Governors State University
OPUS Open Portal to University Scholarship

Fall 2015

Data Framework Management System

Firasat Ali Mohammed
Governors State University

Ahmad Munir Rizwi Syed
Governors State University

Follow this and additional works at: http://opus.govst.edu/capstones

Part of the Databases and Information Systems Commons

For more information about the academic degree, extended learning, and certificate programs of Governors State University, go to
http://www.govst.edu/Academics/Degree_Programs_and_Certifications/

Visit the Governors State Computer Science Department

Table of Contents

1 Project Description

The aim of this project is to design a system for managing multiple data source through networks. The focus is on applying best practices using the ITIL framework to achieve reliability. The IT Infrastructure Library (ITIL) Framework, managed by AXELOS Limited, defines a commonly used, best-practice approach to IT Service Management (ITSM).

1.1 Project Abstract

The goal of this project is to design a system for managing multiple data source through networks. The end user is concerned about the computations that depend on information from several data sources. The access will be from a portal. The system should use grid computing standards to fetch information from the different information sources, consolidates them, presented them as required.

1.2 Competitive Information

Qlik Sense and Snowflake are the major web based cloud systems which are close competitors.

1.3 Relationship to Other Applications/Projects

The application interacts with AWS. Amazon Web Services own and maintain the network-connected hardware required for these application services. Cloud Computing provides a simple way to access servers, storage, databases and a broad set of application services over the Internet.

1.4 Assumptions and Dependencies

General Architecture principles based on past experiences and agile models best practices & methodologies will be used in designing the solution.
The basic TCP/IP (HTTP) protocol will be the only one used to access the application
The web browser will be the primary client used by employees and public users
This system supposed to be a web based application which can be accessed 24/7.
Network connection should be available to use the application.

1.5 Future Enhancements

Future customization will include features that will simplify data storage process. It is very important for the code to be customizable due to the distinct needs of different organization. The application will be integrated with the cloud.

1.6 Definitions and Acronyms

DFMS : Data Framework Management System
DB : Data Base
ITIL : The IT Infrastructure Library (ITIL) Framework
PHP : PHP Hypertext Preprocessor
AWS : Amazon Web Services
HTML: Hyper Text Markup Language
FR – Functional Requirement
SOW – Statement of Work
CS – Computer Science
WAMP: Windows, Apache, MySQL, PHP

2.0 Technical Description

AWS : Cloud Computing provides a simple way to access servers, storage, databases and a broad set of application services over the Internet. Cloud Computing providers such as Amazon Web Services own and maintain the network-connected hardware required for these application services, while you provision and use what you need via a web application.

MySQL Workbench 6.2

MySQL Workbench 6.2 is comprehensive, integrated database system that enables organizations to reliably manage mission-critical information and confidently run today's increasingly complex business applications. MySQL Workbench 6.2 allows companies to gain greater insight from their business information and achieve faster results for a competitive advantage.

MySQL Workbench functionality:

- **SQL Development**: Enables you to create and manage connections to database servers while help in storing data sources. Along with enabling you to configure connection parameters, MySQL Workbench provides the capability to execute SQL queries on the database connections using the built-in SQL Editor.

- **Data Modeling (Design)**: Enables you to create models of your database schema graphically, reverse and forward engineer between a schema and a live database, and edit all aspects of your database using the comprehensive Table Editor. The Table Editor provides easy-to-use facilities for editing Tables, Columns, Indexes, Triggers, Partitioning, Options, Inserts and Privileges, Routines and Views.

- **Server Administration**: Enables you to administer MySQL server instances by administering users, performing backup and recovery, inspecting audit data, viewing database health, and monitoring the MySQL server performance.

- **Data Migration**: Allows you to migrate from Microsoft SQL Server, Microsoft Access, Sybase ASE, SQLite, SQL Anywhere, PostreSQL, and other RDBMS tables,

objects and data to MySQL. Migration also supports migrating from earlier versions of MySQL to the latest releases.

- **MySQL Enterprise Support**: Support for Enterprise products such as MySQL Enterprise Backup and MySQL Audit.

