APPLICATIONS OF BIG DATA PLATFORM USING PUBLIC CLOUD

By

Name Surname

A DISSERTATION

Submitted to

The University of Liverpool

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

24/04/2016

ABSTRACT

APPLICATIONS OF BIG DATA PLATFORM USING PUBLIC CLOUD

By

Name Surname

With the current fast technology development and market change, customer churn has become one of the biggest problems for service providers in the telecom space, usually leading to shrink in market share and lower profit margins for affected service providers. Any smart company that cares about preventing customer churn will proactively care about their customers’ experiences.

Recommended content:

* Introduce into the research or technology area addressed in the dissertation
* Identify specific problems and challenges identified in your work
* Describe what is achieved and proposed in the project
* Briefly describe the structure of the report
* Mention what is the application area and benefits of the proposed solutions

DECLARATION

[check with the recent templates]

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of others.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

ACKNOWLEDGEMENTS

[any acknowledgement you want to express]

Notes about styles

It is recommended to use styles in your DS report.

If you not familiar with using and designing new styles, you can simply copy and paste parts of this documents to include necessary styles in your document.

Some good practice in formatting your text:

Use Code shaded style for your code

Use bulleted lists

* First level
* Second level

TABLE OF CONTENTS

Page

[LIST OF TABLES 3](#_Toc449485225)

[LIST OF FIGURES 4](#_Toc449485226)

[Chapter 1. Introduction 5](#_Toc449485227)

[1.1 Scope 5](#_Toc449485228)

[1.2 Problem Statement 6](#_Toc449485229)

[1.3 Approach 6](#_Toc449485230)

[1.4 Outcome 8](#_Toc449485231)

[Chapter 2. Background and review of Literature 9](#_Toc449485232)

[2.1 Background 9](#_Toc449485233)

[2.2 Literature Review 10](#_Toc449485234)

[2.3 Theory 12](#_Toc449485235)

[Chapter 3. Analysis and Design 14](#_Toc449485236)

[3.1 Design Methodology 14](#_Toc449485237)

[3.2 High Level Architecture 14](#_Toc449485238)

[3.2.1 Big Data 15](#_Toc449485239)

[3.2.2 Real time Correlator 16](#_Toc449485240)

[3.2.3 Analysis Node 16](#_Toc449485241)

[3.2.4 Data Storages 16](#_Toc449485242)

[3.2.5 Visualizations Layer: 17](#_Toc449485243)

[3.3 Use Cases 17](#_Toc449485244)

[Chapter 4. Implementation (Realization) 19](#_Toc449485245)

[4.1 Tier Architecture 19](#_Toc449485246)

[4.1.1 Presentation Tier: 19](#_Toc449485247)

[4.1.2 Business Tier 20](#_Toc449485248)

[4.1.3 Data Tier 20](#_Toc449485249)

[4.2 Environment 21](#_Toc449485250)

[4.3 Development Model 22](#_Toc449485251)

[4.4 Social Media Data sources 23](#_Toc449485252)

[4.5 The Database implementation 24](#_Toc449485253)

[4.6 Data Records and Metrics 26](#_Toc449485254)

[4.6.1 Quality Indicators 27](#_Toc449485255)

[4.6.2 Incidents 28](#_Toc449485256)

[4.6.3 Meta Info Structures 29](#_Toc449485257)

[4.7 Interface implementation 35](#_Toc449485258)

[4.7.1 Dashboard Interface Modules 40](#_Toc449485259)

[Chapter 5. Results and Evaluation 55](#_Toc449485260)

[5.1 The Final Prototype 55](#_Toc449485261)

[5.2 The Evaluation 56](#_Toc449485262)

[Chapter 6. Conclusions 57](#_Toc449485263)

[6.1 Lessons Learned 57](#_Toc449485264)

[6.2 Business Application and Limitations 57](#_Toc449485265)

[6.3 Recommendation for Future Research Work 58](#_Toc449485266)

