Optimizing Business Value in High Performance Engineering Computing:

How IT Centralization Can Improve Productivity, Collaboration and Innovation in the Engineering Enterprise

Srini Chari, Ph.D., MBA

October 2012

chari@cabotpartners.com

Sponsored by ANSYS, Inc.

Executive Summary

Manufacturers are increasingly using systems engineering design and analysis approaches to tackle the ever-growing complexity and sophistication of smart products, expecting to improve quality and reliability and to reduce cost and time to market. High-performance computing (HPC) is ever more critical to model this end-to-end product development and manufacturing process.

However, the performance, scale and capabilities required for today's engineering simulation workloads strain many legacy engineering HPC environments that previously were centered around an individual engineer's desktop or a departmental IT resource. Today, these engineering enterprises need highly parallel, scalable and integrated HPC infrastructures (servers, storage, and networks) that are flexible and responsive to very large and heterogeneous workloads. Engineering enterprises must develop the business case for these HPC investments through a collaborative process involving IT and engineering end users and management. They must also be aware of the subsequent shifting costs of HPC over the past decade or so.

Today, rising operational IT costs (labor, energy, and facilities) have become the Achilles heel in deploying new HPC solutions at many organizations. Centralizing HPC infrastructure has proven to reduce these escalating operational costs while improving reliability and utilization, thereby greatly improving the total cost of ownership (TCO) for HPC. Further, centralizing and co-locating data with compute resources improves engineering collaboration and productivity and better secures sensitive enterprise data and intellectual property (IP).

However, centralizing engineering HPC is a challenge since workloads and data transfer requirements are immense for both computing and visualization, especially when product development teams need to collaborate across multiple sites. A "private cloud" remote interactive-access architecture for engineering simulation, built collaboratively with strategic IT industry participants, is an attractive best practice that has been tested and found to work in real-life engineering computing environments. With the anticipated increased adoption of centralized private clouds for engineering HPC, this remote interactive-access architecture can deliver substantial business value for engineering enterprises.

Greater scale and complexity challenges stress engineering HPC…

To improve quality and reliability and to reduce cost and time-to-market, manufacturers and their supply chain are adopting *engineering design and analysis* approaches to tackle the ever-growing complexity and sophistication of multidisciplinary engineered products. Accurately predicting how complex products will behave in real-world environments requires prediction of how multiple types of coupled physics interact at system and component levels over a wide range of operating conditions.

This, in turn requires *engineering simulation* approaches that are capable of *deep* analysis with advanced technology in one domain *and broad* multidiscipline analyses in structures, fluids, electromagnetics, thermal, etc. It also requires capabilities to perform *comprehensive, high- fidelity multiphysics* simulation with larger detailed geometric models to deliver the high accuracy required. To improve engineering productivity, the end-user experience must be integrated, responsive, and persistent. It is critical to model the interaction between structural mechanics, heat transfer, fluid flow and electromagnetics *all within a single, unified engineering simulation environment.* ANSYS, the leading supplier of engineering simulation software, does this very well through its portfolio of engineering applications.

Design exploration and optimization studies are required in many practical engineering situations. In the near future, these studies will require *integrated and parametric* analyses over thousands of operating scenarios on very large meshes with billion or more cells. HPC and engineering simulation software that optimizes HPC performance are critical *to virtualize this full product development and engineering* process.

All these demands put enormous strains on IT to support engineering at a brand new scale to deliver very *high-performance* computing and data management, and to *protect intellectual property* in an increasingly global *collaborative and mobile* world while improving *IT efficiency.* Today, *operational IT costs* associated with the labor to manage IT, facilities and energy costs, etc. are *rising faster* than hardware or software costs (Fig. 1). Engineering enterprises need a scalable HPC IT environment while reducing costs and complexity.

