

Deploying a Grid Solution with IBM InfoSphere Information Server

IBM Redbooks Solution Guide

Applying grid technology to business IT infrastructures represents an exciting evolution of the virtualization, integration, and support of industry standards that IBM® has long been applying to solve customers' business problems. Grid computing allows a set of disparate computers in an organization to function as one large, integrated computing system. That single system can then be used on problems and processes that are too large and intensive for any single computer to easily handle alone efficiently. Figure 1 illustrates an abstract view of a grid deployment. Additionally, a grid provides processing capacity redundancy in support of high availability objectives, so that if one machine in the grid is offline, work is assigned to another machine.



Figure 1. Grid deployment

The parallel engine of IBM InfoSphere® Information Server enables IBM InfoSphere DataStage®, QualityStage®, and Information Analyzer jobs to run in parallel on either a single symmetric multiprocessor (SMP) server or on multiple servers in a clustered environment. In both cases, a simple configuration file that is supplied at run time defines the degree of parallelism and the corresponding servers or machines to be used by a job. To change the degree of parallelism, or the servers on which a job runs, you must supply a configuration file with the new number of nodes and its associated servers.

To optimize the value of a grid, IBM Lab Services offers the Grid Enablement Toolkit to work with InfoSphere Information Server and resource manager software to dynamically build these configuration files based on the machine reservations that are available. This solution guide explores running InfoSphere Information Server in a grid environment and provides an overview of the Grid Enablement Toolkit.

Did you know?

Enterprise data volumes are growing. In addition, the number of data sources is growing, the number of consumers is growing, and the total number of integration points is growing, making information integration and governance more important than ever before. A grid environment is perfectly suited for keeping up with the pace of data volume growth because it combines lower-cost hardware with a straightforward and repeatable expansion pattern.

Many of the largest InfoSphere Information Server deployments now run on a grid configuration. Some organizations turned to grid because they needed to scale the number of applications to keep up with various business demands. They are now running tens of thousands of applications each day on a grid. Others organizations needed to scale the volume of data they could process in a cost effective manner. They are now processing terabytes per hour on a cluster of 4, 8, or 12 core machines.

Business value

The benefits of grid computing go far beyond power and speed by delivering the following important characteristics:

- Lower total cost of ownership. Grids use lower-cost hardware and operating systems to complete their processing, rather than traditional higher-cost SMP platforms. As a result, they can achieve a lower total cost of ownership and better price/performance advantage.
- Meeting service level agreement (SLA) obligations. Applications are unaware of the computers in the grid that are running a task and the number of computers that are assigned to run the task. This characteristic provides a simple means to quickly scale out the application run time to meet SLA obligations when the business expands and data volume grows.
- Utility usage or chargeback model. Grids work with resource manager software to assign and track the node usage of applications on the grid. Ultimately, the cost of using the capacity on the grid can be recouped by charging a department or customer for the cycles that are required to run their application portfolio over a set period.

Solution overview

The InfoSphere Information Server grid solution has several components that administrators and developers interact with. But first, you must understand the terminology behind the solution.

Grid terminology

The following terms are used in this document within the context of grid computing with the InfoSphere Information Server:

- **Node:** A server within the grid.
- **Partition:** One of the degrees of parallelism in a job. One or more partitions can run on a node.
- **Head or conductor node:** The main server in the grid that provides software (such as InfoSphere DataStage) or services to the compute nodes in the grid. It can also be referred to as the *DataStage conductor node*.
- **Compute node:** A server in the grid (other than the head or conductor node) that typically does not have permanent data storage devices, but provides pure compute power (processor) to a submitted job.
- **Parallel framework:** The InfoSphere Information Server parallel engine.

- **System management tools**: Used to manage servers within the grid environment such as Network Information Service (NIS; user ID and password) and Preboot Execution Environment (PXE) booting (network boot).
- **Resource monitor**: Software used to monitor hardware use of the grid.

Solution components

The InfoSphere Information Server grid solution combines the data integration platform with two other key components: the Grid Enablement Toolkit and the Resource Manager:

- *Grid Enablement Toolkit* is a set of scripts and templates that are used to create a dynamic configuration file (`$APT_CONFIG_FILE`). This configuration file is based on interaction with resource manager software that identifies idle servers in the grid. The toolkit has the following main functions:
 - Coordinate the activities between the InfoSphere Information Server parallel engine and the resource manager.
 - Create the parallel configuration file from a template.
 - Log the activity.
- *Resource Manager software* manages the nodes and the grid and identifies the idle servers from among that pool of servers. It provides the necessary scheduling and monitoring tools to distribute jobs to nodes that are available within the grid.

