Green Computing at Dutch National Supercomputing Center

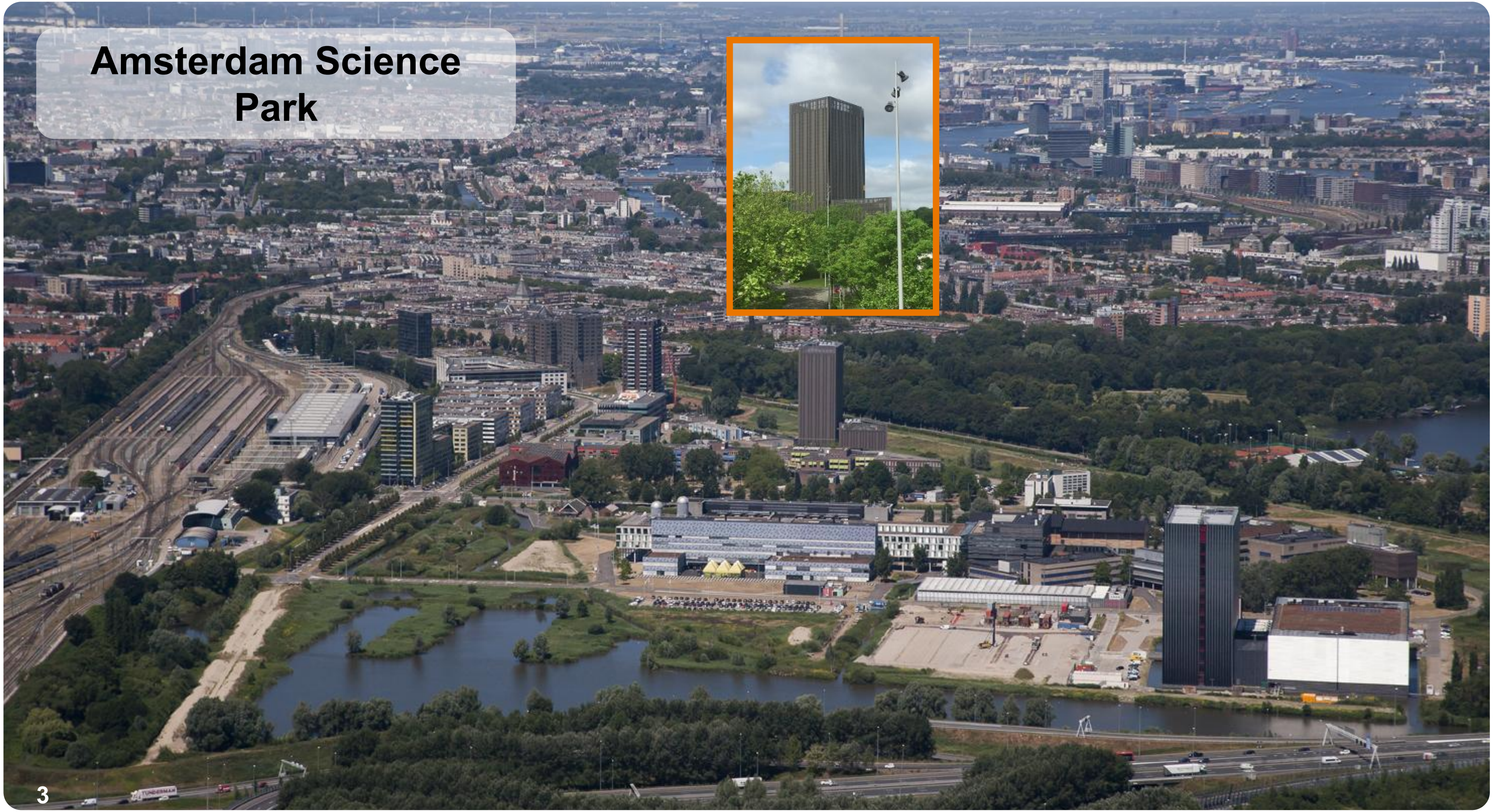
EGI : Green Computing webinar
Introduction session

Sagar Dolas
Program Lead - Future Computing (HPC, Quantum & Networking)
SURF Labs