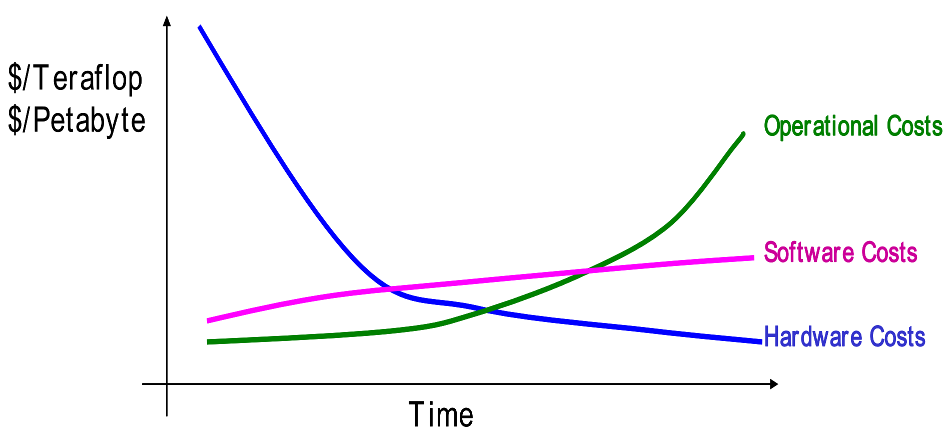


Fig. 1. The shifting costs of high- performance computing[1]

[1] Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies - A Workshop Report and Recommendations , EDUCAUSE Campus Cyberinfrastructure Working Group and Coalition for Academic Scientific Computation, February 2009.

…and IT centralization helps to address engineering HPC challenges

By *centralizing engineering HPC*, organizations benefit from the ability to *improve the utilization* of HPC resources (hardware and software) and get increased *scalability* and *flexibility* to address a range of engineering workloads from highly parallel to high-throughput workloads while *reducing escalating operational costs.* Further, centralizing data management *improves engineering collaboration and productivity* while better *securing sensitive data and IP* in a world with an increasingly mobile and global workforce.

The evolution of HPC – from local to central to private cloud

Over the last 40 years, the pendulum has swung back and forth between centralized and localized computing paradigms for high-performance computing in engineering.

Mainframes – we started from the center. During the 1970s and early 1980s, engineering computing used mainframe computers in a *centralized time-sharing model.* While this worked well for many engineering users, there was a growing demand for more flexible computing capability, and this led to the rise of the distributed computing model.

Client-server -- distributed computing. From the mid-1980s to the early 1990s, the client-server model was popular as end-users, academics, scientists and engineers began to deploy UNIX-based client-server architectures at their own location as opposed to a central location. The Moore's law-powered explosion of microprocessor capability enabled engineering computing users to purchase and *use local computing power* to solve the problems of the day with a new generation of engineering applications, including those from ANSYS. However, as the workloads and the number of degrees of freedom grew, this model started to strain.

Parallel supercomputing. From the mid-1990s onward, as parallel, centralized supercomputers became available, engineers began to solve their complex problems using finer grain grids (increasing problem size and scale) to accurately model intricate physical phenomena. This required engineers and engineering software providers to map such problems onto these parallel, centralized supercomputers. Engineering users in industry, particularly at large enterprises, benefited from *parallel versions* of engineering applications, including tools from ANSYS.

Commodity clusters – local capacity and performance but server sprawl. The continuing growth in microprocessor performance, high-speed interconnects, and innovations in open-source software enabled engineering users to cluster local resources that were largely in their control. Hardware manufacturers as well began shipping low-priced, affordable cluster systems with controlled enterprise distributions of Linux with premium end-user support and deployment services[2]. As a result, utilization levels at centralized supercomputing facilities began to fall as workloads migrated back to local computing environments in the late 1990s to early 2000s. However, over time, many organizations began to see *problems of server sprawl and increased total cost of ownership*, and very large computing problems still required centralized ultra-scale systems.

