

Grid explained

*John Palfreyman
EMEA Leader, Linux and Grid Services
IBM Global Services*

Contents

Executive summary..... 3

 Defining the need for Grid 3

 Making Grid accessible and real..... 4

Grid means business 5

 A business perspective..... 5

 A user’s view of Grid..... 9

 Building a Grid solution 10

Grid terms explained..... 11

 What does Grid do for me? 12

 How does Grid do this? 13

Grid is real and now 18

 Grid clash solution..... 18

 Anti-virus Grid 20

 Provisioning Grid..... 22

Executive summary

Defining the need for Grid

To drive sustainable business growth, organizations are constantly driving their information technology systems to produce business results sooner and make them available to a wider audience, with improved accuracy and usefulness at the point of need. At the same time, organizations are looking to embrace new computing technologies such as provisioning, orchestration and virtualization and the emerging new standards for interoperability to improve their systems' resilience and make more efficient use of human, technology and capital resources.

Unfortunately, most organizations have enormous, unused computing capacity that is widely distributed within a heterogeneous IT environment—all of which inhibits their ability to achieve these objectives. In a typical day, the productivity of an IT environment can be staggeringly low: mainframes are unused 40 percent of the time; UNIX® servers are actually “serving” less than 10 percent of the time; and most individual workstations actually produce work less than 5 percent of the time.¹ Imagine a manufacturer with 40 percent of its assembly plants idle, an airline with 90 percent of its fleet on the ground or a hotel chain with 95 percent of its rooms unoccupied. Grid offers a powerful way of unleashing the compute power and information within an organization to yield sustainable business benefits and overcome inefficiencies by directing the organization's computing power to the business need with the highest priority.

Making Grid accessible and real

Common complaints about Grid range from how it is explained (“When I ask suppliers what Grid is, I get techno-babble.”) to skepticism about its very existence (“Grid will be great when it finally arrives.”). This white paper addresses these complaints by avoiding deep technical discussions and instead focusing on the business needs that can be met by adopting a Grid solution. Its intended audience is the business or IT staff members who are responsible for improving how a company does business by harnessing technology innovation.

The following three main sections cover the fundamentals of Grid and show real-world examples of how IBM clients are using Grid to gain measurable business benefits today:

- *“Grid means business” highlights the kinds of business challenges that can be addressed by a Grid solution and introduces the three most common terms associated with a Grid solution: **Compute Grid**, **Information Grid** and **Enterprise Grid**.*

- *“Grid terms explained” guides readers through Grid terminology by outlining the fundamental functions in any Grid solution, explaining **what** Grid does for a business and **how** Grid achieves it—in language that does not require a deep and specialized knowledge of IT.*
- *“Grid is real and now” presents a few selected projects in the form of user stories to highlight how a Grid solution can help improve the daily business of a typical user. This section intends to show that innovative Grid solutions exist today and can potentially yield significant business benefits.*

This white paper is first in a series from IBM that showcase the value of Grid solutions by focusing on its business benefits. “Getting started with Grid,” the second paper in the series, is intended for those who are interested in Grid solutions but are unsure of where to start and what to do. It outlines alternative approaches, discusses relative benefits and challenges and presents a road map for a successful Grid implementation. “Getting started with Grid” is supported by several technical and business annexes, which describe solutions to typical challenges presented by Grid, such as security and charging and licensing mechanisms.

Grid means business

Many people still associate Grid with early, high-visibility projects such as SETI@home, which harnessed the power of home computers in an attempt to detect extraterrestrial intelligence through the analysis of small slices of radio telescope data. This project started in the late 1990s and Grid has evolved since then from an interesting technology used by the scientific, technical and educational communities to a powerful way of delivering sustainable business benefits to commercial and government organizations.

