IDC
Analyze the Future

I D C T E C H N O L O G Y S P O T L I G H T

The Open HPC Stack Initiative Hits a Milestone

June 2016

Adapted from *The Broader HPC Market: Servers, Storage, Software, Middleware, and Services*, Steve Conway, Earl C. Joseph, Ph.D., Robert Sorensen, 2014, Doc # 255224

Sponsored by Intel

IDC has long maintained that high performance computing (HPC) leadership will be determined more by software advances than hardware progress. The hardware direction is fairly well set, while the HPC software stack—the operating system and other components between the application kernels and the hardware—faces an array of challenges that must be overcome to turn future HPC systems of all sizes into coherent, efficient, productive resources. These challenges are daunting enough that HPC vendors have turned to the open source community to help address them. These partnerships typically result in two versions of the software stack: a free open version and a paid version for users who need added capabilities and vendor support. One major initiative of this kind is OpenHPC, a close collaboration between Intel, acting as catalyst, and a growing number of OEMs and other members of the global HPC community, including open source software developers, ISVs, and end users of HPC systems. Intel has announced the first product of this collaboration: Intel HPC Orchestrator, an integrated software stack aimed at the majority of Top500-class HPC sites. This paper examines the need for HPC open source software stack initiatives and looks at Intel HPC Orchestrator as an examplar of this important trend.

A New Development Model for the HPC Software Stack

A recent IDC study reaffirmed the findings of our prior studies on this topic, namely, that the software stack is the biggest "pain point" for more HPC users than any other technical category (see Table 1).

Which aspect of system complexity is the MOST serious "pain point" for your organization? (Choose ONLY ONE)	
	Percent
System software (OS/middleware) – updating, using,	28.9%
System hardware – heterogeneity (multiple types of processors/components)	20.0%
I/O	13.3%
System hardware – reliability/resiliency	6.7%
Storage – capacity	6.7%
Storage – capability response time (access density)	6.7%
Storage - reliability/resiliency	6.7%
Memory management	2.2%
Other	0.0%
Not sure/don't know	8.9%
	100.1%
Source: IDC, 2016	

Especially in the past decade, the HPC software stack has faced a host of new challenges:

- Burgeoning system sizes
- Heterogeneity: multicores and coprocessors/accelerators
- New environments (e.g., cloud computing)
- Reliability/resiliency at extreme scale
- Energy efficiency/energy costs
- Mix of simulation (floating point) and analytics (integer) workloads
- Mix of synchronous and asynchronous workflows
- SMEs and other commercial users who want "ease-of-everything"
- Commercial analytics users who are new to HPC

More Pain Is On the Way

IDC studies indicate that demands will accelerate as the aforementioned challenges intensify and are joined by new requirements, such as the need to relieve programmers and system administrators by incorporating more automation/intelligence into the stack—especially to increase architecture awareness, energy awareness (a.k.a. "power steering"), and system wellness management.

This escalating rate of change calls into question the existing model for developing the stack. Today, various system OEMs and other vendors employ their own formulas to turn software components into HPC stacks. The results of these initiatives are impressive, but they have also involved considerable duplication of effort that may no longer be tolerable for keeping pace with escalating requirements.

It is increasingly difficult for HPC sites and OEMs to maintain stable, performant versions of all the necessary stack components, given the myriad compatibility and interdependency requirements. It's become harder for them to keep pace with rapid release cycles and version control. On the developer side, many projects must engage in continual triage and debugging when configuring and installing HPC systems.

In short, it seems increasingly unreasonable to expect any single organization to keep pace with the needed rate of change for the HPC software stack, and it seems increasingly counter-productive for so many organizations to spend so much time and money duplicating efforts to create and advance end-to-end stacks.

Benefits of the Open Source Model

More workable going forward, in IDC's opinion, is the development model exemplified by Linux. In this model, stack development is driven primarily by the open source community and vendors offer supported distributions with additional capabilities for customers that require and are willing to pay for them. As the Linux initiative demonstrates, a community-based, vendor-catalyzed model like this has major advantages for enabling software to keep pace with requirements for HPC computing and storage hardware systems.

The Linux model has demonstrated the compelling benefits of the collaborative, "open software ecosystem" approach, for HPC vendors and users alike.

