

# Practices of Business Intelligence

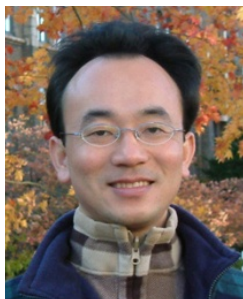
## 人工智慧、大數據與雲端運算

### (ABC: AI, Big Data, and Cloud Computing)

1071BI03

MI4 (M2084) (2888)

Wed, 7, 8 (14:10-16:00) (B217)



Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

<http://mail.tku.edu.tw/myday/>

2018-09-26



# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)   |
|-----------|------------|---|
| 1         | 2018/09/12 | 商業智慧實務課程介紹<br>(Course Orientation for Practices of Business Intelligence)                                   |
| 2         | 2018/09/19 | 商業智慧、分析與資料科學<br>(Business Intelligence, Analytics, and Data Science)  |
| 3         | 2018/09/26 | 人工智慧、大數據與雲端運算<br>(ABC: AI, Big Data, and Cloud Computing)   |
| 4         | 2018/10/03 | 描述性分析I：數據的性質、統計模型與可視化<br>(Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization) |
| 5         | 2018/10/10 | 國慶紀念日 (放假一天) (National Day) (Day off)   |
| 6         | 2018/10/17 | 描述性分析II：商業智慧與資料倉儲<br>(Descriptive Analytics II: Business Intelligence and Data Warehousing)                 |

# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)  |
|-----------|------------|--|
| 7         | 2018/10/24 | 預測性分析I：資料探勘流程、方法與演算法<br>(Predictive Analytics I: Data Mining Process, Methods, and Algorithms) |
| 8         | 2018/10/31 | 預測性分析II：文本、網路與社群媒體分析<br>(Predictive Analytics II: Text, Web, and Social Media Analytics)       |
| 9         | 2018/11/07 | 期中報告 (Midterm Project Report)  |
| 10        | 2018/11/14 | 期中考試 (Midterm Exam)  |
| 11        | 2018/11/21 | 處方性分析：最佳化與模擬<br>(Prescriptive Analytics: Optimization and Simulation)                          |
| 12        | 2018/11/28 | 社會網絡分析<br>(Social Network Analysis)  |

# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)  |
|-----------|------------|--|
| 13        | 2018/12/05 | 機器學習與深度學習<br>(Machine Learning and Deep Learning)  |
| 14        | 2018/12/12 | 自然語言處理<br>(Natural Language Processing)  |
| 15        | 2018/12/19 | AI交談機器人與對話式商務<br>(AI Chatbots and Conversational Commerce)                               |
| 16        | 2018/12/26 | 商業分析的未來趨勢、隱私與管理考量<br>(Future Trends, Privacy and Managerial Considerations in Analytics) |
| 17        | 2019/01/02 | 期末報告 (Final Project Presentation)  |
| 18        | 2019/01/09 | 期末考試 (Final Exam)  |

# Business Intelligence (BI)

①

Introduction to BI and Data Science

2

Descriptive Analytics

3

Predictive Analytics

4

Prescriptive Analytics

5

Big Data Analytics

6

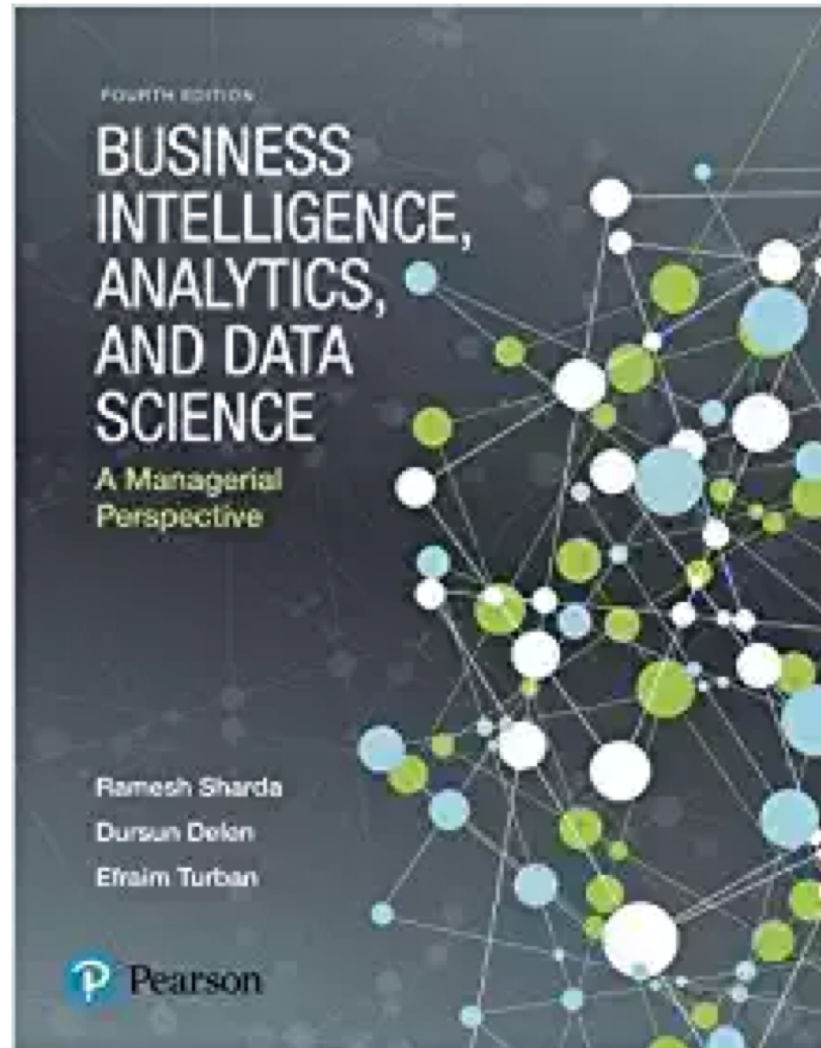
Future Trends

**ABC:  
AI,  
Big Data,  
Cloud Computing**

# Outline

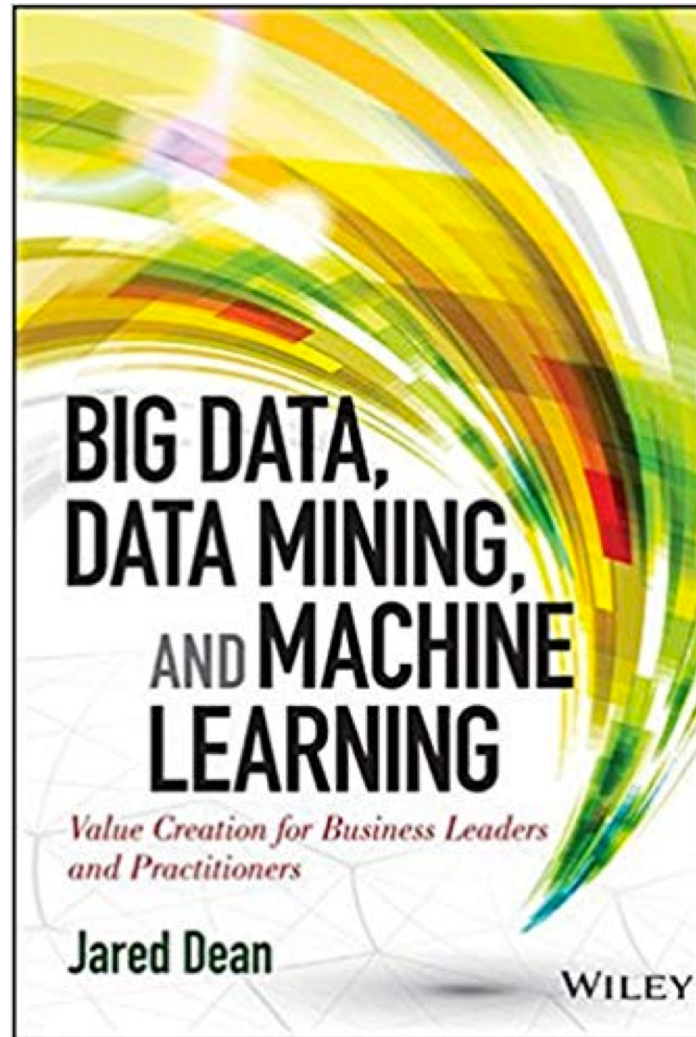
- AI
- Big Data
- Cloud Computing

**Business Intelligence, Analytics, and Data Science:  
A Managerial Perspective, 4th Edition,  
Ramesh Sharda, Dursun Delen, and Efraim Turban,  
Pearson, 2017.**

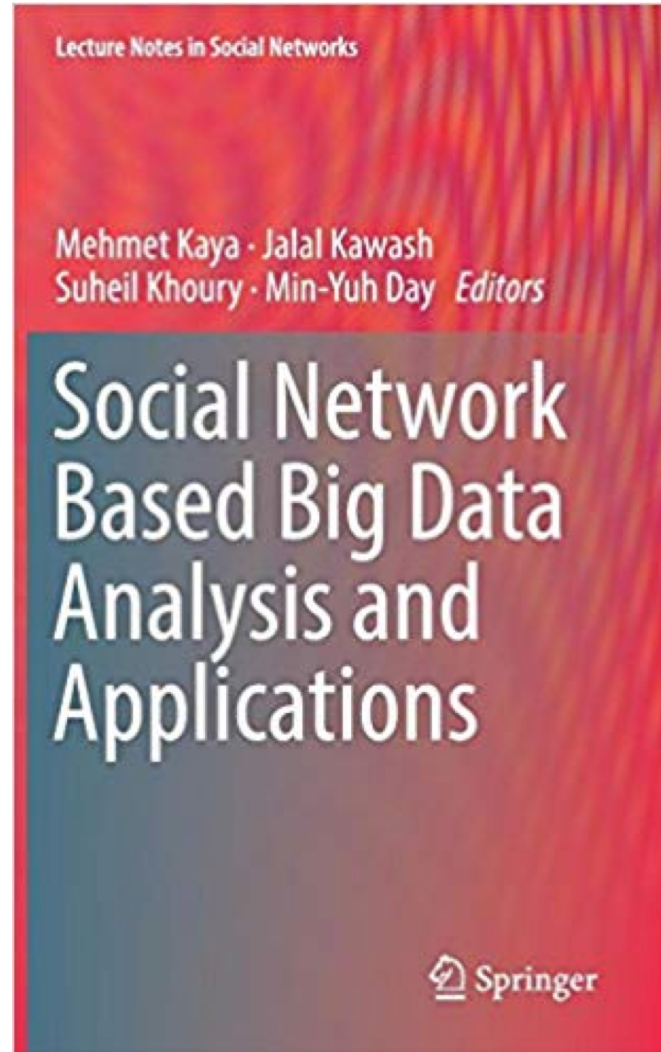




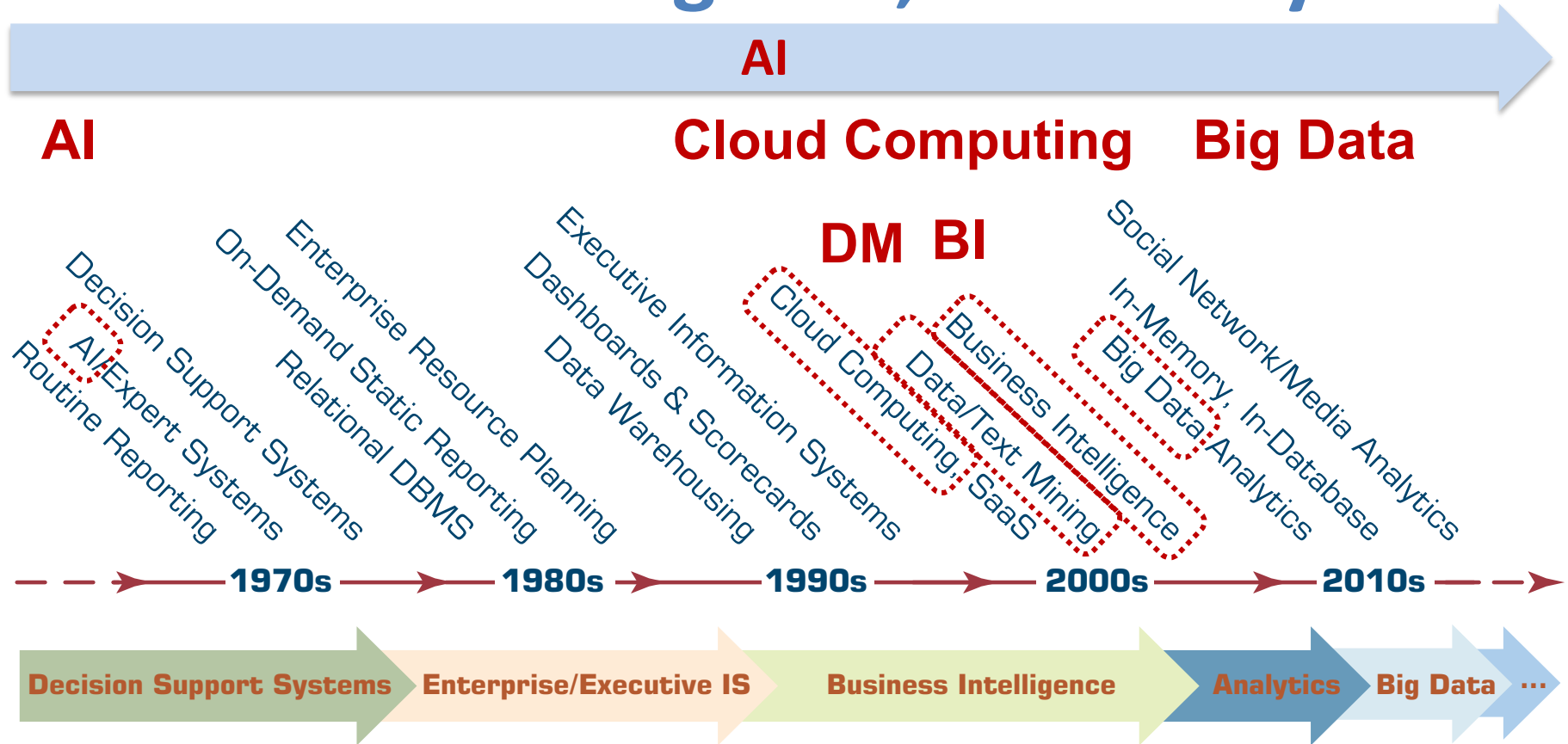
**Big Data, Data Mining, and Machine Learning: Value Creation for  
Business Leaders and Practitioners,  
Jared Dean,  
Wiley, 2014.**



**Social Network Based Big Data Analysis and Applications,  
Lecture Notes in Social Networks,  
Mehmet Kaya, Jalal Kawash, Suheil Khoury, Min-Yuh Day,  
Springer International Publishing, 2018.**

