

Fetch Rewards Data Warehouse Model

1. Overview of Data Structure

The Fetch Rewards dataset consists of three key entities: Users, Receipts, and Brands. These datasets contain unstructured JSON data that require transformation into a structured relational model for efficient querying and analysis in a data warehouse.

2. Structured Relational Data Model

Users Table

Column	Data Type	Description
user_id	STRING	Unique identifier for the user
active	BOOLEAN	User account status (active/inactive)
created_date	TIMESTAMP	Timestamp when user was created
last_login	TIMESTAMP	Last recorded login of the user
state	STRING	State abbreviation (e.g., WI)
role	STRING	User role (always 'consumer')
sign_up_source	STRING	Source where the user signed up

Primary Key: user_id
Relationships: One-to-Many with Receipts

Receipts Table

Column	Data Type	Description
receipt_id	STRING	Unique identifier for each receipt
user_id	STRING	Foreign key referencing Users

purchase_date	TIMESTAMP	Date of purchase
total_spent	FLOAT	Total amount spent on the receipt
rewards_status	STRING	Status of the receipt (Accepted/Rejected)
points_earned	FLOAT	Number of points earned
item_count	INTEGER	Total items purchased

Primary Key: receipt_id

Foreign Key: user_id references Users(user_id)

Relationships: One-to-Many with Receipt_Items

Receipt_Items Table (derived from nested JSON)

Column	Data Type	Description
item_id	STRING	Unique identifier for each item
receipt_id	STRING	Foreign key referencing Receipts
barcode	STRING	Barcode of the purchased item
description	STRING	Item description
final_price	FLOAT	Final price paid
item_price	FLOAT	Original item price
quantity_purchased	INTEGER	Number of units purchased

Primary Key: item_id

Foreign Key: receipt_id references Receipts(receipt_id)

Relationships: Many-to-One with Brands

Brands Table

Column	Data Type	Description
--------	-----------	-------------

brand_id	STRING	Unique identifier for the brand
barcode	STRING	Barcode linked to items
brand_code	STRING	Unique brand code
category	STRING	Category of the product
name	STRING	Brand name
top_brand	BOOLEAN	Whether the brand is a top brand

Primary Key: brand_id

Foreign Key: barcode references Receipt_Items(barcode)

3. JSON Structure Before and After Transformation

Original JSON Structure

```
{
  "users": [
    {
      "_id": { "$oid": "5ff1e194b6a9d73a3a9f1052" },
      "active": true,
      "createdDate": { "$date": 1609687444800 },
      "lastLogin": { "$date": 1609687537858 },
      "role": "consumer",
      "signUpSource": "Email",
      "state": "WI"
    }
  ]
}
```

Transformed JSON Structure

```
{
  "users": [
    {
      "user_id": "5ff1e194b6a9d73a3a9f1052",
      "active": true,
      "created_date": "2021-01-03T15:24:04.800Z",

```

```
"last_login": "2021-01-03T15:25:37.858Z",
"role": "consumer",
"sign_up_source": "Email",
"state": "WI"
}
]
}
```

4. Python Code for Data Processing

```
import pandas as pd
import json
from datetime import datetime

def transform_users(data):
    return [
        {
            'user_id': record['_id']['$oid'],
            'active': record.get('active', False),
            'created_date': datetime.utcfromtimestamp(record['createdDate']['$date'] /
1000).isoformat() + 'Z',
            'last_login': datetime.utcfromtimestamp(record['lastLogin']['$date'] / 1000).isoformat() +
'Z',
            'role': record.get('role'),
            'sign_up_source': record.get('signUpSource'),
            'state': record.get('state')
        }
        for record in data['users']
    ]
```

5. Data Transformation Process

1. Extract:

- Read JSON files using Python/Pandas to handle nested structures.

2. Transform:

- Convert timestamps to standard formats.
- Extract nested receipt items into separate tables.
- Handle missing values and data type conversions.

3. Load:

- Store structured data in a relational database like PostgreSQL or BigQuery.

6. Conclusion

This structured relational data model improves the ability to query and analyze Fetch Rewards data efficiently. The model facilitates easy tracking of user purchases, receipt validations, and brand performance.