IBM offers and recommends IBM Platform™ Symphony as the resource manager in this solution. Some customers have worked with IBM Lab Services to customize a solution that uses other resource manager software.

Job submission and execution flow

Organizations that are already familiar with InfoSphere Information Server will find that the primary touchpoints for job submission are the same. Jobs can be submitted to the grid through the Operations Console or Director clients, or they can be submitted by using the dsjob application programming interface (API). After the job is submitted, the solution affects the changes that are required to use the power of the grid.

The grid implementation works as explained in the following process and as illustrated in Figure 2:

1. When a job is submitted, it is intercepted by a component of the Grid Enablement Toolkit. The component generates and submits a script to the resource manager that gets placed in a Grid Resource Manager queue.

The script remains in the queue until the resources (appropriate number of nodes and specific nodes if constraints are placed for them in one or more stages in the job) that are required by the job become available.
2. After the node or nodes that are requested become available as monitored by the resource manager, the submitted script generates a parallel configuration file based on those specific nodes and starts the job. While the job is running, the resource manager prevents additional jobs from using the same resources until the running job completes.
3. The submitted script waits until the parallel job finishes before performing a final cleanup and release of assigned resources so that those resources can then become available to other jobs.

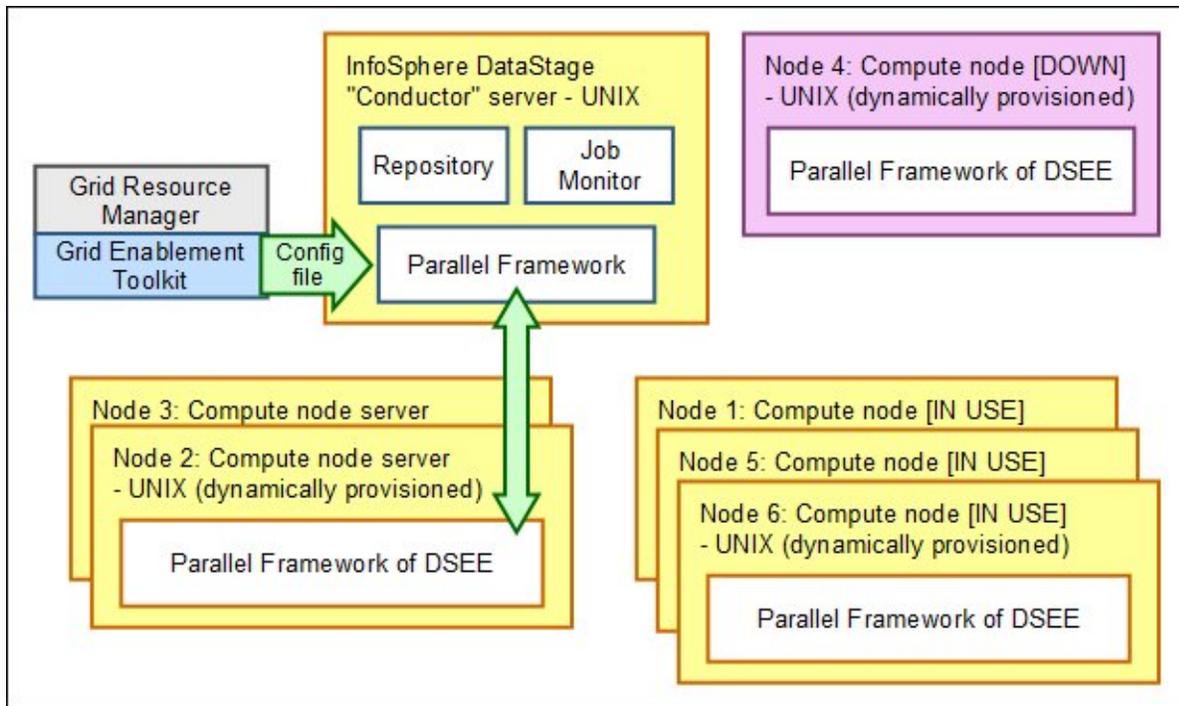


Figure 2. Grid computing in InfoSphere Information Server

Solution architecture

Setting up your InfoSphere Information Server grid computing environment entails the steps that are outlined in Figure 3.

