



Tamkang
University

商業智慧實務

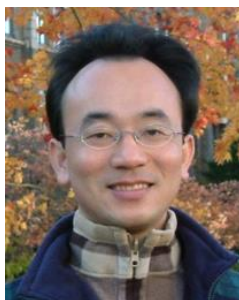
Practices of Business Intelligence

資料科學與巨量資料分析 (Data Science and Big Data Analytics)

1032BI06

MI4

Wed, 9,10 (16:10-18:00) (B130)



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/>

2015-04-15



課程大綱 (Syllabus)

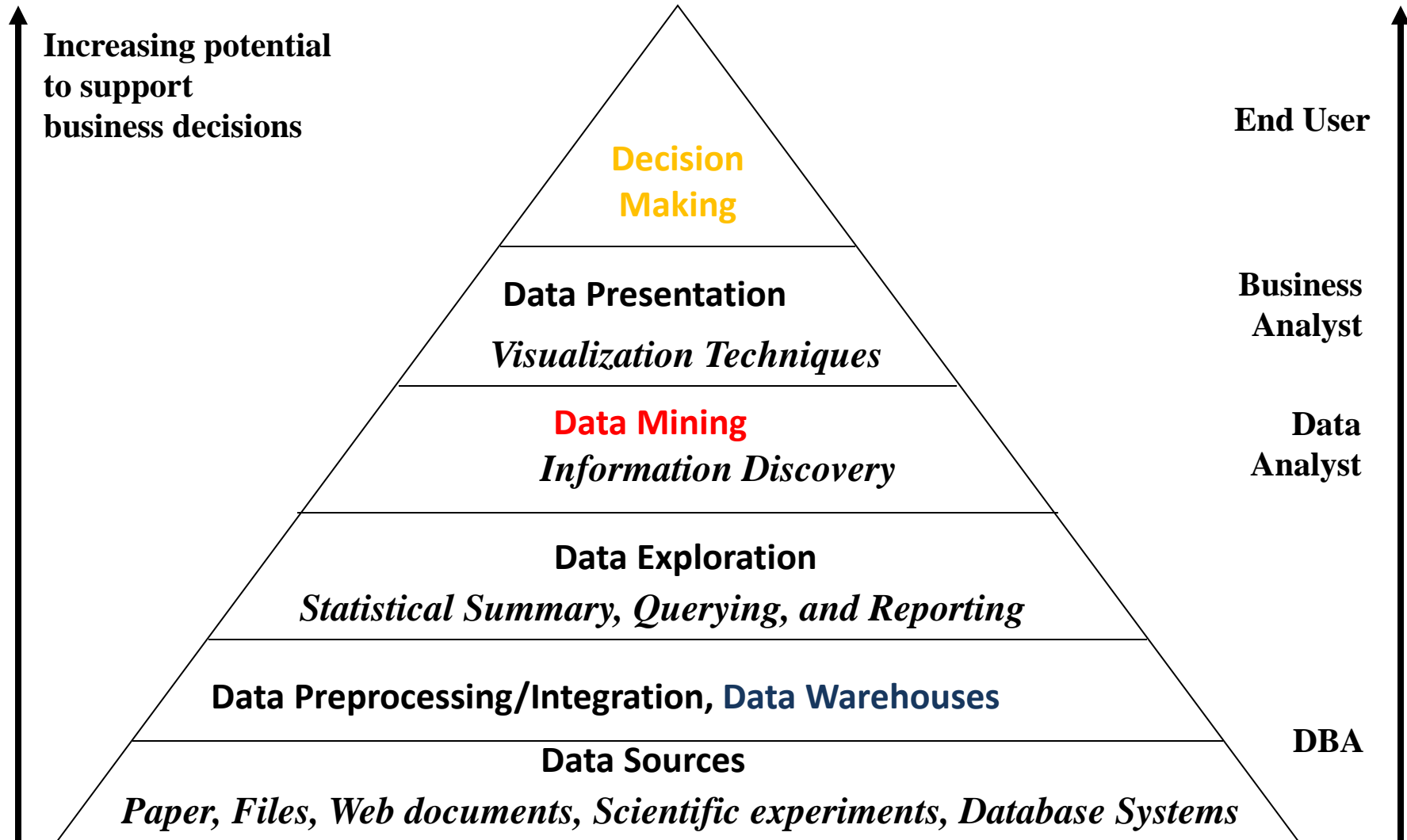
週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2015/02/25	商業智慧導論 (Introduction to Business Intelligence)
2	2015/03/04	管理決策支援系統與商業智慧 (Management Decision Support System and Business Intelligence)
3	2015/03/11	企業績效管理 (Business Performance Management)
4	2015/03/18	資料倉儲 (Data Warehousing)
5	2015/03/25	商業智慧的資料探勘 (Data Mining for Business Intelligence)
6	2015/04/01	教學行政觀摩日 (Off-campus study)
7	2015/04/08	商業智慧的資料探勘 (Data Mining for Business Intelligence)
8	2015/04/15	資料科學與巨量資料分析 (Data Science and Big Data Analytics)

課程大綱 (Syllabus)

