

# Data Engineering Project

## Employee Attendance Monitoring Dashboard on GCP

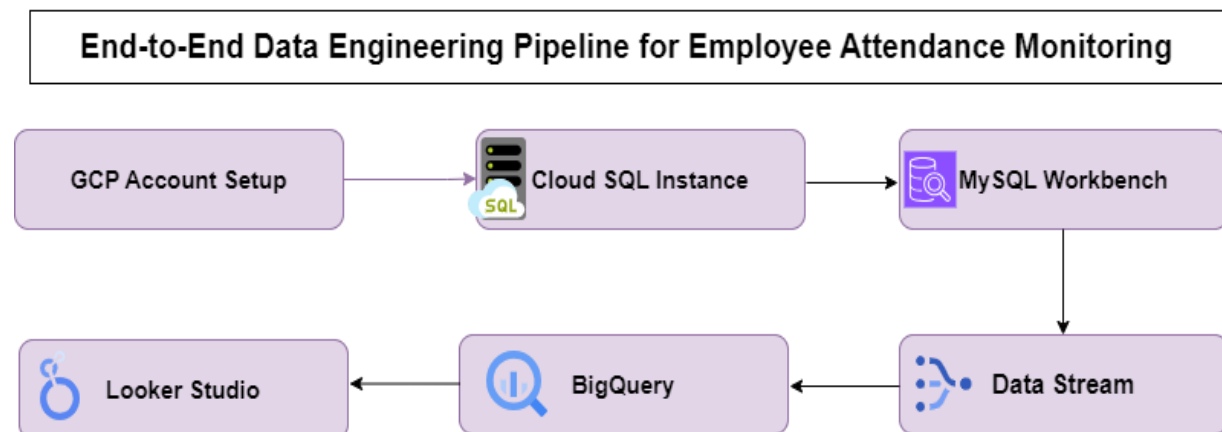
### Introduction:

The Employee Attendance Monitoring Dashboard on Google Cloud Platform (GCP) is a data engineering project designed to efficiently track, manage, and visualize employee attendance data. Leveraging the power of GCP services, the project aims to provide real-time insights and comprehensive reports on employee attendance patterns, ensuring better workforce management and operational efficiency.

### Technologies and Tools:

1. Google Cloud Storage (GCS): For storing raw attendance data files.
2. Big Query: For data warehousing and querying large datasets.
3. Datastream: It enables real-time data integration and streaming between sources and destinations.
4. Looker Studio: For creating interactive and insightful dashboards.
5. MySQL Workbench: For creating database and tables.

### Process (Step by Step):



### Step 1: Create a Google Cloud Platform (GCP) Account

Sign Up:

Visit the Google Cloud Platform website.

Click on the "Get started for free" button.

Follow the instructions to create your GCP account. You will need to provide credit card information, but you will have access to free-tier services (\$300 credits free).

Set Up Billing:

Once your account is created, set up a billing account. This is necessary to use most GCP services. Navigate to the Billing section in the GCP Console and follow the instructions.

## Step 2: Create a New Project in GCP

Access GCP Console: Go to the GCP Console.

Create a Project:

Click on the project drop-down menu at the top of the page.

Select "New Project".

Enter a project name (Attendance Monitoring System)

select an organization (if applicable), and click "Create".

Try Google Cloud for free

### Step 2 of 2 Payment Information Verification

Your payment information helps us reduce fraud and abuse. If using a credit or debit card, you won't be charged until you manually activate your full account.

Payments profile  
Vamsidhar Reddy  
Individual · India · IN · 8055-5564-0956 [Change](#) ⓘ

Your payment information is saved in a payments profile, which is associated with your Google Account and shared across Google services. [Learn more about payments profile](#)

Payment method  
 Mastercard \*\*\*\* 9814 [Change](#)

⚠ Reserve Bank of India requires that cards support automatic payments according to RBI regulations. If your card doesn't support automatic payments, you'll need to make manual payments or use a different card. We'll check your card in the next step. [Learn more about RBI regulation](#)

To make sure taxes for this service are processed correctly, follow the instructions in the email you'll receive from Google after you finish sign-up.

[START FREE](#)

[Privacy policy](#) | [Terms](#)

**Access to all Google Cloud products**

Get everything you need to build and run your apps, websites and services, including Firebase and the Google Maps API.

**\$300 credit for free**

Put Google Cloud to work with \$300 in credit to spend over the next 90 days.

**No autocharge after free trial ends**

We ask you for your credit card to make sure you are not a robot. If you use a credit or debit card, you won't be charged unless you manually activate your full account.

