Cloud Applications

Dr. Prasenjit Chanak
Assistant Professor

Department of Computer Science and Engineering
Indian Institute of Technology (BHU), Varanasi-221005

Scientific Applications

- Scientific applications are a sector that is increasingly using cloud computing systems and technologies
- The immediate benefit seen by researchers and academics is the potentially infinite availability of computing resources and storage at sustainable prices compared to a complete in-house deployment
- 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 in order to leverage cloud resources

Scientific Applications

- The most relevant option is IaaS solutions, which offer the optimal environment for running bag-of-tasks applications and workflows
- Virtual machine instances are opportunely customized to host the required software stack for running such applications and coordinated together with distributed computing middleware capable of interacting with cloud-based infrastructures
- PaaS solutions have been considered as well. They allow scientists to explore new programming models for tack- ling computationally challenging problems
- Applications have been redesigned and implemented on top of cloud programming application models and platforms to leverage their unique capabilities
- Problems that require a higher degree of flexibility in terms of structuring of their computation model can leverage platforms such as Aneka, which supports MapReduce and other programming models

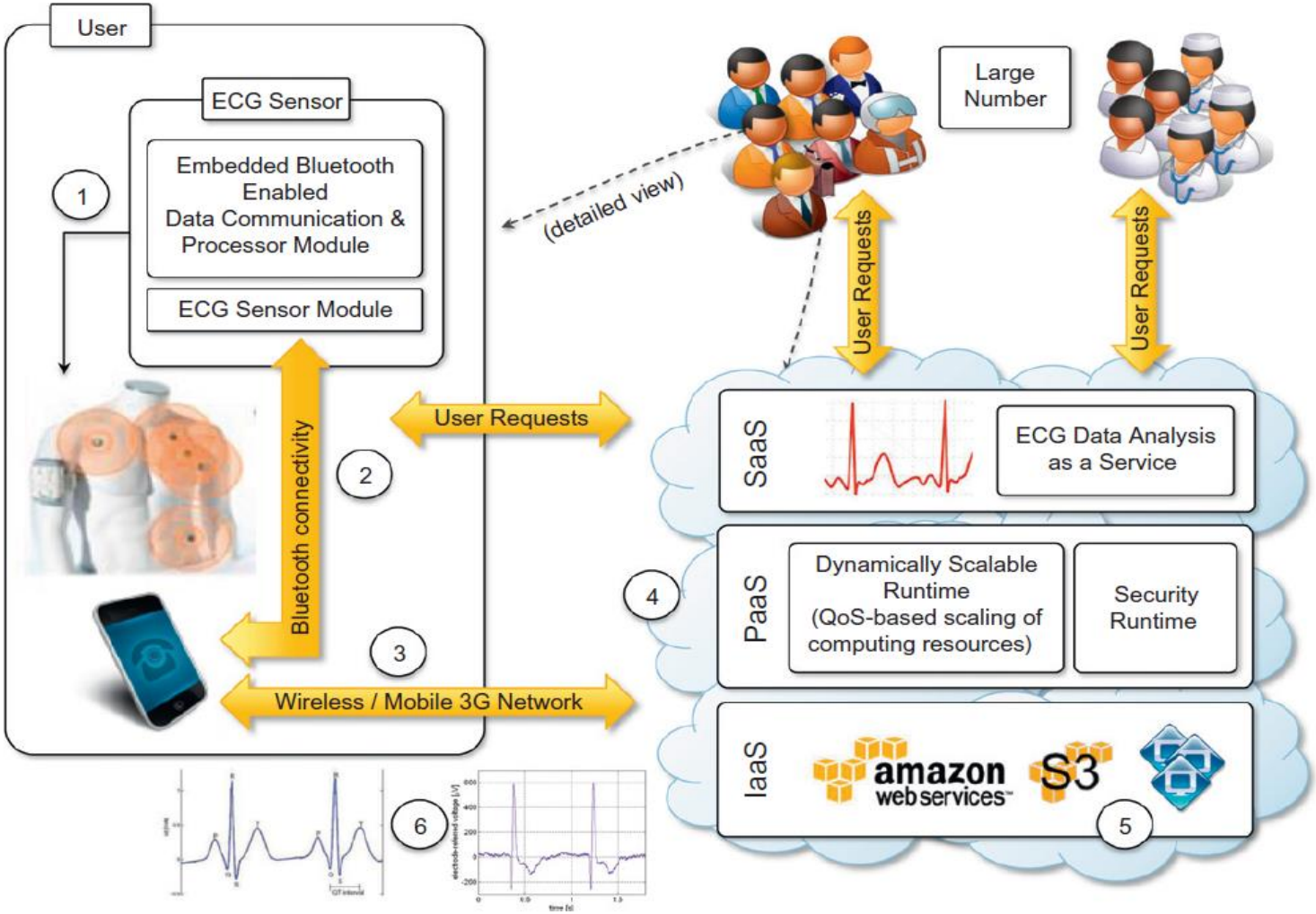
