



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

به نام خدا

آزمایشگاه پایگاه داده
جلسه هشتم
آشنایی با پایگاه داده‌های
NoSQL

NoSQL

- ▶ پایگاه داده‌های NoSQL (مخفف عبارت Not only SQL یا Non-SQL) به اصطلاح یک پایگاه داده‌ی غیر رابطه‌ای (non relational) است که امکانی برای ذخیره و بازیابی اطلاعات به شکلی متفاوت از پایگاه داده‌های رابطه‌ای ایجاد می‌کند.
- ▶ در پایگاه داده‌های رابطه‌ای مرسوم از جبر رابطه‌ای استفاده می‌شود و در آن با جدول سر و کار داریم اما در پایگاه داده‌های غیر رابطه‌ای از داده‌هایی به شکل JSON و XML استفاده می‌کنیم که در بسیاری از مسائل متمرکز بر داده (data intensive) که سرعت و کارایی بالایی را نیاز دارد، عملکرد مناسبی دارد.

انواع NoSQL

- پایگاه داده‌های NoSQL کلید-مقدار (Key-value NoSQL):

▶ در اینگونه پایگاه داده‌ها از یک کلید (Key) به عنوان شناسه‌ی هر داده برای ذخیره و بازیابی مقدار داده (Value) استفاده می‌شود. این دسته به علت سادگی کارکرد مرسوم‌ترین نوع پایگاه داده‌های NoSQL می‌باشد.

- پایگاه داده‌های NoSQL مبتنی بر اسناد (Document NoSQL):

▶ این نوع از پایگاه داده‌ها به منظور ذخیره‌سازی و کار با اسنادی با فرمت های XML, JSON ... به کار می‌روند. از پایگاه داده‌های اسنادی NoSQL جهت ذخیره‌سازی داده های بدون ساختار مشخص با پراکندگی بالا استفاده می‌شود.

- پایگاه داده‌های NoSQL چند ستونه (Wide-column NoSQL):

▶ برخلاف SQL هر ستون می‌تواند شامل داده‌هایی با فرمت و ساختار متفاوت باشد. به عبارتی دیگر نوع تعریف و فرمت یک ستون می‌تواند در هر سطر متفاوت باشد. این پایگاه داده‌ها انعطاف بسیار بالایی در ثبت و کار با داده‌های بسیار عظیم و متفاوت دارند.

انواع NoSQL

- پایگاه داده‌های مبتنی بر گراف (Graph NoSQL):

- ▶ پایگاه داده‌های گرافی به منظور ذخیره سازی حجم زیادی از داده‌های رابطه‌ای (Relational data) طراحی شده‌اند. به زبان ساده می‌توان اینگونه پایگاه داده‌ها را مانند گرافی شامل رئوسی به عنوان داده‌ها و یال‌هایی به عنوان رابطه‌ها در یک گراف در نظر گرفت. از این دسته پایگاه داده‌ها در ذخیره‌سازی انواع معماری‌های داده‌های شبکه‌ای نیز استفاده می‌شود.

- پایگاه داده‌های NoSQL چند مدل (MultiModel NoSQL):

- ▶ پایگاه داده‌های چند مدل می‌توانند تلفیقی از انواع دیگر پایگاه داده‌های NoSQL باشند. (مانند Key-value و مبتنی بر گراف)

برخی از پایگاه داده‌های NoSQL

• Aerospike:

► یک پایگاه داده Key-value بسیار قدرتمند برای کلاسترینگ است که از لحاظ حجم پردازش داده و سرعت بالاترین رتبه را در Benchmark به دست آورده است. ذخیره‌سازی اطلاعات در این پایگاه داده به صورت ادغامی از Ram و SSD صورت می‌گیرد.

• Redis:

► یک پایگاه داده Key-value که برای حجم کار کوچک‌تر به نسبت aerospike مناسب می‌باشد. تمامی اطلاعات این پایگاه داده در Ram ذخیره می‌شوند. این پایگاه داده قابلیت کلاستر شدن ندارد.

