



Informatica® PowerExchange for Web
Content-Kapow Katalyst
10.0

User Guide

© Copyright Informatica LLC 2012, 2018

This software and documentation contain proprietary information of Informatica LLC and are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright law. Reverse engineering of the software is prohibited. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC. This Software may be protected by U.S. and/or international Patents and other Patents Pending.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement and as provided in DFARS 227.7202-1(a) and 227.7702-3(a) (1995), DFARS 252.227-7013(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14 (ALT III), as applicable.

The information in this product or documentation is subject to change without notice. If you find any problems in this product or documentation, please report them to us in writing.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging and Informatica Master Data Management are trademarks or registered trademarks of Informatica LLC in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © Oracle. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © DataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Rogue Wave Software, Inc. All rights reserved. Copyright © Teradata Corporation. All rights reserved. Copyright © Yahoo! Inc. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © Clearpace Software Limited. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © ej-technologies GmbH. All rights reserved. Copyright © Jaspersoft Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © yWorks GmbH. All rights reserved. Copyright © Lucent Technologies. All rights reserved. Copyright (c) University of Toronto. All rights reserved. Copyright © Daniel Veillard. All rights reserved. Copyright © Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright © MicroQuill Software Publishing, Inc. All rights reserved. Copyright © PassMark Software Pty Ltd. All rights reserved. Copyright © LogiXML, Inc. All rights reserved. Copyright © 2003-2010 Lorenzi Davide, All rights reserved. Copyright © Red Hat, Inc. All rights reserved. Copyright © The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright © EMC Corporation. All rights reserved. Copyright © Flexera Software. All rights reserved. Copyright © Jinfonet Software. All rights reserved. Copyright © Apple Inc. All rights reserved. Copyright © Telerik Inc. All rights reserved. Copyright © BEA Systems. All rights reserved. Copyright © PDFlib GmbH. All rights reserved. Copyright © Orientation in Objects GmbH. All rights reserved. Copyright © Tanuki Software, Ltd. All rights reserved. Copyright © Ricebridge. All rights reserved. Copyright © Sencha, Inc. All rights reserved. Copyright © Scalable Systems, Inc. All rights reserved. Copyright © jqWidgets. All rights reserved. Copyright © Tableau Software, Inc. All rights reserved. Copyright © MaxMind, Inc. All Rights Reserved. Copyright © TMate Software s.r.o. All rights reserved. Copyright © MapR Technologies Inc. All rights reserved. Copyright © Amazon Corporate LLC. All rights reserved. Copyright © Highsoft. All rights reserved. Copyright © Python Software Foundation. All rights reserved. Copyright © BeOpen.com. All rights reserved. Copyright © CNRI. All rights reserved.

This product includes software developed by the Apache Software Foundation (<http://www.apache.org/>), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at <http://www.apache.org/licenses/>. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (<http://www.mozilla.org/>), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at <http://www.gnu.org/licenses/lgpl.html>. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at <http://www.openssl.org> and <http://www.openssl.org/source/license.html>.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://curl.haxx.se/docs/copyright.html>. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.dom4j.org/license.html>.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://dojotoolkit.org/license>.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at <http://www.gnu.org/software/kawa/Software-License.html>.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at <http://www.opensource.org/licenses/mit-license.php>.

This product includes software developed by Boost (<http://www.boost.org/>) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http://www.boost.org/LICENSE_1_0.txt.

This product includes software copyright © 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at <http://www.pcre.org/license.txt>.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.eclipse.org/org/documents/epl-v10.php> and at <http://www.eclipse.org/org/documents/edl-v10.php>.

