

# **ComptStat / IN2P3 Machine Learning**

**Networking**



**David Rousseau**

**LAL-Orsay**

**[rousseau@lal.in2p3.fr](mailto:rousseau@lal.in2p3.fr)**

**IN2P3 R&D DATA, 29 janvier 2018**

# Buts



## □ Buts:

- Développer les collaborations entre labos/expériences sur des activités ML à l'IN2P3
- Partage d'expériences
- Documentation/ Tutoriaux

□ MACHINE-LEARNING-L@IN2P3.FR : 65 participants

□ MACHINE-LEARNING-CORE-TEAM-L@IN2P3.FR : Balazs Kegl, David Rousseau, Eric Aubourg, Françoise Bouvet, Emille Ishida, Emmanuel Gangler, Jérôme Pansanel, Vava Gligorov Thomas Vuillaume

□ Pas d'ETP spécifique pour l'instant

□ Pas encore de site web

□ One day workshop at CC-Lyon Jeudi 29 mars 2018

# Partage d'expérience



- **Activité Machine Learning en cours à l'IN2P3**
  - (ne compte pas l'utilisation standard de BDT, qui est toujours la technique recommandée pour classification sur une douzaine de variables)
  - generative models ATLAS calorimeter or LSST
  - anomaly detection LSST or ATLAS
  - Active Learning (LSST)
  - Active Learning for "intelligent" simulation
  - fast tracking ATLAS LHCb
  - deblending with deep learning LSST
  - KM3net event reconstruction/identification
  - CTA event reconstruction/identification
  - Reconstruction de camera imagerie beta/gamma (medical)
  - system administration learning from log files or other information
  
- **Pas (encore) de projet commun identifié (sauf TrackML)**

# Plateforme GPU



- ❑ Pour deep learning accès à plateforme GPU indispensable (libraries Keras+Tensorflow, PyTorch savent utiliser un ou plusieurs GPU, en principe!)
- ❑ Recenser ce qui est dispo à l'IN2P3 (e.g. CC-Lyon, GridCL@GRIF)
- ❑ Recettes d'utilisation
- ❑ Benchmarks HEP pertinents
- ❑ Optimisation (relation avec Decalog)
- ❑ Feedback aux mainteneurs des plateformes

# Tutoriaux



- ❑ « Je veux démarrer avec (par exemple) Deep Learning, qu'est ce que je fais ? »
- ❑ « J'ai mon NN, super, comment je l'utilise en prod ? »
- ❑ Beaucoup de tutoriaux ML dispo en ligne
- ❑ Très peu orientés physique des hautes énergie
- ❑ Objectif
  - Recenser ce qui existe (exemple [SWAN Machine Learning @ CERN, IML @ CERN](#))
  - Créer les tutoriaux qui manquent
  - Jupyter notebook interface de choix
  - → plateforme commune avec REPRISES ?
- ❑ Involvement in 2018 IN2P3 School of Statistics, 28 mai au 1 juin 2018 La Londe les Maures <https://indico.in2p3.fr/event/16931/>
  - Ramped down Root/TMVA, Introduced hands-on with sklearn, keras+tensorflow

# TrackML



- LAL , LPNHE + int'al partners (CERN, Berkeley, Caltech, U Geneve,...)
- Tracking at HL-LHC will be difficult with pile-up reaching 200 compared to  $\sim 30$  today
- Driving the CPU resource for event reconstruction  $\Rightarrow$  x10 speed up looked for
- Large optimisation effort of "classic" algorithms (cf decalog) but looking for new ideas on ML side
- TrackML challenge
  - generic simulation of a LHC silicon detector (ACTS, also relevant to Decalog)
  - release dataset  $1^E5$  event x  $1^E4$  tracks x 10 3D points
  - challenge is to associate the hits into tracks
  - simple 2D version ran as hackathon in March 2017
  - See details [talk CTDWIT 2017 @ LAL](#)
  - full challenge on Kaggle will run from  $\sim$ March 2018 : Accuracy Phase March-May 2018 (high efficiency/low fake rate/no incentive on CPU time), Throughput phase July-Sep 2018 (speed maintaining accuracy)

