

Prof R Madana Mohana



BIG DATA ANALYTICS

Introduction to Big Data

<https://www.youtube.com/c/RASINENIMADANAMOHANA>

HDFS: Outline

Introduction to Big Data

What is Big Data?

Characteristics or V's of Big Data

Examples of the V's in action

Introduction to Big Data

What is Big Data? [According to Bernard Marr]

- Bernard Marr defines **Big Data** as the digital trace that we are generating in this digital era.
- This digital trace is made up of all the data that is captured when we use digital technology.

Introduction to Big Data

What is Big Data? [According to Bernard Marr]

The **basic idea** behind the phrase **Big Data** is that everything we do is **increasingly** leaving a **digital trace (or data)**, which we can **use** and **analyze** to become smarter. The **driving forces** in this brave new world are access to ever increasing **volumes of data** and our ever increasing **technological capability** to **mine** that **data** for **commercial insights**.

Introduction to Big Data

What is Big Data? [The research from **Gartner**]

Big Data is **high-volume**, **high-velocity** and/or **high-variety** information assets that demand *cost effective*, innovative forms of information processing that enable *enhanced insight*, *decision making*, and *process automation*.

Introduction to Big Data

What is Big Data? [Ernst and Young]

Big Data refers to the **dynamic, large and disparate** volumes of data being created by **people, tools and machines**. It requires **new, innovative, and scalable** technology to **collect, host and analytically** process the **vast amount of data** gathered in order to derive **real-time** business insights that relate to **consumers, risk, profit, performance, productivity management and enhanced** shareholder value.

Introduction to Big Data

What is Big Data? [Lisa Arthur, a Forbes contributor]

Big Data is a collection of data from **traditional** and **digital sources** inside and outside a **company** that represent a source of ongoing **discovery** and **analysis**.

Introduction to Big Data

What is Big Data?

- **Big Data** is data whose **scale, distribution, diversity** and/or **timeliness** require the use of **new technical architectures** and **analytics** to enable insights that unlock **new sources of business value**.
 - Requires new data architectures
 - New tools
 - New analytical methods
 - Integrating multiple skills into new role of data scientist.

Introduction to Big Data

What is Big Data?

- Organizations are deriving **business** benefits from **analyzing** ever **large** and more **complex data sets** that increasingly require **real-time capabilities**.

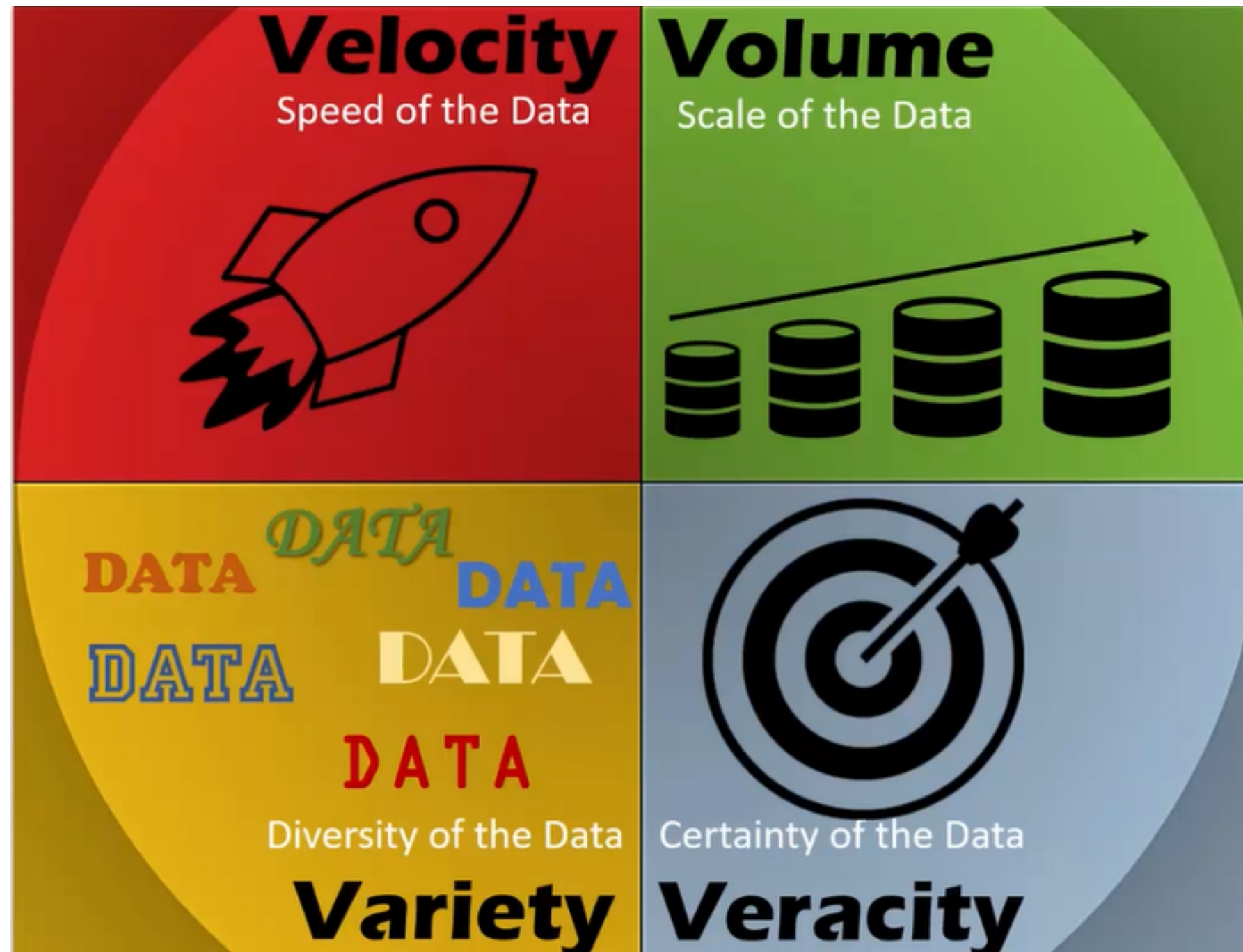
Introduction to Big Data

Characteristics or V's of Big Data:

- There is no one definition of **Big Data**, but there are certain elements that are common across the different definitions, such as *velocity*, *volume*, *variety*, and *veracity*.
- These are the **V's** of **Big Data** or **characteristics** of **Big Data**.

Introduction to Big Data

Characteristics or V's of Big Data:



Introduction to Big Data

Characteristics or V's of Big Data:

- **Velocity** is the speed of the data, or the speed at which data accumulates.
- **Volume** is the scale of the data, or the increase in the amount of data stored.

Introduction to Big Data

Characteristics or V's of Big Data:

- **Variety** is the diversity of the data. We have **structured data** that fits neatly into **rows** and **columns**, or **relational databases** and **unstructured data** that is not organized in a pre-defined way, for example **Tweets, blogposts, pictures, numbers**, and even **video data**.

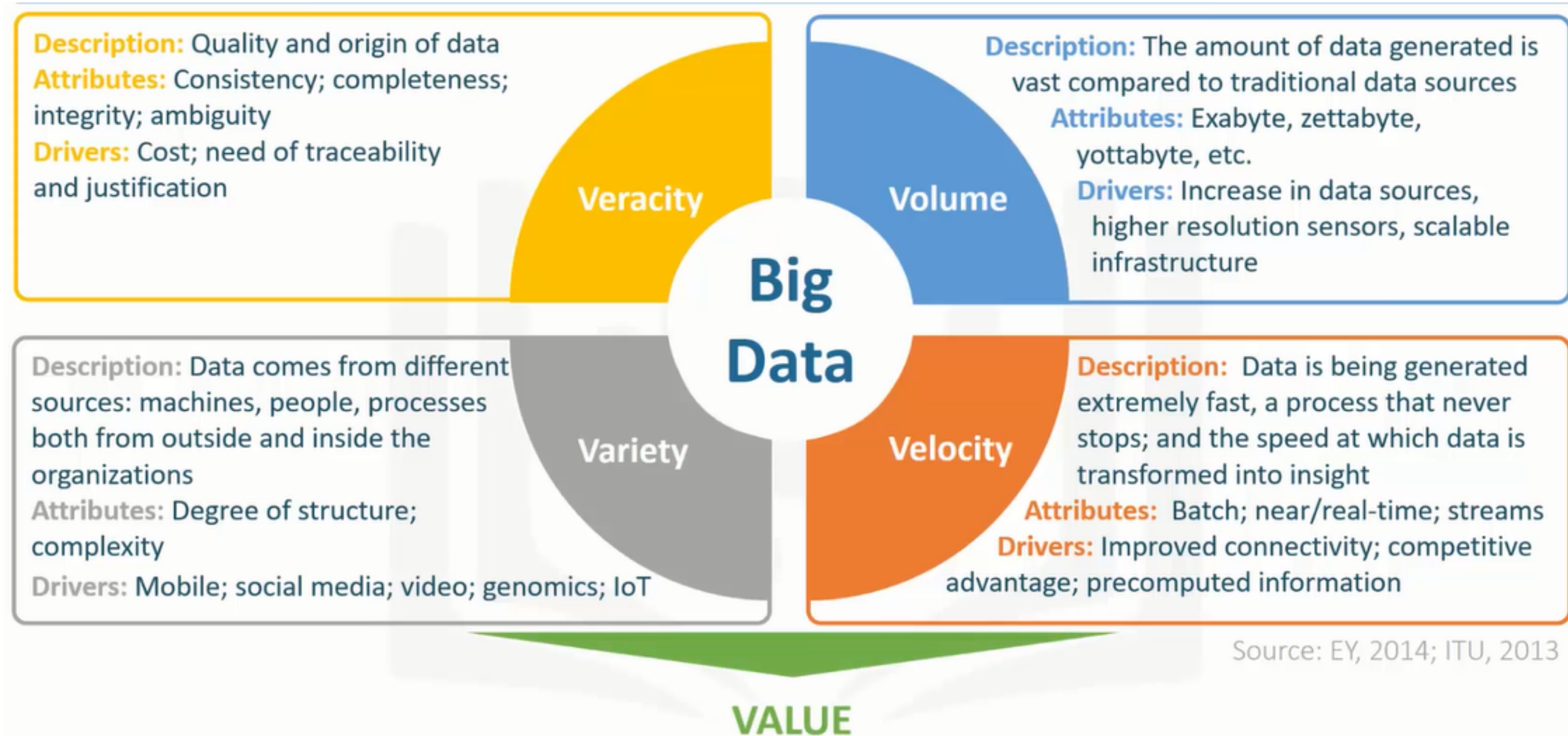
Introduction to Big Data

Characteristics or V's of Big Data:

- **Veracity** is the conformity to **facts** and **accuracy**.
With the large amount of data available, the debate desires on about the **accuracy of data** in the **digital era**. Is the information real, or is it false?

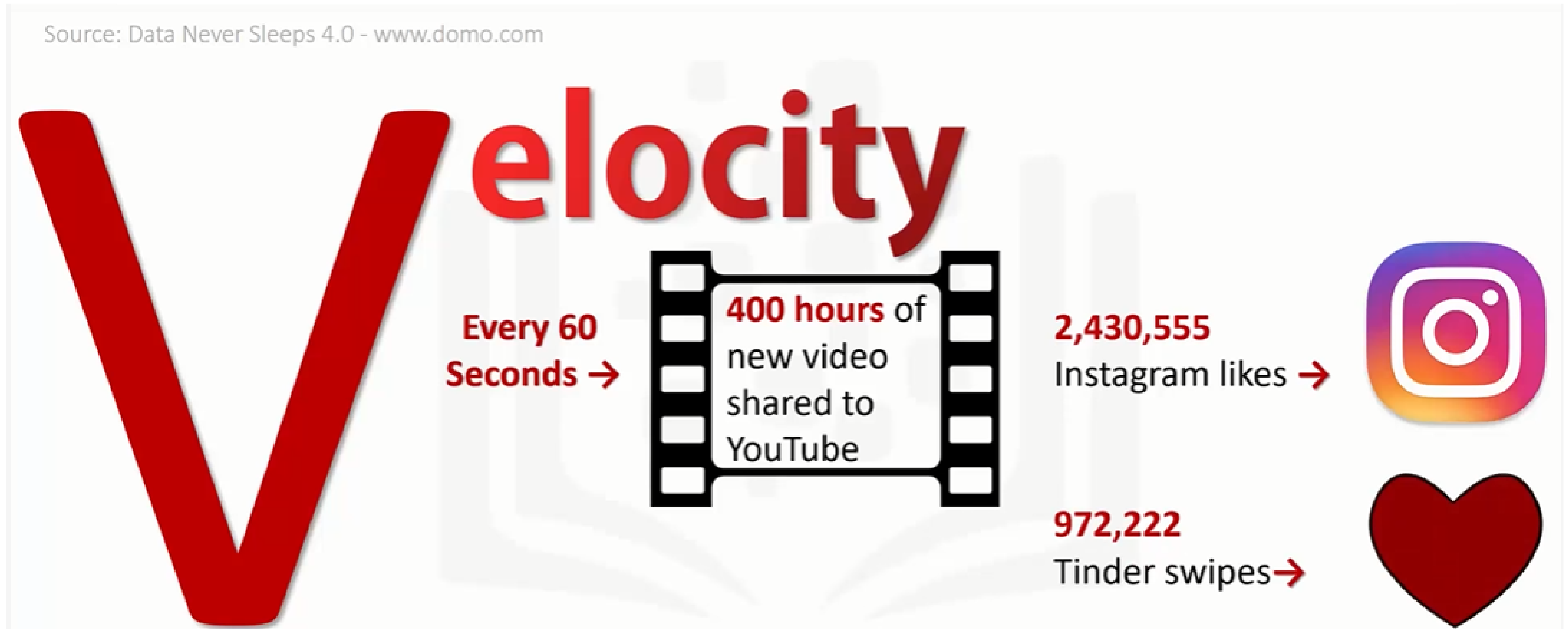
Introduction to Big Data

5 V's of Big Data:



Introduction to Big Data

Examples of the V's in action: Velocity



Introduction to Big Data

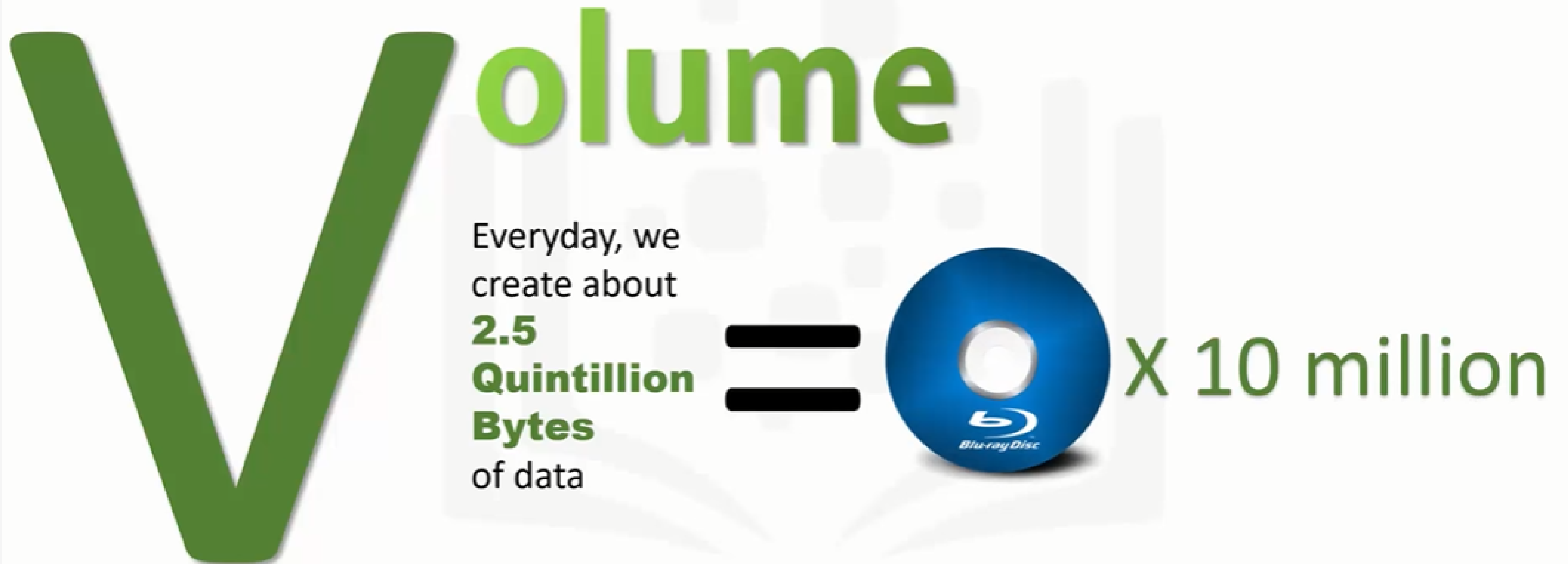
Examples of the V's in action: Velocity

- Every **60** seconds, **400** hours of footage are uploaded to **YouTube**. This amount of data is generated **every minute**.
- So think about how much accumulates over **hours, days, and in years**.

Introduction to Big Data

Examples of the V's in action: Volume

Source: "Bringing big data to the enterprise" - <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> October 2016



Introduction to Big Data

Examples of the V's in action: Volume

- Every day we create approximately **2.5 quintillion** bytes of data. That's **10 million** Blu-ray DVD's every day.
- The **world population** is approximately **seven billion people**, and the vast majority of people are now using digital devices. These devices all generate, capture, and store data. And with more than one device, for example, mobile devices, desktop computers, laptops, etc, we're seeing even more data being produced.

Introduction to Big Data

Examples of the V's in action: **Variety**

Source: "The biggest data challenges that you might not even know you have", Christie Schneider, Content Manager, IBM Watson - <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/>



Introduction to Big Data

Examples of the V's in action: **Variety**

- Let's think about the different types of data: **text**, **pictures**, and **film**.
- What about **sound**, **health data** from **wearable devices**, and many different types of data from **devices** connected to the **Internet of Things (IoT)**.

Introduction to Big Data

Examples of the V's in action: Veracity

Source: "IBM Software Defined Environment", By Dino Quintero, William M Genovese, KiWaon Kim, Ming Jun MJ Li, Fabio Martins, Ashish Nainwal, Dusan Smolej, Marcin Tabinowski, Ashu Tiwary, IBM Redbooks

Veracity

1 in 3 Business leaders don't trust their data



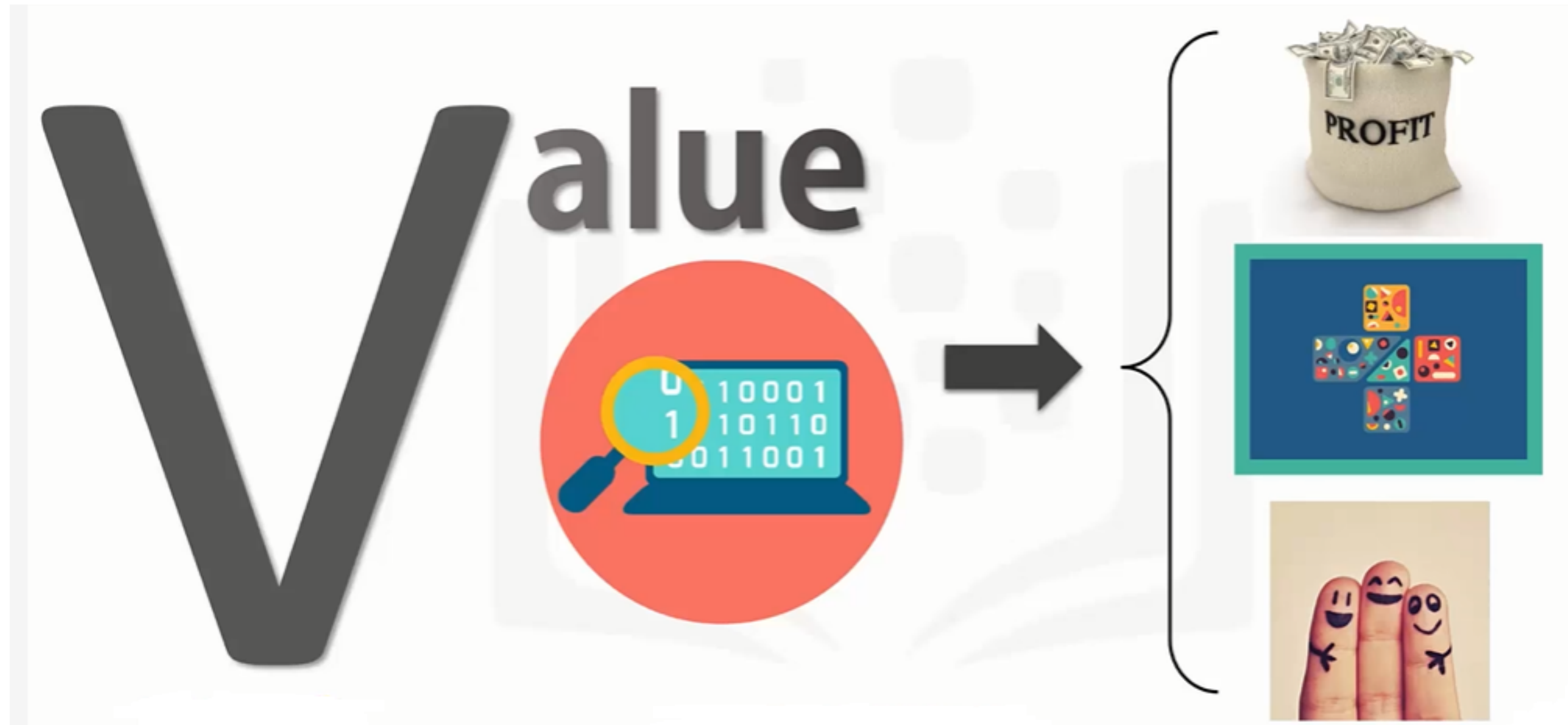
Introduction to Big Data

Examples of the V's in action: Veracity

- **80%** of data is considered to be **unstructured** and we must devise ways to produce **reliable** and **accurate** insights.
- The data must be **categorized, analyzed** and **visualized**.

Introduction to Big Data

Examples of the V's in action: *Emerging V - Value*



Introduction to Big Data

Examples of the V's in action: Value

- The *emerging V* is **value**.
- This **V** refers to our **ability** and need to turn **data** into **value**.
- **Value** isn't just profit.
- It may be **medical** or **social benefits**, or **customer**, **employee**, or **personal *satisfaction***.

Introduction to Big Data

*The main reasons for why people invest time to understand **Big Data** is to derive **value** from it.*

*This is just the beginning of your **Big Data** Course.*

Happy Learning!