Big Data, Big Impact

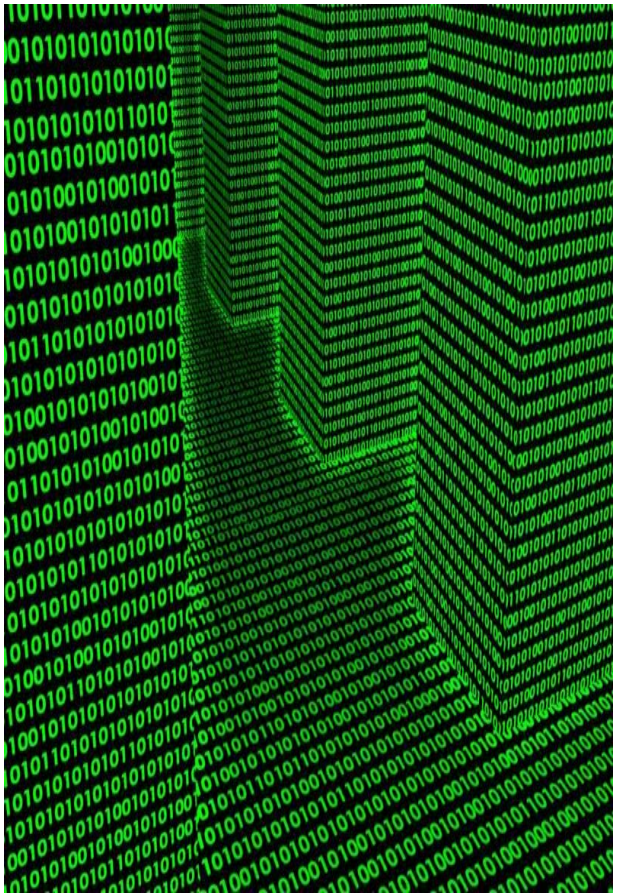
Danny Wang, Ph.D.
Vice President
of Business Strategy and Risk Management
Republic Bank

Agenda

- Overview
- What is ‘Big Data’?
- Accelerates advances in computer & technologies
- Revolutionizes data measurement
- Influences business and economics
- Transforms utility of social networks
- Promotes analytics and decision making
- Introduces new challenges, risks, and pitfalls

Overview

- According to IDC, a leading tech research firm, data is growing exponentially at rate of 50% a year.
- More than doubling every two years.
- We create 2.5 quintillion (10^{18}) bytes of data every day.
- 90% of the data in the world today has been created in the last two years alone.



Overview

- This data comes from everywhere:
  - sensors used to gather climate information
  - posts to social media sites
  - digital pictures and videos
  - purchase transaction records
  - cell phone GPS signals

- McKinsey Global Institute, a consulting firm, projected the U.S. needs
  - 140,000 to 190,000 more workers with “deep analytical” expertise
  - 1.5 million more data literate managers, retrained or hired.

Overview of impact

- The impact of big data is being felt beyond marketing departments.
- The position of chief data officer, data scientist is commonplace in the U.S.
- If used effectively, big data can be a powerful tool.
- Talent, not just technology. People matter!
- The evidence for the power of big data can obscure the difficulties that many companies face when working with it.

Agenda

- Overview
- **What is ‘Big Data’?**
- Accelerates advances in computer & technologies
- Revolutionizes data measurement
- Influences business and economics
- Transforms utility of social networks
- Promotes analytics and decision making
- Introduces new challenges, risks, and pitfalls

What is ‘Big Data’?

