Intro to DBMS & Relational Model

Instructor

- Prateek Narang
- DTU - 2016
- B Tech
- Coding Blocks (2016-20)
- Google L4 SDE
- Coding Minutes
- Scaler

Agenda

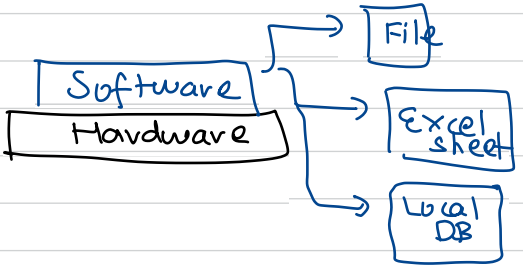
- ~~what is DB~~
- ~~Why you should learn DB~~
- ~~Scaler Curriculum~~
- ~~Types of DBMS~~
- ~~Intro to Relational DB~~
- Intro to Keys (time)

- TODO: Installation (next class)
  (MySQL)

Data

Storage solution

Contacts — Internal / Memory card / Cloud Machine Memory



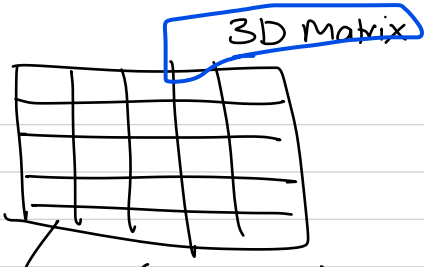
Photos

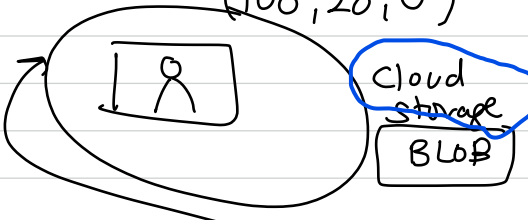
Notes
Docs
PDFs
Videos
APK Files
Expenses
Shopping list

→ Google Cloud
→ "
→ File System
→ Splitwise
→ Whatsapp

Tables

Phone	- 20,000/-
Fruits	- 500/-
⋮	





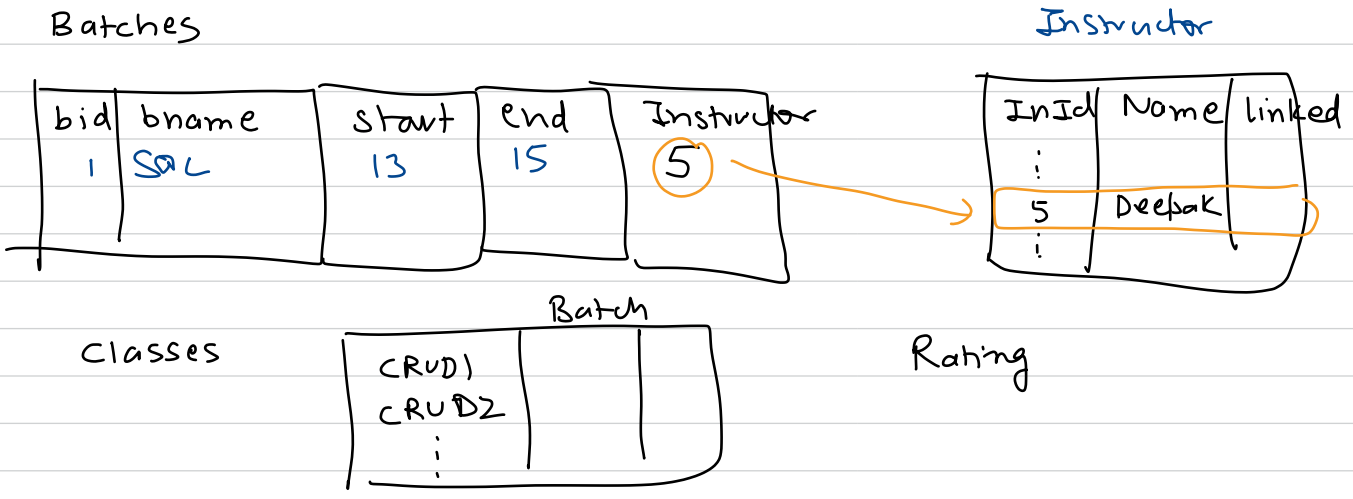
Organisation

User Id	Username	First Name	Last Name	DOB	Job	Pho	Profilepic
1	prateekcn	prateek	narang	270394	teacher	9918	URL of img.
2							
3							

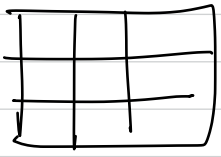
Load

Scaler

Instructor , students , Employees , Mentor , Leaderboard , ...



Programming Language



2D Array

↑

Data Structure
(RAM → volatile)

.txt

name, email, marks
prat, xyz, 80
deepak, abc, 70

File can as
~~as~~ DB

↑

Disk (SSD, HardDisk - Permanent Storage)

TODO. // Open a file
// write text to file
// close the file

Issues with files as DB

→ Comma separate values

students.csv

```
name, batch, psp, attendance, coins, rank
Naman, 1, 94, 100, 0, 1
Amit, 2, 81, 70, 400, 1
Aditya, 1, 31, 100, 100, 2
```

34 120

O(N)

1 Million Records

Marks of Aditya

→ Read line by line

→ Tokenize ,

"Aditya" , " " , " " , 31 ...

→

→ 150 Crore India

150,00,00,000 → 10^{10} / 10^{8} → atleast 100 seconds

(1) Efficiency

→ Too inefficient

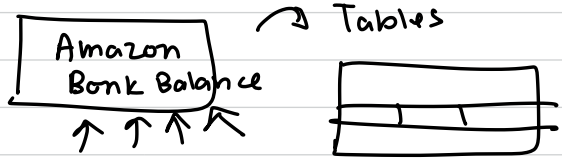
In DB, concept called "Indexing" → O(logN)

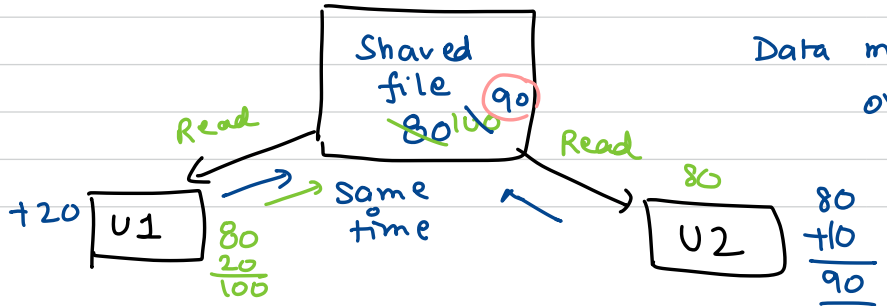
2 Data Integrity → Correctness of data

- you can put batch_id inside student for which no batch exists of that id.
- → you might put up a string in marks colo

→ Data inconsistency are common.

2 Concurrent Acess





Data may get over-written. Expected

$80 + 20 + 10 = 110$

(4) Security

→ sensitive

passwords

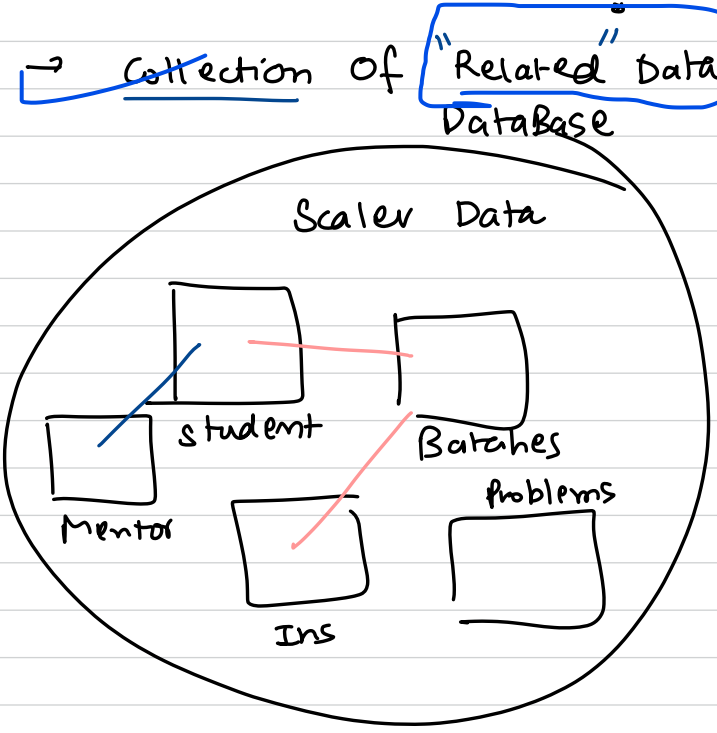
Anyone can read / write on that file.