## Step 3: Create Cloud SQL instance

In the GCP Console, navigate to the SQL section.

Click on "Create instance".

Choose "MySQL" as the database engine. (MySQL, PostgreSQL, SQL server)

Configure the instance (instance ID, password, region, etc.) and click "Create".

Google Cloud | Attendance Monitoring System | Search (/) for resources, docs, products and more

SQL | Create an instance

### Choose your database engine

**MySQL**  
Versions: 8.0, 5.7, 5.6  
[Choose MySQL](#)

**PostgreSQL**  
Versions: 16, 15, 14, 13, 12, 11, 10, 9.6  
[Choose PostgreSQL](#)

**SQL Server**  
Versions: 2022, 2019, 2017  
[Choose SQL Server](#)

Want more context on the Cloud SQL database engines? [Learn more](#)

**Introducing AlloyDB for PostgreSQL**

PostgreSQL plus the best of the cloud: scale-out storage and compute, intelligent caching and AI/ML-powered management. Choose AlloyDB for 4x faster transactional workloads, more efficient analytical and vector queries and an industry-leading 99.99% SLA, inclusive of maintenance.

[CHOOSE ALLOYDB](#) [LEARN MORE](#)

Google Cloud | Attendance Monitoring System | Search (/) for resources, docs, products and more

### Create a MySQL Instance

#### Instance info

Instance ID \*  
mysql

Use lowercase letters, numbers and hyphens. Start with a letter.

Password \*  
root123 [GENERATE](#)

Set a password for the root user. [Learn more](#)

☐ No password

**PASSWORD POLICY**

Database version \*  
MySQL 8.0

[SHOW MINOR VERSIONS](#)

#### Choose a Cloud SQL edition

A Cloud SQL edition determines foundational characteristics of your instance. Choose the best option for your price and performance needs. [Learn more](#)

**Enterprise Plus**

- 99.99% availability SLA
- Sub-second planned maintenance downtime
- Near-zero downtime instance scale-up
- Performance-optimized machines
- Up to 35 days point-in-time recovery window
- Up to 3x higher read throughput with data cache
- Advanced disaster recovery with easy switchback

**Enterprise**

- 99.9% availability SLA
- Less than 60 seconds planned maintenance downtime
- General purpose machines
- Up to 7 days point-in-time recovery window

Choose a preset for this edition. Presets can be customized later as needed.

Production

[COMPARE EDITION PRESETS](#)

#### Choose region and zonal availability

For better performance, keep your data close to the services that need it. Region is permanent, while zone can be changed any time.

Region  
asia-south1 (Mumbai)

**Zonal availability**

☐ Single zone  
In case of outage, no failover. Not recommended for production.

☒ Multiple zones (highly available)  
Automatic failover to another zone within your selected region. Recommended for production instances. Increases cost.

[SPECIFY ZONES](#)

ⓘ Access improved disaster recovery capabilities. Create a cross-region replica and designate for disaster recovery to avoid regional outages on the replicas page. [Learn more](#)

#### Customise your instance

You can also customise instance configurations later.

[SHOW CONFIGURATION OPTIONS](#)

[CREATE INSTANCE](#) [CANCEL](#)

#### Pricing estimate

US\$2.84 per hour (estimated, without discounts)

That's about US\$68.16 per day.

Feature usage and traffic costs aren't included in estimate.

[SHOW COST BREAKDOWN](#)

#### Summary

Cloud SQL edition	Enterprise Plus
Region	asia-south1 (Mumbai)
DB version	MySQL 8.0
vCPUs	8 vCPU
RAM	64 GB
Data cache	Enabled (375 GB)
Storage	250 GB
Connections	Public IP
Backup	Automated
Availability	Multiple zones (highly available)
Point-in-time recovery	Enabled
Network throughput (Mbps)	2,000 of 2,000
Disk throughput (MB/s)	Read: 120.0 of 800.0 Write: 120.0 of 800.0
IOPS	Read: 7,500 of 15,000 Write: 7,500 of 15,000

## Step 4: Allow Network Connections

After the instance is created, go to the instance details.

Under the "Connections" tab, add your IP address to the authorized networks. This will allow you to connect to the instance from your local machine.

The screenshot displays the Google Cloud console interface for a MySQL instance. The left sidebar shows the navigation menu with 'Connections' selected. The main content area shows the 'Connections' tab for a MySQL instance. The 'Instance IP assignment' section has 'Public IP' selected. The 'Authorised networks' section shows a 'New network' form with 'Name: Local Workbench' and 'Network: 35.244.3.107'. The 'Google Cloud services authorisation' section shows 'Enable private path' is unchecked. The 'App Engine authorisation' section shows 'All apps in this project are authorised by default'. The 'SAVE' button is highlighted.

## Step 5: Connect to MySQL Workbench:

Open MySQL Workbench.

Create a new connection with the following details:

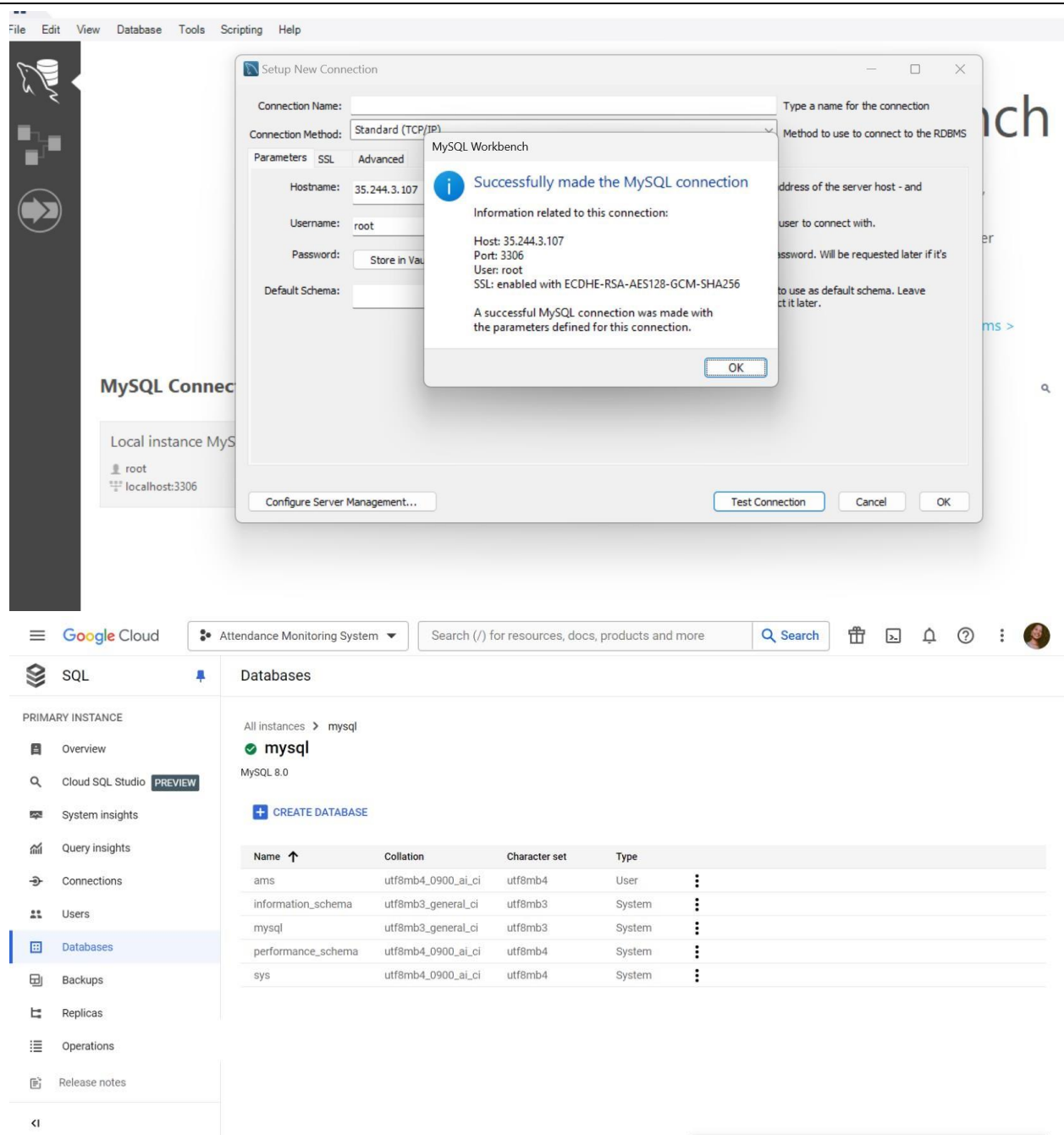
Connection Name: MySQL dB

Hostname: Public IP address of your Cloud SQL instance

Port: 3306

Username: root (or another user you have created)

Password: The password you set during instance creation(root123) Click "Test Connection" to ensure it works, then click "OK".



## Step 6: Export (Replicating) Cloud SQL data to Big Query in real-time by using Datastream

Datastream is a data replication service provided by Google Cloud Platform that allows you to stream real-time changes from your database to Big Query.

Datastream captures the changes (inserts, updates, and deletes) made to the source database and then sends these changes to Big Query in real time

Datastream supports a variety of databases, including MySQL, PostgreSQL and Oracle, and it can be configured to work with databases running on-premises or in the cloud. **Datastream API:**

Enable the Datastream API (If you are using First time) **Create a Datastream stream:**

Go to the Datastream page in the Google Cloud Console.

Click on "Create stream".

Provide a name and description for your stream.

Choose the source type as "My SQL".

Select your Cloud SQL instance and provide connection details (username, password, etc.).

Choose the destination type as "Big Query".

Configure any additional settings as needed (e.g., exclude certain tables, set up transformations).

Review and create the stream.

□ There are 6 steps each step mentioned in the screenshot(follow)

Google Cloud Attendance Monitoring System Search (/) for resources, docs

Create MySQL connection profile

← Create stream

- ✓ Get started  
mysql-bq, MySQL / BigQuery
- 2 Define and test source  
Not configured
- 3 Configure source  
Not configured
- 4 Define destination  
Not configured
- 5 Configure destination  
Not configured
- 6 Review and create  
Not configured

CANCEL

Define MySQL connection

Connection profiles represent the info you've already defined a connection profile. Choose it below. Otherwise, create one.

Source connection profile  
Create a connection profile

Define connection settings

The data location represented by the connection profile.

Connection profile name \*  
mysql-stream  
Must be fewer than 60 characters. 12/60

Connection profile ID \*  
mysql-stream  
Lowercase letters, numbers or hyphens. It must be unique in this project, and cannot be changed later. 12/60

Streams and their associated connection profiles must reside in the same region.

Region  
asia-south1 (Mumbai)  
Choice is permanent.

Connection details

Hostname or IP \*  
35.244.3.107

Port \*  
3306

Username \*  
root

Password \*  
root123

Labels

Labels can help you organise your Datastream resources. [Learn more](#)

+ ADD LABEL

CONTINUE

- Secure your connection to your source  
Choose an encryption type, and you'll see the SSL/TLS details needed.
- Define connectivity method  
The way that Datastream will connect to MySQL.
- Test connection profile  
Ensure that a connection to the source can be established.

CREATE AND CONTINUE BACK

Google Cloud Attendance Monitoring System Search (/) for resources, docs

Create stream

- 1 Get started mysql-bq, MySQL / BigQuery
- 2 Define and test source Not configured
- 3 Configure source Not configured
- 4 Define destination Not configured
- 5 Configure destination Not configured
- 6 Review and create Not configured

CANCEL

Define MySQL connection

Connection profiles represent the info you've already defined a connection profile. Choose it below. Otherwise, create one.

Source connection profile  
Create a connection profile

### Create MySQL connection profile

- ✓ Define connection settings mysql-stream, 35.244.3.107 : 3306
- ✓ Secure your connection to your source None
- Define connectivity method

The way that Datastream will connect to MySQL.

Based on your security and throughput requirements, as well as the location of your source, choose the most fitting connectivity method. [Learn more about the options and trade-offs between connectivity methods.](#)

Connectivity method \* IP allowlisting

IP allowlisting requires you to open your database and/or firewall to incoming connections from Datastream's public, regional IPs.

Copy the public IPs below, and use them to configure the network firewall for MySQL to accept connections from them. [Learn more](#)

asia-south1 (Mumbai) IPs [COPY ALL](#)

34.93.102.253	📋
34.93.10.146	📋
34.93.118.120	📋
35.200.143.183	📋
35.200.183.209	📋

[CONTINUE](#)

- Test connection profile

Ensure that a connection to the source can be established.

[CREATE AND CONTINUE](#) [BACK](#)

### Add these 5 IPs into the Cloud SQL (You need to add one by one)

Go to Cloud SQL

Go to Connections

Add one by one by naming as ds1, ds2, ds3, ds4, ds5(You can see in the below screen shot)



Google Cloud

Attendance Monitoring System

Search (/) for resources, docs, products and more

Search

SQL

PRIMARY INSTANCE

Overview

Cloud SQL Studio ...

System insights

Query insights

Connections

Users

Databases

Backups

Replicas

Release notes

Connections

All instances > mysql

mysql

MySQL 8.0

SUMMARY

NETWORKING

SECURITY

CONNECTIVITY TESTS

Choose how you want your source to connect to this instance, then define which networks are authorised to connect.

You can use the Cloud SQL Proxy for extra security with either option.

Instance IP assignment

Private IP

Assigns an internal, Google-hosted VPC IP address. Requires additional APIs and permissions. Can't be disabled once enabled.