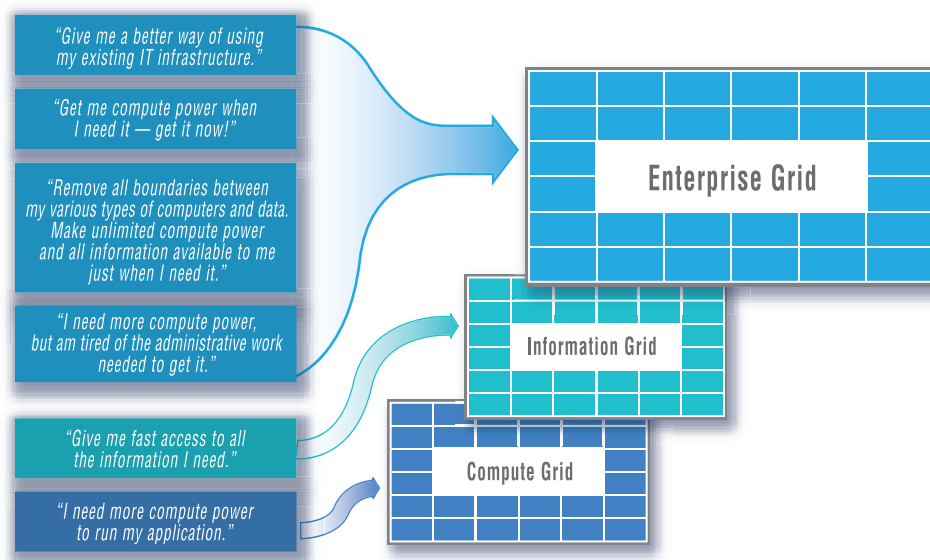
A business perspective

Grid solutions exist to satisfy business needs. Typical requirements that can be satisfied by different types of Grid solutions include:

- *More compute power to run business applications*
- *Affordable analyses performed more accurately and more often*
- *Results returned more quickly—when they are most valuable and relevant*
- *Fast access to all necessary information, irrespective of where it is in an organization*
- *A better way for analysts to use an existing IT infrastructure*
- *Compute power available just when it is needed and users charged for their fair shares*
- *No boundaries between various computers and information*

These different requirements are often associated with common terms for Grid solutions, as shown in Figure 1.

Figure 1: Requirements associated with common Grid terms



The common Grid terms introduced in Figure 1 are further described in the "Grid terms explained" section of this white paper, but essentially a *Compute Grid* offers a way of speeding up a business application by running it sooner (and/or parallelizing the processing work) on shared, dedicated compute capacity or on as-available, spare compute capacity; an *Information* (or *Data*) *Grid* can give rapid information access to all types and sources of business data—from across the enterprise—at the user's workstation or server; and an *Enterprise Grid* brings together Compute Grid and Information Grid resources across organizational divisions to help improve the efficiency of enterprise-wide business functions.

These commonly used terms can be confusing because there is seldom a simple mapping from business needs to Compute Grid, Information Grid or Enterprise Grid; and there is a considerable functional overlap between these Grid types. However, any Grid solution can include one or more of the following functions:

- **Scheduling**—Helps to ensure that compute power is available to run a business application in the minimum time possible
- **Data virtualization**—Can make enterprise-wide information readily available whenever and wherever it is needed
- **Provisioning**—Can make compute power available to users when it is needed and returns it to a central pool when its not
- **Resource management**—Helps to ensure that all elements of a Grid solution work optimally together according to the rules set by the organization

Figure 2 shows that, in practice, a Grid solution can be built by combining different extents of these functions to meet particular business needs.

Figure 2: Grid functions used to build business solutions

Solution Examples	Scheduling	Data Virtualization	Provisioning	Resource Management
Automotive Simulation				
Petroleum Seismic Analysis				
Finance Credit Risk				

With Grid computing, an organization can transform its distributed and difficult-to-manage systems into a large virtual computer that can work on problems and processes too complex for a single computer to handle efficiently. The problems to be solved can involve security, data processing, network bandwidth or data storage. The systems linked in a Grid might be in the same room or distributed across the globe; they might be running different operating systems on many hardware platforms; or they might even be owned by different organizations.

