Trends in "Big Data" Analytics

Philip J Cwynar

University of Pittsburgh
School of Information Sciences

pcwynar@pitt.edu

Outline

- “Big Data” job prospects
- Trends in “Big Data” analytics

“Big Data” Job Prospects

"Big Data" job prospects

"Data scientist" called the "sexiest job of the 21st century"
http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1

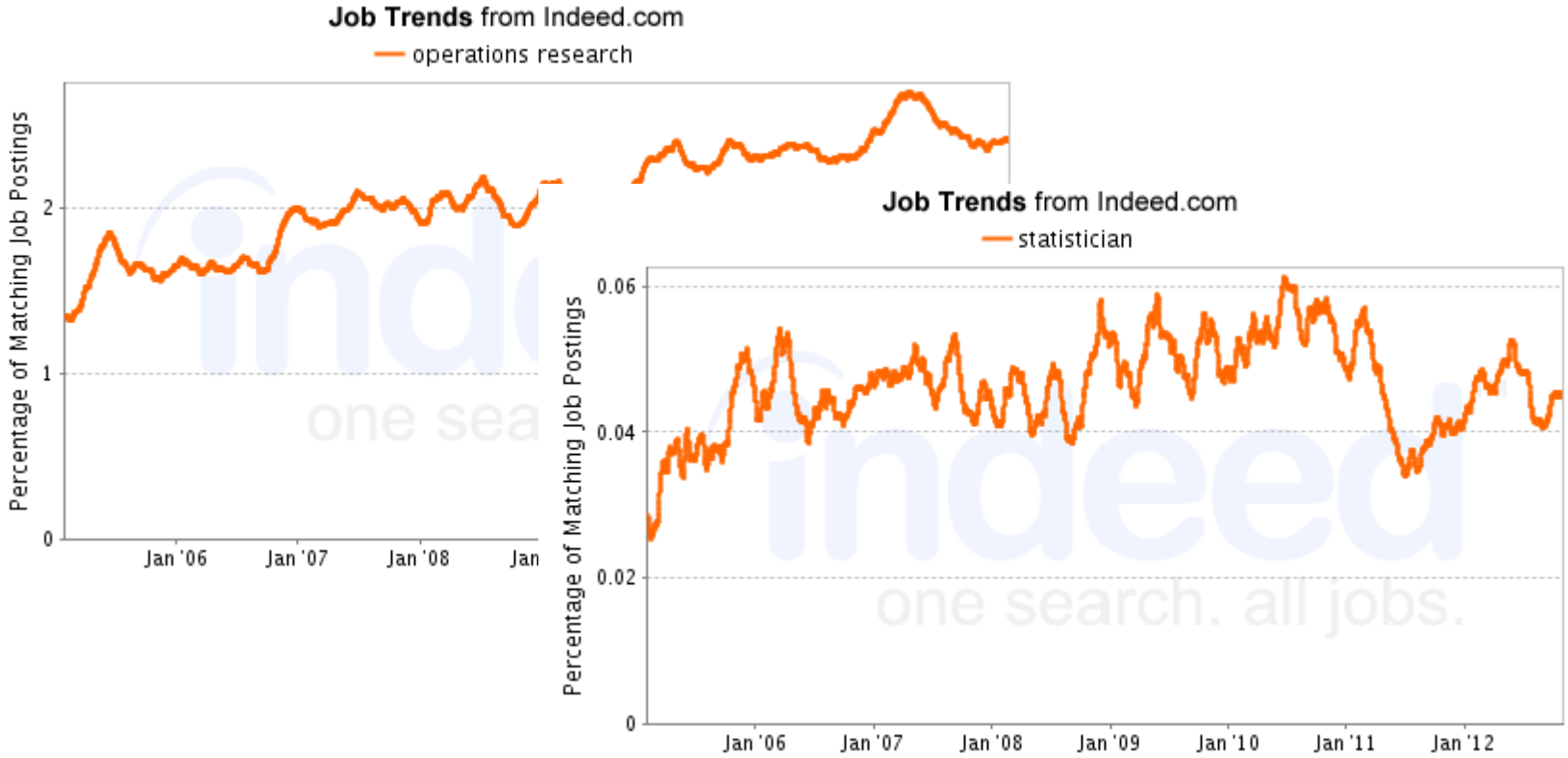
"Big Data" job prospects

In a Computerworld article published in November of 2012, the IT employment firm Gartner estimated that 4.4 million IT jobs will be created in the area of big data between now and 2015. However, Gartner's head of research, Peter Sondergaard, notes a serious shortage of IT professionals with big-data skills: "There is not enough talent in the industry" and that only one-third of the new jobs will be filled.

Why will there be such a significant need for Big Data Analysts and specialists? Because every industry sector and service entity has to deal with Big Data or can benefit from corralling the power of so much information. Obviously, those who work in data-rich disciplines such as astronomy or fields including online retail would depend on the tools and technologies in Big Data management. However, digital data is everywhere and employers from a wide range of sectors (healthcare, finance, place-based retail, manufacturing, and transportation, to name just a few) will be looking to build workforce capacity to enhance their productivity and competitive position in global markets.

"Big Data" job prospects: Traditional analytics

Job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)

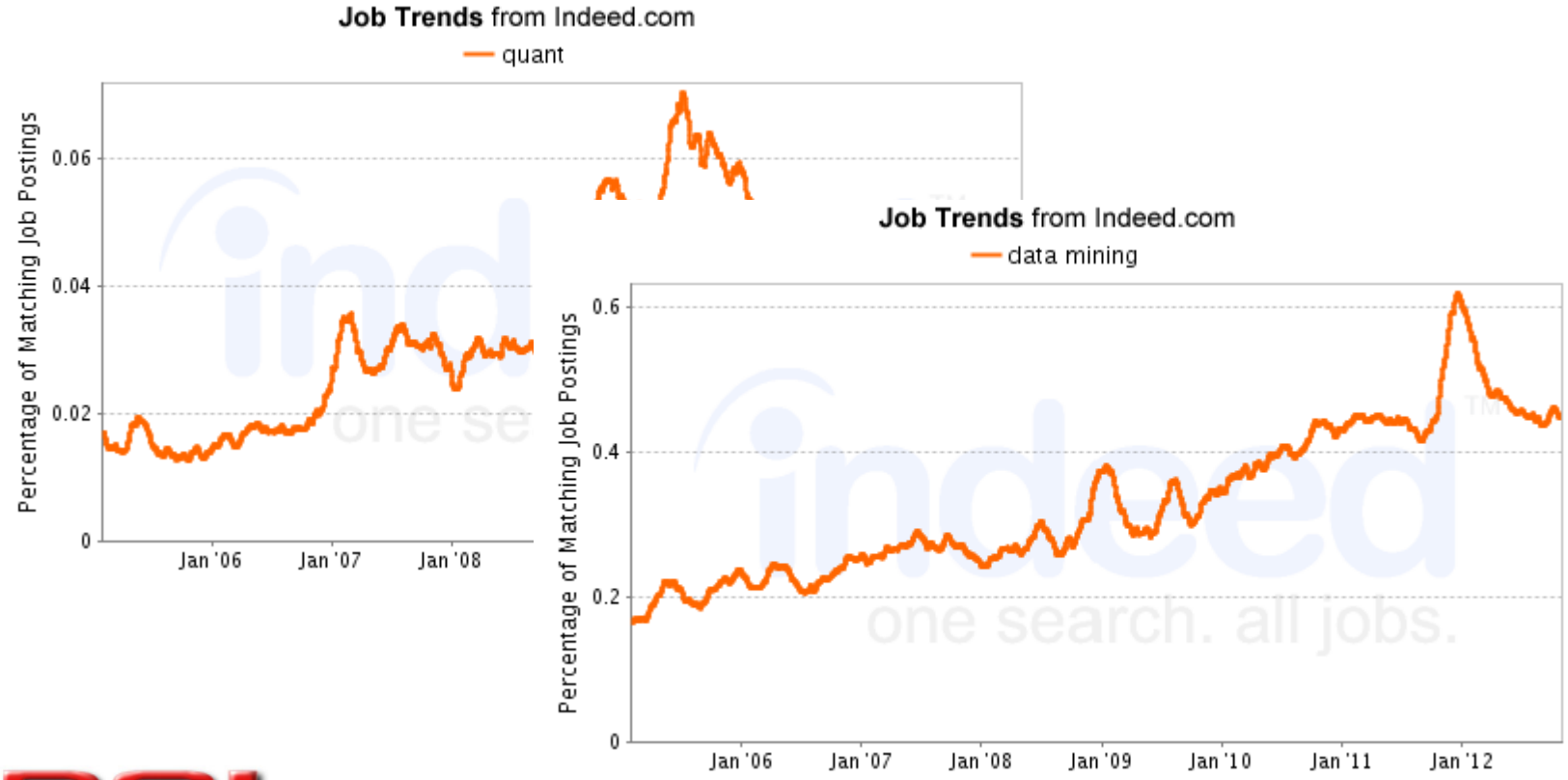


"Big Data" job prospects: Technology

More of the job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)

