

**Retranscription automatique d'interviews  
par un modèle d'apprentissage profond pré-entraîné  
distribué en open-source, et implémenté en Python  
sur accélérateur graphique dans un cluster de calcul intensif.**

[damien.francois@uclouvain.be](mailto:damien.francois@uclouvain.be)

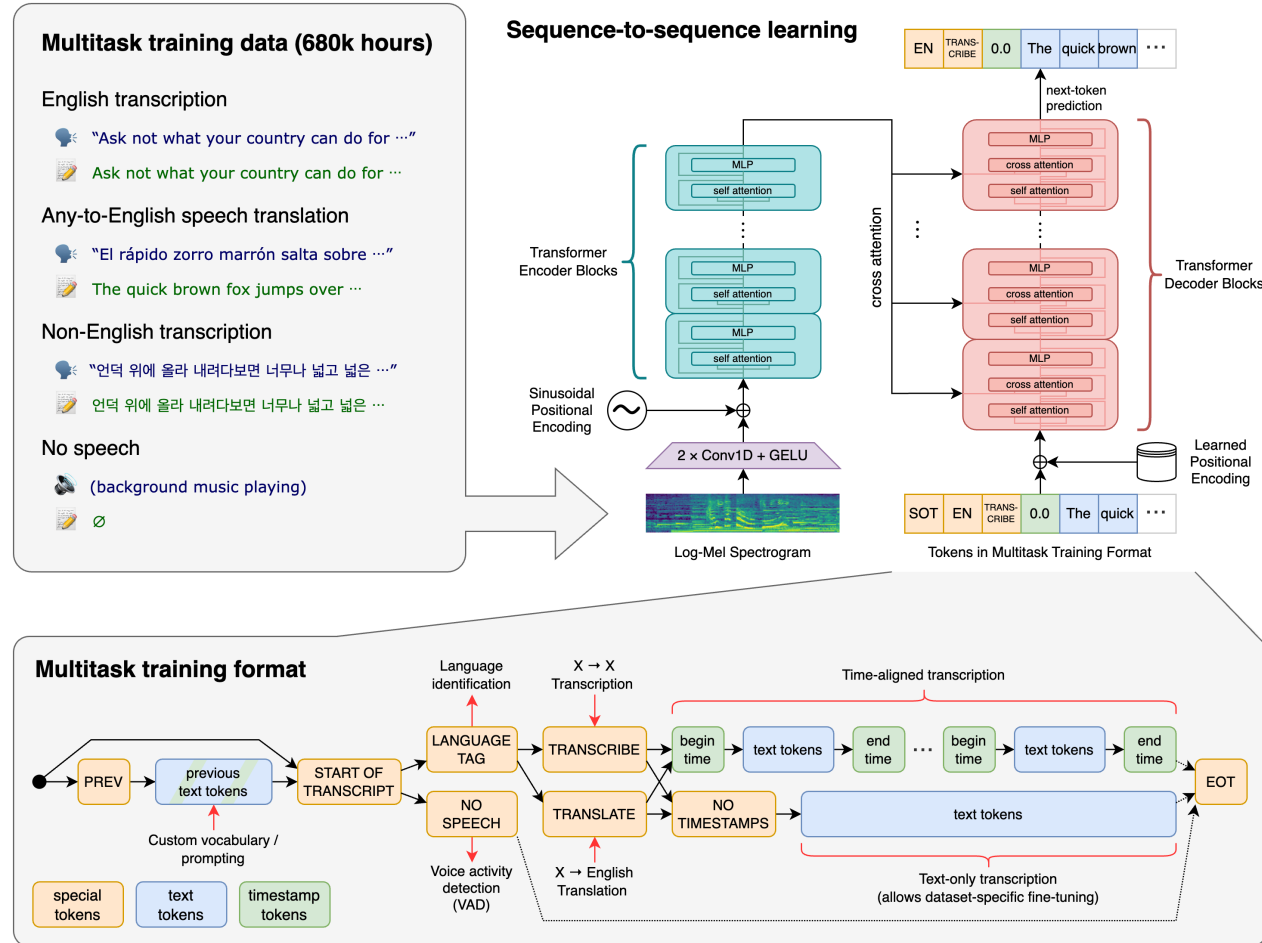
[olivier.mattelaer@uclouvain.be](mailto:olivier.mattelaer@uclouvain.be)

