

Learning Guide Unit 7

Site: [University of the People](#)
Course: CS 3440-01 Big Data - AY2025-T1
Book: Learning Guide Unit 7

Printed by: Mejbaul Mubin
Date: Thursday, 5 September 2024, 2:30 PM

Description

Learning Guide Unit 7

Table of contents

Overview

Introduction

Reading Assignment

Discussion Assignment

Written Assignment

Learning Journal

Self-Quiz

Checklist

Overview

UNIT 7: Trends in Big data

Topics

- Advanced analytics in big data
- MapReduce paradigm in big data
- Artificial Intelligence uses in big data

Learning Objectives

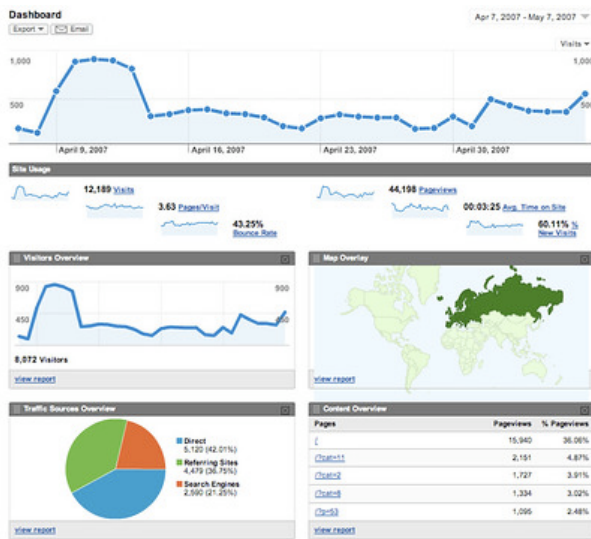
By the end of this Unit, you will be able to:

1. Discuss how advanced analytics has impacted big data analysis
2. Explain the impact of MapReduce paradigm on big data
3. Identify and describe the use of artificial intelligence in big data analysis

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 and submit the Self-Quiz

Introduction



Advanced analytics is a data analysis methodology that employs statistical techniques, deep learning, machine learning algorithms, predictive modeling, and other ways to analyze corporate data from many data sources. Beyond conventional business intelligence (BI) methodologies, advanced analytics employs data science to identify trends and calculate the chance of upcoming events. As a result, a company may be able to respond faster and make decisions with a great deal more precision (Hanna, Burns, & Preslar, 2022). In big data, advanced analytics can help organizations get a better understanding of their company, and ways to improve customer interactions.

The MapReduce paradigm was developed in 2003 to facilitate highly distributed computation of enormous data collections. The MapReduce concept aims to simplify the translation and analysis of huge datasets while allowing programmers to concentrate on algorithms rather than data management. The model makes data-parallel algorithms easy to build. This paradigm has been implemented in several ways, including Google's approach (c++) and Apache's Hadoop implementation (java - Based). Both

applications operate in a shared-nothing, peer-to-peer environment across huge clusters of commodity hardware (IBM, 2022). In big data the MapReduce paradigm is mainly used with the Hadoop architecture in facilitating data translations on very large datasets.

AI makes big data analytics easier by automating and improving complicated analytical activities that would otherwise be time-consuming and labor-intensive, such as data preparation, data visualization, and predictive modeling. AI speeds up analyzing massive, complicated data and generating relevant user insights (QLIK.com, 2022).

In this unit, we will learn how big data impacts advanced data analytics, what the map-reduce paradigm is, and how it has influenced the development and advancement of big data analytics and analysis. We will also touch on artificial intelligence's impact on big data and how it can be used to further data analysis.

References

Hanna, K. T., Burns, E., & Preslar, E. (2022). [What is advanced analytics?](https://www.techtarget.com/searchbusinessanalytics/definition/advanced-analytics#:~:text=Advanced%20analytics%20is%20a%20data,a%20variety%20of%20data%20sources) TechTarget.

<https://www.techtarget.com/searchbusinessanalytics/definition/advanced-analytics#:~:text=Advanced%20analytics%20is%20a%20data,a%20variety%20of%20data%20sources>.

IBM. (2022, Oct 3). [The MapReduce paradigm](https://www.ibm.com/docs/en/netezza?topic=guide-mapreduce-paradigm). Netezza Performance Server for Cloud Pak for Data System 1.0.X from

<https://www.ibm.com/docs/en/netezza?topic=guide-mapreduce-paradigm>

[How Do Big Data and AI Work Together?](https://www.qlik.com/us/augmented-analytics/big-data-ai#:~:text=How%20is%20AI%20used%20with,%2Dintensive%20and%20time%2Dconsuming) (2022). QLIK.com. <https://www.qlik.com/us/augmented-analytics/big-data-ai#:~:text=How%20is%20AI%20used%20with,%2Dintensive%20and%20time%2Dconsuming>.

Reading Assignment

Read through the following to better understand what advanced analytics is and how it impacts big data gathering and analysis, you will also learn the impact of the map-reduce paradigm has on big data, and what impact artificial intelligence has in the big data realm.

Note to students: To access the LIRN resources, you must log in to Moodle and access the Library and Information Resource Network (LIRN) located under the Resources link on the Home page. Click on the Alphabetical View tab at the top of the page and scroll down to the database where JSTOR is located. Copy and paste the title of the resource, into the search bar. A link to the resource will appear. If you have any problems, please contact library@uopeople.edu.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). [Transformational issues of big data and analytics in networked business](https://www.jstor.org/stable/26629677). *Management Information Systems Research Center Quarterly*, 40(4), 807–818. <https://www.jstor.org/stable/26629677>

- This article discusses some issues of big data and analytics in a networked business.

Benke, K., & Benke, G. (2018). [Artificial intelligence and big data in public health](https://www.mdpi.com/1660-4601/15/12/2796/htm). *International Journal of Environmental Research and Public Health*, 15(12), 2796. <https://www.mdpi.com/1660-4601/15/12/2796/htm>

- This article discusses the use of artificial intelligence and big data in the public health realm.

Corbeil, M. E, Corbeil, J.R, & Khan, B. H. (2017). [A framework for identifying and analyzing major issues in implementing big data and data analytics in e-learning: introduction to special issue on big data and data analytics](http://www.jstor.org/stable/44430534). *Educational Technology* 57, (1),3–9. <http://www.jstor.org/stable/44430534>.

- This article some issues related to the implementation of big data and data analytics in an e-learning environment.

Lev-Libfeld, A., & Margolin, A. (2019). [Fast data: Moving beyond big data's map-reduce](https://arxiv.org/ftp/arxiv/papers/1906/1906.10468.pdf). *Journal of Geopython*, (1), <https://arxiv.org/ftp/arxiv/papers/1906/1906.10468.pdf>

- This article discusses moving past big data's MapReduce paradigm.

Prakash, A. A., & Aloysius, A. (2019). [A novel hadoop map reduce paradigm using an effective EPCA approach](http://www.ijaconline.com/wp-content/uploads/2019/04/08-Antony-Prakash.pdf). *Journal of Analysis and Computation*, 12(3), 1-11. <http://www.ijaconline.com/wp-content/uploads/2019/04/08-Antony-Prakash.pdf>

- This article discussed an approach for using the Map Reduce paradigm with Hadoop.

Sinha, S. (2016, November 15). [Fundamentals of mapreduce with mapreduce example](https://medium.com/edureka/mapreduce-tutorial-3d9535ddbe7c). Edureka. <https://medium.com/edureka/mapreduce-tutorial-3d9535ddbe7c>

- This article provides the reader with the fundamentals of MapReduce and provides examples.

Taylor, D. (2022, September 17). [What is mapreduce in hadoop? Big data architecture](https://www.guru99.com/introduction-to-mapreduce.html). Guru99. Retrieved September 27, 2022, from <https://www.guru99.com/introduction-to-mapreduce.html>

- This article discusses MapReduce in Hadoop.

Wedel, M., & Kannan, P. K. (2016). [Marketing analytics for data-rich environments](http://www.jstor.org/stable/44134975). *Journal of Marketing*, 80(6), 97–121. <http://www.jstor.org/stable/44134975>

- This article provides insights into marketing data analytics for data-rich environments.

Video Resource

IMDA Singapore. (2019, June 12). [Tech explainer / what is advanced analytics?\[Video\]](https://www.youtube.com/watch?v=Zf48s5PHMbK). YouTube. <https://www.youtube.com/watch?v=Zf48s5PHMbK>

- This video explains advanced data analytics in a very brief video.

Discussion Assignment

Advanced analytics plays a major role in big data analysis and reporting. It allows analysts to apply advanced techniques to their data analysis resulting in findings that may not have been revealed using legacy data analysis procedures/algorithms.

- Discuss any two techniques employed by advanced analytics and how they have influenced big data analysis and data gathering.

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 other sources.

Use APA citations and references for the textbook and any other sources; 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 your classmates' postings according to the Rating Guidelines. Review the Discussion Forum rating guidelines to see how your classmates will rate 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, it uses external resources to support positions 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

For this week's written assignment, answer the following questions:

In this week's reading by [Lev-Libfeld and Margolin \(2019\)](#), the authors discuss the fact that the MapReduce paradigm is based on several assumptions, which include: the completeness of data, independence of data set calculations, and relevancy distinguishability.

- Describe what each of these assumptions means.
- How will this impact the MapReduce paradigm, if you fail to evaluate the above-said assumptions?
- What will the effect of the impact on big data security be?

You will be assessed based on:

- Description of the completeness of data, independence of data set calculations, and relevancy distinguishability
- Explanation of the impact, on failing to evaluate assumptions noted in the MapReduce paradigm and the effect on big data security.
- Organization and style (including APA formatting)

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 and spelling and be sure you have correctly cited all resources (in APA format) used. Refer to the [UoPeople APA Tutorials in the LRC](#) for help with APA citations.

Reference

Lev-Libfeld, A., & Margolin, A. (2019). [Fast data: Moving beyond big data's map-reduce](#). *Journal of Geopython*, (1), <https://arxiv.org/ftp/arxiv/papers/1906/1906.10468.pdf>

Learning Journal

Artificial intelligence has begun to play a big role in data collection and analytics in the big data realm. It provides alternative ways for organizations to capture data and helps reveal patterns not previously seen with other data generation or collection methods.

- Identify and describe in detail three ways artificial intelligence can be used with big data analysis.

The Learning Journal entry should be a minimum of 400 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