

Khan, Farhan Ahmad

Conference Paper

Practical approach towards big data analytics

2015 Regional Conference of the International Telecommunications Society (ITS): "The Intelligent World: Realizing Hopes, Overcoming Challenges", Los Angeles, USA, 25th-28th October, 2015

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Khan, Farhan Ahmad (2015) : Practical approach towards big data analytics, 2015 Regional Conference of the International Telecommunications Society (ITS): "The Intelligent World: Realizing Hopes, Overcoming Challenges", Los Angeles, USA, 25th-28th October, 2015, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/146316>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

ERICSSON INDIA GLOBAL SERVICES PVT. LTD.

Practical Approach towards Big Data Analytics

Paper Submitted towards ITS Regional
Conference, Los Angeles, 2015

Author:

Farhan Ahmad Khan

RA-FM, BI and Analytics Consultant

Ericsson India Global Services

ASF Insignia, IT SEZ, Gurgaon Faridabad Road, Gwal Pahari

Gurgaon, India - 122003

Contact: +91 7838050512

farhan.ahmad.khan@ericsson.com

Introduction

Data is definitely one of the most strategic assets for Telecom Operators (telcos) today. With the rampant adoption of smartphones and growth in mobile internet, telcos today have access to unprecedented amounts of data sources including – customer profiles, device data, network data, customer usage patterns, location data, apps downloaded, clickstream data so on and so forth. Given the abundance of data at their fingertips, telcos are virtually sitting on a goldmine of information and are in a great position to capitalize on these valuable data sets. In order to gain insights into the avalanche of data that they have at their disposal, telcos are increasingly starting to adopt Hadoop & big data analytics solutions to turn their data into valuable business insights. Operators believe big data will play a critical role in helping them meet business objectives, promote growth, drive efficiencies and profitability across the entire telecom value chain. Today, Service Providers are using Hadoop in a multitude of ways to meet their business objectives. Some Service Providers utilize Hadoop purely as an operational data store to drive operational efficiencies – increasing storage capacity, improving performance and reducing costs while others are building specific data applications on top of Hadoop to drive real-time analytics and actionable insights.

How Big Data technology helps

The emergence of Big Data technologies enables telcos to capture, analyze and monetize enormous volumes of customer information and interaction data across multiple touch points in real time. This provides them with a unique strategic advantage to improve the performance of their core business by offering more targeted products and services and differentiated customer experience. The most valuable customer insights can be monetized, and an array of enabling data and business intelligence services can be offered to companies in a range of industries.

To better address major business issues such as targeting and upselling, telcos are considering new technology platforms that quickly provide access to large amounts of data from multiple sources. Many of these tools revolve around analytics, where telcos lead the way in global industry spending.

The tools enable managers to analyze customer behavior (frequently within a short time frame), uncover patterns in data with powerful data visualization, and respond faster to potential opportunities.

Despite a steady growth in IT spending by the industry overall, all too often a telco's IT infrastructure is represented by twentieth-century legacy systems built for transactions and batch-oriented updates. Their twenty-first-century competitors have moved on to IT infrastructure that accommodates speed through real-time processing and analysis of huge amounts of data, with the ability to engage and collaborate with colleagues for faster innovation.

In recent past, many telcos started implementing big data projects with their business objectives in mind, however, still a big number of telcos wonder if they actually need to go for big data or is their existing data warehouse sufficient.

The first question any tier-II telco asks is: ‘When should we decide to go for big data technology?’

Is it the amount of data that should be the basis of decision making?

Is it because others are doing it that we should also go for it?

Or, is it because big data is the buzz word in telecommunication industry these days?

The size of data available with telcos is huge and no doubt, the variety and velocity of data makes it even more difficult for telcos to meet their need with the current EDW systems.

The need for big data analytics arises due to the changing business models. There has been a paradigm shift in the way products and services are created and managed. It is no more the traditional sample based survey which telcos use to come up with new offerings for the customers.

