Project Tender

Project Name: Data Lake

Client Name:
Willem Greyling Head of Architecture at RMB

Team Name:
IT Admirals

Team Members:
Mr Sboniso Masilela
Mr William Seloma
Miss Martha Mohlala

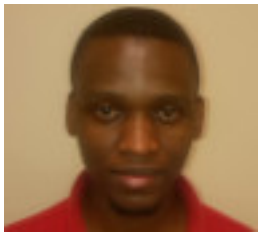
Tender submission Date: 4 May 2015

Contents

1 The Team

1.1 Mr W. Seloma

William Seloma **May 1, 2015**



Phone Number
074 259 3912

E-mail Adress
selomawill@gmail.com

Education

- **Vukuzame F.E.T School** Empumalanga, South Africa
  Grade 12 *2010*
- **University of Pretoria** Pretoria, South Africa
  Final year Bis Multimedia *2015*

Objectives

- I am a dedicated Person who seeks to Use my skills and education to develop good Software that would simplify life for people, I am eager to learn and to be challenged in order to develop my problem solving skill and my programming skills and also to grow as a person.

Experience

1. Web development using HTML, CSS, jquery and Javascript
2. Game development using Unity and C

Skills

1. WEB DEVELOPMENT (LANGUAGES BELOW)
   - HTML 5
   - CSS / CSS5
   - Javascript, Jquery and Ajax
   - PHP

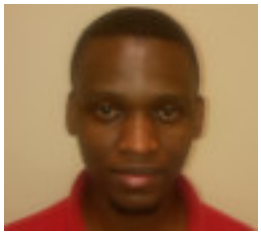
- XML, XSLT, XML Path
- SQL

2. PROGRAMMING (LANGUAGES BELOW)

- Java
- C/C++
- C

1.2 Mr S. Masilela

William Seloma **May 1, 2015**



Phone Number
074 259 3912

E-mail Adress
selomawill@gmail.com

Education

- **Vukuzame F.E.T School** Empumalanga, South Africa
  Grade 12 *2010*
- **University of Pretoria** Pretoria, South Africa
  Final year Bis Multimedia *2015*

Objectives

- I am a dedicated Person who seeks to Use my skills and education to develop good Software that would simplify life for people, I am eager to learn and to be challenged in order to develop my problem solving skill and my programming skills and also to grow as a person.

Experience

1. Web development using HTML, CSS, jquery and Javascript
2. Game development using Unity and C

Skills

1. WEB DEVELOPMENT (LANGUAGES BELOW)
   - HTML 5
   - CSS / CSS5
   - Javascript, Jquery and Ajax
   - PHP
   - XML, XSLT, XML Path
   - SQL

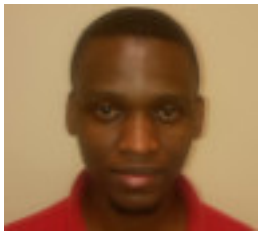
2. PROGRAMMING (LANGUAGES BELOW)

- Java
- C/C++
- C

1.3 Miss M. Mohlala

William Seloma

May 1, 2015



Phone Number
074 259 3912

E-mail Adress
selomawill@gmail.com

Education

- **Vukuzame F.E.T School** — Empumalanga, South Africa
  Grade 12 — *2010*
- **University of Pretoria** — Pretoria, South Africa
  Final year Bis Multimedia — *2015*

Objectives

- I am a dedicated Person who seeks to Use my skills and education to develop good Software that would simplify life for people, I am eager to learn and to be challenged in order to develop my problem solving skill and my programming skills and also to grow as a person.

Experience

1. Web development using HTML, CSS, jquery and Javascript
2. Game development using Unity and C

Skills

1. WEB DEVELOPMENT (LANGUAGES BELOW)
   - HTML 5
   - CSS / CSS5
   - Javascript, Jquery and Ajax
   - PHP
   - XML, XSLT, XML Path
   - SQL

2. PROGRAMMING (LANGUAGES BELOW)

- Java
- C/C++
- C

2 Project execution

2.1 Development Methodology

Agile software development methodology.