Public IP

Assigns an external, Internet-accessible IP address. Requires using an authorised network or the Cloud SQL Proxy to connect to this instance.

Authorised networks

You can specify CIDR ranges to allow IP addresses in those ranges to access your instance.

Local workbench (27.59.155.43)

ds1 (34.93.102.253) (Not saved)

ds2 (34.93.10.146) (Not saved)

ds3 (34.93.118.120) (Not saved)

ds4 (35.200.143.183) (Not saved)

ds5 (35.200.183.209) (Not saved)

ADD A NETWORK

Google Cloud services authorisation

Enable private path

Allows other Google Cloud services like BigQuery to access data and make queries over Private IP.

App Engine authorisation

All apps in this project are authorised by default. You can use Cloud IAM to authorise apps in other projects.

SAVE

DISCARD CHANGES

Google Cloud

Attendance Monitoring System

Search (/) for resources, docs, products and more

Create stream

Get started

mysql-bq, MySQL / BigQuery

Define and test source

Not configured

Configure source

Not configured

Define destination

Not configured

Configure destination

Not configured

Review and create

Not configured

CANCEL

Define MySQL connection

Connection profiles represent the info you've already defined a connection profile choose it below. Otherwise, create one.

Source connection profile

Create a connection profile

Create MySQL connection profile

Define connection settings

mysql-stream, 35.244.3.107 : 3306

Secure your connection to your source

None

Define connectivity method

IP allowlisting

Test connection profile

Ensure that a connection to the source can be established.

Run test to test connectivity to the MySQL source.

RUN TEST

CREATE AND CONTINUE

BACK

7

Google Cloud

Attendance Monitoring System

Search (/) for resources, docs, products and more

Search

1

?

Create stream

Get started  
mysql-bq, MySQL / BigQuery

Define and test source  
mysql-stream, Test passed successfully

Configure source  
Not configured

Define destination  
Not configured

Configure destination  
Not configured

Review and create  
Not configured

CANCEL

Configure stream source

Define which set of database objects (schemas and tables) you'd like Datastream to stream or to exclude from streaming to the destination.

Select objects to include

1 schema

Objects to include \*

Specific schemas and tables

Search schemas and tables

SEARCH

Name	Selection summary
ams	All tables. Future tables on.
Future tables	
attendance_info	All columns SELECT COLUMNS
attendance_timing	All columns SELECT COLUMNS
dept_info	All columns SELECT COLUMNS
designation_info	All columns SELECT COLUMNS
employee_info	All columns SELECT COLUMNS
shift_info	All columns SELECT COLUMNS
information_schema	
mysql	
performance_schema	
sys	

SHOW ADVANCED OPTIONS

CONTINUE

BACK

Google Cloud

Attendance Monitoring System

Search (/) for resources, docs, products and more

Search

2

?

Create stream

Get started  
mysql-bq, MySQL / BigQuery

Define and test source  
mysql-stream, Test passed successfully

Configure source  
All tables from all schemas

Define destination  
bq-dest

Configure destination  
Not configured

Review and create  
Not configured

CANCEL

Configure the connection from Datastream to BigQuery

Specify how Datastream should stream into a BigQuery dataset

For each source schema, Datastream can create a new table within a new dataset or create all new tables within one dataset.

Dataset for each schema  
Let Datastream create a BigQuery dataset for each source schema.

Location type

Region  
Specify a region to colocate your datasets with other Google Cloud services.

Multi-region  
Allow BigQuery to select a region within a group to achieve higher quota limits.

Multi-region \*  
US (multiple regions in United States)

Dataset prefix

Use a prefix to ensure that the dataset is unique in BigQuery. The schema name will be appended to the prefix. Tip: Use a separator, such as an underscore (\_), to separate the prefix from the schema name.

Encryption

Google-managed encryption key  
Keys owned by Google

Cloud KMS key  
Keys owned by customers

Single dataset for all schemas  
For the dataset that you specify, Datastream creates all tables as [schema].[table].

Specify how data is written and processed in BigQuery

Choose Stream write mode

Select your data writing method to align with your stream's purposes. [Learn more](#)

Stream write mode \*  
Merge

Balance data staleness and BigQuery costs

Select a default [table staleness limit](#) to balance BigQuery query performance and cost versus data freshness. Depending on the data pattern and the source table's primary key, BigQuery might read the entire table every time changes are applied. You can use BigQuery reservations to allocate dedicated BigQuery compute resources for CDC row modification operations and set a cap on the cost of performing these operations. [Learn more](#)

Staleness limit  
0 seconds

CONTINUE

BACK

8





Google Cloud




Attendance Monitoring System

Search (/) for resources, docs, products and more

Search



2



Create stream

Get started

mysql-bq, MySQL / BigQuery

Define and test source

mysql-stream, Test passed successfully

Configure source

All tables from all schemas

Define destination

bq-dest

Configure destination

Configured

6 Review and create

Not configured

CANCEL

Review stream details and create

After verifying the stream details, create the stream and start it at a later time, or create and start immediately. Stream configuration will be tested at start.

