Sustainable Computing (at DESY)

Dr Dwayne Spiteri, **Jan Hartmann**, *Yves Kemp*, Konrad Kockler, Martin Gasthuber, Kilian Schwarz et al

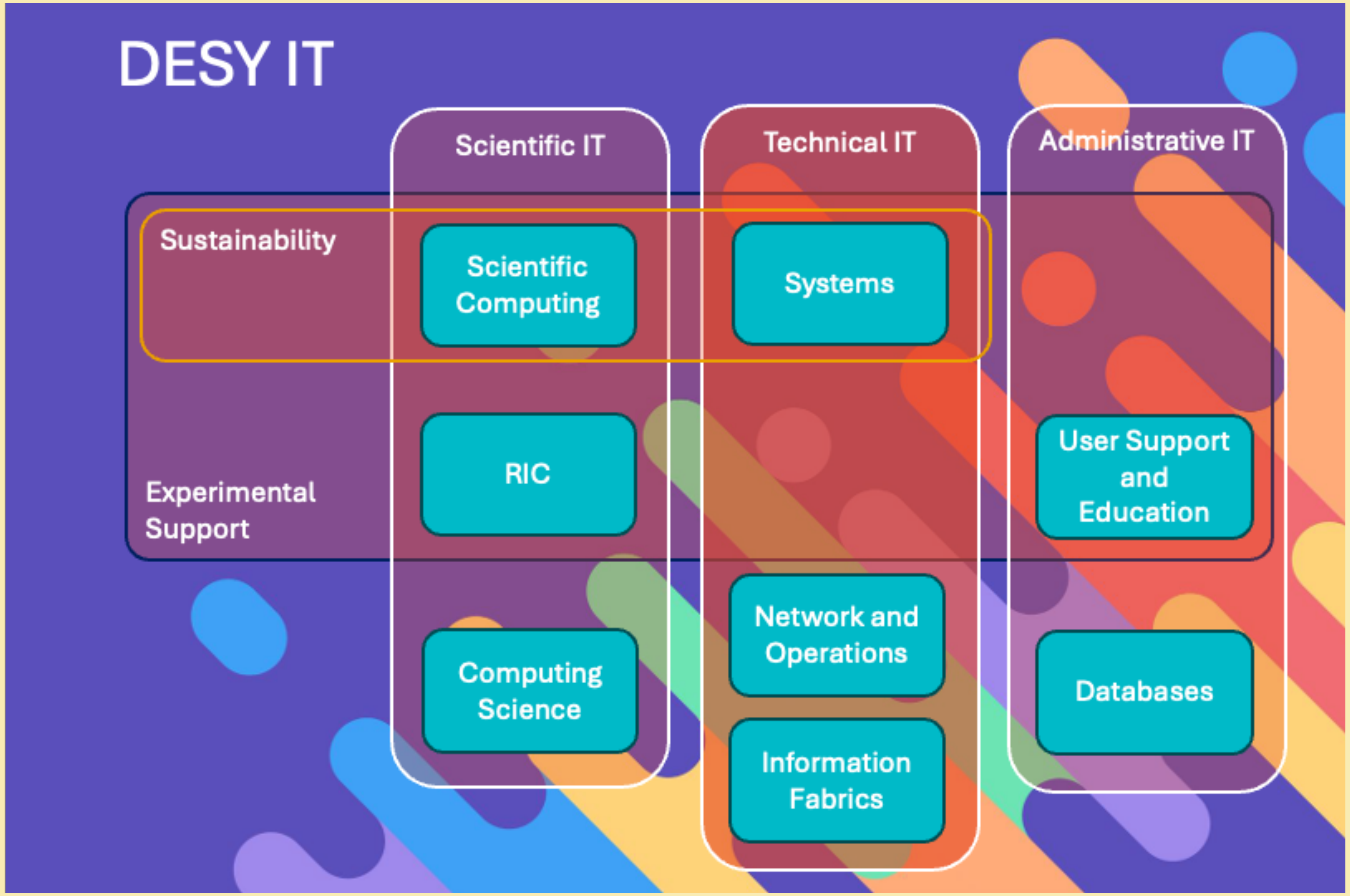
DESY Summer School - 25/07/2025

RESEARCH FACILITY
TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH

Contents

The Sustainable Computing Why

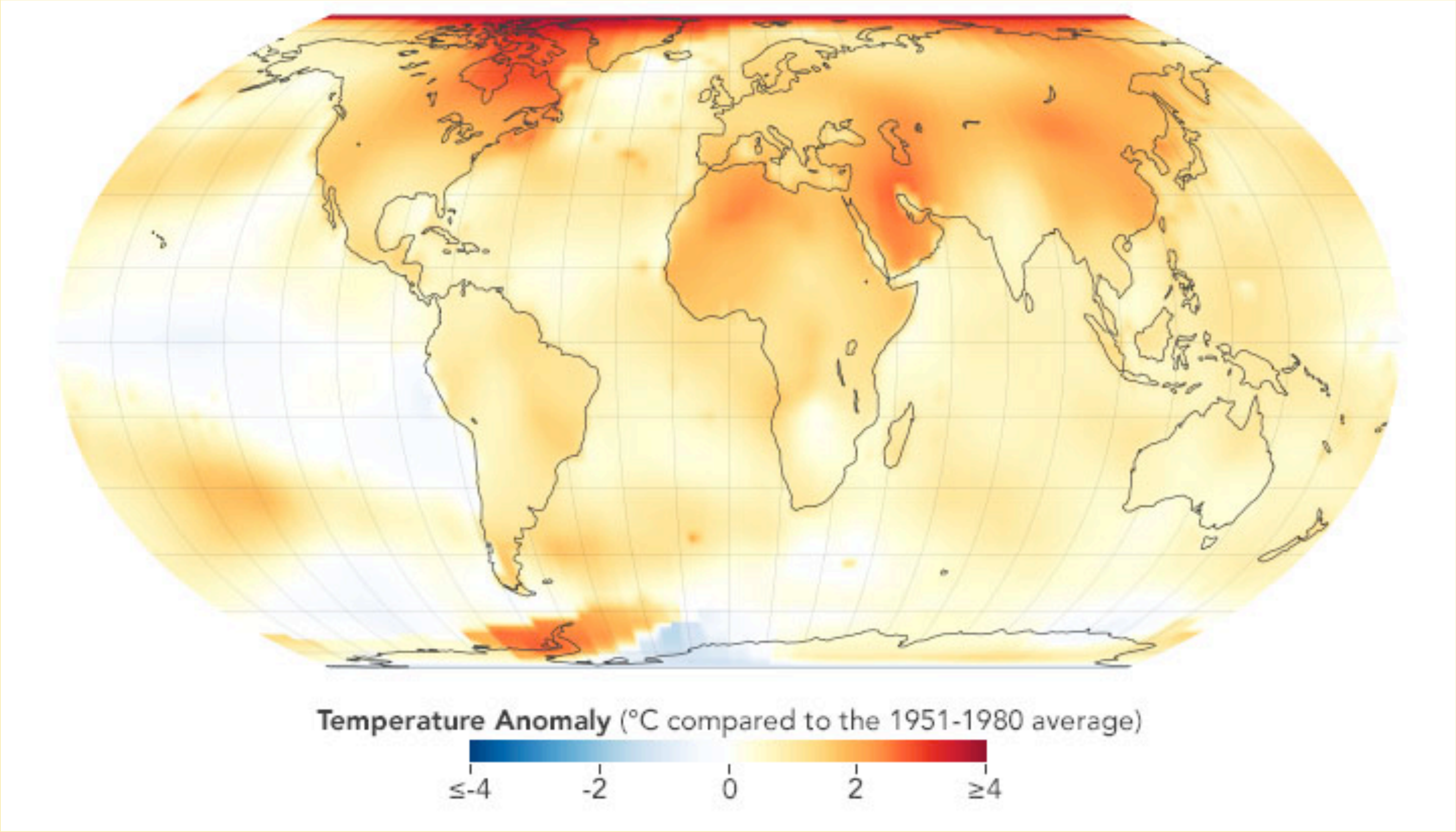
One slide summary



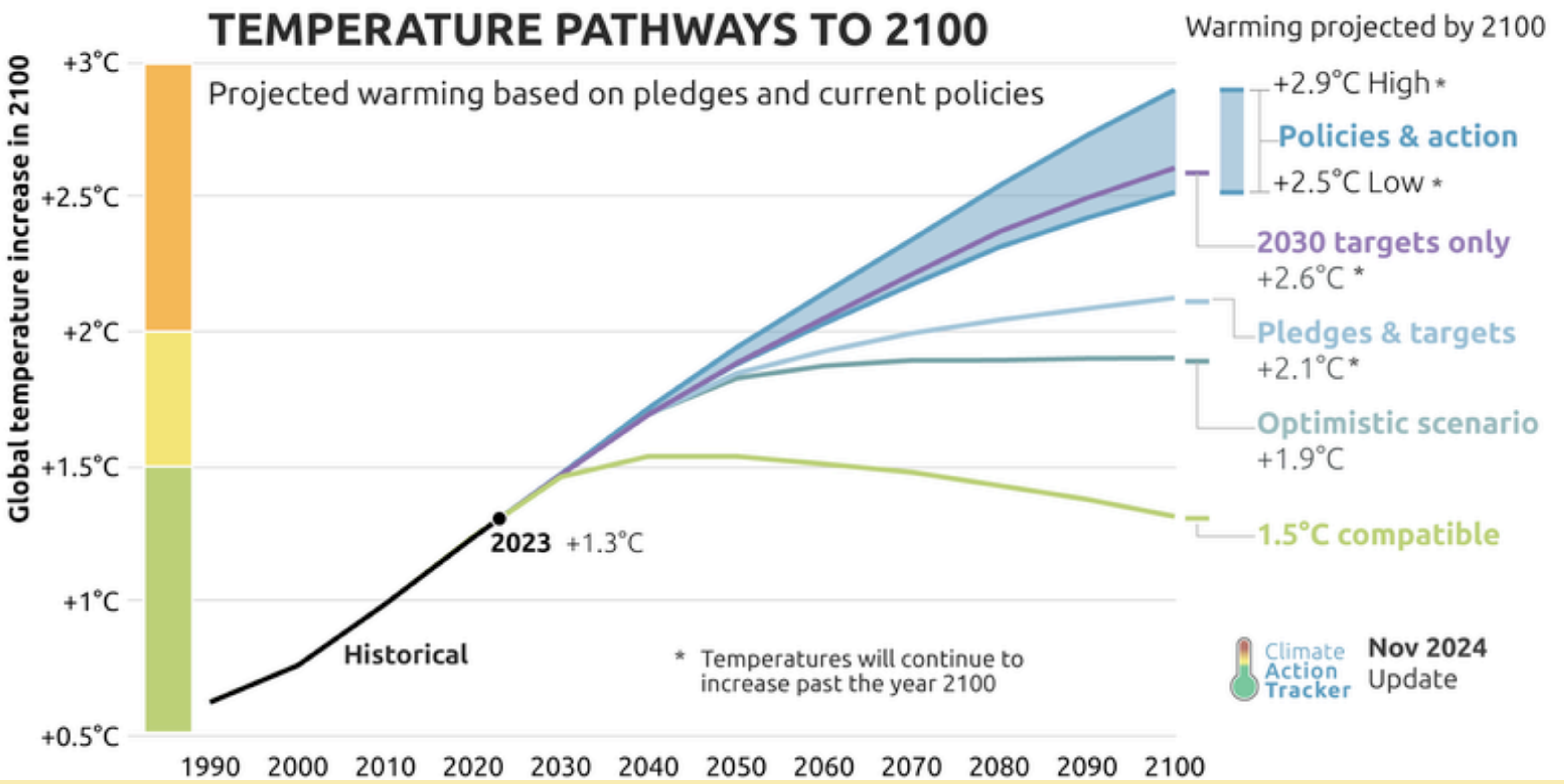
- DESY IT supports both on-site and off-site experiments with:
  live processing (ASAP:O, …),
  storage (dCache, CTA, …),
  and analysis (GRID, NAF, Maxwell)
  of data
- Also actively participate in research, and one of those is areas is IT sustainability.

Do I really need a slide about Climate Change?

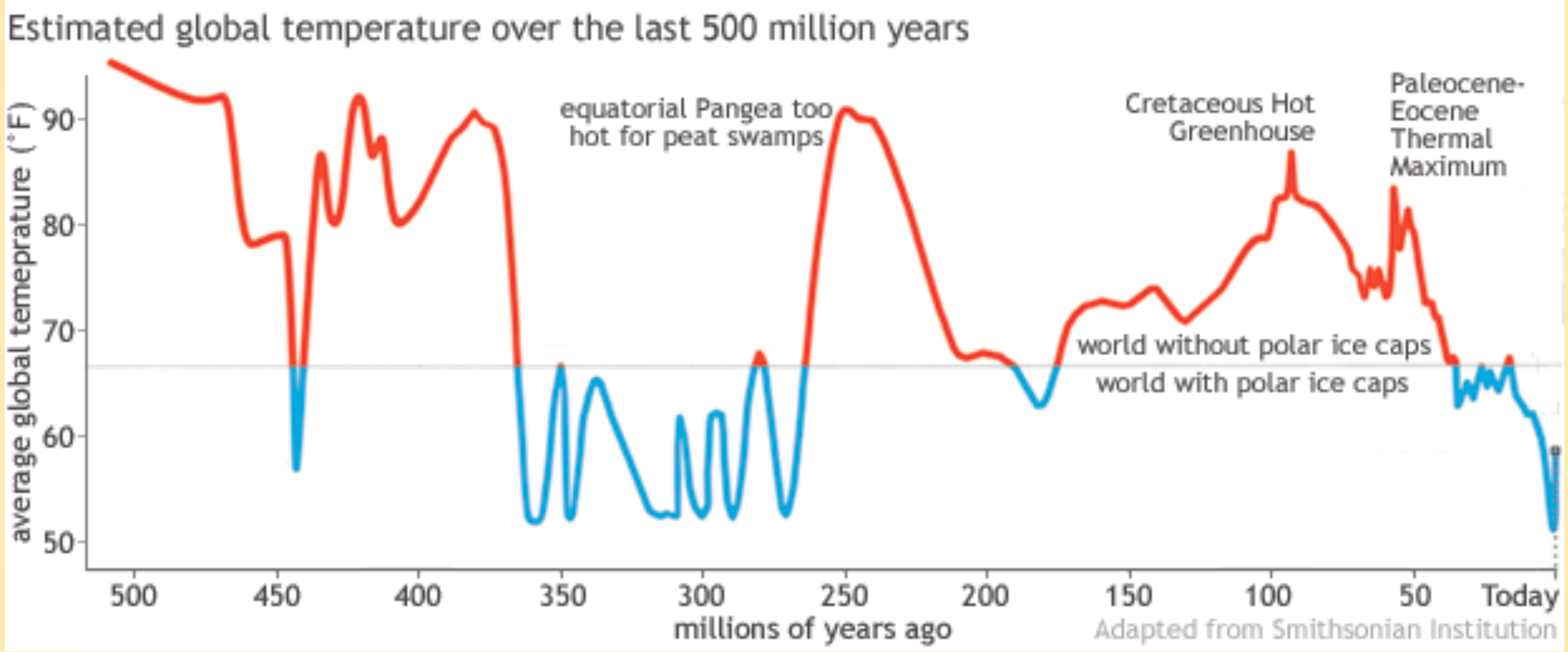
- Increased atmospheric emissions causes global temperatures to rise.
- This current rate-of-change of global temperature has only happened twice in Earth history
- Global averages hide extreme weather problems.
- We need to try and slow this temperature rise.



https://earthobservatory.nasa.gov/images/149321/2021-continued-earths-warming-trend



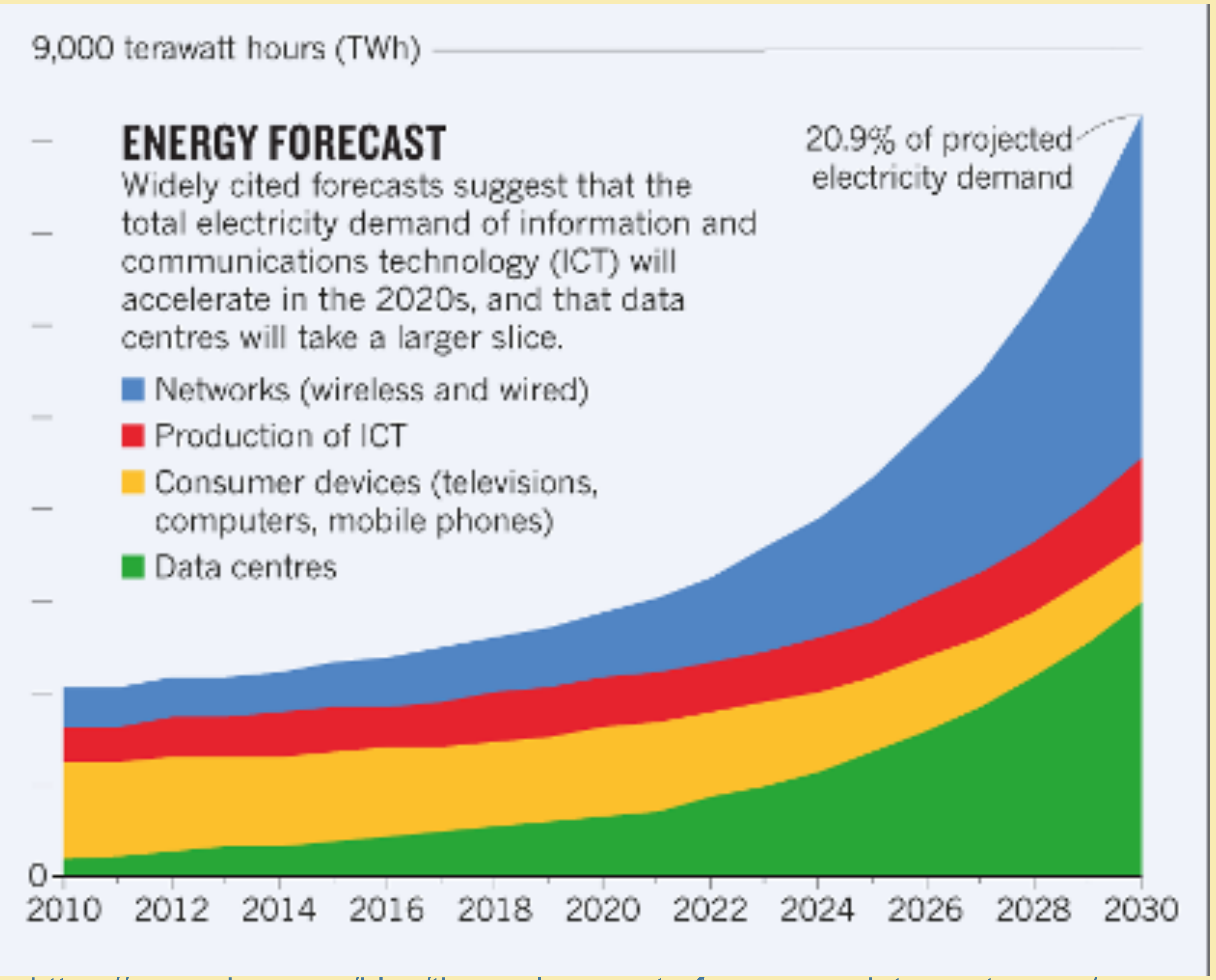
https://climateactiontracker.org/publications/the-climate-crisis-worsens-the-warming-outlook-stagnates/



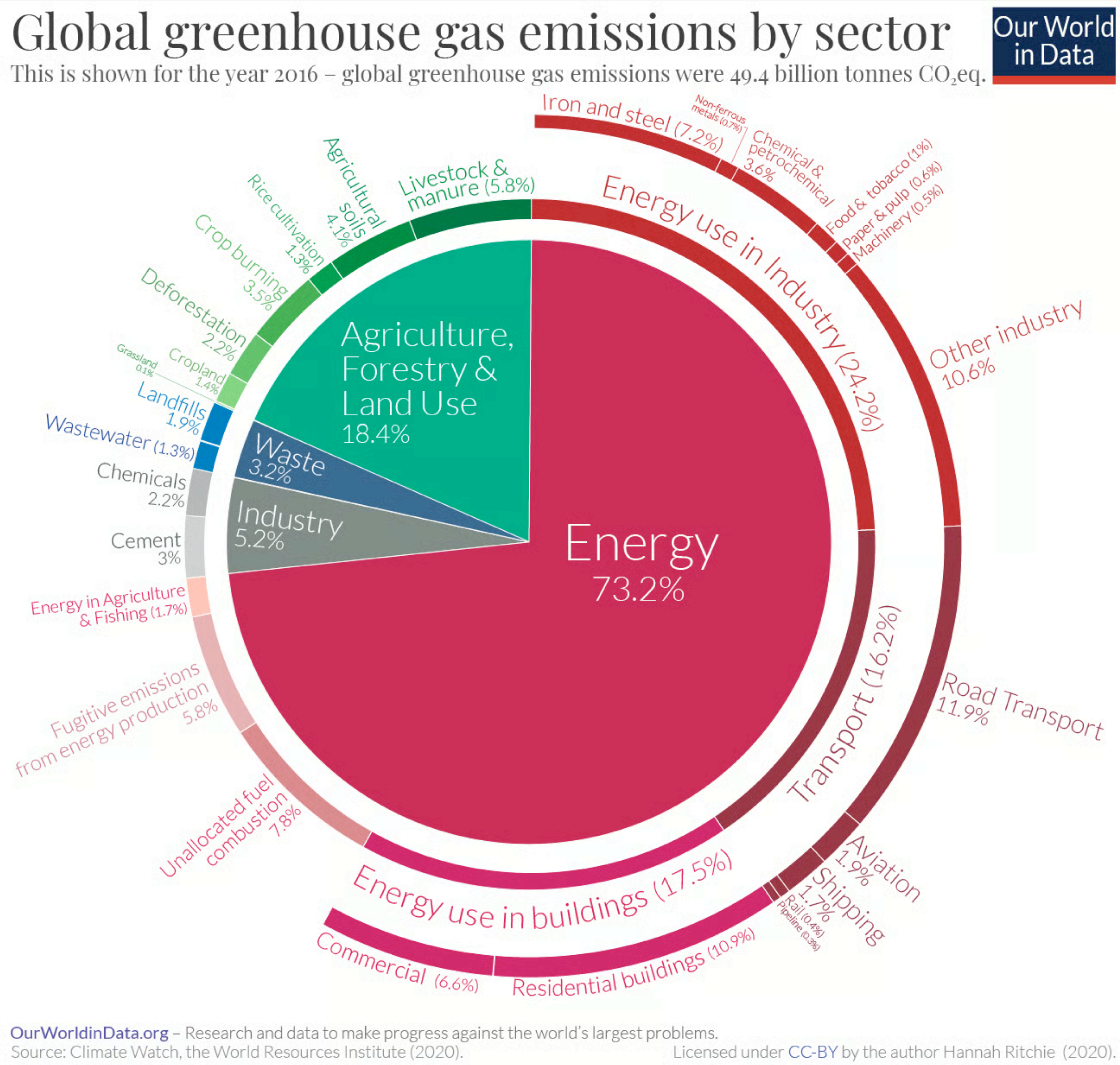
https://blog.practicalethics.ox.ac.uk/2023/09/climate-change-planetary-health-and-the-deep-significance-of-the-anthropocene/

Emissions and IT

- Data-centres, AI and increasingly digital world means that more emissions will be attributed to IT.