This product includes software licensed under the terms at <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib/?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>, <http://antlr.org/license.html>, <http://aopalliance.sourceforge.net/>, <http://www.bouncycastle.org/licence.html>, <http://www.jgraph.com/jgraphdownload.html>, <http://www.jcraft.com/jsch/LICENSE.txt>, http://jotm.objectweb.org/bsd_license.html, <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>, <http://www.slf4j.org/license.html>, <http://nanoxml.sourceforge.net/orig/copyright.html>, <http://www.json.org/license.html>, <http://forge.ow2.org/projects/javaservice/>, <http://www.postgresql.org/about/licence.html>, <http://www.sqlite.org/copyright.html>, <http://www.tcl.tk/software/tcltk/license.html>, <http://www.jaxen.org/faq.html>, <http://www.jdom.org/docs/faq.html>, <http://www.slf4j.org/license.html>, <http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>, <http://www.keplerproject.org/md5/license.html>, <http://www.toedter.com/en/jcalendar/license.html>, <http://www.edankert.com/bounce/index.html>, <http://www.net-snmp.org/about/license.html>, <http://www.openmdx.org/#FAQ>, http://www.php.net/license/3_01.txt, <http://srp.stanford.edu/license.txt>, <http://www.schneider.com/blowfish.html>, <http://www.jmock.org/license.html>, <http://xsom.java.net>, <http://benalman.com/about/license/>, <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>, <http://www.h2database.com/html/license.html#summary>, <http://jsoncpp.sourceforge.net/LICENSE>, <http://jdbc.postgresql.org/license.html>, <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>, <https://github.com/rantav/hector/blob/master/LICENSE>, <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>, <http://jibx.sourceforge.net/jibx-license.html>, <https://github.com/lyokato/libgeohash/blob/master/LICENSE>, <https://github.com/hjiang/jsonxx/blob/master/LICENSE>, <https://code.google.com/p/lz4/>, <https://github.com/jedisct1/libsodium/blob/master/LICENSE>, <http://one-jar.sourceforge.net/index.php?page=documents&file=license>, <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>, <http://www.scala-lang.org/license.html>, <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>, <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>, <https://aws.amazon.com/ssl/>, <https://github.com/twbs/bootstrap/blob/master/LICENSE>, <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>, <https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE>, and <https://github.com/apache/hbase/blob/master/LICENSE.txt>.

This product includes software licensed under the Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), the Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>), the Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (<http://www.opensource.org/licenses/bsd-license.php>), the new BSD License (<http://opensource.org/licenses/BSD-3-Clause>), the MIT License (<http://www.opensource.org/licenses/mit-license.php>), the Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) and the Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://xstream.codehaus.org/license.html>. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit <http://www.extreme.indiana.edu/>.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

See patents at <https://www.informatica.com/legal/patents.html>.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

Publication Date: 2018-09-30

Table of Contents

Preface	6
Informatica Resources.	6
Informatica My Support Portal.	6
Informatica Knowledge Base.	6
Informatica Documentation.	6
Informatica Product Availability Matrixes.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Informatica Global Customer Support.	7
 Chapter 1: Introduction to PowerExchange for Web Content-Kapow Katalyst...	8
Power Exchange for Web Content-Kapow Katalyst Overview.	8
 Chapter 2: PowerExchange for Web Content-Kapow Katalyst Configuration....	9
Prerequisites.	9
Set Environment Variable.	9
Configuring HTTP Proxy Options at Design-Time.	10
Configuring HTTP Proxy Options from the Developer Tool.	10
Configuring HTTP Proxy Options at Run-Time.	11
 Chapter 3: Web Content-Kapow Katalyst Connections.....	12
Web Content-Kapow Katalyst Connection Overview.	12
Web Content-Kapow Katalyst Connection Properties.	12
Creating a Web Content-Kapow Katalyst Connection.	13
 Chapter 4: Web Content-Kapow Katalyst Data Objects.....	14
Web Content-Kapow Katalyst Data Objects Overview.	14
Re-create the Web Content-Kapow Katalyst Data Objects to use Kapow Katalyst 9.2 and 9.3 Libraries.	14
Web Content-Kapow Katalyst Data Object Views.	15
Web Content-Kapow Katalyst Data Object Overview Properties.	15
Web Content-Kapow Katalyst Data Object Read Operation Properties.	16
Source Properties of the Data Object Read Operation.	16
Output Properties of the Data Object Read Operation.	17
Parameterization.	18
Importing a Web Content-Kapow Katalyst Data Object.	19
Creating a Web Content-Kapow Katalyst Data Object Read Operation.	19
 Chapter 5: Web Content-Kapow Katalyst Mappings.....	21
Web Content-Kapow Katalyst Mapping Overview.	21

