

Unidata Python Workshop

Overview

(Unidata, python, data)

21 October 2014

Unidata – Who are we?

- Funded primarily by the U.S. National Science Foundation
- Mission:
To provide data, tools, and community leadership for improving Earth-system education and research
- At the Unidata Program Center, we
 - Provide access to data (via push and pull systems)
 - Develop open source tools and **infrastructure** for data access, analysis, visualization, and data management
 - Advance metadata standards for the Earth system science community
 - Support users of our technologies: faculty, students, and researchers
 - Help to build, represent, and advocate on behalf of our community

Unidata – What we provide

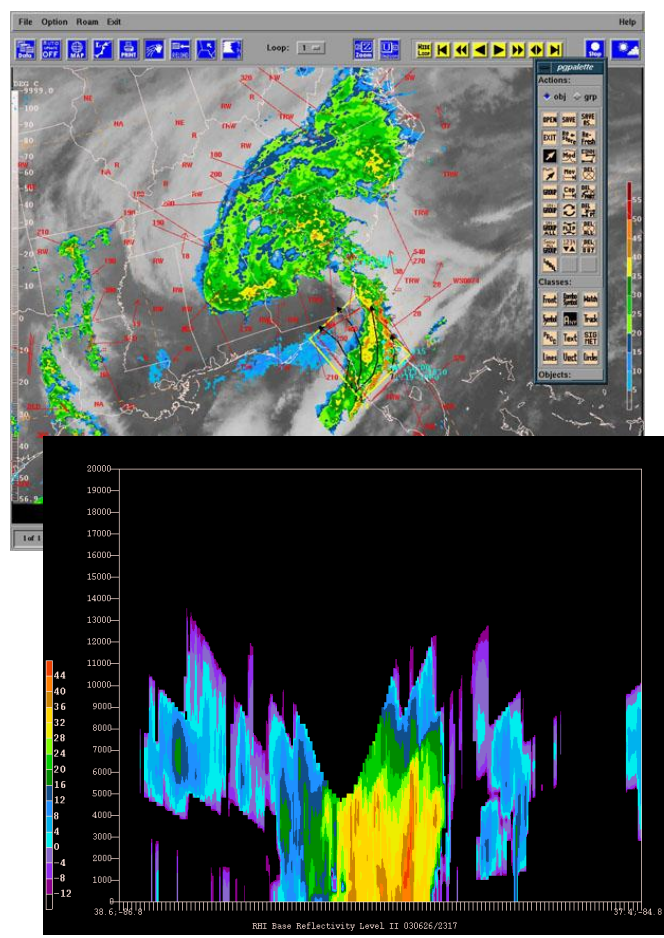
- Real-time data streams over IDD/LDM
 - Radar, satellite, model forecasts, station, etc.
- Visualization and analysis packages
 - IDV, GEMPAK, McIDAS, AWIPS-II
- Rolling archive of real-time data
 - thredds.ucar.edu
- NetCDF data format and libraries

Real-time Data Streams

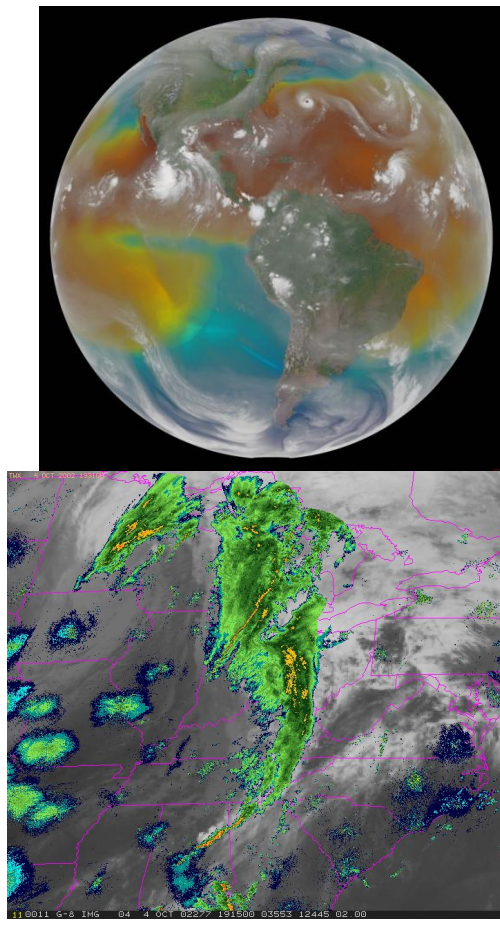
- IDD Real-time Data Flow
 - 30 data feeds provide radar, satellite, text bulletins, lightning, model forecasts, surface and upper air observations, ...
 - LDM routinely handles 10 GB/hour input, with as many as 280,000 products/hour
 - Worldwide collaboration of over 250 institutions running LDM software
- Unidata's LDM
 - Protocol and client/server software
 - Event-driven data distribution
 - Supports subscription to subsets of data feeds

Visualizing and Analyzing Data

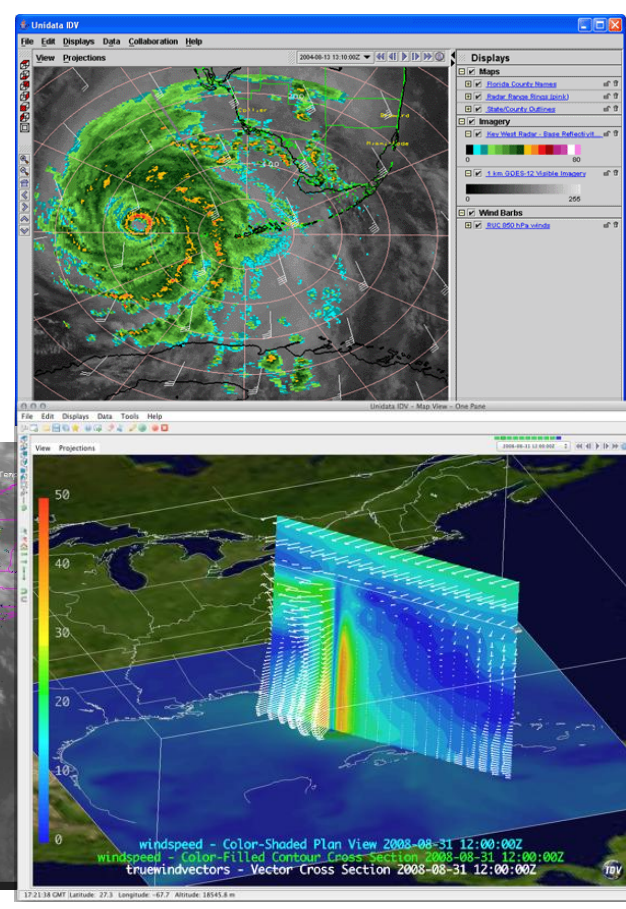
GEMPAK



McIDAS-X



IDV

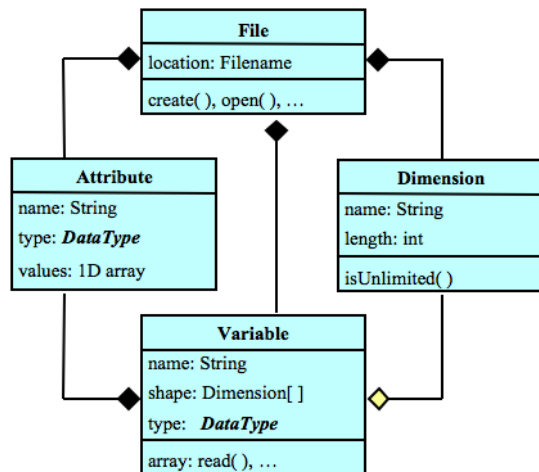


Access to Real-time Data

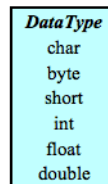
- Unidata TDS: thredds.ucar.edu
 - Serves the last month or so of IDD/LDM data streams
 - Services:
 - OPeNDAP
 - NCSS
 - OGC WCS and WMS
- Unidata ADDE server: adde.ucar.edu
 - Service: McIDAS ADDE

netCDF

- Array-oriented scientific data:
 - Interface for access (C, Fortran, Java, etc.)
 - Machine-independent encoding format
 - Reference libraries (C and Java)

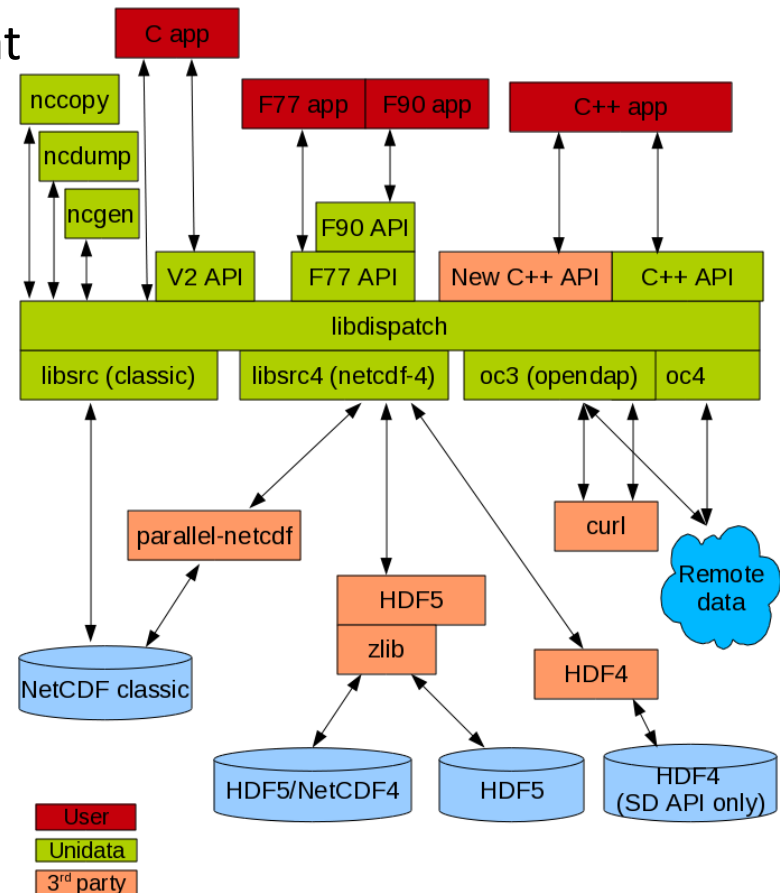


Variables and attributes have one of six primitive data types.



