Semi-Structured Data

Data loading from AWS S3 to Snowflake through SnowPipe

Steps For Migration

1.Creating AWS S3 bucket

2.Creating Database , Schema & Table in SnowFlake

3.Creating File Format

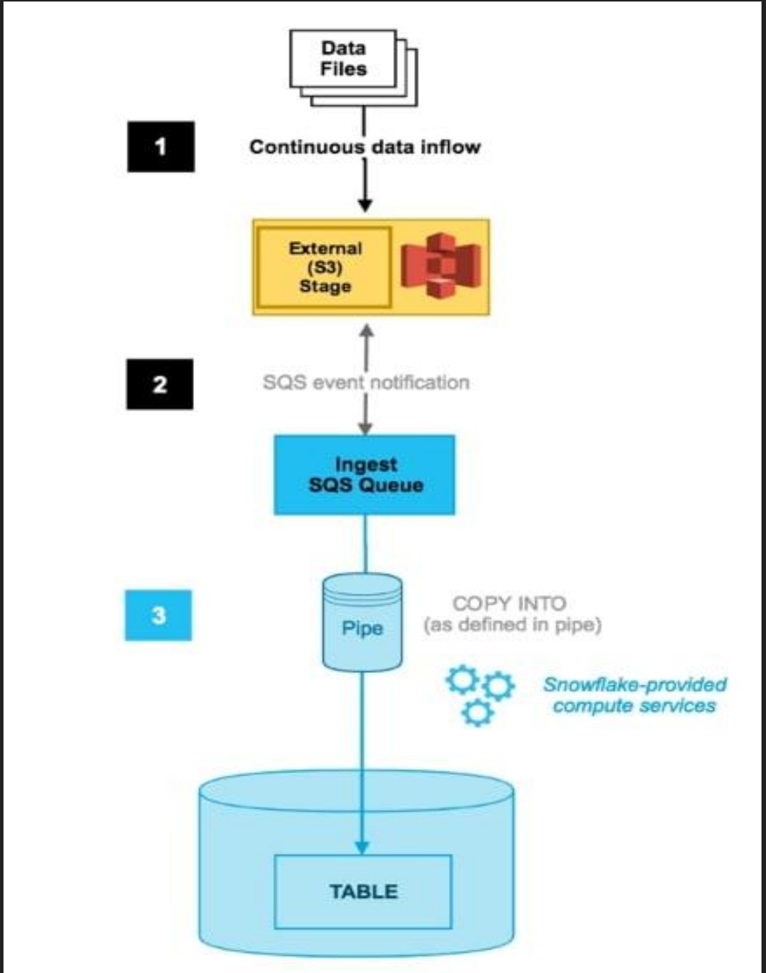
4.Creating Named Stage

5.Creating Pipe

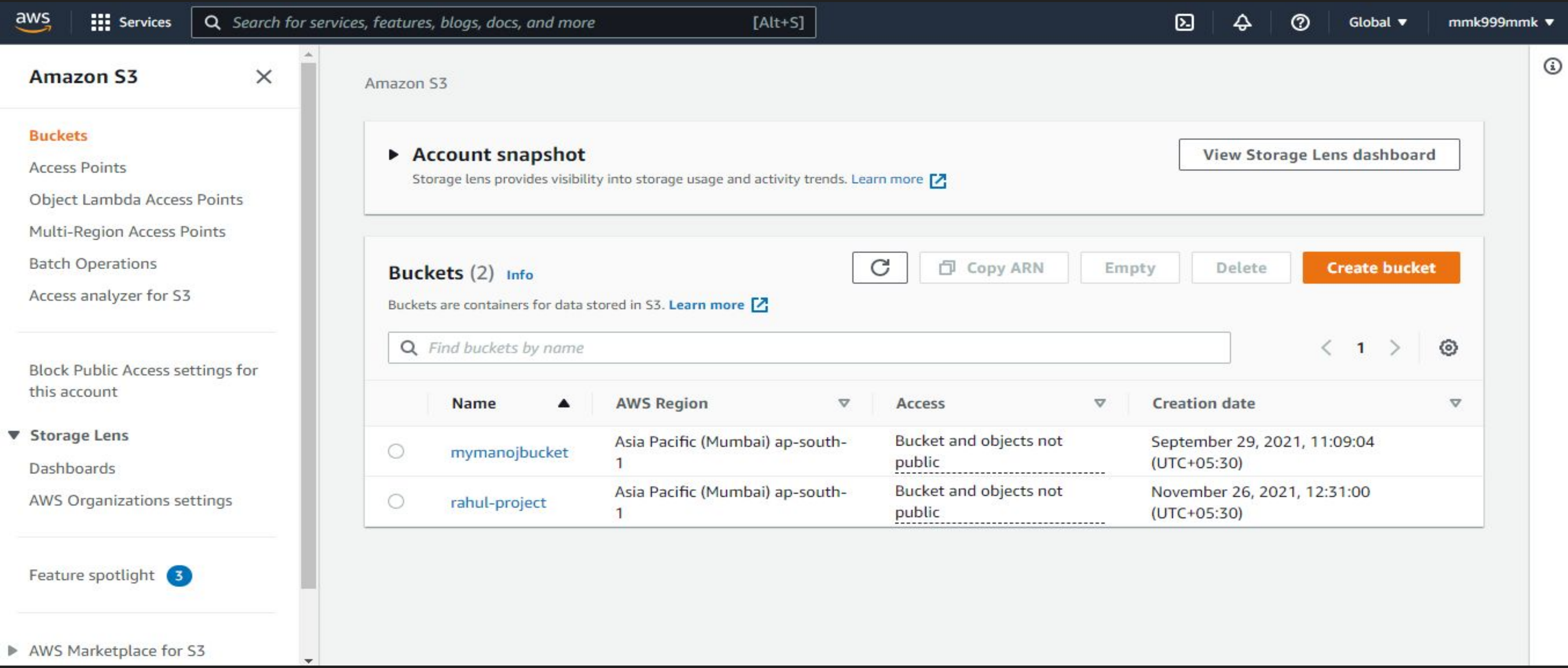
6.Setting notification in AWS

7. Converting loaded data into Relational Table

Architecture Diagram Of Snowpipe



Creating Bucket in AWS S3



aws
Services
Search for services, features, blogs, docs, and more [Alt+S]
Global ▼
mmk999mmk ▼
Amazon S3 ✕
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
▼ Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight 3
▶ AWS Marketplace for S3
Amazon S3

