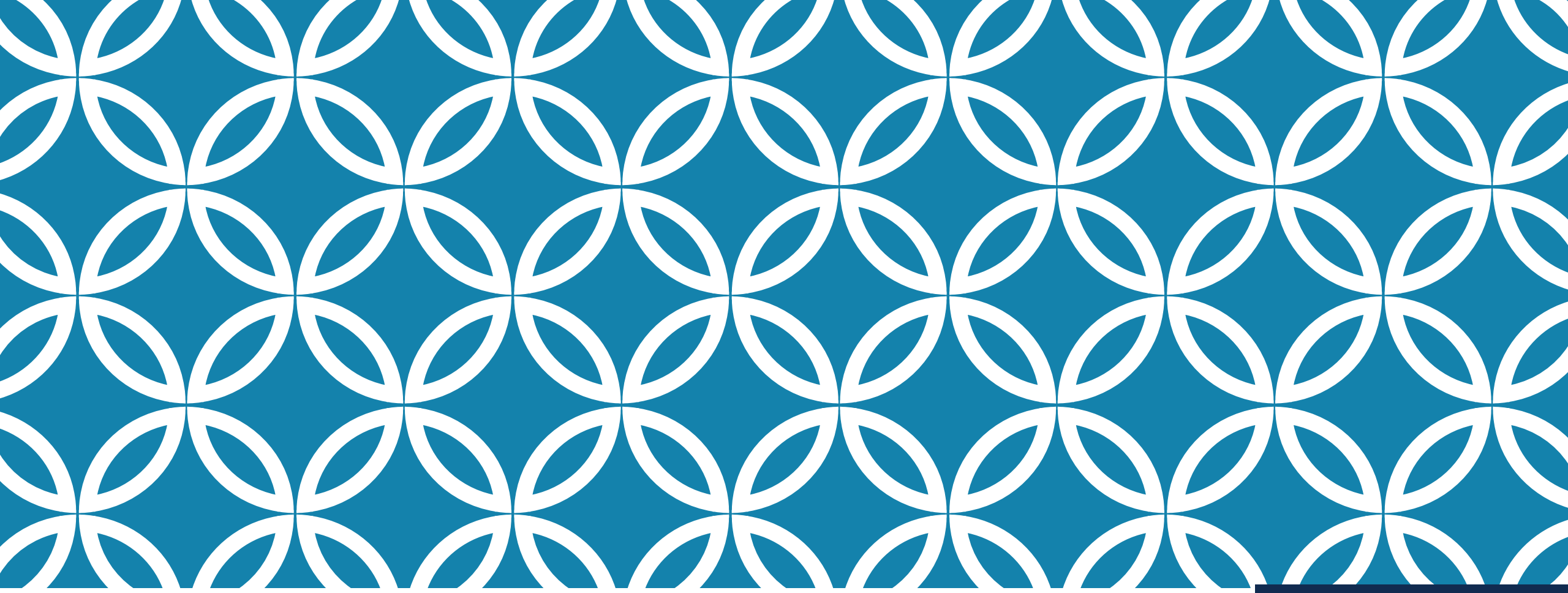
In The Name of Allah, The Most Beneficent, The Most Merciful

BS Computer Science

BIG DATA ANALYTICS

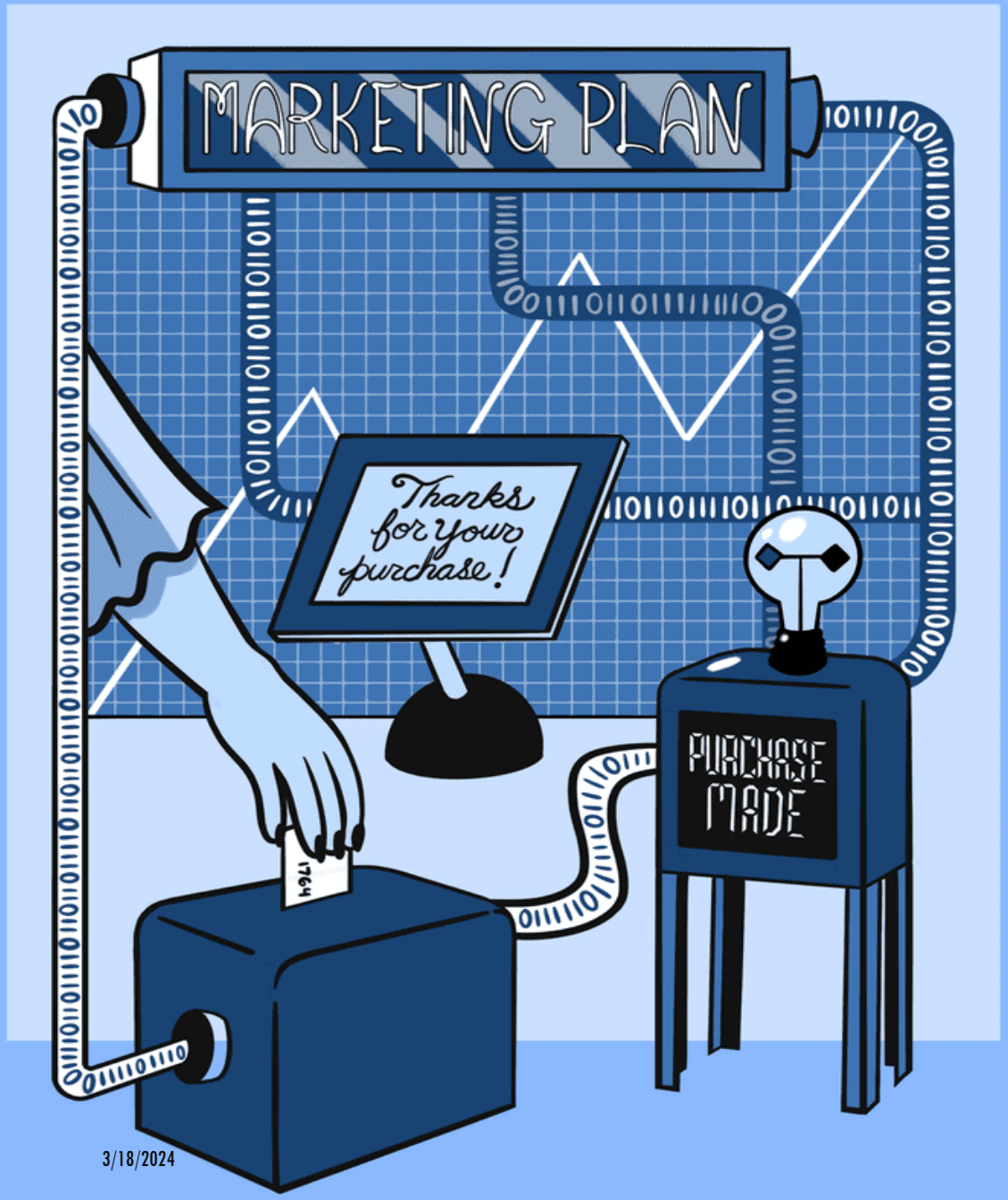
COURSE INSTRUCTOR – HASSAN JAHANGIR



UCP, LAHORE

DEPARTMENT OF COMPUTER SCIENCE

HASSAN JAHANGIR



Big Data

['big 'dā-tə]

Large, diverse sets of information that grow at ever-increasing rates.

Bigdata Analytics

Data is emerging as the world's newest resource for competitive advantage among nations, organizations and business.

1. Big data has few key characteristics such as volume, sources, velocity, variety and veracity.

2. Along with the volume, the number of sources, from where the data is extracted are also growing.

3. Data is increasingly accelerating the velocity at which it is created, as the process are moved from batch to a real time business.

4. The demands of the business from these data also has increased, from an answer next week to an answer in a minute.

BIG DATA

This flood of data is coming from many sources. Consider the following:

- The New York Stock Exchange generates about one terabyte of new trade data per day.
- Facebook hosts approximately 10 billion photos, taking up one petabyte of storage.
- Ancestry.com, the genealogy site, stores around 2.5 petabytes of data.
- The Internet Archive stores around 2 petabytes of data, and is growing at a rate of 20 terabytes per month.
- The Large Hadron Collider near Geneva, Switzerland, will produce about 15 petabytes of data per year.

INTRODUCTION

Big Data.

- Structured and Semi-Structured Data.
- Un Structured Data
- Statistical View of Big Data
- Big data in retail sector.
- ETL (Extraction Transformation Loading)

Types of big data

Structured

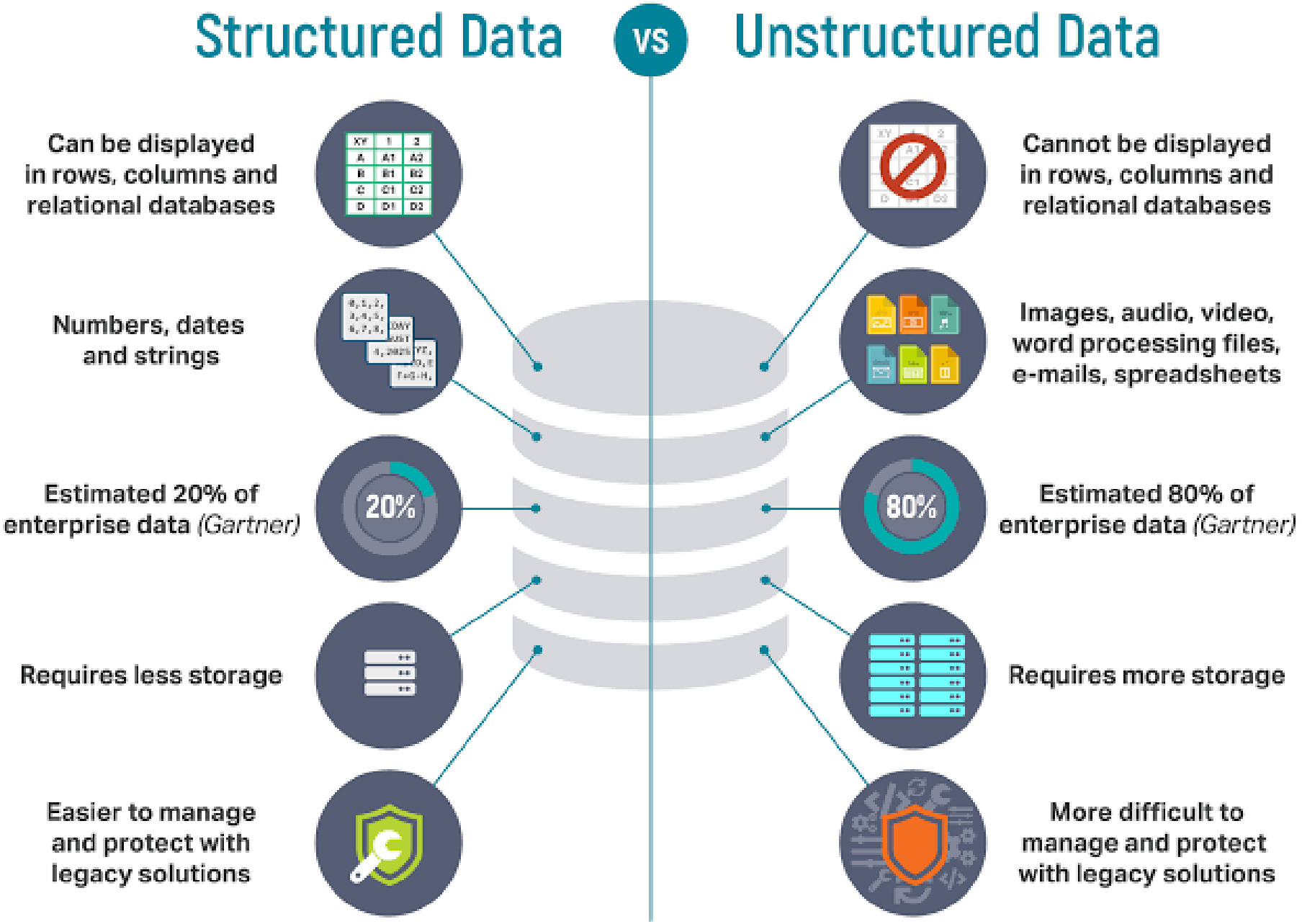
- Standardized files
- Web forms
- Spreadsheets

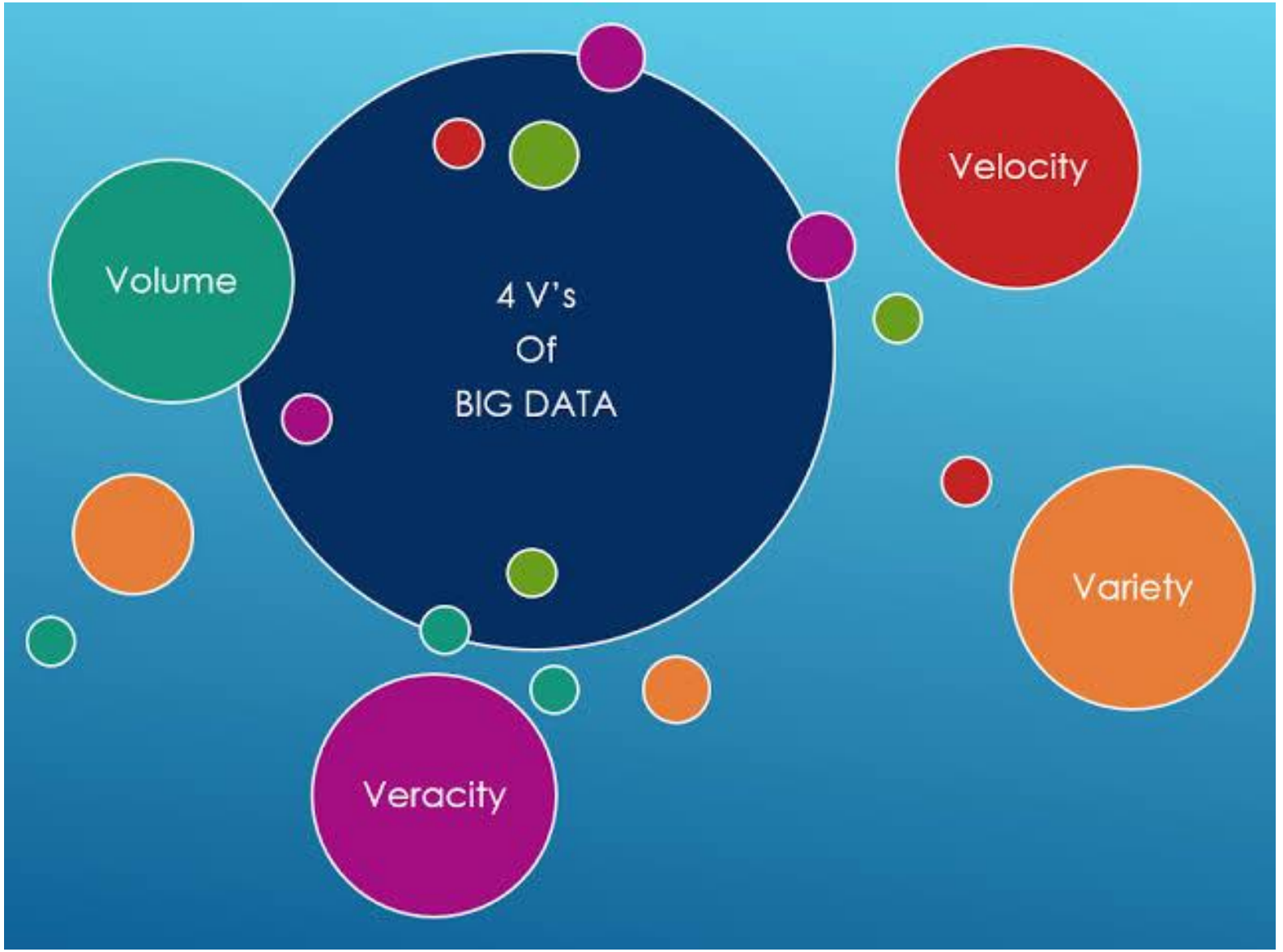
Semi-structured or hybrid

- XML language
- NoSQL databases
- HTML code

Unstructured

- Audios
- Images
- Videos


Structured Data
VS
Unstructured Data
Can be displayed in rows, columns and relational databases
Cannot be displayed in rows, columns and relational databases
Numbers, dates and strings
Images, audio, video, word processing files, e-mails, spreadsheets
Estimated 20% of enterprise data (Gartner)
20%
80%
Estimated 80% of enterprise data (Gartner)
Requires less storage
Requires more storage
Easier to manage and protect with legacy solutions
More difficult to manage and protect with legacy solutions


Velocity
Volume
4 V's
Of
BIG DATA
Variety
Veracity

7 V's of Big Data

2 Variety
It deals with a variety of data forms such as structured, unstructured and semi-structured.

1 Volume
Big data is of no use unless we can turn data into value.

3 Velocity
Speed at which data is generated.

4 Variability
Inconsistency in data can be resolved.

6 Veracity
Accuracy, trustworthiness and quality of data are paramount aspects of data.

5 Visualization
Presentation of data in visually meaningful way.

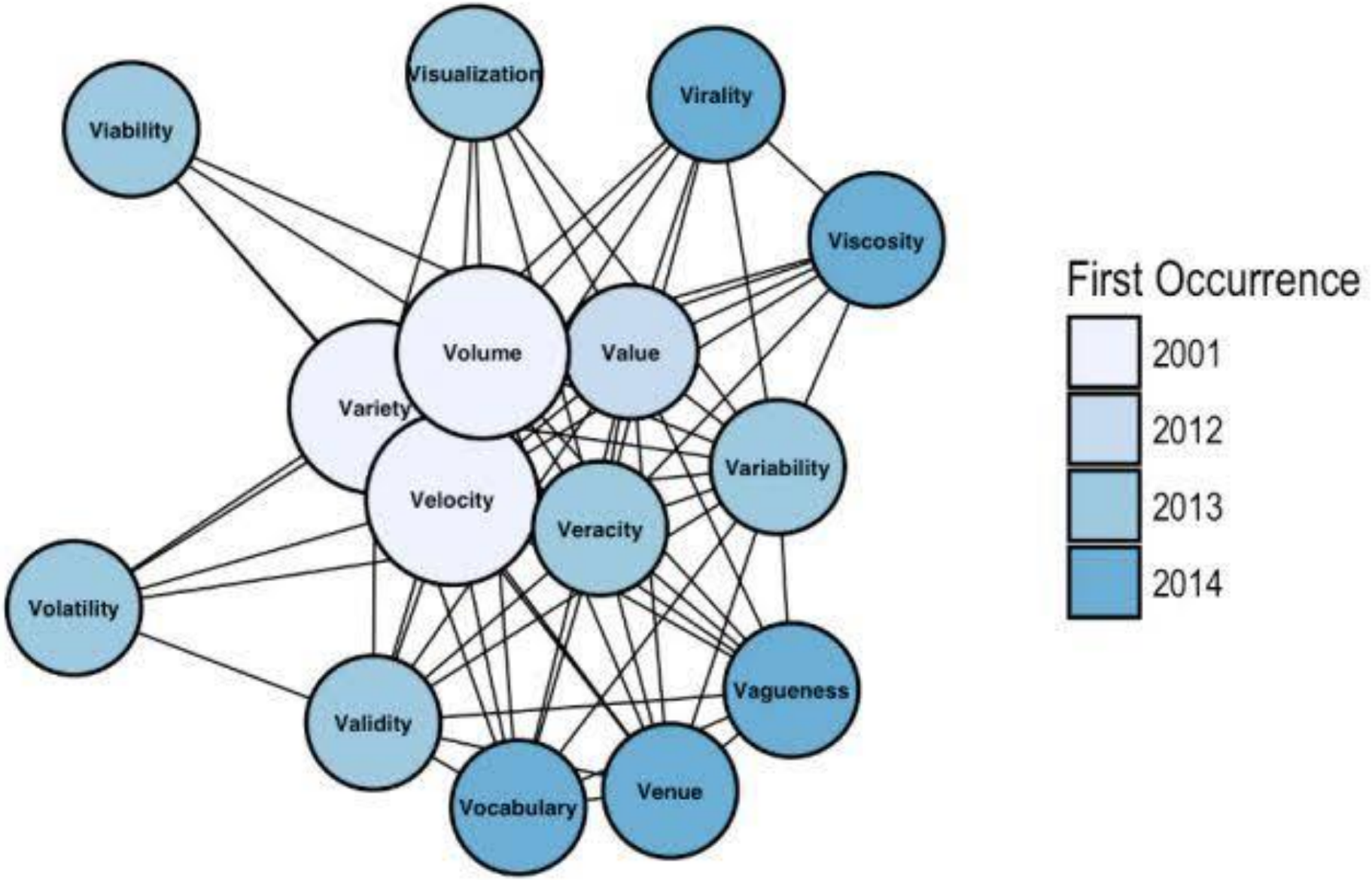
7 Value
Big data is of no use unless we can turn data into value.

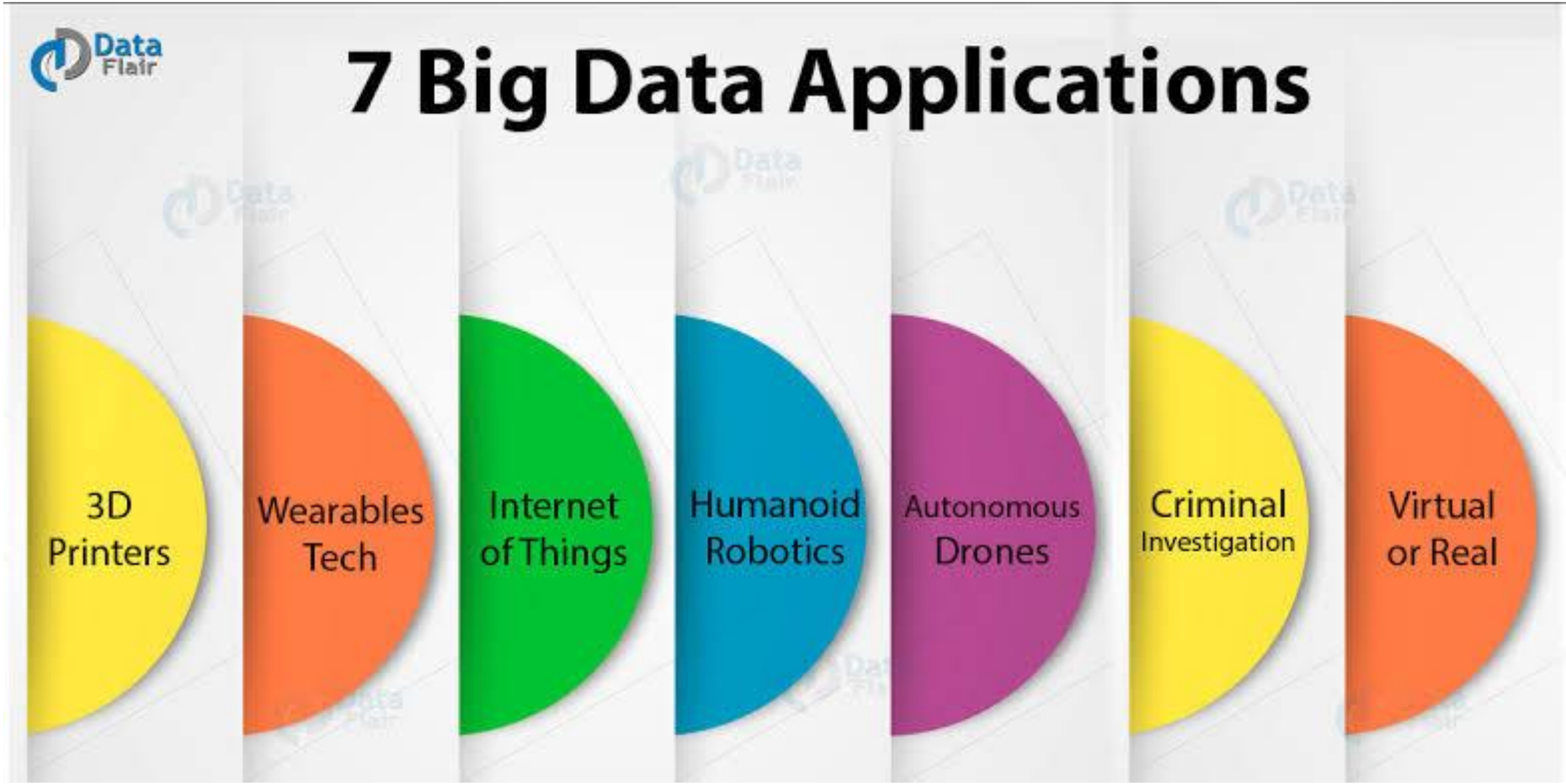
YourTechDiet.com

YourTechDiet

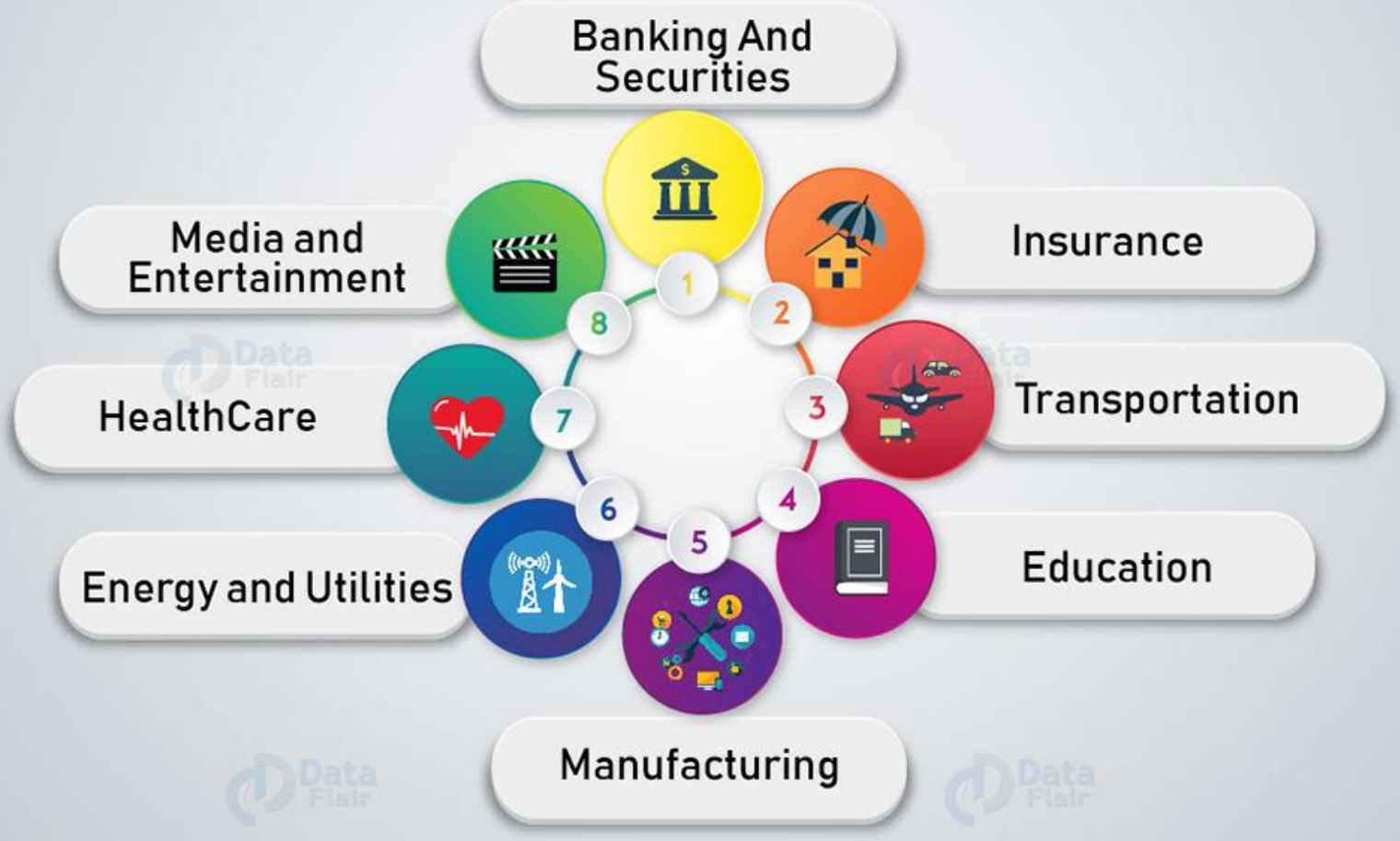
BIG DATA
with 8 V's
01
VOLUME
Can you find the information you are looking for?
02
VALUE
Can you find it when you most need it?
03
VERACITY
Are you dealing with information or disinformation?
04
VISUALISATION
Can you make sense at a glance? Does it trigger a decision?
05
VARIETY
Is a picture worth a 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
AHA to-go? Does it convey a message that can be pasted into a presentation or instagrammed?

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


Visualization
Virality
Viability
Viscosity
Volume
Value
Variety
Variability
Velocity
Veracity
Volatility
Vagueness
Validity
Vocabulary
Venue
First Occurrence
2001
2012
2013
2014


Data Flair
7 Big Data Applications
3D Printers
Wearables Tech
Internet of Things
Humanoid Robotics
Autonomous Drones
Criminal Investigation
Virtual or Real

Big Data Applications



Data Flair

Hadoop

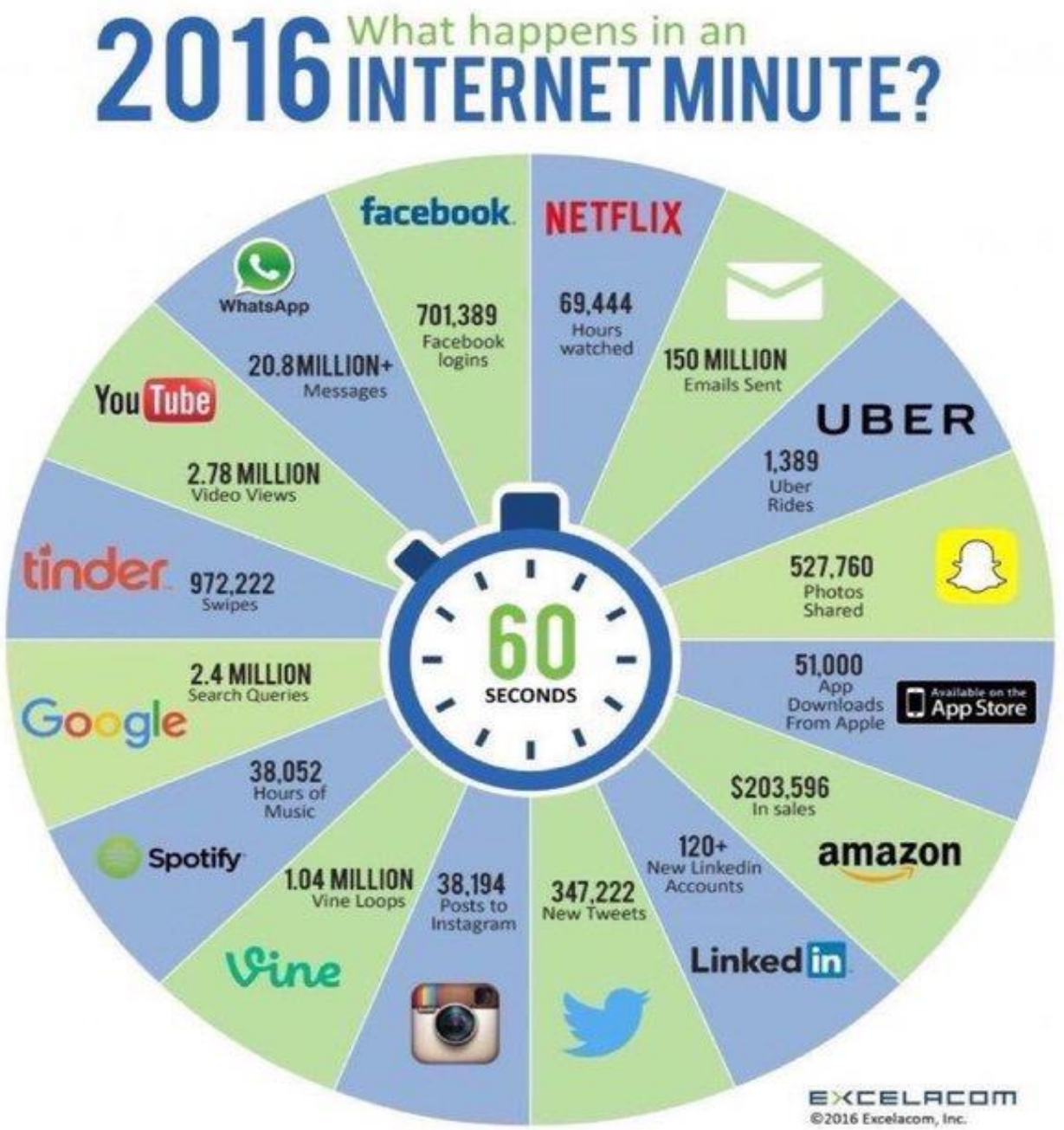
VS

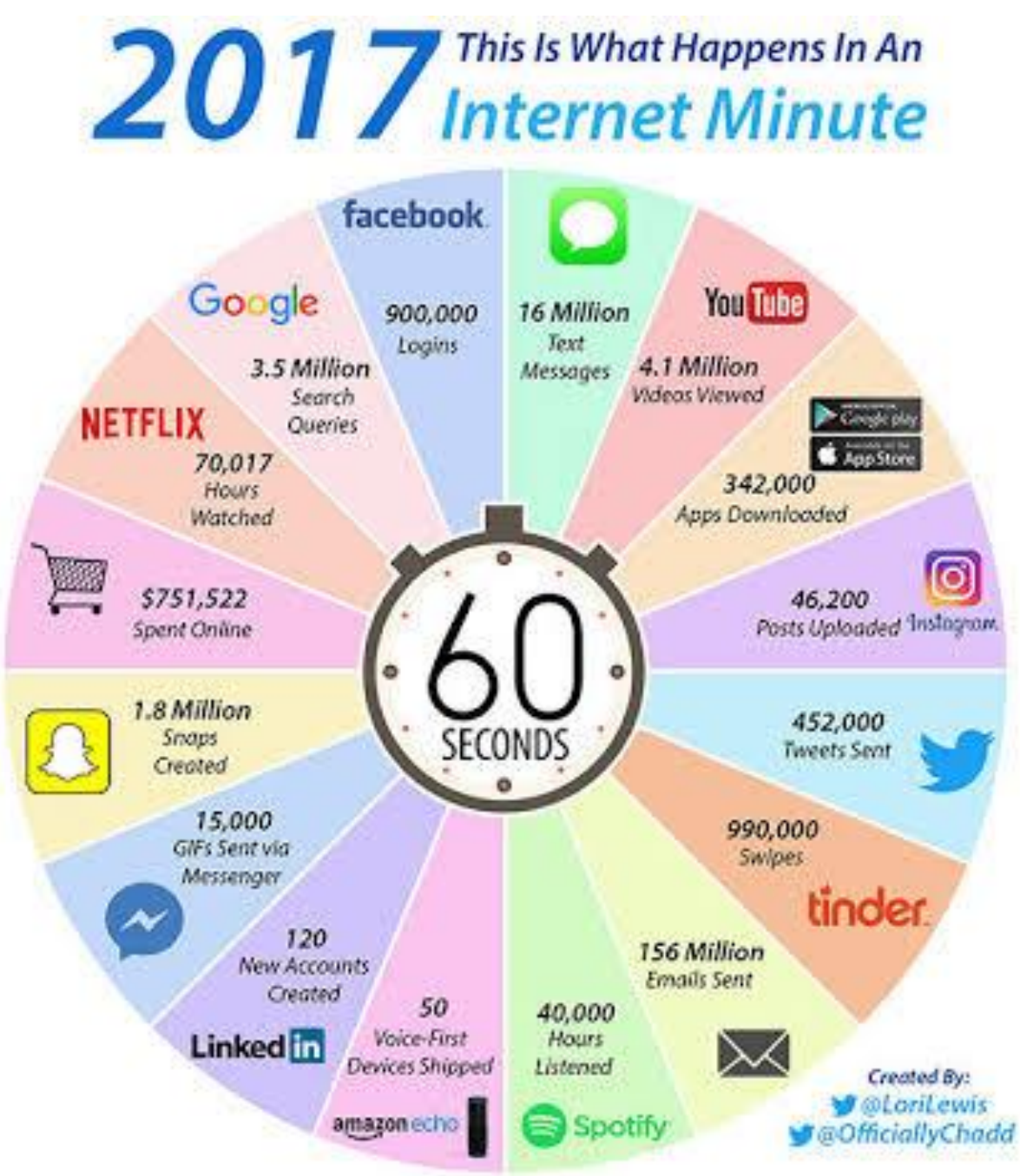
MongoDB

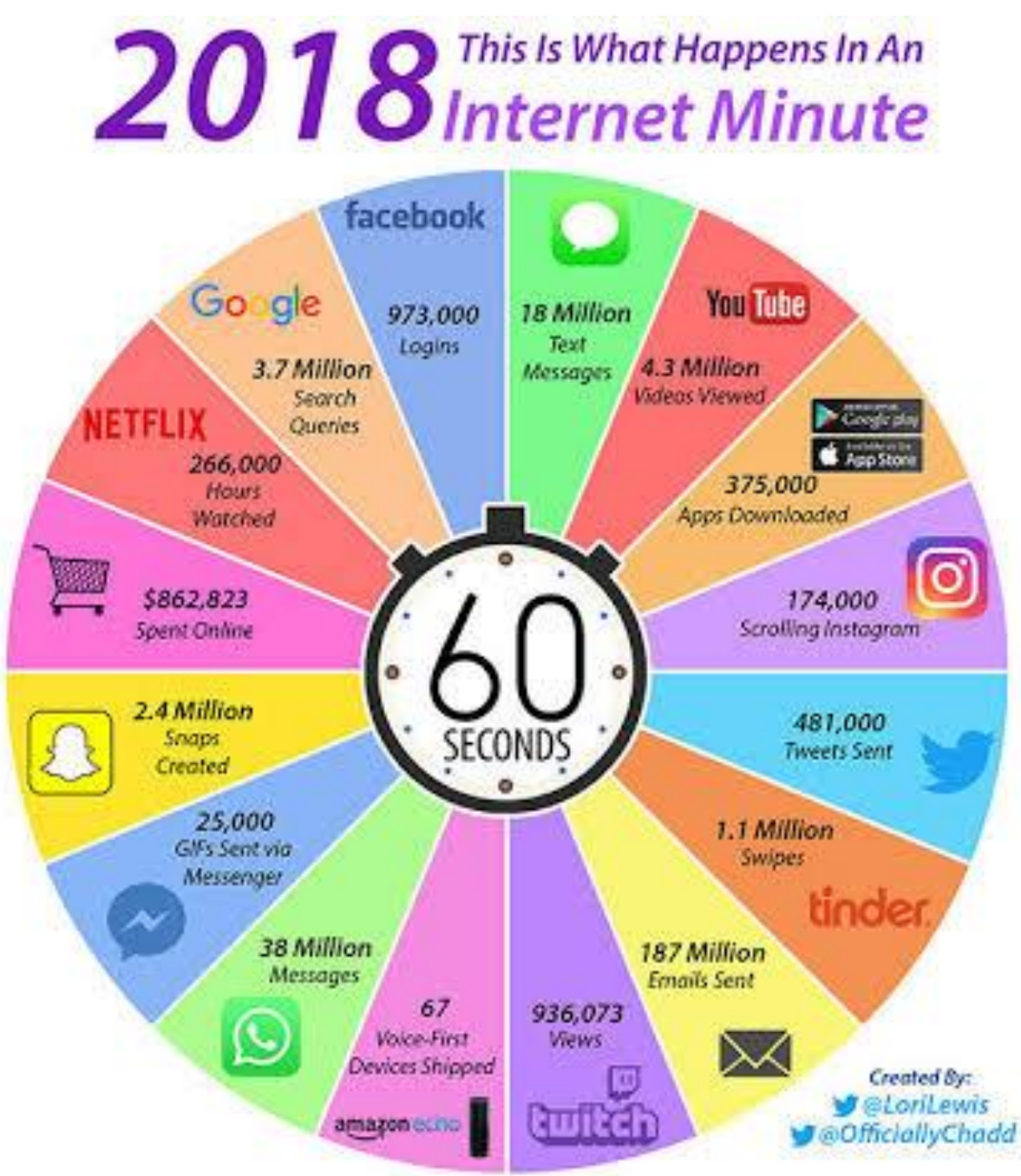
Features	Hadoop	MongoDB
Cross Platform OS	✓	✓
Map-Reduce	✓	✓
Scalable	✓	✓
NoSQL database	X	✓
File System Storage	✓	✓
Server Side Scripting	✓	✓
OLTP	X	✓
Cost effective	X	✓
Handle huge volumes of Data	✓	X

