

# Troubleshooting and Solving Data Join Pitfalls

GSP412



- BigQuery is Google's fully managed, serverless analytics database for querying massive datasets with SQL under a pay-as-you-go model, without infrastructure management.
- This lab focuses on avoiding common pitfalls in data table joins that can corrupt analytical results and insights.
- Key join types include cross join for all row combinations, inner join for matching keys only, left join preserving left table rows, and right join preserving right table rows.
- You'll work with an ecommerce dataset containing millions of Google Analytics records from the Google Merchandise Store loaded into BigQuery.
- Explore the dataset fields and rows while practicing safe join techniques to derive meaningful business insights.

G Build a Data Warehouse with BigQuery +

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:29:51 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

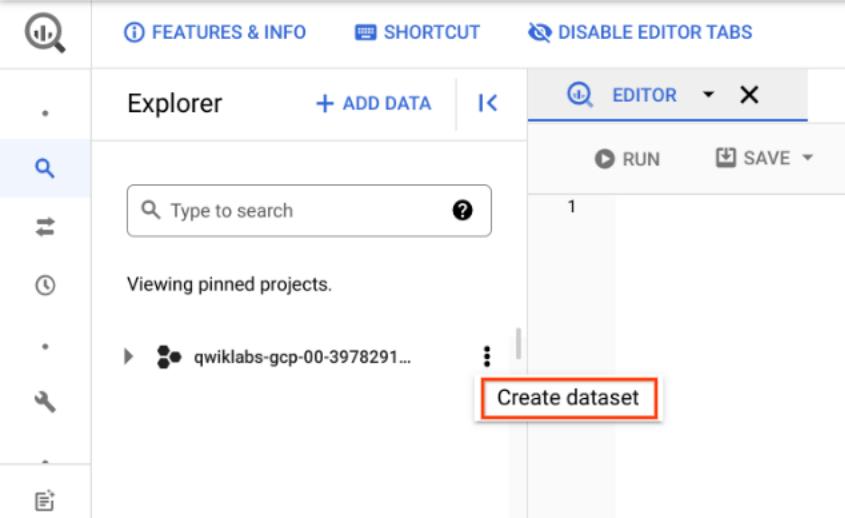
Username: student-02-efbe2bad126b... Password: qQAK3jWEKbic Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

Task 1. Create a new dataset to store your tables

In your BigQuery project, create a new dataset titled ecommerce.

1. Click the three dots next to your Project ID and select **Create dataset**.



FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

Type to search

Viewing pinned projects.

qwiklabs-gcp-00-3978291...

Create dataset

0/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:29:41 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

Explorer + ADD DATA

EDITOR RUN SAVE

Type to search

Viewing pinned projects.

qwiklabs-gcp-00-3978291...

Create dataset

PERSONAL HISTORY

0/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

The **Create dataset** dialog opens.

2. Set the *dataset ID* to **ecommerce**.

3. Leave the other options at their default values, and click **Create dataset**.

In the left pane, you see an **ecommerce** table listed under your project.

Next >

6 15°C Mostly cloudy

Search

ENG IN

01:48 PM 31-12-2025

Dashboard – qwiklabs-gcp-03-253defc4cbc4 x +

https://console.cloud.google.com/home/dashboard?project=qwiklabs-gcp-03-253defc4cbc4&pli=1

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

# Google Cloud

Overview Preview

qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more

Studio Recommendations

Customize

Favorite products appear here

Products

- Billing
- IAM & Admin
- Marketplace
- APIs & Services
- Compute Engine
- Kubernetes Engine
- Cloud Storage
- Security
- BigQuery
- Monitoring
- Cloud Run
- VPC Network

[View all products](#)

Pipelines & Integration

- Data transfers
- Dataform
- Scheduled queries
- Scheduling

Governance

- Sharing (Analytics Hub)
- Policy tags
- Metadata curation

Administration

- Jobs explorer
- Capacity management
- BI Engine
- Disaster recovery
- Recommendations

Storage

- Multi-class multi-region object storage

Recommendations

API APIs

Requests (requests/sec)

0.05/s  
0.04/s  
0.03/s  
0.02/s  
0.01/s  
0

1 PM 1:15 1:30 1:45

● Requests: 0.039/s

Go to APIs overview

Google Cloud Platform status

All services normal

→ Go to Cloud status dashboard

Billing

Estimated charges USD \$0.00  
For the billing period Dec 1 – 31, 2025

Take a tour of billing

→ View detailed charges

Monitoring

Create my dashboard  
Set up alerting policies  
Create uptime checks

15°C 6 Mostly cloudy

Search

1 ENG IN 01:49 PM 31-12-2025

BigQuery – qwiklabs-gcp-03-253 X +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m0

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud Search (/) for resources, docs, products, and more Search

Search BigQuery resources Show starred only

qwiklabs-gcp-03-253defc4... :

Check out what's new in Studio

Create dataset Refresh contents Upload to project Change my default code region

Notebook gallery

Organize and share code assets [Preview](#)

The new Files explorer provides a dedicated view to organize your BigQuery assets (saved queries, notebooks, and more) into folders.

To get started with better project management and easier collaboration, click the Files icon at the top of the left pane. [Learn more](#)

Try it Next

Create new

SQL query Notebook Notebook with Spark ML model Data canvas Pipeline Data preparation Table

Job history Show

6 15°C Mostly cloudy

Search

01:51 PM 31-12-2025 ENG IN

This screenshot shows the Google Cloud BigQuery Studio interface. A modal window titled 'Check out what's new in Studio' is open, highlighting the 'Files' feature. The modal includes a preview image of the BigQuery Assets page with a folder icon circled in blue, and text explaining how it allows users to organize and share code assets. Below the modal, there are sections for creating new datasets, notebooks, and other assets, along with a 'Job history' section. The left sidebar contains various icons for different Google Cloud services, and the bottom navigation bar includes links for search, video, file, and settings.