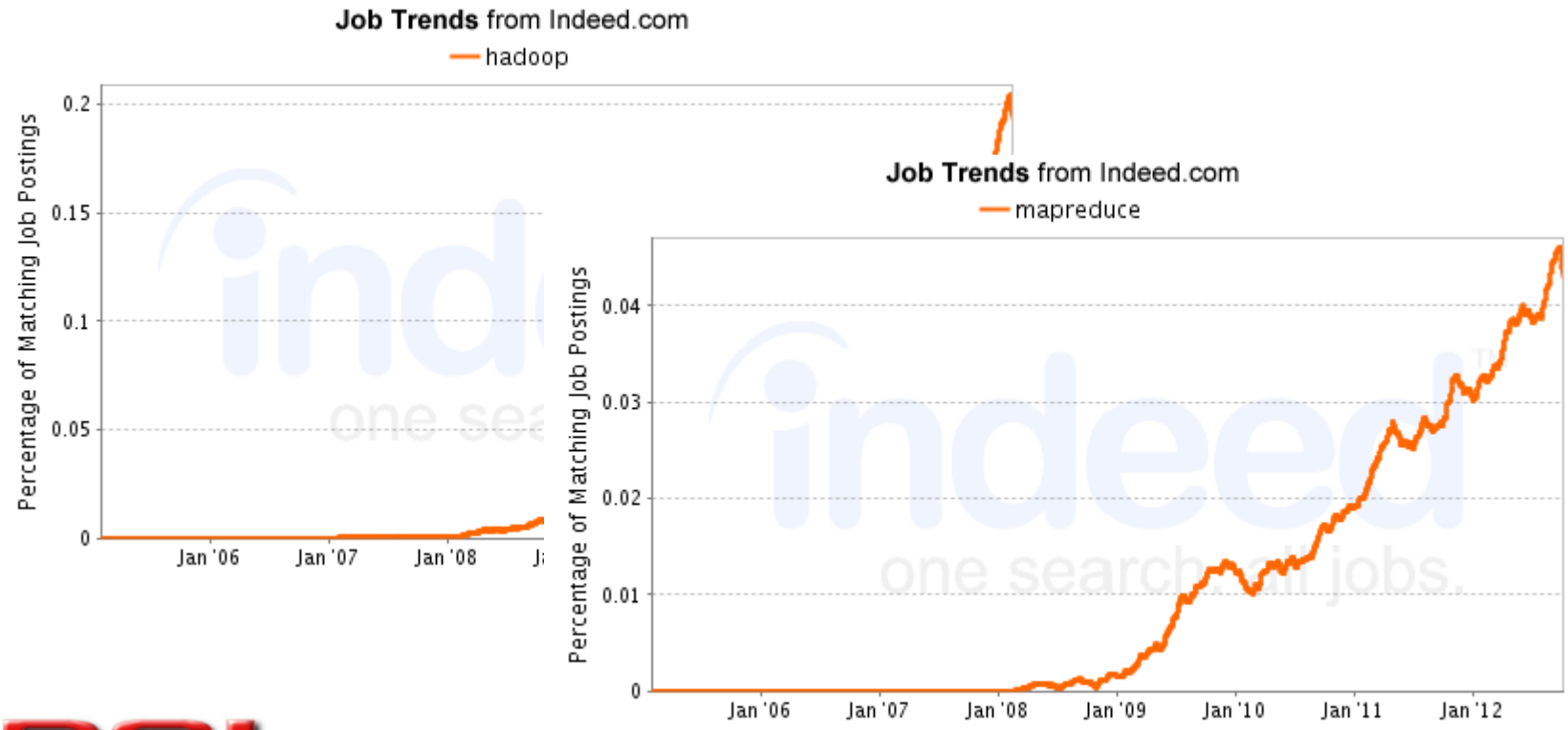
"Big Data" job prospects: Old "buzzwords"

More of the job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)



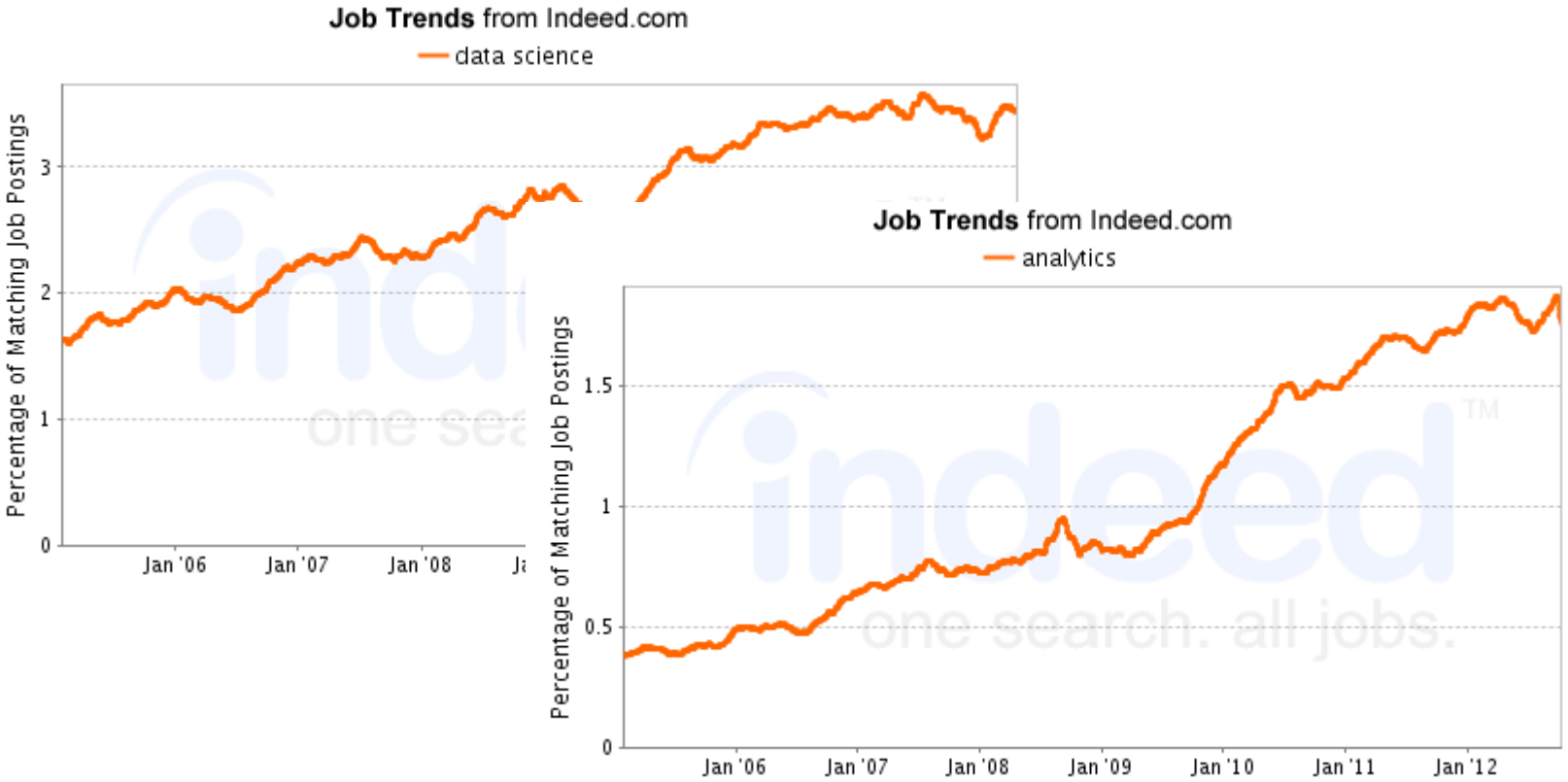
"Big Data" job prospects: "Big Data" technology

More of the job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)