Stream details

Stream name	mysql-bq
Region	asia-south1 (Mumbai)
Source/Destination	MySQL / BigQuery
Encryption	Google-managed

Source details

Connection profile name	mysql-stream
Connection details	35.244.3.107 : 3306
CDC method	Transaction logs
Objects to include	1 schema
Objects to exclude	None
Backfill mode	Automatic

Connection profile name	bq-dest
Destination data set	Dynamic, based on source schema
Connection profile name	bq-dest
Destination data set	Dynamic, based on source schema
Stream write mode	Merge

Validate stream (recommended)

Check your source and destination connectivity and end-to-end stream configuration to ensure that your stream will run successfully.

Validating stream...



CREATECREATE AND STARTBACK

Google Cloud




Attendance Monitoring System

Search (/) for resources, docs, products and more

Search



2



Create stream

Get started

mysql-bq, MySQL / BigQuery

Define and test source

mysql-stream, Test passed successfully

Configure source

All tables from all schemas

Define destination

bq-dest

Configure destination

Configured

6 Review and create

Not configured

CANCEL

Review stream details and create

After verifying the stream details, create the stream and start it at a later time, or create and start immediately. Stream configuration will be tested at start.

Stream details

Stream name	mysql-bq
Region	asia-south1 (Mumbai)
Source/Destination	MySQL / BigQuery
Encryption	Google-managed

Source details

Connection profile name	mysql-stream
Connection details	35.244.3.107 : 3306
CDC method	Transaction logs
Objects to include	1 schema
Objects to exclude	None
Backfill mode	Automatic

Connection profile name	bq-dest
Destination data set	Dynamic, based on source schema
Connection profile name	bq-dest
Destination data set	Dynamic, based on source schema
Stream write mode	Merge

Validate stream (recommended)

Check your source and destination connectivity and end-to-end stream configuration to ensure that your stream will run successfully.

Validation run complete

Connectivity to MySQL database: Validated

Binary logging is enabled: Validated

Binary log format configuration: Validated

Replication permissions: Validated

BigQuery API is enabled: Validated

BigQuery permissions: Validated

BigQuery dynamic data set location: Validated

REVALIDATE

CREATECREATE AND STARTBACK

Google Cloud Attendance Monitoring System Search (/) for resources, docs, products and more Search

**Datastream**

- Streams
- Connection profiles
- Private connectivity

Streams are comprised of configured source and destination connection profiles. They represent the flow of change data captured from source to destination.

Filter Filter streams

Stream name	Status	Details	Source/destination types	Source profile	Destination profile
mysql-bq	Running	Backfill in	MySQL / BigQuery	mysql-stream	bq-dest

Stream started

## Step 7: Big Query

Open Big Query in new tab there you can see your database and tables pushed from Cloud SQL.

Google Cloud Attendance Monitoring System Search (/) for resources, docs, products and more Search

**Explorer**

Viewing resources.

SHOW STARRED ONLY

- Data canvases
- External connections
- rp\_mysqlams
  - attendance\_info
  - attendance\_timing
  - dept\_info
  - designation\_info
  - employee\_info
  - shift\_info

SUMMARY

Nothing currently selected

Welcome to BigQuery Studio.

Create new

- SQL QUERY
- PYTHON NOTEBOOK
- DATA CANVAS

Add your own data

- Local file
  - Upload a local file
  - LAUNCH THIS GUIDE
- Google Drive
  - Google storage service
  - LAUNCH THIS GUIDE
- Google Cloud Storage
  - Google object storage service
  - LAUNCH THIS GUIDE

Show welcome page on startup

Job history

Google Cloud Attendance Monitoring System Search (/) for resources, docs, products and more Search

**Explorer**

Viewing resources.

