ENERGY EFFICIENCY IN DATA BASE DESIGN

M. Mohankumar[1], Dr. M. Anand Kumar[2]

ABSTRACT

Energy efficiency is an increasingly important property of software. A huge number of empirical studies have been conducted on the topic, But current state of the art does not provide empirically validated guidelines for developing energy efficient software .Aims at assessing the impact, in terms of energy savings, of best practices for achieving software energy efficiency, elicited from previous work. By doing so, it identifies which resources are affected by the practices and the possible trade off with energy consumption this paper proposes an explicit definition of green software database design and a tool to support their evaluation. The proposed evaluation tool describes the green efficiency by considering the energy consumption as the main aspect to be studied during the design of the database stage. This approach consist of building a database design by using a green way it reduce the energy consumption pattern of devices at different work load circumstances. The energy consumption data base design is then deployed to estimate the impact of software application based on their resource utilization. Our work has been validated on desktop and server side. The experiments show the effectiveness of the database design that provided relevant information on the energy utilization of software application design of a database in software engineering.

Keywords : Data base design, software application, energy consumption, software engineering, green software engineering.

I. INTRODUCTION

This research work proposes a new algorithm to improve the Energy efficiency of the Green database design. In the previous chapter, the proposed Green based software development architecture was presented. This chapter presents the Green database design analysis of the proposed work. In order to test the design of the front end application design, a set of tests and analysis was performed on back end process of the design.

Green database design analysis was done to trace the sustainability strength of the proposed system. Green database design analysis refers to the study of query behaviors, energy utilization or energy efficiency with a view to finding weaknesses related to energy consumption in the database design with the help of the input and output based queries, queries related to memory utilization and CPU utilization.

[1]Research Scholar, Dept of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India.
Email : mohankumar07@gmail.com
[2]Associate Professor, Dept of CS, CA & IT, Karpagam Academy of Higher Education, Coimbatore, India.
Email : anand2kumarm@gmail.com

The remainder of this paper is organized as follows: Introduction to the problem domain was presented in section I. Section II gives an overview of the literary review in the field of Green Database analysis.The Section III describes about the various tools which are used for finding Green and sustainable software engineering methods. The section IV presents the statement of the problem in SDLC(Software Development Life Cycle). Section V discusses about the present work and it is followed by conclusion in the section VI.

II. Literature Survey

Green Database Analysis

In the last ten years, there was a major usage of paper to maintain the records in Government sectors, Institutions, Factories, industries etc where there is a need to minimize it; this is because to manufacture paper, trees are cut down and this poses a great environmental threat and causes Global warming. But by changing all the data to a storage media, no one is aware that it could also affect the atmosphere in some way or the other. This is the time which has come to realize that just by reducing the usage of paper, one cannot achieve the Green and sustainability i.e. one cannot save the atmosphere from the pressure posed by the maintenance of data. The storage spaces that are used in these days are also doing harm to the atmosphere in some way or the other.

Effect of Data Centers on Environment

Till date, most of the software industries depend on the non- renewable energy sources, that causing sources for the production of energy [1] The data centers will produce heat, for reducing the heat generated by the data centers. The software development companies are using Air-conditioners but the hardware will give maximum performance only in the specified temperature range. Therefore, these air-conditioners consume more energy. The energy used for air-conditioning are produced using non-renewable sources that will produce toxic gases and in turn will lead to ecological warming. The cyclic effect of the data centers on environment. The effect of data centre on environment is multifold and their seven stages are as follows:

1. Running the data centre will produce heat
2. To control the heat cooling system is used
3. Energy will be used to cool the system
4. Production of energy is required for cooling system
5. Use of fuel or Electricity to produce the energy
6. During Energy conversation toxic gases will be emitted and
7. These toxic gases will lead to global warming

Therefore highrisk factors affect the environment directly or indirectly when softwares are developed.

Solution to the Problem "Green Database"

Problem starts with running of the data centre which will produce heat, when the data centre dealing with huge volume of data; this process will create more heat. This issue is directly related to the efficiency of the hardware and software used for production. Therefore, there is this need to involve Green to increase the performance and efficiency of the prevailing hardware and software.

