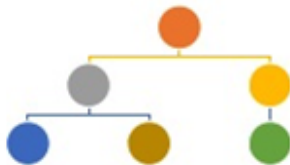# DAY 2 – 4 DOCUMENT MODEL

Emerging Data Modeling and Management Technology:

Key-Value and Document Databases
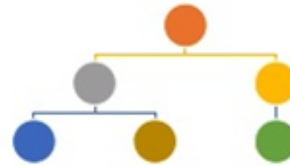

DS&AI : PROFESSIONAL TRAINING COURSE

1 – 3 APRIL 2021 | ASIAN INSTITUTE OF TECHNOLOGY

DS&AI

Document Model

Document Model

A specialized Key-value Store but rather than storing "values," it stores "documents", which are not adhered to schema restrictions.

Provides a way to query the documents based on the contents or metadata.
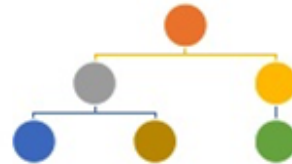
Key/ID/PK

Document

AIT

China

Document Model

A specialized Key-value Store

Designed for storing, retrieving and managing document-oriented information, also known as semi-structured data, such as XML, JSON, BSON

Provides a query/update language that exposes the ability to query or update based on the internal structure in the document.

```json
{
    "FirstName": "Bob",
    "Address": "5 Oak St.",
     ...
}
```

```xml
<contact>
    <firstname>Bob</firstname>
    <lastname>Smith</lastname>
    <phone type="Cell">(123) 555-0178</phone>
    <phone type="Work">(890) 555-0133</phone>
    <address>
     <type>Home</type>
     <street1> 5 Oak St.</street1>
     <city>Boys</city>
     <state>AR</state>
     <zip>32225</zip>
     <country>US</country>
    </address>
  </contact>
```
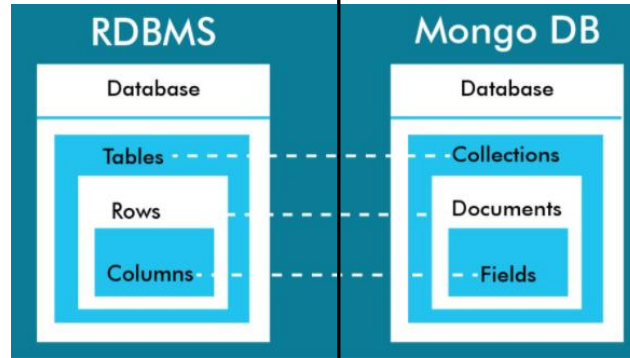
RDBMS vs. Document - What's the difference

Relational data model

Document data model

A Table

A Collection

RDBMS | **Mongo DB**

Database | Database

Tables ----- Collections

Rows ----- Documents

Columns ----- Fields

Relationship : PK, FK, JOIN

One – One Relationship | One – Many Relationship | Many – Many Relationship | Self-Refencing Relationship
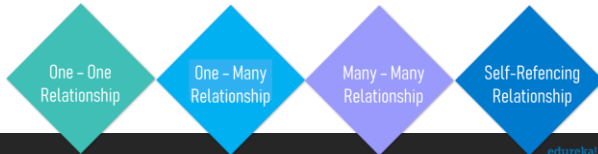
Relationship: embedded & linking, multi-value field types

edureka!

CRUD Operations
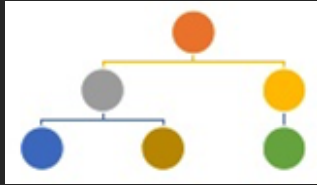
Creation (or insertion)

Retrieval (or query, search, read or find)

Update (or edit)

Deletion (or removal)

About Document Model

Stores data in flexible, JSON-like documents

Fields can vary from document to document and data structure can be changed over time

Queries, indexing, and aggregation

How to model this in JSON?

Size: **12"**

Infield/Outfield/Pitcher model

2-Piece Web pattern

Most popular MLB® pattern among pitchers

Pro Stock® American steerhide leather offers rugged durability and a superior feel

Dual-Welting™ on "exposed edges" of the fingers helps maintain pocket shape and durability

Pro Stock™ hand-designed pattern for unbeatable craftsmanship

Dri-Lex® ultra-breathable wrist lining repels moisture from your hand

Black leather with rich brown embellishments

Pattern: **B212**

Model: **WTA2000BBB212**

Wilson

JSON Object


```
{
    category: "glove",
    model: "PRO112PT",
    name: "Air Elite",
    brand: "Rawlings",
    price: 229.99,
    available: Date("2013-03-31")
}
```

Fields

Values

string

number

date

Field values are typed

Documents are Rich Structures

```
{
    category: "glove",
    model: "PRO112PT",
    name: "Air Elite",
    brand: "Rawlings",
    price: 229.99,
    available: Date("2013-03-31"),
    position: ["infield", "outfield", "pitcher"]
}
```

Fields can contain arrays

Documents are Rich Structures

```
{
        category: "glove",
        model: "PRO112PT",
        name: "Air Elite",
        brand: "Rawlings",
        price: 229.99,
        available: Date("2013-03-31"),
        position: ["infield", "outfield", "pitcher"],
        endorsed: {name: "Ryan Howard",
                           team: "Phillies",
                           position: "first base"},
        history: [{date: Date("2013-03-31"),  price: 279.99},
                  {date: Date("2013-06-01"), price: 259.79},
                  {date: Date("2013-08-15"), price: 229.99}]
}
```

