High Performance Computing with Python

Implementing distance matrices

Rafael Sarmiento and Theofilos Manitaras
ETHZürich / CSCS
CSCS/USI Summer University 2025

ETH zürich
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

NumPy is a Python library that adds support for large multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

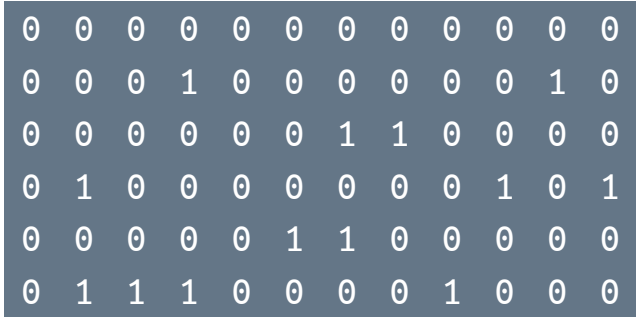
- `numpy.ndarray`: a powerful N-dimensional array object
- Sophisticated functions often written in C
- Linear algebra, Fourier transform, and random number capabilities
- Tools for easy binding to Fortran code (F2PY)
- Compatibility with C
- Many libraries implement the NumPy API, such as Dask and CuPy for graph and GPU computing respectively [NEP 35]

The **SciPy** library provides many user-friendly and efficient numerical routines for operations such as numerical integration, interpolation, optimization, linear algebra and statistics. SciPy builds on the `numpy.ndarray` and expands the set of mathematical functions included in NumPy

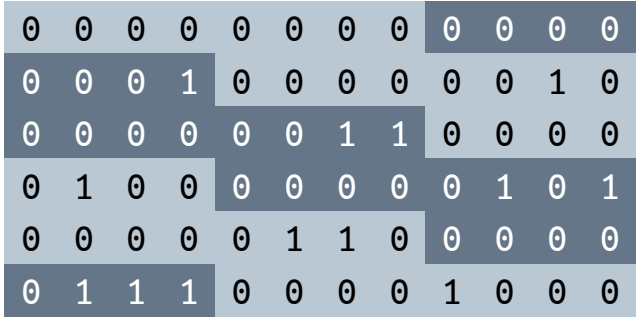
numpy.ndarray

```
>>> x = np.array([[0, 1, 2], [3, 4, 5], [6, 7, 8]], dtype=np.int8)
>>> x
array([[0, 1, 2],
       [3, 4, 5],
       [6, 7, 8]], dtype=int8)
```

numpy.ndarray



numpy.ndarray

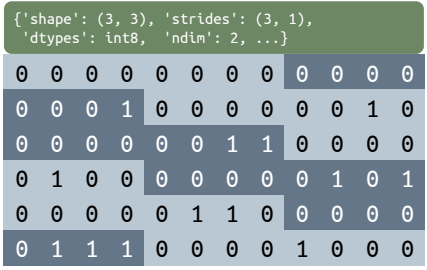


numpy.ndarray

```
{'shape': (3, 3), 'strides': (3, 1),
 'dtypes': int8,  'ndim': 2, ...}
```

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	1	0	0	0	0
0	1	0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	1	1	0	0	0	0	0
0	1	1	1	0	0	0	0	1	0	0	0

numpy.ndarray



- The memory block is called **data buffer**.
- The **metadata** is used to interpret the data buffer within the python context.
- The data buffer is stored in C order (row major) by default.
- All items in the array have the same data type.

numpy.ndarray

#	#	0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15	#	#

numpy.ndarray

Data buffer

#	#	0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15	#	#

NumPy representation

0	1	2	3
4	5	6	7
8	9	10	1
12	12	14	15

```
strides = (4, 1)
shape   = (4, 4)
dtype   = int8
```

numpy.ndarray

Data buffer

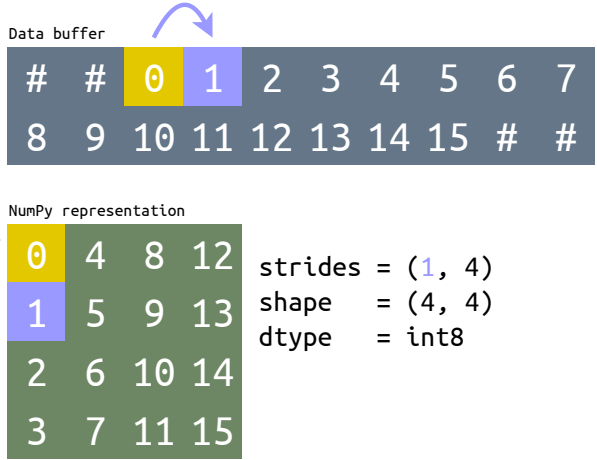
#	#	0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15	#	#

