**Name: Mohamed Ebrahim Mohamed Shehab**

**B.N: 636**

**Date: 4/6/2020**

**Topic:** Big Data

**GitHub link:**

**GitHub page:**

**Application brief:**

Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc. Now day organizations have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format.

1 \ Volume:

It is the volume of data extracted from a source, which determines the value and capabilities of the data to be classified as big data. And it may be the most important characteristic of big data analysis.

Also, describing it as huge does not specify a specific quantity. It is usually measured in bytes or bytes,

2 \ Variety:

It refers to the diversity of data extracted, which helps users, whether they are researchers or analysts, to choose the appropriate data for their field of research and includes structured data in databases and unstructured data that comes from its unstructured nature, such as: pictures, clips, sound recordings, videos, SMS and call logs And GPS data

3 \ Velocity:

It means the speed of production and data extraction to cover the demand for it. As speed is a critical element in making decisions based on these data, which is the time that we take from the moment these data arrive to the moment when the decision is made based on them.

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.

Big data can be collected from publicly shared comments on social networks and websites, voluntarily gathered from personal electronics and apps, through questionnaires, product purchases, and electronic check-ins. The presence of sensors and other inputs in smart devices allows for data to be gathered across a broad spectrum of situations and circumstances.

Big data is most often stored in computer databases and is analyzed using software specifically designed to handle large, complex data sets. Many software-as-a-service (SaaS) companies specialize in managing this type of complex data.

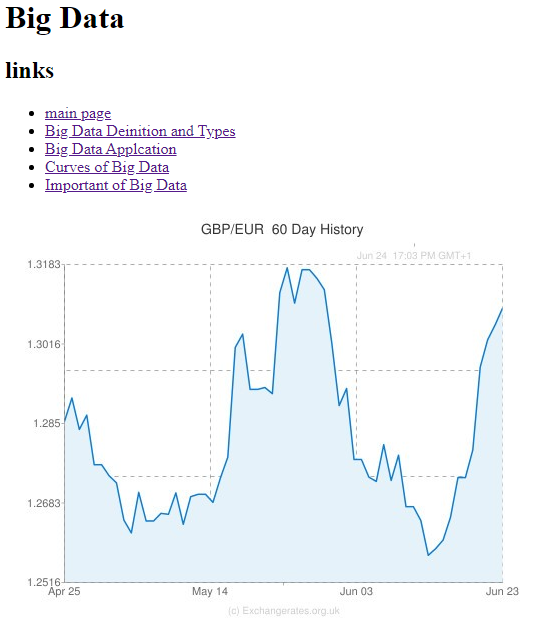
The increase in the amount of data available presents both opportunities and problems.

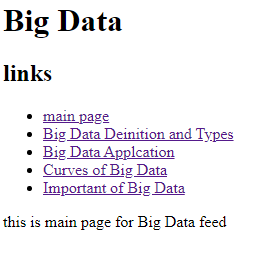
In general, having more data on one’s customers (and potential customers) should allow companies to better tailor their products and marketing efforts in order to create the highest level of satisfaction and repeat business. Companies that are able to collect a large amount of data are provided with the opportunity to conduct deeper and richer analysis.

While better analysis is a positive, big data can also create overload and noise. Companies have to be able to handle larger volumes of data, all the while determining which data represents signals compared to noise. Determining what makes the data relevant becomes a key factor.

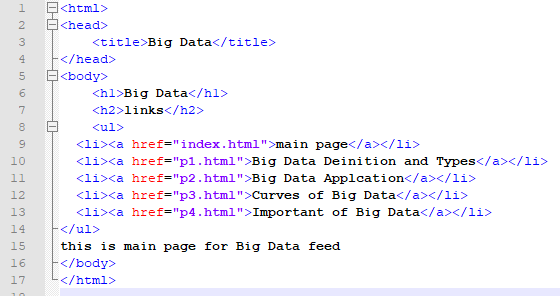
Furthermore, the nature and format of the data can require special handling before it is acted upon. Structured data, consisting of numeric values, can be easily stored and sorted. Unstructured data, such as emails, videos, and text documents, may require more sophisticated techniques to be applied before it becomes useful.

**Screenshots:**

****

****

**Source code:**

****