

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

2

Lecture Topics

- 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!

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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**.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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/>)

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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**.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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”!**

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

Let's Start

9

- ▶ 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?

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}

Still no idea?

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

Comparative Scale of Bytes

10


COMPARATIVE SCALE OF BYTES

Unit	Example	Conversion
Byte	Basic unit of measurement	1 B
Kilobyte (KB)	One page of text	30 KB = 1000 Bytes
Megabyte (MB)	A piece of music	5 MB = 1000 KB
Gigabyte (GB)	A two-hour film	1 GB = 1000 MB
Terabyte (TB)	6 million books	1 TB = 1000 GB
Petabyte (PB)	A stack of DVDs as tall as a 55-storey building	1 PB = 1000 TB
Exabyte (EB)	All the information generated up to 2003	5 EB = 1000 PB
Zettabyte (ZB)	All the data recorded in 2011	1.8 ZB = 1000 EB
Yottabyte (YB)	Storage capacity of the NSA datacenter (92,000 m ² , 2013)	1 YB = 1000 ZB

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

11

What is Big Data?



2.5 quintillion bytes 90% in last two years

“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/>)

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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
 - ▶ ...



Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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)!!!**

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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/>)

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education



- 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.
- Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

How Big is Big?

17

- 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**.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

Data is Growing Exponentially

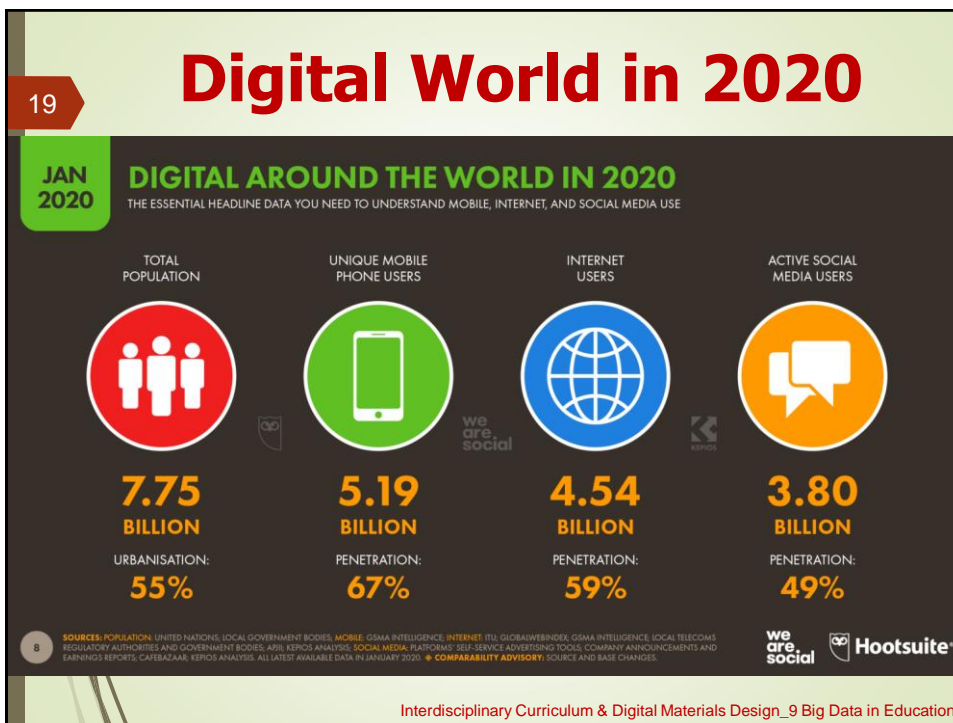
18

Data-Distribution

Year	Structured Data (EB)	Un-structured Data (EB)	Total Data (EB)
2007	~100	~100	~200
2008	~200	~200	~400
2009	~400	~400	~800
2010	~800	~800	~1600
2011	~1600	~1600	~3200
2012	~3200	~3200	~6400
2013	~6400	~6400	~12800
2014	~12800	~12800	~25600
2015	~25600	~25600	~51200
2016	~51200	~51200	~102400
2017	~102400	~102400	~204800
2018	~204800	~204800	~409600
2019	~409600	~409600	~819200
2020	~819200	~819200	~1638400


Legend: ■ Structured Data ■ Un-structured Data

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education




Who's Generating Big Data


20




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

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

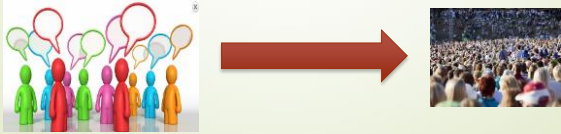
21 **The Model Has Changed...**

➤ The model of generating/consuming data has changed

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

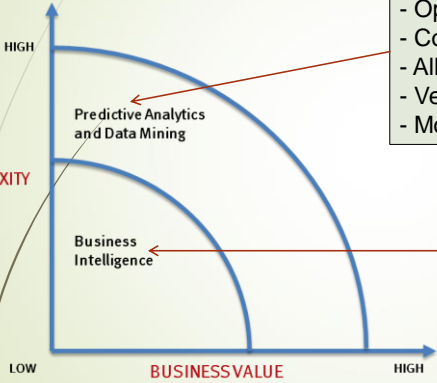


New Model: all of us are generating data, and all of us are consuming data



Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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

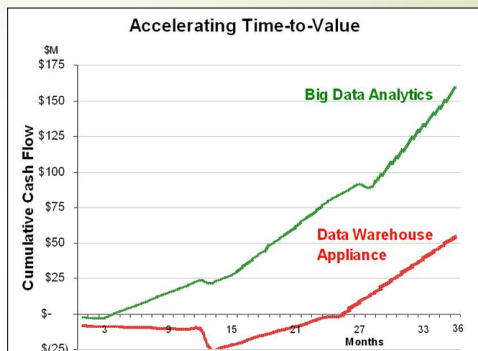
- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

23

Value of Big Data Analytics

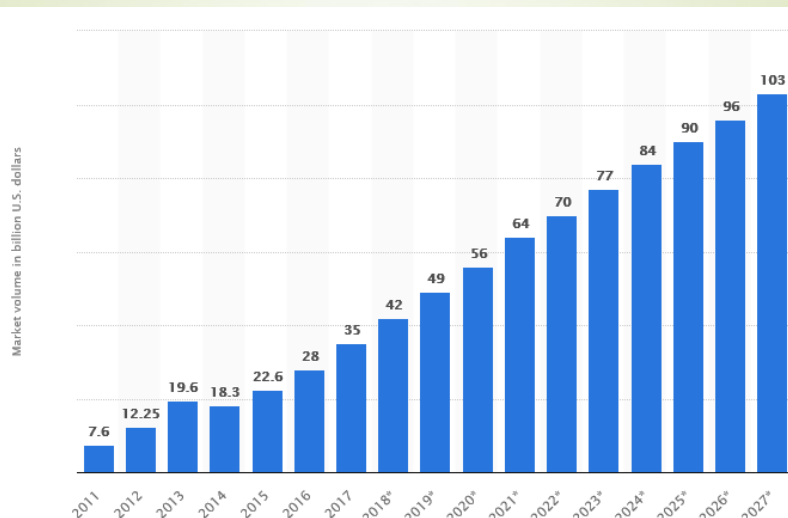
- Big data is more real-time in nature than traditional DW applications
- Traditional DW architectures (e.g. Exadata, Teradata) are not well-suited for big data apps
- Shared nothing, massively parallel processing, scale out architectures are well-suited for big data apps



Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

24

Big Data Market Forecast



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

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

27

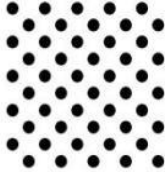
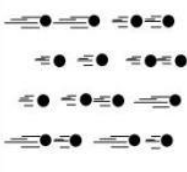
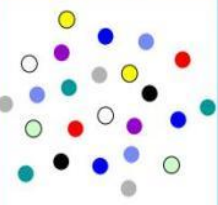

Harnessing Big Data

- **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

Volume	Velocity	Variety	Veracity*
			
<p>Data at Rest</p> <p>Terabytes to exabytes of existing data to process</p>	<p>Data in Motion</p> <p>Streaming data, milliseconds to seconds to respond</p>	<p>Data in Many Forms</p> <p>Structured, unstructured, text, multimedia</p>	<p>Data in Doubt</p> <p>Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations</p>

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

29

4 Vs of Big Data

Big data Expands on 4 fronts

<http://whatis.techtarget.com/definition/3Vs>

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

30

4 Vs of Big Data

40 ZETTABYTES
(40 TRILLION GIGABYTES)
of data will be created by 2020, an increase of 300 times from 2005

6 BILLION PEOPLE have cell phones

VOICE POPULATION 7 BILLION

Volume
SCALE OF DATA

It's estimated that **2.5 QUINTILLION BYTES** (2.5 TRILLION GIGABYTES) of data are created each day

Most companies in the U.S. have at least **100 TERABYTES** (100,000 GIGABYTES) of data stored

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session

Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

By 2015, it is projected there will be **10.9 BILLION NETWORK CONNECTIONS** - almost 2.5 connections per person on earth

Sources: McKinsey Global Institute, Tactix, Cisco, Sarban, EMC, SAS, IBM, MPTTEC, SAS

The FOUR V's of Big Data

From traffic patterns and medical diagnoses to sales history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world lives on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, intelligent customer sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, reduce the cost of operations and infrastructure, and find new sources of revenue.

By 2015, **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.8 million in the United States

As of 2011, the global size of data in healthcare was estimated to be **150 EXABYTES** (151 BILLION GIGABYTES)

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEOS are watched on YouTube each month

30 BILLION PIECES OF CONTENT are shared on Facebook every month

400 MILLION TWEETS are sent per day by about 200 million monthly active users

Variety
DIFFERENT FORMS OF DATA

1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions

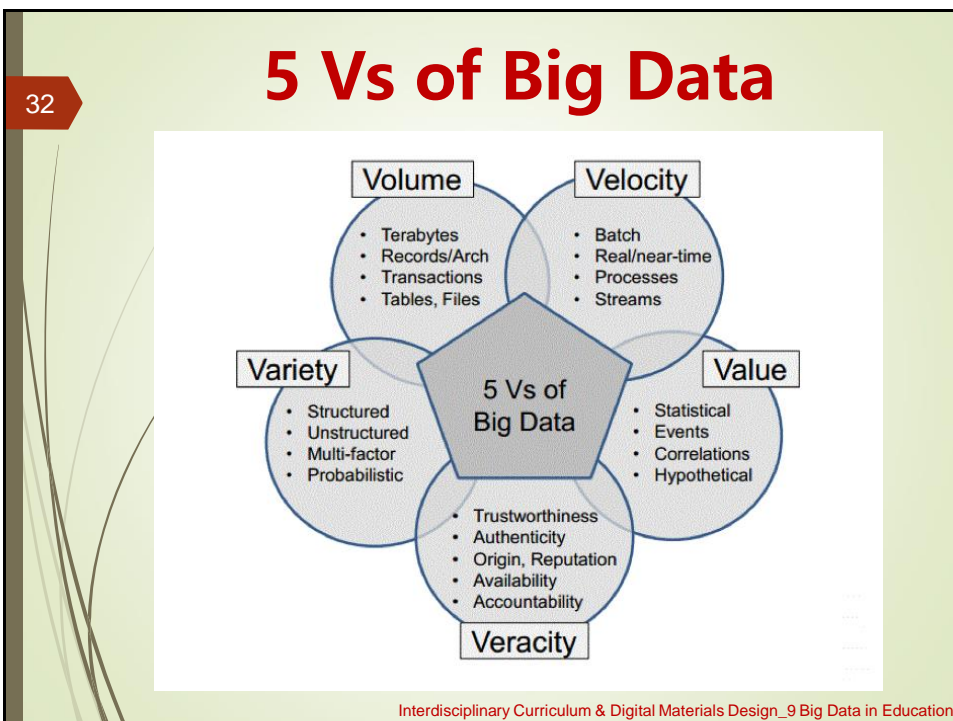
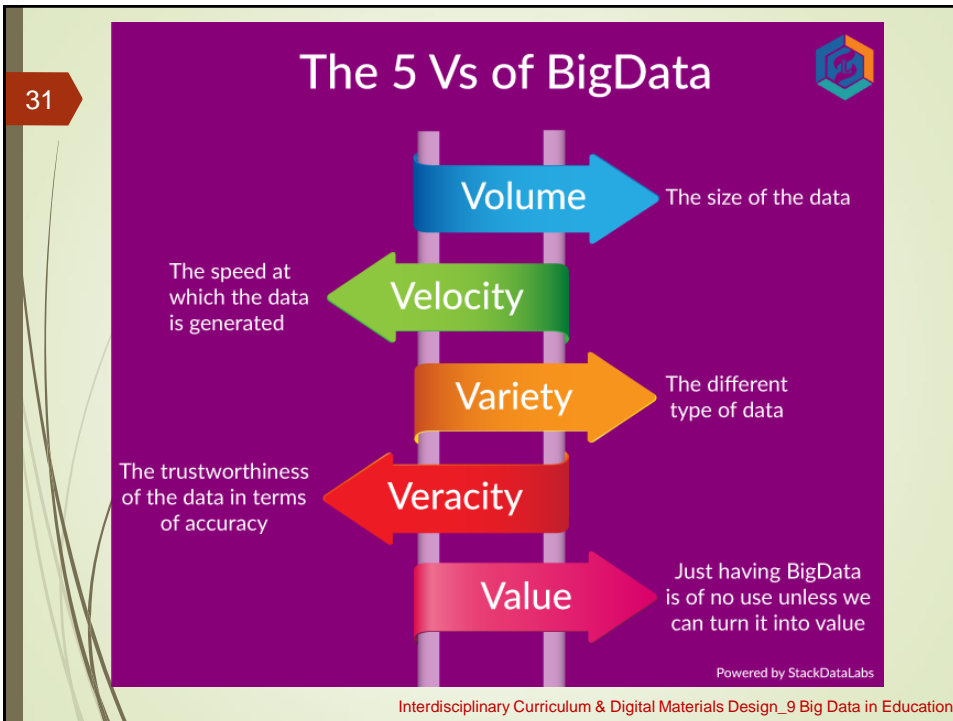
27% OF RESPONDENTS in one survey were unsure of how much of their data was inaccurate

Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**

Veracity
UNCERTAINTY OF DATA

IBM

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education



33

8 Vs of Big Data

BIG DATA
with 8 V's

- 01 VOLUME**: Can you find the information you are looking for?
- 02 VALUE**: Can you find it when you most need it?
- 03 VERACITY**: Are you dealing with information or disinformation?
- 04 VISUALISATION**: Can you make sense at a glance? Does it trigger a decision?
- 05 VARIETY**: Is a picture worth a thousand words in 70 languages? Is your information balanced?
- 06 VELOCITY**: Information gains momentum and crises & opportunities evolve in real time. How is outlook for today?
- 07 VISCOSITY**: Does it stick with you? Does it call for action?
- 08 VIRALITY**: AMA to-go? Does it convey a message that can be passed into a presentation or Instagram?

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

34

10 Vs of Big Data

BigData

- Volume**: Size of Data
- Velocity**: The Speed at which Data is Generated
- Variety**: Different type of Data
- Veracity**: Data Accuracy
- Value**: Useful Data
- Validity**: Data quality, Governance, Master Data Management on Massive
- Variability**: Dynamic, Evolving Behavior in Data Source
- Venue**: Distributed Heterogeneous Data from Multiple Platforms
- Vocabulary**: Data Models, Semantics that describes data Structure
- Vagueness**: Confusion over Meaning of BigData and Tools used

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** !!

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

37

Evolving Big Data Tools

ANALYTICS

<p>DATA ANALYST PLATFORMS</p> <p>Microsoft pentaho alteryx Digital Reasoning guAVUS AYASDI ATTIV/O Datameer Quid incorta. interana ClearStory Origami GENES ENDOR MODE Bottlenose switchboard</p>	<p>DATA SCIENCE PLATFORMS</p> <p>IBM KNIME dataiku DOMINO rapidminer CONTINUUM ANALYTICS ALGORITHMIA DATAWATCH ANGOSS SAS</p>
<p>BI PLATFORMS</p> <p>Microsoft aws DOMC Wave Analytics looker thoughtspot ARCADA DATA AT SCALE GoodData Information Builders MicroStrategy birist</p>	<p>VISUALIZATION</p> <p>tableau SAP Google Cloud CELONIS Qlik Periscope Data ZEPL XCHORDA dioty CHARTIO TOUCAN TECO</p>
<p>COMPUTER VISION</p> <p>Microsoft Azure Amazon Rekognition clarifai EVER AI deepomatic twentybn INNOVIO</p>	<p>HORIZONTAL AI</p> <p>IBM Watson Cortana Face++ 旷视 sentient Voyager AI Kognata Affectiva PROPHESIZ Numenta PETUM narologies CURIOUS AI OSARO BLISS VISION</p>
<p>SEARCH</p> <p>ORACLE ENDECA elasticsearch EXALEAD COVEO Lucidworks ATTIV/O swiftype algolia alphasense MAANA omni:us SINOUIA</p>	<p>LOG ANALYTICS</p> <p>splunk sumologic LOGGLY TIBCO hibano logz.io</p>
<p>LOG ANALYTICS</p> <p>splunk sumologic LOGGLY TIBCO hibano logz.io</p>	<p>SOCIAL ANALYTICS</p> <p>Hootsuite sprinklr NETBASE synthesio tracx simple reach bitly predata SimilarWeb</p>
<p>WEB / MOBILE / COMMERCE ANALYTICS</p> <p>Google Analytics mixpanel AMPITUDE sumAll Airtable RESQI SIGOPT granify custora</p>	<p>MACHINE LEARNING</p> <p>aws Google Cloud H2O DataRobot gamalon ELEMENT VIZENZE bonsai</p>
<p>SEARCH</p> <p>ORACLE ENDECA elasticsearch EXALEAD COVEO Lucidworks ATTIV/O swiftype algolia alphasense MAANA omni:us SINOUIA</p>	<p>SPEECH & NLP</p> <p>Google Cloud twilio amazon alexa narrative IBM Watson Mobile Team Technologies SoundHound Inc. PRIMER Bluebird comcast snips Plexiplex</p>

AI & Big Data Landscape 2019

<p>ADAPTIVE FRONTEND</p> <p>cloudinary MAPR Pivotal IBM IndyPaaS Jethro</p>	<p>INFRASTRUCTURE</p> <p>DESIGN THE CLOUD</p> <p>aws Microsoft Azure Cloud Platform IBM Bluemix Google Cloud</p> <p>OPERATIONS IN-CLOUD</p> <p>IBM Red Hat VMware VMware Tanzu VMware Cloud Foundation VMware vRealize VMware vCloud VMware vSphere VMware vSAN VMware vCenter VMware vCloud Director VMware vCloud Availability VMware vCloud Lifecycle Manager VMware vCloud Lifecycle Manager VMware vCloud Lifecycle Manager</p>	<p>ANALYTICS & MACHINE INTELLIGENCE</p> <p>DATA ANALYST PLATFORMS</p> <p>Microsoft pentaho alteryx Digital Reasoning guAVUS AYASDI ATTIV/O Datameer incorta. interana ClearStory Origami GENES ENDOR MODE Bottlenose switchboard</p> <p>DATA SCIENCE PLATFORMS</p> <p>IBM KNIME dataiku DOMINO rapidminer CONTINUUM ANALYTICS ALGORITHMIA DATAWATCH ANGOSS SAS</p>	<p>DATA</p> <p>IBM Watson Cortana Face++ 旷视 sentient Voyager AI Kognata Affectiva PROPHESIZ Numenta PETUM narologies CURIOUS AI OSARO BLISS VISION</p>	<p>APPLICATIONS - ENTERPRISE</p> <p>BACK OFFICE</p> <p>IBM Watson Cortana Face++ 旷视 sentient Voyager AI Kognata Affectiva PROPHESIZ Numenta PETUM narologies CURIOUS AI OSARO BLISS VISION</p>	<p>APPLICATIONS - INDUSTRY</p> <p>MANUFACTURING</p> <p>IBM Watson Cortana Face++ 旷视 sentient Voyager AI Kognata Affectiva PROPHESIZ Numenta PETUM narologies CURIOUS AI OSARO BLISS VISION</p>
<p>CROSS-INDUSTRY ANALYTICS</p> <p>SAS IOI DATA VISIOPRO TIBCO TRIMBLE ORACLE NETSCOUT SPARKBOX ANAPL CLOUDIFY</p>					
<p>OPEN SOURCE</p> <p>AI / MACHINE LEARNING / DEEP LEARNING</p> <p>TensorFlow PyTorch Keras Caffe MXNet Apache Spark Hadoop Hive Pig Mahout Elasticsearch Logstash Kibana Kafka Zookeeper Cassandra HBase HDFS MapReduce YARN Tez Oozie Ambari Cloudera Databricks Docker Kubernetes Ansible SaltStack Puppet Chef Vagrant Packer Terraform Bash Python Java Scala R Perl PHP JavaScript TypeScript C# F# Go Ruby Python Java Scala R Perl PHP JavaScript TypeScript C# F# Go Ruby</p>					
<p>DATA SOURCES & APIs</p> <p>API / SPACE / DATA</p> <p>IBM Watson Cortana Face++ 旷视 sentient Voyager AI Kognata Affectiva PROPHESIZ Numenta PETUM narologies CURIOUS AI OSARO BLISS VISION</p>					

July 14, 2019 - FINAL 2019 VERSION © Matt Turck (@mattturck), Lisa Xu (@lisaxu92), & FirstMark (@firstmarkcap) mattturck.com/data2019

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.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in 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>)

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

43

Use Cases – Georgia State University

- ▶ 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)

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

45

Use Cases – Arizona State University

- ▶ 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%.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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.”

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education

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.

Interdisciplinary Curriculum & Digital Materials Design_9 Big Data in Education