



**IBM Life Sciences Solutions:
Advancing research and discovery
through information technology.**



Discovering new possibilities

Dramatic advances occurring in the life sciences industry are changing the way we live. These advances fuel the rapid scientific discoveries in genomics, proteomics and molecular biology that serve as the basis for medical breakthroughs, the advent of personalized medicine, and the development of new drugs and treatments. The sequenced human genome has already increased the number of biological drug “targets” that can be explored from about 500 to over 30,000. Soon, the typical life sciences company will need to access and analyze “petabytes” (10^{15} bytes) of data to further their research efforts. In addition to the enormity of the data, there are challenges related to querying non-standard data formats, accessing data assets across global networks and securing data outside firewalls. The competitive advantage belongs to companies that can best use information technology (IT) solutions to capitalize on the opportunities presented by this transformation.

In response to these challenges, life sciences companies are redefining their research methodologies and retooling their IT infrastructures to position themselves for success in this new environment. The traditional trial-and-error approach is rapidly giving way to a more predictive science based on sophisticated laboratory automation and computer simulation. The technology used in newly emerging life sciences discovery models is critical to laboratory productivity and time to market. Key issues include:

- *Sharing and pooling information across global resources while maintaining security*
- *Retrieving and integrating diverse data across a variety of scientific domains*
- *Adding new data sources without new software development or complete redeployment of the solution*
- *Acquiring experimental data from industrial-style laboratory activities 24 hours a day, 7 days a week*
- *Enabling continuous realtime access to data without building and managing database warehouses*
- *Developing new ways to collaborate among research teams using shared research to focus efforts*

Add to these issues the need to work within existing laboratory and business computing environments, and the challenges facing today’s life sciences industry are almost overwhelming.

Flexible infrastructure solutions designed for future growth

End-to-end infrastructure solutions from IBM provide the scalable tools and systems to help life sciences companies access, manage and develop content. The IBM Life Sciences organization offers a comprehensive set of innovative IT infrastructure products and services designed to create value for drug discovery, pharmaceutical and biotech companies, healthcare organizations and academic research institutions. IBM has award-winning Web application servers, advanced computer cluster offerings and powerful database technologies that can help reduce the time and effort required to capture, compile and analyze research data for drug discovery, product development and Web-based clinical trials. In addition, IBM has the experience and the storage, security and systems management products needed to run mission-critical scientific research projects.

Trained professionals from the IBM Global Life Sciences Consulting and Solutions practice can help provide a wide range of consulting, implementation, outsourcing and hosting services to help improve the efficiency of R&D cycles, enhance collaboration within research communities and ensure the successful implementation of life sciences solutions. IBM is developing strategic business partnerships with leading bioinformatics companies to develop end-to-end specialized solutions for the life sciences industry, such as specialized laboratory applications.

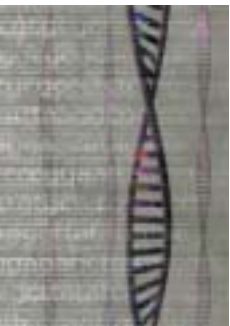


IBM solutions for knowledge management, data integration, high-performance computing, storage, e-business and consulting services deliver the powerful key capabilities needed in life sciences laboratories:

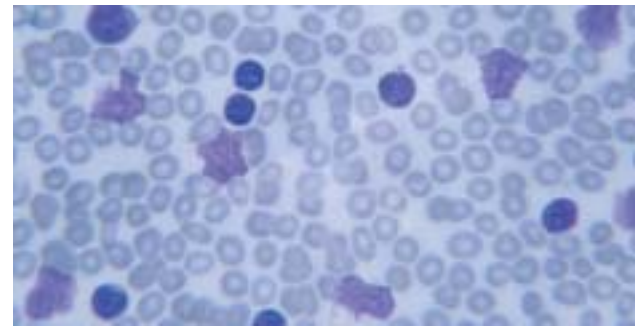
- *Knowledge management tools for transforming life sciences data into knowledge*
- *Data integration for extracting information and identifying patterns from multiple data sources and across diverse data domains*
- *High-performance computing for computational modeling, simulation and visualization*
- *Industry-leading, supercomputing performance for scientific workloads, including genome sequencing, protein structure sequencing and drug target identification*
- *Storage and retrieval technologies and tools for managing data easily*

- *Secure e-business technologies and services for accessing, sharing and processing information across the Internet*
- *Consulting services for assessing requirements and supporting solution design and implementation*
- *Security and data management to help protect the privacy of research data*

The life sciences industry needs flexible, scalable, reliable systems that can easily adapt as needs change. And IBM has an extensive portfolio of proven IT solutions that meet those needs. IBM offers a full range of infrastructure components, including system management and security software, storage technologies, and an impressive selection of high performance servers, such as IBM [^] models.



