

Data-centres of the future: The RF2.0 Project at DESY



Dwayne Spiteri, Jan Hartmann, Sandro Grizzo, Konrad Kockler, Yves Kemp, Martin Gasthuber, Kilian Schwarz et al

ESSRI - 01/10/2026

HORIZON Europe Project Research Facility 2.0

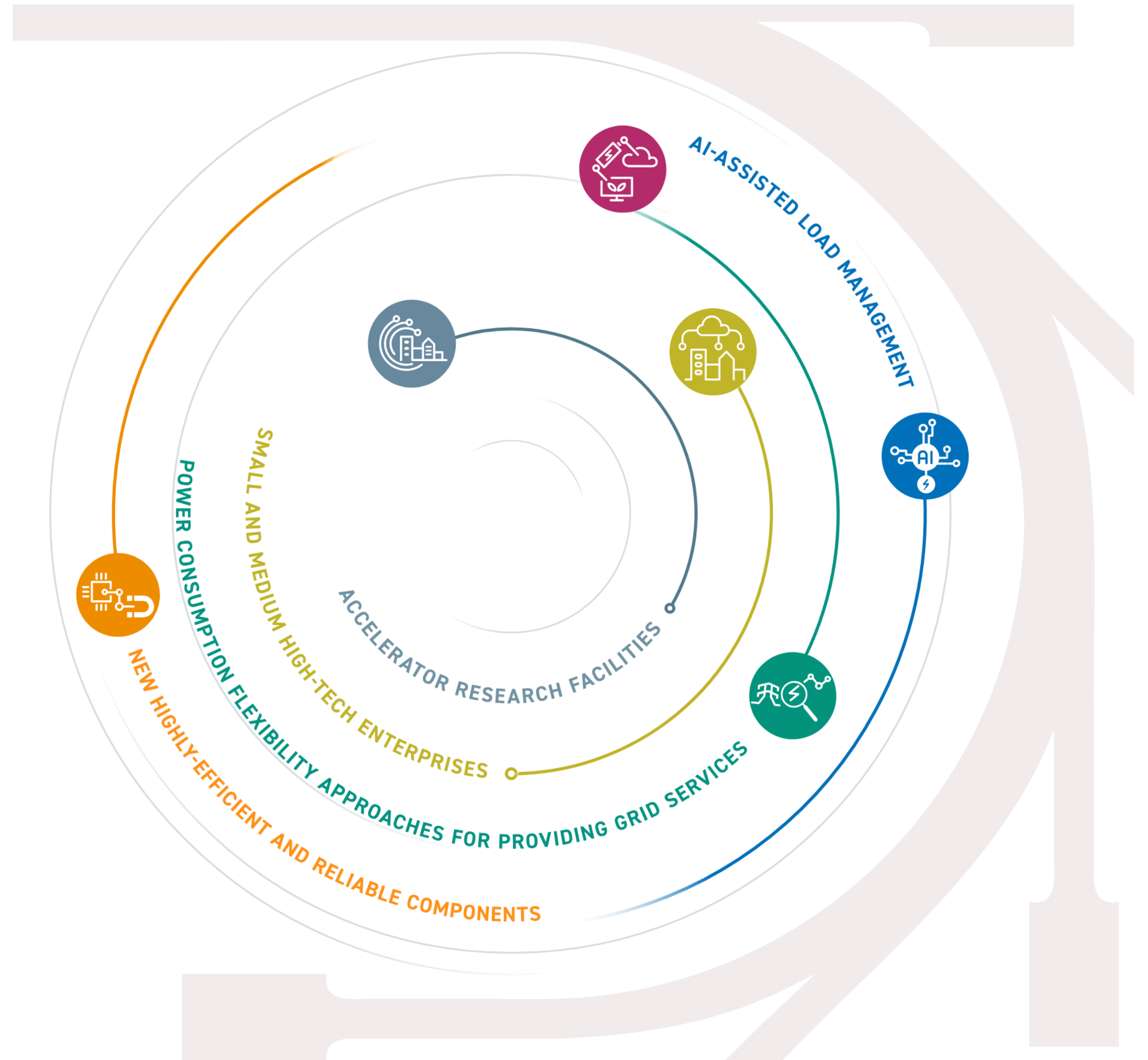


Objective:

Design, develop and validate in-field novel solutions for components and systems, that allow the **transformation of particle accelerators** and **computing centres** into **more sustainable** and **energy efficient research infrastructures**.

Interdisciplinary approach:

At the component and system level, involving both the physics, computer science and energy technology topics.



The RF2.0 Consortium



4 High-Tech SMEs

6 Accelerator Research Facilities

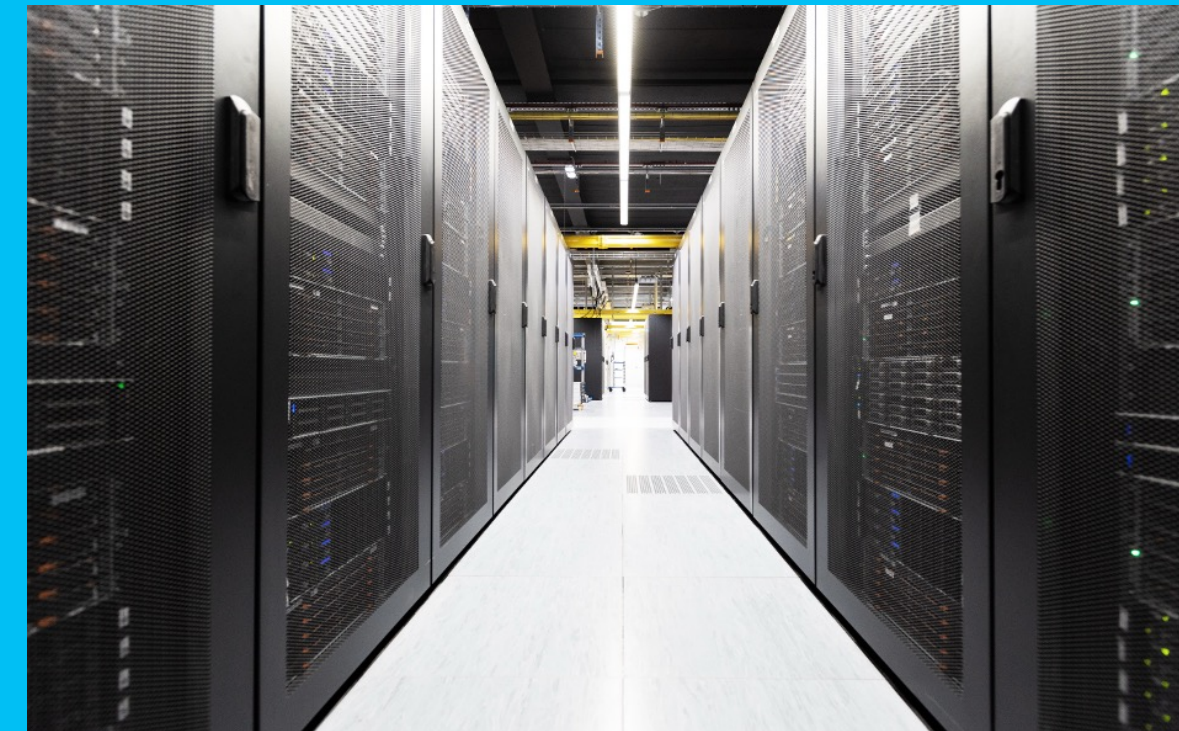
4 Demonstrator Projects

5.6M€ Total Budget

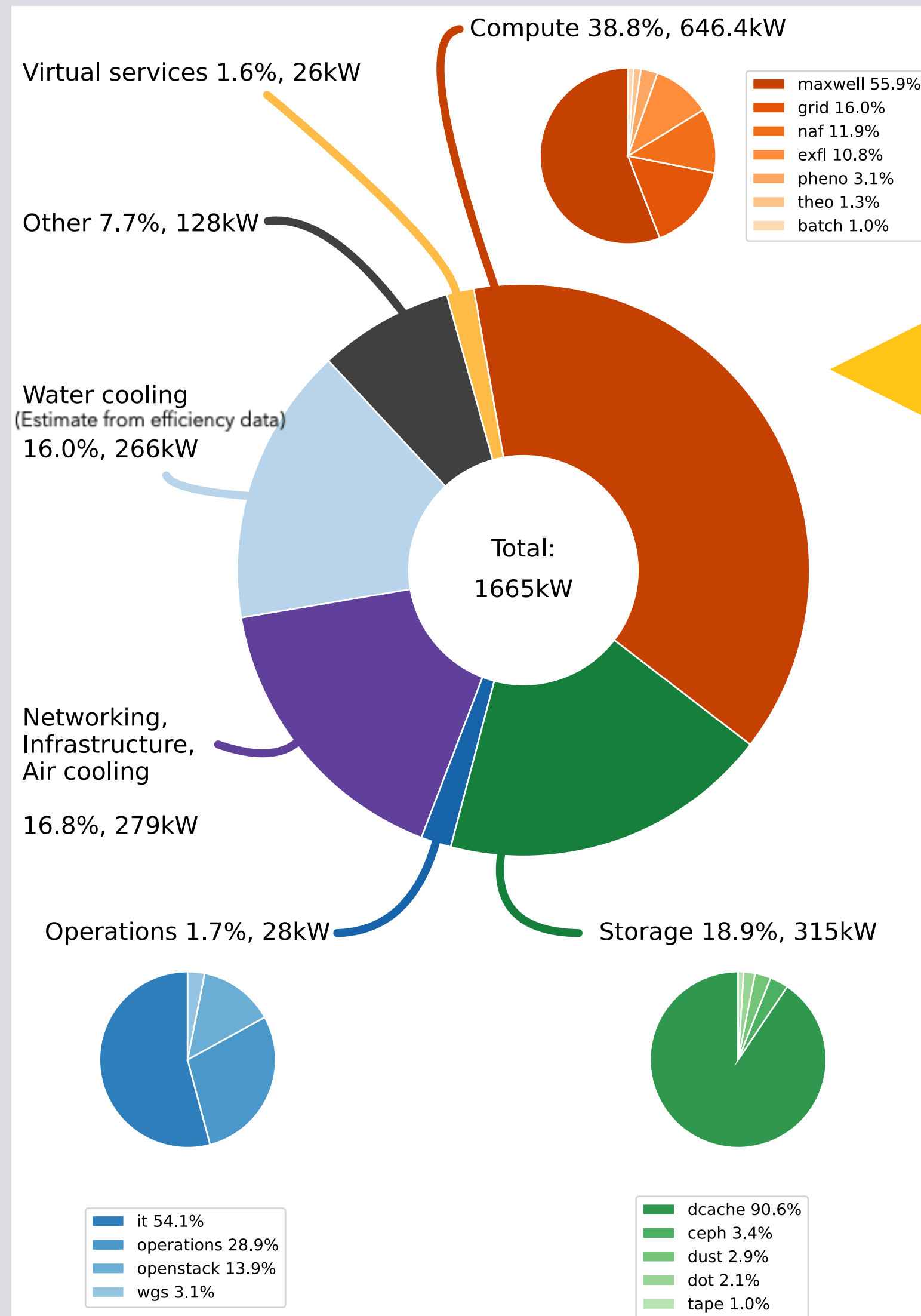


- * EU-funded project tasked with investigating how large research infrastructures can be more sustainable.
- * At DESY we focus on what this means for current and future operation (and construction) of green data-centres.

DESY DATA CENTRE



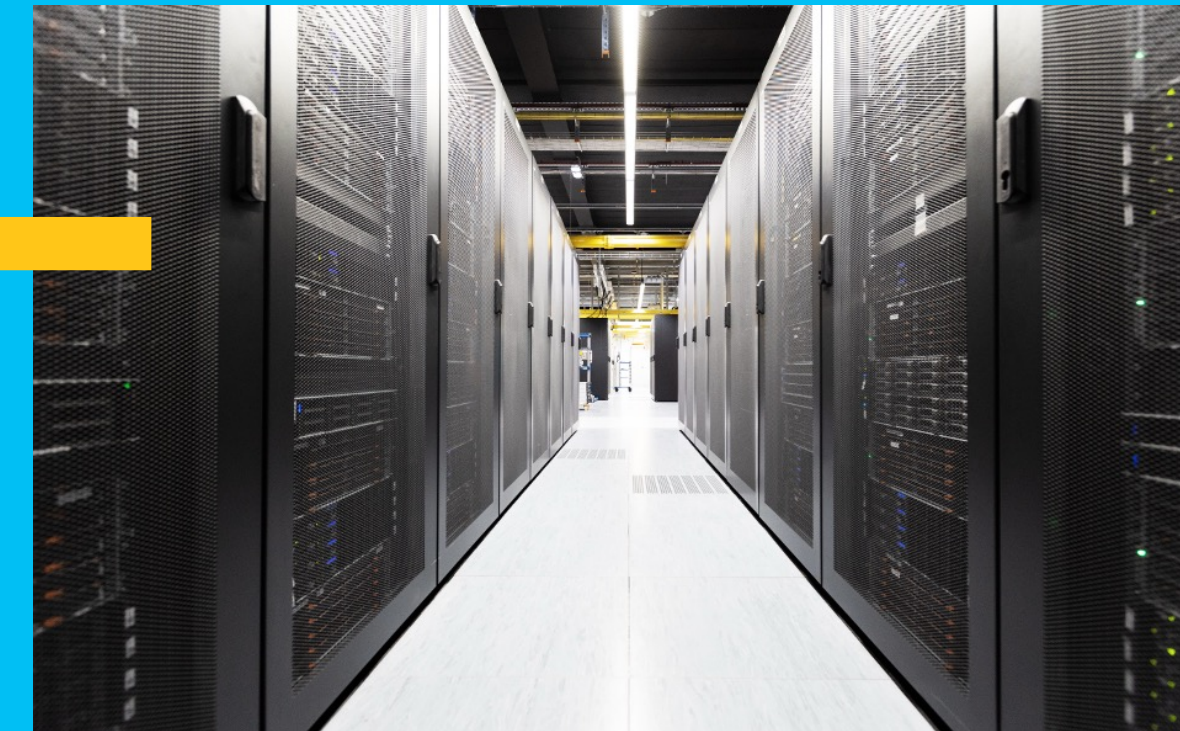
- * EU-funded project tasked with investigating how large research infrastructures can be more sustainable.
- * At DESY we focus on what this means for current and future operation (and construction) of green data-centres.



1. Learn

Take detailed and fine-grained measurements of data-centre components

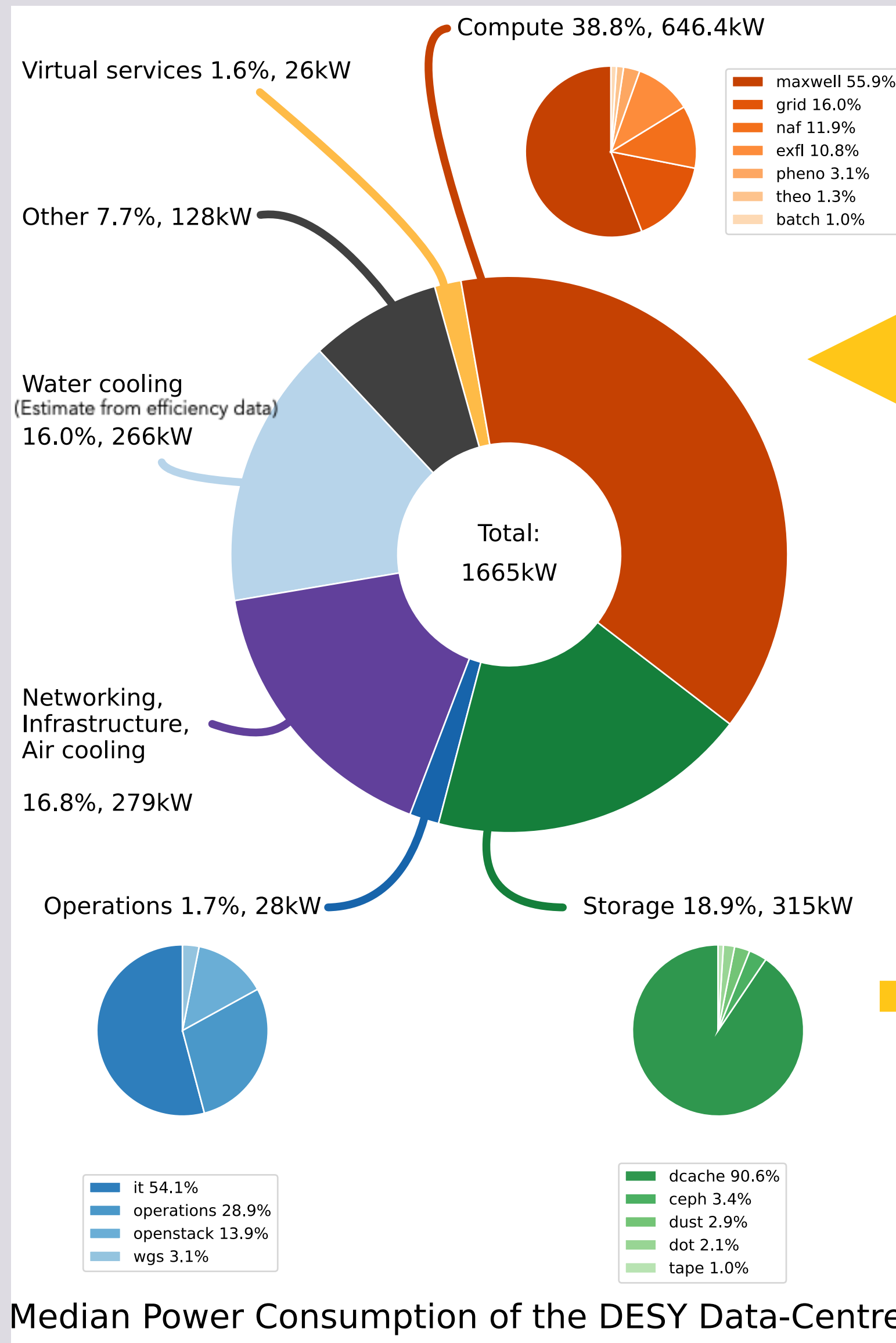
DESY DATA CENTRE



Median Power Consumption of the DESY Data-Centre

* EU-funded project tasked with investigating how large research infrastructures can be more sustainable.

* At DESY we focus on what this means for current and future operation (and construction) of green data-centres.



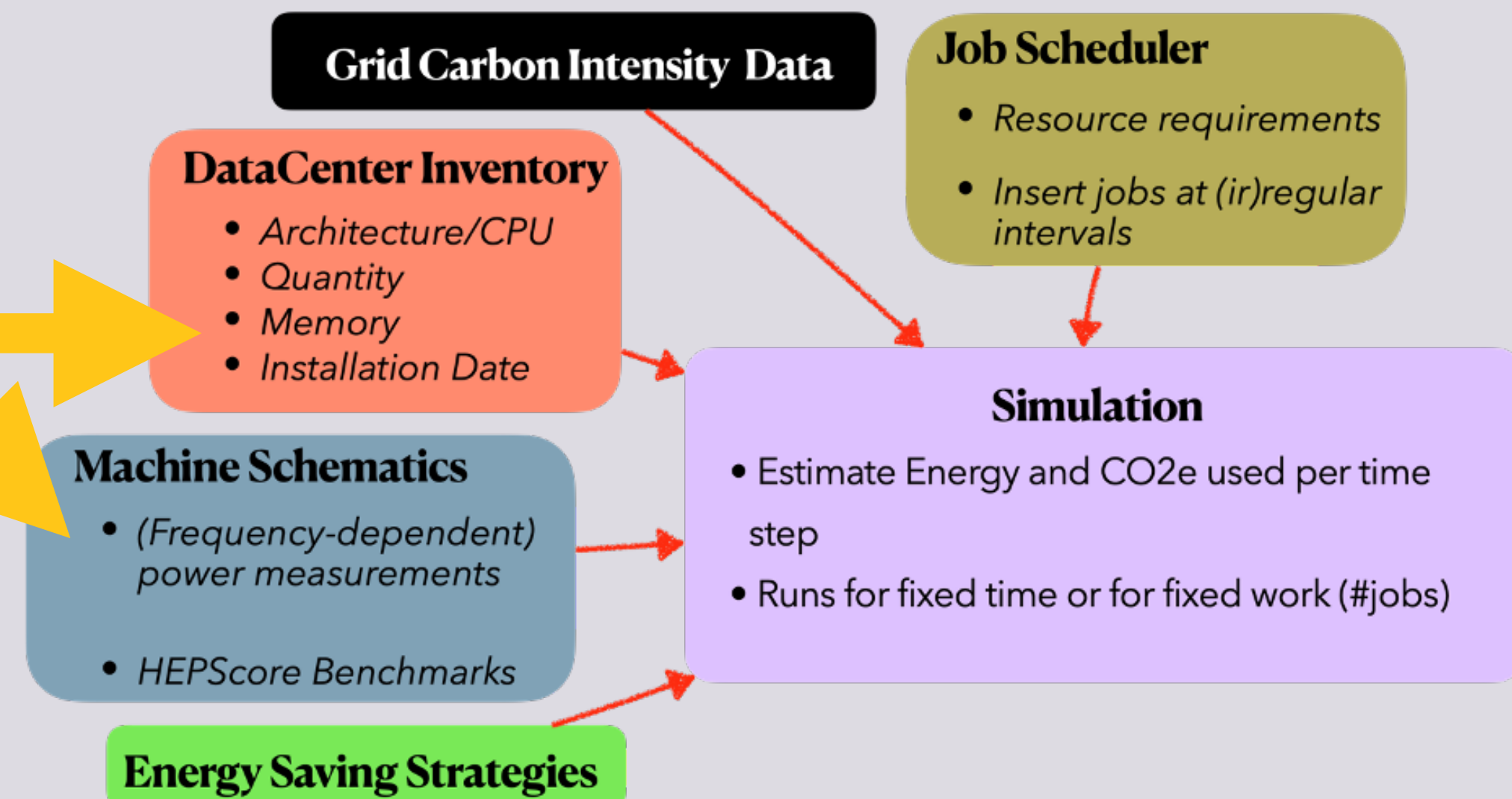
1. Learn

Take detailed and fine-grained measurements of data-centre components



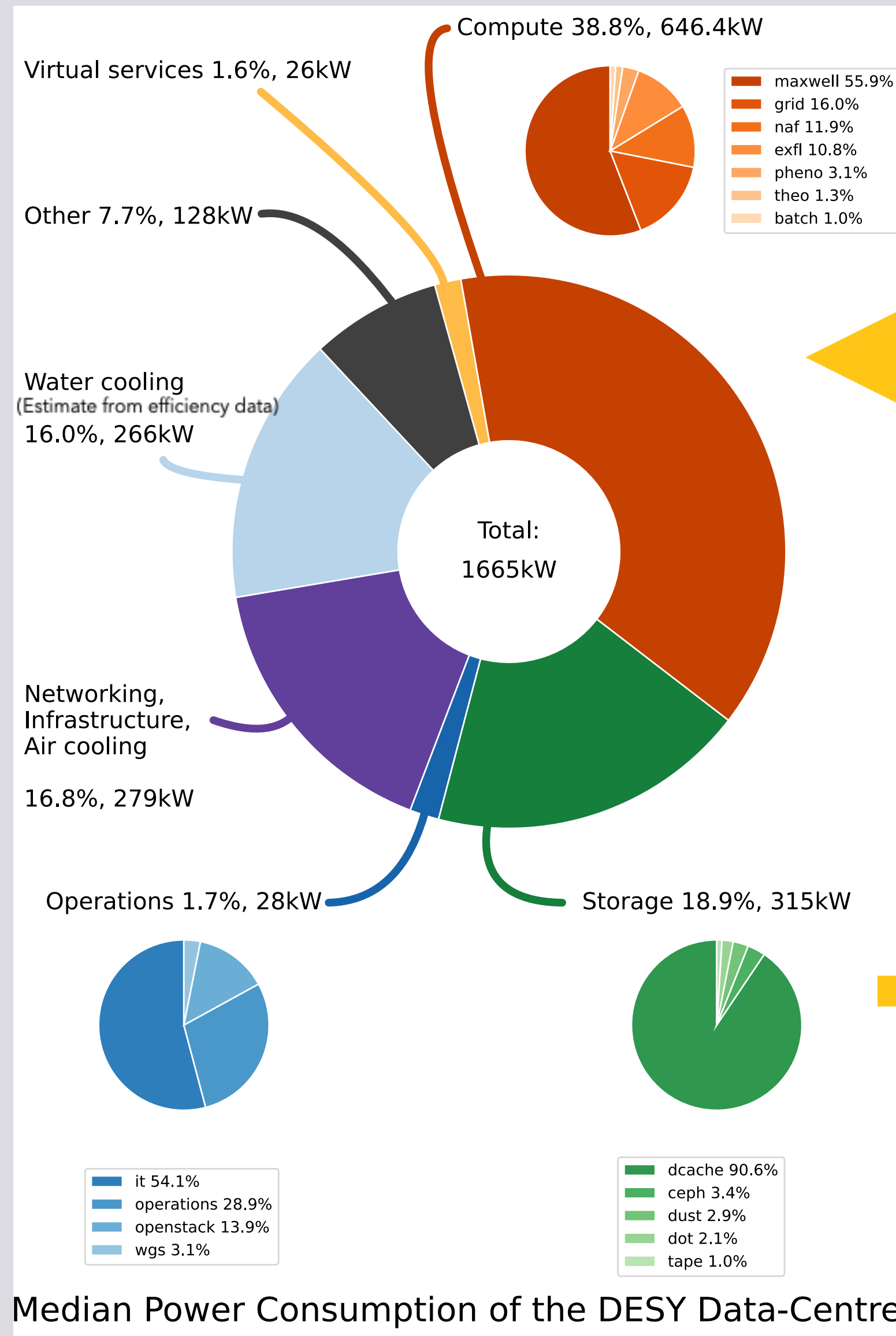
2. Simulate

Test different energy-saving strategies/policies in a simulation of the data-centre



* EU-funded project tasked with investigating how large research infrastructures can be more sustainable.