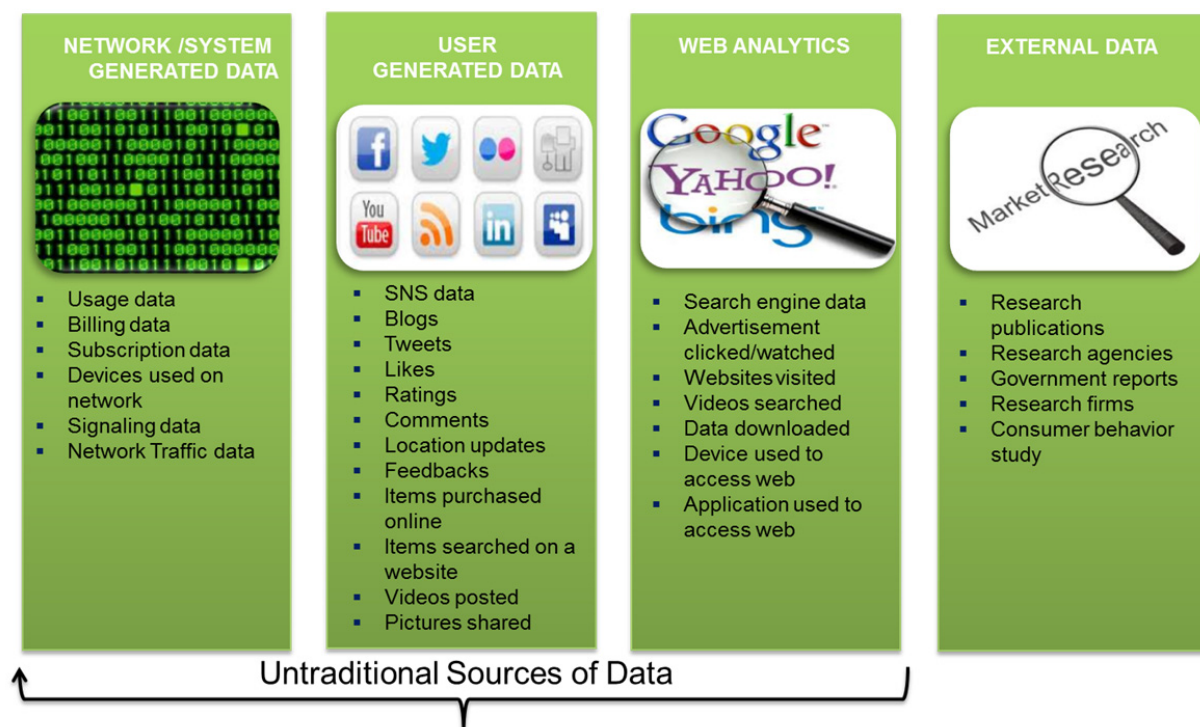


Figure 1: New Sources of Data

Any big data program in a telco is not about managing the data better because of increasing volume, it is about what an operator wants to do with the data available with them.

Starting Point for a Big Data Analytics program in a telco

Every telecom operator is searching for new ways to increase revenues and profits during a time of stagnant growth in the industry, but few have demonstrated the capabilities needed to make the most of this new technology.

That's why operators seeking to make initial inroads with big data are advised to avoid the usual top-down approach, which sets up a business problem to be solved and then seeks out the data that might solve it. This method does have benefits, but it is unlikely to lead to any serendipitous and surprising results — and it is difficult to execute until a company has demonstrated mastery in its use of data.

Instead, operators should begin with the data itself, experimenting with what they have on hand to see what kinds of connections and correlations it reveals. This process must be carried out quickly and iteratively, without the overbearing oversight from which so many business development projects suffer. If it's done right, what emerges can form the basis for more efficient operations and more effective marketing. At its best, this bottom-up method can give operators a more complete, transparent view of customers, enabling new and more profitable ways of capturing and retaining them.

The virtues of big data have been touted in hundreds of articles and reports during the past few years. Yet the benefits have proven elusive for a lot of companies. Indeed, some analysts already see a considerable level of disillusionment regarding big data — an umbrella term encompassing the new methods and technologies for collecting, managing, and analyzing in real time the vast increase in both structured and unstructured data — because too many efforts to implement the technology have not lived up to the high expectations triggered by the hype.

This is particularly true in the telecom sector. Most operators conduct analytics programs that enable them to use their internal data to boost the efficiency of their networks, segment customers, and drive profitability with some success. But the potential of big data poses a different challenge: how to combine much larger amounts of information to increase revenues and profits across the entire telecom value chain, from network operations to product development to marketing, sales, and customer service and even to monetize the data itself.

