Big Data

Big Data's 4 Vs

Big Data

❖ **Big Data's 4 V Big Challenges**

- Volume – Data Size

- Variety – Data Formats

- Velocity – Data Streaming Speeds

- Veracity – Data Trustworthiness

Big Data

❖ **Volume – Data Size**

- 40 Zettabytes (10^{21}) of data is predicted to be created by 2020
- 2.5 Quintillionbytes (10^{18}) of data are created every day
- 6 Billion (10^{9}) people have mobile phones
- 100 Terabytes (10^{12}) of data (at least) is stored by most U.S. companies
- 966 Petabytes (10^{15}) was the approximate storage size of the American manufacturing industry in 2009

Big Data

❖ **Variety – Data Formats**

- 150 Exabytes (10^{18}) was the estimated size of data for health care throughout the world in 2011
- More than 4 Billion (10^{9}) hours each month are used in watching YouTube
- 30 Billon contents are exchanged every month on Facebook
- 200 Million monthly active users exchange 400 Million tweets every day

Big Data

❖ **Velocity – Data Streaming Speeds**

- 1 Terabytes (10^{12}) of trade information is exchanged during every trading session at the New York Stock Exchange

- 100 sensors (approximately) are installed in modern cars to monitor fuel level, tire pressure, etc.

- 18.9 Billion network connections were estimated in 2016

Big Data

❖ **Veracity – Data Trustworthiness**

- 1 out of 3 business leaders have experienced trust issues with their data when trying to make a business decision

- \$3.1 Trillion (10^{12}) a year is estimated to be wasted in the U.S. economy due to poor data quality

Big Data

❖ **New technology is needed to overcome these 4 V Big Data Challenges**

- Volume – Data Size

- Variety – Data Formats

- Velocity – Data Streaming Speeds

- Veracity – Data Trustworthiness

Big Data

How is Big Data being Used?

Big Data

❖ **Wal-Mart**

- Wal-Mart's Data Warehouse
  - Stores approx. 4 petabytes (4×10^{15}) of data

- Records every single purchase
  - Approximately 267 million transactions a day from 6000 stores worldwide is recorded

WAL★MART®

Big Data

❖ **Wal-Mart**

- Wal-Mart's Data Analysis
  - Focused on evaluating the effectiveness of pricing strategies and advertising campaigns

- Seeking for improvement methods in inventory management and supply chains

WAL★MART®

Big Data

❖ **Recommendation System using Big Data**

- Based on data analysis of simple elements
  - What users made purchases in the past
  - Which items do they have in their virtual shopping cart
  - Which items did customers rate and like
  - What influence did the rating have on other customers to make a purchase

Big Data

❖ **Amazon.com**

- Amazon.com's Recommendation System
- Item-to-Item Collaborative Filtering Algorithm
  - Personalization of the Online Store
    ➔ Customized to each customer
  - Each customer's store is based on the customer's personal interest
    - Example: For a new mother, the store will display baby supplies and toys

amazon.com
and you're done.™

Big Data

❖ **Citibank**

- Bank operations in 100 countries

- Big Data analysis on the database of basic financial transactions can enable Global insight on investments, market changes, trade patterns, and economic conditions

- Many companies (e.g., Zara, H&M, etc.) work with Citibank to locate new stores and factories

  citibank

Big Data

❖ **Product Development & Sales**

- For example, a Smartphone takes significant time and money to manufacture

- In addition, the duration of popularity for a new Smartphone is limited

- To maximize sales, a company needs to manufacture just the right amount of products and sell them in the right locations

Big Data

❖ **Product Development & Sales**

- Too much will result in leftovers and a
  big waste for the company!
- Too less will result in a lost opportunity for
  company profit and growth!
- Big Data can help find how many smartphones
  and where the products could be popular based
  on common search terms that people use
  - Use this to also estimate how many products could
    be sold in a certain location
  - But why is this difficult?

Big Data

References

References

- V. Mayer-Schönberger, and K. Cukier, Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013.

- T. White, Hadoop: The Definitive Guide. O'Reilly Media, 2012.

- J. Venner, Pro Hadoop. Apress, 2009.

- S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big Data, Analytics and the Path From Insights to Value," MIT Sloan Management Review, vol. 52, no. 2, Winter 2011.

- B. Randal, R. H. Katz, and E. D. Lazowska, "Big-data Computing: Creating revolutionary breakthroughs in commerce, science and society," Computing Community Consortium, pp. 1-15, Dec. 2008.

- G. Linden, B. Smith, and J. York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan/Feb. 2003.

References

- J. R. GalbRaith, "Organizational Design Challenges Resulting From Big Data," Journal of Organization Design, vol. 3, no. 1, pp. 2-13, Apr. 2014.

- S. Sagiroglu and D. Sinanc, "Big data: A review," Proc. IEEE International Conference on Collaboration Technologies and Systems, pp. 42-47, May 2013.

- M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," Mobile Networks and Applications, vol. 19, no. 2, pp. 171-209, Jan. 2014.

- X. Wu, X. Zhu, G. Q. Wu, and W. Ding, ''Data Mining with Big Data,'' IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, Jan. 2014.

- Z. Zheng, J. Zhu, and M. R. Lyu, ''Service-Generated Big Data and Big Data-as-a-Service: An Overview,'' Proc. IEEE International Congress on Big Data, pp. 403–410, Jun/Jul. 2013.

References

- I. Palit and C.K. Reddy, "Scalable and Parallel Boosting with MapReduce," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 10, pp. 1904-1916, 2012.

- M.-Y Choi, E.-A. Cho, D.-H. Park, C.-J Moon, and D.-K. Baik, "A Database Synchronization Algorithm for Mobile Devices," IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 392-398, May 2010.

- IBM, What is big data?, http://www.ibm.com/software/data/bigdata/what-is-big-data.html [Accessed June 1, 2015]

- Hadoop Apache, http://hadoop.apache.org

- Wikipedia, http://www.wikipedia.org

Image sources

- Walmart Logo, By Walmart [Public domain], via Wikimedia Commons

- Amazon Logo, By Balajimuthazhagan (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons