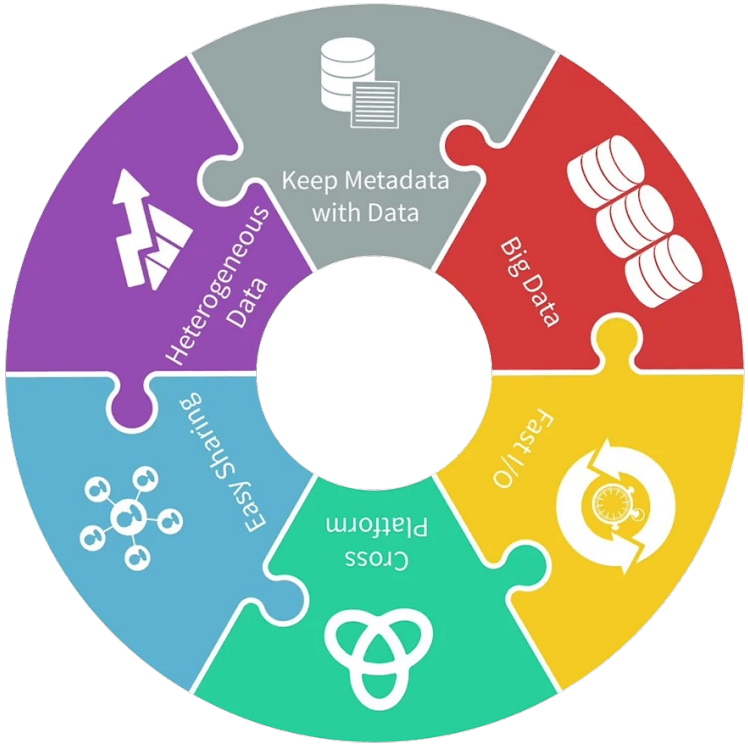
HDF5

Data

- Codes produce data. Need to store
  - Metadata:
    What is being stored
  - Raw data:
    Actual numbers
- Lattice-Group Change in code base around 2015
- Find new data format that:
  - Can be used in C/C++
  - Is readily availabe everywhere
  - Easy to use
  - Selfdescribed (Metadata together with Data)
- HDF5

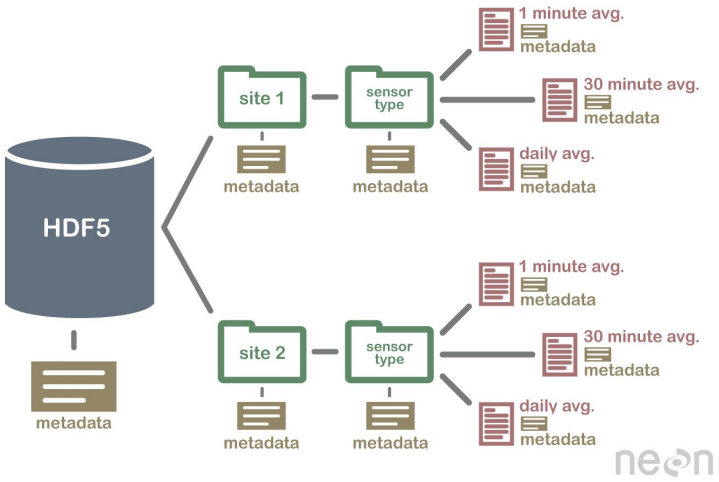
HDF5

- Started by National Center for Supercomputing Applications (NCSA)
- Website
- For me:
  - Easy to use
  - On all platforms I work with available
  - Keep Metadata with the real data
  - Somebody else takes good care of it



Layout

- Two main sets:
  - Group:
    Contains a number of groups or datasets
    Like a folder
  - Dataset:
    Multidimensional Array of fixed type




```
[djukanov@login23 dat]$ h5ls baryon_2pt_H105r001n80
axes                     Group
baryon_2pt               Dataset {4, 2, 4, 17, 179, 96}
infile                   Dataset {SCALAR}
observer_version         Dataset {SCALAR}
qdp_version              Dataset {SCALAR}
qmp_version              Dataset {SCALAR}
```