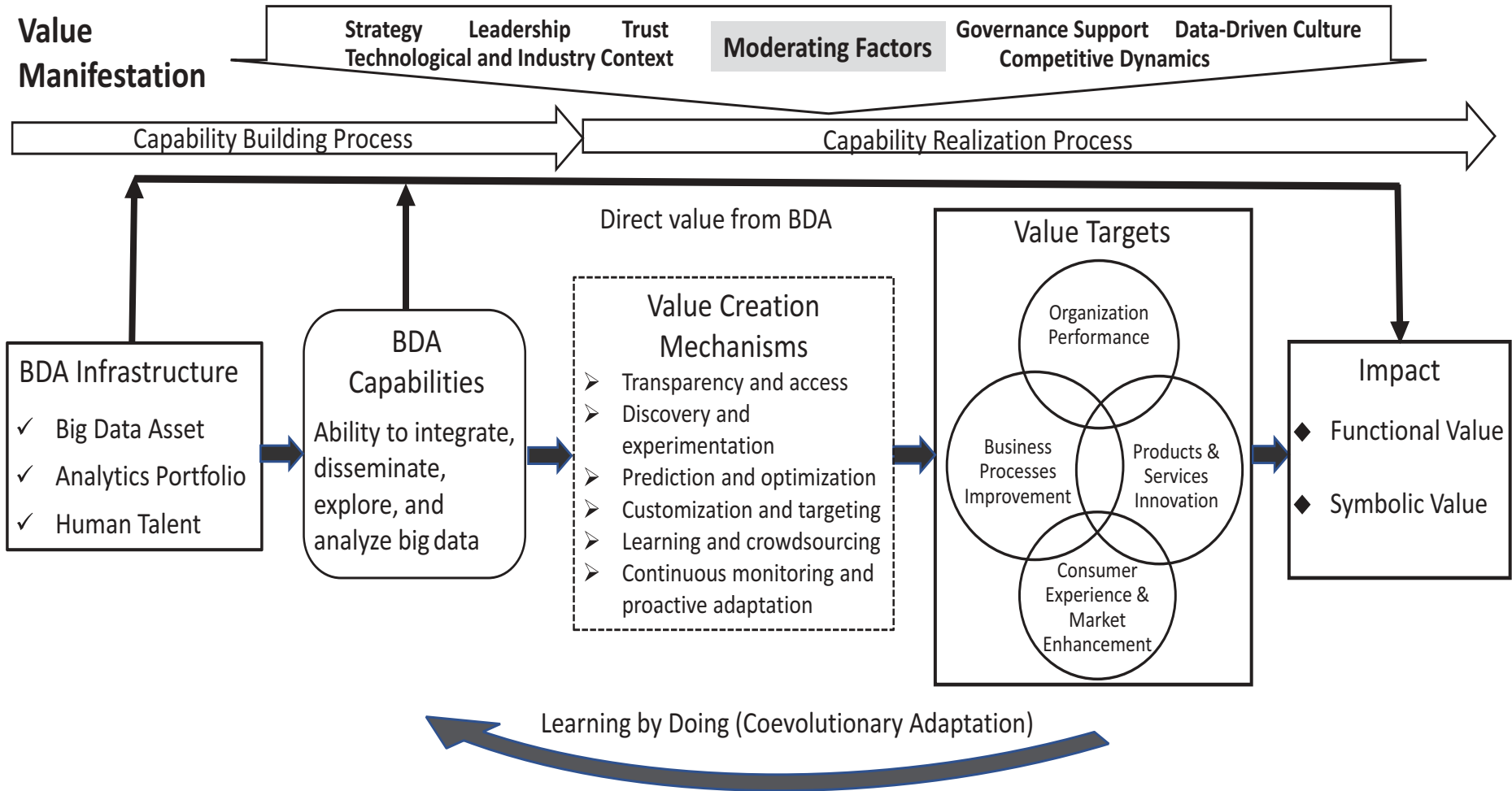


# AI, Big Data, Cloud Computing Evolution of Decision Support, Business Intelligence, and Analytics



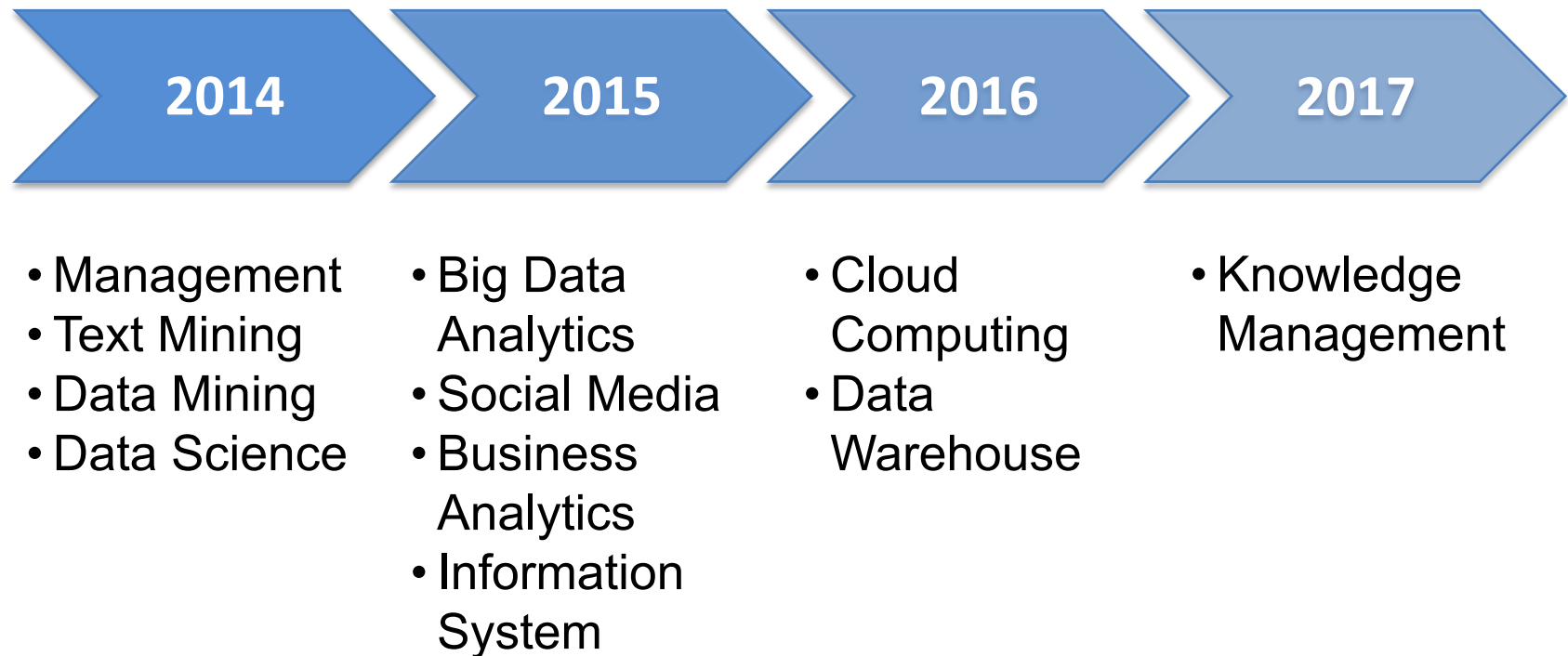
# Value Creation by Big Data Analytics

(Grover et al., 2018)

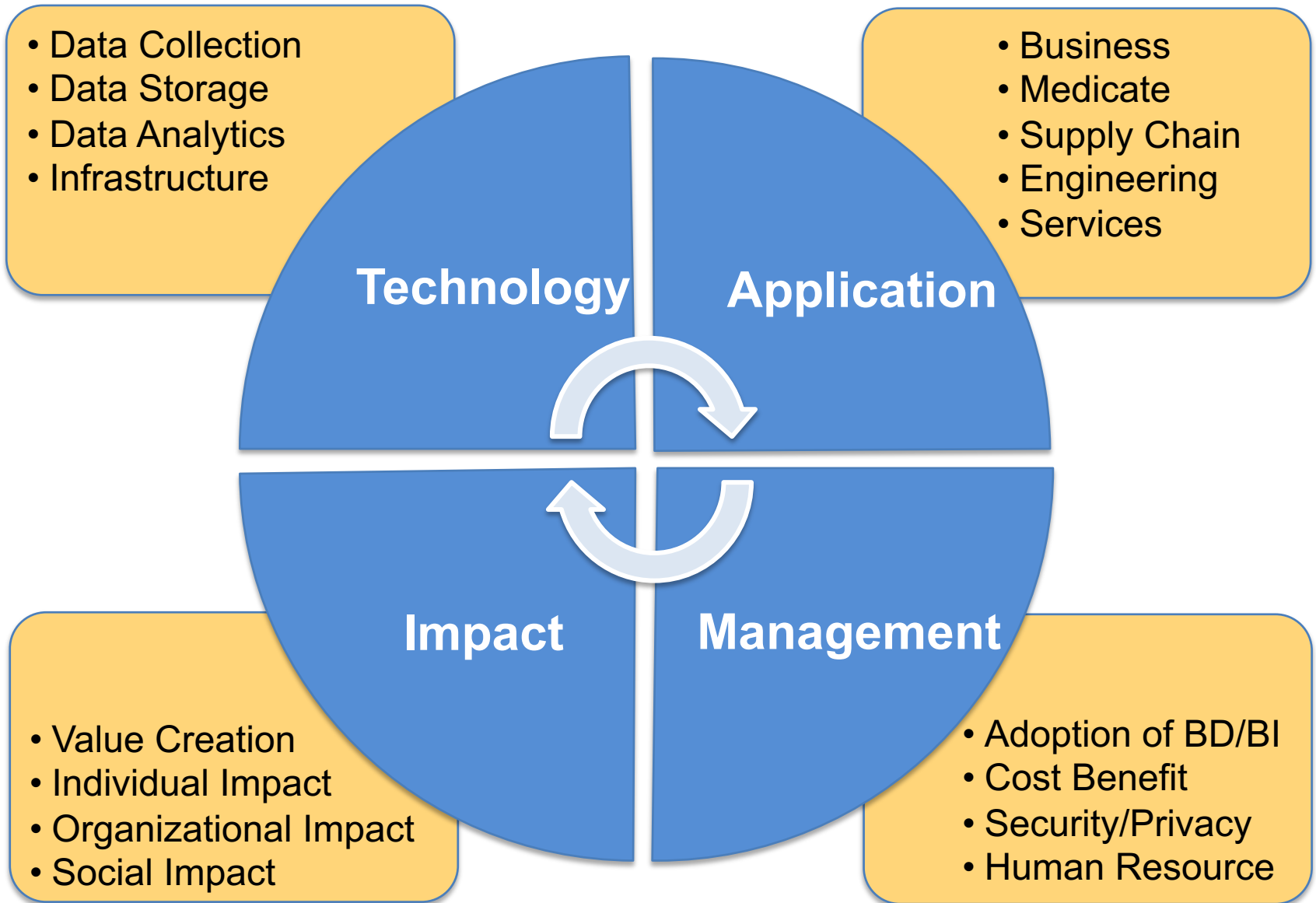


Investments --- Assets ----- Capabilities ----- Applications ----- Targets ----- Impacts ----- Value

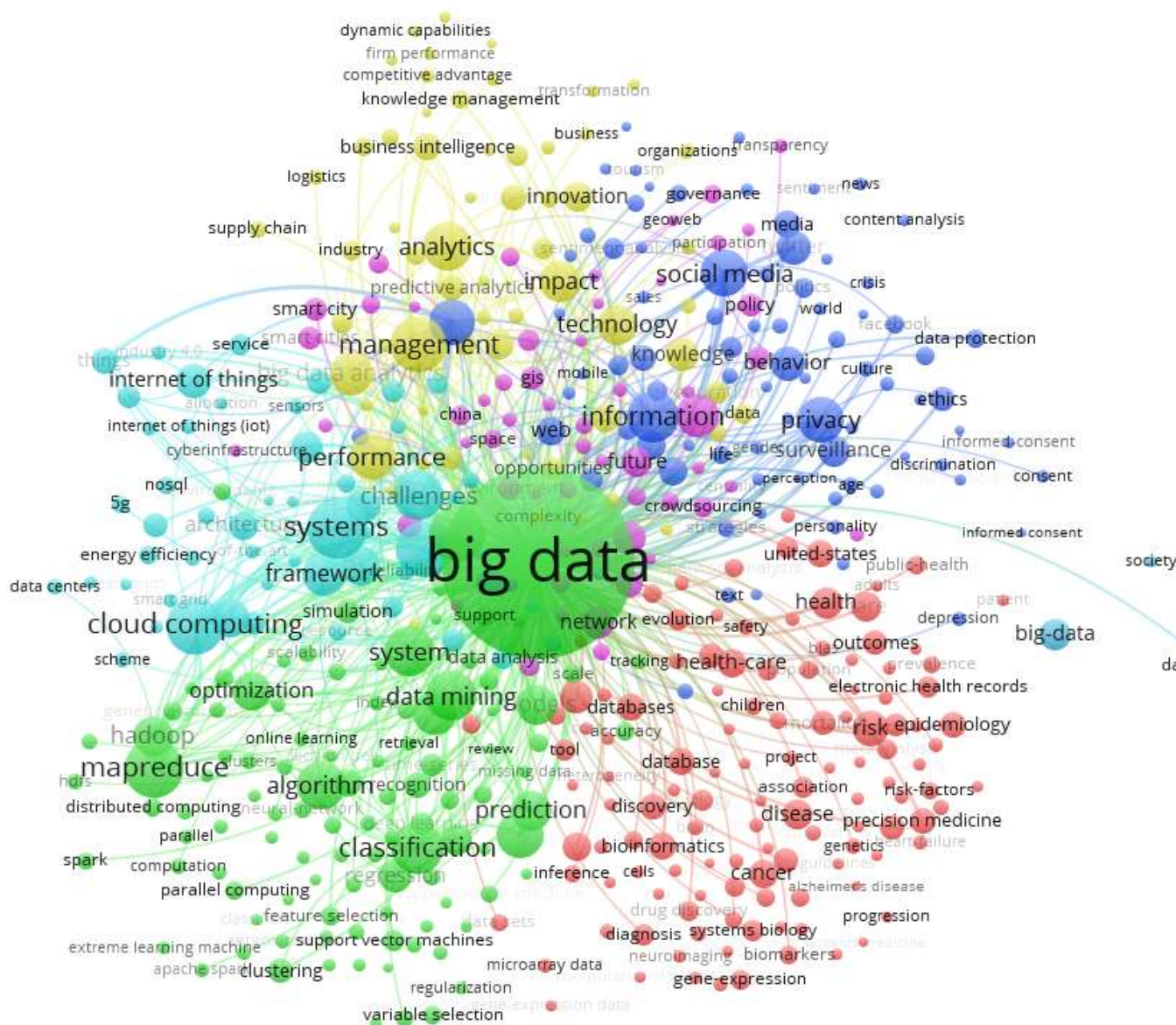
# Evolution of top keywords in “BD & BI” publications



# Framework for BD and BI Research



# Business Intelligence and Big Data analytics



Source: Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, Volume 111, 30, 2018, pp. 2-10

**AI**



# **Definition of Artificial Intelligence (A.I.)**

# Artificial Intelligence

**“... the science and  
engineering  
of  
making  
intelligent machines”  
(John McCarthy, 1955)**

# Artificial Intelligence

**“... technology that  
thinks and acts  
like humans”**

# Artificial Intelligence

**“... intelligence  
exhibited by machines  
or software”**

# 4 Approaches of AI

<b>Thinking Humanly</b>	<b>Thinking Rationally</b>
<b>Acting Humanly</b>	<b>Acting Rationally</b>

# 4 Approaches of AI

**2.**

**Thinking Humanly:  
The Cognitive  
Modeling Approach**

**3.**

**Thinking Rationally:  
The “Laws of Thought”  
Approach**

**1.**

**Acting Humanly:  
The Turing Test  
Approach** (1950)

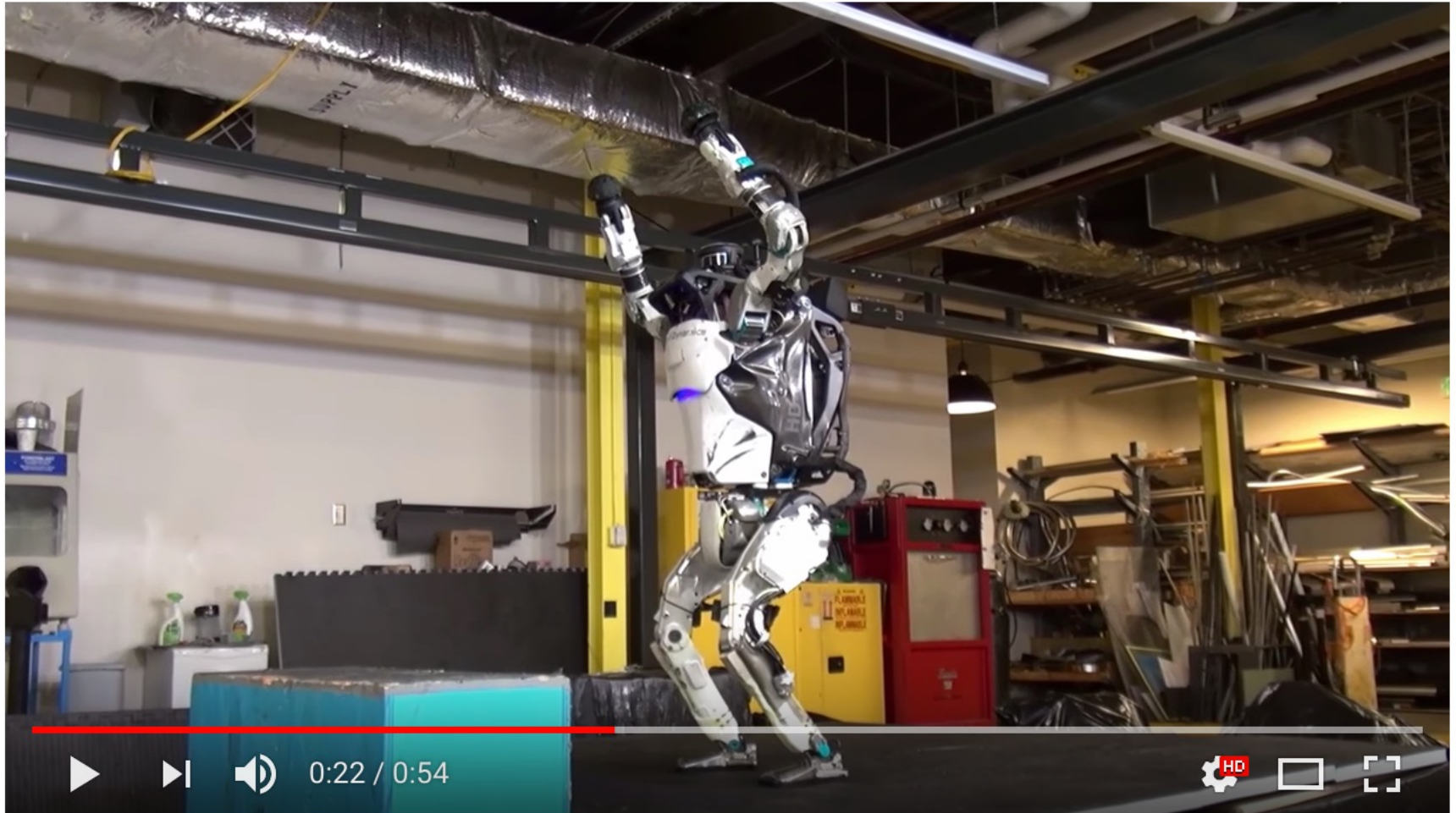
**4.**

**Acting Rationally:  
The Rational Agent  
Approach**

# AI Acting Humanly: The Turing Test Approach (Alan Turing, 1950)

- **Natural Language Processing (NLP)**
- **Knowledge Representation**
- **Automated Reasoning**
- **Machine Learning (ML)**
- **Computer Vision**
- **Robotics**

# Boston Dynamics: Atlas



#13 ON TRENDING

What's new, Atlas?

