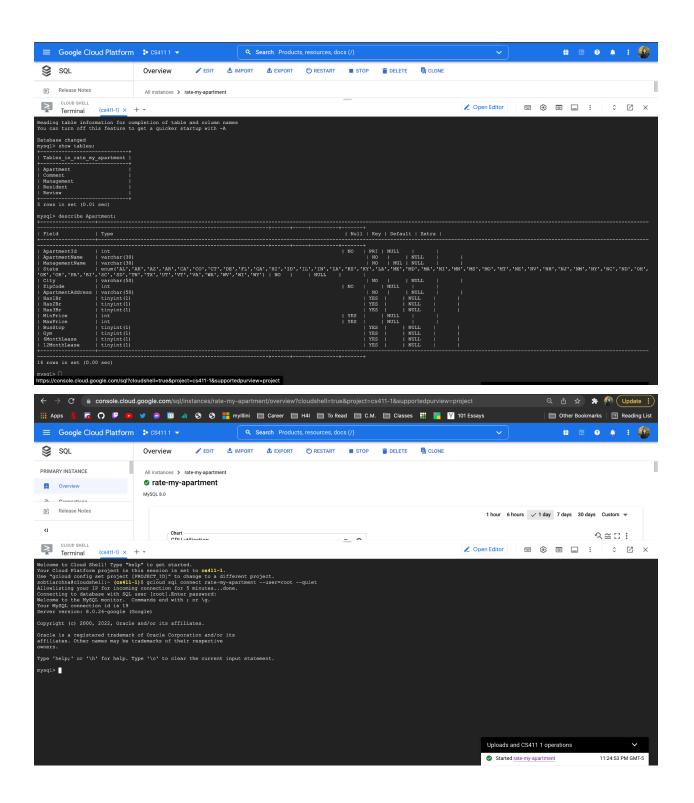
# Stage 3

## **GCP Screenshots of Connection and Table Descriptions**

https://colab.research.google.com/drive/1BDqivQL3J5XUwYTb5XRn6F96They7VRc?usp=sharing - Mock Data script

```
mysql> SELECT COUNT(*) FROM Apartment;
COUNT(*)
| COUNT(*) |
| 1000 |
| 1 row in set (0.05 sec)
l row in set (0.01 sec)
| COUNT(*) |
nysql>
3 rows in set (0.01 sec)
mysql>
```



