

SGI[®] Solutions for BioSciences

Shorten Your Discovery Pipeline
with an Integrated Compute and
Data Management Environment



Pharmaceutical and biochemistry research organizations understand pressure. And that may be truer today than ever before.

Whether your work involves genomics, bioinformatics or computational chemistry, your organization must find ways to shorten your discovery pipeline. Meanwhile, your researchers face a persistently growing workload without a commensurate increase in resources.

In other words, you're expected to conduct more science – but without more scientists.

Productivity Crunch Meets Data Explosion

Requiring scientists to do more with less has made improving productivity a more pressing objective than ever before. Tight budgets, however, can complicate efforts to streamline key processes with new hardware, software and administrative tools. When new computers or storage systems are acquired, many organizations have found them difficult to integrate with existing systems. Those acquisitions make already complex information technology (IT) environments even more difficult to administer and use.

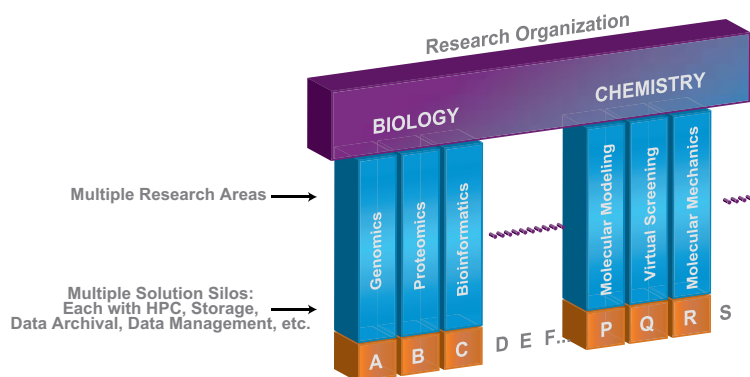
Research organizations face another challenge: the explosive growth of scientific data. Rapidly evolving methods for capturing, analyzing and interpreting data have pushed many institutions to the brink of their ability to compute, store and retrieve the information they need.

The Costs of Disjointed Workflows

As the needs of your organization have evolved, so have its IT resources: its high-performance computing (HPC) systems and high-throughput clusters, its storage and backup systems, and its data management and archiving solutions. But when organizations provision for specific scientific workflows, they create resource “silos” that are often disconnected from others. These silos result in costly and needless redundancies as each department increases its presence in the data center with systems that likely will be used for only a few applications.

Obstacles to Productivity

- Larger and more complex problems slow time to discovery
- Exploding data volumes swamp resources
- Disparate ‘silos’ of systems and data lead to costly, redundant resources
- Disjointed and complex workflows hinder collaboration
- System administration steals time away from science



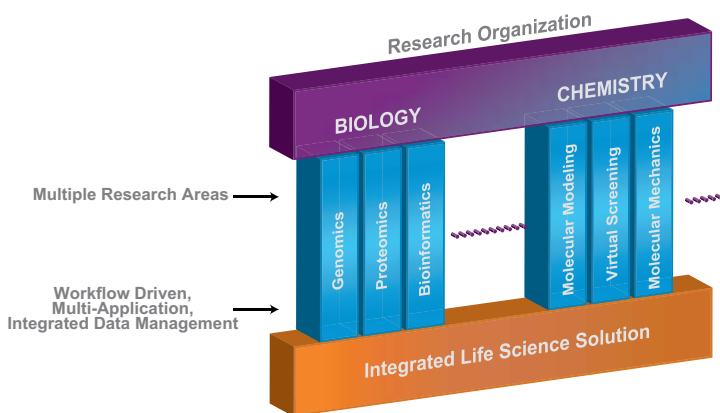
All this has created an increasingly inefficient and complex environment for users. As data sets grow – regularly used proteomics data bases run to 8 TB in size – scientists spend more and more time accessing data and managing where jobs run. They spend more hours trying to gauge which computing and storage resources best meet their needs at any given time. And in an era where institutions are putting a premium on productivity, these complex environments are anything but productive.

