

Gegevensbanken - Benodigdheden werkje

March 22, 2018

1 Inleiding

In dit document geven we kort uitleg bij de tools die je nodig hebt om het werkje van dit vak te voltooien. Waar nodig wordt verwezen naar relevante websites waar de tools in kwestie te vinden zijn.

2 Installatie XAMPP

Om queries te kunnen uitvoeren, is er natuurlijk in eerste instantie een gegevensbank nodig. De typische aanpak bestaat erin om een gegevensbankserver lokaal op te zetten, waarna die gegevensbank kan worden gebruikt om queries te beantwoorden. Een courante keuze hiervoor is XAMPP, cf. <https://www.apachefriends.org/download.html>, wat cross-platform is. **We spreken af dat we de laatste versie gebruiken, i.e. 7.2.3.**

2.1 Installatie

Voor Windows en Mac kan je simpelweg de installer downloaden en uitvoeren.

In linux moet je eerst nog de installer uitvoerbaar maken met het `chmod` commando. Pas daarna kan de installer daadwerkelijk uitgevoerd worden. Concreet zal je iets zeer gelijkaardig aan de volgende commando's moeten uitvoeren in de terminal:

```
sudo chmod +x xampp-linux-x64-7.2.3-0-installer.run
sudo ./xampp-linux-x64-7.2.3-0-installer.run
```

Cf. https://www.apachefriends.org/faq_linux.html voor meer gedetailleerde installatiehulp.

2.2 Instellingen aanpassen

Omdat onze gegevensbank nogal groot is, bestaat de kans dat we handmatig instellingen moeten aanpassen om de gegevensbank in te kunnen laden. Probeer eerst zonder de instellingen aan te passen. Als het niet lukt, kunnen de volgende stappen helpen.

In het `php.ini` bestand¹ van je XAMPP installatie verander je volgende parameters (de exacte waarden maken niet uit, gewoon *significant groter* dan de defaults):

```
memory_limit = 500M
post_max_size = 400M
upload_max_filesize = 400M
max_execution_time = 800
max_input_time = 800
```

En in het `config.default.php` bestand² verander je de volgende instelling:

```
$cfg['ExecTimeLimit']=0;
```

¹De precieze locatie is afhankelijk van je besturingssysteem, cf. <http://stackoverflow.com/questions/6185319/how-to-locate-the-php-ini-file-xampp>, of google zelf, e.g.: 'Where to find php.ini file of XAMPP'

²Zelfde opmerking als bij het `php.ini` bestand

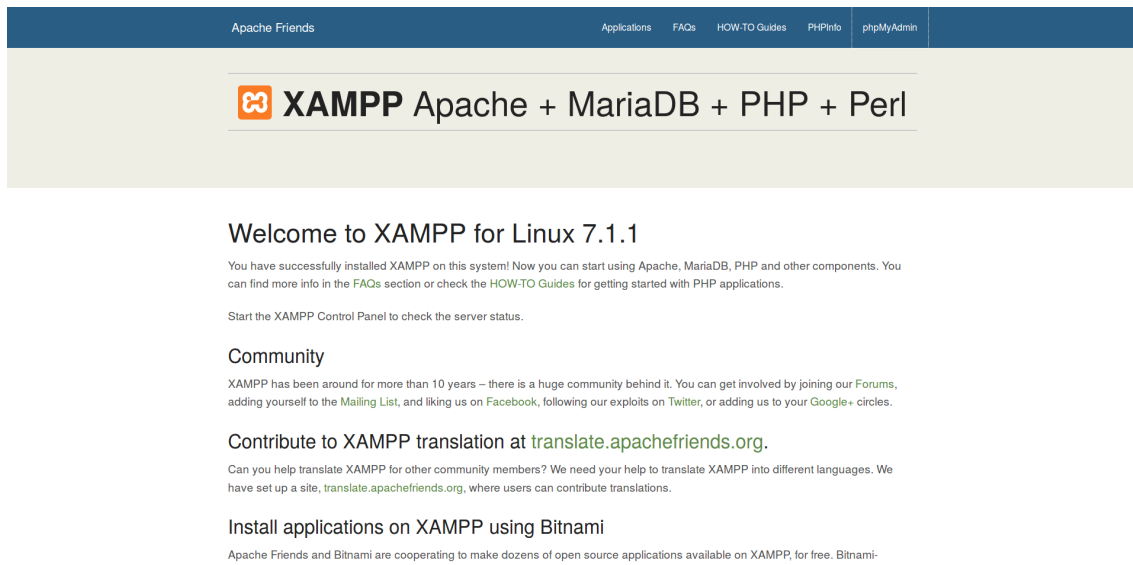


Figure 1: XAMPP startscherm.

