DESIGNING OF A SIMULATOR ARCHITECTURE FOR GREENER DATA CENTER THROUGH KNOWLEDGE TRANSFER BY PARTNERS IN DIFFERENT SECTORS

M. Turker Takci[1], Onur Ozturk[1], A. Aybike Şeker[1], Ahmet Koksoy[1], Tuba Gozel[1], M. Hakan Hocaoglu[1], Habin Lee[2], Youngseok Choi[3], Marios Pournaris[2], Stanimir Yovchev[4], Fatih Karatana[5], Kaan Uzuner[5], Tunç Medeni[5], Halil Yeşilçimen[5]

[1]Gebze Technical University (Turkey)
[2]Brunel University London (UK)
[3]LK Knowledge Engineering LTD (UK)
[4]David Holding (Bulgaria)
[5]Turksat (Turkey)

Abstract

In recent years, increasing demand for internet service providers, cloud services and the information and communications technologies, data centers (DCs) have become a very important place in the energy consumption industry. Furthermore, the energy consumed by the IT industry including DCs reached about 10% of the world's electricity generation. Therefore data centers have become a significant source of CO2 emissions and a crucial player in the electrical power system.

In order to predict energy demand better and reduce energy consumption and CO2 emissions in specific national DCs, the GreenDC Project aims developing a decision support tool. This project was funded by EU through H2020 Marie Skłodowska-Curie. It is progressing by secondment activities that consist of knowledge transfer between academic and industrial partners.

This paper aims to define and explain the architecture of its decision support tool, GreenDC DSS. The GreenDC DSS architecture composed of four interactional layers which are data layer, math model layer, business logic layer, and user interface. In this paper, the design of each layer has been described and the process of entire GreenDC DSS Tool has been examined with an example user scenario which is formed by considering the user requirements of data centre managers. Also, it has been shown that a skeleton of the simulator gives optimum working strategies to data center manager.

Keywords: green data centre, IT, knowledge exchange, decision support systems, CO2 Emission.

1.INTRODUCTION

In recent years, along with an incredible increase in the use of internet service providers and cloud services, the energy used by data centers (DCs) has been skyrocketing due to a large number of hosted servers and associated workload [1],[2]. The cost of providing energy to a typical DC doubles every five years [3]. The annual rate of increase in energy consumed by DCs is higher than the rate of worldwide energy consumption [4]. According to [5], DCs consumed 1.5% of total U.S. electricity consumption and created as many emissions as a mid-sized nation like Argentina. In 2013, approximate 34 power plants in the US produced 91 billion kilowatt hours of electricity to satisfy the energy consumed by DCs. Also, it is estimated that electricity consumption in DC will increase by 53% by 2020 to 140 billion kWh, which will be the equivalent amount of energy generated by 50 coal-fired power plants. Thus it will cause nearly 150 million metric tons of carbon pollution [6]. According to the report published in 2013 [7][8], the IT sector, which includes DCs, is responsible for consuming about 10% of the world's electricity generation. These data show that (DCs) are one of the major energy consumers and a significant source of CO2 emissions globally. The EU's climate strategies and targets, reducing 80-95% greenhouse gas emissions by 2050, have steered all sectors such as power, transport, industry, etc. to contribute to the low carbon transition as to their technical and economic potential [9].

DCs can make a significant contribution to worldwide energy efficiency by reducing the energy consumption of IT ecosystems because it is an important player in the energy consumption industry.

This is one of the most important reasons why most researchers focus on reducing DC power consumption [7], [10] - [15].The optimal distribution of workloads between the servers is targeted by existing studies and commercial solutions. Most researcher generally assume that there is a linear relationship between losses and workload. So, it is assumed that the best solution for distributing workloads is to load servers up to a certain level. After that, the remaining workloads are moved to the next available server. Unfortunately, the losses are not linear. The high load may cause higher losses. The GreenDC Project will use quadratic equation model instead of a simple linear model. Also, the GreenDC takes into account an entire system which includes servers, cooling system, backup power and electrical distribution system from a different perspective than current studies.

