INTERNSHIP PROJECT REPORT

(Project Term August-December 2018)

R&D PERFORMANCE TEAM

Submitted by

Shivansh Sharma **Registration Number: 11508903**

Course Code: CSE 447 - INDUSTRY CO - OP PROJECT - I

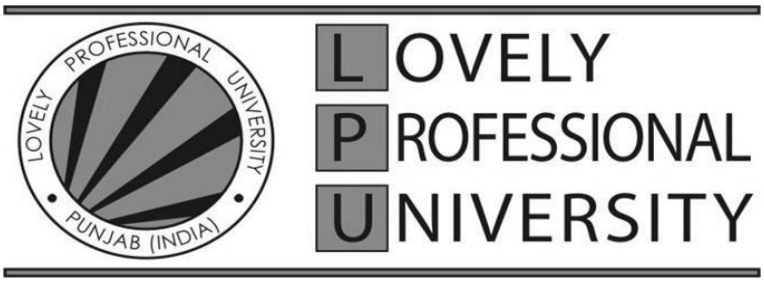
Under the Guidance of

Sami Anand | Associate Professor, LPU

and

Adarsh Praharaj | Sr. Manager IT, Informatica

School of Computer Science and Engineering



CONTINUOUS ASSESSMENT (CA) for INTERNSHIP

(By external internship in-charge from organization)

Name of the Student: Shivansh Sharma Registration Number 11508903

Internship Project Title (if/any): ONE CLICK

Name of Organization & Address: Informatica LLC, Bagmane Tech Park, Bangalore

Name of External Internship in-charge (with mobile number): Adarsh Praharaj

S.No.	Criteria	Marks Obtained	Maximum Marks
1	Student conduct during internship	10	10
2	Punctuality and Enthusiasm	18	20
3	Technical Skill & Knowledge	20	20
4	Internship Project Marks	50	50
	TOTAL	98	100

Date 13/Dec/18 Authorized Signatory Adarsh Praharaj

Name of External Internship in-charge: Adarsh Praharaj

Designation: **Sr. Manager (QA Engineering – R&D)**

Company Seal

INFORMATICA BUSINESS SOLUTIONS PVT. LTD. BANGALORE

Note: The external internship in-charge will mark the continuous assessment only at the time of completion of the internship. Students must ensure that evaluation marks are provided by the organization as per above parameters in the given format during ETP.

DECLARATION

I hereby declare that the project work entitled "**ONE CLICK**" is an authentic record of our own work carried out as requirements of Internship for the award of B.Tech degree in Computer Science and Engineering from Lovely Professional University, Phagwara, under the guidance of Associate Professor Sami Anand and Mr. Adarsh Praharaj (Senior Manager Performance QA team), during (August to December 2018). All the information furnished in this internship project report is based on my own intensive work and is genuine.

Name of Student: Shivansh Sharma

Registration Number: 11508903

SHIVANSH SHARMA

(Signature of Student)

Date: 2 Dec 2018

CERTIFICATE

This is to certify that the declaration statement made by the student is correct to the best of my knowledge and belief. He has progressed well with his internship under my guidance and supervision. The present work is the result of his original investigation, effort and study. No part of the work has ever been submitted for any other degree at any University. The internship is fit for the submission and partial fulfillment of the conditions for the award of B.Tech degree in Computer Science and Engineering(Hons.) from Lovely Professional University, Phagwara.

Signature:

Sami Anand

Associate Professor

School of Computer Science and Engineering,
Lovely Professional University, Phagwara, Punjab.

Date: 2 Dec 2018

ACKNOWLEDGEMENT

I take this opportunity to express our gratitude and respect to all those who have helped me throughout our working period on the real time company environment. Doing internship in a product-based company help us a lot to understand the new technology and how to grow in a corporate world. My special thanks is to our mentor Mr. Adarsh Praharaj (Senior Manager Performance QA team)., who helped me a lot to show us the right path how to work in a company and to Learn the various aspects of application Security

I owe my regards to the entire faculty of the department of Computer Science at LPU from where I learnt the basics of Computer Science and I express my sincere thanks to all our course mates who supported us in the project through various informal discussions which were very valuable to the successful completion of the project.

Shivansh Sharma

TABLE OF CONTENT

CHAPTER 1: INTRODUCTION

1.1 About Organization

Informatica is a data driven company which means making data and analytics part of the business strategy, its systems, processes and culture. It was found in 1993 and has its headquarter in Redwood City, California. The Founder of Informatica is Gaurav Dhillon and Diaz Nesamoney and the current CEO is Anil Chakravarthy.

The main products at core level are Enterprise Cloud Data Management and Data Integration. It offers products for ETL, data masking, data Quality, data replica, data virtualization, master data management, etc. Informatica PowerCenter ETL/Data Integration tool is most widely used tool. The popular clients using Informatica products are U.S Air Force, Allianz, Fannie Mae, ING, Samsung, etc. The popular tools available in the market in competition to Informatica are IBM DataStage, Oracle OWB, Microsoft SSIS and Ab Initio.

1.2 Introduction of Product ONE CLICK

One Click helps the Informatica trainers to easily launch the Informatica domain and its services on amazon machine in 30 minutes, reducing the effort to install the latest build in their machines.

The product is used by R&D team to perform tests on the recent build, without taking a headache of installing the new build and started installing the product, on new servers wasting half of their day.

Chapter 2: Problem Statement

2.1 Problem Statement

In the past, customers or users run applications or programs from software downloaded on a physical computer or server in their building, which, in the case for informatica takes a lot of time and money. It takes 3-4 days just to setup one single product with servers up and running. DEW (Delivery Enablement Workshop) teams were suffered a lot because of this issue. We, as a performance team need to check the performance benchmark of all informatica products and doing this with old solution was a great headache. There is a need to do something which helps everybody and do the work in short time and also saves cost.

2.2 Project Scope

Being a part of Research and Development Performance team, my job was to work on different tools made to help us in monitoring the performance of different products. I was assigned the job of developer for Amazon Web Services, my task was to simplify the use of these web services provided by amazon, for Informatica.

My work as a developer was to build product that can help us in working with amazon services and at the same time cut the cost we were paying for using the services.

One Click provides a framework to run One Click Script to launch Informatica Domain and BDM products.

One Click provides the users with a simple user interface to bring up Informatica products on Amazon EC2 machines in their respective accounts with a single click. The application gives the user the capability to

- Start, stop and terminate EC2 instances.
- Adding or Removing permissions for images/snapshots, copying images/snapshots to a different region.
- Monitoring your AWS account.

Chapter 3: Existing System

3.1 Introduction

The existing system, as an example like EDC (Enterprise Data catalog). EDC is capable of scanning and extracting metadata, perform profiling operations, add data domains, business terms and custom attributes which makes the search and analysis for data easy. At present there is no such strong competitor to EDC in market. Most of the enterprises who need data Integration, analysis and management services, use EDC.