• MongoDB:

► مونگو معروف‌ترین پایگاه داده NoSQL است و برای ذخیره‌سازی اسناد (Documents) استفاده می‌شود. اگر نیاز به ذخیره‌سازی حجم بالایی از داده‌های مختلف و پراکنده را دارید مونگو گزینه مناسبی برای شماست. کار با مونگو به نسبت ساده است و برای شروع کار و آشنایی با NoSQL ها گزینه‌ی مناسبی است.

برخی از پایگاه داده‌های NoSQL

• Cassandra:

► این پایگاه داده از قوی‌ترین پایگاه داده‌های NoSQL می‌باشد و توسط Apache ارائه شده است. پایگاه داده Cassandra در دسته Wide-column قرار دارد. اگر چه سرعت پردازش Cassandra از Aerospike بیشتر است اما قابلیت‌های فراوان و انعطاف بالای Cassandra به عقیده بسیاری آن را قدرتمندترین NoSQL کرده است.

• Neo4j:

► پایگاه داده‌ی بر پایه گراف (Graph) که برای ذخیره سازی ساختمان داده‌های مختلف شبکه‌ای و داده‌های رابطه‌ای بسیار مناسب است.

مقایسه MongoDB و RDBMS

RDBMS	MongoDB
پایگاه داده	پایگاه داده
جدول	مجموعه (Collection)
Tuple / ردیف	سند
ستون	فیلد
جدول پیوست	اسناد پیوست شده
کلید عمومی	کلید عمومی (شناسه کلید عمومی به صورت پیش فرض توسط خود mongodb ارائه شده است)

نحوه‌ی مدل کردن داده در MongoDB

▶ این پایگاه داده همان‌طور که پیش‌تر نیز مطرح شد، یک ساختار ذخیره‌سازی سندگرا است که از قالب BSON (مخفف JSON باینری) برای نمایش سندهای خود بهره می‌برد.

▶ قالب BSON از سوی دیگر، همان قالب JSON است که به‌صورت باینری کدگذاری شده و علاوه‌بر تمامی توانمندی‌ها و کارکردهای قالب JSON، ویژگی‌های منحصربه‌فرد دیگری مانند انواع داده‌ای بیش‌تر، از جمله Date و داده باینری (BinData) که جزء مشخصات JSON نبوده اما برای کارکرد MongoDB از اهمیت بالایی برخوردار هستند را در خود دارد.

▶ ذکر این نکته ضروری است که BinData در واقع معادل نوع داده BLOB در پایگاه‌های داده رابطه‌ای است که برای ذخیره‌سازی محتوای فایل‌ها درون پایگاه داده‌ها مورد استفاده قرار می‌گیرد.

نحوه‌ی مدل کردن داده در MongoDB

► به هر سند در پایگاه داده MongoDB مجموعه‌ای از جفت‌های فیلد-مقدار است که مقدارها می‌تواند یکی از انواع زیر باشد:

- یک نوع BSON
- یک سند
- آرایه‌ای از هر نوع BSON
- آرایه‌ای از سند

► بنابراین، در این پایگاه داده این امکان وجود دارد که یک سند شامل سندهای دیگری باشد. این سندهای درونی بسیار قدرتمند بوده و در عین حال، انعطاف فراوانی را برای شِماهای آن با خود به‌همراه دارند، که البته باید با دقت فراوانی مورد استفاده قرار گیرند؛ چرا که حداکثر اندازه یک سند به طور معمول شانزده مگابایت است.

نحوه‌ی مدل کردن داده در MongoDB

- ▶ پایگاه داده MongoDB با مجموعه‌ها یا Collections کار می‌کند؛ که معادل جدول‌ها در پایگاه‌های داده رابطه‌ای بوده و سند، معادلی برای رکوردهای جدول‌ها در این رده از پایگاه است.
- ▶ به عبارت دیگر هر پایگاه داده MongoDB، متشکل از مجموعه‌هایی است که هر مجموعه خود شامل سندهایی -رکوردها- هستند که همان‌طور که در بخش قبل به آن اشاره شد، هر یک از این سندها می‌توانند خود مجموعه‌ای از سندها را در خود جای دهند.
- ▶ از آن‌جا که هر سند متشکل از یک جفت فیلد اطلاعاتی و مقدار متناظر با آن است، نام مناسب برای شناسایی ویژگی هر سند به‌طور مشخص نام فیلد است که درواقع معادل نام ستون است، که در پایگاه‌های داده رابطه‌ای مورد استفاده قرار می‌گیرد.
- ▶ همچنین هر سند، شامل یک کلید اصلی است که به‌صورت خودکار به فیلد id تخصیص داده می‌شود.

