

Bài giảng: Index trong MongoDB

Tác giả: Đặng Kim Thi

1. Giới thiệu về Index trong MongoDB

1.1. Index là gì?

- Index (chỉ mục) là một cơ chế giúp MongoDB nhanh chóng truy xuất dữ liệu.
- Tương tự như trang mục lục trong sách, thay vì quét từng trang, ta có thể nhảy ngay tới trang mong muốn.
- **Key takeaway:** Index cải thiện hiệu suất truy vấn khi dữ liệu lớn.

1.2. Tại sao cần index?

- **Quá trình quét toàn bộ (COLLSCAN):** Khi không có index, MongoDB phải duyệt qua tất cả tài liệu trong collection.
- **Index giúp giảm tải:**
 - Chỉ duyệt các tài liệu phù hợp.
 - Tăng tốc đối với các collection có hàng triệu tài liệu.

1.3. Nhược điểm của Index

- **Lưu trữ:** Index chiếm thêm không gian bộ nhớ.
 - **Hiệu năng ghi:** Thêm, xóa hay sửa dữ liệu yêu cầu cập nhật index.
-

2. Các loại Index trong MongoDB

2.1. Single Field Index

- **Khái niệm:** Index đơn cho một trường.
- **Ứng dụng:**
 - Tìm sinh viên theo "tên" nhanh chóng.
- **Cú pháp:**

```
db.collection.createIndex({ field: 1 }); // Tăng dần
db.collection.createIndex({ field: -1 }); // Giảm dần
```

- **Ví dụ:**

```
db.students.createIndex({ name: 1 });
```

2.2. Compound Index

- **Khái niệm:** Index cho nhiều trường.
- **Lợi ích:** Giảm thời gian quét khi truy vấn về nhiều tiêu chí.
- **Cú pháp:**

```
db.collection.createIndex({ field1: 1, field2: -1 });
```

- **Ví dụ:**

```
db.orders.createIndex({ customerId: 1, orderDate: -1 });
```

2.3. Multikey Index

- **Khái niệm:** Dùng cho trường chứa dữ liệu mảng.
- **Ví dụ:**

```
db.products.createIndex({ tags: 1 });
```

- **Giải thích:**
 - Mỗi giá trị trong mảng tags sẽ được index.

2.4. Text Index

- **Chức năng:** Tìm kiếm nhanh các chuỗi text.
- **Ví dụ:**

```
db.articles.createIndex({ content: "text" });  
db.articles.find({ $text: { $search: "mongodb indexing" } });
```

- **Giải thích:** Tìm kiếm trong nội dung bài viết.

2.5. Geospatial Index

- **Chức năng:** Hỗ trợ tìm kiếm theo toạ độ và địa lý.
- **Cú pháp:**

```
db.collection.createIndex({ location: "2dsphere" });
```

2.6. TTL (Time to Live) Index

- **Chức năng:** Xóa tài liệu tự động khi hết hạn.
- **Ví dụ:**

```
db.sessions.createIndex({ createdAt: 1 }, { expireAfterSeconds: 3600
});
```

3. Case Study: Phân tích hiệu suất với dữ liệu phức tạp

3.1. Kịch bản thực tế

Một công ty thương mại điện tử có một collection **orders** chứa thông tin như sau:

```
db.orders.insertMany([
  { orderId: 1, customerId: 101, total: 150, items: ["Laptop", "Mouse"],
    orderDate: new Date("2025-01-01") },
  { orderId: 2, customerId: 102, total: 200, items: ["Phone", "Charger"],
    orderDate: new Date("2025-01-02") },
  { orderId: 3, customerId: 103, total: 100, items: ["Tablet"], orderDate:
    new Date("2025-01-03") },
  // Thêm 1 triệu bản ghi ngẫu nhiên...
]);
```

3.2. Truy vấn không có Index

- **Yêu cầu:** Tìm tất cả các đơn hàng có tổng tiền lớn hơn 150.

```
db.orders.find({ total: { $gt: 150 } }).explain("executionStats");
```

- **Kết quả:**

```
{
  "executionStats": {
    "totalDocsExamined": 1000000,
    "executionTimeMillis": 350
  }
}
```

- MongoDB phải quét toàn bộ 1 triệu bản ghi để tìm các đơn hàng phù hợp.

3.3. Truy vấn có Index

- **Tạo Index:**

```
db.orders.createIndex({ total: 1 });
```

- **Thực hiện truy vấn:**

```
db.orders.find({ total: { $gt: 150 } }).explain("executionStats");
```

- **Kết quả:**

```
{
  "executionStats": {
    "totalDocsExamined": 100,
    "executionTimeMillis": 5
  }
}
```

- MongoDB chỉ quét các bản ghi có giá trị **total** phù hợp nhờ Index.

3.4. So sánh hiệu suất

Truy vấn	Không có Index	Có Index
Số tài liệu quét	1,000,000	100
Thời gian thực thi	350 ms	5 ms

3.5. Kết luận

- Index giúp giảm đáng kể số lượng tài liệu cần duyệt và thời gian thực thi.
- Trong các ứng dụng thực tế với dữ liệu lớn, Index là không thể thiếu.

4. Bài tập hướng dẫn

4.1. Bài tập nâng cao: Tạo và So sánh Index

1. Tạo collection **sales** với dữ liệu:

```
db.sales.insertMany([
  { saleId: 1, product: "Laptop", amount: 1200, region: "North", date:
    new Date("2025-01-10") },
  { saleId: 2, product: "Phone", amount: 800, region: "South", date:
    new Date("2025-01-11") },
  { saleId: 3, product: "Tablet", amount: 600, region: "East", date:
    new Date("2025-01-12") },
])
```

```
// Thêm 500,000 bản ghi ngẫu nhiên...  
]);
```

2. Tổ chức truy vấn:

◦ Không có index:

```
db.sales.find({ amount: { $gte: 1000 }  
}).explain("executionStats");
```

◦ Có index:

```
db.sales.createIndex({ amount: 1 });  
db.sales.find({ amount: { $gte: 1000 }  
}).explain("executionStats");
```

3. So sánh thời gian thực thi và số lượng tài liệu được quét.

Tác giả: Đặng Kim Thi