SGI® Solutions for BioSciences

Achieving an Integrated Environment with SGI

Historically, this hasn't been an easy problem to solve. Many solution providers put the burden on researchers to force-fit their workflows to the computing platform, letting the computer's strengths and weaknesses determine how scientists should work.

SGI sees things differently. After a quarter century of focusing on the unique requirements of scientific computing, SGI knows that the solution should be tailored to fit the workflow – not the other way around. SGI helps scientific enterprises overcome their productivity challenges by integrating multiple computing and storage architectures into a flexible solution that can be tailored to a wide range of workflows. By leveraging the award-winning SGI® Altix® and SGI® InfiniteStorage architectures, organizations can easily ensure SGI solutions optimally meet their current needs and evolve them to keep pace with future demands.



Bottom-Line Benefits

SGI's uniquely flexible approach allows researchers in multiple disciplines to share a common set of resources and to more easily collaborate with colleagues. That means you can maximize the return on your organization's IT investments. You can also reduce your ownership costs by serving more workgroups with a single group of assets. And best of all, you can enable your researchers to focus on science, and not on computer system management.

High-Performance and High-Throughput Platforms

SGI provides industry-leading solutions built on the two fundamental types of computing platforms that most research organizations need: high-performance systems and high-throughput clusters.

High performance for large-scale problems. Scientists use high-performance platforms to gain a detailed perspective on a specific – and usually large – problem by leveraging the system's ability to apply many processors working together to shorten time to insight. High-performance systems are commonly used for:

- Genomics
- Proteomics
- Access to large database systems
- Computational chemistry
- Molecular modeling

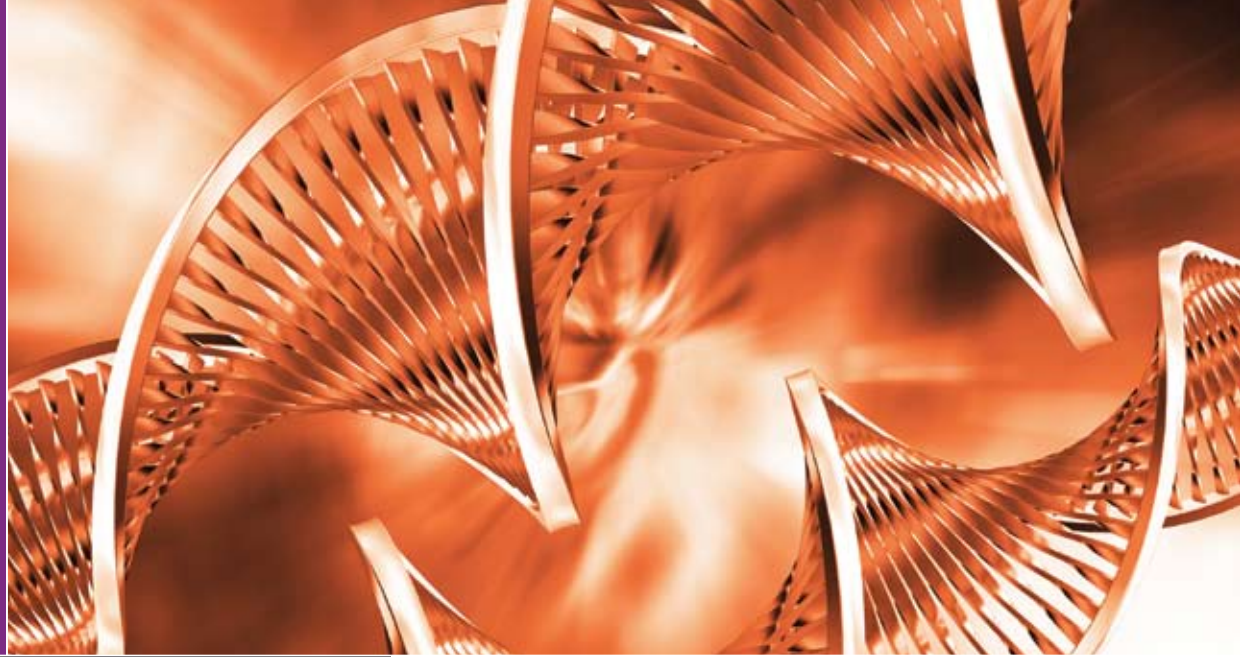
These applications tend to be highly interactive. They require a platform featuring a shared-memory architecture that allows all of the system's memory to be available to the problem. These systems tend to scale to large configurations, up to thousands of processors in a single

