



T.C.
İNÖNÜ ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ
BÜYÜK VERİ TEKNOLOJİLERİ PROJE RAPORU

PROJE EKİBİ:

Ahmet GÜR - 02200201053
Ahmet TAKCI - 02200201055
Nermin UZAY - 02200201019
Ayşe DEVEDEN - 02200201081
Dilara KARATAŞ - 02200201044
Şeymanur ACABEY - 02200201080

İÇİNDEKİLER

MONGODB NEDİR?	3
MongoDB'nin Temel Özellikleri :	3
NOSQL NEDİR?	4
MONGODB MİMARİSİNİN TEMEL BİLEŞENLERİ	5
MONGODB KURULUMU NASIL YAPILIR?.....	6
KODLAMALAR	8

MONGODB NEDİR?

MongoDB, bir belge tabanlı (document-oriented) NoSQL veritabanı yönetim sistemidir. NoSQL (Not Only SQL) terimi, geleneksel SQL tabanlı ilişkisel veritabanlarına alternatif olarak geliştirilen veritabanı türlerini ifade eder.

MongoDB, verileri JSON benzeri belgeler (documents) olarak depolar. Her belge, alan-değer çiftlerinden oluşur ve MongoDB'nin BSON (Binary JSON) formatında saklanır. BSON, JSON benzeri bir veri formatıdır, ancak daha fazla veri türünü destekler ve binary (ikili) olarak depolanır.

MongoDB'nin Temel Özellikleri :

- 1. Esnek Şema:** MongoDB, esnek şema yapısıyla çalışır. Bu, veritabanında depolanacak verinin yapısını önceden tanımlamanız gerekmeyen anlamına gelir. Her belge kendi yapısını tanımlar ve belgeler arasında farklı alanlar ve veri tipleri bulunabilir.
- 2. Yüksek Performans:** MongoDB, verileri belgeler halinde sakladığı için verilere hızlı bir şekilde erişim sağlar. Ayrıca, sorguları optimize etmek için indeksleme ve diğer performans ayarları sunar.
- 3. Yüksek Ölçeklenebilirlik:** MongoDB, yatay ölçeklenebilirlik için tasarlanmıştır. Veritabanını birden fazla sunucuya dağıtarak performansı artırabilir ve yüksek trafikli uygulamalara kolayca uyum sağlayabilirsiniz.
- 4. Replication ve Yedekleme:** MongoDB, replikasyon ve yedekleme özelliklerini destekler. Bu özellikler sayesinde verilerinizi yedekleyebilir, yüksek kullanılabilirlik sağlayabilir ve veri kaybı riskini azaltabilirsiniz.
- 5. Zengin Sorgu Dili:** MongoDB, zengin bir sorgu dili sunar. Bu dille karmaşık sorgular oluşturabilir, filtreleme, sıralama, grublama, birleştirme gibi işlemleri gerçekleştirebilirsiniz.

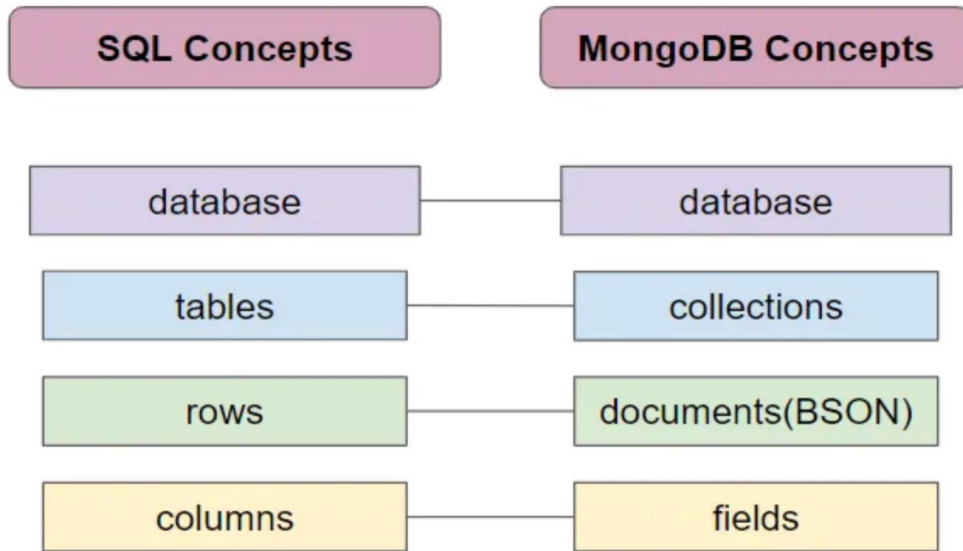
MongoDB, ölçeklenebilirlik, esneklik ve yüksek performans gibi özellikleriyle özellikle büyük ve karmaşık veri setleriyle çalışan uygulamalar için tercih edilmektedir.

