Learning Guide Unit 1

Site: University of the People
Course: CS 3440-01 Big Data - AY2025-T3
Book: Learning Guide Unit 1

Printed by: Ryohei Hayashi
Date: Thursday, 30 January 2025, 5:26 PM

Description

Learning Guide Unit 1

Table of contents

Overview

UNIT 1: Introduction to Big Data

Topics

- Introduction to Big Data
- Five V's of Big Data
- Types of Big Data

Learning Objectives

By the end of this Unit, you will be able to:

1. Describe Big Data and its types with examples.
2. Explain the five V's of Big Data (volume, velocity, variety, veracity, and value).
3. Compare and contrast the three basic characteristics that define Big Data (volume, variety, and velocity).

Tasks

- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Complete and submit the Written Assignment
- Make entries to the Learning Journal
- Take the Self-Quiz

Introduction

Before we move to understand Big Data, we first need to understand what is data. Data is simply the characters and symbols that are performed by a computer and end up as electrical signals recorded on a magnetic disk. Big Data is a collection of that data that is usually very large in volume and grows exponentially over time. The traditional data management tools are often not adequate to store or process it. Some examples of this are the New York stock exchanges, which produce well over terabytes of data daily, and can come in one of three forms: structured, unstructured, or semi-structured.

Background of Big Data

Since the early 1990s, the term 'Big Data' has been in use. Although it is still unknown who first used the term, most people give credit to John R. Mashey for coining the term 'Big Data' popular.

Big Data is neither a completely new concept, even if it has only been around for the previous two decades. People have been attempting to employ data analysis and analytics approaches to support their decision-making process for ages. Around 300 BC, the ancient Egyptians attempted to capture all existing 'data' in Alexandria's library. Furthermore, the Roman Empire used to evaluate military statistics in order to determine the optimal distribution of their forces (Where does Big Data come from, 2019). We will learn more about the background later in our reading assignment.

There are five innate characteristics of big data known as the "5 V's of Big Data", which helps us better understand the essential elements of big data.

The history of big data

Let's reflect on these two questions:

- How does the fact that we have been dealing with big data since the 1960s relate to the importance of having a set of basic principles in place?
- As technology advancements continue to evolve, will big data have a larger role in helping businesses expand?

Although the concept of big data is new, large data sets have their origins in the 1960s and 1970s, when the world of data was just getting started with the first data centers or with the use of relational databases.

People started to notice how much data users generated through YouTube, Facebook, and other online services around 2005. That same year, Hadoop (an open-source platform for storing and analyzing large data collections) was founded. During this time, NoSQL began to gain prominence.

For the growth of big data, the development of open-source frameworks, such as Hadoop was essential, as they make big data easier to deal with and store. The volume of big data has exploded since then. Users are still generating massive amounts of data but it isn't just humans who are doing it (Oracle,2022).

In Unit I, we are going to learn about big data and its type, the five V's of Big Data and the characteristics of Big Data.

References

What is big data? (2022). OCI. https://www.oracle.com/big-data/what-is-big-data/

Where does 'big data' come from? (2019, March 26). Enterprise Big Data Framework. https://www.bigdataframework.org/short-history-of-big-data/

Taylor, D. (2022, March 26). *What is big data? Introduction, types, characteristics, examples.* Retrieved May 16, 2022. Guru99. https://www.guru99.com/what-is-big-data.html

Reading Assignment

Read the following to better understand the basics of big data, and how it is closely related to data quality. You will be introduced to the three base elements of big data (volume, velocity, and variety) to gain a better understanding of how they enhance the quality of big data. You will be introduced to the different types of big data, with examples to get a better understanding of how it is defined and structured.

Dontha, R. (2017). *The Origins of Big Data.* KD nuggets. https://www.kdnuggets.com/2017/02/origins-big-data.html

- This article provides information about how big data came into existence.

Hoeren, T., Kolany-Raiser, B. (2017). *Big data in context: Legal, social and technological insights.* Springer Nature. DOI: 10.1007/978-3-319-62461-7 licensed under CC by 4.0.

● Read the first chapter Big Data and Data Quality (pages 1 - 12) which provides an introduction to how big data and data quality are linked and the importance of having quality data for in-depth analysis.

Mannava, P. (2013). A Study on the challenges and types of big data. *International Journal of Innovative Research in Science, Engineering and Technology, 2*(8). Retrieved from https://www.researchgate.net/publication/342003973_A_Study_on_the_Challenges_and_Types_of_Big_Data

● This article provides a theorem that defines the functions of the Big Data transformation, and suggests a Big Data processing design, from the data-mining point of view.

Taylor, D. (2022, March 26). *What is big data? Introduction, types, characteristics, examples.* Retrieved May 16, 2022Guru99. https://www.guru99.com/what-is-big-data.html

● This article helps to understand the definition of Big Data, its types, characteristics, and examples.

Tyagi, V. (2019, January 10). *5 V's of big data.* GeeksforGeeks. https://www.geeksforgeeks.org/5-vs-of-big-data/

● This article provides information on the 5 V's of Big Data.

What is big data? (2022). OCI. https://www.oracle.com/big-data/what-is-big-data/

● This website provides an introduction to big data and covers the three base elements of volume, velocity, and variety.

Video Resources

Eye on Tech. (2020, February 25). *What is big data and what is it used for?* [Video]. YouTube. https://www.youtube.com/watch?v=jH44SfUNpWw&t=59s

● Watch the video from 1.00. This video describes the types of big data and explains where big data can come from, and what it can be used to accomplish.

Simplilearn. (2019, December 10). *Big data in 5 minutes | what is big data | introduction to big data |big data explained | simplilearn [Video].* YouTube. https://www.youtube.com/watch?v=bAyrObl7TYE

● This video explains about Big Data, 5V's, what Hadoop is and why it came into existence.

Discussion Assignment

We all certainly agree that big data is becoming a big game-changer in mostly all of modern industries. In your own words, describe what do you understand by the term 'big data' and explain its types. Do your own research and provide at least six real-time examples of big data.

Your Discussion should be a minimum of 200 words in length and not more than 300 words. Please include a word count. Following the APA standard, use references and in-text citations for the textbook and any other sources.

Use APA citations and references for the textbook and any other sources used; you should use at least 1 APA citation and reference, but you can use more if needed. Refer to the UoPeople APA Tutorials in the LRC for help with APA citations. You are required to post an initial response to the question/issue presented in the Forum and then respond to at least 3 of your classmates' initial posts. You should also respond to anyone who has responded to you. Don't forget to rate the postings of your classmates according to the Rating Guidelines. Review the Discussion Forum rating guidelines to see how your classmates will be rating your post.

After posting an appropriate, meaningful, and helpful response to your three classmates, you must rate their posts on a scale of 0 (unsatisfactory) to 10 (excellent).

10 (A) - Excellent, substantial, relevant, insightful, enriching, and stimulating contribution to the discussion. Also, uses external resources to support position where required and/or applicable.
8 - 9 (B) - Good, quite substantial and insightful, but missing minor details which would have otherwise characterized it as an excellent response.
6 - 7 (C) - Satisfactory insight and relevance, but required some more information and effort to have warranted a better rating.
4 - 5 (D) - Limited insight and relevance of the material; more effort and reflection needed to have warranted a satisfactory grading.
0 - 3 (F) - Unsatisfactory insight/relevance or failure to answer the question, reflecting a poor or limited understanding of the subject matter and/or the guidelines of the question.

The rating scores are anonymous; therefore, do NOT mention in your remarks the separate rating score you will give the peer. The instructor is the only person who knows which score matches the comment given to a peer. Some classmates may worry that some peers will not provide a fair rating, or be unable to provide accurate corrections for grammar or other errors. It is the instructor's responsibility to ensure fairness and accuracy.

Written Assignment

Write about the five V's of big data. Specifically, describe each item and provide examples that represent each item. Discuss the importance of each of these items to the collection and analysis of big data.

You will be assessed based on:

- Description of five V's of Big data.
- Description of each item and provide examples that represent each item.
- Discussion on the importance of each of these items to the collection and analysis of big data.
- Organization and style (including APA formatting)

This paper should be supported by at least one of your readings. Submit a paper that is at least 2 pages in length exclusive of the reference page, double-spaced using 12-point Times New Roman font. The paper must cite a minimum of two sources in APA format and be well written. Check all content for grammar, spelling and be sure that you have properly cited all resources (in APA format) used. Refer to the UoPeople APA Tutorials in the LRC for help with APA citations.

Learning Journal

Reflect on the learning from this week around the fundamentals of big data and respond to the following:

- Compare and contrast the three base elements of big data (volume, velocity, and variety).
- What role do you feel data quality plays in the overall importance of big data collection and analysis? How does it impact these three base elements?

The Learning Journal entry should be a minimum of 500 words and not more than 750 words. Use APA citations and references if you use ideas from the readings or other sources.

The rubric detailing how you will be graded for this assignment can be found within the unit's assignment on the main course page.

Self-Quiz

The Self-Quiz gives you an opportunity to self-assess your knowledge of what you have learned so far.

The results of the Self-Quiz do not count towards your final grade. However,the quiz is an important part of the University's learning process and it is expected that you will take it to ensure understanding of the materials presented. Reviewing and analyzing your results will help you perform better on future Graded Quizzes and the Final Exam.

Please access the Self-Quiz on the main course homepage; it is listed inside the Unit.

Checklist

- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Complete and submit the Written Assignment
- Make entries to the Learning Journal
- Take the Self-Quiz