by Moises Levy

DATA CENTER MODELING

Modeling data centers helps to estimate and predict key parameters under certain a-priori known conditions. This article describes a systematic, easy-to-follow approach for data center modeling using a cyber-physical systems lens. The approach is aimed at facilitating the modeling of parameters such as quality of service, airflow and power requirements, and key performance indicators, to evaluate data centers under different conditions of workload, and IT resources, and to allow comparison among different scenarios. The results help to communicate and better understand data center behavior, and to evaluate areas of improvement.

INTRODUCTION

Data centers are comprised by information technology equipment (ITE) and supporting infrastructure (e.g., power, telecommunications, environmental control, security, fire protection, and automation). Their main mission is to process and store information securely, and to provide users uninterrupted access to it.

Data centers must satisfy legislation, standards, best practices, and stringent technical requirements in order to guarantee reliability, availability, performance, security, and manage risks. They are frequently provisioned with additional resources to ensure operation under different scenarios. Data centers are evolving from traditional projects to 'software-defined data centers' (SDDC), where ITE services are delivered as a service, enabled mainly by virtualization and the commoditization of ITE. Another trend includes 'software-defined power (SDP)' strategies to deliver a power-aware optimized data centers. Stakeholders are increasingly interested in understanding and predicting data center behavior and risks under multiple scenarios. Modeling provides a quantitative explanation of tradeoffs among the different data center components.

This article describes a comprehensive data center model using a *cyber-physical systems* perspective, meaning the integration of computational and physical components, potentially including human interaction. This approach is an iterative process, since equipment may be upgraded, added, or removed. Legacy and current generation systems may be in use simultaneously throughout the data center life cycle.

The theoretical model serves as the basis to develop simulations, which may be used to predict the behavior of a hypothetical system, or one where physical measurement and experimentation is not feasible. Results help to assess strategies for end-to-end resource management and key performance indicators improvement.

DATA CENTER MODELING

The approach is comprised of three easy-to-follow steps: (1) modeling the cyber components, (2) modeling the physical components, and (3) identifying data center key indicators.

1. Modeling Cyber Components

The first step consists of identifying and modeling the components responsible for processing and scheduling the workloads. These components are defined as 'cyber' components or ITE. The *workload* is the rate and type of jobs processed by the data center as a whole, or by a given piece of ITE. The *processing rate* of the ITE represents the maximum throughput, or the maximum number of jobs that can be processed by the ITE in a unit of time.

Workload processed affects energy consumption, which directly impacts the physical environment. A data center that does not process workload will consume a fixed amount of energy to maintain the availability of all the required resources. As workload increases, power requirement increases, reaching a maximum where processing time may also rise. Parameters such as workloads and ITE specifications are used to estimate quality of service, power, airflow, energy, and key performance indicators through the model.

1.1. Quality of Service

Quality of service includes variables such as queue length, waiting time, and processing time. The power required by the ITE depends on the workload and the quality of service. A data center with poor quality of service indicators may be unsuitable for the desired purpose.

The *workload arrival rate* (W_{in}) can be interpreted as the number of jobs arriving to the data center during a given period. Let *nodes* (N) be the total number of computational nodes. Only the active nodes (n) are available at a given time for processing a workload. The *idle nodes* (n_{idle}), also called 'zombie' nodes, represent the ITE present in the data center requiring power, but which represent a cost without processing a workload.

The total workload arriving at the data center varies with time j and is represented by $W_{in,DC}(j)$. The total workload is distributed among the active nodes through a distribution factor represented by $S(i,j)$. Ideally, individual nodes process specific jobs, and if there are no jobs assigned, these nodes enter the idle mode consuming less power.

The workload evolution at an active node i at time j is shown in Figure 1 and can be described as follows in terms of *workload arrival rate* $W_{in}(i,j)$, *workload departure rate* $W_{out}(i,j)$, *and queue* $L(i,j)$:

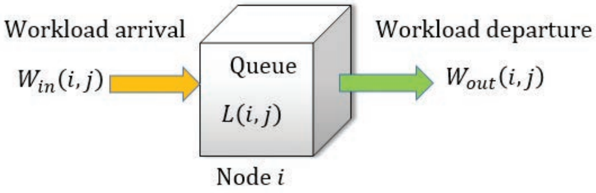

Figure 1: Workload flow at an ITE node

The *workload arrival rate* considers the total workload arriving at the data center distributed among nodes:

$$W_{in}(i,j) = W_{in,DC}(j) * S(i,j)$$

If the ITE capacity is large enough to process the incoming workload, all transactions are processed in real-time, then:

$$W_{out}(i,j) = W_{in}(i,j) + L(i,j-1)$$
$$L(i,j) = 0$$

Otherwise, if the system is receiving more workload that it can process in real-time, the system is overloaded, and a queue is formed, generating congestion for the jobs to be processed, then:

$$W_{out}(i,j) = PR(i)$$
$$L(i,j) = W_{in}(i,j) + L(i,j-1) - PR(i)$$

In summary, input workload arrives to the data center entering the queue, waits for service from a computational node, and eventually receives the service and then leaves, as shown in Figure 2. Quality of service variables are estimated permanently. The model considers open queuing and first-come-first–serve (FCFS) scheduling.

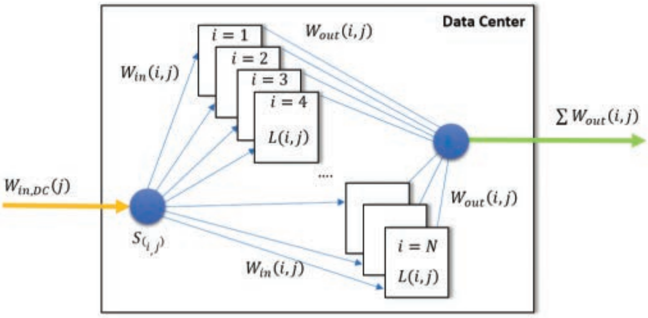

Figure 2: Workload flow at the data center

Data center analysis is concerned with the way IT resource capacity is managed over time. IT node *resource utilization* $U(i,j)$ represents the portion of ITE capacity used to process workload requests:

$$U(i,j) = \frac{W_{out}(i,j)}{PR(i)}$$

The *average and maximum utilization* considers utilization across all nodes. The idle nodes affect *average utilization* since they are not processing workloads.

The *waiting time* (or queuing time) is a measure of the workload remaining to be processed:

$$tw(i,j) = \frac{L(i,j)}{PR(i)}$$

The *average and maximum waiting time* are often related to the quality of service offered to the user. *Response time* may also be estimated, which is the total time that a workload spends in the system from arrival to service completion. Criteria may be established for workload migration, considering types of workloads, ITE requirements, migration policies, scheduling disciplines, and other factors. Further, migration of workloads to different data center sites may be considered.

1.2. Power

In general, processors and memory are the main power consumers for a data center server, followed by power supply loss, storage, PCI (peripheral component interconnect), motherboard, fans, and networking interconnects. Processor power consumption varies by processor, workload, and power management technologies. *CPU power consumption* can be estimated based on processor utilization. *Memory-related power consumption* depends on the specific technology, the idle and active states, as wells as the workloads. Processor and memory cooling are challenging, requiring thermal analysis, which includes power, airflow, temperature, and humidity.

Power supply efficiency depends on the current load. It is typically profiled with high load, which may not be realistic. In a data center, *power supply efficiency loss* is considerable, since server workloads fluctuate, and servers do not perform at full capacity most of the time. *Storage power consumption* can be significant, depending on the technology and number of drives present. Average hard disk power consumption can be estimated as the weighted sum of the power required in idle, write, read, and seek states.

Power consumption characterization for various workloads can be valuable. Studies with different servers and workloads have concluded that ITE power consumption closely follows processor utilization. The *ITE power* requirement over time can be assumed linear from idle (P_{idle}) to maximum (P_{max}) power as workload increases. *ITE power* depends on the workload, and therefore is estimated based on its utilization (U):

$$P_{ITE}(i,j) = P_{idle}(i) + (P_{max}(i) - P_{idle}(i)) * U(i,j)$$

The previous equation supports the development of energy management strategies for various ITE. The *energy consumed* $E(i,j)$ is a function of power (P_{ITE}) and time (t):

$$E(i,j) = \int P_{ITE}(i,t) * dt$$

2. Modeling the physical components

The second step consists of modeling the physical, mainly known as thermal components, including airflow and power requirements. 'Cyber' and thermal components are coupled through the energy consumption of the ITE. Thermal behavior is affected by the power required by the ITE, which depends on the workload processed and quality of service.

2.1. Airflow

ITE must satisfy operating conditions to guarantee their desired performance, reliability, and life expectancy. Airflow predictions involve thermodynamics. Most of the power drawn by ITE is dissipated as heat. In data centers, cold air from the cooling system absorbs heat generated mainly by ITE, and the warm air returns to the cooling system. The heat is then dissipated outside the facility. Figure 3 shows an airflow example, considering a data center with a raised floor and a cold/hot aisle configuration (cold air inlets of the cabinets face the same side, and hot air exhaust faces the same side).

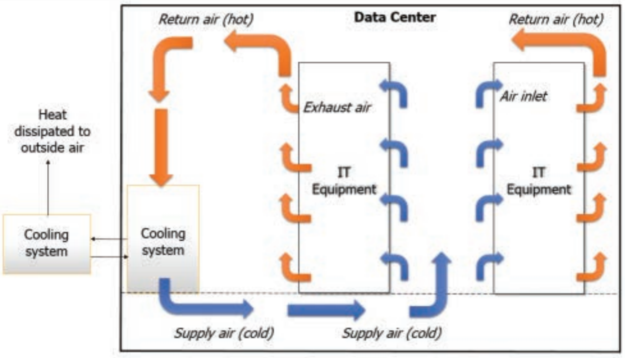

Figure 3: Example of airflow management

The cooling system inside the data center (e.g., CRAC - *computer room air conditioner-* or CRAH *-computer room air handler-*) supplies cold air to the air inlets of the ITE through the raised floor. Exhaust (hot) air is returned to the cooling system unit for further heat exchange through the outside unit (e.g., condenser or chiller plant). It is assumed that there is no mixing between cold air and hot air, or that the mixing can be neglected, as it will not affect the temperature of the return air.

The convective heat transfer at the ITE can be described as:

$$q = Cp * W * \Delta T$$

The *amount of heat transferred (q)* is estimated as the power drawn by the ITE. The *mass flow (W)* is the *airflow rate (Af)* multiplied by the *density of air (ρ)*. The *specific heat (Cp)* and *density of air (ρ)* are dependent of temperature. The *temperature rise (ΔT)* is the difference between the intake air and exhaust air temperatures. The previous equation can be expressed as:

$$P_{ITE}\,(i,j) = Cp * Af(i,j) * \rho * \Delta T(i,j)$$

Therefore, the estimated *airflow* required is:

$$Af(i,j) = \frac{P_{ITE}\,(i,j)}{Cp * \rho * \Delta T(i,j)}$$

Considering a temperature of 25*°C* (or 77*°F*) and standard atmospheric pressure (1 atm.), the specific heat *(Cp)* is *1005 Joule/ (Kg·°C)* and the density of air *(ρ)* is *1.184 Kg/m³*. The airflow (in *CFM* -cubic feet per minute-) is estimated as a constant multiplied by the ratio of the ITE power and the temperature rise (in *°C or °F*):

$$Af(i,j)_{CFM} = 1.78 * \frac{P_{ITE}\,(i,j)}{\Delta T(i,j)_{°C}} = 3.20 * \frac{P_{ITE}\,(i,j)}{\Delta T(i,j)_{°F}}$$

The formulation helps to optimize airflow management and the cooling system. The *total airflow* required must include all the nodes, active and idle. As the heat generated by ITE increases, the airflow rate may also increase to maintain temperature requirements. Most new ITE have variable-speed fans with control algorithms dependent on the utilization of the resource.

2.2. Power

The total cooling capacity of a cooling system can be expressed as the sum of sensible and latent heat removed. The cooling load in data centers is mainly sensible heat, generated by ITE, lights, and motors. Latent heat can be dismissed, as there are few people inside the data center and limited outside air. The sensible heat ratio is the ratio of sensible cooling to total cooling, which usually takes high values, ranging between 0.9 and 1. This is one of the main reasons for using precision cooling systems, designed for highly sensible heat ratios, as opposed to comfort cooling systems. In addition, precision cooling systems operate at higher airflow, satisfy strict temperature and humidity controls, and run continuously.

The *Sensible Coefficient of Performance (SCOP)* is used to estimate the power requirements of the cooling system. The *SCOP* is the ratio of net sensible cooling capacity to the power required to produce that cooling (excluding reheat and humidifier), and depends on the specifications of the cooling system and the supply air temperature. The amount of power required by the cooling system *(P_{AC})* is the ratio of total heat loads to *SCOP*. The total heat loads are the sum of all power delivered to the ITE, lighting, and electrical distribution losses, and are time-dependent.

$$P_{AC} = \frac{\Sigma P_{ITE} + \Sigma P_{Lighting} + \Sigma P_{Losses}}{SCOP}$$

As a simplification, the heat load is considered as the power delivered to the ITE:

$$P_{AC} = \frac{\Sigma P_{ITE}}{SCOP}$$

In addition, the affinity laws for fans are used to estimate the power required by similar fans in relation to airflow. The fan power requirement is proportional to the cube of the airflow supplied:

$$\frac{P_{FAN1}}{P_{FAN2}} = \left(\frac{Af_{FAN1}}{Af_{FAN2}}\right)^3$$

These estimations lead to different energy management strategies when various fans are present in the data center. Consider a data center with just one CRAH unit. If the airflow required by the ITE is reduced by half, the power required by the CRAH unit is reduced by a factor of 8. Further, the data center usually has more than one CRAH unit, and the airflow required by the ITE may be supplied by multiple units, instead of just one unit.