- First and foremost, as noted earlier, this model helps to ensure that the software keeps pace with burgeoning user requirements—thereby making the HPC system a more productive, higher-returning investment.
- This model hypothetically delivers more new capabilities to users, and delivers them sooner.
- Programmers and other developers also gain faster access to new capabilities, including features that can relieve them of explicit, labor-intensive work.
- Vendors can select the best open source components for baseline functions while focusing more of their R&D budgets and efforts on development targeted at providing competitive differentiation. In other words, open source software can act as an offset to vendors' R&D expenses.
- Last but not least, the open versions seeds the community with potential new users of the paid versions who may want additional support or enhanced stack stability for production computing environments. A portion of users who become accustomed to employing the free version will later on want the paid version. (Think, for example, of students who are later employed by HPC sites needing vendor-supported software.)

Desired Attributes of HPC Open Software Stacks

Ideally, an open source-based stack model should aim to benefit the whole HPC community, from the largest, most experienced sites to SMBs that buy only one or two racks of HPC servers.

An effective open source model should include a growing set of reference architectures and standard software modules that make it easier for vendors and users to build custom stacks that work well in practice.

- A fair number of foundational open source HPC software components already exist (e.g., Open MPI, OpenSFS, CentOS, Rocks Cluster Distribution, OpenFOAM, OpenStack, and others). Many HPC community members are already taking advantage of these.

The open source model should also allow vendors and users to insert their own favorite software modules, as long as they obey the conventions of the open source reference architecture. The open source model should leave plenty of room for OEMs and dedicated HPC software stack vendors to leverage their existing investments in system software and to continue to innovate and thrive.

Introducing Intel HPC Orchestrator

As part of the company's Intel Scalable System Framework, an integrated performance platform for HPC that also includes Intel Xeon processors, Intel Xeon Phi coprocessors, Intel Omni-Path interconnect fabric, memory, and other elements. Intel has also been advancing its multi-generational plan to offer commercial versions of the OpenHPC stack. Intel plans to extend HPC Orchestrator implementations to cloud environments. The company has substantial software experience and has been aggressively hiring top software talent for several years and has greatly strengthened its internal expertise.

In June 2016, Intel announced the first of the company's multiple planned commercial versions of the OpenHPC stack, Intel HPC Orchestrator. This initial offering is aimed at the requirements of supercomputers occupying positions 51-500 on the Top500 list. This will be Intel's core solution, the basis for future solutions to address the other parts of the global HPC community.

Intel HPC Orchestrator Follows the Open Source Model

- The Intel HPC Orchestrator software stack is modular and relies on a foundation of proven open source software components Intel has selected, supplemented by new software contributions from the community, Intel, and from collaborating OEMs and other vendors.
- The open source community drives and evolves the foundational layer of the Intel HPC Orchestrator stack. Intel contributes to the open source software and maintains paid, supported distributions of the stack (Intel HPC Orchestrator) that optimize performance on Intel processors, coprocessors, and interconnect fabrics.
- Intel ensures that components of Intel HPC Orchestrator interoperate with strong performance and efficiency across a broad matrix of hardware configurations.

Targeted Benefits of the Intel HPC Orchestrator Product Family

- Intel HPC Orchestrator is the Intel-supported version of the OpenHPC software stack, aimed at the requirements of supercomputers occupying positions 51-500 on the Top500 list. Intel HPC Orchestrator is fully compatible with the open source OpenHPC stack and adds comprehensive Intel support and additional features and functions, especially for operating HPC systems in production computing environments.
- OEMs, resellers, and ISVs can substitute or add their own components in the standard offerings, so that they can continue to differentiate and leverage their investments in these value-added components. The OpenHPC community will create the APIs. OEMs, resellers, and ISVs can act like VARs, adding their own value and selling the commercial versions to end-users.
- OEMs, resellers and ISVs can offload non-differentiating stack R&D to the open community – let the open community maintain the core part of the stack.
- ISVs can develop their software against a common, validated stack. The open stack will absorb new hardware features ISVs can leverage them.
- Intel will maintain the stack, distributing updates and patches so OEMs, resellers, and ISVs don't have to.
- The common core stack will be the same across all participating vendors, creating the basis for greater portability and protection of hardware and software investments over time.

