## Schedule of Course Activities: Session 15

## *[IS 519: Introduction to Cloud Computing Online-Based]*

## *[Instructor: John C. Chan]*

|  |  |
| --- | --- |
| **Overview of Session** |  |
| We will answer the following questions: | 1. Additional subjects on clustering. 2. Concept of a file system. 3. Concept of Load Balancing. 4. … |

**What is a File System?**

In computing, a **file system** (or **filesystem**) is used to control how data is stored and retrieved. Without a file system, information placed in a storage area would be one large body of data with no way to tell where one piece of information stops and the next begins. By separating the data into individual pieces, and giving each piece a name, the information is easily separated and identified. Taking its name from the way paper-based information systems are named, each group of data is called a "file". The structure and logic rules used to manage the groups of information and their names is called a "file system".

File systems can be used on many different kinds of storage devices. Each storage device uses a different kind of media. The most common storage device in use today is a hard drive whose media is a disc that has been coated with a magnetic film. The film has ones and zeros 'written' on it sending electrical pulses to a magnetic "read-write" head. Other media that are used are magnetic tape, optical disc, and flash memory. In some cases, the computer's main memory (RAM) is used to create a temporary file system for short term use.

**Features of a File System:**

### Filenames:

A **filename** (or **file name**) is used to identify a storage location in the file system. Most file systems have restrictions on the length of filenames. In some file systems, filenames are not case sensitive; in others, filenames are case sensitive.

### Directories:

File systems typically have **directories** (also called **folders**) which allow the user to group files into separate collections. This may be implemented by associating the file name with an index in a table of contents or an inode in a Unix-like file system. Directory structures may be flat (i.e. linear), or allow hierarchies where directories may contain subdirectories. The first file system to support arbitrary hierarchies of directories was used in the Multics operating system

### Space management:

### The file system is responsible for organizing files and directories, and keeping track of which areas of the media belong to which file and which are not being used.

File system fragmentation occurs when unused space or single files are not contiguous. As a file system is used, files are created, modified and deleted. When a file is created the file system allocates space for the data. Some file systems permit or require specifying an initial space allocation and subsequent incremental allocations as the file grows. As files are deleted the space they were allocated eventually is considered available for use by other files. This creates alternating used and unused areas of various sizes. This is free space fragmentation. When a file is created and there is not an area of contiguous space available for its initial allocation the space must be assigned in fragments. When a file is modified such that it becomes larger it may exceed the space initially allocated to it, another allocation must be assigned elsewhere and the file becomes fragmented.

### Metadata:

The [length](https://en.wikipedia.org/wiki/File_size) of the data contained in a file may be stored as the number of blocks allocated for the file or as a [byte](https://en.wikipedia.org/wiki/Byte) count. The [time](https://en.wikipedia.org/wiki/System_time) that the file was last modified may be stored as the file's timestamp.

A file system stores all the metadata associated with the file—including the file name, the length of the contents of a file, and the location of the file in the folder hierarchy—separate from the contents of the file.

### Utilities: Initialize, alter parameters of and remove an instance of the file system. Some include the ability to extend or truncate the space allocated to the file system.

* File Access Permissions: Example is to prevent reading or modifying files by a user or group of users. Another reason is to ensure data is modified in a controlled way so access may be restricted to a specific program.

### Maintaining integrity: This includes actions taken if a program modifying data terminates abnormally or neglects to inform the file system that it has completed its activities.

### User data: The most important purpose of a file system is to manage user data. This includes storing, retrieving and updating data.

* …

**Types of File System:**

### Disk file systems: Example, Your PC file system.

* Optical Discs.
* Flash File Systems. Example, Your USB Stick Storage.
* Tape File Systems: Example, the backup file system.
* Database File Systems: Example, IBM’s DB2.
* Transaction File Systems: Example, the travel/flight reservation database.
* Network File System: Example, NFS, SMB etc.
* …

**File Systems and Operating Systems:**

The Linus OS, and its file system is the focus of this class. It is a Unix-like operating systems that create a virtual file system, which makes all the files on all the devices appear to exist in a single hierarchy. This means, in those systems, there is one root directory, and every file existing on the system is located under it somewhere. Unix-like systems can use a RAM disk or network shared resource as its root directory.

Many of the important file system concepts are explained on this video:

<https://www.youtube.com/watch?v=vSMriAHpzQM>

Key Take-A ways:

* Directory vs Folder.
* Root directory (/).
* User directory: /home/john/ etc.
* Current Directory (./); Parent directory (../).
* Absolute vs relative file path.

**What is scalability?**

***Scalability*** is the capability of a system, network, or process to handle a growing amount of work, or its potential to be enlarged in order to accommodate that growth. For example, it can refer to the capability of a system to increase its total output under an increased load when resources (typically hardware) are added.

The concept of scalability is desirable in technology as well as business settings. The base concept is consistent – the ability for a business or technology to accept increased volume without impacting the contribution margin (= revenue −variable costs).

Scalability can be measured in various dimensions, such as:

* *Administrative scalability*: The ability for an increasing number of organizations or users to easily share a single distributed system.
* *Functional scalability*: The ability to enhance the system by adding new functionality at minimal effort.
* *Geographic scalability*: The ability to maintain performance, usefulness, or usability regardless of expansion from concentration in a local area to a more distributed geographic pattern.
* *Load scalability*: The ability for a distributed system to easily expand and contract its resource pool to accommodate heavier or lighter loads or number of inputs. Alternatively, the ease with which a system or component can be modified, added, or removed, to accommodate changing load.
* *Generation scalability* refers to the ability of a system to scale up by using new generations of components. Thereby, *heterogeneous scalability* is the ability to use the components from different vendors.

Scalability, is a main advantage of cloud computing, both in storage, compute powers, and network capacity. The hardware infrastructure is elastic to meet the growing/shrinking demand. Therefore, resources can be efficiently utilized.

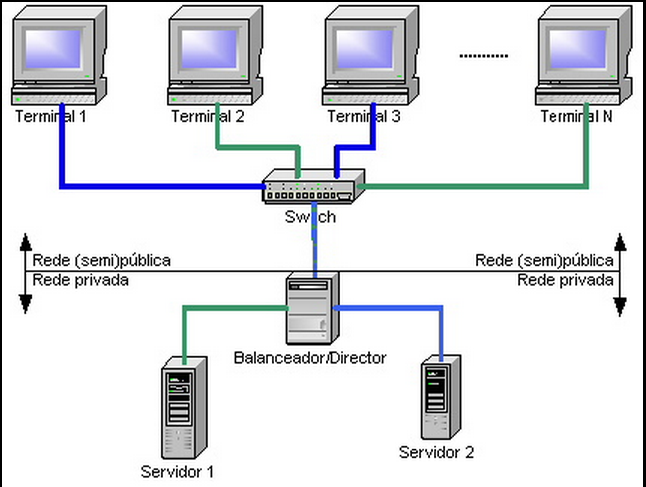
**What is a cluster, why?**

A **computer cluster** consists of a set of loosely or tightly connected computers that work together so that, in many respects, they can be viewed as a single system.

The components of a cluster are usually connected to each other through fast local area networks ("LAN"), with each *node* (computer used as a server) running its own instance of an operating system. In most circumstances, all of the nodes use the same hardwareand the same operating system.

Computer clusters emerged as a result of convergence of a number of computing trends including the availability of low-cost microprocessors, high speed networks, and software for high-performance distributed computing.

Example of a load balancing cluster with two servers, and N-user stations is shown next.



Computer clusters may be configured for different purposes ranging from general purpose business needs such as web-service support, to computation-intensive scientific calculations. In either case, the cluster may use a high-availability approach. Note that the attributes described below are not exclusive and a "computer cluster" may also use a high-availability approach, etc.

**What is load balancing?**

In computing, **load balancing** distributes workloads across multiple computing resources, such as computers, a computer cluster, network links, central processing units or disk drives. Load balancing aims to optimize resource use, maximize throughput, minimize response time, and avoid overload of any single resource. Using multiple components with load balancing instead of a single component may increase reliability and availability through redundancy. Load balancing usually involves dedicated software or hardware, such as a multilayer switch or a Domain Name System server process.

Load balancing is an important metric in cluster configuration and management. The fundamental feature of a load balancer is to be able to distribute incoming requests over a number of backend servers in the cluster according to a scheduling algorithm. (Detail software algorithms, and hardware deployments to facilitate this effort is beyond the scope of the class.)

**What is data redundancy, data recovery?**

In cloud computing, it is important to handle the issue of disk drive failures. Big data means large number of disk drives. It is likely, few of the drives may fail in the course of normal operations.

A **hard disk drive failure** occurs when a hard disk drive malfunctions and the stored information cannot be accessed with a properly configured computer. A disk failure may occur in the course of normal operation, or due to an external factor such as exposure to fire or water or high magnetic fields, or suffering a sharp impact or environmental contamination, which can lead to a head crash.

The specific challenges are: 1) How to recover the data. 2) Replace the failed disk drive, without shutting the system down when doing so.

In most cloud file system, data redundancy are implemented to handle such scenario, where a failed disk drive can be replaced without system shutdowns. Furthermore, the data can be recovered without loss. The level of redundancy can be specified, based on customer needs. The detail mechanism of its implementation, is beyond the scope of this class.

**This video summarize many subjects cover in this session:**

[**https://www.youtube.com/watch?v=9\_qFb1MLjB8**](https://www.youtube.com/watch?v=9_qFb1MLjB8)

**Key Take Ways:**

* + **Scalability.**
  + **File System; Single volume.**
  + **Simplicity of data management.**
  + **…**

**Thought provoking questions about File Systems:**

* + **You have massive amount of data stored in the cloud. How do you locate the data you need? a) On a Window platform. b) On a Linux OS platform.**
  + **Is locating, finding the data become a challenge?**
  + **Since we have ample free storage, we store much more raw data now, e.g. Customer’s purchase habits/patterns. How to you make sense of the vast RAW data, and help to the business?**

**(These questions are beyond the scope of the class).**

End-of-Class Module.

Questions? Please email to me, or post it on Blackboard.

Thank you.