Lars-Håvard Bråten
Ilona Podliashanyk 30/08-2021

Task a:
The three V's of Big Data are volume, variety and velocity. In essence, these are the three keywords which best describe what big data is. Volume defines data size, velocity defines the speed of the incoming data, whether it's streamed, real time or sent in batches. Variety defines the different varieties and formats of data, such as structured and unstructured data.

Volume
Volume defines data size, and can be expressed via terabytes, amount of records/transactions/files etc. Data volume in the IT service requests management system at Florida State University is not explicitly stated in the article. Nevertheless, one can conclude that it is a considerable load for the organization. In a course of 5 years (2011-2016), the organization has received more than 100,000 service requests. Furthermore, each such request is followed by queue management as well as processing (execution) logs. Both requests, queue management and logs result in an extensive amount of database records, transactions, files, free-text records etc.

Velocity
Velocity defines the speed of the incoming data, whether it's streamed, real time or sent in batches. Again, velocity of incoming requests was not described explicitly in the article, but it has characteristics of batch processing. For instance, requests associated with email forwarding, phone services, registering a new device, acquiring access cards etc, are not necessarily executed after the influx. Requests falling under the category of software system malfunctioning/downtime could supposedly be processed near-time (short time after the incoming request). What is prominent of the requests' velocity in the organization, is their changeability, meaning that request's topic prevalence at a given time is affected by time of the year, software updates, platform migrations etc. As a consequence, deriving predictions on request prevalence upcoming week/month by analyzing requests' velocity, can facilitate shorter execution time and more profound experience gains in the organization.

Variety
Variety defines the different varieties and formats of data, as well as data's sources. The organization is likely to exploit both structured, semi-structured and unstructured formats. Sources of data accompanying the requests are both external (provided by the user) and internal (requests queuing, execution logs).
Request's metadata (when, who, where, which reporting channel, manual classification) could be structured and processed in a relational database, or it could be semi-structured and processed in a JSON-format.
Request's execution updates and descriptions provided by the user, are unstructured since they are provided in a form of free text. Analysis of the unstructured data is the focus of the article, as it is considerably more challenging and time consuming to process.

Lars-Håvard Bråten
Ilona Podliashanyk 30/08-2021

Task b:

i) To draw insight from data

One example is the ITS department of Florida State University.
They drew insight from the information they gathered from topic modeling the contents of their incoming service requests, and could, based on their information, catch new trends and changes to both the volume and contents of the service requests (topic prevalence), and put them in relation to other events. As an example, they noticed that the mentions of the topic “E-mail forwarding” suddenly rose, and they could even explain it by an E-mail platform migration happening in the same time period as the number of mentions went up. The same trend was observed with other themes as well, such as the topic “phone services" when the university migrated regular phones over to VOIP.

ii) To make better decisions based on the insight

The ITS department of FSU used the insight they gained from the topic modeling to identify trending topics and how they changed over time. For instance, if there was one topic which stood out more than others, more resources could be put into optimizing that specific service. Or, said in another way, it allowed the ITS to focus their resources on the most pressing matters simply by analyzing incoming requests.

iii) To automate the decision-making and bake it into a business process

An example of this would be the way Iventx optimized their IT services so that the system itself could understand the issue at hand, and even offer a solution. This could be as simple as a password change, but also maybe linking to a wiki page about adding printers if the user had problems connecting to one. The system was capable of solving the issue on its own in 60% of the service requests.

Task c:

- **Dividing all the service tickets into two groups: open and closed requests.** This type of data is well suited to be stored as an enum attribute in a relational database, and can effectively be retrieved with a query.
- **Computing the total number of requests issued in the past week.** This type of data is well suited to be stored as an attribute in a relational database, and can effectively be retrieved with a query. Just count the database entries of issues solved in the last 7 days (from- and to date).
- **Group more than 100,000 service tickets into topics (printer, email forwarding, etc…)** This is a text-data mining task, namely the one that was executed in the IT service requests management system at Florida State University. It also seems more effective to make a program read the contents of the tickets and place them into a category based on the results, than processing data from the relational database.
- **A “recommender” system that suggests the most appropriate expert for solving a problem and the possible solutions based on how requests**

were handled in the past. By both data mining and machine learning. Firstly one could analyze and classify the incoming requests by the best fitting expert (training set), using data mining. Training set can then be used on a machine learning algorithm.