The typical advice offered to telecom operators is to take a top-down approach by focusing on specific business problems that big data might solve, and then gathering the data needed to solve them. But the challenge in this strategy is twofold: First, the business problem often exceeds the

capacity of the available data to solve it, and second, the process of gathering the right data to help solve the problem is poorly understood by many companies.



Figure 2: Big Data Analytics – Starting Point

To circumvent this problem, companies should begin with the inverse approach, viewing the opportunity from the bottom up. In this scenario, you examine the data currently available, and only then determine the business problems the data might help solve, with the help of any additional structured or unstructured data that might be needed. The best way to get started with this approach is through pilot programs. Keeping initial expectations reasonable, a dedicated team gathers all available data, analyzes it to allow new and unexpected opportunities to reveal themselves, and then tests the efficacy of the results in solving one or more real business problems. This tactic offers telecom operators and others a concrete starting point, a more realistic assessment of the benefits of big data, and a better understanding of what is actually needed to achieve those benefits in the long term.

This paper identifies some of the use cases of Big Data Analytics.

Big Data Analytics Use Cases

Reducing churn through real-time Call Data Record (CDR) analysis

Detailed analytics of CDRs by an operator revealed that customers experiencing more than 10 dropped calls in a month had 8 times greater propensity to churn. Furthermore, the operator could pinpoint the exact cause of the issue, from tower to customer's handset. This insight, available in real time, was used to design a churn prevention action specifically tailored for the individual customer. This approach helped the operator reduce churn significantly, especially in their high value segment.

Based on a recent research from Ovum, Telecoms operators globally can be sure to keep only about half of their existing customer base over the next 12 months and about a quarter of all users globally say they will definitely change providers. Given the impact of customer churn affecting the Telco industry today, telcos are effectively using big data analytics to bring together various data points including – quality of service, network performance, subscriber billing information, details on calls to the care centers, and social media sentiment analysis to build an effective model to predict and prevent churn. Churn prediction models allow Telcos to launch retention campaigns that identify and then address “at risk” customers via outbound channels. For example Telcos would be able to proactively reach out to high value customers, who have experienced a series of Quality of Service (QoS) issues or who shared a negative sentiment regarding the service in social media, and address those issues and offer them discounts or service credits to prevent customers from defecting.

Proactive approach to customer experience management

For today’s telcos, improving and optimizing the customer experience is key to maintaining a market differentiation and driving down churn. Telcos are leveraging Hadoop and big data analytics to gain a true 360-degree view of their customers along the customer journey and across all of the diverse interaction channels. Based on the detailed customer profiles, Telcos can then do targeted micro-segmentation of their consumer base, offer a compelling customer experience, develop personalized offer recommendations and predict and prevent churn. Some of the key use cases in this area include:

Targeted Marketing & Personalization: Offer personalized product offerings or derive specific upsell/cross-sell opportunities based on modeling a number of key attributes including subscriber’s usage patterns, device preferences, billing data, customer support requests, purchase history, buying preferences combined with their demographic information, location and socio-economic influences. Telcos can now create targeted customer micro-segments to offer more personalized offers and campaigns. This enables telcos to proactively present the right offer at the right time, in the right context to the right customer in order to improve conversion rates. Examples include – personalized data top-up plans or up-sell recommendations based on data usage, device upgrade campaigns based on specific customer preferences, and discounts or tailored offers based on recent purchases or enquiries or calls into the call center.

Customer Journey Analytics: Real-time analytics that map the user journey and generates actionable insights that can allow Telcos to respond quickly with a “next-best offer” and convert interested prospects into customers. Data such as customer demographics, purchasing behavior

and clickstreams are being combined with attributes such as location and content preferences to for next best offers. This also enables telcos to map specific customer's interactions with the telco at various stages of the lifecycle to promote tailored offerings and campaigns. Journey analytics example could include a real-time analytics model that pull together two personalized offers based on customer's/prospect's recent interactions, overall lifetime value and where they belong in the customer lifecycle.

Proactive Care: Using big data, telcos are building intelligence and analytics tools so as to proactively identify issues and fix it or offer a solution before it impacts the customer. Not only does it provides a compelling customer experience but it also deflects and prevents calls to the customer care centers thereby lowering support costs. Based on a recent survey conducted by Coleman Parkes focused on the Telco industry it found that 84 percent of respondents were more likely to recommend their service provider if the provider was able to identify and pre-emptively resolve potential issues affecting them. Given the impacts, Service Providers are proactively fixing issues or reaching out to customers to help resolve issues before they negatively impact the experience. Telkomsel, in Indonesia, for example, has built a 'proactive dashboard', based on the Cloudera platform, for their broadband services to identify customer experience issues for their high-value customers and proactively fix those issues or engage with customers.