3.2 Existing Software

The basic architecture of EDC is established on following services:

- **MRS**: Model Repository Service which contains the models that are needed to define which model to refer for accepting data from user for resource creation. Information about data domains are also stored in it
- **DIS:** Data Integration Service which performs the integration tasks for Enterprise Information Catalog and external applications like finding similarities.
- **CMS:** Content management Service which manages reference data. It provides reference data information to the Data Integration Service and Informatica Developer. You can use Informatica Developer to import data domains into Model repository.
- **CS:** Catalog Service which that runs Enterprise Unified Metadata and manages connections between service components and external applications. Here in Catalog Service user can make resources in admin part and see their result in

- catalog part. This service is connected to a cluster which stores the metadata fetched.

The EDC basic resource creation works in a no. of stages which are as follows:

3.2.1 Creation of resource

While creating resource we need to provide some of the information which will be used, and a connection will be created with resource. Different kind of resource need different type of information and corresponding option will be provided to fill after choosing resource type. One could verify the details through test connection option provided.

3.2.2 Metadata Extraction

Here the metadata like in Oracle the different datasets, data-elements, their names, source datatype with precision and length gets fetched. This is data about the data present at source.

3.2.3 Staging

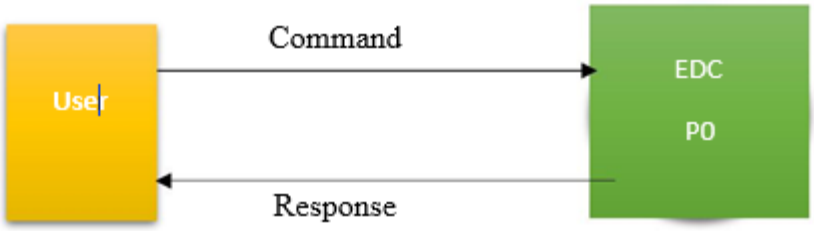
After metadata extraction staging comes into picture where the XDocs of the fetched data are created and get stored in HBase. For each object there is a XDocs which will be created.

3.2.4 Ingestion

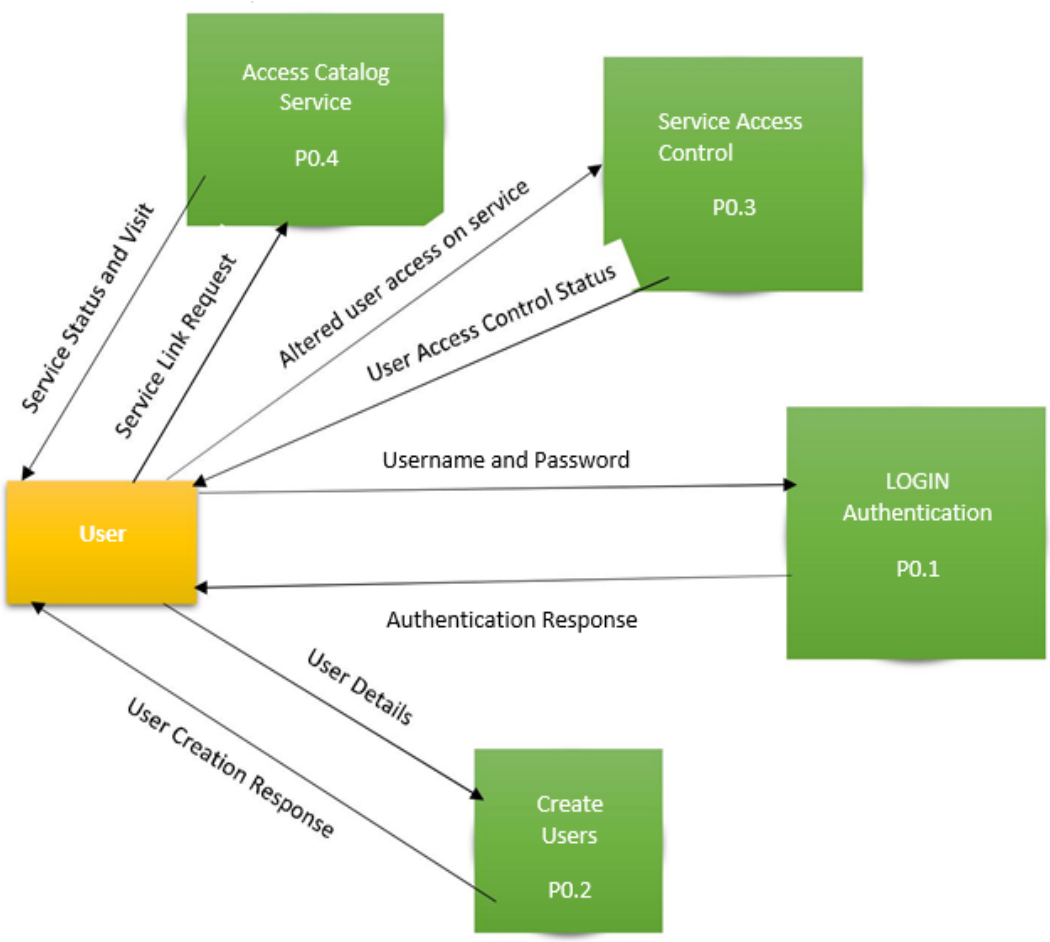
During ingestion the XDocs the are created will be stored in HBase and indexing of it will be made in solar which will enhance the search.

3.3 DFD for present System

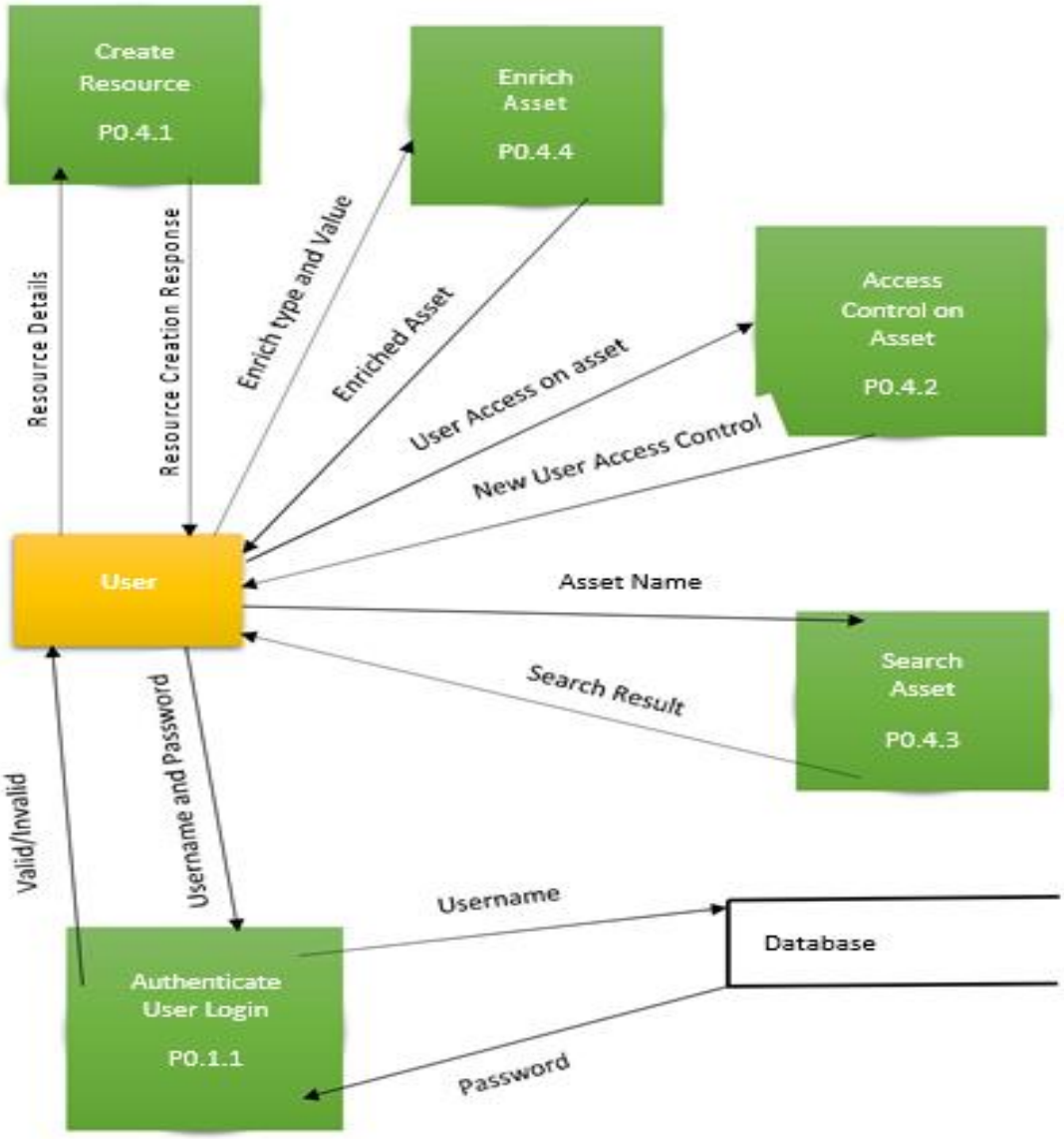
Level 0



Level 1



Level 2


Create Resource
P0.4.1
Enrich Asset
P0.4.4
Access Control on Asset
P0.4.2
Search Asset
P0.4.3
User
Authenticate User Login
P0.1.1
Database
Resource Details
Resource Creation Response
Enrich type and Value
Enriched Asset
User Access on asset
New User Access Control
Asset Name
Search Result
Valid/Invalid
Username and Password
Username
Password

3.4 Upcoming Features

The upcoming features include certification to a dataset or data element where Data Owner or Data Steward or Subject Matter Expert cloud certify the resource for some specialty of the asset. Review and Rating where different users who has access to the asset can provide the rating and reviews to an asset as per their usage. Question and Answer where user can ask any no of question about an asset and any user can answer that question. For good reviews and helpful answers user can even vote for those answers and review. Next is notification i.e. if a resource gets enriched or some collaboration changes are made then if user have followed the asset for desired type, it will be notified to him.

Chapter 4: Problem Analysis

4.1 Product Definition

One Click helps the Informatica trainers to easily launch the Informatica domain and its services on amazon machine in 30 minutes, reducing the effort to install the latest build in their machines.

The product is used by R&D team to perform tests on the recent build, without taking a headache of installing the new build and started installing the product, on new servers wasting half of their day.

4.2 Feasibility Analysis

Feasibility study required deep research on AWS services, how the data on AWS can be accessed and used for our purpose. We had to analyses if it is possible to move further on the idea.

Feedback from employees whether there is requirement for the project if it will be a help to them.

1. Once in place this technology is simple to operate and maintain for a relatively low-cost Marketing

2. This project will allow people to reach large number of customers, and at the same time reducing the pain to install products, thus saving time.