[REFERENCES CITED 59](#_Toc449485267)

LIST OF TABLES

Table 1:Time Plan 67

Table 2: Risks Table 68

LIST OF FIGURES

Page

[Figure 1: High Level Architecture 15](#_Toc449485180)

[Figure 2: Customer Care and Service Operations use cases. 17](#_Toc449485181)

[Figure 3: Tier Architecture (Fella, 2014) 19](#_Toc449485182)

[Figure 4: Presentation Tier (MSDN, 2003) 20](#_Toc449485183)

[Figure 5: Business / Web Tier (MSDN, 2003) 20](#_Toc449485184)

[Figure 6: Data Tier (MSDN, 2003) 21](#_Toc449485185)

[Figure 7: Agile Scrum Framework (Finch, 2012) 22](#_Toc449485186)

[Figure 8: Library to APIs Architecture (Rouviere, 2012) 23](#_Toc449485187)

[Figure 9: Library java code for Facebook API 24](#_Toc449485188)

[Figure 10:SQL Script to create Users table 25](#_Toc449485189)

[Figure 11: MongoDB Database design 26](#_Toc449485190)

[Figure 12: Application Login Interface 35](#_Toc449485191)

[Figure 13: Forgot Password 40](#_Toc449485192)

[Figure 14: Geographical distribution of Worst Experiences 41](#_Toc449485193)

[Figure 15: Executive Dashboard 42](#_Toc449485194)

[Figure 16: System Preferences GUI 43](#_Toc449485195)

[Figure 17: Search by pre-defined Duration 45](#_Toc449485196)

[Figure 18: Search by custom time range 45](#_Toc449485197)

[Figure 19: Customer Care Dashboard Overview 47](#_Toc449485198)

[Figure 20: Incidents by Service by location 48](#_Toc449485199)

[Figure 21: Incidents by Location 49](#_Toc449485200)

[Figure 22: Quality Indicator Menu 50](#_Toc449485201)

[Figure 23:Legends 50](#_Toc449485202)

[Figure 24:Logging Out 52](#_Toc449485203)

[Figure 25:User Account Settings 53](#_Toc449485204)

[Figure 26: Trouble Ticketing 53](#_Toc449485205)

[Figure 27: Incidence by location 70](#_Toc449485206)

[Figure 28: Customer Care Quality Indicators 70](#_Toc449485207)

[Figure 29:Incidents by Service 71](#_Toc449485208)

[Figure 30:Service Operations Dashboard 71](#_Toc449485209)

[Figure 31:Service Operations Incidents 72](#_Toc449485210)

# Introduction

Customers experience is one of the greatest concerns of telecoms services providers nowadays. Just like in every business, rising cost of churn and customer dissatisfaction lead to reduced profit margin and market share. Hence, most telecom services providers are paying much attention to customers’ experience in order to secure the loyalty of their existing customers, attract new customers, and increase market share.

Over the past few decades, several methods have been developed to improve customers experience; one very common method is giving away incentives to customers. However, as the internet becomes more popular and many people are using the web for their day-to-day activities, there is an increasing amount of big data available and it has become obvious that using big data to understand and improve customers’ experience is inevitable and more reliable.

As big data takes the center stage, it becomes more useful in modeling a behavioral-based analytics for customers and this analytics data is targeted at proactively improving the customers experience in real-time, instead of reacting to complaints when it could have been late.

## Scope

This work covers the study on how big data method is applicable in telecoms OSS specifically focusing on the aspect of customer experiences analytics. The study compares the traditional method of customer experience analytics with a new proposed solution. The big data sources that will be applicable and discussed in this study are the popular social media data platforms.