2.1 Project/Application Information flows
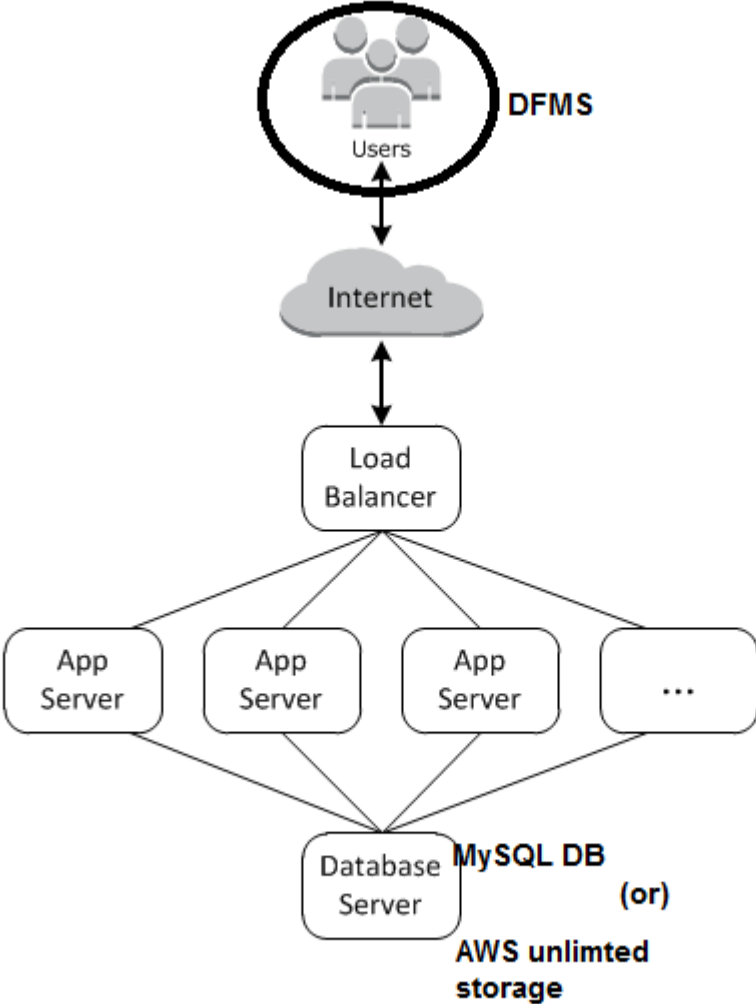
2.2 Interactions with other Projects (if Any)

The data frame work management systems interacts with Amazon Web Services (AWS). AWS is a collection of remote computing services, also called web services, that make up a cloud platform offered by Amazon.com. These services operate from 11 geographical

regions across the world. The most central and well-known of these services arguably include Amazon Elastic Compute Cloud, also known as "EC2", and Amazon Simple Storage Service, also known as "S3". Amazon markets AWS as a service to provide large computing capacity more quickly and more cheaply than a client company building an actual physical server farm.

.

2.3 Interactions with other Applications

The DFM S application interacts with Amazon Web services. Below diagram explains how the interactions between the end user and Amazon RDS are implemented.
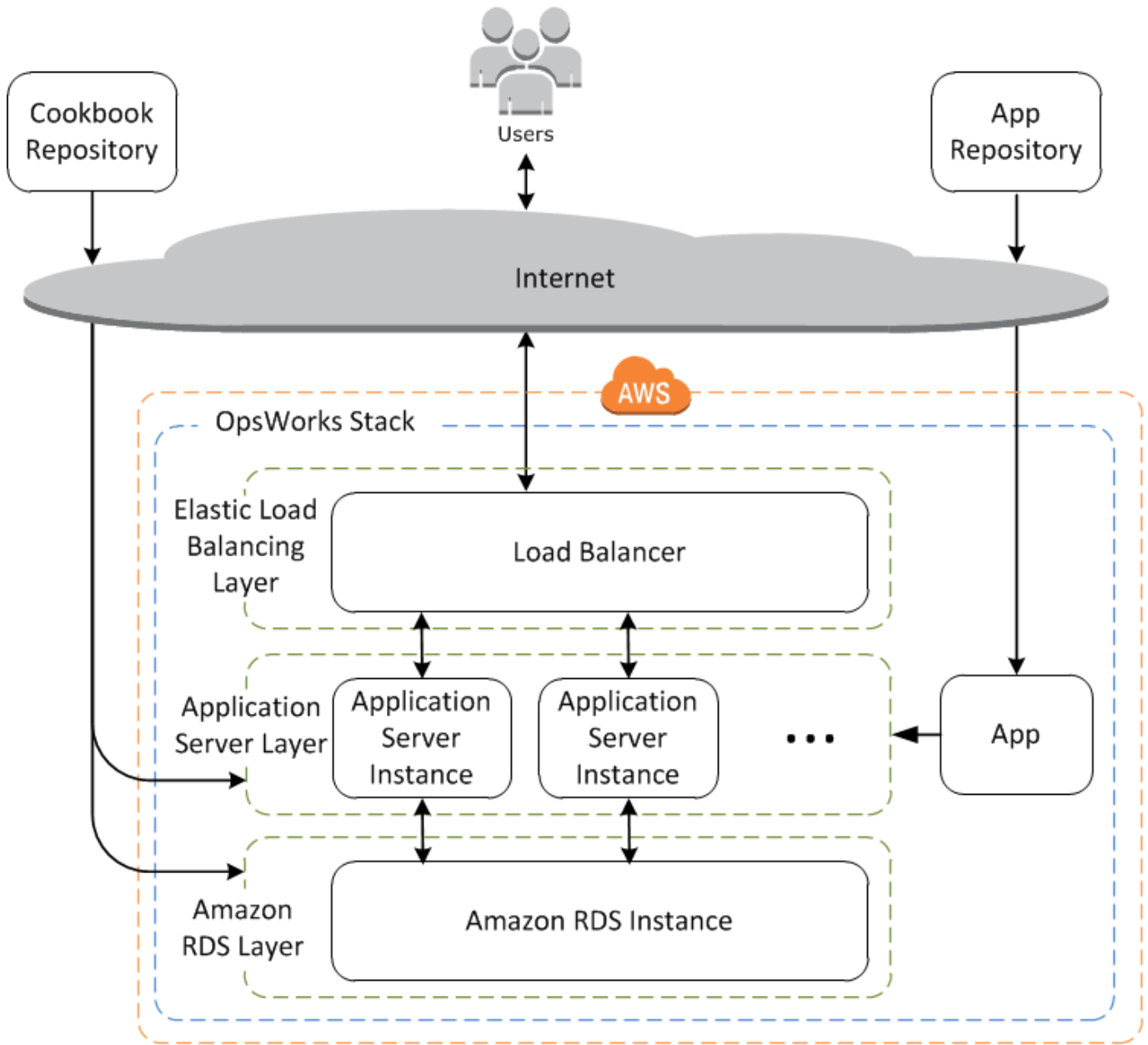
2.4 Capabilities

The DFMS application has the capability to store the data source.
The end user can use the information to merge or analyze the information.
The application has the ability to store the data sources in various formats.
The storing feature of the application is made after careful review of the existing

application available in the market.

3 Project Requirements

3.1 Identification of Requirements:

\<GSU-001-0.1 data source 000101>

The application should have the ability to store data sources.

\<GSU-001-0.1 data source format /000102>

The application should be compatible with various data source formats.

\<GSU-001-0.1 adminstrator/000103>

The webapplication should have the secure login funtionality.

\<GSU-001-0.1 registration /000104>

The user (guest) should be able to register if not a registered user.

\<GSU-001-0.1 view files /000105>

Previously uploaded information visible to end user.

\<GSU-001-0.1 feedback/000106>

The system should be able to capture the user feedback

3.1Operations, Administration, Maintenance and Provisioning (OAM&P)

i) The ability to build and maintain positive relationships with team-members and stakeholders.

ii) Mandatory implementation of the client or supervisor feedback.

iii) The desire and ability to change one's approach to project management and/or course of action in response to business needs.

vii) The ability to get things done efficiently and effectively

3.2 Security and Fraud Prevention

The application coding has been designed to prevent code injection attacks. When the application is interacting with cloud, then the data exchange will happen over HTTPS.

3.3 Release and Transition Plan

Release Process

This statement of work (SOW) shall commence on August 21st, 2015 and shall continue until 29th Nov, 2015.

Below are the timelines and Deliverable details.

- Interim Project I has been given on Nov 03, 2015

- Final Project II been given on Nov 29, 2015

3.4 Transition Schedule

		Member1	Member2
Phase I			
	Project Plan	x	X
	System Requirement Specification		X
	Prototype	x	
	User Manual		
	Presentation	x	X
	Document Formatting	x	
Phase II			
	System Requirement Specification Refinement		X
	Process and Product Specification	x	
	User Manual		X
	Implementation		
	Document Formatting	x	

4 Project Design Description

The purpose of this system is to support the end user or data scientist in storing and analyzing the data sources in multiple formats. Since DFMS is an online system, it can be easily accessed from web-browser with internet access, thus removing any constraints of time or place. The system also sends relevant notifications and information to respective users through emails when they contact for any help. The system will have a user friendly interface which will make it easier for all kinds of participants.

Using Data Frame work management system, you can get your data stored and sorted quickly. In addition to that as the application has been integrated with AWS, it provides you with a comprehensive ecosystem of technology and consulting partners through our AWS Partner Network and AWS Marketplace. Our partners have created innovative big

data analytics solutions or have implemented successful big data platforms for other customers in the AWS Cloud.

5 Project Internal/external Interface Impacts and Specification

i) Structure :

Structure is an internal Factor that impacts operation of Data Framework Management System.

ii) External Communications:

The impact web application has while interacting with cloud services such as AWS have to be closely monitored to achieve performance.

6 Project Design Units Impacts

The Techniques we will be using in our project will be focused on Agile Model driven development, the iterative nature of agile development means features are delivered incrementally, enabling some benefits to be realized early as the product continues to develop. A key principle of agile development is that testing is integrated throughout the lifecycle, enabling regular inspection of the working product as it develops. This allows the product owner to make adjustments if necessary and gives the product team early sight of any quality issues. Agile development principles encourage active 'user' involvement throughout the product's development and a very cooperative collaborative approach. This provides excellent visibility for key stakeholders, both of the project's progress and of the product itself, which in turn helps to ensure that expectations are effectively managed.

6.1 Functional Area/Design Unit A

The objective of the project is to develop a data framework management system within allocated time and specified quality. We combine 2 data sources to generate user desired output.
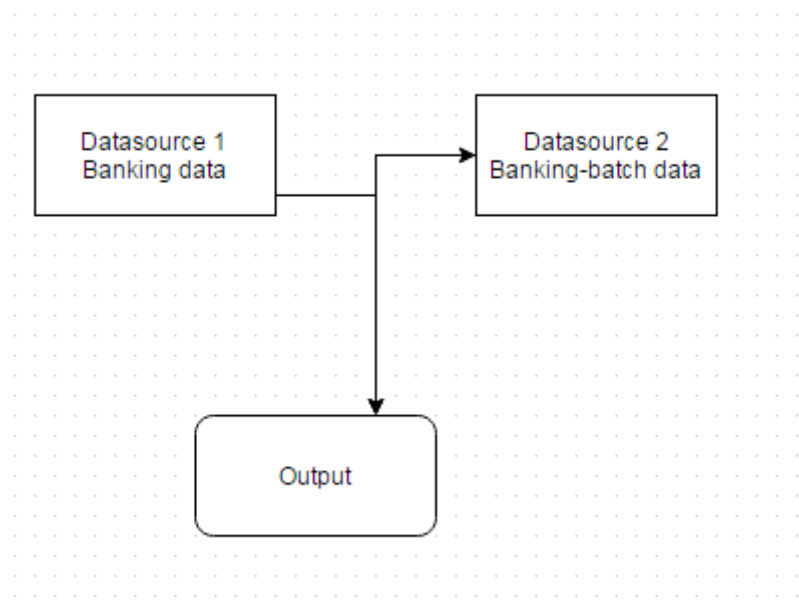
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp_var_	cons_pric	cons_conf	euribor3m
2	44	blue-colla	married	basic.4y	unknown	yes	no	cellular	aug	thu	210	1	999	0	nonexiste	1.4	93.444	-36.1	4.963
3	53	techniciar	married	unknown	no	no	no	cellular	nov	fri	138	1	999	0	nonexiste	-0.1	93.2	-42	4.021
4	28	managem	single	university	no	yes	no	cellular	jun	thu	339	3	6	2	success	-1.7	94.055	-39.8	0.729
5	39	services	married	high.scho	no	no	no	cellular	apr	fri	185	2	999	0	nonexiste	-1.8	93.075	-47.1	1.405
6	55	retired	married	basic.4y	no	yes	no	cellular	aug	fri	137	1	3	1	success	-2.9	92.201	-31.4	0.869
7	30	managem	divorced	basic.4y	no	yes	no	cellular	jul	tue	68	8	999	0	nonexiste	1.4	93.918	-42.7	4.961
8	37	blue-colla	married	basic.4y	no	yes	no	cellular	may	thu	204	1	999	0	nonexiste	-1.8	92.893	-46.2	1.327
9	39	blue-colla	divorced	basic.9y	no	yes	no	cellular	may	fri	191	1	999	0	nonexiste	-1.8	92.893	-46.2	1.313
10	36	admin.	married	university	no	no	no	cellular	jun	mon	174	1	3	1	success	-2.9	92.963	-40.8	1.266
11	27	blue-colla	single	basic.4y	no	yes	no	cellular	apr	thu	191	2	999	1	failure	-1.8	93.075	-47.1	1.41
12	34	housemai	single	university	no	no	no	telephone	may	fri	62	2	999	0	nonexiste	1.1	93.994	-36.4	4.864
13	41	managem	married	university	no	yes	no	cellular	aug	thu	789	1	999	0	nonexiste	1.4	93.444	-36.1	4.964
14	55	managem	married	university	no	no	no	cellular	aug	mon	372	3	999	0	nonexiste	1.4	93.444	-36.1	4.965
15	33	services	divorced	high.scho	no	yes	no	cellular	may	tue	75	5	999	0	nonexiste	-1.8	92.893	-46.2	1.291
16	26	admin.	married	high.scho	no	no	yes	telephone	jun	mon	1021	1	999	0	nonexiste	1.4	94.465	-41.8	4.96
17	52	services	married	high.scho	unknown	yes	no	cellular	jul	thu	117	2	999	0	nonexiste	1.4	93.918	-42.7	4.962
18	35	services	married	high.scho	no	no	no	cellular	apr	thu	1034	2	999	0	nonexiste	-1.8	93.075	-47.1	1.365
19	27	admin.	single	university	no	no	no	telephone	oct	tue	540	1	999	0	nonexiste	-0.1	93.798	-40.4	4.86
20	28	blue-colla	married	basic.9y	unknown	no	no	telephone	may	thu	140	1	999	0	nonexiste	1.1	93.994	-36.4	4.86
21	26	unemploy	single	basic.9y	no	yes	yes	cellular	jul	mon	104	4	999	0	nonexiste	1.4	93.918	-42.7	4.96
22	41	unemploy	married	basic.9y	unknown	yes	no	telephone	apr	fri	246	1	999	1	failure	-1.8	93.075	-47.1	1.405

DataSource 1

14

	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp_var_	cons_pric	cons_conf	euribor3m
2	30	blue-colla	married	basic.9y	no	yes	no	cellular	may	fri	487	2	999	0	nonexiste	-1.8	92.893	-46.2	1.313
3	39	services	single	high.scho	no	no	no	telephone	may	fri	346	4	999	0	nonexiste	1.1	93.994	-36.4	4.855
4	25	services	married	high.scho	no	yes	no	telephone	jun	wed	227	1	999	0	nonexiste	1.4	94.465	-41.8	4.962
5	38	services	married	basic.9y	no	unknown	unknown	telephone	jun	fri	17	3	999	0	nonexiste	1.4	94.465	-41.8	4.959
6	47	admin.	married	university	no	yes	no	cellular	nov	mon	58	1	999	0	nonexiste	-0.1	93.2	-42	4.191
7	32	services	single	university	no	no	no	cellular	sep	thu	128	3	999	2	failure	-1.1	94.199	-37.5	0.884
8	32	admin.	single	university	no	yes	no	cellular	sep	mon	290	4	999	0	nonexiste	-1.1	94.199	-37.5	0.879
9	41	entrepren	married	university	unknown	yes	no	cellular	nov	mon	44	2	999	0	nonexiste	-0.1	93.2	-42	4.191
0	31	services	divorced	profession	no	no	no	cellular	nov	tue	68	1	999	1	failure	-0.1	93.2	-42	4.153
1	35	blue-colla	married	basic.9y	unknown	no	no	telephone	may	thu	170	1	999	0	nonexiste	1.1	93.994	-36.4	4.855
2	25	services	single	basic.6y	unknown	yes	no	cellular	jul	thu	301	1	999	0	nonexiste	1.4	93.918	-42.7	4.958
3	36	self-empl	single	basic.4y	no	no	no	cellular	jul	thu	148	1	999	0	nonexiste	1.4	93.918	-42.7	4.968
4	36	admin.	married	high.scho	no	no	no	telephone	may	wed	97	2	999	0	nonexiste	1.1	93.994	-36.4	4.859
5	47	blue-colla	married	basic.4y	no	yes	no	telephone	jun	thu	211	2	999	0	nonexiste	1.4	94.465	-41.8	4.958
6	29	admin.	single	high.scho	no	no	no	cellular	may	fri	553	2	999	0	nonexiste	-1.8	92.893	-46.2	1.313
7	27	services	single	university	no	no	no	cellular	jul	wed	698	2	999	0	nonexiste	1.4	93.918	-42.7	4.963
8	44	admin.	divorced	university	no	no	no	cellular	jul	wed	191	6	999	0	nonexiste	1.4	93.918	-42.7	4.957
9	46	admin.	divorced	university	no	yes	no	telephone	jul	mon	59	4	999	0	nonexiste	1.4	93.918	-42.7	4.962
0	45	entrepren	married	university	unknown	yes	yes	cellular	aug	mon	38	2	999	0	nonexiste	1.4	93.444	-36.1	4.965
1	50	blue-colla	married	basic.4y	no	no	yes	cellular	jul	tue	849	1	999	0	nonexiste	1.4	93.918	-42.7	4.961
2	55	services	married	basic.6y	unknown	yes	no	cellular	jul	tue	326	6	999	0	nonexiste	1.4	93.918	-42.7	4.962
3	39	techniciar	divorced	high.scho	no	no	no	cellular	mar	mon	222	1	12	2	success	-1.8	93.369	-34.8	0.639
4	29	techniciar	single	university	no	yes	yes	cellular	aug	wed	626	3	999	0	nonexiste	1.4	93.444	-36.1	4.967
5	40	managem	married	high.scho	no	no	yes	cellular	aug	wed	119	1	999	0	nonexiste	1.4	93.444	-36.1	4.965