A user's view of Grid

How does a Grid look and feel from a user perspective? There are two contradicting answers to this question: no different and completely different.

The user should experience improved performance by using the Grid—with significant improvements in the speed of critical business applications and critical business information being available as if it were located on the user's personal workstation. In the case of an Enterprise Grid, users should also notice that their computing resources are “always there, always constant,” with no obvious peaks and troughs in performance and no expensive, frustrating outages.

On the other hand, users can interface with the Grid from their personal workstations in a way that is normal, intuitive and not necessarily different from the way they access business applications in a non-Grid environment.

Building a Grid solution

There are many powerful software and hardware products available today from a broad range of suppliers and the open source community to allow an organization to build a robust and reliable Grid solution with manageable technical risk.

It is important to understand that, despite claims made by some Grid technology vendors, a Grid can only be built—it can never be bought “off the shelf.” It is therefore very important for anyone considering Grid as a solution that the right project approach be taken from the start.

In short, the business problem must be thoroughly understood and analyzed before a technology solution is chosen. The business benefits (in quantitative and qualitative form) also must be understood at an early stage so that a proof-of-concept phase can identify the requirements that can be met in practice with the chosen solution.

In turn, this implies that the right solution partner must be carefully chosen. An ideal choice is a partner who understands the organization's particular business needs, can provide a detailed cost/benefit assessment in advance of the project, can build (or help build) the solution and finally can support the solution during the post-delivery phases. Further, a partner that has a track record of solution delivery in a commercial environment—as identified by customer references and testimonials—is a wise choice and can help decrease project risk at all stages. Such a partner should understand the risks and potential pitfalls involved in building a Grid solution, and advise the best ways to avoid, contain or mitigate such risks.

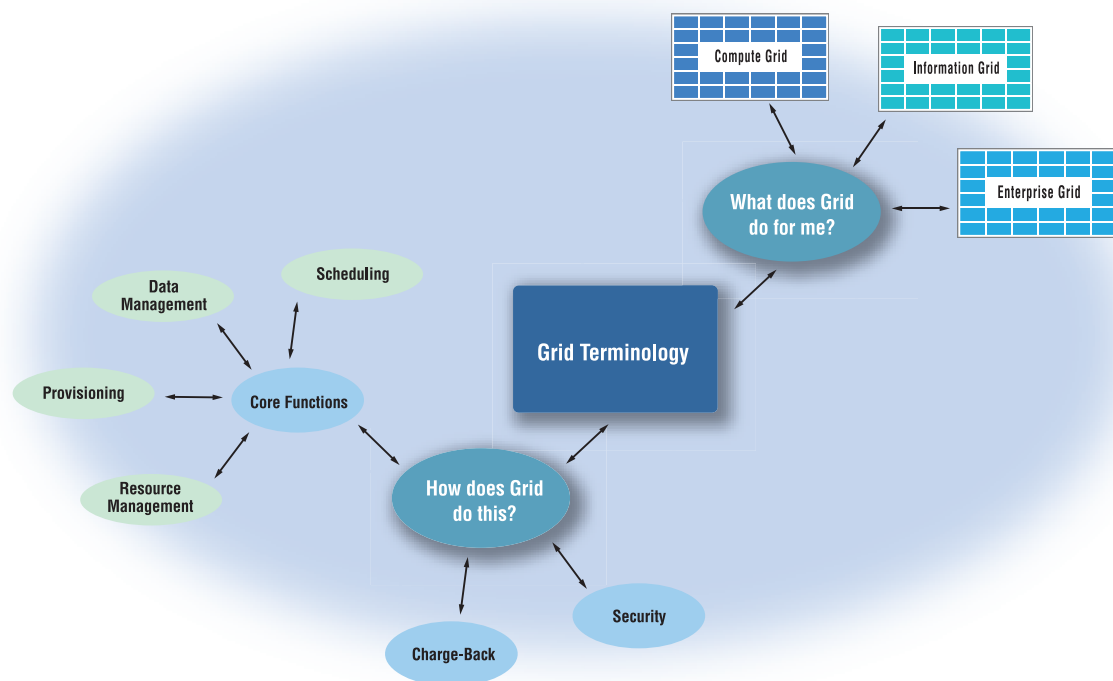
Interestingly, Grid solutions are often built from software products—handling aspects such as scheduling, data virtualization, provisioning and resource management—that could be otherwise considered “normal” in the IT industry. The difficult part involves getting them to work together in a coordinated way. This is considered further in the IBM white paper entitled “Getting started with Grid.”

Grid terms explained

The terminology associated with Grid solutions tends to be both over-technical and confusing. This section explains common Grid terminology in a way that someone with nonspecialist IT knowledge can understand. The section also outlines the business implications of these technical terms.

Grid terminology may be clustered into two distinct groups. The first group of terms are used to explain *what* Grid can do in a business context, and the second are used to explain *how* Grid can be implemented. These terminology groups are shown in Figure 3.

Figure 3: Grid terminology



What does Grid do for me?

The most commonly used terms associated with Grid explain what a Grid can do for an organization. These are as follows:

Compute Grid. This type of Grid is the most common “starter” Grid solution. It offers a way of speeding up a business application by sharing the processing work needed across several computers. Although this Grid works only for certain types of business applications and needs special software or tools, organizations often start by implementing a Compute Grid to illustrate the business benefits associated with implementing a Grid solution.

Information Grid. This Grid type gives rapid information access to all types and sources of business data—from across the enterprise—at the users’ workstations. This is achieved by tying together separate data sources across the organization into a single Information Grid. This gives users considerable productivity benefits if their jobs are information centric, such as that of a research scientist in a pharmaceutical company.

Enterprise Grid. This is the ultimate form of Grid computing. The Enterprise Grid allows users to execute their most demanding business applications in the minimum time possible and get instant access to all information—across the organization. In addition, users experience computing resources that have been described as always there, always constant with no obvious peaks and troughs in performance and no expensive, frustrating outages. The Enterprise Grid provides the functionalities of the Computer and Information Grids—with total visibility and control combined with the minimum possible manual intervention from the IT operations team.

How does Grid do this?

Core functions

Scheduling, data virtualization, provisioning and resource management functions are found—to some degree—at the heart of any Grid solution. Hence, it is important to understand these terms and the role they play in a Grid solution.

Scheduling. In its simplest form, scheduling helps ensure that business applications run as efficiently as possible by marrying them with available compute power. Effective scheduling allows business applications to complete in a fraction of the time taken if “locked” to one computer. Because of the tangible impact that faster results can have for a department—such as faster simulation or risk calculations—scheduling improvements are often where Grid solution experiments start. However, without knowledge of the full spectrum of business potential that Grid offers, this is sometimes where the experiments also stop.

The scope of scheduling within an organization has a major impact on the business benefits that can result. Many early Grid implementations have been restricted to a particular department—perhaps running an engineering simulation on the computers within that department. However, the organization can reap significantly more benefits if peaks in compute load in one department can be matched with lower usage in another. Essentially, the maximum business returns result from enterprise-wide scheduling, which can run a business application on any computer anywhere in the organization. The role that standards plays in realizing this goal is critical, because all machines in the organization are unlikely to be of the same type or running the same operating system. Open standards, such as the Open Grid Services Architecture, underpin the goal of enterprise-wide scheduling and thus can unlock maximum business returns.

There are different types of scheduling. *Task scheduling* enables efficient operation of business applications that typically run in seconds or minutes, while *job scheduling* is the equivalent term for business (typically batch) applications that run for hours or days. In a task scheduling environment, users need to be able to submit information and get the results back quickly. The information must be efficiently transferred to the business application and returned in a way that makes the user think that the application is running locally. Task scheduling is extensively used in financial services where users send customer information to their business applications—such as credit risk predictors—and get back different results for each individual customer.

