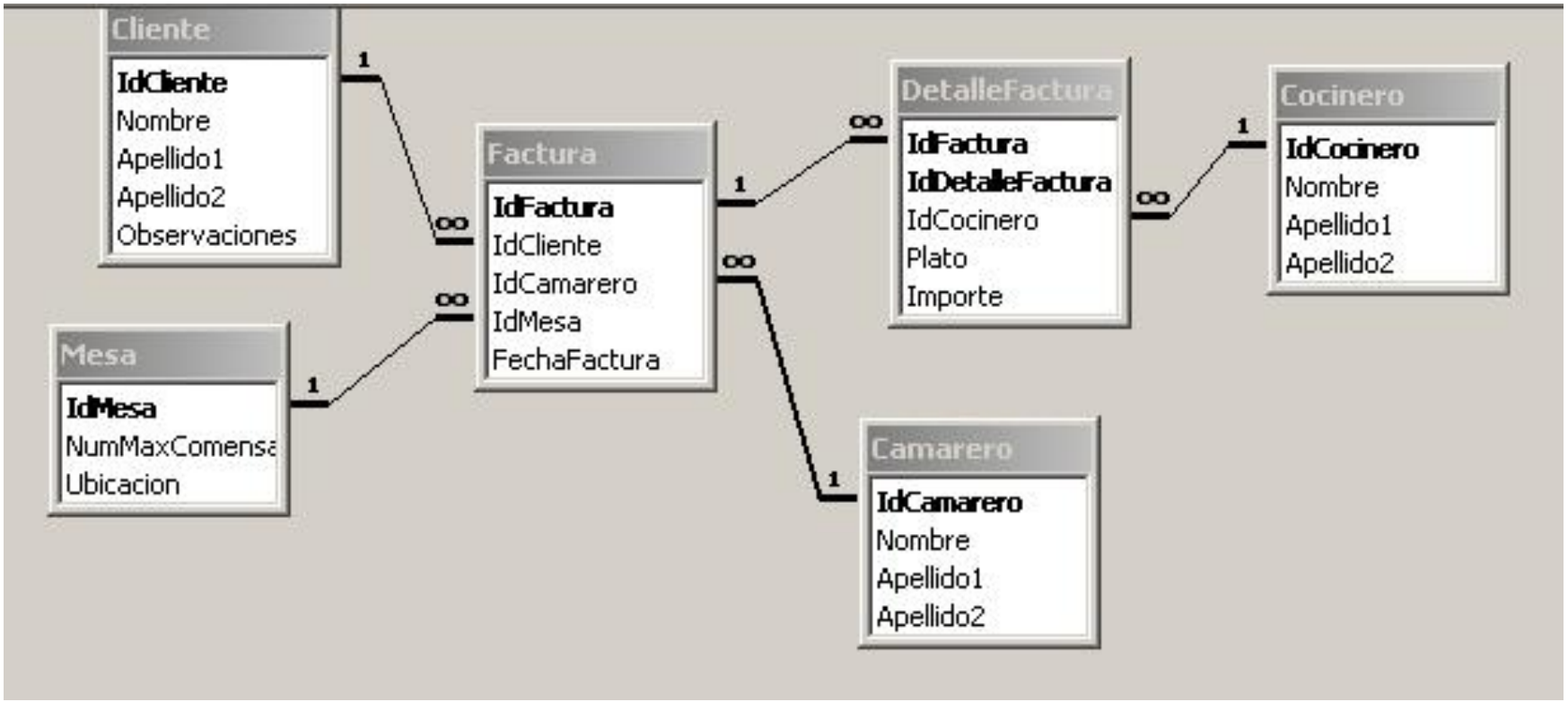
UNIVERSIDAD CATÓLICA de Colombia

Uso de Herramientas Tecnológicas para Data Analytics y Big Data

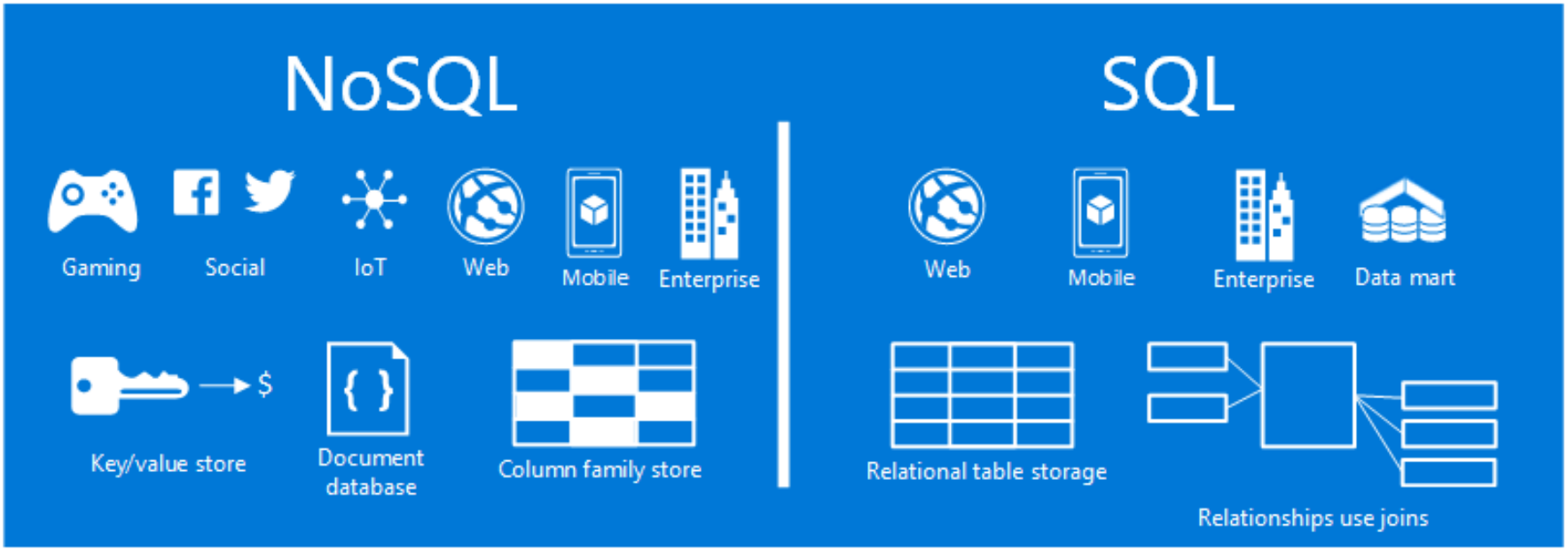
Diego Alberto Rincón Yáñez MSc
darincon@ucatolica.edu.co

Afiliada a la Federación Internacional de Universidades Católicas (FIUC)
w w w . u c a t o l i c a . e d u . c o

Conceptos
Modelo Entidad Relación



Cuando usar NoSQL



Conceptos Diferencias 1.1

RDBMS	MongoDB
Database	Database
Table	Collection
Tuple/Row	Document
column	Field
Table Join	Embedded Documents
Primary Key	Primary Key (Default key _id provided by mongodb itself)

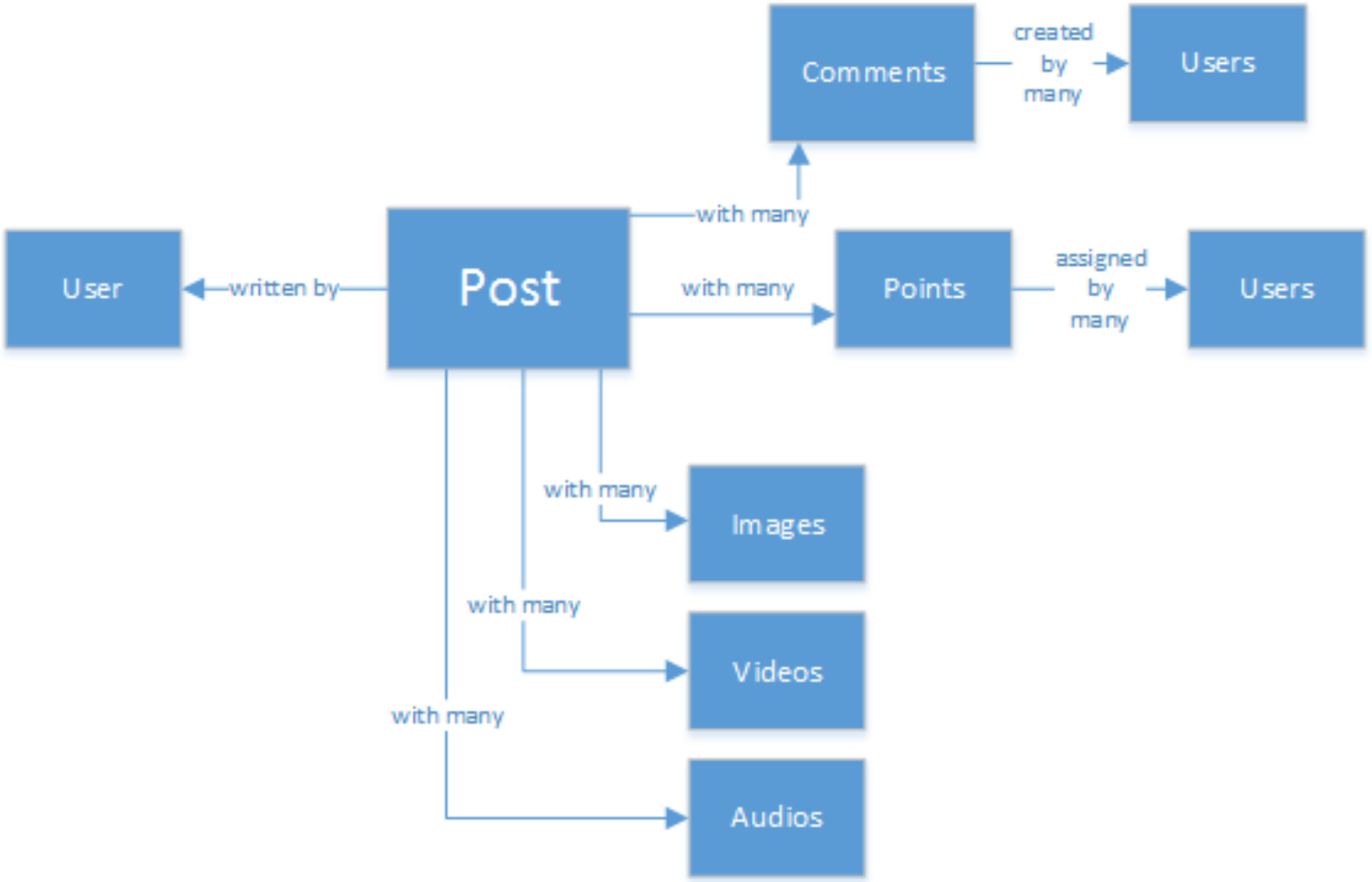
UNIVERSIDAD CATÓLICA de Colombia

Conceptos para NoSQL Mongo DB

- Collecciones (Collections)
  - Entidades (Entities)
    - Documentos (Documents)
      - Campos (Fields)

	NoSQL	SQL
Model	Non-relational	Relational
	Stores data in JSON documents, key/value pairs, wide column stores, or graphs	Stores data in a table
Data	Offers flexibility as not every record needs to store the same properties	Great for solutions where every record has the same properties
	New properties can be added on the fly	Adding a new property may require altering schemas or backfilling data
	Relationships are often captured by denormalizing data and presenting it in a single record	Relationships are often captured in a using joins to resolve references across tables
	Good for semi-structured data	Good for structured data
Schema	Dynamic or flexible schemas	Strict schema
	Database is schema-agnostic and the schema is dictated by the application. This allows for agility and highly iterative development	Schema must be maintained and kept in sync between application and database
Transactions	ACID transaction support varies per solution	Supports ACID transactions
Consistency	Consistency varies per solution, some solutions have tunable consistency	Strong consistency supported
Scale	Scales well horizontally	Scales well vertically

Ejemplo


Comments
created by many
Users
with many
User
written by
Post
with many
Points
assigned by many
Users
with many
Images
with many
Videos
with many
Audios

