

NAME: مصطفى طارق محمد حسانين الخولى

B.N: 913

TOPIC: Big Data

DATE: 31\5\2020

GITHUP LINK: <https://github.com/mostafa-elkholy20/ece006>

GITHUP PAGE: <https://mostafa-elkholy20.github.io/ece006/>

No-one can argue that every corporation is a tech company nowadays. In this globalized age of rapidity; technology is the driver of all the industries and a symbol of progress.

Companies must place technology first, every company must have priority. Key aspect of the company is Information and data. Gathering, processing and understanding how information flows within the business is crucial. Only businesses who really recognize, handle and track data from their industries can be effective. An enterprise is as valuable by its ability to understand and process data to drive business decisions.

Technology without processing information will not yield results; and information without processing ability has no value. It is clear that we need two main work streams, which should work in tandem at all levels, in order to prove the business in future.

One is a data strategy: understanding the market, how data is obtained and processed within the company divisions, knowing the departmental partnerships and creating a network that determines how the business operates depending on how the data flows. From the data perceptive point of view, a company is just a black box that takes inputs and generates outputs, this needs to be refined and broken down until we have a clear understanding of the business data. The second stream is the technology stream: enhancing infrastructure, skills, and tools to process and manage data efficiently.

There's a huge buzz around large data, AI and so on. Understanding the impetus behind these modern concepts is important, and understanding that such innovations are just a normal reaction to a human revolt when we enter the information technology. AI will not succeed by itself, changes in the way the business is run and essential technology changes are required. It is crucial to realize as part of the big data movement that now data is tailored to the data at the heart of the enterprise and the business operations and not the data to the market; this is the central principle behind data lakes. We'll talk about the big data.

To really understand big data, it's helpful to have some historical background. Here is Gartner's definition, circa 2001 (which is still the go-to definition): Big data is data that contains greater variety arriving in increasing volumes and with ever-higher velocity. This is known as the three Vs.

Put simply, big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before.

Big Data Characteristics

The following three are known as “BigData Characteristics”.

- Volume
- Velocity
- Variety

1. Volume:

Volume means “How much Data is generated”. Now-a-days, Organizations or Human Beings or Systems are generating or getting very vast amount of Data say TB(Tera Bytes) to PB(Peta Bytes) to Exa Byte(EB) and more.

VOLUME = Very Large Amount Of Data

2. Velocity:

Velocity means “How fast produce Data”. Now-a-days, Organizations or Human Beings or Systems are generating huge amounts of Data at very fast rate.

VELOCITY = Produce Data at Very Fast Rate

3. Variety:

Variety means “Different forms of Data”. Now-a-days, Organizations or Human Beings or Systems are generating very huge amount of data at very fast rate in different formats. We will discuss in details about different formats of Data soon.

VARIETY = Produce Data in Different Formats

Benefits of Big Data Processing

Ability to process Big Data brings in multiple benefits, such as-

- Businesses can utilize outside intelligence while taking decisions

Access to social data from search engines and sites like facebook, twitter are enabling organizations to fine tune their business strategies.

- Improved customer service

Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.

- Early identification of risk to the product/services, if any
- Better operational efficiency

Big Data technologies can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the data warehouse. In addition, such integration of Big Data technologies and data warehouse helps an organization to offload infrequently accessed data.

SCREEN SHOTS:

Big Data

links

[Big Data Home page](#)

[Big Data Definition](#)

[Big Data Works and types](#)

[Big Data Advantages and Disadvantages](#)

[Big Data Applications](#)

Big Data Definition

There are way too many definitions on big data, almost everyone is talking about big data. The primary idea behind big data is that "(Almost) everything we do nowadays are leaving digital footprints and tracers (i.e. data) which can be used for analysis.

Some examples of big data:

- Analyzing **medical record** to prevent epidemic.
- Analyzing transport and traffic records to optimize **traffic flow** (think Waze).
- Data modelling for **Prediction** and **Personalization**.

The not so obvious big data:

- Analyzing your **writing style**, browsing pattern to identify your personality.
- Study **web log** to analyze your shopping preference.
- Analyzing your **typing speed** to identify your character.



