Module-5: Cloud Tools and Applications

1   Aneka PaaS

2   Open stack:

2.1 Introduction to open stack

2.2 Components of open stack

3   Scientific Applications:

3.1 Healthcare: ECG Analysis in the Cloud

3.2 Biology: Protein Structure Prediction

3.3 Biology: Gene Expression Data Analysis for Cancer Diagnosis

3.4 Geo-Science: Satellite Image Processing

4   Business and Consumer Applications:

4.1 CRM and ERP

4.2 Productivity

4.3 Social networking

5   Amazon web services

**6   Google AppEngine**

7   Microsoft Azure

1. Aneka PaaS

Aneka is an Application Platform-as-a-Service (Aneka PaaS) for Cloud Computing. It acts as a framework for building customized applications and deploying them on either public or private Clouds. One of the key features of Aneka is its support for provisioning resources on different public Cloud providers such as Amazon EC2, Windows Azure and GoGrid.

- Aneka is the product of Manjarasoft.
- Aneka is used for developing, deploying, and managing cloud applications.
- Aneka can be integrated with existing cloud technologies.
- Aneka includes extensible set of APIs associated with programming models like MapReduce.
- These APIs supports different types of cloud models like private, public, hybrid cloud.

1.1 Aneka framework:
- Aneka is a software platform for developing cloud computing applications.
- In Aneka cloud applications are executed.
- Aneka is a pure PaaS solution for cloud computing.
- Aneka is a cloud middleware product.
- Aneka can be deployed on a network of computers, a multicore server, datacenters, virtual cloud infrastructures, or a mixture of these.
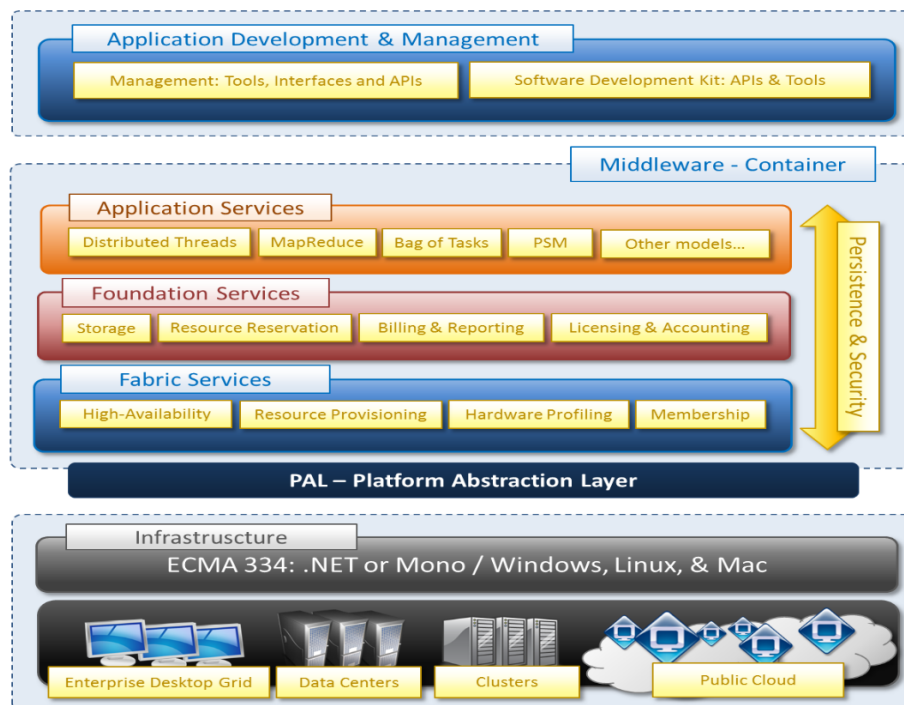
Aneka framework overview

1.2 Aneka container can be classified into three major categories:

1. Fabric Services

2. Foundation Services

3. Application Services

1. Fabric services:

Fabric Services define the lowest level of the software stack representing the Aneka Container. They provide access to the resource-provisioning subsystem and to the monitoring facilities implemented in Aneka.

2. Foundation services:

Fabric Services are fundamental services of the Aneka Cloud and define the basic infrastructure management features of the system. Foundation Services are related to the logical management of the distributed system built on top of the infrastructure and provide supporting services for the execution of distributed applications.

3. Application services:

Application Services manage the execution of applications and constitute a layer that differentiates according to the specific programming model used for developing distributed applications on top of Aneka.

2. Open Stack

2.1 Introduction to open stack

- It is a free open standard cloud computing platform that first came into existence on July 21′ 2010.
- It was a joint project of Rackspace Hosting and NASA to make cloud computing more universal in nature.
- It is deployed as Infrastructure-as-a-service (IaaS) in both public and private clouds where virtual resources are made available to the users.
- The software platform contains interrelated components that control multi-vendor hardware pools of processing, storage, networking resources through a data center.

In OpenStack, the tools which are used to build this platform are referred to as "projects". These projects handle a large number of services including computing, networking, and storage services. Unlike virtualization, in which resources such as RAM, CPU, etc are abstracted from the hardware using hypervisors, OpenStack uses a number of APIs to abstract those resources so that users and the administrators are able to directly interact with the cloud services.

2.2 Components of OpenStack

Apart from various projects which constitute the OpenStack platform, there are nine major services namely Nova, Neutron, Swift, Cinder, Keystone, Horizon, Ceilometer, and Heat. Here is the basic definition of all the components which will give us a basic idea about these components.

1. **Nova (compute service):** It manages the compute resources like creating, deleting, and handling the scheduling. It can be seen as a program dedicated to the automation of resources that are responsible for the virtualization of services and high-performance computing.

2. **Neutron (networking service):** It is responsible for connecting all the networks across OpenStack. It is an API driven service that manages all networks and IP addresses.

3. **Swift (object storage):** It is an object storage service with high fault tolerance capabilities and it used to retrieve unstructured data objects with the help of Restful API. Being a distributed platform, it is also used to provide redundant storage within servers that are clustered together. It is able to successfully manage petabytes of data.

4. **Cinder (block storage):** It is responsible for providing persistent block storage that is made accessible using an API (self- service). Consequently, it allows users to define and manage the amount of cloud storage required.

5. **Keystone (identity service provider):** It is responsible for all types of authentications and authorizations in the OpenStack services. It is a directory-based service that uses a central repository to map the correct services with the correct user.

6. **Glance (image service provider):** It is responsible for registering, storing, and retrieving virtual disk images from the complete network. These images are stored in a wide range of back-end systems.

7. **Horizon (dashboard):** It is responsible for providing a web-based interface for OpenStack services. It is used to manage, provision, and monitor cloud resources.

8. **Ceilometer (telemetry):** It is responsible for metering and billing of services used. Also, it is used to generate alarms when a certain threshold is exceeded.

9. **Heat (orchestration):** It is used for on-demand service provisioning with auto-scaling of cloud resources. It works in coordination with the ceilometer.

These are the services around which this platform revolves around. These services individually handle storage, compute, networking, identity, etc. These services are the base on which the rest of the projects rely on and are able to orchestrate services, allow bare-metal provisioning, handle dashboards, etc.

Advantages of using OpenStack

- It boosts rapid provisioning of resources due to which orchestration and scaling up and down of resources becomes easy.
- Deployment of applications using OpenStack does not consume a large amount of time.
- Since resources are scalable therefore they are used more wisely and efficiently.
- The regulatory compliances associated with its usage are manageable.

Disadvantages of using OpenStack

- OpenStack is not very robust when orchestration is considered.
- Even today, the APIs provided and supported by OpenStack are not compatible with many of the hybrid cloud providers, thus integrating solutions becomes difficult.
- Like all cloud service providers OpenStack services also come with the risk of security breaches.

3. Scientific applications

Scientific applications are a sector that is increasingly using cloud computing systems and technologies.

Cloud computing systems meet the needs of different types of applications in the scientific domain: high-performance computing (HPC) applications, high-throughput computing (HTC) applications, and data-intensive applications.

The opportunity to use cloud resources is even more appealing because minimal changes need to be made to existing applications to leverage cloud resources.

3.1 Healthcare: ECG analysis in the cloud

3.2 Biology: protein structure prediction

3.3 Biology: gene expression data analysis for cancer diagnosis

3.4 Geoscience: satellite image processing

3.1 Healthcare: ECG Analysis in the Cloud

Healthcare is a domain in which computer technology has found several and diverse applications: from supporting the business functions to assisting scientists in developing solutions to cure diseases. An illustration of the infrastructure and model for supporting remote ECG monitoring is shown in Figure 10.1. Wearable computing devices equipped with ECG sensors constantly monitor the patient's heartbeat. Such information is transmitted to the patient's mobile device, which will eventually forward it to the cloud-hosted Web service for analysis.
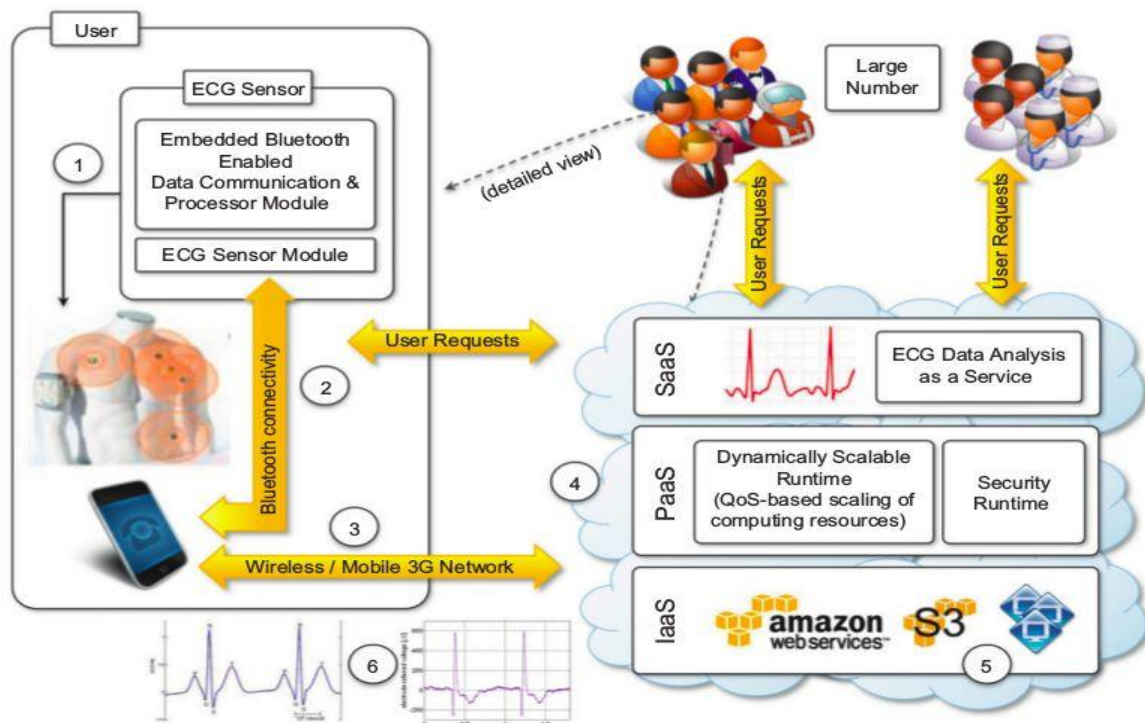
FIGURE 10.1

An online health monitoring system hosted in the cloud.

The Web service forms the front-end of a platform that is hosted in cloud and leverages three layers of cloud computing stack: SaaS, PaaS, and IaaS.

The Web service constitute SaaS application that will store ECG data in the Amazon S3 service and issue a processing request to the scalable cloud platform.

The runtime platform is composed of a dynamically sizable number of instances running the workflow engine and Aneka.

The number of workflow engine instances is controlled according to the number of requests in the queue of each instance, while Aneka controls the number of EC2 instances used to execute the single tasks defined by the workflow engine for a single ECG processing job.

Advantages

1. The elasticity of cloud infrastructure that can grow and shrink according to the requests served. As a result, doctors and hospitals do not have to invest in large computing infrastructures designed after capacity planning, thus making more effective use of budgets.

2. Ubiquity. Cloud computing technologies are easily accessible and promise to deliver systems

with minimum or no downtime. Computing systems hosted in cloud are accessible from any Internet device through simple interfaces (such as SOAP and REST-based Web services). This makes systems easily integrated with other systems maintained on hospital's premises.

3. Cost savings. Cloud services are priced on a pay-per-use basis and with volume prices for large numbers of service requests.

3.2 Biology: Protein Structure Prediction

Applications in biology require high computing capabilities and operate on large data-sets that cause extensive I/O operations.

Therefore, biology applications have made use of supercomputing and cluster computing infrastructures.

Similar capabilities can be leveraged using cloud computing technologies in a more dynamic fashion, thus opening new opportunities for bioinformatics applications.

Protein structure prediction is a computationally intensive task that is fundamental to different types of research in the life sciences.

The geometric structure of a protein cannot be directly inferred from the sequence of genes that compose its structure, but it is the result of complex computations aimed at identifying the structure that minimizes the required energy.

This task requires the investigation of a space with a massive number of states, consequently creating many computations for each of these states.

One project that investigates the use of cloud technologies for protein structure prediction is Jeeva - an integrated Web portal that enables scientists to offload the prediction task to a computing cloud based on Aneka (Figure 10.2).
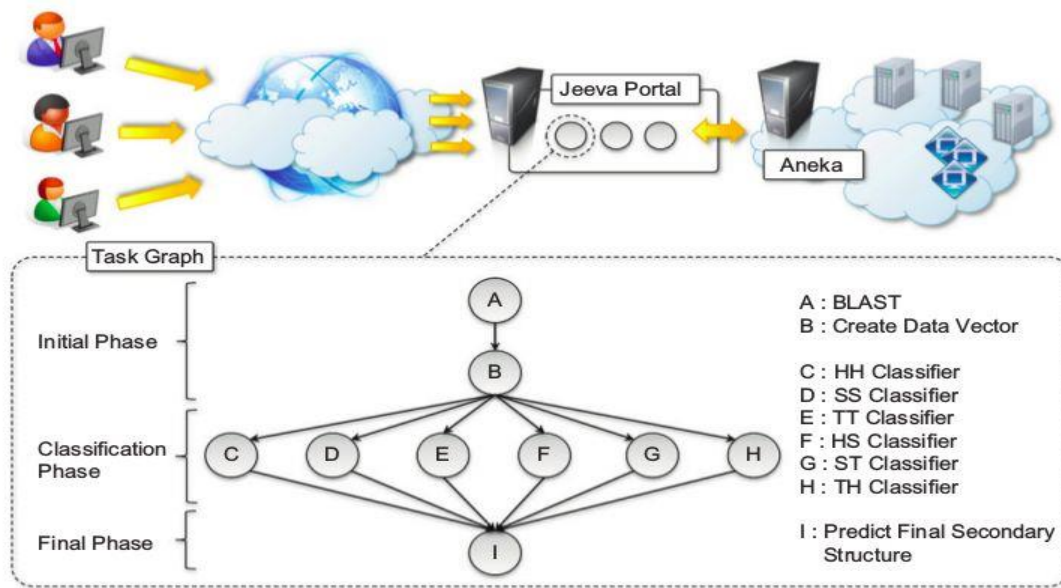
FIGURE 10.2

Architecture and overview of the Jeeva Portal.

The prediction task uses machine learning techniques (support vector machines) for determining the secondary structure of proteins. These techniques translate problem into one of pattern recognition, where a sequence has to be classified into one of three possible classes (E, H, and C). A popular implementation based on support vector machines divides the pattern recognition problem into three phases: initialization, classification, and a final phase. These three phases have to be executed in sequence, we can perform parallel execution in the classification phase, where multiple classifiers are executed concurrently. This reduces computational time of the prediction. The prediction algorithm is then translated into a task graph that is submitted to Aneka. Once the task is completed, the middleware makes the results available for visualization through the portal.

The **advantage** of using cloud technologies is the capability to leverage a scalable computing infrastructure that can be grown and shrunk on demand.

3.3 Biology: Gene Expression Data Analysis for Cancer Diagnosis

Gene expression profiling is the measurement of the expression levels of thousands of genes at once. It is used to understand the biological processes that are triggered by medical treatment at a cellular level. Important application of gene expression profiling is cancer diagnosis and treatment. Cancer is a disease characterized by uncontrolled cell growth and proliferation. This behavior occurs because genes regulating the cell growth mutate.

Gene expression profiling is utilized to provide a more accurate classification of tumors.

The dimensionality of typical gene expression datasets ranges from several thousands to over tens of thousands of genes.

This problem is approached with learning classifiers, which generate a population of condition-action rules that guide the classification process. The eXtended Classifier System (XCS) has been utilized for classifying large datasets in bioinformatics and computer science domains.

A variation of this algorithm, CoXCS [162], has proven to be effective in these conditions. CoXCS divides the entire search space into subdomains and employs the standard XCS algorithm in each of these subdomains. Such a process is computationally intensive but can be easily parallelized because the classifications problems on the subdomains can be solved concurrently.

Cloud-CoXCS (Figure 10.3) is a cloud-based implementation of CoXCS that leverages Aneka to solve the classification problems in parallel and compose their outcomes. The algorithm is controlled by strategies, which define the way the outcomes are composed together and whether the process needs to be iterated.
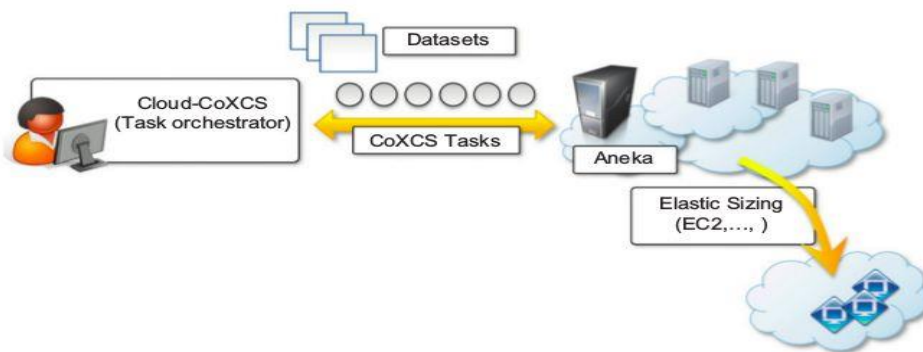
FIGURE 10.3
Cloud-CoXCS: An environment for microarray data processing on the cloud.

3.4 Geoscience: satellite image processing

Geoscience applications collect, produce, and analyze massive amounts of geospatial and nonspatial data. As the technology progresses and our planet becomes more instrumented, volume of data that needs to be processed increases significantly.

Geographic information system (GIS) applications capture, store, manipulate, analyze, manage, and present all types of geographically referenced data.

Cloud computing is an attractive option for executing these demanding tasks and extracting meaningful information to support decision makers.

Satellite remote sensing generates hundreds of gigabytes of raw images that need to be further processed to become the basis of several different GIS products. This process requires both I/O and compute-intensive tasks. Large images need to be moved from a ground station's local storage to compute facilities, where several transformations and corrections are applied.

The system shown in Figure 10.4 integrates several technologies across the entire computing stack.

A SaaS application provides a collection of services for such tasks as geocode generation and data visualization.

At the PaaS level, Aneka controls the importing of data into the virtualized infrastructure and the execution of image-processing tasks that produce the desired outcome from raw satellite images.

The platform leverages a Xen private cloud and the Aneka technology to dynamically provision the required resources.
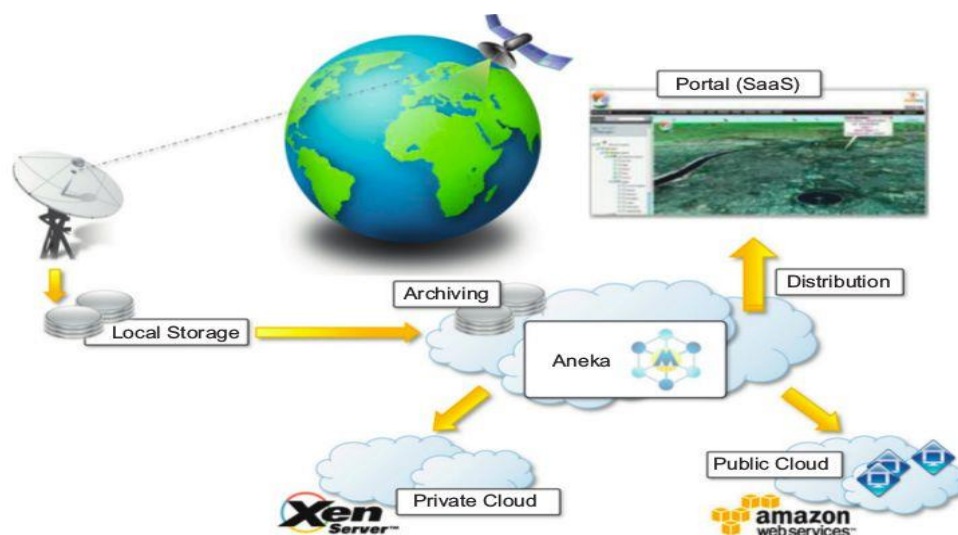
FIGURE 10.4

A cloud environment for satellite data processing.

4. Business and Consumer Applications

The business and consumer sector is the one that benefits the most from cloud computing technologies. On one hand, the opportunity to transform capital costs into operational costs makes clouds an attractive option for all enterprises that are IT-centric. On the other hand, the sense of ubiquity that the cloud offers for accessing data and services makes it interesting for end users. The combination of all these elements has made cloud computing the preferred technology for a wide range of applications.

4.1 CRM and ERP

4.2 Productivity

4.3 Social networking

4.1 CRM and ERP

Customer relationship management (CRM) and enterprise resource planning (ERP) applications are market segments that are flourishing in the cloud.

Cloud CRM applications constitute a great opportunity for small enterprises and start-ups to have fully functional CRM software without large up-front costs and by paying subscriptions. Your business and customer data from everywhere and from any device, has fostered the spread of cloud CRM applications.

ERP solutions on the cloud are less mature and have to compete with well-established in-house solutions. ERP systems integrate several aspects of an enterprise: finance and accounting, human resources, manufacturing, supply chain management, project management, and CRM.

1 Salesforce.com

Salesforce.com is most popular and developed CRM solution available today. As of today more than 100,000 customers have chosen Safesforce.com to implement their CRM solutions. The application provides customizable CRM solutions that can be integrated with additional features developed by third parties.

Salesforce.com is based on the Force.com cloud development platform.

This represents scalable and high-performance middleware executing all operations of all Salesforce.com applications.

The architecture of the Force.com platform is shown in Figure 10.5. At the core of the platform resides its metadata architecture, which provides the system with flexibility and scalability.

Application core logic and business rules are saved as metadata into the Force.com store.

Both application structure and application data are stored in the store. A runtime engine executes application logic by retrieving its metadata and then performing the operations on the data.

A full-text search engine supports the runtime engine. This allows application users to have an effective user experience The search engine maintains its indexing data in a separate store.
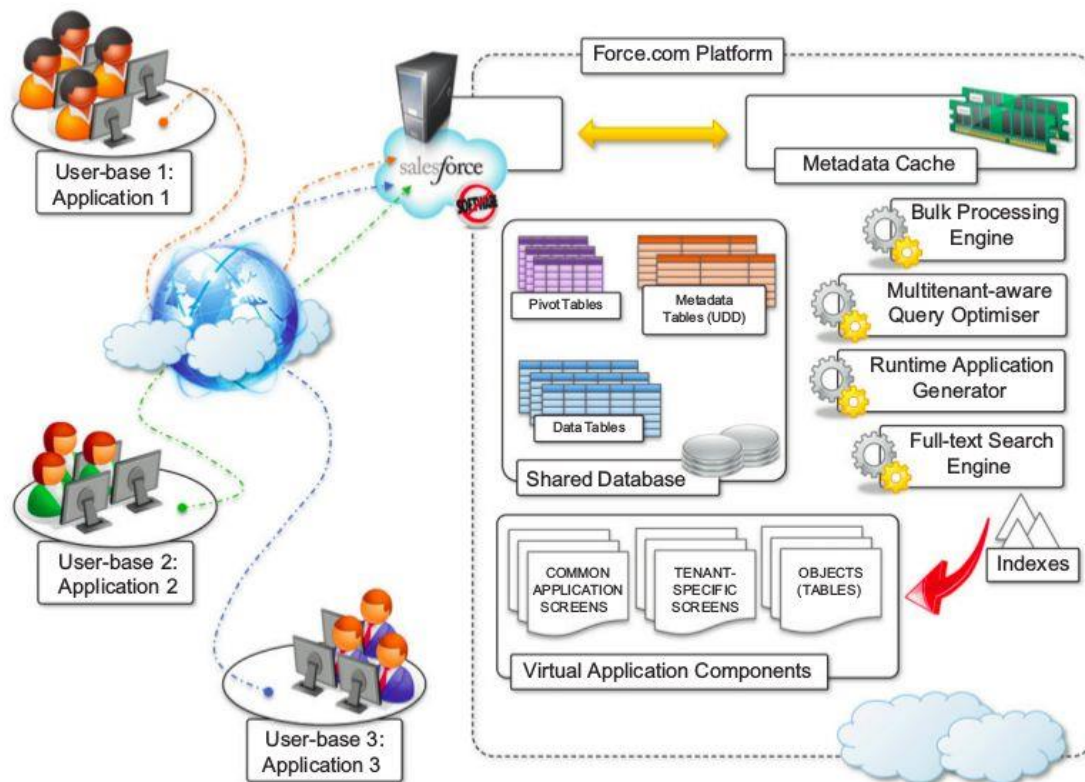
FIGURE 10.5

Salesforce.com and Force.com architecture.

2 Microsoft dynamics CRM

The system is completely hosted in Microsoft's datacenters across the world and offers to customers a 99.9% SLA.

Each CRM instance is deployed on a separate database, and application provides users with facilities for marketing, sales, and advanced customer relationship management.

Dynamics CRM Online features can be accessed either through a Web browser interface or by means of SOAP and RESTful Web services.

This allows Dynamics CRM to be easily integrated with both other Microsoft products and line-of-business applications.

Dynamics CRM can be extended by developing plug-ins that allow implementing specific behaviors triggered on the occurrence of given events.

3 NetSuite

NetSuite provides a collection of applications that help customers manage every aspect of the business enterprise.

Its offering is divided into three major products: NetSuite Global ERP, NetSuite Global CRM1 , and NetSuite Global Ecommerce.

Moreover, an all-in-one solution: NetSuite One World, integrates all three products together.

The services NetSuite delivers are powered by two large datacenters on the East and West coasts of the United States, connected by redundant links.

This allows NetSuite to guarantee 99.5% uptime to its customers.

The NetSuite Business Operating System (NS-BOS) is a complete stack of technologies for building SaaS business applications that leverage the capabilities of NetSuite products.

On top of the SaaS infrastructure, the NetSuite Business Suite components offer accounting, ERP, CRM, and ecommerce capabilities.

4.2 Productivity

Productivity applications replicate in cloud. The most common tasks that we are used to performing on our desktop: from document storage to office automation and complete desktop environments hosted in the cloud.

1 Dropbox and iCloud

Online storage solutions have turned into SaaS applications and become more usable as well as more advanced and accessible.

The most popular solution for online document storage is **Dropbox**, that allows users to synchronize any file across any platform and any device in a seamless manner (Figure 10.6). Dropbox provides users with a free storage that is accessible through the abstraction of a folder. Users can either access their Dropbox folder through a browser or by downloading and installing a Dropbox client, which provides access to the online storage by means of a special folder. All the modifications into this folder are silently synched so that changes are notified to all the local instances of the Dropbox folder across all the devices.

Another interesting application in this area is **iCloud**, a cloud-based document-sharing application provided by Apple to synchronize iOS-based devices in a completely transparent manner. Documents, photos, and videos are automatically synched as changes are made, without any explicit operation. This allows the system to efficiently automate common operations without any human intervention. This capability is limited to iOS devices, and currently there are no plans to provide iCloud with a Web-based interface that would make user content accessible from even unsupported platforms.
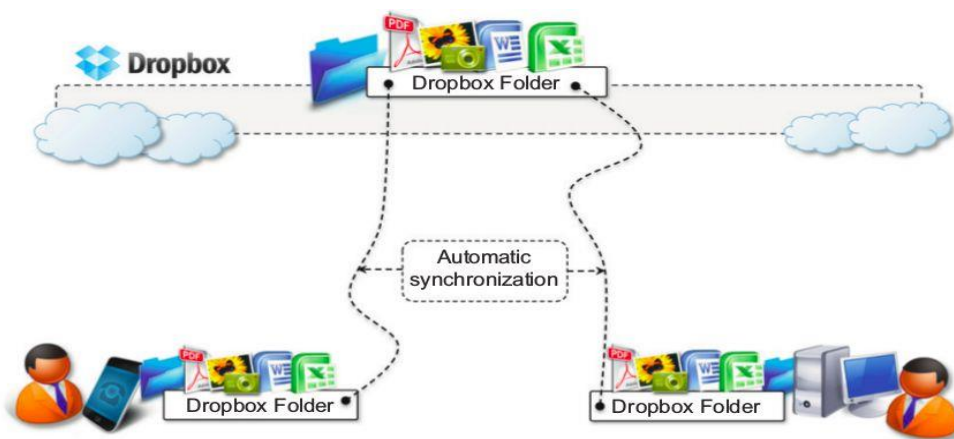
FIGURE 10.6

Dropbox usage scenario.

2 Google docs

Google Docs is a SaaS application that delivers the basic office automation capabilities with support for collaborative editing over the Web.

Google Docs allows users to create and edit text documents, spreadsheets, presentations, forms, and drawings. It aims to replace desktop products such as Microsoft Office and OpenOffice and provide similar interface and functionality as a cloud service.

By being stored in the Google infrastructure, these documents are always available from anywhere and from any device that is connected to the Internet.

Google Docs is a good example of what cloud computing can deliver to end users: ubiquitous access to resources, elasticity, absence of installation and maintenance costs, and delivery of core functionalities as a service.

3 Cloud desktops: EyeOS and XIOS/3

EyeOS is one of the most popular Web desktop solutions based on cloud technologies. It replicates the functionalities of a classic desktop environment and comes with pre-installed applications for the most common file and document management tasks (Figure 10.7). Single users can access the EyeOS desktop environment from anywhere and through any Internet-connected device, whereas organizations can create a private EyeOS Cloud on their premises to virtualize the desktop environment of their employees and centralize their management.
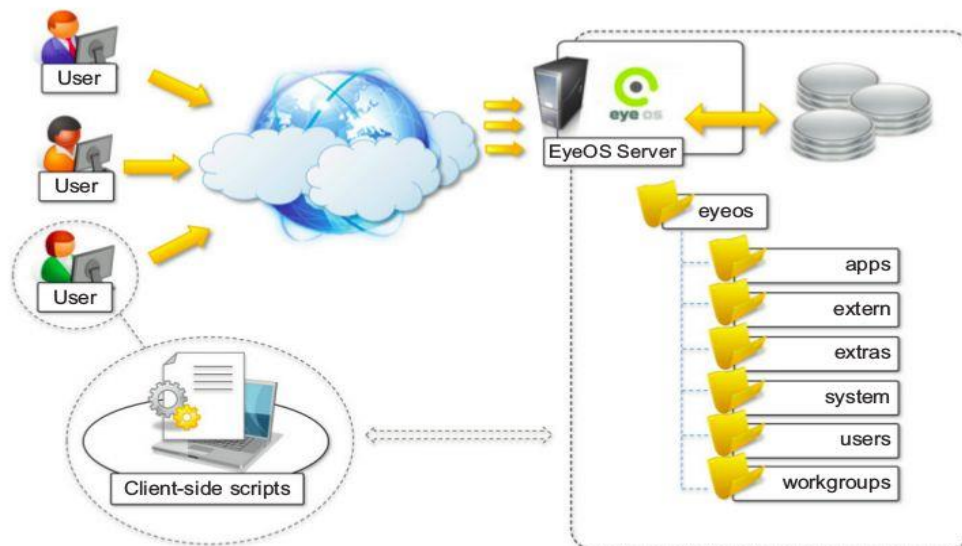
FIGURE 10.7

EyeOS architecture.

The EyeOS architecture is quite simple: On the server side, the EyeOS application maintains the information about user profiles and their data, and the client side constitutes the access point for users and administrators to interact with the system. EyeOS stores the data about users and applications on the server file system. Once the user has logged in by providing credentials, the desktop environment is rendered in the client's browser by downloading all the JavaScript libraries required to build the user interface and implement the core functionalities of EyeOS. EyeOS also provides APIs for developing new applications and integrating new capabilities into the system. EyeOS applications are server-side components that are defined by at least two files (stored in the eyeos/apps/appname directory): appname.php and appname.js. The first file defines and implements all the operations that the application exposes; the JavaScript file contains the code that needs to be loaded in the browser in order to provide user interaction with the application.

Xcerion XML Internet OS/3 (XIOS/3) is another example of a Web desktop environment. The service is delivered as part of the CloudMe application, which is a solution for cloud document storage. The key differentiator of XIOS/3 is its strong leverage of XML, used to implement many of the tasks of the OS: rendering user interfaces, defining application business logics, structuring file system organization, and even application development. XIOS/3 is released as open-source software and implements a marketplace where third parties can easily deploy applications that can be installed on top of the virtual desktop environment. It is possible to develop any type of application and feed it with data accessible through XML Web services: developers have to define the user interface, bind UI components to service calls and operations, and provide the logic on how to process the data.

4.3 Social networking

Social networking applications have grown considerably in the last few years to become the most active sites on the Web. To sustain their traffic and serve millions of users seamlessly, services such as Twitter and Facebook have leveraged cloud computing technologies.

1 Facebook

Facebook is probably the most evident and interesting environment in social networking. With more than 800 million users, it has become one of the largest Websites in the world.

To sustain this incredible growth, it has been fundamental that Facebook be capable of continuously adding capacity and developing new scalable technologies and software systems while maintaining high performance to ensure a smooth user experience.

Currently, the social network is backed by two data centers that have been built and optimized to reduce costs and impact on the environment.

On top of this highly efficient infrastructure, built and designed out of inexpensive hardware, a completely customized stack of opportunely modified and refined open-source technologies constitutes the back-end of the largest social network.

The reference stack serving Facebook is based on LAMP (Linux, Apache, MySQL, and PHP). This collection of technologies is accompanied by a collection of other services developed in-house.

These services are developed in a variety of languages and implement specific functionalities such as search, news feeds, notifications, and others.

While serving page requests, the social graph of the user is composed.

The social graph identifies a collection of interlinked information that is of relevance for a given user. Most of the user data are served by querying a distributed cluster of MySQL instances, which mostly contain key-value pairs.