

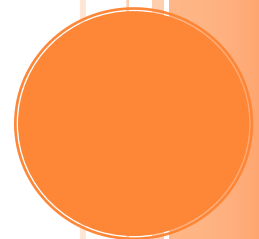
DISTRIBUTED SYSTEMS AND CLOUD COMPUTING

A Comparative Study

Geographically distributed resources, such as storage devices, data sources, and computing power, are interconnected as a single, unified resource and available for users around the world. Repetitive or resource-intensive IT tasks are outsourced to service providers, who execute the task and often provide the results at a lower cost. This report looks into a new emerging model that offers computing as a utility by third parties and bills the user only for consumption. Here we compare the concept of distributed computing for resource-intensive work and offering computing as a utility.

Grace Ramamoorthy

3/4/2011



Distributed Systems and Cloud Computing

A Comparative Study

SECTION: 1 INTRODUCTION

These days computing is categorized according to their usage pattern. Given below are some forms of computing:

1.1 Parallel Computing

It is a form of computation where a large task is divided into unrelated smaller tasks in such a way that these smaller tasks can be concurrently computed. This requires multicore, multiprocessor and massively parallel processing computers within a single machine.

1.2 Cluster Computing

A group of linked computers that are tightly coupled with high speed networking and work closely together is called Cluster Computing. In many respects they form a single computer, provide a single system image illusion and operate mostly in shared memory mode. Computers that form a cluster are homogeneous in their operating systems and hardware specifications. They are normally contained in a single location and are not distributed.

A form of computation where smaller chunks of a large task are performed by a single machine simultaneously is parallel computing.

A homogeneous group of linked computers working together form a cluster.

A heterogeneous collection of computers that work together form a distributed system.

1.3 Distributed Computing

A collection of hardware and software systems that contain more than one processing or storage element but appearing as a single coherent system running under a loosely or tightly controlled regime is called Distributed Computing. The computers in the distributed system do not share a memory instead they pass messages asynchronously or synchronously between them. This is a type of segmented or parallel computing that runs on a heterogeneous system.

1.3.1 Grid Computing

A cluster of networked, loosely coupled computers that work together to perform large tasks is called Grid Computing. Grids provide a dependable, consistent and inexpensive access to computational capabilities. They harness the unused processing cycles of all computers in the network to solve complex problems that are too intensive for a single stand-alone computer. Grids that combine multiple high-capacity resources into a single virtual supercomputer is called a Distributed Supercomputing. When the grid schedules a large number of loosely coupled tasks in unused processor cycles, it is called High-Throughput Computing. When the grid meets short-term resource requirements especially for real-time computing demands, it is On-Demand Computing. Grids are inherently heterogeneous and

distributed. The computers in the grid can run different operating systems and can have different hardware.

1.3.2 Utility Computing

Utility Computing uses distributed systems for less challenging and resource-demanding tasks. Most in-house data centers need to over-provision essential resources to handle peak loads. This leads to idle resources at off-peak times. Therefore, rather than operating servers in-house, organizations subscribe to an external utility computing service provider and pay only for the hardware and software resources as and when they use. Utility Computing relies on the principle of consolidation, where physical resources are shared by a number of applications and users. The utility data centers provide virtual computing environments and charge the user on usage per hour; and data transfer and storage capacity charged per GB or TB used.

Grid Computing is a form of distributed system with non-interactive workloads.

Utility Computing is a business model where computing resources are metered.

Cloud Computing is a form of utility computing that allows developers to develop, deploy and run applications.

1.3.3 Cloud Computing

This is a style of Grid Computing where dynamically stable and virtualized resources are available as a service over the internet. IT services are hosted in the data centers and commercialized. This is a form of specialized Distributed and Utility Computing.

SECTION: 2 DISTRIBUTED COMPUTING

2.1 Explained

Vast improvements in networking, communication protocols and component technology led the way to distributed computing. This replaced some of the huge glass walled computer systems with thousands of workstations and personal computers. These were linked up to do the same or more intensive computing that the large single systems. Distributed Computing in the real sense does not mean one way data-exchange between computers but more intelligent interactions between the systems where the computation and data are distributed.

