cloudera®

INDUSTRY BRIEF

Driving Value for Telcos with an Enterprise Data Hub

Data as a Strategic Asset for Telcos

Communication Service Providers (CSPs) are among the world's biggest aggregators of consumer data and work under the most uncertain regulatory and market conditions. With the growth and rampant adoption of smartphones and mobile internet, CSPs today generate unprecedented amounts of valuable data including – customer profiles, device data, network data, usage patterns, location data, apps downloaded, content preferences, clickstream data so on and so forth. For example, a medium size wireless company generates approximately 1-2 billion CDRs (call detail records) and terabytes of data every day. Significant volumes of data are being generated every second and with the rapid adoption of concepts such and Internet of Things (IoT) and Machine-to-Machine (M2M), these data sources are continuing to grow exponentially.

Where is the Telecom Data??

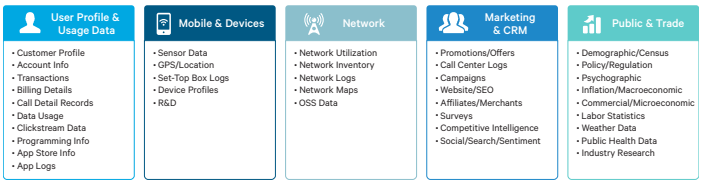


Figure 1: Where is the Telecom Data?

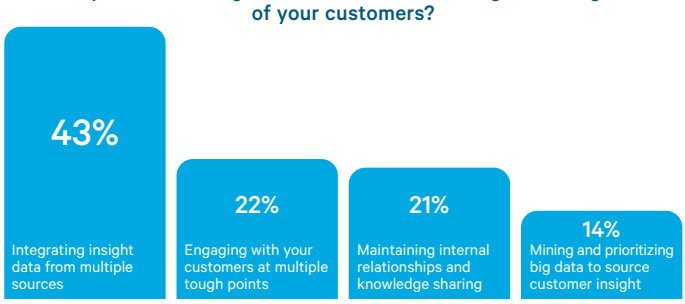
With this abundance of available data, CSPs are virtually sitting on a goldmine of information are in a great position to capitalize on all this data. However, the question really is – How can CSPs leverage this treasure trove of data that they have at their disposal and turn it into something meaningful that will help them drive down operational costs, deliver a more compelling and personalized customer experience, reduce churn or help drive new revenue streams?

Key Data Management Challenges for Service Providers

1) Data Silos: Even though Telcos have access to a wide variety of disparate data sources, one of the biggest challenges that they face today is breaking down the data silos that have been built up over the years. Driven by consolidation and rapid expansion, Telcos, like many large enterprises, have seen their various systems grow unchecked over the years to a point where they store overlapping and often conflicting information. For most large operators, data tends to reside across hundreds of systems and platforms, that cuts across organizational boundaries including – multiple billing & rating platforms, varied point of sale and customer care systems, CRM systems, sales and ordering systems, several network, inventory and OSS systems and whole suite of legacy and enterprise platforms as well. And the complexity is only multiplied as CSPs start to expand into additional lines of business and offer bundled product offerings. Each business unit maintains its own set of customer, product, and inventory data across it own network of internal and legacy systems and data marts.

According to the results of a Telecoms IQ survey, released in 2014, nearly half of telecoms operators believe that *integrating data from multiple sources* is their greatest barrier to obtaining a 360-degree view of their customers[1].

What do you find is the greatest barrier to obtaining a 360 degree view of your customers?



Source: Customer Insight & Analytics in Telecoms Market Survey conducted by Telecoms IQ – Sept 2014

Figure 2: Customer Insight & Analytics in Telecoms Market Survey, Telecoms IQ, Sept 2014

For CSPs, it is virtually impossible today to bring together all or even some of these data sets together to build a single, unified view of the customer and their business. Telcos need to be able to combine historical data with new data, from diverse sources, both structured and un-structured, and be able to interact and analyze with data in multiple ways to gain insights.

> CSPs need to be able to bring together data from disparate data sources (both structured & unstructured) to generate meaningful value out of it.

2) New & Diverse Data Sources: Apart from data silos, the second key challenge that Telcos now need to address is the new and ever-increasing data sources and types. CSPs currently employ a variety of systems to support their diverse data goals: data warehouses for operational reporting; storage systems to keep data available and safe; specialized massively parallel databases for large-scale analytics; and search systems for finding and exploring data. While these systems are suitable for traditional data and workloads, they are not equipped to handle today's exponential growth in data volume and variety, or the range of users who seek insights from that data. With an increasing volume of the data coming in unstructured or semi-structured formats from emerging data sources (such as clickstream

[1] Source: Telecoms IQ - Customer Insight & Analytics in Telecoms Market Survey, Sept 2014

data, social media, streaming sensor data, data from network logs and log files, documents, call center records), CSPs need to be able to bring together multiple types of data from these disparate data sources to generate meaningful value out of it.