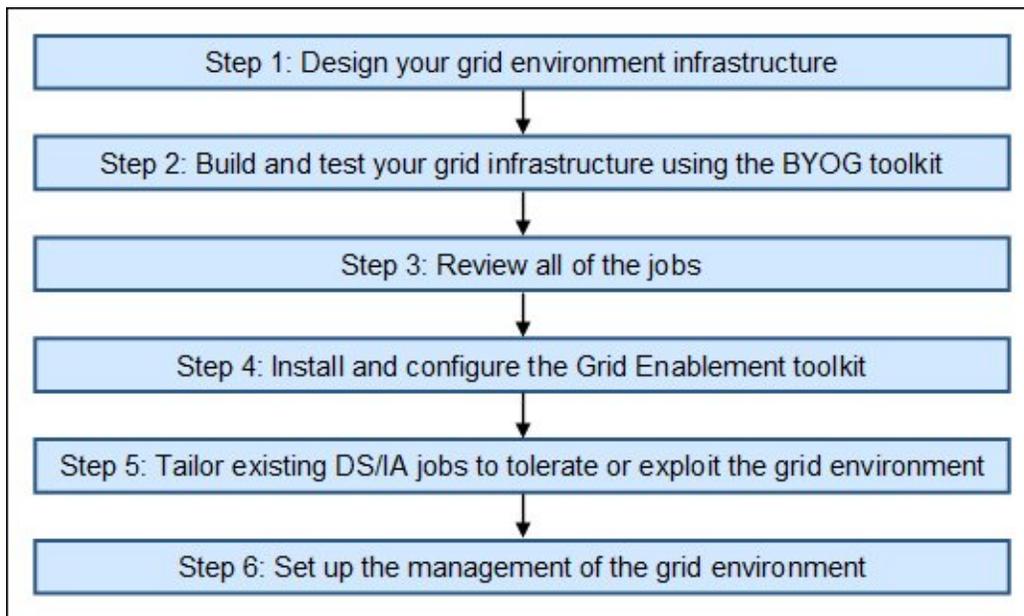


Figure 3. Setting up the InfoSphere Information Server grid environment

Setting up your InfoSphere Information Server grid computing environment entails the following steps.

Step 1: Design your grid environment infrastructure

After choosing to implement a grid environment as a solution to your business requirements, decide on the following elements of the grid infrastructure:

- Operating system (OS) for the compute node

If you have an existing InfoSphere Information Server, that OS platform influences the decision because all operating systems must be the same (homogeneous) in the grid environment. Most current grid customers are deployed on Red Hat Linux or SUSE Linux.

- Number and capacity of compute nodes in the grid environment

Each concurrent job and job sequence requires its own slots within the set available across the compute nodes. The cumulative of all these compute nodes is what is required as a starting point and are influenced by the number of total projects and the data volumes to be processed.

- Head node capacity

Certain activities of the platform (such as log collection, sequencer execution, and operations monitoring) run entirely on the head node. Size the head node adequately to accommodate the workload of the hundreds or thousands of jobs to be coordinated through this node.

- High availability (HA) configuration

If an HA solution is desired, decide whether a dedicated or shared standby node configuration is appropriate. If a shared standby node is selected, you must determine whether this node must be one of the compute nodes, which requires the compute node to have an identical capacity as the head node has.

- Storage type

Determine whether to use a network-attached storage (NAS)-based configuration or a storage area network (SAN)-based configuration. Your existing IT infrastructure will most likely influence this decision.

- Network type

Determine whether to have a separate (private) network for the connections between the front-end node and the compute nodes. This decision determines whether additional network interface cards (NICs) are required.

- Resource Manager

A grid solution uses dynamic workload management functions and, therefore, requires a resource manager. Having a resource manager as a standard part of the solution, as the IBM Platform Symphony is with an InfoSphere Information Server grid solution, can also influence this decision.

- Resource Monitor

Cost and functionality determine this decision.

- Compute node image configuration.

Decide whether to use software, such as the PXE boot process, to configure the compute nodes.

Step 2: Building and testing your grid infrastructure by using the BYOG toolkit

The Build Your Own Grid (BYOG) toolkit is a set of scripts and templates that are delivered by IBM Lab Services. The BYOG toolkit helps you to configure a Red Hat or SUSE Linux Grid environment. It provides the necessary tools to build the head node and the compute nodes without system administration on each compute node.

Step 3: Reviewing jobs

Your existing InfoSphere Information Server has a number of InfoSphere DataStage, QualityStage, and Information Analyzer jobs. Identifying and understanding which jobs will be migrated from an existing environment to the grid helps define the grid configuration limits:

- Grid environment global variable `MaxNodesPerJob`
- Job-specific values such as `COMPUTENODES` and `PARTITIONS`.

