

# DAY 4

# Database Design

Conceptual  
Model

Entity Relationship  
Diagram (ERD)

Database  
Schema



**movies**

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...

**movie\_actor**

movie_id	actor_id
101	50
101	51
102	52
102	53

**actors**

actor_id	name	...
50	Yash	...
51	Sanjay Dutt	...
52	Benedict C	...
53	Elizabeth O	...

**why do we  
need multiple  
tables?**

## movies

movie_id	title	actor_1	birth_year_1	actor_2	birth_year_2
101	KGF2	Yash	1986	Sanjay Dutt	1989
102	Doctor Strange Multiverse	Benedict C	1976	Elizabeth Olsen	1989

Less flexible

## movies

movie_id	title	actor	birth_year
101	KGF2	Yash	1986
101	KGF2	Sanjay Dutt	1959
102	Doctor Strange Multiverse	Benedict Cumberbatch	1976
102	Doctor Strange Multiverse	Elizabeth Olsen	1989

Data duplication

Poor data integrity



Data  
Integrity??

## movies

movie_id	title	actor	birth_year
101	KGF2	Yash	1986
101	KGF99	Sanjay Dutt	1959
102	Doctor Strange Multiverse	Benedict Cumberbatch	1976
102	Doctor Strange Multiverse	Elizabeth Olsen	1989

Data integrity is accuracy and consistency of data over its life cycle



Less  
duplication

Better data  
integrity

Flexible  
design

## movies

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...

## actors

actor_id	name	...
50	Yash	...
51	Sanjay Dutt	...
52	Benedict C	...
53	Elizabeth O	...

**movies**

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...

**movie\_actor**

movie_id	actor_id
101	50
101	51
102	52
102	53

**actors**

actor_id	name	...
50	Yash	...
51	Sanjay Dutt	...
52	Benedict C	...
53	Elizabeth O	...

⌚ Link table

**Normalization:** Is a process of organizing database so that we can avoid duplication and increase data integrity

1 NF

2 NF

3 NF



Budget: 1 Billion INR

Revenue: 12.5 Billion INR



Budget: 200 Million USD

Revenue: 954 Million USD

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...



Budget: 1 Billion INR

Revenue: 12.5 Billion INR



Budget: 200 Million USD

Revenue: 954 Million USD

movie_id	title	...	budget	revenue
101	KGF2	...	1 billion INR	12.5 billion INR
102	Doctor Strange Multiverse	...	200 million USD	954 million USD
220	Bahubali 3	...		
221	THOR 10	...		

**movies**

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...
220	Bahubali 3	...
221	THOR 10	...

**financials**

movie_id	budget	revenue	...
101	1 billion INR	12.5 billion INR	...
102	200 million USD	954 million USD	...

## movies

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...
220	Bahubali 3	...
221	THOR 10	...

## languages

language
Hindi
Telugu
Kannada
English

## movies

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...
220	Bahubali 3	...
221	THOR 10	...

## movie\_actor

movie_id	actor_id
101	50
101	51
102	52
102	53

## actors

actor_id	name	...
50	Yash	...
51	Sanjay Dutt	...
52	Benedict C	...
53	Elizabeth O	...

## financials

movie_id	budget	revenue	...
101	1 billion INR	12.5 billion INR	...
102	200 million USD	954 million USD	...

## languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

# Takeaways

- Database design is a critical step which consists of 3 stages: **Conceptual Model, Entity Relationship Diagram (ERD) and Database Schema**
- **Data Integrity** is the measure of consistency and accuracy of data over its life cycle
- ‘**Link table**’ is a term used to describe a table that acts as a link between two tables
- Normalization is a process of organizing a database to **avoid duplication and improving data integrity**

# Database Design

Conceptual  
Model

Entity Relationship  
Diagram (ERD)

Database  
Schema



movies

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...
220	Bahubali 3	...
221	THOR 10	...

movie\_actor

movie_id	actor_id
101	50
101	51
102	52
102	53

actors

actor_id	name	...
50	Yash	...
51	Sanjay Dutt	...
52	Benedict C	...
53	Elizabeth O	...