Web Content-Kapow Katalyst Mapping Example.	21
Mapping Input.	22
Mapping Output.	22
Appendix A: Datatype Reference.	23
Datatype Reference Overview.	23
Web Content-Kapow Katalyst Datatypes and Transformation Datatypes.	23
Index.	25

Preface

The *Informatica PowerExchange for Web Content-Kapow Katalyst User Guide* describes how to use PowerExchange for Web Content-Kapow and Kapow Katalyst with Informatica Data Services and Informatica Data Explorer to extract information from the Web and move that information to targets for analysis. The guide is written for database administrators and developers who are responsible for developing mappings and workflows. This guide assumes that you have knowledge of Kapow Katalyst and Informatica.

Informatica Resources

Informatica My Support Portal

As an Informatica customer, the first step in reaching out to Informatica is through the Informatica My Support Portal at <https://mysupport.informatica.com>. The My Support Portal is the largest online data integration collaboration platform with over 100,000 Informatica customers and partners worldwide.

As a member, you can:

- Access all of your Informatica resources in one place.
- Review your support cases.
- Search the Knowledge Base, find product documentation, access how-to documents, and watch support videos.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

As an Informatica customer, you can access the Informatica Knowledge Base at <https://mysupport.informatica.com>. Use the Knowledge Base to search for documented solutions to known technical issues about Informatica products. You can also find answers to frequently asked questions, technical white papers, and technical tips. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team through email at KB_Feedback@informatica.com.

Informatica Documentation

The Informatica Documentation team makes every effort to create accurate, usable documentation. If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com. We will use your feedback to improve our documentation. Let us know if we can contact you regarding your comments.

The Documentation team updates documentation as needed. To get the latest documentation for your product, navigate to Product Documentation from <https://mysupport.informatica.com>.

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. You can access the PAMs on the Informatica My Support Portal at <https://mysupport.informatica.com>.

Informatica Velocity

You can access Informatica Velocity at <https://mysupport.informatica.com>. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where developers and partners can share solutions that augment, extend, or enhance data integration implementations. By leveraging any of the hundreds of solutions available on the Marketplace, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at <http://www.informaticamarketplace.com>.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or through the Online Support.

Online Support requires a user name and password. You can request a user name and password at <http://mysupport.informatica.com>.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <http://www.informatica.com/us/services-and-training/support-services/global-support-centers/>.

CHAPTER 1

Introduction to PowerExchange for Web Content-Kapow Katalyst

This chapter includes the following topic:

- [Power Exchange for Web Content-Kapow Katalyst Overview, 8](#)

Power Exchange for Web Content-Kapow Katalyst Overview

PowerExchange for Web Content-Kapow Katalyst uses a third party tool called Kapow Katalyst to extract data from any web site.

You can create a robot in Kapow Katalyst to extract web site data. Use PowerExchange for Web Content-Kapow Katalyst to read the data from a robot into a Web Content-Kapow Katalyst data object through the Data Integration Service. You can load the extracted data to a target and then use the data for data mining and analysis.

For example, you can track product announcements from your top competitors. You can create robots using Kapow Katalyst and have them run everyday on the competitor's web pages, product web pages, or blogs to extract relevant data. You can use PowerExchange for Web Content-Kapow Katalyst to import the data from the robot, and run mappings. Results can be stored in a database to be used for data mining and analysis.

For information about Kapow Katalyst, see the Kapow Katalyst documentation.

CHAPTER 2

PowerExchange for Web Content-Kapow Katalyst Configuration

This chapter includes the following topics:

- [Prerequisites, 9](#)
- [Set Environment Variable , 9](#)
- [Configuring HTTP Proxy Options at Design-Time , 10](#)
- [Configuring HTTP Proxy Options at Run-Time, 11](#)

Prerequisites

PowerExchange for Web Content-Kapow Katalyst requires Kapow Katalyst to be installed and robots to be created.

Before you use PowerExchange for Web Content-Kapow Katalyst, complete the following tasks.

1. Install Informatica services and clients.
 - Create a Data Integration Service and Model Repository Service.
2. Install and configure Kapow Katalyst with a valid license.

