# Moduł 7
# Narzędzia do analizy Big Data, część 3:
## Narzędzia do analizy Big Data I

iBigWorld:
Innovations for Big Data in a Real World

ULSIT team



iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Cele nauczania

Wykład ma na celu ukazanie różnorodności technologii analizy Big Data i ich segmentacji, a także rozważenie niektórych popularnych rozwiązań programistycznych, ich funkcji, zalet, ograniczeń i możliwości szkoleniowe.



Co-funded by the
Erasmus+ Programme
of the European Union

# Struktura wykładu

- Przegląd narzędzi do analizy Big Data
- Popularne oprogramowanie do analizy danych
  - Cloudera
  - Oracle Analytics Cloud
  - SAP HANA
  - The Alteryx platform
  - SAS Viya
  - Apache Spark
  - Inne produkty

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the
Erasmus+ Programme
of the European Union

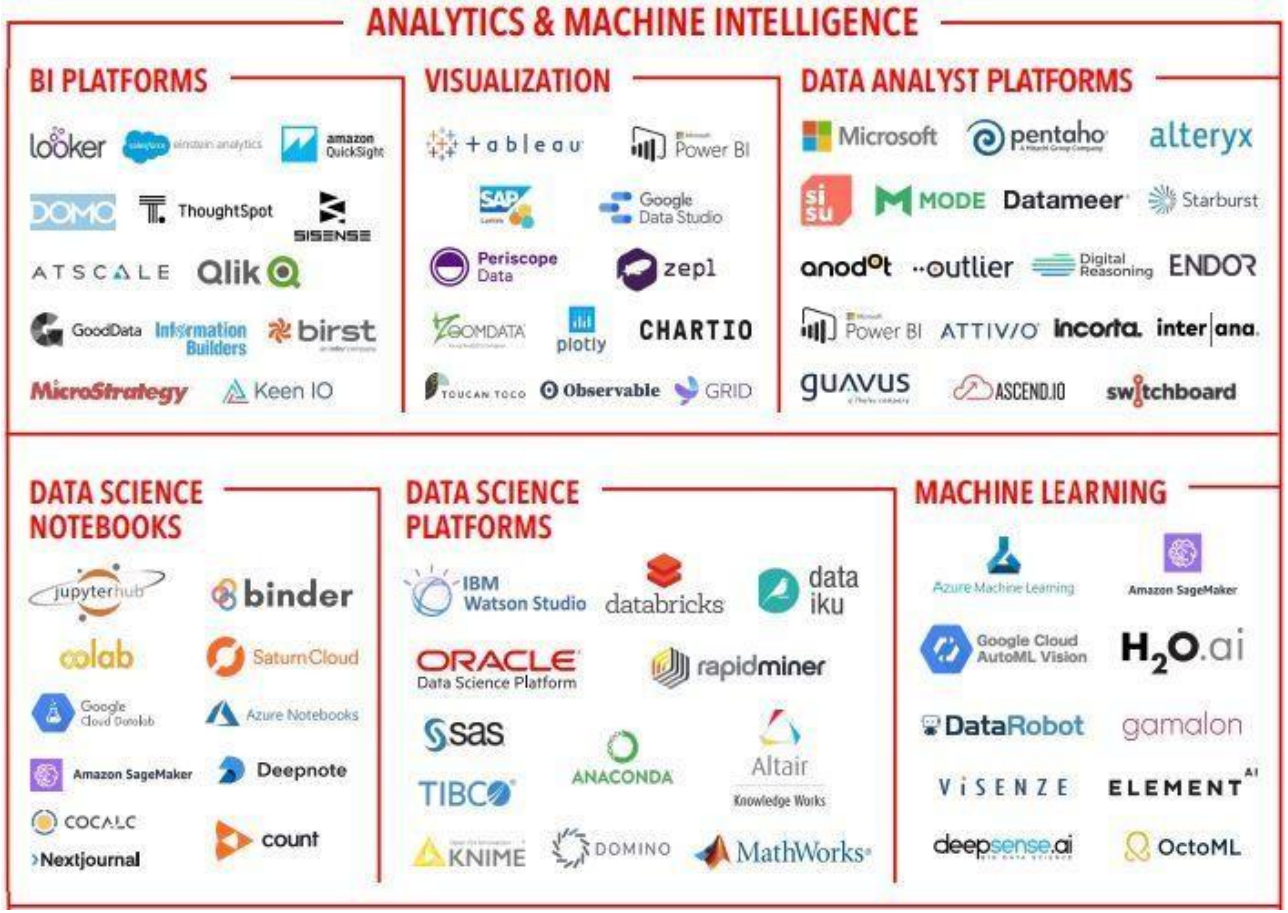
# Wyniki nauczania:

- Poznać główne kategorie technologii analizy Big Data
- Poznać cel, funkcjonalność, zalety i ograniczenia oprogramowania, które zapewnia analitykę danych
- Poznać ogólne zalety narzędzi do analizy Big Data



Co-funded by the Erasmus+ Programme of the European Union

# Dziedzina analityki i inteligencji maszynowej dla Big Data




iBigWorld:2020-1-PL01-KA203-082197

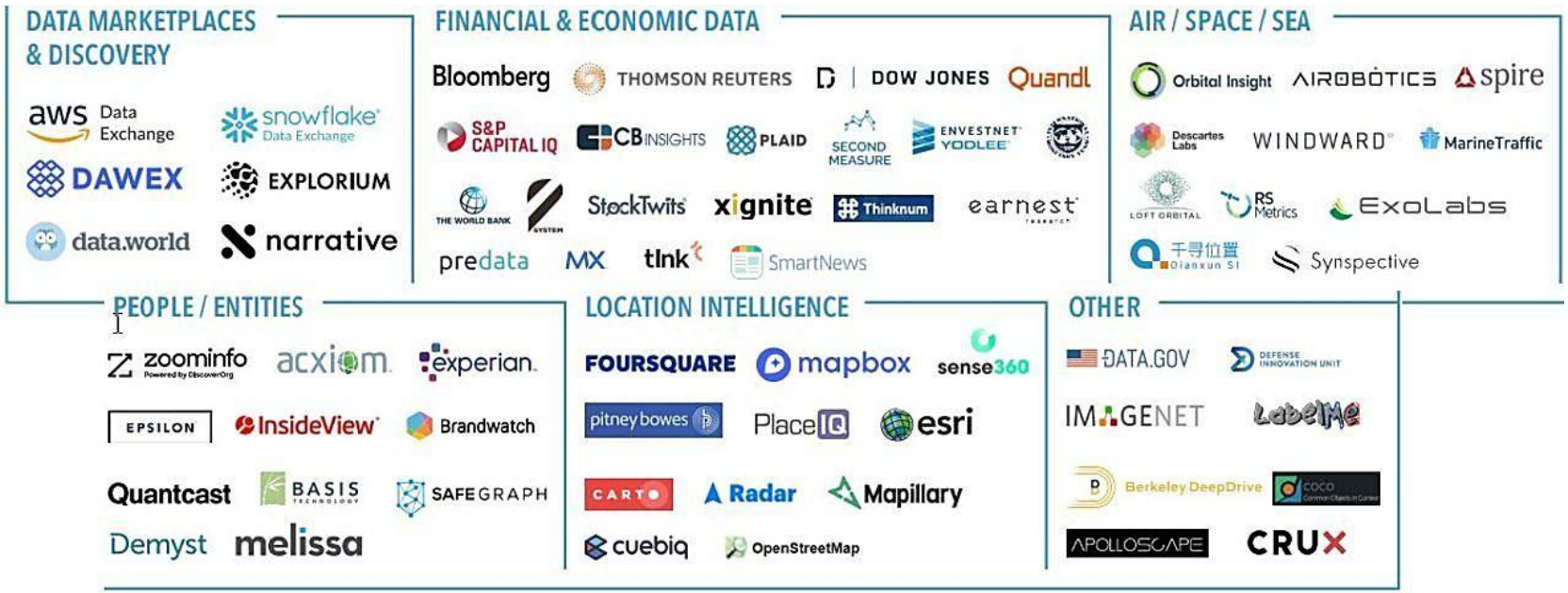
Co-funded by the Erasmus+ Programme of the European Union

# Dziedzina analityki i inteligencji maszynowej dla Big Data


iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Obszar analityki Big Data: źródła danych i aplikacje




iBigWorld:2020-1-PL01-KA203-082197

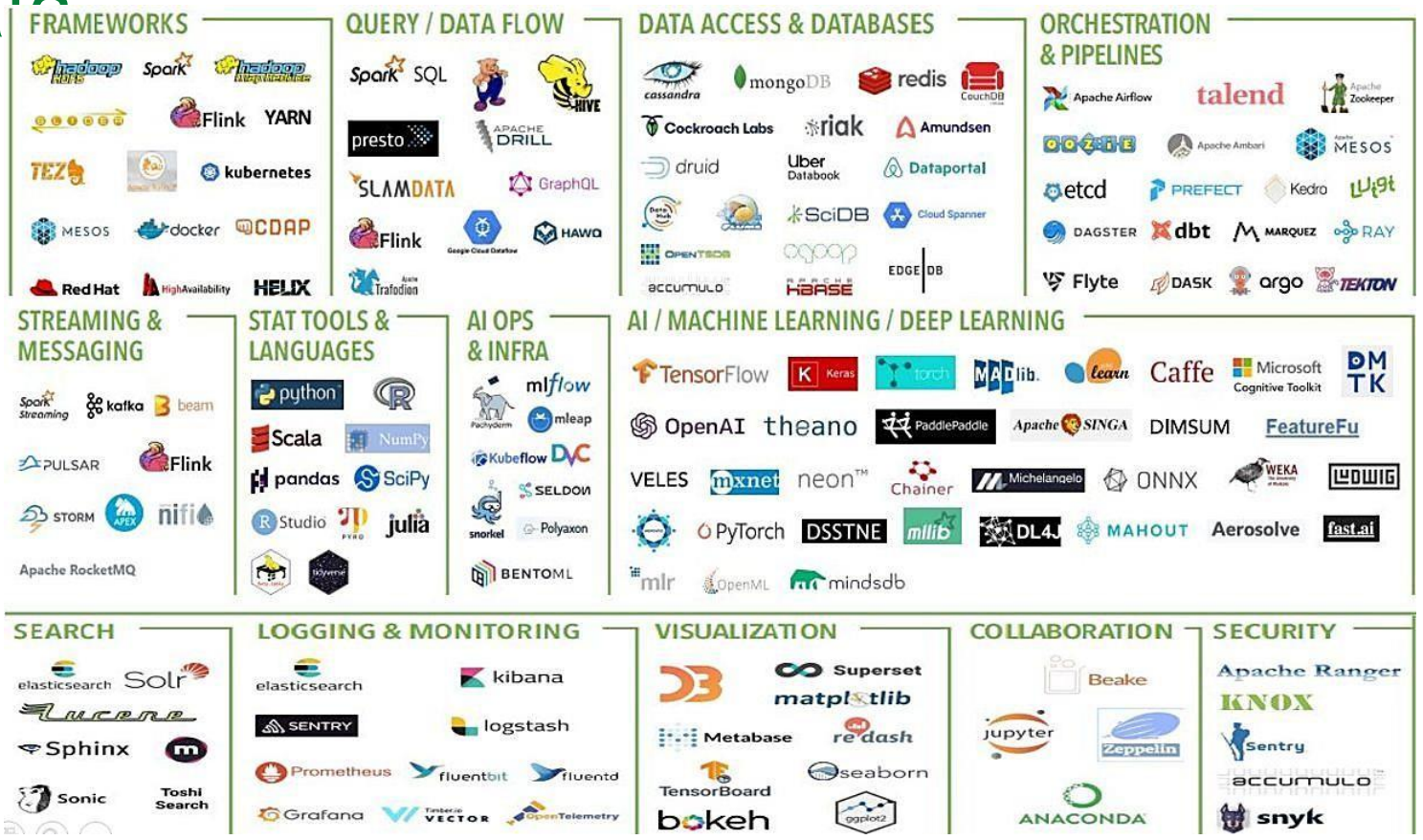
Co-funded by the Erasmus+ Programme of the European Union