financials

movie_id	budget	revenue	...
101	1 billion INR	12.5 billion INR	...
102	200 million USD	954 million USD	...

languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

# Takeaways

- You can create an entity relationship diagram (**ERD**) in **MySQL**
- **Numeric, String and Date** are the major data types which have further subtypes

# SQL Data Types

Numeric

Text

Date, Time

Others

# Numeric

## Whole Numbers

5    -89    2501

## Numbers with decimal point

5.2    -2.1    988.331

## Integers

- TINYINT
- SMALLINT
- INT
- BIGINT

## Floating Point Types

- FLOAT
- DOUBLE
- DECIMAL

Type	Storage (Bytes)	Signed		Unsigned	
		Min	Max	Min	Max
TINYINT	1	-128	127	0	255
SMALLINT	2	-32768	32767	0	65535
MEDIUMINT	3	-8388608	8388607	0	16777215
INT	4	-2147483648	2147483647	0	4294967295
BIGINT	8	$-2^{63}$	$2^{63} - 1$	0	$2^{64} - 1$

## Floating Point Types

**FLOAT**

9.12345678912345

9.12346

4 bytes

**DOUBLE**

9.12345678912345

8 bytes



**179.52 cm –  
Low Precision  
is okay**

**25.6775464-**  
**High Precision**  
**required**



## Floating Point Types

<b>FLOAT</b>	9.12346	4 bytes
9.12345678912345		
<b>DOUBLE</b>	9.12345678912345	8 bytes
<b>DECIMAL</b>	DECIMAL(5, 3) 89.124	
	DECIMAL(3, 1) 10.1 2.9	

# Floating Point Types

**FLOAT**

**DOUBLE**



Approximal Value

Scientific  
Computation

**DECIMAL**

Exact Value

Monetary,  
business use  
cases

# Numeric

## Whole Numbers

5    -89    2501

## Numbers with decimal point

5.2    -2.1    988.331

## Integers

- TINYINT
- SMALLINT
- INT
- BIGINT

## Floating Point Types

- FLOAT
- DOUBLE
- DECIMAL

# Takeaways

- **Integers** and **Floating** points are the major classifications of numerical data types
- Each of these classifications has subtypes such as **TINYINT**, **SMALLINT**, **FLOAT**, **DECIMAL** etc.
- Different data types occupy different **storage space**
- Knowing the data type and its storage space will **enable you to design an efficient database**
- Float and double are the data types you can use for **storing approximate values** such as scientific calculations
- Decimal type is used to store accurate values such as **financial transactions**

# SQL Data Types

Numeric

String

Date, Time

Others

# String

Fixed Length

Variable Length

**CHAR(3)**

“USD”

“INR”

“A ”

**VARCHAR(150)**

“KGF 2”

“The Shawshank Redemption”

# String

Fixed Length

Variable Length

ENUM

ENUM('small', 'medium', 'large', 'x-large')

ENUM('billions', 'millions', 'thousands')

financials

movie_id	budget	revenue
101	1 billion INR	12.5 billion INR
102	200 million USD	954 million USD

# Takeaways

- **Fixed and variable length strings** are two major types of string
- **ENUM** is a string object data type used to specify a fixed number of options for column values
- It is a best practice to define the data type for all the columns
- One can add rows manually to the table in the SQL editor
- **BLOB** is a data type that will enable you to store images as binary text in a table

# SQL Data Types

Numeric

String

Date, Time

Others

# **stock\_transactions**

<b>date_time</b>	<b>customer</b>	<b>stock</b>	<b>qty</b>	<b>order_type</b>
2020-07-01 09:30:59	Thor Hathodawala	TSLA	100	BUY
2020-07-01 12:30:07	Lauki Lal	RIL	20	SELL
2020-07-02 10:12:34	Bruce Hariyali	MSFT	200	BUY



**DATETIME**

**yyyy mm dd hh:mm:ss**

# daily\_sales\_summary

date	customer	product	qty
2020-07-02	Croma	iPhone 14	200
2020-07-02	Croma	LG TV 2300	159
2020-07-01	Ebay	Logitech Mouse	700



