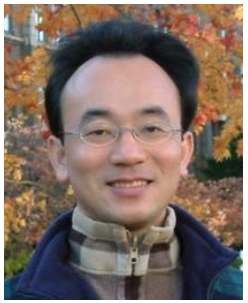
Tamkang University
淡江大學

Big Data Mining
巨量資料探勘

Tamkang University

Course Orientation for Big Data Mining
(巨量資料探勘課程介紹)

1042DM01
MI4 (M2244) (3094)
Tue, 3, 4 (10:10-12:00) (B216)



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/

2016-02-16

淡江大學104學年度第2學期
課程教學計畫表
Spring 2016 (2016.02 - 2016.06)

- 課程名稱：巨量資料探勘 (Big Data Mining)
- 授課教師：戴敏育 (Min-Yuh Day)
- 開課系級：資管四P (TLMXB4P) (M2244) (3094)
- 開課資料：選修 單學期 2 學分 (2 Credits, Elective)
- 上課時間：週二 3,4 (Tue 10:10-12:00)
- 上課教室： B216

課程簡介

- 本課程介紹巨量資料探勘 (Big Data Mining) 的基礎概念及應用技術。
- 課程內容包括
  - 巨量資料探勘 (Big Data Mining)
  - 巨量資料基礎：MapReduce典範、Hadoop與Spark生態系統 (Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem)
  - 關連分析 (Association Analysis)
  - 分類與預測 (Classification and Prediction)
  - 分群分析 (Cluster Analysis)
  - SAS企業資料採礦實務 (SAS EM)
  - 巨量資料探勘個案分析與實作
  - Google TensorFlow 深度學習 (Deep Learning with Google TensorFlow)

Course Introduction

- This course introduces the fundamental concepts and applications technology of big data mining.
- Topics include
  - Big Data Mining
  - Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem
  - Association Analysis
  - Classification and Prediction
  - Cluster Analysis
  - Data Mining Using SAS Enterprise Miner (SAS EM)
  - Case Study and Implementation of Big Data Mining
  - Deep Learning with Google TensorFlow

課程目標
(Objective)

- 瞭解及應用巨量資料探勘基本概念與技術。

- Understand and apply the fundamental concepts and technology of big data mining

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2016/02/16	巨量資料探勘課程介紹 (Course Orientation for Big Data Mining)
2	2016/02/23	巨量資料基礎：MapReduce典範、Hadoop與Spark生態系統 (Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem)
3	2016/03/01	關連分析 (Association Analysis)
4	2016/03/08	分類與預測 (Classification and Prediction)
5	2016/03/15	分群分析 (Cluster Analysis)
6	2016/03/22	個案分析與實作一 (SAS EM 分群分析)：Case Study 1 (Cluster Analysis – K-Means using SAS EM)
7	2016/03/29	個案分析與實作二 (SAS EM 關連分析)：Case Study 2 (Association Analysis using SAS EM)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
8	2016/04/05	教學行政觀摩日 (Off-campus study)
9	2016/04/12	期中報告 (Midterm Project Presentation)
10	2016/04/19	期中考試週 (Midterm Exam)
11	2016/04/26	個案分析與實作三 (SAS EM 決策樹、模型評估)：Case Study 3 (Decision Tree, Model Evaluation using SAS EM)
12	2016/05/03	個案分析與實作四 (SAS EM 迴歸分析、類神經網路)：Case Study 4 (Regression Analysis, Artificial Neural Network using SAS EM)
13	2016/05/10	Google TensorFlow 深度學習 (Deep Learning with Google TensorFlow)
14	2016/05/17	期末報告 (Final Project Presentation)
15	2016/05/24	畢業班考試 (Final Exam)

教學方法與評量方法

- 教學方法
  - 講述、討論、賞析、模擬、實作、問題解決
- 評量方法
  - 紙筆測驗、實作、報告、上課表現

教材課本

- 教材課本
  - 講義 (Slides)
  - 資料採礦運用：以SAS Enterprise Miner為工具，李淑娟，2015，SAS賽仕電腦軟體
- 參考書籍
  - Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, Jared Dean, Wiley, 2014
  - Data Science for Business: What you need to know about data mining and data-analytic thinking, Foster Provost and Tom Fawcett, O'Reilly, 2013
  - Applied Analytics Using SAS Enterprise Mining, Jim Georges, Jeff Thompson and Chip Wells, SAS, 2010
  - Data Mining: Concepts and Techniques, Third Edition, Jiawei Han, Micheline Kamber and Jian Pei, Morgan Kaufmann, 2011

作業與學期成績計算方式

- 作業篇數
  - 3篇
- 學期成績計算方式
  - ☑期中評量：30％
  - ☑期末評量：30 ％
  - ☑其他（課堂參與及報告討論表現）： 40 ％

Team Term Project

- Term Project Topics
  - Big Data mining
  - Web and Text mining
  - Business Intelligence
  - Big Data Analytics
  - Social Computing
- 3-4 人為一組
  - 分組名單於 2016/02/23 (二) 課程下課時繳交
  - 由班代統一收集協調分組名單

2016/02/23

巨量資料基礎：MapReduce典範、Hadoop與Spark生態系統

(Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem)

2016/05/10

Google TensorFlow

深度學習

(Deep Learning

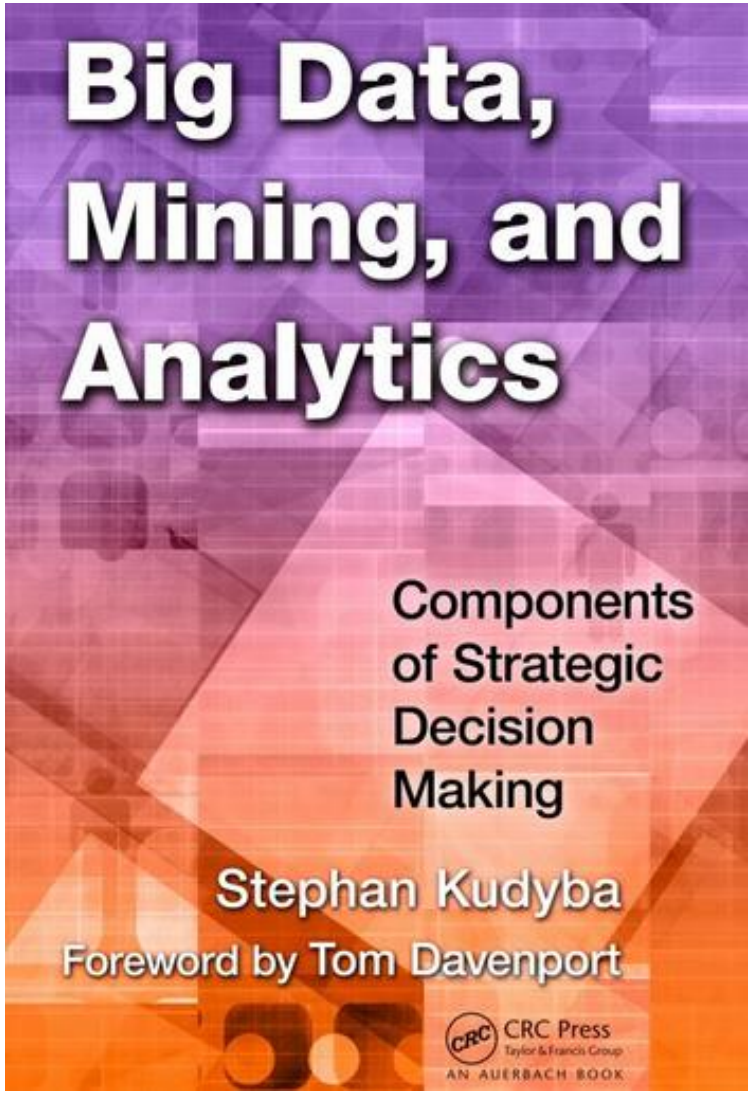
with

Google TensorFlow)

Big Data Analytics and Data Mining

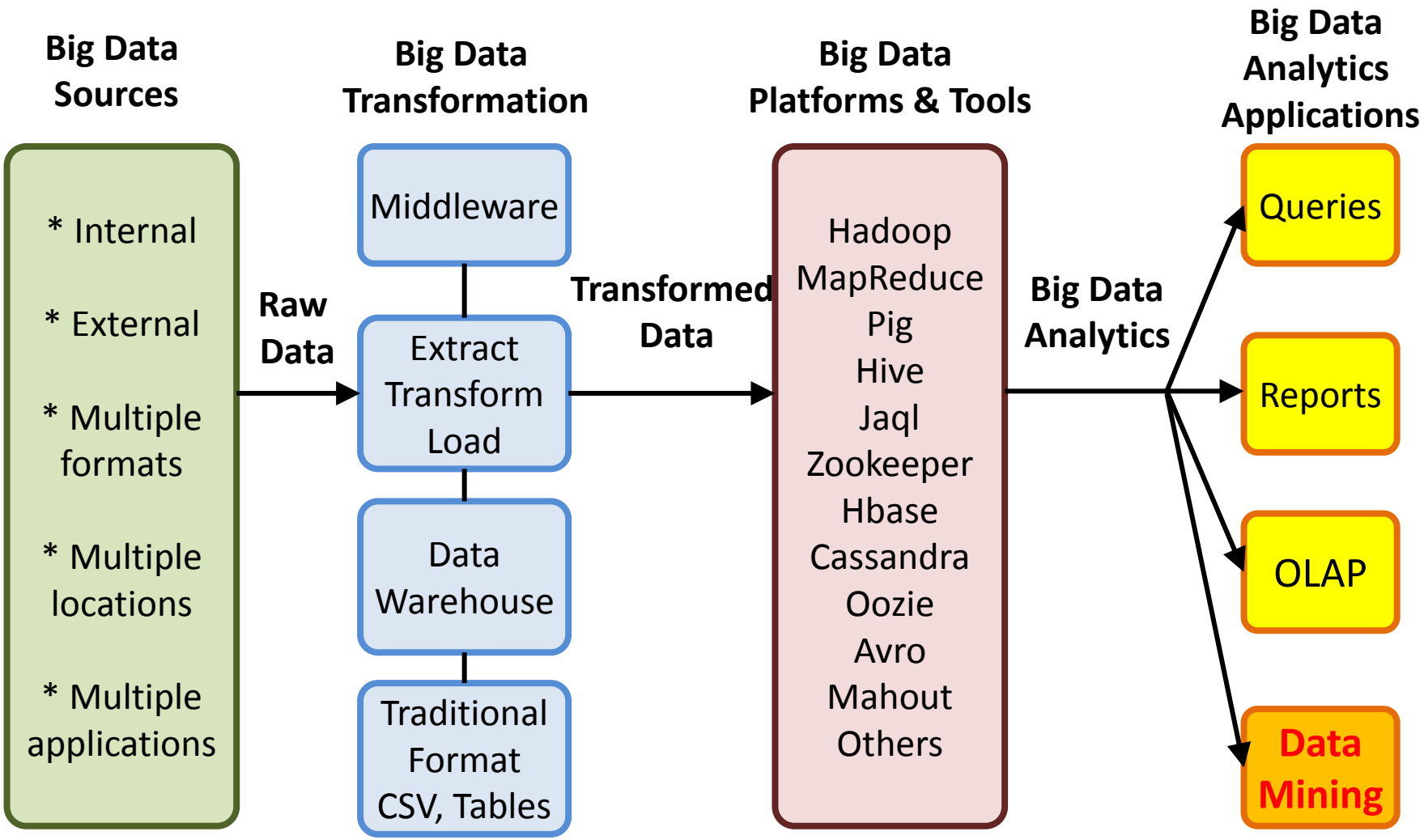
Stephan Kudyba (2014),

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



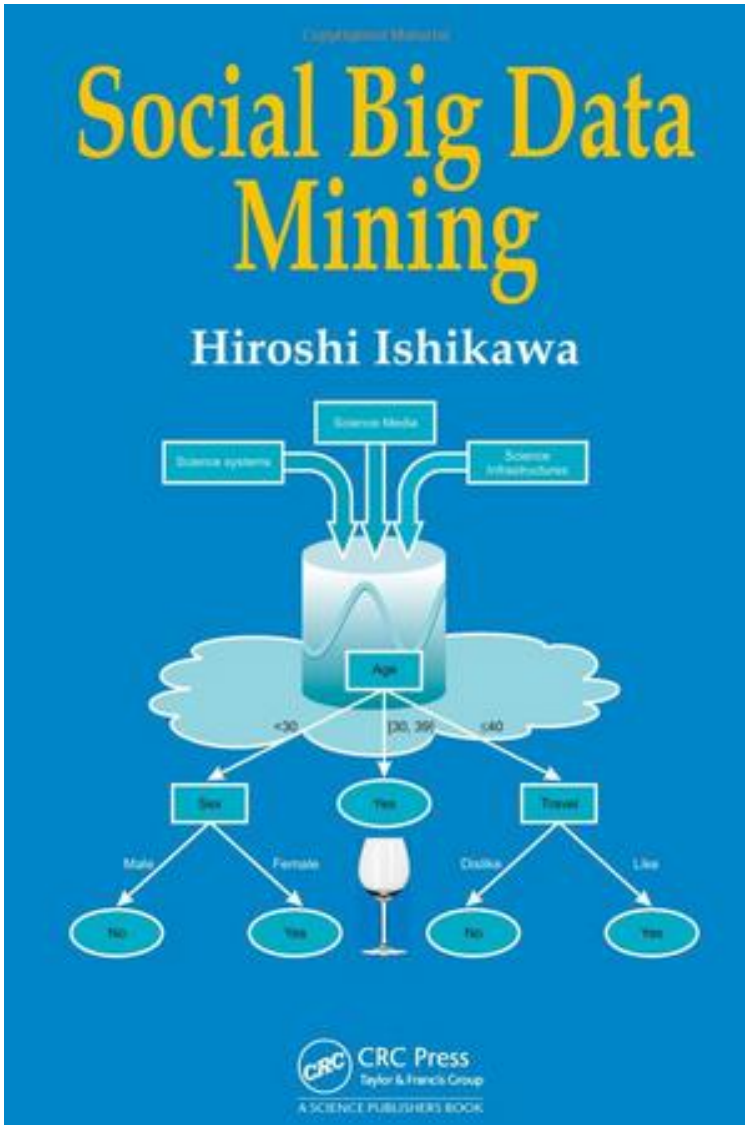
Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Architecture of Big Data Analytics

