DENERGIUM

Digital Energy Optimum

HPC-AI computation optimization based on the energy criterion

GREENDAYS 2024
The DENERGIUM summary

Energy-to-Solution

Increase the computation capacity

GLOBAL OPTIMIZATION WITH A GAIN UPTO 20%

Decrease the hardware cost

Decrease the financial cost

HPC – AI Market

Birthdate
02/2023
The HPC-IA workflow

Applications → cluster management and job scheduling system → Servers → Results
The HPC-IA workflow

Applications ➔ cluster management and job scheduling system ➔ Servers ➔ Results

Energy efficiency histogram
(x axis from 13% to 97%)

Energy efficiency gauge of computing nodes
The EnergyScopium Software Suite

**RECORD**
A software sensor that measures the energy consumption of different components reliably (CPU, GPU, RAM, etc.)

**MONITOR**
A model aggregates and transforms the raw data to track the global consumption of applications. Everything is then nicely displayed in energy dashboards and reports.

**OPTIMIZE**
A quasi-real-time access to the energy profile of applications is provided to highlight suboptimal phases in a program. Tuning is allowed and one can then optimize non-efficient phases.
EnergyScopium for users and developers

energyscopium jobid<10166268> nodes<['n231', 'n313']> ncpus <64> duration(hour)<22.51> estimated energy(kWh)<10.58> with adding cooling (PUE=1.3): (kWh)<13.76> and (gCO2 in FR)<687.78>
How EnergyScopium works?

Virtual Machine

Supported HW: Intel, AMD Epyc, Nvidia GPU, AMD GPU

compute nodes
## Optimization strategy

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
</table>
## Optimization strategy

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
</table>

**GAIN**

![Diagram showing GAIN with axes for Energy and Time]
## Optimization strategy

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
</table>

**Same application result**  
**Almost same duration (+2%)**  
**Energy savings (-22%)**

**Graphs:**
- Two graphs showing power consumption over time for CPU & GPU in different scenarios.
Optimization strategy

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
</table>

Same application result without AVX-512 - with AVX-512
Duration (/ 2), Energy (-40%)
Optimization strategy

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Time (sec)</th>
<th>Consumption (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graph showing consumption over time with high and low frequency states.
Optimization strategy

Application energy cost versus application duration

Energy (Wh) versus Duration (sec)
Optimization strategy

Application configurations versus server classes

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Matrice d'efficacité application/matériel**

<table>
<thead>
<tr>
<th></th>
<th>serveur 1</th>
<th>serveur 2</th>
<th>serveur 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1</td>
<td>50%</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>v2</td>
<td>60%</td>
<td>50%</td>
<td>20%</td>
</tr>
<tr>
<td>v3</td>
<td>x</td>
<td>x</td>
<td>80%</td>
</tr>
</tbody>
</table>

**Serveur type 1**
- processeurs=2, coeurs=24
- TDP=150W
- mémoire=8GB, GPUx0

**Serveur type 2**
- processeurs=4 coeurs=32
- TDP=120W
- mémoire=92GB, GPUx0

**Serveur type 3**
- processeurs=2, coeurs=16
- TDP=120W
- mémoire=16GB, GPUx1(TDP=300W)
### Optimization strategy

<table>
<thead>
<tr>
<th>Application</th>
<th>Server</th>
<th>Cluster</th>
</tr>
</thead>
</table>

#### Batch Scheduler

- **Unknown**
- **Known**
  > 80% of job submitted

- **legacy batch scheduler strategy**

- **Batch scheduler strategy based on energy/time, efficiency software/hardware**

- **Known**

- **Improve Knowledge**
Optimization strategy - Conclusion
Credits:

- https://mediatheque.inria.fr/Mediatheque/media/38324 (© Inria / Photo C. Morel)
- https://www.dell.com/fr-fr/dt/ai-technologies/index.htm#tab0=0
- https://www.flaticon.com/free-icons/deep-learning icons created by Becris

Flash the code to contact us
Contact@DENERGIUM.com