“...infrastructure solutions from IBM provide scalable tools and systems to help life sciences companies access, manage, and develop content.”



Leveraging the value of data: high-performance computing

An important, fast-growing business within the life sciences industry is focused on the compilation of genomic information into databases and the sale of information through subscriptions to drug companies and biotech research institutions. To help identify and analyze patterns within genetic data for viability as diagnostics and pharmaceutical products, drug discovery companies need these powerful, high-performance solutions.

High-speed, high-performance computing power and industrial-strength databases perform a wide range of data-intensive computing functions. Mapping genetic and proteomic information. Data mining to identify patterns and similarities. Text mining using huge libraries of information. These activities require high-speed computer infrastructures with well-integrated storage systems.

IBM delivers complete solutions for searching vast quantities of genomic data from a large number of sources and executing thousands of jobs simultaneously. A wide range of server solutions enable drug discovery companies and biotechnical researchers to leverage the value of the data while maintaining control over the analysis phase of the process.

Using Linux® technology, researchers can formulate complex queries, search multiple genomic databases, create, submit and execute computational jobs and manage the computing server resources. Because the Linux operating system is not tied to any single architecture or proprietary set of tools, it is singularly well-suited for the development and deployment of applications in a heterogeneous data environment. IBM offers the broadest selection of Linux-enabled software and services available to ensure secure Linux application development environments for the life sciences.

“IBM high-performance computing incorporates the processing power and deep computing technology necessary to help solve the most complex challenges of science.”



IBM servers for Life Sciences Solutions

Server	Capabilities	Benefits
IBM ^ pSeries	<ul style="list-style-type: none"> • Large-scale server with world-leading IBM RISC System/6000® and UNIX® performance • Industry-leading clustering advantages • Microsoft® Windows NT® and UNIX workload interoperability • Industry-leading storage area network solutions 	<ul style="list-style-type: none"> • Large symmetric multiprocessor scalability • Flexible horizontal growth
IBM ^ xSeries	<ul style="list-style-type: none"> • Rack-optimized Intel® technology-based servers • Scalable • Provides an infrastructure for e-business applications 	<ul style="list-style-type: none"> • Easy to cluster • Support for dynamic e-business data and transaction workloads • Low cost of ownership
IBM RS/6000® Scalable POWERparallel Systems® (SP™)	<ul style="list-style-type: none"> • Designed with speed, scalability, flexibility, manageability and availability • Can handle demanding technical and commercial workloads without disrupting critical operations • Versatile, distributed memory, multinode server can support up to 512 nodes 	<ul style="list-style-type: none"> • As a large scale server, provides a central point-of-control for serial, symmetric multiprocessor (SMP) and parallel workloads • Simplified system management
IBM ^ zSeries	<ul style="list-style-type: none"> • Large-scale server is the first e-business enterprise server • Traditional IBM computing strengths of scalability, flexibility, high availability 	<ul style="list-style-type: none"> • Support for dynamic e-business workloads for collaborative online research • Low cost of ownership

IBM AIX®, IBM ^ and IBM RS/6000 Scalable POWERparallel Systems (SP) solutions offer the flexibility and reliability required to handle mission-critical and data-intensive applications as well as the scalability needed to handle large workload demands today and tomorrow.

DiscoveryLink: Solving the data integration dilemma

Today's life sciences businesses require solutions with the flexibility to adapt and extend mission-critical applications to meet customer demands as well as the stability to smoothly absorb these changes. IBM's data integration strategy provides super computers, software and services to enable successful research and development in life sciences laboratories. The DiscoveryLink™ solution from IBM Life Sciences includes the combined resources of DiscoveryLink middleware and IBM Life Sciences Services. Using this versatile new software, IBM Life Sciences Services can create new components that allow specialized databases—for proteomics, genomics, combinatorial chemistry or high-throughput screening—to be accessed and integrated quickly and easily.

