

IBM Grid Offering for Analytics Acceleration: Agricultural Chemicals



Highlights

- *Helps improve the use of computing and storage resources across an organization*
- *Accelerates time-to-results and increases the number of calculations processed*
- *Offers an affordable, flexible IT environment for fast, large-scale analysis*

The challenges facing agricultural chemical analysis

Dramatic advances in biotechnology are changing the way we live—from the food we eat to the clothes we wear. These advances fuel the rapid scientific discoveries in genomics, proteomics and molecular biology that serve as the basis for creating crop protection chemicals, high-value seeds that can be adapted for various growing conditions, greater produce yields and, ultimately, higher quality produce for consumers around the globe. To help ensure that these advances continue, agri-science companies are constantly adapting to embrace new data volumes and requirements, emerging IT approaches and innovations in

distributed computing. The industry also must continue to meet several challenges: global competition, strict environmental regulations and changing distribution patterns.

To facilitate bioinformatics and molecular modeling, companies in the agricultural chemical industry use a range of tools, including open source codes, in-house tools and applications from independent software vendors (ISVs). These applications may need to run on a heterogeneous infrastructure of PCs, workstations, clusters and high-performance computing systems. Modelers often swap between high-throughput/capacity calculations and high-precision/capability calculations.

Today's agri-science environment requires quick turnaround for large volumes of calculations and simulations so that companies can compete in a fast-changing world. Therefore, the IT environment must be robust, adaptable, flexible and extensible.

Introducing Grid computing

Grid computing is a services-oriented architectural approach that uses



Grid Computing

open standards to enable distributed computing over the Internet, a private network, or both. This approach can help agricultural chemical companies aggregate disparate IT elements such as compute resources, data storage and filing systems to create a single, unified system and address fluctuating workload requirements.

At its core, Grid computing enables devices—regardless of their operating characteristics—to be virtually shared, managed and accessed across an enterprise, industry or workgroup. Although the physical resources that compose a Grid may reside in multiple locations, users have seamless and uninterrupted access to these resources. This resource virtualization provides the necessary access, data and processing power to rapidly solve complex business problems, conduct compute-intensive research and data analysis and engage in real-time business on demand. Grid computing helps to promote the efficient utilization of technology resources and fosters the creation of cost-effective, resilient IT infrastructures that are adaptable to change.

Grid computing and analytics acceleration in agricultural chemicals

In response to the challenges of the industry, agricultural chemical companies are redefining their research methodologies and retooling their IT infrastructures to position themselves for success. Grid technologies can greatly accelerate the execution of sequence comparison algorithms—such as BLAST, FastA, CLUSTW, Smith-Waterman and HMMR—that are used in bioinformatics and molecular modeling.

For example, the size of data volumes being generated in bioinformatics is growing at an exponential rate. With the growth of the dataset size, computations that mine these large datasets for information are taking longer, thus decreasing productivity and inhibiting the pace of research dependent on this productivity. This large data size coupled with a spiky demand function—in which the demand for large compute power cycles occurs in spurts—makes the case for aggregating compute resources and then sharing them as a large, virtual resource in a systematic and secure compute Grid.

Depending on the customer's analytics applications, flexibility requirements and existing infrastructure, two other techniques can be used in combination with Grid computing to facilitate analytics acceleration: application parallelization and hardware acceleration. Application parallelization divides an application into independent tasks that can then be executed concurrently, thereby reducing the time to completion. Hardware acceleration applies additional processing power by establishing or growing a cluster solution, or by using faster processors.

IBM Grid Offering for Analytics Acceleration

IBM has a comprehensive approach to help determine the most appropriate combination of technologies for bioinformatics and molecular modeling acceleration in the agricultural chemical industry. Award-winning WebSphere® Web application servers, advanced computer cluster offerings, powerful database technologies, application integration approaches and knowledge management techniques from IBM can help reduce the time and effort required to capture, compile and analyze research data and can

accelerate product development. IBM Grid Offerings include a Grid Innovation Workshop, a pilot implementation and a complete portfolio of IBM Global Services offerings.