NOSQL NEDİR?

Son yıllarda verinin inanılmaz boyutlara ulaşması ve katlanarak artması sonucunda mevcut olarak kullanılan ilişkisel veritabanı sistemleri yerine ortaya atılmış bir kavramdır. İlişkili veritabanı sistemleri ile arasındaki en büyük fark ilişkisel veritabanı sistemlerinde veriler tablo ve sütunlar ile ilişkili bir şekilde tutulurken NoSQL’de json bir yapıda tutulmasıdır.

NoSQL sistemlerin avantajlarına değinmek gerekirse ilk olarak performans gösterilebilir. Okuma ve yazma işlemleri ilişkisel veritabanılarına göre çok daha hızlı olmaktadır. İkinci olarak ise NoSQL sistemler yatay olarak genişletilebilirler. Binlerce sunucu bir arada çalışarak inanılmaz derecedeki veriler üzerinde işlemler yapılabilir.

Ayrıca eklemekte fayda var ki günümüzde Büyük Veri alanında yapılan çalışmalarda NoSQL sistemler yoğun olarak kullanılmaktadır.



MONGODB MİMARİSİNİN TEMEL BİLEŞENLERİ

Belge Yapısı (Document Structure): MongoDB'de veri, JSON benzeri BSON belgeleri içinde saklanır. Her belge, alan-değer çiftleri içerir ve JSON benzeri bir formata sahiptir. Belge tabanlı yapı, esneklik sağlar ve aynı koleksiyon içinde farklı alanlara sahip belgelerin saklanabilmesine olanak tanır.

Koleksiyonlar (Collections): Koleksiyonlar, belgelerin gruplandığı yapıdır. Her bir koleksiyon, benzer veri türlerini veya mantıksal olarak birbirine bağlı belgeleri içerir. Örneğin, bir "users" koleksiyonu kullanıcı bilgilerini içerebilir.

Sharding: MongoDB sharding, büyük veri setlerini yönetmek ve sorgu performansını artırmak için kullanılan bir tekniktir. Veri kümesi parçalara bölünerek farklı düğümlere dağıtılır. Bu, yüksek performans ve ölçeklenebilirlik sağlar.

Replikasyon (Replication): Replikasyon, birincil ve ikincil düğümler arasında veri kopyalarının tutulmasını sağlar. Bu, yüksek kullanılabilirlik ve veri güvenliği sağlar. Birincil düğüm, yazma işlemleri için sorumludur, ikincil düğümler ise veri kopyalarını içerir ve başka bir düğümün yerine geçebilir.

Endeksleme (Indexing): MongoDB, belirli alanlarda endeksleme yaparak sorgu performansını artırır. Endeksler, belirli alanlarda hızlı erişim sağlar ve sorguların daha etkili çalışmasını sağlar.

Aggregation Framework: Aggregation Framework, MongoDB'nin güçlü sorgulama ve analiz yeteneklerini sağlayan bir özelliktir. Bu, veritabanındaki belgeleri filtrelemek, düzenlemek, gruplamak ve analiz etmek için kullanılır.

Gizleme (Capping): MongoDB'de belirli bir koleksiyonun maksimum belge sayısını sınırlamak için kullanılır. Bu, belirli bir koleksiyondaki belgelerin sınırlı bir süre boyunca tutulmasına olanak tanır.

Küme (Cluster): MongoDB kümesi, birden çok MongoDB sunucusundan oluşan bir yapıdır. Bu sunucular, veri dağıtımı, yedekleme ve yüksek kullanılabilirlik gibi amaçlar için bir araya getirilebilir. Bu temel bileşenler, MongoDB'nin esnek, ölçeklenebilir ve performans odaklı mimarisini oluşturan ana unsurlardır. Bu bileşenlerin kombinasyonu, çeşitli kullanım senaryolarına ve büyük veri işleme ihtiyaçlarına uyum sağlar.

MONGODB KURULUMU NASIL YAPILIR?