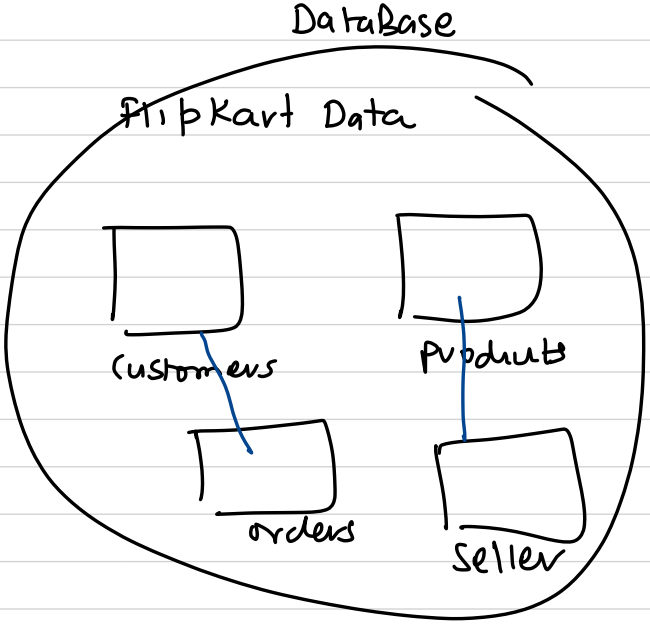
- in DB, user-level → Admin / co-admin / user.

What is Database ?

→ Collection of "Related" Data

Table → entity





DBMS (Database Management System)