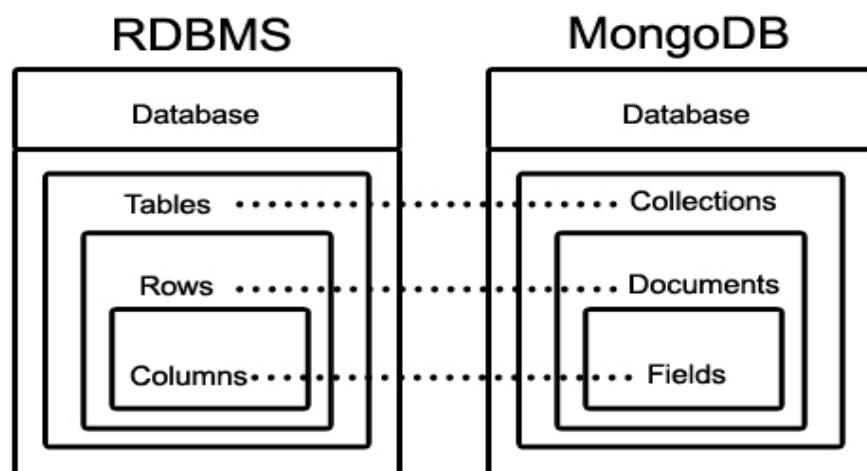
نحوه‌ی مدل کردن داده در MongoDB

► نکته‌ی حائز اهمیت درمورد مجموعه‌ها در پایگاه داده غیر رابطه‌ای MongoDB آن است که اگرچه مجموعه‌ها، به‌نوعی با مفهوم جدول‌ها هم‌ارز هستند که سندها را در خود جای می‌دهند، اما واقعیت آن است که این مجموعه‌ها هیچ ساختاری را بر سندها تحمیل نمی‌کنند و این بزرگ‌ترین تفاوت این رده از پایگاه‌های داده با انواع رابطه‌ای آن است.

► به‌عبارت دیگر، برخلاف جدول‌های پایگاه‌های رابطه‌ای که هر رکورد آن، ستون‌های یکسانی دارند، سندهای موجود در یک مجموعه می‌توانند ساختارهای – فیلدهای اطلاعاتی – کاملاً متفاوتی داشته باشند. اگرچه در کاربردهای معمول، سندهای موجود در یک مجموعه، معمولاً ساختار پایه‌ای مشترکی دارند.

