

[Google Discussions](#)

Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

[Go to Exam](#)

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 48 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 48

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

Your company is loading comma-separated values (CSV) files into Google BigQuery. The data is fully imported successfully; however, the imported data is not matching byte-to-byte to the source file. What is the most likely cause of this problem?

- A. The CSV data loaded in BigQuery is not flagged as CSV.
- B. The CSV data has invalid rows that were skipped on import.
- C. The CSV data loaded in BigQuery is not using BigQuery's default encoding.
- D. The CSV data has not gone through an ETL phase before loading into BigQuery.

[Show Suggested Answer](#)

by [deleted] at March 21, 2020, 8:36 a.m.

Comments

[Submit](#)

[Removed] [Highly Voted](#) 3 years, 7 months ago

Answer: C

Description: Bigquery understands UTF-8 encoding anything other than that will result in data issues with schema

upvoted 26 times

🗄️ 👤 **YAS007** Highly Voted 👍 2 years, 2 months ago

Answer : C :

" If you don't specify an encoding, or if you specify UTF-8 encoding when the CSV file is not UTF-8 encoded, BigQuery attempts to convert the data to UTF-8. Generally, your data will be loaded successfully, but it may not match byte-for-byte what you expect."

https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv#details_of_loading_csv_data

👍 🔄 🚩 upvoted 17 times

🗄️ 👤 **desertlotus1211** Most Recent 🕒 1 month, 4 weeks ago

Selected Answer: B

The byte-to-byte mismatch is more consistent with invalid rows being skipped during the load process (due to format or parsing issues), rather than an encoding issue.

Answer B

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **samdhimal** 9 months, 3 weeks ago

SITUATION:

- Your company is loading comma-separated values (CSV) files into Google BigQuery.
- Data is fully imported successfully.

PROBLEM:

- Imported data is not matching byte-to-byte to the source file. Reason?

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **samdhimal** 9 months, 3 weeks ago

A. The CSV data loaded in BigQuery is not flagged as CSV.

Since BigQuery support multiple formats it could be that maybe avro or json was selected.

But the file import was successful hence csv was selected. Either manually or it was left as is since the default file type is csv. Lastly, this is WRONG.

B. The CSV data has invalid rows that were skipped on import.

-> Since the data was successfully imported there were no invalid rows. Hence, This is wrong answer too.

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **samdhimal** 9 months, 3 weeks ago

C. The CSV data loaded in BigQuery is not using BigQuery's default encoding.

-> "BigQuery supports UTF-8 encoding for both nested or repeated and flat data. BigQuery supports ISO-8859-1 encoding for flat data only for CSV files."

Source: <https://cloud.google.com/bigquery/docs/loading-data>

Default BQ Encoding: UTF-8

This is probably the correct answer because if the csv file encoding was not UTF-8 and instead it was ISO-8859-1 then we would have to tell bigquery that or else it will assume it is UTF-8. Hence, Imported data is not matching byte-to-byte to the source file. CORRECT ANSWER!

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **samdhimal** 9 months, 3 weeks ago

D. The CSV data has not gone through an ETL phase before loading into BigQuery.

-> ETL means Extract, Transform and Load and this is actually very important content for Cloud Data Engineers.

Look into it if interested! But getting back to the topic: ETL is usually required when the source format and target format are different. You need to extract source file and the transform it before loading the data to fit the target. This is also not a viable option. Also Data is imported successfully and the question doesn't mention anything regarding ETL.

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **medeis_jar** 1 year, 10 months ago

Selected Answer: C

A is not correct because if another data format other than CSV was selected then the data would not import successfully.

B is not correct because the data was fully imported meaning no rows were skipped.

C is correct because this is the only situation that would cause successful import.

D is not correct because whether the data has been previously transformed will not affect whether the source file will match the BigQuery table.

👍 🔄 🚩 upvoted 6 times

🗄️ 👤 **MaxNRG** 1 year, 11 months ago

Selected Answer: C

C is correct because this is the only situation that would cause successful import.

A is not correct because if another data format other than CSV was selected then the data would not import successfully.

B is not correct because the data was fully imported meaning no rows were skipped.

D is not correct because whether the data has been previously transformed will not affect whether the source file will match the BigQuery table.

https://cloud.google.com/bigquery/docs/loading-data#loading_encoded_data

👍 ↩ 🚩 upvoted 2 times

🗄 👤 **NicolasN** 10 months, 3 weeks ago

Exactly ↑

The updated link (Dec. 2022) and the quote:

<https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv#encoding>

"If you don't specify an encoding, or if you specify UTF-8 encoding when the CSV file is not UTF-8 encoded, BigQuery attempts to convert the data to UTF-8. Generally, your data will be loaded successfully, but it may not match byte-for-byte what you expect."

👍 ↩ 🚩 upvoted 1 times

🗄 👤 **anji007** 2 years ago

Ans: C

👍 ↩ 🚩 upvoted 3 times

🗄 👤 **sumanshu** 2 years, 4 months ago

Vote for 'C'

👍 ↩ 🚩 upvoted 3 times

🗄 👤 **sumanshu** 2 years, 3 months ago

A is not correct because if another data format other than CSV was selected then the data would not import successfully.

B is not correct because the data was fully imported meaning no rows were skipped.

C is correct because this is the only situation that would cause successful import.

D is not correct because whether the data has been previously transformed will not affect whether the source file will match the BigQuery table.

👍 ↩ 🚩 upvoted 2 times

🗄 👤 **naga** 2 years, 9 months ago

Correct C

👍 ↩ 🚩 upvoted 2 times

🗄 👤 **haroldbenites** 3 years, 2 months ago

C is correct

👍 ↩ 🚩 upvoted 3 times

🗄 👤 **saarabh1805** 3 years, 2 months ago

C is correct answer, Refer below link for more informaiton.

https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv#details_of_loading_csv_data

👍 ↩ 🚩 upvoted 6 times

🗄 👤 **[Removed]** 3 years, 7 months ago

Answer: C

👍 ↩ 🚩 upvoted 10 times



Platform

> Home

> All Exams

> Examtopics PRO

> Training Courses