NumPy representation

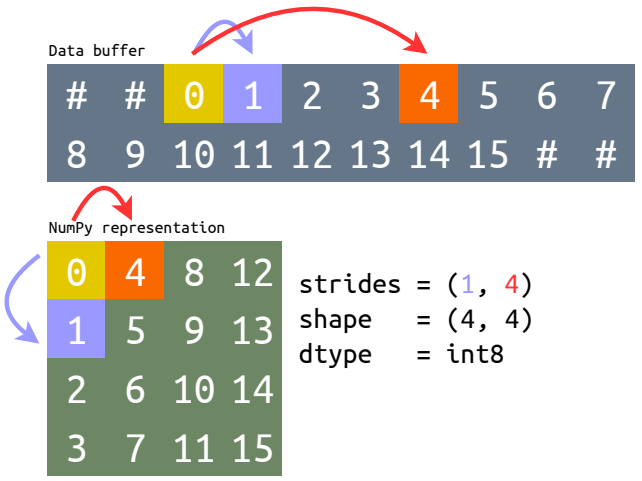
0	4	8	12
1	5	9	13
2	6	10	14
3	7	11	15

```
strides = (1, 4)
shape   = (4, 4)
dtype   = int8
```

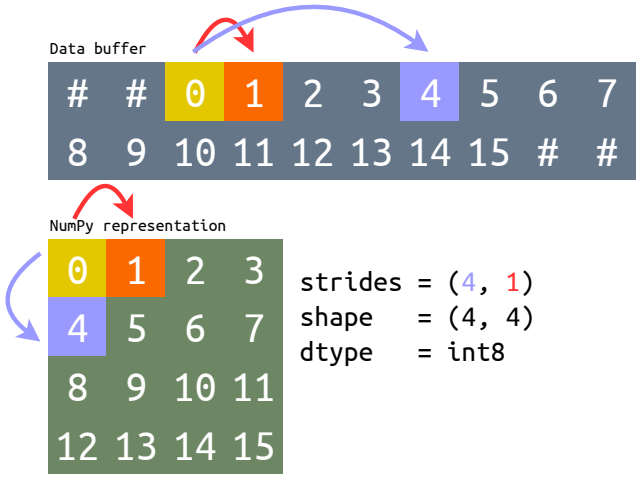
numpy.ndarray



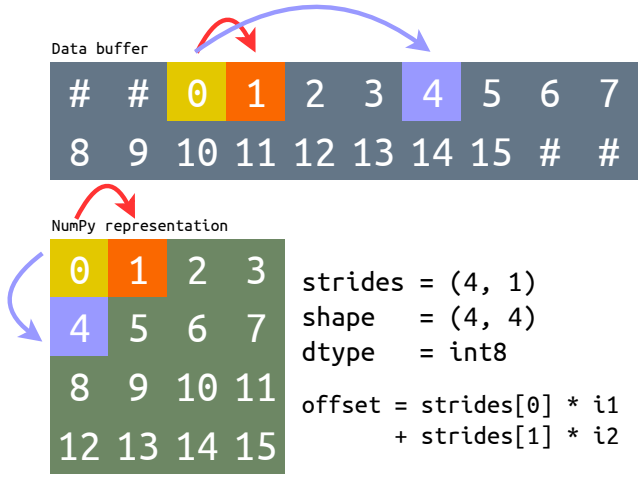
numpy.ndarray



numpy.ndarray



numpy.ndarray



Vectorization

- Use operations over the whole array instead of over single elements.

```
z = x + y              # x = np.array([...])
                       # y = np.array([...])
```

Vectorization

- Use operations over the whole array instead of over single elements.

```
z = x + y           # x = np.array([...])
                    # y = np.array([...])
```

- When working with arrays, use *ufuncs* and general NumPy's functions.

```
y = np.exp(x)       # x = np.array([...])
z = x @ y           # equivalent to np.matmul(x, y)
```

Vectorization

- Use operations over the whole array instead of over single elements.

```
z = x + y            # x = np.array([...])
                     # y = np.array([...])
```

- When working with arrays, use *ufuncs* and general NumPy's functions.

```
y = np.exp(x)        # x = np.array([...])
z = x @ y            # equivalent to np.matmul(x, y)
```

- Adapt your solutions to use the two points above.