Create dataset - BigQuery - qwiklabs

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m0

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud Project ID: qwiklabs-gcp-03-253defc4cbc4

Search (/) for resources, docs, products, and more

Check out what's new in Studio

- Files explorer
- Gemini in queries
- Apache Spark
- Visualization Cell
- Notebook gallery

Create new

- SQL query
- Notebook
- Notebook with Spark
- ML model

Job history

6 15°C Mostly cloudy

Search

1 ENG IN 01:52 PM 31-12-2025

## Create dataset

Project ID \* **qwiklabs-gcp-03-253defc4cbc4** [Change](#)

Dataset ID \* **ecommerce**  
Letters, numbers, and underscores allowed

Some locations have been restricted due to a policy set by your organization. [Learn more about restricting locations.](#)

Data location

Tags

Advanced options

**Create dataset** Cancel

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:26:05 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

Task 2. Pin the lab project in BigQuery

Scenario: Your team provides you with a new dataset on the inventory stock levels for each of your products for sale on your ecommerce website. You want to become familiar with the products on the website and the fields you could use to potentially join on to other datasets.

The project with the new dataset is **data-to-insights**.

- In the Google Cloud console, in the **Navigation menu** (≡) click **BigQuery**.

The Welcome to BigQuery in the Cloud Console message box opens.

**Note:** The Welcome to BigQuery in the Cloud Console message box provides a link to the quickstart guide and UI updates.

- Click **Done**.
- BigQuery public datasets are not displayed by default. To open the public datasets project, copy **data-to-insights** (to paste in a dialog in the next step).

0/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next

6 15°C Mostly cloudy

Search

01:52 PM 31-12-2025

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:25:33 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

25/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

1. In the Google Cloud console, in the **Navigation menu** (≡) click **BigQuery**.  
The Welcome to BigQuery in the Cloud Console message box opens.  
**Note:** The Welcome to BigQuery in the Cloud Console message box provides a link to the quickstart guide and UI updates.

2. Click **Done**.

3. BigQuery public datasets are not displayed by default. To open the public datasets project, copy **data-to-insights** (to paste in a dialog in the next step).

4. Click **+ Add data > Star a project by name** then paste the data-to-insights name.

5. Click **Star**.

The **data-to-insights** project is listed in the **Explorer** section.

## Task 3. Examine the fields

Previous Next

6 15°C Mostly cloudy

Search

01:52 PM 31-12-2025

BigQuery – qwiklabs-gcp-03-253 x +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m0

Free AI Paraphrasing... Transcript of Case St... GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Explorer + Add data

Search for resources

Home Starred Shared with me Job history qwiklabs-gcp-03-253defc4cbc4 Datasets Connections Queries (Classic) Queries Notebooks Data canvases Data preparations Pipelines Repositories

Check out what's new in Studio

Files explorer Gemini in queries Apache Spark Visualization Cell Notebook gallery

Organize and share code assets [Preview](#)

The new Files explorer provides a dedicated view to organize your BigQuery assets (saved queries, notebooks, and more) into folders.

To get started with better project management and easier collaboration, click the Files icon at the top of the left pane. [Learn more](#)

Try it Next

Create new

SQL query Notebook Notebook with Spark ML model Data canvas Pipeline Data preparation Table

Job history Show

6 15°C Mostly cloudy

Search

01:53 PM 31-12-2025 ENG IN

This screenshot shows the Google Cloud BigQuery Studio interface. On the left, there's a sidebar with various navigation options like Home, Starred, and Project-specific sections. The main area features a 'Check out what's new in Studio' modal. This modal highlights the 'Files explorer' feature, which is currently selected. It includes a preview image showing a file structure with a folder icon circled in blue, and descriptive text explaining how it helps organize BigQuery assets. Below the modal, there are buttons for 'Try it' and 'Next'. At the bottom, there are buttons for creating new resources like SQL queries, notebooks, and pipelines, along with links for job history and data preparation. The bottom navigation bar includes standard browser controls and system status indicators like weather and battery level.

Add data – BigQuery – qwiklabs

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m0

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for re

Explorer + Add data

Search for resources

- Home
- Starred
- Shared with me
- Job history
- qwiklabs-gcp-03-253defc4cbc4
  - Datasets
  - Connections
  - Queries
  - (Classic) Queries
  - Notebooks
  - Data canvases
  - Data preparations
  - Pipelines
  - Repositories

Check out what's new in BigQuery

- Files explorer
- Gemini in queries
- Apache Spark
- Visualization Cell
- Notebook gallery

Create new

- SQL query
- Notebook

Recent Job history

Filter By

Add data

Search for data sources

Most popular data sources

- Databases (19)
- Marketing (38)
- Business Applications (30)
- Storage/Data Lakes (6)
- Data Warehouses (3)
- Data Streaming (4)

Public Datasets

Sharing (Analytics Hub)

Star a project by name

Close

15°C Mostly cloudy

Search

ENG IN 01:53 PM 31-12-2025

Google Cloud Storage Data lake storage and analysis.

Local File Data ingestion from local files.

Amazon Redshift Data warehousing and business intelligence.

Amazon S3 Data lake storage and analysis.

Apache Spark Unified engine for large-scale data analytics.

Azure Blob Storage Data lake storage and analysis.

BigQuery – qwiklabs-gcp-03-253 X +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m0

Free AI Paraphrasing... Transcript of Case St... GCP-LAB Hydra datawarehouse Products

Google Cloud Search (/) for resources, docs, products, and more Search

Explorer + Add data

Search for resources

Home Starred Shared with me Job history qwiklabs-gcp-03-253defc4cbc4 Datasets Connections Queries (Classic) Queries Notebooks Data canvases Data preparations Pipelines Repositories

Check out what's new in Studio

Files explorer Gemini in queries Apache Spark Visualization Cell Notebook gallery

Star a project Project name \* data-to-insights Cancel Star Try it Next

Organize and share code assets Preview

The new Files explorer provides a dedicated view to organize your BigQuery assets (saved queries, notebooks, and more) into folders.