1. When a new high performance hardware is introduced, it may produce less heat, but this approach is difficult to maintain in all the scenario as it depends on the cost of the hardware.
2. To reduce the heat production in the data centre, several techniques are already available and it will be produced in the near future. The available techniques are Virtualization, De-duplication, and Cloud Computing. These techniques are used for increasing the performance of the software development companies.
3. If that data centre is located in a cool place, then heat generation will also be less. Thus energy utilization and atmosphere effect will be reduced.
4. Designing of the data centre is to consume less power.

Energy Production for the data centers:

In the modern world people all over the world are utilizing large amount of data centre. They are utilizing the high volume of energy; for example, one of the research result done by the Microsoft shows the daily power consumption of the data centre is equal to the power consumed by the thousands of homes. Such a result shows challenges to the energy production technologies in the present scenario. Suppose the efficiency of the data centre is increased, it is possible to do in two ways; that is, to locate the data centre at a cold place but the need is electricity is must to run the data centre to operate. This will lead to the need to think about the power generation strategies. Some of the options include,

1. Setting up a nuclear power plant for data centre electricity utilization is a good practice but it has its own drawbacks
2. Use of non-conventional sources of energy to supply electricity to the data centre, is not much efficient.
3. Use of photovoltaic cells, evaporative cooling, and waste recycling are some of the procedures that can be practiced to achieve Green.

III. Green Database Steps For Methodology

To build a Green database design it is necessary to follow the steps given below, for Cost effective and efficiency of the database related process

- Understanding: to understand the energy requirement for the data centre
- Design: Based on the understanding, to develop a plan to build the Green database design.
- Virtualization: To increase the performance of the database it is must to follow the virtualization option in the server.

- Measure, Manage and Report: to analyze the developed system and report to the database administrator.

Energy Saving Options and ROI:

Energy Saving Options and ROI [2]

Energy Saving Action	Saving Independent of Other Actions		Energy Saving With the Cascade Effects			ROI
	Saving (KW)	Savings (%)	Savings (KW)	Savings (%)	Cumulative Savings (KW)	
Lower Power Processors	111	10%	111	10%	111	12-18 months
High-efficiency power supplies	141	12%	124	11%	235	5 to 7 months
Power Management features	125	11%	86	8%	321	Immediate
Blade Servers	8	1%	7	1%	328	TCO reduced 38%
Server virtualization	156	14%	86	8%	414	TCO reduced 63%
415v AC power distribution	34	3%	20	2%	434	2 to 3 months
Cooling best practices	24	2%	15	1%	449	4 to 6 months
Variable capacity cooling: Variable speed fan drives	79	7%	49	4%	498	4 to 10 months
Supplemental cooling	200	18%	72	6%	570	10 to 12 months
Monitoring & optimization: Cooling units synchronized	25	2%	15	1%	585	3 to 6 months

Power consumption is nowadays constraining any type of IT device, from mobile phones to supercomputers. IT energy consumption keeps rising, now reaching almost 5% of world wide electricity consumption (920 TWh) [3]. Data centers are responsible for about one-third of this amount and reducing their energy impact is more and more a priority not only for environmental reasons, but also (if not mostly) for their economic sustainability: operational electricity costs account for 20% of a datacenter Total Cost of Ownership (TCO) [4]. Traditionally, energy efficiency has been mostly a concern of hardware experts and very specific computing fields, such as embedded systems or high-performance computing, due to the peculiar energy requirements of such contexts. Recently, the Software Engineering community has defined Energy efficiency has a relevant, general property of software [5],[6,7]. In this regard, envisioning the role of researchers in Software Energy Efficiency and providing guidelines to address Software Energy Efficiency concerns, is based on sound empirical evidence.

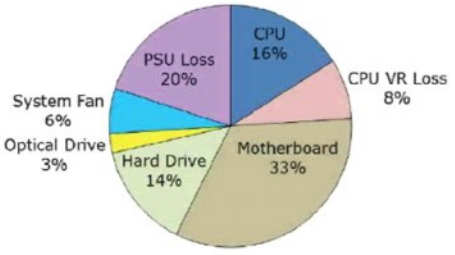
In particular, this study aims at evaluating the energy impact of Education Relation based software Mapping (ERM) frameworks, widely used in education based software applications typically running in data center environments. An empirical experiment is performed aiming at determining the trade-offs in terms of performance and energy consumption of two different types of servers, they are normal server and Green server, with respect to using plain SQL queries in the source code

Green Database Design and Metrics:

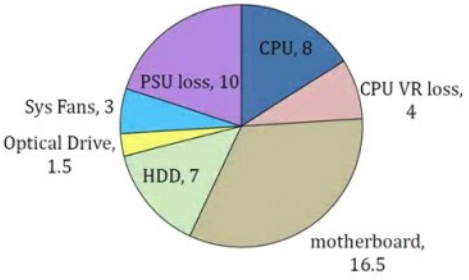
The extensive research activities take place only in the data centers which are based on the models of Green approaches. But it is only useful for the huge volume of record sending and receiving companies. When considering the medium and small size projects, the data center approach is not suitable because it is costlier and virtualization of a database rate is also high. Hence the Green database design

model is executed only for a medium size project based companies. The energy management becomes a serious feature in designing and operating of Database Management Systems (DBMS). During a design, the designer should consider the Greenness of the software because the energy efficiency is so important like the database design for software. Software engineer plays a significant role in reducing the power utilization of the application while they write; database designer and database administrator also possibly can reduce the power consumption of the application by the help of normalization of query and procedure. Present center of attention areas on improving the data center energy competence are the energy competent application design and application co-design When the study thinks the design of the database that is required will give more importance for the below table points.

The proposed solution resolves the power awareness of the database as in Table 2. The only setback is that the database management system needs to be revisited or redesigned and the initial efforts are to be invested. In the design stage, approximately each and every one of the actions are energy associated and most of the investigator are paying attention on the energy utilization and energy competent and only in other aspects. But this work gives an accurate solution in the Green design point of view. When a design is created on the basis of the satisfactory stage of sustainability, then one bestow the star credit for those companies. This method is corresponding to the Capability Maturity show organize procedure of the repeatable second stage. It is equivalent to the standard technique in completely through the plan of the structures; database, design the engineer, database overseer. In this way, the Architect needs to work for this strategy and meet the clients' level of expectations. The following figure shows a breakdown of the idle consumption of a typical desktop.

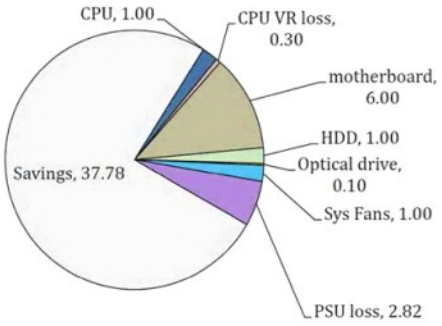


Typical PC Idle Consumption by Percentage[10]



Idle Mode of a Desktop System in Watts[10]

The above figure shows the translates the ITI break-down for a typical computer with an idle of about 50 watts



Idle Mode Power Consumption With Efficient Components, Watts[10]

Table: 2 Common Parameters Details

Parameters	Description
Hardware	Hardware Optimization
	Power Management
Software	Software Optimization
	Query Management and Optimization
Architecture	Architecture Maneuvering
	Distributed System

According to the parameters shown in the Table 2 above, when the hardware setting was different from the other Power Management and Energy Utilization, they may also differ from one to one. The same in Software setting, Query Management and Optimization give different energy utilization, and even the architecture varies in the energy consumption and utilization. These aspects raise the following research questions:

- How queries are executed in the different hardware settings in Green Database?
- What are the energy saving possibilities in the Green Database Design?
- What are the procedures involved in employing the energy efficiency via Green Computing and Green Database System?
- How could these query determinants be the best implemented for energy efficiency and Green Database Design?

Therefore it is considered as an improved solution to show the benefits of energy efficiency through the Green database design[8]. The selection of the best way to partition the data in the distributed environment is a critical physical database design problem[9]. Dramatic performance improvements are achieved through the distributed execution of queries across many nodes. Query Optimization for such system is a challenging and important problem. This study obtains certain queries based on the I/O operations related to CPU and memory and for further development of the research.

From the above study helps to find the database design in software development life cycle phases the

Here we are discussing about the design phase how the Database design plays an important role energy efficiency of a SDLC

Design Phase: But this wont gave a correct solution for the Green design point of view. Because data base administrator also plays an important role in design phase if the data base designer design a data base in an optimized manner it will give through the life cycle of the SDLC phase it becomes green and sustainable , here we are indentifying some quires related to I/O, Memory, and CPU utilization ,these queries will help to data base designer to optimize the application, and reduce the risk in design of a data base.

IV. Operation Related Queries

I/O operation related Queries are

Query :1

```
SELECT [Drive],
CASE
WHEN num_of_reads = 0 THEN 0
ELSE (io_stall_read_ms/num_of_reads)
END AS [Read Latency],
CASE
WHEN io_stall_write_ms = 0 THEN 0
ELSE (io_stall_write_ms/num_of_writes)
END AS [Write Latency],
CASE
WHEN num_of_reads = 0 THEN 0
ELSE (num_of_bytes_read/num_of_reads)
END AS [Avg Bytes/Read],
CASE
WHEN io_stall_write_ms = 0 THEN 0
ELSE
(num_of_bytes_written/num_of_writes)
END AS [Avg Bytes/Write],
CASE
WHEN (num_of_reads = 0 AND
num_of_writes = 0) THEN 0
ELSE ((num_of_bytes_read +
num_of_bytes_written)/(num_of_reads +
num_of_writes))
END AS [Avg Bytes/Transfer]FROM
(SELECT
LEFT(UPPER(mf.physical_name),
2) AS Drive, SUM(num_of_reads) AS
num_of_reads,SUM(io_stall_read_ms) AS
io_stall_read_ms, SUM(num_of_writes)
AS num_of_writes,SUM(io_stall_write_ms)
AS io_stall_write_ms,
SUM(num_of_bytes_read)
AS
num_of_bytes_read,SUM(num_of_bytes_w
ritten)
AS num_of_bytes_written, SUM(io_stall)
sys.dm_io_virtual_file_stats(NULL, NULL)
```

```
AS io_stall FROM
sys.dm_io_virtual_file_stats(NULL,
NULL)

AS vfs INNER JOIN sys.master_files

AS mf WITH (NOLOCK) ON
vfs.database_id = mf.database_id
AND vfs.file_id = mf.file_id GROUP
BY
LEFT(UPPER(mf.physical_name), 2))

AS tab ORDER BY [Overall Latency]
OPTION (RECOMPILE);
```

Look for I/O requests taking longer than 15 seconds in the five most recent SQL Server Error Logs

Description: Finding 15 second I/O warnings in the SQL Server Error Log is useful evidence of— poor I/O performance (which might have many different causes)

Query: 2 Calculates average stalls per read, per write, and per total input/output for each database file

```
SELECT DB_NAME(fs.database_id) AS
[Database Name],

CAST(fs.io_stall_read_ms/(1.0 +
fs.num_of_reads)AS NUMERIC(10,1))

AS
[avg_read_stall_ms],CAST(fs.io_stall_writ
e_ms/(1.0 + fs.num_of_writes)

AS NUMERIC(10,1)) AS
[avg_write_stall_ms],CAST((fs.io_stall_rea
d_ms + fs.io_stall_write_ms)/(1.0 +
fs.num_of_reads + fs.num_of_writes)
```

```
AS NUMERIC(10,1)) AS
[avg_io_stall_ms],CONVERT(DECIMAL(1
8,2), mf.size/128.0) AS [File Size (MB)],
mf.physical_name, mf.type_desc,
fs.io_stall_read_ms, fs.num_of_reads,
fs.io_stall_write_ms, fs.num_of_writes,
fs.io_stall_read_ms + fs.io_stall_write_ms
AS [io_stalls], fs.num_of_reads +
fs.num_of_writes AS [total_io] FROM
sys.dm_io_virtual_file_stats(null,null) AS
fsINNER JOIN sys.master_files AS mf
WITH (NOLOCK) ON fs.database_id =
mf.database_id AND fs.[file_id] =
mf.[file_id] ORDER BY avg_io_stall_ms
DESC OPTION (RECOMPILE);
```

Description:

Helps determine which database files on the entire instance have the most I/O bottlenecks

— This can help you decide whether certain LUNs are overloaded and whether you might

— want to move some files to a different location or perhaps improve your I/O performance

CPU Related

Query:1 Get CPU utilization by database

```
WITH DB_CPU_Stats
AS
(SELECT DatabaseID, DB_Name(DatabaseID) AS [Database Name], SUM(total_worker_time) AS [CPU_Time_Ms]
 FROM sys.dm_exec_query_stats AS qs
 CROSS APPLY (SELECT CONVERT(int, value) AS [DatabaseID]
        FROM sys.dm_exec_plan_attributes(qs.plan_handle)
        WHERE attribute = N'dbid') AS F_DB
 GROUP BY DatabaseID)
SELECT ROW_NUMBER() OVER(ORDER BY [CPU_Time_Ms] DESC) AS [CPU Rank],
    [Database Name], [CPU_Time_Ms] AS [CPU Time (ms)],
    CAST([CPU_Time_Ms] * 1.0 / SUM([CPU_Time_Ms]) OVER() * 100.0 AS DECIMAL(5, 2)) AS [CPU Percent]
FROM DB_CPU_Stats
WHERE DatabaseID<> 32767 -- ResourceDB
ORDER BY [CPU Rank] OPTION (RECOMPILE);
```

// Helps determine which database is using the most CPU resources on the instance

Query :2 Get CPU Utilization History for last 256 minutes

```
DECLARE @ts_nowbigint = (SELECT cpu_ticks/(cpu_ticks/ms_ticks) FROM sys.dm_os_sys_info WITH (NOLOCK));
SELECT TOP(256) SQLProcessUtilization AS [SQL Server Process CPU Utilization],
SystemIdle AS [System Idle Process], 100 - SystemIdle - SQLProcessUtilization AS [Other Process CPU Utilization], DATEADD(ms, -1 * (@ts_now - [timestamp]), GETDATE()) AS [Event Time] FROM (SELECT record.value('(./Record/@id)[1]', 'int') AS record_id,
        record.value('(./Record/SchedulerMonitorEvent/SystemHealth/SystemIdle)[1]', 'int') AS [SystemIdle], record.value('(./Record/SchedulerMonitorEvent/SystemHealth/ProcessUtilization)[1]', 'int') AS [SQLProcessUtilization], [timestamp] FROM (SELECT [timestamp], CONVERT(xml, record) AS [record] FROM sys.dm_os_ring_buffers WITH (NOLOCK)
        WHERE ring_buffer_type = N'RING_BUFFER_SCHEDULER_MONITOR' AND record LIKE N'%<SystemHealth>%') AS x) AS y ORDER BY record_id DESC OPTION (RECOMPILE);
```

Description: Look at the trend over the entire period — Also look at high sustained Other Process CPU Utilization values.