Data virtualization. Although scheduling improvements can offer more visible benefits for the adoption of Grid technologies, seamless access to enterprise-wide information for improved productivity and collaboration can be more compelling in the long term than being able to speed up a business process.

If a small amount of information is associated with a particular business application, then it is possible for the combined package (data and application) to be moved to the computer in the Grid that can process it. Although this is similar to the way that SETI@home works, it is unmanageable for many commercial applications. Another way of achieving this is to use a storage area network, which distributes information directly to the business applications that need it at the computer where it can be processed, thus achieving the first type of data virtualization. This technique is important because it underpins several Grid solutions.

In a traditional IT system, business information is stored on hard disk drives connected to either a desktop workstation or a server. It is relatively simple for business applications to have rapid access to this information if they are running on the “local” computer, but access becomes more difficult as the infrastructure extends across several geographically dispersed locations and includes computer and storage devices from different manufacturers. At the same time, managing information access is often critical to daily operations and business strategy. Everything must be instantly accessible from anywhere and at any time.

Storage virtualization technology allows the organization to manage these widely distributed resources as if they were one very large “information reservoir.” Balancing overutilized storage in one area with underutilized storage in another area can be taken care of with minimal intervention by IT operations and without end users being aware. The solutions are designed to scale and adjust capacity without disrupting business, thus helping to ensure that the business is satisfied with the service it receives from IT.

Significant and sustainable improvements in business efficiencies can be gained by embracing these types of data, storage and file virtualization across the enterprise, either in isolation or coupled with enterprise scheduling or by moving toward even more advanced concepts such as federated distributed databases (for example, IBM DB2® Information Integrator) as the third and most abstract level of data virtualization.

Provisioning. This function helps ensure that the right computing resource is available at the right time and in the right place to run a business application within the timeframe required by the user. One way to envision this is to imagine an organization having all of its computers in a “resource pool” in the basement and an army of “provisioners” who rush to the user’s desk with just the right number of computers to run their business process in the required time. Once this is completed, the provisioners rush back into the basement returning the computers to the resource pool, waiting for the next user to call. Automated provisioning achieves the same goals without the need for a fast-moving IT operations staff by switching compute power automatically to the users who need it most, thus optimizing the cost-effectiveness and performance of the organization’s computing infrastructure.

Resource management. This is a core functional area because everything in Grid computing and virtualization revolves around the optimal use of resources. There are several aspects to resource management as follows.

Systems management gives IT operations the ability to manage, monitor and understand what is happening with all the key computing and information storage resources and, as such, is critical to the acceptance of Grid computing in the commercial world.

Workload management proactively manages the priority of business applications. This becomes more challenging across a complex network of heterogeneous computers running different operating systems, which is likely to be found in a Grid.

From a business standpoint, workload management helps ensure that service level agreements can be met. These may take the form of “We must have the results from application X within the business day” or “I need access to information Y within 5 seconds,” and can become challenging when multiple business applications are running on one Grid.

A Grid combines a large number of compute and information resources into a larger shared resource. The Grid can then offer a *resource balancing* effect by moving Grid-enabled applications to computers with lower utilization to absorb unexpected peaks in activity. This can help address occasional peak activity loads in parts of a larger organization by better using idle machines in the Grid; or if the Grid is already fully utilized, the lowest-priority work can be temporarily suspended to make room for the higher-priority work.

Other functions

In addition to the core functions just described, there are certain aspects of a Grid solution that can become a blockage to full solution exploitation if not proactively addressed. The most common of these are *charge-back* (also referred to as billing) and *security*.

Charge-back. When many users have common access to a Grid, tracking their compute and information usage may be important to the organization. In this case, charge-back solutions are deployed to accurately track how resources and software licenses are used within the Grid environment so that different parts of the organization can charge for the use of their assets.