With single-query data access, the IBM DiscoveryLink solution allows researchers to extract information from large volumes of heterogeneous life sciences research and clinical data sources including the integration of important legacy data. The DiscoveryLink solution includes the DB2® Life Sciences DataConnect and DB2 RelationalConnect set of software tailored specifically to life sciences research and development requirements for integrating data from multiple sources. DB2 Life Sciences DataConnect provides the interfaces (wrappers) through which the database communicates with the federated data sources. DB2 RelationalConnect provides the robust, scalable communication structure for connecting to relational data sources, such as Oracle, and providing secure access to legacy data.



“The DiscoveryLink solution combines innovative middleware and integration services to extract information from large volumes of heterogeneous life sciences and clinical data sources.”



IBM systems and systemware for Life Sciences Solutions

Operating system	Capabilities	Benefits
IBM AIX	<ul style="list-style-type: none"> • Top-rated UNIX operating system • Can handle more data and users efficiently — across e-business, business intelligence, enterprise resource planning and other critical business applications • Runs across the entire range of RS/6000 systems • IBM AIX/L[®] Version 5.1 runs Linux applications 	<ul style="list-style-type: none"> • Extensive cluster management utilities for SP systems • Best-of-breed system and network security • Scalability — across systems of different sizes, across different technologies and hardware platforms
Linux	<ul style="list-style-type: none"> • Complements open architecture, cross-platform application framework, Linux-enabled software and servers (including ^ xSeries, ^ zSeries and ^ pSeries) • Runs approximately 500 Linux applications 	<ul style="list-style-type: none"> • Choice of servers • Top performance and flexibility • Support for dynamic e-business and transaction workloads
Systemware and tools	Capabilities	Benefits
IBM High Availability Cluster Multi-Processing (HACMP)	<ul style="list-style-type: none"> • An industry-leading, high-availability solution for mission-critical applications, running on the RS/6000 platform 	<ul style="list-style-type: none"> • Systems management capabilities for monitoring all RS/6000 servers and other components within the cluster, from a single system • Flexibility and reliability
IBM Engineering and Scientific Subroutine Library (ESSL)	<ul style="list-style-type: none"> • A library of mathematical subroutines that provides high-performance mathematical functions for scientific and engineering applications 	<ul style="list-style-type: none"> • ESSL for AIX contains over 400 high-performance mathematical subroutines tuned to RS/6000 hardware
IBM Mathematical Acceleration SubSystem (MASS)	<ul style="list-style-type: none"> • A library of tuned mathematical intrinsic functions that improve performance for many technical applications 	<ul style="list-style-type: none"> • Improves application performance
IBM Parallel System Support Programs for AIX (PSSP)	<ul style="list-style-type: none"> • A collection of administrative and operational software applications, running on each node of an RS/6000 SP system and SP-attached servers • Single point of control 	<ul style="list-style-type: none"> • Scalability across the range of system configurations • RS/6000 cluster technology for event monitoring and recovery coordination of SP systems and their environments
IBM LoadLeveler	<ul style="list-style-type: none"> • A distributed, network-wide, job management program for dynamically scheduling work on IBM and non-IBM processors 	<ul style="list-style-type: none"> • Balance workloads for efficient management and distribution of job flow in the network and across all LoadLeveler-defined hardware resources
IBM General Parallel File System (GPFS)	<ul style="list-style-type: none"> • Provides service to parallel and serial applications running on AIX and Linux 	<ul style="list-style-type: none"> • High-performance, distributed file systems meet the demands of scientific research • Allows scalable access to unlimited file sizes and clients

DiscoveryLink is unique among existing systems because it enables easy creation of wrappers for nonrelational sources and provides the capability to add new sources dynamically. It also includes query optimization technology that automatically searches for the most efficient means of executing a query and assembling the results. IBM DB2® Universal Database™, the industry's first multimedia, Web-ready, federated database, provides the industry-leading performance and scalability required to drive the most demanding life sciences applications. Although it is built on DB2 Universal Database technology, DiscoveryLink is designed to be used along with most other databases. Functioning as the critical bridge between R&D applications and diverse data sources, DiscoveryLink delivers a single-table virtual database federation of multiple heterogeneous data sources. While these data sources appear as a single format to the applications and end users, the functionality, integrity, location and form of the original data remains unchanged.

