



Informatica® PowerExchange for Microsoft
Azure Cosmos DB SQL API
10.5

User Guide

© Copyright Informatica LLC 2019, 2021

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. 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.

Informatica, the Informatica logo, PowerExchange, and Big Data Management are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

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 © 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/hsqldbLicense.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/license.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/asl/>; <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.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2021-03-31

Table of Contents

Preface	6
Informatica Resources.	6
Informatica Network.	6
Informatica Knowledge Base.	6
Informatica Documentation.	6
Informatica Product Availability Matrices.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Informatica Global Customer Support.	7
 Chapter 1: Introduction to PowerExchange for Cosmos DB SQL API.....	8
PowerExchange for Microsoft Azure Cosmos DB SQL API Overview.	8
Introduction to Cosmos DB.	9
 Chapter 2: PowerExchange for Microsoft Azure Cosmos DB SQL API Installation and Configuration.....	10
PowerExchange for Cosmos DB SQL API Configuration Overview.	10
Prerequisites.	10
Installing a TLS certificate.	11
 Chapter 3: PowerExchange for Cosmos DB SQL API Connections.....	12
Microsoft Azure Cosmos DB SQL API Connections Overview.	12
Microsoft Azure Cosmos DB SQL API Connection Properties.	12
Creating a PowerExchange for Cosmos DB SQL API Connection.	13
 Chapter 4: PowerExchange for Microsoft Azure Cosmos DB SQL API Data Objects.....	14
PowerExchange for Cosmos DB SQL API Data Objects Overview.	14
PowerExchange for Cosmos DB SQL API Data Object Properties.	15
PowerExchange for Cosmos DB SQL API Data Object Read Operation Properties.	15
Column Projection Properties.	16
Microsoft Azure Cosmos DB SQL API Data Object Write Operation Properties.	17
Column Projection Properties.	17
Parameterization of Cosmos DB Data Objects.	18
Creating a Cosmos DB Data Object.	18
Creating a Cosmos DB Data Object Operation.	19
 Chapter 5: PowerExchange for Microsoft Azure Cosmos DB SQL API Mappings.....	21
Microsoft Azure Cosmos DB Mappings Overview.	21
Mapping Validation and Run-time Environments.	21

Appendix A: Cosmos DB Data Types Reference.....	23
Cosmos DB Data Types Reference Overview.	23
Azure Cosmos DB and Transformation Data Types.	24
Data Types Parsing for Cosmos DB.	24
Rules and Guidelines for Data Types.	25
Index.	26

Preface

Use the *Informatica® PowerExchange® for Microsoft Azure Cosmos DB SQL API User Guide* to learn how to read and write documents to a collection in the Cosmos DB database by using the developer tool. Learn to create a Microsoft Azure Cosmos DB SQL API connection and develop and run mappings in an Informatica domain.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. 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 you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

CHAPTER 1

Introduction to PowerExchange for Cosmos DB SQL API

This chapter includes the following topics:

- [PowerExchange for Microsoft Azure Cosmos DB SQL API Overview, 8](#)
- [Introduction to Cosmos DB, 9](#)

PowerExchange for Microsoft Azure Cosmos DB SQL API Overview

PowerExchange for Microsoft Azure Cosmos DB SQL API provides connectivity between Informatica and Cosmos DB. Use PowerExchange for Microsoft Azure Cosmos DB SQL API to extract and load Cosmos DB documents through the Data Integration Service.

You can use PowerExchange for Cosmos DB SQL API to read JSON documents from and write JSON documents to a collection in the Cosmos DB database. You can use PowerExchange for Cosmos DB SQL API to process large volumes of data.

You can use PowerExchange for Cosmos DB SQL API for the following data integration scenarios:

- Create a Cosmos DB data warehouse. You can aggregate data from Cosmos DB and other source systems, transform the data, and write the data to Cosmos DB.
- Migrate data from a relational database or other data sources to Cosmos DB. For example, you want to migrate data from a relational database to Cosmos DB. You can write data from multiple relational database tables with different schemas to the same Cosmos DB collection. A Cosmos DB collection contains the data in a Cosmos DB database.
- Move data between operational data stores to synchronize data. For example, an online marketplace uses a relational database as the operational data store. You want to use Cosmos DB instead of the relational database. However, you want to maintain the relational database along with Cosmos DB for a period of time. You can use PowerExchange for Cosmos DB SQL API to synchronize data between the relational data store and the Cosmos DB data store.
- Migrate data from Cosmos DB to a data warehouse for reporting. For example, your organization uses a business intelligence tool that does not support Cosmos DB. You must migrate the data from Cosmos DB to a data warehouse so that the business intelligence tool can use the data to generate reports.