## **DDL Script for Tables**

```
CREATE TABLE Comment(
    CommentId INTEGER PRIMARY KEY,
    ResidentId INTEGER NOT NULL,
    ApartmentId INTEGER NOT NULL,
    FeedbackText VARCHAR (255),
   OverallScore INTEGER,
   NoiseScore INTEGER,
   LocationScore INTEGER,
   WifiScore INTEGER,
    FOREIGN KEY (ResidentId) REFERENCES Resident(ResidentId) ON DELETE
CASCADE,
    FOREIGN KEY (ApartmentId) REFERENCES Apartment (ApartmentId) ON DELETE
CASCADE
);
CREATE TABLE Resident(
   ResidentId INTEGER PRIMARY KEY,
   ApartmentId INTEGER NOT NULL,
    FirstName VARCHAR(30) NOT NULL,
   LastName VARCHAR(30) NOT NULL,
   Email VARCHAR(50) NOT NULL,
   OauthId INTEGER NOT NULL,
   ProfilePicture VARCHAR (150),
    SchoolGrade ENUM('F', 'So', 'J', 'S') NOT NULL,
   FOREIGN KEY (ApartmentId) REFERENCES Apartment (ApartmentId) ON DELETE
CASCADE
);
CREATE TABLE Apartment(
    ApartmentId INTEGER PRIMARY KEY NOT NULL,
    ApartmentName VARCHAR (30) NOT NULL,
   ManagementName VARCHAR(30) NOT NULL,
    State ENUM('AL', 'AK', 'AZ', 'AR', 'CA', 'CO', 'CT', 'DE', 'FL', 'GA',
'HI', 'ID', 'IL', 'IN', 'IA', 'KS', 'KY', 'LA', 'ME', 'MD', 'MA', 'MI',
'MN', 'MS', 'MO', 'MT', 'NE', 'NV', 'NH', 'NJ', 'NM', 'NY', 'NC', 'ND',
'OH', 'OK', 'OR', 'PA', 'RI', 'SC', 'SD', 'TN', 'TX', 'UT', 'VT', 'VA',
'WA', 'WV', 'WI', 'WY') NOT NULL,
   City VARCHAR (50) NOT NULL,
   ZipCode INTEGER NOT NULL,
```

```
ApartmentAddress VARCHAR(50) NOT NULL,
   Has1Br BOOLEAN,
   Has2Br BOOLEAN,
   Has3Br BOOLEAN,
   MinPrice INTEGER,
   MaxPrice INTEGER,
   BusStop BOOLEAN,
   Gym: BOOLEAN,
   6MonthLease: BOOLEAN,
   12MonthLease: BOOLEAN,
   FOREIGN KEY (ManagementName) REFERENCES Management (Name) ON DELETE
CASCADE
);
CREATE TABLE Review(
   ReviewId INTEGER PRIMARY KEY,
   ApartmentId INTEGER NOT NULL,
   OverallAvgScore: REAL DEFAULT 0,
   NoiseAvgScore: REAL DEFAULT 0,
   LocationAvgScore: REAL DEFAULT 0,
   WifiAvgScore: REAL DEFAULT 0,
   FOREIGN KEY (ApartmentId) REFERENCES Apartment (ApartmentId) ON DELETE
CASCADE
);
CREATE Table Management(
   Name VARCHAR (30) PRIMARY KEY,
   Owner VARCHAR(100) NOT NULL,
   HQ: Enum('AL', 'AK', 'AZ', 'AR', 'CA', 'CO', 'CT', 'DE', 'FL', 'GA',
'HI', 'ID', 'IL', 'IN', 'IA', 'KS', 'KY', 'LA', 'ME', 'MD', 'MA', 'MI',
'MN', 'MS', 'MO', 'MT', 'NE', 'NV', 'NH', 'NJ', 'NM', 'NY', 'NC', 'ND',
'OH', 'OK', 'OR', 'PA', 'RI', 'SC', 'SD', 'TN', 'TX', 'UT', 'VT', 'VA',
'WA', 'WV', 'WI', 'WY') NOT NULL
```

#### Queries

[Query 1]: Selects all of the managements and scores from managements with an average rating above 1.25

SELECT Management.Name, AVG(OverallAvgScore) AS MgmtAvg

FROM Management JOIN Apartment ON Management.Name = Apartment.ManagementName NATURAL JOIN Review GROUP BY Management.Name HAVING MgmtAvg > 1.25;

```
mysql> SELECT Management.Name, AVG(OverallAvgScore) AS MgmtAvg
    -> FROM Management JOIN Apartment ON Management.Name = Apartment.ManagementName NATURAL JOIN Review
   -> GROUP BY Management.Name
   -> HAVING MgmtAvg > 1.25;
| Name
                              | MamtAva
                              | 1.441580756013746 |
| Cole-Dougherty
| Pollard-Mendoza
                              | 1.606529209621993
| Ayala, Stevens and Chapman | 1.5105263157894737
| Singh Inc
| 1.6432748538011697
| Garza-Smith
                              1.5344202898550723
| Thomas, Gutierrez and Alvarez | 1.8061056105610562
| Martinez, Johnson and Campos | 1.4714285714285713
| Fobby's Apartments
| Zobby's Apartments
                            | 1.33333333333333333
| 1.83333333333333333
| Cobby's Apartments
| Robby's Apartments
| Mobby's Apartments
| Nobby's Apartments
| Sobby's Apartments
                                               2.5
17 rows in set (0.01 sec)
```

### [Query 1 but with a limit of 15]

SELECT Management.Name, AVG(OverallAvgScore) AS MgmtAvg
FROM Management JOIN Apartment ON Management.Name = Apartment.ManagementName
NATURAL JOIN Review
GROUP BY Management.Name
HAVING MgmtAvg > 1.25
LIMIT 15;

```
mysql> SELECT Management.Name, AVG(OverallAvgScore) AS MgmtAvg
    -> FROM Management JOIN Apartment ON Management.Name = Apartment.ManagementName NATURAL JOIN Review
    -> GROUP BY Management.Name
    -> HAVING MgmtAvg > 1.25
   -> LIMIT 15;
| Name
                                 | MgmtAvg
                                 | 1.5105263157894737
 Ayala, Stevens and Chapman
 Cobby's Apartments
                                  1.3333333333333333
 Cole-Dougherty
                                  1.4415807560137455
 Fobby's Apartments
 Garza-Smith
                                 1 1.5344202898550723
 Gutierrez LLC
                                  1.6253205128205126
 Martinez, Johnson and Campos
                                  1.4714285714285713
 Mobby's Apartments
Nobby's Apartments
 Pollard-Mendoza
                                  1.6065292096219932
                                  1.833333333333333
 Robby's Apartments
 Singh Inc
                                  1.6432748538011694
 Sobby's Apartments
                                                  2.5
 Thomas, Gutierrez and Alvarez | 1.8061056105610562
                                  1.382302405498282
 Wilson, Gonzalez and Kim
15 rows in set (0.01 sec)
```

## [Query 2]: Select all apartment comments from AK

SELECT CommentId, OverallScore, NoiseScore, LocationScore, WifiScore, FeedbackText, State

FROM Comment c JOIN Resident r ON r.ResidentId = c.ResidentId JOIN Apartment b ON b.ApartmentId = r.ApartmentId

WHERE r.ApartmentId IN (SELECT a.ApartmentId

FROM Apartment a WHERE a.State = 'AK');

```
mysql> SELECT Commentid, OverallScore, NoiseScore, JocationScore, WifiScore, FeedbackText, State

-> FROM Comment c JOIN Resident r ON r.Residentid = c.ResidentId JOIN Apartment b ON b.ApartmentId = r.ApartmentId

-> WEERE r.ApartmentId IN (SELECT a.ApartmentId IN (SELECT a.ApartmentId = r.ApartmentId = r.ApartmentId |

-> WEERE r.ApartmentId IN (SELECT a.ApartmentId IN (SELECT a.ApartmentId = r.ApartmentId = r.ApartmentId |

-> WEERE r.ApartmentId IN (SELECT a.ApartmentId |

FROM Apartment a |

-> WEERE r.ApartmentId IN (SELECT a.ApartmentId |

FROM Apartment a |

-> WEERE r.ApartmentId IN (SELECT a.ApartmentId |

FROM Apartment a |

-> WEERE r.ApartmentId IN (SELECT a.ApartmentId |

FROM Apartment a |

FROM Apartme
```

#### [Query 2 but with a limit of 15]

SELECT CommentId, OverallScore, NoiseScore, LocationScore, WifiScore, FeedbackText, State

FROM Comment c JOIN Resident r ON r.ResidentId = c.ResidentId JOIN Apartment b ON b.ApartmentId = r.ApartmentId

# WHERE r.ApartmentId IN (SELECT a.ApartmentId FROM Apartment a WHERE a.State = 'AK')

LIMIT 15;

```
eyeq's SERECT Commented. Overall Score, NoiseScore, LocationScore, WidiScore, PembesCart, State

-> MIRRE r.Apartmentid IN (SCLECT a.Apartment)
-> MIRRE r.Apartmentid IN (SCLECT a.Apartment)
-> MIRRE gr.Apartment of IN (SCLECT a.Apar
```

# Indexing

To demonstrate performance before and after, here are the results of running EXPLAIN ANALYZE on *Query 1* (as denoted above)

## **BEFORE** indexing:

## **AFTER indexing on Management Name in Management:**

- You can note [time differences] (i.e. filter 100+ -> 13)
  - Indexing on Management Name in Apartment yields significantly better results than indexing on Management Name in Management. To reference the times, the query executed before indexing had an actual time of 124.031..124.038 whereas after indexing on the management name in Apartment, this query had an actual time of 0.001..0.005. While we somewhat expected indexing management name on Management to perform better because of it being a clustered index, we found indexing management name on Apartment to perform better to be a surprise. We understand how indexing is utilized in WHERE statements, however, our complex query relies on GROUP BY and having. Because of this and the way the query joins with the Apartment table, we believe that query 2 is performing because of this group by and having condition. Additionally, because the strength of indexing is amplified when doing ranged querying, this query was the perfect opportunity to index on the management name of the apartment table, because it would be easy for the B+ tree to navigate down to the management average and find the averages in the correct range.

