



Informatica® Cloud Application Integration  
July 2024

# Simple RAG Consumption with Pinecone

© Copyright Informatica LLC 2024

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, Informatica Master Data Management, and Live Data Map 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 © 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 © Teleric 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](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.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://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://www.php.net/license/3_01.txt); <http://srp.stanford.edu/license.txt>; <http://www.schneier.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.

Publication Date: 2024-07-31

# Table of Contents

- Preface ..... 5**
  
- Chapter 1: Introduction to Simple RAG Consumption with Pinecone recipe.... 6**
  - Prerequisites for creating an index in Pinecone. . . . . 6
  
- Chapter 2: Recipe contents..... 9**
  - Simple RAG Consumption with Pinecone recipe assets. . . . . 9
  
- Chapter 3: Using the Simple RAG Consumption with Pinecone recipe..... 11**
  - Copying and accessing the recipe. . . . . 11
  - Configuring and publishing the GeminiRAGConsumption connection. . . . . 12
  - Configuring and publishing the PineconeRAGConsumption connection. . . . . 12
  - Configuring and publishing the processes. . . . . 13
  - Invoking the process. . . . . 13

# Preface

Use *Simple RAG Consumption with Pinecone* to learn how to receive a query from the user, convert it into vectors, form the context, and return a comprehensive response using a Large Language Model (LLM). The recipe is based on REST and SOAP APIs and you use an HTTP request to call the process.

## CHAPTER 1

# Introduction to Simple RAG Consumption with Pinecone recipe

The Simple Retrieval Augmented Generation (RAG) Consumption with Pinecone recipe is based on REST and SOAP APIs.

The process submits a query that is received from the user, converts it into a vector, and uses it to search for similar vectors in a database. The top K matches are retrieved, filtered by a cutoff score, and used to form a context. This context including the original query is passed to a Large Language Model (LLM) to generate and return a comprehensive response.

## Prerequisites for creating an index in Pinecone

To ensure that the process works correctly, you must create an index in Pinecone and add text for the context. The context is then converted into vectors using the Gemini Embeddings process.

1. Open the **Create a new index** page in Pinecone.
2. In the **Default** field, enter an index name and in the **Dimensions** field, enter the value **768** as shown in the following image:

### Create a new index

Default /

#### Configuration

The dimensions and metric depend on the model you select.

Dimensions  Metric  [Setup by model](#)

#### Capacity mode

SERVERLESS PODS

##### Serverless

Charges based on data storage, reads, and writes.

You are currently on the **Starter Plan**



Cancel [Create index](#)

3. Click **Create index**.

After creating the index, you can use the **HOST** value without `https://` in the **Index\_Host** input parameter as shown in the following image:

[← Back to indexes](#)

### somenameindex ●

METRIC	DIMENSIONS	HOST
cosine	768	<a href="https://somenameindex-38s14f2.svc.aped-4627-b74a.pinecone.io">https://somenameindex-38s14f2.svc.aped-4627-b74a.pinecone.io</a>
CLOUD	REGION	TYPE
 AWS	us-east-1 	Serverless

[BROWSER](#) METRICS NAMESPACES (0)

#### Index records

### No Records Yet

A record is an object you add to an index containing a vector and, optionally, its metadata

[+ Add a Record](#)