# Pole analizy Big Data: Zasoby


iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

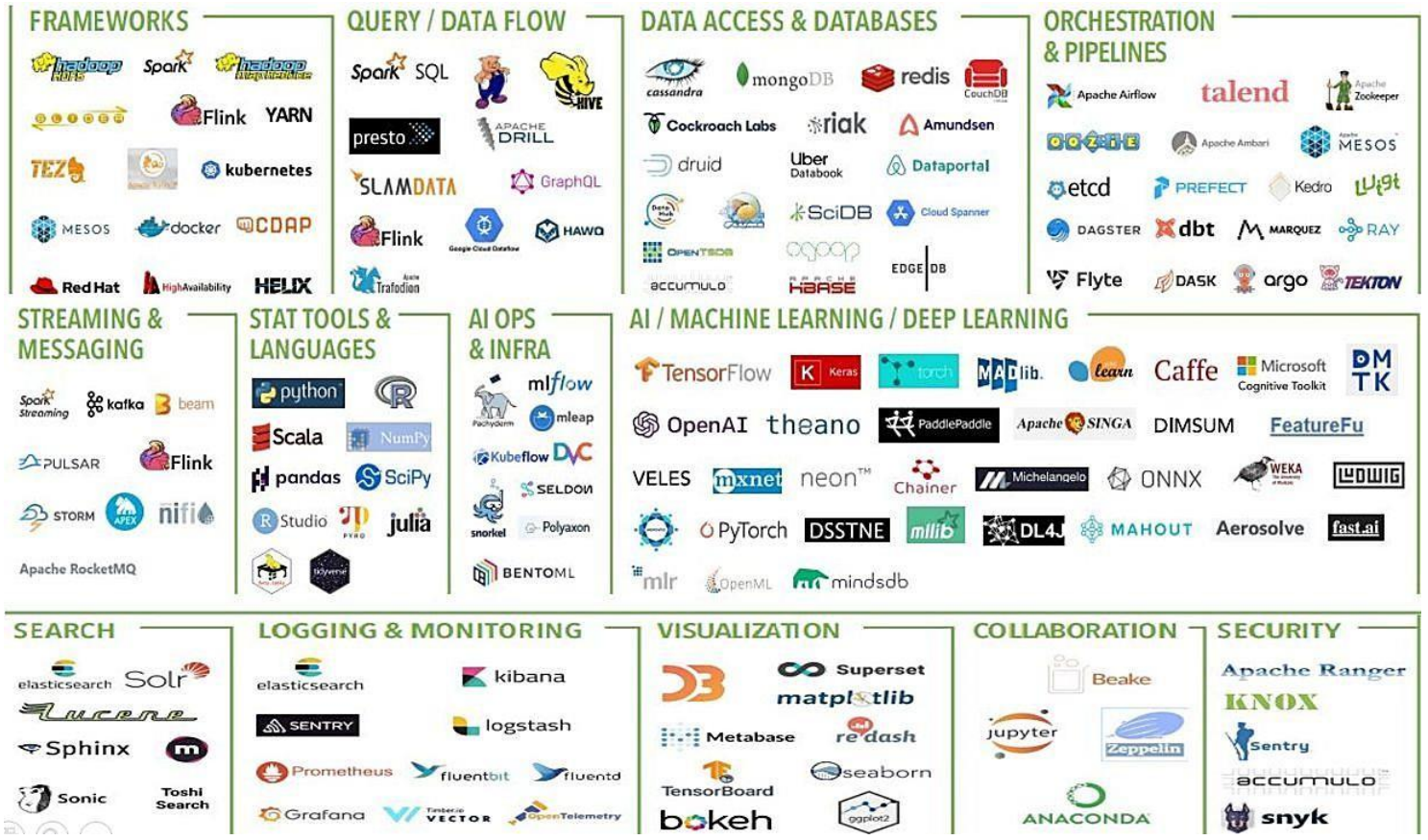
# Dziedzina analityki Big Data: Otwarte zasoby przetwarzania Big Data




iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Dziedzina analityki Big Data: Otwarte zasoby przetwarzania Big Data




iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do rozwiązywania problemów Big Data

- Cloudera
- Oracle Analytics Cloud
- SAP HANA
- The Alteryx platform
- SAS Viya
- Apache Spark
- i wiele innych

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do Big Data Cloudera & Oracle Analytics Cloud

**Cloudera** oferuje wszystko, czego potrzebujesz do interakcji z centrum danych przedsiębiorstwa, w tym oprogramowanie do zadań z zakresu krytycznych danych biznesowych, takich jak przechowywanie, dostęp, zarządzanie, analiza, bezpieczeństwo i wykrywanie.

**Oracle Analytics Cloud** oferuje szeroki zakres możliwości raportowania i analizy z chmury. Przygotowuje i analizuje dane w celu zidentyfikowania trendów, a następnie przekształca je w intuicyjne wizualizacje, które użytkownicy mogą eksplorować i udostępniać.

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do Big Data SAP HANA & Alteryx

**SAP HANA** to baza danych w pamięci dla platformy technologicznej SAP do wsparcia biznesowego w czasie rzeczywistym. Jest dostępne lokalnie, w chmurze i jako rozwiązanie hybrydowe do zaawansowanej analizy rzeczywistych danych transakcyjnych w celu wyświetlania praktycznych wniosków

**Platforma Alteryx** to pakiet 5 produktów oferujących kompleksowe wsparcie, od zbierania danych z głębokich pul danych po zautomatyzowane operacje przedsiębiorstwa, analizy finansowe i przemysłowe. Umożliwia tworzenie powtarzalnych przepływów pracy ETL, z językiem programowania lub bez niego.

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do Big Data SAS Viya & Apache Spark

**SAS Viya** to kompleksowy, samoobsługowy silnik oparty na chmurze, który łączy analitykę wizualną, statystyki i naukę o danych dla przedsiębiorstw. Wykorzystuje ustandaryzowaną bazę kodu z obsługą programowania w R, Python, SAS, Java i Lua.

**Apache Spark** to ujednolicone oprogramowanie analityczne typu open source do rozproszonego szybkiego przetwarzania. Dystrybuuje dane pomiędzy klastrami w czasie rzeczywistym. Obsługuje Python, R, Scala, SQL i Java i może działać samodzielnie lub łatwo integrować się z szerszymi przepływami pracy.

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do Big Data Apache Spark

**Zalety:**

- Bezpłatne i otwarte oprogramowanie
- Zaawansowane przetwarzanie
- Uniwersalne funkcjonowanie
- Łatwość użycia
- Tolerancja błędów

**Kluczowe cechy:**

- Tryb autonomiczny
- GraphX: obliczenia równoległe i budowa diagramów w systemie.
- Nauczanie maszynowe
- Rozproszone zbiory danych
- Przesyłanie strumieniowe danych
- Integracja

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do Big Data Apache Spark

**Ograniczenia:**

- Odpowiedzi działu obsługi klienta mogą być opóźnione
- Ciężkie szkolenie dla nowych użytkowników
- Niektóre funkcje czasami działają wolno

**Strona:**

www.apache.org

**Nauczanie:**

- Istnieje dokumentacja dla wszystkich wersji
- **Możesz zadawać pytania na forum**
- Forum StackOverflow Apache


iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie dla Big Data Cloudera

**Zalety:**

- Skalowalność
- Wysoka i łatwa rozszerzalność poprzez API dla rozszerzeń wtyczek
- Intuicyjny interfejs użytkownika
- Bezpłatne i otwarte oprogramowanie
- Ciągła integracja usług
- Komponenty wielokrotnego użytku

**Cechy:**

- Portal internetowy i serwer KNIME
- Obsługuje rozszerzenia
- Obsługuje integrację
- Przepływy pracy importu/eksportu
- Wykonywanie równoległe w systemach wielordzeniowych
- Wersja wiersza poleceń

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie do Big Data Cloudera

**Nauczanie:**

- Darmowa edukacja
- Kształcenie odpłatne - w klasach, w klasach wirtualnych, osobiście
- Trzydniowe szkolenie ról z instruktorem – na całym świecie

**Strona WWW:**

www.cloudera.com

cloudera®

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Popularne oprogramowanie dla BigData

- **Apache Hadoop** to platforma typu open source do pracy z dużymi ilościami danych.
- **Apache Pig** to platforma open source, która jest idealna do analizowania dużych zbiorów danych i prezentowania ich w strumieniach danych.
- Platforma **RapidMiner** to oparta na chmurze seria ofert analizy danych, która obsługuje wszystkie poziomy ekosystemu Big Data.