↳ Software that allows to do operations

C - Create
R - Read
U - Update
D - Delete

} efficient

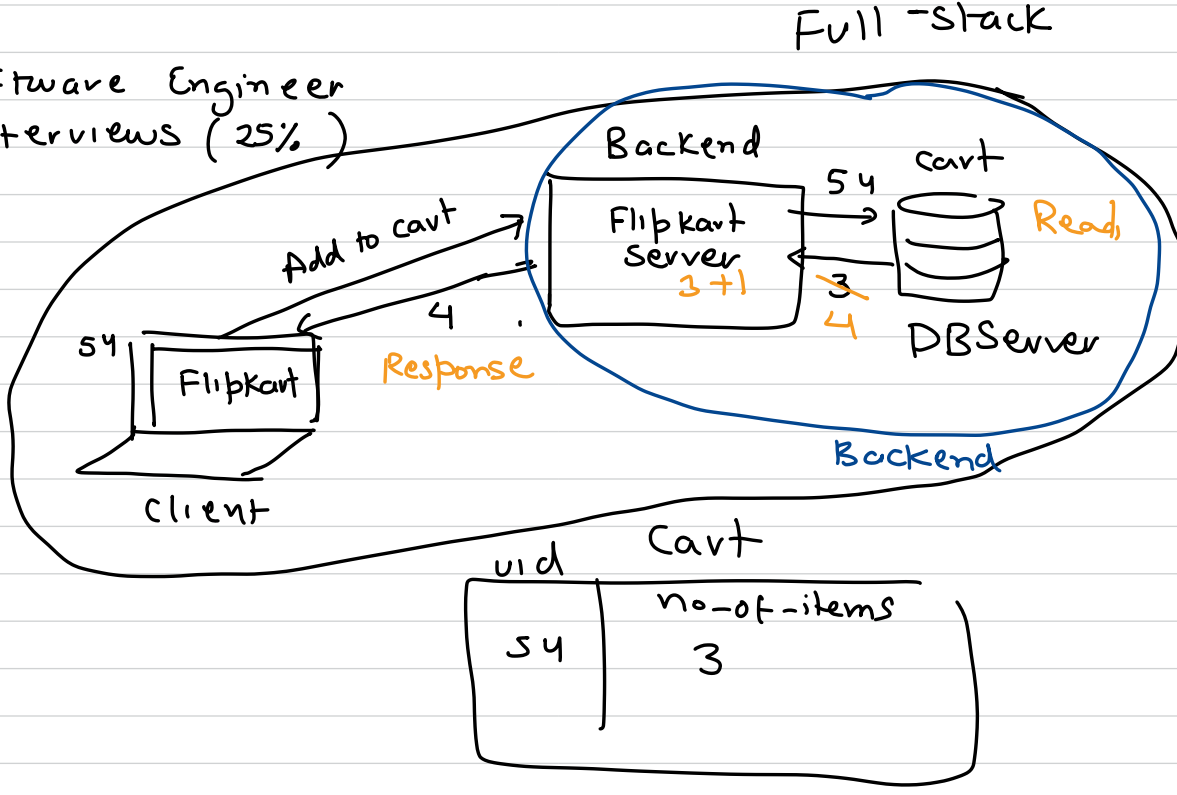
on a database along with ensuring

→ data integrity
→ security
→ concurrency

Why ?

↳ Software Engineer

↳ Interviews (25%)



uid	cart
	no-of-items
54	3

SQL Module Scaler Curriculum

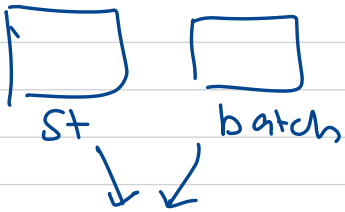
12 lectures

1. How DB work
2. SQL queries & Schema Design

HLD Module (later) after your LLD.

3. Scalability & Distributed Databases

- CRUD
- Joins
- Aggregated
- Subqueries
- Indexing (Trees)
- Misc
- Schema Design

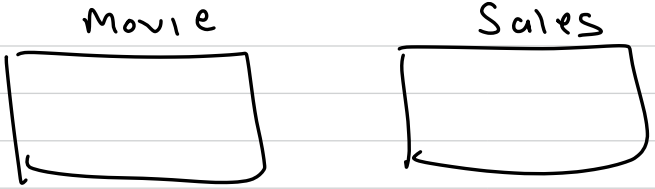


Interviews.

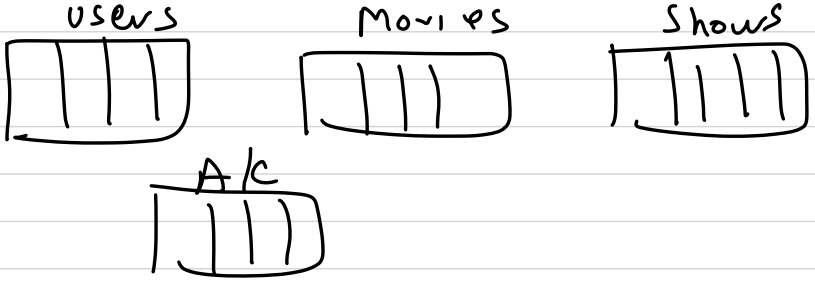
Questions.

10.25 PM

1. Write a query to find the most popular movie acc to revenue.



2. Design a DB for app like Netflix.



Types of DataBases :

Relational DB

→ Table (Rows & columns)

→ RDBMS :

- SQL
- MySQL ← widely used, free, good support
- Microsoft SQL
- PostgresSQL
- Oracle

Non - Relational DB

→ No tables

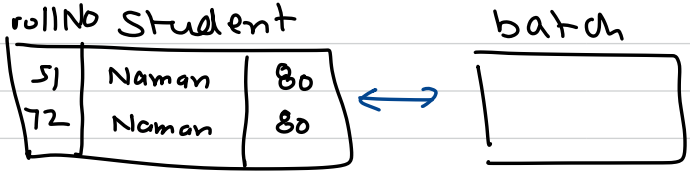
→ graph database,
K-V pairs,
: JSON

→ MongoDB
→ Redis
→ Firebase
→ Elastic Search

HLD Module

Properties of RDBMS:

1. RDBMS represent database as collection of tables with each table denoting a entity and we define relationships b/w entities



1:1
1:m
m:m
> Schema Design

2. Every Row must be unique

3. Every value in col should have same datatype.

DATA INTEGRITY

④ Value inside each (row, col) cell should be atomic

SQL standards

id	username
71	Prateek

1	iphone
2	gopro
3	insta360
4	samsung

Product

Shopping card

id	oid	items
71	5	1, 3

list not allowed in cell as it is not atomic.