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

Social Big Data Mining

(Hiroshi Ishikawa, 2015)



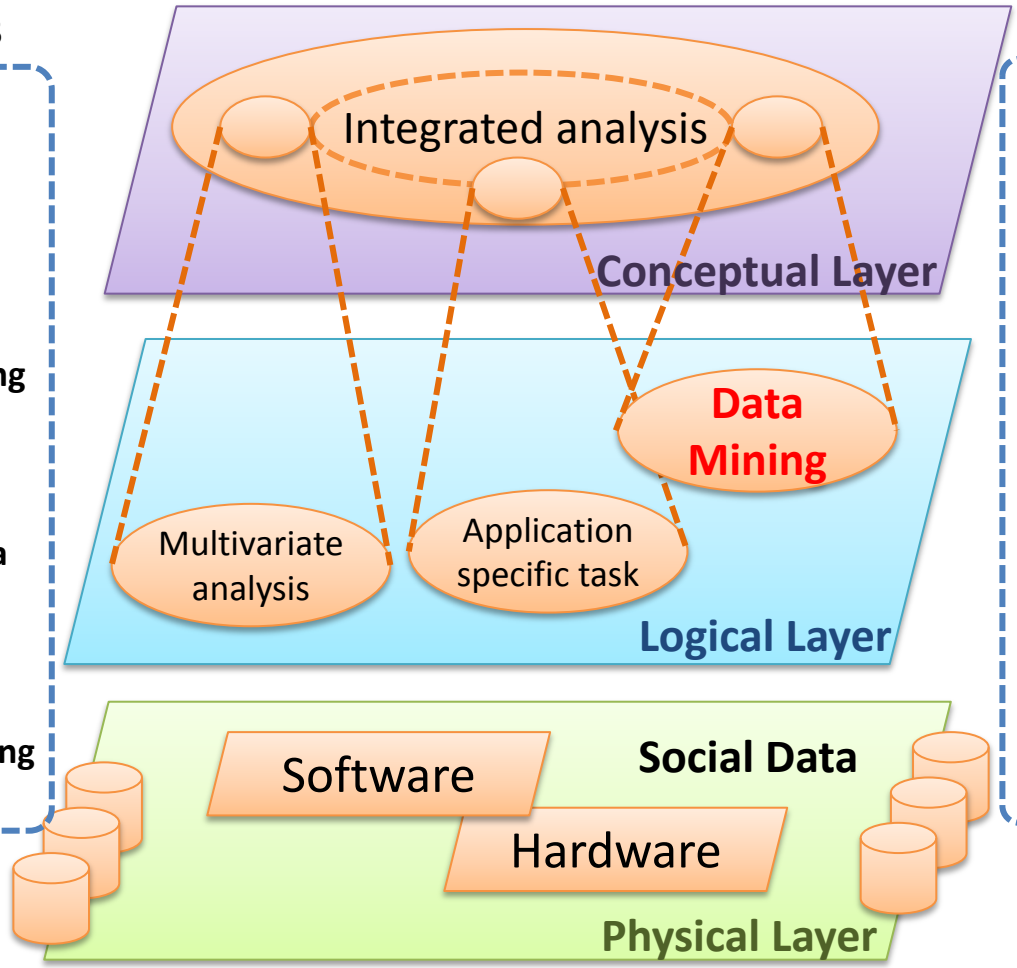
Source: http://www.amazon.com/Social-Data-Mining-Hiroshi-Ishikawa/dp/149871093X

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 distrusted processing

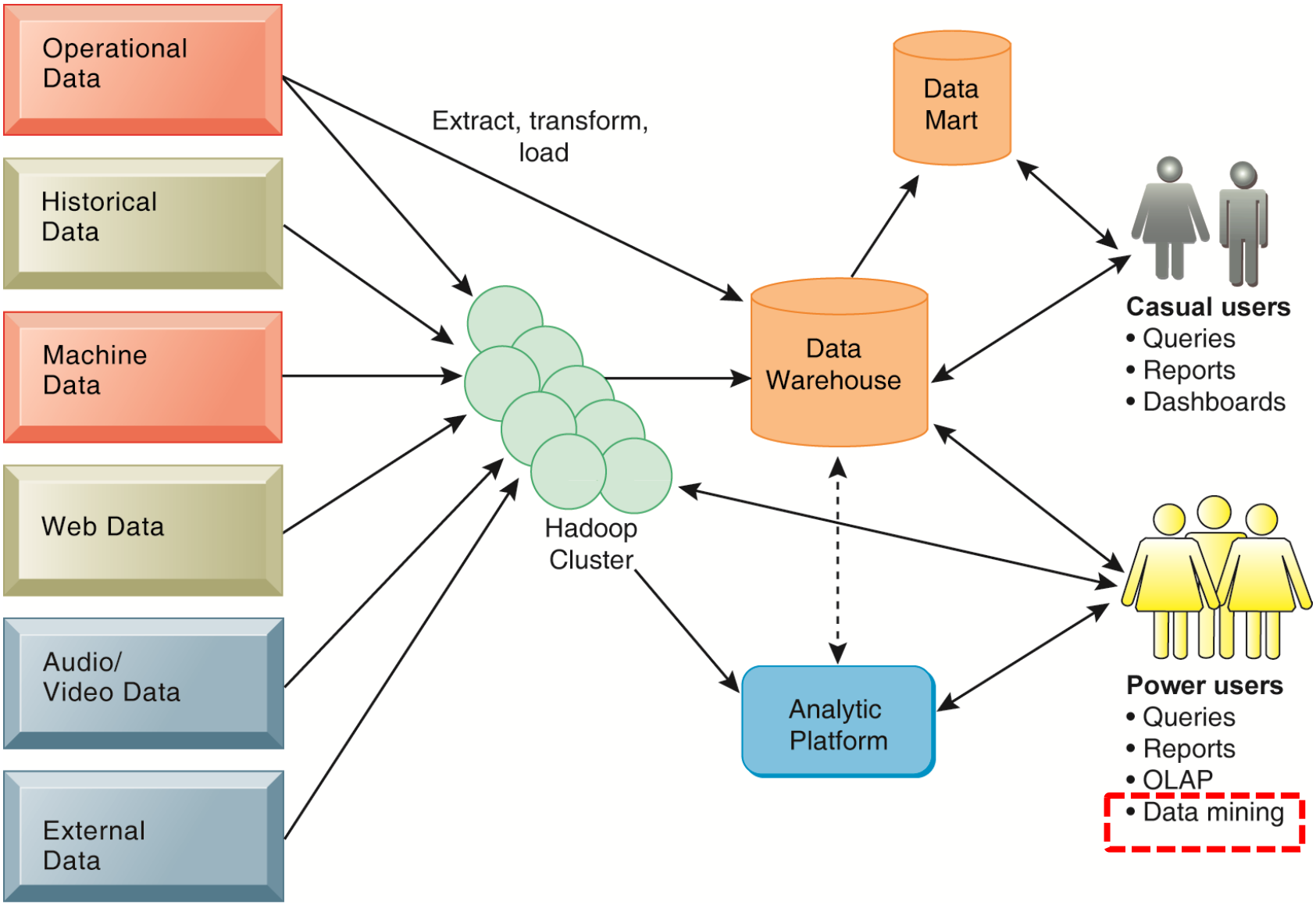


Analysts

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

Source: Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press

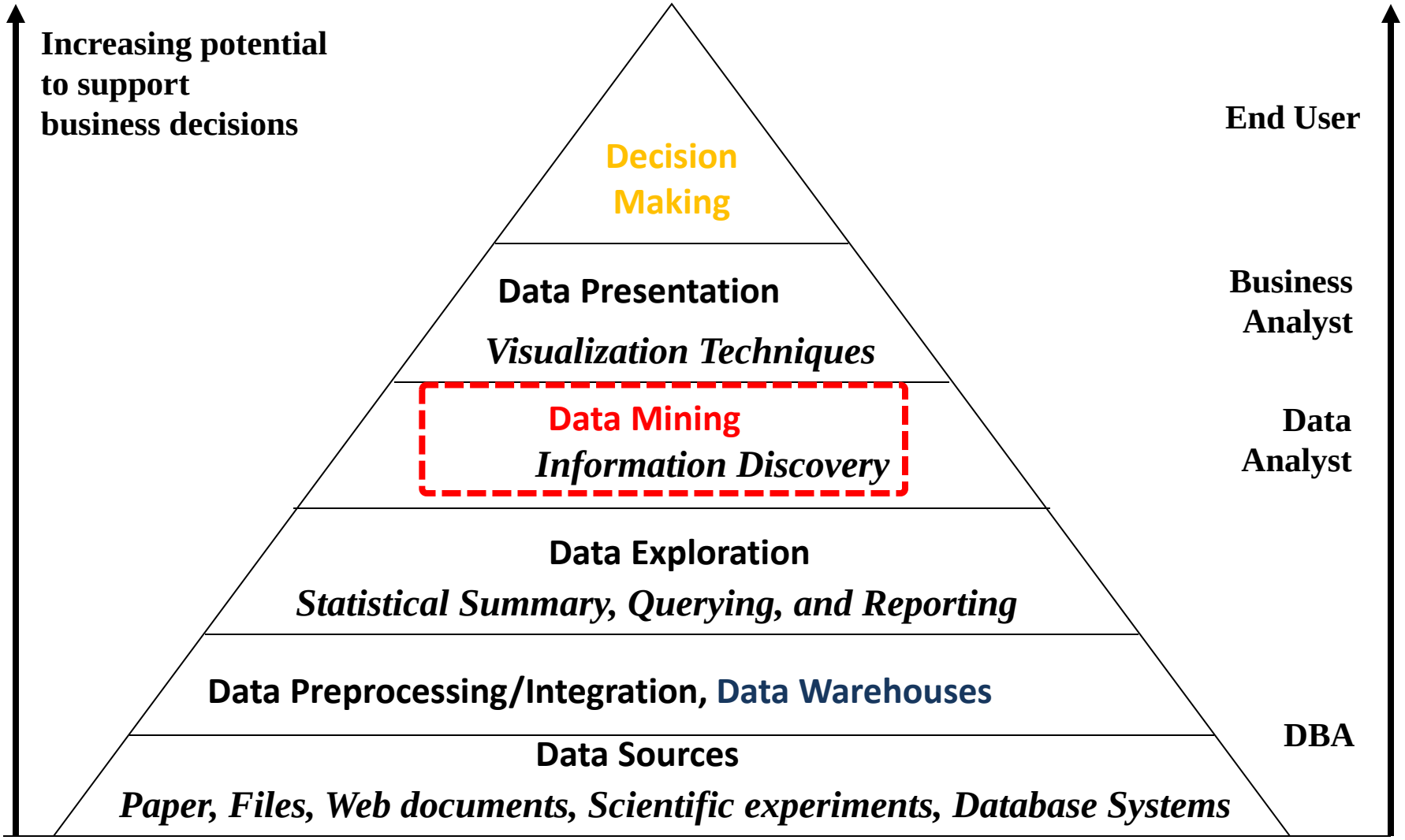
Business Intelligence (BI) Infrastructure



Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.

Data Warehouse

Data Mining and Business Intelligence

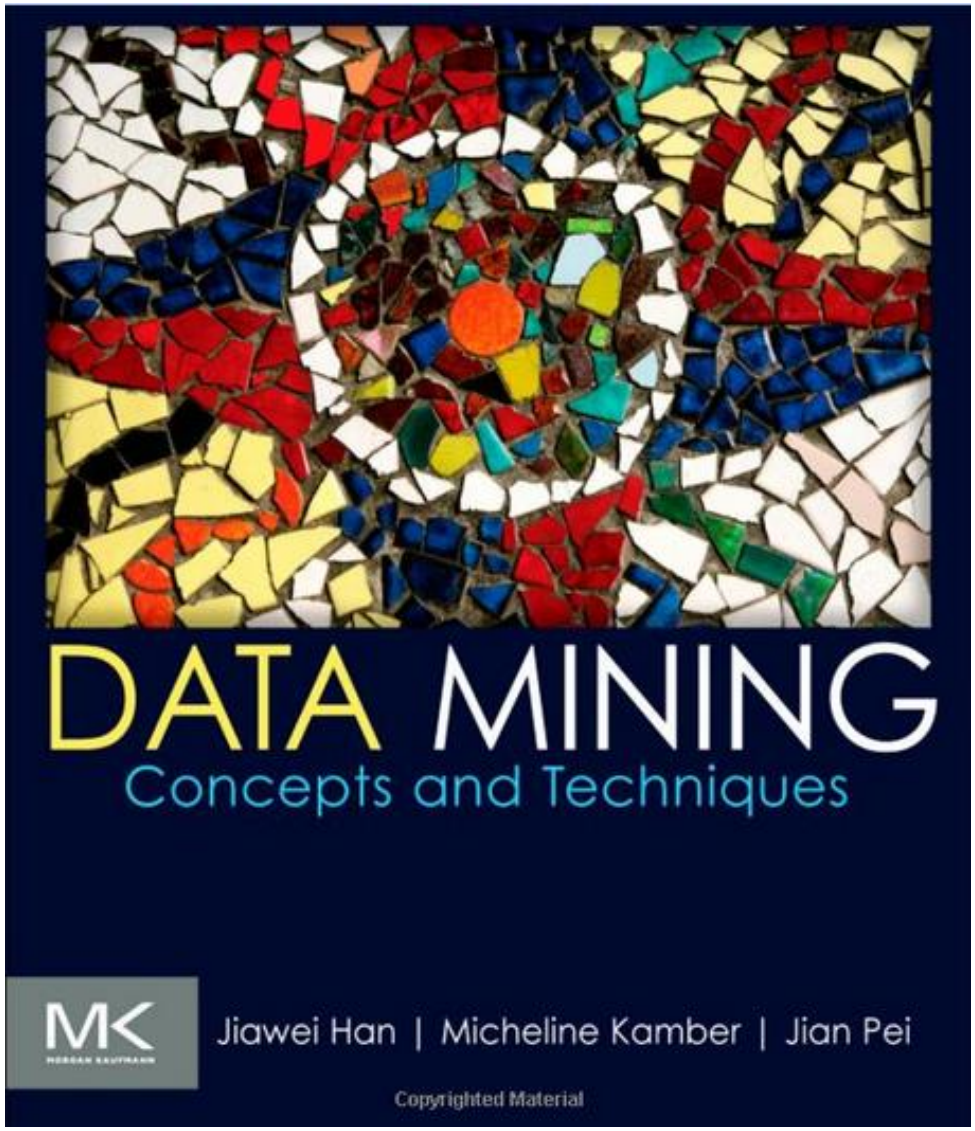


Source: Jiawei Han and Micheline Kamber (2006), Data Mining: Concepts and Techniques, Second Edition, Elsevier