Dan (her)start je de XAMPP services en ben je klaar om de gegevensbank te importeren. Op Windows en Mac wordt er doorgaans een icoontje aangemaakt waar je op kan klikken. Op linux moet je enkele terminalcommando's uitvoeren³, wederom cf. apachefriends.org/faq_linux.html.

3 Importeren van de database

Nu de server met correcte instellingen er is, rest ons nog het effectief opzetten van de gegevensbank die we voor dit werkje zullen nodig hebben. De eerste stap bestaat erin om de .sql files van de gegevensbank te downloaden. De sql-file die de database bevat is te vinden op: seanlahman.com/baseball-archive/statistics/. We werken op de 2016 versie.

Start de XAMPP services en open je browser. Daar surf je naar localhost. Je komt dan op een pagina terecht die er gelijkaardig moet uitzien aan figuur 1.

In de navigatiebalk bovenaan ga je vervolgens naar het PHPMYAdmin tabblad, cf. figuur 2. Maak nu eerst een nieuwe gegevensbank aan. In de linkerbalk staat er een knop **new** waarmee je dit kan doen.

Eens de gegevensbank is aangemaakt verschijnt ze aan de linkerkant. Klik op de naam van deze gegevensbank, en navigeer naar het **import** tabblad (bovenaan) en daar kan je dan het .sql bestand van onze gegevensbank importeren. Nadat dit geïmporteerd is, staat de gegevensbank op je lokale server.

Nu rest ons nog het aanmaken van een gebruiker die kan interageren met de gegevensbank. In het menu aan de linkerkant zie je nu de zopas geïmporteerde gegevensbank staan. Als je hierop klikt, krijg je de verschillende tabellen waaruit een gegevensbank bestaat te zien, voor een voorbeeld cf. figuur 3. In de navigatiebalk bovenaan het scherm is er een tabblad **privileges**. Daar heb je dan de mogelijkheid om een nieuwe gebruiker aan te maken die bepaalde rechten heeft voor deze gegevensbank.

³Tip: om je leven in de toekomst makkelijker te maken kan je een alias aanmaken voor een terminalcommando. Voor meer info google: 'How to make alias for terminal command linux'.

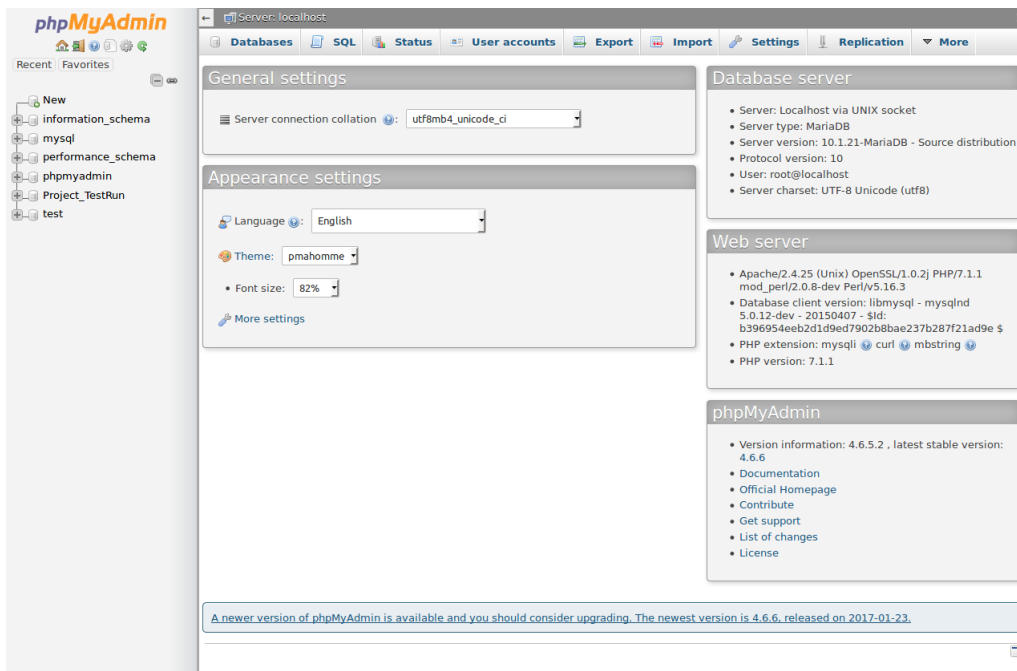


Figure 2: PHPMyAdmin startscherm.

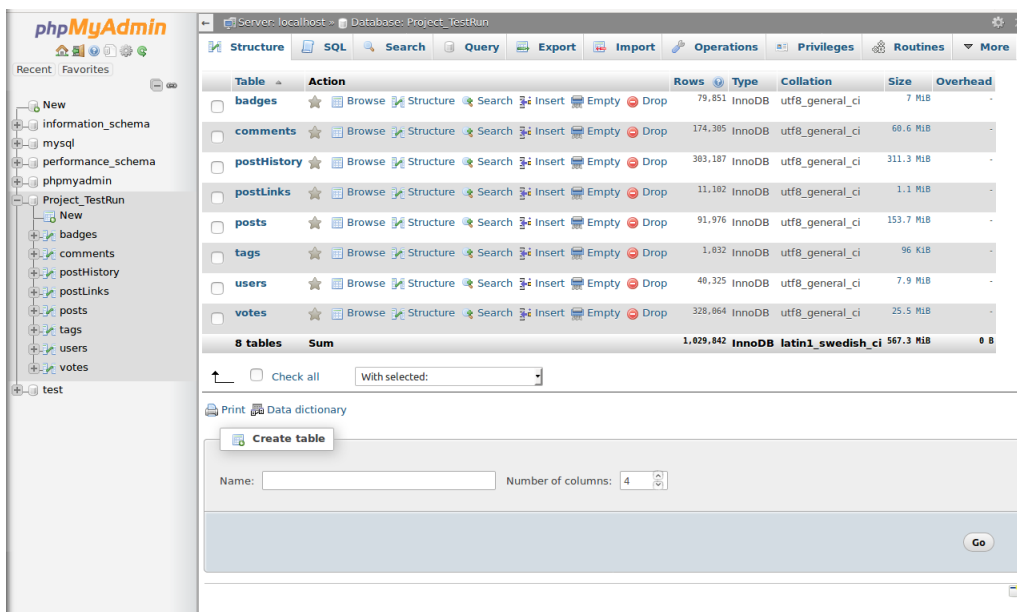


Figure 3: De tabellen in de gegevensbank. Het tabblad privileges bovenaan is waar je nieuwe users kan aanmaken.

4 Jupyter notebooks

4.1 Intro

Een handig platform (i.e. een zeer uitgebreide (python) package manager) om in te werken is anaconda. Dit platform wordt veel gebruikt in de datascience community en heeft enkele handige functionaliteiten die het leven iets makkelijker maken. Anaconda is ook cross-platform en kan gedownload worden via <https://www.continuum.io/downloads>. Jupyter notebook (zelf een collectie van python packages) zit standaard bij anaconda, dus daar moet je je verder niet om bekommeren. Meer gedetailleerde installatie-instructies vind je hier: <https://docs.continuum.io/anaconda/install>.

4.2 Extra packages installeren

Naast de standaard packages die bij anaconda zitten, hebben we enkele specifieke packages nodig om verbinding te maken met onze gegevensbank. Open de terminal/command prompt en voer de volgende commando's uit:

```
conda install sqlalchemy
conda install mysql
conda install mysql-connector-python
conda install pandas
```

Voor meer informatie over de mogelijkheden van deze typische conda commando's, kan je een kijkje nemen op de *cheat sheet*, cf. <https://conda.io/docs/using/cheatsheet.html>. Voor nog meer informatie over het anaconda platform kan je je natuurlijk altijd wenden tot de officiële documentatie, cf. <https://conda.io/docs/>.

4.3 Start jupyter notebook

Nu dat dat alles erop zit, kunnen we beginnen met het openen van Jupyter en het openen/aanmaken van een Jupyter notebook. Hiertoe volstaat het om een terminal/command prompt te openen en het volgende commando in te voeren:

```
jupyter notebook
```

In je browser wordt dan een nieuw tabblad geopend waar je de bestanden van je huidige map ziet staan. Als je dan navigeert naar een `.ipynb` bestand, wordt het geopend, en kan je beginnen met programmeren in een notebook omgeving.

Voor een korte tutorial i.v.m. jupyter notebooks, kan je (bijvoorbeeld) hier terecht: jupyter-notebook-beginner-guide.readthedocs.io/en/latest/.