1. Resmi web sitesine gidilir. <https://www.mongodb.com/try/download/community>
2. İndirme sayfasında, işletim sisteminize uygun olan MongoDB Community Server sürümünü seçin ve indirin. İşletim sisteminiz Windows ise, MSI dosyasını indirin. Linux kullanıyorsanız, uygun paketi indirin.
3. Kurulum sihirbazında, varsayılan ayarları genellikle kabul edebilirsiniz. İsteğe bağlı olarak, kurulum dizini veya veritabanı dosyalarının konumunu değiştirebilirsiniz.
4. Kurulum tamamlandığında, MongoDB hizmeti otomatik olarak başlatılacaktır. Eğer başlatılmazsa, başlatılmasını sağlamak için MongoDB hizmetini manuel olarak başlatabilirsiniz.

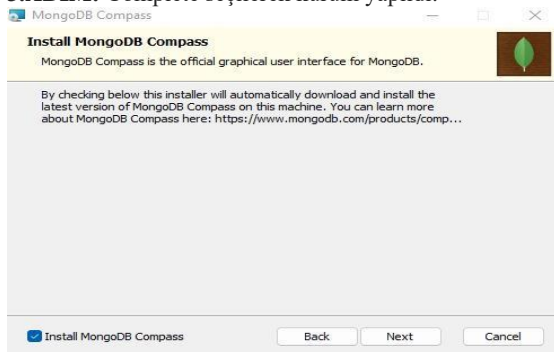
1.ADIM: MONGODB.com adresine gidilerek 7.05 version indirildi.



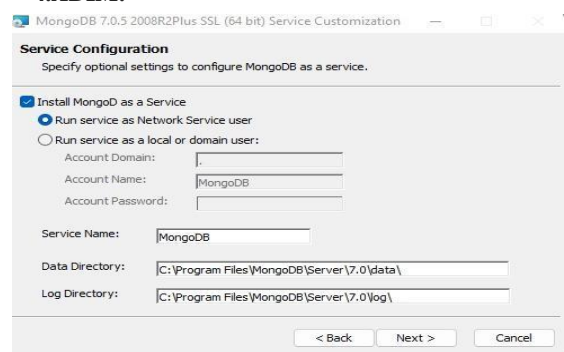
2.ADIM: İndirme işlemi bittikten sonra kurulumu başlanır.



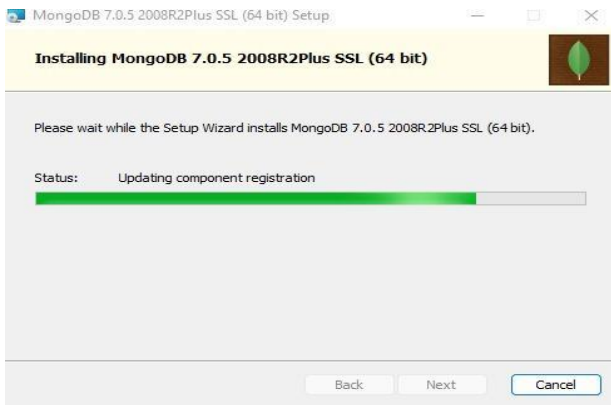
3.ADIM: Complete seçilerek kurum yapıldı.



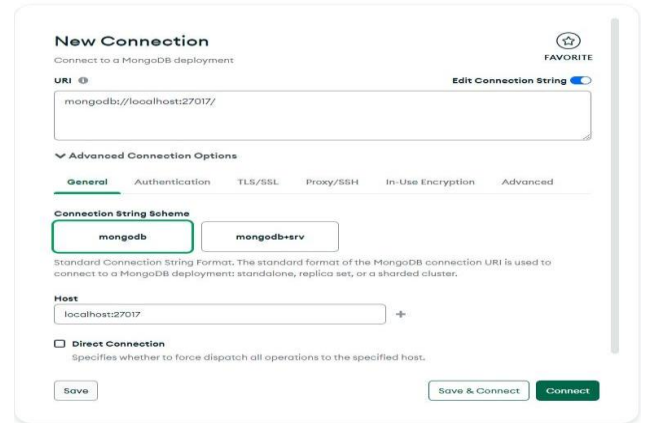
4.ADIM:



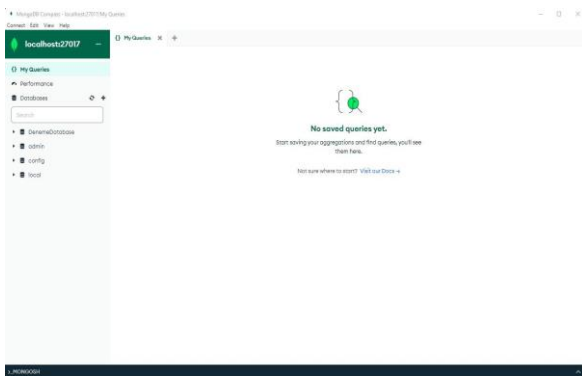
5.ADIM: MongoDB'nin indirme aşaması



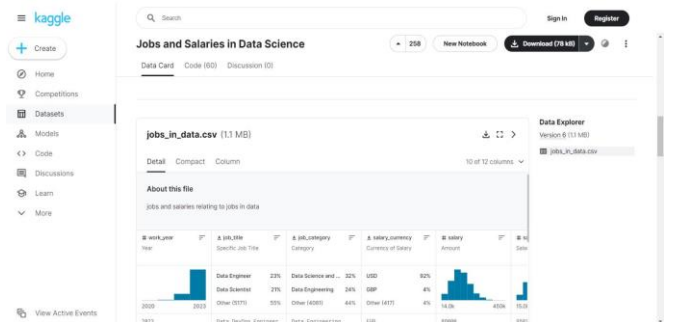
6.ADIM: MongoDB kurulduktan sonraki ekran



7.ADIM: MongoDB seçilerek connect'e devam edilir.



8.ADIM: Kaggle üzerinden jobs_in_data.csv datasetin indiriyoruz.



KODLAMALAR

```
(base) (venv) PS C:\Users\nermin.uzay\PycharmProjects\pythonProject1> pip install pymongo
Collecting pymongo
  Downloading pymongo-4.6.1-cp310-cp310-win_amd64.whl.metadata (22 kB)
  Downloading dnspython-2.5.0-py3-none-any.whl.metadata (5.8 kB)
  Downloading pymongo-4.6.1-cp310-cp310-win_amd64.whl (472 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 472.7/472.7 kB 2.1 MB/s eta 0:00:00
```

- Pip paket yöneticisi aracılığıyla pymongo kütüphanesini indiriyoruz ve projemize ekliyoruz.
-

```
1 import pymongo
2 import pandas as pd
3 import json
```

- Projemizde kullanılacak olan gerekli kütüphaneleri, Python dilinde bulunan import anahtar kelimesi ile içe aktarıyoruz.
-

```
# Connection to MongoDB
client = pymongo.MongoClient('mongodb://localhost:27017/')
db = client['bigdata'] # Database name
collection = db['bigdata'] # Collection name
```

- Projemizdeki Python kodu içerisinde, MongoDB veritabanına bağlantı sağlıyoruz.
-

```
# Read CSV file
csv_file_path = 'c:/Users/Ahmet/Desktop/BuyukVeri/dataset.csv'
df = pd.read_csv(csv_file_path)
```

- .csv formatındaki veri setini projemizdeki Python kodu aracılığıyla okuyoruz ve dataframe olarak df değişkenine kaydediyoruz.


```
# Convert dataframe to dictionary
records = df.to_dict(orient='records')
```

→ df değişkeninde tuttuğumuz dataframe'i dictionary formatına çeviriyoruz.

```
# Insert collection
collection.insert_many(records)
```

→ insert_many() fonksiyonu ile koleksiyona bir veya daha fazla belge eklenir.

In 11 1 df.head()

Executed in 46ms, 5 Feb at 20:49:23

Out 11 5 rows × 12 columns pd.DataFrame

	work_year	job_title	job_category	salary_currency	salary	salary_in_usd	employee_residence	experience
0	2023	Data DevOps Engineer	Data Engineering	EUR	88000	95012	Germany	Mid-Le
1	2023	Data Architect	Data Architecture and Modeling	USD	186000	186000	United States	Senior
2	2023	Data Architect	Data Architecture and Modeling	USD	81800	81800	United States	Senior
3	2023	Data Scientist	Data Science and Research	USD	212000	212000	United States	Senior
4	2023	Data Scientist	Data Science and Research	USD	93300	93300	United States	Senior

→ df.head() komutuyla datanın ilk 5 satırını görüntülüyoruz.

