

## IBM @server® Blue Gene Solution



---

### Highlights

---

- *Leadership performance in a space-saving, power-efficient package for the most demanding high performance computing applications*
- *Design innovations for advancement of science across an important set of vital workloads*
- *Technology accessible to a wide range of scientists, researchers and developers*

### Supercomputing leadership

The IBM® @server® Blue Gene® Solution is the result of an IBM supercomputing project begun over five years ago dedicated to building a new family of supercomputers optimized for bandwidth, scalability and the ability to handle large amounts of data while consuming a fraction of the power and floor space required by today's high performance systems.

Blue Gene ranks as the number one and number two supercomputers on the TOP500 list<sup>1</sup> along with three other entries in the top 10. The world's fastest supercomputer is installed at the Department of Energy's/National Nuclear Security Administration's Lawrence Livermore National Laboratory (LLNL). The fully configured 64 rack, 130,000 PowerPC® Blue Gene system at LLNL has achieved an astonishing 367 peak teraflops. The DOE entered into a partnership with IBM in funding research and development of this machine in 2000 in order to explore designs for highly cost-effective computers.

But Blue Gene is not just a supercomputer that delivers ultrascaleable performance. It is also extremely efficient. Because of unique design points that allow dense packaging of processors, memory and interconnect, Blue Gene offers leadership efficiency in the areas

## Unsurpassed performance, ultrascalable computing

of power and floor space consumption. With most new HPC application development designed for clusters, efficient use of power and space now rivals scalability as the biggest challenge. Available in configurations ranging from one to 64 racks, Blue Gene is the innovative new solution from IBM that delivers an ultrascalable solution without sacrificing efficiency.

### Applicable to important workloads

When first chartered over five years ago, the Blue Gene project had as its goal to develop a massively parallel computer applied to the study of biomolecular phenomena such as protein folding. The effort would advance the understanding of the mechanisms behind protein folding via large-scale simulation, and explore novel ideas in massively parallel machine architecture and software. The level of performance provided by Blue Gene can enable a tremendous increase in the scale of simulations beyond what is possible with other supercomputers. Successful simulation studies of protein folding on this scale are expected to advance the techniques, models, and algorithms used in biomolecular simulation.

Hands-on experiences with many differing applications have shown that the Blue Gene architecture is applicable to a number of parallel workloads found across a variety of disciplines. Today, IBM and its collaborators are exploring a growing list of high-performance computing (HPC) applications including life sciences, financial modeling, hydrodynamics, quantum chemistry, molecular dynamics, astronomy and space research, materials science and climate modeling.

Some examples of early Blue Gene usage are summarized below:

- *Blue Gene will help researchers at the National Center for Atmospheric Research (NCAR) perform atmospheric modeling and ensure that the center maintains a leadership position in the field. By using supercomputers like the Blue Gene system, NCAR can model and analyze data faster and can estimate how natural factors and human induced changes to the atmosphere are affecting our climate.*
- *ASTRON, a leading astronomy organization in The Netherlands will use Blue Gene as the central processing engine for a new type of radio telescope called LOFAR, capable of examining the beginnings of the earliest stars and galaxies after the formation of the universe.*
- *Computational Biology Research Center at Japan's National Institute of Advanced Industrial Science and Technology (AIST) will use the extreme computational power of Blue Gene for structure prediction and dynamics study on disease-related proteins, protein-ligand binding analysis, and virtual screening for searching new lead compounds. These are the key techniques for new generation e-drug design and discovery.*
- *Boston University intends to use Blue Gene to help tackle a host of difficult scientific problems ranging from subnuclear physics through genetics and cellular biology to the modeling of space weather and ocean systems. For example, BU researchers intend to use Blue Gene to predict how activities on the surface of the Sun, such as solar flares, affect the Earth's radiation belt, its upper atmosphere, and the ionosphere.*

As greater numbers of scientists and researchers apply large-scale cluster computing to a diverse set of complex problems and build a collective expertise in parallel program development, the relevance of the Blue Gene architecture becomes clearer. The design innovations of Blue Gene offer the promise of advancing vital science across numerous disciplines.

### **Widely accessible**

Blue Gene is a powerful, multi-teraflop system constructed from technology that can lead to petaflop (1000 teraflops) performance, and many interested scientists and engineers might conclude that harnessing Blue Gene is simply beyond their reach. In response, IBM has taken steps to make Blue Gene accessible to a wide range of users.

