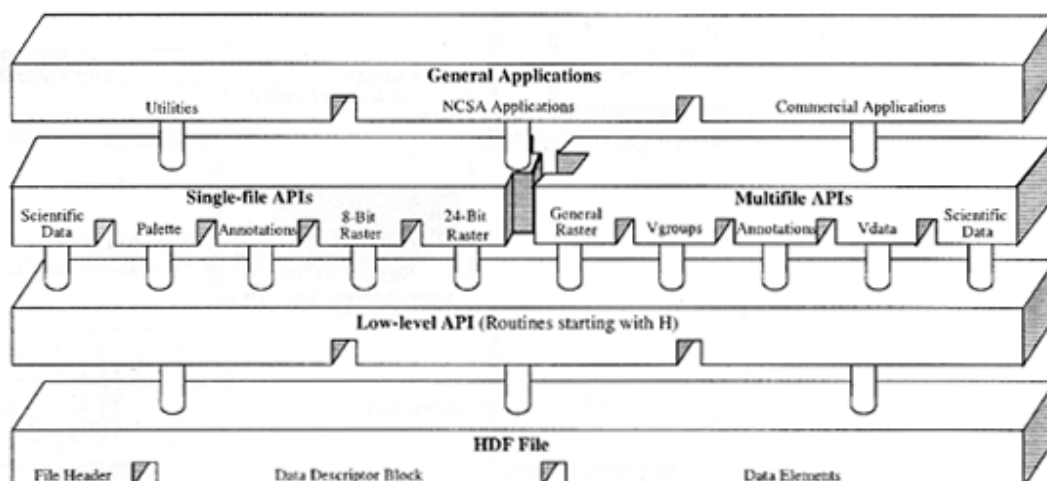


附图 2.3-1 HDF 数据集数据类型

2.4 HDF 文件的 3 层交互

HDF 文件可以在几个交互层次中可视。在最底层，HDF 是一个存储科学数据的物理文件格式。在它的最高层，HDF 是集工具和应用于一体的数据文件，可以对 HDF 文件中的数据进行修改、显示和分析。在这两个层次之间，HDF 是一个能提供高层和底层编程接口的软件库。附图 2.4-1 为这些接口层的示意图。

Three Levels of Interaction with the HDF File



附图 2.4-1 HDF 文件的 3 层交互

基本接口层，或称为底层的应用编程接口（API）是为软件开发者保留的。它是为数据流的直接文件 I/O、错误处理、内存管理和物理存储而设计的。它是一个为有经验的 HDF 程序员提供的软件工具。比较目前从高层接口得到的功能，通过使用这些基本接口层，HDF 程序员创建 HDF 文件时能够做更多的事。底层接口例程只提供 C 语言。

HDF APIs（HDF 应用编程接口）分为两类：多文档接口和单文档接口。多文档接口是提供从一个应用中同时连接几个 HDF 文件的接口，这点很重要，但单文档接口并不支持这点。用户在开发新的接口和界面时，需注意是在一个改进了的新接口版本下开发的。

HDF APIs 包含几个独立的例程集，每个例程集是专门为简化一种数据类型的存储处理而设计的。这些接口在附图 2.4-1 中作为单文档文件和多文档文件层。尽管每个接口都要求程序调用，但所有底层细节都可以忽略。大多数情况下，只须在正确的时间调用正确的函数，剩下的事就由接口程序处理。多数 HDF 接口例程都有 FORTRAN-77 和 C 语言，也有用于 Java 程序员访问 HDF 文件的 Java HDF 接口程序。

2.5 FY-3A 卫星数据文件中使用的 HDF 对象

2.5.1 文件（全局）属性

文件属性又称全局属性。HDF 除了描述每一个科学数据对象 SDS 的特征外，也在总体上描述文件的内容。《风云三号》A 星数据文件（全局）属性记录了扫描时间、轨道参数、扫描线数和仪器状态等信息。

2.5.2 科学数据集 (SDS)

在科学数据集中保存的是探测数据、定位数据、定标数据等信息。科学数据集有一些属性参数，可以对数据对象进行自我描述。附图 2.5-1 是科学数据集对象的组成结构。



附图 2.5-1 科学数据集 (SDS) 的必选属性和可选属性示意图

科学数据集 (SDS) 是多维结构的数据。

如附图 2.5-1，左列是必选属性，右列是可选属性。

◇ 必要属性，描述科学数据集 (SDS) 的基本属性：

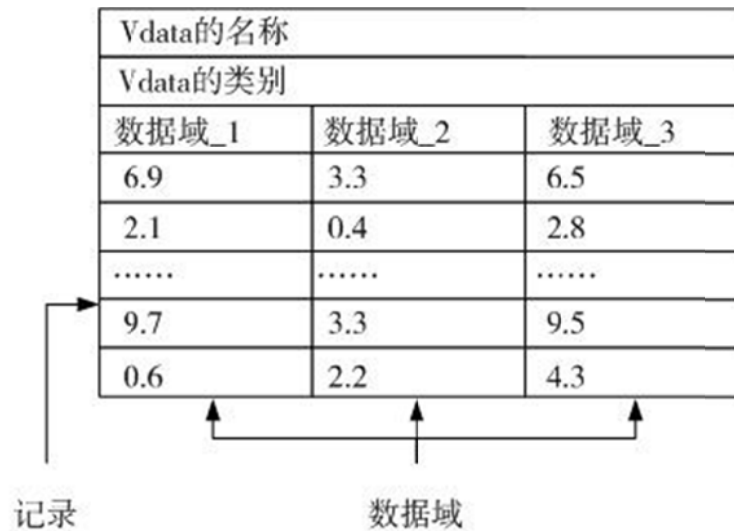
- 名称：该科学数据集具有唯一性的标识名称。
- 数据类型：数据存储类型。
- 数据维数：描述数据各维的大小。

◇ 可选属性：

- 预定义属性，补充描述以下信息：
 - 各维及数据的标记
 - 数据的单位 (units)
 - 值域 (Valid_range)
 - 填充值 (_FillValue)
 - 用于解释或显示数据的坐标系
- 用户定义属性：FY-3A 产品用这类属性 (不同于类似的预定义属性) 指明数据缩放比例 (Slope) 和偏移量 (Intercept)。
- 坐标缩放比例：在解释和显示数据时使用的各维的缩放比例。

2.5.3 虚拟数据 (Vdata)

虚拟数据 (Vdata) 是有固定数据域的表格，FY-3 数据文件中使用 Vdata 保存表格化信息。附图 2.5-2 是其结构示意图。



附图 2.5-2 由 HDF 定义的 Vdata 各项示意图

虚拟数据（Vdata）由名称、类别、数据域名唯一地定义。

- ◇ 名称：从总体上描述数据内容的标记
- ◇ 类别：通过指明数据用途进一步区分 Vdata
- ◇ 数据域名：各数据域的标记
- ◇ 数据域：保存元数据的单元，每列为一类数据域，在 Vdata 生成时要指明其数据类型和级别数（order）。数据类型可以是字符，8 比特、16 比特或 32 比特的有符号的或无符号的整型数、32 比特或 64 比特的浮点数。级别数（order）则指明在数据域中特定数据类型所使用的字的个数。