

Teaching Case

Who Renews? Who Leaves? Identifying Customer Churn in a Telecom Company Using Big Data Techniques

Daniel A. Asamoah

Information Systems and Supply Chain Management Department
Raj Soin College of Business
Wright State University
Dayton, OH 45431, USA
daniel.asamoah@wright.edu

Ramesh Sharda

Pankush Kalgotra

Management Science and Information Systems Department
William Spears School of Business
Oklahoma State University
Stillwater, OK 74074, USA
ramesh.sharda@wright.edu, pankush@okstate.edu

Mark Ott

Teradata Corporation
Dayton, OH 45342, USA
mark.ott@teradata.com

ABSTRACT

Within the context of the telecom industry, this teaching case is an active learning analytics exercise to help students build hands-on expertise on how to utilize Big Data to solve a business problem. Particularly, the case utilizes an analytics method to help develop a customer retention strategy to mitigate against an increasing customer churn problem in a telecom company. Traditionally, the forecast of customer churn uses various demographic and cell phone usage data. Big Data techniques permit a much finer granularity in the prediction of churn by analyzing specific activities a customer undertakes before churning. The authors help students to understand how data from customer interactions with the company through multiple channels can be combined to create a “session.” Subsequently, the authors demonstrate the use of effective visualization to identify the most relevant paths to customer churn. The Teradata Aster Big Data platform is used in developing this case study.

Keywords: Big data, Data analytics, Telecommunication, Active learning, Business intelligence, User satisfaction

1. INTRODUCTION

The purpose of this teaching case is to introduce students to some of the ways in which big data analytics can be utilized to make business decisions. In this paper, we present big data as data that is defined by multiple characteristics including volume, velocity, variety, and veracity (Chen, Chiang, and Storey, 2012). Whereas some data utilized for big data applications may possess all the above mentioned

characteristics, others may possess only some of the characteristics mentioned above, with volume occurring frequently as a dominant characteristic. Specifically, we focus on the volume and variety components of the definition of big data and show how to leverage those characteristics to build business-driven, big data applications. The case exposes students to the use of big data analytics to reduce customer churn in the telecom industry. Churning occurs when a business loses a customer to another

competitor, leading to loss of sales (Adwan et al., 2014). For privacy's sake, the authors use the pseudonym New Horizon Telecom (NHT) for the telecom company on the basis of which this case was developed. This written case is presented as a non project-based exercise to facilitate a class discussion about the business application of big data analytics (Cappel and Schwager, 2002).

By following one of the most prominent Business Intelligence (BI) model curriculum guidelines (Gupta, Goul, and Dinter, 2015), we focus on designing a teaching case that supports the teaching of a Level 2 Management Information Systems (MIS) course as designed in the *MSIS 2006 Model Curriculum* (Gorgone et al., 2006). In the *MSIS 2006 Model Curriculum*, Level 2 courses are those that focus on sophisticated and emerging topics such as Big Data. Whereas previous efforts to teaching BI in the Business School has focused on the design of comprehensive analytics programs, this teaching case supports Asamoah et al.'s (in press) effort in designing content at the micro level where specific courses are utilized to teach BI.

In subsequent sections of this paper, we first situate the teaching case within a telecommunications background and present the use of big data analytics to solve a business problem. Next, we present a case story in which the authors first explain the challenges NHT was facing in terms of retaining its core customer base and its motivation for analyzing customer churn as part of its business operations. The case text also presents the opportunities that the vast amount of NHT's customer data presented in their quest to utilize customer communication data to reduce churn. This involved an exposition of NHT's data sources, data management, and analysis. The case text also describes how and why NHT sessionized the data from multiple channels and identified potential churn based on customer activities. A Big Data assignment derived from the case is presented in the subsequent section. A feedback survey of a pre-test of the assignment in an actual Big Data course is also presented. In the conclusion, we discuss how structural changes were made in NHT based on the results of the churn analysis. The objectives of the case are for students to:

1. Understand the technical and operational environment for a Big Data-driven approach to solving business problems.
2. Use an analytics approach to analyze customer interactions data and reduce customer churn in a business organization such as a telecommunications company.
3. Study and understand how to utilize a Big Data Analytics tool, using Teradata Aster as a sample platform.

2. BACKGROUND

NHT is a nationwide telecommunications service provider that offers fourth generation mobile cellular services to its customers. NHT operates in a highly volatile telecommunication industry where customer churn is a common problem. For instance, worldwide, the rate of customer churn in the telecom service industry ranges from 20% to 40% per year (Ahn, Han, and Lee, 2006). For prepaid services, which are common in emerging markets, churn

rates are as high as 70% per year (De, 2014). In most countries, regulations, such as those that govern number portability, facilitate easy customer churn and make for keen competition in customer retention. This teaching case describes the use and implementation of Big Data Analytics (BDA) techniques to solve a business problem related to customer churn in a telecommunications service company such as NHT.

The authors show how NHT utilized BDA to manage and analyze very large amounts of telecom customer data from multiple sources and in multiple formats to provide deeper insights into how a telecom business can reduce churn. In this teaching case, the authors apply and demonstrate customer churn analysis on this high-volume, multi-structured data. Specifically, we present the use of a comprehensive Big Data management framework, including an apache Hadoop platform, an SQL-like data query platform, and a high availability data storage platform.