Agenda

- *Introduction to SURF, Netherlands*
- *Energy as a design point for supercomputing operations*
- *Energy experiments on experimental systems*
- *EAR design & data collections*
- *Questions*

Amsterdam Science Park



SURF

Reliable, secure and innovative ICT infrastructure

Digital innovation and transformation of education and research

Knowledge exchange and trainings

Services development and integration with EU initiates

Snellius : Dutch National Supercomputer

Snellius

Phase 1:

- 76,832 cores (1.6 ×)
- 144 GPUs (3 Pflop/s, 14 ×)
- Total peak: 6.1 Pflop/s (3.4 ×)

Phase 1 and 2

- Total peak: 11.2 Pflop/s (6.2 ×)

- Full system, based on choice for Phase 3:
  - > 200,000 cores (> 4 ×)
  - Total peak: 13.6 – 21.5 Pflop/s (7.6 – 11.9 ×)

Snellius – Energy Consumption

- HPL Energy Consumption (typical use: 85%, idle use: ~ 25%)
  - Snellius
    - Phase 1: 620 kW (0.7 ×)
    - Phase 1 and 2: 1200 kW (1.4 ×)
    - Full system (“worst case”: Phase 3 GPU): 1430 kW (1.6 ×)
- Average energy consumption based on phasing
  - 2021–2022: ~ 1 ×
  - 2023: 1.5 ×
  - 2024 and later: 1.6 ×

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse
Through hot & cold well

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling
Hot & cold island

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

Application analysis & tuning using EAR

User awareness

Platform for energy & performance visualisation

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

Application monitoring, analysis & tuning

User awareness

Platform for energy & performance visualisation

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling
Hot & cold island

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

Application monitoring, analysis & tuning

User awareness

Platform for energy & performance visualisation

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling
Hot & cold island

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

Application monitoring, analysis & tuning

User awareness

Platform for energy & performance visualisation

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling
Hot & cold island

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

Application monitoring, analysis & tuning

User awareness

Platform for energy & performance visualisation

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling
Hot & cold island

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

Application monitoring, analysis & tuning

User awareness

Platform for energy & performance visualisation

Energy as a design point for supercomputing operations

Datacenter

PUE < 1.19

Waste heat reuse

Use of Hydroelectricity

Infrastructure Level

Rear door heat exchanger + Direct water cooling

Hot & cold island

CPU + GPU architecture

System Level

Energy & power management, energy capping, Energy accounting

Application Level

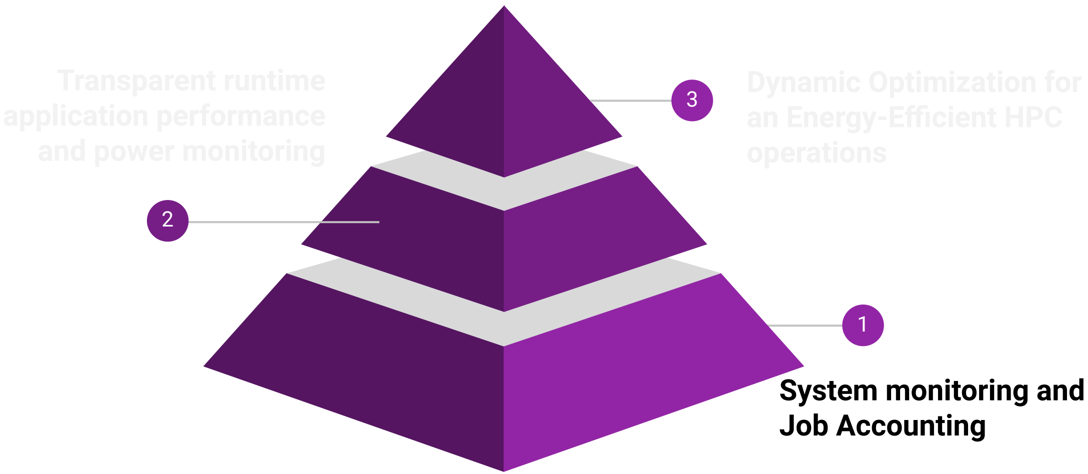
Application monitoring, analysis & tuning

User awareness

Platform for energy & performance visualisation

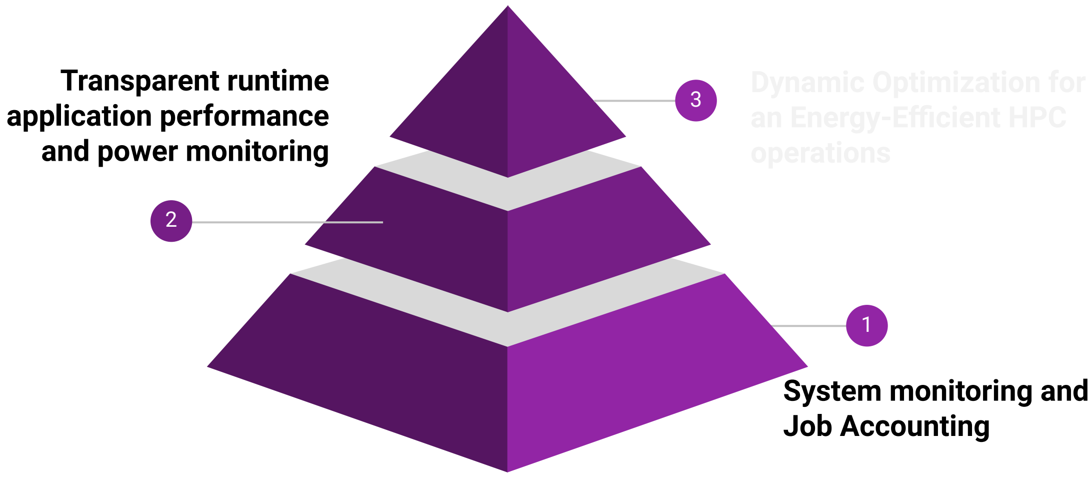
EAR software

EAS main values

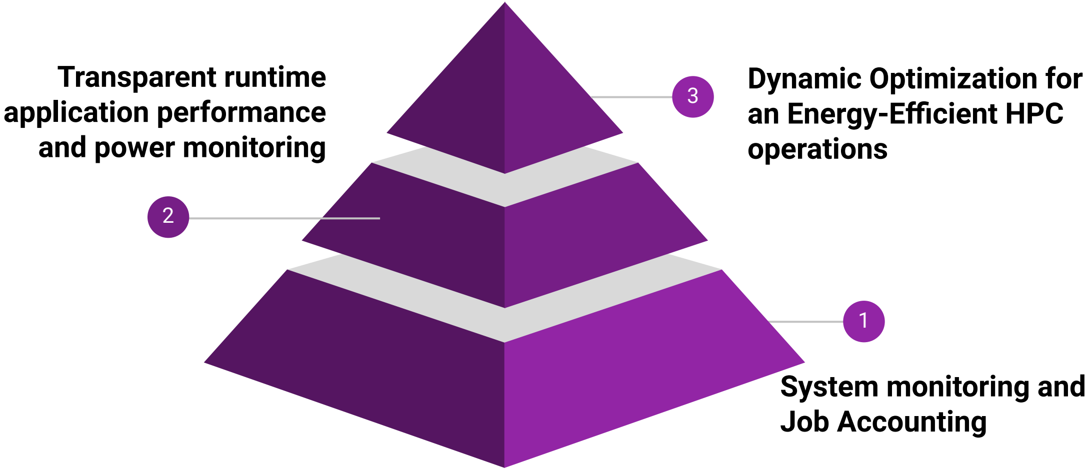


EAR software

EAS main values



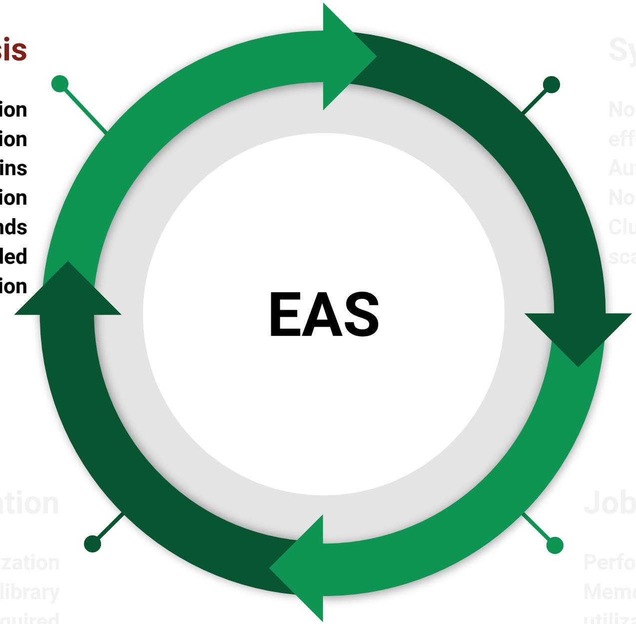
EAS main values



Energy-efficiency process

Analysis

Hints for application analysis and optimization
Hints for system analysis and optimization
Reporting system, extensible based on plugins
Trace file generation
Energy accounting and reporting commands included
Graphana visualization



System Monitoring and Control

Nodes temperature, power (Node, DRAM. Package, GPU), effective frequency …
Automatic reporting of run time hardware issues
Node powercap: heterogeneous nodes, power balance
Cluster powercap : Hierarchical architecture for system scalability

Optimization

Runtime application energy optimization
Transparent, dynamic and lightweight runtime library with no user intervention required
Automatic energy savings according to energy policies

Job Accounting

Performance metrics monitoring: CPI, CPU Gflops, Memory bandwidth, MPI hints, IO bandwidth, GPU utilization
Granularity: jobid, stepid, “loop”, user, node

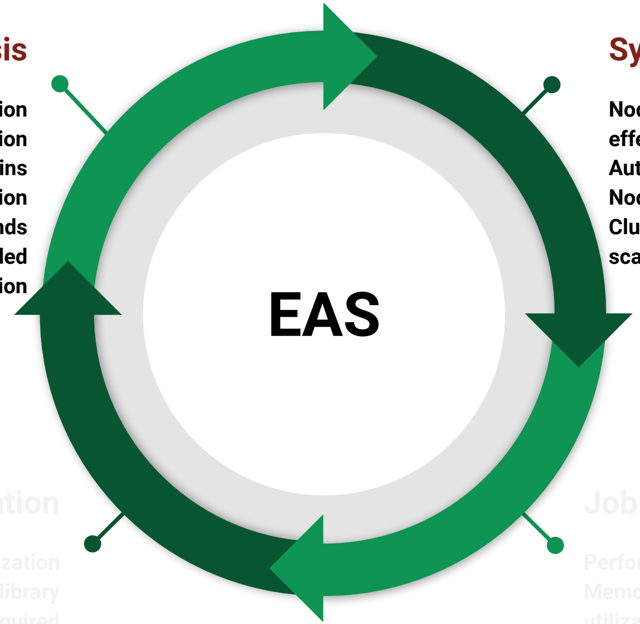
Energy-efficiency process with EAS technology

Analysis

Hints for application analysis and optimization
Hints for system analysis and optimization
Reporting system, extensible based on plugins
Trace file generation
Energy accounting and reporting commands included
Graphana visualization

System Monitoring and Control

Nodes temperature, power (Node, DRAM. Package, GPU), effective frequency …
Automatic reporting of run time hardware issues
Node powercap: heterogeneous nodes, power balance
Cluster powercap : Hierarchical architecture for system scalability



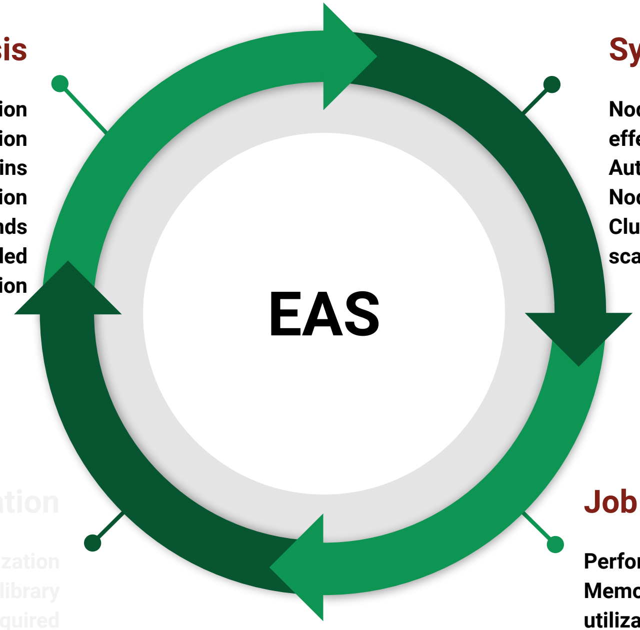
Optimization

Runtime application energy optimization
Transparent, dynamic and lightweight runtime library with no user intervention required
Automatic energy savings according to energy policies

Job Accounting

Performance metrics monitoring: CPI, CPU Gflops, Memory bandwidth, MPI hints, IO bandwidth, GPU utilization
Granularity: jobid, stepid, "loop", user, node

Energy-efficiency process with EAS technology



Analysis

Hints for application analysis and optimization
Hints for system analysis and optimization
Reporting system, extensible based on plugins
Trace file generation
Energy accounting and reporting commands included
Graphana visualization

System Monitoring and Control

Nodes temperature, power (Node, DRAM. Package, GPU), effective frequency …
Automatic reporting of run time hardware issues
Node powercap: heterogeneous nodes, power balance
Cluster powercap : Hierarchical architecture for system scalability

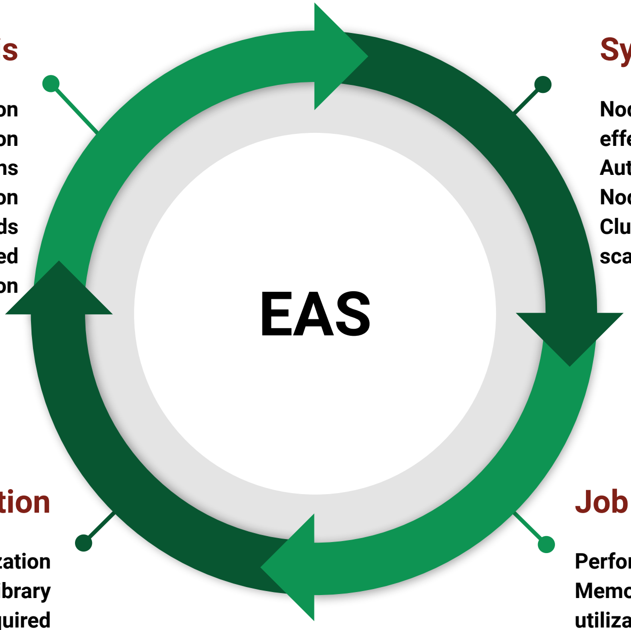
Optimization

Runtime application energy optimization
Transparent, dynamic and lightweight runtime library with no user intervention required
Automatic energy savings according to energy policies

Job Accounting

Performance metrics monitoring: CPI, CPU Gflops, Memory bandwidth, MPI hints, IO bandwidth, GPU utilization
Granularity: jobid, stepid, “loop”, user, node

Energy-efficiency process with EAS technology



Analysis

Hints for application analysis and optimization
Hints for system analysis and optimization
Reporting system, extensible based on plugins
Trace file generation
Energy accounting and reporting commands included
Graphana visualization

System Monitoring and Control

Nodes temperature, power (Node, DRAM. Package, GPU), effective frequency …
Automatic reporting of run time hardware issues
Node powercap: heterogeneous nodes, power balance
Cluster powercap : Hierarchical architecture for system scalability

Optimization

Runtime application energy optimization
Transparent, dynamic and lightweight runtime library with no user intervention required
Automatic energy savings according to energy policies

Job Accounting

Performance metrics monitoring: CPI, CPU Gflops, Memory bandwidth, MPI hints, IO bandwidth, GPU utilization
Granularity: jobid, stepid, “loop”, user, node

Energy optimization

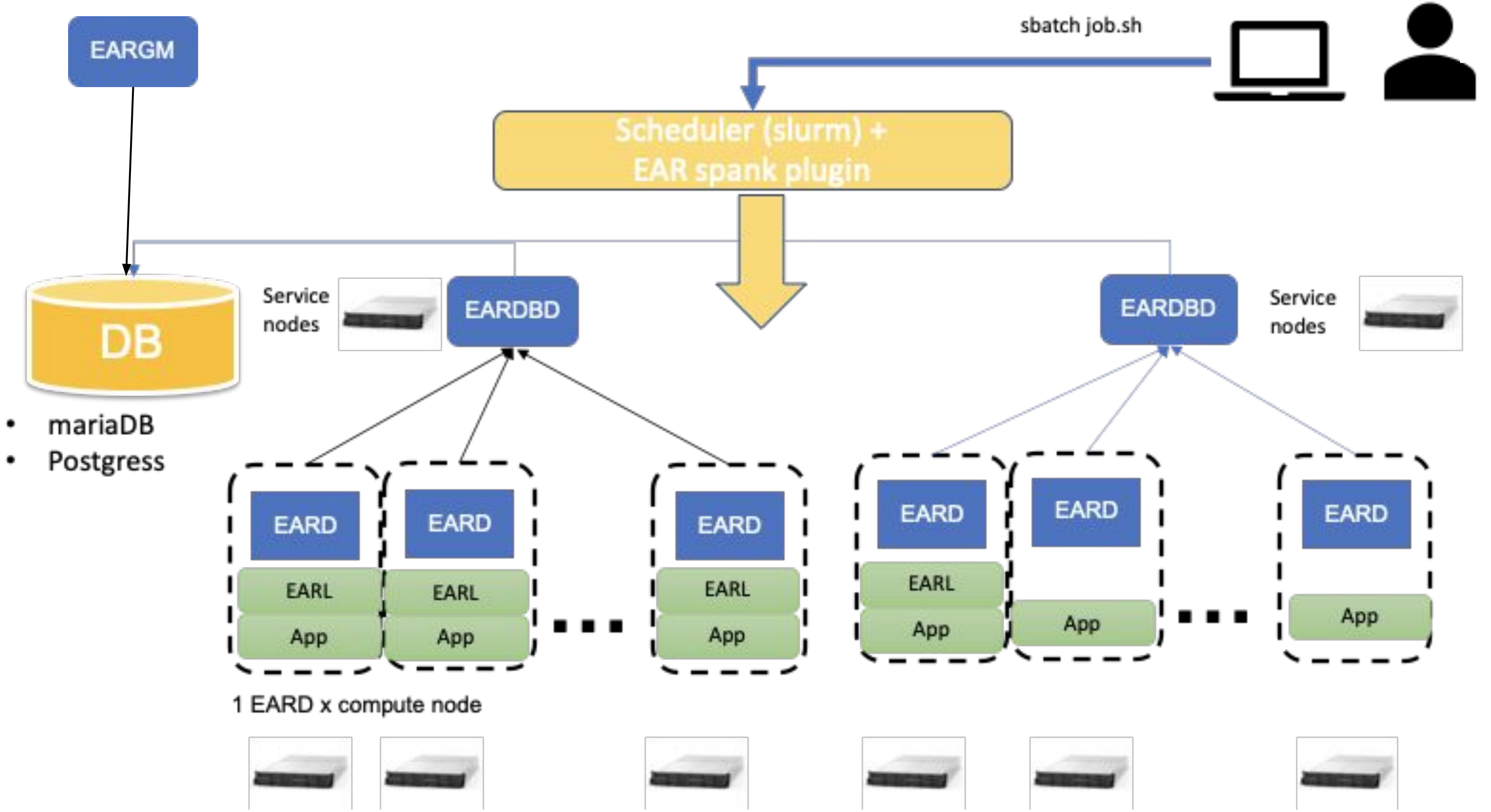
- Common to all the policies
    - At runtime, loop is detected and loop signature is computed
    - EAR uses time and power models for frequency selection
- **EAR-min_time** policy
    - Applications start at default policy frequency lower than nominal
    - **Loops with “enough” energy efficiency are accelerated** (compute bound)
    - Policy detects changes and applies again the policy
- **EAR- min_energy policy**
    - Applications start at default policy frequency (nominal)
    - **Frequency is reduced up to a maximum performance penalty**

Energy optimization

- Common to all the policies
  - At runtime, loop is detected and loop signature is computed
  - EAR uses time and power models for frequency selection
- **EAR-min_time** policy
  - Applications start at default policy frequency lower than nominal
  - **Loops with “enough” energy efficiency are accelerated** (compute bound)
  - Policy detects changes and applies again the policy
- **EAR- min_energy policy**
  - Applications start at default policy frequency (nominal)
  - **Frequency is reduced up to a maximum performance penalty**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EAR architecture overview

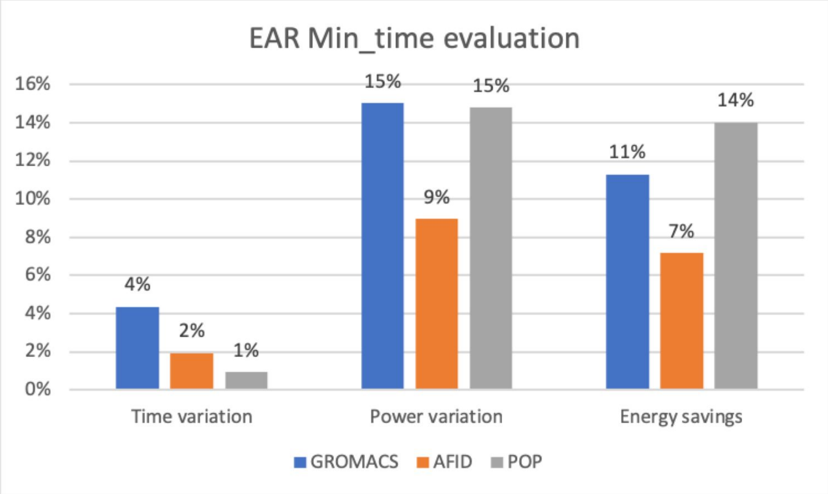


Experimental System for energy experiments

- Experiments system have 4x Lenovo SR650 system
- Skylake 6148 @2.5GHz 40c nodes with 100 g HDR
- 2x V100 nodes for GPU tuning experiments
- Default frequency=2.5GHz
- Easybuild software stack
- Jenkins for automated software builds
- User access for community
- EAR as energy management framework

Energy experiments with real use cases

- Application use cases provided by SURF Open Innovation Lab (SOIL) collaboration.
  - Experiments executed in Lenovo SD530 system
  - Skylake 6148 @2.4GHz 20c nodes with EDR network
  - Default frequency=2.0GHz
  - **GROMACS** 640 processes. 16 nodes
  - **AFID** 600 processes. 15 nodes
  - **POP** 400 processes. 10 nodes



– ***Average energy savings of 10%***

Visualisation Prototype

Other approaches for achieving energy efficiency

- ML based acceleration for HPC application
- Optimizing the number of simulation required to achieve particular outcome on HPC system
- Exploring Neuromorphic computing for scientific analysis & experimentation
- Introduce the concept of energy accounting

What next ?

1. In progress to operationalise on the Dutch national supercomputer
2. Large number of experiments will be carried out to fully understand tunable parameters
3. Assessing impact, operational and user requirements
4. Pilots will be carried out at SURF infrastructure jointly with member institutions.
5. Most important : knowledge dissemination : Joint hackathons, workshops and more.

What next ?

1. In progress to operationalise on the Dutch national supercomputer
2. Large number of experiments will be carried out to fully understand tunable parameters
3. Assessing impact, operational and user requirements
4. Pilots will be carried out at SURF infrastructure jointly with member institutions.
5. Most important : knowledge dissemination : Joint hackathons, workshops and more.

What next ?

1. In progress to operationalise on the Dutch national supercomputer
2. Large number of experiments will be carried out to fully understand tunable parameters
3. Assessing impact, operational and user requirements
4. Pilots will be carried out at SURF infrastructure jointly with member institutions.
5. Most important : knowledge dissemination : Joint hackathons, workshops and more.

What next ?

1. In progress to operationalise on the Dutch national supercomputer
2. Large number of experiments will be carried out to fully understand tunable parameters
3. Assessing impact, operational and user requirements
4. Pilots will be carried out at SURF infrastructure jointly with member institutions.
5. Most important : knowledge dissemination : Joint hackathons, workshops and more.

What next ?

1. In progress to operationalise on the Dutch national supercomputer
2. Large number of experiments will be carried out to fully understand tunable parameters
3. Assessing impact, operational and user requirements
4. Pilots will be carried out at SURF infrastructure jointly with member institutions.
5. Most important : knowledge dissemination : Joint hackathons, workshops and more.

Thank you !