Standalone Data-Center Sizing Combating the Over-provisioning of Both the IT and Electrical Parts

M. Benaissa, J.-M. Nicod, G. Da Costa

FEMTO-ST (Besançon) & IRIT (Toulouse) – DATAZERO2 ANR PROJECT

December 13th, 2022

Data Centers and energy efficiency

Does using IT technologies have any consequences ?

- IT consumes a huge amount of energy : 56,5 TWh in France (2015)
  - annual consumption of 8,282,000 French households
  - Expected increase of +25% in 2030
  - 2.5% of France's carbon footprint

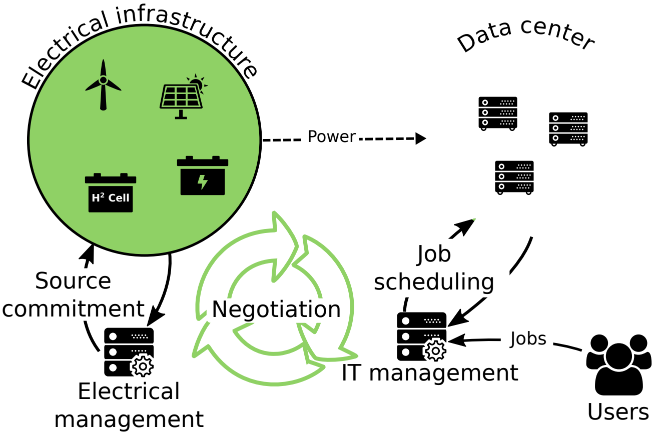
Electricity consumption of French datacenters in 2015 reached approximately 10 TWh ≈ 17.6%

Only 50 %-60 % of the datacenter ressources are used

increasing the energy efficiency of data-centers

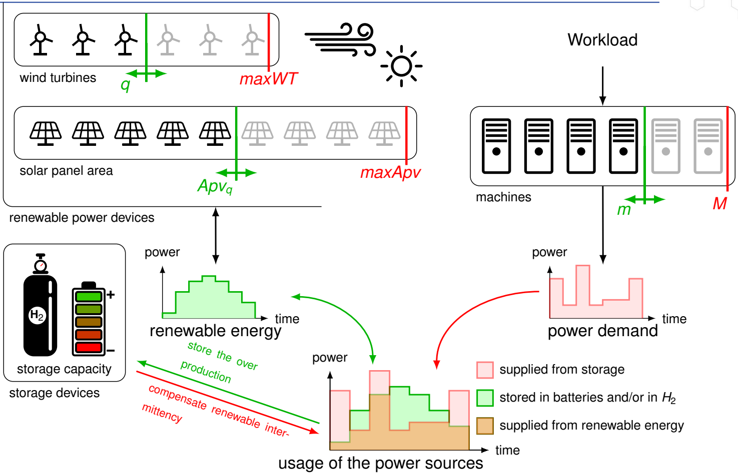
supplying data-centers with only green energy

Context



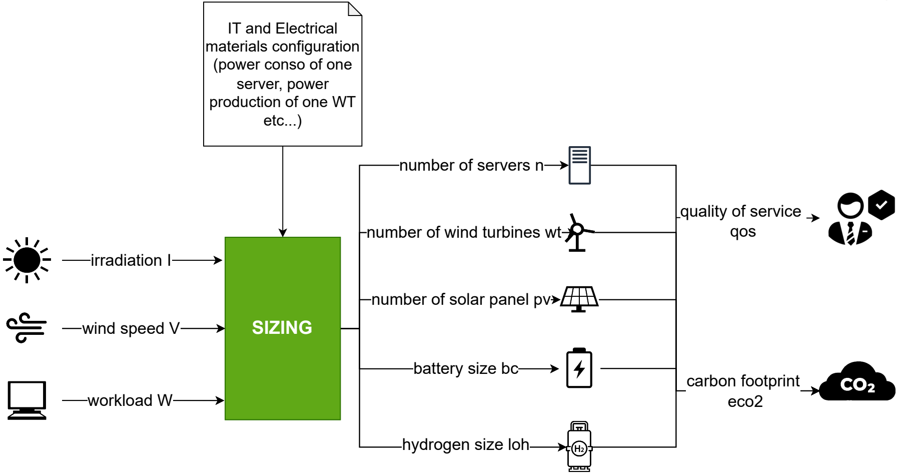
Electrical infrastructure
H² Cell
Power
Data center
Job
scheduling
Source
commitment
Negotiation
Jobs
Electrical
management
IT management
Users

Sizing Overview



outputs: configurations (m, q, Apv_q, storage capacity) + associated metrics, QoS, etc.

Input/Output



Objectives

Objectives

- The datacenter is 100% autonomous (not connected to the electrical grid), and depends only on renewable energy (solar and wind) and electrical storage (batteries and hydrogen)
- Requested Quality of Service (QoS) reached
- Electrical sizing adapted to supply the infrastructure for a given power demand + weather condition (neither under-provisioning nor over-provisioning)
- Robust (Objectives are reached even with unexpected conditions : Failure, low wind speed, request overload...)

Robustness

The sizing is based on one year of workload and weather data, that should "represent" the 30 years of datacenter life. A mean sizing of different senarios may give a robust suggestion, given a workload and a location.

Hypothesis

Hypothesis

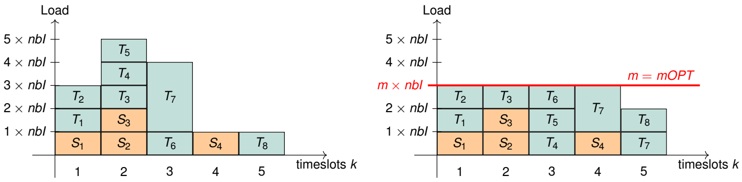
- The datacenter is 100% autonomous (not connected to the power grid) only powered by renewable energy (solar and wind) thank to energy storage (batteries and hydrogen)
- Homogeneous system (machines, wind turbines, solar panels, etc.)
- Only the main components are considered (cooling details, network structure... are not considered)
- Redundancy not considered
- Mainly Cloud, but can be adapted to HPC
- No failure, no maintenance

➠ Rather restrictive assumptions that need to be refined in this work in progress

Analysis Results

IT Sizing

This part uses EDF algorithm to schedule tasks. The number of machines m is found by binary search.



- All tasks of timeslot k are moved into a waiting queue and sorted by deadline and size. Tasks are scheduled next, until reaching the limit m. Remaining tasks are scheduled into the next timeslot, unless the deadline is reached.

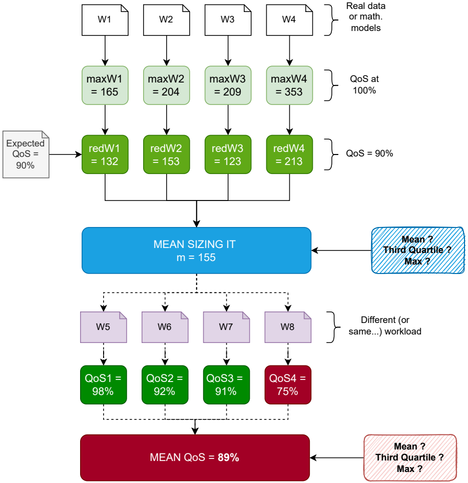
Quality of Service vs Sizing

Quality of Service

The Quality of Service (QoS) is measured by the percentage of rejected tasks : A task is rejected if it can not be scheduled without exceeding its deadline. *mOPT* is the minimum number of machines required to guarantee an optimal QoS (100% of tasks successfully scheduled).

➠ Next analysis : Decrease the number of machines and observe the impact on the QoS

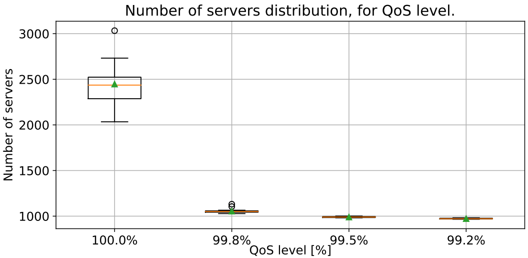
IT sizing analysis



- 50 workloads generated, following distributions law of Google trace workloads.
- For each workload, an initial sizing (100% QoS) is computed.
- For each initial sizing, the QoS is negocied (99.8%, 99.5% and 99.2%)

The first idea was to observe the distribution of each result, for each QoS, and decide then how to propose the best sizing, depending of these results.

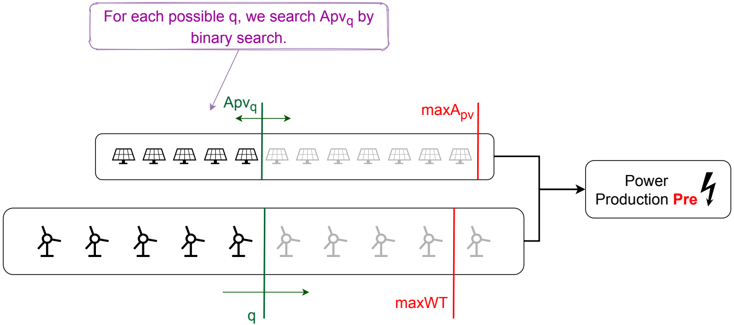
Results of QoS variation



- ... But this dispersion is drastically reduced as soon as the QoS reaches 99.8%. This phenomenon is explained by the fact that the rejections "smooth" the workload, by discarding the epiphenomena. The reduction is also remarkable because the 0.2% reduction is enough to almost halve the number of servers.

