

# Quiz 1: grote bedrijven

Amazon  
Facebook  
Google  
Twitter  
Youtube  
Wikipedia

?	JavaScript	C, C++, Go, <sup>[3]</sup> Java, Python	BigTable, <sup>[4]</sup> MariaDB <sup>[5]</sup>
?	Flash, JavaScript	C/C++, Python, Java, <sup>[6]</sup> Go <sup>[7]</sup>	MySQL, BigTable
?	JavaScript	Java, C++, Perl <sup>[13]</sup>	
?	JavaScript	C++, Java, Scala, Ruby on Rails <sup>[15]</sup>	MySQL <sup>[16]</sup>
?	JavaScript	PHP	MySQL, MariaDB <sup>[14]</sup>
?	JavaScript	Hack, PHP, C++, Java, Python, Erlang, D, <sup>[8]</sup> Xhp <sup>[9]</sup>	MySQL, <sup>[10]</sup> HBase

# Populariteit van databases

Rank	Change	Database	Share	Trend
1		Oracle	36.89 %	-0.4 %
2		MySQL	23.03 %	-1.4 %
3		SQL Server	18.35 %	+0.1 %
4		PostgreSQL	2.9 %	+0.2 %
5	↑↑	MongoDB	2.45 %	+0.4 %
6		SQLite	2.27 %	-0.1 %
7	↓↓	DB2	2.27 %	-0.0 %
8		FileMaker	1.16 %	-0.4 %
9	↑↑	Redis	1.07 %	+0.2 %
10		Teradata	1.02 %	+0.2 %
11	↑	Splunk	1.02 %	+0.2 %
12	↑↑↑↑↑	elasticsearch	0.96 %	+0.5 %
13	↑↑	Apache Hive	0.96 %	+0.3 %
14	↓↓↓↓↓	Sybase	0.83 %	-0.1 %
15	↓↓	Solr	0.76 %	-0.0 %

16	↓↓	Firebird	0.68 %	-0.0 %
17	↓	SAP HANA	0.6 %	+0.2 %
18	↑↑	MariaDB	0.41 %	+0.1 %
19		hbase	0.34 %	-0.0 %
20	↑↑	Ingres	0.32 %	-0.0 %
21	↓↓↓	memcached	0.29 %	-0.1 %
22	↓	Informix	0.27 %	-0.0 %
23		Neo4j	0.23 %	-0.0 %
24	↑	Netezza	0.19 %	-0.0 %
25	↑↑	CouchBase	0.17 %	-0.0 %
26	↓↓	CouchDB	0.17 %	-0.0 %
27	↓	Riak	0.15 %	-0.0 %
28	↑	DynamoDB	0.13 %	-0.0 %
29	↓	dBase	0.11 %	-0.0 %

Bron: ppyl, gebaseerd op google trends

# SQL is nog steeds de standaard voor databases



- SQL ondersteunt alle soorten data en biedt goede performance voor een heel veel soorten acties. Het is een heel goede all-round database
- SQL-databases zijn ACID: data wordt niet vergeten, verwisseld of verkeerd verwerkt
- Dankzij SQL is er geen vendor lock-in: het is mogelijk (met kleine aanpassingen) om de database te veranderen.

# Meest populaire SQL databases

Database	Licentiekosten	Wanneer gebruikt
PostgreSQL	Nee (open source)	Populair alternatief voor MySQL
MS SQLserver	Ja	Meestgebruikte database in .NET omgevingen
MySQL	Nee (open source)	Gekocht door Oracle. Onderdeel van LAMP architectuur voor websystemen: Linux Apache MySQL PHP
Oracle	Ja	Gratis/goedkope licenties voor kleine toepassingen