Centralization - cost concerns and scale up of demand for HPC spur a return to center. Today, the engineering community must deliver innovative designs better, faster and cheaper,

[2] Al Gillen and Dan Kusnetzky, "Enterprise Linux: It 's All About Support", http://www.netoasis.net/plan/Research2/idc_research.pdf

design high-quality products with fewer resources, across a distributed ecosystem of partners, and respond to increased cost control of engineering IT. Centralized computing environments offer a way to overcome these issues. They provide *interactive and batch* remote access, *shared and centralized* engineering IT, and an *integrated business and technical* environment. This unlocks designer skills from a location, provides greater access to compute and storage resources, and aligns resource to project priorities.

Centralization also addresses data security and emerging big data challenges. Apart from operational and cost efficiencies, IT centralization enables companies to keep *sensitive data in central and secure locations* – a key requirement in an era in which users access applications and data from offices, homes, and on the road through both desktops and multiple mobile devices.

Globally, companies are storing over seven exabytes of new data every year[3]. Managing and drawing actionable business insights from this big data requires companies to deploy centralized infrastructures with *computing close to the data* with more sophisticated analysis applications.

Private clouds - improving HPC centralization. Cloud computing is expected to be a $199 billion phenomenon by 2015, and 40% of IT solutions are expected to be delivered over the cloud in the next few years[4]. Cloud computing delivers IT as a service including infrastructure, platform, and applications. It has the potential to reduce IT capital and operational expenses with the ability to *rapidly access* and *pool IT resources* and then *dynamically allocate* them from the resource pool to improve load balancing. This could yield a much higher utilization rate and, thus, better economics — sometimes improving system utilization from *15 percent to 90 percent*.

HPC can benefit from these overall IT trends to move to the cloud and manage big data. However, the HPC cloud will be a unique environment, and the HPC cloud for engineering will have unique requirements. For many organizations, *private clouds will be the best match* for the needs of high-performance engineering computing. A private cloud is inside an enterprise firewall or securely hosted by a service provider, providing users with all the benefits of cloud computing, but with less exposure to Internet security risks. Typically, private clouds are client owned, managed by the client or a service provider. Clients usually define their customized requirements for cloud access and usage policies. Compared to public cloud options, advantages include *better security, privacy, customization, and control* of the HPC environment. Private clouds are the next frontier for enhancing engineering innovation, productivity, and collaboration.

IT and Engineering must use a cost–benefit framework for HPC investments

Engineering enterprises must develop the business case for their HPC investments through a collaborative process involving IT and engineering end-users and management. This collaborative process helps to jointly justify investment decisions, improves the IT organization's effectiveness, and deepens collaboration between engineers and the IT organization, as they all work toward common goals. They should compute the total cost of ownership (TCO) over several years to make objective decisions while evaluating various HPC infrastructure options. However, reducing the TCO alone is inadequate. The team must use a framework of inter-related drivers and associated metrics that minimizes the *total costs* incurred while maximizing the *value* delivered by engineering simulation solutions.

---

[3] McKinsey Global Institute Research, "Big Data: The Next Frontier for Innovation, Competition, and Productivity", May 2011.
[4] Gartner 2010 CIO Survey

Business value. The business value derived from an investment in HPC for engineering simulation comes in one of several forms:

- *Strategic value:* faster time to market, increased profits, improved brand equity, better partnerships with stakeholders, ability to attract and retain top global engineering talent
- *Engineering value:* breakthrough engineering and innovation, broader and deeper engineering collaboration, greater engineering insights, more accurate understanding, better products
- *Operational Value:* faster time to results, reduced cost of product development, more users supported, improved user productivity, better capacity planning
- *IT Value:* improved system management, administration, and provisioning; enhanced security; higher utilization; scalability; reduced downtime; access to robust proven technology and infrastructure management expertise

IT costs. It is essential to recognize costs from a range of sources, including:

- *Data center capital*: new servers, storage, networks, power distribution units, chillers, software purchase, etc.
- *Data center facilities:* land, buildings, containers, facilities maintenance, etc.
- *Operational costs:* labor (salaries for end-users and IT staff), energy, IT hardware and software maintenance, software license, etc.
- *Other costs:* deployment and training, downtime, bandwidth, etc.