The main purpose of the GreenDC Project is to develop a decision support tool, GreenDC DSS that gives operation options to data centre managers to minimize energy waste and CO2 emissions. GreenDC DSS is based on dynamic simulation model and nonlinear energy forecasting model. In the GreenDC project, implemented through the exchange of information between two academic partners and three industrial partners, knowledge of dynamic simulation and non-linear energy demand forecasting are transferred academic partners to industrial partners while industrial partners transfer their knowledge on data centre operations. The GreenDC Project is progressing by secondment activities. The main outcome of the GREENDC activities are expected to reduce CO2 emissions and energy waste due to non-optimised energy load balancing from large number of data centres across Europe.

The remaining of this paper is structured as follows: Section 2 describes the main architecture of the GreenDC DSS. In this section, each layer is described in detailed. In section 3, daily life working condition of a data centre was enacted by user personas and scenarios.

2.GREENDC DSS ARCHITECTURE

The GreenDC DSS has data, business logic, user interface and math model layers as seen in Figure 1. Each layer provides interfaces for upper layer to make them independent of each other.

Data layer has two sublayers: data collection layer and data normalization layer. Data collection layer includes components that interact with devices or third-party tools that generate or provide data from DC devices including IT and non-IT facilities. Data normalization layer components convert the collected data into data model of the GreenDC DSS .

Math Model layer will contain components that conduct estimation and optimisation of energy use of DCs. Estimation component will provide services related to estimating energy consumption of devices and DC as a whole. Optimisation component will provide services related to optimising the operation of DCs by manipulating intervention mechanisms including optimal temperature for a given predicted workloads, number of VMs to be shut down for a given time period, the duration and which workloads to be postponed in case of high workloads, and identifying the best load balancing mechanisms in different circumstances.

The business logic layer contains components that will deal with requests coming from user interface layer components. The components include user manager, DC Monitoring, and DC Simulator. User manager component will deal with authorization and authentication of users as well as user profile management. DC Monitoring component will provide data on the DC usage for a given period including energy consumption, workloads, the status of facilities (IT and non-IT), and management reports. Finally, DC Simulator will allow users conduct diverse simulations to test the effectiveness of different intervention mechanisms.

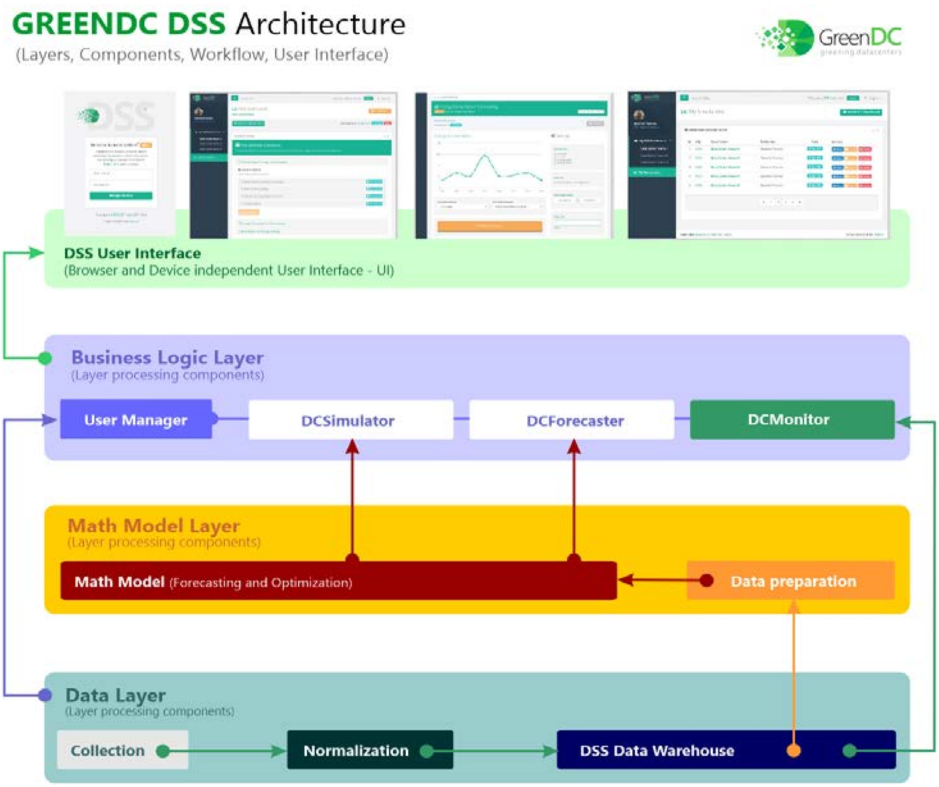

Figure 1. GREENDC DSS Architecture

The user interface layer includes user interface components for easy and efficient interaction with users. The user interface components will be designed to minimise users' interactions with the system. The components in the user interface layer will use services provided by the business logic layer components.

2.1. User Interface Layer

The interface of the DSS platform solution is based on contemporary WEB standards for user usability, device independency and lightweight programming. The main intention of the user interface (UI) designers are to keep its simplicity and consistency, delivering maximum possible results to end-user, with minimum interface interactions.

The panel represents basic interaction for:

- Users profile info
- User role info
- Profile management
- Log-out
- Quick Access to many DC's **Dashboards** (based on user access rights)
- Quick Access to **Simulations archive** performed by specific user
- Hidden on mobile devices

2.2 Data Layer

Data Layer is split into two segments above the physical segments: Data Collection Layer, Data Normalization Layer

- Data Collection Layer

In Data collection layer, the data is provided upon request payload parameters. Collection interface communicates physical devices and collect consumption data from data center' resources by given parameters. Data Collection layer collects data from variant devices such as UPS, Air Conditioners, IT devices(switches, firewalls, LB, etc.) and also gathers data via 3rd party software APIs such Zabbix or

DCIM softwares. Serialize interface serializes the data which means verify and correct the data to store into database as much as pure, clean and secure.

- Data Normalization Layer

In data normalization layer, The data is converted available for any authorised request. Therefore, it is needed to implement a generic-data normaliser and publish it into an API or SOA object. It is not needed sophisticated and complicated methods or entities in this architecture because there is an opinion to store the data in a schema-free structured data in NoSQL datastore.By this layer, There is an ORM(Object Relational Mapping) library such as Hibernate(Java) or SqlAlchemy(Python) or Doctrine(Php). This library helps us to manage connection operations like open-close connections; execute SQL statements, handle result-sets retrieved from database; map object models(classes) with database entities.

2.3.Business logic layer

There are three Interfaces to serve for the requests coming from user interface layer components.

- The UserManager Interface

In this interface, operations such as registering new users or deleting existing user, if there is a registered user on the system, checking the entered user id and password, controlling which services a registered user can access are performed.

- The MonitorDC Interface

Monitoring energy consumption of data centres is performed in this interface. In order to prepare server capacity and required energy sources and cope with the change of the energy consumption in DC, workloads and energy consumption for a given list of devices (or entire DC) are forecasted. Information about forecasting are monitored and returned as a list in this interface.

- SimulateDC Interface

SimulateDC interface allow data centre managers to test impacts of different interventions for saving energy by DC devices.

2.4. Math Model Layer

Math Model Layer is composed of two interfaces as EstimationModel and OptimisationModel. The historical data which include device type, number of device, device settings, room temperature, humidity etc. are provided by Data Layer to Data preparation section. Data preparation for math layer section provide the required information to EstimationModel and OptimizationModel Interfaces.

- EstimationModel Interface

In this interface, the required information, is used forecasting, about data centre configuration is provided by getting the information of devices concerning the historical data such as device type, number of the device, device settings etc. and the information about data with reference to devices and the other measured data (i.e. room temperature and humidity) from Data Layer. This information helps to specify historical data type to define the forecasting parameters. Data type of IT devices (i.e. CPU usage, CPU temperature, network traffic etc.) and the related other parameters such as temperature, day of week and time will be considered for defining. The power consumption of IT devices, cooling and power systems will be evaluated separately. Parameters will be defined related with each part. The historical data which is taken from Data Layer, consists of the regarding data types of the forecast workloads or the energy consumption of devices and data centre.

- OptimisationModel Interface

The information to optimise data centre energy cost is evaluated in The OptimisationModel interface. At first, the information about device and data settings are provided. Then the Optimization parameters are determined. After that, the estimated data related to the Optimization parameters are taken from EstimationModel Interface. At the end of the process, the Optimization results are presented.

Required info on parameters varies by working condition which is selected by data center manager. For example, cooling set temperature info, humidity info, outside temperature info can be given by

OptimizationModel Interface to learn the impact of shutting down VMs in order to find an optimal option to minimise energy consumption.

3. IMPLEMENTATION OF GREENDC DSS ACCORDING TO USER SCENARIO

This section contains of system scenarios by taking into account the user requirements of GREENDC DSS. In order to compose the system scenarios, roles and personas are defined.

Roles combine a set of typical actions that will fulfill tasks in scenarios. The role concept typecasts users in different categories. However, the implementation is supposed to make all tools easy to use for a better usability of the GREENDC DSS.

- **Group Leader:** The very first responsibility for a group leader starts from the energy consumption forecast and he needs to assign proper actions and tasks to group members based on the forecasting result. GreenDC DSS provides the long-term and short-term perspectives on the energy consumption to the group leader so that he/she can cope with any circumstances. Also, GreenDC DSS support the simulation for the energy consumption of data centre so that a leader can choose appropriate intervention option with the support of group members.
- **Group Member:** Group members usually conduct overall management action regarding the control of energy consumption level based on the supervision of a leader. The fundamental role of the members is to monitor the level of energy consumption.

Using scenarios gives the personas defined roles for a current situation. A persona illustrates the reality of life with a multi-dimensional view of the possible men /women behind a role in GREENDC DSS. The persona concept facilitates system architects to meet the requirements of this analysis to put oneself in the position of possible system users with their diverse social backgrounds.

Johan Green (Group Leader)

- He has to make sure that every group member does his/her own duties right on time and complete as defined by the regulations.
- He makes an assignment and coordinates the group member duties.

Melina Silverwood (Group Member)

- Manage and distribute all IP addresses which belong to XX Data centre.
- Manage and control all internal IP blocks which are used in Data Centre LAN and E-government network
- Manage and control all active network devices and infrastructure of Data Centre and E-government network
- Take and manage backups of all active network devices and infrastructure of Data Centre and E-government network
- Manage and configure VPN and other network configurations between/inside E-government services, Data Centre regional branches and headend, call centres, DRC(Disaster Recovery Centre), TV and Satellite broadcast units.
- Design system (network and infrastructure) architecture for required governmental agencies and manage configured system(s).

3.1. The Monitoring and Reporting of Energy Consumption

Melina Silverwood is a Group Member of XX Data Centre. Melina's main task is to ensure that data center operations are performed in an efficient and timely manner by monitoring the energy consumption in the data center in real time. Melina monitors the real-time monitoring control panel in GreenDC DSS, showing the interface merged with the DCIM tool (data center infrastructure management tool) in every day. Melina checks the real-time line chart presenting energy consumption of every minute. Melina checks whether the daily consumption summary of today's energy consumption is consistent with the previous few days.

The management team of the XX data center holds a regular weekly meeting and evaluates the weekly energy consumption report, which is one of the most important materials for in-depth analysis of the data center maintenance. To get the periodic report, Melina or John open the Periodic Reporting dashboard in the GREENDC DSS. It offers daily, weekly, monthly, and quarterly reporting including energy consumption stats, major events/intervention, and future consumption trend report. The reporting results also can be exported into doc or pdf file format.

3.2 Energy Consumption Forecasting

John Green, Group Leader, has a lot of concerns as the energy consumption of data centre has been incredible increased since last few years. Furthermore, in terms of long-term trend, the increasing digitization in all areas of the economy and society is resulting in an increasing need for processing power. Based on his experience, he is also recognizing seasonal trend (short-term) of energy consumption. To build-up the plan to cope with this micro and macro change of energy consumption, John Green wants to estimate the future energy consumption for long-term and short-term perspective.

John Green chooses the “Energy Consumption Forecasting” menu from GREEN DC DSS and the dashboard for consumption forecasting is shown to him. The dashboard shows the daily/weekly/monthly-based energy consumption forecasting using the forecasting model that has been fitted through the whole historical data of data centre. The forecasting result says the consumption will increase so John tries to tune some variables in the forecasting model through the simulation interface to see how these variables can affect to the future consumption.

3.3. Skeleton of the GreenDC DSS Simulator for Energy Saving

Whilst Melina monitors the energy consumption in real-time, she realizes that the current consumption is higher than the same time in previous few days. To sort out this situation, Melina discusses with John Green, data centre manager, regarding intervention options. To choose one of the options, they conduct the simulation of each possible intervention option. All these actions can be done through the Simulation Dashboard that supports the energy consumption simulation and real-time intervention interface.

- **Simulation for Controlling the Variable in the Energy Consumption Model**

GREENDC DSS embeds some energy consumption models and tunes their parameters using historical data. One directive way to minimise the energy consumption is to adjust some tuneable variables in the energy consumption model. Melina can choose one of embedded energy consumption model to verify the main factor cause higher consumption. Then she can check tune some variables that can be directly adjustable from DSS level. Also, Melina can compare the performance of energy consumption model among embedded ones as DSS shows the performance comparison tables among existing energy consumption models. After this comparison, she can apply the variable into real-time service level.

Different parameters for the operation of data centres are used and the optimization of settings is critical for the overall energy consumption and meeting service level agreement. For example, the temperature has bidirectional impacts between the cooling system and servers. Keeping a DC in the higher temperature will make the energy consumption by cooling system less while the performance of servers worse therefore more energy consumption. How much positive impacts for the cooling system and negative impacts for servers are now known and finding the optimal temperature for a given forecasted workloads.

- **Simulation for the impact of shutting down VMs**

XX data centre management team know that how to effectively balance the power load in data centres is an issue that every data centre manager is familiar with. When done correctly, a properly balanced data centre helps to secure uptime and is often an important avenue for the facility to utilize extra power capacity. When improperly balanced, available power can become stranded, and the chance of damage to vital infrastructure increases. Taking the time to optimize power distribution when installing or refitting a data centre is well worth the effort and is another crucial step toward maximizing its performance.

To help avoid stranding power, Melina can conduct simulation for shutting down the idle VMs to reach the optimal balanced status through the Load Balancing interface in Simulation Dashboard. If the simulation result shows the expected level of consumption, then she can push the “Apply System” button with the setting used for simulation.

- **Simulation for the impact of different load balancing tools**

How to use servers and VMs within DC affects the overall energy consumption as overloaded devices are expected to consume energy unnecessarily. Different load balancing tools are used to have even usage of servers and VMs for a high level of workloads. However, the performance of load balancing tools is different for different tools therefore data centre managers would like to know what are the net amount saved (or wasted) for using alternative load balancing tools.

- **Simulation for the impact of postponing workloads**

Different workloads have different priorities depending on who the requests come from and when they are created. Therefore, postponing workloads with low priority can prevent overloads of servers and VMs and save energy consumption. Melina would be able to set the rules for delaying certain workloads. This setting includes the level of priority, the duration of postponing, in certain case designate such workloads to certain VMs or servers. Melina then checks how much energy was saved and the impact to overall service level agreement.

- **Simulation for Setting the Network Configuration**

The last option for Melina is to set the network configuration again. She can open the network configuration menu on the Simulation Dashboard. The menu presents the graphic user interface with network structure and enables her to configure the parameter of each network node to minimise the energy consumption so that she can check the expected energy consumption according to the network configuration setting. If the simulation result shows the expected level of consumption, then she can push the “Apply System” button with the setting used for simulation.

5.CONCLUSIONS

The GreenDC project, funded by EU H2020 Marie Skłodowska-Curie, aims to develop a simulator for data center managers to make decisions to reduce energy consumption and CO2 emissions through predicting energy demand. In this paper, the architecture of the simulator named GreenDC DSS is explained. The skeleton of an optimization and estimation tool of GreenDC DSS which is developed to create optimum working conditions to provide energy efficiency is shown.