Layout

- Command-line tools:
  h5ls, h5diff …

```
[djukanov@login23 dat]$ h5ls baryon_2pt_H105r001n80/axes
channel                  Dataset {4}
momentum                 Dataset {179, 3}
polarizations            Dataset {17}
propagator_list          Dataset {2, 2}
source_position          Dataset {4, 4}
```

- Bindings for most analysis langs
  Python, R, matlab, Mathematica
- I use the Python package h5py regularly

C++ Interface

- Use in our code

```
void hdf5_writer::init(const char *filename, const bool mpi)
{
  use_mpi = mpi;
  complex_t = complex_type();
  if (use_mpi || QDP::Layout::nodeNumber() == 0)
  {
    hid_t fapl = H5P_DEFAULT;
    if (use_mpi)
    {
      fapl = H5Pcreate(H5P_FILE_ACCESS);
      H5Pset_fapl_mpio(fapl, MPI_COMM_WORLD, MPI_INFO_NULL);
    }
    file = H5Fcreate(filename, H5F_ACC_EXCL, H5P_DEFAULT, fapl);
    if (file < 0)
    {
      error_root(1, 1, "[hdf5_output.cc]", "File %s could not be created.\nFile may already exist!", filename);
    }
    axes_group = H5Gcreate(file, "/axes", H5P_DEFAULT, H5P_DEFAULT, H5P_DEFAULT);

    if (!use_mpi)  //RM after test with MPI
    {  // testing for now //RM
      H5LTmake_dataset_string(file, "/infile", observer.infile_contents);
      H5LTmake_dataset_string(file, "/observer_version", observer_RELEASE);
      H5LTmake_dataset_string(file, "/qdp_version", qdp_RELEASE);
      H5LTmake_dataset_string(file, "/qmp_version", qmp_RELEASE);
    } //RM
```

Python Interface

- Python Package h5py

```
In [1]:
1 import h5py
2 import numpy as np

In [2]:
1 data=h5py.File('./baryon_2pt_H105r001n80','r')

In [3]:
1 data.keys()

Out[3]: [u'axes',
 u'baryon_2pt',
 u'infile',
 u'observer_version',
 u'qdp_version',
 u'qmp_version']
```

- Let‘s look at the data

```
In [4]:
1 data['baryon_2pt']

Out[4]: <HDF5 dataset "baryon_2pt": shape (4, 2, 4, 17, 179, 96), type "<c16">
```

- Multidimensional Array of complex numbers

Python Interface

- Selfdescribing data:

```
In [4]:  1  data['baryon_2pt']
Out[4]: <HDF5 dataset "baryon_2pt": shape (4, 2, 4, 17, 179, 96), type "<c16">
```

What do the axes mean?

```
In [5]:  1  [i.label for i in data['baryon_2pt'].dims]
Out[5]: [u'source position',
         u'propagator list',
         u'channel',
         u'polarizations',
         u'momentum',
         u'T']
```

```
In [6]:  1  data['axes']['source_position'][:]
Out[6]: array([[ 0,  0,  0, 40],
               [16, 16,  0, 40],
               [ 0, 16, 16, 40],
               [16,  0, 16, 40]], dtype=int32)
```

Slicing the n-dim Array

- Perform mean over source pos, for specific momenta

In [16]:
```
mask=np.array(map(lambda x: x[0]**2+x[1]**2+x[2]**2==1,
    data['axes']['momentum'][:]))
```

In [19]:
```
data['baryon_2pt'][:,0,0,0,mask,0:10].shape
```

Out[19]:
```
(4, 6, 10)
```

In [20]:
```
np.mean(np.mean(data['baryon_2pt'][:,0,0,0,mask,0:10],axis=0),axis=0)
```

Out[20]:
```
array([0.00000000e+00+0.00000000e+00j, 3.72372107e-15+5.72185928e-15j,
       1.76389472e-14-3.24649408e-15j, 7.29352979e-14-1.12613155e-14j,
       3.15909270e-13+3.71719799e-14j, 7.76175048e-13-1.89283775e-14j,
       1.30915252e-12+1.95093299e-13j, 1.75146680e-12+2.73272223e-13j,
       3.13868875e-12-2.23982839e-13j, 2.85284823e-12-8.13861207e-13j])
```

HELMHOLTZ
Helmholtz-Institut Mainz