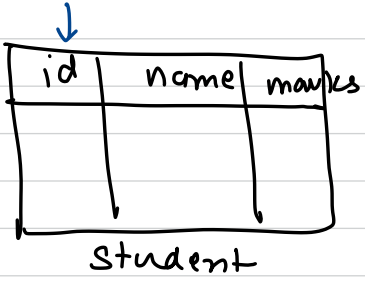
list is not atomic.

Correct way

uid	oid	items
71	5	1
71	5	3

✓ atomic

5 Col seq is not guvanteed (SQL Standard)




```
SELECT *
FROM
  student ;
```

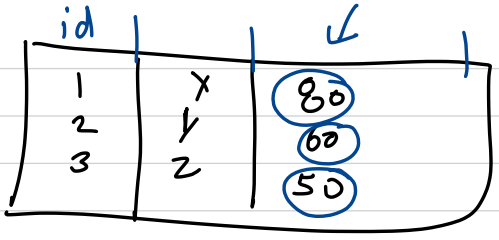
id , name , marks

0 1 2

```
SELECT      id, marks
  FROM
    student
```

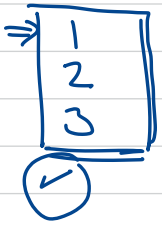
id , marks , name

6 Row seq is also not guvanteed.




```
Select *
FROM
  students
ORDER BY marks;
```

→ Default acc to Primary key



3 2 1 | 2 3 1 |

Custom Ordering =

(7) ~~(7)~~ Name of every col should be unique.

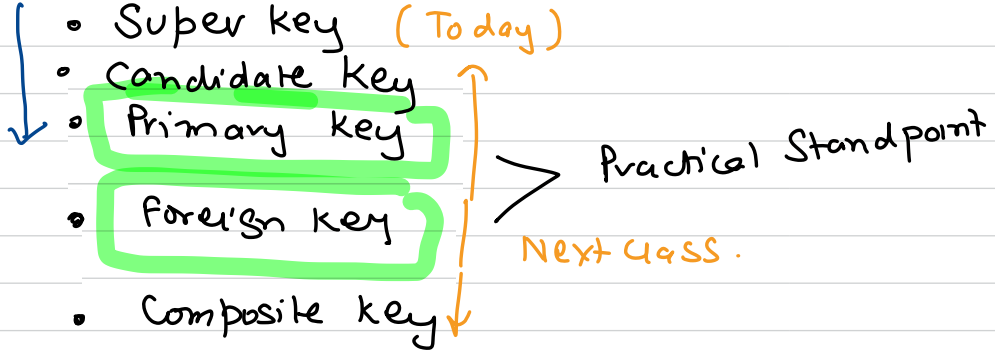
name	marks	marks
X	70	60

Ambiguity?

[whatsapp group link → PIN]

Keys in Relational Database

↳ Foundational Concept



- **Super-Key**

Student

Name	Email	Phone	Marks	Batch-id
X	abc	9918	70	100
Y	|	|	60	25
Z			50	100

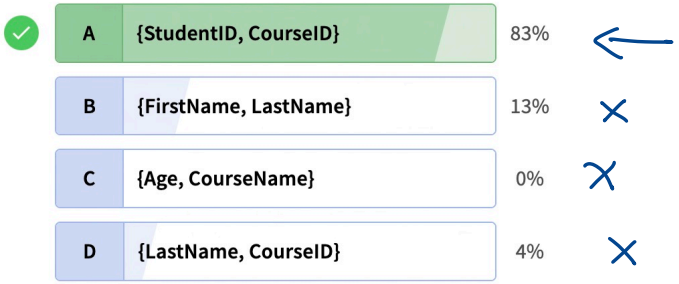
Super-keys.

- { email, batch-Id }
- { email }
- { phone No }
- { name, email }
- { name, phone No }
- { email, phone No }

Any combination of cols that can uniquely identify a Row inside a table is a SK.

Which of the following is a Super Key for the Student table?