Step 4: Installing and configuring the Grid Enablement Toolkit

The Grid Enablement Toolkit that is provided modifies the InfoSphere Information Server components to enable the creation of a dynamic configuration file. You must install and configure this toolkit by using the IBM Lab Services engagement to set up your grid environment.

Step 5: Tailoring existing jobs to use the grid environment

Although most of the configuration is done on the environment, you are likely to have a set of jobs for which you want to take a specific set of capacity at run time that is different than the default set. In those cases, you need to override at the job or sequencer level the resource quantity to be requested of the Resource Manager. The grid toolkit defines two new environment variables that allow the user to make job-specific requests:

- `$APT_GRID_COMPUTENODES`: This value must be between 1 and the `MaxNodesPerJob` value, where `MaxNodesPerJob` is specified in the `grid_global_values` or `DSParams` file for each project.
- `$APT_GRID_PARTITIONS`: Specifies the number of partitions per compute node. It can have a value of 1 through *n*.

Step 6: Setting up the management of the grid environment

Setting up the management of the grid environment entails the following tasks:

- Administering and managing the resource manager such as IBM Tivoli® Workload Scheduler LoadLeveler®
- Administering and managing the resource monitor such as Ganglia
- Testing the failover and failback scenarios of your HA configuration
- Performance monitoring and tuning of the grid environment

The following sections provide more details about several of the topics described in the previous steps.

Note: IBM offers services engagement to help you design the optimal grid environment for your organization. Contact your local IBM marketing representative for details about building an Information Platform and Solutions data integration grid.

Grid node configurations

Grid environments with InfoSphere Information Server vary depending upon your particular current configuration or IT environment, such as the use of NAS or SAN. Regardless of the configuration you implement, a server must be designated as the primary server or head node, and multiple servers must be designated as compute nodes.

- All InfoSphere DataStage, QualityStage, and Information Analyzer jobs are initiated from the head node.
- Compute nodes are where the jobs are processed.

The head node must be able to accommodate messages from all concurrently submitted jobs. Data is moved between the various nodes based on job partitioning requirements by the Parallel Framework components using TCP/IP (ports $\geq 11,000$). The ports in the 10,000 range are used for conductor-to-section-leader communication. The topology of the network with 1-Gb switches must be managed to isolate jobs to a gigabit switch whenever possible.

Network storage configuration

NAS configuration is the recommended solution over SAN for grid implementations for the following reasons:

- NAS configuration is simple and allows for multiple head nodes that share the compute resources (which can be valuable when accommodating development, test, production, or all three environments within one grid). A SAN configuration requires something similar to a Global File System (GFS) to support multiple head nodes.
- A SAN configuration requires multiple NICs on the front end of the NFS to share the file systems to the compute nodes, given the limitation of 1-GB limit of each NIC card. Based on I/O requirements, you might require two or more NICs for the NFS mounts.
- In a SAN configuration with more than one head node, the compute nodes cannot identify the specific head node `/home` directory to use without GFS. With GFS, however, the compute nodes can still only share the one `/home` directory between both head nodes. With a NAS configuration, each user has only one `/home` directory. Different users can access different `/home` directories. Software mount points are defined based on a specific head node, avoiding any confusion about the specific `/home` directory to be used. As mentioned previously, you can support different uses such as development and regression testing from the same compute nodes.

Private networks

Large amounts of data and messages are moved between the head node and the compute nodes, which can place significant bandwidth demands on the gigabit switches. A separate private network is recommended for this data movement by using multiple NICs for public and private network connections. This network is applicable to SAN-configured grid environments.

When compute nodes have multiple NICs for public and private network connections, there is a chance that the actual host name of the nodes is not in the private network, but in the public network. Based on the resource manager used, this situation causes the routing of the activity to occur on the public network rather than on the desired private network. To avoid this situation, by using InfoSphere Information Server, you can translate the node name into a private network name by using a node translation process that is run whenever the dynamic `APT_CONFIG_FILE` is created.

High availability grid environments

HA in a grid environment (Figure 4) is similar to HA in a non-grid environment.

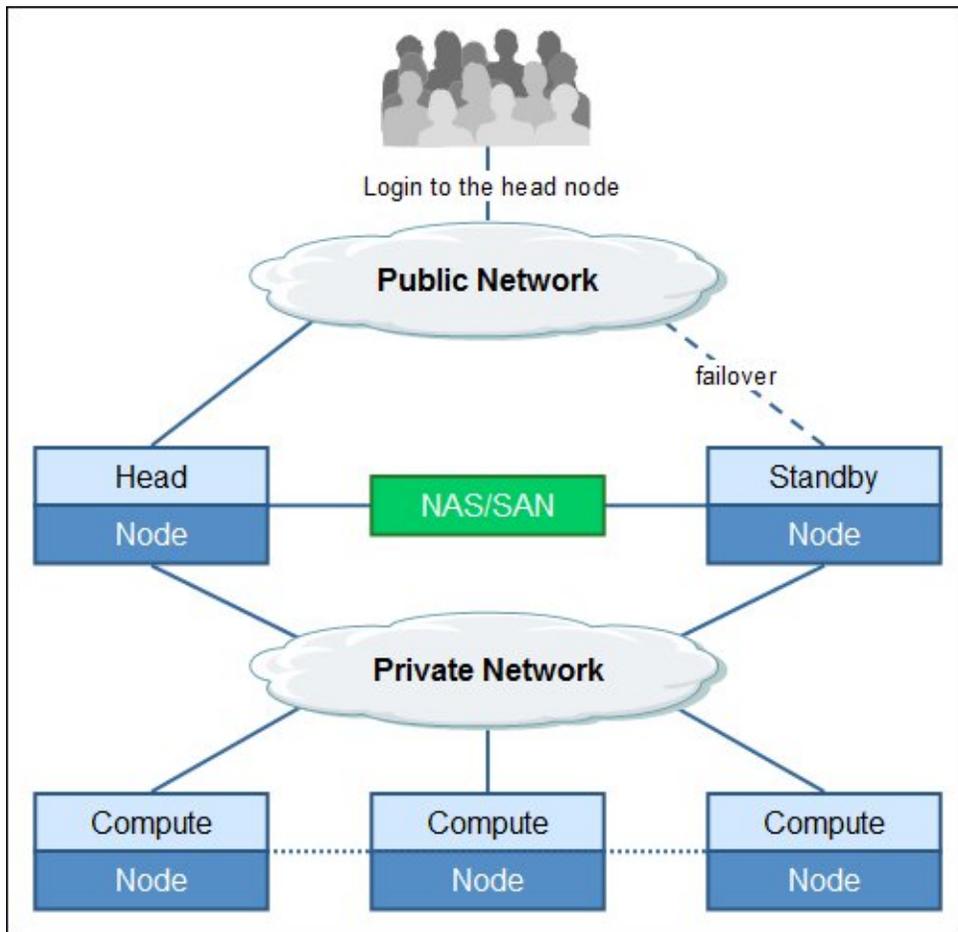


Figure 4. Grid configuration with high availability

An HA grid environment has the following considerations:

- A dedicated standby node (as shown in Figure 2) or a shared standby node can be used. For a shared standby node, the workload that runs on the shared node can be a non-grid workload or one of the compute nodes that is running the grid workload. In most cases, one of the compute nodes in the grid environment serves as the standby shared node.
- The standby node requires you to enforce the following factors:
 - o HA software runs on the standby compute node and the head node.
 - o Processor, memory, and disk requirements on the standby node must match that of the head node. Deciding which architecture to choose.

Working with resource managers

With a resource manager, the InfoSphere Information Server grid computing solution can use a resource (node) without knowing which resource (node) is providing the service. This way allows a process to use a resource (node) today that was not available yesterday or that might become unavailable tomorrow. The resource manager supports this scenario by keeping track of resources, identifying the servers that are down, and monitoring the system load.

IBM offers and recommends IBM Platform Symphony as the resource manager in the Information Server grid solution. Some customers have worked with IBM Lab Services to customize a solution that uses other resource manager software, including IBM LoadLeveler, Oracle Grid Engine, and DataSynapse GridServer.

The resource manager can be used to manage project workloads by using queues. The queues provide the following management capabilities:

- Prioritize workloads. Different queues can be assigned a priority to use more of the resources on the grid than others. This feature allows projects with strict SLA requirements to get the proportion of time and resource required to adhere to those commitments.
- Act as a license restrictor. You can restrict the types of jobs that can be started for any queue. For example, you might only license 4 compute nodes for InfoSphere QualityStage, but the grid has 20 nodes in total. The resource manager can track which machines have InfoSphere QualityStage installed and launch only those jobs on the appropriate nodes.
- Alter priority based on time of day. In grids that host production and non-production environments, the time of day is critical to determine the priority of workloads and the degree of resources that are available. The resource allocations can be changed through simple scripting commands to accommodate this flexibility.

