

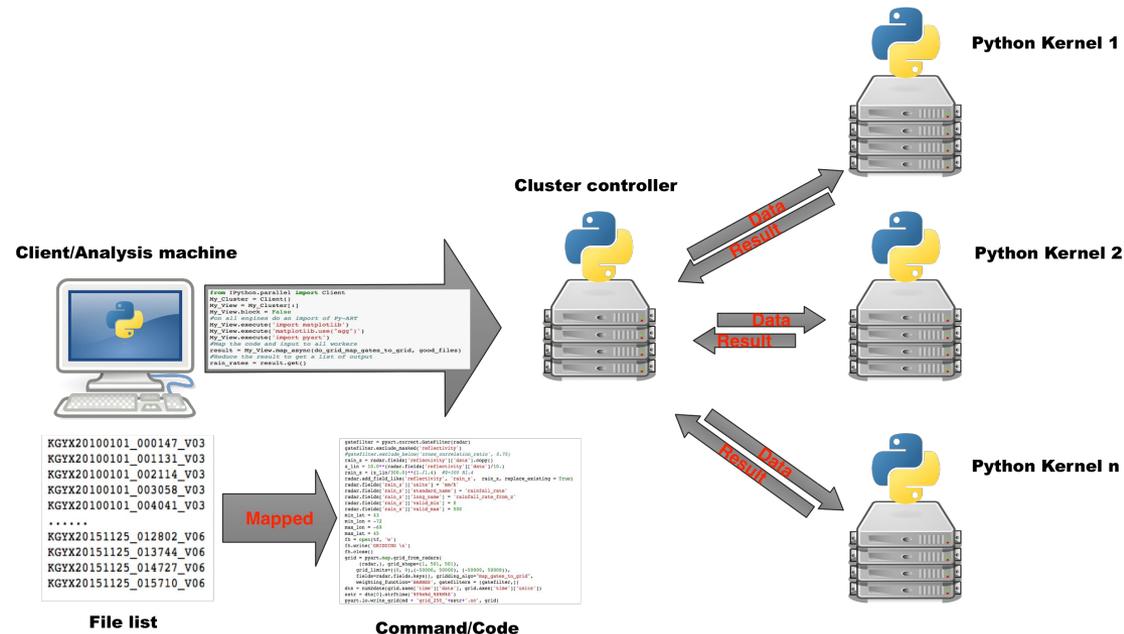
Compute requirements for Radar VAPS

ARM

CLIMATE RESEARCH FACILITY

Current methodology of dealing with large data sets

- Radar networks (and even single radar sites) produce “moderately large data”. Pales in comparison to cat videos.
- In general most radar experts are not big data experts, even those within this program consider themselves dabblers at best.
- Fortunately part of Project Jupyter has made it easier for us to scale our algorithms assuming independence from scan to scan. This is the IPython Cluster.
- IPCluster works by distributing worker kernels to machine cores and then mapping commands and reducing on input parameters (eg file list).



Requirements for radar processing

- We run out of IO bandwidth first, memory second.
- Simple process and grid takes ~4GB memory, complex tasks (meshing networks together can take more).
- This is on a “moderately optimized” codebase. We do not waste memory but everything is done in floats, no packed integer math.
- In a recent experiment on a GPFS system we used 1024 cores to process 13 years of NEXRAD data over the SGP in 10 wall hours. This was a simple Z-R relationship and meshing to a 500x500 grid.
- **Requirements: Submission system (PBS), 4+ GB per core (v. happy if 8). 300+ core machine (Say 18 nodes @16GB/node). V. Good FS, GPFS for example. Seamless ordering and transport from HPSS.**