2.2 Motivation

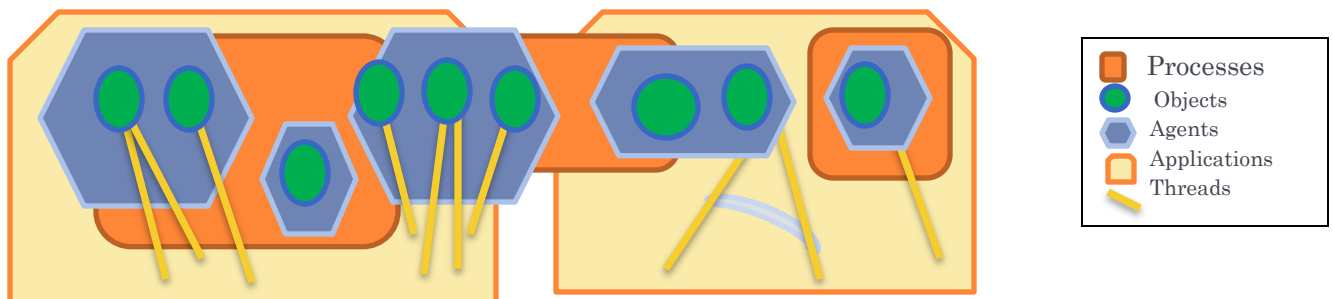
There are many reasons to shift towards Distributed Computing. Some of the common motivations:

- The ability to perform larger tasks by commodity computers instead of more-expensive large dedicated computers.
- Ability to access large and critical data sets remotely and perform tasks on them.

- Ability to have redundant processing of high-availability tasks in multiple computers when fault tolerance is critical.
- To better utilize under-utilized server systems.

2.3 Anatomy

Distributed application is built upon several layers. The lowest layer is the network layer that enables a host of computers to talk to each other. Higher layers provide directory services and security protocols. Distributed applications run on top of these layers. This distributed layer can be broken into process, thread, objects and agents. Each computer can run several processes at once and each process can either be dedicated to one application or shared by many applications. Each process could have one or many threads that work independently or in a synchronized way. Objects are a group of related data with methods that operate on that data. A process can comprise of many objects and each of these objects can be accessed by one or more threads within each process. In a Distributed environment these objects can be logically placed on multiple processes on multiple computers. Agents are functional elements in any application and can be made up of multiple objects and threads and can be distributed across multiple processes. A distributed application is a coordinated group of agents that work together to accomplish a common goal. Each of these agents can belong to more than one application and, therefore, can be distributed across multiple processes on remote hosts and can control multiple objects or threads.



2.4 Future:

The main goal of a Distributed Computing system is to connect users and IT resources in a transparent, open, cost-effective, reliable and scalable way.

When geographically distributed yet interconnected resources, such as storage devices, data sources, and supercomputers, are exploited by users around the world as single, unified resource, distributed computing becomes a service. This approach from organizations offering a large portfolio of scalable and flexible services promises to deliver IT as a utility. Grids, Utility and cloud Computing are the outcome of this trend.

SECTION: 3 CLOUD COMPUTING

3.1 Explained

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services are referred to as Software as a Service (SaaS). The datacenter hardware and software is classified as Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, it is a Public Cloud and the service that is sold is Utility Computing. The term Private Cloud is used to refer to internal datacenters of a business or other organizations not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. People can be users or providers of SaaS, or users or providers of Utility Computing.

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services.

3.2 Core Properties

Cloud Computing means different things to different people – sometimes encompassing incompatible ideas. However, we can broadly identify Cloud Computing as:

- An illusion of infinite computing resources available on demand
- An elimination of an up-front commitment by Cloud users
- An ability to pay for use of computing resources on a short-term basis as needed
- A facility to provide local application response times and 100 percent uptime
- An abstraction of hardware failures and peak load requirements

3.3 Comparison

| | Personal Computers | Centralized Computing | Cloud Computing |
|---------------------|---|---|---|
| Applications | Run by the user on their own computer, hard disk etc., | Run for the user by their IT center on the IT center's central server. | Run for the user by a third party on their shared hardware. |
| Connections | Connect to local hardware. | Connect to LAN to IT datacenter. | Connected through Web to somewhere. |
| Pros | Everything is local. The user owns it and runs it as much as they want. | IT center owns it and runs it for the user. Local connectivity and user needs to know how to use the service. | Third party owns it and user pays by the usage. No worries about backup, storage, power usage or restore. |
| Cons | Users have to be at | Users use what they get. | Connectivity is an |

| | | | |
|----------------|---|--|--|
| | their PC to use it. Users responsible for the backup, power use and storage. | Users don't have control over the corporate decisions. | issue. Everything – data and tools are stored elsewhere. Users don't have control over security. |
| Example | Using PowerPoint 2007 on user's PC with Microsoft Office licensed software. | Microsoft Exchange Server in the company's IT center. | Using Facebook through a web interface – on hardware and software that user knows nothing about. |
| Analogy | Owning your own boat. | Renting a boat for a holiday. | Riding around on someone else's boat by paying for fuel and food. |

3.4 Features

The various major players of Cloud Computing offer SAAS, PAAS, IAAS, HAAS, DAAS etc., This is a simple comparison of the services provided by some of these players.

| | AWS | AppEngine | MobileMe |
|---------------------|--|---|---|
| Availability | Provides raw computing resources through EC2. Higher level services through S3, SQS and SimpleDB. | Software development and deployment. | Does not offer software development and deployment. |
| Languages | Users can configure the raw resources according to their need, use any OS and write applications using any language and any framework. | Users can write any program as long as they use AppEngine framework, Python and BigTable. Google might allow other languages as well. | |
| Services | EC2 is hardware as a service. S3, SQS and SimpleDB are platform as a service. | Platform as a service that also offers deployment as a service. . | Software as a service. |

SECTION: 4 COMPARING CLOUD AND DISTRIBUTED COMPUTING

Distributed Computing is a group of networked computers that work on a smaller portion of a large task simultaneously. This group can consist of any kind of computer including virtual servers from a cloud. Distributed Systems are open and scalable. The computers in the system can come and go and the distributed system server must assign the work and collect the output from the systems as they arrive and leave.

On the other hand Cloud Computing is a specialized form of Distributed Computing. In Cloud Computing the underlying resources, such as storage, processors, memory, are completely abstracted from the consumer. Therefore the vendor of the cloud service is responsible for the reliability, performance, scalability and security of the service. For the developer this is a big advantage. For the companies, this reduces capital expenditure. The disadvantages are the loss of control over the outsourced resources, latency, network and cloud integration issues.

SECTION: 5 ADVANTAGES OF CLOUD COMPUTING

Cloud Computing is a form of Utility Computing with the following advantages:

5.1 User- Centric

Cloud services are simple to use. Users are not required to change their work environments. It is similar to accessing a traditional public utility such as water, electricity or gas.

User-centric, On-demand provisioning, autonomous system, scalability and flexibility are the features of Cloud Computing.

5.2 On-Demand Provisioning

Cloud provides resources, infrastructure and services according to users' demands.

5.3 Autonomous System

In contrast to the distributed systems, the computing system is managed transparently to users. Hardware, software, and data in the cloud are configured to present a single platform image.

5.4 Scalability and flexibility

These are the most important features of Distributed as well as Cloud Computing. The services and platforms are flexible and can be scaled to meet user requirements.

5.5 Cost Reduction

For small and medium-sized enterprises, the ability to outsource IT services and applications offer the potential to reduce overall costs. It also enables them to try processing-intensive activities by eliminating the need for up-front capital investment and reduces maintenance cost on dedicated infrastructure. By using applications from the cloud, the users save on software license fees and maintenance and update costs.

SECTION: 6 CHALLENGES TO CLOUD AND DISTRIBUTED COMPUTING

The continuous and successful deployment of computing as a utility presents challenges, including issues of privacy, security, liability, access, and regulation. Distributed Computing models operate across international borders, and raise jurisdiction and law enforcement issues similar to those of the Internet. Some of the challenges of the distributed system have filtered down to Cloud

The challenges include issues of privacy, security, liability, access, and regulation.

Computing. As Gartner summarizes, the seven issues cloud customers should address before migrating from in-house infrastructure to external resources are privileged user access, regulatory compliance, data location, data segregation, data recovery, investigative support, and long-term viability.

6.1 Network Reliability

Users will expect the cloud to be reliable especially if the cloud is running a mission-critical application. Data integrity and correctness of results depend on network reliability. However, network is not always reliable. When we add all the points of failure for all the various services and APIs that any distributed application uses, the odds of failure are not just high, they are a given. Instead, it is important to build an app that can function at reduced capacity when a given service is offline.

6.2 Latency

Geographic distribution affects performance in a big way. While building distributed and cloud applications, it is good to factor in latency and distribute services closer to customers by building cloud availability zones and content delivery networks.

6.3 Bandwidth

These days applications are accessed by heterogeneous devices and so bandwidth is not a guarantee all the time. Restricting the quantity of data that is shipped across by the application will help handle bandwidth problems.

6.4 Network Security

Network is not secured. The levels of privacy and anonymity available to the user of a cloud is lower than the user of desktop applications. Organizations may be concerned about the security of client data and proprietary algorithms; researchers may be concerned about unintended release of discoveries; individuals may fear the misuse of sensitive personal information. Since the physical infrastructure in a distributed computing environment is shared among its users, any doubts about data security have to be overcome. Therefore, to protect the privacy of users, care must be taken to guard the users' data and applications for manipulating that data.

6.5 Access and Usage

The possibility of storing and sharing data in clouds raises concerns about copyright, licenses, and intellectual property rights. **Clouds can be accessed at any time, from any place, by any user with an internet connection.** Licensing, usage agreements and intellectual property rights may vary in different countries, but the cloud hides these differences, which can cause problems.

6.6 Topology

The biggest attraction of cloud and distributed computing is the ability to add hardware at will and thereby changing the topology of the infrastructure. Therefore, it is important not to develop applications that expect static infrastructure design.

6.7 Administrative Control

With cloud and distributed systems, our applications interact with systems outside our administrative control. These systems may have performance, availability or security issues that we have no direct influence over. Staying mindful that these systems are beyond our control can help ensure minimal impact on the services when they are unresponsive.

6.8 Transport Cost

Transport costs are purchased per transaction, per gigabyte, per compute hour, etc. Cloud storage and transportation costs are a major component of application hosting costs. These are priced like any other commodity.

6.9 Standardization

Clouds face a huge hurdle in creating an open environment between clouds. Applications need to be easily portable so that users can move their data between clouds of different providers without much effort. There should be interoperability so that users can employ infrastructure and services from different providers. Standardized interfaces would allow users to use the same code on different distributed computing solutions, which could additionally decrease the risk of a total loss of data.

SECTION 7: CONCLUSION

Cloud Computing is a new term for a long-held dream of computing as a utility. What originally started as Grid Computing, temporarily using remote supercomputers or clusters of mainframes to address large and complex scientific problems that could not be solved on in-house infrastructures, has evolved into a service-oriented business model that offers physical and virtual resources on a pay as you go basis. While Cloud Computing takes Distributed Computing to a utility stage, ubiquitous and unmetered access to broadband Internet is the key to its success. In addition, better standardization, portability and interoperability of its distributed components will help move Cloud Computing to its full potential.

SECTION 8: REFERENCES

<http://cloudcomputing.sys-con.com/node/761678>

http://news.cnet.com/8301-13846_3-10025122-62.html#ixzz1Fdmg68zU

<http://www.slideshare.net/sudarsun/distributed-computing>

<http://cyberaide.googlecode.com/svn/trunk/papers/08-cloud/vonLaszewski-08-cloud.pdf>

http://www.jatit.org/research/introduction_grid_computing.htm

<http://blog.newrelic.com/2011/01/06/the-fallacies-of-distributed-computing-reborn-the-cloud-era/>

http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000090001PDFE.pdf

<http://cyberaide.googlecode.com/svn/trunk/papers/08-cloud/vonLaszewski-08-cloud.pdf>