IBM product foundation: supported platforms for Grid computing

IBM offers several products to build a Grid computing platform for analytics acceleration, including IBM **@server**[™] pSeries[™] and zSeries[®] servers, the IBM **@server** Cluster 1600 and the IBM **@server** Cluster 1350.

The innovative pSeries servers are designed to lower costs, improve efficiency and speed the transformation to a Grid-enabled infrastructure. The zSeries can provide on demand virtual node and hardware capacity, controlled access to enterprise data, increased resiliency to the Grid infrastructure and development tools for rapid Grid application prototyping. The Cluster 1600 brings together cluster-ready pSeries servers, cluster management software and a choice of IBM or industry-standard cluster interconnect technologies. The Cluster 1350 combines the power of rack-optimized IBM **@server** xSeries[®] servers with

IBM Cluster Systems Management for Linux, optional IBM storage products and leading third-party networking components to create powerful, flexible and secure solutions for high-performance computing and commercial application environments.

IBM product foundation: Grid-enabling software

The open source Globus Toolkit[™], developed through The Globus Project[™], includes a full range of Grid-based commands and programming tools to quickly develop Grid applications or to run existing applications in a Grid environment. It provides the underlying elements for Grid security, communication, information infrastructure, resource management and portability.

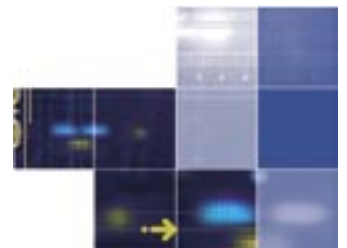
The IBM Grid Toolbox is based on the Globus Toolkit, but also includes enhanced documentation, installation tools for AIX[®] and Linux[®] environments and optional IBM service and support offerings. The Toolbox can be deployed on various IBM **@server** products—the pSeries, xSeries, zSeries and iSeries[™] servers—and on the IBM Cluster 1600 and Cluster 1350 platforms.

See Figure A, which outlines available technologies and IBM software that enable Grid computing for analytics acceleration.

Parallelization software

Application parallelization can facilitate Grid computing when used for analytics acceleration. IBM uses industry-standard messaging paradigms, such as the Message Passing Interface (MPI) and Parallel Virtual Machine (PVM), to help parallelize your applications.

Clearly, parallelizing the application is just the first step—running the code on a distributed set of virtualized resources is required to significantly increase the computational throughput. IBM's relationships with Platform Computing, DataSynapse and United Devices allow customers to use a single product for application parallelization and distributed execution. These companies offer products that provide job- and task-level scheduling for low-overhead job submission as well as for interactive computations. IBM also offers job- and task-level schedulers, along with the Globus Toolkit, to support commercial application environments.





Grid Computing

ISV Grid technologies

IBM has established strong relationships with Grid middleware ISVs. The technologies provided by these ISVs offer innovative solutions to some of today's complex business problems. IBM has assessed these technologies in detail, helping to ensure that they adopt appropriate standards and provide interoperability, scalability, security and resilience.

IBM services to support Grid technologies

Grid computing can be adapted to serve unique business environments, and IBM Global Services (IGS) support all elements of this architecture—servers, operating systems, middleware and networks—with a full range of Grid-related services. For example, IGS can provide a business-case analysis, cost justification and a strategic perspective for leveraging Grid computing across the enterprise. For other customers, IGS may review potential applications, select the appropriate technologies

and help enable the selected applications to use Grid computing effectively.

IGS also can design and deploy Grid infrastructures that can help provide outstanding performance, scalability, security, resiliency and reduced operational costs. Or, if you need an end-to-end Grid solution, IGS can integrate all of these capabilities into a coordinated approach that delivers a comprehensive Grid implementation.