High throughput for multiple jobs. High-throughput systems are typically designed to emphasize throughput over computational speed. This proves enormously beneficial for a range of BioSciences applications, including:

- Gene sequence analyses
- Bioinformatics statistical analyses
- Virtual screening
- Large genomic or proteomic database searches

Advantages of an Integrated SGI Environment

- Lower cost of ownership from greater utilization and simplified management
- Faster return on investment
- Ability to solve larger problems
- Capacity to solve more problems
- More flexibility to support changing workloads
- Simplified access to large data volumes
- Researchers focus on science, not system administration



A Common Computing Environment

SGI bases all its solution components on a common environment. This allows your researchers to move from one platform to another, and to manage projects that span multiple systems and technologies.

This SGI environment provides:

- A common Linux operating system across all platforms so users need learn only one environment
- One job scheduling and system management environment to manage and provision workflows across all platforms
- Comprehensive storage solutions to simplify data sharing, access, movement, and management across systems
- Optional performance and tuning enhancements

Typically, these systems take the form of clusters that combine industry-standard nodes offering small amounts of internal scaling. The ability of high-throughput systems to run a large number of independent jobs tends to be more important than their ability to accelerate the completion of a single, larger job.

SGI's Modular Solution Platform for BioSciences

Scientific research can be unpredictable. That's why SGI developed a modular solution platform that keeps pace with the dynamic nature of research. This platform integrates SGI's full line of scalable shared-memory Linux® systems for high-performance applications, "scale-out" Linux clusters and servers for high-throughput applications, and storage capabilities in a way that can be tuned to match specific requirements.

Only this type of modular, fully integrated platform can address the changing needs of today's researchers.

Domain Expertise = Solutions that Make Sense

Unlike many vendors, SGI not only has the deep domain expertise to optimize workflow solutions for your unique environment, but the commitment to share that expertise no matter how large or small your implementation. With SGI, you'll have direct access to Ph.D.-level BioSciences experts who engage hands-on with your organization to design an integrated solution that meets your current requirements, while leaving the necessary headroom to accommodate future needs.

The result is a lower cost of ownership that only true BioSciences expertise and ongoing professional support can bring.

SGI® Solutions for BioSciences

SGI® Altix®: A High-Performance Shared-Memory Platform

SGI® Altix® servers provide industry-leading compute performance and scalability for computationally intensive correlations and simulations, and large-scale scientific databases and sequence assemblies.

Built using SGI's advanced NUMAflex® technology, these mid-range and high-end systems utilize Intel® Itanium® Processors. Altix systems can scale from four to 1,024 CPU cores in a single, shared-memory environment. This design minimizes system overhead and makes it easy for your applications to run faster. The result is that large scientific databases, complex sequence assembly and large model molecular dynamics scale to new levels of performance, enabling faster insights and a deeper understanding of chemical and genomic processes. And with mid-range systems scaling up to 76 CPU cores and over 400GB of memory, even smaller organizations can solve problems that, until recently, were all but unsolvable.

SGI® Altix® XE: A High-Throughput Cluster Platform

SGI® Altix® XE cluster systems provide industry-leading high-throughput capability for virtual screening, cluster analysis, statistical analyses, and a wide range of other bioscience applications. Many of these applications are optimized to take advantage of the superior performance of Dual-Core or Quad-Core Quad-Core Quad-Core Intel® Xeon® Processors.

