The Opportunity for Data Wrangling in Telecommunications
Big Data: New Driver of Growth in Telecommunications

In today’s globalized markets, quality of service and overall organizational productivity play critical roles in the growth and competitiveness of any business. Business success is dependent upon organizations staying well tuned to the needs of their customers. Remaining focused on the unique needs and behaviors of individuals and households is critical to survival in telecommunications given the industry’s maturity and fiercely competitive nature, as well as the largely commoditized nature of the industry’s core service offerings. Amidst the rapid global expansion of mobile technology, Telecommunications organizations face unprecedented competitive pressure. Prepaid and short-term contracts for cell phone users worldwide, along with free Internet phone services, have become mainstays in the industry, creating an even more challenging environment for Communications Service Providers (CSP), given the growing breadth of communications options available to customers that may entice churn.

However, one company’s churn is another company’s opportunity for growth and innovation. The savvy use of big data is helping industry leaders overcome their customer retention challenges by helping to create personalized engagement strategies that work in real time, and it is creating new business models that tap into growing markets where data is becoming the actual product.

CSP’s have been engaged in data analytics for decades utilizing company-wide data warehouse and business intelligence (BI) environments. Some have created data marts and applied data mining techniques for customer experience management. Over time, however, these legacy CSP data architectures have not been able to adjust to the analytics requirements of today, such as the need to incorporate semi-structured and unstructured data sources and to cope with growing volumes of information generated within the organization and from 3rd party sources. The demand for immediate results, a higher degree of automation and the need to provide the highest quality of services enabled through extreme personalization, is driving CSPs to invest heavily in big data tools to develop new architectures, for data preparation, and for analytics. Big data requires the ability to integrate, store and transform vast stores of structured, unstructured and semi-structured data in a single “data lake” for various analytic and visualization initiatives. The return on the data lake investment, initially, can be measured in productivity gains and cost savings from simplicity gained in data access and manipulation. But ultimately, the true value lies in developing an internal data organization that can provide deeper insights faster by enabling front-line employees to utilize data to provide real-time offers and campaigns, resulting in a truly customer-oriented organization.

Need for Data Wrangling in Telecommunications

Despite the treasure trove of available data on customer behavior generated within CSPs, even the most innovative CSP organizations are challenged with the task of empowering their data analysts, IT managers and data scientists to access diverse data and build end to end views of their subscribers.
Relatively new sources like social media data are different than traditional legacy data because the format of the data is raw, unstructured and often complex. Sources lacking traditional structure are complex because they are not generated in neat tables of columns and rows. Data is further complicated given networks are constantly evolving with new protocols, new nodes and new services; all of which add to the amount of raw data created.

The opportunity big data platforms like Hadoop offer is the promise of a common environment to land, combine and analyze data. However, these projects often fail because of the challenges in defining the analysis workflow for new exploratory analytics initiatives and in providing tooling that empowers a diverse set of users to collaboratively work with diverse big data sources. The analysis of these new data types requires leveraging a combination of structured, semi-structured and unstructured data sources; meaning they have to be wrangled extensively before the actual analysis can even take place.

Common Challenges Faced by Telecommunications Organizations:

Skills gap: Performing exploratory analytics on large data sets requires a different set of skills, technologies and techniques not commonly found in existing CSP business and data analytics teams. Finding and hiring talent with the right mix of statistics, programming and telecommunications domain expertise is difficult. Once they are on board, even then, their skills are often marginalized as they spend a majority of their time in low-level cleaning tasks or not being able to gain access to the data they need. Some estimate that data cleaning and preparation tasks constitute 50-80% of the development time and cost in data warehousing and analytics projects.
Burden on IT Teams: In many organizations, the IT organization is responsible for managing and disseminating data resources. As the volume and variety of data increases over time, IT teams struggle to meet the analytics needs of the data scientists and business users they support.