3. CASE TEXT

3.1 Overview of NHT's Sales Problem

The Chief Executive Officer (CEO) at NHT called for an immediate meeting among the C-level executives on the Monday after the release of the company's 1st quarter financial report in 2012. The CEO was particularly concerned that sales had dropped drastically in the previous quarter. The company had rolled out an aggressive marketing campaign at the beginning of the year with the expectation that sales would increase beginning in the first quarter. The CEO began by trying to seek answers from the Chief Marketing Officer (CMO) about the current state of affairs. Although the CMO maintained that the marketing campaign had been effective, he accepted that it may be interesting to specifically identify what contributed to the low sales. He sought the involvement of the Chief Information Officer (CIO) in determining the common behaviors and actions of customers prior to service cancellation. Service cancellation had been identified by the customer relation management (CRM) department as a key contributor to low sales in the first quarter. The CFO's report had already indicated that:

1. The churn activities led to an overall decrease in the customer core base which resulted in decreased overall revenue.
2. As churning occurred, longtime and loyal customers were lost. New customers mostly started with basic services. Add-on services, which create more revenue, took time to be adopted by these customers, resulting in slow growth of company revenue.

As an action item for a follow-up meeting, a decision was made by management to identify the triggers that led to customer churn and to find a lasting solution to this problem. It was quickly identified that these triggers were numerous as is typical in such a turbulent telecommunication market. They also identified that not all the triggers lent themselves to immediate and actionable remedial strategies since some were external and therefore outside the control of the organization. Some factors that led to these triggers were cost of switching, service usage, and length of service. Several factors were also related to customer service and

relationship activities. In a follow-up meeting, the CIO suggested that as an immediate measure to boost sales in the next quarter, they should utilize internal factors to identify and predict service cancellation based on the history of a customer's communication with the company. This would enable NHT to design effective strategies to curb the customer churn.

The CIO further argued that the rich sets of communication between the customers and CRM department could help identify behavioral patterns that led to churn. The CRM department communicated with the customer through multiple channels and in different formats. These multiple means of communication presented data management challenges to the churn analysis phase. The challenges posed are discussed in the next section.

The executives agreed that exploring churn analysis from the perspective of the customer relations department was key to reducing cancellations. This was because many of the cancellations involved a series of back-and-forth communications with the customer service department. Hence, the NHT executives set up a task force that reported directly to the CIO to look further into this problem and devise a means of tackling it. The task force was made up of members from the customer relations management (CRM), the marketing, and the information technology (IT) departments. Initially, the team explored multiple ways to study the churn problem. Various methods were proposed to prevent customer churn and boost revenue. These solutions ranged from creating reactive market strategies to reduce churn to building decision support tools that provide real-time ability to generate insights about churn trends. The task force also considered contemporary approaches, such as the use of social network theory as a means of studying churn analysis as was used in Dasgupta et al. (2008). Such approaches utilize the strength of ties and other communication patterns between individuals as a way of analyzing churn activities among telecom customers.

Even though such methods could have been used, the task force members recognized that the exponential increase in the amount and variety of data they are collecting gave them an opportunity to go far beyond current analytics techniques. For example, NHT has multiple sources of data about customer interactions. When customers have an issue about a bill, payment, call quality, or the device itself, they can interact with the company in many different ways: visiting a physical store or a company website or calling the company call center. Each channel creates its own records of customer interactions. Thus, a lot of data is collected about customer interactions. However, given the different formats and data stores in which these interactions are recorded, a single view of the interactions by the same customer may not be visible. Thus, not much analysis is done to identify the common sequences of customer attempts to resolve their issues before cancelling their services. The task force realized that existing predictive analytics approaches have not been able to utilize all these sources and volume of data in a holistic manner to generate actionable insights that are both efficient and cost-effective.

3.2 The Challenge of Data

The CRM department at NHT had customer communication records for each of three different channels of communication: the website, physical stores, and call centers. The challenges posed by the data structure to the analysis of service cancellation were categorized into three areas: data from multiple sources, data variety, and data volume.

1. **Data from Multiple Sources:** As noted above, a customer could connect with NHT by accessing their accounts on the company's website, allowing NHT to generate weblog information on customer activity. For instance, weblog tracking allowed the company to know if and when a customer reviewed his current plan, submitted a complaint, or checked his bill online. Customers could also walk into an NHT customer service center to lodge a service complaint, request a plan change, or cancel the service. Lastly, a customer could call the customer service center on the phone and transact business just like he or she would do in person at a customer service center. Any one of these sources can support the analysis of churn trends among customers using that channel. However, only limited insight would be generated and, hence, a remediation strategy would not be comprehensive enough to solve the churn problem. The task force deemed a combination of data from all three sources as the best way of gaining a broad scale and in-depth insights into the churn problem. In effect, whatever methodology was implemented had to leverage customer data from all three channels of data in order to generate rich insights.
2. **Data Variety:** An attendant issue, based on the multiple sources of data, was the variety of data that was generated by each means of customer communication. That is, data derived from online, call center, and walk-in activities were all of different structures. The original data from the three sources were in an unstructured format. For instance, the call center data was originally audio recordings which had to be first transcribed into text before being transformed into a structured format. In order to perform meaningful analysis on data from all sources, the individual data sets had to be converted into similar structured formats.
3. **Data Volume:** The third hurdle that was created by the data was the vast amount of data that had to be analyzed. Previous data analytics projects mostly utilized a small sample set of data for analysis. However, the company wanted to leverage the multiple variety and sources of data to generate as much insights as possible.

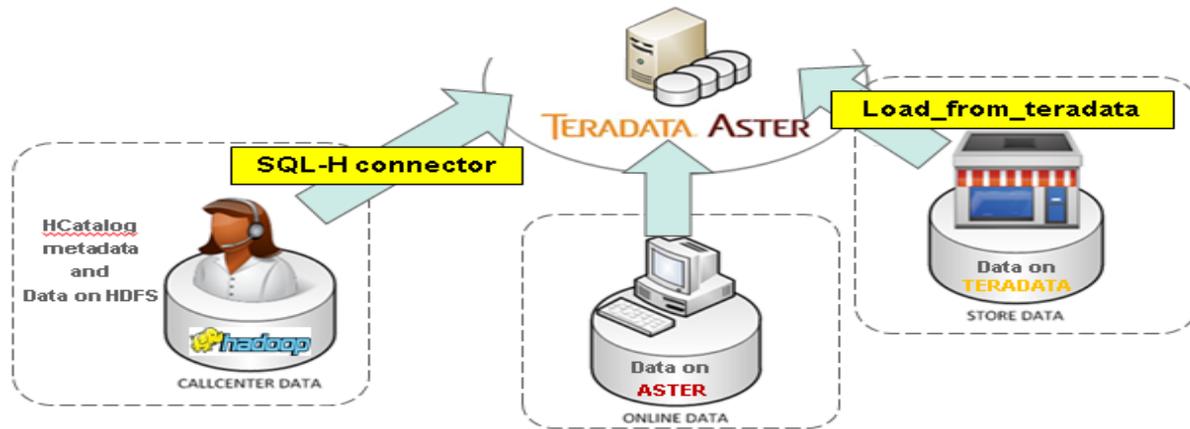


Figure 1. Aster Architecture

As is common in other large scale data analytics projects, traditional platforms for data management are not versatile enough to handle advanced data analysis as is required by the three characteristics of data described above (Thusoo, Shao, and Anthony, 2010). As such, a distributed Big Data framework was used for data storage and analysis. The next section presents the Big Data environment.

3.3 Distributed Big Data Environment

Based on the recommendation of the CIO, the task force utilized Teradata Aster, a unified BDA architecture that can handle large volumes of unstructured data in a distributed environment for data analysis. Teradata Aster consists of components that possess several algorithms, tools, and processes that allow for data cleaning, formatting, transformation, storage, querying, and analysis. Aster has capabilities for developing and running Map-Reduce (MR) jobs (Ghazal et al., 2013). Specifically, Aster has enabled MR analytics functions such as time series analysis, associative mining, and predictive analysis.

A unique capability of Aster is its ability to connect to other databases and data management platforms such as Hadoop and traditional data warehouses such as Teradata data warehouse and Oracle DB. From Teradata Aster, one can connect to and access other databases using specific connectors developed by the Teradata Corporation. Figure 1 shows this same point graphically. Data can be loaded into Aster from multiple sources using various connectors.

Analysis for this particular teaching case can be completed using the express version of Teradata Aster. The express version is free to download and includes all available functionalities. In the express version, the maximum size of the dataset that can be processed is 17 GB. One has to fill a form to begin the download. URL links that provide information on where to download Aster and how to get started are listed in Table A.1 in Appendix A.

4. DATA PREPARATION AND ANALYSIS

4.1 Data

The task force obtained unstructured datasets from the three channel sources. These were converted into structured form using Teradata Aster's data management functions. The

variables in the three structured datasets are listed in Table 1. The variables in each dataset include a unique customer identification, a specific channel of interaction used by the customer, the action performed by the customer, and the date and time of the action.

Variable Name	Description
Customer ID	A unique identification number for each customer.
Channel	The channel used by the customers to interact with the company (online, call center or physical store)
Action	Different activities performed by customers such as bill dispute, view bill, account summary, etc.
Date	The data and time of a particular action.

