Joshua Corelli Big Data

Large collections data can be analyzed, processed, and stored, which is what the Big Data field is all about (Chapter 1, page 4). In the video, “Secret Life of Big Data | Intel”, the speaker gives an example of the King of England, William, and how he surveyed his people, created the Doomsday book from these surveys, and judged his people with this knowledge. This is a great example of Big Data and how large collections of data can be stored. The surveys collected the data to be analyzed, the Doomsday book organized these surveys to be processed, made sense of, and then stored. This book or data could then be used for whatever purpose the handler decides, which in this case judges used this book to extract value from the data and question people about their property and objects, as well as keeping records.

In Chapter 1 of the textbook it states the wide range of insights and benefits of the resulting Big Data. The resulting Big Data provides intelligence and knowledge that can allow a business to make accurate predictions, fault and fraud detections, improve decision-making, and provide other discoveries that were not known before (Page 5). The Doomsday book in the “Secret Life of Big Data | Intel” video provided these same insights and benefits.

In the video, “Anyone can harness the power of Big Data | Haley Wixom | TEDxConcordCarlisleHighSchool”, she states that the three steps to a data project are data collection, data analysis, and value creation. This also ties into what the textbook says about Big Data collection in Chapter 1 where the Big Data field is large collections of data that can be store or collected, analyzed, and processed. Her first data project was to optimize her drive to school. Based off her data results she determines that on Tuesdays her drive to school is the longest and on Fridays her drive to school is the shortest. Her second data project was to determine how she could fall asleep faster. Based off her data results she discoveries that finishing her work earlier allowed her to fall asleep faster. Both data results provided a wide range of insights and benefits and concur with the insights and benefits said in Chapter 1, providing her with accurate predictions, new discoveries, and improved decision-making of when to leave for school and when to stop working (page 5).

Data analytics is the process of identifying, procuring, preparing, and analyzing large amounts of raw unstructured data to get meaningful information that can then be used for data-decision making. Data analytics prevents data-decision making from being based on made up stories that can be twisted and changed, but rather on factual data by using scientific backing (Chapter 1, page 7). In the video, “Secret Life of Big Data | Intel”, she explains that the Doomsday book is messy because it starts with humans who make mistakes. She says we make up stories and lie, struggling to tell the facts. For example, when we put information on dating cites, we do not always tell the truth. She also says the under different circumstances our answers may be different. For example, when we are asked questions by someone with a sword vs without a sword are answers may differ. She adds that how you visualize data can make facts suddenly and abundantly clear. For example, the cholera outbreak in London. They did not look for where people died just when they died. Once they started looking for where people died, they found that it was all on one street where everyone used the same water source. With data

analytics enabling data-driven decision-making being backed by scientific knowledge rather than stories and new ways of visualizing, data is more accurate and factual taking out this human error (Chapter 1, page 7).

There are 9 stages of Big Data analytic lifecycle: Business Case Evaluation, Data Identification, Data Acquisition and Filtering, Data Extraction, Data Validation and Cleaning, Data Aggregation and Representation, Data Analysis, Data Visualization, and Utilization of Analysis Results (Chapter 3, Page 55). In the video, "Supercharge Your Marketing with Google Cloud (Cloud Next '19 UK)", the last two speakers said to start your Big Data projects small. Analyze, process, and store that data and then make bigger projects from your data results. Their point was to not do too much at once or you will make your life harder and your data results could be sloppy, take it slow and then expand. The reason Big Data analytics is laid out into 9 categories as a step-by-step methodology is because each step has a specific purpose and value that is needed to organize the activities and tasks involved in analyzing, processing, and storing data (Chapter 3, page 55). There are no skipping steps when adopting Big Data and its planning perspective (Chapter 3, page 48).

For a collection of data to be considered Big Data there are five characteristics to be considered. These five characteristics are volume, velocity, variety, veracity, and value (Chapter 1, page 13). The velocity of data is the amount of time it takes for the data to be processed once it enters the enterprise's perimeter and value is the usefulness of data for an enterprise and depends on the velocity of data (Chapter 1, page 14-16). In the video, "Google Cloud Next Madrid '17- Robert Saxby: 'Introduction to Big Data'", the speaker explains the importance of event time and processing time through Apache beam. He states that data whose event time is way earlier than the processed time is bad. This supports what is stated in Chapter one of the textbook. Since the value of data is dependent on velocity of data, the longer it takes to process the data the lower the value of the data. This is because analytic results have a shelf-like and data that can be analyzed quickly has more value to business, allowing them to make decisions quickly (Chapter 1, page 16).

The value of data is also dependent on the veracity of data, which refers to the quality or accuracy of the data collected. Data the enters Big Data needs to be checked for quality. For data to have a good quality, invalid data or noise data needs to be removed from the data in order for the data to have a good signal-to-noise ratio and thus a higher value (Chapter 1, page 16). In the video, "Supercharge Your Marketing with Google Cloud (Cloud Next '19 UK)", the last two speakers talk about their journey to using Google Cloud and how it has benefited them. In Google Cloud they built an engine for understanding their customers, an engine for activating their customer segments automatically, segments that are up to date every day, and reaching all google marketing platform for activation. By doing this, data was specific for that customer and always accurate due to the day to day recalculations, getting rid of all data that was invalid or noise data that did not pertain to the customer. In conclusion, they were able to provide their customers the best quality and relevant data pertaining to that specific customer, which based of the information of veracity and value of data in Chapter 1, their data has a high veracity and value to their business.