# Beperkingen van SQL hebben geleid tot alternatieven

- Snelheid: Veel websites hebben 1 bepaald soort queries heel vaak en moeten daarvoor worden geoptimaliseerd.
- Efficiency: SQL is goed maar niet altijd efficient. Grote bedrijven als google kunnen miljoenen besparen op database-hosting met efficiëntere databases.
- Structuur: relationele databases gaan uit van veel relaties tussen data. Soms is data minder of anders gestructureerd
- Gebruiksgemak: Andere databases hebben interfaces die beter aansluiten op webstandaarden

# Het CAP theorem onderbouwt waarom er verschillende types database zijn

Elk database-systeem zou drie eigenschappen moeten hebben:

- C: Consistency
- A: Availability: elke server in het systeem is altijd beschikbaar voor lees en schrijf-acties
- P: Partitioning: de data zou ivm schaalbaarheid te verdelen moeten zijn in verschillende partities, met behoud van A en C

Het CAP- theorem (bewezen in 2002) zegt dat een schaalbaar database-systeem niet alle drie de eigenschappen kan hebben. Als een partitie optreedt (door wegvallen netwerkverbinding) kan het systeem consistent of available blijven, maar niet allebei.

# Alternatieve modellen geven in bepaalde omgevingen een hogere performance

Data Model	Performance	Scalability	Flexibility	Complexity
Key-Value Store	high	high	high	none
Column-Oriented Store	high	high	moderate	low
Document-Oriented Store	high	variable (high)	high	low
Graph Database	variable	variable	high	high
<b>Relational Database</b>	<b>variable</b>	<b>variable</b>	<b>low</b>	<b>moderate</b>

Bron: Ben Scofield, tech lead living social, presentation at codemash 2010

# Besproken onderwerpen

- Column-database: Hbase
  - MapReduce paradigma
- CouchDB
  - Gebruik Json als interfacetaal
  - REST
- Graph database: Neo4J
- Aanbevolen literatuur