To get started with better project management and easier collaboration, click the Files icon at the top of the left pane. [Learn more](#)

Create new SQL query Notebook Notebook with Spark ML model Data canvas Pipeline Data preparation Table

Job history Show

6 15°C Mostly cloudy

Search

01:53 PM 31-12-2025 ENG IN

BigQuery – qwiklabs-gcp-03-253 X +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m0

Free AI Paraphrasing... Transcript of Case St... GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Explorer + Add data

Search for resources

- Home
- Starred
- Shared with me
- Job history
- qwiklabs-gcp-03-253defc4cbc4
- data-to-insights
  - Datasets
  - Connections
  - (Classic) Queries

Check out what's new in Studio

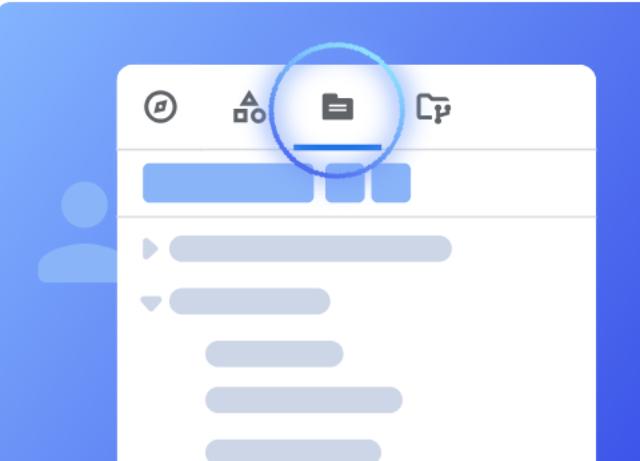
Files explorer

Gemini in queries

Apache Spark

Visualization Cell

Notebook gallery



Try it Next

Organize and share code assets [Preview](#)

The new Files explorer provides a dedicated view to organize your BigQuery assets (saved queries, notebooks, and more) into folders.

To get started with better project management and easier collaboration, click the Files icon at the top of the left pane. [Learn more](#)

Create new

SQL query Notebook Notebook with Spark ML model Data canvas Pipeline Data preparation Table

Job history Show

6 15°C Mostly cloudy

Search

01:54 PM 31-12-2025 ENG IN

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:24:08 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAK3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

name.

5. Click Star.

The `data-to-insights` project is listed in the **Explorer** section.

25/100

Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

## Task 3. Examine the fields

Next, get familiar with the products and fields on the website you can use to create queries to analyze the dataset.

- In the left pane in the Resources section, navigate to `data-to-insights > ecommerce > all_sessions_raw`.
- On the right, under the Query editor, click the **Schema** tab to see the Fields and information about each field.

## Task 4. Identify a key field in your ecommerce dataset

Previous Next

6 15°C Mostly cloudy

Search

01:54 PM 31-12-2025

The screenshot shows the Google BigQuery web interface. The left sidebar contains navigation icons and a search bar for 'BigQuery resources'. The main area displays the 'ecommerce' dataset under 'data-to-insights / Datasets / ecommerce'. The 'Overview' tab is selected, showing a table of tables with columns: Table ID, Type, Create time, Expiration time, and Label. The 'Tables' tab is active, showing 13 tables: all\_sessions, all\_sessions\_raw, categories, checkout\_nudge, classification\_model\_2\_results, days\_with\_rain, partition\_by\_day, partitions, product\_list, products, and rev\_transactions. Each table row includes a link to its details. The 'Details' tab is also visible. At the top right, there are buttons for 'Create Table', 'Share', 'Copy', 'Delete', and 'Refresh'. A 'Job history' section is at the bottom.

Table ID	Type	Create time	Expiration time	Label
<a href="#">all_sessions</a>	Table	May 29, 2018, 5:40:55P...	None	None
<a href="#">all_sessions_raw</a>	Table	May 29, 2018, 1:57:04A...	None	None
<a href="#">categories</a>	Table	Jul 9, 2018, 4:18:53AM ...	None	None
<a href="#">checkout_nudge</a>	Table	Jul 12, 2018, 2:27:00AM ...	None	None
<a href="#">classification_model_2_results</a>	Table	Mar 18, 2020, 2:00:29A...	None	None
<a href="#">days_with_rain</a>	Table	Apr 21, 2019, 5:02:20AM...	None	None
<a href="#">partition_by_day</a>	Table	Jul 9, 2018, 6:14:41AM ...	None	None
<a href="#">partitions</a>	Table	Apr 23, 2019, 3:59:32AM...	None	None
<a href="#">product_list</a>	Table	Jun 21, 2018, 10:46:39P...	None	None
<a href="#">products</a>	Table	Jul 9, 2018, 4:18:40AM ...	None	None
<a href="#">rev_transactions</a>	Table	May 22, 2018, 8:56:18P...	None	None

BigQuery – qwiklabs-gcp-03-253 X +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m5!1m4!4m3!1sdata-to-insights!2secommerce!3sall\_s...

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Untitled...ery X all\_sessions\_raw X +

data-to-insights / Datasets / ecommerce / Tables / all\_sessions\_raw

all\_sessions\_raw Query Open in Share Copy Snapshot Delete Export Refresh

Show starred only

Show more

- data-to-insights
  - Connections
  - advanced
  - customer\_insights
- ecommerce
  - Models (2)
  - all\_sessions
  - all\_sessions\_raw
  - categories
  - checkout\_nudge
  - classification\_m
  - days\_with\_rain
  - partition\_by\_da
  - partitions
  - product\_list

Search BigQuery resources

Filter Enter property name or value

Field name	Type	Mode	Description	Key	Collation	Default Value	Policy Tags	Data Policies
fullVisitorId	STRING	NULLABLE	-	-	-	-	-	-
channelGrouping	STRING	NULLABLE	-	-	-	-	-	-
time	INTEGER	NULLABLE	-	-	-	-	-	-
country	STRING	NULLABLE	-	-	-	-	-	-
city	STRING	NULLABLE	-	-	-	-	-	-
totalTransactionRevenue	INTEGER	NULLABLE	-	-	-	-	-	-
transactions	INTEGER	NULLABLE	-	-	-	-	-	-
timeOnSite	INTEGER	NULLABLE	-	-	-	-	-	-
pageviews	INTEGER	NULLABLE	-	-	-	-	-	-
sessionQualityDim	INTEGER	NULLABLE	-	-	-	-	-	-

Edit schema View row access policies

Job history Show

6 15°C Mostly cloudy

Search

01:56 PM 31-12-2025



Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc4



Student Resources

[Get Meaningful Insights with](#)[Previous](#)

## Task 4. Identify a key field in your ecommerce dataset

Examine the products and fields further. You want to become familiar with the products on the website and the fields you could use to potentially join on to other datasets.

### Examine the records

In this section you find how many product names and product SKUs are on your website and whether either one of those fields is unique.