SHOW STARRED ONLY

- Data canvases
- External connections
- rp\_mysqlams
  - attendance\_info
  - attendance\_timing
  - dept\_info
  - designation\_info
  - employee\_info
  - shift\_info

SUMMARY

Nothing currently selected

Untitled query

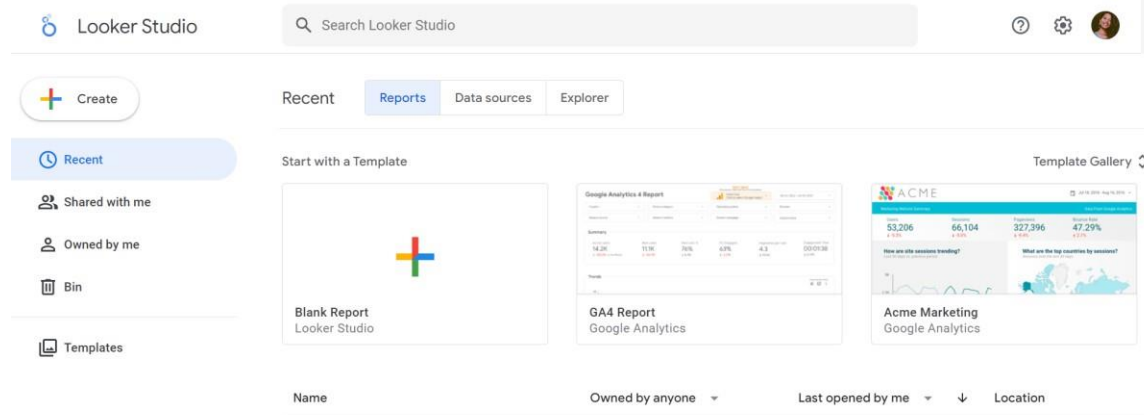
1 SELECT \* FROM 'rp\_mysqlams.employee\_info'

Query results

Row	emp_id	emp_name	dept_name	designation	datastream_metade
1	285	Mannu Pal	BOTAN SINGH WORKERS	HELPER	a734ffff1-4e63-4324
2	291	Ram Sukhmdra	BOTAN SINGH WORKERS	HELPER	a734ffff1-4e63-4324
3	127	Kaushal Sen	BOTAN SINGH WORKERS	HELPER	a734ffff1-4e63-4324
4	13	Ganour Ram	SLITTING	SLITTING INCHARGE	a734ffff1-4e63-4324
5	299	Sonu Pal	BOTAN SINGH WORKERS	HELPER	a734ffff1-4e63-4324
6	39	Sandeep yadav	TUBE MILL	COC OPERATOR	a734ffff1-4e63-4324
7	2009	Sharad Hanmant	TUBE MILL	PRODUCTION SUPERVISOR	a734ffff1-4e63-4324
8	319	Bhupendra	BOTAN SINGH WORKERS	HELPER	a734ffff1-4e63-4324

Results per page: 50 1 - 50 of 61

Job history



## Step 8: Looker Studio

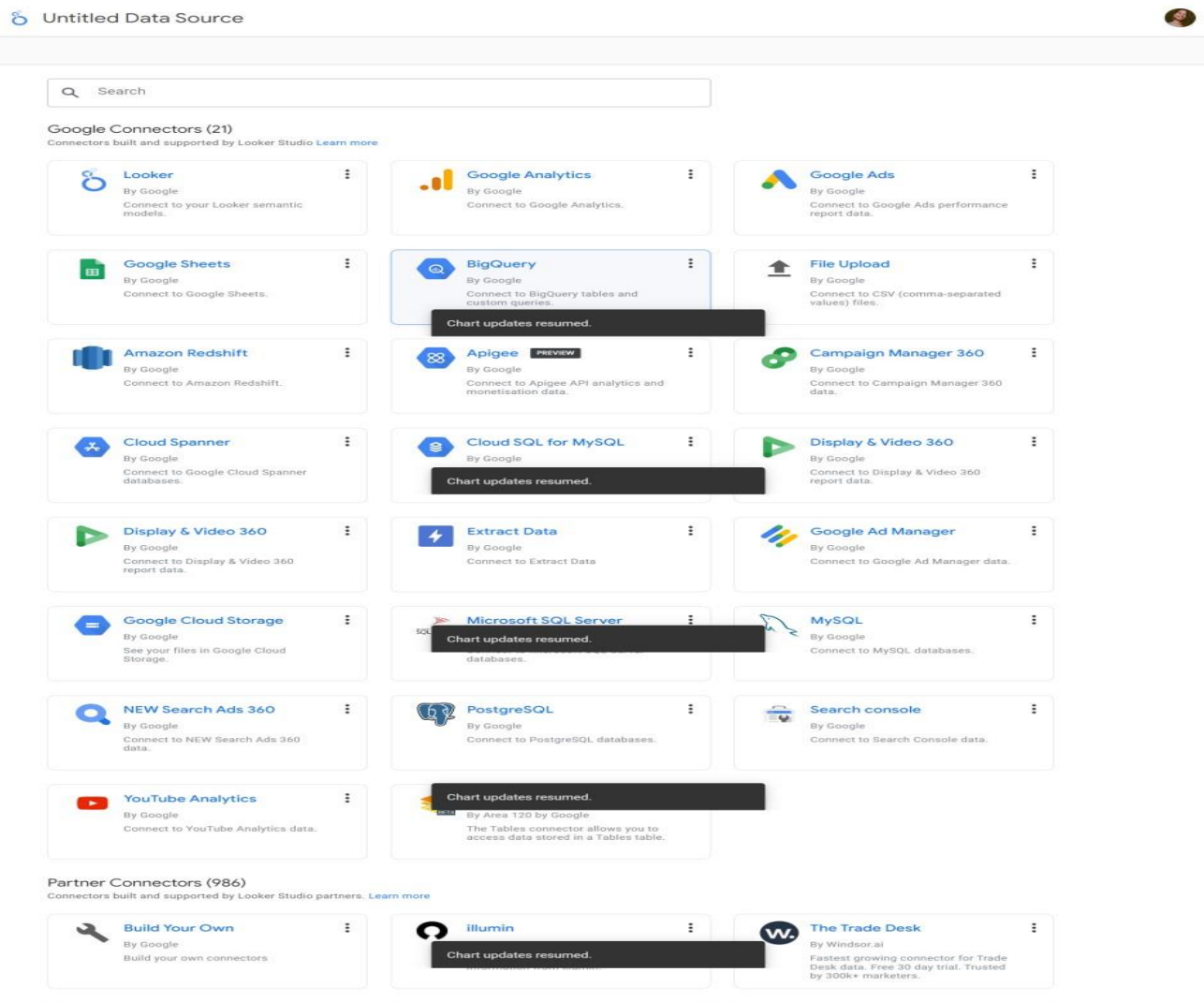
Open looker studio in (chrome)