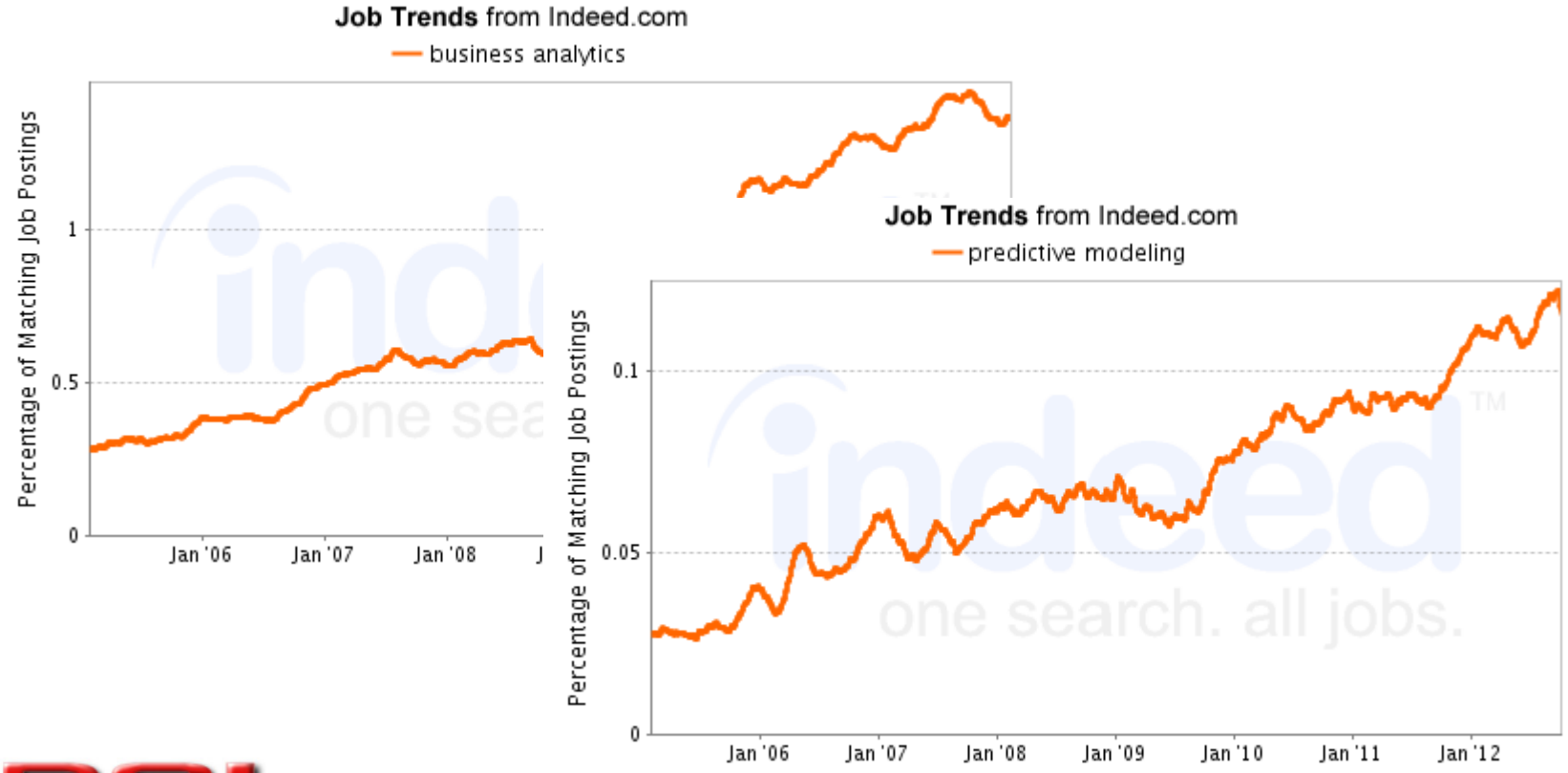
"Big Data" job prospects: Buzzwords

More of the job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)



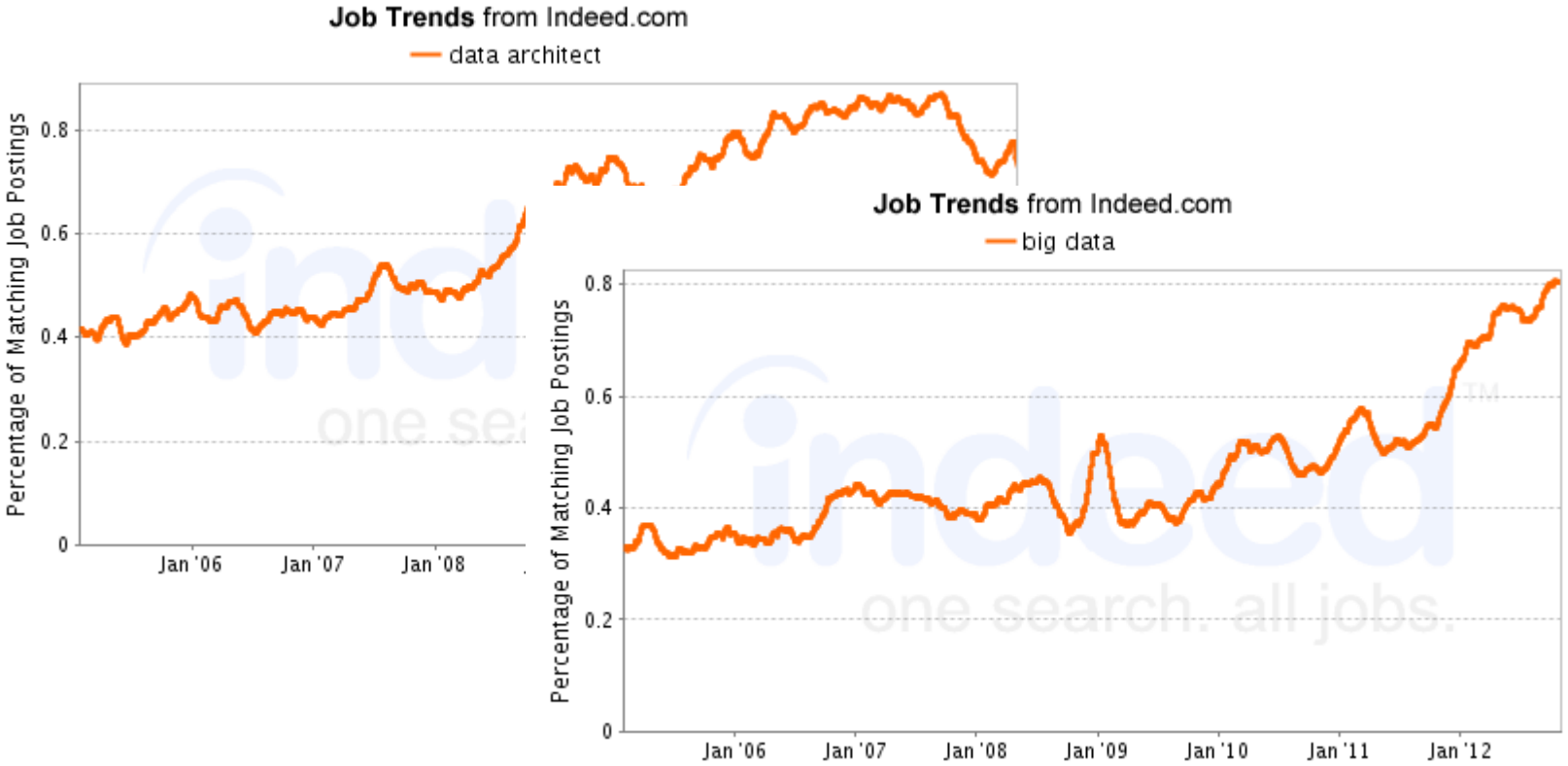
"Big Data" job prospects: Buzzwords

More of the job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)



"Big Data" job prospects: Buzzwords

More of the job trends (
http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/job-trends-for-data-science-big-data-web-analytics-etc)



"Big Data" job prospects: Salary levels

Salary levels for selected analytical job titles (http://www.analyticbridge.com/group/salary-trends-and-reports/forum/topics/salary-report-for-selected-analytical-job-titles)

Senior Data Architect	$121,000
Director of Analytic	$113,000
Quantitative Analyst	$96,000
SEO Manager	$78,000
Web Analytics Specialist	$77,000
Web Analyst	$67,000
Predictive Analytics Expert	$61,000
Marketing Analyst	$54,000

Trends in “Big Data” Analytics

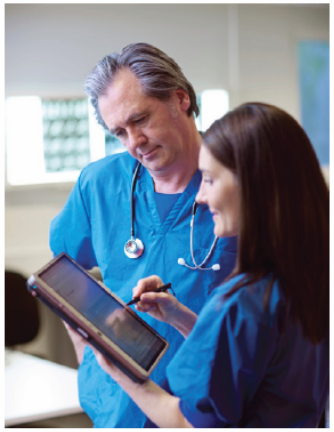
Trends in "Big Data"

Reading:

"Big data: The next frontier for innovation, competition, and productivity."

June 2011, Free report available at:
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McKinsey Global Institute

June 2011

McKinsey & Company, Inc.: a global management consulting firm focusing on solving issues of concern to senior management; advises many businesses, governments, and institutions.

McKinsey Global Institute: an economics "think tank" on globalization, corporate strategy and governance established in 1990

Trends in "Big Data:" Key insights

(1) Data have swept into every industry and business function and are now an important factor of production

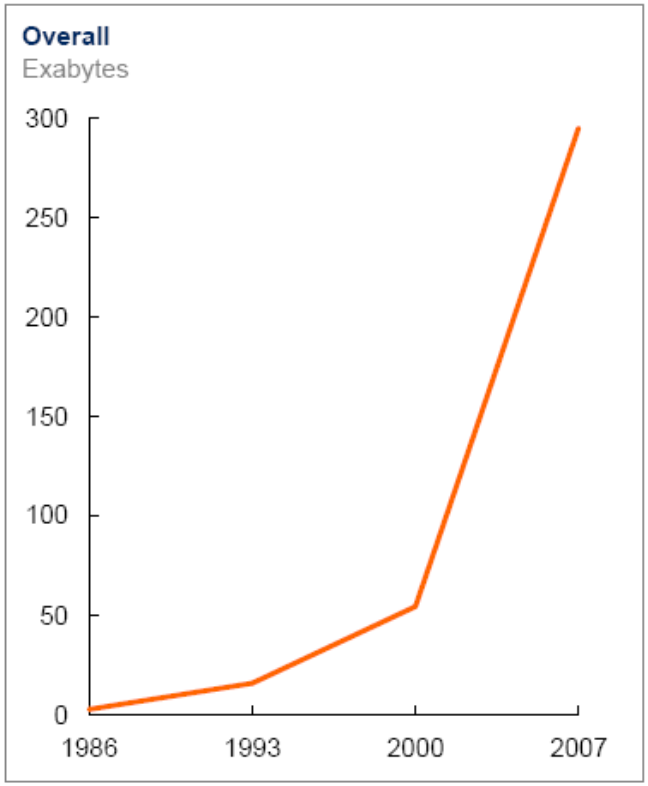
Big data has now reached every sector in the global economy.

