

Moyens de calcul du Criann

Focus sur l'utilisation pour l'IA

Marie-Sophie Cabot & Benoist GASTON

Journée LLM de la Fédération Normastic - Caen - 28 Novembre 2024

Agenda

Présentation générale du Criann

- Contexte
- Moyens de calcul, modalités d'accès

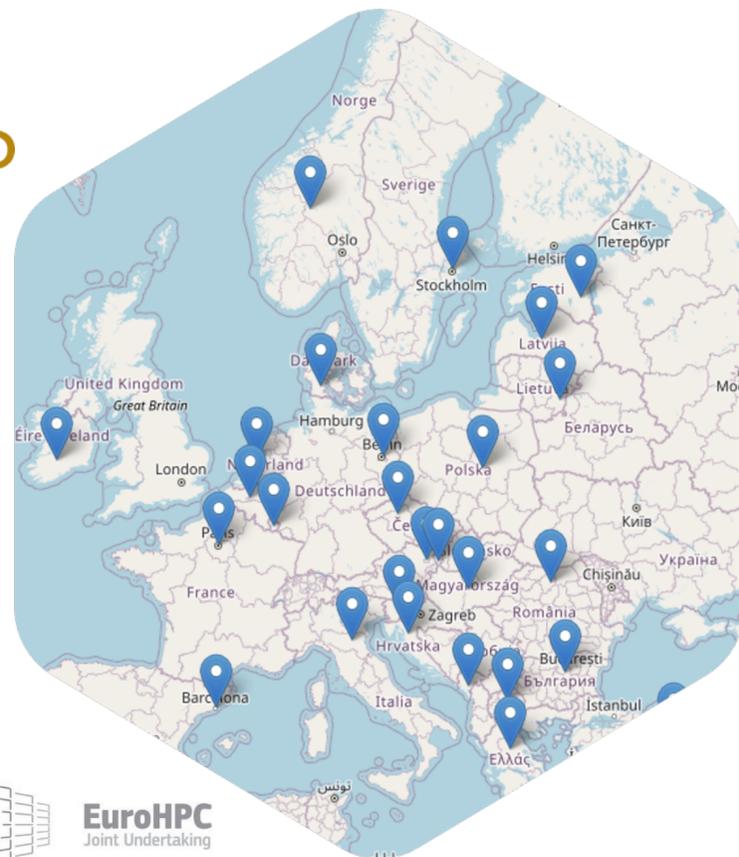
Focus IA sur Austral

- Démo Jupyter
- Apprentissage Distribué
 - Data Parallelism avec Pytorch - Optimisation avec ZeRO
 - Model Parallelism avec Pytorch

Contexte

EuroCC

Centre de compétence HPC



This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101101903. The JU receives support from the Digital Europe Programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia

Partenaires du Centre de compétence français :



Avec la participation de  le mésocentre des mésocentres

- Réseau de 33 centres de compétence nationaux
 - Favoriser l'usage du HPC et des technologies associées (HPDA, IA & Quantum)
 - Fédérer l'écosystème
 - Développer de formations
 - Accompagner les besoins et les demandes
- Programme d'accompagnement à l'usage du HPC
 - Porté par Criann & Romeo, partenaires MesoNET
 - A destination du secteur public et du secteur privé
 - Collectivités, administrations
 - Industrie, PME, startups

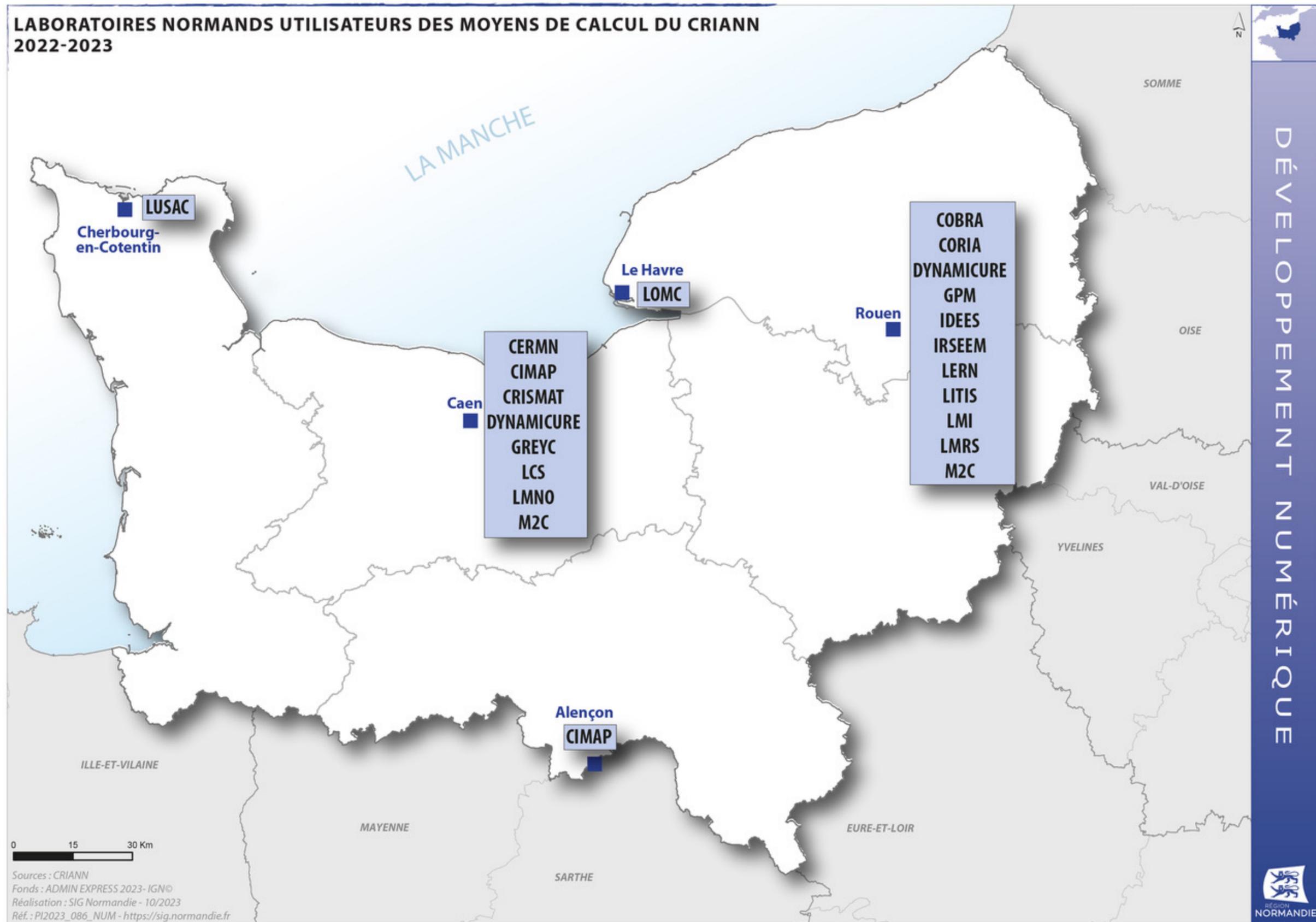


CRIANN

Présentation

- Structure créée en 1992
 - Établissements ESR (sur le périmètre de Normandie Université)
 - Rectorat, CHU et établissements de santé
 - Collectivités territoriales
- Mutualisation d'équipements et de services à haut niveau de performance
 - Réseau régional Syvik
 - Centre de données régional
 - Calcul intensif (HPC)
- Equipe : ~15 ETP
- Certification ISO27001 et HDS