UClouvain \ SST \ CISM

**Objectif: Utiliser l'outil "Whisper" pour accélérer le travail de retranscription d'interviews**

# Whisper : points forts

- derniers développements en *Deep Learning* (robuste, efficace)
- travail en batch (plusieurs retranscriptions en même temps)
- open-source (gratuit)
- pré-entraîné (utilisable *as is*)



## Whisper points faibles ❌

- Pas parfait ("*All models are wrong, but some are useful*")
- Gourmand en ressources
- Pas tout simple à mettre en place

```
!whisper "Aya-Nakamura.mp4" --model large-v2 --language fr
[00:00.000 --> 00:20.000] J'entends des bails à trois sur moi, à ce qu'il paraît je te cours après, mais ça va pas mais t'es taré
[00:20.000 --> 00:25.000] Mais comment ça, le monde est typé, tu croyais quoi, qu'on se voit plus jamais
[00:25.000 --> 00:30.000] J'pourrais t'afficher mais c'est pas mon délire, d'après les rumeurs tu m'as eue dans ton lit
[00:30.000 --> 00:49.000] Oh Jah Jah, y'a pas moyen Jah Jah, j'suis pas ta catin Jah Jah, genre en cotonna baby tu dettes ça
[00:49.000 --> 00:54.000] Tu penses à moi, j'pense à faire de l'argent, j'suis pas ta daronne et j'te fais pas la morale
[00:54.000 --> 00:59.000] Tu parles sur moi, y'a air, crache encore, y'a air
[00:59.000 --> 01:04.000] Tu voulais m'avoir, tu savais pas comment faire, tu jouais un rôle, tu finiras aux enfers
[01:04.000 --> 01:09.000] T'façon n'a qu'amour, je l'ai couché, le jour où on se croise faut pas touffer
[01:09.000 --> 01:14.000] Tu jouais le grand frère pour me salir, tu cherches des problèmes sans faire exprès
[01:14.000 --> 01:19.000] Putain mais tu déconnes, c'est pas comme ça qu'on fait les choses
[01:19.000 --> 01:28.000] Putain mais tu déconnes, c'est pas comme ça qu'on fait les choses
[01:28.000 --> 01:48.000] Oh Jah Jah, y'a pas moyen Jah Jah, j'suis pas ta catin Jah Jah, genre en cotonna baby tu dettes ça
[01:48.000 --> 02:07.000] Oh Jah Jah, y'a pas moyen Jah Jah, j'suis pas ta catin Jah Jah, genre en cotonna baby tu dettes ça
[02:07.000 --> 02:27.000] Oh Jah Jah, y'a pas moyen Jah Jah, j'suis pas ta catin Jah Jah, genre en cotonna baby tu dettes ça
[02:27.000 --> 02:37.000] En cotonna baby tu dettes ça
[02:37.000 --> 02:44.000] Oh Jah Jah
```

Transcription de [Djadja, Aya Nakamura, 2018, 2min 55](#) sur Google Colab.

Mis à part quelques approximations (« à trois » au lieu de « atroces », « typé » au lieu de « tit-pe » ou Djadja qui donne « Jah Jah », etc.), Whisper s'en sort bien.