Customer Sentiments Analysis: A telecom operator monitors twitter and social network activity and then tailors marketing campaigns using the customers' colloquial terms, rather than the traditional marketing or technical lingo. In Telkomsel, after interacting with the customers in their own language, the click rate for the online marketing campaign increased by 30% and conversion rate by 20%.

Network Optimization and Analytics

In order to keep up with the explosive growth in mobile data, telcos will need to continue to invest heavily in their networks, pumping in as much as 18 - 20% of their revenues every year into CAPEX. Network capacity is a highly valuable resource and Telcos are starting to leverage big data & analytics to effectively monitor and manage network capacity, build predictive capacity models and use it for prioritizing and planning network expansion decisions.

Network Capacity Planning & Optimization: By correlating network usage, subscriber density, along with traffic and location data, telcos can more accurately monitor, manage and forecast network capacity and plan effectively for potential outages. Using real-time capacity data, they can visualize and pinpoint highly congested areas where network usage is nearing its capacity thresholds, in order to prioritize expansion for new capacity roll out. Similarly for areas with excess network capacity, they can plan on running specific customer campaigns or promotions to increase uptake. Effectively optimizing and utilizing network capacity can mean millions of dollars in savings for telcos every year. Based on real-time analytics and traffic, they can also develop predictive capacity forecasting models, track actual versus forecasted traffic to fine-tune the model and plan for supplemental capacity in case of outages.

Network Expansion & Investment Planning: Planning and prioritizing network expansion projects can be a tricky and a balancing act for telcos. With so many dependencies and considerations, they need to be able to effectively prioritize their investments and resources based on – future connectivity needs, strategic objectives, projected RoI, forecasted traffic, customer experience etc. all while ensuring that your highest valued customers get to benefit from these investments as well. Telcos need to be able to effectively combine network traffic data, customer experience metrics, revenue potential and location data along with customer value data to ensure they are investing their CAPEX in the right spots. A number of telcos are already using Hadoop and big data analytics tools to aid in their network expansion and planning purposes. BT, for example, is using Hadoop and big data analytics to help them prioritize how and where they can expand high-speed broadband services to customers within the UK.

Real-Time Network Analytics: Telcos are also using big data and analytical tools to build real time capacity heat maps that continually monitor the quality of user experience and alert the teams for network congestion or potential outages. They used to rely on historical data for their network management but Big data analytics can enhance these processes by enabling real-time processing of network data to continuously monitor and manage the network and help them model network activity and map future demand. As a result, network engineers can get a holistic view of events occurring in the network and can proactively respond to network failures and outages helping them save millions. For example, Service Providers can now model the potential impact, in real-time, of a particular cell site goes down based on the number of subscriber and capacity in the adjacent sites. Similarly based on real-time data collected from the cell towers, engineers can monitor any drop in service performance at a specific location and send in crews, if need be, for a proactive resolution.

Telco Operational Analytics

Another key area of application for Telcos is the use of big data around driving internal efficiencies, process improvements and cost savings around the core Telco operations. Telcos are starting to adopt big data solutions powered by Hadoop for everything from plugging and minimizing revenue leakage, managing network and cyber security, driving down order-to-activation lead-times to proactively identifying and fixing customer issues in order to minimize truck rolls. Some of the more prominent use cases include:

Revenue leakage & Revenue Assurance: Based on industry estimates, Telcos lose approx. 2.8% of their revenues to leakage & fraud annually – costing the industry approximately US \$40 Billion every year, which means telcos could be adding \$ 40 Billion to their bottom-line without selling any additional products or services. Leveraging Hadoop and big data solutions enables the telcos to examine and plug dozens of actual or potential leakage points through the network and customer-facing systems, and to correct data before it reaches the billing system. Hadoop based solutions can help Service Providers to process and analyze both structured and unstructured data going back several years, rather than just a few months, enabling them to gain

a better understanding of the behavior of customers. Crucially, Hadoop has made it cost-effective to use deep packet inspection (DPI) to detect fraud and revenue leaks, as well as identify new revenue opportunities. DPI generates vast amounts of data – up to millions of records per second, which simply wouldn't be possible to collect or analyze without Hadoop.