IBM provides access to Blue Gene through the Deep Computing Capacity on Demand (DCCoD) center. Clients with constrained budgets and limited requirements for access to a Blue Gene system could request time on the system and just pay for the amount of capacity reserved. In this way, clients

could contract for variable capacity and services to help satisfy short term planned or unplanned peak workloads. Remote access would be provided via a dedicated Virtual Private Network connection between the client site and IBM's facility. Accessing Blue Gene through the DCCoD center could help clients quickly tap Blue Gene supercomputing power while helping to reduce financial and technical risk.

Blue Gene systems can be leased and outright purchases can be financed through IBM Global Financing (IGF). IGF can help clients control costs with highly competitive rates, powerful asset-management tools, and end-of-lease or end-of-life options that maximize flexibility while minimizing risk.

Finally, IBM has installed the world's most powerful privately owned supercomputer, the Watson Blue Gene system, nicknamed BGW, at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. With a processing speed of over 90 teraflops, BGW is comprised of 20 racks. IBM plans to use the system to explore how BGW's unprecedented power might enable an extraordinary period of

progress in a range of technical fields and business applications. Depending on availability, selected researchers outside IBM will be granted access to some of the capacity of BGW through IBM's Deep Computing Capacity on Demand offering. In this way, access to an extremely large-scale supercomputer can be made available that would otherwise be beyond the reach of most researchers and academics.

### **Innovative design for ultrascale computing**

The Blue Gene system is built out of a very large number of compute nodes, each of which has a relatively modest clock rate contributing to both low power consumption and low cost. Blue Gene utilizes IBM PowerPC® embedded processors, embedded DRAM and system-on-a-chip techniques that allow for integration of all system functions including compute processor, communications processor, three cache levels, and multiple high speed interconnection networks with sophisticated routing onto a single ASIC. Because of a relatively modest processor cycle time, the memory is

## Unsurpassed performance, ultrascale computing

close, in terms of cycles, to the processor. This is also advantageous for power consumption and enables construction of dense packages in which 1024 dual-processor compute nodes can be placed within a single rack. Blue Gene can be scaled up to 65,536 compute nodes yielding a peak performance of 367 Teraflops with extremely cost-effective characteristics and low-power, cooling and floor space requirements.

Blue Gene is built up as follows: two nodes per compute card, 16 compute cards per node board, 16 node boards per 512-node midplane and two midplanes in a 1024-node rack. Each processor can perform four floating point operations per cycle. Depending on the nature of the application to be run on Blue Gene, the programmer may choose to employ both processors in a compute node for computation or have one processor dedicated to handling message passing operations. In addition to the compute nodes, the system

provides for a flexible number of dual-processor I/O nodes which handle communication between compute nodes and other systems.

The nodes are interconnected through five networks: a 3-dimensional torus network for point-to-point messaging between compute nodes, a global collective network for collective operations over the entire application, a global barrier and interrupt network, a gigabit Ethernet for machine control, and another gigabit Ethernet network for connection to other systems.

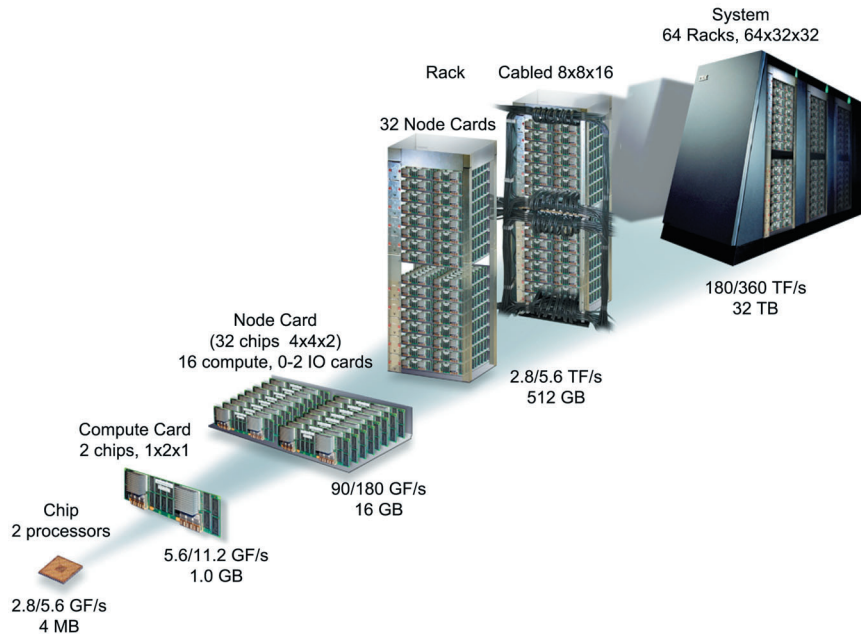
The 3D torus allows for each node to have low-latency, high-bandwidth interconnect with its six nearest neighbors and it supports general point-to-point communication. The torus network is particularly effective for applications with locality of communication. The global collective network is useful for speeding up commonly used MPI collective communications constructs. And the global barrier network quickly synchronizes state across all processors in the system.