Broadcasting

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_n \end{bmatrix}$$

$(n, 1)$ $\qquad\qquad$ $(1, n)$

Broadcasting

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_0 + y_0 & x_0 + y_1 & \cdots & x_0 + y_n \\ x_1 + y_0 & x_1 + y_1 & \cdots & x_1 + y_n \\ x_2 + y_0 & x_2 + y_1 & \cdots & x_2 + y_n \\ \vdots & \vdots & \cdots & \vdots \\ x_n + y_0 & x_n + y_1 & \cdots & x_n + y_n \end{bmatrix}$$

$(n, 1)$ $(1, n)$ (n, n)

[lab] Broadcasting

- Let's open the notebook `numpy/broadcasting.ipynb` and go over the cells and the questions. The goal of this notebook is to understand the broadcasting operations presented there.

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

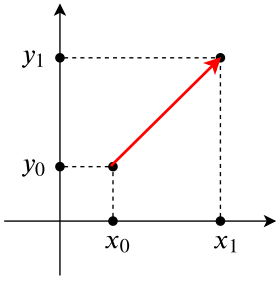
Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Euclidean distance matrix

$$d_{\mathbf{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) = \begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

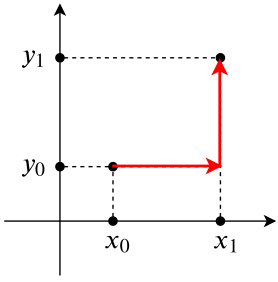
Euclidean distance matrix



$$d_{\mathrm{e}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) =$$

$$\begin{bmatrix} \sum (x_{1i} - y_{1i})^2 & \sum (x_{1i} - y_{2i})^2 & \dots & \sum (x_{1i} - y_{ni})^2 \\ \sum (x_{2i} - y_{1i})^2 & \sum (x_{2i} - y_{2i})^2 & \dots & \sum (x_{2i} - y_{ni})^2 \\ \dots & \dots & \dots & \dots \\ \sum (x_{ni} - y_{1i})^2 & \sum (x_{ni} - y_{2i})^2 & \dots & \sum (x_{ni} - y_{ni})^2 \end{bmatrix}$$

Cityblock distance matrix



$$d_{\text{cb}}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2} & y_{n3} \end{bmatrix}\right) =$$

$$\begin{bmatrix} \sum |x_{1i} - y_{1i}| & \sum |x_{1i} - y_{2i}| & \dots & \sum |x_{1i} - y_{ni}| \\ \sum |x_{2i} - y_{1i}| & \sum |x_{2i} - y_{2i}| & \dots & \sum |x_{2i} - y_{ni}| \\ \dots & \dots & \dots & \dots \\ \sum |x_{ni} - y_{1i}| & \sum |x_{ni} - y_{2i}| & \dots & \sum |x_{ni} - y_{ni}| \end{bmatrix}$$

```
def euclidean_distance_matrix(x, y):
    num_samples = x.shape[0]
    dist_matrix = np.empty((num_samples,
                            num_samples))
    for i, xi in enumerate(x):
        for j, yj in enumerate(y):
            diff = xi - yj
            dist_matrix[i][j] = diff @ diff

    return dist_matrix
```

```
def euclidean_distance_matrix(x, y):
    num_samples = x.shape[0]
    dist_matrix = np.empty((num_samples,
                            num_samples))
    for i, xi in enumerate(x):
        for j, yj in enumerate(y):
            diff = xi - yj
            dist_matrix[i][j] = diff @ diff

    return dist_matrix
```

✗ Use operations over the whole array instead of over single elements.

✓ When working with arrays, use ufuncs and general NumPy's functions.