Cyber Security & Information Management: As device proliferation continues, Cyber security takes center stage for Telcos as they race to ensure their networks and associated systems are secure from malicious attacks. Legacy event detection capabilities are unable to collect and analyze all the data sources necessary for identifying & responding to advanced threats due to the sheer cost and complexity. Security professionals need to be able to access and analyze an avalanche of data (including logs, events, packets, flow data, asset data, configuration data etc.) in real-time in order to mitigate risk, detect incidents, and respond to breaches. Communication Service Providers are increasingly starting to rely on Hadoop-based big data platforms to collect and analyze log data, to find anomalies that will in turn fire an alert when detecting unusual activity and creates an event for a security analyst. The Hadoop based data hubs can provide a cost-effective platform for storage and advanced analytics capabilities to support deep packet analysis, behavior analytics and profiling and threat modeling.

Big Data Operating model for Telcos

Big Data analytics provides an immense opportunity for companies to rediscover the scientific way of working – hypothesis-driven and based on rigorous data analytics. Implementing the most optimal operating model is key for organizations striving to maximize the benefits of Big Data.

Critical aspects of best practice operating models are:

1. Advanced Data and Business Intelligence (D&BI) capabilities.
2. Clear outline of the accountabilities across the D&BI value chain.
3. Organizational emphasis on the D&BI function.

For operators, there is a range of options for operating models. At one end of the spectrum is the fully Distributed Model, in which the roles and responsibilities across the Data and BI value chain are dispersed across the organization. On the other end is the Centralized Model in which a central function takes ownership from strategy to implementation of Big Data initiatives and acts as a service provider to internal, as well as external customers, on data, insights and advanced analytical offerings.

An operating model for Big Data should be selected based on some key considerations:

- Maturity level of Data & Business Intelligence processes – operators with lower maturity of traditional BI functions should aim at greater centralization.

- Ambition with respect to 3rd party monetization – operators aiming at significant focus on 3rd party monetization should aim at greater centralization
- Availability of advanced analytics skill sets in the organization – operators with scarce resources should aim at greater centralization to build capabilities sustainably in the long run.

The Big Data operating model should reflect a clear definition of responsibilities in the Data & BI value chain with complete centralization of advanced analytics, BI and Information Management activities to foster standardization and transparency of information on one hand, and organization-wide dispersion of end-use activities on the other hand. Telecom companies should embrace the role of Chief Data Officer (CDO). A CDO is dedicated to mining, analyzing and managing data, and coordinating its use throughout the whole organization across use cases. The CDO is in-charge of the central Data & BI function which should be the engine to deliver analytics products roadmap, implement the chosen business models for internal and external monetization, implement robust data governance and promote a data driven mindset in the organization.

