

# Dataflow reads JSON files in Storage but randomly writes 0 records

Last updated by | Nelson Gomes | Mar 9, 2023 at 11:00 PM PST

---

## Contents

- [Description](#)
- [Root Cause](#)
- [Solution](#)
- [Some options to consider](#)

## Description

When customer uses a wild card path e.g. (concat(\$Folder,'/',\$tableName,'/WorkplanTask\*')) dataflow reads the data from the files but fails to write to sink and only results to empty 0kb files. Issue occurs randomly.

```
AdmsClient.transformedEvent for JobId: 44429283-f479-4d88-9e31-9cca56a1f7c0, rowCount: 0 metric:
{ "stage": 34,
"target": "ActivityCache",
"time": 4880,
"bytesRead": 759720997,
"bytesWritten": 0,
"recordsRead": 100,
"recordsWritten": 0,
},
```

## Root Cause

Customer may have empty JSON files in their defined wildcard path. With wildcard path, Dataflow infer schema by randomly sampling the source files instead of the imported schema of the dataset. I.e., if empty files are randomly chosen for the sampling, then an empty schema will be used resulting to zero records written.

## Solution

Advise customer to filter out empty files before using the wildcard path as the best practice and will make their pipeline runs faster (requiring less compute time). Wild card path processes all files that meets the search condition.

## Some options to consider

1. Use get metadata activity prior to the dataflow to get list of the files with their size and write the non-empty files to a text file, then use the List of files option in dataflow instead of the wild card path to use

that list.

2. The get metadata could be used with a dataflow inside a foreach to run for each non-empty file.

### Additional information

- **IdM reference:**
- **Ava:** 2211180030001513
- **Author:** aliyuahmad
- **Reviewer:** negome
- **Keywords:** empty file, 0 records, zero records.

How good have you found this content?