<https://www.youtube.com/watch?v=fRj34o4hN4I>



# Humanoid Robot: Sophia



<https://www.youtube.com/watch?v=S5t6K9iwcdw>

# Can a robot pass a university entrance exam?

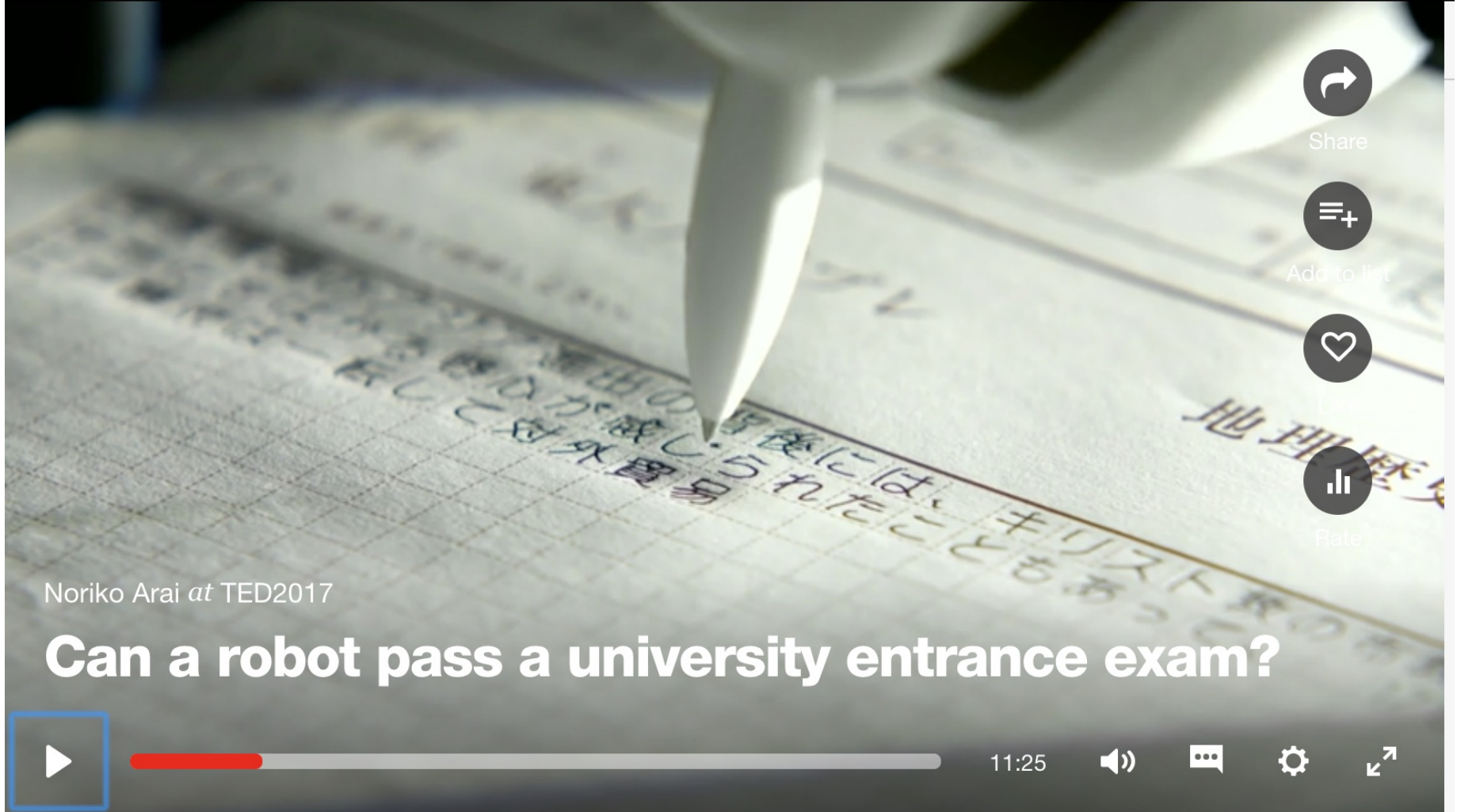
Noriko Arai at TED2017

**TED** Ideas worth spreading

WATCH

DISCOVER

ATT



Share



Add to list



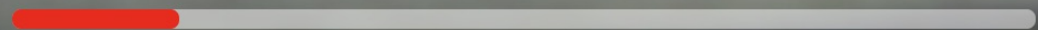
Like



Rate

Noriko Arai at TED2017

## Can a robot pass a university entrance exam?



11:25



[https://www.ted.com/talks/noriko\\_arai\\_can\\_a\\_robot\\_pass\\_a\\_university\\_entrance\\_exam](https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam)

<https://www.youtube.com/watch?v=XQZjkPyJ8KU>

# Artificial Intelligence (A.I.) Timeline

S/Z/Y/G/

## A.I. TIMELINE

1950

### TURING TEST

Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking it is human, then it has intelligence

1955

### A.I. BORN

Term 'artificial intelligence' is coined by computer scientist, John McCarthy to describe "the science and engineering of making intelligent machines"

1961

### UNIMATE

First industrial robot, Unimate, goes to work at GM replacing humans on the assembly line

1964

### ELIZA

Pioneering chatbot developed by Joseph Weizenbaum at MIT holds conversations with humans

1966

### SHAKY

The 'first electronic person' from Stanford, Shakey is a general-purpose mobile robot that reasons about its own actions

A.I. WINTER

Many false starts and dead-ends leave A.I. out in the cold

1997

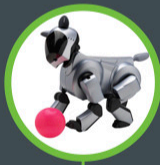
### DEEP BLUE

Deep Blue, a chess-playing computer from IBM defeats world chess champion Garry Kasparov

1998

### KISMET

Cynthia Breazeal at MIT introduces Kismet, an emotionally intelligent robot insofar as it detects and responds to people's feelings



1999

### AIBO

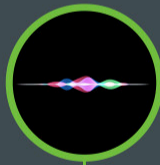
Sony launches first consumer robot pet dog AIBO (AI robot) with skills and personality that develop over time



2002

### ROOMBA

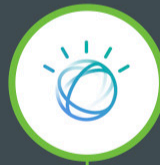
First mass produced autonomous robotic vacuum cleaner from iRobot learns to navigate and clean homes



2011

### SIRI

Apple integrates Siri, an intelligent virtual assistant with a voice interface, into the iPhone 4S



2011

### WATSON

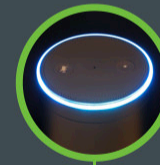
IBM's question answering computer Watson wins first place on popular \$1M prize television quiz show Jeopardy



2014

### EUGENE

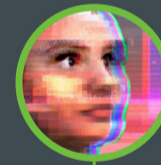
Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is human



2014

### ALEXA

Amazon launches Alexa, an intelligent virtual assistant with a voice interface that completes shopping tasks



2016

### TAY

Microsoft's chatbot Tay goes rogue on social media making inflammatory and offensive racist comments



2017

### ALPHAGO

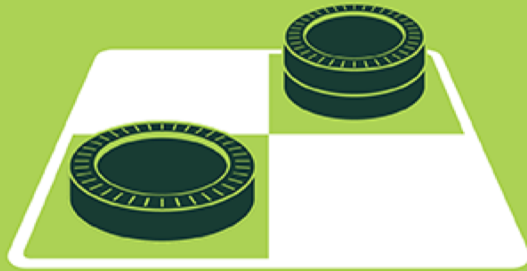
Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number (2<sup>170</sup>) of possible positions

# Artificial Intelligence

## Machine Learning & Deep Learning

### ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



### MACHINE LEARNING

Machine learning begins to flourish.



### DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

1980's

1990's

2000's

2010's

Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# AI, ML, DL

## Artificial Intelligence (AI)

### Machine Learning (ML)

Supervised  
Learning

Unsupervised  
Learning

### Deep Learning (DL)

CNN

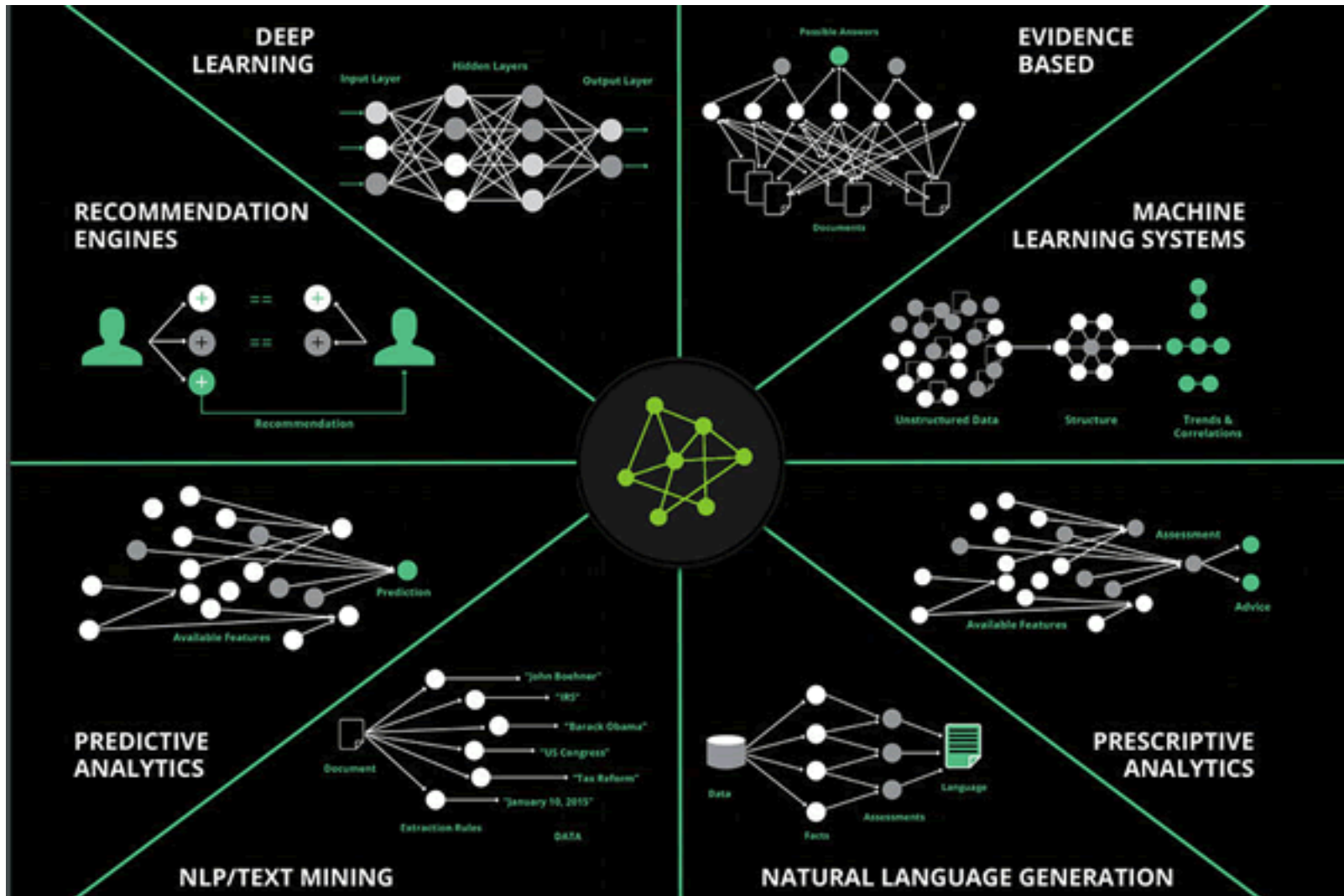
RNN LSTM GRU

GAN

Semi-supervised  
Learning

Reinforcement  
Learning

# Artificial Intelligence (AI) is many things

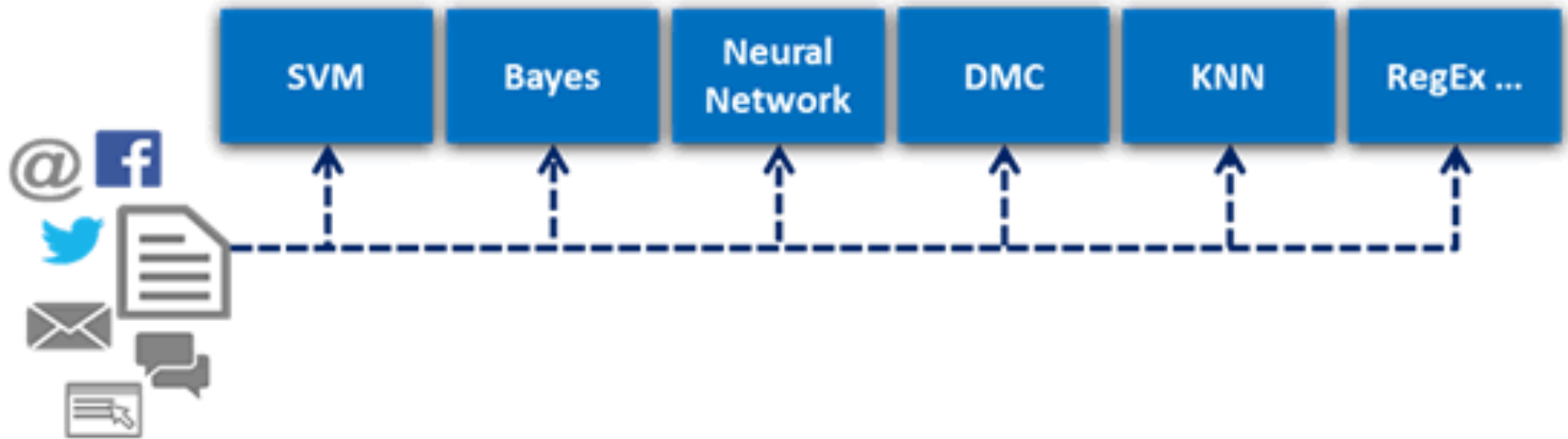


## Ecosystem of AI

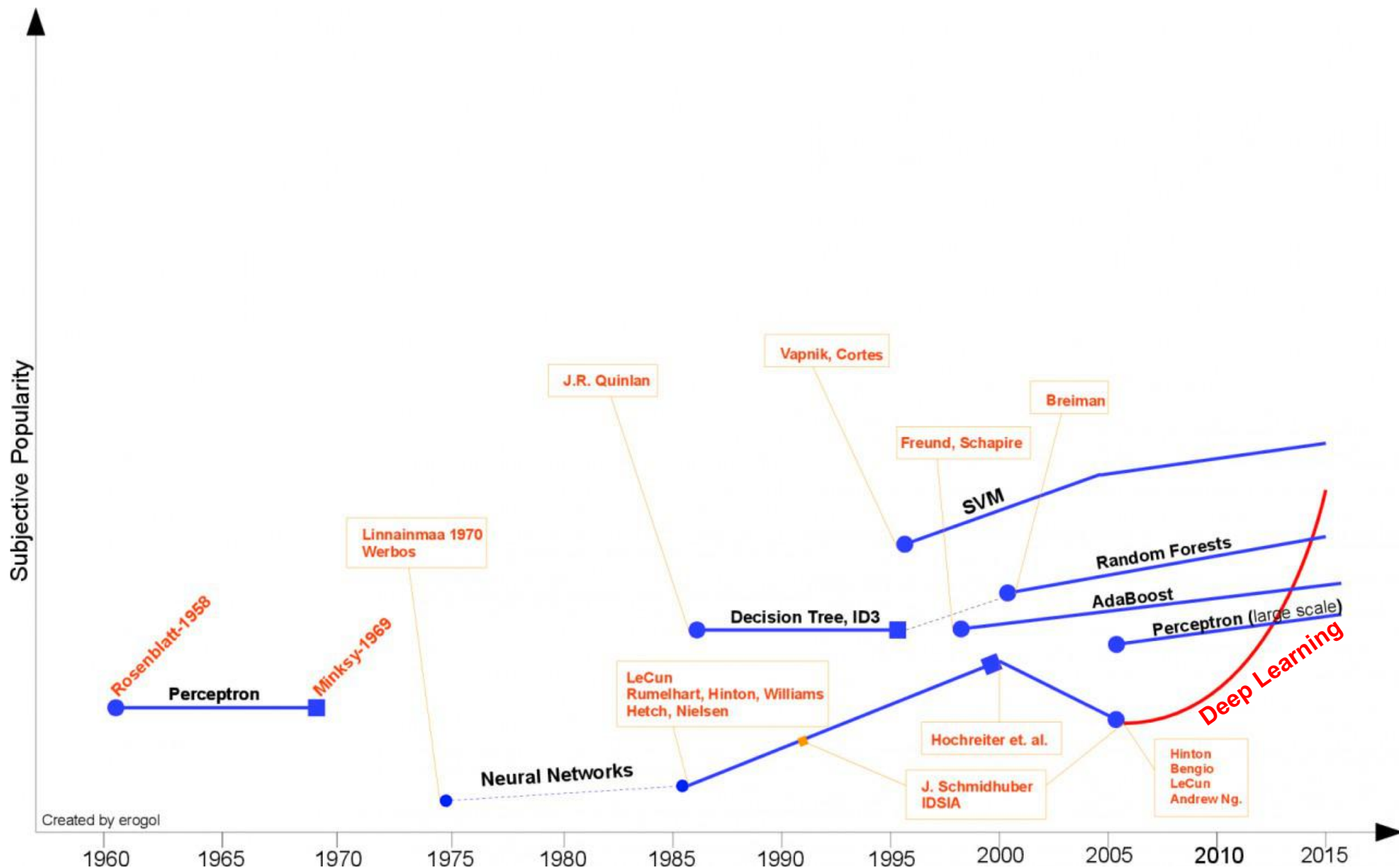
Source: <https://www.i-scoop.eu/artificial-intelligence-cognitive-computing/>