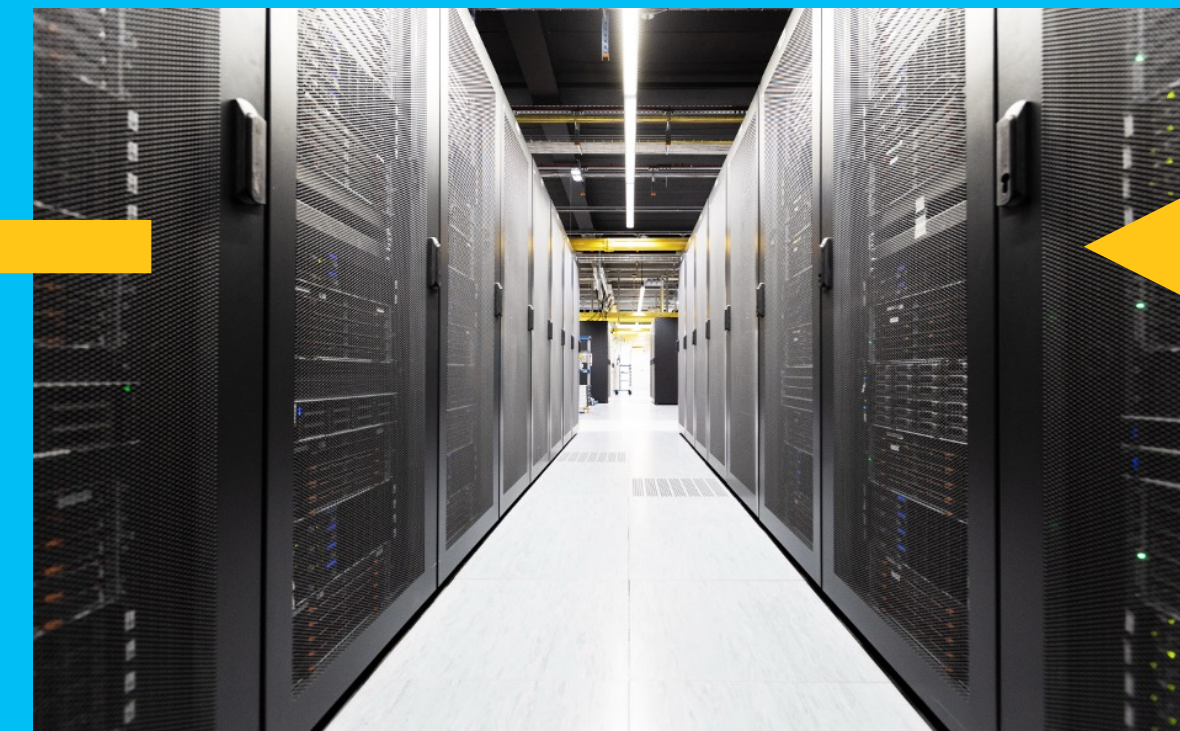
* At DESY we focus on what this means for current and future operation (and construction) of green data-centres.



1. Learn

Take detailed and fine-grained measurements of data-centre components

DESY DATA CENTRE

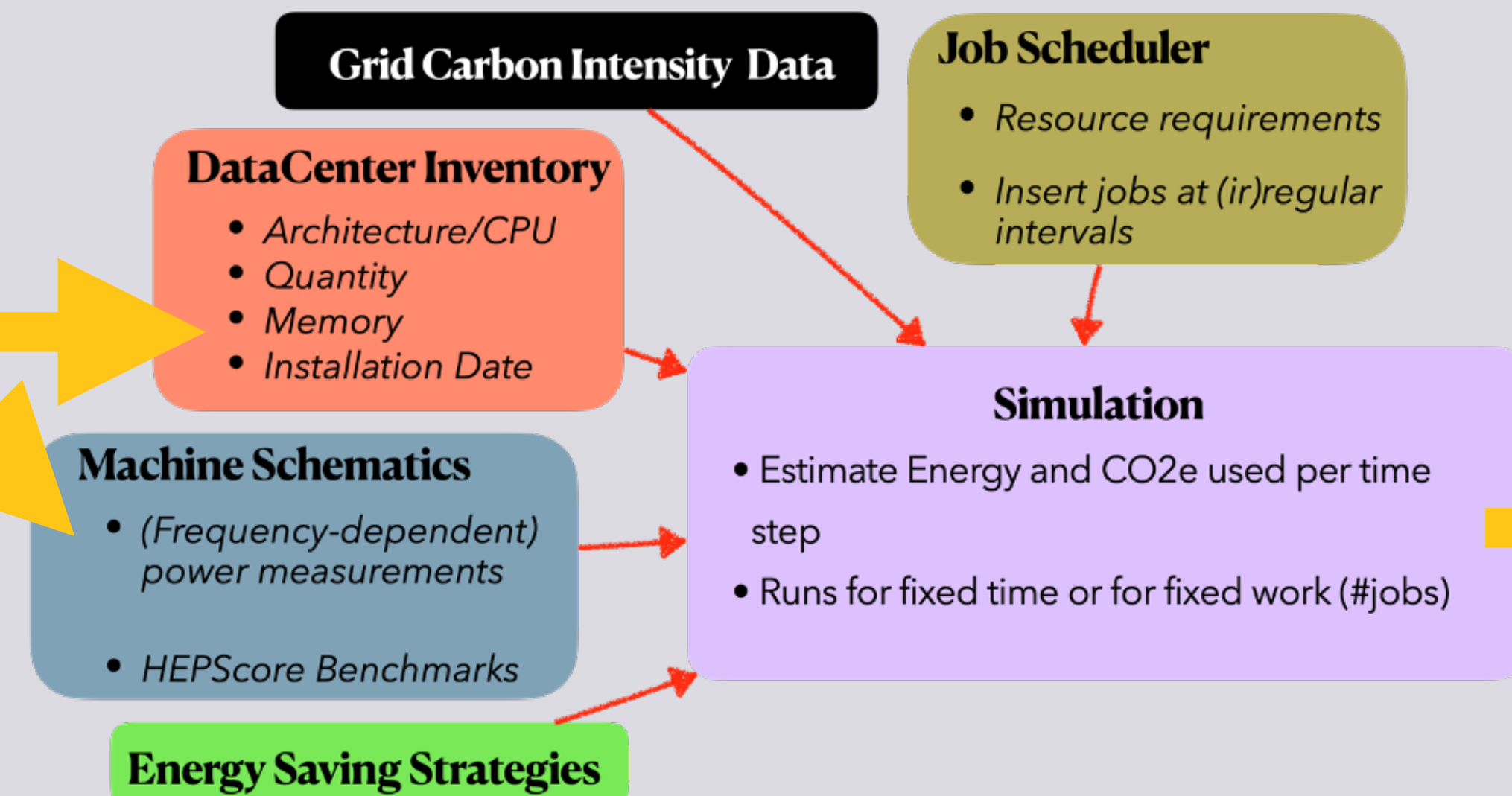


3. Refine

Take simulation findings and dynamically implement them at the datacenter

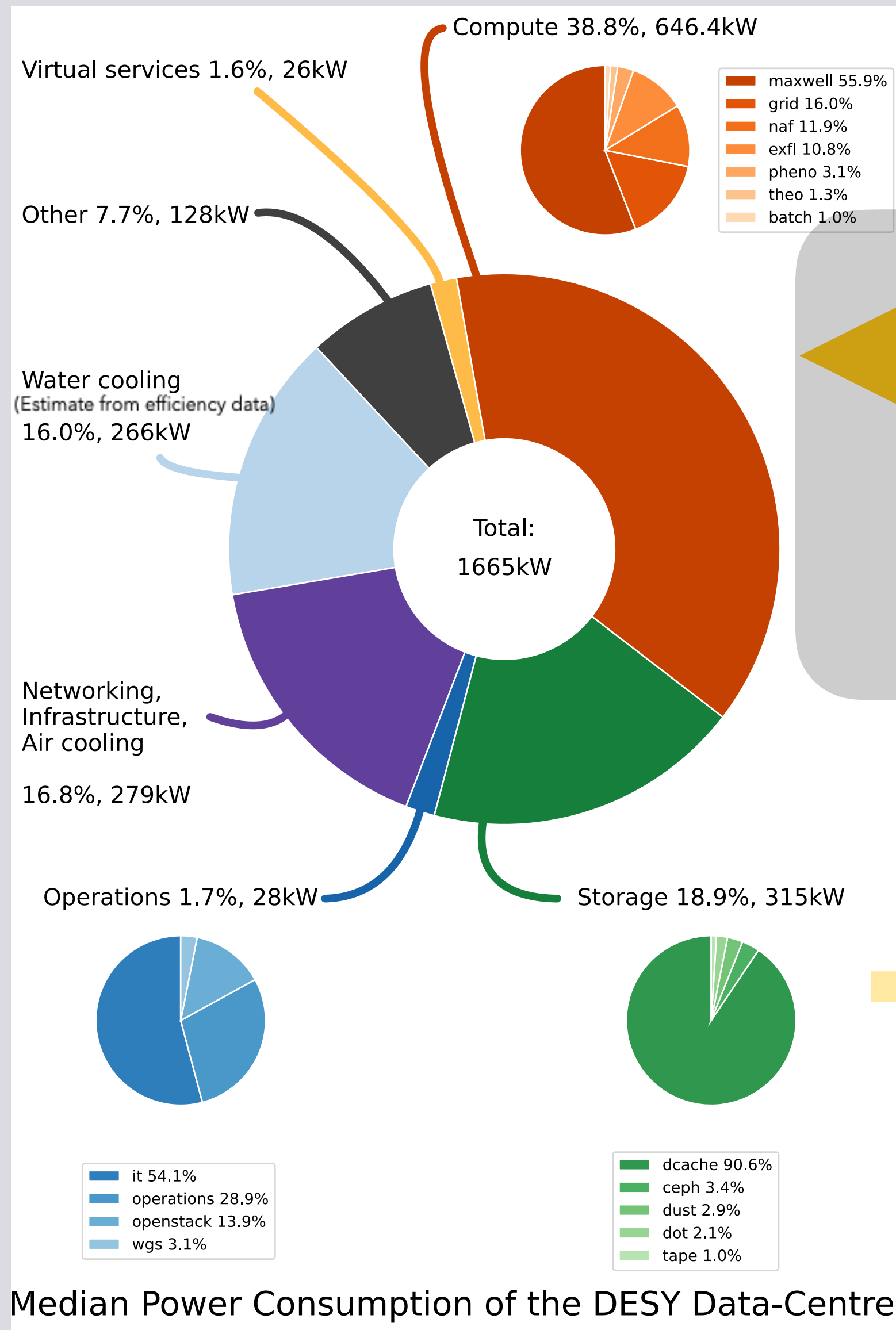
2. Simulate

Test different energy-saving strategies/policies in a simulation of the data-centre



* EU-funded project tasked with investigating how large research infrastructures can be more sustainable.

* At DESY we focus on what this means for current and future operation (and construction) of green data-centres.



1. Learn

Take detailed and fine-grained measurements of data-centre components

DESY DATA CENTRE

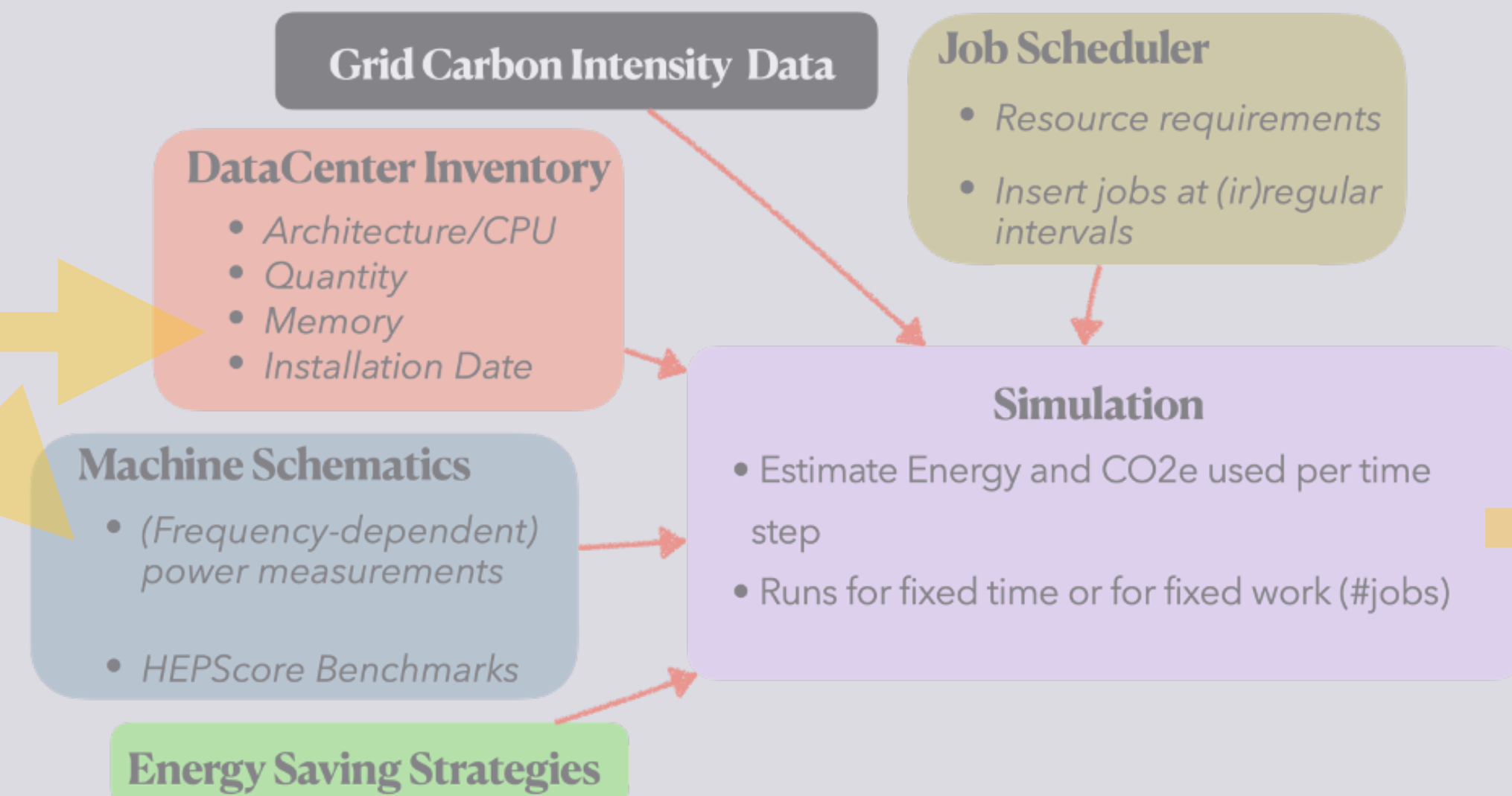


3. Refine

Take simulation findings and dynamically implement them at the datacenter

2. Simulate

Test different energy-saving strategies/policies in a simulation of the data-centre



1. Learn

Static Measurements

Machine Identifiers

Manufacturer

CPU type and architecture

Installation Date

Work Capability

Number of Cores

Memory Available

Operational Frequencies

Idle Power Draw



(Frequency-dependent)

HEPScore values

Typical Power Draw



Rack Power Used



Power Measurements



Performance Measurements

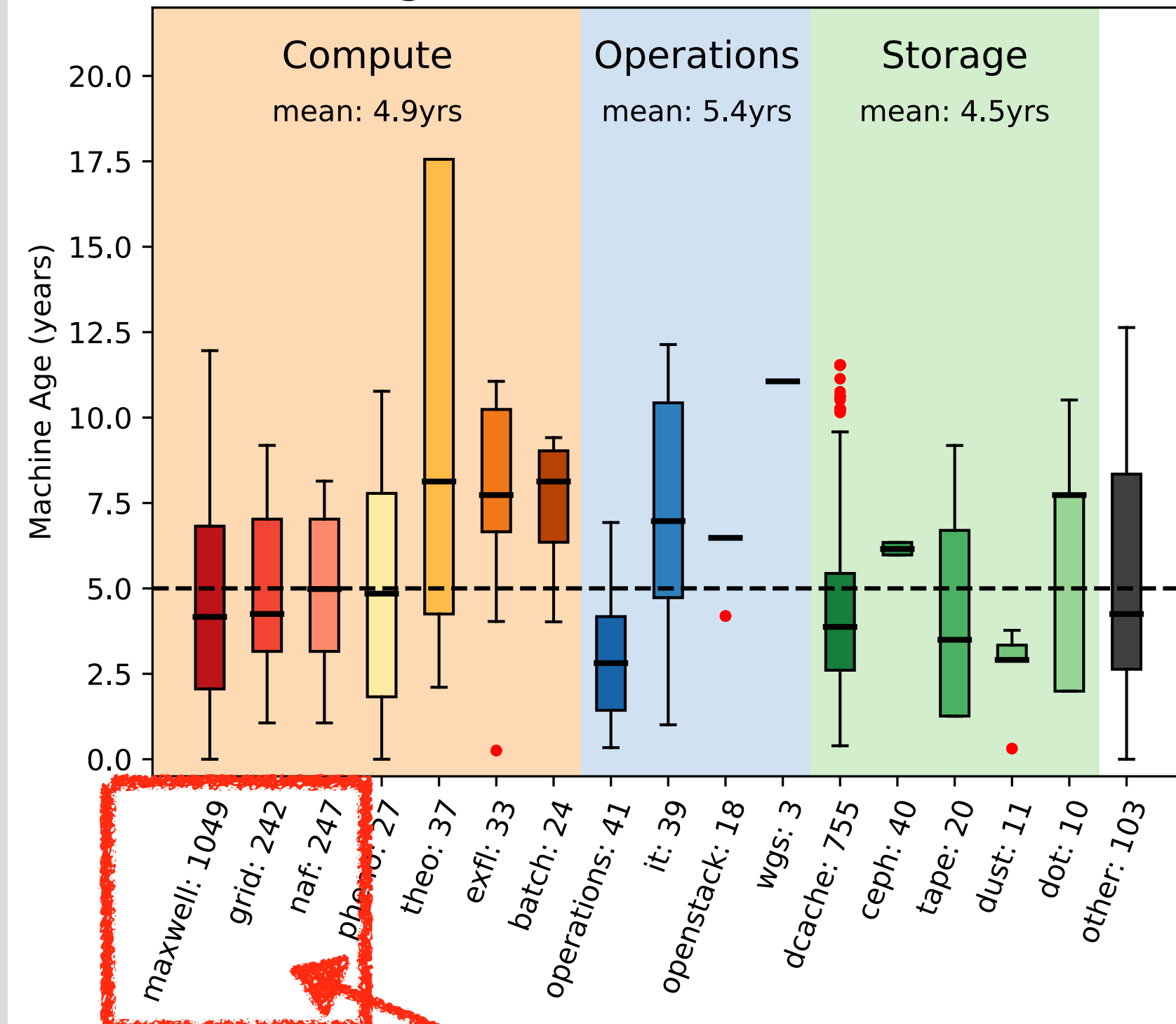
- Relative performance of machines → Benchmarking (HEPScore)

- Benchmark created and maintained by a CERN-led team designed to calculate how well compute servers can handle HEP Workloads.

- Bigger number better

Data-Centre Data Initial Analysis

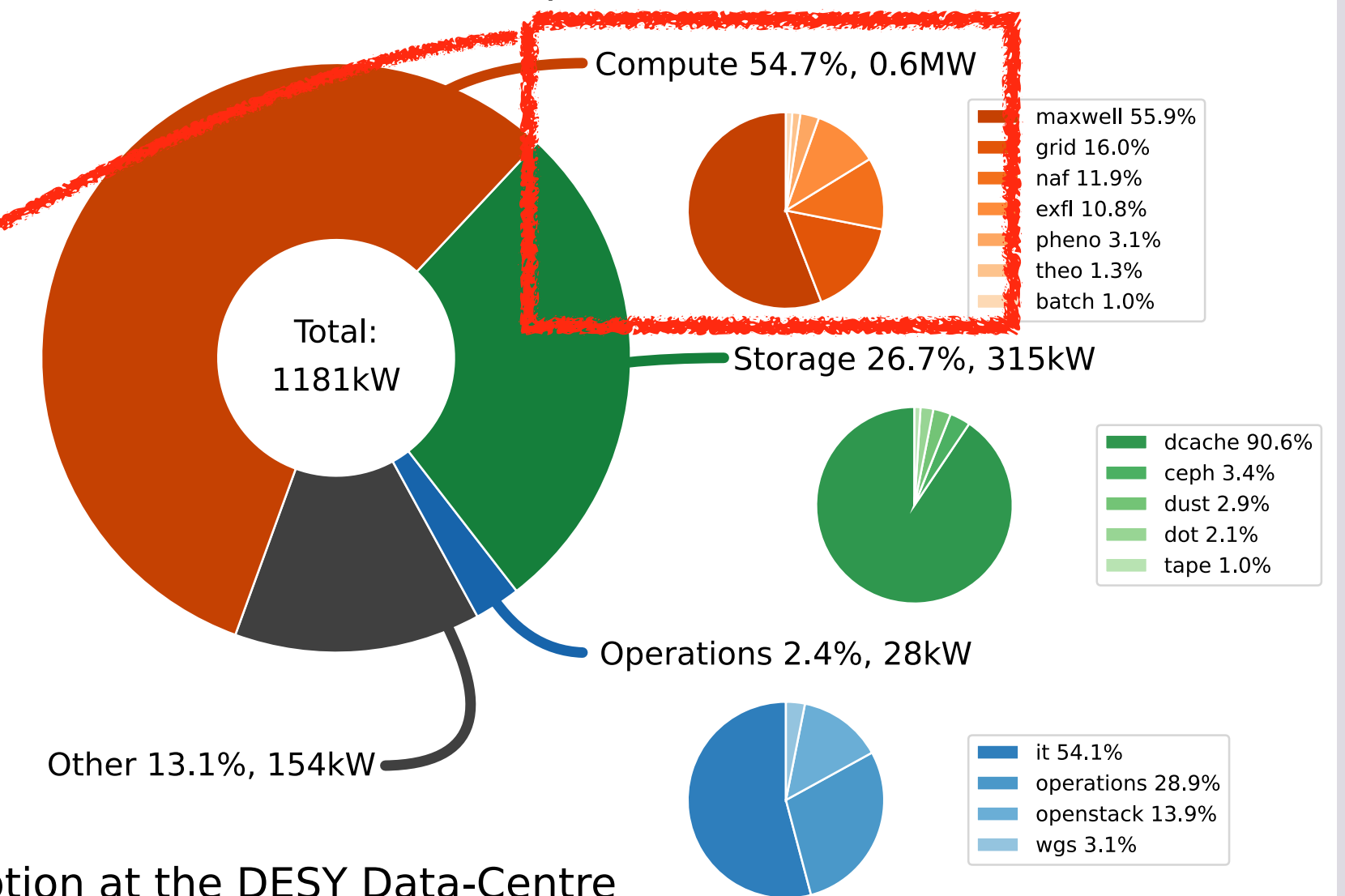
Current IQR Age Profiles of Datacentre Machines



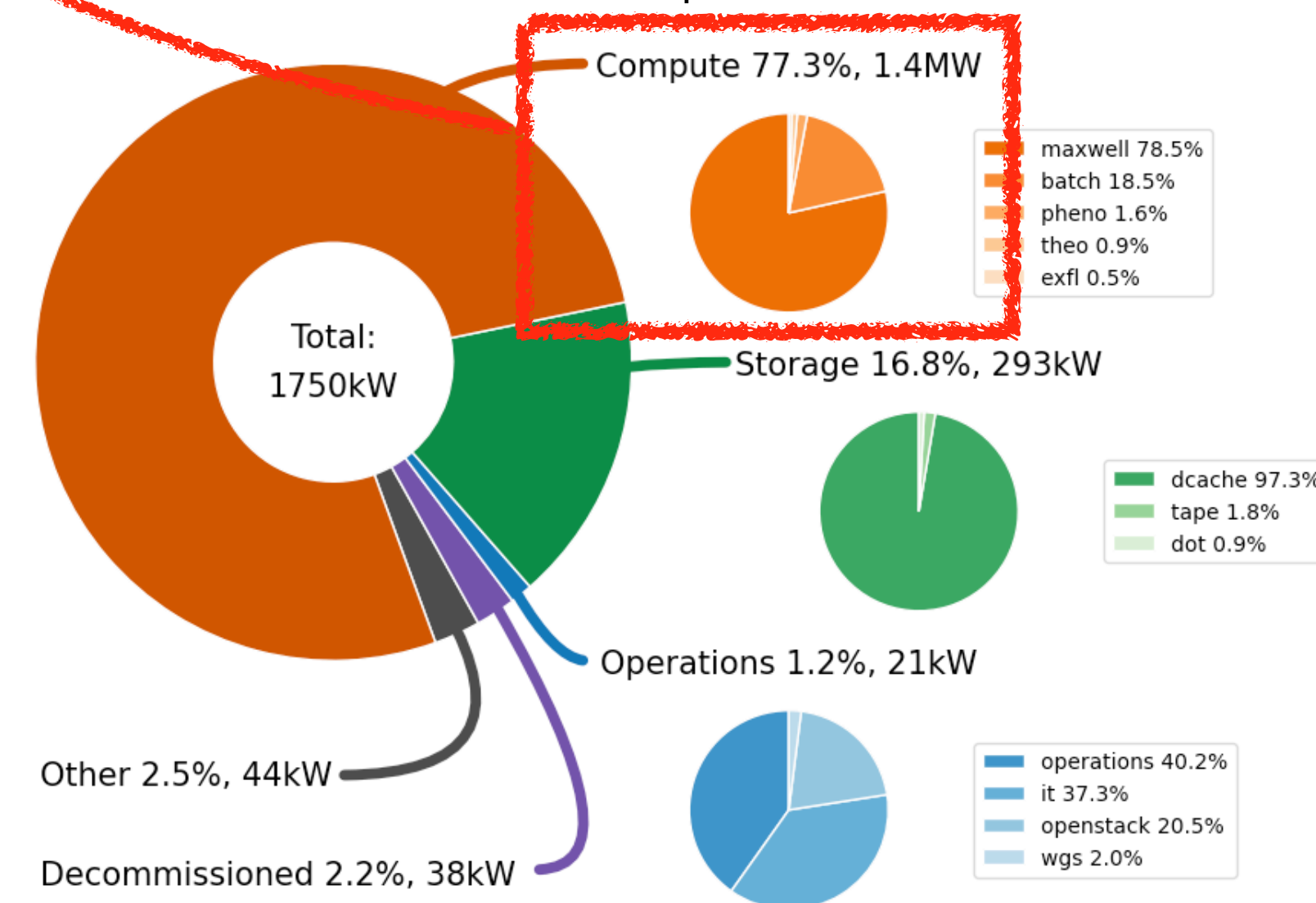
- We have three main computing clusters with different types of users

- Largest Fluctuations in power is in Compute

Median Power Consumption at the DESY Data-Centre



Maximum Power Consumption at the DESY Data-Centre



The tools at our disposal

Policies/Actions to take to have more control over when we consume energy

Computing is a natural starting point

- It's the largest part of the pie
- It's the most varied in energy use
- Changing parameters on a machine affects the effectiveness of running jobs

Frequency
scaling

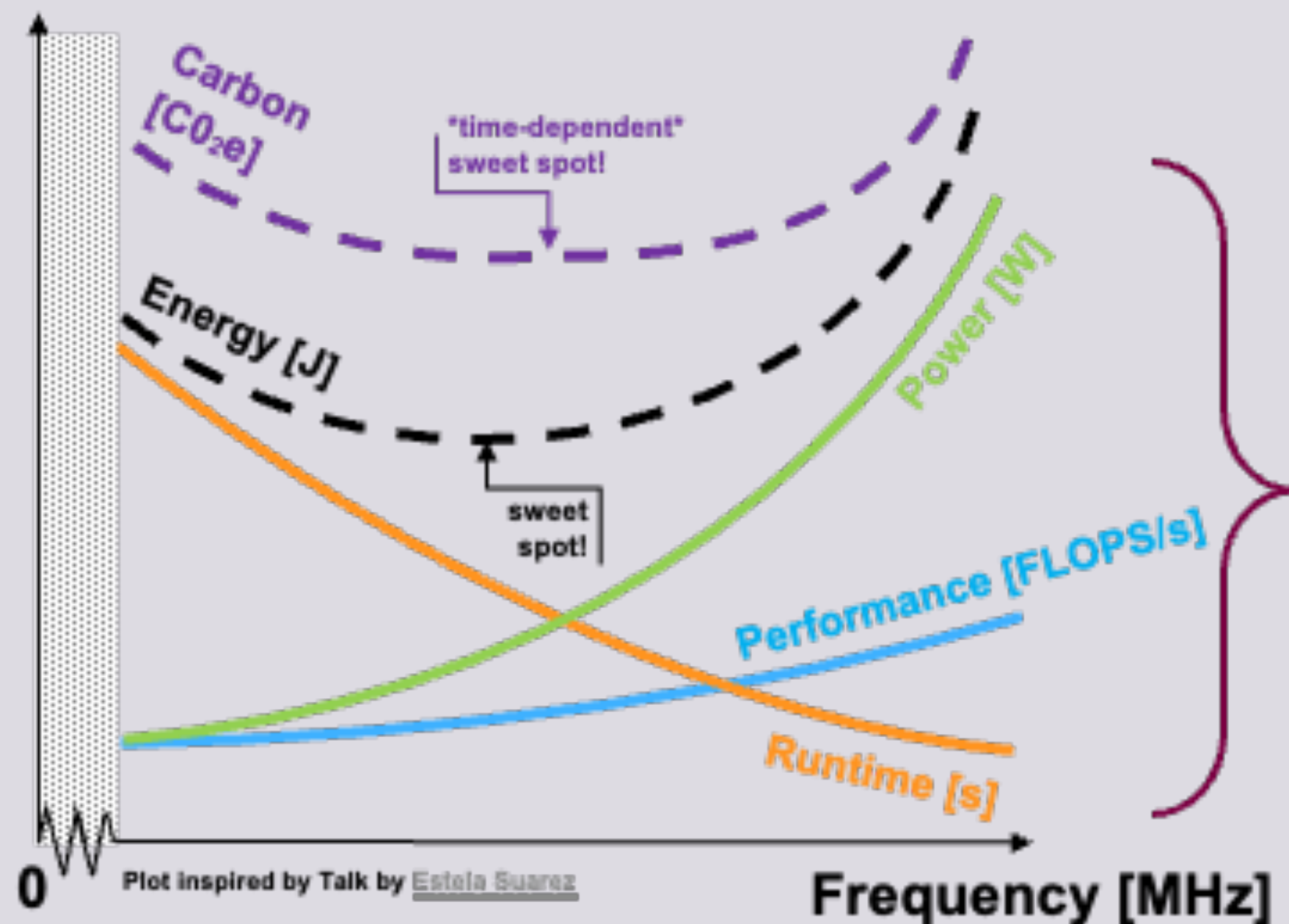
(Nice) Node
shutdown

Need a metric to determine how well a machine is running

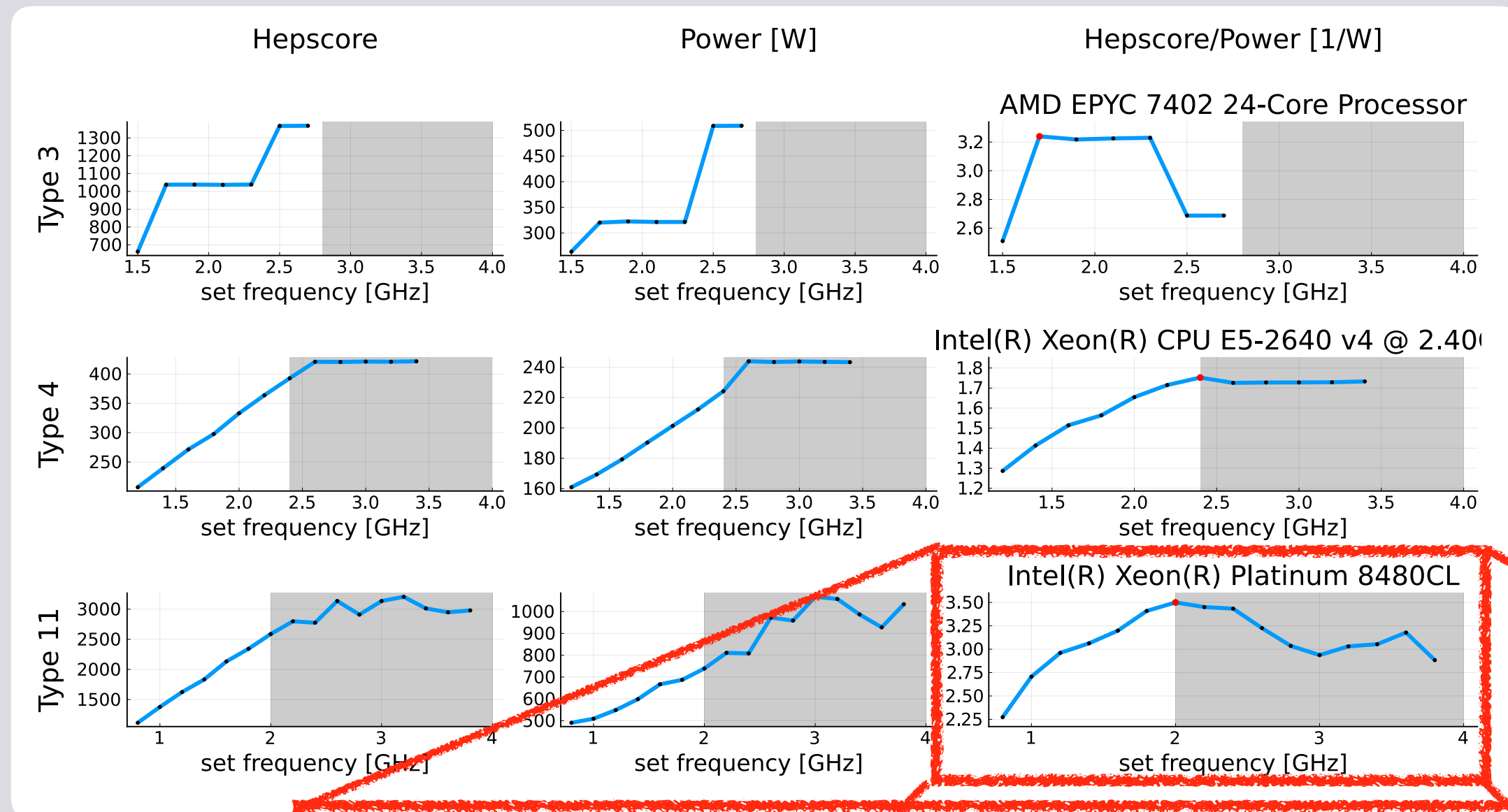
- Needs to run with workloads that are realistic to our users
- Should be a well-maintained standard reference

HEPScore23 (HEPScore)

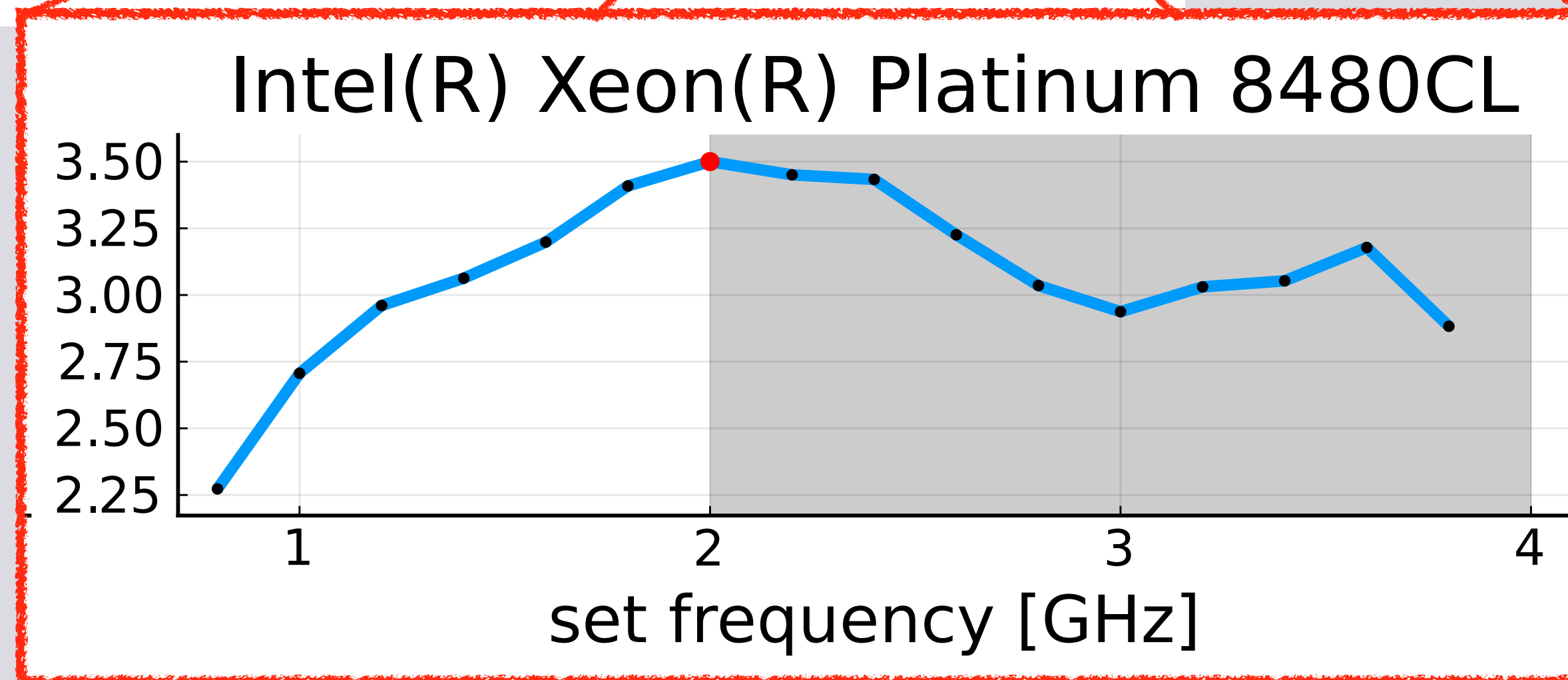
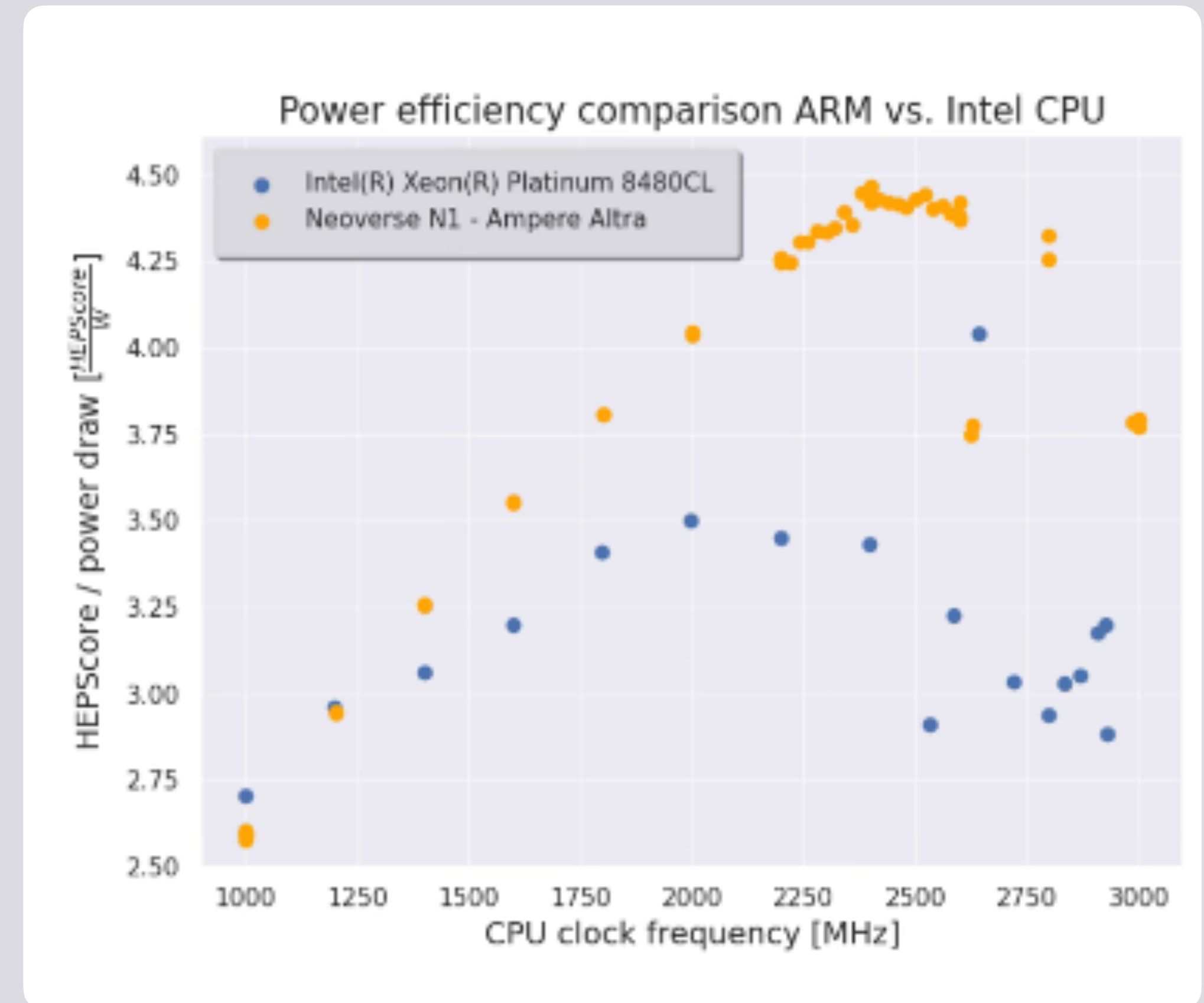
- Uses workloads common to particle physics
- Has power plug-ins
- Maintained by CERN
- Single number related to number of events per second that can be processed



Data-Centre Data Summary Analysis



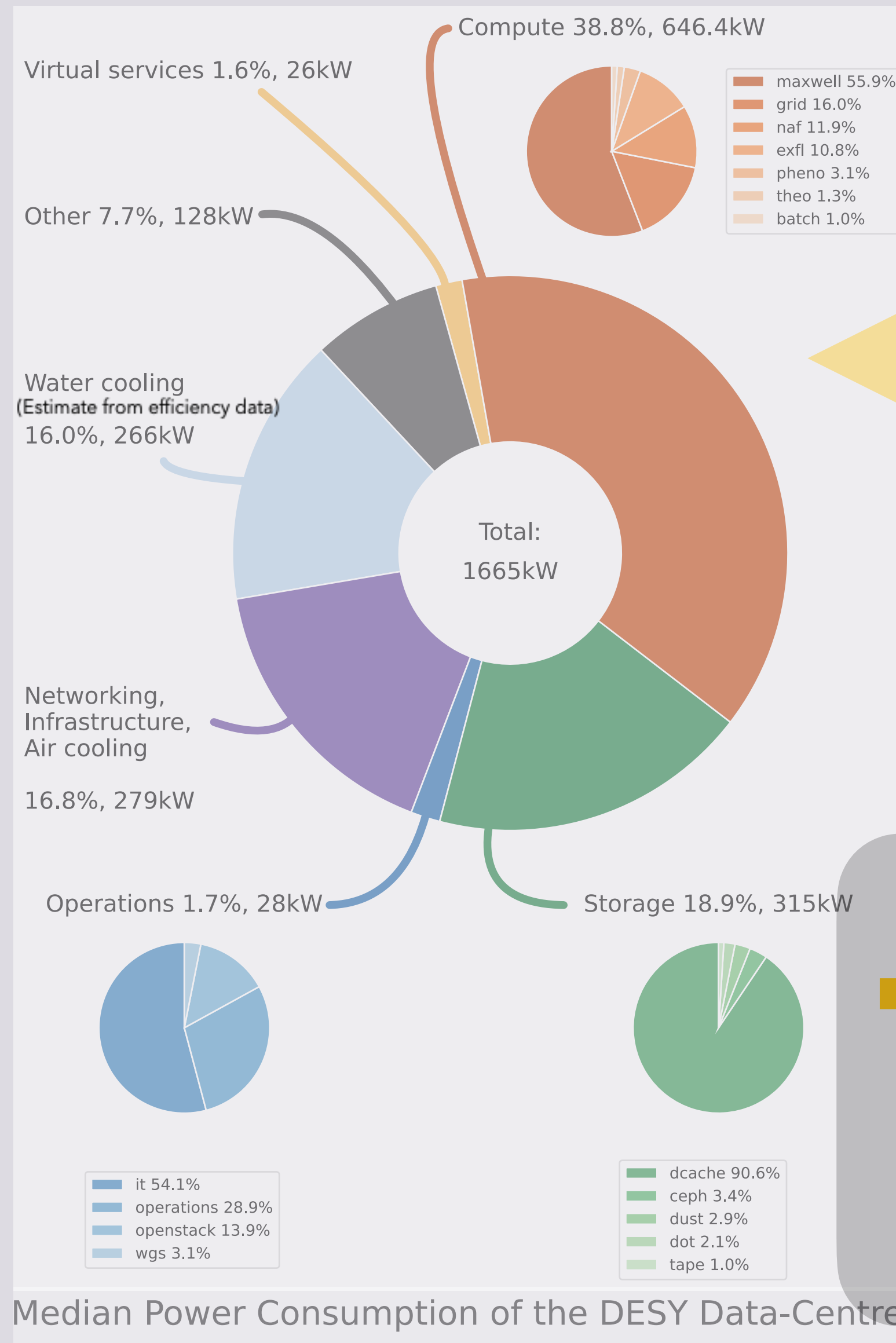
- HS/W for different computer chip architectures



- HEPscore Value/Watt against Clock speed for each type of node = **HS/W**

* EU-funded project tasked with investigating how large research infrastructures can be more sustainable.

