Modeling of an Efficient Low Cost, Tree based Data Service Quality Management for Mobile Operators using In-Memory Big Data Processing and Business Intelligence Use cases

Kingsley A. Ogudo and Dahj Muwawa Jean Nestor

December 14, 2018

Modeling of an Efficient Low Cost, Tree based Data Service Quality Management for Mobile Operators using In-Memory Big Data Processing and Business Intelligence Use cases

***Abstract*—Network Operators are shifting their business interest towards Data services in a geometric progression manner, as Data services is becoming the major source of Telco revenue. The wide use of Data platforms; such as WhatsApp, Skype, Hangout and other Over the Top (OTT) voice applications over the traditional voice services is a clear indication that Network Operators need to adjust their business model and needs. And couple with the adoption of Smartphones usage which grows continuously year by year, which means more subscribers to manage, large amount of transactions generated, more network resources to be added and evidently more human technical expertise required to ensure good service quality. With the large amount of transactions generated by data traffic and the high demanding service qualities, Mobile Network Operators are spending millions of rands/dollars to deploy Robust Service Quality Management (SQM) and Customer Experience Management (CEM) to stay competitive in the market. These solutions' high cost is justified by the integration of Big Data Solutions, Machine Learning capabilities and good visualization of insight data. However, the Return on Investment (ROI) of the expensive systems are not as conspicuous as the provided functionalities and business rules. Therefore, in this paper an efficient model for low cost Service Quality Management system is presented, using the advantages of the In-Memory Big Data processing and simple low cost business Intelligence tools to showcase how a good Service Quality Management can be implemented with no big investment.**

Keywords—Service Quality Management, In-Memory Big Data, Business Intelligence, Service Quality Index, Over The Top Application (OTT), Data Traffic and ROI.

I. INTRODUCTION

One of the biggest concerns of the Cellular Network Operators is to ensure continuous service availability and quality for a good subscriber retention index. However, with the growth and high adoption of Smart phones and Smart devices alike, increase in Data services, data traffic is increasing exponentially. Tracking Service Quality becomes a tedious and complicated exercise, requiring not only telecommunications knowledge but also the knowledge required to find useful information in the billions of transactions records generated by mobile subscribers. Thus, the involvement of Data Analytics, Big Data processing and Business Intelligence in the area of Telecommunications. The big step in terms of innovation in the area of technology is the advancement of computer technology which has changed completely the way of handling Data, Analyzing Data and visualizing data [1]. The data collected from the network contains a lot of subscribers' information that can help network operators in decision-making processes and business improvement methodology. Cellular Network Operators aim to improve service quality taking into consideration the time constraint. The justification of the big money being invested in Service Quality and Customer Experience Managements is scaled on multiple levels including Domain expertise, technology awareness and the investment in Science, Engineering, Mathematics and Technology. In this paper, a model of a Tree-based Service Quality Management system is investigated and illustrated using data collected from an Operator, SparkSQL with In-memory Big Data processing and cloud based BI tool for visualization.

II. THE CONCEPT OF IN-MEMORY BIG DATA

The concept of Big Data is related to the amount of data to be processed. The processing of such data requires a lot computational power and advanced technologies. The characteristics of Big Data emphasizes on the Volume, Velocity and Variety "VVV" techniques, called the three "Vs" to make the difference with the traditional data handling methodologies, Volume, Velocity and Variety [2].

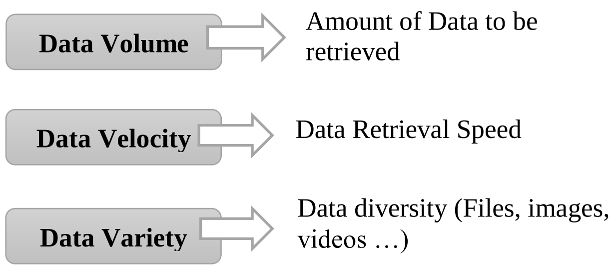

Figure 1: Big Data Characteristics, the 3 Vs

A. *In-Memory Data Processing*

For many years data processing has been based on data stored in disks including relational databases such as Structured Query Language (SQL). However, with the growth of data and the high demand of low-latency services, the concept of in-memory Database storage has developed. Because data stored in the memory are accessed quickly, in-memory processing becomes the most suitable methodology for real time analytics [3]. In-Memory data processing has the advantage of reducing the Hardware footprint while improving the performance. CPU power, Memory and disk storage are considered and are subject to variation to meet the business need.

