



Odisee  
DE CO-HOGESCHOOL



# Big Data - Afspraken

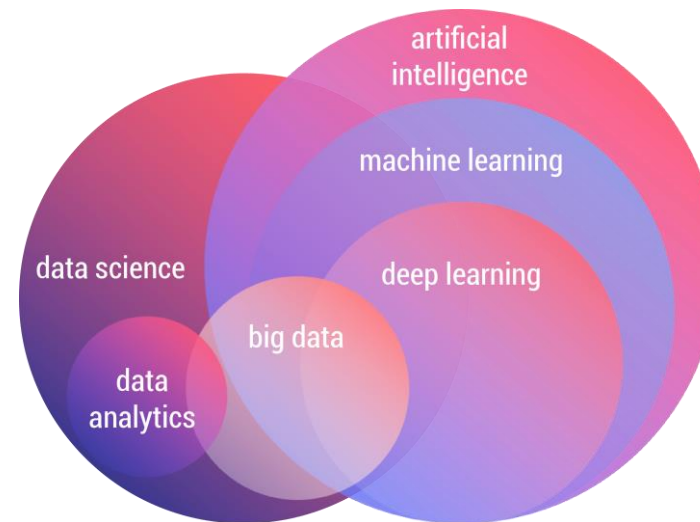
Jens Baetens

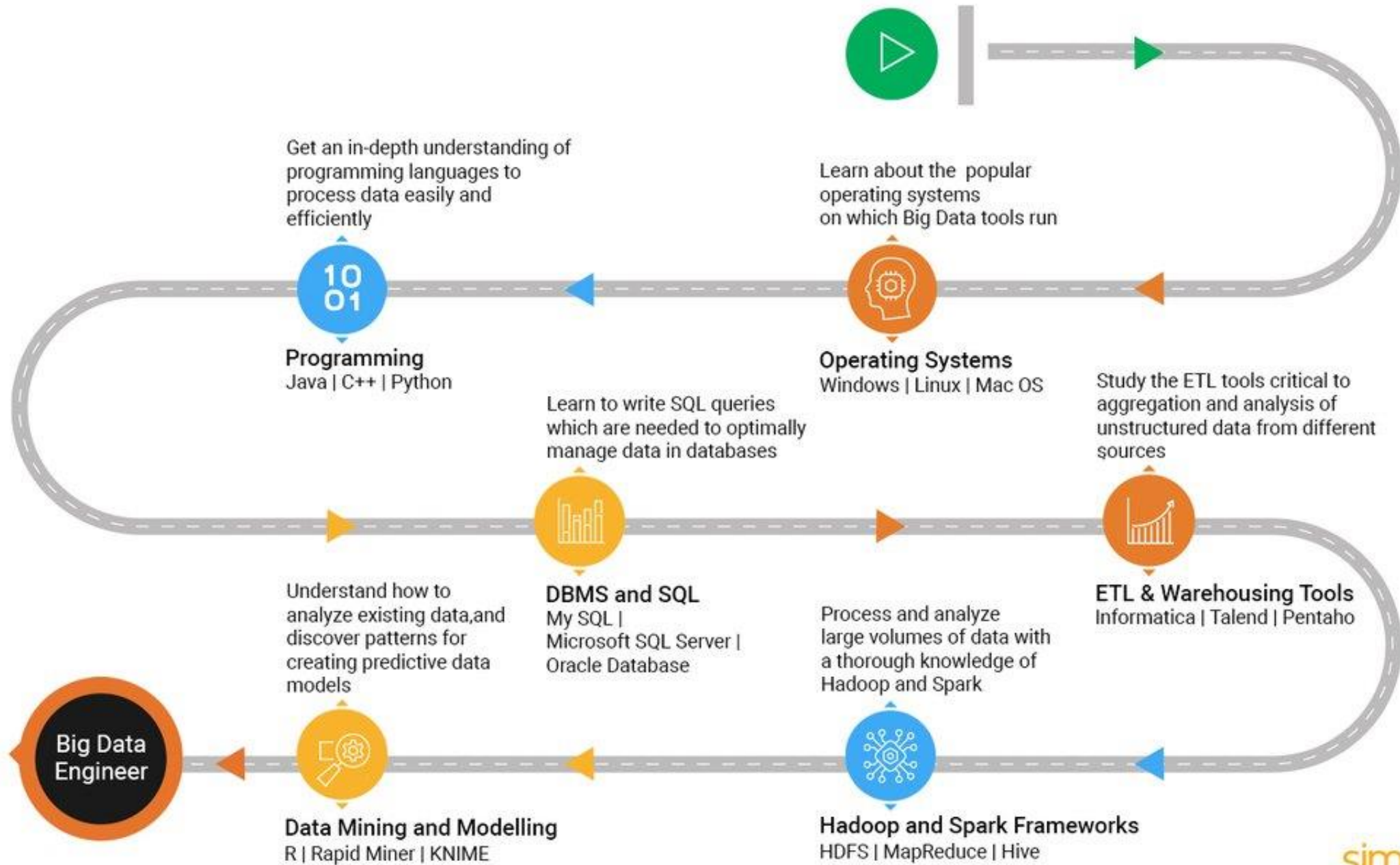


# Verloop en inhoud van het vak

## Vakken keuzerichting

- ▣ Data Science – 5 studiepunten
- ▣ Big Data – 5 studiepunten
- ▣ Machine Learning – 6 studiepunten







# Vacature Data Scientist KBC

## Data Scientist (NL/ENG)

KBC Bank & Verzekering · Flemish Region, Belgium (Hybrid) 12 hours ago · 15 applicants

Together, we develop artificial intelligence and machine learning solutions that **transform our business**. We work on them using, among others, data platforms like HDFS, Spark and AWS.

### What do we expect from you?

- You develop AI/ML solutions that transform our business. In order to roll them out successfully, you work closely with your colleagues in IT.
- You stay up to date with data science and AI.
- You help to strengthen the foundations within the KBC Group in the areas of knowledge and technology.
- You network with potential academic and commercial partners to set up partnerships.
- You have experience with data analytics and know how to use the more common tools for statistical analysis and machine learning.
- You have insight into learning algorithms, data-processing strategies, machine learning and natural language processing and know how to implement them in a business context.
- You're more than able when it comes to Python and data science toolsets (pandas, scikit-learn or H2O). Any experience you have with any of the following will be a benefit:
  - Learning frameworks (TensorFlow, PyTorch or Keras)
  - Big data analyses via cloud computing (AWS or MS Azure)
  - Apache Spark, Hadoop, Kubernetes
  - Cloudera
  - Git.

es

applicants. [Try Premium for free](#)

+8 more



- ▣ Eigenschappen van Big Data
- ▣ Mogelijke vormen van data
- ▣ Distributed Filesystemen
- ▣ Distributed Computing
- ▣ Cloud platforms

# Verloop

- ▣ # lesblokken van 4 uur
  - ▬ online in het derde kwartaal (donderdag namiddag)
  - ▬ op campus in het vierde kwartaal (maandag voormiddag)
- ▣ Evaluatie op basis van
  - ▬ 4 oefeningen
  - ▬ Project





## Studiemateriaal

- ▣ Slides en voorbeeld code via github repository
- ▣ Opdrachten en project via Toledo / Github classroom
- ▣ Tip: Hou een goed overzicht bij van het Hadoop eco-systeem en waarvoor de verschillende geziene toepassingen gebruikt kunnen worden.





## Afspraken

- ▣ Wees op tijd
- ▣ Vragen buiten de lessen mag steeds via Teams of mail
- ▣ Actief meewerken in de les beste leermethode





# Evaluatie

## Evaluatie

- ▣ 4 oefeningen: samen 50%
- ▣ Project met mondelinge verdeling: 50%
  - In groepen van 2 studenten
  - Opgave volgt nog
  - Respecteer de tussentijdse deadlines
- ▣ Respecteer de deadlines, te laat = 0 voor die oefening
- ▣ Elkaar helpen mag maar oplossingen overnemen niet



# Tools

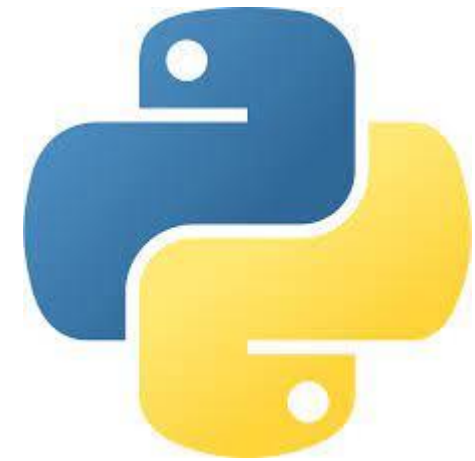
## Virtuele machine

- ▣ Virtuele machine kan gedownload worden via de link op Toledo



# Python

- ▣ Scripting programmeertaal
- ▣ Te installeren via academic software of anaconda
- ▣ Bevat een reeks handige packages



# Hadoop

- ▣ Distributed file system met daarboven op functionaliteiten voor distributed applicaties / computing /

## Core Hadoop Ecosystem