Barriers to Organizational Collaboration: CSPs are often not structured as holistic organizations. Data is generated across a variety of disparate systems that run the networks, operation customer support, marketing, billing and more. They tend to be walled off from each other, with no easy way to transfer data between them. Within this kind of data environment, it is hard to detect customer dissatisfaction and ultimately predict customer churn before it occurs.

Empowering Telecommunications Organizations to Leverage Diverse Data to Drive Innovation

Trifacta was founded to provide an agile, broadly accessible and highly productive software platform for people who work with data. By removing friction stemming from a reliance on IT to access data and time spent “wrangling” frustrating, complex data sets, Trifacta accelerates the process of making data usable.

Trifacta is focused on empowering every organization to radically accelerate the process of preparing data of all shapes and sizes for analysis by providing an entirely new approach for how analysts access, transform and blend diverse sources of data.

Wrangling is this preparation process of converting diverse data from their raw formats into a structured and consumable format for business intelligence or statistical modeling tools. Trifacta offers an experience to non-technical users in a seamless and self-service way. Trifacta automatically discovers the data, structures it in a familiar grid interface, identifies potential invalid data and suggests the best ways to clean and transform the data. Trifacta learns from the user’s interaction, providing immediate feedback to the interactions, to better guide the user in enriching and validating the data at scale so it can be published with confidence to the next stage of the analytical process.

“Trifacta is an essential component of data preparation for further analysis, where you can complete a transformation as fast as a couple minutes, while using scripts will take much longer.”

XAVIER QUINTUNA
Principal Big Data Architect, Orange
Trifacta sits between the data storage layer, standard systems or a Hadoop platform and the visualization or machine learning applications used downstream in the process.

With Trifacta, CSPs can leverage their diverse data footprint including customer data, network data and billing data through a series of simplified interactions that have proven to save organizations substantial amounts of time and resources traditionally allocated to data preparation.

We bridge IT and business groups by providing a common platform where the contents of diverse data are discovered and analytic requirements are defined and executed at scale.

**Trifacta in Action**

**Leveraging Diverse Data to Predict Customer Churn**

Historically, customer churn has been a lagging indicator of customer satisfaction. The combination of Hadoop, Trifacta and downstream analytics tools have allowed telecommunications companies to build a 360 degree view of the customer by integrating external data with data housed within silos in the CSP. The objective of collecting this data to build new churn analysis models is two fold: CSPs can understand customer behaviors and priorities, as well as build predictive models on top of the data. Leveraging these analytic techniques enables CSPs to be in a stronger position to predict when and which customers are likely to leave, triggering more effective marketing campaigns and proactive customer outreach. This enables new leading indicators of churn from certain behaviors that may not have been available in the past, as well as driving new sources of revenue for the first time.

Trifacta is utilized in churn analysis by empowering analysts to access and transform diverse data sources for analysis, without solely relying on IT. In the following churn example, an analyst can take customer, account activity (e.g. voice and text data) and social media (e.g. Twitter handles) data sets from a Hadoop data lake, access, transform and blend them together using Trifacta. Subsequently, they can leverage an analytics package to conduct sentiment analysis, identifying negative and positive tweets to determine if existing customers have commented in one particular direction. Trifacta empowers analysts to transform messy raw data into a clean structured view of the data. Using other approaches, this process of preparing diverse big data for analysis can often take months, but with Trifacta it can now be performed in a matter of hours.

Trifacta can accelerate the process of data analysis and significantly improve the productivity of the people in organizations that work with diverse big data sources.
Creating Actionable Data Through Data Wrangling

Trifacta presents telecommunications analytics teams with a new approach to dealing with the challenges of working with the scale, complexity and diversity of data today. Trifacta’s approach to data wrangling utilizes the latest techniques in machine learning, data visualization and human-computer interaction to allow IT teams, data scientists, data analysts and business analysts to become more productive in wrangling data themselves—allowing them to build and manage data products and transformation scripts more effectively and on demand.

Benefits of Trifacta:

• **Empowering the people who understand the data best**: By providing a breakthrough user experience, Trifacta enables business users to access any form of data and prepare it in a consumable layout for their analysis. This can be done by themselves in a self-service approach, breaking the dependencies from IT or complex hand coding languages.