Datasource 2

education	age	job	marital	default	housing	loan
high.school	39.0	services	single	no	no	no
high.school	25.0	services	married	no	yes	no
high.school	36.0	admin.	married	no	no	no
high.school	29.0	admin.	single	no	no	no
high.school	39.0	technician	divorced	no	no	no

Output

6.1.2 Functional Overview

With this system, users can do the following major functions:
I.   Store data source.
II. Download stored data source.
III. View data source

6.1.3 Impacts

Impacts of the system are:
I.   User should be able to access the system over the network.
II. Participants can be a normal user, researcher or data scientist.
III. Participants must have the valid email for the authentication purpose.

6.1.3 Requirements

A	B
1 DFMS Requirement	
2 The user needs to provide the correct user ID and password	
3 The user needs to provide the correct username	
4 The application should have the capability to capture the news in the Information technology world	
5 The application should have the capability to capture the different datasources such as CSV and Txt etc	
6 The appliation should have the capability to check the different data sources which have been already been uploaded	
7 The application should have the capability to let the end user provide the feeback or contact the management	
8 The application should have the interface to select the datasource type which is being uploaded	

6.2   Functional Area/Design Unit B

Impacts

  I.   The AWS service needs at least two data sources for the functionality to work.

 II.   The user should be able to gain access to AWS console by submitting valid credentials.

7 Open Issues:

   i)   Third party – Relating to issues with Amazon Web services or MySQL. The team encountered connectivity issues with MySQL and AWS as they are different project which the application is interacting.

  ii)   Technical – This segment deals with the technical aspects of the project. It is important for the team to decide the compatible cloud service.

 iii)   Business process – Relating to the project's design. Here the focus is towards how the design helps in achieving a better business model

 iv)   Change management – This segment deals with the end user change requests. The team needs to discuss if the requested change is valid and then implement it.

8 Acknowledgements:

As a team we would like to thank our supervisor for providing feedbacks and necessary information to help the team complete the project.

All the team members of group 4 express special gratitude and thanks to the computer science department and industry persons for giving such attention and time.

9 References

- https://software.intel.com/sites/default/files/Mechanisms%20to%20Protect%20Data%20in%20the%20Open%20Cloud.pdf

- http://download.cnet.com/WampServer/3000-10248_4-10797035.html

- Learning PHP, MySQL & JavaScript: With jQuery, CSS & HTML5 (Learning Php, Mysql, Javascript, Css & Html5) 4th Edition

- https://d0.awsstatic.com/whitepapers/aws_cloud_adoption_framework.pdf

- https://d0.awsstatic.com/whitepapers/itil-event-management-in-the-cloud.pdf

- http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/monitoring_ec2.html

- ITIL Service Operation Publication, Office of Government Commerce, 2007.

