TCSL-70130 Lecture 09: Big Data in Education
大數據之教育應用

Shiow-yang Wu 吳秀陽

CSIE 資訊工程學系
NDHU 國立東華大學

Lecture Topics

2

- What is big data? How big?
- Examples of big data
- Characteristics of big data
- Big data systems and analytics
- Big data in education
- Big data in education use cases

3

What is Big Data?

- The growth of data in volume(數量), velocity(速度), variety(種類) and veracity(準確性) are in such an unprecedented scale that traditional data management systems can no longer handle it properly.
- We need new technologies, new systems and new tools to deal with extremely large data sets.
- Big data ⇒ New applications ⇒ New business!

4

Examples of Big Data

- Walmart(沃爾瑪), the world's biggest retailer with over 11,000 stores in 27 countries.
- More than 260 million customers visit the distribution network weekly.
- Walmart Data Café is one of the world's largest private cloud.
- Process 2.5 PB of data every hour. (1PB=1024TB)
- More than 90% of requests are analyzed within the first 2 seconds.

5

Examples of Big Data

- Facebook(FB) is another good example.
- Over 2.41 billion monthly active users(MAU), among them, 1.59 billion are daily active users(DAU) as of June 2019.
- Photo uploads total 300 million per day.(Source: Gizmodo)
- Every 60 seconds, 136,000 photos are uploaded, 510,000 comments are posted, and 293,000 status updates are posted. That amounts to 1000+ TB of data generated per day.(Source: The Social Skinny)

(https://zephoria.com/top-15-valuable-facebook-statistics/)

6

Examples of Big Data

- Internet of Things(IoT) is expected to push the need for big data even further.
- 26.66 billion IoT devices in 2019 are active. By 2025 there will be 75 billion IoT devices in the World.
- 127 new devices are connected every second to the Internet.
- By 2020, every person will generate 1.7 MB in just 1 second.
- Global market will grow up to 457 billion by 2020.

7

Why Big Data?

From business point of view:

- Big data can unlock significant value by making information transparent.
- Big data can help organizations collect more accurate and detailed operational information to expose variability and boost performance.
- Big data allows ever-narrower segmentation of customers and therefore much more precisely tailored products or services.

8

Why Big Data?

- Sophisticated analytics can substantially improve decision-making, minimize risks, and unearth valuable insights that would otherwise remain hidden.
- Big data can be used to develop the next generation of products and services.
- Big Data is a "Big Deal"!

9

Let's Start

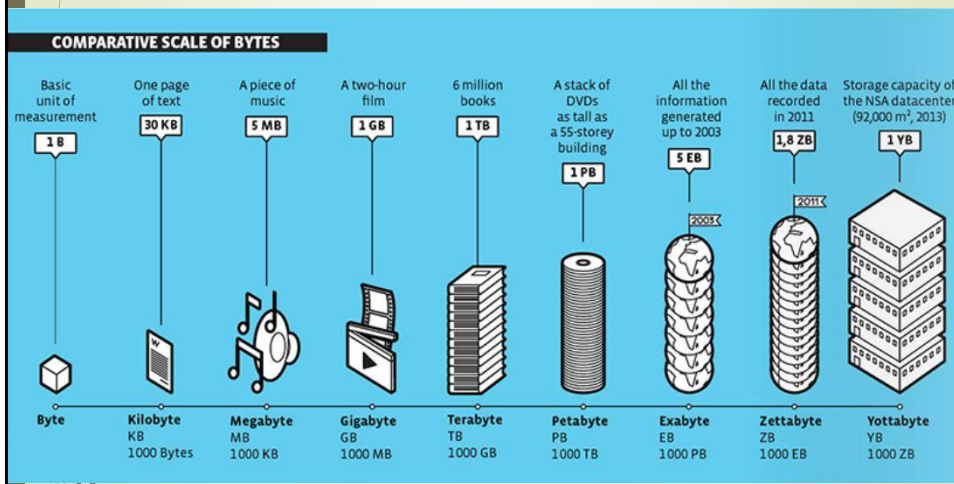
- What is the maximum file size of your files?
  - Movies/Files/Streaming video that you have used?
  - What have you observed?
- Simple questions:
  - What's the size of your HD?
  - You know GB, TB, right?
  - PB, EB, ZB, or YB?

Still no idea?