## Problem Statement

There have been several customer experience analytics (CEA) solutions claiming to utilize and use big data methods to assist telecom network operators in their customer experience management demands but these solutions simply make use of the data generated by network equipment and users’ devices. They do not really represent the actual experience of the customers or indicate the actual area of services the customers will like to see improvements. Hence, these solutions are not quite efficient and could be misleading, causing operators to spend CAPEX and OPEX in the wrong places. This study was born out of the need to improve on customer experiences analytics and make it more efficient, evolve the solutions from using traditional data generated by customers’ devices and other mobile network equipment to using big data generated from the actual feedback and comments from consumers through the various social media and internet platforms.

## Approach

This research work is on customer experience analytics; which is one of the applications of big data methods in telecoms OSS subsystem, starting this study requires an in-depth study of the telecoms OSS architecture in a mobile network environment.

Extensive research is done through the following:

* Reviews and Research of journals and books on big data analytics, and customer experience analytics.
* Reviews and research on existing customer experience analytics solutions
* Reviews and research on TMF forum e-journals of big data building blocks
* Study on cloud platform architecture of big data solutions

The analysis and design of the new proposed system is done using information gathered from the various studies and researches. Data from social media and other internet platforms is used as input data.

|  |  |
| --- | --- |
| Step | Short Description |
| Hypothesis | Big Data methods and tools can be used by telecom service providers to improve customer experiences within their wireless networks and allow for both advanced business reporting and predictive analytics for customer facing services.  Correct and careful selection of various big data sources from different social media and internet platforms can enrich and improve analysis and prediction of customers’ experiences and behavioral patterns. With Big Data Advanced Analytics, Telco can be able to predict future behavioral pattern of their customers and the networks helping the Telco to take the right business decisions that will suite the business. |
| Research Methods | Review related literature on Big Data Analytics and its application to telecommunication systems management  Review literature on application of Social and internet search data for Big Data Analytics  Case study and industry Research.  Development of software prototype to solve specific case study application |
| IT Artefact | Taxonomy of Data Analytics methods and tools applicable for telecom analysis  Enhances architecture of a traditional Operations Support System Solution.  Software prototype of the customer experience analytics solution |
| Evaluation | The evaluation will be based on the IT Artefact (prototype) which will be able to present an application of Data Analytics for Telecoms in the customer experience use cases  The proposed big data use cases will be evaluated against selected known use cases. |

## Outcome

This research aims at ensuring the correct representation of customers’ experiences of the services and products delivered by telecom service providers. At the end of the study, a detailed analysis and design of the new solution will be available; to show how to implement the new approach to customer experience analytics solution.

In additional, aim of the project also includes a comparative study of the traditional customer management analytics and the new proposed method; this also reveals the efficiency of the new solution.

# Background and review of Literature

## Background

With the fast and continuous evolution of telecom mobile networks, services providers and operators are facing new challenges to deliver the quality of experience expected by their customers. The costs for customer care are growing much higher on a yearly basis and the percentage number of issues resolved during the first customer care call are getting lower, while the percentage of dissatisfied customers are growing.

## Literature Review

Organizations are getting to recognize the importance of big data more and more, not just from the regular traditional data sources but also from other sources. Chen et al. (2012), based on a survey of about 4,000 information technology professionals working in 93 countries across 25 different industries, concluded that big data in business analytics is one of the four major technology trends in the 2010s.

## Theory

In today’s telecom industry, there is a rising need to focus more on the customers’ experiences ever than before, the telecom operator businesses need the right customer centric solution for customer experience analytics, traditional customer experience analytics solutions only focused on mobile network equipment and devices to obtain the required big data sources for customer experience analytics solution. One may argue that the traditional solutions for customer experience analytics are not exploring the potentials of the big data properly, there is a lot more that we can achieve from the whole lot of big data already available.

# Analysis and Design

## Design Methodology

In this project, we will make use of object-oriented methodology in the design, this will ensure the system design is easy to understand and relate with, since humans too references to everything as objects.

## High Level Architecture

This section describes the high-level architecture of this customer experience analytics. This solution is, designed to gather metrics and events from social media and internet platforms. It looks at what issues the customers are facing and can categorize issues in terms of geographical location, services types, timeline etc. and monitors improvements or degradation of the customer experiences and perceptions. The system gives the customer care the ability to correlate incidents recorded or reported via social media platforms with incidents reported within the network through data generated by the wireless network equipment. Figure 1 below shows the high-level architecture of the solution.

### Big Data

The big data component of the architecture consists of data from different sources such as social media platforms and internet data sources; the input at this stage is unstructured data, which will be sent to the real time correlator for further processing.

## Use Cases

This customer experience analytic solution supports three main use cases; these are:

* Customer care use cases
* Service Operations use cases
* Marketing analytics use cases

# Implementation (Realization)

## Tier Architecture

This section describes the architectural implementation of the proposed solution of the customer experience analytics. The solution is implemented as a 3-tier architecture, namely the presentation tier, business tier and data tier. The 3-tier architecture ensures an efficient client-server computing system, each tier can be developed independently and evolve without without having impact on the other tier. Figure 3 (Fella, 2014) below describes the 3-tiers

### Presentation Tier:

The first tier is the presentation tier, otherwise known as the client tier. This tier deals with the UI and visualizations functionalities. The users or clients of the software solution interacts through the presentation tier which communicates and transacts directly with the middle tier. Implementation of the presentation layer is done using HTML, CSS and JSP technologies.

Figure deleted

Figure 4: Presentation Tier (MSDN, 2003)

### Business Tier

The middle tier is the Business tier; this is responsible for managing the business logic of the end to end functionalities of the software system, it is responsible for translating and enforcing every pre-defined rules of the system.

## Environment

The server environment is a virtual machine, installed and maintained using Amazon Web Services (AWS) cloud infrastructure. The client side will access the server side using HTTP and SSH protocols as it may be required. JDBC is used for all database connections to the server.

Hardware Setup:

The hardware infrastructure instance is baseline to a T2 instance called the t2.micro of the Amazon EC2. This enables the VM instance to burst or scale up when it is required to use more than the available resources and at the same time start up with minimal hardware resources. The t2.micro configuration contains:

* 1GB of Memory
* 1vCPU
* 32/64-bit Platform

Operating System Server Environment:

The software setups in the operating system environment are listed below;

CentOS 7 x64

Glassfish 4.1.1

JDK 8 U 66

Firefox 43.0.4

MySQL DB server

MongoDB NoSQL server

## The Database implementation

The MySQL server contains only one database called **HALO\_DB,** it is used to store structured data such as user access policies and management data, and other system sensitive data which do not change often but are required for continuous operation of the system. For example, the users table containing users access and privilege details is a structured data which does not require heavy indexing and suitable to be created and stored in MySQL schema. This table **Users** is created in the MySQL schema using the SQL script shown in figure 10 below:

CREATE TABLE `users` (

`userID` int(11) NOT NULL AUTO\_INCREMENT,

`userName` varchar(50) NOT NULL,

`password` varchar(45) DEFAULT NULL,

`privilege` varchar(45) DEFAULT NULL,

`fullName` varchar(45) DEFAULT NULL,

`lastLoginDate` datetime DEFAULT NULL,

`status` tinyint(1) NOT NULL DEFAULT '1',

`description` text,

PRIMARY KEY (`userID`),

UNIQUE KEY `userName\_UNIQUE` (`userName`)

) ENGINE=InnoDB AUTO\_INCREMENT=1003 DEFAULT CHARSET=utf8;

Figure 11: Example code with Code Shaded style

### Incidents

Incidents represent Quality of Service (QoS) degradations. Based on KPI values, and reference data, the system flags incidents which identify what is the issue, why the issue was raised, the impact, and the recommended action to mitigate the problem. This information can be displayed on the GUIs.

### Meta Info Structures

Below describes some of the meta info structures of the prototypes solution.