Memory Related

Query1: Good basic information about OS memory amounts and state -You want to see "Available physical memory is high"

```
SELECT physical_memory_in_use_kb/1024 AS
[SQL Server Memory Usage (MB)],
large_page_allocations_kb,
locked_page_allocations_kb,
page_fault_count,
        memory_utilization_percentage,
available_commit_limit_kb,
        process_physical_memory_low,
process_virtual_memory_low
FROM    sys.dm_os_process_memory    WITH
(NOLOCK) OPTION (RECOMPILE);
```

Description:

This will not return any rows if you have — not had any memory dumps (which is a good thing)

Query 2: Good basic information about OS memory amounts and stateYou want to see "Available physical memory is high"

```
SELECT total_physical_memory_kb/1024
AS [Physical Memory (MB)],
available_physical_memory_kb/1024 AS
[Available Memory (MB)],
total_page_file_kb/1024 AS [Total Page
File (MB)],
        available_page_file_kb/1024   AS
[Available Page File (MB)],
        system_cache_kb/1024          AS
[System Cache (MB)],
system_memory_state_desc AS [System
Memory State]
FROM    sys.dm_os_sys_memory    WITH
(NOLOCK) OPTION (RECOMPILE);
```

Description: This indicates that you are not under external memory pressure

V. Discussion

This paper discussed about the Green database design analysis aimed to evaluate the various levels of the proposed queries the Input and Output operation was performed and the analysis shows that the different query was run in different scenario it was very helpful to trace the information from the database. This means that there is the possibility that can help to reduce the environmental impacts. CPU related queries are performed to measure the strength of the proposed algorithm. The result of these tests shows that different application databases are taken to analyze the CPU utilization on the proposed algorithm in milliseconds. Memory usage, higher order differential memory analysis and auto correlation test are executed in the proposed

algorithm. The test result shows that with the five different databases in the similar application it was very useful to identify the memory usage

VI. Conclusion

The advantages of these queries in the Green and Sustainable Database design approaches in SDLC with regard to the traditional SDLC. If the database designer implements this approach, it will diminish the carbon discharge during the development instance; additionally, it will augment the competence of the artifact with no disturbing the environment dimension. Besides it will give a sustainable software product to the world. This work is meant for the betterment of the world from natural calamities and it is taken for consideration as a serious threat for the survival of human being and proceeds for further research in this specified field.

VII. Reference:

[1]. Sachin Chaudhary , Devendra Prasad Murala and V.K. Shrivastava Global Journal of Business Management and Information Technology. Volume 1, Number 2 (2011), pp. 105-111

[2]. Rick Baur. (2011, September) http://net.educause.edu. [Online]. http://net.educause.edu/ir/library/pdf/bauer.pdf

[3]. E. Gelenbe and Y. Caseau. The impact of information technology on energy consumption and carbon emissions. Ubiquity, 2015(June):1:1–1:15, June 2015.

[4]. C. Gough, I. Steiner, and W. Saunders. Energy Efficient Servers: Blueprints for Data Center Optimization, chapter Data Center Management, pages 307–318. Apress, Berkeley, CA, 2015.

[5]. A. Hindle. Green mining: Investigating power consumption across versions. In ICSE, pages 1301–1304, jun 2012.

[6]. P. Lago, R. Kazman, N. Meyer, M. Morisio, H. A. Muller, F. Paulisch, G. Scanniello, B. Penzenstadler, ¨ and O. Zimmermann. Exploring initial challenges for green software engineering: summary of the first GREENS workshop, at ICSE 2012. SIGSOFT Softw. Eng. Notes, 38(1):31–33, Jan. 2013.

[7]. P. Lago, N. Meyer, M. Morisio, H. Muller, and others. ¨ Leveraging energy efficiency to software users: summary of the second GREENS workshop, at ICSE 2013. ACM SIGSOFT Software, 2014

[8]. Cheng Chen , Bingsheng He , Xueyan Tang , Changbing Chen , Yubao Liu," Green Databases Through Integration of Renewable Energy" 6th Biennial Conference on Innovative Data Systems Research(CIDR 13),jan6- 9,2013,doi=10.1.1.366.2184 Asilomar, California, USA.

[9]. Rashid M, Ardito L, Torchiano M. Energy Consumption Analysis of Algorithms Implementations. ESEM, 2015, 2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). 2015; 1–4. Doi:10.1109/ESEM.2015.7321198

[10]. http://www.energy.ca.gov/