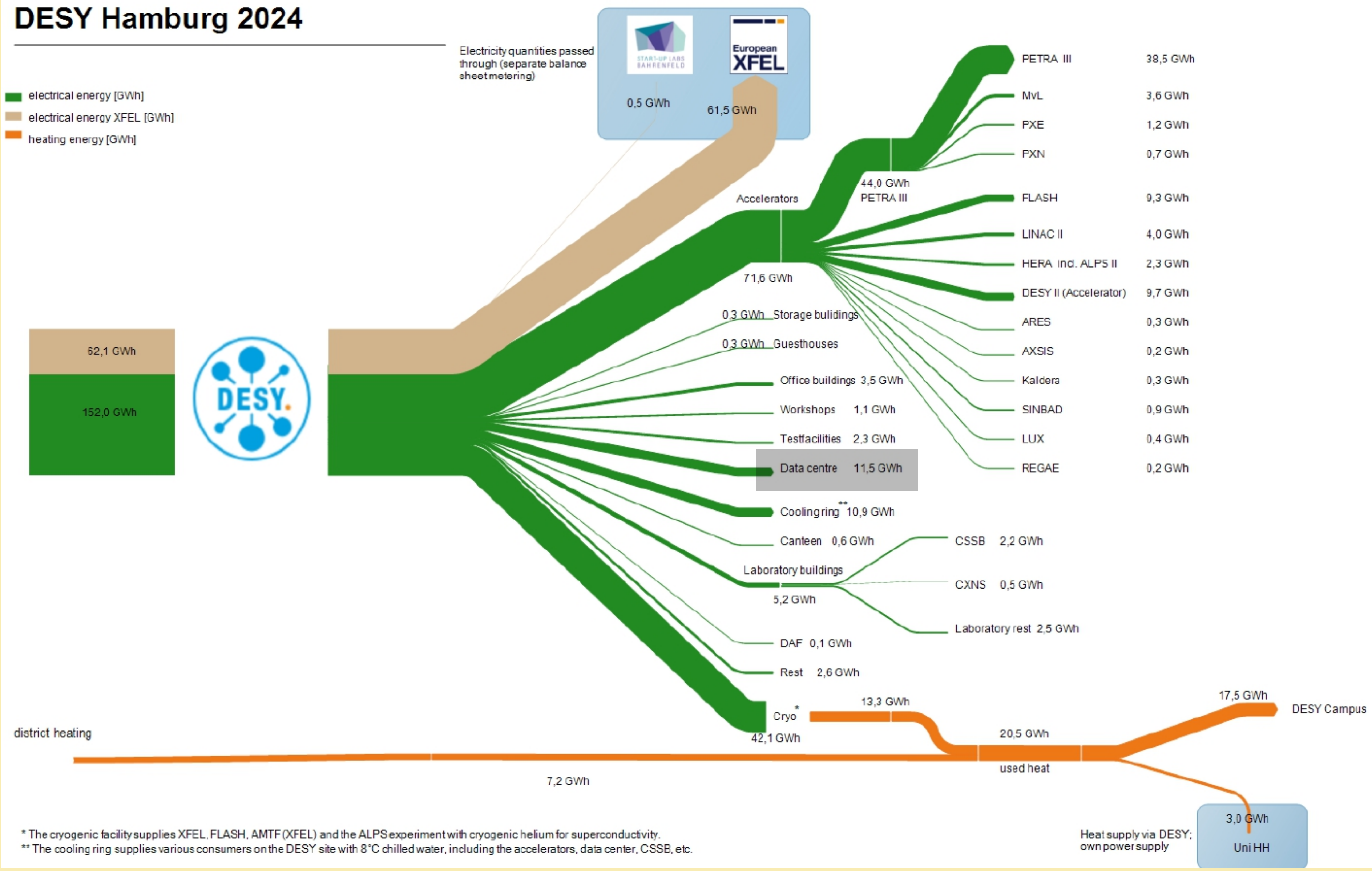


https://www.akcp.com/blog/the-real-amount-of-energy-a-data-center-use/



Context about climate crisis

And the role of IT



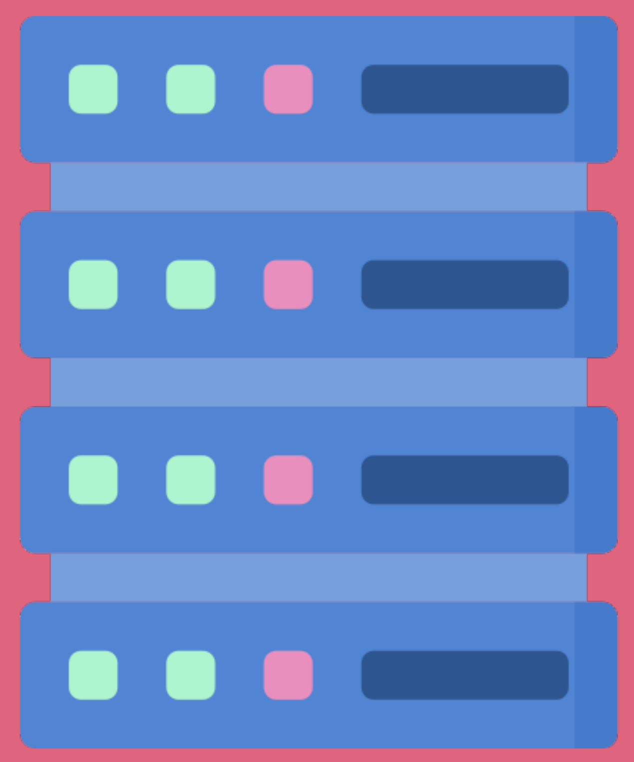
https://www.akcp.com/blog/the-real-amount-of-energy-a-data-center-use/

- At DESY the data centre consumes 11.5GWh of power a year, and that's around 8% of the total power draw at DESY.
- Future data-centres will be even more power-hungry and are likely to form part of more energy-intensive ecosystems
- We should be leading by example - research-led solutions

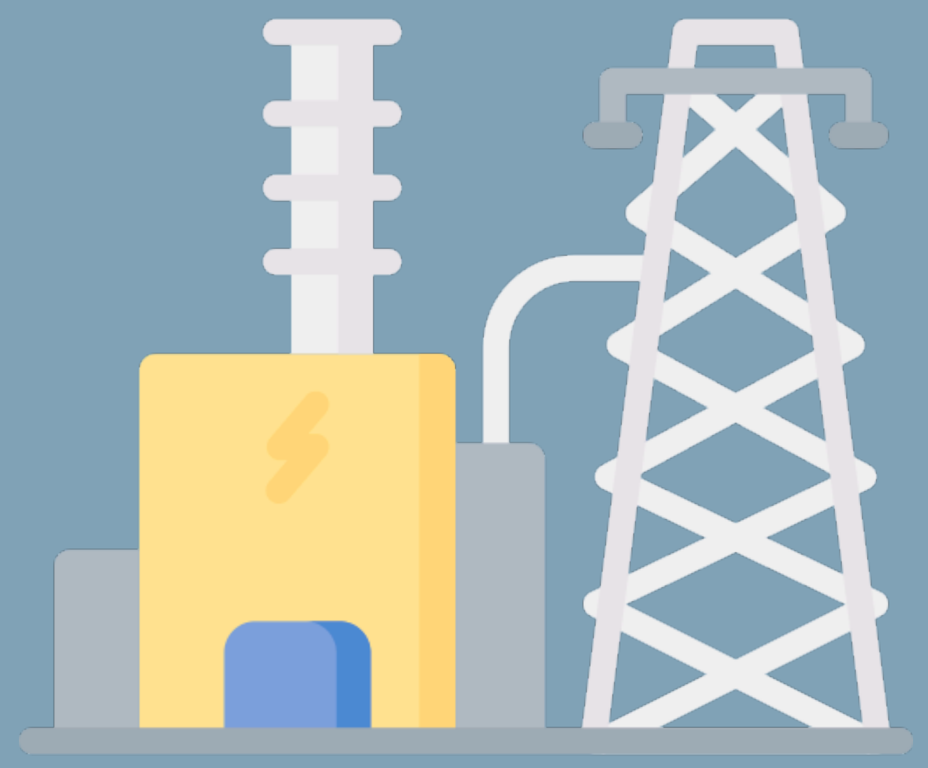
What is IT Sustainability

- Often we talk about it in the context of the resources we use, but I see it as three parts.

Materials/Construction
(+ EOL - Raw material impact)



Operation
(Upkeep - Running Costs)



Personpower
(Growth, Knowledge Transfer)

- If something is to be truly sustainable, you need to understand the part all of these play in your ecosystem.

- What does this mean with respect to the IT at DESY.

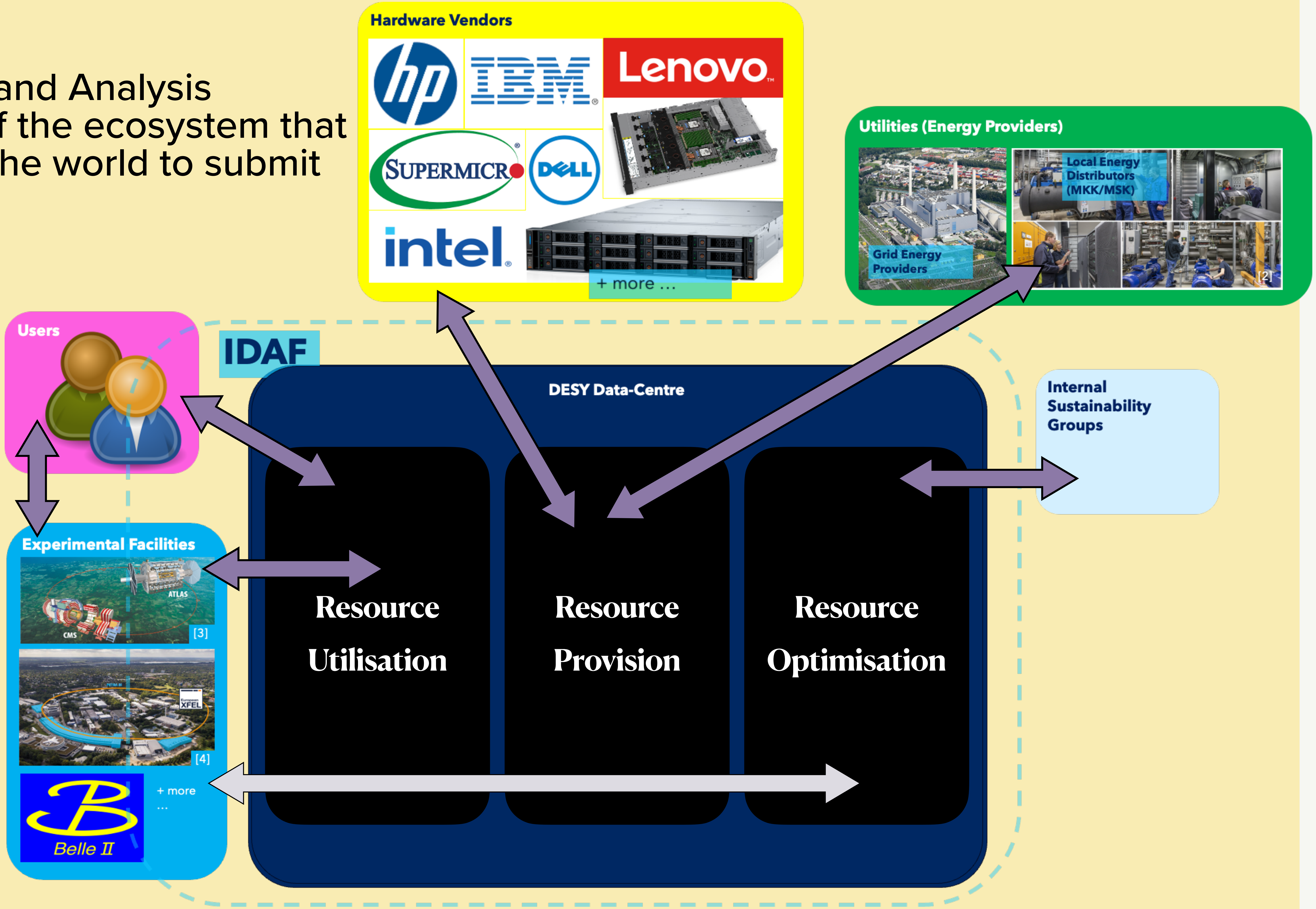
Images from: https://www.flaticon.com

The IDAF at DESY

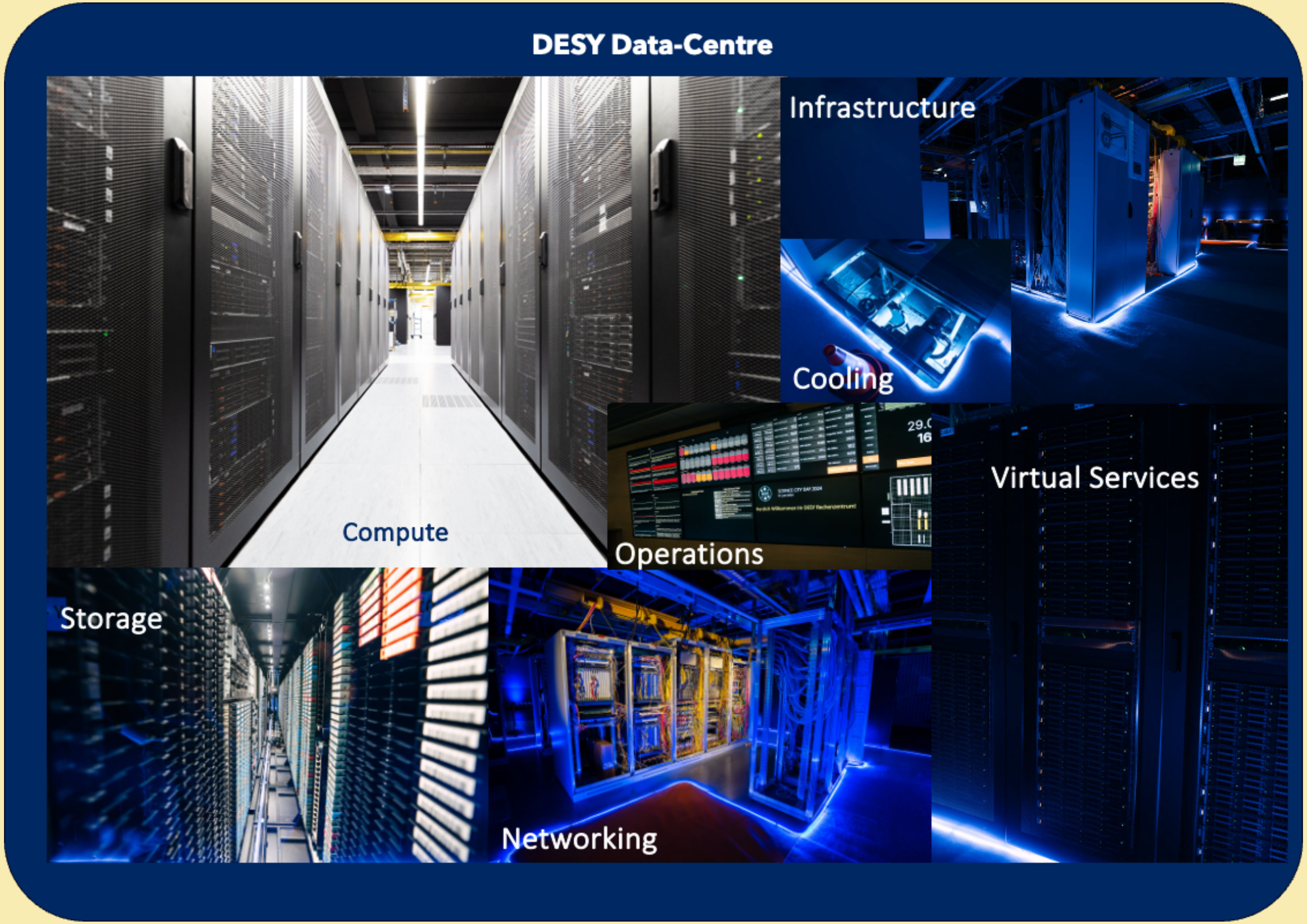
- The Interdisciplinary Data and Analysis Facility (IDAF) forms part of the ecosystem that allows users from around the world to submit scientific work to DESY

- While the data-centre is at it's heart, sustainability efforts will be limited if the parts of this wider ecosystem don't talk to each other

- Cooling is operated centrally, and is managed by a different department



DESY Data Centre



- Made up of three clusters

  NAF - Local Particle Physics Users

  GRID - Remote Particle Physics* Users

  MAXWELL - Local Photon Science users

- Each cluster provides unique services in terms of compute and storage to a different types of users.

- The resources therefore are used in different ways

- **Research required to making this ecosystem sustainable**

RF2.0 RESEARCH FACILITY TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH

The Sustainable Computing How

RF2.0 RESEARCH FACILITY TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH at DESY.

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

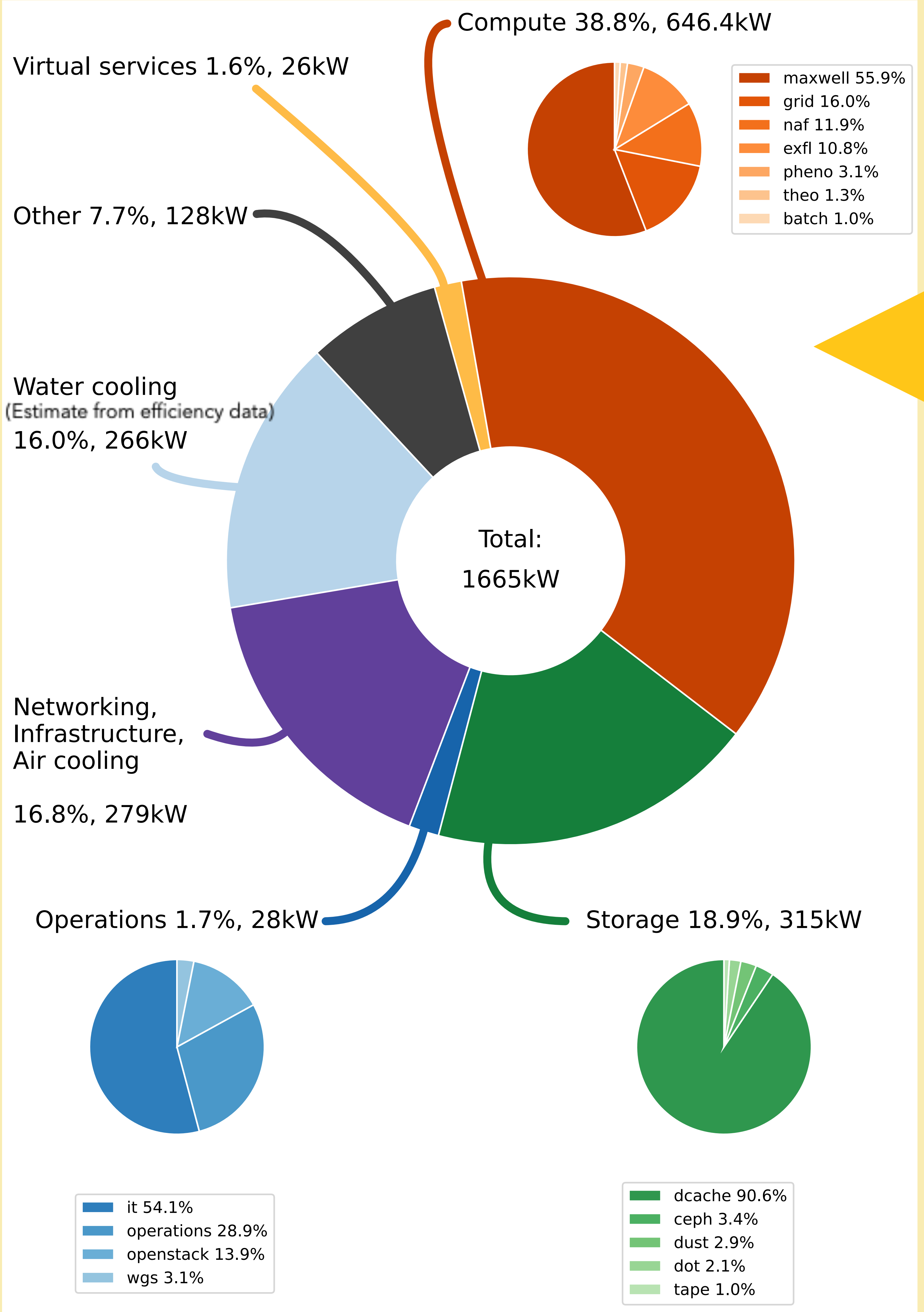
- At DESY we focus on what this means for current and future operation (and construction) of green data-centres.