The Evolution of BI Capabilities

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

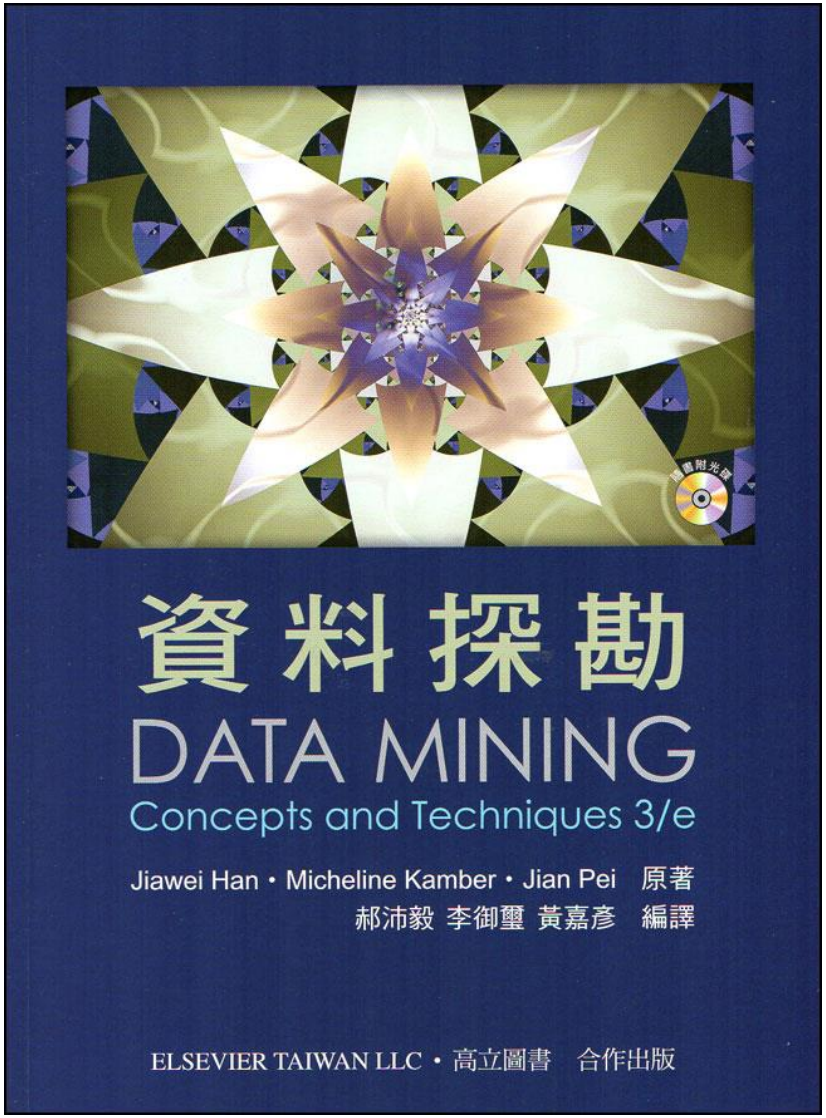
Data Mining



Source: http://www.amazon.com/Data-Mining-Concepts-Techniques-Management/dp/0123814790

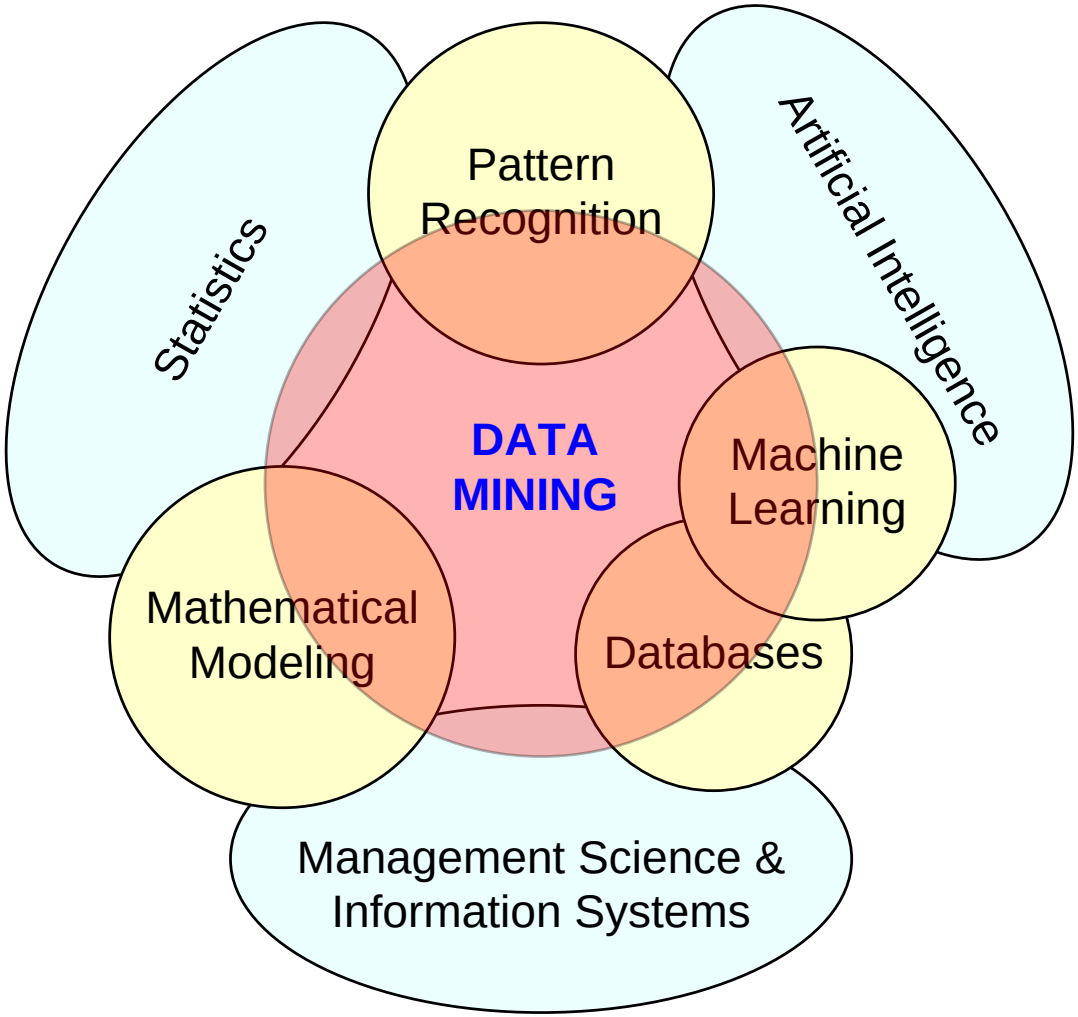
郝沛毅, 李御璽, 黃嘉彥 編譯, 資料探勘

(Jiawei Han, Micheline Kamber, Jian Pei, Data Mining - Concepts and Techniques 3/e), 高立圖書, 2014



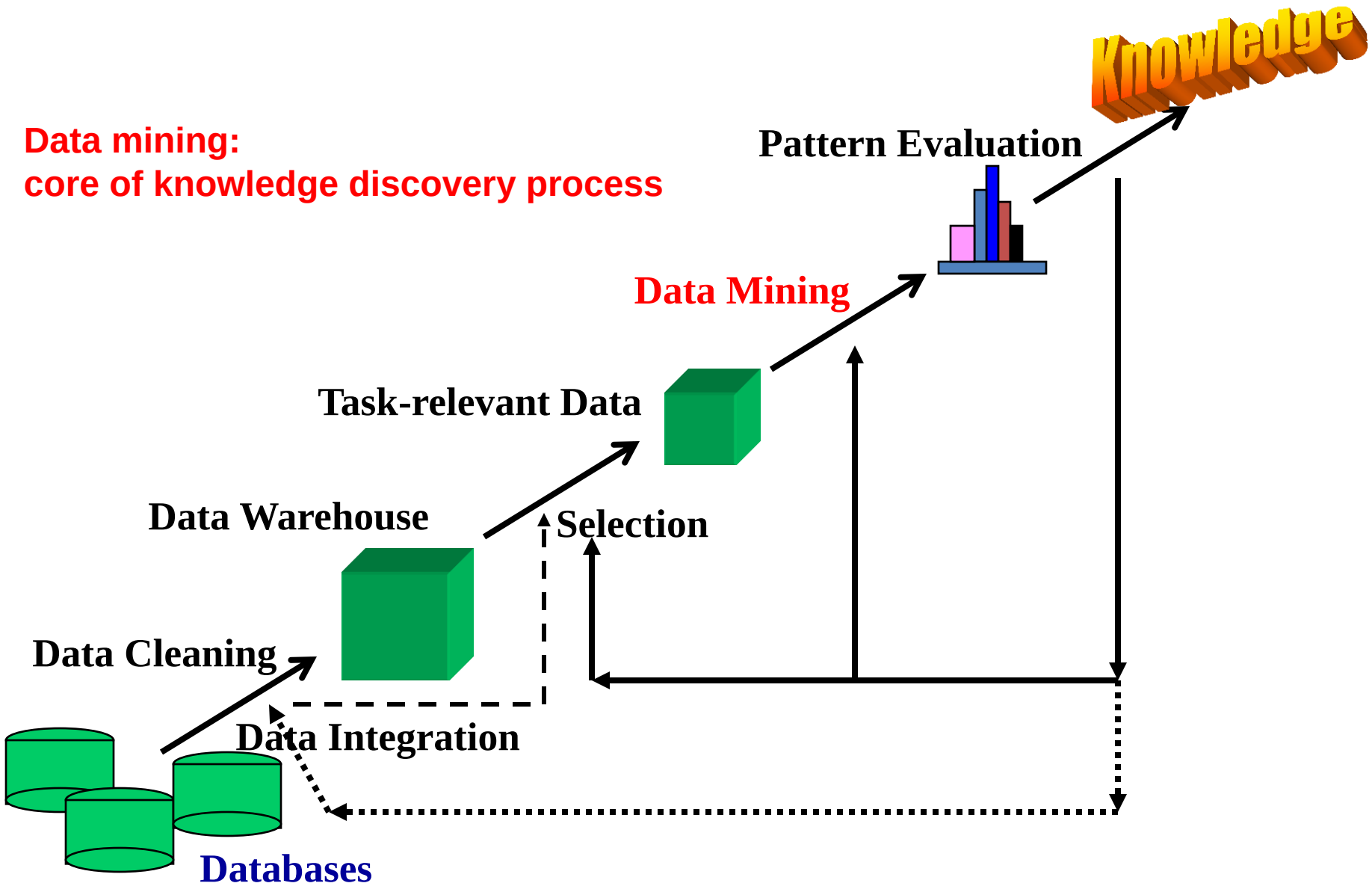
Source: http://www.books.com.tw/products/0010646676

Data Mining at the Intersection of Many Disciplines



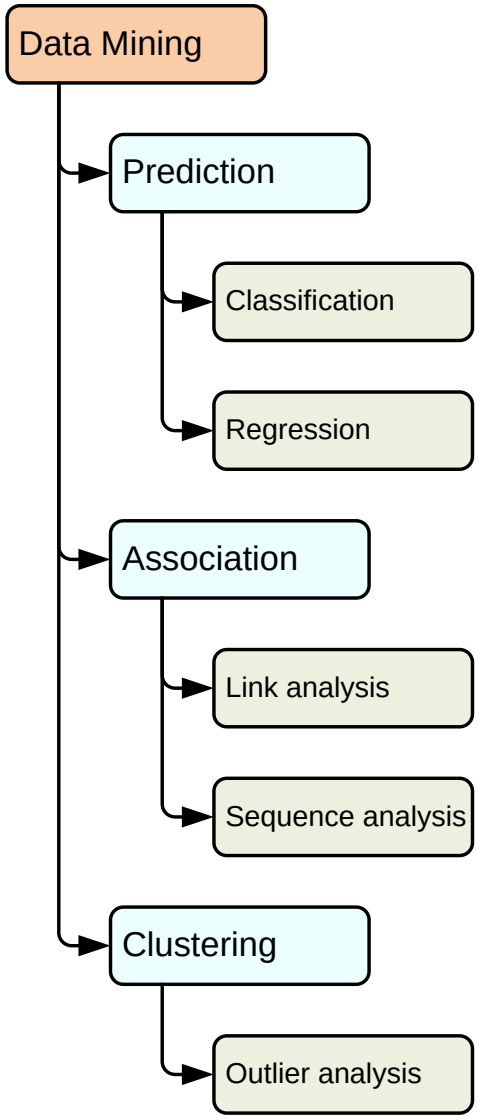
Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Knowledge Discovery (KDD) Process