✗ Adapt your solutions to use the two points above.

Implementation #1

```
def euclidean_distance_matrix(x, y):
    diff = x[:, np.newaxis, :] - y[np.newaxis, :, :]

    return (diff * diff).sum(axis=2)
```

Implementation #2

$$\sum_k (x_{ik} - y_{jk})^2 = (\vec{x}_i - \vec{y}_j) \cdot (\vec{x}_i - \vec{y}_j) = \vec{x}_i \cdot \vec{x}_i + \vec{y}_j \cdot \vec{y}_j - 2\vec{x}_i \cdot \vec{y}_j$$

Implementation #2

$$\sum_k (x_{ik} - y_{jk})^2 = (\vec{x}_i - \vec{y}_j) \cdot (\vec{x}_i - \vec{y}_j) = \vec{x}_i \cdot \vec{x}_i + \vec{y}_j \cdot \vec{y}_j - 2\vec{x}_i \cdot \vec{y}_j$$

$\vec{x}_i \cdot \vec{y}_j \rightarrow$ `x @ y.T` : Matrix product of $\{\vec{x}\}$ and $\{\vec{y}\}$

Implementation #2

$$\sum_k (x_{ik} - y_{jk})^2 = (\vec{x}_i - \vec{y}_j) \cdot (\vec{x}_i - \vec{y}_j) = \vec{x}_i \cdot \vec{x}_i + \vec{y}_j \cdot \vec{y}_j - 2\vec{x}_i \cdot \vec{y}_j$$

$\vec{x}_i \cdot \vec{y}_j \rightarrow$ `x @ y.T` : Matrix product of $\{\vec{x}\}$ and $\{\vec{y}\}$

$\vec{x}_i \cdot \vec{x}_i \rightarrow$ `(x * x).sum(axis=1)` : A vector of elements $\sum_j x_{ij} x_{ij} \equiv \sum_j x_{ij}^2$

$\vec{y}_j \cdot \vec{y}_j \rightarrow$ `(y * y).sum(axis=1)` : A vector of elements $\sum_j y_{ij} y_{ij} \equiv \sum_j y_{ij}^2$

Implementation #2

$$\sum_k (x_{ik} - y_{jk})^2 = (\vec{x}_i - \vec{y}_j) \cdot (\vec{x}_i - \vec{y}_j) = \vec{x}_i \cdot \vec{x}_i + \vec{y}_j \cdot \vec{y}_j - 2\vec{x}_i \cdot \vec{y}_j$$

$\vec{x}_i \cdot \vec{y}_j \rightarrow$ `x @ y.T`

$\vec{x}_i \cdot \vec{x}_i \rightarrow$ `(x * x).sum(axis=1)[:, np.newaxis]`

$\vec{y}_j \cdot \vec{y}_j \rightarrow$ `(y * y).sum(axis=1)[np.newaxis, :]`

Implementation #2

```
def euclidean_distance_matrix(x, y):
    x2 = (x * x).sum(axis=1)[:, np.newaxis]
    y2 = (y * y).sum(axis=1)[np.newaxis, :]
    xy = x @ y.T

    return np.abs(x2 + y2 - 2. * xy)
```

[lab] Euclidean distance matrix with NumPy

- Let's open the notebook `numpy/euclidean-distance-matrix.ipynb` and check step by step what the function `euclidean_trick` does.

dask is a flexible library for parallel computing in Python. It provides dynamic task scheduling optimized for computation as well as big data collections like parallel arrays, dataframes, and lists that extend common interfaces like NumPy and Pandas to larger-than-memory or distributed environments.

- `dask.delayed` can be used to parallelize custom algorithms by creating computational graphs.

```
# regular code
x = func1(<args>)
y = func2(<args>)
z = func3(x, y)
```

```
# with dask
x = dask.delayed(func1)(<args>)
y = dask.delayed(func2)(<args>)
z = dask.delayed(func3)(x, y)
#       x          y
#        \        /
#       func1   func2
#         \     /
#          func3
#            |
#            z
z.compute(scheduler='threads')
```

- `dask.delayed` can be used to parallelize custom algorithms by creating computational graphs.

```
list_delayed = [dask.delayed(func1)(<args>),
                dask.delayed(func2)(<args>),
                dask.delayed(func3)(<args>)]

dask.compute(*list_delayed, scheduler='threads')
```

- `dask.array` implements a subset of the NumPy array interface using blocked algorithms, cutting up the large array into chunks of small arrays.
- `dask.bag` parallelizes computations across a large collection of generic Python objects.
- `dask.dataframe` is a large parallel DataFrame composed of many smaller Pandas DataFrames which may live on disk for larger-than-memory computing on a single machine or a cluster.