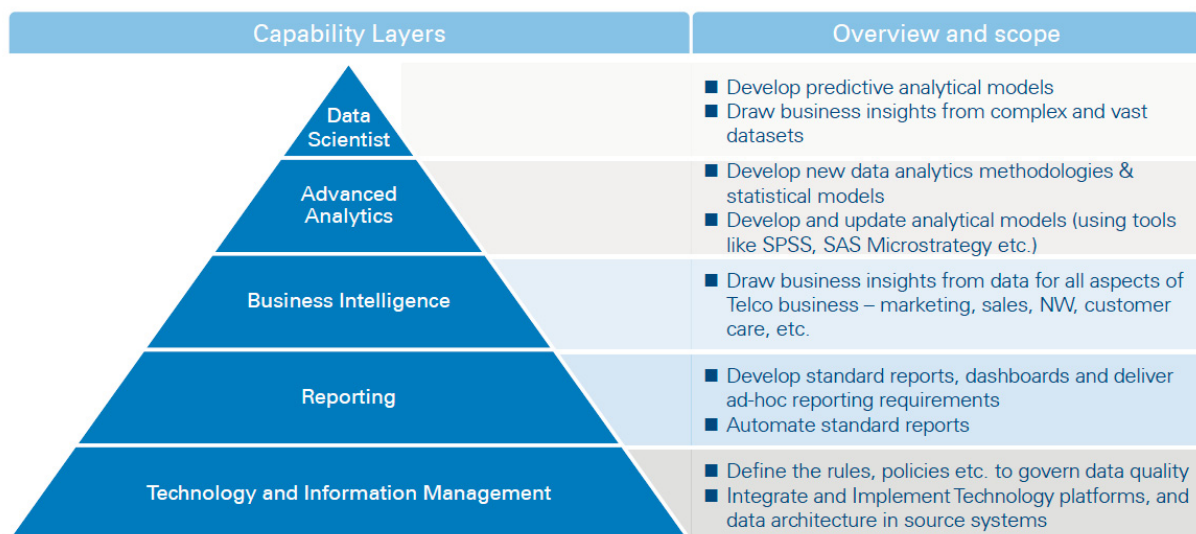


Figure 3: Telcos aspiring to monetize Big Data investments should develop capabilities across five layers of D&BI value chain

Hadoop as an Enterprise Data Hub

Some telcos have identified their own Hadoop use cases, but there are similar patterns in the Hadoop data architectures that they all build. Those data architectures allow telcos to store new types of data, retain that data longer, and join diverse datasets together to derive new insight.

The following reference architecture diagram represents a generic approach adopted by the telcos.

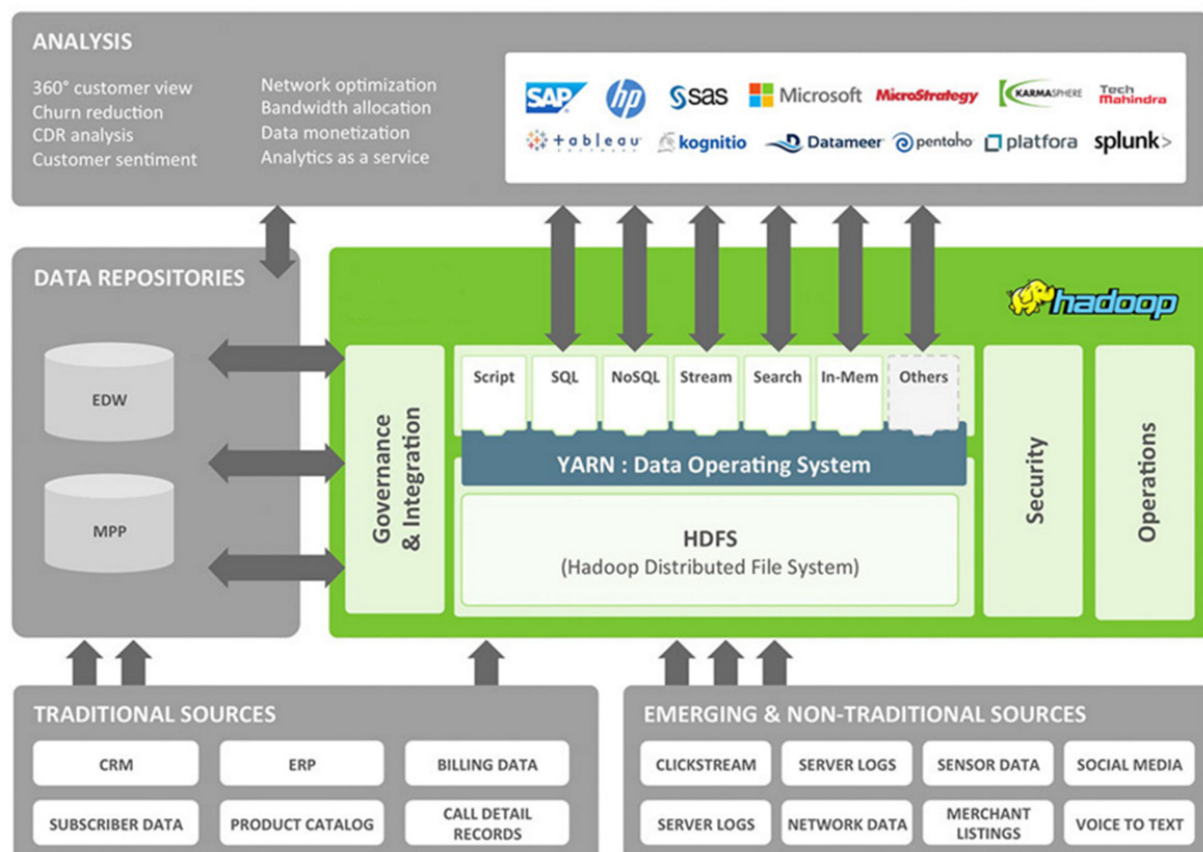


Figure 4: Generic Big Data Architecture in Telco

Data Warehouse vs. Hadoop

This rapid (and irreversible) pace of telecommunications data growth requires more efficient, scalable storage. So telcos are using Apache Hadoop to turn their burgeoning storage liabilities into strategic information assets.