With its innovative database technology and multiplatform hardware support, the DB2 DiscoveryLink technology can complement and extend the capabilities of existing data warehouses and object frameworks. This enables drug discovery, biotechnology and pharmaceutical research companies to boost productivity while protecting their IT investments. DiscoveryLink allows integration of multiple diverse sources, including text search engines and industry-specific repositories to make data management and application development easy.

Knowledge management: Turn research data into insight

Many drug discovery processes—including clinical trials—require maximum efficiency for data sharing and knowledge management functions like patient record mining across companies. IBM has the end-to-end open source infrastructure solutions to help optimize data sharing and information management.

IBM knowledge management tools for Life Sciences Solutions		
Tools	Capabilities	Benefits
Lotus® K-Station	<ul style="list-style-type: none"> With open, compliant, standards-based architecture provides single-point access to critical business information, collaboration functionality and a knowledge portal 	<ul style="list-style-type: none"> Increased online laboratory productivity between individuals and across global and organizational communities
IBM Text Knowledge Miner	<ul style="list-style-type: none"> Extracts information and insight from large collections of documents by search, analysis and classification 	<ul style="list-style-type: none"> Ability to quickly categorize and ascertain the factual content of document collections and explore relationships among selected categories
IBM Intelligent Miner™ for Text	<ul style="list-style-type: none"> Text analysis tools, full-text search engine and a Web search solution provide a wide variety of text mining applications 	<ul style="list-style-type: none"> Ability to extract information from unstructured information Ability to classify incoming e-mail by predefined topics in log files Automatic detection of deadlines

With IBM knowledge management software and services, laboratory research teams can capture, manage and share information and create new information relationships. Life sciences teams can build virtual research communities using secure Web portals for online research collaboration in realtime. And meet the security requirements of the entire R&D process, including privacy issues concerning genetic information of donors in research studies, patient records in clinical trials and confidential business information. IBM provides a full range of knowledge management infrastructure solutions and development tools to enhance the overall effectiveness of the research process.

Storage: Share centralized data

A key component of a high-quality life sciences solution is reliable, disaster-proof storage. IBM Web planning, designing and hosting services can help life sciences research teams develop custom solutions for managing clinical trials, personalized medicine and FDA approvals in a secure and stable environment. IBM storage hardware, software and services can help maximize laboratory productivity and minimize

operating costs. Researchers can store results obtained from collaborative research and data mining in “pools” of commonly shared knowledge that are administered from a centralized point. Laboratories can increase capacity without interruptions using these scalable storage systems and reduce backup time because only modified data needs to be transferred. And IBM consultants can design a flexible disaster-tolerance strategy—with no single point of failure—to accommodate laboratory data needs as they evolve.

e-business: Manage and share data in a secure, stable environment

R&D organizations can use IBM data integration and management tools to create easy-to-use Web portals to leverage the Internet for collaborative global life sciences research. Award-winning Web application servers, advanced computer cluster systems and powerful database technologies can help reduce the time and effort required to capture, compile and analyze research data for drug discovery, product development and Web-based clinical trials.



“Reliable, disaster-proof IBM storage hardware, software and services help maximize laboratory productivity.”