Healthcare: ECG Analysis in the Cloud

- 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 important application is the use of cloud technologies to support doctors in providing more effective diagnostic processes. In particular, here we discuss electrocardiogram (ECG) data analysis on the cloud
- The capillary development of Internet connectivity and its accessibility from any device at any time has made cloud technologies an attractive option for developing health-monitoring systems

Healthcare: ECG Analysis in the Cloud

- ECG data analysis and monitoring constitute a case that naturally fits into this scenario
- ECG is the electrical manifestation of the contractile activity of the heart's myocardium. This activity produces a specific waveform that is repeated over time and that represents the heartbeat
- The analysis of the shape of the ECG waveform is used to identify arrhythmias and is the most common way to detect heart disease. Cloud computing technologies allow the remote monitoring of a patient's heartbeat data, data analysis in minimal time, and the notification of first-aid personnel and doctors should these data reveal potentially dangerous conditions
- This way a patient at risk can be constantly monitored without going to a hospital for ECG analysis
- At the same time, doctors and first-aid personnel can instantly be notified of cases that require their attention

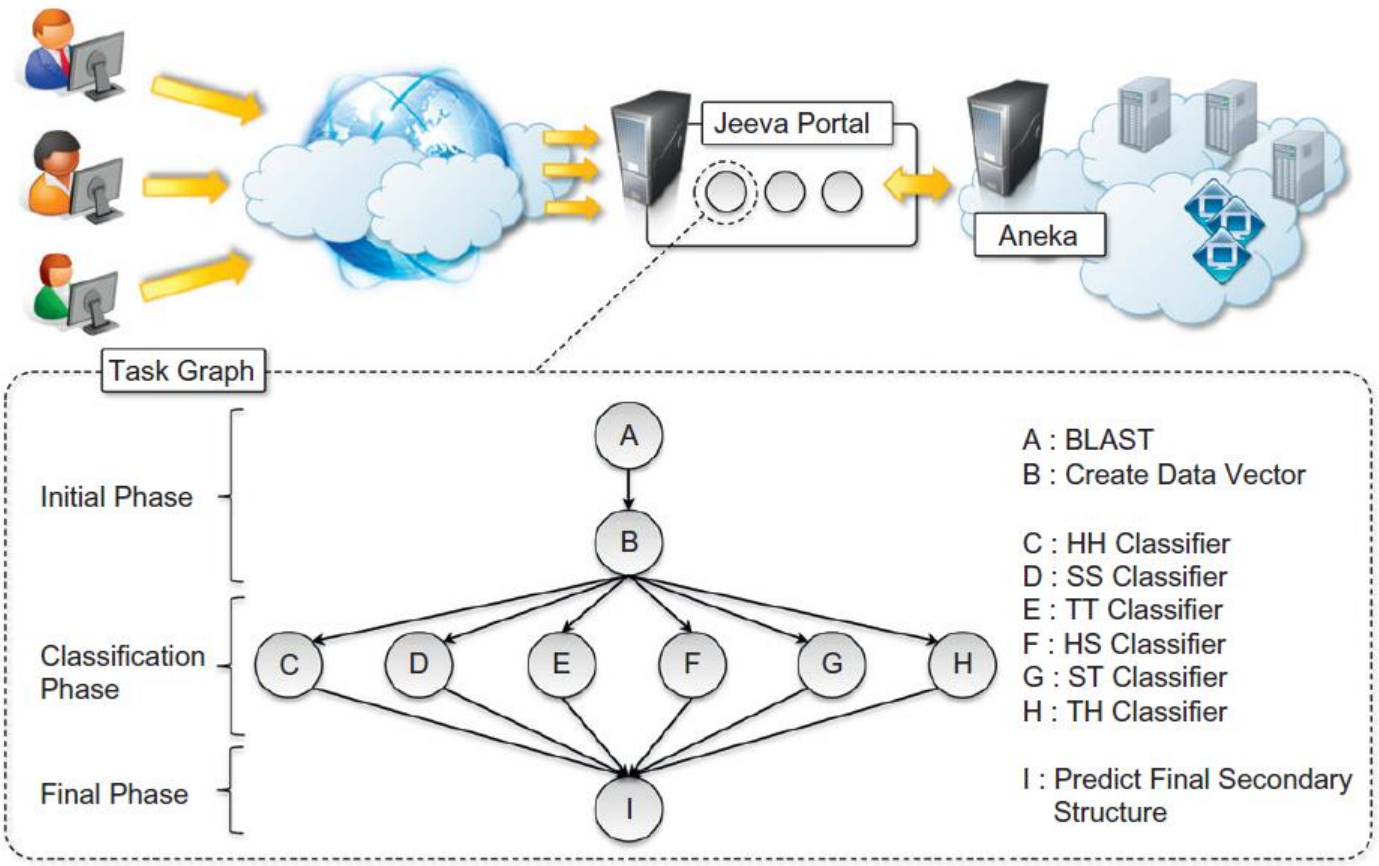
An Online Health Monitoring System Hosted in the Cloud



Biology: Protein Structure Prediction

- Applications in biology often require high computing capabilities and often operate on large data- sets that cause extensive I/O operations
- Protein structure prediction is a computationally intensive task that is fundamental to different types of research in the life sciences. Among these is the design of new drugs for the treatment of diseases
- 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 a large number of computations for each of these states. The computational power required for protein structure prediction can now be acquired on demand, without owning a cluster or navigating the bureaucracy to get access to parallel and distributed computing facilities. Cloud computing grants access to such capacity on a pay-per-use basis

Architecture and overview of the Jeeva Portal



Biology: Protein Structure Prediction

- 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
- The prediction task uses machine learning techniques (support vector machines) for determining the secondary structure of proteins. These techniques translate the 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
- Even though these three phases have to be executed in sequence, it is possible to take advantage of parallel execution in the classification phase, where multiple classifiers are executed concurrently

Biology: Protein Structure Prediction

- This creates the opportunity to sensibly reduce the 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 (i.e., Aneka as scalable cloud middleware) versus conventional grid infrastructures is the capability to leverage a scalable computing infrastructure that can be grown and shrunk on demand
- This concept is distinctive of cloud technologies and constitutes a strategic advantage when applications are offered and delivered as a service

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
- Together with protein structure prediction, this activity is a fundamental component of drug design, since it allows scientists to identify the effects of a specific treatment
- Another 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. This means that all the cancerous cells contain mutated genes. In this context, gene expression profiling is utilized to provide a more accurate classification of tumors
- The classification of gene expression data samples into distinct classes is a challenging task

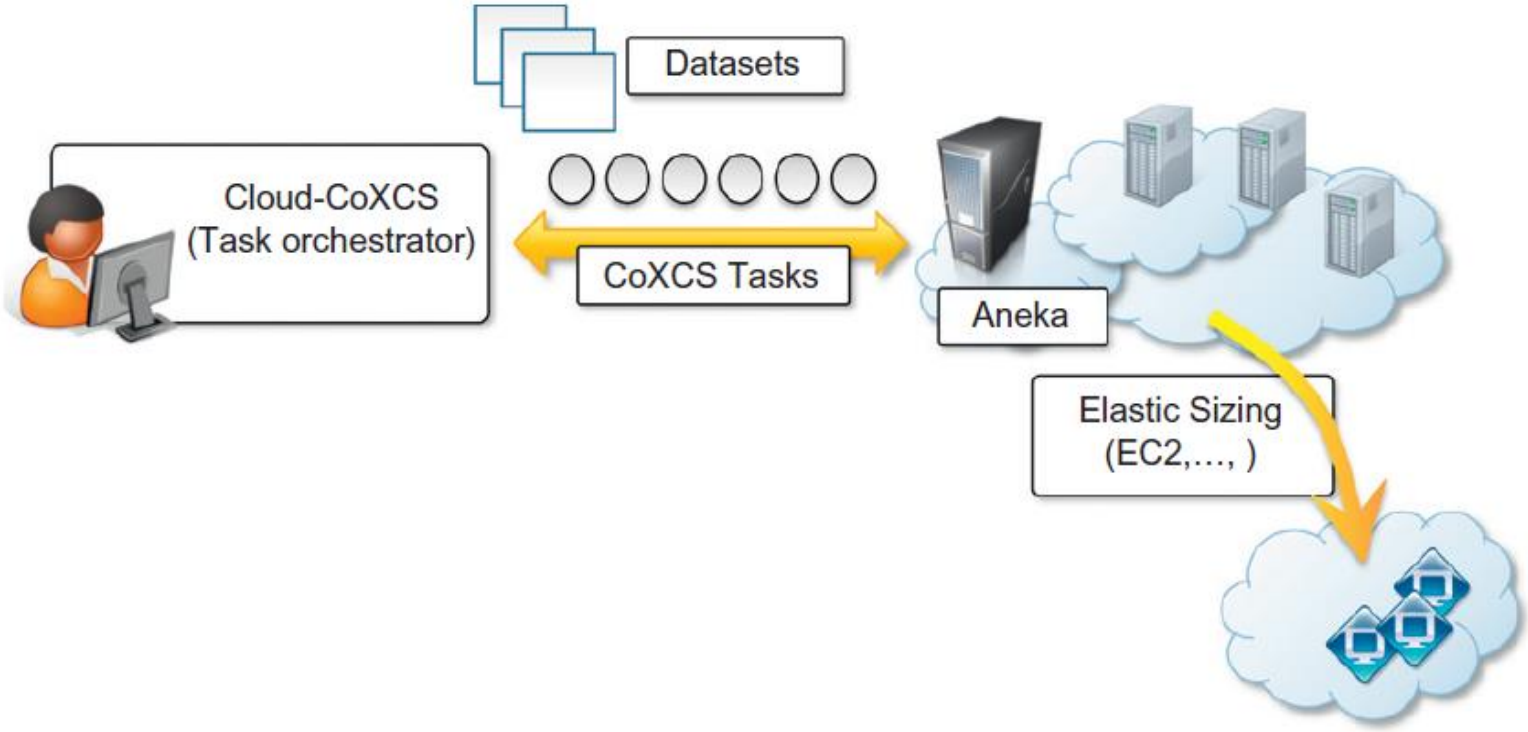
Biology: Gene Expression Data Analysis for Cancer Diagnosis

- The dimensionality of typical gene expression datasets ranges from several thousands to over tens of thousands of genes. However, only small sample sizes are typically avail- able for analysis
- This problem is often approached with learning classifiers, which generate a population of condition-action rules that guide the classification process. Among these, the eXtended Classifier System (XCS) has been successfully utilized for classifying large datasets in the bioinformatics and computer science domains
- However, the effectiveness of XCS, when confronted with high dimensional datasets (such as microarray gene expression data sets), has not been explored in detail. A variation of this algorithm, CoXCS, 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

Biology: Gene Expression Data Analysis for Cancer Diagnosis

- Such a process is computationally intensive but can be easily parallelized because the classifications problems on the subdomains can be solved concurrently
- Cloud-CoXCS 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
- Because of the dynamic nature of XCS, the number of required compute resources to execute it can vary over time
- Therefore, the use of scalable middleware such as Aneka offers a distinctive advantage

Biology: Gene Expression Data Analysis for Cancer Diagnosis



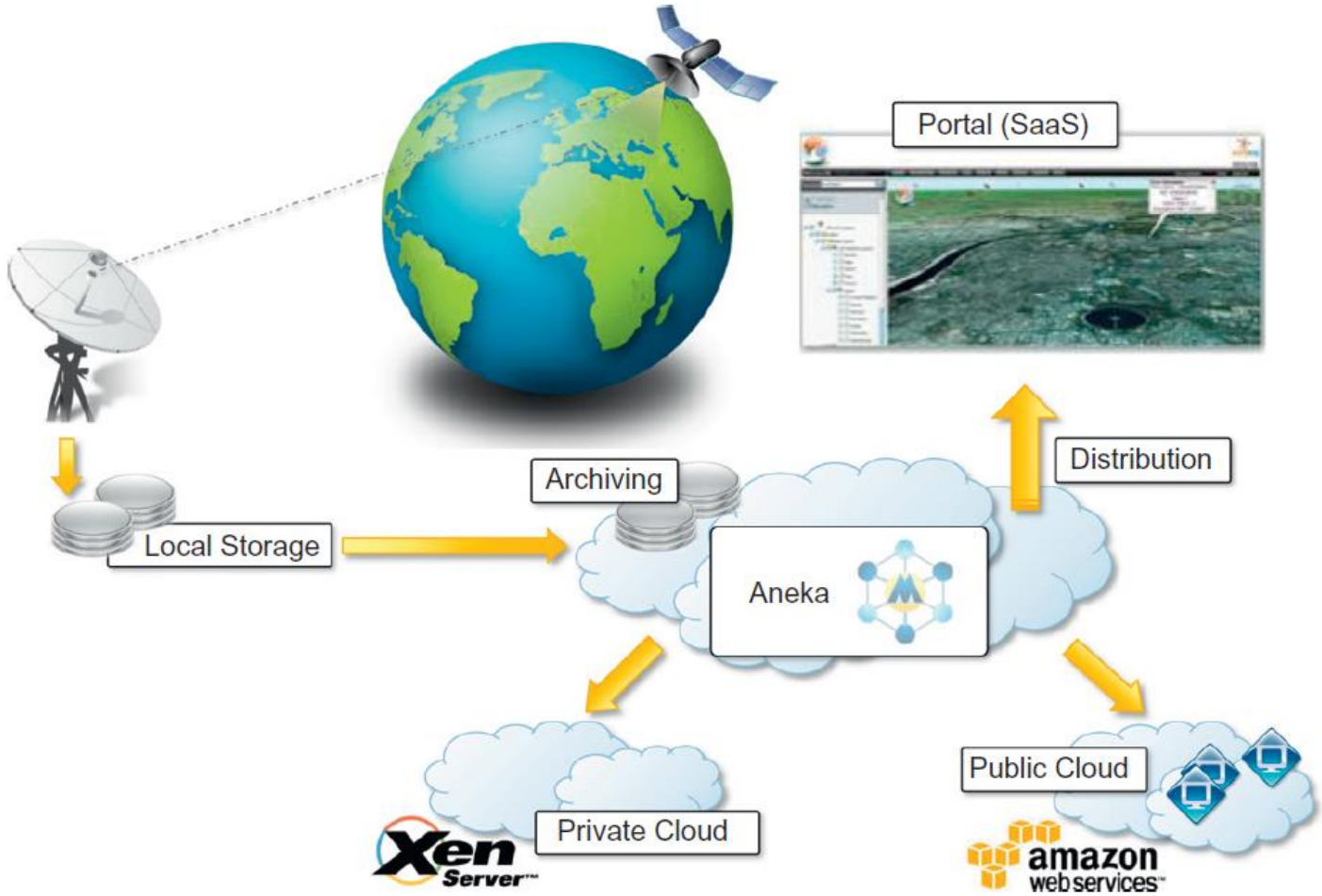
Geoscience: Satellite Image Processing

- In particular, the geographic information system (GIS) is a major element of geoscience applications. GIS applications capture, store, manipulate, analyze, manage, and present all types of geographically referenced data
- This type of information is now becoming increasingly relevant to a wide variety of application domains: from advanced farming to civil security and natural resources management. As a result, a considerable amount of geo-referenced data is ingested into computer systems for further processing and analysis
- 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

Geoscience: Satellite Image Processing

- Large images need to be moved from a ground station's local storage to compute facilities, where several transformations and corrections are applied. Cloud computing pro- vides the appropriate infrastructure to support such application scenarios
- A cloud-based implementation of such a workflow has been developed by the Department of Space, Government of India
- 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 (i.e., grow or shrink) on demand

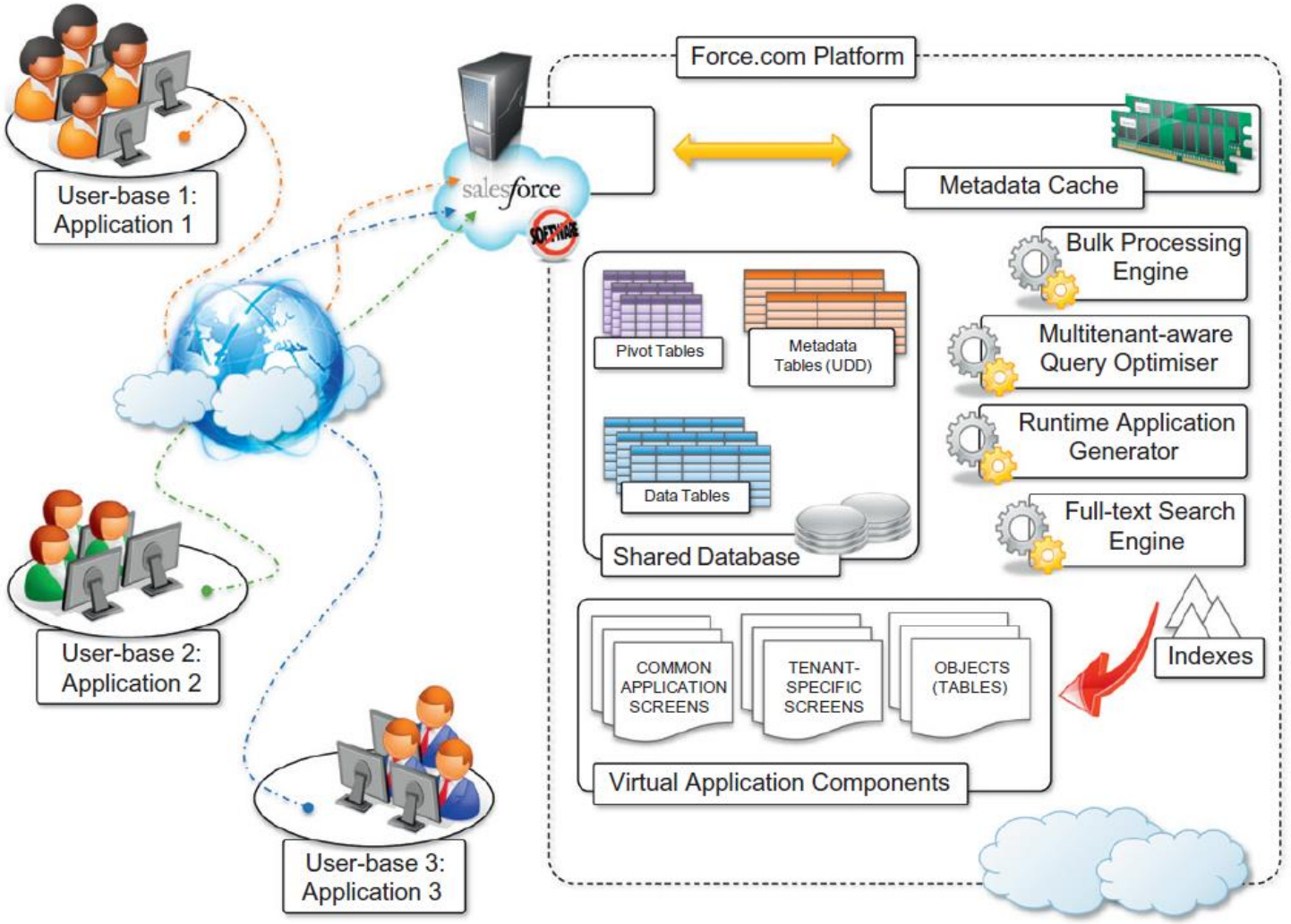
A Cloud Environment for Satellite Data Processing