The architecture of the GreenDC DSS has 4 layers: data, math model, business logic and user interface layers. In Data Layer, the Data Collection interface collects data from variant devices though communicating the physical devices such as UPS, air conditioners, IT devices(switches, firewalls, etc.) then the collected data is normalized for any authorized request. In Math Model Layer, firstly the historical data for estimationModel Interface and OptimisationModel interface are provided. In EstimationModel Interface, the power consumption of IT devices, cooling and power systems will be forecasted by using historical data. After that, the optimum working conditions are presented upon optimization results. In Business logic layer, basic parameters and the main operation processes to manage GreenDC DSS simulator are performed. Finally, in user interface layer, a well-designed user-friendly panel is developed as DSS platform interface for end-users. That panel contains of user profile and role info, log-out menu, dashboards access and simulations results etc.

Also, the processes of daily life working condition of GreenDC DSS to offer optimum working condition to data center manager and group member are performed.

ACKNOWLEDGEMENTS

This work has been funded by the GREENDC project from the European Union's Horizon 2020 research and innovation programme under grant agreement No 734273.

REFERENCES

[1] L. Rao, X. Liu, L. Xie, W. Liu, R. Lei, L. Xue, X. Le, and L. Wenyu, “Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment,” *Proc. - IEEE INFOCOM*, pp. 1–9, 2010.

[2] J. Li, Z. Bao, Z. Li, and S. Member, “Modeling demand response capability by internet data centers processing batch computing jobs,” *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 737–747, 2015.

[3] R. Buyya, C. Vecchiola, and S. T. Selvi, "*Mastering Cloud Computing: Foundations and Applications Programming", 1st edition*. 2013.

[4] W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, and P. Demeester, “Trends in worldwide ICT electricity consumption from 2007 to 2012,” *Comput. Commun.*, vol. 50, pp. 64–76, 2014.

[5] V. Mathew, R. K. Sitaraman, and P. Shenoy, “Energy-aware load balancing in content delivery networks,” *2012 Proc. IEEE INFOCOM*, pp. 954–962, 2012.

[6] P. Delforge, “America ’ s Data Centers Are Wasting Huge Amounts of Energy,” *Nat. Resour. Def. Counc.*, vol. IB:14-08-A, no. August, pp. 1–5, 2014.

[7] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, “Cutting the Electric Bill for Internet-Scale Systems,” *Sigcomm 2009*, pp. 123–134, 2009.

[8] M. P. Mills, "The cloud begins with coal", . Technical report, Digital Power Group, 2013.

[9] D. Papers, S. R. K. Neuhoff and S. Schwenen, and D. Papers, “Power Market Design beyond 2020: Time to Revisit Key Elements?,” 2015.

[10] Y. Yao, L. Huang, A. Sharma, L. Golubchik, M. Neely, Y. Yuan, H. Longbo, A. Sharma, L. Golubchik, M. Neely, Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely, “Data centers power reduction: A two time scale approach for delay tolerant workloads,” *Proc. - IEEE INFOCOM*, pp. 1431–1439, 2012.

[11] J. Yao, X. Liu, W. He, and A. Rahman, “Dynamic control of electricity cost with power demand smoothing and peak shaving for distributed internet data centers,” *Proc. - Int. Conf. Distrib. Comput. Syst.*, pp. 416–424, 2012.

[12] A. Rahman, X. Liu, and F. Kong, “A Survey on Geographic Load Balancing Based Data Center Power Management in the Smart Grid Environment,” *IEEE Commun. Surv. Tutorials*, vol. 16, no. 1, pp. 214–233, 2014.

[13] A. Gandhi, M. Harchol-Balter, and M. A. Kozuch, “Are sleep states effective in data centers?,” *2012 Int. Green Comput. Conf. IGCC 2012*, pp. 1–10, 2012.

[14] R. Stanojevic and R. Shorten, “Distributed dynamic speed scaling,” *Proc. - IEEE INFOCOM*, vol. 2006, pp. 1–5, 2010.

[15] J. L. J. Liu, F. Z. F. Zhao, X. L. X. Liu, and W. H. W. He, “Challenges Towards Elastic Power Management in Internet Data Centers,” *2009 29th IEEE Int. Conf. Distrib. Comput. Syst. Work.*, 2009.

View publication stats