The options of a particular resource manager determine the specific management capabilities. For more information, see the relevant product documentation as listed in "Related information." For a description of the grid solution and more configuration details, see the IBM Redbooks® publication *Deploying a Grid Solution with the IBM InfoSphere Information Server*, SG24-7625.

Integration

InfoSphere Information Server offers a collection of product modules and components that work together to achieve business objectives within the information integration domain. The product modules provide business and technical functionality throughout the entire initiative from planning through design to implementation and reporting phases.

InfoSphere Information Server consists of the following product modules and components:

- IBM InfoSphere DataStage
- IBM InfoSphere QualityStage
- IBM InfoSphere Information Analyzer
- IBM InfoSphere Business Glossary
- IBM InfoSphere Metadata Workbench
- IBM InfoSphere FastTrack
- IBM InfoSphere Information Services Director
- IBM InfoSphere Blueprint Director

Supported platforms

InfoSphere Information Server tiers are available on the following platforms:

- The installable client tier components, which provide the user interface, are available only on Microsoft Windows platforms. Business Glossary and Metadata Workbench require only a supported web browser. These two Information Server modules do not have any client installable components.
- The server tiers (services, engine, and repository) are available on the Linux, UNIX, and Windows platforms (Microsoft Windows Server, Red Hat Linux, SUSE Linux, IBM AIX®, Oracle Sun Solaris, and Hewlett Packard HP-UX). Although each services tier component can be deployed on a separate host, the services and engine tiers should be deployed on the same platform type.
- The database for the InfoSphere Information Server repository can be implemented by using IBM DB2®, Oracle, or SQL Server.

Ordering information

IBM InfoSphere Information Server is available only through IBM Passport Advantage®. It is not available as a shrink wrapped product.

The InfoSphere products can be sold only directly by IBM or by authorized IBM Business Partners for Software Value Plus. For more information about IBM Software Value Plus, go to:

http://www.ibm.com/partnerworld/page/svp_authorized_portfolio

To locate IBM Business Partners for Software Value Plus in your geographic region for a specific Software Value Plus portfolio, contact your IBM representative.

For ordering information, see the IBM Offering Information page (announcement letters and sales manuals) at:

http://www.ibm.com/common/ssi/index.wss?request_locale=en

On this page, enter *InfoSphere Information Server*, select the information type, and then click **Search**. On the next page, narrow your search results by geography and language.

Related information

For more information, see the following documents:

- *Deploying a Grid Solution with the IBM InfoSphere Information Server*, SG24-7625
<http://www.redbooks.ibm.com/redbooks/pdfs/sg247625.pdf>
- *IBM InfoSphere Information Server Deployment Architectures*, SG24-8028
<http://www.redbooks.ibm.com/abstracts/sg248028.html>
- *IBM InfoSphere Information Server Installation and Configuration Guide*, REDP-4596
<http://www.redbooks.ibm.com/abstracts/redp4596.html>
- IBM InfoSphere Information Server
http://www.ibm.com/software/data/integration/info_server
- IBM InfoSphere Information Server Enterprise Edition
http://www.ibm.com/software/data/integration/info_server/enterprise-edition
- IBM Offering Information page (announcement letters and sales manuals)
http://www.ibm.com/common/ssi/index.wss?request_locale=en
- IBM InfoSphere Information Center
<http://publib.boulder.ibm.com/infocenter/iisinfo/v9r1/index.jsp>
- The IBM Publications Center
<http://www.ibm.com/shop/publications/order>

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service. IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you. This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk. IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

© Copyright International Business Machines Corporation 2013. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

This document was created or updated on January 17, 2013.

Send us your comments in one of the following ways:

- Use the online **Contact us** review form found at:
ibm.com/redbooks
- Send your comments in an e-mail to:
redbook@us.ibm.com
- Mail your comments to:
IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400 U.S.A.

This document is available online at <http://www.ibm.com/redbooks/abstracts/tips0965.html> .

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>. The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both: <http://www.ibm.com/legal/copytrade.shtml>

AIX®
DataStage®
DB2®
IBM®
InfoSphere®
LoadLeveler®
Passport Advantage®
Platform™
QualityStage®
Redbooks®
Redbooks (logo)®
Tivoli®

The following terms are trademarks of other companies:

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

Microsoft, Windows, and the Windows logo 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, or service names may be trademarks or service marks of others.