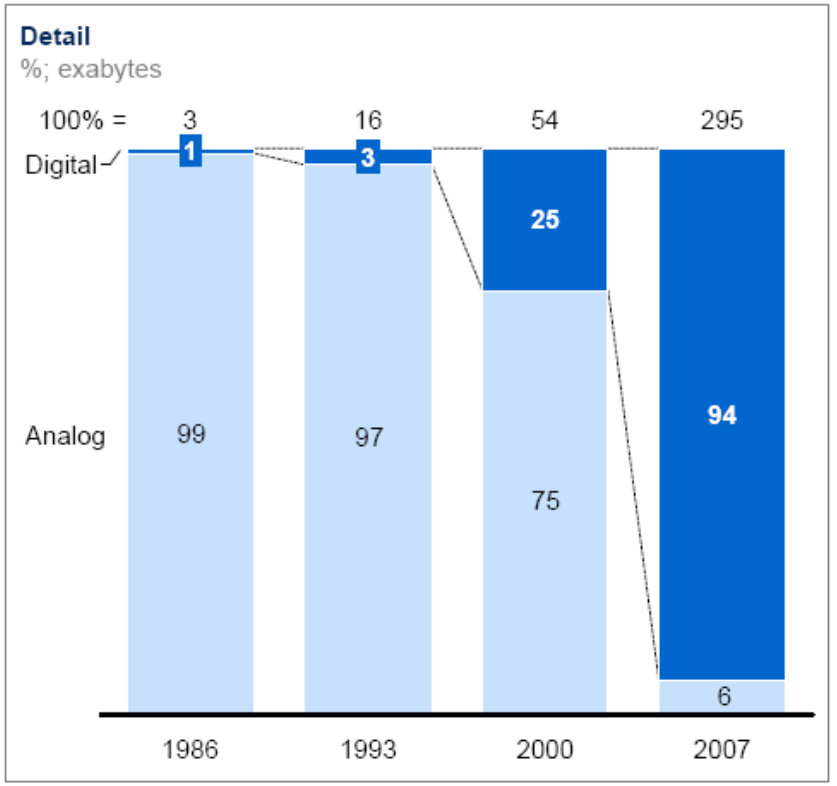
The possibilities of big data continue to evolve rapidly, driven by innovation in the underlying technologies, platforms, and analytic capabilities for handling data, as well as the evolution of behavior among its users as more and more individuals live digital lives.

Trends in "Big Data:" Key insights

Data storage has grown significantly, shifting markedly from analog to digital after 2000

Global installed, optimally compressed, storage





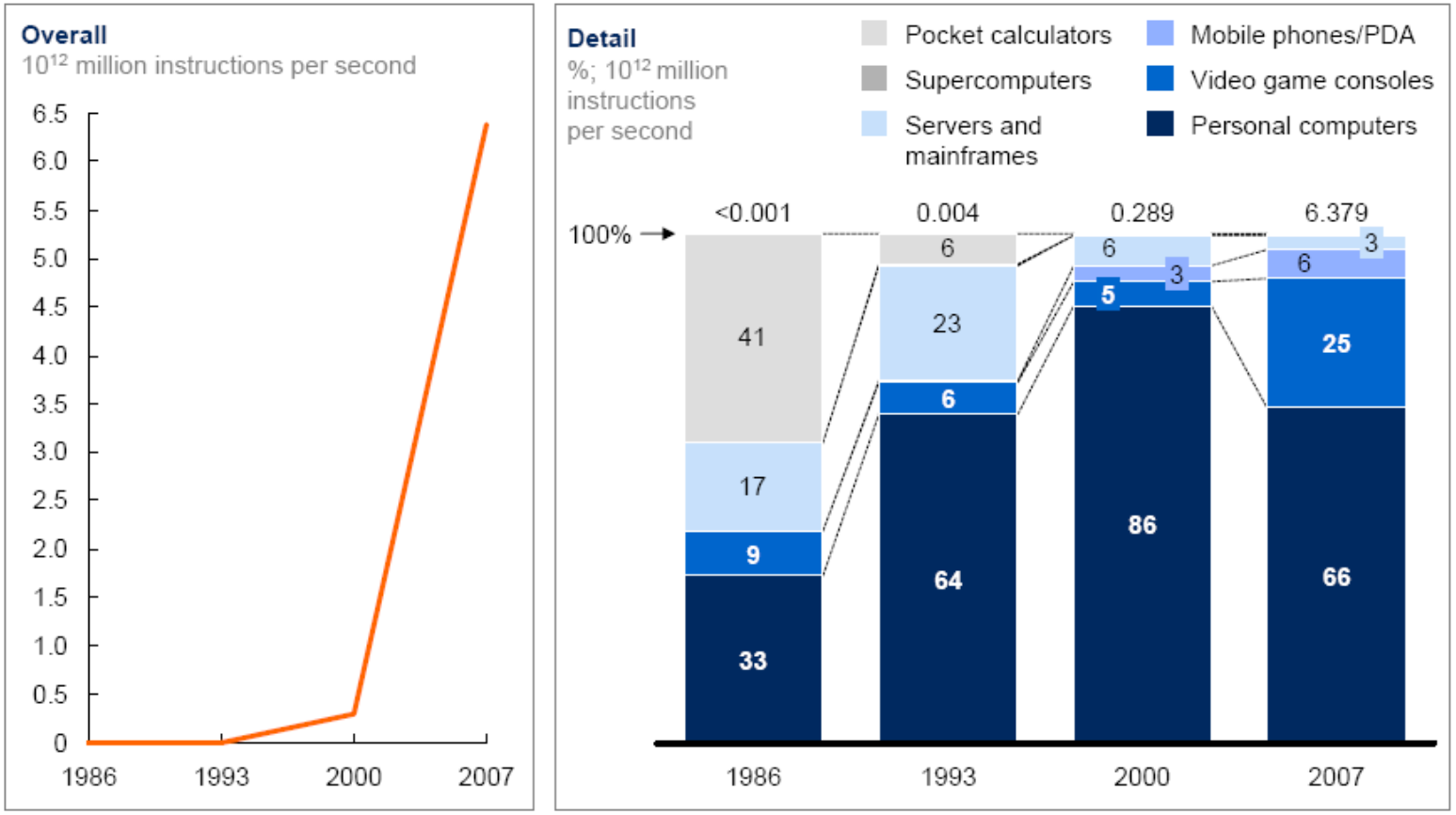
NOTE: Numbers may not sum due to rounding.

SOURCE: Hilbert and López, "The world's technological capacity to store, communicate, and compute information," *Science*, 2011

