

Onderzoeksverslag MongoDB

ISE-PROJECT GROEP B2

Auters: Bram Bakker (599907)

Datum: 28-04-22

Versie: 1.0

Docent: Michel Koolwaaij • Frank Tempelman • Helen Visser

Inhoudsopgave

1. Inleiding
2. Onderzoeksvragen
 - 2.1. Wat is MongoDB?
 - 2.2. Hoe werkt MongoDB?
 - 2.3. Hoe kunnen we data importeren naar MongoDB?
 - 2.4. Hoe zorgen we ervoor dat we de data vanuit SQL kunnen exporteren in deze vorm?
3. Conclusie
4. Bibliografie

Inleiding

Het bedrijf NUTMEG wil gebruik maken van een staging area/API om de data op te kunnen halen voor andere partijen. Zij willen dat andere gebruikers de data kunnen inzien via een andere database dan MSSQL zelf. Zij geven aan hier MongoDB te voor willen te gebruiken. Dit zou inhouden dat MongoDB de data vanuit MSSQL op haalt, en deze vervolgens kan tonen aan een gebruiken van bijvoorbeeld een website. Omdat wij geen kennis hebben van MongoDB en de connectie tussen deze twee databases hebben wij de volgende hoofdvraag opgesteld: "Hoe kunnen we MongoDB gebruiken als staging area voor onze database?" Om deze vraag goed te kunnen beantwoorden hebben wij ook een reeks deelvragen opgesteld, namelijk:

- Wat is MongoDB?
- Hoe werkt MongoDB?
- Hoe kunnen we data importeren naar MongoDB?
- Hoe zorgen we ervoor dat we de data vanuit SQL kunnen exporteren in deze vorm?

Naast dit onderzoeksverslag hebben we ook al een aantal dingen met MongoDB moeten implementeren. Een aantal van deze stukken zullen ook in dit verslag terugkomen.

Onderzoeksvragen

In dit hoofdstuk zullen we de deelvragen van dit document beantwoorden.

Wat is MongoDB?

MongoDB is een open-source database die werkt met documenten. Op de website van MongoDB staat het volgende: "MongoDB is an open-source document database built on a horizontal scale-out architecture that uses a flexible schema for storing data" (MongoDB, z.d.). MongoDB is goed te vergelijken met een relationele database. In een relationele database heb je tabellen, deze heten in MongoDB 'collections'. Deze zijn in principe te vergelijken. Over hoe MongoDB precies in zijn werk gaat straks meer. MongoDB is, in 1 term, een document-oriented database. Ook dit is een vorm van een NoSQL database. Een tweetal voorbeelden van de voordelen van MongoDB zijn: "It is a natural form to store data." (MongoDB, z.d.) en "It is human-readable." (MongoDB, z.d.).

Hoe werkt MongoDB?

Zoals in het vorige hoofdstuk vermeld is MongoDB een document-oriented database. Eerder is ook al aangegeven dat het gebruik van tabellen en collections overeen komen. Het grootste verschil zit hem in hoe de data wordt opgeslagen. In MongoDB wordt alle data opgeslagen als een bestand of een document. Deze zijn vervolgens weer uit te lezen in JSON formaat (of, als gewenst, csv formaat). "MongoDB stores data objects in collections and documents instead of the tables and rows used in traditional relational databases. Collections comprise sets of documents, which are equivalent to tables in a relational database. Documents consist of key-value pairs, which are the basic unit of data in MongoDB." (Purestorage, z.d.).

Hoe kunnen we data importeren naar MongoDB?

Het importeren van data in MongoDB kan op meerdere manieren. Afhankelijk van hoe we de data ophalen en vervolgens weer verwerken zijn er verschillende manieren hoe we deze weer in MongoDB kunnen stoppen. Tijdens het implementeren van MongoDB zijn hier een aantal punten naar boven gekomen. Allereerst hebben we geprobeerd een connectie op te zetten tussen MSSQL en MongoDB met gebruik van Polybase. Dit is uiteindelijk niet gelukt. We hebben hierna een makkelijkere weg gekozen. Als we de data als CSV of JSON hebben kunnen we dit bestand importeren in MongoDB. Dit kan makkelijk via MongoDB Compass. MongoDB Compass is een soort van GUI die iets minder kan dan de shell, maar in principe hetzelfde resultaat geeft voor deze basis dingen. Tijdens het implementeren zijn we ook op een aantal tools gekomen, waaronder mongoimport. Dit is een extensie van MongoDB die je bestanden in CSV of JSON formaat laat importeren via de commandprompt. Je maakt een connectie met de database, geeft aan welke collectie (tabel) je wil gebruiken, en welke data je daarin wilt hebben. Door onze implementatie met MongoDB op te zetten met deze tool zijn wij er vrij zeker van dat dit, voor nu, de beste manier is.

Hoe zorgen we ervoor dat we de data vanuit SQL kunnen exporteren in deze vorm?

Tijdens onze implementatie zijn we steeds een stapje verder gegaan met het automatiseren van het overzetten van data. Om te beginnen zijn we begonnen met het handmatig aanmaken van CSV bestanden gevuld met testdata. Al snel kwamen we op het idee om het maken van een CSV te laten doen door een programmeertaal. Voor nu hebben wij de taal PHP gekozen. Hierdoor zit er gelijk een website bij die gebruikt zou kunnen worden voor het downloaden van alle bestanden in JSON. In php zijn er functies beschikbaar die een bestand kunnen aanmaken als csv en deze vervolgens opslaan op een locatie. Doordat het maken van

een CSV bestand verder veel automatisch gebeurt is het met het oplossen van een paar puzzels snel gebeurt. Zoals eerder aangegeven kunnen we de gegevens daarna makkelijk in MongoDB importeren.

Conclusie

Door het onderzoek zijn we in korte tijd genoeg te weten gekomen over de database MongoDB, hoe deze in elkaar zit en werkt, en hoe wij deze in kunnen zetten voor onze opdracht. Op een vrij eenvoudige en snelle manier kunnen wij de database zo opzetten dat deze de data kan leveren waar nodig. Door het gebruik van MongoDB staat de data die getoond wordt ook los van de data die in de uiteindelijke database staat. Met 1 klik op de knop kan de data ververs worden, om zo de gebruikers de meest accurate data te bieden.

Bibliografie

- MongoDB. (z.d.). Why Use MongoDB And When To Use It? Geraadpleegd op 28 april 2022, van <https://www.mongodb.com/why-use-mongodb>
- Polybase. (z.d.). What Is MongoDB? Purestorage. Geraadpleegd op 28 april 2022, van <https://www.purestorage.com/fr/knowledge/what-is-mongodb.html>