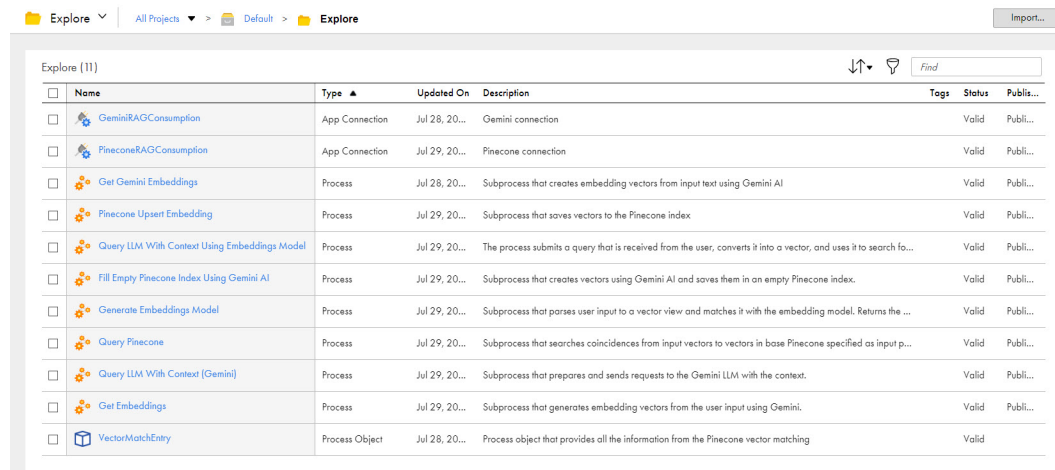


# CHAPTER 2

## Recipe contents

The Simple RAG Consumption with Pinecone recipe contains app connections, processes, and a process object.

The following image shows the assets that the Simple RAG Consumption with Pinecone recipe package contains:



## Simple RAG Consumption with Pinecone recipe assets

The following table lists the assets that the Simple RAG Consumption with Pinecone recipe package contains:

Asset Name	Asset Type	Description
PineconeRAGConsumption	App connection	Pinecone connection.
GeminiRAGConsumption	App connection	Gemini connection.
VectorMatchEntry	Process object	Provides all the information from the Pinecone vector matching.

Asset Name	Asset Type	Description
Get Gemini Embeddings	Process	Subprocess that creates embedding vectors from input text using Gemini AI.
Pinecone Upsert Embedding	Process	Subprocess that saves vectors to the Pinecone index.
Fill Empty Pinecone Index Using Gemini AI	Process	Subprocess that creates vectors using Gemini AI and saves them in an empty Pinecone index.
Generate Embeddings Model	Process	Subprocess that parses user input into a vector view and matches it with the embedding model. Returns the matching context score and metadata.
Query Pinecone	Process	Subprocess that searches coincidences from input vectors to vectors in the base Pinecone index specified as input parameter Index_Host_Pinecone. Returns a result with metadata.
Query LLM With Context (Gemini)	Process	Subprocess that prepares and sends requests to the Gemini LLM with the context.
Get Embeddings	Process	Subprocess that generates embedding vectors from the user input using Gemini.
Query LLM with Context using Embeddings Model	Process	Submits a query that is received from the user, converts it into a vector, and uses it to search for similar vectors in a database. The top K matches are retrieved, filtered by a cutoff score, and used to form a context. This context including the original query is passed to a Large Language Model (LLM) to generate and return a comprehensive response.

## CHAPTER 3

# Using the Simple RAG Consumption with Pinecone recipe

To use the Simple RAG Consumption with Pinecone recipe, you must perform the following steps manually:

Step 1: Copy and access the recipe

Step 2: Configure and publish the GeminiRAGConsumption connection

Step 3: Configure and publish the PineconeRAGConsumption connection

Step 4: Configure and publish the processes

Step 5: Invoke the process

## Copying and accessing the recipe

To copy and access the recipe content, perform the following steps:

1. Open the **Simple RAG Consumption with Pinecone** recipe and click **Use**.
2. Select the location where you want to copy the recipe, and then click **Continue**.
3. In the **Copying the recipe** dialog box, click **OK**.

It might take some time for the recipe to get copied. You will receive a notification when the recipe is ready for use.

4. After the recipe is copied, click **Explore** to access the recipe content.

- Navigate to the project or folder where you copied the recipe or enter the recipe name in the **Find** box. All the assets in the recipe are displayed as shown in the following image:

The screenshot shows the 'Explore' interface with a table of assets. The table has columns for Name, Type, Updated On, Description, Tags, Status, and Publish. The assets listed include connections and various processes related to Gemini AI and Pinecone.