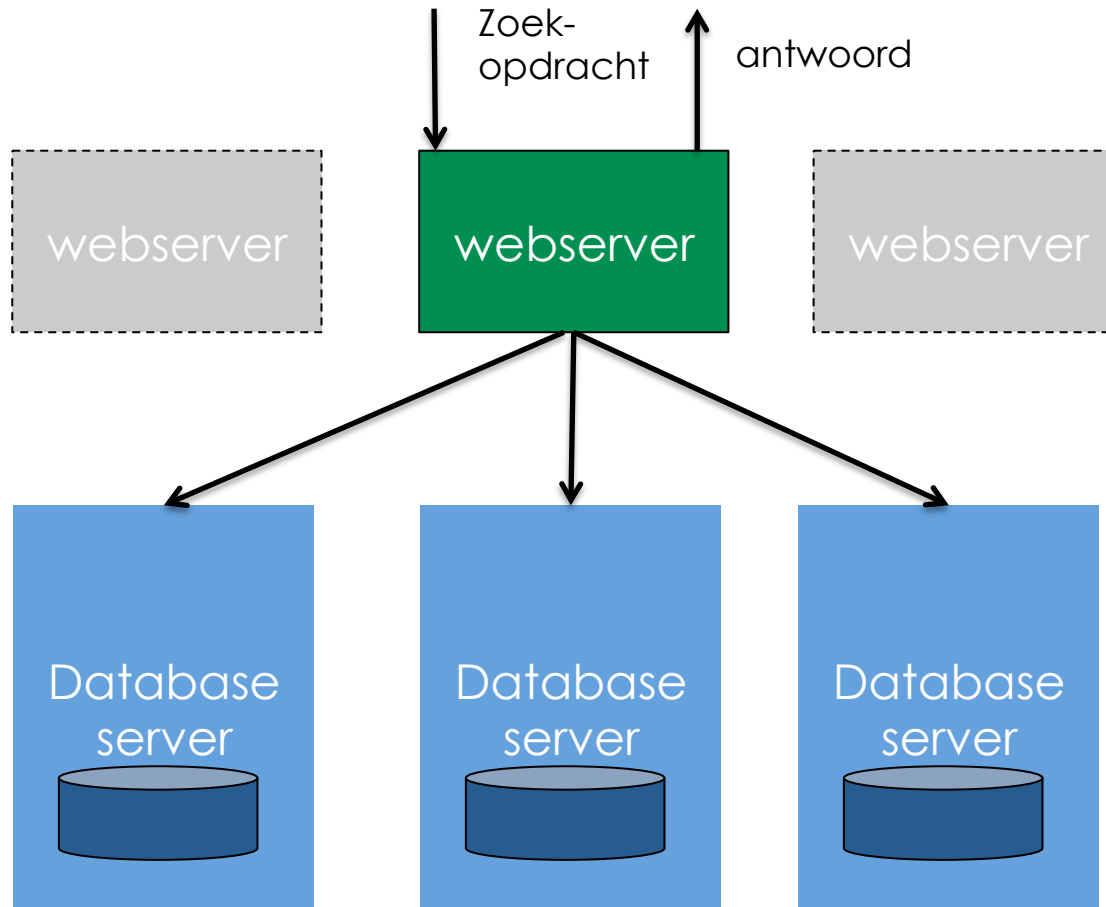
# HBase

- Ontwikkeld op basis van ideeën van google om werk beter te verdelen over meer goedkope computers
- Is bedoeld voor 'big data': goede performance op grote verzamelingen, minder voor kleine verzamelingen
- Is het meest bruikbaar in combinatie met HDFS, Hadoop en Zookeeper

# Voordelen HBase

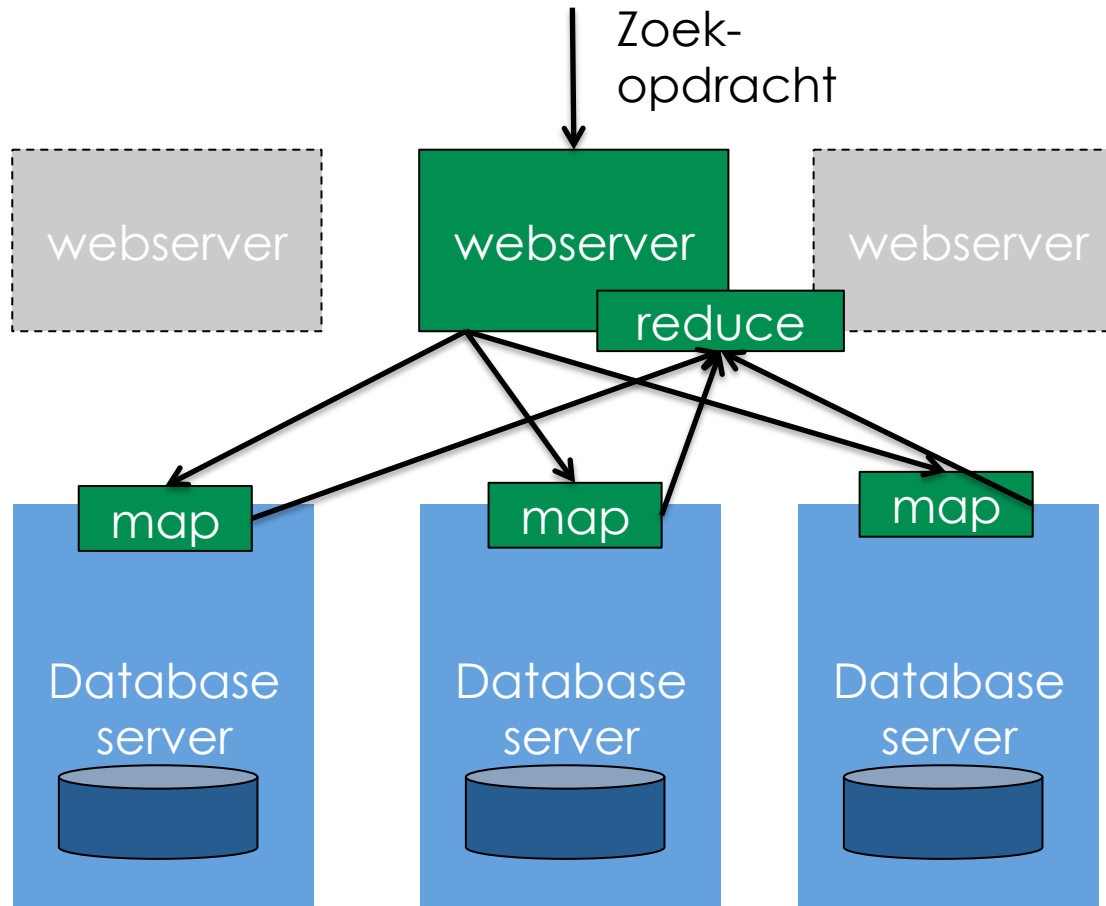
- Inspired by Google's BigTable, nu onderdeel van Apache
- Geschikt voor Big Data, bijvoorbeeld door compressie op kolomniveau, partiitionering.
- Automatische versioning: elke write krijgt een timestamp toegevoegd, en eerdere waarden zijn opvraagbaar
- Onderdeel van Hadoop, de 'familie' van Big Data technology van Apache

