Google Cloud Platform
Nested, Repeated, and Nested Repeated Fields

BigQuery for Data Analysts

V1.2

Approximate timing: 30 minutes

Agenda

1 Nested Fields

2 Lab

3 Repeated Fields

4 Lab

5 Nested Repeated Fields

6 Quiz & Lab

Nested Fields

- BigQuery supports importing and exporting nested fields in JSON and AVRO files
- A nested record field adds a named substructure to a row of data
- Useful as a mechanism to organize related information

Example: JSON Nested Schema and Data

Nested schema

```
{"name": "name", "type": "string", "mode": "REQUIRED"},
   {"name": "book", "type": "RECORD", "fields":
      [
      {"name": "title", "type": "string"},
      {"name": "ISBN",  "type": "string"}
      ]
   }
```

Nested data

```
{"name": "randolph", "book": {"title": "The Beginning", "ISBN": "213423422" } }
{"name": "charles", "book": {"title": "Fortunate Few", "ISBN": "993032933" } }
{"name": "james", "book": {"title": "Homeward Bound", "ISBN": "884920039" } }
```

Notes:
There is no comma "," between the JSON data items. The delimiter is a
NEW_LINE character.

Querying Nested Fields

- BigQuery automatically flattens nested fields when querying
  - `SELECT * ...` results in columns `<record_name>_<nested_field_name>`
- To query a specific nested field:
  - `SELECT name, book.title FROM [dataset.table]`

Query results

```
+----------+----------------+------------+
|   name   |   book_title   | book_ISBN  |
+----------+----------------+------------+
| randolph | The Beginning  | 213423422  |
| charles  | Fortunate Few  | 993032933  |
| james    | Homeward Bound | 884920039  |
+----------+----------------+------------+
```

Notes:
There is no comma "," between the JSON data items. The delimiter is a
NEW_LINE character.

Agenda

1. Nested Fields
2. **Lab**
3. Repeated Fields
4. Lab
5. Nested Repeated Fields
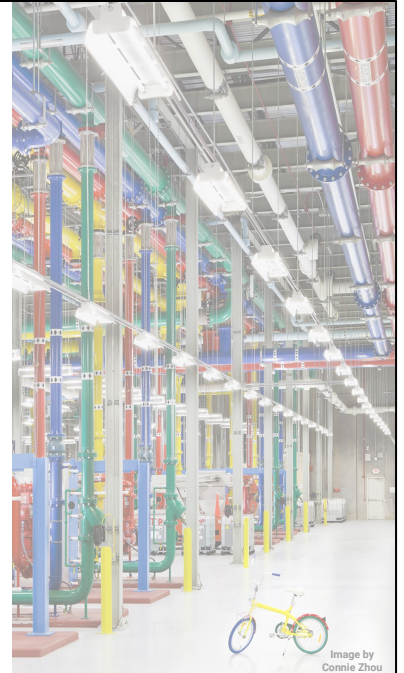6. Quiz & Lab

Lab, Part 1

Nested fields

Agenda

1. Nested Fields
2. Lab
3. Repeated Fields
4. Lab
5. Nested Repeated Fields
6. Quiz & Lab

Repeated Fields

- BigQuery supports importing and exporting repeated fields in JSON and AVRO files
- A repeated field adds an array of data inside a single field or record
- Useful as a mechanism to denormalize a foreign table

Example: JSON Repeated Schema and Data

Repeated schema

```
{"name": "name", "type": "string", "mode": "REQUIRED" },
    {"name": "city", "type": "string", "mode": "REPEATED" },
    {"name": "book", "type": "string", "mode": "REPEATED" }
```

Repeated data

```
{"name": "randy", "city": ["Tucson", "Houston", "Seattle"], "book":
["The Fudge"]}
{"name": "charlie", "city": ["Tucson", "Seattle", "Redmond"], "book":
[]}
{"name": "cynthia", "city": ["Houston", "Austin"          ], "book":
["The Fudge", "Outlaws"]}
```

Notes:
There is no comma "," between the JSON data items. The delimiter is a
NEW_LINE character.

