

Data Qualities Issues:

Script for data loading from json to tabular form.

```
/*-----*/
Users json data dump into tabular format using below SQL SERVER script.
-----*/
DECLARE @JSON VARCHAR(MAX)
SELECT @JSON = BulkColumn
FROM OPENROWSET (BULK 'C:\Users\priya\Downloads\Fetch-Rewards\users.json', SINGLE_CLOB) AS i

DROP TABLE IF EXISTS #Users
CREATE TABLE #Users
(
    id          VARCHAR(MAX)
    ,active      BIT
    ,createdDate DATETIME
    ,lastLogin   DATETIME
    ,role        VARCHAR(MAX)
    ,signUpSource VARCHAR(10)
    ,state       VARCHAR(2)
)

INSERT INTO #Users
SELECT --DISTINCT
id
,active
,DATEADD(S, CONVERT(INT,LEFT(createdDate, 10)), '1970-01-01') AS createdDate
,DATEADD(S, CONVERT(INT,LEFT(lastLogin, 10)), '1970-01-01') AS lastLogin
,role
,signUpSource
,state
FROM OPENJSON (@JSON, '$.Users')
WITH (id          VARCHAR(MAX) '$._id.$oid'
      ,active      BIT          '$.active'

INSERT INTO #Users
SELECT --DISTINCT
id
,active
,DATEADD(S, CONVERT(INT,LEFT(createdDate, 10)), '1970-01-01') AS createdDate
,DATEADD(S, CONVERT(INT,LEFT(lastLogin, 10)), '1970-01-01') AS lastLogin
,role
,signUpSource
,state
FROM OPENJSON (@JSON, '$.Users')
WITH (id          VARCHAR(MAX) '$._id.$oid'
      ,active      BIT          '$.active'
      ,createdDate VARCHAR(100) '$.createdDate.$date'
      ,lastLogin   VARCHAR(100) '$.lastLogin.$date'
      ,role        VARCHAR(100) '$.role'
      ,signUpSource VARCHAR(100) '$.signUpSource'
      ,state       VARCHAR(100) '$.state'
)
```

1. Duplicates Records

After executing above script run below script and you will see the below output.

83 %

Results Messages

	iD	active	createdDate	lastLogin	role	signUpSource	state
1	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI
2	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI
3	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI
4	5ff1e1eacfcfc399c274ae6	1	2021-01-03 15:25:30.000	2021-01-03 15:25:30.000	consumer	Email	WI
5	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI
6	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI
7	5ff1e1e8cfcfc399c274ad9	1	2021-01-03 15:25:28.000	2021-01-03 15:25:28.000	consumer	Email	WI
8	5ff1e1b7cfcfc399c274a5a	1	2021-01-03 15:24:39.000	2021-01-03 15:24:39.000	consumer	Email	WI
9	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI
10	5ff1e1f1cfcfc399c274b0b	1	2021-01-03 15:25:37.000	2021-01-03 15:25:37.000	consumer	Email	WI
11	5ff1e194b6a9d73a3a9f1052	1	2021-01-03 15:24:04.000	2021-01-03 15:25:37.000	consumer	Email	WI

Query executed successfully. DESKTOP-3GBI49Q\MSSQLSERVER... DESKTOP-3GBI49Q\priya ... master 00:00:00 495 rows

- When you run below script you will see unique records. Another thing is Null vlaues in state,signUpSource and lastLogin. Null value is not good for signUpSource field which is used to identify sign up source of user. This will lead Data security issues.

```
/* if you see the output by running above query you will see around 495 records but there are lot of duplicate records you will find in this output.
So to over come from this you can use distinct to get unique records.
Another thing is Null vlaues in state,signUpSource and lastLogin.
Null value is not good for signUpSource field which is used to identify sign up source of user.
This will lead Data security issues.
There should be some default value based on business requirement to avoid Null values.
For unique records run below query*/
SELECT DISTINCT * FROM #Users
```

3 %

Results Messages

	iD	active	createdDate	lastLogin	role	signUpSource	state
1	54943462e4b07e684157a532	1	2014-12-19 14:21:22.000	2021-03-05 16:52:23.000	fetch-staff	NULL	NULL
2	55308179e4b0eabd8f99caa2	1	2015-04-17 03:43:53.000	2018-05-07 17:23:40.000	consumer	NULL	WI
3	5964eb07e4b03efd0c0f267b	1	2017-07-11 15:13:11.000	2021-03-04 19:07:49.000	fetch-staff	NULL	IL
4	59c124bae4b0299e55b0f330	1	2017-09-19 14:07:54.000	2021-02-08 16:42:58.000	fetch-staff	NULL	WI
5	5a43c08fe4b014fd6b6a0612	1	2017-12-27 15:47:27.000	2021-02-12 16:22:37.000	consumer	NULL	NULL
3	5e27526d0bdb6a138c32b556	1	2020-01-21 19:35:09.000	NULL	consumer	Google	WI
7	5f2068904928021530f8fc34	1	2020-07-28 18:04:00.000	2021-02-04 15:30:05.000	fetch-staff	Email	WI
3	5fa32b4d898c7a11a6bcebf3e	1	2020-11-04 22:29:33.000	2021-03-04 07:21:58.000	fetch-staff	Google	AL
3	5fa41775898c7a11a6bcebf3e	1	2020-11-05 15:17:09.000	2021-03-04 16:02:02.000	fetch-staff	Email	NULL
10	5fb0a078be5fc9775c1f3945	1	2020-11-15 03:28:56.000	NULL	consumer	Google	AL
11	5fbc35711d967d1222cbf3e	1	2020-11-23 22:19:29.000	2021-02-26 04:25:51.000	fetch-staff	Email	NULL