A file has named variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One dimension may be of unlimited length.

NetCDF Library Architecture



Unidata User Community

- Support the community
 - User Workshops
 - Training Workshops
 - Mailing lists
 - For specific software packages
 - “community” email list – for Unidata community announcements
- Represent and advocate for the community
- More: <http://www.unidata.ucar.edu/>

Why Unidata and Python?

- Embraced by the earth science community
 - Language popularity measured by search hits on [AMS web site](#)

Year	Python	Java	Fortran
2011	19	2	4
2012	57	9	2
2013	60	12	9

- Requests from Unidata community for Python support

What is Python?

- General purpose, high-level language invented by Guido van Rossum.
- Multiple paradigms
- scripting
- object-oriented
- imperative
- functional
- "Python is executable pseudocode"



Unidata and Python

"I have used a combination of Perl, Fortran, NCL, Matlab, R and others for routine research, but found out this general-purpose language, Python, can handle almost all in an efficient way from requesting data from remote online sites to statistics, graphics." - UCAR Scientist

- Avenue for exploring, and leveraging netCDF and THREDDS Data Server technologies.
- Embraced by earth science community for analysis and exploration (see table).
- Publication quality graphics and visualization which are improving all the time.

Python Environment

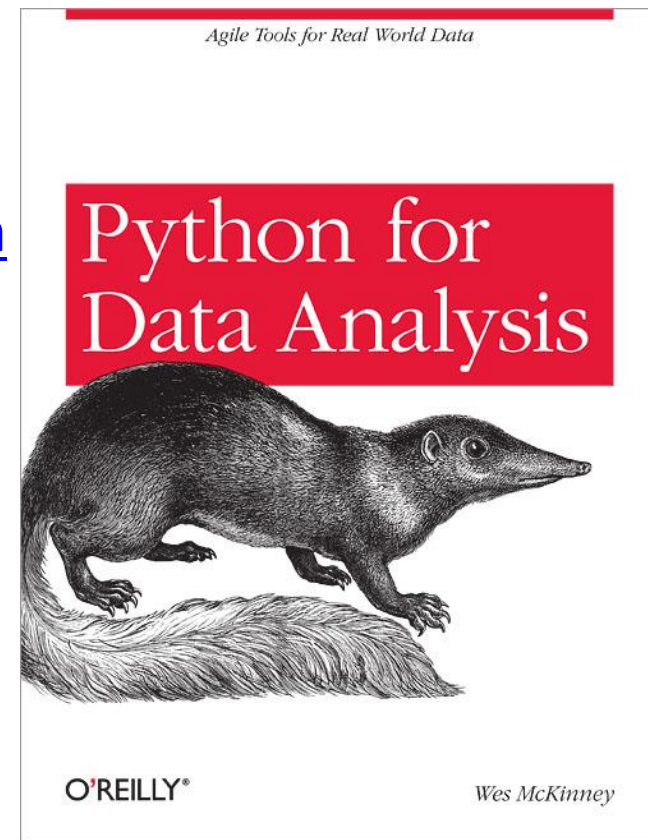
- Setting up a Python environment can be challenging. Some libraries require native dependencies. There are various solutions to this problem.
 - Use a commercial Python distribution with a package manager. Academic use is typically free. This is a good option for beginners but has limitations.
 - “Roll your own” Python library management strategy as you gain experience.
- For various reasons the scientific Python world is still at Python 2.7.x so that is what we will be using.

Python Editor

- Today, we will be using the ipython notebook. It is good for presentations and sharing finished code. It is not so good for code development. Here are some Python IDEs:
 - [Python Tools for Visual Studio](#)
 - [Spyder](#) (Scientific Python Developement EnviRonment)
 - [Emacs Ipython Notebook](#)
 - [Enthought Canopy Editor](#)
 - [Wakari](#), a hosted Python data analysis environment

Python Background material

- [A Hands-On Introduction to Using Python in the Atmospheric and Oceanic Sciences](#)
- [Lectures on scientific computing with Python](#)
- [Why Python is the Next Wave in Earth Sciences Computing](#)
- [Oceanographic Analysis with Python - Rich Signell](#)
- [Python Scientific Lecture Notes](#)
- [Enthought](#)



Where to ask for Help



Tag your questions with `python`, `netcdf`, `thredds`, etc.

Let's Get Started

Today We Will ...

- Read and write netCDF files
- Use matplotlib to visualization geoscience data
- Read model and station data from a TDS NCSS
- Accessing data with PyDAP
- Read Radar Level 2 data
- Request maps from an OGC WMS server