Table 1. Data Description

Figures 2 through 4 show a sample structure of the call center, online, and physical store tables. The variable, ACTION, determines the customer's activities when interfacing with NHT. The team identified twelve different ACTION types for analysis. Across all three channels of communication, a customer could present a bill dispute, request for plan upgrade, request for plan downgrade, perform profile update, view account summary, make a service complaint, view bill, review contract, access store locator function on the website, access frequently asked questions section on the website, or browse devices. Ultimately, a customer could also cancel all services.

customer_id	channel	action	timestamp
352257	CALLCENTER	BILL DISPUTE	2010-05-22 01:18:51.000000
352257	CALLCENTER	SERVICE COMPLAINT	2010-05-30 22:18:51.000000

Figure 2. Sample Call Center Dataset

customer_id	channel	action	timestamp
408115	ONLINE	BILL DISPUTE	2010-06-11 11:18:39.000000
485592	ONLINE	BILL DISPUTE	2010-06-17 14:41:10.000000

Figure 3. Sample Weblogs Dataset

customer_id	channel	action	timestamp
352257	STORE	NEW ACCOUNT	2010-05-21 07:18:51.000000
352257	STORE	BILL DISPUTE	2010-06-03 03:18:51.000000

Figure 4. Sample Physical Store Dataset

Based on previous actions from all channels of communications, the objective of the task group set was to identify the most common paths that resulted in the final cancellation of services. Three datasets from the three channels of communication are available at: http://tunweb.teradata.ws/tunstudent/dataset/Telco_churn_datasets.zip. The data are in structured form (i.e., comma or tab separated). Each dataset contains approximately one million records.

4.2 Data Preparation

The task force wanted to focus on the sequence of activities that resulted in a final cancellation of services by a customer. In order to do this, multiple approaches could be used to prepare, manage, and present the data in such a way that would allow sequential analyses to be done. In this case, the analytics team joined all three tables so that activities across the channels could be compared. Hence, a 3-way UNION view was created, appending all three datasets.

4.3 Sessionizing

In order to determine if a sequence of actions was related and collectively responsible for a customer’s final churn decision, sequence of activities was classified into sessions. The concept of a session is highly used in web analytics to connect a series of user events online. A session is defined in multiple ways by the researchers in web analytics (Gayo-Avello, 2009; Spiliopoulou et al., 2003). To name a few, a sequence of events or queries are considered as one session if:

1. These occur within one day.
2. No more than *t time* passes between successive events. This *t time* is known as the “period of inactivity” or the “timeout threshold.”
3. These occur once the user starts the browser until he or she quits.
4. The successive events are directly connected. An example would be the Amazon product recommendation system in which the user accesses multiple products, one after the other, by clicking on the direct links.

Whereas the first and second definitions are time-oriented, the third and fourth definitions are navigation-oriented (Spiliopoulou et al., 2003). Generally, for most

sessionizing operations, emphasis is on online activities. However, in NHT’s churn analysis problem, a user can use any channel: online, call-center, or face-to-face store visit to interact with the company. Therefore, NHT’s definition of a session accommodated multiple channels of communication. In this particular context, NHT defined a session as *a sequence of actions through multiple channels, with the timeout threshold of five days, which lead to the cancellation of a service*. Take an example of a user A, who interacted with the company through various channels before cancelling the service. On Monday, the customer made a phone call to register his complaint. The customer was not satisfied with the response and went to a physical store on Wednesday to cancel the service. The customer used different channels, but the team considered this one session because both activities were related to the same issue.

Users’ actions within a five-day period could be grouped into various sessions (with a unique identifier) based on which sequential analysis can be made. A sample of such a sessionized table is shown in figure 5. The first two actions by the same customer were on the same day and therefore have the same sessionid (0). On the other hand, because the third action was performed after more than five days, the session id is different i.e. 1.

customer_id	channel	action	timestamp	sessionid
351218	CALLCENTER	SERVICE COMPLAINT	2010-05-04 07:45:55.000000	0
351218	ONLINE	BILL DISPUTE	2010-05-04 07:49:13.000000	0
351218	CALLCENTER	SERVICE COMPLAINT	2010-05-12 03:56:10.000000	1

Figure 5. Sample Output from Creating Sessions

4.4 Identifying Churn (Aster nPath analysis)

In order to analyze actions that lead to service cancellation, the nPath function (Nejmeh, 1988) was used. The nPath function is operationalized in Aster’s SQL-MapReduce programming framework. Aster’s SQL-MapReduce framework enables the use of standard SQL to write very powerful functions in other programming languages such as Java and Python (Teradata Corporation, 2016). Aster, therefore, provides SQL capabilities to nPath (Woo, 2013) such that by use of regular expression, the nPath function performs regular pattern matches in sequences of rows. The pattern-matching step should identify instances where customers had actions other than service cancellation followed by a cancelling of the service within the same session. For instance, an event where a customer starts by viewing a bill online, then lodges a service complaint, and thereafter cancels a service should be identified and classified as a churn event. On the contrary, if a customer views a bill online and later enters a customer service center to lodge a complaint, such an activity should not be flagged as a churn event.

4.5 Focus on Key Paths (Golden Path Iteration Process)

The team realized that preliminary analysis could produce thousands of potential routes that a customer may take to cancel service. Targeting all the individual routes and designing strategies to curb the upward trend of customer

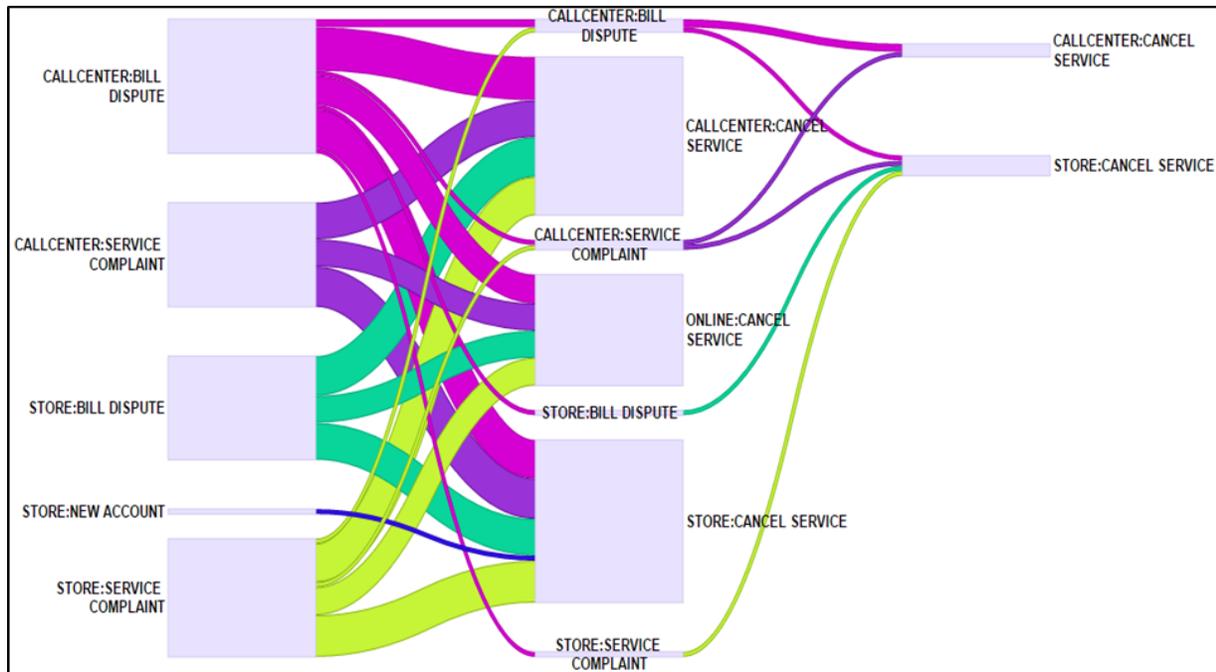


Figure 6. Visualization Sample of Top 20 Paths Using AppCenter

churn is impracticable for both short and long terms goals. This led to the next set of analysis, which was to zoom into the preliminary analysis and help create actionable insights. This set of analysis was intended to be more actionable for the CEO and other upper-level management.

For the next phase, the team particularly identified what details to focus on as far as generating deeper insights into the preliminary results was concerned. This involved several iterations in order to identify better insights. The first consideration was to identify the most occurring routes of interest (also termed as the golden path) and design necessary business strategies to curb churning. In zooming into actionable insights, the team could focus on the top 20 frequently occurring paths spanning all the channels. For instance, Figure 6 shows that cancellations can either be effected at the store or the call center. The actions on the left indicate the common path followed by customers who would eventually cancel their service.

AppCenter, a highly interactive web-based visualization tool created by the Teradata Aster, was used to visualize and focus on the 20 most occurring paths. After this phase, the team presented the model to the CIO. He found the results to be very good. At the next meeting with the other company executives, the CIO presented the model, based on which a decision was made to test it during the third quarter of the year. In the company's end-of-year report, they realized that, indeed, the trend of customer churn had decreased beginning in the third quarter of the year. This trend also helped boost sales within the same period.

5. CASE STUDY DISCUSSION QUESTIONS

In order for students to fully appreciate the basis of the business case and the role of analytics in solving the related problems, students should identify specific reasons and

situations that may limit the type of approach they would propose. Some open-ended sample questions that could guide group discussions are listed below.

1. What unique problems did customer churn pose to NHT's business operations?
2. Explain how the characteristics of NHT's data posed as a challenge to existing methods of data analysis.
3. In creating sessions, why did NHT adopt a time-oriented approach?
4. Identify other approaches for sessionizing a customer's trend of communication with NHT.
5. What do you think a customer can do before cancelling the service?

6. HANDS-ON TELECOM CHURN EXERCISE

In this section, we demonstrate the process a student should follow to complete the assignment as performed by the task force in NHT. The datasets used are subsets of the big datasets compiled by NHT. We elected to use subsets of the data in describing this case, rather than the whole dataset, so that students would be able to smoothly run the analysis using the express (virtual image) version of the Teradata Aster platform. Whereas relatively small sizes of datasets are used in this case, students are introduced to hands-on big data analytics concepts and experience by following the steps described and answering specific related questions. This exercise will expose students to a business problem in the telecom domain. It will further introduce students to an analytics approach to solving the business problem using the Teradata Aster platform, nPath analysis, and the AppCenter visualization tool. The following steps are required to complete the assignment:

1. Download the express (virtual image) version of the Teradata Aster from the link below. <https://aster-community.teradata.com/community/download>. One has to complete a form to begin the download.
2. To setup Teradata Aster, follow instructions at the link: <https://goo.gl/Dy62U4>
3. The three datasets containing data from multiple channels of communications can be found on the link http://tunweb.teradata.ws/tunstudent/dataset/Telco_churn_datasets.zip
4. Upload the datasets in the Aster database to run queries. This can be done using an Aster function called `ncluster_loader`. The instructions to use it can be found on the instructions guide provided in step 2.
5. Combine all three datasets using a simple SQL Join query.
6. Sessionize the result from the step 5 using a time threshold of five days, as done by the NHT task force. To do so, an Aster function called "sessionize" could be used.
7. After sessionizing the data, run `nPath` function to find the sequences of types of questions asked and the contact channel used by customers.
8. Finally, create visualizations using AppCenter and use them to answer questions 1 to 5 below.

The steps are presented in Figure 7. Figure 7 also shows the number of records in each dataset and at each step.

Based on these steps and analysis, students should be able to answer the following questions:

- Question 1: How many different paths or sequences ended with the cancellation of service?
- Question 2: Provide the top 10 sequences of actions that started with SERVICE COMPLAINT as an action at any channel and ended with cancellation of a service. Provide an AppCenter visualization for the top 10 paths.

Question 3: Provide an AppCenter visualization of the top ten paths found in question 2.

Question 4: Based on your analysis, suggest at least one thing the telecom company can do to avoid churn in terms of handling customer complaints. (An open-ended question).

Question 5: What were the two subsequent common actions performed by the customers after they performed an action of "BILL DISPUTE" through the "Call Center"?

Question 6: How many customers performed the specific sequence of actions given below before the cancellation? These actions can be performed through any channel.

- i) REVIEW CONTRACT
- ii) BILL DISPUTE
- iii) BILL DISPUTE
- iv) SERVICE COMPLAINT

7. TELCO ASSIGNMENT SUMMARY FEEDBACK

To test the effectiveness of this Telco Churn case on its ability to help students utilize the capability of BDA to solve a business problem, the authors distributed a short survey (Table B.1 in Appendix B) to the class of 20 students during the following class meeting. Most students found the purpose of the assignment very useful. Some students commented that the assignment not only trained them on how to perform path analysis, but also taught them to think about a business problem and design a solution to it from a business intelligence perspective. One student commented, "This [exercise] gave me a better picture of how Big Data Analytics would help in identifying problems and making decisions in business."

The assignment was simple to understand, which helped students focus on understanding the concepts rather than tussling with the technical platform. The students found the exercise very interesting and beneficial and recommended

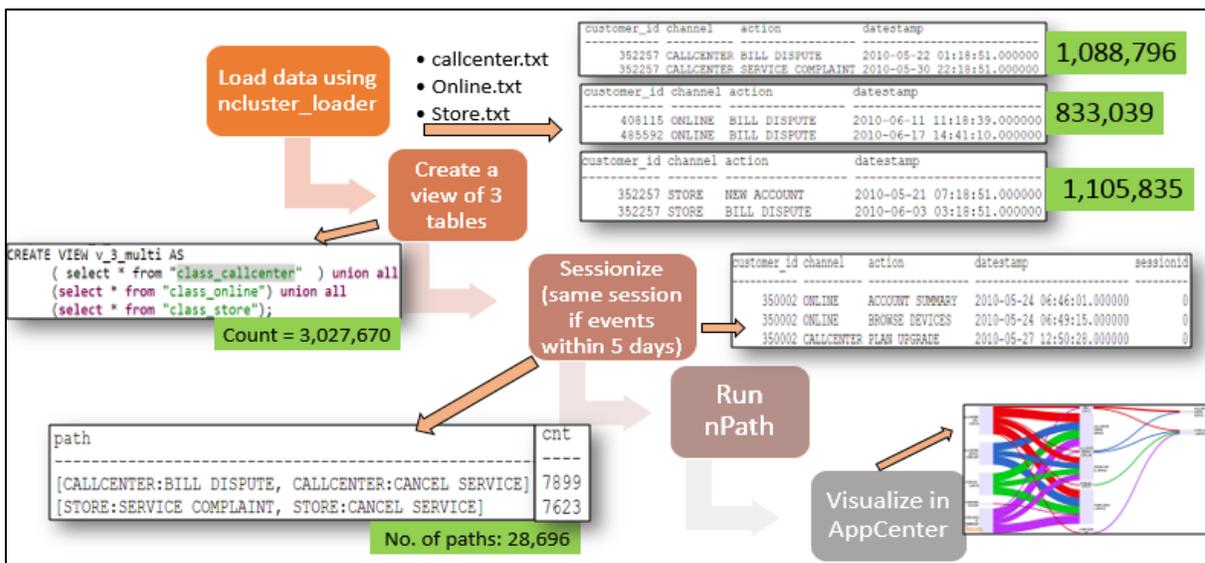


Figure 7. Hands-on Exercise Steps

having more exercises on path analysis. The average time of the assignment was two hours. However, some students took a little bit longer to complete given that this was their first introduction to a big data analytics platform. Based on the survey instrument, we realize that the use of data from multiple sources could pose a challenge for students. It may be worthwhile to dedicate a preceding class time to get students exposed to the concept of data management on the Teradata Aster Big Data platform. This prior session could encourage students to practice with sample snippets of data with multiple structures and from different sources. The guides to learn and use Teradata Aster and SQL-MR are listed in the Appendix A. The survey we performed, although limited, helps to indicate the general trend of the effectiveness of the case study in teaching a BDA course in a business setting.