Connected with Gigabit Ethernet and/or InfiniBand, and accessing a common set of storage through high-performance SGI NAS platforms, these systems can support up to 160 CPU cores with almost 2 TFLOPS of raw performance in a single rack, with almost unlimited high-throughput scalability. Altix XE clusters come pre-configured and tested with a software stack of tools and applications, so researchers can spend their time pursuing science rather than making computers work. When integrated as a scientific workflow solution for genomics and proteomics, bioinformatics, or computational chemistry, Altix XE cluster systems allow high-throughput jobs to be run on systems which are explicitly tailored to the task, allowing high-performance jobs (or those that require large memory) and any "overflow" high-throughput jobs to be run on SGI® Altix® systems.

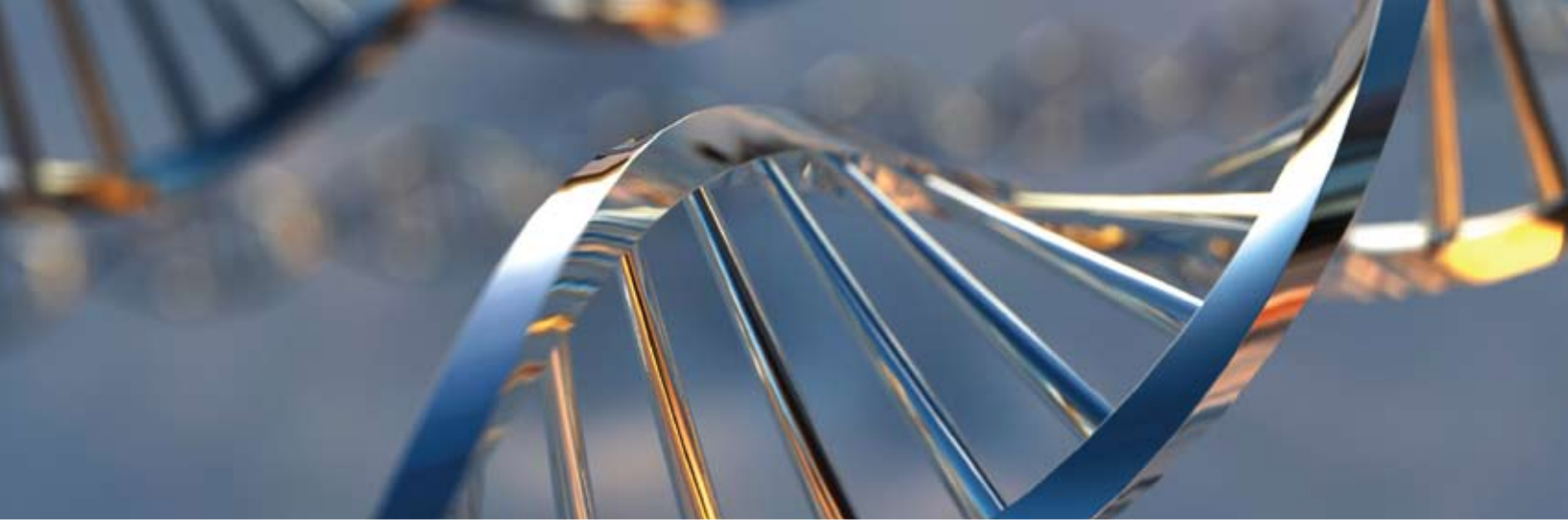
SGI® Altix® ICE: High Throughput in a Small Footprint

SGI® Altix® ICE is a breakthrough line of bladed servers purpose-built to handle large BioSciences problems and scale-out workloads. SGI® Altix® ICE enables high-throughput environments while minimizing their demands on your data center's space and power. Its ultra-dense rack architecture allows a single SGI® Altix® ICE rack to be powered by as many as 512 Quad-Core Intel® Xeon® Processors cores and deliver 6 TFLOPS of performance. This packs more compute power onto every floor tile, so your organization can make the most of its valuable data center space.