#### Quality Indicator Meta Info Structure:

This structure is used to store information about the Quality Indicators (otherwise known as KPIs). These Quality Indicators are values typically stored in KPI table and associated with an aspect of the end customers’ service experience, for example, Video Speed/Quality.

# Results and Evaluation

## The Final Prototype

The final prototype’s source code will be stored in a remote git repository available online at [www.gitlab.com](http://www.gitlab.com). Name of the git project is **UoL-Project** and can be accessed as a private project only.

This proposed method will be able to increase revenues, reduce customer service costs, help to get first hand customer feedbacks on products and services and improve business brand.

5.x Hypothesis Evaluation Summary (last section in the chapter)

[It is expected that you confirm your hypothesis by means of demonstrating working prototype and/or other methods such as expert assessment/review]

# Conclusions

## Lessons Learned

An obvious lesson learned in this project is that the design and development of the prototype is a lot of work and requires much hard work and time than anticipated. For future work, more time or reduced scope approach is advisable for the prototype design and development.

## Business Application and Limitations

Application of data analytics will no doubt become critical tools across different works of life, ranging from medicine, education, engineering and many more. Though its application right now is still growing, the time will come when data science and analytics will become a necessity for professionals in the nearest future.

## Recommendation for Future Research Work

This research will form the beginning of many more innovative applications of big data methods to come. A future recommendation is to include reporting systems that can be used by the Management C-level, a reporting system that is easily consumable by these group of management level will be valuable, and will give more feasibility to the top management on customer experiences regarding the company’s products and services.

REFERENCES CITED

Acquino, J. (2015). Transforming Customer Experience Design with Data [Online], Available at <http://www.1to1media.com/view.aspx?docid=35462> (Accessed March 08, 2016).

Alexandrov, A., Bergmann, R., Ewen, S., Freytag, J., Hueske, F., Heise, A., Kao, O., Leich, M., Leser, U., Markl, V., Naumann, F., Peters, M., Rheinländer, A., Sax, M., Schelter, S., Höger, M., Tzoumas, K., Warneke, D. (2014). The Stratosphere platform for big data analytics, Vol. 23, Issue 6. pp. 939-964.

APPENDICES

###### Approved DS Proposal

[This appendix is mandate]

###### Design interfaces

The interface of the prototype was coded and developed using combination of html, javascripts and css, below shows some of the developed interfaces for the project prototype.

###### HALO Code Listing

This project consists of prototype developed in java, html, and css codes.

**MongoDB.Java –** This code handles connection between the application layer and the Data layer towards the MongoDB databases.

package com.halo.db;

import java.lang.reflect.Field;

import java.util.ArrayList;

import java.util.Iterator;

import java.util.List;

import java.util.Map.Entry;

import org.bson.Document;

import org.slf4j.Logger;

import org.slf4j.LoggerFactory;

import com.google.gson.Gson;

import com.google.gson.GsonBuilder;

import com.google.gson.JsonElement;

import com.google.gson.JsonObject;

import com.mongodb.MongoClient;

import com.mongodb.client.FindIterable;

import com.mongodb.client.MongoCollection;

import com.mongodb.client.MongoDatabase;

import static org.bson.Document.parse;

import static com.halo.FIELD\_REQUEST\_ID;

import static com.halo.FIELD\_REQUEST\_ACTION;

import static com.halo.FIELD\_DEVICE;

public class MongoDB {

private static MongoClient client;

private static MongoDatabase database;

private static Logger log = LoggerFactory.getLogger(MongoStub.class);

private static Gson gson = new GsonBuilder().serializeNulls().create();

public static final String HEADER = "header";

public static final String PAYLOAD = "payload";

public static void init(){

log.info("\ninit()\n");

if(client == null)

client = new MongoClient("127.0.0.1", 27017);

if(database == null)

database = client.getDatabase("ncs");

}

@Override

public void removeJSONAdapter(Class type) {

// TODO Auto-generated method stub

}

}