LABORATOIRES NORMANDS UTILISATEURS DES MOYENS DE CALCUL DU CRIANN
2022-2023

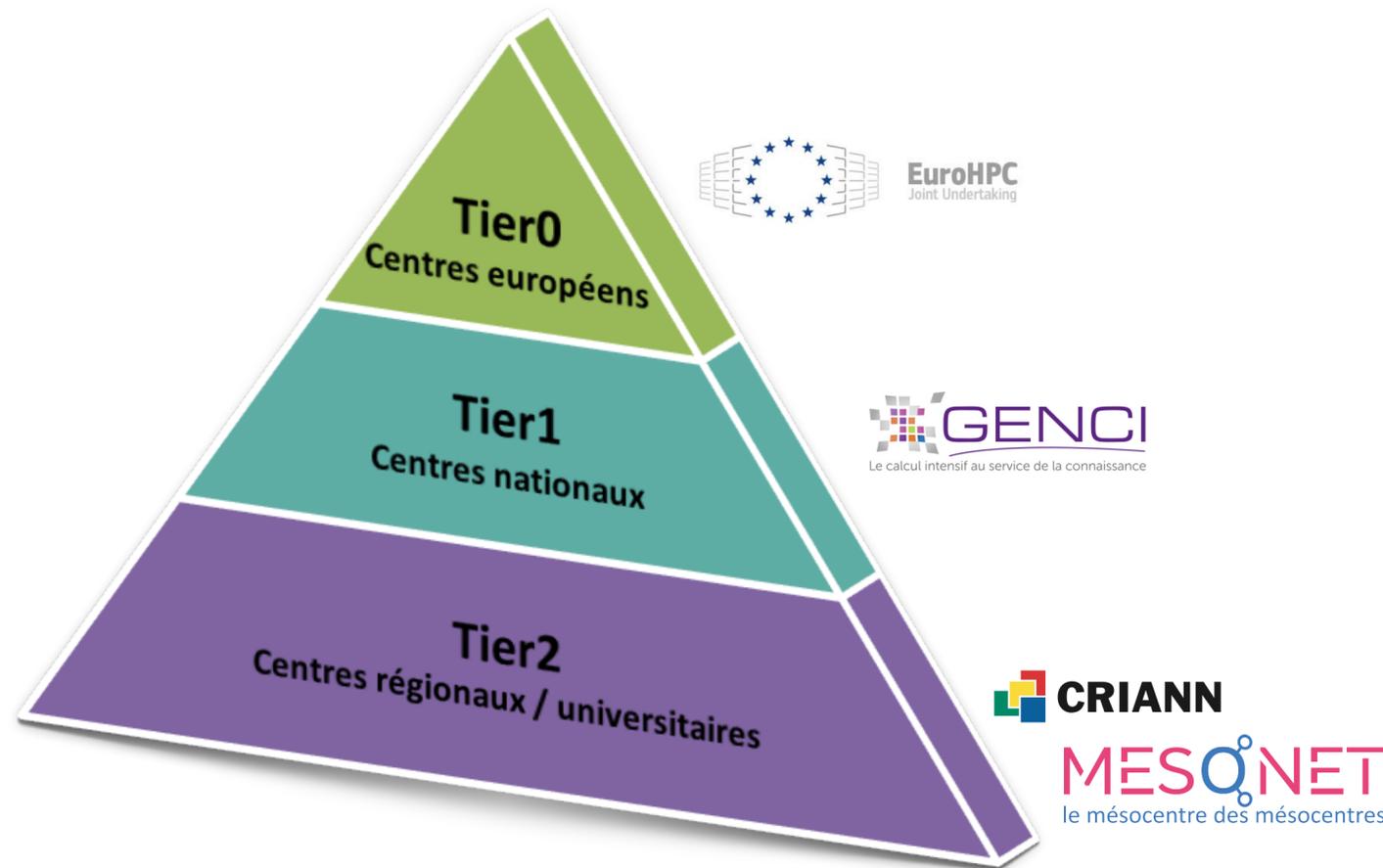


DÉVELOPPEMENT NUMÉRIQUE

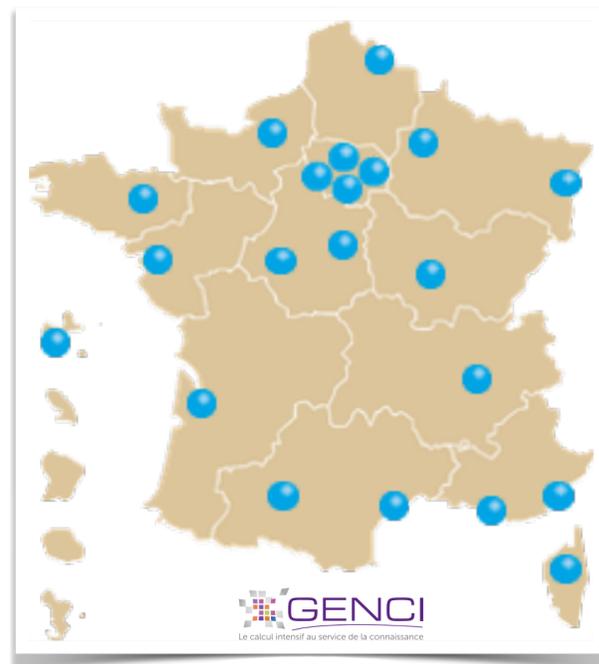


Service de calcul intensif

Dans l'écosystème du calcul intensif pour la Recherche



Organisation de moyens de calcul intensif pour la recherche



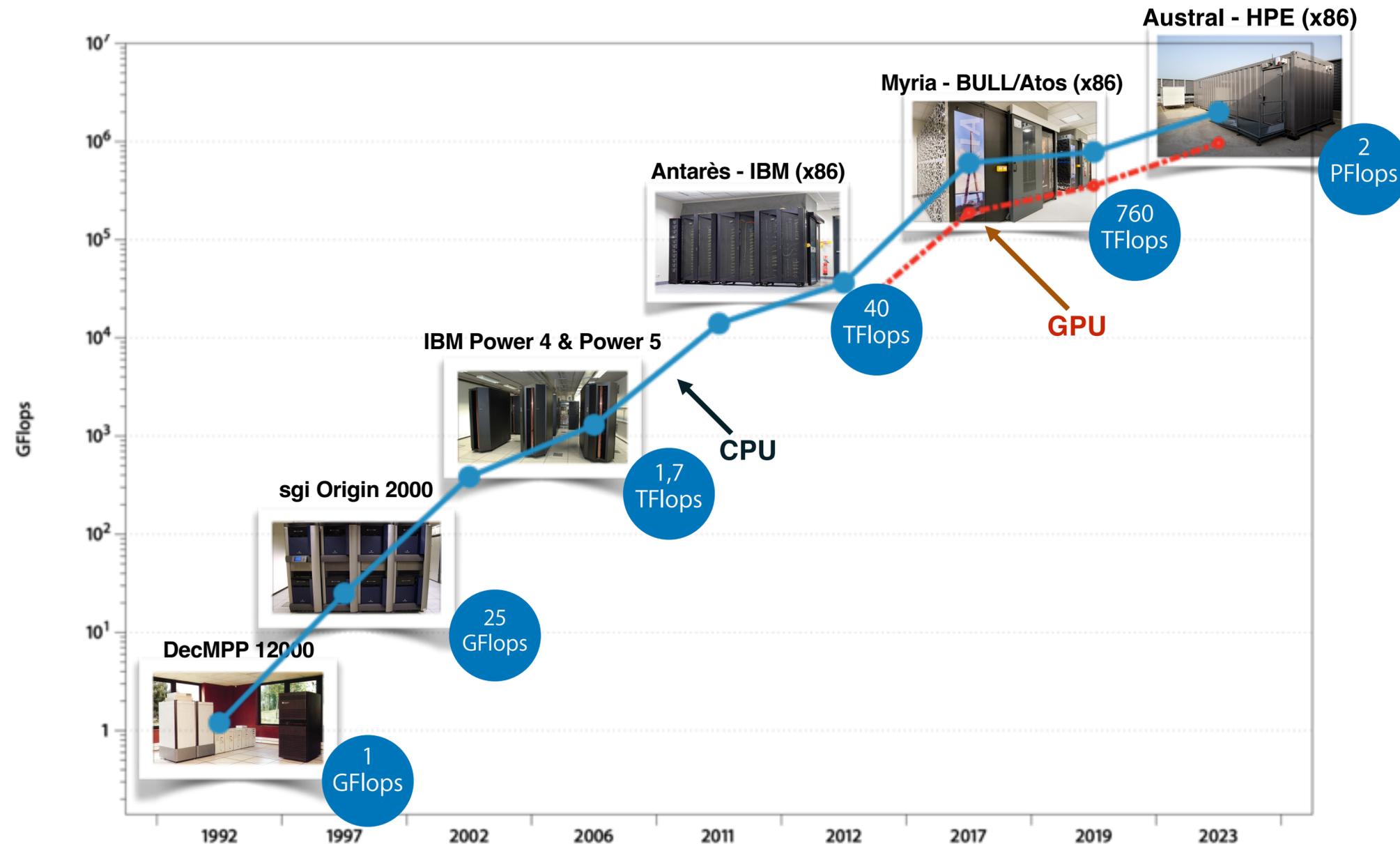
MESONET
le mésocentre des mésocentres

- Mésocentre à l'échelle régionale
 - Multi-disciplinaire / multi-laboratoires
 - Assure la **prise en main** du calcul intensif doctorants, nouvelles communautés scientifiques
- Partenaire de **MesoNET**
 - Equipex+[2021-2027] ANR pilotage Genci
 - Coordination de 21 centres de calcul universitaires régionaux
 - 8 centres (dont le Criann) hébergeurs d'une infrastructure distribuée (calcul, stockage)
 - Ouverture des ressources à la formation
 - Haut niveau de support
 - Objectif d'infrastructure de recherche

Moyens de calcul
Modalités d'accès
Services associés

CRIANN - Service de calcul intensif

Évolution des moyens de calcul du Criann



Calculateur Austral - Caractéristiques [oct 2023 - 2024]



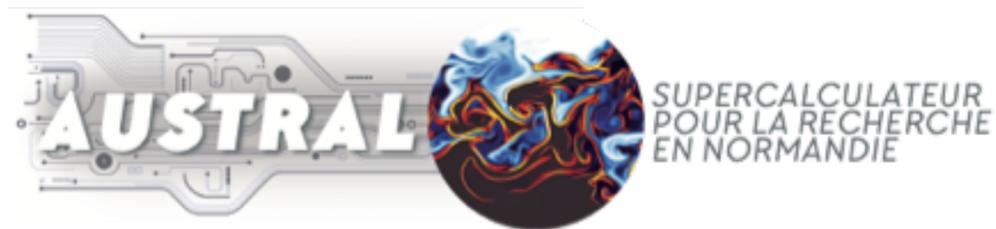
Nœuds de calcul

- 88 GPUs Nvidia A100 80 GB SXM4**
11 nœuds - Apollo 6500 Gen 10+ DLC nodes
- 23 808 cœurs@2.4 GHz - 95 TB RAM**
124 nœuds AMD Genoa - 2x96 cores - 768 GB RAM DDR5
- Cray XD 2000 Gen11 DLC
- Nœud SMP 6 TB de RAM DDR4**
Superdome Flex - 224 cœurs@2.6 GHz Intel Cooper Lake
- 8 GPUs AMD Mi210 (veille technologique)**
2 nœuds 4 x GPU - Apollo 6500 Gen 10+

Interconnexion Slingshot 200 Gbit/s	Stockage 2 Po (1 Po NVMe)
---	-------------------------------------

Accès à distance 100 Gbit/s
5 frontales de connexion
5 serveurs de visualisation
Environnements interactifs pour l'IA

RedHat - Slurm - Lustre



Austral - Extension 2025

Dans le cadre de MesoNET



A100-SMX—(88 devices sur Austral depuis 2023)	
Mémoire	80 GB HBM2e
Bandwith	2 TB/s
Interconnect	NVLink 600GB/s
Flops DP	9.7 TFlops / 19.5 Tflops TC
Flops SP	19.5 TFlops / 156 Tflops TC
Flops HP	312 Tflops with TC

H200-SMX (extension 16 devices sur Austral Q1.2025)	
Mémoire	141 GB HBM3e
Bandwith	4.8 TB/s
Interconnect	NVLink 900GB/s
Flops DP	34 TFlops / 67 Tflops TC
Flops SP	67 TFlops / 987 Tflops TC
Flops HP	1979 Tflops with TC

- Ajout d'un tranche GPU H200 à Austral
 - 2 nœuds x 8 GPU H200
 - Opérationnel courant Q1.2025
