

# Data Streaming

Data Streaming is a platform where data is processed continuously generated by number of sources. There are so many platforms which provide data streaming but MemSQL(Free Tier) has wide variety of sources to stream data into MemSQL(Free Tier) database.

As we know MemSQL(Free Tier) has gained good popularity in today's real-time analytics market. Its data streaming feature allow us to stream data from number of sources. These are:

- ✓ AWS S3
- ✓ Kafka
- ✓ Filesystem
- ✓ HDFS

In this slide show we will only show you the demo of data streaming through MemSQL(Free Tier)(Community) Pipeline feature from AWS S3 Source. AWS S3 is a storage platform offered by Amazon Web Services that provides object storage through a web.

After this slide show it is obvious you will have queries in your mind. Please feel free to write us and we will share our experience as much as possible.

# Data Streaming through MemSQL(Free Tier) Pipeline (Source AWS S3)

To stream data into MemSQL(Free Tier) from AWS S3 Source you will have to consider some pre-requisite and follow some steps.

## Pre-requisite

1. AWS Account with IAM role
2. MemSQL(Free Tier)(Community) installed
3. Linux (RedHat/CentOS 7)

## Steps:

1. Install S3 Cli on your machine to produce data to process into MemSQL(Free Tier) database through Pipeline

✓ `curl "https://d1vvhvl2y92vvt.cloudfront.net/awscli-exe-linux-x86_64.zip" -o "awscliv2.zip"`

✓ `unzip awscliv2.zip`

✓ `sudo ./aws/install`

✓ How to Check: type `aws` <ENTER>

It should not display command not found

# Data Streaming through MemSQL(Free Tier) Pipeline (Source AWS S3)

## Steps:

2. Write a bash shell that can produce data and send to S3 Storage

```
#!/bin/bash
rm /root/mem/*.csv
dt=`date +"%Y-%m-%d %H:%M:%S"`
dt1=`date +"%Y-%m-%d-%H-%M-%S"`
for i in $(seq 10);
do
echo "1,TOSY,100,^$dt^" | awk -v v=$i 'BEGIN {FS=OFS=","} {$1=v$1;$3=v$3}{print}' >> /root/mem/csvfile_`dt1`.csv 2>&1
done
aws s3 cp /root/mem/*.csv s3://dataloadmem
```

3. Set this script in cron to produce file every min and send to S3.

```
* * * * * sh /root/mem/pipelinefile.sh
```

(Please disable cron setting after test otherwise it will kill S3 and MemSQL(Free Tier) memory)

# Data Streaming through MemSQL(Free Tier) Pipeline (Source AWS S3)

## Steps:

4. Create a database table to accept data from pipeline in MemSQL(Free Tier) using below query

```
CREATE TABLE `ts` (  
  `id` int(11) DEFAULT NULL,  
  `name` varchar(50) CHARACTER SET utf8 COLLATE utf8_general_ci DEFAULT NULL,  
  `amt` int(11) DEFAULT NULL,  
  `enroll_date` datetime DEFAULT NULL  
  /*!90618 , SHARD KEY () */  
)
```

5. Create a Pipeline in MemSQL(Free Tier) to establish data process from S3 to MemSQL(Free Tier) Database table (ts)

```
memSQL>create pipeline ts as load data s3 'dataloadmem' config '{"region" : "ap-south-1"}' credentials  
'{"aws_access_key_id": "xxxxxxxxxxxxxxxxxxxx", "aws_secret_access_key":  
"xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"}' into table ts fields terminated by ',' optionally enclosed by '^' lines  
terminated by '\n';
```

# Data Streaming through MemSQL(Free Tier) Pipeline (Source AWS S3)

## Steps:

6. Start the Pipeline to actually process the data

```
memsql> start pipeline ts;
```

7. Check the data load into table by MemSQL(Free Tier) S3 Pipeline Streaming

```
memsql > select count(*),enroll_date from ts group by enroll_date order by enroll_date limit 2;
```

```
+-----+-----+  
| count(*) | enroll_date      |  
+-----+-----+  
|      10 | 2019-12-06 13:31:14 |  
|      10 | 2019-12-06 13:33:58 |  
+-----+-----+
```

```
2 rows in set (0.00 sec)
```

# Data Streaming through MemSQL(Free Tier) Pipeline (Source AWS S3)

## Steps:

8. Sources to see Pipeline processing details, errors etc.

```
+-----+
| Tables_in_information_schema |
+-----+
| PIPELINES                    |
| PIPELINES_BATCHES           |
| PIPELINES_BATCHES_METADATA  |
| PIPELINES_BATCHES_SUMMARY   |
| PIPELINES_CURSORS           |
| PIPELINES_ERRORS            |
| PIPELINES_FILES             |
| PIPELINES_OFFSETS           |
+-----+
```