# Artificial Intelligence (AI)

## Intelligent Document Recognition algorithms



# Deep Learning Evolution





# Machine Learning Models

Deep Learning

Association rules

Decision tree

Clustering

Bayesian

Kernel

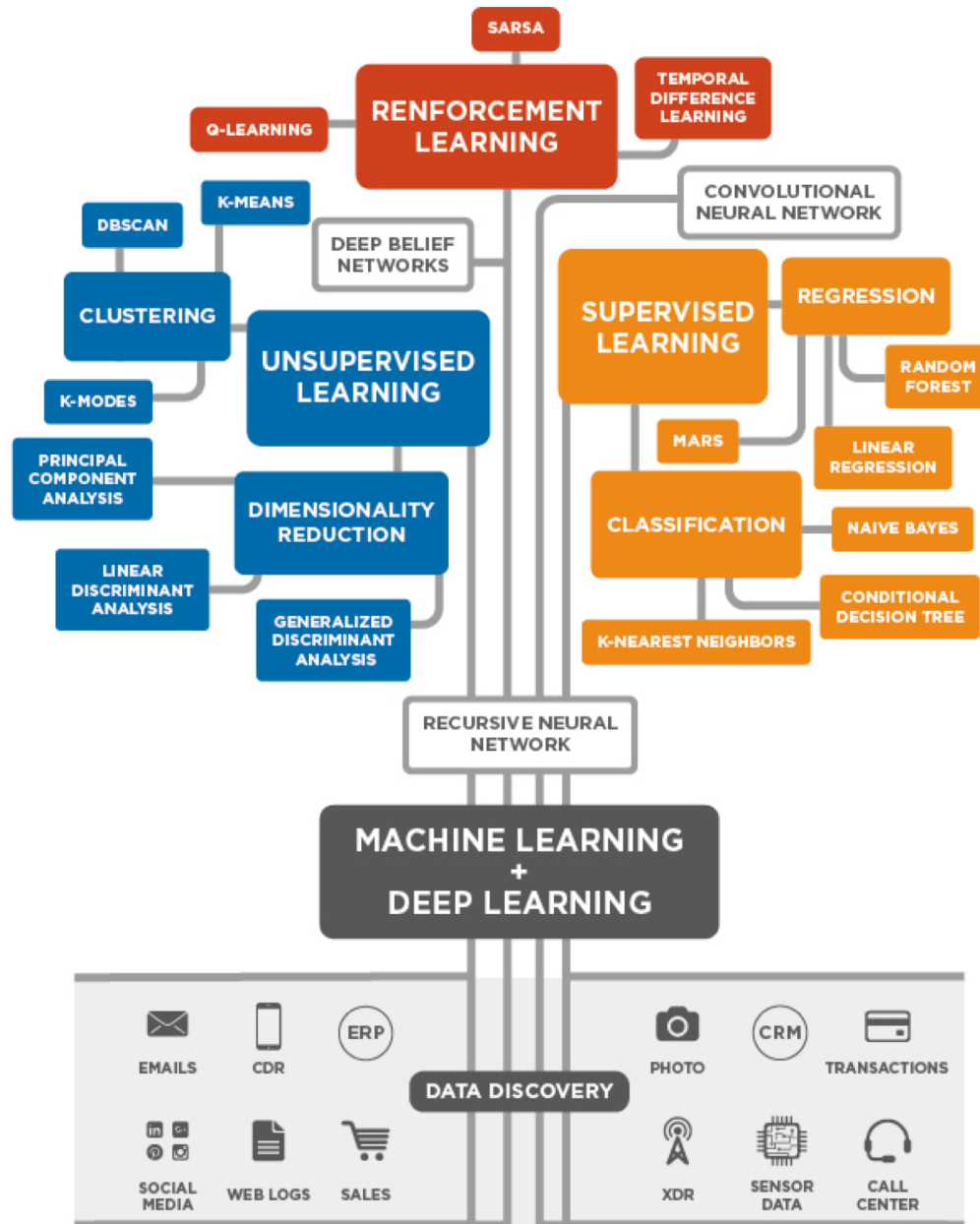
Ensemble

Dimensionality reduction

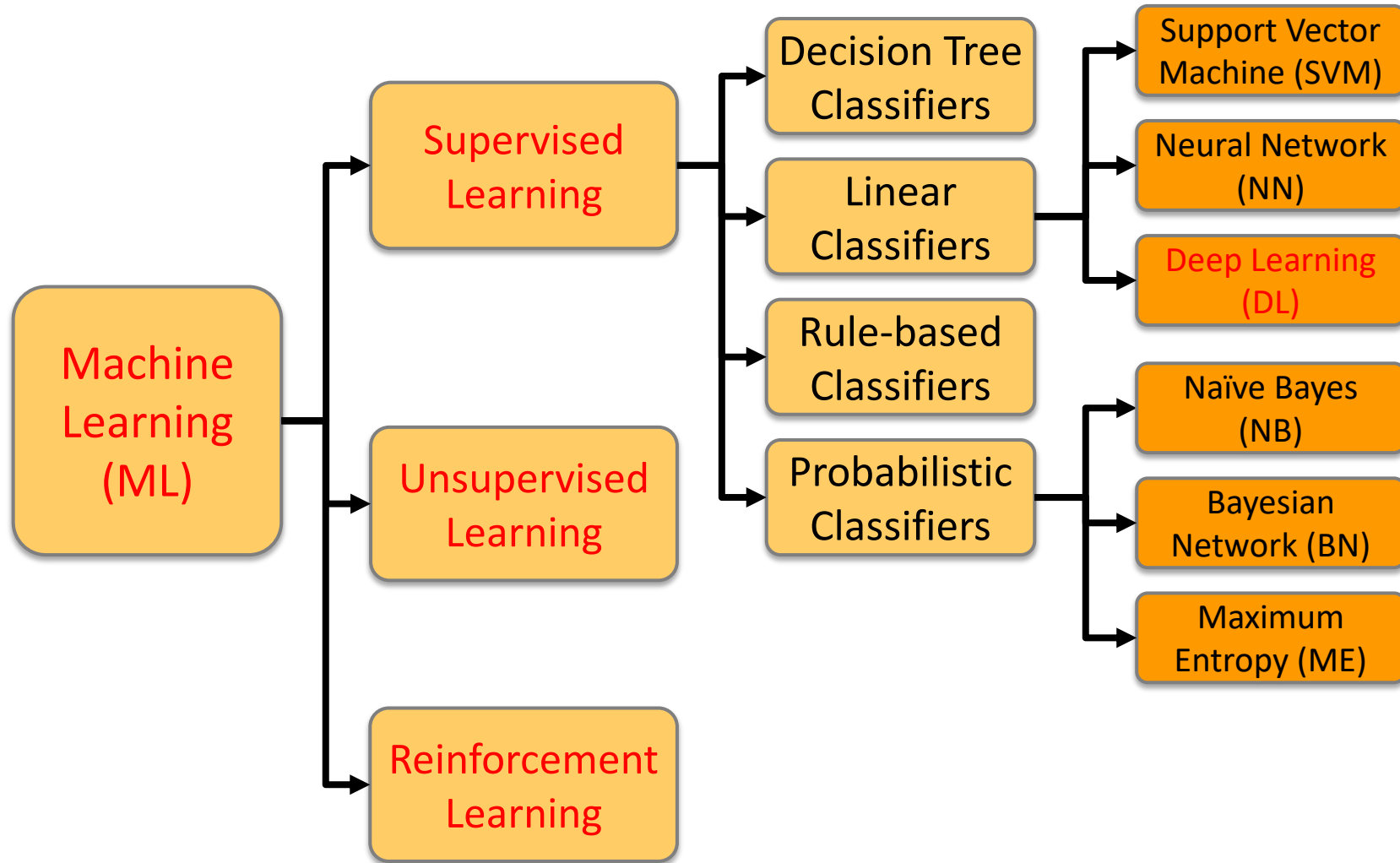
Regression Analysis

Instance based

# 3 Machine Learning Algorithms

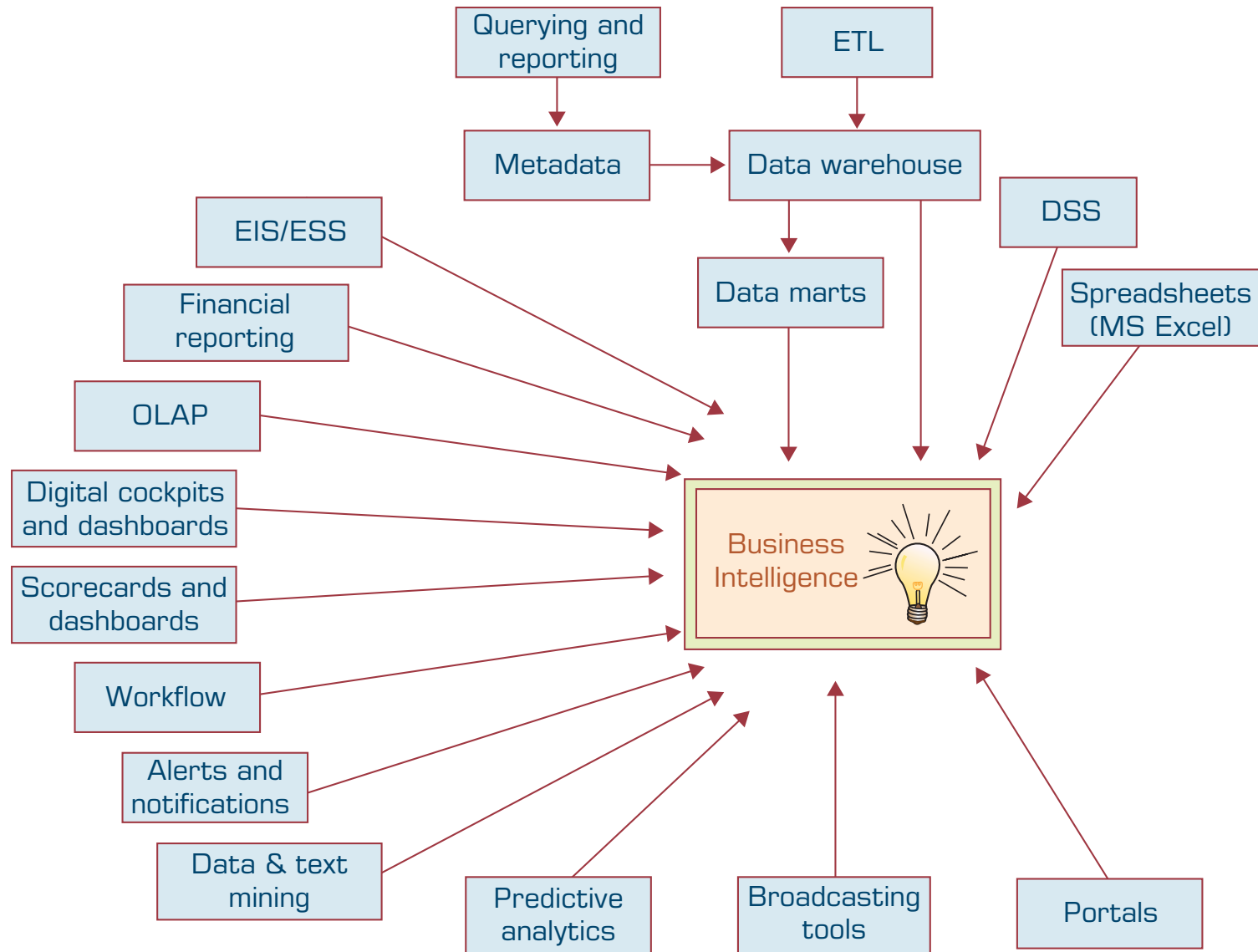


# Machine Learning (ML) / Deep Learning (DL)

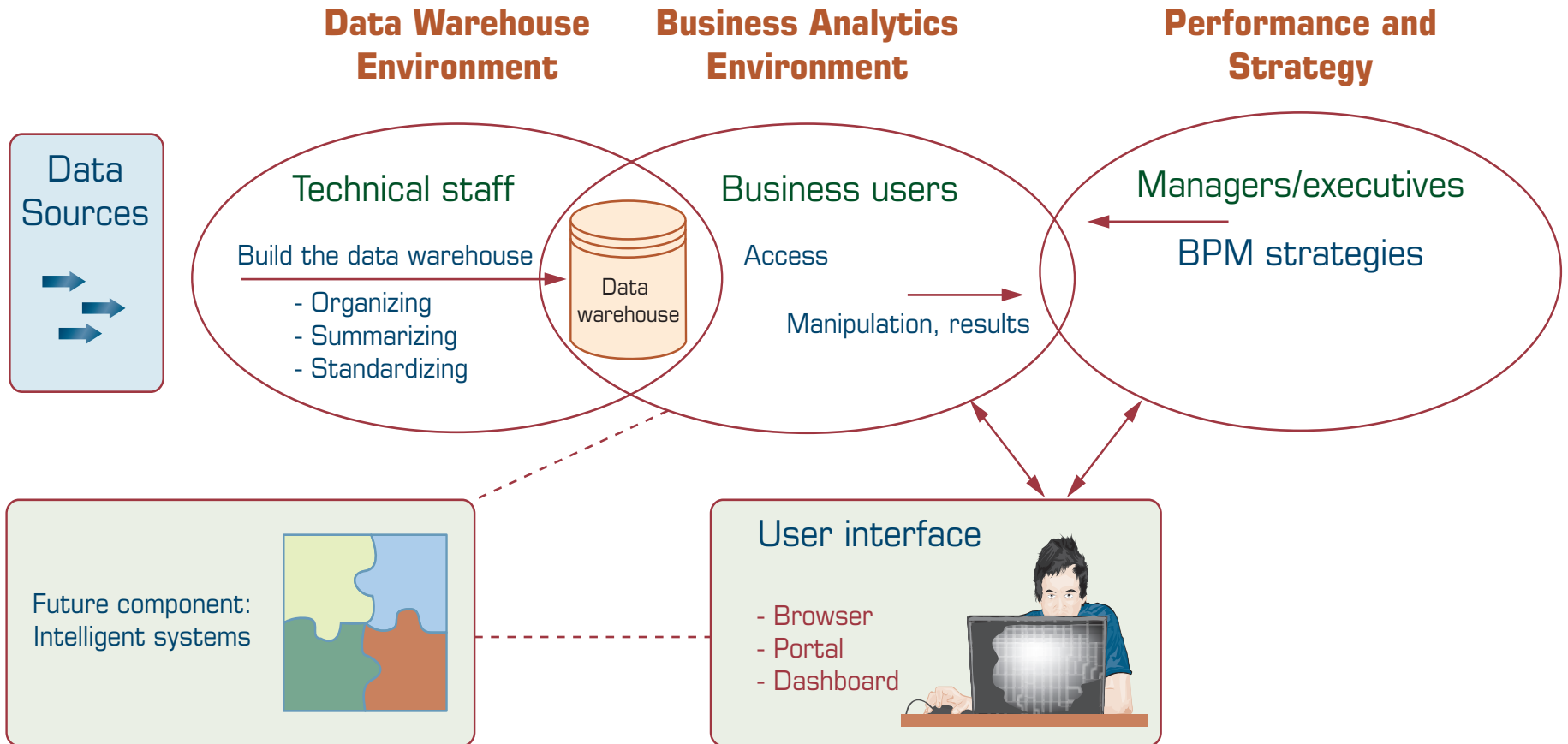


# Big Data

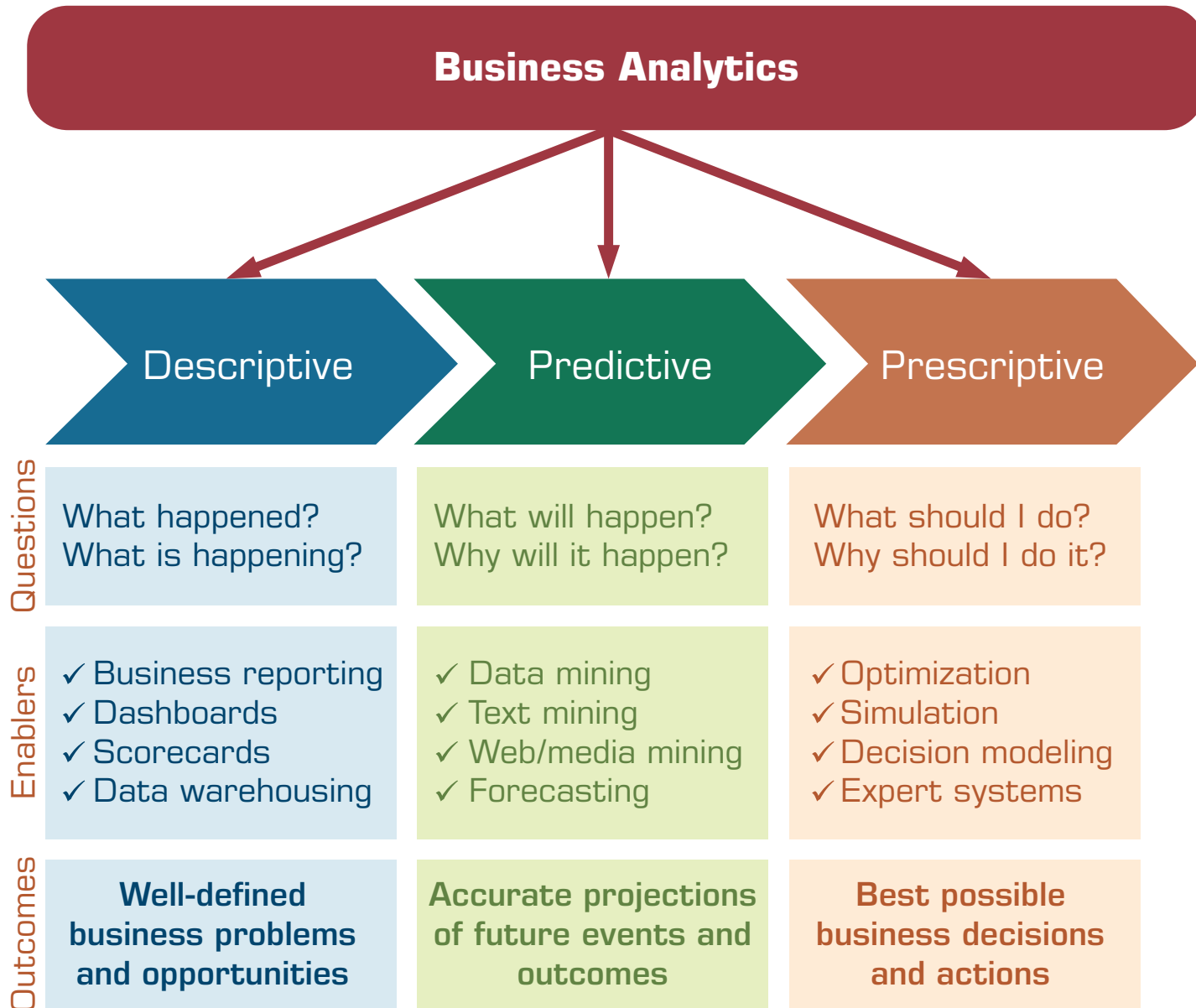
# Evolution of Business Intelligence (BI)



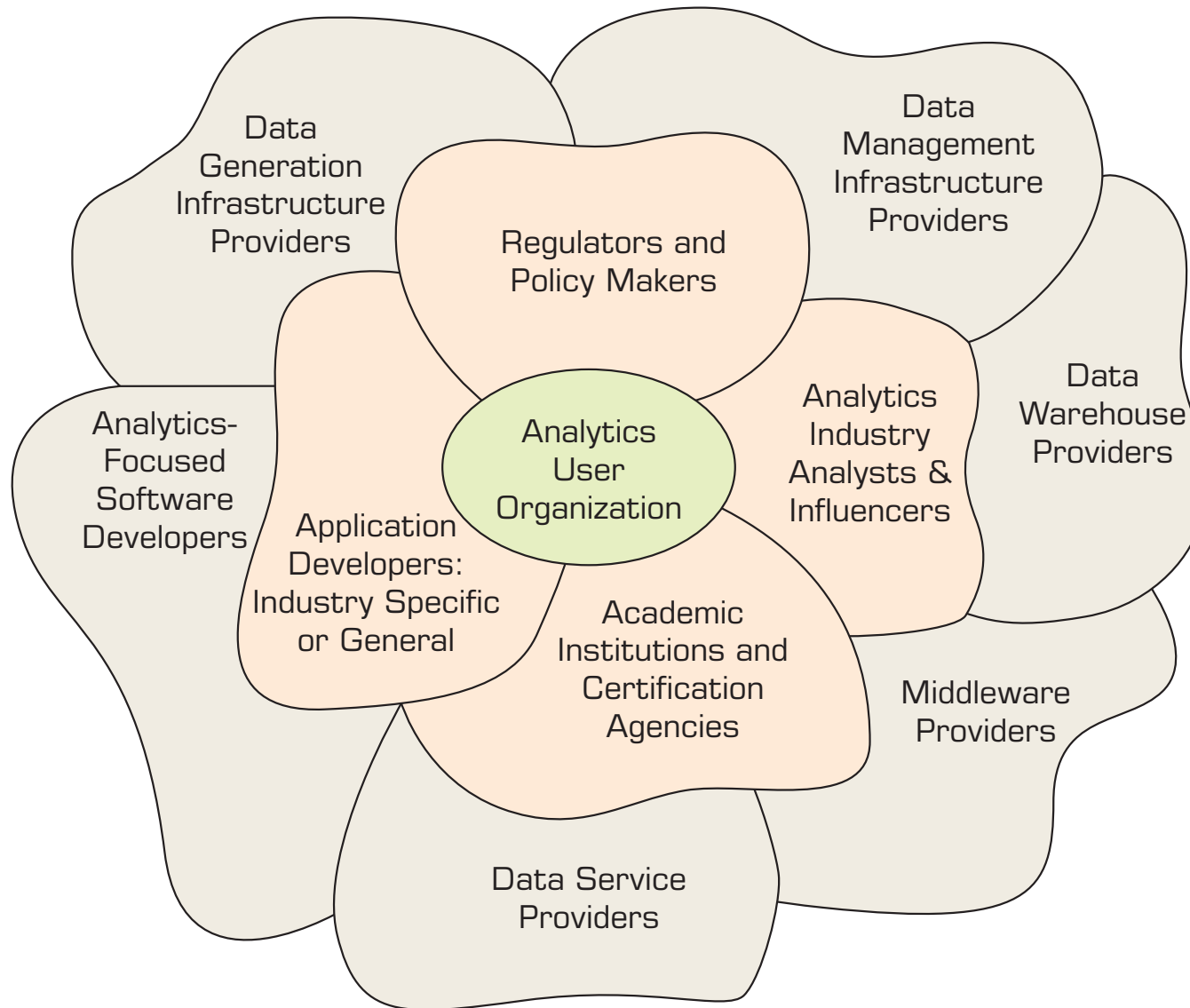
# A High-Level Architecture of BI



# Three Types of Analytics



# Analytics Ecosystem

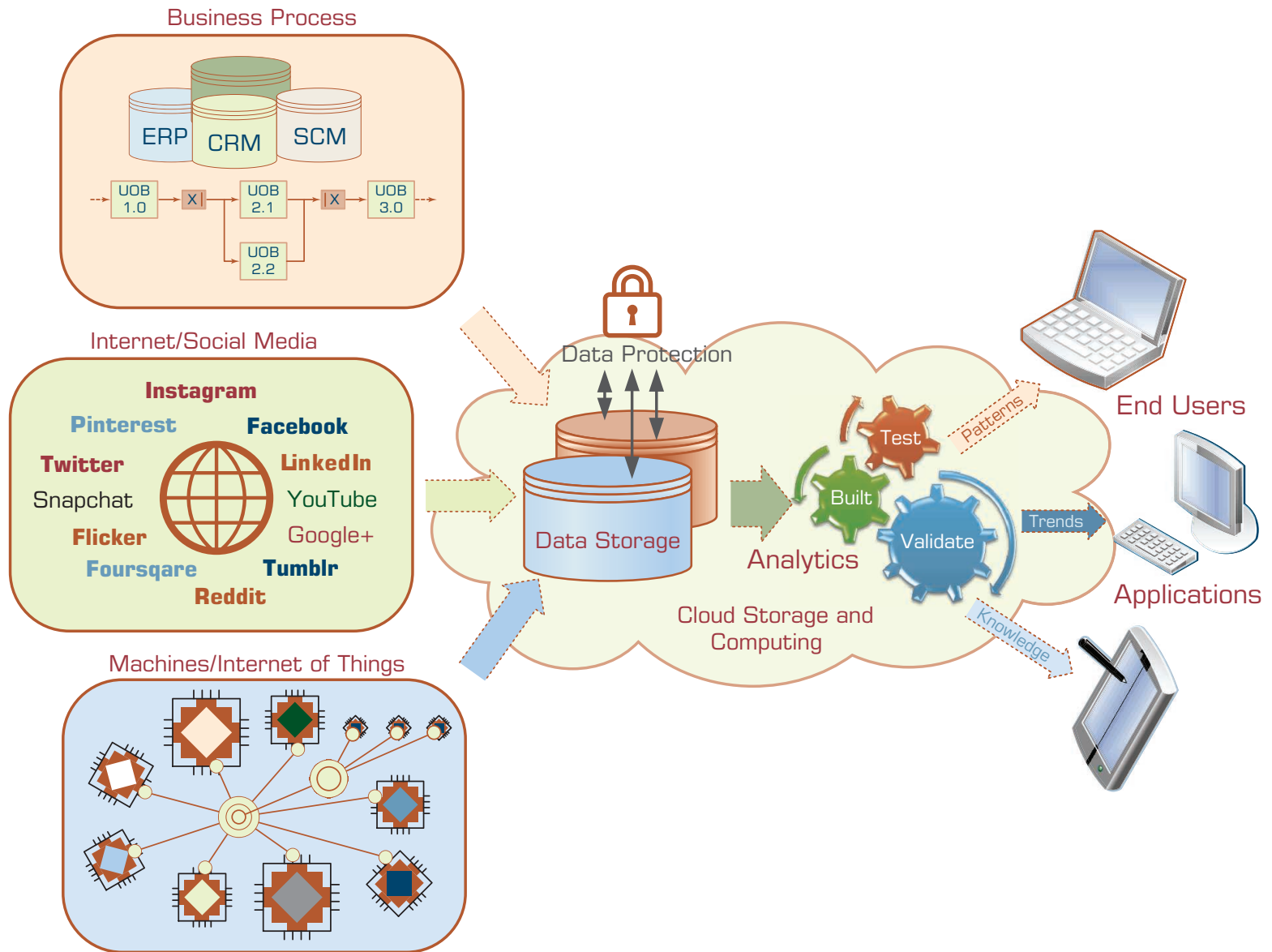