- http://www.allaboutagile.com/10-good-reasons-to-do-agile-development/

- http://www.allaboutagile.com/10-good-reasons-to-do-agile-development/

Description: Using the above screen where user has to enter login credentials to get inside the system. If the user has provided as correct user credentials then he will be access the DFMS dash board to upload the data sources and view files, RSS feed to browser the latest information technology news.
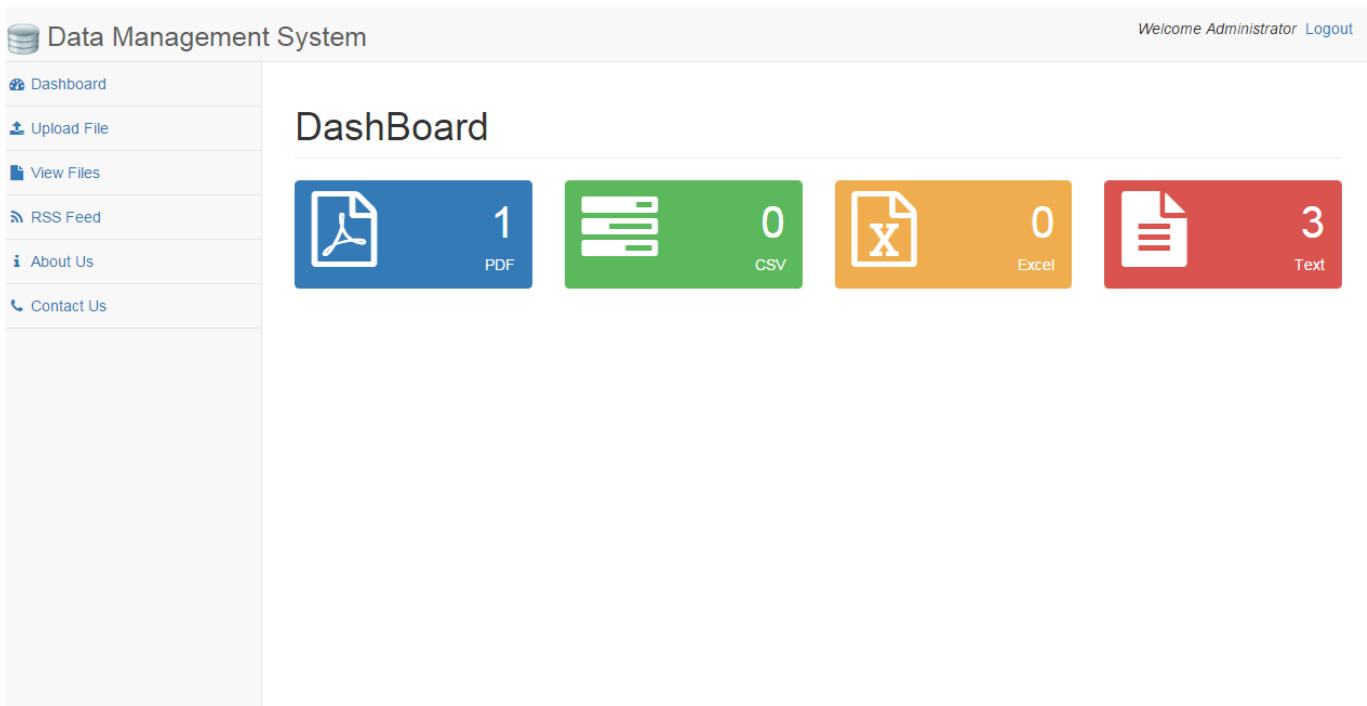
Description: The above screen shot is the dashboard of the DFMS application. It offers the interface using which the user can upload different files and then can also view the Feed news. In addition to that the user can view the files which have been already uploaded them. The user can also provide the feedback using the contact us tab available.

