

A large yellow hexagon that serves as a background for the main title text.

# BIG DATA

## INTRODUCTION

A smaller blue hexagon positioned at the bottom right, overlapping the yellow one.

Data  
Sheet

# Big Data Introduction

## What is Data?

The numbers, characters, or symbols on which a machine performs operations that can be stored and transmitted as electrical signals and captured on magnetic, optical, or mechanical recording devices.

## What is Big Data?

Big data is just data but it has an immense scale. Big Data is a term used to describe a compilation of data which is small in volume but grows exponentially over time. In short, such data is so massive and complex that it cannot be stored or processed effectively by any of the conventional data management tools.

## Structured

Any data that can be stored, accessed, and processed in fixed format is considered 'structured' data. Over time, computer science talent has become more active in developing techniques for working with and deriving value from such data (where the format is well understood in advance). Nowadays, however, we are foreseeing problems when a scale of such data grows exponentially, standard sizes are in the range of multiple zettabytes.

## Unstructured

Any data with a format or structure unknown is labelled as unstructured data. As well as the scale being immense, un-structured data faces several challenges in terms of its processing to extract value from it. A typical example of unstructured data is a heterogeneous source of data which contains a combination of simple text files, images, videos etc. Nowadays, companies have plenty of data available with them, but sadly they don't know how to extract value from it as it is in its raw form or unstructured format.

## Semi-structured

Semi-structured data may contain data in both types. We may see semi-structured data as structured in form but it is not specifically defined in relational DBMS with e.g. a table definition. Semi-structured data for example is a data described in an XML format.

## Characteristics of Big Data

- **Volume** - The term Big Data itself is linked to an immense scale. Data size plays a very important role in assessing meaning from the results.
- **Variety** - Variety applies to both structured and unstructured, heterogeneous sources and the nature of data.
- **Velocity** - The word 'velocity' refers to the rate at which data is produced.
- **Variability** - This applies to the uncertainty that can often be seen from the data.

## Benefits of Big Data Processing

- Companies should make use of outside knowledge when making decisions
- Effective Customer Support
- Risk detection for the product / services early, if any
- Better quality of service