To configure Kapow Katalyst, complete the following tasks:

 - Create a robot.
 - Run the RoboServer.
 - Upload the robot to the Local Management Console.
 - Verify that the robot was uploaded.

Set Environment Variable

Set the environment variable to any Kapow Katalyst version.

Set the environment variable on each machine that hosts the Developer tool and on the machine that hosts Informatica Services in the following format:

```
INFA_KAPOW_DIST_DIR=<Kapow Katalyst Installation Directory>\API\robosuite-java-api
```

Note: If the path ends with a semicolon or colon, the path is not valid and you cannot run mappings.

Configuring HTTP Proxy Options at Design-Time

If your organization uses a proxy server to access the internet, you can configure the HTTP proxy server authentication settings at design time. You can configure the HTTP proxy server authentication from the developerCore.ini file.

Configuring HTTP Proxy Options from the Developer Tool

If your organization uses a proxy server to access the internet, you can configure the HTTP proxy server authentication settings from the developerCore.ini file.

- Ensure that you enable the proxy server settings from your web browser.
- Access the developerCore.ini file from the following location:
`<Informatica Installation Location>\clients\DeveloperClient`
- Add the following properties to the developerCore.ini file:

Property	Description
-Dhttp.proxyHost=	Name of the HTTP proxy server.
-Dhttp.proxyPort=	Port number of the HTTP proxy server.
-Dhttp.proxyUser=	Authenticated user name for the HTTP proxy server. This is required if the proxy server requires authentication.
-Dhttp.proxyPassword=	Password for the authenticated user. This is required if the proxy server requires authentication. Note: The password is in plain text and not encrypted.
-Dhttp.nonProxyHosts=	List of host names or IP addresses for which you must not use the proxy server. Separate the list of IP addresses or host names with a pipe symbol (). For example, localhost:10.20.30.40 myHost Specify the IP address or name of the machine on which the Informatica gateway node runs so that the Developer tool connects to the domain.
-Dhttps.proxyHost=	Name of the HTTPS proxy server.
-Dhttps.proxyPort=	Port number of the HTTPS proxy server.

Configuring HTTP Proxy Options at Run-Time

If your organization uses a proxy server to access the internet, you must configure the HTTP proxy server authentication settings for the Data Integration Service.

1. Open the Administrator tool.
2. Click the **Administration** tab, and then select the Data Integration Service.
3. Click the **Properties** tab.
4. Click **Edit** in the HTTP Proxy Server Properties section.
5. Configure the following properties:

Property	Description
HTTP Proxy Server Host	Name of the HTTP proxy server.
HTTP Proxy Server Port	Port number of the HTTP proxy server. Default is 8080.
HTTP Proxy Server User	Authenticated user name for the HTTP proxy server. This is required if the proxy server requires authentication.
HTTP Proxy Server Password	Password for the authenticated user. This is required if the proxy server requires authentication.
HTTP Proxy Server Domain	Domain for authentication.

CHAPTER 3

Web Content-Kapow Katalyst Connections

This chapter includes the following topics:

- [Web Content-Kapow Katalyst Connection Overview, 12](#)
- [Web Content-Kapow Katalyst Connection Properties, 12](#)
- [Creating a Web Content-Kapow Katalyst Connection, 13](#)

Web Content-Kapow Katalyst Connection Overview

Use a Web Content-Kapow Katalyst connection to access and import robots in Kapow Katalyst.

Create a Web Content-Kapow Katalyst connection to create data objects, preview data, and run mappings.

Web Content-Kapow Katalyst Connection Properties

Use a Web Content-Kapow Katalyst connection to access robots in Kapow Katalyst. This is a social media type connection. You can create and manage a Web Content-Kapow Katalyst connection in the Administrator tool or the Developer tool.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes Web Content-Kapow Katalyst connection properties:

Property	Description
Name	Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? /
ID	String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name.

Property	Description
Description	The description of the connection. The description cannot exceed 765 characters.
Location	The Informatica domain where you want to create the connection.
Type	The connection type. Select Web Content-Kapow Katalyst.
Management Console URL	URL of the Management Console where the robot is uploaded. The URL must start with http or https. For example, http://localhost:50080.
RQL Service Port	The port number where the socket service listens for the RQL service. Enter a value from 1 through 65535. Default is 50000.
Username	User name required to access the Local Management Console.
Password	Password to access the Local Management Console.