Description: The DFMS application uses WampServer for hosting purpose. WampServer is a Windows web development environment
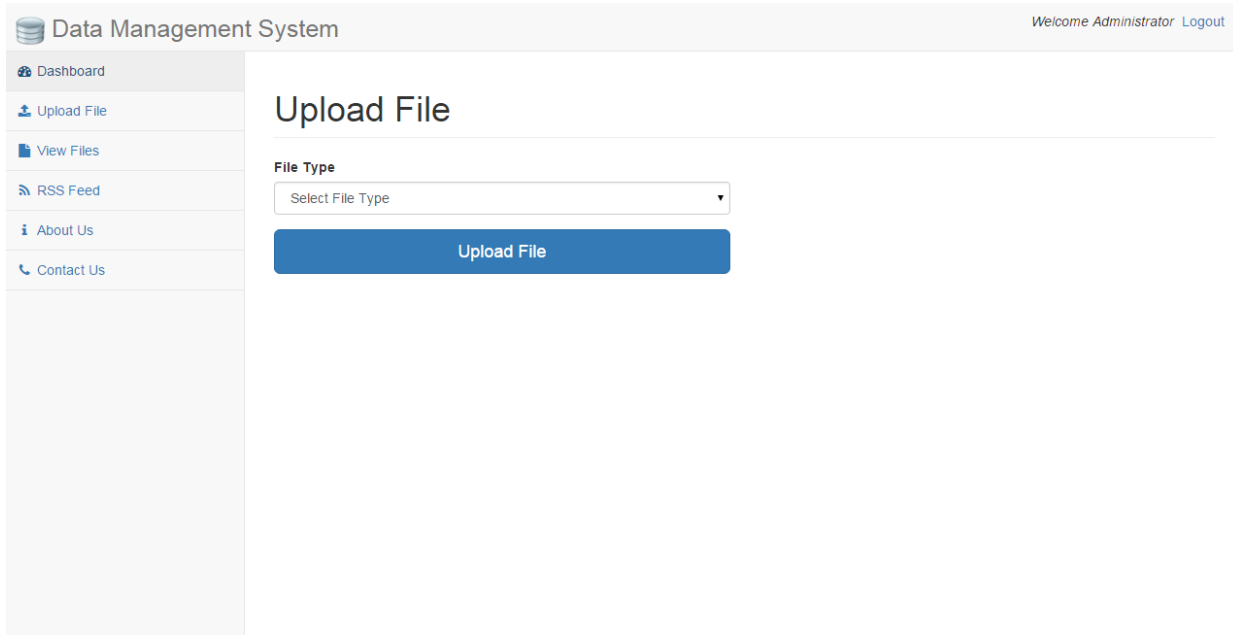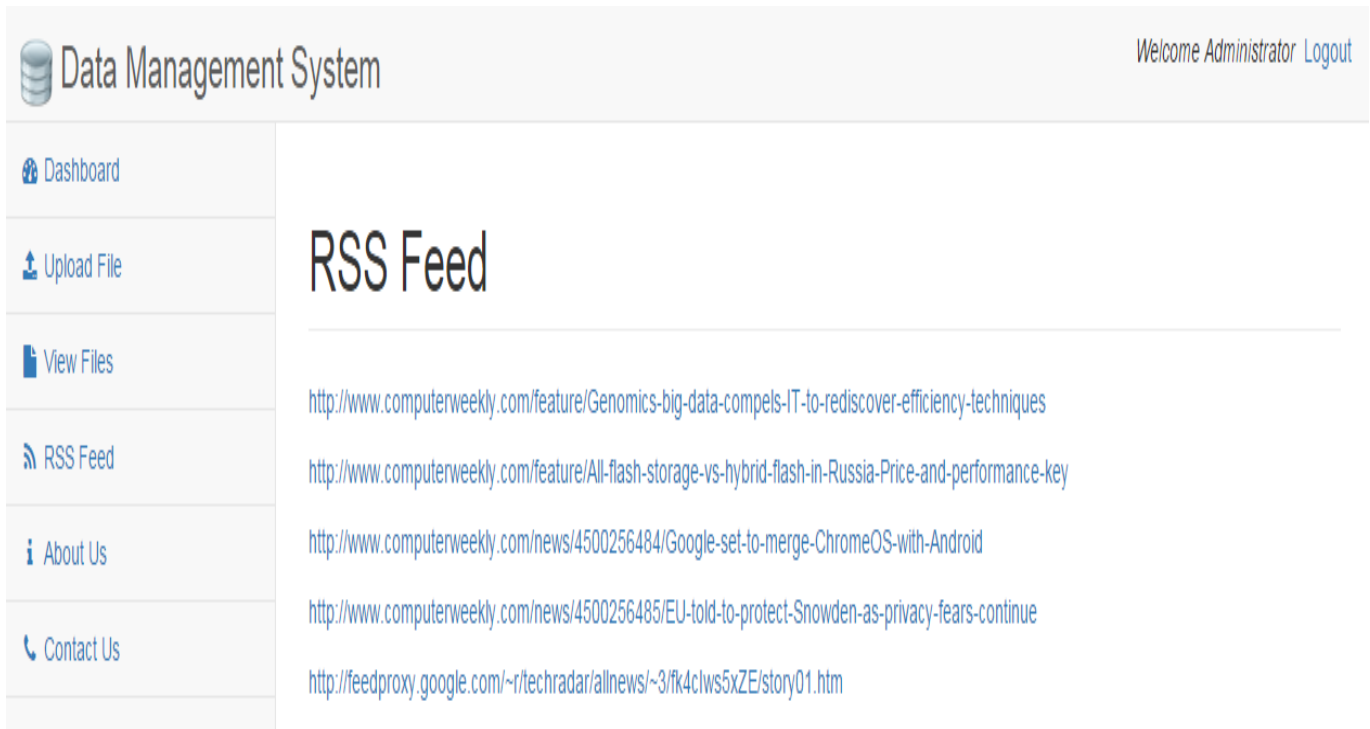
Description: This interface lets the user to select the data source type such as CSV, Txt etc which the end user selects before uploading the file into the database which in our case local host.

Page Description: Using the above webpage the user can view the previously uploaded data sources.

Description: Using the above displayed webpage the end user will be able directly visit other website WebPages.