Engineering companies must continually evaluate the costs and value of HPC within this broad cost–benefit framework. They must focus on maximizing *value* not just for one application instance but a collection of workloads typical in their production environments. This objective analysis is driving many HPC engineering enterprises to centralize and consider cloud computing.

Why centralize engineering HPC?

A combination of high-performance computing systems, massive storage systems, visualization, applications, and middleware, all connected by high-speed networks, is needed for today's engineering HPC infrastructure. There are several key reasons to centralize engineering HPC:

- **Allows you to balance requirements for large-scale computing with high throughput capacity.** Engineering simulation workloads typically consist of a mix of a few long-running large jobs and many smaller jobs. Consolidating this workload on a flexible yet economical centralized infrastructure best optimizes this heterogeneous workload.

- **Can reduce the TCO.** An independent 2005 TCO[5] study (Fig. 2) for campus HPC conducted by Indiana University illustrates these TCO drivers for two configuration instances: one teraflop (a medium-scale system at the time) and hundred teraflops of peak performance (a very large-scale system at the time). The study analyzed three cases for each instance: worst-case distributed, best-case distributed, and completely centralized.

[5] Bradley Wheeler and Thomas Hacker, "Centralize Research Computing to Drive Innovation…Really", http://net.educause.edu/ir/library/powerpoint/EDU05163.pps.

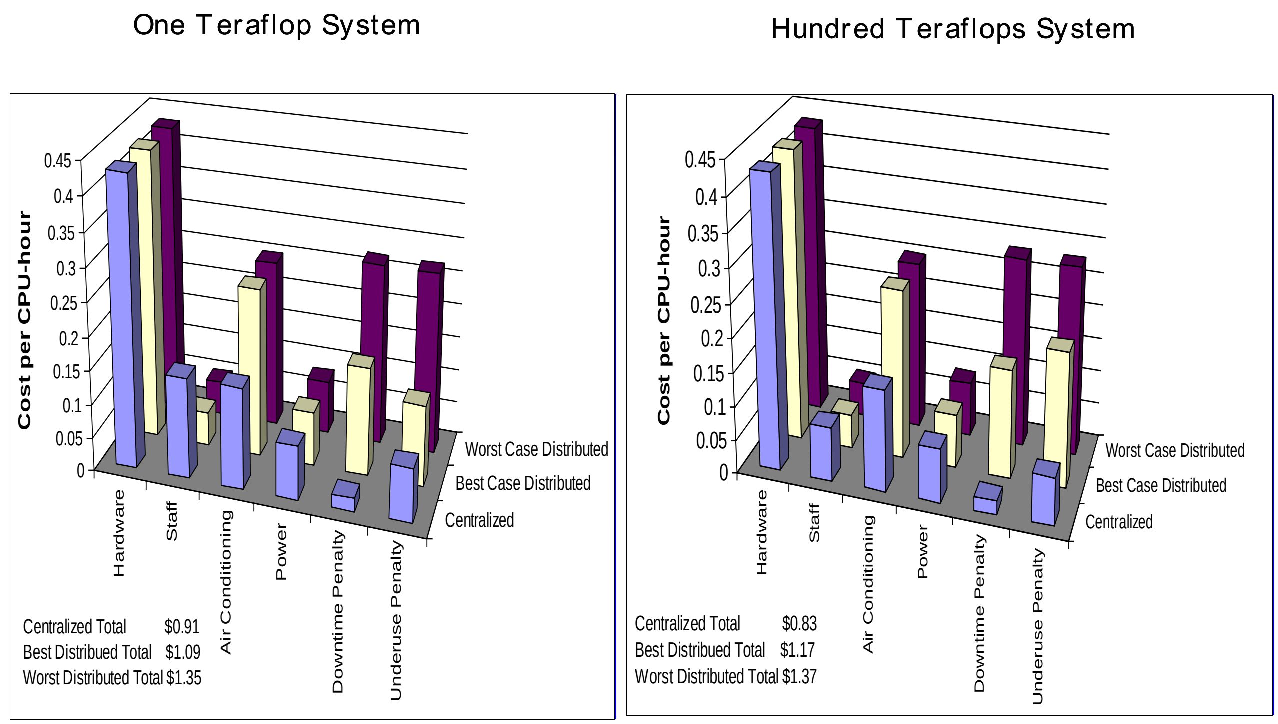


Fig. 2. A 2005 Indiana University TCO analysis for distributed vs. centralized HPC[4]

