Schema on Read vs Schema on Write

Prepared By : Prof. Shital Pathar

ETL

- Data cannot be directly loaded into the target system, if data must be loaded(write) then it must match with the structure(schema) of the system.

Schema on write

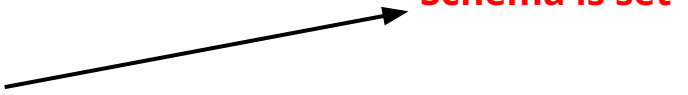
- In Order to load(write) the data into the system,the data must match the structure(schema) of the target system. If the data does not match the target system then the data can not be written into the system.
- **Example:-**
- Create table Employee

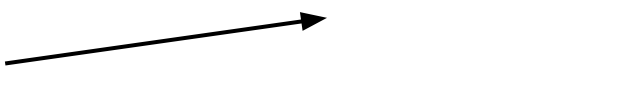
  (id int primary key,

  name varchar, dob date,

  email varchar)

- insert into Employee(id,name,dob,email) values(1,'shital','01/01/9999','abc@xyz.com')



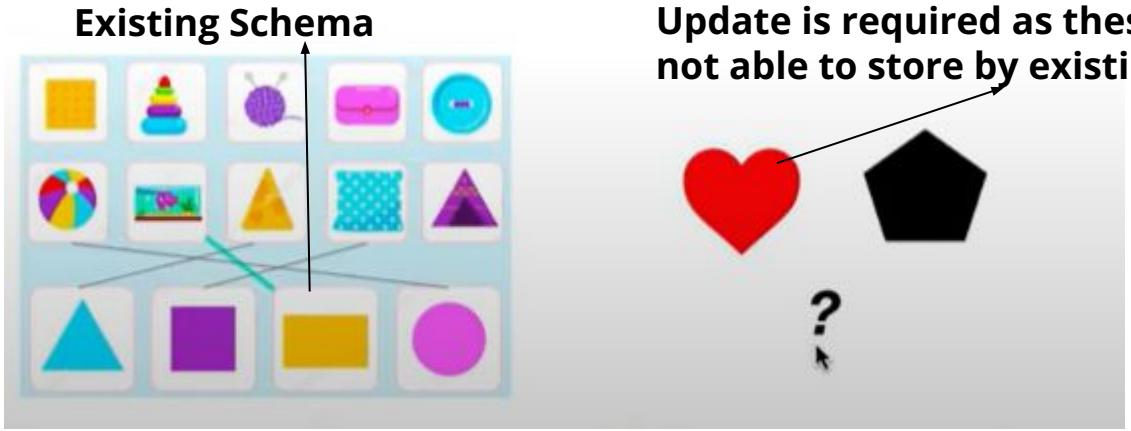


Advantages of schema on write

- Since the data is well structured , read operation is quick and seamless.
- Queries get executed faster.
- If one schema fits for all use cases then schema on write is good.

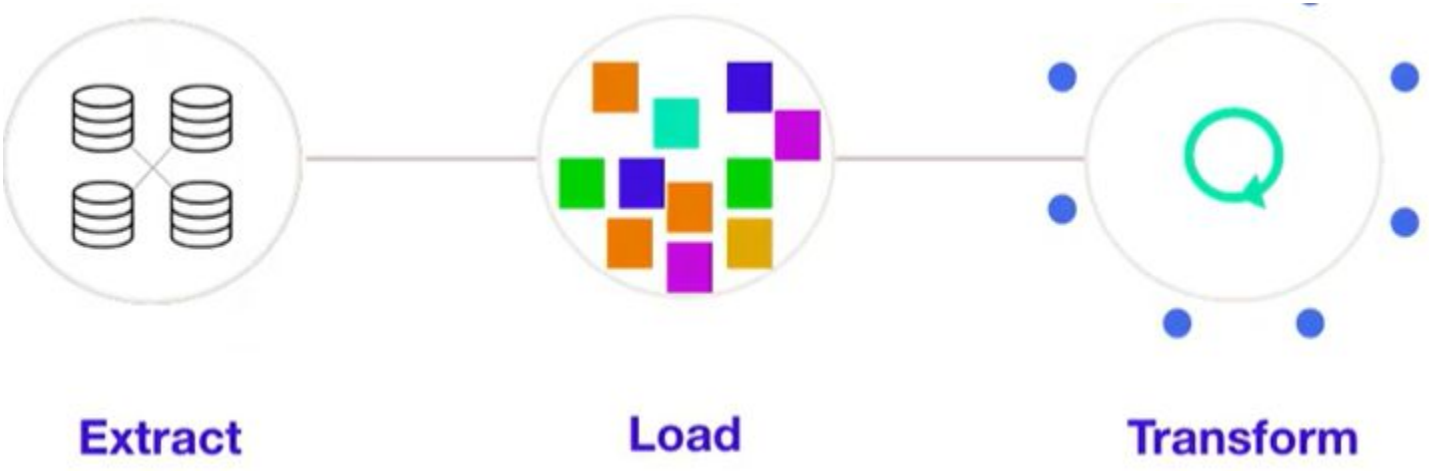
Challenges of Schema on Write

- If the incoming data from data sources does not matches the schema then the schema needs to be altered, which causes several problems.
- Altering the existing schema is difficult and time consuming process.
- **Example** :- what if the schema is required to be updated if different types of requirements are arised.it is very difficult to manage this problem.