Bij SQL worden alle berekeningen door 1 server gedaan, ook al is data verspreid



- De webserver doet een query op alle database servers
- De database servers leveren benodigde data op
- De webserver verwerkt de data tot een antwoordpagina

Bij Mapreduce wordt rekenwerk verdeeld over meerdere servers

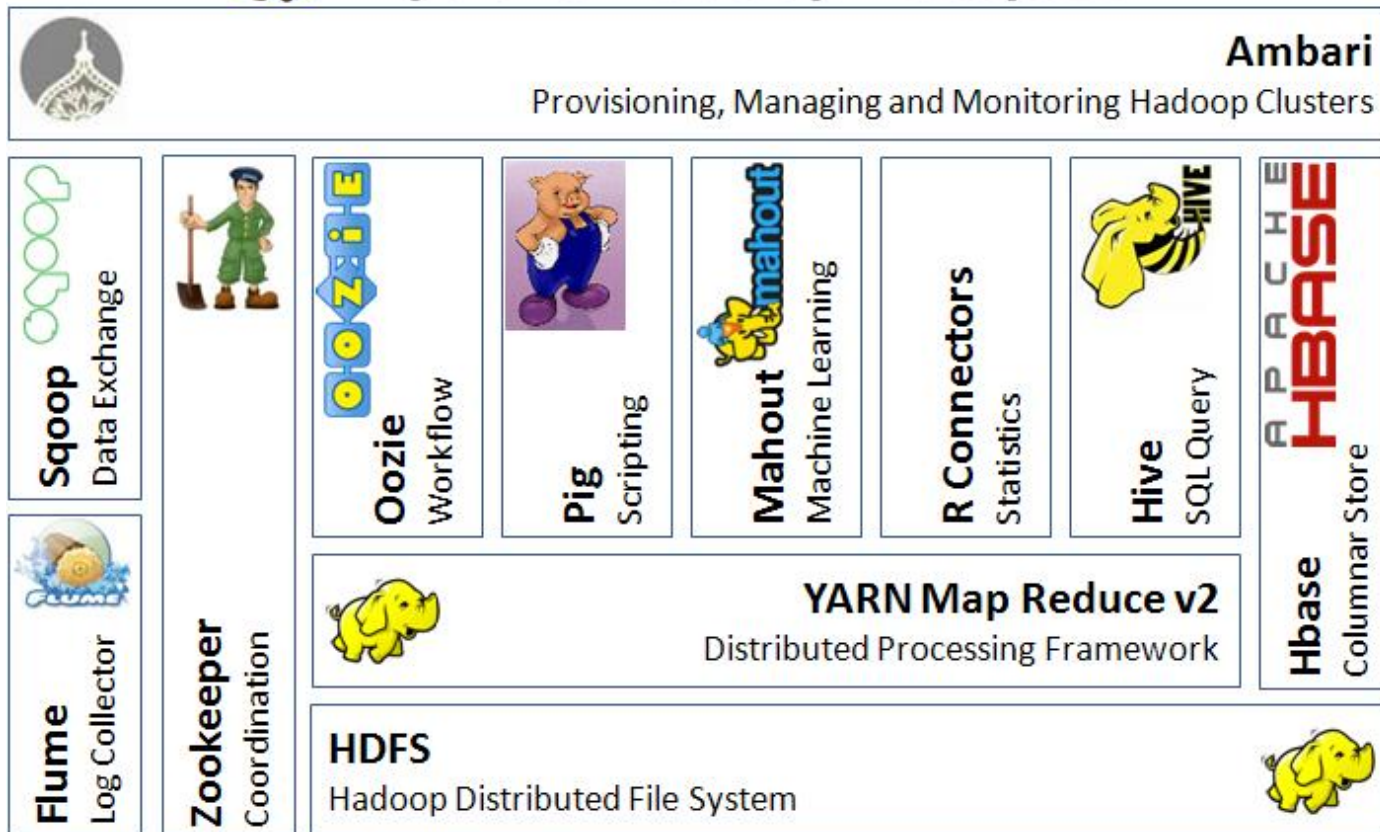


- De webserver stuurt de code voor de “map” stap naar elke database server
- Elke databaseserver runt de map-code op de data en stuurt deelresultaat naar webserver
- De webserver combineert resultaten tot totaalantwoord (“reduce”)

# Hadoop als ecosysteem

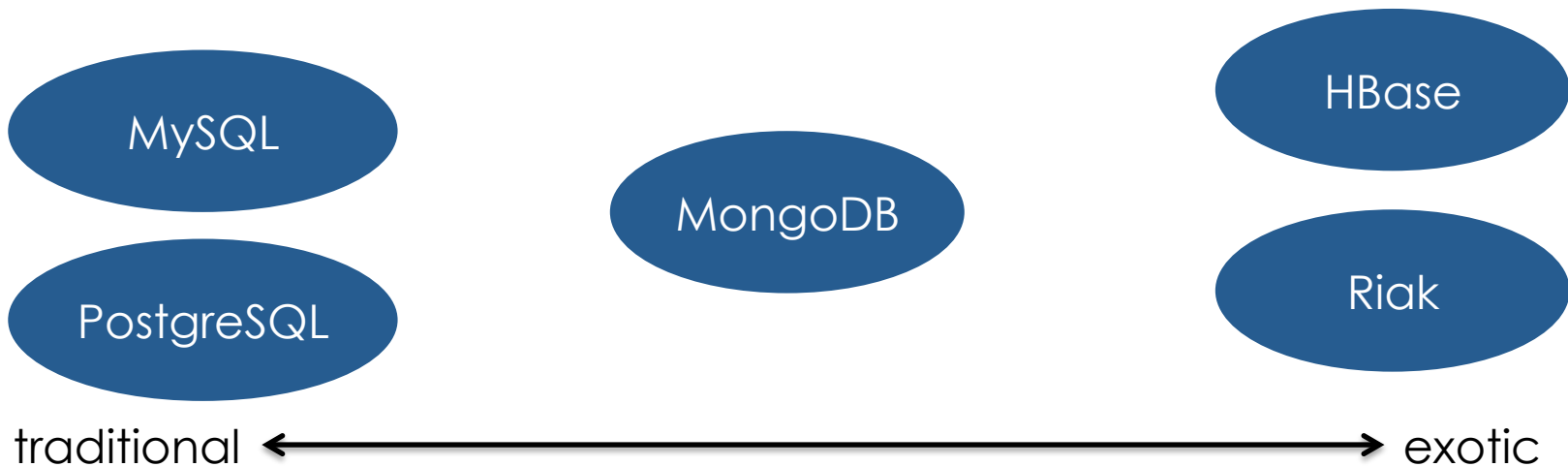


## Apache Hadoop Ecosystem



# MongoDB

- First released in 2009, one of the most popular databases for web systems
- Designed by Dwight Merriman, former CTO of doubleclick
- Open source
- Less different from SQL than other NoSQL database



# Creating with MongoDB

```
$ mongo book
```

```
> db.towns.insert({
  name: "New York",
  population: 22200000,
  last_census: ISODate("2009-07-31"), famous_for:
  [ "statue of liberty", "food" ], mayor : {
  name : "Michael Bloomberg",
  party : "I" }
})

> db.towns.find()
{
  "_id" : ObjectId("4d0ad975bb30773266f39fe3"),
  "name" : "New York",
  "population": 22200000,
  "last_census": "Fri Jul 31 2009 00:00:00 GMT-0700 (PDT)",
  "famous_for" : [ "statue of liberty", "food" ],
  "mayor" : { "name" : "Michael Bloomberg", "party" : "I" } }
```

- MongoDB is schemaless: tables and columns are created by inserting data
- Easy to use, easy to mess up: typos create new tables and columns

# Defining function in MongoDB

```
function insertCity(  
name, population, last_census, famous_for,  
mayor_info  
) { db.towns.insert({  
  name: name,  
  population: population,  
  last_census: ISODate(last_census),  
  famous_for: famous_for,  
  mayor : mayor_info  
}); }
```

```
insertCity("Punxsutawney", 6200, '2008-31-01', ["phil the  
groundhog"], { name : "Jim Wehrle" }  
)
```

- MongoDB speaks Javascript: functions can be defined in javascript



# MongoDB allows for nested queries, like SQL

Find towns with independent mayors...

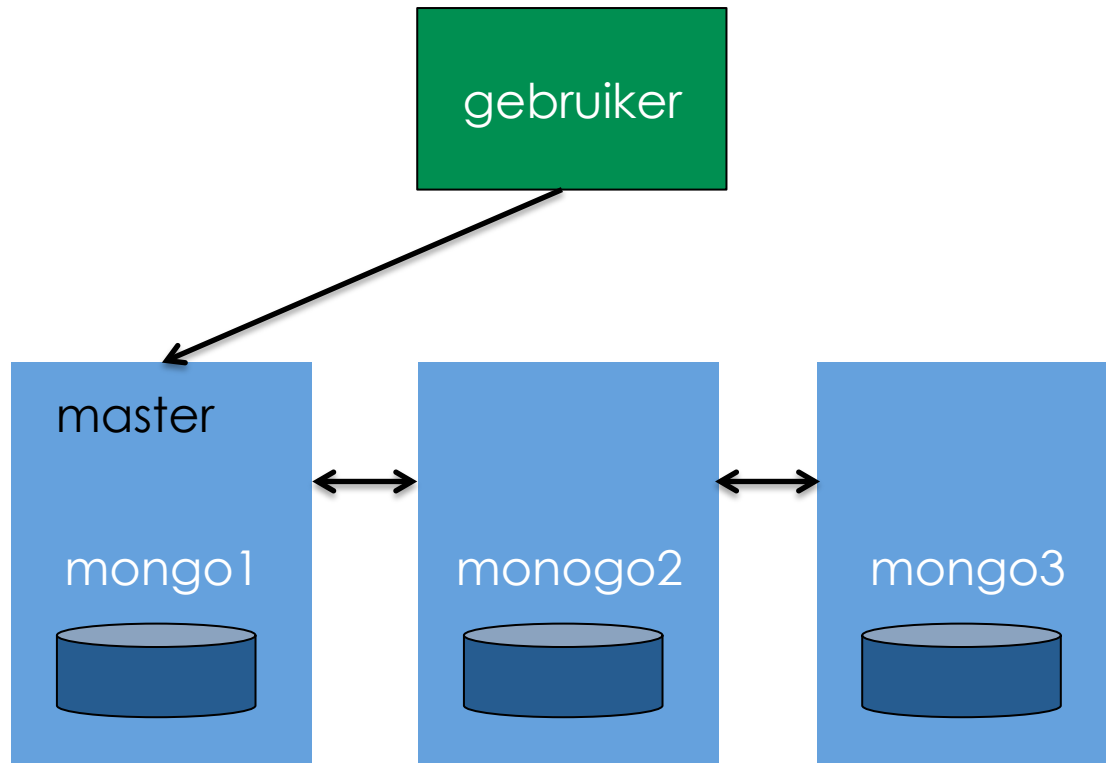
```
db.towns.find(  
  { 'mayor.party' : 'I' },  
  { _id : 0, name : 1, mayor : 1 }  
)
```