▶ **Account snapshot**

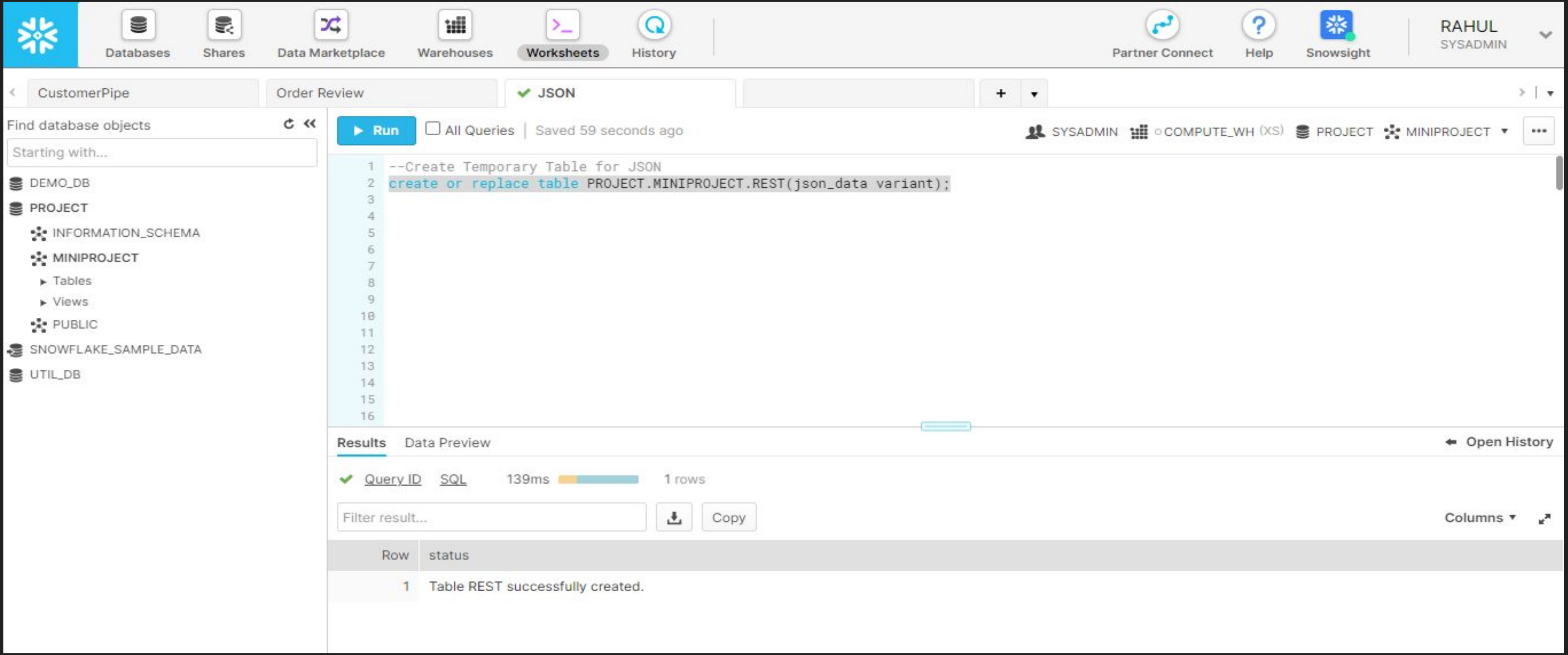
Storage lens provides visibility into storage usage and activity trends. Learn more

Buckets (2) Info

Buckets are containers for data stored in S3. Learn more

	Name ▲	AWS Region	Access	Creation date
○	mymanojbucket	Asia Pacific (Mumbai) ap-south-1	Bucket and objects not public	September 29, 2021, 11:09:04 (UTC+05:30)
○	rahul-project	Asia Pacific (Mumbai) ap-south-1	Bucket and objects not public	November 26, 2021, 12:31:00 (UTC+05:30)

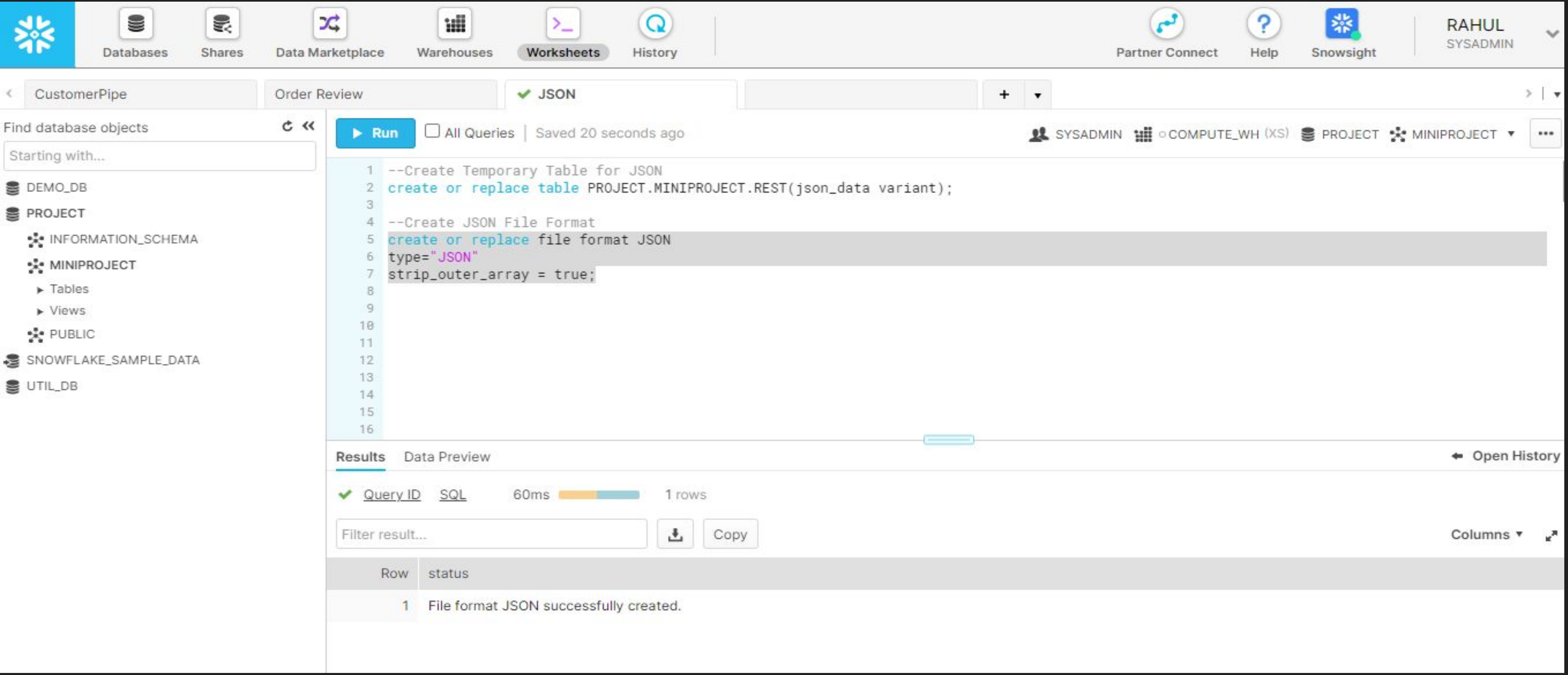
Create Database , Schema and Table



```
--Create Temporary Table for JSON
create or replace table PROJECT.MINIPROJECT.REST(json_data variant);
```

Row	status
1	Table REST successfully created.

