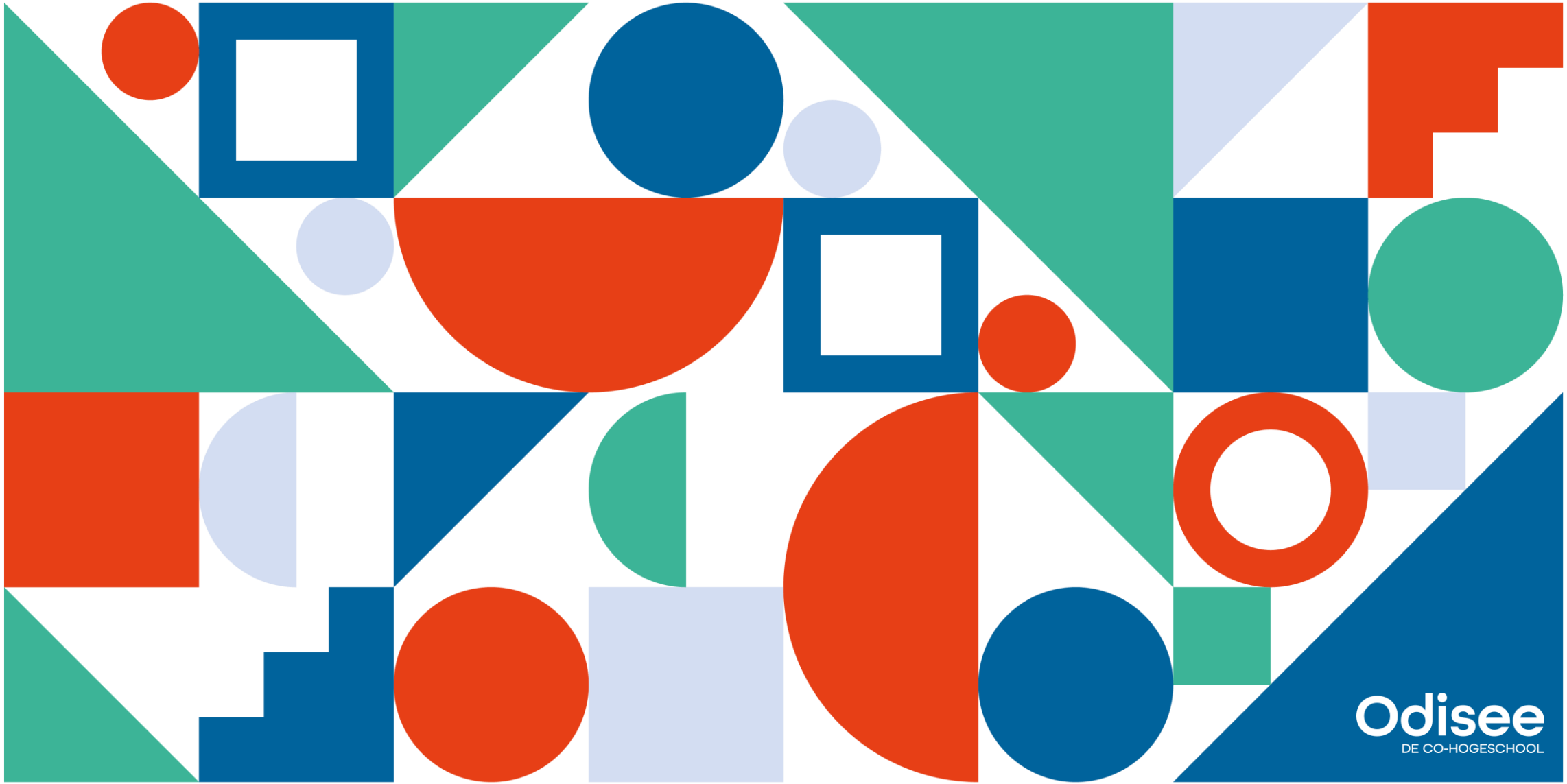

Odisee
DE CO-HOGESCHOOL

Big Data - Introductie

Jens Baetens

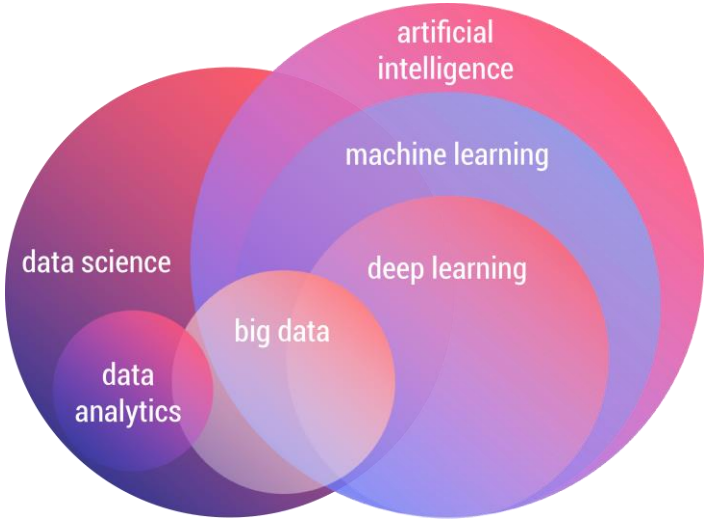
Structuur

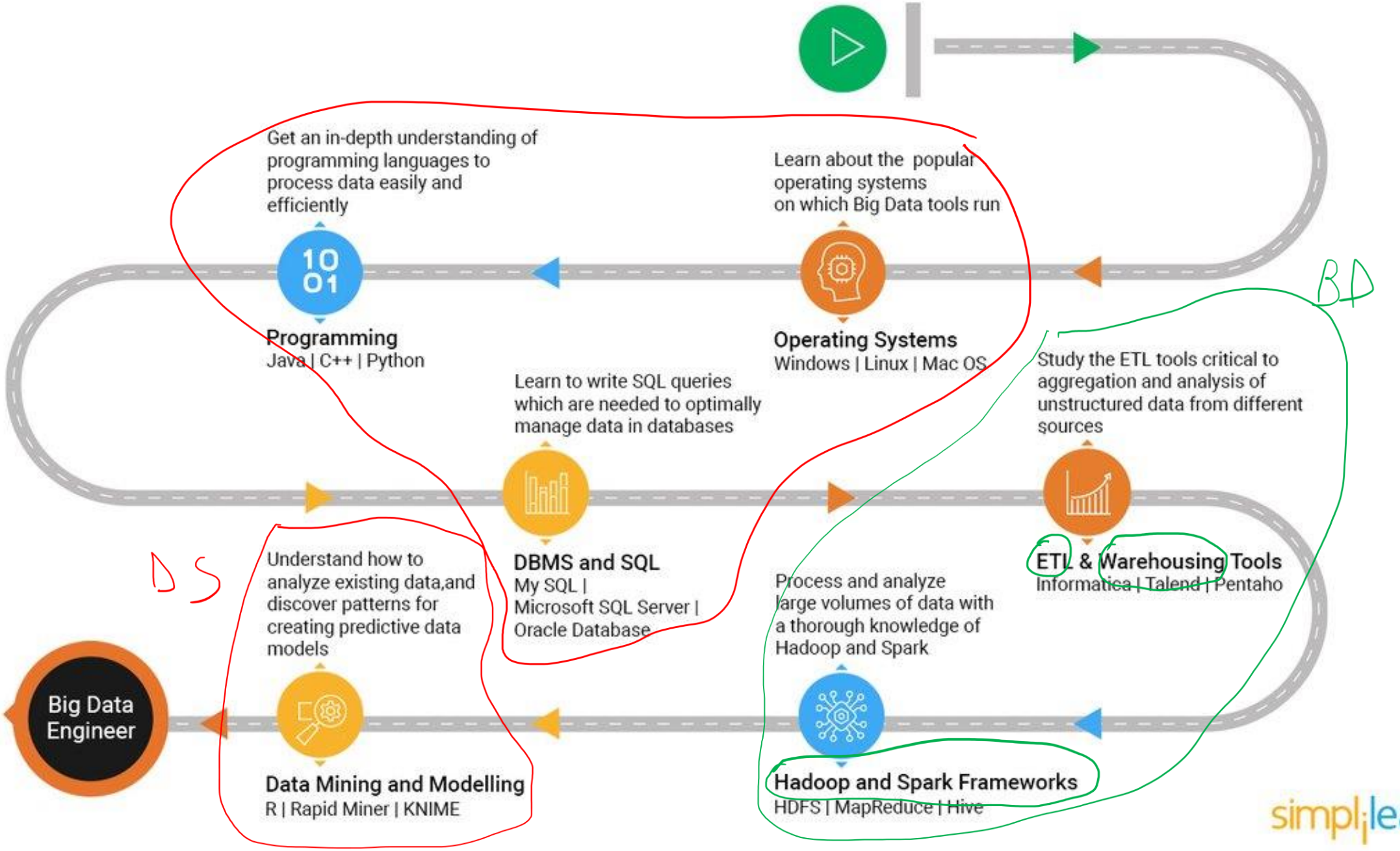
- Structuur van het vak
- Evaluatie
- Recap – Data Science
- Kenmerken van Big Data
- Distributed Storage
- Distributed Computing
- Tools

Structuur van het vak

Vakken keuzerichting

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




Get an in-depth understanding of programming languages to process data easily and efficiently
Programming
Java | C++ | Python
Learn about the popular operating systems on which Big Data tools run
Operating Systems
Windows | Linux | Mac OS
Learn to write SQL queries which are needed to optimally manage data in databases
DBMS and SQL
My SQL | Microsoft SQL Server | Oracle Database
Study the ETL tools critical to aggregation and analysis of unstructured data from different sources
ETL & Warehousing Tools
Informatica | Talend | Pentaho
Understand how to analyze existing data,and discover patterns for creating predictive data models
Data Mining and Modelling
R | Rapid Miner | KNIME
Process and analyze large volumes of data with a thorough knowledge of Hadoop and Spark
Hadoop and Spark Frameworks
HDFS | MapReduce | Hive
Big Data Engineer
DS
BD
simplilearn

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.

es

plicants. Try Premium for free

+8 more

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 of MS Azure)
  - Apache Spark, Hadoop, Kubernetes
  - Cloudera
  - Git.

