Machine Learning for Energy Efficient and Carbon-Neutral Cloud Data Centers

Shashikant Ilager, Ivona Brandic
TU Wien, Austria

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

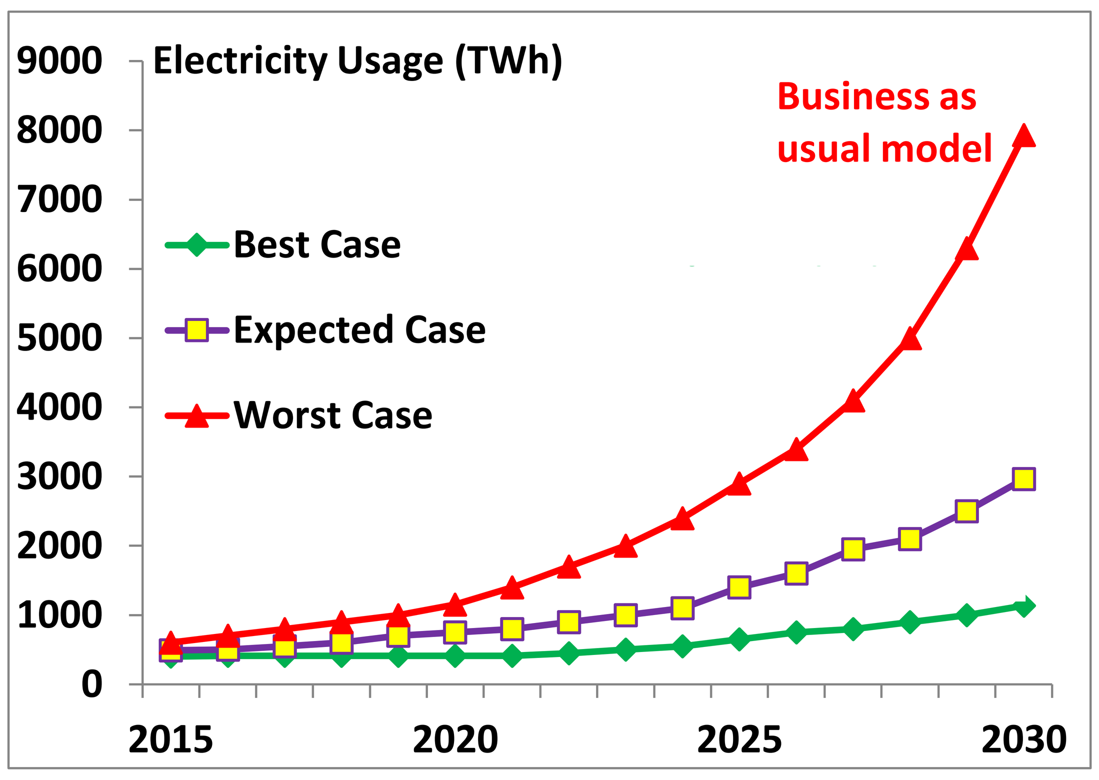
12 RESPONSIBLE CONSUMPTION AND PRODUCTION

1 Introduction

- Cloud computing
  - Allows businesses to outsource their IT facilities to third party providers
  - Avoids expensive up-front investments of establishing their own infrastructure
  - Computing as a utility service
  - On-demand delivery
  - Customers pay for what they use
  - Virtualized resources

2 Impact of Cloud Computing on Environment

- Cloud Data Centers (DCs) require energy for powering computing clusters and their cooling systems.
- Remarkably, DCs account for one-fifth of the world's electrical consumption, surpassing the airline industry[6].