Furthermore, IBM's array of strategic relationships with Grid technology developers allows you to review and select the appropriate technologies. By offering these options and capabilities, IGS can provide you with a low-risk, efficient and effective Grid solution.

IBM as a leader in Grid computing

IBM is committed to open standards and is working with The Globus Project open source development community, the Global Grid Forum (GGF) and the Interoperable Informatics Infrastructure

Consortium (I3C) to promote the adoption of open standards and accelerate the availability of Grid technology.

IBM is a strong supporter of The Globus Project, a multi-institutional research and development effort to address the technical and business challenges of Grid computing. Founded by a team of technicians and researchers, The Globus Project has defined an open source Grid reference architecture and a set of tools to assist Grid deployment.

IBM Design Centers use the latest Grid technologies, including the Open Grid Services Architecture (OGSA), which merges the open protocols used for Grid computing with the protocols used for Web services. IBM Design Centers provide access to the latest software from leading Grid software companies, such as United Devices and Platform Computing, and the latest open source Grid technologies from The Globus Project.

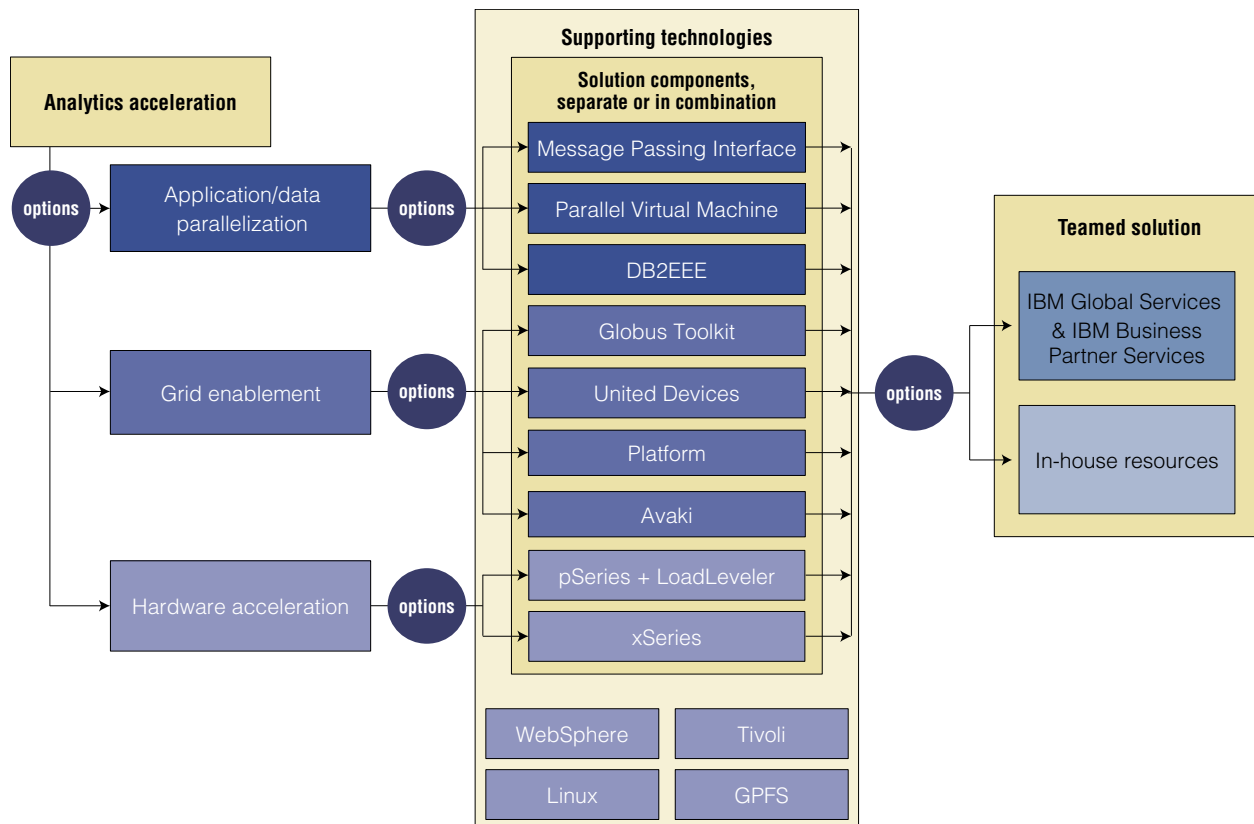


Figure A. Analytics acceleration decision landscape





Grid Computing

Figure B. Engagement points for Grid enablement

Grid Innovation Workshop for Agricultural Chemicals	Grid Pilot Implementation Services	Grid Rollout Implementation Services
<p>IBM can help you decide which option is right for you. When appropriate, IBM Global Services Grid Innovation Workshop can help to develop the vision, plan and execution model for your Grid project.</p>	<p>IBM can help define a pilot implementation project that will help you to validate the benefits of Grid computing. Piloting services include:</p> <ul style="list-style-type: none"> • Introduction, education, installation and support of the Globus Toolkit • Selection of Grid middleware partners relevant to your needs • Training on Grid concepts and implementation techniques • Creation of a test Grid • Access to public data sources • Projection of the benefits of a full Grid implementation • Timely and significant analysis of the data Grid technology 	<p>IBM Global Services can assist with your Grid implementation project by providing an array of implementation services such as:</p> <ul style="list-style-type: none"> • Project management support • Business-process consulting services and training • Help desk • Data integration services through federation • Portal development • Linux implementation services • High-performance clusters • Engagement services • IBM Security Services • Systems integration services