Introduction to Cosmos DB

Cosmos DB is a globally distributed, document based, NoSQL database that maintains multiple data models.

A Cosmos DB database contains a set of collections. A collection is a set of documents and is similar to a table in a relational database. Cosmos DB stores records as documents that are similar to rows in a relational database. A document contains fields that are similar to columns in a relational database. A document can have a dynamic schema. A document in a collection does not need to have the same set of fields or structure as another document in the same collection. A document can also contain nested documents.

CHAPTER 2

PowerExchange for Microsoft Azure Cosmos DB SQL API Installation and Configuration

This chapter includes the following topics:

- [PowerExchange for Cosmos DB SQL API Configuration Overview, 10](#)
- [Prerequisites, 10](#)
- [Installing a TLS certificate, 11](#)

PowerExchange for Cosmos DB SQL API Configuration Overview

You must configure PowerExchange for Cosmos DB SQL API before you can extract data from or load data to a Cosmos DB collection.

Prerequisites

You must perform the following prerequisite before you can use PowerExchange for Microsoft Azure Cosmos DB SQL API:

- Install and configure the Informatica services.
- Install and configure the Developer tool. You can install the Developer tool when you install Informatica clients.
- Create a Data Integration Service and a Model Repository Service in the Informatica domain.
- Verify that a cluster configuration is created in the domain.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica Network:

<https://network.informatica.com/community/informatica-network/product-availability-matrices>

Installing a TLS certificate

If the domain is TLS-enabled, download the certificate first and then add the certificate in the trust store.

Perform the following steps from Developer tool host machine:

1. Run the following command to download the certificate and convert it to PEM format:

```
openssl s_client -servername {yourdomain}.documents.azure.com -connect  
{yourdomain}.documents.azure.com:{port number} < /dev/null | openssl x509 -outform pem >  
cosmos.pem.
```

For example, `openssl s_client -servername bswar.documents.azure.com -connect
bswar.documents.azure.com:443 < /dev/null | openssl x509 -outform pem`

2. Run the following command to import the certificate in the trust store:

```
keytool -import -noprompt -trustcacerts -alias cosmo -file /tmp/test/cosmos.pem -keystore  
<INFA_HOME>/services/shared/security/infa_truststore.jks
```

CHAPTER 3

PowerExchange for Cosmos DB SQL API Connections

This chapter includes the following topics:

- [Microsoft Azure Cosmos DB SQL API Connections Overview, 12](#)
- [Microsoft Azure Cosmos DB SQL API Connection Properties, 12](#)
- [Creating a PowerExchange for Cosmos DB SQL API Connection, 13](#)

Microsoft Azure Cosmos DB SQL API Connections Overview

To connect to a Cosmos DB collection, you must create a Microsoft Azure Cosmos DB SQL API connection.

Use the Developer tool or Administrator tool to create a Microsoft Azure Cosmos DB connection.

Microsoft Azure Cosmos DB SQL API Connection Properties

Use a Microsoft Azure Cosmos DB connection to connect to the Cosmos DB database. When you create a Microsoft Azure Cosmos DB connection, you enter information for metadata and data access.

The following table describes the Microsoft Azure Cosmos DB connection properties:

Property	Description
Name	Name of the Cosmos DB connection.
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.
Description	Description of the connection. The description cannot exceed 765 characters.

Property	Description
Location	The project or folder in the Model repository where you want to store the Cosmos DB connection.
Type	Select Microsoft Azure Cosmos DB SQL API.
Cosmos DB URI	The URI of Microsoft Azure Cosmos DB account.
Key	The primary and secondary key to which provides you complete administrative access to the resources within Microsoft Azure Cosmos DB account.
Database	Name of the database that contains the collections from which you want to read or write JSON documents.

Note: You can find the Cosmos DB URI and Key values in the **Keys** settings on Azure portal. Contact your Azure administrator for more details.

Creating a PowerExchange for Cosmos DB SQL API Connection

Create a connection to import a Cosmos DB collection into the Developer tool.