Querying Repeated Fields

- A query such as `SELECT *` ... produces an error:
  - `Cannot output multiple independently repeated fields at the same time`
- Querying one repeated field, yields automatically flattened result
  - Example: `SELECT name, book FROM [dataset.table]`

<div>

Query results

```
+---------+-----------+
|  name   |   book    |
+---------+-----------+
| randy   | The Fudge |
| charlie | NULL      |
| cynthia | The Fudge |
| cynthia | Outlaws   |
+---------+-----------+
```

</div>

Notes:
This is a simple query showing how BigQuery automatically flattens one repeated field.

Querying Multiple Repeated Fields

- FLATTEN operator unrolls multiple repeated fields
  - One record for each of the values
- Example: `SELECT * FROM (FLATTEN ([dataset.table], city))`

```
+---------+-----------+-----------+        Query results
|  name   |   city    |   book    |
+---------+-----------+-----------+
| randy   | Maracaibo | The Fudge |
| randy   | Houston   | The Fudge |
| randy   | Seattle   | The Fudge |
| charlie | Maracaibo | NULL      |
| charlie | Seattle   | NULL      |
| charlie | Redmond   | NULL      |
+---------+-----------+-----------+
```

Agenda

1 Nested Fields

2 Lab

3 Repeated Fields

4 Lab

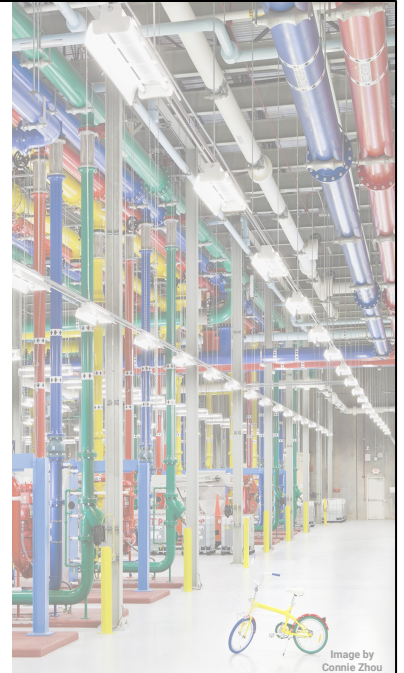5 Nested Repeated Fields

6 Quiz & Lab

Lab, Part 2

Repeated fields

Image by
Connie Zhou

Agenda

1 — Nested Fields

2 — Lab

3 — Repeated Fields

4 — Lab

5 — Nested Repeated Fields

6 — Quiz & Lab

Nested Repeated Fields

- BigQuery supports importing and exporting nested repeated fields in JSON and AVRO files
- Combine nested and repeated fields to denormalize a one-to-many relationship
- Useful as a mechanism to organize related information

Example: JSON Nested Repeated Schema

```
                                                    Nested repeated schema
{"name": "author", "type": "string", "mode": "REQUIRED"},
  {"name": "book", "type": "RECORD", "mode": "REPEATED", "fields":
    [
      {"name": "title", "type": "string", "mode": "REQUIRED"},
      {"name": "checked_out", "type": "timestamp", "mode": "REPEATED"}
    ]
  },
  {"name": "citiesLived", "type": "RECORD", "mode": "REPEATED", "fields":
    [
      {"name": "place", "type": "string"},
      {"name": "yearsLived", "type": "integer", "mode": "REPEATED"}
    ]
  }
```

Example: JSON Nested Repeated Data

Nested repeated data

{"author": "melville", "book": [{"title": "Moby Dick", "checked_out":
["2014-12-12 14:23", "2013-04-03 12:13"] }], "citiesLived": [{"place":
"Denver, CO", "yearsLived" : ["1986", "1987"] }] }
{"author": "hardy", "book": [{"title": "Return of the Native",
"checked_out": ["1984-05-30 12:12", "1986-03-12 00:00", "1992-05-03
04:32"] }, {"title": "The Mayor of Casterbridge", "checked_out":
["1983-06-23 12:12", "1986-03-12 00:00", "1992-05-03 04:32"] }],
"citiesLived": [{"place": "Austin, TX", "yearsLived" : ["1982", "1983",
"1984"] }, {"place": "Dublin, CA", "yearsLived" : ["1992", "1999", "2000"]
}] }
{"author": "koontz", "book": [{"title": "Velocity", "checked_out":
["1990-06-10 12:10", "2000-03-11 10:00", "1992-05-03 04:32"] }, {"title":
"Intensity", "checked_out": ["2003-06-23 02:12", "2004-03-12 20:00",
"1992-05-03 04:32"] }] }