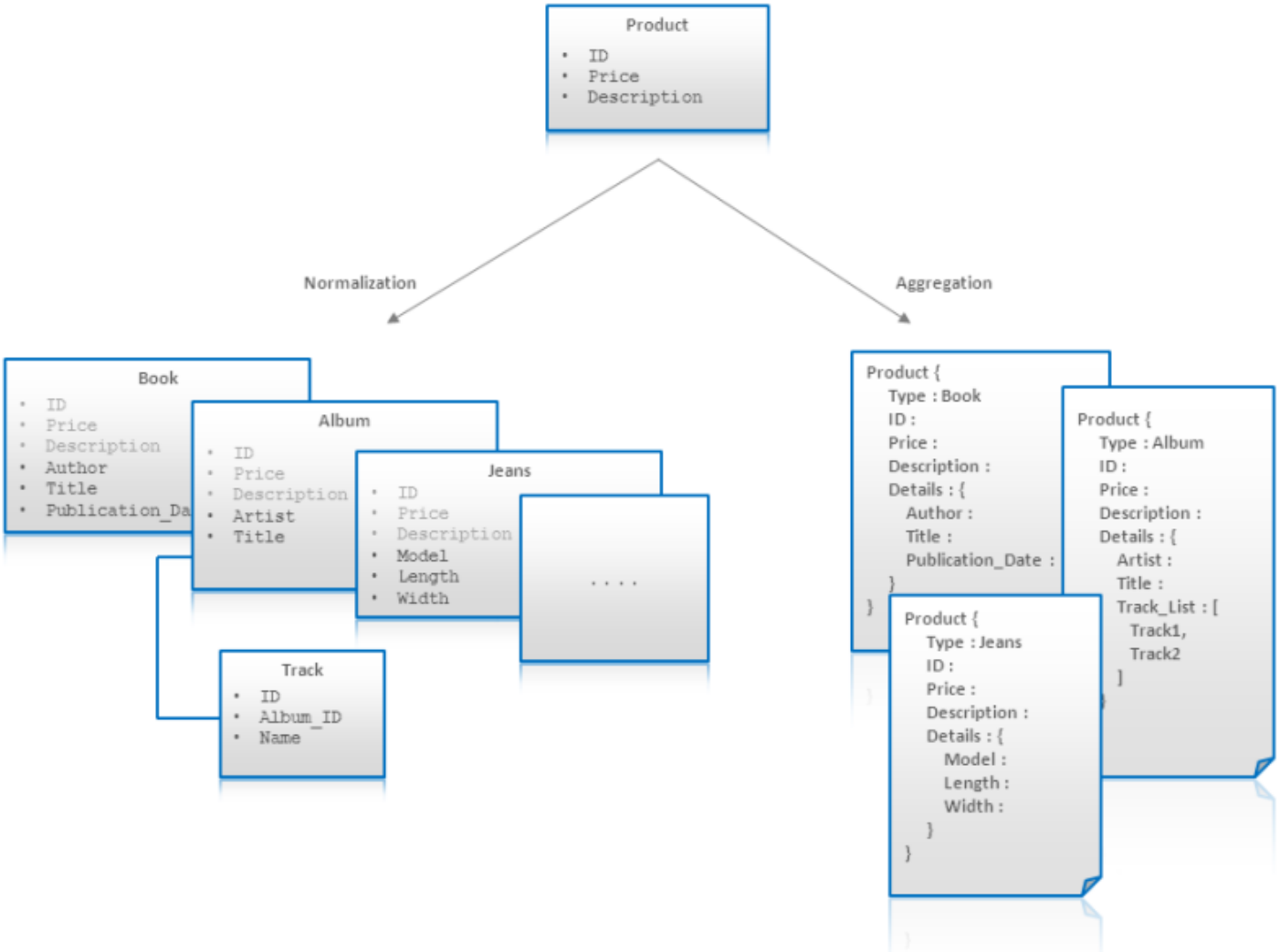
```
{
    "id":"ew12-res2-234e-544f",
    "title":"post title",
    "date":"2016-01-01",
    "body":"this is an awesome post stored on NoSQL",
    "createdBy":User,
    "images":[
        "http://myfirstimage.png",
        "http://mysecondimage.png"
          ],
    "videos":[
         {"url":"http://myfirstvideo.mp4", "title":"The first video"},
        {"url":"http://mysecondvideo.mp4", "title":"The second video"}
          ],
    "audios":[
          {"url":"http://myfirstaudio.mp3", "title":"The first audio"},
          {"url":"http://mysecondaudio.mp3", "title":"The second audio"}
    ]
}
```