Figure C. Possible technical components: IBM Grid Offering for Analytics Acceleration in Agricultural Chemicals

Component	Features/Benefits
PARALLELIZATION	
Message Passing Interface (MPI)	<ul style="list-style-type: none"> • Uses open standards and helps ensure portability over many platforms • Provides point-to-point communication, collective communication, virtual process topologies and support for writing heterogeneous parallel applications • Can be attained as either a shareware or vendor-supplied implementation • Can provide significant flexibility in performance tuning and programming or in a message-passing environment
Parallel Virtual Machine (PVM)	<ul style="list-style-type: none"> • Uses open source software • Can be used with heterogeneous systems (running UNIX® or Microsoft® Windows® operating systems) • Permits systems to be coupled together across a network and used as a single, large parallel computer
GRID	
Globus Toolkit/IBM Grid Toolbox	<ul style="list-style-type: none"> • Facilitates quick development of Grid applications • Can be used with various IBM @server products
IBM General Parallel File System (GPFS)	<ul style="list-style-type: none"> • Provides shared access to files across multiple disk drives on multiple nodes • Provides a common file system abstraction for data shared among all the nodes in a Grid • Allows applications to easily access files using standard POSIX (Portable Operating System Interface for UNIX) file system interfaces • Enables parallel applications to simultaneously access either the same or different files
SOFTWARE	
IBM WebSphere platform	<ul style="list-style-type: none"> • Provides infrastructure software for dynamic e-business. Founded on open standards such as J2EE and Web services, WebSphere has evolved from a Web application server to a full set of related products and offerings with a common base. IBM expects WebSphere to be the hosting environment for customers wishing to exploit the Open Grid Services Architecture • Delivers value in three areas: <ul style="list-style-type: none"> – Reach and user experience. Enhances customer loyalty by delivering a rich user experience through offerings such as WebSphere Portal – Business integration. Helps slash process cycle time and costs, and thus can help improve return on investment (ROI), through offerings such as WebSphere MQ – Foundation and tools. Provides a reliable, scalable and flexible infrastructure for dynamic e-business through offerings such as WebSphere Application Server and WebSphere Studio