The OpenHPC Initiative Has Strong Momentum

Nearly all major OEMs are participants, as are many key HPC ISVs. Dozens of the world's top HPC user sites have also joined this collaboration as participants. (Figure 1).

Source: Intel

Continuous Integration Environment

In addition, Intel is donating hardware to the Texas Advanced Computing Center (TACC) to build a continuous integration environment for developing OpenHPC open source versions. This build-and-test environment integrates new elements into the foundational layer of OpenHPC and uses the OpenHPC tests suite to ensure that new elements are compliant and performant.

Opportunities

- **Leverage open source software innovation on behalf of HPC users.** Intel and its OpenHPC collaborators can select the best software developments from the vibrant open source community and incorporate them into free and paid variants of OpenHPC, including Intel HPC Orchestrator and any future Intel offerings in this category.
- **Further exploit IDC-predicted HPC market growth by helping to make HPC systems easier to use and more productive.** The collaborative development model underpinning Intel HPC Orchestrator can help HPC systems—most of which are based today on Intel Xeon x86 processors—to be more productive and therefore more attractive to buyers. Ease-of-use advances will help tap the deep reservoirs of SMBs and commercial analytics users who could benefit from HPC but have not moved up to it yet.
- **Address system software requirements that will later benefit the enterprise server market.** Some of the most forward-thinking large commercial firms are already adopting HPC systems for advanced analytics challenges that their enterprise servers can't handle well alone. By

addressing these requirements in the software stack, Intel can pave the way for larger numbers of commercial companies to adopt HPC. IDC forecasts that the market for commercial firms using HPC resources—we call this the high performance data analysis (HPDA) market—will grow at about three times the rate of the overall HPC market through 2020.

Challenges

- **Motivate the open source community to help drive HPC stack innovation.** Intel will need to catalyze the open source community without being overly intrusive. It turns out that the company has worked closely with the open source community for years and seems to have mastered this balancing act.
- **Collaborate with OEMs who differentiate in part based on their own HPC software stacks.** A major challenge for Intel's OpenHPC-based initiative is to gain acceptance from HPC system vendors (OEMs) and ISVs who have invested considerable time and money in developing and marketing their own system software offerings. Intel consciously designed Intel HPC Orchestrator to allow vendors to maintain strong competitive differentiation by inserting their own software components into the stack. The fact that many major vendors are collaborating with Intel in the OpenHPC initiative speaks for itself (see Figure 1).
- **Maintain a release cadence that meets the needs of the ecosystem.** An open HPC stack, once created, needs to evolve in concert with the rest of the HPC ecosystem. Intel, working with its partners in the open source and HPC vendor communities, will need to release stack revisions and updates on a regular, scheduled cadence that makes sense for the global HPC community. Intel seems well aware of this need.

Conclusion

Future (HPC) leadership will be determined more by software advances than hardware progress. IDC studies show that software stack challenges are the biggest source of technology pain many HPC sites feel today. It is fast becoming impractical for any single vendor to keep pace with this growing array of challenges. As a result, many vendors are now involved in software stack development initiatives that rely on the open source community for help. These initiatives follow the development model pioneered by the Linux community, which results in free open source versions of software alongside paid, vendor-supported distributions. Intel has been acting as a catalyst for the OpenHPC initiative, which now counts many OEMs, ISVs, and known-name user organizations among its participants. Intel HPC Orchestrator is the first vendor-supported product of this collaboration and is aimed at the majority of Top500-class users (positions 51-500 on the list). Intel HPC Orchestrator creates an opportunity for Intel and its OEM and ISV collaborators to accelerate the healthy growth rate of the HPC market by making these systems more productive and easier to use. In doing thiis, Intel and its partners could also speed the existing adoption of HPC by commercial firms with daunting analytics challenges. The challenges for Intel include maintaining strong relationships with the open source community and with OEMs and ISVs that need to continue differentiating from competitors, along with maintaining a release cadence that meets the needs of the global HPC community. IDC believes that Intel has established strong momentum for addressing these challenges and is well positioned to exploit IDC's predicted growth for the worldwide HPC and HPDA markets.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com