IBM storage for Life Sciences Solutions

Storage hardware	Capabilities	Benefits
IBM TotalStorage™ IP Storage 200i	<ul style="list-style-type: none"> High-performance, low-cost iSCSI storage using Internet protocols 	<ul style="list-style-type: none"> Scalability Low cost
IBM TotalStorage Networked Attached Storage 300G	<ul style="list-style-type: none"> A LAN/SAN gateway incorporating different storage system devices on the Storage Area Networks (SANs) to provide scalability to IP-based clients and servers Open system 	<ul style="list-style-type: none"> Interoperability Ability to leverage features and performance of SAN
Disk storage systems	Capabilities	Benefits
IBM 7133 Serial Disk System	<ul style="list-style-type: none"> Disk storage among systems High availability to safeguard data access 	<ul style="list-style-type: none"> Scalability Flexibility Investment protection
IBM Modular Storage Server	<ul style="list-style-type: none"> Scalable storage server for Fibre Channel SANs 	<ul style="list-style-type: none"> Easy to manage Flexibility
IBM Enterprise Storage Server	<ul style="list-style-type: none"> Storage pooling of disk media for multiple, heterogeneous systems High availability to safeguard data access 	<ul style="list-style-type: none"> Flexibility Scalability Investment protection
Tape and optical storage	Capabilities	Benefits
LTO Ultrium System	<ul style="list-style-type: none"> Standards-based, high performance Optimized for streaming environments High data transfer rates 	<ul style="list-style-type: none"> Scalability Investment protection
MagStar 3590 Tape Subsystem	<ul style="list-style-type: none"> Basic tape drive Multisystem connectivity 	<ul style="list-style-type: none"> High-capacity performance Reliability
MagStar 3494 Tape Library	<ul style="list-style-type: none"> Handles multiple MagStar 3590 drives to support large volumes of data 	<ul style="list-style-type: none"> Variety of connectivity options Scalability
MagStar Virtual Tape Server	<ul style="list-style-type: none"> Combines tape and disk storage for improved response times 	<ul style="list-style-type: none"> Improved efficiency Enhanced data protection
IBM Network Storage Server	<ul style="list-style-type: none"> Network-attached storage using tape media 	<ul style="list-style-type: none"> Centralized backup for network clients and servers
IBM General Parallel File System	<ul style="list-style-type: none"> A scalable, parallel file system that runs under AIX and Linux operating systems 	<ul style="list-style-type: none"> Scalability High availability Online system management, scientific computing, data mining, digital library and scalable network file servers
Tivoli® Storage Manager	<ul style="list-style-type: none"> High-performance adaptive differencing technology and LAN-free client data transfer provides backup/restore, archive and hierarchical file management 	<ul style="list-style-type: none"> Scalability—spans palmtops to mainframes over 35 platforms High-speed capabilities

IBM Web tools for Life Sciences Solutions

Web middleware	Capabilities	Benefits
IBM WebSphere® Application Server	<ul style="list-style-type: none">• Provides a robust multiplatform Java™ technology-based Web application environment• Combines the control and portability of server-side applications with Java technologies	<ul style="list-style-type: none">• Fast development, deployment and integration of e-business applications• Easy-to-use single Web-based interface
IBM WebSphere Portal Server	<ul style="list-style-type: none">• Scalable, single point-of-access for data, people and applications• Permits federated search against all forms of data	<ul style="list-style-type: none">• Ease of use• Ability to leverage Web site tools
IBM MQSeries® Integrator	<ul style="list-style-type: none">• Market's leading message-queuing product applies intelligent routing to seamlessly integrate applications, databases and networks	<ul style="list-style-type: none">• Rapid application deployment• Scalability



“Create easy-to-use Web portals to harness the power of the Internet for collaborative, global research.”



IBM WebSphere Application Server is an e-business application deployment environment built on open, standards-based technology. The WebSphere software platform provides the foundation for building and expanding a life sciences e-business infrastructure on the Web. With a stable Web foundation life sciences researchers can access multiple online data sources, personalize data queries and ensure the secure management of research and medical information.

Services and consulting: Optimize your solution

IBM has the skills, resources and infrastructure management support to meet the laboratory and business needs of the life sciences industry. IBM specialists can provide consulting, systems integration, strategic outsourcing services and hosting services. Along with data management, knowledge management and e-business solutions that work with existing systems and software. As technology needs change and grow, IBM consultants can help optimize systems to accommodate every level of expansion.



IBM is working directly and in concert with IBM Business Partners to provide highly effective approaches to IT management and strategy that are specifically tailored for life sciences companies—including data integration query management, data mining and data interpretation—using advanced algorithms and deep computing technologies.

