Design and Development of Business Intelligence Portal in Cloud Computing

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Software Engineering

Submitted By
Sonam Chawla
(800931020)

Under the supervision of:
Dr. Inderveer Chana
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

June 2011

Certificate

I hereby certify that the work which is being presented in the thesis entitled, "*Design and Development of Business Intelligence Portal in Cloud Computing*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Inderveer Chana* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

(Sonam Chawla)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

17/06/11

(Dr. Inderveer Chana)
Assistant Professor,
Computer Science and Engineering Department

Countersigned by

(Dr. Maninder Singh)
Head 19/6/11
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. K. Mohapatra)
Dean (Academic Affairs)
Thapar University
Patiala

Acknowledgement

I would like to acknowledge everyone who supported me during my experience at Thapar University and my work on this thesis.

First of all I would like to thank my family who stayed with me and supported me all the time while I made this dream come true.

I extend my thanks to my thesis supervisor, Dr. Inderveer Chana, who helped me to materialize my ideas on this thesis. Her enthusiasm and optimism made this experience both rewarding and enjoyable. Most of the novel ideas and solutions found in this thesis are the result of our numerous stimulating discussions. Her feedback and editorial comments were also valuable for writing this thesis.

I am also very thankful to the faculty and staff members of Computer Science and Engineering Department, especially Mrs. Anju Bala and Ms. Rajni for their direct – indirect help and cooperation.

I would also like to thank my friends especially Ipneet Kaur, Aradhana Majithia, Vaneet Kaur, Arpita Sharma, Aarti Sharma, Ravneet Kaur Grewal, Amandeep Kaur Johar and Ravneet Kaur Chawla for their support.

Sonam Chawla

(800931020)

Abstract

Cloud Computing is a buzzword for web infrastructure. There is a shift in business from desktops towards Cloud Computing. Cloud Computing is not a completely new concept; it has intricate connection to the established Grid Computing paradigm, and other relevant technologies such as utility computing, cluster computing and distributed systems in general. Business Intelligence (BI) refers to computer-based techniques used in spotting, digging-out, and analyzing business data, such as sales revenue by products and/or departments or associated costs and incomes. Business Intelligence technologies provide historical, current, and predictive views of business operations. Common functions of Business Intelligence technologies are reporting, online analytical processing (OLAP), analytics, data mining, business performance management, benchmarking, text mining, and predictive analytics. Business Intelligence is, nowadays, given a high priority in top enterprise applications. Business Intelligence technologies provide historical, current, and predictive views of business operations.

Integrating BI with Cloud Computing will bring revolutionary changes in business analytics. Implementing BI in Cloud Computing allows users to access data from anywhere in the world. To access business operations in a Cloud in a simple way Business Intelligence applications are required. Some of the cloud providers have implemented BI in Cloud Computing, but still there are lot of challenges that need to be handled while implementing BI Cloud.

In this thesis available BI Clouds have been compared, benefits of BI cloud and challenges while implementing a BI Cloud have been discussed. A BI cloud portal has also been designed, developed and presented in this thesis. This portal has been deployed on cloud environment provided by Google App Engine.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

This chapter gives brief introduction of Cloud Computing, Business Intelligence (BI), BI Cloud and overview of thesis.

1.1 Background

The general idea behind the technology dates back to the 1960s, when John McCarthy wrote that "computation may someday be organized as a public utility." Then, grid computing, a concept that originated in the early 1990s as an idea for making computer power as easy to access as an electric power grid also contributed to cloud computing [1]. Figure 1.1 shows the evolution of Cloud from Clustering.

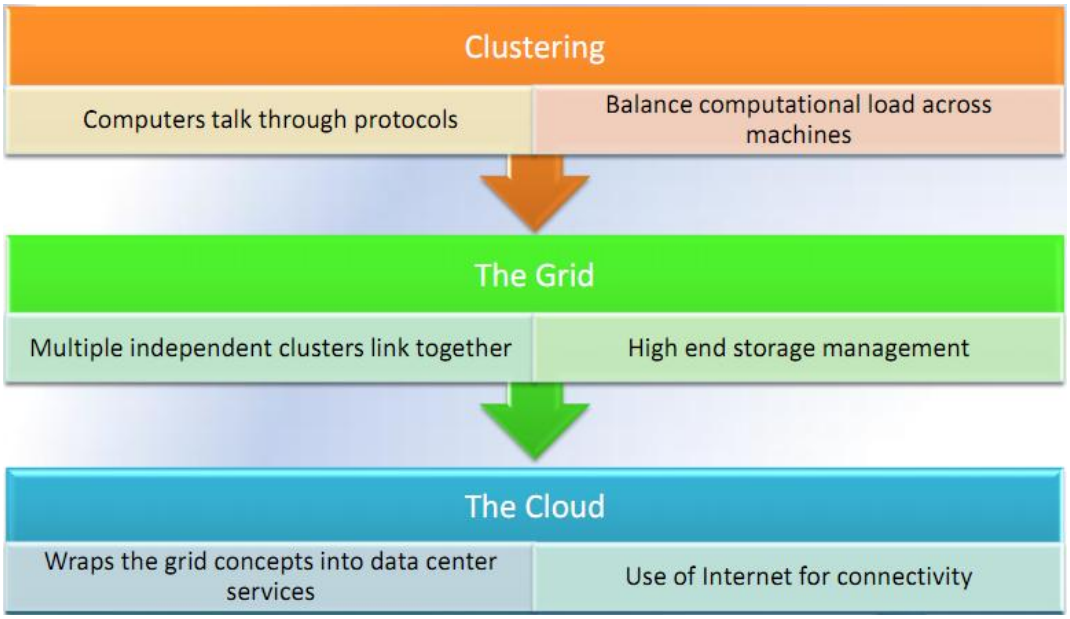


Figure 1.1 Evolution of Cloud Computing from Clustering [2]

Cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers and other devices on-demand, like the electricity grid. Cloud computing customers do not own the physical infrastructure, instead avoiding capital expenditure by renting usage from a third-party provider. They consume resources as a service and pay only for resources that they use [4].

Cloud computing has recently emerged as a new paradigm for hosting and delivering services over the Internet. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows

enterprises to start from the small and increase resources only when there is a rise in service demand [5].

Typical cloud computing providers deliver common business applications online that are accessed from another Web service or software like a Web browser, while the software and data are stored on servers. Most cloud computing infrastructures consist of services delivered through common centers and built-on servers. Clouds often appear as single points of access for consumers' computing needs. Commercial offerings are generally expected to meet quality of service (QoS) requirements of customers, and typically include service level agreements (SLAs) [32]. The key characteristic of cloud computing is that the computing is "in the cloud"; that is, the processing (and the related data) is not in a specified, known or static place(s) [33]. This is in contrast to a model in which the processing takes place in one or more specific servers that are known [6].

1.1.1 Classification of Clouds

Clouds can be classified according to who the owner of the Cloud data centres is. Clouds can be classified as –

- Private Clouds
- Public Clouds
- Hybrid Clouds
- Federation of Clouds

Private Clouds

Private Clouds are internal Clouds. A Private Cloud is fully owned by a single company who has total control over the applications run on the infrastructure, the place where they run, and the people or organizations using it. A Private Cloud relies on virtualization of an organization’s existing infrastructure. The key advantage of a Private Cloud is to gain all advantages of virtualization, while retaining full control over the infrastructure [7].

Public Clouds

A Public Cloud is a data centre, hardware and software run by third parties, e.g. Google, which expose their services to companies and consumers via the Internet. A Public Cloud is not restricted to a limited user base. It is made available in a pay-as-you-go manner to the general public [7].

Hybrid Clouds

Hybrid Clouds combine Public and Private Clouds and allow an organization to both run some applications on an internal Cloud infrastructure and others in a Public Cloud. Thus, companies can benefit from scalable IT resources offered by external Cloud providers while keeping specific applications or data inside the firewall. A mixed Cloud environment adds complexity regarding the distribution of applications across different environments, monitoring of the internal and external infrastructure involved, security and privacy, and may therefore not be suited for applications requiring complex databases or synchronization [7].

Federations of Clouds

The terms Federated Clouds or Federation of Clouds denote collaboration among Public Clouds even though Private Clouds may be involved. Federated Clouds are a collection of single Clouds that can interoperate, i.e. exchange data and computing resources through defined interfaces. According to basic federation principles, in a Federation of Clouds each single Cloud remains independent, but can interoperate with other Clouds in the federation through standardized interfaces. There is no common Cloud interoperability standard [7].

1.1.2 Different Levels of Cloud Computing

Cloud computing has three levels of service offerings: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a service (IaaS). These levels support virtualization and management of differing levels of the solution stack.

Software as a Service (SaaS)

A SaaS provider hosts and manages a given application in their own data center and makes it available to multiple users over the Web. Some SaaS providers run on another cloud provider's PaaS or IaaS service offerings [8].

Platform as a Service (PaaS)

Platform as a Service is an application development and deployment platform delivered as a service to developers over the Web. It facilitates development and deployment of applications without the cost and complexity of buying and managing the underlying infrastructure, providing all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely available from the Internet. This platform consists of infrastructure software, and typically includes a database, middleware and development tools [8].

Infrastructure as a Service (IaaS)

Infrastructure as a Service is the delivery of hardware (server, storage and network), and associated software (operating systems virtualization technology, file system), as a service. It is an evolution of traditional hosting that does not require any long term commitment and allows users to provision resources on demand. Unlike PaaS services, the IaaS provider does very little management other than keep the data center operational and users must deploy and manage the software services themselves--just the way they would in their own data center. Amazon Web Services Elastic Compute Cloud (EC2) and Secure Storage Service (S3) are examples of IaaS offerings [8].

1.1.3 Cloud Computing Advantages

Following advantages of Cloud Computing have been identified by Dr. Wendy A. Warr [9]:

- Economies of scale – Reduces cost of setting up physical data center
- Software as a Subscription (SaaS) - SaaS deployments usually take less time than in-house ones, upgrades are easier, and users are always using the most recent version of the application.
- Pay-As-You-Go - Little or no capital investment and maintenance cost is needed.
- Portability/Accessibility - Files and softwares are available anywhere provided that there is an active internet connection.
- Efficient Use of Computer Resources - Users no longer require separate servers for different applications. With virtualization multiple server technologies can run from a single server.
- Elasticity – Cloud has the ability to add capacity or applications almost at a moment's notice.

1.2 Introduction to Business Intelligence

In a 1958 article, IBM researcher Hans Peter Luhn used the term business intelligence. He defined intelligence as: "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal." Business intelligence has evolved from the decision support systems which began in the 1960s and developed throughout the mid-80s. In 1989 Howard Dresner (later a Gartner Group analyst) proposed "business intelligence" as an umbrella term

to describe concepts and methods to improve business decision making by using fact-based support systems [3].

Business Intelligence (BI) refers to computer-based techniques used in spotting, digging-out, and analyzing business data, such as sales revenue by products and/or departments or associated costs and incomes [37]. Business Intelligence technologies provide historical, current, and predictive views of business operations. Common functions of Business Intelligence technologies are reporting, online analytical processing (OLAP), analytics, data mining, business performance management, benchmarking, text mining, and predictive analytics [10].

Business Intelligence was coined by Gartner in the early 1990s, the term BI denotes on the one hand an analytic process that transforms internal and external data into information about capabilities, market positions, activities, and goals that the company should pursue in order to stay competitive. On the other hand, BI stands for Information System concepts like Online Analytical processing (OLAP), Querying and Reporting, or Data Mining that provides different methods for a flexible goal-driven analysis of business data, provided through a central data pool. BI may facilitate the connections in the new-form organization, bringing real-time information to centralized repositories and support analytics that can be exploited at every horizontal and vertical level within and outside the firm [11] [29].

The promise of BI is in specific and timely knowledge about customers, products and markets. This knowledge can help boost profits, reduce costs and support better, more effective management. Based on comprehensive, detailed and relevant information, this knowledge is crucial to achieving and sustaining a competitive advantage. BI requires information on demand—achieved by combining multidimensional data analysis and data mining with advanced statistical and analytical functions in a real-time, integrated environment [12] [30].

1.2.1 BI Architecture

Figure 1.2 shows Business Intelligence Architecture. It shows that data is retrieved from different databases (like Oracle, MySQL, DB2, Sybase, MS Access, etc.). These different types of resources give data in different format which then is integrated in one format using any DBMS tool (like SQL Server, Oracle, etc.). This data is stored in data warehouse which is then extracted for analysis. After analysing the data, reports are generated in the format (like pie chart, table, graphical representation, etc.)

required by the clients. The reports are then presented to the clients on presentation layer by using softwares (like SharePoint, MS Office, etc.).

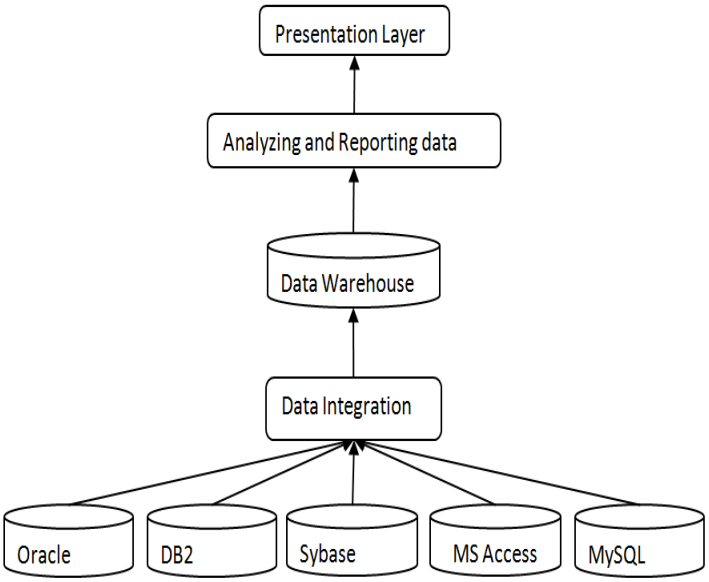


Figure 1.2 Business Intelligence Architecture

1.2.2 BI Advantages

Following advantages of Business Intelligence have been identified [2]:

- Accurate sensing of customers' needs and requirements
- Ability to adapt to any market changes or shifts, including economic recessions
- Accurate business and sales forecasting
- Help business managers and owners develop competitive business strategies based on accurate and timely information
- Recognition of customer needs
- Optimisation of operations and resources
- Increases Cost-effectiveness
- Quality analysis as the basis for future projections.

1.2.3 Challenges : Implementing BI

Organizations implementing a BI solution may face several challenges [3]:

- Integrating complex data from heterogeneous hardware platforms and software environments
- Managing distributed systems that have no single point of control and time-sensitive operations
- Improving data access while reducing expenses

- Performing frequent updates across already overtaxed networks with rapidly increasing traffic
- Backing up, recovering and archiving data within diminishing availability windows
- Incorporating efficiencies of new technologies without requiring massive downtime, costs or retraining
- Providing scalable servers to run multi-terabyte applications
- Providing storage that protects data and is scalable, open and manageable

1.3 BI Cloud

Implementing Business Intelligence in Cloud computing is supposed to change the economics of Business Intelligence by making available the hardware, networking, security and software needed to create data marts and data warehouses on demand with a pay-as-you-go approach to usage and licensing [13]. Business Intelligence operations can be simplified by integrating Business Intelligence with Cloud Computing.

According to a recent Gartner survey, the most important drivers for investing in business intelligence are related to providing faster response to user's needs for data, decreasing costs, and enhancing the user's methods for data sharing and self-service. Cloud computing makes sense to business intelligence solutions only if it offers benefits to customers [14].

Although there are many benefits of implementing business intelligence in cloud, there are many challenges which need to be addressed while implementing BI Cloud like security, performance, availability, integration, etc [31].

1.4 Organisation of Thesis

The chapters in this thesis are organized as follows –

Chapter 2 – This chapter describes in detail the literature survey, how to implement Business Intelligence in Cloud Computing and comparison of available BI Clouds.

Chapter 3 – This chapter describes the problem statement of the thesis. It gives the gap analysis, requirement analysis and need for BI Cloud Portal.

Chapter 4 – This chapter describes the solution of problem, design of solution, Cloud Environment Alternatives available and implementation of solution.

Chapter 5 – This chapter gives the Experimental results – installation of Python 2.7, Installation of Google App Engine, Snapshots of BI Cloud Portal developed and benefits of BI Cloud Portal.

Chapter 6 – This chapter describes the conclusion, contributions of work done and future research work possible.

Chapter 2

Literature Review

This chapter describes in detail the literature survey, how to implement Business Intelligence (BI) in Cloud Computing and comparison of available BI Clouds.

2.1 Implementing Business Intelligence in Cloud Computing

According to Brian Gentile, CEO of Jaspersoft, after outsourcing the entire hardware and software cycle to a third-party provider (such as Amazon's EC2, Google's App Engine, and solutions from GoGrid and AppNexus) companies can focus on applications and solutions. They can also scale their business up and down quickly, without the delays and costs typically associated with hardware and software acquisition [15].

Implementation of Business Intelligence in Cloud Computing enable users to access information from anywhere in the world for making sales decisions and to predict and act on business opportunities [34]. A sales executive, for example, can query real-time data to identify sales opportunities and understand how many sales in their region are ready to close. On the development side, a manufacturing process engineer can analyze real-time data from a plant floor to improve yield and reduce shipment delivery times [16].

2.1.1 Benefits of Implementing Business Intelligence in Cloud Computing

According to a Gartner survey, the most important drivers for investing in BI are related to providing faster response to user's needs for data, decreasing cost, and enhancing the user's methods for data sharing and self-service [14].

The main benefits of implementing business intelligence in cloud computing include lower cost, multiple redundant sites, and scalable provisioning of resources, which in turn allow for business continuity, disaster recovery and on-demand performance improvements [17].

Lower cost

Implementation of BI in cloud would lower the cost incurred by companies to purchase and maintain hardware as well as software. Companies do not need to invest large amounts of money for hardware, software, licences and knowledge to put BI infrastructure up and running. They would only have to pay for the resources they

need or used. Thus, unnecessary cost paid for the time in which no user accesses the application and computing resources would cut down [28].

Multiple redundant sites

One of the major concerns of business intelligence professionals is to keep the solution available the longest time possible. One way to achieve this is to have multiple sites that offer redundancy. Since cloud computing providers have sites geographically dispersed, this characteristic is achieved [28].

Scalable provisioning of resources

The work load of BI solutions during a day does not remain constant. Therefore, at certain points in time, some servers could be idle while others may be reaching their peaks on processing, memory usage or I/O operations. With cloud computing resources can automatically scale up and scale down according to the requirements [28].

On-demand performance improvement

One of the most common problems that are seen on BI applications is when the customers need to expand their data warehousing capabilities without affecting daily operations. This task is usually very complex because it requires high investments in new hardware, storage, licences, and human effort to perform the migration to the new environment. By implementing BI in Cloud, this problem would be addressed almost instantaneously and transparently for users, by taking advantage of existing hardware and software resources [28].

Usage billing

By using cloud computing, companies pay for a service as they use or pay on a monthly or yearly basis. With this policy, the expenses move from Capital Expenditure (CapEx) to Operational Expenditure (OpEx) [28].

Fast deployment

Instead of spending long time preparing and installing required hardware and software, the platforms can be up and running in just minutes, ready to configure applications and start populating the data warehouse [28].

Easy maintenance

Most of the maintenance needed for hardware and software, like firmware, updates and upgrades, are done by the cloud computing provider. Also, since these applications are accessed through internet browsers, maintenance on client computers reduces dramatically [28].

2.1.2 Challenges of Implementing BI in Cloud

Although the benefits that cloud computing brings to the industry have been mentioned above but there are several challenges that have been identified by several authors [35]. The main challenges faced by BI solutions in the cloud are discussed in this section.

Cyrus Golkar of B-eye-Network identifies these CIO concerns [18] –

- Security
- Performance
- Availability
- Integration
- Customization

Mukund Deshpande and S. Joshi of B-eye-Network identify these issues [19] -

- Moving data to the cloud
- Storing data in the cloud
- BI components as a service
- Integration with on-premise data

W. Eckerson of Data Warehouse Institute identifies these constraints [20] –

- Data transfers
- Data security
- Due diligence

S. Dine of Datasource Consulting, LLC identifies these challenges [21] -

- The ability to scale-up is limited
- Difficult to quell security concern
- Viability of moving large amounts of data
- Performance of physical data access
- Reliability of service concerns
- Pricing is variable and complex

Dave Wells of B-eye-Network explains these factors as the downside of cloud computing [22] –

- Security
- Privacy
- Compliance
- Control

- Governance

Tom Lounibos of Eclipse Developer's Journal mentioned the next challenges according to SOASTA, a company based on the cloud that offers products for web application testing [23] –

- Security
- Governance
- Performance

Thus, it can be seen that there are several challenges while implementing BI in Cloud. Thus main challenges can be concluded from the recurring challenges from above sources. Main challenges are security, moving large amount of data, performance, integration and reliability.

2.2 Cloud Providers and Business Intelligence: State-of-the-art

There are many Cloud service providers (such as Amazon, Google, Informatica, etc.). Table 2.1 shows the list of the cloud providers and that provide the Business Intelligence Cloud. Amazon, GoGrid, Google, Sun Microsystems, and Rackspace do not provide BI Cloud. Informatics provides only data integration of BI in Cloud using Amazon EC2 cloud. IBM has private BI Cloud called Blue Insight. RIGHTSCALE also provides BI Cloud using the technologies of JasperSoft, Vertica and talend.

Table 2.1 Cloud Providers providing BI Clouds

Cloud Provider	**BI Cloud**
Amazon	NO
GoGrid	NO
Google	NO
Informatica	YES (data integration via Amazon EC2)
IBM	YES (private cloud Blue Insight)
Salesforce.com	YES (sales and service cloud)
RightScale	YES
Sun Microsystem	NO
Rackspace	NO

Table 2.2 shows the BI Cloud providers and their features. It can be construed that currently no BI Cloud utilizes the full features of BI and Cloud Computing integration. Table shows the comparison of various BI Cloud available according to the features provided by them. Public/Private Cloud shows that whether the cloud provided by them is publicly available or only for their organisation. Open Source means that the source code of BI Cloud is available or not. Full BI means whether all functions of Business Intelligence are addressed or not. Common functions of business intelligence technologies are reporting, online analytical processing,

analytics, data mining, benchmarking, and predictive analytics. Data Management shows the data management technique used by the BI Cloud providers. Flexible shows the expandability of BI Cloud.

Table 2.2 BI Cloud Comparison

Features	**RightScale**	**IBM (Blue Insight)**	**Salesforce.com (Sales Cloud 2)**	**Salesforce. com (Service Cloud 2)**	**Informatica**
Public/ Private	Public	Private	Public	Public	Public
Open Source	No	No	No	No	No
Full BI	Yes	Yes (Business Analytics)	Yes (Sales Service)	No (only Customer Service)	No (only data integration)
Data Management	Specialized analytic databases	More than Petabyte data storage	Automated data management	No	Data migration, replication, & Archiving
Forecasting	No	Yes	Yes	No	No
Scalability	Yes	Yes	Less	Less	Less
Flexible	Yes	Yes	No	No	No

Above table concludes that RightScale BI Cloud is better among available Public BI Clouds whereas Private BI Cloud has been implemented by IBM only.

Chapter 3

Problem Statement

This chapter discusses about the gap analysis, requirement analysis and need for BI Cloud.

3.1 Gap Analysis

In today's state-of-the-art corporate and scientific environments enormous amounts of data are being generated. Collecting and analyzing this data is getting increasingly complex, because of the amount of data, and the complexity of the analysis. Business intelligence architecture should be flexible and able to adapt to changing business needs and objectives. A rigid architecture will not serve the needs of the business in a fast-changing environment. Thus, Business Intelligence is being implemented in Cloud. From the literature survey following gaps have been found:

- Portal of BI Cloud is not available.
- BI only covers one part of business process i.e. analysing data. Other processes like management processes, operational processes are not available on cloud.
- Service oriented Business Intelligence platform could be developed and implemented in cloud.
- Real time BI in cloud is not available.

3.2 Cloud Computing requirements for BI solutions

One of the objectives that originated business intelligence and data warehousing was to develop tasks that were not able to be executed on transactional systems. This forced a rethinking of different hardware and software architectures, like database optimized for reading instead of writing and high interactive and visual applications.

The basic architecture needed to run business intelligence solutions in the cloud is depicted in Figure 3.1.

