

Introduction to netCDF

Arvind Rajendra Pagire

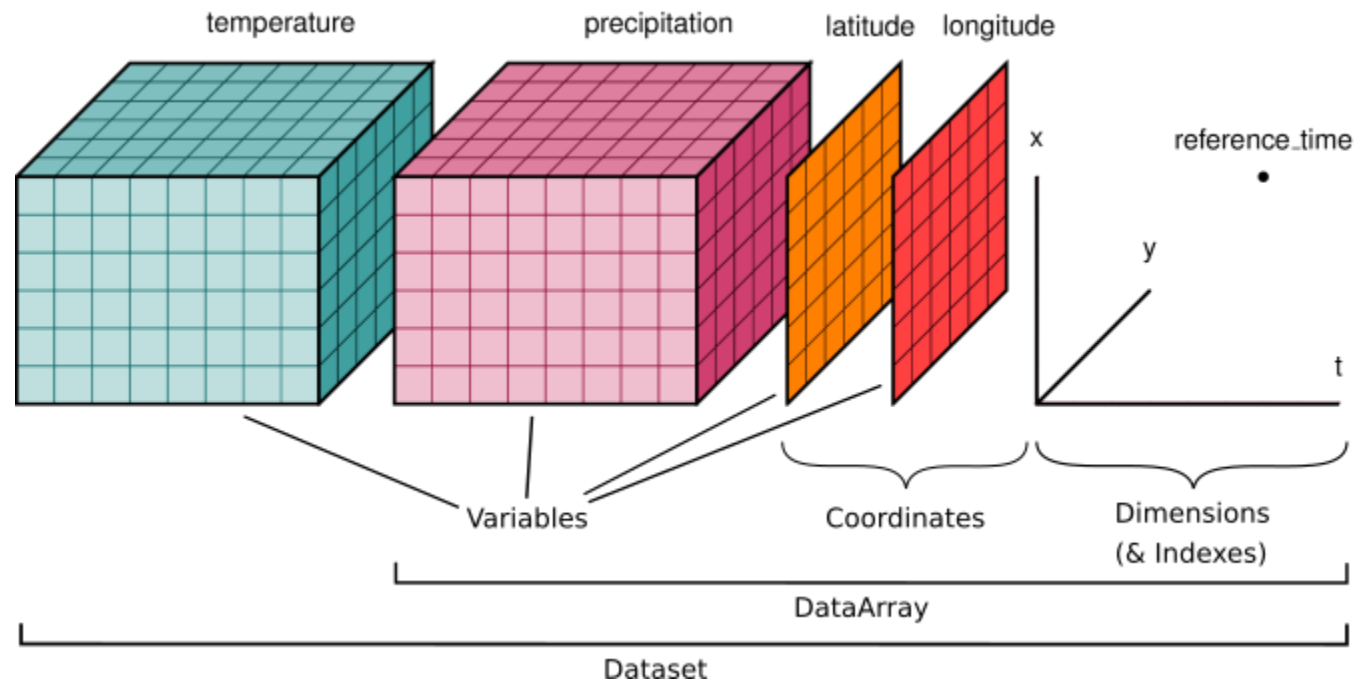
Junior research fellow

Climate monitoring and Prediction Group,

Climate Research and Services,

India Meteorological Department, Pune

Python training at CRS Pune (11-15 September 2023)



- ❖ The Network Common Data Form, or netCDF, is an interface to a library of data access functions for storing and retrieving data in the form of arrays.
- ❖ An array is an n -dimensional (where n is 0, 1, 2, ...) rectangular structure containing items which all have the same data type (e.g., 8-bit character, 32-bit integer).
- ❖ A scalar (simple single value) is a 0-dimensional array.

1 D ARRAY:

C	O	D	I	N	G	E	E	K
0	1	2	3	4	5	6	7	8

← single row of elements

2 D ARRAY:

		col 0	col 1	col 2
	i \ j	0	1	2
row 0	0	A	A	A
row 1	1	B	B	B
row 2	2	C	C	C

← column

} array elements

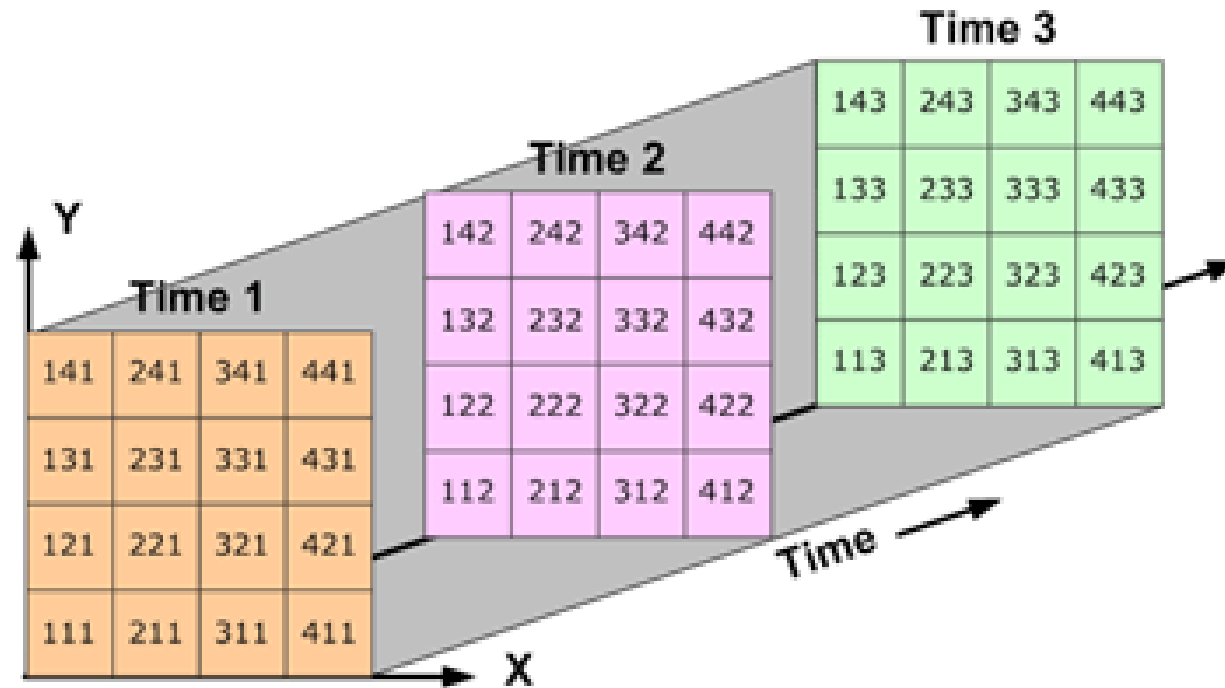
↑ rows

The data in a netCDF file is **stored in the form of arrays**. The data stored in an array needs to be of the same type (homogeneous).

Temperature varying over time at a location is stored as a **one-dimensional array**. You can think of it as a list containing elements of the same data type (i.e. integers, floats).

An example of a **2-dimensional array** is temperature over an area for a given time. A Pandas DataFrame is also a 2-dimensional data structure, but it differs from an array: a DataFrame can store heterogeneous data elements, and you can access it as a spreadsheet (using the column names and rows).

Three-dimensional (3D) data, like temperature over an area varying with time. Think of this as a Pandas DataFrame where the “columns” (variables) have more than one dimension.



1. Spatial information — Location on the surface of the Earth.
2. Time information — At what time of the day and year, the measurements were taken.
3. Scientific values — Temperature, Rainfall, etc.