Comme c'est gourmand en ressources, il faut utiliser des outils comme les

**Clusters de calcul intensif et accélérateurs graphiques**

# Calcul intensif



100 €

1000 €

5 0000 €

500 000 €

$10^6$  flops

$10^9$  flops

$10^{10}$  flops

$10^{12}$  flops

kBytes

GBytes

TBytes

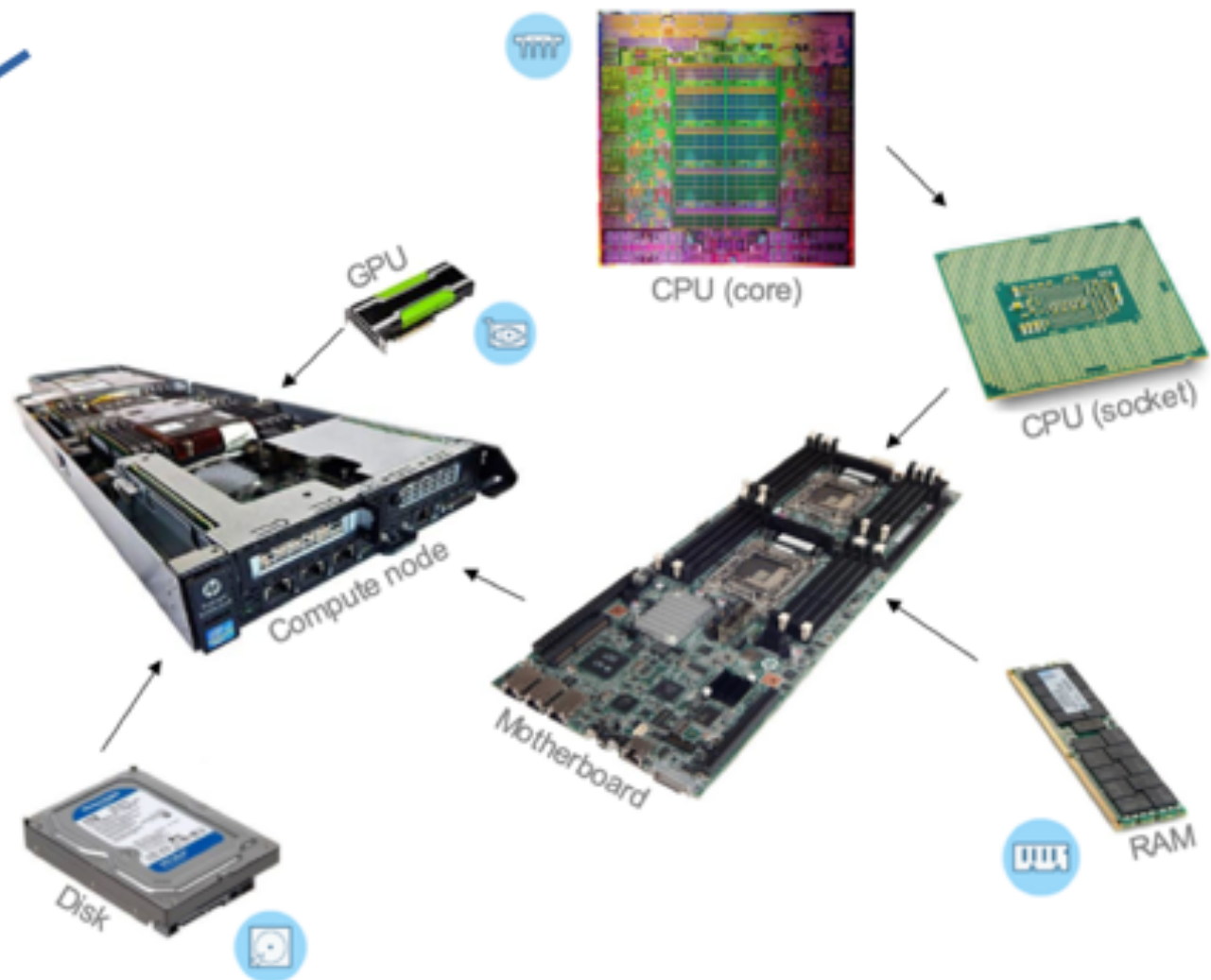
PBytes

Calc UI

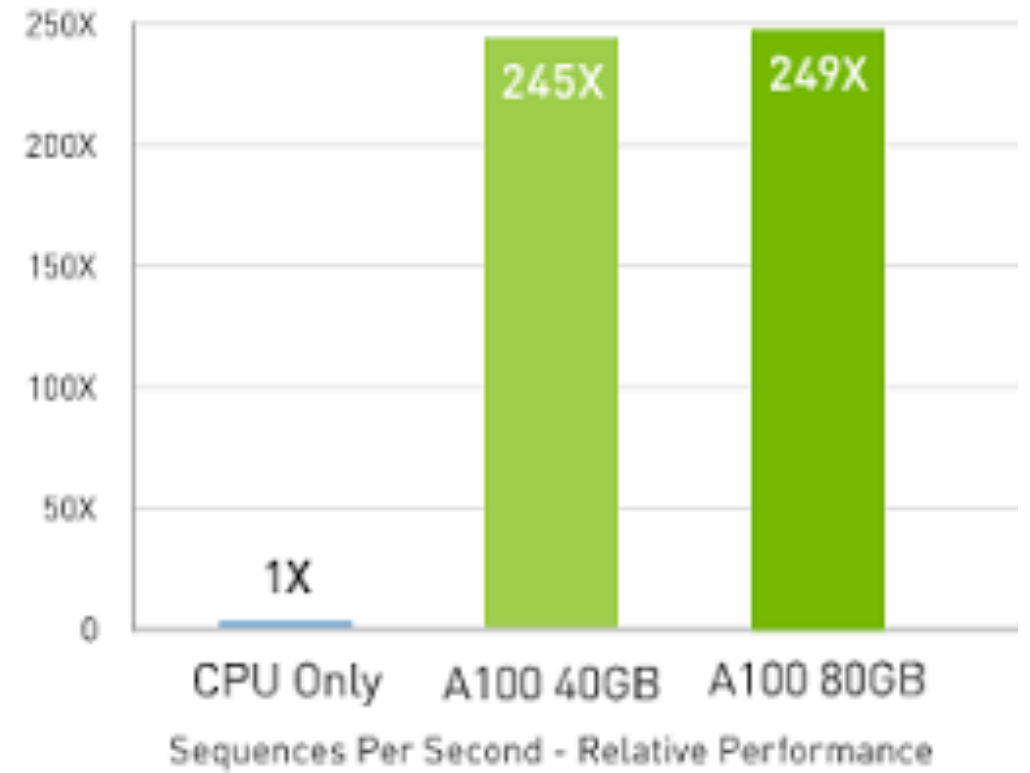
Excel, etc.

SAS, etc.

Fortran, Python, etc.

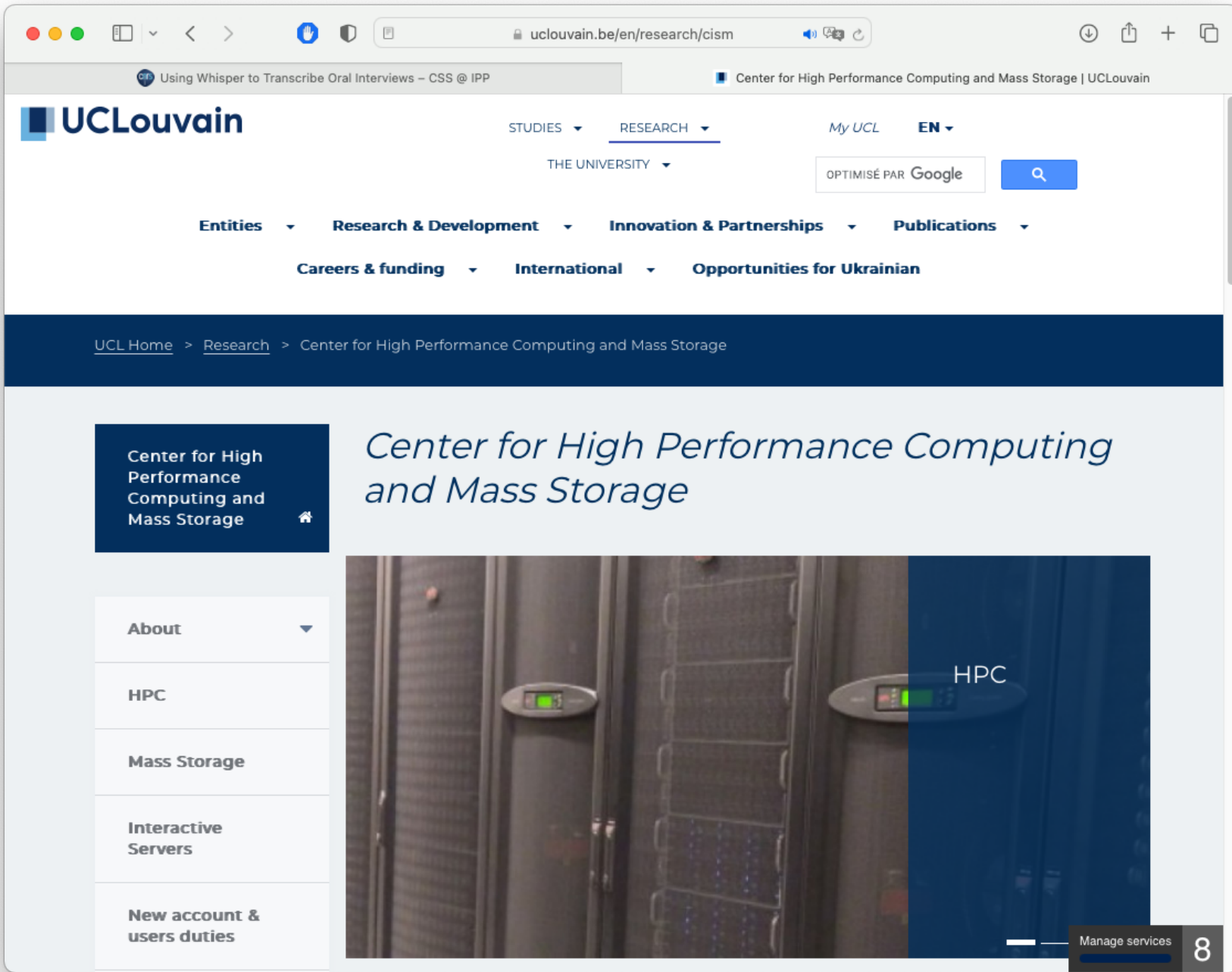


# Accélérateurs graphiques



Whisper n'est pas simple à mettre en place mais vous pouvez compter sur

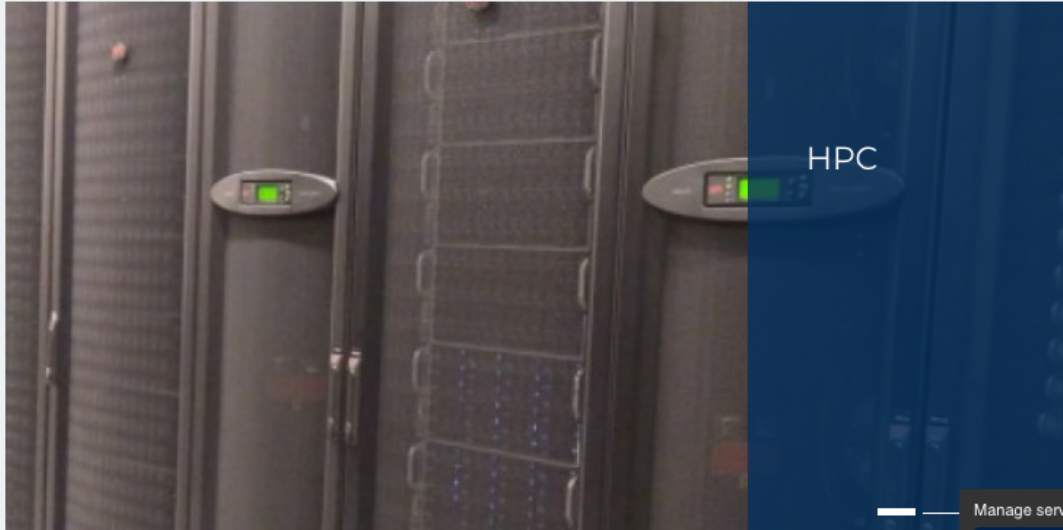
**La plate-forme technologique CISM**



Center for High Performance Computing and Mass Storage

# Center for High Performance Computing and Mass Storage

- About
- HPC
- Mass Storage
- Interactive Servers
- New account & users duties













# CISM

Quelques chiffres clés:

- 2 clusters de calcul intensif
- 1 cloud privé
- ~200TFlops de puissance cumulée
- ~1PB de stockage cumulé
- ~300 kW de consommation



# CISM



Quelques chiffres clés:

- ~20 heures de formation /an
- 50 à 100 personnes formées /an
- ~500 pages de doc, 100+ heures de vidéo
- ~150 tickets de support /an
- ~1000 installations de logiciel /an

# CISM

Quelques chiffres clés:

- ~70 publications d'utilisateurs / an
- ~20 groupes de recherche ~350 utilisateur·trice·s UCLouvain  
( 800+ en tout )

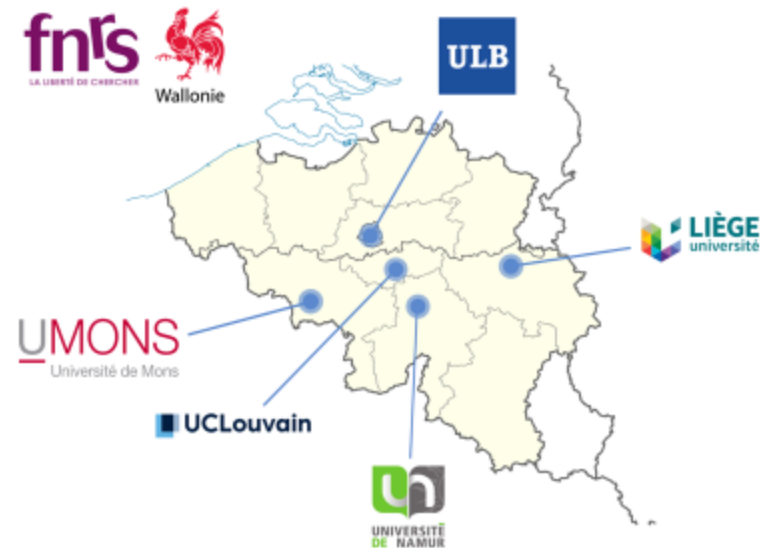






## About

CÉCI is the 'Consortium des Équipements de Calcul Intensif'; a consortium of high-performance computing centers of [UCLouvain](#), [ULB](#), [ULiège](#), [UMons](#), and [UNamur](#). The CÉCI is supported by the [F.R.S-FNRS](#) and the [Walloon Region](#). [Read more.](#)



## LUMI consortium

LUMI, one of the EuroHPC pre-exascale supercomputers, is located at CSC's data center in Kajaani, Finland. The supercomputer is hosted by the LUMI consortium including ten European countries.



# CISM

Quelques chiffres clés:

- 2 clusters
  - Lemaitre3 (CÉCI)
  - Manneback
    - Réservé UCLouvain
    - utilisation gratuite
    - 20+ GPUs

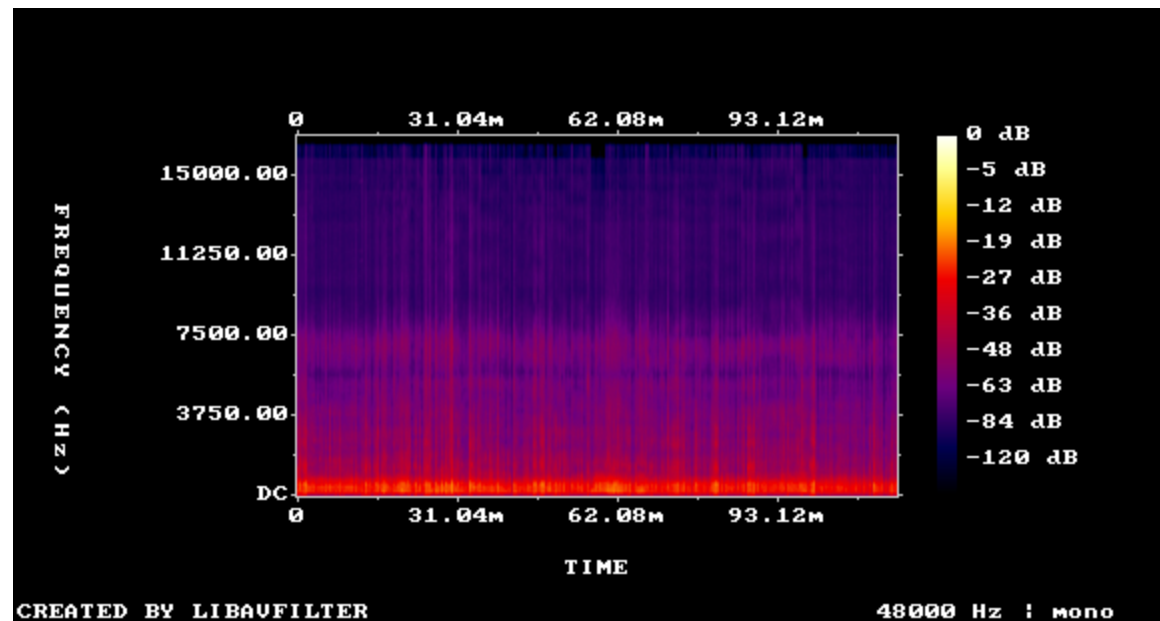


**Résultat: 2 heures d'interview retranscrites en 30 minutes**

# Enregistrement

Fichier audio `Julie.aac` (prénom d'emprunt):

```
$ ffprobe Julie/Julie.aac
Duration: 01:56:24.13, bitrate: 98 kb/s
  Stream #0:0: Audio: aac (LC), 48000 Hz, mono, fltp, 98 kb/s
```





# Retranscription

"small" (10 min de calcul)

| "large" (30 min de calcul)

00:11.120 --> 00:16.120

À moins que ce soit au volume boîte de nuit, normalement, ça devrait pas poser de problème sur le registrement.

00:16.120 --> 00:18.620

Si nous on s'entend normalement, il nous entend aussi.

00:18.620 --> 00:19.120

Ok.

00:19.120 --> 00:20.120

Donc euh...

00:20.120 --> 00:22.120

Put\*\*, j'ai pas un GSM comme ça moi.

00:22.120 --> 00:24.120

J'ai des jeux de trucs périmés.

00:25.120 --> 00:27.120

C'est un GSM de question, si.

00:27.120 --> 00:27.620

Ouais, ouais, ouais.

00:11.000 --> 00:13.000

À moins que ce soit au volume boîte de nuit, normalement ça ne devrait pas poser de problème sur l'enregistrement.

00:16.000 --> 00:17.000

On parle plus faux.

00:17.000 --> 00:19.000

Si nous on s'entend, normalement il nous entend aussi.

00:19.000 --> 00:20.000

Donc...

00:20.000 --> 00:22.000

Par pitié, j'ai pas un GSM comme ça moi.

00:22.000 --> 00:24.000

J'ai des vieux trucs périmés.

00:25.000 --> 00:27.000

C'est un GSM d'occasion aussi.

00:27.000 --> 00:28.000

Oui, oui, oui.

**Maintenant: 2 heures de formations pour faire la même chose**

## La bonne nouvelle:

pour vous faire la même chose, il ne faut pas

- devenir expert machine learning
- devenir développeur Python

## La "mauvaise" nouvelle

il faudra

- apprendre comment se connecter aux clusters ( `SSH` )
- apprendre Linux et la ligne de commande ( `bash` )
- apprendre comment "activer" des programmes pré-installés ( `lmod` )
- apprendre comment soumettre des jobs sur le cluster ( `slurm` )