DATA MODELING

Embedded (Denormalized) Data

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

Embedded sub-document

Embedded sub-document

Embedded: When to Use

- Contains relationship: one-to-one

```
{
   _id: "joe",
   name: "Joe Bookreader",
   address: {
              street: "123 Fake Street",
              city: "Faketon",
              state: "MA",
              zip: "12345"
            }
}
```

Embedded: When to Use

- Contains relationship: always queried with parent

```
{
 "ID": 1,
 "ItemName": "hamburger",
 "ItemDescription": "cheeseburger, no cheese",
 "Category": "sandwiches",
 "CategoryDescription": "2 pieces of bread + filling",
 "Ingredients": [
     {"ItemName": "bread", "calorieCount": 100, "Qty": "2 slices"},
     {"ItemName": "lettuce", "calorieCount": 10, "Qty": "1 slice"}
     {"ItemName": "tomato","calorieCount": 10, "Qty": "1 slice"}
     {"ItemName": "patty", "calorieCount": 700, "Qty": "1"}
}
```

Embedded: When to Use

- Contains relationship: child data intrinsic to parent

```
{
  "id": "Order1",
  "customer": "Customer1",
  "orderDate": "2018-09-26",
  "itemsOrdered": [
      {"ID": 1, "ItemName": "hamburger",
        "Price":9.50, "Qty": 1}
      {"ID": 2, "ItemName": "cheeseburger",
        "Price":9.50, "Qty": 499}
  ]
}
```

51

51

Embedded: When to Use

- Similar rate of updates
- One-to-Few relationship

```
{   "_id": "joe",
    "name": "Joe Bookreader",
    "addresses": [
                  { "street": "123 Fake Street",
                    "city": "Faketon",
                    "state": "MA", "zip": "12345"
                  },
                  {"street": "1 Some Other Street",
                    "city": "Boston",
                    "state": "MA", "zip": "12345"
                  }
                ]
 }
```

52

52

Embedded: When to Use

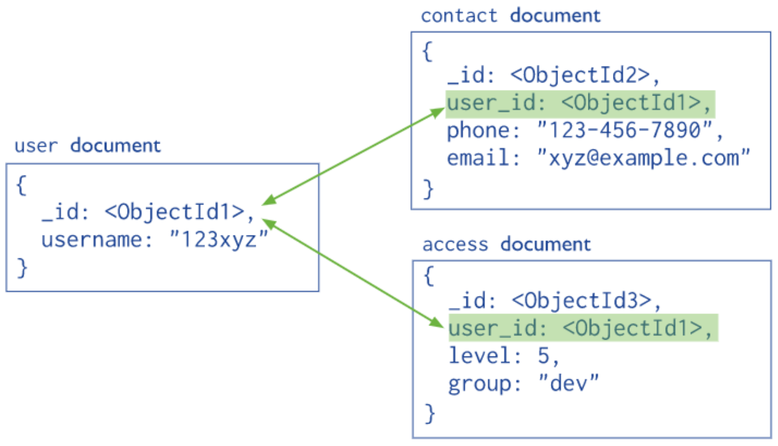
- One-to-One
- Always queried with parent
- Child data intrinsic to parent
- Similar rate of updates
- One-to-Few

- Better read performance
- Single atomic write for related data

53

53

Linked (Normalized) Data



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"
}
```

54

54

Linked (Normalized) Data

- 1 : many (unbounded relationship)

```
{
"_id": "1",
"name": "Alice",
"email": "alice@contoso.com",
"Orders": [
  {
       "_id": "Order1",
       "orderDate": "2018-09-18",
       "itemsOrdered": [
           {"ID": 1, "ItemName": "hamburger", "Price":9.50, "Qty": 1}
           {"ID": 2, "ItemName": "cheeseburger", "Price":9.50, "Qty":
           499}]
  },
  ...
  {
       "_id": "OrderNfinity",
       "orderDate": "2018-09-20",
       "itemsOrdered": [
           {"ID": 1, "ItemName": "hamburger", "Price":9.50, "Qty": 1}]
  }]
}
```

Linked (Normalized) Data

- Data changes at different rates

```
{
"id": "1",
"name": "Alice",
"email": "alice@contoso.com",
"stats":[
 {"TotalNumberOrders": 100},
 {"TotalAmountSpent": 550}]
}
```

Linked (Normalized) Data

- Many : Many relationships

```
{ "_id": "speaker1",
  "name": "Alice",
  "email": "alice@contoso.com",
  "sessions":[
        {"id": "session1"},
        {"id": "session2"}
    ]
}
```

```
{ "_id": "speaker2",
  "name": "Bob",
  "email": "bob@contoso.com",
  "sessions":[
        {"id": "session1"},
        {"id": "session4"}
    ]
}
```

```
{
  "_id": "session1",
  "name":
     "Modelling Data 101",
  "speakers":[
      {"id": "speaker1"},
      {"id": "speaker2"}
   ]
}
```

Linked (Normalized) Data

- Heavily referenced by many others

```
{ "_id": "speaker1",
  "name": "Alice",
  "email": "alice@contoso.com",
  "sessions":[
        {"id": "session1"},
        {"id": "session2"}
  ]
}
```

```
{ "_id": “attendee1",
  "name": “Eve",
  "email": “eve@contoso.com",
  "bookmarkedSessions":[
        {"id": "session1"},
        {"id": "session4"}
  ]
}
```

```
{
  "_id": "session1",
  "name":
    "Modelling Data 101",
  "speakers":[
      {"id": "speaker1"},
      {"id": "speaker2"}
  ]
}
```

Normalized: When to Use

- 1-to-many (unbounded relationship)
- Many-to-many relationships
- Data changes at different rates
- What is referenced, is heavily referenced by many others

59

59

Denormalized Many-To-One

```
{
   title: "MongoDB",
   author: "Kristina
Chodorow",
published_date: ...,
   pages: 216,
   language: "English",
   publisher:
    { name: "O'Reilly",
      founded: 1980,
      location: "CA"
    }
}
```

```
{
   title: "50 Tips and
Tricks",
   author: "Kristina
Chodorow",
   published_date: ...,
   pages: 68,
   language: "English",
   publisher:
    { name: "O'Reilly",
      founded: 1980,
      location: "CA"
    }
}
```

60

60

Normalized One-To-Many

```
{
   name: "O'Reilly",
   founded: 1980,
   location: "CA",
   books: [123456789,
           234567890, ...]
}

{
    _id: 123456789,
    title: "MongoDB",
    author: [ "Kristina
Chodorow" ],
    published_date: ...,
    pages: 216,
    language: "English"
}
```

```
{
   _id: 234567890,
   title: "50 Tips and Tricks",
   author: [ "Kristina
Chodorow" ],
   published_date: ...,
   pages: 68,
   language: "English"
}
```

61

61

Normalized Many-To-One

```
{
   _id: "oreilly",
   name: "O'Reilly",
   founded: 1980,
   location: "CA"
}

{
   _id: 123456789,
   title: "MongoDB",
   author: [ "Kristina
Chodorow" ],
   published_date: ...,
   pages: 216,
   language: "English",
   publisher_id: "oreilly"
}
```

```
{
   _id: 234567890,
   title: "50 Tips and Tricks",
   author: "Kristina Chodorow",
   published_date: ...,
   pages: 68,
   language: "English",
   publisher_id: "oreilly"
}
```

62

62

DBRefs

- Manual Reference
  - include document's _id field in another
  - What if more than one collection?
- DBRef:
  - $id: _id of referenced document
  - $ref: collection
  - $db: database (optional)

63

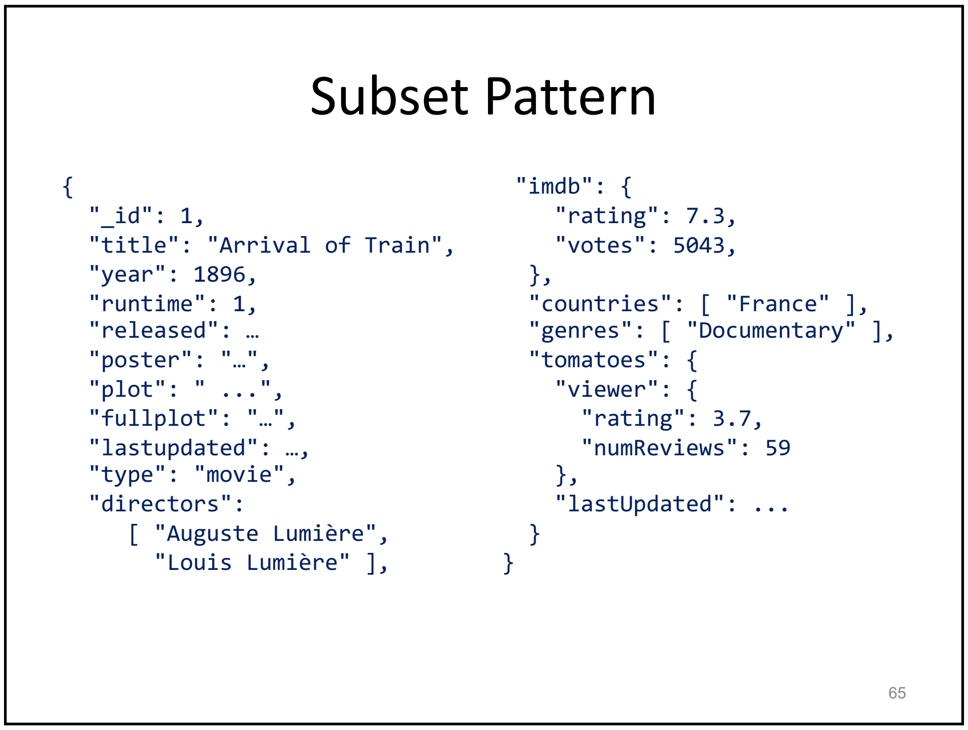
63

DBRefs

```
{
  "_id" : ObjectId("5126bbf64aed4daf9e2ab771"),
  // .. application fields
  "creator" : {
      "$ref" : "creators",
      "$id" : ObjectId("5126bc054aed4daf9e2ab772"),
      "$db" : "users",
      "extraField" : "anything"
               }
}
```

64

64



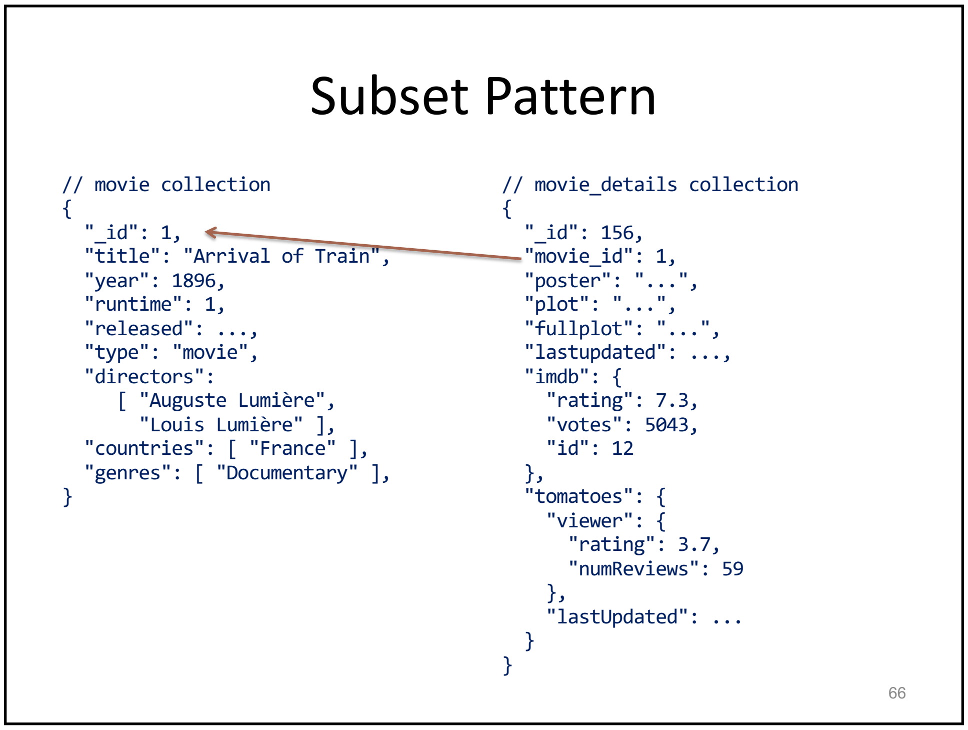
Subset Pattern

```
{
  "_id": 1,
  "title": "Arrival of Train",
  "year": 1896,
  "runtime": 1,
  "released": …
  "poster": "…",
  "plot": " ...",
  "fullplot": "…",
  "lastupdated": …,
  "type": "movie",
  "directors":
     [ "Auguste Lumière",
       "Louis Lumière" ],
```

```
 "imdb": {
    "rating": 7.3,
    "votes": 5043,
  },
  "countries": [ "France" ],
  "genres": [ "Documentary" ],
  "tomatoes": {
    "viewer": {
      "rating": 3.7,
      "numReviews": 59
    },
    "lastUpdated": ...
  }
}
```

65

65



Subset Pattern

```
// movie collection
{
  "_id": 1,
  "title": "Arrival of Train",
  "year": 1896,
  "runtime": 1,
  "released": ...,
  "type": "movie",
  "directors":
     [ "Auguste Lumière",
       "Louis Lumière" ],
  "countries": [ "France" ],
  "genres": [ "Documentary" ],
}
```

```
// movie_details collection
{
  "_id": 156,
  "movie_id": 1,
  "poster": "...",
  "plot": "...",
  "fullplot": "...",
  "lastupdated": ...,
  "imdb": {
    "rating": 7.3,
    "votes": 5043,
    "id": 12
  },
  "tomatoes": {
    "viewer": {
      "rating": 3.7,
      "numReviews": 59
    },
    "lastUpdated": ...
  }
}
```

66

66

Subset Pattern

- Only embed most frequently used data
- Benefit: Improved read performance
- Cost: Additional reads, normalization
- Cost: Duplicate storage
  - Maintain consistency between copies
- Cost: Correct embedding
  - Ex: Highest-rated comments on blog post

67

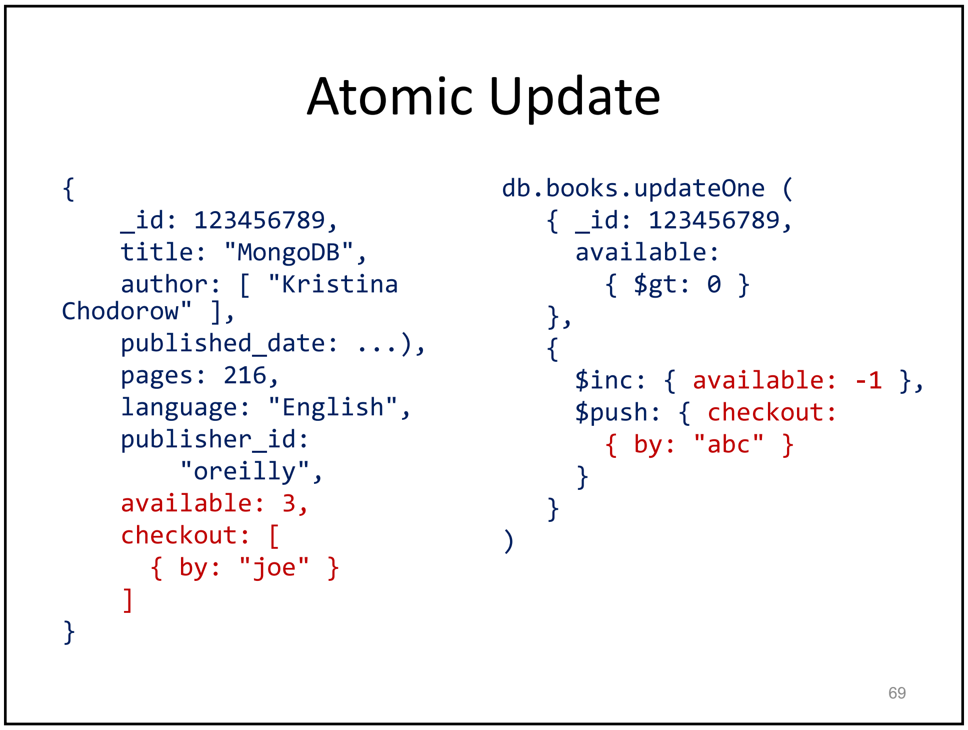
67

Operational Concerns

- Atomic update
  - Within a document: atomic write
  - Multiple documents: transactions
- Sharding
  - choice of shard key
- Large number of Collections
- Large number of Small Documents
  - "Roll up" logical grouping of small docs

68

68



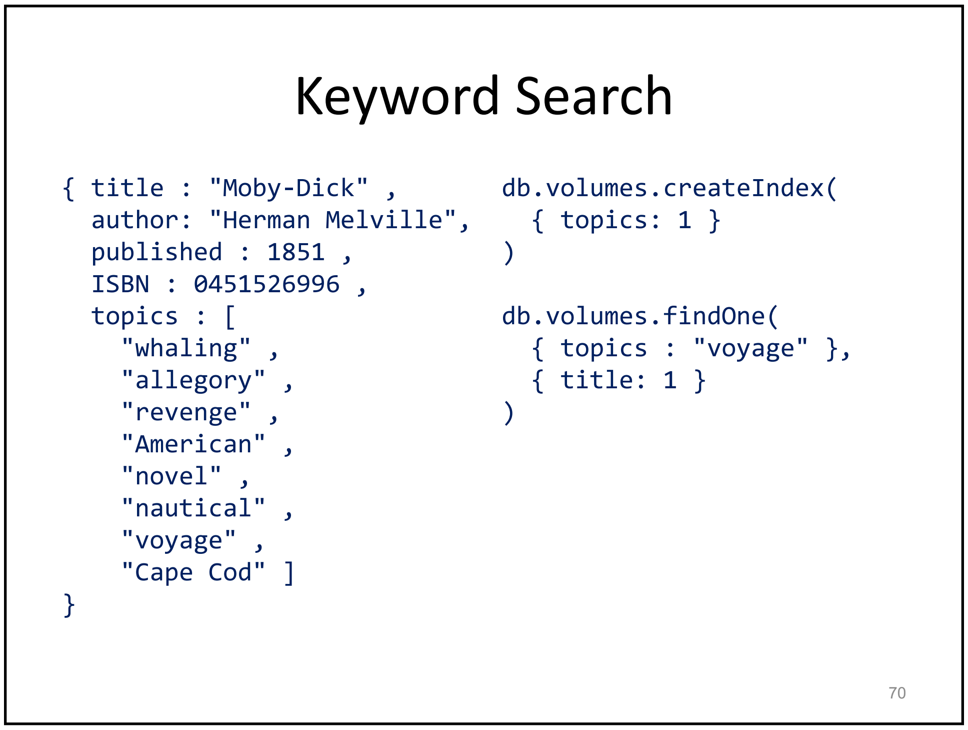
Atomic Update

```
{
    _id: 123456789,
    title: "MongoDB",
    author: [ "Kristina
Chodorow" ],
    published_date: ...),
    pages: 216,
    language: "English",
    publisher_id:
        "oreilly",
    available: 3,
    checkout: [
      { by: "joe" }
    ]
}
```

```
db.books.updateOne (
   { _id: 123456789,
     available:
       { $gt: 0 }
   },
   {
     $inc: { available: -1 },
     $push: { checkout:
       { by: "abc" }
     }
   }
)
```

69

69



Keyword Search

```
{ title : "Moby-Dick" ,
  author: "Herman Melville",
  published : 1851 ,
  ISBN : 0451526996 ,
  topics : [
    "whaling" ,
    "allegory" ,
    "revenge" ,
    "American" ,
    "novel" ,
    "nautical" ,
    "voyage" ,
    "Cape Cod" ]
}
```

```
db.volumes.createIndex(
  { topics: 1 }
)

db.volumes.findOne(
  { topics : "voyage" },
  { title: 1 }
)
```

70

70