As the high performance computing (HPC) community pushes towards the exascale horizon, the importance and prevalence of fine-grained parallelism in new computer architectures is increasing. This is perhaps most apparent in the proliferation of so-called “accelerators” such as the Intel Xeon Phi or GPGPUs, but the trend also holds for CPUs, where serial performance has grown slowly and effective use of hardware threads and vector units are becoming increasingly important to realizing high performance. This has significant implications for weather, climate, and Earth system modeling codes, many of which display impressive scalability across MPI ranks but take relatively little advantage of threading and vector processing. In addition to increasing parallelism, next generation codes will also need to address increasingly deep hierarchies for data movement: multiple types of user-addressable memory, many NUMA domains and caches, on node vs. off node, local vs. wide neighborhoods on the interconnect, and even in the I/O system. We will discuss these hardware architectural trends and some approaches (grounded in experiences with the Intel Xeon and Xeon Phi architectures) for restructuring Earth science codes to maximize concurrency across multiple levels (vectors, threads, MPI ranks), and also discuss some novel approaches for minimizing expensive data movement/communication.