**Unit 1**

**Reading Task:**

* [What is Big Data?](https://www.forbes.com/video/4857597029001)
  + This video explains what big data are and how companies successfully leverage their use.
* [Big Data and Hadoop](https://www.youtube.com/watch?v=FHVuRxJpiwI)
  + This video explores how Hadoop can help process large, complex data sets.
* [Massively Parallel Processing and Big Data](https://www.oracle.com/technetwork/database/bi-datawarehousing/twp-hadoop-oracle-194542.pdf)
  + This white paper investigates how large and complex data can be processed effectively.
* [Data Warehouse or Hadoop?](http://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB-6448.pdf)
  + This white paper provides insight into when to use data warehousing and when to use Hadoop.
* [Introduction to Streaming Data](https://towardsdatascience.com/introduction-to-stream-processing-5a6db310f1b4)
  + This article provides an introduction to stream processing.
* [Encrypting Data at Rest](https://d1.awsstatic.com/whitepapers/AWS_Securing_Data_at_Rest_with_Encryption.pdf)
  + This white paper provides an overview of different methods for encrypting your data at rest, which are available today.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Unit 1 - Discussion Board** | | | **Task Type:** | Discussion Board | | **Deliverable Length:** | A minimum of 100 words | | **Points Possible:** | 10 | | **Description:**  **Reminder: Primary Response posts are due by Thursday.**  Students will be expected to post their primary Discussion Board posting by Thursday of each week. Discussion posts will be graded, and late submissions will be assigned a late penalty per the Late Penalty Policy in the syllabus. NOTE: All submission posting times are based on midnight Central Time.  This is a terrific time to meet your classmates, learn a little bit about them, and for them to know about you. Please introduce yourself in the Introductory Discussion Board, accessed via the Discussion Board link to the left. If you are unsure how to start, begin by discussing your background and why you are in school. Feel free to mention your proudest accomplishments and offer advice to others. Say something unique or fun about yourself! Additionally, please review the course outline and share at least one area you look forward to learning more about with your classmates.  Your introduction should be a minimum of 100 words.  You may also upload a picture by clicking "Post File" when responding to the Discussion Board. JPEG images work well, but it will accept other formats, too.  **Get started today. For purposes of your attendance, please post your introduction to this discussion by Thursday.**  **You are encouraged to participate in this thread throughout the term.** | | | **Course Objectives:** | | | **Model Answer:** | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Unit 1 - Individual Project** | | | **Task Type:** | Individual Project | | **Deliverable Length:** | 3-5 pages (not including cover page and resource page) | | **Points Possible:** | 125 | | **Description:**  **Assignment Details**  During your first Individual Project (IP), you will set up the Python environment Anaconda that will be used for subsequent IPs. The Spyder IDE will be installed during this task, and a simple program will be created to demonstrate that it works correctly. [This link will provide guidance.](https://docs.anaconda.com/anaconda/install/)  Once the IDE is installed, research Python data streaming libraries. Identify 1 of interest, and you will utilize it for the Unit 1 activity. Write the needed Python code based on the [Titanic data set](https://resources.careered.com/LCMSFileSharePreview/Resources/ZippedFiles/titanic_train.zip) that assumes records will be streaming and predictions could be made for survival.  You will not write the predictions for this activity, but the streaming data architecture will be defined.  The project deliverables include the following:   * [Download the Titanic data set](https://resources.careered.com/LCMSFileSharePreview/Resources/ZippedFiles/titanic_train.zip). * Utilize Python and a streaming library to consume the data set. * Provide a screenshot of the ingested data. * Submit your solution for grading in a Microsoft Word document.   **Individual Project Rubric**  The Individual Project (IP) Grading Rubric is a scoring tool that represents the performance expectations for the IP. This Individual Project Grading Rubric is divided into components clearly describing what should be included within each IP element. It is the roadmap that can help you develop your IP.   |  |  |  | | --- | --- | --- | | **Points Possible** | **Points Earned** | **Comments** | | 30 |  |  | | 30 |  |  | | 30 |  |  | | 25 |  |  | | 10 |  |  | | 125 |  |  | |  |  |  |   **Reference**  Anaconda. (2022). *Installation*. <https://docs.anaconda.com/anaconda/install/> | | | **Course Objectives:**  ·        Analyze the preparation, collecting, storing, securing, analyzing, interpreting, processing, and reporting of streaming data.  ·        Examine the most common challenges of managing streaming data analytics-a mix of streaming and at-rest data collection, measurement, tracking, analysis, security, and reporting.  ·        Understand the differences between streaming data analytics and at-rest data analytics, as well as the foundation, techniques, and processes of streaming data analytics. | | | **Model Answer:**  Student approaches may vary, but all should reflect the following:   * + The data should be loaded into Python. | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Unit 1 - Live Chat Extra Credit Summary** | | | **Task Type:** | Live Chat | | **Deliverable Length:** | 200 words | | **Points Possible:** | 15 | | **Description:**  You can receive credit in this class for attending Live Chats or reviewing the chat archive. This opportunity is available once per week. To receive these points, you can either:  1.    Attend the Live Chat session. or  2.    Review the archived chat session and submit a 200-word content summary. The archive review summary must meet the expectations described below.   * Summarize the content of the chat. Some ideas for what to include are a description of what information was covered and how you will apply this information to your academic or professional work. * Label the "Chat Credit" document when the Live Chat was presented. * A summary must be submitted within one week of the Live Chat to be considered for credit.   Your instructor has the discretion to determine whether to award the points. A summary must meet each of these guidelines. Points are awarded on an all-or-nothing basis.  To submit your summary, head to the assignment list and select Unit 1 - Live Chat Extra Credit Summary.  View the [Individual Project Demo](https://careered.libguides.com/ctu/success/classroom#s-lg-box-wrapper-26053866) for guidance on submitting your chat summary as an Individual Project via the Virtual Classroom or the CTU Mobile app. | | | **Course Objectives:**  ·        Analyze the preparation, collecting, storing, securing, analyzing, interpreting, processing, and reporting of streaming data.  ·        Examine the most common challenges of managing streaming data analytics-a mix of streaming and at-rest data collection, measurement, tracking, analysis, security, and reporting.  ·        Understand the differences between streaming data analytics and at-rest data analytics, as well as the foundation, techniques, and processes of streaming data analytics. | | | **Model Answer:**  Students must meet all of the expectations outlined in the assignment. | | |

**Optional Live Session Talking Points:**

* The Three *V*s. of Big Data
* Structured Versus Unstructured Data
* Common Applications of Big Data
* Streaming Data Versus At-Rest Data
* Utilizing the Python Programming Environment With Anaconda