- A meme and a marketing term
  - Meme -- “an idea, behavior, style, or usage that spreads from person to person within a culture” Source: Merriam Webster
  - A lot of people talk about it, especially in IT, analytics world.
- Southland for advancing trends in technology
- Open the door to a new approach to understanding the world and making decisions
- Big data is more than simply a matter of size
- It is an opportunity to find insights in new and emerging types of data and content
- Makes your business more agile, and answers questions that were previously considered beyond your reach

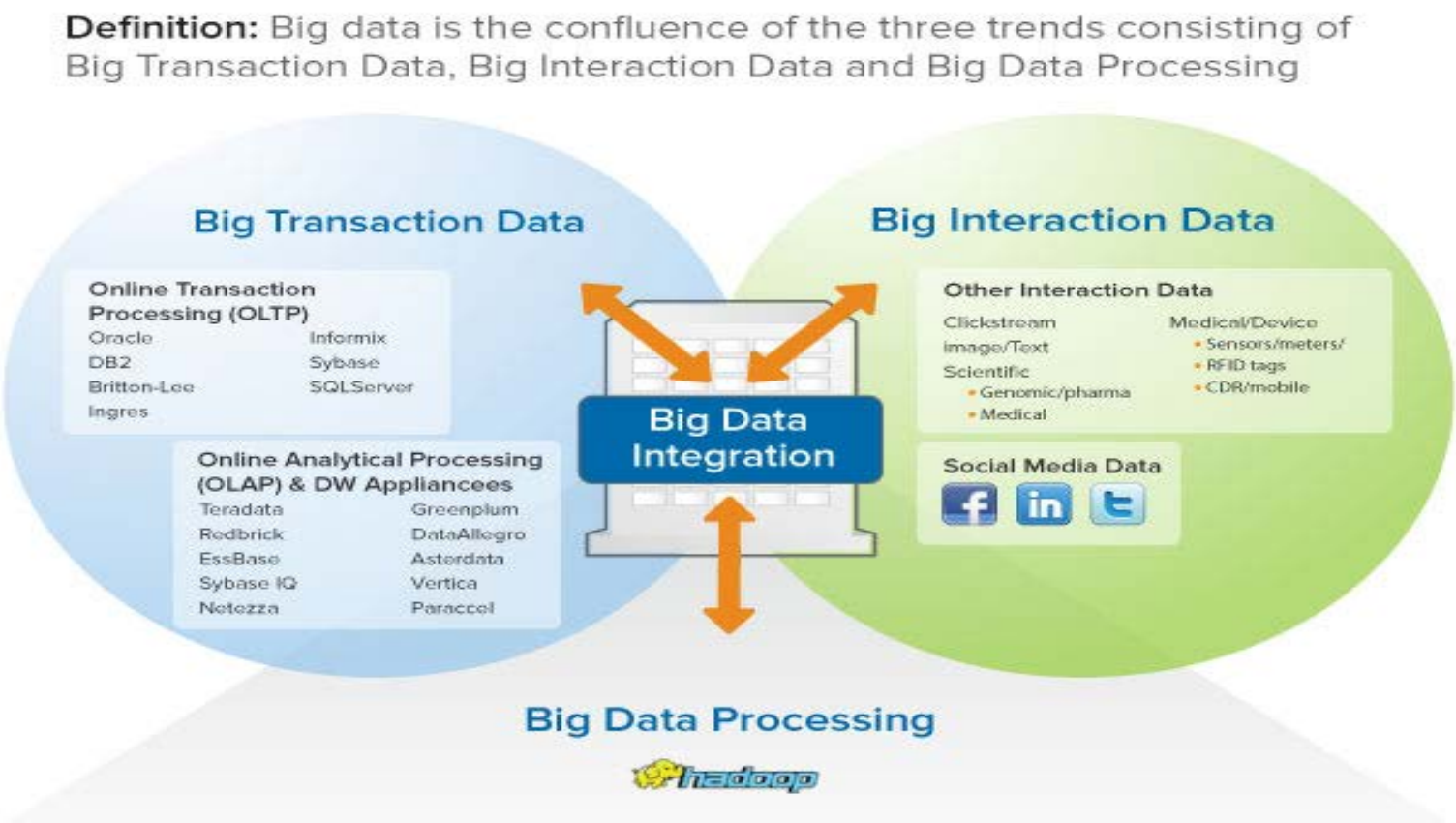
Examples

- Wal-Mart handles more than a million customer transactions each hour and imports those into databases estimated to contain more than 2.5 petabytes of data.
- Radio frequency identification (RFID) systems used by retailers and other scan generate 100 to 1,000 times the data of conventional bar code systems.
- Facebook handles more than 250 million photo uploads and the interactions of 800 million active users with more than 900 million objects (pages, groups, etc.) – each day.
- More than 5 billion people are calling, texting, tweeting and browsing on mobile phones worldwide.