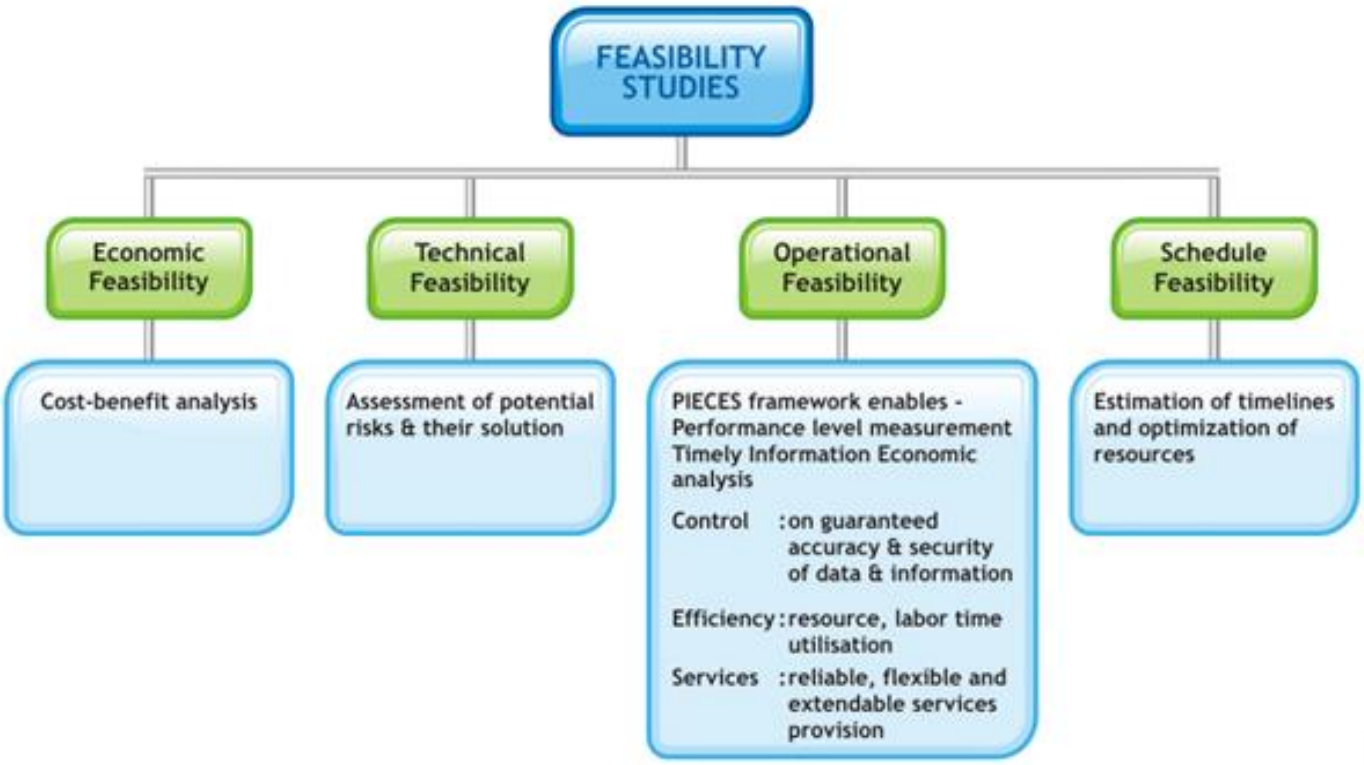


Fig 3.1

3. Using AWS services like EC2, RDS and EBS volumes save money as the servers are on demand. When there is need, we use the resources, thus no need to buy servers.

4. User Interface to create config file from the data in user's AWS account and then launching the required instances with one click, will reduce the effort to write a config file AMI id from the amazon and then run the One Click scripts.

5. It will be easy to get the data from AWS, by doing SDK calls using predefined functions in AWS, so the implementation is quite feasible technically.

Chapter 5: Software Requirement Analysis

5.1 Introduction

The requirement analysis based on the feedback from teams were:

1. A single platform to launch AWS instances, perform operation like start, stop and terminate instances from one web-app.

2. For launching instance using One Click script the user had to write ids from AWS console, which is time consuming, because each time a new launch is done, we make a new configuration. Finding a way in which the user can see all security groups, key pairs and images on one single page, from which they can select the best options.

3. The main concern when using AWS services is the cost we must pay when the user launches instances. They don't know what the cost of the launch will be, so the task is to find out a way in which the user can see the approximate cost of the instances launched from UI.

5.2 General Description

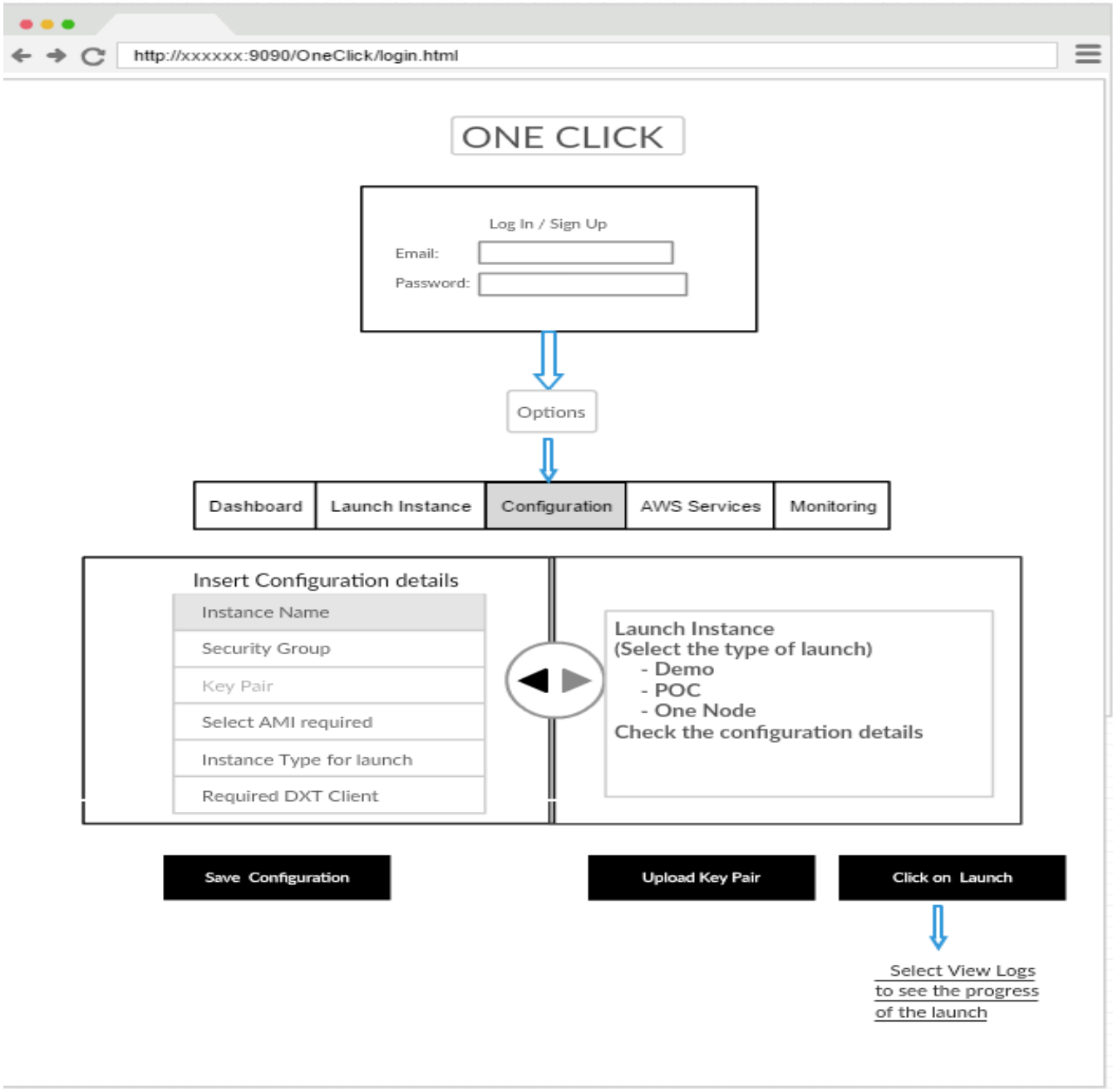
In general, this is performance enhancement application which is used by Informatica to optimize and stabilize their business. It can work on different platform as no native installation is required, an internet browser is all we need.

5.3 Specific Requirements

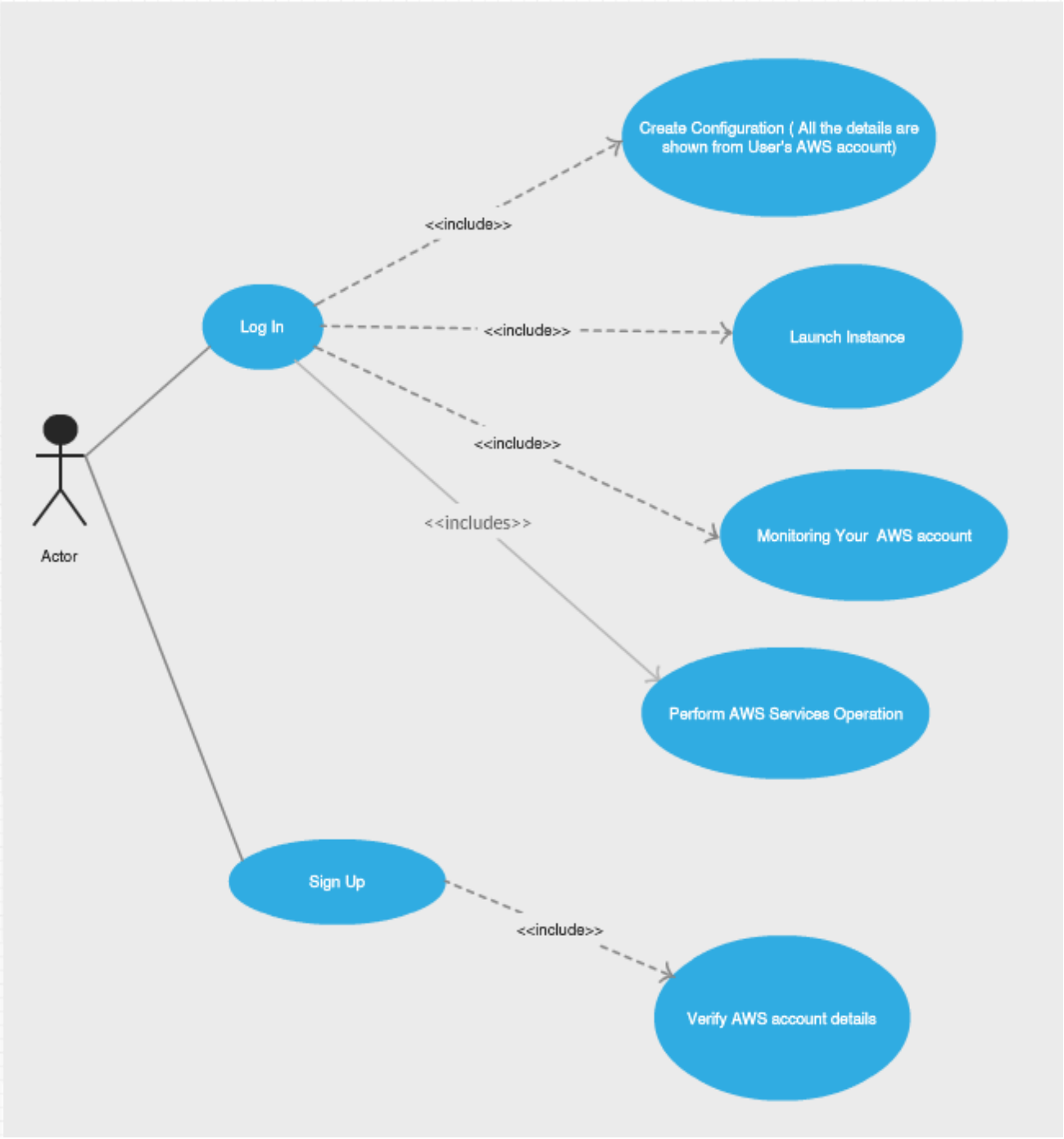
It is necessary that user must keep his AWS credential details handy all the time for login purpose. ONE CLICK is all operated on AWS services, so it requires AWS access key Id and AWS access Key, only then user will be able to monitor or launch instances from this application.

Chapter 6: Design

6.1 System Design



6.2 Detailed Design



6.3 Pseudo Code

The languages used in this project are

6.3.1 User Interface (HTML, CSS, JavaScript, jQuery)

The main component of jQuery used is jQxgrid .

Creating adapters to get data in the form of json from the servlet request.

```
function getSummaryDataAdaptor(url)
{
        var source =
        {
        datatype : "json",
        datafields : [
        {
        name : 'INSTANCENAME',
        type : 'string',
        },
        {
        name : 'INSTANCEID',
        type : 'string',
        },
        {
        name : 'INSTANCETYPE',
        type : 'string',
        },
        {
        name : 'COSTHOUR',
```

```
            type : 'number',

            }],
            root:'data',
            url : url,
            pagesize : 20
            };

            var dataAdapter = new $.jqx.dataAdapter(source);
            return dataAdapter;
    }
```

Where URL is the servlet request to get the json. Json that the server is sending will have key as the name field specified in the adapter.

This data adapter is used to populate the grid.

The grid has a field of data adapter. The grid should have the same datafields to get the data from servlet request.

```
width : getContentWidth(),
autoHeight : true,
sortable : true,
autoshowcolumnsmenubutton: false,
filtermode : 'excel',
source : getSummaryDataAdaptor(summaryURL),
showtoolbar : true,
renderstatusbar : function(statusbar)
{



},
selectionmode : 'checkbox',
columnsresize : true,
pageable : true,
pagesizeoptions : [ '5', '10', '15', '20' ],

columns : [
{
text : 'Instance Name',
datafield : 'INSTANCENAME',
width : '12%'

},
{
text : 'Instance Type',
datafield : 'INSTANCETYPE',
width : '10%'
},
{
text : 'Rank',
datafield : 'STATUS',
width : '5%'
```

6.3.2 Server code – JAVA, RESTful API, AWS SDK for java

Developing an API in java

Example: To verify AWS credentials there will be a GET request and its return type is plain text.

```
@Path("/checkAwsCredentials")
		@GET
@Produces(MediaType.TEXT_PLAIN)

	public String check(@Context HttpServletRequest sessionToken){


		if (!SecurityRestEndPoint.isSessionValid(sessionToken)) {
			return Utilities.genericMessage("You are not allowed to perform this operation. Session has been invalidated.");
		}

UserProfileDataModel userProfile = SecurityRestEndPoint.getUserProfile(sessionToken);


	if(Utilities.isNullOrEmpty(userProfile.getAccessKey()) || Utilities.isNullOrEmpty(userProfile.getAccessKeyId())){
				return "false";
			}

		return "true";

		}
```

6.3.3 Using AWS SDK java

To use AWS SDK, we must make a connection using built in functions provide by the SDK. We should make a Basic Authentication to further connect to AWS Services.

```
String accessKey = userProfile.getAccessKey();
        String accessKeyId = userProfile.getAccessKeyId();
        BasicAWSCredentials awsCreds = new BasicAWSCredentials(accessKeyId, accessKey);
```

We can now connect to different services using Example :

```
AmazonEC2Client ec2Client = new AmazonEC2Client(awsCreds);
```

6.3.4 Connect Client and Server

To call this API from the client there is a simple get request from JavaScript

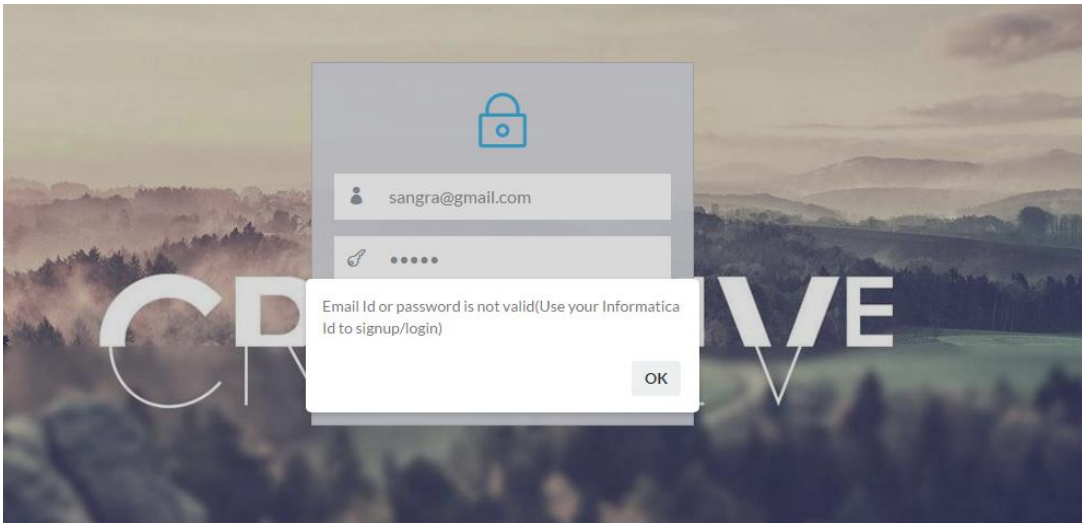
Example

```
$.get("servlets/aws/checkIfValidCredentials?accesskey="+access_id+"&accesskeyid="+access,function(data){
        if(data == "true"){

//do something

}
```

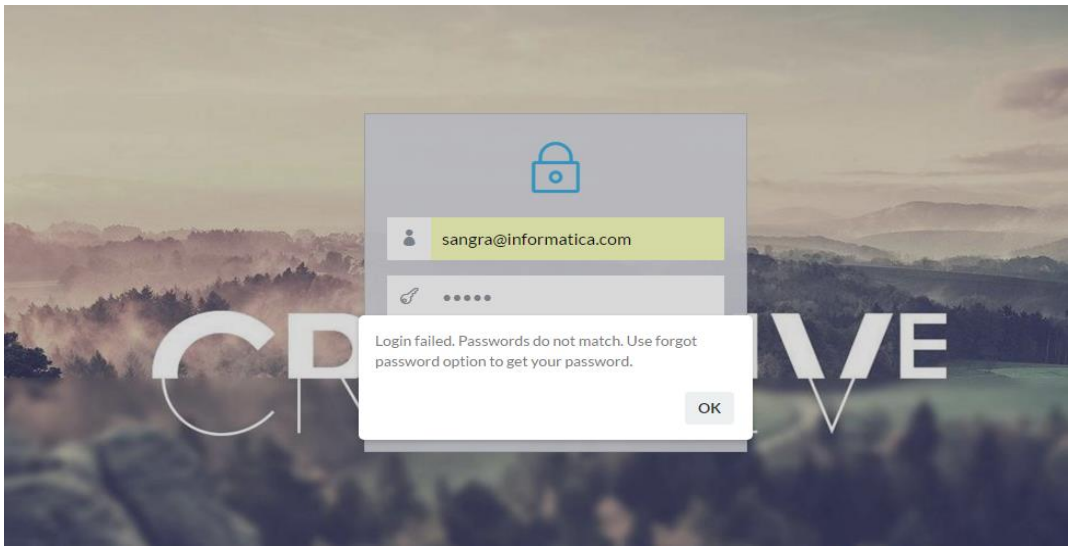
Chapter 7: Testing

Login and Sign Up Testing

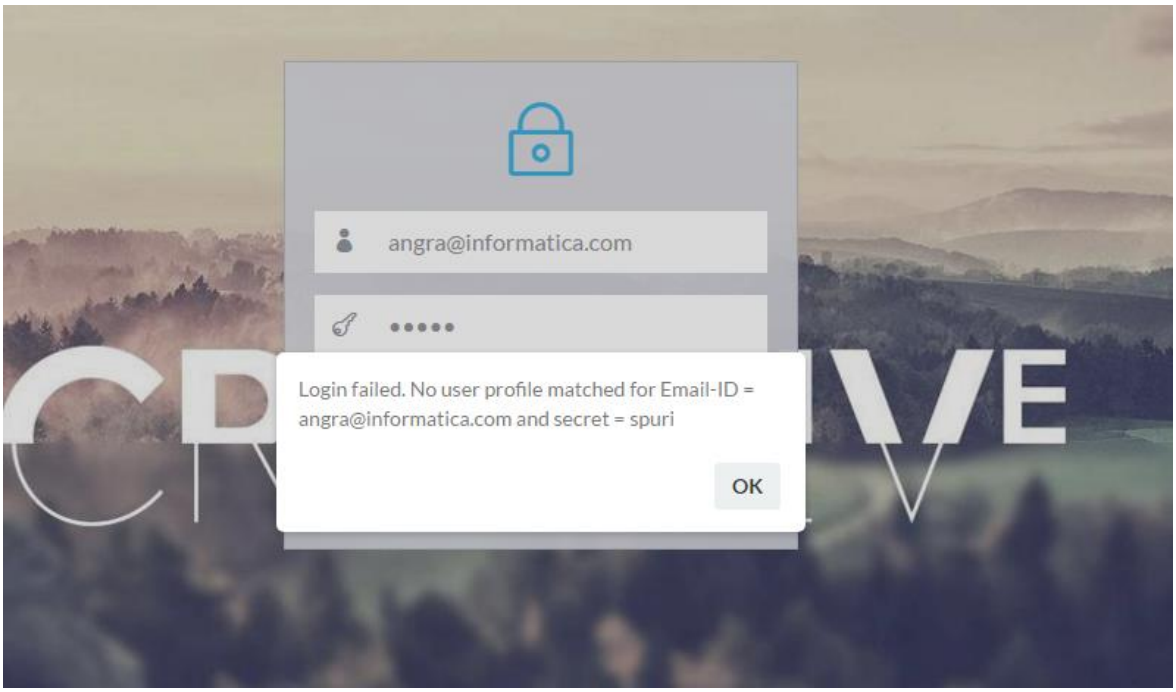
1. Check if the email id is a valid Informatica id.



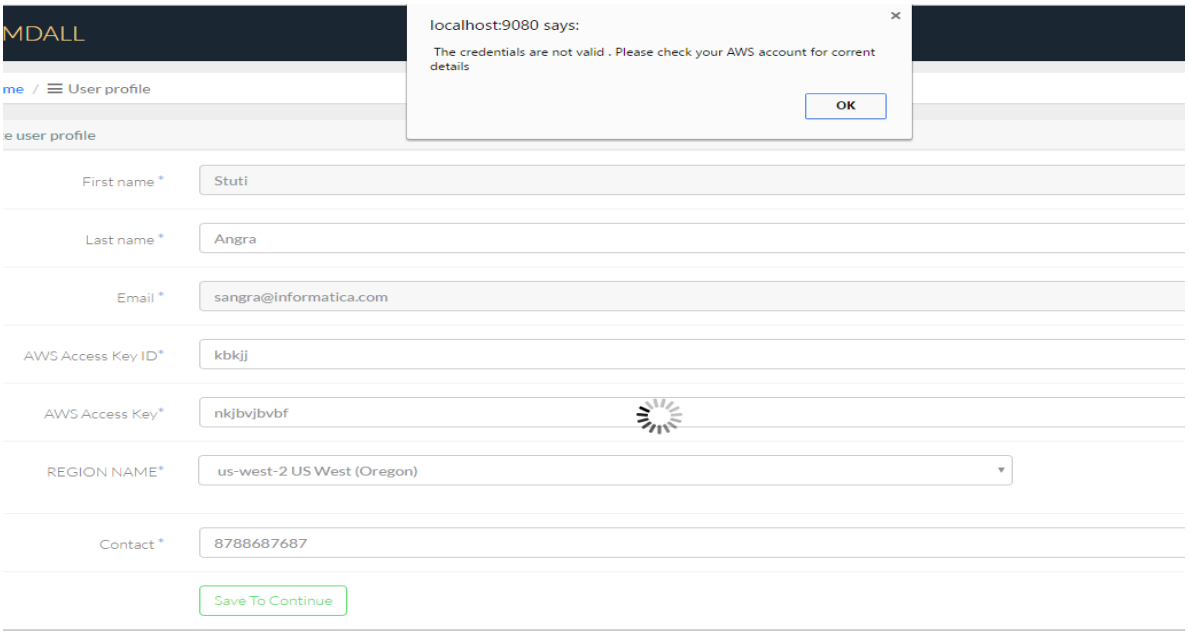
2. To authenticate the sign in, check if password is checked when logging and showing appropriate message.



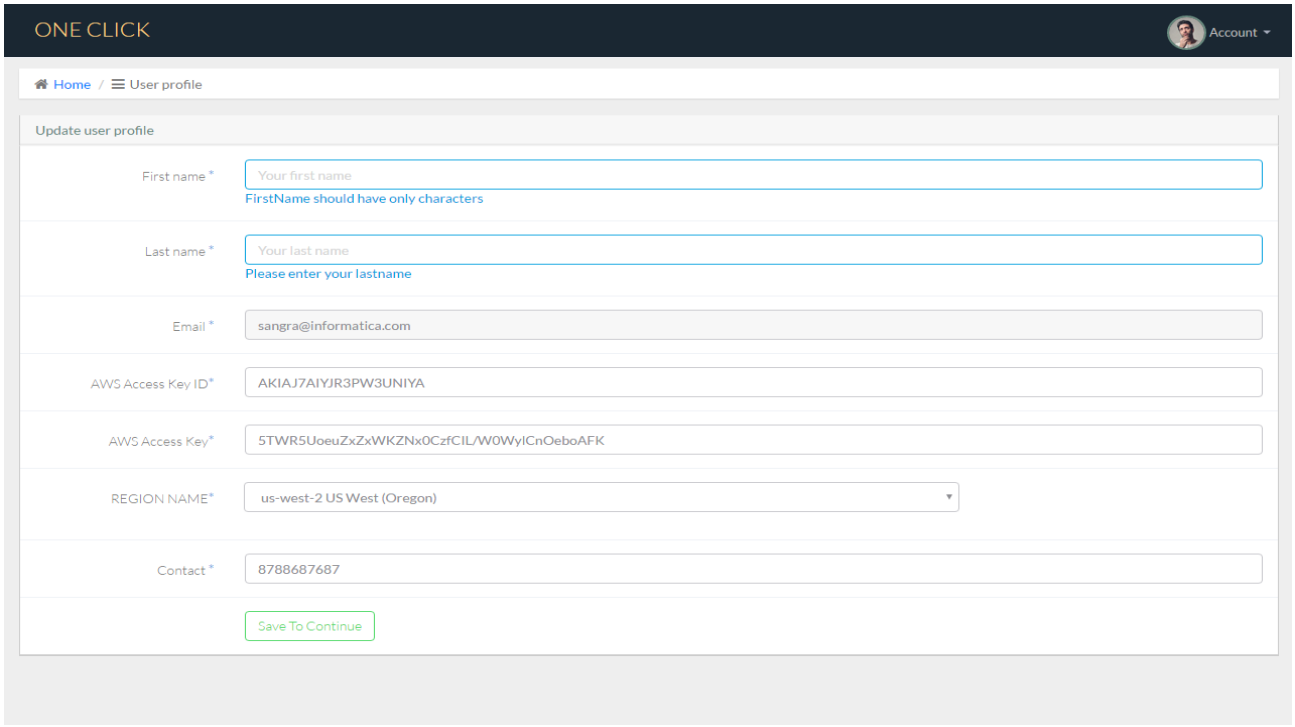
3. If the user is not present and a new user tries to login. Don't allow the user to log In.



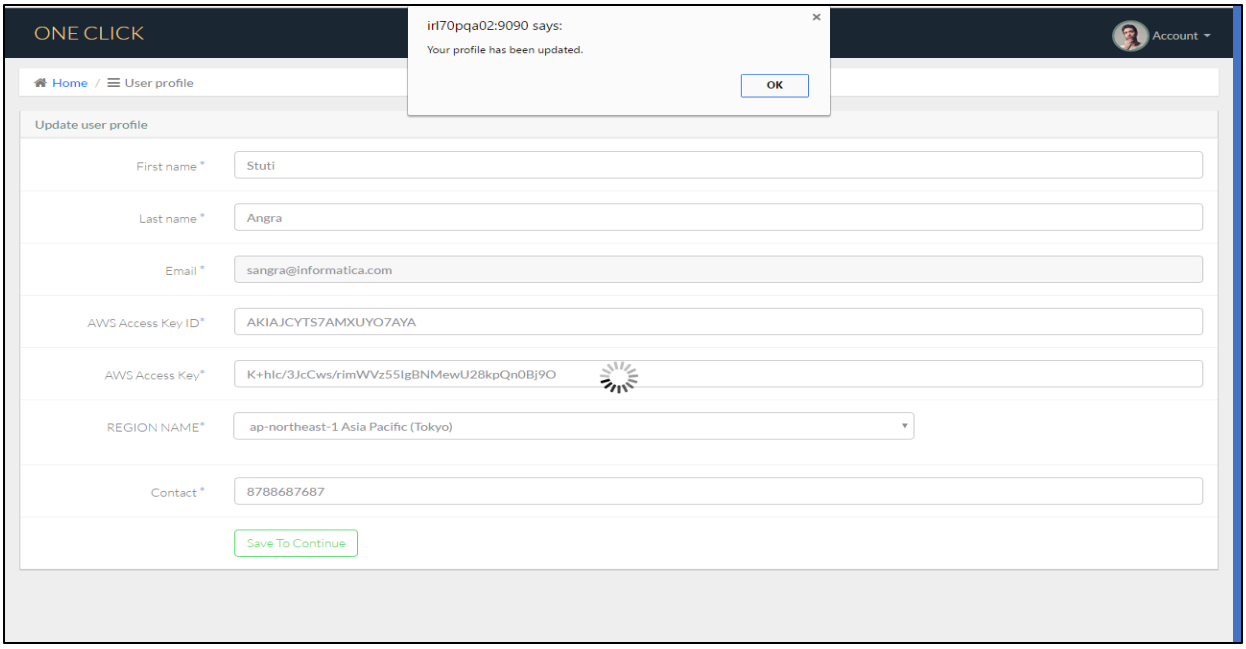
4. The whole application depends on AWS access key and Id when the user tries to save data AWS authentication is done to check if it is a valid id or not.



