



### ACTIVIDAD # 1

### Tipo actividad: Reading: "Grasping the fundamentals of Big Data. (Book-Big Data for Dummies)"

Reading: "Grasping the fundamentals of Big Data. (Book-Big Data for Dummies)"

### Book:

https://jan.newmarch.name/IoT/BigData/Big%20Data%20For%20Dummies.pdf

Chapter One: Grasping the Fundamentals of Big Data.pdf

### 4) Multiple-choice questions about the previous reading.

### **Multiple-Choice Questions:**

- 1. What is the primary challenge organizations face in managing and analyzing data?
  - a. Finding a pragmatic approach
  - b. Dealing with structured data only
  - c. Ignoring customer information
  - d. Using brute-force methods
- 2. What is a characteristic of unstructured data mentioned in the text?
  - a. Stored in relational databases
  - b. Easily recognizable patterns
  - c. Binary large objects (BLOBs)
  - d. Independent of structured query language (SQL)
- 3. \*\*What is the definition of big data, according to the text?\*\*

- a. Any large dataset
- b. Data with extremely high velocity only
- c. Data with at least three shared characteristics
- d. A stand-alone technology

4. What does the evolution of data management include, apart from software advances?





- a. Social media data
- b. Hardware, storage, networking
- c. Predictive analytics
- d. Traditional business models
- 5. Which wave of data management introduced the relational data model and RDBMS?
  - a. Wave 1: Creating manageable data structures
  - b. Wave 2: Web and content management
  - c. Wave 3: Managing big data
  - d. None of the above

6. What technology emerged in response to managing unstructured data in the second wave?

- a. Data warehouses
- b. Data marts
- c. Object Database Management System (ODBMS)
- d. Entity-Relationship (ER) model
- 7. What does the text suggest is the defining factor for the big data era?
  - a. Reduction in costs
  - b. Business process management
  - c. Focus on unstructured data
  - d. Invention of cloud computing
- 8. What is the importance of data virtualization in big data management?
  - a. Storing data in flat files
  - b. Cost-effective storage
  - c. Unstructured data elements
  - d. Snapshot-based storage
- 9. What is an example of data in motion usage mentioned in the text?
  - a. Business analyst understanding customer patterns
  - b. Analyzing the human genome
  - c. Anti Terrorist activities
  - d. Analyzing product quality during manufacturing





- 10. What is emphasized as crucial for big data success in the text?
  - a. Isolated technologies
  - b. Utilizing one tool or technology
  - c. Managing massive amounts of data
  - d. Virtualization of data

# 5) Gap fill activity about the reading: "Grasping the fundamentals of Big Data. (Book-Big Data for Dummies)"

### Gap Filling Exercises:

1. Big data is not a single \_\_\_\_\_ but a combination of old and new technologies that helps companies gain actionable insight.

2. Volume, velocity, and variety are the three main characteristics used to break down big data. Another critical characteristic often considered is \_\_\_\_\_.

3. A big data management architecture involves several key functions: capture, organize, integrate, analyze, and \_\_\_\_\_.

4. To support an unanticipated or unpredictable volume of data, a physical infrastructure for big data is based on a \_\_\_\_\_\_ computing model.

5. Interfaces and feeds, redundant physical infrastructure, security infrastructure, and operational data sources are components of the big data \_\_\_\_\_.

6. MapReduce, Hadoop, and Big Table are innovations that address the capability to process massive amounts of data efficiently, cost-effectively, and in a timely \_\_\_\_\_.

7. Analytical data warehouses and data marts provide compression, multilevel partitioning, and a massively parallel processing \_\_\_\_\_.

8. Big data analytics involve the capability to manage and analyze \_\_\_\_\_\_ of data to deal with clusters of information that could impact the business.

9. With big data, reporting and data visualization become tools for looking at the \_\_\_\_\_\_ of how data is related.







10. Big data applications are designed to take advantage of the unique characteristics of big data, relying on huge volumes, velocities, and varieties of data to transform the behavior of a \_\_\_\_\_.

### 6) Discussion questions about the previous reading.

What are the implications of the three traditional characteristics of big data—volume, velocity, and variety? How does the addition of the fourth V, veracity, impact the reliability of big data analysis?

The reading mentions the importance of validation when combining data sources. Can you provide examples of situations where data validation is critical, and what challenges might arise in ensuring the accuracy and context of combined data?

Explore the role of MapReduce, Hadoop, and Big Table in the evolution of big data. How do these technologies address the fundamental challenge of processing massive amounts of data efficiently, and what specific benefits do they bring to organizations?

# 7) Socialization key vocabulary reading activity #3 "Defining Big Data Analytics. (Book-Big Data for Dummies)"

### Keywords:

**Big Data Analytics**: The process of examining and uncovering insights from large and complex datasets, often involving advanced analytics techniques.

**Amazon**: A multinational technology and e-commerce company known for its successful use of big data analytics, particularly in applications like the recommendation engine.

**Google**: A global technology company recognized for its expertise in big data analytics, search engines, and various online services.

**Recommendation Engine**: A system that utilizes algorithms to analyze user preferences and behaviors, providing personalized suggestions or recommendations, as exemplified by Amazon.

**Predictive Modeling**: The use of statistical algorithms and data-mining techniques to analyze current and historical data, making predictions about future outcomes.









**Text Analytics**: The process of extracting meaningful information from unstructured text data, often involving techniques from computational linguistics and statistics.

**Basic Analytics:** Initial and fundamental analysis of data, including activities like slicing and dicing, reporting, simple visualizations, and basic monitoring.

**Advanced Analytics:** In-depth and sophisticated analysis of data, encompassing techniques such as predictive modeling, machine learning, neural networks, and other advanced data-mining approaches.

**Data Mining:** The exploration and analysis of large datasets to identify patterns, using techniques from statistics, artificial intelligence, and database management.

**Predictive Attributes:** Variables and features within a dataset that are used in predictive modeling to forecast future outcomes.

