SUDHEERA THARUNIE DE SILVA

WHAT IS BIG DATA

Collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications.

Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

MAGNITUDE OF DATA

By 2020, about 1.7 megabyte of new information will be created every second for every human being on the planet.

By 2020, accumulated digital universe of data will grow from 4.4 zettabyets in 2015 to around 44 zettabytes, or 44 *trillion* gigabytes.

By 2015, we performed 40,000 Google queries every second which makes it 3.5 searches per day and 1.2 trillion searches per year.

1.5 billion smartphone were shipped in 2016 - all packed with sensors capable of collecting all kinds of data, not to mention the data the users create themselves.

Between 2015-2020, it is expected that 50 billion smart connected devices will come online, all developed to collect, analyze and share data.

By 2015, less than 0.5% of all data is ever analysed and used, just imagine the potential !!

MORE FACTS

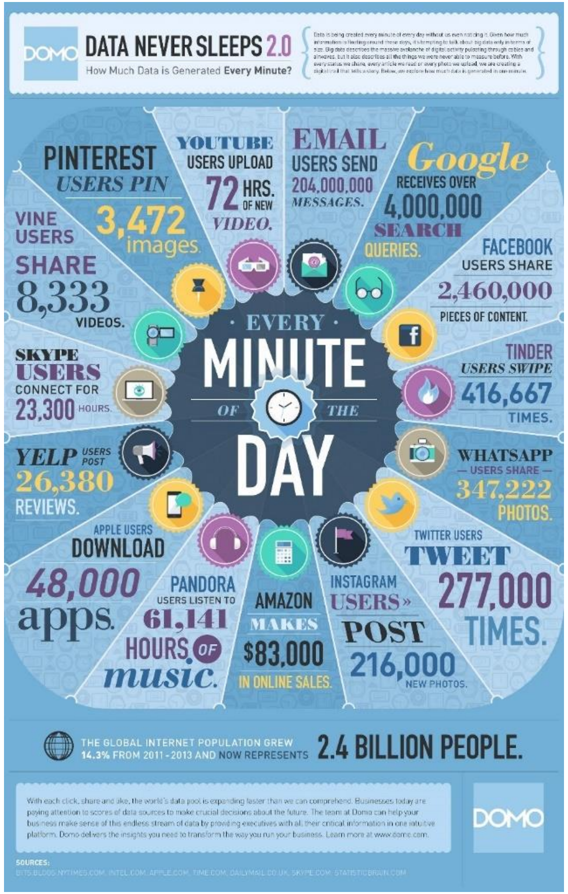
Google processes 24 PB a day (2009)

Wayback Machine has 3 PB + 100 TB/month (2009)

Facebook has 2.5 PB of user data + 15 TB/day (2009)

eBay has 6.5 PB of user data + 50 TB/day (2009)

eBay has 6.5 PB of user data + 50 TB/day (2009)



DATA NEVER SLEEPS 2.0

How Much Data is Generated **Every Minute?**

Data is being created every minute of every day without us even noticing it. Given how much information is floating around these days, it's tempting to talk about big data only in terms of size. Big data describes the massive avalanche of digital activity pulsating through cables and airwaves, but it also describes all the things we were never able to measure before. With every status we share, every article we read or every photo we upload, we are creating a digital trail that tells a story. Below, we explore how much data is generated in one minute.

DOMO
EVERY MINUTE OF THE DAY
PINTEREST USERS PIN 3,472 images.
YOUTUBE USERS UPLOAD 72 HRS. OF NEW VIDEO.
EMAIL USERS SEND 204,000,000 MESSAGES.
Google RECEIVES OVER 4,000,000 SEARCH QUERIES.
VINE USERS SHARE 8,333 VIDEOS.
FACEBOOK USERS SHARE 2,460,000 PIECES OF CONTENT.
SKYPE USERS CONNECT FOR 23,300 HOURS.
TINDER USERS SWIPE 416,667 TIMES.
YELP USERS POST 26,380 REVIEWS.
WHATSAPP — USERS SHARE — 347,222 PHOTOS.
APPLE USERS DOWNLOAD 48,000 apps.
PANDORA USERS LISTEN TO 61,141 HOURS OF music.
AMAZON MAKES $83,000 IN ONLINE SALES.
INSTAGRAM USERS» POST 216,000 NEW PHOTOS.
TWITTER USERS TWEET 277,000 TIMES.

THE GLOBAL INTERNET POPULATION GREW 14.3% FROM 2011 - 2013 AND NOW REPRESENTS 2.4 BILLION PEOPLE.

With each click, share and like, the world's data pool is expanding faster than we can comprehend. Businesses today are paying attention to scores of data sources to make crucial decisions about the future. The team at Domo can help your business make sense of this endless stream of data by providing executives with all their critical information in one intuitive platform. Domo delivers the insights you need to transform the way you run your business. Learn more at www.domo.com.

SOURCES:
BITS.BLOGS.NYTIMES.COM, INTEL.COM, APPLE.COM, TIME.COM, DAILYMAIL.CO.UK, SKYPE.COM, STATISTICBRAIN.COM

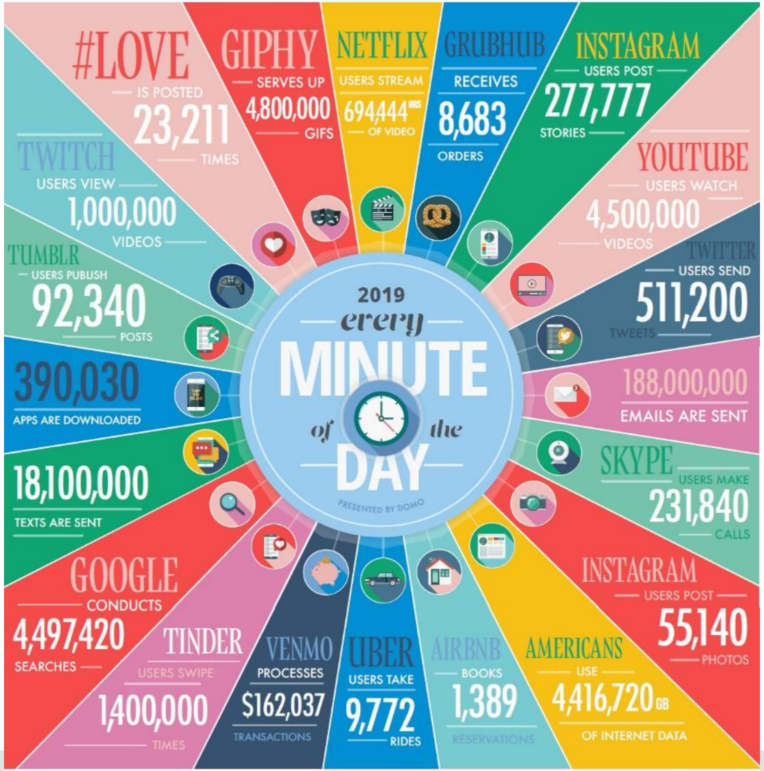
Source: https://aci.info/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic/

DOMO

DATA NEVER SLEEPS 7.0

How much data is generated *every minute?*

There's no way around it: big data just keeps getting bigger. The numbers are staggering, and they're not slowing down. By 2020, there will be 40x more bytes of data than there are stars in the observable universe. In our 7th edition of Data Never Sleeps, we bring you the latest stats on how much data is being created in every digital minute.



