Chapter 8

Future Trends, Privacy and Managerial Considerations in Analytics

Learning Objectives (1 of 2)

8.1 Explore some of the emerging technologies that may impact analytics, business intelligence (BI), and decision support

8.2 Describe the emerging Internet of Things (IoT) phenomenon, potential applications, and the IoT ecosystem

8.3 Describe the current and future use of cloud computing in business analytics

8.4 Describe how geospatial and location-based analytics are assisting organizations

Learning Objectives (2 of 2)

8.5 Describe the organizational impacts of analytics applications

8.6 List and describe the major ethical and legal issues of analytics implementation

8.7 Identify key characteristics of a successful data science professional

Opening Vignette

Analysis of Sensor Data Helps Siemens Avoid Train Failures

Discussion Questions

1. In industrial equipment such as trains, what parameters might one measure on a regular basis to estimate the equipment's current performance and future repair needs?
2. How would weather data be useful in analyzing a train's equipment status?
3. Estimate how much data you might collect in one month using, say, 1,000 sensors on a train. Each sensor might yield 1 KB data per second.
4. How would you propose to store such data sets?

Internet of Things (IoT) (1 of 2)

- I o T is an area with explosive growth
- Connecting physical world to the Internet
- Social Network versus I o T
  - human-to-human vs. machine-to-machine
- Enablers: sensors and sensing devices
- Example
  - Self driving cars
  - Fitness trackers
  - Smartbin – trash detectors detecting fill levels
  - Smart refrigerators, and other appliances

Internet of Things (IoT) (2 of 2)

- By 2020, besides computing and communication devices (tablets, phones, and PCs), another 38B things will be connected to the Internet
- Reasons for incredible growth in IoT:
  - Hardware – smaller, affordable, more powerful
  - Availability of BI tools – more capable and cheaper
  - Emergence of new and innovative use cases
- There isn't a universal agreement on the term IoT
  - Web of Things
  - Internet of Systems, …

Application Case 8.1

SilverHook Powerboats Uses Real-Time Data Analysis to Inform Racers and Fans

Questions for Discussion

1. What type of information might the sensors on a race boat generate that would be important for the racers to know? What about for the fans?
2. Which other sports might benefit from similar technologies?
3. What technological challenges might you face in building such systems?

Application Case 8.2

Rockwell Automation Monitors Expensive Oil and Gas Exploration Assets

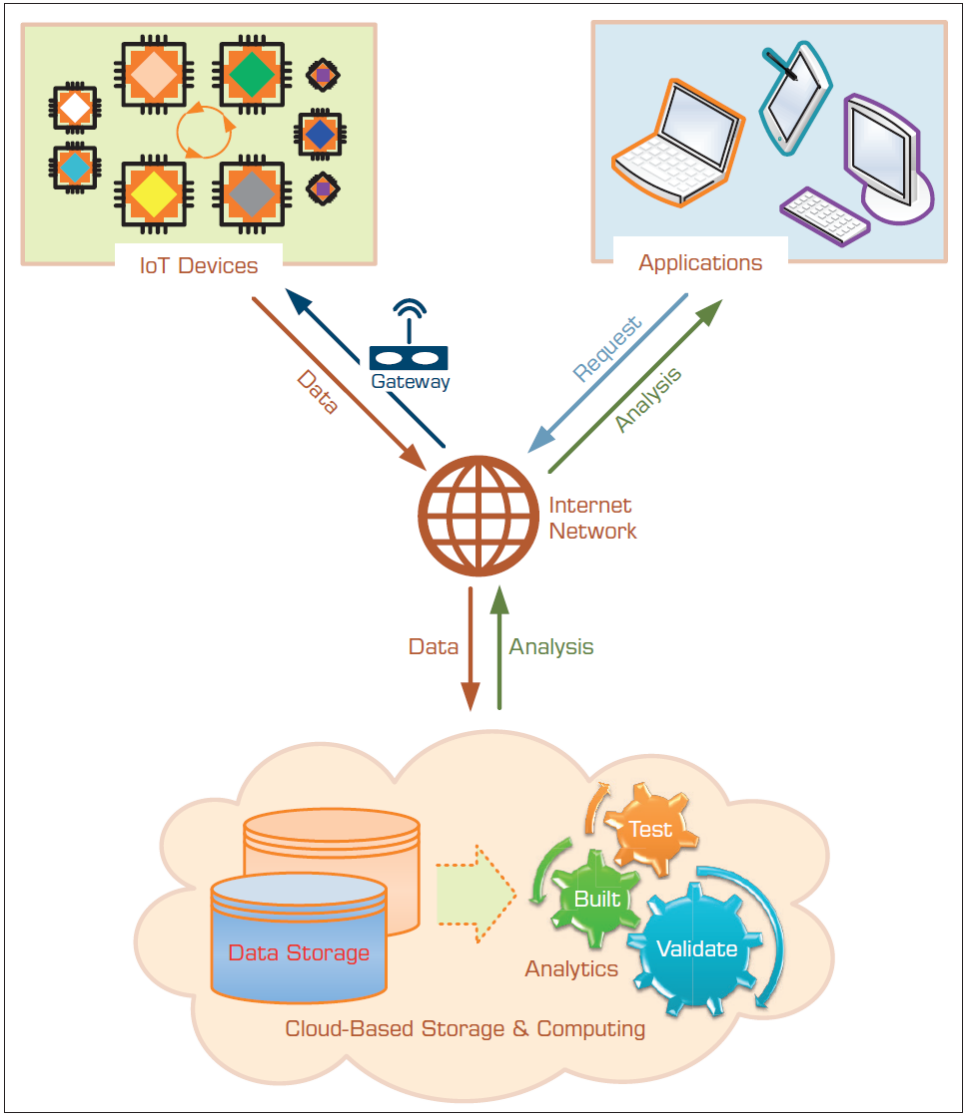
Questions for Discussion

1. What type of information would likely be collected by an oil and gas drilling platform?
2. Does this application fit the three V's of Big Data (volume, variety, velocity)? Why or why not?
3. Which other industries could use similar operational measurements and dashboards?