Memory unit	Size	Binary size
kilobyte (kB/KB)	10^3	2^{10}
megabyte (MB)	10^6	2^{20}
gigabyte (GB)	10^9	2^{30}
terabyte (TB)	10^{12}	2^{40}
petabyte (PB)	10^{15}	2^{50}
exabyte (EB)	10^{18}	2^{60}
zettabyte (ZB)	10^{21}	2^{70}
yottabyte (YB)	10^{24}	2^{80}

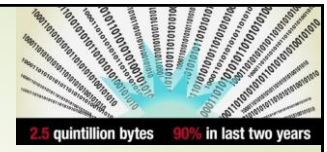
10

Comparative Scale of Bytes



11

What is Big Data?



"Every day, we create 2.5 quintillion (10^{18} or EB) bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few."

This data is "**big data**."

(https://www-01.ibm.com/software/in/data/bigdata/)

(http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/)

12

Big Data EveryWhere!

- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - Purchases at department/ grocery stores
  - Bank/Credit Card transactions
  - Social Network
  - Sensors
  - Internet of things
  - …

13

Huge Amount of Data

- There are huge volumes of data in the world:
  - From the beginning of recorded time until 2003, We created 5 billion gigabytes (exabytes) of data.
  - In 2011, the same amount was created every two days.
  - In 2013, the same amount of data is created every 10 minutes.
  - By 2020, every person will generate 1.7 MB in just a second (146.88 GB a day)!!!

14

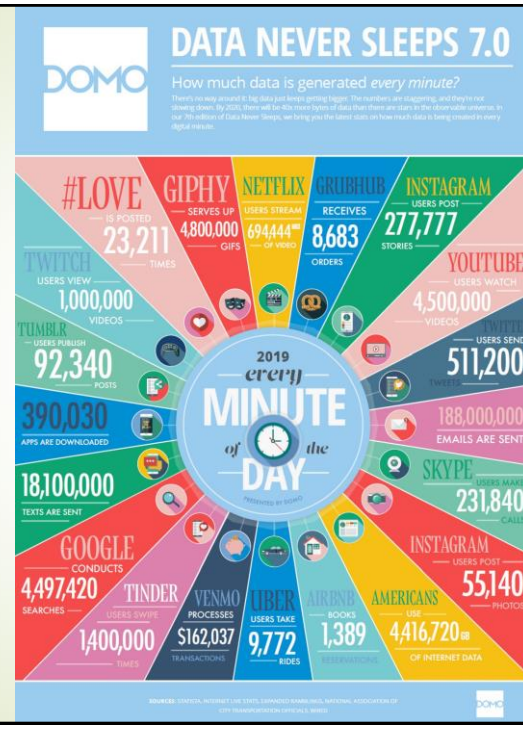
Data/Minute on the Internet

- 2014, 2.4 billion users. 3.4 billion by 2016. 3.8 billion in 2017. June 2019, over **4.4 billion internet users**.
- **Every minute**, the following happens on the internet:
  - Netflix users stream 694,444 hours of video
  - Instagram users post 277,777 stories
  - Youtube users watch 4,500,000 videos
  - Twitter users send 511,200 tweets
  - Skype users make 231,840 calls
  - Uber users take 9,772 rides
  - Google conducts 4,497,420 searches
  - Facebook users send 3 million posts, 510,000 comments, 293,000 status updates, 136,000 photos, 4+ million likes
  - Giphy serves up 4,800,000 gifs

(https://www.visualcapitalist.com/big-data-keeps-getting-bigger/)

15

Data Never Sleeps



16

How much data per day?

- 1,209,600 new data on social media
- 682 million tweets
- 4+ million hours of video upload, 5.97 billion hours watched on Youtube
- 67,305,600 Instagram posts
- 2+ billion monthly active and 1.58 billion daily active Facebook users (Q2 2019)
- 4.3 billion Facebook messages posted
- 5.76 billion Facebook likes
- 5.6 billion daily Google searches in 2019.

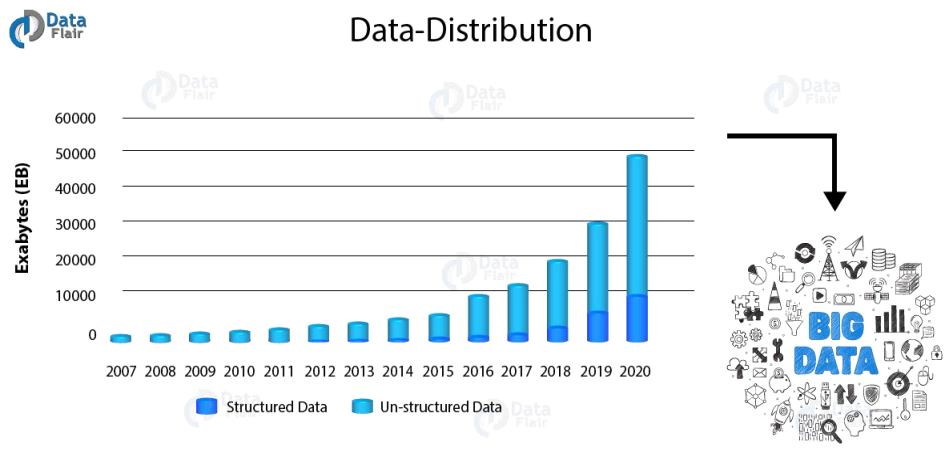
17

How Big is Big?

- Stu Feldman (Chief Scientist of Schmidt Futures, ex Google VP) says at least 10TB in terms of data rate.
  - **PB** level is now commonly recognized as big
- "Please call your work Small Data, NOT Big Data."
- Luckily, Google agrees that their work is Extreme Data.

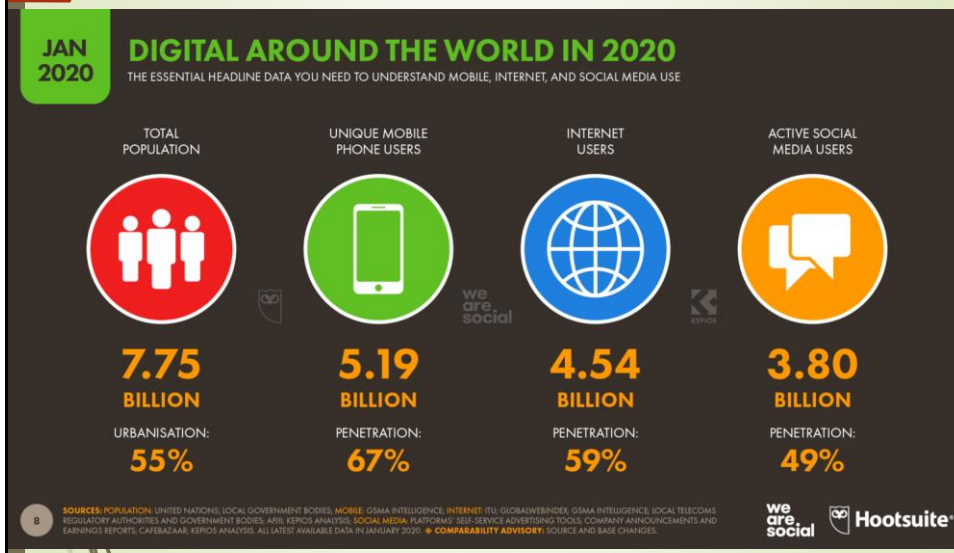
18

Data is Growing Exponentially



19

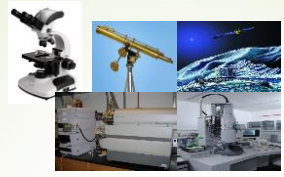
Digital World in 2020



20

Who's Generating Big Data

Social media and networks
(all of us are generating data)



Scientific instruments
(collecting all sorts of data)

Mobile devices
(tracking all objects all the time)

Sensor technology and networks
(measuring all kinds of data)

- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

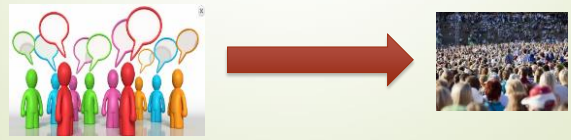
21

The Model Has Changed…

- The model of generating/consuming data has changed

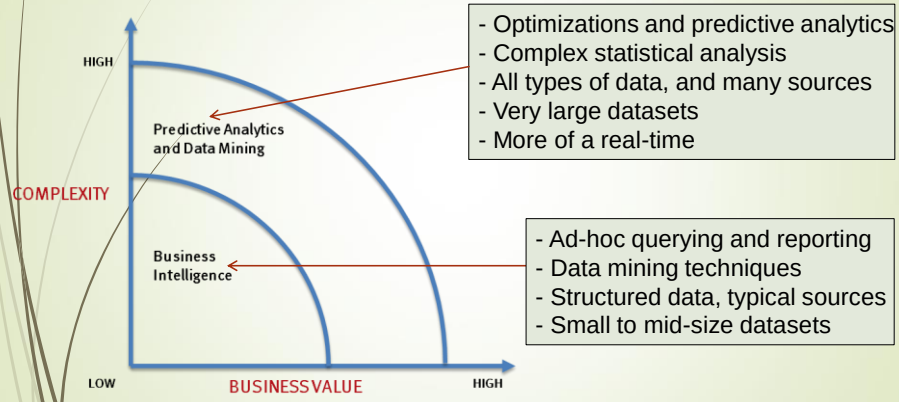
Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data



22

What's Driving Big Data



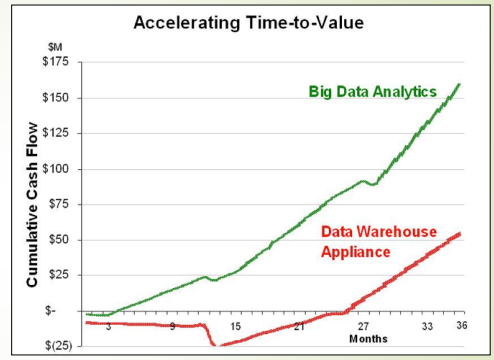
- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of a real-time

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

23

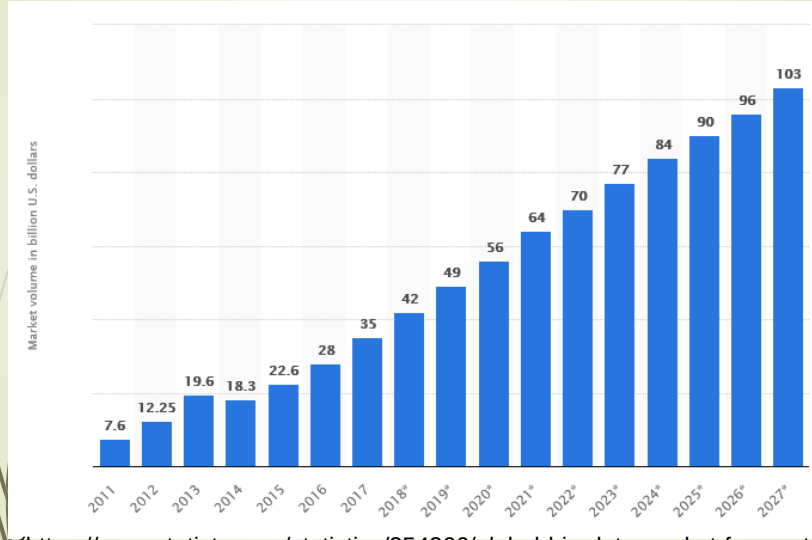
Value of Big Data Analytics

- Big data is more real-time in nature than traditional DW applications
- Traditional DW architectures (e.g. Exadata, Teradata) are not well-suited for big data apps
- Shared nothing, massively parallel processing, scale out architectures are well-suited for big data apps



24

Big Data Market Forecast



(https://www.statista.com/statistics/254266/global-big-data-market-forecast/)

25

Types of Data

- Relational data (Tables/Transaction/Legacy Data)
- Text data (Web)
- Semi-structured data (XML)
- Unstructured data
- Multimedia data (images, audio, video)
- Graph data
  - Social Network, Semantic Web (RDF), …
- Streaming data
  - You can only scan the data once

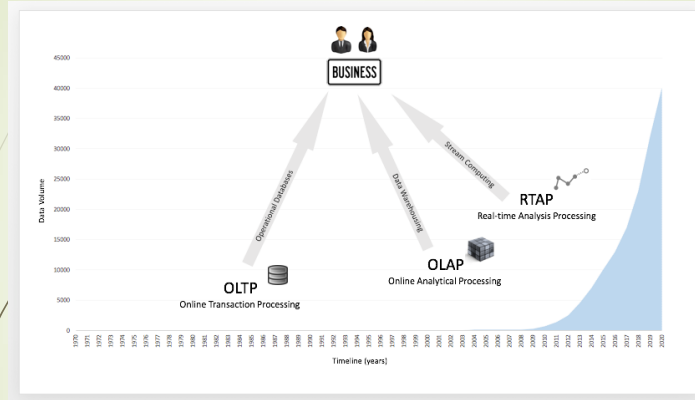
26

What to do with these data?

- Aggregation and Statistics
  - Data warehouse and OLAP
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling
- Machine learning

27

Harnessing Big Data

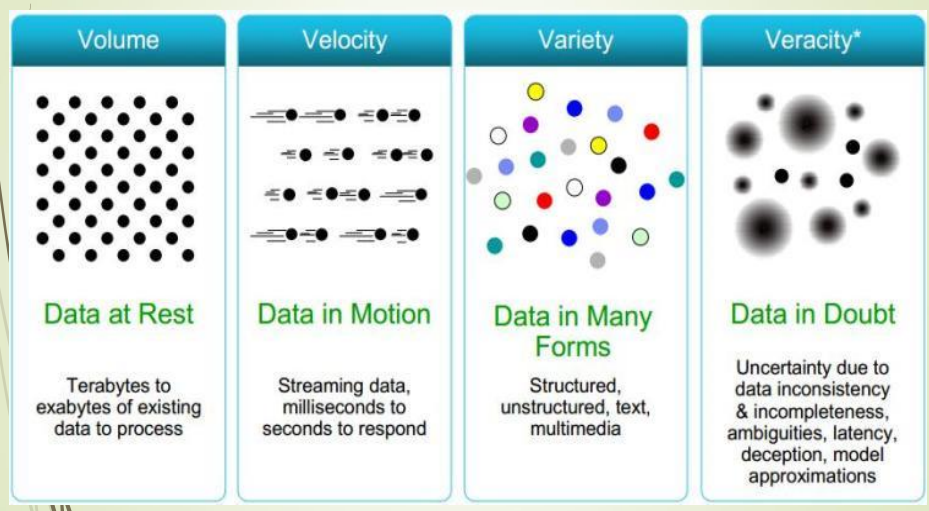


- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data warehousing)
- **RTAP:** Real-Time Analytics Processing (Big data technology)

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

28

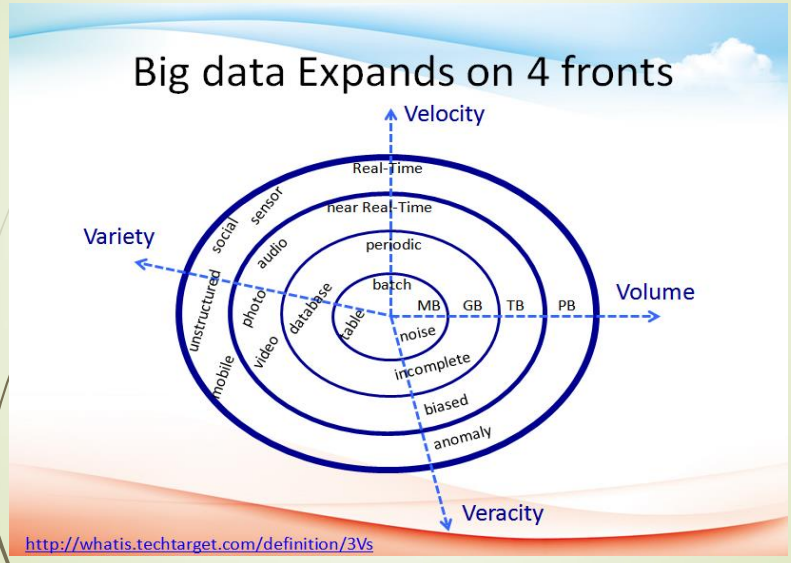
4 Vs of Big Data



Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

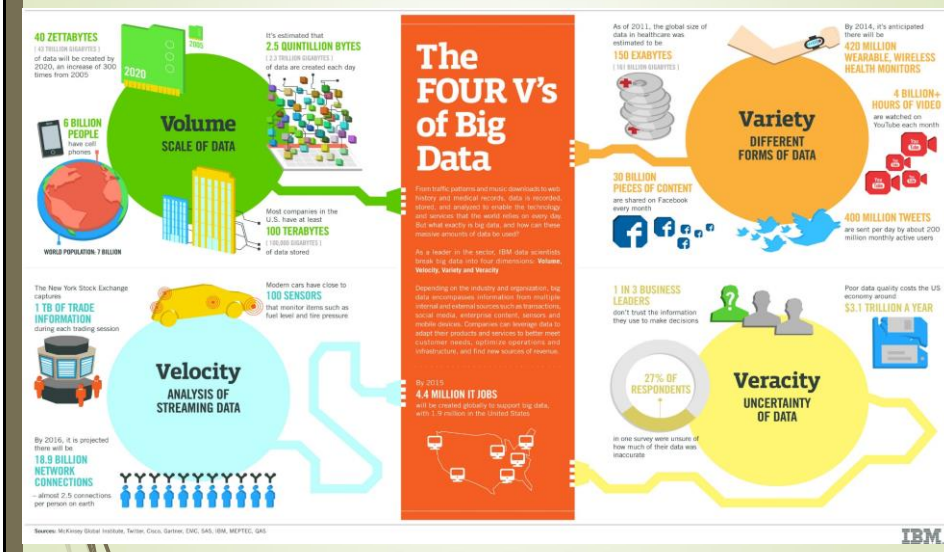
29

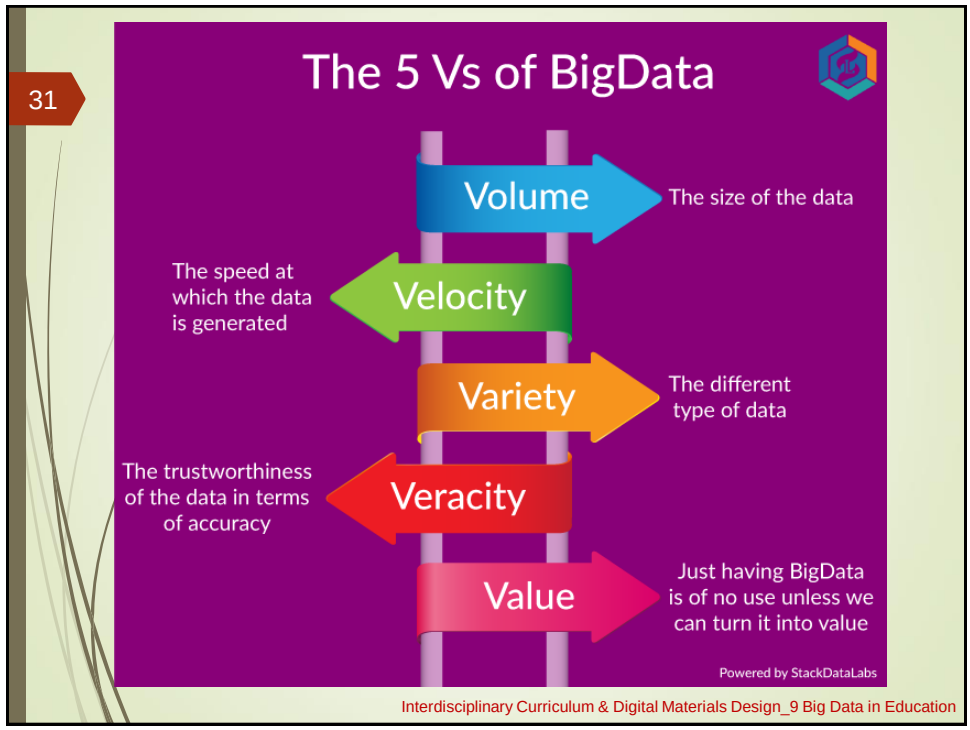
4 Vs of Big Data

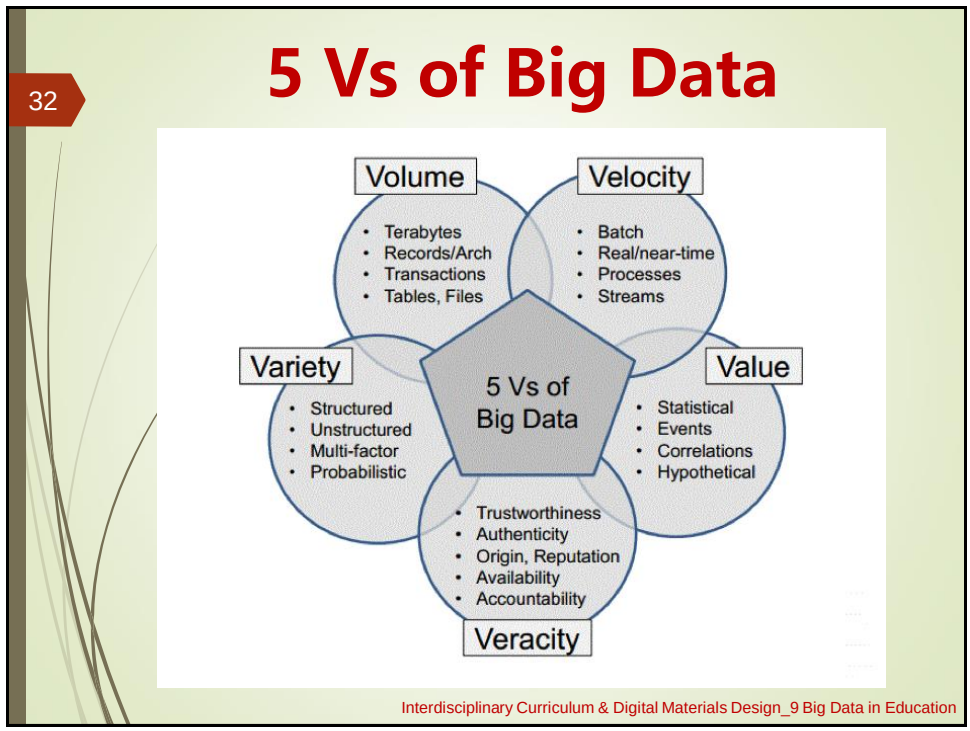


30

4 Vs of Big Data




31
The 5 Vs of BigData
Volume
The size of the data
The speed at which the data is generated
Velocity
Variety
The different type of data
The trustworthiness of the data in terms of accuracy
Veracity
Value
Just having BigData is of no use unless we can turn it into value
Powered by StackDataLabs
Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education


32
5 Vs of Big Data
Volume
Terabytes
Records/Arch
Transactions
Tables, Files
Velocity
Batch
Real/near-time
Processes
Streams
Variety
Structured
Unstructured
Multi-factor
Probabilistic
5 Vs of Big Data
Value
Statistical
Events
Correlations
Hypothetical
Trustworthiness
Authenticity
Origin, Reputation
Availability
Accountability
Veracity
Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

33

8 Vs of Big Data

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

34

10 Vs of Big Data

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

35

Meanings of Some Vs

- **Validity**: The issue of collecting data which is correct and accurate for the intended use.
- **Volatility**: How long is data valid and how long should it be stored.
- **Variability**: Big data is variable, i.e. variance in meaning, changing of meaning (rapidly).\
- **Visualization**: Making data comprehensible, easy to understand and read.
- The list keeps growing to **42 Vs** !!

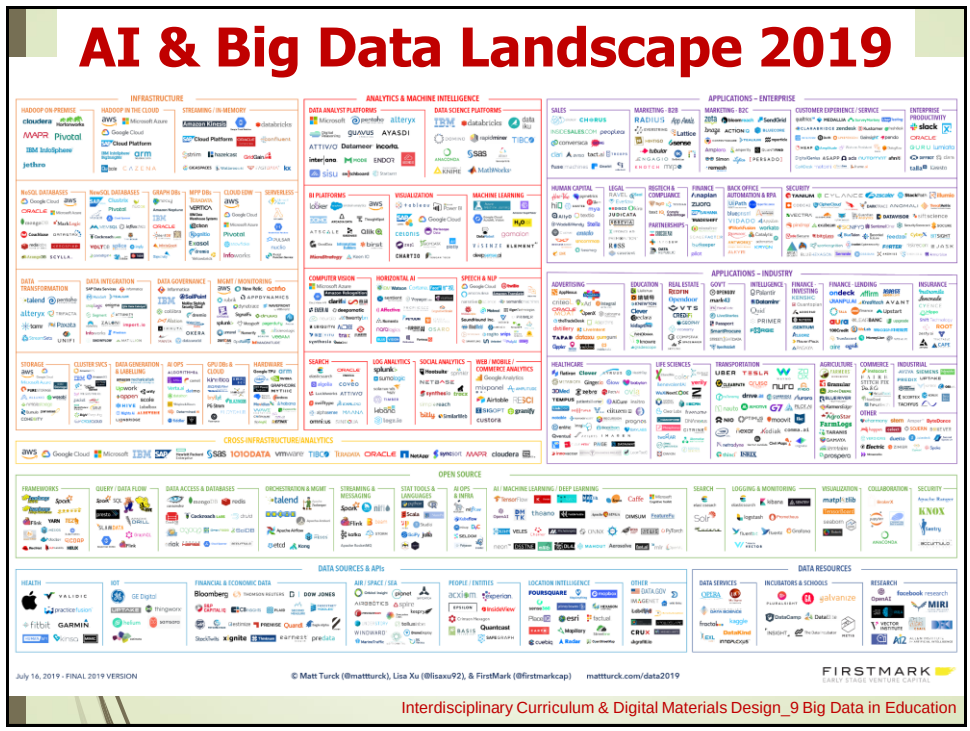
36

Emerging Technologies for Managing Big Data

- Architecture
- Storage
- Computing
- Graph
- Database/Data warehousing
- Stream processing
- Real-time Analytics & Business knowledge
- Big data as a service

37

Evolving Big Data Tools



How Big Data can help in Education 1

39

- Empowers better decision-making
  - From experience-driven to data/analytics-driven
- Students' results
  - Detail tracking and analysis of students' progress
  - Large scale comparison of students' performance
- Career prediction
  - With analysis of a student's weaknesses and strengths, can offer career recommendation.

How Big Data can help in Education 2

40

- The mapping concept
  - Provide evidence to comprehend the learning models of students.
  - Prompt the situations of creative learning, self-learning or group learning.
- Enhance learning experience
  - Learning can be more effective by tailoring toward individual student's progress and needs with analytics and mapping.

41

How Big Data can help in Education 3

- Analytics for educators
  - Educators can improve their teaching skills based on analytics of the effectiveness of different subjects and methods.
- Increase participation
  - With big data analytics and customization of learning content, more students are entering MOOCs and distance learning courses.
  - Democratization of education

42

Use Cases - Khan Academy

- Khan Academy provides users and teachers with extensive amounts of data that can be used to teach more effectively:
  - to know when a concept is mastered
  - to know when a student needs more help
  - to know how students can teach one another, etc
- To shape the interactions between teachers and students based on data is foundational in their system of teaching.
- A widely viewed TED talk by Salman Khan (https://youtu.be/nTFEUsudhfs)

43

Use Cases – Georgia State Uinversity

- Many students were collecting credits but not getting closer to graduation.
- They set up a system that track, "the thousands of decisions that students make every day and predict the likelihood of their academic success on that basis."
- Using 10 years' worth of data on every student, grade and course to develop and calibrate the model.
- The average time to graduation has fallen by half a semester (saving $15 million tuition fees)

44

Use Cases – University of Tasmania

- With over 26,000 students, they deployed a learning management systems (LMS) to know how exactly a student is using the facility to learn.
- The system helps in tracking student's learning progress and troubles.
- By offering customized guidance, the overall progress of the student is ascertained.

45

Use Cases – Arizona State Uinversity

- They used big data analytics to customize their learning process and to create individual profiles for each student.
- The idea was to spend more time supporting students individually and less time on generalized lecturing.
- The result was excellent — ASU reduced the number of students dropping out by 50% and increased passing rates by 10%.

46

Use Cases – Purdue University

- PU developed a system called ***Course Signals*** which helps with detecting academic and behavioral issues among students.
- Through data mining, it can measure the engagement of students, academic preparation and effort levels and create a profile for each student with attention on the risk of failing or dropping out.
- "As and Bs have increased by as much as 28% in some courses. In most cases, the greatest improvement is seen in students who were initially receiving Cs and Ds in early assignments, and pull up half a letter grade or more to a B or C."

47

Concluding Remarks

- Students are faced with serious academic challenges which are often recognized too late.
- Successful use cases of BDA show that BDA can be applied to both online education and more traditional learning environments.
- Using big data can help educational institutions and teachers to detect and correct problems early before it's too late.