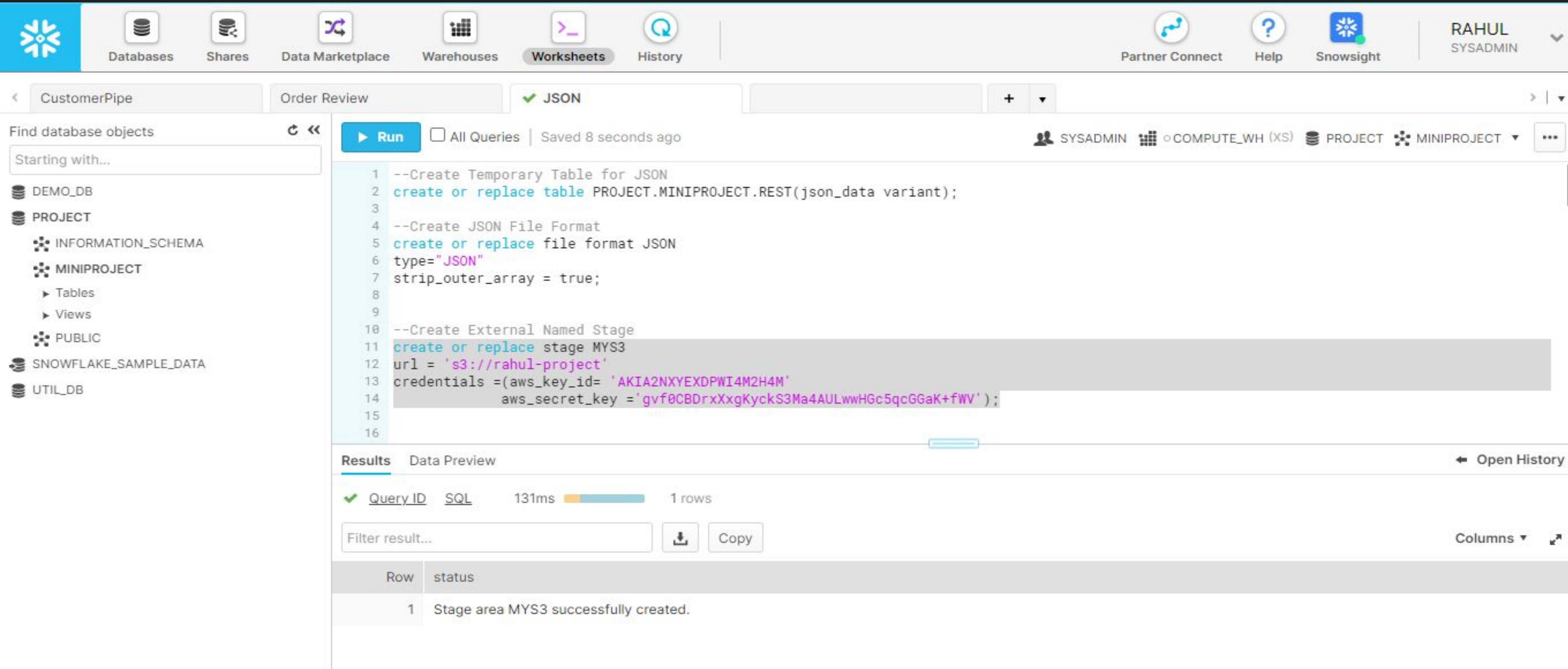
Create File Format



```
--Create Temporary Table for JSON
create or replace table PROJECT.MINIPROJECT.REST(json_data variant);

--Create JSON File Format
create or replace file format JSON
type="JSON"
strip_outer_array = true;
```

Row	status
1	File format JSON successfully created.

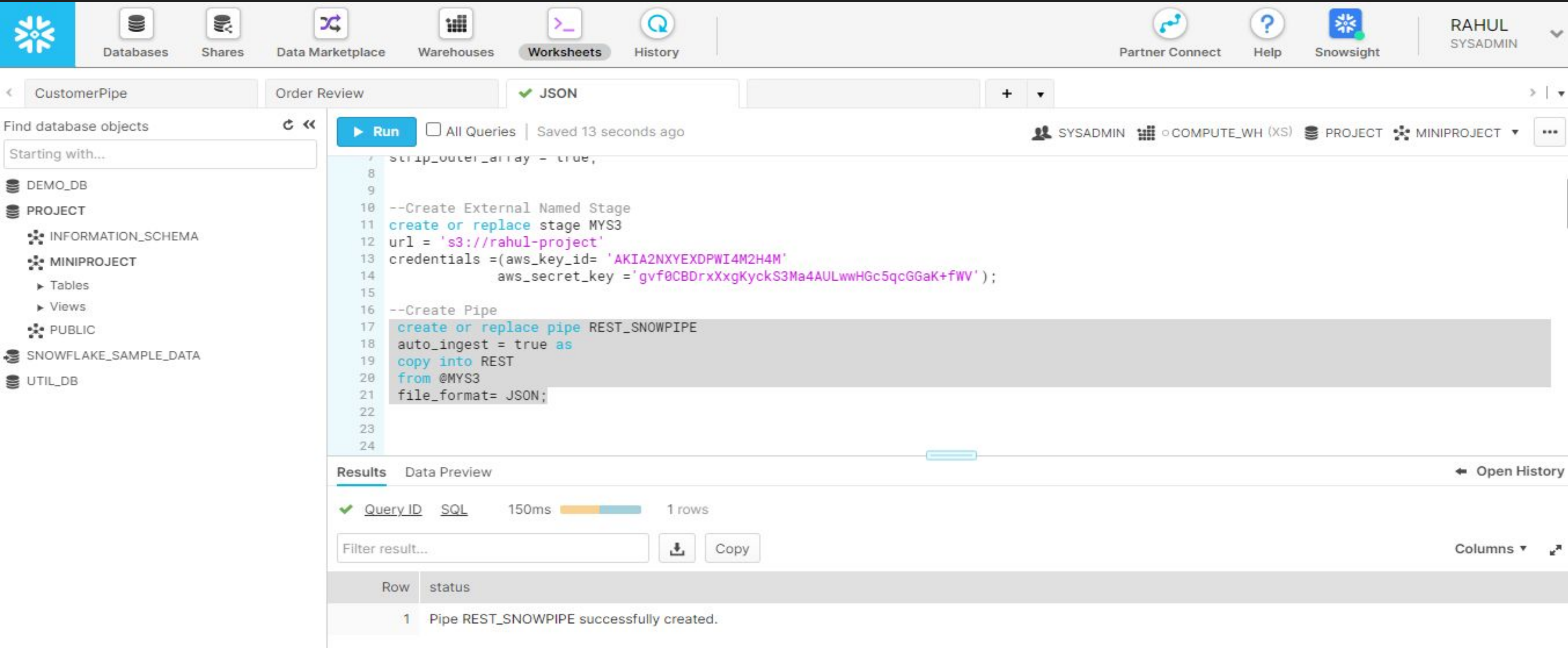
Create Named Stage



```
--Create Temporary Table for JSON
create or replace table PROJECT.MINIPROJECT.REST(json_data variant);

--Create JSON File Format
create or replace file format JSON
type="JSON"
strip_outer_array = true;


--Create External Named Stage
create or replace stage MYS3
url = 's3://rahul-project'
credentials =(aws_key_id= 'AKIA2NXYEXDPWI4M2H4M'
             aws_secret_key ='gvf0CBDrxXxgKyckS3Ma4AULwwHGc5qcGGaK+fWV');
```

Row	status
1	Stage area MYS3 successfully created.