Trends in "Big Data:" Key insights

Computation capacity has also risen sharply

Global installed computation to handle information

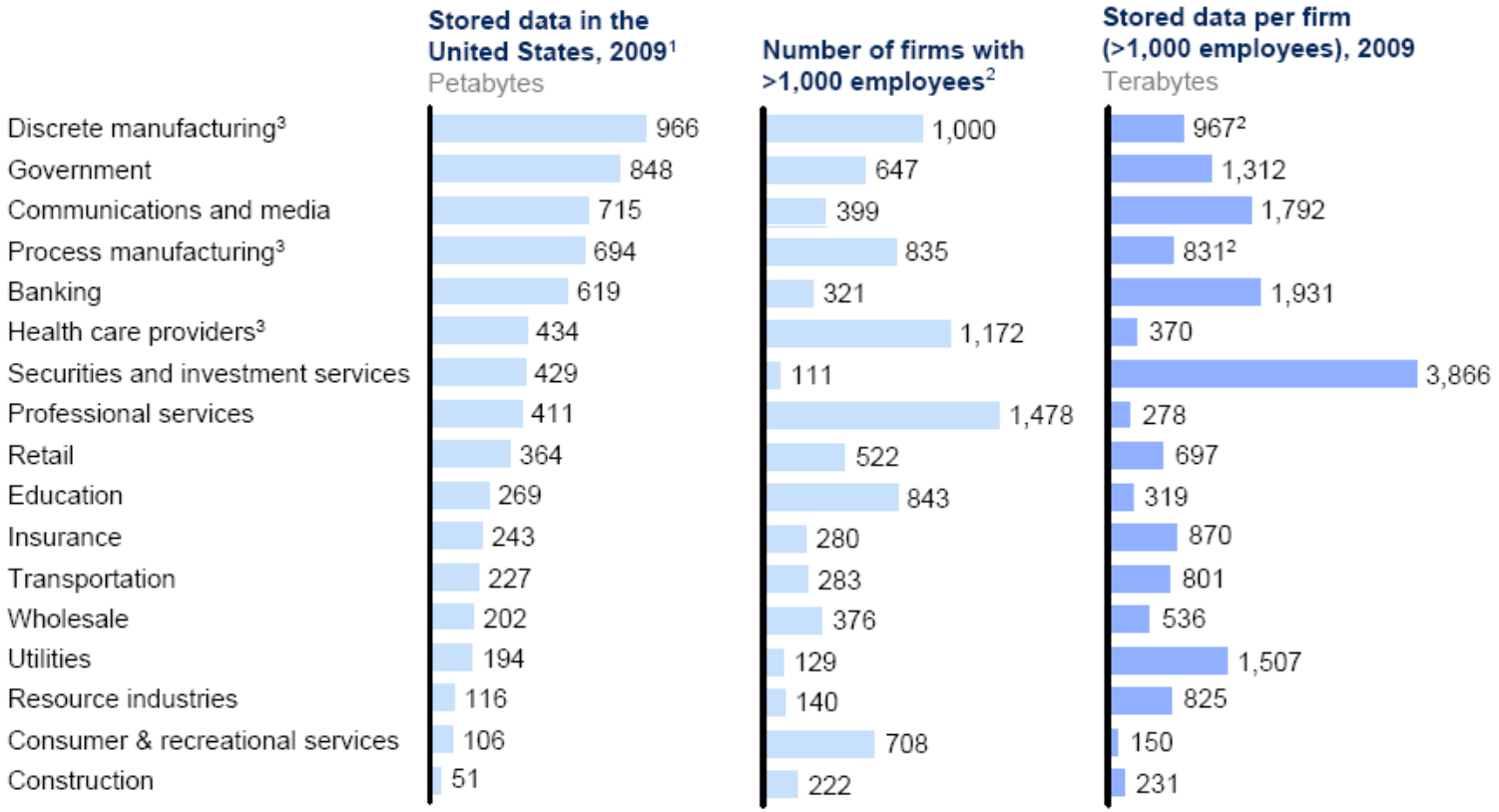


NOTE: Numbers may not sum due to rounding.

SOURCE: Hilbert and López, "The world's technological capacity to store, communicate, and compute information," *Science*, 2011

Trends in "Big Data:" Key insights

Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



1 Storage data by sector derived from IDC.
2 Firm data split into sectors, when needed, using employment
3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

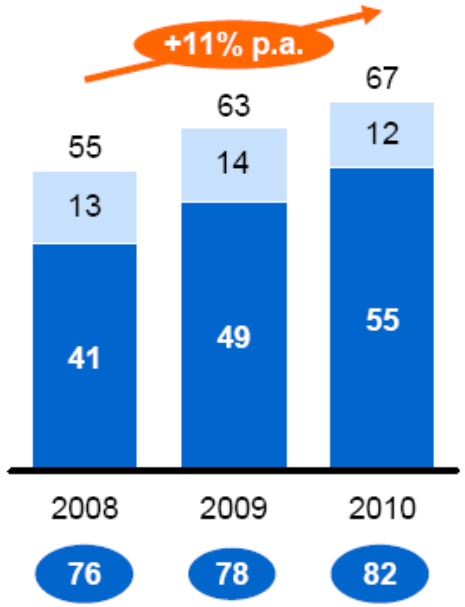
Trends in "Big Data:" Key insights

The penetration of social networks is increasing online and on smartphones; frequent users are increasing as a share of total users[1]

Frequent user[2]

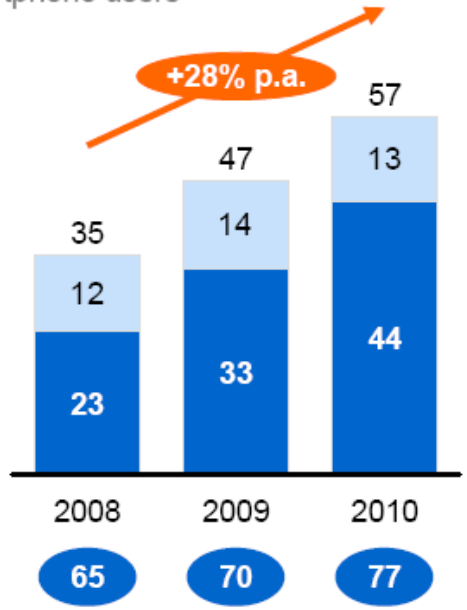
Social networking penetration on the PC is slowing, but frequent users are still increasing

% of respondents



Social networking penetration of smartphones has nearly dodoubled since 2008

% of smartphone users



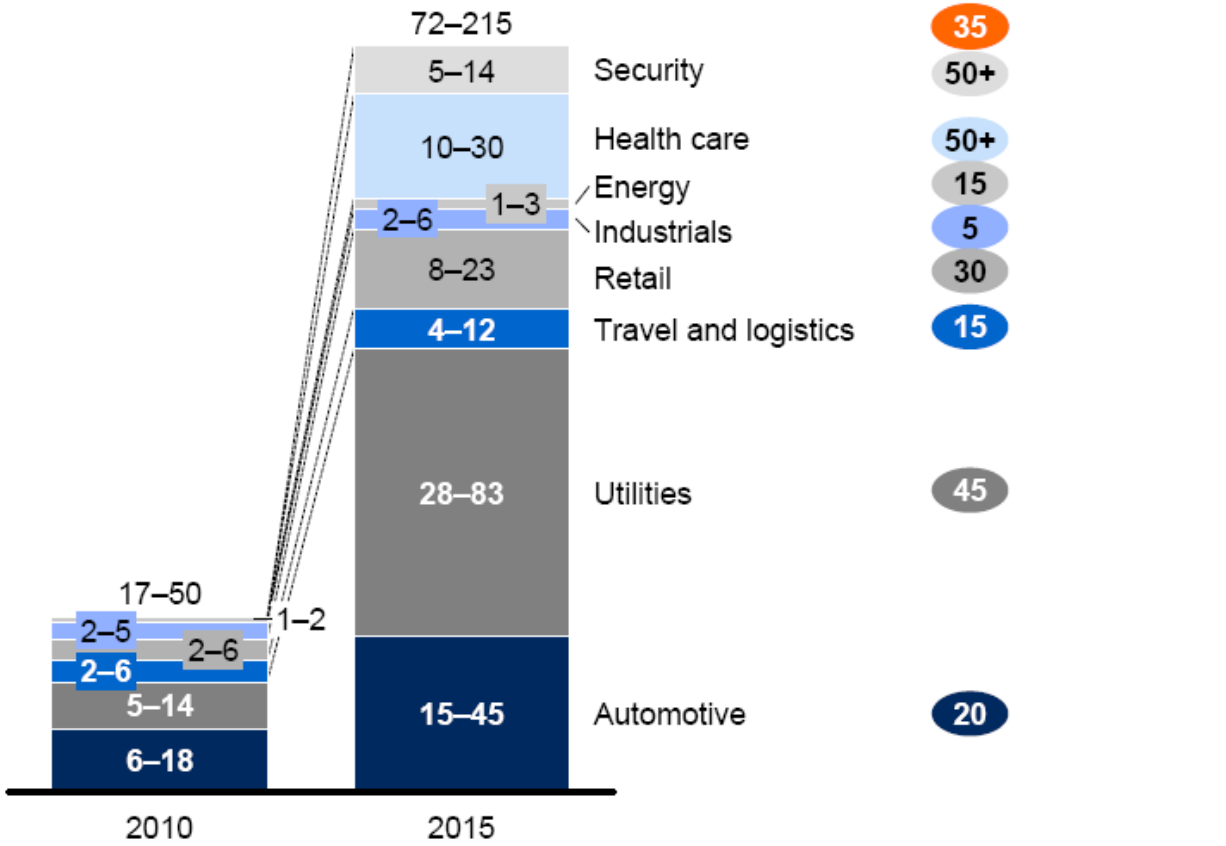
1 Based on penetration of users who browse social network sites. For consistency, we exclude Twitter-specific questions (added to survey in 2009) and location-based mobile social networks (e.g., Foursquare, added to survey in 2010).
2 Frequent users defined as those that use social networking at least once a week.

SOURCE: McKinsey iConsumer Survey

Trends in "Big Data:" Key insights

Data generated from the Internet of Things will grow exponentially as the number of connected nodes increases

Estimated number of connected nodes
Million



NOTE: Numbers may not sum due to rounding.
SOURCE: Analyst interviews; McKinsey Global Institute analysis

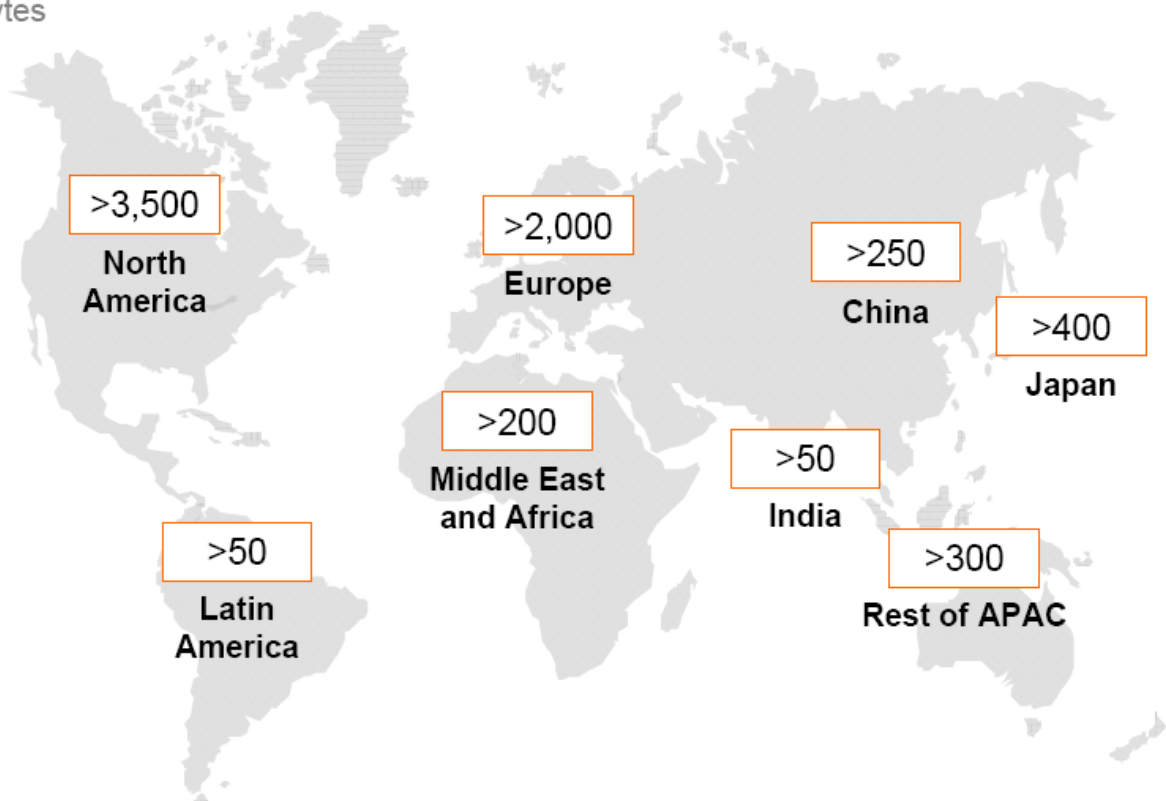
Trends in "Big Data:" Key insights

(1) Data have swept into every industry and business function and are now an important factor of production (continued)

There is considerable asymmetry among regions in the World and industries.

Amount of new data stored varies across geography

New data stored[1] by geography, 2010
Petabytes



1 New data stored defined as the amount of available storage used in a given year; see appendix for more on the definition and assumptions.

SOURCE: IDC storage reports; McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

(1) Data have swept into every industry and business function and are now an important factor of production (continued)

There is considerable asymmetry among regions in the World and industries.

The type of data generated and stored varies by sector[1]

	Video	Image	Audio	Text/ numbers
Banking				
Insurance				
Securities and investment services				
Discrete manufacturing				
Process manufacturing				
Retail				
Wholesale				
Professional services				
Consumer and recreational services				
Health care				
Transportation				
Communications and media[2]				
Utilities				
Construction				
Resource industries				
Government				
Education				

Penetration: High, Medium, Low

1 We compiled this heat map using units of data (in files or minutes of video) rather than bytes.
2 Video and audio are high in some subsectors.
SOURCE: McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

(2) Big data creates value in several ways

The report identifies five broadly applicable ways to leverage big data that offer transformational potential to create value and have implications for how organizations will have to be designed, organized, and managed:

Creating transparency (making big data more easily accessible to relevant stakeholders in a timely manner can create tremendous value)

Enabling experimentation to discover needs, expose variability, and improve performance (Organizations can collect more accurate and detailed performance data, in real or near real time, on everything from product inventories to personnel sick days. IT enables organizations to instrument processes and then set up controlled experiments. Using data to analyze variability in performance—that which either occurs naturally or is generated by controlled experiments—and to understand its root causes can enable leaders to manage performance to higher levels.)

Trends in "Big Data:" Key insights

(2) Big data creates value in several ways (continued)

Segmenting populations to customize actions (organizations create highly specific segmentations and to tailor products and services precisely to meet those needs.)

Replacing/supporting human decision making with automated algorithms (Sophisticated analytics can substantially improve decision making, minimize risks, and unearth valuable insights that would otherwise remain hidden.)

Innovating new business models, products, and services (Big data enables companies to create new products and services, enhance existing ones, and invent entirely new business models.)

Trends in "Big Data:" Key insights

(3) Use of big data will become a key basis of competition and growth for individual firms

The use of big data is becoming a key way for leading companies to outperform their peers. They estimate, for example, that a retailer embracing big data has the potential to increase its operating margin by more than 60 percent.

Big data will also help to create new growth opportunities and entirely new categories of companies, such as those that aggregate and analyze industry data. For example, medical clinical information providers, which aggregate data and perform the analyses necessary to improve health care efficiency, could compete in a market worth more than $10 billion by 2020.

Trends in "Big Data:" Key insights

(4) The use of big data will underpin new waves of productivity growth and consumer surplus

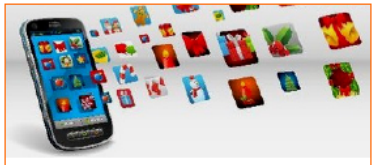
Across five domains studied, they identified many big data levers that will underpin substantial productivity growth.

These opportunities have the potential to improve efficiency and effectiveness, enabling organizations both to do more with less and to produce higher-quality outputs, i.e., increase the value-added content of products and services

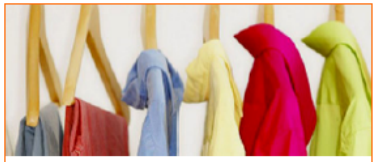
Big data can generate significant financial value across sectors

US health care
- $300 billion value per year
- ~0.7 percent annual productivity growth

Europe public sector administration
- €250 billion value per year
- ~0.5 percent annual productivity growth



Global personal location data
- $100 billion+ revenue for service providers
- Up to $700 billion value to end users



US retail
- 60+% increase in net margin possible
- 0.5–1.0 percent annual productivity growth

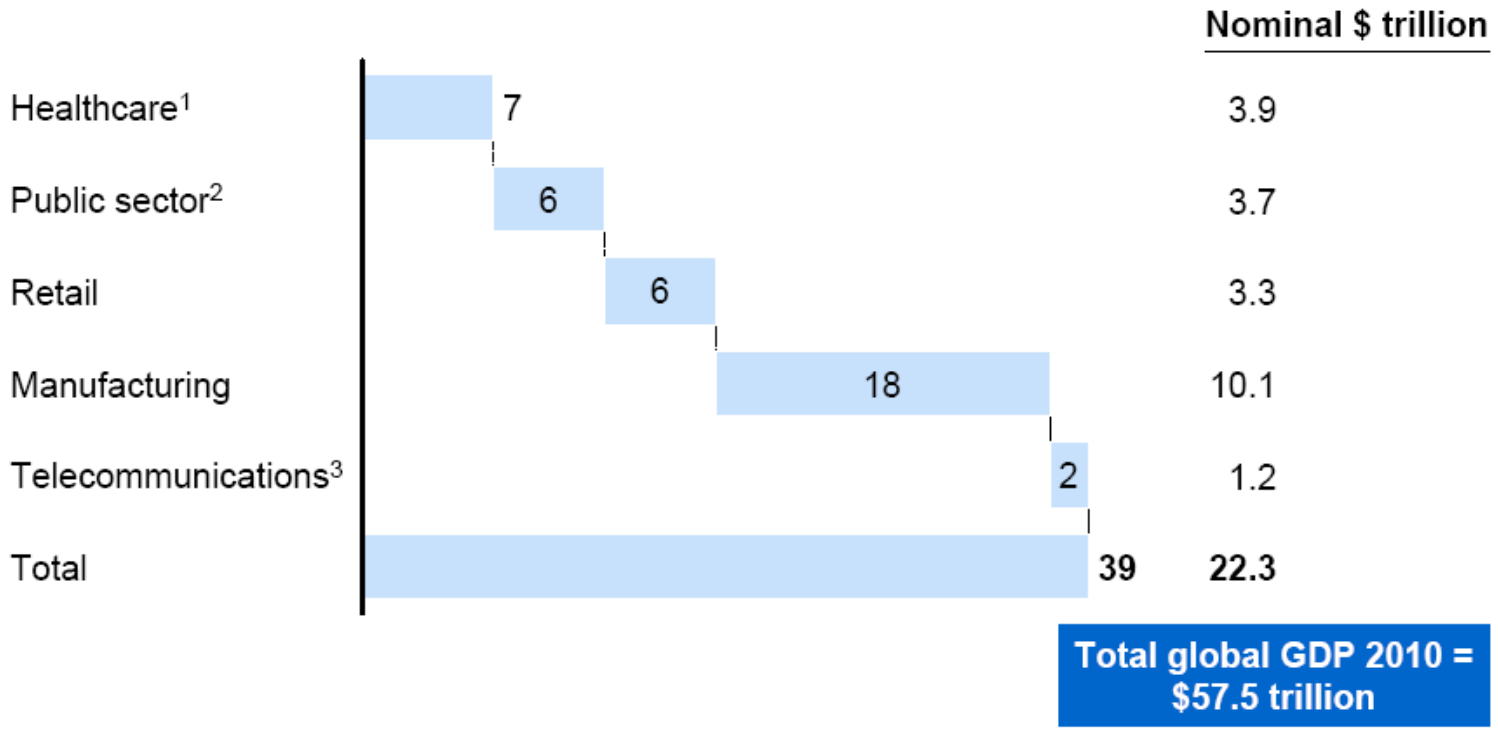
Manufacturing
- Up to 50 percent decrease in product development, assembly costs
- Up to 7 percent reduction in working capital