Electricity Use by Data Centres

3 Objective and Challenges

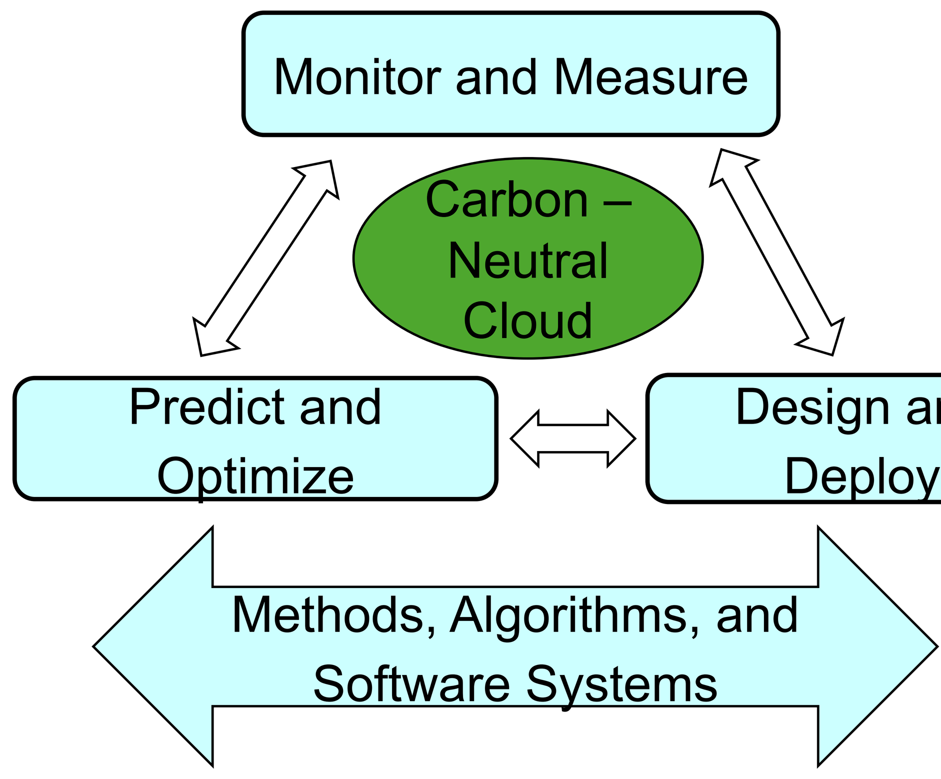
Objective: Increasing energy efficiency of Cloud DCs by managing computing and cooling systems, while satisfying user's Service Level Agreements (SLA) requirements.

- **Challenges**
  - Trade-offs between subsystems
  - Multi-tenancy and Virtualisation
  - Heterogenous workloads
  - Non-linear characteristics of resources and workloads
  - Stringent Service Level Agreements (SLAs)
  - *Complex non-linear dependencies*

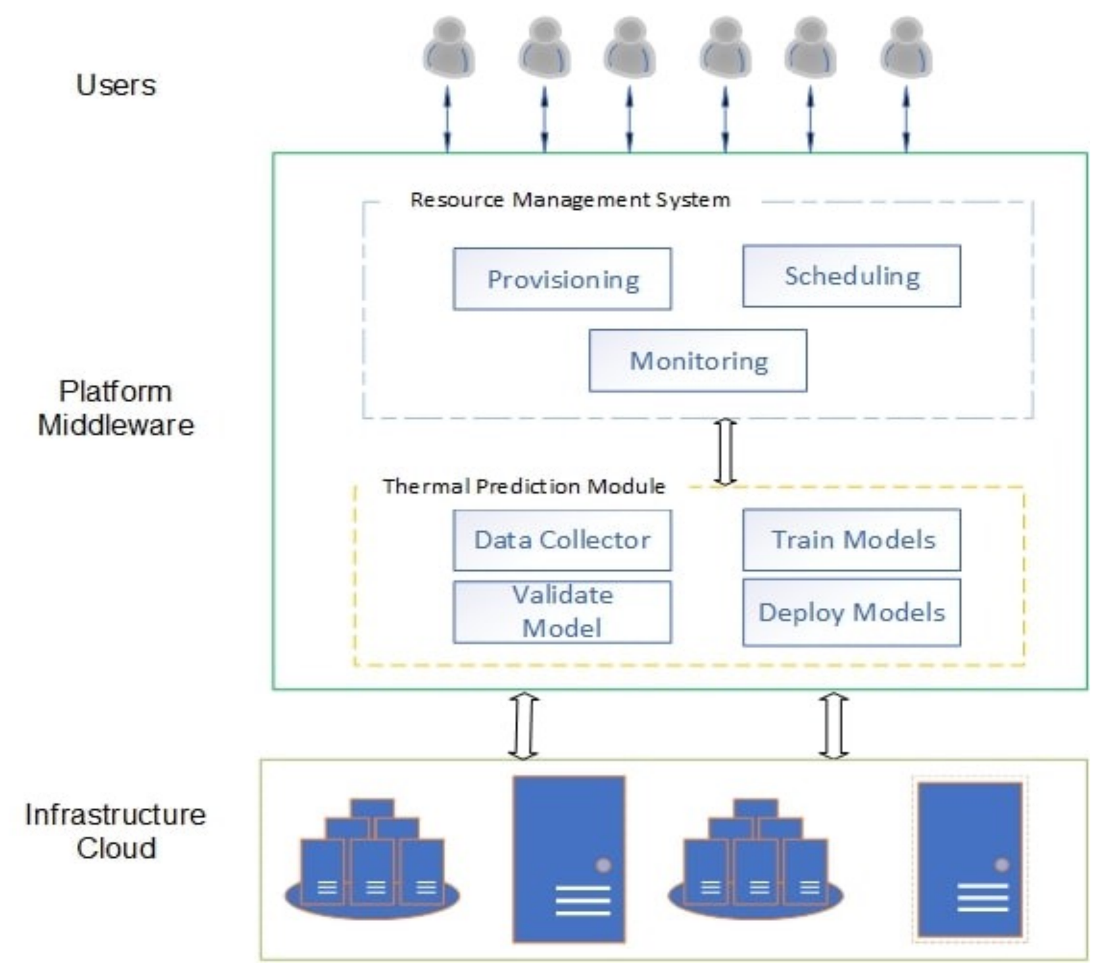
4 Hypothesis

Learning-based resource management solutions can learn complex dependencies between DC resource and application parameters and help to develop adaptive and energy efficient methods in DCs.

5 Proposed Approach



ML-based Resource Management: A System Model



6 Scientific Works on Cloud Sustainability

1. Developed ML techniques for optimizing GPU frequencies, ensuring energy efficiency and deadline-compliant scheduling[5].
2. Explored dynamic VM consolidation methods for energy savings and hotspot mitigation, maintaining SLA compliance[2].
3. Created thermal prediction models for precise server temperature forecasts, introducing a scheduling algorithm to reduce data center peak temperatures[1,4].
4. Designed tools for monitoring, data analysis, and ML model development[1,5].

7 Impact and Outcome

Paradigm shift: from "*time to solution*" to "*kW to solution*"

- A **quantum leap** in the sustainability DCs and ICT in general
- Significant reduction in energy & CO_2 footprint of DCs.
- Produced open-source datasets and software systems.

References

1. **Shashikant Ilager,** A. Toosi, M. Jha, I. Brandic, R. Buyya, **"A Data-driven Analysis of a Cloud Data Center: Statistical Characterization of Workload, Energy and Temperature",** UCC, 2023. **['Best Paper Award'].**
2. **Shashikant Ilager,** R. Kotagiri, and R. Buyya**, " ETAS: Energy and thermal-aware dynamic virtual machine consolidation in cloud data center with proactive hotspot mitigation", CCPE 2019.**
3. **Shashikant Ilager,** R. Kotagiri, and R. Buyya**, "Thermal Prediction for Efficient Energy Management of Clouds Using Machine Learning",** *IEEE TPDS*, 2021.
4. **Shashikant Ilager,** R. Muralidhar R. Kotagiri, and R. Buyya, **"A Data-Driven Frequency Scaling Approach for Deadline-aware Energy Efficient Scheduling on Graphics Processing Units (GPUs)",** CCGRID, 2020 **['Best Paper Award'].**
5. **Shashikant Ilager, J.** Fahringer, S. Dias, , I. Brandic**, "DEMon: Decentralized Monitoring for Highly Volatile Edge Environments",** UCC 2022
6. **Shashikant Ilager,** "**Machine Learning-based Energy and Thermal Efficient Resource Management Algorithms for Cloud Data Centres"**, Phd Thesis, 2021.

HPC Lab
High Performance Computing Laboratory