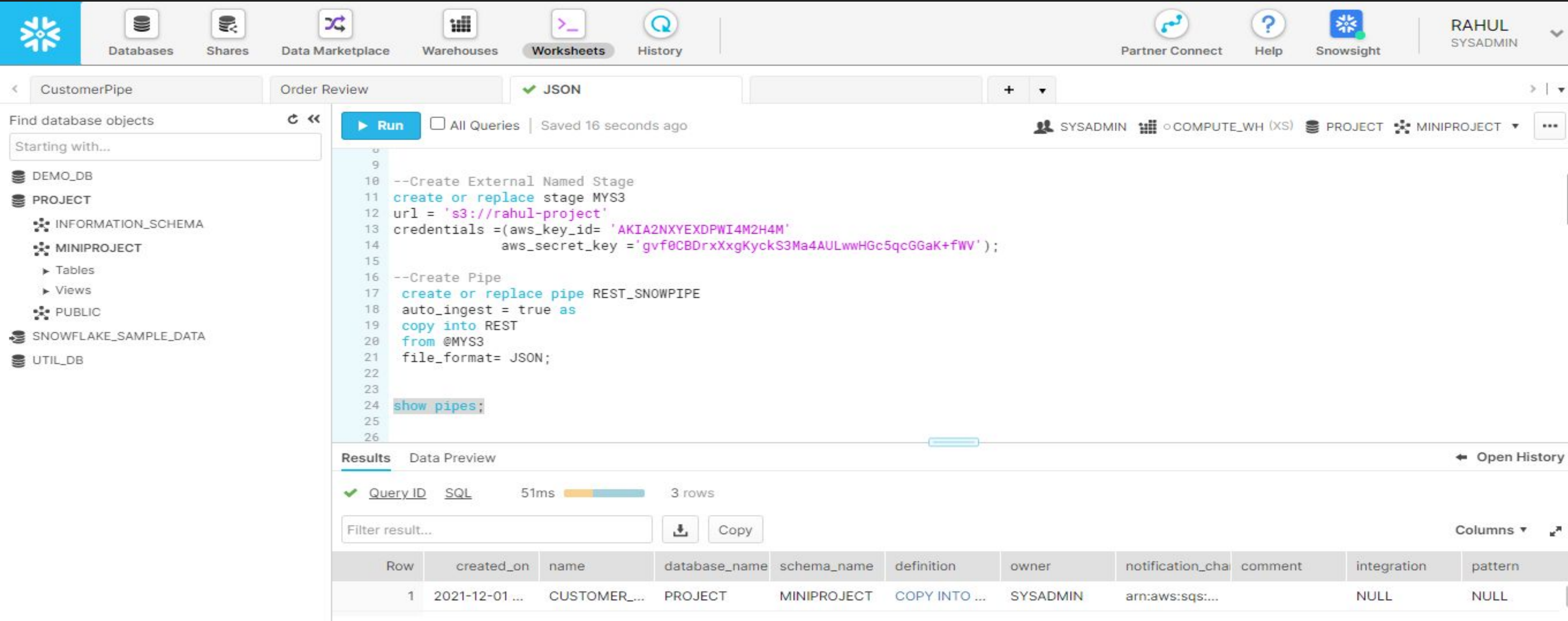
Create Pipe



```
strip_outer_array = true;


--Create External Named Stage
create or replace stage MYS3
url = 's3://rahul-project'
credentials =(aws_key_id= 'AKIA2NXYEXDPWI4M2H4M'
             aws_secret_key ='gvf0CBDrxXxgKyckS3Ma4AULwwHGc5qcGGaK+fWV');

--Create Pipe
 create or replace pipe REST_SNOWPIPE
 auto_ingest = true as
 copy into REST
 from @MYS3
 file_format= JSON;
```

Row	status
1	Pipe REST_SNOWPIPE successfully created.

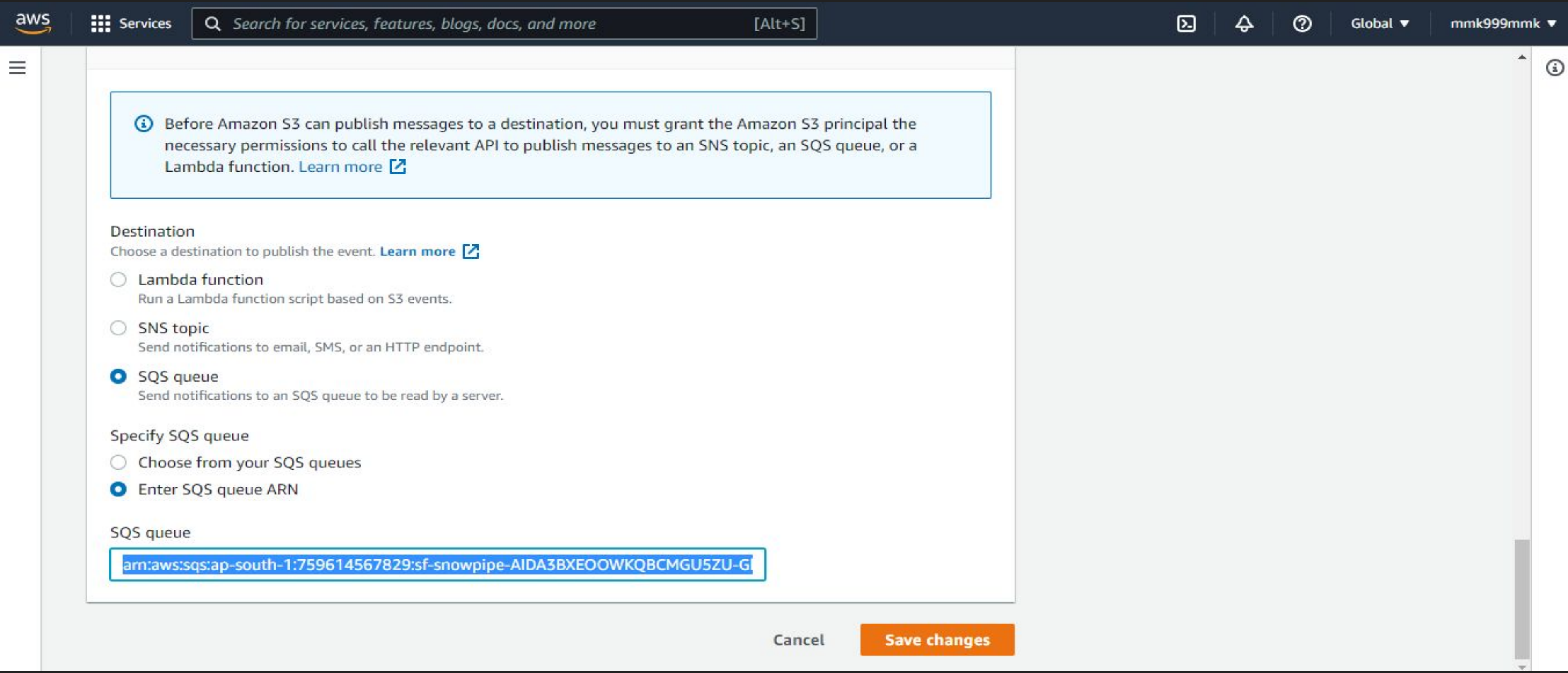
Use ‘show pipes’ command to see pipe and copy notification_channel



```
--Create External Named Stage
create or replace stage MYS3
url = 's3://rahul-project'
credentials =(aws_key_id= 'AKIA2NXYEXDPWI4M2H4M'
              aws_secret_key ='gvf0CBDrxXxgKyckS3Ma4AULwwHGc5qcGGaK+fWV');

--Create Pipe
 create or replace pipe REST_SNOWPIPE
 auto_ingest = true as
 copy into REST
 from @MYS3
 file_format= JSON;


show pipes;
```