Security. Although Grid optimizes the usage of an organization's computational and information storage resources, this capability should not compromise the integrity of the IT infrastructure and business applications and must protect sensitive information from inadvertent disclosure or unauthorized modification. These vital aspects must be addressed by appropriate levels of security—from an end-to-end systemic viewpoint—applied to the various domains within the Grid.

Grid is real and now

This section illustrates the usage of Grid in real-world implementations by outlining how using the Grid solution affects users in practical terms, and explaining how the Grid was constructed from the solution components introduced in previous sections.

Grid clash solution

The first example involves the application of a Compute Grid in an automotive design application. This project was undertaken by IBM for a major automotive manufacturer in Europe called Magna Steyr.

User story

It is a bright September morning and Kurt is preparing to leave for work. Kurt has been working as a design engineer since he left college three years ago and enjoys designing automobiles—and is much enthused by his latest assignment with the new cabriolet design team.

As Kurt cycles along, he remembers stories of how things used to be in the design team. Kurt's job is to ensure that all parts of the new cabriolet fit together correctly by running a computer-aided design package on his design workstation. This software is called a clash simulation and uses so much compute power that performing a clash simulation of the whole vehicle used to take about three days to run. This caused considerable frustrations for the design team. Kurt remembers the amount of manual effort needed to divide the car design into sections so that the simulation could be completed in a reasonable amount of time. They could afford to run the whole-car simulation only at the end of the design process.

At a recent conference, Kurt heard horror stories of the considerable costs and timescale delays incurred by a competitor when this complete clash simulation showed that parts did not fit together correctly. It was rumored that they even had to delay a new model launch date. Kurt could not imagine what this must have cost in both revenue and reputation.

As Kurt locks up his bicycle, he thinks about how different things are now. The whole-car clash simulation that used to take three days can now run in just four hours. Kurt does not really understand how this was done, but his friend Peter on the Information Technology team raves about the innovative use of Grid computing technology introduced to his company by IBM.

What Kurt does understand is how this has revolutionized his job. Shortly after logging on to his workstation, Kurt is delighted to see that the changes he made to the cabriolet front subassembly yesterday evening have been successfully incorporated into the whole-car design. Feeling the warm glow of achievement, Kurt cannot wait to get started on his next design task!

IBM enabled the clash environment at Magna Steyr to run with Platform Computing as a Grid middleware—allowing the Dassault Systemes CATIA DMU application to be run on several workstations within the design office rather than being constrained to one user's workstation. In this way, the clash simulation was split up and distributed across a pool of 20 workstations (Solaris, AIX® and IRIX operating systems) by the Platform LSF package. As a result, the clash simulation could be run as a daily overnight batch rather than a lengthy, multi-day operation.

Anti-virus Grid

This hypothetical example involves the application of an Enterprise Grid for virus detection, removal and central reporting/control and was undertaken by IBM as an internal technology demonstration project only.

User story

Wendy works for the CIO of a major multinational organization and has special responsibility for system security. In preparation for a meeting with her boss, Wendy thinks back to a conference a few months ago where several organizations were discussing how to protect their systems from external malicious attack by the many viruses that are ever more frequently reported in the press.

Wendy was surprised that most of the delegates had the same problems as she did in her organization. Anti-virus programs were installed on all desktop and server machines, but many delegates suspected (but could not prove) that these were disabled or that the users stopped the weekly virus scan before it was complete. One delegate explained, “If I’m in the middle of a customer presentation and the virus scan starts, I kill the scan as quickly as I can. I’m sure all users do the same!”

Wendy decided to take action to address this problem as soon as she returned, because she knew that important company information could be compromised, or in the worst-case scenario, the entire corporate infrastructure could be rendered useless by viral attack. She knew that the IBM team was working on a Managed Desktop proposal for her colleague Peter, and Wendy decided to discuss possible radical solutions with them.

The lead architect from IBM made a surprising suggestion. Why not tie all desktop and server machines together into a company-wide Grid, and then “wrap” the corporate anti-virus package into a service-oriented architecture. This would enable the anti-virus program to run in the quiet times of each computer (rather than once a week) and also provide one view of the organization’s anti-virus status to the CIO. Wendy was so impressed with this idea that she immediately involved her boss and together they commissioned a prototype with IBM that is now in use across their head office IT infrastructure. That was three months ago, and now Wendy is looking forward to reviewing new projects.

The meeting starts with a very positive and constructive atmosphere. Her boss is very pleased to have a complete overview of the virus protection status of the HQ team, and has already found several people who had not run an anti-virus scan in months and has paid them a personal visit! The user representatives are equally pleased. They tell Wendy that they thought about her once a week at 12 P.M. on Monday (when their virus scan used to run) because there was no frustrating downtime waiting for it to complete. Instead, the virus scan runs when they are not using their computer—and they very soon forgot about it.

The meeting concludes with Wendy being given the go-ahead to extend the project to cover the entire organization—“the sooner, the better,” as her boss puts it. All she has to do now is to get a good price from IBM for the full project implementation. She feels confident that she can do this!

In this example, IBM made innovative use of the Globus Toolkit and the service-oriented architecture so that the anti-virus scan could run on the workstations and servers during quiet times and not disturb users' working patterns. The use of the Globus Toolkit helped ensure that a single "security map" of the virus scan status for the entire organization could be compiled and the Grid services on the network could be extended to other applications.

Provisioning Grid

This example involves the application of provisioning technology to the complex problem of software configuration and testing, based on a project undertaken by IBM for a major European automotive manufacturer.

User story

Robert has worked as software test manager at a major enterprise for the past three years. He leads a team of software engineers in the end-to-end testing of the system software that forms the heart of his company's most modern product. As with many complex products, Robert knows that the system software is critical from many points of view. Without it, the product would be unusable and useless. Robert also knows that there are about 15 separate computers within the latest product—and all of them must work together in an error-free manner in order for his company to continue to enjoy its prestige brand and market leadership position. Hence, the role of Robert's team—to configure, test and fix faults in the system software prior to product launch—is critical to the success of the entire organization.

Robert is always keen to exploit the latest software technology to help his team do their job more effectively. He is justifiably delighted with a new automated provisioning system—designed and built by the IBM Grid Team—which allows his team to concentrate on the most important job of finding and fixing faults in the system software.

Before the new system was installed, Robert's team would spend up to five working days to configure the 14 servers in the system test center. During this time, no testing could be done, and team members became very frustrated and impatient to get on with their main testing job. These frustrations were further compounded by the manual configuration errors that were inevitable in such a complex configuration process.

Things are very different now. The IBM Tivoli® Provisioning Manager (TPM) product takes care of the entire configuration activity. Using TPM, Robert's team builds a graphical representation of the test environment on a workstation screen and then automatically configures all the servers accordingly—and

does so quickly. The configuration task is completed in about four hours, avoiding the manual, error-prone processes of the past. The secondary consequences are considerable. Robert can now commit with confidence to aggressive service levels for his managers and users (the software developers in his company) and can use his test hardware much more efficiently, requiring less capital investment—a fact much appreciated by the finance department.

But what pleases Robert the most is that the new system frees up his team to spend more time doing what the organization needs them to do—that is, to fix faults and optimize the end-to-end system software. And he is proud that, by adopting such innovative new technology, his company no longer needs to worry about its competition catching up.

In this project example, IBM made extensive use of the Tivoli Provisioning Manager product to automatically configure a multiple-server installation to represent the changing software development infrastructure of the customer.



© Copyright IBM Corporation 2005

IBM Corporation
Systems and Technology Group
Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
January 2005
All Rights Reserved.

IBM, the IBM logo, AIX, DB2 and Tivoli are trademarks of International Business Machines 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 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.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only. ALL INFORMATION IS PROVIDED ON AN "AS-IS" BASIS, WITHOUT ANY WARRANTY OF ANY KIND.

¹Source: IBM/Ogilvy IT Influencer Briefing Document, September 2004.