Table of Contents

Task 1. Schema Validation. Validate the provided JSON against a defined schema. The validation should allow for some flexibility (e.g., optional fields or additional fields not explicitly defined in the schema).

1. **Schema Validation**
   Validate the provided JSON against a defined schema. The validation should allow for some flexibility (e.g., optional fields or additional fields not explicitly defined in the schema).

Link to GitHub:

https://github.com/Ivasteel/tasks-leany-labs/blob/master/Task1_in_BigQuery.sql

```sql
-- Task 1. Schema Validation.
-- Validate the provided JSON against a defined schema. The validation should allow for some
flexibility (optional fields or additional fields not explicitly defined in the schema).

/*
BigQuery SQL does not have built-in JSON schema validation like some other database systems.
So we can't directly perform schema validation within BigQuery using a defined schema.
But, we can perform some basic checks/COALESCE to ensure the JSON structure is as expected.
*/

WITH
-- Initial JSON
 with_json_data_initial AS (
    SELECT '''
    {
        "message": "Blob with id 1332003122532507648 does not exist",
        "_links": {
            "self": {
                "href": "/1332003122532507648",
                "templated": false
            }
        }
    }
    ''' AS raw_json
 )
 -- Initial JSON but, templated field is null
,with_json_data_templated_null AS (
    SELECT '''
```

```sql
        {
            "message": "Blob with id 1332003122532507648 does not exist",
            "_links": {
                "self": {
                    "href": "/1332003122532507648",
                    "templated": null
                }
            }
        }
        ''' AS raw_json
    )
    -- Initial JSON, but templated field is missed at all
    ,with_json_data_templated_missed AS (
        SELECT '''
        {
            "message": "Blob with id 1332003122532507648 does not exist",
            "_links": {
                "self": {
                    "href": "/1332003122532507648"
                }
            }
        }
        ''' AS raw_json
    )
    -- Initial JSON, but templated field is missed at all, however, we need to have default value for
such cases (data quality)
    ,with_json_data_templated_missed_but_need_default_value AS (
        SELECT '''
        {
            "message": "Blob with id 1332003122532507648 does not exist",
            "_links": {
                "self": {
                    "href": "/1332003122532507648"
                }
            }
        }
        ''' AS raw_json
    )
SELECT
    'Initial JSON' as description,
    JSON_VALUE(raw_json, '$.message') AS message,
    JSON_VALUE(raw_json, '$._links.self.href') AS self_href,
    SAFE_CAST(JSON_VALUE(raw_json, '$._links.self.templated') AS BOOL) AS self_templated,
    CASE WHEN JSON_VALUE(raw_json, '$.message') IS NOT NULL THEN 'message field exists' ELSE 'message
field missing' END AS message_check,
    CASE WHEN JSON_VALUE(raw_json, '$._links.self.href') IS NOT NULL THEN 'self.href field exists'
ELSE 'self.href field missing' END AS self_href_check,
    CASE WHEN JSON_VALUE(raw_json, '$._links.self.templated') IS NOT NULL THEN 'self.templated field
exists' ELSE 'self.templated field missing' END AS self_templated_check
FROM with_json_data_initial

UNION ALL
```

```sql
SELECT
  'Initial JSON but templated is null' as description,
  JSON_VALUE(raw_json, '$.message') AS message,
  JSON_VALUE(raw_json, '$._links.self.href') AS self_href,
  SAFE_CAST(JSON_VALUE(raw_json, '$._links.self.templated') AS BOOL) AS self_templated,
  CASE WHEN JSON_VALUE(raw_json, '$.message') IS NOT NULL THEN 'message field exists' ELSE 'message
field missing' END AS message_check,
  CASE WHEN JSON_VALUE(raw_json, '$._links.self.href') IS NOT NULL THEN 'self.href field exists'
ELSE 'self.href field missing' END AS self_href_check,
  CASE WHEN JSON_VALUE(raw_json, '$._links.self.templated') IS NOT NULL THEN 'self.templated field
exists' ELSE 'self.templated field missing' END AS self_templated_check
FROM with_json_data_templated_null

UNION ALL

SELECT
  'Initial JSON but without templated field' as description,
  JSON_VALUE(raw_json, '$.message') AS message,
  JSON_VALUE(raw_json, '$._links.self.href') AS self_href,
  SAFE_CAST(JSON_VALUE(raw_json, '$._links.self.templated') AS BOOL) AS self_templated,
  CASE WHEN JSON_VALUE(raw_json, '$.message') IS NOT NULL THEN 'message field exists' ELSE 'message
field missing' END AS message_check,
  CASE WHEN JSON_VALUE(raw_json, '$._links.self.href') IS NOT NULL THEN 'self.href field exists'
ELSE 'self.href field missing' END AS self_href_check,
  CASE WHEN JSON_VALUE(raw_json, '$._links.self.templated') IS NOT NULL THEN 'self.templated field
exists' ELSE 'self.templated field missing' END AS self_templated_check
FROM with_json_data_templated_missed

UNION ALL

SELECT
  'Initial JSON but without templated field however we need to have default value instead of NULL'
as description,
  JSON_VALUE(raw_json, '$.message') AS message,
  JSON_VALUE(raw_json, '$._links.self.href') AS self_href,
  COALESCE(SAFE_CAST(JSON_VALUE(raw_json, '$._links.self.templated') AS BOOL), FALSE) AS
self_templated,
  CASE WHEN JSON_VALUE(raw_json, '$.message') IS NOT NULL THEN 'message field exists' ELSE 'message
field missing' END AS message_check,
  CASE WHEN JSON_VALUE(raw_json, '$._links.self.href') IS NOT NULL THEN 'self.href field exists'
ELSE 'self.href field missing' END AS self_href_check,
  CASE WHEN COALESCE(SAFE_CAST(JSON_VALUE(raw_json, '$._links.self.templated') AS BOOL), FALSE) IS
NOT NULL THEN 'self.templated field exists' ELSE 'self.templated field missing' END AS
self_templated_check

FROM with_json_data_templated_missed_but_need_default_value;
```

Query results:

Row	description	message	self_href ▾	self_templa	message_check ▾	self_href_check ▾	self_templated_check ▾
1	Initial JSON	Blob with id 133200312253250...	/1332003122532507648	false	message field exists	self.href field exists	self.templated field exists
2	Initial JSON but templated is null	Blob with id 133200312253250...	/1332003122532507648	null	message field exists	self.href field exists	self.templated field missing
3	Initial JSON but without templa...	Blob with id 133200312253250...	/1332003122532507648	null	message field exists	self.href field exists	self.templated field missing
4	Initial JSON but without templated field however we need to have default value instead of NULL	Blob with id 133200312253250...	/1332003122532507648	false	message field exists	self.href field exists	self.templated field exists

Evidence of script: