

Odissee  
DE CO-HOGESCHOOL



# Big Data - afspraken

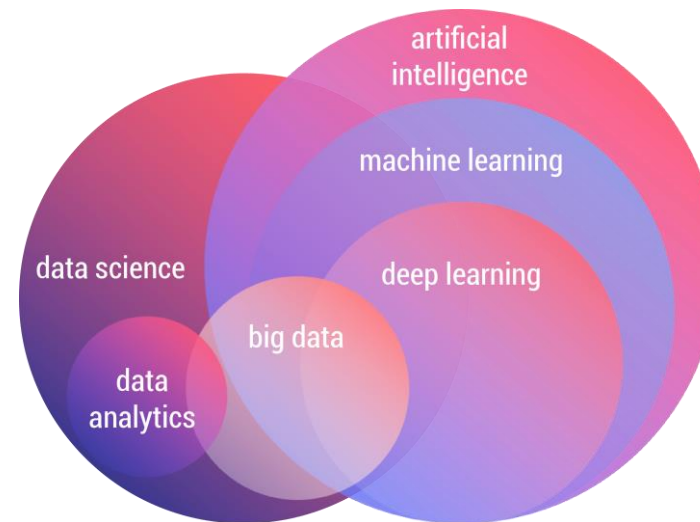
Jens Baetens

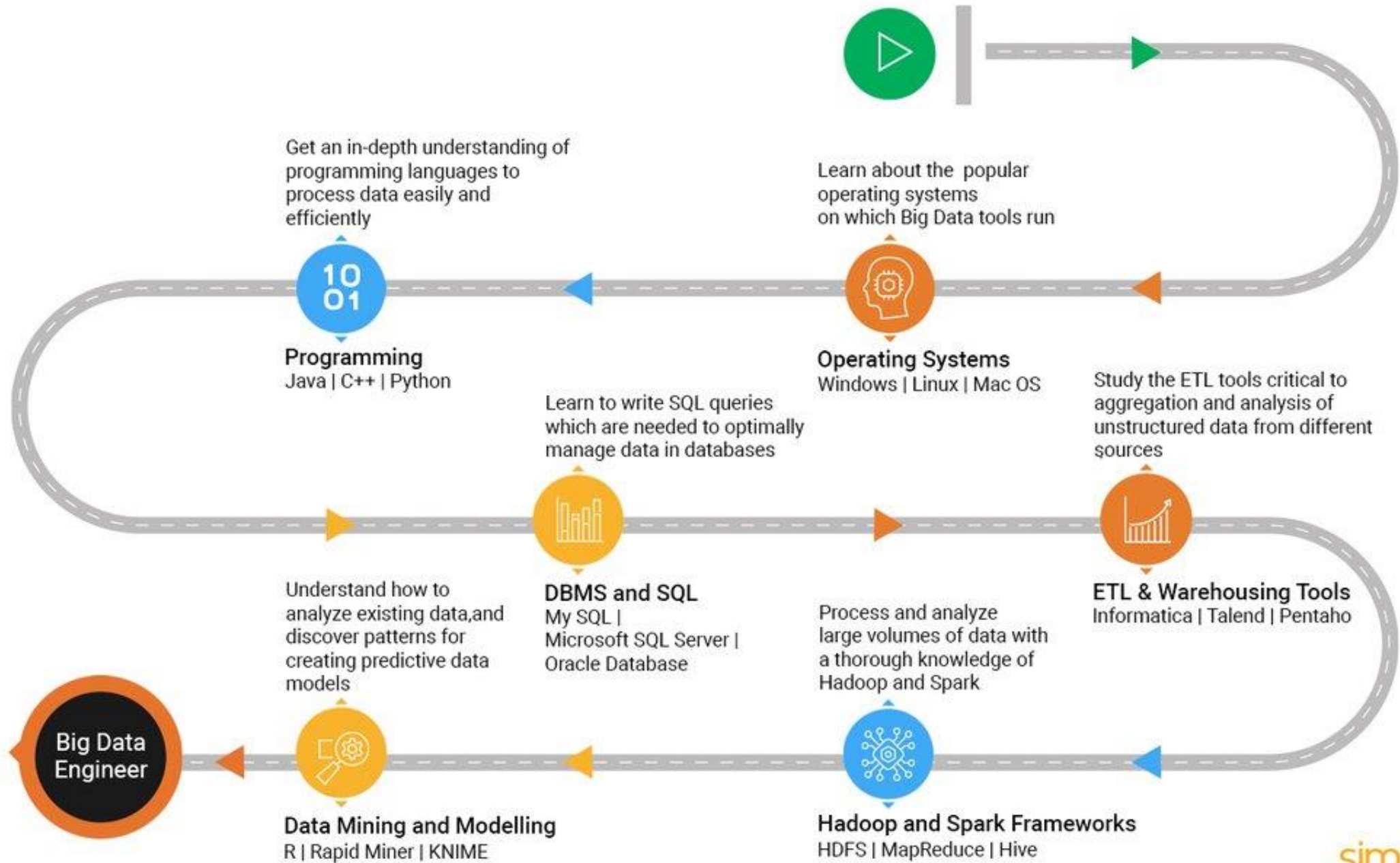


# Verloop en inhoud van het vak

## Vakken keuzerichting

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





- ▣ 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 voormiddag)
  - ▬ op campus in het vierde kwartaal (dinsdag voormiddag)
- ▣ Evaluatie op basis van
  - ▬ 4 oefeningen
  - ▬ Project



## Studiemateriaal

- ▣ Slides en voorbeeld code op Toledo en 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