IoT Technology Infrastructure

- IoT related technology components can be divided into four major blocks:
  1. **Hardware**
     - physical devices, sensors, and actuators
  2. **Connectivity**
     - Collecting and sending sensory data to the cloud
  3. **Software**
     - Integrating, and processing data for patterns
  4. **Applications**
     - Creating context specific alerts, actionable insight

Building Blocks of IoT Technology Infrastructure



RFID Sensors (1 of 2)

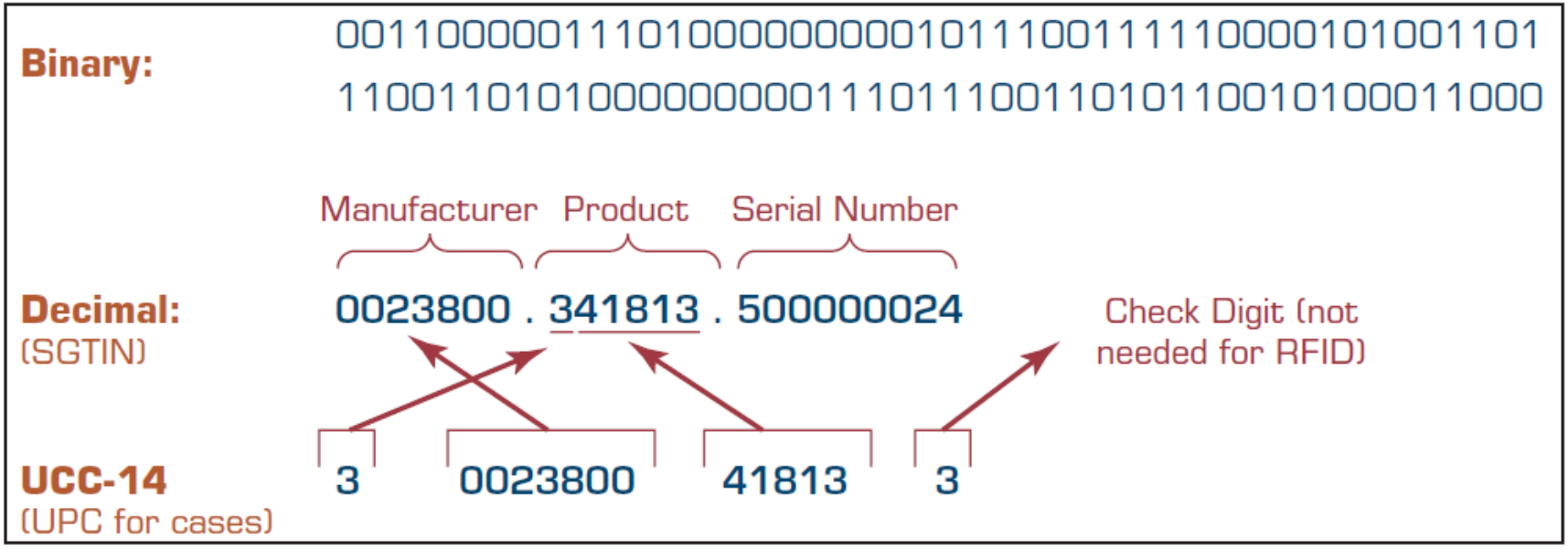
- RFID: radio-frequency identification
- One of the earliest/disruptive sensor technologies
- Part of a family of automatic identification technologies
  - Including ubiquitous barcodes and magnetic strips
- The goal is to use radio-frequency waves to accurately and quickly identify objects
- Use of RFID is led/promoted by large retailers
  - Wal-Mart, Target, Dillard's

RFID Sensors (2 of 2)

- How does RFID work?
  - **Tag** – a circuit attached to the product to be identified
  - **Interrogator** (i.e., reader) – with antennas and a computer to detect objects, store the data, and take due actions
- Tags can be passive or active
  - **Passive tag** – small, inexpensive, no power source
  - **Active tag** – larger, more expensive, has power source
- Which one is better?
  - Retail uses passive tags, others may use active tags
- RFID + Sensors can be used for perishable goods

Data Representation in RFID

- Data representation for a given application domain
- For Retail: Electronic Product Code (EPC)
- RFID tags contain 96 bits of data



Fog Computing

- Data produced by IoT is huge in size (problem)
- Fog computing is to address the issue by
  - Proposing fog nodes to process the data close to IoT
  - Fog nodes – any device including routers or switches