Técnicas Conceptuales

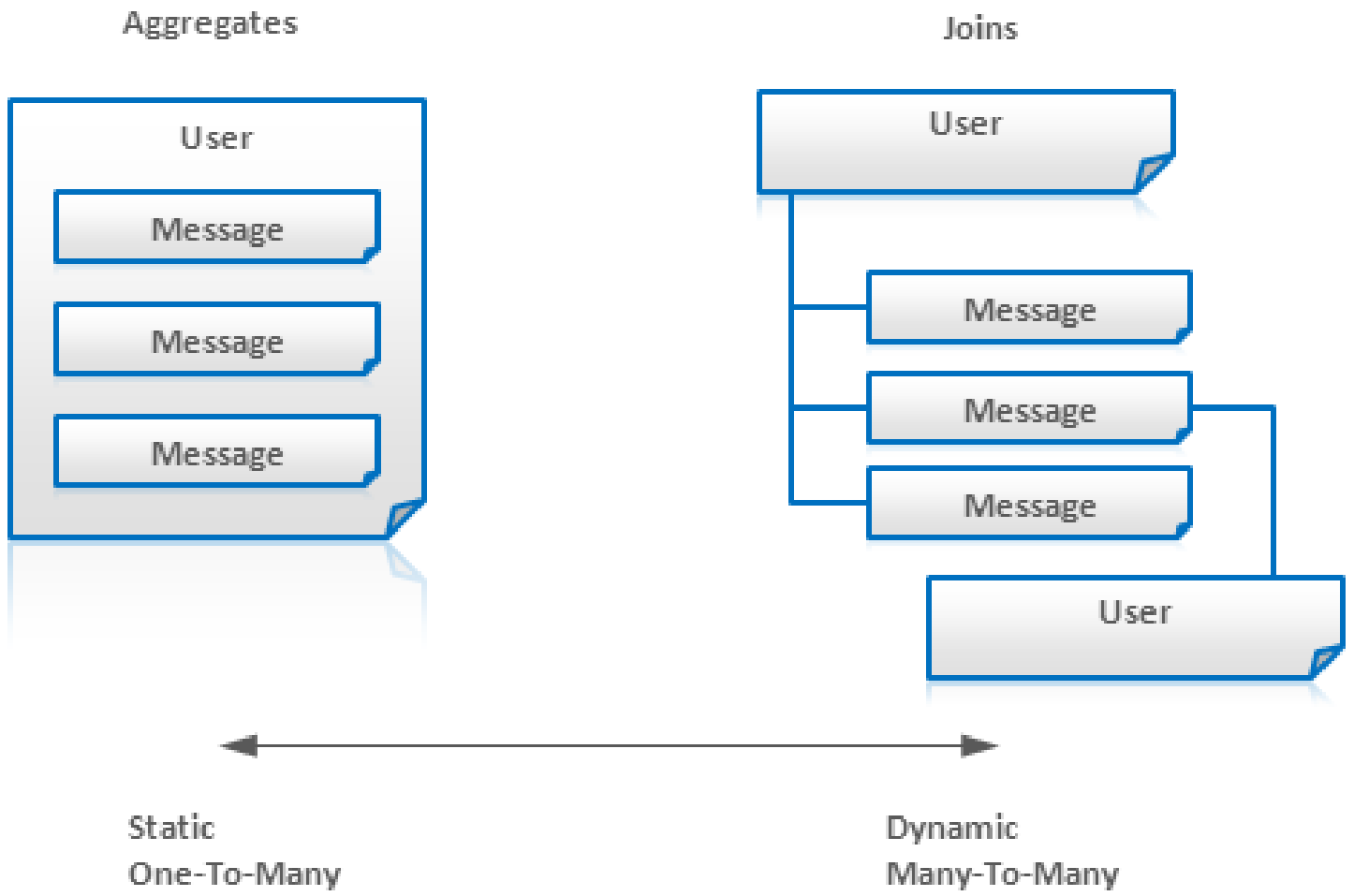
UNIVERSIDAD CATÓLICA de Colombia

- Denormalization
- Aggregates
- Application Side Joins
- Atomic Aggregates
- Dimensionality Reduction

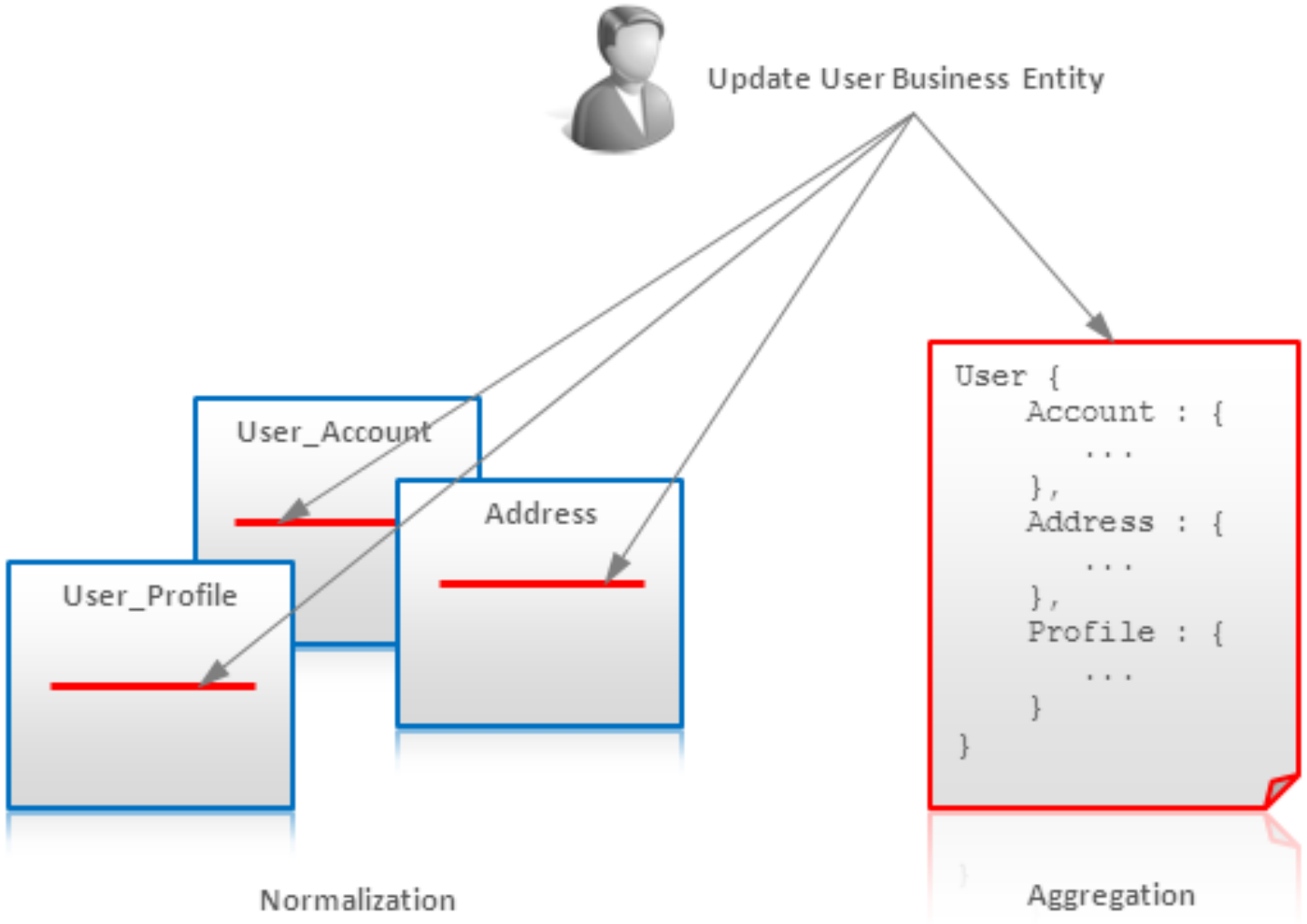
Agregación


Product
ID
Price
Description
Normalization
Aggregation
Book
ID
Price
Description
Author
Title
Publication_Da
Album
ID
Price
Description
Artist
Title
Jeans
ID
Price
Description
Model
Length
Width
. . . .
Track
ID
Album_ID
Name
Product {
Type : Book
ID :
Price :
Description :
Details : {
Author :
Title :
Publication_Date :
}
}
Product {
Type : Album
ID :
Price :
Description :
Details : {
Artist :
Title :
Track_List : [
Track1,
Track2
]
}
Product {
Type : Jeans
ID :
Price :
Description :
Details : {
Model :
Length :
Width :
}
}