INTRODUCTION

Statistical View of Big Data 2016




2017 This Is What Happens In An Internet Minute
facebook
900,000 Logins
16 Million Text Messages
YouTube
4.1 Million Videos Viewed
342,000 Apps Downloaded
46,200 Posts Uploaded
Instagram
452,000 Tweets Sent
990,000 Swipes
tinder
156 Million Emails Sent
40,000 Hours Listened
Spotify
50 Voice-First Devices Shipped
amazon echo
120 New Accounts Created
LinkedIn
15,000 GIFs Sent via Messenger
1.8 Million Snaps Created
$751,522 Spent Online
70,017 Hours Watched
NETFLIX
3.5 Million Search Queries
Google
60 SECONDS
Created By:
@LoriLewis
@OfficiallyChadd


2018 This Is What Happens In An Internet Minute
facebook
973,000 Logins
18 Million Text Messages
YouTube
4.3 Million Videos Viewed
375,000 Apps Downloaded
174,000 Scrolling Instagram
481,000 Tweets Sent
1.1 Million Swipes
tinder
187 Million Emails Sent
936,073 Views
twitch
67 Voice-First Devices Shipped
amazon echo
38 Million Messages
25,000 GIFs Sent via Messenger
2.4 Million Snaps Created
$862,823 Spent Online
266,000 Hours Watched
NETFLIX
3.7 Million Search Queries
Google
60 SECONDS
Created By:
@LoriLewis
@OfficiallyChadd

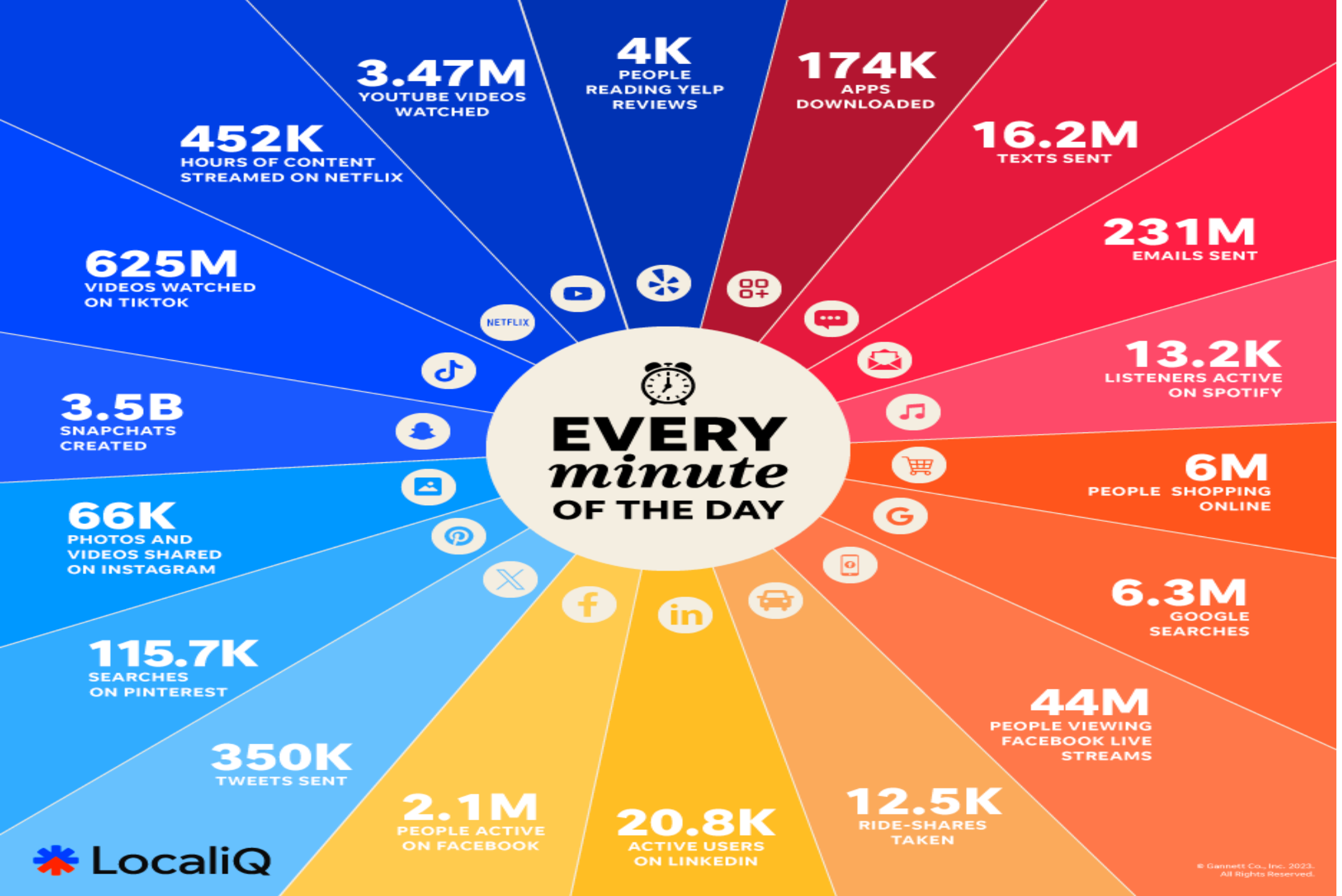
2018 This Is What Happens In An Internet Minute
facebook
973,000 Logins
18 Million Text Messages
4.3 Million Videos Viewed
375,000 Apps Downloaded
174,000 Scrolling Instagram
481,000 Tweets Sent
1.1 Million Swipes
187 Million Emails Sent
936,073 Views
67 Voice-First Devices Shipped
38 Million Messages
25,000 GIFs Sent via Messenger
2.4 Million Snaps Created
$862,823 Spent Online
266,000 Hours Watched
3.7 Million Search Queries
60 SECONDS
Created By:
@LoriLewis
@OfficiallyChadd

2019 This Is What Happens In An Internet Minute
facebook
1 Million Logging In
18.1 Million Texts Sent
4.5 Million Videos Viewed
390,030 Apps Downloaded
347,222 Scrolling Instagram
87,500 People Tweeting
1.4 Million Swipes
188 Million Emails Sent
1 Million Views
41 Music Streaming Subscriptions
180 Smart Speakers Shipped
4.8 Million Gifs Served
41.6 Million Messages Sent
2.1 Million Snaps Created
$996,956 Spent Online
694,444 Hours Watched
3.8 Million Search Queries
60 SECONDS
Created By:
@LoriLewis
@OfficiallyChadd

3.8 MILLION
search queries
1 MILLION
logging in
18.1 MILLION
texts sent
694,444
hours watched
4.5 MILLION
videos viewed
2.1 MILLION
snaps created
390,030
apps downloaded
41.6 MILLION
messages sent
60
SECONDS
347,222
scrolling
$996,956
spent online
87,500
people tweeting
4.86 MILLION
impressions served
exoClick
1.4 MILLION
swipes
2 MILLION
RTB queries
exoClick
41
music streaming
subscriptions
1.8 MILLION
emails sent

2020

AIRBNB
GUESTS BOOK
747 STAYS
AMAZON
SHOPPERS SPEND
$455K
X
USERS SEND
360K
TWEETS
6.3M
SEARCHES
HAPPEN ON
GOOGLE
WHATSAPP
USERS SEND
41.6M
MESSAGES
LINKEDIN
USERS SUBMIT
6,060
RESUMES
VIEWERS WATCH
43 YEARS
OF STREAMING
CONTENT
3,720
USERS DOWNLOAD
INSTAGRAM
THREADS
CYBER-
CRIMINALS
LAUNCH 30 DDOS ATTACKS
EVERY MINUTE
01:00
OF THE DAY
PRESENTED BY
DOMO
CHATGPT
USERS SEND
6,944 PROMPTS
INSTAGRAM
USERS SEND
694K REELS VIA DM
FACEBOOK
USERS LIKE
4M POSTS
DOORDASH
DINERS PLACE
$122K
IN ORDERS
FANS STREAM A
TAYLOR SWIFT
SONG 69.4K TIMES
VENMO
USERS SEND
$463K
IN PAYMENTS
GLOBAL
INTERNET
USERS SPEND
25.1M
HOURS ONLINE
THE AVERAGE
PERSON
PRODUCES
102 MB
OF DATA
PEOPLE TRADE
$398M
IN TREASURY
BONDS
PEOPLE SEND
241M
EMAILS
TWITCH
USERS WATCH
48K
HOURS OF CONTENT


EVERY minute OF THE DAY
3.47M YOUTUBE VIDEOS WATCHED
4K PEOPLE READING YELP REVIEWS
174K APPS DOWNLOADED
16.2M TEXTS SENT
231M EMAILS SENT
13.2K LISTENERS ACTIVE ON SPOTIFY
6M PEOPLE SHOPPING ONLINE
6.3M GOOGLE SEARCHES
44M PEOPLE VIEWING FACEBOOK LIVE STREAMS
12.5K RIDE-SHARES TAKEN
20.8K ACTIVE USERS ON LINKEDIN
2.1M PEOPLE ACTIVE ON FACEBOOK
350K TWEETS SENT
115.7K SEARCHES ON PINTEREST
66K PHOTOS AND VIDEOS SHARED ON INSTAGRAM
3.5B SNAPCHATS CREATED
625M VIDEOS WATCHED ON TIKTOK
452K HOURS OF CONTENT STREAMED ON NETFLIX
NETFLIX
LocaliQ
© Gannett Co., Inc. 2023. All Rights Reserved.

2024

INTRODUCTION

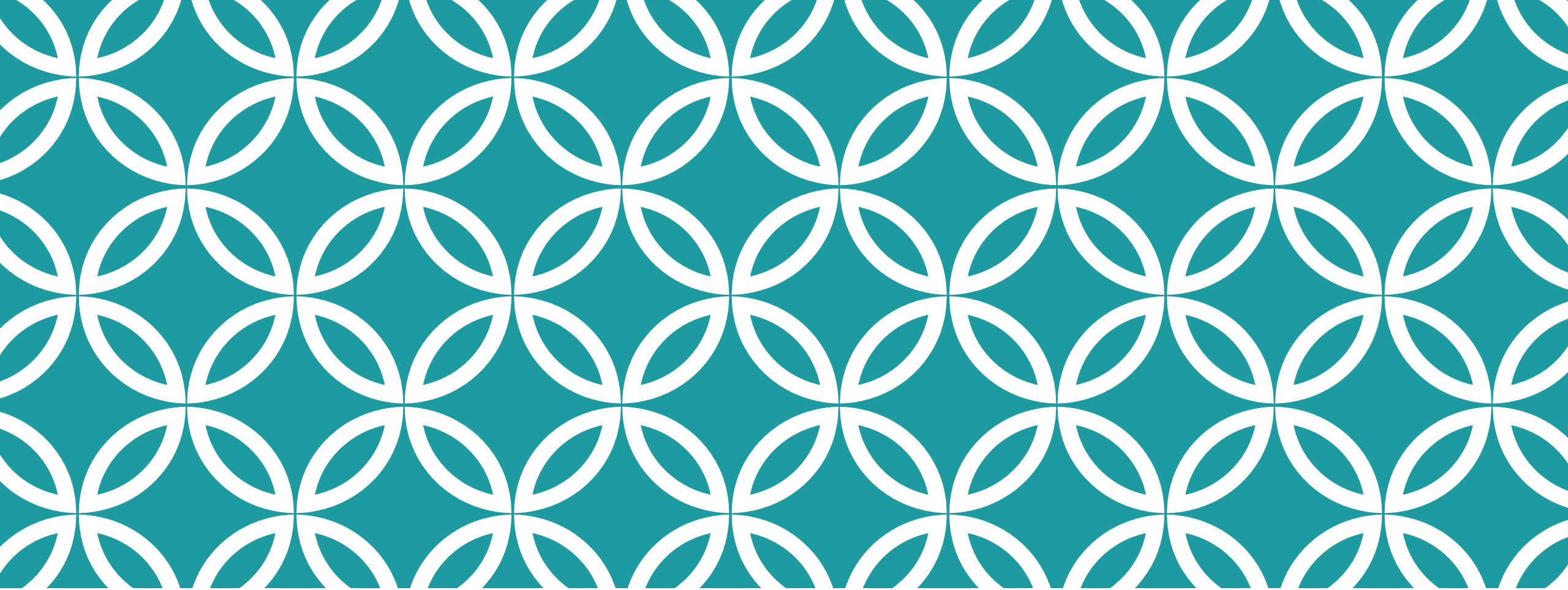
90% of the available data in the world is collected in previous two years and 80% of the data comes from unstructured sources of data like social media.

INTRODUCTION

Big data in retail Sector with multi dimensional.

Fact

Dimensions



Thank you all so much for your attention

QUESTIONS ?