Table 8.1 Difference between Fog Nodes and a Cloud Platform

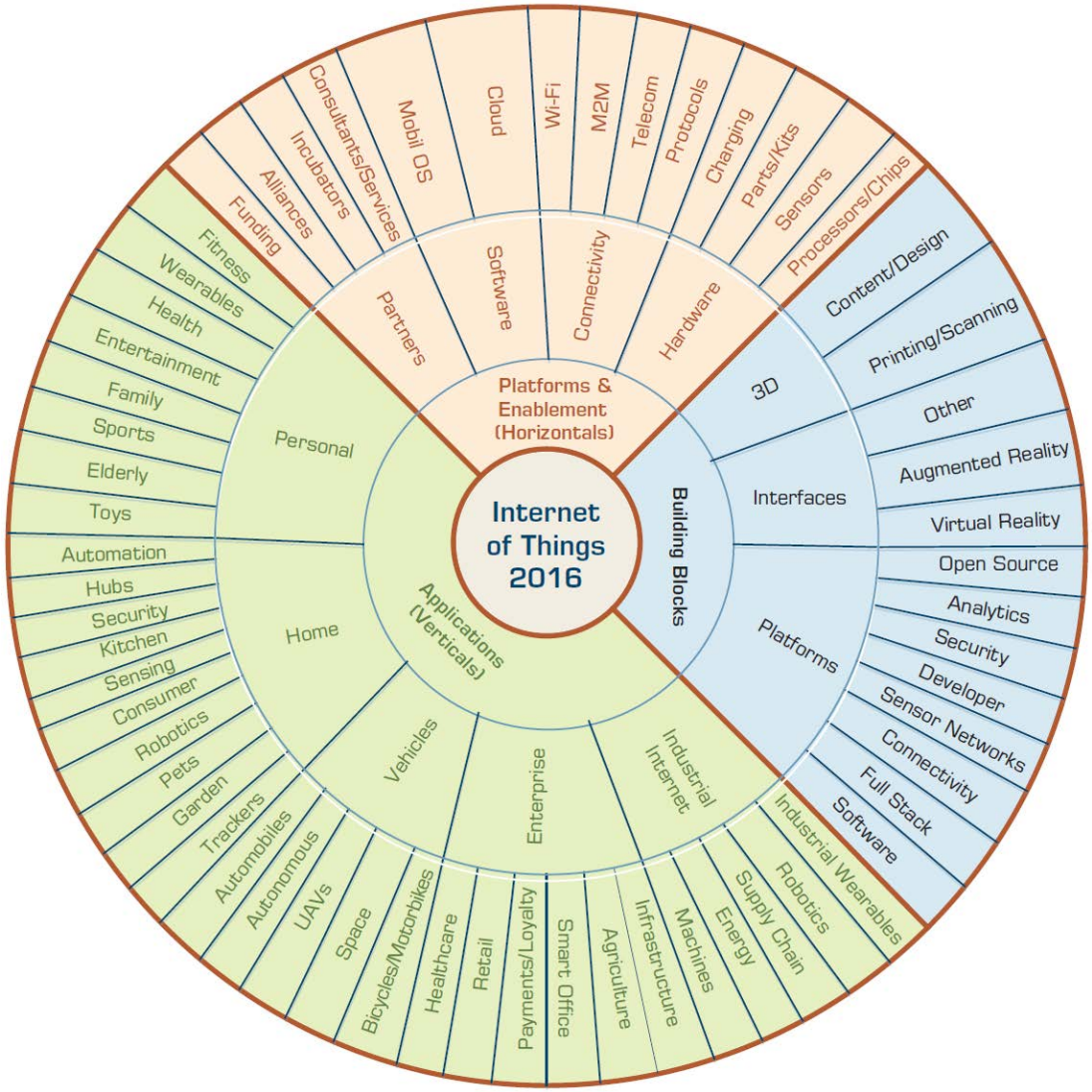
Fog Nodes	Cloud Platform
Receive data from IoT devices	Receives and aggregates data from fog nodes
Run IoT real-time analytics in millisecond response time	Analysis is performed on huge amounts of business data and can take hours or weeks

  - Data Center/Cloud → Fog Device → Physical Device (i.e., sensors that generate the data)

Other IoT Considerations

- IoT Platforms – many large companies are in it
  - Amazon AWS IoT, Microsoft Azure IoT Suite, Predix IoT, Platform by General Electric (GE), IBM Watson IoT solutions, and Teradata Unified Data Architecture
- IoT Start-up Ecosystem
  - Many start-up companies are emerging in the field of IoT
  - Examples include Sigfox, 3D Robotics, Canary, Athos, Greenwave, Jawbone, FreedomPop, Razer, and Ring
  - **Fitbit** – one of the most successful IoT startups
- See Figure 8.3 for a pictorial representation of IoT Ecosystem

Internet of Things (IoT) Ecosystem



Managerial Considerations in the Internet of Things

1. Organizational Alignment
2. Interoperability Challenges
3. Security

- Emerging growth of IoT and its potential to help us achieve the vision of **smart cities**, **smart grid**, smart **anything**

Cloud Computing and Business Analytics

- A style of computing in which dynamically scalable and often virtualized resources are provided over the Internet.
- Users need not have knowledge of, experience in, or control over the technology infrastructures in the cloud that supports them.
- Cloud computing = utility computing, application service provider grid computing, on-demand computing, software-as-a-service (SaaS), …
  - Cloud = Internet
  - Related “-as-a-services”: infrastructure-as-a-service (Iaa S), platforms-as-a-service (PaaS)

Cloud Computing Example (1 of 2)

- Web-based e-mail → cloud computing application
  - Stores the data (e-mail messages)
  - Stores the software (e-mail programs)
  - Centralized hardware/software/infrastructure
  - Centralized updates/upgrades
  - Access from anywhere via a Web browser
  - e.g., Gmail
- Web-based general application = cloud application
  - Google Docs, Google Spreadsheets, Google Drive,…
  - Amazon.com's Web Services

