IBM InfoSphere DataStage and QualityStage
Version 8 Release 7

Connectivity Guide for IBM Red Brick Warehouse

IBM

IBM InfoSphere DataStage and QualityStage
Version 8 Release 7

Connectivity Guide for IBM Red Brick Warehouse

IBM

> **Note**
>
> Before using this information and the product that it supports, read the information in "Notices and trademarks" on page 17.

Contents

Chapter 1. Introduction

InfoSphere® DataStage® provides the ability to load tables in Red Brick warehouses in bulk. These topics provide an introduction to the Red Brick Load stage as well as installation instructions and configuration information.

Audience

This guide is intended for InfoSphere DataStage designers who create or modify jobs that use the Red Brick Load stage.

Introduction to the Red Brick Load Stage

Red Brick Load enables you to rapidly and efficiently prepare and load streams of tabular data from any InfoSphere DataStage stage (for example, the ODBC stage, the Sequential File stage, and so forth) into tables of the target Red Brick Warehouse.

Red Brick Load is a passive stage that supports one or more stream input links (but no output links). Each input link corresponds to a different bulk loading session within an InfoSphere DataStage job.

Bulk loading can occur automatically only if both the InfoSphere DataStage job and Red Brick server are on the same physical machine.

Configuring the Environment

There is no requirement for special environment variables in support of the Red Brick Load stage.

Red Brick Load and the Parallel Canvas

Red Brick Load stages can run on the Parallel Canvas. The default for *all* stages is Sequential. "In Parallel" mean you can set it to run in parallel, but this is *not* the default. Table 1 provides detailed information.

Table 1. Availability on the Parallel Canvas

Connectivity Stage	Available on the Parallel Canvas in Windows	Available on the Parallel Canvas in UNIX	Used as a Source or a Target or for Processing	Runs Sequentially or In Parallel
Red Brick Load		Yes	Target	Sequential

Mapping String Data
About this task

The purpose of the NONE map on the Server canvas is to turn off mapping of string data in any stage in which the map is set, that is, to pass the data through verbatim. This feature is handled differently on the Parallel Canvas. When you

define string data (char, varchar, and so on), there is an additional field in the stage's Columns grid called Extended. This can be set to blank or Unicode. If this option is set to blank, no mapping occurs (that is, "NONE"); the map specified on the NLS tab is ignored. If this option is set to Unicode, the NLS map is applied. In order to read or write Japanese data, for example, set Extended to Unicode. When the job compiler detects this combination (char, varchar, and so on, and Unicode), it generates the appropriate run-time code.

Chapter 2. Red Brick Load Stages

This topic describes the following:

Functionality

Red Brick Load has the following functionality:
- Support for data files that exceed the 2-GB file size limit for 64-bit file systems.
- Generation and automatic execution of Red Brick commands to load data into the Red Brick Warehouse.
- Generation of intermediate data files in delimiter-separated ASCII format.
- An option to drop and create simple target tables as defined in the column information of the corresponding link (for example, tables that have no referential constraints, triggers, or stored procedures).
- An option to generate the data and script files only for manual execution.
- An option to let you specify commands to run the BeforeCommand and the AfterCommand, for example, to initialize the database.
- Support of NLS (National Language Support).

The following functionality is not supported:
- Complex TMU or RISQL scripts (you can create these using any text editor, such as WordPad, Notepad, and so on)
- Binary data types
- Generation of fixed-record format for data files

Terminology

The following table explains the Red Brick terms used in this document

Table 2. Terminology in this Book

Term	Description
Bulk Load Plug-in	A passive stage whose role in a DataStage job is to take streams of tabular data and load them into tables of a target database.
control file	A file of commands that bulk loads a table from a single link. The control file acts as an input file to the TMU utility. See "Input Properties" for details.
data file	An ASCII file of row/column data from an input link that is to be loaded.
RISQL	An SQL entry tool with a set of business analysis extensions to SQL for warehouse databases.
SQL file	An input file to RISQL, consisting of standard SQL or RISQL extension statements.

Table 2. Terminology in this Book (continued)

Term	Description
TMU	The Table Management Utility that loads data into the Red Brick Warehouse and maintains the tables, indexes, and referential integrity.

:

Mode of Operation

Red Brick Load has two modes of operation:

- **Automatic Mode.** The stage executes the table creation command, the Before Command, the data loading command, the After Command, and the command for removing intermediate data files.
- **Manual Mode.** If you set **Load Automatically** to **N**, the stage stores the commands in a batch file without executing them. You can run the batch file later to execute the commands.

Red Brick Load generates intermediate data files, the TMU control file, the SQL file for table creation, and a batch file. The batch file, stored in the control file directory, contains the BeforeCommand, the AfterCommand, the RISQL command for table creation, the TMU command for data loading, and the Windows command for intermediate data file removal, if applicable.

Note: You must run the batch file from the project directory if the directories entered contain relative pathnames.

Control Files

The Red Brick Load stage can generate up to nine files per link, including the following:

- A control file for data loading if no CustomCntlFile is entered
- An SQL file for table creation
- An intermediate data file to store data from the (input) link
- A batch file to store the spawn commands
- A log file for each of the spawn commands if applicable: table creation, BeforeCommand, data loading, AfterCommand, and removal of the intermediate data file

The following table lists the file name used if a custom control file is entered, the default file name used if no custom control file is entered, and the description for the file:

Table 3. Naming Conventions and Descriptions of Control Files

File Name (CustomCntlFile)	File Name (No CustomCntlFile)	Description
CustomCntlFile.tmu	database_table.tmu	Control file for data loading
CustomCntlFile.sql	database_table.sql	SQL file for table creation
CustomCntlFile.dat	database_table.dat	Data file to store data from the input link

Table 3. Naming Conventions and Descriptions of Control Files (continued)

File Name (CustomCntlFile)	File Name (No CustomCntlFile)	Description
CustomCntlFile_table.bat	database_table.bat	Batch file to store the spawn commands
CustomCntlFile_create.log	database_table_create.log	Log file for table creation
CustomCntlFile_load.log	database_table_load.log	Log file for data loading
CustomCntlFile_before.log	database_table_before.log	Log file for the BeforeCommand
CustomCntlFile_after.log	database_table_after.log	Log file for the AfterCommand
CustomCntlFile_removal.log	database_table_removal.log	Log file for the removal of the intermediate data file

Loading a Red Brick Warehouse

About this task

Using the IBM® InfoSphere DataStage Designer,

Procedure

1. Add a Red Brick Load stage to your InfoSphere DataStage job
2. Link the Red Brick stage to its data source
3. Specify column definitions using the **Columns** tab
4. Determine the appropriate load mode, as documented in "Mode of Operation"
5. Add the appropriate property values on the **Inputs** tab, as documented in "Input Properties"
6. Compile the job
7. If the job compiles correctly, you can choose one of the following:
   - Run the job from within Designer
   - Run or schedule the job using the InfoSphere DataStage Director
8. If the job does not compile correctly, correct the errors and recompile

Input Properties

Use the **General** tab to specify the load operation.

RedBrick Load supports the following input properties that are visible from the IBM InfoSphere DataStage Designer. The following table includes these column heads:

- **Property** is the text that the job designer sees in the stage editor user interface.
- **Default** is the text used if the job designer does not supply any value.
- **Help Text** describes the properties.

Table 4. Input properties

Property	Default	Help Text
Database Name	None	(Required) Database name.
User ID	None	Name used to connect to the Red Brick® database.

Table 4. Input properties (continued)

Property	Default	Help Text
Password	None	Password for the database.
TableName	None	Name of the Red Brick database table to be loaded. (Required if Custom Control File Name is not specified.)
Create a new table?	Y	Y=Yes, N=No. Creates the table, using the column definitions for the link.
TMU Mode	insert	Valid values for operation of the InfoSphere DataStageManager: append, replace, modify, update, and insert.
Delimiter	| (vertical bar)	Field delimiter character used to separate columns in the intermediate data file.
Stop loading after a number of discarded rows	1	Number (0 or more) of discarded rows after which to stop loading.
Custom Control File Name	None	Name of the custom control file. Otherwise, defaults to the generated control file named *database_table.tmu*.
Before Command	None	Windows commands executed before table creation and loading.
After Command	None	Windows commands executed after table creation and loading.
Remove intermediate data file	Y	Y=Yes, N=No. Y removes the intermediate data file after the stage executes the appropriate After Command.
Data File Directory	None	Name of the directory containing the intermediate data files that store data from the input link. Otherwise, defaults to the name of the directory for the InfoSphere DataStage project.
Control File Directory	None	Name of the directory containing the custom control files that load data, create tables, and so forth. Otherwise, defaults to the name of the directory for the InfoSphere DataStage project.

Table 4. Input properties (continued)

Property	Default	Help Text
Log File Directory	None	Name of the directory containing logs files with information about table creation, Before Command, data loading, After Command, or removal of the intermediate data file. Otherwise, defaults to the name of the directory for the InfoSphere DataStage project.
Discard File Directory	None	Name of the directory containing the rows that are not loaded by the TMU. Otherwise, defaults to the name of the directory for the InfoSphere DataStage project.
Load Automatically	Y	Y=Yes, N=No. If N is entered, the stage only prepares the intermediate data file and the script files. You can later load the data by running the batch file stored in the control file directory. You must run this batch file from the project directory if the specified directories, such as CntlDirectory and LogDirectory, do not use an absolute path.

Contacting IBM

You can contact IBM for customer support, software services, product information, and general information. You also can provide feedback to IBM about products and documentation.

The following table lists resources for customer support, software services, training, and product and solutions information.

Table 5. IBM resources

Resource	Description and location
IBM Support Portal	You can customize support information by choosing the products and the topics that interest you at www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server
Software services	You can find information about software, IT, and business consulting services, on the solutions site at www.ibm.com/businesssolutions/
My IBM	You can manage links to IBM Web sites and information that meet your specific technical support needs by creating an account on the My IBM site at www.ibm.com/account/
Training and certification	You can learn about technical training and education services designed for individuals, companies, and public organizations to acquire, maintain, and optimize their IT skills at http://www.ibm.com/software/sw-training/
IBM representatives	You can contact an IBM representative to learn about solutions at www.ibm.com/connect/ibm/us/en/

Providing feedback

The following table describes how to provide feedback to IBM about products and product documentation.

Table 6. Providing feedback to IBM

Type of feedback	Action
Product feedback	You can provide general product feedback through the Consumability Survey at www.ibm.com/software/data/info/consumability-survey

Table 6. Providing feedback to IBM (continued)

Type of feedback	Action
Documentation feedback	To comment on the information center, click the Feedback link on the top right side of any topic in the information center. You can also send comments about PDF file books, the information center, or any other documentation in the following ways: • Online reader comment form: www.ibm.com/software/data/rcf/ • E-mail: comments@us.ibm.com

Accessing product documentation

Documentation is provided in a variety of locations and formats, including in help that is opened directly from the product client interfaces, in a suite-wide information center, and in PDF file books.

The information center is installed as a common service with IBM InfoSphere Information Server. The information center contains help for most of the product interfaces, as well as complete documentation for all the product modules in the suite. You can open the information center from the installed product or from a Web browser.

Accessing the information center

You can use the following methods to open the installed information center.

- Click the **Help** link in the upper right of the client interface.

  Note: From IBM InfoSphere FastTrack and IBM InfoSphere Information Server Manager, the main Help item opens a local help system. Choose **Help > Open Info Center** to open the full suite information center.

- Press the F1 key. The F1 key typically opens the topic that describes the current context of the client interface.

  Note: The F1 key does not work in Web clients.

- Use a Web browser to access the installed information center even when you are not logged in to the product. Enter the following address in a Web browser: http://host_name:port_number/infocenter/topic/ com.ibm.swg.im.iis.productization.iisinfsv.home.doc/ic-homepage.html. The host_name is the name of the services tier computer where the information center is installed, and port_number is the port number for InfoSphere Information Server. The default port number is 9080. For example, on a Microsoft® Windows® Server computer named iisdocs2, the Web address is in the following format: http://iisdocs2:9080/infocenter/topic/ com.ibm.swg.im.iis.productization.iisinfsv.nav.doc/dochome/ iisinfsrv_home.html.

A subset of the information center is also available on the IBM Web site and periodically refreshed at http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r7/ index.jsp.

Obtaining PDF and hardcopy documentation

- A subset of the PDF file books are available through the InfoSphere Information Server software installer and the distribution media. The other PDF file books are available online and can be accessed from this support document: https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1.

- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at http://www.ibm.com/e-business/linkweb/publications/ servlet/pbi.wss.

Providing feedback about the documentation

You can send your comments about documentation in the following ways:
- Online reader comment form: www.ibm.com/software/data/rcf/
- E-mail: comments@us.ibm.com

How to read syntax diagrams

The following rules apply to the syntax diagrams that are used in this information:

- Read the syntax diagrams from left to right, from top to bottom, following the path of the line. The following conventions are used:
  - The >>--- symbol indicates the beginning of a syntax diagram.
  - The ---> symbol indicates that the syntax diagram is continued on the next line.
  - The >--- symbol indicates that a syntax diagram is continued from the previous line.
  - The ---\>< symbol indicates the end of a syntax diagram.
- Required items appear on the horizontal line (the main path).

  ►►—*required_item*————————————————————————————————————►◄

- Optional items appear below the main path.

  ►►—*required_item*—————————————————————————————————————►◄
   └*optional_item*┘
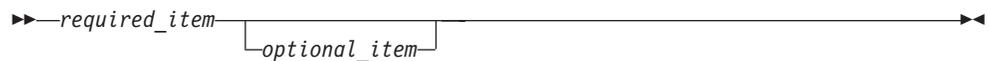
  If an optional item appears above the main path, that item has no effect on the execution of the syntax element and is used only for readability.

         ┌*optional_item*┐
  ►►—*required_item*——————————————————————————————————————►◄

- If you can choose from two or more items, they appear vertically, in a stack.

  If you must choose one of the items, one item of the stack appears on the main path.

  ►►—*required_item*—┬*required_choice1*┬———————————————————►◄
                └*required_choice2*┘
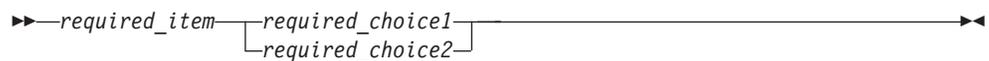
  If choosing one of the items is optional, the entire stack appears below the main path.

  ►►—*required_item*———————————————————————————————————————►◄
            ├*optional_choice1*┤
            └*optional_choice2*┘
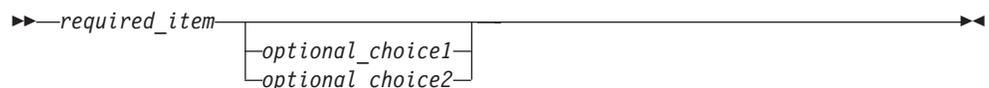
  If one of the items is the default, it appears above the main path, and the remaining choices are shown below.

         ┌*default_choice*┐
  ►►—*required_item*—┼——————————————————┼——————————————————►◄
           ├*optional_choice1*┤
           └*optional_choice2*┘
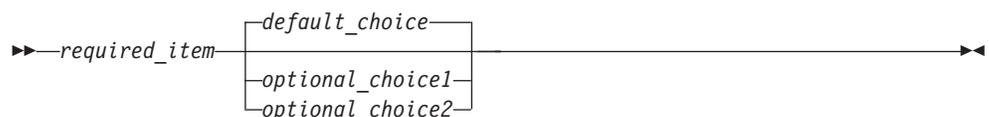
- An arrow returning to the left, above the main line, indicates an item that can be repeated.

```
                           ┌───────────────┐
                           │               │
►►──required_item──────────┴─repeatable_item─┴──────────────────────────────►◄
```

If the repeat arrow contains a comma, you must separate repeated items with a comma.

```
                           ┌──,────────────┐
                           │               │
►►──required_item──────────┴─repeatable_item─┴──────────────────────────────►◄
```

A repeat arrow above a stack indicates that you can repeat the items in the stack.

- Sometimes a diagram must be split into fragments. The syntax fragment is shown separately from the main syntax diagram, but the contents of the fragment should be read as if they are on the main path of the diagram.

```
►►──required_item──│ fragment-name │──────────────────────────────────────►◄
```

Fragment-name:

```
├──required_item──────────────────────────────────────┤
          └─optional_item─┘
```

- Keywords, and their minimum abbreviations if applicable, appear in uppercase. They must be spelled exactly as shown.
- Variables appear in all lowercase italic letters (for example, `column-name`). They represent user-supplied names or values.
- Separate keywords and parameters by at least one space if no intervening punctuation is shown in the diagram.
- Enter punctuation marks, parentheses, arithmetic operators, and other symbols, exactly as shown in the diagram.
- Footnotes are shown by a number in parentheses, for example (1).

Product accessibility

You can get information about the accessibility status of IBM products.

The IBM InfoSphere Information Server product modules and user interfaces are not fully accessible. The installation program installs the following product modules and components:

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

For information about the accessibility status of IBM products, see the IBM product accessibility information at http://www.ibm.com/able/product_accessibility/index.html.

Accessible documentation

Accessible documentation for InfoSphere Information Server products is provided in an information center. The information center presents the documentation in XHTML 1.0 format, which is viewable in most Web browsers. XHTML allows you to set display preferences in your browser. It also allows you to use screen readers and other assistive technologies to access the documentation.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.

Notices and trademarks

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

Notices

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 grant 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.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

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.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

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 measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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.

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

This information is for planning purposes only. The information herein is subject to change before the products described become available.

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.

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. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml.

The following terms are trademarks or registered trademarks of other companies:

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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

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

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office

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

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACS^{Link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS^{Link} licensee of the United States Postal Service.

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

Index

A
automatic mode   4, 7

C
configuration requirements   1
control files   4, 6
customer support
   contacting   9

F
functionality   3

G
grid editor   5

I
input links   1
input properties   5
introduction   1

L
legal notices   17

M
manual mode   4, 7
mapping string data   1

P
prerequisites   1
product accessibility
   accessibility   15
product documentation
   accessing   11

S
software services
   contacting   9
support
   customer   9

T
terminology   3
trademarks
   list of   17

W
web sites
   non-IBM   13

IBM ®

Printed in USA