Big Data, Deep Learning and OpenPOWER Foundation

OAS - Days 2018

N. Parmiggiani, A. Bulgarelli

Computing Science Challenges for Physical Science

From **“OpenPOWER Foundation for Physical Science Workgroup charter”**:

- Today **the scientific community** is facing an **enormous increase in data volume, rate and dimensionality**
- The **reason** for this explosion of data is in the advance of **science** itself.
- **Processing** of the huge amount of data generated by experiments and simulations, and the complex and hard-to-discover relationships inside them, become a **challenge**.

Authors:

Andrea Bulgarelli, INAF
Luca Graziani, INAF
Raffaella Schneider, INAF
Ugo Becciani, INAF
Valentina Fioretti, INAF
Adriano De Rosa, INAF
Lionel Clavien, Groupe T2i
Cecilia Carniel, IBM

Cross-cutting technologies

- OpenPOWER Foundation for Physical Science workgroup + collaboration with CERN
- Big Data
- Deep Learning

OpenPOWER Foundation (OPF) for Physical Science WG (OPF-PS)

- A workgroup focused on establishing an interface between ICT industries and the Physical Science community (around the Power CPUs on Linux platforms).
- **Build a network of relationships**: e.g. CERN/OpenLAB
  - master thesis on Geant 4
- Leaded by A. Bulgarelli (member of the OpenPOWER Foundation Technical Steering Committee)

Main OPF industrial members (160+). Between them: IBM, NVIDIA, Google, Mellanox, Hitachi, Linux distributions, Samsung, Xilinx, ...

Main academic members (120+). Between them: INAF, Caltech, Univ. of Illinois, CINECA, Rice Univ., OAK Ridge Nat. Lab., Lawrence Livermore Nat. Lab., ...

Current OPFPS members (13+ companies, 28 people): INAF, CINECA, IBM, Julich supercomputing Center, E4, european universities, ISTO, ...

Why Big Data?

- Big Data: data that can't be managed with usual technologies
- Increase Telescope size and number lead to Big Data
- Key technology for future observatories. Some examples: SKA and CTA

Archive Data: 700 TB/y
Data Stream: 1 GB/s

Raw Data: 1000 PB/y
Archive Data: 3 PB/y
Data Stream: 1-5 GB/s

Raw Data: 16 TB/s
Archive Data: 600 PB/y
Analysis Stream: 5 TB/s

Tools and Technologies

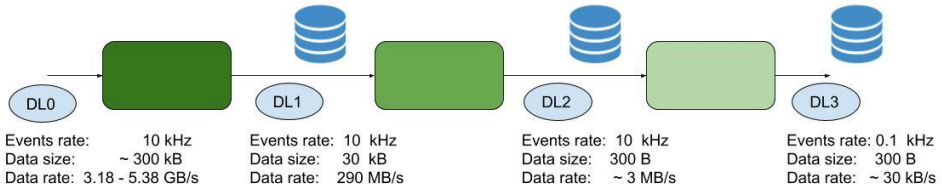
Relational Database

Cluster Computing Framework

In-Memory non relational database

Big Data for LST1 (and AGILE)

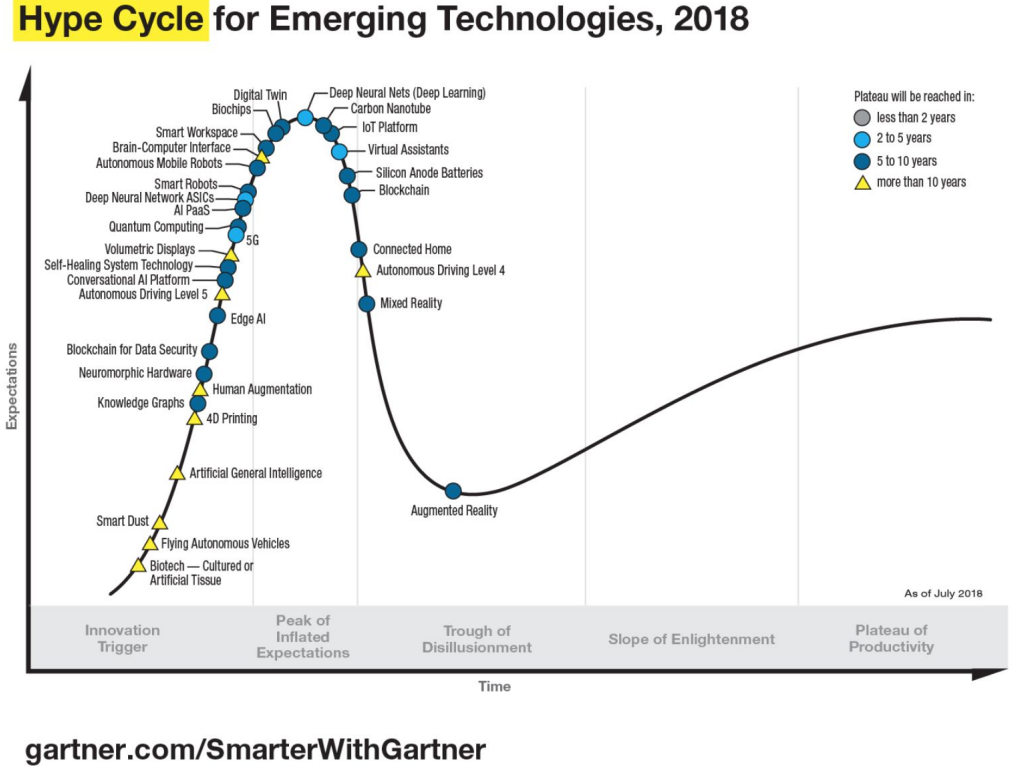
- We are using key Big Data technologies to manage the LST1 Data



- Our software can insert LST1 data inside a Redis or MySQL database satisfying DL1-DL3 requirement
- The library can also push data through a communication channel for the Graphical User Interface
- Big Data for AGILE: To analyze 1 year of data with current tools requires 2 months. We are studying the use of Big Data technologies to analyze AGILE data and reduce the analysis time
- Available resources:
  - PhD on Big Data (N. Parmiggiani)
  - Master Thesis Student is testing this library (G. Zollino)

Why Deep Learning?

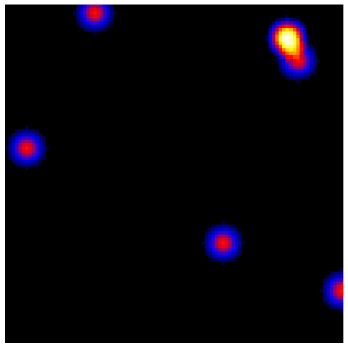
- Models based on artificials neural networks, class of machine learning algorithms
- Deep Learning is the new field of the Machine Learning
- It requires dedicated (and costly) hardware: GPU -> we have used a dedicated IBM machine (for free) for 2 years
- Our purpose: to understand if this new technologies can resolve our problems in a more effective way.
- Resources: 1 PhD on machine learning (L. Baroncelli)



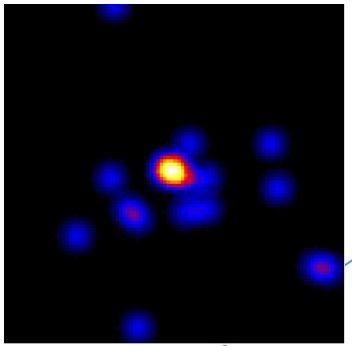
Deep Learning for AGILE

- Starting from a physical model based on a list of GRB seen by AGILE and Fermi we simulated 400k maps for training and test
- The maps are simulated with a Monte Carlo simulator and a Gaussian smoothing is applied before start the classification.
- After few minutes of training the accuracy reached on the test dataset is 0.974%.
- We simulated 10M maps to calculate the false alarm rate. Evaluating this algorithm on real AGILE GRBs we got detections about 5-6 sigma.

Background



Source



Conclusions

- Application of leading ICT technologies to high-energy astrophysical projects (AGILE, CTA, ASTROGAM)
- In connection with
  - the market (OpenPOWER Foundation)
  - international research center (e.g. CERN/OpenLAB)
  - Software engineering departments of italian universities
    - University of Bologna
    - University of Modena
    - University of Padova
- Master thesis and PhDs in progress. It is important to establish a more formal connection with these Software engineering departments