2019 *every* MINUTE *of the* DAY

PRESENTED BY DOMO

#LOVE IS POSTED 23,211 TIMES

GIPHY SERVES UP 4,800,000 GIFS

NETFLIX USERS STREAM 694444 HRS OF VIDEO

GRUBHUB RECEIVES 8,683 ORDERS

INSTAGRAM USERS POST 277,777 STORIES

YOUTUBE USERS WATCH 4,500,000 VIDEOS

TWITTER USERS SEND 511,200 TWEETS

188,000,000 EMAILS ARE SENT

SKYPE USERS MAKE 231,840 CALLS

INSTAGRAM USERS POST 55,140 PHOTOS

AMERICANS USE 4,416,720 GB OF INTERNET DATA

AIRBNB BOOKS 1,389 RESERVATIONS

UBER USERS TAKE 9,772 RIDES

VENMO PROCESSES $162,037 TRANSACTIONS

TINDER USERS SWIPE 1,400,000 TIMES

GOOGLE CONDUCTS 4,497,420 SEARCHES

18,100,000 TEXTS ARE SENT

390,030 APPS ARE DOWNLOADED

TUMBLR USERS PUBLISH 92,340 POSTS

TWITCH USERS VIEW 1,000,000 VIDEOS

SOURCES: STATISTA, INTERNET LIVE STATS, EXPANDED RAMBLINGS, NATIONAL ASSOCIATION OF CITY TRANSPORTATION OFFICIALS, WIRED

DOMO

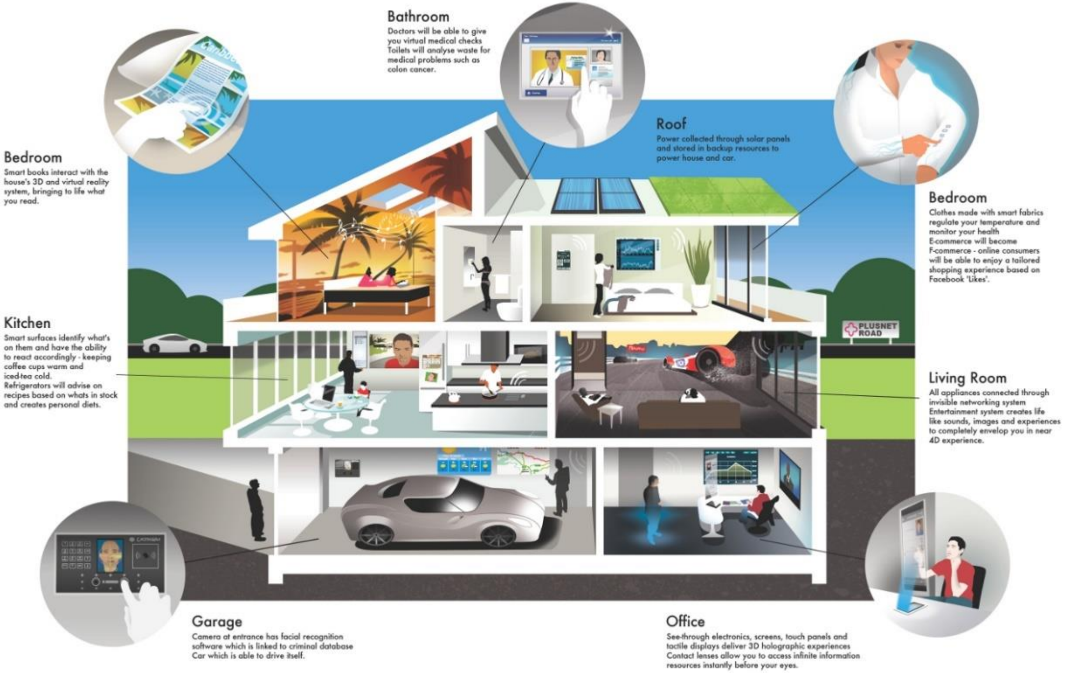
DATA GENERATION

SOCIAL MEDIA

MOBILE DEVICES

SENSOR TECHNOLOGY AND NETWORKS

PROCESSING POWER AND SENSORS HAVE BECOME AFFORDABLE



Bedroom
Smart books interact with the house's 3D and virtual reality system, bringing to life what you read.

Kitchen
Smart surfaces identify what's on them and have the ability to react accordingly - keeping coffee cups warm and iced-tea cold.
Refrigerators will advise on recipes based on whats in stock and creates personal diets.

Bathroom
Doctors will be able to give you virtual medical checks
Toilets will analyse waste for medical problems such as colon cancer.

Roof
Power collected through solar panels and stored in backup resources to power house and car.

Bedroom
Clothes made with smart fabrics regulate your temperature and monitor your health
E-commerce will become F-commerce - online consumers will be able to enjoy a tailored shopping experience based on Facebook 'Likes'.

Living Room
All appliances connected through invisible networking system
Entertainment system creates life like sounds, images and experiences to completely envelop you in near 4D experience.

Garage
Camera at entrance has facial recognition software which is linked to criminal database
Car which is able to drive itself.

Office
See-through electronics, screens, touch panels and tactile displays deliver 3D holographic experiences
Contact lenses allow you to access infinite information resources instantly before your eyes.