We intend on releasing new software at the end of every iteration and only Agile software development is best suited for this and the reviewing of software priorities at the end of every iteration will keep us focused on the bigger picture(Data Lake).

2.2 Informing the client about the project

The client has planed to meet on a monthly bases, that being stated we will ensure to keep in touch via email and Skype if the client is willing to have extra meetings for us to show our progress since were meeting only on a monthly bases.

Providing feed back to our client is essential to the type of development methodology we have chosen, and this we feel ensures a fully functional software which is in line with the clients liking at the end of the development of the software.

2.3 Solving technical challenges

We noticed that we are dealing with big data where we receive a lot data from our sources which is unstructured(in this case) and of different formats, we intend on building a data system that will be able to structure and store the data(using technologies like Hadoop) into a data lake and present a logically structured version of the data.
The data system will be:

- Fault tolerant
- Cost effective
- Flexible

Hadoop is designed to deal with large clusters of data and of any type and all the above points are catered for by Hadoop.
The solution we propose will allow any application to interact with our data system and the data system will provide organised information to the application.

This will make the application interacting with our data system able to adhere to the following usability goals:

1. Effectiveness - make the product good at what its supposed to do.
2. Efficiency - help to increase productivity.
3. Utility - provide the functionality that the users want/need.
4. Learnability - Make it easy for user to learn and use the product.
5. Intuitive - Make it easy to use and understand the application.

2.4 Technologies we will use

1. Schema-less databases, or NoSQL databases: this is one of the technologies suited to handle big data.
2. MapReduce also suited for handling large clusters of data.

2.5 Implementation Plan

This aims to outline the System Development Life-Cycle (SDLC) of the proposed project solution. The project team will be submitting the following documents (further documentation to be included at a later stage):

Requirements Specification document

- This document will provide a comprehensive description of the data lake.
  This requirement specification document will give a detailed description of the purpose of the system, system features and interfaces and formally stipulate the systems functional and non-functional requirements, its data requirements, quality requirements and the constraints under which it should operate.

Architectural Specification Document

- This document will provide a comprehensive description of data lake system. This architectural specification will give a detailed view of the purpose of the data lake system with regard to its overall architecture and architectural features. This will then formally stipulate the subsystem views, policies, its data requirements, as well as the constraints under which the system operates.

Design Specification Document

- This document will provide a design specification for the Data Lake system, which will specify the design and layout of the system as specified in the related documents below. This design specification should help with the planning and implementation of a product which adheres to the following usability goals:

  - Effectiveness - is the product good at what its supposed to do?
  - Efficiency - does the product help to increase productivity?
  - Utility - does the product provide the functionality that the users want/need?
  - Learnability - is it easy to learn to use the product?
  - Intuitive - is it easy to use and understand? etc.

Included in all the above stipulated documentation will be:

1. Unified Modelling Language (UML) diagrams to further illustrate the flows, operations, sequences etc. that form part of the system's operation.
2. Database Entity Relationship (ER) diagrams to illustrate the system entities that will form part of the system's persistent data.

Users and installation manual

These documents will provide detailed instructions of the installation and the use of the system to ensure effective and efficient use of the system.

3 References

1. Beal, Vangie. 'What Is Audit Trail? Webopedia'. Webopedia.com. N.p., 2015. Web. 9 Mar. 2015.

2. Rodrigues, T. (2012). 10 emerging technologies for Big Data. [online] TechRepublic. Available at: `http://www.techrepublic.com/blog/big-data-analytics/10-emerging-technologies-for-big-data/` [Accessed 1 May 2015].

3. Www-01.ibm.com, (2015). IBM What is Hadoop? United States. [online] Available at: `http://www-01.ibm.com/software/data/infosphere/hadoop/` [Accessed 1 May 2015]