8. CONCLUSION

This paper presents a case study that focused on the use of BDA to solve a business problem. The case study can be used in teaching a Business Intelligence and Analytics (BI&A) course. This case has been implemented in teaching a graduate level BDA course offered in a Management Information Systems program.

The case is based on an actual consultancy work for a telecommunications service company that wanted to reverse the tide of increasing customer churn. A key structural change was made at NHT to address how CRM associates responded to customer complaints. For instance, the resultant decision support system that was built enabled associates to determine if a customer who was being attended to had the profile of a potential customer who would churn. Based on this information, a decision can be made by the associate to offer deals or bonuses that would help incentivize the customer to maintain his service with NHT. This is in addition to a more proactive attitude by the associate to swiftly resolve whatever issue the customer may have. By so doing, NHT improved its sales over the remainder of the year.

The feedback received from the students through a survey indicated that the students generally found the active learning exercise to be informative in terms of helping them build their skills in big data analytics. The students also found the assignment to be easy to understand and execute.

As is common with teaching cases, the level of realism in its presentation may have an impact on how well students can imbibe and implement the intended learning objectives. Given that this case is based on a real world scenario, we hope that this potential limitation is diminished. We also anticipate that the case can be used in other domain areas besides the Information Systems (IS) discipline.

9. REFERENCES

Adwan, O., Faris, H., Jaradat, K., Harfoushi, O., & Ghatashah, N. (2014). Predicting Customer Churn in Telecom Industry Using Multilayer Preceptron Neural Networks: Modeling and Analysis. *Life Science Journal*, 11(3), 75–81.

- Ahn, J.-H., Han, S.-P., & Lee, Y.-S. (2006). Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry. *Telecommunications Policy*, 30(10), 552–568.
- Asamoah, D. A., Sharda, R., Zadeh, A. H., & Kalgotra, P. (in-press). Preparing a Data Scientist: A Pedagogic Experience in Designing a Big Data Analytics Course. *Decision Sciences Journal of Innovative Education*.
- Cappel, J. J. & Schwager, P. H. (2002). Writing IS Teaching Cases: Guidelines for JISE Submission. *Journal of Information Systems Education*, 13(4), 287–293.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
- Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjee, S., Nanavati, A. A., & Joshi, A. (2008). Social Ties and their Relevance to Churn in Mobile Telecom Networks. In *Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology* (668–677). Nantes, France.
- De, A. (2014). Revenue Enhancement and Churn Prevention for Telecom Service Providers. Retrieved from <http://www.wipro.com/documents/revenue-enhancement-and-churn-prevention-for-telecom-service-providers.pdf>
- Gayo-Avello, D. (2009). A Survey on Session Detection Methods in Query Logs and a Proposal for Future Evaluation. *Information Sciences*, 179(12), 1822–1843.
- Ghazal, A., Rabl, T., Hu, M., Raab, F., Poess, M., Crolotte, A., & Jacobsen, H.-A. (2013). BigBench: Towards an Industry Standard Benchmark for Big Data Analytics. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (1197-1208). New York, NY.
- Gorgone, J. T., Gray, P., Stohr, E. A., Valacich, J. S., & Wigand, R. T. (2006). MSIS 2006: Model Curriculum and Guidelines for Graduate Degree Programs in Information Systems. *Communications of the Association for Information Systems*, 17, Article1.
- Gupta, B., Goul, M., & Dinter, B. (2015). Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs. *Communications of the Association for Information Systems*, 36(23), 449–476.
- Nejmeh, B. A. (1988). NPATH: A Measure of Execution Path Complexity and its Applications. *Communications of the ACM*, 31(2), 188–200.
- Spiliopoulou, M., Mobasher, B., Berendt, B., & Nakagawa, M. (2003). A Framework for the Evaluation of Session Reconstruction Heuristics in Web-Usage Analysis. *Inforns Journal on Computing*, 15(2), 171–190.
- Teradata Corporation. (2016). MapReduce and Teradata Aster SQL-MapReduce. Retrieved March 21, 2016, from <http://www.teradata.com/Teradata-Aster-SQL-MapReduce/>.
- Thusoo, A., Shao, Z., & Anthony, S. (2010). Data Warehousing and Analytics Infrastructure at Facebook. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (1013–1020). Indianapolis, IN.

Woo, J. (2013). Market Basket Analysis Algorithms with MapReduce. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(6), 445–452.

AUTHOR BIOGRAPHIES

Daniel Adomako Asamoah is an Assistant Professor in Management Information Systems in the Raj Sooin School of Business at the Wright State University. He received his Ph.D. in Management Information Systems from the Spears School of Business at the Oklahoma State University.