• **Accelerate time to value**: productivity is drastically enhanced by an order of magnitude. Instead of frustrating cycles between IT and business to deliver and refine the data, the data analyst can explore and validate the data with immediate feedback for the insight he’s looking for.

• **Lower business risk with more accurate data**: by allowing the end user to clean and validate data, the entire process is made more accurate and trustworthy. With correct information and trust in ones data, users lower the business risk in making decisions or implementing efficient data products.

• **Unlock innovation using a wider variety of data**: less time cleaning data means more time analyzing data. Trifacta liberates users from the cumbersome process of preparing data so they can focus on higher value analytics problems. With a simplified data preparation process, users can
The Opportunity for Data Wrangling in Telecommunications

augment the number of data points to unleash business innovation, take competitive lead, optimize operational efficiency and reduce the cost of processes with a noticeable impact to the company’s bottom line.

A Disruptive New Approach to Preparing Data

Experience a new way for working with diverse data—empowering analysts to interact with data in ways they never thought possible.

Interactive Exploration: Trifacta presents the user with automated visual representations of the data based upon the inferred data type of each attribute of the data. These profiles require no specification by the user and Trifacta automatically present each data type in the most compelling visual representation—geographic elements are presented as maps; time-oriented elements are presented according the common hierarchies such as day, month, year, etc. Every profile is completely interactive—allowing the user to simply select certain elements of the profile to prompt transformation suggestions.

Predictive Transformation: Upon pulling up a data set within Trifacta users are presented with a visual representation of the data set they are working with. These visual representations are interactive—enabling the user to click, drag or select over the specific elements or attributes of the data they’d like to manipulate. Every interaction within Trifacta leads to a prediction—the system evaluates the data you’re working with and the specific interaction applied against the data to then recommend a ranked list of suggested transformation for the user to evaluate or even edit depending upon what they’re trying to do.

As users browse through the different suggested transformations presented to them, the system will also present a preview of how each transformation, when applied to the data, will impact the data itself. This iterative feedback loop is always occurring throughout the use of Trifacta—constantly taking inputs from the data and the user to intelligently recommend ways to manipulate the data and giving the user the ability to validate their work with previews of each transform.

Intelligent Execution: Every transformation step defined by the user in the application is logged in Trifacta’s domain specific language called Wrangle, allowing the application to take the finished script the user is defining in Trifacta and compile that down into the appropriate execution framework, based upon the scale of the data the user is working with and the type of transformation. Depending upon the data, Trifacta can compile down to Pig, Spark and Trifacta’s own execution engine for jobs that can run on a single machine. This is all done behind the scenes—abstracting the user from the underlying execution framework.

Collaborative Data Governance: Although the core focus of Trifacta is enabling the people who know the data best to be able to access and transform it themselves, we recognize that organizations require having centralized processes for determining who has access to data, how metadata and lineage
About Trifacta
Trifacta, the leading data wrangling solution for exploratory analytics, significantly enhances the value of an enterprise’s big data by enabling users to easily transform and enrich raw, complex data into clean and structured formats for analysis. Leveraging decades of innovative work in human-computer interaction, scalable data management and machine learning, Trifacta’s unique technology creates a partnership between user and machine, with each side learning from the other and becoming smarter with experience. Trifacta is backed by Accel Partners, Greylock Partners and Ignition Partners.

For Additional Questions, Contact Trifacta
www.trifacta.com
844.332.2821

Experience the Power of Data Wrangling Today
www.trifacta.com/start-wrangling

are tracked, how transformation jobs are operationalized and how data sets and transformation scripts are shared with other users. Instead of creating a completely separate governance framework in Trifacta, we have built support for the existing enterprise standard frameworks on Hadoop for security, user authentication, access controls, job scheduling and so forth. This enables Trifacta customers to simply implement existing governance policies in Hadoop instead of creating a new, entirely separate governance framework for Trifacta.

Sources
1 New York Times, For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights, August 2014