

به نام یکتا



طراحی پایگاه داده‌ها

تمرین ۵

استاد

مهدی آخی

تهیه و تدوین

تیم دستیاران درس - بخش تمارین

بهار ۱۴۰۳

فهرست

2.....	ضوابط تمرین
2.....	سیاست‌های جزئی تمرین ۵
3.....	سوال اول (۲۵ نمره)
4.....	سوال دوم (۲۵ نمره)
5.....	سوال سوم (۵۰ نمره)
6.....	سوال چهارم (۳۰ نمره)
9.....	سوال پنجم (۴۰ نمره)

ضوابط تمرین

- پرسش و پاسخ

- برای هر تمرین یک پست در کوئرا ایجاد خواهد شد که می‌توانید سوالات و ابهامات احتمالی خود را در زیر همان پست بپرسید.

- سیاست تاخیر

- تاخیر به صورت ساعتی محاسبه می‌شود، هر تمرین را تا حداکثر ۲ روز پس از ددلاین می‌توانید ارسال کنید.
- در مجموع هر فرد 5 روز تاخیر مجاز دارد.
- به ازای هر ساعت تاخیر غیر مجاز ۲ درصد از نمره‌ی آن تمرین کم خواهد شد.

- سیاست تقلب

- در صورت مشاهده هرگونه مشابهت نامتعارف، بار اول نمره‌ی کل تمرین صفر شده و بار دوم ادامه‌ی درس میسر نخواهد بود. (برای هر دو طرف درگیر)
- مشورت و ایده گرفتن از یکدیگر در صورتی که راه حل دیگری را مشاهده نکنید ایرادی ندارد اما باید در پاسخستان صراحتاً ذکر شود.

سیاست‌های جزئی تمرین ۵

- این تمرین مربوط به اسلایدهای NoSQL درس می‌باشد.
- ددلاین ارسال پاسخ: 2 تیر
- نمره‌ی تمرین: 1.5 نمره
- برای گرفتن نمره‌ی کامل تمرین تنها به 150 نمره نیاز دارید. (20 نمره اضافی است)

سوال اول (۲۵ نمره)

MongoDB به عنوان یک پایگاه داده NoSQL، ساختارهای داده‌ای متفاوتی را ارائه می‌دهد که هر کدام می‌توانند برای نیازهای خاصی به کار روند. یکی از این ساختارها، Capped Collection است.

- تعریف کنید Capped Collection چیست و چه خصوصیات منحصر به فردی دارد که آن را از سایر collectionها در MongoDB متمایز می‌سازد.
- بررسی کنید استفاده از Capped Collections چه مزایا و معایبی دارد. تمرکز خود را بر روی عملکرد و محدودیت‌های مدیریت داده‌ها قرار دهید.
- در چه سناریوهای کاربردی استفاده از Capped Collections توصیه می‌شود؟ به توضیحات خود مثال‌های عملی اضافه کنید.
- نحوه ایجاد یک Capped Collection جدید در MongoDB را شرح دهید. چه پارامترهایی باید هنگام ایجاد این نوع Collection تعیین شود و چه تأثیراتی بر روی عملکرد و مدیریت داده‌های آن خواهد داشت؟

سوال دوم (۲۵ نمره)

MongoDB به عنوان یک سیستم پایگاه داده NoSQL، ابزارها و تکنیک‌های مختلفی را برای مدیریت داده‌ها فراهم می‌کند، که شامل backup, export, import و restore می‌شود. این امکانات به خصوص برای تضمین دسترسی پایدار و امن به داده‌ها حیاتی هستند.

- توضیح دهید MongoDB چگونه امکان import و export داده‌ها را فراهم می‌کند. مشخص کنید کدام ابزارها و فرمت‌های داده‌ای برای این منظور استفاده می‌شوند.
- بررسی کنید چه روش‌هایی برای backup گیری از داده‌ها در MongoDB وجود دارد. توضیح دهید هر یک از این روش‌ها چه مزایا و معایبی دارند و در چه شرایطی توصیه می‌شوند.
- شرح دهید چگونه می‌توان داده‌ها را در MongoDB restore کرد. این فرایند چه تأثیری بر روی دسترسی به داده‌ها و عملکرد کلی سیستم دارد؟
- ارزیابی کنید که چگونه این ابزارها و تکنیک‌ها می‌توانند به بهبود فرایندهای نگهداری داده‌ها و اطمینان از دسترسی سریع و مطمئن به داده‌ها کمک کنند.

سوال سوم (۵۰ نمره)

یک فایل دیتاست به نام `AirlineTwitterSentiment.json` در اختیار شما قرار داده شده است که حاوی داده‌های مربوط به توییت‌های کاربران درباره شرکت‌های هواپیمایی مختلف است. از این داده‌ها می‌خواهیم تجزیه و تحلیل‌های مختلفی را انجام دهیم:

- **نمایش تمام توییت‌ها:** دستوری بنویسید که تمام مستندات توییت‌ها را نمایش دهد.
- **توییت‌های منفی:** دستوری بنویسید که تمام توییت‌های با احساس منفی را نمایش دهد.
- **توییت‌ها بر اساس شرکت و احساس:** دستوری بنویسید که توییت‌های مثبت مربوط به شرکت خاصی (مثل Delta) را نمایش دهد.
- **تجزیه و تحلیل احساسات:** دستوری بنویسید که تعداد توییت‌های مثبت، منفی و خنثی برای هر شرکت هواپیمایی را محاسبه کند.
- **جستجو بر اساس کلمات کلیدی:** دستوری بنویسید که توییت‌هایی را که شامل کلمات کلیدی خاصی (مثل «تأخیر» یا «لغو») هستند، فیلتر و نمایش دهد.
- **میانگین زمان پاسخ:** دستوری بنویسید که میانگین زمان پاسخ شرکت‌های هواپیمایی به توییت‌ها را محاسبه کند.

