Renewable Energy-Aware Data Centre Operations for Smart Cities – the DC4Cities Approach

Preparation of Camera-Ready Contributions to SCITEPRESS Proceedings

Sonja Klingert¹, Florian Niedermei², Corentin Dupont³, Giovanni Giuliani⁴, Thomas Schulze¹, and Hermann de Meer²

¹Software Engineering Group, University of Mannheim, Germany
²Computer Networking and Computer Communications Group, University of Passau, Germany
³Create-Net, Italy
⁴HP Italy Innovation Center, Italy
/klingert, schulze]@informatik.uni-mannheim.de, {niederme, demeer}@fim.uni-passau.de
³corentin.dupont@create-net.org, ⁴giuliani@hp.com

Keywords: Data Center, Energy-Aware, Renewable Energy Source, Smart Cities, Workload Scheduling

Abstract: Data centres are important players in smart cities both as IT service providers and as energy consumers. Integrating intermittent renewable energy sources into the local power grid is one challenge in future smart cities aiming at an IT based low carbon economy. The project DC4Cities takes up this challenge by offering a both technical and business related solution for optimizing the share of local renewable power sources when operating data centres in smart cities. To this end, power management options between the data centre and the smart city together with internal adaptation strategies for data centres are introduced. Finally, an implementation of the suggested approach is presented and evaluated in a simulation.

1 INTRODUCTION

In 1516 Sir Thomas More published his book “Utopia”, a vision of a perfect society which became the blueprint of all future visions for an ideal society. Today, facing the risk of climate change processes we cannot control, we are again badly in need of visions.

In this paper we offer a vision with regards to an eco-friendly data centre (DC) management in smart cities. The challenge is a society that runs on big data, especially in smart cities whose organisation is based on sensor networks along the dimensions of smart governance, smart economy, smart mobility, smart environment, smart people, and smart living. All data created by these sensors need to be collected, processed and put into effect by DCs, most of them within the geographical boundaries of the smart city. However, in order to maintain a level of CO2 emissions which is in line with the EU’s emission goals of 2020, 2030, and beyond (EC, 2009), this data management needs to be done with as little impact on the smart city’s CO2 balance as possible.

The general setting for such a vision is to enable a near real-time communication between the smart city, all DCs participating in this scheme and specific customers of the DCs. The smart city can be represented by a so-called Energy Management Authority of the Smart City (EMA-SC) which acts as a mediator between the DCs and the energy system. As such, it overviews and analyses the current state of the power grid and the availability of renewable energy resources. From these data and under some other constraints the DC then computes an ideal power plan. Depending on how accurately this ideal power plan is implemented, the DC is either rewarded or might need to pay a fine.

This paper gives more insight into the DC4Cities research and associated prototype that implements it. Section 2 will give a more detailed overview on the general research idea, section 3 will introduce the architecture of the DC4Cities implementation, and section 4 will present first simulation results. Finally, section 5 will position our approach in relation to others’ work.
2 GENERAL APPROACH

Operating a DC at a level of 100% of renewable energy is not a problem if its energy provider is only using hydro or geothermal sources. To save CO\textsubscript{2} emissions caused by DCs one might suggest to move them to places offering corresponding infrastructures like Norway. Even though large DC owners like Google are considering this as a solution for parts of their DCs, it is not practicable in all scenarios. Some DCs need to be close to users and inside cities for manifold reasons: Companies often see security risks in outsourcing data or other IT services, especially to foreign countries. Network latency is another aspect that makes moving services to distant locations unfeasible (e.g. near real-time stock exchange services). So, if moving DCs to locations with a high availability of renewable energy is not an option, DCs need to become more energy aware, energy efficient, and energy adaptive. However, running a DC at high levels of renewable energy sources in the city is a great challenge. A main reason for this is the lack of availability of locally produced renewable energy due to space limitations.

![Diagram: DC4Cities Impact on Renewable Energy Utilization](image1)

In order to tackle this problem, DCs need to try to adapt better to the availability of renewable energy, minimize their energy consumption for specific tasks in general and adhere to constraints of a higher directive which is managing energy aspects in a smart city – the aforementioned EMA-SC. Figure 1 schematically shows a renewable-energy optimized and non-optimized DC. On the left hand side, the energy consumption of a traditional, i.e. non-optimized DC is shown, which clearly does not correspond to the generation curve of renewable energy. The right hand side depicts the energy profile of the same DC after DC4Cities has been applied. Here the production and consumption curves are much more alike, and therefore the amount of renewable energy used by the DC is much higher. In order to reach this goal, an active coordination between EMA-SC and a DC is mandatory for setting reasonable goals as well as technique for controlling energy adaptation within the DC.

The EMA-SC is responsible for the energy coordination within a smart city and the communication with DCs. It monitors the energy consumption and computes the desired energy consumption both for the smart city as a whole and individual DCs, and it sets certain targets for DCs and other large consumers inside the city. In case some of the targets cannot be reached, it also tries to resolve such conflicts by facilitating the negotiation of objectives with the respective consumers’ organizations. In the case of a DC such an escalation could for example lead to workload relocation within a federation of DCs having different energy sources or simply the payment of a fine.

For the DC internal adaptation, two objectives are pursued: on the one hand re-organizing workload (and thus the energy consumption profile) in order to match the shape of the renewable supply curve, and on the other hand minimizing energy consumption in order to meet the current level of renewable supply.

![Diagram: DC4Cities High-Level Architecture](image2)

The DC4Cities controller (in the middle of figure 2) will retrieve forecasts about the energy source mix directly from the providers or indirectly through forecasting models. This is done through the “Renewable Energy Adaptive Interface”. For example for local solar panels, we start from weather forecasting data.

This information will be processed periodically (e.g. every 15 minutes) and merged with the power objectives of the DC in order to compute the ideal power plan for the DC (see section 2.1). Inside the DC, this “energy budget” will then be split onto the different services; for each service a so-called “Energy Adaptive Software Controller” (EASC) will use DC specific monitoring and automation tools to correctly schedule and tune the SW/HW resource usage of this service in line with the directives received by the controller through the “Energy
Adaptive DC Operation Interface”. The expected result is that the actual power consumption of the whole DC will be considerably closer to the previously computed ideal power plan, thus meeting the DC power objectives received by the Smart City.

2.1 Coordination between Smart City and Data Centre

Coordination between the smart city and DCs under its energy management scheme is managed by EMA-SC. In our scenario we assume EMA-SC to be located inside the smart city (a stand-alone solution is also possible). We also assume that the smart city has certain goals regarding the share of renewable energy in the energy mix of its big consumers. These goals are quite high-level and therefore hardly immediately technically enforceable. However, by subdividing these high-level goals into more concrete objectives, the smart city goals can be translated into constraints regarding power/energy usage. Once these constraints are in place and information on future renewable energy availability has been obtained, EMA-SC is able to calculate power budgets for the large consumers inside the city. To this end, the DC4Cities software includes a component named Ideal Power Planner (IPP). The IPP is responsible for transforming renewable energy forecasts and power/energy usage constraints imposed by EMA into a power plan for a DC. The IPP executes the following steps:

1. Calculate the total amount of renewable power available to the DC from the renewable energy availability forecast(s)
2. Using information on the minimum (DCPowerMIN) and maximum (DCPowerMAX) power demand of the DC, scale the values obtained in step 1 so that the difference of the maximum and minimum corresponds to the difference of DCPowerMAX and DCPowerMIN
3. Shift the result of step 2 so that the minimum aligns with DCPowerMIN
4. Apply any constraints from SC side, e.g., power and energy constraints

These power plans are then communicated to the DCs, which will follow them as closely as possible. While the basic compliance of power/energy objectives has to be secured via trial runs at the configuration time of the software system, everyday changes in energy supply may lead to a non-compliance regarding objectives from DC side. To handle this case, DC4Cities includes an escalation mechanism.

A DC will trigger an escalation in case it cannot comply with objectives by means of internal adaptation. In case of such an escalation, EMA-SC may seek further external solutions such as suggesting and supporting a federation with other DCs or updating/adding objectives. As these escalations are exceptional cases, they can also be handled in a face-to-face manner between the DC and EMA-SC. Escalation may be triggered manually or automatically in case no DC operation compliant with the IPP can be calculated or if a threshold agreed on in the contract is violated. Depending on the time until the occurrence of the predicted non-compliance, two stages of escalation are envisioned: A ‘warning’ which marks a possible exceeding of thresholds in the future (e.g., > 6 hours), and an ‘alarm’ (e.g., < 6 hours). Possible reactions of EMA-SC depend on the underlying RenEnergy contract with the DC (see section 2.3). The following de-escalations are envisioned:

- Tolerate violation of objectives because of unforeseen temporary events either at DC side or originating from energy side
- The RenEnergy contract might allow a certain number and frequency of violations.
- Incur penalty according to an agreed reward/punishment scheme
- Change objectives (temporarily/permanently)
- Set temporary objectives in order to prevent decrease in goal achievement
- Offer federation assistance

Obviously, the occurrence of an escalation depends on the metrics used by the DC4Cities system in order to assess the “fitness” of the DC with respect to its goals. The metrics are for instance used to measure the distance between the ideal power plan and the actual power utilization of the DC. Based on the distance and a threshold value, escalation events might be determined.

2.2 Energy Adaptation within a Data Centre

Following the communication with EMA-SC, the compliance with its goals needs to be implemented inside the DC.
The novelty of DC4Cities is to propose a multilevel API able to allow each level of a modern DC to follow energy directives (see Figure 4). There are three levels: IaaS, PaaS and individual applications.

Firstly, we need to reschedule some tasks performed in the DC to match the hours where renewable energy is available. This scheduling requires detailed knowledge about the applications running in the DC and should therefore be tackled at the level of PaaS, where we have the knowledge of the applications running. Additionally, the PaaS layer provides a uniform interface to interact with the applications, and to scale them up and down.

The secondary objective of DC4Cities is to save energy in the DC. Previous projects such as FIT4Green\(^1\) showed that it is most efficient to use the IaaS layer in order to consolidate virtual machines (VMs) on the most efficient servers, and then switch off the unused servers.

Thirdly, we need to control specific applications in the DC, such as maintenance jobs. Applications such as virus scan or database compression are ideal candidates because they perform regular tasks that can be rescheduled if needed.

\[\text{Figure 4: Components of the DC Subsystem.}\]

### 2.2.1 IaaS Optimization

At IaaS level, we adapt key parameters of an existing VM scheduler, such as the consolidation factor and the VM migration frequency. This allows the VM scheduler to follow energy directives provided by the EMA, while, of course still respecting the SLA of each VM. For instance, when there is little renewable energy available, we increase the consolidation factor. This will cause the migration of VMs and in a second step the shutdown of unused servers. If there is a high level of renewable energy available, we relax the VM scheduler constraints, so that VMs will not need to be migrated to be more consolidated. At this level the VMs are “anonymous”: we don’t know which software is running inside them nor at what time they might be shut down, since this is at the discretion of the client. A tool such as Plug4Green (Dupont et al., 2014) or OpenStack Neat (Beloglazov et al., 2014) can be used.

### 2.2.2 PaaS Optimization

A PaaS manager allows automating the deployment of applications in VMs. It is backed up by the IaaS framework running the VMs. A PaaS manager such as Cloud Foundry\(^2\) also provides a certain number of common services for all applications. The PaaS optimization in DC4Cities then consists in (1) consolidating and tuning up and down the performance of those services, (2) issuing commands to scale up and down the controlled applications. This scaling can happen in two different dimensions: either vertically or horizontally. Vertical scaling allows, for example, to allocate more CPUs or more RAM to a particular VM. Horizontal scaling of an application allows creating new VMs containing new instances of the same application. The PaaS layer is then offering us several opportunities for optimization of applications according to renewable energy availability. In particular, the common interface for horizontal scaling of applications is promising.

### 2.2.3 Application Optimization

As a third adaptation strategy, DC4Cities ultimately controls some key applications and processes of a DC. This includes virus scanning, database cleaning, but also physical server scheduled maintenance. Indeed, server maintenance is a recurring activity in a DC and it has a significant energetic impact (Soundararajan et al., 2010).

There are two kinds of applications directly controlled by DC4Cities: task-oriented and service-oriented. In the first case tasks are scheduled and performed by the applications, such as offline video transcoding or web crawling. In some cases, those tasks can be post- or pre-poned. We take advantage of this possibility to schedule tasks at the right moment (obviously compliant with the underlying contract). The DC4Cities prototype then monitors the progress of the task and its KPIs. If it is e.g. too late, it will modify its schedule to finish faster.

Service-oriented applications, on the other hand, have a continuous service to perform, such as serving web pages. In this case, we tune the performance of the application within an identified range while respecting the SLA. For example, in the

\(^{1}\) http://www.fit4green.eu

\(^{2}\) http://www.cloudfoundry.org
case of a web server, we augment/reduce the number of client threads within the boundaries of the SLA.

2.2.4 Coordination Strategies

The optimization in DC4Cities happens at two stages. First, relatively autonomous application managers (the EASCs) allow enforcing energy budgets at IaaS, PaaS and individual application levels. Indeed, a situation with fully autonomous energy-aware applications, following the same energy directives, enables an increased usage of renewables. Secondly, DC4Cities offers an additional level of coordination: the prototype communicates with each EASC to retrieve their scheduling options and can then arbitrate between them. This level of coordination is necessary for better results: the energy directives need to be adapted for each EASC, because each underlying application has a different level of flexibility. Furthermore, the execution patterns provided by each application need to be consolidated in a central system in order to make decisions. This would e.g. avoid the creation of power demand peaks.

The central system should also be able to split the energy budget granted by the EMA between the overlapping systems. Indeed, the energy allocation for the EASC-Apps is overlapping the energy allocation of the EASC-PaaS, because some applications might be operated by PaaS managers. Similarly, the energy allocation for the EASC-PaaS is overlapping the energy allocation of the EASC-IaaS, because VMs created by a scaling command on a PaaS manager are hosted by an IaaS manager.

2.3 Incentives and Monitoring

The previous sections showed that from a technical point of view it is possible to create a consistent vision on how to integrate DC energy management into a smart city relying heavily on renewable energy. However, in order for this idea to disseminate and for the vision to become a reality in future decades, this technical approach must be viable from a business point of view, both for participating DCs and the smart city.

As the first goal is not to ‘save’ energy, i.e. to reduce the amount of energy needed, but rather to rearrange the power profile of a DC, under today’s energy tariffs there is no direct power cost reduction. Here the smart city comes into play as mediator between energy system and DC – a mediator with pre-defined goals regarding the share of renewable energy at the city’s energy mix.

Ignoring the option of a ‘dictator’ smart city, the EMA-SC needs to incentivize DCs to adhere to the DC4Cities scheme. Both (real-time) dynamic energy prices and green tariffs are not applicable: dynamic prices e.g. formed by the timely interaction on the energy wholesale market, reflect the scarcity of all power in the power grid and not the availability of renewables. On the other hand, green tariffs today are offered only on certificate basis, i.e. they define “green” by renewable power being fed into the power grid anywhere and thus regularly level out the volatility of – locally produced – renewable power.

Therefore, we propose a so-called ‘RenEnergy’ contract between EMA-SC and the DC that contains both a model for which behaviour to reward or penalize and how to monitor this behaviour.

As shown in section 2.1 an ideal power plan IPP is calculated as reference power profile and power plan for the DC. The deviation between the IPP and the realized power profile of the DC is the basis for a new metric, DCAdapt, whose parameter value needs to be determined in the RenEnergy contract. Additionally for monitoring and rewarding the DC’s effort to adapt to the IPP, the new metric RenPercent represents the share of renewable energy consumption of the power profile of a DC. This is the second pillar of RenEnergy contracts.

However, the DC’s options to adapt the workload - and with it its power profile - to the IPP are limited due to contracts with its customers laid down in service level agreements, SLA. Former work by the authors suggested how to turn these into GreenSLAs that reward the customers for collaboration whenever an adaption process in the DC is necessary. These GreenSLAs are implemented in the DC4Cities system through ‘green points’ that are granted to those customers that permit the DC to shift its workload when required by the IPP.

3 ARCHITECTURE

This section presents the high-level architecture of the DC4Cities approach and provides details on the prototype that has been developed to validate it.

Figure 5: DC4Cities Architecture.
The overall technical architecture is in strict relation with the pattern described in the general approach (at the beginning of section 2) and is organized in a set of modules with clear responsibilities. The following diagram graphically represents the different modules and the interactions during the main control loop:

1. DC4cities process controller retrieves the next 24 hours energy forecasts for each electricity provider of the DC through the ERDS handler
2. The Max/Ideal power plan (see section 2.1) is computed
3. The power plan is split into different plans, one for each service hosted by the DC (see section 2.2)
4. Multiple splitting policies can be configured to better tailor the system to the DC business needs
5. The controller will request EASC to create specific power budgets for the next 24 hours for each service
6. The Option plan collector will receive a set of possible alternatives by each EASC
7. All Option plans will be consolidated and globally optimized to achieve the best usage of renewable energy source
8. If a good solution is found, the different EASC will receive the confirmation of which option plan to enact. If no solution is found, an escalation process is triggered [8x]
9. EASC will use automation tools to control the SW/HW resources of the service in line with the received plan (Working Mode).
10. Finally the controller will share the DC power plan with the energy provider, to enable some form of demand/response cooperation

Besides the external interfaces described in section 2, additional internal interfaces connect the controller to the web presentation of the dashboard (on the right side of Figure 5) and to the historical database subsystem (left) used to store all monitored data as well as to support predictions using on correlation models based on the collected data.

The prototype of DC4Cities is implemented in Java and hosted inside a Tomcat Application Server; the external interfaces are defined and supported using JSON/Rest technology.

4 TRIALS AND EXPECTATIONS

The EU project DC4Cities includes three trials to validate the system using different services inside DCs who have different power providers with various characteristics:

- CSUC and IMI; Barcelona (Spain), powered by Gas Natural; CPU intensive video conversion tasks; federation option
- CN/APSS; Trento (Italy), powered by Italian grid, located with lots of hydro production; report generation tasks for health system
- HP; Milan (Italy), powered both by local photovoltaics (PVs) and the Italian grid; test lab for a Web E-learning platform offering a world-wide service (HP LIFE, an HP Living Progress initiative)

When writing the initial version of this paper a small model was developed based on preliminary measures in Trento and HP, to provide first ideas about expected results. Now, the first phase of the trial has just been completed, and collected results are in line with these expectations. The detailed report of the trial will be published as public project deliverable D6.2 on DC4Cities project website.

The model considers a batch oriented application that needs to produce 4320 reports/day (similar to the Trento trial), initially running with power only from the Italian grid. On an average day, the renewable energy percentage varies between 29.21% and 49.18% (average 37.16%). If the report generation is spread with uniform distribution over 24 hours, obviously the average RenPercent (see section 2.3) of the power spent by the HW resources of this service is 37.16% as depicted in Figure 6.

![Figure 6: Uniform Workload distribution over 24 hours.](image)

If the computing jobs are concentrated on a few hours (running several reports in parallel) exactly at the time when the grid has the highest share of renewable energy, and if unused servers are then turned off, the RenPercent for the service becomes 42.20%, still producing the same amount of work – in terms of reports generated (Figure 7).

Adding 8 solar panels to the previous setting, each one generating a maximum of 250Wh, with a hourly distribution similar to the one measured at HP solar lab (as shown in Figure 8), and...
concentrating the computing when the highest power is provided by the PVs, without exceeding their production limits, the expected RenPercent will reach 79.41% (see Figure 9).

5 RELATED WORK

Our work is an amalgamation of three different areas: research about integrating renewable energy into smart city energy supply and distribution, also related to the integration of renewables into the smart grid. The second area is energy-aware operation of DCs via workload control, and the third concerns with energy aware contracts of and with DCs.

(Brenna et al., 2012) look into energy challenges of smart cities, including how to increase the share of distributed energy sources at the smart city’s power consumption; however, the distribution of available power on consumers is not a topic. The same applies to (Sioshansi, 2012) who deal with integrating intermittent energy sources into the smart grid, using demand response. Also here, the notion that power is a limited resource and therefore may require maximum consumption objectives is not analysed. (Liu et al. 2014) suggest energy budget optimization for households with shiftable and non-shiftable loads by scheduling the shiftable load and buying/selling energy under different information assumptions. Although the value of this work is high for DC4Cities as regarding shiftable load the setting is comparable, the contract option was preferred due to the high reliability of reaction.

In (Goiri et al., 2013), the authors propose GreenSwitch, a model-based approach for scheduling dynamically the workload and selecting the source of energy to use. The authors focus on the trade-offs involved in powering DCs with solar and/or wind energy, and propose an implementation of their solar powered mini DC. With contrast to this approach, we propose the possibility to schedule the workload more fine-grained, at application level.

In (Dupont et al., 2015), the authors present Plug4Green, an energy aware VM manager with a focus on flexibility and adaptability to new scenarios. Its main capacity is to migrate VMs and manage servers on/off state to save energy while respecting an extensive SLA. Plug4Green is re-used as a IaaS VM manager in the presented work (see section 2.2.1). A similar tool is OpenStack Neat.

Finally, the question of how to incentivize DCs in order to modify their behaviour according to the availability of renewable energy arises: It relates to...
both, to work in the area of SLA induced workload planning as well as energy aware contracts of DCs with their power providers. There is some research of the authors about how to turn SLAs into GreenSLAs in order to increase the flexibility of the DC, e.g. (Botero et al., 2013); and we build on this knowledge base to a high degree. Using SLAs as constraints to operate DC services like in (Sakellarriou et al. 2009) helped a lot understanding the problem of job scheduling but fell short of the incentive aspect. Apart from the relation between the data center and its customers, in order to make DCs be interested in following renewable power patterns, they need to have corresponding contracts with their energy service providers. There is a lot of literature about different options how to make any type of customer comply with demand response schema like in (Aalami et al., 2010, Chao, 2011). But regarding the specific also technical challenges of integrating DCs in demand response schemes, again the authors need to refer to their own work, e.g. (Basmadjian et al., 2013, Berl et al., 2013).

6 CONCLUSIONS

This paper presented a vision on how to maximize the share of renewable energy sources when operating a DC given the conditions of a smart city aiming at a local low-carbon power supply. It was shown that this is not only technically feasible but that there are also design options for the relation between the DC and the smart city which offer incentives for the DC to participate in the scheme. However, the approach is obviously dependent on the individual technical infrastructure and the real applications running in the DC as well as on financial and geographical conditions of the smart city where it is located. In order to evaluate the feasibility of the presented approach complementing trials in a real environment are necessary. These are planned for the second phase of the project.

ACKNOWLEDGEMENTS

This work was carried out within the European Project DC4Cities (FP7-ICT-2013.6.2).

REFERENCES