# Job Titles of Analytics

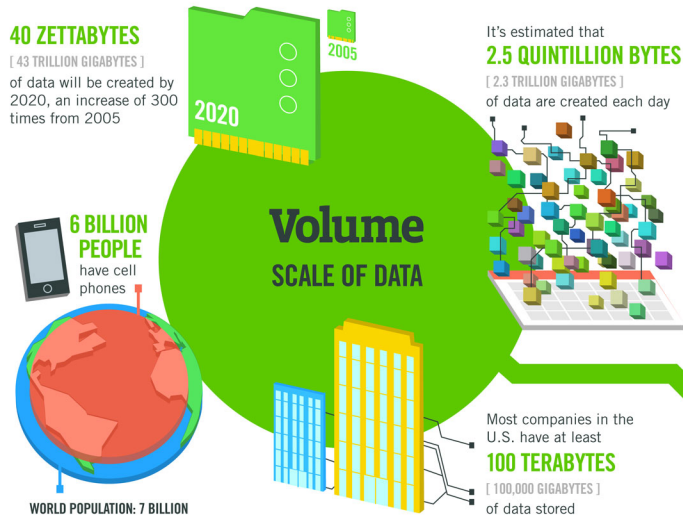


# A Data to Knowledge Continuum



**Big Data**  
**Analytics**  
and  
**Data Mining**

# Big Data 4 V



## The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES**  
[161 BILLION GIGABYTES]



**30 BILLION PIECES OF CONTENT** are shared on Facebook every month



By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO** are watched on YouTube each month



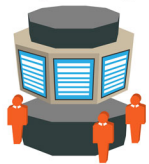
**400 MILLION TWEETS** are sent per day by about 200 million monthly active users



## Variety

DIFFERENT FORMS OF DATA

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

## Velocity

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** – almost 2.5 connections per person on earth