* At DESY we focus on what this means for current and future operation (and construction) of green data-centres.



1. Learn

Take detailed and fine-grained measurements of data-centre components

DESY DATA CENTRE



3. Refine

Take simulation findings and dynamically implement them at the datacenter

2. Simulate

Test different energy-saving strategies/policies in a simulation of the data-centre

Grid Carbon Intensity Data

DataCenter Inventory

- Architecture/CPU
- Quantity
- Memory
- Installation Date

Machine Schematics

- (Frequency-dependent) power measurements
- HEPScore Benchmarks

Energy Saving Strategies

Job Scheduler

- Resource requirements
- Insert jobs at (ir)regular intervals

Simulation

- Estimate Energy and CO2e used per time step
- Runs for fixed time or for fixed work (#jobs)

How do you simulate a datacenter?

- Data-centres power budget are largely compute - so start there

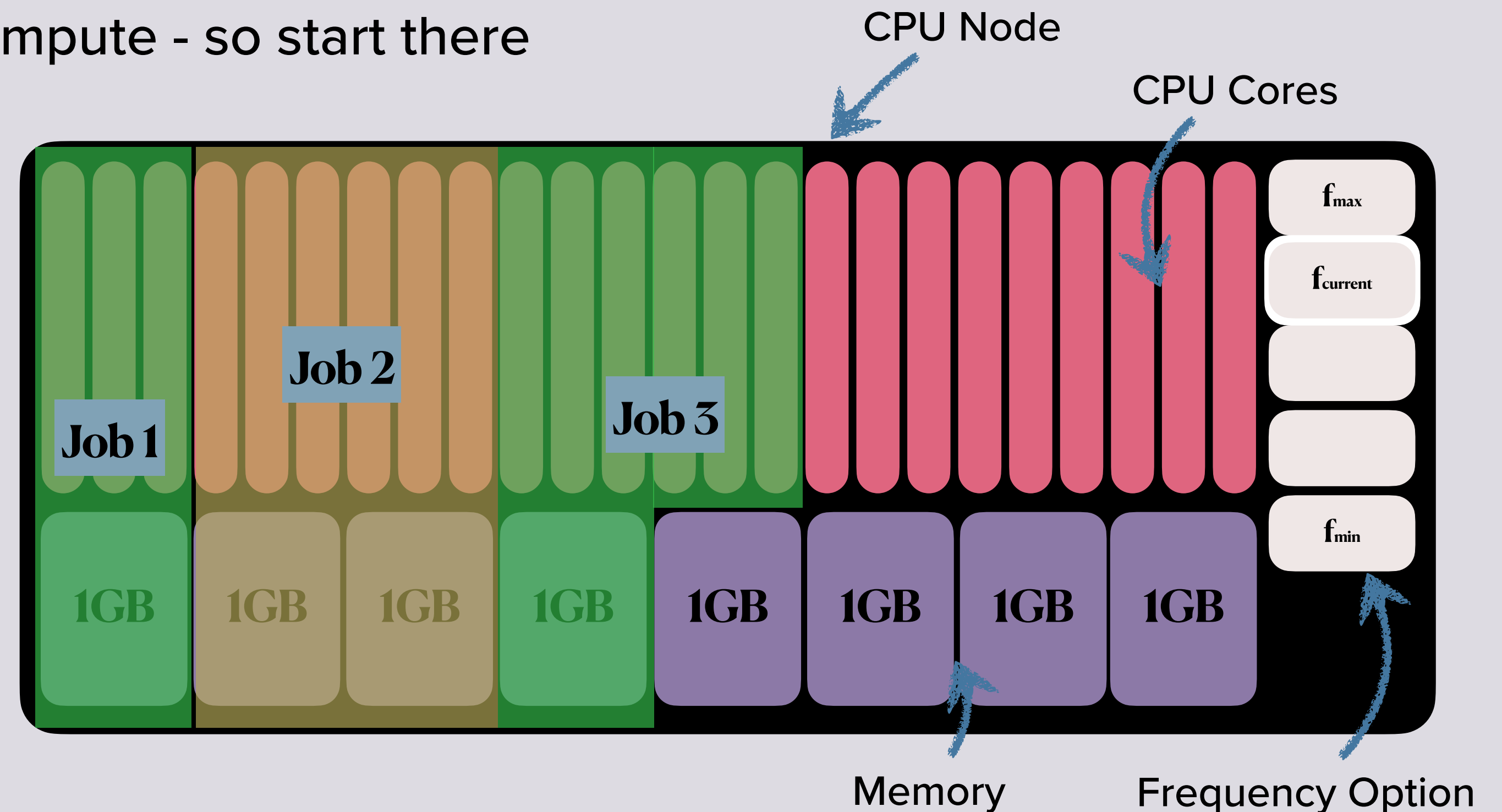
- Machines/Servers have nodes which have cores, and each core can be run in parallel to perform tasks.

- Jobs come in and request (use) cores and memory

- Assume that power used is

$$P_{Total} = P_{Idle} + \frac{\text{Cores Used}}{\text{Cores available}} * (P_{Max} - P_{Idle}) \quad \text{* For non hyper-threaded nodes}$$

$$P_{Total} = P_{Idle} + \left(\frac{\text{Cores Used}}{\text{Cores available}} * (P_{Max} - P_{Idle}) \right) * \frac{f_{current}}{f_{max}}$$

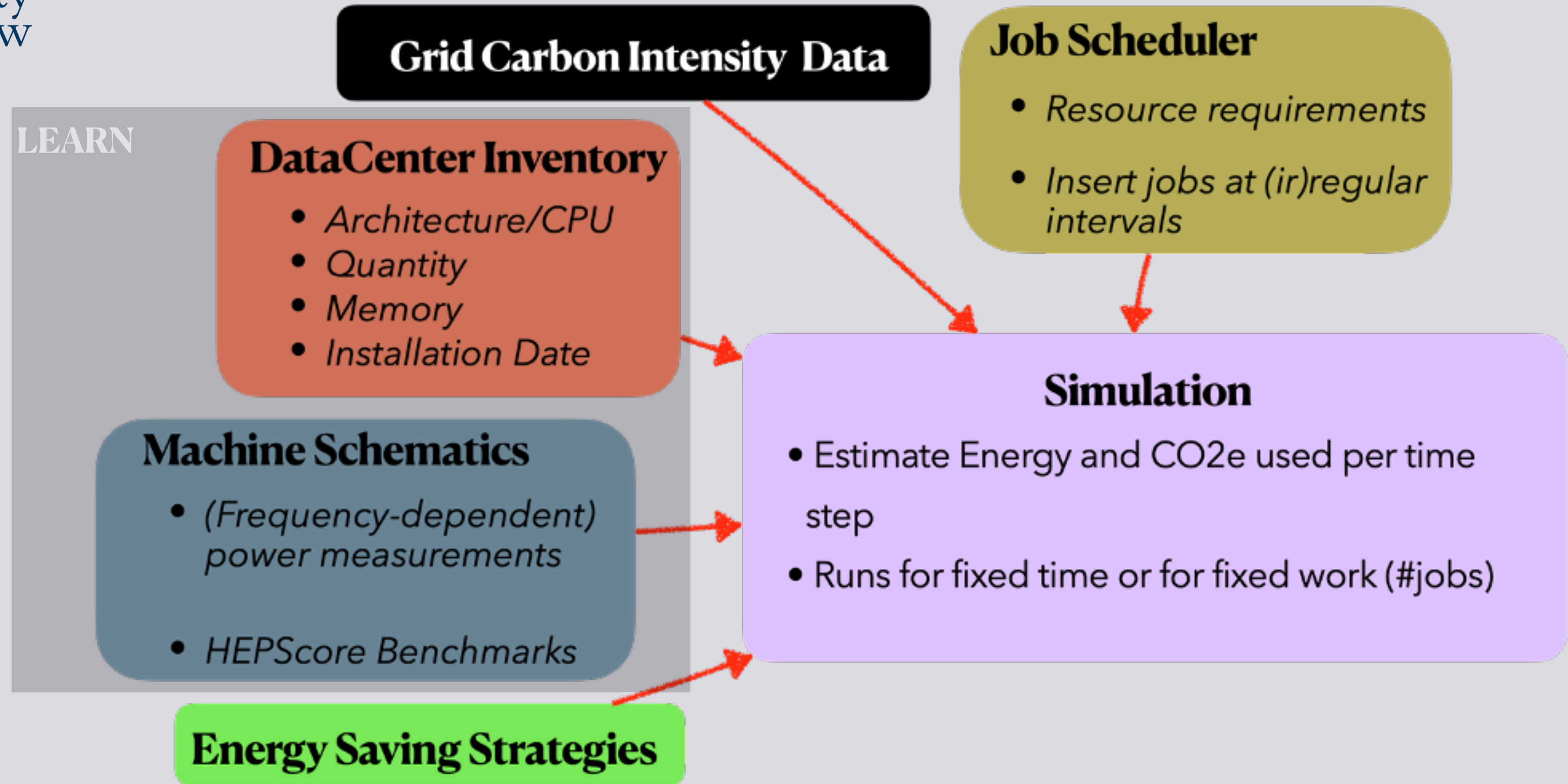


ORACLE-D

Optimised Resource Analysis and Carbon Legacy Estimator - for Data centres



- Initially created at the University of Glasgow - Currently being expanded at DESY using RF2.0 funding
- Mainly aimed at simulating data-centre **compute** and outputting carbon usage data



Output

- Each time the simulation is called, a file gets produced with the following information

```
=====
Summary
=====

Total Simulated-time Duration : 87.0 hours
Total Real-time Duration      : 51.8 minutes

Jobs Started                  : 300000
Jobs Finished                  : 300000

Total CPU duration            : 1462488.0 hours
Average CPU duration           : 4.87 hours
Average Occupancy of all clusters : 72.7 %

Total energy consumed by compute : 10696.31 kWh
Peaktime (5-9pm) energy consumption: 1936.24 kWh
Average energy consumption per job : 35.65 Wh

Estimated CO2e emissions       : 5629.292 kg
Estimated Peaktime CO2e emissions : 1062.283 kg
Average CO2e emissions per job   : 18.764 g
Peaktime CO2e emissions percentage: 18.871 %
```

Simulated and Real-time duration of the simulation

Job information

Total and Average CPU duration + Cluster Occupancy

Estimated energy used in total, during peak times and job-average

Estimated CO₂ (e)quivalent emissions for said work

Use Case 1 - Can you save carbon by shifting work?

- Insert jobs to run for 7 days of simulated time. Do you save carbon by clocking down nodes when the carbon intensity of the grid is forecast to be high?

No Changes

```
=====
Summary
=====

Total Simulated-time Duration      : 168.0 hours
Total Real-time Duration           : 156.0 minutes

Jobs Started                       : 466536
Jobs Finished                      : 450576

Total CPU duration                 : 2285107.9 hours
Average CPU duration              : 4.90 hours

Total energy consumed by compute   : 10339.39 kWh
Peak time (5-9pm) energy consumption: 1649.79 kWh
Average energy consumption per job : 22.55 Wh

Estimated CO2e emissions          : 688.678 kg
Estimated Peak time CO2e emissions : 118.386 kg
Average CO2e emissions per job    : 1.502 g
Peak time CO2e emissions percentage: 17.190 %
```



University
of Glasgow

2024 ScotGrid Tier 2

Forecasted Clock-down

Each job produces 3% less CO₂

```
=====
Summary
=====

Total Simulated-time Duration      : 168.0 hours
Total Real-time Duration           : 174.3 minutes

Jobs Started                       : 392392
Jobs Finished                      : 376432

Total CPU duration                 : 2313757.3 hours
Average CPU duration              : 5.90 hours

Total energy consumed by compute   : 8613.15 kWh
Peak time (5-9pm) energy consumption: 1243.06 kWh
Average energy consumption per job : 22.41 Wh

Estimated CO2e emissions          : 560.153 kg
Estimated Peak time CO2e emissions : 88.647 kg
Average CO2e emissions per job    : 1.457 g
Peak time CO2e emissions percentage: 15.825 %
```

17% reduction in jobs

25% peak time energy reduction

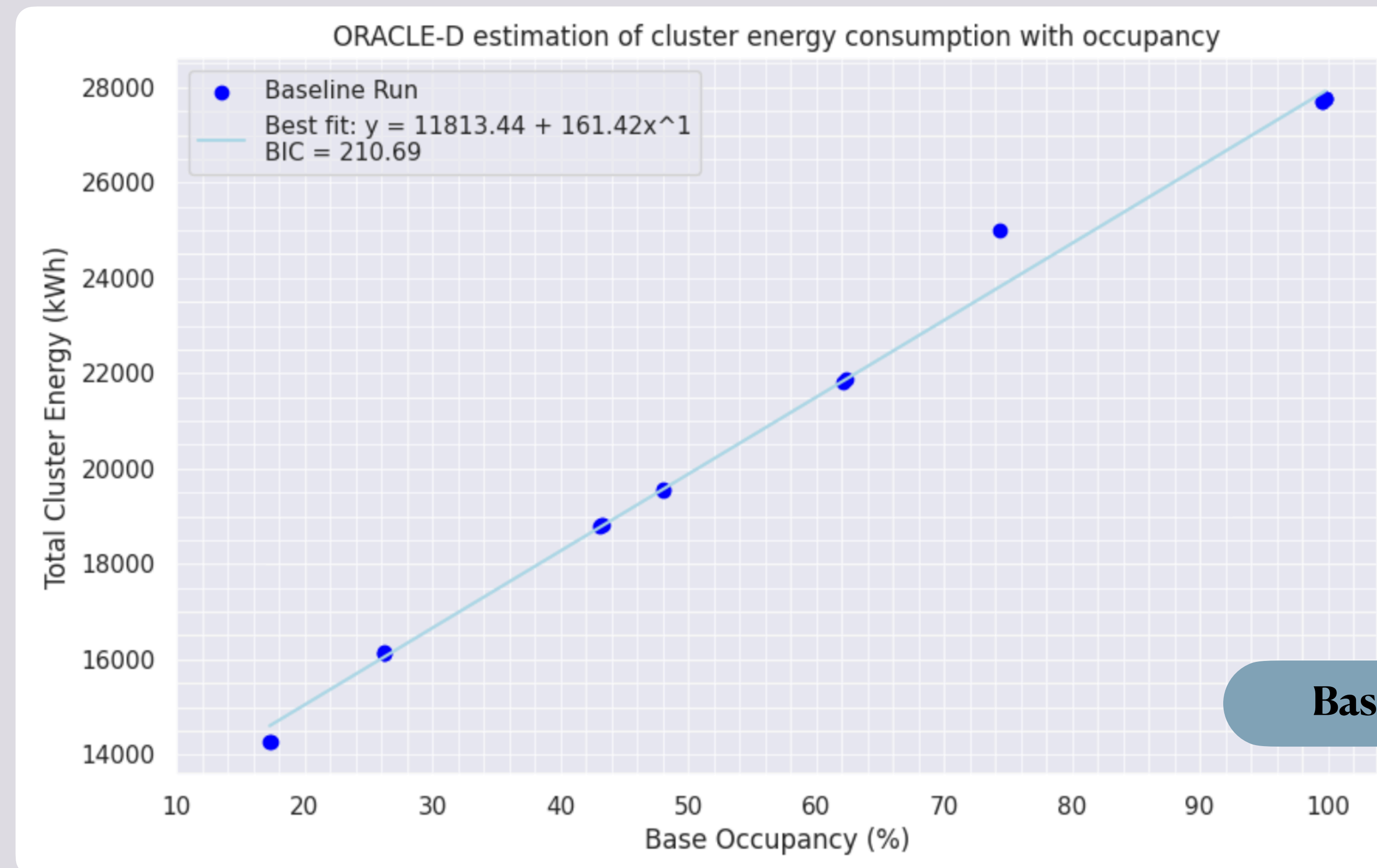
20% overall CO₂ reduction

Use Case 2 - How much energy can be saved by clocking down nodes?



- Insert jobs to run for 10 days of simulated time on our 'Grid' cluster.
- Energy Strategy - *How much energy is used (can be saved) by reducing power to during peak times (12:00-15:00)*

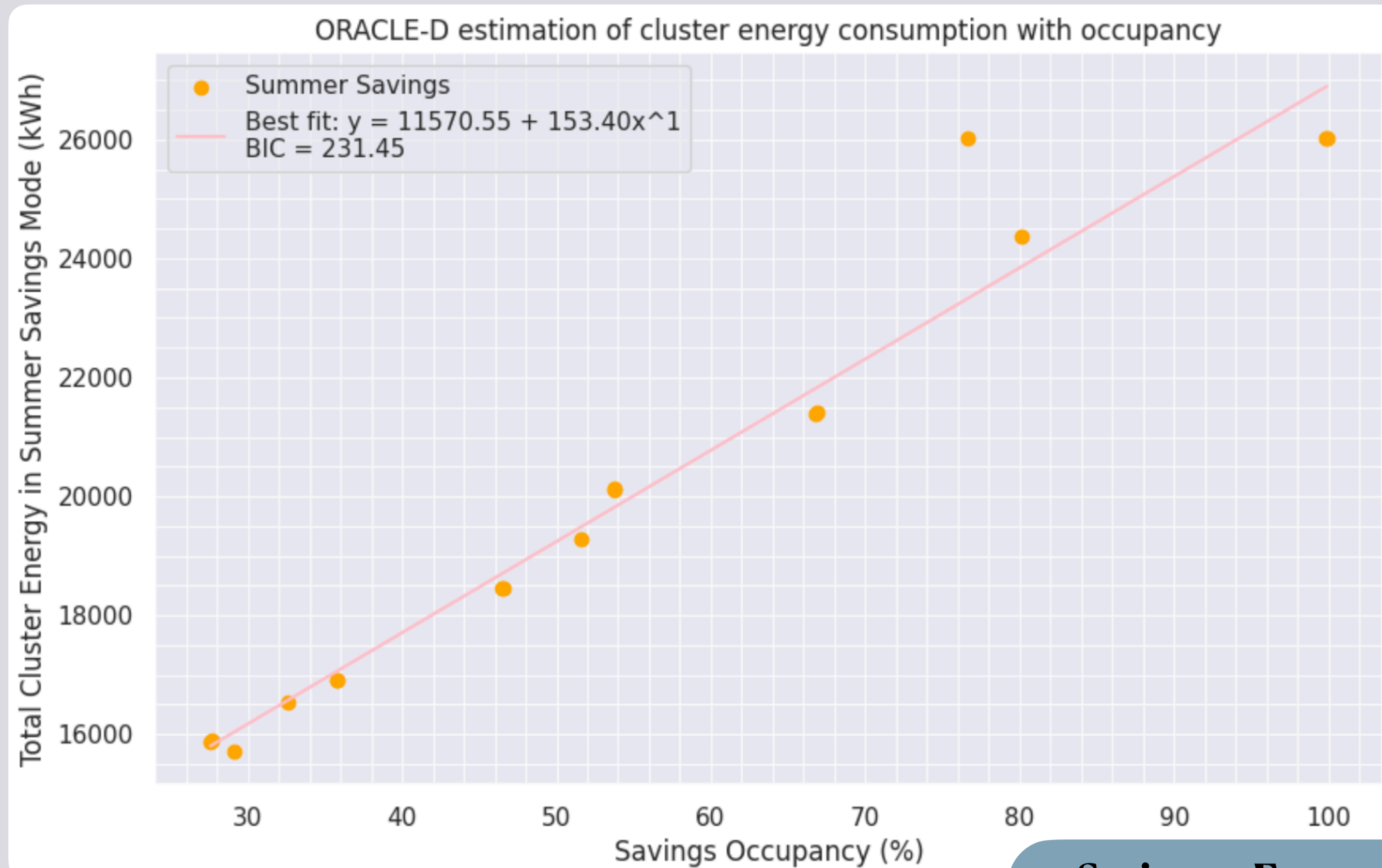
- How much power is used depends on how 'full' the cluster is - Occupancy
- The 'job' mixture sent to the simulation is changed to simulate different occupancies



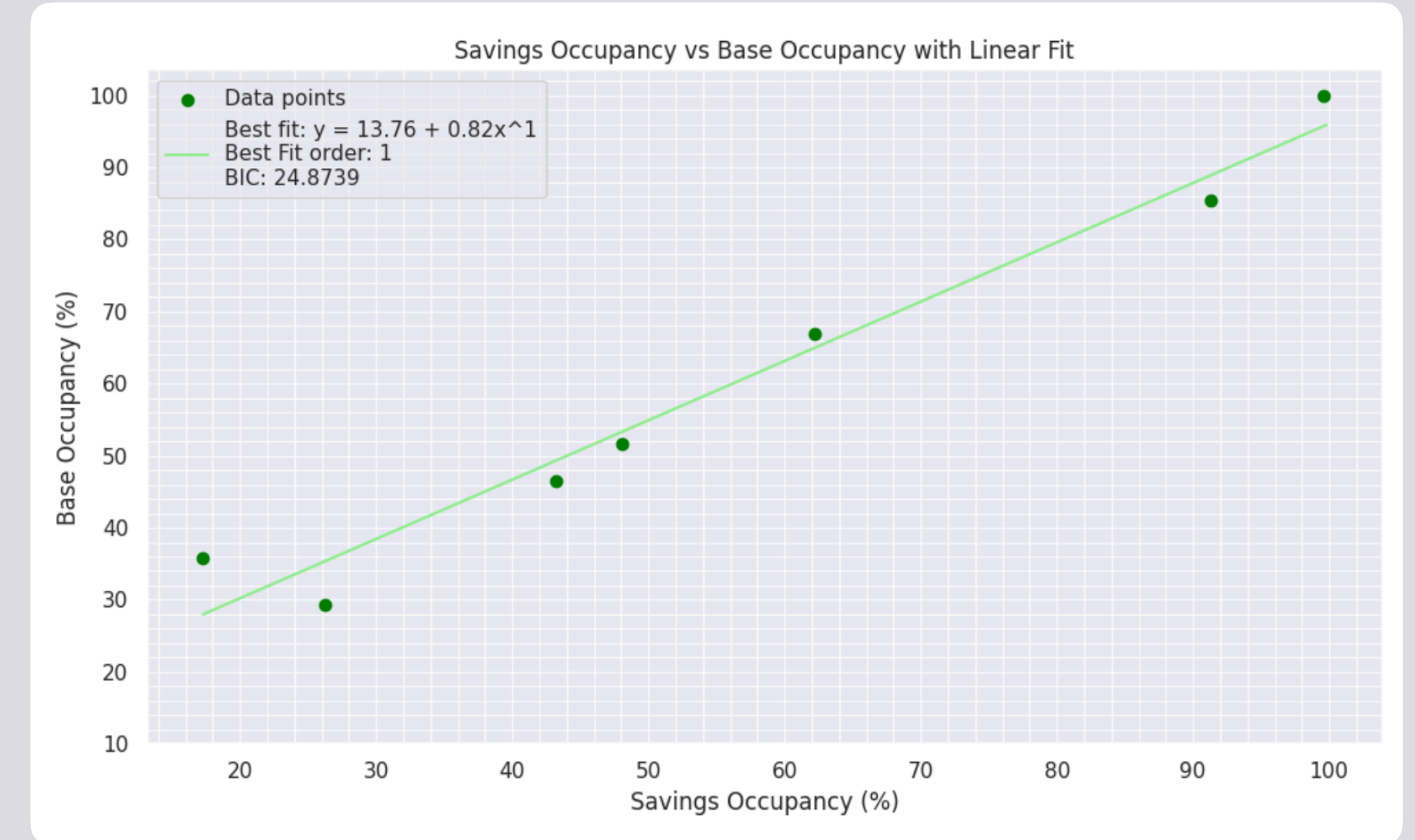
Base = Normal Operation

Use Case 2 - How much energy can be saved by clocking down nodes?

- Run simulation in Base mode and Savings mode with the same job mixes.



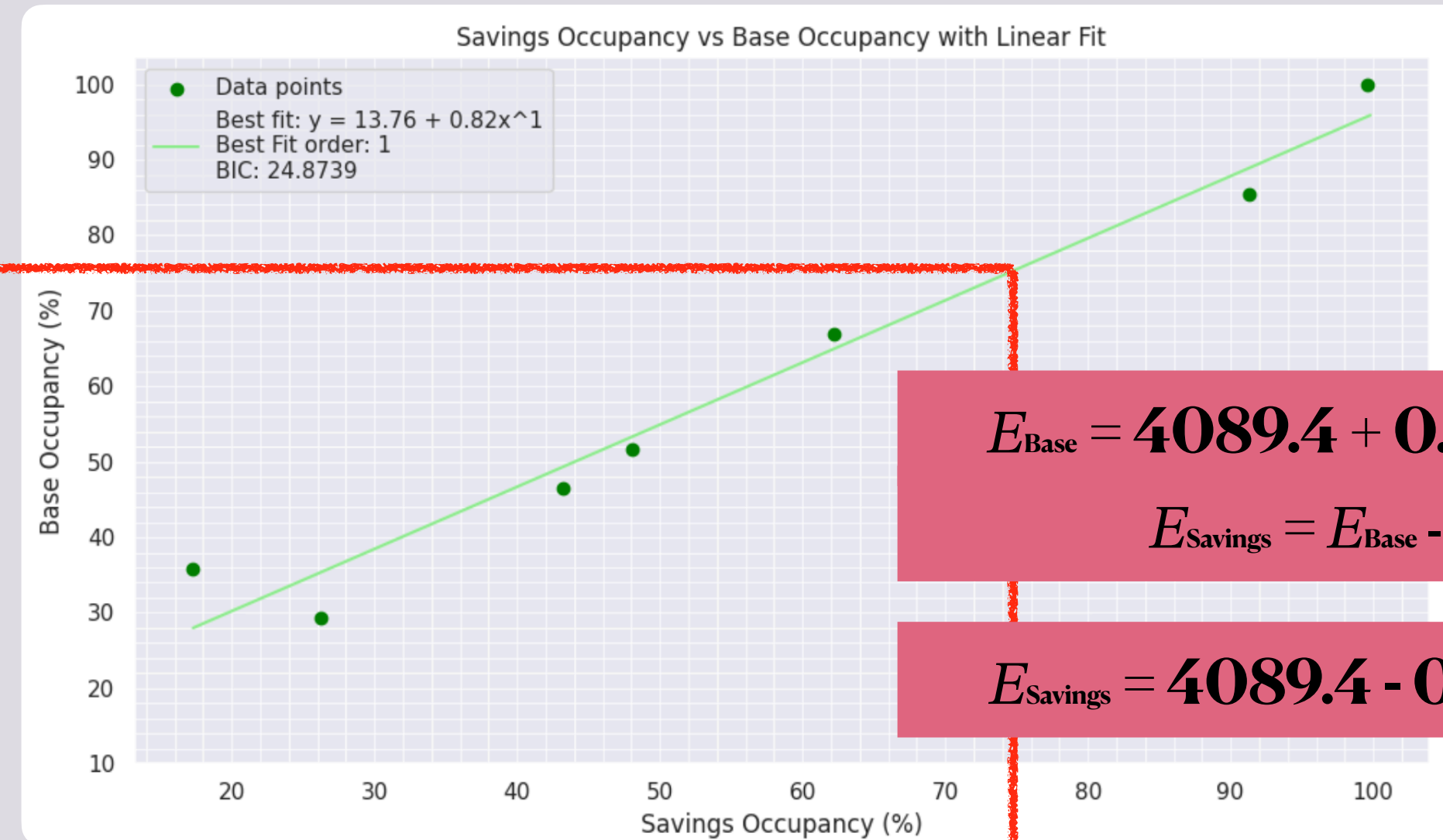
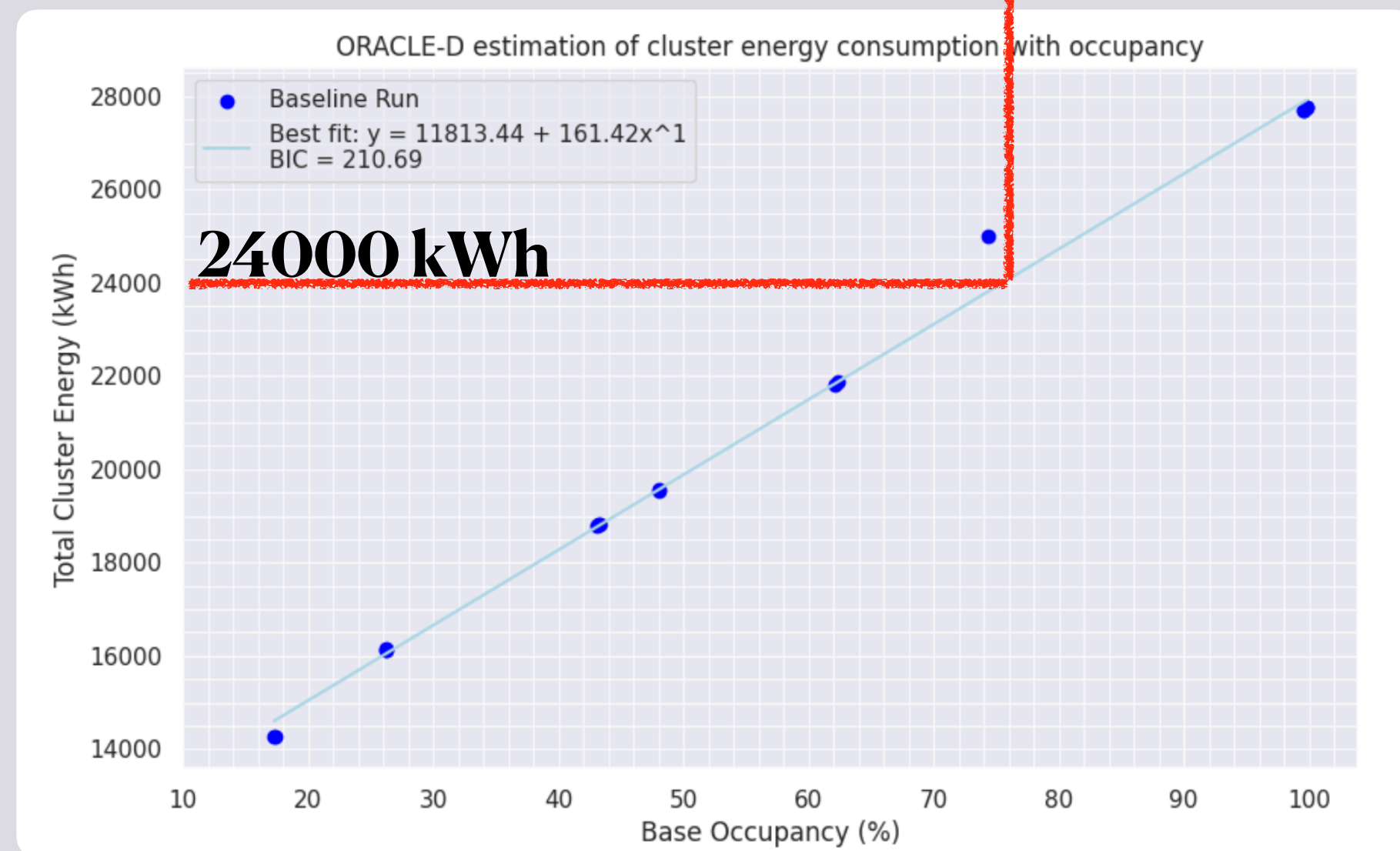
Savings = Energy Saving Operation



- Occupancy of the cluster for the same jobs is higher on the savings options and work is stretched out over time.
- Fit the occupancy of the Base and Savings operation to find estimated energy savings

Use Case 2 - How much energy can be saved by clocking down nodes?

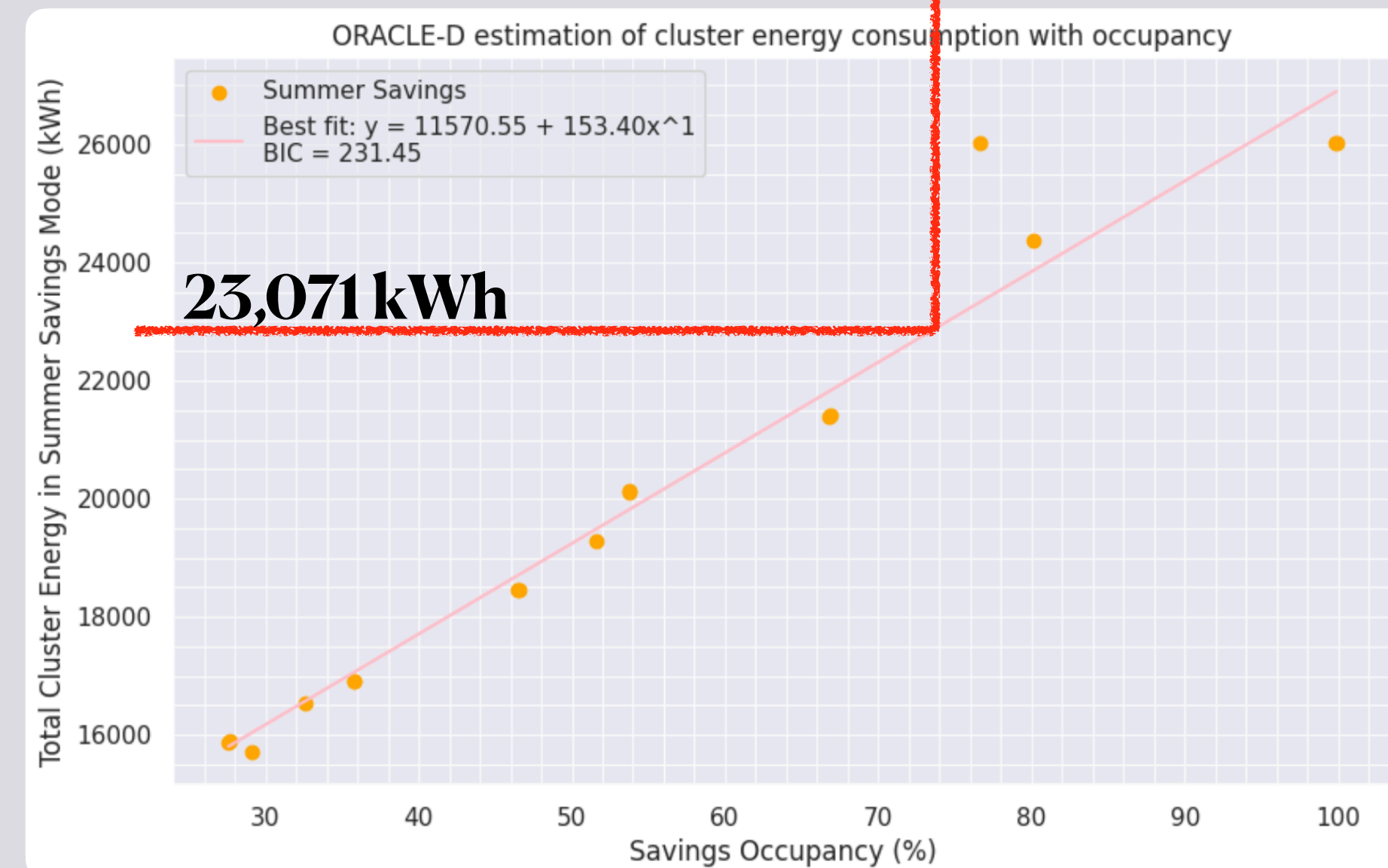
- All three linear fits can be combined to create a formula that converts the energy of Energy Policy run to it's equivalent Base run and visa versa.



$$E_{\text{Base}} = 4089.4 + 0.863 * E_{\text{Policy}}$$

$$E_{\text{Savings}} = E_{\text{Base}} - E_{\text{Policy}}$$

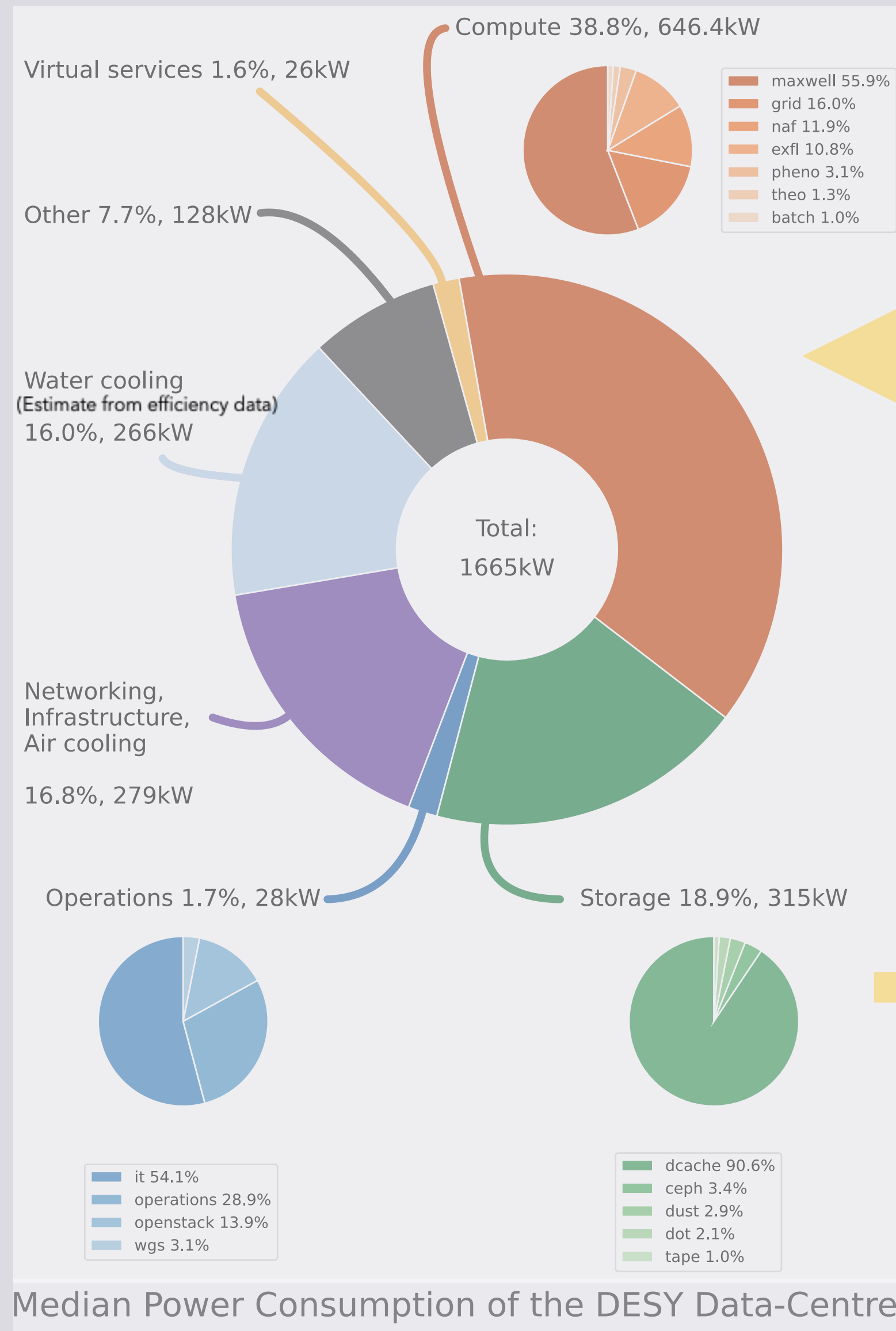
$$E_{\text{Savings}} = 4089.4 - 0.137 * E_{\text{Policy}}$$



- Example:
Normal operation - 24,000 kW
Energy Policy run - 23,071 kW
Expected savings 928 kW

* EU-funded project tasked with investigating how large research infrastructures can be more sustainable.

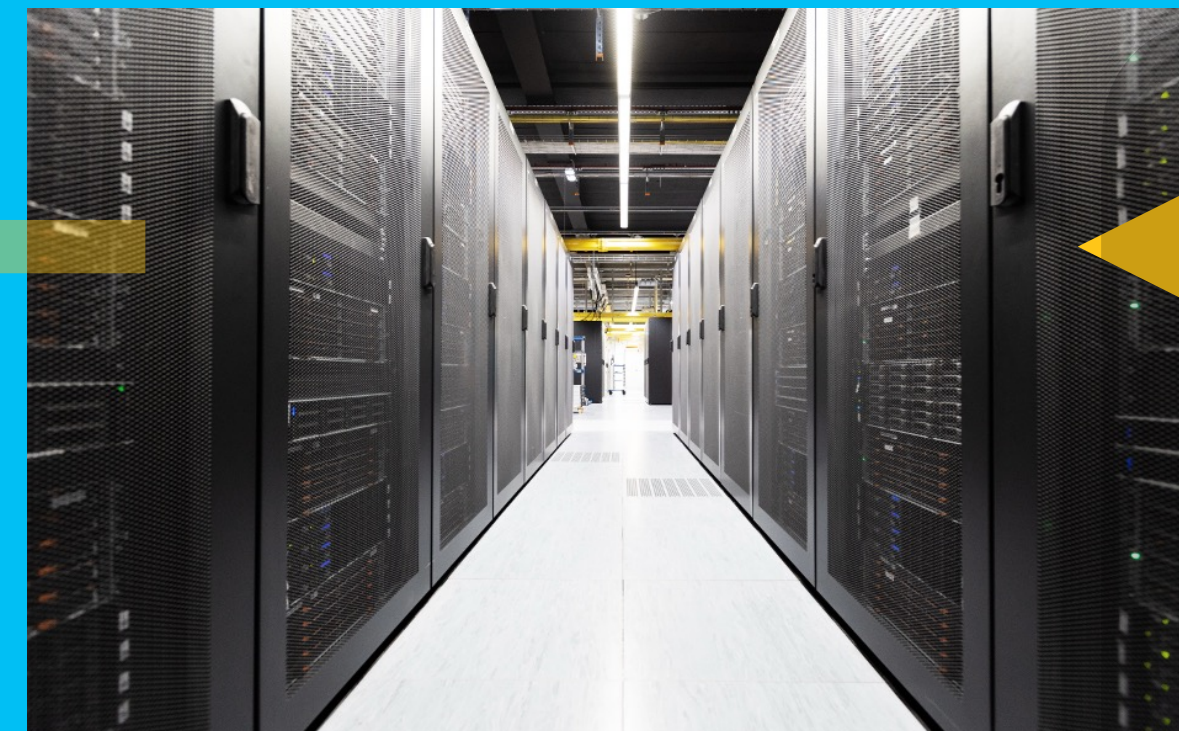
* At DESY we focus on what this means for current and future operation (and construction) of green data-centres.



1. Learn

Take detailed and fine-grained measurements of data-centre components

DESY DATA CENTRE

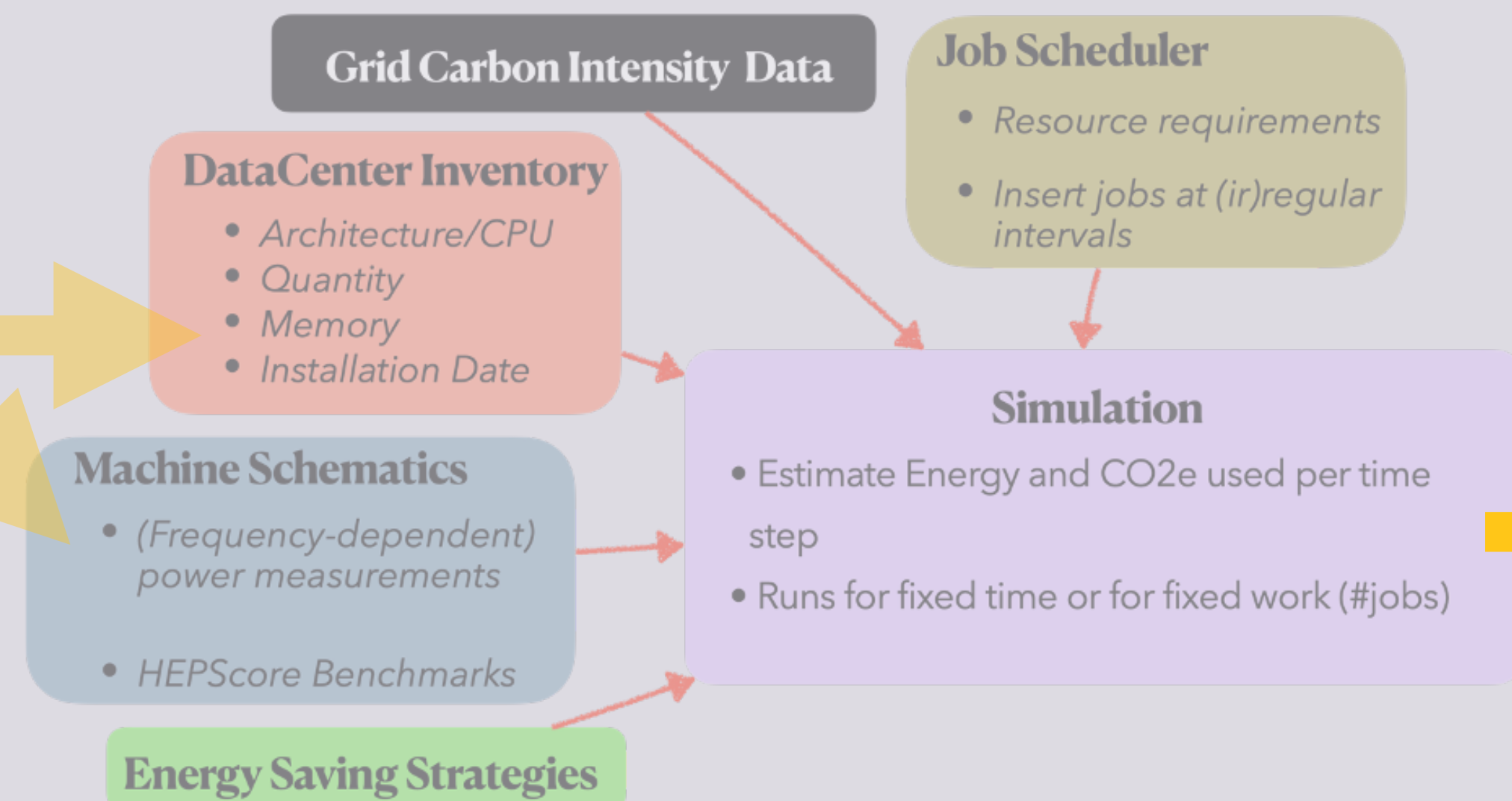


3. Refine

Take simulation findings and dynamically implement them at the datacenter

2. Simulate

Test different energy-saving strategies/policies in a simulation of the data-centre



3. Refine

Take what we learn and apply it - Summer Savings Programme at DESY

- We were asked to reduce the total power output of the data center by 15% during hottest part of the summer
- Different clusters serve different communities, and each have their own needs, and therefore their own demands - no one size fits all solution.

	GRID	NAF	MAXWELL
Primary Userbase	Remote Particle Physics Users	Local & National Particle Physics Users	Local & Guest Photon Science Users
Typical Usage	~95% utilisation Production jobs	70-75% utilisation, end user analysis	70-75% utilisation, HPC & memory intensive jobs
Scheduling Type	Partial Node (cores)	Partial Node (cores)	Whole/Multiple - Node
Preferred Energy Saving Measure	Frequency scaling	(Nice) Node shutdown	(Nice) Node shutdown

Summer Savings at DESY: Results

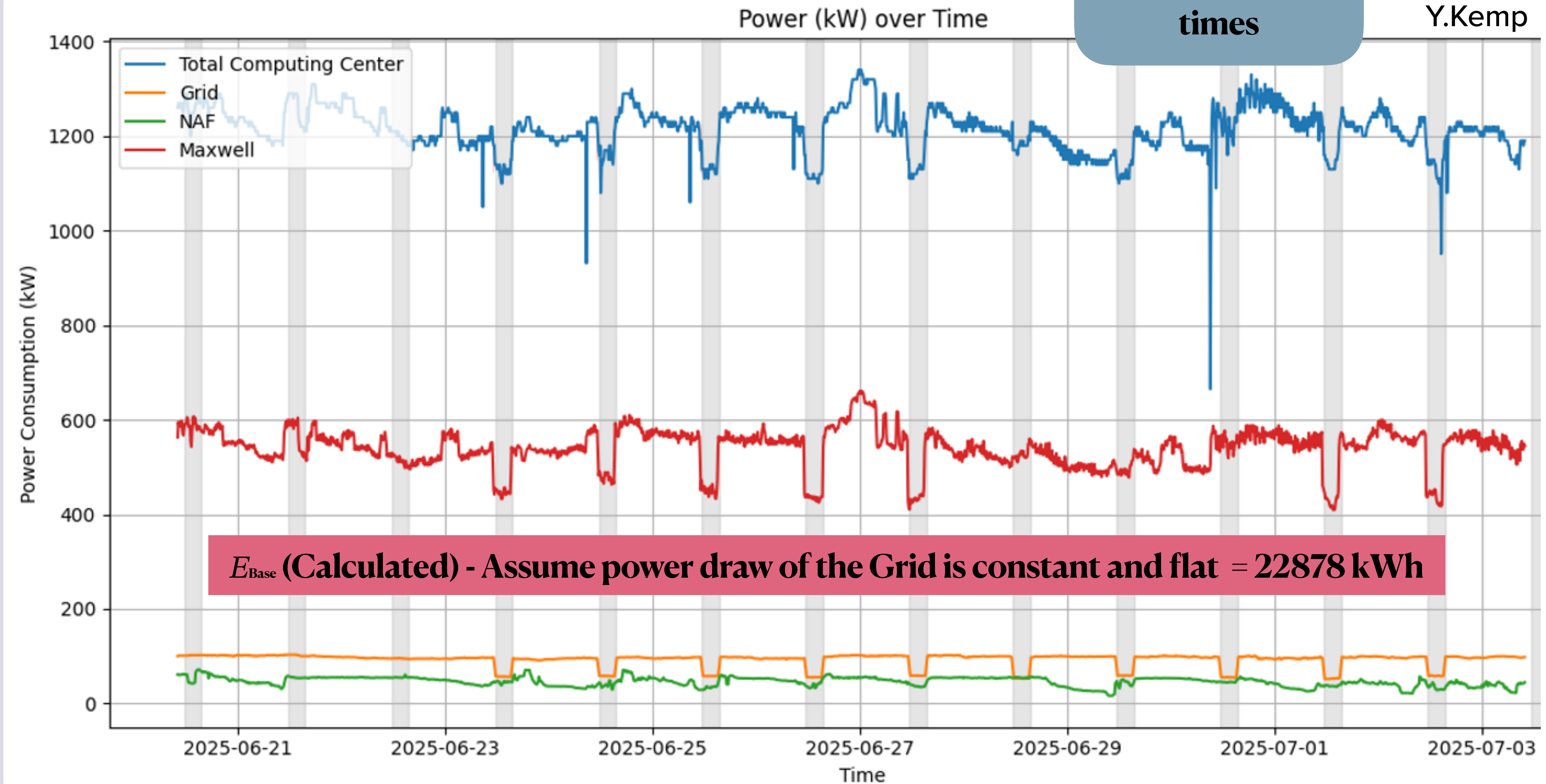
- Amount saved different for different clusters.
- In the Saving scheme, how much energy was saved doing this for the Grid cluster?

Total Measured Energy

$$E_{\text{Policy}} = 21818 \text{ kWh}$$

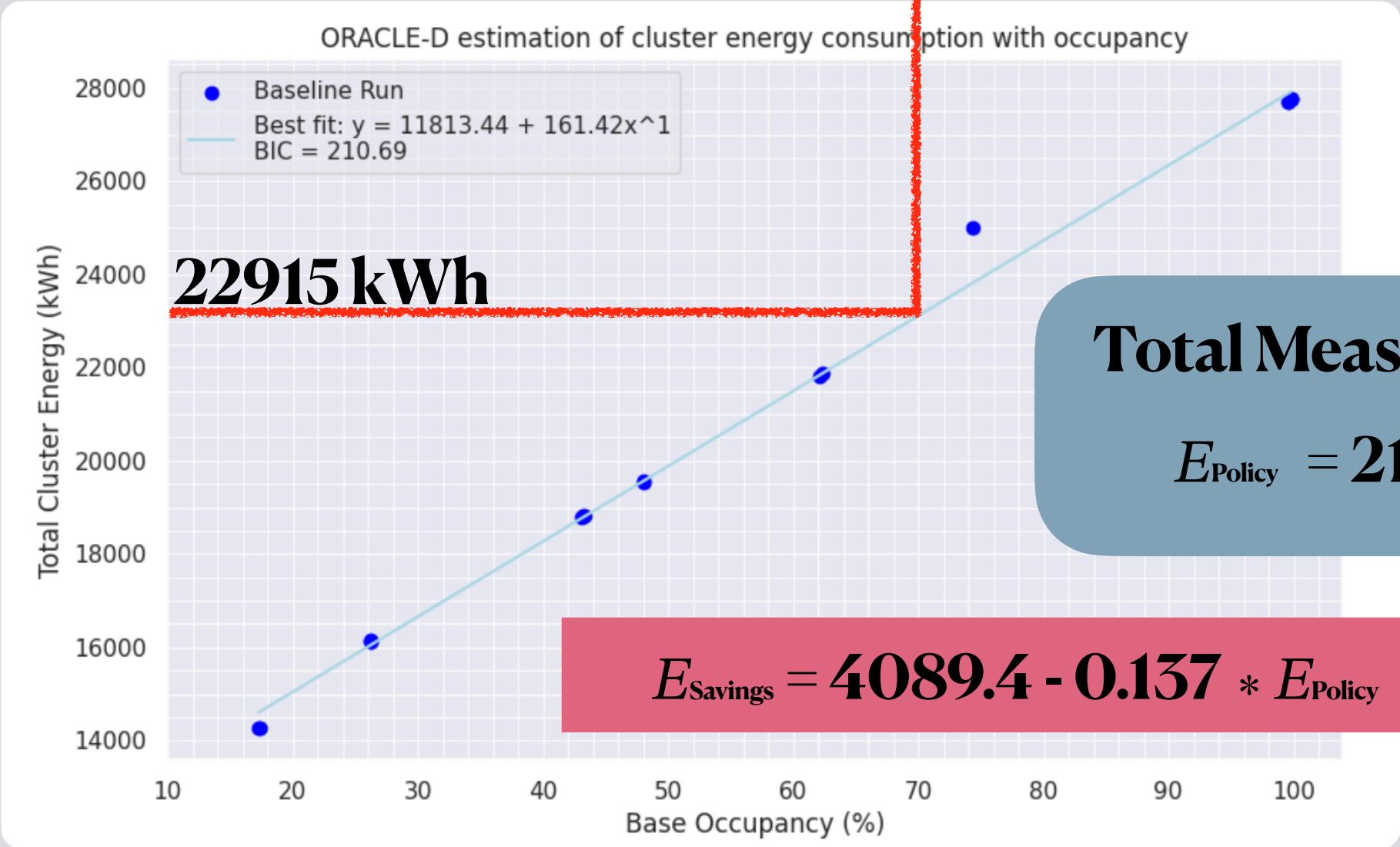
Grid Summer Savings

$$E_{\text{Savings}} = 1060 \text{ kWh}$$



Summer Savings at DESY: ORACLE-D Prediction

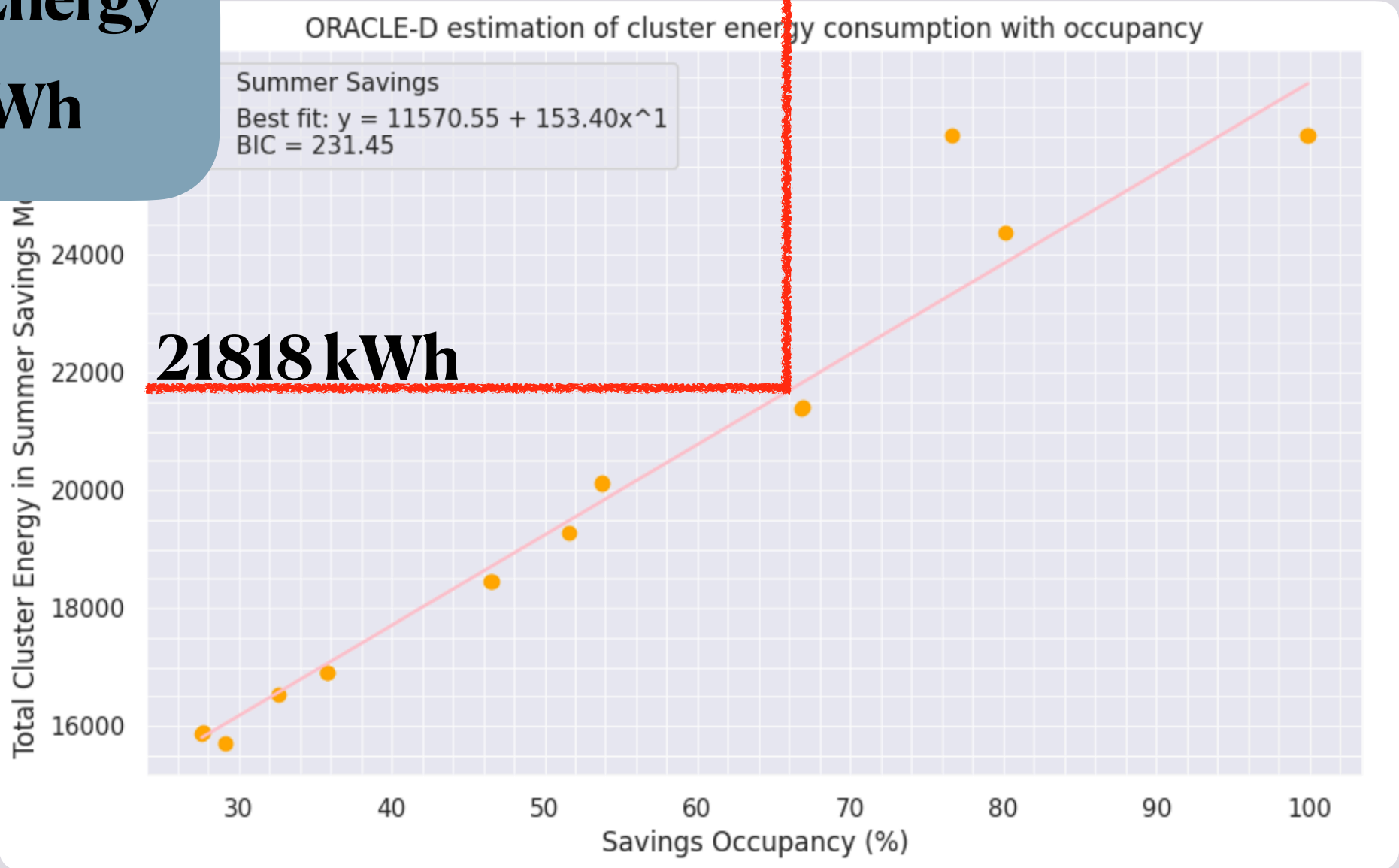
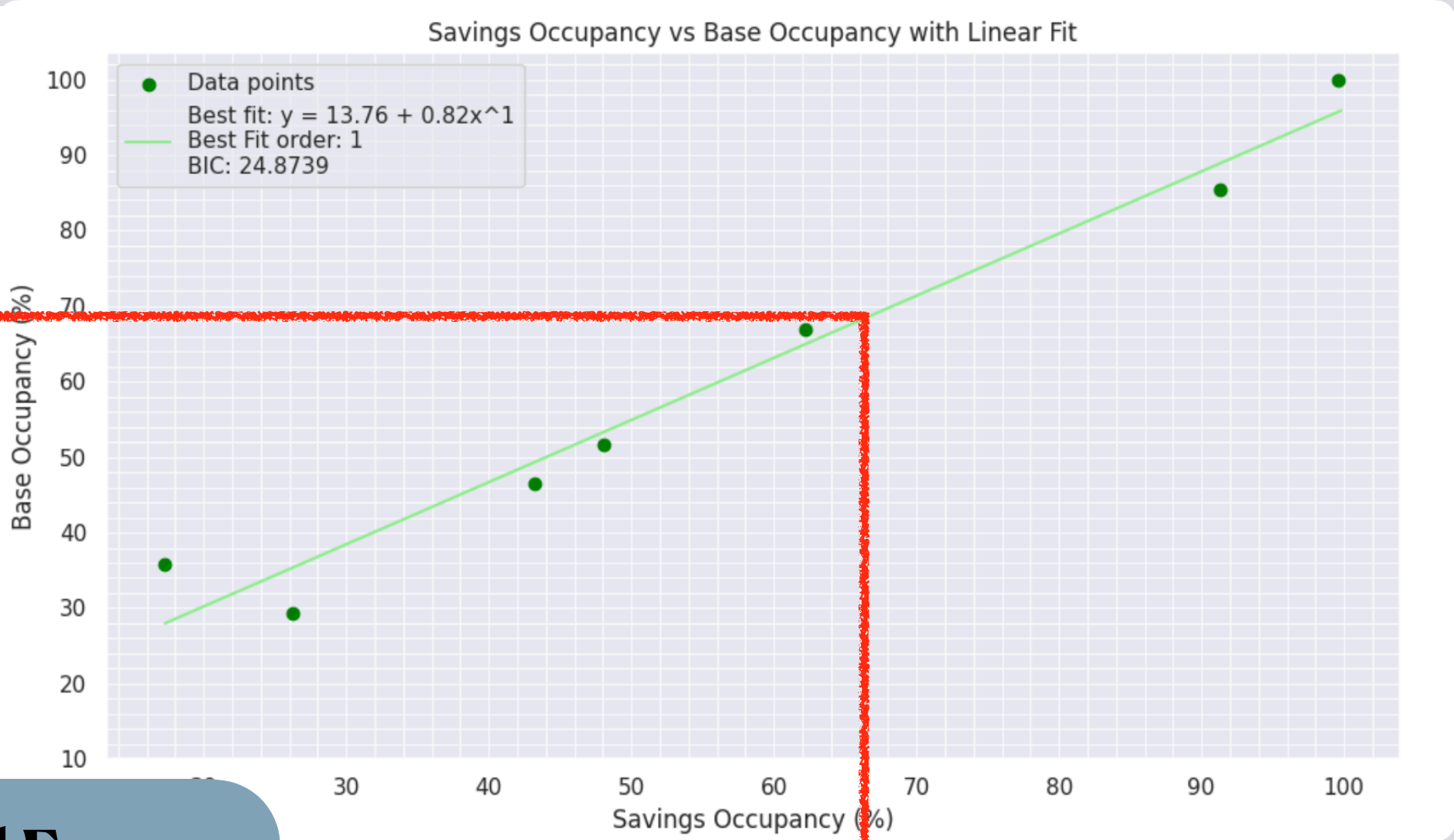
Insert whatever value of E_{Base} or E_{Policy} was measured, and subtract them to get the estimated savings



Total Measured Energy
 $E_{\text{Policy}} = 21818 \text{ kWh}$

$$E_{\text{Savings}} = 4089.4 - 0.137 * E_{\text{Policy}}$$

Grid (Simulated) Summer Savings
 $E_{\text{Savings}} = 1097 \text{ kWh}$

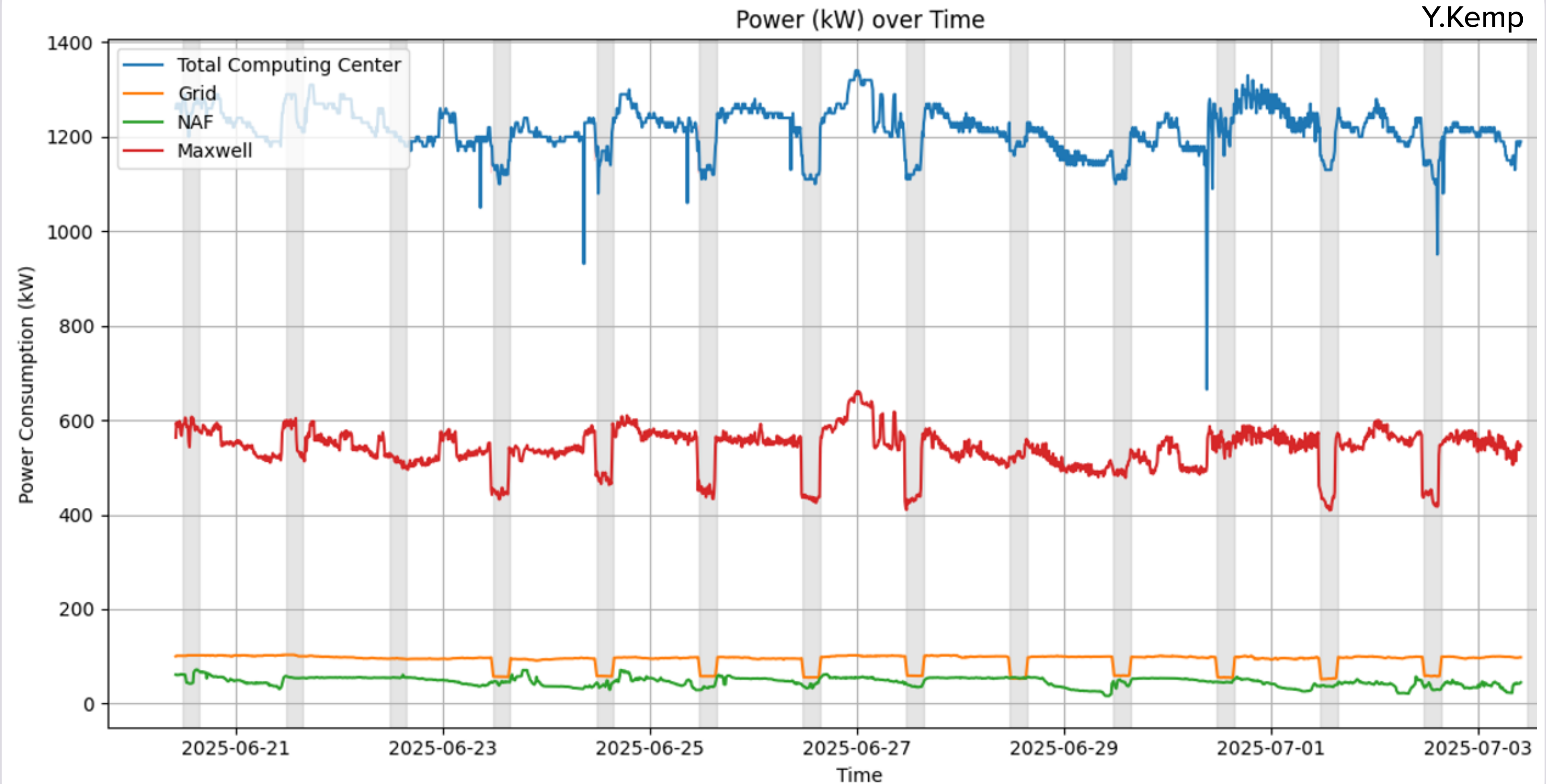


Summer Savings at DESY: Results

Across the week
saw **~3300 kWh**
saved across all
clusters

With minimal?
disruption to
research

GRID policy was
simulated by
ORACLE-D
beforehand and
predicted the
energy saving
amount to about 3%



What Next?



- We wanted to test some of the limits of what is currently possible
- A mini data centre to test the future of flexible computing
- We therefore need
 - Machines we own and control
 - to perform power-draw shaping based on availability of (green) energy
 - and automatically respond to external stimuli

POCCET

Proof of Concept Cluster for Emissions Tracking



Instances/services
rf20-wnXX

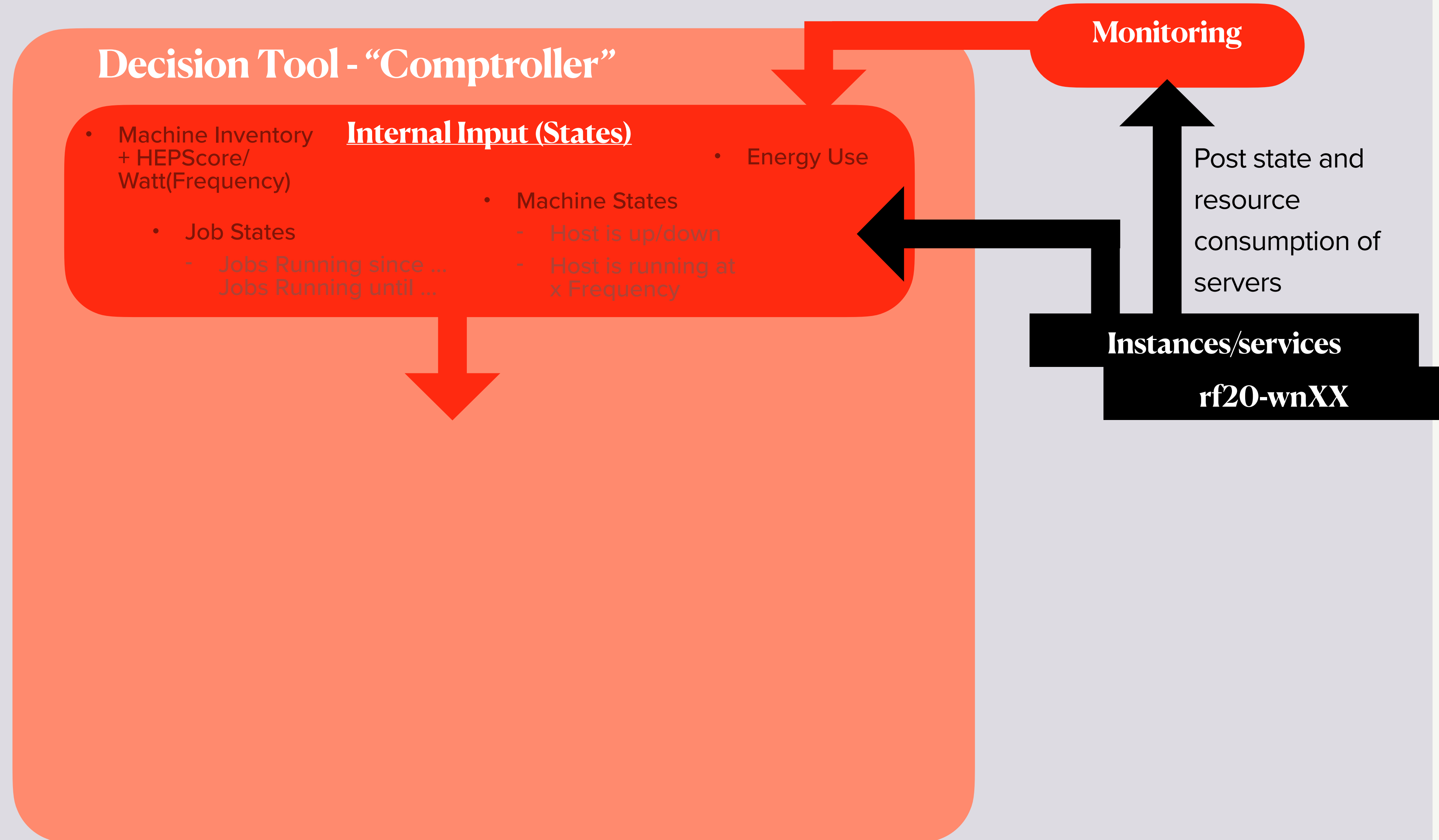
- 17 Dell PowerEdge R430 Machines in two racks
- 2 Ampere Altra (ARM based machines)
- 2 Power metering capable smart PDUs installed, reporting to *Graphite*, visualized in *Grafana*
- These PDU's have the ability to switch each port individually, and allow for more accurate energy/power draw measurements
- The hosts have been configured to be part of a separate VLAN, to not interfere with the rest of the data centre
- A basic HTCondor setup is provisioned via puppet, consisting of central manager, access point and execution points



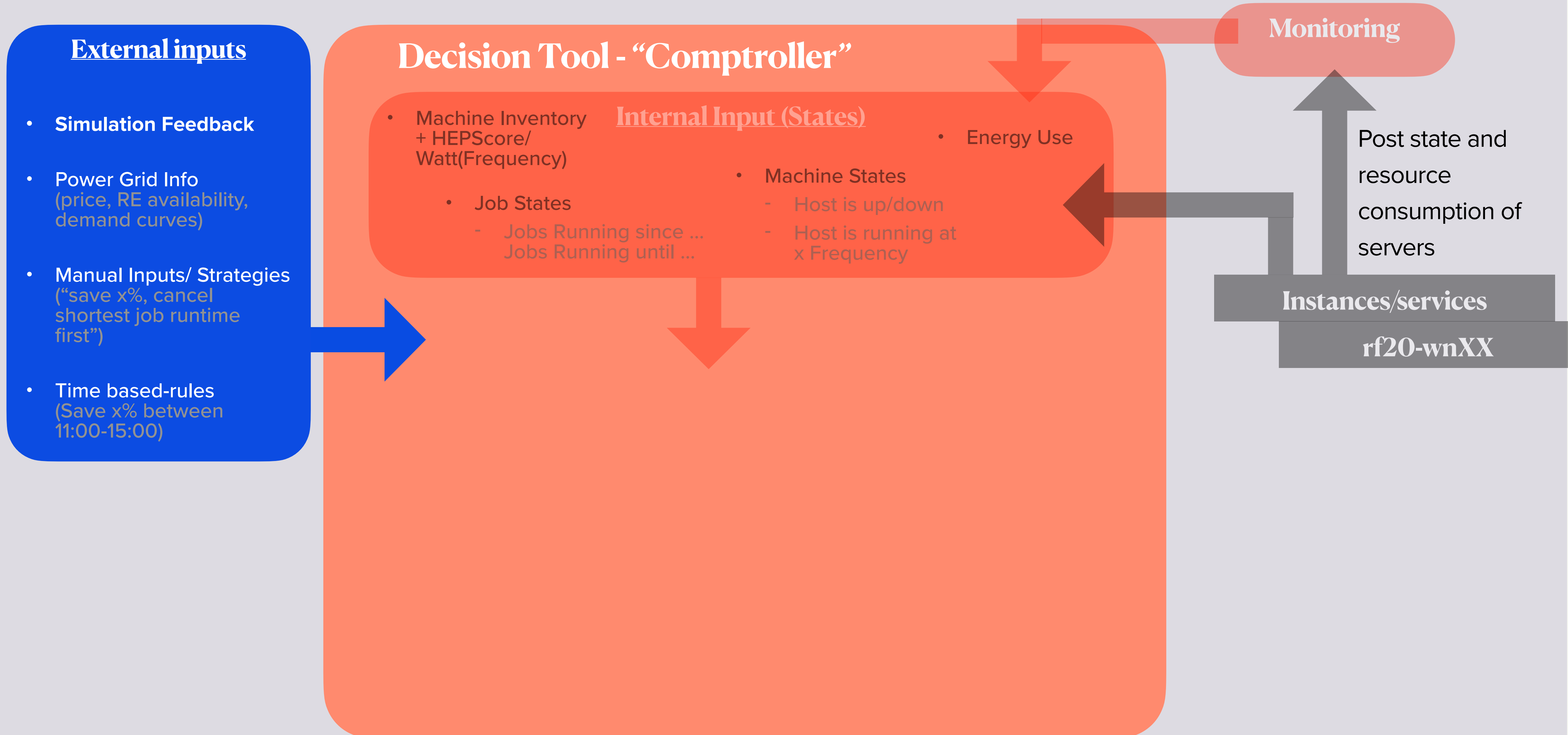
“The brains” of POC CET

Decision Tool - “Comptroller”

“The brains” of POC CET



“The brains” of POC CET

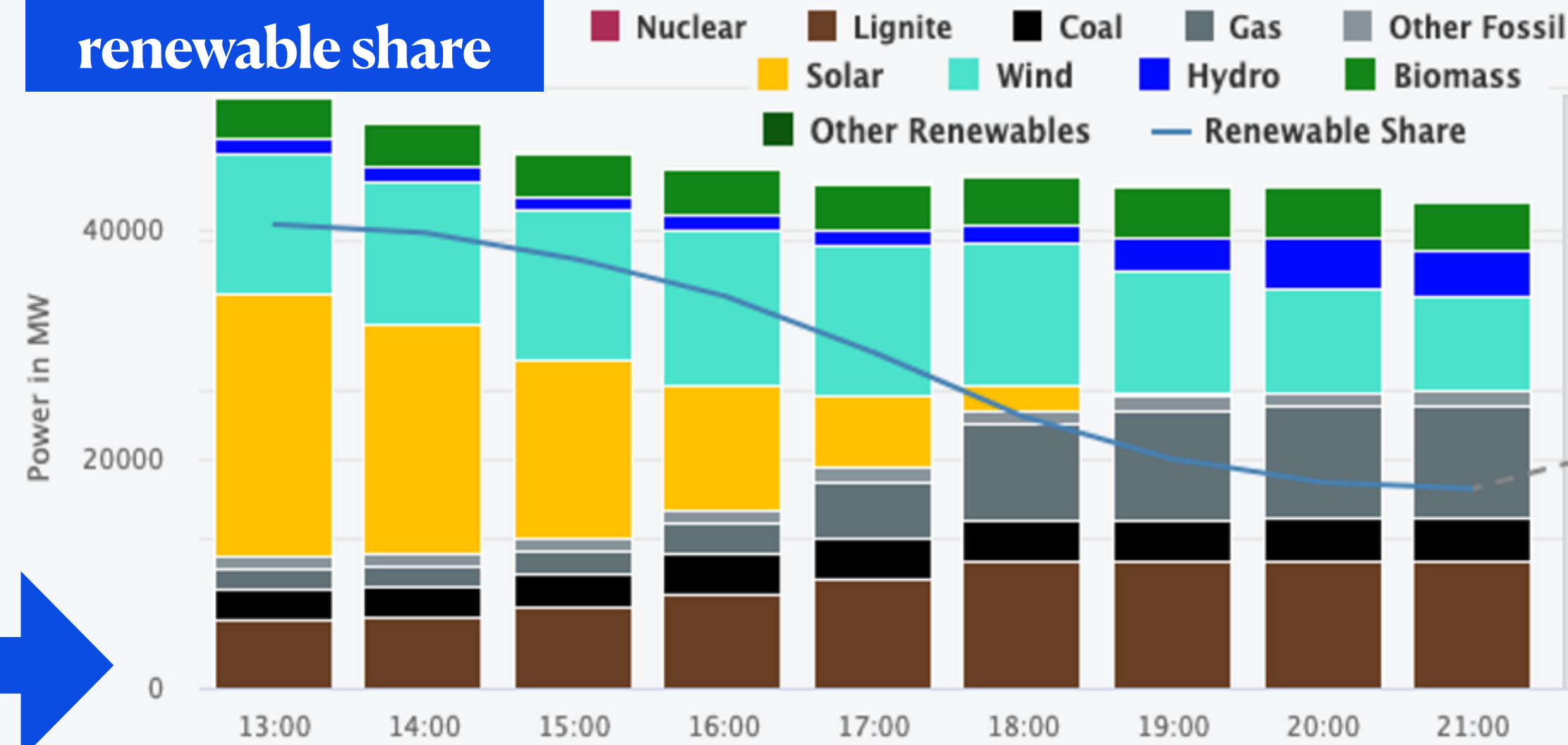


“The brains” of POC CET

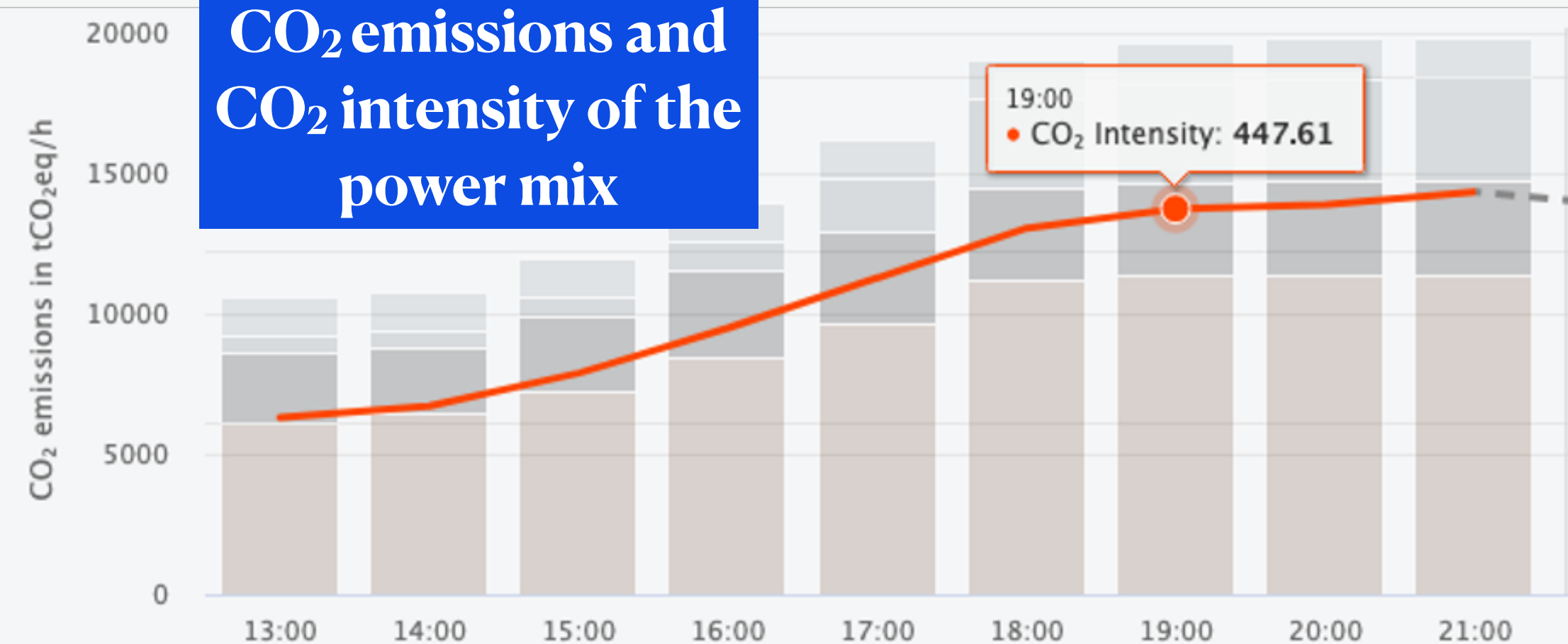
External inputs

- Simulation Feedback
- Power Grid Info (price, RE availability, demand curves)
- Manual Inputs/ Strategies (“save x%, cancel shortest job runtime first”)
- Time based-rules (Save x% between 11:00-15:00)

Power mix and renewable share



CO₂ emissions and CO₂ intensity of the power mix



Monitoring

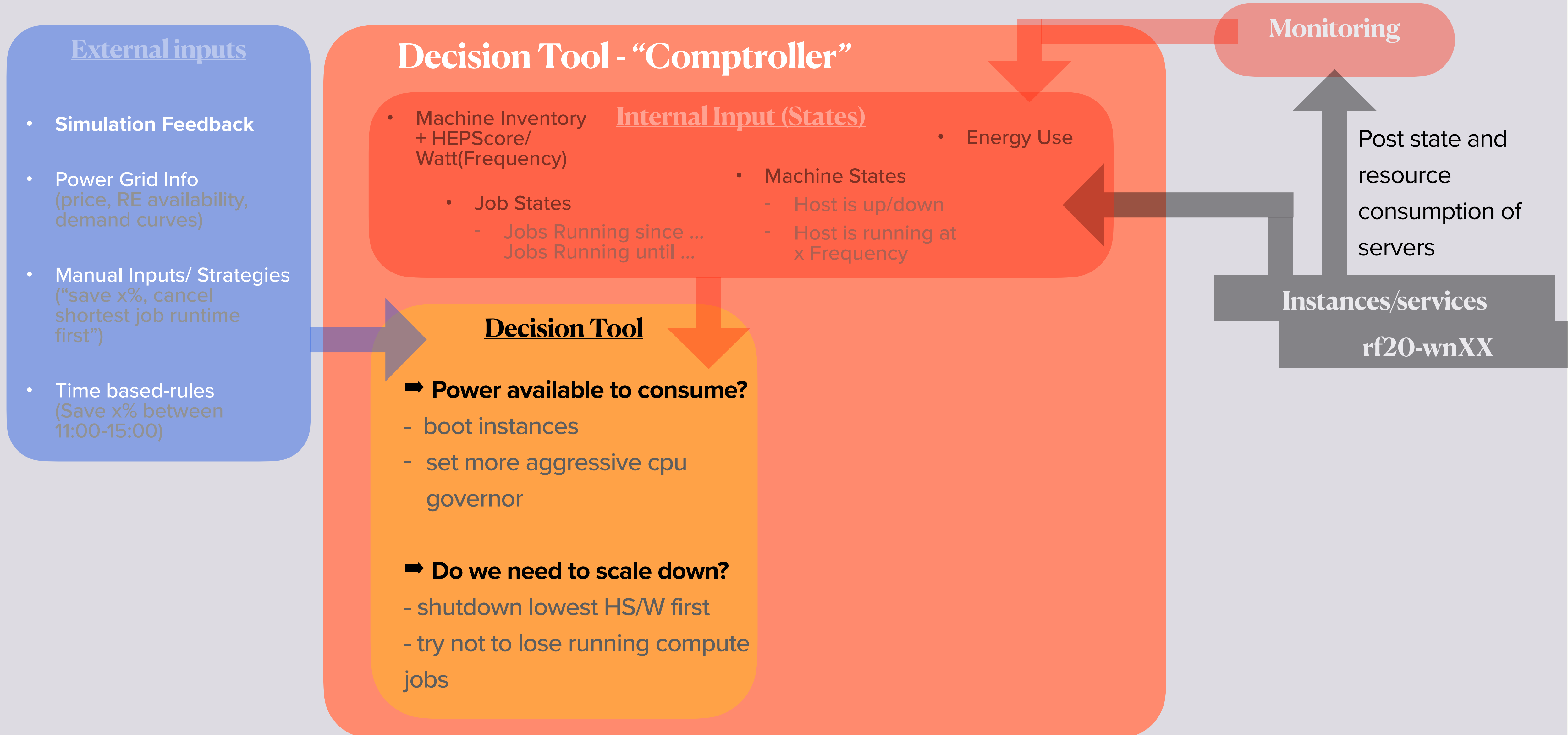
Post state and resource consumption of servers

Instances/services

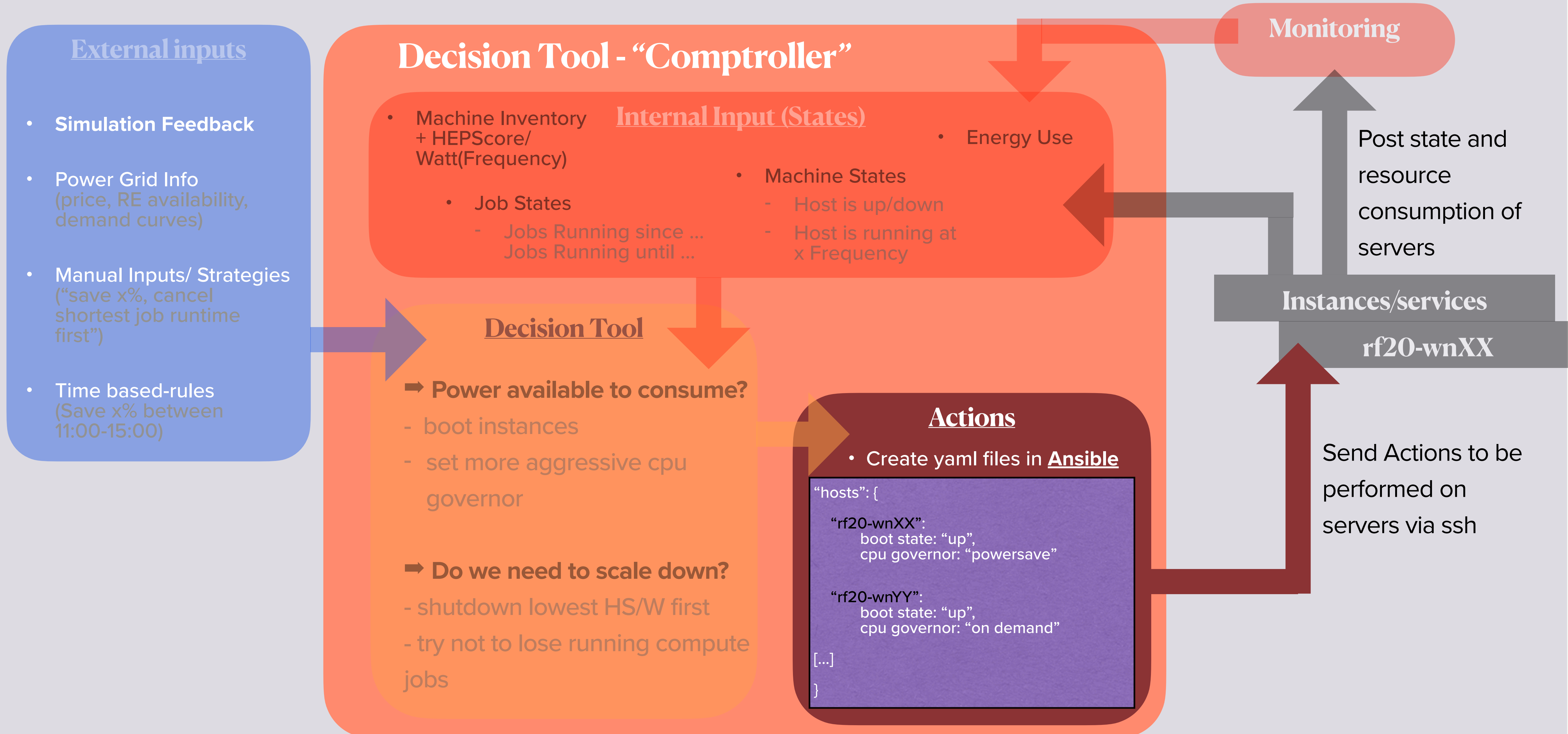
rf20-wnXX

•Both images adapted from: https://www.greengrid-compass.eu/zone/DE_LU

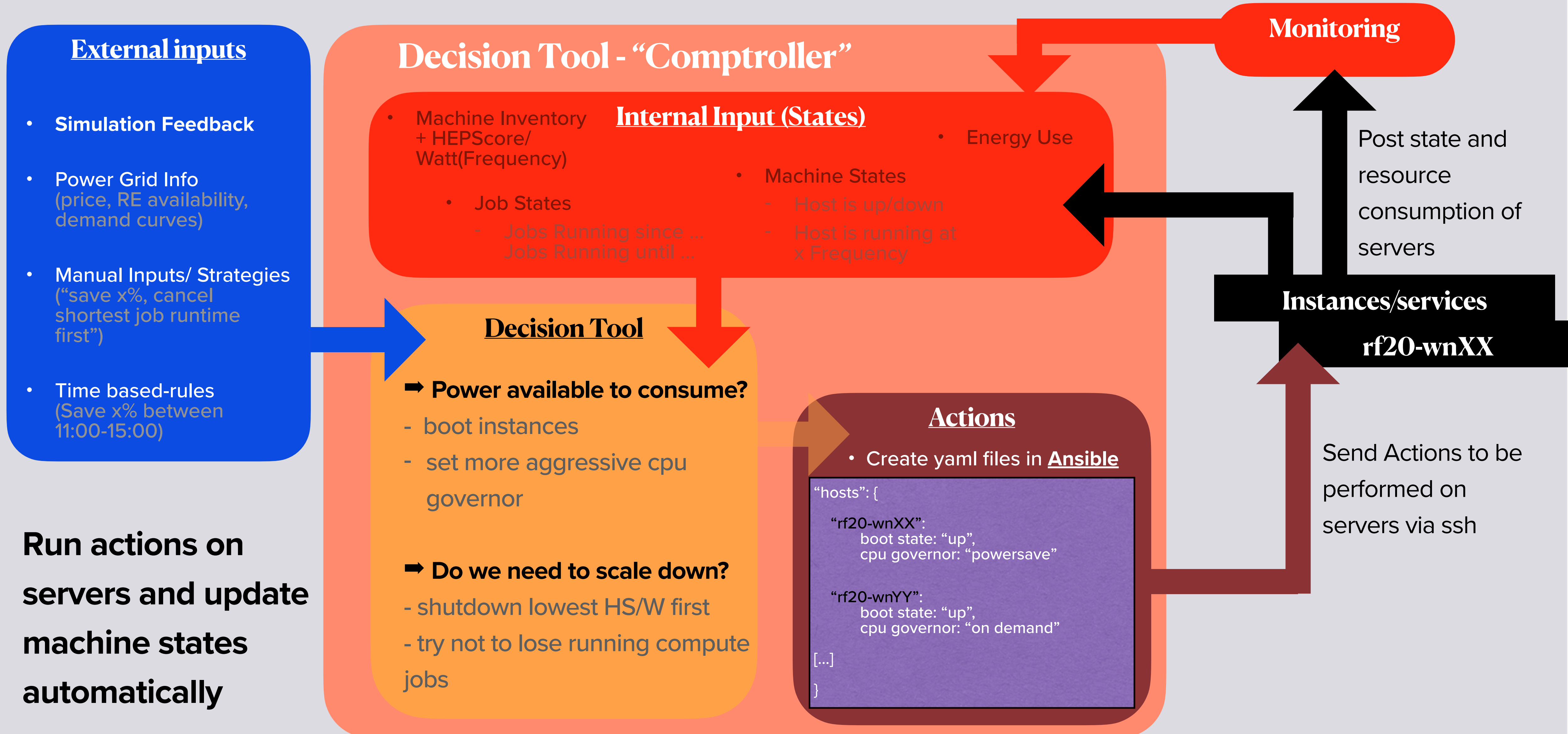
“The brains” of POC CET



“The brains” of POC CET

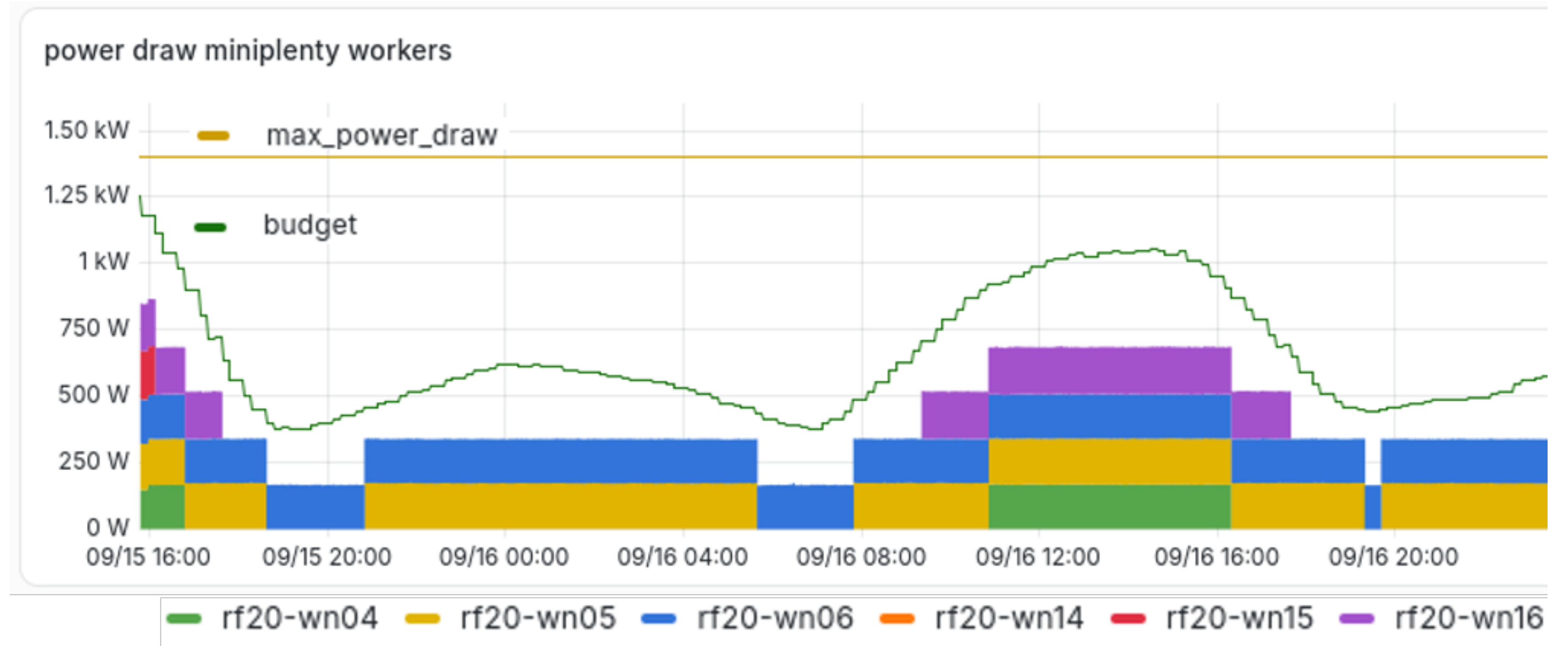


“The brains” of POC CET



Current Results

- Sculpting simple loads on the cluster possible
- Following a budget that depends on the cleanliness of the energy grid.



To Do

- More complicated loads
- Runtime 'buckets' to varying policies dependent on predicted job length
- More policies to turn into cluster actions

Conclusion

- DESY aims to achieve its scientific goals while **minimizing its environmental impact**
- The DESY RF2.0 Team works towards designing the **scientific data centres of the future**
- Within RF2.0 we have:
 - performed benchmarks and tests on machines to create a **variety of policies** that can be enacted at data centres to **reduce their climate impact**
 - expanded a **modular simulation framework** to simulate the operation of HTC clusters at DESY - and tested a first version against real data and **validated to within 3%****
 - built a test cluster POC CET to **test implementations** of these and more radical future policies **under real conditions**
- Long-Term Goal: We aim to **become an active participant in the power grid**



For further information
and to follow our
project progress visit
www.rf20.eu



and our Social Media accounts: RF2.0 Project @rf20_project



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101131850 and from the Swiss State Secretariat for Education, Research and Innovation (SERI).



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
**State Secretariat for Education,
Research and Innovation SERI**

Backup

Rethink

- The equipment we buy from vendors
- [Users] Do I submit this job?
- [Users] Attend FH Training Sessions
- [Exp. Facilities] Flexible work campaigns
- [Exp. Facilities] Quota changes
- [Exp. Facilities] Pilot lengths/functionality
- [Utilities] The times we run work

Refuse

- Unregistered/untrained users
- To run wasteful work

Reduce

- Compute node clock speeds
- [Vendors] Embedded carbon from servers
- [Utilities] Total energy used during high demand

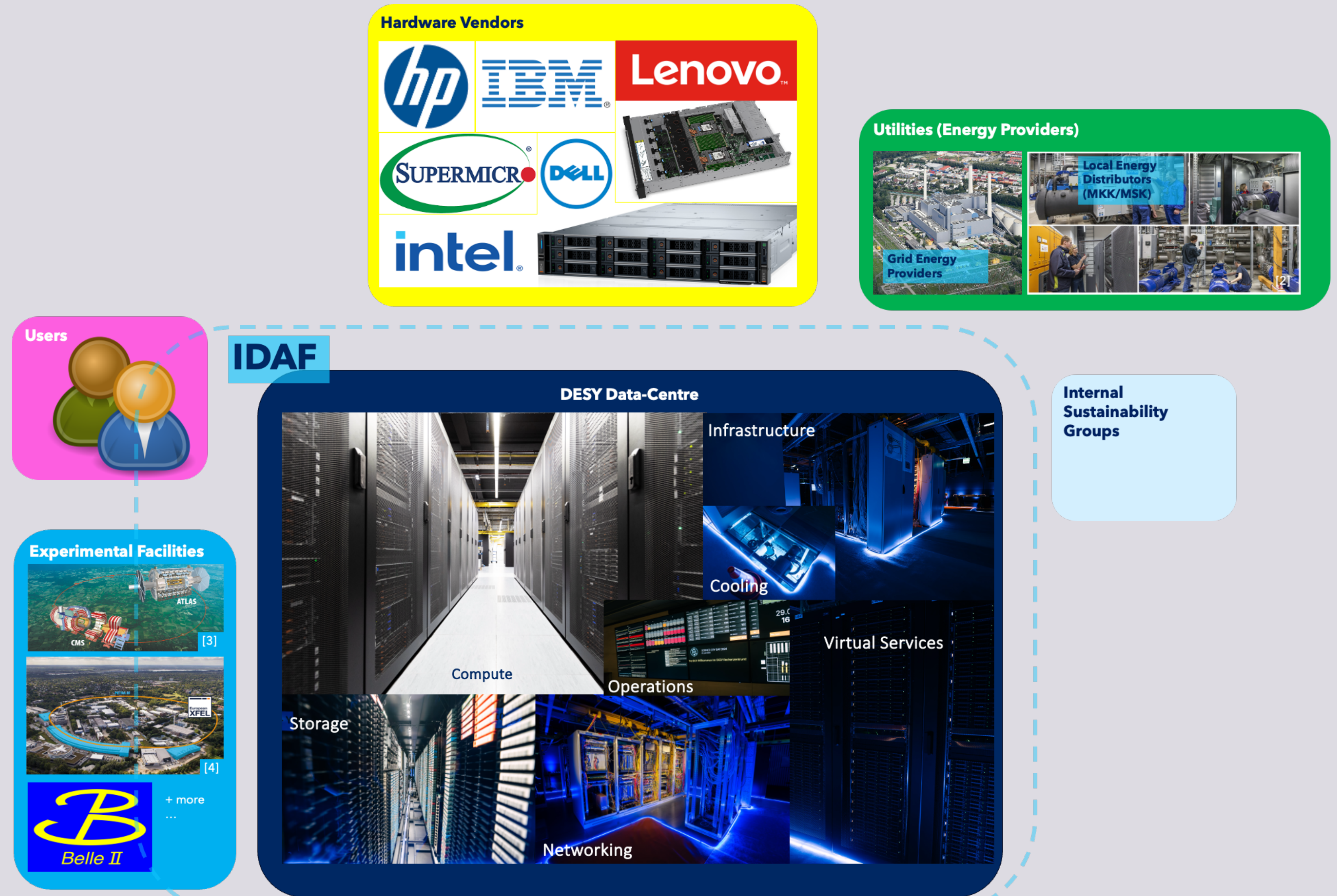
Reuse

- [Sust. Group] waste heat from servers
- Functioning parts from older machines

Repair/Regift/Recycle

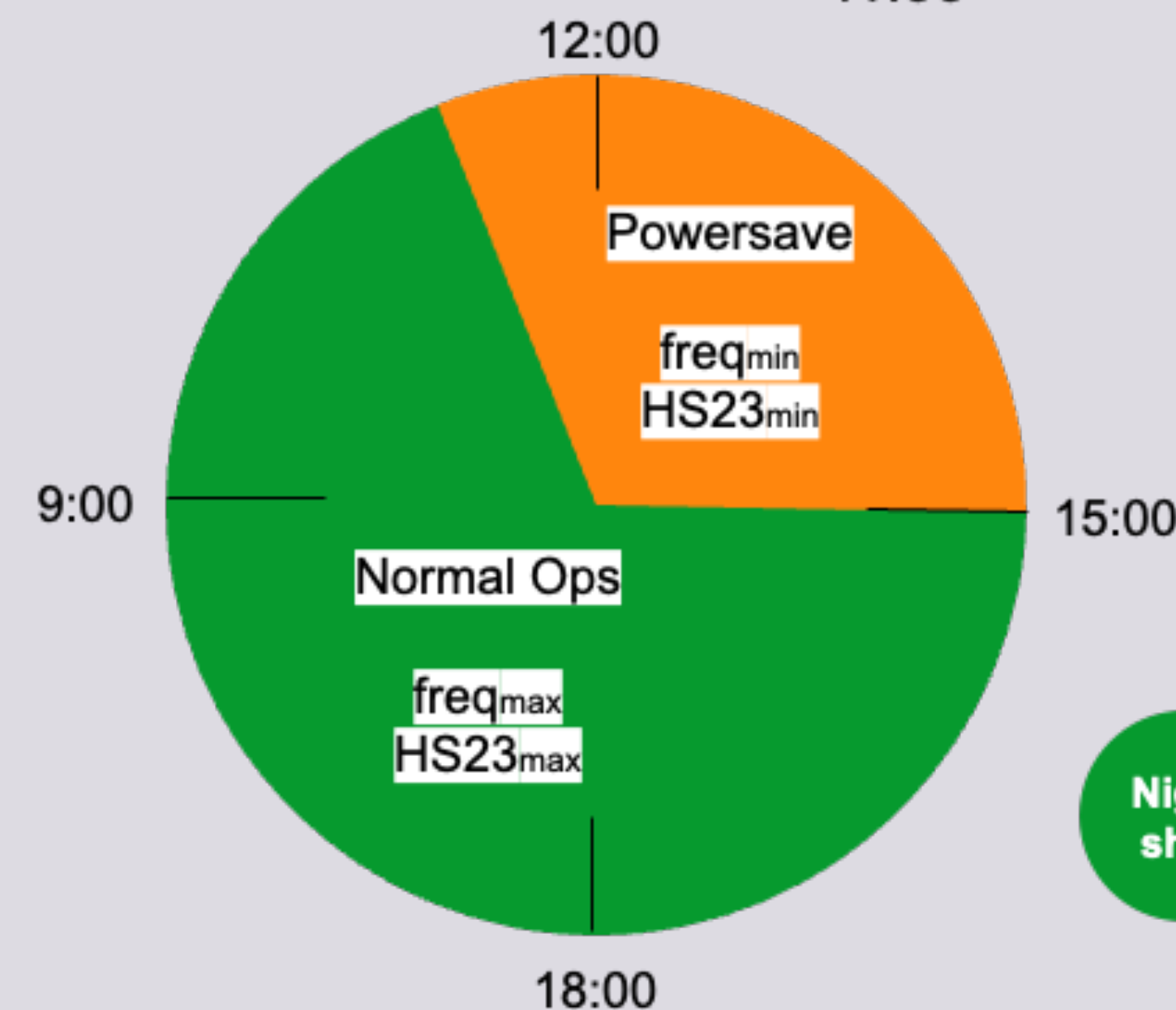
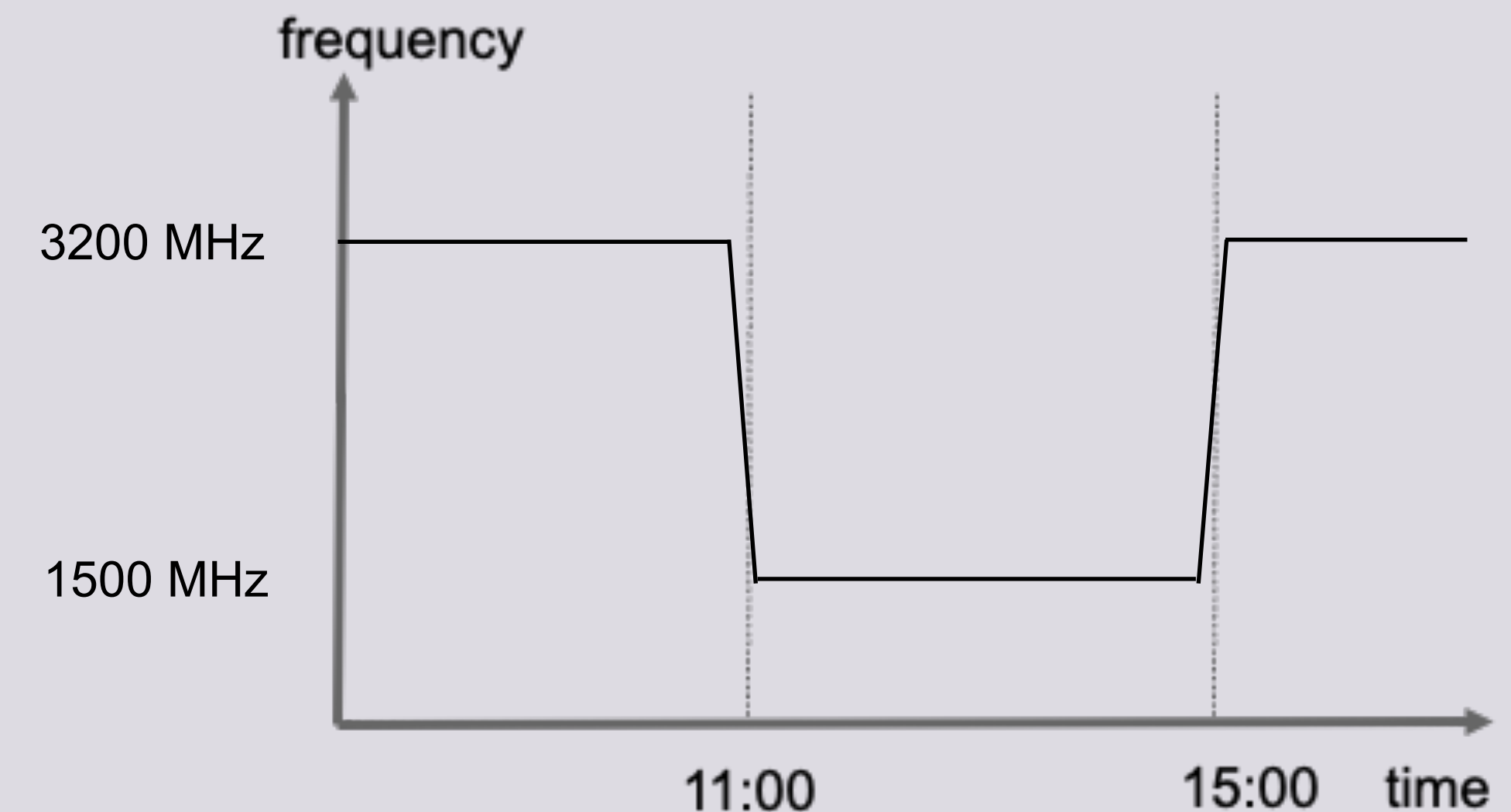
- Use old machines past warranty
- Donate old machines to other parts of the cluster
- [Vendors] Give old machines back to manufacturers

What can we do for Sustainability?



Summer Savings at DESY: The Grid Cluster

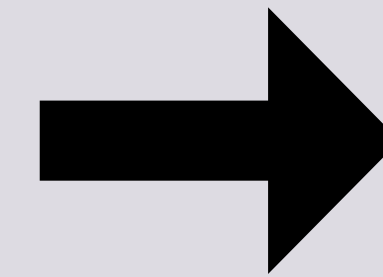
- Switch CPU Governors to powersave/min frequency and maybe pin CPU's to their minimum stepping
 - 40kW less power draw projected for a frequency capped grid cluster
- External resource - we end up delivering less computing.
 - Accounting-wise a rough correction is required for how much “HEPScore-Hours” we’re delivering to experiments.
 - Base Site Compute Delivery = $HS23_{\max}$
 - Summer Savings Site Compute Delivery = $20/24 HS23_{\max} + 4/24 HS23_{\min}$



T.Hartmann et al

Summer Savings at DESY: The NAF Cluster

- Refuse to run work that would be running between 11:00 and 15:00
- In addition - usage is lower typically. Hibernate instances when they are not used in HTCondor
- Typically happens at night
- Issues, to external monitoring, the site looks smaller. Turn-on only as needed?



Implement Condor Rooster Config to shut idle instances down, boot offline instances
→ use only what we need

- **Worker node config: when to go to sleep**
- **Rooster Daemon** (running on master): when and how to **wake nodes up**

Summer Savings at DESY: The Maxwell Cluster

- Whole nodes are selected by users, less fine controls on the servers themselves
- Take 20% of the cluster (older nodes) and schedule “maintenance slots” in advance between 11:00 and 15:00
- If no work run on machines in this time -> they can be shut down.
- Any work that would run during this time is not scheduled.