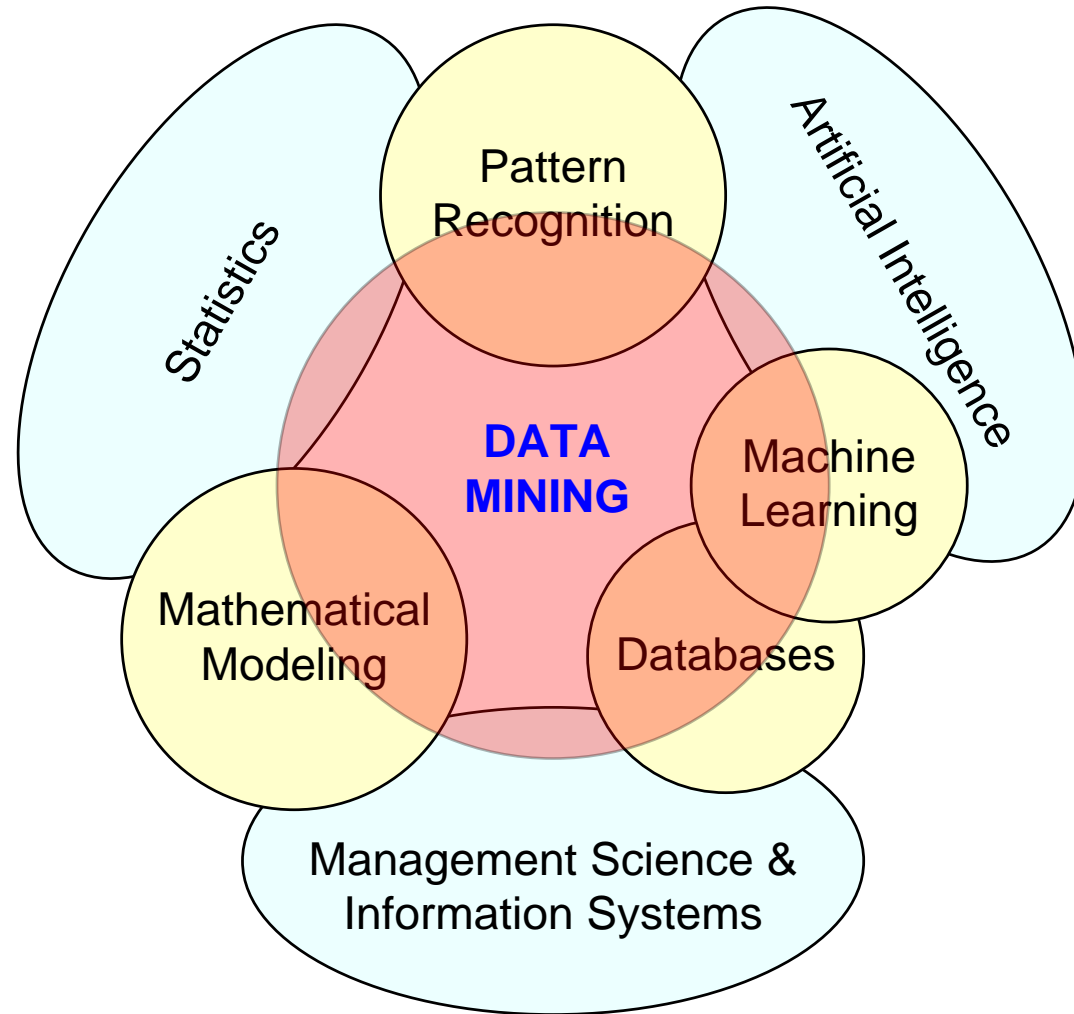
週次	日期	內容 (Subject/Topics)
9	2015/04/22	期中報告 (Midterm Project Presentation)
10	2015/04/29	期中考試週 (Midterm Exam)
11	2015/05/06	文字探勘與網路探勘 (Text and Web Mining)
12	2015/05/13	意見探勘與情感分析 (Opinion Mining and Sentiment Analysis)
13	2015/05/20	社會網路分析 (Social Network Analysis)
14	2015/05/27	期末報告 (Final Project Presentation)
15	2015/06/03	畢業考試週 (Final Exam)

