

SOL - vs NOSOL -

① Structured data →

Ebbinghaus Theory -

- 15% in 2 hours

→ 2/3 ~~an~~ will forget in 1 day.

- 3/4 you will forget 1 week.

25% of your knowledge.

SOLID
Outputs longer!

→ 9-6.15 or 2 hours
Average student / Topper -

Topper - Spends his/her time on
revision / concepts rather than
solving same questions - 4h

→ Average student - directly jumps to
a question (9-6).
↑ studying.

Any traiter there is no schedule.
(work 24x7).

① Structured data - Tables, with schemes.

② Semi-structured data - xml, json.

③ Unstructured data \Rightarrow varying len.

} volume and
its
growing.

{ 100 GB can be big data for a startup,
10,000 GB

Eg. - 1000 GB - FB }

Tag,

IoT - varying len

images, audio, video.

id	name	last name	salary
----	------	-----------	--------

Schem X - 1 Analyze

① Structured data → Tech, HR, (Traditionally system) Normalisation forms
Normalise into 3NF forms. (GNK) -

RDBMS - (table, column) - Already have the schema.
MySQL Schemas - | Structure of table

① Create the table at runtime. GNK.

② Create the table/view using script and the program will just insert the data.

① we can have an option - runtime created (if not present) or else it should be present.

② Two ways to create a table
→ fixed schema

RDBMS - create table (table name) (
column name column datatype, constraint

① SQL script.

② Terraform - (language).
Backward

(2) Unstructured data / semi structured data. (NoSQL databases)

① Not having fixed schema. ← (json file)

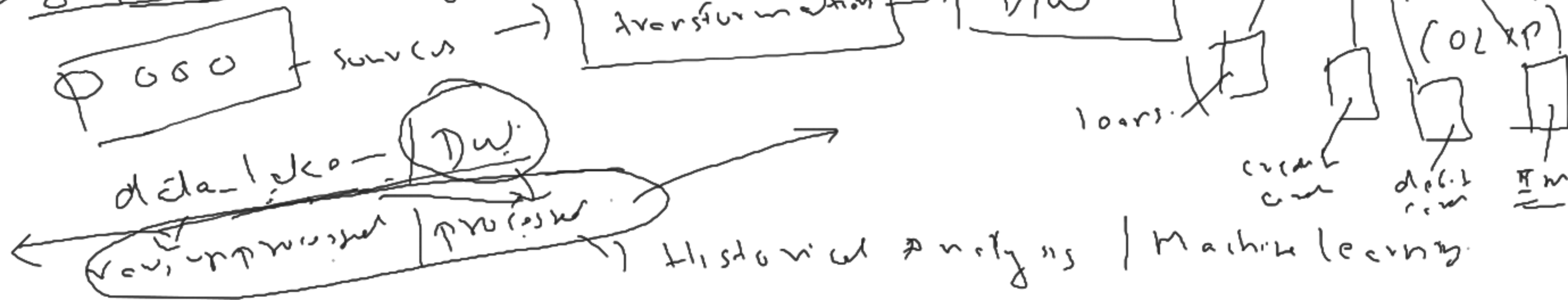
Four kinds-

① Key-Value pair → Redis,

② Document → MongoDB, documentDB, Firestore.
⚡ (ACID)

③ column based database -> Hbase, Cassandra (

④ Graph based - intology



① Online Analytical Processing → historical data / → getting every on.

② Online Transaction Processing → day to day, transactional data.

slowest process.



Hard disk - I/O operations.

① Bank statements.

Regular updates.

Fast enough

manually.

Joltoper (poripunil)

Desktop - Hard disk → I/O operations every

slow.

SSD - I/O operations are faster.

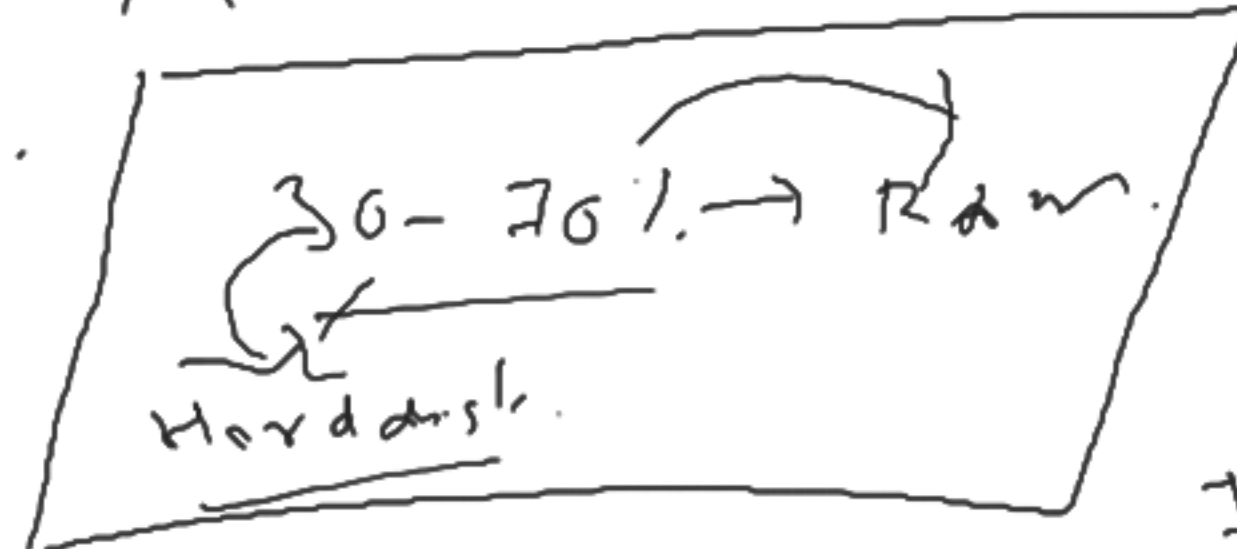
pushed - date - 10%.

RAM - (B+ tree, B-tree) → (MIT).

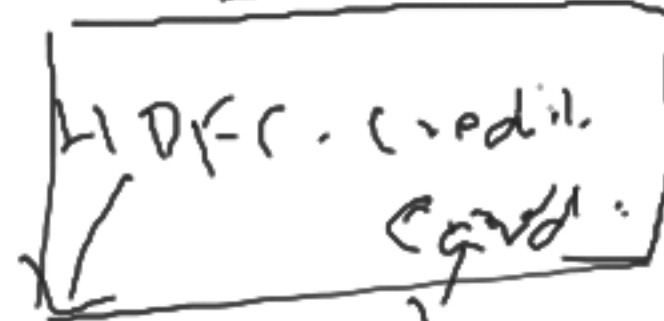
Knitting.



system.



30-40 sec.

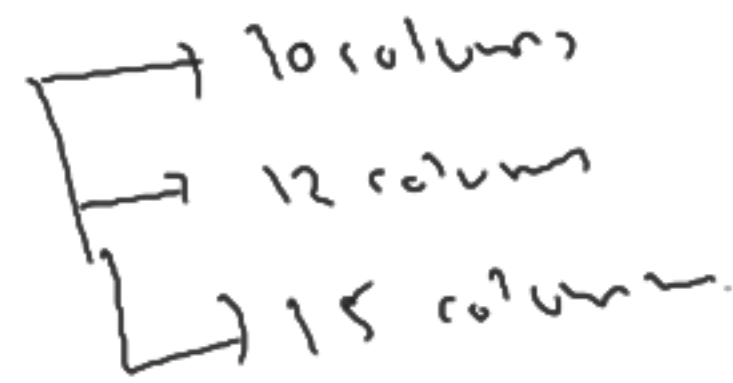


Inter

Digital Services.

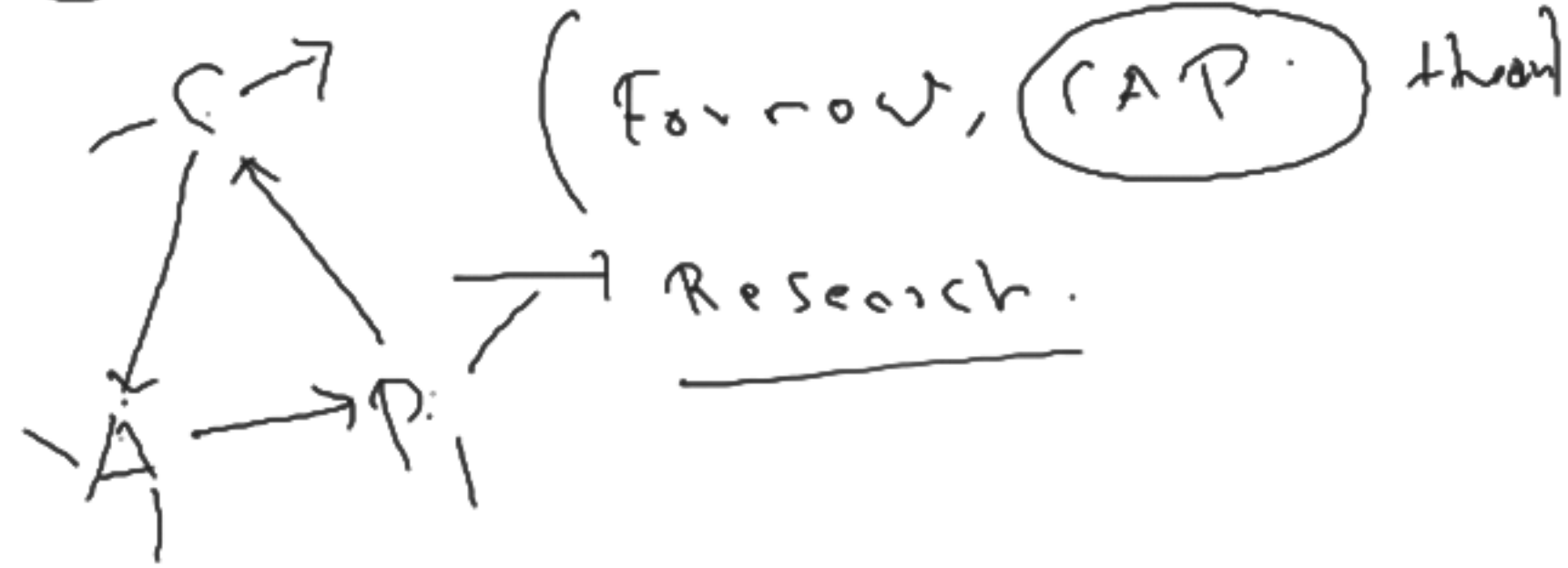
Infogy - Alimda (Finance Minister). → Income tax low.

