
Article

[sween](#) · Jul 5, 2022 4m read

IRIS Data to Google Big Query - InterSystems Cloud SQL via Dataflow



How to include IRIS Data into your Google Big Query Data Warehouse and in your Data Studio data explorations. In this article we will be using Google Cloud Dataflow to connect to our InterSystems Cloud SQL Service and build a job to persist the results of an IRIS query in Big Query on an interval.

If you were lucky enough to get access to Cloud SQL at Global Summit 2022 as mentioned in "[InterSystems IRIS: What's New, What's Next](#)", it makes the example a snap, but you can pull this off with any publicly or vpc accessible listener you have provisioned instead.

Prerequisites

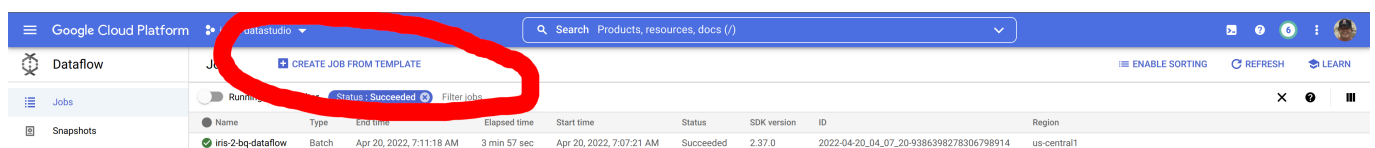
Provision InterSystems Cloud SQL for temporary use

Setup Google Cloud Platform

Google Dataflow Job

If you followed the steps above you should have the following in your inventory to execute the job to read your InterSystems IRIS data and ingest it into Google Big Query using Google Dataflow.

In the Google Cloud Console, head over to Dataflow and select "Create Job from Template"



This is a rather unnecessary/exhaustive illustration on how to instruct you to fill out a form with the generated pre-requisites, but it calls out the source of the components...

Google Cloud Platform iris-2-datastudio Search Products, resources, docs (/)

Dataflow Create job from template

Job name *
iris-2-bq-dataflow
Must be unique among running jobs

Regional endpoint *
us-central1 (Iowa)
Choose a Dataflow regional endpoint to deploy worker instances and store job metadata. By using the worker region endpoint, you can select any available Google Cloud region or zone as the Dataflow regional endpoint. [Learn more](#)

Dataflow template *
Jdbc to BigQuery
Template that reads from a Jdbc source and writes to a BigQuery table. Jdbc connection string must be provided. The connection string is created using the Google Cloud KMS API. [OPEN TUTORIAL](#)

Read from JdbcIO

Write to BigQuery

Required parameters

Jdbc connection URL string. Connection string can be passed in as plaintext or as ...
jdbc:IRIS://k8s-c5ce7068-a4244044-265532e16d-2be47d3d6962f6cc.elb.us-east-1.amazonaws.com:1521/sampledb

Jdbc driver class name *
com.intersystems.jdbc.IRISDriver
Jdbc driver class name. E.g. com.mysql.jdbc.Driver

Jdbc source SQL query. *
SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, TABLE_TYPE, SELF_REFERENCING
Query to be executed on the source to extract the data. E.g. select * from sampledb.sample_table

BigQuery output table *

... to round it out, make sure you expand the bottom section and supply your credentials for IRIS.

Google Cloud Platform iris-2-datastudio Search Products, resources, docs (/)

Dataflow Create job from template

BigQuery output table *
iris-2-datastudio:irisdata.dataflowtable
BigQuery table location to write the output to. The table's schema must match the source query schema. Ex: your-project:your-dataset:your-table-name

GCS paths for Jdbc drivers *
gs://iris-2-datastudio/intersystems-jdbc-3.3.0.jar
Comma separate GCS paths for Jdbc drivers. E.g. gs://your-bucket/driver_jar1.jar,gs://your-bucket/driver_jar2.jar

Temporary directory for BigQuery loading process *
gs://iris-2-datastudio/input
Example: gs://your-bucket/your-files/temp_dir

Temporary location *
gs://iris-2-datastudio/tmp
Path and filename prefix for writing temporary files. Ex: gs://your-bucket/temp

Encryption

☒ Google-managed encryption key
No configuration required

☐ Customer-managed encryption key (CMEK)
Manage via Google Cloud Key Management Service

Optional parameters

Jdbc connection property string
Properties string to use for the Jdbc connection. E.g. unicode=true&characterEncoding=UTF-8