- Find how many product names and product SKUs are on the website. **Copy and Paste** the below query in bigquery **EDITOR**:

```
#standardSQL
# how many products are on the website?
SELECT DISTINCT
productSKU,
v2ProductName
FROM `data-to-insights_ecommerce_all_sessions_raw`
```

[Next >](#)

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:21:52 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources Get Meaningful Insights with

Examine the records

In this section you find how many product names and product SKUs are on your website and whether either one of those fields is unique.

1. Find how many product names and product SKUs are on the website. **Copy and Paste** the below query in bigquery **EDITOR**:

```
#standardSQL  
# how many products are on the website?  
SELECT DISTINCT  
productSKU,  
v2ProductName  
FROM `data-to-insights.ecommerce.all_sessions_raw`
```

2. Click **Run**.

Look at the pagination results in the console for the total number of records returned.

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.8 sec elapsed, 1,013 MB processed)

25/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next





G Build a Data Warehouse with BigQuery +

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:21:05 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. [Learn more.](#)

Open Google Cloud console

Username: student-02-efbe2bad126b (copy)

Password: qQAk3jWEKbic (copy)

Project ID: quicklabs-gcp-03-253defc (copy)

Student Resources: Get Meaningful Insights with

2,205 products and SKUs  
1,925 products and SKUs

Submit

But...do the results mean that there are that many unique product SKUs? One of the first queries you will run as a data analyst is looking at the uniqueness of your data values.

3. Clear the previous query and run the below query to list the number of distinct SKUs are listed using DISTINCT:

```
#standardSQL
# find the count of unique SKUs
SELECT
DISTINCT
productSKU
FROM `data-to-insights.ecommerce.all_sessions_raw`
```

How many DISTINCT SKUs are returned?

25/100

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

01:57 PM 31-12-2025



01:19:59  
Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc4



Student Resources

[Get Meaningful Insights with](#)

How many DISTINCT SKUs are returned?

 119 distinct SKUs 1,909 distinct SKUs 2,273 distinct SKUs[Submit](#)

There are fewer DISTINCT SKUs than the SKU & Product Name query had before. Why do you think that is?

 The first query was excluding some Product Names. The first query showed that only one Product Name can belong to a SKU. The first query also returned Product Name. It appears multiple Product Names can have the same SKU.

Lab instructions and tasks

25/100

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

[Previous](#)[Next](#)

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:19:51 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with...

Examine the relationship between SKU & Name

Now determine which products have more than one SKU and which SKUs have more than one Product Name.

1. Clear the previous query and run the below query to determine if some product names have more than one SKU. The use of the STRING\_AGG() function to aggregate all the product SKUs that are associated with one product name into comma separated values.

```
SELECT v2ProductName,
       COUNT(DISTINCT productSKU) AS SKU_count,
       STRING_AGG(DISTINCT productSKU LIMIT 5) AS SKU
FROM `data-to-insights.ecommerce.all_sessions_raw`
WHERE productSKU IS NOT NULL
GROUP BY v2ProductName
HAVING SKU_count > 1
ORDER BY SKU_count DESC
```

2. Click Run.

25/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

01:58 PM 31-12-2025



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:17:47 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: qwiklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

Do some product names have more than one SKU? Look at the query results to confirm

No Yes

Submit

Which product has the most SKUs associated?

Google Sunglasses Waze Womens Typography Short Sleeve Tee Android Womens Short Sleeve Badge Tee Dark Heather

25/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

02:00 PM 31-12-2025



BigQuery – qwiklabs-gcp-03-253 X +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m10!1m4!4m3!1sdata-to-insights!2secommerce!3sall... | 1

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Untitled query Run Save Download Share Schedule Open in More

```
1 SELECT
2   productSKU,
3   COUNT(DISTINCT v2ProductName) AS product_count,
4   STRING_AGG(DISTINCT v2ProductName LIMIT 5) AS product_name
5 FROM `data-to-insights.ecommerce.all_sessions_raw`
6 WHERE v2ProductName IS NOT NULL
7 GROUP BY productSKU
8 HAVING product_count > 1
9 ORDER BY product_count DESC
```

Query completed Using on-demand processing quota

Query results Save results Open in

Row	productSKU	product_count	product_name
1	9181664	3	Waterproof Gear Bag,Waterpoof Gear Bag,Google Small Waterproof Duffel
2	GGOEGCLB020832	3	Softsided Travel Pouch Set/Set of 3 Nested Travel

Results per page: 50 1 – 50 of 347 < > >>

Job history Show

Low visibility Now

Search

ENG IN 02:02 PM 31-12-2025

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:14:23 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b6f2c [copy]  
Password: qQAk3jWEKbic [copy]  
Project ID: qwiklabs-gcp-03-253defca [copy]

Student Resources: Get Meaningful Insights with

Job History, Query History, Saved Queries

Note: Try replacing STRING\_AGG() with ARRAY\_AGG() instead. Pretty cool, right? BigQuery natively supports nested array values. You can learn more from the [Work with arrays guide](#).

When you look at the query results, are there single SKU values with more than one product name associated? What do you notice about those product names?

No, the Product SKUs match the Product Names one-for-one.  
 Yes, most of the product names are similar but not exactly the same.

Submit

You will see why this many-to-many data relationship will be an issue in the next section.

Click **Check my progress** to verify the objective.

25/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Low visibility Now

Search

02:04 PM 31-12-2025



Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defca



Student Resources

[Get Meaningful Insights with](#)

## Task 5. Pitfall: non-unique key

In inventory tracking, a SKU is designed to uniquely identify one and only one product. For us, it will be the basis of your JOIN condition when you lookup information from other tables. Having a non-unique key can cause serious data issues as you will see.

1. Write a query to identify all the product names for the SKU 'GG0EGPJC019099'.

Possible solution:

```
SELECT DISTINCT
    v2ProductName,
    productSKU
FROM `data-to-insights.ecommerce.all_sessions_raw`
WHERE productSKU = 'GG0EGPJC019099'
```

2. Click Run.

v2ProductName	productSKU
Product A	GG0EGPJC019099

[Previous](#)

50/100

Lab instructions and task

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

[Next](#)



01:13:07  
End Lab

Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc



Student Resources

[Get Meaningful Insights with](#)

```
SELECT v2ProductName,  
       productSKU  
FROM `data-to-insights.ecommerce.all_sessions_raw`  
WHERE productSKU = 'GG0EGPJC019099'
```

50/100

Lab instructions and task

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

2. Click Run.

v2ProductName	productSKU
7&quot; Dog Frisbee	GG0EGPJC019099
7" Dog Frisbee	GG0EGPJC019099
Google 7-inch Dog Flying Disc Blue	GG0EGPJC019099

What do you notice about the product names?

- They are exactly the same.  
 They are mostly the same except for a few characters.

[Submit](#)[Previous](#)[Next](#)



Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAK3jWEKbic



Project ID

qwiklabs-gcp-03-253defc



Student Resources

[Get Meaningful Insights with](#)

From the query results, it looks like there are three different names for the same product. In this example, there is a special character in one name and a slightly different name for another:

50/100

Lab instructions and task

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

## Joining website data against your product inventory list

Now see the impact of joining on a dataset with multiple products for a single SKU. First explore the product inventory dataset (the `products` table) to see if this SKU is unique there.

- Clear the previous query and run the below query:

```
SELECT
SKU,
name,
stockLevel
FROM `data-to-insights.ecommerce.products`
WHERE SKU = 'GGOEGPJ019099'
```

Is the SKU unique in the product inventory dataset?

[Previous](#)[Next](#)



The screenshot shows a Google Cloud lab interface. On the left, there's a sidebar with 'Contents' and a dropdown menu for 'Troubleshooting and Solving Data Join Pitfalls'. The main area has a timer at 01:08:40 and a red 'End Lab' button. A 'Caution' message reminds users not to deviate from instructions. Below it are input fields for 'Username' (student-02-efbe2bad126b), 'Password' (qQAk3jWEKbic), and 'Project ID' (qwiklabs-gcp-03-253defc). A 'Student Resources' section includes a link to 'Get Meaningful Insights with'. The main content area contains two questions:

**Is the SKU unique in the product inventory dataset?**

- Yes, just one record is returned.
- No, there are duplicate SKUs in the inventory dataset.

**How many dog frisbees do you have in inventory?**

- 0
- 154
- 10,540

Both questions have a 'Submit' button below them. On the right side, there's a vertical navigation bar with sections like 'Lab instructions and tasks', 'GSP412', 'Overview', 'Setup and requirements', 'Task 1.', 'Task 2.', 'Task 3.', 'Task 4.', 'Task 5.', 'Task 6.', and a progress bar at 50/100.

G Build a Data Warehouse with BigQuery +

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Join pitfall: Unintentional many-to-one SKU relationship

End Lab 01:08:32 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

50/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

SELECT DISTINCT website.v2ProductName, website.productSKU, inventory.stockLevel FROM `data-to-insights.ecommerce.all\_sessions\_raw` AS website JOIN `data-to-insights.ecommerce.products` AS inventory ON website.productSKU = inventory.SKU WHERE productSKU = 'GGOEGPJC019099'

What happens when you join the website table and the product inventory table on SKU? Do you now have inventory stock levels for the product?

Previous Next

6 15°C Mostly cloudy

Search

02:09 PM 31-12-2025



01:07:54  
End Lab

Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)

[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc4



Student Resources

[Get Meaningful Insights with](#)

```
SELECT DISTINCT
    website.v2ProductName,
    website.productSKU,
    inventory.stockLevel
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
JOIN `data-to-insights.ecommerce.products` AS inventory
    ON website.productSKU = inventory.SKU
WHERE productSKU = 'GGOEGPJC019099'
```

What happens when you join the website table and the product inventory table on SKU? Do you now have inventory stock levels for the product?

- No, there is no inventory data, the join did not work.
- Yes, there are inventory levels but the stockLevel is showing three times (one for each record).
- Yes, there is inventory data and everything looks fine.

[Submit](#)

Lab instructions and tasks

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

[Previous](#)[Next](#)



Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc



Student Resources

[Get Meaningful Insights with](#)

2. Clear the previous query and run the below query:

```
WITH inventory_per_sku AS (
    SELECT DISTINCT
        website.v2ProductName,
        website.productSKU,
        inventory.stockLevel
    FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
    JOIN `data-to-insights.ecommerce.products` AS inventory
        ON website.productSKU = inventory.SKU
    WHERE productSKU = 'GG0EGPJC019099'
)

SELECT
    productSKU,
    SUM(stockLevel) AS total_inventory
FROM inventory_per_sku
GROUP BY productSKU
```

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

[Previous](#)[Next](#)