Name	Type	Updated On	Description	Tags	Status	Publish
GeminiRAGConsumption	App Connection	Jul 28, 20...	Gemini connection		Valid	Publi...
PineconeRAGConsumption	App Connection	Jul 29, 20...	Pinecone connection		Valid	Publi...
Get Gemini Embeddings	Process	Jul 28, 20...	Subprocess that creates embedding vectors from input text using Gemini AI		Valid	Publi...
Pinecone Upsert Embedding	Process	Jul 29, 20...	Subprocess that saves vectors to the Pinecone index		Valid	Publi...
Query LLM With Context Using Embeddings Model	Process	Jul 29, 20...	The process submits a query that is received from the user, converts it into a vector, and uses it to search fo...		Valid	Publi...
Fill Empty Pinecone Index Using Gemini AI	Process	Jul 29, 20...	Subprocess that creates vectors using Gemini AI and saves them in an empty Pinecone index.		Valid	Publi...
Generate Embeddings Model	Process	Jul 29, 20...	Subprocess that parses user input to a vector view and matches it with the embedding model. Returns the ...		Valid	Publi...
Query Pinecone	Process	Jul 29, 20...	Subprocess that searches coincidences from input vectors to vectors in base Pinecone specified as input p...		Valid	Publi...
Query LLM With Context (Gemini)	Process	Jul 29, 20...	Subprocess that prepares and sends requests to the Gemini LLM with the context.		Valid	Publi...
Get Embeddings	Process	Jul 29, 20...	Subprocess that generates embedding vectors from the user input using Gemini.		Valid	Publi...
VectorMatchEntry	Process Object	Jul 28, 20...	Process object that provides all the information from the Pinecone vector matching		Valid	

## Configuring and publishing the GeminiRAGConsumption connection

To configure and publish the GeminiRAGConsumption connection, perform the following steps:

- Open the **GeminiRAGConsumption** connection.
- In the **Type** field, select **Gemini**.
- In the **Run On** field, select **Cloud Server or any Secure Agent**.
- In the **Connection Properties** section, enter the API key in the **API\_Key** property. The **API\_Key** property authenticates Gemini connection requests.
- Save, test, and publish the connection.

## Configuring and publishing the PineconeRAGConsumption connection

To configure and publish the PineconeRAGConsumption connection, perform the following steps:

- Open the **PineconeRAGConsumption** connection.
- In the **Type** field, select **Pinecone**.
- In the **Run On** field, select **Cloud Server or any Secure Agent**.
- In the **Connection Properties** section, enter the API key in the **API\_Key** property. The **API\_Key** property authenticates Pinecone connection requests.
- Save, test, and publish the connection.

# Configuring and publishing the processes

The Get Gemini Embeddings process, Pinecone Upsert Embedding process, and Fill Empty Pinecone Index Using Gemini AI process consume the user's input text, modify it to a vector representation, and save it to the Pinecone index.

To configure and publish the processes, perform the following steps:

1. Open the following processes in the order specified below:
  1. Get Gemini Embeddings
  2. Pinecone Upsert Embedding
  3. Fill Empty Pinecone Index Using Gemini AI
  4. Get Embeddings
  5. Query Pinecone
  6. Generate Embeddings Model
  7. Query LLM With Context (Gemini)
  8. Query LLM with Context using Embeddings Model
2. For each process, on the **Start** tab of the Start step, select **Cloud Server** from the **Run On** list.
3. Optionally, for the **Query LLM with Context using Embeddings Model** process, in the **Set LLM Models** step, in the **Assignments** tab, you can update the values in the **Set\_Context\_Model** and **Set\_Embedding\_Model** fields.
4. Save and publish all the processes.

## Invoking the process

When you invoke the Query LLM With Context Using Embeddings Model process, the user sees the answer matching the context that was used in the LLM request.

You can run the process using one of the following options:

- REST or SOAP API endpoints in any API client such as cURL, Postman, SOAP UI, or any programming language
- Web browser by passing the input parameters