Blue Gene also requires a service node where the system administrator manages the complex, front-end nodes where end-users compile and launch jobs, and file servers for storing data.

Encompassing numerous design innovations, Blue Gene can be scaled-up to previously unachievable levels of performance while staying within the practical limitations clients face with power consumption, thermal displacement and available floor space.

### **Familiar software environment tuned for Blue Gene**

Three fundamental principles were followed when the system software was designed for Blue Gene: simplicity, performance and familiarity. Driving toward simplicity in the software design has allowed development of software that takes advantage of hardware features to deliver high performance without compromising stability and security. And by creating a programming and administration environment based on



familiar programming languages, libraries, job management tools and parallel file systems, clients benefit from the innovative design elements of Blue Gene without facing a steep learning curve.

The front-end nodes of a Blue Gene complex are the portals through which programmers access the computational core of the system. The front-end nodes run a standard SUSE SLES9 Linux® distribution which provides a familiar platform from which users compile and debug programs and submit jobs.

Blue Gene systems are supported by standard IBM XL Fortran, C and C++ compilers for PowerPC that have been augmented with a backend that takes advantage of the dual floating-point unit that is unique to Blue Gene.

Programmers can employ the popular IBM Engineering and Scientific Subroutine Library (ESSL), a state-of-the-art collection of over 400 mathematical subroutines that provide optimum performance for floating-point engineering and scientific applications

written in Fortran, C or C++. Many of these routines have been tuned for the Blue Gene architecture.

In support of parallel programming, Blue Gene is offered with an MPI solution that leverages the MPICH2 library from Argonne National Laboratory to produce an implementation that exploits the communication technology of the compute nodes.

For job submission and workload balancing, IBM LoadLeveler® provides support for Blue Gene. LoadLeveler, which has been used for years by large-scale cluster clients, provides a facility for building, submitting and processing jobs, and is designed to match application processing needs with available resources.

In a Blue Gene environment, LoadLeveler coordinates with a special scheduler function that selects a set of compute nodes to form a partition that meets the size and shape requirements specified by the user.

Further enhancing the Blue Gene software environment is the inclusion of IBM General Parallel File System (GPFS). GPFS is a high-performance,

## Unsurpassed performance, ultrascalable computing

shared-disk file system that will provide fast data access from all nodes in a Blue Gene complex. Applications can readily access shared files using standard file system interfaces, and the same file can be accessed concurrently from multiple nodes.

The Blue Gene system also includes a service node where the system administrator manages the complex. The key functions available to the administrator are system configuration, initialization, monitoring and operation. These functions have been integrated into the IBM Cluster Systems Management product to provide additional management capabilities and a single point of control for management across the platforms supported by CSM. Also running on the service node is a DB2® relational database that is a repository for static and dynamic state information.

In summary, the Blue Gene system software is a collection of end-user components tuned for performance and reliability that will also be familiar to many existing cluster clients. In this way, clients will be able to become productive on Blue Gene right away while benefiting from its architectural features.

### Backed by IBM

Blue Gene is backed by one-year warranty package that covers all hardware and software components. Support is provided Monday through Friday 8 A.M. to 5 P.M. local time with a maximum two-hour response time objective from the local IBM support team. The hardware warranty covers three years of parts replacement. Customers may also purchase maintenance contracts for years following the first year of coverage supplied through the warranty.

Clients will be able to call IBM or go online to report Blue Gene problems. Technical documentation, descriptions of Blue Gene problem resolutions, how-to information and defect fix lists will be accessible via the Web. Optional on-site assistance will be available on a per-incident basis.