1. In the Developer tool, click **Window > Preferences**.
2. Select **Informatica > Connections**.
3. Expand the domain in the **Available Connections** list.
4. Select the connection type as **NoSQL > Microsoft Azure Cosmos DB SQL API**, and then click **Add**.
5. Enter a connection name.
6. Optionally, enter a connection ID, description, location, and the type of the connection.
7. Click **Next**.
8. Specify the **Cosmos DB URI**, **Key**, and **Database** you want to connect to.
9. Click **Test Connection** to verify if the connection to the Cosmos DB collection is successful.
10. Click **Finish**.

CHAPTER 4

PowerExchange for Microsoft Azure Cosmos DB SQL API Data Objects

This chapter includes the following topics:

- [PowerExchange for Cosmos DB SQL API Data Objects Overview, 14](#)
- [PowerExchange for Cosmos DB SQL API Data Object Properties, 15](#)
- [PowerExchange for Cosmos DB SQL API Data Object Read Operation Properties, 15](#)
- [Microsoft Azure Cosmos DB SQL API Data Object Write Operation Properties, 17](#)
- [Parameterization of Cosmos DB Data Objects, 18](#)
- [Creating a Cosmos DB Data Object, 18](#)
- [Creating a Cosmos DB Data Object Operation, 19](#)

PowerExchange for Cosmos DB SQL API Data Objects Overview

Use a Cosmos DB data object to import a Cosmos DB collection. Then, add the Cosmos DB collection to a Cosmos DB data object operation, and add the operation to a mapping to read or write data.

When you create a Cosmos DB connection, select the connection type as **NoSQL > Microsoft Azure Cosmos DB SQL API** and define the connection properties. Then, create a Cosmos DB data object for the Cosmos DB collection.

PowerExchange for Cosmos DB SQL API Data Object Properties

You can configure the Cosmos DB data object properties when you create the data object.

General Properties

The following table describes the general properties that you configure for Cosmos DB data objects:

Property	Description
Name	Name of the Cosmos DB data object.
Location	The project or folder in the Model repository where you want to store the Cosmos DB data object.
Connection	Name of the Cosmos DB connection that you created to connect to the Cosmos DB collection.

Object Properties

The following table describes the general properties that you configure for the Cosmos DB object:

Property	Description
Name	Name of the collection in the Model repository.
Description	Description of the collection in the Model repository. By default, the link to the collection in the Cosmos DB database appears.
Native Name	Name of the collection in the Cosmos DB database.
Path Information	Relative path of the collection in the Cosmos DB database.
Partition Key	The field name to identify the partition to perform the read operation.

PowerExchange for Cosmos DB SQL API Data Object Read Operation Properties

Cosmos DB data object read operation properties include run-time properties that apply to the Cosmos DB collection you add in the Cosmos DB data object.

The Developer tool displays advanced properties for the Cosmos DB data object operation in the **Advanced** view.

The following table describes the advanced property for a Cosmos DB data object read operation:

Property	Description
Throughput (RU/s)	<p>A positive integer and a multiple of 100. Request units processing per second. If you specify -1, the throughput is not altered during the read operation.</p> <p>400 is the minimum throughput from the third party for a non-partition collection. 1000 is the minimum throughput from the third party for a partition collection.</p> <p>Note: If a Cosmos DB database has throughput shared among its collections, you cannot use the collections with Cosmos DB connector. You must use non-shared throughput collections with Cosmos DB connector.</p>
Partition Key	<p>The field name to identify the partition to perform the read operation. You can specify any of the following values:</p> <ul style="list-style-type: none"> - <All>. Reads data from all partitions. - Field name. Reads data from the field name partition. For example, you can read data from partition named on the City field, Boston. You can specify comma separated multiple field names. - null. Reads data from the partition named null. - <null>. Reads data from the null partition. Applicable for string data type. You should specify a string value with double quotes in a filter condition.
Page Size	Number of documents to read per request. Default is 50.
Filter Query Override	<p>A case-sensitive filter query with conditional and logical operators. Use the following syntax:</p> <p><objectName>.<columnName>="conditionValue"</p> <p>Example, Address.City="Boston"</p> <p>For more information about logical operators syntax, see <i>Microsoft Azure Cosmos DB SQL API Documentation</i>.</p>

Column Projection Properties

The Developer tool displays the column projection properties for intelligent structure model, Avro, JSON, and Parquet Cosmos DB sources in the Properties view of the **Read** operation.