Grid Computing

Component	Features/Benefits
SOFTWARE	
IBM Tivoli®	<ul style="list-style-type: none"> • Provides highly scalable, cross-platform systems management • Enables efficient management of e-business infrastructures from a single location • Can help maximize ROI and speed deployment • Can simplify management of separate devices without sacrificing productivity, security or performance
ISV Grid technologies	• Offer innovative solutions to today's business problems
Linux	• Provides an open source code base for Grid development efforts
IBM DB2 Universal Database™	• Provides shared-nothing architecture in which each machine in a cluster has exclusive access to its own disks and memory and does not compete with other machines for resources
Enterprise—Extended Edition (DB2EEE)	• Facilitates complex parallel processing operations on platforms such as AIX, HP-UX, Microsoft Windows NT®, Linux and Sun Solaris
IBM LoadLeveler®	<ul style="list-style-type: none"> • Provides a high-performance computing cluster job scheduler for xSeries and pSeries servers • Enables checkpoint restart and backfill scheduling
HARDWARE	
IBM @server pSeries	<ul style="list-style-type: none"> • Can help to lower costs, improve efficiency and speed transformation to a Grid-enabled infrastructure • pSeries 690: Features 1.1 or 1.3GHz POWER4™ processors, the first “server-on-a-chip” design for high-end UNIX servers, with up to 32 processors per server; delivers true logical partitioning; can be divided into as many as 16 virtual servers, each with its own set of system resources such as processors, memory and I/O • pSeries 670: Features 1.1 or 1.3GHz POWER4 processors, with up to 16 processors per server and the same logical partitioning capabilities as the pSeries 690 • pSeries 660: Features a range of three types of powerful, midrange, rack-mounted servers; offers optimal flexibility and outstanding power in a small footprint • pSeries 630: Features up to four 1.45GHz POWER4+™ processors; offers the lowest entry price for a pSeries four-way cluster building block

Component	Features/Benefits
HARDWARE	
IBM @server zSeries	<ul style="list-style-type: none"> • Dynamic server provisioning. Dynamically deploys and configures new virtual Grid servers; allows enterprises to tap zSeries capacity to address new business opportunities; includes a customer-initiated upgrade feature that can enable customers to add hardware capacity through a simple online interface • High-availability features. Increase the service level for commercial Grids—while facilitating high-speed access to mission-critical enterprise data and application resources—by housing critical Grid scheduler and directory software in the zSeries environment; includes Geographically Dispersed Parallel Sysplex™ (GDPS®), which can maintain mission-critical data for both Linux and z/OS™ systems, helping to protect critical Grid data resources • z/VM® virtualization technology. Supports the creation of up to hundreds of virtual Linux-based Grid nodes on a single server; can help develop, test, service and create prototypes for Grid computing applications and infrastructures; provides a fast-track, proof-of-concept sandbox to minimize startup costs while accelerating time to market of new applications • Server consolidation capability. Offers cost advantages, particularly for customers seeking to speed ROI by consolidating distributed servers on the zSeries • Grid data repository. Leverages reliable database management systems with security-rich access to shared data; allows localized data to be accessed among virtualized Grid nodes using HiperSockets memory-speed bandwidth with near-zero latency; can optimize parallel processing of terabyte databases for intensive data searching on GDPS
IBM @server Cluster 1600	<ul style="list-style-type: none"> • Provides a highly scalable AIX 5L™ or Linux cluster and Grid computing solution for large-scale computational modeling, multi-terabyte databases and cost-effective data center, server and workload consolidation • Brings together cluster-ready pSeries servers, cluster management software and a choice of IBM or industry-standard cluster interconnect technologies • Extends the benefits of IBM RS/6000® SP™ technology to additional hardware building blocks • Can be built using the latest, most powerful pSeries servers, including the pSeries 690, 670, 655 and 630 servers, to deliver high-performance computing solutions and reduce complexity and total cost of ownership (TCO) for managing multiple systems • Provides a choice of Parallel System Support Programs (PSSP) or Cluster Systems Management (CSM) for single-point-of-management control • Provides a choice of industry-standard (10/100Mbps or 1Gbps Ethernet), SP Switch or SP Switch2 interconnect technologies • Includes the AIX 5L operating system and a complete cluster software suite for high-end technical computing • Includes GPFS, a high-performance POSIX file system that provides fast data access across the cluster environment • Includes High Availability Cluster Multiprocessing (HACMP) for AIX 5L, which is designed to keep business-critical applications and systems operational 24x7. HACMP can automatically detect system and application failures; provide the capability to recover system hardware, applications, data and users while keeping recovery time to an absolute minimum; and enable hardware and software upgrades without interrupting cluster operations