Business and Consumer Applications

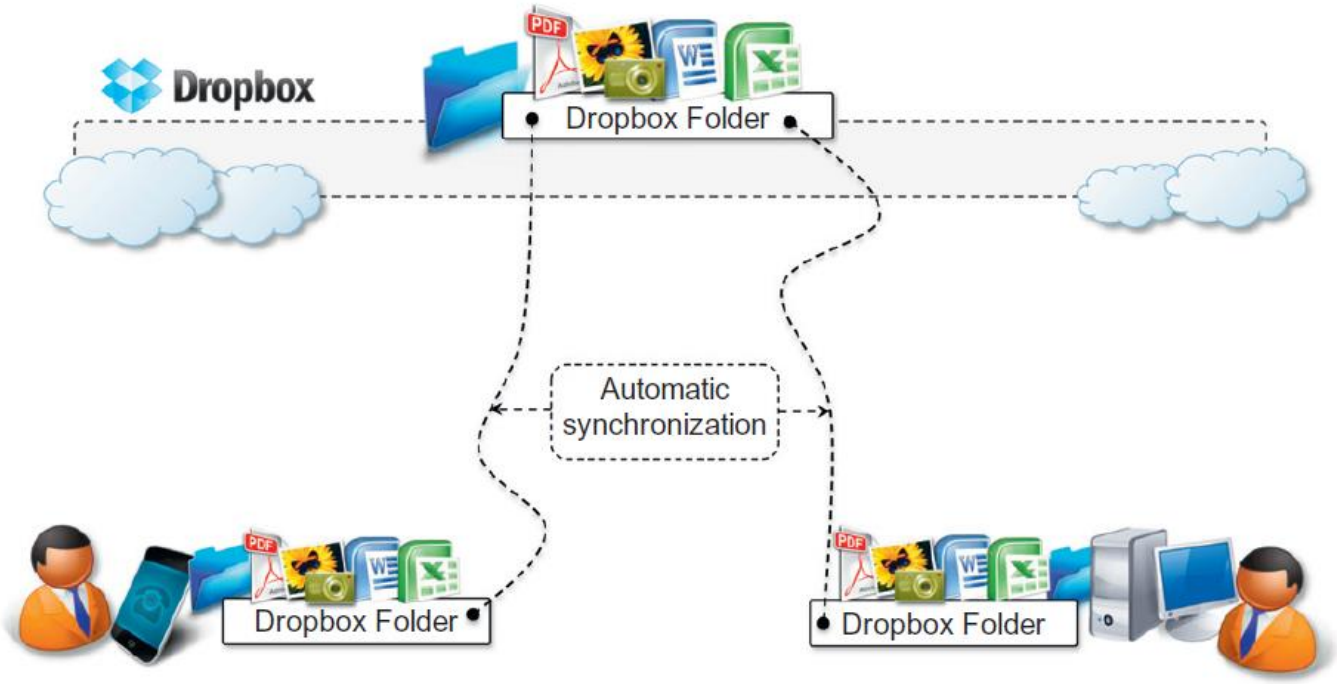
- The business and consumer sector is the one that probably benefits the most from cloud computing technologies. On one hand, the opportunity to transform capital costs into operational costs makes clouds an attractive option for all enterprises that are IT-centric
- On the other hand, the sense of ubiquity that the cloud offers for accessing data and services makes it interesting for end users as well. Moreover, the elastic nature of cloud technologies does not require huge up-front investments, thus allowing new ideas to be quickly translated into products and services that can comfortably grow with the demand
- The combination of all these elements has made cloud computing the preferred technology for a wide range of applications, from CRM and ERP systems to productivity and social-networking applications

Salesforce.com and Force.com Architecture



Productivity

- Productivity applications replicate in the cloud some of 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



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
- The possibility of continuously adding capacity while systems are running is the most attractive feature for social networks, which constantly increase their user base

Multiplayer Online Gaming

- Online multiplayer gaming attracts millions of gamers around the world who share a common experience by playing together in a virtual environment that extends beyond the boundaries of a normal LAN
- Online games support hundreds of players in the same session, made possible by the specific architecture used to forward interactions, which is based on game log processing. Players update the game server hosting the game session, and the server integrates all the updates into a log that is made available to all the players through a TCP port
- The client software used for the game connects to the log port and, by reading the log, updates the local user interface with the actions of other players

Scalable Processing of Logs for Network Games