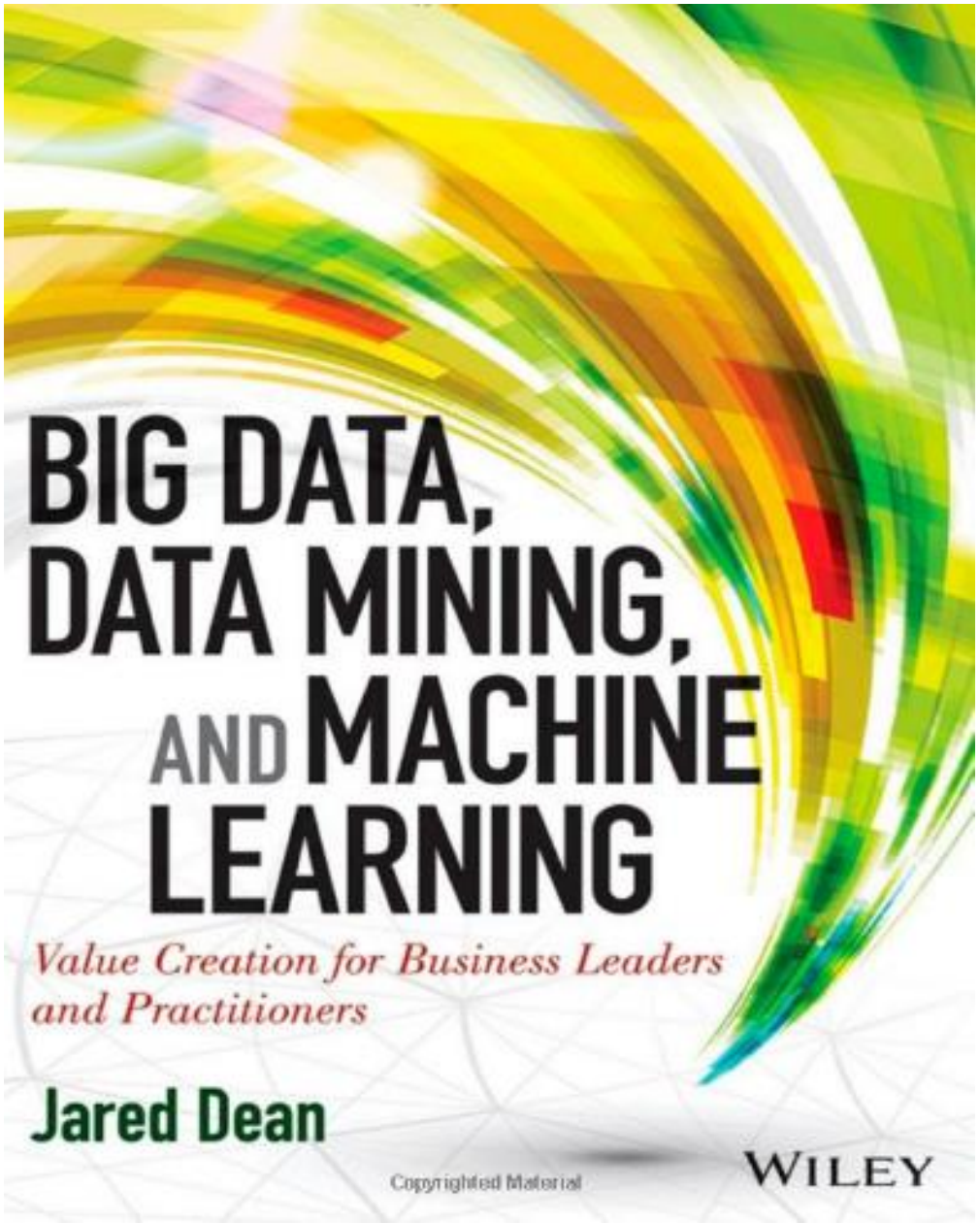
Source: Han & Kamber (2006)

A Taxonomy for Data Mining Tasks



Learning Method	Popular Algorithms
Supervised	Classification and Regression Trees, ANN, SVM, Genetic Algorithms
Supervised	Decision trees, ANN/MLP, SVM, Rough sets, Genetic Algorithms
Supervised	Linear/Nonlinear Regression, Regression trees, ANN/MLP, SVM
Unsupervised	Apriory, OneR, ZeroR, Eclat
Unsupervised	Expectation Maximization, Apriory Algorithm, Graph-based Matching
Unsupervised	Apriory Algorithm, FP-Growth technique
Unsupervised	K-means, ANN/SOM
Unsupervised	K-means, Expectation Maximization (EM)

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

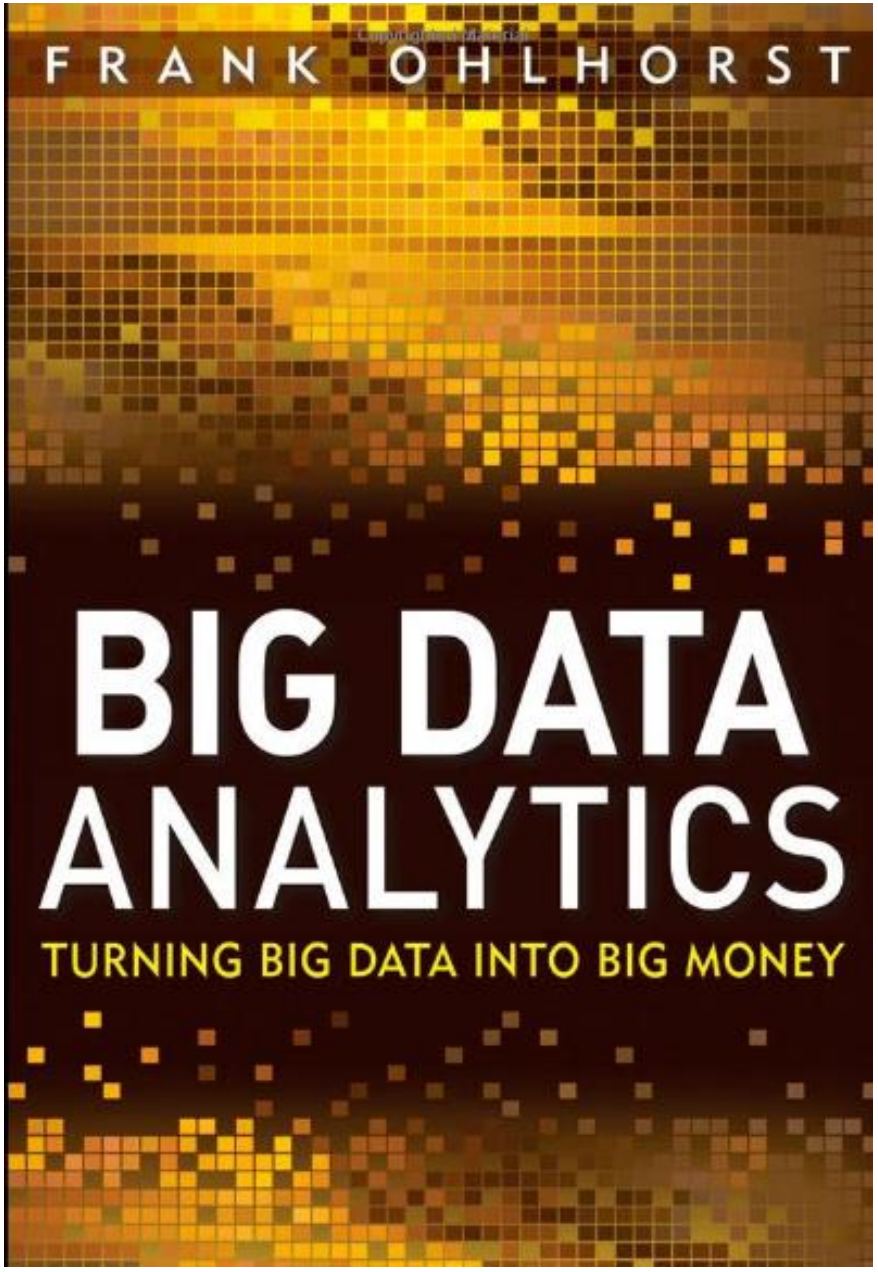


Source: http://www.amazon.com/Data-Mining-Machine-Learning-Practitioners/dp/1118618041

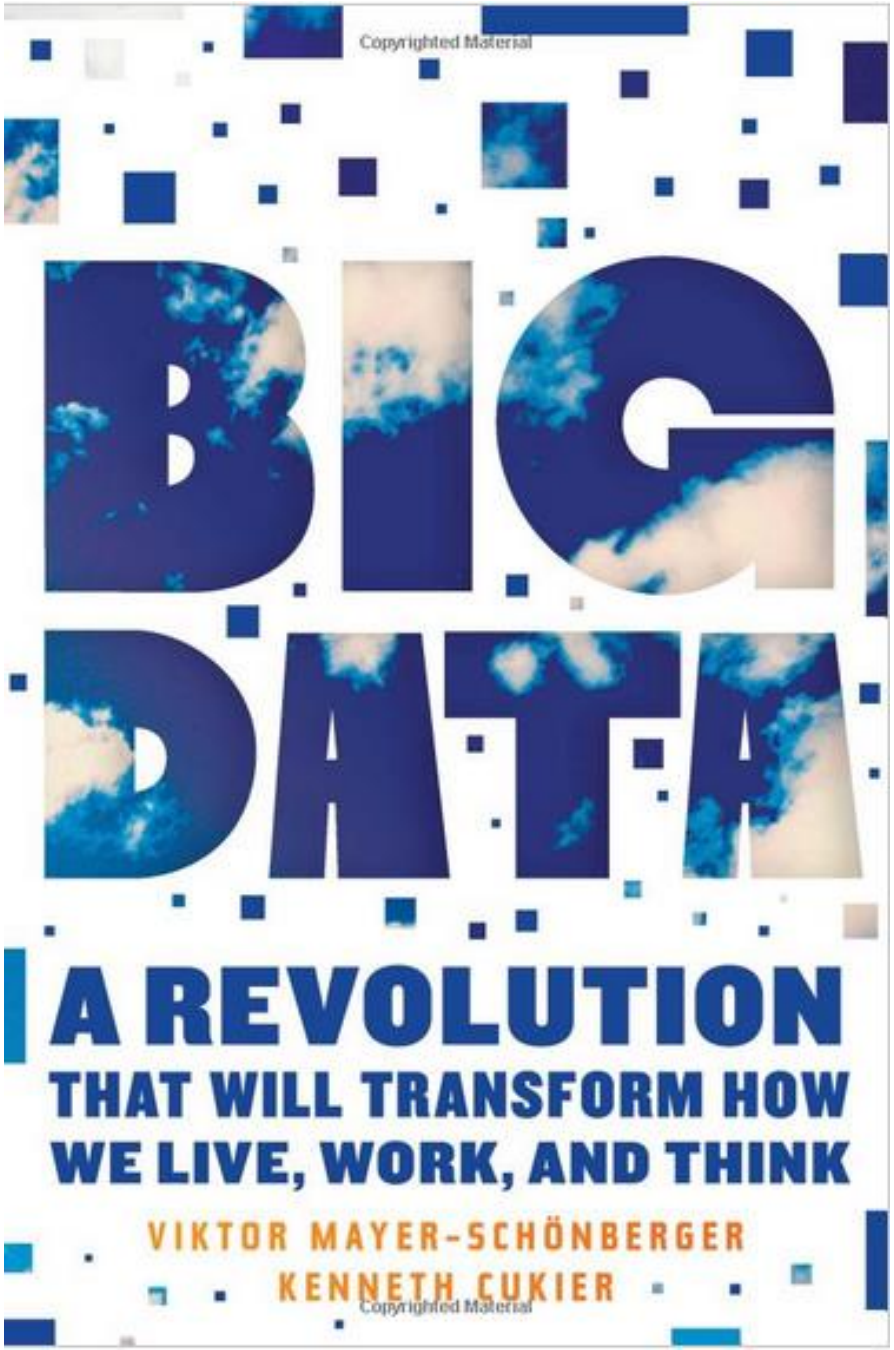
Deep Learning

Intelligence from Big Data

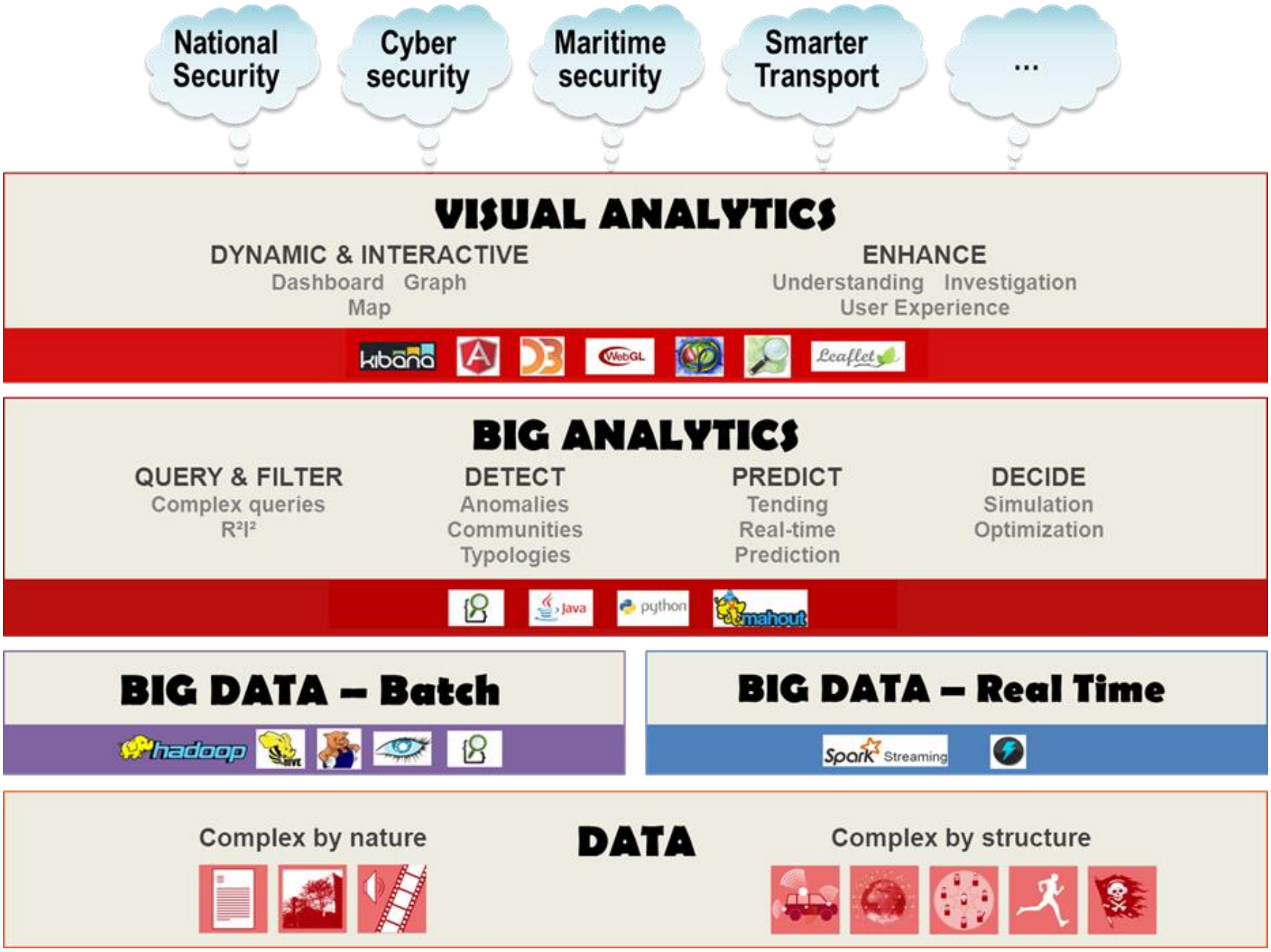
Source: https://www.vlab.org/events/deep-learning/