Row	created_on	name	database_name	schema_name	definition	owner	notification_cha	comment	integration	pattern
1	2021-12-01 ...	CUSTOMER_...	PROJECT	MINIPROJECT	COPY INTO ...	SYSADMIN	arn:aws:sqs:...		NULL	NULL

Setting Notification



Before Amazon S3 can publish messages to a destination, you must grant the Amazon S3 principal the necessary permissions to call the relevant API to publish messages to an SNS topic, an SQS queue, or a Lambda function. Learn more

Destination

Choose a destination to publish the event. **Learn more**

- ○ Lambda function
  Run a Lambda function script based on S3 events.
- ○ SNS topic
  Send notifications to email, SMS, or an HTTP endpoint.
- ● SQS queue
  Send notifications to an SQS queue to be read by a server.

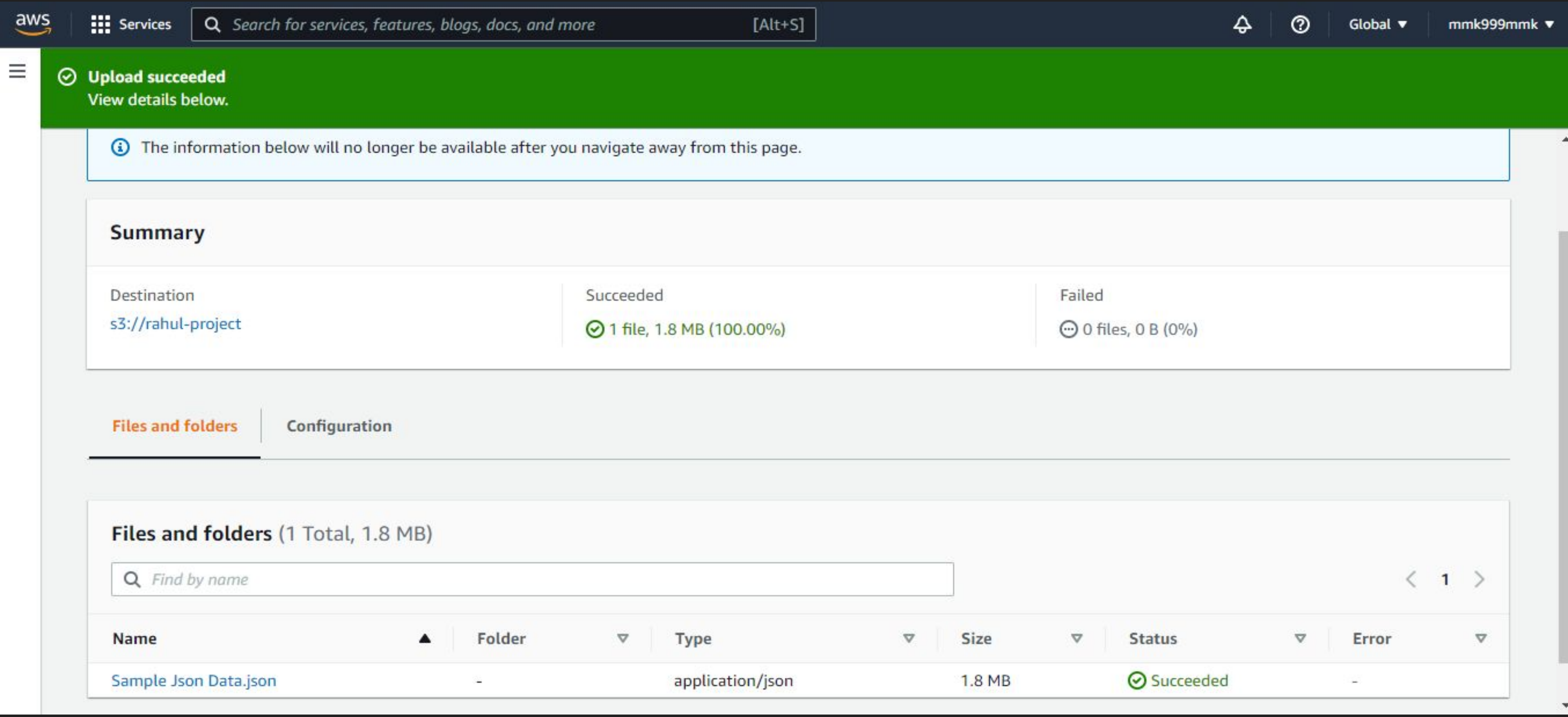
Specify SQS queue

- ○ Choose from your SQS queues
- ● Enter SQS queue ARN

SQS queue

arn:aws:sqs:ap-south-1:759614567829:sf-snowpipe-AIDA3BXEOOWKQBCMGU5ZU-G

Uploading Data



Upload succeeded
View details below.

The information below will no longer be available after you navigate away from this page.