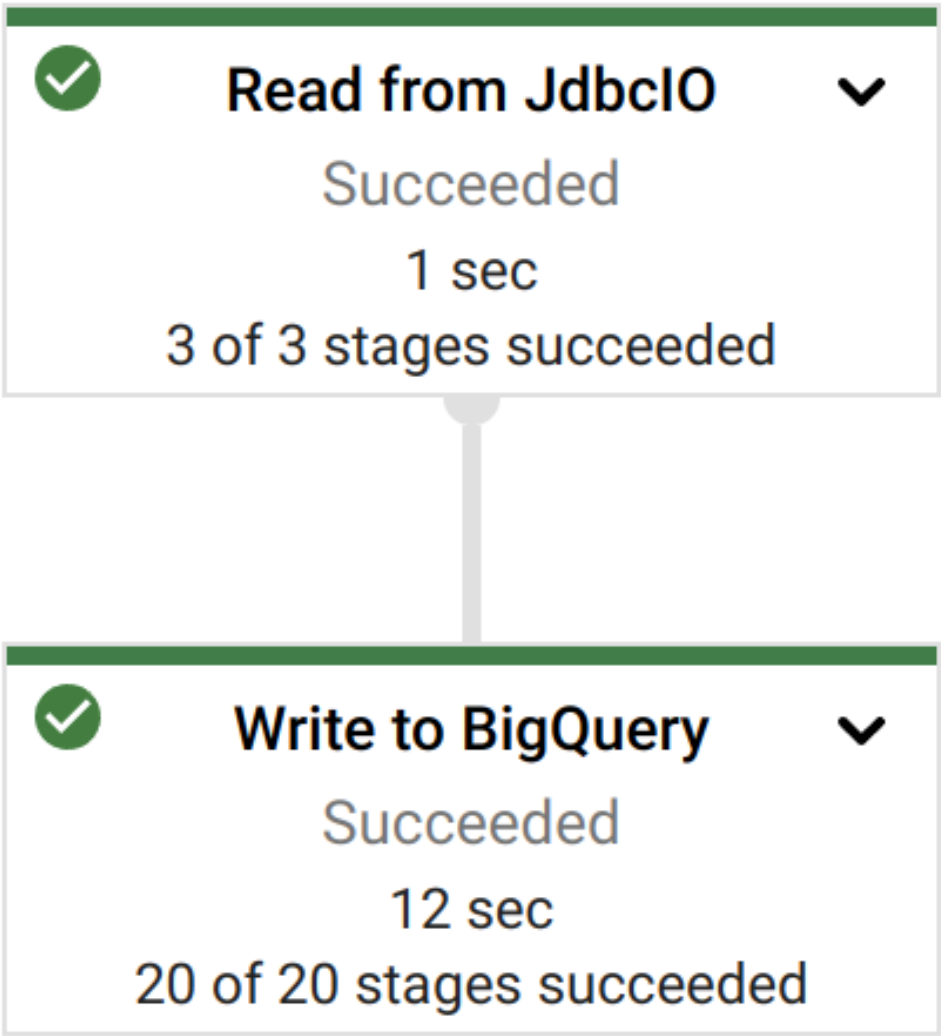
Jdbc connection username
SQLAdmin
User name to be used for the Jdbc connection. User name can be passed in as plaintext or as a base64 encoded string encrypted by Google Cloud KMS

Jdbc connection password
Testing12!

For the ones who found those screenshots offensive to your intelligence, here is the alternate route to go to keep you inside your comfort zone in the CLI to run the job:

```
gcloud dataflow jobs run iris-2-bq-dataflow \
--gcs-location gs://dataflow-templates-us-centrall/latest/Jdbc_to_BigQuery \
--region us-centrall --num-workers 2 \
--staging-location gs://iris-2-datastudio/tmp \
--parameters connectionURL=jdbc:IRIS://k8s-
c5ce7068-a42
44044-265532e16d-2be47d3
d6962f6cc.elb.us-east-1.amazonaws.com
:1972/USER,driverClassName=
com.intersystems.jdbc.IRISDriver
,query=SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, TABLE_TYPE, SELF_REFERENCING_C
OLUMN_NAME, REFERENCE_GENERATION, USER_DEFINED_TYPE_CATALOG, USER_DEFINED_TYPE_SCHEMA
, USER_DEFINED_TYPE_NAME, IS_INSERTABLE_INTO, IS_TYPED, CLASSNAME, DESCRIPTION, OWNER
, IS_SHARDED FROM INFORMATION_SCHEMA.TABLES;,outputTable=iris-2-datastudio:irisdata.d
ataflowtable,driverJars=gs://
iris-2-datastudio/intersystems-
jdbc-3.3.0.jar
,bigQueryLoadingTemporaryDirectory=gs://iris-2-datastudio/input,username=SQLAdmin,pas
sword=Testing12!
```

Once you have kicked off your job, you can bask in the glory a successful job run:



Results

Taking a look at our source data and query in InterSystems Cloud SQL...

