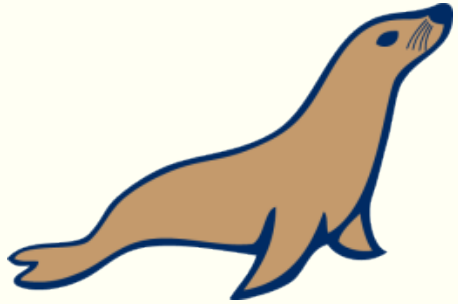
ADVANCED DATABASE MANAGEMENT SYSTEMS PROJECT

MARIADB



By:
Priyanka Prabakarrao
Rithika Kandimalla
Sai Surya Prakash Tamminedi
Yesha Desai

WALMART RETAIL DATA ANALYSIS USING MARIADB

MariaDB

- MariaDB is an open-source relational database management system.
- A fork of MySQL created by original MySQL developers.
- Designed as a drop-in replacement for MySQL, maintaining compatibility.
- Widely used across a variety of applications and industries.

Key features of MariaDB

- Open Source and Licensing: Released under the GNU General Public License (GPL).
- Compatibility with MySQL: Compatible with MySQL's APIs and command syntax.
- Performance Optimization: Focus on high performance and query optimization.
- Storage Engines: Supports multiple storage engines for different use cases.
- Security and Replication: Enhanced security features and advanced replication options.

Functions of MariaDB

- Data Storage: Stores data in structured tables, making it easy to organize and retrieve information.
- Data Retrieval: Enables the retrieval of data through SQL queries.
- Data Manipulation: Supports CRUD operations (Create, Read, Update, Delete).
- Data Security: Provides features like user account management and encryption.
- Replication: Supports data replication for high availability and fault tolerance.

How is MariaDB better than other Databases

Database	Feature	Advantage of MariaDB
Oracle DBMS	Closed-source, proprietary	MariaDB is open-source and free to use.
MySQL	Fork of MySQL, compatible with MySQL applications	MariaDB has additional features, such as Galera Cluster.
PostgreSQL	Open-source and high performance	MariaDB is generally considered to be easier to use.
SQLite	Embedded RDBMS	MariaDB is better suited for larger applications.
IBM Db2 on Cloud	Cloud-based version of IBM Db2	MariaDB is more cost-effective.
Microsoft Azure SQL	Cloud-based version of Microsoft SQL Server	MariaDB is more cost-effective.
Amazon Redshift	Cloud-based data warehouse	MariaDB is designed for transactional applications, while Amazon Redshift is optimized for analytical workloads and data warehousing. .
Redis	In-memory data store	MariaDB can also be used for caching and messaging.
Google BigQuery	Cloud-based data warehouse	MariaDB is more cost-effective
Google AlloyDB for PostgreSQL	Cloud-based version of PostgreSQL	MariaDB is more cost-effective
Amazon Aurora	Cloud-based relational database	MariaDB is more cost-effective

MariaDB vs MySQL

Feature	MariaDB	MySQL
Performance	Faster	Slower
Scalability	More scalable	Less scalable
Security	More secure	Less secure
Community Support	Larger and more active	Smaller and less active
Licensing	Fully open-source	Dual licensing model
Ease of Upgrading	Backward compatible	Not backward compatible

About the Dataset: Walmart Sales Forecast

Aim:
To assess the distribution of store sizes across different store types ('A', 'B', 'C') within Walmart and to understand the influence of markdowns on sales during holiday and non-holiday periods, enabling data-driven decisions to optimize store sizes and markdown strategies for increased profitability.