B. SparkSQL

The advance of technology comes with the challenge of skills development. With the high success in the market of traditional relational database systems (RDBMS), companies have invested a lot in relational databases such as SQL. Migrating completely away from the Relational Databases to new architectures becomes a huge technological challenge.

SparkSQL is an Apache Spark module developed to process Relational Data Structure with Spark [4]. It allows the usage of SQL queries to retrieve information. Thus, SQL Users to take advantage of the Big Data processing. SparkSQL presents three main capabilities [5]:

- Data frame abstraction in different programing languages including Java, Python and Scala to efficiently work with structured data sets.
- Read and write many popular structured data formats such as JASON, Parquet, and Tables).
- Using SQL, query data from inside Spark program and from external tools that use database connectors to communicate with Spark (JDBC/ODBC). This scenario is used in this paper since the connection to the dataset is via JDBC connector.

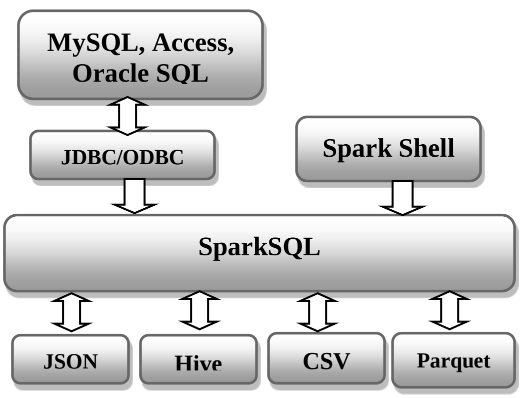

Figure 2: SparkSQL Scope of Usage

C. SparkSQL In-Memory Caching and Data Access

Spark SQL provides the possibility to cache data in a structured way in the memory. Applying columnar compression techniques on the structured data dictionary encoding scheme, it reduces memory footprint. Data cached in memory can be retrieved for different applications including visualization and Machine Learning.

III. The Concept of Business Intelligence

Although the concept of BI (Business Intelligence) has been given different façades by different people, the simplest definition is the presentation of clear and relevant dashboards or reports which transform data into actionable information [6]. BI saves time in terms of reporting and improving productivity. The key concepts to consider when delivering corporate BI are KPI (Key Performance Indicators), Scorecards and dashboards. Another important aspect of BI is the drilling down and navigation between different dashboards. In this paper, the concept of Service Quality Index (SQI) is introduced along the mentioned key aspects.

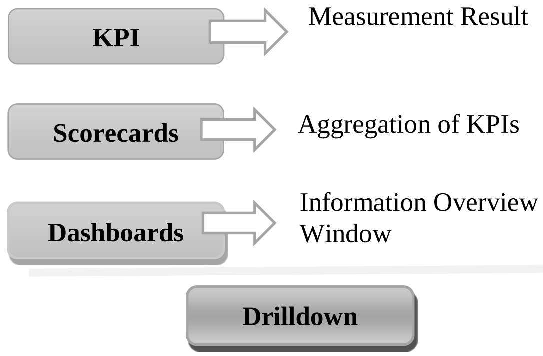

Figure 3: Key Concepts of Business Intelligence

IV. Problematic and Objectives

Using the advantages of SparkSQL with in-memory data caching and the BI tool, this paper demonstrates the building of a low cost tree based Service Quality Management for Network evaluation and monitoring. The objective is to construct a basic framework that leverages on the current business demand of Network Operator. The business needs of Operators differ based on the business models, the processes, competitors and maturity in the business. Nevertheless, the main goal and objective is to have good Service Quality and improve Customer Experience.

A. Relating the Problem to Big Data

Let us assume that a Network Operator is launching a new video service and is interested in promoting the package to a certain group of subscribers who are already high users of video applications based on their past customer experience data. The solution to the puzzle is related to the choice of the data to be used and the methodology to extract such information. This could be achieved by classifying customers with high volume of data usage on video-streaming. Using the applicable algorithm to classify the customers based on the amount of data bundles used on video-streaming, individuals subscribers can be selected; and this relates already to a real Big Data problem due to the fact that the search data in the Network Infrastructure combines both text and numbers (Variety), and usually large (Volume, Velocity) [7]. However, the interesting property of Big Data is the amount of records which in this case could be in the order of hundreds of millions of transactions done by customers provided from each part of the Network.

B. Relating the Problem to Business Intelligence

In order to take effective business decisions based on Data Analysis, the transformed data must be presented in an understanding, meaningful and clear way. Let us assume that the

Network Operator in section A identifies the subscribers likely to adopt the new video service launched and presents the data in form of reports and dashboards, classifying the users per Region, District, Town and Neighborhood to easily contact them. This relates already to Business Intelligence because the production of the dashboards and reports have translated the Network data into actionable information [8].