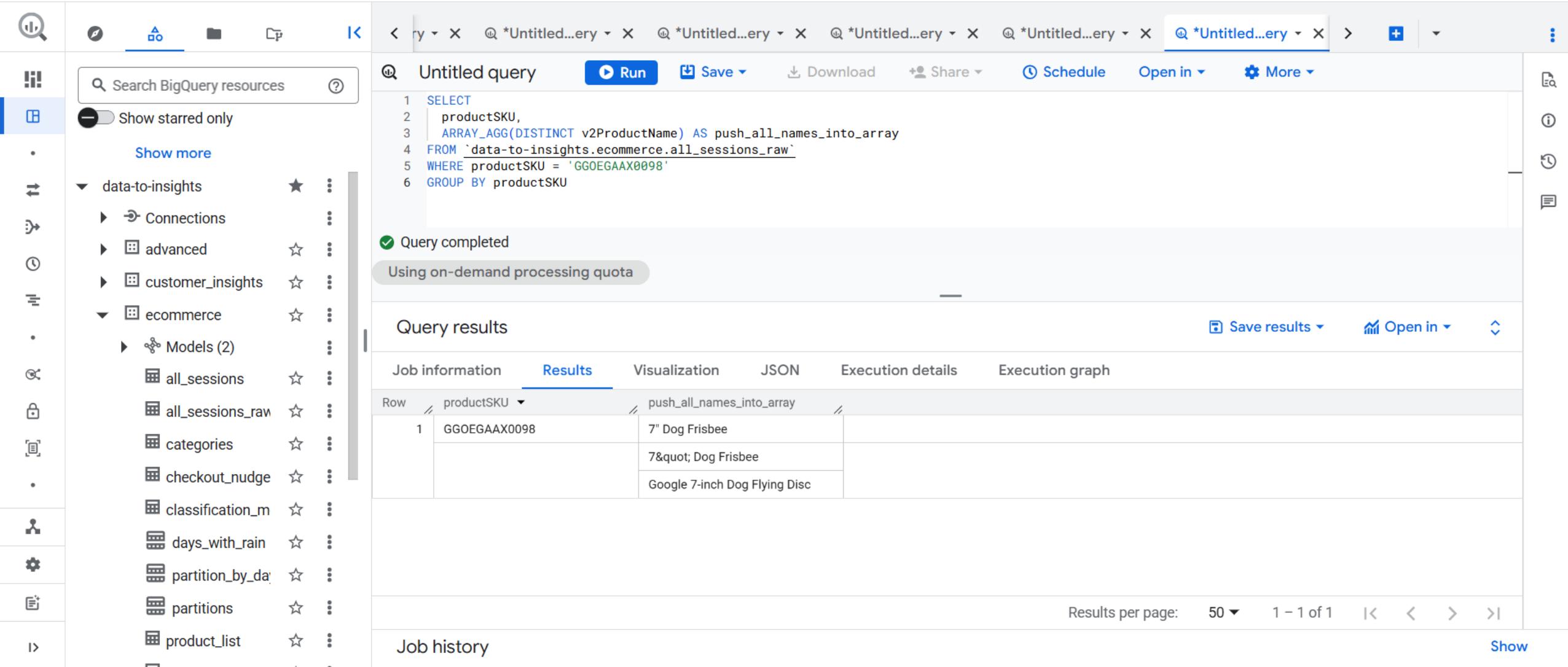


☰ Google Cloud

 [qwiklabs-gcp-03-253defc4cbc](#)

Search (/) for resources, docs, products, and more

 Search





Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc



Student Resources

[Get Meaningful Insights with](#)[Previous](#)

Now instead of having a row for every Product Name, you only have a row for each unique SKU.

2. If you wanted to deduplicate the product names, you could even LIMIT the array like so:

```
SELECT
    productSKU,
    ARRAY_AGG(DISTINCT v2ProductName LIMIT 1) AS
    push_all_names_into_array
FROM `data-to-insights.ecommerce.all_sessions_raw`
WHERE productSKU = 'GGOEGAAX0098'
GROUP BY productSKU
```

## Join pitfall: losing data records after a join

Now you're ready to join against your product inventory dataset again.

1. Clear the previous query and run the below query:

[Next](#)



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:05:28 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

Join pitfall: losing data records after a join

Now you're ready to join against your product inventory dataset again.

1. Clear the previous query and run the below query:

```
#standardSQL
SELECT DISTINCT
website.productSKU
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
JOIN `data-to-insights.ecommerce.products` AS inventory
ON website.productSKU = inventory.SKU
```

How many records were returned? All 1,909 distinct SKUs?

✓ No, just 1,090 records  
✗ Yes, all 1,909 records

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

02:12 PM 31-12-2025



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:05:15 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources Get Meaningful Insights with

It seems 819 SKUs were lost after joining the datasets. Investigate by adding more specificity in your fields (one SKU column from each dataset):

2. Clear the previous query and run the below query:

```
#standardSQL  
# pull ID fields from both tables  
SELECT DISTINCT  
website.productSKU AS website_SKU,  
inventory.SKU AS inventory_SKU  
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website  
JOIN `data-to-insights.ecommerce.products` AS inventory  
ON website.productSKU = inventory.SKU  
# IDs are present in both tables, how can you dig deeper?
```

It appears the SKUs are present in both of those datasets after the join for these 1,090 records. How can you find the missing records?

Join pitfall solution: selecting the correct join type and filtering for NULL

The default JOIN type is an INNER JOIN which returns records only if there is a SKU

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Next >

6 15°C Mostly cloudy

Search

02:13 PM 31-12-2025



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:04:50 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources Get Meaningful Insights with

1. Rewrite the previous query to use a different join type to include all records from the website table, regardless of whether there is a match on a product inventory SKU record. Join type options: INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN, CROSS JOIN.

Possible solution:

```
#standardSQL
# the secret is in the JOIN type
# pull ID fields from both tables
SELECT DISTINCT
website.productSKU AS website_SKU,
inventory.SKU AS inventory_SKU
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
LEFT JOIN `data-to-insights.ecommerce.products` AS inventory
ON website.productSKU = inventory.SKU
```

2. Click Run.

You have successfully used a LEFT JOIN to return all of the original 1,909 website SKUs in your results.

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next

6 15°C Mostly cloudy

Search

02:13 PM 31-12-2025





Contents

Troubleshooting and Solving Data Join Pitfalls

Dashboard

Catalog

Paths

Collections

01:04:21  

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.  
[Learn more.](#)

[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defca



Student Resources

[Get Meaningful Insights with](#)

You have successfully used a LEFT JOIN to return all of the original 1,909 website SKUs in your results.

True or False: Many inventory SKU values are NULL.

- True  
 False

[Submit](#)

How many SKUs are missing from your product inventory set?

1. Write a query to filter on NULL values from the inventory table.

Possible solution:

```
#standardSQL
# find product SKUs in website table but not in product
inventory table
```

- 75/100
- Lab instructions and tasks
- GSP412
- Overview
- Setup and requirements
- Task 1. Create a new dataset to store your tables
- Task 2. Pin the lab project in BigQuery
- Task 3. Examine the fields
- Task 4. Identify a key field in your ecommerce dataset
- Task 5. Pitfall: non-unique key
- Task 6. Join pitfall solution: use distinct

[Previous](#)[Next](#)

Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:04:13 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: qwiklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with...

Submit

How many SKUs are missing from your product inventory set?

1. Write a query to filter on NULL values from the inventory table.

Possible solution:

```
#standardSQL
# find product SKUs in website table but not in product
inventory table
SELECT DISTINCT
website.productSKU AS website_SKU,
inventory.SKU AS inventory_SKU
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
LEFT JOIN `data-to-insights.ecommerce.products` AS inventory
ON website.productSKU = inventory.SKU
WHERE inventory.SKU IS NULL
```

2. Click Run.

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

02:14 PM 31-12-2025



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:03:35 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAK3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

website.productSKU AS website\_SKU,  
inventory.SKU AS inventory\_SKU  
FROM `data-to-insights.ecommerce.all\_sessions\_raw` AS website  
LEFT JOIN `data-to-insights.ecommerce.products` AS inventory  
ON website.productSKU = inventory.SKU  
WHERE inventory.SKU IS NULL

2. Click Run.

Question: How many products are missing?

