Faculty of engineering - Shoubra
Benha University

Literature Review

in fulfillment of the requirements of

Department	**Engineering Mathematics and Physics**
Division	**----------**
Academic Year	**2019-2020 Preparatory**
Course name	**Computer**
Course code	**ECE001**

Title: -

Big Data

By:

	Name	Edu mail	B.N
1	منار محمود السيد احمد الصباغ	manar196064@feng.bu.edu.eg	952

Approved by:

1	Dr.Ahmed Bayoumi
2	Dr.Shady Elmashad
3	Dr. Abdelhamid Attaby

Project link : https://github.com/manarmahmod/html-project

Website link : https://manarmahmod.github.io/html-project/index.html

Abstract

Data is The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in volume and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.

Table of contents

Introduction

Big data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years. The term has been in use since the 1990s, with some giving credit to John Mashey for popularizing the term. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data philosophy encompasses unstructured, semi-structured and structured data, however the main focus is on unstructured data. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many zettabytes of data. Big data requires a set of techniques and technologies with new forms of integration to reveal insights from data-sets that are diverse, complex, and of a massive scale.

Literature Review

To preview the results data about big data I have created a website to show it.
This is screenshots from the website pages :

Big Data

home

Characteristics

applications

usage

problems

The term has been in use since the 1990s, with some giving credit to John Mashey for popularizing the term. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data philosophy encompasses unstructured, semi-structured and structured data, however the main focus is on unstructured data. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many zettabytes of data. Big data requires a set of techniques and technologies with new forms of integration to reveal insights from data-sets that are diverse, complex, and of a massive scale.

Big Data

home

Characteristics

applications

usage

problems

Big data can be described by the following characteristics:

1. Volume

The quantity of generated and stored data. The size of the data determines the value and potential insight, and whether it can be considered big data or not.

2. Variety

The type and nature of the data. This helps people who analyze it to effectively use the resulting insight. Big data draws from text, images, audio, video; plus it completes missing pieces through data fusion.

3. Velocity

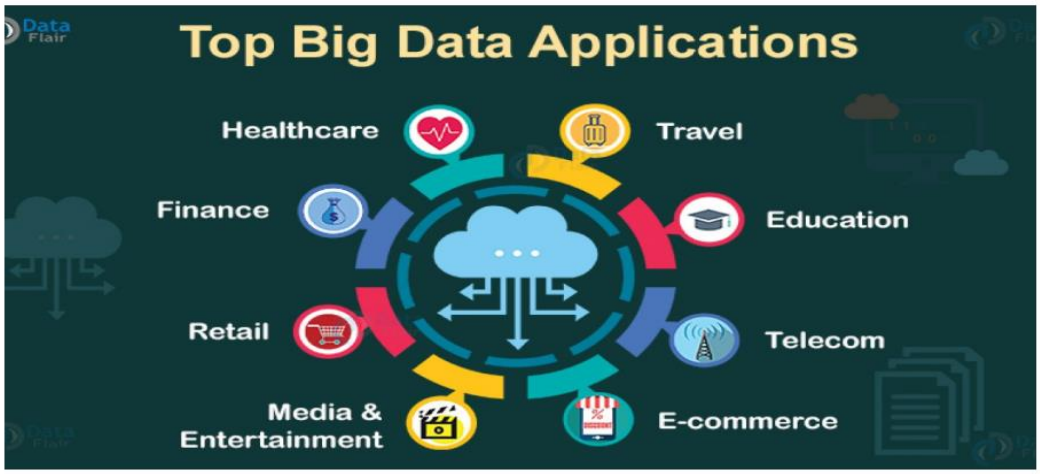
The speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Big data is often available in real-time. Compared to small data, big data are produced more continually. Two kinds of velocity related to big data are the frequency of generation and the frequency of handling, recording, and publishing.

4. Veracity

It is the extended definition for big data, which refers to the data quality and the data value. The data quality of captured data can vary greatly, affecting the accurate analysis.

Big Data

home
Characteristics
applications
usage
problems



Big Data

home
Characteristics
applications
usage
problems

there's common probelms facing big data such as:

1. Lack of Understanding
2. High Cost of Data Solutions
3. Too Many Choices
4. Complex Systems for Managing Data
5. Security Gaps
6. Low Quality and Inaccurate Data
7. Lack of Skilled Workers
8. Data Integration
9. Processing Large Data Sets

Snapshots from source code :

```
<html>
<title> Big Data</title>
<body>
<div style="background:#dcf2ec;">
<br>
<h1 style = "text-align: center">Big Data</h1>
<br>
</div>
<div style="float:left; width:200px; background:#9bbfb5; height: auto; text-align:center;">
<br>
<h2><a href="index.html">home</a></h2>
<br>
<h2><a href="Characteristics.html">Characteristics</a></h2>
<br>
<h2><a href="applications.html">applications</a></h2>
<br>
<h2><a href="usage.html"> usage</a></h2>
<br>
<h2><a href="problems.html">problems</a></h2>
<br>
</div>
<div style="text-align:left; height:auto;">
<img src="bigdata.jpg" style = "float:right; margin-left:30px; margin-right:20px;">
<p style="font-size:130%; margin-top: 3em; margin-bottom: 1em; margin-left: 12em; ">
The term has been in use since the 1990s, with some giving credit to John Mashey for popularizing the term.
Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture,
curate, manage, and process data within a tolerable elapsed time. Big data philosophy encompasses unstructured,
semi-structured and structured data, however the main focus is on unstructured data.
Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many zettabytes of data.
Big data requires a set of techniques and technologies with new forms of integration to reveal insights from data-sets that are diverse,
complex, and of a massive scale.
</p>
</div>
</body>
</html>
```

```
<div style="background:#dcf2ec;"> <br>
<h1 style = "text-align: center">Big Data</h1> <br>
</div>
<div style="float:left; width:200px; background:#9bbfb5; height:270%; text-align:center;"> <br>
<h2><a href="index.html">home</a></h2>
<br>
<h2><a href="Characteristics.html">Characteristics</a></h2>
<br>
<h2><a href="applications.html">applications</a></h2>
<br>
<h2><a href="usage.html"> usage</a></h2>
<br>
<h2><a href="problems.html">problems</a></h2>
<br>
</div>
<div style="text-align:left; height: 100%;">
<div style="font-size:130%; margin-top: 3em; margin-bottom: 1em; margin-left: 12em; margin-right:5em">

<h2>Big data can be described by the following characteristics:</h2>
<ol>
<li style="color:red;font-size:30px">Volume</li>
<p>The quantity of generated and stored data. The size of the data determines the value and potential insight,
and whether it can be considered big data or not.</p>

<li style="color:red;font-size:30px">Variety</li>
<p>The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.
Big data draws from text, images, audio, video; plus it completes missing pieces through data fusion.</p>
<br>
<li style="color:red;font-size:30px">Velocity</li>
<p>The speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.
Big data is often available in real-time. Compared to small data, big data are produced more continually.
Two kinds of velocity related to big data are the frequency of generation and the frequency of handling, recording, and publishing.</p>
<br>
<li style="color:red;font-size:30px">Veracity</li>
<p>It is the extended definition for big data, which refers to the data quality and the data value.
```

Results and discussion

Data sets are considered “big data” if they have a high degree of the following three distinct dimensions: volume, velocity, and variety. Value and veracity are two other “V” dimensions that have been added to the big data literature in the recent years. Additional Vs are frequently proposed, but these five Vs are widely accepted by the community and can be described as follows:

- **Velocity**: the speed at which the data is been generated
- **Volume**: the amount of the data that is been generated
- **Variety**: the diversity or different types of the data
- **Value**: the worth of the data or the value it has
- **Veracity**: the quality, accuracy, or trustworthiness of the data

Large volumes of data are generally available in either structured or unstructured formats. Structured data can be generated by machines or humans, has a specific schema or model, and is usually stored in databases. Structured data is organized around schemas with clearly defined data types. Numbers, date time, and strings are a few examples of structured data that may be stored in database columns. Alternatively, unstructured data does not have a predefined schema or model. Text files, log files, social media posts, mobile data, and media are all examples of unstructured data.

Conclusions

To summarize, we are generating a massive amount of data in our everyday life, and that number is continuing to rise. Having the data alone does not improve an organization without analyzing and discovering its value for business intelligence. It is not possible to mine and process this mountain of data with traditional tools, so we use big data pipelines to help us ingest process, analyze, and visualize these tremendous amounts of data.

References

(1) https://www.ntnu.no/iie/fag/big/lessons/lesson2.pdf

(2) https://arxiv.org/ftp/arxiv/papers/1705/1705.04928.pdf

(3) https://www.guru99.com/what-is-big-data.html