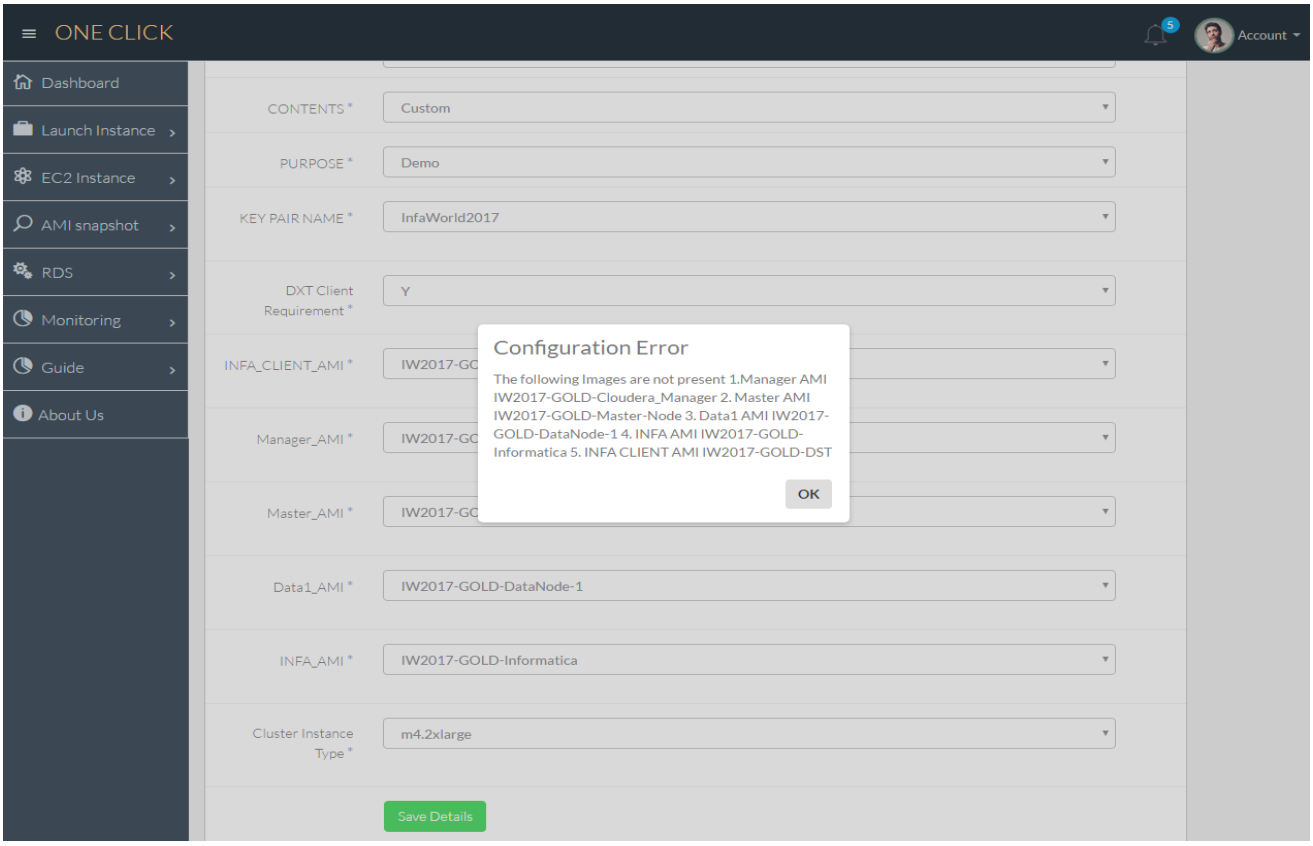
5. Checks on required not null values in the user profile data model



6. Only allowing the user with correct AWS credentials to update data



7. Only saving the configuration if the images are present in the user's account.



Chapter 8: Implementation

8.1 Implementation of Project

This product is accessible through any browser, it will be accessible to users in the organization. This is not a public product. It is used by enterprises and need to possess the account in AWS.

8.2 Post Implementation and Maintenance

Cloud Computing provides us a means by which we can access the applications as utilities, over the Internet. It allows us to create, configure, and customize applications online. Cloud Computing refers to manipulating, configuring, and accessing the applications online. It offers online data storage, infrastructure and application.

1.Our future goal is to incorporate all the Informatica products inside the setup. Teams use a common platform to launch their instances, install their products using one Click.

2.To enhance the monitoring, to show more details about the AWS accounts, their resource usage and the cost to company.

3.A Launch tab where the user can launch instances other than Demo, POC and One Node which will then be used to monitor resources easily (We must run a script on all instances to provide this feature).

4.The same platform will now be used to launch instances on Microsoft Azure.

Chapter 9: Project Legacy

9.1 Project Current Status

Currently the project is completed for AWS users. It's been on testing continuously, we have created a help desk which can help users to mail directly to us whenever there is any issue.

9.2 Technical and Managerial Lessons Learn

- I have learnt how to work effectively under a team in an organization.
- Manage your work in such a way that it will complete within ETA.
- Analyze the time requirement for an assignment.
- Make the use of your team members and synergies the output of whole team.
- Working on different technology like AWS, hibernate, SDKs etc.
- I have learnt about EDC product of Informatica. Its working, usage, purpose, etc.
- Testing manual and automation for REST APIs for the product EDC.

Chapter 10: Bibliography

- www.javatpoint.com
- www.wikipedia.com
- http://docs.aws.amazon.com/sdk-for-java/v1/developer-guide/welcome.html
- https://stackoverflow.com