Creating a Web Content-Kapow Katalyst Connection

Create a connection in the Developer tool before you import Web Content-Kapow Katalyst data objects. When you create a Web Content-Kapow Katalyst connection, you enter information such as a connection ID and the URL of the Local Management console where the robot is uploaded.

1. Click **Window > Preferences**.
2. Select **Informatica > Connections**.
3. Expand the domain in the **Available Connections**.
4. Select **Social Media > Web Content-Kapow Katalyst** and click **Add**.
5. Enter a connection name.
6. Enter an ID for the connection.
7. Optionally, enter a connection description.
8. Select the domain where you want to create the connection.
9. Select a Web Content-Kapow Katalyst connection type.
10. Click **Next**.
11. Configure the connection properties.
12. Click **Test Connection** to verify the connection to Kapow Katalyst.
13. Click **Finish**.

CHAPTER 4

Web Content-Kapow Katalyst Data Objects

This chapter includes the following topics:

- [Web Content-Kapow Katalyst Data Objects Overview, 14](#)
- [Re-create the Web Content-Kapow Katalyst Data Objects to use Kapow Katalyst 9.2 and 9.3 Libraries, 14](#)
- [Web Content-Kapow Katalyst Data Object Views, 15](#)
- [Web Content-Kapow Katalyst Data Object Overview Properties, 15](#)
- [Web Content-Kapow Katalyst Data Object Read Operation Properties, 16](#)
- [Parameterization, 18](#)
- [Importing a Web Content-Kapow Katalyst Data Object, 19](#)
- [Creating a Web Content-Kapow Katalyst Data Object Read Operation, 19](#)

Web Content-Kapow Katalyst Data Objects Overview

A Web Content-Kapow Katalyst data object is a physical data object that uses a robot as a source. A Web Content-Kapow Katalyst data object represents data that is based on a robot.

Import a robot into the Developer tool to create a Web Content-Kapow Katalyst data object. After you create a data object, create a data object read operation. You can use the data object read operation as a source in a mapping or mapplet.

Re-create the Web Content-Kapow Katalyst Data Objects to use Kapow Katalyst 9.2 and 9.3 Libraries

Web Content-Kapow Katalyst data objects that you created with Kapow Katalyst 8.3 libraries are incompatible with Kapow Katalyst 9.2 or 9.3 libraries.

If you use Kapow Katalyst 9.2 or 9.3 libraries, you must re-create the Web Content-Kapow Katalyst data objects that you created with the Kapow Katalyst 8.3 libraries.

After you re-create a Web Content-Kapow Katalyst data object, create a data object read operation, and update the run-time properties. Use the data object read operation in the mappings.

Web Content-Kapow Katalyst Data Object Views

The Web Content-Kapow Katalyst data object contains views to edit the object name and the properties.

After you create a Web Content-Kapow Katalyst data object, you can change the data object and data object operation properties in the following data object views:

- **Overview** view. Use the **Overview** view to edit the Web Content-Kapow Katalyst data object name, description, and robots.
- **Data Object Operation** view. Use the **Data Object Operation** view to view and modify the properties that the Data Integration Service uses when it reads data from a robot.

When you create mappings that use Web Content-Kapow Katalyst sources, you can view the data object properties in the **Properties** view.

Web Content-Kapow Katalyst Data Object Overview Properties

The **Overview** view displays general information about the Web Content-Kapow Katalyst data object and detailed information about the robot that you imported.

The following table describes the general properties that you configure for a Web Content-Kapow Katalyst data object:

Property	Description
Name	Name of the Web Content-Kapow Katalyst data object.
Description	Description of the Web Content-Kapow Katalyst data object.
Connection	Name of the Web Content-Kapow Katalyst connection.

The following table describes the properties of the robot that you import:

Property	Description
Name	Name of the robot.
Type	Native datatype of the robot.
Description	Description of the robot.

Web Content-Kapow Katalyst Data Object Read Operation Properties

The Data Integration Service reads data from a robot based on the data object read operation. The Developer tool displays the data object read operation properties of the Web Content-Kapow Katalyst data object in the **Data Object Operation** view.

You can view or configure the data object read operation from the source and output properties.

- **Source properties.** Represents data that the Data Integration Service reads from the robot. Select the source properties to view data such as the name and description of the robot and the column properties.
- **Output properties.** Represents data that the Data Integration Service passes into the mapping pipeline. Select the output properties to edit the port properties of the data object read operation. You can also enter the Kapow input type and attributes and specify the Kapow cluster name.

Source Properties of the Data Object Read Operation

The source properties are populated based on the robot that you added when you created the data object. The source properties of the data object read operation include general and column properties that apply to the robot.

You can view the source properties of the data object read operation from the **General**, **Column**, and **Advanced** tabs.

General Properties

The general properties display the name and description of the robot.

Column Properties

The column properties display the datatypes, precision, and scale of the source property in the data object read operation.

The following table describes the source column properties of the data object read operation:

Property	Description
Name	Name of the column.
Type	Native datatype of the column.
Precision	Maximum number of significant digits for numeric datatypes, or maximum number of characters for string datatypes. For numeric datatypes, precision includes scale.
Scale	Maximum number of digits after the decimal point for numeric values.
Description	Description of the column.

Output Properties of the Data Object Read Operation

The output properties represent data that the Data Integration Service passes into the mapping pipeline. Select the output properties to edit the port properties of the data object read operation. You can also enter the Kapow input type and attributes and specify the Kapow cluster name.

The output properties of the data object read operation include general properties that apply to the data object operation. They also include port, source, and advanced properties that apply to the robot.

You can view and change the output properties of the data object read operation from the **General**, **Ports**, **Sources**, and **Advanced** tabs.

General Properties

The general properties display the name and description of the data object read operation.

Ports Properties

The output ports properties display the datatypes, precision, and scale of the data object read operation.

The following table describes the output port properties that you configure in the data object read operation:

Property	Description
Name	Name of the port.
Type	Datatype of the port.
Precision	Maximum number of significant digits for numeric datatypes, or maximum number of characters for string datatypes. For numeric datatypes, precision includes scale.
Scale	Maximum number of digits after the decimal point for numeric values.
Description	Description of the port.

Sources Properties

The source properties list the robots in the data object read operation. You can add and delete robots.

Advanced Properties

The advanced properties allow you to specify the Kapow input type and attributes and the Kapow cluster name in the data object read operation.

The following table describes the advanced properties that you configure in the data object read operation:

Property	Description
Kapow Input Type Name	The name of a datatype returned or used as input by the robot. Enter the name of the Kapow Katalyst datatype specified when you created the robot. If the robot has two input types, separate them with a pipe symbol (). For example, InputType1 InputType2.
Kapow Input Attributes	A variable that is used as input to the robot while running the robot. You can enter multiple attribute values as input to the robot. If the robot has two input variables, separate them with a pipe symbol. For example, enter a list of input attributes for InputType1 and InputType2 in the following format: InputType.someattribute=value InputType.someattribute=value SonyInput.queryparam=cybershot LGInput.queryparam=TV For more information about variable configuration see the Kapow Katalyst documentation.
Kapow cluster name	The name of the cluster on which the robot runs.
Ends After	Amount of time that the Data Integration Service runs the mapping. Enter the duration in the following format: hh:mm Default is blank, which indicates that the Data Integration Service runs the mapping until you stop it.

Parameterization

You can parameterize the Web Content-Kapow Katalyst connection and read operation properties to override the properties at run time.

The following table lists the read operation properties that you can parameterize and the type of parameterization supported:

Property	Type of Parameterization Supported
Kapow Input Type Name	Full
Kapow Input Attributes	Full

Property	Type of Parameterization Supported
Kapow Cluster Name	Full
Ends After	Full

Importing a Web Content-Kapow Katalyst Data Object

Import a Web Content-Kapow Katalyst data object to read data from a robot.

Configure a Web Content-Kapow Katalyst connection before you import a Web Content-Kapow Katalyst data object. Ensure that RoboServer is running and you have uploaded the robot to the Local Management Console. You can access the URL of the Local Management Console to confirm that the robot is loaded.