The following table describes the column projection properties that you configure for the Cosmos DB sources:

Property	Description
Enable Column Projection	Displays the column details of the Cosmos DB sources.
Schema Format	Displays the schema format that you selected while creating the Cosmos DB object. You can change the schema format and provide respective schema.
Schema	<p>Displays the schema associated with the Cosmos DB object. You can select a different schema.</p> <p>Note: If you disable the column projection, the schema associated with the Cosmos DB object is removed. If you want to associate schema again with the Cosmos DB object, enable the column projection and click Edit Schema.</p>
Column Mapping	<p>Displays the mapping between input and output ports.</p> <p>Note: If you disable the column projection, the mapping between input and output ports is removed. If you want to map the input and output ports, enable the column projection and click Edit Schema to associate a schema to the Cosmos DB object.</p>

Microsoft Azure Cosmos DB SQL API Data Object Write Operation Properties

Cosmos DB data object write operation properties include run-time properties that apply to the Cosmos DB collection you add in the Cosmos DB data object.

The Developer tool displays advanced properties for the Cosmos DB data object operation in the **Advanced** view. The following table describes the advanced properties for a Cosmos DB data object write operation:

Property	Description
Throughput (RU/s)	<p>A positive integer and a multiple of 100. Request units processing per second. If you specify -1, the throughput is not altered during the write operation.</p> <p>400 is the minimum throughput from the third party for a non-partition collection. 1000 is the minimum throughput from the third party for a partition collection.</p> <p>Note: If a Cosmos DB database has throughput shared among its collections, you cannot use the collections with Cosmos DB connector. You must use non-shared throughput collections with Cosmos DB connector.</p>
Automatic ID Generation	<p>Generate ID for the documents written to the target. Specify any of the following values:</p> <ul style="list-style-type: none">- Enabled. Cosmos DB generates IDs for the documents.- Disabled. The source object provides IDs for the documents. The Data Integration Service rejects the rows if a Null value is provided to the ID port in the target or if you do not connect a port to the ID port in target.
Treat Source Rows As	<p>Operation to perform. You can select Insert, Update, Upsert, or Delete.</p> <p>Note: You must connect the ID port and Partition Key port to perform Update, Upsert, or Delete operations. The Data Integration Service displays an exception if you do not connect the ID port or the partition key for Upsert and Update operations, whereas, the Delete operation fails. For Upsert and Update operations, in addition to the ports you want to update, you must connect all other ports a document contains. The values for the unconnected ports get deleted during Upsert and Update operations.</p> <p>If you have a duplicate Partition Key and ID combination in the source rows, the mapping fails at the first occurrence of a bad record. The azure-cosmosdb-spark library throws an error and stops the mapping execution immediately.</p>

Column Projection Properties

The Developer tool displays the column projection properties for intelligent structure model, Avro, JSON, and Parquet Cosmos DB targets in the Properties view of the **Write** operation.

The following table describes the column projection properties that you configure for the Cosmos DB targets:

Property	Description
Enable Column Projection	Displays the column details of the Cosmos DB targets.
Schema Format	Displays the schema format that you selected while creating the Cosmos DB object. You can change the schema format and provide respective schema.

Property	Description
Schema	Displays the schema associated with the Cosmos DB object. You can select a different schema. Note: If you disable the column projection, the schema associated with the Cosmos DB object is removed. If you want to associate schema again with the Cosmos DB object, enable the column projection and click Edit Schema .
Column Mapping	Displays the mapping between input and output ports. Note: If you disable the column projection, the mapping between input and output ports is removed. If you want to map the input and output ports, enable the column projection and click Edit Schema to associate a schema to the Cosmos DB object.

Parameterization of Cosmos DB Data Objects

You can parameterize the Cosmos DB connection and the Cosmos DB data object operation properties.

You can parameterize the following data object properties for Cosmos DB data objects:

- Connection
- Native Name to override the collection name.

You can parameterize the following data object read operation advanced properties for Cosmos DB data objects:

- Partition Key
- Page Size
- Throughput (RU/s)
- Filter Query Override

You can parameterize the following data object write operation advanced properties for Cosmos DB data objects:

- Throughput (RU/s)
- Automatic ID Generation
- Treat Source Rows As

For more information about mapping parameters, see the *Informatica Developer Mapping Guide*.

Creating a Cosmos DB Data Object

Create a Cosmos DB object to specify the Cosmos DB collection that you want to access to read or write data.

1. Select a project or folder in the **Object Explorer** view.
2. Click **File > New > Data Object**.
3. Select **Microsoft Azure Cosmos DB SQL API Data Object** and click **Next**.

The **New Microsoft Azure Cosmos DB SQL API 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. Click **Browse** next to the **Connection** option and select a Cosmos DB connection from which you want to import the Cosmos DB collection.
7. To add a Cosmos DB collection to the data object, click **Add** next to the **Resource** option.
The **Add Resource** dialog box appears.
8. Select the required connection under Package Explorer.
The list of collections appears.
9. Click **Go** and then click **Advanced**. The **Schema Document ID** field appears.
10. Specify the document ID in the **Schema Document ID** field. The schema is fetched from the Cosmos DB collection based on the document ID you provide here.
11. Select the collection that contains the document ID you specified in **Schema Document ID**.
12. Click the selected collection row.
The document details appear under **Entity Information**.
13. Click **OK**.
14. Click **Finish**.

The data object appears under Data Objects in the project or folder in the **Object Explorer** view. The Data Object Read and Write operations are created by default.

Note: The schema or metadata for the read or write operations is derived based on the ID provided in the **Schema Document ID** field using the best match logic. For the provided ID, if a column 'C1' contains a value '123', the Data Integration Service interprets the value as Integer. If the derived data types do not match your requirements, you can modify the data types in read or write operations using the **Edit Schema** option in the column projection properties.

You can use the **Name** field to search for a name in the collection. The text that you provide in the **Name** field must be a part of the collection name and is case-sensitive.

Creating a Cosmos DB Data Object Operation

Create a Cosmos DB data object operation from a Cosmos DB data object that contains a Cosmos DB collection.

Before you create a Cosmos DB data object operation, you must create a Cosmos DB data object with the Cosmos DB collection.

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 operation. You can choose to create a read operation or a write operation.
5. Click **Add**.
The **Select a resource** dialog box appears.
6. Select the Cosmos DB collection for which you want to create the data object operation and click **OK**.

7. Click **Finish**.

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

CHAPTER 5

PowerExchange for Microsoft Azure Cosmos DB SQL API Mappings

This chapter includes the following topics:

- [Microsoft Azure Cosmos DB Mappings Overview, 21](#)
- [Mapping Validation and Run-time Environments, 21](#)

Microsoft Azure Cosmos DB Mappings Overview

After you create a Cosmos DB data object read or write operation, you can create a mapping.

You can create an Informatica mapping containing a Cosmos DB data object read operation as the input, and a relational or flat file data object operation as the target. You can create an Informatica mapping containing objects such as a relational or flat file data object operation as the input, transformations, and a Cosmos DB data object write operation as the output to load data to Cosmos DB.

Validate and run the mapping. You can deploy the mapping and run it or add the mapping to a Mapping task in a workflow. You cannot define a Lookup transformation for the Microsoft Azure Cosmos DB data object.

Mapping Validation and Run-time Environments

You can validate and run mappings in the native environment or in a non-native environment, such as Hadoop or Databricks.

When you validate a mapping, you can validate it against one or all of the engines. The Developer tool returns validation messages for each engine.

When you run a mapping, you can choose to run the mapping in the native environment or in a non-native environment, such as Hadoop or Databricks. When you run mappings in the native environment, the Data Integration Service processes and runs the mapping. When you run mappings in a non-native environment, the Data Integration Service pushes the processing to a compute cluster, such as Hadoop or Databricks.

You can run standalone mappings, mappings that are a part of a workflow in a non-native environment. When you select the Hadoop environment, the Data Integration Service pushes the mapping logic to the Spark engine.

When you select the Databricks environment, the Integration Service pushes the mapping logic to the Databricks Spark engine, the Apache Spark engine packaged for Databricks.