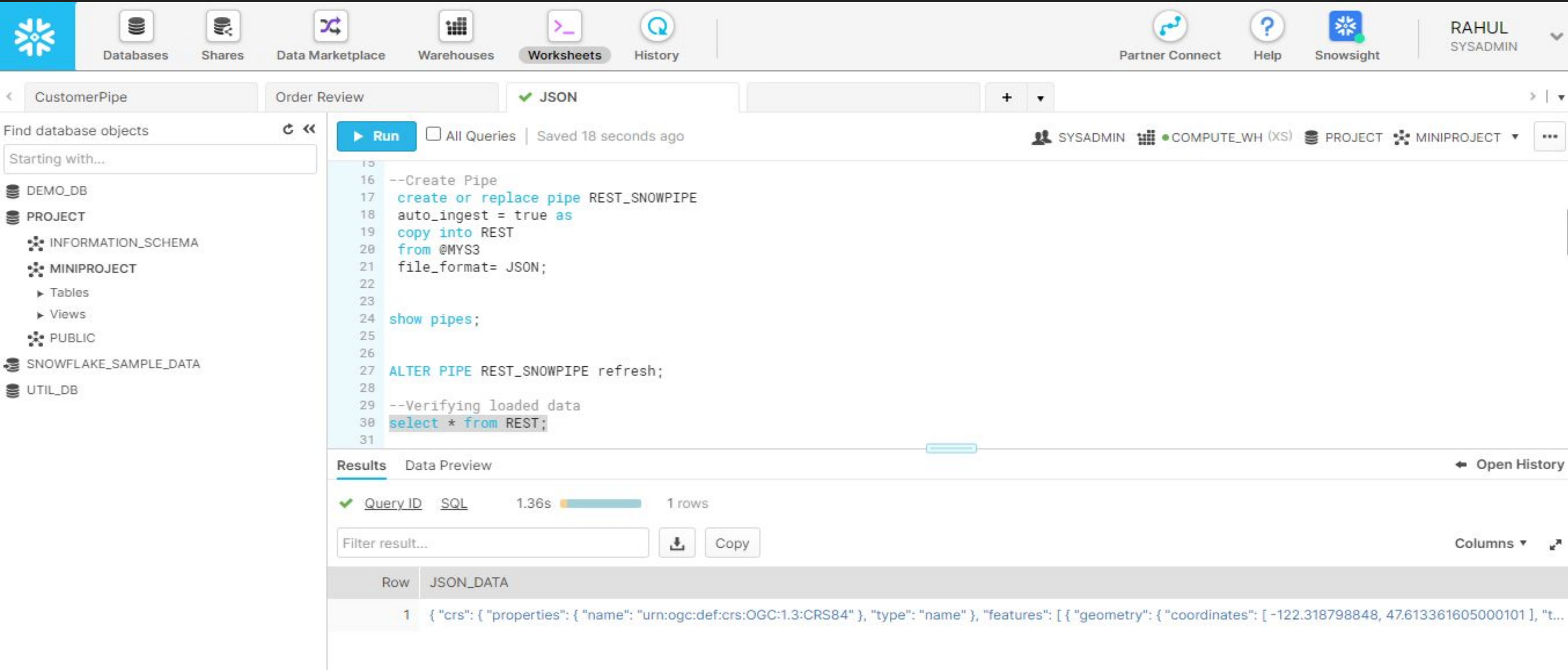
Summary

Destination	Succeeded	Failed
s3://rahul-project	✓ 1 file, 1.8 MB (100.00%)	0 files, 0 B (0%)

Files and folders (1 Total, 1.8 MB)

Name	Folder	Type	Size	Status	Error
Sample Json Data.json	-	application/json	1.8 MB	✓ Succeeded	-

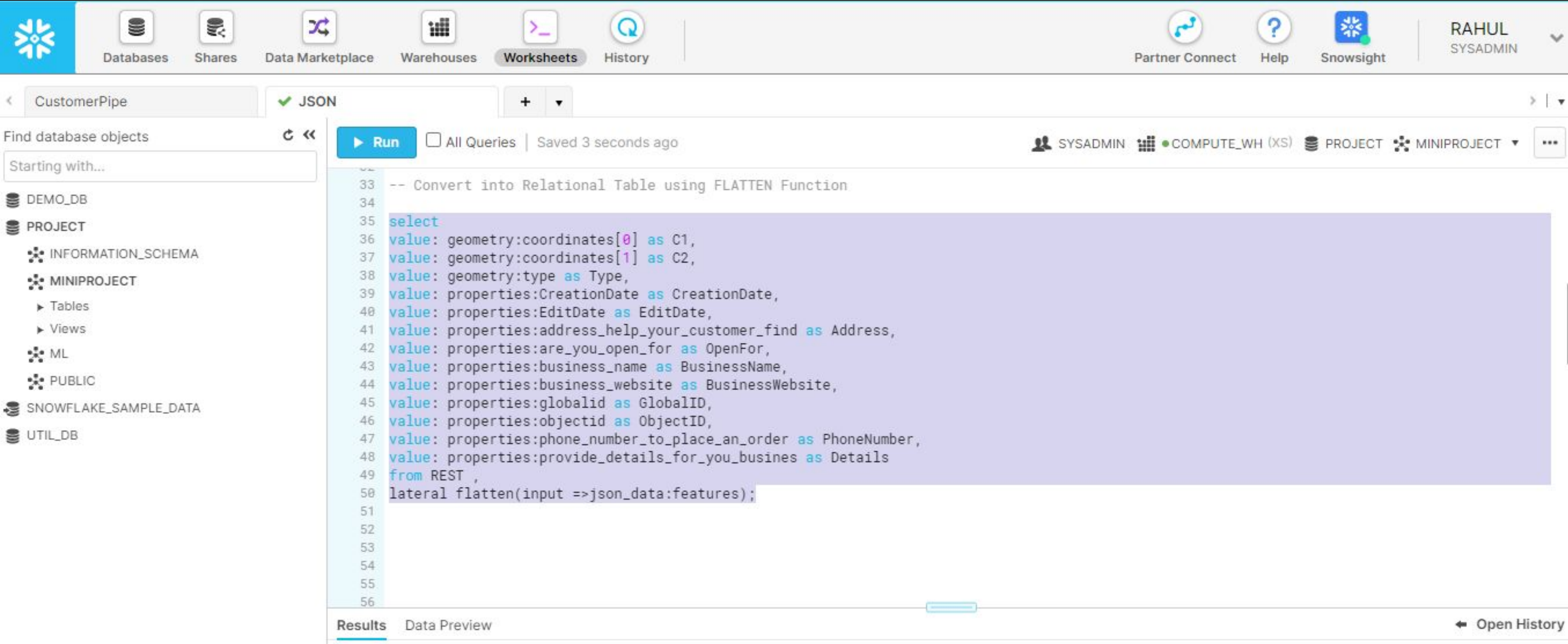
Verify loaded data



```
--Create Pipe
 create or replace pipe REST_SNOWPIPE
 auto_ingest = true as
 copy into REST
 from @MYS3
 file_format= JSON;


show pipes;


ALTER PIPE REST_SNOWPIPE refresh;

--Verifying loaded data
select * from REST;