NATURE OF DATA

DATA MODEL

THEN

Few companies were generating data and all others were consuming

Traditional media – News papers, TV, Radio channels

Government

NOW

All of us are generating data and all of us are consuming

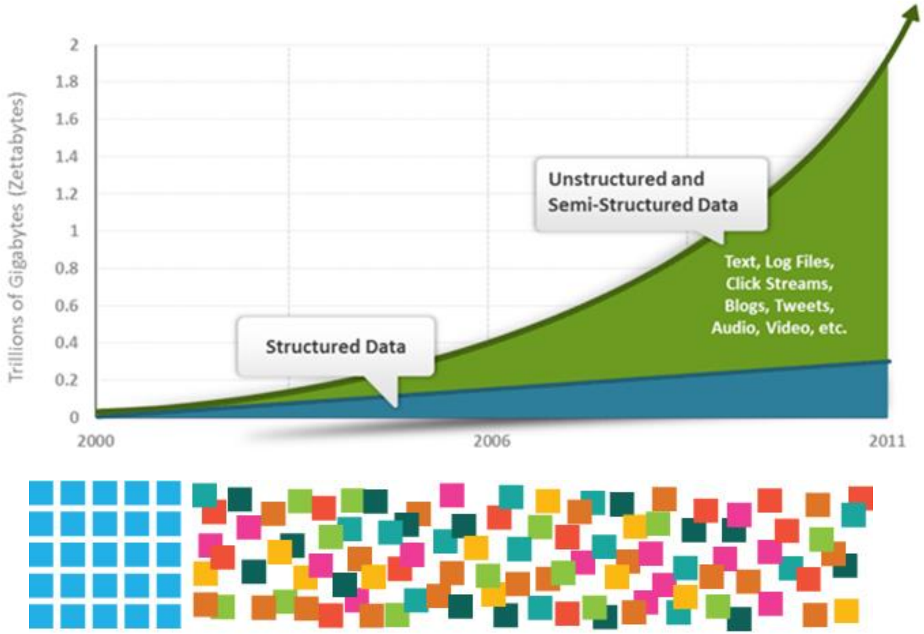
- Social media – FB, Instagram
- Crowd sourcing – Google maps

DATA TYPES

- **Structured Data**
  - Data is organized according to a specific structure or schema.
  - Eg: Data stored in a relational database.
- **Unstructured Data**
  - Data that does not have any consistent organization, associated metadata or schema.
  - Eg: Certain documents, images, audio and video files
- **Semi-Structured Data**
  - Data that has some form of organization.
  - Eg: Server logs, XML / email documents, sensor data.

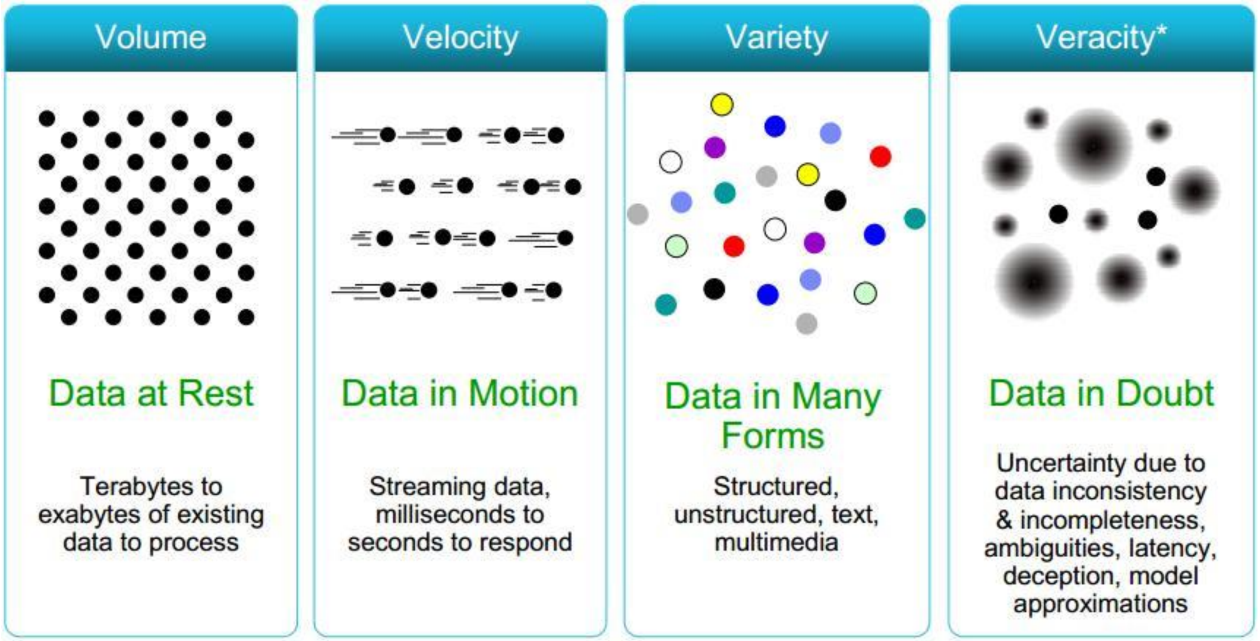
BIG DATA?

- Data volume has grown exponentially over the past decade.



80% of the world's data is **unstructured**

CHARACTERISTICS



Volume

Data at Rest

Terabytes to exabytes of existing data to process

Velocity

Data in Motion

Streaming data, milliseconds to seconds to respond

Variety

Data in Many Forms

Structured, unstructured, text, multimedia

Veracity*

Data in Doubt

Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

The FOUR V's of Big Data

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

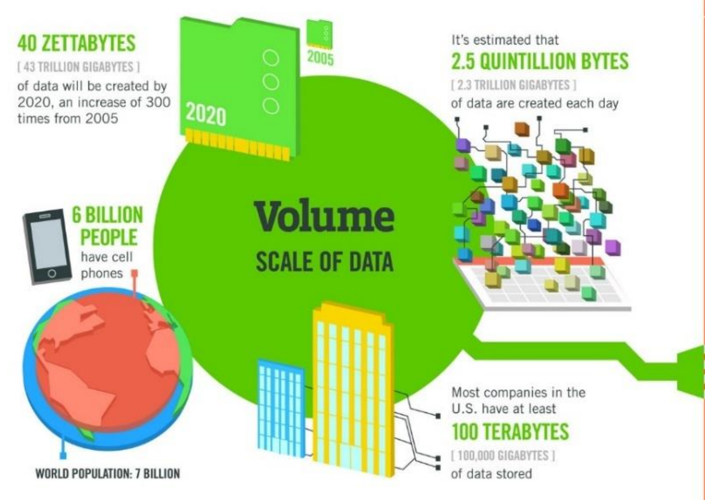
As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize 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 1.9 million in the United States



Volume

SCALE OF DATA

40 ZETTABYTES

[43 TRILLION GIGABYTES]

of data will be created by 2020, an increase of 300 times from 2005

It's estimated that

2.5 QUINTILLION BYTES

[2.3 TRILLION GIGABYTES]

of data are created each day

6 BILLION PEOPLE

have cell phones

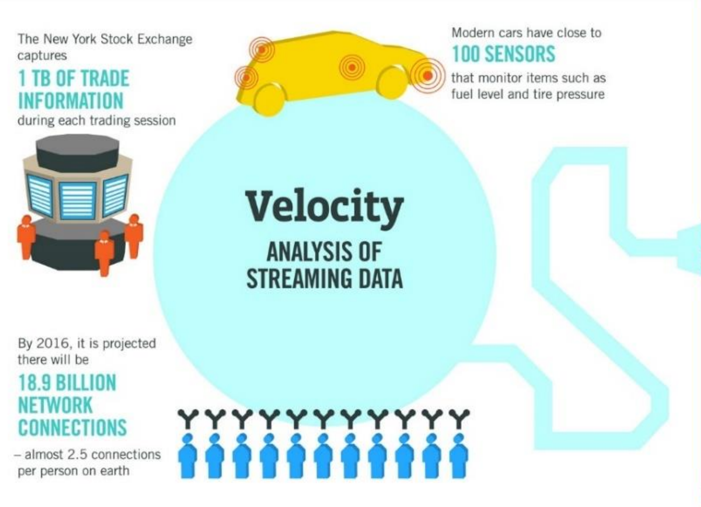
WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least

100 TERABYTES

[100,000 GIGABYTES]

of data stored



Velocity

ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session

Modern cars have close to

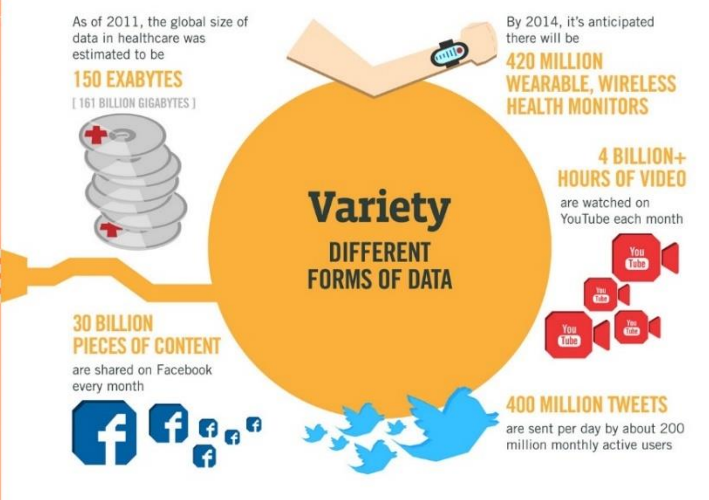
100 SENSORS

that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth



Variety

DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO

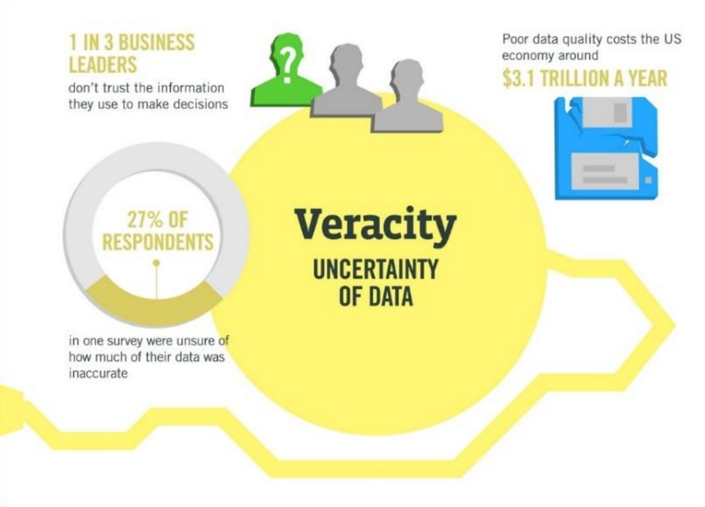
are watched on YouTube each month

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



Veracity

UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions

Poor data quality costs the US economy around

$3.1 TRILLION A YEAR

27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

IBM

BIG DATA ANALYTICS

Big data analytics is the process of examining **large data**sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information.

Source: http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics

ML/AI AND BIG DATA

- Machine Learning and AI is becoming increasingly popular when generating output from Big Data.

- AI/ML technologies are used to find patterns more effectively and ML is used for predictive analytics using big data

BIG DATA ANALYITICAL PLATFORMS

BIG DATA

ALSO
BIG
DATA
TOOLS
DISK
TARGET
APPLIED
SENSOR
MAY
SET
USE
RELATIONAL
SOCIAL
LARGER
CONTINUES
EVERY
EXAMPLES
SHARED
DEFINITION
CURRENT
MOVING
WITHIN
RECONSIDER
PRACTITIONERS
CAPTURE
ZETTABYTES
COMPLEX
TENS
INDEXING
CITATION
INTERNET
DISTRIBUTED
BUSINESS
SETS
GENOMICS
COMPLEXITY
ANALYTICS
DESCRIBING
RADIO-FREQUENCY
MANAGEMENT
TERABYTES
CASE
ABILITY
NOW
BIOLOGICAL
PROCESSING
UBIQUITOUS
CAPACITY
BURIED
WORLD'S
SOLID
TYPES
PETABYTES
INCLUDE
ELAPSED
SYSTEMS
TOLERABLE
HUNDREDS
NETWORKS
DATABASES
AMOUNT
SIZE
DESKTOP
RECORDS
CURRENTLY
GARTNER
WORKING
STORAGE
MPP
OPPORTUNITIES
SAN
QUALITIES
SINCE
PARALLEL
SEARCH
MASSIVELY
GROW
DIFFICULTY

Q & A