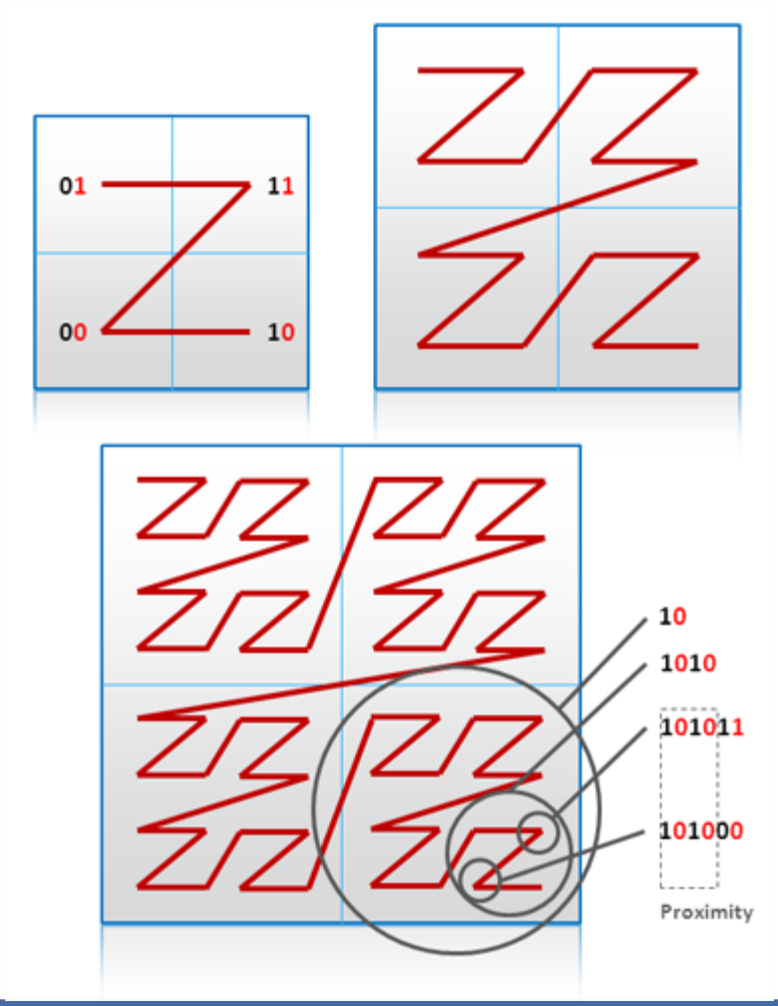
“Joins” Laterales



Agregados Atomicos



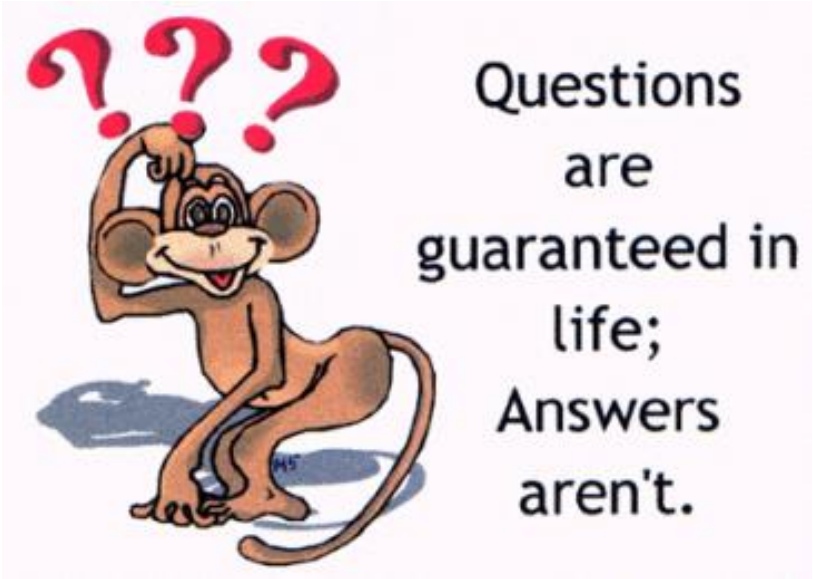
Reducción de Dimensión



UNIVERSIDAD CATÓLICA de Colombia

Ejercicio No. 3 - Fase 3

- Describa las Entidades en MongoDB
- ¿Llene una?



¿Preguntas?

Diego Alberto Rincón Yáñez MCSc.
Twitter: @d1egoprog.