# I Am Data!

Edition II

## The World without Language and Organizations without Data is Impossible

**Mustafa Qizilbash**

*https://www.linkedin.com/in/mustafaisonline*

*If you don't want to learn Data but want to know What Data is, then Read ME!*

PERDANA LEADERSHIP FOUNDATION
YAYASAN KEPIMPINAN PERDANA

# IMAGES LOCATION

Please SCAN below QR Code to refer all the relevant images for each topic. Each image is named based on topic name.





ISBN: 979-8-3748-3788-9

# DEDICATION

"I Am Data! Edition II" is dedicated to all those audiences, colleagues, friends, and family members who motivated me to publish the second edition. This book addresses three types of audiences: the educational sector, where the aim is to help students understand the data world with simple examples and definitions; fresh graduates and individuals who are starting their careers and trying to find the right domain as a career path; and senior management who are struggling to keep up with the jargon used by salespeople. They can refer to this book, where each topic is no longer than 2-3 pages, before attending meetings.

Respected Dr. Mahathir Mohamad,
With my deepest respect and
gratitude for your
enduring leadership & vision -

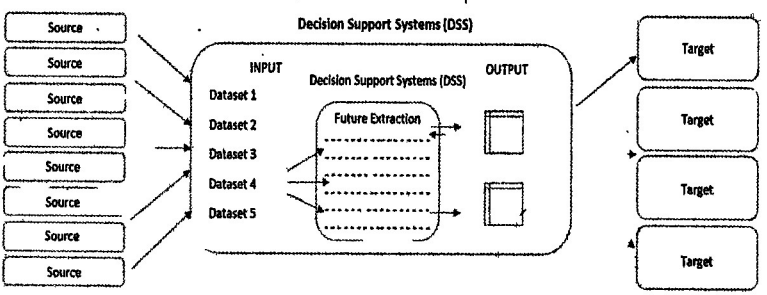
MUSTAFA OZKUBASH


PERDANA
LEADERSHIP
FOUNDATION
YAYASAN
KEPIMPINAN
PERDANA

# Table of Contents

# 1. DECISION SUPPORT SYSTEM [DSS] AND KPI(S)

A question can come in mind why DSS and KPI is the first topic in this book, why not What is Data? The response is, Data is nothing, if it's not useful, if it's not required by any DSS, if it's not fulfilling any KPI. The whole Data domain is dependent on the intention of DSS and KPI productivity.




The complete abbreviation of DSS is Decision Support System. To understand DSS, one must understand KPI i.e., Key Performance Indicator which is a quantifiable measure to evaluate the success of an organization, employee, etc.

In the Corporate World, Decision Support System (DSS) is dependent on Key Performance Indicators (KPI).

Decision Support System (DSS) with Key Performance Indicators (KPI) orchestrates Business Approach. Organizations without defined KPI(s) are without direction and bound to fail. For example, corporate KPI(s) can show 200% annual profit growth, reduced OPEX cost by 5% vs Annual Revenue, Budget Utilization vs ROI etc., but doing so is a question mark on its existence.

The first and foremost KPI for any employee is to align with the organization's KPI(s). At the same time, companies, and departments without employee evaluation KPI(s) will always have high turnover. For example, an employee's annual bonus can be dependent on a few standard KPI(s) like one must do a certain number of certifications or achieve certain deliverables of a Project or Product or increase sales target by 20%-30% etc.

Key Performance Indicators (KPI) are also the means to challenge your Business Planning & Strategy e.g.

- If a new product doesn't show quarterly revenue growth vs certain Profit % growth, it will be discontinued.
- If the new Data Lake or Cloud implementation doesn't increase Revenue by a certain %, it will be only used as cold storage or only online data processing.

In this era, data has become a major player in the organization's Decision Support System. Due to a drastic increase in data volume, every now and then we see new storage frameworks.

Organizations that don't stay up to date with new technologies are threatened with examples like Nokia, Kodak etc.

DSS has been there since inception. Whether human or animal all use DSS in their daily life, the only difference humans have made with time, they have improved their decision making. In the second half of the 20th century, we realized the importance of information

required for DSS and we decided to store it, electrically, for future reference in the shape of a database.

---

*The Williams-Kilburn Tube, invented in 1947, featured the first fully electronic form of data storage. The device was 16 ½ inches long, 6 inches wide, and stored data by displaying a grid of dots on cathode ray tubes and sending a static charge through the tubes.*

---

Since the 90's it has gained popularity and we started to use more and more datastores like Data Warehouse, Big Data, Data Lake, Delta Lake, Data Lakehouse, Hadoop, Object Storage, NoSQL databases etc.

Now businesses are looking towards Real-Time DSS, which means the moment a customer swaps his/ her card and before entering the password, may with OTP code, there comes a promotion on the mobile that same product is available in the same area with 10% more discount.

OTP means **One Time Password**: It's a temporary secure PIN-code sent to you via SMS or e-mail that is valid only for one session. If you cannot receive and confirm the OTP code you will not be able to continue with the transaction.

So, before we jump into the journey of exploring data terms in the rest of the book, we must establish the fact that all Databases, Data Warehouses, Big Data, Data Lake, Delta Lake, Data Lakehouse, Hadoop, Object Storage, NoSQL databases, Business Intelligence Solution etc., in fact every system is there JUST AND JUST to support management decision making which is why it's called DSS (Decision Support Systems).

# 2. WHAT IS DATA?

**D**ata is THE most powerful word nowadays. But the question is, 'What is the exact definition of Data?'

*By Mustafa Qizilbash*

***'Properties and Behavior of an Entity is Data.'***

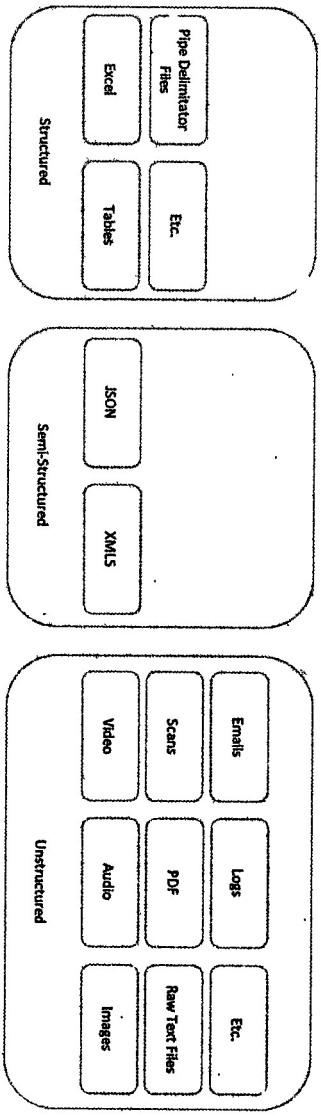
One of my students once asked. What is Data?

He expected to respond with the Wikipedia definition about data which is: Data are individual facts, statistics, or items of information, often numeric. In a more technical sense, data is a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum is a single value of a single variable.

Then he turned towards me and repeated his question but now with more curiosity. I replied to him with my definition of data i.e., '*Properties* and *Behavior* of an *Entity* is *Data*'. His next fair question was then what Entity, Property or Behavior are?

So here was my response, to understand data, one must first understand Entity, Attribute/ Property and Behavior.

- **Entity:** *Something*, which has *Attributes.*
- **Attributes (Properties)**: *Various identification properties of an Entity are Attributes.* Example, Car has a name, it has an engine, seats, tires, lights etc. These are the details about cars which hardly change like the name of the car hardly changes.
- **Behavior:** *What you do, is your behavior or actions*. Example, speed of a car can be measured in km/hour or miles/hour, cars can be sold, cars can be bought etc. These are the Attributes of a Car that can't change, like if Car stops moving, it's a useless product. Cars must get sold, else what's the point of making it! All this information which is generated by Car is its behavior which is ultimately used for planning and decision making.




Data is something that also generates data, it is something which must have some value to the organization and have ROI (Return on Investment). In other words, if a data set

is not assisting Decision Support Systems (DSS) then it is not worth keeping, storing, managing, and processing.

***For Corporate, 'Data is MUST tag with ROI.'***

**Robert M. Solow's (Nobel Economics 1987)**
"Software productivity doesn't show up in the numbers."

**Peter Drucker at one point observed**
"To treat systems & accounting/finance as two separate academic & professional disciplines is unacceptable."

**HT Johnson (a student of AD Chandler) & R Kaplan's**
"Relevance Lost"

**Alan S. Michael**
"MUST tag with ROI of at least one line of business". "What is a line of business?" the answer is "an industry in which the company competes". "What is an industry" the answer (unfortunately) has many answers today. Some industry classification systems suggest the global economy has about 150 industries, some say a few hundred, the U.S. government suggests just over 1,000 - and my firm (based on Michael Porter's five forces model) believe there are more than 23,000 industries. In short, a company = one or more lines of business; and each line of business has data + an ROI model. Industry classification systems with different "industry" definitions - https://en.wikipedia.org/wiki/Industry_classification

**David**
That (allegedly) laudable objective has so far proved elusive to the closed (Newtonian) world view of finance.

The general concept about data is the DIKW model i.e., Data → Information → Knowledge → Wisdom. Though there is nothing wrong with this model but first I don't agree that Knowledge comes

after Information, secondly it's limiting the Data Processing Cycle. The data processing cycles are never-ending i.e., one cycle's data output is an input for another cycle till it's ready for decision. Then once a decision is taken, it becomes the input for the next decisioning cycle, and these cycles keeps rotating. Data Science makes use of these Data Cycles for Deep Learning, Machine Learning, Data Mining etc.

Data can be anything like your personal or family details, banking transactions, mobile calls, education, experience, health, exercise, food, training etc. all these data elements can be used by someone to make some decisions.

Data resides in different forms such as metadata, master data, behavioral data, etc. It can be in three forms i.e., data-at-rest, data-in-trans and data-in-use. It can be for Management, Operational teams, Auditing, Compliance, Regulatory bodies etc.

Any organization which does not make use of its data is bound to fail. In this era, data has become the most powerful asset of your organization, and if one doesn't have a Data Strategy in place then better pack your bags and don't waste your time.

The CEO(s) of the current era are heavily dependent on data like they want to see what a particular customer has done in the last 5 minutes. DSS has become more and more dependent on Data.

**Unique Properties of Data**

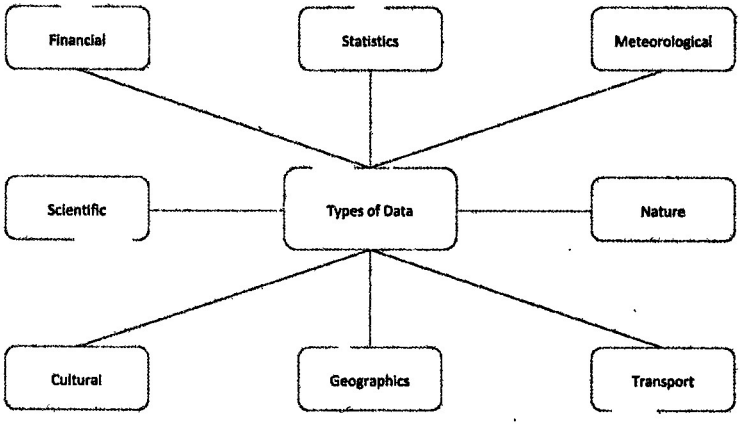
Traditional assets, when used, are consumes, devalued, depreciated, reduces quality and quantity.

But Data, when used:

- Generates more data i.e., how much it's used, who used, when used, where used etc.
- Its value increased i.e., the more a data set is used, shows it's used more frequently for decision making.
- It has no depreciation date; max it can become warm and then cold, but you can need it any time in future. Deleting or purging data is big decision by any stakeholder or even by Regulatory bodies.

# 3. TYPES OF DATA

There are three types of data i.e., structured, semi and unstructured which can come from any type of source e.g., databases, mobile applications, social media, sensors, logs, CCTV cameras, Radio etc. Most of us are pretty much clear about structure and unstructured where semi-structured is still a debatable candidate.




## Structured data

The most common definition of structured data is the data that is sitting in a predefined format of Rows and Columns.

For example, there is a table: employee-contact-details and fields are employee-code, email, mobile, LinkedIn URL, Instagram, and Twitter. Not every employee will have all these details but when it's exported, all columns will be sent for DSS with all the rows whether the row is empty or not. As there are 6 columns with e.g., 10 rows, there will be 10 rows and each row will have 6 columns either empty or with data.

In todays' world, the examples of structured data are tables, excel, CSV and text files etc.

Structured data is something.

- which doesn't randomly change its format.
- which you already know before you enter.
- which once entered can easily be auto extracted.

*'Which does not change its format'* is the most important attribute. As mentioned above, we are told that ONLY data in rows and columns are structured as ['it doesn't randomly change its format'] but what about a picture of a face of someone? Can we extract structured information from it? Yes, we can as e.g., in a 4X4 face picture we know which areas are eyes or ears, nose, forehead etc. Then we can write a code that can differentiate face parts from unlimited pictures. But at the same time, all the pictures must be face pictures which can be distinguished by metadata.

**Semi-structured data**

Now, there are only 2 types of semi-structured data formats i.e., XML & JSON. As a semi-structured data set, the number of columns is not consistent.

Referring to the same example, there is a table: employee-contact-details and fields are employee-code, email, mobile, LinkedIn URL,

Instagram, and twitter. Not every employee will have all these details but this time when it's sent for DSS, only those rows + columns will be exported which have data. For example, there are 6 columns with e.g., 10 rows, now if column 2 is empty in row 8 then when row 8 will be exported, it will have 5 columns instead of 6. This is one way of calling it semi-structured. Another reality is, in Facebook or Twitter, no one knows what kind of data a user will enter right so this also defines it as Semi-structured data.

Semi-structured data is something.

- which doesn't randomly change its format.
- which you don't already know before you enter.

**Unstructured data**

In general, the definition of unstructured data is information that either does not have a pre-defined data model or is not organized in a predefined manner. Then what about the example given above? Yes, face parts can be structured information *only* if all the pictures are about faces but it becomes unstructured if the picture contains multiple faces, male & female, human & animal, young or elder, etc. so in other words metadata can play a great role in defining the structure, semi & unstructured data.

Unstructured data is something.

- which is not defined,
- which you never know up ahead,
- which once stored still need assumptions to understand,
- which one must keep analyzing forever,
- which you can keep score forever with the use of ML algorithms.

Author with 21+ years of experience, spread across countries like Singapore, New Zealand, Malaysia, Myanmar, and Pakistan, shares his knowledge around Data.

Edition I: Jan. 22 [Topics = 72] | Edition II: Feb. 23 [Topics = 201]

**This book is an effort to address three types of Target Audiences**

1) Students: To understand Data field with simple examples and definitions
2) Fresh Graduates: Who has just started their careers and still struggling to find right domain as their career path
3) Senior Management: Who are struggling to catchup with jargons used by salespeople before they jump into meetings, now they can refer this book where each topic is no more than 2-3 pages

"An inspiring comment from one of the Buyer of 1st Edition, 'If you haven't read, I am Data! you are missing a lot.'"

ISBN 979-8-3748-3788-9


PERDANA LEADERSHIP FOUNDATION
YAYASAN KEPIMPINAN PERDANA