

MongoDB

Mimanshu Maheshwari

August 31, 2024

Contents

1	Introduction	4
1.1	Getting Started with MongoDB Atlas	4
1.1.1	Creating and Deploying on Atlas Cluster	4
1.2	Introduction to MongoDB	5
1.2.1	Key Terms	5
1.3	MongoDB and The Document Model	5
1.3.1	The MongoDB Document Model	5
1.4	Managing Databases, Collections, and Documents in Atlas Data Explorer	5
2	Data Modeling	6
2.1	Introduction to Data Modeling	6
2.2	Types of Data Relationships	7
2.2.1	Different types of relationships data can have:	7
2.2.2	Two main ways to model these relationships:	7
2.3	Modeling Data Relationships	7
2.3.1	Embedding Data in Documents	7
2.3.2	Referencing Data in Documents	8
2.4	Scaling a Data Model	8
2.5	Using Atlas Tools for Schema Help	8
3	Connection to MongoDB Database	10
3.1	Using MongoDB Connection Strings	10
3.1.1	Connection String	10
3.2	Connecting to a MongoDB Atlas Cluster with the Shell	10

List of Figures

Listings

1.1	account test data	4
1.2	filter on account id	4
1.3	load cluster	5

Chapter 1

Introduction

1.1 Getting Started with MongoDB Atlas

1.1.1 Creating and Deploying on Atlas Cluster

1. Create a Organization
2. Create a Project under Organization
3. Create a cluster → Advanced Settings → Shared M0 cluster → Add username and password → Add you ip as allowed origin.
4. Open Database tab and navigate to your cluster
5. Load default data
6. Navigate to collections tab to see loaded clusters
7. Sample database fetch from analytics accounts table

```
1 {
2   "_id": {
3     "$oid": "5ca4bbc7a2dd94ee5816238c"
4   },
5   "account_id": {
6     "$numberInt": "371138"
7   },
8   "limit": {
9     "$numberInt": "9000"
10  },
11  "products": [
12    "Derivatives",
13    "InvestmentStock"
14  ]
15 }
```

Listing 1.1: account test data

8. To filter on account id:

```
1 {
2   account_id: 794875
3 }
```

Listing 1.2: filter on account id

9. To login into atlas login use `atlas auth login`

10. create cluster with user and password:

```
1 atlas setup --clusterName myAtlasClusterEDU --provider WS --currentIp --skipSampleData --  
  ↪ username myAtlasDBUser --password myatlas-001 --projectId 66d19clf26ef8b512df3b41 |  
  ↪ tee atlas_cluster_details.txt
```

Listing 1.3: load cluster

11. Load sample data: `atlas clusters smapleData load myAtlasClusterEDU`

1.2 Introduction to MongoDB

1.2.1 Key Terms

Document

The basic using of data in MongoDB

Collections

A grouping of documents. Documents in collection are typically similar they don't necessarily have same structure.

Database

A container for Collections

1.3 MongoDB and The Document Model

1.3.1 The MongoDB Document Model

- MongoDB stores data in structures known as documents

Document

Docuement are displayed in JavaScript Object Notation (JSON) format but stored in Binary JavaScript Object Notation (BSON) format. BSON also supports additional data types that are unavailable in JSON

It can support All JSON data types as well as Dates, Numbers, Object Id's, and more!

ObjectId is used to create unique identifires. Every Document requires an *_id* field, which acts as a primary key. If an inserted document doesn't include the *_id* field, MongoDB automatically generates an ObjectId for the *_id* field. MongoDB supports flexible schema model and polymorphic data, this allows us to store documents with different structure together. Documents may contain fields. Fields may contain different types.

Optional Schema validation

Set of constraints on the structure of documents.

1.4 Managing Databases, Colections, and Documents in Atlas Data Explorer

Explore Atlas

Chapter 2

Data Modeling

2.1 Introduction to Data Modeling

Data Modeling is process of defining how data is stored and the relationships and the relationships that exists among different entities in your data.

We refer to Organization of data inside a database as a **Schema**.

To develop schema think about your database and you need to ask these questions:

- What does my application do?
- What data will I store?
- How will users access this data?
- What data will be most valuable to me?

Asking these questions will help you to:

- describe your task as well those of users.
- What you data looks like?
- the relationships among the data.
- The tooling you plan to have.
- The access patterns that might emerge.

A good data model can:

- Make it easier to manage data
- Make queries more efficient
- Use less memory and Computer Processing Unit (CPU)
- Reduce cost

As a general principal of MongoDB data that is accessed together should be stored together. Collections do not enforce any document structure by default. That means document even those in the same collection can have different structures.

Each document at MongoDB can be different which is known as **Polymorphism**.

Embedded document model enables us to build complex relationships among data. You can normalize your data by using database references.

The key principal here is how your application will use the data rather than how it's stored in the database.

2.2 Types of Data Relationships

2.2.1 Different types of relationships data can have:

- One-to-one
- One-to-many
- Many-to-Many

One-to-One

A relationship where a data entity in one set is connected to exactly one data entity in another set.

One-to-Many

A relationship where a data entity in one set is connected to any number of entities in another set.

Many-to-Many

A relationship where any number of data entities in one set are connected to any number of data entities in another set.

2.2.2 Two main ways to model these relationships:

- Embedding
- Referencing

Embedding

We take related data and insert it into our document.

Referencing

When we refer to documents in another collection in our document. Reference is made by their respective `ObjectID`.

2.3 Modeling Data Relationships

2.3.1 Embedding Data in Documents

Also known as nested document.

- Avoids application joins.
- Provides better performance for read operations.
- Allows developers to update related data in a single write operations.

Issues:

- Embedding data into a single document can create large documents.
- Large documents have to be read into memory in full, which can result in a slow application performance for you end user.
- Continuesly adding data without limit creates **unbounded documents**.
- Unbounded Documents may exceed the BSON document threshold of 16MB. This is schema anti-patterns which you should avoid.

2.3.2 Referencing Data in Documents

When we want to store data in two different collections but also ensure that those collections are related, we can use **References**. References save the *_id* field of one document in another document as a link between the two. They are simple and sufficient for most use cases. Using references is called **linking** or **data normalization**. It avoids duplication of data and in most cases smaller documents, however referencing data you'll need to query from multiple documents costing extra resources and impacts read performance.

Embedding	Referencing
✓ Single query to retrieve data	✓ No duplication
✓ Single operation to update/delete data	✓ Smaller Documents
! Data Duplication	! Need to join data from multiple documents
! Large documents	

2.4 Scaling a Data Model

Optimum efficiency of:

- query result depends on
- memory usage
- CPU usage
- storage
- When creating a data model, avoid documents that are unbounded. Document size grows infinitely. Problems as the array grows larger in embedded document.
 - It will take up more space in memory
 - may impact write performance as entire document is rewritten into MongoDB data storage.
 - it will be difficult to perform pagination of that array.
 - Maximum document size of 16MB will lead to storage problems.
- benefit is that we can retrieve all document in single read.
- Avoid:
 - More than the document size limit of 16 MB.
 - Poor query performance.
 - Poor write performance.
 - Too much memory being used.

2.5 Using Atlas Tools for Schema Help

Schema design patterns are guidelines that help developers plan, organize, and model data. Schema anti-patterns result in suboptimal performance and non-scalable solutions. Eg:

- massive arrays.
- Massive number of collections.
- Bloated documents.
- Unnecessary indexes.

- Queries without indexes.
- Data that's accessed together, but stored in different collections.

Tools in Atlas are **Data Explorer** and **Performance Advisor**

Chapter 3

Connection to MongoDB Database

3.1 Using MongoDB Connection Strings

Connection string can be used to connect from MongoShell, MongoDB Compass, or any other application. It provides standered fromat and DNS seed list format.

Standered Format

Used to connect to standalone clusters, replica sets, or sharded clusters.

DNS Seed List Format

Released in MongoDB 3.6.

3.1.1 Connection String

It allows to provide a DNS server list to connection string. Gives more flexibility of deployment. Ability to change servers in rotation without reconfiguring clients. In atlas it is present in Database tab → Connect → Connect to application.

Connection String: *mongodb+srv://<username>:<password>@cluster0.usqsf.mongodb.net/?retryWrites=true&w=majority*

If port number is not specified MongoDB will default to 27017.

3.2 Connecting to a MongoDB Atlas Cluster with the Shell