1. Select a project or folder in the **Object Explorer** view.
2. Click **File > New > Data Object**.
3. Select **Web Content-Kapow Katalyst Data Object** and click **Next**.
The **New Web Content-Kapow Katalyst Data Object** dialog box appears.
4. Enter a name for the data object.
5. Click **Browse** next to the **Location** option and select the target project or folder.
6. Select the default data object type.
7. Click **Browse** next to the **Connection** option and select the Web Content-Kapow Katalyst connection from which you want to import the robots.
8. To add a robot, click **Add** next to the **Selected Resource(s)** option.
The **Add Resource** dialog box appears.
9. Select a robot. You can search for it or navigate to it.
 - Navigate to the Web Content-Kapow Katalyst robot that you want to add and click **OK**.
 - To search for the robot, enter the name or description of the robot you want to add. Click **OK**.
10. If required, add more robots to the Web Content-Kapow Katalyst data object.
You can also add robots to a Web Content-Kapow Katalyst data object after you create it.
11. Click **Finish**.

The data object appears as a physical data object in the project or folder in the **Object Explorer** view.

Creating a Web Content-Kapow Katalyst Data Object Read Operation

You can create the data object read operation for one or more Web Content-Kapow Katalyst data objects. You can add a Web Content-Kapow Katalyst data object read operation to a mapping or mapplet as a source.

Before you create a Web Content-Kapow Katalyst data object read operation, you must create the Web Content-Kapow Katalyst data object with at least one robot.

1. Select the data object in the Object Explorer view.
2. Right-click and select **New > Data Object Operation**.

The **Data Object Operation** dialog box appears.

3. Enter a name for the data object operation.
4. Select the type of data object read operation.
5. Click **Add**.

The **Select a resource** dialog box appears.

6. Select the robot for which you want to create the read data object operation and click **OK**.
7. Click **Finish**.

The Developer tool creates the data object operation for the selected data object.

CHAPTER 5

Web Content-Kapow Katalyst Mappings

This chapter includes the following topics:

- [Web Content-Kapow Katalyst Mapping Overview, 21](#)
- [Web Content-Kapow Katalyst Mapping Example, 21](#)

Web Content-Kapow Katalyst Mapping Overview

After you create a Web Content-Kapow Katalyst data object read operation, you can develop a mapping.

You can define the following objects in the mapping:

- Web Content-Kapow Katalyst data object read operation as the input to read data from a robot
- Relational data object as the output

Validate and run the mapping to extract the Web Content-Kapow Katalyst data and load it to a relational or flat file target.

Web Content-Kapow Katalyst Mapping Example

Your organization needs to track product announcements from its competitors. Create robots in Kapow Katalyst and configure them to run every day on the competitor web pages and product pages to extract relevant data.

Create a robot called New Products with the following criteria:

- URLs of the competitors web sites
- Tag the Latest News section on the competitors web sites

Create a mapping that reads data from the web sites and writes those records to a table.

Mapping Input

The mapping source is a Web Content-Kapow Katalyst data object that reads the information extracted from the web sites using a robot.

You import a physical data object by adding the robot called New Products. Create a read data object operation for the Web Content-Kapow Katalyst data object you have created. Add the data object read operation to the mapping.

Mapping Output

Add a relational data object to the mapping as an output.

When you run the mapping, the Data Integration Service writes the extracted web site data to the target table. You can use the results stored in the database for data analysis.

APPENDIX A

Datatype Reference

This appendix includes the following topics:

- [Datatype Reference Overview, 23](#)
- [Web Content-Kapow Katalyst Datatypes and Transformation Datatypes, 23](#)

Datatype Reference Overview

The Developer tool uses the following datatypes in PowerExchange for Web Content-Kapow Katalyst mappings.

- **Web Content-Kapow Katalyst native datatypes.** Web Content-Kapow Katalyst native datatypes appear in the physical data object column properties.
- **Transformation datatypes.** Set of datatypes that appear in the transformations. They are internal datatypes based on ANSI SQL-92 generic datatypes, which the Data Integration Service uses to move data across platforms. Transformation datatypes appear in all transformations in a mapping.