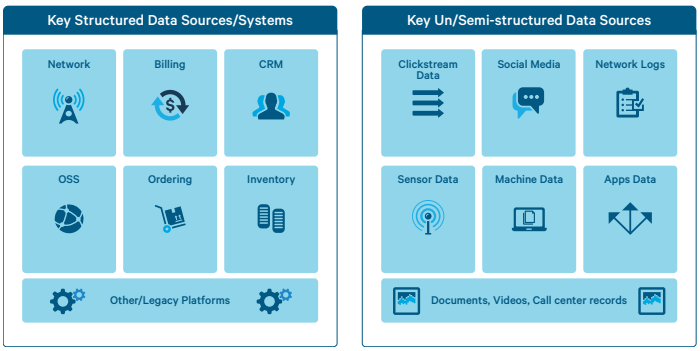


Figure 3: Key Structured & Unstructured Data Sources for Telcos

As data volumes continue to grow at ~100% year-over-year, Telcos will need to augment or transform their traditional data management mechanisms in order to take advantage of this new variety and speed of data.

3) Cost of Data Storage & Processing: For Telcos, securely storing billions of records and providing transparent, real-time customer access has been cost-prohibitive as it requires multiple expensive systems to handle the huge size, complexity, and variety of data. A recent survey of large enterprise firms including telecommunications carriers indicated that storing data in relational database systems typically runs between $30,000 and $100,000 (USD) per terabyte (TB) per year in total costs. Due to the considerable growth in data generated and managed, most Telcos employ a 'multi-temperature data management' strategy to reduce costs. In this type of infrastructure, "hot data" that is frequently accessed for reporting and analysis is kept in the data warehouse. "Warm" or "cold data" — which is typically older, considered less valuable, and accessed less frequently — is offloaded to a data storage system that offers cheaper storage and is not readily available for reporting or analysis. Given these challenges, Telcos are gravitating towards a more effective and hybrid data management mechanisms, more suited for petabyte scale, multi-structured data storage and analytics.

> Storing data in relational database systems typically runs between $30,000 and $100,000 (USD) per terabyte (TB) per year in total costs

The SQL based traditional data warehouses have served the Telcos very well over the years, however as the volumes and variety of data starts to multiply at an accelerated pace, storing data in warehouses might not be a scalable and cost-effective approach. A new hybrid model is required, which pragmatically extends the value of existing investments while enabling fundamentally new ways of delivering value from data.

A New & Effective Data Management Architecture for Telcos

Hadoop is rapidly emerging as the preferred data processing framework tailored for Telcos due to its flexibility, scalability and low cost of storing and processing raw data. More and more CSPs are utilizing Cloudera's Enterprise Data Hub (EDH), powered by Apache Hadoop, to complement their existing data warehouses to improve performance, reduce costs, and enable new insights. With an enterprise data hub, telecom operators can easily bring together user, transaction, network, and service data across multiple sources into a single, unified platform at considerably lower cost. And because Hadoop is built on a highly scalable and flexible file system, any type of data (including structured data from network, BSS and OSS systems or streaming data from social media or machine logs) can be loaded into the EDH without altering its format, preserving data integrity and delivering complete analytic flexibility. Data generated by machines and sensors, including application and web log files, can be collected in real time and streamed directly into an enterprise data hub instead of being staged in a temporary file systems or data marts. And because Hadoop uses industry standard hardware, the cost per terabyte of storage is, on average, 10x cheaper than a traditional relational data warehouse system.

> With an enterprise data hub, telecom operators can easily bring together user, transaction, network, and service data across multiple sources into a single, unified platform at considerably lower cost.

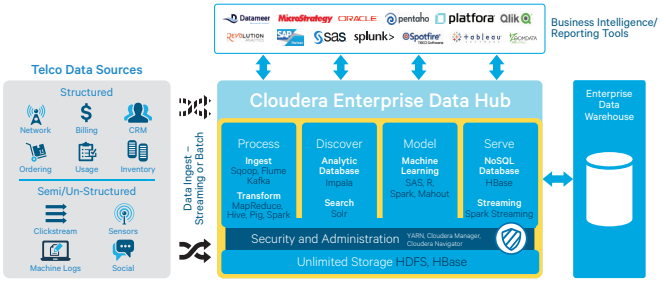


Figure 4: EDH based Data Management Architecture for Telcos

The Enterprise Data Hub loads, stores, and transforms source data to feed the warehouse with clean, high value data for enterprise reporting. To achieve full value from all your data, an enterprise data hub builds on Hadoop's principal benefits by delivering an array of powerful analytics and processing engines certified to work seamlessly with your existing infrastructure and systems. More importantly, it is connected and compatible with the systems and BI/ reporting tools you already rely on and thereby complement your existing investments.

Thus, the introduction of an enterprise data hub at the core of Telco's information architecture promotes the centralization of all data, in all formats, available to all business users and groups, with full fidelity and security at up to 99% lower capital expenditure per terabyte compared to a traditional data management mechanisms.

The Enterprise Data Hub promotes the centralization of all data, in all formats, available to all business users and groups, with full fidelity and security at up to 99% lower capital expenditure per terabyte

The Enterprise Data Hub provides the flexibility to run a variety of workloads – including batch processing, interactive SQL, search, advanced analytics, and machine learning – together with the robust security, governance, data protection, and management that Telcos require. Cloudera's Impala enables users to run native, real-time SQL queries directly against Hadoop data, sidestepping real-time query limitations of Hive and MapReduce to explore, visualize, and analyze data to discover interesting patterns and trends. To overcome latency, Apache Flume and Apache Kafka —Hadoop's service for efficiently collecting, aggregating, and moving large amounts of log data—can load billions of data points into HDFS, the distributed file system and primary storage layer for Hadoop, within a few seconds. The enterprise data hub also helps analyze this massive data and even run models on streaming data using the in-memory capabilities of Apache Spark—the next-generation, open-source processing engine that combines batch, streaming, and interactive analytics on all the data in HDFS.

Thus, with an Enterprise Data Hub, CSPs are rethinking data, transforming it from a cost center to a crucial asset they can use derive new insights, lower data management costs while meeting the data demands of the Telecom industry.

Removing Data Silos

One of the most compelling benefits of applying a data hub strategy to the enterprise architecture of a Telco is the ability to consolidate all the different data types, from many different sources, into a single, central, active repository. So all of the structured data from your diverse billing, customer care, ordering, network and legacy systems along with semi/un-structured data including network logs, clickstream data, social media and sensor data can now be stored within a single platform for comprehensive analytics. Data can be collected, stored, processed, explored, modeled, and served in one unified platform enabling Service Providers to stitch together a true 360-degree view of their customers and their business.

Service Providers can now effectively combine customer usage information with sentiment analysis from social media feeds to identify, analyze and prevent churn. CSPs can bring together the customer usage patterns with customer lifetime value data and network performance data to 'proactively' address any Quality of Service (QoS) issues for customers based on their value to the business. When information is freed from silos, secured, and made available to the business to answer key questions about the market - as they need it, in its original form, and accessed via familiar tools - everyone in the C-suite can rest assured that they have a complete view of the business and their customers, perhaps for the first time.

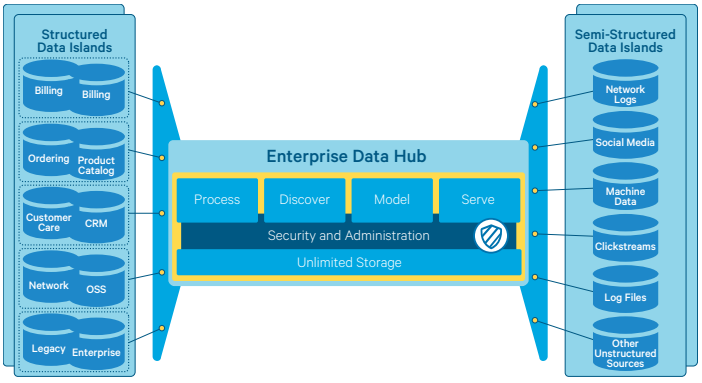


Figure 5: Bridging islands of data to generate unique value

With an Enterprise Data Hub, data can be collected, stored, processed, explored, modeled, and served in one unified platform enabling Service Providers to stitch together a true 360-degree view of their customers

Most importantly, centralizing and bringing compute to all your data enables new information-driven business competencies that were previously too expensive or complex for most enterprises. Thus harnessing Hadoop & Enterprise Data Hub presents Telcos with endless opportunities to reuse both historical and real-time data to gain a true 360-degree view of their customers and business, drive down churn, optimize network performance and launch new revenue streams.

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified Platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, access, process, secure, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Cloudera's open source big data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 22,000 individuals worldwide. Over 1,400 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production.

For additional information, please visit us at: www.cloudera.com

cloudera.com

1-888-789-1488 or 1-650-362-0488
Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2015 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.