

3 pillars of big data :

Structured Data

Semi Structured Data

Unstructured Data

By **Prowebscraper**

What is big data ?

- Big data stands for large volumes of data.
- It could be both structured and unstructured.
- It's not about the amount of data but what businesses do with the given data that counts.
- Businesses leverage on big data for insights that can propel their growth.

Example of Big Data

Following are some of the examples of 'Big Data' :


- Facebook users send roughly 31.25 million messages and watch 2.77 million videos, every minute!!
- Walmart customers' transactions provide the company with about 2.5 petabytes of data, every hour.

Example of Big Data

- YouTube usage jumped three times from 2014-2016 with users uploading 400 hours of new video each minute of every day! Now, in 2017, users are watching 4,146,600 videos every minute.
- Instagram users upload 46,740 million posts every minute!
- 5.2 BILLION daily Google Searches in 2017.



As you can see, there are 3 pillars of Big Data :

1. Structured data
 2. Unstructured data
 3. Semi structured data
- 

1. Structured data

- Structured Data refer to the data which is already stored in databases, in an ordered manner.
- It accounts for about 20% of the total existing data, and is used the most in programming and computer-related activities.

Example of Structured data :

- Meta-data (Time and date of creation, File size, Author etc.)
- Library Catalogues (date, author, place, subject, etc)
- Census records (birth, income, employment, place etc.)
- Economic data (GDP, PPI, ASX etc.)

2. Unstructured data :

- Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, unstructured data poses multiple challenges in terms of its processing for deriving value out of it.
- The rest of the data created, about 80% of the total account for unstructured big data.

Example of Unstructured data :

- Media (MP3, digital photos, audio and video files)
- Text files (Word processing, spreadsheets, presentations etc.)
- Social Media (Data from Facebook, Twitter, LinkedIn)

3. Semi-Structured data :

- Semi-structured data is a form of structured data that does not conform with the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self-describing structure.

Example of Semi-Structured data :

- Personal data stored in a XML file-

```
<rec><name>Harry</name><sex>Male</sex><age>35</age></rec>
```

```
<rec><name>Justin</name><sex>Female</sex><age>41</age></rec>
```

```
<rec><name>Shawn</name><sex>Male</sex><age>29</age></rec>
```

```
<rec><name>Ed sheeran</name><sex>Male</sex><age>26</age></rec>
```

```
<rec><name>Drake</name><sex>Male</sex><age>35</age></rec>
```

- As you can notice, Internet is a maze of unbounded data.
- But it can be understood, interpreted and used under these three categories.
- You can benefit from each type of data if you understand their characteristics and strengths.

- Businesses worldwide construct their empire on these three pillars and capitalize on their limitless potential.

The question is,
do you???

References :

<http://www.prowebscraper.com/blog/structured-vs-unstructured-data-best-thing-you-need-to-know/>