Miguel Villaplana

- ATLAS - Milano
- Supervisor: Laura Perini

ATLAS EventIndex (in collaboration with Dario Barberis)

- **A database with the references to the files including each event in every stage of processing**
  - fast and efficient selection of events of interest, based on various criteria, from the billions of events recorded
  - an indexing system that points to those events in millions of files scattered in a world-wide distributed computing system
  - contains records of all events processed by ATLAS, in all processing stages

- **Contribution to Functional Tests and User support**
  - Based on previous work from A. Favareto: bash script
  - Redesigned from scratch using python
    - modular design, exception handling, code documentation
  - Added new functionality
    - test MC samples
    - resubmission of failed tests

- **Status and plans**
  - Short term: machinery is mostly in place but needs polishing
  - Long term: design and implementation of the functional tests for the Event Whiteboard (EI's evolution)

Harvester (in collaboration with Alessandro De Salvo)

- The current production and distributed analysis system in ATLAS (PanDA) relies on a server-pilot paradigm
  - A server maintains state and manages workflows with various granularities
  - Pilots are job-centric and run independently on worker nodes with a limited view of local resources

- PanDA itself has no means of managing and monitoring cloud utilisation

- **Harvester is a resource-facing service between the PanDA server and the collection of pilots**
  - It is a stateless service with knowledge of the resources
  - It can act as an intermediate communication channel between PanDA server and pilots

- **Current activity**:
  - Our goal is to improve site deep resource knowledge
  - Start by using the available information sources (GLUE 1.2, GLUE 2)
    - Understand if the available information sources could be reliable and useful w.r.t. what we currently have in Panda
      - Compare the values obtained from Glue 2 and what we have in Panda
      - Many problems found: misconfigured sites, only a fraction (< 40%) of sites/CEs have matching values between Glue2 and Panda
    - In summary, GLUE values do not really seem reliable enough

- **Future plans:** grabbing values directly from jobs
  - Initial prototype of a collector designed by A. De Salvo
  - Plan to contribute to the development and testing of the prototype

Machine learning

- **H -> Tau Tau mass reconstruction (in collaboration with Attilio Andreazza)**
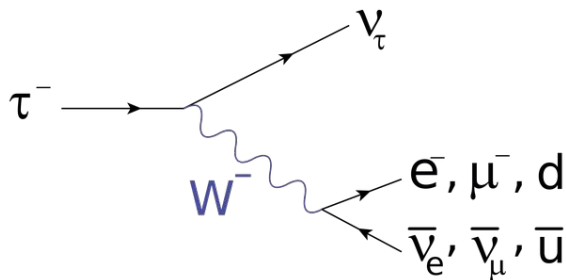
  - Tesi di laurea di Aldo Materassi

  - **Goal**: predicting the invariant mass of ditau system using the visible tau decay products kinematics
    - Difficult final state: many neutrinos

  - **Status and plans:**
    - Now learning basics about ATLAS software and how to get/process the data
    - We will first reproduce previous results
      - using BDTs (and random forests)
      - can we train at truth level and test at reco level?
      - best way to use ML frameworks in the market (scikit-learn, pyTorch,...) inside ATLAS software
    - Try other ML techniques

Share resources across groups at UniMi

- **Groups are encouraged to organise their resources under HTCondor pools**
  - Execute machines report to the central manager of their own pool

- **We add an additional central manager to which all execute machines report too**
  - This provides usage accounting across all the resources together
  - Serves as a top-level pool to submit jobs to when users want to access all possible resources

- **Users get the quality of service they were already enjoying, but excess jobs may be conveniently sent to the other resources**
  - Group pools remain the default pool for job submission, but with the super-pool added to their FLOCK_TO list
  - We give the group's negotiator priority over super-pool's to guarantee high priority to group users on their own machines

- **Status and plans**
  - The system is in place and shows good behaviour
  - ATLAS use-case:
    - Trivially parallel jobs. CVMFS is the only requirement.
  - Currently researching how to run MPI jobs encapsulated in Docker containers via HTCondor
    - Tesi di laurea triennale di Massimo Miserendino (sup. F. Prelz, D. Rebatto and A. Andreazza)
      http://infn.it/thesis/thesis_dettaglio.php?tid=12211
  - Monitoring and accounting
    - Based on Filebeat+Elasticsearch+Kibana
  - Documentation and user support
    - Scripts and conf files in GitLab @ INFN (baltig.infn.it)
    - Instructions for sys. andmins growing both in GitLab and in a local twiki
    - User support: twiki + mailing lists for support and announcements in construction