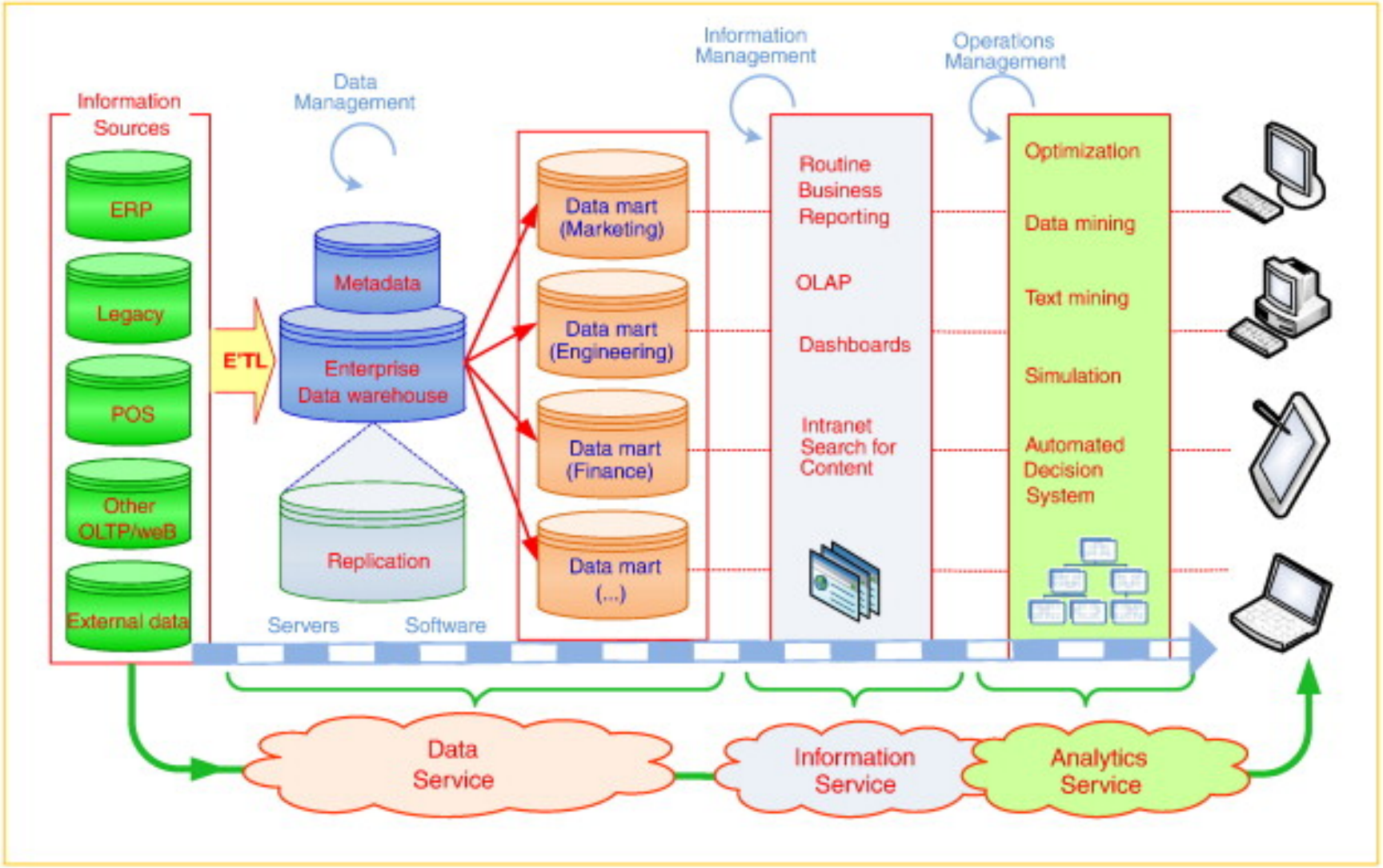
Cloud Computing Example (2 of 2)

- Cloud computing is used in
  - e-commerce, BI, CRM, SCM, …
- Business model
  - Pay-per-use
  - Subscribe/pay-as-you-go
- Companies that offer cloud-computing services
  - Google, Yahoo!, Salesforce.com
  - IBM, Microsoft (Azure)
  - Sun Microsystems/Oracle

Cloud Computing and Service-Oriented Thinking

- Service-oriented thinking is one of the fastest-growing paradigms today
- Toward building agile data, information, and analytics capabilities as services
- Service orientation + DSS/BI
- Component-based service orientation fosters
  - Reusability, Substitutability, Extensibility, Scalability, Customizability, Reliability, Low Cost of Ownership, Economy of Scale,…

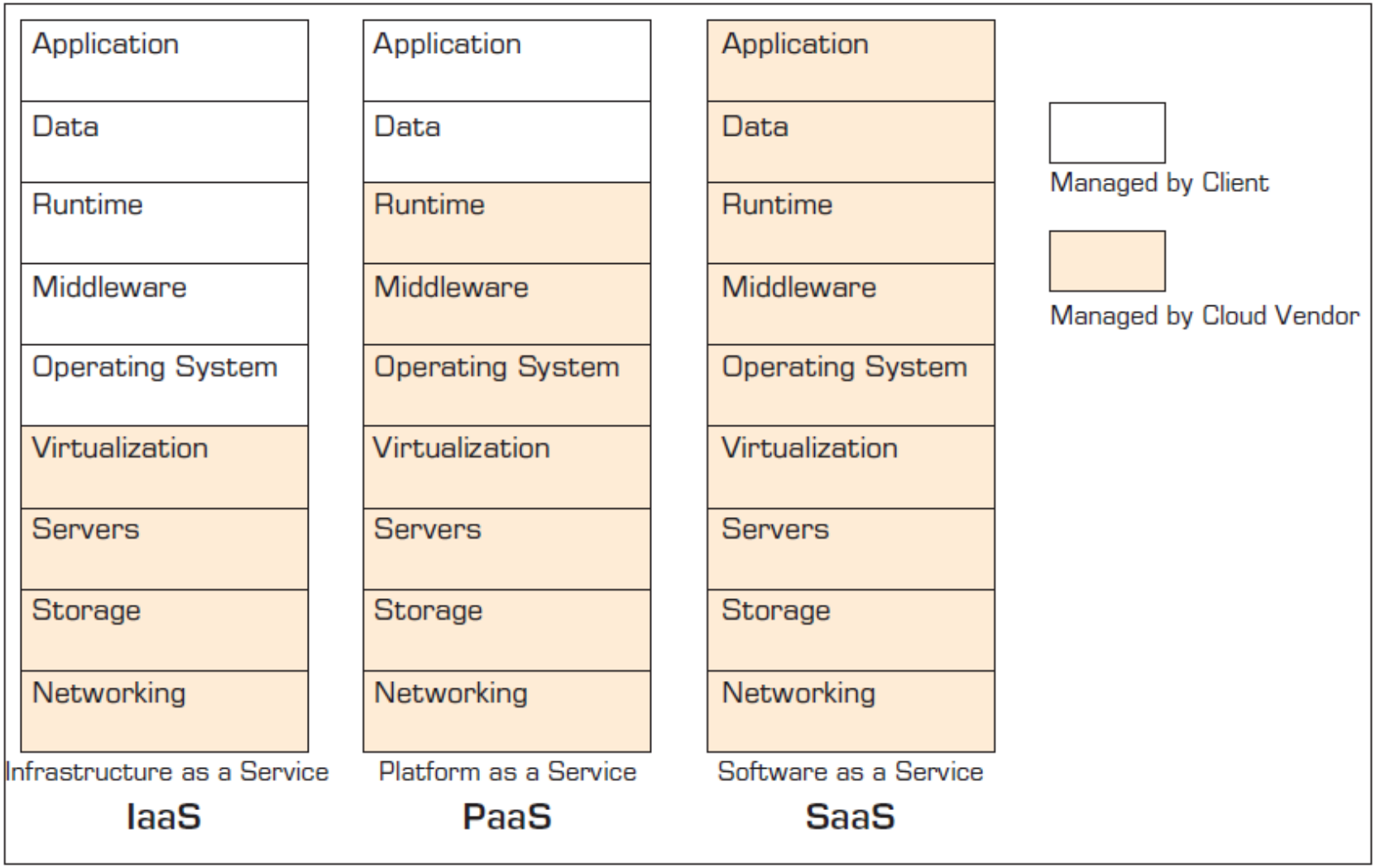
Service-Oriented DSS/BI



Variations of Service-Oriented Architecture and the Cloud

- Data as a Service (DaaS)
- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)
- …
- Why so many .aaS is emerging?

Different Types of Cloud Offerings



Essential Technologies for Cloud Computing

- Virtualization
  - Creation of a virtual version of something like an operating system or server
  - Example: logical division of a hard drive to create two separate hard drives in a computer
- Levels of virtualization
  - Network virtualization
  - Storage virtualization
  - Server virtualization
- Relates to which cloud service is employed

Cloud Deployment Models

- Private cloud
- Public cloud
- Hybrid cloud
- Which cloud model is good for you?
- Major cloud platform providers in analytics:
  - Amazon Elastic Beanstalk
  - I B M Bluemix
  - Microsoft Azure
  - Google App Engine
  - OpenShift

Representative Analytics as a Service Offering

- Teradata - Aster Analytics as a Service
- IBM Watson Analytics
- MineMyText.com
- SAS Visual Analytic and Visual Statistics
- Tableau
- Showflake
- Predix by General Electric

→ **Most of these have free/restricted/trial offerings**

Illustrative Analytics Applications Employing the Cloud Infrastructure (1 of 2)

- MD Anderson Cancer Center Utilizes Cognitive Computing Capabilities of IBM Watson to Give Better Treatment to Cancer Patients
- Public School Education in Tacoma, Washington, Uses Microsoft Azure Machine Learning to Predict School Dropouts
- Dartmouth-Hitchcock Medical Center Provides Personalized Proactive Healthcare Using Microsoft Cortana Analytics Suite

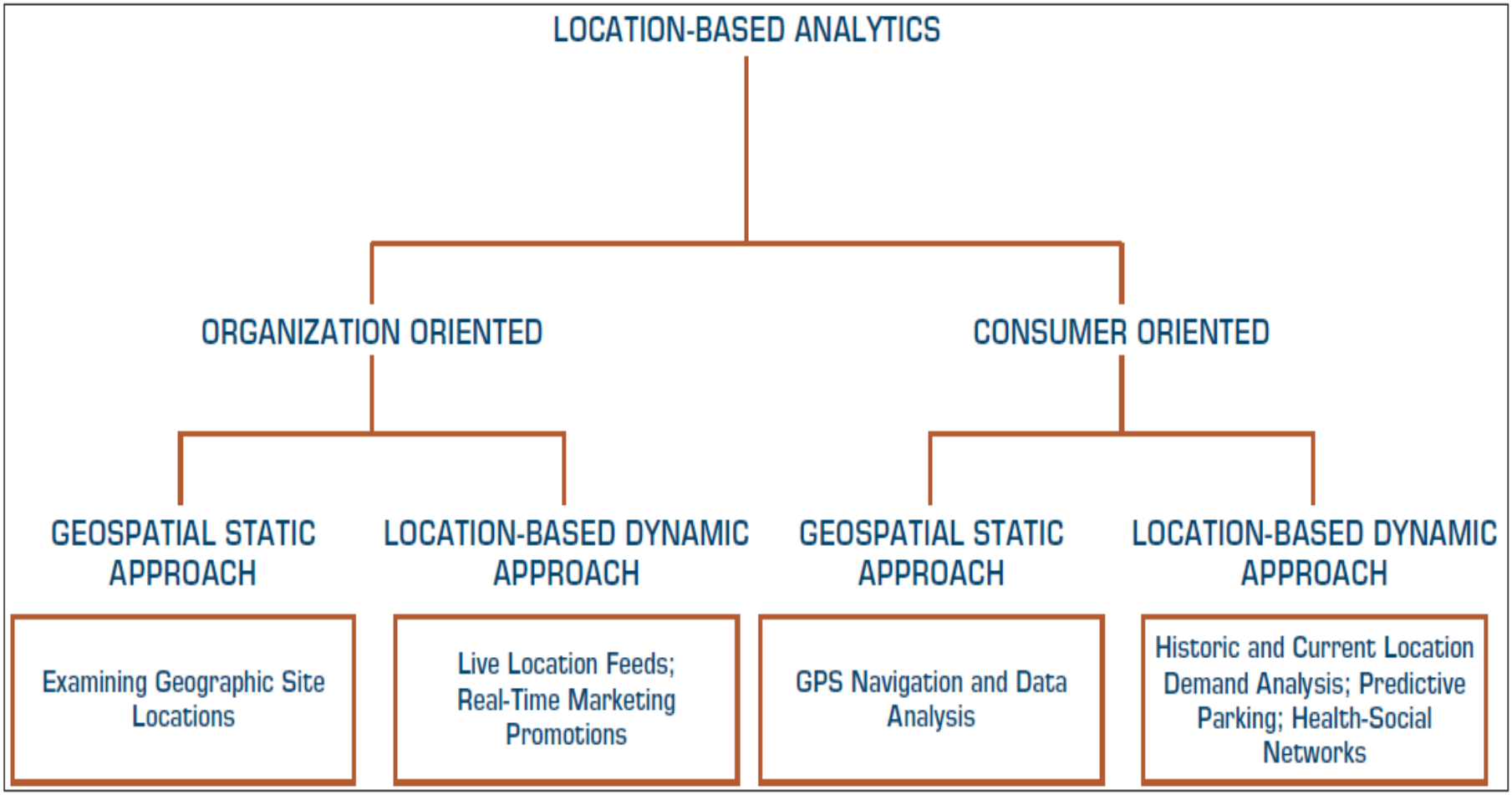
Illustrative Analytics Applications Employing the Cloud Infrastructure (2 of 2)