With Altix ICE, SGI took advantage of energy-conscious and field-proven innovations originally developed for its Intel® Itanium® processor-based Altix® line, equipping the system with an energy-smart power architecture that costs less to operate. And to keep large and dense configurations running cool, SGI® Altix® ICE boasts SGI's new, third-generation water-cooled door design. The optional cooling feature enables chilled water running through the system's unique hinged door to carry away up to 95 percent of the heat generated by an SGI® Altix® ICE system.

And like Altix XE clusters, SGI® Altix® ICE systems come pre-configured with SGI's comprehensive software solution stack, enabling fast and easy deployment upon arrival.





Institute of Biophysics-UFRJ: Modelling Bio-Molecular 3D Structures with Altix XE

The natural next step after conducting genomics studies is to determine of three-dimensional structures of proteins that are possible targets for new drugs, vaccines or bio-technological processes. In Brazil and other developing countries, modelling the 3D structures of new HIV mutants and increasingly drug-resistant strains of malaria could lead to breakthroughs that dramatically improve the quality of life for millions of people.

No one knows this better than researchers at the Institute of Biophysics at Federal University of Rio de Janeiro (UFRJ), which recently acquired an SGI® Altix® XE cluster powered by 96 Intel Xeon processor cores and equipped with 96GB of memory. In work supported by Brazilian research funding agencies from the Ministry of Science and Technology and the State of Rio de Janeiro, UFRJ scientists aim to computationally test a large number of leading molecules against different mutant targets of HIV and malaria. Under the direction of Dr. Paulo M. Bisch, a professor at the UFRJ Institute of Biophysics, and the institute's Dr. Pedro G. Pascutti, researchers are using GROMACS molecular dynamics simulations to model the interaction of potential drugs and selected targets to screen and identify the best matches for local mutant proteins. The team will then identify leading compounds for use in pre-clinical experiments to be deployed in carefully supervised HIV and malaria programs across the country.

Using the SGI® Altix® XE system – Brazil's first to be powered by quad-core Intel Xeon processors – scientists have run very large molecular dynamics queries. The large queries reduce the time needed to test different sequences and target structures. They also increase the number of leading compounds to be queried. The Altix XE cluster also integrates with the GRID-UFRJ project, part of an effort to combine the power of all cluster systems in the Federal University of Rio de Janeiro to create a resource capable of handling even larger computational experiments. The team has also been developing, in collaboration with Dr. Marta Mattoso of the UFRJ Graduate School of Engineering, a new distributed workflow to enable molecular dynamics simulations using UFRJ's GRID architecture.

The Institute of Biophysics also is using the SGI® Altix® XE cluster to benefit research beyond its own: The institute takes part in the e-science research grid network EELA (E-infrastructure shared between Europe and Latin America). UFRJ's biophysics research team is charged with running GRID-BIOMOL 3D, one of the key EELA applications.

SGI® RASC™: Targeted Acceleration on a Dramatic Scale

When the goal is to run large or multiple BLAST-n nucleotide searches or to quickly produce simulations of complex proteins over specific timescales, researchers benefit from a significant speedup in processing. Field Programmable Gate Arrays (FPGA) is a breakthrough technology that enables dramatic speedup in modeling and query run times. SGI® Altix® servers with SGI® RASC™ (Reconfigurable Application Specific Computing) computation blades leverage the power of FPGAs, utilizing gate array technology that can be reconfigured by the user for optimal performance on a specific algorithm.

By integrating FPGA based SGI RASC technology into Altix systems, users can develop capabilities that run 10, 100, or even 1,000 times faster than traditional solutions. Yet these enhancements come with only minimal increases in power consumption. With SGI RASC, system price and footprint are significantly lower than large clustered servers with comparable processing power.

Data Solutions: High-Performance Storage and Information Life Cycle Management

As data volumes grow, high-performance data access is the limiting factor in several BioSciences environments. In some fields, including computational chemistry, advanced versioning and data archiving techniques are required for regulatory approval.