Fields can contain an array of sub-documents

12

Variation is Easy!

```
{
category: bat,
model: B1403E,
name: Air Elite,
brand: "Rip-IT",
price: 399.99

diameter: "2 5/8",
barrel: R2 Alloy,
handle: R2 Composite,
type: composite,

}
```

```
{
category: glove,
model: PRO112PT,
name: Air Elite,
brand: "Rawlings",
price: "229.99"

size: 11.25,
position: outfield,
pattern: "Pro taper",
material: leather,
color: black

}
```

```
{
category: ball,
model: ROML,
name: MLB,
brand: "Rawlings",
price: "6.99"

cover: leather,
core: cork,
color: white

}
```

Easy Query! (MongoDB Query)

```
> db.products.find( { "position" : "infield",
                       "endorsed.team" : "Phillies" } )
         {
                category: "glove",
                model: "PRO112PT",
                name: "Air Elite",
                brand: "Rawlings",
                price: 229.99,
                available: Date("2013-03-31"),
                position: ["infield", "outfield", "pitcher"],
                endorsed: {name: "Ryan Howard",
                           team: "Phillies",
                           position: "first base"},
         }
```

Object Relationships

Referencing & Embedding

EX.1

contact document
```
{
  _id: <ObjectId2>,
  user_id: <ObjectId1>,
  phone: "123-456-7890",
  email: "xyz@example.com"
}
```

user document
```
{
  _id: <ObjectId1>,
  username: "123xyz"
}
```

access document
```
{
  _id: <ObjectId3>,
  user_id: <ObjectId1>,
  level: 5,
  group: "dev"
}
```

```
{
  _id: <ObjectId1>,
  username: "123xyz",
  contact: {
          phone: "123-456-7890",
          email: "xyz@example.com"
        },
  access: {
          level: 5,
          group: "dev"
        }
}
```
Embedded sub-document

Embedded sub-document

https://docs.mongodb.com/manual/core/data-modeling-introduction/

EX.2

Modeled in 2 possible ways

Embed

Patients

```
{
    _id: 2,
    first: "Joe",
    last: "Patient",
    addr: { ...},
    procedures: [
        {
            id: 12345,
            date: 2015-02-15,
            type: "Cat scan",
            ...},
        {
            id: 12346,
            date: 2015-02-15,
            type: "blood test",
            ...}]
}
```

Reference

Patients

```
{
    _id: 2,
    first: "Joe",
    last: "Patient",
    addr: { ...},
    procedures: [12345, 12346]}
```

Procedures

```
{
    _id: 12345,
    date: 2015-02-15,
    type: "Cat scan",
    ...}
{
    _id: 12346,
    date: 2015-02-15,
    type: "blood test",
    ...}
```
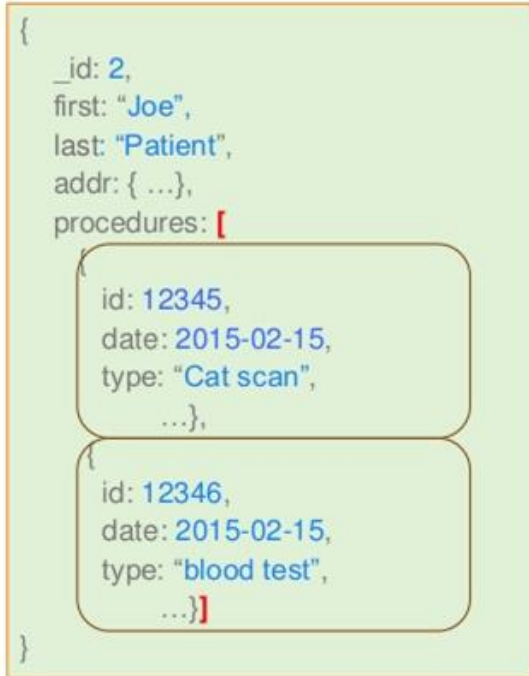
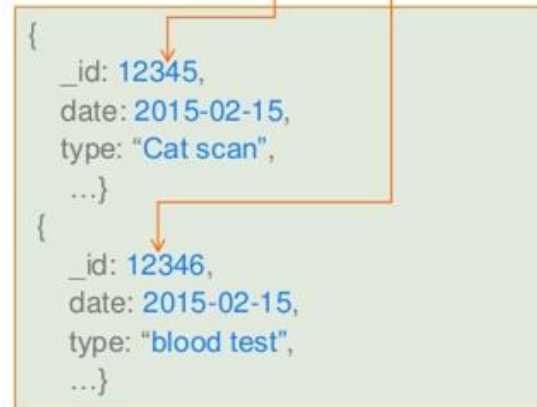**EX.3**

Embedding **Physicians** in **Hospitals** collection

```
{
  _id: 1,
  name: "Oak Valley Hospital",
  city: "New York",
  beds: 131,
  physicians: [
    {
      id: 12345,
      name: "Joe Doctor",
      address: {...},
        ...},
    {
      id: 12346,
      name: "Mary Well",
      address: {...},
        ...}]
}
```

```
{
  _id: 2,
  name: "Plainmont Hospital",
  city: "Omaha",
  beds: 85,
  physicians: [
    {
      id: 63633,
      name: "Harold Green",
      address: {...},
        ...},
    {
      id: 12345,
      name: "Joe Doctor",
      address: {...},
        ...}]
}
```

Data Duplication

the REFERENCE

○ P. Sadalage and M. Fowler: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley Professional, 2013

○ Jan L. Harrington: Relational Database Design and Implementation, 4th edition, Morgan Kaufmann, 2016

○ A. Makris, K. Tserpesa, V. Andronikou Dimosthenis Anagnostopoulos: A Classification of NoSQL Data Stores Based on Key Design Characteristics, Procedia Computer Science, Vol. 97, 2016, pp. 94-103.

○ MongoDB Schema Design: Practical Applications and Implications [https://www.slideshare.net/mongodb/mongodb-schema-design-practical-applications-and-implications]