### Blue Gene—a commitment to Deep Computing

Blue Gene joins IBM's broad portfolio of Deep Computing solutions that includes POWER™ processor-based UNIX® symmetric multiprocessor (SMP) systems, Linux clusters, high-speed interconnects, storage, workstations, visualization solutions and an extensive collection of software tools. By leveraging the many choices, IBM has been the revenue marketshare leader in the industry for large-scale HPC computing every year since 1999.<sup>2</sup> The addition of an innovative solution like Blue Gene,

## Blue Gene at a glance

	Details	Benefits
<b>Processor</b>	PowerPC 440 700 MHz; two per node	Low power allows dense packaging; better processor-memory balance
<b>Memory</b>	512MB SDRAM-DDR per node	
<b>Networks</b>	1) 3D Torus - 175 MBps in each direction 2) Collective Network—350 MBps; 1.5 $\mu$ sec latency 3) Global Barrier/Interrupt 4) Gigabit Ethernet (I/O & connectivity) 5) Control (system boot, debug, monitoring)	Special networks speed up internode communications; designed for MPI programming constructs; improve systems management
<b>Compute nodes</b>	Dual processor; 1024 per rack	Double FPU improves performance
<b>I/O nodes</b>	Dual processor; 16, 32, 64 or 128 per rack	Facilitates job launch and I/O, raising efficiency of compute nodes
<b>Operating systems</b>	Compute Node—Lightweight proprietary kernel I/O Node—Embedded Linux Front End and Service Nodes—SUSE SLES 9 Linux	Kernel tailored to processor design; industry-standard distribution on front-end and service nodes preserves familiarity to end users and administrators
<b>Performance</b>	Peak performance per rack—5.73 teraflops Linpack performance per rack—4.71 teraflops	Highest available performance benefits capability customers
<b>Power</b>	27.6 kW power consumption per rack (maximum) 12 kW power consumption per rack (idle) 208 VAC 3-phase; 100 amp service per rack	Low power draw enables dense packaging
<b>Cooling</b>	Air conditioning 8 tons/rack (minimum) 2800 CFM (compute rack); 350 CFM (power supplies)	Low cooling requirements enable extreme scale-up
<b>Acoustics</b>	9.0 LwAD and 8.7 LwAm	
<b>Dimensions (includes air duct)</b>	Height—77" Width—36" Depth—36" Weight—1810 lbs. Service clearances—30" front and back Raised floor height—16" minimum	Design allows dense floor plan layout for better floor space utilization

accompanied by a product roadmap that stretches to petaflop performance before the end of the decade, demonstrates a significant commitment to the Deep Computing community and the researchers who are pursuing solutions to the important challenges facing humankind.

The Blue Gene project has progressed far in the past five years. Hands-on experience has clarified the extent to which Blue Gene can be applied to various computationally intensive workloads. Programs have been developed that expand the opportunities for researchers to access the power of Blue Gene. Software packages that have brought high value to IBM cluster solutions will be enabled for Blue Gene, helping to preserve the familiarity of the operating environment.

Collaborations with government laboratories, institutions of higher learning, applications and tools vendors, and industrial clients are today producing new insights into the potential of Blue Gene.

Since the formal launch in November 2004, the evidence is clear that Blue Gene has become a growing contributor to the advancement of science in a way that is cost-effective, resource-efficient and extremely scalable.

## For more information

To learn more about the IBM **@server**® Blue Gene Solution, please contact your IBM marketing representative or visit the following Web site:

**ibm.com/servers/deepcomputing/bluegene.html**

<sup>1</sup> Benchmark as of June 30, 2005 is available at <http://www.top500.org>. The Blue Gene results were submitted to TOP500 organization on June 10, 2005.

<sup>2</sup> IDC High-Performance Computer QView published quarterly. IBM has been number one in terms of annual revenue every year since 1999, in the Capability competitive segment, systems configured and purchased to solve the largest, most demanding problems.



© Copyright IBM Corporation 2005

IBM Corporation  
Integrated Marketing Communications  
Systems and Technology Group  
Route 100  
Somers, NY 10589

Produced in the United States  
August 2005  
All Rights Reserved

This publication was developed for products and/or services offered in the United States. IBM may not offer the products, features or services discussed in this publication in other countries.

The information may be subject to change without notice. Consult your local IBM business contact for information on the products, features and services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

IBM, the IBM logo, the e-business logo, **@server**, Blue Gene, DB2, LoadLeveler, POWER and PowerPC are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both. A full list of U.S. trademarks owned by IBM may be found at: **ibm.com/legal/copytrade.shtml**.

UNIX is a registered trademark of The Open Group in the United States, other countries, or both.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product and service names may be trademarks or service marks of others.

IBM hardware products are manufactured from new parts, or new and used parts. In some cases, the hardware product may not be new and may have been previously installed. Regardless, IBM warranty terms apply.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

Photographs show engineering and design models. Changes may be incorporated in production models.

Copying or downloading the images contained in this document is expressly prohibited without the written consent of IBM.