**1 IN 3 BUSINESS LEADERS** don't trust the information they use to make decisions



Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



**27% OF RESPONDENTS**

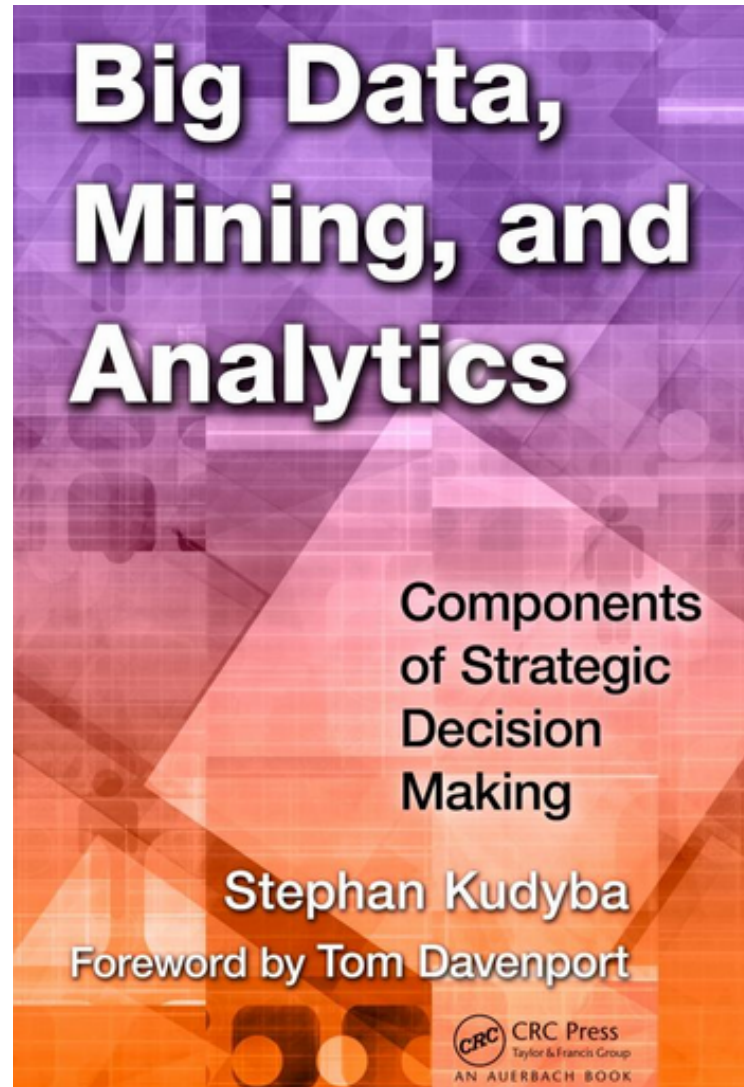
## Veracity

UNCERTAINTY OF DATA

in one survey were unsure of how much of their data was inaccurate

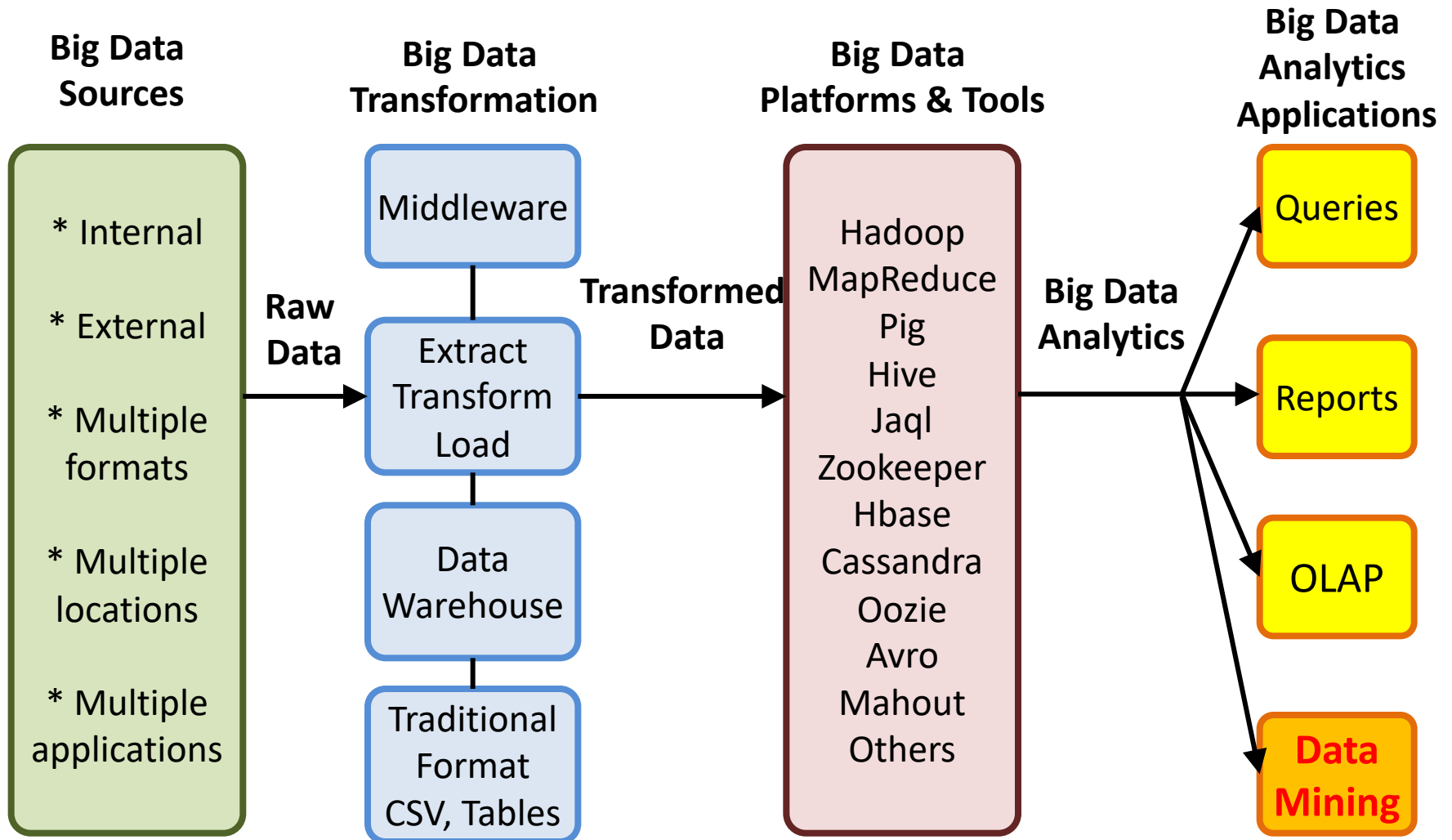
**value**

Stephan Kudyba (2014),  
**Big Data, Mining, and Analytics:**  
**Components of Strategic Decision Making**, Auerbach Publications



Source: <http://www.amazon.com/gp/product/1466568704>

# Architecture of Big Data Analytics



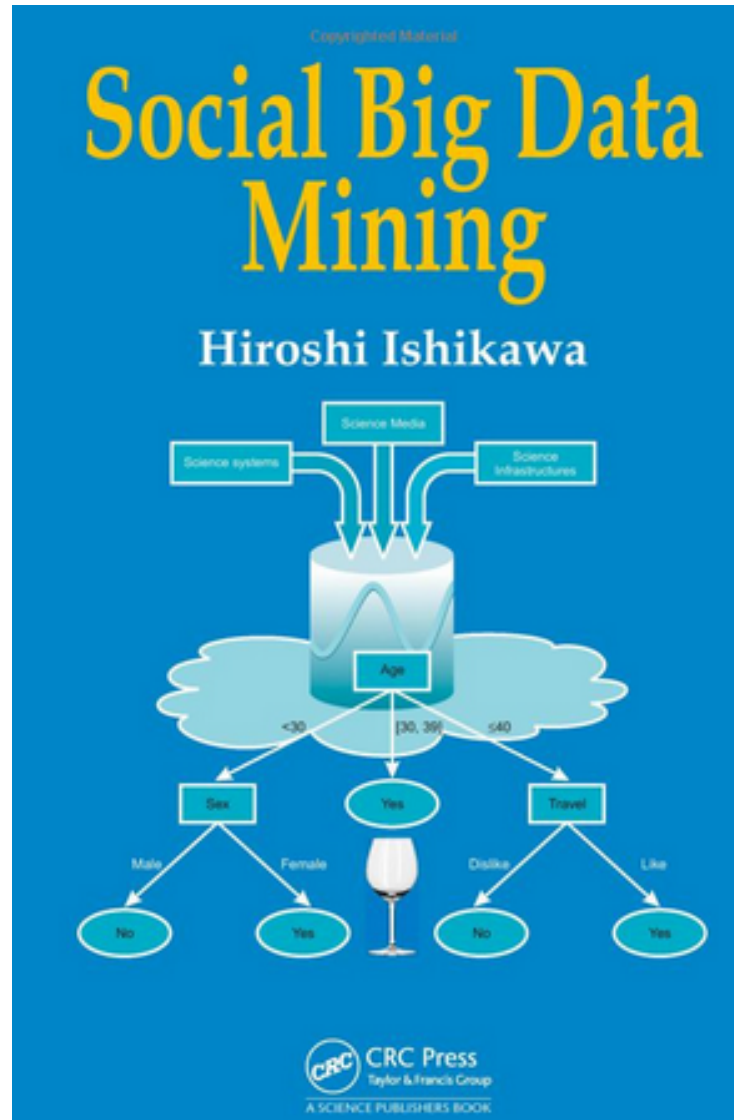
# Architecture of Big Data Analytics





# Social Big Data Mining

(Hiroshi Ishikawa, 2015)

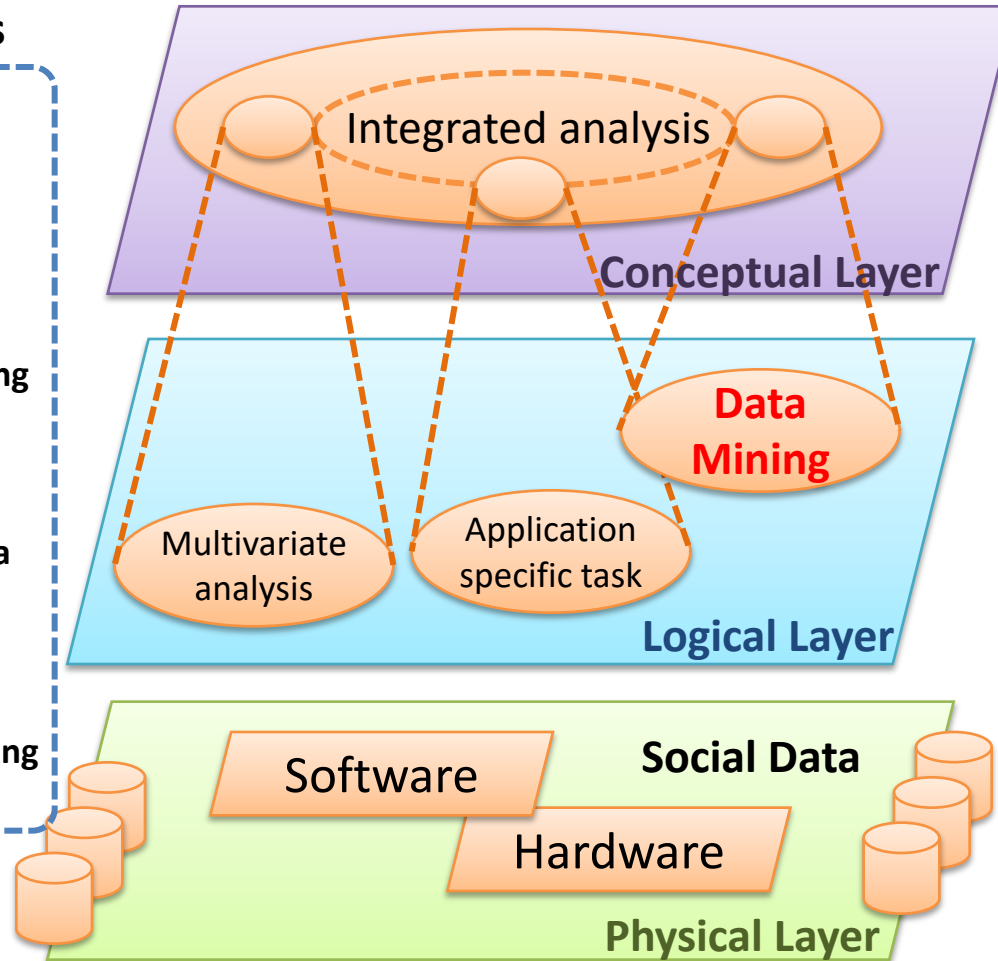


# Architecture for Social Big Data Mining

(Hiroshi Ishikawa, 2015)

## Enabling Technologies

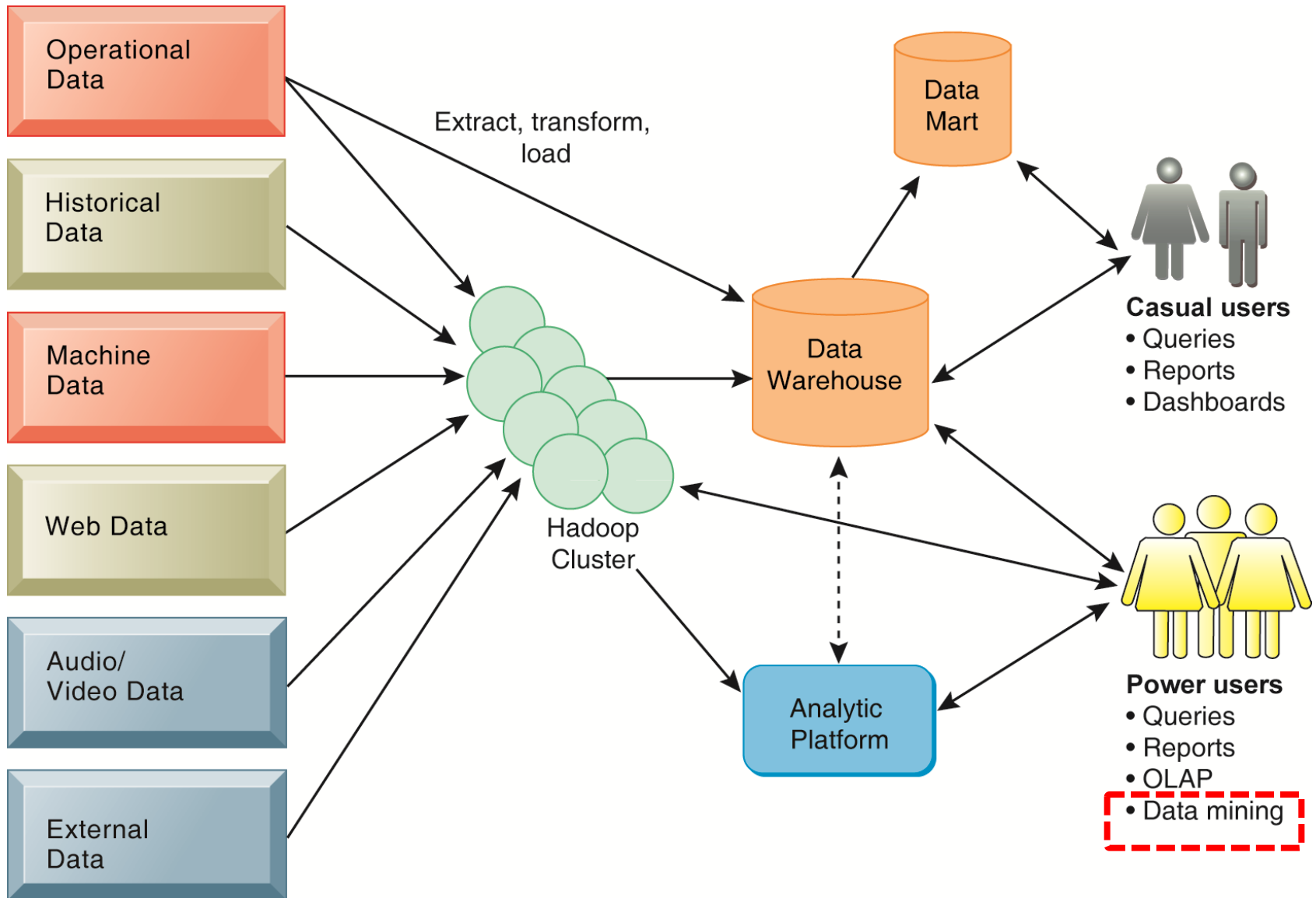
- Integrated analysis model
- Natural Language Processing
- Information Extraction
- Anomaly Detection
- Discovery of relationships among heterogeneous data
- Large-scale visualization
- Parallel distributed processing



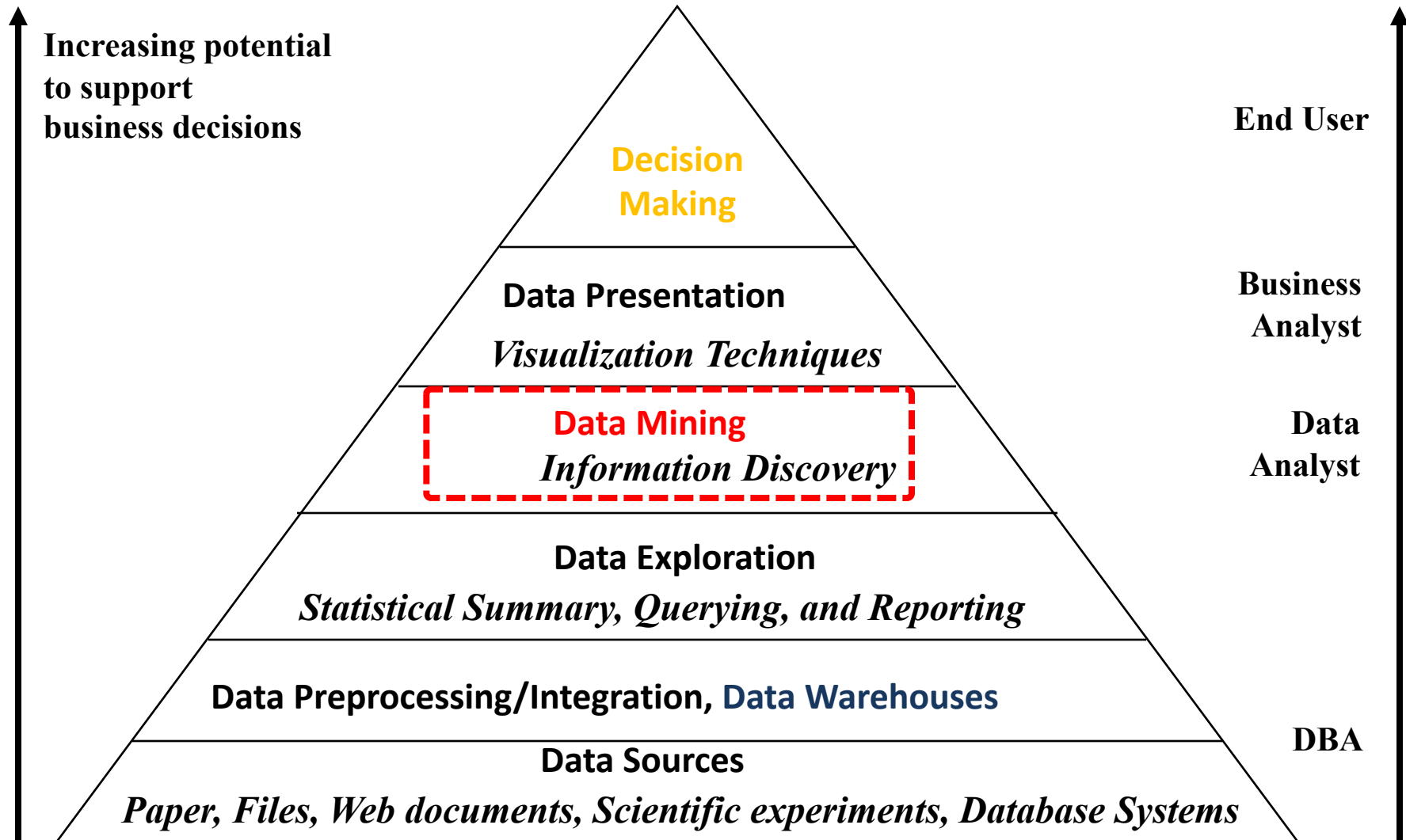
## Analysts

- Model Construction
- Explanation by Model
- Construction and confirmation of individual hypothesis
- Description and execution of application-specific task