What is ‘Big Data’?

- “Big Data refers to relatively large amounts of structured and unstructured data that require machine-based systems and technologies in order to be fully analyzed.” – Ad Age
- “Big data usually includes data sets with sizes beyond the ability of commonly-used software tools to capture, curate, manage, and process the data within a tolerable elapsed time.” – Wikipedia
- “Big Data is the massive amounts of data that collect over time that are difficult to analyze and handle using common database management tools.” – PCMac

Definition



Source: www.informatica.com

4 Vs

- **Volume:** Enterprises are awash with ever-growing data of all types, easily amassing terabytes—even petabytes—of information.
  - Turn 12 terabytes of Tweets created each day into improved product sentiment analysis
  - Convert 350 billion annual meter readings to better predict power consumption

- **Velocity:** Sometimes 2 minutes is too late. For time-sensitive processes such as catching fraud, big data must be used as it streams into your enterprise in order to maximize its value.
  - Scrutinize 5 million trade events created each day to identify potential fraud
  - Analyze 500 million daily call detail records in real-time to predict customer churn faster

- **Variety**: Big data is any type of data - structured and unstructured data such as text, sensor data, audio, video, click streams, log files and more. New insights are found when analyzing these data types together.
  - Monitor 100's of live video feeds from surveillance cameras to target points of interest
  - Exploit the 80% data growth in images, video and documents to improve customer satisfaction

- **Veracity**: 1 in 3 business leaders don't trust the information they use to make decisions. How can you act upon information if you don't trust it? Establishing trust in big data presents a huge challenge as the variety and number of sources grows.

Source: www.ibm.com

Agenda

- Overview
- What is ‘Big Data’?
- **Accelerates advances in computer & technologies**
- Revolutionizes data measurement
- Influences business and economics
- Transforms utility of social networks
- Promotes analytics and decision making
- Introduces new challenges, risks, and pitfalls

Volume increasing, scope expanding

- In addition to an increasing volume of data, the scope of data sources is also expanding.
- Particularly, unstructured data such as words, images, video on the web and those streams of sensor data.
- Unstructured data is not typically grist for traditional database.

