

National Aeronautics and
Space Administration



Exploring Options for a Bespoke Supercomputer Targeted for Weather and Climate Workloads

*Bob Sorensen, Hyperion Research
Conducted for
Tsengdar Lee, NASA*

November 2019

Background and Focus of Study

NASA's earth system models have unique high performance computing (HPC) requirements, which can differ greatly from standard industry offerings


- Moreover, the gap between vendor HPC solutions and earth system models has been growing, such that these models can exploit less and less of the peak computing capability of current HPC systems

The primary focus of this study was to gather key insights, through a series of surveys with weather and climate users and potential HPC suppliers, on options available to NASA, and others, to develop a bespoke HPC system specifically targeted for weather/climate research

Study Structure

The study was divided into two major

- The first phase centered on a series of interviews held with expert researchers and users in the HPC-based climate and weather community to collect thoughts and insights on current and future operation requirements as well as the specific HPC hardware, software, and architectures needed to meet those workloads
 - For phase one, 15 different weather/climate organizations in the US and overseas were surveyed including ECMWF, LANL, NOAA, ORNL, UCAR, and the University of Delaware
- Phase two consisted of taking the results of phase one to generate a second survey of HPC suppliers and independent HPC designers to assess the challenges and opportunities of developing a bespoke HPC to meet the phase one requirements
 - HPC suppliers providing input included Cray Inc, Dell EMC, HPE, and IBM



Phase One Study Highlights

Weather/Climate User Current and Future HPC Concerns and Insights

Limitations with Current and Planned HPCs

Researchers and users in the HPC-based climate and weather community had a broad list of limitations with current HPCs and with commercial HPC offerings in the next few years

- Key concerns centered on
 - Drawbacks in memory and storage latency and bandwidth
 - Inability to use full processor capabilities
 - A lack of diversity in processor option/designs
 - The current mainstream reliance on GPUs that are not well suited to current weather/climate community workloads
 - The trend towards vendor specific interconnect options at the highest levels of computing
- Other concerns specifically mentioned:
 - Inappropriate node/CPU designs that lead to low (near 1%) efficiency on current atmosphere and ocean models
 - Overall high, and increasing, system capital and operating costs

Leveraging AI into Weather/Climate Research

Another major response theme centered on missing opportunities for bringing AI, particularly machine learning capabilities, into the programming mix, *but that opinion was not universally supported*

- Key prospects were seen for mixing machine learning (ML) with traditional computation techniques running concurrently (asynchronously) with the simulation
 - However, such usage would require a shared memory mechanism for communication between the ML and traditional computation instances
- Researchers expect to see growth in ML and increasing requirements for supporting ML software frameworks, such as TensorFlow and Keras
- GPUs might be (although are not necessarily) a significant platform for ML applications in atmospheric research

Many Future Architectural Considerations

Survey respondents had a long, ambitious, and somewhat diverse list of architectural considerations for any future bespoke HPC

- Heterogeneous systems should consist of mostly CPU only nodes with a small fraction of CPU-GPU nodes,
 - Otherwise the interconnect can become a bottleneck
- A key gating performance factor will be memory bandwidth per core, a measure that is at best flat, or in some cases declining, for DRAM-based CPU systems
- Processors will likely have hundreds to thousands of compute cores that will require more fine-grain parallelism in the applications
 - This will drive requirements for more memory bandwidth per processor, lower latency communications, parallel I/O with high bandwidth storage and file systems
- Application scalability is another key gating factor for performance improvements
 - To support scalability, the interconnect needs to be balanced relative to on node memory bandwidth and the MPI tasks must be mapped to minimize off node communications

Future Software Considerations

Respondents offered a wide list of software considerations as well. Two fundamental themes:

- A wide range of existing applications need to be rewritten to take advantage of new hardware and software trends
- There needs to be a strong base of software development tools to support those activities
 - Many expect that this process will be complex and met with some resistance
- Key comments:
 - In general, existing applications are poorly architected and largely designed for traditional, low-core count CPU processors
 - Need reliable performance analysis and debugging toolsets, assuming that existing code can be ported

Phase Two Study Highlights

HPC Vendor Inputs: Concerns and Suggestions

High Interest in Exploring a Flexible Bespoke HPC Design...Within Limits

All HPC suppliers surveyed indicated an interest in exploring the development of *some form* of a bespoke HPC for NASA

- However, the level of commitment and the degree to which a system would use special purpose or customized hardware and software varied greatly
- Most indicated that the option for building a special purpose one-off system or even a series of such a system over time could not be economically justified regardless of the amount of NRE
 - Unless that particular design had value within the wider weather community at a minimum or across a number of complementary verticals
- Some respondents indicated that no amount of NRE would justify the development of a one-off system as the opportunity costs would always be too high

HPC Design Flexibility Will be Required

A key point made repeatedly was that flexibility in HPC design through either a reliance on open systems components or carefully chosen proprietary designs could help significantly