- Pour information : tranche 1 MesoNET au Criann
 - Architecture spécialisée vectorielle /veille technologique du projet

Calculateur Boreale (MesoNET Criann)
Puissance crête vectorielle 177 TFlop/s

Nœuds de calcul

Total 72 Vector Engines
 NEC SX Aurora Tsubasa 20b
 Par VE : 8 cœurs@1.6 GHz et 48 Go HBM2
 Par cœur : 64 registres vectoriels de 256 éléments DP

Sur 9 nœud Intel Xeon Icelake
 2 x 16 cœurs@2.9 GHz - 256 Go DDR4 - 960 Go NVMe

- Autres machines du projet, modalités d'accès sur www.mesonet.fr

Calculateur Austral

Modalités d'accès

- Tout compte utilisateur est ouvert dans un Projet
- Accès recherche : demande d'heures de calcul par Projet portée par le resp. de projet (membre permanent d'un laboratoire de recherche)
 - <https://gramc.criann.fr/>
- Accès au fil de l'eau pour les nouveaux projets
 - Renouvellement annuel (novembre), demandes d'extensions à mi-année
 - Justification des heures : rapport d'activité, publications, mention des moyens du Criann dans les publications
- Accès entreprises : contractualisation et paiement à l'usage

THÉMATIQUE SCIENTIFIQUE	NOM DU LOGICIEL
SIMULATION ATOMISTIQUE ET OUTILS CONNEXES	CHARMM
	GROMACS
	NAMD
	MOLPRO
	VASP
	PSI4
	DALTON
	AMF
	LAMMPS
	QCORE
	ASE
	ICMR-GAUSSIAN
	Quantum Espresso
	VMD
	BIOLOGIE
Guppy	
Dorado	
MÉCANIQUE DES FLUIDES	Star CCM+
	SWASH
	MODULEF
	DUALPHYSICS
	YADE
	TELEMAC-MASCARET
	OPENFOAM
	FOAM-EXTEND
	CODE_SATURNE
	MODÉLISATION ATMOSPHÉRIQUE, CLIMATOLOGIE, OCÉANOGRAPHIE
NCL	
WGRIB	
GEOS	
GDAL	
CDO	
R_TERRA	
NCO	
SIRANE	

THÉMATIQUE SCIENTIFIQUE	NOM DU LOGICIEL
MÉCANIQUE, ACOUSTIQUE	Code ASTER
	CAST3M
	HYPERWORKS
	LS-DYNA
	NASTRAN
	SALOME-MECA
	MATHÉMATIQUES, STATISTIQUES
OCTAVE	
SCILAB	
R	
Python/dask	
Python/pandas	
MACHINE LEARNING, DEEP LEARNING	
	TensorFlow/Keras
	Horovod
	Scikit-learn
	OpenCV
	MAILLAGES
Oasis	
COUPLEURS	Precice
VISUALISATION	Paraview
	Ferret
	Xmgrace
	Molden
	Ncview

Logiciels disponibles sur Austral
(mars 2024)

Service de calcul intensif

Support scientifique

- **Logithèque applicative**

- Installations à la demande, optimisées pour l'architecture
- Mais possibilité d'effectuer sa propre installation

- **Support avancé**

- Aide au portage et à l'optimisation
- Veille technologique

- **Formations**

- Prise en main jusqu'à utilisation avancée
- Techniques et outils du HPC (MPI, openMP, compilation, profilage, ...)

Agenda des formations

<https://indico.criann.fr/category/3/>

Formations calcul	Sessions 2024	Modalités	Durée	Formateur
<i>Austral usage généraliste</i>	<i>fev. et oct. 2024</i>	<i>Online</i>	<i>1/2 journée</i>	<i>PBM / Criann / MesoNET</i>
<i>Austral usage Deep Learning</i>	<i>mars et oct. 2024</i>	<i>Online</i>	<i>1/2 journée</i>	<i>BGn / Criann / CC-FR</i>
<i>Prise en main du calculateur vectoriel Boreale</i>	<i>Juillet 2024</i>	<i>mixte</i>	<i>1/2 journée</i>	<i>PBM / Criann / MesoNET</i>
<i>Programmation parallèle OpenMP</i>	<i>mars 2024</i>	<i>Face to face</i>	<i>1.5 jours</i>	<i>PBM / Criann / MesoNET</i>
<i>Programmation parallèle MPI</i>	<i>mai 2024</i>	<i>Face to face</i>	<i>1.5 jours</i>	<i>PBM / Criann / MesoNET</i>
<i>Python pour le HPC</i>	<i>juin 2024</i>	<i>Face to face</i>	<i>2 jours</i>	<i>BGn / Criann / CC-FR</i>
<i>Python pour le HPDA</i>	<i>12&13 dec. 24</i>	<i>Face to face</i>	<i>2 jours</i>	<i>BGn / Criann / CC-FR</i>
<i>Introduction au calcul quantique</i>	<i>9 & 10 oct. Rouen et Caen</i>	<i>Mixte</i>	<i>1/2 journée</i>	<i>O. Hess Eviden via MesoNET</i>
<i>MLDE user Workshop</i>	<i>26 & 27 nov. 2024</i>	<i>Online</i>	<i>2 x 2heures</i>	<i>Expert HPE</i>
<i>CPE Cray Programming Environment</i>	<i>14 & 15 janv. 25</i>	<i>Face to Face</i>	<i>2 jours</i>	<i>Expert HPE</i>
<i>Algorithmes quantiques pour l'optimisation combinatoire</i>	<i>9 janvier 2025</i>	<i>Face to face</i>	<i>1 jour</i>	<i>Expert Eviden via MesoNET</i>

Formations récurrentes

Focus Agenda 2025

Formation etc.

- Possibilités de sessions sur le calcul quantique
 - Financement MesoNET, formateur expert Eviden
 - Intérêt à Caen de sessions sur la programmation quantique ?
 - Basics (2 jours) Advanced (3 jours)
- Possibilité de Boot Camp avec Nvidia
 - AI for science envisagé
 - Autres sujets d'intérêt ?
- Journée scientifique des utilisateurs du Criann en juin
 - Contributions des utilisateurs bienvenues

Démo Jupyter

Apprentissage Distribué

Apprentissage Distribué

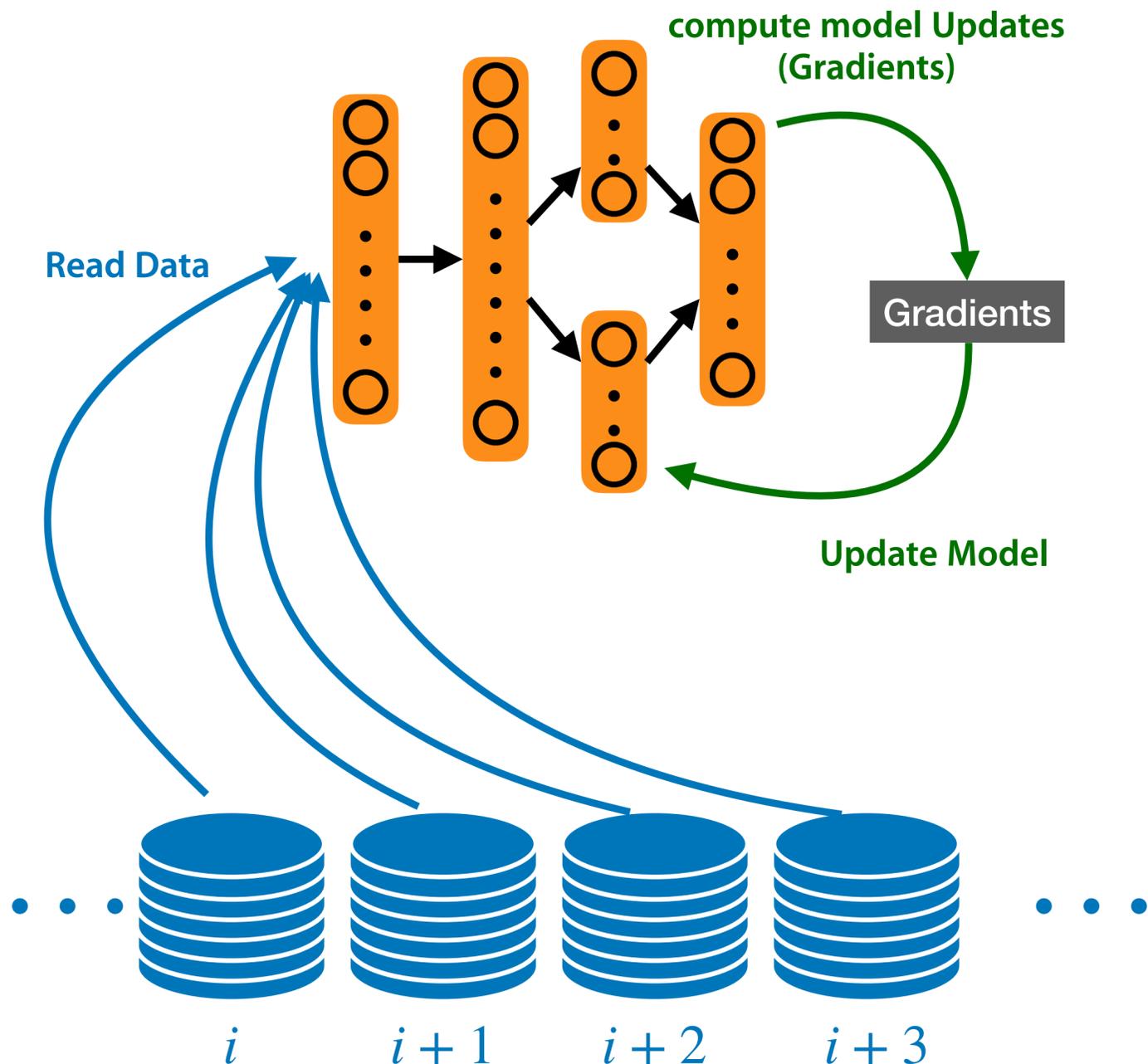
Entraînement de modèle d'IA du point de vue HPC

- Problèmes majeurs

1. Modèle, Optimizer, Gradients : forte consommation mémoire GPUs => goulot d'étranglement
2. Temps d'entraînement d'un modèle très long
Généralement résolu avec le *Data Parallelism (DP)*
3. Modèle trop volumineux pour tenir sur une unité de calcul (typiquement LLM)
Généralement résolu avec la distribution de modèle *aka Model Parallelism (MP)*

- Au Criann

- Data parallelism : régulièrement utilisé par quelques utilisateurs depuis plusieurs années
- Model Parallelism : des demandes occasionnelles mais de plus en plus fréquentes

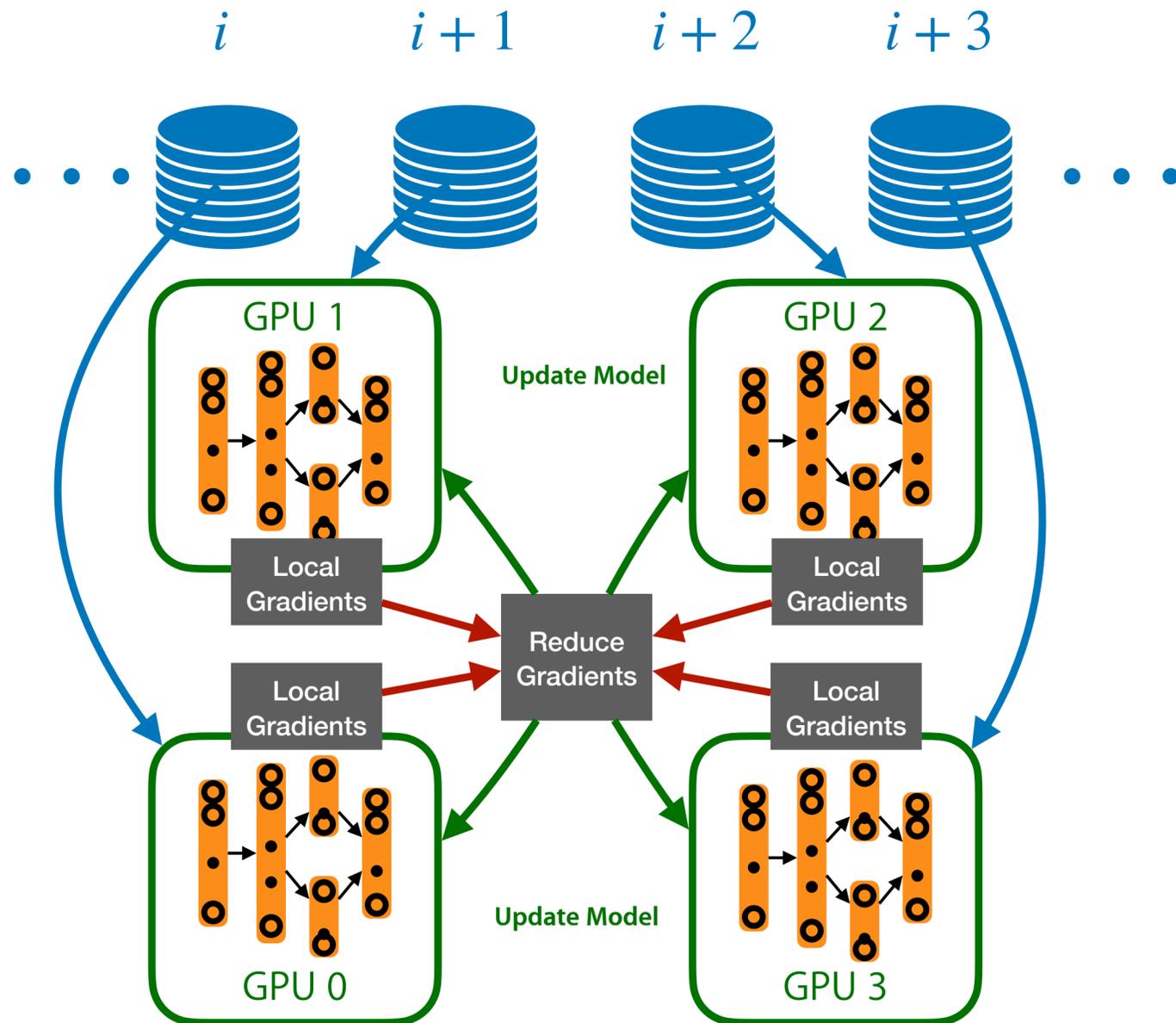


Data Parallelism

Data Parallelism (DP)

Principes

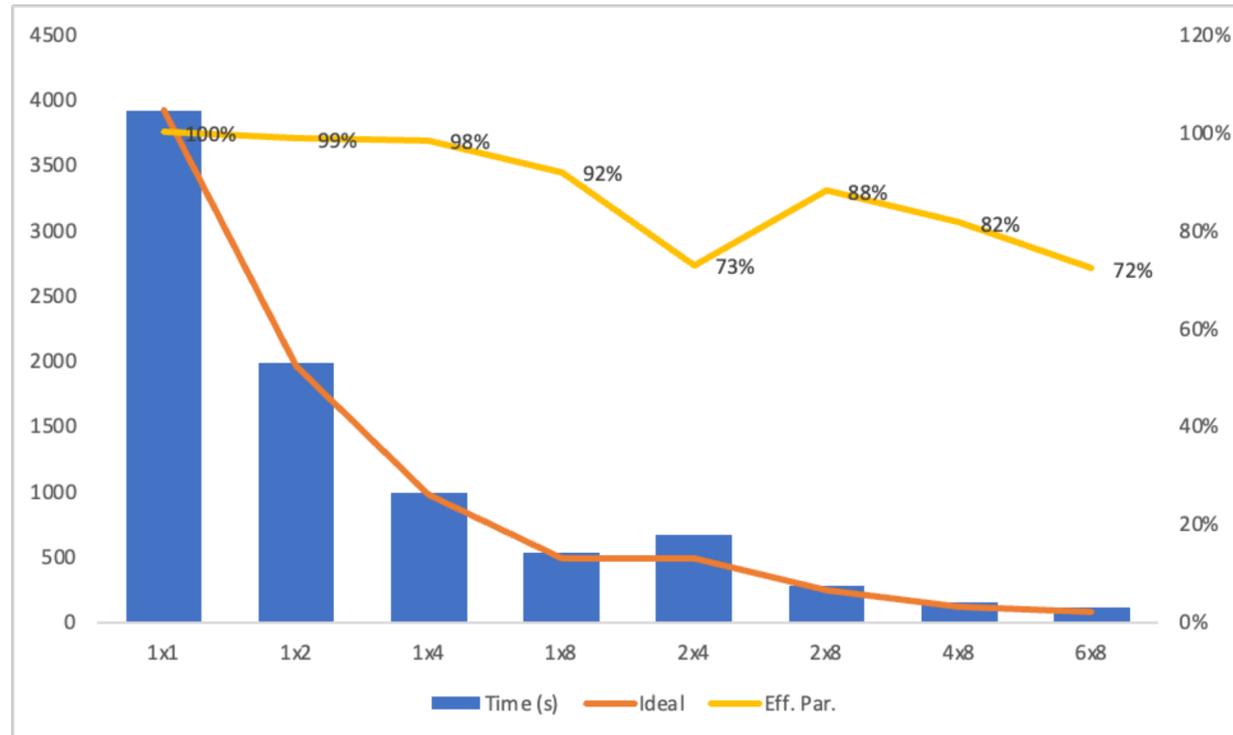
- Traitement des *batches* données en parallèle
 - Distribution des données
 - Réplication du modèle sur chaque ressource
 - Calcul des gradients locaux
 - Agrégation des gradients à l'aide d'une réduction
- Avantage
 - Bonne efficacité de calcul => traitement de gros volumes de données
- Inconvénient
 - Mauvaise efficacité mémoire => taille des modèles limitée