Schema on Read(ELT)

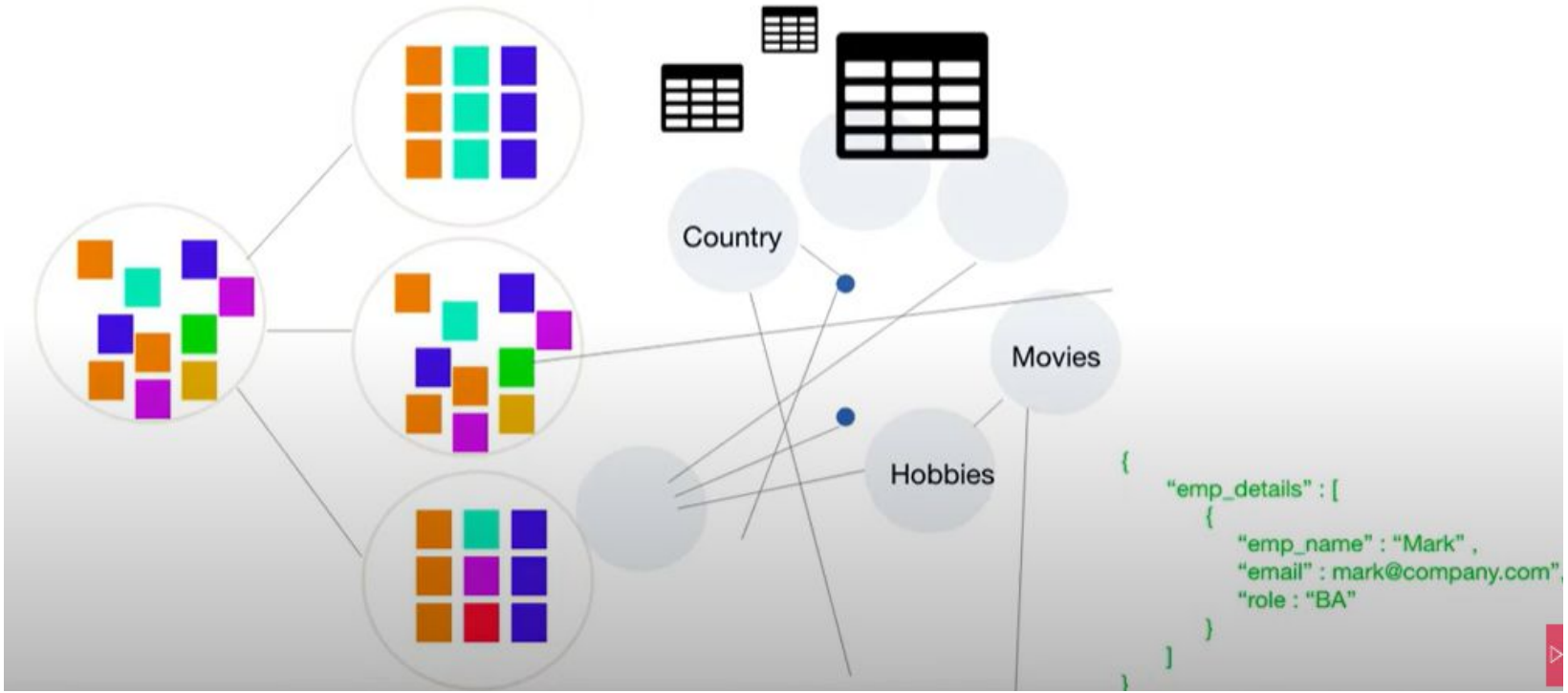
- The aim is not how well the data is organized, the aim is how well the data is organized to get valuable insights.
- An organization may have multiple use cases, each use case may require the specific data.



Schema on Read(ELT)

- In this approach as per the read query request the schema will be built.
- Schema on read gives the benefit of retrieving the most updated data(structurted, semi-structured or unstructured).
- Schema on read will apply a model based on read request and data type.
- Schema on read allows for more flexibility and a richer data exploration experience because analysts can pull in fields as needed.

Schema on read(ELT)



Schema on Read vs Schema on Write

	Schema on Write	Schema on Read
Schema	User has to define a schema	Schema is inferred from the data
Data	Structured and relational	Unstructured and Structured
End User Experience	The only queryable data is pre-selected	Allows richer data exploration
Positive Features	Lightweight	Adaptable