

**STANDARDIZED COMMUNICATION FOR BIGDATA  
ANALYTICS THROUGH JSON**

K.L.K Madushanka

(168244U)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

February 2020

**STANDARDIZED COMMUNICATION FOR BIGDATA  
ANALYTICS THROUGH JSON**

K.L.K Madushanka

(168244U)

Dissertation submitted in partial fulfilment of the requirements for the degree  
Master of Science specializing Data Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

February 2020

## DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or another medium. I retain the right to use this content in whole or part in future works.

.....

Kasun Madushanka Liyanage

.....

Date

I certify that the declaration above by the candidate is true to the best of my knowledge and the above candidate has carried out research for the Masters Dissertation under my supervision.

.....

Dr. Amal Shehan Perera

.....

Date

## **Abstract**

Big data is not a new terminology in the Information Technology sector anymore. With the emergence of big data, arise the need for analyzing large amounts of data that consist trillions of records. Additionally, big data have already penetrated multiple areas in data analytics. Therefore, different technological solutions were developed to handle these big data complexities. However, even after decades, contemporary solutions are unable to address complex issues and overcome several limitations.

Lack of a common communication standard has resulted in many issues in big data analytics. Presently, all the big data solution companies are using their in-house ad hoc communication methods to perform analytics. Unfortunately, this leads to limitations in integration and reusability of the solutions built. To overcome this, Microsoft introduced the XMLA (XML for Analysis), an industry standard for accessing data in analytical systems, namely OLAP (online analytical processing) systems. XMLA was well standardized and well designed for accessing data through Multi-Dimensional Expressions (MDX). Development of tailor-made query languages to access and analyze the stack of scattered data stores has caused the creation of different standards. This leads to the state where almost all big data services offering their proprietary query languages and APIs for data analysis.

This research is to propose a methodology for addressing the ad-hoc integration of these big data analytics endpoints through a JSON based specification by reusing XMLA structures. The research components are publishing a communication model using JSON specification and proposing to adopt the standards to existing stores. This solution will enable frontend tools to be fully independent of the backend storage model. Also, this will allow existing JSON standardized frontend tools to easily integrate with big data analytics through eliminating the necessity of a specific frontend tool aiming a data store.

Keywords: Big Data Communication, JSON Based Communication, JQA Specification

## **ACKNOWLEDGEMENT**

I would like to take this time to sincerely appreciate the people who helped and guided me in this research. First, Dr. Amal Shehan Perera, my supervisor, for the supervision and guidance provided throughout the research and all the MSc module lecturers for encouraging and motivating me to complete this dissertation.

For Zone24x7.inc and the management for offering me flexibilities and support for the MSC initiation and related work and for my family for always trusting me and supporting me in my educational journey.

I also wish to thank all my colleagues and friends for all their help, support, interest and valuable advice. Finally, I would like to thank all others whose names are not listed particularly but have given their support in many ways and encouraged me to make this a success.

# TABLE OF CONTENT

DECLARATION	i
Abstract	ii
ACKNOWLEDGEMENT	iii
TABLE OF CONTENT	iv
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF ABBREVIATIONS	viii
INTRODUCTION	1
1.1    Big Data and Analytics	2
1.2    Big Data Analytics and Communication	3
1.3    Research Problem	6
1.4    Research Objectives	7
1.5    Summary	8
LITERATURE SURVEY	9
2.1    JavaScript Object Notation	10
2.2    XML for Analysis and OLAP	12
2.3    JavaScript Object Notation and Big Data Stores	13
2.3.1    Apache Kylin	14
2.3.2    Druid	15
2.3.3    ElasticSearch	17
2.3.4    Data Services	18
2.4    Summary	19
METHODOLOGY	20
3.1    JSON Queries for Analysis	21

3.2	JQA Implementation	23
3.3	Summary	25
	CASE STUDY, RESULTS AND OBSERVATIONS	26
4.1	Case Study	27
4.2	Dashboard, Results and Observations	28
4.3	Summary	32
	CONCLUSION	33
5.1	Challenges and Limitation	35
5.2	Future Work	36
	REFERENCES	38

## LIST OF FIGURES

	Page
Figure 1.1 : OLAP Design with SQL Data Sources Separating each Layer	5
Figure 2.1 : Simple JSON object with key/value pair	11
Figure 2.2 : Simple JSON List	11
Figure 2.3 : Multi Types JSON Object	11
Figure 2.4 : Sample Discovery request [23]	13
Figure 2.5 : Sample Execute request [23]	13
Figure 2.6 : Apache Kylin sample request [24]	14
Figure 2.7 : Apache Kylin sample response [24]	15
Figure 2.8 : Druid sample request [13]	16
Figure 2.9 : Druid sample response [13]	16
Figure 2.10 : Elasticsearch sample request [7]	17
Figure 2.11 : Elasticsearch sample response [7]	17
Figure 3.1 : JQA communication overview	22
Figure 3.2 : JQA store Meta method	23
Figure 3.3 : Search API list method	24
Figure 3.4 : Aggregation API Syntax	24
Figure 3.5 : Aggregation Count Syntax	25
Figure 4.1 : Dashboard Overview	29
Figure 4.2 : Widget Configurator	31



## LIST OF TABLES

	Page
Table 4.1 Frontend development effort in hours	29
Table 4.2 Backend development effort in hours	30

## LIST OF ABBREVIATIONS

Abbreviation	Description
<b>API</b>	Application Programming Interface
<b>ETL</b>	Extract, Transform, and Load
<b>HTTP</b>	Hypertext Transfer Protocol
<b>IoT</b>	Internet of Things
<b>JSON</b>	JavaScript Object Notation
<b>MDX</b>	Multidimensional Expressions
<b>NoSQL</b>	Not Only SQL
<b>OLAP</b>	Online Analytical Processing
<b>RPC</b>	Remote Procedure Call
<b>SOAP</b>	Simple Object Access Protocol
<b>SQL</b>	Structured Query Language
<b>SSAS</b>	SQL Server Analysis Services
<b>WSP</b>	Web-Service Protocol
<b>WWW</b>	World Wide Web
<b>XML</b>	Extensible Markup Language
<b>XMLA</b>	XML for Analysis
<b>JQA</b>	JSON Queries for Analysis
<b>DaaS</b>	Data As a Service