- Mankind Pharma Uses IBM Cloud Infrastructure to Reduce Application Implementation Time by 98%
- Gulf Air Uses Big Data to Get Deeper Customer Insight
- Chime Enhances Customer Experience Using Snowflake

Location-Based Analytics (1 of 3)

- Geospatial Analytics
- Geocoding
  - Visual maps
  - Postal codes
  - Latitude & Longitude
- Enables aggregate view of a large geographic area
- Integrate “where” into customer view

Location-Based Analytics (2 of 3)



Location-Based Analytics (3 of 3)

- Location-based databases
- Geographic Information System (GIS)
  - Used to capture, store, analyze, and manage the data linked to a location
  - Combined with integrated sensor technologies and global positioning systems (GPS)
- Location Intelligence (LI)?
  - Interactive maps that further drill down to data/information details about any location

Use of Location-Based Analytics

- Retailers – location + demographic details combined with other transactional data can help …
  - determine how sales vary by population level
  - assess locational proximity to other competitors and their offerings
  - assess the demand variations and efficiency of supply chain operations
  - analyze customer needs and complaints
  - better target different customer segments
  - …

GIS Applications

- In addition to business/retail applications, GIS based analytics are being used in
  - Agricultural applications
  - Crime analysis
  - Disease spread prediction
- For more applications, look at
  - esri.com (producer of ArcGIS)
  - grindgis.com
- LI can be combined with weather and environmental data to create a richer data/information infrastructure

Application Case 8.4

Great Clips Employs Spatial Analytics to Shave Time in Location Decisions

Questions for Discussion

1. How is geospatial analytics employed at Great Clips?
2. What criteria should a company consider in evaluating sites for future locations?
3. Can you think of other applications where such geospatial data might be useful?

Application Case 8.5

Starbucks Exploits GIS and Analytics to Grow Worldwide

Questions for Discussion

1. What type of demographics and GIS information would be relevant for deciding on a store location?
2. It has been mentioned that Starbucks encourages its customers to use its mobile app. What type of information might the company gather from the app to help it better plan operations?
3. Will the availability of free Wi-Fi at Starbucks' stores provide any information to Starbucks for better analytics?

A Multimedia Exercise in Analytics Employing Geospatial Analytics

- Go To Teradata University Network (TUN)
- Find the BSI Case video on “The Case of the Dropped Mobile Calls”
- **Watch the video via TUN or at YouTube** youtube.com/watch?v=4WJR_Z3exw4
- **Also, look at the slides at** slideshare.net/teradata/bsi-teradata-the-case-of-the-dropped-mobile-calls
- Discuss the case

Real-Time Location Intelligence

- Many devices are constantly sending out their location information
  - Cars, airplanes, ships, mobile phones, cameras, navigation systems, …
    - GPS, Wi-Fi, RFID, cell tower triangulation
- Reality mining?
  - Real-time location information = real-time insight
  - Path Intelligence (pathintelligence.com)
    - Footpath – movement patterns within a city or store
    - How to use such movement information

Application Case 8.6

Quiznos Targets Customers for Its Sandwiches

Questions for Discussion

1. How can location-based analytics help retailers in targeting customers?
2. Research similar applications of location-based analytics in the retail domain.

Analytics Applications for Consumers

- Explosive growth of the apps industry
  - i O S, Android, Windows, Blackberry, Amazon, …
  - Directly used by consumers (not businesses)
  - Enabling consumers to become more efficient
  - Interesting Examples
    - CabSense – finding a taxi in New York City
      - Rating of street corners; interactive maps, …
    - Park P G H – finding a parking spot
      - Downtown Pittsburgh, Pennsylvania
  - App industry is already 25B in size and growing (wsj.com/apps)

Issues of Legality, Privacy, and Ethics (1 of 3)

- **Legal** issues to consider
  - What is the value of an expert opinion in court when the expertise is encoded in a computer?
  - Who is liable for wrong advice (or information) provided by an intelligent application?
  - What happens if a manager enters an incorrect judgment value into an analytic application?
  - Who owns the knowledge in a knowledge base?
  - Can management force experts to contribute their expertise?

Issues of Legality, Privacy, and Ethics (2 of 3)

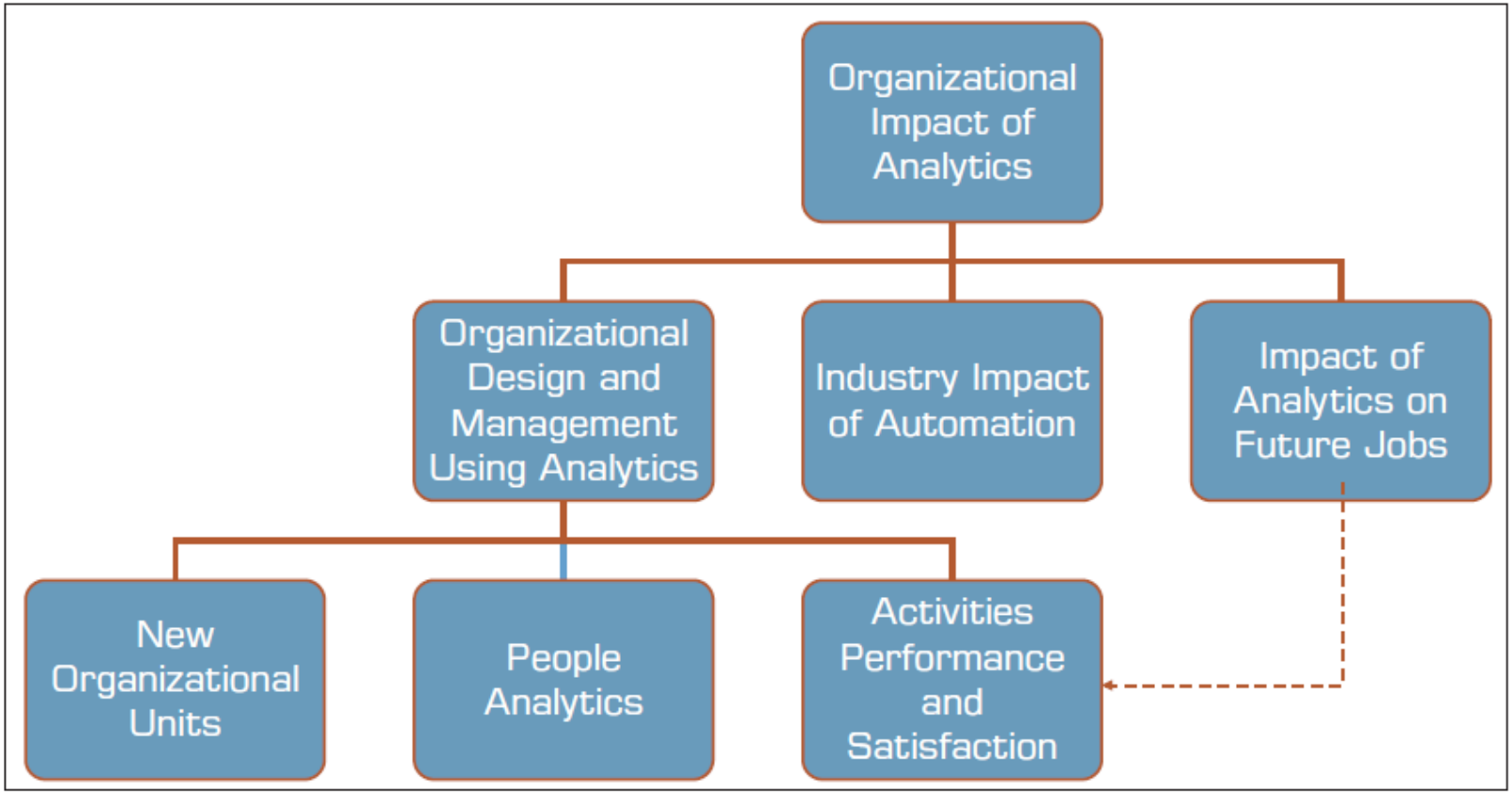
- **Privacy -** The right to be left alone and the right to be free from unreasonable personal intrusions
  - Collecting information about individuals
    - How much is too much?
  - Mobile User Privacy
    - Location-based analysis/profiling
  - Homeland Security and Individual Privacy
  - Recent Issues in Privacy and Analytics
    - "What They Know" about you (wsj.com/wtk)
    - Rapleaf (rapleaf.com), X + 1 (xplusone.com), Bluecava (bluecava.com), reputation.com, sociometric.com…
  - Who owns our private data?

Issues of Legality, Privacy, and Ethics (3 of 3)

- **Ethics i**n Decision Making and Support
  - Electronic surveillance
  - Software piracy
  - Invasion of individuals' privacy
  - Use of proprietary databases
  - Use of knowledge and expertise
  - Accessibility for workers with disabilities
  - Accuracy of data, information, and knowledge
  - Protection of the rights of users
  - Accessibility to information
  - Personal use of corporate computing resources
  - … more ….

Impacts of Analytics in Organizations (1 of 2)

- Analytics revolution → Cultural transformation



Impacts of Analytics in Organizations (2 of 2)

- New Organizational Units
  - BI department
  - Data science department
- Redesign of an Organization through the Use of Analytics
  - People analytics
  - HR analytics
- Analytics Impact on Managers' Activities, Performance, and Job Satisfaction
  - Data/fact/analytics driven decision

Potential Impacts of Analytics on Managers

- Less expertise/experience is requirement
- Faster decision making (augmented with analytics)
- Less reliance on experts and analysts (data rules!)
- Power is being redistributed among managers
- Support for complex decisions makes them faster to develop and be of better quality
- Information needed for high-level decision making is expedited or even self-generated
- Automation of routine decisions or phases in the decision-making process may eliminate some managers

Impacts of Analytics in Organizations

- Industrial Restructuring
  - AI, analytics, and cognitive computing can change the industry in a fundamental way
- Automation's Impact on Jobs
  - Data science and AI will change the nature of human jobs (another wave of automation is in the horizon)
- Unintended Effects of Analytics
  - Social and long-term effects of the models
  - "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy"

Data Scientist as a Profession

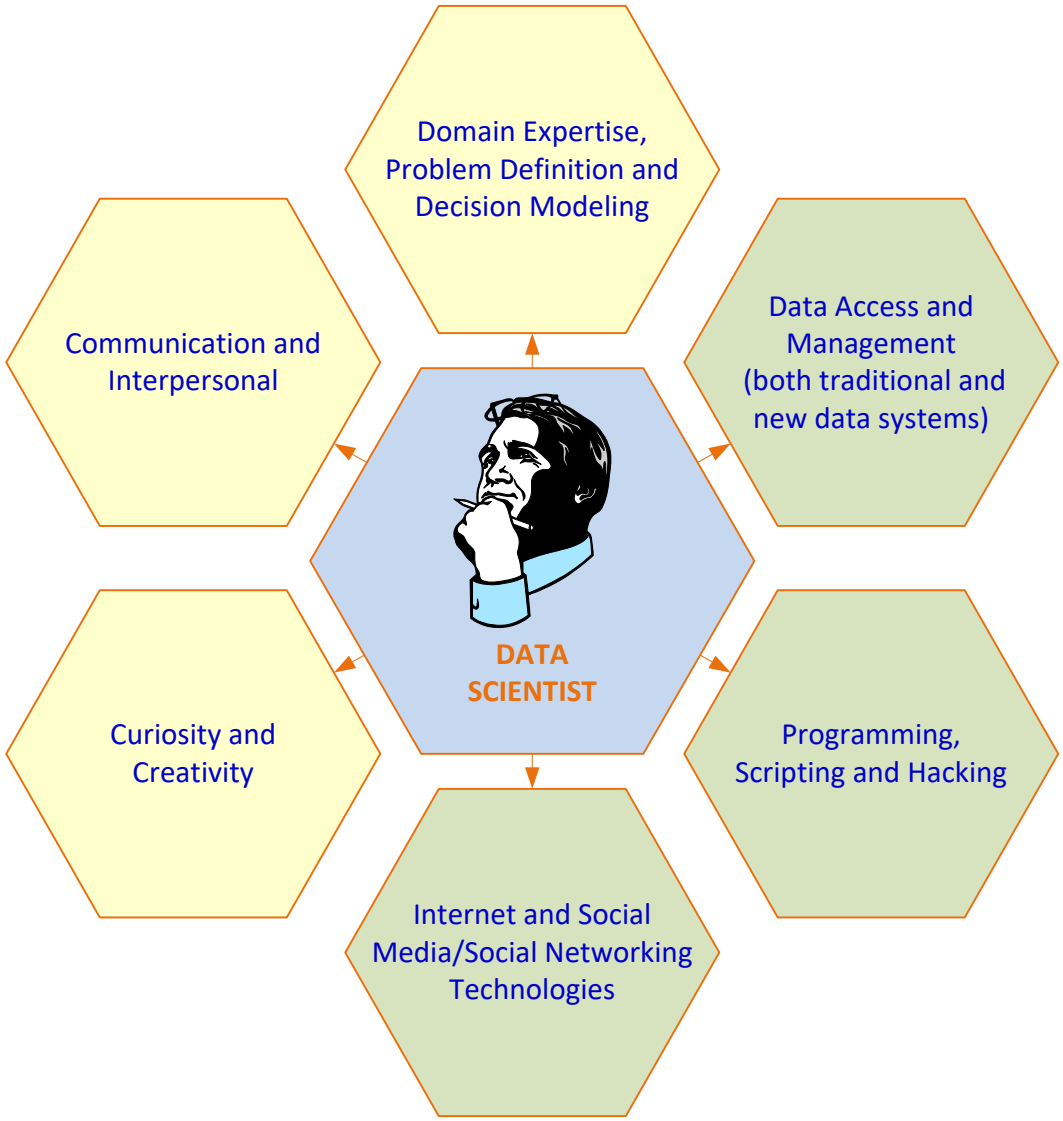
"The Sexiest Job of the 21st Century"

Thomas H. Davenport and D. J. Patil

Harvard Business Review, October 2012

- Data Scientist = Big Data guru
  - One with skills to **investigate** Big Data
- Very high salaries, very high expectations
- Where do Data Scientists come from?
  - M.S./Ph.D. in MIS, CS, IE,… and/or Analytics
  - There is not a specific degree program for DS!
  - PE, PML, … DSP (Data Science Professional)

Skills That Define a Data Scientist



A Typical Job Post for Data Scientist

TECHNOLOGY INSIGHTS 8.1
A Typical Job Post for Data Scientists

[Some company] is seeking a Data Scientist to join our Big Data Analytics team. Individuals in this role are expected to be comfortable working as a software engineer and a quantitative researcher. The ideal candidate will have a keen interest in the study of an online social network and a passion for identifying and answering questions that help us build the best products.

Responsibilities

- Work closely with a product engineering team to identify and answer important product questions
- Answer product questions by using appropriate statistical techniques on available data
- Communicate findings to product managers and engineers
- Drive the collection of new data and the refinement of existing data sources
- Analyze and interpret the results of product experiments
- Develop best practices for instrumentation and experimentation and communicate those to product engineering teams

Requirements

- MS or PhD in a relevant technical field, or 4+ years of experience in a relevant role
- Extensive experience solving analytical problems using quantitative approaches
- Comfort with manipulating and analyzing complex, high-volume, high-dimensionality data from varying sources
- A strong passion for empirical research and for answering hard questions with data
- A flexible analytic approach that allows for results at varying levels of precision
- Ability to communicate complex quantitative analysis in a clear, precise, and actionable manner
- Fluency with at least one scripting language such as Python or PHP
- Familiarity with relational databases and SQL
- Expert knowledge of an analysis tool such as R, Matlab, or SAS
- Experience working with large data sets, experience working with distributed computing tools a plus (MapReduce, Hadoop, Hive, etc.)

Statements about Data Scientists

- Data scientists turn Big Data into big value, delivering products that delight users and insight that informs business decisions.
- A data scientist is not only proficient in working with data, but also appreciates data itself as an invaluable asset.
- By 2020 there will be 4.5 million new data scientist jobs, of which only one-third will be filled because of the lack of people available to fill them.
- Today’s data scientists are the quants of the financial markets of the 1980s.