Data center layout, cooling system and ITE must be considered to understand airflow management requirements. Different strategies such as air economizers or free cooling may be implemented to reduce the energy consumption of the cooling system.

3. Identifying data center key indicators

The last phase consists of recognizing key indicators for the data center. Several metrics can help to examine efficiency, productivity, sustainability, operations, and risk, indicators for which the user must determine an acceptable level. For the sake of the explanation, the *Power Usage Effectiveness (PUE)* metric is used as an example of key indicator, since it is one of the most widely used energy efficiency metrics. Figure 4 shows the energy flow in a data center.

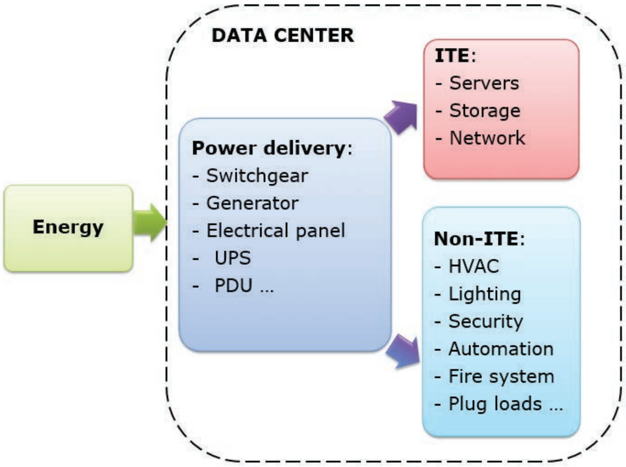

Figure 4: Data center energy flow diagram

Power Usage Effectiveness (PUE) is the ratio of energy used in a facility to energy delivered to ITE. The ideal PUE is one. In terms of the average power required by the data center (P_{DC}) and the ITE (P_{ITE}), using the same measurement period, it can be expressed as:

$$PUE = \frac{P_{DC}}{\Sigma P_{ITE}}$$

The power required to operate a data center can be simplified as the summation of the power required by the ITE and the cooling system, then:

$$PUE = \frac{\Sigma P_{ITE} + \Sigma P_{AC}}{\Sigma P_{ITE}}$$

Considering a cooling system comprising CRAC units, the PUE metric can be expressed as:

$$PUE = \frac{\Sigma P_{ITE} + \Sigma P_{CRAC}}{\Sigma P_{ITE}} = 1 + \frac{1}{SCOP}$$

where P_{CRAC} represents the power required by the cooling system, and $SCOP$ represents the sensible coefficient of performance of the CRAC unit.

Considering a cooling system comprising a chiller water plant and CRAH units, the PUE metric can be expressed as:

$$PUE = \frac{\Sigma P_{ITE} + \Sigma P_{Chiller} + \Sigma P_{Fan}}{\Sigma P_{ITE}} = 1 + \frac{1}{SCOP} + \frac{\Sigma P_{Fan}}{\Sigma P_{ITE}}$$

where $P_{Chiller}$ represents the power required by the chiller water plant, P_{Fan} represents the power required by the CRAH units, and $SCOP$ represents the sensible coefficient of performance of the chiller.

The behavior of the $SCOP$ is nonlinear and usually decreases with lower temperatures. To supply colder air, the cooling system consumes more energy. Therefore, as we increase the temperature of the air supplied by the cooling system, the $SCOP$ increases, and the PUE decreases. There are limits imposed by climate analysis, cooling system type, ITE requirements and airflow management.

CONCLUSIONS

Modeling contributes to better understanding data center behavior, tradeoffs and impacts of different scenarios. The model described, using a cyber-physical systems lens, involves a comprehensive view of a data center. Depending on each specific data center configuration and operations, the model assumptions may need to be revised. Results help to identify strategies to improve end-to-end resource management and key performance indicators.

Since right-sizing the data center based on modeling of resource performance may not be accurate, and specific needs may change rapidly, real-time monitoring of the data center may help to calibrate models. Nowadays, this is made possible given that data center monitoring and management systems collect data in real-time.