**DAILY ASSESSMENT FORMAT**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date:** | **08/07/2020** | **Name:** | **Lavanya B** |
| **Course:** | **Matlab** | **USN:** | **4al17ec043** |
| **Topic:** | **Loops**  **Arrays**  **Importing data….** | **Semester & Section:** | **6th A** |
| **Github Repository:** | **Lavanya B** |  |  |

|  |
| --- |
| **FORENOON SESSION DETAILS** |
| **Image of session** |
|  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date:** | **08/07/2020** | **Name:** | **Lavanya B** | |
| **Course:** | **IoT** | **USN:** | **4al17ec043** | |
| **Topic:** | **Big data processing** | **Semester & Section:** | **6th A** | |
| **AFTERNOON SESSION DETAILS** | | | |
| **Image of session** | | | |
| **Report**  **Big data**  **What is Big Data?**  **Data is information that comes from a variety of sources, such as people, pictures, text, sensors, and web sites. Data also comes from technology devices like cell phones, computers, kiosks, tablets, and cash registers. Most recently, there has been a spike in the volume of data generated by sensors. Sensors are now installed in an ever growing number of locations and objects. These include security cameras, traffic lights, intelligent cars, thermometers, and even grape vines!**  **Here are three characteristics that indicate an organization may be dealing with Big Data:**  **They have a large amount of data that increasingly requires more storage space (volume).**  **They have an amount of data that is growing exponentially fast (velocity).**  **They have data that is generated in different formats (variety).**  **How much data do sensors collect? Here are some estimated examples:**  **Sensors in one autonomous car can generate 4,000 gigabits (Gb) of data per day.**  **An Airbus A380 Engine generates 1 petabyte (PB) of data on a flight from London to Singapore.**  **Safety sensors in mining operations can generate up to 2,4 terabits (TB) of data every minute.**  **Sensors in one smart connected home can produce as much as 1 gigabyte (GB) of information a week.**  **While Big Data does create challenges for organizations in terms of storage and analytics, it can also provide invaluable information to fine-tune operations and improve customer satisfaction.**  **Large datasets**  **Companies do not necessarily have to generate their own Big Data. Smaller organizations might not have the sensors, the volume of customers, or the ability to generate the variety of information that could benefit their company. There are sources of free data sets available, ready to be used and analyzed by anyone willing to look for them.**  **Many companies of various sizes believe they have to collect their own data to see benefits from big data analytics, but it is simply not true.**  **Distributed process**  **Distributed Processing**  **From a data management perspective, analytics were simple when only humans created data. The amount of data was manageable and relatively easy to sift through. However, with the explosion of business automation systems and the exponential growth of web applications and machine-generated data, analytics is becoming increasingly more difficult to manage. In fact, 90% of data that exists today has been generated in just the last two years. This increased volume within a short period of time is a property of exponential growth. This high volume of data is difficult to process and analyze within a reasonable amount of time.**  **Rather than large databases being processed by big and powerful mainframe computers and stored in giant disk arrays (vertical scaling), distributed data processing takes the large volume of data and breaks it into smaller pieces. These smaller data volumes are distributed in many locations to be processed by many computers with smaller processors. Each computer in the distributed architecture analyzes its part of the Big Data picture (horizontal scaling).**  **Most distributed file systems are designed to be invisible to client programs. The distributed file system locates files and moves data, but the users have no way of knowing that the files are distributed among many different servers or nodes. The users access these files as if they were local to their own computers. All users see the same view of the file system and are able to access data concurrently with other users.**  **Hadoop was created to deal with these Big Data volumes. The Hadoop project started with two facets: The Hadoop Distributed File System (HDFS) is a distributed, fault tolerant file system, and MapReduce, which is a distributed way to process data. Hadoop has now evolved into a very comprehensive ecosystem of software for Big Data management.**  **Hadoop is open-source software enabling the distributed processing of large data sets that can be terabytes in size and that are stored in clusters of computers. Hadoop is designed to scale up from single servers to thousands of machines, each offering local computation and storage. To make it more efficient, Hadoop can be installed and run on many VMs. These VMs can all work together in parallel to process and store the data.**  **Hadoop has two main features that have made it the industry standard for handling Big Data:**  **Scalability - Larger cluster sizes improve performance and provide higher data processing capabilities. With Hadoop, cluster size can easily scale from a five node cluster to a one thousand node cluster without excessively increasing the administrative burden.**  **Fault tolerance – Hadoop automatically replicates data across clusters to ensure data will not be lost. If a disk, node, or a whole rack fails, the data is safe.** | | | |