## **AFTER indexing on Management Name in Apartment:**

```
Sysql> CREATE INDEX amm on Apartment (ManagementName);
Query OK, 0 rows affected (0.17 sec)
Records: 0 Duplicates: 0 Management Name, ANG(OverallAugGoore) AS Mymthy

> STROM Management ON Management Name = Apartment ManagementName NATURAL JOIN Review

-> CREATE INDEX and Management Name = Apartment ManagementName NATURAL JOIN Review

-> RAVING Mymthy

-> RAVING Mymthy

-> NAVING Mymthy
-1.25;

| EXPLAIN | Management Name | Apartment ON ManagementName NATURAL JOIN Review

-> Filter: (Mymthy
-1.25) (cost=074.03 rows=1000) (actual time=0.63. 5.528 rows=17 loops=1)

-> Croup aggregate: avg (Review.OverallAugGoore) (cost=074.03 rows=1000) (actual time=0.534.5.510 rows=20 loops=1)

-> Nested loop inner join (cost=124.03 rows=1000) (actual time=0.032.5.038 rows=1000 loops=1)

-> Nested loop inner join (cost=124.03 rows=1000) (actual time=0.072.0.038 rows=1000 loops=1)

-> Index lookup on Apartment using PRIMARY (cost=2.25 rows=2) (actual time=0.017.0.038 rows=20 loops=1)

-> Index lookup on Review using ApartmentId (ApartmentId=Apartment.ApartmentCd) (cost=0.25 rows=1) (actual time=0.033.0.004 rows=1000)

1 row in set (0.01 sec)

mysql>[]
```

Here are the results of running EXPLAIN ANALYZE on **Query 2** (as denoted above)

## **BEFORE** indexing:

```
mysql> EXPLAIN ANALYZE SELECT Commented. OverallScore, NoiseScore, LocationScore, WifiScore, FeedbackText, State

-> FDOX Comment count Was indicated a Apartment of DNA Apartment of DNA Apartment of TROM Apartm
```

## AFTER indexing on ApartmentId and State from the Apartment table:

AFTER indexing on ApartmentId and ResidentId from the Resident table:

```
mysql> CHEATE INDEX ap_res_id ON Resident(Apartmentid, Residentid);

Gouly OF, Orawa Edected (6.1 sec)

Eygql> EXPLAIN ANALYZE SELECT Commentid, OverallScore, NoiseScore, LocationScore, NifiScore, PeedBackText, State

> FROM Comment colon Resident ON F.Residentid = c.Residentid JOIN Apartmentid = r.Apartmentid = r.Apartmentid |

> WHERE r.Apartmentid IN (SELECT a.Apartmentid |

> Nested loop inner join (cost-458.72 rows-246) (actual time-0.091.0.095 rows-24 loops-1)

-> Nested loop inner join (cost-458.72 rows-246) (actual time-0.090.0.095 rows-24 loops-1)

-> Nested loop inner join (cost-458.72 rows-29) (actual time-0.090.0.095 rows-29 loops-1)

-> Pilter: (a.State = "AK") (cost-033.33 rows-1000 (actual time-0.095.0.063 rows-29 loops-1)

-> Index stond loop inner join (cost-456.77 rows-00) (actual time-0.056.0.063 rows-100 loops-1)

-> Index lookup on r using ap_reg_id (Apartmentid-s.Apartmentid (cost-0.25 rows-2) (actual time-0.003.0.000 rows-1 loops-21)

-> Index lookup on c using Residentid (Residentid-r.Residentid) (cost-0.25 rows-2) (actual time-0.003.0.000 rows-1 loops-21)

-> Index lookup on c using Residentid (Residentid-r.Residentid) (cost-0.55 rows-2) (actual time-0.003.0.000 rows-1 loops-23)
```

## AFTER indexing on CommentId and ResidentId from the Comment table:

```
Database changed
mysely CREATE INDEX con_res_id ON Comment(Comment), Residentid);
Casey ON, Or magnetic EXPLAIN

Rectain: Deplicates: O Northings On Principle On
```

Clearly, the second index is going to be the most efficient index to use for query 2
because within the WHERE clause of condition 2 we are finding when apartment IDs
and resident IDs are equal. By creating an index on both of these attributes, we
determined that the query will be most optimized.