[lab] Simple dask graphs

- Let's run the notebook `dask/intro.ipynb`.
  - The goal is the go over the cells and questions and annotate the code with `dask.delayed` to make the execution lazy.
  - Before running predict how much time it will take.
  - The processor on Piz Daint's 'gpu' nodes have 24 threads. Try a number of tasks higher and lower than 24 a see what happens.

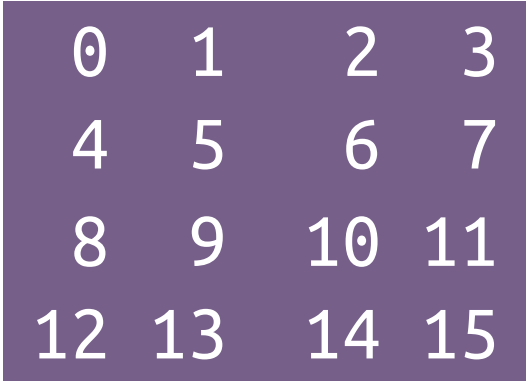
[lab] Cityblock distance matrix with SciPy and dask

- Let's run the notebook `dask/cityblock-distance-matrix-scipy.exercise.ipynb`.
- `scipy.spatial.distance.cdist` can be used to compute the cityblock distance matrix. It is fast but doesn't use OpenMP threads. We can easily write a distributed Cityblock distance matrix function based on `cdist` with the help of dask.
  - Same as the previous exercise, go over the cells and annotate the code to execute it lazily. This time you have to go over the notebook and find what needs to be changed.
  - While timing `cdist` check with `top` that it runs on a single thread.
  - Why is it relevant for the implementation of such distributed function that `cdist` runs on a single thread?
  - Check that when we create the list of delayed functions the execution is deferred to when `compute` is called.
- Run all cells and compare the execution times of the different approaches.
- While running the `%timeit` function calls, you may open a terminal and check with `top` that the new distributed function is runnig in multiple threads.

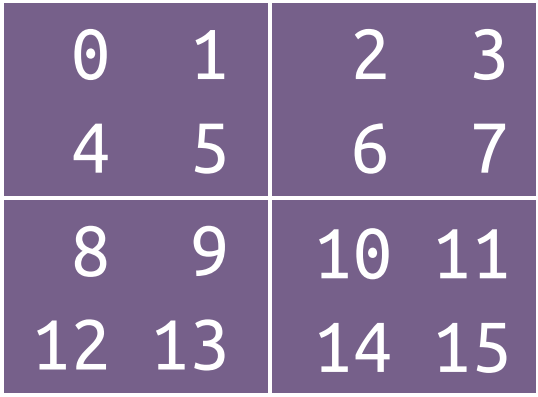
[lab] `dask.delayed` threads vs processes

- Let's run the notebook `dask/processes-vs-threads.ipynb`.
  - Run the notebook and reply the questions.

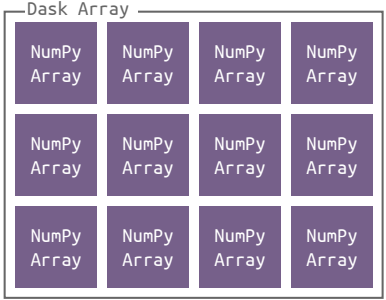
dask array



dask array

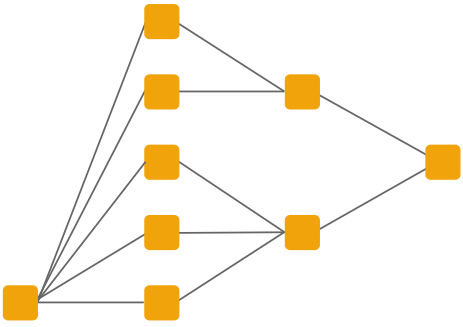


ETH zürich
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre



- A dask array consists of many NumPy arrays arranged into a grid
- Those NumPy arrays may live on memory, disk or remote machines
- `dask.array` implements many of the numpy functions but in block-wise fashion and are executed through a graph.
- For equal sizes, operations on dask arrays are in general slower than the corresponding NumPy ones.