V. QUALITY OF SERVICE OVERVIEW

The objective of QoS is to provide a satisfying end user experience when using a specific service. The perception of service quality differs based on the service requirements. For instance listening to music on the internet requires a lower delay compared to sending an email, while sending an email requires high reliability compared to the audio streaming. The QoS as used in the scope of this paper is quantitative, involving definition of metrics such as throughput, latency, and retransmission [9].

A. Service Class Categorization

Service Quality is categorized based on their requirements. The limitation of the air interface is taken into consideration when categorizing services also referred to as QoS Classes defined by the 3GPP [10]. The key differentiator between the service classes is the sensitivity to delay.

Table 1: QoS Service Class

Service Class	Characteristics	Applications
Conversational	• Low delay. • Preserve time variation	• Voice Call • VoIP
Streaming	• Preserve time variation	Audio & Video Streaming
Interactive	• Request response pattern • Preserve payload content	• Web browsing • Instant Messaging
Background	• Not delay sensitive • Preserve payload content	• Email • Files download

B. Service Application Model Used in this Research

Based on the theoretical background elaborated in the previous section, the service applications used in the SQM system are shown in Figure 6 organized in a Tree, providing drilldown mechanism from top to Bottom, Layer 0 to Layer 3. To the layer tree, a set of Performance Indicators is applied.

VI. HARDWARE PRE-REQUISITES

The Dataset is made available through JDBC connection. A good understanding of the dataset is very crucial into building the use cases, which requires knowledge of Telecommunications. A high performing computer is used, Core i7 with 16Gb RAM, 2.40 GHz CPU. The processing power of the server determines and limits the amount of data that can be analyzed. The dataset to be loaded in memory must be less than the memory capacity. Apache Spark must be installed in the server with Spark SQL for processing. A cloud BI tool is used for clear and meaningful visualization of processed data. SQL language is a pre-requisite as although the Spark is used o cache and access data in memory, the queries are written in SQL which are properly understood by the SparkSQL Module.

VII. DESIGN PROCESS METHODOLOGY

A database connection is established between the original data source and the processing server. The processing server caches the big data frame in memory for fast processing using Spark SQL. SQL is used to query information from the dataset and display on the dashboard. The SQM system is separated into two parts: the Back-End and the Front-End as shown on Figure 5. In its simple format, relating the process to a Data Mining problem, data is collected and understood, then transformed to the correct format for the mediation layer, cached in memory for fast processing and smartly visualized to improve business decisions as illustrated on Figure 4

VIII. UNDESRTANDING THE DATASET AND KPI

The dataset contains service information from the network. It is made of categorical fields that are used as keys for queried information and numerical fields from which KPIs and Score cards are deduced. The data fields used in the SQM database system contains user information, location information, application information and Key Performance Indicators.

A. Key Performance Indicators

The exchange of information in the network is done through transport and application protocols which carry packet. TCP (Transmission Control Protocol) is used as the transport protocol for connection-oriented packets. TCP applications are very sensitive to the conditions of the communication pipes. The selected performance metrics used for the SQM are as follow:

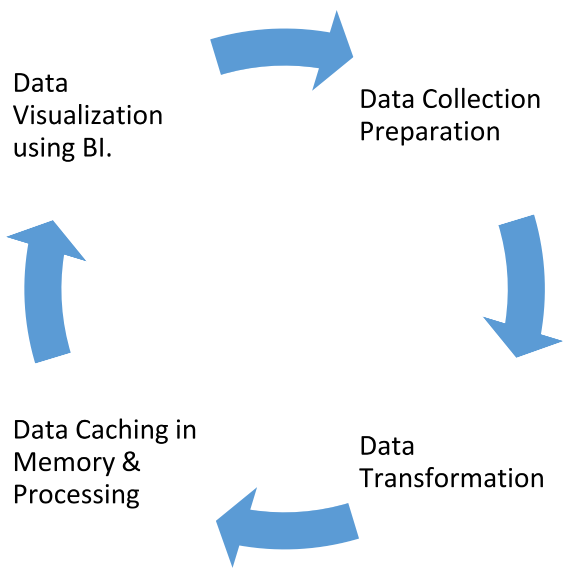

Figure 4: Design Methodology

- Latency: time required for packets to be transmitted across the Network. When the Round Trip Time (RTT) is relatively high, this can affect End-to-End Service Quality. Measured in Milliseconds.
- Packet Loss/Packet Retransmission.
- Throughput: the amount of Bits that can be sent in one second (Data rate).
- Domain Name Server Performance:

B. Service and Key Quality Index Logic

In order to give the overall quality of service for the entire network, the concepts of Key Quality Index (KQI) and Service Quality Index (SQI) are introduced. KQI is a combination of Key Performance Indicators (KPI) with weighted inputs to the formula. Depending on the Operators, the KQI formula can be modified.

$$KQI = \beta \sum_{i=1}^{n} KPI_i \qquad (1)$$

$$SQI = \alpha \sum_{j=1}^{m} KQI_j \qquad (2)$$

Substitute (1) in (2) provides the general equation of the global SQI of the Network:

$$SQI = \alpha.\beta \sum_{j=1}^{m} \left(\sum_{i=1}^{n} KPI_{ij} \right) \qquad (3)$$

Where α and β are weighted coefficient of the performance Index. Which can be defined by the Operator or the Data Scientist doing the Analytics of the Service Quality.

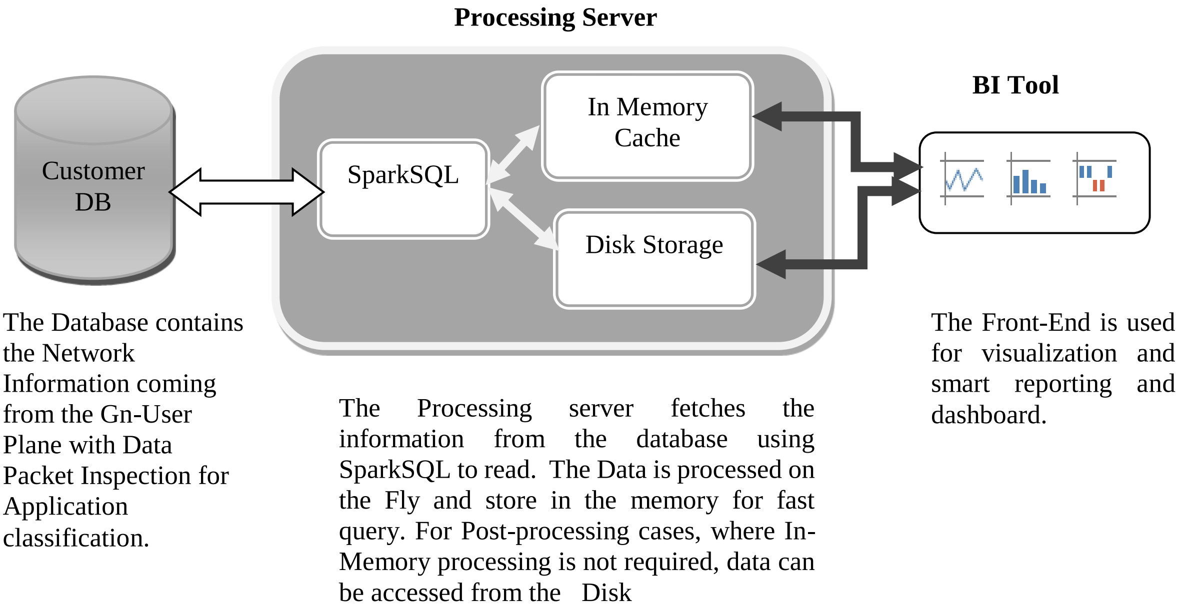

Figure 5: In-House SQM System Architecture

C. Performance Aggregation Model

One of the key aspects of the SQM is to be able to determine the points of quality degradation in the network. A poor service can be linked to Network problems, device problems or single subscriber related. Thus the need of multi-level aggregation. The aggregation levels used are Network, Device and to a certain level Customer. The objective is to identify good and bad performing cells or regions, devices and customers. For each Layer of the Service Quality System, Network performance, Device Performance and Customer Performance are summarized.

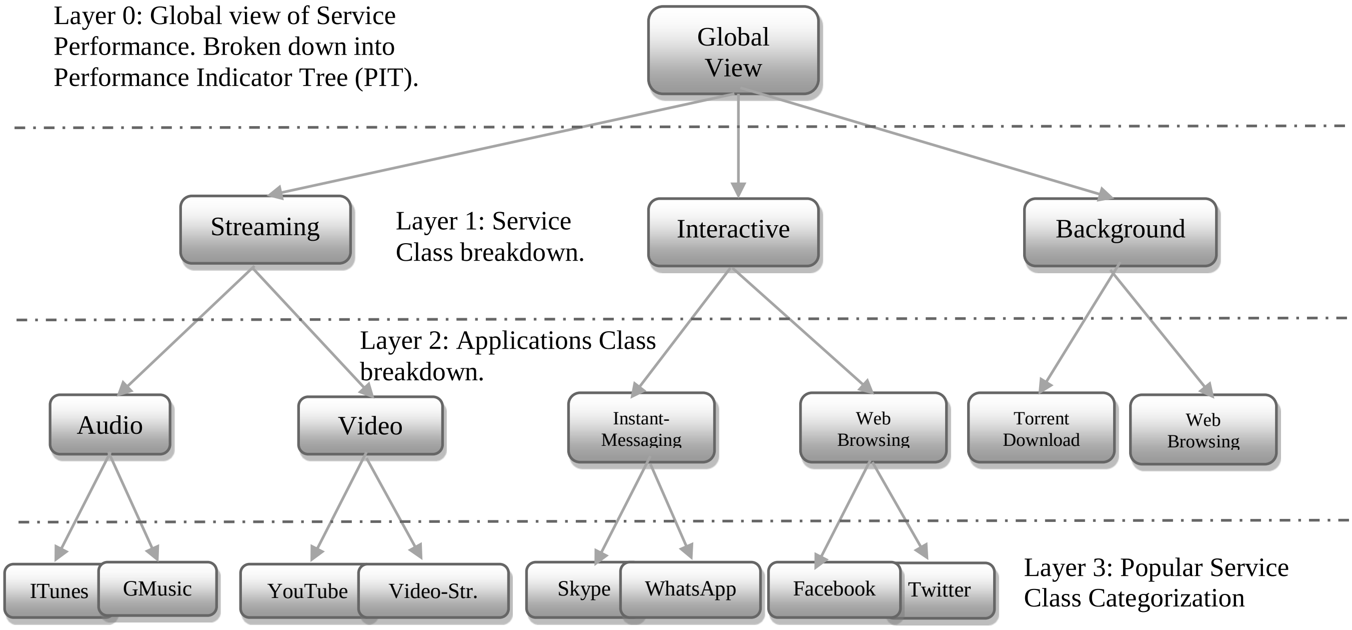

Figure 6: SQM Service Application Tree Model

For each layer, the aggregation model constitutes the framework of the BI. It indicates the impact of service quality on network cells, devices and on customers.

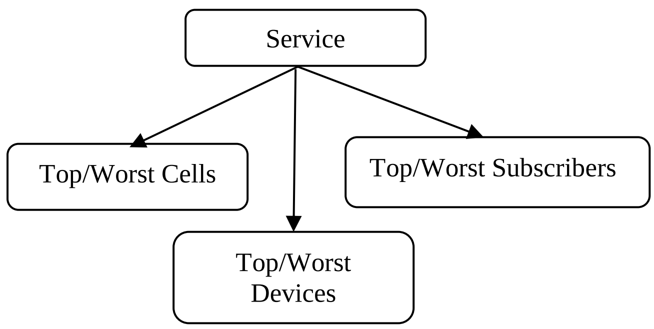

Figure 7: SQM Aggregation Model Tree

IX. Visualization of The Result

The visualization is built on the idea to show multiple views of the Network Service data. Combination of different reports elements are combined to give a deep insight of performance data in a single view for a powerful decision making platform. Mathematical models or algorithms are applied to the BI system to reduce calculation of percentages and totals that can be graphically represented by Charts, Gauges, and Maps. Taking into consideration the objectives to keep the system cost low, in other word the cost effectiveness. The visualization of the End result is done in a simple and concise way. From Layer 0 which is the Global view, the Operator can drilldown to Layer 1, and from Layer 1 to Layer 2, Layer 2 to Layer 3 for more insight on the information. The SQM Dashboard is shown in the below sections.

A. Layer 1 Use Case: Overall Network Service Quality.

In order to engage to Network Optimization and troubleshooting, it is important to have an overview of the entire network performance. The benefits of the use case is to identify underperforming services, services with poor SQI. Figure 8.

Figure 8: Layer 1 Use Case Illustration

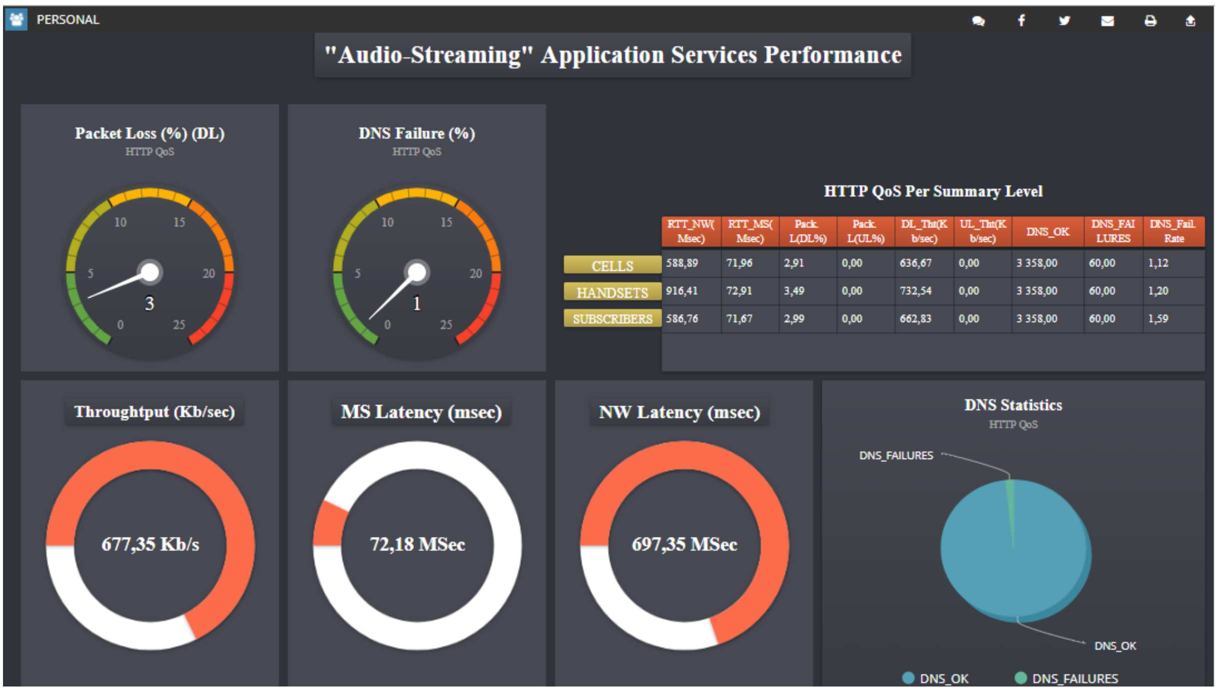

Figure 6: Use Case Layer 2 Illustration

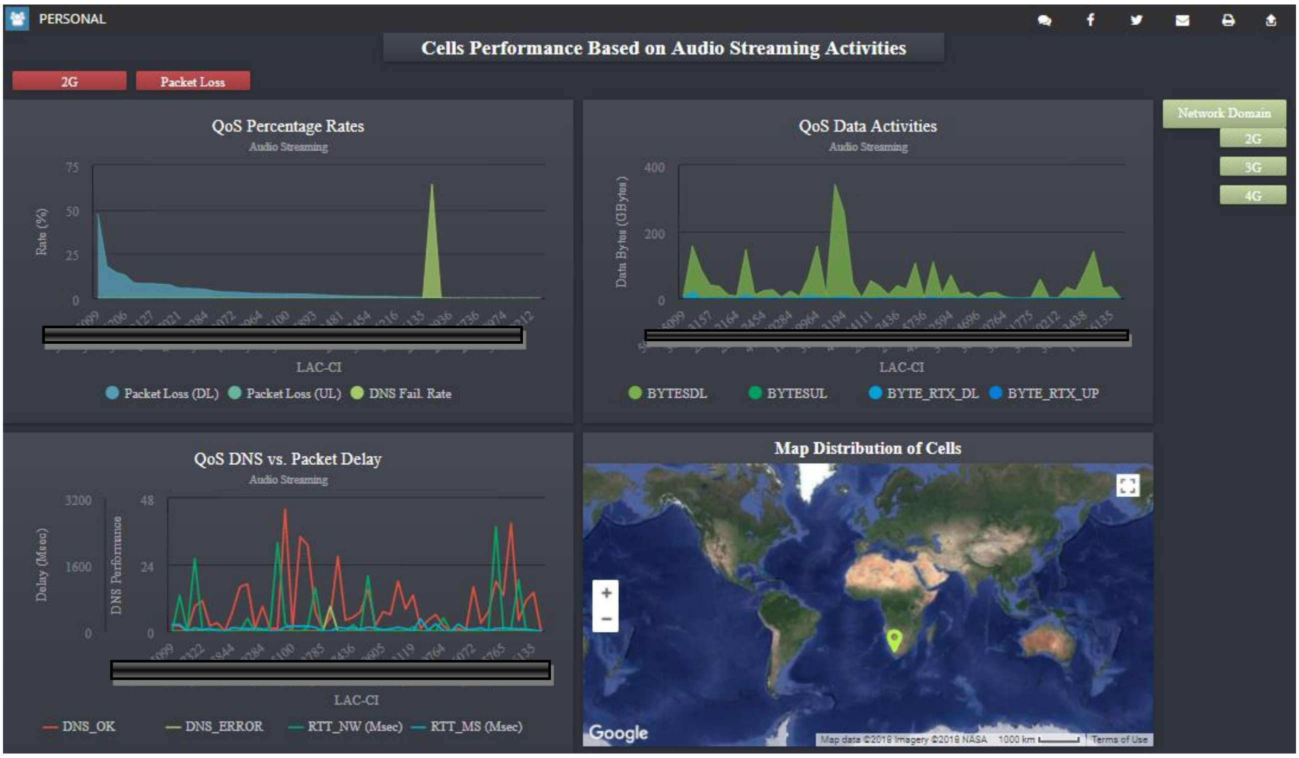

Figure 10: Layer 3 Use Case Illustration

LAC_CI	#TAC	#SUBS	RTX_DL(GB)	RTX_UL(GB)	DL(GB)	UL(GB)	DNS_OK	DNS_ERRORS	DNS_Fail.Rate%)	DL_Pack.Loss	UL_Pack.Loss	THT_DL(Kb/s)	THT_UL(Kb/s)	NW Latency(Msec)	MS Latency(Msec)
099	1,00	1,00	3,03	0,01	6,38	0,00	2,00	0,00	0,00	47,46	0,00	502,62	0,00	0,00	165,00
006	1,00	1,00	4,70	0,07	26,10	0,00	17,00	0,00	0,00	18,02	0,00	404,94	0,00	309,83	57,92
874	1,00	1,00	9,07	0,05	61,87	0,00	1,00	0,00	0,00	14,66	0,00	1 377,33	0,00	2 128,00	9,00
206	1,00	1,00	20,71	0,04	157,35	0,00	2,00	0,00	0,00	13,16	0,00	3 580,58	0,00	860,67	163,33
024	1,00	1,00	12,59	0,04	145,51	0,00	0,00	0,00	0,00	8,65	0,00	2 979,98	0,00	0,00	1,50
173	1,00	1,00	2,76	0,11	32,98	0,00	18,00	0,00	0,00	8,37	0,00	252,02	0,00	2 505,67	67,50
127	1,00	1,00	0,27	0,02	3,24	0,00	1,00	0,00	0,00	8,31	0,00	118,61	0,00	0,00	65,75
844	1,00	1,00	0,81	0,05	10,22	0,00	7,00	0,00	0,00	7,92	0,00	222,66	0,00	0,00	84,63