Schema evolution.



[At regular interval,] → Cassandra.

① Task - (ACID, CAP).
ACID - CAP.



S21

Cloud provider Implementation BigQuery (GCP)

Serverless

High level automatically load a new server for you.

scale up
scale down.

3 server → 5 server - 1/2/1/1/1

mysql

1, 2

Round Robin logic

computation policy → 3 days

T1 - cognized - resigned - 3 months

2 months / 1.5 months

admin, user

S21 2m

7.7.2024

SQL commands -

DDL - Data Definition Language → create, drop, Alter, truncate

DML = " Manipulation language → insert, update, delete, call, explain, lock

TCL = Transaction control language → commit, rollback, savepoint, set transaction, set constraint.

DQL = Data Query language → select.

DCL = Data control language → grant, revoke.

select → many builder,

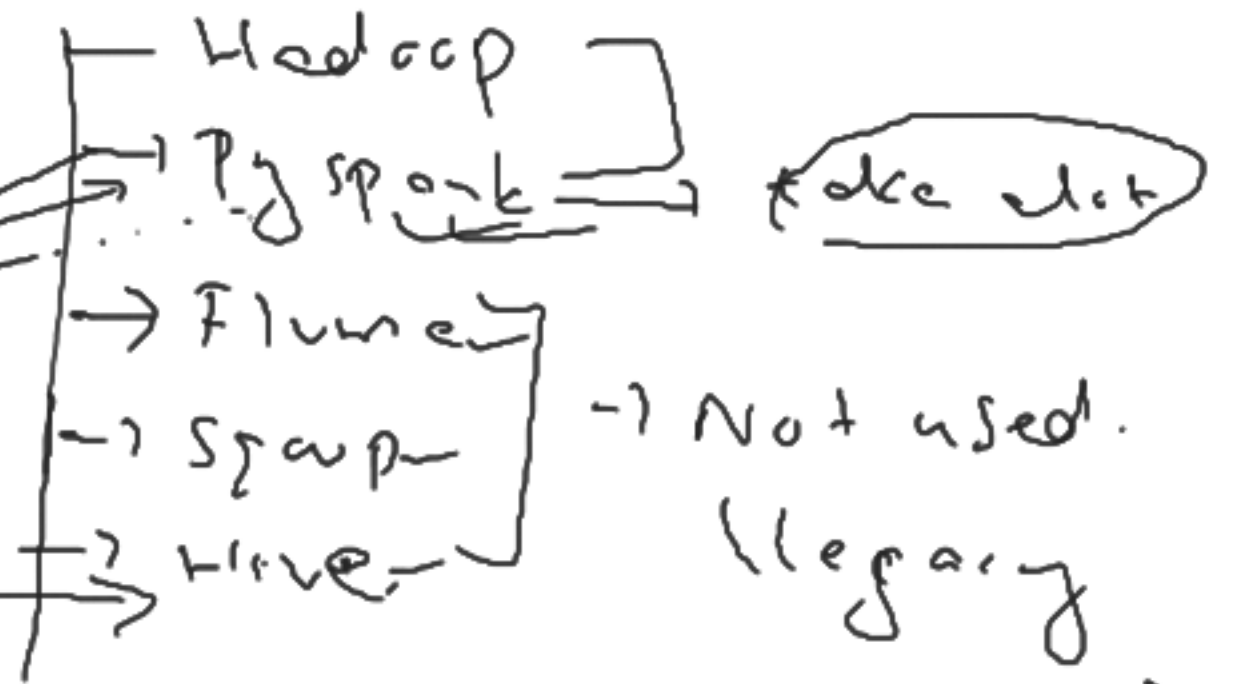
- Case when
- string
- Arithmetic operation
- date time
- if

Hadoop → (Share the link of Hadoop) + configuration. (Youtube video!)

↓ open-source

(PySpark).

Kafka



Stable -