In each instance, the TCO as measured by $/CPU-hour is *lowest for centralization.* This clearly illustrates better economics with centralization: *improved energy efficiency, utilization and, reliability as well as lower operational costs.* Better still, for the larger 100 teraflops system, this TCO savings margin was even greater ($.83/CPU-hour vs. $.91/CPU-hour). The marginal TCO gains get better as centralized HPC systems scale to even higher levels of performance.

Next, we examine individual cost components in detail. As should be expected, hardware costs are the same for each case. The power costs for facilities and IT equipment are lowest for centralized. It takes only one large central IT facility to centralize versus several smaller IT facilities for a distributed architecture. As electricity and real estate costs continue to escalate, these costs will become a larger fraction of the TCO.

In this study, IT staff costs for centralized are higher, because in distributed scenarios, graduate students managed many of these systems, and their costs were not explicitly accounted for. This could also happen at engineering enterprises, as these staff costs could be "hidden" under the individual department budget and not explicitly included in the enterprise IT costs for labor and systems administration. IT managers at engineering enterprises must accurately account for these hidden staff costs during cost-benefit analysis.

The downtime and underuse penalties for centralized systems are lowest, and this differential is greater on the larger system — clearly indicating that centralized HPC infrastructures are *more reliable* and can *better optimize utilization* for a range of workloads. This may be counterintuitive to many engineers who are accustomed to using local resources. However, an engineer who has dedicated access to only one system suffers complete loss of productivity for the duration of time when that system is down. In contrast, an engineer who can access a centralized system continues to be productive even when many components of this centralized system are unavailable.

While software-licensing costs were not considered in this Indiana University study, it is important to note that with centralization, cost benefits are likely to accrue to engineering enterprises from the ability to *improve the utilization* of engineering simulation software licenses, and these enterprises could potentially optimize license fees and availability of software needed for engineering productivity.

- **Enables more effective data management, improving engineering collaboration and productivity.** When teams of engineers collaborate on a single system design, maintaining data and context centrally is critical to improve collaboration and productivity throughout the engineering enterprise. Preserving both the data and the context of a simulation provides valuable understanding of the inputs used to produce the simulation result, such as details of the design of the part or system, the operating/boundary conditions, and what simulation tools, assumptions, and software versions were used. This significantly improves engineering productivity and collaboration and allows re-use of the simulation data for future, related design projects.

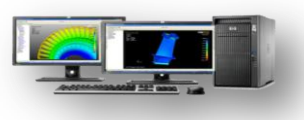
- **Enhances efficiency and better secures data and intellectual property.** A single engineering simulation produces *up to hundreds of GBs of data.* Engineers study these large data sets and produce summary reports and recommendations, which are the engineering insights and key IP associated with simulation. Typically, the simulation data is scattered throughout an enterprise from corporate databases to shared drives to an engineer's personal hard drive, *creating a security exposure for data and other IP.* Systematic means of securing, tracking and using simulation data enhances efficiency, and centralization is key. Further, as storage continues to become more affordable, it is possible to keep and manage work-in-progress simulation results centrally, saving engineering time and network bandwidth associated with moving files between the end-user and the compute resource.

- **Is the best weapon to combat network limitations.** Today, network limitations are getting worse even as computing and storage become *more economical. Storage capacity* per dollar continues to double every 12 months, and computing performance doubles every 18 months[6]. Therefore, it is possible to store and analyze relevant engineering simulation data even though it is growing considerably. However, it is very *hard to effectively share* this data over collaborating locations, as s*ite-to-site network performance doubles only every* 4 years. From the 1970s to today, the ratio of computing performance to wide-area circuit bandwidth has grown from 20 flops/bit to 1,000 flops/bit, hurting the balance between computing performance to network bandwidth[7]. The ratio of data set size to network bandwidth is far worse!