784	1,00	1,00	2,80	0,10	36,80	0,00	27,00	0,00	0,00	7,60	0,00	392,51	0,00	0,00	91,50
021	1,00	1,00	0,63	0,04	11,08	0,00	4,00	0,00	0,00	5,72	0,00	333,79	0,00	0,00	71,17
605	1,00	1,00	6,12	0,09	109,63	0,00	7,00	0,00	0,00	5,58	0,00	1 089,90	0,00	13,43	33,86
775	1,00	1,00	0,17	0,01	3,08	0,00	0,00	0,00	0,00	5,39	0,00	210,40	0,00	0,00	50,33
284	1,00	1,00	1,08	0,04	22,03	0,00	9,00	0,00	0,00	4,91	0,00	501,29	0,00	0,00	52,67
625	1,00	1,00	0,42	0,04	10,66	0,00	2,00	0,00	0,00	3,94	0,00	222,82	0,00	0,00	64,40
438	1,00	1,00	2,73	0,19	77,22	0,00	39,00	0,00	0,00	3,54	0,00	371,23	0,00	54,62	68,12

Figure 11: Layer 4 Use Case Illustration

B. Layer 2 Use Case: Specific Application Service Performance

Layer 2 Use case provides a specific service performance as illustrated on Figure 9. The selected service is Audio-Streaming, the dashboard shows the overall throughput, and the total volume of data used, the latency and the packet loss of the audio-streaming application. Based on this analysis, the worst

Key Performance Indicator (KPI) can be evaluated quickly to take optimization decision.