Source: http://www.amazon.com/Big-Data-Analytics-Turning-Money/dp/1118147596

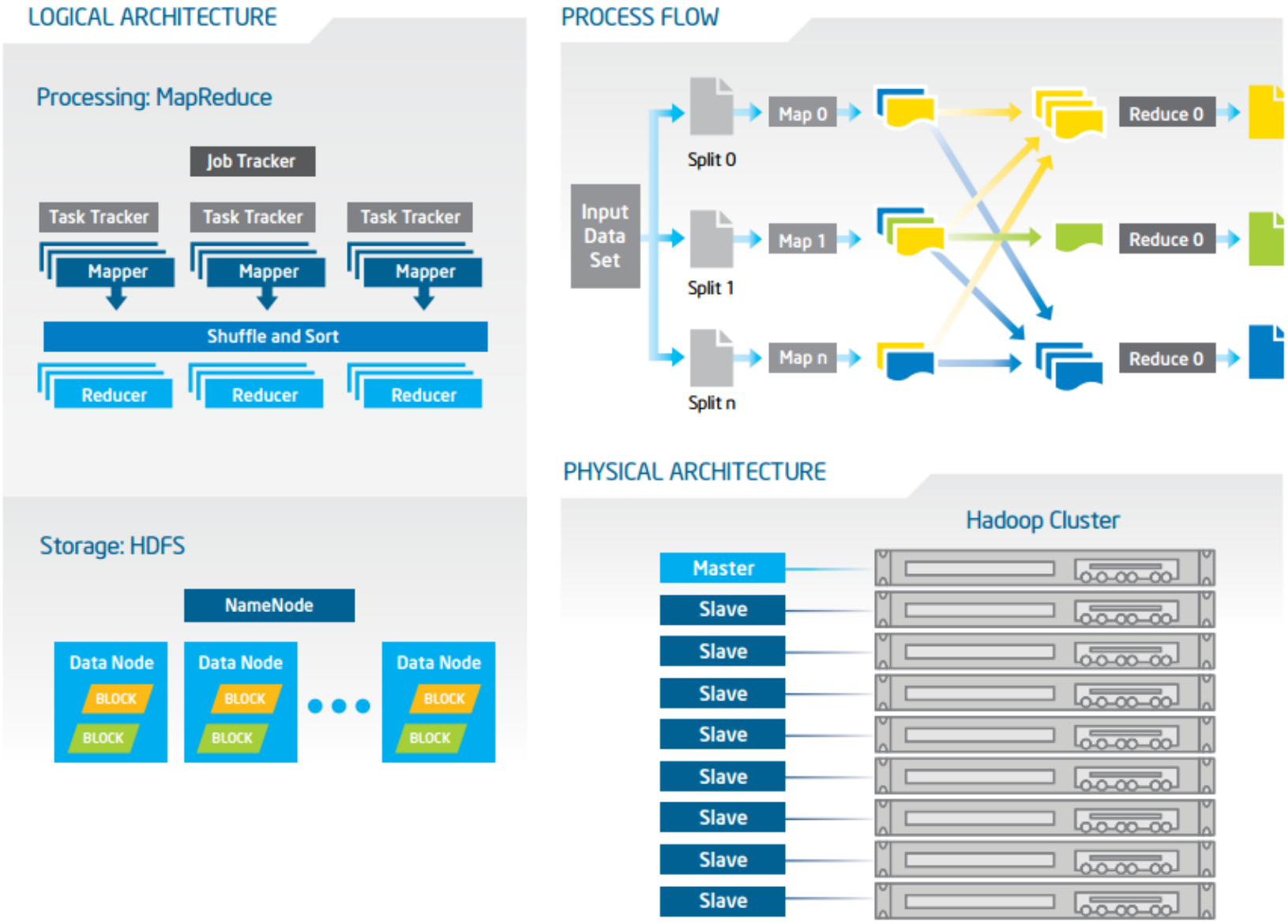


Source: http://www.amazon.com/Big-Data-Revolution-Transform-Mayer-Schonberger/dp/B00D81X2YE



Source: https://www.thalesgroup.com/en/worldwide/big-data/big-data-big-analytics-visual-analytics-what-does-it-all-mean

