Normalization and Relationships

data_pt_rmt_jun21

Normalized Data in our Database

account_id	district_id	frequency	date
3328	1	POPLATEK MESICNE	970317
5120	1	POPLATEK MESICNE	930621
9472	1	POPLATEK TYDNE	930706
2305	1	POPLATEK MESICNE	931211
3585	2	POPLATEK MESICNE	960623
2	2	POPLATEK MESICNE	930226
2050	2	POPLATEK MESICNE	931022
3842	2	POPLATEK MESICNE	970312
1539	2	POPLATEK PO OBRATU	930103

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107
2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674
3	Beroun	central Bohemia	75232	55	26	4	1	5	41.7	8980	1.95	2.21	111	2824	2813
4	Kladno	central Bohemia	149893	63	29	6	2	6	67.4	9753	4.64	5.05	109	5244	5892
5	Kolin	central Bohemia	95616	65	30	4	1	6	51.4	9307	3.85	4.43	118	2616	3040
6	Kutna Hora	central Bohemia	77963	60	23	4	2	4	51.5	8546	2.95	4.02	126	2640	3120

It the tables were not Normalized
in our Database

account_id	district_id	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	frequency	date
3328	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK MESICNE	970317
5120	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK MESICNE	930621
9472	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK TYDNE	930706
2305	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK MESICNE	931211
3585	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	960623
2	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	930226
2050	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	931022
3842	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	970312
1539	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK PO OBRATU	930103

Advantages of Normalization

1) A smaller database can be maintained as normalization eliminates duplicate data.

2) As databases have smaller size, the operations and queries becomes faster and shorter thereby improving response time and speed.

3) Narrower tables are possible as normalized tables will be fine-tuned and will have lesser columns which allows for more data records per page.

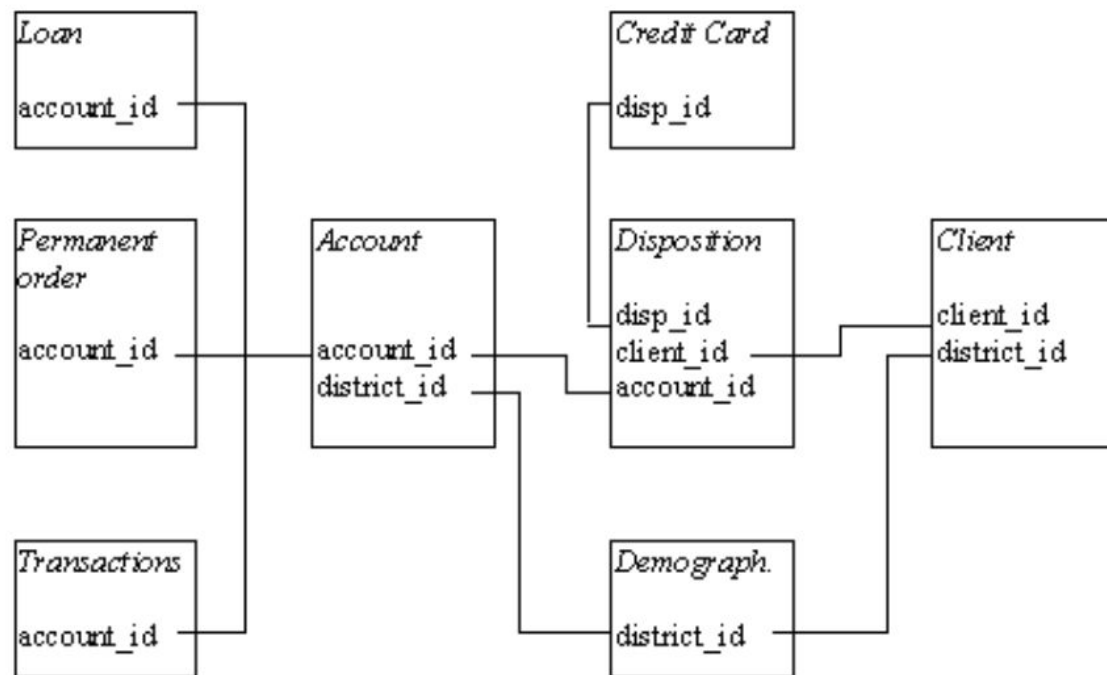4) Easier to detect and contain anomalies.

Anomalies

Insert Anomaly: An insertion anomaly occurs when we want to add a new record to a relation but not all the data are available to allow us to add the new record. (Ex.: bank account with no transactions yet)

Update Anomaly: An update anomaly occurs when an update of an attribute value needs to be made at multiple places.

Delete Anomaly: A deletion anomaly occurs when the deletion of a record leads to a loss of data.

account_id	district_id	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	frequency	date
3328	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK MESICNE	970317
5120	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK MESICNE	930621
9472	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK TYDNE	930706
2305	1	Hl.m. Praha	Prague	1204953	0	0	0	1	1	100	12541	0.29	0.43	167	85677	99107	POPLATEK MESICNE	931211
3585	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	960623
2	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	930226
2050	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	931022
3842	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK MESICNE	970312
1539	2	Benesov	central Bohemia	88884	80	26	6	2	5	46.7	8507	1.67	1.85	132	2159	2674	POPLATEK PO OBRATU	930103

PRIMARY KEY

- The primary key is a **UNIQUE identifier** for each record in a database's table (ie: client_id)

- It must be one column of the table which contains UNIQUE values.

- It can't be NULL.

- Is automatically incremented each time a new record is inserted.

FOREIGN KEY

- They are used to establish a "link" between two tables.

Customers

Column name	Characteristics
customer_id	PRIMARY KEY
first_name	
last_name	

Orders

Column name	Characteristics
order_id	PRIMARY KEY
order_date	
customer_id	FOREIGN KEY

CONSTRAINTS

Data Integrity: Set of rules in order to ensure the accuracy and consistency of data over it entire lifecycle. They can be:

- **Entity Integrity:** Set of rules to ensure that every table MUST have a PRIMARY KEY to make sure that every record is unique and not NULL

- **Referential Integrity:** The foreign keys can't take values which are not contained in the primary keys of the table which is referenced from.

- **Domain Integrity:** All the values inside a column MUST be of the same type.

https://www.w3resource.com/mysql/creating-table-advance/constraint.php

ENTITY INTEGRITY CONSTRAINTS COUNTER EXAMPLE

Name	Last_name	Course	Country
Philippo	Caneto	DA	IT
~~Philippo~~	~~Caneto~~	~~WD~~	~~IT~~
Emet	Johnson	CY	PO

EXAMPLE OF REFERENTIAL ENTITY

Countries

country_id	country_name
US	United States
MEX	Mexico
AR	Argentina
FR	France

country_id is PRIMARY KEY

Doctors

doctor	country_id
Joseph	US
Paul	US
~~Melanie~~	~~TK~~
Cynthia	FR

country_id is FOREIGN KEY

DOMAIN INTEGRITY CONSTRAINTS EXAMPLE

State
Arkansas
~~FL~~
Maine
California
North Dakota

Command Types

DDL – Data Definition Language

DQL – Data Query Language

DML – Data Manipulation Language

DCL – Data Control Language

TCL - Transaction Control Language

https://www.geeksforgeeks.org/sql-ddl-dql-dml-dcl-tcl-commands/

IRON
HACK

THANKS !