ECTS

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

Verloop

- # lesblokken van 4 uur
  - online in het derde kwartaal
  - op campus in het vierde kwartaal
- Evaluatie op basis van
  - 2 Type A evaluaties (40% per, theoretisch en praktisch gedeelte)
  - 4 Type B evaluaties (5% per)

Studiemateriaal

- Slides en voorbeeld code via github repository
- Opdrachten 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

Tools

Containers

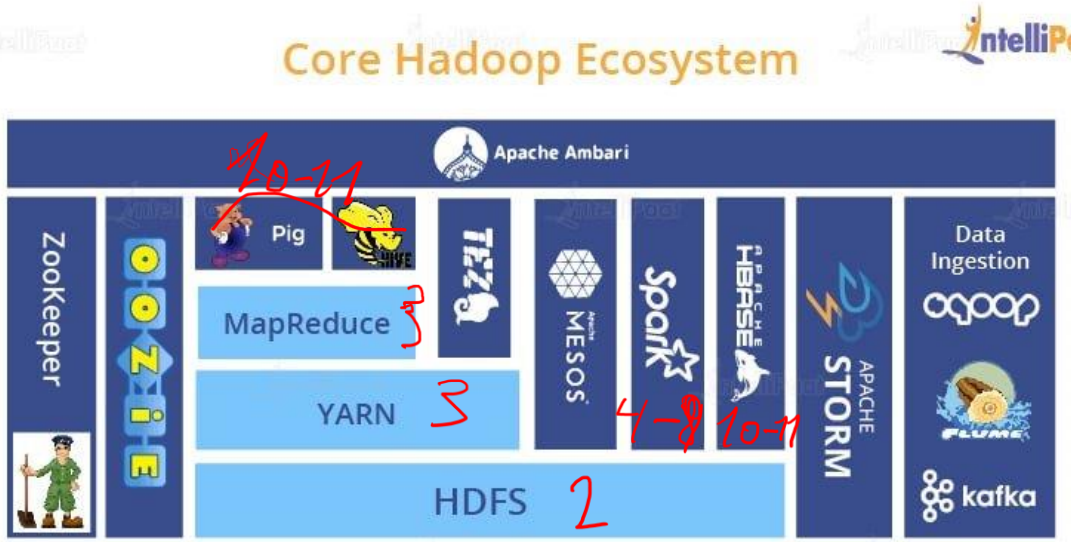
- Docker desktop – linux containers
  - Zie installatie.pdf op Toledo om de containers voor dit vak te gebruiken
  - Cluster met 4 datanodes

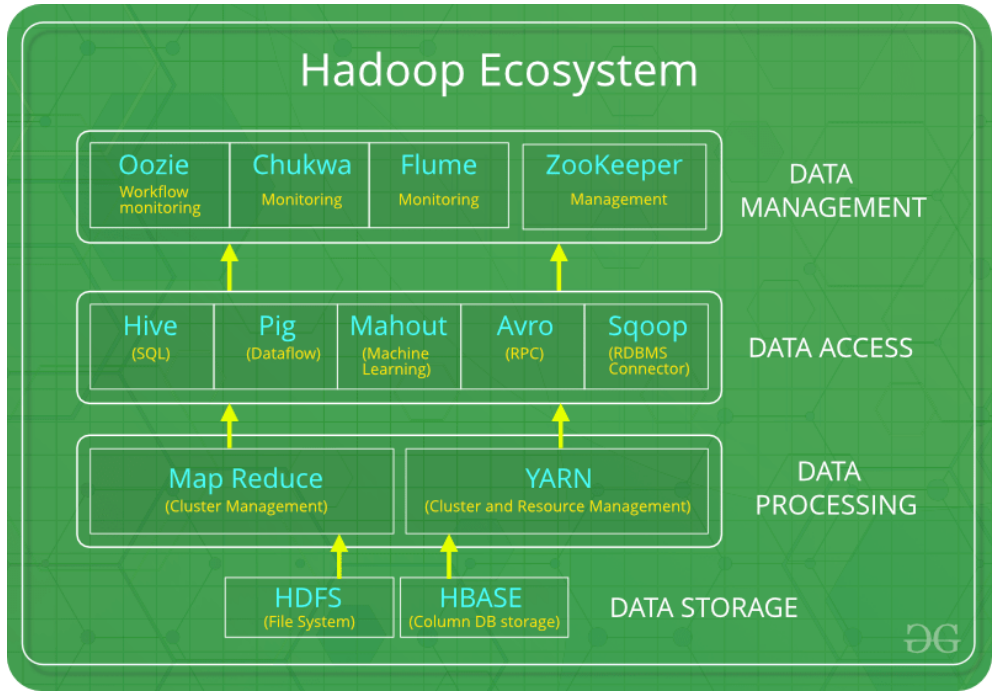
Python

- 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 /





NoSQL databases

- Allerlei implementaties maar vooral MongoDB heel populair