نحوه‌ی مدل کردن داده در MongoDB

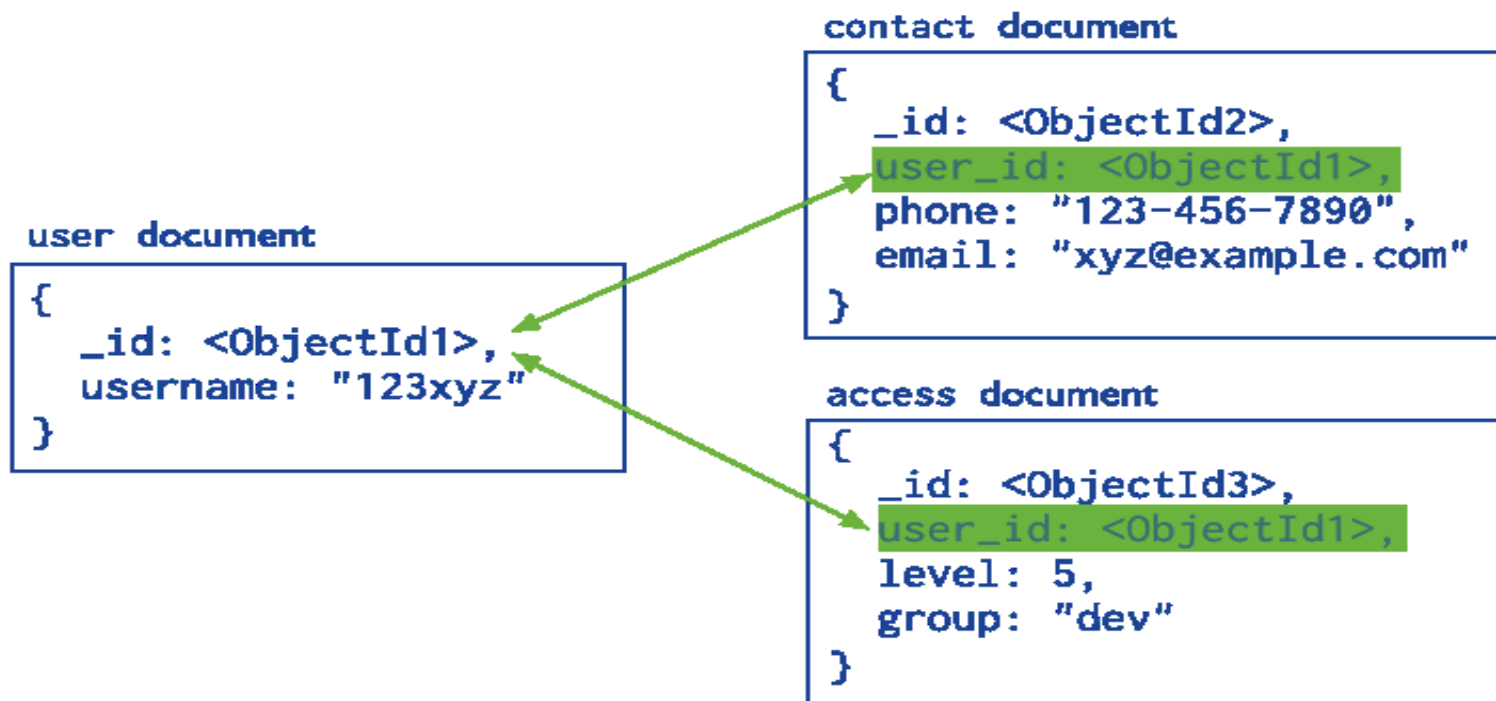
► به‌طور خلاصه می‌توان ساختار پایگاه داده‌ای MongoDB را متشکل از مجموعه‌هایی دانست که هر مجموعه شامل سندهایی است و هر سند، معرف یک زوج فیلد اطلاعاتی و مقدار فیلد است که مقدار می‌تواند یکی از مقادیر مطرح شده در بخش‌های قبل را در خود جای داده و ساختار سند‌گرایی تو در تو را فراهم آورد. هر سند با نام فیلد آن شناخته شده و کلید اصلی آن به‌صورت خودکار تخصیص داده می‌شود.



انواع مدل داده در MongoDB

مدل داده‌ی نرمالسازی شده (Normalized) یا ارجاعی

► در این شیوه ارتباطات میان سندها از طریق ایجاد یک اتصال میان سندها برقرار می‌شود. در واقع با افزودن یک شناسه‌ی یک سند درون سند دیگر یک اشاره‌گر به سند بالادستی ایجاد می‌شود.



انواع مدل داده در MongoDB

مدل داده‌ی جاسازی شده (Embedded)

► در این شیوه سندها به صورت تو در تو ایجاد می‌شوند. در واقع سندی که قرار است به سند دیگری متصل شود، مستقیماً به عنوان یک جزء در سند دیگر قرار داده می‌شود.

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

Embedded sub-document

Embedded sub-document

تمرین

1. مزایای پایگاه داده‌های NoSQL نسبت به SQL چیست؟
2. در چه صورت از هر کدام از روش‌های مدل‌سازی داده (Embedded و Normalized) استفاده می‌کنیم؟ به صورت خلاصه بنویسید.