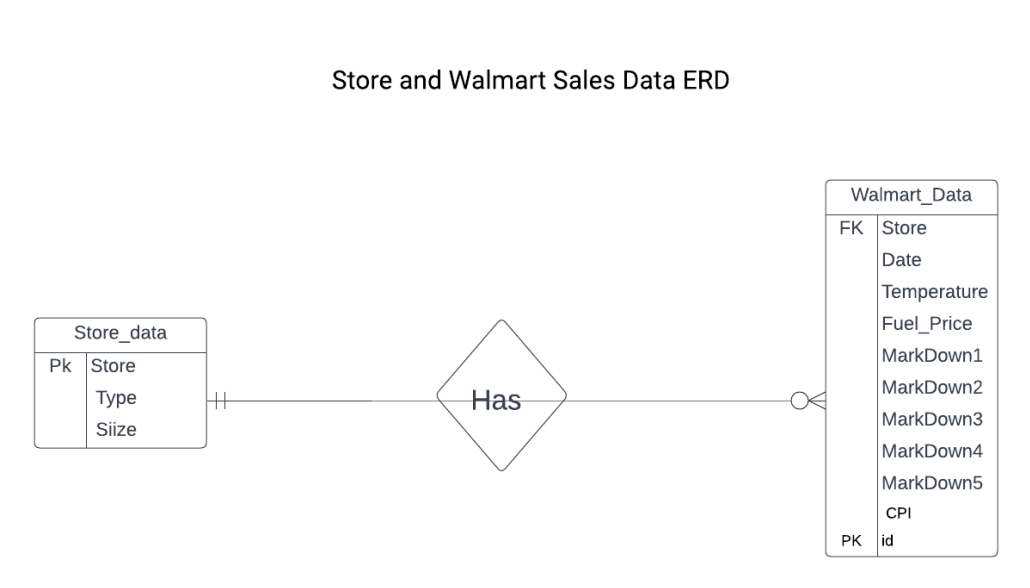
Walmart Dataset:
Store - Store number
Date - Week
Temperature - Average temperature in the region
Fuel_Price - Cost of fuel in the region
MarkDown's - Anonymized data related to promotional markdowns that Walmart is running.
CPI - The consumer price index
Unemployment - The unemployment rate
IsHoliday - Whether the week is a special holiday week

Store Dataset
Store - Stores numbered from 1 to 45
Type - Store type has been provided, there are 3 types — A, B and C.
Size - Stores size has provided

Transactional Nature of the product

- The database is designed for transactional purposes, handling data such as sales, store information, and markdowns. It supports frequent data updates, including inserts, updates, and deletes. The use of primary keys, foreign keys, and data validation checks ensures data integrity and consistency. The insertion of new records indicates that the database is used to record day-to-day operations.
- This database is designed for transactional data and is well-suited for handling the day-to-day operations of a business.

Entity and Relationship Descriptions



Cardinality: Many-to-One (M:1)

- Each record in the ‘Walmart_data’ table is associated with one and only one store from the ‘Store_data’ table.
- Each store in the ‘Store_data’ table can be associated with multiple records in the ‘Walmart_data’ table.

Modality: Mandatory from the ‘Walmart_data’

- Every record in the ‘Walmart_data’ table must be associated with a store in the ‘Store_data’ table. There can be no "orphaned" records in ‘Walmart_data’.

CRUD Operations

CREATE

Store_data Table

Store(PK), Type, Size

Walmart_data

Table

id(PK),Store(FK),Date,Temperature,Fuel_Price,Markdown1,Markdown2,Markdown3,Markdown4,Markdown5,CPI,Unemployment, IsHoliday

READ

SQL INSERT & LOAD DATA LOCAL INFILE are two methods to import data into MariaDB

•SQL INSERT method for few rows

•Two tables are populated using the LOAD DATA LOCAL INFILE

UPDATE

Updated Type for a store and Temperature for a Particular Date using SQL UPDATE Method

DELETE

Deleted few duplicate records in both the tables using DELETE method

Business Question

1. **What is the total store size for each store type (e.g., 'A', 'B','C')**

```sql
-- total store size for each store type
SELECT
    Type,
    SUM(Size) AS TotalStoreSize
FROM store_data
GROUP BY Type;
```

store_data (3r × 2c)

Type	TotalStoreSize
A	3,899,450
B	1,760,242
C	243,250

2. **What are the total markdown amounts for all stores during holiday vs non-holiday periods.**

```sql
SELECT
   Store,IsHoliday,
    SUM(
        COALESCE(CAST(MarkDown1 AS DECIMAL(10, 2)), 0) +
        COALESCE(CAST(MarkDown2 AS DECIMAL(10, 2)), 0) +
        COALESCE(CAST(MarkDown3 AS DECIMAL(10, 2)), 0) +
        COALESCE(CAST(MarkDown4 AS DECIMAL(10, 2)), 0) +
        COALESCE(CAST(MarkDown5 AS DECIMAL(10, 2)), 0)
    ) AS total_markdown
FROM walmart_data
GROUP BY IsHoliday,Store;
```

walmart_data (91r × 3c)

	IsHoliday	total_markdown
1	FALSE	1,539,571.19
2	FALSE	1,882,503.7
3	FALSE	453,638.19
4	FALSE	2,043,528.02
5	FALSE	542,019.39

Thank you

References

- KDD Nuggets Dataset Link: https://www.kaggle.com/datasets/aslanahmedov/walmart-sales-forecast?select=train.csv
- https://www.kaggle.com/datasets/aslanahmedov/walmart-sales-forecast?select=stores.csv
- https://mariadb.org/
- https://mariadb.org/documentation/
- Teams Recording link: Team Recording