In 2008 there was a global recession and companies need new ways to improve their efficiency and effectiveness. To do this, businesses reduced costs of products to maintain their corporate bottom line. As companies began to emerge from this recession they began to focus outward looking for new customers as well as keeping their existing customers. They did this by innovation rather than transformation, introducing new products and services and delivering increased value to customers (Chapter 2, page 30-31). This is an important thought process, now businesses are more customer-centric and with Information and Communication Technology is has never been easier.

In the video, "Supercharge Your Marketing with Google Cloud (Cloud Next '19 UK)", many of the speaker's businesses have shifted to this approach of providing new products and services that are special to the customer. The first speaker says that enterprises need to be customer-centric to grow and bring in more revenue. The third speaker says that their business is growing and to maintain this growth they need to do things smarter. One way they maintain this growth is by improving their marketing. By marketing she means the customer journey and the workload efficiency. They started by collecting data, such as customer information and order information and online behavioral data. With this data they made prediction models pertaining to their customers. With this collected data and prediction models they activated this data for the customer to see, such as recommended products on a webpage. The last two speakers as mentioned previously used Google Cloud to make their data specific to a certain customer. For example, if a customer was vegan they would change what their site viewed to the customer to not advertise meat, but instead advertise vegan friendly things. In conclusion, Businesses are moving to a customer-centric focus and changing how they deliver their products through the use Information and Communication Technology (Chapter 2, page 37).

Information and Communication Technology have accelerated the pace of Big Data adoption through the use of data analytics and data science, digitization, affordable technology and commodity hardware, social media, hyper-connected communities and devices, and cloud computing (Chapter 2, page 37). With the shift to technology that is affordable, it saves both cost and time for businesses and evens out the playing field. Businesses can no longer outspend smaller businesses and technology no longer delivers a competitive advantage (Chapter 2, page 38). In the videos, "Google Cloud Next Madrid '17- Robert Saxby: 'Introduction to Big Data'" and "Supercharge Your Marketing with Google Cloud (Cloud Next '19 UK)", they both explain the use of Google Cloud, which has a free version available. Google Cloud is an example of technology that uses Big Data. It has made life easier and reduced the cost and time for businesses to analyze, process, and store large amounts of Big Data.

Social media has forced businesses to consider customer feedback, therefore increasing the involvement of customers in businesses and adopting this customer-centric approach. Social media has enabling targeting marketing and increased the accuracy and speed of data as well as forcing businesses to listen to the voice of the customer by providing new services and products (Chapter 2, page 39). In the video, "Secret Life of Big Data | Intel", surveys were used to collect data. Social media acts in the same way, collecting useful information and data as mentioned before by customer feedback. Google cloud also collects customer data so that

businesses can give customers what they want and information the pertains to that specific customer.

The increasing coverage of the internet and number of devices people can access the internet on has enabled more people to be continuously active in virtual communities and businesses (Chapter 2, page 40). In the videos, “Google Cloud Next Madrid '17- Robert Saxby: ‘Introduction to Big Data’” and “Supercharge Your Marketing with Google Cloud (Cloud Next ’19 UK)”, the use of Google Cloud allows businesses to collect real time data from customer interaction and provide customers with the best relevant data in the form of products and services.

Cloud computing is a very important in the Big Data field. Cloud computing advancements have led to the creation of environments that are highly scalable, on-demand IT resources that can be leased via pay-as-you-go models (Chapter 2, page 40). Cloud computing is what Google Cloud is all about. In the videos, “Google Cloud Next Madrid '17- Robert Saxby: ‘Introduction to Big Data’” and “Supercharge Your Marketing with Google Cloud (Cloud Next ’19 UK)”, they discuss the use of google cloud as a cloud management software. They talk about numerous benefits in both videos.

Google Cloud solves the complex analytics challenges by allowing businesses to analyze complex data at scale, bring together all of a businesses marketing and customer data to one singular place, real time and up to date analytics for businesses teams, leverage Google’s ML right in BigQuery, and Connects to the marketing tools businesses use. It gives your business user self-service insights when and where they need it without having to wait long periods of time to get it from somebody else. Google Cloud allows scaling out when you need it for a bigger data project and offers this on-demand analysis via its self-service insights as described in Chapter 2 of the textbook.

In chapter 3 of the textbook, Big Data Adoption and Planning Considerations are analyzed. With adoption of Big Data a businesses need to be organized, be able to hold a lot of data and keeping that data private and secure, process that data, offer real time support, conquer performance challenges, pass government requirements, have a methodology and clouds. Google Cloud solves these challenges with the adoption of Big Data. It allows you to put all your information and data into one place, can be scaled to hold a lot of data, has quick processing of data when the data event occurred, allows for changes and fixes to data once it has been into and out of Big Data, and fits the definition of a cloud. As mentioned before, Chapter 3 describes the Big Data Analytics lifecycle. To address the adoption and planning considerations above this step-by-step methodology is needed to organize the activities and tasks involved in analyzing, processing, and storing data (Chapter 3, page 55). Google Cloud helps businesses organize their work and helps them through this step-by-step process.