**DATE**

**yyyy mm dd**



**“yyyy mm dd”** is one kind of date format.  
Format depends upon the user or country

#codebasicsday

## ANNOUNCING

5<sup>TH</sup> SEP'22 as CODEBASICS DAY

Event from 17:00pm to 21:30pm IST

- 1 17:30 IST | The Teacher Who Doesn't Want to Retire
- 2 18:00 IST | Interview with **Nitin Aggarwal**  
Head of Cloud AI Solutions, Google India
- 3 19:00 IST | A New Initiative to Help Learners
- 4 19:30 IST | Live Fun Data Quiz
- 5 20:30 IST | Teachers Who Influenced Us
- 6 21:00 IST | New Series: Project Portfolio for Beginners

PARTICIPATE, LEARN AND WIN EXCITING PRIZES!



codebasics.io

## event\_schedule

start_time	end_time	title	category
17:30:00	17:59:00	The Teacher Who Doesn't Want to Retire	Inspirational Video
18:00:00	18:59:00	Interview With Nitin Aggarwal	Interview
19:30:00	20:00:00	Live Fun Data Quiz	Live Quiz

TIME

hh:mm:ss

# movies

movie_id	title	...	release_year
101	KGF2	...	2022
102	Doctor Strange	...	2022
104	Thor: Ragnarok	...	2017



YEAR

yyyy

Type	Storage (Bytes)	Min	Max
TINYINT	1	0	255
SMALLINT	2	0	65535
YEAR	1	1901	2155

# actors

actor_id	name	birth_year	last_update
50	Yash	1986	2021-02-15 04:34:33
51	Sanjay Dutt	1959	2022-04-07 12:00:45
52	Benedict Cumberbatch	1976	2022-05-10 15:23:58

Type	Min	Max
DATETIME	1000-01-01 00:00:00	9999-12-31 23:59:59
TIMESTAMP	1970-01-01 00:00:01	2038-01-19 03:14:07

**TIMESTAMP**

**yyyy mm dd hh:mm:ss**

## Date, Time

<b>DATETIME, TIMESTAMP</b>	<b>yyyy mm dd hh:mm:ss</b>	2020-07-01 12:30:07
<b>DATE</b>	<b>yyyy mm dd</b>	2020-07-01
<b>TIME</b>	<b>hh:mm:ss</b>	12:30:07
<b>YEAR</b>	<b>yyyy</b>	2020

# Takeaways

- **DATE, YEAR AND DATETIME** are the major date types under Date, Time category
- Defining correct data types will enable you to prevent errors in the database
- You can use '**TIMESTAMP**' data type to automatically enter the current timestamp at which the record is created

