*Journal of Advances and Scholarly Researches in Allied Education*
*Vol. 19, Issue No. 1, January-2022, ISSN 2230-7540*

# Infrastructure storage Cost: A myth while implementing Enterprise Content Management

**Sirfraz Ul Haque[1]*, Dr. Deepak Shukla[2]**

[1] Research Scholar, Sai Nath University

***Abstract -***

***Infrastructure sizing is very critical for the fast performance of the application in a company where thousands of users try to access to information simultaneously. Every company wants to harness the maximum out of their investment so accurate sizing is required with projected growth rate.***

***Most of the companies feel that hardware cost takes lion share in it but actually it is a myth.***

***Purpose: The purpose of this document is to outline the various cost involved while implementing the Enterprise Content Management in a manufacturing company. Most of the decision making revolves around the infrastructure cost while there are other costs which are involved in managing the data inside an organization. To provide good system performance initially and over next five years, this analysis will put some empirical data to reach a conclusion.***

***Design/methodology/approach: The paper presents and analyses an empirical study of an ECM project where SAP is used as leading ERP solution.***

***Findings: The findings of the paper include a definition of a process model for sizing infrastructure which will help companies to make a right decision while implementing ECM.***

***Research limitations/implications: An overview of sizing infrastructure for ECM system implementation in manufacturing industry where SAP is used as leading ERP solution. Due to data privacy, the client name is not mentioned.***

***Practical implications: This paper is applicable for those organization who are planning to implement ECM and Archiving and hovering around storage cost which looks big to them.***

***Originality/value: There are many articles which deal with Sizing guideline. But this article deals with storage cost and administration cost of managing the data in manufacturing industry and the myth around it.***

***Keywords - Hardware Sizing, System Performance, CPU, RAM***

----------X----------

## 1. INTRODUCTION

The introduction of information system in a manufacturing company where ECM is going to be implemented is a complex process where infrastructure sizing is very important. The investment on information system is continuously expanding in all the companies across the verticals as they are directly responsible for the improvement of productivity. In addition, rapid transitions are taking place in the solution side from mainframe to client/server, Internet, Intranet, which resulted in putting more importance on the management of system performance and capacity management due

to the complexity and increasing usage of the system. In other words, it is very important to calculate the required hardware resources accurately prior to the introduction of the system, because failure of managing the system performance and capacity sizing can result in high cost and resources waste and lower the productivity and cause distrust and credibility breakdown due to the poor service. In general, resource capacity sizing of information system has been conducted by system suppliers or internal resources based on the non-standardized methodology, even though the cost of hardware generally takes up 30%~50% of total project cost. Particularly there is no standardized methodology on resource capacity sizing and CPU performance evaluation. Also, the price of server varies depending on the CPU spec. Therefore, it is necessary to establish a methodology for required resources capacity. The empirical analysis has been conducted based on the survey by working level experts in the organizations and discussion with groups of experts using existing capacity sizing framework and hardware capacity sizing guidelines. This paper is organized as follows.

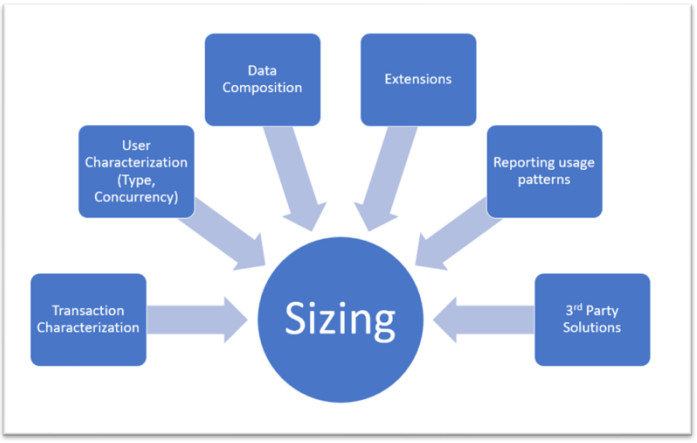
- The research designs
- Factors that affect sizing of ECM
- Data Storage Cost Estimation
- Conclusion

We normally use different terminology for the identical implication of hardware resource capacity calculation, such as Capacity Planning, Capacity Sizing, and System Sizing, and there are few differences among them. System capacity sizing determines system requirements such as CPU type or number, Disk type or volume and memory volume based on the concepts defined by organizations such as TPC(Transaction Processing Performance Council), SPEC(Standard Performance Evaluation Corporation) and IDEAs. In other words, the system capacity sizing uses mathematical methodology based on business process and applications, which is different from the one which decides the required capacity sizing decided by system architecture and application. Hardware sizing is studied when software vendor notify the most suitable hardware size for their package. Many research for software sizing, such as LOC(Line Of Code), FP(Function Point), etc. are exist, but research for hardware sizing is very rare.