IBM's commitment to life sciences

The IBM commitment to the life sciences industry is defined by the establishment of the specialized IBM Life Sciences Solutions Group dedicated to rapidly bringing leading-edge technology out of the laboratory and into the marketplace for customers and Business Partners in the fields of pharmaceutical research, biotechnology, genomics, proteomics, health and other life sciences. In addition, a dedicated IBM Global Life Sciences Consulting and Solutions practice has been established to focus IBM service capabilities and expertise, as well as intellectual capital on helping customers and Business Partners to migrate their R&D units into even more efficient and competitive operations.

IBM Research continues to pursue strategic exploration of technologies applicable to life sciences research and product development. Pattern discovery and matching functional and structural genomics and proteomics. Visualization, technical knowledge management and nanotechnology. The IBM Computational Biology Center and the IBM Deep Computing Institute are two key research centers housing teams of scientists working on projects involving computational biology, chemistry and material science. The long term projects at the IBM Computational Biology Center foster IBM collaboration with life sciences companies to bring scientific expertise directly into the development of life sciences solutions. Detailed studies are in progress aimed at providing new clues for medical diagnostics, the synthesis and design of novel materials and the analysis of genes and their relationships.

The IBM Deep Computing Institute provides business decision-making capabilities to analyze and develop solutions for complex and difficult problems. Combining these capabilities with advances in algorithms, analytic methods, modeling and simulation, data management and software infrastructures enables valuable scientific, engineering and business solutions. Joining life sciences research with information technology, IBM is uniquely positioned to help the industry solve the challenges of R&D information management.

IBM is actively pursuing strategic business partnerships with life sciences companies whose complementary skills, knowledge and resources can help build value-rich solutions for the industry.

There are extraordinary challenges and opportunities ahead for the companies engaged in the life sciences industry. The scientific challenges in this emerging industry are matched equally by the challenges associated with managing data integration and developing the computing technology and tools needed to provide solutions for the laboratory.

Combined with IBM's core strength in providing robust technology and global services, the IBM Life Sciences business unit delivers innovative, scalable, infrastructure solutions that are unique within the industry.

For more information

To learn more about IBM Life Sciences solutions, visit our Web site at ibm.com/solutions/lifesciences. Or contact an IBM Life Sciences specialist at ls@us.ibm.com.



“Joining life sciences research with information technology, IBM is uniquely positioned to help solve the challenges of R&D information management.”

Why IBM?

Today, IBM systems include 215 of the world's 500 most powerful high-performance computers; the world's fastest supercomputer (ASCI White, 12.3 trillion calculations/sec); the world's largest established database (116 terabits); advanced storage management; and a world-renowned computational biology center. IBM announced a \$100 million initiative to build what promises to be the world's most powerful supercomputer—"Blue Gene"—which will be capable of more than one quadrillion operations per second (one "petaflop").

The Blue Gene project is an IBM Research project with three goals:

- *Explore the challenges associated with cellular hardware architectures for massively parallel computers.*
- *Conduct research into software application development of a computer system with interprocessor communications capabilities necessary for optimized performance.*
- *Exploit the Blue Gene supercomputer to address the important problems of protein science, in particular, the phenomenon of protein folding, but ultimately, specific issues such as drug-protein interaction.*

IBM is committed to bringing powerful products and expertise to help researchers and scientists speed their processes and optimize their results.





© Copyright IBM Corporation 2001

IBM Corporation
Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
05-01
All Rights Reserved

AIX, AIX/L, DB2, DB2 Universal Database, DiscoveryLink, IBM, the IBM logo, Intelligent Miner, MQSeries, pSeries, RISC System/6000, RS/6000, Scalable POWERparallel Systems, SP, TotalStorage, WebSphere, xSeries and zSeries are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Lotus is a trademark or registered trademark of Lotus Development Corporation in the United States, other countries or both.

Tivoli is a trademark or registered trademark of Tivoli Systems, Inc. in the United States, other countries or both.

Intel is a trademark of the Intel Corporation in the United States, other countries or both.

Microsoft and Windows NT are trademarks or registered trademarks of Microsoft Corporation in the United States, other countries or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries or both.

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

Linux is a registered trademark of Linus Torvalds.

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

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.



Printed in the United States on recycled paper containing 10% recovered post-consumer fiber.



G580-0475-02