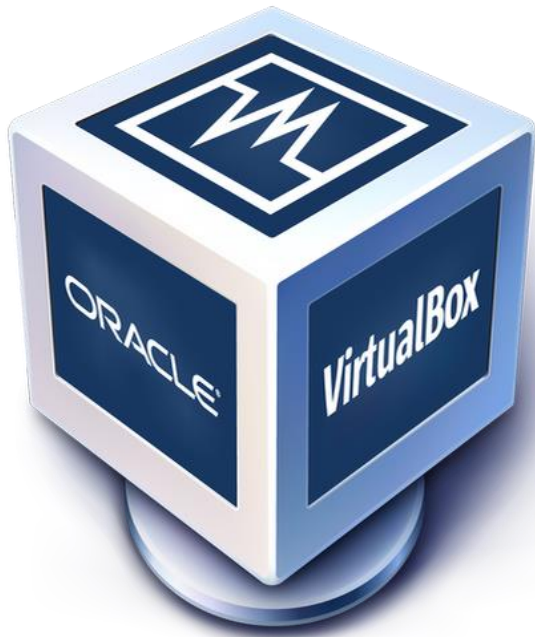
- ▣ 4 oefeningen: samen 50%
- ▣ Project met mondelinge verdeling: 50%
  - In groepen van 2 studenten
  - Opgave volgt nog
  - Respecteer de tussentijdse deadlines
- ▣ Respecteer de deadlines, te laat = 0 voor die oefening
- ▣ Elkaar helpen mag maar oplossingen overnemen niet



# Tools

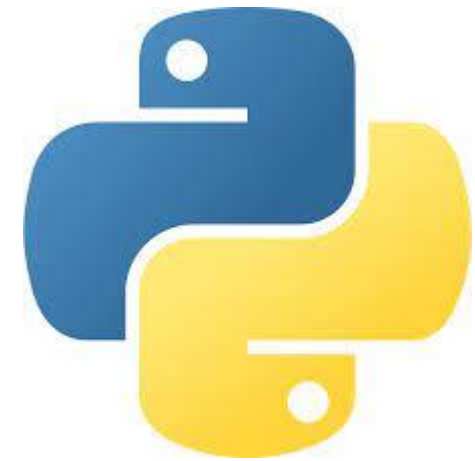
## Virtuele machine

- ▣ Virtuele machine kan gedownload worden via de bestanden van Teams voor dit vak



# 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