Business Intelligence

Data Mining, Data Warehouses



Data Mining at the Intersection of Many Disciplines



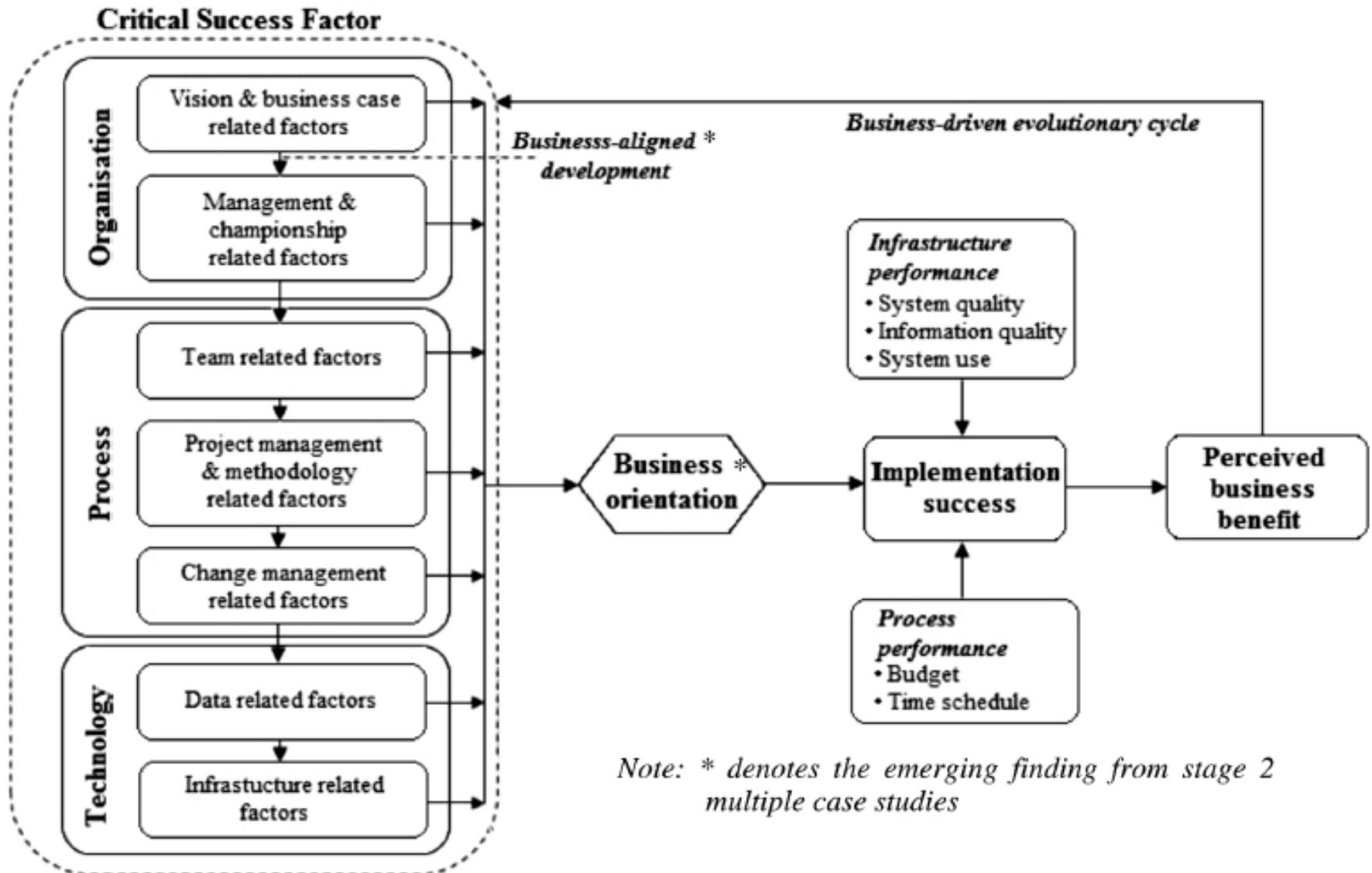
Outline

- Business Intelligence **Implementation**
- Business Intelligence **Trends**
- **Data Science**
- **Big Data Analytics**
 - **Big Data, Big Analytics:**
Emerging Business Intelligence and Analytic
Trends for Today's Businesses

Business Intelligence Implementation

Business Intelligence Implementation

CSFs Framework for Implementation of BI Systems



Critical Success Factors of Business Intelligence Implementation

- **Organizational dimension**
 - Committed management support and sponsorship
 - Clear vision and well-established business case
- **Process dimension**
 - Business-centric championship and balanced team composition
 - Business-driven and iterative development approach
 - User-oriented change management.
- **Technological dimension**
 - Business-driven, scalable and flexible technical framework
 - Sustainable data quality and integrity

Business Intelligence Trends

Business Intelligence Trends

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

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

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

Data Science

“Data science
is
the study of the
generalizable extraction of
knowledge from data ”

Data Science

- A common epistemic requirement in assessing whether new knowledge is **actionable for decision making** is its **predictive power**, not just its ability to explain the past.

Data Scientist

- A data scientist requires an integrated skill set spanning **mathematics, machine learning, artificial intelligence, statistics, databases, and optimization,** along with a **deep understanding** of the craft of problem formulation to engineer effective solutions.

Data Scientist: **The Sexiest Job** **of the 21st Century**

(Davenport & Patil, 2012)(HBR)

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."

Data Scientist

Profile of a Data Scientist



Data Science

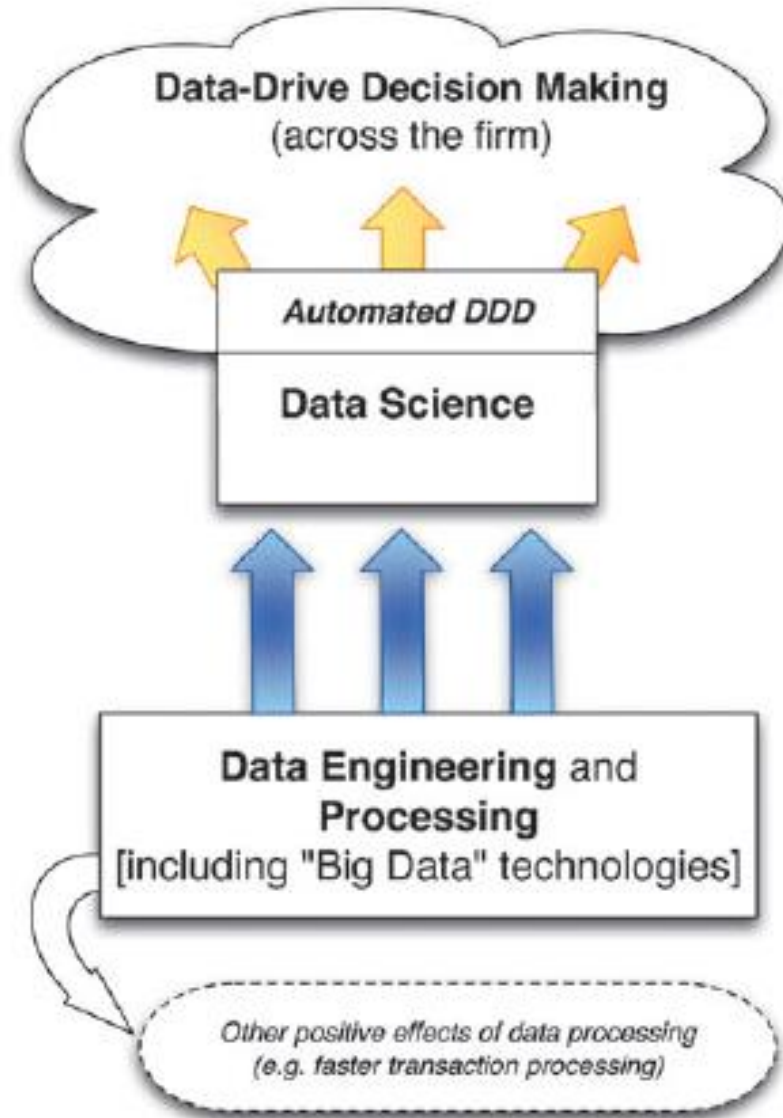
and its Relationship to

Big Data

and

Data-Driven Decision Making

Data science in the organization



Big Data Analytics

Big Data, Big Analytics:

**Emerging Business Intelligence
and Analytic Trends
for Today's Businesses**

Big Data:

The Management Revolution

HBR.ORG

Harvard Business Review



OCTOBER 2012
REPRINT R1210C

SPOTLIGHT ON BIG DATA

Big Data: The Management Revolution

Exploiting vast new flows of information can radically improve your company's performance. But first you'll have to change your decision-making culture.
by Andrew McAfee and Erik Brynjolfsson

Copyrighted Material



ENTERPRISE ANALYTICS

Optimize Performance, Process, and
Decisions through Big Data



EDITED BY
THOMAS DAVENPORT

Copyrighted Material

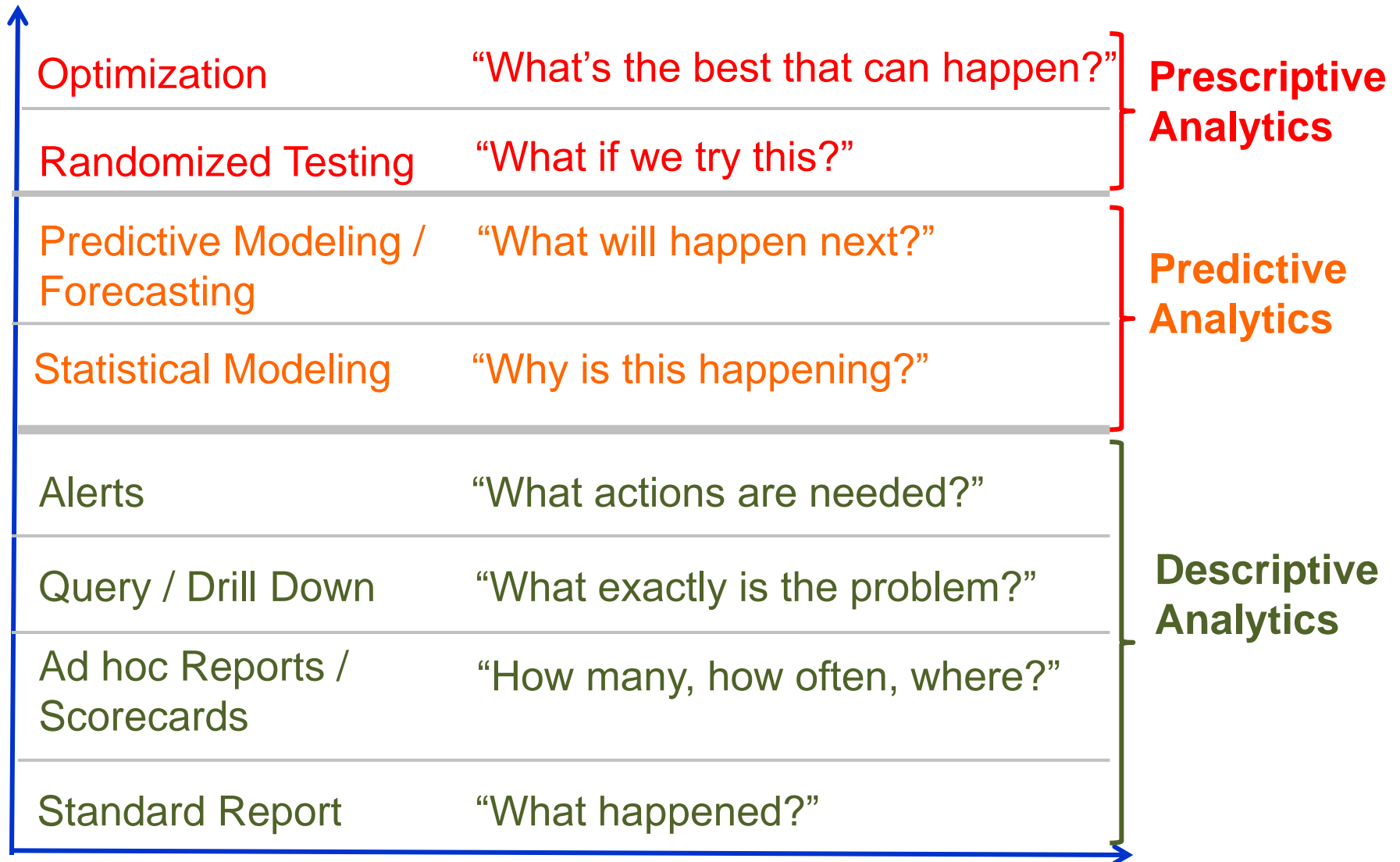
Business Intelligence and Enterprise Analytics

- Predictive analytics
- Data mining
- Business analytics
- Web analytics
- **Big-data** analytics

Three Types of Business Analytics

- Prescriptive Analytics
- Predictive Analytics
- Descriptive Analytics

Three Types of Business Analytics



Big-Data Analysis

- **Too Big,
too Unstructured,
too many different source
to be manageable through traditional
databases**

The Rise of “Big Data”

- “Too Big” means databases or data flows in **petabytes (1,000 terabytes)**
 - Google processes about 24 petabytes of data per day
- “Too unstructured” means that the data isn’t easily put into the traditional rows and columns of conventional databases

Examples of Big Data

- Online information
 - Clickstream data from Web and social media content
 - Tweets
 - Blogs
 - Wall postings
- Video data
 - Retail and crime/intelligence environments
 - Rendering of video entertainment
- Voice data
 - call centers and intelligence intervention
- Life sciences
 - Genomic and proteomic data from biological research and medicine



Source: <http://www.amazon.com/Big-Data-Analytics-Intelligence-Businesses/dp/111814760X>

Wiley CIO Series

Copyrighted Material

Foreword by
JIM STOGDILL
General Manager
Radar,
O'Reilly Media

BIG DATA BIG ANALYTICS

EMERGING BUSINESS INTELLIGENCE AND
ANALYTIC TRENDS FOR TODAY'S
BUSINESSES

Michael Minelli • Michele Chambers • Ambiga Dhiraj

Copyrighted Material

Big Data, Big Analytics:

Emerging Business Intelligence and Analytic Trends for Today's Businesses

- What Big Data is and why it's important
- Industry examples (Financial Services, Healthcare, etc.)
- Big Data and the New School of Marketing
- Fraud, risk, and Big Data
- Big Data technology
- Old versus new approaches
- Open source technology for Big Data analytics
- The Cloud and Big Data

Big Data, Big Analytics:

Emerging Business Intelligence and Analytic Trends for Today's Businesses

- Predictive analytics
- Crowdsourcing analytics
- Computing platforms, limitations, and emerging technologies
- Consumption of analytics
- Data visualization as a way to take immediate action
- Moving from beyond the tools to analytic applications
- Creating a culture that nurtures decision science talent
- A thorough summary of ethical and privacy issues

What is **BIG Data**?

Volume

Large amount of data

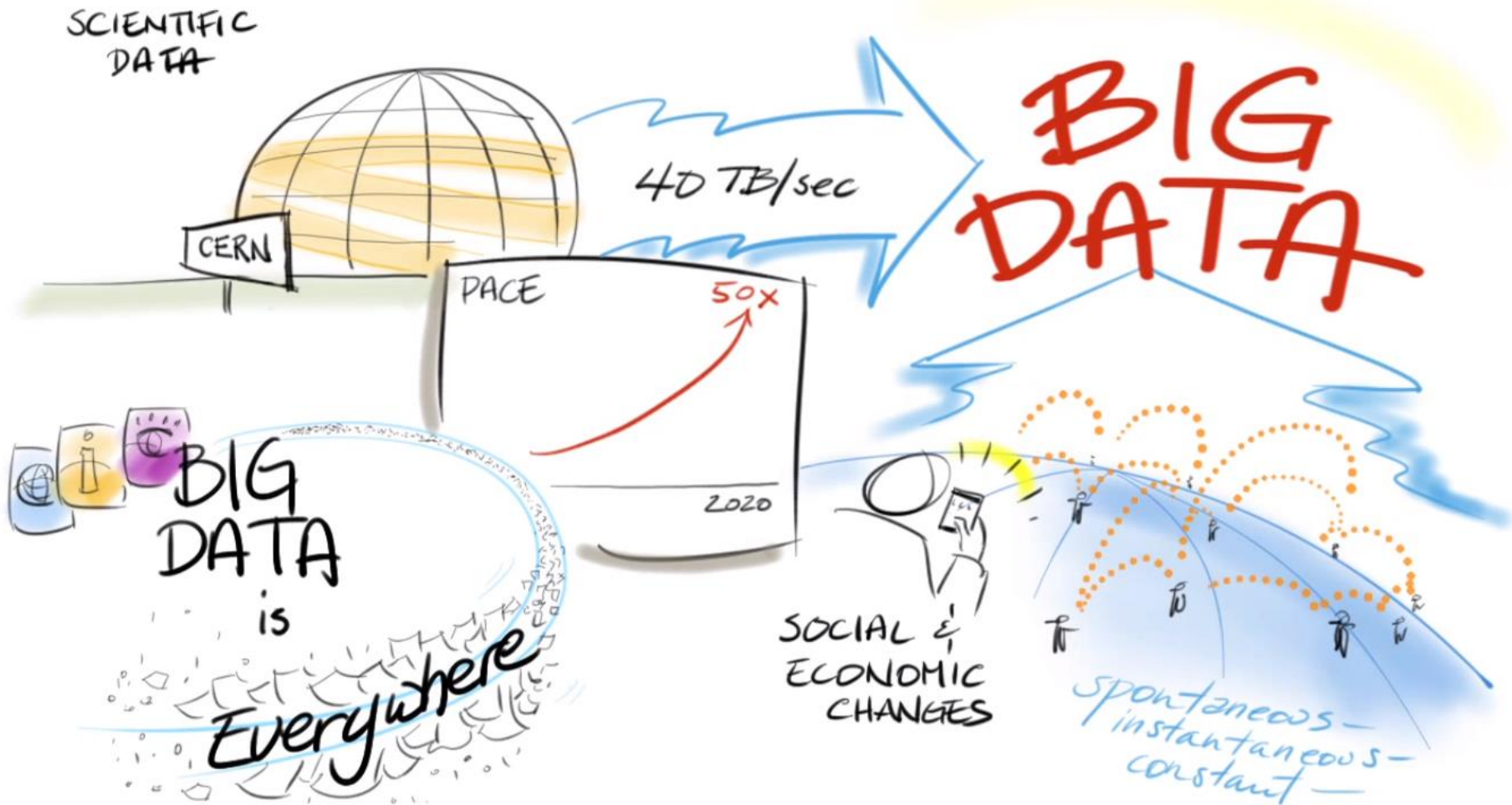
Velocity

Needs to be analyzed **quickly**

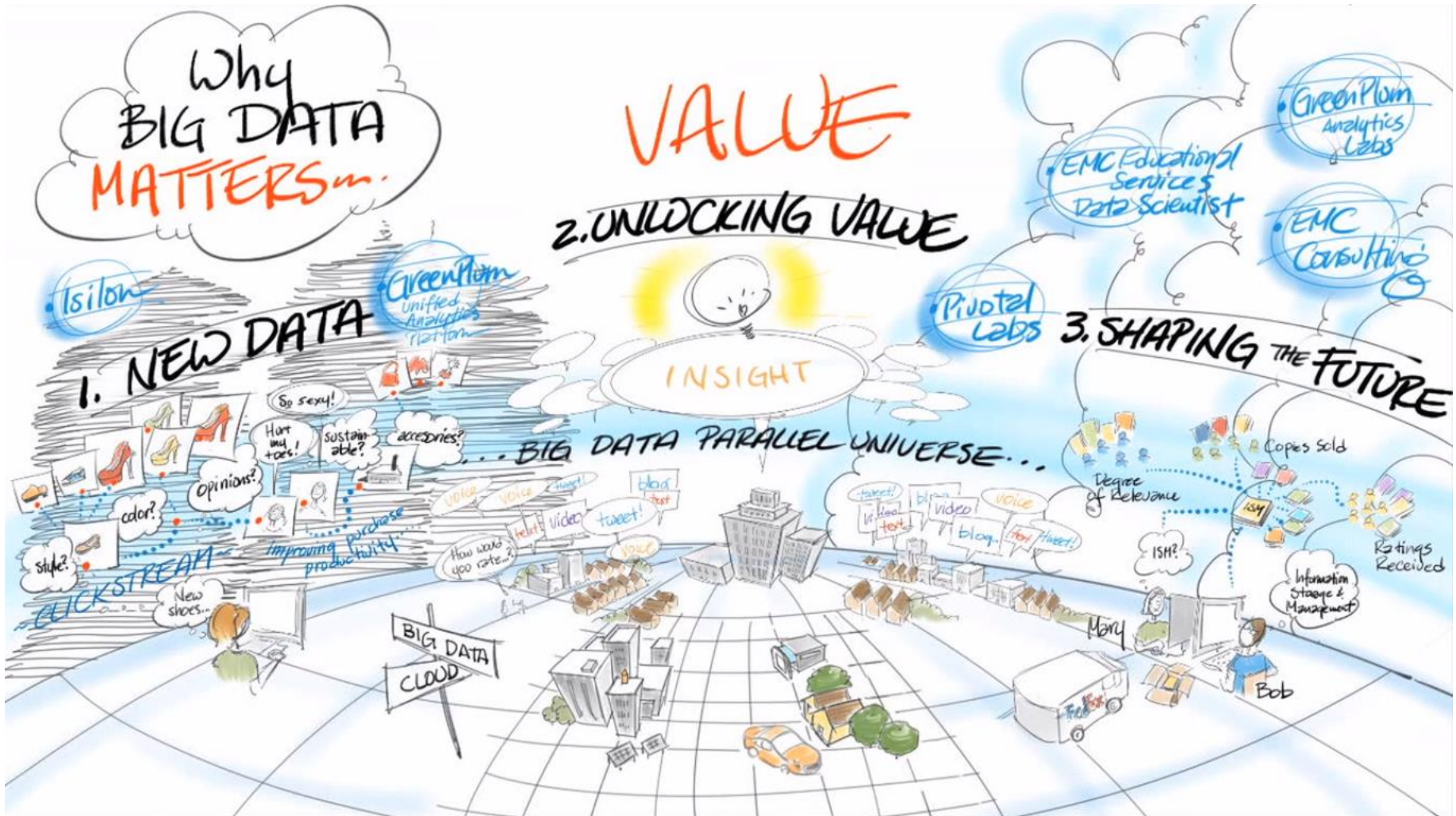
Variety

Different types of structured and unstructured data

Big Ideas: How Big is **Big Data**?



Big Ideas: Why **Big Data** Matters



Key questions enterprises are asking about **Big Data**

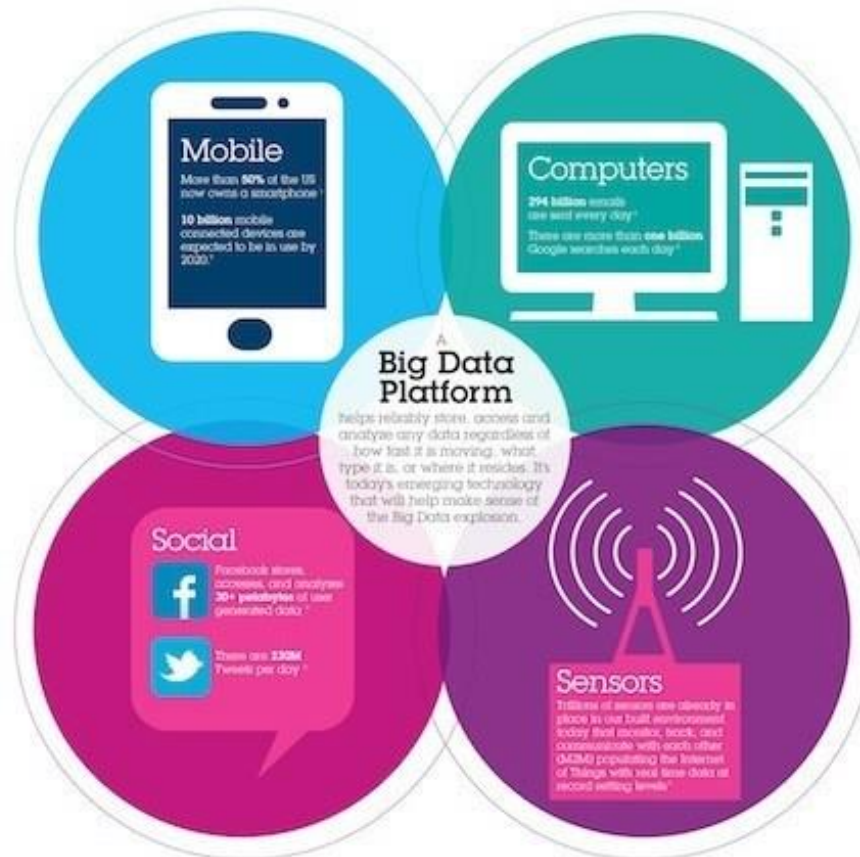
- How to store and protect big data?
- How to backup and restore big data?
- How to organize and catalog the data that you have backed up?
- How to keep costs low while ensuring that all the critical data is available when you need it?

Volumes of Data

- Facebook
 - **30 billion pieces of content** were added to Facebook this past month by 600 million plus users
- Youtube
 - **More than 2 billion videos** were watch on YouTube yesterday
- Twitter
 - **32 billion searches** were performed last month on Twitter

Big Data: Making the World go Round

Big Data is growing and moving fast from a variety of sources; are you keeping up?



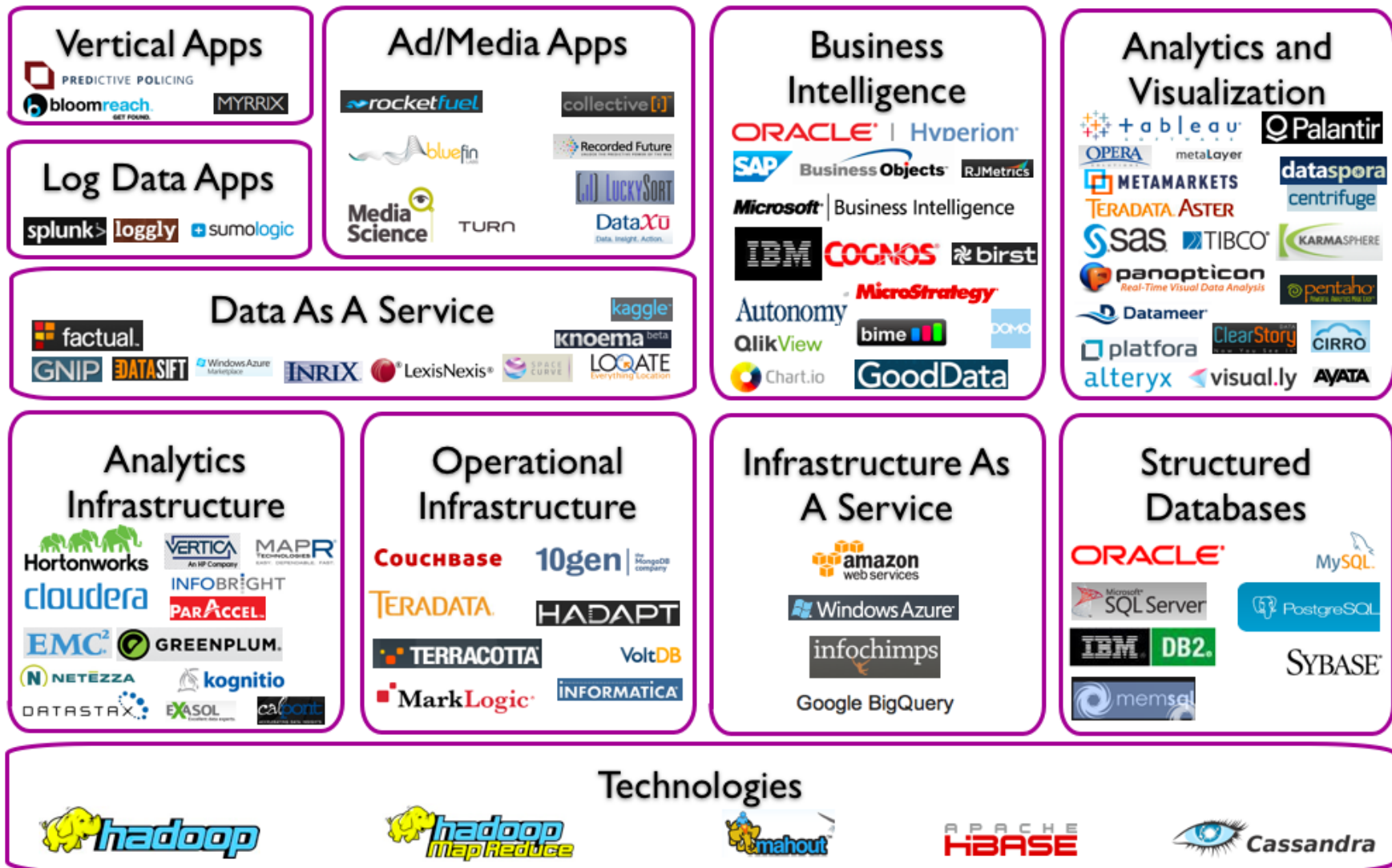
Information gathered by IBM:

1. Clarks Analytics Consulting - US Mobile Data Market Update Q3 2012
2. 2011 SensorNet
3. IBM - Managing the Big Flood of Big Data in Digital Marketing
4. Google - How Google Search Works
5. Wikiboo - Taming Big Data
6. IBM - Managing the Big Flood of Big Data in Digital Marketing
7. IBM

IBM, the IBM logo, and the IBM name are trademarks of International Business Machines Corporation. Other product names and service trademarks belong to their respective owners. IBM and the IBM logo are trademarks of International Business Machines Corporation. © International Business Machines Corporation 2012. All rights reserved.



Big Data Landscape



Big Data Landscape (Version 2.0)

Infrastructure



Analytics



Applications



Cross Infrastructure / Analytics



Open Source Projects



© Matt Turck (@mattturck) and ShivonZilis (@shivonz) Bloomberg Ventures

Source: <http://mattturck.com/2012/10/15/a-chart-of-the-big-data-ecosystem-take-2/>

BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or IPO

Infrastructure

NoSQL Databases
 FOUNDATIONDB, DATSTAX, mongoDB, COUCHBASE, ZEROSPIKE, HYPERKTABLE, aqri, CLOUDANT, OhnData, Neo4j, sones

Hadoop On Prem
 HADAPT, cloudera, splice, Zettaset, amazon, MAPR, Microsoft, Hortonworks, Pivotal, IBM

Big Data
 MORTAR, infochimps, Quobole, JETRO, altiscale, AMAZON REDSHIFT

NewSQL Databases
 MarkLogic, TRANSATLICE, Plain, paradigm, memsql, deep db, skySQL, NUODB, Clustrix, VoltDB, SQLFire

Cluster Service
 LexisNexis, HPCC Systems, mesosphere, Acunu

Management / Monitoring
 BUTER, THOUGHT, New Relic, metator, StackIQ, tidemark, appnomic, oceanSYNc, DATADOG, boundy

Graph Databases
 Neo4j, aster data, InfiniteGraph

Data Transformation
 TRIFACTA, Paxata, KALIDO, Evelytix, TRANSIRON, syncsoft

Security
 DATAGUISE, Stormpath, IMPERIA

Storage
 Cleversafe, panasas, sinbstrorage, Compuverde

App Dev
 CONTINUITY, CONCURRENT, wibedata

Crowd-sourcing
 microTASK, servio, mobileworks

Analytics

Analytics Platforms
 databricks, QuantCell, PERSASIVE, guavus, Datameer, KARMA360, collectiveIQ, PRECOG, dataspora

For Business Analysts
 STAT WING, CIRRO, TREPAREL, OrigamiLogic, ClearStory, DataGravity

Data Science Platforms / Tools
 domino, Alpine, Sense, MORTAR, CONTINUUM, ploty, yhat, MODE

Unstructured Data
 BASIS, ATTIVO, GENERAL SENTIMENT, semantria, crimson hexagon, ai, Quid, Palantir

Data Visualization
 tableau, ZavenData, visual.ly, Roambi, Chart.io, looker, Ayasdi, ISS, DataHero, TICKBOARD

Machine Learning
 SKYTREE, big ml, YOUTAINE ANALYTICS, wise.io, contact relevant

Location / People / Events
 RADIUS, Fliptop, LOCATE, Rocu, PlaceIQ

Statistical Computing
 SAS, MATLAB, Revolution, SPSS

Log Analysis
 splunk, loggly, sumologic, Kibana

Crowd-Sourced
 kaggle, METAMARKETS, DataKind

Real Time
 amato, causita

SMB
 RJMetrics, retention, sumail, GoSquared, custora

Applications

Ad Optimization
 aggregate knowledge, rocketfuel, TAPAD, ai Match, MediaMath, 33across

Publisher Tools
 Chartbeat, Yieldex, yieldbot

Marketing
 LATTICE ENGINES, Sailthru, spinnkr, gainsight, Kontera, RelateIQ, Tell apart, persado, bloomreach, CLICKFOX, Pursuway

Finance
 Lenddo, BILL GUARD, wonga, cignifi, LendUp, KENSHO, OnDeck

Human Capital
 evolv, Centelo, gild

Legal
 JUDICATA, RAVEL, Lex Machina

Government / Regulation
 mark43, enigma, FORTSCALE, feedzai

Security
 SCIFYD, sift science

Education / Learning
 KNEWTON, @eclara, PANORAMA, Clever

Health
 Recombine, 23andMe, Ginger.io, FLATIRON, Counsyl

Industries
 tubular, OPOWER, SIGHT MACHINE, THE CLIMATE CORPORATION, NEXT BIG SOUND

Cross Infrastructure /

SAP, SAS, IBM, Google, Microsoft, vmware, amazon, 1010data, talend, TERAdata, hp, NetApp

Open Source

Framework
 Hadoop, YARN, HDFS, Spark

Query / Data Flow
 MapReduce, Pig, Hive, Tez

Data Access
 Cassandra, SciDB, ORACLE, HBASE, mongoDB, riak, Sqoop

Coordination / Work-flow
 ZooKeeper, talend

Real Time
 Storm

Stat Tools
 SciPy, IP

Machine Learning
 MLlib

Cloud Deploy
 AWS, Azure, GCP

Search
 Solr, ELASTICSEARCH, LUCENE

Data Sources

Data Mkts
 Windows Azure Marketplace, blueka, DataMarket, factual, knoema

Data Sources
 DATA GOV, premise, YODLEE, VALIDIC, xignite, ploid, quandt, STANDARD TREASURY, human/api

Sensor Data
 kinsa, STREETLINE, fitbit, RunKeeper, JAWBONE, LUMASENSE, Withings, BASIS, estimate

Incubators & School
 zipfian, GA, INSIGHT, DataLine

Big Data Vendors and Technologies



<p>Data Acquisition</p> <p><i>Including Complex Event Processing (CEP) tools</i></p>	<p>VLDW and BI Appliances</p>	<p>Analytics</p>	<p>BPM & Action</p>
<p>Data Providers</p> <p><i>And all your own data</i> <i>And your partners data</i></p>	<p>No SQL</p>	<p>Data Virtualization</p>	<p>Microsoft</p> <p>Capgemini - Capping IT off Manuel Sevilla - 2012</p>
<p>Data Governance</p>	<p>Content Management</p>	<p>BI Tools</p>	<p>BI Tools</p>

Source: <http://www.capgemini.com/blog/capping-it-off/2012/09/big-data-vendors-and-technologies-the-list>

Processing Big Data

Google



Source: http://whatsthebigdata.files.wordpress.com/2013/03/google_datacenter.jpg

Processing Big Data, Facebook



Summary

- Business Intelligence **Implementation**
- Business Intelligence **Trends**
- **Data Science**
- **Big Data Analytics**
 - **Big Data, Big Analytics:**
Emerging Business Intelligence and Analytic Trends for Today's Businesses

References

- Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*, 50(3), 23.
- Lim, E. P., Chen, H., & Chen, G. (2013). Business Intelligence and Analytics: Research Directions. *ACM Transactions on Management Information Systems (TMIS)*, 3(4), 17
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*.
- Davenport, T. H., & Patil, D. J. (2012). Data Scientist. *Harvard business review*.
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59.
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73.
- Thomas H. Davenport,
Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data, FT Press, 2012
- Michael Minelli, Michele Chambers, Ambiga Dhiraj,
Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley, 2013
- Viktor Mayer-Schonberger, Kenneth Cukier,
Big Data: A Revolution That Will Transform How We Live, Work, and Think, Eamon Dolan/Houghton Mifflin Harcourt, 2013