```

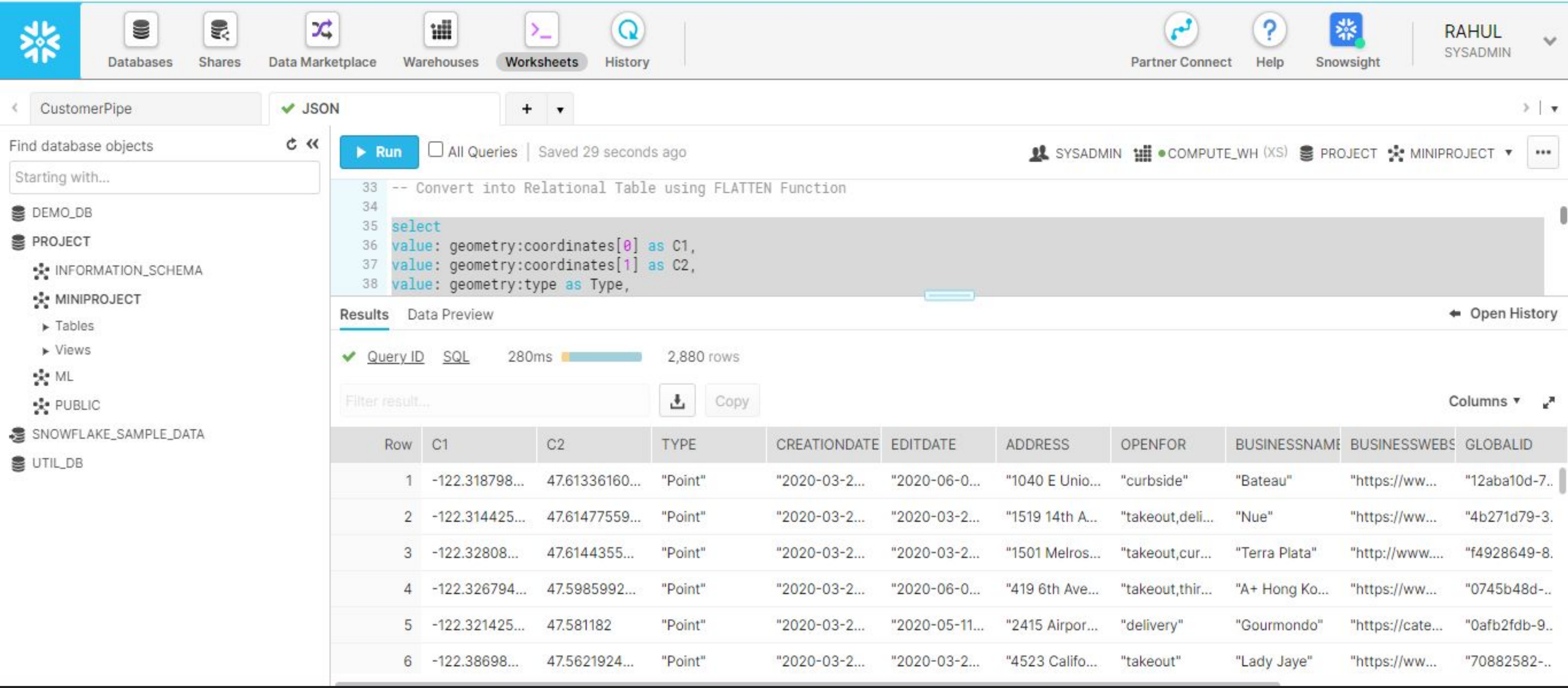
Row	JSON_DATA
1	{ "crs": { "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" }, "type": "name" }, "features": [{ "geometry": { "coordinates": [-122.318798848, 47.613361605000101], "t...

Transforming loaded data into Relational Table



```sql
-- Convert into Relational Table using FLATTEN Function

select
value: geometry:coordinates[0] as C1,
value: geometry:coordinates[1] as C2,
value: geometry:type as Type,
value: properties:CreationDate as CreationDate,
value: properties:EditDate as EditDate,
value: properties:address_help_your_customer_find as Address,
value: properties:are_you_open_for as OpenFor,
value: properties:business_name as BusinessName,
value: properties:business_website as BusinessWebsite,
value: properties:globalid as GlobalID,
value: properties:objectid as ObjectID,
value: properties:phone_number_to_place_an_order as PhoneNumber,
value: properties:provide_details_for_you_busines as Details
from REST ,
lateral flatten(input =>json_data:features);
```

Relational Table

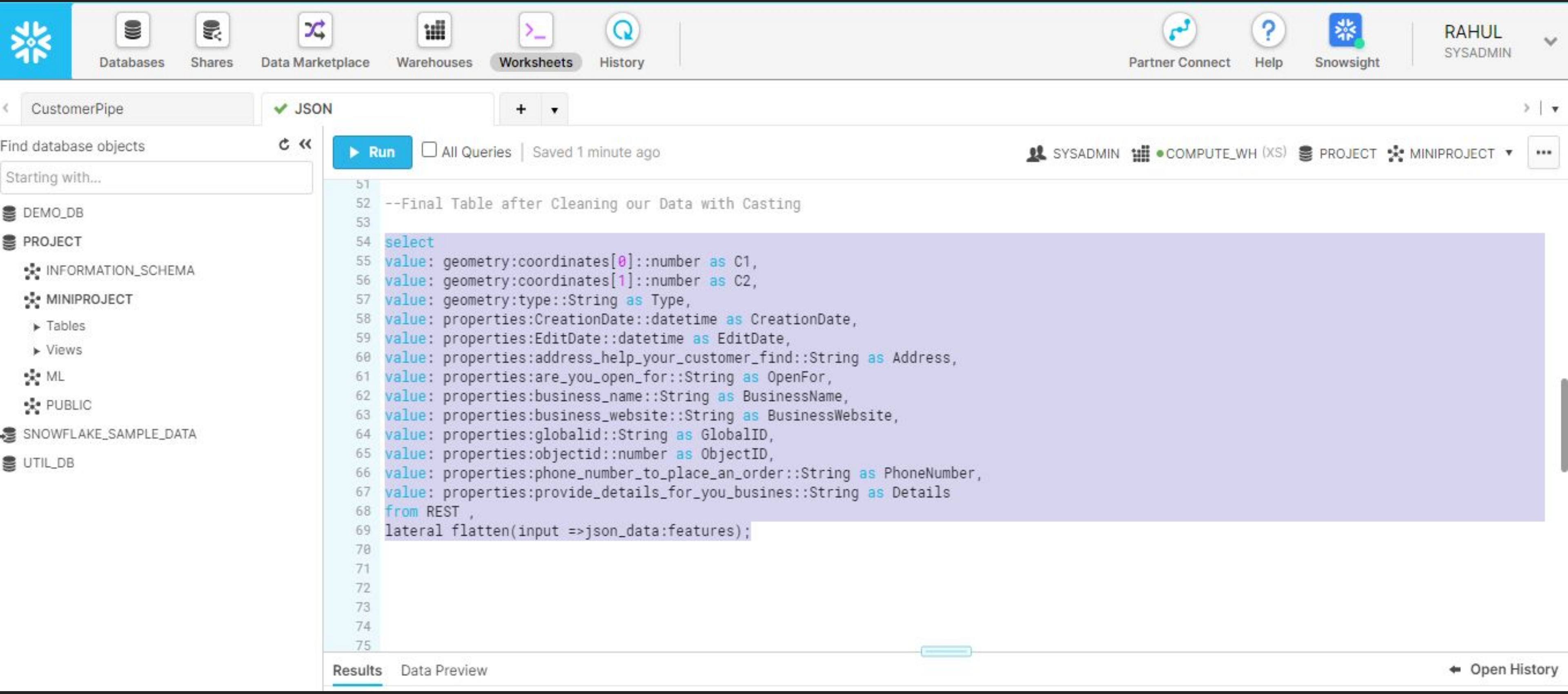


```
-- Convert into Relational Table using FLATTEN Function

select
value: geometry:coordinates[0] as C1,
value: geometry:coordinates[1] as C2,
value: geometry:type as Type,
```