Answer: 819 products are missing (SKU IS NULL) from your product inventory dataset.

• Clear the previous query and run the below query to confirm using one of the specific SKUs from the website dataset:

```
#standardSQL  
# you can even pick one and confirm  
SELECT * FROM `data-to-insights.ecommerce.products`  
WHERE SKU = 'GGOEGATJ060517'  
# query returns zero results
```

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

02:14 PM 31-12-2025





01:03:05

Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc



Student Resources

[Get Meaningful Insights with](#)

```
WHERE SKU = 'GGOEGATJ060517'  
# query returns zero results
```

Why might the product inventory dataset be missing SKUs?

- Some SKUs could be digital products that you do not store in warehouse inventory
- Old products you sold in past website orders are no longer offered in current inventory
- There is legitimate missing data from inventory and should be tracked

All of the above

[Submit](#)

Now, what about the reverse situation? Are there any products in the product inventory dataset but missing from the website?

1. Write a query using a different join type to investigate.

[Previous](#)[Next](#)

- 75/100
- Lab instructions and tasks
- GSP412
  - Overview
  - Setup and requirements
  - Task 1. Create a new dataset to store your tables
  - Task 2. Pin the lab project in BigQuery
  - Task 3. Examine the fields
  - Task 4. Identify a key field in your ecommerce dataset
  - Task 5. Pitfall: non-unique key
  - Task 6. Join pitfall solution: use distinct



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

End Lab 01:02:57 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

Submit

Now, what about the reverse situation? Are there any products in the product inventory dataset but missing from the website?

1. Write a query using a different join type to investigate.

Possible solution:

```
#standardSQL
# reverse the join
# find records in website but not in inventory
SELECT DISTINCT
website.productSKU AS website_SKU,
inventory.SKU AS inventory_SKU
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
RIGHT JOIN `data-to-insights.ecommerce.products` AS inventory
ON website.productSKU = inventory.SKU
WHERE website.productSKU IS NULL
```

75/100

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:02:32 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: qwiklabs-gcp-03-253defc...

Student Resources Get Meaningful Insights with

Answer: Yes. There are two product SKUs missing from the website dataset

Next, add more fields from the product inventory dataset for more details.

- Clear the previous query and run the below query:

```
#standardSQL  
# what are these products?  
# add more fields in the SELECT STATEMENT  
SELECT DISTINCT  
website.productSKU AS website_SKU,  
inventory.*  
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website  
RIGHT JOIN `data-to-insights.ecommerce.products` AS inventory  
ON website.productSKU = inventory.SKU  
WHERE website.productSKU IS NULL
```

Why would the below products be missing from the ecommerce website dataset?

website_SKU	SKU	name	orderedQuantity	stockLevel	restockingLeadTime	sentimentScore	sentimentMagnitude
		USB wired					

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next

6 15°C Mostly cloudy

Search

02:15 PM 31-12-2025



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

simply just do a LEFT JOIN and switch the ordering of the tables.

What if you wanted one query that listed all products missing from either the website or inventory?

1. Write a query using a different join type.

Possible solution:

```
#standardSQL
SELECT DISTINCT
    website.productSKU AS website_SKU,
    inventory.SKU AS inventory_SKU
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
FULL JOIN `data-to-insights.ecommerce.products` AS inventory
ON website.productSKU = inventory.SKU
WHERE website.productSKU IS NULL OR inventory.SKU IS NULL
```

2. Click Run.

You have your 819 + 2 = 821 product SKUs.

End Lab 01:02:10 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: qwiklabs-gcp-03-253defca

Student Resources: Get Meaningful Insights with

Dashboard Catalog Paths Collections

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

02:16 PM 31-12-2025



Build a Data Warehouse with BigQuery

https://partner.skills.google/course\_templates/624/labs/597936

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:01:46 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAK3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

regardless of matching join keys. You then filter out where you have mismatches on either side

Join pitfall: unintentional cross join

Not knowing the relationship between data table keys (1:1, 1:N, N:N) can return unexpected results and also significantly reduce query performance.

The last join type is the CROSS JOIN.

Create a new table with a site-wide discount percent that you want applied across products in the Clearance category.

1. Clear the previous query and run the below query:

```
#standardSQL
CREATE OR REPLACE TABLE ecommerce.site_wide_promotion AS
SELECT .05 AS discount;
```

In the left pane, `site_wide_promotion` is now listed in the Resource section under your project and dataset.

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

15°C Mostly cloudy

Search

02:16 PM 31-12-2025



BigQuery – qwiklabs-gcp-03-253 x +

https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m10!1m4!4m3!1sdata-to-insights!2secommerce!3sall... | 1

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Untitled query Run Save Download Share Schedule Open in More

```
1 #standardSQL
2 CREATE OR REPLACE TABLE ecommerce.site_wide_promotion AS
3 SELECT .05 AS discount;
```

Search BigQuery resources Show starred only

site\_wide\_promotion Click to open in current tab (Ctrl + Click - open in new tab, Shift + Ctrl + Click - open in split tab)

Show more Show more

data-to-insights Connections advanced customer\_insights ecommerce Models (2) all\_sessions all\_sessions\_raw categories checkout\_nudge classification\_model\_2\_resu

Query completed

Query results Save results Open in

Job information Results Execution details Execution graph

This statement created a new table named site\_wide\_promotion. Go to table

Job history Show

6 15°C Mostly cloudy

Search

02:17 PM 31-12-2025 ENG IN



## Troubleshooting and Solving Data Join Pitfalls

In the left pane, `site_wide_promotion` is now listed in the Resource section under your project and dataset.

- Clear the previous query and run the below query to find out how many products are in clearance:

```
SELECT DISTINCT
productSKU,
v2ProductCategory,
discount
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
CROSS JOIN ecommerce.site_wide_promotion
WHERE v2ProductCategory LIKE '%Clearance%'
```

How many products are on clearance?

82

52

0

Lab instructions and tasks

75/100

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

< Previous

Next >



G Build a Data Warehouse with BigQuery +

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Skills Partner

Build a Data Warehouse with BigQuery > Troubleshooting and Solving Data Join Pitfalls

Contents Troubleshooting and Solving Data Join Pitfalls

Dashboard Catalog Paths Collections

End Lab 01:00:25 Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Cloud console

Username: student-02-efbe2bad126b...

Password: qQAk3jWEKbic

Project ID: quicklabs-gcp-03-253defc...

Student Resources: Get Meaningful Insights with

See the impact of unintentionally adding more than one record in the discount table.

3. Clear the previous query and run the below query to insert two more records into the promotion table:

```
INSERT INTO ecommerce.site_wide_promotion (discount)  
VALUES (.04),  
(.03);
```

Next, view the data values in the promotion table.

4. Clear the previous query and run the below query:

```
SELECT discount FROM ecommerce.site_wide_promotion
```

How many records were returned?

Answer: 3

What happens when you apply the discount again across all 82 clearance products?

75/100 Lab instructions and tasks

GSP412 Overview Setup and requirements Task 1. Create a new dataset to store your tables Task 2. Pin the lab project in BigQuery Task 3. Examine the fields Task 4. Identify a key field in your ecommerce dataset Task 5. Pitfall: non-unique key Task 6. Join pitfall solution: use distinct

Previous Next

6 15°C Mostly cloudy

Search

02:18 PM 31-12-2025









Free AI Paraphrasin...

Transcript of Case St...

GCP-LAB

Hydra

datawarehouse

Products



Google Skills

Partner



6472

5

[Build a Data Warehouse with BigQuery](#) > [Troubleshooting and Solving Data Join Pitfalls](#)

Contents

Troubleshooting and Solving Data Join Pitfalls

Dashboard

Catalog

Paths

Collections

00:59:21

Time limit

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defc



Student Resources

[Get Meaningful Insights with](#)

```
SELECT discount FROM ecommerce.site_wide_promotion
```



How many records were returned?

**Answer:** 3

What happens when you apply the discount again across all 82 clearance products?

5. Clear the previous query and run the below query:

```
SELECT DISTINCT
productSKU,
v2ProductCategory,
discount
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
CROSS JOIN ecommerce.site_wide_promotion
WHERE v2ProductCategory LIKE '%Clearance%'
```

Copied!



How many products are returned?

**Answer:** Instead of 82, you now have 246 returned which is more records than your original table started with

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

[Previous](#)[Next](#)

BigQuery – qwiklabs-gcp-03-253 X + https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m10!1m4!4m3!1sqwiklabs-gcp-03-253defc4cbc4!2se... | 1

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Untitled query Run Save Download Share Schedule Open in More

```
1 SELECT DISTINCT
2 productSKU,
3 v2ProductCategory,
4 discount
5 FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
6 CROSS JOIN ecommerce.site_wide_promotion
7 WHERE v2ProductCategory LIKE '%Clearance%'
```

Query completed Using on-demand processing quota

Query results Job information Results Visualization JSON Execution details Execution graph

Row	productSKU	v2ProductCategory	discount
1	GGOEGAAX0168	Home/Clearance Sale/	0.04
2	GGOEGAAX0329	Home/Clearance Sale/	0.03
3	GGOEGAAX0329	Home/Clearance Sale/	0.05
4	GGOEGAAX0334	Home/Clearance Sale/	0.04
5	GGOEGAAX0340	Home/Clearance Sale/	0.03
6	GGOEGAAX0340	Home/Clearance Sale/	0.05
7	GGOEGAAX0344	Home/Clearance Sale/	0.04

Results per page: 50 1 – 50 of 246 Show

Job history

6 15°C Mostly cloudy Search

ENG IN 02:19 PM 31-12-2025



Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.

[Learn more.](#)[Open Google Cloud console](#)

Username

student-02-efbe2bad126b



Password

qQAk3jWEKbic



Project ID

qwiklabs-gcp-03-253defca



Student Resources

[Get Meaningful Insights with](#)

How many products are returned?

**Answer:** Instead of 82, you now have 246 returned which is more records than your original table started with.

Now investigate the underlying cause by examining one product SKU.

6. Clear the previous query and run the below query:

```
#standardSQL
SELECT DISTINCT
productSKU,
v2ProductCategory,
discount
FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
CROSS JOIN ecommerce.site_wide_promotion
WHERE v2ProductCategory LIKE '%Clearance%'
AND productSKU = 'GGOEGOLC013299'
```

What was the impact of the CROSS JOIN?

**Answer:** Since there are 3 discount codes to cross join on, you are multiplying the original dataset by 3.

Lab instructions and tasks

GSP412

Overview

Setup and requirements

Task 1. Create a new dataset to store your tables

Task 2. Pin the lab project in BigQuery

Task 3. Examine the fields

Task 4. Identify a key field in your ecommerce dataset

Task 5. Pitfall: non-unique key

Task 6. Join pitfall solution: use distinct

[Previous](#)[Next](#)

BigQuery – qwiklabs-gcp-03-253 X + https://console.cloud.google.com/bigquery?project=qwiklabs-gcp-03-253defc4cbc4&ws=!1m10!1m4!4m3!1sqwiklabs-gcp-03-253defc4cbc4!2se... | 1

Free AI Paraphrasing... Transcript of Case Study GCP-LAB Hydra datawarehouse Products

Google Cloud qwiklabs-gcp-03-253defc4cbc4 Search (/) for resources, docs, products, and more Search

Untitled query Run Save Download Share Schedule Open in More

```
1 #standardSQL
2 SELECT DISTINCT
3 productSKU,
4 v2ProductCategory,
5 discount
6 FROM `data-to-insights.ecommerce.all_sessions_raw` AS website
7 CROSS JOIN ecommerce.site_wide_promotion
8 WHERE v2ProductCategory LIKE '%Clearance%'
9 AND productSKU = 'GGOEGOLC013299'
```

Query completed Using on-demand processing quota

Query results Job information Results Visualization JSON Execution details Execution graph

Row	productSKU	v2ProductCategory	discount
1	GGOEGOLC013299	Home/Clearance Sale/	0.04
2	GGOEGOLC013299	Home/Clearance Sale/	0.03
3	GGOEGOLC013299	Home/Clearance Sale/	0.05

Results per page: 50 1 – 3 of 3 Show

Job history

6 15°C Mostly cloudy Search

ENG IN 02:19 PM 31-12-2025