SOURCE: McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

The five sectors or domains we have chosen to study in depth make important contributions to the global economy

Estimated global GDP of sectors in 2010
% of total GDP



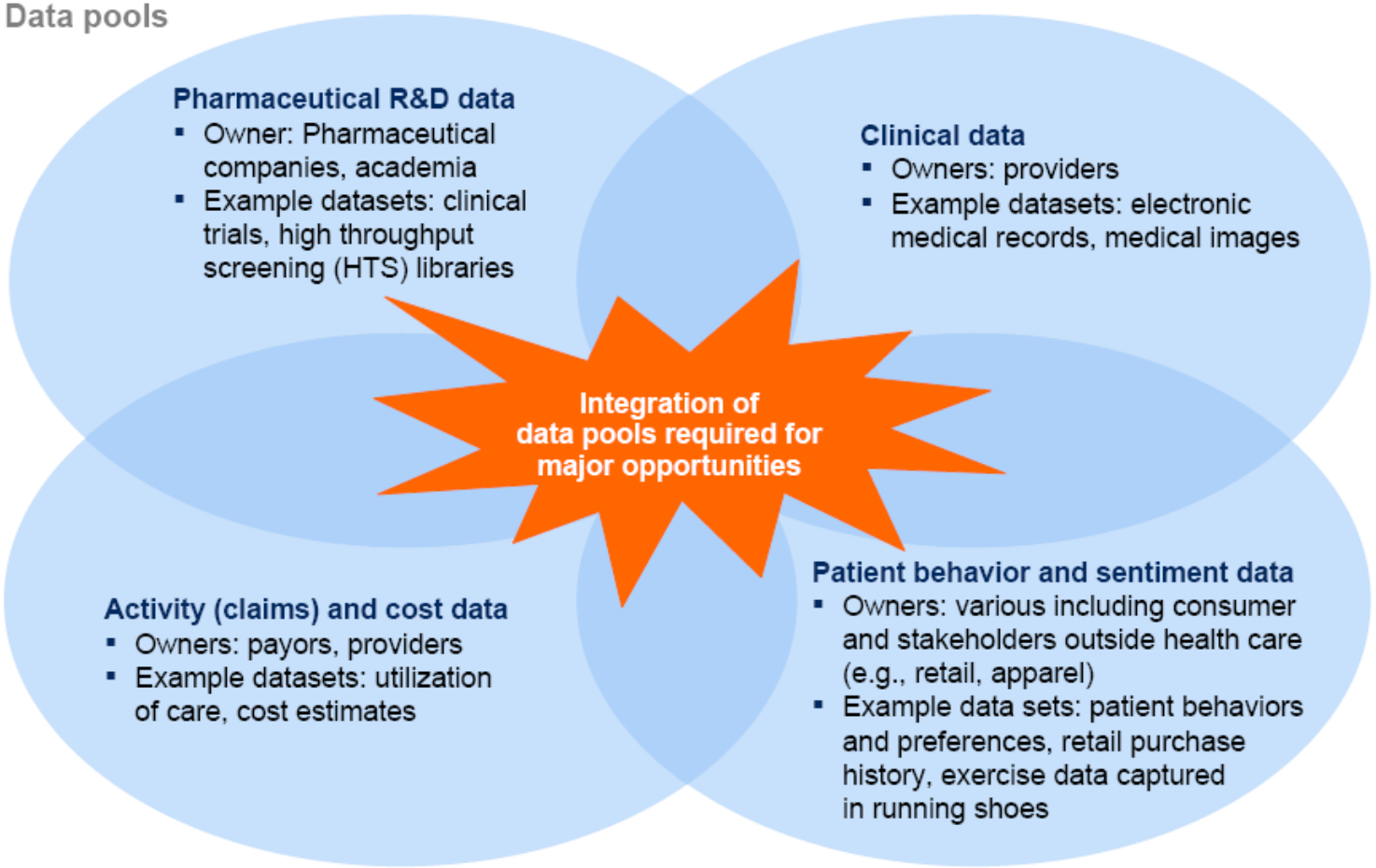
1 Includes health and social services, medical and measuring equipment, and pharmaceuticals.
2 Refers to public sector administration, defense, and compulsory social security (excludes education).
3 Since personal location data is a domain and not a sector, we've used telecom as a comparison for GDP.
NOTE: Numbers may not sum due to rounding.

SOURCE: Global Insight; McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

Four distinct big data pools exist in the US health care domain today with little overlap in ownership and low integration

Data pools



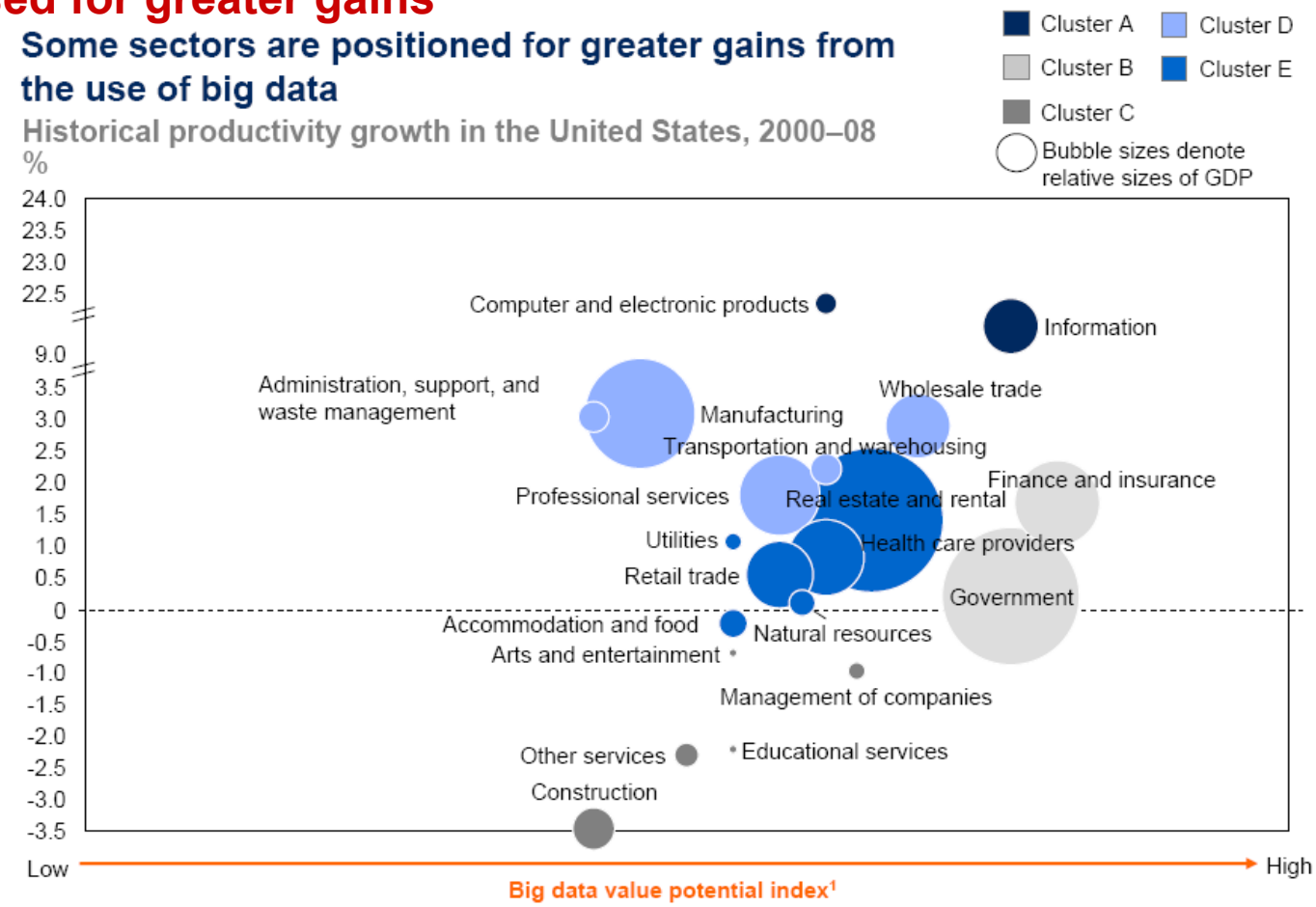
SOURCE: McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

(5) While the use of big data will matter across sectors, some sectors are poised for greater gains

Some sectors are positioned for greater gains from the use of big data

Historical productivity growth in the United States, 2000–08
%



1 See appendix for detailed definitions and metrics used for value potential index.
SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

(6) There will be a shortage of talent necessary for organizations to take advantage of big data