SGI has spent more than a decade solving some of the world's most daunting storage and data management problems. SGI's family of storage platforms and data management environments leverages these innovations to underpin of all SGI scientific workflow solutions. They allow multiple computer room systems to access a single copy of data in an efficient, high-performance environment and extend that single view of data directly to desktops. They also address your organization's data archiving needs with information lifecycle management software and high-performance, low-cost tape alternatives.

Consider how SGI helps organizations manage the archiving and retrieval of scientific data with the SGI InfiniteStorage Data Migration Facility (DMF). DMF quickly and efficiently moves older data files to a tape archive, making the much faster disk space self-managing.

Working with the intuitive DMF software interface, your organization can more easily respond to your changing data access patterns. In a typical data life-cycle, new data is accessed intensively, while aging data is needed less frequently. In fact, older data may remain dormant for many months, or even years, and still occasionally undergo brief periods of renewed access. Without DMF, most sites would have to either maintain a huge pool of inactive data on expensive disk storage or manually archive that data to tape, where it is difficult or impossible to access on demand.

With DMF, organizations can define the data migration policy necessary to always keep high priority information close at hand, while lower priority information is seamlessly migrated to tape. This gives administrators control over how data moves from near-line storage to tape archives, ensuring that data migration dovetails with the unique workflow of your scientists.

The SGI Advantage for BioSciences

With an integrated SGI solution capable of supporting a range of multiple research workflows, your organization can improve productivity and remain flexible to not only meet today's challenges, but the challenges that lie ahead. An environment based on SGI's award-winning compute, storage and data management can help you shorten your discovery pipeline while still meeting the heightened expectations of management and the marketplace. Meanwhile, you'll better protect your IT investments with an open-standards environment that scales along with your needs.

UC-Riverside: Mapping the Barley Genome with Altix

Is barley poised to be the next super-plant? It's already an important crop in the United States, both as animal feed and for use in foods ranging from bread and pastries to beer and soup. And in climates too cold for raising maize or rice, barley is fast emerging as a source of bio-fuel.

With a 64-core, 128GB SGI® Altix® 4700 system, the University of California-Riverside (UC-Riverside) is building accurate maps of the barley genome. The project is funded by the National Science Foundation and United States Department of Agriculture, and aims to leverage the tools of genomics to develop new barley cultivar, such as a high-protein variety or one especially resistant to disease. Under the direction of Dr. Stefano Lonardi and Dr. Tim Close, UC-Riverside researchers are constructing a high-density genetic map of barley. The team will then identify genes linked to important traits with molecular markers, to be deployed in 10 breeding programs across the country.

With UC-Riverside's SGI® Altix® system, which is also equipped with two SGI RASC boards, scientists have run very large BLAST queries. This has cut down the time needed to update the barley database each time a new release of related genomic resources, such as rice gene models, becomes available. The team has also been developing new algorithms to build better physical and genetic maps. Yonghui Wu, a graduate student in Dr. Lonardi's lab, has developed a novel algorithm to assign genes to genomic clones that exploits the physical map. To achieve this, the Altix system must solve massive linear optimization programs with tens of thousands of variables and constraints.

UC-Riverside's Department of Computer Science and Engineering acquired the SGI system with funding from the NSF. The resource is used for a range of scientific and engineering projects, not only at UC-Riverside but throughout the UC System.

LEARN MORE

LEARN HOW SGI CAN HELP YOU OVERCOME THE CHALLENGES OF DATA-INTENSIVE SCIENTIFIC WORKFLOWS AND SHORTEN YOUR DISCOVERY PIPELINE.

**FOR MORE INFORMATION,
VISIT: SGI.COM/INDUSTRIES/BIOSCI**

Corporate Office
1140 E. Arques Avenue
Sunnyvale, CA 94085
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5185.2860
Europe +44 118.912.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000

© 2008 SGI. All rights reserved. Specifications subject to change without notice. Silicon Graphics, SGI and Altix are registered trademarks of Silicon Graphics, Inc., in the U.S. and/or other countries worldwide. Intel and Itanium are registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. All other trademarks mentioned herein are the property of their respective owners.
4069 [01.2008]

J15350