Click on blank report (AMS)

Click on add data, you can see different data sources.

Click on Big Query

Add tables one by one (you can see this in screen shot)



Untitled Report

FileEditingViewInsertPageArrangeResourceHelp

ResetShareView

Pause updates

Add pageAdd dataAdd a chartAdd a control

Theme and layout

+ Add quick filter

Reset

Data

← Add data to report

Data credentials: Tejaswi Nagarjuna

By Google

BigQuery is Google's fully managed, petabyte scale, low-cost analytics data warehouse. BigQuery charges for querying/processing data. Those queries are charged to the credit card of the billing project.

LEARN MOREREPORT AN ISSUE

RECENT PROJECTS

ProjectData set

bigquery is Google's fully managed, petabyte scale, low-cost analytics data warehouse. BigQuery charges for querying data. Those queries are charged to the credit card of the billing project.

LEARN MOREREPORT AN ISSUE

RECENT PROJECTS

ProjectData set

MY PROJECTS

Enter Project ID manuallyrp\_mysqlslams

SHARED PROJECTS

Attendance Monitoring System

CUSTOM QUERY

My First Project

My First Project

PUBLIC DATA-SETS

Table

attendance\_infoattendance\_timingdept\_infodesignation\_infoemployee\_infoshift\_info

CancelAdd

Untitled Report

FileEditingViewInsertPageArrangeResourceHelp

ResetShareView

Pause updates

Add pageAdd dataAdd a chartAdd a control

Theme and layout

+ Add quick filter

Reset

Data

Let's get started

Drag a field from the Data Panel to the canvas to add a new chart or select a component on the report canvas to edit it.

Data

Search

attendance\_infodept\_infodesignation\_infoemployee\_infoshift\_infoattendance\_timing

Add Data

## Step 9 : Building dashboard by using Looker Studio



### YouTube Links:

#### 1.Creating Google Cloud Platform (GCP) account [Free Tier]

<https://youtu.be/FPd-jC4QIPk?si=Zf8HQcYCb66SKrPE>

#### 2.Creating MySQL instance

<https://youtu.be/xH93izCfKPw?si=NDoSmySW0fO7bG8>

#### 3.Connecting Google Cloud SQL database with Local Workbench

<https://youtu.be/S-PexX3-M2w?si=bhCT22rYrWtSLV4a>

#### 4.Load data from Cloud SQL to Big Query by using Datastream

<https://youtu.be/LOR8najS884?si=ptEZ9yVQaO4jnAfM>

#### 5.Building AMS dashboard by using Looker Studio

<https://youtu.be/mYet7g4CvFQ?si=hjquICbL72rQUryv>