**Module 2: Real time Data**

**Provide an example in which real-time data can be used as a competitive advantage for corporations. Research and provide your thoughts on the following: Why do you feel this real-time data is an advantage? How long can the advantage last?**

We've all heard that data is the secret to Amazon and Google's revenue success. All of these businesses have been able to turn the massive amounts of data they collect from their customers into significant revenue streams, whether it's through their search patterns, the posts they share, the products they buy, or the music they listen to. It's the fact that those businesses have been able to effectively use that information to better understand and market to their customers.

Artificial intelligence and Deep Learning is being used by all of these organizations to accomplish this. Many firms, large and small, will begin to turn to AI in order to develop better data strategies, achieve client uptake, and better compete against their competitors.

The advantages of real-time analytics are innumerable. There are a few distinctive advantages to which no company can afford to ignore are:

* KPI Visualization in Real Time
* Increased market competitiveness
* Data-driven demand sensing
* Management of Business Processes in a Proactive Manner
* Customer profiling that is accurate

Every piece of data you collect on a customer or potential consumer is thus another piece of data you can utilize to develop a successful marketing or advertising strategy. Artificial intelligence, big data, and data analytics are all intertwined. Data, reams and reams of information, is required for artificial intelligence and, by extension, deep learning. Deep learning can only be useful for your company if you provide it with a constant stream of data to feed it.

**Reference:**

[1] Martyn Barnett. (Feb, 2014). The Importance of Real Time Data & Five Reasons Why You Need it. *Fourth Source*. Retrieved from https://www.fourthsource.com/data/importance-real-time-data-five-reasons-need-22014

**Module 3: Real life scenario**

**You are the Data Analytics manager for a medium sized health care software provider. Your manager is asking you to create a system using Apache Hadoop and some of the components we learned about in the lab last week. Here are the requirements:**

* **Incorporate data from 45 million rows in a MySQL corporate database containing patient records, histories, visits, and prescriptions.**
* **Incorporate unstructured data captured from patient monitoring equipment present at hundreds of patient care facilities**

**You are being asked to answer the question:**

* **What is the health trend of patients being treated at our customer locations, given the measurement of vitals measured with our monitoring equipment for these 3 indicators (blood sugar, pulse, temperature) ?**

**As the Data Analytics manager, how would you go about completing your assignment?**

If I were the Data Analytics Manager, I would collect the data and do some analysis on top of it. A deep review and discussions with the teammates and corresponding stakeholders will better serve the purpose of understanding on them. It will give me a deeper insight on the what kind of data, what are the attributes present and their types, what are necessary and what are not necessary for the business. If I had a small amount of time to put up the system, I would start with the MySQL database. It will take a significant amount of effort to include unstructured data from hundreds of patient care facilities.

As a result, starting with this dataset is faster and easier. In order to process such a vast dataset, Apache Hadoop will come in handy. As a result, computer technology and infrastructure must be capable of delivering a cost-effective implementation of limited Parallel Data Processing. It can take advantage of the MapReduce method to handle data more quickly by using Apache Hadoop. The fundamental benefit of migrating from a relational database like MySQL to an HDFS is increased performance. Furthermore, the HDFS can provide redundancy, allowing computer nodes to remain up even if one goes down. Provide storage for unstructured data collections in the billions and trillions if needed for a system.

In healthcare, big data is utilized to cut costs, cure diseases, increase profitability, predict epidemics, and improve the quality of life by preventing deaths. This will considerably enhance the throughput of large datasets, such as the patient records, histories, visits, and prescriptions described in the problem statement. We can now start tracking and analyzing the data. We're probably checking blood sugar, pulse, temperature and other vital signs and could just query the new system and perform some data analyses on top of it. I would perform some analytics by clustering patients with comparable vitals and create a healthy and unwell patient model to examine overall patient health by geographic area with in a period of time,

**References:**

[1] Dash, S., Shakyawar, S.K., Sharma, M. et al. (2019). Big data in healthcare: management, analysis and future prospects. *J Big Data 6, 54*. Retrieved from https://doi.org/10.1186/s40537-019-0217-0

[2] Sahil Dhankhad. (Apr 8, 2019). A Brief Summary of Apache Hadoop: A Solution of Big Data Problem and Hint comes from Google. *Towards Data Science, Medium Blog.* Retrieved from https://towardsdatascience.com/a-brief-summary-of-apache-hadoop-a-solution-of-big-data-problem-and-hint-comes-from-google-95fd63b83623

**Module 4: Big Data Trends**

**Please read the two articles below, using the provided URL. Asses the information, check the author profiles and how suitable those profiles or other activities to the Big Data Management topic.**

* **Provide cons and pros for each article.**
* **Clearly provide your personal opinion, based on the information from both article's :**
  + **Do you think Sensor Data structured or unstructured?**
  + **Explain why do you think so.**

**Article 1:**

**https://blog.westerndigital.com/examples-of-unstructured-data/#:~:text=Unstructured%20data%20is%20data%20that,isn't%20in%20a%20database.&text=Examples%20of%20unstructured%20data%20includes,sensor%20or%20social%20media%20posts**