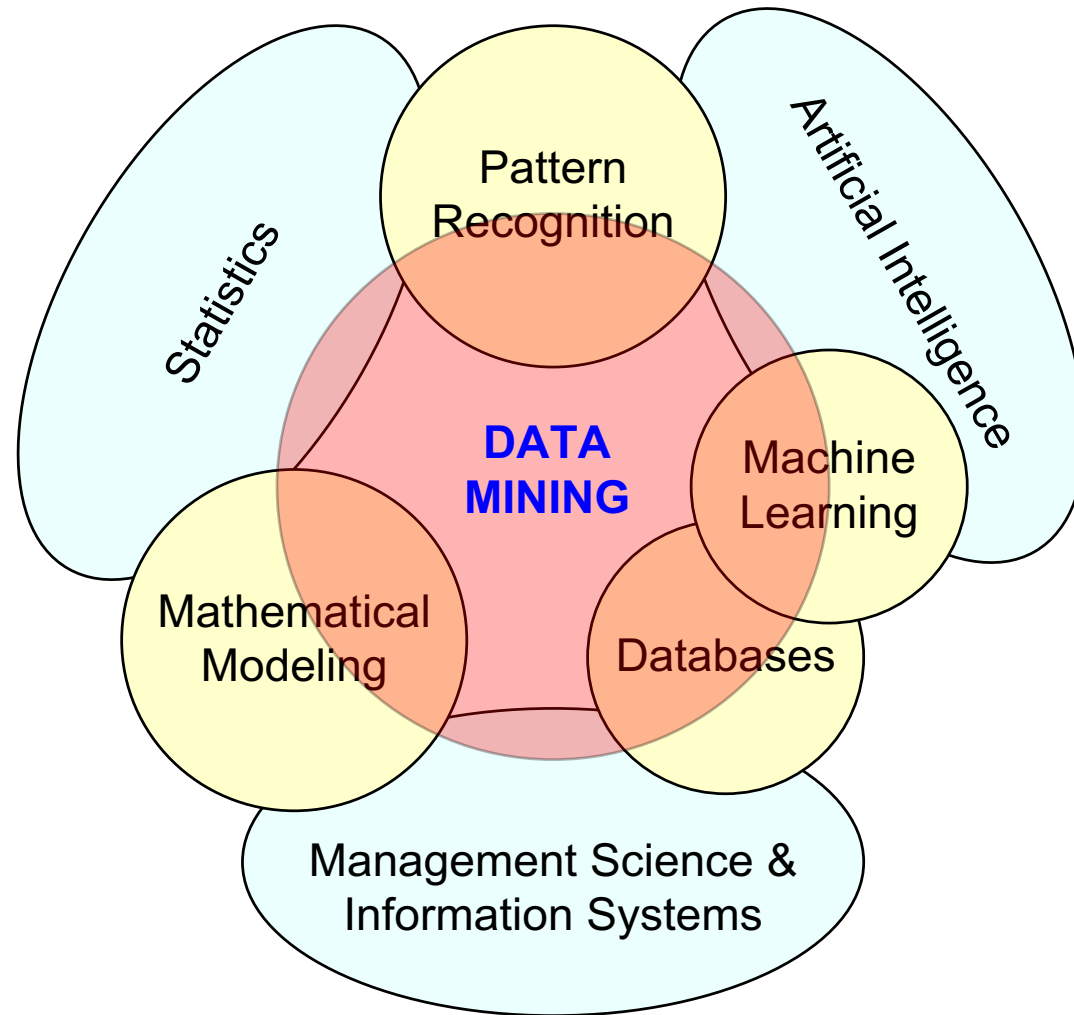
# Business Intelligence (BI) Infrastructure



# Business Intelligence and Data Mining



# Data Mining at the Intersection of Many Disciplines





# Data Mining:

Core **Analytics** Process

The **KDD** Process for  
Extracting Useful **Knowledge**  
from Volumes of **Data**

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996).

The **KDD Process** for  
Extracting Useful **Knowledge**  
from Volumes of **Data**.

Communications of the ACM, 39(11), 27-34.

Knowledge Discovery in Databases creates the context for developing the tools needed to control the flood of data facing organizations that depend on ever-growing databases of business, manufacturing, scientific, and personal information.

## The KDD Process for Extracting Useful Knowledge from Volumes of Data

AS WE MARCH INTO THE AGE of digital information, the problem of data overload looms ominously ahead. Our ability to analyze and understand massive datasets lags far behind our ability to gather and store the data. A new generation of computational techniques and tools is required to support the extraction of useful knowledge from the rapidly growing volumes of data. These techniques and tools are the subject of the emerging field of knowledge discovery in databases (KDD) and data mining.

Large databases of digital information are ubiquitous. Data from the neighborhood store's checkout register, your bank's credit card authorization device, records in your doctor's office, patterns in your telephone calls,

Usama Fayyad,  
Gregory Piatetsky-Shapiro,  
and Padhraic Smyth

and many more applications generate streams of digital records archived in huge databases, sometimes in so-called data warehouses.

Current hardware and database technology allow efficient and inexpensive reliable data storage and access. However, whether the context is business, medicine, science, or government, the datasets themselves (in raw form) are of little direct value. What is of value is the knowledge that can be inferred from the data and put to use. For example, the marketing database of a consumer

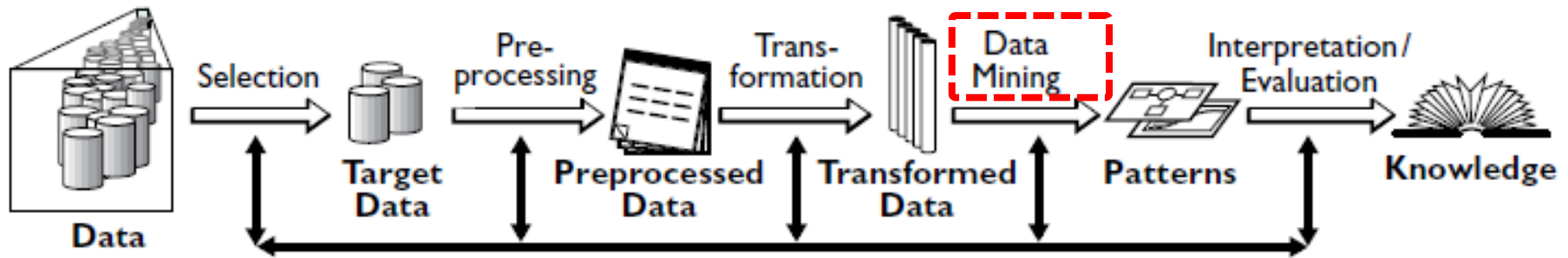


TEHRAN UNIVERSITY

# Data Mining

## Knowledge Discovery in Databases (KDD) Process

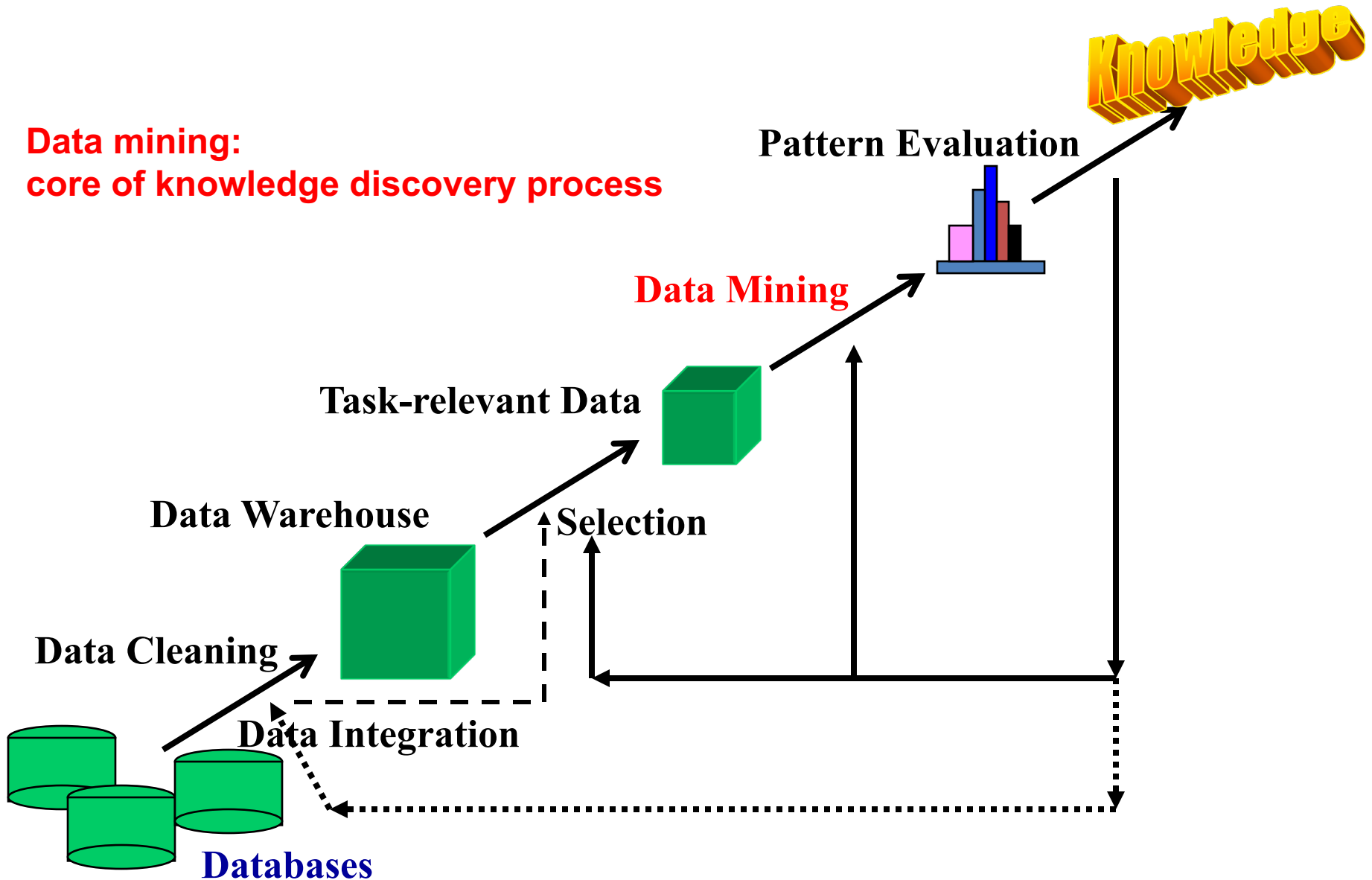
(Fayyad et al., 1996)





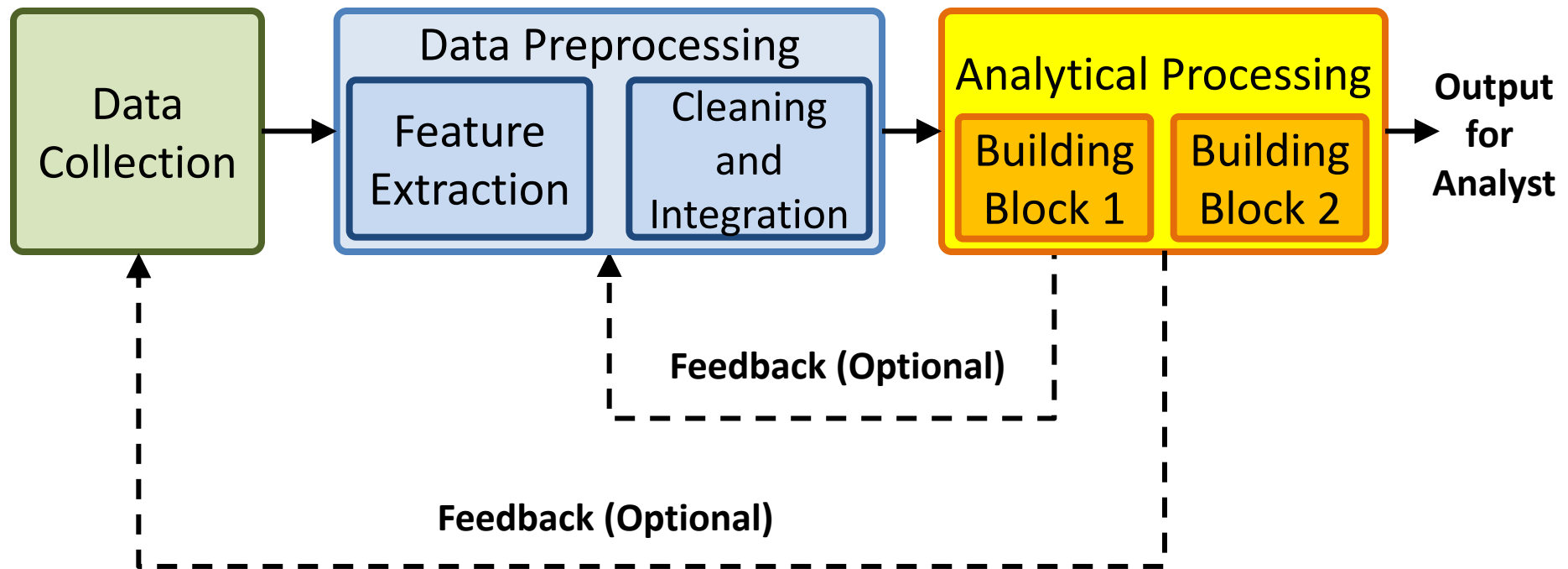
# Knowledge Discovery (KDD) Process

**Data mining:**  
core of knowledge discovery process

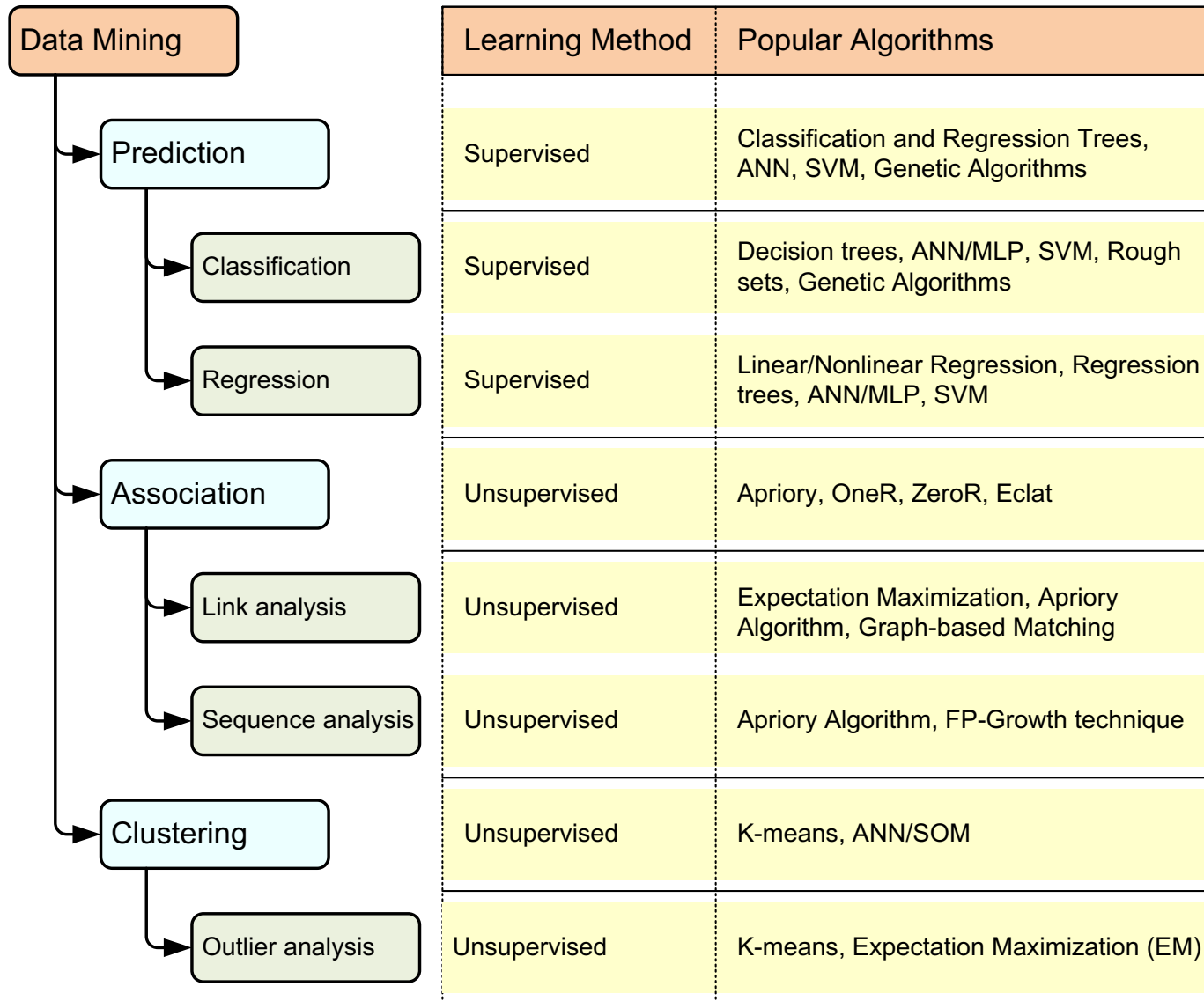


# Data Mining Processing Pipeline

(Charu Aggarwal, 2015)