Big Data with Hadoop Architecture

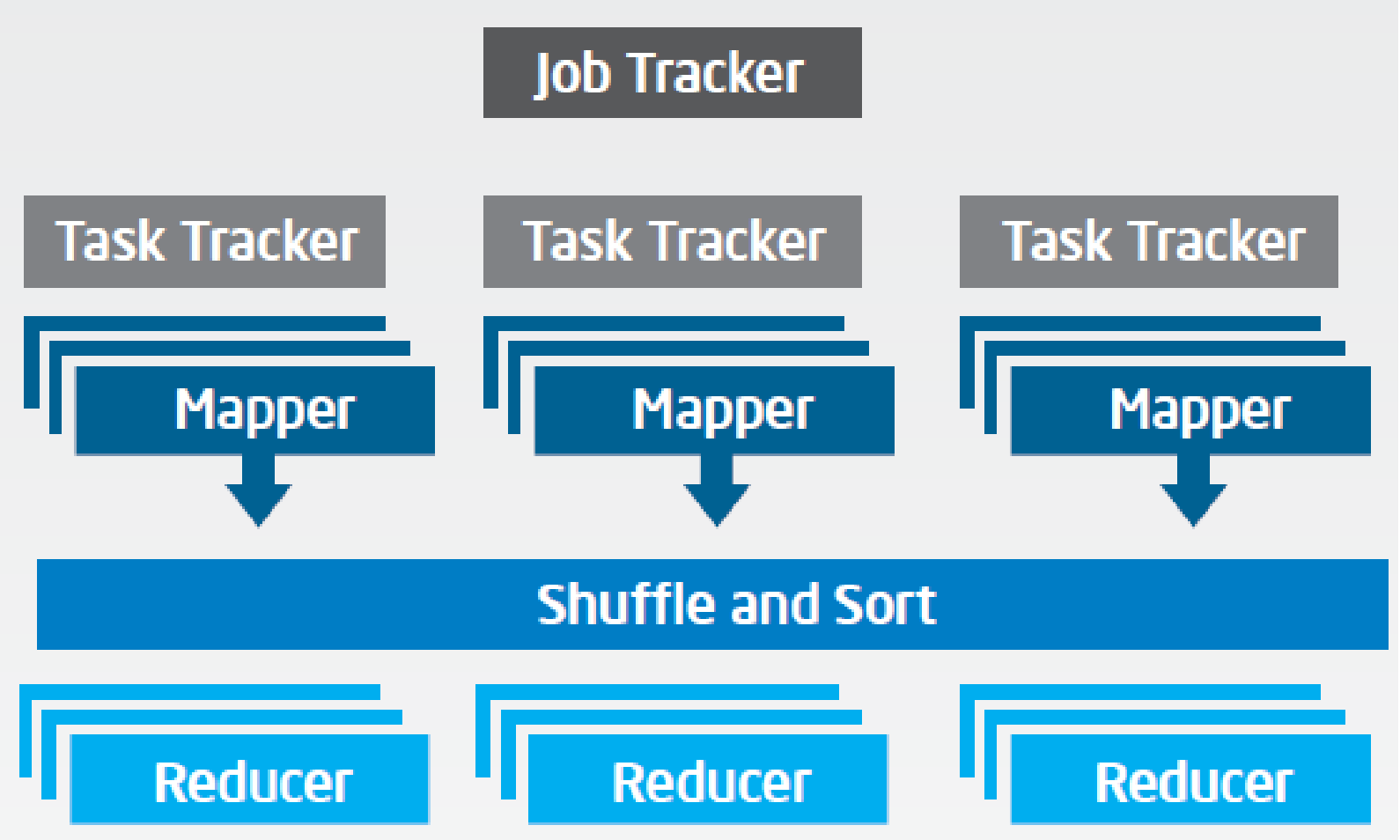


Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

Big Data with Hadoop Architecture

Logical Architecture

Processing: MapReduce

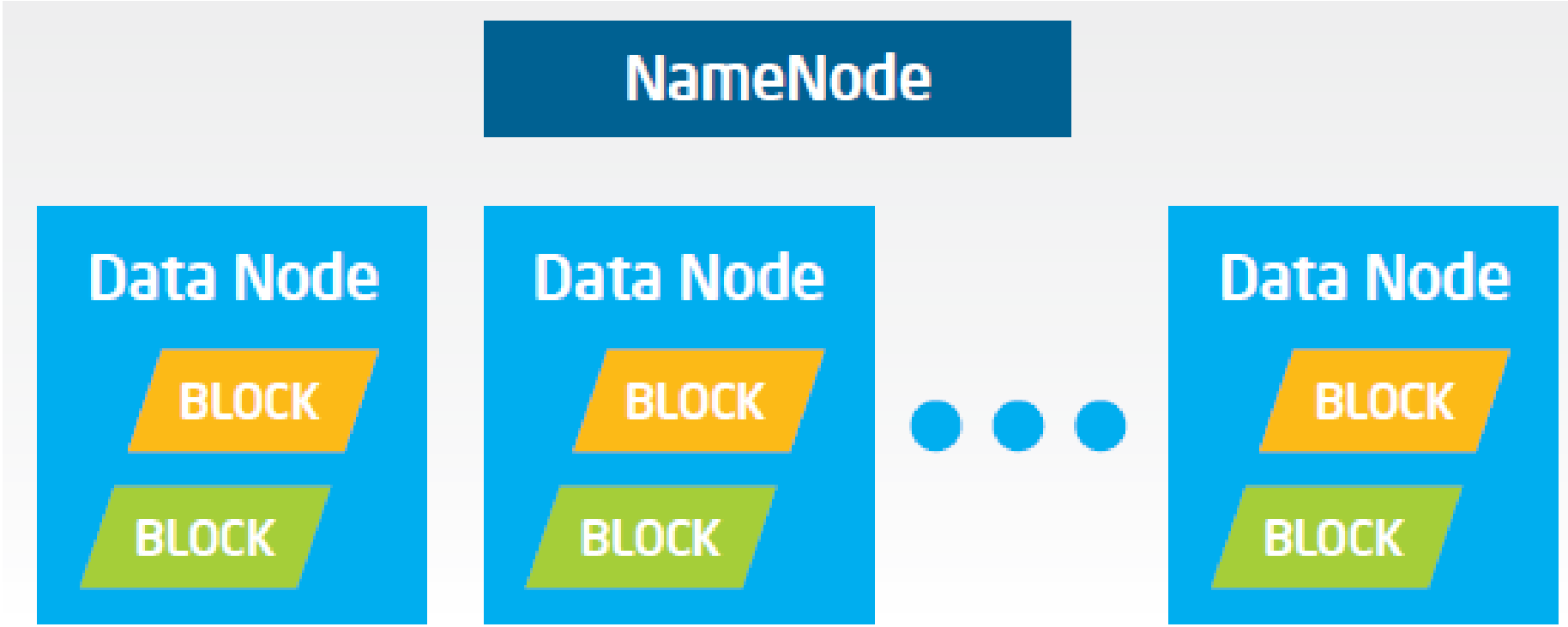


Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

Big Data with Hadoop Architecture

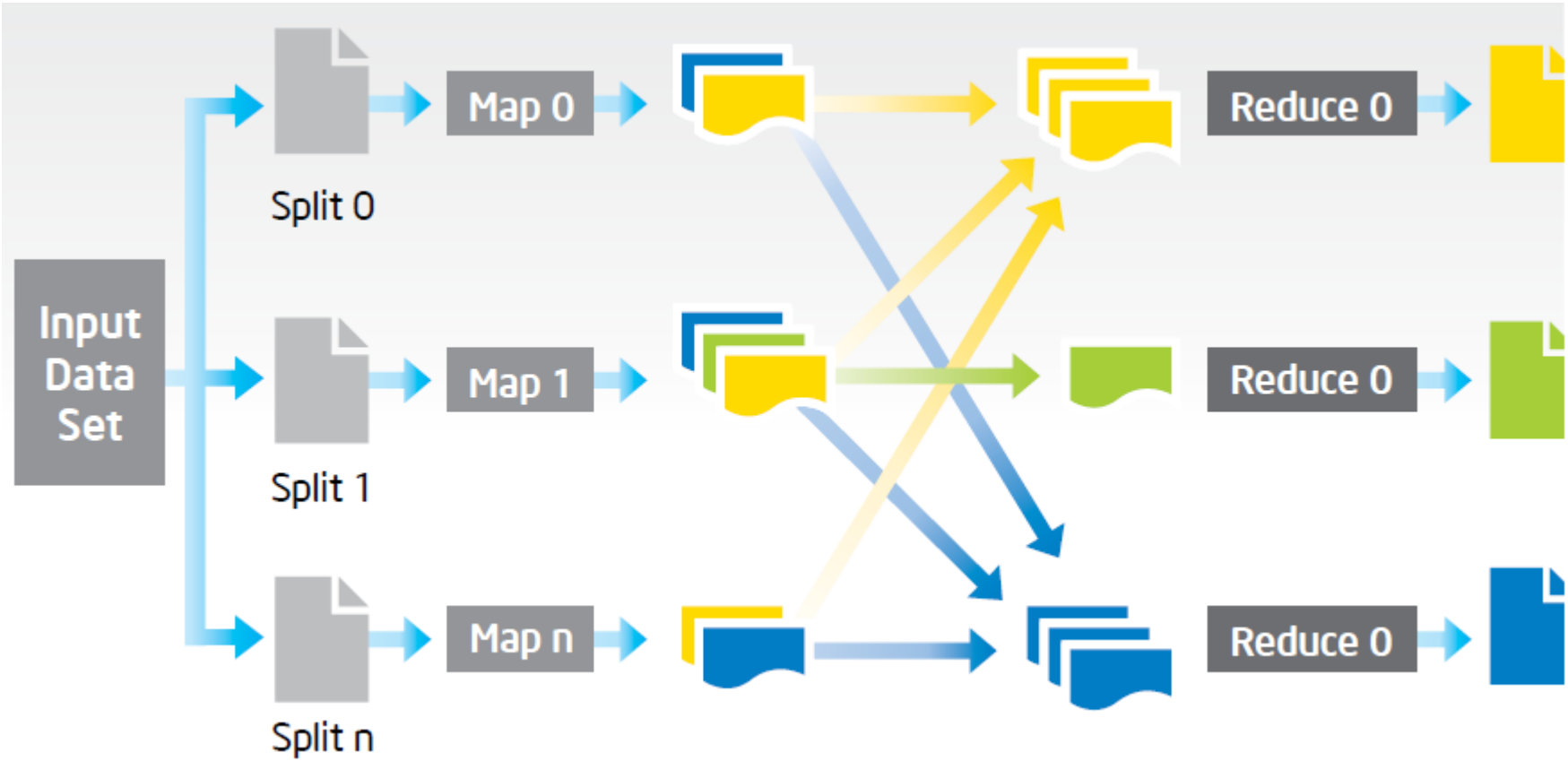
Logical Architecture

Storage: HDFS



Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

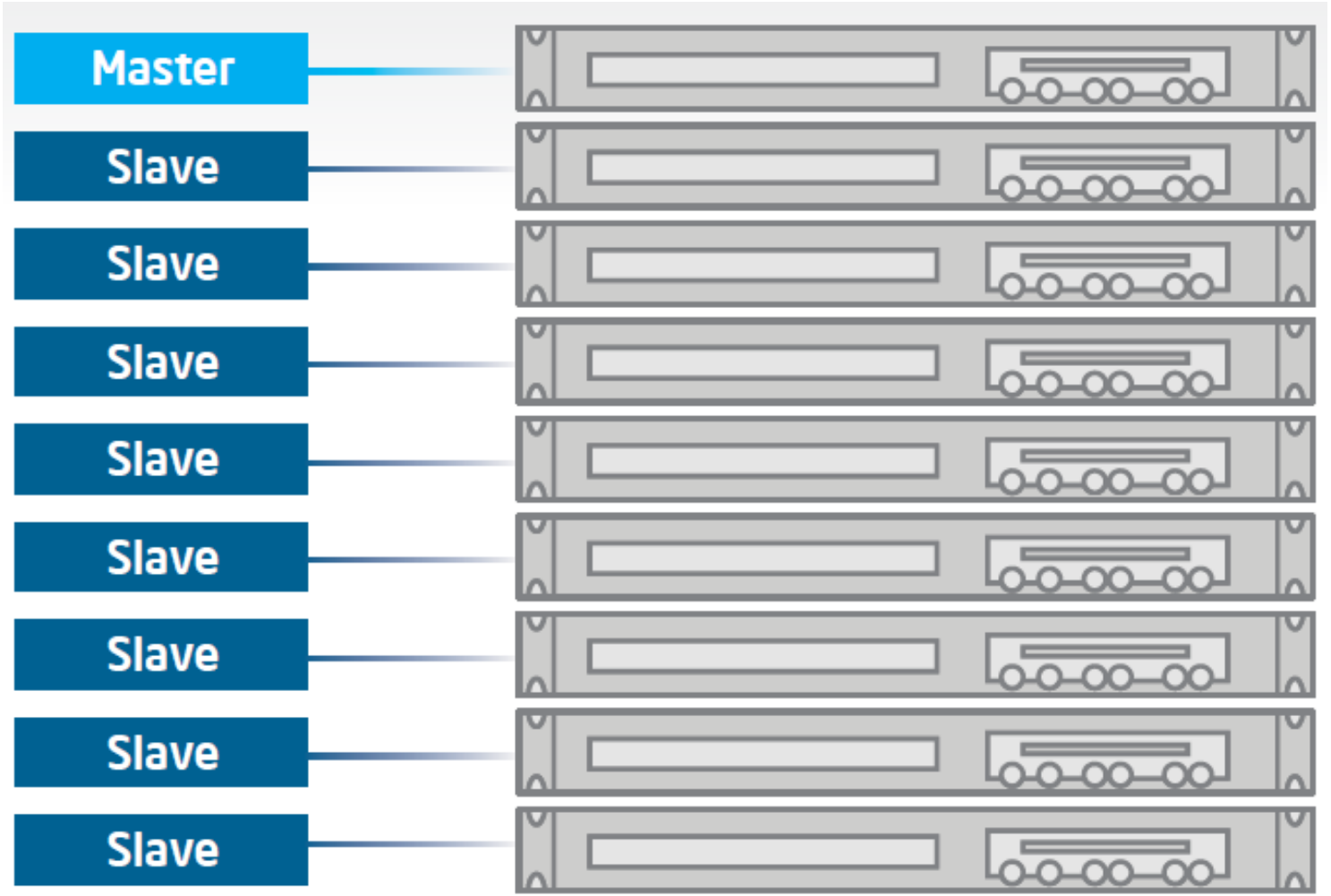
Big Data with Hadoop Architecture
Process Flow



Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

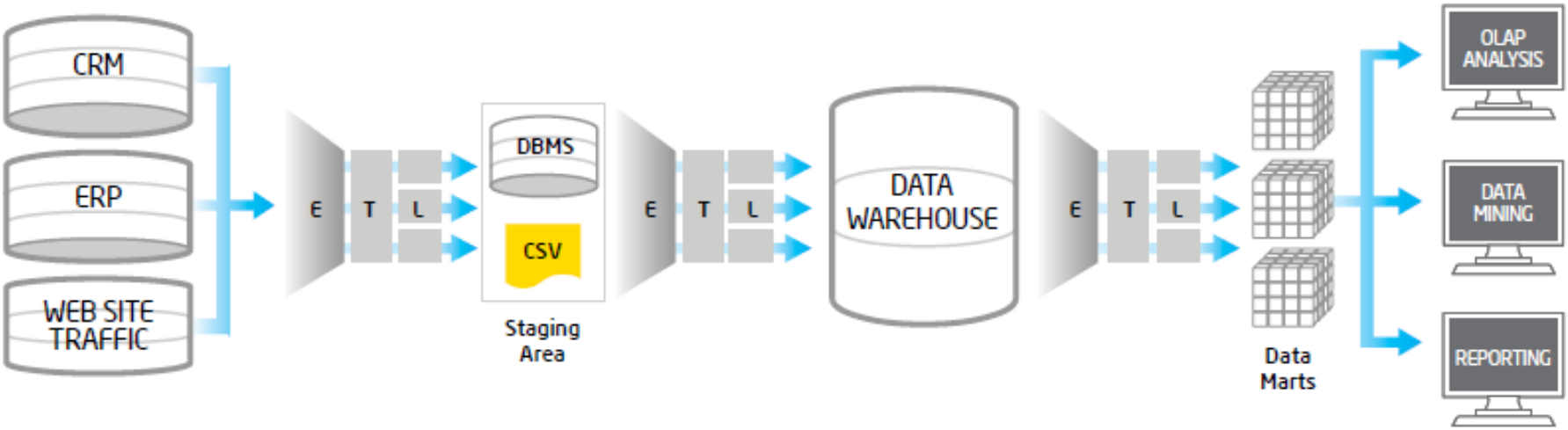
Big Data with Hadoop Architecture

Hadoop Cluster



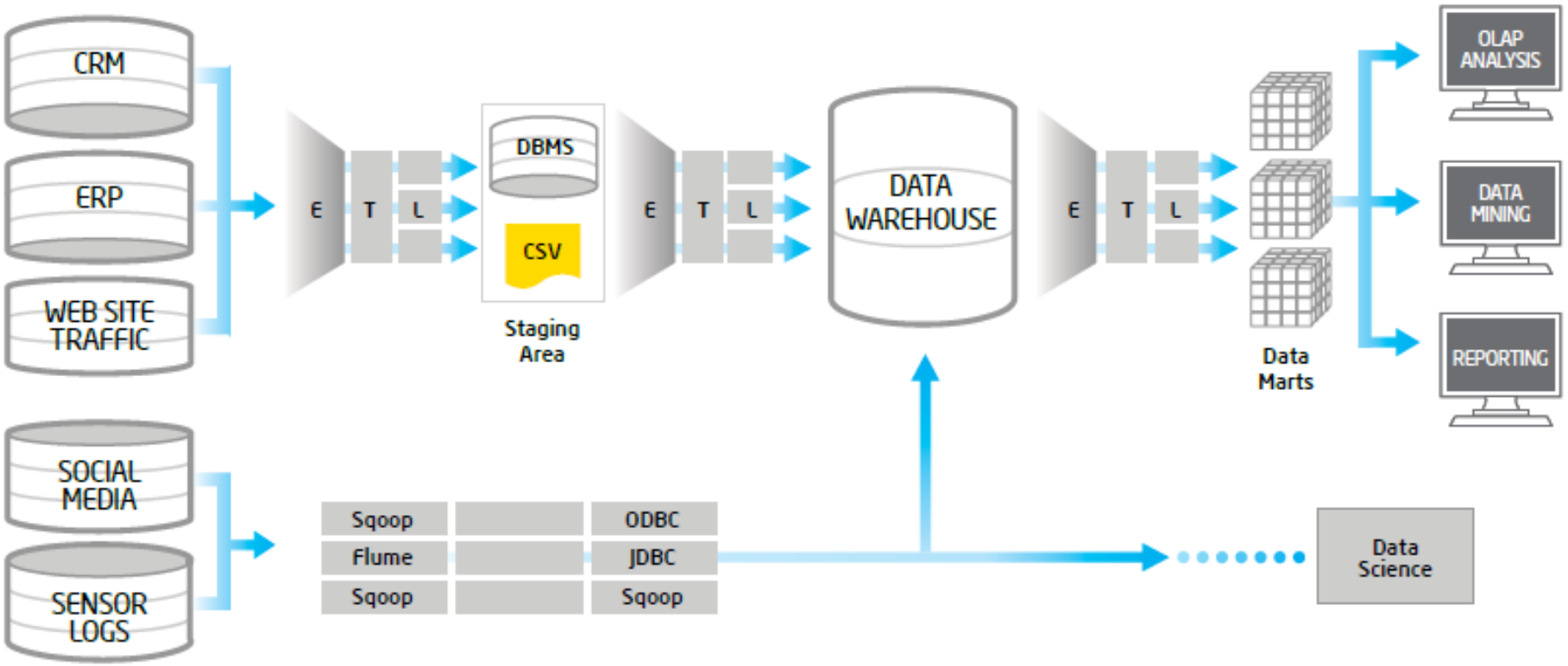
Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

Traditional ETL Architecture



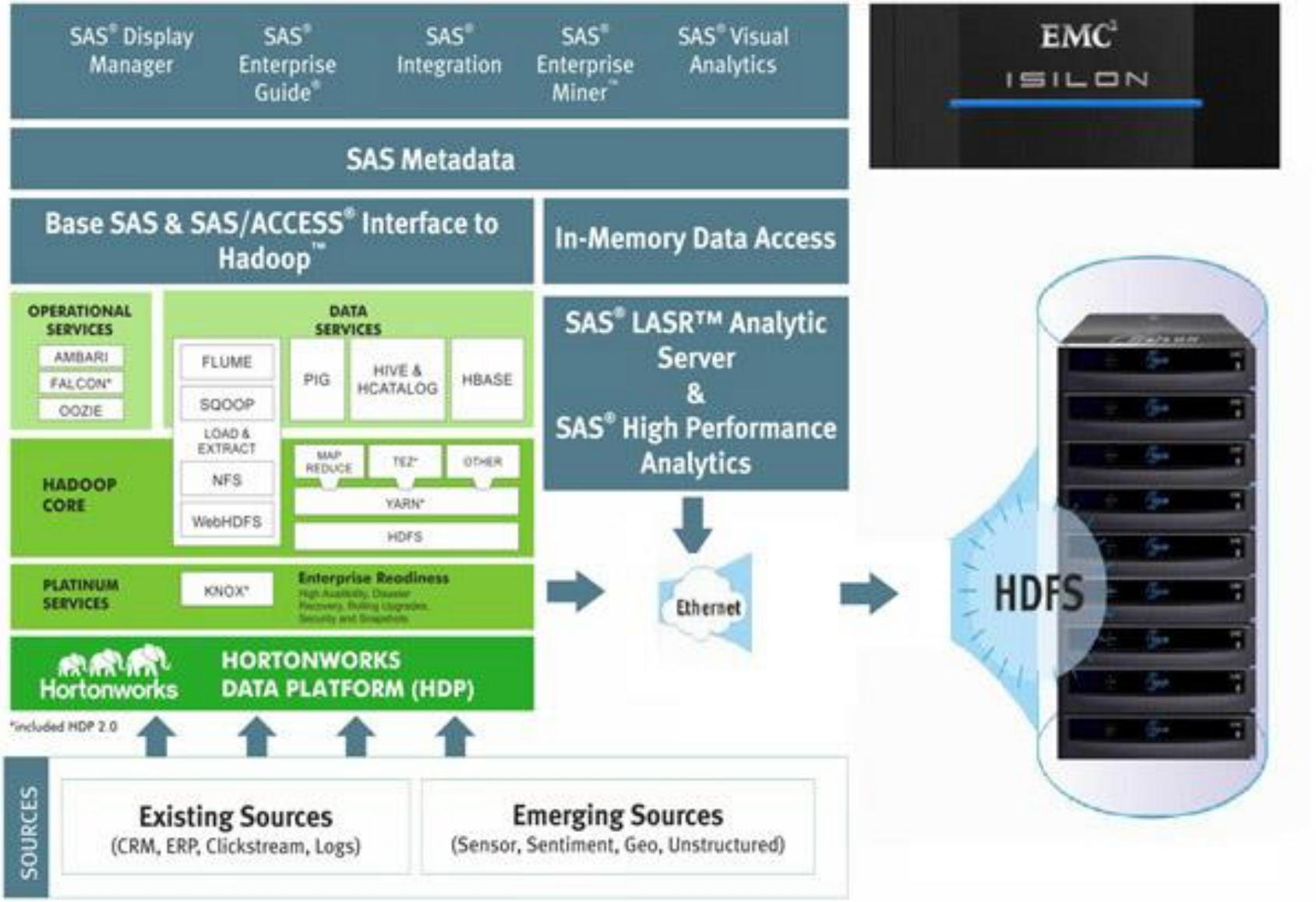
Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

Offload ETL with Hadoop (Big Data Architecture)