In 18 1 collection.insert_many(records)

Executed in 107ms, 5 Feb at 20:49:55

Out 18 InsertManyResult([ObjectId('65c11fc3371cf15df4ed9d5e'), ObjectId('65c11fc3371cf15df4ed9d5f'), ObjectId('65c11fc3371cf15df4ed9d60'), ObjectId('65c11fc3371cf15df4ed9d61'), ObjectId('65c11fc3371cf15df4ed9d62'), ObjectId('65c11fc3371cf15df4ed9d63'), ObjectId('65c11fc3371cf15df4ed9d64'), ObjectId('65c11fc3371cf15df4ed9d65'), ObjectId('65c11fc3371cf15df4ed9d66'), ObjectId('65c11fc3371cf15df4ed9d67'), ObjectId('65c11fc3371cf15df4ed9d68'), ObjectId('65c11fc3371cf15df4ed9d69'), ObjectId('65c11fc3371cf15df4ed9d6a'), ObjectId('65c11fc3371cf15df4ed9d6b'), ObjectId('65c11fc3371cf15df4ed9d6c'), ObjectId('65c11fc3371cf15df4ed9d6d'), ObjectId('65c11fc3371cf15df4ed9d6e'), ObjectId('65c11fc3371cf15df4ed9d6f'), ObjectId('65c11fc3371cf15df4ed9d70'), ObjectId('65c11fc3371cf15df4ed9d71'), ObjectId('65c11fc3371cf15df4ed9d72'), ObjectId('65c11fc3371cf15df4ed9d73'), ObjectId('65c11fc3371cf15df4ed9d74'), ObjectId('65c11fc3371cf15df4ed9d75'), ObjectId('65c11fc3371cf15df4ed9d76'), ObjectId('65c11fc3371cf15df4ed9d77'), ObjectId('65c11fc3371cf15df4ed9d78'), ObjectId('65c11fc3371cf15df4ed9d79'), ObjectId('65c11fc3371cf15df4ed9d7a'), ObjectId('65c11fc3371cf15df4ed9d7b'), ObjectId('65c11fc3371cf15df4ed9d7c'), ObjectId('65c11fc3371cf15df4ed9d7d'), ObjectId('65c11fc3371cf15df4ed9d7e'), ObjectId('65c11fc3371cf15df4ed9d7f'), ObjectId('65c11fc3371cf15df4ed9d80'), ObjectId('65c11fc3371cf15df4ed9d81'), ObjectId('65c11fc3371cf15df4ed9d82'), ObjectId('65c11fc3371cf15df4ed9d83'), ObjectId('65c11fc3371cf15df4ed9d84'), ObjectId('65c11fc3371cf15df4ed9d85'), ObjectId('65c11fc3371cf15df4ed9d86'), ObjectId('65c11fc3371cf15df4ed9d87')])

→ Son olarak collection.insert_many(records) komutuyla collection oluşturuyoruz.

→ Tüm bu işlemlerden sonra entegre ettiğimiz veritabanını MongoDB üzerinden görüntülüyoruz ve burada gerek duyulan işlemleri sağlayabiliriz.

The screenshot shows the MongoDB Compass interface for a local instance at localhost:27017. The database 'bigdata' is selected, and the 'Documents' tab is active. Three documents are displayed in the collection:

```

{
  "_id": ObjectId("65c11e59a3357bae74dd4e75"),
  "work_year": 2023,
  "job_title": "Data Architect",
  "job_category": "Data Architecture and Modeling",
  "salary_currency": "USD",
  "salary": 81800,
  "salary_in_usd": 81800,
  "employee_residence": "United States",
  "experience_level": "Senior",
  "employment_type": "Full-time",
  "work_setting": "In-person",
  "company_location": "United States",
  "company_size": "M"
}

{
  "_id": ObjectId("65c11e59a3357bae74dd4e76"),
  "work_year": 2023,
  "job_title": "Data Scientist",
  "job_category": "Data Science and Research",
  "salary_currency": "USD",
  "salary": 212000,
  "salary_in_usd": 212000,
  "employee_residence": "United States",
  "experience_level": "Senior",
  "employment_type": "Full-time",
  "work_setting": "In-person",
  "company_location": "United States",
  "company_size": "M"
}

{
  "_id": ObjectId("65c11e59a3357bae74dd4e77"),
  "work_year": 2023,
  "job_title": "Data Scientist",
  "job_category": "Data Science and Research",
  "salary_currency": "USD",
  "salary": 93300,
  "salary_in_usd": 93300
}

```

The interface includes a sidebar with 'My Queries', 'Performance', 'Databases', and a list of databases (admin, bigdata, config, local). The 'bigdata' database is expanded, showing the 'bigdata' collection. The main area has tabs for 'Documents', 'Aggregations', 'Schema', 'Indexes', and 'Validation'. A search bar and a 'Filter' button are present. Below the search bar are buttons for 'ADD DATA', 'EXPORT DATA', 'UPDATE', and 'DELETE'. The bottom status bar shows 'MONGOSH' and system information like '12°C' and 'Çok bulutlu'.