|  |
| --- |
| CphBusiness |
| Gutenberg project |
| Alexander Steen, Jonas Simonsen, Kasper Worm, & Martin Karlsen. |

|  |
| --- |
| Github: https://github.com/Databasserne  24-05-2017 |

# Database

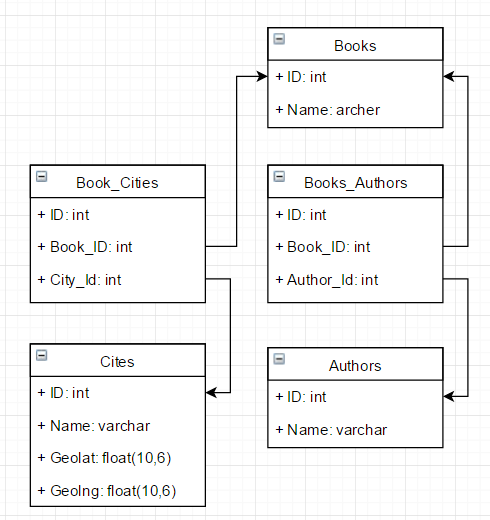
Vi har valgt at bruge en relationel database samt en graph database, i form af MySQL og Neo4J.  
Vi valgte MySQL fordi det er en database vi kender, i forvejen, og vi har valgt Neo4J fordi det var den som virkede mest interessant, at lære.

# Data modelling

## MySQL

Vi har lavet tabeller til hver type objekt (City, Author og Book).  
Vi har lavet many-to-many relationship mellem City og Book, samt Author og Book, hvorfra vi har valgt at lave to tabeller til at danne den relation (Books\_Cities og Books\_Authors).

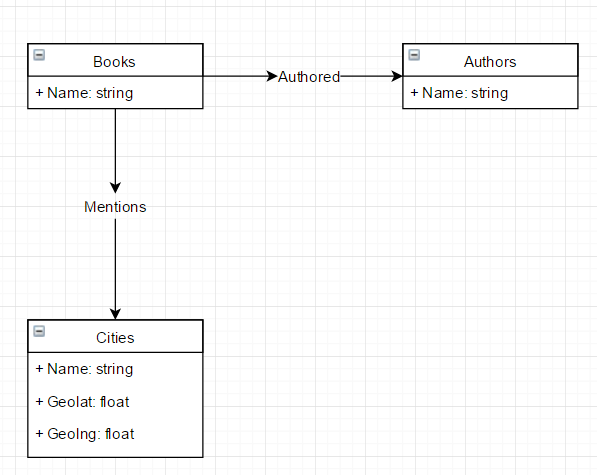
*Mysql relational diagram*



## Neo4J

Vi har lavet en Node for hver type object (Book, Author, City) og har lavet relationerne: Authored og Mentions.

*Neo4J Node diagram*



# Data importing

Data er blevet importeret gennem vores Scraber project på Github.  
Vi har hentet alle bøgerne ned på vores computere og bruger så Scraber til at hente alle byer og indsætte i databaserne, hvorefter den læser igennem alle bøger for at finde ord som matcher de byer, som vi har i databasen. For hver bog, som læses, finder programmet også bogens metafil til at få Author og Title, som indsættes i databasen.

Grunden hastighed finder vi kun byerne, fra bøgerne, som kun består af et ord.

# Query test

|  |  |  |  |
| --- | --- | --- | --- |
| **Query** | **Parameter** | **MySQL (ms)** | **Neo4J (ms)** |
| City | Florence | 2463 | 429 |
| City | Naples | 2510 | 285 |
| City | London | 1583 | 1218 |
| City | Berlin | 1736 | 248 |
| City | Washington | 2889 | 925 |
| Book | La Fiammetta | 948 | 120 |
| Book | The Mystery | 1138 | 78 |
| Book | The King James Version of the Bible | 1153 | 81 |
| Book | A Beautiful Possibility | 1099 | 78 |
| Book | The Warriors | 1127 | 72 |
| Author | Shakespeare, William | 366 | 654 |
| Author | Unknown | 200 | 753 |
| Author | Anonymous | 444 | 2730 |
| Author | Lindsay, Anna Robertson Brown | 76 | 71 |
| Author | Newnes, George | 71 | 110 |
| Location | 37/-122 | 2924 | 175 |
| Location | 42/-81 | 3305 | 314 |
| Location | 44/-70 | 3276 | 596 |
| Location | 55/12 | 2511 | 312 |
| Location | 51/5 | 3393 | 310 |

Alle bøgerne var importeret i Neo4J mens kun 1/3 af bøgerne var importeret i MySQL. Vi danner derfor en hypotese om at MySQL vil være 2-3 gange langsommere end den er nu, hvis alle bøgerne blev importeret.

# Recommendation

Efter vi har kørt vores test, kan vi se at langt de fleste queries er hurtigere i Neo4j. Den eneste som er langsommere på Neo4J er når vi søger på en Author, men vi forventer at kunne lave den hurtigere, hvis det skulle bruge i production.

Til et projekt af sådan type, anbefaler vi at man bruger Neo4J da den overordnet set, er langt hurtigere.