# SQL Data Types

Numeric

String

Date, Time

Others

JSON

Spatial



Laptops &amp; Accessories

TV &amp; Home Entertainment

Audio

Cameras

Computer Peripherals

Smart Technology

Musical Instruments

## Electronics &amp; accessories



PAY ON DELIVERY



EASY RETURNS



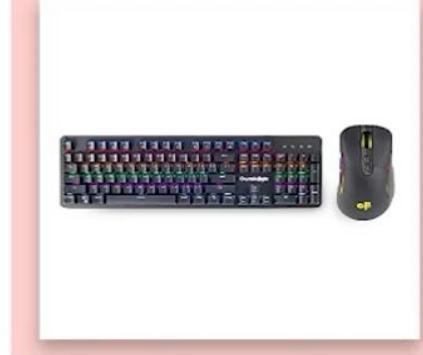
FAST DELIVERY

100%  
ORIGINAL

Headphones



Fitness trackers &amp; smartwatches



Computer accessories



Pen drives



Bought by customers this season



## Product information

### Technical Details

Brand	LG
Manufacturer	LG, LG Electronics India Pvt. Ltd.   Plot No. A5, MIDC Ranjangaon, Tal. Shirur, Pune - 412220   serviceindia@lge.com   TEL (TOLL FREE) : 1800 315 9999 / 1800 180 9999   SMS : 5757554
Model	22MP68VQ-P.BTR
Model Name	22MP68VQ
Product Dimensions	9.7 x 49.1 x 29.3 cm; 2.8 Kilograms
Item model number	22MP68VQ-P.BTR
Hardware Interface	HDMI
Response Time	5 Milliseconds
Resolution	FHD 1080p
Special Features	Anti Glare Screen, Adaptive Sync, Tilt Adjustment, Flicker-Free
Mounting Hardware	Power Cable, HDMI Cable
Number Of Items	1
Display Technology	LCD
Standing screen display size	54.6 Centimetres
Display Type	LCD
Viewing Angle	178 Degrees
Image Aspect Ratio	16:9
Image Resolution	250
Image Contrast Ratio	1000:1

Image source: amazon.in

### Additional Information

ASIN	B09GWB35XV
Customer Reviews	 8,150 ratings 4.4 out of 5 stars
Best Sellers Rank	#7,728 in Electronics (See Top 100 in Electronics) #41 in Monitors
Date First Available	22 September 2021
Packer	LG Electronics India Pvt. Ltd.   Plot No. A5, MIDC Ranjangaon, Tal. Shirur, Pune - 412220   serviceindia@lge.com   TEL (TOLL FREE) : 1800 315 9999 / 1800 180 9999   SMS : 5757554
Importer	LG Electronics India Pvt. Ltd.   Plot No. A5, MIDC Ranjangaon, Tal. Shirur, Pune - 412220   serviceindia@lge.com   TEL (TOLL FREE) : 1800 315 9999 / 1800 180 9999   SMS : 5757554
Item Dimensions LxWxH	9.7 x 49.1 x 29.3 Centimeters
Net Quantity	1 Piece
Generic Name	PC Monitor

### Feedback

Would you like to [tell us about a lower price?](#)

# products

product_id	name	brand	features
115	Sandisk SD Card	Sandisk	{ "storage": 64, "color": "black"}
116	LG Monitor	LG	{"size": 27, "display_type": "LCD"}
117	Airpodes 141	boAt	{"connector": "wireless"}



# items

item_id	name	brand	category	price	weight	size
1	Regular Butter	Amul	Dairy	240	500 gram	
2	White Butter	Amul	Dairy	235	500 gram	
3	T Shirt	Adidas	Clothes	800		XL
4	T Shirt	Adidas	Clothes	1200		S

# products

product_id	name	brand	features
115	Sandisk SD Card	Sandisk	{ "storage": 64, "color": "black"}
116	LG Monitor	LG	{"size": 27, "display_type": "LCD"}
117	Airpodes 141	boAt	{"connector": "wireless"}

# items

item_id	name	brand	category	price	properties
1	Regular Butter	Amul	Dairy	240	{"weight": 500, "is_salts": 1}
2	White Butter	Amul	Dairy	235	{"weight": 400, "is_salts": 0}
3	T Shirt	Adidas	Clothes	800	{"size": "XL", "color": "Blue"}
4	T Shirt	Adidas	Clothes	1200	{"size": "S", "color": "Red"}

↑  
**JSON**

**JAVASCRIPT Object Notation**

# Takeaways

- **JSON** is a popular and efficient data type to store massive amount of data
- ‘→’ operator is used to extract a JSON object
- **SPATIAL** datatype is used to represent geospatial data types like latitude, longitude etc.

# Primary Key

---





# customers

name	email	phone	SSN	address	...
Dhaval Patel	dp@fakedomain.com	+91798912823	123123123	32 Patli Gali, Wonderland	...
Thanos Thakur	tt@fakedomain.com	+1732999111	999111333	21 Abe O road, Titan	...
Peter Pandey		+1732111222	888222444	5 forest hills, Queens, New York	...

## customers

customer_id	name	email	phone	SSN	...
150	Dhaval Patel	dp@fakedomain.com	+91798912823	123123123	...
151	Thanos Thakur	tt@fakedomain.com	+1732999111	999111333	...
152	Peter Pandey		+1732111222		...

