PROCESSING AND PERFORMANCE OF TEXT FILE FORMAT ON DIFFERENT DATA STORAGE SYSTEMS

R. Mariappan[1] , Dr. B. Parthasarathy[2]

[1]Professor, Department of Master of Computer Applications
[2]Professor, Department of Master of Computer Applications
[1,2]Mailam Engineering College, Mailam, Villupuram Distric, Tamil Nadu, India
Email : [1] rm_mca@sify.com

ABSTRACT

Text is the Data, that consists of characters representing the words and symbols of human speech and usually, characters coded according to the ASCII standard, which assigns numeric values to numbers, letters, and certain symbols. Text is obviously the simplest of data type and required the learnt amount of storage in addition that text data can he made fields in a data issue that can be indexed, searched and sorted. Text is also basic building block of a document. This analysis mainly focuses on the text file format to identify the right file to right applications for the performance of storing and accessing speed with less memory space along with the apt file format on specific storage devices, RAM and processors. The file format system is one of the most commonly used read/write process. This analysis evolves based on different types of file format on different storage devices taking into account different research perspectives.

Keywords - Random Access Memory,ASCII,File Format

I. INRODUCTION

File format is the structure of a file that defines the way it is stored and laid out on the screen or in print. The format can be fairly simple and common, as are files stored as “plain” ASCII text, or it can be quite complex and include various types of control instructions and codes used by programs, printers, and other devices. Examples include RTF (Rich Text Format), DCA (Document Content Architecture), PICT, DIF (Data Interchange Format), DXF (Data Exchange File), TIFF (Tagged Image File Format), and EPSF (Encapsulated PostScript Format).

The file system resides permanently on secondary storage, which is designed to hold a large amount of data permanently. The most common secondary storage medium is the disk. The data storage in the system that stores data for easy and fast access at any time and any where. This data storage and retrieval system in is reliable, efficient, errorless and fast access time and read/write speed in different storage devices. The main goal of traditional file systems are provide a comfortable interface for file access to the user and to make efficient use of storage media.

II. RELATED WORK

Regan W.Moore [1], That Digital entities are images of reality, stored as data (bits) that must be interpreted for display by applying information and knowledge. The preservation of digital entities requires the ability to characterize data, information, and knowledge.

Anzelmo, F.D[2]. In developing its, Chemical Abstracts Service (CAS) has built files which require hundreds of millions of bytes of storage. To cope with this size and variability, CAS has developed an internal standardized storage format and has implemented it on the IBM System/360.

Abrahma Silberschatz [3], Observed that Several types of data storage exist in most computer systems. Among the media available are cache, main memory, flash memory, magnetic disks , optical disk and magnetic tapes.

Cyber all [4] It is meant to operate in a COMOHO (commercial office, mobile office, and home office) environment, i.e. computing anywhere, anytime. The main desktop office machine (CO) holds all files.

Gopalan Sivathanu[5], Data integrity is a fundamental aspect of storage security and reliability. The reliable access to data is prerequisite for most computer system and applications. File system consistency is one of the common ones. Data storage patterns are being employed for efficient retrieval of data from secondary storage devices.

A. N. Papadopoulos, and D. Zhang, [6], observed that ,Storage media are of vital importance in order for computer systems to store and organize information

effectively and efficiently. Storage media compose a hierarchy with primary storage (registers, cache and main memory) tertiary storage (tapes, optical disks and flash memory) and secondary storage (hard disks).

Joseph C.Slater [7] observed the Data must be stored and documented in a meaningful way that is absolutely clear. Regarding graphics, use EPS or PDF for all graphics that are not a pure photographs.

There are several file formats that are used in information storage and retrieval system. Data is organized in a data storage hierarchy or increasingly complex form such as bit, bytes/character, fields, records, files and databases.

- DOC [Document]
- PDF [Portable Document File]
- XLS [Microsoft Excel Worksheet]
- RTF [Rich Text File format]
- PPT [Power Point Template]
- HTML [Hyper Text Markup Language]

Process and Performance of Text File Format on different storage system

The process and performance of text file formats and file sizes are shown in the following Table 1 and Fig. 1,

S.No	System Configuration	Process Speed	File Type	File Size
1	P - IV	1.60 G.Hz	DOC	8.4
2	P - IV	1.60 G.Hz	PDF	0.165
3	P - IV	1.60 G.Hz	XLS	8.39
4	P - IV	1.60 G.Hz	RTF	37
5	P - IV	1.60 G.Hz	PPT	8.41
6	P - IV	1.60 G.Hz	HTML	0.5

Table 1. Process and performance of Text File format and File size

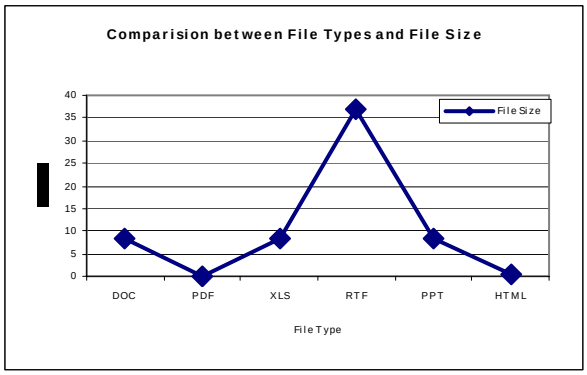

Fig 1 : Process and performance of Text File format and File size

Process and Performance of various text files reading access time on HDD

The process of reading access time on Hard disk with different memory capacity of RAM and Processors with various types of text file format as shown in the following Table 2 and Fig 2.

S.No	System Configuration	Process Speed	File Type	512 MB	1GB
1	P – IV	1.60 G.Hz	DOC	2.3	1.7
2	P – IV	1.60 G.Hz	PDF	1	0.8
3	P – IV	1.60 G.Hz	XLS	1.4	0.9
4	P – IV	1.60 G.Hz	RTF	4.4	3.5
5	P – IV	1.60 G.Hz	PPT	1.3	1.1
6	P – IV	1.60 G.Hz	HTML	0.61	0.52

Table 2: Process and Performance of Text File Format Reading Access Time on HDD

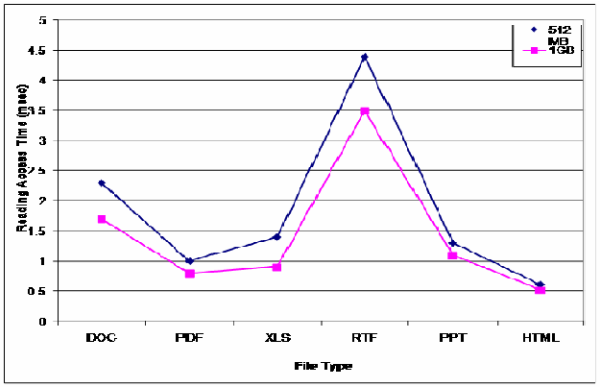

Fig. 2: Process and performance of Text File Format Reading Access Time on HDD

Process and Performance of various text files reading access time on Pen Drive

The process of reading access time on Pen Drive with different memory capacity of RAM and Processors with various types of text file format as shown in the following Table 3 and Fig 3.

S. No	System Configuration	Process Speed	File Type	512 MB	1 GB
1	P – IV	1.60 G.Hz	DOC	2	1.9
2	P – IV	1.60 G.Hz	PDF	1.6	1.42
3	P – IV	1.60 G.Hz	XLS	1.5	1.38
4	P – IV	1.60 G.Hz	RTF	4.2	3.94
5	P – IV	1.60 G.Hz	PPT	1.8	1.71
6	P – IV	1.60 G.Hz	HTML	0.9	0.84

Table 3: Process and Performance Text File Format Reading Access Time on Pen Drive

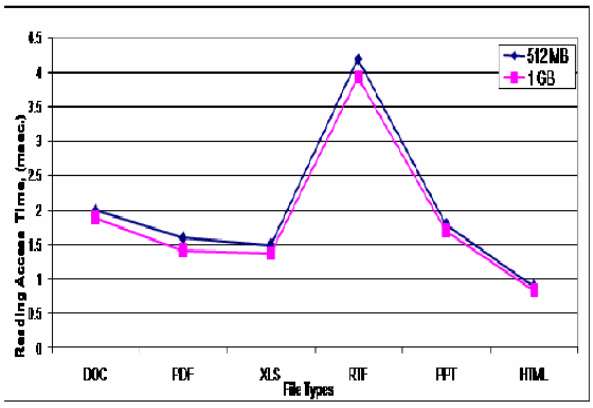

Fig 3: Process and performance of Text File Format Reading Access Time on Pen Drive

DOC [Document]

In computing, DOC or doc an abbreviation of 'document' is a file extension for word processing documents, most commonly for Microsoft Word. File Format .doc is the word binary File Format is a collection of records and structures that specify text, tables, fields, pictures, embedded XML markup, and other document content. A document file format is a text or binary file format for storing documents on a storage media, especially for use by computers.