Source: https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf

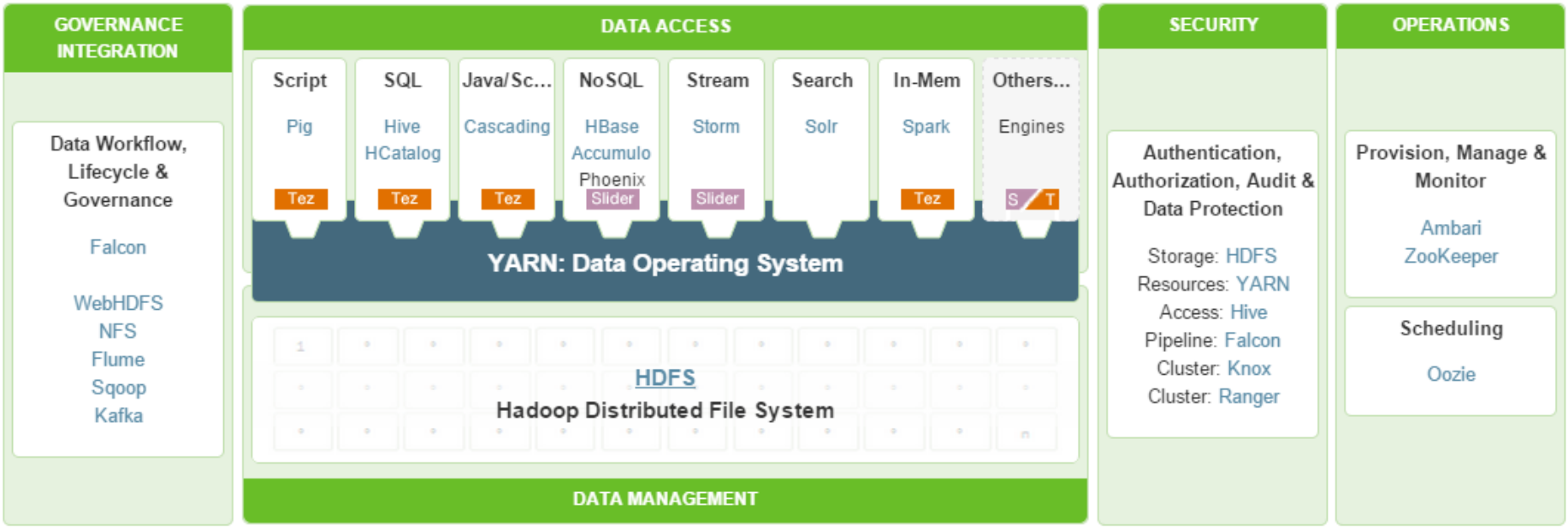
Big Data Solution



Source: http://www.newera-technologies.com/big-data-solution.html

HDP

A Complete Enterprise Hadoop Data Platform

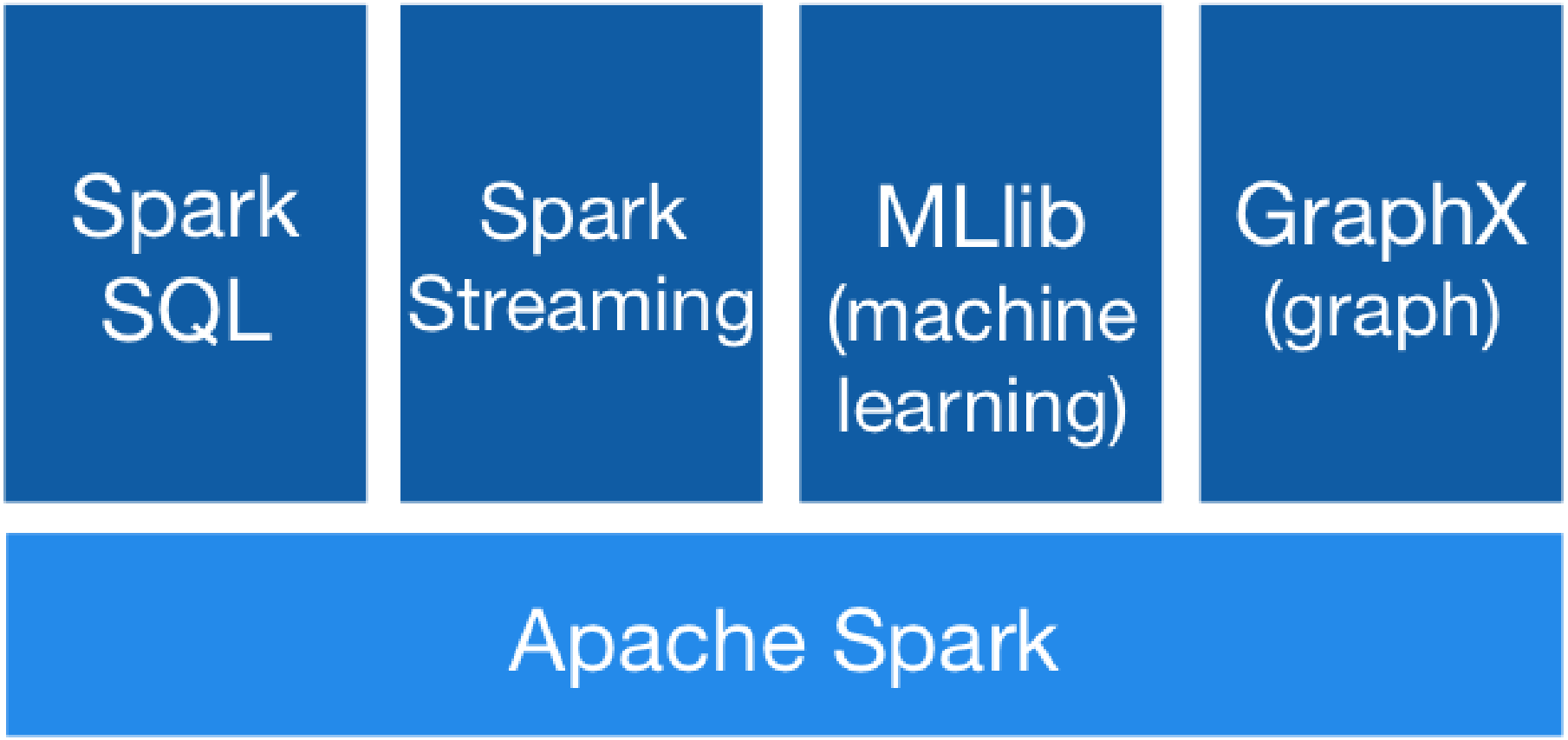


Source: http://hortonworks.com/hdp/

Spark and Hadoop

Source: http://spark.apache.org/

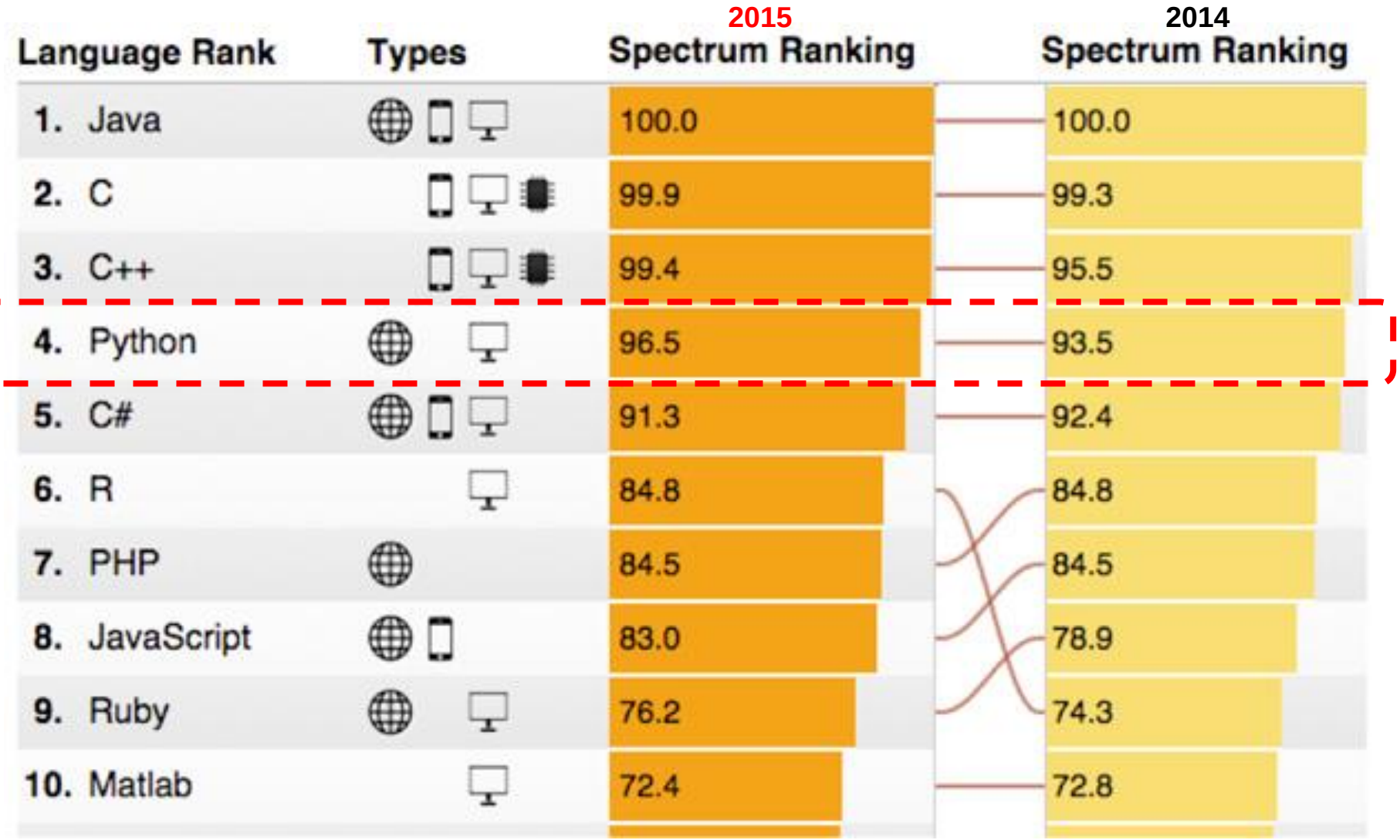
Spark Ecosystem



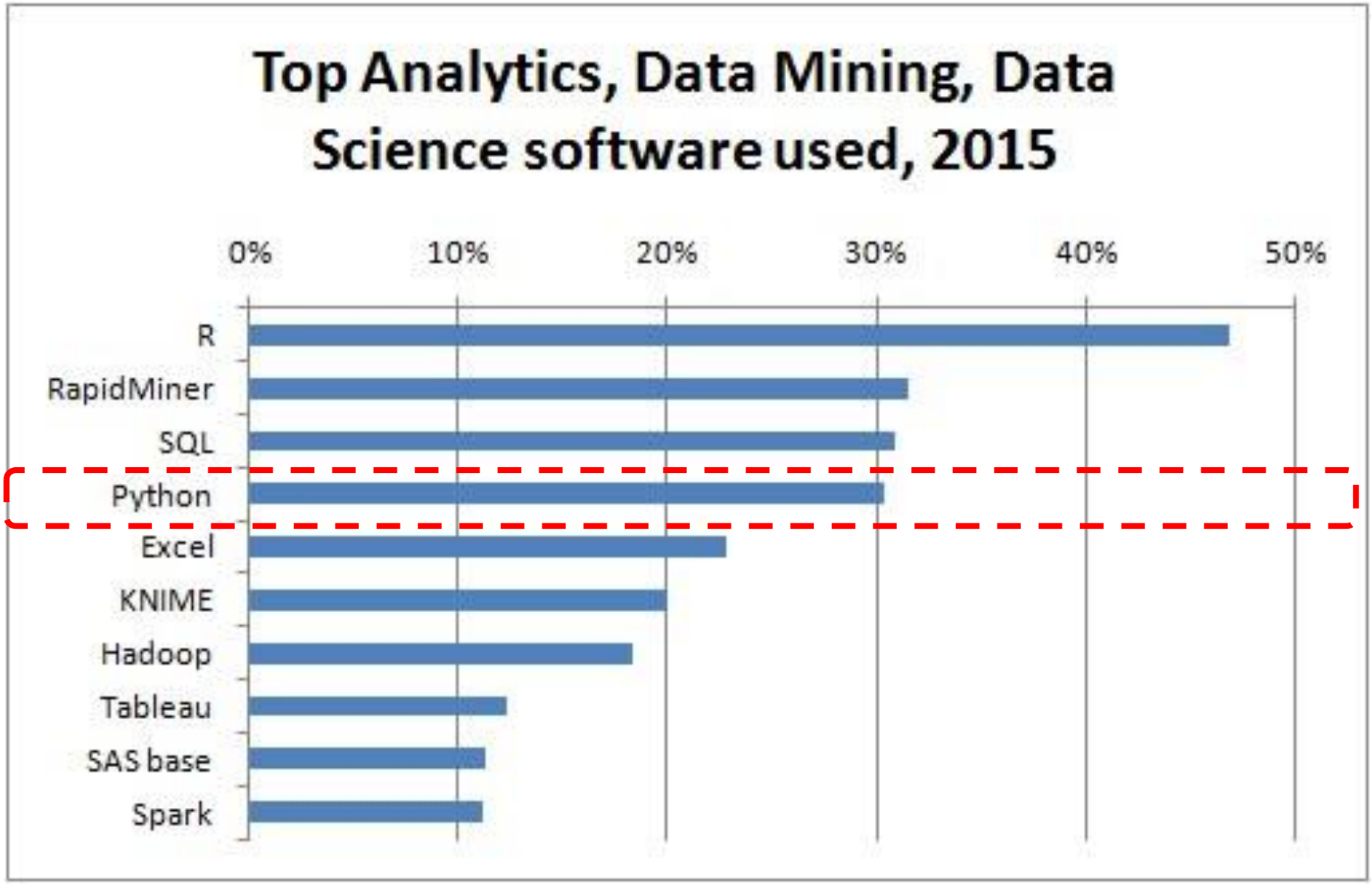
Source: http://spark.apache.org/

Python for Big Data Analytics

(The column on the left is the 2015 ranking; the column on the right is the 2014 ranking for comparison

