

## Accueil des thématiques du traitement de données, IA et Deep-Learning au Criann

Bilan de la période juillet 2018 - août 2019

Ce document présente un bilan synthétique des activités de l'année écoulée en lien avec la thématique scientifique du traitement de données – Intelligence Artificielle et en particulier réseaux de neurones profonds (Deep Learning).

Ce document fait suite au bilan rédigé mi-2018 et au rapport d'activités du projet Feder MNSN de février 2019.

### Sommaire

1.	Bilan de l'utilisation de Myria en IA & DL (août 2019) .....	2
2.	L'environnement de Myria pour le DL – actualisation 2019.....	3
2.1.	Nouveaux matériels intégrés au calculateur (2019) .....	3
2.2.	Logithèque de Myria pour l'IA (août 2019) .....	4
2.3.	Environnements interactifs pour le Deep Learning .....	4
2.4.	Files d'attente pour les travaux GPU, partage des ressources .....	5
1.1.	Support aux utilisateurs, formation des nouveaux utilisateurs .....	5
2.	L'environnement de test (« Deca »).....	5
3.	Travaux de veille technologique .....	6
4.	Autres actions et évolutions internes .....	6
4.1.	Évolution dans l'administration de la logithèque .....	6
5.	Datalab Normandie .....	6
6.	Traitement de données sur Myria.....	7
6.1.	Logithèque de Myria pour les traitements statistiques (août 2019) .....	7

#### CRIANN

**Centre Régional Informatique et d'Applications Numériques de Normandie**  
 Technopôle du Madrillet – 745 avenue de l'Université – 76800 SAINT-ÉTIENNE-DU-ROUVRAY  
 Tél. : 02 32 91 42 91 – admin@criann.fr – www.criann.fr  
 SIRET N° 383 599 990 00025 – CODE APE 7219Z

## 1. Bilan de l'utilisation de Myria en IA & DL (août 2019)

Le précédent rapport, mi-2018, faisait état de 5 projets scientifiques de chercheurs en Deep Learning sur le calculateur Myria, ainsi que d'un projet porté par le Criann et visant à exploiter les processeurs KNL de Myria. On comptait alors 10 utilisateurs actifs hors Criann.

L'évolution de l'usage des ressources du Criann par les chercheurs de cette thématique est présenté sur les figures 1 et 2.

Le nombre de projets (figure 1) continue d'augmenter, avec quatre nouveaux projets depuis le précédent bilan. Deux d'entre eux ne font pas appel aux techniques du Deep Learning, mais concernent le traitement de données : l'un en statistiques et l'autre en théorie des graphes. Ces projets sont regroupés sous le vocable « IA » dans les figures ci-dessous. Il est à noter que l'un des deux nouveaux projets en Deep Learning (pour du traitement d'images médicales) n'a pas encore consommé d'heures.

Le nombre d'utilisateurs actifs (ayant consommé des heures) a augmenté de façon très significative, puisque l'on compte maintenant 22 utilisateurs.

Le nombre d'heures consommées par ces projets (figure 2) est également en forte augmentation et atteint 250 000 pour les 8 premiers mois de 2019 (contre 50 000 en 2017 et 130 000 en 2018).

Précision :

- Les outils de Deep Learning sont aussi utilisés par un projet du laboratoire Coria ; celui-ci n'apparaît pas dans les éléments présentés ici (il est difficile de faire ressortir la part des heures « Deep Learning » de ce projet, par ailleurs fort consommateur d'heures de calcul sur des processeurs « classiques ») ;
- Quelques autres projets issus des laboratoires de mathématiques ne sont pas comptabilisés non plus ; il s'agit de projets utilisateurs de longue date du Criann, qui concernent la résolution d'équations et non le traitement de données.

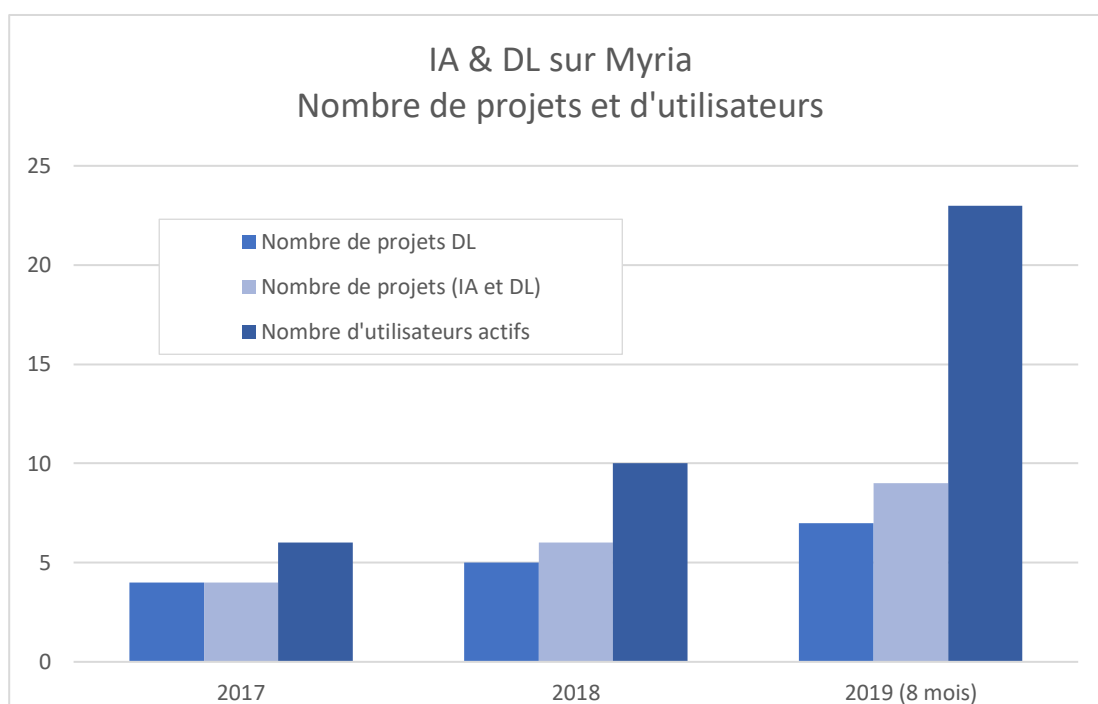


Figure 1 : Évolution du nombre de projets et d'utilisateurs actifs en IA – DL sur Myria

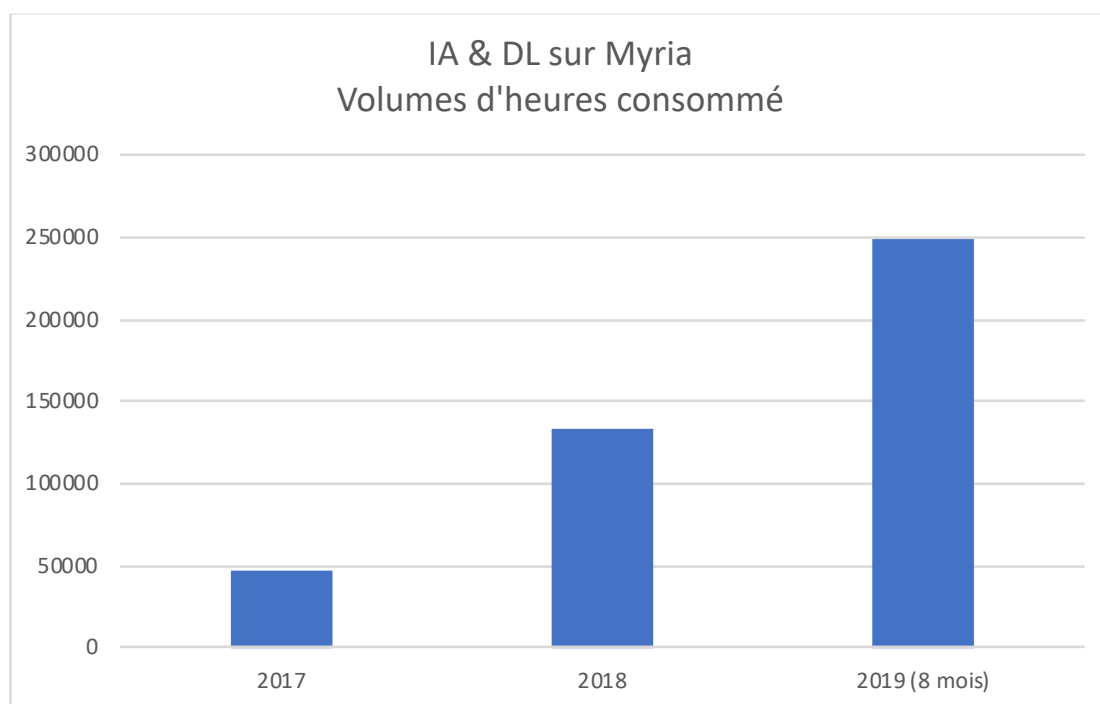


Figure 2 : Évolution du volume d'heures consommé en IA – DL sur Myria

## 2. L'environnement de Myria pour le DL – actualisation 2019

### 2.1. Nouveaux matériels intégrés au calculateur (2019)

Dans le cadre du programme RIN - plateformes de recherche porté par la Région Normandie et cofinancé par l'Union Européenne (Feder), un financement a été obtenu pour intégrer des GPU de dernière génération au calculateur Myria.

Cette extension est composée de cinq serveurs dotés chacun des composants suivants :

- CPU : 32 cœurs à 2,1 Ghz et 192 Go de RAM DDR4 (bi-socket Skylake G-6130) ;
- GPU : 4 cartes Tesla V100 SXM2 à 32 Go de mémoire GDDR5 ;
- réseau d'interconnexion NVLink 300 Go/s entre les cartes GPU.

La puissance crête théorique totale de cette extension est de 166 TFlop/s en double précision (dont 10 TFlop/s CPU).

Les gains attendus pour la communauté du Deep Learning sont conséquents : par rapport aux cartes P100, le constructeur annonce un gain de l'ordre d'un facteur 3 du temps de restitution sur le cas test de reconnaissance d'images Resnet-50, en apprentissage (avec MXNet) comme en inférence.

Ces nouvelles cartes présentent également la capacité à traiter des problèmes de plus grande dimension par rapport aux cartes P100, avec un saut de mémoire par GPU de 12 à 32 Go et une augmentation significative de la bande passante mémoire, de 549 Go/s à 900 Go/s.

Il sera possible d'exécuter des travaux multi-GPU (jusqu'à 4) en bénéficiant de l'interconnexion à 300 Go/s du NVLink (nouveau sur Myria).

La fiche complète décrivant le calculateur Myria après cette extension est disponible en annexe.

L'arrivée de ces matériels va mobiliser le support scientifique du Criann. En effet, bénéficier des performances annoncées requiert un niveau d'expertise élevé.

## 2.2. Logithèque de Myria pour l'IA (août 2019)

Après la mise à jour du système de Myria en août, et la montée de version de Cuda en version 10.0, les environnements de Deep Learning ont été également mis à jour.

Les utilisateurs ont accès aux piles logicielles décrites dans le tableau 1.

Outil	Python 3.6 (Cuda 9.0)	Python 3.6 (Cuda 10.0)
<i>Bibliothèque NVIDIA</i>		
<b>CuDNN</b>	7.1	7.2
<i>Environnements de Deep Learning</i>		
<b>Caffe</b>	1.0.0	Cf. pyTorch 1.2.0
<b>pyTorch</b>	1.0.1	1.2.0
<b>Keras</b>	2.2.4	2.2.5
<b>TensorFlow</b>	1.8.0	1.14
<i>Environnement de Machine Learning</i>		
<b>Scikit-learn</b>	0.20	0.21
<i>Bibliothèque de calcul parallèle MPI pour le DL</i>		
<b>Horovod</b>	0.16.1	0.18.0
<i>Bibliothèque pour le traitement d'images</i>		
<b>opencv</b>	3.4.0	4.1.0

Tableau 1 : piles logicielles IA-DL sur Myria (août 2019)

NB : les anciennes versions basées sur Python 2 restent accessibles mais il est conseillé d'utiliser les dernières versions pour une meilleure performance.

## 2.3. Environnements interactifs pour le Deep Learning

Suite aux demandes exprimées par les chercheurs du Litis et du Greyc, plusieurs environnements interactifs ont été mis en place sur Myria :

- TensorBoard
- Jupyter notebook et son successeur Jupyter Lab

Ces environnements facilitent la mise au point des modèles, pour les travaux de grande taille.

Pour un usage rationnel des ressources du supercalculateur, les utilisateurs seront invités à privilégier :

- l'utilisation de leur poste de travail local s'il s'agit de travaux de petite taille,
- l'utilisation du mode batch, sans interface graphique, pour les travaux de production.

Ces environnements interactifs ont été utilisés sur Myria lors de la journée d'initiation au Deep Learning (le 14 juin 2019), avec des ressources GPU réservées à cet effet.

Pour ne pas impacter les travaux de production, il a été décidé d'attendre la mise en service des matériels GPU de dernière génération (cf. 2.1) pour déployer le mode interactif plus largement en utilisant pour cela une partie des serveurs de la plus ancienne génération (K80).

Les modalités d'utilisation sont les suivantes :

- connexion depuis les locaux du laboratoire (mis en place pour le Litis et pour le Greyc).
- commandes `starttensorboard`, `jupyternb` et `jupyterlab` pour lancer les sessions interactives depuis Myria.

## 2.4. Files d'attente pour les travaux GPU, partage des ressources

Le précédent rapport faisait état d'un besoin d'amélioration du partage des ressources GPU, et du besoin d'accès rapide pour les travaux de calibrage des modèles.

En réponse à cette demande, le Criann a procédé à un ajustement des paramètres du logiciel de traitement par lots (seuls 2 travaux par utilisateur sont pris en compte à chaque affectation de ressources). Le nombre de ressources GPU par utilisateur n'a pas été limité pour le moment afin d'optimiser l'usage de la machine (le nombre d'utilisateurs GPU est encore relativement faible et leur activité présente généralement des pics, ponctuels mais d'intensité élevée). Cette situation pourra être revue si besoin.

Pour les travaux de calibrage, une file d'attente `gpu_court` limitée à 4 heures a été créée. Cette file d'attente s'appuie actuellement sur les mêmes machines que les autres files d'attente GPU.

Aucune évolution n'a été apportée concernant l'affichage des ressources disponibles (ce point avait été soulevé par les utilisateurs en 2018). L'information est accessible depuis les frontales de connexion, au travers des commandes appropriées, décrites dans la documentation (commandes du gestionnaire de travaux Slurm).

## 1.1. Support aux utilisateurs, formation des nouveaux utilisateurs

Jusqu'à présent, les nouveaux utilisateurs de la communauté étaient sensibilisés aux spécificités du calculateur lors de demi-journée « Deep Learning sur Myria » ou lors de rencontres avec le support scientifique.

Il s'avère qu'avec l'augmentation du nombre d'utilisateurs, l'arrivée de stagiaires « au fil de l'eau » dans les équipes, un effort doit être mené pour faciliter un usage optimal des ressources par les nouveaux arrivants. En effet, cet usage optimal requiert que les utilisateurs aient un minimum conscience de l'architecture matérielle et logicielle sous-jacente.

C'est pourquoi une documentation spécifique les débutants, sous forme de tutoriels, sera prochainement mise en place avec pour objectif pour répondre aux cas d'usage les plus fréquents.

## 2. L'environnement de test (« Déca »)

Mi-2018, le Criann a remis en service quelques nœuds de l'ancien calculateur Antares pour constituer une petite machine de test, baptisée « Déca ».

La machine Déca a permis en particulier de prototyper les outils interactifs décrits précédemment avant de les déployer en production sur Myria. Ce travail s'est effectué en collaboration avec les chercheurs et doctorants du Litis.

Suite à cette collaboration, l'équipe du Litis a demandé à ce qu'un nouveau serveur GPU, acquis sur les fonds propres du laboratoire, soit intégré à Déca plutôt que d'être installé de façon autonome, afin de bénéficier d'un environnement logiciel maintenu et d'un environnement de partage de ressources au travers du logiciel Slurm.

Dans une deuxième étape, la même équipe s'est rapprochée du Criann afin que le même environnement soit proposé aux élèves ingénieurs (cycle master) pour leurs Travaux Pratiques et projets de Deep Learning, sur du matériel en cours d'acquisition par l'Insa de Rouen. Une réunion entre l'équipe technique du Criann, la direction des services informatiques de l'Insa et les enseignants concernés a permis de préciser :

- Les caractéristiques techniques des matériels requises pour leur intégration au Criann ;
- La répartition des travaux d'exploitation entre les services du Criann et de l'Insa (création des comptes étudiants).

### 3. Travaux de veille technologique

Dans un contexte où les outils évoluent très vite, une veille technologique régulière est effectuée sur les outils pour Deep Learning en lien avec le HPC. L'outil Horovod, par exemple, a ainsi été testé puis proposé aux chercheurs, avant qu'ils en expriment la demande.

Au niveau national, Genci met en service actuellement des ressources spécifiques pour l'IA. Il s'agit d'une tranche de « Jean Zay », la nouvelle machine de l'Idris. Actuellement en test, elle sera ouverte aux utilisateurs en fin d'été 2019. La synergie entre centres nationaux et centres régionaux est essentielle pour faciliter le passage des chercheurs du centre régional vers les moyens de plus grande taille. Le Criann participe ainsi au groupe de travail Tier 1 – Tier 2 du projet Equip@meso, et est en particulier impliqué dans l'évaluation d'environnements interactifs pour l'IA, qui seront potentiellement déployés sur la machine nationale.

L'arrivée de l'IA (du Deep Learning) au Criann remonte à l'installation de Myria en 2017. La croissance de cette activité laisse augurer que lors du prochain renouvellement de ressources, un benchmark « IA » devra être ajouté au jeu de codes de test. Ce point fait également l'objet d'un travail spécifique, en lien avec un chercheur du Litis.

Plus généralement, c'est également dans ce contexte de convergence grandissante entre les usages du calcul intensif, du Big data et de l'intelligence artificielle que le plateau de calcul intensif du Criann se prépare à évoluer à l'horizon 2021. Les ressources matérielles et logicielles devront ainsi pouvoir être utilisées de manière optimales selon des modalités quelque peu différentes (traitement par lot, interactif) et bénéficier des ressources transversales au calculateur (réseau d'interconnexion, stockage, environnement logiciel).

### 4. Autres actions et évolutions internes

#### 4.1. Évolution dans l'administration de la logithèque

Le précédent rapport mentionnait la difficulté à administrer une logithèque IA complexe, avec un foisonnement de versions, et des versions « sur-étagère » non adaptée aux spécificités du calculateur.

La question de l'utilisation de conteneurs était alors une piste envisagée.

L'administration de la logithèque a finalement pu être rationalisée en s'appuyant sur l'environnement virtuel Python et en le couplant à l'outil module déjà en place. Cela permet de proposer différents environnements optimisés sans risque de conflits.

### 5. Datalab Normandie

Le Criann participe au projet « Datalab Normandie » et est impliqué dans l'élaboration de la plateforme de travail collaboratif pour le traitement de données (Deep Learning, traitements statistiques).

Il est prévu que l'environnement de travail mis à disposition par l'entreprise Saagie soit à terme être hébergé sur l'infrastructure du Criann.

## 6. Traitement de données sur Myria

### 6.1. Logithèque de Myria pour les traitements statistiques (août 2019)

Pour l'analyse des données, les utilisateurs ont bien sûr accès au logiciel R, ainsi qu'à son alternative pandas, une bibliothèque python. Les versions installées sont décrites dans le tableau 2.

<i>Logiciels et environnements pour l'analyse de données</i>	
<b>Panda</b>	Version 0.24.2
<b>R</b>	Versions 3.3.2 – 3.5.1 – 3.5.2

*Tableau 2 : piles logicielles statistiques sur Myria (août 2019)*