Executing code on columnar data: the translation problem and formats that help

Jim Pivarski

Princeton – DIANA

February 8, 2017

Scope of computability

Three types of data transformations:

Flat: apply N-argument function to each element of N aligned arrays.

Known in the Numpy community as a "ufunc."

Explode: emulate (nested) for loops by replicating data in one array so that it becomes aligned with another array.

Reduce: emulate reducer functions (sum, mean, max...) by combining elements of an array so that it becomes aligned with an outer level of structure.

Scope of computability

diana hep

Three types of data transformations:

Flat: apply N-argument function to each element of N aligned arrays.

Known in the Numpy community as a "ufunc."

Explode: emulate (nested) for loops by replicating data in one array so that it becomes aligned with another array.

Reduce: emulate reducer functions (sum, mean, max...) by combining elements of an array so that it becomes aligned with an outer level of structure.

What's missing?

"Repeat until convergence." Whatever determines "convergence" may be different for each element of the array: they'd all wait for the last one, anyway.

Flat transformations

Any C function can be wrapped up as a ufunc.

```
import ctypes, numpy, numba

libMathCore = ctypes.cdll.LoadLibrary("libMathCore.so")
chi2_ctypes = libMathCore._ZN5TMath17ChisquareQuantileEdd # c++filt!
chi2_ctypes.argtypes = (ctypes.c_double, ctypes.c_double)
chi2_ctypes.restype = ctypes.c_double

# compile to pure-C ufunc
@numba.vectorize(["f8(f8, f8)"], nopython=True)
def chi2_ufunc(p, ndf):
    return chi2_ctypes(p, ndf)

p = numpy.random.uniform(0, 1, int(1e6)) # million random numbers
result = chi2_ufunc(p, 100)              # call ufunc on all of them
# 3.22 seconds

import ROOT
result = [ROOT.TMath.ChisquareQuantile(pi, 100) for pi in p]
# 9.32 seconds
```

(Performance comparison is just to show that the ufunc computes `ChisquareQuantile` in C, not in Python. Simpler functions show a more dramatic difference.)