  Engineering simulation applications produce very large data sets (results and intermediate files). Moving these files stresses network bandwidth and creates long wait times and lost engineering productivity. Interactive applications work best if wide-area network (WAN) *latency effects are minimized*, further motivating the need to centralize data and all facets of the computing workload.

[6] "The data deluge", the *Economist*, February 2010.
[7] Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies - A Workshop Report and Recommendations , EDUCAUSE Campus Cyberinfrastructure Working Group and Coalition for Academic Scientific Computation, February 2009.

- **Supports virtual desktop initiatives and mobile platform strategies.** Currently, most engineering HPC users work using a combination of a local desktop to build the model and then upload the file in a remote batch configuration to an HPC resource for computation. Then, the result file is sent back to the local desktop machine for visualization. In reality, the local machine may be inadequate for the size of job that the HPC resource can handle. Therefore, the local machine impedes the adoption of high-fidelity simulation. Centralization, in conjunction with an effective remote and interactive user experience, provides an alternative to this batch workflow as it retains the data in-place, co-located with the computing resources (Fig. 3).



Local computing

- Pre-processing/ solve/ post-processing on local desktop system
- Files stored locally under individual control
- Inherent capacity limitations; also limits collaboration and data management

Centralized computing with interactive remote access

- Solver execution conducted on central remote HPC resource
- Pre-processing/post-processing also conducted remotely, utilizing thin client technology
- Simulation files kept centrally, so bottlenecks related to file transfer minimized
- Limitations of local hardware minimized such as the inability to post-process large files on the local machine
- Emerging remote access and job management solutions enhance collaboration and data management

Fig. 3. Advantages of centralized and interactive process for engineering HPC workloads

What is needed to centralize engineering HPC?

To achieve operational efficiency, a strategy to centralize IT infrastructure must include:

- Centralized management of applications, workloads, and data to improve system utilization and efficiency and to ensure IP protection
- Workstation access and/or virtual desktop and mobile platform strategies that ensure end-user productivity across both interactive and computational phases of the engineering workload.

There are three key capabilities needed to achieve best practice:

- **Maintain data local to the HPC resource.** Engineering enterprises must evolve to a *data-centric* view of resources and build up shared data repositories that are *co-located* with the compute resource. This is vital to eliminate network bottlenecks, optimize the remote workflow, and enable collaboration.

- **Implement common integrated tools for workload scheduling and management software.** Typically, multiple heterogeneous workloads run in production at most engineering firms. Workload scheduling and management solutions available from several software firms can significantly *increase overall system utilization, efficiency and throughput while reducing time to results* in centralized cluster/grid and cloud configurations. When implemented in a common unified environment, these tools can further enhance engineering productivity and IT efficiency.

- **Implement a remote display strategy for remote interactive access and mobile platform support.** For users to actually work in a mode that is remote from the data, they need full interactive and graphical remote work processes. This enhances engineering productivity and eliminates network bottlenecks associated with moving files between the end-user and the computational resources.

A remote simulation architecture for ANSYS

Leading enterprises are already working to deploy an improved remote workflow for the centralized engineering HPC/private cloud. The user is remote from the resource and remote from the data, which now is managed by a central server (Fig. 4). The files reside at the HPC resource leading to better efficiency, data security, enhanced collaboration, and data access. Engineering enterprises can now conduct a full simulation via remote access using graphics servers co-located with the central computing and data storage.