Answer:

```
{  
  "name" : "New York",  
  "mayor" : {  
    "name" : "Michael Bloomberg",  
    "party" : "I"  
  }  
}
```

# MongoDB is ontworpen om redundant gebruikt te worden

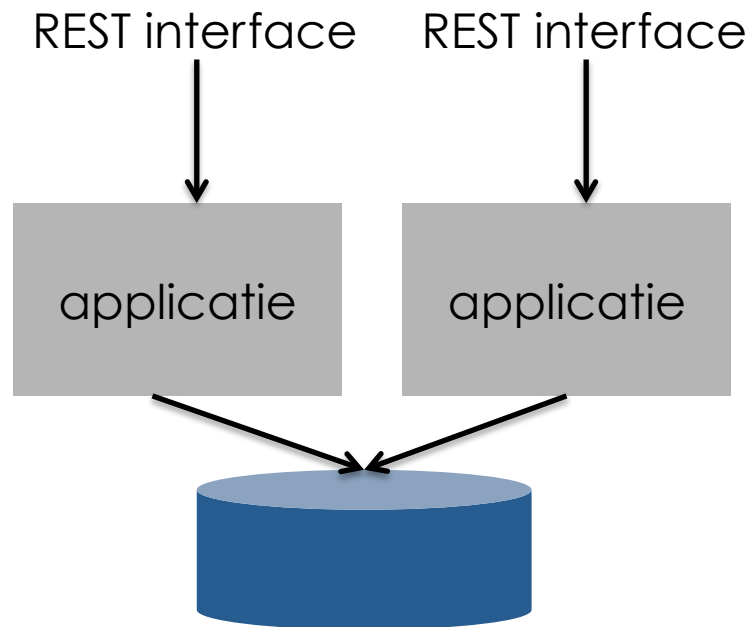
```
$ mkdir ./mongo1 ./mongo2 ./mongo3  
$ mongod --replSet book --dbpath ./mongo1 --port 27011 --rest  
$ mongod --replSet book --dbpath ./mongo2 --port 27012 --rest  
$ mongod --replSet book --dbpath ./mongo3 --port 27013 --rest
```



# Json is een nieuwe interfacetaal

- Plain text of CSV (comma seperated values) was altijd de simpelste interface-taal, maar is foutgevoelig
- XML is ontwikkeld als generieke taal, maar is inefficiënt en moeilijk leesbaar
- Json is een populair alternatief geworden in web-omgevingen.
- Voorbeelden:
  - {naam:"stad1","kleur":"rood",{positie:{x:144,y:16}}}