**Article 2:**

[**https://searchdatamanagement.techtarget.com/blog/The-Wondrous-World-of-Data/Big-Data-Myth-4-Big-Data-is-Unstructured-Data#:~:text=Many%20of%20the%20sensor%2Ddriven,this%20data%20is%20highly%20structured.&text=Almost%20all%20of%20this%20machine,analytical%20purposes%20and%20nothing%20else.**](https://nam12.safelinks.protection.outlook.com/?url=https:%2F%2Fsearchdatamanagement.techtarget.com%2Fblog%2FThe-Wondrous-World-of-Data%2FBig-Data-Myth-4-Big-Data-is-Unstructured-Data%23:~:text%3DMany%2520of%2520the%2520sensor-driven%2Cthis%2520data%2520is%2520highly%2520structured.%26text%3DAlmost%2520all%2520of%2520this%2520machine%2Canalytical%2520purposes%2520and%2520nothing%2520else.&data=04%7C01%7Cv.shevchenko%40northeastern.edu%7Cb0e7f4aad7a4472ebe4c08d8c97f1f54%7Ca8eec281aaa34daeac9b9a398b9215e7%7C0%7C0%7C637480895116999779%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=zbiizcOvfd9fUr0r0D6HT7L05rQkW1uIWMfiV0ocHlw%3D&reserved=0)

**Article 1:**

Now-a-days everything is connected to one or the other device in and around us, thanks to advancements of IoT, Sensors, Analytics, Data, and AI. This article talks about the real-time enterprise examples and their use cases. In my opinion large enterprise systems and machines continuously output information in some or the other format. Unstructured data is frequently observed and classified as qualitative data. It is incapable of being processed and analyzed using standard tools and procedures. Because unstructured data isn't well suited to relational databases, it's more commonly found in object-oriented databases like NoSQL. Combining IoT data sources for cross-referential analysis is a difficult task if the data can’t be identical in a relational manner.

**Unstructured Data Pros:**

* A huge variety of formats and Data Models are flexible
* Stored in Data Lakes and uses NoSQL Non-relational Databases
* Provides better insights and more opportunities to turn our data into a competitive advantage

**Unstructured Data Cons:**

* Challenging to analyze and leverage unstructured data
* Special Tools are necessary for effective Data Management

**Article 2:**

Let’s take an example of Amazon e-commerce website where people often purchase items on it. When a user is buying a product, Amazon tracks every action taken by users in order to determine which adverts to show and which products to recommend. So, in this product id, pricing data, customer data are structured in nature and customer behavior, spending patterns, reviews, ratings, social media mentions, media, customer service, and logs are in unstructured manner. Both the data are important for finding patterns of a customer. The main problem is to figure out how to use and handle it for maximum benefit.

**Structured Data Pros:**

* Several formats and Easy to search
* Requires less processing in comparison to unstructured data and is easier to manage.
* Several tools available for structured data that simplify usage, management, and analysis

**Structured Data Cons:**

* Structured data is stored in data warehouses which are built for space saving but are difficult to change and not very scalable/flexible
* Limited usability because of its pre-defined structure/format

**References:**

[1] Eric Ottem. (Sept 16, 2019). Examples of Unstructured Data ⁠— 4 Enterprise Use Cases. *Western Digital Blog*. Retrieved from https://blog.westerndigital.com/examples-of-unstructured-data/

[2] Rick van der Lans. (12 Oct 2015). Big Data Myth 4: Big Data is Unstructured Data. *Search Data Management Tech Target.* Retrieved from https://searchdatamanagement.techtarget.com/blog/The-Wondrous-World-of-Data/Big-Data-Myth-4-Big-Data-is-Unstructured-Data

**Module 5: Streaming Data**

**Discuss the following questions:**

* **What new Analytics tool needs to be invented and why?**
* **What personal needs do you have that would benefit from such a new tool?**

**What new Analytics tool needs to be invented and why?**

Decision makers should be the focus of the new application platform that needs to be made. When data delivers critical info to take measures. When this comes to future of predictive analytics, the source of evidence we use will determine our decisions and actions. There will only be an ever-increasing volume of data as time passes. We can analyze data and turn it into visuals so that it can be read with the help of these analytical tools. There are 4 sorts of analytics, they are diagnostics, descriptive, predictive, and prescriptive. It's a lot easier to mesh product data with internal company data, it's a more practical long-term solution for a specific organization, and it offers a lot more specific in-depth research, particularly predictive modeling. However, one disadvantage is that data integrity may fall partly against you to transmit the correct events. So, a tool with ML/ AI enablement for easy data management and followed by all the integrities to follow while implementing things would be handy for Analysts.

**What personal needs do you have that would benefit from such a new tool?**

Given what we've learned about the main key metrics that predict a team's growth, an advanced analytic solution must include these insights and enable a limited set of organized activities and things for various sectors right away. For instance, if I had important data indicating that one of my company's processes might fail, this newly designed tool should be able to present me with the facts and information I need, as well as other options I can take based on the data. This, to be honest, removes the most of the human parts and focuses more on machine learning and AI. When it comes to personal needs, the type of advantages that I would personally receive from such a new tool would be viewing the various options available to me, as well as the potential results of each option. I'd then be able to make a decision based on what I believe is best for me, given the situation I'm in, how I want to go, and so on.

**Reference:**

[1] Sandra Khvoynitskaya. (January 30, 2020). The future of big data: 5 predictions from experts for 2020-2025. *Itransition* *Blog*. Retrieved from https://www.itransition.com/blog/the-future-of-big-data