Querying Nested Repeated Fields

- A query such as SELECT * ... produces an error:
  - `Cannot output multiple independently repeated fields at the same time`
- Example: `SELECT author, book.title, book.checked_out FROM [dataset.table]`

```
                                                        Query results
+----------+---------------------+------------------------+
|  author  |     book_title      |    book_checked_out    |
+----------+---------------------+------------------------+
| melville | Moby Dick           | 2014-12-12 14:23:00 UTC |
| melville | Moby Dick           | 2013-04-03 12:13:00 UTC |
| hardy    | Return of the Native | 1984-05-30 12:12:00 UTC |
| hardy    | Return of the Native | 1986-03-12 00:00:00 UTC |
| hardy    | Return of the Native | 1992-05-03 04:32:00 UTC |
+----------+---------------------+------------------------+
```
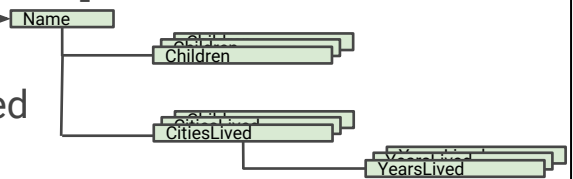
Using the WITHIN Keyword (1 of 2)

- WITHIN keyword works with aggregate functions
- Example:
  - ```
    SELECT fullName,
    COUNT(children.name) WITHIN
    RECORD FROM [dataset.tableId]
    ```
- WITHIN RECORD
  - Aggregates data in the repeated values within the record

Name

Children

CitiesLived

YearsLived

Using the WITHIN Keyword (2 of 2)

- WITHIN <node>
  - Aggregates data in the repeated values within a node
- Example:
  - ```
    SELECT fullName, count(citiesLived.place) WITHIN RECORD,
    citiesLived.place, count(citiesLived.yearsLived) WITHIN
    citiesLived FROM [dataset.tableId]
    ```

```
|- fullName: string (required)                                    Schema
  +- children: record (repeated)
       |  |- name: string
       |  |- age: integer
   +- citiesLived: record (repeated)
       |  |- place: string
       |  +- yearsLived: integer (repeated)
```

Agenda

1. Nested Fields
2. Lab
3. Repeated Fields
4. Lab
5. Nested Repeated Fields
6. Quiz & Lab

Module Review

Which of the following statements is true?
*(select **1** of the available options)*

- ❏ A nested field is useful to help de-normalize a foreign table
- ❏ If there are two or more repeated fields you must use a FLATTEN operator
- ❏ You can load nested field data using a CSV file
- ❏ Nested repeated data can be exported to a CSV file
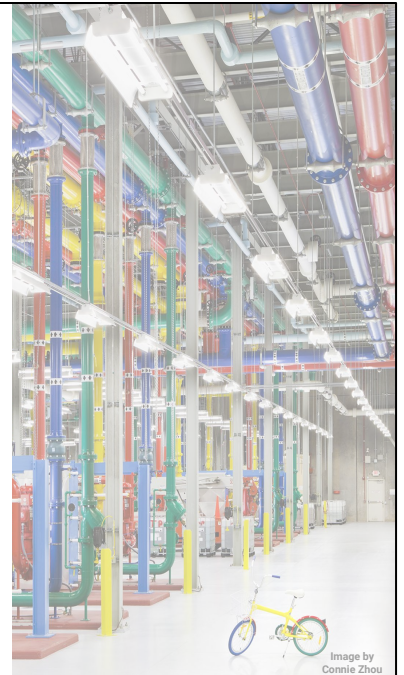
Lab, Part 3

Nested repeated fields

Resources

- Dealing with Data
  https://cloud.google.com/bigquery/docs/data
- Query Reference
  https://cloud.google.com/bigquery/query-reference
- Nested and Repeated Data
  https://cloud.google.com/bigquery/docs/data#nested

Module Review Answers

Which of the following statements is true?
*(select **1** of the available options)*

❏ A nested field is useful to help de-normalize a foreign table

✓ If there are two or more repeated fields you must use a FLATTEN operator

❏ You can load nested field data using a CSV file

❏ Nested repeated data can be exported to a CSV file

cloud.google.com