# REST is een populaire conventie voor webservices



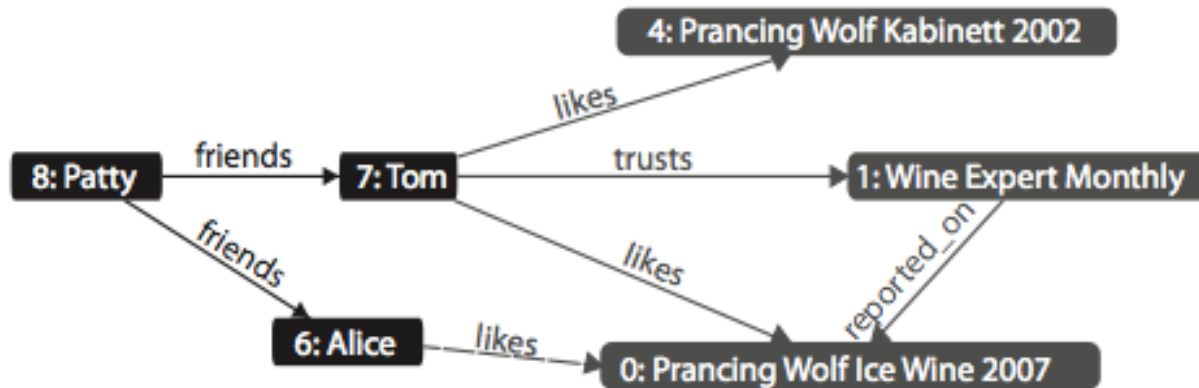
## REST conventies:

- GET-opdrachten veranderen geen data
- Alle opdrachten kunnen veilig worden herhaald
- Alle HTML-foutcodes worden gebruikt zoals bedoeld
- Data blijft niet achter in de applicatie
- REST vaak gebruikt met Json

# Graph databases: Neo4j

- Graph databases slaan alles op als nodes met relaties ertussen.
- Dit kan in theorie ook in SQL, maar dan krijg je alles in 1 tabel met lange queries.
- Ideaal voor semigestructureerde data, of bijvoorbeeld sociale netwerken

# Voorbeeld



Vind de vrienden  
van Alice:

```
alice.bothE('friends').bothV.except([alice]).name
```

Hoeveel wijnen  
vind iedereen  
leuk?

```
gremlin> wines_count = [:]  
gremlin> g.V.outE('likes').outV.name.groupCount( wines_count )  
gremlin>  
wines_count  
==> Alice=1  
==> Tom=2
```

# Aanbevolen literatuur:

Seven databases in seven weeks:

1. SQL: PostgreSQL
2. Web-oriented: Riak
3. Column-store: Hbase
4. Json document database: MongoDB
5. Web-documenten: CouchDB
6. Graph db: Neo4J
7. Fast key-value store: Redis

# Vragen

- Wordt de keuze voor een database-technologie weer een strategische keuze?
- Valt de keuze en inrichting van een database ook onder auteursrecht?
- Wat als een systeem alleen uit een database bestaat?



# Voorbeeld van een API-dienst

The screenshot shows the homepage of the EU VAT API service. The header includes the logo 'EU VAT API' with a percentage icon, and navigation links for 'SIGN UP', 'MY ACCOUNT', 'DOCUMENTATION', and 'BLOG'. The main content area has a dark background with white text. It states: 'Make your online shop EU VAT compliant. Fast and easy.' followed by 'Because of the B2C 2015 EU VAT changes you have to charge your clients the applicable VAT of their country. Our API automates this for you. So you can stick to your core business and still obey this particular law.' Below this is a yellow call-to-action: 'It works perfectly! Try it yourself:'. A light gray box contains a form with three fields: 'I sell' with a dropdown menu showing 'Digital Downloads', 'To' with a dropdown menu showing 'Austria', and 'VAT Rate' showing '20%'. At the bottom, there is a white text 'GET THIS API TODAY!' and a yellow button with 'SIGN UP NOW'.

EU VAT API

SIGN UP MY ACCOUNT DOCUMENTATION BLOG

Make your online shop EU VAT compliant. Fast and easy.

Because of the B2C 2015 EU VAT changes you have to charge your clients the applicable VAT of their country.  
Our API automates this for you. So you can stick to your core business and still obey this particular law.

**It works perfectly! Try it yourself:**

I sell	To	VAT Rate
Digital Downloads	Austria	<b>20%</b>

**GET THIS API TODAY!** **SIGN UP NOW**

Bron: <https://vatinfo.eu>, initiatief van radish concepts en legal ICT