



La plateforme SCIGNE : Une équipe et des services pour le *Machine Learning*

Sébastien GEIGER et Jérôme PANSANEL

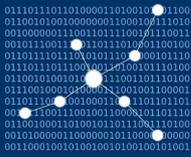
Journée Machine Learning – 1^{er} octobre 2020



```
011011101101000011010010 01100
01100101001000000011000100111010
001000000110011011100101110011
00101110011 0101101001100100
011011101110010110 000101110
0110111011000000000101110100
01100101001010 11001101110100
0111001000100000000001100001
01101110 00100011 1101101101
00 110011100100110000010110110
0110010001101001011011100110100
0010100000110000001011000 0000
00110001001010010010100100101001
```

Sommaire

- **Présentation de la plateforme SCIGNE**
- **Exemple de réalisation :
classification et segmentation de séries
temporelles**



La plateforme SCIGNE

En quelques mots

- Services pour l'analyse et le stockage des données scientifiques
- Issue de l'évolution de la grille de calcul et du Cloud Computing
- Labellisée par l'IN2P3 en 2017
- 8 ingénieurs / 3,3 FTE
- Calcul HTC (anciennement grille de calcul)
- Cloud IaaS / conteneur as a service
- Gestion et archivage de données
- Disponibilité > 99 %
- Sources variées d'investissement (laboratoire, IN2P3, France Grilles, CPER, IFB, H2020)

→ <https://www.grand-est.fr>

Expertises

- Formation et rédaction de documentation
- Technologies (archivage, grille & Cloud, stockage distribué, ...)
- Gestion des *workflows* de calcul et de données
- FAIRisation et mise en ligne des données
- Développement GPU et parallélisation des calculs
- Sécurité des systèmes et des applications
- Conseil acquisition
- Recrutement pour fin 2020 d'un ingénieur dédié à l'IA / ML (EOSC-Pillar)

Cloud computing

- Service de Cloud Computing basé sur OpenStack
- Fédéré avec les Clouds France Grilles, IFB et EGI
- > 500 coeurs disponibles
- Gabarit de serveurs jusqu'à 48 coeurs et 512 Go de RAM
- 500 To de stockage disponible basé sur CEPH (stockage objet)
- Réseau dédié et isolé
- 1 GPU V100 disponible
- 8 GPUs RTX 2080Ti en cours d'intégration
- Nouvelle acquisition de GPUs en cours d'acquisition

```
0110111011010000110100101100
0110010100100000001100010111010
00100000011001101110011110011
00101110011101110101100100100
0110111101110101110000101110
01101110110000100100101110100
011001010010111001101110100
0111001000100010001001100001
01101110001000111101101101
0110011001100110000101101110
0110010001101001011011101100
00101000001100000010110000000
00110001001010010010100100101001
```

Gestion des données

Gestion des données de recherche (iRODS, CEPH)

- 500 To distribués
- Disponibilité > 99 %
- Basé sur CEPH
- Utilisation par OpenStack et iRODS
- Intégration avec FG-iRODS
- Implémentation de DMP

0111011101101000011010010
0110010100100000001100010111010
00100000011001101110011110011
00101110011101110101100100100
0110111101110101110000101110
011101110110000100100101110100
011001010010111001101110100
01110100010000000001100001
011011100000011101101101
0110011001100110000101101110
011001000110100101101110100
0010100000110000001110000000
00110001001010010010100100101001

Infrastructures partenaires

Accompagnement vers d'autres centres

- Ferme GPU du CC-IN2P3 :
<https://doc.cc.in2p3.fr/fr/Computing/job-types/job-gpu.html>
- Mésocentre de l'Université de Strasbourg :
<https://hpc.unistra.fr>
- Centre de calcul ROMEO :
<https://romeo.univ-reims.fr/>
- Calculateur Jean Zay :
<http://www.idris.fr/jean-zay>

- **Présentation de la plateforme SCIGNE**
- **Exemple de réalisation :
classification et segmentation de séries
temporelles**

011011101101000011010010 01100
0110010100100000001100010111010
00100000011001011100101110011
00101110011 01011010101100100
0110111011101010110 000101110
01101110110000100100101110100
01100101001010 11001101110100
01110100010000100010001100001
01101110 00100011 1101101101
00 11001100100110000101101110
011001000110100101101101110100
001010000011000000111000 0000
00110001001010010010100100101001

Merci de votre attention !