Big Data

links

[Big Data Home page](#)

[Big Data Definition](#)

[Big Data Works and types](#)

[Big Data Advantages and Disadvantages](#)

[Big Data Applications](#)

The Definition of Big Data

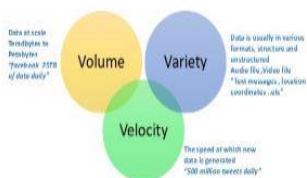
There is little agreement as to how to describe big data as one of the most 'hyped' words on the industry today. The term is also synonymously used with similar terms such as Business Intelligence (BI) and data mining. It is accurate because these three words are for analyzing data and in many cases advanced analytics. Yet the definition of big data is different from the other two because data sizes, number of transactions and the amount of data sources are so big and complicated that specific approaches and techniques are needed to obtain information from data (for example, conventional data warehouse solutions can be limited when working with big data). This also forms the basis for the most used definition of big data, the three V: Volume, Velocity and Variety as shown in.

- Volume: Large amounts of data, from datasets with sizes of terabytes to zettabyte.

- Velocity: Large amounts of data from transactions with high refresh rate resulting in data streams coming at great speed and the time to act on the basis of these data streams will often be very short. There is a shift from batch processing to real time streaming.

- Variety: Data come from different data sources. For the first, data can come from both internal and external data source. More importantly, data can come in various format such as transaction and log data from various applications, structured data as database table, semi-structured data such as XML data, unstructured data such as text, images, video streams, audio statement, and more. There is a shift from sole structured data to increasingly more unstructured data or the combination of the two.

Three Characteristics of Big Data 3Vs



This leads us to the most widely used definition in the industry. Gartner (2012) defines Big Data in the following. Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. It is also coming at you fast, it is coming at you in complex format, and it is coming at you from a variety of sources. Around now it should be obvious that the "big" in big data is not just about quantity. Although big data is definitely about getting a lot of data, big data is not about data volume alone. What it means is not just that you get a ton of info. It is also important to note that the definition of an absolute threshold for what constitutes big data might not have too much value. The big data of today might not be the big data of tomorrow because of the technology progress. It's a relative concept, on the whole. From the viewpoint of others, whether the company facing major problems (and opportunities) about the data's volume, velocity, and variety it is the big data challenge. Such problems usually carry together the need for distinct systems and strategies for data collection and process.

Big Data

links

- [Big Data Home page](#)
- [Big Data Definition](#)
- [Big Data Works and types](#)
- [Big Data Advantages and Disadvantages](#)
- [Big Data Applications](#)

How Big Data Works

Big data gives you new insights that opens up new opportunities and business models. Getting started involves three key actions:

1. **Integrate:** Big data puts data from many various sources and applications. Traditional data integration mechanisms, such as ETL (extract, transform, and load) generally aren't up to the task. The study of large data sets at terabyte, or even petabyte, size involves different techniques and technology. You need to put in the data during integration, process it, and make sure it is structured and accessible in a way that your market analysts can get started.
2. **Manage:** Big data requires storage. You may choose the storage system at the cloud, at the premises, or both. You can store your data in any type you choose and on-demand add bring processing specifications and appropriate process engines to such data sets. Most people pick their option for storage due to where their data is currently residing. The cloud is gradually gaining popularity because it meets the existing computing needs and enables you to spin up resources as needed.
3. **Analyze:** when you analyze and act on your data, the investment in big data pays off. Provide fresh insights for the diverse data sets virtually analyzed. Explore the data further to make new discoveries. Tell others about your findings. Build data models with machine learning and artificial intelligence. Put your data to work.

Types of big data (Big Data could be found in three forms).

1. Structured

Example: An 'Employee' table in a database.

Employee ID	Employee Name	Gender	Department	Salary (\$)
1001	John Doe	Male	Finance	10000
1002	Jane Smith	Female	HR	12000
1003	Michael Chen	Male	Marketing	15000
1004	Sarah Lee	Female	Operations	18000
1005	David Kim	Male	IT	20000

2. Unstructured

Example: The output returned by 'Google Search'.



3. Semi-structured

Example: Personal data stored in an XML file

```
<?xml version='1.0' encoding='UTF-8'?>
<person>
  <name>John Doe</name>
  <age>35</age>
  <gender>Male</gender>
  <department>Finance</department>
  <salary>10000</salary>
  <email>john.doe@company.com</email>
  <phone>123-456-7890</phone>
  <address>
    <street>123 Main St</street>
    <city>New York</city>
    <state>NY</state>
    <zip>10001</zip>
  </address>
  <hobbies>
    <hobby>Reading</hobby>
    <hobby>Golfing</hobby>
  </hobbies>
  <pets>
    <pet name='Fluffy' type='cat' age='5' color='white' breed='Persian' />
    <pet name='Buddy' type='dog' age='3' color='black' breed='Labrador' />
  </pets>
</person>
```

Big Data

links

- [Big Data Home page](#)
- [Big Data Definition](#)
- [Big Data Works and types](#)
- [Big Data Advantages and Disadvantages](#)
- [Big Data Applications](#)

Big Data Advantages and Disadvantages

Advantages

- Big data analysis derives innovative solutions.
- It helps in improving science and research.
- It improves healthcare and public health with availability of record of patients.
- It helps in financial tradings, sports, polling, security/law enforcement etc.
- Any one can access vast information via surveys and deliver answer of any query.
- Every second additions are made.
- One platform carry unlimited information.

Disadvantages

- Traditional storage can cost lot of money to store big data.
- Lots of big data is unstructured.
- Big data analysis violates principles of privacy.
- It can be used for manipulation of customer records.
- It may increase social stratification.
- Big data analysis is not useful in short run.
- Big data analysis results are misleading sometimes.

Big Data

links

[Big Data Home page](#)

[Big Data Definition](#)

[Big Data Works and types](#)

[Big Data Advantages and Disadvantages](#)

[Big Data Applications](#)

Big Data Applications

- **Banking and Securities** : The Securities Exchange Commission (SEC) is using Big Data to monitor financial market activity. They are currently using network analytics and natural language processors to catch illegal trading activity in the financial markets. Retail traders, Big banks, hedge funds, and other so-called 'big boys' in the financial markets use Big Data for trade analytics used in high-frequency trading, pre-trade decision-support analytics, sentiment measurement, Predictive Analytics, etc.

- **Communications, Media and Entertainment**: Simultaneously, companies in this field evaluate consumer data coupled with behavioral data to construct detailed customer profiles that can be used for:

- Create content for the different target audiences
- Suggest on-demand content
- Test efficiency of the content

- **Healthcare Providers**: Some hospitals use data from millions of patients that have been collected from a cell phone app to allow doctors to use evidence-based medicine as opposed to administering multiple medical / lab tests to all patients going to the hospital. A research battery may be efficient but it can be inefficient and typically unreliable as well. The University of Florida has used free public health data and Google Maps to create visual data that allows for faster identification and efficient analysis of health care information, used to track chronic disease spread.

- **Education**: The use of big data in higher education is quite significant. The University of Tasmania for example. An Australian university of more than 26,000 students has implemented a learning and management program that monitors, among other aspects, as a student signs in to the System, how much time is spent in various sites in the system, as well as a student's average success over time. It is also used in a different use case of the use of Big Data in education to measure the effectiveness of teachers to ensure a pleasant experience for both students and teachers. The success of teachers may be fine-tuned and evaluated against numbers of students, subject matter, profiles of students, expectations of students, description of behaviors, and numerous other variables.

- **Manufacturing and Natural Resources**: Big Data allows for predictive modeling in the natural resource industry to support decision-making that has been used to ingest and integrate large amounts of data from geospatial data, graphic data, text, and temporal data. Areas of interest where this was used include; seismic interpretation and characterization of the reservoir. Big data has already been used to address today's industrial problems and, gain a competitive advantage, among other benefits.

- **Government**: Big Data has a wide array of applications in public services, including energy exploration, financial market analysis, fraud detection, health-related research, and environmental protection. Some more specific examples are as follows: Large amounts of social disability claims made to the Social Security Administration (SSA) that arrive in the form of unstructured data are being analyzed using big data. The analytics are used to quickly and efficiently process medical information for faster decision-making, and to detect suspicious or fraudulent claims. Homeland Security Department uses Big Data for various use cases. Big data is analyzed from various government agencies and is used to protect the country.

C:\Users\mte\Desktop\New folder\index.html > ...

```
1 </html>
2 <head>
3   <title>Big Data</title>
4 </head>
5 <body>
6
7 </body>
8 <h1><FONT SIZE="+4">Big Data</FONT></h1>
9 <h2>links</h2>
10 <ul><a href=index.html class="href"> <FONT SIZE="+3">Big Data Home page</FONT></a></ul>
11 <ul><a href="Big Data Definition.html" class="herf"><FONT SIZE="+3">Big Data Definition</FONT></a></ul>
12 <ul><a href="Big Data Works and types.html" class="herf"><FONT SIZE="+3">Big Data Works and types</FONT></a></ul>
13 <ul><a href="Big Data Advantages and Disadvantages.html" class="herf"><FONT SIZE="+3">Big Data Advantages and Disadvantages</FONT></a></ul>
14 <ul><a href="Big Data Applications.html" class="herf"><FONT SIZE="+3">Big Data Applications</FONT></a></ul>
15
16 <img src=https://www.drhanlau.com/wp-content/uploads/2016/08/big-data-definition-cher-han.jpg alt="img 2">
17
18 <img src=https://aadamov.files.wordpress.com/2015/04/3v-big-data-1.jpg alt="img 2">
19
20
21
22 </body>
23
24
25
26
27
28
29 </html>
```

```
1 </html>
2 <head>
3   <title>Big Data</title>
4 </head>
5 <body>
6
7 </body>
8 <h1><FONT SIZE="+4">Big Data</FONT></h1>
9 <h2>links</h2>
10 <ul><a href=index.html class="href"> Big Data Home page</a></ul>
11 <ul><a href="Big Data Definition.html" class="herf">Big Data Definition</a></ul>
12 <ul><a href="Big Data Works and types.html" class="herf">Big Data Works and types</a></ul>
13 <ul><a href="Big Data Advantages and Disadvantages.html" class="herf">Big Data Advantages and Disadvantages</a></ul>
14 <ul><a href="Big Data Applications.html" class="herf">Big Data Applications</a></ul>
15 <h2><FONT SIZE="+3">The Definition of Big Data</FONT></h2>
16 <h3><FONT SIZE="+2">There is little agreement as to how to describe big data as one of the most 'hyped' words on the industry today.
17   The term is also synonymously used with similar terms such as Business Intelligence ( BI) and data mining.
18   It is accurate because these three words are for analyzing data and in many cases advanced analytics.
19   Yet the definition of big data is different from the other two because data sizes, number of transactions and the amount of data sources are so big and comp
20   This also forms the basis for the most used definition of big data, the three V: Volume, Velocity and Variety as shown in.
21 </FONT></h3>
22 <h4><FONT SIZE="+2">- Volume: Large amounts of data, from datasets with sizes of terabytes to zettabyte.
23 </FONT></h4>
24 <h5><FONT SIZE="+2">- Velocity: Large amounts of data from transactions with high refresh rate resulting in data streams coming at great speed and the time to act
25   There is a shift from batch processing to real time streaming.</FONT></h5>
26
```


c:\Users\mte\Desktop\New folder\Big Data Definition.html body

```
27 <h6><FONT SIZE="+2">- Variety: Data come from different data sources.
28     For the first, data can come from both internal and external data source.
29     More importantly, data can come in various format such as transaction and log data from various applications, structured data as database table, semi-structured data as log files, etc.
30     There is a shift from sole structured data to increasingly more unstructured data or the combination of the two.
31     
32 </FONT></h6>
33
34 <h8> <FONT SIZE="+2">
35     This leads us to the most widely used definition in the industry. Gartner (2012) defines Big Data in the following.
36     Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable the capture, management, analysis and visualization of data.
37     It is also coming at you fast, it is coming at you in complex format, and it is coming at you from a variety of sources.
38     Around now it should be obvious that the "big" in big data is not just about quantity. Although big data is definitely about getting a lot of data, big data is also about getting it fast, it is coming at you in complex format, and it is coming at you from a variety of sources.
39     It is also important to note that the definition of an absolute threshold for what constitutes big data might not have too much value. The big data of today might be the structured data of tomorrow.
40     Such problems usually carry together the need for distinct systems and strategies for data collection and process.
41 </FONT></h8>
42
43
44 </body>
45
46
47 </html>
```

C:\Users\mte\Desktop\New folder\Big Data Works and types.html head title

```
1 </html>
2 <head>
3     <title>Big Data</title>
4 </head>
5 <body>
6 </body>
7 <h1><FONT SIZE="+4">Big Data</FONT></h1>
8 <h2>links</h2>
9 <ul><a href=index.html class="href"> Big Data Home page</a></ul>
10 <ul><a href="Big Data Definition.html" class="herf">Big Data Definition</a></ul>
11 <ul><a href="Big Data Works and types.html" class="herf">Big Data Works and types</a></ul>
12 <ul><a href="Big Data Advantages and Disadvantages.html" class="herf">Big Data Advantages and Disadvantages</a></ul>
13 <ul><a href="Big Data Applications.html" class="herf">Big Data Applications</a></ul>
14 <h2><FONT SIZE="+3">How Big Data Works</FONT></h2>
15 <h3><FONT SIZE="+2">Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions:
16 </h3></FONT>
17 <h4><FONT SIZE="+2">1. Integrate:
18     Big data puts data from many various sources and applications. Traditional data integration mechanisms, such as ETL (extract, transform, and load) generally require a lot of time and effort.
19     You need to put in the data during integration, process it, and make sure it is structured and accessible in a way that your market analysts can get started.
20 </FONT></h4>
21 <h5><FONT SIZE="+2">2. Manage:
22     Big data requires storage. You may choose the storage system at the cloud, at the premises, or both.
23     You can store your data in any type you choose and on-demand add bring processing specifications and appropriate process engines to such data sets. Most people use a combination of the two.
24 </FONT></h5>
25 <h6><FONT SIZE="+2">3. Analyze:
26     When you analyze and act on your data, the investment in big data pays off. Provide fresh insight for the diverse data sets visually analyzed.
27     Explore the data further to make new discoveries. Tell others about your findings. Build data models with machine learning and artificial intelligence. Put your data to work.
28 </FONT></h6>
29
```

```

<h7><FONT SIZE="+3">Types of big data
</h7>

<h8><FONT SIZE="+2">(Big Data could be found in three forms):</FONT>
<ul>1. Structured
  <h1><FONT SIZE="+2">Example: An 'Employee' table in a database.
  </FONT></h1>
  <img src=data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAQABAAQ/2wCEAAkGBxAQEAE8QDxEWDXUPFQ8PDw8QFhUPEA8VFRYWfUWFRUVISggGBomHRUVITEhJSkrLjAwFx8zODMsNyoutLi0BCgoKDQ0OG
</ul>
<ul>2. Unstructured
  <h1><FONT SIZE="+2">Example: The output returned by 'Google Search'.</FONT></h1>
  <img src=https://www.guru99.com/images/Big_Data/061114_0759_WhatIsBigDa5.png alt="img 3">
</ul>
<ul>3. Semi-structured
  <h1><FONT SIZE="+2">Example: Personal data stored in an XML file</FONT></h1>
  <img src=https://www.ibm.com/support/knowledgecenter/SSBRAM_8.8.0/com.ibm.classify.workbench.doc/graphics/Catalog.gif alt="img 3">
</ul>
</ul>
</h8>
</body>
</html>

```

```

c: > Users > mte > Desktop > New folder > <> Big Data Advantages and Disadvantages.html > head > title
1  </html>
2  <head>
3  <title>Big Data</title>
4  </head>
5  <body>
6  </body>
7  <h1><FONT SIZE="+4">Big Data</FONT></h1>
8  <h2>links</h2>
9  <ul><a href=index.html class="href"> Big Data Home page</a></ul>
10 <ul><a href="Big Data Definition.html" class="herf">Big Data Definition</a></ul>
11 <ul><a href="Big Data Works and types.html" class="herf">Big Data Works and types</a></ul>
12 <ul><a href="Big Data Advantages and Disadvantages.html" class="herf">Big Data Advantages and Disadvantages</a></ul>
13 <ul><a href="Big Data Applications.html" class="herf">Big Data Applications</a></ul>
14 <h2><FONT SIZE="+3">Big Data Advantages and Disadvantages</FONT></h2>
15 <table style="width:100%">
16 <tr>
17 <th><h3><FONT SIZE="+2">Advantages</FONT></h3> </th>
18 <th><h3><FONT SIZE="+2">Disadvantages</FONT></h3> </th>
19 </tr>
20 <tr>
21 <td><h4>• Big data analysis derives innovative solutions.</h4> </td>
22 <td><h4>• Traditional storage can cost lot of money to store big data.</h4> </td>
23 </tr>
24 <tr>
25 <td><h4>• It helps in improving science and research.</h4> </td>
26 <td><h4>• Lots of big data is unstructured.</h4> </td>
27 </tr>
28 <tr>
29 <td><h4>• It improves healthcare and public health with availability of record of patients.</h4> </td>
30 <td><h4>• Big data analysis violates principles of privacy.</h4></td>
31 </tr>
32 <tr>

```


C:\Users\nte\Desktop\New folder\Big Data Applications.html > head > title

```
1 </html>
2 <head>
3   <title>Big Data</title>
4 </head>
5 <body>
6
7 </body>
8 <h1><FONT SIZE="+4">Big Data</FONT></h1>
9 <h2>links</h2>
10 <ul><a href=index.html class="href"> Big Data Home page</a></ul>
11 <ul><a href="Big Data Definition.html" class="herf">Big Data Definition</a></ul>
12 <ul><a href="Big Data Works and types.html" class="herf">Big Data Works and types</a></ul>
13 <ul><a href="Big Data Advantages and Disadvantages.html" class="herf">Big Data Advantages and Disadvantages</a></ul>
14 <ul><a href="Big Data Applications.html" class="herf">Big Data Applications</a></ul>
15 <H2><FONT SIZE="+3">Big Data Applications</FONT></H2>
16 <H3><FONT SIZE="+2"><ul>- Banking and Securities : The Securities Exchange Commission (SEC) is using Big Data to monitor financial market activity. They are currently using Big Data to monitor the trading patterns of
17   Retail traders, Big banks, hedge funds, and other so-called 'big boys' in the financial markets use Big Data for trade analytics used in high-frequency trading
18   </ul>
19 <ul> - Communications, Media and Entertainment: Simultaneously, companies in this field evaluate consumer data coupled with behavioral data to construct detailed
20   <ul>• Create content for the different target audiences</ul>
21   <ul>• Suggest on-demand content</ul>
22   <ul>• Test efficiency of the content </ul>
23   </ul>
24 <ul> - Healthcare Providers: Some hospitals use data from millions of patients that have been collected from a cell phone app to allow doctors to use evidence-based medicine.
25   The University of Florida has used free public health data and Google Maps to create visual data that allows for faster identification and efficient analysis of
26   </ul>
27 <ul>- Education: The use of big data in higher education is quite significant. The University of Tasmania for example. An Australian university of more than 26,000 students.
28   It is also used in a different use case of the use of Big Data in education to measure the effectiveness of teachers to ensure a pleasant experience for both students and teachers.
29   </ul>
30 <ul>- Manufacturing and Natural Resources: Big Data allows for predictive modeling in the natural resource industry to support decision-making that has been used in the past.
31   Big data has already been used to address today's industrial problems and, gain a competitive advantage, among other benefits.
32   </ul>
```