PDF [Portable Document File]

The PDF, or Portable Document Format, is a file format created by Adobe Systems. This format represents a way to transmit brochures, booklets, tax forms, restaurant menus, software documentation, or virtually any other type of word processing or desktop publishing document. A PDF file can describe documents containing any combination of text, graphics, and images in a device independent and resolution independent format.

XLS [Microsoft Excel Worksheet]

Excel is the most popular spreadsheet program in the world and Excel worksheets can contain data such as numbers, formulas and graphs compiled from multiple sources. The .xls files use a Binary Interchange File Format to store spreadsheet data. XLS files contain data in spreadsheet format which is organised in rows and columns. XLS files are used primarily to store numerical data, and can also contain formulas, charts and graphs, which alter dynamically to reflect the source data. XLS files are widely used in any application that involves the storing and manipulation of numerical data, such as budgeting, accountancy and scientific research.

RTF [Rich Text File format]

Rich Text format (RTF) is file format allows embedding graphics and other file format within a document. The RTF is an important because it is used to attach, embed or link other text file or even binary files such as executables, audio files and video files. The Rich-Text Format extended the range of information carried through from one word-processor application or desktop publishing system to another.

PPT [Power Point Template]

PowerPoint is widely used by business people, educators, students, and trainers and among the most prevalent forms of persuasive technology.PowerPoint presentations consist of a number of individual pages or "slides". The Slides may contain text, graphics, movies, and other objects, which may be arranged freely on the slide. PowerPoint, The presentation can be printed,

displayed live on a computer, or navigated through at the command of the presenter.

HTML [Hyper Text Markup Language]

HTML, which stands for Hypertext Markup Language, is the predominant markup language for web pages. HTML markup consists of several key components, including elements and their attributes, character-based data types, and character references and entity references. Another important component is the document type declaration, which specifies the Document Type Definition. It also can include or can load scripts in languages such as JavaScript, which affect the behavior of HTML processors like Web browsers, and Cascading Style Sheets (CSS) to define the appearance and layout of text and other material. The use of CSS is encouraged over explicit presentational markup.

III. CONCLUSION

This article mainly analyzes the different types of file format system, storage devices, storage and retrieval systems which are most frequently used in the various applications. This analysis mainly focuses on the text file format to right applications for the process of storing and accessing. The file format system is one of the most commonly used read/write process. The storage media and file formats are all vital importance in order for computer systems to store and organize the various kinds of data in effectively and efficiently in data processing and Multimedia applications.

IV. REFERENCE

[1] Regan W.Moore, The preservation of Data, Information and knowledge, University of California, sandiego,2000.

[2] Anzelmo, F.D. Computers, IEEE Transactions on A Data-Storage Format for Information System Files Volume: C-20, Issue: 1 On page(s): 39–43 Jan. 1971.

[3] Abrahma Silberschatz, Hentry F. Korth and S. Sudharssan. “Database System Concept”, 5th Edition, Tata McGraw Hill. 2002.

[4] Cyberall : A Personal Store for Everything, 2000.

[5] Gopalan Sivathanu ,Charles P.Wright, and Erez Zadok Stony Brook University , Computer Science Department “Ensuring Data Intergrity in Storage Techniques and Applications” 2005.

[6] Marios Hadjieleftheriou, A. N. Papadopoulos, and D. Zhang, "Disk Storage and Basic File Structures", Chapter 19 of Handbook of Database Technologies, J. Hammer and M. Schneider (editors), CRC Press, 2007.

[7] Joseph C.Slater, How to archive your/Thesis/Dissertation/Project Data, Write State University, January 6, 2007.

[8] Weir, R. Internet Computing, IEEE Open Document Format: The Standard for Office Documents Volume: 13 , Issue: 2 On page(s): 83 – 87 (2009)

[9] P.C.Bhatt, An Introduction to Operating Systems-Concepts and Practice, Prentice Hall Of India, 2004

[10] Abraham Silberschatz, Hentry F.Korth and S.Sudharssan,”Database System Concepts”, 4th Edition, Tata McGraw Hill, 2002

[11] Raghu Ramakrishnan & Johannesgerhrke, ”Data Base Management Systems”, Mc Graw Hill International Edition, 2000