

A large, curved view of the Earth from space, showing blue oceans and white clouds, occupies the top-left corner of the slide.

Infrastructure, organisation et cas d'usage autour de l'IA au CNES

Guillaume Eynard-Bontemps JDEV 2020

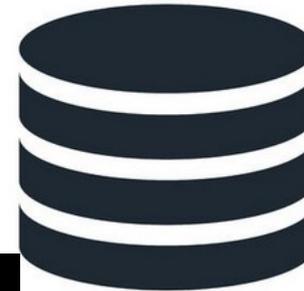
CNES Computing Center

- ❖ User oriented service offer
- ❖ Component & Services
 - **Computing platform** (currently HPC), with 12K cores and 8,5PB of storage
 - Storage platform : Datalake, currently in development, only 16PB of tapes
 - Services & applications:
 - Software modules,
 - Gitlab and associated CI,
 - Pangeo/Jupyterhub,
 - Containers (Singularity),
 - **Artificial Intelligence (GPUs and Python 😊)**
 - User support and expertise (scientific computing, big data processing, AI)



Computing Platform

Support / Expertise



Datalake

Services & applications



HAL

- CPU : 500 Tflops CPU
- GPU : 240 Tflops GPGPU / 2 Pflops Deep Learning
- 480 batch servers / 12000 cores
- 8 interactive servers pre/post processing w/ GPU
- 8.5 PB GPFS / 300TB burst buffer/ 100GBs bandwidth

Ethernet 1 Gbs / 10 Gbs



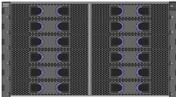
Admin nodes
2x X3550 M5



Login Nodes
2x X3550 M5



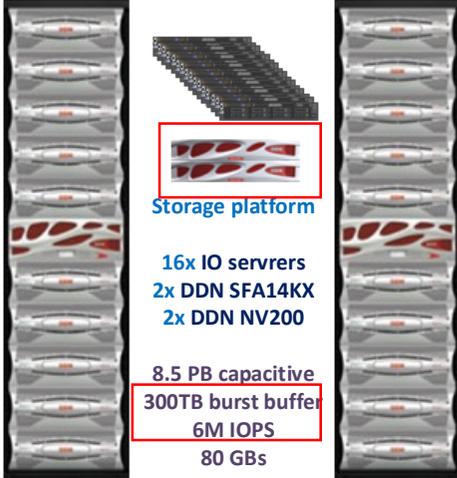
Development Nodes
6x SR650 + GPU Quadro
336 cores
1.5TB RAM



CPU compute nodes
276x NX360 M5
100x SD530
104x C6220
12000 cores
60 TB RAM



GPU compute nodes
2x quad T4
4x quad V100
1x octo v100
28 GPU
832 GB RAM



Storage platform
16x IO servers
2x DDN SFA14KX
2x DDN NV200
8.5 PB capacitive
300TB burst buffer
6M IOPS
80 GBs

Infiniband 56 Gbs / 100 Gbs

GPU and IA Softwares

- ❖ GPU addressed through our job scheduler PBS
- ❖ Cuda drivers, several versions on GPU nodes
- ❖ Cuda library as modules
- ❖ Singularity images ready to use or extend
- ❖ Python environments ready to use (Conda)
- ❖ Wiki and tutorials as Notebooks
- ❖ Some GPUs usable from our Jupyterhub

Spawner Options

Select a job profile:

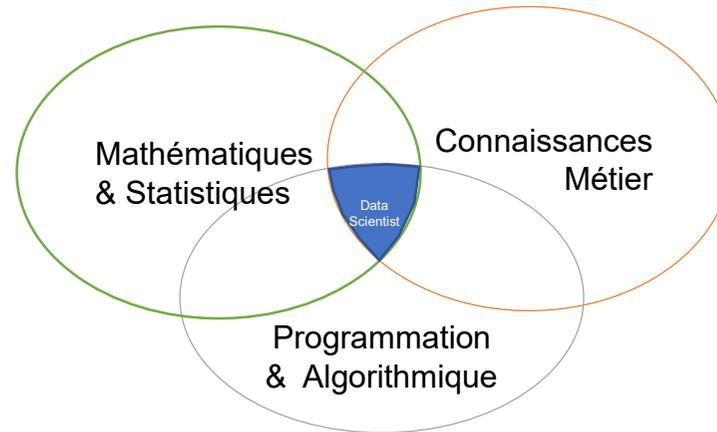
Qdev - 1 cores, 4 GB, 12 hours -- Default to use
Qdev - 1 cores, 4 GB, 12 hours -- Default to use
Qdev - 4 cores, 15 GB, 12 hours
Qdev full node - 16 cores, 60GB
Batch - 1 cores, 5 GB, 12 hours
Batch - 1 cores, 5 GB, 72 hours
Batch - 4 cores, 20 GB, 12 hours
Batch full node - 24 cores, 120 GB, 12 hours
Batch 2019 full node - 40 cores, 184 GB, 12 hours
GPGPU - 1 gpgpu T4 -- Default to use for GPU, 9 cores, 94 GB, 10 hours
GPGPU - 1 gpgpu V100, 6 cores, 94 GB, 10 hours

Coordination IA

- **Fédérer les acteurs CNES liés à l'IA**
- **Améliorer la communication entre les métiers**
- **Améliorer la formation (Manager, Ingénieur, Expert, MOOC, Mini-formation interne)**
- **Coordonner les animations COMET sur le sujet**
- **Proposer, mettre en place et maintenir les outils et compétences nécessaires à l'utilisation de l'IA (GPU, CONDA, Jupyter...)**
- **Echanger sur les projets internes et de nos partenaires.**

Cellule Support en AI, Qui?

- **5 personnes formées à DNO/ISA/CID : Fabien Meslet-Millet, Felipe Alves Suana, Christophe Taillan, Orphée Faucoz, Denis Standarovski.**



- **L'équipe de support du centre de calcul en particulier pour l'utilisation de l'infra et des logiciels**
- **La possibilité d'activer ponctuellement des membres de la coordination d'autres directions du CNES pour des sujets spécifiques en fonction de leur disponibilité**

Cellule Support en AI, Pourquoi?

- **Apporter le support aux métiers du CNES souhaitant utiliser l'IA**
- **Accompagnement la montée en compétence**
- **2020: ouverture du support aux laboratoires.**

Cellule Support en AI, Comment?

- Aide au choix de l'outillage (ex: Concours CESBIO, Sauvegarde vol au CSG)
- Aide au choix des modèles (ex: Sauvegarde vol au CSG)
- Optimisation de modèles (ex: Mesures aberrantes en altimétrie)
- Prototypage (ex: Orbitographie précise, ballons...)
- Choix/Accompagnement des R&T et Thèses
- **Accompagnement au passage en production**
- **Former**

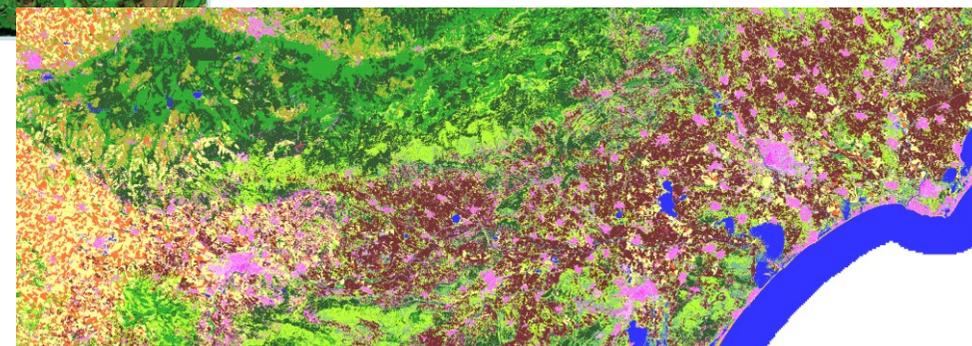
Domaines d'application

- **Extraction d'information**
- **Détection d'anomalies**
- **Prédiction**
- **Simplification de modèles physiques**
- **Autonomie**

Deep Learning : Land cover map

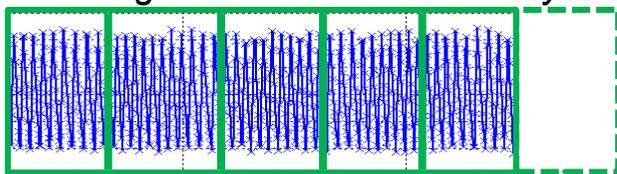


❖ Comparaison à la chaine Iota2 du CESBIO

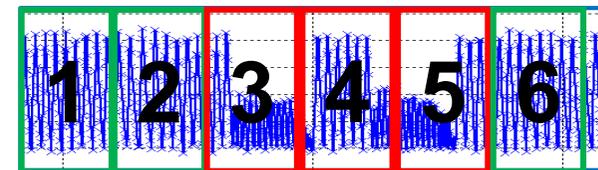


NOSTRADAMUS: faults detection

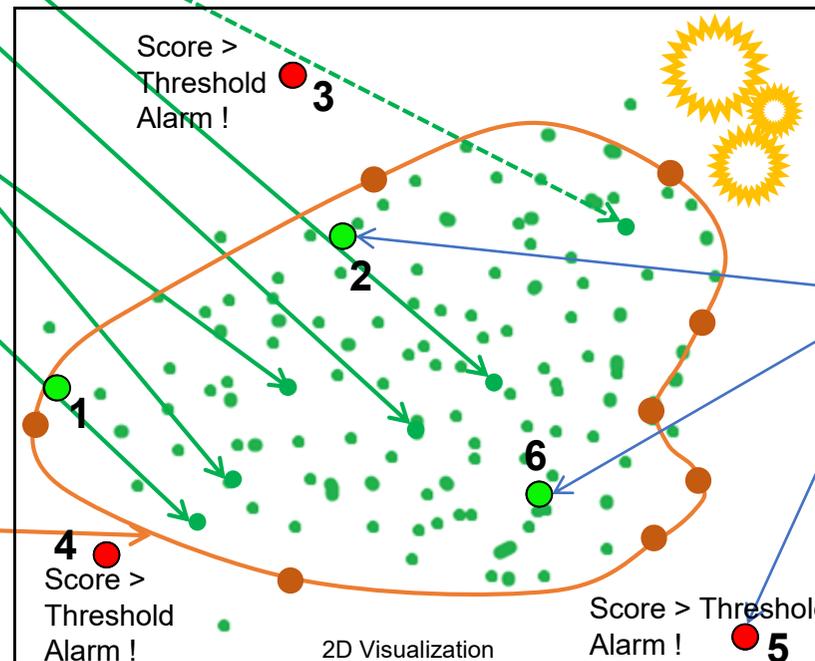
Training data: nominal Telemetry



Telemetry to check



$$X_i = \begin{pmatrix} \min \\ \text{avg} \\ \max \\ \text{Standard deviation} \\ \dots \end{pmatrix}$$



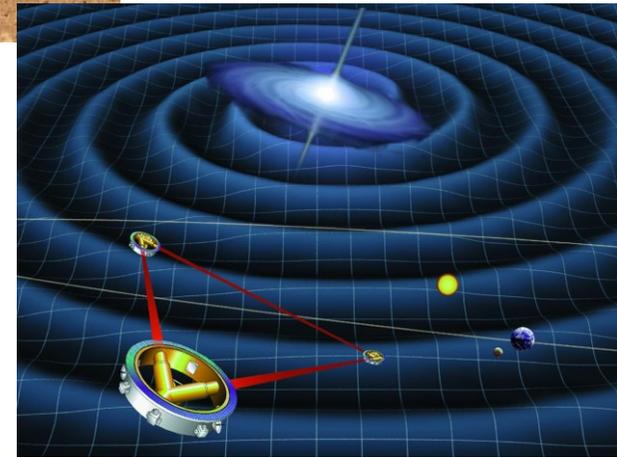
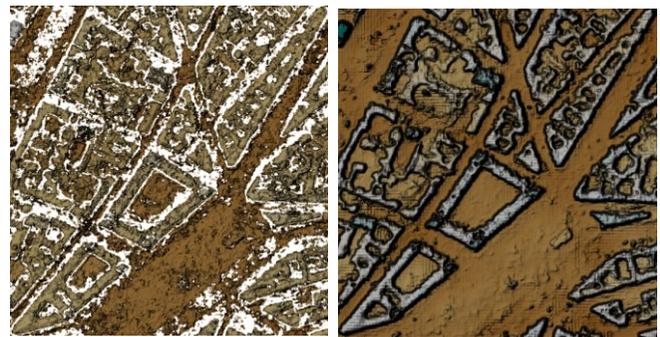
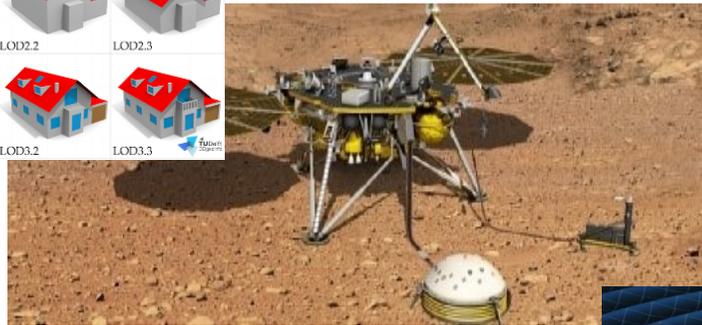
Decision boundary

Tested Vectors

Supports aux projets

- AI4GEO
- EOLAB
- LISA
- Sismoc
- SpaceShip

	LOD x.0	LOD x.1	LOD x.2	LOD x.3
LOD0				
LOD1				
LOD2				
LOD3				



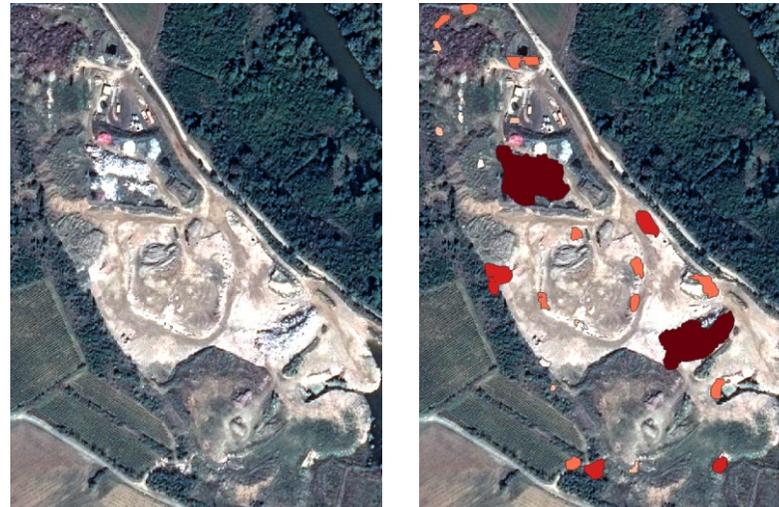
Décharges illégales (ex. Toulouse)

- **Analyse de l'impact environnemental:**

- Couplage des détections avec des bases de données géographiques

- Calcul de distance à des lieux d'habitation, d'écoles...

- Utilisation de modèle numérique de terrain pour identifier les écoulement pouvant générer de la pollution



EchoFinder

- Avec le CADMOS
- Echographe capable de faire le tracking en temps réel de 5 organes (Carotide, Nerf Optique, veine porte, jugulaire et cœur)



MAGELLAN – Echographe pour COVID-19

- Avec le CHU de toulouse
- Détecter, compter, quantifié les facteurs d'atteintes du poumons
- Prédire l'évolution de la maladie