- Hardware flexibility enables the right choice of memory fabric, so that any processing element can be used and scaled as long as it can speak to the fabric
 - Gen-Z was cited as a promising open system bus that can support a wide range of memory access modalities
- Processor selection was viewed as a topic open for discussion
 - Intel and AMD X86, ARM, and related ARM vector processors such as those available from Fujitsu were all named as potential processor choices
 - ARM presents some attractive options for a custom design that would allow for targeted core count to memory bandwidth ratios
- Regardless of processor choice, a careful selection of processor SKU would enable the tuning of memory and interconnect bandwidth per core to maximize these when customer workloads or benchmarks show that these are performance limiters
 - The careful selection of processor SKUs to provide needed memory bandwidth would enable NASA to exploit new, rapidly emerging storage technologies including high bandwidth memory (HBM), on package memory, and nonvolatile memory as they become more available

Gleaning NASA Requirements Are Problematic

HPC providers indicated that determining exactly what such a system would look like could be challenging

- NASA has a wide and varied range of workloads that would need to be accurately characterized and combined into a comprehensive set of hardware and software requirements
- Some respondents expressed concern that there would be 'too many cooks in the kitchen' to develop a well-defined set of vendor requirements for this community
- For any system under consideration, vendors indicated they would be chasing a moving target, and vendors would want a better understanding of what the end goals were and how best they can meet those requirements

Widespread Concerns About Existing Code Base -- Modernization is Key

By far, the major concern of respondents interested in working with NASA centered on NASA's need to modernize their codes to better capture improvements in both high end hardware and software.

- Almost all respondents cited the critical need for code modernization and refactorization, while acknowledging that such a task will not be easy
- There will only be escalating costs to support an increasingly obsolete programming paradigm, and the first step to addressing that reality is to 'bite the bullet' and embrace a comprehensive software rewrite plan
- Cloud service providers were cited as leading the way in driving continual code refactoring as they provide new software releases on a regular basis, allowing for the quick harnessing of any new technology gains

Improved Benchmarking Efforts Are Vital

On performance metrics, HPC supplier respondents indicated that benchmarks like LINPACK and HPCG were not sufficient to guide them in an overall system design

- Many indicated that benchmarks using mini application suites, testcases, or even full applications would be the best way to determine the various strengths and opportunities of any HPC design under consideration
- Larger questions to be addressed included: what are your current problems, what are you trying to accomplish, what steps have you taken to profile your ecosystem for bottlenecks, etc.?
- Respondents stressed the need for better analysis of existing NASA workflows and key applications to trace the use and movement of data to design systems that maximized the value of data while minimizing its movement

Upfront Codesign Efforts are Essential

All respondents strongly indicated that the first and perhaps most important step in the development of a bespoke HPC should center on pre-RFI vendor and NASA codesign team activities

- Host a design conference open to all vendors that fosters frank dialogues about existing and planned codes, hardware and software requirements for those codes, and any other special considerations unique to the NASA workloads
- Target codesign activities not on current technology or products, but instead consider what will be available three years or more
- Efforts should be made to engage the wider weather/climate community in any such codesign activities as well as to extend outreach to those outside the sector who are also looking to acquire systems that are not 'straight out of the catalog'



Recommendations for Next Steps

Improve Benchmarks

NASA has a wide and varied range of workloads that would need to be accurately characterized and combined into a comprehensive set of hardware and software requirements

- Key considerations involve an analysis of
 - Processor core counts and memory bandwidth
 - Special memory frameworks such as HBM and persistent memory
 - GPUs and other accelerator use
- The results of a comprehensive requirements analysis can then be used to compose benchmarks that consist of mini application suites, testcases, or even full applications to help determine the various strengths and opportunities of any HPC design under consideration

Embrace New Workloads Composition

NASA should move to extend the composition of their HPC workloads beyond traditional modeling and simulation to increase performance on existing applications and to expand capabilities in new ones. They include:

- AI based programming not only in machine learning applications, but also to accelerate traditional modern and simulation jobs
- Big data jobs, both batch and near real time that could be a mix of large storage volumes along with high bandwidth data inputs from large arrays of edge-connected sensors
- Hybrid on premise/cloud computing platforms to help absorb large fluctuations in workload, to test out and perhaps take advantage of new hardware opportunities quickly, and to tap into the growing base of CSP resident HPC software

Conduct Codesign Conferences

NASA could organize a pre-competitive codesign conference or continuing series of conferences that bring together interested commercial vendors to discuss options and opportunities based on the analysis described above

- A key goal of these co-design conferences would be to build long-term relationships with key vendors to generate a strategic multi-generation procurement plan for HPCs instead of targeting a one-time procurement
- This would be an excellent opportunity to reach out to USG organizations that are also in the process of refactoring/modernizing their code as well as facing similar HPC hardware issues
- Likewise, similar collaborations could be established with both domestic and foreign weather/climate HPC centers

Codesign Agenda Items:

- Generate relevant projections of advances in future HPC hardware and software from a wide base of HPC vendors and users that will better align with NASA requirements but that are in keeping with larger commercial trends
- Work with vendors, many that have significant expertise in profiling and tuning applications to systems using future technologies, in helping to develop a robust methodology for making performance predictions from NASA benchmarks
- Explore opportunities to leverage any NASA hardware or software developments into the larger weather/climate HPC ecosystem or even other HPC verticals to help vendors justify the use of non-COTS hardware or software
- Enlist vendor insights or other USG organizations on the best ways for NASA to begin refactoring and modernizing their current software base to best take advantage of existing and planned commercial technology and HPC products

Thank You

Questions?

bsorensen@hyperionres.com