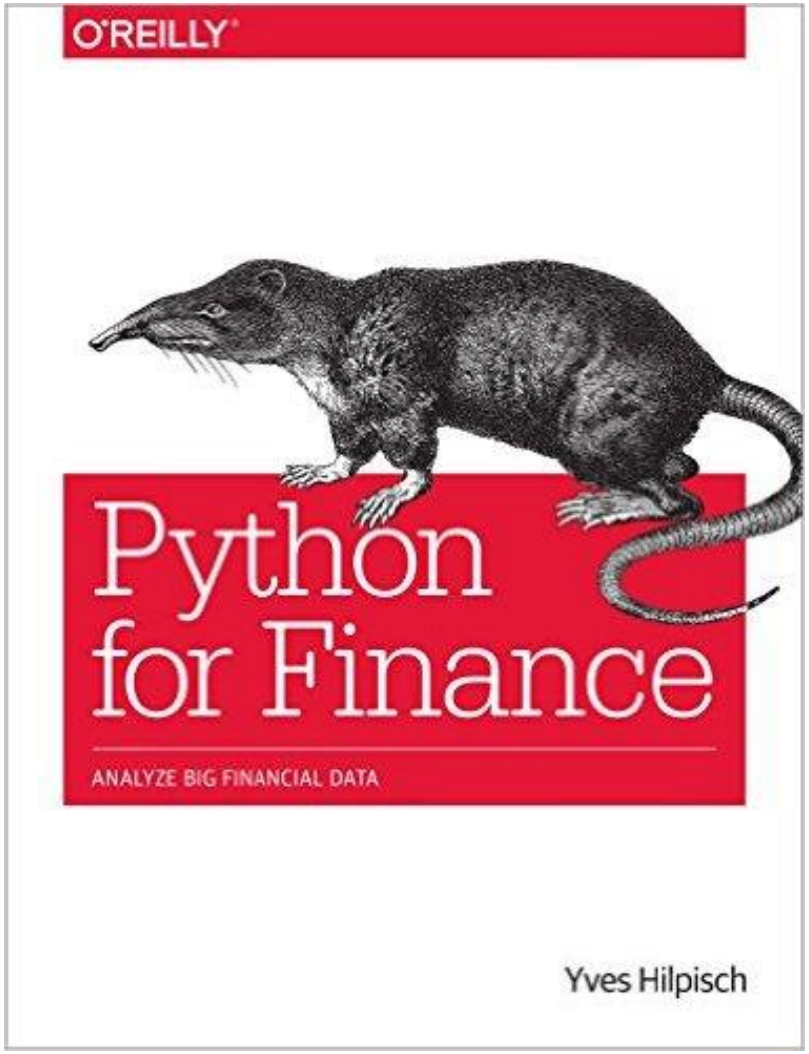


Source: http://spectrum.ieee.org/computing/software/the-2015-top-ten-programming-languages



Source: http://www.kdnuggets.com/2015/05/poll-r-rapidminer-python-big-data-spark.html

Yves Hilpisch, Python for Finance: Analyze Big Financial Data, O'Reilly, 2014

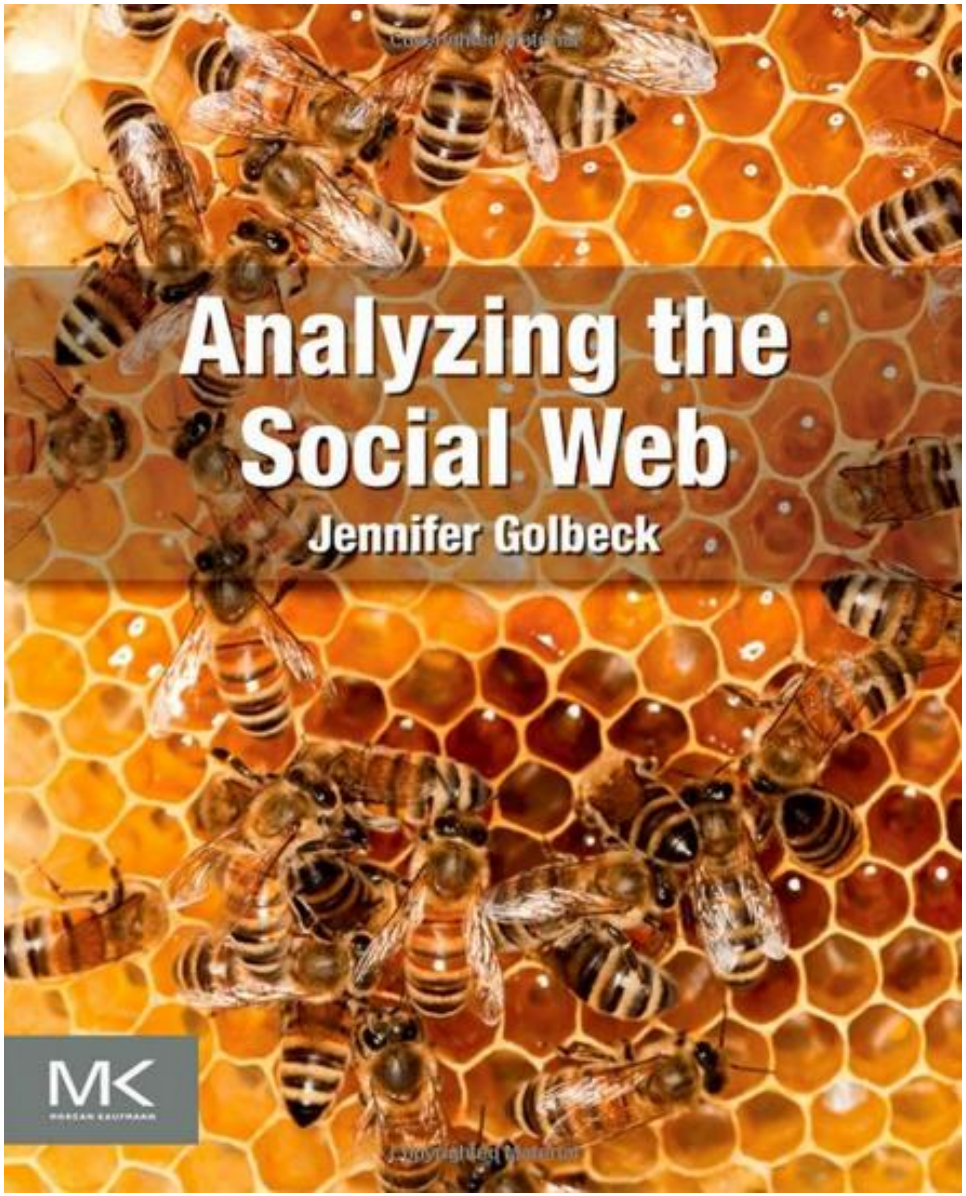


Source: http://www.amazon.com/Python-Finance-Analyze-Financial-Data/dp/1491945281

Business Insights with Social Analytics

Analyzing the Social Web:
Social Network Analysis

Jennifer Golbeck (2013), **Analyzing the Social Web**, Morgan Kaufmann



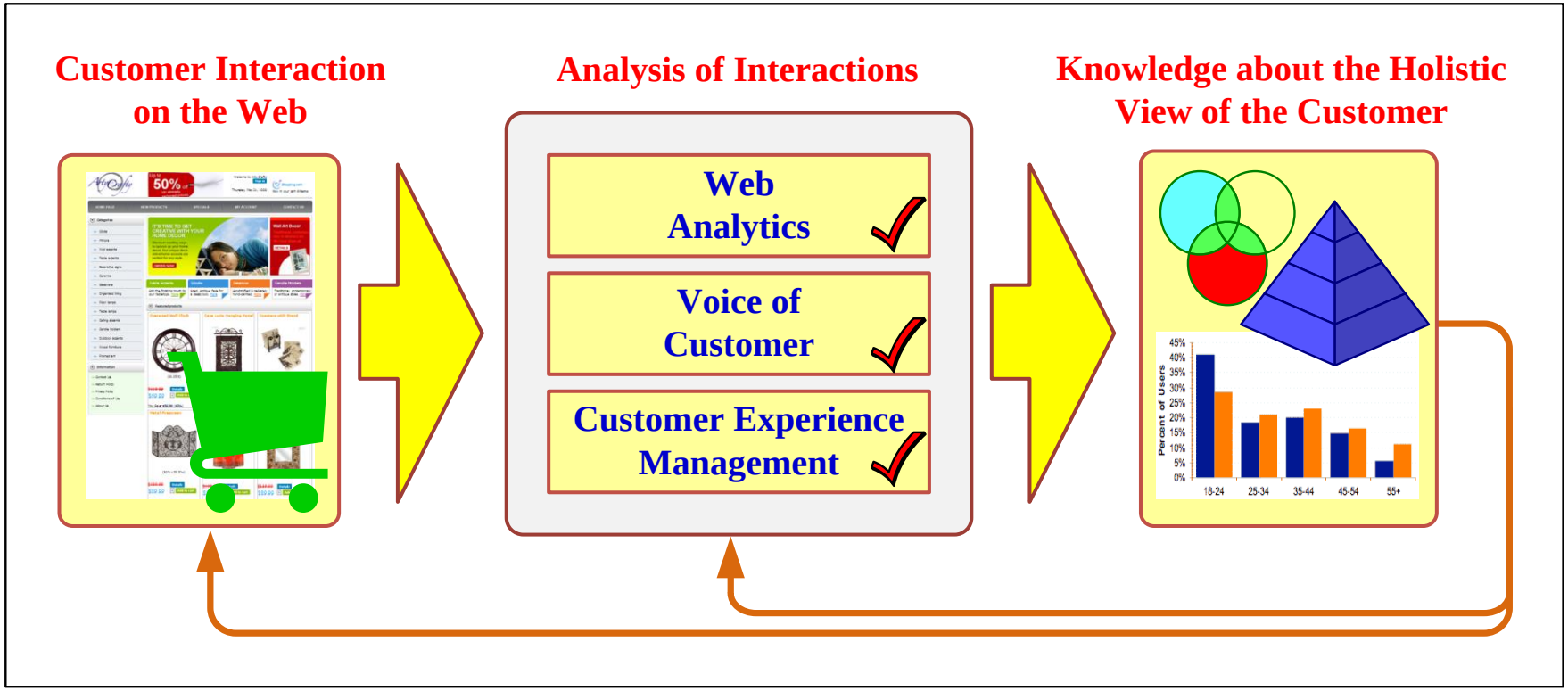
Source: http://www.amazon.com/Analyzing-Social-Web-Jennifer-Golbeck/dp/0124055311

Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

Source: http://www.amazon.com/Mining-Social-Web-Analyzing-Facebook/dp/1449388345

Web Mining Success Stories

- Amazon.com, Ask.com, Scholastic.com, …
- Website Optimization Ecosystem



Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Business Intelligence Trends

1. Agile Information Management (IM)
2. Cloud Business Intelligence (BI)
3. Mobile Business Intelligence (BI)
4. Analytics
5. Big Data

Source: http://www.businessspectator.com.au/article/2013/1/22/technology/five-business-intelligence-trends-2013

Business Intelligence Trends: Computing and Service

- Cloud Computing and Service
- Mobile Computing and Service
- Social Computing and Service

Business Intelligence and Analytics

- Business Intelligence 2.0 (BI 2.0)
  - Web Intelligence
  - Web Analytics
  - Web 2.0
  - Social Networking and Microblogging sites
- Data Trends
  - Big Data
- Platform Technology Trends
  - Cloud computing platform

Source: Lim, E. P., Chen, H., & Chen, G. (2013). Business Intelligence and Analytics: Research Directions. *ACM Transactions on Management Information Systems (TMIS)*, *3*(4), 17

Business Intelligence and Analytics: Research Directions

1. Big Data Analytics
   - Data analytics using Hadoop / MapReduce framework
2. Text Analytics
   - From Information Extraction to Question Answering
   - From Sentiment Analysis to Opinion Mining
3. Network Analysis
   - Link mining
   - Community Detection
   - Social Recommendation

Source: Lim, E. P., Chen, H., & Chen, G. (2013). Business Intelligence and Analytics: Research Directions. *ACM Transactions on Management Information Systems (TMIS)*, *3*(4), 17

Data Scientist:

The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

Source: Davenport, T. H., & Patil, D. J. (2012). Data Scientist. *Harvard business review*

SAS第五屆大數據資料科學家競賽
文字分析與數位行銷大賽

文字分析&
數位行銷大賽

在這個巨量資料的時代，懂得巨量分析的專業人才「資料科學家」（Data Scientist）將成為未來炙手可熱的明日之星。
SAS 希望學生以創意無限及發掘新商機的角度出發，搭配巨量資料分析實例主題，鼓勵全國大學以分組專案及簡報競賽方式，
分析高達兩千萬筆的巨量資料，親身體驗巨量分析的神奇魔力!

http://saschampion.com.tw/

SAS第五屆大數據資料科學家競賽
文字分析與數位行銷大賽

§ 活動時間與地點

1. 報名日期：**2016年2月22日**（一）至**2016年3月18日**（五）額滿為止
2. 起跑說明會：**2016年3月25日**（五）下午六點半至八點半止 **(每組皆須指派隊員出席，須事先報名)**
3. 玉山銀行玉山人力發展中心**1樓** 登峰廳 **(台北市中山區撫順街41巷13號1樓)**
4. 初賽資料分析訓練課程**(Enterprise Guide)**：**2016年3月26日**（六）至**3月30日**（三），每梯次為期**1天(每梯次名額有限，依名額順序額滿為止，活動執行單位將通知參賽者可參加場次)**

【台中場】地點:逢甲大學 (台中市西屯區文華路100號)

時間：上午九點至下午五點，每梯次共計一天

日期	名額
2016年3月26日 (星期六)	120 人
2015年3月27日 (星期日)	120 人

【台北場】地點:台北大學 民生校區 (台北市中山區民生東路三段67號)

時間：上午九點至下午五點，每梯次共計一天台北場

日期	名額
2016年3月26日 (星期六)@東吳大學城中校區	120 人
2016年3月27日 (星期日)	120 人

http://saschampion.com.tw/

Summary

- This course introduces the fundamental concepts and applications technology of big data mining.
- Topics include
  - Big Data Mining
  - Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem
  - Association Analysis
  - Classification and Prediction
  - Cluster Analysis
  - Data Mining Using SAS Enterprise Miner (SAS EM)
  - Case Study and Implementation of Big Data Mining
  - Deep Learning with Google TensorFlow

Contact Information

戴敏育 博士 (Min-Yuh Day, Ph.D.)

專任助理教授
淡江大學 資訊管理學系

電話：02-26215656 #2846
傳真：02-26209737
研究室：B929
地址： 25137 新北市淡水區英專路151號
Email： myday@mail.tku.edu.tw
網址：http://mail.tku.edu.tw/myday/