When the Data Integration Service reads source data, it converts the native datatypes to the comparable transformation datatypes before transforming the data. When the Data Integration Service writes to a target, it converts the transformation datatypes to the comparable native datatypes.

Web Content-Kapow Katalyst Datatypes and Transformation Datatypes

The following table lists the Web Content-Kapow Katalyst datatypes that Data Integration Service supports and the corresponding transformation datatypes:

Web Content-Kapow Katalyst Datatype	Range and Description	Transformation Datatype
Binary	Any sequence of bytes. Default size is 10 MB. Maximum size is 100 MB.	Binary
Boolean	A boolean (true/false) attribute.	String

Web Content-Kapow Katalyst Datatype	Range and Description	Transformation Datatype
Character	A single character.	String
Country	Country code, as defined by the ISO-3166 standard, For example, DE for Germany.	String
Currency	Currency code, as defined by the ISO-4217 standard. For example, EUR for Euro.	String
Date	A date in the format yyyy-mm-dd hh:mm:ss.n.	Date/Time
HTML	An HTML clip similar to long text. You can preview the clip in a browser window. Default size is 10 MB. Maximum size is 100 MB.	Text
Image	Similar to binary data, except that you can preview the image. Default size is 10 MB. Maximum size is 100 MB.	Binary
Integer	You can specify maximum and minimum values from -9,223,372,036,854,775,808 to 9,223,372,036,854,775,807.	Bigint
Language	Language code, as defined by the ISO-639 standard. For example, DE for German.	String
Long Text	Long text displayed in a multiline text field. Long text is similar to a text area. 1 to 45,825,598 characters. Default size is 1 MB. Maximum size is 10 MB.	Text
Number	A number. You can specify between $\pm 2.2 \times 10^{-308}$ to $\pm 1.8 \times 10^{308}$.	Decimal
PDF	Similar to binary data, except that you can preview the PDF document. Default size is 10 MB. Maximum size is 100 MB.	Binary
Properties	List of name/value pair properties. This is the same as long text, except that the text represents a list of properties, where each property is a name/value pair.	Text
Short text	Short text displayed in a one-line text field. It will accept data with multiple lines. Scroll to view more information. 1 to 5,245,616 characters. Default size is 10 KB. Maximum size is 1 MB.	String
XML	An XML document similar to long text. Well-formed XML documents are allowed. Default size is 10 MB. Maximum size is 100 MB.	Text

INDEX

A

advanced properties [18](#)

C

column properties [16](#)
configuring
 HTTP proxy options [11](#)
configuring HTTP proxy options
 Developer tool [10](#)
creating
 Web Content-Kapow Catalyst connection [13](#)
 Web Content-Kapow Catalyst data object read operation [19](#)

D

data object operation
 output properties [17](#)
 source properties [16](#)
datatype reference overview
 description [23](#)
datatypes
 Transformation [23](#)
 Web Content-Kapow Catalyst [23](#)

G

general properties [16](#), [17](#)

I

importing
 Web Content-Kapow Catalyst data object [19](#)

O

overview
 PowerExchange for Web Content-Kapow Catalyst [8](#)

overview (*continued*)
 Web Content-Kapow Catalyst data object [14](#)
 Web Content-Kapow Catalyst mapping [21](#)

P

PowerExchange for Web Content-Kapow Catalyst
 overview [8](#)
prerequisites
 Web Content-Kapow Catalyst [9](#)
properties
 Web Content-Kapow Catalyst data object [15](#)

S

sources properties [18](#)

W

Web Content-Kapow Catalyst
 environment variable [9](#)
 mapping example [21](#)
 views [15](#)
Web Content-Kapow Catalyst connection
 creating [13](#)
 overview [12](#)
Web content-Kapow Catalyst connections
 properties [12](#)
Web Content-Kapow Catalyst data object
 importing [19](#)
 overview [14](#)
 properties [15](#)
Web Content-Kapow Catalyst data object operation
 column properties [16](#)
 general properties [16](#)
 ports properties [16](#)
 sources properties [16](#)
Web Content-Kapow Catalyst data object read operation
 creating [19](#)
Web Content-Kapow Catalyst mapping
 overview [21](#)