`dask.array` graph

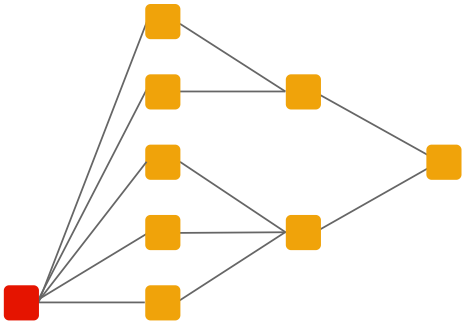


waiting

computing/allocated

memory freed

`dask.array` graph

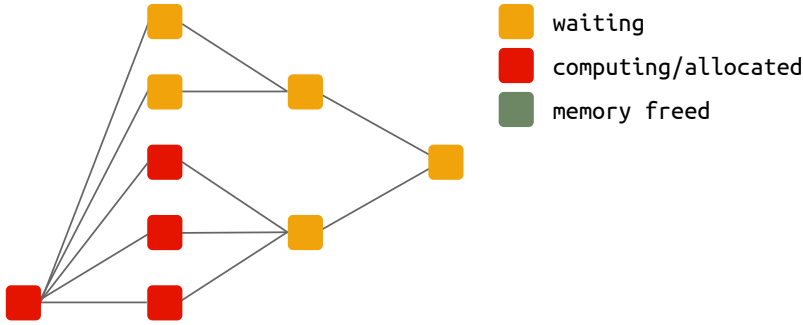


waiting

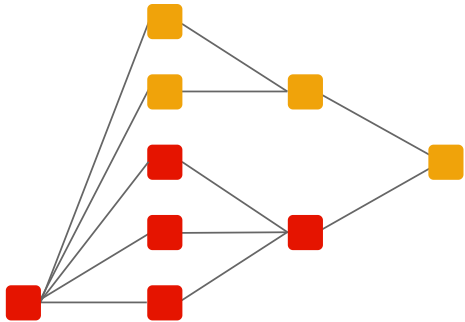
computing/allocated

memory freed

`dask.array` graph



`dask.array` graph

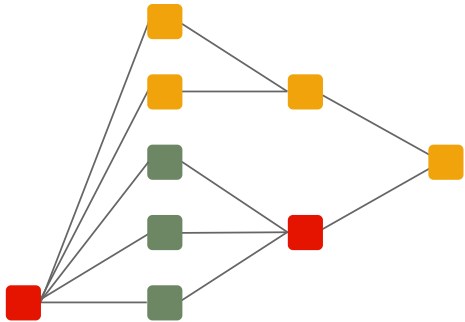


waiting

computing/allocated

memory freed

`dask.array` graph

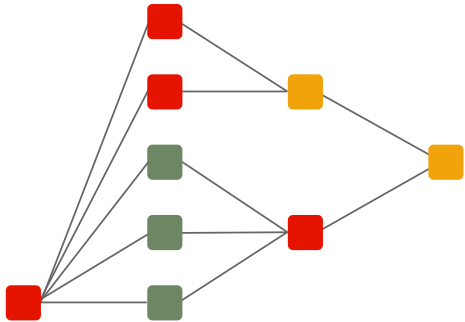


waiting

computing/allocated

memory freed

`dask.array` graph

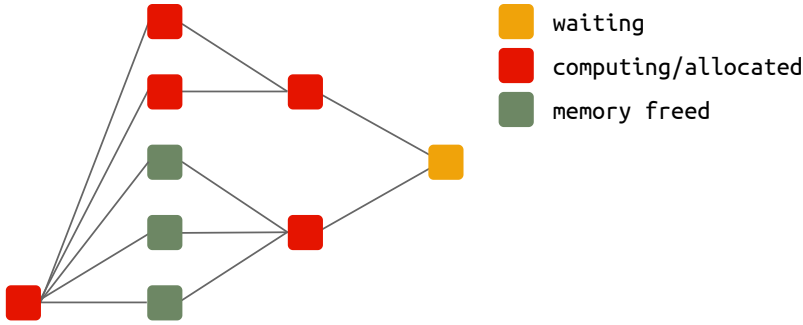


waiting

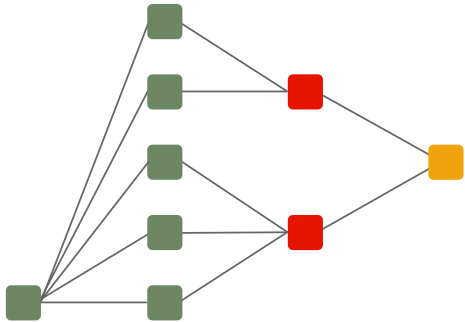
computing/allocated

