PT1 Stage 3

Tables

-- Users
CREATE TABLE IF NOT EXISTS User (
  user_id INT AUTO_INCREMENT PRIMARY KEY,
  name VARCHAR(100),
  email VARCHAR(100) UNIQUE
);

-- City, *Populated from Ticketmaster API*
CREATE TABLE IF NOT EXISTS City (
  city_id INT AUTO_INCREMENT PRIMARY KEY,
  city_name VARCHAR(100) NOT NULL,
  state VARCHAR(50) NOT NULL,
);

-- Venue, *Populated from Ticketmaster API*
CREATE TABLE IF NOT EXISTS Venue (
  venue_id BIGINT AUTO_INCREMENT PRIMARY KEY,
  venue_name VARCHAR(255) NOT NULL,
  city_id INT NOT NULL,
  latitude DECIMAL(9,6),
  longitude DECIMAL(9,6),
  city_name VARCHAR(255) NOT NULL
  FOREIGN KEY (city_id) REFERENCES City(city_id)
);

-- Event, *>1000 rows from Ticketmaster API*
CREATE TABLE IF NOT EXISTS Event (
  event_id BIGINT AUTO_INCREMENT PRIMARY KEY,
  name VARCHAR(255) NOT NULL,
  date DATE,
  ticket_price DECIMAL(10,2),
  venue_name VARCHAR(255) NOT NULL,
  venue_id BIGINT
  FOREIGN KEY (venue_id) REFERENCES Venue(venue_id)
  );

```
mysql> SHOW TABLES;
+------------------+
| Tables_in_getlitty |
+------------------+
| AirbnbListing    |
| City             |
| Event            |
| Event_backup     |
| Nearby           |
| User             |
| Venue            |
| Venue_backup     |
| WantsToAttend    |
+------------------+
9 rows in set (0.00 sec)
```

```
mysql> SELECT COUNT(*) AS row_count
    -> FROM Event;
+-----------+
| row_count |
+-----------+
|      1188 |
+-----------+
1 row in set (0.00 sec)
```

-- AirbnbListing,   *>1000 rows from InsideAirbnb CSV*
CREATE TABLE IF NOT EXISTS AirbnbListing (
  listing_id BIGINT PRIMARY KEY,
  city_id INT NOT NULL,
  latitude DECIMAL(9,6),
  longitude DECIMAL(9,6),
  price_per_night DECIMAL(10,2),
  availability_365 INT,
  room_type VARCHAR(50),
  city_name VARCHAR(255) NOT NULL
  FOREIGN KEY (city_id) REFERENCES City(city_id)
);

```
mysql> SELECT COUNT(*) AS row_count
    -> FROM AirbnbListing;
+-----------+
| row_count |
+-----------+
|      7071 |
+-----------+
1 row in set (0.01 sec)
```

-- WantsToAttend (bridge)
CREATE TABLE IF NOT EXISTS WantsToAttend (
  user_id INT,
  event_id BIGINT,
  PRIMARY KEY (user_id, event_id),
  FOREIGN KEY (user_id) REFERENCES User(user_id),
  FOREIGN KEY (event_id) REFERENCES Event(event_id)
);

-- Nearby (Event <-> Airbnb),   *> 1000 rows, Populated from Event and Airbnb*
CREATE TABLE IF NOT EXISTS Nearby (

event_id BIGINT,
listing_id BIGINT,
total_cost DECIMAL(12,2),
distance DECIMAL(8,2),
PRIMARY KEY (event_id, listing_id),
FOREIGN KEY (event_id) REFERENCES Event(event_id),
FOREIGN KEY (listing_id) REFERENCES AirbnbListing(listing_id)
);

```
mysql> SELECT COUNT(*) AS row_count
    -> FROM Nearby ;
+-----------+
| row_count |
+-----------+
|   8400348 |
+-----------+
1 row in set (9.44 sec)
```

Queries

QUERY 1: Cheapest Airbnb option per event within a mile
- Filtering by max distance and cost ranking for each event
- Concepts: Subquery, Join, Aggregation via GROUP BY

SELECT e.event_id, e.name AS event_name, c.city_name, c.state,
   MIN(n.total_cost) AS cheapest_total_cost
FROM Event e
JOIN Venue v ON e.venue_id = v.venue_id
JOIN Nearby n ON e.event_id = n.event_id
JOIN AirbnbListing a ON n.listing_id = a.listing_id
JOIN City c ON v.city_id = c.city_id
WHERE n.distance <= 1
GROUP BY e.event_id, e.name, c.city_name, c.state
ORDER BY cheapest_total_cost ASC
LIMIT 15;

```
+----------+------------------------------------------+-----------+----------+--------------------+
| event_id | event_name                               | city_name | state    | cheapest_total_cost |
+----------+------------------------------------------+-----------+----------+--------------------+
|      513 | Chicago Blackhawks vs. Detroit Red Wings | Chicago   | Illinois |              35.00 |
|      627 | Chicago Blackhawks vs. Philadelphia Flyers | Chicago | Illinois |              35.00 |
|      399 | Chicago Blackhawks vs. Anaheim Ducks     | Chicago   | Illinois |              35.00 |
|      666 | Chicago Blackhawks vs. Pittsburgh Penguins | Chicago | Illinois |              35.00 |
|      721 | Chicago Blackhawks vs. Vegas Golden Knights | Chicago | Illinois |             35.00 |
|      740 | Chicago Blackhawks vs. St. Louis Blues   | Chicago   | Illinois |              35.00 |
|      957 | Chicago Blackhawks vs. San Jose Sharks   | Chicago   | Illinois |              35.00 |
|       60 | Chicago Blackhawks vs. Ottawa Senators   | Chicago   | Illinois |              35.00 |
|      213 | Chicago Blackhawks vs. New Jersey Devils | Chicago   | Illinois |              35.00 |
|      270 | Chicago Blackhawks vs. Calgary Flames    | Chicago   | Illinois |              35.00 |
|      287 | Chicago Blackhawks vs. Seattle Kraken    | Chicago   | Illinois |              35.00 |
|      330 | Chicago Blackhawks vs. Colorado Avalanche | Chicago  | Illinois |              35.00 |
|      336 | PAUL McCARTNEY: GOT BACK                 | Chicago   | Illinois |              35.00 |
|      345 | PAUL McCARTNEY: GOT BACK                 | Chicago   | Illinois |              35.00 |
|      357 | Chicago Blackhawks vs. Minnesota Wild    | Chicago   | Illinois |              35.00 |
+----------+------------------------------------------+-----------+----------+--------------------+
15 rows in set (4.29 sec)
```

<u>Indexing</u>

Index	Attributes Indexed	Query Cost	Notes
Baseline	N/A	83,755	N/A
idx_nearby	Nearby (distance, event_id)	36,652	Improvement
idx_event	Event (venue_id)	83,755	No improvements
idx_venue	Venue (city_id)	83,755	No improvements

Adding the index on Nearby(distance, event_id) reduced the cost from 83,755 to 36,652. This is because MySQL could better filter rows using an index range scan on distance <= 1, instead of scanning the entire Nearby table. Indexes on Event(venue_id) and Venue(city_id) made no difference since those joins were already optimized by primary keys.

QUERY 2: Top 5 cheapest concert trips in Illinois
- Supports our "filter by state" and affordability ranking features
- Concepts: JOIN, Aggregation via GROUP BY

SELECT E.name AS event_name, C.city_name, C.state, E.date,
   MIN(N.total_cost) AS cheapest_total_cost
FROM Event E
JOIN Venue V ON E.venue_id = V.venue_id
JOIN City C ON V.city_id = C.city_id
JOIN Nearby N ON E.event_id = N.event_id
WHERE C.state = 'Illinois'
GROUP BY E.event_id, C.city_name, C.state, E.date, E.name
ORDER BY cheapest_total_cost ASC
LIMIT 15;

```
+-------------------------------------------+-----------+----------+------------+--------------------+
| event_name                                | city_name | state    | date       | cheapest_total_cost |
+-------------------------------------------+-----------+----------+------------+--------------------+
| Chicago Blackhawks vs. Ottawa Senators    | Chicago   | Illinois | 2025-10-28 |              13.00 |
| Chicago Bears vs. New York Giants         | Chicago   | Illinois | 2025-11-09 |              13.00 |
| Chicago Blackhawks vs. New Jersey Devils  | Chicago   | Illinois | 2025-11-12 |              13.00 |
| Chicago Blackhawks vs. Calgary Flames     | Chicago   | Illinois | 2025-11-18 |              13.00 |
| Chicago Blackhawks vs. Seattle Kraken     | Chicago   | Illinois | 2025-11-20 |              13.00 |
| Chicago Bears vs. Pittsburgh Steelers     | Chicago   | Illinois | 2025-11-23 |              13.00 |
| Chicago Blackhawks vs. Colorado Avalanche | Chicago   | Illinois | 2025-11-23 |              13.00 |
| PAUL McCARTNEY: GOT BACK                   | Chicago   | Illinois | 2025-11-24 |              13.00 |
| PAUL McCARTNEY: GOT BACK                   | Chicago   | Illinois | 2025-11-25 |              13.00 |
| Chicago Blackhawks vs. Minnesota Wild     | Chicago   | Illinois | 2025-11-26 |              13.00 |
| Chicago Blackhawks vs. Nashville Predators| Chicago   | Illinois | 2025-11-28 |              13.00 |
| Chicago Blackhawks vs. Anaheim Ducks      | Chicago   | Illinois | 2025-11-30 |              13.00 |
| Chicago Blackhawks vs. Detroit Red Wings  | Chicago   | Illinois | 2025-12-13 |              13.00 |
| Chicago Bears vs. Cleveland Browns        | Chicago   | Illinois | 2025-12-14 |              13.00 |
| Chicago Bears vs. Green Bay Packers       | Chicago   | Illinois | 2025-12-20 |              13.00 |
+-------------------------------------------+-----------+----------+------------+--------------------+
15 rows in set (0.99 sec)
```

Indexing

Index	Attributes Indexed	Query Cost	Notes
Baseline	N/A	4081	N/A
idx_city	City (state)	553	Improvement
idx_nearby	Nearby (event_id)	4081	No improvement
idx_event	Event (venue_id)	4081	No improvement
idx_venue	Venue (city_id)	4081	No improvement

Indexing City(state) reduced the cost from 4081 to 553, since MySQL could quickly find only Illinois rows before joining and grouping. Other indexes, such as on Nearby(event_id) or Venue(city_id), didn't allow for any improvement because those joins were already efficiently handled with key lookups.

QUERY 3: Events with most available Airbnb Listings (within 5 miles)
-   Focuses on listing availability count
-   Concepts: JOIN, Aggregation via Group by

SELECT E.event_id, E.name AS event_name, C.city_name, C.state, COUNT(A.listing_id) AS
num_available_listings, ROUND(AVG(A.price_per_night), 2) AS avg_price_per_night,
MIN(N.distance) AS closest_listing_distance
FROM Event E
JOIN Venue V ON E.venue_id = V.venue_id
JOIN City C ON V.city_id = C.city_id
JOIN Nearby N ON E.event_id = N.event_id
JOIN AirbnbListing A ON N.listing_id = A.listing_id
WHERE N.distance <= 5 AND A.availability_365 > 0
GROUP BY E.event_id, E.name, C.city_name, C.state
ORDER BY num_available_listings DESC, avg_price_per_night ASC
LIMIT 15;

```
+----------+-------------------------------------------+-----------+----------+------------------------+---------------------+-------------------------+
| event_id | event_name                                | city_name | state    | num_available_listings | avg_price_per_night | closest_listing_distance |
+----------+-------------------------------------------+-----------+----------+------------------------+---------------------+-------------------------+
|       60 | Chicago Blackhawks vs. Ottawa Senators    | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      213 | Chicago Blackhawks vs. New Jersey Devils  | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      270 | Chicago Blackhawks vs. Calgary Flames     | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      287 | Chicago Blackhawks vs. Seattle Kraken     | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      740 | Chicago Blackhawks vs. St. Louis Blues    | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      330 | Chicago Blackhawks vs. Colorado Avalanche | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      336 | PAUL McCARTNEY: GOT BACK                   | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      345 | PAUL McCARTNEY: GOT BACK                   | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      357 | Chicago Blackhawks vs. Minnesota Wild     | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      374 | Chicago Blackhawks vs. Nashville Predators | Chicago  | Illinois |                   4771 |              534.35 |                    0.15 |
|      399 | Chicago Blackhawks vs. Anaheim Ducks      | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      513 | Chicago Blackhawks vs. Detroit Red Wings  | Chicago   | Illinois |                   4771 |              534.35 |                    0.15 |
|      721 | Chicago Blackhawks vs. Vegas Golden Knights | Chicago | Illinois |                   4771 |              534.35 |                    0.15 |
|      666 | Chicago Blackhawks vs. Pittsburgh Penguins | Chicago  | Illinois |                   4771 |              534.35 |                    0.15 |
|      627 | Chicago Blackhawks vs. Philadelphia Flyers | Chicago  | Illinois |                   4771 |              534.35 |                    0.15 |
+----------+-------------------------------------------+-----------+----------+------------------------+---------------------+-------------------------+
15 rows in set (4.35 sec)
```

Indexing

Index	Attributes Indexed	Query Cost	Notes
Baseline	N/A	83,755	N/A
idx_nearby	Nearby (distance, event_id)	44,283	Improvement
idx_airbnb	AirbnbListing (availability_365)	83,755	No improvement
idx_event	Event (venue_id)	83,755	No improvement
idx_venue	Venue (city_id)	83,755	No improvement

Indexing Nearby(distance, event_id) reduced the query cost from 83,755 to 44,283, which was the most significant performance improvement for this query. The index helps to efficiently filter listings within 5 miles and join on N.event_id, so we avoid a full scan of the Nearby table. Without it, the database must check all rows, which is costly given the large number of event-listing pairs. The other indexes did not improve performance, as the filters and joins they support are either not highly selective or already handled. Overall, the index on Nearby directly supports the query's main filtering and join conditions, resulting in the largest cost reduction.

QUERY 4: Find Chicago concerts with below-average lodging costs (within a mile of the venue)
- Represents affordability within a specific city (Chicago) by identifying events that have below-average lodging costs within a set distance range
- Concepts: Multiple JOINs, Aggregation via GROUP BY, Subquery

```
SELECT E.event_id, E.name AS event_name, C.city_name, C.state, E.date,
   MIN(A.price_per_night) AS cheapest_airbnb_price
FROM Event E
JOIN Venue V ON E.venue_id = V.venue_id
JOIN City C ON V.city_id = C.city_id
JOIN Nearby N ON E.event_id = N.event_id
JOIN AirbnbListing A ON N.listing_id = A.listing_id
WHERE C.city_name = 'Chicago'
  AND N.distance <= 1
  AND A.availability_365 > 0 -- available listings only
GROUP BY E.event_id, E.name, C.city_name, C.state, E.date
HAVING MIN(A.price_per_night) < (
   SELECT AVG(A2.price_per_night)
   FROM AirbnbListing A2
   JOIN Nearby N2 ON A2.listing_id = N2.listing_id
   JOIN Event E2 ON N2.event_id = E2.event_id
   JOIN Venue V2 ON E2.venue_id = V2.venue_id
   JOIN City C2 ON V2.city_id = C2.city_id
   WHERE C2.city_name = 'Chicago'
      AND N2.distance <= 1
)
ORDER BY cheapest_airbnb_price ASC
LIMIT 15;
```

```
+----------+-------------------------------------------+-----------+----------+------------+-----------------------+
| event_id | event_name                                | city_name | state    | date       | cheapest_airbnb_price |
+----------+-------------------------------------------+-----------+----------+------------+-----------------------+
|       60 | Chicago Blackhawks vs. Ottawa Senators    | Chicago   | Illinois | 2025-10-28 |                 35.00 |
|      213 | Chicago Blackhawks vs. New Jersey Devils  | Chicago   | Illinois | 2025-11-12 |                 35.00 |
|      270 | Chicago Blackhawks vs. Calgary Flames     | Chicago   | Illinois | 2025-11-18 |                 35.00 |
|      287 | Chicago Blackhawks vs. Seattle Kraken     | Chicago   | Illinois | 2025-11-20 |                 35.00 |
|      330 | Chicago Blackhawks vs. Colorado Avalanche | Chicago   | Illinois | 2025-11-23 |                 35.00 |
|      336 | PAUL McCARTNEY: GOT BACK                   | Chicago   | Illinois | 2025-11-24 |                 35.00 |
|      345 | PAUL McCARTNEY: GOT BACK                   | Chicago   | Illinois | 2025-11-25 |                 35.00 |
|      357 | Chicago Blackhawks vs. Minnesota Wild     | Chicago   | Illinois | 2025-11-26 |                 35.00 |
|      374 | Chicago Blackhawks vs. Nashville Predators| Chicago   | Illinois | 2025-11-28 |                 35.00 |
|      399 | Chicago Blackhawks vs. Anaheim Ducks      | Chicago   | Illinois | 2025-11-30 |                 35.00 |
|      513 | Chicago Blackhawks vs. Detroit Red Wings  | Chicago   | Illinois | 2025-12-13 |                 35.00 |
|      627 | Chicago Blackhawks vs. Philadelphia Flyers| Chicago   | Illinois | 2025-12-23 |                 35.00 |
|      666 | Chicago Blackhawks vs. Pittsburgh Penguins| Chicago   | Illinois | 2025-12-28 |                 35.00 |
|      721 | Chicago Blackhawks vs. Vegas Golden Knights| Chicago  | Illinois | 2026-01-04 |                 35.00 |
|      740 | Chicago Blackhawks vs. St. Louis Blues    | Chicago   | Illinois | 2026-01-07 |                 35.00 |
+----------+-------------------------------------------+-----------+----------+------------+-----------------------+
15 rows in set (0.24 sec)
```

<u>Indexing</u>

Index	Attributes Indexed	Query Cost	Notes
Baseline	N/A	8678	N/A
idx_city	City (city_name)	1183	Improvement
idx_nearby	Nearby (distance, event_id)	4115	Less improvement
idx_airbnb	AirbnbListing (availability_365)	8678	No improvement
idx_event	Event (venue_id)	8678	No improvement
idx_venue	Venue (city_id)	8678	No improvement

Indexing City (city_name) greatly improved cost because the query filters specifically on city_name = 'Chicago' in both the main query and the subquery. Without an index, MySQL had to perform a full table scan of the City table each time to find matching rows, which is inefficient even for moderately sized tables. By adding an index, MySQL could quickly locate only the rows for Chicago using the index lookup instead of scanning the entire table. This allowed the database to reduce the number of rows joined with other tables early in the execution, significantly lowering the overall query cost.