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

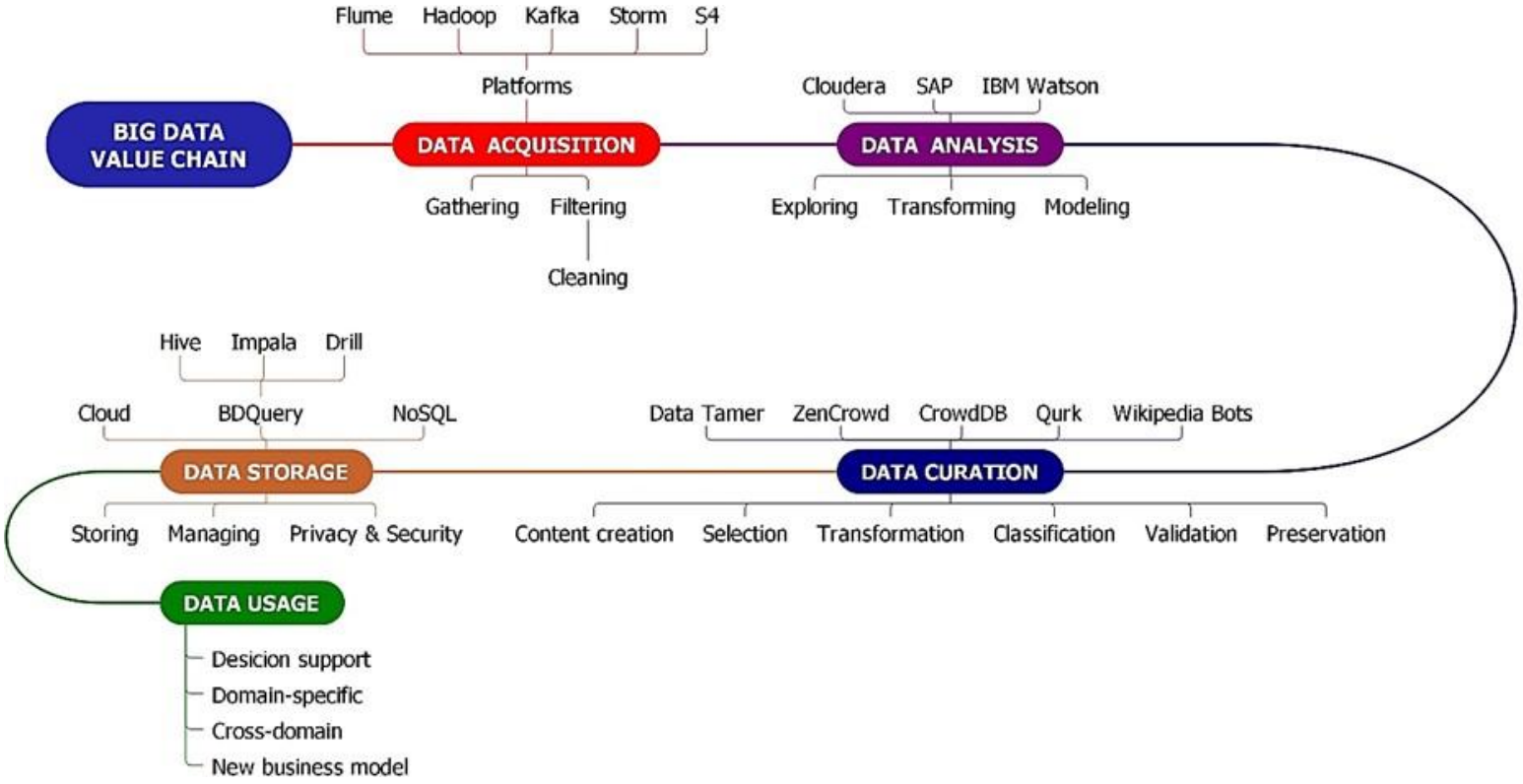
# Popularne oprogramowanie dla BigData

- **MATLAB** to numeryczna platforma obliczeniowa i programistyczna, która pozwala użytkownikom opracowywać i wdrażać algorytmy matematyczne, tworzyć modele i analizować dane.
- **R** to język programowania i środowisko programowania typu open source do analiz statystycznych, tworzenia wykresów i raportowania.
- **Python** to wysokopoziomowy, szybki, wydajny, przenośny, prosty, open-source, język programowania ogólnego przeznaczenia, który obsługuje inne technologie.

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Procesy i decyzje w planowaniu strategicznym Big Data (Big Data Value Chain)




iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Literatura


- Matt Turck, https://mattturck.com/data2020.
- Gartner Peer Insights, https://www.gartner.com/reviews/home.
- G2: Business Software and Services Reviews, https://www.g2.com.
- Select Hub, https://www.selecthub.com.
- TrustRadius, https://www.trustradius.com.
- Capterra, https://www.capterra.com.
- FinancesOnline, https://financesonline.com.
- Software Advice, https://www.softwareadvice.com.




Co-funded by the Erasmus+ Programme of the European Union

# Pytania?

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the
Erasmus+ Programme
of the European Union