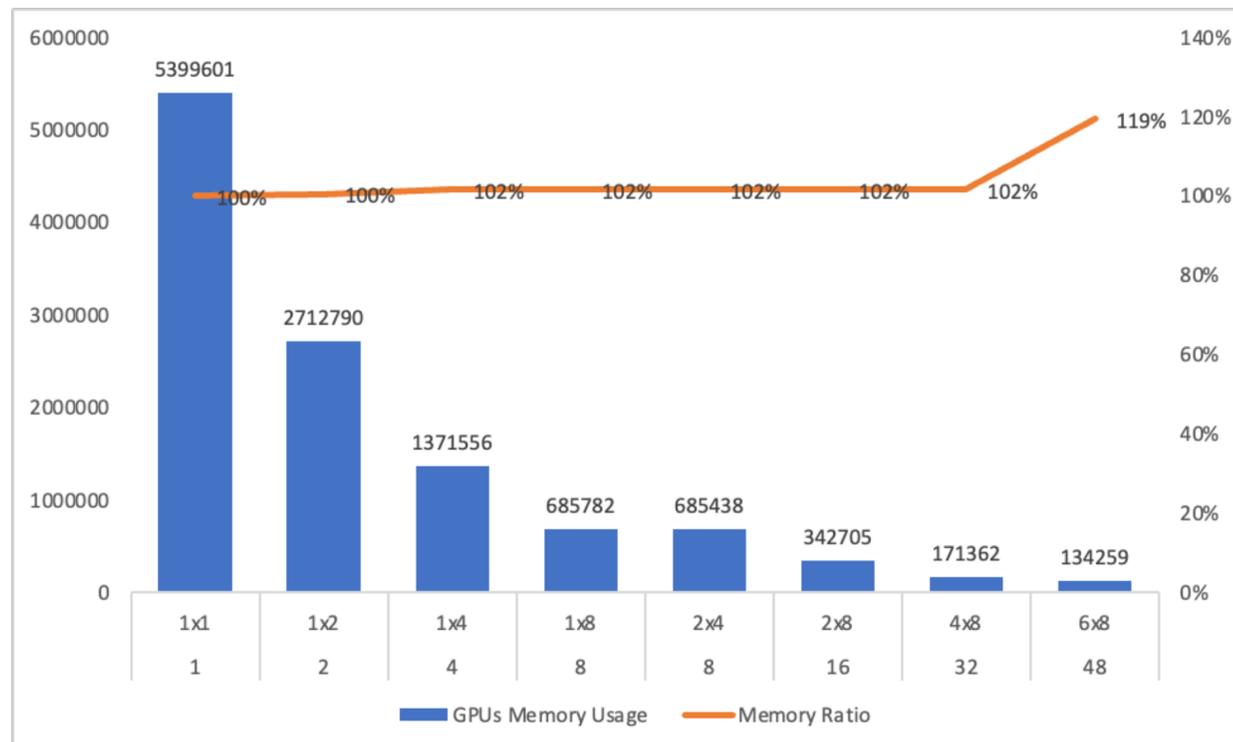
Data Parallelism (DP)

Tests et résultats

- Framework pytorch avec solution intégrée *DistributedDataParallel (DDP)*
 - Approche multi-gpus et multi-nœuds
 - 1 -> 48 GPUs A100
 - Intégration de SLURM (sous forme d'un package python)
 - Modèle gpu-2-XL Community
 - 1,5 Milliard de paramètres
 - Dataset
 - 10 000 items de Wikitext Graphcore (extrait des articles vérifiés de Wikipedia)



Temps de restitution et efficacité parallèle en fonction du nombre de GPUs

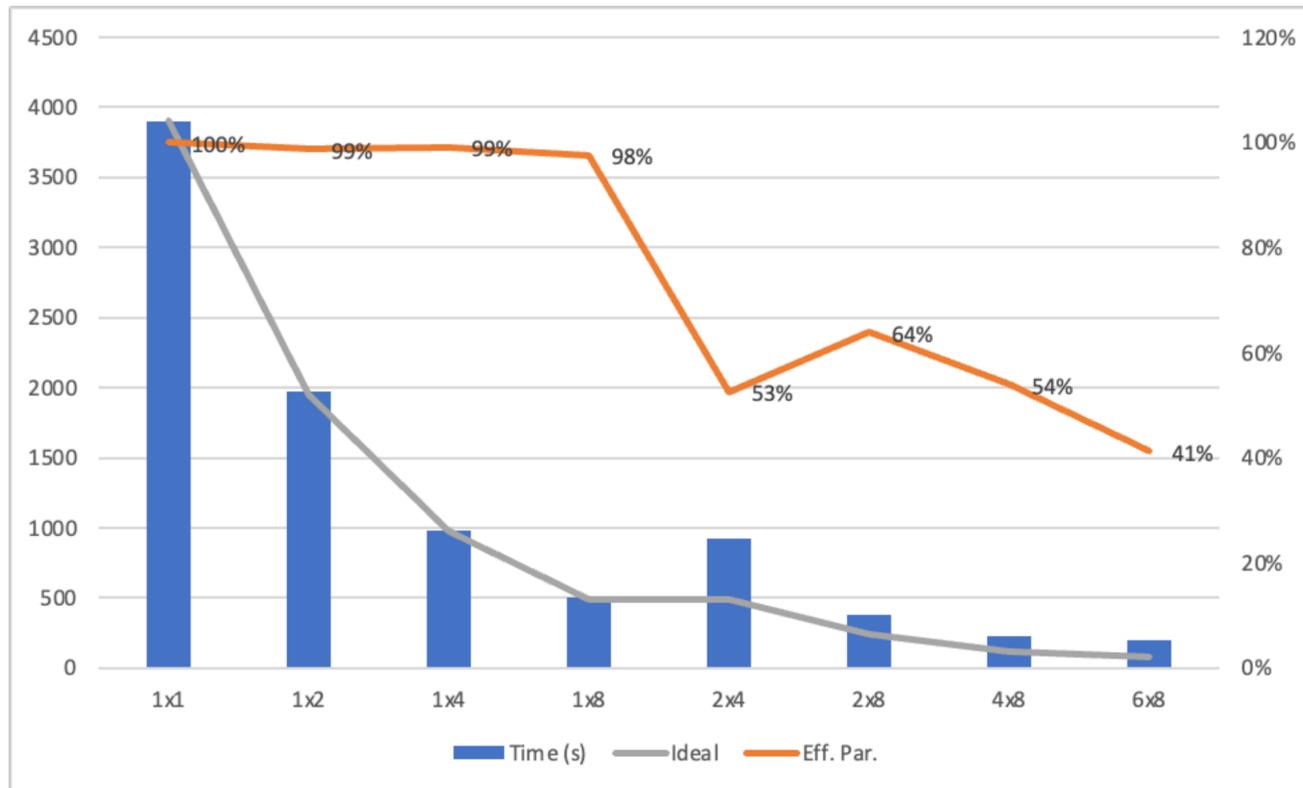


Consommation mémoire en fonction du nombre de GPUs

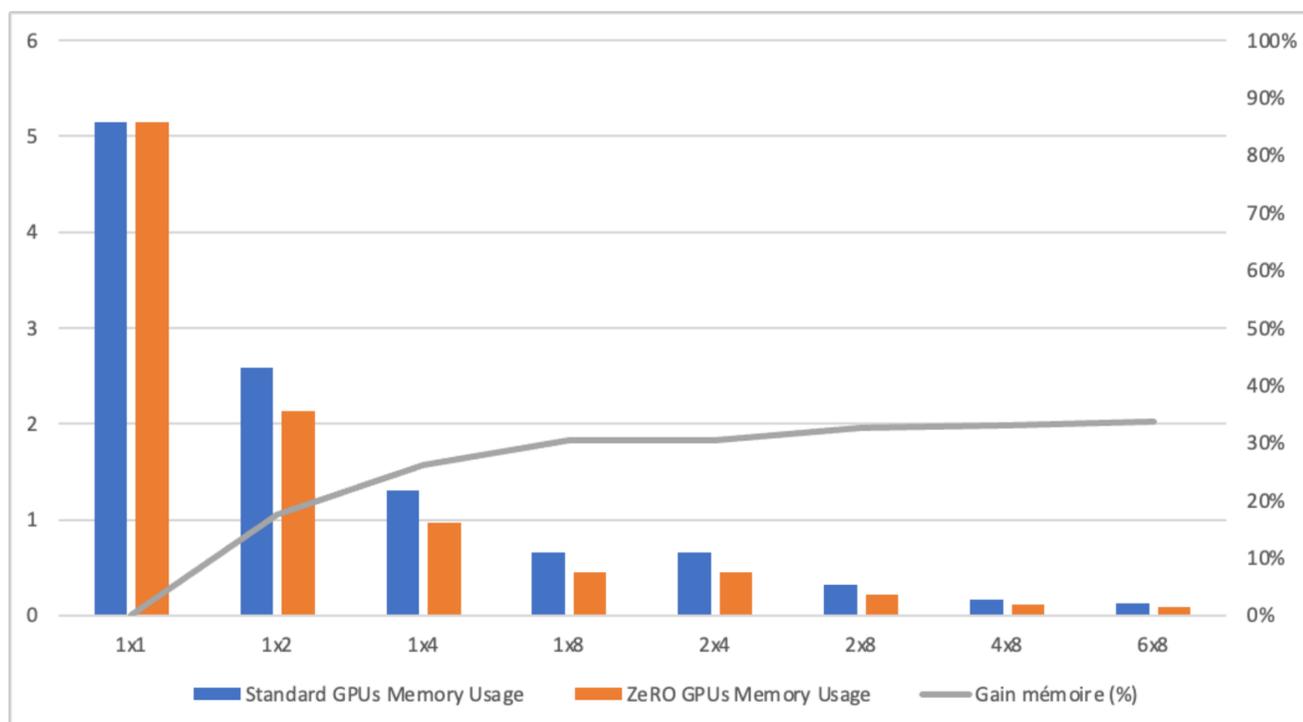
Data Parallelism optimisation

ZeRO (Zero Redundancy Optimizer)

- Optimiser l'entraînement de modèles massifs en machine learning
- Bibliothèque DeepSpeed (Microsoft)
 - Réduction de la consommation mémoire
 - Partitionnement / Distribution de l'*optimizer* et des gradients
- Tests et résultats
 - Comparaison avec le *Data Parallelism*



Temps de restitution et efficacité parallèle en fonction du nombre de GPUs



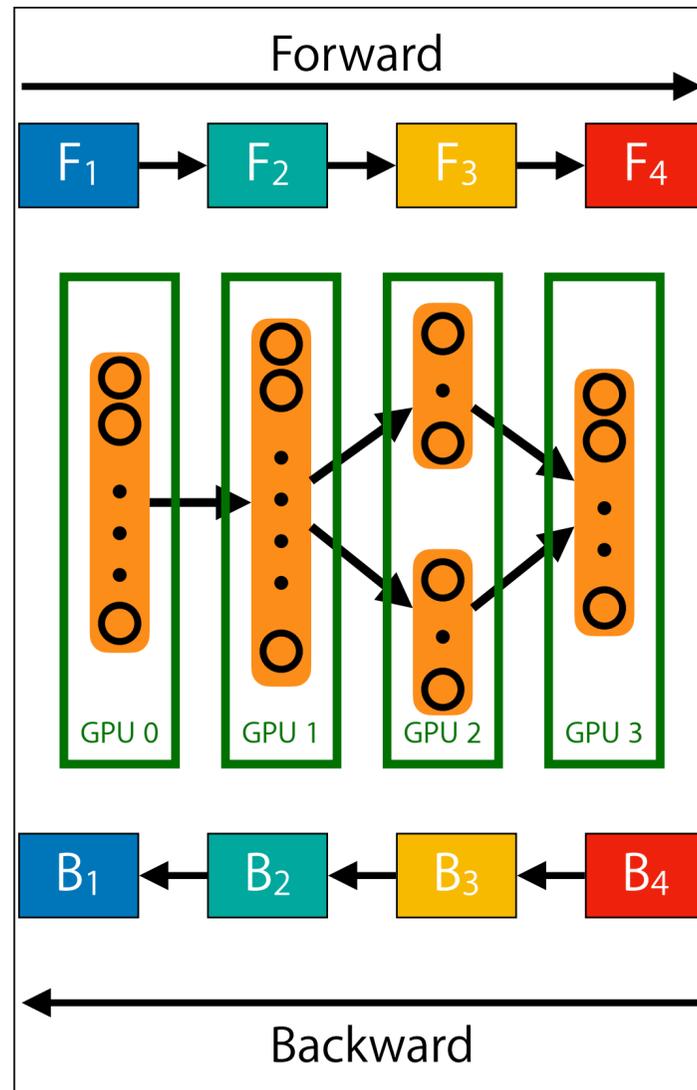
Consommation mémoire / méthode standard

Model Parallelism

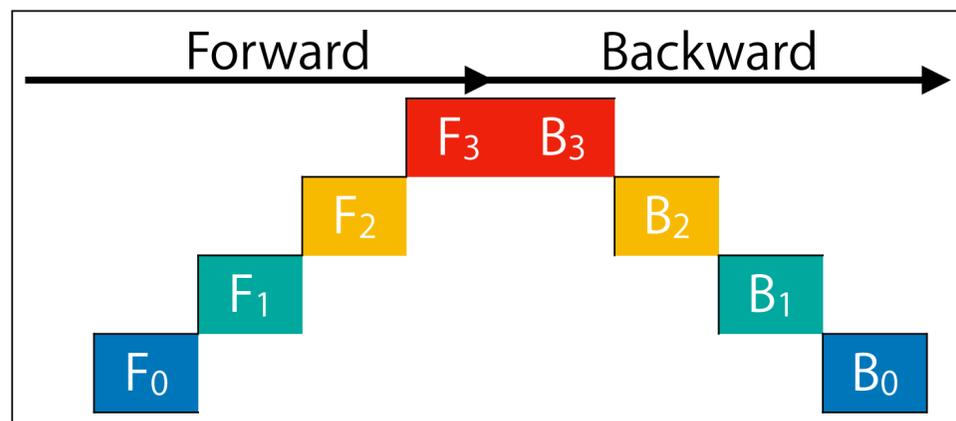
Model Parallelism (MP)

Principe

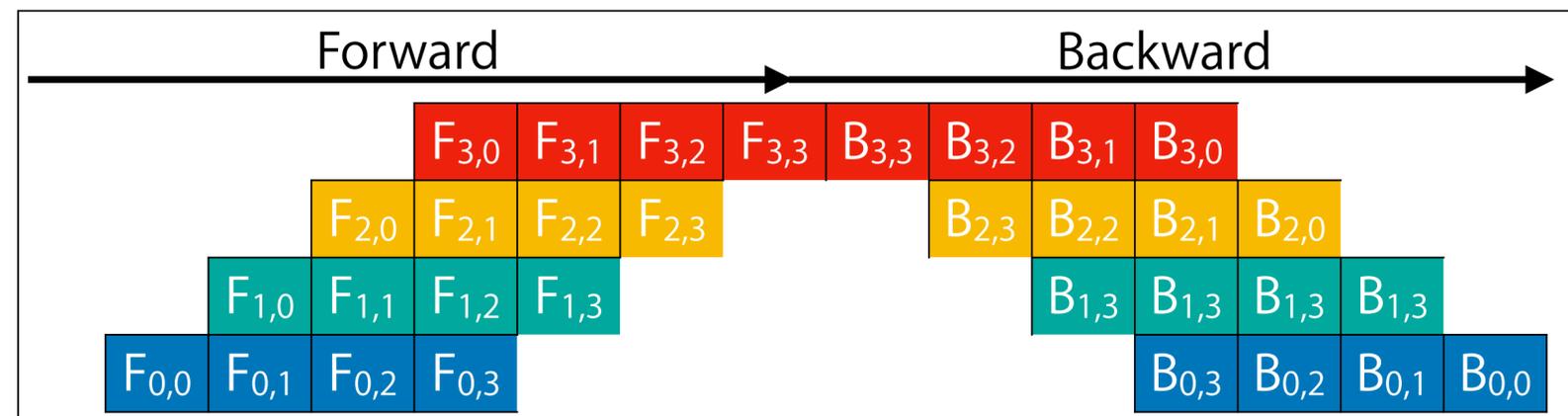
- Distribution d'un modèle sur plusieurs gpus
 - Traiter des modèles plus gros
 - Généralement découpage par couche
- Parallélisme par *pipelining* :
 - Division d'un *batch* de données en différent *chunks*
- Avantage
 - Bonne efficacité mémoire => traitement de très gros modèles
- Inconvénient
 - Faible parallélisation



Distribution du modèle sur plusieurs GPUs



Pipelining

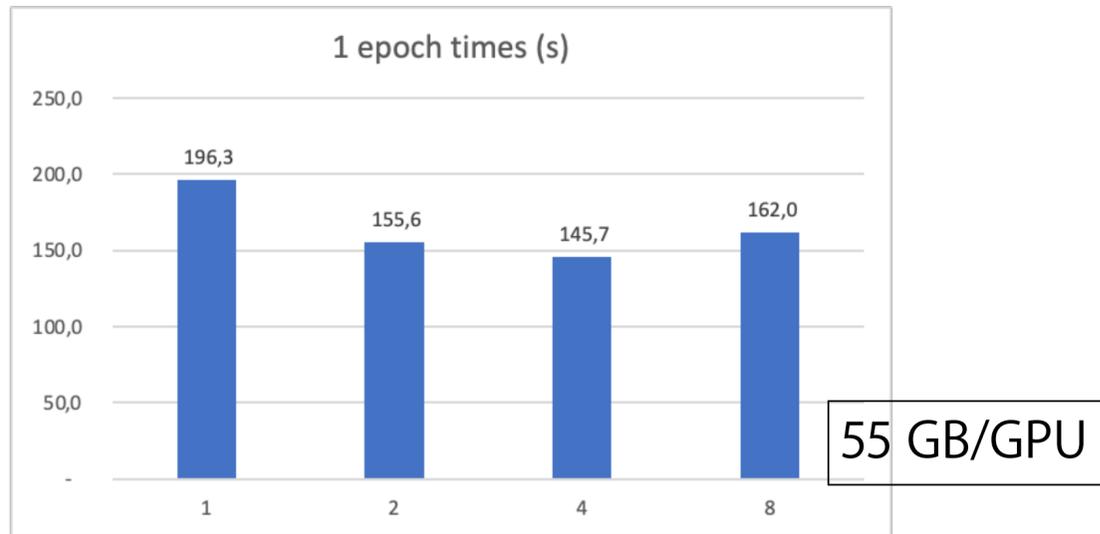


Pipeline Parallèle

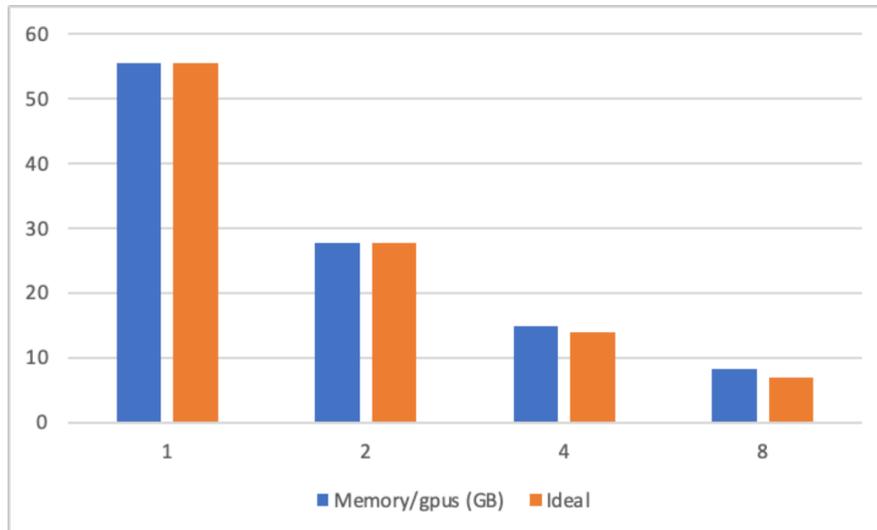
Model Parallelism

Tests et résultats

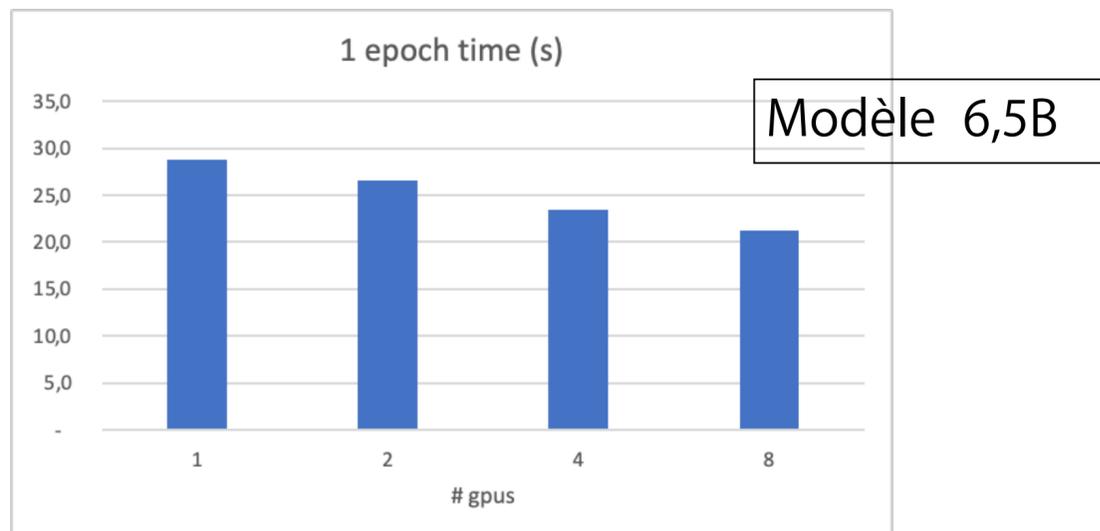
- Framework pytorch (+Pipe pour le *pipelining*)
- Intranœud uniquement
 - 1, 2, 4, 8 GPUs A100
- Modèle-jouet avec architecture *transformers* de taille variable (nombre de couches)
 - 6.5, 13, 26, 52 milliards de paramètres
- Dataset
 - 2096 items de *Wikitext Graphcore* (extrait des articles vérifiés de Wikipedia)



Temps par epoch en fonction du chunk : 8 gpus 52B paramètres



Empreinte mémoire par gpu - 6.5B paramètres



Temps par epoch en fonction du nombre de GPUs - 6.5B paramètres

Questions ?

Le plateau de calcul intensif du Criann est cofinancé par la Région Normandie, l'État français et l'Union européenne (Fonds Feder).
MesoNET bénéficie d'un financement de l'Agence nationale de la recherche au titre des Investissements d'avenir.
Le Centre de Compétence EuroCC français est cofinancé par l'Union européenne et par l'État français.
Le réseau régional Syvik est cofinancé par la Région Normandie et par l'Union européenne (fonds Feder).
Le fonctionnement du Criann bénéficie du soutien de la Région Normandie.



Centre Régional Informatique et d'Applications Numériques de Normandie
www.criann.fr