Electrical Sizing

For a given power demand (provided by the IT sizing) and a given wind speed and irradiation, gives a set of several configurations $Q = (q, Apvq, maxBCq, maxLOHq)$



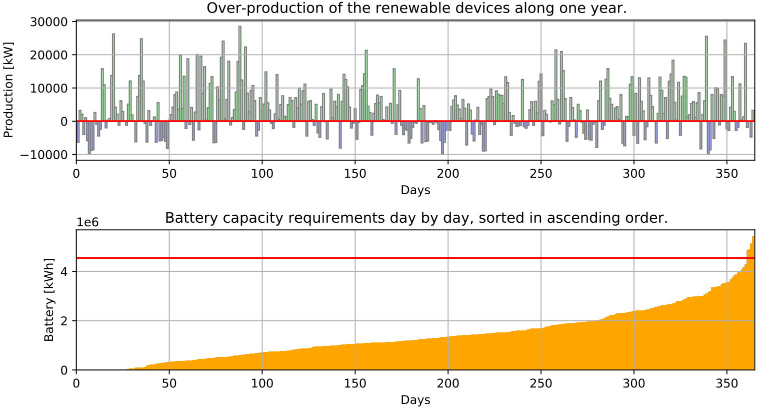
Battery reduction

Battery sizing

Batteries are used to manage the intermittency of solar energy, day by day.

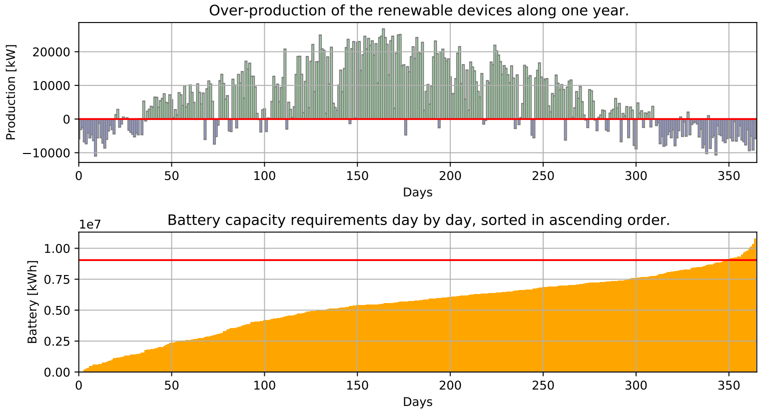
- Next analysis : Decrease the battery size and compensate the gap with solar panel.

Electrical sizing analysis



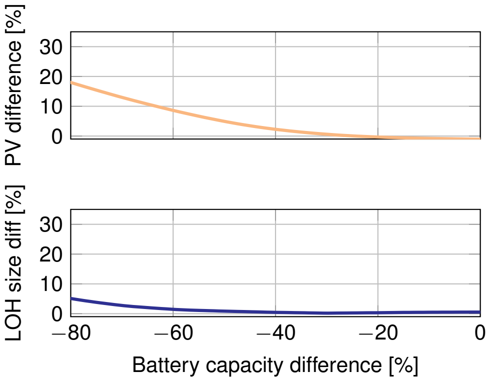
- Electrical sizing on Minneapolis (USA) - 2021, for the configuration $(WT, PV) = (4, 3277)$

Electrical sizing analysis



- Electrical sizing on Minneapolis (USA) - 2021, for the configuration $(WT, PV) = (0, 27583)$

Results of Battery reduction



- PV and H_2 evolution, depending on the battery capacity reduction. The reduction is computed on the basis of the reference configuration.

Batteries capacity reduction	-16%	-20%	-30%
Number of impacted days	18	27	68
PV increasing [%]	-0.52%	-0.3%	0.69%
LOH increasing [%]	0.39%	0.31%	0.18%

Conclusion

- For all workloads studied, a reduction in the number of servers by up to half had little impact on QoS. A resilient reduction would then be possible: Indeed, for a QoS of 95% for example, all the workloads seem to be able to cope with less than half the servers needed for a QoS of 100%.

⚠ The workloads studied follow the same jobs distribution, and that the total load remains modest. An analysis of more workloads, with different profiles, may lead to different results.

- Similarly, for electrical sizing, a clear decrease in battery capacity had little impact on the rest of the configurations, both for solar panels and hydrogen storage, due to the number of impacted days.

⚠ This gives only the intuition that lowering the battery is possible. This experiment must be tested with more inputs.

➠ In both cases, a life cycle (LCA) and a budget analysis of the infrastructure are needed.

➠ More experiments with the aim of reducing the cost of the infrastructure, in terms of economic and environmental impacts.

Thank you for your attention