The lower layers are formed by hardware and software systems. These are the minimum elements that have to be offered by the cloud computing provider. Hardware refers to processing, storage, and networks, while software refers to the operating systems and drivers required to handle the hardware.

The data integration refers to tools needed to perform the ETL and data cleansing processes. The database refers to the relational or multidimensional database systems

that administer the information. Data warehousing tools are the set of applications that allow the creation and maintenance of the data warehouse. BI tools are the set of front-end applications that enable the final users to access and analyze the data.

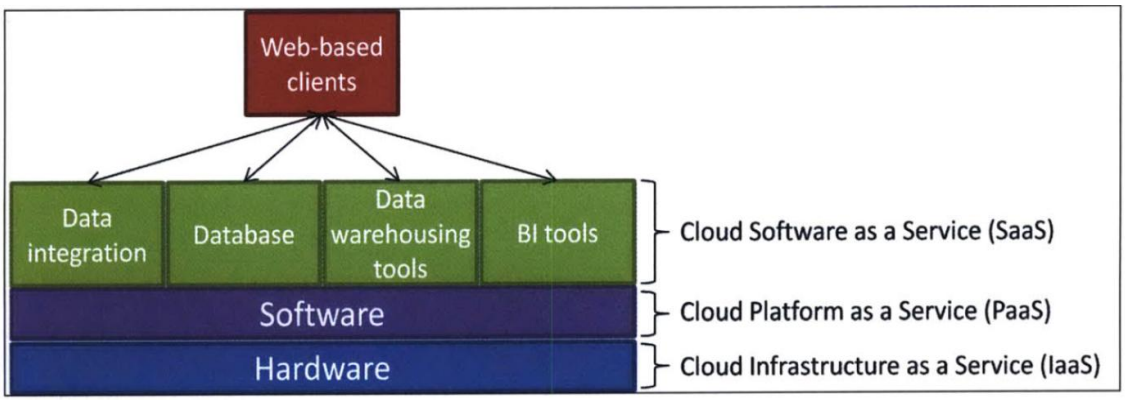


Figure 3.1 Basic Business Intelligence Architecture in Cloud Computing

3.3 Need for BI Cloud

Business Intelligence needs to be implemented on an environment which can store data securely. Data management is the main issue for Business Intelligence as there is a huge amount of data which is being stored and accessed for BI solutions. This data is increasing exponentially and thus storing and managing such a large amount of data is becoming a challenge. Storing such a large amount of data is also challenging as need for memory is increasing with increasing amount of data. Implementing business intelligence in cloud can resolve this issue as data is stored on third party servers and managed by them using various data management techniques at less cost.

As BI Cloud provides a cost efficient BI solution; decrease the load of storing and managing a huge amount of data; elasticity in resources required like memory, processors, etc.; increases scalability; can assessed anywhere from the world; and data is available at distributed sites, thus solution is available for the longest time possible. Thus, there is a need for implementing Business intelligence on cloud which will also result in increasing adoption of BI solutions worldwide.

BI Cloud Portal is a need in today's environment as the client might need a robust and agile system that allows him for advance manipulation of the website content with the ability to deploy new changes in short time.

Chapter 4
Solution of Problem

This chapter discusses about how the problem stated in previous chapter can be solved with the help of Data Flow Diagram and UML diagrams.

4.1 Design of Solution

The solution to the problem (implementation of business intelligence on cloud) has been designed through Data Flow Diagram and UML diagrams (use case, activity, sequence, deployment). Following section presents the design of the solution using Data Flow Diagram and UML diagrams.

4.1.1 Data Flow Diagram

Data Flow Diagrams (DFD) are the graphical representation of flow of data through an information system. A DFD provides no information about the timing of processes, or about whether processes will operate in sequence or in parallel. DFDs of BI Portal designed for this thesis are shown from figure 4.1 – figure 4.4. Figure 4.1 shows the context level (level 0) DFD of the portal with four entities – user (visitor of portal), new user (want to register), member (registered user) and administrator. Figure 4.2 shows the level 1 DFD for user, Figure 4.3 shows the level 1 DFD for new user, Figure 4.4 shows the level 1 DFD for member.

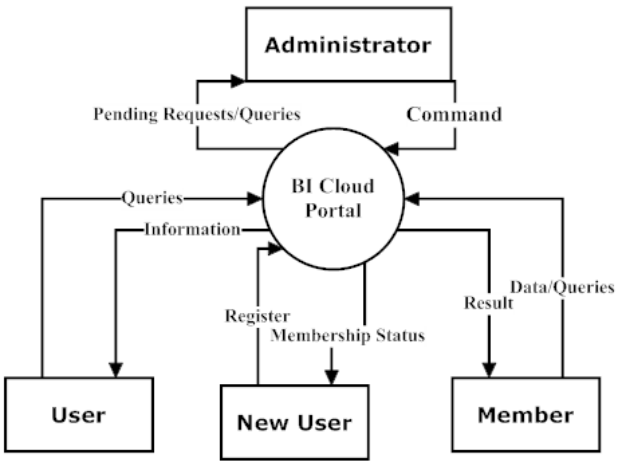


Figure 4.1 Context-level DFD

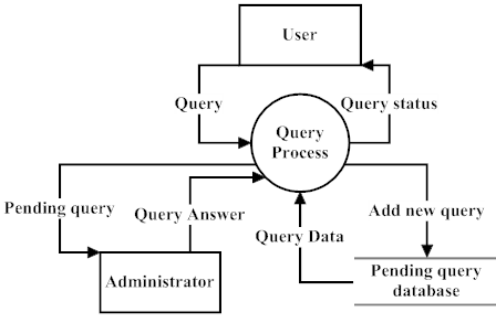


Figure 4.2 Level - 1 DFD for user

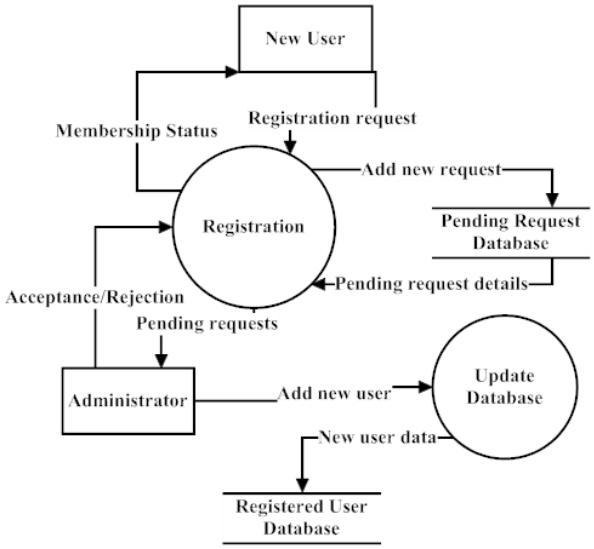


Figure 4.3 Level - 1 DFD for new user

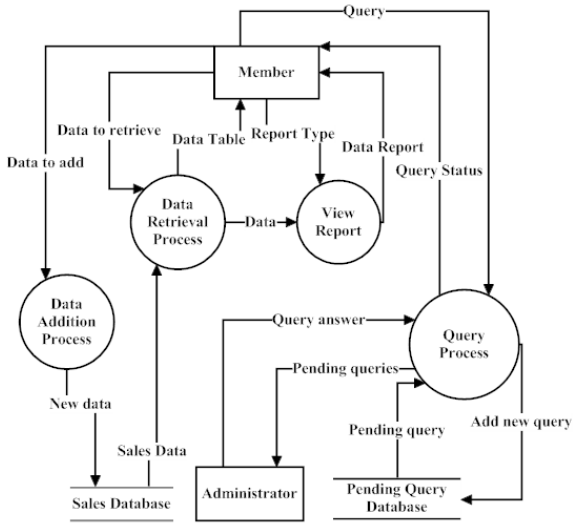


Figure 4.4 Level - 1 DFD for member

4.1.2 UML Diagram

Each UML diagram is designed to let developers and customers view a software system from a different perspective and in varying degrees of abstraction.

Use Case Diagrams

Use case diagrams present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. It is used to gather requirements of the system including internal and external influences. Figure 4.5 shows the use case diagram for BI Cloud Portal.

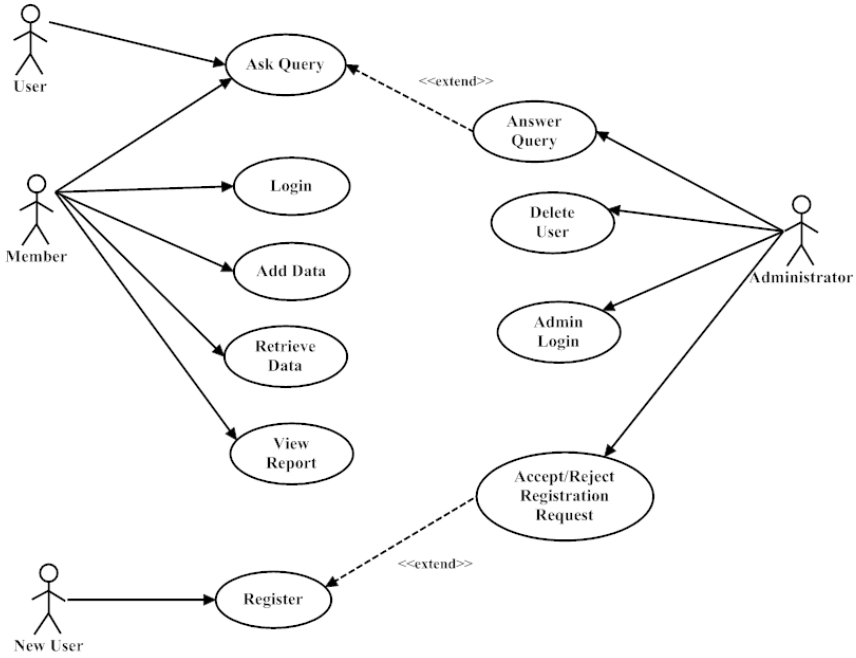


Figure 4.5 Use case diagram for BI Cloud Portal

Sequence Diagram

A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner. Figure 4.6 – Figure 4.8 show sequence diagrams for BI Cloud portal for different objects – user, new user and member. Figure 4.6 shows the interaction between user and administrator. Figure 4.7 shows the interaction between new user and administrator. Figure 4.8 shows the interaction of member with server and administrator.

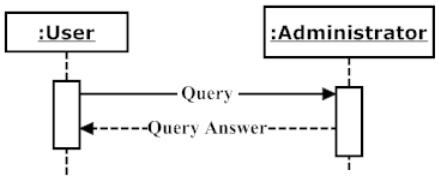


Figure 4.6 Sequence diagram of user and administrator

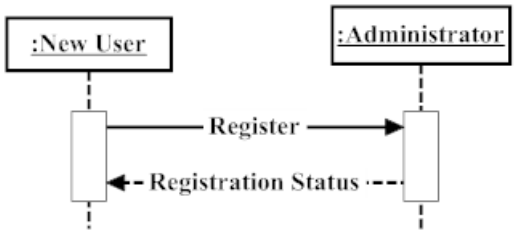


Figure 4.7 Sequence diagram of new user and administrator

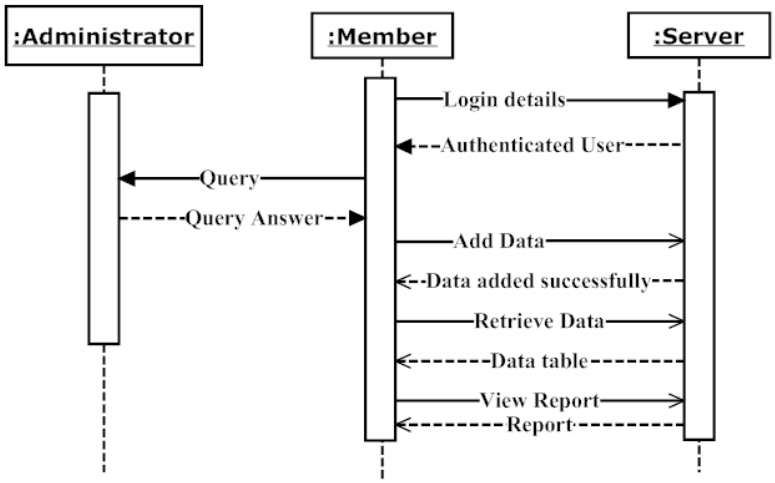


Figure 4.8 Sequence diagram of member with server and administrator

Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. Activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control. Figure 4.9 – figure 4.11 shows the activity diagram of various functions of BI Cloud Portal.

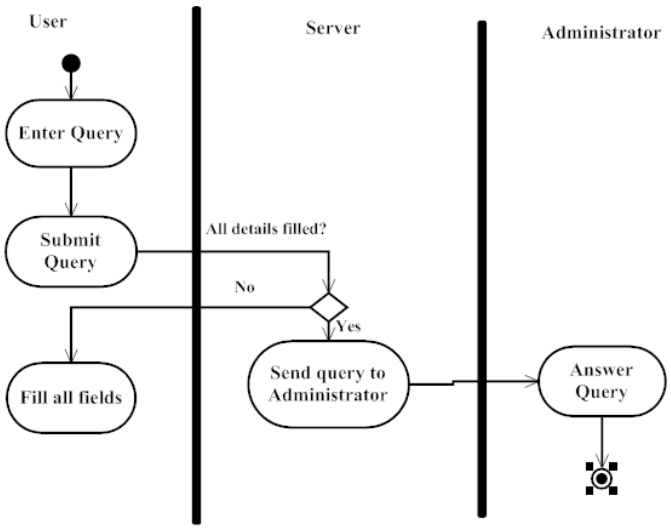


Figure 4.9 Activity diagram for query processing

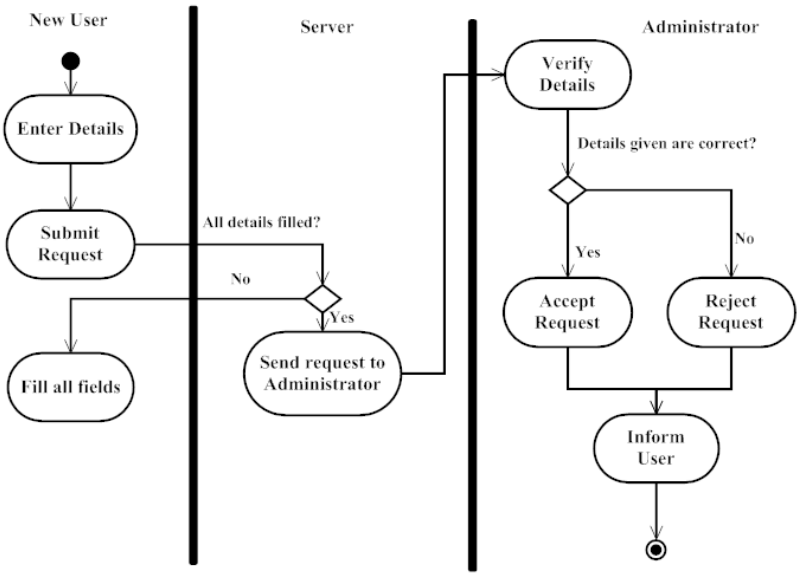


Figure 4.10 Activity diagram showing the registration of new user

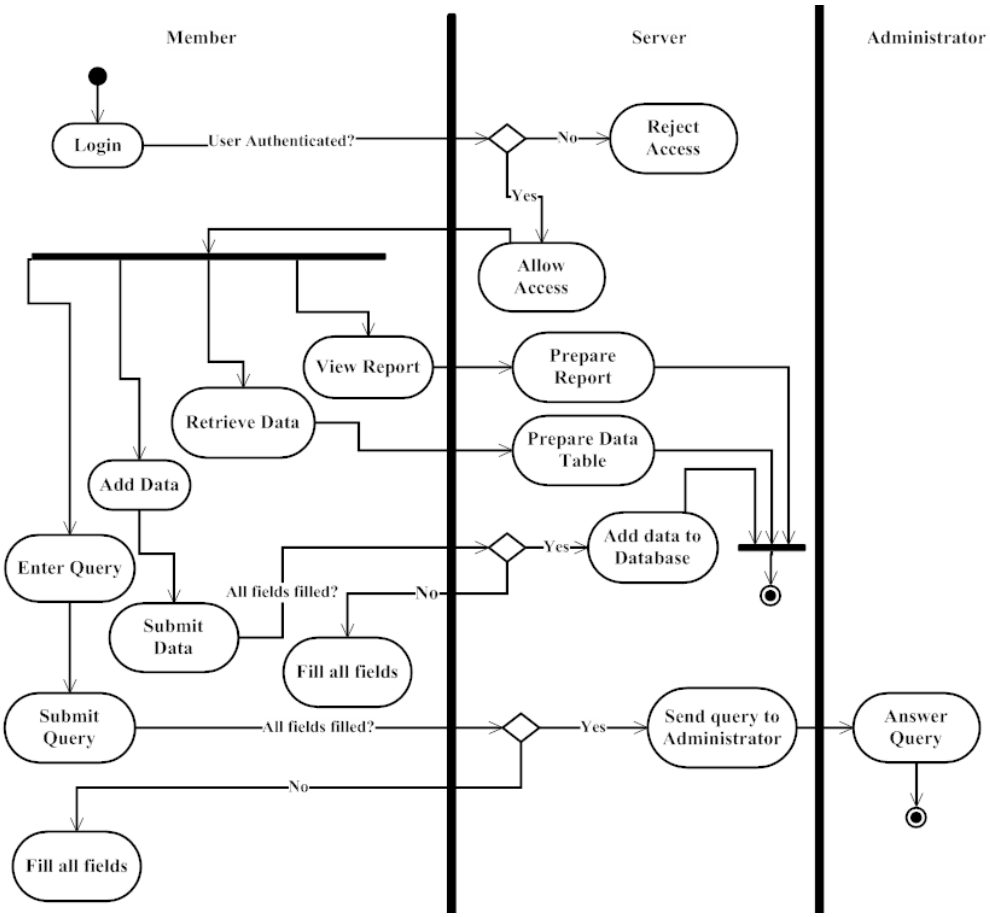


Figure 4.11 Activity diagram showing activities after a member log in.

Deployment Diagram

Deployment Diagram models the physical deployment of artifacts on nodes. Deployment diagram depicting hardware components where BI Cloud Portal deployed is shown in figure 4.12.

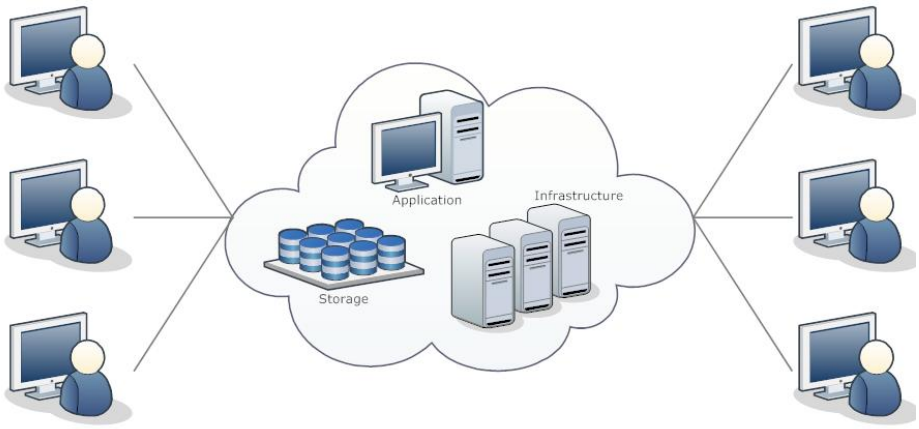


Figure 4.12 Deployment Diagram

4.2 Tool Alternatives for setting Cloud Environment

Cloud applications have different composition, configuration, and deployment requirements. So, different tools are available for setting cloud environment. Following section gives brief description of popular tools commonly utilized for creating Cloud environment.

4.2.1 CloudSim

Specifically in the case of Cloud computing, where access to the infrastructure incurs payments in real currency, simulation-based approaches offer significant benefits to Cloud customers by allowing them to - test their services in repeatable and controllable environment free of cost; and tune the performance bottlenecks before deploying on real Clouds [24].

CloudSim is an extensible simulation toolkit that enables modelling and simulation of Cloud computing environments. The CloudSim toolkit supports modelling and creation of one or more virtual machines (VMs) on a simulated node of a Data Center, jobs, and their mapping to suitable VMs. It also allows simulation of multiple Data Centers to enable a study on federation and associated policies for migration of VMs for reliability and automatic scaling of applications [24].

4.2.2 OpenNebula

OpenNebula is one of the key technologies of reservoir plan and the flagship research project in virtualization infrastructure and cloud computing of European Union. Like nimbus, OpenNebula is also an open source cloud service framework. It allows user deploy and manage virtual machines on physical resources and it can set user's data centers or clusters to flexible virtual infrastructure that can automatically adapt to the change of the service load. The main difference of OpenNebula and nimbus is that

nimbus implements remote interface based on EC2 or WSRF through which user can process all security related issues, while OpenNebula does not [25].

OpenNebula is also an open and flexible virtual infrastructure management tool, which can use to synchronize the storage, network and virtual techniques, and let users dynamically deploy services on the distributed infrastructure according to the allocation strategies at data center and remote cloud resources. Through the interior interfaces and OpenNebula data center environment, users can easily deploy any types of clouds. OpenNebula is mainly used to manage the data center of private cloud and infrastructure of cluster and it also support hybrid cloud to connect the local and public infrastructure. This is very useful to build high scalable cloud computing environment. Besides, OpenNebula also supports public cloud platform by providing interfaces and functions to virtual machines, storage and network management and so on. Through the control interfaces, users can access services provided by OpenNebula cloud computing platform [25].

4.2.3 Hadoop

The Apache™ Hadoop™ software library is a framework that allows for the distributed processing of large data sets across clusters of computers using a simple programming model. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures. The project includes these subprojects [26]:

Hadoop Common

The common utilities that support the other Hadoop subprojects. It provides access to the filesystems supported by Hadoop. The Hadoop Common package contains the necessary JAR files and scripts needed to start Hadoop. The package also provides source code, documentation, and a contribution section which includes projects from the Hadoop Community.

Hadoop Distributed File System (HDFS)

A distributed file system that provides high throughput access to application data. It is a distributed, scalable, and portable filesystem written in Java for the Hadoop framework. Each node in a Hadoop instance typically has a single datanode; a cluster of datanodes form the HDFS cluster. The HDFS stores large files across multiple

machines. It achieves reliability by replicating the data across multiple hosts, and hence does not require RAID storage on hosts.

Hadoop MapReduce

A software framework for distributed processing of large data sets on compute clusters. MapReduce is a framework for processing huge datasets on certain kinds of distributable problems using a large number of computers (nodes), collectively referred to as a cluster (if all nodes use the same hardware) or as a grid (if the nodes use different hardware). Computational processing can occur on data stored either in a filesystem (unstructured) or within a database (structured).

- "Map" step: The master node takes the input, partitions it up into smaller sub-problems, and distributes those to worker nodes. A worker node may do this again in turn, leading to a multi-level tree structure. The worker node processes that smaller problem, and passes the answer back to its master node.
- "Reduce" step: The master node then takes the answers to all the sub-problems and combines them in some way to get the output – the answer to the problem it was originally trying to solve.

The advantage of MapReduce is that it allows for distributed processing of the map and reduction operations.

4.2.4 Google App Engine

Google App Engine is “a system that exposes various pieces of Google’s scalable infrastructure so that you can write server-side applications on top of them”. Simply this is a platform which allows users to run and host their web applications on Google’s infrastructure. These applications are easy to build, easy to maintain and easy to scale whenever traffic and data storage needed. By using Google’s App Engine, there are no servers to maintain and no administrators needed. The idea is user just to upload his application and it is ready to serve its own customers. User has a choice either his product to be served by the free domain appspot.com or to allow Google Apps to serve it from domain chosen by the customer. Google also provide the user with the option to limit the access of the application within the members of his own organization or to share it with the rest of the world. The starting packet is free of charge and additional obligation. All the user have to do is to sign up for a free account, and then to develop and upload his own application. The starting package

includes up to 500MB of storage and enough CPU power and bandwidth to serve 5 million page views per month [27].

The Application Environment

Several key features are included in the environment:

(i) dynamic web serving, with full support for common web technologies

(ii) persistent storage with queries, sorting and transactions

(iii) automatic scaling and load balancing

(iv) APIs for authenticating users and sending email using Google Accounts

(v) a fully featured local development environment that simulates Google App Engine on user's computer

The Python Runtime Environment

The Runtime environment provided by App Engine uses Python and Java programming language, but other languages as well are taken into account to be supported in the near future. The current supported version of the Python runtime environment is 2.5 and 2.7, which includes the standard Python library. All methods, except those ones which violates the sandbox restriction, like attempting to open a socket or write into a file, are included in the library. The modules which features are not supported by the standard are disabled, and the part of code which imports them will raise an error. All applications code must be written entirely only in Python language and code with extensions written in C is not supported. Rich APIs for the Google Accounts, URL fetch, Datastore and email servers are provided by the Python environment.

Simple web application framework called "webapp" which makes easy to start building applications is provided as well by App Engine. In addition Django Web application framework, version 0.96.1 is also included in App Engine, but it should be take into account that App Engine datastore is not a relational database, which is required by some Django components. The option to be included other third-party libraries is also supported. As long as they are implemented in Python programming language and if they not require any unsupported library modules, it should not be a problem to be used.

Chapter 5

Experimental Results

This chapter focuses on implementation of BI Cloud portal on Google App Engine and experimental results of this approach. This chapter also describes the benefits of BI Cloud Portal designed and developed for this thesis.

5.1. Implementation

BI Cloud Portal has been developed using Google App Engine using Python and HTML. Google App Engine is a PaaS which provides development platform to develop Cloud applications. Data is stored in Cloud environment i.e. at distributed sites provided by Google. The BI Cloud Portal can be assessed anywhere from the world by visiting the website "bicloudportal.appspot.com". Its basic functionalities are shown as snapshots below. Figure 5.1 shows the home page of portal.

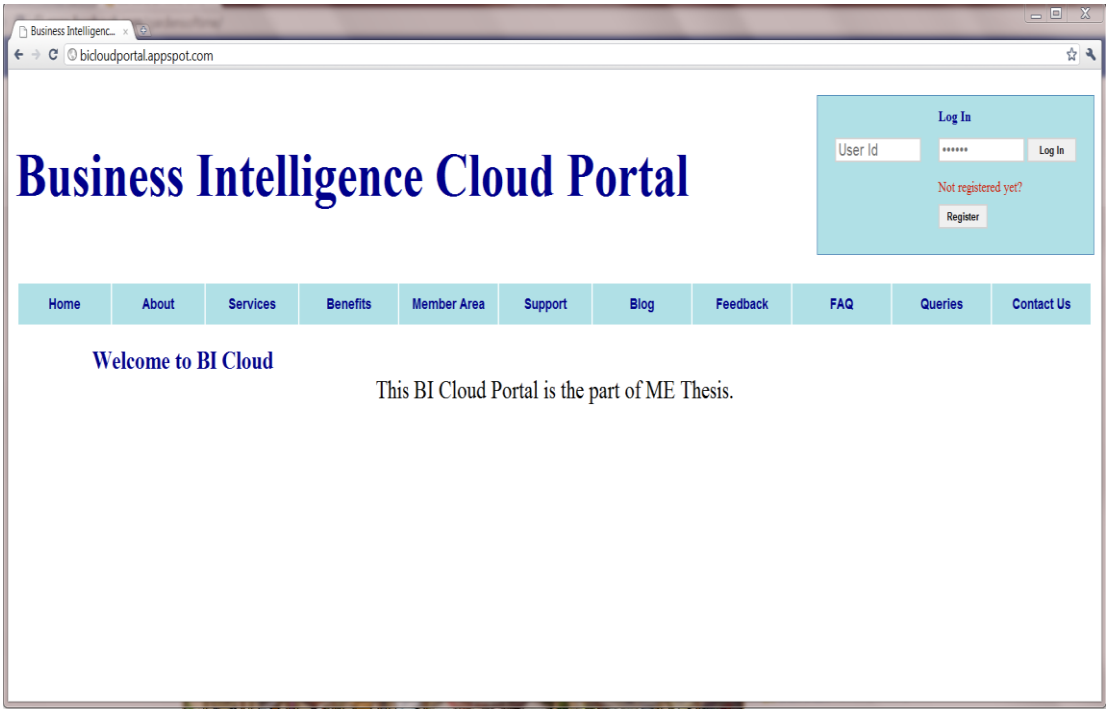


Figure 5.1 Home page of Portal

Figure 5.2 - 5.4 shows the queries being asked by a non-member user. Figure 5.2 shows the query page for non-member user. Figure 5.3 shows the query entered by the user. On clicking submit button from Figure 5.4, query is submitted to the admin and user is informed that the query will be answered within 24 hrs.

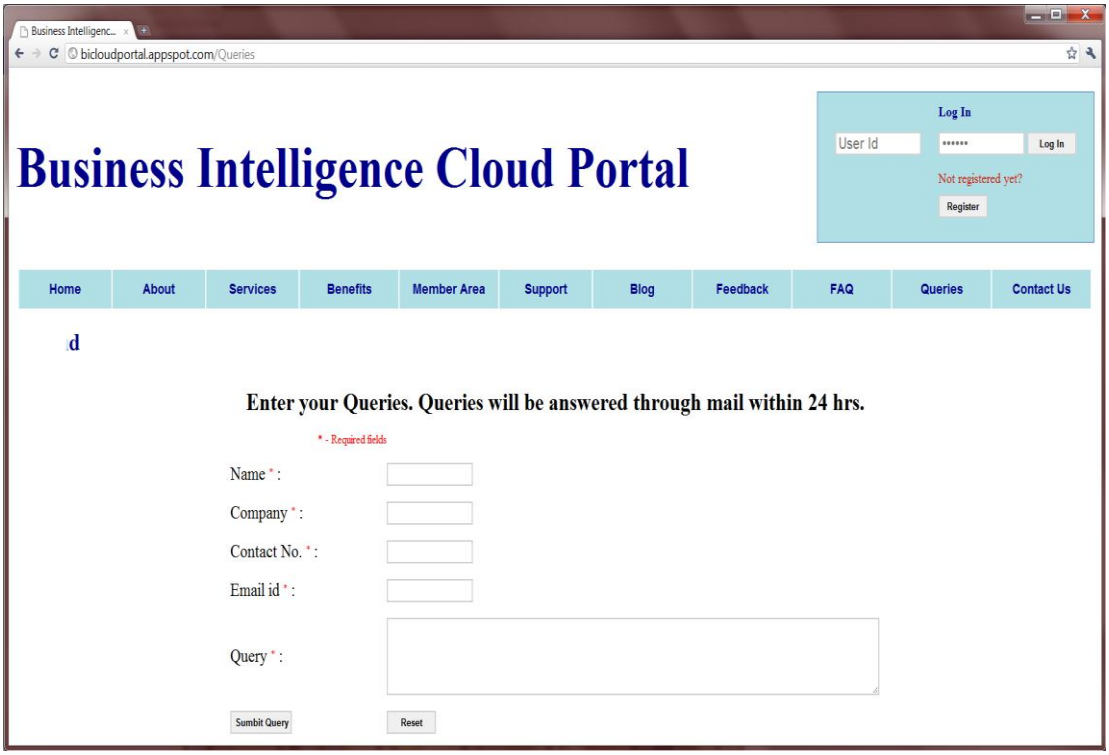


Figure 5.2 Query page for non-member user.

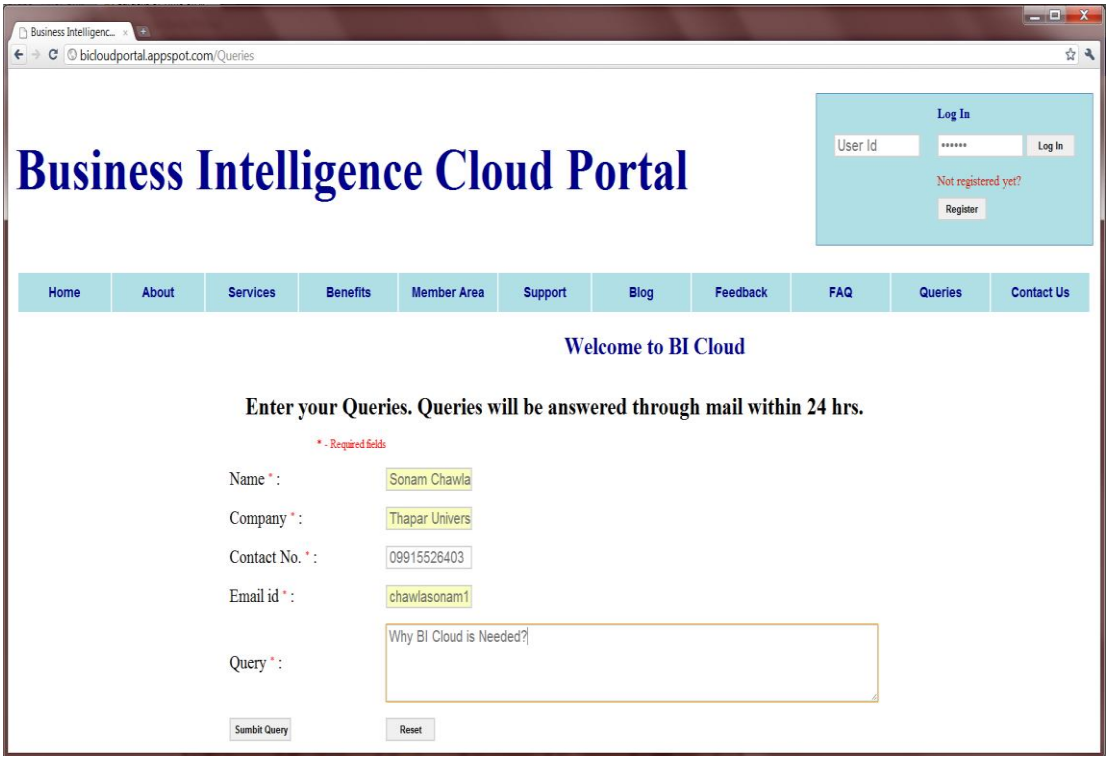


Figure 5.3 Query entered by non-member user.

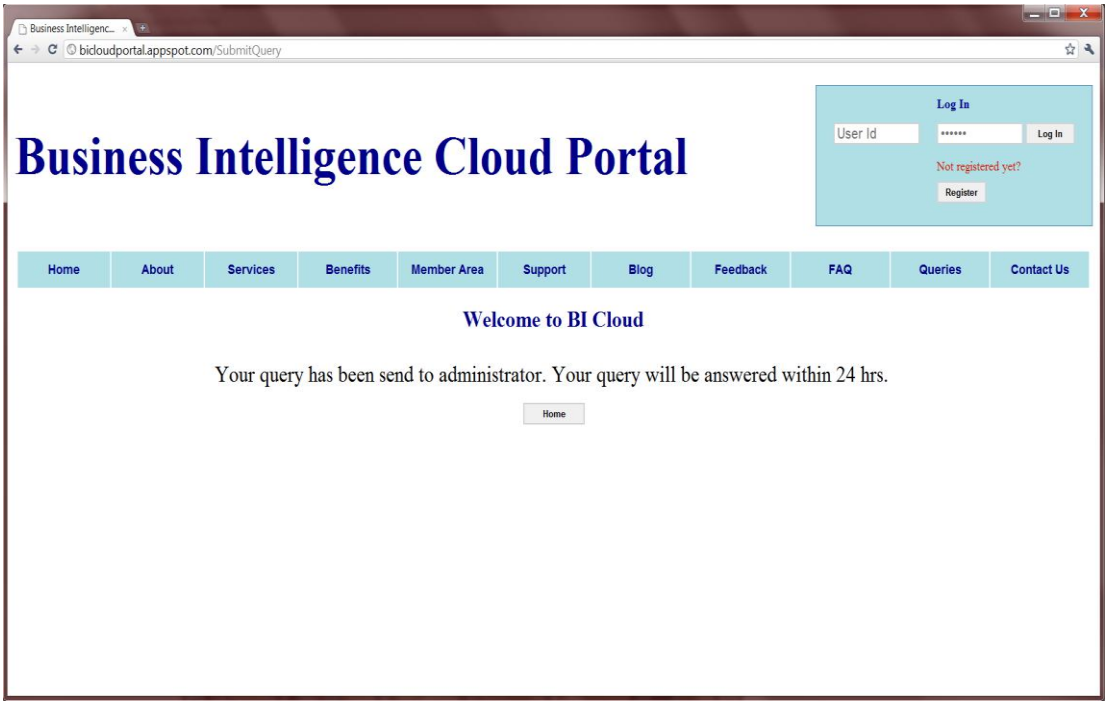


Figure 5.4 Message send by server regarding query answer.

Figure 5.5 – 5.12 show pages for registration to BI Cloud Portal. Figure 5.5 shows the registration page with full details entered by the new user. On clicking Register button of figure 5.5, registration request is send to the administrator for verification and user is asked to come back after 24 hrs as shown in figure 5.6.

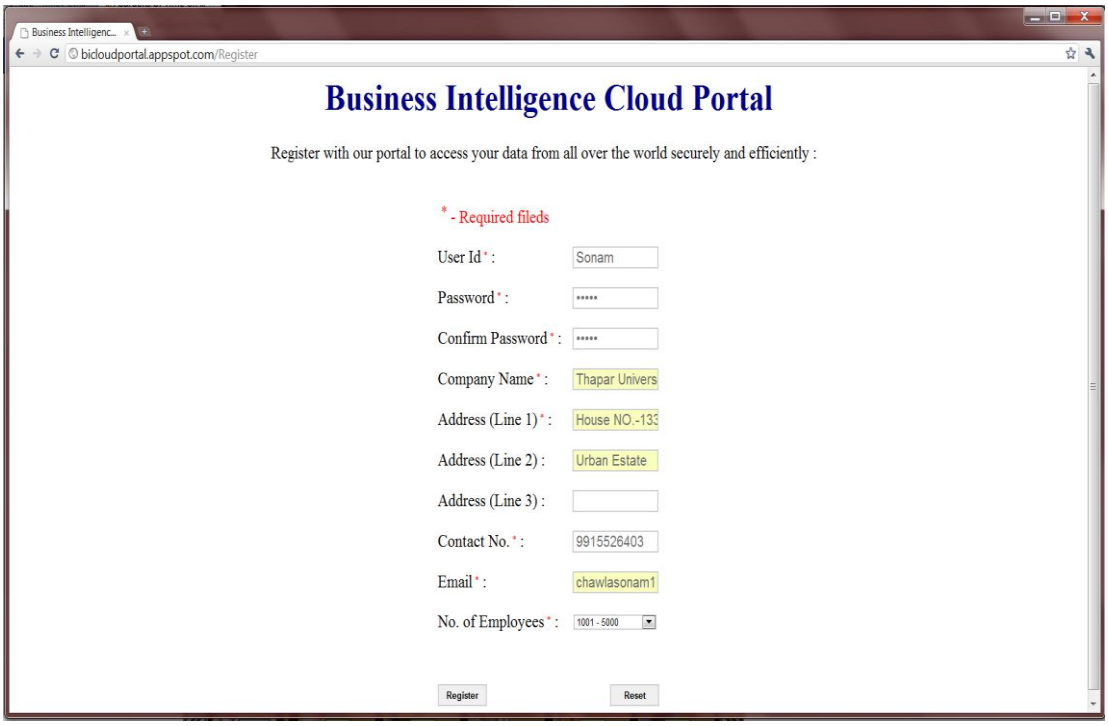


Figure 5.5 Registration page with new user's data filled.

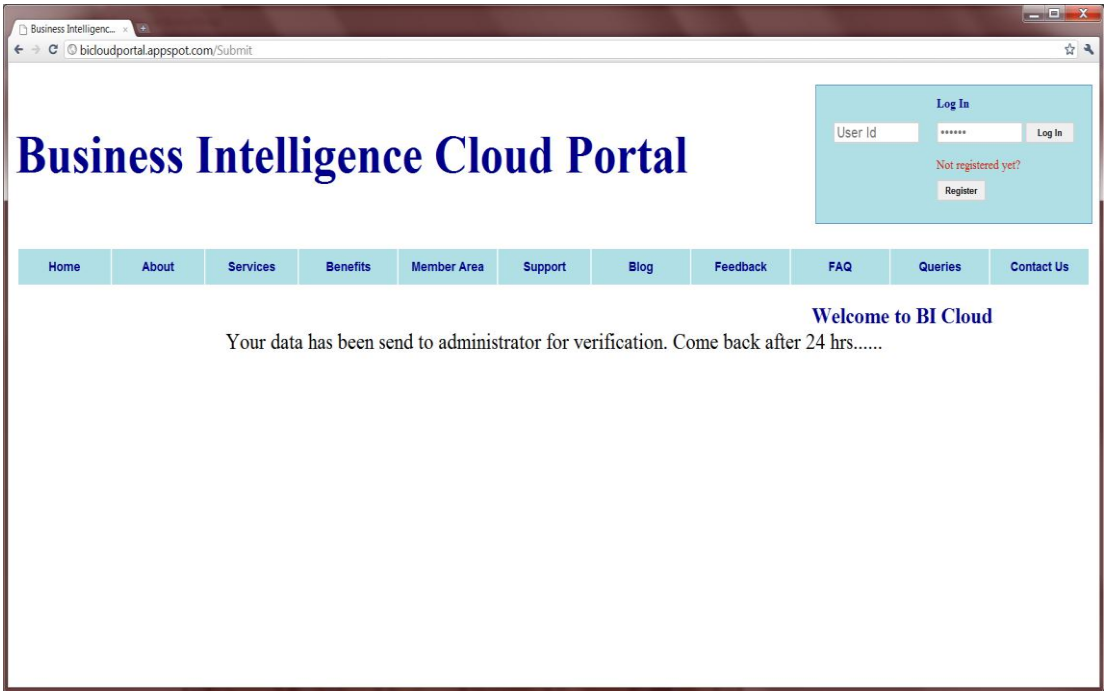


Figure 5.6 Message send by server regarding registration.

Figure 5.7 shows the page when user is trying to register with existing User Id. On clicking Register button of figure 5.7, server replies with registration error and user is asked to choose a new User Id as shown in figure 5.8.

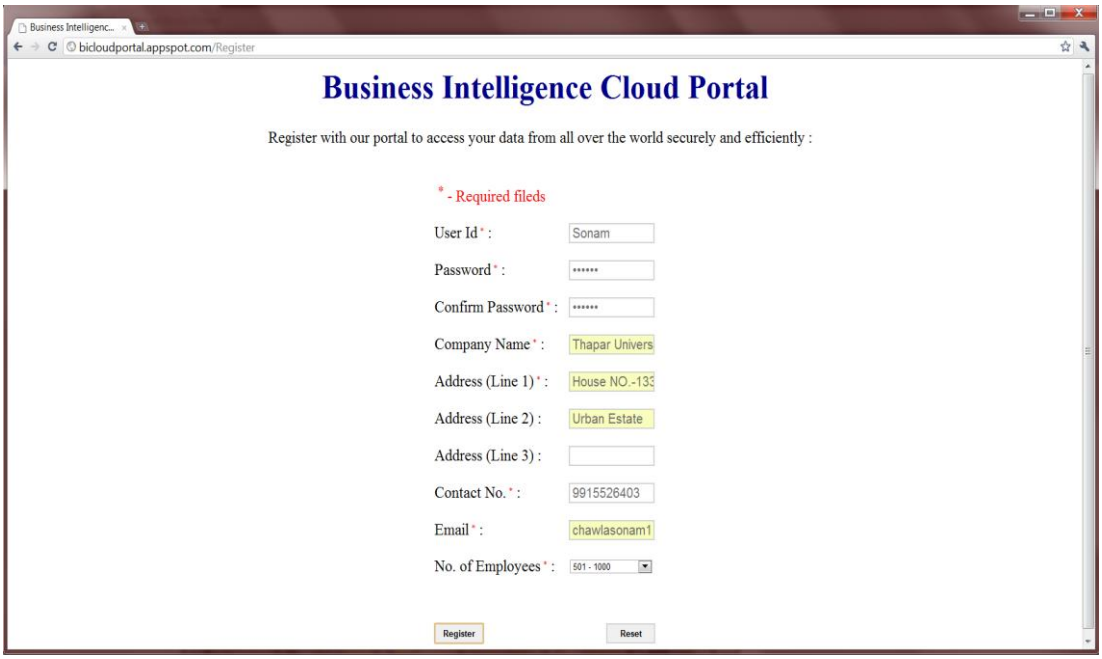


Figure 5.7 User trying to register with existing User Id

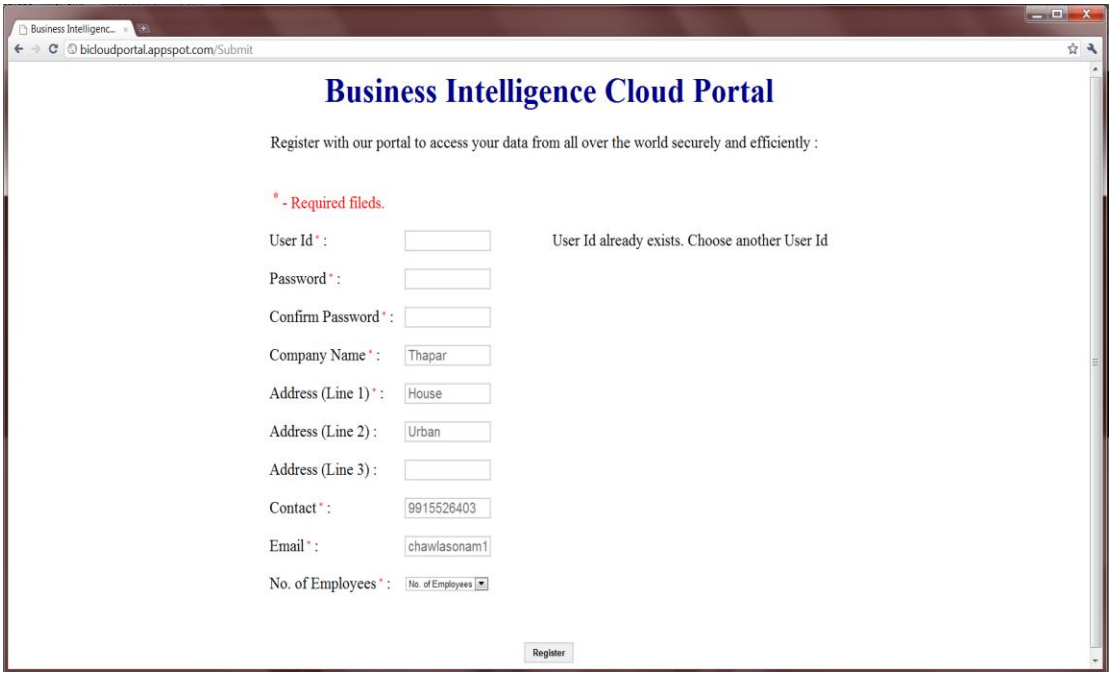


Figure 5.8 Registration error - User Id already exists

Figure 5.9 shows the page when user is trying to register with different values for Password and Confirm Password fields. On clicking Register button of figure 5.9, server replies with registration error and user is asked to enter same values for both fields as shown in figure 5.10.

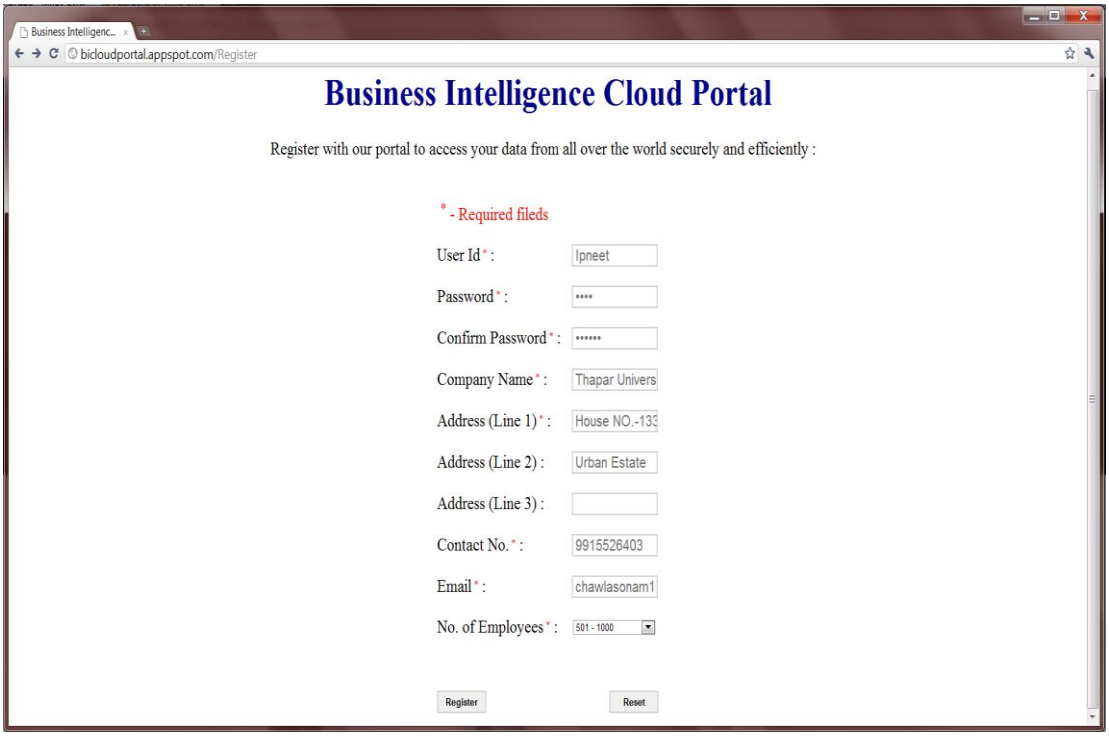


Figure 5.9 User trying to register with unmatched passwords

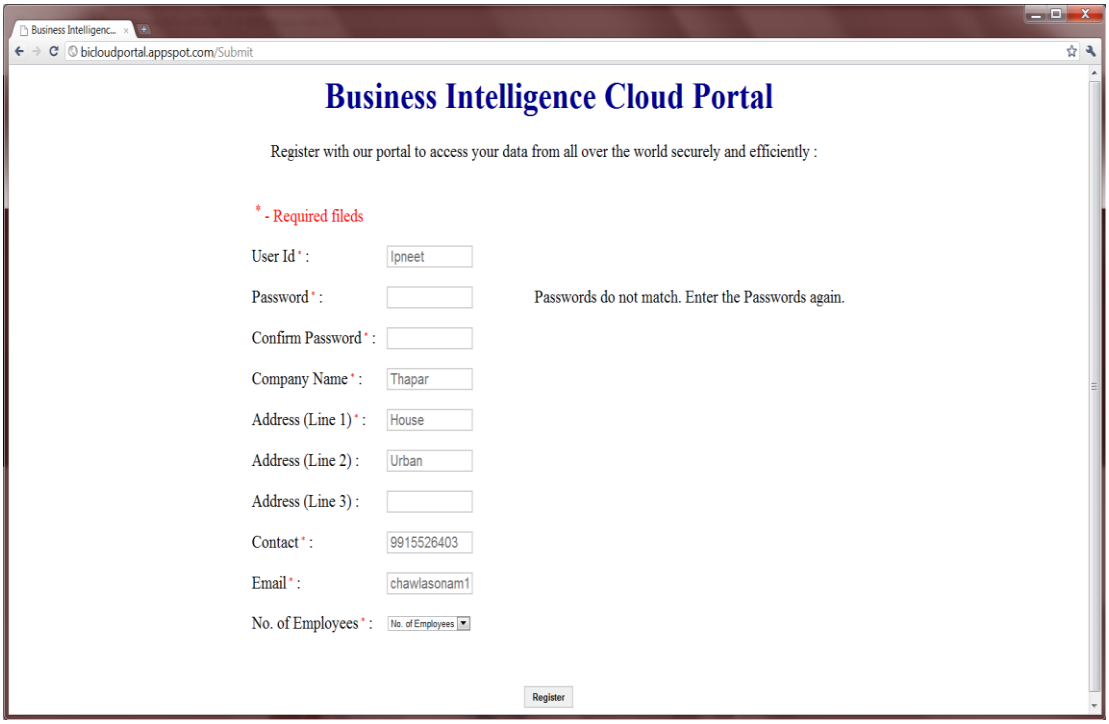


Figure 5.10 Registration error – Passwords do not match

Figure 5.11 shows the page when user is trying to register with some of the required fields blank. On clicking Register button of figure 5.11, server replies with registration error and user is asked to enter values for all the required fields as shown in figure 5.12.

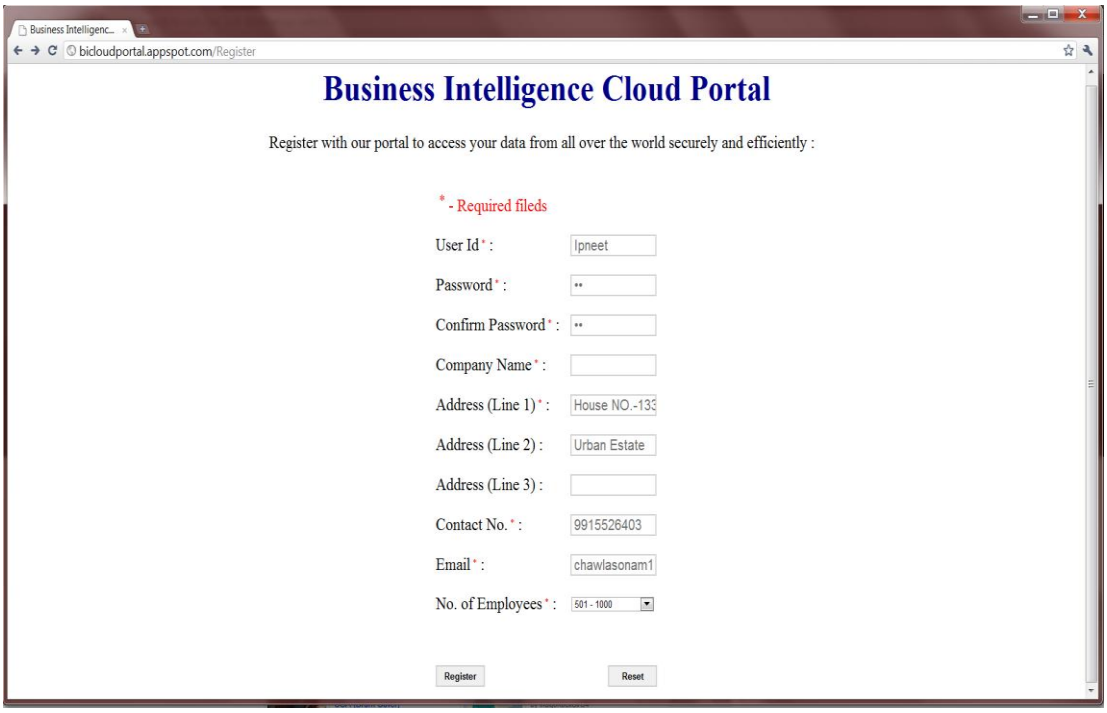


Figure 5.11 User trying to register with empty required fields.

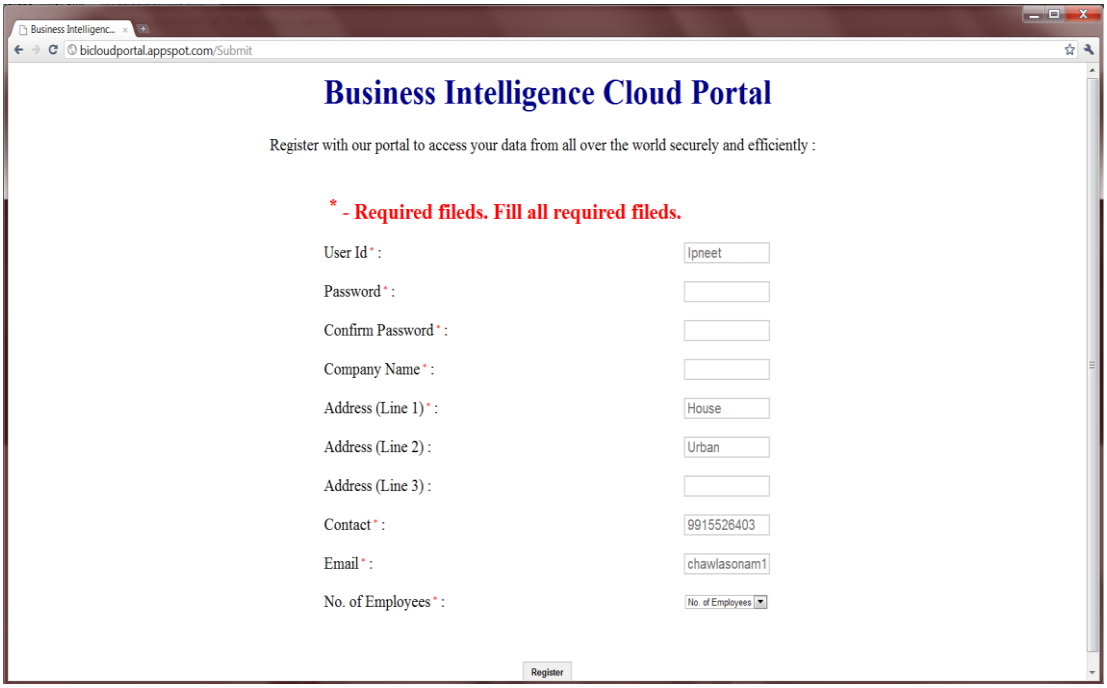


Figure 5.12 Registration error – Some required fields are left blank.

Figure 5.13 shows the home page after admin log-in using right credentials. Home page also shows the pending registration request. Figure 5.14 shows the selected action for the selected user. Action can be accept or reject.

Figure 5.13 Home page after admin log-in showing pending registration requests.

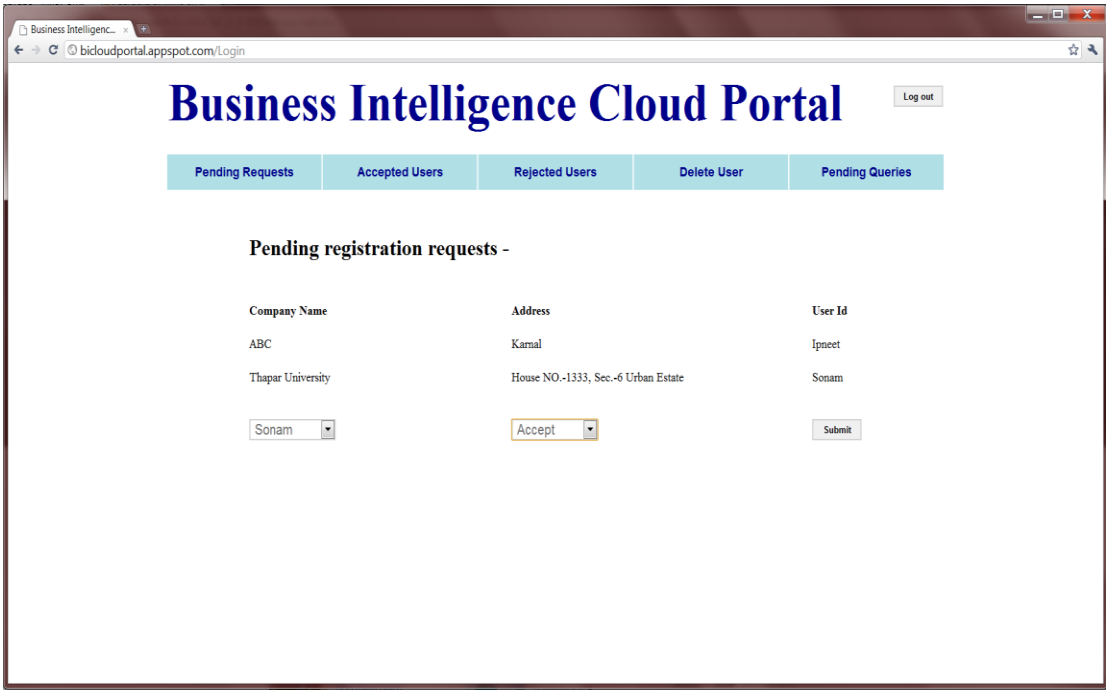


Figure 5.14 Selected action is accept for user

On clicking submit button, acceptance or rejection page opens up. Figure 5.15 shows the acceptance page which asks for Role of the user, role can be user or administrator. On clicking submit button, acceptance notification is mailed to the accepted user on given email-id as shown in Figure 5.16.

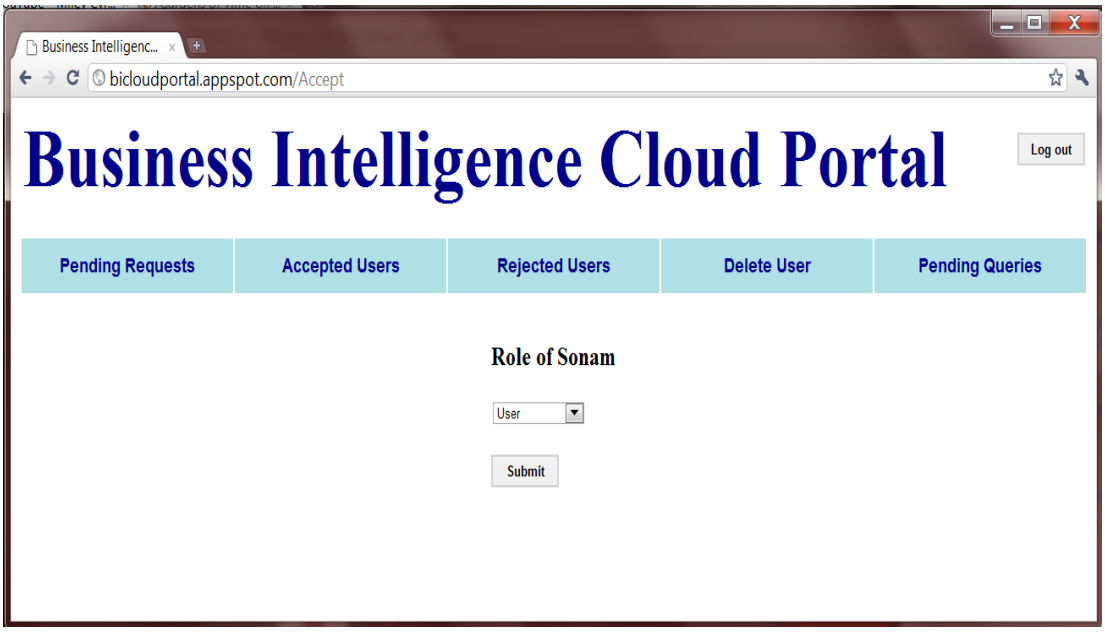


Figure 5.15 Selecting role for accepted user

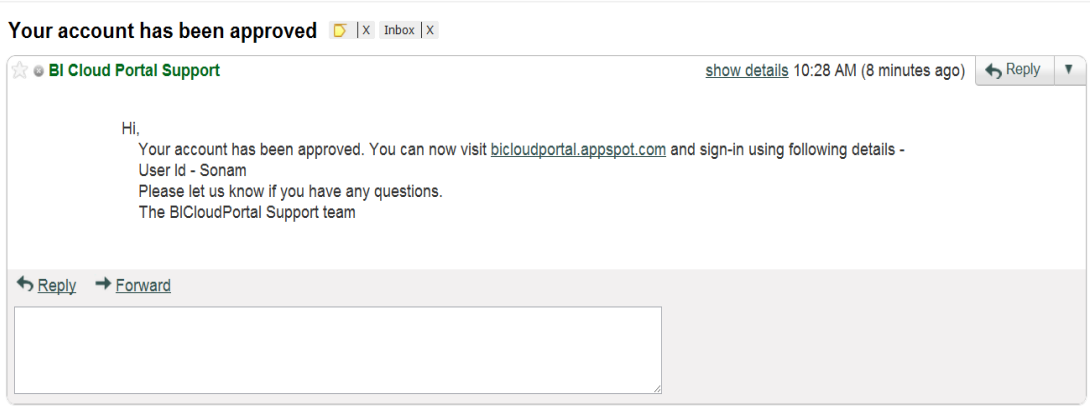


Figure 5.16 Email notification showing the acceptance of the registration request.

Figure 5.17 shows the selected action (i.e. reject) for the selected user. Figure 5.18 shows the rejection page which asks the reason for rejection.

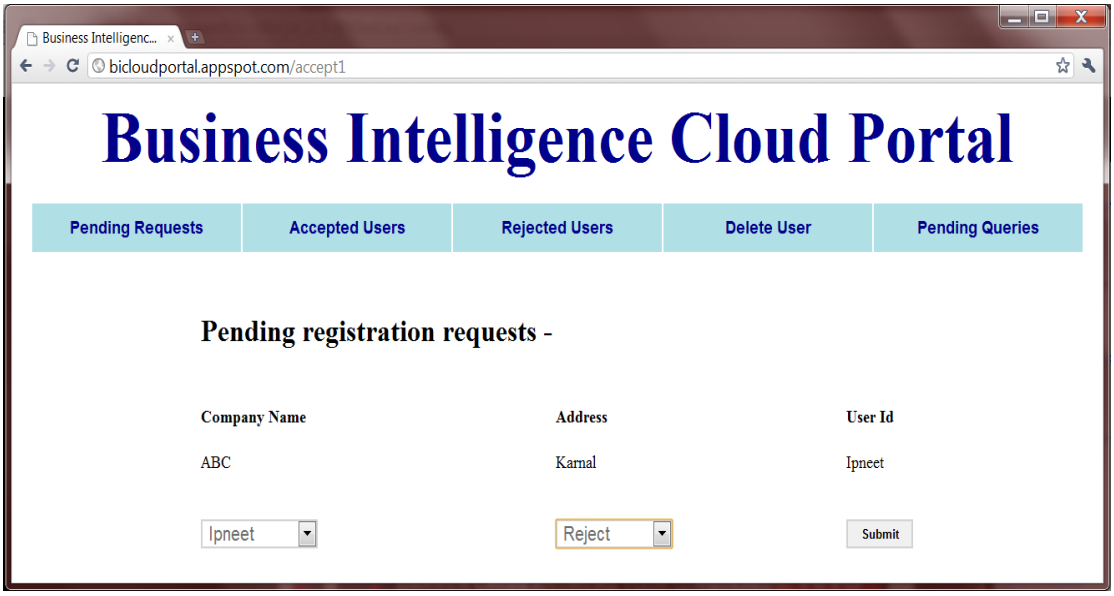


Figure 5.17 Selected action is reject for the user

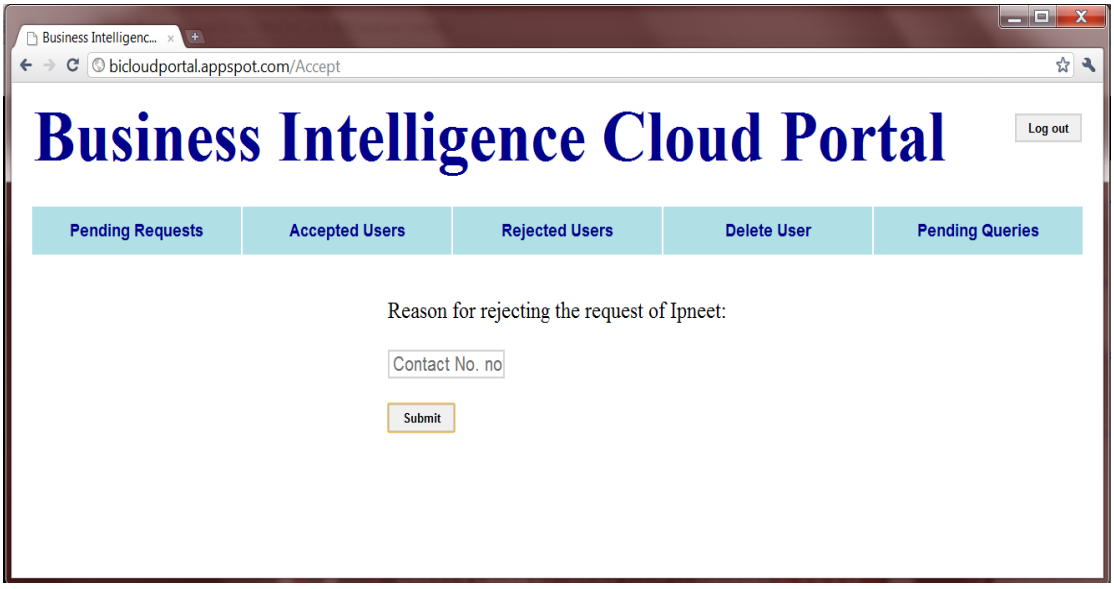


Figure 5.18 Reason for rejecting registration request.

Figure 5.19 shows the list of accepted users (members) of BI Cloud Portal and their details. Figure 5.20 shows the list of rejected users and the reason for rejecting them.

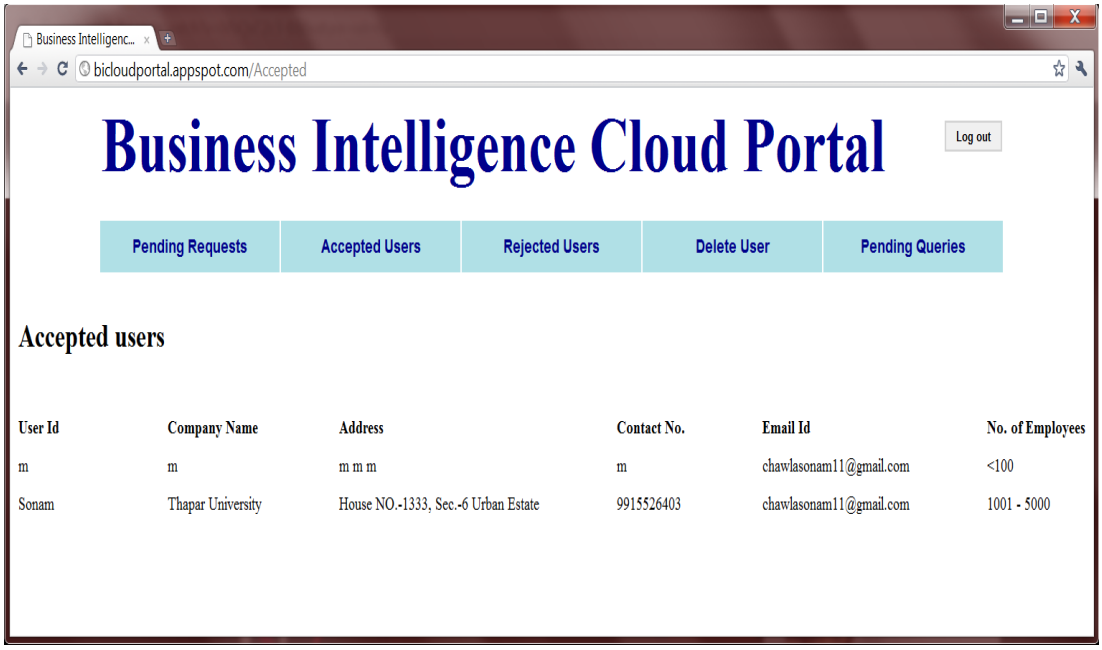


Figure 5.19 List of members (accepted users)

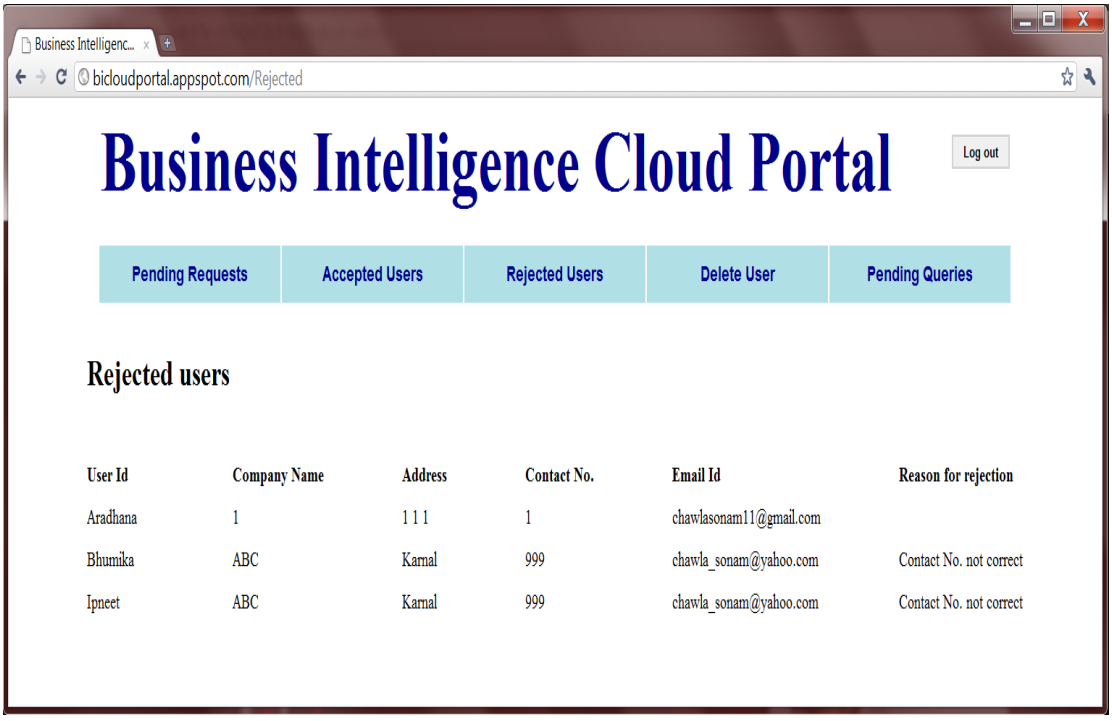


Figure 5.20 List of rejected users

Figure 5.21 shows the list of members and an option to delete an accepted user. On clicking delete button, selected user will be deleted from members list.

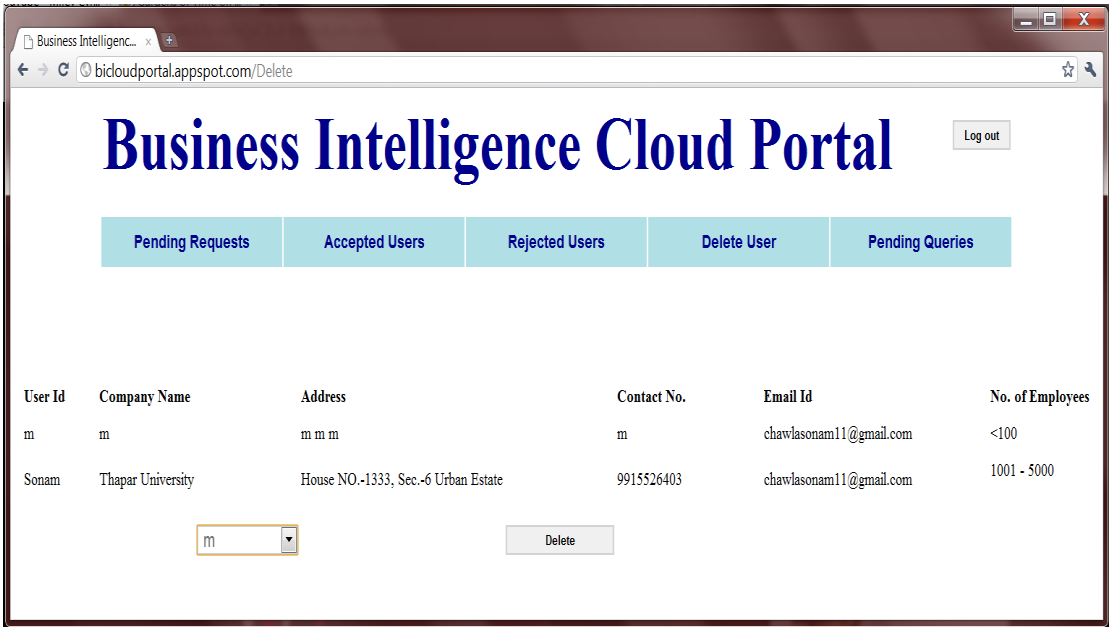


Figure 5.21 Page to delete a selected member from accepted users list.

Figure 5.22 shows the list of queries to be answered and selected user whose query is being answered. On clicking Answer query button, page for answering the query is opened as shown in figure 5.23. The answer to the query is mailed to the user on given email-id as shown in figure 5.24.

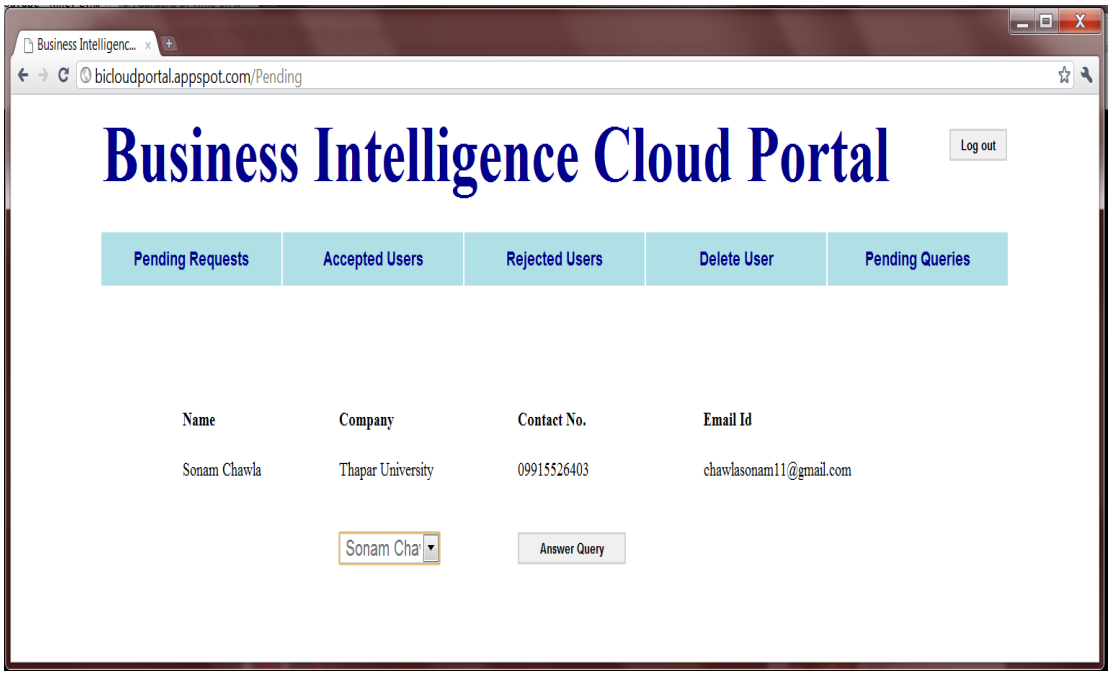


Figure 5.22 Pending queries and their details

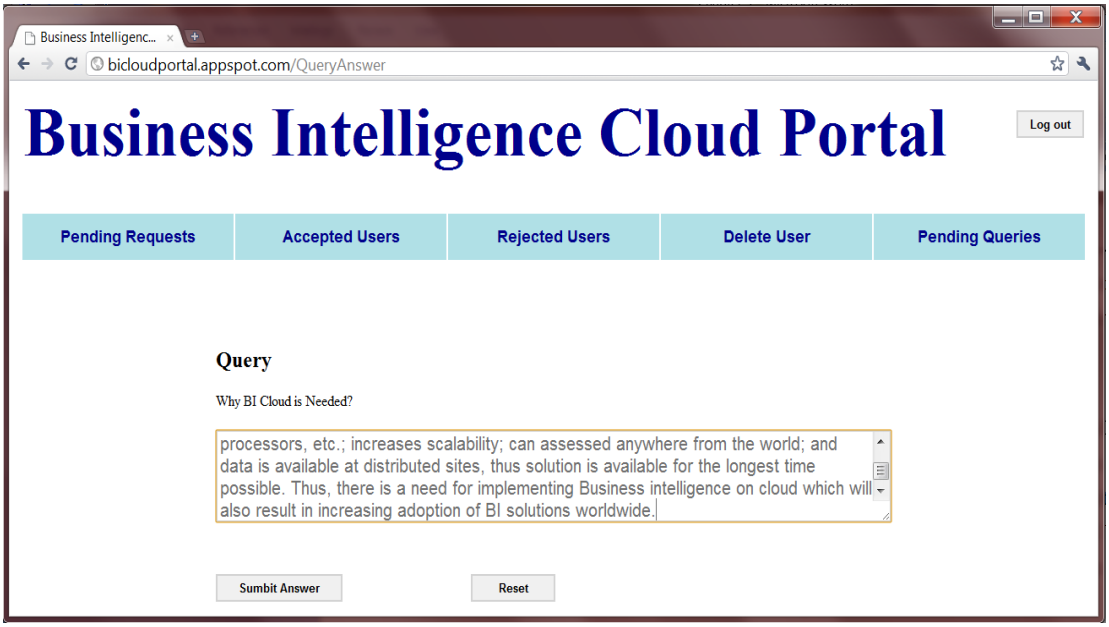


Figure 5.23 Query and answer of the query being answered by the admin.

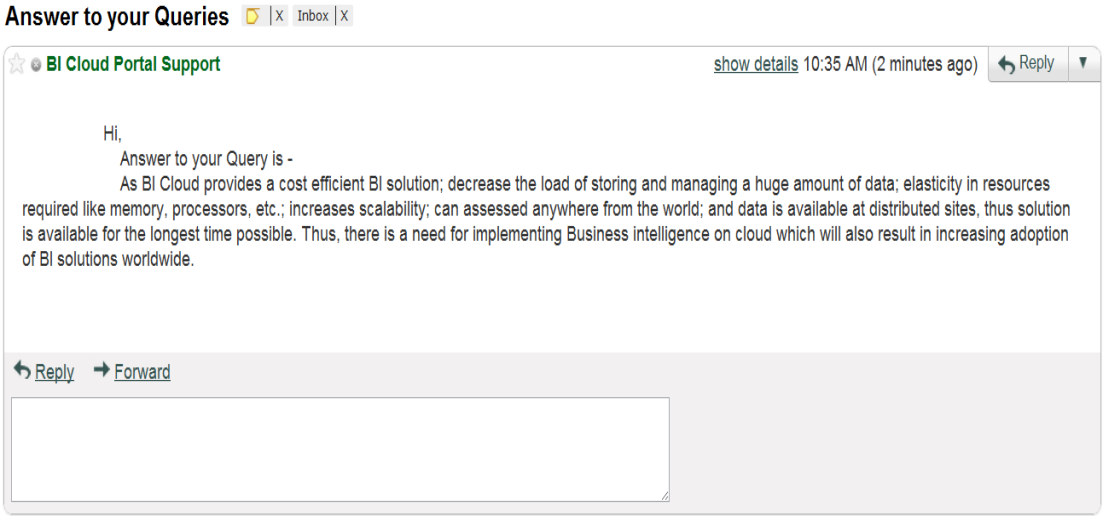


Figure 5.24 Answer to the query mailed to the user by admin.

Figure 5.25 shows the home page after a member log-in. It also shows the page to add data to the database. Figure 5.26 shows the data added by the member. On clicking add data button, data is added to the database and page to add new data opens up as shown in figure 5.27.

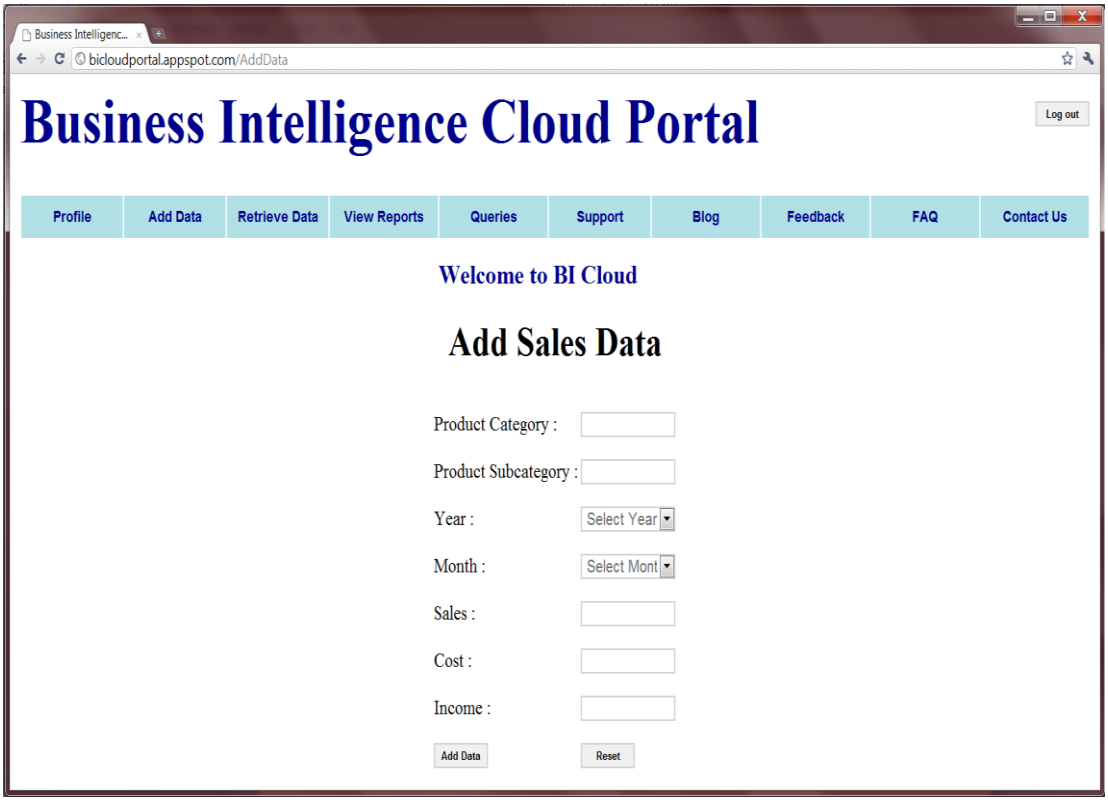


Figure 5.25 Home page for member and to add data in the database.

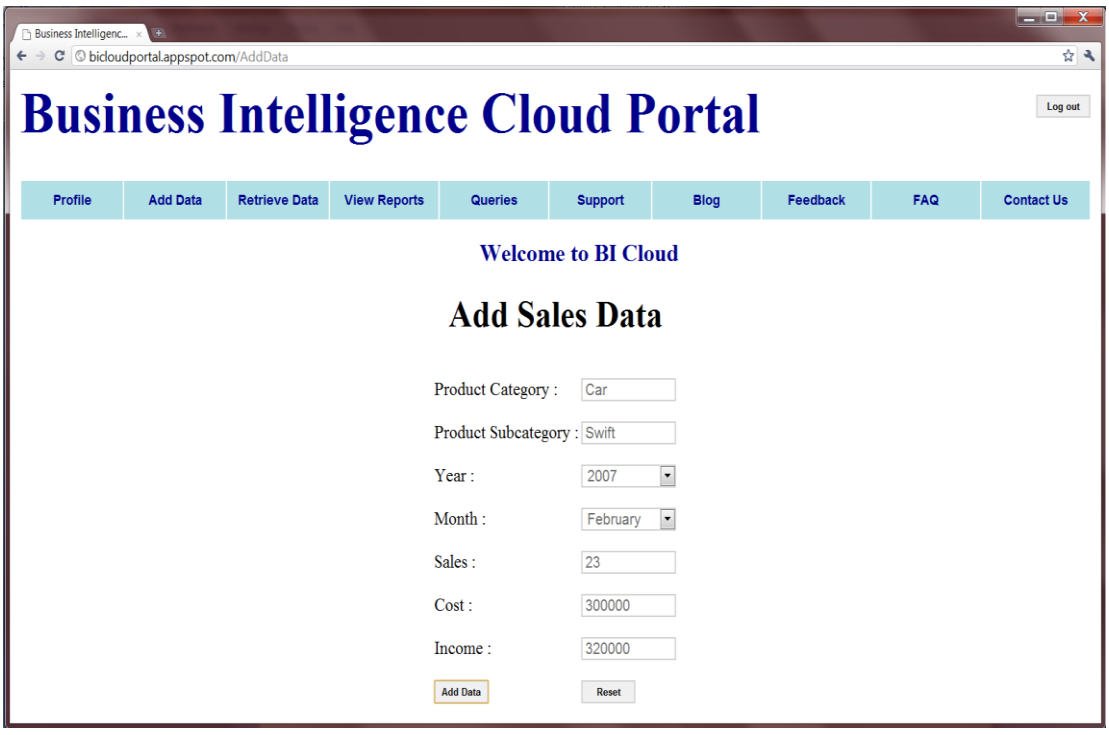


Figure 5.26 Data entered by the member to add to the database.

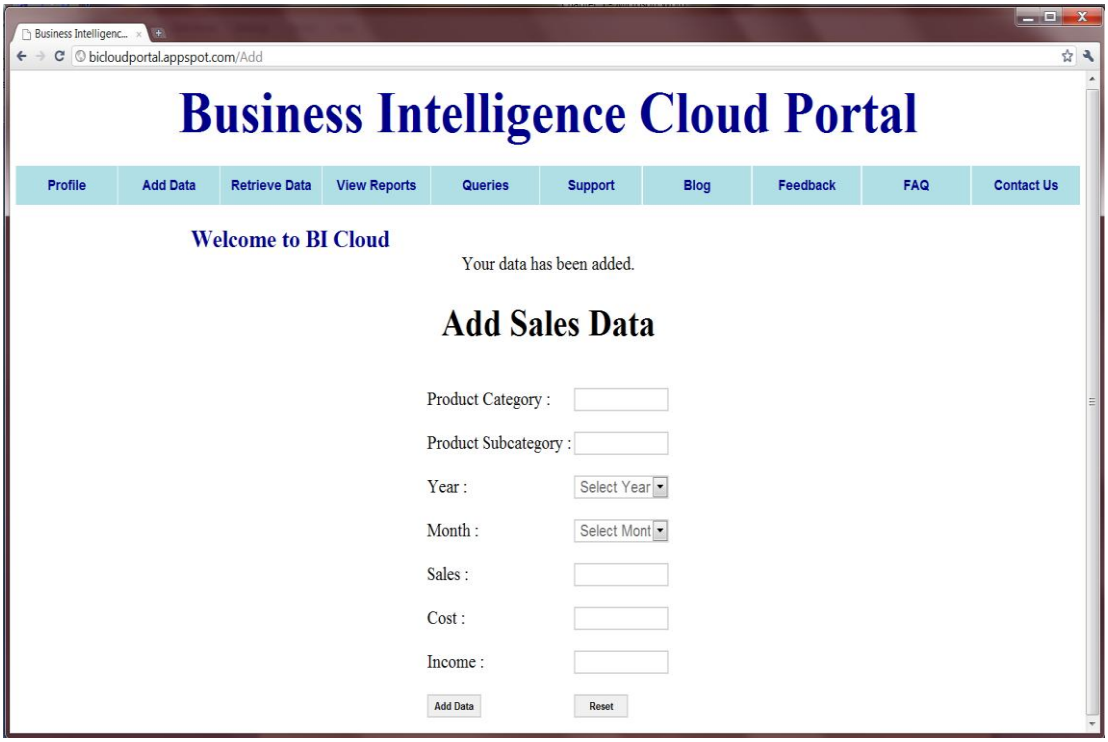
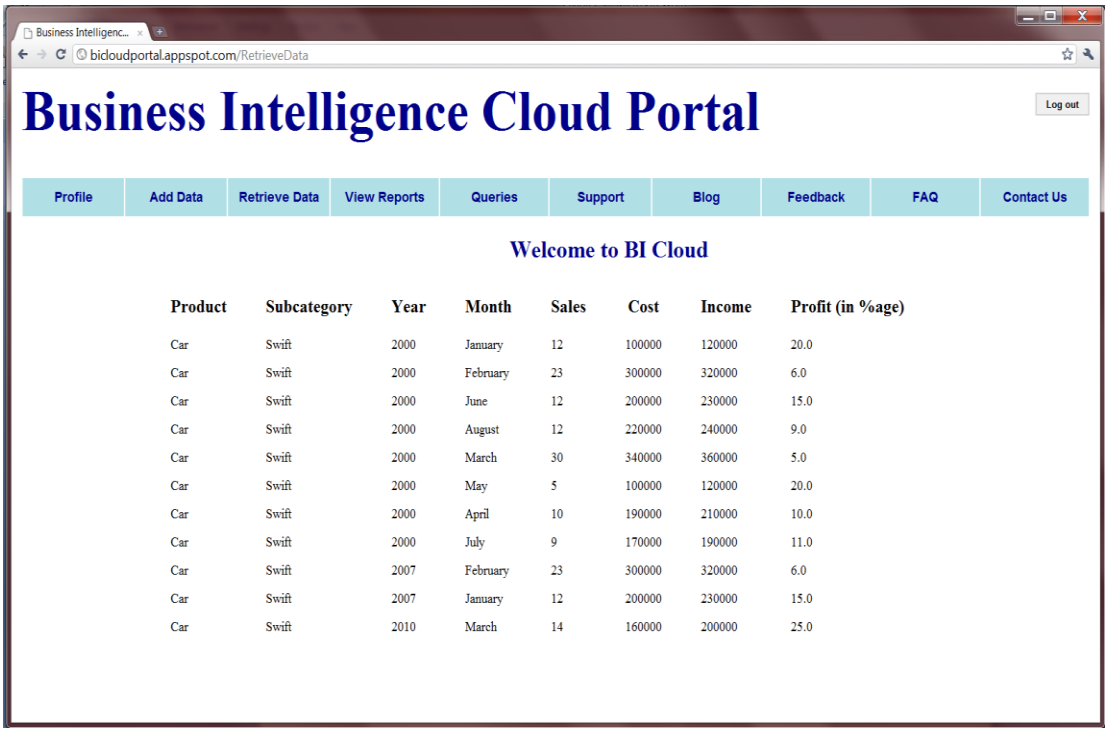


Figure 5.27 Page to add new data.

Data stored in the database can be retrieved by selecting the Retrieve Data option. Figure 5.28 shows the data stored in the database. Data is stored and retrieved from Google database at distributed sites.



Product	Subcategory	Year	Month	Sales	Cost	Income	Profit (in %age)
Car	Swift	2000	January	12	100000	120000	20.0
Car	Swift	2000	February	23	300000	320000	6.0
Car	Swift	2000	June	12	200000	230000	15.0
Car	Swift	2000	August	12	220000	240000	9.0
Car	Swift	2000	March	30	340000	360000	5.0
Car	Swift	2000	May	5	100000	120000	20.0
Car	Swift	2000	April	10	190000	210000	10.0
Car	Swift	2000	July	9	170000	190000	11.0
Car	Swift	2007	February	23	300000	320000	6.0
Car	Swift	2007	January	12	200000	230000	15.0
Car	Swift	2010	March	14	160000	200000	25.0

Figure 5.28 Stored data in the database.

Figure 5.29 – 5.44 show different types of reports (Cost v/s Income, Sales and Profit Report) using different charts (Bar, Pie and Area Chart) that can be generated through

this portal. Figure 5.29 shows the selected option by the user to generate monthly bar chart of cost v/s income. Figure 5.30 shows the bar chart generated according to the user requirements.

Figure 5.29 Selection of type of report and chart.

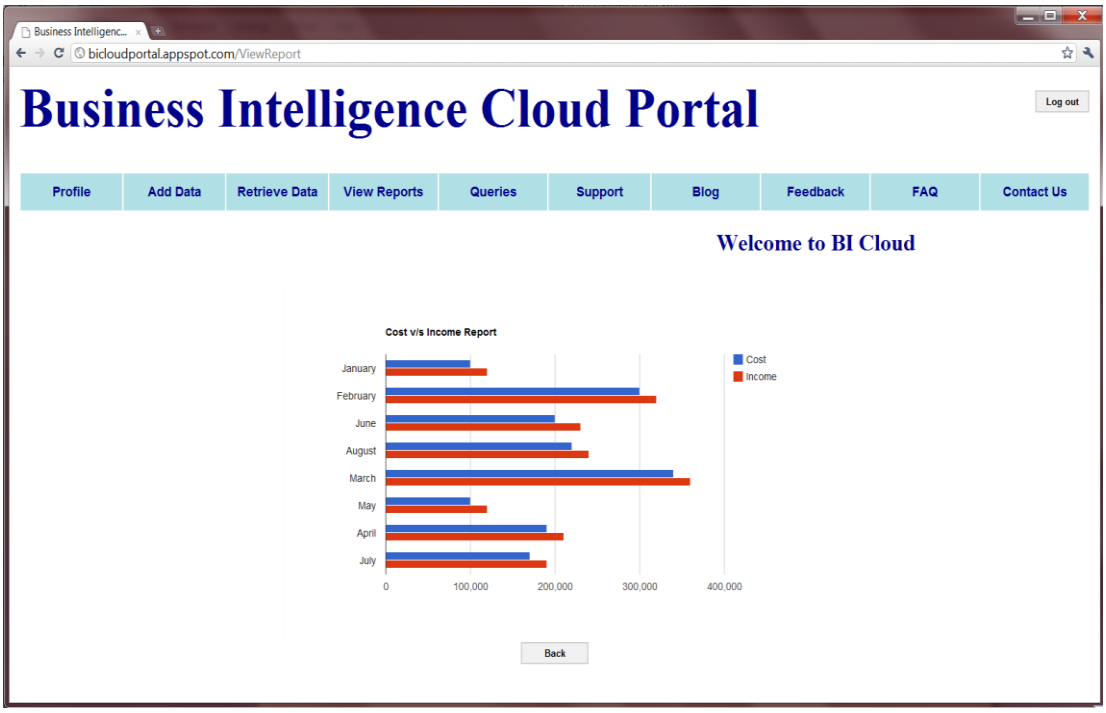


Figure 5.30 Monthly cost v/s income report shown in the form of Bar Chart as selected in figure 5.29.

Figure 5.31 shows the selected option by the user to generate monthly area chart of cost v/s income. Figure 5.32 shows the area chart generated according to the user requirements.

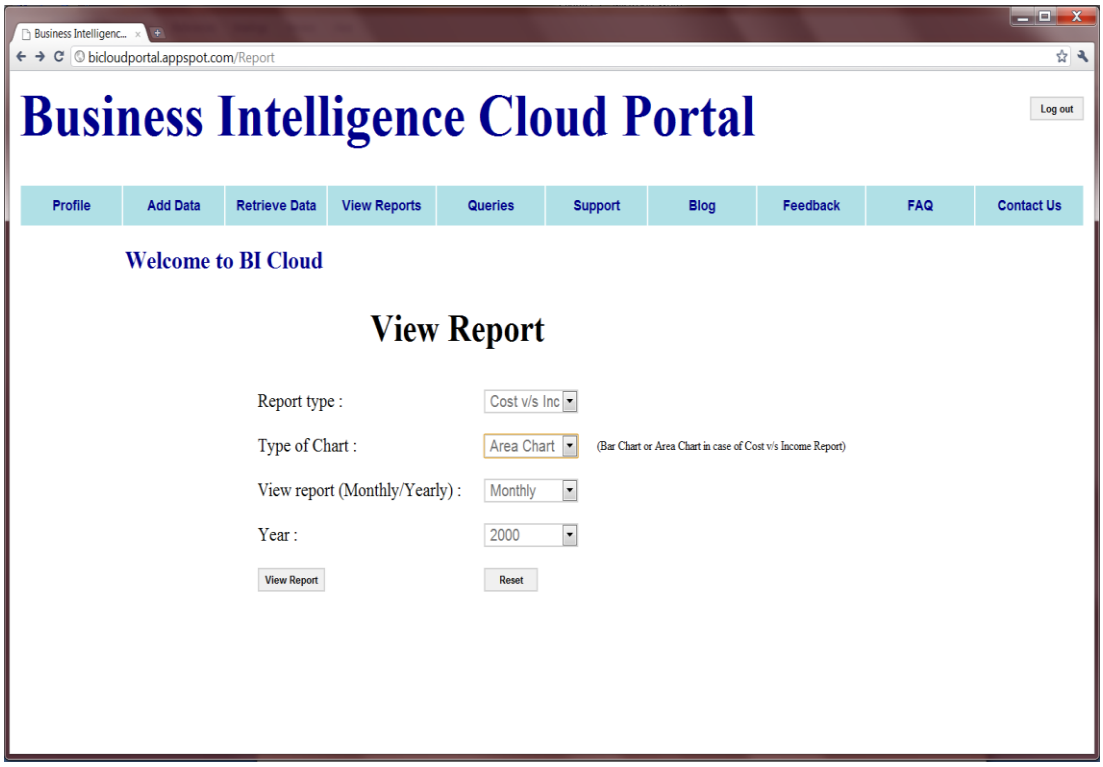


Figure 5.31 Selection of type of report and chart.

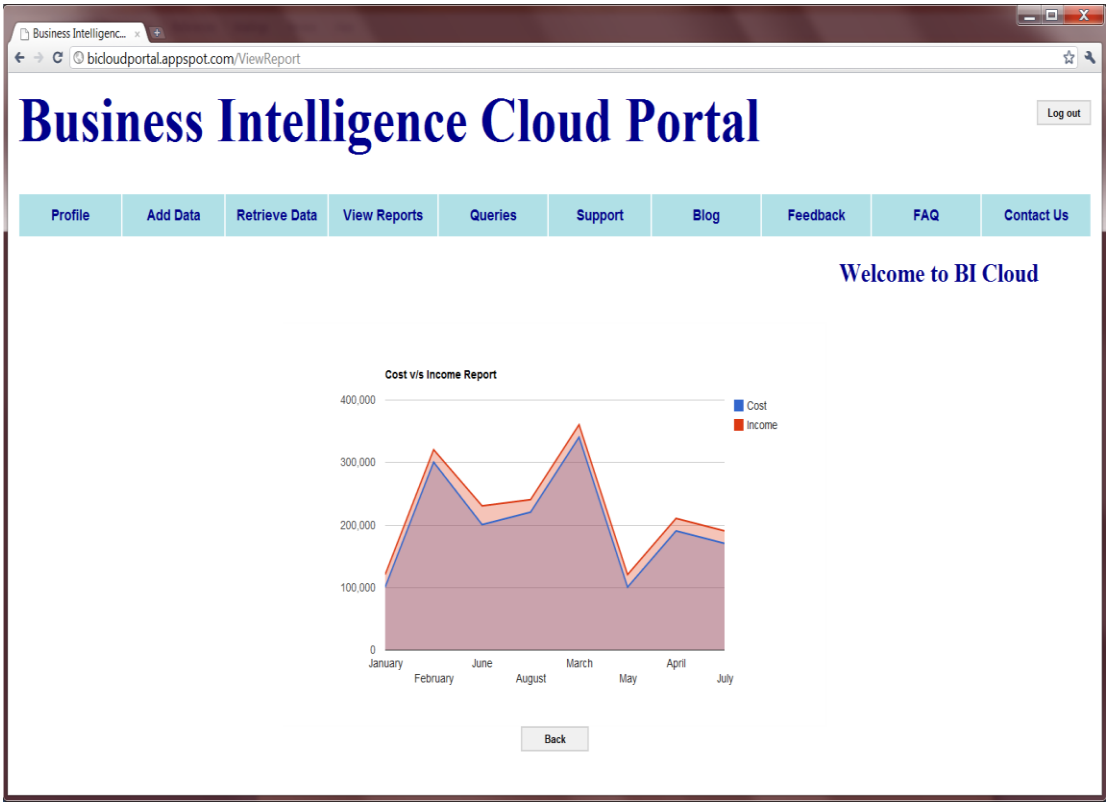


Figure 5.32 Monthly cost v/s income report shown in the form of Area Chart as selected in figure 5.31.

Figure 5.33 shows the selected option by the user to generate yearly area chart of cost v/s income. Figure 5.34 shows the area chart generated according to the user requirements.

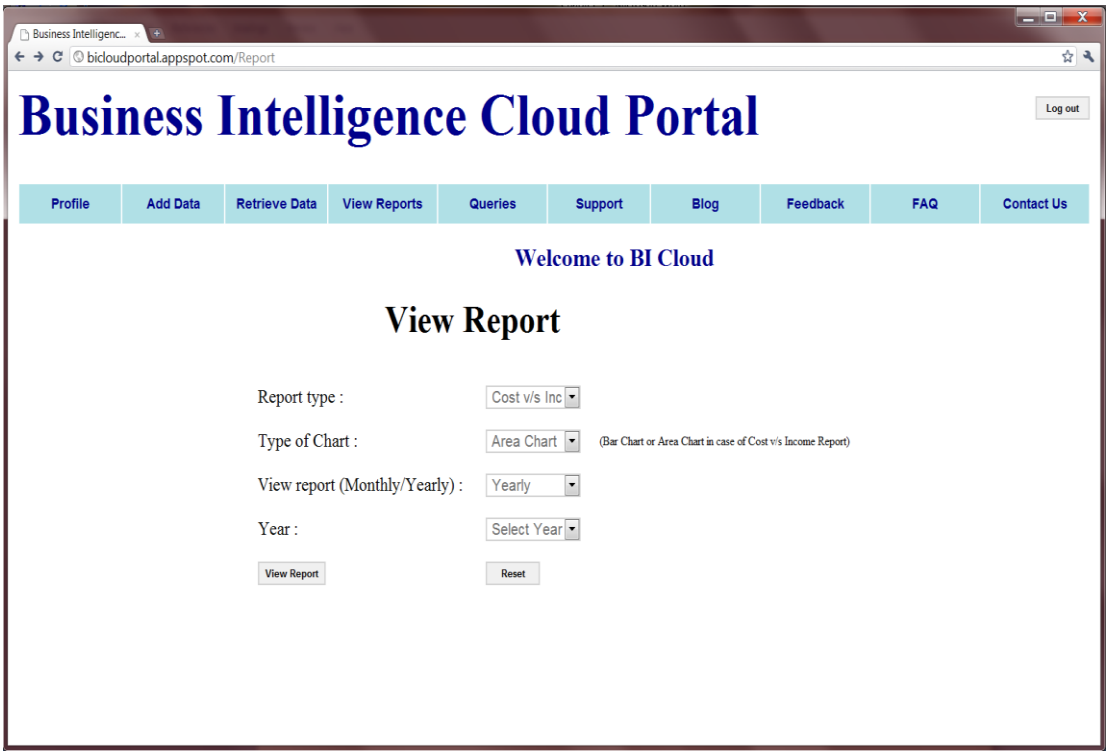


Figure 5.33 Selection of type of report and chart.

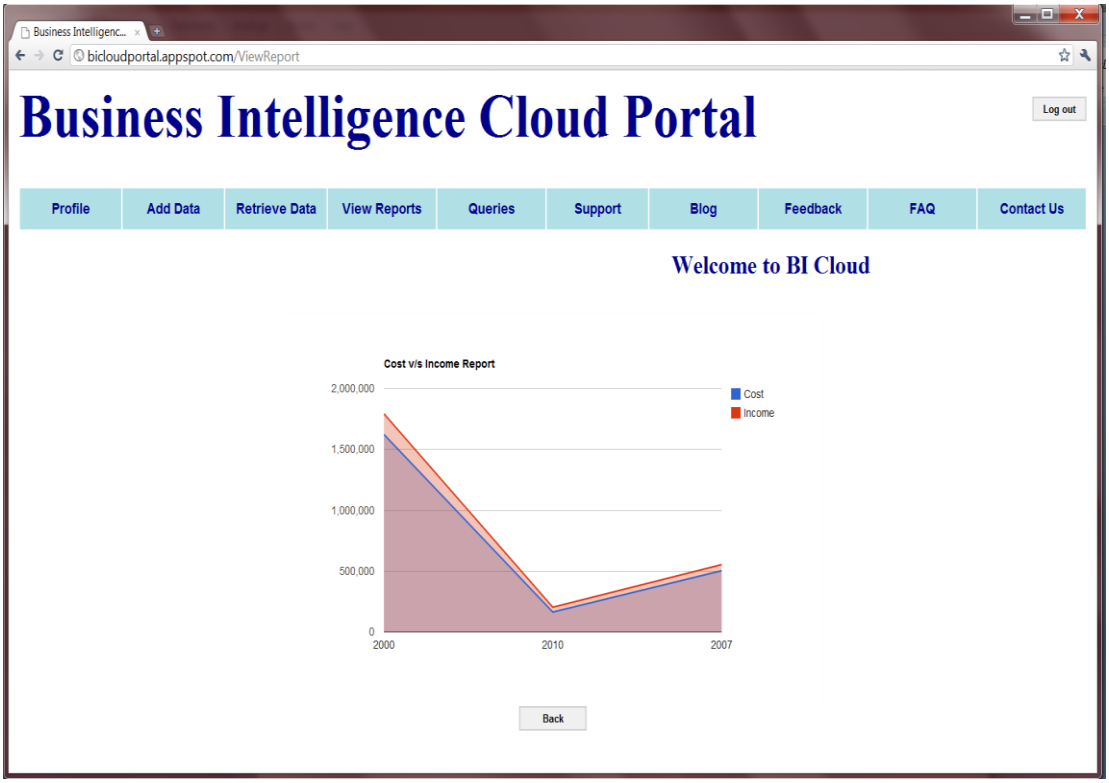


Figure 5.34 Yearly cost v/s income report shown in the form of Area Chart as selected in figure 5.33.

Figure 5.35 shows the selected option by the user to generate yearly bar chart of cost v/s income. Figure 5.36 shows the bar chart generated according to the user requirements.

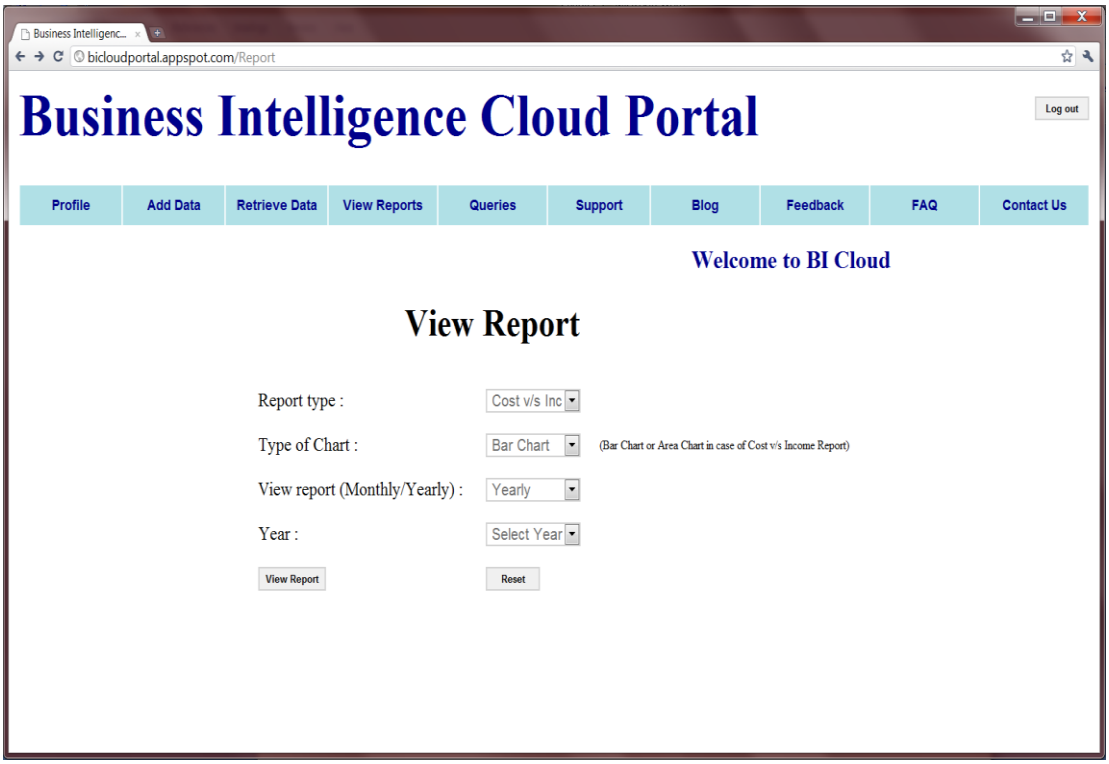


Figure 5.35 Selection of type of report and chart.

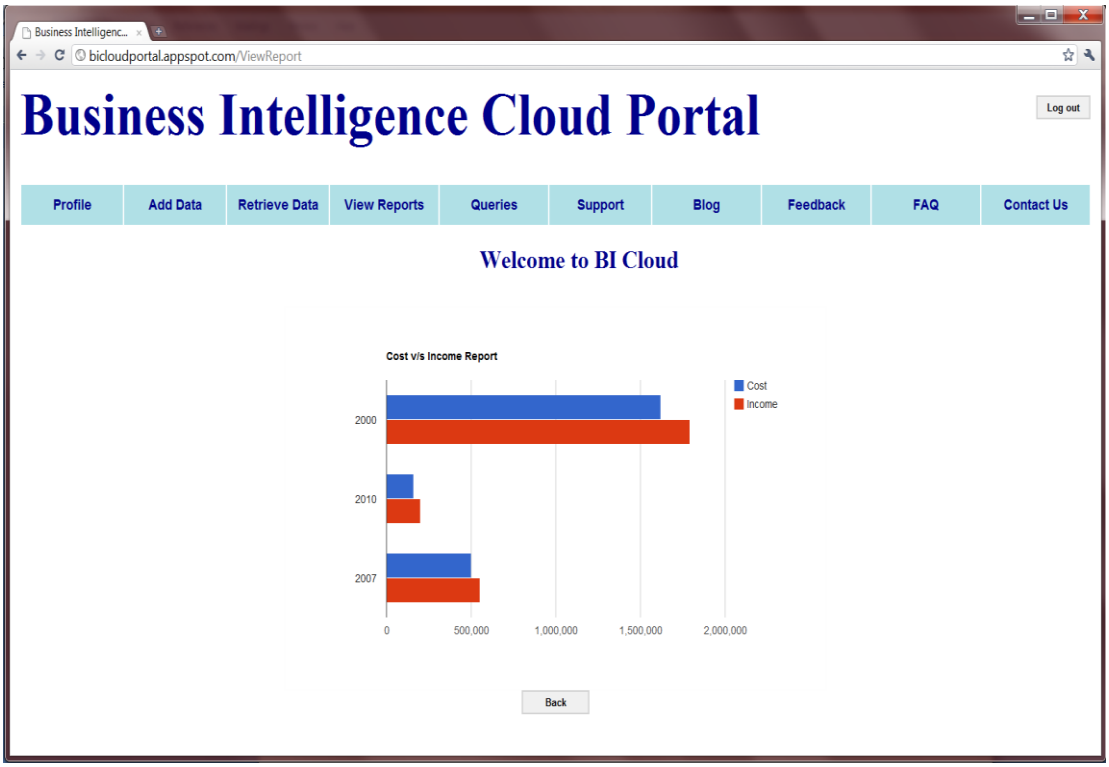


Figure 5.36 Yearly cost v/s income report shown in the form of Bar Chart as selected in figure 5.35.

Figure 5.37 shows the selected option by the user to generate yearly bar chart of sales. Figure 5.38 shows the bar chart generated according to the user requirements.

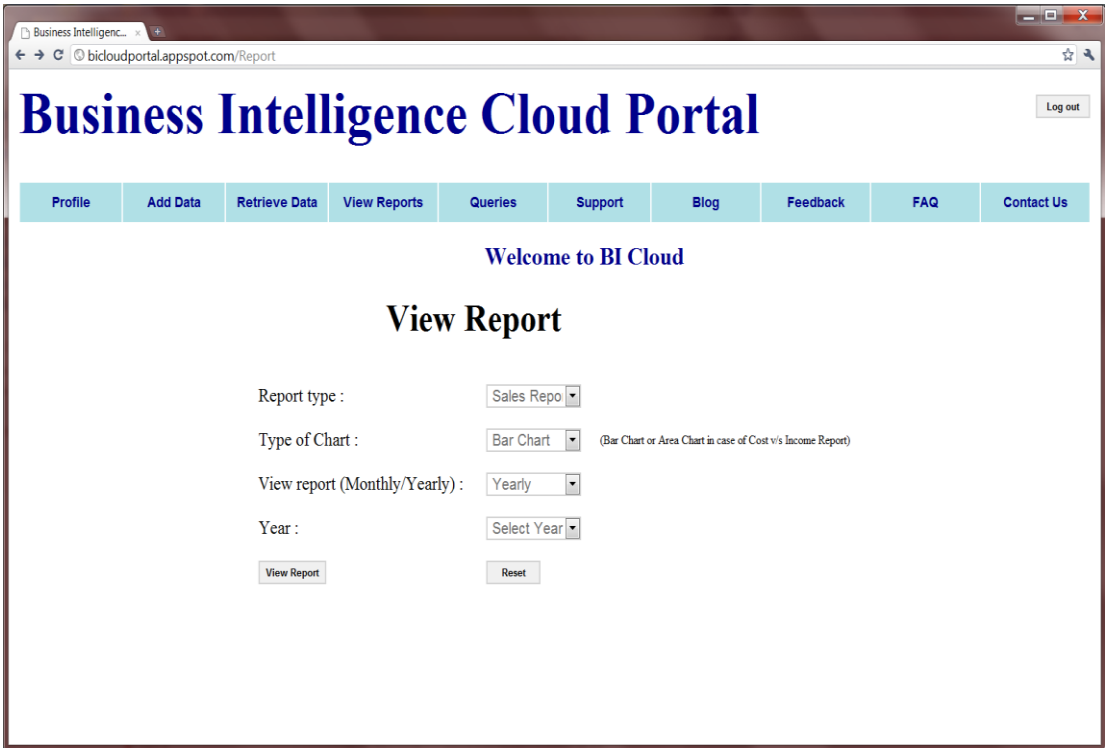


Figure 5.37 Selection of type of report and chart.

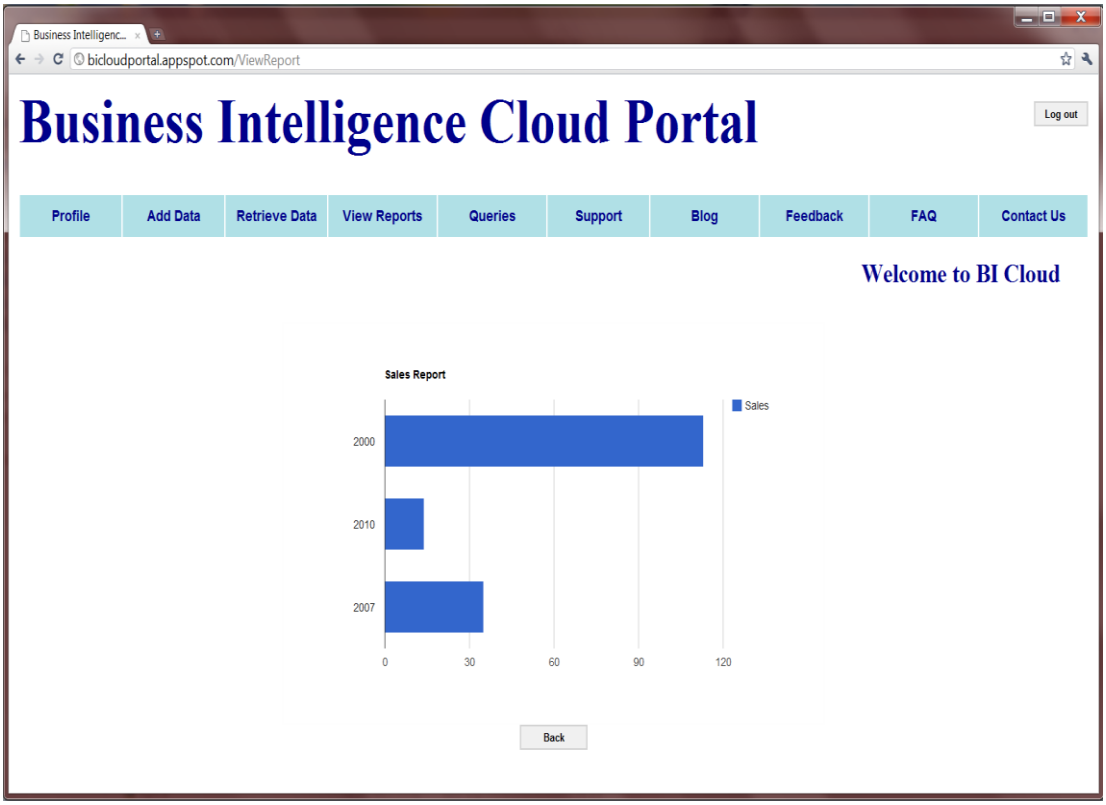


Figure 5.38 Yearly Sales report shown in the form of Bar Chart as selected in figure 5.37.

Figure 5.39 shows the selected option by the user to generate monthly pie chart of sales. Figure 5.40 shows the bar chart generated according to the user requirements.

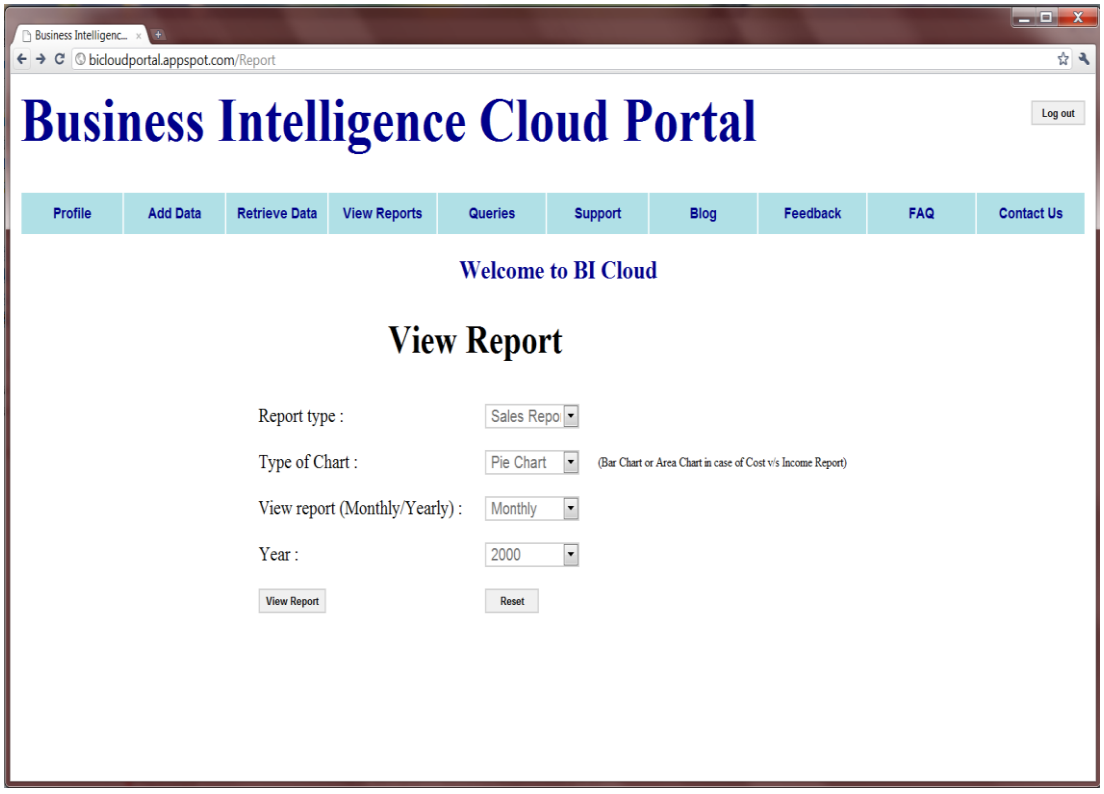


Figure 5.39 Selection of type of report and chart.

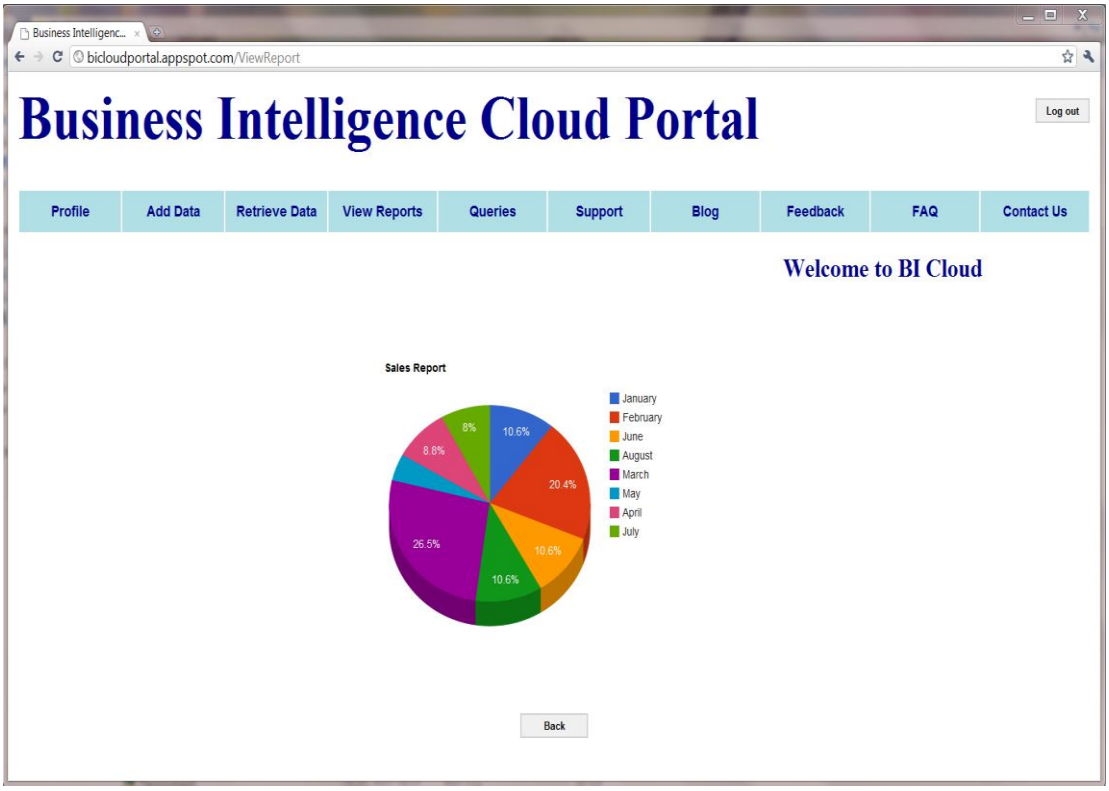


Figure 5.40 Monthly Sales report shown in the form of Pie Chart as selected in figure 5.39.

Figure 5.41 shows the selected option by the user to generate monthly bar chart of profit. Figure 5.42 shows the bar chart generated according to the user requirements.

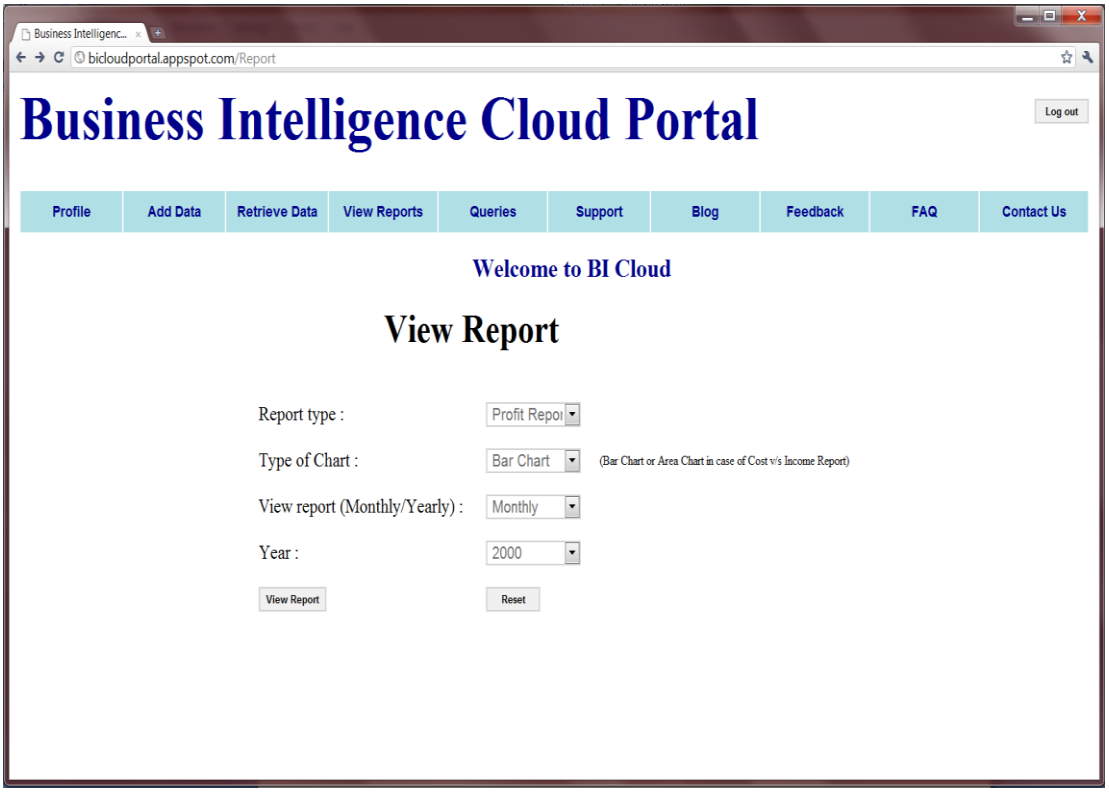


Figure 5.41 Selection of type of report and chart.

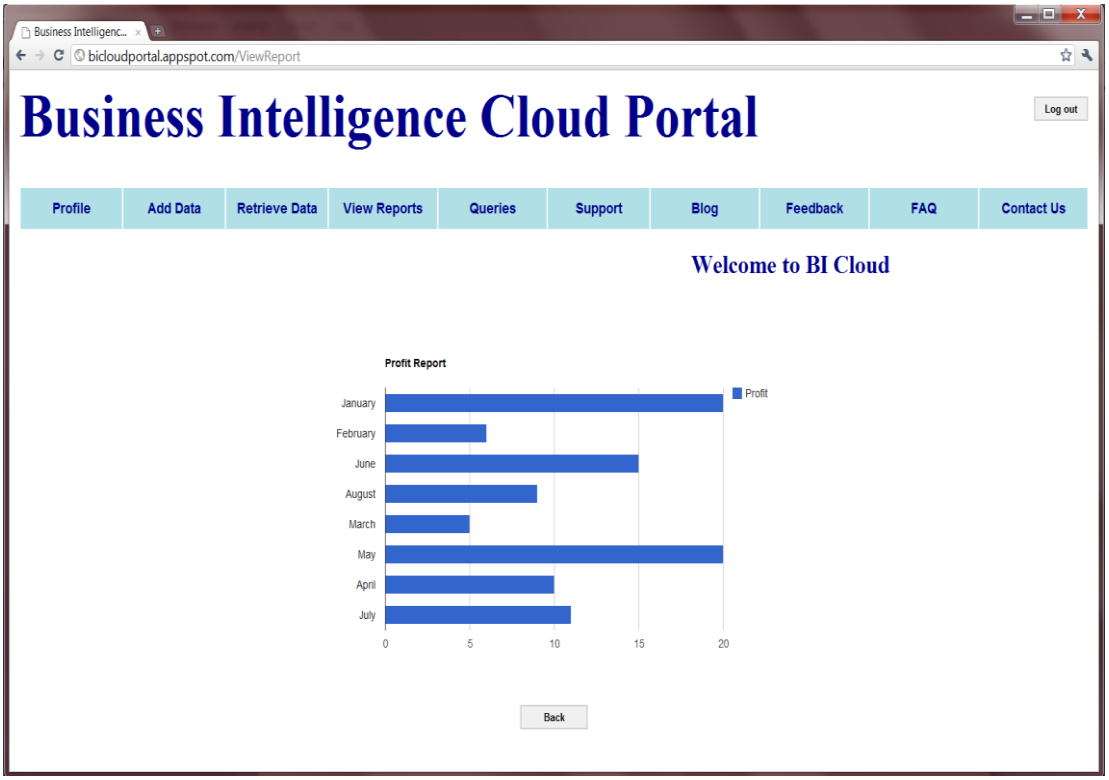


Figure 5.42 Monthly Profit report shown in the form of Bar Chart as selected in figure 5.41.

Figure 5.43 shows the selected option by the user to generate yearly bar chart of profit. Figure 5.44 shows the bar chart generated according to the user requirements.

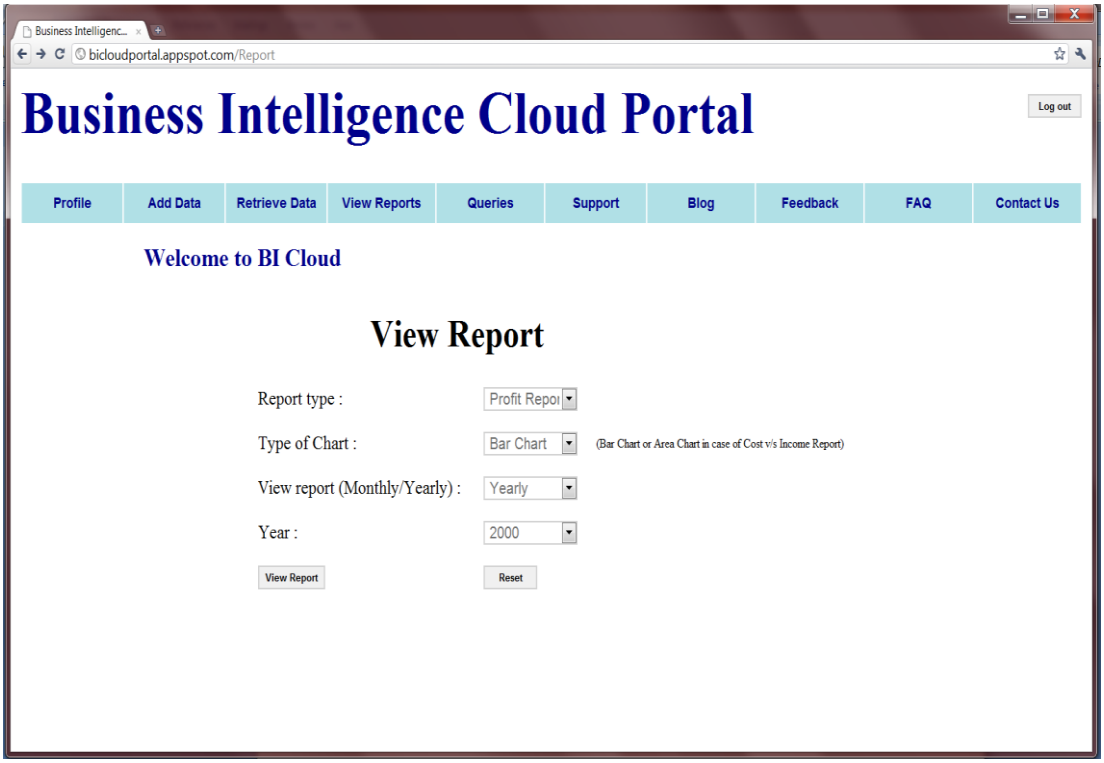


Figure 5.43 Selection of type of report and chart.

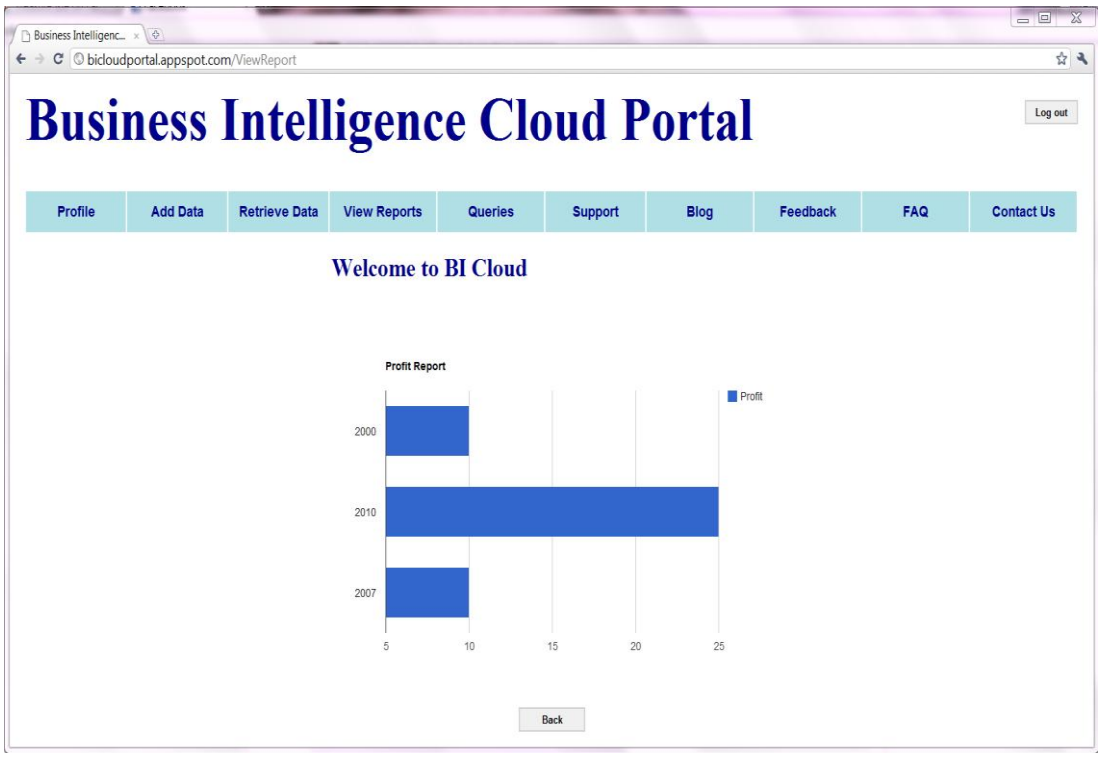


Figure 5.44 Yearly Profit report shown in the form of Bar Chart as selected in figure 5.43.

Chapter 6

Conclusions and Future Scope

This chapter discusses the conclusions of work presented in this thesis. The chapter ends with a discussion of the future direction which can be taken further.

6.1 Conclusions

This thesis gives the introduction and background of Business Intelligence and Cloud Computing. Then it discusses about BI Cloud and the comparison between available BI Clouds. Benefits and challenges of BI cloud are also discussed in this thesis. A BI cloud portal has been designed, developed and presented in this thesis. The Portal has been developed in python and HTML and has been deployed on Google App Engine which provides PaaS for developing applications on cloud.

6.2 Thesis Contributions

a) In this thesis available BI Clouds provided by various cloud providers (like IBM, Salesforce.com, RightScale, etc.) have been analyzed and compared according to their features.

b) A BI Cloud portal has been designed and the design of this portal has been presented through Data Flow Diagram and UML Diagrams.

c) The BI Cloud Portal has been implemented in python.

d) The portal has been deployed on the cloud environment provided by Google using Google App Engine.

6.3 Future Scope

a) This work shows the reporting function of business intelligence. Data integration and data mining can be included in this portal for implementing full features of business intelligence.

b) More efficient BI Cloud model can be designed and validated by developing new BI tools that provide real-time BI.

References

[1] Sourya, “A History of Cloud Computing”, Feb 2011, http://www.cloudtweaks.com/2011/02/a-history-of-cloud-computing/

[2] IPRO, “Cloud Computing - What It Is and What You Need to Know, Megan Wilcox”, May 2009, http://www.iprotech.com/Downloads/Files/Conferences/ May-2009/Cloud_Computing_What_it_Is.pdf

[3] H. P. Luhn, "A Business Intelligence System", IBM Journal of Research and Development, Vol. 2, No. 4, pp. 314-319, Oct. 1958.

[4] Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, “Cloud Computing and Grid Computing 360-Degree Compared”, Grid Computing Environments workshop, 2008, pp. 1-10, Nov. 2008.

[5] Qi Zhang, Lu Cheng, Raouf Boutaba, “Cloud computing: state-of-the-art and research challenges”, Springer, pp 7-18, Apr 2010

[6] Buyya, Rajkumar; Chee Shin Yeo, Srikumar Venugopal, “Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities”, IEEE Xplore, pp. 5-13, Sept. 2008

[7] Katarina Stanoevska-Slabeva, Thomas Wozniak, Santi Ristol, “Cloud Basics – An Introduction to Cloud Computing”, Grid and Cloud Computing - A Business Perspective on Technology and Applications, Springer, 2010, ch. 4, pp. 47 – 62.

[8] “Architectural Strategies for Cloud Computing”, An Oracle White Paper in Enterprise Architecture, August 2009

[9] Dr. Wendy A. Warr, Wendy Warr & Associates, “Cloud Computing”, November 2009, http://www.warr.com

[10] IBM, “Basics of business intelligence systems—an executive overview”, June 2003.

[11] Andreas Seufert, Josef Schiefer, “Enhanced Business Intelligence - Supporting Business Processes with Real-Time Business Analytics”, 16th International Workshop on Database and Expert Systems Applications (DEXA'05), Aug 2005.

[12] “Basics of business intelligence systems—an executive overview”, June 2003

[13] “Cloud Computing and Business Intelligence - The Road Ahead”, Cloud Computing Journal, Aug 2009, http://cloudcomputing.sys-con.com/node/1086696

[14] McMurphy, Neil, “Survey of BI Purchase Drivers shows need for new approach to business intelligence”, July, 2008

[15] “BI and Cloud Computing Are an Optimal Match”, Cloud Computing Journal, March 2010, http://cloudcomputing.sys-con.com/node/1329121

[16] IBM, “Smart Analytics Cloud for System z”.

[17] Randolph Potter and Brendon Bezuidenhout, “Matching business intelligence with cloud computing”, October 2009, http://xqrx.com/writing/a_cloud.php

[18] Golkar, Cyrus, “Top business and technology questions in cloud computing”, Jul, 2009, http://www.b-eye-network.com/channels/1550/view/10905

[19] Deshpande, mukund & joshi, shreekanth, “Incorporating BI in the cloud”, Aug 2009, http://www.b-eye-network.com/channels/1550/view/11143

[20] Eckerson, Wayne, “Implementing BI in the Cloud”, June 2009. http://portals.tdwi.org/blogs/wayneeckerson/2009/06/implementing-bi-in-the-cloud.aspx

[21] Recombinant Data Corp., “Cloud Computing for Healthcare and Life Sciences Data Warehousing”, 2009, http://www.b-eye-network.com/files/CloudComputing_WP.pdf

[22] Wells, Dave, “What’s up with cloud analytics”, December 2009

[23] Lounibos, Tom “SOASTA’s 10,000 Hours in the Cloud”, December 2009, http://eclipse.syscon.com/node/1150203

[24] Rajkumar Buyya, Rajiv Ranjan and Rodrigo N. Calheiros, “Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities”, High Performance Computing & Simulation, 2009.

[25] Junjie Peng, Xuejun Zhang, Zhou Lei, Bofeng Zhang, Wu Zhang, Qing Li, “Comparison of Several Cloud Computing Platforms”, Second International Symposium on Information Science and Engineering

[26] D. Orgizovie, B. Svilicic, E. Tijan. “Open Source Science Clouds”. MIPRO 2010, Opatija, May 24-28, 2010

[27] Alexander Zahariev, “Google App Engine”, Apr 20

[28] Eumir P. Reyes, “A Systems Thinking Approach to Business Intelligence Solutions Based on Cloud Computing”, M.S. thesis, SDM, MIT, Cambridge, Massachusetts, February 2010.

[29] Gartner, “Gartner Identifies the Top 10 Strategic Technologies for 2009”, October 2009, http://www.gartner.com/it/page.jsp?id=777212

[30] Gartner, “Gartner Identifies the Top 10 Strategic Technologies for 2010”, October 2010, http://www.gartner.com/it/page.jsp?id=1210613

[31] Eckerson, Wayne W., “The Five Dimensions of Business Intelligence”, 2005

[32] Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing v15”, National Institute of Standards and Technology, October 2009.

[33] Gartner, “Five Refining Attributes of Public and Private Cloud Computing”, May 2009.

[34] Potter, Randolph & Bezuidenhout, Brendon, “Matching Business Intelligence with Cloud Computing”, October 2009, http://xqrx.com/writing/a_cloud.php

[35] Stephen Dine, “B.I. in The Cloud”, August 2009.

[36] Tony Bradley, “Rackspace Outage Has Limited Impact”, http://www.pcworld.com/businesscenter/article/185171/rackspace_outage_has_limited_impact.html

List of Papers Accepted

Sonam Chawla, Inderveer Chana, "Business Analytics: Achieving Business Intelligence in Cloud Computing", accepted in Mini Conference on Contemporary Intelligent Systems (CCIS-2010), November 11-12, 2010.

Appendix

Installation of Google App Engine for Python

This appendix shows the installation of Google App Engine and Python 2.7.

A.1. Installation of Google App Engine

Google App Engine can be installed by downloading the .msi file of Google App Engine SDK for Python from http://code.google.com/appengine/. Steps for installing the Google App Engine are shown as figure A.1 – A.6.

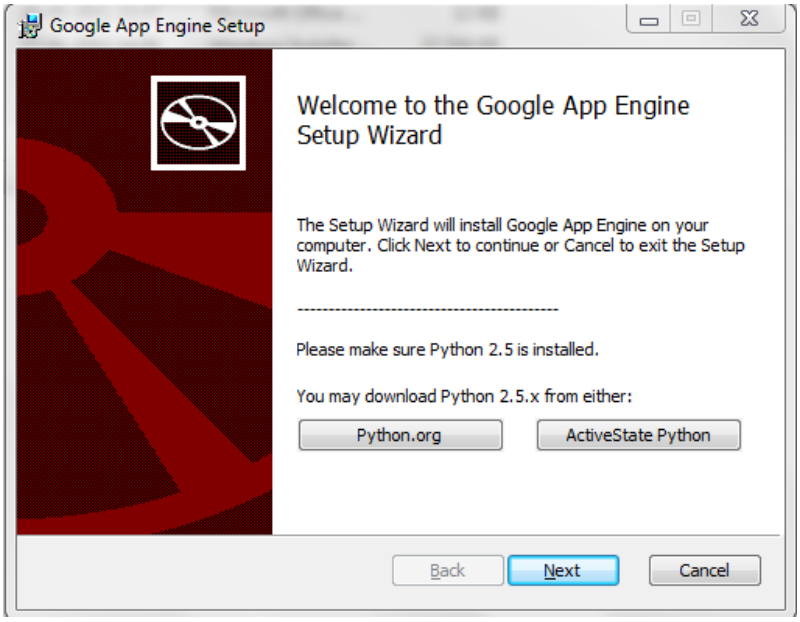


Figure A.1 Home page of the Google App Engine setup wizard.

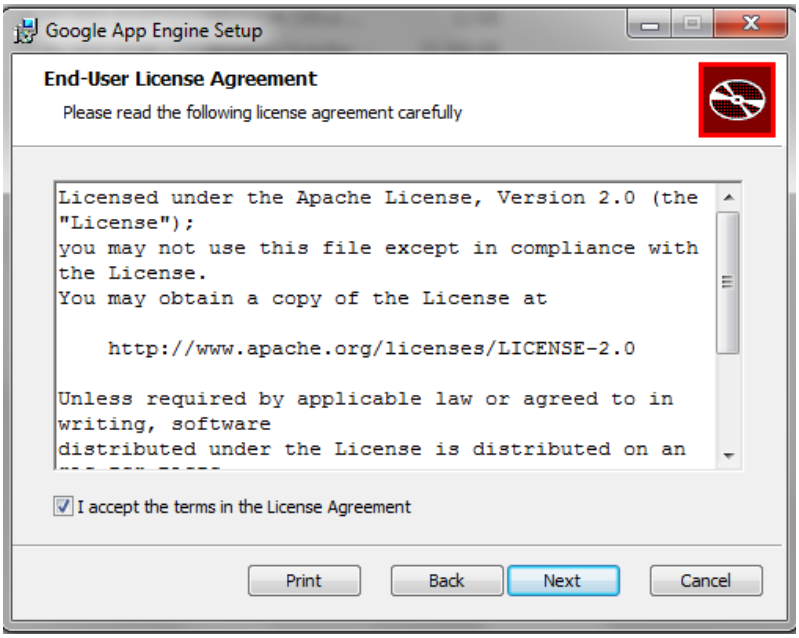


Figure A.2 End-user license agreement.

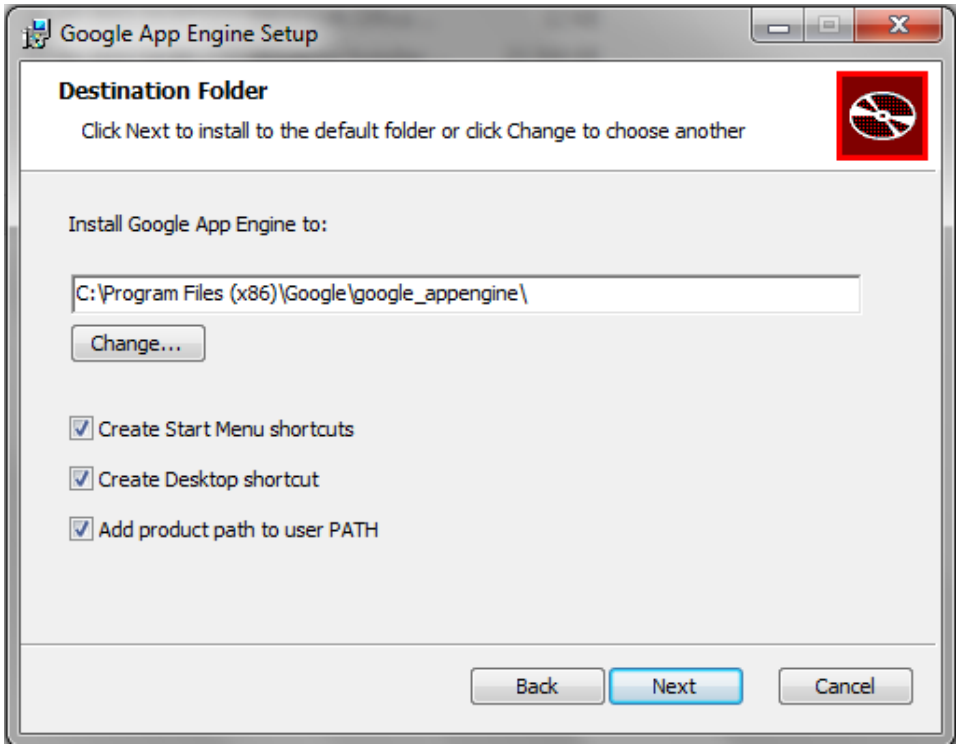


Figure A.3 Selection of destination folder.

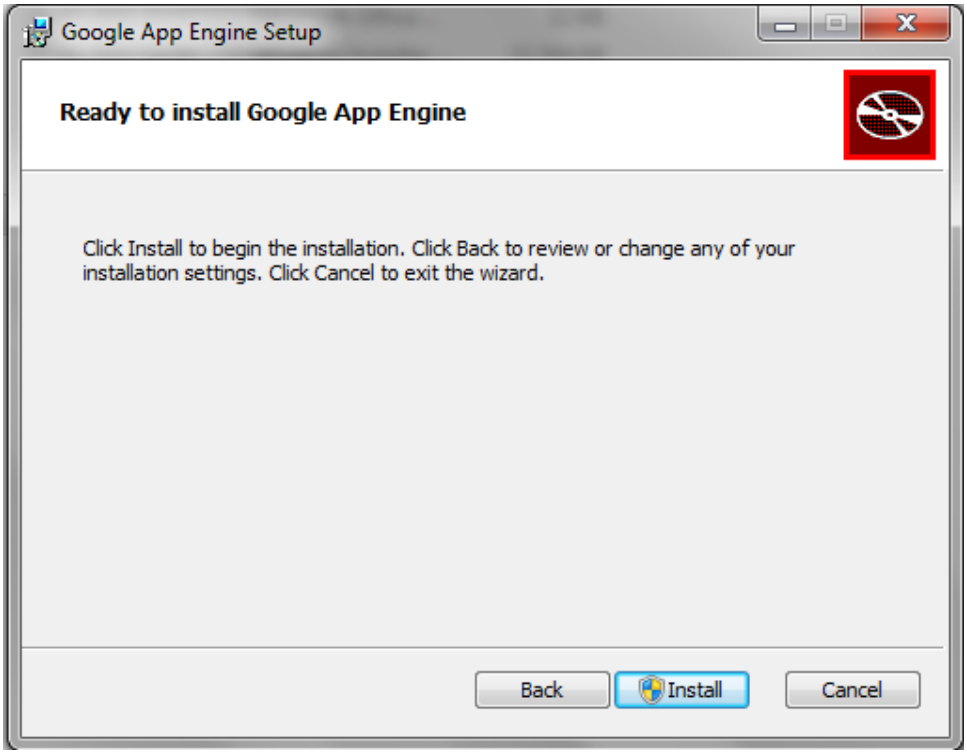


Figure A.4 Page to start the installation.

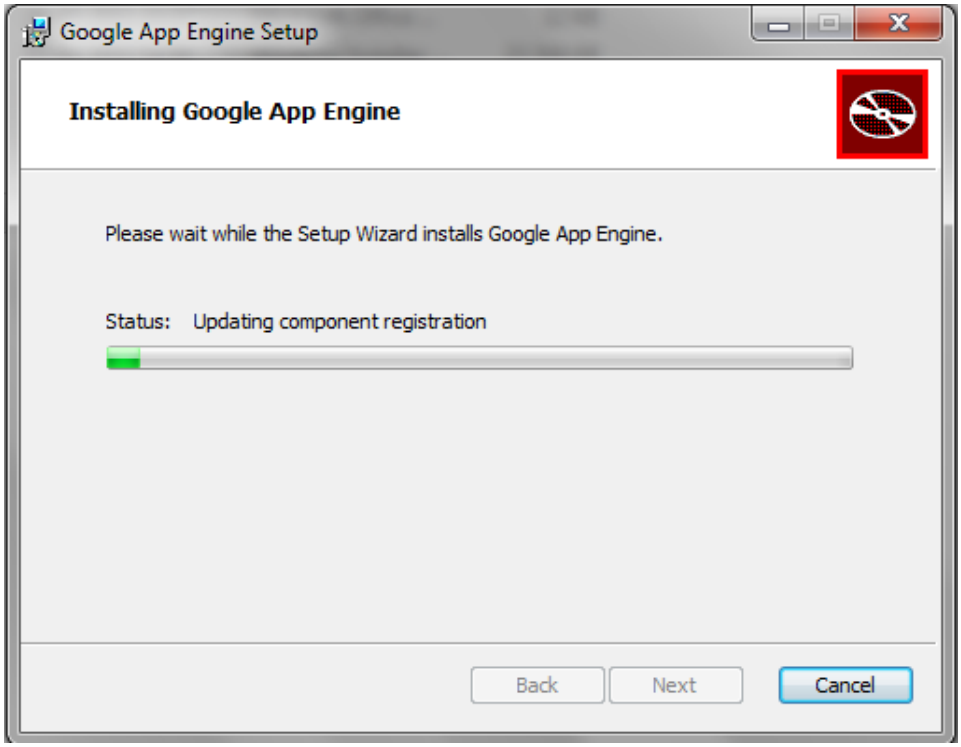


Figure A.5 Installing Google App Engine.

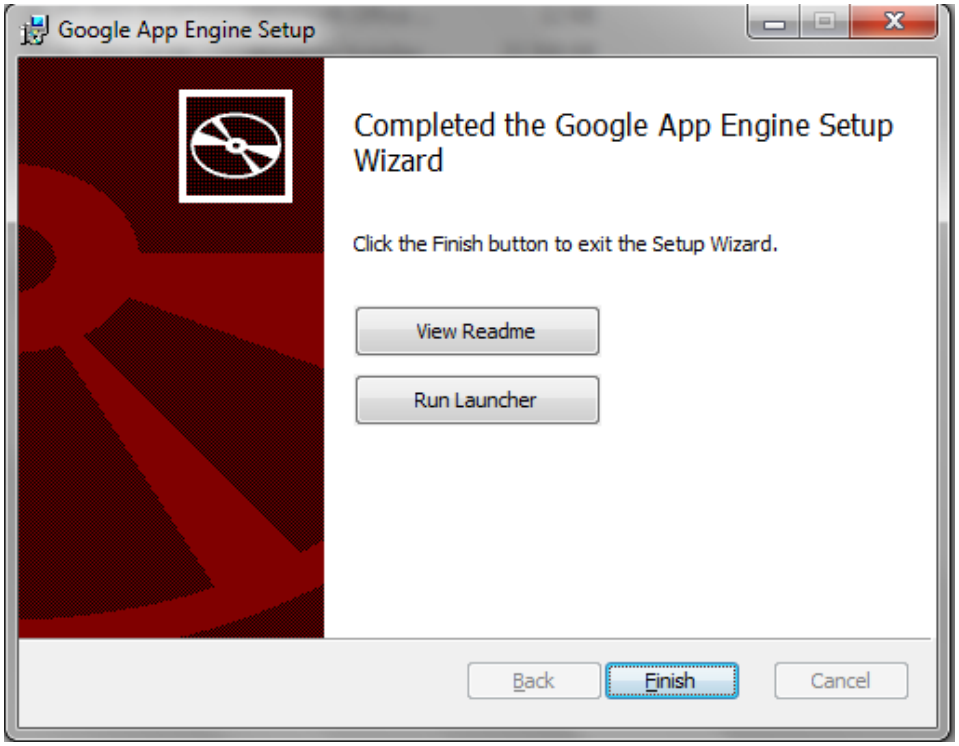


Figure A.6 Completed the Google App Engine Setup.

A.2. Installation of Python 2.7

Python 2.7.msi file can be downloaded from http://www.python.org/getit/. Installation steps are shown in figure A.7 to A.

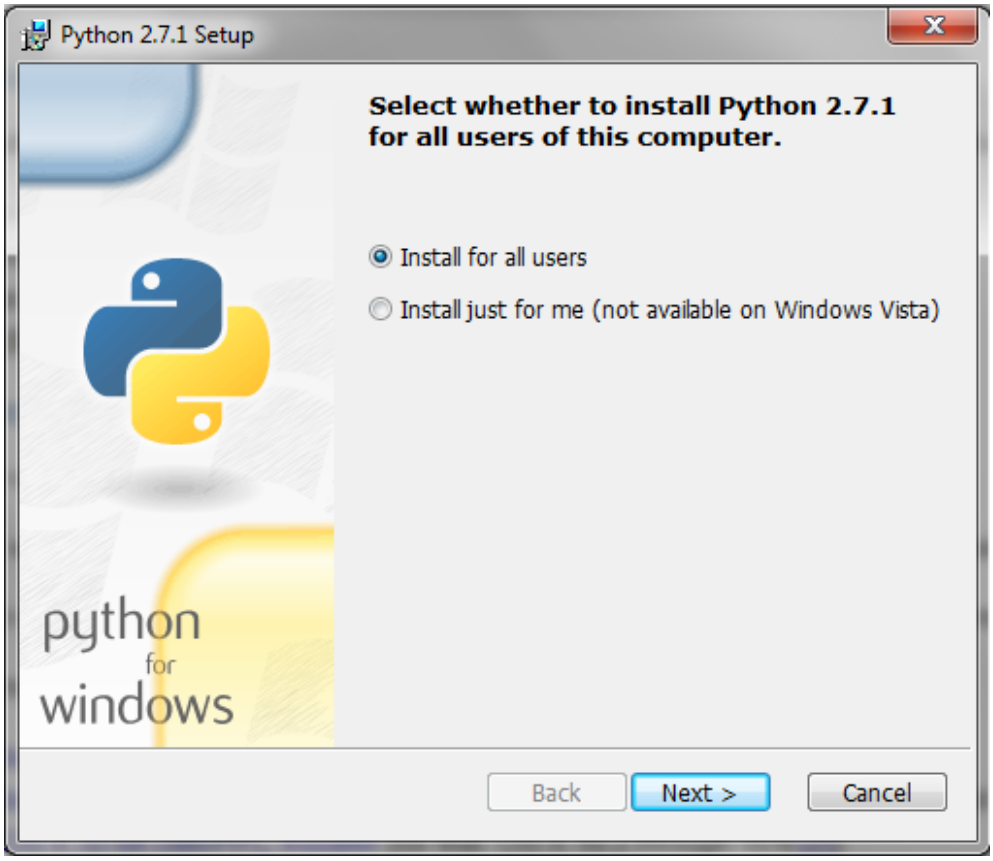


Figure A.7 Start page of installation of python 2.7

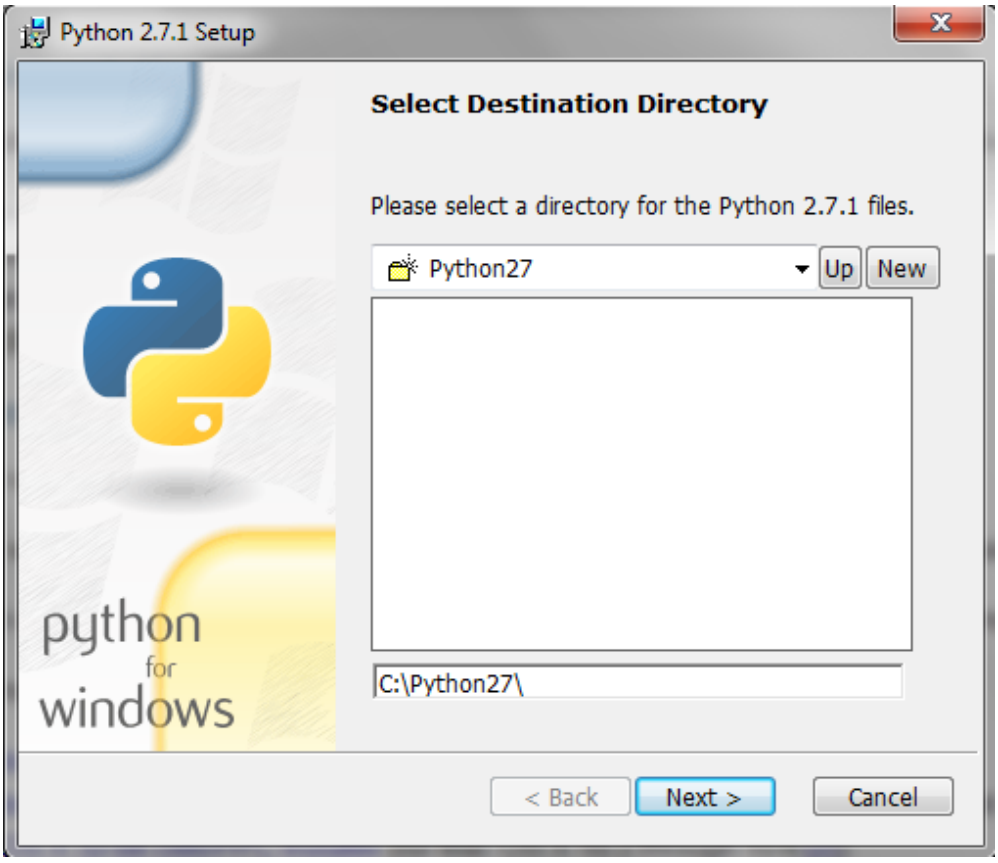


Figure A.8 Select destination directory.

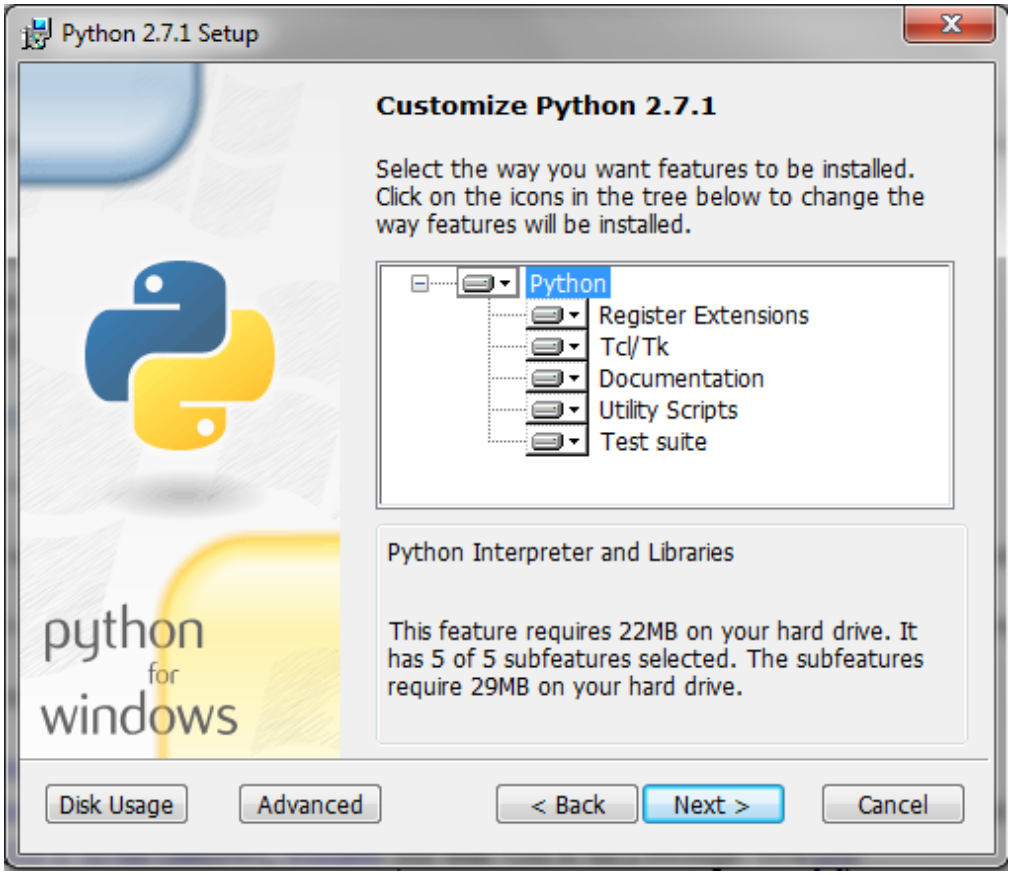


Figure A.9 Customize Python 2.7

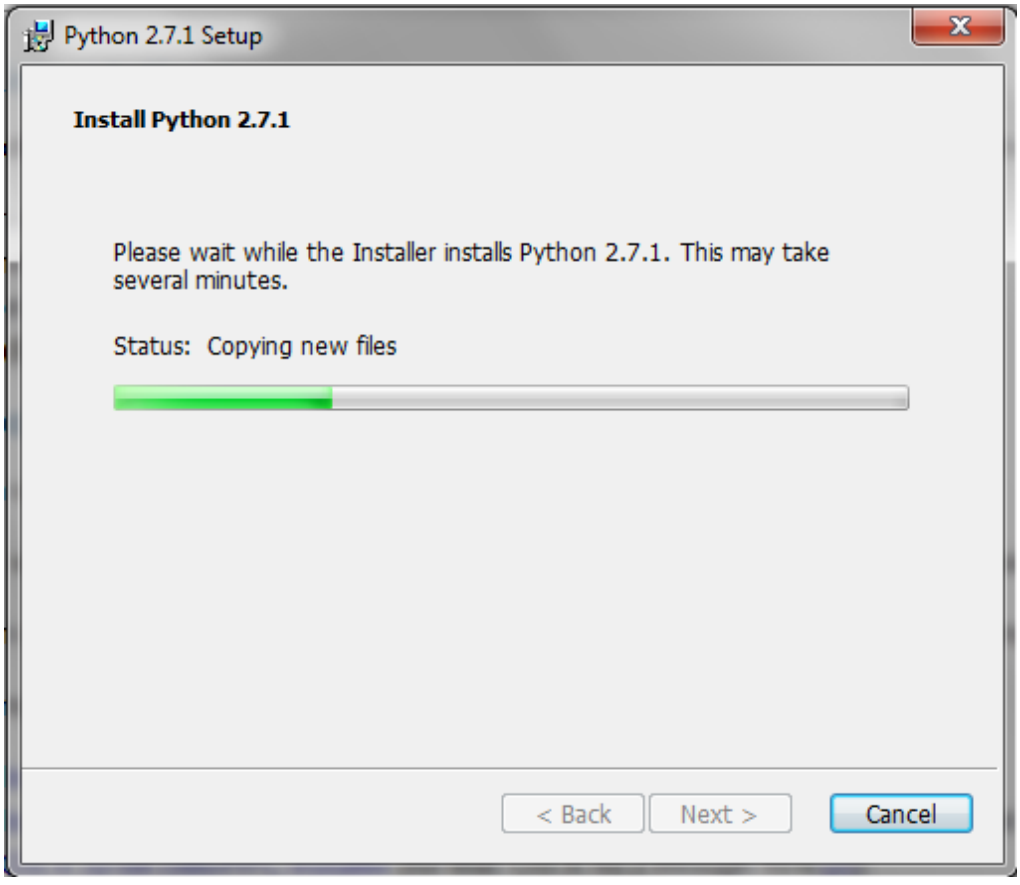


Figure A.10 Installing Python

Figure A.11 Completing Installation