# customers

name	email	phone	SSN	address	...
Dhaval Patel	dp@fakedomain.com	+91798912823	123123123	32 Patli Gali, Wonderland	...
Thanos Thakur	tt@fakedomain.com	+1732999111	999111333	21 Abe O road, Titan	...
Peter Pandey		+1732111222	888222444	5 forest hills, Queens, New York	...

Primary Key

natural key

## customers

customer_id	name	email	phone	SSN	...
150	Dhaval Patel	dp@fakedomain.com	+91798912823	123123123	...
151	Thanos Thakur	tt@fakedomain.com	+1732999111	999111333	...
152	Peter Pandey		+1732111222		...



**Primary Key**      surrogate key

# addresses

street	city	state	country	postal_code	...
32 Jalebi Marg	Mumbai	Maharashtra	India	1114	...
21 Abe O Street	Los Angeles	California	USA	4351	...
9 Teri Yaki Road	Kyoto	Honshu	Japan	7732	...

Primary Key

# addresses

street	city	state	country	postal_code	...
32 Jalebi Marg	Mumbai	Maharashtra	India	1114	...
21 Abe O Street	Los Angeles	California	USA	4351	...
9 Teri Yaki Road	Kyoto	Honshu	Japan	7732	...

Primary Key

composite key

natural key

## customers

customer_id	name	email	phone	SSN	...
150	Dhaval Patel	dp@fakedomain.com	+91798912823	123123123	...
151	Thanos Thakur	tt@fakedomain.com	+1732999111	999111333	...
152	Peter Pandey		+1732111222		...



Primary Key

# Takeaways

- Primary key is a **unique identifier** which cannot have any duplicates
- Primary key that already exists in database is called **natural key**
- Primary key that is generated by user artificially is called **surrogate key**
- Composite key is a primary key that is generated by **combining multiple columns**
- **Auto Increment option in schema settings** will enable you to auto assign numerical values on records incrementally

# Foreign Key

---



## movies

movie_id	title	...	language_id
101	KGF2	...	1
102	Doctor Strange Multiverse	...	5
220	Bahubali 3	...	2
221	THOR 10	...	5

## languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

## movies

movie_id	title	...	language_id
101	KGF2	...	1
102	Doctor Strange Multiverse	...	5
220	Bahubali 3	...	2
221	THOR 10	...	5
222	Cars 7	...	55

## languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

# movies

movie_id	title	...	language_id
101	KGF2	...	1
102	Doctor Strange Multiverse	...	5
220	Bahubali 3	...	2
221	THOR 10	...	5

# languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

Foreign Key

# movies

movie_id	title	...	language_id
101	KGF2	...	1
102	Doctor Strange Multiverse	...	5
220	Bahubali 3	...	2
221	THOR 10	...	5

Child table

Foreign Key

# languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

Parent table

# Takeaways

- **Foreign key column** is a way to establish relationship between two tables
- The parent table contains the primary key which is connected to the child table which contains the foreign key
- The key benefit of creating a relationship is to prevent having undesirable records in the database
- By creating a relationship you can also update or delete records automatically in the child table based on the action you perform in the parent table

# Takeaways

- ‘Forward engineer’ is the option to create a database from a data model
- ‘Reverse engineer’ is the option to create / edit a data model from database
- One can also make changes directly in the database which is a popular practice
- However in some cases **reverse engineering** is performed where it is required to understand or document the model

# Takeaways

- Backfilling is a process of adding records in bulk to a database
- MySQL provides an option to map column names between the destination table and the table from which the data is imported

# Takeaways

- **Updating the database** is a task that is typically performed by data engineers
- **INSERT, UPDATE and DELETE** are the primary database modification clauses of SQL
- '**INSERT**' is the clause used in SQL to add records
- '**VALUES**' clause will enable you to add single records or multiple records

# Takeaways

- **UPDATE, SET and WHERE** are the clauses used to update a value to an existing record or multiple records
- **DELETE and WHERE** can be used together to delete single or multiple records
- One can specify in the parameters passed to the WHERE clause to define if the action should **impact single or multiple records**