280ms 2,880 rows

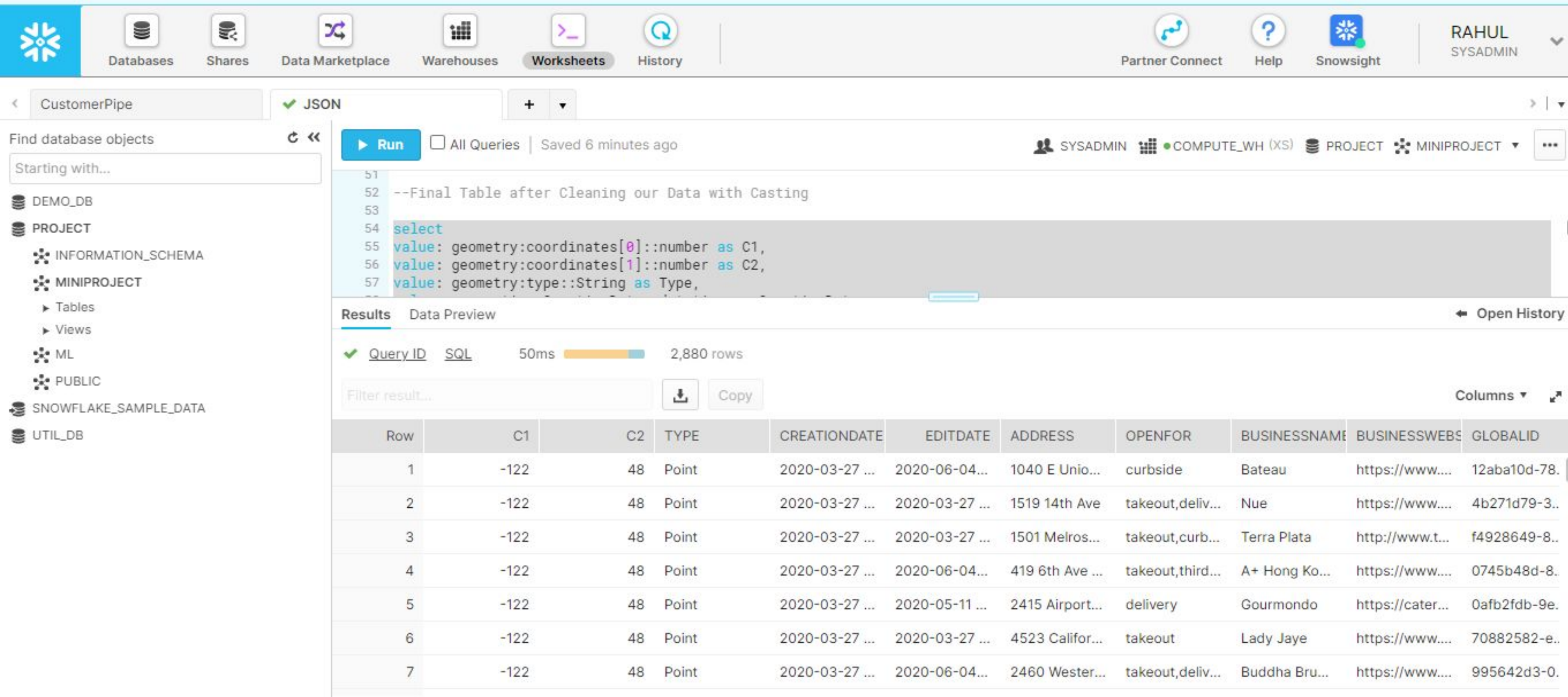
Row	C1	C2	TYPE	CREATIONDATE	EDITDATE	ADDRESS	OPENFOR	BUSINESSNAME	BUSINESSWEBS	GLOBALID
1	-122.318798...	47.61336160...	"Point"	"2020-03-2...	"2020-06-0...	"1040 E Unio...	"curbside"	"Bateau"	"https://ww...	"12aba10d-7..
2	-122.314425...	47.61477559...	"Point"	"2020-03-2...	"2020-03-2...	"1519 14th A...	"takeout,deli...	"Nue"	"https://ww...	"4b271d79-3.
3	-122.32808...	47.6144355...	"Point"	"2020-03-2...	"2020-03-2...	"1501 Melros...	"takeout,cur...	"Terra Plata"	"http://www....	"f4928649-8.
4	-122.326794...	47.5985992...	"Point"	"2020-03-2...	"2020-06-0...	"419 6th Ave...	"takeout,thir...	"A+ Hong Ko...	"https://ww...	"0745b48d-..
5	-122.321425...	47.581182	"Point"	"2020-03-2...	"2020-05-11...	"2415 Airpor...	"delivery"	"Gourmondo"	"https://cate...	"0afb2fdb-9..
6	-122.38698...	47.5621924...	"Point"	"2020-03-2...	"2020-03-2...	"4523 Califo...	"takeout"	"Lady Jaye"	"https://ww...	"70882582-..

Cleaning Transformation



```
--Final Table after Cleaning our Data with Casting

select
value: geometry:coordinates[0]::number as C1,
value: geometry:coordinates[1]::number as C2,
value: geometry:type::String as Type,
value: properties:CreationDate::datetime as CreationDate,
value: properties:EditDate::datetime as EditDate,
value: properties:address_help_your_customer_find::String as Address,
value: properties:are_you_open_for::String as OpenFor,
value: properties:business_name::String as BusinessName,
value: properties:business_website::String as BusinessWebsite,
value: properties:globalid::String as GlobalID,
value: properties:objectid::number as ObjectID,
value: properties:phone_number_to_place_an_order::String as PhoneNumber,
value: properties:provide_details_for_you_busines::String as Details
from REST ,
lateral flatten(input =>json_data:features);
```

Final Relational Table



SYSADMIN COMPUTE_WH (XS) PROJECT MINIPROJECT

```
--Final Table after Cleaning our Data with Casting

select
value: geometry:coordinates[0]::number as C1,
value: geometry:coordinates[1]::number as C2,
value: geometry:type::String as Type,
```

✓ 50ms 2,880 rows

Row	C1	C2	TYPE	CREATIONDATE	EDITDATE	ADDRESS	OPENFOR	BUSINESSNAME	BUSINESSWEBS	GLOBALID
1	-122	48	Point	2020-03-27 ...	2020-06-04...	1040 E Unio...	curbside	Bateau	https://www....	12aba10d-78.
2	-122	48	Point	2020-03-27 ...	2020-03-27 ...	1519 14th Ave	takeout,deliv...	Nue	https://www....	4b271d79-3..
3	-122	48	Point	2020-03-27 ...	2020-03-27 ...	1501 Melros...	takeout,curb...	Terra Plata	http://www.t...	f4928649-8..
4	-122	48	Point	2020-03-27 ...	2020-06-04...	419 6th Ave ...	takeout,third...	A+ Hong Ko...	https://www....	0745b48d-8.
5	-122	48	Point	2020-03-27 ...	2020-05-11 ...	2415 Airport...	delivery	Gourmondo	https://cater...	0afb2fdb-9e.
6	-122	48	Point	2020-03-27 ...	2020-03-27 ...	4523 Califor...	takeout	Lady Jaye	https://www....	70882582-e..
7	-122	48	Point	2020-03-27 ...	2020-06-04...	2460 Wester...	takeout,deliv...	Buddha Bru...	https://www....	995642d3-0.