48 users have participated

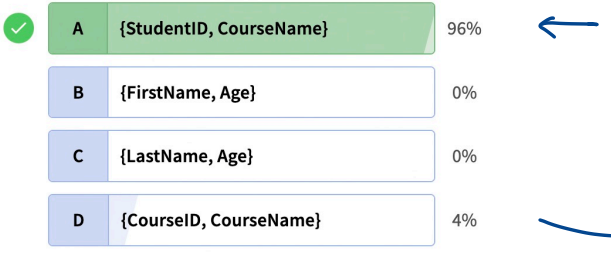


Naman Kr
Naman Kr
29, ML
29, ML

Kr, 6
Kr, 6

Which of these combinations could also be a Super Key for the Student table?

46 users have participated



7, ML
11, ML

Ram → 7, ML
Kush 7, ML

Given the uniqueness of the StudentID, which of these could be a potential Super Key for the Student table?

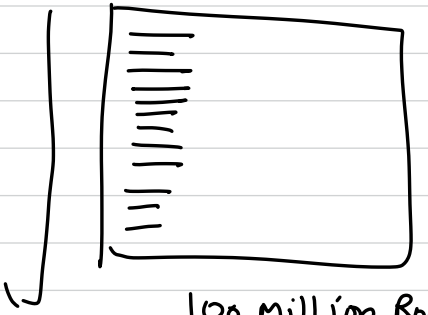
42 users have participated

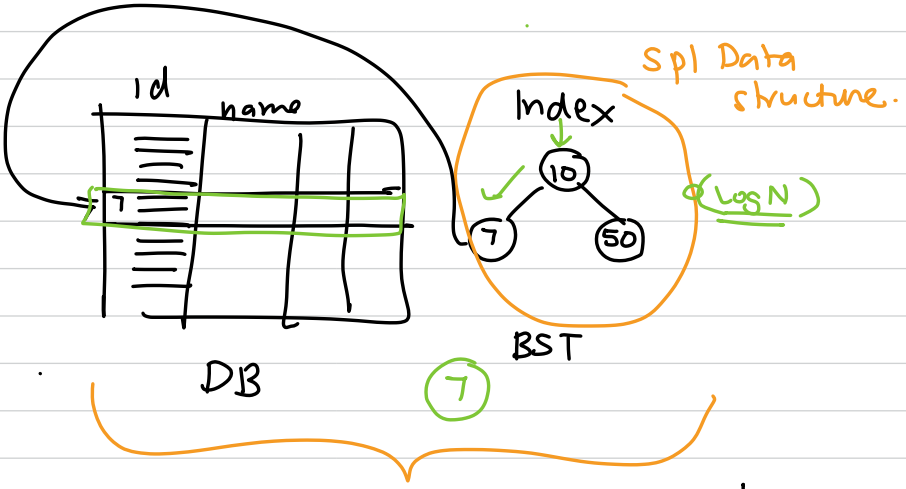
A	{StudentID, FirstName}	14%
B	{StudentID, Age}	0%
C	{StudentID, LastName}	2%
✓ D	All of the above	83%

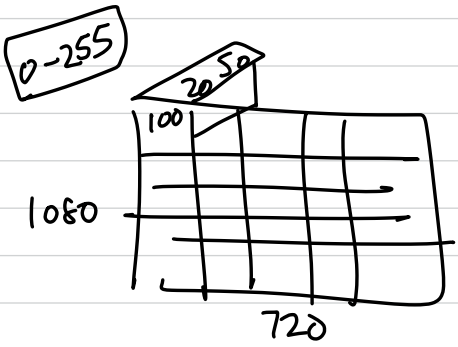
Super-keys.

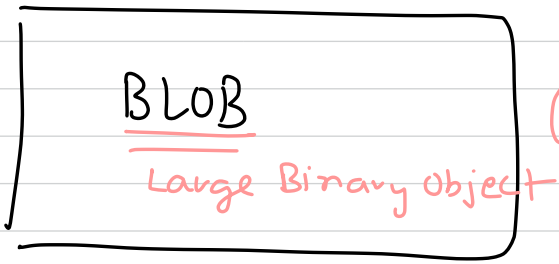
→ 8, Navang

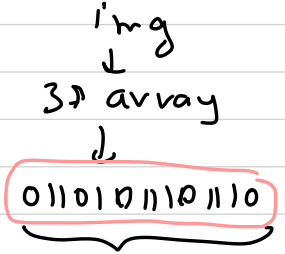
→ 8, ML











1080 X 720 pixels
= 1080 X 720 X 3 integer

$= 1080 \times 720 \times 3 \times 8$

$=$ _____ bit.

huge.