# A Taxonomy for Data Mining Tasks

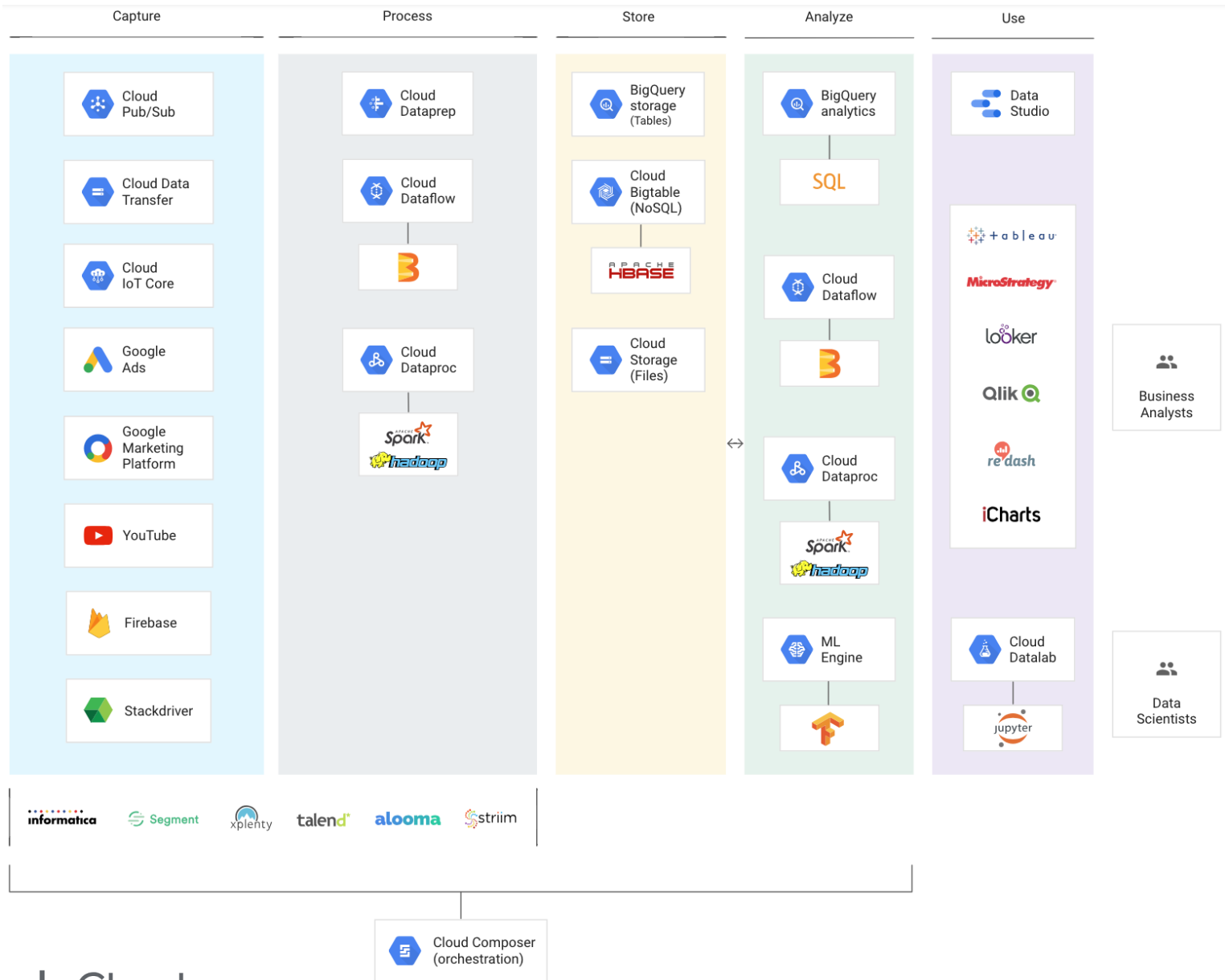


# Cloud Computing



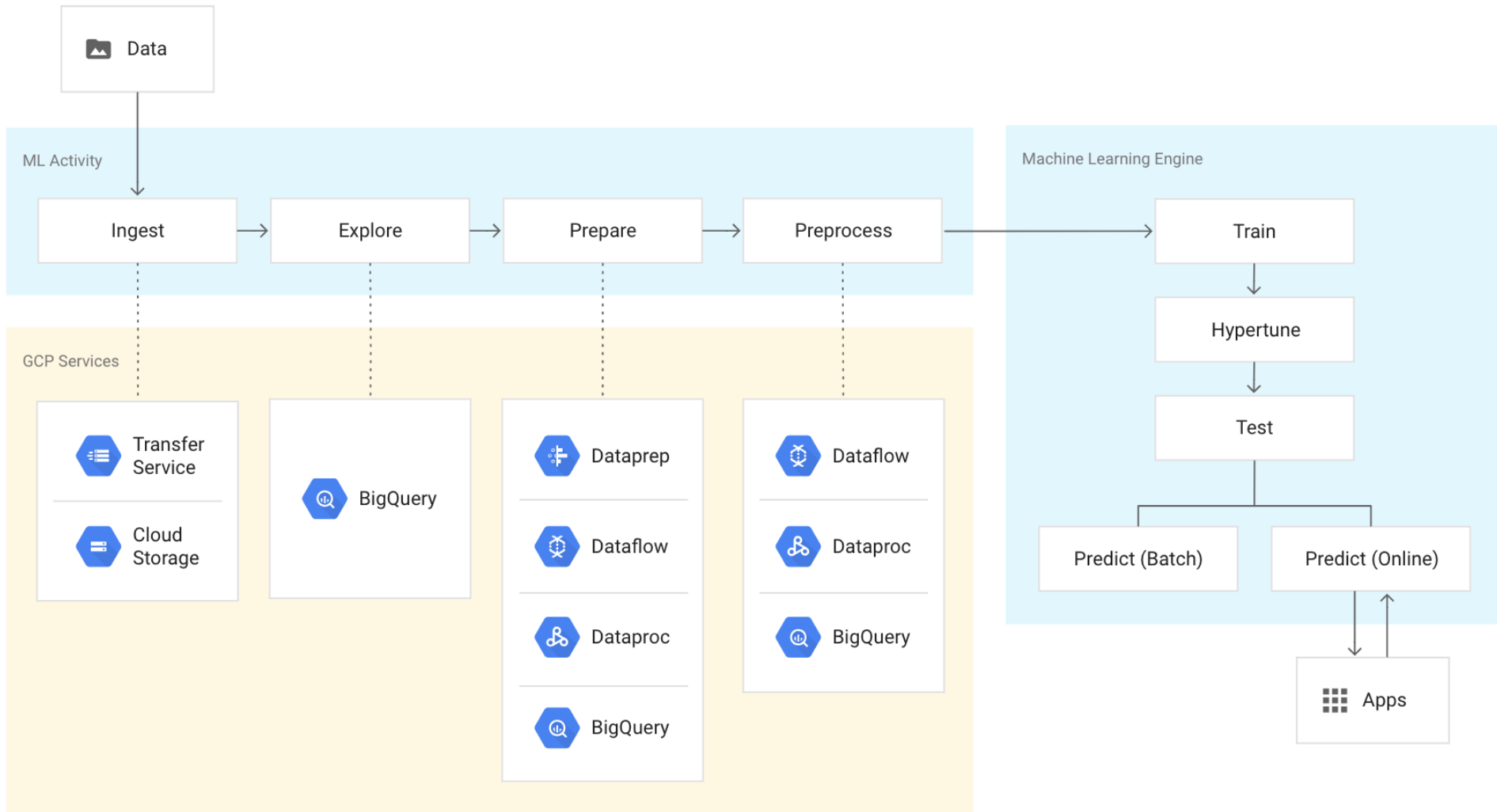
Google Cloud

# Google Cloud Big Data Analytics



# Google Cloud

## Machine learning and Cloud AI



# Google Colab

The screenshot shows the Google Colaboratory web interface. At the top, the browser address bar displays the URL <https://colab.research.google.com/notebooks/welcome.ipynb>. The main header includes the Colab logo, the text "Hello, Colaboratory", and a menu with options: File, Edit, View, Insert, Runtime, Tools, and Help. On the right side of the header, there are "SHARE" and "CONNECT" buttons, along with a user profile icon. Below the header, a toolbar contains buttons for "CODE", "TEXT", "CELL", "COPY TO DRIVE", and "EDITING". A left-hand sidebar contains a "Table of contents" with sections: "Getting Started", "Highlighted Features", "TensorFlow execution", "GitHub", "Visualization", "Forms", "Examples", and "Local runtime support". The main content area features a large "Welcome to Colaboratory!" message with the Colab logo and a brief description. Below this, a "Getting Started" section lists several links: "Overview of Colaboratory", "Loading and saving data: Local files, Drive, Sheets, Google Cloud Storage", "Importing libraries and installing dependencies", "Using Google Cloud BigQuery", "Forms, Charts, Markdown, & Widgets", "TensorFlow with GPU", and "Machine Learning Crash Course: Intro to Pandas & First Steps with TensorFlow". A "Highlighted Features" section is partially visible, starting with a "Seedbank" subsection that mentions a link to discover machine learning examples. The "TensorFlow execution" subsection is also partially visible, mentioning that Colab allows executing TensorFlow code in the browser.





# Cloud Computing

## AWS

### Amazon Web Services



Compute



Storage



Database



Migration



Networking & Content Delivery



Developer Tools



Management Tools



Media Services



Security, Identity & Compliance



Analytics



Machine Learning



Mobile Services



AR & VR



Application Integration



Customer Engagement



Business Productivity



Desktop & App Streaming



Internet of Things

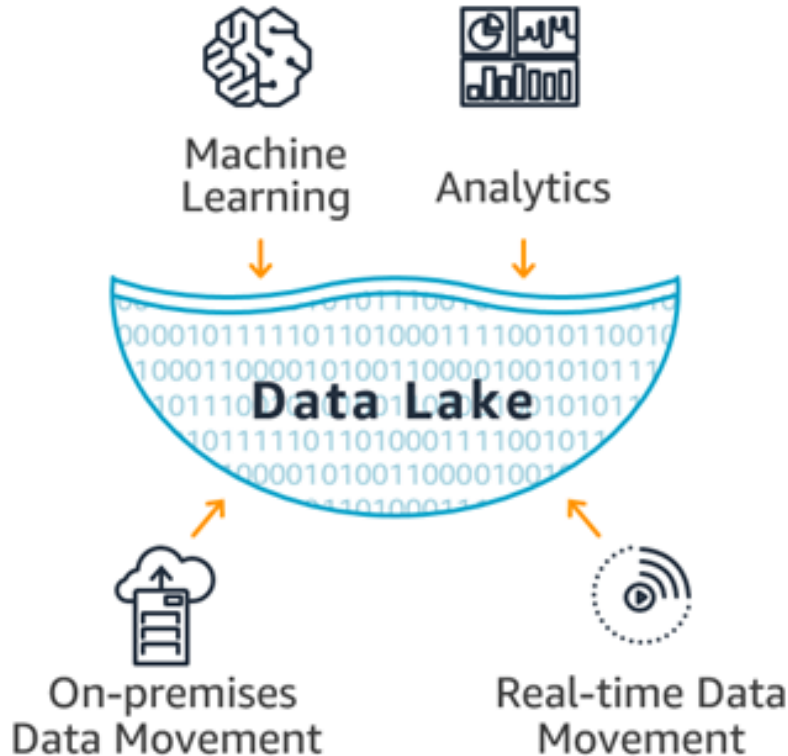


Game Development



AWS Cost Management

# Data Lakes and Analytics on AWS



## Data Movement

Import your data from on-premises, and in real-time.

## Data Lake

Store any type of data securely, from gigabytes to exabytes.

## Analytics

Analyze your data with a broad selection of analytic tools and engines.

## Machine Learning

Forecast future outcomes, and prescribe actions.



# AWS Products

## Analytics

- **Amazon Athena**
  - Query data in S3 using SQL
- **Amazon CloudSearch**
  - Managed search service
- **Amazon EMR**
  - Hosted Hadoop framework
- **Amazon Elasticsearch Service**
  - Run and scale Elasticsearch clusters
- **Amazon Kinesis**
  - Analyze real-time video and data streams
- **Amazon Redshift**
  - Fast, simple, cost-effective data warehousing
- **Amazon QuickSight**
  - Fast business analytics service
- **AWS Data Pipeline**
  - Orchestration service for periodic, data-driven workflows
- **AWS Glue**
  - Prepare and load data



# Machine Learning on AWS

## Machine learning in the hands of every developer and data scientist



### Build

Connect to other AWS services and transform data in SageMaker notebooks



### Train

Use SageMaker's algorithms and frameworks, or bring your own, for distributed training



### Tune

SageMaker automatically tunes your model by adjusting multiple combinations of algorithm parameters



### Deploy

Once training is completed, models can be deployed to SageMaker endpoints, for real-time predictions



# Cloud Computing

## AWS Cloud Practitioner

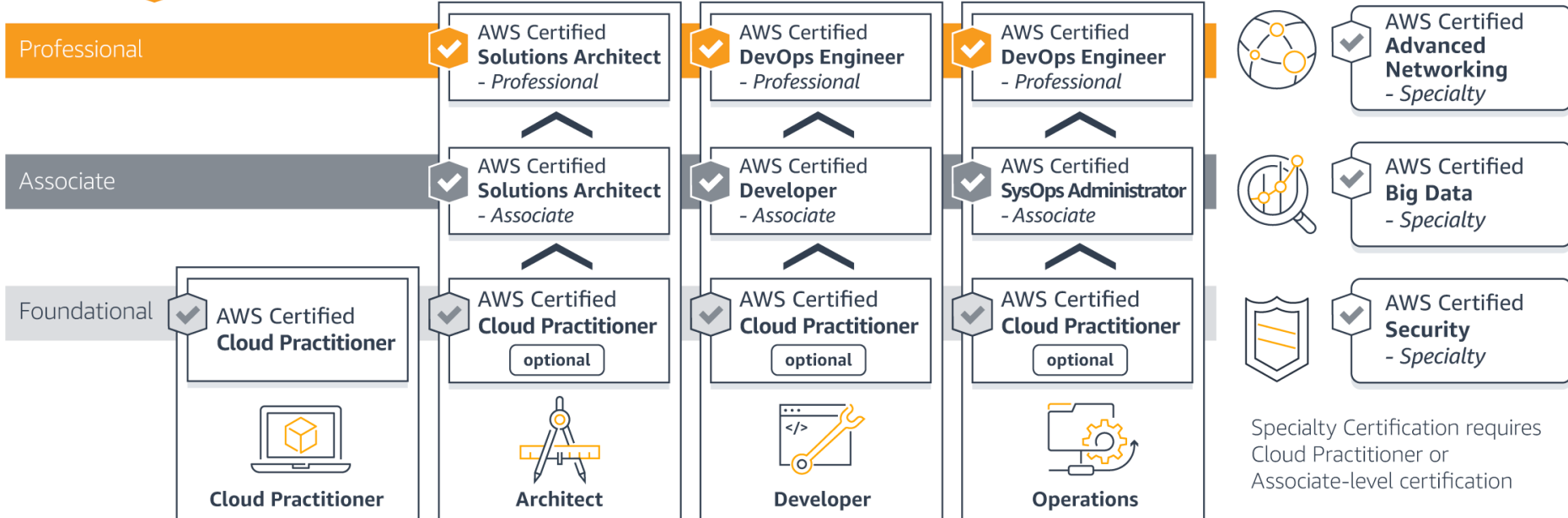
## AWS Solutions Architect

## AWS Certified Big Data Specialty

### aws CERTIFIED

#### Role-Based Certifications

#### Specialty Certifications



# Summary

- AI
- Big Data
- Cloud Computing

# References

- Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson.
- Jared Dean (2014), Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, Wiley.
- Mehmet Kaya, Jalal Kawash, Suheil Khoury, and Min-Yuh Day (2018), Social Network Based Big Data Analysis and Applications, Lecture Notes in Social Networks, Springer International Publishing.
- Varun Grover, Roger HL Chiang, Ting-Peng Liang, and Dongsong Zhang (2018), "Creating Strategic Business Value from Big Data Analytics: A Research Framework", Journal of Management Information Systems, 35, no. 2, pp. 388-423.
- Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, 111, no. 30, pp. 2-10.
- Stuart Russell and Peter Norvig (2016) , Artificial Intelligence: A Modern Approach, 3rd Edition, Pearson International.
- Javier Mata, Ignacio de Miguel, Ramón J. Durán, Noemí Merayo, Sandeep Kumar Singh, Admela Jukan, and Mohit Chamania (2018), "Artificial intelligence (AI) methods in optical networks: A comprehensive survey", Optical Switching and Networking, 28, pp. 43-57
- Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications.