|  |  |  |
| --- | --- | --- |
| **Unigrb** | UNIVERZITET U NIŠU  ELEKTRONSKI  FAKULTET | **logo_1960_4** |

SISTEMI ZA UPRAVLJANJE

BAZAMA PODATAKA

SEMINARSKI RAD

**MONGODB ATLAS KAO DATABASE-AS-A-SERVICE NA GOOGLE CLOUD-U**

|  |  |
| --- | --- |
| Mentor:  Aleksandar Stanimirović | Student:  Dragana Korunović  1220 |

Niš, jun 2021. godine

**Sadržaj**

[1. Uvod 1](#_Toc74426486)

[2. Database as a Service (DBaaS) 2](#_Toc74426487)

[2.1. Prednosti DbaaS 2](#_Toc74426488)

[2.1.1. Nema dodatnog lokalnog hardvera 2](#_Toc74426489)

[2.1.2. Brz razvoj baze podataka 3](#_Toc74426490)

[2.1.3. Operacije sa podacima spremne za upotrebu 3](#_Toc74426491)

[2.2. Izbor database-as-a-service 3](#_Toc74426492)

[2.3. MongoDB Atlas 3](#_Toc74426493)

[3. Google Cloud platforma 4](#_Toc74426494)

[3.1. Prednosti 5](#_Toc74426495)

[3.2. Primena 5](#_Toc74426496)

[4. MongoDB Atlas na Google Cloud-u 6](#_Toc74426497)

[4.1. Prednosti 6](#_Toc74426498)

[4.2. Primena 6](#_Toc74426499)

[5. Rad u MongoDB Atlas-u 7](#_Toc74426500)

[5.1. Upravljanje dokumentima 9](#_Toc74426501)

[5.2. Upravljanje indeksima 10](#_Toc74426502)

[5.2.1. Dinamičko mapiranje 10](#_Toc74426503)

[5.2.2. Statičko mapiranje 10](#_Toc74426504)

[5.2.3. Kreiranje indeksa 10](#_Toc74426505)

[5.2.4. Pretraga dokumenata 14](#_Toc74426506)

[5.3. Pristup 18](#_Toc74426507)

[5.3.1. MongoDB CLI 18](#_Toc74426508)

[5.3.2. MongoDB Compass 20](#_Toc74426509)

[6. Zaključak 24](#_Toc74426510)

[7. Literatura 25](#_Toc74426511)

# Uvod

Sa porastom količine podataka generisanih od strane rešenja koja koriste kako pojedinci, tako i kompanije, razvila se potreba za sistemima koji će moći da se izbore sa količinom i brzinom generisanih podataka. Pošto skladište podataka potrebno za neko rešenje može biti ogromno ili nepredvidivog rasta, kao i kapacitet obrade, rešenje ovog problema je ponuđeno preko database-as-a-service servisa u oblaku.

Database-as-a-service (DBaaS) pruža resurse kao što su infrastruktura, hardver, operativni sistem i softver nekog sistema za upravljanje bazama podataka, pri čemu se oni nalaze na udaljenim mašinama (“u oblaku”). Tako je omogućeno skladištenje i pristup ogromnim količinama podataka, nad kojima je lako vršiti različite operacije. Jedan od pružalaca servisa u oblaku je Google Cloud platforma.

U database-as-a-service baze podataka spada i MongoDB Atlas. On, kao nestruktuirana baza podataka, pruža bolju skalabilnost u odnosu na struktuirane, dok replikacijom osigurava otpornost na otkaze. Preko MongoDB Atlas-a omogućen je lak pristup i ažuriranje udaljenih podataka.

Google Cloud platforma pruža alate za upravljanje i servise kao što su izračunavanje, skladištenje i analiza podataka i mašinsko učenje, dok MongoDB Atlas omogućava distribuirane transakcije dokumenata, uz dobru automatizaciju. Pri tome se za vizuelizaciju podataka mogu koristiti MongoDB Charts. Pri izboru Google Cloud platforme kao pružaoca usluga MongoDB Atlas-u, povećava se moć izračunavanja, elastičnost resursa i skalabilnost, a mogu se iskoristiti i njene usluge vezane za mašinsko učenje i veštačku inteligenciju.

MongoDB Atlas-u može se pristupiti preko MongoDB Shell-a i MongoDB Compass-a. Prvi podrazumeva CLI, a drugi grafički korisnički interfejs. MongoDB Shell ili mongocli omgućava izvršenje osnovnih komandi upravljanja ovom bazom podataka. MongoDB Compass predstavlja najbolji način da se podaci vizuelno istraže i pokrenu ad hoc upiti u nekoliko sekundi. Interakcija sa podacima podrazumeva potpunu CRUD funkcionalnost i može se vršiti preko ugrađenog vizuelnog alata za ažuriranje.

Sledeće poglavlje posvećeno je DbaaS, njegovim karakteristikama i prednostima, dok se treće bavi Google Cloud platformom i razlozima njenog korišćenja. U četvrtom poglavlju biće reči o MongoDB Atlas-u na Google Cloud platformi, a u petom o radu sa MongoDB Atlas-om, koji se nalazi na pomenutoj platformi. Biće reči o upravljanju dokumentima i indeksima i pristupu Atlas-u preko korisničkih interfejsa. Poslednje poglavlje posvećeno je zaključku rada.

# Database as a Service (DBaaS)

Database-as-a-service je servis računarstva u oblaku. Ovaj servis olakšava svoje korišćenje time što se nalazi na udaljenoj mašini, pa nije potrebno nabavljanje hardvera i instaliranje softvera. Takođe, pružalac ovog servisa preuzima akcije održavanja baze podataka.

Jezgro DBaaS baze podataka je instalirano na udaljenom centru podataka, pa se ne primećuje velika razlika u rukovanju u odnosu na lokalno instaliranu bazu. Glavna razlika je fizička infrastruktura baze podataka u oblaku. U slučaju da se koristi javno Infrastructure as a Service (IaaS)[[1]](#footnote-1) okruženje u oblaku (Microsoft Azure, Amazon Web Services (AWS) ili Google Cloud Platform), baza podataka se pokreće na zajedničkoj hardverskoj platformi. Ovaj pristup povećava moć izračunavanja, elastičnost resursa i skalabilnost za povećanje skladišta podataka i obradu zahteva.

Razlika između DBaaS i upravljanja bazama podataka je jednostavnost korišćenja koju prve pružaju i ušteda vremena i resursa, pošto pružalac servisa upravlja infrastrukturom i obezbeđuje da fizički i aplikacioni slojevi dobro funkcionišu i budu optimizovani. Za razliku od lokalnog centra podataka, računarstvo u oblaku pruža praktično neograničen prostor za skladištenje podataka.

Korišćenje DbaaS polazi od manjih sajtova, gde se one koriste za backend operacije i deo su njegovog web hosting-a, sve do velikih organizacija. Ove organizacije koriste SQL i NoSQL baze podataka za skladištenje ogromnih količina podataka za analizu u realnom vremenu preko kompleksnih real-time aplikacija. Generalno, DbaaS su pogodne za aplikacije kojima je potrebna skalabilnost i fleksibilnost baze podataka. U nastavku su navedene neke od glavnih prednosti korišćenja DbaaS.

## Prednosti DbaaS

### Nema dodatnog lokalnog hardvera

Razvoj modernih aplikacija zahteva stalno ulaganje u dodatnu memoriju i kapacitet obrade. DbaaS preko modela računarstva u oblaku pruža uslove za neograničen razvoj, bez početnih ulaganja i brige o količini memorije i dodatnim hardverskim resursima. Servisima koje DbaaS pruža (infrastruktura, hardver, operativni sistem, softver) upravlja njihov pružalac, kao što je MongoDB, i tako oslobađa programera brige o njima.

### Brz razvoj baze podataka

Model database-as-a-service i poboljšanje operacija sa podacima je pojednostavljeno. Takođe, programeri mogu da se koncentrišu na rad sa podacima, bez razmišljanja o mogućem nedostatku memorijskog prostora. Omogućeno je brzo kreiranje baza podataka i lako kloniranje skupova podataka i konfiguracija. Korišćenje API-eva u sklopu servisa u oblaku omogućava izgradnju modernih i naprednih aplikacija koje pomažu dostizanju poslovnih i strateških ciljeva.

### Operacije sa podacima spremne za upotrebu

Database-as-a-service NoSQL baza podataka kao što je MongoDB Atlas je pogodna za rad sa big data. Moguće je kreirati operacioni sloj podataka (Operation Data Layer) nad ovom bazom i tako pružiti podatke kao servis (data as a service). Ovaj sloj može da pomogne kompaniji da omogući pristup svojim podacima, koje će da koriste aplikacije koje se kreiraju da budu inteligentne, brze i rade u realnom vremenu.

## Izbor database-as-a-service

Izbor DbaaS zavisi od potreba njenog korišćenja. Često se vrši migracija postojeće lokalne baze podataka na njenu verziju u oblaku. Ako to nije slučaj, dobro je odlučiti se za nestruktuiranu NoSQL bazu podataka, zbog fleksibilnosti koju poseduje. Takođe, nestruktuirane baze podataka pružaju bolju skalabilnost u odnosu na struktuirane, nauštrb garancije transakcija. MongoDB je ovakva baza, koja može da čuva izvorne i neizmenjene podatke. Odluka o korišćenju se treba doneti u timu ljudi sa iskustvom koji imaju uvid u osnovne potrebe aplikacije koju nameravaju da kreiraju.

## MongoDB Atlas

MongoDB Atlas je jedan od NoSQL DbaaS servisa sa opcijom hosting-a baze podataka. Prilikom pretplate na ovaj DBaaS se dobija mogućnost rezervisanja baze podataka u oblaku na nekom od IaaS okruženja (Microsoft Azure, Amazon Web Services (AWS) ili Google Cloud Platform), kao i dozvole, podrška i održavanje koji su potrebni za upravljanje njome. Pruženi su i API-ji na oblaku pogodni za kreiranje aplikacija, pristup i izmenu podataka preko koda.

Pri korišćenju MongoDB DBMS-ova moguće je instalirati ih i pokrenuti lokalno u privatnom okruženju u oblaku. Pri tome se ne dobijaju iste prednosti, pa se, ako je potrebno, mora kreirati i održavati sopstvena infrastruktura, pri čemu je poželjno da bude otporna na otkaze i visoko dostupna. Neke od prednosti MongoDB Atlas-a su sledeće:

* *Multi-cloud distribucija podataka* – podaci se particionišu ili distribuiraju na preko 75 regiona u oblaku na pomenutim IaaS servisima.
* *Bezbednost osetljivih podataka* – poseduje kontrolu bezbednosti i naprednu funkcionalnost da zadovolji nove i postojeće mere privatnosti i saglasnosti.
* *Dizajniran za produktivnost programera* – omogućava brži razvoj aplikacija preko drajvera, integracija i alata za rukovanje, vizuelizaciju i analizu podataka.
* *Pogodan za kritična opterećenja* – karakteriše ga visoka dostupnost sa distribuiranom otpornošću na otkaze i rezervnim opcijama koje vrše oporavak podataka.
* *Izgrađen za optimalne performanse* – moguće je skaliranje na zahtev, dostupni su alati za optimizaciju resursa i uvid u realnom vremenu u performanse baze podataka.
* *Upravljiv za efikasnost operacija* – omogućava potpuno automatizovano rezervisanje infrastrukture, podešavanje baze podataka, održavanje i unapređivanje verzija.

# Google Cloud platforma

Google Cloud javna infrastruktura u oblaku podrazumeva Google Cloud platformu, Google Workspace (G Suite), enterprise verzije [Android](https://en.wikipedia.org/wiki/Android_(operating_system)) i [Chrome OS](https://en.wikipedia.org/wiki/Chrome_OS)-a, API-je za mašinsko učenje, kao i servise za mapiranje preduzeća.

Google Cloud platforma (Google Cloud Platform), je skup servisa računarstva u oblaku koji je pokrenut na infrastrukturi koju kompanija Google već koristi za svoje proizvode, kao što su [Google Search](https://en.wikipedia.org/wiki/Google_Search), [Gmail](https://en.wikipedia.org/wiki/Gmail), [skladište](https://en.wikipedia.org/wiki/Google_Drive) dokumenata i [YouTube](https://en.wikipedia.org/wiki/YouTube). Pruža alate za upravljanje i servise kao što su izračunavanje, skladište i analiza podataka i mašinsko učenje. Google Cloud platforma nudi i [infrastructure as a service](https://en.wikipedia.org/wiki/Infrastructure_as_a_service), [platform as a service](https://en.wikipedia.org/wiki/Platform_as_a_service)[[2]](#footnote-2) i [serverless computing](https://en.wikipedia.org/wiki/Serverless_computing)[[3]](#footnote-3) okruženja.

Prvi Google Cloud servis, App Engine, nastao je 2008. godine. To je platforma za razvoj i hosting web aplikacija. Danas postoji niz drugih Google Cloud servisa. U nastavku su date glavne prednosti Google Cloud platforme.

## Prednosti

* *Brži razvoj aplikacija* – Google Cloud se posvećuje open-source, multicloud i hybrid cloud okruženjima, zbog čega se ubrzava razvoj aplikacija.
* *Pametnije donošenje odluka u organizaciji* – svaki član razvojnog tima može da ima uvid u moguće aktivnosti, preko platforme za analitiku bez servera, koja je potpuno upravljiva.
* *Povezivanje i saradnja* – dostupni su integrisani video pozivi, email, čet i aplikacije za saradnju preko dokumenata. Google Workspace je jedno od rešenja koje pomaže produktivnosti na radnom mestu i sadrži aplikacije kao što su Gmail, Docs, Drive i Meet.
* *Modernizacija infrastrukture i aplikacije –* omogućena jemodernizacija infrastrukture u oblaku i pokretanje kritičnih opterećenja sistema, kao što su VMware, SAP, Oracle i Windows. Modernizacija postojeće aplikacije je omogućena preko servisa koji pokreću kontejnere i druge mogućnosti računarstva u oblaku.
* *Upravljanje bazama podataka* – moguća je migracija i upravljanje podacima kompanije uz bezbednost, pouzdanost, visoku dostupnost i potpuno upravljive servise podataka.
* *Veštačka inteligencija i mašinsko učenje* – dostupna su rešenja koja se zasnivaju na veštačkoj inteligenciji i mašinskom učenju, koja se koriste da bi se povećala efikasnost u brizi o korisnicima, zapošljavanju i dr.
* *Bezbednost* – zaštita podataka omogućena je preko naprednih servisa za bezbednost, koji nisu skupi.

Navedene prednosti platforme podstakle su njenu primenu, koja je data u sledećem odeljku.

## Primena

Servisi Google Cloud platforme mogu se iskoristiti za automatizovanje procesa u izradi poslovnih aplikacija, kao i u sistemima za maloprodaju i finansijske usluge. Primena je pronađena i u zdravstvu, gde se može personalizovati iskustvo pacijenta i modernizovati istraživanja. U javnom sektoru se preko ovih servisa mogu poboljšati usluge pružene građanima i uvesti različite inovacije. Takođe, ne treba izostaviti medije i industriju igara, gde se mogu iskoristiti uvidi u podatke korisnika, dostupni u realnom vremenu.

# MongoDB Atlas na Google Cloud-u

MongoDB Atlas u kombinaciji sa Google Cloud platformom pruža potpuno upravljiv servis na Google-ovoj infrastrukturi, koja je skalabilna na globalnom nivou i pouzdana. Upis i čitanja su brzi, iako se klasteri mogu nalaziti bilo gde na svetu. Lako je upravljati bazama podataka preko ovog servisa i izvršiti migraciju na njega, dok je model podataka fleksibilan i omogućava kompleksne upite i indeksiranje.

Servisi sa kojima može da se vrši integracija su Cloud Key Management Service[[4]](#footnote-4), za upravljanje osetljivim podacima i Cloud Provider Snapshots[[5]](#footnote-5), za kratak oporavak i minimizaciju troškova.

## Prednosti

MongoDB Atlas omogućava da se iskoriste najnovija MongoDB svojstva, kao što su distribuirane transakcije dokumenata, pri čemu se vodi računa o operacijama sa bazom podataka i njenom održavanju. MongoDB Atlas se lako može nabaviti na Google Cloud Marketplace-u, uz jednostavan program plaćanja. Pri tome poseduje ugrađenu bezbednost, visoku skalabilnost i dostupnost.

MongoDB se koristi u kombinaciji sa Google Cloud platformom kada su nekom sistemu potrebne prethodno navedene osobine. MongoDB Stitch funkcije omogućavaju pristup bazama podataka i kolekcijama Atlas-a preko front-end aplikacionog koda. Za vizuelizaciju podataka se mogu koristiti MongoDB Charts. Takođe se mogu iskoristiti usluge Google-a vezane za mašinsko učenje i veštačku inteligenciju.

## Primena

Sve je više onih koji se odlučuju da koriste Atlas pokrenut na Google Cloud platformi. Razlozi za to su različiti, kao što je upravljanje katalozima proizvoda popularnih sajtova za online trgovinu i spajanje raznovrsnih korisničkih podataka da bi im se pružilo bolje iskustvo. Popularno je koristiti pomenuti servis pri izradi globalnih mobilnih i web aplikacija. Takođe, upisi i čitanje su brzi, pa se Atlas može koristiti kada je potrebno u kratkom vremenskom periodu izvršiti veliki broj transakcija, na primer 1000 po minutu.

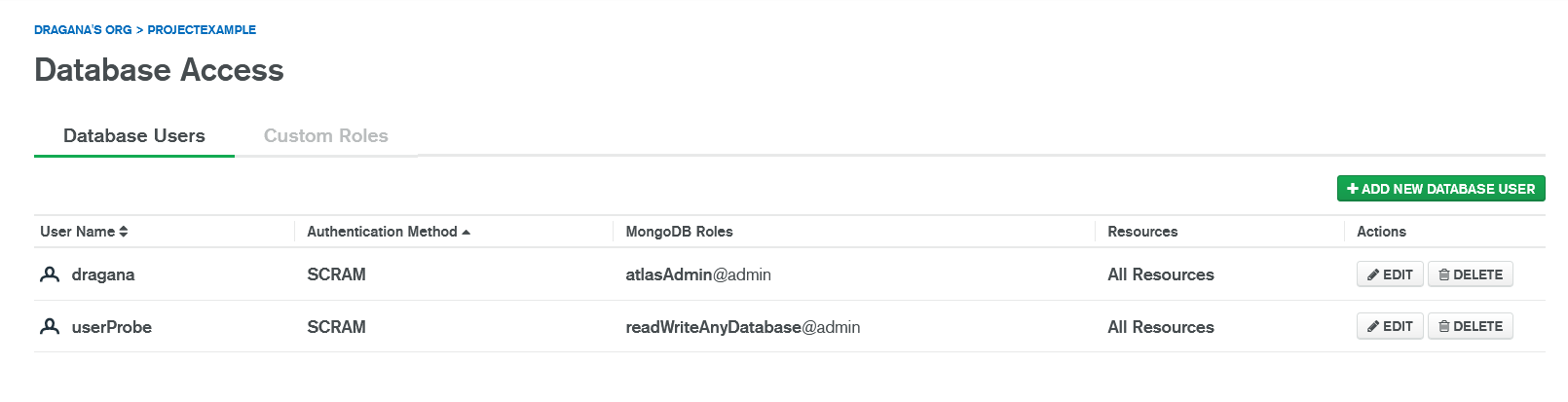
U sledećem poglavlju će na jednostavnim primerima biti prikazan način rada u okruženju MongoDB Atlas. Verzija baze podataka koja je korišćena je najnovija, MongoDB 4.4.6.

# Rad u MongoDB Atlas-u

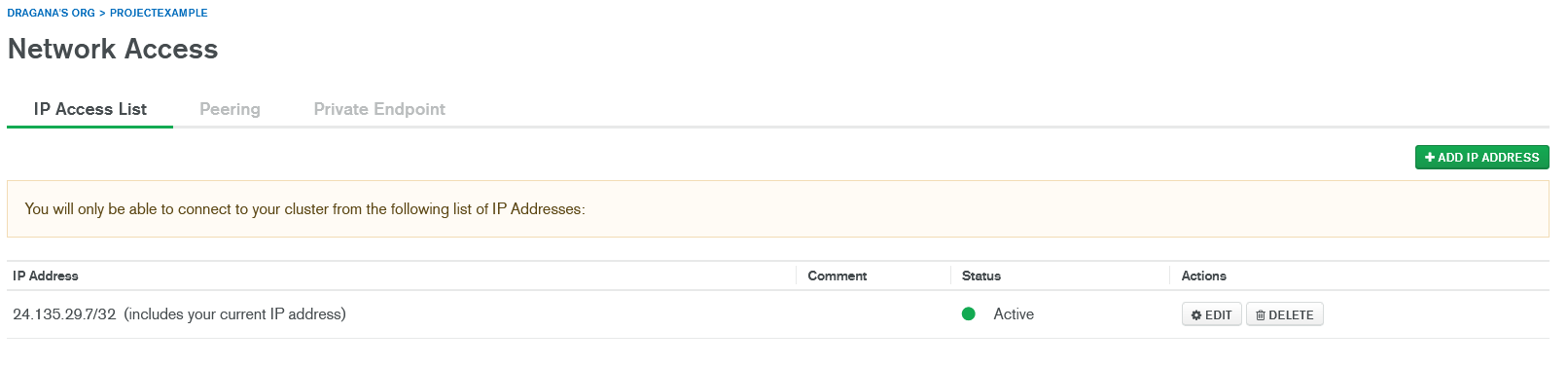
Prvi korak u radu sa MongoDB Atlas-om je učlanjivanje, gde je izabrano da ono bude automatsko, preko Google naloga. Zatim je bilo potrebno kreirati organizaciju i projekat, jednostavnom dodelom imena. Efekti ovih akcija su da korisnik koji ih je izvršio poseduje najviše privilegije u radu sa ovim okruženjem (dobija ulogu *Organization Owner*). Nakon toga je kreiran i klaster, gde je iz ponude pružaoca servisa AWS, Google Cloud i Azure izaran drugi. Kao lokacija izabrana je Belgija, najbliža Srbiji, radi dostizanja maksimalne brzine rada sa bazom. Vrsta klastera koja je izabrana je M0. Ona nije pogodna za produkciju, ali je zato besplatna i poslužiće demonstraciji rada baze. Klaster tipa M0 pruža deljeni RAM sa 512 MB skladišta podataka i ograničeni broj baza podataka i konekcija. Poslednji korak kreiranja klastera je dodela naziva, koji se naknadno ne može promeniti. Kreirani klaster nazvan je *TestCluster*. Na stranici klastera treba izabrati opciju *Load Sample Dataset*, zaubacivanje uzorka skupa podataka, koji će biti korišćen u primerima.

Zatim će biti prikazan glavni prozor pod nazivom *Data Explorer*,nakon čega je potrebno podesiti pristup bazi podataka preko kartice *Database Access*. Klikom na *Add New Database User* se dobija mogućnost unosa korisničkog imena i načina autentifikacije. Prvo korisničko ime je *dragana*, a autentifikacija će se vršiti preko lozinke (izabrana opcija *password*), a zatim od mogućih privilegija korisnika (*Admin[[6]](#footnote-6), Read and write to any database[[7]](#footnote-7), Only read any database[[8]](#footnote-8)*) biće izabrana *Admin*. Isti postupak je ponovljen za korisnika *userProbe*, s tim što su njegove privilegije *Read and write to any database* (Slika 1).

Za rad sa MongoDB Atlas-om neophodno je podesiti i pristup preko mreže kreiranom klasteru. Izabrana je opcija *IP Access List*, što znači da se samo preko navedenih IP adresa može povezati sa *TestCluster*-om. Za dati primer izabrana je opcija izbora adresa u opsegu koji uključuje trenutnu javnu adresu korišćene mašine na mreži. Rezultat dodavanja pomenutog opsega prikazan je u kartici *IP Access List*, sekcije *Network Access* (Slika 2).



Slika – podaci o korisnicima koji imaju pristup projektu



Slika – lista IP adresa sa kojih se može pristupiti projektu

Takođe, treba odrediti prava pristupa organizaciji, klikom na karticu *Organization Access*, a zatim izabrati opciju *Create API Key*. Najpre treba odrediti naziv ključa i dozvole koje korisnik ima prilikom njegovog korišćenja za pristup organizaciji (Slika 3). Na slici 4 prikazani su generisani privatni i javni API ključ, kao i opcija dodavanja IP adresa preko kojih se uz pomoć API ključeva može pristupiti organizaciji (*Add Access List Entry*).

|  |  |
| --- | --- |
|  |  |
| Slika – kreiranje API ključa, prvi deo | Slika – kreiranje API ključa, drugi deo |

## Upravljanje dokumentima

Dokumentima[[9]](#footnote-9) u okviru kolekcija klastera se upravlja preko sekcije *Data Explorer*, koja je deo grafičkog korisničkog interfejsa MongoDB Atlas-a. Njihov pregled daje uvid u šemu baze podataka. Da bi se pregledali dokumenti kolekcije, potrebno je kliknuti na željenu bazu podataka u okviru klastera, a zatim na željenu kolekciju. Nakon klika na ime kolekcije, može se izvršiti pretraga uz opcioni filter pretrage. Rezultati se prikazuju tako da ih na svakoj stranici ima najviše 20.

Pri pretrazi se mogu definisati uslovi filtriranja, gde je uslov jednakosti dat u obliku:

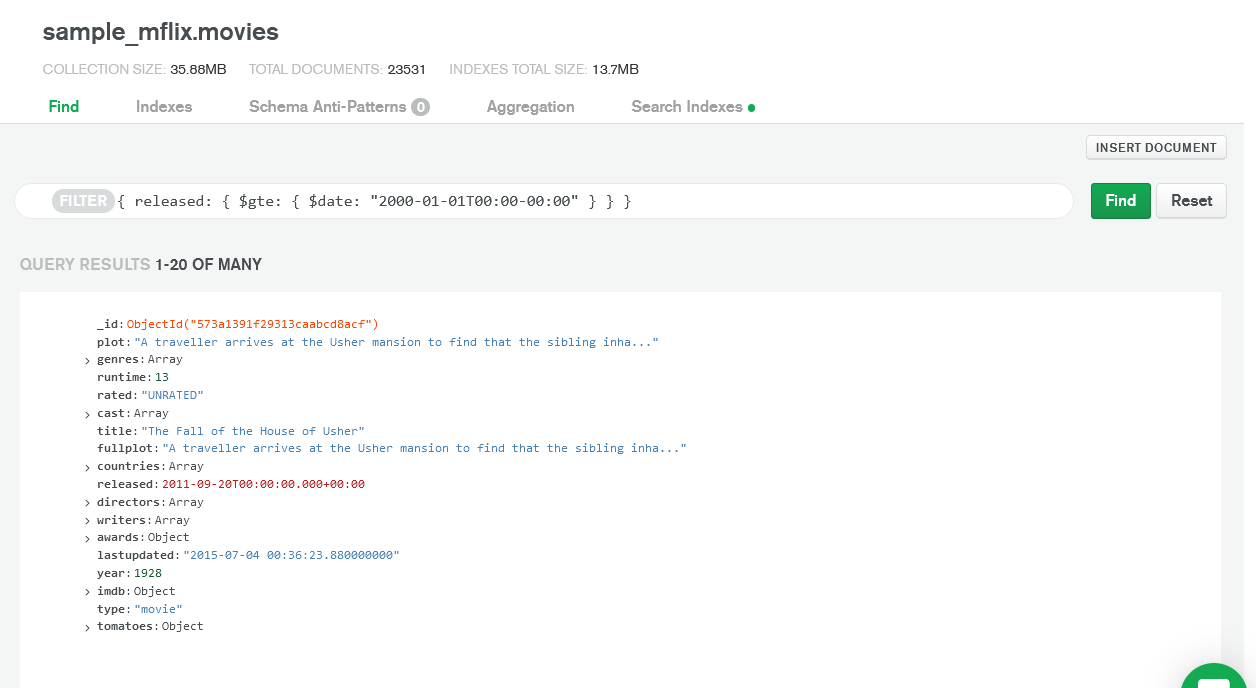
*{ <polje1>: <vrednost1>, ... }*,

a uslov koji koristi operatore za vrednosti polja piše se u obliku:

*{<polje1>: {<operator\_upita>: <vrednost1>}, ...}.*

Upit sa slike 5 koristi operator poređenja nad poljem *date* kolekcije *movies* i vraća sve filmove koji su izašli nakon ponoći 1. januara 2000. godine, po koordinisanom univerzalnom vremenu. Takođe, moguće je dodati potpuno nov dokument ili klonirati postojeći i izmeniti ga po potrebi. U okviru *Find* kartice za pretragu dokumenata nalazi se dugme *Insert Document*. Na slici 6 se vidi prozor u kome se vrši ažuriranje dokumenta, čijem je polju *\_id* automatski dodeljena jedinstvena vrednost. Više dokumenata se može dodati izborom opcije *JSON View*, a zatim njihovim unosom u JSON formatu.

Opcije sa slike 7 se pojavljuju pored svakog dokumenta iz liste dokumenata u kolekciji i redom označavaju njegovo ažuriranje, kopiranje saržaja, kloniranje i brisanje. Razlika u kloniranju i ažuriranju je to što se dokument koji se klonira može i ne mora dodatno ažurirati.



Slika – upit nad kolekcijom movies

|  |  |
| --- | --- |
|  | Clone document icon |
| Slika – kreiranje dokumenta za dodavanje u kolekciju | Slika – opcije za svaki iz liste dokumenata |

## Upravljanje indeksima

Indeksi za pretragu služe za olakšavanje pretrage dokumenata po vrednostima određenih polja. Atlas Search indeks može da koristi statičko i dinamičko mapiranje.

### Dinamičko mapiranje

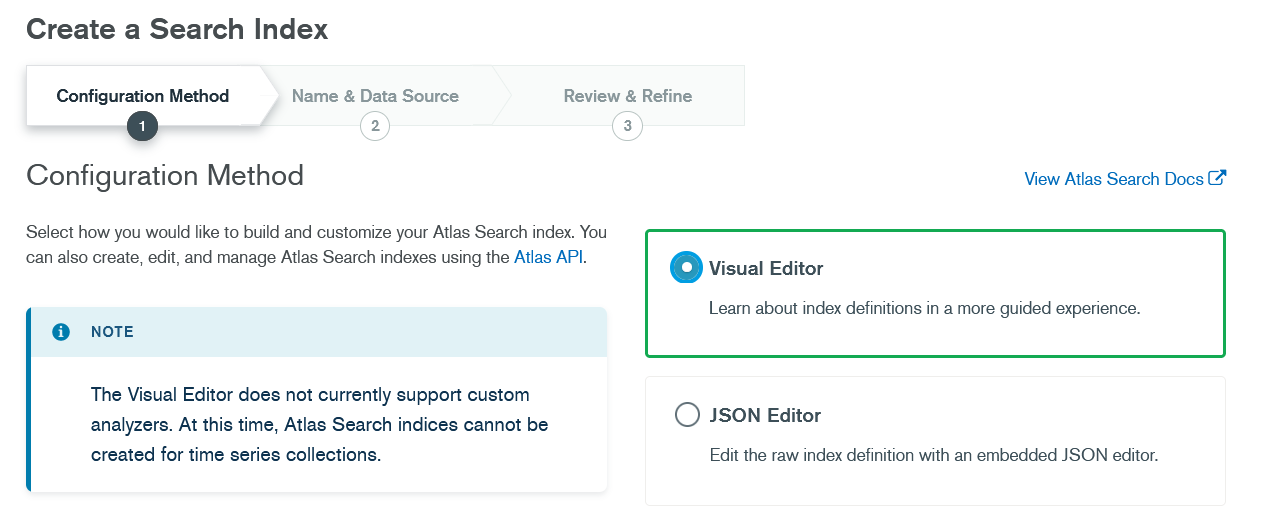
Pod dinamičkim mapiranjem se podrazumeva indeksiranje svih polja dokumenata koja su podržanog tipa i nisu eksplicitno isključena. Moguće je indeksirati i samo određena polja na ovaj način. Dinamičko mapiranje se koristi kada se šema baze podataka često menja ili je nepoznata. Mana dinamičkih indeksa je zauzeće većeg prostora na disku i slabije performanse. Ova mana nije toliko izražena kada se koristi servis računarstva u oblaku.

### Statičko mapiranje

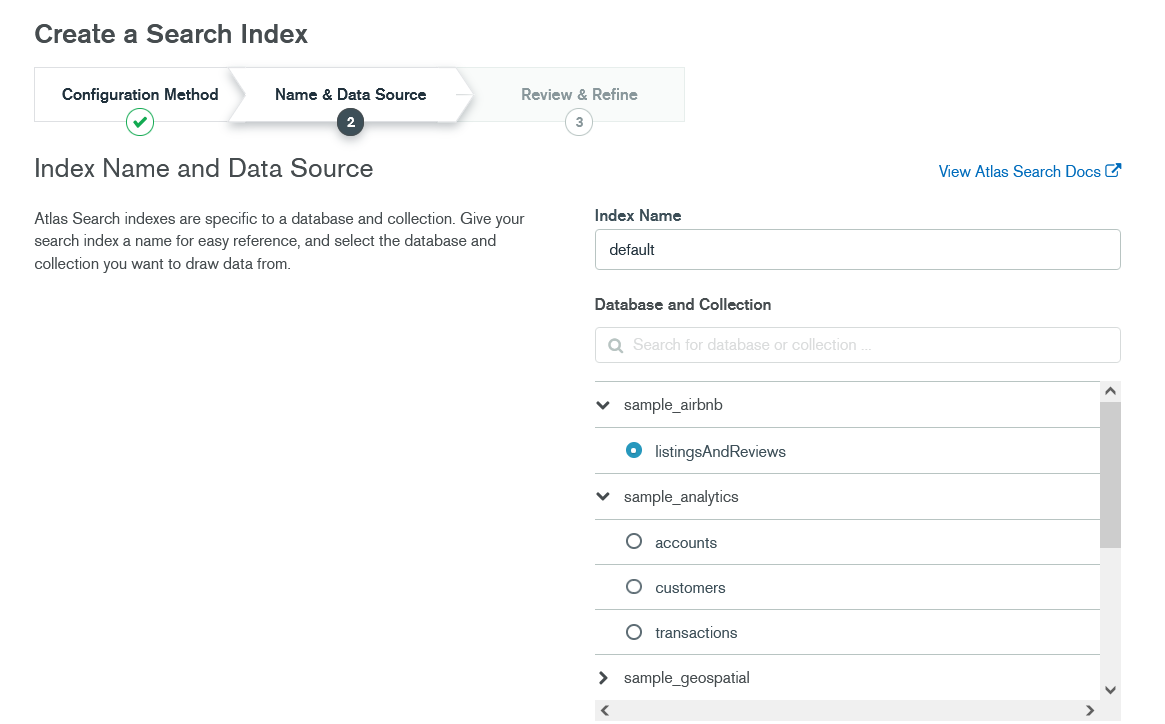
Statičko mapiranje se koristi zbog uštede prostora, za indeksiranje samo navedenih polja dokumenata. Dati primer vezan je za ovu vrstu indeksiranja.

### Kreiranje indeksa

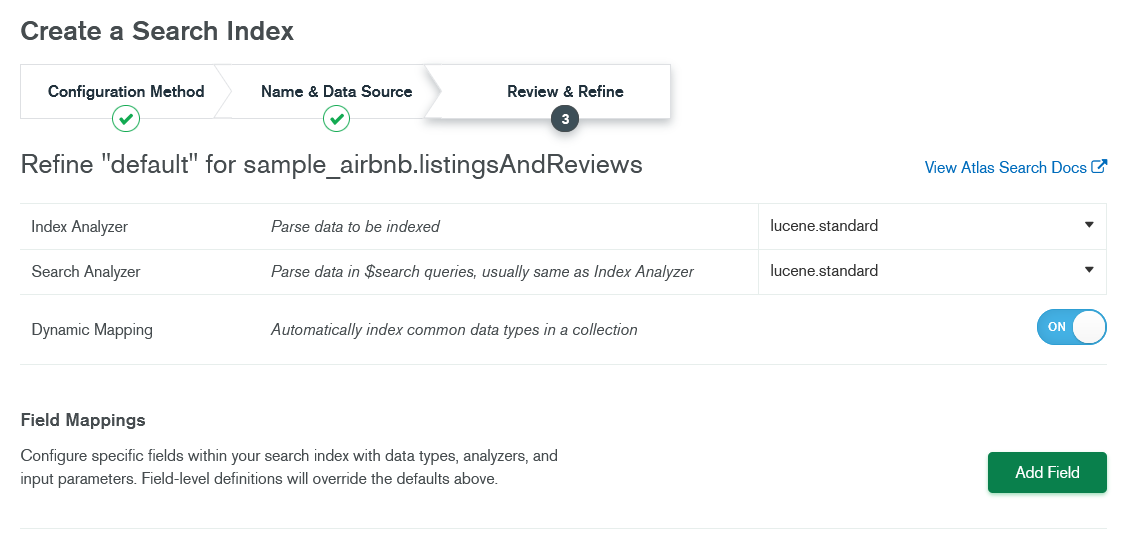
Na stranici klastera *Data Explorer*-a najpre treba izabrati željeni klaster, a zatim preći na karticu *Search*. Nakon toga treba izabrati opciju *Create Search Index*. Indeks za pretragu se može definisati preko vizuelnog editora ili pomoću JSON koda. U nastavku će biti opisan prvi slučaj. Najpre treba uneti naziv indeksa (u ovom primeru to je *default*), koji mora biti jedinstven za dati namespace. Uz to treba izabrati naziv baze podataka i kolekcije. Dati primer vezan je za kolekciju *listingsAndReviews* u bazi podataka *sample\_airbnb*,iz ubačenog uzorka skupa podataka*.* Uključena je opcija *dynamic mapping,* a polja *description* i *notes* su izabrana za statičko mapiranje i njihov tip je *String*. Sam indeks i polja koja ga čine koriste standardni analizator[[10]](#footnote-10) (slike 8 – 11).



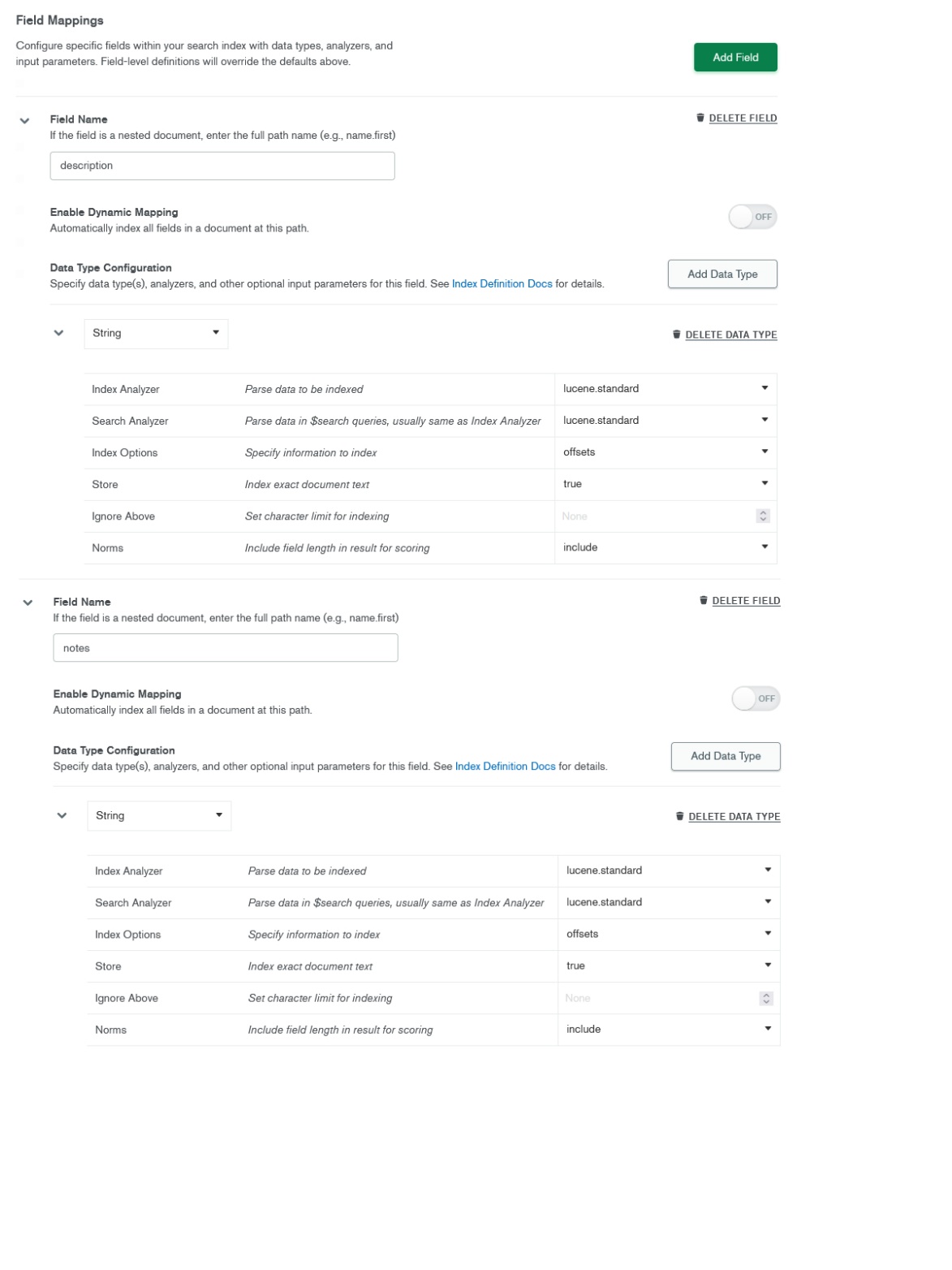
Slika – kreiranje indeksa za pretragu, prvi korak



Slika – kreiranje indeksa za pretragu, drugi korak

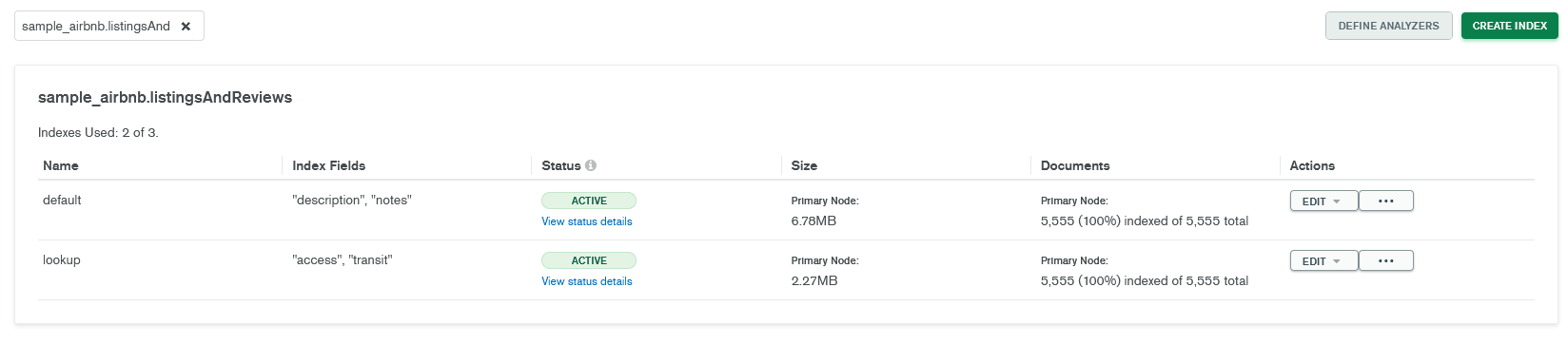


Slika – kreiranje indeksa za pretragu, treći korak

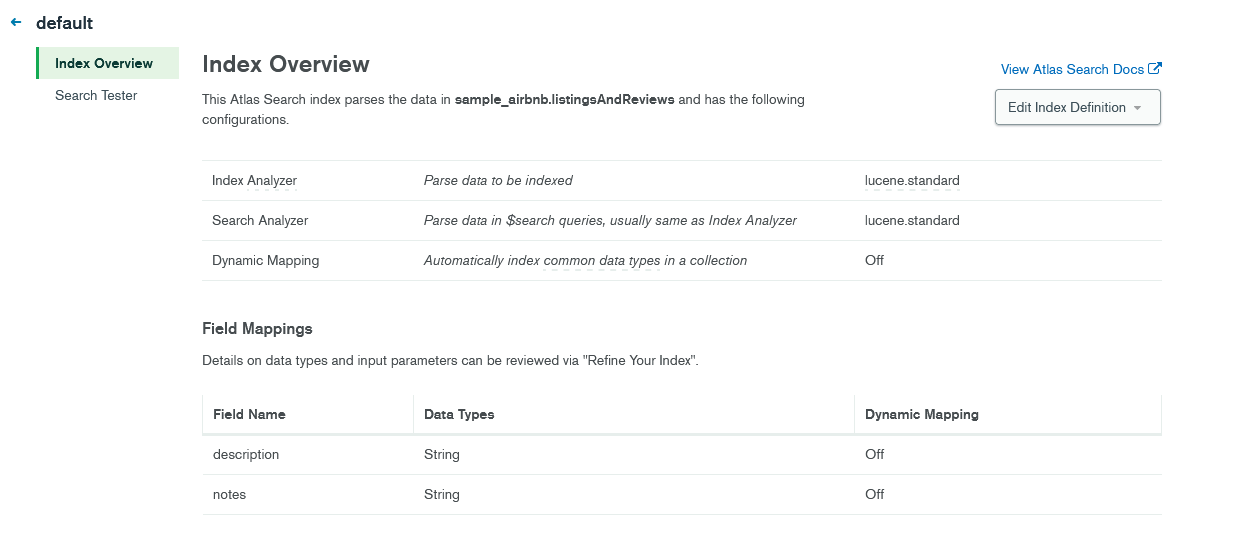


Slika – kreiranje indeksa za pretragu, treći korak

Na slici 12 prikazana je lista indeksa u kolekciji *listingsAndReviews*. Pošto nema smisla dinamički mapirati indeks koji se sastoji od dva statička polja, tip indeksa *default* je promenjen na *static*, preko njegove opcije *edit.* Nakon promene mogu se videti detalji o datom indeksu (Slika 13).



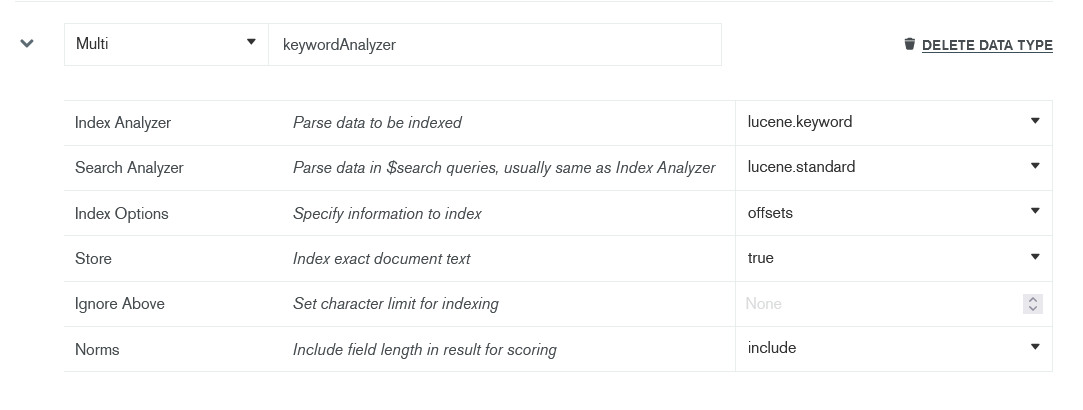
Slika – indeksi kolekcije listingsAndReviews



Slika – podaci o indeksu

Postupak kreiranja indeksa (nazvanog *defaultIndex*) je ponovljen i za kolekciju *movies* u bazi podataka *sample\_mflix*,iz istog uzorka skupa podataka*.* Pri tome je potrebno isključiti opciju *dynamic mapping* i dodati polja *genres, plot* i *title*, tipa string (po kojima se vrši indeksiranje). Takođe, treba dodati i tip *multi* za poslednje polje, koje koristi multi analizator, kome dodeljujemo ime *keywordAnalyzer* (Slika 14). Analizator indeksa se treba promeniti na *lucene.keyword*.

Kreirani indeks specificira standardni analizator kao podrazumevani za data polja. Za polje *title*, keyword analizator je alternativni i služi za indeksiranje celog polja u vidu jedinstvenog termina, što znači da se termin iz upita mora potpuno poklapati sa pomenutim terminom u indeksu. Polju *genres*, koje predstavlja niz stringova, dodeljen je standardni analizator uz tip elemenata niza, string. Pri tome nije potrebno posebno naglašavati da je polje tipa *array* (niz).

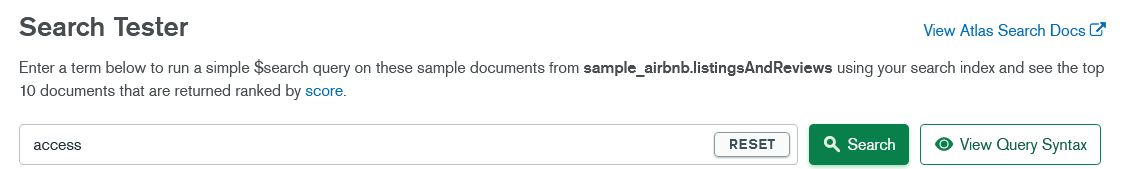


Slika – podešavanje tipa polja i analizatora

### Pretraga dokumenata

Pretraga svih MongoDB klastera moguća je od verzije 4.2. Atlas Search pruža više analizatora teksta, rangiranje rezultata po score-u i bogat upitni jezik. Nakon kreiranja indeksa, treba testirati pretragu dokumenata.

Da bi se testirala pretraga dokumenata kolekcije, može se izvršiti pretraga preko njene *Search* kartice (Slika 15). Unosom željenog termina može se generisati sintaksa upita, klikom na *View Query Syntax*, kako u Mongo Query Language-u (Slika 16), tako i u izabranom programskom jeziku (Slika 17). Mongo Query Language je jezik koji se koristi za pisanje MongoDB upita, baziranih na JavaScript-u i JSON strukturi dokumenata. Deo generisanih rezultata prikazan je na slici 18 i predstavlja sve dokumente koji sadrže termin *access* u bilo kom od svojih polja.



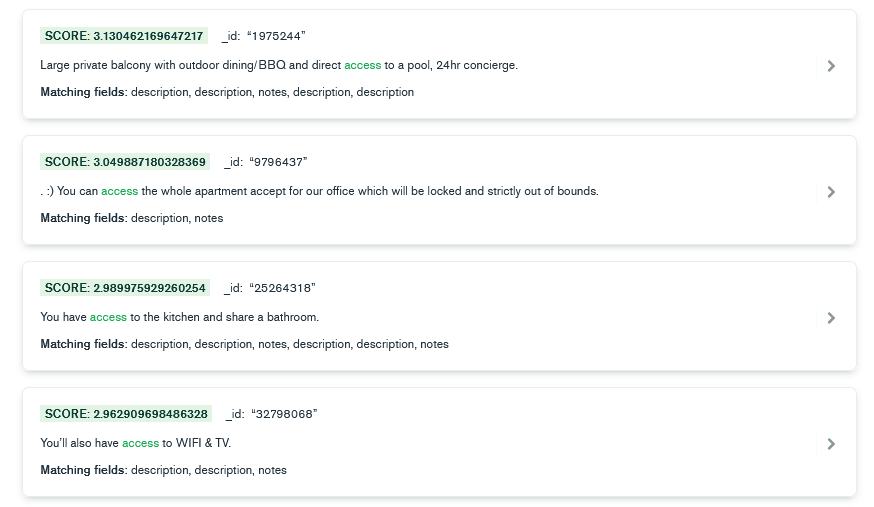
Slika – mogućnost pretrage dokumenata

|  |  |
| --- | --- |
|  |  |
| Slika – sintaksa upita u MongoDB API-u | Slika – sintaksa upita u izabranom programskom jeziku |

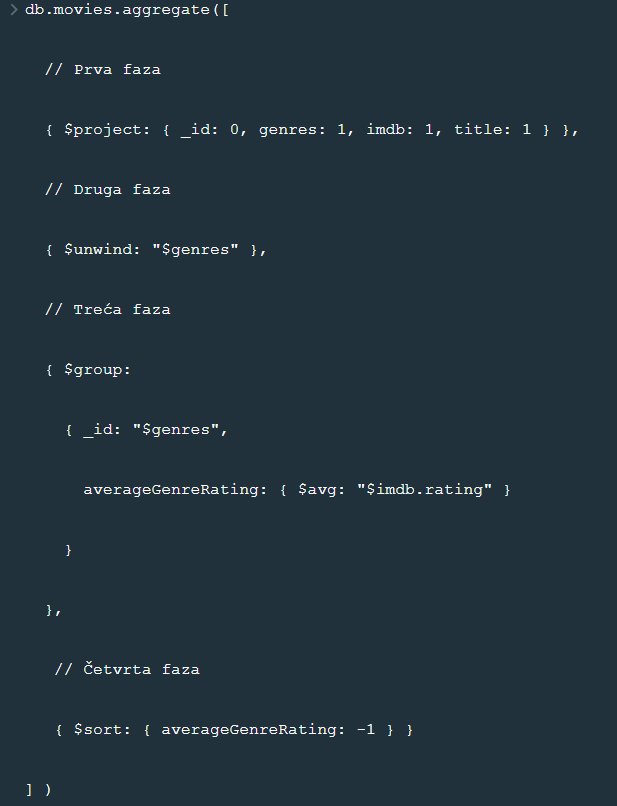
* **Izvršavanje pipeline-ova agregacije**

Pipeline-ovi agregacije (aggregation pipelines) služe za pretragu dokumenata, tj. oni na osnovu izabranih pipeline faza vraćaju transformisane dokumente iz određene kolekcije i izvršavaju se, kako u MongoDB Atlasu, tako i u Compass-u i Shell-u (predstavljenim u odeljku 5.3). Agregacija se koristi za grupisanje podataka po određenom izrazu, računanje rezultata na osnovu više polja i njihovom čuvanju u novom, kao i filtriranje i sortiranje podataka.

MongoDB pipeline agregacija se sastoji od faza, gde svaka menja dokumente, pri čemu ih može i generisati ili filtrirati. Na primeru sa slike 19 se mogu videti neke od brojnih faza agregacije koje se mogu primeniti. Primer koristi uzorak podataka *sample dataset*, bazu podataka *sample\_mflix* i kolekciju *movies*. U nastavku je opisan način izvršenja izabranih faza.



Slika – deo rezultata pretrage upita sa slike 16



Slika 19 – pipeline agregacije nad kolekcijom movies u mongocli-u

|  |
| --- |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |

* ***Prva faza***

Faza [*$project*](https://docs.mongodb.com/manual/reference/operator/aggregation/project/#mongodb-pipeline-pipe.-project) prosleđuje u narednu fazu samo one dokumente koji sadrže polja *genres*, *imdb* i *title*. Definisanje *\_id: 0* služi da naglasi da se ovo polje zanemaruje pri prelasku dokumenata u narednu fazu. Izgled dokumenta koji se prosleđuje u narednu fazu je sledeći:

|  |
| --- |
|  |
| { | |
|  | |

|  |
| --- |
| imdb: { rating: 8.2, votes: 419589, id: 83658 }, |
|  |

|  |
| --- |
| genres: [ 'Sci-Fi', 'Thriller' ], |
|  |

|  |
| --- |
| title: 'Blade Runner'  } |
|  |

* ***Druga faza***

Faza [*$unwind*](https://docs.mongodb.com/manual/reference/operator/aggregation/unwind/#mongodb-pipeline-pipe.-unwind) u sledeću fazu prosleđuje novi dokument za svaki od elemenata *genres.* To su sledeći dokumenti:

|  |
| --- |
| {  imdb: { rating: 8.2, votes: 419589, id: 83658 }, |
|  |

|  |
| --- |
| genres: 'Sci-Fi', |
|  |

|  |
| --- |
| title: 'Blade Runner' |
|  |

|  |  |
| --- | --- |
| } | |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| imdb: { rating: 8.2, votes: 419589, id: 83658 }, |
|  |

|  |
| --- |
| genres: 'Thriller', |
|  |

|  |
| --- |
| title: 'Blade Runner'  } |
|  |

* ***Treća faza***

Faza [*$group*](https://docs.mongodb.com/manual/reference/operator/aggregation/group/#mongodb-pipeline-pipe.-group) najpre vraća distinktne vrednosti dokumenata iz prethodne faze, zatim za svaku takvu vrednost polja *genre* kreira dokument čiji \_id je jednak toj vrednosti. Na kraju dodaje polje *averageGenreRating* svakom novom dokumentu, koje sadrži prosečnu vrednost polja *imdb.rating* onih dokumenata koji pripadaju tom istom žanru. Ovi dokumenti se prosleđuju u sledeću fazu. Neki od dokumenata nastalih u ovoj fazi su sledeći:

|  |
| --- |
| { \_id: 'History', averageGenreRating: 7.202306920762287 }, |
|  |

|  |
| --- |
| { \_id: 'Biography', averageGenreRating: 7.097142857142857 }, |
|  |

|  |
| --- |
| { \_id: 'Adventure', averageGenreRating: 6.527788649706458 }, |
|  |

|  |
| --- |
| { \_id: 'Family', averageGenreRating: 6.36096256684492 }, |
|  |

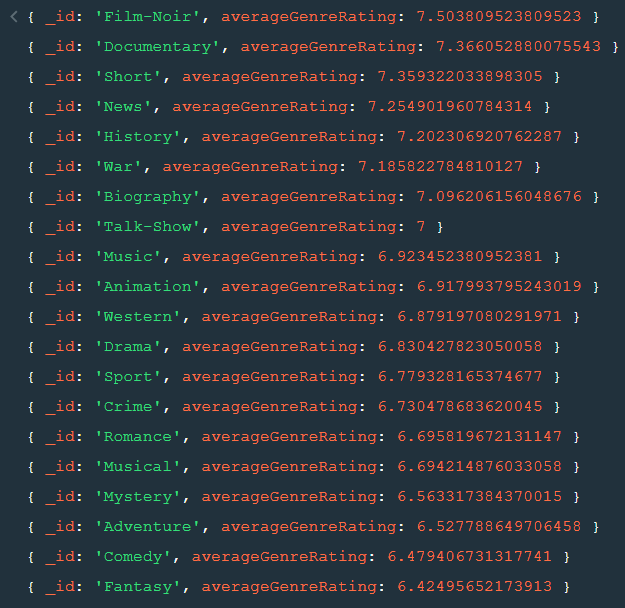
|  |
| --- |
| { \_id: 'Crime', averageGenreRating: 6.730478683620045 }, |
|  |

|  |
| --- |
| { \_id: 'Western', averageGenreRating: 6.879197080291971 }, |
|  |

|  |
| --- |
| { \_id: 'Fantasy', averageGenreRating: 6.42495652173913 }, |
|  |

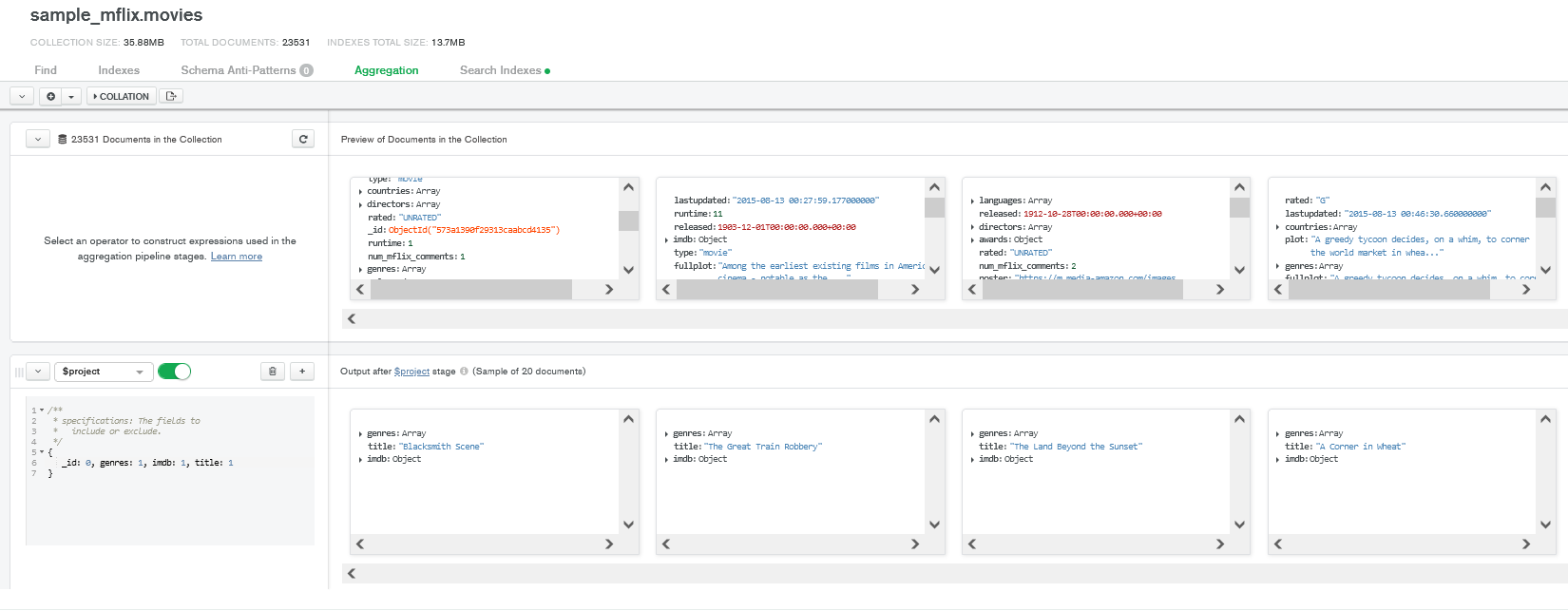
* ***Četvrta faza***

Faza [*$sort*](https://docs.mongodb.com/manual/reference/operator/aggregation/sort/#mongodb-pipeline-pipe.-sort) sortira dokumente iz prethodne faze u opadajućem redosledu vrednosti polja *averageGenreRating* field. Konačno, dobija se rezultat prikazan na slici 20.



Slika 20 – rezultat pipeline-a agregacije nad kolekcijom movies u mongocli-u

*Data Explorer* MongoDB Atlas-a pruža interfejs za kreiranje pipeline-ova agregacije, uz sugestije polja svake od faza i prikaz dokumenata koji su rezultat svake od njih. Na slici 21 se vidi kako bi izgledalo kreiranje prve faze prethodno objašnjenog pipeline-a.



Slika 21 – kreiranje pipeline-a agregacije u Data Explorer-u Atlas-a

## Pristup

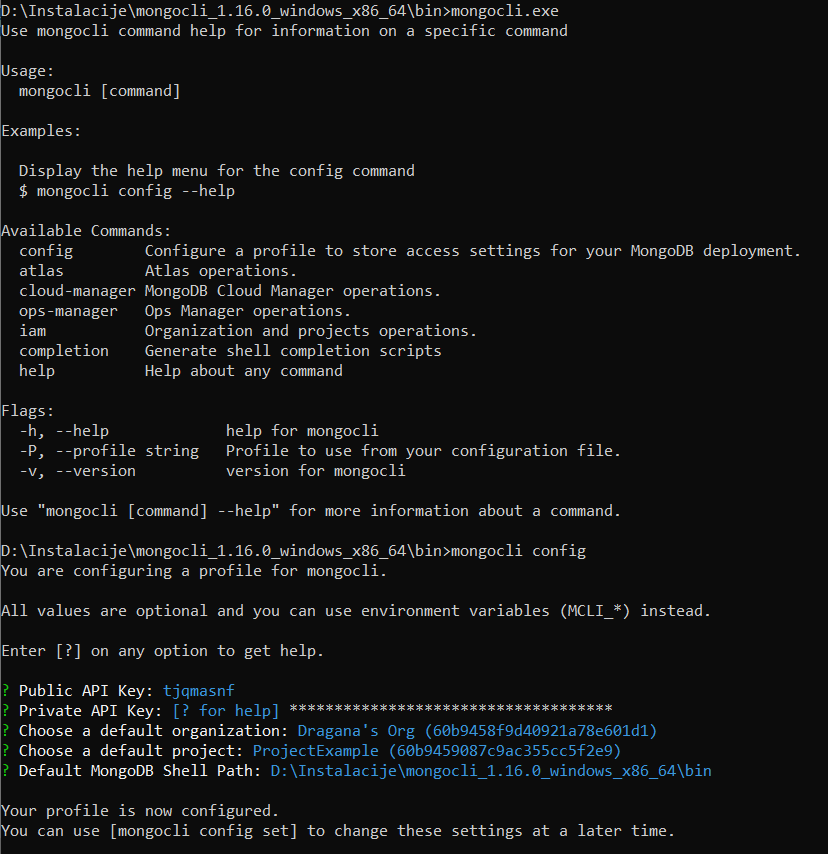
MongoDB Atlas-u moguće je pristupiti i preko MongoDB CLI-a, kao i grafičkog korisničkog interfejsa MongoDB Compass, koji će biti opisani u narednim odeljcima.

### MongoDB CLI

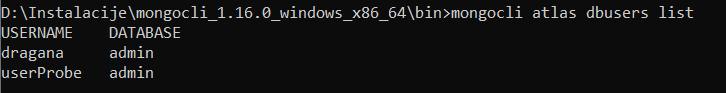
MongoDB Command Line Interface (mongocli ili MongoDB Shell) služi za kreiranje i ažuriranje MongoDB baze u oblaku preko MongoDB Atlas-a, ili lokalne baze podataka preko MongoDB Cloud Manager-a i MongoDB Ops Manager-a. MongoDB CLI ubrzava interakciju i testiranje MongoDB servisa.

Korišćenje mongocli-a biće demonstrirano u nastavku. Nakon kreiranja API ključeva organizacije (Slika 4) omogućen je pristup preko ovog interfejsa. Najpre se treba pozicionirati u *bin* folder mongocli instalacije i pokrenuti datoteku *mongocli.exe*. Zatim treba definisati konfiguraciju podrazumevanog profila. Mongocli će zatražiti javni i privatni API ključ, a nakon njihovog unosa ponuditi organizacije kojima odgovarajuća instanca Atlas-a pristupa. Treba izabrati organizaciju i projekat u okviru nje, a zatim i podrazumevanu putanju do MongoDB Shell-a (Slika 22).

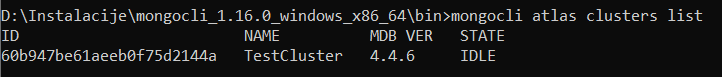
Preko mongocli-a moguće je pristupiti korisnicima baze sa svojim privilegijama (Slika 23), kao i osnovnim informacijama o klasterima prisutnim u organizaciji (Slika 24). Rezultat sa slike 7 daje i komanda *mongocli atlas cluster describe TestCluster*, pošto se njome opisuje jedini prisutni klaster, pod imenom *TestCluster*. Komande za kreiranje i ažuriranje klastera nisu dostupne za M0, M2 i M5 klastere.



Slika – konfiguracija podrazumevanog profila u mongocli-u

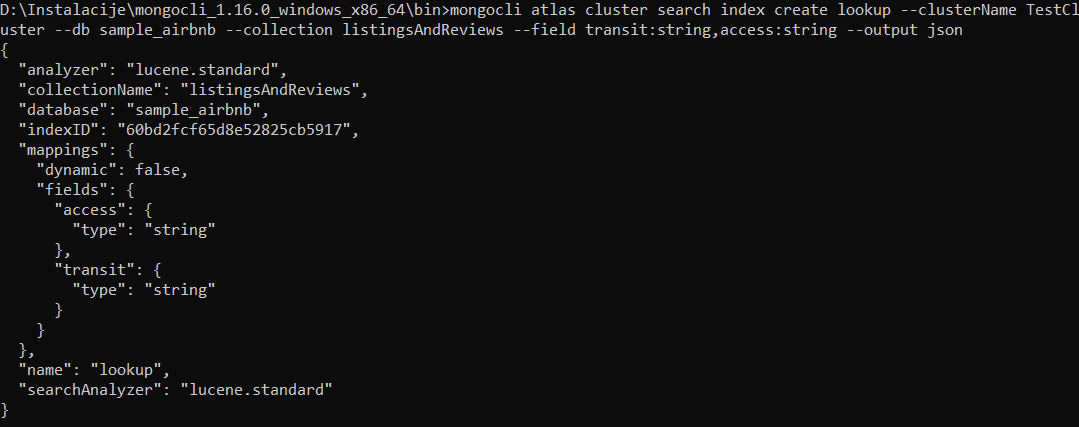


Slika – lista korisnika baze podataka



Slika – lista klastera MongoDB-a

Preko MongoDB Shell-a može se kreirati indeks za pretragu, navođenjem imena klastera, baze i kolekcije, polja po kome se vrši indeksiranje i programskog jezika u kome će biti prikazan kreirani indeks. Podrazumevani analizator je standardni analizator (*lucene standard analyzer*), mapiranje statičko, a ID se automatski dodeljuje. Dat je primer kreiranja indeksa pod nazivom *lookup* (Slika 25).



Slika – kreiranje indeksa lookup preko mongocli-a

### MongoDB Compass

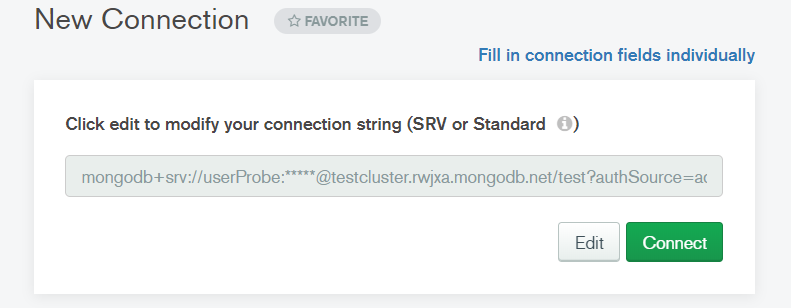
MongoDB Compass je GUI MongoDB-a i predstavlja najbolji način da se pristupi i manipuliše podacima u ovoj bazi podataka. Moguće je vizuelno istražiti podatke, pokrenuti ad hoc upite u nekoliko sekundi, kao i pregledati i optimizovati performanse upita. Takođe se može proširiti funkcionalnost dodavanjem plugin-ova. MongoDB Compass pruža mogućnost za lakše donošenje odluka o indeksiranju, validaciji dokumenata i dr. Neke od glavnih prednosti MongoDB Compass-a su sledeće:

* *Podaci se mogu istražiti putem ugrađene vizuelizacije šeme*– Compass prikazuje bogate strukture unutar kolekcija preko intuitivnog GUI-ja, koje su nastale na osnovu analize dokumenata u bazi. Omogućava brzu vizuelizaciju i istraživanje šeme, da bi se razumela učestalost, tipovi i opsezi polja u skupu podataka.
* *Uvid u status servera i performanse upita*– real-time server statistike omogućavaju da se vide ključne metrike servera i operacije nad bazom podataka. Može se dobiti jednostavan uvid i u najkorišćenije kolekcije (Slika 26).

|  |  |
| --- | --- |
|  |  |
| Slika – vizuelizacija operacija, upisa i čitanja | Slika – modifikacija dokumenta preko GUI-a |

* *Bolji pristup CRUD operacijama olakšava interakciju sa podacima*– modikacija postojećih dokumenata je omogućena i putem intuitivnog vizuelnog alata za izmenu, a mogu se ubaciti, klonirati ili obrisati dokumenti u samo par klikova (Slika 27).
* *Razumevanje problema u performansama preko vizuelnih objašnjenja* – prikaz kako se upiti izvršavaju preko GUI-ja, koji pomaže u identifikaciji i rešavanju problema u performansama (Slika 30).
* *Uvid u iskorišćenost baze podataka i upravljanje indeksima* – može da se razume tip i veličina indeksa, njihova iskorišćenost i specijalne karakteristike. Dodavanje i uklanjanje indeksa je jednostavno.
* *Jednostavnija validacija podataka –* moguće je napisati pravila validacije JSON šeme u editoru koji vrši sugestiju imena polja, BSON tipova podataka i ključnih reči za validaciju. Takođe se može videti koji dokumenti zadovoljavaju pravilo validacije.
* *Olakšane agregacije –* omogućeno jekreiranje agregacionih pipeline-ova u intuitivnom korisničkom interfejsu. Faze se kreiraju uz pomoć automatske dopune, pri čemu se vidi njihovo dejstvo u pregledu dokumenta. One mogu da se dodaju, brišu i preraspoređuju, a kreirani pipeline se eksportuje u odgovarajući kod i zatim koristi u aplikaciji.

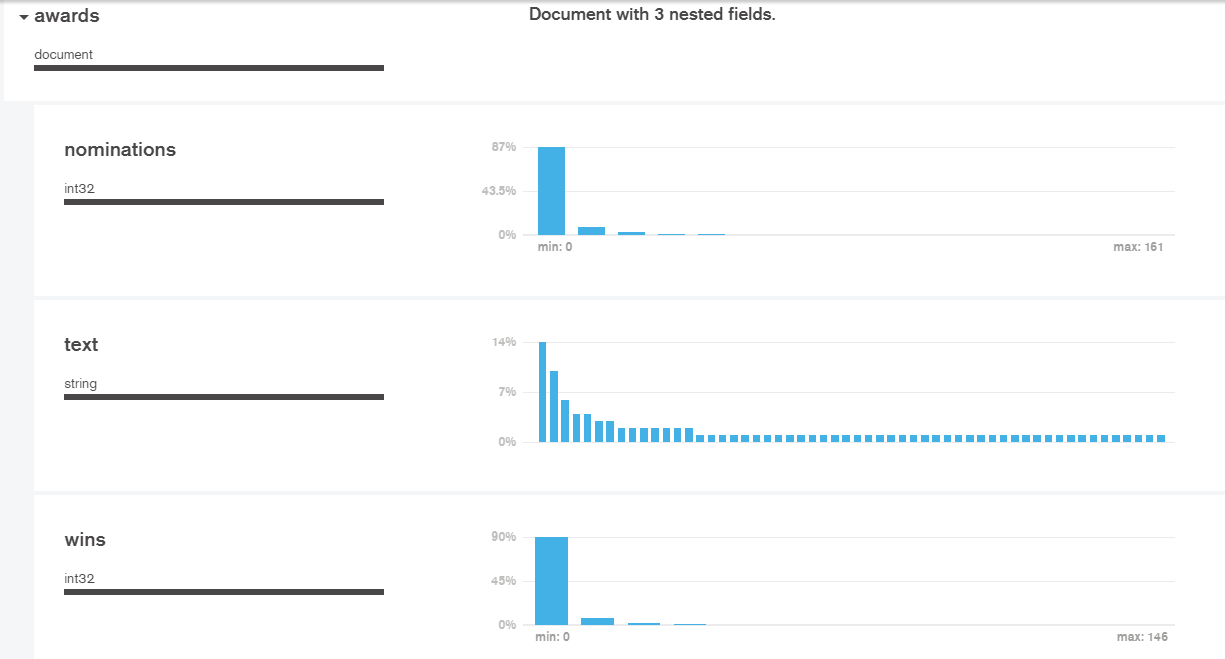
Da bi se povezalo sa klasterom MongoDB Atlas-a preko Compass-a, u dijalog *New Connection* unosi se konekcioni string, generisan u ovom servisu (Slika 28). Glavni prozor Compass-a pruža uvid u klaster na koji je povezan, njegove kolekcije i baze podataka.



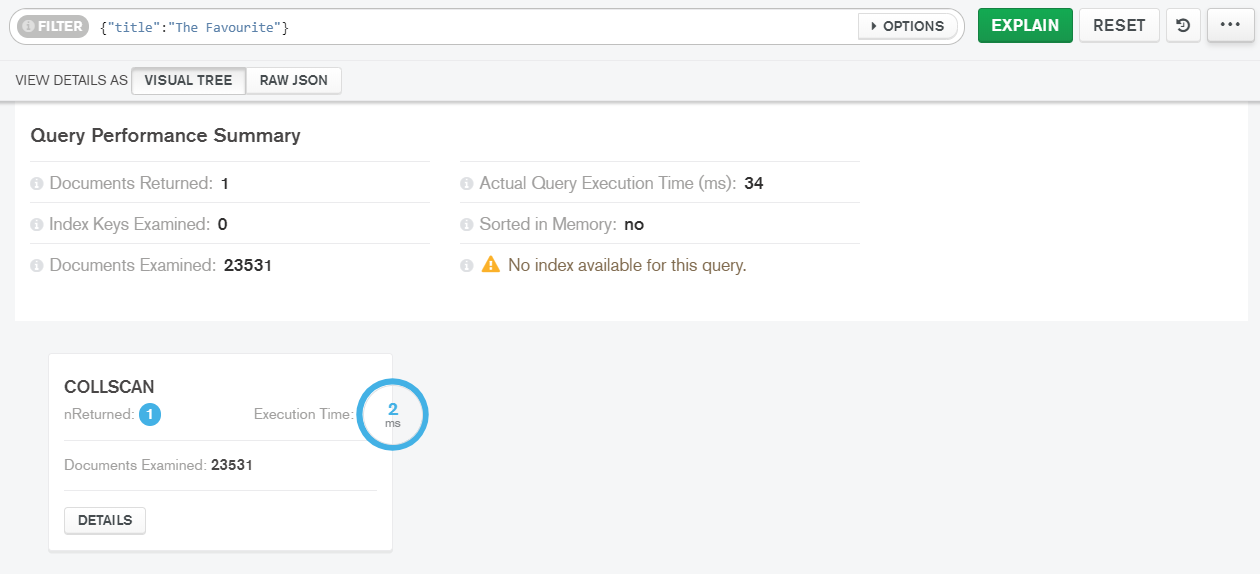
Slika – povezivanje preko konekcionog string-a

Preko kartice Shema izabrane kolekcije može se videti vizuelizacija šeme podataka, koja je zasnovana na nasumično odabranih 1000 dokumenata. Na slici 29 prikazan je deo statističkih podataka o šemi kolekcije *movies*, baze podataka *sample\_mflix*, gde se vidi procenat zastupljenih vrednosti polja odabranih dokumenata (sampling). Kartica *Explain Plan* služi za prikaz podataka o upitu i njegovim performansama (Slika 30).