DESY DATA CENTRE

RESEARCH FACILITY
TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH
at DESY.

* 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.



Median Power Consumption of the DESY Data-Centre

1. **Learn**

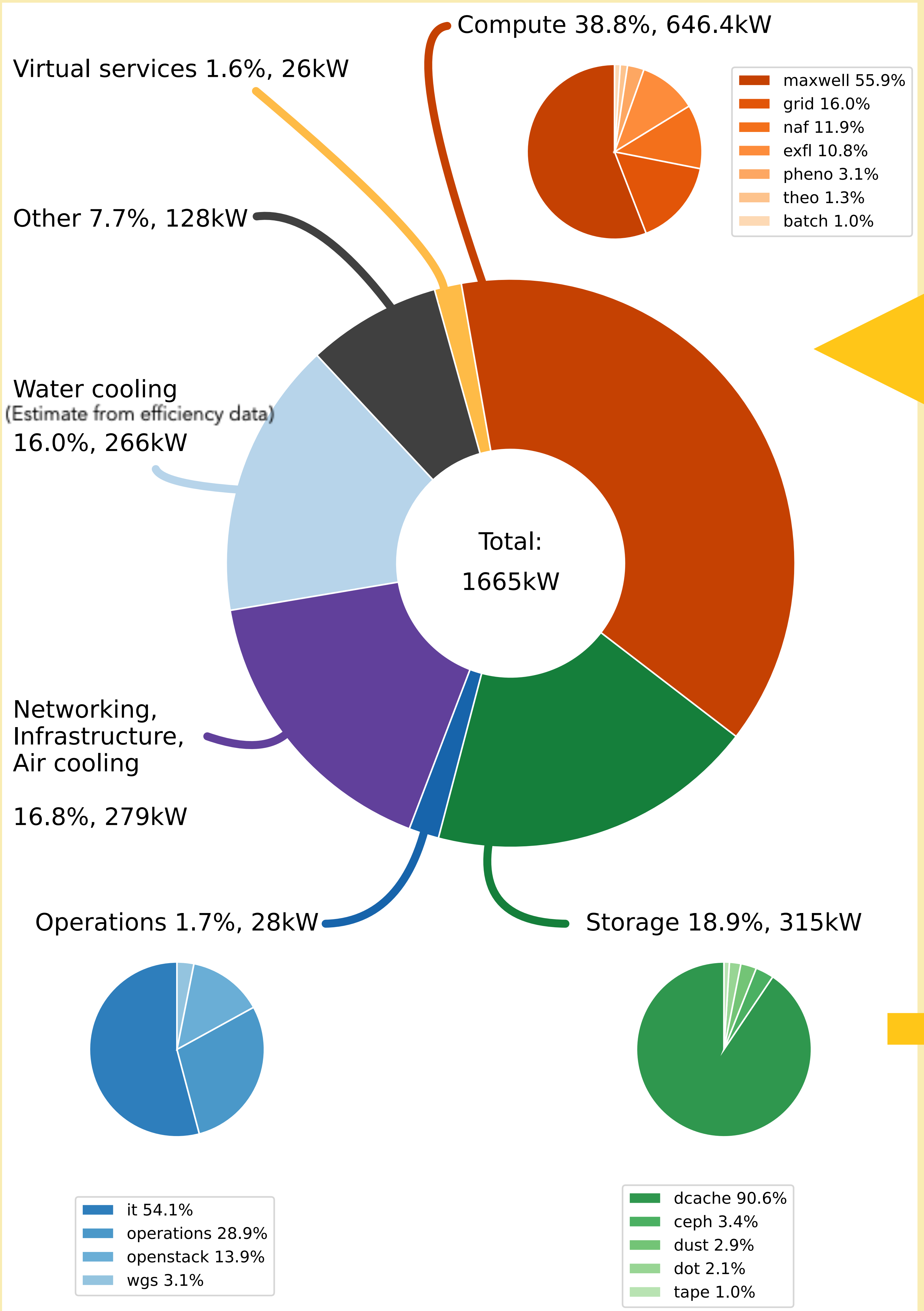
Take detailed and fine-grained measurements of data-centre components

DESY DATA CENTRE

RESEARCH FACILITY 2.0 at DESY

TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH

- 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.



Median Power Consumption of the DESY Data-Centre

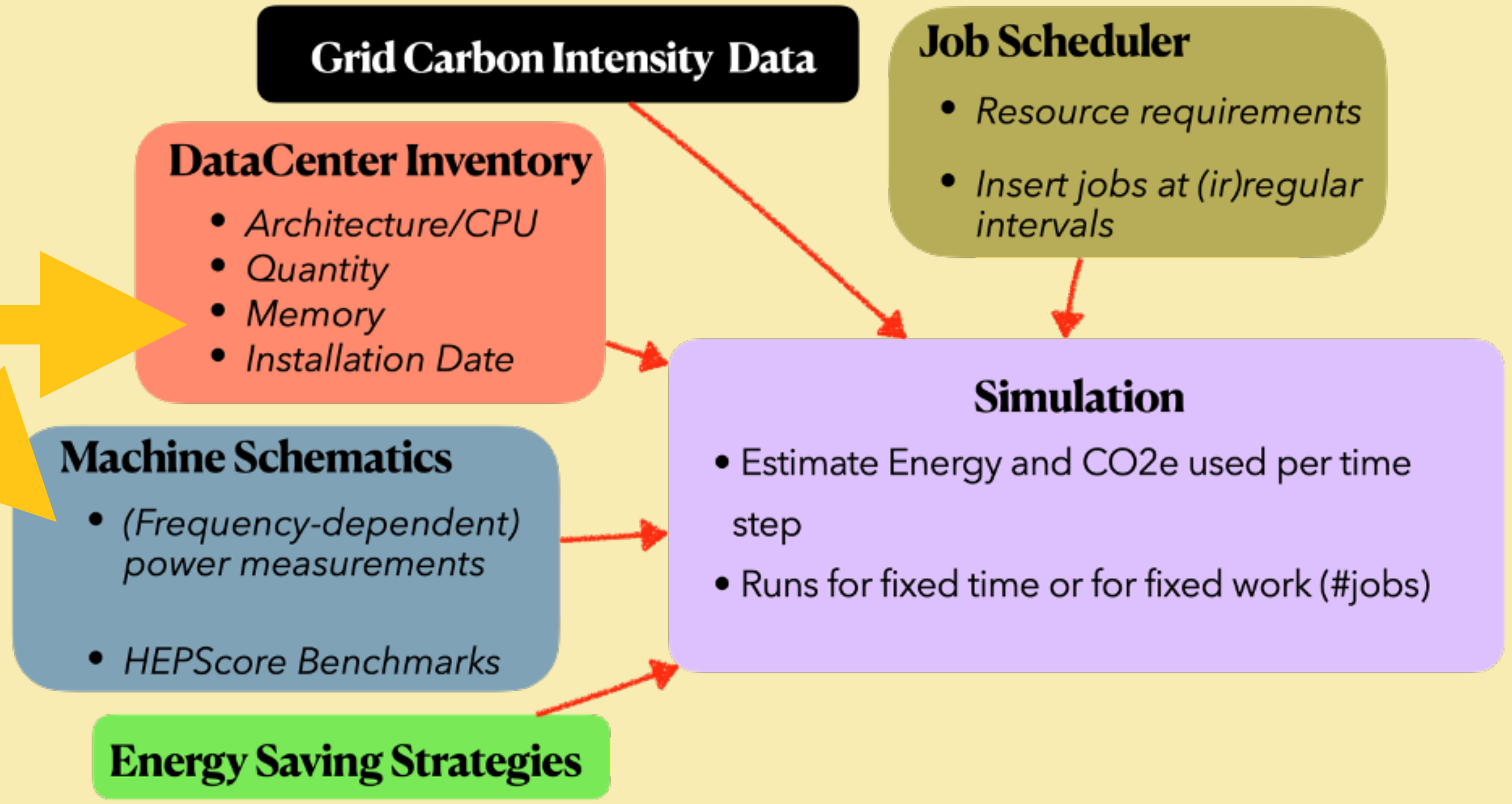
1. Learn

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

DESY DATA CENTRE

2. Simulate

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

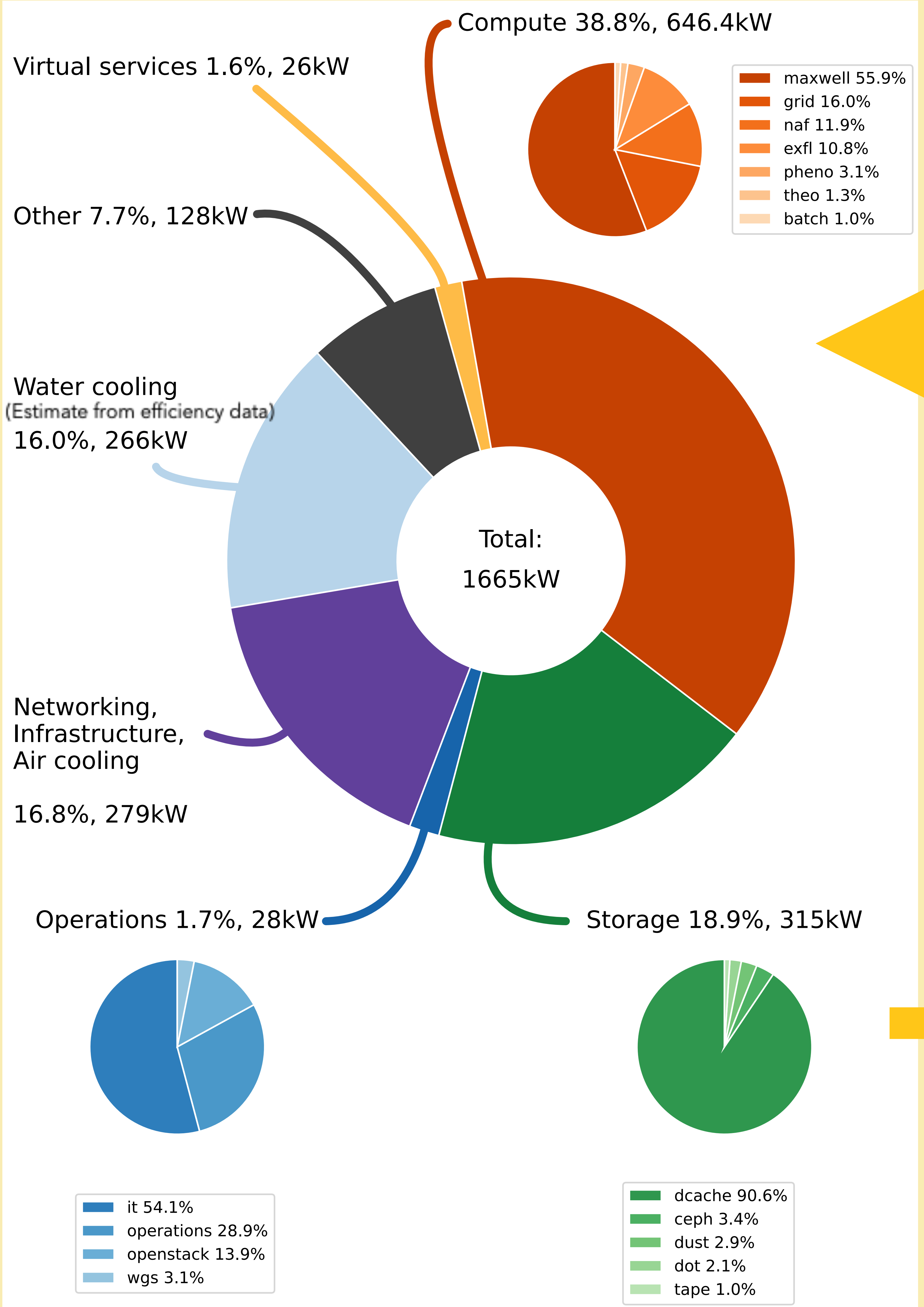


RESEARCH FACILITY
TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH

RF 2.0 at DESY

* 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.



Median Power Consumption of the DESY Data-Centre

DESY DATA CENTRE

1. Learn

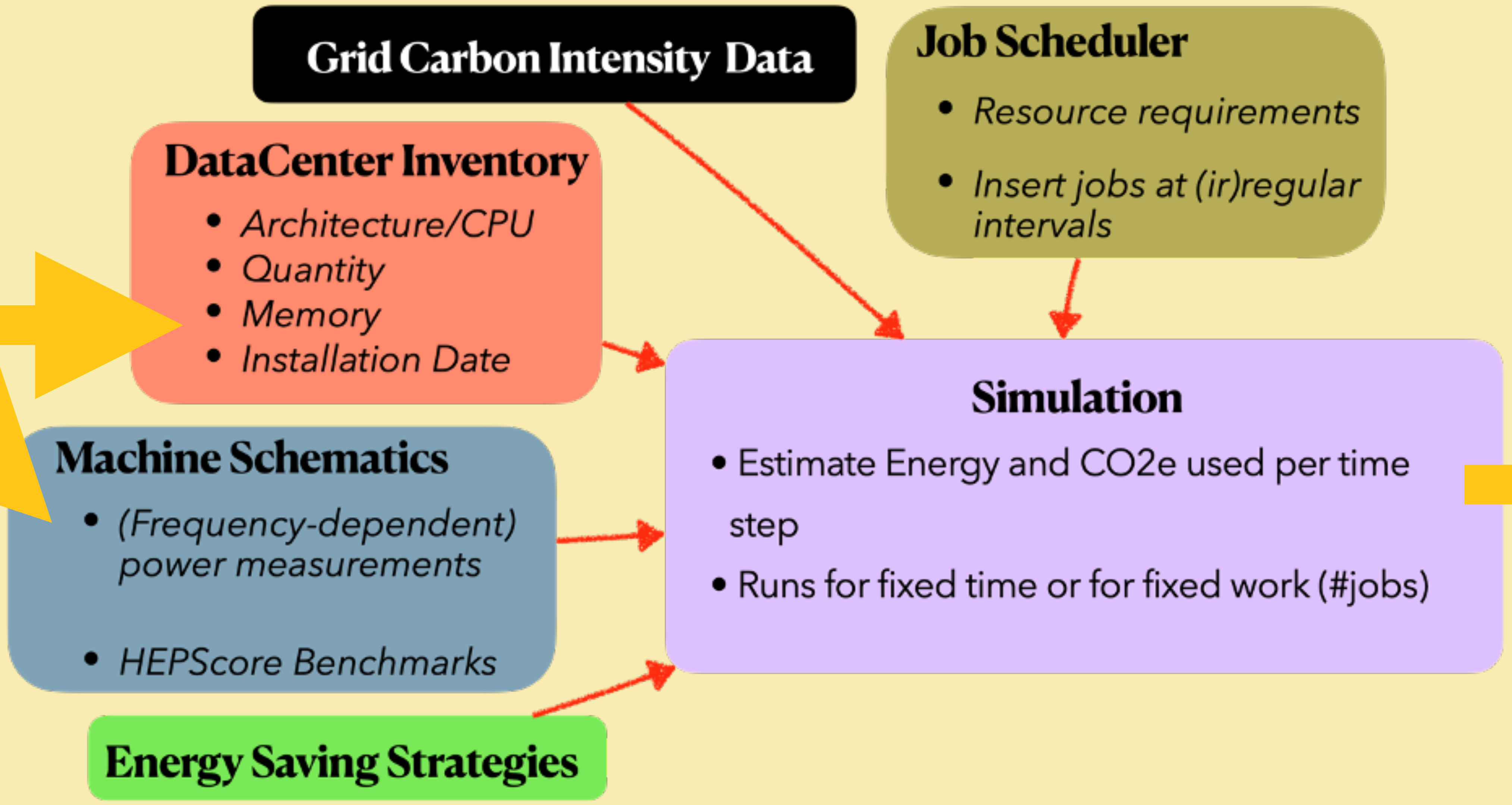
Take detailed and fine-grained measurements of data-centre components

3. Refine

Take simulation findings and dynamically implement them at the datacenter

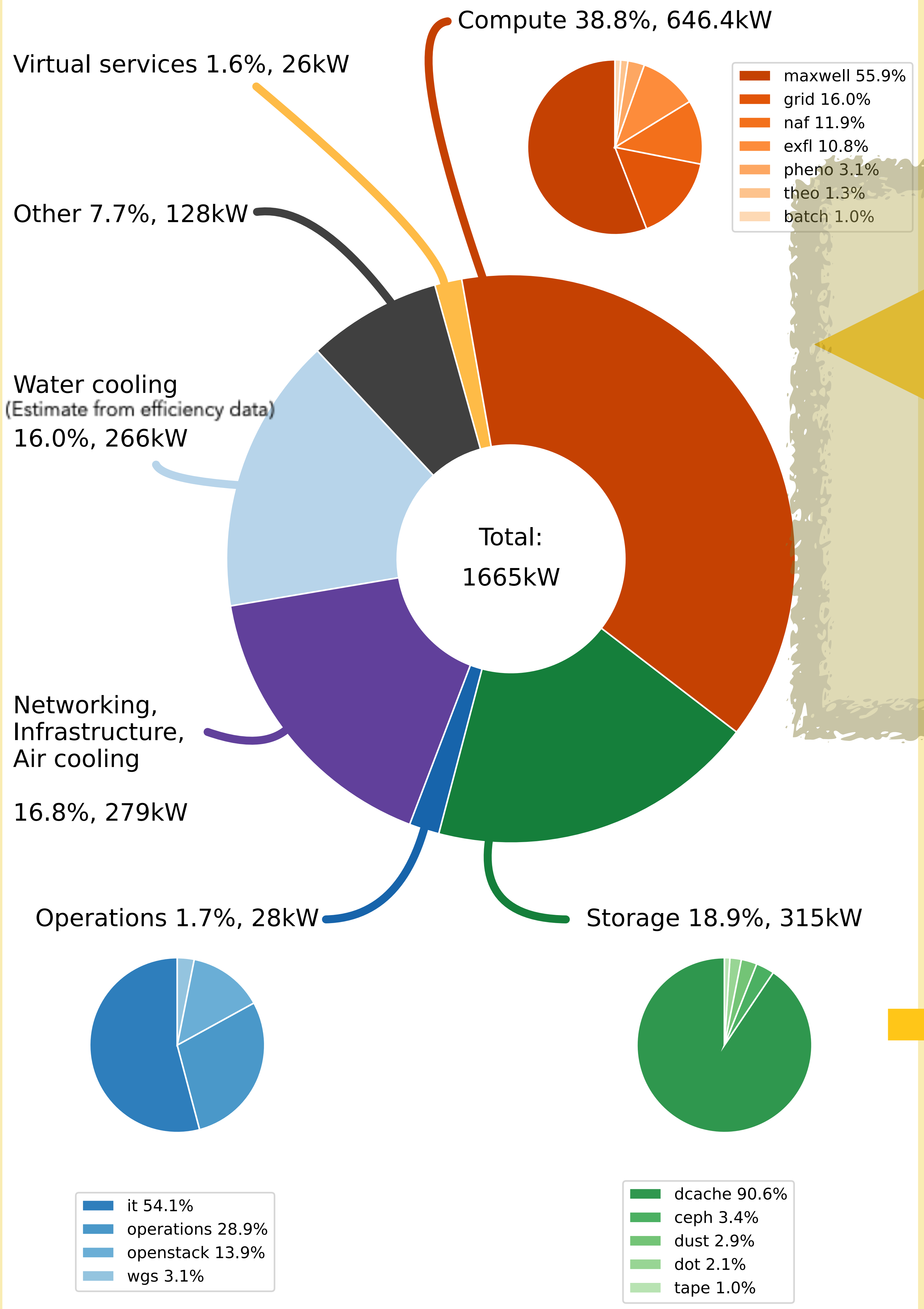
2. Simulate

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



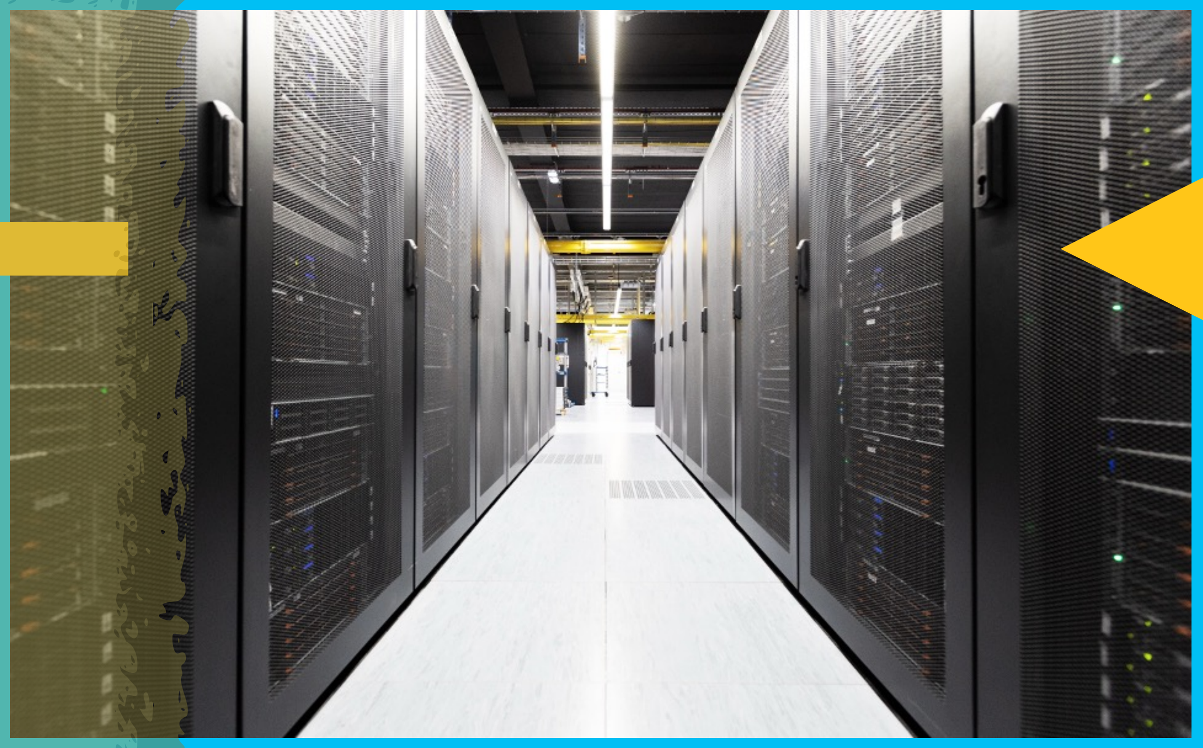
RF2.0 RESEARCH FACILITY
TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH
at DESY.

- 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.



Median Power Consumption of the DESY Data-Centre

DESY DATA CENTRE



1. Learn

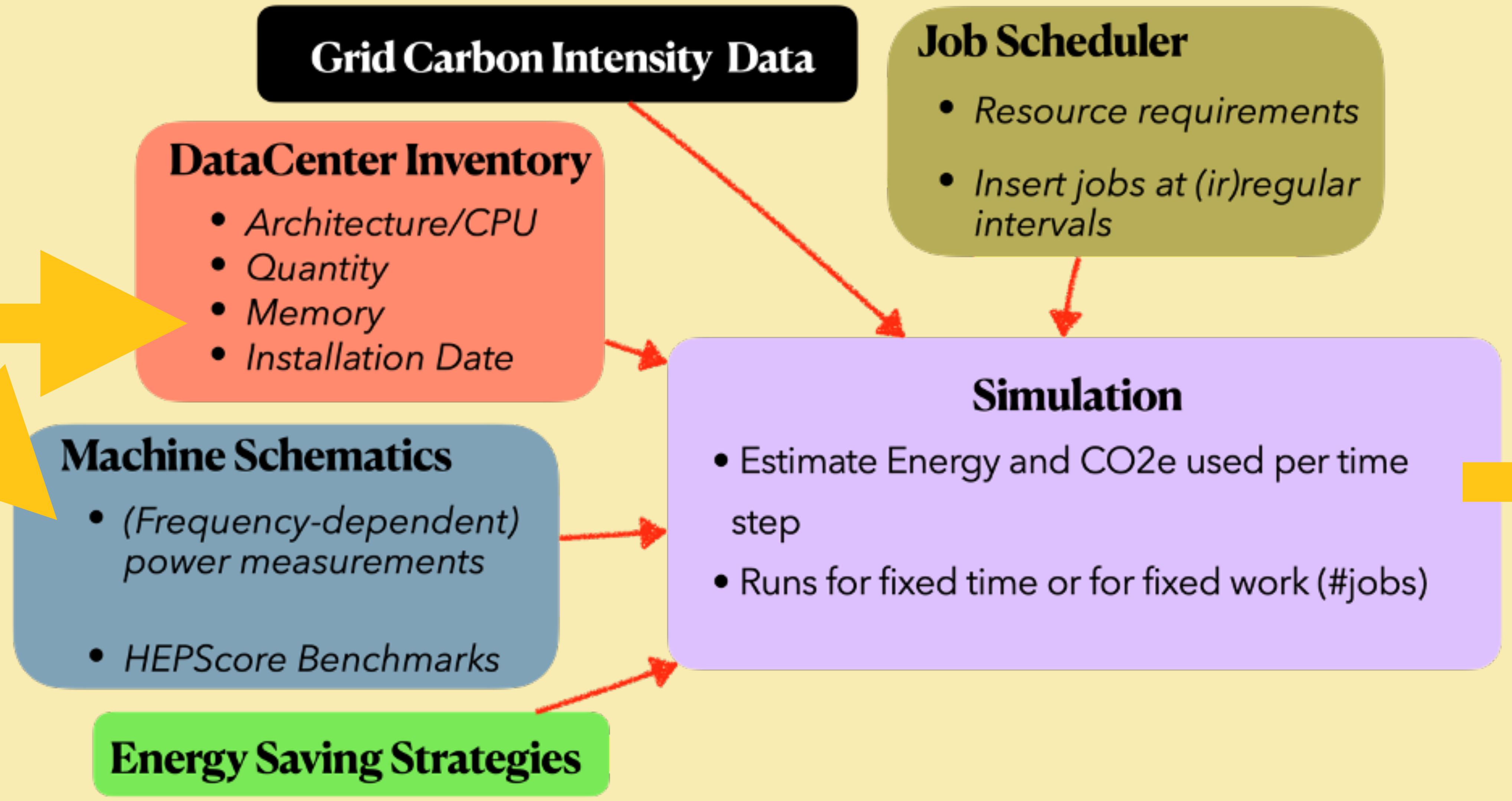
Take detailed and fine-grained measurements of data-centre components

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

Data Gathering and Analysis

Static Measurements

Machine Identifiers

Manufacturer

CPU type and architecture

Installation Date

Work Capability

Number of Cores

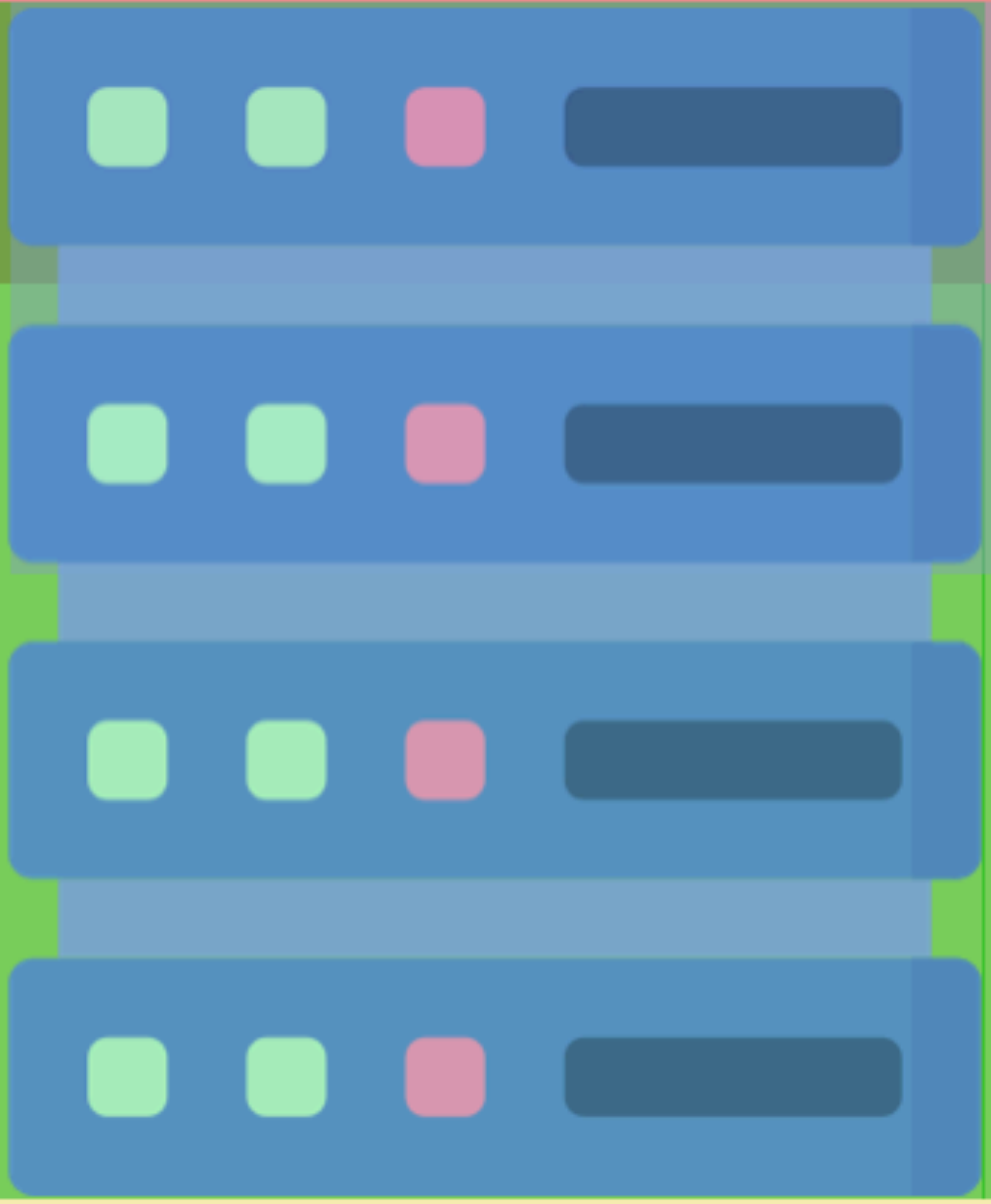
Memory Available

Operational Frequencies

(Frequency-dependent)

HEPScore values

Idle Power Draw



Rack Power Used

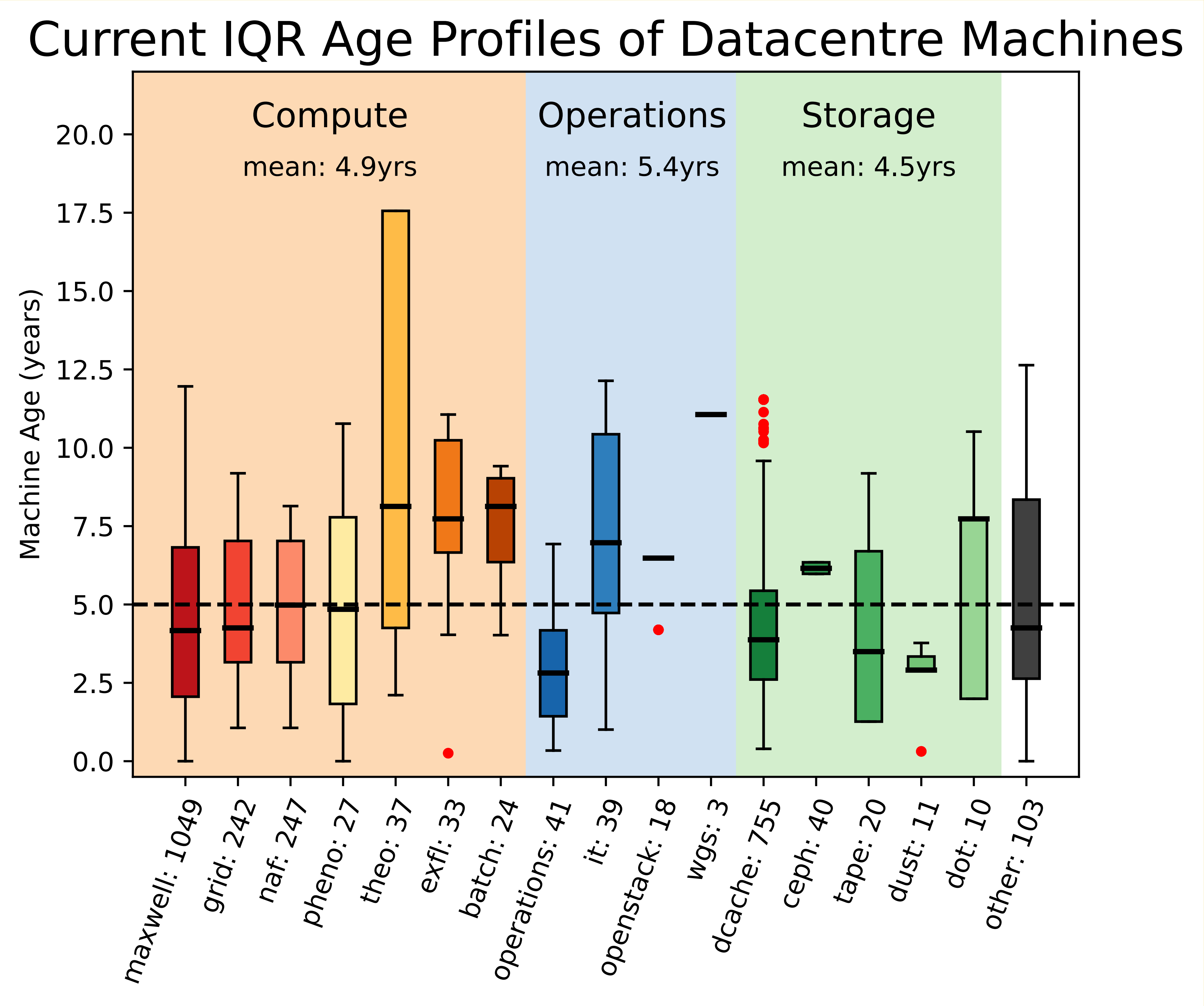
Typical Power Draw

Power Measurements

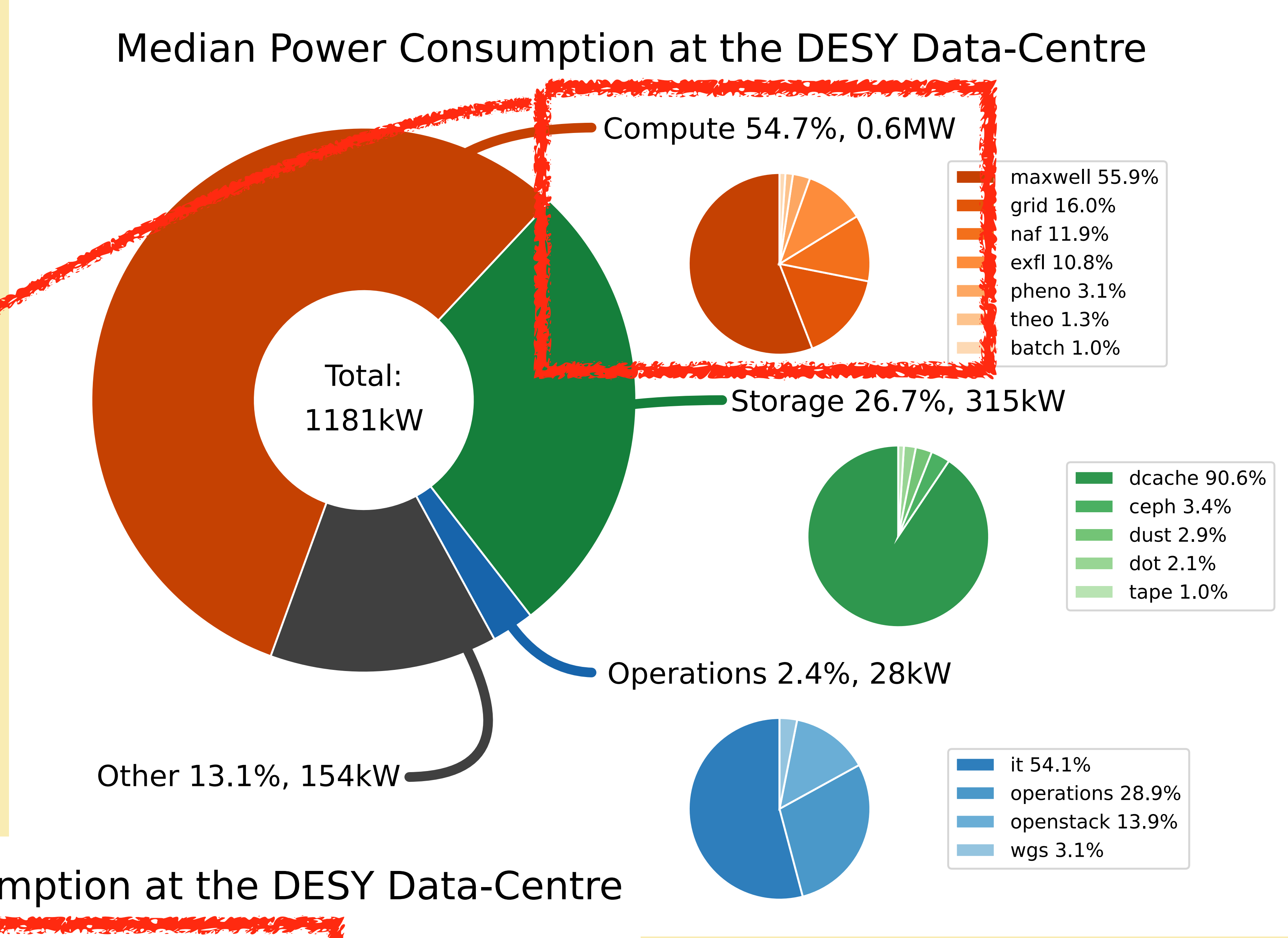
Performance Measurements

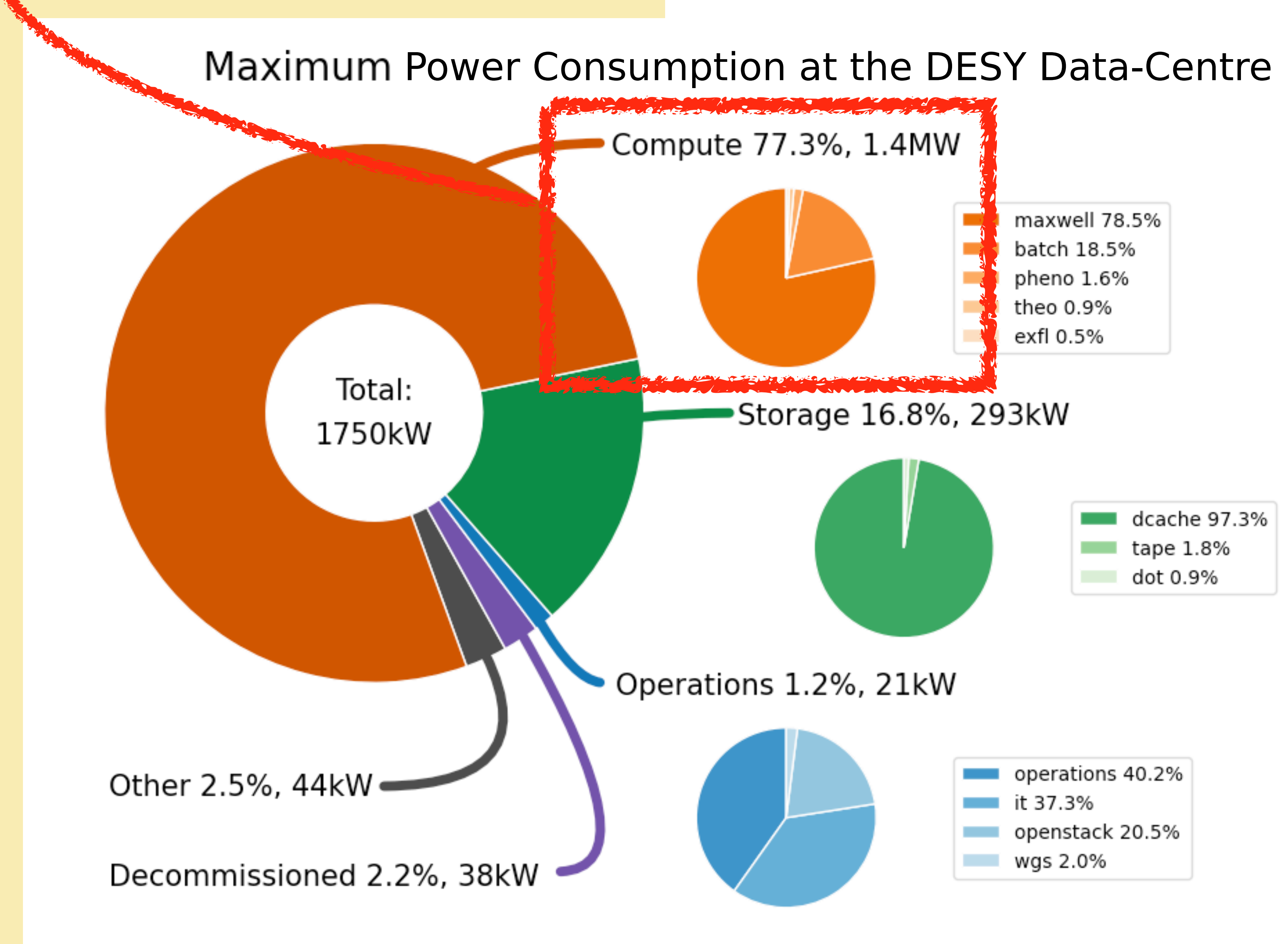
- Relative performance of machines -> Benchmarking (HEPScore)
- Benchmark created and maintained by a CERN-led team designed to calculate how well compute servers can handle HEP Workloads.
- Bigger number better

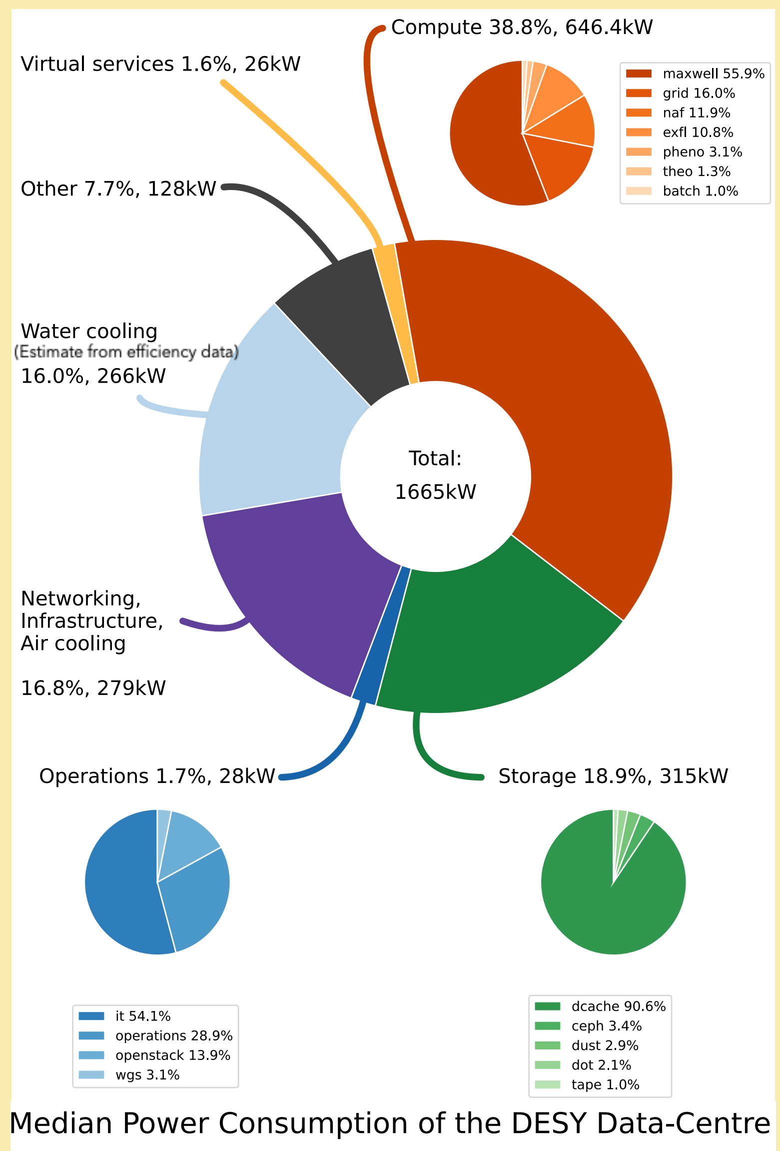
Data-Centre Data Summary Analysis



- Largest Fluctuations in power is in Compute



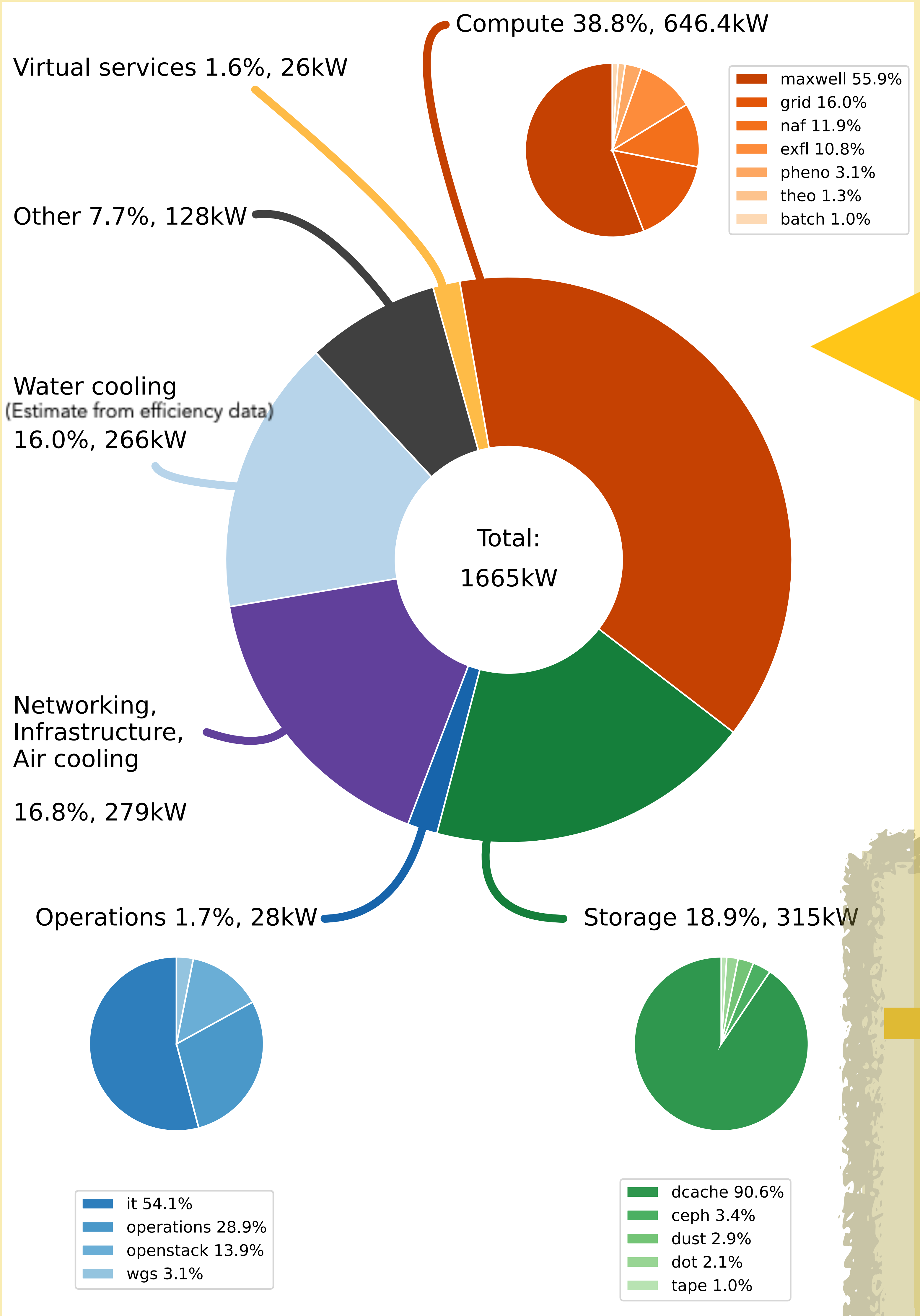




- Learning the types of lifetimes different machines have, then learn why?

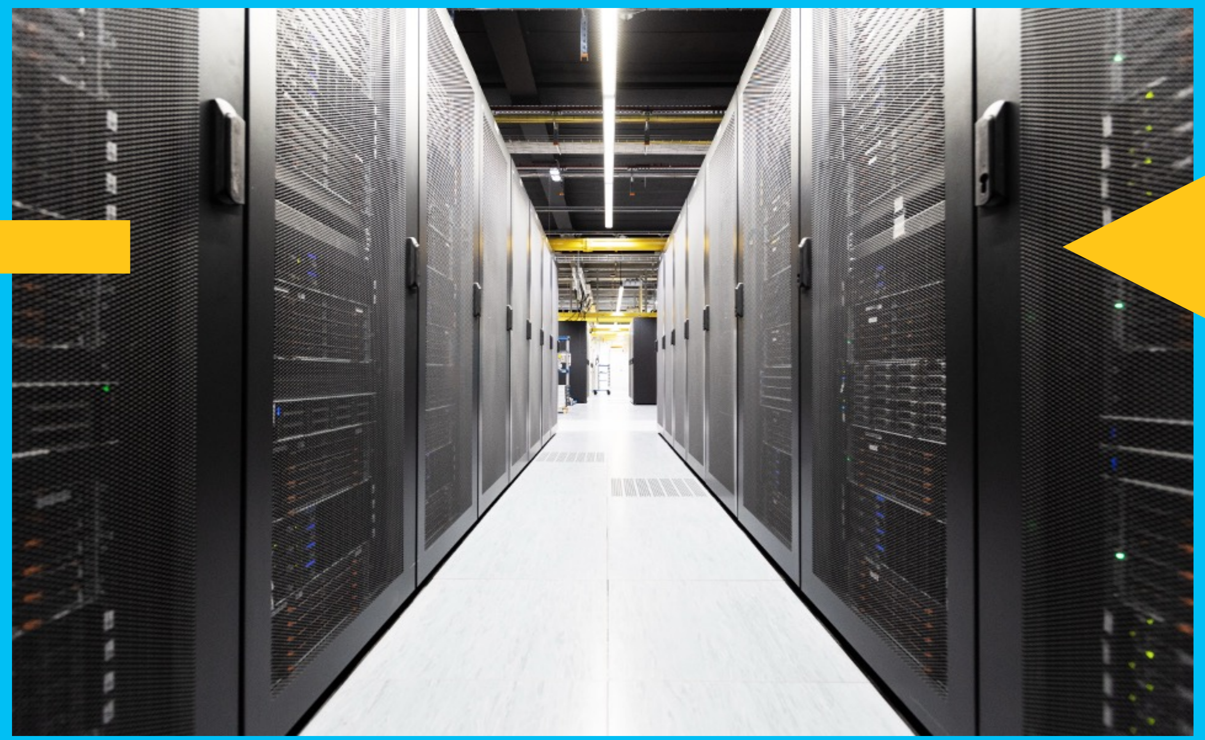
RF 2.0 RESEARCH FACILITY TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH at DESY.

- 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.



Median Power Consumption of the DESY Data-Centre

DESY DATA CENTRE



1. Learn

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

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

Job Scheduler
- *Resource requirements*
- *Insert jobs at (ir)regular intervals*

DataCenter Inventory
- *Architecture/CPU*
- *Quantity*
- *Memory*
- *Installation Date*

Machine Schematics
- *(Frequency-dependent) power measurements*
- *HEPScore Benchmarks*

Energy Saving Strategies

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

2. Simulate

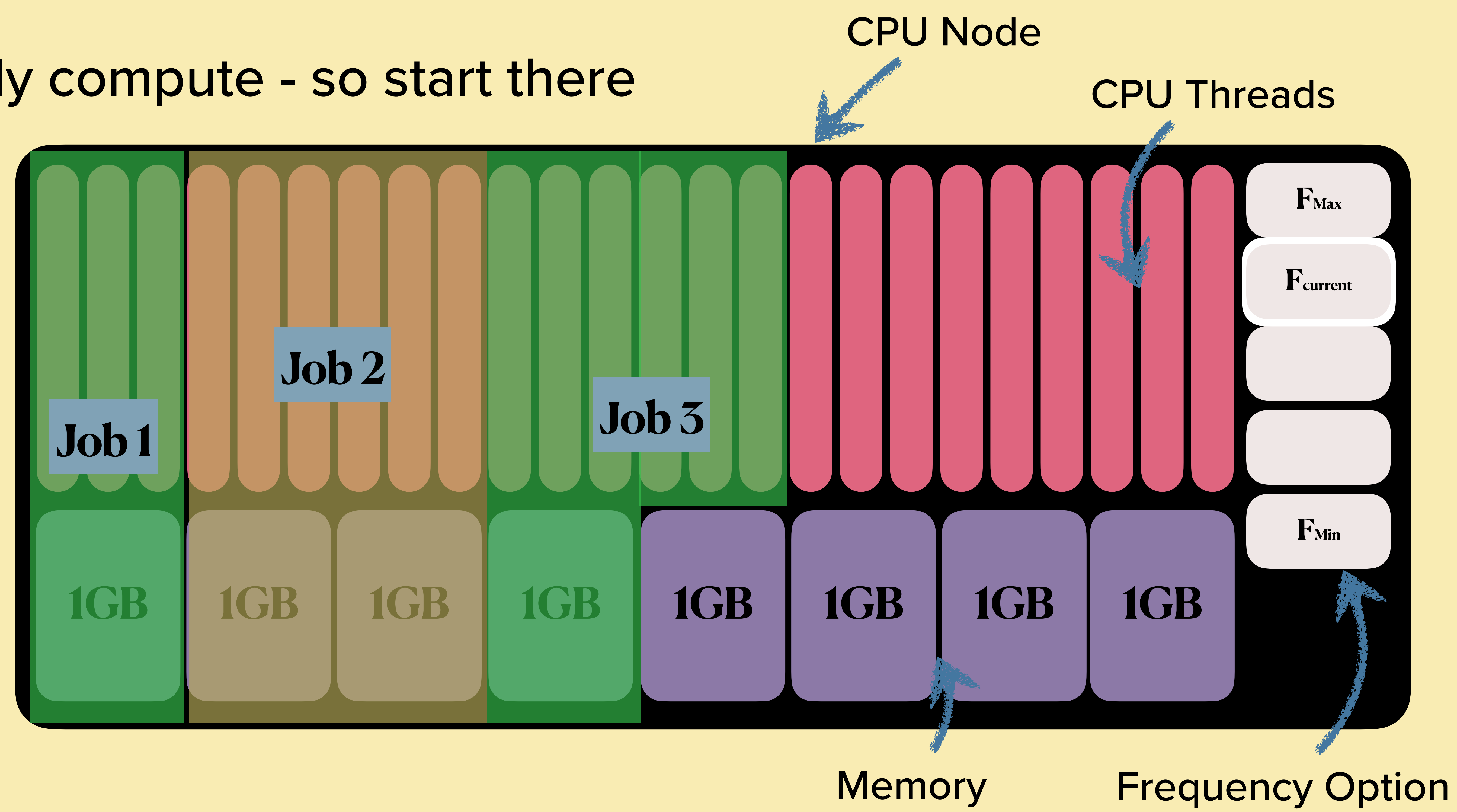
Estimate and Trial

- The data-centre is in continuous operation. Trialling out things that can be advantageous should be done offline if possible
    - (Don’t develop on the main branch)

- A simulation
    - should test results of policies faster than implementing them at a datacenter
    - can run a variety of energy scenarios in parallel
    - doesn’t disturb the everyday running of a data-centre

How do you simulate a datacenter?

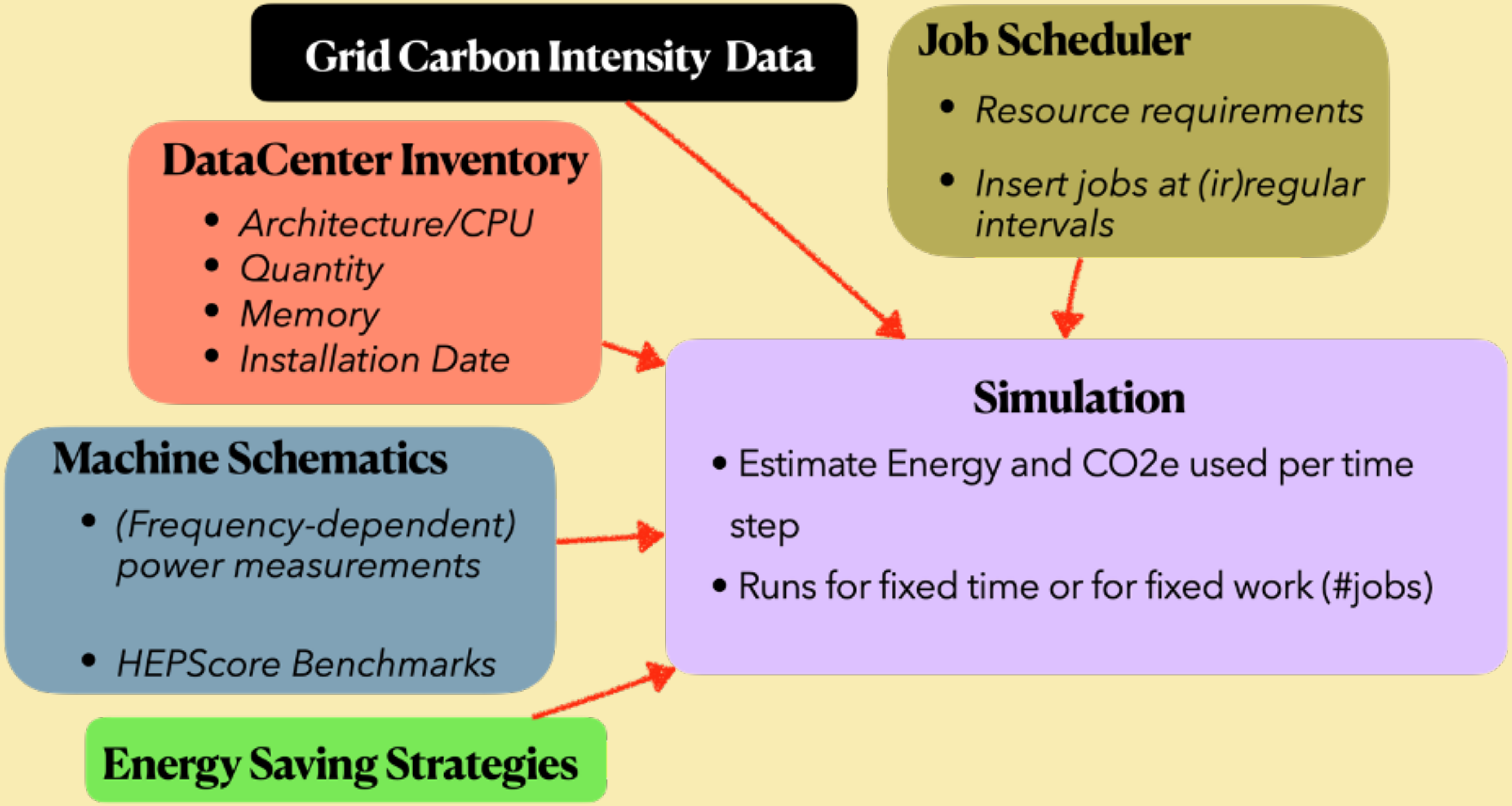
- Data-centres power budget are largely compute - so start there
- Machines/Servers have nodes which have threads, and each thread can be run in parallel to perform tasks.
- Jobs come in and request (use) threads and memory
- Assume that power used is



- $P_{Total} = P_{Idle} + \frac{\text{Threads Used}}{\text{Threads available}} * (P_{Max} - P_{Idle})$ * For non hyper-threaded nodes
- $P_{Total} = P_{Idle} + \left(\frac{\text{Threads Used}}{\text{Threads available}} * (P_{Max} - P_{Idle})\right) * \frac{F_{current}}{F_{max}}$

The Data-Centre Simulation Framework

- Initially created at the University of Glasgow - Currently being expanded 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 CO_2 (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 %
```

17% reduction in jobs

25% peak time energy reduction

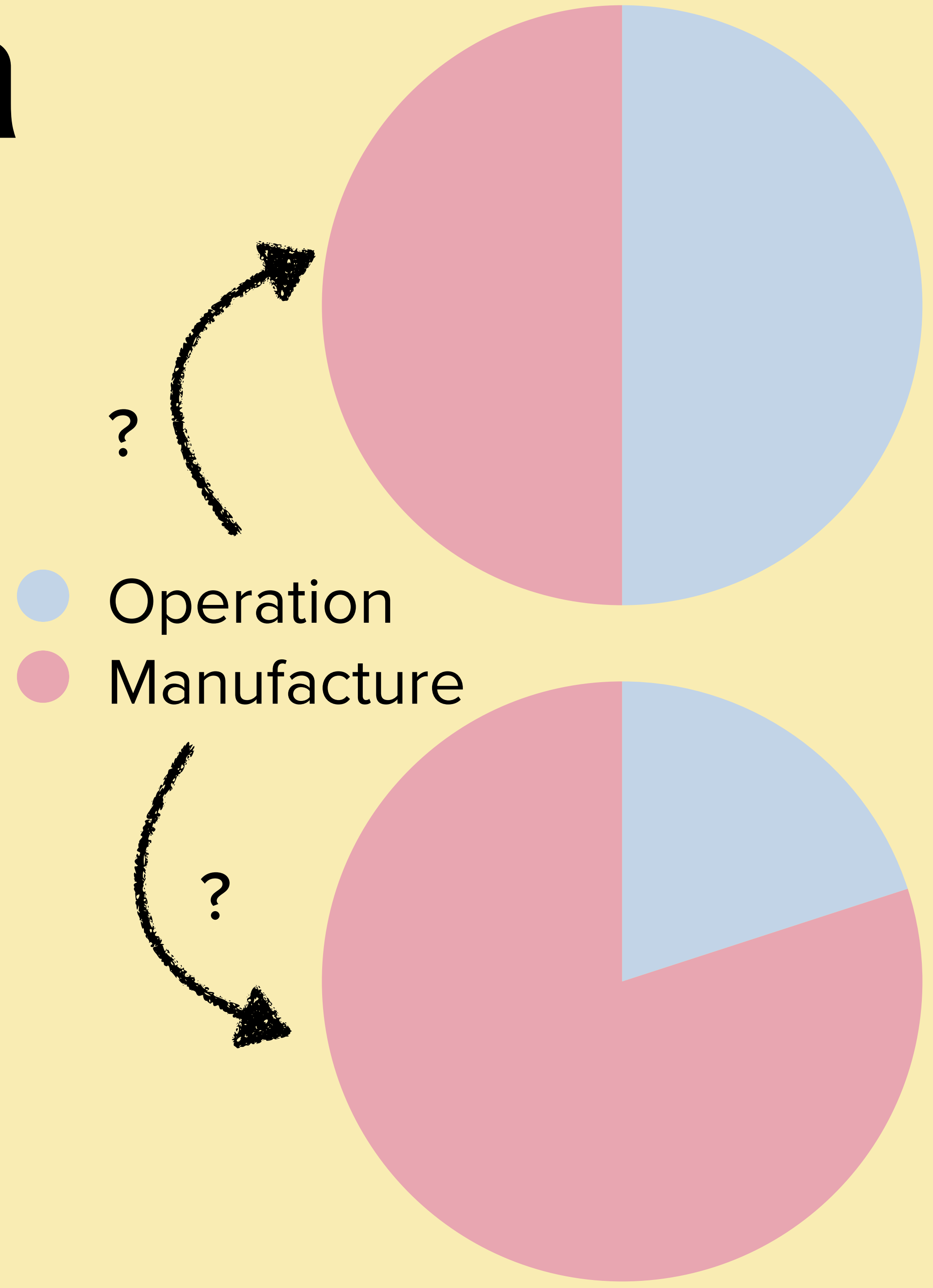
20% overall CO_2 reduction

Forecasted Clock-down

Each job produces 3% less CO_2

```
========
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 %
```

Side Note: Embedded Carbon



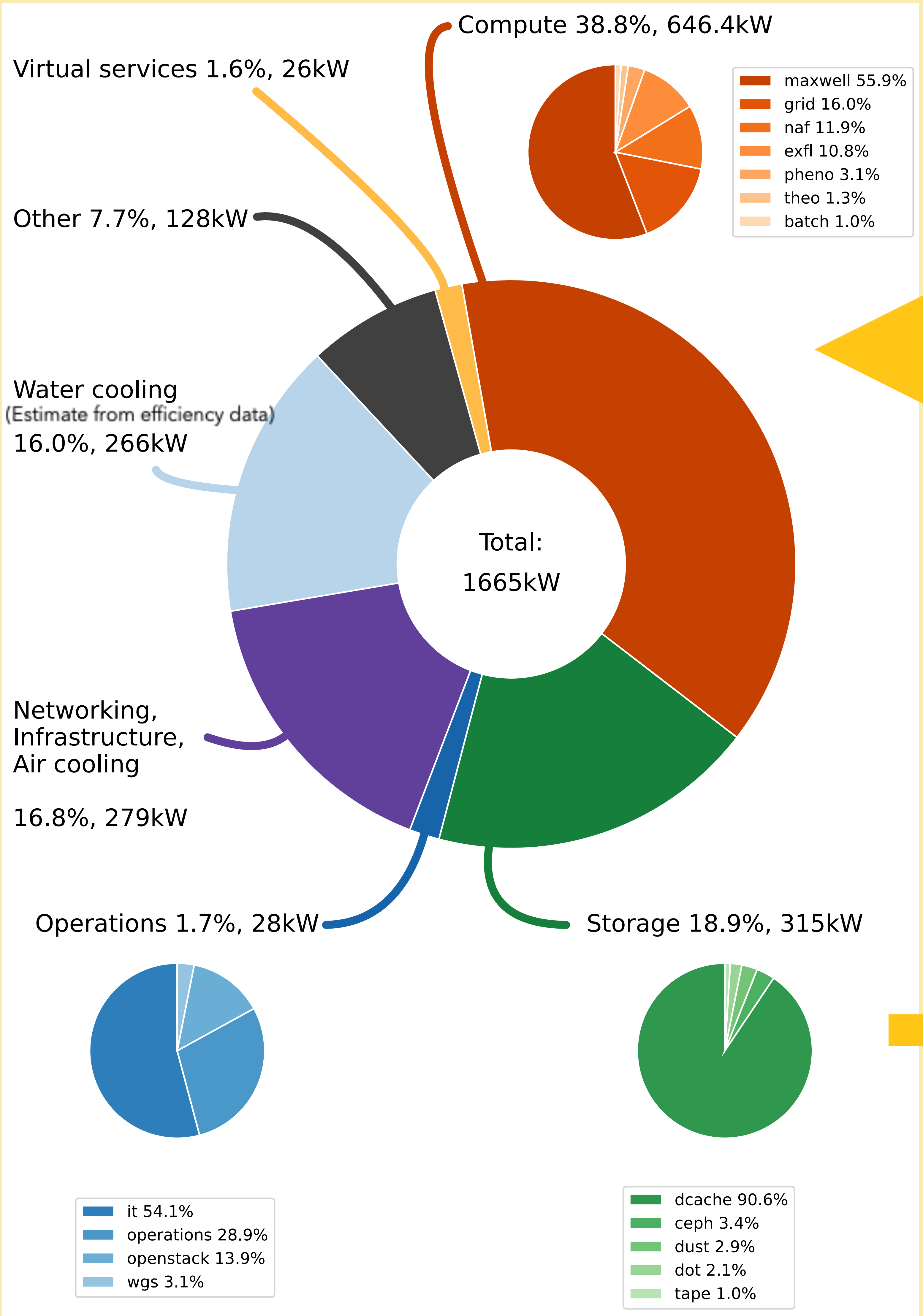
- A significant component of carbon in a servers lifetime is in the embedded carbon

- Need to start pressuring hardware vendors to give us or produce some carbon lifecycle analyses - Procurement?

- The improvements tested are only on the carbon opportunity cost of RUNNING work. Assume an total operational carbon cost of Y and an embedded carbon cost of X

RF2.0 RESEARCH FACILITY TOWARDS A MORE ENERGY-EFFICIENT AND SUSTAINABLE PATH at DESY

- 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.



Median Power Consumption of the DESY Data-Centre

DESY DATA CENTRE

1. Learn

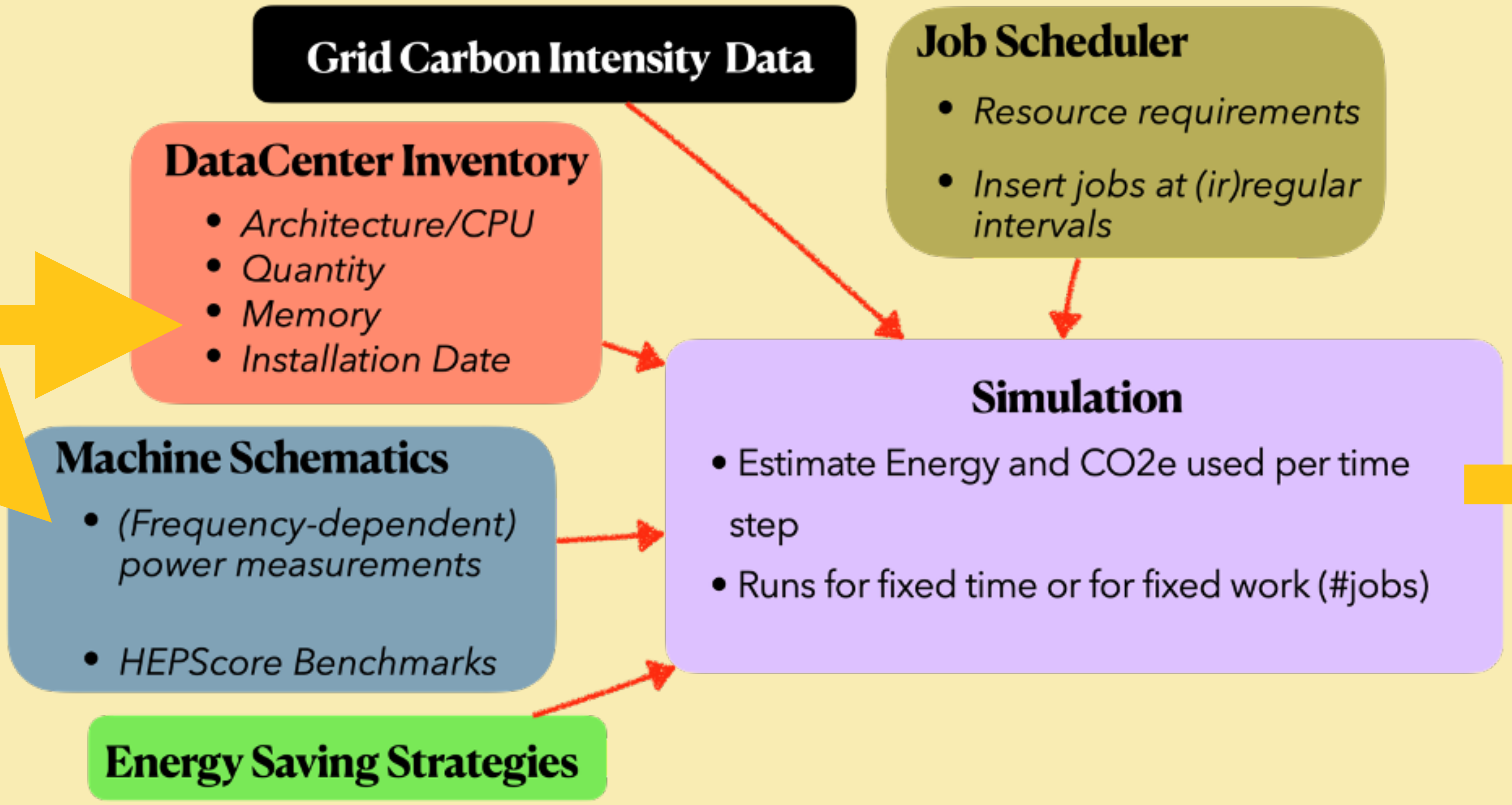
Take detailed and fine-grained measurements of data-centre components

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

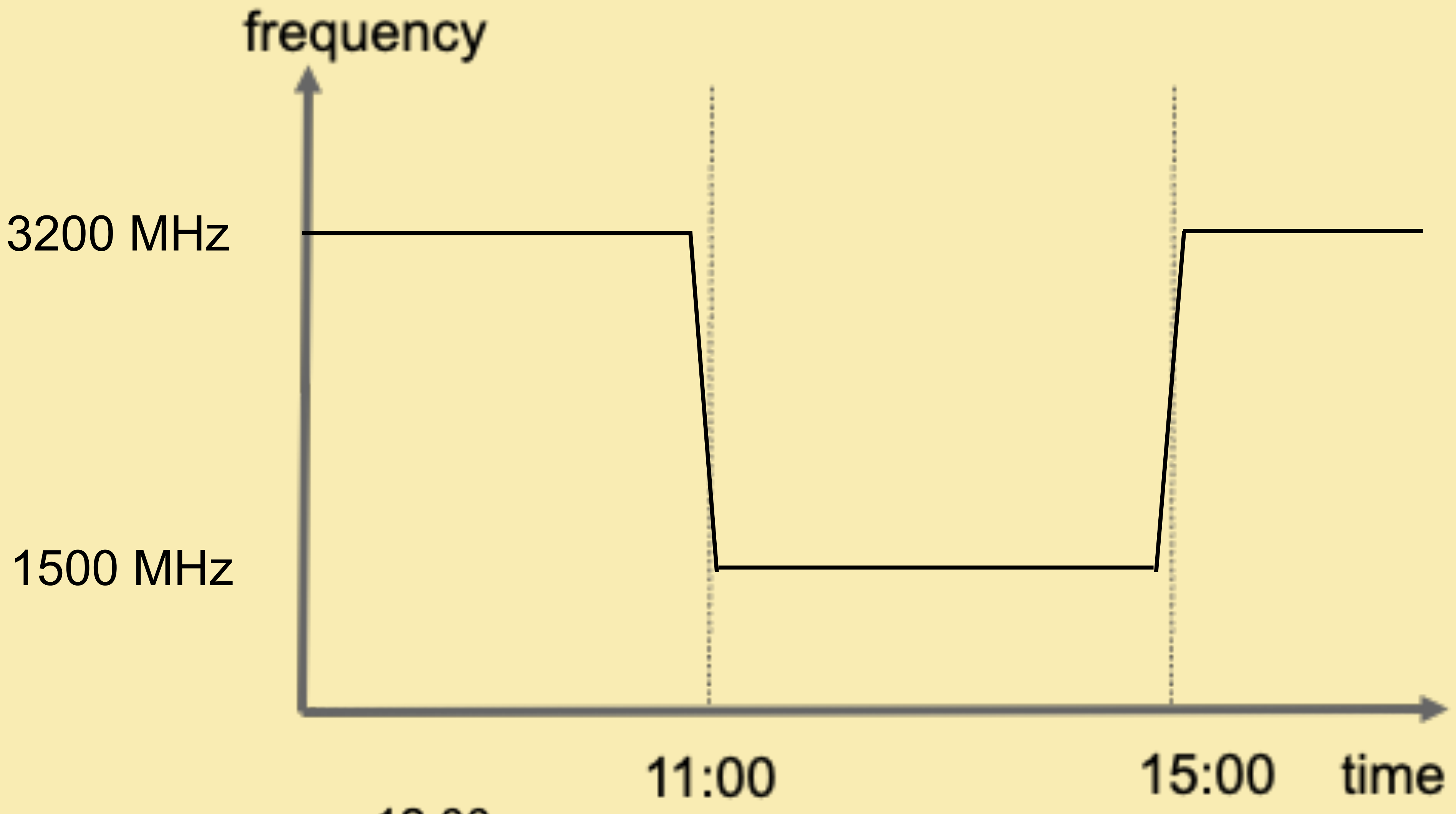
Take what we learn and apply it

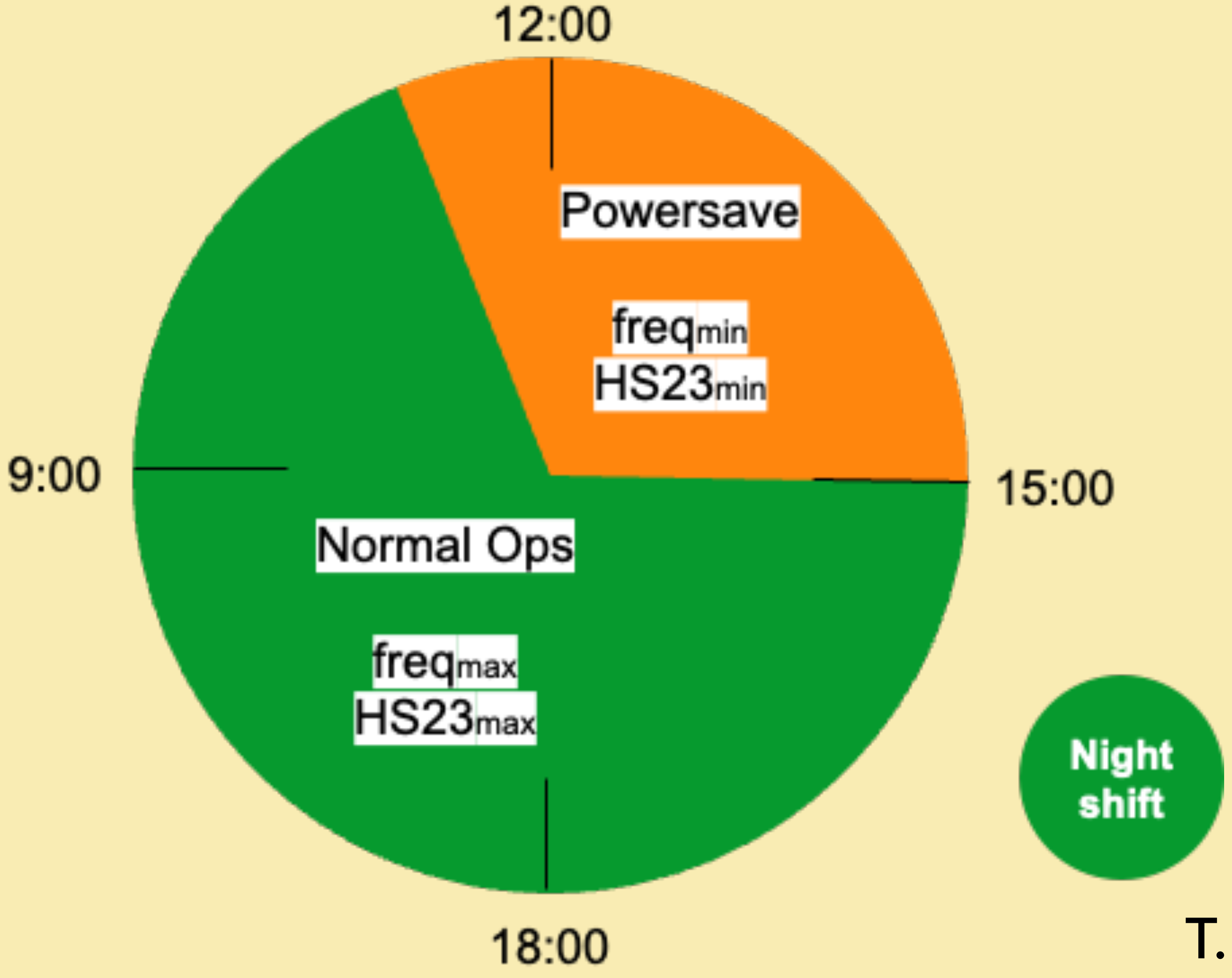
- Ultimate goal is to have the data centre be more flexible on demand.
    - Consume electricity when green energy is available, do not consume electricity when green energy is sparse
    - Try to run on the most effective frequencies
- Different clusters serve different communities, and each have their own needs, and therefore their own demands - no one size fits all solution.

Summer Savings at DESY:
Test of Frequency Scaling Capabilities

<u>Grid HTC Cluster</u>

- 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.
  - = 20/24 $HS23_{max}$ + 4/24 $HS23_{min}$





T.Hartmann et al

Summer “Savings" at DESY:
Upcoming nodes-on-demand test

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?

→

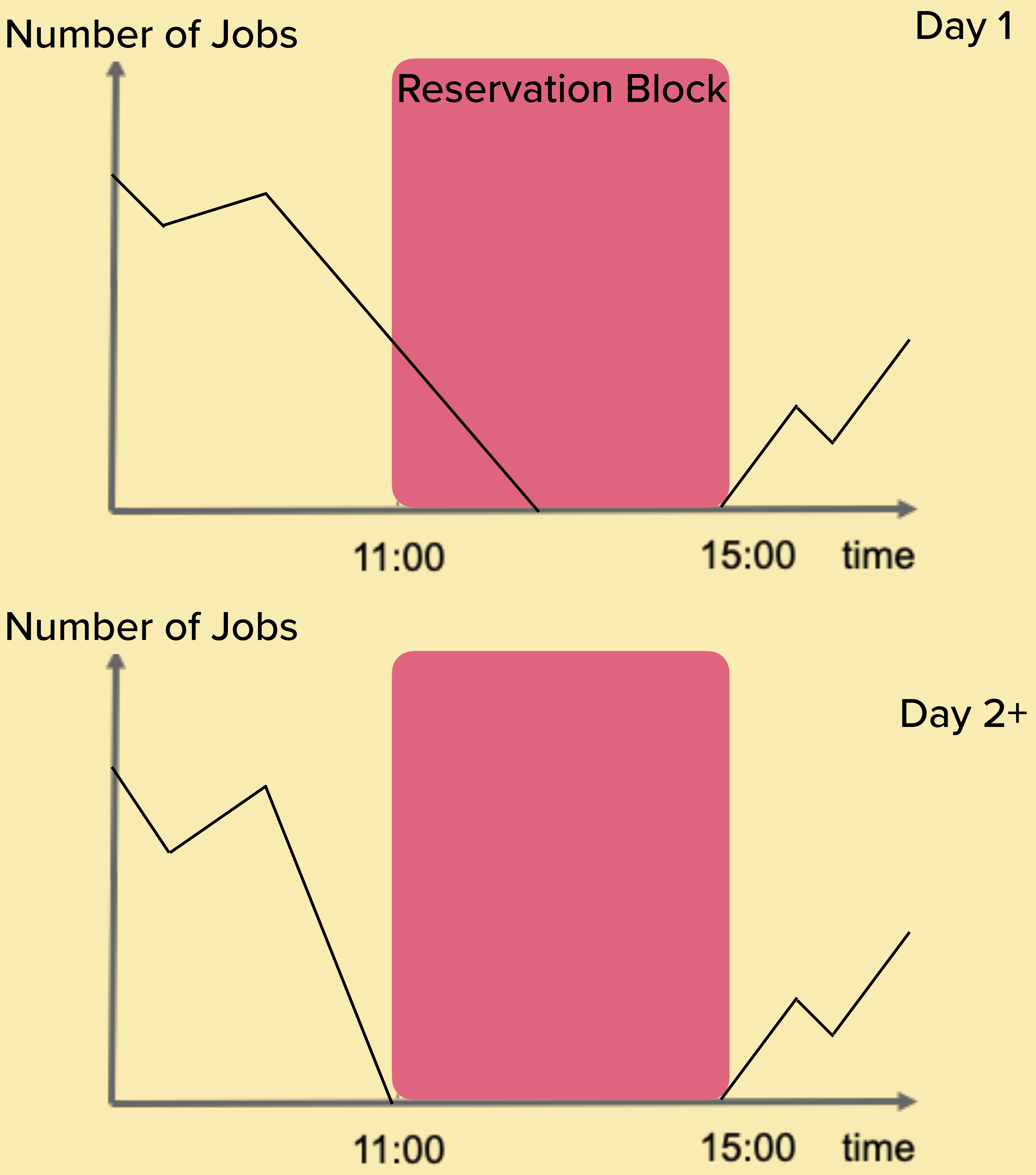
Implement Condor Rooster Config to shut idle instances down, boot offline instances
➡ use only what we need

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

Summer "Savings" at DESY:
Upcoming nodes-on-demand test

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.

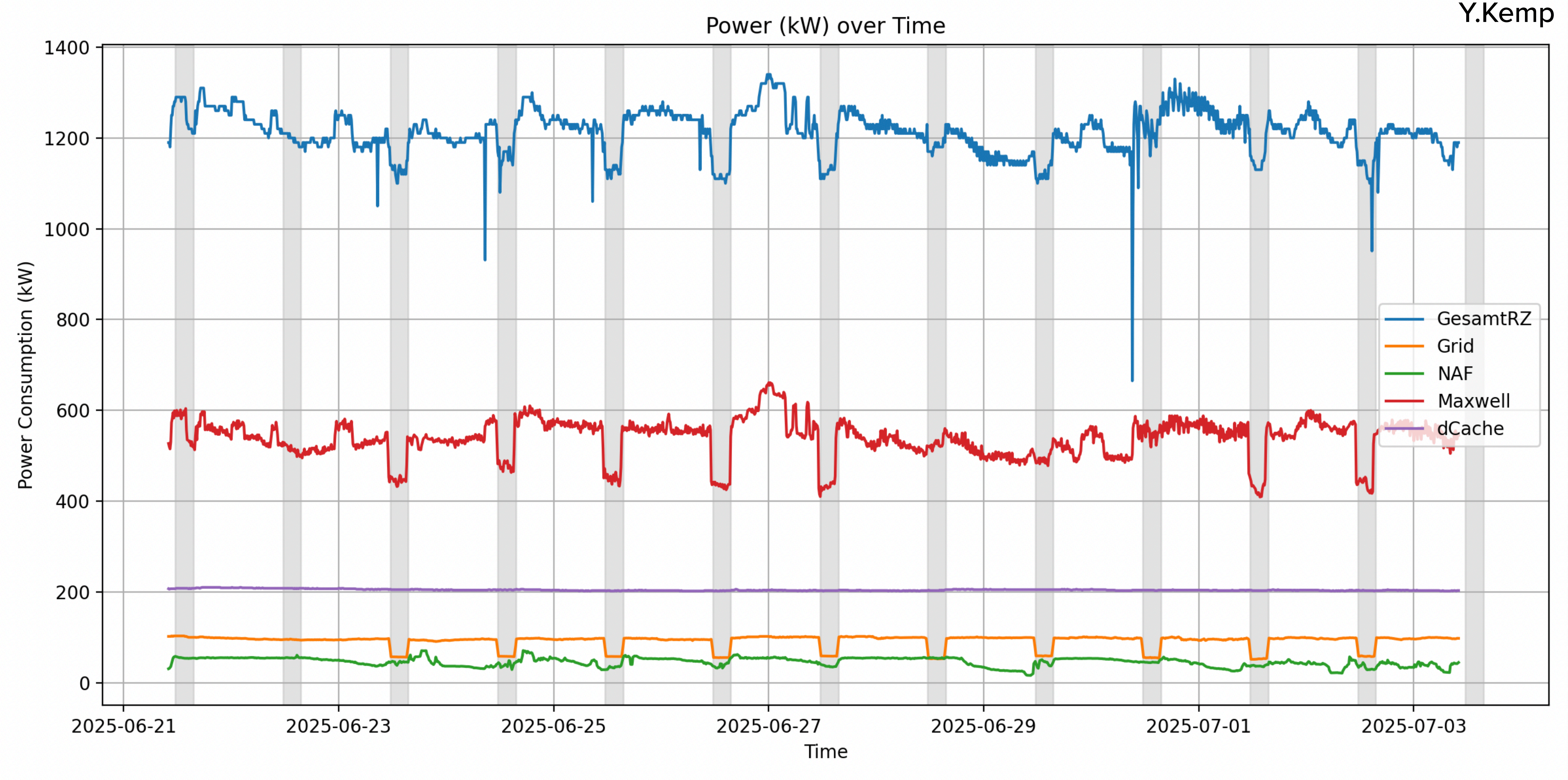


Summer Savings at DESY: Provisional Results

Plotting the power reported by the machines in each of the clusters

Across the week saw ~3300 kWh saved

With minimal? disruption to research



The Sustainable Computing You

What can we do for Sustainability?

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

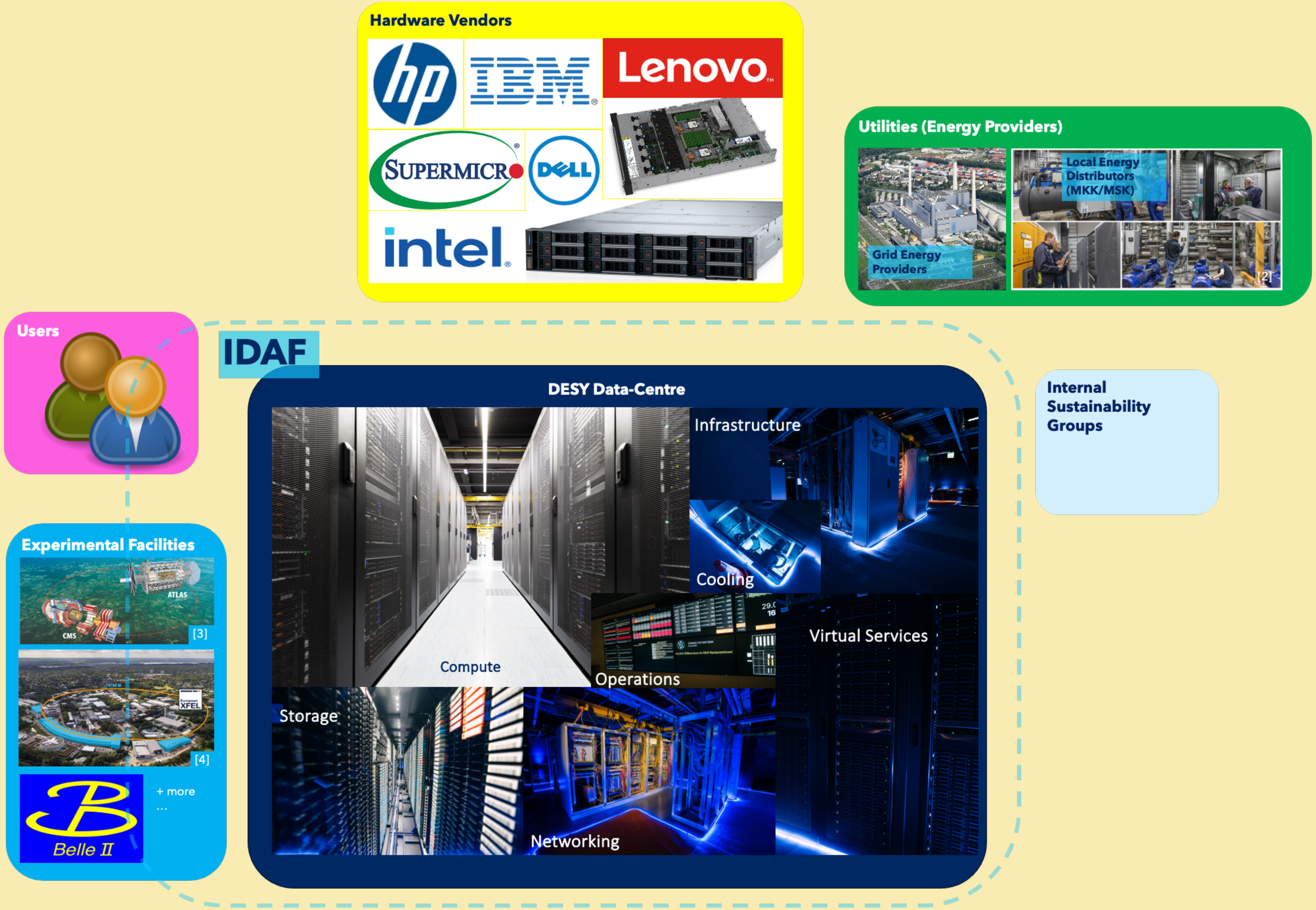
- Compute node clock speeds
- [Vendors] Embedded carbon from servers
- [Utilities] Total energy used during high demand

Reuse

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

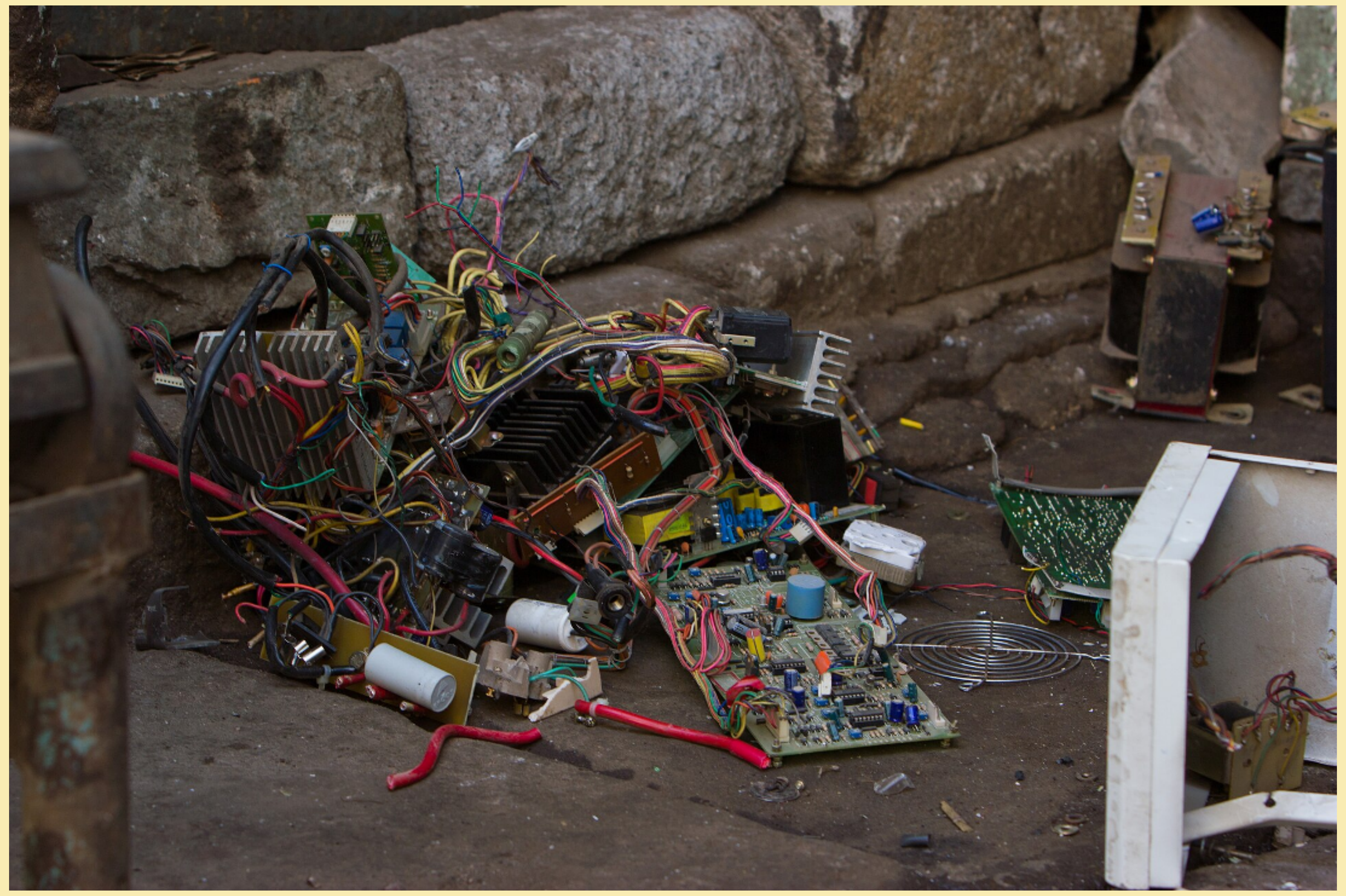
Repair/Regift/Recycle

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



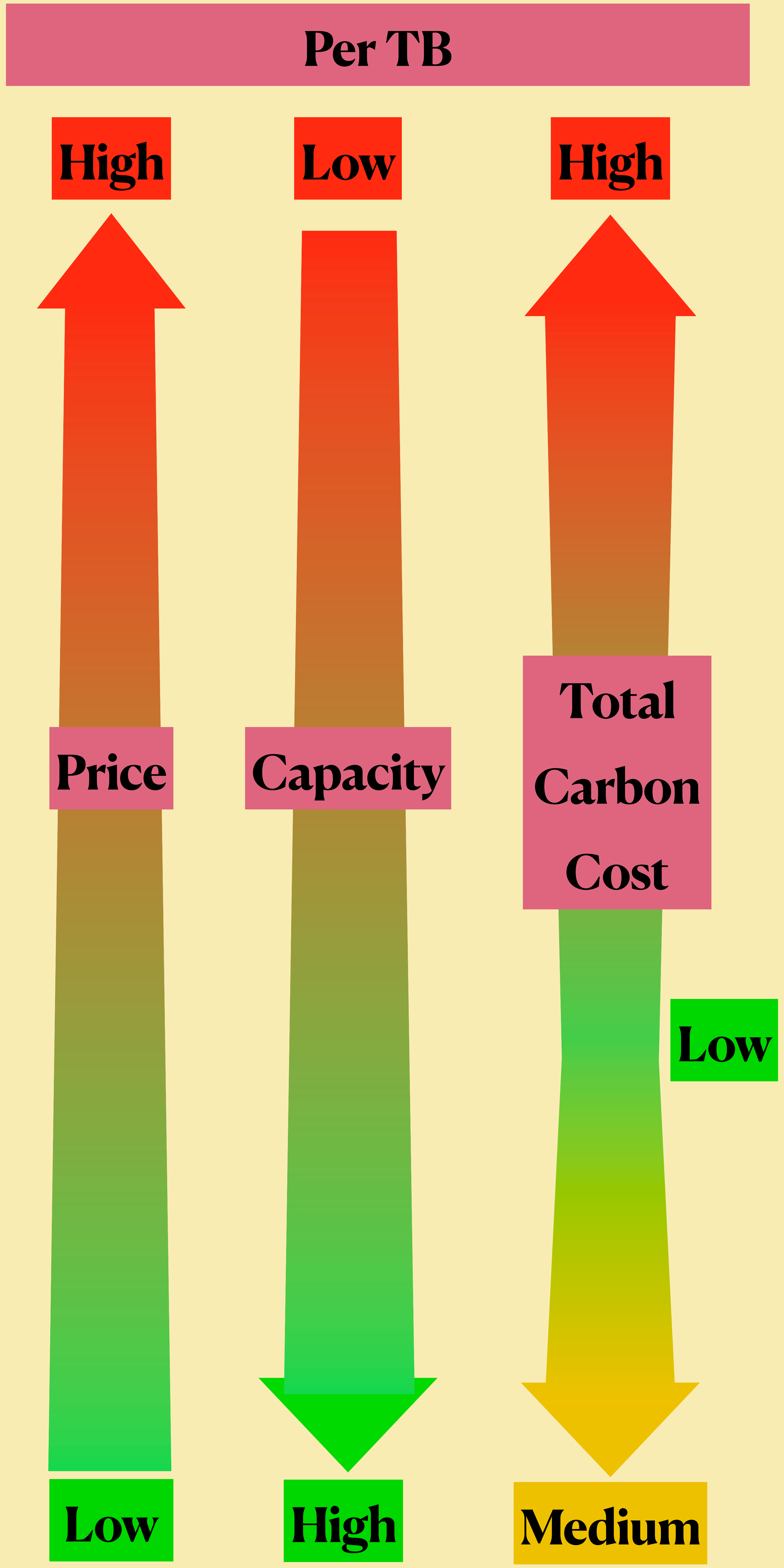
What can you do for sustainability

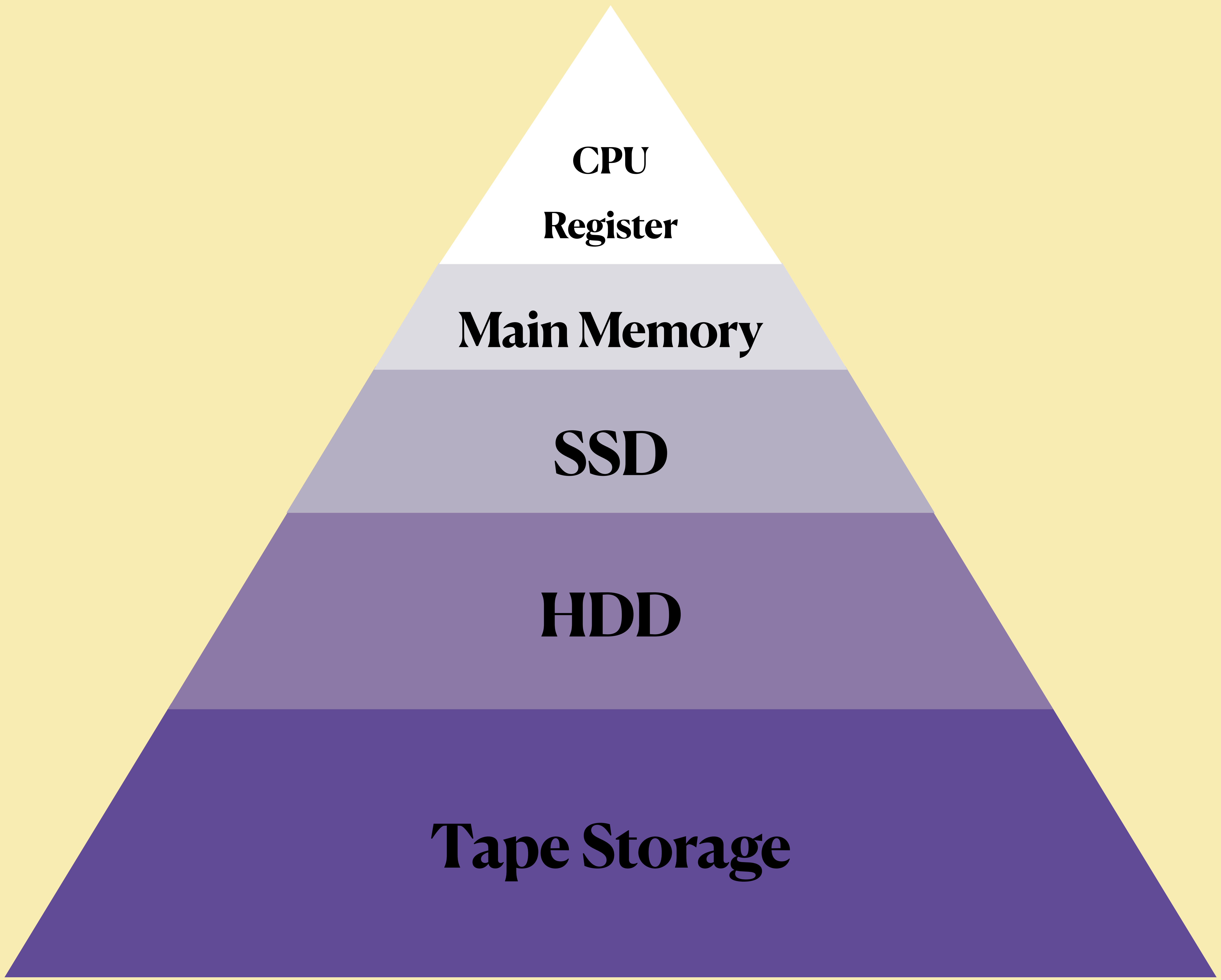
(E-)waste



- Datasets Generation - Do you need to save multiple copies of your data?
  - How long to need to retain them for?
  - Are they stored elsewhere in the data-centre?
  - Are they annotated, could other people use it?
- Dataset Storage - Where are you storing data, Disk, tape?
- Local Equipment - Laptops have higher embedded carbon than servers - use them as long as possible and dispose of them properly so that most of it is recycled
  - Can you buy second-hand? Do you need the latest and greatest technology?
- Search Engines - Use a better search engine on your computer and your phone : don’t use AI by default when a simple search can do. Think before you AI
- Don’t fly to Venice on your private jet.

Storage Technology Hierarchy

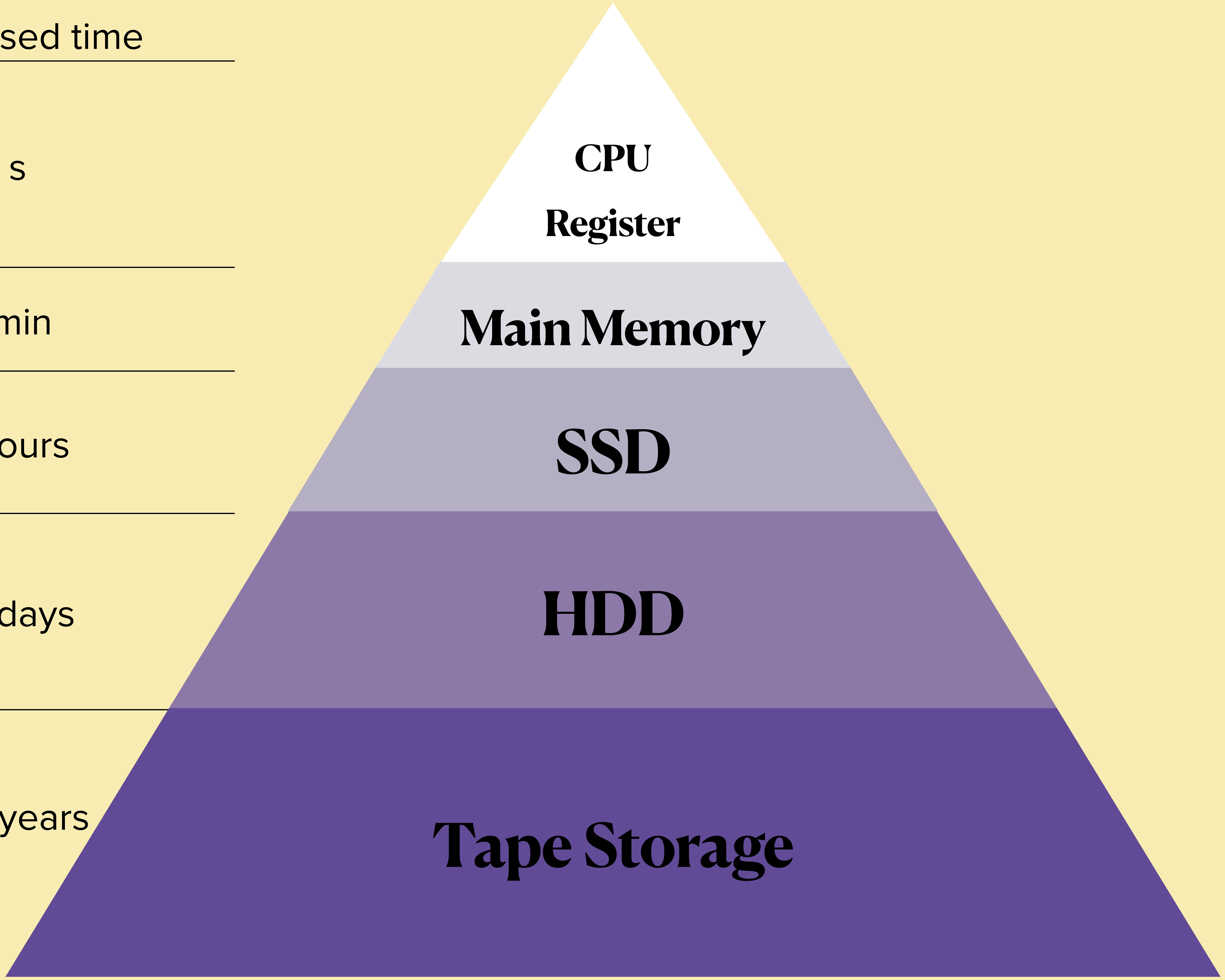




Slide adapted from T. Mykrchyan

Storage Technology Hierarchy

Access Speed (ns)	Normalised time
0.5	1 s
100	3 min
16,000	9 hours
5,000,000	116 days
5,500,000,000	349 years



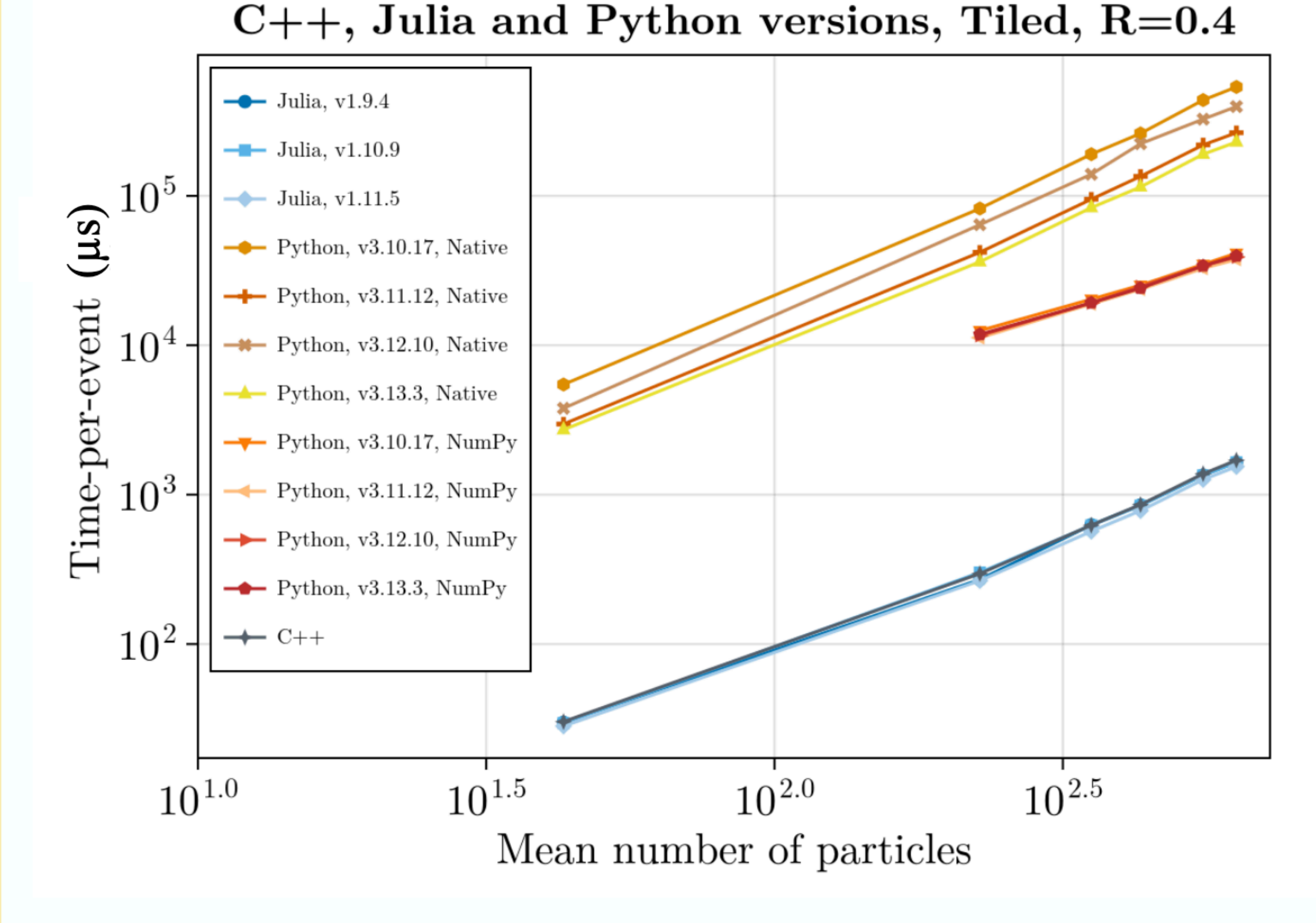
Slide adapted from T. Mykrchyan

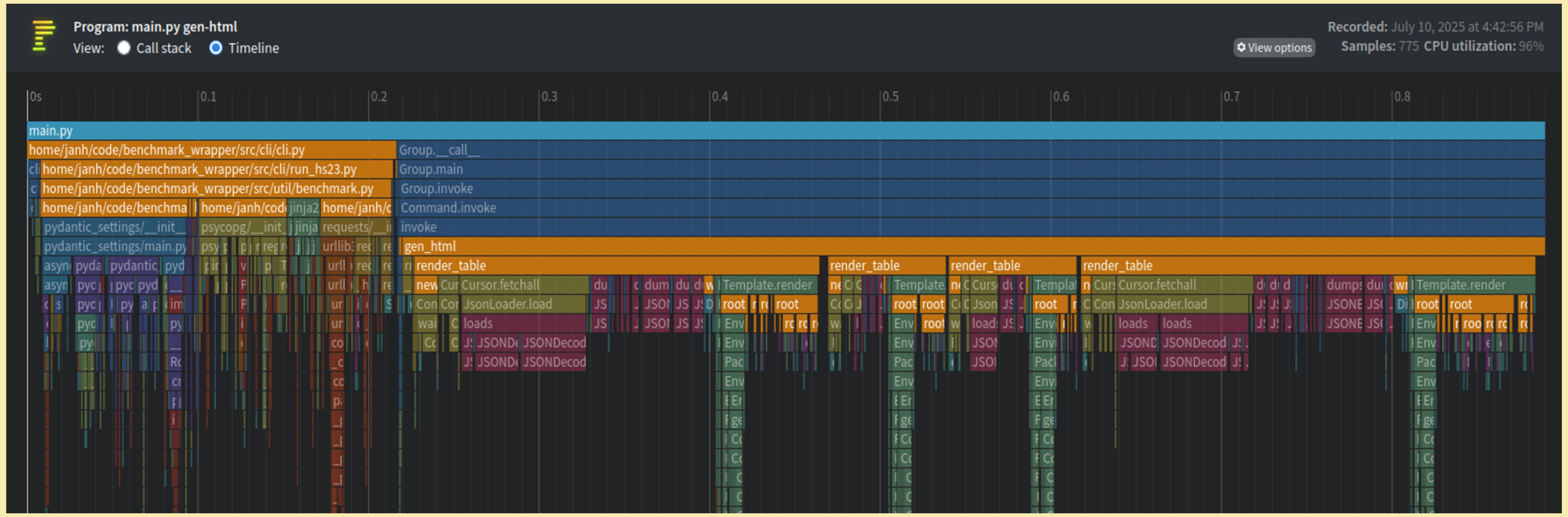
What can you do for sustainability

Code-writing

- Can you reduce the time your code runs for?
- Use a debugger and Profiling tools! Is a lot of processing time used in a place you didn't expect? (IT department can help!)
- Use checkpointing for longer computations ➜ reduce CPU consumption in case of failures!
- Write in more performant language from the start.
- Documentation!
- Learn to write better code, check in and upgrade your coding skills with a course every now and again.
- Multi-threaded coding (warning, here be dragons)

[Graeme Stewarts talk (https://indico.cern.ch/event/1338689/contributions/6009700/attachments/2951318/5189394/Julia%20in%20HEP.pdf)](https://indico.cern.ch/event/1338689/contributions/6009700/attachments/2951318/5189394/Julia%20in%20HEP.pdf)





What can you do for sustainability

Running and Submission

- Rethink: Run a small subset (10 events when you have 1000 to do) on your laptop to bug-check.

- Reduce: Do you need to use our compute resources? Can you run your work on your laptop?
  - When should you want to access DESY data-centre compute resource?
  - How does one get access?
    - When requesting resource you usually can specify RAM and #cores/threads - do you need that much? Checking log files to see what you actually used and then adjusting.
  - **Matters not what you run but how long it runs for**

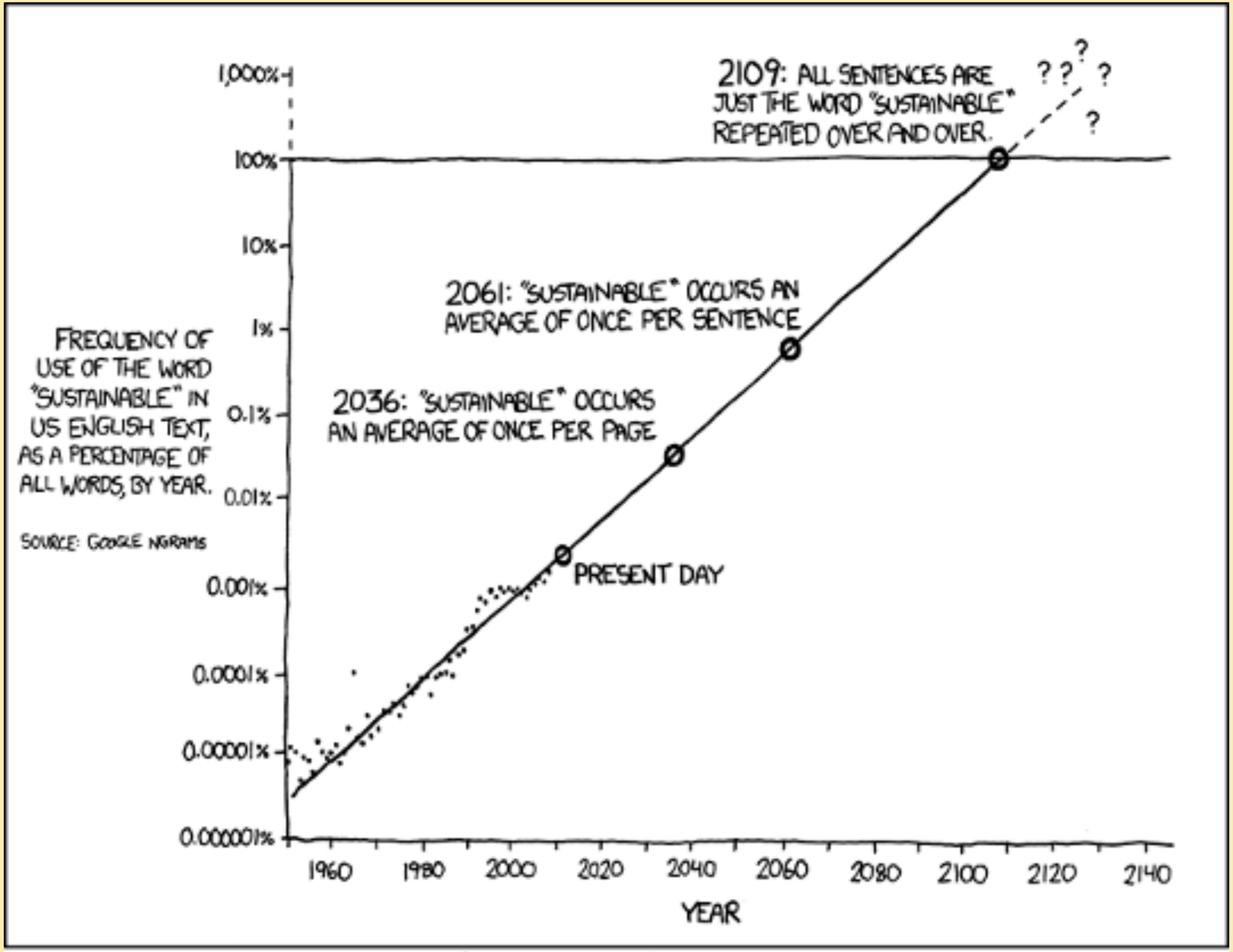
- Recycle: Try to run your work where your data is - avoids unnecessary copies or expensive file transfers

https://www.meme-arsenal.com/en/create/template/11824993

Conclusions

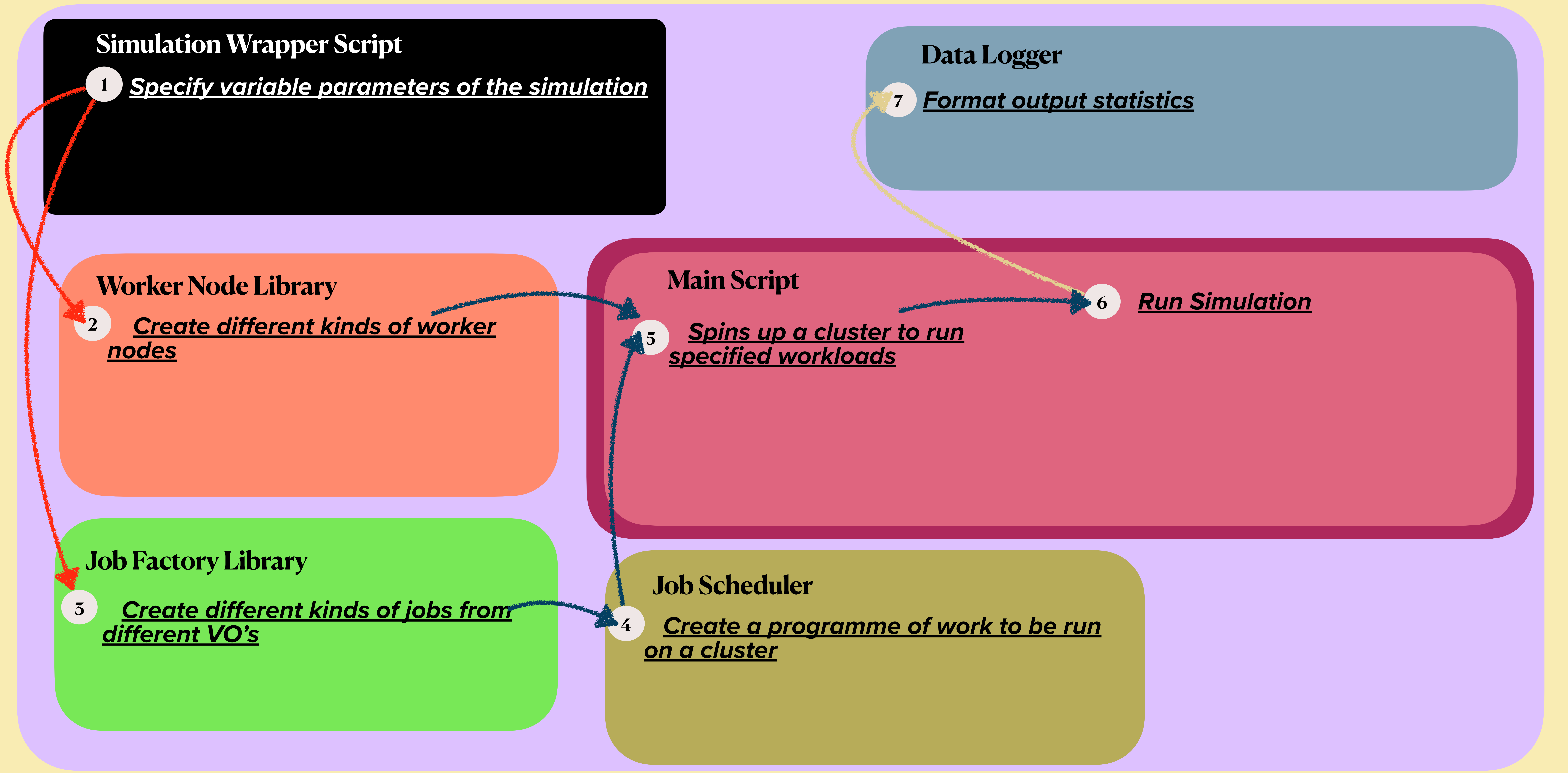
- IT waste will be important in the future, and we should all work to reduce our impact. We at DESY are actively researching what we can do help
    - Work is underway/mostly complete to collect, monitor and measure data-centre nodes for simulation and analysis.
    - A simulation framework has been created to try and test different kinds of operation of data-centres. It is currently private, but the plan is to make it freely available soon
    - Implementing findings from the above at the datacenter.
- Fostering more involved relationships with communities to change behaviours.
- You can help now (while you are here at DESY) by being educated and interested, and in the future by taking these practices to future research institutions.

Backup



THE WORD "SUSTAINABLE" IS UNSUSTAINABLE.

Simple Simulation Schematic



Use Case 2 - What do different procurements look like?

- An example type of recommendation - Running fixed work of 50,000 jobs, what new machines will lower your impact? (Same number of new cores each)

No Changes (With Old Kit)

```
Total Simulated-time Duration        : 27.8 hours
Total Real-time Duration             :  1.0 minutes

Jobs Started                         : 50000
Jobs Finished                        : 50000

Total CPU duration                   : 250451.5 hours
Average CPU duration                 : 5.01 hours

Total energy consumed by compute     : 1362.10 kWh
Peaktime (5-9pm) energy consumption: 292.48 kWh
Average energy consumption per job : 27.24 Wh

Estimated CO2e emmissions            : 94.188 kg
Estimated Peaktime CO2e emmissions : 19.462 kg
Average CO2e emmissions per job      : 1.884 g
Peaktime CO2e emmissions percentage: 20.663 %
```

30% CO_2 reduction →

Replacing older nodes w/ x86 - AMD Siena

```
Total Simulated-time Duration        : 20.0 hours
Total Real-time Duration             :  0.6 minutes

Jobs Started                         : 50000
Jobs Finished                        : 50000

Total CPU duration                   : 259273.7 hours
Average CPU duration                 : 5.19 hours

Total energy consumed by compute     : 969.80 kWh
Peaktime (5-9pm) energy consumption: 211.61 kWh
Average energy consumption per job : 19.40 Wh

Estimated CO2e emmissions            : 66.048 kg
Estimated Peaktime CO2e emmissions : 13.810 kg
Average CO2e emmissions per job      : 1.321 g
Peaktime CO2e emmissions percentage: 20.909 %
```

32% CO_2 reduction →

Replacing older nodes w/ ARM - AltraMax M128-30

```
Total Simulated-time Duration        : 18.0 hours
Total Real-time Duration             :  0.5 minutes

Jobs Started                         : 50000
Jobs Finished                        : 50000

Total CPU duration                   : 252801.8 hours
Average CPU duration                 : 5.06 hours

Total energy consumed by compute     : 939.53 kWh
Peaktime (5-9pm) energy consumption: 217.55 kWh
Average energy consumption per job : 18.79 Wh

Estimated CO2e emmissions            : 63.599 kg
Estimated Peaktime CO2e emmissions : 14.197 kg
Average CO2e emmissions per job      : 1.272 g
Peaktime CO2e emmissions percentage: 22.323 %
```

Dynamic control of power usage

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)

Decision Tool - “Comptroller”

Internal Input (States)

Decision Tool

➡ **Power available to consume?**
- boot instances
- set more aggressive cpu governor

➡ **Do we need to scale down?**
- shutdown lowest HS/W first
- try not to lose running compute jobs

Actions

- /instances/batchYYYY
  - Set governor “powersave”
- /instances/batchZZZZ
  - shutdown
- /instances/controller
  - Boot “batchWWWW”

Monitoring

Post state and resource consumption of servers

Instances/controller-agents

batchXXXX-agent

Send Actions to be performed on servers

Run actions on servers and update machine states

How to build and test this Comptroller?

<u>Work "Backwards"</u>

- Create scripts that can automatically change the states of machines.
- Test various capabilities individually
- Evaluate effectiveness using monitoring