His research focuses on business analytics and intelligence, big data applications, decision support systems in health care and electronic social media. His research has been published in multiple journals, including *Communications of the Association of Information Systems*, *Decision Support Systems*, *Journal of Computing and Information Technology* and *Simulation*. Major conferences such as the annual meetings of the Americas Conference on Information Systems (AMCIS), the Decision Sciences Institute (DSI), the Institute for Operations Research and the Management Sciences Healthcare (INFORMS Healthcare) and the Pre-ICIS Business Analytics Congress have accepted his papers. He has won multiple awards including the prestigious Phoenix Doctoral Award and the Philips Dissertation Fellowship Award. As a principal investigator, his research has been funded by the Wright State University Research Council. He is also an editorial board member of the *International Journal of Experimental Algorithms (IJE)*.

Ramesh Sharda is the Vice Dean of the Watson Graduate School of Management, Watson/ConocoPhillips Chair and a Regents Professor of Management Science and Information Systems in the Spears School of Business at the Oklahoma State University. He has co-authored two textbooks



(Business Intelligence and Analytics: Systems for Decision Support, 10th edition, Pearson and Business Intelligence, Analytics, and Data Science: A Managerial Perspective on Analytics, 4th Edition, Pearson). His research has been published in major journals in management science and information systems including *Management Science*, *Operations Research*, *Information Systems Research*, *Decision Support Systems*, *Interfaces*, *INFORMS Journal on Computing*, and many others. He is a member of the editorial boards of journals such as the *Decision Support Systems* and *Decision Sciences Journal*. He is currently serving as the Executive Director of Teradata

University Network and received the 2013 INFORMS HG Computing Society Lifetime Service Award.

Pankush Kalgotra is a doctoral candidate majoring in Management Science and Information Systems at the Oklahoma State University (OSU). His research interests include healthcare analytics, network science, dark side of IT and neuroimaging in Information Systems. He has more than five years of experience with Data Mining, Texting Mining,



Sentiment Analysis and Big Data Analytics. He is proficient in using and teaching Teradata Aster, a Big Data Platform. He is a SAS® certified Predictive Modeler and was awarded with the SAS Student Scholar Award in 2013 and the SAS Student Ambassador Award in 2014. His team won the SAS Shootout competition in 2014. For his teaching effectiveness, he received the Spears School of Business Outstanding Graduate Teaching Associate Award in 2015. He was also awarded the Distinguished Graduate Fellowship in 2015 and 2016 at OSU.

Mark Ott is a 30-year veteran of Teradata Corporation, starting off as Microsoft/Cisco instructor. In the last 5 years he has focused on Aster and Hadoop technologies. Previously, he was a publishing editor for Windows IT magazine and holds certifications in Microsoft, Cisco, Teradata and Hortonworks products. He travels the world teaching Big Data



Analytics to both Teradata customers and internal employees.

APPENDIX A: TEACHING RESOURCES

Resource Name	Resource Location
Teradata Aster Express Download Link	https://aster-community.teradata.com/community/download
Teradata Aster Page on Teradata University Network	http://www.teradatauniversitynetwork.com/Software/Downloadable/Teradata-Aster/ *Create a free login to access the page
Getting Started Guide	https://goo.gl/Dy62U4
Guide to learn SQL-MR and Analytics Guide	https://goo.gl/fBFCIC
Regular Expressions	http://www.regular-expressions.info/tutorial.html

Table A.1. Links to Teaching Resources

APPENDIX B: POST-ASSIGNMENT SURVEY

<i>Please rate the following statements on a scale of 1 to 5, where 1 = disagree; 2 = somewhat disagree; 3 = neutral; 4 = somewhat agree; 5 = agree.</i>		
Statement	Average Rating	Standard Deviation
The Teradata Aster was easy to use.	4.6	0.80
The Teradata Aster was easy to understand.	4.6	0.70
The Teradata Aster was easy to run SQL and SQL-MapReduce (SQL-MR) queries.	4.3	0.75
The Teradata Aster SQL-MR query was easy to use to sessionize the time-stamped events.	4.5	0.80
The nPath analysis helped me to better understand the actions performed by the customers before cancelling their services.	4.3	0.97
The nPath analysis helped me to generate interesting insights that cannot be gained through traditional data analytics.	4.4	0.99
This exercise helped me better understand the concept of linking multiple events (weblogs, call-center and physical store events) into a session.	4.3	1.07
The Telco Churn exercise helped me to better understand the sequence of actions performed by the customers.	4.5	0.90
The Telco Churn exercise provided me with experience for managing the data collected from multiple sources.	3.6	1.44
This exercise gave me a good awareness of the potential of big data analytics technologies.	4.4	0.79

Table B.1. Post-Assignment Survey Instrument



No matter how sophisticated the technology, it still takes people!™



STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

Copyright ©2016 by the Education Special Interest Group (EDSIG) of the Association of Information Technology Professionals. Permission to make digital or hard copies of all or part of this journal for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial use. All copies must bear this notice and full citation. Permission from the Editor is required to post to servers, redistribute to lists, or utilize in a for-profit or commercial use. Permission requests should be sent to Dr. Lee Freeman, Editor-in-Chief, Journal of Information Systems Education, 19000 Hubbard Drive, College of Business, University of Michigan-Dearborn, Dearborn, MI 48128.

ISSN 1055-3096