CLOUD SERVICES PORTAL

Cloud Services

Deployments

List

Overview

Import Files

Manage Files

SQL Query Tools

Tenants

Settings

Resource Center

Reports

Administration

Deployments > Google Dataflow Big Query > SQL Query Tools

Schema Tree

SQL Editor

SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, TABLE_TYPE, SELF_REFERENCING_COLUMN_NAME, REFERENCE_GENERATION, USER_DEFINED_TYPE_CATALOG, USER_DEFINED_TYPE_SCHEMA, USER_DEFINED_TYPE_NAME, IS_INSERTABLE_INTO, IS_TYPED, CLASSNAME, DESCRIPTION, OWNER, IS_SHARDED FROM INFORMATION_SCHEMA.TABLES

Response timeout (ms)

25000

Clear

< Previous

Execute

Next >

History

Results

TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE	SELF_REFERENCING_COLUMN_NAME	REFERENCE_GENERATION	USER_DEFINED_TYPE_CATALOG	USER_D
null	%SQL_Diag	Message	SYSTEM TABLE	null	null	null	null
null	%SQL_Diag	Result	SYSTEM TABLE	null	null	null	null
null	INFORMATION_SCHEMA	CHECK_CONSTRAINTS	SYSTEM TABLE	null	null	null	null
null	INFORMATION_SCHEMA	COLUMNS	SYSTEM TABLE	null	null	null	null
null	INFORMATION_SCHEMA	CONSTRAINT_COLUMN_USAGE	SYSTEM TABLE	null	null	null	null
null	INFORMATION_SCHEMA	CONSTRAINT_TABLE_USAGE	SYSTEM TABLE	null	null	null	null
null	INFORMATION_SCHEMA	INDEXES	SYSTEM TABLE	null	null	null	null

... and then Inspecting the results in Big Query, it appears we do in fact, have InterSystems IRIS Data in Big Query.

The screenshot shows the Google Cloud Platform Data Studio interface. A SQL query is entered in the editor: `SELECT * FROM `iris-2-datastudio.irisdata.dataflowtable` LIMIT 1000`. The query is executed, and the results are displayed in a table. A red arrow points to the 'Query results' tab, which shows a table with columns: JOB INFORMATION, RESULTS, JSON, and EXECUTION. The table contains 6 rows of data, including system tables like %SQL_Diag, INFORMATION_SCHEMA.CHECK_CONSTRAINTS, INFORMATION_SCHEMA.COLUMNS, INFORMATION_SCHEMA.CONSTRAINT_COLUMN_USAGE, and INFORMATION_SCHEMA.CONSTRAINT_TABLE_USAGE.

Row	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE	SELF_REFERENCING_COLUMN_NAME	REFERENCE_GENERATION	USER_DEFINED_TYPE_CATALOG	USER_DEFINED_TYPE_SCHEMA	USER_DEFINED_TYPE
1	null	%SQL_Diag	Message	SYSTEM TABLE	null	null	null	null	null
2	null	%SQL_Diag	Result	SYSTEM TABLE	null	null	null	null	null
3	null	INFORMATION_SCHEMA	CHECK_CONSTRAINTS	SYSTEM TABLE	null	null	null	null	null
4	null	INFORMATION_SCHEMA	COLUMNS	SYSTEM TABLE	null	null	null	null	null
5	null	INFORMATION_SCHEMA	CONSTRAINT_COLUMN_USAGE	SYSTEM TABLE	null	null	null	null	null
6	null	INFORMATION_SCHEMA	CONSTRAINT_TABLE_USAGE	SYSTEM TABLE	null	null	null	null	null

Once we have the data in Big Query, it is trivial to include our IRIS data into Data Studio by selecting Big Query as the data source... this example below is missing some flair, but you can quickly see the IRIS data ready for manipulation in your Data Studio project.

The screenshot shows a Google Data Studio report titled 'Untitled Report'. The report displays a table of data with columns: TABLE_CATALOG, CLASSNAME, OWNER, IS_SHARDED, and Record Count. The table contains 10 rows of data. A scorecard is also visible, showing a 'Record Count' of 64. The right sidebar shows the 'Data' panel with a list of fields, including DESCRIPTION, IS_INSERTABLE_IN..., IS_SHARDED, IS_TYPED, OWNER, REFERENCE_GENE..., SELF_REFERENCIN..., TABLE_CATALOG, TABLE_NAME, TABLE_SCHEMA, TABLE_TYPE, USER_DEFINED_TY..., USER_DEFINED_TY..., USER_DEFINED_TY..., and Record Count.

TABLE_CATALOG	CLASSNAME	OWNER	IS_SHARDED	Record Count
8.. null	INFORMATION.S...	.PUBLIC	false	2
9.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2
1.. null	INFORMATION.S...	.PUBLIC	false	2

[#Analytics](#) [#Best Practices](#) [#Cloud](#) [#integration-required](#) [#InterSystems IRIS](#)

Source URL: <https://community.intersystems.com/post/iris-data-google-big-query-intersystems-cloud-sql-dataflow>