memory freed

`dask.array` graph



`dask.array` graph

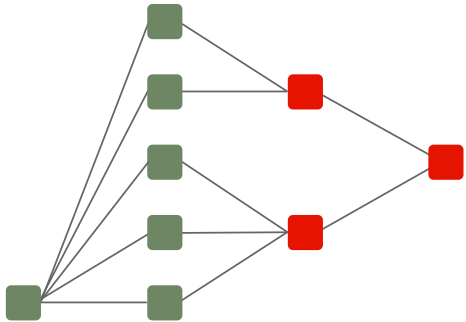


waiting

computing/allocated

memory freed

`dask.array` graph

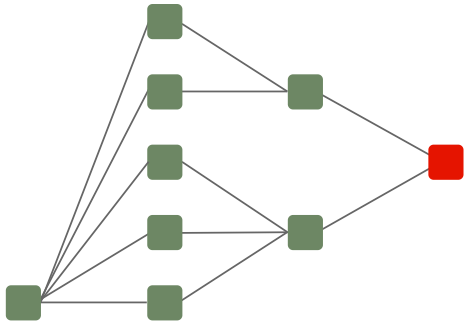


waiting

computing/allocated

memory freed

`dask.array` graph

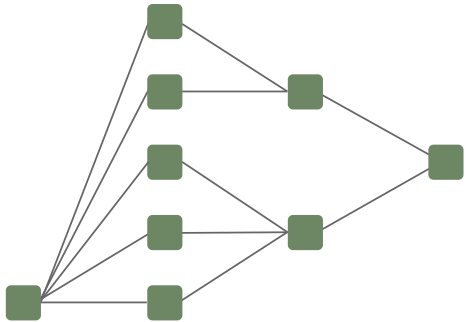


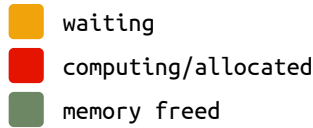
waiting

computing/allocated

memory freed

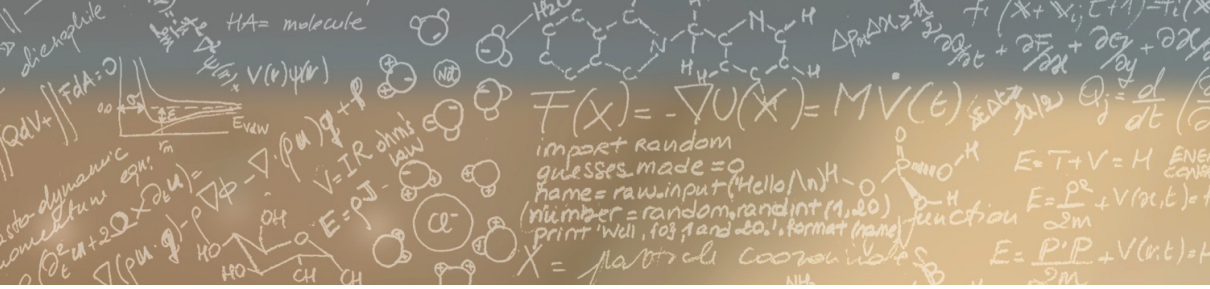
`dask.array` graph





[lab] dask arrays

- Let's run together the notebooks `dask/dask-arrays.ipynb` and `04-dask-array-from-file.ipynb`



Thank you for your attention!