

Netflix Case study

This case study focuses on leveraging Apache Cassandra to manage a Netflix dataset containing user subscriptions, which includes details such as user demographics, subscription types, revenue, and device usage. It explores key aspects like creating a keyspace with appropriate replication strategies, designing a schema with suitable data types, and defining primary and clustering keys for efficient data retrieval. The study also covers CRUD operations (insert, update, delete), batch operations for optimizing data handling, and deriving actionable analytics.

Key Analytics

1. Total Revenue by Country: Analyzing total revenue contribution from different countries.
2. User Count by Device: Understanding which devices are most popular among users.
3. Subscription Type Popularity: Tracking the popularity of different subscription plans.
4. Churn Rate Analysis: Identifying users who haven't made a recent payment (using last_payment_date).
5. Demographic Insights: Understanding the age distribution and gender ratios of users.

| User ID | Subscription Type | Monthly Revenue | Join Date | Last Payment Date | Country | Age | Gender | Device | Plan Duration |
|---------|-------------------|-----------------|-----------|-------------------|----------------|-----|--------|------------|---------------|
| 1 | Basic | 10 | 15-01-22 | 10/6/2023 | United States | 28 | Male | Smartphone | 1 Month |
| 2 | Premium | 15 | 5/9/2021 | 22-06-23 | Canada | 35 | Female | Tablet | 1 Month |
| 3 | Standard | 12 | 28-02-23 | 27-06-23 | United Kingdom | 42 | Male | Smart TV | 1 Month |
| 4 | Standard | 12 | 10/7/2022 | 26-06-23 | Australia | 51 | Female | Laptop | 1 Month |
| 5 | Basic | 10 | 1/5/2023 | 28-06-23 | Germany | 33 | Male | Smartphone | 1 Month |
| 6 | Premium | 15 | 18-03-22 | 27-06-23 | France | 29 | Female | Smart TV | 1 Month |
| 7 | Standard | 12 | 9/12/2021 | 25-06-23 | Brazil | 46 | Male | Tablet | 1 Month |
| 8 | Basic | 10 | 2/4/2023 | 24-06-23 | Mexico | 39 | Female | Laptop | 1 Month |
| 9 | Standard | 12 | 20-10-22 | 23-06-23 | Spain | 37 | Male | Smartphone | 1 Month |
| 10 | Premium | 15 | 7/1/2023 | 22-06-23 | Italy | 44 | Female | Smart TV | 1 Month |
| 11 | Basic | 10 | 16-05-22 | 22-06-23 | United States | 31 | Female | Smartphone | 1 Month |
| 12 | Premium | 15 | 23-03-23 | 28-06-23 | Canada | 45 | Male | Tablet | 1 Month |
| 13 | Standard | 12 | 30-11-21 | 27-06-23 | United Kingdom | 48 | Female | Laptop | 1 Month |
| 14 | Basic | 10 | 1/8/2022 | 26-06-23 | Australia | 27 | Male | Smartphone | 1 Month |
| 15 | Standard | 12 | 9/5/2023 | 28-06-23 | Germany | 38 | Female | Smart TV | 1 Month |
| 16 | Premium | 15 | 7/4/2022 | 27-06-23 | France | 36 | Male | Tablet | 1 Month |
| 17 | Basic | 10 | 24-01-22 | 25-06-23 | Brazil | 30 | Female | Laptop | 1 Month |
| 18 | Standard | 12 | 18-10-21 | 24-06-23 | Mexico | 43 | Male | Smartphone | 1 Month |
| 19 | Premium | 15 | 15-02-23 | 23-06-23 | Spain | 32 | Female | Smart TV | 1 Month |
| 20 | Basic | 10 | 27-05-23 | 22-06-23 | Italy | 41 | Male | Tablet | 1 Month |
| 21 | Premium | 15 | 10/6/2023 | 22-06-23 | United States | 26 | Female | Laptop | 1 Month |
| 22 | Basic | 10 | 22-07-22 | 28-06-23 | Canada | 34 | Male | Smartphone | 1 Month |
| 23 | Standard | 12 | 5/12/2021 | 27-06-23 | United Kingdom | 49 | Female | Smart TV | 1 Month |
| 24 | Standard | 12 | 3/4/2022 | 26-06-23 | Australia | 31 | Male | Tablet | 1 Month |
| 25 | Basic | 10 | 14-03-23 | 28-06-23 | Germany | 40 | Female | Laptop | 1 Month |

1. Keyspace Creation

A keyspace in Cassandra is similar to a database in other RDBMS. It defines the scope for data replication and provides a namespace for tables.

Example:

```
CREATE KEYSPACE netflix_analytics
```

```
WITH replication = {'class': 'SimpleStrategy', 'replication_factor': '1'};
```

Explanation:

class: Defines the replication strategy (e.g., SimpleStrategy or NetworkTopologyStrategy).

replication_factor: The number of replicas for each data point.

2. Create a Table

Creating a table involves defining the data types for each column.

Example:

```
CREATE TABLE netflix_analytics.subscriptions (  
    user_id UUID PRIMARY KEY,  
    subscription_type TEXT,  
    monthly_revenue DECIMAL,  
    join_date DATE,  
    last_payment_date DATE,  
    country TEXT,  
    age INT,  
    gender TEXT,  
    device TEXT,  
    plan_duration INT  
);
```

Explanation:

UUID: Universally Unique Identifier, A 128-bit unique identifier ideal for globally unique values.

TEXT: Stores variable-length strings for text-based data. Used for string data (subscription_type, country, gender, device).

DECIMAL: Represents fixed-point numbers, suitable for precise financial values. For monetary values (monthly_revenue).

DATE: Stores only the date component (year, month, day). For date fields (join_date, last_payment_date).

INT: A 32-bit signed integer for numerical data without decimals. For integers (age, plan_duration).

3. Defining Primary Key, Partition Key, Clustering

Primary Key: A unique identifier for rows in a table.

Partition Key: Determines the distribution of data across nodes.

Clustering Columns: Define the order of data storage within a partition.

Example:

```
CREATE TABLE netflix_analytics.subscriptions (  
    user_id UUID,  
    country TEXT,  
    device TEXT,  
    subscription_type TEXT,  
    monthly_revenue DECIMAL,  
    join_date DATE,  
    last_payment_date DATE,  
    age INT,  
    gender TEXT,  
    plan_duration INT,  
    PRIMARY KEY ((country, device), user_id)  
);
```

Explanation:

Partition Key: (country, device) ensures data distribution based on country and device.

Clustering Key: user_id defines the ordering within the partition.

4. Insert Command and Batch Insert

Inserts data into the table.

Single Insert Example:

```
INSERT INTO netflix_analytics.subscriptions (  
    user_id, country, device, subscription_type, monthly_revenue,  
    join_date, last_payment_date, age, gender, plan_duration  
) VALUES (  
    uuid(), 'USA', 'Mobile', 'Premium', 15.99, '2023-01-01', '2024-01-01', 25, 'M', 12
```

);

Batch Insert Example:

BEGIN BATCH

INSERT INTO netflix_analytics.subscriptions (user_id, country, device, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration)

VALUES (uuid(), 'USA', 'Mobile', 'Premium', 15.99, '2023-01-01', '2024-01-01', 25, 'M', 12);

INSERT INTO netflix_analytics.subscriptions (user_id, country, device, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration)

VALUES (uuid(), 'Canada', 'Laptop', 'Standard', 12.99, '2023-02-01', '2024-02-01', 30, 'F', 6);

APPLY BATCH;

5. Update Command and Batch Update

Updates existing records.

Single Update Example:

UPDATE netflix_analytics.subscriptions

SET monthly_revenue = 19.99

WHERE country = 'USA' AND device = 'Mobile' AND user_id = some-uuid;

Batch Update Example:

BEGIN BATCH

UPDATE netflix_analytics.subscriptions

SET monthly_revenue = 19.99

WHERE country = 'USA' AND device = 'Mobile' AND user_id = some-uuid;

UPDATE netflix_analytics.subscriptions

SET plan_duration = 18

WHERE country = 'Canada' AND device = 'Laptop' AND user_id = another-uuid;

APPLY BATCH;

6. Delete Command

Removes records from the table.

Example:

```
DELETE FROM netflix_analytics.subscriptions  
WHERE country = 'USA' AND device = 'Mobile' AND user_id = some-uuid;
```

7. Analytics Derived from the Dataset

You can derive various insights such as:

Monthly Revenue Per Country:

```
SELECT country, SUM(monthly_revenue) AS total_revenue  
FROM netflix_analytics.subscriptions  
GROUP BY country, device;
```

Active Users by Device Type:

```
SELECT device, COUNT(*) AS user_count  
FROM netflix_analytics.subscriptions  
GROUP BY country, device;
```

Average Age of Subscribers:

```
SELECT AVG(age) AS average_age  
FROM netflix_analytics.subscriptions;
```

Subscription Types Distribution:

```
SELECT subscription_type, COUNT(*) AS subscription_count  
FROM netflix_analytics.subscriptions  
GROUP BY country, device;
```

Revenue Trends Over Time:

```
SELECT join_date, SUM(monthly_revenue) AS daily_revenue  
FROM netflix_analytics.subscriptions
```

GROUP BY country, device;

User-Based Analysis within Country and Device

```
SELECT user_id, subscription_type, monthly_revenue, join_date
FROM netflix_analytics.subscriptions
WHERE country = 'USA' AND device = 'Mobile'
ORDER BY user_id;
```

Tracking User Subscription Changes Over Time

```
SELECT user_id, subscription_type, monthly_revenue, join_date
FROM netflix_analytics.subscriptions
WHERE country = 'USA' AND device = 'Mobile'
ORDER BY user_id DESC;
```

Few Records for the analytics

[illegible]

```
INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('Australia', 'Mobile', uuid(), 'Premium', 16.99, '2023-04-22', '2023-12-01', 29, 'F', 12);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('USA', 'Tablet', uuid(), 'Premium', 16.99, '2022-10-30', '2023-11-25', 33, 'M', 18);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('Canada', 'Mobile', uuid(), 'Standard', 12.99, '2023-02-07', '2023-11-26', 36, 'M', 12);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('UK', 'Smart TV', uuid(), 'Premium', 16.99, '2023-03-09', '2023-11-15', 44, 'F', 12);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('Australia', 'Laptop', uuid(), 'Basic', 8.99, '2021-12-19', '2023-11-25', 31, 'M', 36);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('India', 'Smart TV', uuid(), 'Basic', 8.99, '2023-05-20', '2023-11-29', 28, 'M', 6);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('USA', 'Tablet', uuid(), 'Standard', 12.99, '2023-08-15', '2023-12-01', 26, 'F', 12);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('Canada', 'Laptop', uuid(), 'Premium', 16.99, '2021-09-13', '2023-12-02', 45, 'M', 24);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('UK', 'Tablet', uuid(), 'Standard', 12.99, '2023-03-03', '2023-12-05', 33, 'F', 18);

INSERT INTO netflix_analytics.subscriptions (country, device, user_id, subscription_type, monthly_revenue, join_date, last_payment_date, age, gender, plan_duration) VALUES ('USA', 'Smart TV', uuid(), 'Basic', 8.99, '2023-04-09', '2023-11-26', 22, 'M', 6);
```