C. Layer 3 Use Case: Specific Aggregation Performance Impact of a specified Service

Layer 3 use case illustrates the performance of Audio streaming application impact on cells. In the same fashion, the handsets or customer aggregation could have been chosen. Figure 10 shows an example of such a Use case.

D. Layer 4: Technology based Performance based on a specific Service

From Layer 4, the lowest element of the model is analyzed still based on the Audio streaming service application. In a tabular format, the report shows for each cell, representing a geographical region, the amount of subscribers seen in the network, the amount of devices, and the relevant KPIs. Illustration on Figure 11.

X. CONCLUSION

This paper proposes a simple efficient, and low cost Service Quality Management (SQM) system for Network Operators using new technologies. The motivation towards this model is the constant high investment of Network Operators on plug-in tools. The approach in this paper is to leverage Cellular Network Operators Skills on new technology adoption.

With the data services emerging with the likes of twitter, Facebook, WhatsApp and other instant messaging, Online Streaming, downloading applications, combined with the increase of Mobile devices users in recent times create a driving force of Network Operator business models. In order to also stay competitive, Operators need to transform the data from the Network to information that can be used to make intelligent cost effective business decisions. We recommend that Operators need to redefine their business and organization operational strategies while taking advantage of new technologies and digital transformation. Big Data analytics, Machine Learning, BI are the technologies that Operators need to give attention to in other to stay competitive in the market and most ultimately increase good return on investment.

Different Network Operators will take different values from the network data depending on two factors:

- Business transformation, cost reduction by increasing efficiency in processes.
- Faster decision making based on real-time Network data information.

The SQM model developed in this article is a combination of Telecommunications and Data Science disciplines. The model is an introduction to an in-house SQM development that can help Network Operators to reduce the cost on Network Monitoring tools.

There are billions of transactions generated in the Network by Mobile Subscribers. With the advantage of mobility and advance Mobile functionalities, subscribers can keep activity for a long time, thus generating different kinds of signaling information in the network. All these data in the network can be a great source of information if mined correctly.

The proposals to Telecommunications Business transformation consist in leveraging the expertise where necessary, leveraging the technology in a competitive and cost effective way, making data a priority and accessible to the entire Organization.

REFERENCES

[1] A. Azzalini and B. Scarpa, Data Analysis & Data Mining, and Introduction, Oxford University Press, 2012, pp 2-8.

[2] S. Kaisler, F. Armour, J. Alberto Espinoza, W. Money, Big Data: Issues and Challenges Moving Forward

[3] Plattner, Hasso; Zeier, Alexander (2012). In-Memory Data Management: Technology and Applications. Springer Science & Business Media. ISBN 9783642295744.

[4] Hao Zhang, Gang Chen, Beng Chin Ooi, Kian-Lee Tan and Meihui Zhang, In-Memory Big Data Management and Processing: A Survey, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 7, JULY 2015.

[5] Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia, Learning Spark LIGHTNING-FAST DATA ANALYSIS, ISBN: 978-1-449-35862-4, O'Reilly, pp161-183.

[6] Adam Aspin, Business Intelligence with SQL Server Reporting Services, Apress Copyright © 2015, ISBN-13(pbk): 978-1-4842-0533-4, ISBN-13(electronic): 978-1-4842-0532-7.

[7] Jay liebowitz, Big Data and Business Analytics, CRC Press Taylor & Francis Group, FL 33487-2742 © 2013 by Taylor & Francis Group, International Standard Book Number-13: 978-1-4665-6578-4 (Hardback) pp 2- 50.

[8] Business Intelligence: Data Mining and Optimization for Decision Making Carlo Vercellis © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-51138-1.

[9] G. Gomez & R. Sanchez, End-to-End Quality of Service over Cellular Networks, Copyright © 2005 John Wiley & Sons Ltd, ISBN-13 978-0-470-01180-5 (HB), ISBN-10 0-470-01180-7 (HB), pp 10-287

[10] 3GPP TS 23.107 version 5.4.0 Release 5, Universal Mobile Telecommunications System (UMTS); Quality of Service (QoS) concept and architecture, Technical Specification.

[11] Marie-Line AM. , Sabine R. Nokia Bell Labs, Nozay, France, Quality of Experience-aware enhanced Inter-Cell Interference Coordination for Self Organized HetNet.

[12] D. Wu, Q. Wu, Y. Xu, J. Jing, and Z. Qin, ”QoE-Based Distributed Multichannel Allocation in 5G Heterogeneous Cellular Networks: A Matching-Coalitional Game Solution”, IEEE ACCESS, vol. 5, pp. 61-71, 2017