Research Design This study had empirical analysis with group of experts from the manufacturing company and suppliers in order to obtain and generalize the appropriateness of capacity sizing method. In this study we have studied the client existing leading application SAP and its data volume growth to reach to a definite conclusion.

## 2. FACTORS THAT AFFECT SIZING OF ECM

Below are some factors which generally influence the sizing. First and foremost is the study of data volume inside the company is very important. The more detailed information that is collected, the more precisely you can determine sizing. If sufficient time is not spent while doing the sizing it will be ikely to be inaccurate. The absolute minimum requirement for necessary data is the peak transaction line load per hour.




Above picture depicts the factors which are used to accurately estimate the sizing of infrastructure. We will discuss the above factors one by one briefly.

**Transactions –** Inside a manufacturing company there will be various department where number of transactions are different than others. For accurately determining the infrastructure sizing it is important to measure the overall transaction and on top of that which department is hitting more transactions per day. Typically, some transactions have certain peaks throughout the day/week. This mostly depends on the transaction type. HR department would be having huge transactions during payroll time but Sales and Marketing department is having more transactions per day through out the month.

**Number of concurrent users –** This is also important factor while determining the sizing of the environment. How many users at the same time creating the same transaction and accessing the same information directly impacts the memory requirement inside the company. To find out about their average frequency is a tedious task. To estimate an approximate number for correct sizing concurrent users, play a very important role. Criteria

for concurrent users is that the user meets all the following criteria:

- Number of Logged on user per hour
- On what transactions they are working on
- There should not be idle session

**Data composition –** This is mostly concerned with the setup and configuration of your system. For instance, how many legal organizations will you have, how many BOM levels will you have, and how complex will your security setup be? Each of these issues may have a minor impact on performance, but they may be mitigated by making wise infrastructure decisions.

**Extensions –** Inside a company there are loads of customization done to make the process working as per its requirements. It can simple sometime and it can be very complex too. Sizing of the infrastructure required varies depending on the number of customizations and the nature of complexity and usage.

Performance assessments are recommended for complex adaptations to guarantee that they are not only checked for efficiency but also to help understand the infrastructure requirements. This is especially more important if the extensions aren't coded according to performance and scalability best practices.

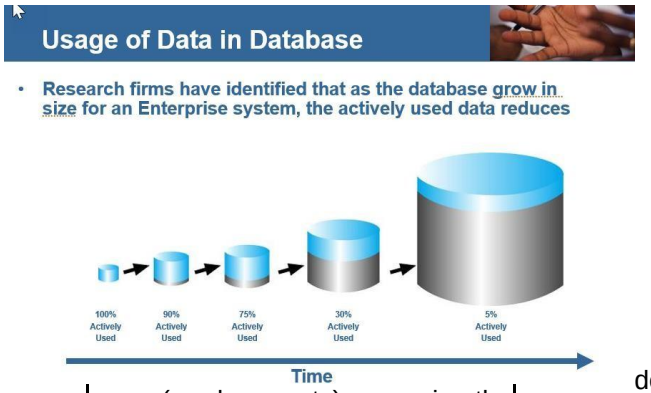
**Reporting and analytics** – Reporting is the next important parameter which plays a vital role while determining the infrastructure required. To run a customized report which directly hits the database to fetch data to present it to business for a meaning decision. It is also important like how frequently these reports are run in the business hour.

**Sizing your environment**

To understand your sizing requirements, you need to know the peak volume of transactions that you need to process. Below table explains all the breakdown in micro level for accurate sizing of the infrastructure while implementing ECM.

| | |
|---|---|
| Planning period | Planning period should cover 3 or 5 years. |
| | After approx. 4 years hardware is out-dated and technology has changed. |
| Business scenario | Try to understand the business scenario: document flow and users involved. |
| | E.g. for data archiving, the number of users is irrelevant. |
| | One archive server can be used for different business scenarios. |
| Disk space | GB are now standard for partitions! Don't care about MBs. |
| Performance | Generally: fast disk I/O is more important than a high number of processors and large memory. |
| Main Memory | When using ISO images (typical size 4-10 GB): |
| | Adding memory improves write performance (max 4 GB usefull) |
| SIA & Compression | Single instance archiving (SIA): for groupware only |
| | SIA or compression requires processor capacity. |
| Single documents | Are documents stored as single files or in container files (ISO image). |
| ISO Image | For high volumes ISO images are recommended. |
| | May depend on storage platform used. |
| Local backup | Is local backup sufficient, or is more data security required (keep copy of the data in a remote location) |
| Remote Standby | |
| No. of documents to be archived per day | Don't calculate with more than 10 hours! The rest of the day must be reserved for administration, backup or trouble shooting.<br>8 hours: time available for document input<br>12 hours: time available for archiving (write to storage system) |

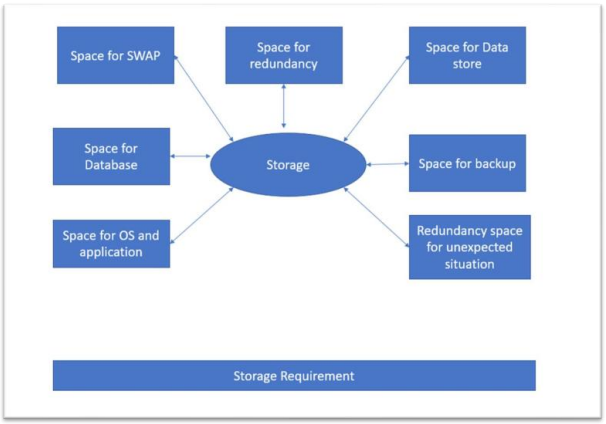
| | |
|---|---|
| No.of documents to be archived per year | Use 220 days per year |
| No. of users | No. of users (licenses):<br><br>every user which may access archived documents<br><br>No. of parallel requests: |
| | users (read requests) accessing the archive server – not the leading application - at the same time! |
| Cache areas | Cache may not be needed for hard-disk based storage systems.<br><br>The disk buffer can be used as cache, if document access rates are high shortly after archiving. |
| | Cache areas are recommended for scan/index scenarios. |
| Compression | Compression factor depends on documents content:<br><br>High for text, word documents or COLD documents<br><br>Low for powerpoints or graphics |
| Typical document sizes | - TIFF: 50 kB per scanned page<br>- PDF: 20 kB per page<br>- ASCII: 5 kB per page |




## 3. DATA STORAGE COST FACTORS IN ECM

Above figure explains about the different factors which play crucial role for accurate sizing of the space requirement.

While implementing ECM major four key factors are



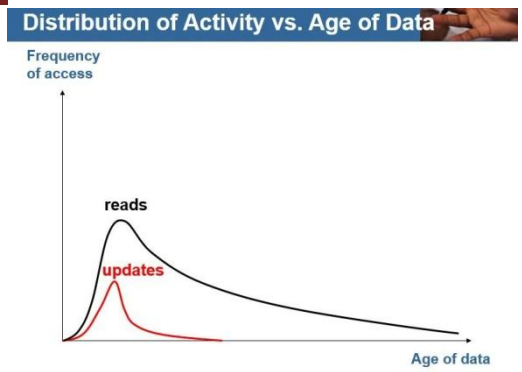

dealt with.

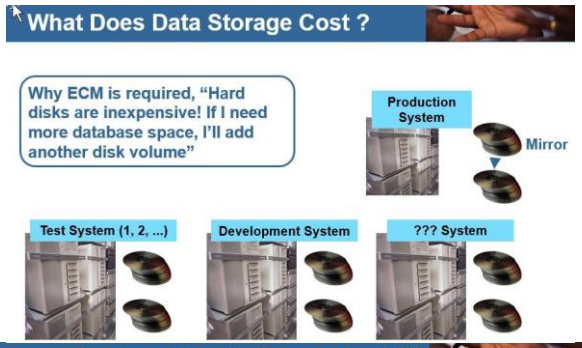
- Data Management
- Document Management
- Taxation and Reporting
- Compliance

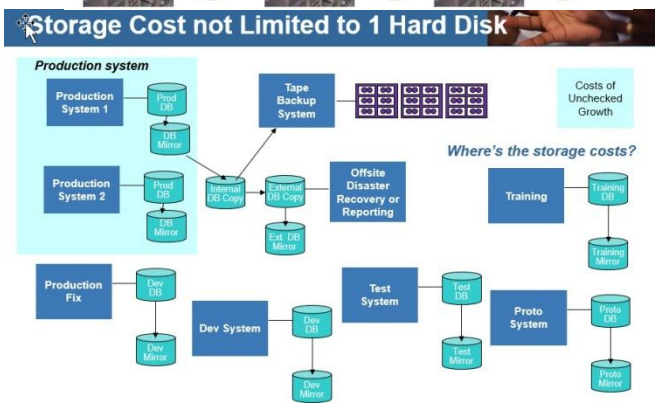
**Data Management**

It is observed across the organizations that data usage gets decreasing as per span of time. The frequency of accessing past data is very low and therefore Data Archiving is done.

There is much research which reach to a conclusion that volume of frequently used data gets decreasing as the time pass and 90% of the total data becomes passive and makes system slow. Below graph about the distribution of activity against data age can be understood easily. Due to this reason while doing the infrastructure sizing the existing data load is analyzed very dedicatedly.

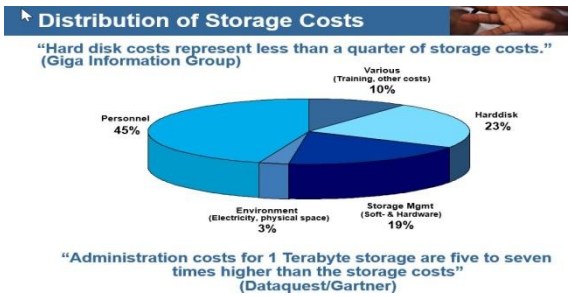


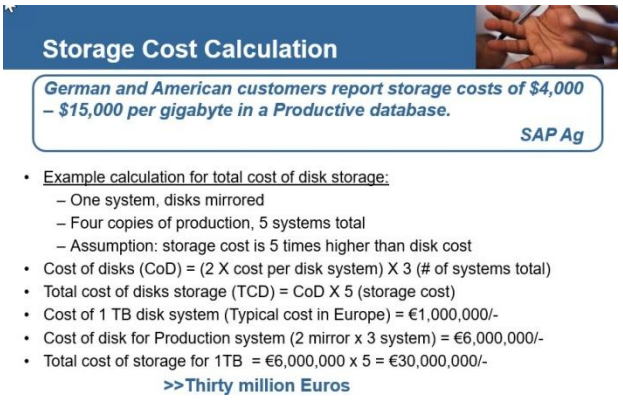




## 4. DATA STORAGE COST ESTIMATION

When distribution cast of data is deeply analyzed, it is found that Hard disk costs just represents 23% of the sizing cost and rest 77% is due to other factors. As per Gartner below is the distribution of maintenance cost of Data inside any company.




As per German and American company the storage cost of 1GB data in production environment is between $4,000 – $15,000. This was found as per research conducted by SAP Ag across various companies. In below slide storage cost is analyzed with breakdown. This also proves that ECM is very critical to any company to reduce the maintenance cost of data and optimize the output.




Document Management: Under Document management below factors are studies in depth so that accurate sizing can be done.

**Printlist Archiving**

All the SAP Generated reports can be stored in Content Repositories for any length of time. Implementing a Printlist (or Report) Archiving will help business to view a report multiple time without executing it again. It is cost effective to store reports for a longer period rather than keeping the data in the database. Example of a printlist could be General Ledger Report for a period.

**Outbound document archiving**

This involves capturing the copy of the documents that are generated by the system in an electronic format such as output type from different process like Quotation, Sales Order, Billing document and shipping documents.

**Inbound document archiving**

This covers capturing the external documents like vendor invoices, delivery notes, to the respective SAP documents.

**Compliance**

There are over 4000 compliance regulations today in the US alone. Important Federal regulations include

- SEC 17a-4
- Sarbanes-Oxley (SOX)

- Health Insurance Portability and Accountability Act (HIPAA)

The legal compliance requires that data and documents that are stored external to the SAP Database in WORM (Write Once Read Many) devices which will make sure that data is not modifiable after the data is created. Many SAP Certified archiving vendors like OpenText/IXOS, Documentum, Easy Archive can store and manage the data in WORM devices like, WORM Juke Boxes or WORM complaint disk storage like, EMC's Centera, IBM's Total Storage, NetApps's SnapLock. Many organizations goes through this syndrome like storage device are cheap so whenever data volume is high just add more hard disk to the system but as per analysis of data in a company the same data gets replicated multiple times.

## 6. CONCLUSION

Above is a detailed and comprehensive roadmap while taking a decision to go for sizing. When the reasons and concepts were given to the management, most individuals involved in the discussion understood and accepted them.

Storage Cost is just one factor while determining the sizing of infrastructure in a company whereas the major portion of money is spent in maintaining the data inside the company.

The value, effort, and cost for doing the sizing exercise inside the company requires a deep understanding of the process involved.

Accurate sizing not only very important for the company but it also allows hassle free running of the business in an efficient manner.

**REFERENCES**


1. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems by Martin Kleppmann
2. Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Services by Brendan Burns
3. Clean Architecture: A Craftsman's Guide to Software Structure and Design by Robert C Martin
4. Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Services Kindle Edition by Brendan Burns
5. Sizing and Estimating Software in Practice: Making MK II Function Points Work (International Software Engineering S.) by Stephen Treble
6. https://sap.com
7. https://gartner.com
8. http://www.spec.org/
9. https://github.com/



**Corresponding Author**

**Sirfraz Ul Haque***

Research Scholar, Sai Nath University

**Email Id-** sirfraz.haque@gmail.com