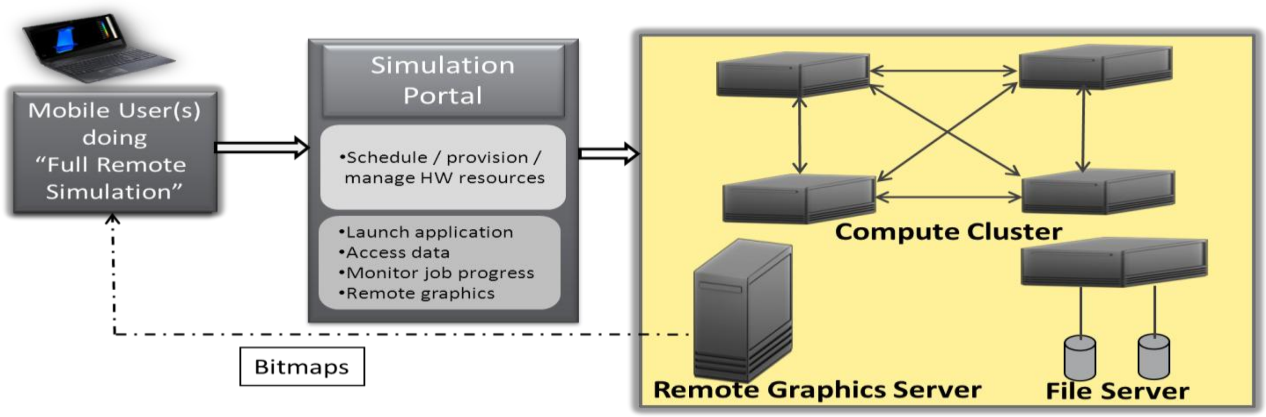


Fig. 4. Engineering simulation through remote interactive access

There are three components to this architecture:

Common integrated tools for end-user access. The simulation portal, which might be based on third-party tools and/or custom in-house tools, is integrated with ANSYS applications. This portal provides a remote user with the ability to *manage* the hardware, *launch* ANSYS applications, and *access* data and data management tools.

Remote visualization and graphics software and servers. A remote 3-D visualization tool allows users to run interactive ANSYS applications on the server side, with simple bitmaps sent back to the remote user, enabling users to *visualize results remotely* while keeping the data centralized.

Data management. Users have access to the centrally stored data, which is managed via intelligent data management tools that provide high-performance access in a controlled manner—in this case, ANSYS Engineering Knowledge Manager (EKM). This product is a comprehensive solution for simulation-based process and data management, and has several features that enable engineering project managers to *manage* and *secure* data centrally:

- Actual simulation files (large results files) reside on a normal scalable file store, enabling standard IT processes for backup and archive.
- EKM is database agnostic, so IT can use whatever database it currently uses or is familiar with.
- EKM is scalable – IT can start with a single site and can take on more complex federated file structures as it grows.

Putting it all together. The processes and best practices described here are achievable goals that can add measureable benefits to your overall simulation investment. While neither ANSYS nor any other engineering analysis application provider offers all aspects of the integrated HPC, IT, and engineering simulation solution as an off-the-shelf product today, it can be implemented using solutions from ANSYS and its strategic partners. In fact, leading ANSYS enterprise customers *are also investing* today to develop such capabilities.

Conclusions

Engineering HPC architectures started out with centralized mainframes in the 1970s. Now they have come a full-circle, after going through client-server and cluster computing distributed architectures. There are *four* triggers for this centralization:

- Engineering workloads have become large and no longer can be run efficiently on local servers. IT must balance this large-scale computing capability with high-throughput capacity.
- Shrinking budgets mean that IT departments need to do more with less, and centralization is proven to reduce TCO for HPC workloads and increase license and hardware utilization.
- Wide-area network bandwidth performance is inadequate to transfer large engineering simulation data sets back and forth during collaborative analysis.
- Centralized data management improves engineering collaboration and productivity. It also is better at protecting IP and securing data with the proliferation of remote globally distributed users and mobile computing devices.

However, centralizing engineering HPC is a challenge since workloads and data transfer requirements are massive for both computing and visualization, and product and design teams need to collaborate across widely scattered locations. The ANSYS *remote interactive access* architecture, built collaboratively with strategic IT partners, is an attractive option that has been tested and found to work in real-life engineering computing environments. With this centralized or private cloud approach, engineering enterprises can optimize their business value from HPC and benefit from improved productivity, collaboration, and innovation.