APPENDIX A

Cosmos DB Data Types Reference

This appendix includes the following topics:

- [Cosmos DB Data Types Reference Overview, 23](#)
- [Azure Cosmos DB and Transformation Data Types, 24](#)
- [Data Types Parsing for Cosmos DB, 24](#)
- [Rules and Guidelines for Data Types, 25](#)

Cosmos DB Data Types Reference Overview

When you run the session to read data from or write data to Cosmos DB, the Data Integration Service converts the transformation data types to comparable native Cosmos DB data types.

Informatica Developer uses the following data types in Cosmos DB mappings:

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

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

Azure Cosmos DB and Transformation Data Types

The following table compares the JSON data type to the transformation data type:

JSON Data Type	Transformation Data Type	Range and Description
boolean	integer	The default transformation type for boolean is integer. You can specify string data type with values of True and False. True is equivalent to the integer 1 and False is equivalent to the integer 0.
Number (double)	double	-1.79769313486231570E+308 to +1.79769313486231570E+308. Precision 15.
Number (float)	double	-1.79769313486231570E+308 to +1.79769313486231570E+308. Precision 15.
Number (int)	integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Number (long)	bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0.
string	string	1 to 104,857,600 characters.

See [“Data Types Parsing for Cosmos DB” on page 24](#) for more details.

Data Types Parsing for Cosmos DB

During the read or write operations, the Data Integration Service parses data based on the data types defined in the schema. If the data values do not match the data types defined in the schema, the Data Integration Service rejects the document.

The following table lists the data types allowed at run time for the numeric data types specified in the schema:

Data Type in Schema	Allowed Run-time Data
Integer	Short, Integer
BigInt or Long	Short, Integer, Long (Maximum precision 19)
Float	Short, Integer, Long, Float
Double	Short, Integer, Long, Float, Double
Decimal (Maximum precision 28)	Short, Integer, Long, Float, Double, Long
String	String

Note: PowerExchange for Cosmos DB SQL API does not support JSON nested document type field or array type field.

Rules and Guidelines for Data Types

Consider the following rules and guidelines when you use the data types for Microsoft Azure Cosmos DB:

- When a field in the schema is of Integer data type and you write a BigInt data to that field, the azure-cosmosdb-spark library converts the BigInt data to an integer and writes it to the Integer field.
- When a field in the schema is of BigInt data type and you write a String value to it, the mapping fails.
- When a field in the schema is of Double data type and you try to read a BigInt value from it, junk values are read.
- When the Decimal data type has precision greater than 18, the mapping fails in Spark mode.
- When one of the row values exceeds the range of BigInt data type, the mapping fails in Spark mode and all the rows are rejected.
- When one of the row values exceeds the range of Integer data type, the mapping reads an invalid value in Spark mode.
For example, create a document with Int data type and use the value as 9999999999 which exceeds the range of Int data type. When you run a mapping, the target contains an invalid value 1410065407.
- You cannot edit the schema for a JSON file at the object level. To edit the schema for a JSON file, click **Column Projection** for the read or write operation and then click **Edit Schema**.

INDEX

C

- Cosmos DB connection
 - creating [12](#)
- Cosmos DB connections
 - creating [13](#)
 - overview [12](#)
- Cosmos DB data object
 - creating [18](#)
- Cosmos DB data object operation
 - creating [19](#)
 - read properties for Cosmos DB [15](#)
 - write properties for Cosmos DB [17](#)
- Cosmos DB data objects
 - general properties [15](#)
 - object properties [15](#)
 - overview [14](#)
- creating
 - Cosmos DB connection [12](#)
 - Cosmos DB connections [13](#)
 - Cosmos DB data object [18](#)
 - Cosmos DB data object operation [19](#)

D

- data types
 - parsing [24](#)

I

- Introduction
 - Cosmos DB [9](#)

N

- native data type [24](#)

O

- overview
 - Cosmos DB [12](#)

- overview (*continued*)
 - Cosmos DB data objects [14](#)

P

- PowerExchange for Cosmos DB data types
 - overview [23](#)
- PowerExchange for Cosmos DB mappings
 - overview [21](#)
- PowerExchange for Cosmos DB SQL API
 - configuration overview [10](#)
 - overview [8](#)

R

- read operation properties
 - column projection properties [16](#)
- run-time properties for Cosmos DB
 - Cosmos DB data object read operation [15](#)
 - Cosmos DB data object write operation [17](#)

S

- Spark engine
 - mappings [21](#)

T

- transformation data type [24](#)

W

- write operation properties
 - column projection properties [17](#)