سوال چهارم (۳۰ نمره)

یک پایگاه داده‌ی NoSQL را در نظر بگیرید که داده‌ی زیر، یک نمونه از اعضای این پایگاه داده است. با توجه به ساختار این عضو، مشخص کنید هرکدام از دستورات داده شده به زبان MongoDB چه کاری انجام می‌دهند. نحوه‌ی عملکرد هرکدام از آن‌ها را به صورت کامل توضیح دهید:

```
{
  "_id": ObjectId("61548f58009b8a95e0c1eb2d"),
  "name": "John Smith",
  "email": "john.smith@example.com",
  "age": 35,
  "address": {
    "street": "123 Main St",
    "city": "Anytown",
    "state": "CA",
    "zip": "12345"
  },
  "friends": [
    {
      "name": "Jane Doe",
      "email": "jane.doe@example.com"
    },
    {
      "name": "Bob Johnson",
      "email": "bob.johnson@example.com"
    }
  ],
  "createdAt": ISODate("2021-09-29T20:15:04Z")
}
```

- `db.students.find({
 "age": 20, "friends.name": "Jane Doe"
})`
- `db.students.find({
 "createdAt": {
 $gt: ISODate("2021-01-01T00:00:00Z")
 }
})`
- `db.students.find({
 "address.city": "Anytown"
}, {
 "name": 1, "email": 1, "_id": 0
})`
- `db.students.find({
 "name": {
 $regex: /^J/
 }
})`
- `db.students.find({
 "address.state": "CA"
}).sort({
 "age": 1
})`
- `db.students.find().sort({
 "age": 1
}).limit(1)`
- `db.students.find({
 "email": {
 $exists: false
 }
})`

- `db.students.find({
 "friends": {
 $size: {
 $gt: 1
 }
 }
})`
- `db.students.find({
 "age": {
 $gte: 20, $lte: 30
 }
})`
- `db.students.aggregate([{
 $group: {
 _id: null, averageAge: {
 $avg: "$age"
 }
 }
}])`
- `db.students.aggregate([{
 $project: {
 name: 1, email: 1, numberOfFriends: {
 $size: "$friends"
 }, nameLength: {
 $strLenCP: "$name"
 }
 }
}])`

سوال پنجم (۴۰ نمره)

به سوالات زیر پاسخ کامل دهید و پاسخ هرکدام را به مقدار لازم شرح دهید:

- چهار نوع اصلی از پایگاه‌های داده‌ی NoSQL (مستندگرا، ستون‌گرا، گراف، و کلید-مقدار) را توضیح دهید و برای هر کدام یک مثال کاربردی ارائه دهید.
- چرا پایگاه‌های داده‌ی NoSQL معمولاً برای برنامه‌هایی که نیاز به مقیاس‌پذیری بالا دارند انتخاب می‌شوند؟ مثال بزنید.
- محدودیت‌های پایگاه‌های داده‌ی NoSQL در تجزیه و تحلیل داده‌های پیچیده و تراکنش‌ها چیست؟ چه راه‌حلهایی برای این محدودیت‌ها پیشنهاد می‌دهید؟
- اهمیت پایگاه‌های داده توزیع‌شده در معماری NoSQL چیست و چه مزایایی دارند؟ چالش‌های رایج در مدیریت آنها چیست؟
- مفهوم CAP theorem را توضیح دهید و بگویید چگونه پایگاه‌های داده‌ی NoSQL این سه معیار را متعادل می‌کنند؟
- مزایا و معایب شاخص‌گذاری در پایگاه‌های داده‌ی NoSQL چیست و چگونه می‌توان آن را بهینه کرد؟
- چگونه پایگاه‌های داده‌ی NoSQL با تغییرات سریع در ساختار داده‌ها سازگار می‌شوند و این ویژگی چه مزایایی به همراه دارد؟
- مزایای استفاده از قوام نهایی (eventual consistency) در پایگاه‌های داده‌ی NoSQL چیست و در چه مواردی ممکن است این رویکرد مناسب نباشد؟
- تفاوت‌های اصلی بین مدل‌های ACID و BASE را توضیح دهید و مثال‌هایی از پایگاه‌های داده‌ی NoSQL که هر یک از این مدل‌ها را دنبال می‌کنند، بیاورید.
- چگونه پایگاه‌های داده‌ی NoSQL می‌توانند برای ذخیره و تحلیل داده‌های جغرافیایی (geospatial data) استفاده شوند؟ مزایا و چالش‌های مرتبط با این کاربرد چیست؟