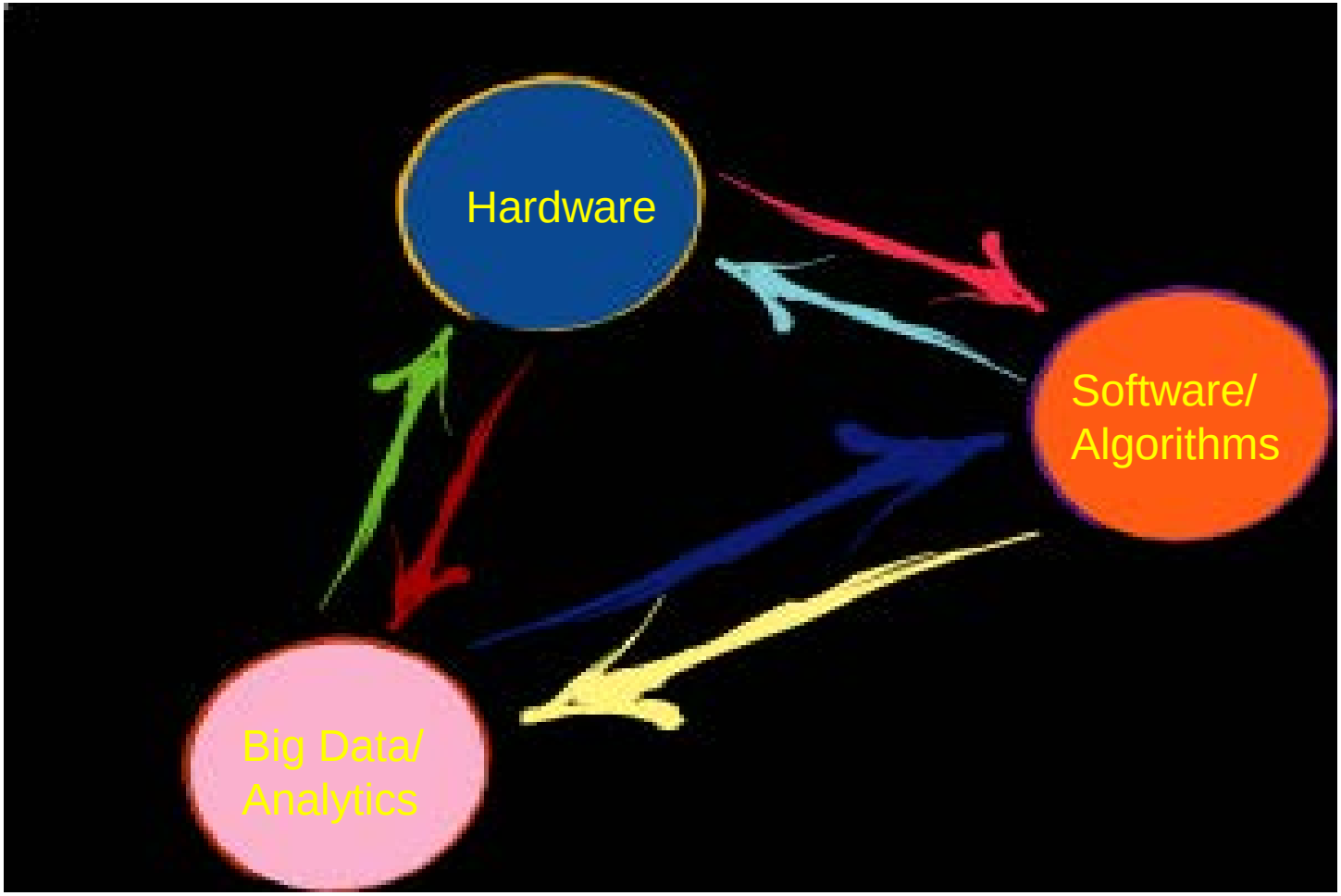
Hardware advancements

- Cheap, abundant storage and server processing capacity
- Faster processors
- Affordable large-memory capabilities, such as Hadoop
- New storage and processing technologies designed specifically for large data volumes, including unstructured data
- Parallel processing, clustering, MPP, virtualization, large grid environments, high connectivity and high throughputs
- Cloud computing and other flexible resource allocation arrangements

Software advancements

- Artificial Intelligence technologies
  - Natural language processing
  - Pattern recognition
  - Machine perception, computer vision, and speech recognition
  - Machine learning
- Advanced algorithms

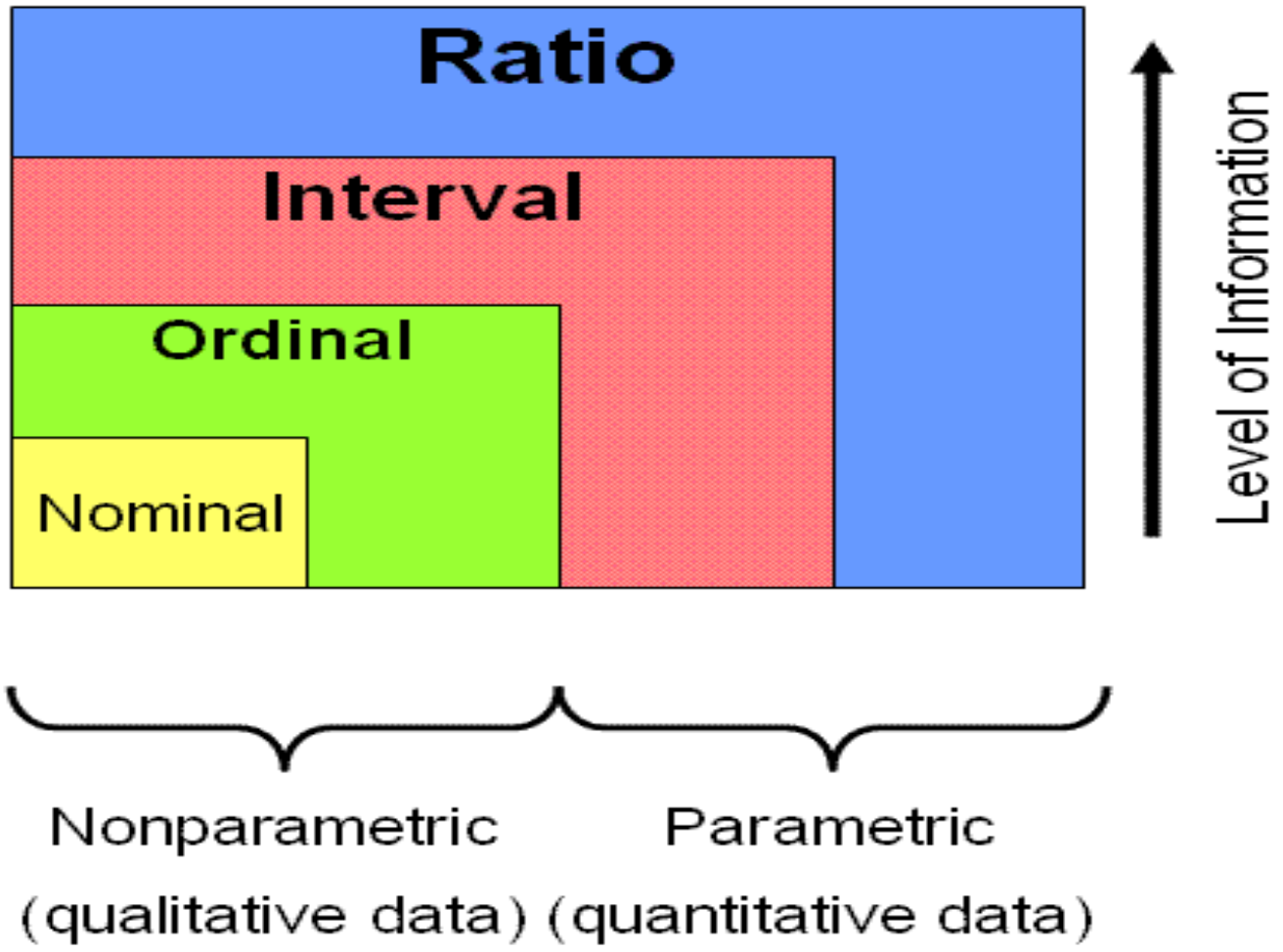
A virtuous circle



Agenda

- Overview
- What is ‘Big Data’?
- Accelerates advances in computer & technologies
- **Revolutionizes data measurement**
- Influences business and economics
- Transforms utility of social networks
- Promotes analytics and decision making
- Introduces new challenges, risks, and pitfalls

Traditional data measurement



***Nonparametric statistics may be used to analyze interval and ratio data measurements.**

Fine-grained measurement

- Prof. Erick Brynjolfsson at MIT's Sloan School of Management, thinks data measurement is the modern equivalent of the microscope.

- The microscope, invented four centuries ago, allowed people to see and measure things as never before – at the cellular level.

- Big data, such as Google search, Facebook posts and Twitter messages, make it possible to measure behavior and sentiment in fine details.

- Big data can change how we view the world.

Agenda

- Overview
- What is ‘Big Data’?
- Accelerates advances in computer & technologies
- Revolutionizes data measurement
- **Influences business and economics**
- Increases utility of social networks
- Promotes analytics and decision making
- Introduces new challenges, risks, and pitfalls

How does big data impact business?

- Big data has a robust influence on business, economics and other fields.
- Decisions will increasingly be based on data and analytics rather than on experience and intuition.
- Plenty of evidence now exists of the payoff from data-first thinking.
- Using analytical input to increase profit and make the most astute decisions is key to success in current business environments.

Research highlights

» There is a strong link between financial performance and the effective use of big data.

» Companies become successful at exploiting data by focusing on business priorities.

» Talent matters as much as technology

» Social media analytics and web-tracking technologies can transform the way business collect data about customers

Source: Big data Lesson from the leaders, The Economist Intelligence Unit Limited 2012

Agenda

- Overview
- What is ‘Big Data’?
- Accelerates advances in computer & technologies
- Revolutionizes data measurement
- Influences business and economics
- **Transforms utility of social networks**
- Promotes analytics and decision making
- Introduces new challenges, risks, and pitfalls

Impact on social networks

- Big data is already transforming the study of how social networks function.
- Huge digital data sets of collective online behavior can be collected and unitized.
- Patterns of influence and peaks in communication on a subject can be found.
- Actionable results can be yielded.
- These patterns and results also can be used to make future predictions.

Agenda

- Overview
- What is ‘Big Data’?
- Accelerates advances in computer & technologies
- Revolutionizes data measurement
- Influences business and economics
- Increases utility of Social networks
- **Promotes analytics and decision making**
- Introduces new challenges, risks, and pitfalls

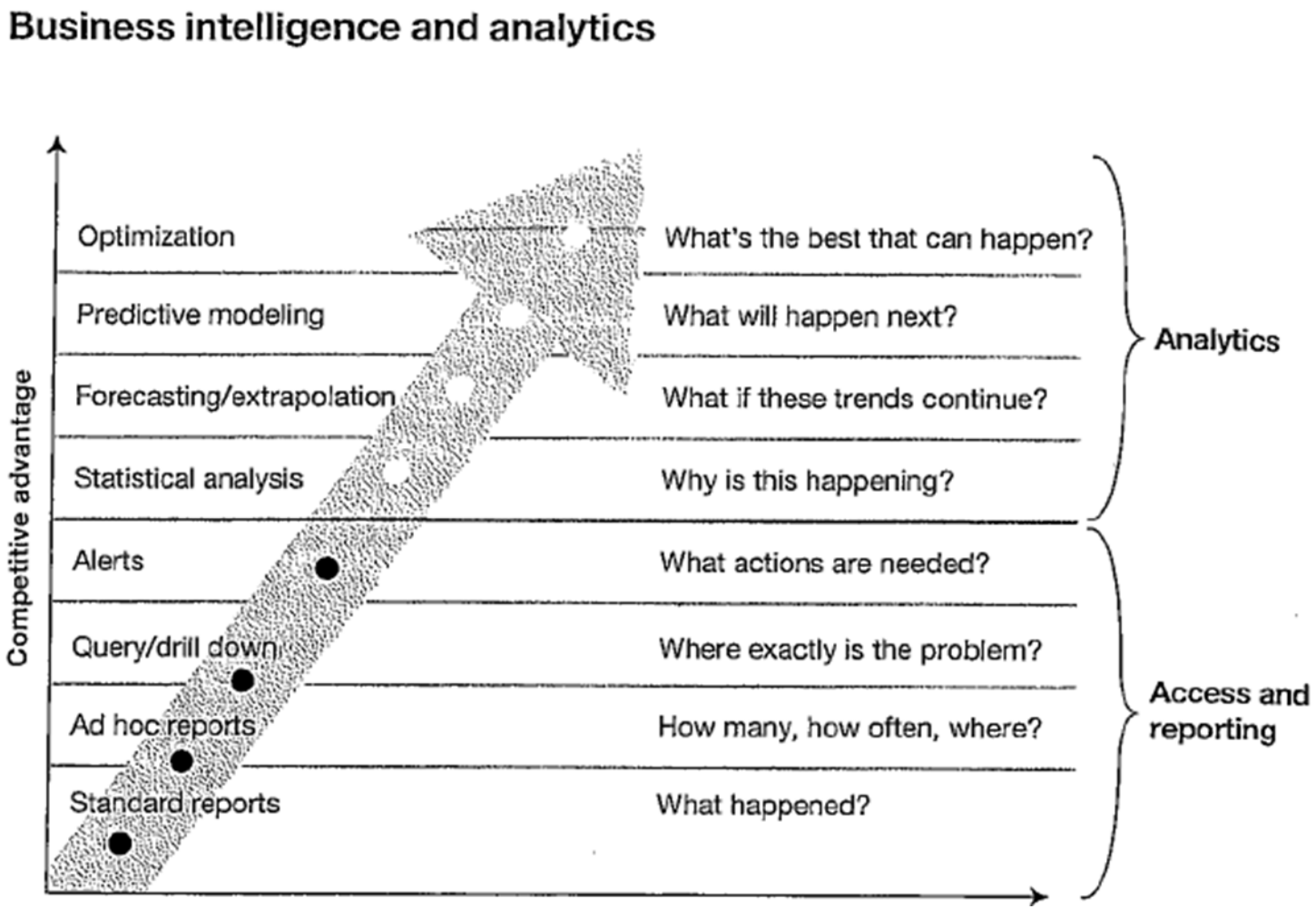
Impact on analytics

- Big data has an impact on analytics through the development and utilization of new and enhanced procedures and tools.
- Data is tamed and understood using computer algorithms and mathematical models.
- The predictive power of big data is being explored – and shows promise – in fields like public health, economic development and forecasting, etc.
- Retailers, such as Walmart and Kohl's, use big data driven approach for their businesses.
- Logistics such as UPS, FedEx, mine big data to get a optimal solution.

Compete on analytics

- Companies in many industries offer similar products and use comparable technologies.
- High-performance business processes are among the last remaining points of differentiation.
- Make the smartest business decision possible.
- Execute the business with maximum efficiency and effectiveness.

Big Data analytics



Agenda

- Overview
- What is ‘Big Data’?
- Accelerates advances in computer & technologies
- Revolutionizes data measurement
- Influences business and economics
- Increases utility of social networks
- Promotes analytics and decision making
- **Introduces new challenges, risks, and pitfalls**

Challenges in the financial market

»How can we improve our risk forecasting at the point of origination?

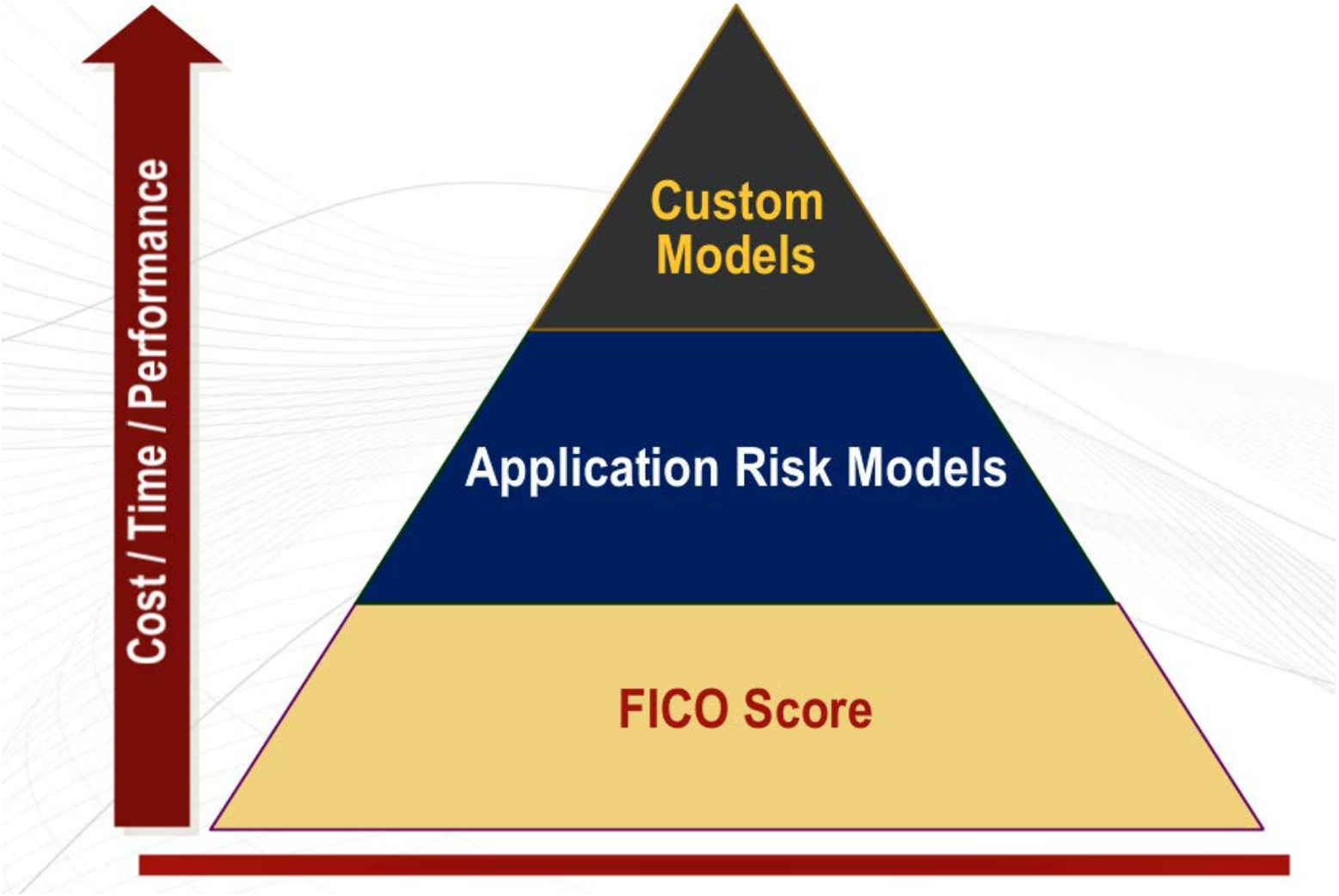
»How can we increase the number of approved loan/credit card/mortgage applications?

»How can we reduce the level of risk?

» How can we streamline operations for increased cost efficiencies?

» How can we maintain regulatory compliance?

Scoring tools



The bigger, the better?

- We see mountains of data today.
- The existing mountains are continually growing, and new ones are being built daily.
- Are we collecting the right data? In a good manner?
- Sometimes, we are swamped with more data that we can use.
- No one addresses the cost of data mining
  - System/software/computers
  - People/staff infrastructure
  - Maintaining the hardware and software
  - Data storage

Cost benefit analysis

- Data collection, storage, analysis and use in government and the private sector is necessary.
- It is also important to ask ourselves if a project is truly needed.
- Does the data collection and analysis activities and expenditures really provide tangible benefits?
- Numbers tell us something, but don’t tell us everything we need to know.
- We need to engage people and human intellect and insight.

Don’t head in a wrong direction

» Large benefits can be realized from a data driven approach.

» However, with huge data sets and fine-grained measurement, there is also an increased risk of false discoveries.

» The challenge of data mining is to find a meaningful needle in massive haystacks.

» The trouble is that there are a lot of straws that look like needles.

THANK YOU

dwang@republicbank.com