Given that Hadoop is on such an exponential rise in adoption and its ecosystem is expanding in both depth and breadth, it is natural to ask whether Hadoop's rise will cause the demise of traditional data warehousing solutions.

Another way to put this question is to look at it in a bigger context: To what extent is big data changing the traditional data analytics landscape?

Data warehousing is a set of techniques and software to enable the collection of data from operational systems, the integration and harmonization of that data into a centralized database and then the analysis, visualization and tracking of key performance indicators on a dashboard.

A key difference between data warehousing and Hadoop is that a data warehouse is typically implemented in a single relational database that serves as the central store. In contrast, Hadoop and the Hadoop File System are designed to span multiple machines and handle huge volumes of data that surpass the capability of any single machine.

Furthermore, the Hadoop ecosystem includes a data warehousing layer/service built on top of the Hadoop core. Those services on top of Hadoop include SQL (Presto), SQL-Like (Hive) and NoSQL (Hbase) type of data stores. In contrast, over the last decade, large data warehouses shifted to use custom multiprocessor appliances to scale to large volumes like those from Netezza (bought by IBM) and Teradata. Unfortunately, those appliances are very expensive and out of reach for most small- to medium-sized businesses.

With this background and context it's natural to ask: Is Hadoop the death of data warehousing?

To answer this question, it's important to divide the techniques of data warehousing from the implementation. Hadoop (and the advent of NoSQL databases) will auger the demise of data warehousing appliances and the "traditional" single database implementation of a data warehouse.

Evidence of this can be seen with Hadoop vendors like Cloudera billing its platform as an "enterprise data hub," in essence subsuming the need for traditional data management solutions.

And the Hadoop or NoSQL ecosystem is still evolving. Many big data environments are choosing hybrid approaches that span NoSQL, SQL and even NewSQL data stores. Additionally, there are changes and potential improvements to the MapReduce parallel processing engine on the horizon like Apache's Spark project. So, while this story is far from over, it is safe to say that traditional, single server relational databases or database appliances are not the future of big data or data warehouses.

On the other hand, the techniques of data warehousing to include Extract-Transform-and-Load (ETL), dimensional modeling and business intelligence will be adapted to the new Hadoop/NoSQL environments. Furthermore, those technologies will also morph to support more hybrid environments. The key principle seems to be that not all data is equal, so IT managers

should choose the data storage and access mechanism to best suit the usage of the data. Hybrid environments could include key-value stores, relational databases, graph stores, document stores, columnar stores, XML databases, metadata catalogs and others.

In general, while big data will change the implementation of data warehousing over the next five years, it will not obsolete the concepts and practice of data warehousing.

References

1. "Big data analytics and Telecom Analytics Market: Business Case, Market Analysis & Forecasts 2014 – 2019". Mind Commerce. Retrieved 6-02-2015
2. "Benefitting from Big data analytics". Booz and Company. 2013
3. "Data Science in Telecommunications". www.mastersindatascience.org. Retrieved 02-02-2015.
4. "From big data analytics to knowledge: analytic use cases for CSPs". HP. Retrieved 28-01-2015.
5. "BIG DATA ANALYTICS Mine for revenue". Page 50-51. Voice & Data, September 2014.
6. "Data Science as a Service". www.analyzecorp.com. Retrieved 02-02-2015.
7. "What is a data scientist". www.ibm.com. Retrieved 04-02-2015
8. Jeff Leek (2013-12-12). "The key word in "Data Science" is not Data, it is Science". Simply Statistics.
9. "Hadoop vs data warehousing". www.gcn.com. Retrieved 20-10-2015.
10. "Telcos increase Customer Engagement with Big Data insights". Harvard Business Review. Retrieved 15-10-2015.
11. "Big Data use cases for telcos". Cloudera. Retrieved 15-10-2015.
12. "Big Data: A goldmine in telcos' backyard". Arthur D Little. Retrieved 15-10-2015.
13. "Hadoop is transforming telecommunications". Hortonworks.com. Retrieved 22-10-2015.