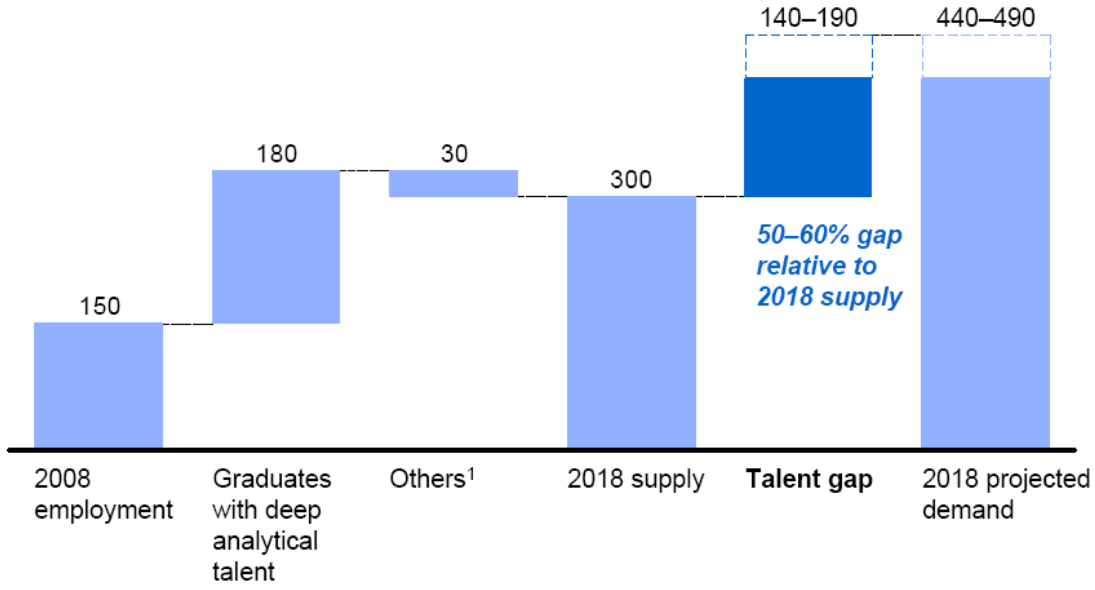
A significant constraint on realizing value from big data will be a shortage of talent, …

… particularly of people with deep expertise in statistics and machine learning, and the managers and analysts who know how to operate companies by using insights from big data.

Demand for deep analytical positions in a big data world could exceed the supply being produced on current trends by 140K to 190K positions.

Demand for deep analytical talent in the United States could be 50 to 60 percent greater than its projected supply by 2018

Supply and demand of deep analytical talent by 2018
Thousand people

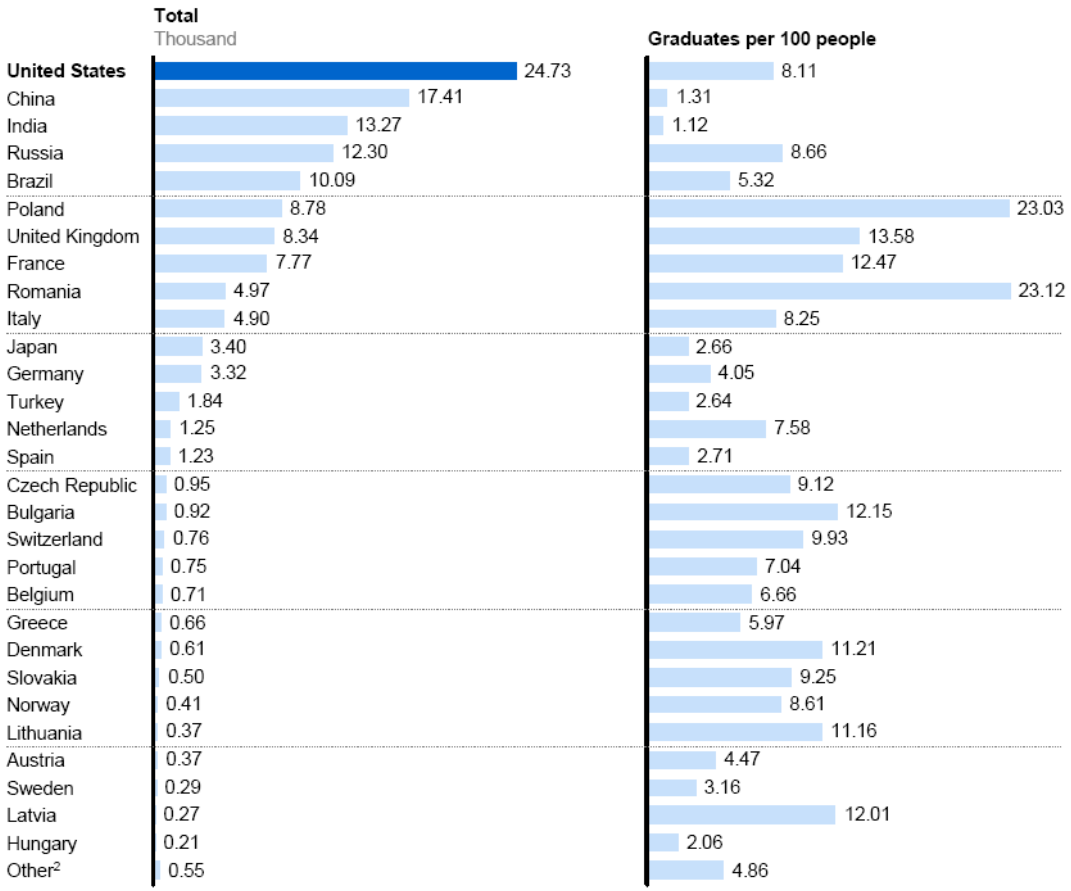


1 Other supply drivers include attrition (-), immigration (+), and reemploying previously unemployed deep analytical talent (+).
SOURCE: US Bureau of Labor Statistics; US Census; Dun & Bradstreet; company interviews; McKinsey Global Institute analysis

Trends in "Big Data:" Key insights

The United States graduates the largest number of people with deep analytical training

Number of graduates with deep analytical training in 2008[1]



1 These data count new graduates, i.e., a flow of deep analytical talent, which we define as people with advanced training in statistics and/or machine learning and who conduct data analysis.
2 Other includes Finland, Estonia, Croatia, Slovenia, Iceland, Cyprus, Macedonia, and Malta.
SOURCE: Eurostat; Russia Statistics; Japan Ministry of Education; India Sat; NASSCOM Strategic Review 2005; China Statistical Yearbook; China Education News; IMF World Economic Outlook Database

Trends in "Big Data:" Key insights

(7) Several issues will have to be addressed to capture the full potential of big data

Data policies (As an ever larger amount of data is digitized and travels across organizational boundaries, there is a set of policy issues that will become increasingly important, including, but not limited to, *privacy, security, intellectual property, and liability*.)

Technology and techniques (To capture value from big data, organizations will have to deploy new technologies (e.g., storage, computing, and analytical software) and techniques (i.e., new types of analyses).)

Organizational change and talent. (Organizational leaders often lack the understanding of the value in big data as well as how to unlock this value.)

Trends in "Big Data:" Key insights

(7) Several issues will have to be addressed to capture the full potential of big data (continued)

Access to data (To enable transformative opportunities, companies will increasingly need to integrate information from multiple data sources. In some cases, organizations will be able to purchase access to the data.)

Industry structure (Sectors with a relative lack of competitive intensity and performance transparency, along with industries where profit pools are highly concentrated, are likely to be slow to fully leverage the benefits of big data.)

Trends in "Big Data:" The human side

"A wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."

Herbert A. Simon, "Designing organizations for an information-rich world," in Martin Greenberger, Computers, Communication, and the Public Interest, Baltimore, MD: The Johns Hopkins Press, 1971

"Big Data"

http://www.ted.com/talks/kenneth_cukier_big_data_is_better_data (15'51")

http://www.ted.com/talks/susan_etlinger_what_do_we_do_with_all_this_big_data (12'23")

http://www.ted.com/talks/joel_selanikio_the_surprising_seeds_of_a_big_data_revolution_in_healthcare (16'14")

http://www.ted.com/talks/ben_wellington_how_we_found_the_worst_place_to_park_in_new_york_city_using_big_data (11'48")

http://www.ted.com/watch/ted-institute/ted-state-street/jessica-donohue-the-up-side-of-data (12'11")

"Big Data"

http://www.youtube.com/watch?v=B27SpLOOhWw (17'46")
http://www.youtube.com/watch?v=W2Vnke8ryco (16'38")

“Business Intelligence”

http://www.youtube.com/watch?v=LhZX0MAYKp8 **(19’14”)**

Trends in “Business Analytics”

http://www.youtube.com/watch?v=nfMnILQVZXo (19’35”)

Real Time Analytics for Big Data: A Facebook Case Study

Recommended for watching at home (if not now, the later in the semester):

http://www.youtube.com/watch?v=viPRny0nq3o (1h 17'46")

Nice classification of analytics from the point of view of time.

Required reading

- http://www.datanami.com/2016/01/22/data-scientists-the-myth-and-the-reality/

- http://www.datanami.com/2016/01/06/what-does-2016-mean-for-data-science/