MongoDB Compass takođe poseduje ugrađen MongoDB Shell, koji se aktivira preko opcije *\_MONGOSH BETA*. Na taj način se može pristupiti MongoDB Atlas klasteru preko odgovarajućih komandi, pisanih u programskom jeziku JavaScript. Na slici 31 dat je primer ubacivanja dokumenta preko komande u JavaScript-u, gde je struktura dokumenta u JSON formatu. Ovoj komandi mora da prethodi komanda *use sample\_mflix*, da se naglasi da se promenljiva *db* odnosi na ovu bazu podataka. Nakon uspešno izvršenog dodavanja prikazaće se potvrda uspeha i ID ubačenog dokumenta. Preko funkcije *find*, uz navođenje filtera pretrage, dobija se kao odgovor dokument sa traženim naslovom (Slika 32). U polju za pretragu MongoDB Compass-a dovoljno je navesti filter pretrage i dobiti identičan rezultat kao na slici 32 (Slika 33).



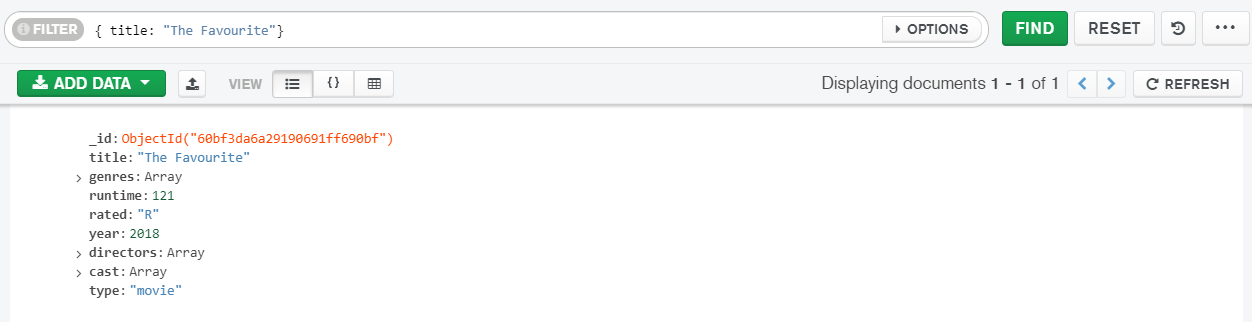
Slika – kartica Shema MongoDB Compass-a



Slika – uvid u performanse izvršenog upita

Pošto je u odeljku 5.2.3. kreiran statički indeks *defaultIndex* nad kolekcijom *movies*, omogućena je pretraga po standardnom ili *keyword* analizatoru za polje *title*. Može se naglasiti korišćenje drugog, da bi se pronašao naslov identičan tekstu pretrage (Slika 34). Ovaj upit se može izvršiti samo prilikom statičkog mapiranja indeksa. Rezultati upita, svi filmovi pod nayivom *The Count of Monte Cristo*, prikazani su na slici 35.

|  |  |
| --- | --- |
|  |  |
| Slika – ubacivanje dokumenta u kolekciju | Slika – pronalaženje dokumenta po naslovu |



Slika – pronalaženje dokumenta po naslovu preko Compass-a

|  |  |
| --- | --- |
|  |  |
| Slika – pronalaženje dokumenta na osnovu keyword polja | Slika – rezultati upita sa slike 34 |

# Zaključak

Nakon upoznavanja sa okruženjima i konceptima kojima se rad bavi, sledi osvrt na njih, kao i na zaključke koji se mogu izvesti nakon njihove obrade. Najpre je bilo reči o database-as-a-service, platformi za hosting podataka u oblaku. Infrastruktura, skladište podataka, softver baze podataka, licence, replikacija, oporavak od otkaza i backup informacije su uključene u pretplatu na ovu platformu.

Računarstvo u oblaku omogućava rast baze podataka i pristup dodatnom kapacitetu obrade, kada je potrebno. Ovo je posebno korisno kada je u pitanju nerelaciona baza podataka, kao što je MongoDB, koja omogućava horizontalno skaliranje podataka. Računarstvo u oblaku čini izgradnju skalabilnih baza podataka mogućom. Database-as-a-service baze podataka, kao što je MongoDB, oslanjaju se na platforme u oblaku.

MongoDB Atlas je nerelaciona baza podataka koja se nalazi na izabranoj platformi u oblaku (Amazon Web Services (AWS), Microsoft Azure ili Google Cloud Platform). MongoDB Atlas pridružen Google Cloud platformi je svojim korisnicima omogućio veću agilnost u skaliranju i upravljanju resursima. Takođe, njegovi klasteri se mogu nalaziti u više različitih regiona.

U radu su dati primeri konfigurisanja klastera i rada sa dokumentima i indeksima iz korisničkog interfejsa Atlas-a. Ovom bazom podataka može se upravljati preko alata MongoDB Shell i MongoDB Compass. Prvi je koristan za jednostavno i brzo izvršavanje osnovnih operacija sa bazom podataka, dok drugi pruža dodatne mogućnosti pregleda ponašanja podataka. Moguće je identifikovati tačke zastoja ili spor upit koji remeti performanse, što omogućava dolaženje do bržeg i boljeg rešenja. Takođe, Compass pruža mogućnost ažuriranja indeksa.

Zbog jednostavne konfiguracije klastera, zauzimanja resursa i dobrih performansi, korišćenje MongoDB Atlas-a na Google Cloud platformi je preporučljivo za projekte koji zahtevaju efikasnost i veliko skladište podataka. Omogućeno je skaliranje ovih podataka, kao i dobra dostupnost. Takođe je garantovana bezbednost podataka i pristup rešenjima koja se zasnivaju na veštačkoj intelilgenciji i mogu da optimizuju rad različitih delova projekta.

# 

# Literatura

1. *Database As A Service (DBaaS) Explained*, dostupno na: https://www.mongodb.com/database-as-a-service (pristupljeno 3. 6. 2021)
2. *MongoDB Atlas*, dostupno na: <https://www.mongodb.com/cloud/atlas> (pristupljeno 5. 6. 2021)
3. *Google Cloud Platform*, dostupno na: <https://en.wikipedia.org/wiki/Google_Cloud_Platform> (pristupljeno 5. 6. 2021)
4. *Accelerate your transformation with Google Cloud*, dostupno na: <https://cloud.google.com/> (pristupljeno 5. 6. 2021)
5. *MongoDB Atlas on Google Cloud*, dostupno na: <https://cloud.google.com/mongodb> (pristupljeno 5. 6. 2021)
6. *MongoDB Compass*, dostupno na: <https://www.mongodb.com/products/compass> (pristupljeno 5. 6. 2021)
7. *Configure the MongoDB CLI*, dostupno na: <https://docs.mongodb.com/mongocli/master/configure/#std-label-mcli-configure> (pristupljeno 7. 6. 2021)
8. *Manage Documents in Data Explorer*, dostupno na: <https://docs.atlas.mongodb.com/data-explorer/documents/#std-label-data-explorer-docs> (pristupljeno 7. 6. 2021)
9. *Mongocli atlas cluster search index*, dostupno na: listhttps://docs.mongodb.com/mongocli/v1.8/reference/atlas/search-index-list/ (pristupljeno 7. 6. 2021)
10. *Run Atlas Search Queries*, dostupno na: <https://docs.atlas.mongodb.com/reference/atlas-search/tutorial/run-query/#std-label-fts-tutorial-run-query> (pristupljeno 8. 6. 2021)
11. *Run Aggregation Pipelines*, dostupno na: <https://docs.atlas.mongodb.com/data-explorer/cloud-agg-pipeline/> (pristupljeno 9. 6. 2021)
12. *Run Aggregation Pipelines*, dostupno na: <https://docs.mongodb.com/mongodb-shell/run-agg-pipelines/> (pristupljeno 9. 6. 2021)

1. Infrastructure as a service (IaaS) je tip servisa računarstva u oblaku koji pruža ključne resurse za izračunavanja, skladištenje i mrežno povezivanje na zahtev korisnika. [↑](#footnote-ref-1)
2. Platform as a service (PaaS) je kategorija servisa računarstva u oblaku koja omogućava zauzimanje, kreiranje, pokretanje i upravljanje celinom koju čini platforma za izračunavanje i jedna ili više aplikacija. Pri tome se izbegava kompleksnost kreiranja i održavanja potrebne infrastrukture i omogućava kreiranje, razvoj i pripremanje za korišćenje ovakvih celina. [↑](#footnote-ref-2)
3. Serverless computing predstavlja arhitekturu gde pružalac usluga u oblaku u potpunosti upravlja izvršenjem koda, umesto da se aplikacije smeštaju na servere na tradicionalan način. [↑](#footnote-ref-3)
4. Servis za upravljanje klučevima koji omogućava upravljanje simetričnim i asimetričnim kriptografskim ključevima servisa na oblaku, na isti način kao i ključevima lokalnih servisa. [↑](#footnote-ref-4)
5. Cloud Provider Snapshots su kopije podataka na oblaku koje omogućavaju potpuno upravljivo rezervno skladište i mogućnost oporavka podataka preko pružaoca servisa u oblaku. [↑](#footnote-ref-5)
6. Omogućava pristup prozoru *Data Explorer* i akcijama kao što su: pregled, kreiranje i brisanje baza podataka, kolekcija i indeksa; pregled, ažuriranje i brisanje dokumenata. [↑](#footnote-ref-6)
7. Omogućava pristup prozoru *Data Explorer* i akcijama kao što su: pregled i kreiranje baza podataka i kolekcija; pregled, ažuriranje i brisanje dokumenata; pregled indeksa. [↑](#footnote-ref-7)
8. Omogućava pristup prozoru *Data Explorer* i akcijama kao što su: pregled baza podataka i kolekcija; pregled dokumenata i indeksa. [↑](#footnote-ref-8)
9. Dokument je osnovna jedinica podataka u MongoDB bazi i skladišti se u JSON formatu. [↑](#footnote-ref-9)
10. Analizatori kontrolišu način konstruisanja indeksa i može se izabrati njihov tip tako da se rečnik i gramatika određenog jezika najbolje iskoristi. Standardni analizator je podrazumevan za svaki indeks za pretragu MongoDB Atlas-a. Neutralan je za jezike jer svaki termin odgovara jednoj reči, pri čemu se velika slova zamenjuju malim i uklanja se interpunkcija. [↑](#footnote-ref-10)