Grid Computing

Component	Features/Benefits
HARDWARE	
IBM @server Cluster 1350	<ul style="list-style-type: none">• Runs on the Linux operating system• Can operate in a standalone environment for dedicated application workloads or in a Grid computing environment for high-throughput resources and high performance• Combines rack-optimized IBM @server xSeries servers, CSM for Linux, optional IBM storage products and leading third-party networking components to create powerful, flexible solutions for high-performance computing and commercial application environments• Can be used in a wide range of configurations, including single-rack systems of up to 32 cluster nodes and large-scale systems of up to 512 cluster nodes (larger configurations are possible through special order)• Uses xSeries 335 or xSeries 345 as the cluster nodes: one- or two-way x335 servers (1U), with Intel® Xeon™ processors running at either 2.4 or 2.8GHz; may include up to 32 one- or two-way x345 storage servers (2U), with Intel Xeon processors running at 2.4 or 2.8GHz. Each cluster requires a two-way x345 management node• Includes GPFS, a high-performance POSIX file system that provides fast data access across the cluster environment• Can be managed by CSM for Linux Version 1.2 or 1.3, which provides resource monitoring, automated operations, remote hardware control and command execution, configuration file management and parallel network installation; CSM for Linux helps ease administration and may reduce life-cycle costs by allowing management of an entire Cluster 1350 system from a single management node
SERVICES	
IBM Global Services	<ul style="list-style-type: none">• Support all elements of Grid computing with a full range of services• Allow customers to review and select the appropriate technologies

For more information

To learn more about Grid computing for analytics acceleration in agricultural chemicals, contact your IBM representative or IBM Business Partner, or visit ibm.com/grid.

AMERICAS

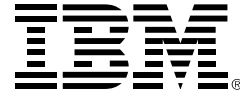
Basu Hurkadli
e-mail: bnhurka@us.ibm.com

ASIA PACIFIC

Takayuki Takano
e-mail: ttakano@jp.ibm.com

EUROPE/MIDDLE EAST/AFRICA

Philippe Bricard
e-mail: bricard@fr.ibm.com



© Copyright IBM Corporation 2003

IBM Systems Group
Route 100
Somers, NY 10589

Produced in the United States
04-03
All Rights Reserved

IBM, the IBM logo, the e-business logo, AIX, AIX 5L, DB2 Universal Database, eServer, GDPS, Geographically Dispersed Parallel Sysplex, iSeries, LoadLeveler, POWER4, POWER4+, pSeries, RS/6000, SP, Tivoli, WebSphere, xSeries, z/OS, z/VM and zSeries are trademarks or registered trademarks of International Business Corporation in the United States, other countries, or both.

Globus Toolkit and The Globus Project are trademarks of the University of Chicago.

Intel and Xeon are trademarks of Intel Corporation in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds.

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

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

Other company, product and service names may be trademarks 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.

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, our warranty terms apply. All information in these materials is subject to change without notice.

ALL INFORMATION IS PROVIDED ON AN "AS IS" BASIS, WITHOUT ANY WARRANTY OF ANY KIND.

G522-2567-00